Domain Decomposition Methods in Science and Engineering

Thirteenth International Conference on Domain Decomposition Methods

Lyon, France

Edited by: N. Debit
M.Garbey
R. Hoppe
J. Périaux
D. Keyes
Y. Kuznetsov

Published by ddm.org

Domain Decomposition Methods in Science and Engineering

Thirteenth International Conference on Domain Decomposition Methods

Lyon, France

Domain Decomposition Methods in Science and Engineering

Thirteenth International Conference on Domain Decomposition Methods, Lyon, France

Edited by

N. Debit
Lyon, France

M.Garbey
Houston, USA

R. Hoppe
Augsburg, Germany

J. Périaux
Paris, France

D. Keyes
Norfolk, USA

Y. Kuznetsov
Houston, France

Published by ddm.org

Visit our Home Page at http://www.ddm.org

Produced from LaTeXmanuscripts submitted by the authors.

Printed and bound in France.

ISBN xxx

Preface

This volume captures 53 of 100 the presentations of the Thirteenth International Conference on Domain Decomposition Methods, which was hosted by the University of Lyon in the Champfleuri Conference Center in the Province of Rhone-Alps, France, October 9-12, 2000. Approximately 117 mathematicians, engineers, physical scientists, and computer scientists from 22 countries came to this nearly annual gathering.

Since three parallel sessions were employed at the conference in order to accommodate as many presenters as possible, attendees and non-attendees alike may turn to this volume to keep up with the diversity of subject matter that the topical umbrella of "domain decomposition" inspires throughout the community. The interest of so many authors in meeting the editorial demands of this proceedings volume demonstrates that the common thread of domain decomposition continues to justify a regular meeting. "Divide and conquer" may be the most basic of algorithmic paradigms, but theoreticians and practitioners alike are still seeking - and finding - incrementally more effective forms, and value the interdisciplinary forum provided by this proceedings series.

Besides inspiring elegant theory, domain decomposition methodology satisfies the architectural imperatives of high-performance computers better than methods operating only on the finest scale of the discretization and over the global data set. These imperatives include: concurrency on the scale of the number of available processors, spatial data locality, temporal data locality, reasonably small communication-to-computation ratios, and reasonably infrequent process synchronization (measured by the number of useful floating-point operations performed between synchronizations). Spatial data locality refers to the proximity of the addresses of successively used elements, and temporal data locality refers to the proximity in time of successive references to a given element. Spatial and temporal locality are both enhanced when a large computation based on nearest-neighbor updates is processed in contiguous blocks. On cache-based computers, subdomain blocks may be tuned for workingset sizes that reside in cache. On message-passing or cache-coherent nonuniform memory access (cc-NUMA) parallel computers, the concentration of gridpoint-oriented computations - proportional to subdomain volume - between external stencil edge-oriented communications - proportional to subdomain surface area, combined with a synchronization frequency of at most once per volume computation, gives domain decomposition excellent parallel scalability on a per iteration basis, over a range of problem size and concurrency. In view of these important architectural advantages for domain decomposition methods, it is fortunate, indeed, that mathematicians studied the convergence behavior aspects of the subject in advance of the wide availability of these cost-effective architectures, and showed how to endow domain decomposition iterative methods with algorithmic scalability, as well.

Domain decomposition has proved to be an ideal paradigm not only for execution on advanced architecture computers, but also for the development of reusable, portable software. Since the most complex operation in a Schwarz-type domain decomposition iterative method - the application of the preconditioner - is logically equivalent in each subdomain to a conventional preconditioner applied to the global domain, software developed for the global problem can readily be adapted to the local problem, instantly presenting lots of "legacy" scientific code to be harvested for parallel implementations. Furthermore, since the majority of data

sharing between subdomains in domain decomposition codes occurs in two archetypal communication operations - ghost point updates in overlapping zones between neighboring subdomains, and global reduction operations, as in forming an inner product - domain decomposition methods map readily onto optimized, standardized message-passing environments, such as MPI.

Finally, it should be noted that domain decomposition is often a natural paradigm for the modeling community. Physical systems are often decomposed into two or more contiguous subdomains based on phenomenological considerations, such as the importance or neglibility of viscosity or reactivity, or any other feature, and the subdomains are discretized accordingly, as independent tasks. This physically-based domain decomposition may be mirrored in the software engineering of the corresponding code, and leads to threads of execution that operate on contiguous subdomain blocks, which can either be further subdivided or aggregated to fit the granularity of an available parallel computer, and have the correct topological and mathematical characteristics for scalability.

Organizing the contents of an interdisciplinary proceedings is an interesting job, and our decisions will inevitably surprise a few authors, though we hope without causing offense. It is often difficult to assign a paper to just one of the categories of theory, algorithms, and applications. Readers are encouraged not to take the primary divisions very seriously, but to trace all the connections.

These proceedings will be of interest to mathematicians, computer scientists, and computational scientists, so we project its contents onto some relevant classification schemes below.

American Mathematical Society (AMS) 1991 subject classifications include:

Optimal control
Numerical simulation, modeling
Iterative methods for linear systems
Multigrid methods, domain decomposition for IVPs
Finite elements, Rayleigh-Ritz and Galerkin methods, finite elements
Spectral, collocation and related methods
Multigrid methods, domain decomposition for BVPs
Integral equations
Parallel computation
Mathematical software

Association for Computing Machinery (ACM) 1998 subject classifications include:

Programming environments, reusable libraries
Analysis and complexity of numerical algorithms
Numerical linear algebra, optimization, differential equations
Mathematical software, parallel implementations, portability
Applications in physical sciences and engineering

Applications for which domain decomposition methods have been specialized in this proceedings include:

Stokes, Euler, Navier-Stokes, multiphase flow, reacting flow
Porous media, atmospheric transport
Phase change, free surface phenomena
Semiconductor device physics
Linear and nonlinear elasticity
Acoustics, electromagnetics

The Neumann-Neumann method - a substructuring preconditioner typically employing

Additive Schwarz on the resulting interface problem - remains a topic of theoretical development and diverse applications [Giraud *et al.*, Alart *et al.*, Pavarino & Widlund], as odes the related Finite Element Tearing and Interconnection (FETI) method [Brenner, Dostal *et al.*]. Primal-dual formulations of FETI were heavily featured in the twelfth symposium in Chiba; primal-dual formulations emerge in further contexts in this proceedings [Klawonn & Widlund, Hoppe *et al.*].

Mortar methods, a nonoverlapping form of domain decomposition permitting flexibility in the form of nonmatching grids, were also a very active area in the Chiba symposium and continue to draw attention [Bjørstad *et al.*, Braess & Dahmen, Oswald & Wohlmuth, Shyy *et al.*, Tai *et al.*]. Another active area in nonoverlapping domain decomposition that is closely tied to the discretization is the optimal parametrization of Robin interface conditions [Bounaim, Gander, Gander *et al.*, Faille *et al.*, Dolean *et al.*, Rapin & Lube, Knopp *et al.*]. Related interface developments are presented under the rubric of optimal control and virtual control [Gervasio *et al.*, Pironneau *et al.*].

Overlapping domain decomposition methods continue to be refined, as well. This volume features two papers that shore up the highly effective Restricted Additive Schwarz (RAS) method. One [Cai *et al.*] shows how RAS, with its asymmetrical communication-saving restriction and extension operators can be rendered symmetric in an appropriate subspace and produces new theoretical bounds that mirror its observed superiority with respect to standard Additive Schwarz. The other [Frommer *et al.*] adopts a purely algebraic approach of oblique projections to produce the same ranking of additive Schwarz variants over the class of M-matrices, and also considers a restricted multiplicative Schwarz.

Two papers on the Aitken-Schwarz method introduced in Chiba [Baranger *et al.*, Garbey *et al.*] extend this overlapping technique, whose analysis depends upon Fourier decomposition of interface modes to nonlinear problems and less regular meshes. Meanwhile, nonlinear Additive Schwarz preconditioning [Cai *et al.*] has been applied to problems with shocks and has been shown to greatly improve the domain of convergence of Newton's method.

A novel purely algebraic method known as "multigraph", providing an algorithmic "spectrum" between exact Gaussian elimination and blocked iteration is presented in [Bank & Smith]. At an opposite extreme, waveform relaxation, a method that avoids forming discrete algebraic problems at common intermediate timesteps is advocated in [Daoud & Gander].

The implications for domain decomposition of several discretization techniques, apart from the customary conforming finite element and finite difference techniques on a single partitioned grid, are taken up by various authors. We mention especially fictitious domain methods [Feng & Karakashian, Lasser & Toselli], spectral methods [Azaiez *et al.*], and the increasingly theoretically supported discretization technique of finite volumes [Cautres *et al.*]. Apart from these methods rooted in differential equation formulations, there is a paper on domain decomposition for integral equation-based boundary element methods [Boubendir & Bendali].

These highlighted contributions only begin to call attention to technical points of interest in the current proceedings. We also note, sadly, a point of personal interest to all applied mathematicians, whether working in domain decomposition or not: this proceedings contains two of the last contributions of Jacques-Louis Lions.

For the convenience of readers coming recently into the subject of domain decomposition methods, a bibliography of previous proceedings is provided below, along with some major recent review articles and related special interest volumes. This list will inevitably be found embarrassingly incomplete. (No attempt has been made to supplement this list with the larger

and closely related literature of multigrid and general iterative methods, except for the books by Hackbusch and Saad, which have significant domain decomposition components.)

References

Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors. *Domain Decomposition Methods*, Philadelphia, PA, 1989. SIAM. Proceedings of the Second International Symposium on Domain Decomposition Methods, Los Angeles, California, January 14 - 16, 1988.

Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors. *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*, Philadelphia, PA, 1990. SIAM.

Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors. *Domain Decomposition Methods*, Bergen, 2001. ddm.org. Proceedings of the Twelfth International Conference on Domain Decomposition Methods, Chiba, 2000.

Charbel Farhat and François-Xavier Roux. Implicit parallel processing in structural mechanics. In J. Tinsley Oden, editor, *Computational Mechanics Advances*, volume 2 (1), pages 1–124. North-Holland, 1994.

R. Glowinski, J. Périaux, and Z. Shi, editors. *Domain Decomposition Methods in Sciences and Engineering*. John Wiley & Sons, 1997. Proceedings of Eighth International Conference, Beijing, P.R. China.

Roland Glowinski, Gene H. Golub, Gérard A. Meurant, and Jacques Périaux, editors. *Domain Decomposition Methods for Partial Differential Equations*, Philadelphia, PA, 1988. SIAM. Proceedings of the First International Symposium on Domain Decomposition Methods for Partial Differential Equations, Paris, France, January 1987.

Roland Glowinski, Yuri A. Kuznetsov, Gérard A. Meurant, Jacques Périaux, and Olof Widlund, editors. *Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations*, Philadelphia, PA, 1991. SIAM. Held in Moscow, USSR, May 21–25, 1990.

Wolfgang Hackbusch. *Iterative Solution of Large Sparse Linear Systems of Equations*. Springer-Verlag, Berlin, 1994.

David E. Keyes, Tony F. Chan, Gérard A. Meurant, Jeffrey S. Scroggs, and Robert G. Voigt, editors. *Fifth International Symposium on Domain Decomposition Methods for Partial Differential Equations*, Philadelphia, PA, 1992. SIAM. Held in Norfolk, VA, May 6–8, 1991.

David E. Keyes, Youcef Saad, and Donald G. Truhlar, editors. *Domain-Based Parallelism and Problem Decomposition Methods in Computational Sciences and Engineering*. SIAM, 1995.

David E. Keyes and Jinchao Xu, editors. *Domain Decomposition Methods in Science and Engineering*, volume 180 of *Contemporary Mathematics*, Providence, R.I., 1994. AMS. Proceedings of the Seventh International Conference on Domain Decomposition, October 27-30, 1993, The Pennsylvania State University.

Choi-Hong Lai, Petter E. Bjørstad, Mark Cross, and Olof Widlund, editors. *Eleventh International Conference on Domain Decomposition Methods*, 1998. Proceedings of the 11th International Conference on Domain Decomposition Methods in Greenwich, England, July 20-24, 1998.

Patrick Le Tallec. Domain decomposition methods in computational mechanics. In J. Tinsley Oden, editor, *Computational Mechanics Advances*, volume 1 (2), pages 121–220. North-Holland, 1994.

Jan L. Mandel, editor. *Domain Decomposition Methods 10 - The tenth International Conference on Domain Decomposition Methods*, Providence, RI, 1998. AMS. Proceedings of the tenth International Conference on Domain Decomposition Methods , August 10-14, 1997, Boulder, Colorado, U. S. A.

Magine S. Espedal Petter E. Bjørstad and David E. Keyes, editors. *Ninth International Conference on Domain Decomposition Methods*, 1997. Proceedings of the 9th International Conference on Domain Decomposition Methods in Bergen, Norway.

Alfio Quarteroni, Yuri A. Kuznetsov, Jacques Périaux, and Olof B. Widlund, editors. *Domain Decomposition Methods in Science and Engineering: The Sixth International Conference on Domain Decomposition*, volume 157 of *Contemporary Mathematics*. AMS, 1994. Held in Como, Italy, June 15–19,1992.

Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, 1996.

Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

Jinchao Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, December 1992.

We also mention the homepage for domain decomposition on the World Wide Web, `www.ddm.org`, maintained by Professor Martin Gander of McGill University. This site features links to conference, bibliographic, and personal information pertaining to domain decomposition, internationally.

The technical direction of the Thirteenth International Conference on Domain Decomposition Methods in Scientific and Engineering Computing was provided by a scientific committee consisting of: Petter E. Bjørstad, Tony F. Chan, Peter J. Deuflhard, Roland Glowinski, Ronald Hoppe, Hideo Kawarada, David E. Keyes, Yuri A. Kuznetsov, Jacques Périaux, Olivier Pironneau, Alfio Quarteroni, Zhong-Ci Shi, Olof B. Widlund, and Jinchao Xu.

Local organization was undertaken by the following members of the faculty and staff of the Center for Distributed and Parallel Scientific Computing at the University of Lyon, Claude Bernard: Jacques Baranger, Naïma Debit, Dominique Eyheramendy, Marc Garbey, Jean-François Maître, Fabienne Oudin, and Damien Tromeur-Dervout.

The scientific and organizing committees, together with all attendees, are grateful to the following agencies, organizations, corporations, and departments for their financial and logistical support of the conference: Compaq France, Alstom, Hewlett-Packard, Silicon Graphics Inc, Platform Computing, France Telecom, Région Rhône-Alpes, Direction Générale de l'Armement, ECCOMAS, CNRS, Fédération Calcul Scientifique Lyon, Conseil Général du Rhône, Mairie de Villeurbanne, SMAI, and GAMNI.

CIMNE, UPC Spain, is publishing the proceedings of the International Conference on Domain Decomposition Methods for the second time. The editors are very grateful to Eugenio Oñate for his patience with our process.

N. Debit
Lyon, France

M.Garbey
Houston, USA

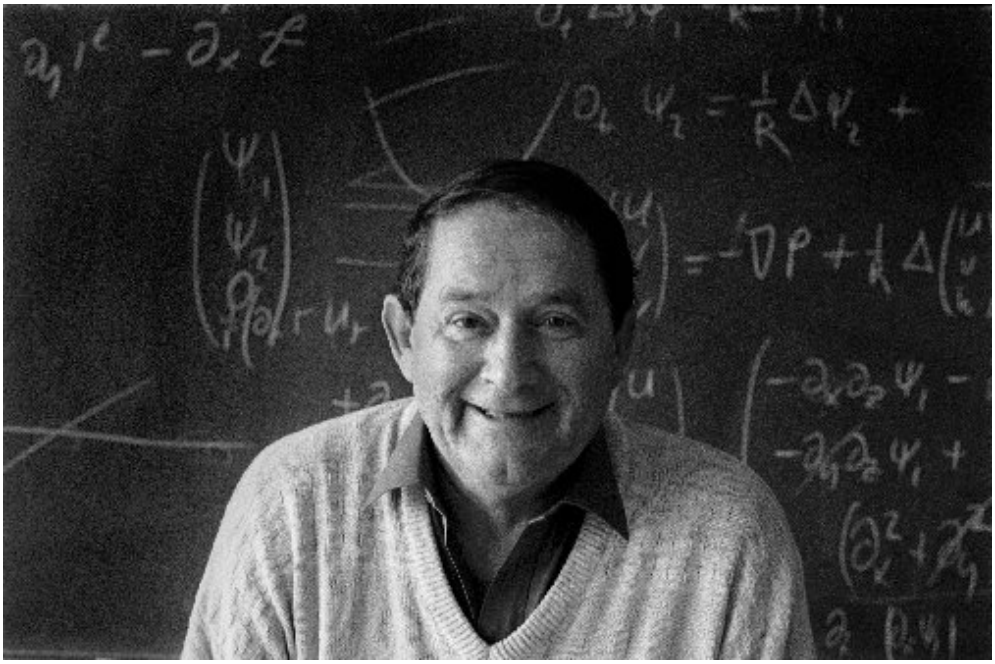
R. Hoppe
Augsburg, Germany

J. Périaux
Paris, France

D. Keyes
Norfolk, USA

Y. Kuznetsov
Houston, France

November 2001



This conference has been dedicated to the memory of Wiktor Eckhaus who was a great applied Mathematician, and a good man. His contribution to the matching asymptotic theory in the 70's was in nature a domain decomposition approach to the construction of uniform asymptotic expansion for singular perturbed problems.

Contents

Part I

INVITED LECTURES

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

1 Additive Schwarz method for nonsymmetric problems : application to frictional multicontact problems

P. Alart [1], M. Barboteu [2], P. Le Tallec [3], M. Vidrascu [4]

Introduction

In this paper, we present a generalization of a Neumann-Neumann domain decomposition method for solving nonsymmetric elliptic systems in a scalable way. It uses the theoretical framework of Schwarz additive decomposition method and introduces a coarse space well adapted to nonsymmetric cases. The efficiency of this method is evaluated on nonsymmetric frictional contact problems.

In iterative substructuring, the parallel solution of a complex structural problem is achieved by splitting the original domain of computation in smaller nonoverlapping simpler subdomains, and by reducing the initial problem to an interface system to be solved by a parallel two-level preconditioned conjugate gradient method. Many variants of this approach have been proposed and investigated in the recent literature, all associated to different choices of preconditioners and of coarse spaces [BPS86], [Smi92], [LTDRV91].

Up to now, the main objectives when developing such preconditioners were to achieve efficiency and scalability even in presence of complex geometries, strongly heterogeneous coefficients, general elliptic operators (3D anisotropic elasticity, shells, etc ..) and arbitrary meshes (unstructured, nonmatching, etc ..). These objectives cannot be reached without an adequate coarse solver [DW92]. For FETI preconditioners, this coarse solver is introduced by strongly imposing a kinematic constraint at each iteration (rigid body modes in FETI1 [FR94], rigid and corner modes in FETI2, corner modes only in FETI DP [FLL$^+$01]). In balanced Neumann-Neumann techniques, this solver appears while imposing orthogonality to an adequate coarse space of singular modes. The recent applications have introduced two new key dimensions in the development of such a coarse solver, namely its ability to handle nonsymmetric operators, and its industrial feasibility (automatic construction and cost efficiency). In our case, this new perspective is motivated by multicontact frictional problems.

This evolution requires complete review of the construction process of such coarse solvers, which is done hereafter in the framework of the Neumann-Neumann Domain Decomposition Method. The key point is the construction of the local spaces $\bar{Z}_i$ of rigid motions. For symmetric problems, the space $\bar{Z}_i$ is the kernel $Ker\mathbf{S}^i$ of the local Schur operators, with the possible addition of corner modes for fourth order problems. For advection diffusion problems, the good choice is based on constants. In the general case, the choice of $\bar{Z}_i$ must both set the arbitrary constants to zero in the solutions of the local Neumann problems (thus ensuring a scale invariance of the related energy norm), and regularize these local problems. For this purpose, we will introduce dual rigid modes obtained by solving local adjoint regularized Neumann problems.

[1]Université Montpellier, alart@lmgc.univ-montp2.fr
[2]Université Perpignan, barboteu@univ-perp.fr
[3]Ecole Polytechnique, patrick.letallec@polytechnique.fr
[4]INRIA Rocquencourt, Marina.Vidrascu@inria.fr

The paper is organized as follows. The basic domain decomposition methodology is first reviewed (§2), with an application to frictional contact problems illustrating the difficulties arising in presence of nonsymmetric operators (§3). Such nonsymmetric problems are handled in (§4) by reformulating the two level Neumann-Neumann preconditioner to an additive Schwarz algorithm, and by defining an appropriate coarse space by duality. In the last section (§5), we test the efficiency of this updated general Neumann-Neumann preconditioner on the numerical solution of nonsymmetric structural problems with contact and friction.

Balancing method for symmetric systems

The basic idea in nonoverlapping domain decomposition methods is to split the domain Ω of study into N small nonoverlapping subdomains $\Omega^n (n = 1, N)$ and interfaces defined by : $\Omega = \bigcup_{n=1}^{N} \Omega^n \cup \bigcup_{n=1}^{N} \Gamma^n \quad \text{with} \quad \Gamma^n = \partial\Omega^n \cap \left(\bigcup_{\substack{p=1 \\ p \neq n}}^{N} \partial\Omega^p \right) - \partial\Omega.$
Substructuring techniques consist then in reducing the original global system to an interface problem by a block Gaussian elimination of the internal degrees of freedom and in iteratively solving the resulting variational interface problem :

$$\exists \bar{\mathbf{u}} \in \bar{V} \quad / \quad < \mathbf{S}\bar{\mathbf{u}}, \bar{\mathbf{v}} > = < \bar{\mathbf{f}}, \bar{\mathbf{v}} > \qquad \forall \bar{\mathbf{v}} \in \bar{V} = \mathrm{Tr}\, H(\Omega)|_{\Gamma}. \tag{1}$$

The matrices $\mathbf{S} = \sum_{i=1}^{N} \mathbf{R}^i \mathbf{S}^i (\mathbf{R}^i)^t$ and $\mathbf{S}^i$ denote respectively the global Schur complement matrix (defined on Γ) and the local Schur complement matrices (defined on Γ^i by $\mathbf{S^i} = \bar{\mathbf{K}}^{\mathbf{i}} - (\mathbf{B^i})^{\mathbf{t}} (\mathring{\mathbf{K}}^{\mathbf{i}})^{-\mathbf{1}} \mathbf{B^i}$). Above, $(\mathbf{R}^i)^t$ is the restriction operator which goes from Γ to Γ^i, and $\mathbf{K^i} = \begin{pmatrix} \mathring{\mathbf{K}}^i & \mathbf{B^i} \\ (\mathbf{B^i})^{\mathbf{t}} & \bar{\mathbf{K}}^i \end{pmatrix}$ denotes the subdomain stiffness matrix, the first block corresponding to the internal degrees of freedom $\mathring{\mathbf{X}}^i$, the second one corresponding to the interface degrees $\bar{\mathbf{X}}^i$. The interface problem (1) can be solved by a preconditioned conjugate gradient method (symmetric cases) or the GMRES method (nonsymmetric cases). Hereafter, we use the multilevel Neumann-Neumann preconditioner. This iterative technique never requires the explicit calculation of the matrix $\mathbf{S}$. We have just to form the matrix vector products $\mathbf{S}\bar{\mathbf{p}}$ and $\mathbf{M^{-1}}\bar{\mathbf{r}}$ by solving independent auxiliary Dirichlet and Neumann problems on the local subdomains and a global coarse problem defined on a space of singular (rigid body) motions. Altogether, the product of the preconditioner $\mathbf{M}^{-1}$ and of the residual gradient $\bar{\mathbf{r}}$ has the following form,

$$\mathbf{M}^{-1}\bar{\mathbf{r}} = \sum_{i=1}^{N} \left\{ \mathbf{D}^i \, (\tilde{\mathbf{S}}^i)^{-1} \, (\mathbf{D}^i)^t \, \bar{\mathbf{r}} \right\} - \mathbf{G}\, \gamma,$$

where $\mathbf{D}^i$ is a weighting matrix, defining a local partition of unity on the interface and $(\tilde{\mathbf{S}})^{-1}$ denotes an regularized inverse of $\mathbf{S}^i$. Moreover, $\mathbf{G}\, \gamma$ is linear combination of subdomain rigid body motions over the interface obtained by projection of the residual onto this set of rigid body motions. In practice, the projection $\mathbf{G}\, \gamma$ is obtained by solving a global optimization problem over the interface Γ in order to minimize the residual [LT94] :

$$\min_{\mathbf{G}\gamma} \|\bar{\mathbf{r}}\|_{\Gamma}^2 := \min_{\mathbf{G}\gamma} \left\{ (\mathbf{S}\, (\mathbf{M}^{-1} - \mathbf{S}^{-1})\, \bar{\mathbf{r}})^t \, (\mathbf{M}^{-1} - \mathbf{S}^{-1})\, \bar{\mathbf{r}} \right\}. \tag{2}$$

This balanced preconditioner is very general and can be efficiently applied to linear or nonlinear three-dimensional elasticity problems using either matching or nonmatching grids [TSV94], to nonlinear plate or shell problems [TMV98].

A first "mechanical" nonsymmetric extension

As constructed above, the basic balanced Neumann Neumann preconditioner is not well adapted to nonsymmetric problems. Indeed the minimization problem (2) is not well defined for nonsymmetric Schur complement matrices. The numerical experiments [BAV01] also show that the behaviour of the iterative Schur complement solver (GMRES algorithm) is strongly perturbed when applied to structural problems with friction, i.e. when nonsymmetry is introduced in the tangent matrices [4]. The first idea is to replace the matrix $\mathbf{S}$ by the symmetrized matrix $\mathbf{S}^s$ ($\mathbf{S}^s = \mathbf{S} + \mathbf{S}^t$). Another choice is to use a symmetric matrix which has a mechanical meaning [BAV01] considering the interface reduced matrix $\mathbf{S}^*$ with a zero friction coefficient ($\mathbf{S}^* = \mathbf{S}_{\mu=0}$) to evaluate the norm of the difference between $\mathbf{M}^{-1}$ and $\mathbf{S}^{-1}$ and so to formulate the coarse problem. Then the minimization problem takes the following form :

$$\min_{\mathbf{G}\gamma} |||\bar{\mathbf{r}}|||_{\Gamma}^2 := \min_{\mathbf{G}\gamma} \left\{ \left(\mathbf{S}^* \left(\mathbf{M}^{-1} - \mathbf{S}^{-1}\right) \bar{\mathbf{r}}\right)^t \left(\mathbf{M}^{-1} - \mathbf{S}^{-1}\right) \bar{\mathbf{r}} \right\}. \tag{3}$$

This minimum is reached for the function $\mathbf{G}\,\gamma$ which cancels its gradient, which defines $\mathbf{G}\,\gamma$ as the solution of the following equality :

$$\left(\mathbf{G}^t\ \mathbf{S}^*\ \mathbf{G}\right) \gamma = -\mathbf{G}^t\ \mathbf{S}^* \sum_{i=1}^{N} \left(\mathbf{D}^i \left(\tilde{\mathbf{S}}^{\mathbf{i}}\right)^{-1} \left(\mathbf{D}^i\right)^t\right)\ \bar{\mathbf{r}}, \tag{4}$$

which defines the coarse problem specially adapted to the nonsymmetry of the friction [BAV01]. As we will see later, the dependence due to nonsymmetry is reduced, but it is nonoptimal. So, to establish a general nonsymmetric preconditioner, we now introduce a generalisation of this preconditioner by viewing it as an additive Schwarz method.

Interpretation as additive Schwarz methods and general extension to nonsymmetric problems

The Neumann-Neumann preconditioner can in fact be viewed as an additive Schwarz technique [TV97] iteratively solving an interface problem with operator $A = \mathbf{S}$ on the interface space $\bar{V}$ using the preconditioner

$$\mathbf{M}^{-1} = \tilde{\mathbf{A}}_o^{-1} + \sum_i \mathbf{I}^i (\tilde{\mathbf{A}}^i)^{-1} (\mathbf{I}^i)^t.$$

Above, the operator $\tilde{\mathbf{A}}_0 = \tilde{\mathbf{S}}$ (resp. $\tilde{\mathbf{A}}^i = \tilde{\mathbf{S}}^{\mathbf{i}}$) denotes an approximate restriction of the original operator $\mathbf{S}$ onto the coarse space $\bar{V}_G = \sum_{i=1}^{N} D_i \bar{Z}_i \subset \bar{V}$ (resp. onto the local spaces $\bar{V}_i^{\perp}$), the local spaces $\bar{V}_i^{\perp} \subset \bar{V}_i = \mathrm{Tr}\, H(\Omega)|_{\Gamma^i}$ are locally defined by duality

$$\bar{V}_i^{\perp} = \{\bar{\mathbf{v}}_f \in D_i \bar{V}_i, \quad < \mathbf{S}\bar{\mathbf{v}}_f, \bar{\mathbf{v}}_G >= 0, \qquad \forall \bar{\mathbf{v}}_G \in \bar{V}_G\}$$

and the extension from local to global space is given by $\mathbf{I}^i = (\mathbf{I} - \mathbf{P}_G)\mathbf{D}^i$, with $\mathbf{P}_G : \bar{V} \to \bar{V}_G$ the orthogonal $\tilde{\mathbf{S}}$ projection. This extension operator is in fact the key originality of the Neumann-Neumann preconditioner. With this notation, the additive Schwarz preconditioner reduces to the previous preconditioner

$$\mathbf{M}^{-1} = \tilde{\mathbf{S}}_0^{-1} + \sum_i (\mathbf{I} - \mathbf{P}_G)\mathbf{D}^i(\tilde{\mathbf{S}}^i)^{-1}(\mathbf{D}^i)^t(\mathbf{I} - \mathbf{P}_G)^t. \tag{5}$$

operating within the orthogonal of the coarse space, that is the image of the projection $(\mathbf{I} - \mathbf{P}_G)$.

The basic question is now to properly construct the local component $\bar{Z}_i$ of the coarse space $\bar{V}_G$. The objective is that its **orthogonal** complement (where the preconditioner lives) be nice. With a detailed examination, it can be observed that being nice means in fact that:
- the local Neumann solutions $\mathbf{w}^i$ must be scale invariant in energy norm, which requires to put all constants to zero in the local Neumann subproblems,
- the local Neumann subproblems must be regularized by adding a few boundary conditions.

Altogether, one only needs to impose implicitly that a few constants or boundary conditions $\mathbf{C}_\alpha^i$ be equal to zero for the solutions $\mathbf{w}^i$ of the local Neumann problems. We therefore need them to satisfy

$$\langle \mathbf{w}^i, \mathbf{C}_\alpha^i \rangle = 0, \forall \alpha,$$

that is

$$\langle \mathbf{K}^i\mathbf{w}^i, (\mathbf{K}^i)^{-t}\mathbf{C}_\alpha^i \rangle = 0, \forall \alpha,$$

or equivalently, since $\mathbf{w}^i$ is solution of a local Neumann problem with matrix $\mathbf{K}^i$

$$\langle \mathbf{D}^i\mathbf{S}^i\bar{\mathbf{v}}, (\mathbf{K}^i)^{-t}\mathbf{C}_\alpha^i \rangle = 0, \forall \alpha, \bar{\mathbf{v}} \in \bar{V}_G^{\perp}.$$

This is automatically guaranteed if $\bar{\mathbf{v}}$ is orthogonal to the function $\mathbf{D}^i(\mathbf{K}^i)^{-T}\mathbf{C}_\alpha^i$, that is if the local space is generated by the so called dual rigid modes as follows

$$\bar{Z}_i = \text{vect}((\mathbf{K}^i)^{-t}\mathbf{C}_\alpha^i).$$

Detailed algorithm

The adapted strategy which generalizes the approach of both the symmetric and the advection case, is thus given by the following steps [PAV00] :

1. Identify the local degrees of freedom $(P_{i\alpha})_{\alpha=1,N_i}$ which cancel all N_i rigid modes of subdomain i. In practice, this is done by identification of the small pivots in the factorization of the associated local stiffness matrix, with the possibility of choosing more degrees of freedom than necessary. For plate and shell problems, we can simply choose the degrees of freedom which lie on subdomain corners.

2. Introduce a regularization $\mathbf{K}_R^i$ of the local stiffness $\mathbf{K}^i$ on $V_i = H(\Omega^i)$

$$< \mathbf{K}_R^i\mathbf{v}^i, \hat{\mathbf{v}}^i > = < \mathbf{K}^i\mathbf{v}^i, \hat{\mathbf{v}}^i > + \sum_{\alpha,\beta} \mathbf{M}_{\alpha\beta}^i \mathbf{v}^i(P_{i\alpha})\hat{\mathbf{v}}^i(P_{i\beta}), \quad \forall \mathbf{v}^i, \hat{\mathbf{v}}^i \in V_i,$$

   the matrix $\mathbf{M}^i$ being a definite positive arbitrary matrix. For nonsymmetric problems, the matrices $\mathbf{K}^i$ and $\mathbf{K}_R^i$ are nonsymmetric.

3. Compute **dual rigid modes** $(\mathbf{v}^i_{G\alpha})_{\alpha=1,N_i}$ by solving local regularized Neumann problems set on the space V_i of subdomain displacement functions defined on subdomains i,

$$< (\mathbf{K}^i_R)^t \mathbf{v}^i_{G\alpha}, \hat{\mathbf{v}}^i >= \hat{\mathbf{v}}^i(P_{i\alpha}), \forall \hat{\mathbf{v}}^i \in V_i, \mathbf{v}^i_{G\alpha} \in V_i. \tag{6}$$

   For advection-diffusion problems or for unsteady problems, we must also introduce the dual constant mode defined by,

$$< (\mathbf{K}^i_R)^t \mathbf{v}^i_G, \hat{\mathbf{v}}^i >= \int_{\Omega^i} \hat{\mathbf{v}}^i, \quad \forall \hat{\mathbf{v}}^i \in V_i, \mathbf{v}^i_G \in V_i, \tag{7}$$

   in order to achieve scale invariance in the Neumann subproblems.

4. Introduce the local rigid space $Z_i = \text{vect}\Big(\mathbf{v}^i_{G\alpha}, \alpha = 1, N_i\Big)$.

The last construction leads to the local rigid spaces already introduced for symmetric cases [TV97] or for the advection-diffusion case [ATNV00]. The space Z_i does not depend on the choice of the regularized matrix $\mathbf{M}^i$ because all elements $\mathbf{v}^i$ of Z_i verify by construction,

$$< (\mathbf{K}^i)^t \mathbf{v}^i, \hat{\mathbf{v}}^i >= 0, \forall \hat{\mathbf{v}}^i \in V_i \text{ such that } \hat{\mathbf{v}}^i(P_{i\beta}) = 0, \forall \beta.$$

With this choice, the 2-level Neumann-Neumann preconditioner takes the form defined in (5)

$$\mathbf{M}^{-1}\mathbf{S} = \mathbf{P}_G + \sum_{i=1}^{N} (\mathbf{I} - \mathbf{P}_G)\mathbf{D}^i (\tilde{\mathbf{S}}^i)^{-1} (\mathbf{D}^i)^t (\mathbf{I} - \mathbf{P}_G)^t \mathbf{S}. \tag{8}$$

Above, the regularized Schur inverse $(\tilde{\mathbf{S}}^i)^{-1}$ acting on a given linear form L_i defined on the local interface space $\bar{V}_i'$ yields the interface vector $(\tilde{\mathbf{S}}^i)^{-1} L_i = Tr(\mathbf{w}^i)_{\Gamma^i}$ obtained by solution of the local regularized Neumann problem :

$$< \mathbf{K}^i_R \mathbf{w}^i, \hat{\mathbf{v}}^i >= L_i(Tr(\hat{\mathbf{v}}^i)_{|\Gamma^i}), \quad \forall \hat{\mathbf{v}}^i \in V_i, \mathbf{w}^i \in V_i. \tag{9}$$

Our construction ensures that the solutions $\mathbf{w}^i = (\tilde{\mathbf{S}}^i)^{-1} (\mathbf{D}^i)^t (\mathbf{I} - \mathbf{P}_G)^t \bar{\mathbf{r}}$ of the local Neumann problems have rigid constants $\mathbf{w}^i(P_{i\alpha})$ fixed to zero. Indeed, by definition of the dual rigid modes $\mathbf{v}^i_{G\alpha}$ and by the construction of $\mathbf{w}^i$ and by the projection $\mathbf{P}_G$, we have :

$$\mathbf{w}^i(P_{i\alpha}) =< (\mathbf{K}^i_R)^t \mathbf{v}^i_{G\alpha}, \mathbf{w}^i >=< \bar{\mathbf{r}}, (\mathbf{I} - \mathbf{P}_G)\mathbf{D}^i \mathbf{v}^i_{G\alpha} >= 0. \tag{10}$$

This value of the rigid constant on $\mathbf{w}^i$ cancels the effect of the regularization. We have indeed: $< \mathbf{K}^i_R \mathbf{w}^i, \mathbf{w}^i >=< \mathbf{K}^i \mathbf{w}^i, \mathbf{w}^i >$, which guarantees in some way the optimality of our algorithm.

Application to frictional contact problems

Nonsymmetric frictional contact problems

The behaviour of multicontact structures is characterized by a multiplicity of contact interfaces between deformable structure bodies. These large nonlinear problems constitute a class

of problems well suited to the use of the above numerical substructuring techniques. The modelling of the frictional contact problem is first based on a hybrid formulation presented in Alart and Curnier [PC91]. Following this augmented Lagrangian approach [PC91], the equilibrium of a discretized contact bodies system is governed by the system of nonlinear equations

$$\left\{ \begin{array}{l} F_{int} - F_{ext} + \mathbf{N}^t \mathcal{F}(\mathbf{u}, \lambda) = 0, \\ -\frac{1}{r}(\lambda - \mathcal{F}(\mathbf{u}, \lambda)) = 0, \end{array} \right. \tag{11}$$

where $\mathbf{N}$ is a restriction operator from Ω to Γ_c (Γ_c is the contact boundary). The notation $\mathbf{u}$ stands for kinematic variables (displacements or rotations) and λ for the static variables (contact forces or torques). Moreover, $\mathcal{F}(\mathbf{u}, \lambda)$ defines the discretized contact operator, with r the corresponding penalty coefficient, F_{int} and F_{ext} denote respectively the internal and the external discretized forces,

$$< F_{int}(\mathbf{u}), \hat{\mathbf{v}} >= \int_\Omega E\sigma(\nabla_{sym}\mathbf{u}) : \nabla_{sym}\hat{\mathbf{v}} \quad \text{and} \quad < F_{ext}, \hat{\mathbf{v}} >= \int_\Omega f \cdot \hat{\mathbf{v}},$$

and $\mathcal{F}(\mathbf{u}, \lambda)$ is the assembly of elementary contributions according to the notion of contact element [PC91]. For sake of simplicity, the local contact operator is presented for a contact between a deformable body and a rigid obstacle in a bidimensional modelling. Consequently the displacement $\mathbf{u}$ concerns only the node of the body on Γ_c and λ the contact force exerted by Γ_c on the obstacle. It is convenient to split it into normal and tangential components $\lambda = \lambda_n \mathbf{n} + \lambda_t$ and to express $\mathcal{F}^e(\mathbf{u}, \lambda)$ in this local frame :

$$\mathcal{F}^e(\mathbf{u}, \lambda) = \sigma_n^- \mathbf{n} + Proj_{C(\sigma_n^-)}\sigma_t, \tag{12}$$

where $\sigma = \sigma_n \mathbf{n} + \sigma_t$, $\sigma_n = \lambda_n + r u_n$, $\sigma_t = \lambda_t + r\delta\mathbf{u}_t$, $\sigma_n^- = min(\sigma_n, 0)$ and $C(\sigma_n^-)$ the Coulomb set $[\mu\sigma_n^-, -\mu\sigma_n^-]\mathbf{t}$) (where μ is the Coulomb coefficient and $\delta\mathbf{u}_t$ is a displacement increment). If the contact status is sliding, the tangent matrix of this operator is non symmetric and takes the tensorial form

$$\partial_\lambda \mathcal{F}^e(\mathbf{u}, \lambda) = (\mathbf{n} - \mu\mathbf{t}) \otimes \mathbf{n}, \quad \partial_u \mathcal{F}^e(\mathbf{u}, \lambda) = r(\mathbf{n} - \mu\mathbf{t}) \otimes \mathbf{n}.$$

For more complex contact elements, this type of local matrix is distributed on all contact nodes of target contactor areas.

We have chosen to treat both variables $\mathbf{u}$ and λ simultaneously through Newton's method. The system of equations is then split into two parts involving the pair $\mathbf{x} = (\mathbf{u}, \lambda)$, i.e. a differentiable elastic part G and a nondifferentiable frictional contact one $\mathcal{F}$

$$\mathbf{G}(\mathbf{x}) + \mathcal{F}(\mathbf{x}) = 0. \tag{13}$$

To overcome the nondifferentiability of the equation (13), Newton's method may be extended to the following iterative form [PC91]:

$$(\mathbf{K}^m + \mathbf{A}_c^m)\,\Delta\mathbf{x}^m = -\,(\mathbf{G}(\mathbf{x}^m) + \mathcal{F}(\mathbf{x}^m)) \quad \text{where} \quad \Delta\mathbf{x}^m = \mathbf{x}^{m+1} - \mathbf{x}^m, \tag{14}$$

to be solved at each iteration m by the previously introduced generalized Neumann-Neumann domain decomposition method. The matrix $\mathbf{K}^m = \partial\mathbf{G}(\mathbf{x}^m)$ is the usual elastic stiffness matrix and $\mathbf{A}_c^m \in \partial\mathcal{F}(\mathbf{x}^m)$ represents the generalized Jacobian of $\mathcal{F}$ at $\mathbf{x}^m$. The nonsymmetry of the matrix $\mathbf{A}_c^m$ is due to the friction terms. The contact interface is discretized by contact finite elements which yields elementary nonsymmetric tangent matrices if the contact status is "in friction situation".

"Multi-contact" structures

The efficiency of these different multilevel preconditioners will be assessed on two examples of "multicontact" structures :
- collections of deformable grains with contact interfaces between the grains.
- rolling shutters composed by many slats jointed by a hinge with play and eventually rotative friction.

Collection of deformable grains

Our motivation here is to study in granular media modelling the behaviour of a collection of deformable grains submitted to classical solicitations such as shear or compression. This problem is an interesting and delicate "multicontact" problem : the proportion of contact is very large. The interactions between the grains are governed by the frictional contact laws (Signorini unilateral contact law and Coulomb friction law).

At a discrete level, the interactions between grains are modelled by a frictional contact

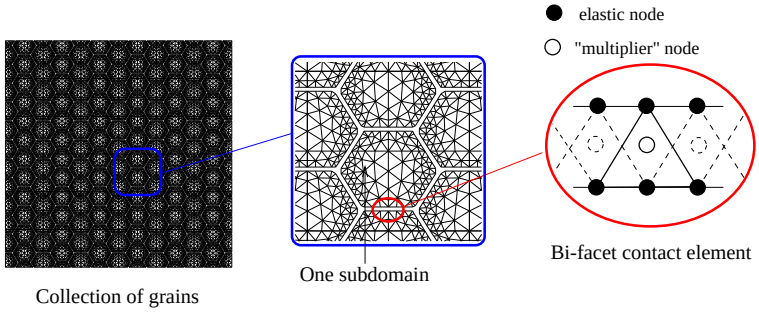


Figure 1: Deformable grains, one subdomain and a bi-facet contact element.

element (Figure 1) which takes into account large slip over the contact interface. This bi-facet contact element has 5 nodes : 4 elastic nodes which contain the displacement $\mathbf{u}$ (u_x, u_y) and a multiplier node containing the frictional contact forces λ. Moreover the contactor node can slip over two target facets. A generalization to more facets can be carried out easily.

Rolling shutters composed by many hinged slats

The aim of this problem is to simulate the quasi-static behavior of such shutters submitted to strong winds [ABLM99]. A rolling shutter is a specific case of multi-contact structure. The rolling shutters for shops, stores and hangars are formed by a succession of slats jointed by a hinge [ABLM99]. Such a structure is then composed by an assembly of elastic structures (plates in flexion and torsion) which leads to consider a large number of contact zones. The edges of the slats are designed in such a way that the slats fit into each other. To facilitate the rolling of the shutters at the opening, the profile of the slat requires a gap or a play in the hinge. We must then develop a specific model which takes into account the play $(-g, +g)$ in the hinge and eventually the friction in the rotations $(\delta\beta)$ of the hinges between the slats. The contact and friction laws are more complicated than the usual case. For more details on the modelling, see [ABLM99].

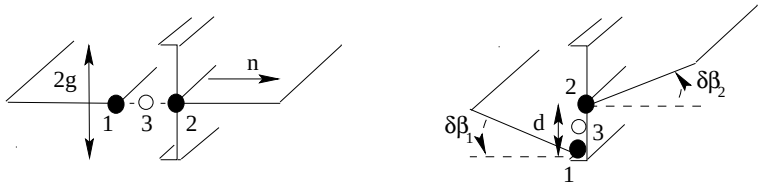


Figure 2: hinge contact element.

Substructuring strategy

One feature of this nonlinear nonsymmetric domain decomposition strategy consists in putting the numerical subdomain interfaces away from the physical contact interfaces [BAV01]. Contrary to current approaches we therefore suggest to treat the physical contact interfaces as internal surfaces : the contact interfaces (hinges for shutters and contact area for deformable grains) must be inside the subdomains and do not constitute decomposition interfaces. Thus, the decomposition is not forced to respect the geometry of its components; such a subdomain is shown in Figure 1. This allows a better balance of the size of the subdomains and leads to an optimal decomposition for parallel efficiency.

Numerical behaviour of Neumann-Neumann preconditioners

In this section, we analyse the convergence behaviour of the interface solver (GMRES) with the multi-level Neumann-Neumann preconditioners. We test their efficiency as a function of the friction coefficient and the number of subdomains (scalability properties). As previously observed, the nonsymmetry is due to our formulation of frictional contact problems. The considered preconditioners are :

- The standard Neumann-Neumann preconditioner with coarse space (2-level),
- The specific Neumann-Neumann preconditioner which uses a symmetrized matrix $\mathbf{S}^*$ (with a friction coefficient equal to zero),
- The new nonsymmetric Neumann-Neumann preconditioner introduced in this paper.

The first result, described in Figure 3, gives the evolution of average number of GMRES

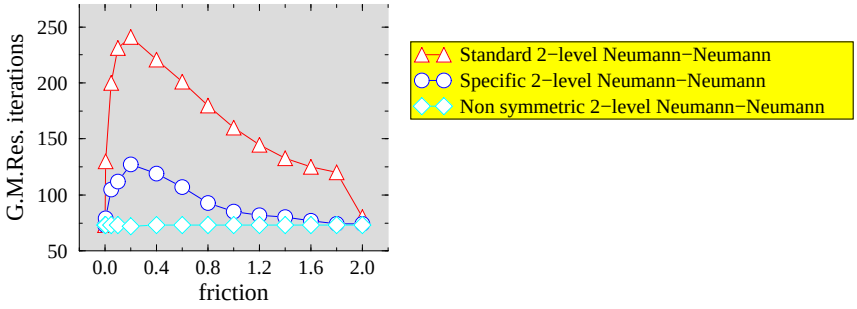


Figure 3: Influence of the friction coefficient on the preconditioners.

iterations (per Newton iterations) for different values of the friction coefficient varying from

0 to 2 for a rolling shutters with 16 slats and 30 subdomains (26 floating subdomains), respectively. We observe the inefficiency of the solver using the standard Neumann-Neumann preconditioner (curve $\triangle$) for values of friction coefficient close to $\mu = 0,2$. This is due to the large increase of the ratio of slip status and so to the large proportion of nonsymmetry. The first extension procedure (curve $\circ$) improves this dependance but does not cancel it. On the other hand, the new nonsymmetric preconditioner (curve $\diamond$) makes the interface solver insensitive to the nonsymetry.

Next, we analyse the scalability properties of the different Neumann-Neumann preconditioners for the problems of rolling shutters and collections of deformable grains. For the rolling shutters (figure 4), we can verify that for a problem without friction ($\mu = 0$, symmetric problem), the 2-level Neumann-Neumann preconditioner has a classical behaviour : independence from subdomain number (curve $*$). But with friction, the standard procedure

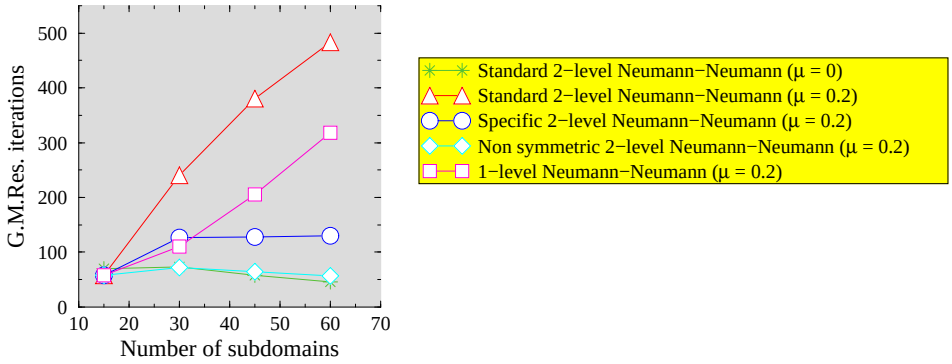


Figure 4: Numerical scalability of the preconditioners (rolling shutters).

leads to a high increase of the number of iterations (curve $\triangle$) with the number of subdomains. The results are even worse than without coarse solver (curve). The first extension strategy (curve $\circ$) improves the convergence but is not optimal. On the other hand, the 2-level nonsymmetric Neumann-Neumann preconditioner (curve $\diamond$) leads to a full recovery of the numerical scalability properties obtained with a symmetric problem.

We finally present for the collection of deformable grains the influence of the number of sub-domains (Figure 5) on the number of iterations. The good behaviour of the nonsymmetric preconditioner is confirmed when the number of floating subdomains increases. This nonsymmetric procedure is more efficient than the standard and specific balancing method specially in presence of shear. Indeed, the friction (and then the nonsymmetry) plays a more important role in shear than in compression (Figure 5). Thus the strategy developed in this paper extends to large scale nonsymmetric (frictional contact) problems.

References

[ABLM99]K. Ach, P. Alart M. Barboteu, F. Lebon, and B. MBodji. Parallel frictional contact algorithms and industrial applications. *Comp. Meth. Appl. Mech. Engng*, 177:169–181, 1999.

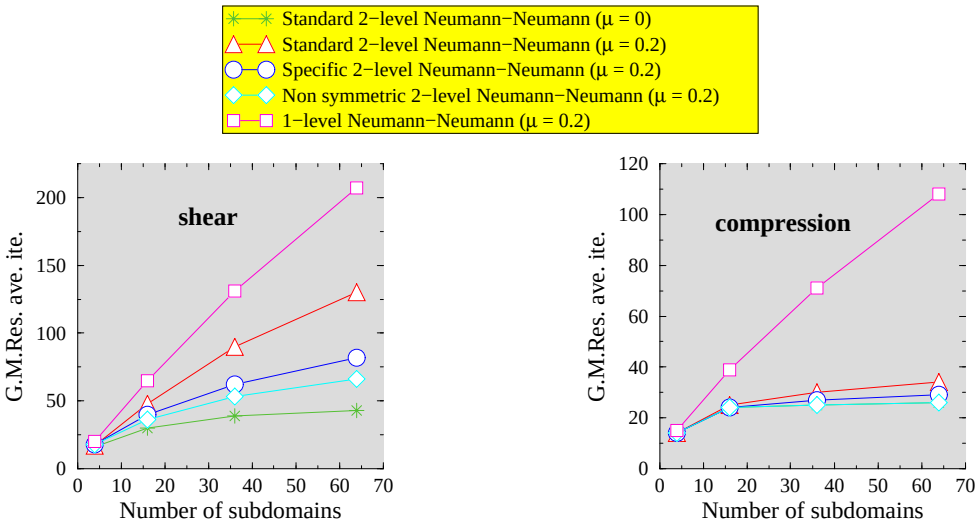


Figure 5: Numerical scalability of the preconditioners (deformable grains).

[ATNV00]Y. Achdou, P. Le Tallec, F. Nataf, and M. Vidrascu. A domain decoposition preconditioner for an advection-diffusion problem. *Comp. Meth. Appl. Mech. Engng*, 184:145–170, 2000.

[BAV01]M. Barboteu, P. Alart, and M. Vidrascu. A domain decomposition strategy for non classical frictional multicontact problems. *Comp. Meth. Appl. Mech. Eng.*, 2001. to appear.

[BPS86]James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. The construction of preconditioners for elliptic problems by substructuring, I. *Math. Comp.*, 47(175):103–134, 1986.

[DW92]Maksymilian Dryja and Olof B. Widlund. Additive Schwarz methods for elliptic finite element problems in three dimensions. In David E. Keyes, Tony F. Chan, Gérard A. Meurant, Jeffrey S. Scroggs, and Robert G. Voigt, editors, *Fifth International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 3–18, Philadelphia, PA, 1992. SIAM.

[FLL+01]Charbel Farhat, Michel Lesoinne, Patrick LeTallec, Kendall Pierson, and Daniel Rixen. FETI-DP: A dual-primal unified FETI method - part I: A faster alternative to the two-level feti method. *Int. J. Numer. Meth. Engng.*, 50:1523–1544, 2001.

[FR94]C. Farhat and F.X. Roux. Implicit parallel processing in structural mechanics. *Comput. Mech. Adv.*, 2:1–124, 1994.

[LT94]Patrick Le Tallec. Domain decomposition methods in computational mechanics. In J. Tinsley Oden, editor, *Computational Mechanics Advances*, volume 1 (2), pages 121–220. North-Holland, 1994.

[LTDRV91]Patrick Le Tallec, Yann-Hervé De Roeck, and Marina Vidrascu. Domain-decomposition methods for large linearly elliptic three dimensional problems. *J. of Computational and Applied Mathematics*, 34, 1991.

[PAV00]P. Le Tallec P. Alart, M. Barboteu and M. Vidrascu. Méthode de schwarz additive avec solveur grossier pour problèmes non symétriques. *C. R. Acad. Sci.*, 331(I):399–404,

2000.

[PC91]P.Alart and A. Curnier. A mixed formulation for frictional contact problems prone to newton like solution methods. *Comp. Meth. Appl. Mech. Eng.*, 92:353–375, 1991.

[Smi92]Barry F. Smith. An optimal domain decomposition preconditioner for the finite element solution of linear elasticity problems. *SIAM J. Sci. Stat. Comput.*, 13(1):364–378, January 1992.

[TMV98]P. Le Tallec, J. Mandel, and M. Vidrascu. A Neumann-Neumann domain decomposition algorithm for solving plate and shell problems. *SIAM J. Numer. Math.*, 35:836–867, 1998.

[TSV94]P. Le Tallec, T. Sassi, and M. Vidrascu. Three-dimensional domain decomposition methods with nonmatching grids and unstructured coarse solvers. In D. Keyes and J. Xu, editors, *Seventh International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 133–139. AMS, 1994.

[TV97]Patrick Le Tallec and Marina Vidrascu. Generalized Neumann-Neumann preconditioners for iterative substructuring. In Petter E. Bjørstad, Magne Espedal, and David Keyes, editors, *Domain Decomposition Methods in Sciences and Engineering*. John Wiley & Sons, 1997. Proceedings from the Ninth International Conference, June 1996, Bergen, Norway.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

2 Multigraph Algorithms Based on Sparse Gaussian Elimination

R. E. Bank[1], R. K. Smith[2]

Introduction

In this work, we describe a multilevel-multigraph algorithm. An excellent recent survey on algebraic approaches to multilevel iterative methods is given in Wagner [Wag99]. This article also contains an extensive bibliography. The algorithm discussed here is described more fully in [BS00]. Our goal is to develop an iterative solver with the simplicity of use and robustness of general sparse Gaussian elimination, and at the same time to capture the computational efficiency of classical multigrid algorithms. While we do not believe that the current algorithm achieves this goal, it represents an important step in this direction. To guarantee robustness, general sparse Gaussian elimination with minimum degree ordering is a point in the parameter space of our method. This is a well known and widely used method, among the most computationally efficient of general sparse direct methods [GL81].

To obtain simplicity of use and implementation, our algorithms incorporate many technologies and algorithms originally developed for general sparse Gaussian elimination. Besides the minimum degree algorithm, the Reverse Cuthill-McKee ordering is the basis of our coarsening procedure. Our sparse matrix data structures are a generalization of those first introduced in the symmetric Yale Sparse Matrix Package [EGSS82], and our (incomplete) factorization procedure is a generalization of the sparse row elimination scheme used there. To gain computational efficiency, our method offers the possibility to compute an incomplete factorization with the user able to specify a drop tolerance and an absolute bound on the total fill-in. This factorization becomes the smoother in a multilevel procedure similar to the classical multigrid method.

Sparse direct methods typically have two phases. In the *initialization* phase, equations are ordered, and symbolic and numerical factorizations are computed. In the *solution* phase, the solution of the linear system is computed using the factorization. Our procedure, as well as other algebraic multilevel methods, also breaks naturally into two phases. The initialization consists of ordering, incomplete symbolic and numeric factorizations, and the computation of the transfer matrices between levels. In the solution phase, the preconditioner computed in the initialization phase is used to compute the solution using the preconditioned Composite Step Conjugate Gradient (CSCG) or the Composite Step Biconjugate Gradient (CSBCG) method [BC93].

In the spirit of general sparse Gaussian elimination, we have tried to minimize the number of user specified control parameters. In the initialization phase, there are three parameters. The most important is the drop tolerance ($dtol$) for the incomplete factorization. Because the fill-in for the ILU tends to be a very nonlinear and unpredictable function of the drop tolerance, we also allow the user to specify an upper bound on the amount to fill-in the be allowed in the

[1]University of California at San Diego, La Jolla CA 92093, rbank@ucsd.edu. The work of this author was supported by the National Science Foundation under contract DMS-9706090.
[2]Agere Systems, Murray Hill, NJ 07974, kentsmith@agere.com.

incomplete factorization ($maxfil$). Finally, the maximum number of levels in the multilevel procedure ($maxlvl$) can be specified. In the solution phase, the user can specify only two control parameters: the maximum number of iterations ($maxcg$) and an error tolerance (tol) for the convergence criterion.

Our main interest is in developing a solver for discretizations of scalar elliptic problems as in the finite element code *PLTMG* [Ban98]. However, our solver was developed as a stand-alone linear equations solver, and can formally be applied to any structurally symmetric, nonsingular, sparse matrix. By structurally symmetric, we mean that the pattern of nonzeros in the matrix is symmetric, although the numerical values of the matrix elements may render it nonsymmetric. Many problems arising in practice naturally have structural symmetry, and of course all can be *made* structurally symmetric by storing some extra zeroes. For certain problems handled by *PLTMG*, the matrices are symmetric and positive definite, but for others, the linear systems are highly nonsymmetric and/or indefinite. Thus in practice, this represents a very broad class of behavior.

Structural symmetry allows for some important simplifications in the implementation. In particular, we can handle linear systems involving symmetric matrices A and nonsymmetric matrices A and A^t within a single, unified code, rather than developing specialized subroutines for each of these three cases. In the nonsymmetric case, linear systems involving A^t arise naturally in the context of the CSBCG algorithm, and hence are important for our solver. This limits the complexity of the code, and also eliminates additional parameters that might be needed to further classify a given matrix. On the other hand, it seems clear that a specialized solver directed at a specific problem or class of problems, and making use of additional knowledge, is likely to outperform our algorithm on that particular class of problems. Although we do not think our method is provably "best" for any particular problem, we believe its generality and robustness, coupled with reasonable computational efficiency, make it an interesting and useful approach for solving linear systems.

Matrix Formulation

Let A be a large sparse, nonsingular $N \times N$ matrix. We assume that the sparsity pattern of A is symmetric, although the numerical values need not be. We consider the solution of the linear system

$$Ax = b. \tag{1}$$

Let B be an $N \times N$ nonsingular smoothing matrix. In our case, B is an approximate factorization of A, i.e.,

$$B = (L + D)D^{-1}(D + U) \approx P^t AP, \tag{2}$$

where L is (strict) lower triangular, U is (strict) upper triangular with the same sparsity pattern as L^t, D is diagonal, and P is a permutation matrix.

Given an initial guess x_0, m steps of the smoothing procedure produce iterates x_k, $1 \le k \le m$, given by

$$\begin{aligned} r_{k-1} &= P^t(b - Ax_{k-1}), \\ B\delta_{k-1} &= r_{k-1}, \\ x_k &= x_{k-1} + P^t\delta_{k-1}. \end{aligned} \tag{3}$$

The second component of the two-level preconditioner is the *coarse grid correction*. Here we assume that the matrix A can be partitioned as

$$\hat{P}A\hat{P}^t = \begin{pmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{pmatrix}, \tag{4}$$

where the subscripts f and c denote *fine* and *coarse*, respectively. Similar to the smoother, the partition of A in fine and coarse blocks involves a permutation matrix $\hat{P}$. The $\hat{N} \times \hat{N}$ coarse grid matrix $\hat{A}$ is given by

$$\begin{aligned} \hat{A} &= \begin{pmatrix} V_{cf} & I_{cc} \end{pmatrix} \begin{pmatrix} A_{ff} & A_{fc} \\ A_{cf} & A_{cc} \end{pmatrix} \begin{pmatrix} W_{fc} \\ I_{cc} \end{pmatrix} \\ &= V_{cf}A_{ff}W_{fc} + V_{cf}A_{fc} + A_{cf}W_{fc} + A_{cc}. \end{aligned} \tag{5}$$

The matrices V_{cf} and W_{fc}^t are $\hat{N} \times N$ matrices, with identical sparsity patterns; thus $\hat{A}$ has a symmetric sparsity pattern. If $A^t = A$, we require $V_{cf} = W_{fc}^t$, so $\hat{A}^t = \hat{A}$.

Let

$$\hat{V} = \begin{pmatrix} V_{cf} & I_{cc} \end{pmatrix} \hat{P}, \qquad \hat{W} = \hat{P}^t \begin{pmatrix} W_{fc} \\ I_{cc} \end{pmatrix}. \tag{6}$$

In standard multigrid terminology, the matrices $\hat{V}$ and $\hat{W}$ are called *restriction* and *prolongation*, respectively. Given an approximate solution x_m to (1), the coarse grid correction produces an iterate x_{m+1} as follows:

$$\begin{aligned} \hat{r} &= \hat{V}(b - Ax_m), \\ \hat{A}\hat{\delta} &= \hat{r}, \\ x_{m+1} &= x_m + \hat{W}\hat{\delta}. \end{aligned} \tag{7}$$

In typical multilevel fashion, the linear system $\hat{A}\hat{\delta} = \hat{r}$ in (7) is solved by recursion, in our case a multilevel V-cycle. One the coarsest level, we apply the iteration (3). A single cycle takes an initial guess x_0 to a final guess x_3 as follows: x_1 is defined using (3), x_2 is defined using (7), and x_3 is defined using (3). Note in particular that we use only one pre-smoothing and one post-smoothing iteration.

Some Implementation Details

To complete the definition of the method, we must provide algorithms to:

- Compute the incomplete factorization matrix B in (2).
- Compute the permutation matrix P in (2).
- Compute the fine-coarse partitioning matrix $\hat{P}$ in (4).
- Compute the sparsity patterns and numerical values in the prolongation and restriction matrices in (6).

ILU Factorization

Our incomplete $(L+D)D^{-1}(D+U)$ factorization is similar to the row elimination scheme developed for the symmetric YSMP codes [EGSS82, GL81]. Without loss of generality, assume that the permutation matrix $P = I$, so that $A = (L+D)D^{-1}(D+U) + E$, where E is the error matrix.

After k steps of elimination, we have the block factorization

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} D_{11}+L_{11} & 0 \\ L_{21} & I \end{pmatrix} \begin{pmatrix} D_{11}^{-1} & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} D_{11}+U_{11} & U_{12} \\ 0 & I \end{pmatrix} + \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & 0 \end{pmatrix},$$

where A_{11} is $k \times k$ and A_{22} is $N-k \times N-k$. We assume that at this stage the blocks D_{11}, L_{11}, L_{21}, U_{11}, and U_{12} have been computed.

Our goal for step $k+1$ is to compute the first row and column of the approximate Schur complement S, given by

$$\ell = Se_1 = A_{22}e_1 - L_{21}(D_{11}^{-1}U_{12}e_1),$$
$$u = S^t e_1 = A_{22}^t e_1 - U_{12}^t(D_{11}^{-1}L_{21}^t e_1).$$

This is done by a procedure similar to the row elimination scheme employed by the symmetric YSMP codes. After the complete (sparse) vectors ℓ and u are computed, certain entries are dropped (assigned to the error matrix E). In particular, we neglect a pair of off-diagonal elements if

$$\max |L_{ij}|, |U_{ji}| \le dtol\sqrt{|D_{jj}A_{ii}|}, \tag{8}$$

where $j = k+1$; D_{ii} has not yet been computed. The drop tolerance $dtol$ is applied in a symmetric fashion to maintain a symmetric sparsity pattern in the factorization.

It is well known that the fill-in generated through the application of a criterion such as (8) is a highly nonlinear and matrix dependent function of $dtol$. This is especially problematic in the present context, since control of the fill-in is necessary in order to control the work per iteration in the multilevel iteration. Thus, in addition to the drop tolerance $dtol$, the user sets the parameter $maxfil$, which specifies that the total number of nonzeros in U is not larger than $maxfil \cdot N$.

Our basic strategy is to compute the incomplete decomposition using the given drop tolerance. If it fails to meet the given storage bound, we increase the drop tolerance and begin a new incomplete factorization. We continue in this fashion until we complete a factorization within the given storage bound. Of course, such repeated factorizations are computationally expensive, so we develop heuristics which allow us to predict a drop tolerance which will satisfy the storage bound. Thus, should the original factorization fail to satisfy the storage bound, usually only one additional ILU factorization is needed. This is discussed in detail in [BS00].

Finally, we note that there is no comprehensive theory regarding the stability of incomplete triangular decompositions. For certain classes of matrices (e.g., M-matrices), the existence of certain incomplete factorizations has been established; however, in the general case, with potentially indefinite and/or highly nonsymmetric matrices, one must contend in a practical way with the possibility of failure or near failure of the factorization. In our implementation, a failure is revealed by some diagonal entries in D becoming close to zero. Off-diagonal elements L_{ji} and U_{ij} are multiplied by D_{ii}^{-1}, and the solution of $(L+D)D^{-1}(D+U)x = b$ also

involves multiplication by D_{ii}^{-1}. For purposes of calculating the factorization and solution, the value of D_{ii}^{-1} is modified near zero as follows:

$$D_{ii}^{-1} = \begin{cases} 1/D_{ii} & \text{for } |D_{ii}| > \alpha \\ D_{ii}/\alpha^2 & \text{for } |D_{ii}| \leq \alpha \end{cases} .$$

Here α is a small constant; in our implementation, $\alpha = \mu\|A\|$, where μ is the machine epsilon. Although many failures could render the preconditioner well-defined but essentially useless, in practice we have noted that D_{ii}^{-1} is rarely modified for a the large class of finite element matrices which are the main target of our procedure.

Ordering

The minimum degree ordering is used to compute the permutation matrix P in (2). Intuitively, if one is computing an incomplete factorization, an ordering which tends to minimize the fill-in in a complete factorization should tend to minimize the error E in the incomplete factorization. For particular classes of matrices, specialized ordering schemes have been developed; for example, for matrices arising from convection dominated problems, ordering along the flow direction has been used with great success. However, in this general setting, we prefer to use just one strategy for all matrices, to reduce the complexity of the implementation, and to avoid the issue of deciding among various ordering possibilities. We remark that for convection dominated problems, minimum degree orderings perform comparably well to the specialized ones, provided some (modest) fill-in is allowed in the incomplete factorization. For us, this seems to be a reasonable compromise.

Our minimum degree ordering is a standard implementation. We have implemented two small enhancements to the minimum degree ordering; as a practical matter, both involve changes to the input graph data structure that is provided to the minimum degree code. First, we have implemented a drop tolerance similar to that used in the factorization. In particular, the edge in the graph corresponding to off-diagonal entries A_{ij} and A_{ji} is not included in the input data structure if

$$\max |A_{ij}|, |A_{ji}| \leq dtol\sqrt{|A_{jj}A_{ii}|}.$$

This excludes many entries which are likely to be dropped in the subsequent incomplete factorization.

The second modification involves some modest *a priori* diagonal pivoting designed to minimize the number failures (near zero diagonal elements) in the subsequent factorization. This procedure is described in detail in [BS00].

Fine-Coarse Partitioning

Our coarsening scheme is based upon another well-known sparse matrix ordering technique, the Reverse Cuthill-McKee algorithm. This ordering tends to yield reordered matrices with minimal bandwidth, and is widely used with generalized band elimination algorithms [GL81]. Our coarsening procedure is just a simple post-processing step of the basic ordering routine, in which the N vertices of graph are marked as *COARSE* or *FINE*. Initially, all vertices are *UNMARKED*. We proceed through the vertices in RCM order; each *UNMARKED* vertex we

encounter is relabeled *COARSE*, and all of its neighbors are labeled *FINE*. This implicitly defines the matrix $\hat{P}$ given in (4).

Under this procedure, all coarse vertices are surrounded by fine vertices. This implies that the matrix A_{cc} in (4) is a diagonal matrix. For the sparsity patterns of matrices arising from discretizations of scalar partial differential equations as in *PLTMG*, the number of coarse unknowns $\hat{N}$ is typically on the order of $N/4$ to $N/5$.

Computing the Transfer Matrices

We now define the matrices V_{cf} and W_{fc}^t of (5). To define the sparsity structure, we take all the connections of each coarse grid vertex to its fine grid neighbors; that is, the sparsity structures of V_{cf} and W_{fc}^t are the same as the block A_{cf}.

We chose numerical values for V_{cf} and W_{fc} according to the formulae

$$W_{fc} = -R_{ff} D_{ff}^{-1} A_{fc},$$
$$V_{cf} = -A_{cf} D_{ff}^{-1} \tilde{R}_{ff}.$$

Here D_{ff} is a diagonal matrix with diagonal entries equal to those of A_{ff}. In this sense, the nonzero entries in V_{cf} and W_{fc} are chosen as multipliers in Gaussian Elimination. The nonnegative diagonal matrices R_{ff} and $\tilde{R}_{ff}$ are chosen such that nonzero rows of W_{fc} and columns of V_{cf}, respectively, have unit norms in ℓ_1.

Finally, if necessary, the coarsened matrix $\hat{A}$ of (5) is "sparsified" using the drop tolerance and a criterion like (8) to remove small off-diagonal elements. Empirically, applying a drop tolerance to $\hat{A}$ at the end of the coarsening procedure has proved more efficient, and more effective, than trying to independently sparsify its constituent matrices.

Numerical Illustrations

In this section, we present a few numerical illustrations. The problems are all of the form $\mathcal{L}_i u = 1$ in $\Omega = (0,1) \times (0,1)$ with $u = 0$ on $\partial\Omega$. The operators $\mathcal{L}_i$, $1 \le i \le 7$, are given by

$$\begin{aligned}
\mathcal{L}_1 u &= -\Delta u, \\
\mathcal{L}_2 u &= -\Delta u - 1000\, u_x, \\
\mathcal{L}_3 u &= -\Delta u - 1000\, u_x - 1000\, u_y, \\
\mathcal{L}_4 u &= -\Delta u - 1000\, u, \\
\mathcal{L}_5 u &= -\Delta u + 1000\, u, \\
\mathcal{L}_6 u &= -.001 u_{xx} - u_{yy}, \\
\mathcal{L}_7 u &= -\Delta u - 1000\{(y - .5)\, u_x - (x - .5)\, u_y\}.
\end{aligned}$$

These problems are standard PDE's chosen to reflect a wide variety of behavior. We solved these problems on $n \times n$ uniform meshes with $n = 51, 101, 201$; the resulting linear systems are of order $N = n^2$. Uniform meshes were used for standardization, although these problems could be more effectively solved in *PLTMG* using adaptive meshes. A 5×5 mesh, as well as the solutions to the seven problems are shown in Figure 1. Continuous piecewise linear

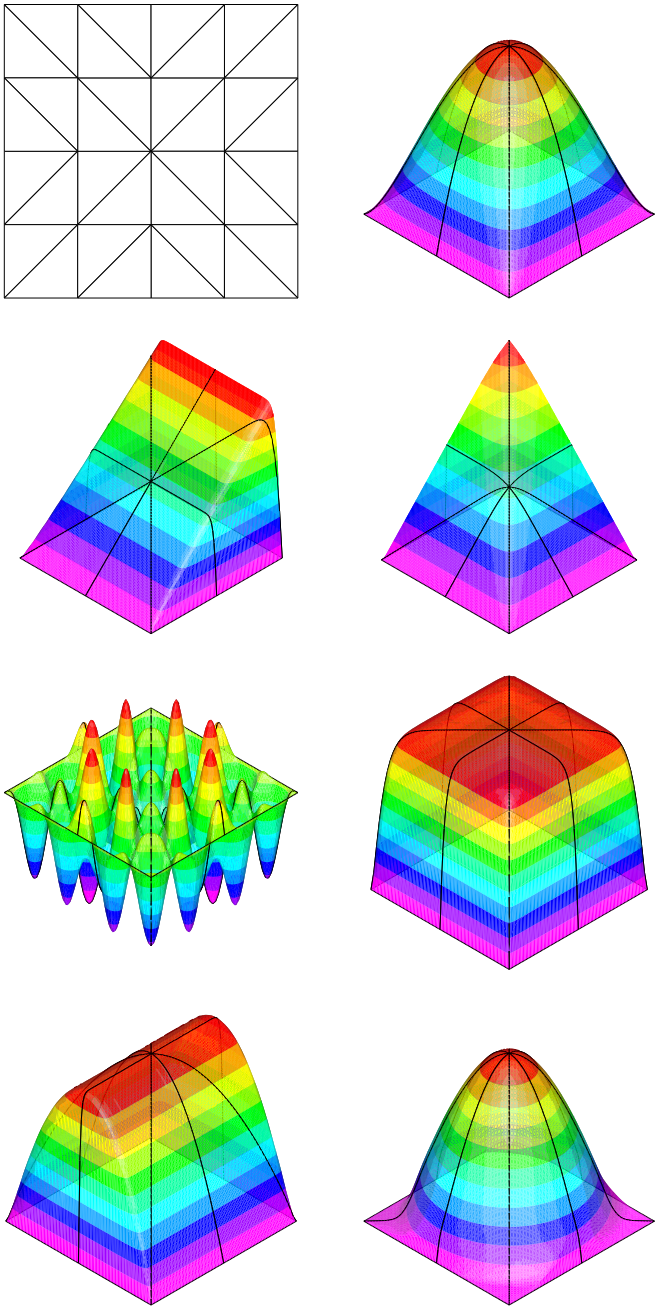


Figure 1: A 5×5 uniform triangulation, and solutions to problems 1-7.

Table 1: Performance comparison.

n	N	Levels	Digits	Cycles	Init.	Solve
Problem 1, $dtol = 10^{-2}$						
51	2601	7	8.4	3	2.1e-1	5.1e-2
101	10201	7	6.8	3	1.0e 0	2.9e-1
201	40401	9	6.2	3	4.5e 0	1.4e 0
Problem 2, $dtol = 10^{-3}$						
51	2601	7	7.5	1	2.5e-1	6.4e-2
101	10201	8	6.1	1	1.2e 0	3.7e-1
201	40401	9	10.3	3	6.5e 0	4.6e 0
Problem 3, $dtol = 10^{-3}$						
51	2601	7	11.6	1	5.8e-1	1.2e-1
101	10201	8	6.5	1	4.9e 0	5.1e-1
201	40401	8	10.0	2	6.8e 0	3.7e 0
Problem 4, $dtol = 10^{-4}$						
51	2601	6	6.7	1	3.7e-1	3.8e-2
101	10201	7	6.1	3	2.3e 0	4.9e-1
201	40401	7	7.0	4	13.4e 0	2.9e 0
Problem 5, $dtol = 10^{-2}$						
51	2601	7	7.6	2	2.7e-1	4.1e-2
101	10201	7	6.4	2	1.2e 0	2.3e-1
201	40401	8	6.7	2	4.4e 0	1.0e 0
Problem 6, $dtol = 10^{-4}$						
51	2601	6	8.6	1	2.0e-1	2.4e-2
101	10201	7	8.2	1	8.9e-1	1.4e-1
201	40401	8	7.5	1	4.1e 0	6.7e-1
Problem 7, $dtol = 10^{-3}$						
51	2601	6	10.3	2	2.9e-1	1.0e-1
101	10201	7	7.7	2	1.6e 0	6.6e-1
201	40401	8	6.6	2	8.2e 0	3.1e 0

finite elements and the usual nodal basis functions are used in *PLTMG* to construct the linear systems.

In Table 1, we summarize the results of the calculation. Here *Levels* refers to the number of levels used in the calculation. In this test, the parameter $maxlvl$ was sufficiently large that it had no effect on the computation. The fill-in control parameter $maxfil$ was also sufficiently large that it had no effect on the computation. The drop tolerance was set as indicated; although not carefully optimized, the tolerance was crudely chosen according to the difficulty of the problem to produce roughly comparable results for all problems. The initial guess for all problems was $x_0 = 0$.

The parameter *Digits* refers to

$$Digits = -\log \frac{\|r_k\|}{\|r_0\|}. \tag{9}$$

In these experiments, we asked for 6 digits of accuracy. The column labeled *Cycles* indicates the number of multigrid cycles (accelerated by composite step conjugate gradients or biconjugate gradients) that were used to achieve the indicated number of digits. Finally, the last two columns, labeled *Init.* and *Solve*, record the CPU time for the initialization and solution phases of the algorithm, respectively. Initialization includes all the orderings, incomplete factorizations, and computation of transfer matrices used in the multigraph preconditioner. Solution includes the time to solve (1) to at least 6 digits given the preconditioner. These experiments were run on an SGI Octane R10000 250mhz, using double precision arithmetic and the f90 compiler.

In analyzing these results, it is clear that our procedure does reasonably well on all of the problems. Although it appears that the rate of convergence is not always independent of N, it seems apparent that the work is growing no faster than logarithmically. CPU times for larger values of N are affected by cache performance as well as the slightly larger number of cycles.

In our next experiment, we illustrate the effect of the parameters $maxlvl$ and $dtol$. For the $\mathcal{L}_1$, $\mathcal{L}_4$, and $\mathcal{L}_7$ and $N = 40401$, we solved the problem for $dtol = 10^{-k}$, $1 \le k \le 4$ and $1 \le maxlvl \le 3$. $\mathcal{L}_4$ and $\mathcal{L}_7$ are the two most challenging problems in this suite. The results are given in Table 2. Although we expect all iterations to eventually converge (at least in exact arithmetic), we terminated the iteration after $maxcg = 25$ steps or when the solution had 6 digits, as measured by (9).

Here we see that, in general, decreasing the drop tolerance or increasing the number of levels improves the convergence behavior of the method. On the other hand, the timings do not always follow the same trend. For example, increasing the number of levels from $maxlvl = 1$ to $maxlvl = 2$ often decreases the number of cycles but increases the time. This is because for $maxlvl = 1$, our method defaults to the standard CG of BCG iteration with the incomplete factorization preconditioner. When $maxlvl > 1$, one pre-smoothing and one post-smoothing step are used for the largest matrix. With the additional cost of the recursion, the overall cost of the preconditioner is more than double the cost for the case $maxlvl = 1$.

We also note that, unlike the classical multigrid method, where the coarsest matrix is solved exactly, in our code we have chosen to approximately solve the coarsest system using just one smoothing iteration using the incomplete factorization. When the maximum number of levels are used, as in Table 1, the smallest system is typically 1×1 or 2×2, and this is an irrelevant remark. However, in the case of Table 2, the fact that the smallest system is not solved exactly significantly influences the convergence.

Table 2: Dependence of convergence of $dtol$ and $maxlvl$.

$dtol$	$maxlvl$	Digits	Cycles	Init.	Solve
Problem 1, $N = 40401$					
10^{-1}	1	3.7	25	1.1	2.7
	2	4.3	25	2.6	6.3
	3	6.2	22	3.7	7.6
10^{-2}	1	5.6	25	1.5	3.3
	2	6.2	13	3.5	4.3
	3	6.1	8	4.2	3.3
10^{-3}	1	6.0	12	2.2	1.8
	2	6.2	6	4.9	2.3
	3	7.0	4	5.6	2.0
10^{-4}	1	6.5	5	3.7	1.0
	2	6.5	2	7.9	1.2
	3	8.5	2	5.6	1.5
Problem 4, $N = 40401$					
10^{-1}	1	3.1	25	1.2	3.0
	2	3.6	25	2.7	7.2
	3	3.9	25	3.8	10.2
10^{-2}	1	3.7	25	1.5	4.4
	2	4.2	25	3.5	11.0
	3	2.5	25	4.2	10.0
10^{-3}	1	6.1	25	3.4	5.0
	2	3.7	25	7.7	11.8
	3	5.7	25	9.6	14.1
10^{-4}	1	6.7	5	8.1	1.3
	2	6.7	4	12.1	2.4
	3	7.0	4	12.9	2.8
Problem 7, $N = 40401$					
10^{-1}	1	3.1	25	1.4	7.9
	2	3.1	25	3.2	18.9
	3	4.5	25	4.0	22.3
10^{-2}	1	5.5	25	1.9	8.7
	2	6.2	10	4.5	9.5
	3	6.4	7	5.2	6.9
10^{-3}	1	7.0	6	2.9	2.3
	2	7.5	4	6.9	3.6
	3	7.7	3	7.8	3.8
10^{-4}	1	6.2	3	3.8	1.2
	2	8.8	2	9.4	2.4
	3	7.4	1	10.5	2.2

Finally we note biconjugate gradient iteration used for nonsymmetric problems requires two matrix multiplies and two preconditioning applications (for the matrix and its transpose), so the overall cost per step is about twice that of the regular conjugate gradient iteration.

References

[Ban98]Randolph E. Bank. *PLTMG: A Software Package for Solving Elliptic Partial DifferentiaEquations, Users' Guide 8.0l*, volume 5 of *Software, Environments and Tools*. SIAM, Philadelphia, 1998.

[BC93]Randolph E. Bank and Tony F. Chan. An analysis of the composite step biconjugate gradient method. *Numerische Mathematik*, 66:295–319, 1993.

[BS00]Randolph E. Bank and R. Kent Smith. An algebraic multilevel multigraph algorithm. Technical report, University of California at San Diego, 2000. submitted to SIAM J. on Scientific Computing.

[EGSS82]S. C. Eisenstat, M. C. Gursky, M. H. Schultz, and A. H. Sherman. Algorithms and data structures for sparse symmetric Gaussian elimination. *SIAM J. Sci. Stat. Comput.*, 2:225–237, 1982.

[GL81]Alan George and Joseph Liu. *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1981.

[Wag99]Christian Wagner. Introduction to algebraic multigrid. Technical report, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen,Universität Heidelberg, 1999.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

3 The Mortar element method revisited – What are the right norms?

D. Braess[1], W. Dahmen[2]

Introduction

A number of investigations have recently been devoted to the mortar method as a domain decomposition method with non-overlapping subdomains. Its attraction comes from its great flexibility due to the fact that different types of discretization are possible on different subdomains. The best experience is with H^1-elliptic problems. In contrast to standard conforming elements, there may be jumps across the interfaces between adjacent subdomains, and the continuity conditions are replaced by weak matching conditions the so called *mortaring conditions*. Our guiding question here will be to what extent there still remain "interdomain-constraints" on the discretizations which are possibly imposed by stability and accuracy requirements, in particular, when dealing with highly *non-quasi-uniform meshes*.

There is by now almost a standard way to treat mortar elements in the framework of nonconforming elements, where it was originally analyzed, see e.g. [BMP94]. However, since it may be technically cumbersome to eliminate the constraints imposed by the matching conditions and since fast solvers are by now available for mixed formulations, the analysis as a saddle point problem has recently attracted interest, see e.g. [BB99, BDW99, Woh99b]. Moreover, on a principal level the inf-sup condition is also often hidden in the analysis of mortar elements based on the nonconforming theory. If the inf-sup condition holds, the error of approximation by functions with and without the mortaring conditions are of the same order [Bra01, Remark III.4.10]. This tool is frequently used for estimating the term that represents the approximation error in the lemma of Berger, Scott, and Strang. Therefore we believe that the understanding of the saddle point formulation is at the heart of the matter which will be the point of view taken in this paper.

The fact that the framework for the saddle point formulation is still less well established in comparison with the nonconforming method is due to the subtle difference between (at least) two trace spaces in the scale of Sobolev spaces with index 1/2. To be specific, let Γ_{kl} denote the (typical) interface between the subdomains Ω_k and Ω_ℓ. When the variational problem is considered in the Sobolev space $H^1(\Omega)$ or $H^1_0(\Omega)$, then the trace space $H^{1/2}_{00}(\Gamma_{kl})$ endowed with the norm $\|g\|_{H^{1/2}_{00}(\Gamma_{kl})} := \|\chi_{\Gamma_{kl}} g\|_{1/2,\partial\Omega_k}$ (where $\chi_{\Gamma_{kl}}$ is the standard indicator function) turns out to be an appropriate function space for the jumps over the interior boundary Γ_{kl}. In the 2D case this can be realized by forcing the trial functions to be continuous at the cross points, which is a mild constraint. However, for 3D problems the jumps would have to vanish along the boundaries of the interfaces, and this would entail severe restrictions on the discretizations for neighboring subdomains. Thus jumps living in the larger space $H^{1/2}(\Gamma_{kl})$, are usually admitted in actual computations with mortar elements.

[1]Ruhr-Universität Bochum, braess@num.ruhr-uni-bochum.de
[2]RWTH Aachen, dahmen@igpm.rwth-aachen.de
The work of this author has been supported in part by the TMR network "Wavelets in Numerical Simulation" funded by the European Commission

This discrepancy (gap) prohibits the use of Brezzi's theory with the standard Sobolev spaces and their norms. For a rigorous treatment one had to resort to nonstandard methods. One possibility is to introduce *mesh-dependent norms* as done, e.g. in [BDW99, DFG$^+$01, Woh99b]. Continuity, ellipticity, and the inf-sup condition as required by Brezzi's theory are then available. Another concept can be found in [BB99, Woh99a] where the analysis is performed in a two-stage process. In a first step merely the direct variables are estimated by the nonconforming theory. In the second step only the inf-sup condition and no ellipticity is required for achieving an error estimate of the Lagrange multipliers.

A principal objective of this paper is to narrow this gap somewhat. Specifically, we will explore to what extent and under what circumstances one can dispense with mesh-dependent norms. Some mesh-dependence still turns out to remain but only for one variable and in a weaker form no longer involving an explicit mesh size parameter. Moreover, the new norms can be bounded by $\|\cdot\|_{H_{00}^{1/2}(\Gamma_{kl})}$ if applied to a function in the space $H_{00}^{1/2}(\Gamma_{kl})$. It models a function space in which $H_{00}^{1/2}(\Gamma_{kl})$ has a finite codimension, while it differs from $H^{1/2}(\Gamma_{kl})$ by an infinite dimensional space. It is now easily understandable why all the different concepts have one point in common. They all make use of the fact — in an open or hidden way — that the subset of finite element functions whose jumps belong to $H_{00}^{1/2}(\Gamma_{kl})$, is sufficiently thick.

Aside from these theoretical considerations there is the following practical reason for addressing the above issue. Nonoverlapping domain decomposition appears to be particularly suitable for problems with complicated domains or jumping coefficients so that one expects solutions with singular behavior. Therefore the use of highly non-quasi-uniform or adaptively refined meshes in different subdomains should be covered by the theory. However, the mesh-dependent norms from [BDW99, BD98, Woh99b] only work well when using quasi-uniform meshes. In fact, in connection with error estimates the mesh sizes should not even differ too much from one subdomain to the other one, see [DFG$^+$01] for an extension to mesh-dependent norms with suitable local mesh size functions.

So the core question is how independently from each other can the discretizations on different subdomains be chosen so as to retain stability and overall accuracy even when the individual meshes are highly non-quasi-uniform.

Recently, an error analysis has been performed in [KLPV01] for the mortar method on meshes that are only *locally* quasi-uniform. The price that has been paid there is that the meshes on adjacent subdomains have to *match along the boundary* of the interface which in the three dimensional case severly imposes on the mesh generator. Our approach allows us to abundon this constraint to restore full mortar flexibility. We still obtain error estimates of the same type as in [KLPV01], where the constants now depend only on *one sided mesh size ratios*. Cleary, local refinements on or near an interface would result from a singular behavior of the approximated solution on or near that interface affecting both adjacent subdomains. Thus conditions of this type (even two sided versions) tend to be satisfied automatically by reasonable mesh adaptation strategies.

The paper is organized as follows. In Section 3 we describe the continuous problem. Section 3 is concerned with the discrete counterparts. Specifically, we formulate several requirements to be met by the discretizations. These are similar in spirit (and in fact closely related) to those in [KLPV01] and have been recognized to play a pivotal role in many preceding investigations [BB99, BMP94, BDW99, BD98, Woh99b]. Section 3 is devoted to the stability analysis for this setting. In contrast to [KLPV01] we work here in a saddle point context for a choice of norms that is different from prior investigations. In Section 3 we dis-

cuss error estimates from different point of views. The concepts are then applied in Section 3 to the so called dual basis mortar method from [BP99, KLPV01, Woh99a]. In particular, we establish standard types of error estimates for locally quasi-uniform meshes without the above mentioned interface boundary matching condition from [KLPV01].

The continuous problem

Consider the second order elliptic boundary value problem

$$\begin{aligned} -\operatorname{div} a(x)\operatorname{grad} u(x) &= f(x) && \text{in } \Omega, \\ a(x)\tfrac{\partial u}{\partial n} &= g(x) && \text{on } \Gamma_N \subset \partial\Omega, \\ u &= 0 && \text{on } \Gamma_D := \partial\Omega \setminus \Gamma_N, \end{aligned} \tag{1}$$

where $a(x)$ is a piecewise sufficiently smooth and uniformly positive definite matrix defined for x in the bounded domain $\Omega \subset \mathbb{R}^d$, Γ_D is a subset of the boundary Γ of Ω (with positive measure relative to Γ), and $\Gamma_N := \Gamma \setminus \Gamma_D$. $H^1_{0,D}(\Omega)$ denotes the closure in $H^1(\Omega)$ of all C^∞-functions vanishing on Γ_D.

Suppose that Ω is decomposed into non-overlapping subdomains $\Omega_k, k = 1, \ldots, k_{\max}$, i.e.,

$$\bar\Omega = \bigcup_{k=1}^{k_{\max}} \bar\Omega_k, \quad \Omega_k \cap \Omega_l = \emptyset \qquad \text{for } k \neq l. \tag{2}$$

For simplicity we will assume throughout the rest of the paper that the domain $\Omega \subset \mathbb{R}^d$ and that the subdomains Ω_k in (2) are polyhedral. If the closures of Ω_k and Ω_l have a $(d-1)$-dimensional intersection, we set $\bar\Gamma_{kl} := \bar\Omega_k \cap \bar\Omega_l$. However, we do not insist on the partition to be geometrically conforming, i.e., Γ_{kl} need *not* be a full common face of both subdomains. The Γ_{kl} form the *skeleton*

$$\mathcal{S} := \bigcup_{k,l} \Gamma_{kl}.$$

Γ_{kl}, Γ_N, and Γ_D will always be assumed to be the union of polyhedral subsets of the boundaries of the Ω_k.

The mortar method is based on a variational formulation of (1) with respect to the product space

$$X_\delta := \{v \in L^2(\Omega) : v|_{\Omega_k} \in H^1(\Omega_k),\ k = 1, \ldots, k_{\max},\ v|_{\Gamma_D} = 0\},$$

endowed with the norm

$$\|v\|_{1,\delta} := \left(\sum\nolimits_{k=1}^{k_{\max}} \|v\|^2_{H^1(\Omega_k)} \right)^{1/2}.$$

The space $H^1_{0,D}(\Omega)$ is characterized as a subspace of X_δ determined by appropriate constraints on jumps across interfaces.

This suggests the following weak formulation of (1): For a *suitable* pair of spaces X, M, *find* $(u, \lambda) \in X \times M$ *such that*

$$\begin{aligned} a(u,v) + b(v,\lambda) &= (f,v)_{0,\Omega} + (g,v)_{0,\Gamma_N} \quad \text{for all } v \in X, \\ b(u,\mu) &= 0 \qquad\qquad\qquad\qquad\quad \text{for all } \mu \in M, \end{aligned} \tag{3}$$

where $(u,v)_{0,\Omega}$ and $(g,v)_{0,\Gamma_N}$ denote the L^2 inner products on Ω and Γ_N, respectively.

$$\begin{aligned} a(u,v) &:= \textstyle\sum_k \int_{\Omega_k} (a(x)\nabla u(x)) \cdot \nabla v(x) dx, \\ b(v,\mu) &:= \textstyle\sum_{\Gamma_{kl} \subset \mathcal{S}} (\mu, [v])_{0,\Gamma_{kl}}. \end{aligned}$$

The *jump* $[v]$ of a function $v \in X$ is defined on $\mathcal{S}$ by $[v] := v|_{\overline{\Omega}_k} - v|_{\overline{\Omega}_l}$ on Γ_{kl} (see [BDW99] for further background information). We note that each interface Γ_{kl} appears *only once* in the sum over $\mathcal{S}$.

Discretization

In order to describe next the mortar method as a *discrete* version of (3), we choose for each subdomain Ω_k a (conforming) triangulation $\mathcal{T}_k$ subject to the following assumptions:
(a) Each triangulation is completely *independent* of those on neighboring subdomains. This means that the nodes in $\mathcal{T}_k$ which belong to Γ_{kl} need *not* match with any of the nodes of $\mathcal{T}_l$.
(b) The $\mathcal{T}_k$ will always be shape regular but only *locally* quasi-uniform, i.e., the ratios of maximal and minimal diameters of the elements in $\mathcal{T}_k$ need not remain bounded.

With each $\mathcal{T}_k$ we associate a finite element space $\mathcal{S}(\mathcal{T}_k) \subset H^1(\Omega_k) \cap H^1_{0,D}(\Omega)$. In principle, this could have any fixed polynomial order, but for simplicity we will refer in most cases to spaces of piecewise linear finite elements on $\mathcal{T}_k$. We set

$$X_h := \prod_{k=1}^{k_{\max}} \mathcal{S}_1(\mathcal{T}_k) \subset X_\delta, \tag{4}$$

where the index h indicates the dependence on the discretization.

The next crucial step is to fix the Lagrange multipliers for each Γ_{kl} (i.e. the space M in (3)). In this context, we stress the following implicit notational convention to be used throughout the rest. The indexing of the interface Γ_{kl} (as opposed to Γ_{lk}) always expresses that Ω_k has been chosen as the non-mortar side. This distinction is important because the Lagrange multipliers will only depend on the non-mortar side in a way that will be specified later in more detail. Whenever Γ_{kl} is a full common face of both adjacent subdomains, the choice of the mortar side is completely arbitrary. If Γ_{kl} is strictly contained in at least one of the faces, the following provision has to be taken. We will always assume that $\partial\Gamma_{kl}$ is covered by the faces of the cells in Γ_{kl} induced by at least one of the triangulations $\mathcal{T}_k$ or $\mathcal{T}_l$. If only one of these triangulations has this property, the corresponding subdomain has to be chosen as the non-mortar side and hence will be denoted by Ω_k.

Meanwhile several types of Lagrange multiplier spaces have been considered in the literature, see e.g. [BMP94, BD98, KLPV01, Woh99a]. Instead of considering any specification we formulate first some requirements on the multiplier spaces that can be extracted from the above mentioned studies. To this end, let $\mathcal{T}_{kl}$ denote the restriction of the mesh $\mathcal{T}_k$ to Γ_{kl} and set $S^0_{kl} := \mathcal{S}(\mathcal{T}_{kl}) \cap H^1_0(\Gamma_{kl}) \subseteq H^{1/2}_{00}(\Gamma_{kl})$. Given S^0_{kl}, we will employ finite dimensional spaces $M_{kl} \subset L_2(\Gamma_{kl})$ with the following properties:

P.1 The spaces S^0_{kl} and M_{kl} have the same dimension

$$\dim S^0_{kl} = \dim M_{kl}. \tag{5}$$

P.2 Whenever $\mathcal{S}(\mathcal{T}_k)$ has *approximation order* n, then M_{kl} should have approximation order at least $n-1$, i.e.,

$$\inf_{w\in M_{kl}} \|v-w\|_{0,\Gamma_{kl}} \le c\bar{h}^{n-1}|v|_{n-1,\Gamma_{kl}}, \tag{6}$$

where $\bar{h}$ is the maximal mesh size of $\mathcal{T}_{kl}$. More precisely, defining for every vertex i of $\mathcal{T}_k$ the local mesh size $h_i := \max\{\operatorname{diam}\tau : \tau\in\mathcal{T}_{kl},\ i\in\tau\}$, we set

$$\bar{h} := \max_{i\in\mathcal{T}_{kl}} h_i, \qquad \underline{h} := \min_{i\in\mathcal{T}_{kl}} h_i.$$

Thus for piecewise linear finite elements on Ω_k one has $n=2$ and (6) requires first order convergence.

P.3 The pair (S^0_{kl}, M_{kl}) is L_2*-stable*, i.e.,

$$\inf_{w\in M_{kl}} \sup_{v\in S^0_{kl}} \frac{(v,w)_{0,\Gamma_{kl}}}{\|v\|_{0,\Gamma_{kl}}\|w\|_{0,\Gamma_{kl}}} \ge c_0 \tag{7}$$

for some fixed constant c_0 (depending on Γ_{kl}).

P.4 It is well-known that (7) implies that

$$(Q_{kl}w, v)_{0,\Gamma_{kl}} = (w,v)_{0,\Gamma_{kl}}, \quad \forall\ v\in S^0_{kl}, \tag{8}$$

uniquely defines a projector $Q_{kl} : L_2(\Gamma_{kl}) \to M_{kl}$ such that

$$\|Q_{kl}w\|_{0,\Gamma_{kl}} \le c_0^{-1}\|w\|_{0,\Gamma_{kl}}, \quad w\in L_2(\Gamma_{kl}). \tag{9}$$

Here we require in addition that the adjoint $Q^*_{kl} : L_2(\Gamma_{kl}) \to S^0_{kl}$ of Q_{kl} is also bounded on $H^{1/2}_{00}(\Gamma_{kl})$

$$\|Q^*_{kl}v\|_{H^{1/2}_{00}(\Gamma_{kl})} \le c_1\|v\|_{H^{1/2}_{00}(\Gamma_{kl})}. \tag{10}$$

The pair (S^0_{kl}, M_{kl}) is called *admissible* if **P.1 – P.4** hold.

Remark 1 *When $\mathcal{T}_{kl}$ is quasi-uniform,* **P.4** *is a consequence of (9) and the approximation property (6) in* **P.2** *provided that the spaces M_{kl} also satisfy a standard inverse property. Only if the meshes are merely locally quasi-uniform, requirement* **P.4** *requires attention.*

The space of discrete multipliers is now defined as

$$M_h := \prod_{\Gamma_{kl}\subset S} M_{kl} \tag{11}$$

where, again, the index h indicates the dependence on $\mathcal{T}_{kl}$ and should not be viewed as mesh size parameter when used as a subscript. Moreover, the finite element functions that satisfy the mortaring conditions, form the space

$$V_h := \{v_h \in X_h,\, b(v_h, \mu) = 0 \quad \forall \mu_h \in M_h\}. \tag{12}$$

The discrete counterpart to (3) now reads

$$\begin{array}{lcll} a(u_h, v_h) + b(v_h, \lambda_h) & = & (f, v_h)_{0,\Omega} + (g, v_h)_{0,\Gamma_N}, & v_h \in X_h, \\ b(u_h, \mu_h) & = & 0, & \mu_h \in M_h. \end{array} \tag{13}$$

We will show that, (13) is a stable and accurate discretization of (3), if the pairs (S^0_{kl}, M_{kl}), $\Gamma_{kl} \in \mathcal{S}$ are (uniformly) admissible in the above sense.

Stability

First we address the stability of (13). In contrast to [KLPV01] we treat (13) as a saddle point problem. Thus, one has to show that the operators

$$\mathcal{L}_h := \begin{pmatrix} A_h & B_h^T \\ B_h & 0 \end{pmatrix} \;:\; X_h \times M_h \to X_h' \times M_h' \tag{14}$$

induced by (13) are *uniformly* bounded and have *uniformly bounded inverses* with respect to the underlying meshes. Of course, this depends on the norms for X_h and M_h which have yet to be specified. As explained in [BDW99], due to the subtle differences between the trace spaces $H^{1/2}(\Gamma_{kl})$ and $H^{1/2}_{00}(\Gamma_{kl})$, standard (broken Sobolev norms) turn out to be inappropriate. While for quasi-uniform grids appropriate *mesh-dependent norms* offer a cure [BDW99, BD98, Woh99b, Woh99a] we wish to reduce the mesh-dependence of norms in favor of mesh flexibility.

Our main deviation from previous studies therefore lies in the choice of the norms. Recall that the jumps $[v_h]$ are *not* required to lie in the spaces $H^{1/2}_{00}(\Gamma_{kl})$ which naturally arise in the analysis of the continuous problem. However, it will be seen that it suffices to measure their *projection* into the trace spaces $S^0_{kl} \subset H^{1/2}_{00}(\Gamma_{kl})$ in the norm $\|\cdot\|_{H^{1/2}_{00}(\Gamma_{kl})}$. In fact, for any $v_h \in X_h$ we define

$$\|v_h\|^2_{1,h} := \|v_h\|^2_{1,\delta} + \sum_{\Gamma_{kl} \in \mathcal{S}} \|Q^*_{kl}[v_h]\|^2_{H^{1/2}_{00}(\Gamma_{kl})}, \tag{15}$$

while for $\mu \in M_h \subset M := \prod_{\Gamma_{kl} \subset \mathcal{S}} (H^{1/2}_{00}(\Gamma_{kl}))'$ we take the natural dual norm

$$\|\mu\|^2_{-1/2} := \sum_{\Gamma_{kl} \subset \mathcal{S}} \|\mu\|^2_{(H^{1/2}_{00}(\Gamma_{kl}))'}. \tag{16}$$

Note that any mesh-dependence of $\|\cdot\|_{1,h}$ enters only implicitly through the projectors Q^*_{kl}.

First we address the continuity of the bilinear forms $a(\cdot,\cdot)$, $b(\cdot,\cdot)$ with respect to these norms. Since for $v_h \in X_h$ and $\mu_h \in M_h$

$$|(v_h, \mu_h)_{0,\Gamma_{kl}}| = |(Q^*_{kl} v_h, \mu_h)_{0,\Gamma_{kl}}| \le \|Q^*_{kl} v_h\|_{H^{1/2}_{00}(\Gamma_{kl})} \|\mu_h\|_{(H^{1/2}_{00}(\Gamma_{kl}))'},$$

one has, in view of (15), that

$$|a(u_h, v_h)| \lesssim \|u_h\|_{1,h}\|v_h\|_{1,h}, \quad |b(v_h, \mu_h)| \lesssim \|v_h\|_{1,h}\|\mu_h\|_{-1/2}, \tag{17}$$

holds for any $v_h, u_h \in X_h$, $\mu_h \in M_h$, where the constants depend on the constant c_0 in **P.3**.

The first step towards confirming stability of the discretization is to confirm the ellipticity of the bilinear form $a(\cdot,\cdot)$ on the kernel

$$V_h := \{v \in X_h : b(v, \mu) = 0 \quad \text{for } \mu \in M_h\}$$

of the constraints.

Proposition 1 *The bilinear form $a(\cdot,\cdot)$ is elliptic on V_h, i.e.,*

$$a(v, v)\|v\|^2_{1,h} \qquad \textit{for all } v \in V_h. \tag{18}$$

Proof The inequality $a(v, v)\|v\|^2_{1,\delta}$ for $v \in V_h$, can be inferred by the analysis in [BMP94]. So the desired ellipticity estimate stated in the theorem follows as soon as we have proved that also $\sum_{\Gamma_{kl}\subset S} \|[v]\|^2_{1/2,h,\Gamma_{kl}} \lesssim \|v\|^2_{1,\delta}$ for $v \in V_h$. But this is obviously true since by definition of Q_{kl} one has for $v_h \in V_h$ and any $w \in L_2(\Gamma_{kl})$ that $(Q^*_{kl}[v_h], w)_{0,\Gamma_{kl}} = ([v_h], Q_{kl}w)_{0,\Gamma_{kl}} = 0$. Thus $Q^*_{kl}[v_h] = 0$ which completes the proof. ∎

Since the continuity (17) and ellipticity (18) have already been established, it remains to verify the validity of the *LBB-condition* to ensure the stability of the discretization (13), i.e., the uniform bounded invertibility of the mappings $\mathcal{L}_h$ from (14); see, e.g. [BF91].

Theorem 1 *Assume that the pairs (S^0_{kl}, M_{kl}) are admissible (i.e., that* **P.1 – P.4** *hold) and that the meshes $\mathcal{T}_k$ are shape regular and locally quasi-uniform. Then there exists a constant $\beta > 0$ depending only on the mesh parameters and on the constants c_0, c_1 in* **P.3** *and* **P.4**, *respectively, such that the pairs of spaces X_h, M_h defined above satisfy the LBB-condition*

$$\inf_{\mu\in M_h} \sup_{v\in X_h} \frac{b(v, \mu)}{\|v\|_{1,h}\|\mu\|_{-1/2}} \geq \beta. \tag{19}$$

The main ingredient in the proof of Theorem 1 is the following observation.

Lemma 1 *Under the hypotheses of Theorem 1 there exists for every $\mu \in M_{kl}$ an element $v^* \in S^0_{kl}$ such that*

$$(v^*, \mu)_{0,\Gamma_{kl}} \geq c\left(\|v^*\|^2_{H^{1/2}_{00}(\Gamma_{kl})} + \|\mu\|^2_{(H^{1/2}_{00}(\Gamma_{kl}))'}\right) \tag{20}$$

holds for some constant $c > 0$ independent of v^* and μ.

Proof Given any $\mu \in M_{kl}$, one can find, by definition, a $v \in H^{1/2}_{00}(\Gamma_{kl})$ such that

$$\|\mu\|_{(H^{1/2}_{00}(\Gamma_{kl}))'} \quad \leq \quad 2\frac{(v, \mu)_{0,\Gamma_{kl}}}{\|v\|_{H^{1/2}_{00}(\Gamma_{kl})}} = 2\frac{(v, Q_{kl}\mu)_{0,\Gamma_{kl}}}{\|v\|_{H^{1/2}_{00}(\Gamma_{kl})}} = 2\frac{(Q^*_{kl}v, \mu)_{0,\Gamma_{kl}}}{\|v\|_{H^{1/2}_{00}(\Gamma_{kl})}}.$$

Thus, setting $v^* := Q^*_{kl}v \in S^0_{kl}$, we conclude, in view of (10),

$$c_1^{-1}\|v^*\|_{H^{1/2}_{00}(\Gamma_{kl})}\|\mu\|_{(H^{1/2}_{00}(\Gamma_{kl}))'} \leq \|v\|_{H^{1/2}_{00}(\Gamma_{kl})}\|\mu\|_{(H^{1/2}_{00}(\Gamma_{kl}))'} \leq 2(v^*, \mu)_{0,\Gamma_{kl}},$$

which completes the proof. ∎

We are now ready to complete the

Proof of Theorem 1. Given $\mu \in M_h$ let μ_{kl} denote its component corresponding to $\Gamma_{kl} \subset \mathcal{S}$. We define a suitable $v \in X_h$ as follows. For each Γ_{kl} let $v_{kl} \in S^0_{kl}$ be the function constructed in Lemma 1 satisfying (20). Bearing in mind, that, by our notational convention, Ω_k denotes the non-mortar side of Γ_{kl}, we define $\hat{v}_{kl}$ to be the harmonic extension of the boundary data

$$\hat{v}_{kl}(x) = \begin{cases} v_{kl}(x) & \text{if } x \in \Gamma_{kl}, \\ 0 & \text{if } x \in \partial\Omega_k \setminus \Gamma_{lk}, \end{cases}$$

and define $v|_{\Omega_k} := \sum_{\Gamma_{kl} \subset \partial\Omega_k} \hat{v}_{kl}$ as the superposition of these extensions. In particular, v vanishes identically on any subdomain Ω_l that is never a non-mortar side. Hence

$$([v], \mu)_{0,\Gamma_{kl}} = (v_{kl}, \mu)_{0,\Gamma_{kl}} \gtrsim \|v_{kl}\|^2_{H^{1/2}_{00}(\Gamma_{kl})} + \|\mu\|^2_{(H^{1/2}_{00}(\Gamma_{kl}))'}. \tag{21}$$

This therefore implies also

$$\begin{aligned} \sum_{\Gamma_{kl} \subset \partial\Omega_k} ([v], \mu)_{0,\Gamma_{kl}} &\gtrsim \sum_{\Gamma_{kl} \subset \partial\Omega_k} \|\hat{v}_{kl}\|^2_{H^1(\Omega_k)} + \|\mu\|^2_{(H^{1/2}_{00}(\Gamma_{kl}))'} \\ &\gtrsim \|v\|^2_{H^1(\Omega_k)} + \sum_l \|\mu\|^2_{(H^{1/2}_{00}(\Gamma_{kl}))'}. \end{aligned} \tag{22}$$

Since clearly $\|Q^*_{kl} v\|_{H^{1/2}_{00}(\Gamma_{kl})} = \|v_{kl}\|_{H^{1/2}_{00}(\Gamma_{kl})}$ it follows that

$$\left(\sum_{\Gamma_{kl} \subset \mathcal{S}} \|v\|^2_{H^1(\Omega_k)} \right)^{1/2} + \left(\sum_{\Gamma_{kl} \subset \mathcal{S}} \|v_{kl}\|^2_{H^{1/2}_{00}(\Gamma_{kl})} \right)^{1/2} \geq \|v\|_{1,h}.$$

Combining (21) and (22), we have

$$\left(\sum_{\Gamma_{kl} \subset \mathcal{S}} \|\mu\|^2_{(H^{1/2}_{00}(\Gamma_{kl}))'} \right)^{1/2} \left\{ \left(\sum_{\Gamma_{kl} \subset \mathcal{S}} \|v\|^2_{H^1(\Omega_k)} \right)^{1/2} + \left(\sum_{\Gamma_{kl} \subset \mathcal{S}} \|v_{kl}\|^2_{H^{1/2}_{00}(\Gamma_{kl})} \right)^{1/2} \right\} \lesssim \sum_{\Gamma_{kl} \subset \mathcal{S}} ([v], \mu)_{0,\Gamma_{kl}},$$

and conclude that $b(v, \mu) \gtrsim \|v\|_{1,h} \|\mu\|_{-1/2}$. This establishes the validity of the LBB-condition. ∎

Error Estimates

We wish to discuss next the accuracy of the above discretizations. For simplicity we confine the discussion to piecewise linear trial spaces on the subdomains so that the approximation order is $n = 2$. Accordingly the approximation order of the multipliers is assumed to be $n - 1 = 1$. The higher order case can be treated analogously provided the solution u of the continuous problem (3) has enough regularity on each Ω_k. Moreover, we will always assume that the pairs (S^0_{kl}, M_{kl}) are admissible and that the meshes $\mathcal{T}_k$ are shape regular and locally

quasi-uniform. Let us denote by $\bar{h}_k$ the maximal mesh size in Ω_k. If $u|_{\Omega_k} \in H^2(\Omega_k)$, then one hopes that the discrete solution u_h of (13) satisfies an estimate of the type

$$\|u - u_h\|_{1,h}^2 \leq c \sum_{k=1}^{k_{max}} \bar{h}_k^2 \|u\|_{2,\Omega_k}^2. \tag{23}$$

We want to identify the essential obstructions encountered when going about an estimate of the type (24). The usual point of departure is Strang's second lemma, see e.g. [Bra01], p. 107 or [BDW99], which says that

$$\|u - u_h\|_{1,h} \leq c \left(\inf_{v_h \in V_h} \|u - v_h\|_{1,h} + \sup_{v_h \in V_h} \frac{\int_{\mathcal{S}} a \frac{\partial u}{\partial n} [v_h] ds}{\|v_h\|_{1,h}} \right). \tag{24}$$

Since $v_h \in V_h$, due to the orthogonality relation (12) we may subtract an arbitrary element $\mu_h \in M_h$ from the conormal derivative of u in the consistency error so that

$$\begin{aligned} \sum_{\Gamma_{kl} \in \mathcal{S}} (a \frac{\partial u}{\partial n}, [v_h])_{0,\Gamma_{kl}} &= \sum_{\Gamma_{kl} \in \mathcal{S}} (a \frac{\partial u}{\partial n} - \mu_h, [v_h])_{0,\Gamma_{kl}} \\ &\leq \sum_{\Gamma_{kl} \in \mathcal{S}} \|a \frac{\partial u}{\partial n} - \mu_h\|_{(H^{1/2}(\Gamma_{kl}))'} \|[v_h]\|_{1/2,\Gamma_{kl}}. \end{aligned} \tag{25}$$

We know from the trace theorem that $\|[v_h]\|_{1/2,\Gamma_{kl}} \leq c(\|v_h\|_{1,\Omega_k} + \|v_h\|_{1,\Omega_l})$. Moreover, when the M_{kl} have approximation order $n - 1 = 1$, a standard duality argument ensures that we can find a $\mu_h \in M_{kl}$ such that

$$\left\| a \frac{\partial u}{\partial n} - \mu_h \right\|_{(H^{1/2}(\Gamma_{kl}))'} \leq c \bar{h}_k \left\| a \frac{\partial u}{\partial n} \right\|_{1/2,\Gamma_{kl}} \leq c \bar{h}_k \|\nabla u\|_{1,\Omega_k} \leq c \bar{h}_k \|u\|_{2,\Omega_k},$$

where we have used the trace theorem again. Therefore, by using the Cauchy–Schwarz inequality, one obtains

$$\begin{aligned} \sum_{\Gamma_{kl} \in \mathcal{S}} (a \frac{\partial u}{\partial n}, [v_h])_{0,\Gamma_{kl}} &\leq c \left(\sum_{\Gamma_{kl} \in \mathcal{S}} \bar{h}_k^2 \|u\|_{2,\Omega_k}^2 \right)^{1/2} \|v_h\|_{1,\delta} \\ &\leq c \left(\sum_{\Gamma_{kl} \in \mathcal{S}} \bar{h}_k^2 \|u\|_{2,\Omega_k}^2 \right)^{1/2} \|v_h\|_{1,h}, \end{aligned}$$

so that the quotient in (24) is bounded by $c \left(\sum_{\Gamma_{kl} \in \mathcal{S}} \bar{h}_k^2 \|u\|_{2,\Omega_k}^2 \right)^{1/2}$.

It remains to establish an analogous bound for the approximation error $\inf_{v_h \in V_h} \|u - v_h\|_{1,h}$ in (24). To this end, note first that $[u - v_h] = -[v_h]$ and for $v_h \in V_h$ one has $Q_{kl}^*[u - v_h] = 0$. Hence,

$$\inf_{v_h \in V_h} \|u - v_h\|_{1,h} = \inf_{v_h \in V_h} \|u - v_h\|_{1,\delta}. \tag{26}$$

The right hand side of (26) has indeed been shown in [KLPV01] to be bounded by the right hand side of (23) under a certain assumption **M1**. This condition requires that the meshes

induced by $\mathcal{T}_k$ and $\mathcal{T}_l$ *match along* the boundary $\partial\Gamma_{kl}$ of the interface Γ_{kl}. Condition **M1** allows one to employ local extensions from $H_{00}^{1/2}(\Gamma_{kl})$ to deal with the constraints.

In order to avoid this constraint we prefer an alternative and start with an unconstrained approximation of u on each subdomain Ω_k. In fact, from the inf-sup condition in Theorem 1 above and Fortin's general argument [Bra01, Remark III.4.10] we conclude that the estimate in X_h yields an upper bound for the approximation in the kernel V_h,

$$\inf_{v_h\in V_h}\|u-v_h\|_{1,h}\le c\inf_{v_h\in X_h}\|u-v_h\|_{1,h}. \tag{27}$$

In this case, however, the full norm has to be used, i.e., the terms $\|Q_{kl}^*[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})}$ have to be estimated as well (in particular, when $[u-v_h]\notin H_{00}^{1/2}(\Gamma_{kl})$). Since $u\in H^2(\Omega)$, there are many ways to construct an approximation v_h in X_h such that

$$\|u-v_h\|_{1,\delta}^2\le c\sum_{k=1}^{k_{max}}\bar h_k^2\|u\|_{2,\Omega_k}^2, \tag{28}$$

e.g. Lagrange interpolants or Clément's quasi-interpolants would do. Thus, it remains to estimate the terms $\|Q_{kl}^*[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})}$ which are actually more problematic. Of course, the problem is that under the above assumptions $[u-v_h]$ is not necessarily in $H_{00}^{1/2}(\Gamma_{kl})$ so that one cannot directly bound $\|Q_{kl}^*[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})}$ by $\|[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})}$ (see (10) in **P.4**) and use then the trace theorem in order to ensure ultimately that

$$\|Q_{kl}^*[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})}\le c\left(\|u-v_h\|_{1,\Omega_k}+\|u-v_h\|_{1,\Omega_l}\right), \tag{29}$$

thereby obtaining again the same bound as in (28) and thus confirming (23). Therefore we will discuss next some instances where (29) is indeed true which incidentally will shed some light on the type of obstructions arising in the general case.

First of all, since $[u]=0$ we have $[u-v_h]\in H_{00}^{1/2}(\Gamma_{kl})$ if and only if $[v_h]\in H_{00}^{1/2}(\Gamma_{kl})$. For $d=2$ this can always be arranged by choosing v_h to interpolate u at the cross points (without requiring the whole mortar discretization to enforce continuity at cross points!) In the case $d=3$ this is not possible. This is exactly where condition **M1** in [KLPV01] comes into play which requires that the meshes in Ω_k and Ω_l match along $\partial\Gamma_{kl}$ so that $[v_h]$ can indeed be arranged to be in $H_{00}^{1/2}(\Gamma_{kl})$, e.g., by choosing v_h as the nodal interpolant. In this case (10) in **P.4** can be invoked to estimate

$$\|Q_{kl}^*[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})}\le c\|[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})}$$

which indeed leads to (29) and thus is an instance where (23) can be confirmed. Hence in summary, one way to ensure an estimate of the type (23) is to sacrifice some of the *mortar flexibility* by enforcing interface boundary matching condition **M1**.

On the other hand, one could hope that $[u-v_h]$ fails to be in $H_{00}^{1/2}(\Gamma_{kl})$ by such a small deviation so that the smoothing caused by the application of Q_{kl}^* keeps $\|Q_{kl}^*[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})}$ comparable to $\|[u-v_h]\|_{1/2,\Gamma_{kl}}$ which would again lead to (29). One way to pursue this line is to apply an inverse inequality

$$\|g_h\|_{H_{00}^{1/2}(\Gamma_{kl})}\le c\|h^{-1/2}g_h\|_{0,\Gamma_{kl}},\quad g_h\in S_{kl}^0, \tag{30}$$

which is to be understood as follows. Following [DFG+01] we denote by h a *mesh function*, namely the unique piecewise linear function that interpolates the maximal diameter of all triangles sharing the corresponding nodal point. Then estimates of the form (30) are established in [DFG+01]. Now denoting by h_k, h_l the mesh size functions induced on Γ_{kl} by the mesh on Ω_k and Ω_l, respectively, we have

$$\|Q_{kl}^*[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})} \leq c\|h_k^{-1/2}Q_{kl}^*[u-v_h]\|_{0,\Gamma_{kl}}. \tag{31}$$

Now if Q_{kl}^* were sufficiently local in the sense that the following condition
P.5:

$$\|h_k^{-s}Q_{kl}^*v\|_{0,\Gamma_{kl}} \leq c\|h_k^{-s}v\|_{0,\Gamma_{kl}}, \tag{32}$$

holds, this combined with (32) would allow us to infer from (31) that

$$\|Q_{kl}^*[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})} \leq c\|h_k^{-1/2}[u-v_h]\|_{0,\Gamma_{kl}}. \tag{33}$$

We recall that Ω_k is the non-mortar side. Arranging now for $v_k := v_h|_{\Omega_k}$, $v_l := v_h|_{\Omega_l}$ that the restrictions $v_k|_{\Gamma_{kl}}, v_l|_{\Gamma_{kl}}$ to Γ_{kl} are suitable local Clément approximations, standard arguments yield

$$\begin{aligned}\|h_k^{-1/2}[u-v_h]\|_{0,\Gamma_{kl}}^2 \quad \leq \quad & c\sum_{\tau\in\mathcal{T}_{kl}}\Big\{(h_{k,\tau}^{-1/2}h_{k,\tau}^{3/2})^2|u|_{3/2,\hat{\tau}}^2 + \\ & c\sum_{\tau'\in\mathcal{T}_l|_{\Gamma_{kl}},\tau'\cap\tau\neq\emptyset}(h_{k,\tau}^{-1/2}h_{l,\tau'}^{3/2})^2|u|_{3/2,\hat{\tau}'}^2\Big\},\end{aligned} \tag{34}$$

where $\hat{\tau}$ is the union of supports of basis functions overlapping τ. Thus introducing

$$\rho_{kl} := \max_{x\in\Gamma_{kl}}\sqrt{h_l(x)/h_k(x)}, \tag{35}$$

we obtain from (34) by summing over τ

$$\|h_k^{-1/2}[u-v_h]\|_{0,\Gamma_{kl}}^2 \leq \rho_{kl}^2\left(\bar{h}_k^2+\bar{h}_l^2\right)\|u\|_{3/2,\Gamma_{kl}}^2.$$

This estimate combined with the trace theorem

$$\|Q_{kl}^*[u-v_h]\|_{H_{00}^{1/2}(\Gamma_{kl})} \leq c\rho_{kl}\left(\bar{h}_k\|u\|_{2,\Omega_k}+\bar{h}_l\|u\|_{2,\Omega_l}\right), \tag{36}$$

also yields the estimate (23) upon summing over k.

Theorem 2 *Suppose that all the meshes $\mathcal{T}_k$ are shape regular and locally (not globally) quasi-uniform. If* **P.5** *holds, then the matching condition* **M1** *from [KLPV01] can indeed be abundoned to obtain still an estimate of the form (23), provided that there exist a uniform bound*

$$\rho_{kl} \leq c. \tag{37}$$

Of course, if the meshes are quasi-uniform, the above argument simplifies and one arrives at the situation considered in [BDW99, BD98]. Note also that (37) is a weak constraint that tends to be satisfied automatically when the meshes are determined by reasonable error

estimators since a possible singularity on or near an interface will affect a neighborhood on both sides.

Finally, it should be noted that estimates of the form (23) are ultimately of limited value when dealing with highly non-quasi-uniform meshes. In fact, they would be only useful when the solution u has deficient regularity so that the local H^2 norms (or even H^s-norms for $s < 2$) are not appropriate. This issue is beyond the scope of the present discussion and will be addressed elsewhere.

Dual Bases

We wish to apply our approach to the so called *dual bases mortar method* that has been proposed in [KLPV01] for $d = 3$ and in [Woh99a] for $d = 2$, see also [BP99] for a similar approach in the wavelet context. Note that the assumptions in [KLPV01] are phrased in a somewhat different way but Lemma 3.2 in [KLPV01] relates the requirements there closely to the present formulation **P.4**. Specifically, in [KLPV01] two types of Lagrange multiplier spaces M_h are discussed for piecewise linear trial functions in X_h. Finite volume discretizations on dual meshes are shown to satisfy **P.1 – P.4** where, however, **P.4** can only be ensured to hold under certain restrictions on local mesh size ratios.

In contrast, the so called dual bases mortar method realizes **P.1 – P.4** for *any* locally quasi-uniform shape regular meshes without any quantitative mesh constraints. Let us briefly recall the main ingredients.

The multiplier space M_{kl} is most conveniently defined with the aid of the following mapping F_{kl}. Let τ be any triangle in $\mathcal{T}_{kl}$ and let for any $v \in S^0_{kl}$ the values of v at the nodes x_i of τ be denoted by v_i. Then $F_{kl}v = w$ is defined as the unique piecewise linear function on $\mathcal{T}_{kl}$ whose restriction to τ is determined by its nodal values w_i, for $d = 3$ as follows:

(i) $w_i := 3v_i - v_r - v_s$ for all vertices $i \neq r \neq s$ of τ when none of these vertices belongs to $\partial\Gamma_{kl}$;

(ii) If exactly one vertex, say x_i lies on $\partial\Gamma_{kl}$ set $w_i := (v_r + v_s)/2$, $w_r := (5v_r - 3v_s)/2$, $w_s := (5v_s - 3v_r)/2$;

(iii) If exactly two vertices x_r, x_s belong to $\partial\Gamma_{kl}$ let $w_i = w_r = w_s := v_i$;

(iv) If all vertices of τ belong to $\partial\Gamma_{kl}$ set $w_i = w_r = w_s = v_q$ where x_q is the nearest interior node to τ.

Now let us denote by ϕ_i, $x_i \in \mathcal{N}_{kl}$, the standard piecewise linear basis functions for S^0_{kl} normalized by $\phi_i(x_r) = \delta_{i,r}$, where $\mathcal{N}_{kl}$ is the set of *interior* nodes of $\mathcal{T}_{kl}$. Let

$$\psi_i := F_{kl}\phi_i, \quad x_i \in \mathcal{N}_{kl}$$

and define $M_{kl} := \text{span}\,\{\psi_i : x_i \in \mathcal{N}_{kl}\}$. This yields

$$(\phi_i, \psi_j)_{0,\Gamma_{kl}} = 0, \quad i \neq j, \quad (\phi_i, \psi_i)_{0,\tau} = \frac{|\tau|}{3}, \quad x_i \in \tau, \tag{38}$$

so that

$$\dim M_{kl} = \dim \mathcal{S}^0_{kl}, \tag{39}$$

which is **P.1**. More precisely, one concludes from the above relations

$$(\phi_i, \psi_j)_{0,\Gamma_{kl}} = a_i \delta_{i,j}, \quad a_i := \frac{1}{3} \sum_{\tau : x_i \in \tau} |\tau| \sim h_i^{d-1}. \tag{40}$$

Thus one has the explicit representations

$$Q_{kl} w = \sum_{x_i \in \mathcal{N}_{kl}} (w, \phi_i)_{0,\Gamma_{kl}} a_i^{-1} \psi_i, \quad Q_{kl}^* v = \sum_{x_i \in \mathcal{N}_{kl}} (v, \psi_i)_{0,\Gamma_{kl}} a_i^{-1} \phi_i. \tag{41}$$

Since constants are easily seen to be locally reproduced in M_{kl}, it is now easy to verify the approximation property **P.2**. Likewise biorthogonality (40) easily leads to **P.3**, see also [KLPV01] while **P.4** has also been established already in Lemma 3.2 of [KLPV01]. Hence all the requirements **P.1 – P.4** hold in this case. Thus to apply the above reasoning we only have to discuss (32). For any $\tau \in \mathcal{T}_{kl}$ one has for $\sigma_i := \operatorname{supp} \phi_i = \operatorname{supp} \psi_i$, $\hat{\tau} := \bigcup \{\sigma_i : x_i \in \tau\}$

$$\|h^{-1/2} Q_{kl}^* g\|_{0,\tau} \quad \leq \quad c h_\tau^{-1/2} \sum_{x_i \in \tau} a_i^{-1} \|g\|_{0,\sigma_i} \|\psi_i\|_{0,\sigma_i} \|\phi_i\|_{0,\sigma_i} \leq c h_\tau^{-1/2} \|g\|_{0,\hat{\tau}},$$

where we have used that, in view of the L_∞ normalization of the basis functions ϕ_i, ψ_i, the local quasi-uniformity of $\mathcal{T}_{kl}$ and (40), $a_i^{-1} \|\psi_i\|_{0,\sigma_i} \|\phi_i\|_{0,\sigma_i} \sim 1$. Hence (32) follows from summing over $\tau \in \mathcal{T}_{kl}$. This confirms that **P.5** holds as well.

Corollary 1 *The error estimate (23) holds for the dual basis mortar method provided the meshes satisfy (37).*

If the Lagrange multiplier spaces did not have local dual bases either ϕ_i or ψ_i in (40) would have global support but would, by Demko's theorem, exhibit a certain exponential decay. This would still entail the validity of condition **P5** but under certain constraints on the local mesh size ratios, as expected.

References

[BB99]Faker Ben Belgacem. The mortar finite element method with Lagrange multipliers. *Numer. Math.*, 84(2):173–197, 1999.

[BD98]D. Braess and W. Dahmen. Stability estimates of the mortar finite element method for 3-dimensional problems. *East-West J. Numer. Math.*, 6(4):249–264, 1998.

[BDW99]Dietrich Braess, Wolfgang Dahmen, and Christian Wieners. A multigrid algorithm for the mortar finite element method. *SIAM J. Numer. Anal.*, 37:48–69, 1999.

[BF91]F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods.* Springer-Verlag, New-York, 1991.

[BMP94]Christine Bernardi, Yvon Maday, and Anthony T. Patera. A new non conforming approach to domain decomposition: The mortar element method. In Haim Brezis and Jacques-Louis Lions, editors, *Collège de France Seminar*. Pitman, 1994. This paper appeared as a technical report about five years earlier.

[BP99]Silvia Bertoluzza and Valerie Perrier. The mortar wavelet method. Technical Report 99–17, Istituto di Analisi Numerica del C.N.R. Pavia, 1999. ENUMATH 99, to appear.

[Bra01]Dietrich Braess. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, Cambridge, 2001. Second Edition.

[DFG+01]Wolfgang Dahmen, Birgit Faermann, Ivan Graham, Wolfgang Hackbusch, and Stefan Sauter. Inverse inequalities for non-quasiuniform meshes and applications to the mortar element method. Technical Report 201, IGPM, RWTH Aachen, 2001.

[KLPV01]Chisup Kim, Raytcho Lazarov, Joseph Pasciak, and Panayot Vassilevski. Multiplier spaces for the mortar finite element method in three dimensions. *SIAM J. Numer. Anal.*, 39:519–538, 2001.

[Woh99a]B. Wohlmuth. Discretization methods and iterative solvers based on domain decomposition. Technical report, Habilitation, Department of Mathematics, Augsburg, 1999.

[Woh99b]B. Wohlmuth. Hierarchical a posteriori error estimators for mortar finite element methods with Lagrange multipliers. *SIAM J. Numer. Anal.*, 36:1636–1658, 1999.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

4 A New Look at FETI

Susanne C. Brenner[1]

Introduction

The Finite Element Tearing and Interconnecting (FETI) Method is usually formulated in terms of matrices and vectors (cf. [FR91], [MT96], [PJF97], [Tez98], [RF99], [MTF99], [KW01] and the references therein). In this paper we give a coordinate-free formulation of the FETI method and construct a new FETI preconditioner in terms of this formulation, which enable us to analyze it within the additive Schwarz framework. We will present the ideas for a second-order model problem on a polyhedral domain $\Omega \subset \mathbb{R}^3$. Details of the analysis can be found in [Bre00] (which deals with the 2D analog) and a forthcoming paper on the 3D FETI preconditioner.

Let $\Omega_1, \ldots, \Omega_J$ be tetrahedra which form a quasi-uniform triangulation of Ω with mesh-size H. Each of these subdomains is the union of tetrahedra from the quasi-uniform triangulation $\mathcal{T}$ of Ω, the mesh-size of which is denoted by h. Let $V(\Omega) \subset H_0^1(\Omega)$ be the P_1 finite element space associated with $\mathcal{T}$. The model problem is:

Find $u \in V(\Omega)$ such that

$$a(u,v) = \int_\Omega fv\,dx \qquad \forall\, v \in V(\Omega)\,, \tag{1}$$

where $f \in L_2(\Omega)$ and the variational form $a(\cdot,\cdot)$ is defined by

$$a(v,w) = \sum_{j=1}^{J} a_j(v,w) \quad \text{and} \quad a_j(v,w) = \alpha_j \int_{\Omega_j} \nabla v \cdot \nabla w\,dx\,. \tag{2}$$

The coefficients $\alpha_1, \ldots, \alpha_J$ in (2) are positive constants.

For simplicity we assume that $\partial\Omega_j \cap \partial\Omega$ is not zero-dimensional. We say that Ω_j is (i) *anchored* if $\partial\Omega_j \cap \partial\Omega$ contains a face of Ω_j, (ii) *hinged* if $\partial\Omega_j \cap \partial\Omega$ contains an edge of Ω_j but no faces, and (iii) *floating* if $\partial\Omega_j \cap \partial\Omega = \emptyset$.

Remark 1 *The construction and analysis of the 3D preconditioner in this paper can be applied (with modifications) to the general case where $\Omega_1, \ldots, \Omega_J$ are nonoverlapping polyhedral subdomains which do not necessarily form a triangulation of Ω and whose boundaries can intersect $\partial\Omega$ in zero-dimensional sets.*

A Coordinate-Free Formulation of FETI

Let $\Gamma_j = \partial\Omega_j \setminus \partial\Omega$ and $\Gamma = \bigcup_{j=1}^{J} \Gamma_j$ be the interface of the subdomains. The space $V(\Gamma)$ ($\subset V(\Omega)$) of discrete harmonic functions is the orthogonal complement of the space $\{v \in V(\Omega) : v = 0 \text{ on } \Gamma\}$ with respect to $a(\cdot,\cdot)$. By solving (in parallel) a discrete Poisson equation on each subdomain, the problem (1) can be reduced to the following problem on the interface:

[1]Department of Mathematics,University of South Carolina, Columbia, SC 29208, USA.

Find $\bar{u} \in V(\Gamma)$ such that

$$a(\bar{u}, v) = \int_\Omega f v \, dx \qquad \forall\, v \in V(\Gamma)\,. \tag{3}$$

Let $V(\Gamma_j)$ be the space of discrete harmonic functions on Ω_j which vanish on $\partial\Omega_j \cap \partial\Omega$, and

$$\tilde{V} = V(\Gamma_1) \times V(\Gamma_2) \times \cdots \times V(\Gamma_J)\,.$$

Let $\mathcal{N}_j$ (resp. $\mathcal{N}_e$ or $\mathcal{N}_F$) be the set of nodes on Γ_j (resp. the open edge e or the open face F of a subdomain) and $\mathcal{N} = \bigcup_{j=1}^J \mathcal{N}_j$. For each $p \in \mathcal{N}$ we define σ_p to be the index set of the subdomains neighboring p, i.e.,

$$\sigma_p = \{1 \le j \le J : p \in \partial\Omega_j\}\,,$$

and for each $k, \ell \in \sigma_p$ we define $\mu_{p,k,\ell} \in \tilde{V}'$ (the dual space of $\tilde{V}$) by

$$\mu_{p,k,\ell}(\tilde{v}) = v_\ell(p) - v_k(p) \qquad \text{for} \quad \tilde{v} = (v_1, \ldots, v_J) \in \tilde{V}\,. \tag{4}$$

The subspace of $\tilde{V}'$ spanned by all such $\mu_{p,k,\ell}$'s is denoted by M_p and the space of Lagrange multipliers is $M = \sum_{p\in\mathcal{N}} M_p$.

In terms of M, which enforces the continuity along Γ, the interface problem (3) can be reformulated as:

Find $(\tilde{w}, \phi) \in \tilde{V} \times M$ such that

$$\sum_{j=1}^J a_j(w_j, v_j) + \langle \phi, \tilde{v}\rangle = \sum_{j=1}^J \int_{\Omega_j} f v_j \, dx \qquad \forall\, \tilde{v} \in \tilde{V}\,, \tag{5}$$

$$\langle \mu, \tilde{w}\rangle = 0 \qquad \forall\, \mu \in M\,, \tag{6}$$

where $\tilde{w} = (w_1, \ldots, w_J)$, $\tilde{v} = (v_1, \ldots, v_J)$ and $\langle\cdot,\cdot\rangle$ is the canonical bilinear form between a vector space and its dual space.

The solution of (3) is related to $\tilde{w}$ by $\bar{u}\big|_{\Omega_j} = w_j$ for $1 \le j \le J$.

Remark 2 *Throughout this paper we always keep elements of $V(\Gamma_j)'$ or $\tilde{V}'$ on the left-hand side of the canonical bilinear form $\langle\cdot,\cdot\rangle$, and members of $V(\Gamma_j)$, $\tilde{V}$ or their quotient spaces on the right-hand side.*

The FETI method solves (5)–(6) in the following way.

Let the Schur complement operator $S_j : V(\Gamma_j) \longrightarrow V(\Gamma_j)'$ be defined by

$$\langle S_j z_1, z_2\rangle = a_j(z_1, z_2) \qquad \forall\, z_1, z_2 \in V(\Gamma_j)\,. \tag{7}$$

Let $\mathrm{Ker}S_j = \{v \in V(\Gamma_j) : S_j v = 0\}$, $(\mathrm{Ker}S_j)^\perp = \{\psi \in V(\Gamma_j)' : \langle\psi, v\rangle = 0 \ \forall\, v \in \mathrm{Ker}S_j\}$, and the pseudo-inverse $S^+ : (\mathrm{Ker}S_j)^\perp \longrightarrow V(\Gamma)/\mathrm{Ker}S_j$ be defined by the following properties:

$$\begin{aligned}
\langle \psi_1, S_j^+ \psi_2\rangle &= \langle \psi_2, S_j^+ \psi_1\rangle && \forall\, \psi_1, \psi_2 \in (\mathrm{Ker}S_j)^\perp\,,\\
S_j S_j^+ \psi &= \psi && \forall\, \psi \in (\mathrm{Ker}S_j)^\perp\,,\\
S_j^+ S_j z &= \pi_j z && \forall\, z \in V(\Gamma_j)\,,
\end{aligned} \tag{8}$$

where $\pi_j : V(\Gamma_j) \longrightarrow V(\Gamma_j)/\mathrm{Ker}S_j$ is the canonical projection.

Remark 3 $\mathrm{Ker}\, S_j$ *is the space of constant functions for a floating* Ω_j *and* $\mathrm{Ker}\, S_j = \{0\}$ *for an anchored or a hinged* Ω_j*, in which case* $S_j^+ = S_j^{-1}$.

Let $\tilde{S} : \tilde{V} \longrightarrow \tilde{V}'$ be the product of the S_j's. Then

$$\mathrm{Ker}\tilde{S} = \mathrm{Ker}S_1 \times \cdots \times \mathrm{Ker}S_J$$

and the pseudo-inverse $\tilde{S}^+ : (\mathrm{Ker}\tilde{S})^\perp \longrightarrow \tilde{V}/\mathrm{Ker}\tilde{S}$ is the product of the S_j^+'s.

Let $\gamma_f \in \tilde{V}'$ be defined by

$$\langle \gamma_f, \tilde{v} \rangle = \sum_{j=1}^{J} \int_{\Omega_j} f v_j \, dx \qquad \forall\, \tilde{v} = (v_1, \ldots, v_J) \in \tilde{V}\,, \tag{9}$$

and let $\phi_0 \in M$ satisfy

$$\langle \phi_0, \tilde{v} \rangle = \langle \gamma_f, \tilde{v} \rangle \qquad \forall\, \tilde{v} \in \mathrm{Ker}\tilde{S}\,. \tag{10}$$

It follows from (5), (7), (9) and (10) that $\phi_* = \phi - \phi_0 \in M \cap (\mathrm{Ker}\tilde{S})^\perp$. Moreover, equation (5) can be written as

$$\tilde{S}\tilde{w} + \phi_* = \gamma_f - \phi_0\,. \tag{11}$$

Since $\gamma_f - \phi_0 \in (\mathrm{Ker}\tilde{S})^\perp$ by (10), we obtain from (8) and (11) the relation

$$\pi\tilde{w} + \tilde{S}^+\phi_* = \tilde{S}^+(\gamma_f - \phi_0)\,, \tag{12}$$

where $\pi : \tilde{V} \longrightarrow \tilde{V}/\mathrm{Ker}\tilde{S}$ is the product of the π_j's. Equation (6) and (12) then imply

$$\langle \psi, \tilde{S}^+\phi_* \rangle = \langle \psi, \tilde{S}^+(\gamma_f - \phi_0) \rangle \qquad \forall\, \psi \in M \cap (\mathrm{Ker}\tilde{S})^\perp\,. \tag{13}$$

Equation (13) is a symmetric positive definite (SPD) system on

$$F = M \cap (\mathrm{Ker}\tilde{S})^\perp$$

which determines ϕ_*. Once we have found ϕ_* (and hence $\phi = \phi_* + \phi_0$), then we can recover $\tilde{w}$ (and hence $\bar{u}$) in two steps. In the first step we determine (by a parallel solve) $\tilde{w}_* \in \tilde{V}$ with the property that

$$\pi\tilde{w}_* = \tilde{S}^+(\gamma_f - \phi)\,. \tag{14}$$

In the second step we find $\tilde{w}_0 \in \mathrm{Ker}\tilde{S}$ such that

$$\langle \mu, \tilde{w}_0 \rangle = -\langle \mu, \tilde{w}_* \rangle \qquad \forall\, \mu \in M\,. \tag{15}$$

Then $\tilde{w} = \tilde{w}_* + \tilde{w}_0$ and $\phi = \phi_* + \phi_0$ satisfy the system (5)–(6).

Equation (13) can be rewritten as

$$\mathbb{S}^+\phi_* = g_0\,,$$

where $\mathbb{S}^+ : F \longrightarrow F'$ and $g_0 \in F' = \tilde{V}/(M^\perp + \mathrm{Ker}\,\tilde{S})$ are defined by

$$\langle \psi, \mathbb{S}^+\eta \rangle = \langle \psi, \tilde{S}^+\eta \rangle \qquad \forall\, \psi, \eta \in F\,, \tag{16}$$

and $\langle \psi, g_0 \rangle = \langle \psi, \tilde{S}^+(\gamma_f - \phi_0) \rangle \quad \forall\, \psi \in F$. The operator $\mathbb{S}^+$ is therefore at the heart of the FETI method.

Remark 4 *The underdetermined system* (10) *is solvable since*

$$M^{\perp} \cap \operatorname{Ker} \tilde{S} = \{0\} . \tag{17}$$

The overdetermined system (15) *is solvable because* (13) *and* (14) *imply* $\langle \mu, \tilde{w}_* \rangle = 0 \; \forall \, \mu \in M \cap (\operatorname{Ker} \tilde{S})^{\perp}$. *Its solution is also unique because of* (17).

Additive Schwarz FETI Preconditioners

Let $R_j : \tilde{V}' = V(\Gamma_1)' \times \cdots \times V(\Gamma_J)' \longrightarrow V(\Gamma_j)'$ be the restriction map. Then R_j maps $(\operatorname{Ker} \tilde{S})^{\perp}$ into $(\operatorname{Ker} S_j)^{\perp}$ and we can express (16) as

$$\mathbb{S}^{+} = \sum_{j=1}^{J} (R_j \Phi)^t S_j^{+} (R_j \Phi) ,$$

where $\Phi : F \longrightarrow (\operatorname{Ker} \tilde{S})^{\perp}$ is the natural injection. It is therefore natural to precondition $\mathbb{S}^{+}$, which is a sum of SPD operators, by the sum of the "inverses" of the SPD operators, i.e., the FETI preconditioner $\mathbb{T} : F' \longrightarrow F$ should have the form

$$\mathbb{T} = \sum_{j=1}^{J} I_j S_j I_j^t , \tag{18}$$

where $I_j : (\operatorname{Ker} S_j)^{\perp} \longrightarrow F$ is "inverse" to $R_j \Phi$ in the sense that $\sum_{j=1}^{J} I_j R_j \Phi =$ the identity operator on F, i.e.,

$$\sum_{j=1}^{J} I_j R_j \phi = \phi \qquad \forall \, \phi \in F . \tag{19}$$

Remark 5 *It follows from* (18) *that* $\mathbb{T}$ *is an additive Schwarz preconditioner and hence can be analyzed by the well-known additive Schwarz theory (cf. [SBG96] and the references therein).*

We will define the operator I_j in terms of three operators.
For each $p \in \mathcal{N}_j$, we define $\delta_{p,j} \in V(\Gamma_j)$ by

$$\delta_{p,j}(q) = \begin{cases} 1 & \text{if } q = p , \\ 0 & \text{if } q \in \mathcal{N}_j \setminus \{p\} . \end{cases}$$

Note that $\{\delta_{p,j} : p \in \mathcal{N}_j\}$ is a basis of $V(\Gamma_j)$. For $t \in [1/2, \infty)$ we define $\mathbb{E}_j : V(\Gamma_j)' \longrightarrow M$ by

$$\mathbb{E}_j \psi = \sum_{p \in \mathcal{N}_j} \langle \psi, \delta_{p,j} \rangle \sum_{\ell \in \sigma_p} \alpha_\ell^t \mu_{p,\ell,j} , \tag{20}$$

and $\mathbb{D}_j : V(\Gamma_j)' \longrightarrow V(\Gamma_j)'$ by

$$\langle \mathbb{D}_j \eta, \delta_{p,j} \rangle = \langle \eta, \delta_{p,j} \rangle / \Big(\sum_{\ell \in \sigma_p} \alpha_\ell^t \Big) \qquad \forall \, p \in \mathcal{N}_j . \tag{21}$$

The operators $\mathbb{E}_j$ and $\mathbb{D}_j$ form a partition of unity with R_j on the space M:

$$\sum_{j=1}^{J} \mathbb{E}_j \mathbb{D}_j R_j \mu = \mu \qquad \forall\, \mu \in M\,. \tag{22}$$

Remark 6 *The operator $\mathbb{D}_j$ is a diagonal scaling operator which together with $\mathbb{E}_j$ forms an averaging process that yields (22) and also ensures the bound for the condition number of $\mathbb{T}\mathbb{S}^+$ is independent of the coefficients $\alpha_1, \ldots, \alpha_J$. This scaling technique is well-known.*

Finally we note that if $\mathbb{Q} : M \longrightarrow F$ is a projection operator and the map $I_j : (\mathrm{Ker}\, S_j)^\perp \longrightarrow F$ is defined by

$$I_j v_j = \mathbb{Q}\mathbb{E}_j \mathbb{D}_j v_j \qquad \forall\, v_j \in (\mathrm{Ker}\, S_j)^\perp\,, \tag{23}$$

then (19) follows from (22). Hence the crucial step in defining the additive Schwarz FETI preconditioner $\mathbb{T}$ is the construction of the projection $\mathbb{Q}$.

Let $\mu \in M$ and

$$\mathbb{Q}\mu = \mu - \mu_*\,. \tag{24}$$

Then $\mathbb{Q}$ is a projection from M onto F provided that $\mu \to \mu_*$ is linear and

$$\langle \mu_*, \tilde{v} \rangle = \langle \mu, \tilde{v} \rangle \qquad \forall\, \tilde{v} \in \mathrm{Ker}\, \tilde{S}\,, \tag{25}$$

$$\mu_* = 0 \qquad \text{if } \mu \in F\,. \tag{26}$$

Remark 7 *Once we have chosen a solution space for (25), we can take μ_* to be the member of the solution space that minimizes an appropriate inner product norm. This will automatically guarantee that $\mu \to \mu_*$ is linear and (26) is satisfied.*

A New 3D Preconditioner

Let $\mathcal{V}$ (resp. $\mathcal{E}$) be the set of vertices (resp. (open) edges) of floating subdomains and $M_{\mathcal{V}} = \sum_{p\in\mathcal{V}} M_p$. For $e \in \mathcal{E}$ we define

$$\sigma_e = \{1 \le j \le J : e \subset \partial\Omega_j\}\,,$$

and for $k, \ell \in \sigma_e$, we define

$$\mu_{e,k,\ell} = \frac{1}{|\mathcal{N}_e|} \sum_{p\in\mathcal{N}_e} \mu_{p,k,\ell}\,. \tag{27}$$

Let the space $M_e \subset M$ be generated by all such $\mu_{e,k,\ell}$'s and $M_{\mathcal{E}} = \sum_{e\in\mathcal{E}} M_e$.

The solution space of the projection equation (25) is then chosen to be $M_{\mathcal{W}} = M_{\mathcal{V}} + M_{\mathcal{E}}$, where the subscript $\mathcal{W}$ stands for *wire-basket*.

Remark 8 *Equation (25) is solvable in $M_{\mathcal{W}}$ because 0 is the only $\tilde{v} \in \mathrm{Ker}\, \tilde{S}$ annihilated by all $\mu_* \in M_{\mathcal{W}}$.*

According to Remark 7, we still need to introduce an appropriate inner product on $M_{\mathcal{W}}$ in order to complete the definition of $\mathbb{Q}$.

Let p (resp. e or F) be a vertex (resp. an open edge or an open face) of Ω_j, and $\tilde{1}_{p,j} \in \tilde{V}$ (resp. $\tilde{1}_{e,j} \in \tilde{V}$ or $\tilde{1}_{F,j} \in \tilde{V}$) be characterized by (i) the j-th component of $\tilde{1}_{p,j}$ (resp. $\tilde{1}_{e,j}$ or $\tilde{1}_{F,j}$) equals 1 at p (resp. the nodes in e or F) and vanishes at all the other nodes in $\mathcal{N}_j$, and (ii) all the other components of $\tilde{1}_{p,j}$ (resp. $\tilde{1}_{e,j}$ or $\tilde{1}_{F,j}$) vanish.

For $p \in \mathcal{V}$, $\mu_p \in M_p$ and $\tilde{v} = (c_1, \ldots, c_J) \in \operatorname{Ker} \tilde{S}$, where the c_j's are constants, we have

$$\langle \mu_p, \tilde{v} \rangle = \langle \mu_p, \sum_{j \in \sigma_p} c_j \tilde{1}_{p,j} \rangle \,. \tag{28}$$

It is easy to see (cf. (2) and (44) below) that

$$a\big(\sum_{j \in \sigma_p} c_j \tilde{1}_{p,j}, \sum_{j \in \sigma_p} c_j \tilde{1}_{p,j} \big) \approx h \sum_{j \in \sigma_p} \alpha_j c_j^2 \,. \tag{29}$$

Remark 9 *To avoid the proliferation of constants, we use the notation $A \lesssim B$ to represent the statement that $A \leq$ constant $\times B$, where the constant is independent of the mesh-sizes, the number of subdomains and the coefficients $\alpha_1, \ldots, \alpha_J$. The notation $A \approx B$ is equivalent to $A \lesssim B$ and $B \lesssim A$.*

In view of (28) and (29), it is natural to define

$$\|\mu_p\|_{M_p} = \sup_{\sum_{j \in \sigma_p} c_j^2 > 0} \frac{\langle \mu_p, \sum_{j \in \sigma_p} c_j \tilde{1}_{p,j} \rangle}{(h \sum_{j \in \sigma_p} \alpha_j c_j^2)^{1/2}} \qquad \forall\, \mu_p \in M_p \,. \tag{30}$$

Similarly, for $\mu_e \in M_e$ and $\tilde{v} = (c_1, \ldots, c_J) \in \operatorname{Ker} \tilde{S}$, we have

$$\langle \mu_e, \tilde{v} \rangle = \langle \mu_e, \sum_{j \in \sigma_e} c_j \tilde{1}_{e,j} \rangle \,. \tag{31}$$

It is again easy to see (cf. (2) and (44) below) that

$$a\big(\sum_{j \in \sigma_e} c_j \tilde{1}_{e,j}, \sum_{j \in \sigma_e} c_j \tilde{1}_{e,j} \big) \approx H \sum_{j \in \sigma_e} \alpha_j c_j^2 \,. \tag{32}$$

In view of (31) and (32), we define

$$\|\mu_e\|_{M_e} = \sup_{\sum_{j \in \sigma_e} c_j^2 > 0} \frac{\langle \mu_e, \sum_{j \in \sigma_e} c_j \tilde{1}_{e,j} \rangle}{(H \sum_{j \in \sigma_e} \alpha_j c_j^2)^{1/2}} \qquad \forall\, \mu_e \in M_e \,. \tag{33}$$

Since $\| \cdot \|_{M_p}$ and $\| \cdot \|_{M_e}$ are dual to inner product norms, they are also norms of inner products. If we define

$$\|\mu\|_{M_{\mathcal{W}}}^2 = \sum_{p \in \mathcal{V}} \|\mu_p\|_{M_p}^2 + \sum_{e \in \mathcal{E}} \|\mu_e\|_{M_e}^2 \qquad \forall\, \mu \in M_{\mathcal{W}} \,, \tag{34}$$

where $\mu = \sum_{p \in \mathcal{V}} \mu_p + \sum_{e \in \mathcal{E}} \mu_e$, $\mu_p \in M_p$ and $\mu_e \in M_e$, then $\| \cdot \|_{M_{\mathcal{W}}}$ is also an inner product norm.

We can now define the projection operator $\mathbb{Q}$ by (24), where $\mu_* \in M_{\mathcal{W}}$ is the solution of (25) with the minimum $M_{\mathcal{W}}$ norm. The preconditioner $\mathbb{T}$ is then given by (18), (20), (21) and (23).

Remark 10 *The computation of the minimum norm solution μ_* of (25) in the space $M_{\mathcal{W}}$ is the "coarse problem" that provides global communication among the subdomains.*

Note that both $\|\cdot\|_{M_p}$ and $\|\cdot\|_{M_e}$ can be computed without knowing the triangulation $\mathcal{T}$ if we use appropriate bases. The case of $\|\cdot\|_{M_e}$ is clear from (33) if we use the basis $\{\mu_{e,\ell_*,\ell} : \ell \in \sigma_e \setminus \{\ell_*\}, \ell_* = \min \sigma_e\}$. For $\|\cdot\|_{M_p}$ we use the basis $\{h^{1/2}\mu_{p,\ell_*,\ell} : \ell \in \sigma_p \setminus \{\ell_*\}, \ell_* = \min \sigma_p\}$. Let $\mu = \sum_{\ell\in\sigma_p\setminus\{\ell_*\}} \gamma_\ell(h^{1/2}\mu_{p,\ell_*,\ell})$. Using (4) we can rewrite the right-hand side of (30) as

$$\sup_{\sum_{j\in\sigma_p} c_j^2>0} \frac{h^{1/2}\sum_{\ell\in\sigma_p\setminus\{\ell_*\}} \gamma_\ell(c_\ell - c_{\ell_*})}{(h\sum_{j\in\sigma_p}\alpha_j c_j^2)^{1/2}} = \sup_{\sum_{j\in\sigma_p} c_j^2>0} \frac{\sum_{\ell\in\sigma_p\setminus\{\ell_*\}} \gamma_\ell(c_\ell - c_{\ell_*})}{(\sum_{j\in\sigma_p}\alpha_j c_j^2)^{1/2}},$$

which shows that $\|\cdot\|_{M_p}$ can indeed be computed without knowledge of $\mathcal{T}$.

Hence the coarse problem is mesh-independent and the coefficient matrix for the coarse problem can be computed and factorized once the Ω_j's and the α_j's are given. This process can be carried out prior to or simultaneous with the meshing of the subdomains, and the same factorization of the coarse problem can be applied to any triangulation of the subdomains.

Remark 11 *The complexity of the computation of the coefficient matrix for the coarse problem is the same as that of a finite element stiffness matrix, where each floating subdomain corresponds to a node and two such nodes are neighbors if they share a common vertex.*

Remark 12 *By construction, we have*

$$\|\mu_*\|_{M_{\mathcal{W}}} \le \|\mu_\dagger\|_{M_{\mathcal{W}}}\,, \tag{35}$$

where $\mu_\dagger \in M_{\mathcal{W}}$ is any solution of (25). In Section 4 we will construct a solution $\mu_\dagger = \sum_{p\in\mathcal{V}}\mu_{\dagger,p} + \sum_{e\in\mathcal{E}}\mu_{\dagger,e}$, where each $\mu_{\dagger,p} \in M_p$ (resp. $\mu_{\dagger,e} \in M_e$) depends only on the restriction of μ to $\bigcup_{j\in\sigma_p}\Gamma_j$ (resp. $\bigcup_{j\in\sigma_e}\Gamma_j$). Then $\|\mu_\dagger\|_{M_{\mathcal{W}}}$ provides local estimates for μ_* (and hence $\mathbb{Q}\mu$) which ensure that the condition number estimate for $\mathbb{T}\mathbb{S}^+$ is independent of J.

Condition Number Estimates

There is a simple estimate for $\lambda_{\min}(\mathbb{T}\mathbb{S}^+)$. Let $\phi \in F$ be arbitrary. We have

$$\langle\phi, \tilde{S}^+\phi\rangle = \sum_{j=1}^J \langle R_j\phi, S_j^+ R_j\phi\rangle$$

and hence, in view of (16) and (19),

$$\langle\phi, \mathbb{S}^+\phi\rangle = \sum_{j=1}^J \langle\nu_j, S_j^+\nu_j\rangle\,, \tag{36}$$

where $\nu_j = R_j\phi \in (\mathrm{Ker}\, S_j)^\perp$ and $\sum_{j=1}^J I_j\nu_j = \phi$. It then follows from (18), (36) and the additive Schwarz theory that

$$\lambda_{\min}(\mathbb{T}\mathbb{S}^+) \ge 1\,. \tag{37}$$

The following are useful formulas (cf. [MT96]) for $\langle\cdot, \tilde{S}^+\cdot\rangle$ and $\langle\cdot, S_j^+\cdot\rangle$.

Lemma 1 *The following estimates hold*:

$$\langle \mu, \tilde{S}^{+}\mu\rangle^{1/2} \approx \sup_{\tilde{v}\in\tilde{V}\setminus\operatorname{Ker}\tilde{S}} \frac{\langle \mu, \tilde{v}\rangle}{\left(\sum_{j=1}^{J}\alpha_j |v_j|^2_{H^1(\Omega_j)}\right)^{1/2}} \qquad \forall\, \mu \in (\operatorname{Ker}\tilde{S})^{\perp},$$

where $\tilde{v} = (v_1, v_2, \ldots, v_J)$, *and*

$$\langle \nu_j, S_j^{+}\nu_j\rangle^{1/2} \approx \sup_{v_j\in V(\Gamma_j)\setminus\operatorname{Ker} S_j} \frac{\langle \nu_j, v_j\rangle}{\alpha_j^{1/2}|v_j|_{H^1(\Omega_j)}} \qquad \forall\, \nu_j \in (\operatorname{Ker} S_j)^{\perp}.$$

Remark 13 *If* Ω_j *is a floating subdomain, then without loss of generality we may assume the* v_j *in Lemma 1 satisfy* $\int_{\Omega_j} v_j\, dx = 0$.

Lemma 1 enables us to employ the following estimates, which have been established in the study of 3D domain decomposition preconditioners (cf. [Dry88], [BPS89], [BX91], [DSW94], [DW95], [KW01] and the references therein).

Lemma 2 *Let* D *be a regular tetrahedron of diameter* H *and* v *be any discrete harmonic function with respect to a quasi-uniform subdivision of* D *with mesh-size* h. *Let* p (*resp.* e *or* F) *be a vertex* (*resp. an open edge or an open face*) *of* D, *and* v_p (*resp.* v_e *or* v_F) *be the discrete harmonic function that coincides with* v *at* p (*resp. the nodes in* e *or* F) *and vanishes at all the other nodes on* ∂D.

Then the following estimates hold provided v *vanishes on one of the faces or* $\int_D v\, dx = 0$:

$$\|v_e\|^2_{L_2(e)} \lesssim [1+\ln(H/h)]\,|v|^2_{H^1(D)}, \tag{38}$$

$$|v_p|^2_{H^1(D)} \lesssim |v|^2_{H^1(D)}, \tag{39}$$

$$|v_e|^2_{H^1(D)} \lesssim [1+\ln(H/h)]\,|v|^2_{H^1(D)}, \tag{40}$$

$$|v_F|^2_{H^1(D)} \lesssim [1+\ln(H/h)]^2|v|^2_{H^1(D)}. \tag{41}$$

If v *vanishes on one of the edges, then* (38) *and* (41) *remain valid, and the following estimates hold*:

$$|v_p|^2_{H^1(D)} \lesssim [1+\ln(H/h)]\,|v|^2_{H^1(D)}, \tag{42}$$

$$|v_e|^2_{H^1(D)} \lesssim [1+\ln(H/h)]^2|v|^2_{H^1(D)}. \tag{43}$$

In the special case where v *is the constant function* 1, *we have*

$$|1_p|^2_{H^1(D)} \approx h\,,\ |1_e|^2_{H^1(D)} \approx H \quad \textit{and} \quad |1_F|^2_{H^1(D)} \approx H[1+\ln(H/h)]\,. \tag{44}$$

Let $\nu_j \in (\operatorname{Ker} S_j)^{\perp}$ and $\phi = \sum_{j=1}^{J} I_j\nu_j$. From (23), (24) and Lemma 1 we have

$$\langle \phi, \tilde{S}^{+}\phi\rangle \lesssim \mathcal{S}_1 + \mathcal{S}_2\,, \tag{45}$$

$$\mathcal{S}_1 = \sup_{\tilde{v}\in\tilde{V}\setminus\operatorname{Ker}\tilde{S}} \frac{\langle \sum_{j=1}^{J}\mathbb{E}_j\mathbb{D}_j\nu_j, \tilde{v}\rangle^2}{\sum_{j=1}^{J}\alpha_j|v_j|^2_{H^1(\Omega_j)}}, \tag{46}$$

$$\mathcal{S}_2 = \sup_{\tilde{v}\in\tilde{V}\setminus\operatorname{Ker}\tilde{S}} \frac{\langle \mu_*, \tilde{v}\rangle^2}{\sum_{j=1}^{J}\alpha_j|v_j|^2_{H^1(\Omega_j)}}, \tag{47}$$

where $\mu_* \in M_{\mathcal{W}}$ is the minimum $M_{\mathcal{W}}$ norm solution of (25) with

$$\mu = \sum_{j=1}^{J} \mathbb{E}_j \, \mathbb{D}_j \, \nu_j \, . \tag{48}$$

The term $\mathcal{S}_1$ can be estimated by (20), (21), Remark 13, (39)–(43) and (46). The result is

$$\mathcal{S}_1 \lesssim [1 + \ln(H/h)]^2 \sum_{j=1}^{J} \langle \nu_j, S_j^+ \nu_j \rangle \, . \tag{49}$$

In view of Remark 12, we will estimate $\mathcal{S}_2$ by constructing a local solution of (25). For $e \in \mathcal{E}$ we define $\mu_{\dagger,e} \in M_e$ by

$$\langle \mu_{\dagger,e}, \tilde{1}_{e,j} \rangle = \langle \mu, \tilde{1}_{e,j} + \frac{1}{3} \sum_{F \in \mathcal{F}_{e,j}} \tilde{1}_{F,j} \rangle \qquad \forall \, j \in \sigma_e \, , \tag{50}$$

where μ is given by (48) and $\mathcal{F}_{e,j}$ is the set of the two faces of Ω_j which have e as an edge. We also define $\mu_{\dagger,p} = \mu_p$ for $p \in \mathcal{V}$, or equivalently,

$$\langle \mu_{\dagger,p}, \tilde{1}_{p,j} \rangle = \langle \mu, \tilde{1}_{p,j} \rangle \qquad \forall \, j \in \sigma_p \, . \tag{51}$$

Note that, for a floating subdomain Ω_j, the j-th component of the sum of $\tilde{1}_{e,j} + \frac{1}{3} \sum_{F \in \mathcal{F}_{e,j}} \tilde{1}_{F,j}$ over all six edges of Ω_j and $\tilde{1}_{p,j}$ over all four vertices of Ω_j is exactly the constant function 1. Hence $\mu_\dagger = \sum_{p \in \mathcal{V}} \mu_{\dagger,p} + \sum_{e \in \mathcal{E}} \mu_{\dagger,e}$ satisfies (25).

Remark 14 *Since μ_e and μ belong to the space M and the functions $\sum_{j \in \sigma_e} \tilde{1}_{e,j}$ and $\sum_{j \in \sigma_e} \big(\tilde{1}_{e,j} + \sum_{F \in \mathcal{F}_{e,j}} \frac{1}{3} \tilde{1}_{F,j} \big)$ are continuous on the interface Γ, we have*

$$\langle \mu_e, \sum_{j \in \sigma_e} \tilde{1}_{e,j} \rangle = 0 = \langle \mu, \sum_{j \in \sigma_e} \big(\tilde{1}_{e,j} + \sum_{F \in \mathcal{F}_{e,j}} \frac{1}{3} \tilde{1}_{F,j} \big) \rangle \, .$$

Therefore the overdetermined system (50) is consistent and, in view of (4) and (27), has a unique solution.

It follows from (20), (21), (30), (33), Lemma 1, (44), (48), (50), and (51) that

$$\begin{aligned} \|\mu_{\dagger,p}\|_{M_p}^2 &\lesssim \sum_{\partial\Omega_j \ni p} \langle \nu_j, S_j^+ \nu_j \rangle \, , \\ \|\mu_{\dagger,e}\|_{M_e}^2 &\lesssim [1 + \ln(H/h)] \sum_{\partial\Omega_j \supset e} \langle \nu_j, S_j^+ \nu_j \rangle \, , \end{aligned}$$

and hence, by (34),

$$\|\mu_\dagger\|_{M_{\mathcal{W}}}^2 \lesssim [1 + \ln(H/h)] \sum_{j=1}^{J} \langle \nu_j, S_j^+ \nu_j \rangle \, . \tag{52}$$

Let $\tilde{v} = (v_1, \ldots, v_J) \in \tilde{V}$ be arbitrary. Definitions (27), (30), (33), (34) and the Cauchy-Schwarz inequality imply

$$\langle \mu_*, \tilde{v} \rangle \lesssim \Big(\sum_{p \in \mathcal{V}} \|\mu_{*,p}\|^2_{M_p} \Big)^{1/2} \Big[h \sum_{p \in \mathcal{V}} \sum_{j \in \sigma_p} \alpha_j v_j(p)^2 \Big]^{1/2} + \Big(\sum_{e \in \mathcal{E}} \|\mu_{*,e}\|^2_{M_e} \Big)^{1/2} \Big[H \sum_{e \in \mathcal{E}} \sum_{j \in \sigma_e} \alpha_j \bar{v}^2_{e,j} \Big]^{1/2}, \tag{53}$$

where $\bar{v}_{e,j} = |\mathcal{N}_e|^{-1} \sum_{p \in \mathcal{N}_e} v_j(p)$ is the mean nodal value of v_j on e. It follows from Remark 13, (38), (39), (42), (44) and the Cauchy-Schwarz inequality that

$$h v_j(p)^2 \lesssim |(v_j)_p|^2_{H^1(\Omega_j)} \lesssim [1 + \ln(H/h)] \, |v_j|^2_{H^1(\Omega_j)} , \tag{54}$$

$$H \bar{v}^2_{e,j} \lesssim \|(v_j)_e\|^2_{L_2(e)} \lesssim [1 + \ln(H/h)] \, |v_j|^2_{H^1(\Omega_j)} . \tag{55}$$

Combining (53), (54) and (55), we find

$$\langle \mu_*, \tilde{v} \rangle^2 \lesssim [1 + \ln(H/h)] \, \|\mu_*\|^2_{M_{\mathcal{W}}} \sum_{j=1}^{J} \alpha_j |v_j|^2_{H^1(\Omega_j)} ,$$

which together with (35), (47) and (52) yield

$$\mathcal{S}_2 \lesssim [1 + \ln(H/h)]^2 \sum_{j=1}^{J} \langle \nu_j, S_j^+ \nu_j \rangle . \tag{56}$$

Finally we conclude from (16), (45), (49) and (56),

$$\langle \phi, \mathbb{S}^+ \phi \rangle \lesssim [1 + \ln(H/h)]^2 \sum_{j=1}^{J} \langle \nu_j, S_j^+ \nu_j \rangle \tag{57}$$

whenever $\nu_j \in (\operatorname{Ker} S_j)^\perp$ for $1 \le j \le J$ and $\phi = \sum_{j=1}^{J} I_j \nu_j$. It then follows from (18), (57) and the additive Schwarz theory that

$$\lambda_{\max}(\mathbb{T}\mathbb{S}^+) \lesssim [1 + \ln(H/h)]^2. \tag{58}$$

Combining (37) and (58), we have the following theorem on the condition number $\kappa(\mathbb{T}\mathbb{S}^+)$.

Theorem 1 *There exists a positive constant C, independent of h, H, J and the α_j's, such that*

$$\kappa(\mathbb{T}\mathbb{S}^+) \le C[1 + \ln(H/h)]^2.$$

Acknowledgment

This work was supported in part by the National Science Foundation under Grant No. DMS-0074246.

References

[BPS89]James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. The construction of preconditioners for elliptic problems by substructuring, IV. *Math. Comp.*, 53:1–24, 1989.

[Bre00]Susanne C. Brenner. An additive Schwarz preconditioner for the FETI method. Technical Report IMIP 2000:08, University of South Carolina, Department of Mathematics, 2000.

[BX91]James H. Bramble and Jinchao Xu. Some estimates for a weighted L^2 projection. *Math. Comp.*, 56:463–476, 1991.

[Dry88]Maksymilian Dryja. A method of domain decomposition for 3-D finite element problems. In Roland Glowinski, Gene H. Golub, Gérard A. Meurant, and Jacques Périaux, editors, *First International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 43–61, Philadelphia, PA, 1988. SIAM.

[DSW94]Maksymilian Dryja, Barry F. Smith, and Olof B. Widlund. Schwarz analysis of iterative substructuring algorithms for elliptic problems in three dimensions. *SIAM J. Numer. Anal.*, 31(6):1662–1694, December 1994.

[DW95]Maksymilian Dryja and Olof B. Widlund. Schwarz methods of Neumann-Neumann type for three-dimensional elliptic finite element problems. *Comm. Pure Appl. Math.*, 48(2):121–155, February 1995.

[FR91]Charbel Farhat and Francois-Xavier Roux. A Method of Finite Element Tearing and Interconnecting and its Parallel Solution Algorithm. *Int. J. Numer. Meth. Engng.*, 32:1205–1227, 1991.

[KW01]Axel Klawonn and Olof B. Widlund. FETI and Neumann–Neumann Iterative Substructuring Methods: Connections and New Results. *Comm. Pure Appl. Math.*, 54:57–90, January 2001.

[MT96]Jan Mandel and Radek Tezaur. Convergence of a Substructuring Method with Lagrange Multipliers. *Numer. Math.*, 73:473–487, 1996.

[MTF99]Jan Mandel, Radek Tezaur, and Charbel Farhat. A Scalable Substructuring Method by Lagrange Multipliers for Plate Bending Problems. *SIAM J. Numer. Anal.*, 36(5):1370–1391, 1999.

[PJF97]K.C. Park, M.R. Justino, and C.A. Felippa. An algebraically partitioned FETI method for parallel structural analysis: algorithm description. *Int. J. Num. Meth. Engng.*, 40:2717–2737, 1997.

[RF99]Daniel Rixen and Charbel Farhat. A simple and efficient extension of a class of substructure based preconditioners to heterogeneous structural mechanics problems. *Int. J. Numer. Meth. Engng.*, 44:489–516, 1999.

[SBG96]Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[Tez98]Radek Tezaur. *Analysis of Lagrange Multiplier Based Domain Decomposition*. PhD thesis, University of Colorado at Denver, Department of Mathematics, 1998.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

5 Aitken-Schwarz algorithm on Cartesian grid

M. Garbey , D. Tromeur-Dervout [1] [2]

Introduction

This paper is devoted to the generalization of the Aitken-Schwarz (AS) domain decomposition method (resp. Steffensen-Schwarz (SS)) method introduced in [GTD01]. A solver was first designed to solve linear (resp. nonlinear) elliptic problems in metacomputing framework with a slow communication network. In [GTD01] the domain decomposition was one dimensional domain decomposition of multidimensionnal problems. We extend this domain decomposition to multilevel one dimensional AS (resp. SS) domain decomposition. The AS (resp. SS) method is recursively applied in one different direction at each level. The difficulty is to generate homogeneous Dirichlet boundary conditions at each level of domain decomposition. This problem is solved in AS domain decomposition with the superposition principle when linear problems are solved. A similar shifting technique is also adopted to solve nonlinear problems with SS. Some results on 2D linear and nonlinear problems are given as examples.

The arithmetical complexity of AS is investigated when the inner solver has linear or nonlinear complexity. Notably, a comparison with the best implementation of a fast solvers such as FFT on Poisson problem are given. The stability of the Aitken-Schwarz and Steffensen-Schwarz multilevel domain decomposition methods is investigated with an extensive sensitivity analysis experiment that measures the influence on the convergence history when one systematically perturbed randomly the subproblem solution at the end of each subdomain solve.

The plan of this paper is as follows: section 1 recalls the principles of the Aitken-Schwarz domain decompositions, section 2 describes the extension of the methodology from one dimensional domain decomposition to domain decomposition in several space directions, section 3 comments on the arithmetical complexity of the method, and section 4 comments on the stability of the method. Finally, section 5 gives the conclusions and perspectives.

1 Principles of the Aitken-Schwarz method

We are going to describe briefly the numerical ideas behind the Aitken Schwarz method. We refer to [GTD01] for more details.

For simplicity, we illustrate the concept with the discretized Helmholtz operator $L[u] = \Delta u - \lambda u,\ \lambda > 0$, with a grid that is a tensorial product of one dimensional grids, and a square domain decomposed into strip subdomains.

Let us consider the homogeneous Dirichlet problem $L[U] = f\ in\ \Omega = (0,1),\ U_{|\partial\Omega} = 0$, in one space dimension. We restrict ourselves to a decomposition of Ω into two overlapping

[1] This work was supported by the Région Rhône Alpes
[2] CDCSP/ISTIL - University Lyon 1, 69622 Villeurbanne, France
{garbey,dtromeur}@cdcsp.univ-lyon1.fr, http://cdcsp.univ-lyon1.fr

subdomains $\Omega_1 \bigcup \Omega_2$ and consider the additive Schwarz algorithm [Sch80, Lio88, Lio89].

$$L[u_1^{n+1}] = f \; in \; \Omega_1, \; u_{1|\Gamma_1}^{n+1} = u_{2|\Gamma_1}^{n}, \; L[u_2^{n+1}] = f \; in \; \Omega_2, \; u_{2|\Gamma_2}^{n+1} = u_{1|\Gamma_2}^{n}. \tag{1}$$

with given initial conditions $u_{1|\Gamma_2}^0, u_{2|\Gamma_1}^0$ to start this iterative process.

To simplify the presentation, we assume implicitly in our notations that the homogeneous Dirichlet boundary conditions are satisfied by all intermediate subproblems. This algorithm can be executed in parallel on two computers [Kuz91]. At the end of each subdomain solve, the artificial interfaces $u_{2|\Gamma_1}^n$ and $u_{1|\Gamma_2}^n$ have to be exchanged between the two computers.

In order to avoid as much as possible redundancy in the computation we fix once and for all the overlap between subdomains to be the minimum, i.e of size one mesh. This algorithm can be extended to an arbitrary number of subdomains and is nicely scalable, because the communications linked only subdomains that are neighbors.

However it is one of the worst numerical algorithms to solve the problem, because the convergence is extremely slow. We introduce thereafter a modified version of this Schwarz algorithm so called Aitken-Schwarz that transforms this dead slow iterative solver into a direct fast solver while keeping the scalability of the Schwarz algorithm for a moderate number of subdomains. The idea is as follows.

We observe that the interface operator T,

$$(u_{1|\Gamma_1}^n - U_{\Gamma_1}, u_{2|\Gamma_2}^n - U_{\Gamma_2})^t \rightarrow (u_{1|\Gamma_1}^{n+1} - U_{\Gamma_1}, u_{2|\Gamma_2}^{n+1} - U_{\Gamma_2})^t \tag{2}$$

is *linear*.

Therefore, the sequence $(u_{1|\Gamma_1}^n, u_{2|\Gamma_2}^n)$ has pure linear convergence that is, it satisfies the identities:

$$u_{1|\Gamma_2}^{n+1} - U_{|\Gamma_2} = \delta_1 (u_{2|\Gamma_1}^n - U_{|\Gamma_1}), \; u_{2|\Gamma_1}^{n+1} - U_{|\Gamma_1} = \delta_2 (u_{1|\Gamma_2}^n - U_{|\Gamma_2}), \tag{3}$$

where δ_1 (resp. δ_2) is the damping factor associated to the operator L in subdomain Ω_1 (resp. Ω_2) [GH97]. Consequently

$$u_{1|\Gamma_2}^2 - u_{1|\Gamma_2}^1 = \delta_1 (u_{2|\Gamma_1}^1 - u_{2|\Gamma_1}^0), \; u_{2|\Gamma_1}^2 - u_{2|\Gamma_1}^1 = \delta_2 (u_{1|\Gamma_2}^1 - u_{1|\Gamma_2}^0), \tag{4}$$

So except if the initial boundary conditions match with the exact solution U at the interfaces, the amplification factors can be computed from the linear system(4). Since $\delta_1 \delta_2 \neq 1$ the limit $U_{|\Gamma_i}, i = 1, 2$ is obtained as the solution of the linear system (3). Consequently, this generalized Aitken acceleration procedure gives the *exact* limit of the sequence on the interface Γ_i based on two successive Schwarz iterates $u_{i|\Gamma_i}^n, \; n = 1, 2$, and the initial condition $u_{i|\Gamma_i}^0$. An additional solve of each subproblem (1) with boundary conditions $u_{\Gamma_i}^\infty$ gives the final solution of the ODE problem. We can further improve this first algorithm as follows.

Let (v_1, v_2) be the solution of

$$L[v_1] = 0 \; in \; \Omega_1, \; v_{|\Gamma_1} = 1; \; L[v_2] = 0 \; in \; \Omega_2, \; v_{|\Gamma_2} = 1. \tag{5}$$

We have then $\delta_1 = v_{|\Gamma_2}, \; \delta_2 = v_{|\Gamma_1}$. Consequently δ_1 and δ_2 can be computed before-hand numerically or analytically.

Once (δ_1, δ_2) are known, we need only *one* Schwarz iterate to accelerate the interface and an additional solves for each subproblems. This is a total of two solves per subdomain. The Aitken acceleration thus transforms the additive Schwarz procedure into an *exact* solver

regardless of the speed of convergence of the original Schwarz method, and in particular with a minimum overlap.

This Aitken-Schwarz algorithm can be reproduced for multidimensional problems. As a matter of fact, it can be shown [GTD01] that the coefficients of each wave number of the sine expansion of the trace of the solution generated by the Schwarz algorithm has its own rate of *exact* linear convergence.

We can then generalize the one dimensional algorithm to two space dimensions as follows:

- step1 : compute analytically or numerically in parallel each damping factor δ_j^k for each wave number k from the two point one D boundary value problems analogues of (5) with the operator

$$L_k \equiv u_{xx} - (4/h_y^2 \sin^2(k\frac{h_y}{2}) + \lambda)u,$$

  with h_y being the space step in y direction.

- step2: apply one additive Schwarz iterate to the Helmholtz problem with subdomain solver of choice (multigrid, fast Fourier transform, PDC3D, etc...)

- step3:

  - compute the sine expansion $\hat{u}^n_{k|\Gamma_i},\ n = 0,1,\ k = 1..N$ of the traces on the artificial interface $\Gamma_i, i = 1..2$ for the initial boundary condition $u^0_{|\Gamma_i}$ and the solution given by *one* Schwarz iterate $u^1_{|\Gamma_i},\ i = 1,2$.

  - apply generalized Aitken acceleration *separately* to each wave coefficients in order to get $\hat{u}^\infty_{j|\Gamma_i}$.

  - recompose the trace $u^\infty_{j|\Gamma_i}$ in physical space.

- step4: compute in parallel the solution in each subdomains $\Omega_i,\ i = 1,2$ with new inner BCs and subdomain solver of choice.

So far, we have restricted ourselves to domain decomposition with two subdomains. We show in [GTD01], that a generalized Aitken acceleration technique can be applied to an arbitrary number $q > 2$ of subdomains with strip domain decomposition. Our main result is that no matter is the number of subdomains, the total number of subdomain solves required to produce the final solution is still **two**.

However the generalized Aitken acceleration of the vectorial sequences of the sine expansion coefficients of the interface introduces a coupling between all interfaces.

To be more specific, we obtain a given linear system for each wave number k,

$$\tilde{u}^\infty = (Id - P_k)^{-1}(\tilde{u}^{n+1} - P_k\tilde{u}^n). \tag{6}$$

and P_k has the following pentadiagonal structure:

$$\begin{vmatrix}
0 & \delta_1^r & 0 & 0 & .... & & & \\
\delta_2^{l,l} & 0 & 0 & \delta_2^{l,r} & ... & & & \\
\delta_2^{r,l} & 0 & 0 & \delta_2^{r,r} & ... & & & \\
 & & & ... & \delta_{q-1}^{l,l} & 0 & 0 & \delta_{q-1}^{l,r} \\
 & & & ... & \delta_{q-1}^{r,l} & 0 & 0 & \delta_{q-1}^{r,r} \\
 & & & ... & 0 & 0 & \delta_q^r & 0
\end{vmatrix}$$

But we observe first that this generalized Aitken acceleration processes independently each waves coefficients of the sinus expansion of the interfaces. Second the highest is the frequency k the smallest are the damping factors $\delta_j^{l,l}$, $\delta_j^{l,r}$, $\delta_j^{r,l}$, $\delta_j^{r,r}$. A careful stability analysis of the method shows that

- for low frequencies, we should use the generalized Aitken acceleration coupling all the subdomains.
- for intermediate frequencies, we can neglect this global coupling and implement only the local interaction between subdomains that overlap.
- for high frequencies, we do not use Aitken acceleration because one iteration of the Schwarz algorithm damps the high frequencies error enough.

The algorithm has then the same structure than the two subdomains algorithm presented above. Step 1 and step 4 are fully parallel. Step 2 requires only local communication and scales well with the number of processors. Step 3 requires global communication of interfaces in Fourier space for low wave numbers, and local communications for intermediate frequencies. In addition for moderated number of subdomains, the arithmetic complexity of step 3 that is the kernel of the method is negligible compared to step 2.

Our algorithm can be extended successfully to grids that are tensorial product of one dimensional grids with arbitrary (irregular) space step [BGO00], iterative domain decomposition method such as Dirichlet-Neumann procedure with non-overlapping subdomains or red/black subdomains iterative procedure.

For nonlinear elliptic problems, the Aitken acceleration is no longer exact. the so-called Steffensen-Schwarz variant is then a very efficient numerical method for low order perturbation of constant coefficient linear operators - once again we refer to [GTD01] for more details. We will proceed now with the description of the generalization of the method to domain decomposition in more than one space directions.

2 Multilevel Aitken-Schwarz and Steffensen-Schwarz Domain Decomposition

Let us consider first the linear case and denote L the discrete linear differential operator. For simplicity, we will restrict this presentation to problems in two space dimensions. Once again, we assume homogeneous Dirichlet Boundary conditions on domain Ω. We introduce a first level of domain decomposition into strips in direction x

$$\Omega = \bigcup_{i=1,n_x} \Omega_i,$$

where the $\Omega_i = (x_{i,l}, x_{i,r}) \times (0, \pi)$ are the overlapped rectangles represented in Figure 1.

To proceed with a two dimensional domain decomposition, we introduce a second level of domain decomposition and decompose each subdomain Ω_i into a set of overlapping rectangles in direction y,

$$\Omega_i = \bigcup_{j=1,n_y} \Omega_{i,j},$$

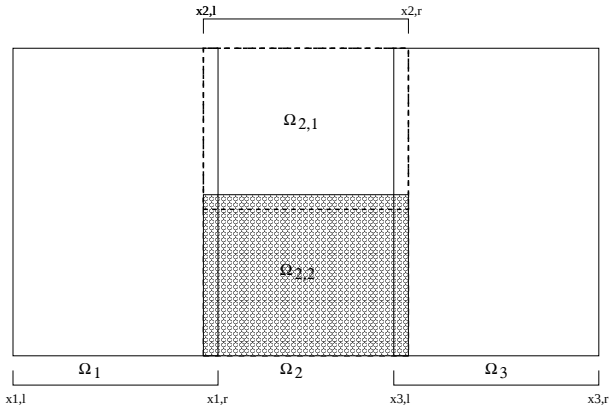


Figure 1: Multilevel Aitken-Schwarz Method principle

The main idea is to apply recursively on each subdomain decomposition level the Aitken-Schwarz algorithm. The difficulty comes from the fact that the Dirichlet boundary conditions of the subdomain at the first level are no more homogeneous Dirichlet boundary conditions. Consequently, the sine expansion operator should not be applied directly to the trace of the interfaces solution generated by this second level of the Schwarz algorithm.

We introduce therefore a shift denoted v_i in each subdomain Ω_i, in order to retrieve the homogeneous Dirichlet boundary conditions problem on each strip Ω_i.

Let $\otimes$ be the notation for the Kronecker product. In each strip Ω_i, we solve with Aitken-Schwarz the modified problem

$$L[w_i^{n+1}] = f - L[v_i] \ in \ \Omega_i \tag{7}$$

$$w_i^{n+1} = 0 \ in \ \partial\Omega_i \tag{8}$$

where v_i in matrix notation is defined as

$$v_i = d_i^{-1}.X_l \otimes u_{x=x_{i,l}} + d_i^{-1}.X_r \otimes u_{x=x_{i,r}}, \tag{9}$$

with d_i the size of the strip Ω_i in x direction: $d_i = x_{i,l} - x_{i,r}$, $x_i = (x_{i,l}, .., x_{i,r})$ row vector of the x coordinates of the grid points in $\bar{\Omega}_i$ in increasing order, $X_l = x_i - (x_{i,l}, .., x_{i,l})$ and $X_r = x_i - (x_{i,r}, .., x_{i,r})$, and $u_{x=x_{i,l}}$, $u_{x=x_{i,r}}$ are the column vectors containing the artificial boundary condition.

Table 1 gives the error between the Aitken-Schwarz solution and the discretized exact solution $u(x, y) = (x^2 - 0.25)y(y - 1)$ in maximum norm for a number of subdomains n_x in x-direction varying from 1 to 16 and a number of subdomains n_y in y-direction varying from 2 to 16 and for four global size meshes varying from 34×34 to 258×258 points for the Poisson problem. It exhibits that :

- the methodology gives accurate results close to the machine accuracy (we recall that the test are done with the Matlab software),

- the accuracy reached increases with the number of subdomains especially for large size problem. This is due to the fact that the local system are smaller leading to smaller conditioning number and then round off error in the LU factorization are smaller than in the few-subdomain case.

34×34 points	n_x subdomains				
n_y subdomains	1	2	4	8	16
2	1.9920e-13	2.9296e-14	9.3051e-15	4.0246e-16	7.0777e-16
4	1.2676e-13	2.2225e-14	2.8172e-15	4.0246e-16	7.0777e-16
8	4.2848e-15	1.1623e-15	4.7184e-16	4.0246e-16	7.0777e-16
16	8.5522e-16	2.2421e-16	4.0246e-16	4.0246e-16	7.0777e-16

66×66 points	n_x subdomains				
n_y subdomains	1	2	4	8	16
2	1.0693e-12	3.8589e-13	6.1952e-14	9.1593e-15	1.1380e-15
4	9.3924e-13	4.1277e-13	4.2577e-14	1.1005e-14	1.1380e-15
8	5.3798e-13	1.0418e-13	3.6227e-14	2.2204e-15	1.1102e-15
16	4.8329e-15	2.1164e-15	1.2906e-15	2.1649e-15	1.1380e-15

130×130 points	n_x subdomains				
n_y subdomains	1	2	4	8	16
2	7.7432e-12	2.3652e-12	1.2857e-12	8.1442e-14	1.6535e-14
4	6.2317e-12	1.6914e-12	6.3427e-13	1.3916e-13	1.9729e-14
8	2.6878e-12	1.2031e-12	3.5410e-13	1.7667e-13	7.2026e-15
16	1.0170e-12	4.1206e-13	5.3402e-14	2.3787e-14	5.9119e-15

258×258 points	n_x subdomains				
n_y subdomains	1	2	4	8	16
2	4.3664e-11	1.3513e-11	2.1562e-12	1.6181e-12	1.6694e-13
4	2.9052e-11	6.6053e-12	2.8467e-12	1.1419e-12	3.3179e-13
8	2.0213e-11	6.7309e-12	4.5298e-12	1.3563e-12	2.1001e-13
16	1.9160e-11	4.4557e-12	1.8232e-12	3.2840e-13	8.4238e-14

Table 1: Error with respect of the number of subdomains

Secondly, let us consider the nonlinear case. The problem to be solved can be written formally as

$$A(u)u = F \tag{10}$$

We do not have anymore the superposition principle as in the linear case, but we can still use the same shift to recover at the first level of domain decomposition homogeneous Dirichlet boundary conditions. We set:

$$A[w_i^{n+1} + v_i] = F \; in \; \Omega_i \tag{11}$$

$$w_i^{n+1} = 0 \; in \; \partial\Omega_i \tag{12}$$

where v_i is defined as in (9).

The solution for one Schwarz iterate on the subdomain Ω_i is obtained as

$$u_i^{n+1} = w_i^{n+1} + v_i \tag{13}$$

To illustrate the two level domain decomposition algorithm, we consider the Bratu problem which represents a simplified model of combustion written as follows:

$$\Delta u(x,y) + exp(\lambda u(x,y)) = 0, \;\; (x,y) \in \Omega = [0,1]^2, \lambda \geq 0, \tag{14}$$

$$u(x,y) = 0, \;\; (x,y) \in \partial\Omega. \tag{15}$$

The discrete operator on a regular stencil of space step h_x in x direction and h_y in y direction is:

$$-\frac{u_{i+1,j} + u_{i-1,j} - 2u_{i,j}}{h_x^2} - \frac{u_{i,j+1} + u_{i,j-1} - 2u_{i,j}}{h_y^2} + exp(\lambda u_{i,j}).$$

This operator is a nonlinear and nonseparable discrete operator. We use a Newton scheme to solve each nonlinear subdomain problem. The solution of the linear systems inside the Newton loop are obtained either by sparse LU or Preconditioned Conjugate gradient method with uncomplete LU. The acceleration procedure is described in [GTD01]. To be more precise, since the nonlinearity of the discrete operator is a second order perturbation of the Laplacian, we use the same acceleration procedure as in the Poisson problem case; that is, we compute the diagonal approximation of the matrix of acceleration P [GTD01] based on three successive Schwarz iterates.

Figures 2 and 3 give the convergence history of the Steffenson-Schwarz multilevel domain decomposition on the 2D Bratu problem with $n_x \times n_y = 3 \times 4$ subdomains and $\lambda = 6$. The convergence history is given for two problem sizes, namely 30×29 and 90×89. For the smaller problem the size of overlap is one mesh cell, but for the 3 times larger problem we have used 3 mesh cells overlap. The convergence to the exact discrete solution of the problem, at the outer loop level i.e the Steffensen-Schwarz iteration between Ω_i strips, -see 2- and inside each strips -see 3- with the second level of Steffensen-Schwarz iteration seems to be almost independent of the number of grid points provided that the size of the overlap between subdomains in each space direction stays the same.

The stop criterion for the Newton loop (resp. the Steffensen-Schwarz iterative procedure inside strips) was that the difference between two successive iterates was less than 10^{-9} (resp. 10^{-7}).

3 Arithmetical complexity

For the Helmholtz or Poisson operator case, the arithmetic complexity of the Aitken Schwarz method can be easily given analytically, provided the arithmetic complexity of the linear solver used in each subdomain is given.

Let us assume for simplicity that the arithmetic complexity of a fast sinus transform or its inverse of a vector of size N is $5Nlog_2(N)$. For strip domain decomposition with n_x subdomains, and a problem of global size $N_x \times N_y$, the Aitken acceleration requires the sinus transform and its inverse of the artificial interfaces at two iteration levels. It results into $20\ (n_x - 1)\ N_y log_2(N_y)$ operations. The solution of the pentadiagonal linear system

corresponding to the acceleration procedure itself cost $36\ N_y\ (n_x - 1)$ operations. We recall that we need to solve each subdomain problem twice.

If one uses a sparse Gaussian elimination for each subdomain linear solve, the overall arithmetic complexity is therefore approximately

$$6\ n_x\ N_x\ N_y\ (\frac{N_x}{n_x} + 3)^2\ +\ 20\ (n_x - 1)\ N_y log_2(N_y)\ +\ 36\ N_y\ (n_x - 1)$$

If one uses a fast Poisson or Helmholtz solver, the arithmetic complexity becomes approximately

$$20\ n_x\ N_y\ (\frac{N_x}{n_x}+2)\left(\log_2(\frac{N_x}{n_x} + 2) + \log_2(N_y)\right) + 20\ (n_x-1)\ N_y log_2(N_y) + 36\ N_y\ (n_x-1).$$

This complexity analysis can be extended to the two level domain decomposition method described in this paper. We have summarized in Figure 4 and Figure 5 the result of this analysis. The efficiency of our solver increases when the number of subdomains n_y in the second space direction increases from 1 to 16. Our two level domain decomposition method speedup significantly the sparse Gaussian solver, but stays at best 50% slower than a fast Poisson solver.

Our methodology is not therefore the best Poisson solver in terms of arithmetic complexity, but as shown in [BGH$^+$] its parallel efficiency in distributed computing with slow network is very good, as opposed to the parallel efficiency of fast Poisson solver based on Fast Fourier Transform algorithm that requires global transpose of matrices.

We proceed now with some experimental measurement of the arithmetic complexity of our two levels domain decomposition with the Bratu problem. We have compared different iterative procedures for the same final global accuracy of the solution: the error in maximum norm between the exact solution of the discrete problem and the final iterate is about 10^{-4}.

The linear subdomain solver inside the Newton loop is either sparse Gaussian elimination or conjugate gradient with incomplete LU preconditioning. We select the most efficient solver in our experiments, and typically the direct linear solver is preferred when the subdomains are narrow strips.

Figure 6 reports on the domain decomposition performance with $n_x = 8$ or $n_x = 16$ strip subdomains compared to the iterative solver with no domain decomposition i.e $n_x = 1$. The problem's size is $N_x \times N_y$ with $N_x = 81$, and $N_y = 11,\ 21,\ 41,\ 81$. We get good performances only if the strips are narrow enough and N_y is large. Once again the Steffensen-Schwarz algorithm for such problem becomes a very efficient algorithm for large problems. The two level domain decomposition efficiency follows the same principle. In addition, the parallel efficiency of this algorithm in metacomputing situation has been demonstrated in [BGH$^+$].

Now we proceed with some remarks on the stability of this method.

4 Sensitivity analysis

For the linear case, and when the acceleration matrix P_k are known analytically, the additional source of unstabilities in the Aitken Schwarz algorithm may come from the linear solve of (6). Let us restrict ourselves to uniform strip domain decomposition with minimum overlap and

denote

$$\begin{pmatrix} \delta_1 & 0\ 0\ \delta_2 \\ \delta_2 & 0\ 0\ \delta_1 \end{pmatrix} \tag{16}$$

the generic subblock of P_k for a given wave number k. The conditioning number of $Id - P_k$ for the Helmholtz operator, is bounded by [GTD]:

$$cond(Id - P_k) \leq 2(\frac{1}{1-\delta_1} + \frac{\delta_2\ (1-\delta_1)^{-2}}{1-\delta_2\ (1-\delta_1)^{-1}})$$

with

$$\delta_1 = \sinh(\sqrt{\lambda}h_x)/\sinh(\sqrt{\lambda}d_x),\ \delta_2 = \sinh(\sqrt{\lambda}(d_x - h_x))/\sinh(\sqrt{\lambda}d_x),$$

where d_x is the size of the Ω_i strip in x direction. The conditioning number is then of order h_x^{-1} for $\lambda = 0(1)$. A direct numerical simulation to test the sensitivity of our algorithm to perturbation on the RHS of the linear differential problem confirms the good stability properties of the one-level and two-level Aitken-Schwarz method. The linear stability of the solvers deteriorates very slowly as the number of subdomains increases, as expected.

The sensitivity analysis of the Steffensen-Schwarz method for nonlinear elliptic problems is more challenging, because P_k^i is approximately reconstructed from the sequence of 3 Schwarz iterates:

$$\begin{pmatrix} \hat{u}_{i-1}^{r,n+3} - \hat{u}_{i-1}^{r,n+2} & \hat{u}_{i-1}^{r,n+2} - \hat{u}_{i-1}^{r,n+1} \\ \hat{u}_{i+1}^{l,n+3} - \hat{u}_{i+1}^{l,n+2} & \hat{u}_{i+1}^{l,n+2} - \hat{u}_{i+1}^{l,n+1} \end{pmatrix} = P_k^i \begin{pmatrix} \hat{u}_i^{l,n+2} - \hat{u}_i^{l,n+1} & \hat{u}_i^{l,n} - \hat{u}_i^{l,n} \\ \hat{u}_i^{r,n+2} - \hat{u}_i^{r,n+1} & \hat{u}_i^{r,n} - \hat{u}_i^{r,n} \end{pmatrix} \tag{17}$$

where the $\hat{u}_i^{l,n}$ and $\hat{u}_i^{r,n}$ stand for the sine expansion coefficients of the left and right interfaces solution in Ω_i.

In particular there is no guarantee that (17) system is well posed. In our implementation, the Steffensen acceleration algorithm is applied only to waves for which this system is not badly conditioned or possibly singular. We have undertaken an extensive sensitivity analysis experiment that measures the influence on the convergence history of our algorithm when one systematically perturbed randomly the subproblem solution at the end of each subdomain solve. The test for a given domain decomposition and a given number of grid points was realized 50 times, and we checked by doubling the number of runs the sensitivity of the result. Figure 7 shows a representative average measure of the error that was obtained as a function of the norm of the perturbation. We looked at square domains with $n_x = 2$ ('o' curves), $n_x = 4$ ('+' curves), $n_x = 8$ ('v' curves) and $n_x = 16$ ('*' curves). We checked the influence of the number of points in y direction, for these four different cases. It should be noticed that the standard deviation from the mean in these experiments are of the same order than the mean of errors. These results seems to provide some confidence in the robustness of our method.

5 Conclusion

We have extended our result on Aitken like acceleration of the Schwarz algorithm presented in [GTD01], to two level domain decomposition and further investigated the arithmetic complexity and stability of our algorithm.

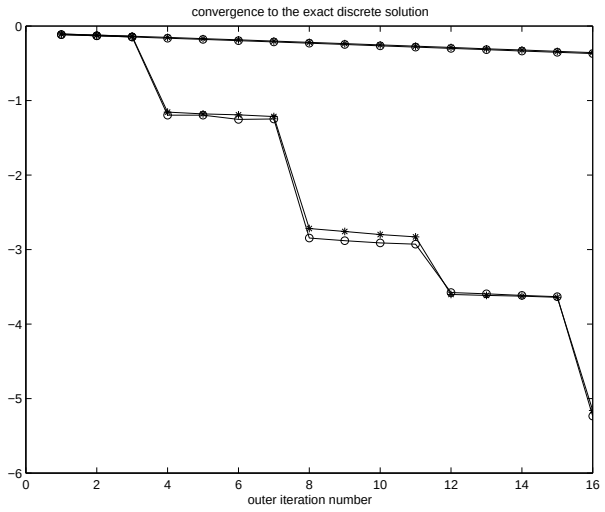


Figure 2: Convergence of the first level of Steffensen-Schwarz iterative solver, 'o' for 30×29 problem size, '*' for 90×89 problem size.

Further extension of this method to irregular meshes or non-matching grids are presently under investigation -see [BGO00] for example. We have shown in this paper that our technique is robust and numerically efficient, for the Helmholtz operator or weakly nonlinear high order perturbation of this operator such as the operator in Bratu problem. The main interest of our methodology lies however in its application to large scale metacomputing. The LIONS project [BGH$^+$] demonstrates the rather unique ability of our algorithm to provide numerical and parallel efficiency for a PDE solver with several hundred of processors distributed on several heterogeneous large-scale parallel computers in Europe linked with a slow network.

References

[BGH$^+$]N. Barberou, M. Garbey, M. Hess, M. Resch, T. Rossi, J. Toivanen, and D. Tromeur-Dervout. Scalable numerical algorithm and software tools for efficient metacomputing of PDEs. *submitted*.

[BGO00]J. Baranger, M. Garbey, and F. Oudin. Recent development on Aitken-Schwarz method. In *DD13 proceeding*, 2000.

[GH97]M. Garbey and H.G.Kaper. Heterogeneous domain decomposition methods for singular perturbation problems. *SIAM J. Num. Anal.*, 34:1513–1544, 1997.

[GTD]M. Garbey and D. Tromeur-Dervout. Aitken-Schwarz domain decomposition method. *in preparation*.

[GTD01]M. Garbey and D. Tromeur-Dervout. Two level domain decomposition for multi-

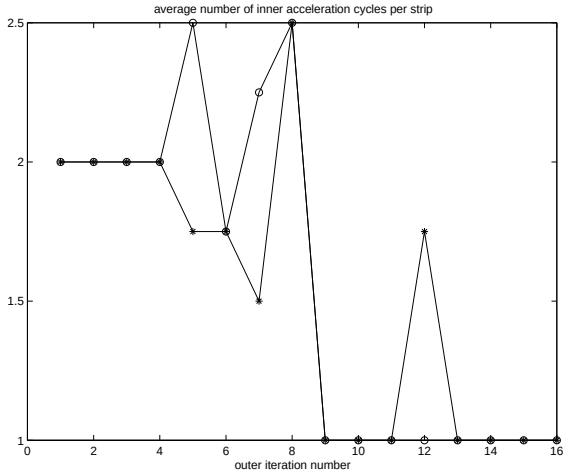


Figure 3: Averaged number of Steffensen acceleration cycles inside each strip subdomain at each iteration number of the outer loop, 'o' for 30×29 problem size, '*' for 90×89 problem size.

clusters. In published by DDM.org T. Chan & al editors, editor, *Domain Decomposition in Sciences and Engineering*, pages 325–339, 2001.

[Kuz91]Yu. A. Kuznetsov. Overlapping domain decomposition methods for fe-problems with elliptic singular perturbed operators. In PA SIAM, Philadelphia, editor, *Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations (Moscow, 1990)*, pages 223–241, 1991.

[Lio88]P.-L. Lions. On the Schwarz alternating method. i. In PA SIAM, Philadelphia, editor, *First International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 1–42, 1988.

[Lio89]P.-L. Lions. On the Schwarz alternating method. ii. stochastic interpretation and order properties. In *Domain Decomposition Methods (Los Angeles, CA, 1988)*, pages 47–70, 1989.

[Sch80]H. A. Schwarz. Gesammelte mathematische abhandlungen. *Springer Berlin*, 2(2nd ed.):133–143, 1980.

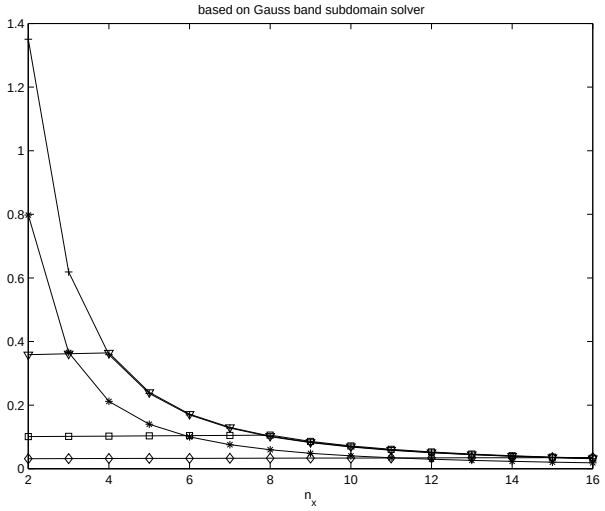


Figure 4: Arithmetic complexity for the linear case assuming sparse Gauss subdomain solver, '*' for $n_y = 1$, '+' for $n_y = 2$, triangles for $n_y = 4$, square for $n_y = 8$, diamond for $n_y = 16$.

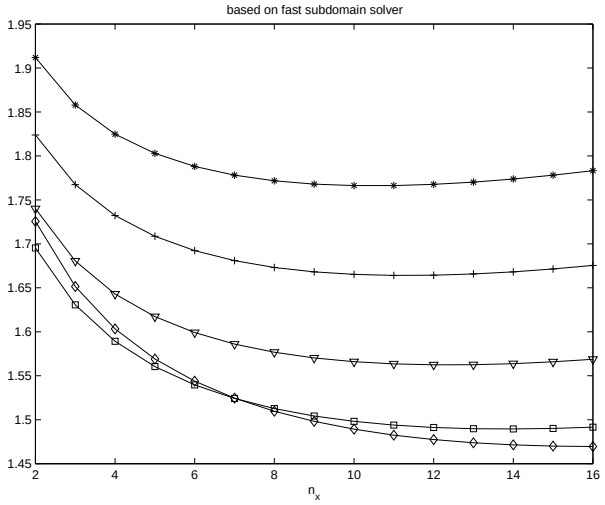


Figure 5: Arithmetic complexity for the linear case assuming fast Poisson subdomain solver '*' for $n_y = 1$, '+' for $n_y = 2$, triangles for $n_y = 4$, square for $n_y = 8$, diamond for $n_y = 16$.

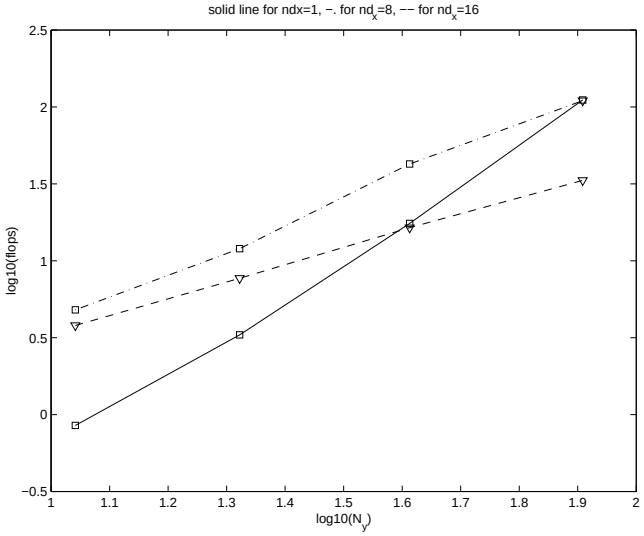


Figure 6: Arithmetic complexity for the nonlinear case, solid line for $n_x = 1$, '.-' for $n_x = 8$, dashed line for $n_x = 16$.

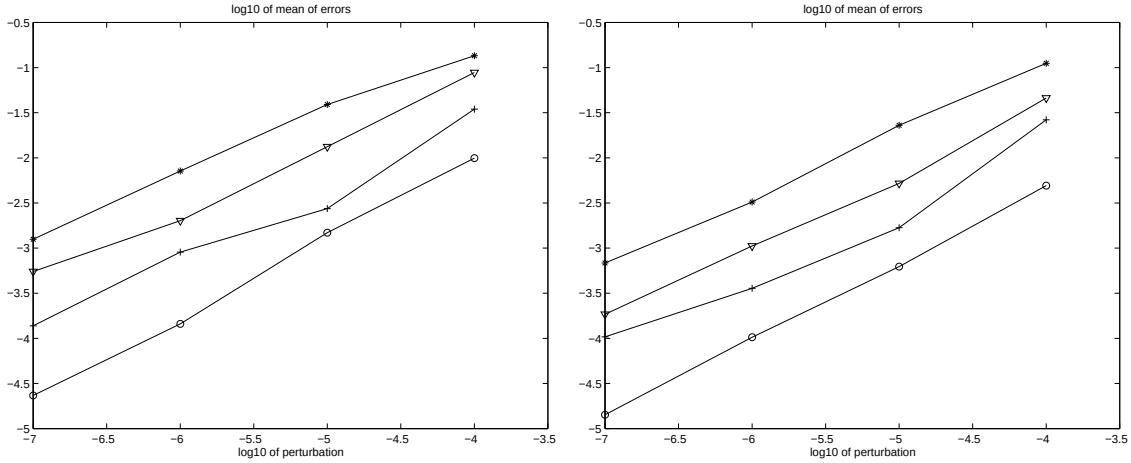


Figure 7: $N_x = 80,\ N_y = 80$, overlap is one mesh cell.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

6 Domain decomposition and fictitious domain methods with distributed Lagrange multipliers

Yu.A. Kuznetsov[1]

Introduction

In this paper we consider three applications of the distributed Lagrange multiplier technique [DGH+92, GHJ+97, GK98] to design new domain decomposition and fictitious domain methods for the diffusion equation

$$-\nabla(a\nabla u) = f, \qquad x \in \Omega, \tag{1}$$

in a bounded 2D/3D polygonal domain with the homogeneous Dirichlet boundary condition

$$u = 0, \qquad x \in \partial\Omega, \tag{2}$$

and a piece-wise constant diffusion coefficient a.

The above restrictions are imposed for the sake of simplicity. The generalizations of the algorithms and theoretical results to more complicated equations, domains, and boundary conditions are obvious.

Let Ω_h be a triangular/tetrahedral partitioning of Ω, and V_h be the corresponding piece-wise linear finite element subspace of $H_0^1(\Omega)$. We shall always assume in this paper that Ω_h is a shape-regular mesh. Then the classical finite element method

$$u^h \in V_h: \quad a(u_h, v) = l(v) \qquad \forall v \in V_h \tag{3}$$

where

$$a(u, v) = \int_\Omega a\nabla u \cdot \nabla v \, \mathrm{d}x \quad \text{and} \quad l(v) = \int_\Omega f v \, \mathrm{d}x,$$

results in the system of linear algebraic equations

$$A\bar{u} = \bar{f} \tag{4}$$

with a symmetric positive definite matrix $A \in \Re^{n\times n}$, $n = \dim V_h$, and a vector $\bar{f} \in \Re^n$. We also denote by M the mass matrix and by $\hat{M}$ the lumped mass matrix, i.e. $\hat{M}$ is diagonal and $M\bar{e} = \hat{M}\bar{e}$, $\bar{e}^T = (1, \ldots, 1)$, $\bar{e} \in \Re^n$.

For $\Omega_{1,h}$ and $\Omega_{2,h}$ being non-overlapping subdomains of Ω_h such that $\Omega_h = \Omega_{1,h} \cup \Omega_{2,h}$, we denote by A_1 and A_2 the corresponding stiffness matrices and by $M_1(\hat{M}_1)$ and $M_2(\hat{M}_2)$ the corresponding mass (lumped mass) matrices. The matrices A, M and $\hat{M}$ can be introduced by subassembling of matrices A_i, M_i, $\hat{M}_i$ with the same subassembling matrices N_i, $i = 1, 2$, respectively. For instance,

$$\begin{aligned} A &= N_1 A_1 N_1^T + N_2 A_2 N_2^T, \\ \hat{M} &= N_1 \hat{M}_1 N_1^T + N_2 \hat{M}_2 N_2^T. \end{aligned}$$

[1]Department of Mathematics, University of Houston, Houston, TX, 77204-3008, USA, e-mail: kuz@math.uh.edu

Domain decomposition for composite materials

Let Ω be a rectangle and ω_i, $i = \overline{1,m}$, $m \geq 1$, be open non-overlapping polygonal subdomains of Ω, i.e. $\omega_i \cup \omega_j = \emptyset$ for $i \neq j$ and $\partial\omega_i \cap \partial\Omega = \emptyset$, $i, j = \overline{1,m}$. An example of Ω is given in Figure 1. We assume that ω_i are shape-regular, $c_1 d \leq \mathrm{diameter}(\omega_i) \leq c_2 d$ and $\mathrm{distance}(\omega_i, \partial\Omega) \geq c_3 d$ with some positive constants c_1, c_2, and c_3 where $d > 0$ is given. We also assume that $a = 1 + \frac{1}{\delta_i}$, $\delta_i \equiv const \in (0, 1]$ in ω_i, $i = \overline{1,m}$, and $a \equiv 1$ in the rest of Ω. We shall call this model example a "composite material".

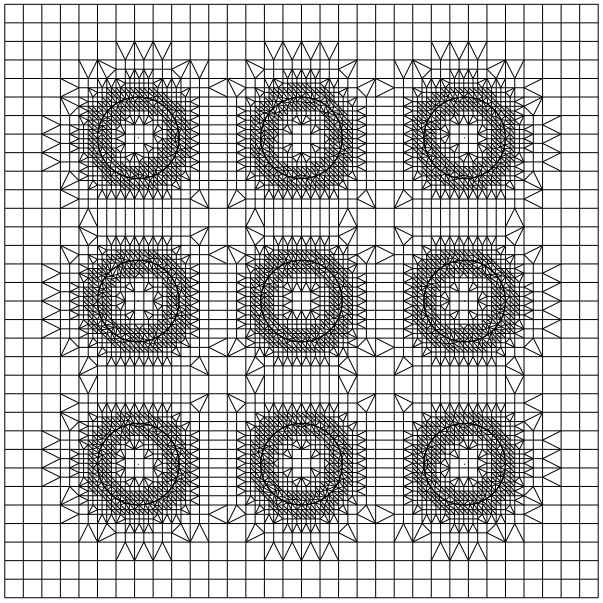


Figure 1: The computational grid.

The stiffness matrix A of system (4) can be presented in the form

$$A = A_0 + \sum_{i=1}^{m} \frac{1}{\delta_i} B_i \tag{5}$$

where

$$(B_i \bar{v}, \bar{w}) = \int_{\omega_i} \nabla v_h \cdot \nabla w_h \, \mathrm{d}x \qquad \forall v_h, w_h \in V_h,$$

and

$$(A_0 \bar{v}, \bar{w}) = \int_{\Omega} \nabla v_h \cdot \nabla w_h \, \mathrm{d}x \qquad \forall v_h, w_h \in V_h.$$

It is obvious that with an appropriate permutation matrix P_i we have

$$P_i B_i P_i^T = \begin{pmatrix} A_i & 0 \\ 0 & 0 \end{pmatrix}$$

where $-A_i$ is the stiffness matrix of the Laplacian for the subdomain ω_i, $1 \le i \le m$.

In [Kuz00] was proposed to replace system (4) with A in (5) by a saddle point system

$$\mathcal{A}\begin{pmatrix}\bar{u}\\ \bar{\lambda}\end{pmatrix} = \begin{pmatrix}A_0 & B^T\\ B & -C\end{pmatrix}\begin{pmatrix}\bar{u}\\ \bar{\lambda}\end{pmatrix} = \begin{pmatrix}\bar{f}\\ 0\end{pmatrix} \tag{6}$$

with

$$B^T = (B_1\ B_2\ \dots\ B_m) \in \Re^{n\times(mn)}$$

and the block diagonal matrix

$$C = \begin{pmatrix}\delta_1 B_1 & & \\ & \ddots & \\ & & \delta_m B_m\end{pmatrix} \in \Re^{(mn)\times(mn)}.$$

System (6) is equivalent to system (4) in the sense that the solution vector $\bar{u}$ to (4) coincides with the solution subvector $\bar{u}$ to (6) and vice versa. Moreover,

$$\bar{\lambda}_i - \frac{1}{\delta_i}\bar{u} \in \ker B_i$$

for any solution subvector $\bar{\lambda}_i$ to (6), $i = \overline{1, m}$.

Let a matrix $H_A = H_A^T > 0$ be spectrally equivalent to A_0^{-1}, i.e

$$c_4(H_A\bar{v},\, \bar{v}) \le (A_0^{-1}\bar{v},\, \bar{v}) \le c_5(H_A\bar{v},\, \bar{v}) \qquad \forall \bar{v} \in \Re^n$$

with positive constants c_4 and c_5 independent of the mesh Ω_h. Then the matrix

$$\mathcal{H} = \begin{pmatrix}H_A & 0\\ 0 & H_\lambda\end{pmatrix} \tag{7}$$

with

$$H_\lambda = \mathrm{diag}\{B_1^+,\, B_2^+,\, \dots\, ,\, B_m^+\},$$

where B_i^+ denotes the generalized inverse to B_i, $i = \overline{1, m}$, was proposed in [Kuz00] as an effective preconditioner for the matrix $\mathcal{A}$ in (6). To justify the latter statement we have to consider the matrix $\mathcal{A}\mathcal{H}$ in its invariant subspace $im\mathcal{A}$ supplied with the scalar product generated by the matrix

$$\mathcal{D} = \begin{pmatrix}H_A & 0\\ 0 & D_\lambda\end{pmatrix},$$

where

$$D_\lambda = \mathrm{diag}\{B_1,\, B_2,\, \dots\, ,\, B_m\}.$$

It can be easily shown that $\mathcal{A}\mathcal{H}$ is a symmetric operator in $im\mathcal{A}$ with respect to the $\mathcal{D}$-scalar product. Moreover, $im\mathcal{A} = im(\mathcal{A}\mathcal{H})$. To this end, all non-zero eigenvalues of the matrix $\mathcal{A}\mathcal{H}$ belong to the union of two segments $[d_1;\, d_2]$ and $[d_3;\, d_4]$ with end points

$$d_1 \le d_2 < 0 < d_3 \le d_4.$$

The condition number of $\mathcal{AH}$ with respect to the subspace $im\mathcal{A}$ and the $\mathcal{D}$-scalar product is defined by

$$\mathrm{Cond}_{\mathcal{D}}(\mathcal{AH}) = \frac{\max\{d_4;\ |d_1|\}}{\min\{d_3;\ |d_2|\}}.$$

Under all the above assumptions the following result was proved in [Kuz00].

Proposition 1

$$\mathrm{Cond}_{\mathcal{D}}(\mathcal{AH}) \le c_6, \tag{8}$$

where c_6 is a positive constant independent of the values $\delta_1,\ \delta_2,\ \ldots,\ \delta_m$ and the mesh Ω_h.

Remark 1 *In general, the constant c_6 depends on the constants c_i, $i = \overline{1,5}$.*

The implementation procedure of the preconditioner $\mathcal{H}$ is based on a simple observation that

$$B_i B_i^+ = \begin{pmatrix} Q_i & 0 \\ 0 & 0 \end{pmatrix} \tag{9}$$

where

$$Q_i \equiv A_i A_i^+.$$

The results of numerical experiments for the geometry given in Fig. 1 are presented in Table 1. For numerical experiments H_A was chosen to be the BPX-preconditioner [BPX90].

Table 1. The number of PCG iterations.

δ	13×13	34×34	76×76	160×160
1	15	16	18	18
10^{-1}	17	22	25	27
10^{-2}	19	23	27	29
10^{-3}	19	23	27	29
10^{-4}	19	23	27	29

The vectors λ_i, $i = \overline{1,m}$, in (6) can be called the discrete distributed Lagrange multipliers. They have a very simple connection with the continuous/differential distributed Lagrange multiplier. System (6) can be obtained by the straightforward finite element discretization of the variational problem: find $u \in H_0^1(\Omega)$, $\lambda_i \in H^1(\omega_i)$, $i = \overline{1,m}$, such that

$$\begin{aligned} &\int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}x + \sum_{i=1}^{m} \int_{\omega_i} \nabla \lambda_i \cdot \nabla v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x, \\ &\int_{\omega_i} \nabla u \cdot \nabla \mu_i \,\mathrm{d}x - \delta_i \int_{\omega_i} \nabla \lambda_i \cdot \nabla \mu_i \,\mathrm{d}x = 0, \qquad i = \overline{1,m}, \end{aligned} \tag{10}$$

$\forall v \in H_0^1(\Omega)$, $\mu_i \in H^1(\omega_i)$, $i = \overline{1,m}$.

Fictitious domain method

The name "fictitious domain method" was originally suggested by V.K. Saul'ev in [Sau63]. The Saul'ev's idea is to replace differential problem (1)–(2) by the problem

$$\begin{aligned} -\nabla(a_\delta \nabla u_\delta) &= f_\delta, \qquad x \in \Pi, \\ u_\delta &= 0, \qquad x \in \partial\Pi, \end{aligned} \tag{11}$$

where Π is a rectangle containing the original simply-connected domain Ω,

$$a_\delta = \begin{cases} a, & x \in \Omega, \\ 1 + \frac{1}{\delta}, & x \in \Pi \setminus \bar{\Omega}, \end{cases} \qquad f_\delta = \begin{cases} f, & x \in \Omega, \\ 0, & x \in \Pi \setminus \bar{\Omega}. \end{cases}$$

It was proved that $\|u_\delta - \hat{u}\|_{H_0^1(\Omega)} \to 0$ as $\delta \to 0$ where

$$\hat{u} = \begin{cases} u, & x \in \Omega, \\ 0, & x \in \Pi \setminus \bar{\Omega}. \end{cases}$$

The form of the equation in (1) reminds us the situation considered in the previous section. If we introduce the distributed Lagrange multiplier by

$$\lambda = \frac{1}{\delta} u \tag{12}$$

in $\omega = \Pi \setminus \bar{\Omega}$, then the weak saddle point formulation reads as follows: find $u \in H_0^1(\Pi)$, $\lambda \in H^1(\omega)$, $\lambda = 0$ on $\partial\omega \cap \partial\Pi$, such that

$$\begin{aligned} &\int_\Pi \nabla u \cdot \nabla v \, \mathrm{d}x + \int_\omega \nabla\lambda \cdot \nabla v \, \mathrm{d}x = \int_\Pi f_\delta v \, \mathrm{d}x, \\ &\int_\omega \nabla u \cdot \nabla \mu \, \mathrm{d}x - \delta \int_\omega \nabla\lambda \cdot \nabla\mu \, \mathrm{d}x = 0, \end{aligned} \tag{13}$$

$\forall v \in H_0^1(\Pi)$, $\mu \in H^1(\omega)$, $\mu = 0$ on $\partial\omega \cap \partial\Pi$.

The interesting observation is that with $\delta = 0$ formulation (13) coincides with the distributed Lagrange multiplier fictitious domain method invented by R. Glowinski (see [DGH+92, GHJ+97]). Thus, the Glowinski's method is the closure with respect to the parameter δ of the Saul'ev's method.

The finite element discretization to (13) results in the algebraic system

$$\mathcal{A} \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \bar{\lambda} \end{pmatrix} \equiv \begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & B_{22} \\ 0 & B_{22} & -\delta B_{22} \end{pmatrix} \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \bar{\lambda} \end{pmatrix} = \begin{pmatrix} \bar{f}_1 \\ \bar{f}_2 \\ 0 \end{pmatrix} \tag{14}$$

where B_{22} stays for the stiffness matrix in subdomain ω, and

$$A_0 = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

stays for the stiffness matrix in the rectangle Π. If we present $\mathcal{A}$ in a different block form:

$$\mathcal{A} = \begin{pmatrix} A_0 & B^T \\ B & -\delta C \end{pmatrix}, \qquad C = B_{22},$$

and assume that a matrix H_A is spectrally equivalent to A_0^{-1}, then the preconditioner for $\mathcal{A}$ can be proposed in the form of the block diagonal matrix

$$\mathcal{H} = \begin{pmatrix} H_A & 0 \\ 0 & H_\lambda \end{pmatrix} \tag{15}$$

where $H_\lambda = B_{22}^{-1}$.

Assume that the norm preserving finite element extension theorem for the subdomain ω with respect to the rectangle Π holds. Then,

$$\mathrm{Cond}_{\mathcal{H}}(\mathcal{A}\mathcal{H}) \le c_7$$

where c_7 is a positive constant independent of the mesh Π_h and value of $\delta \in [0; 1]$. In the case $\delta = 0$ the result was proved in [GK98]. For the case $\delta > 0$ one has to use technique from [Kuz00].

Overlapping domain decomposition

Let Ω_h be partitioned into two subdomains $\Omega_{1,h}$ and $\Omega_{2,h}$ such that $G_h = \Omega_{1,h} \cap \Omega_{2,h}$ is nonempty. We assume that $\mathrm{meas}(\partial G_h \cap \partial\Omega) \ge const > 0$, and the norm preserving finite element extension results from G_h into $\Omega_{1,h}$ and $\Omega_{2,h}$ hold [Wid87]. Later we shall give the algebraic interpretation of this assumption.

Let the bilinear form $a(u, v)$ be split into two bilinear forms [Kuz97]:

$$a(u, v) = a_1(u, v) + a_2(u, v) \tag{16}$$

and the linear form $l(v)$ be also splitted into two linear forms:

$$l(v) = l_1(v) + l_2(v) \tag{17}$$

where

$$a_i(u, v) = \int_{\Omega_i} a_i \nabla u \cdot \nabla v \, \mathrm{d}x$$

with

$$a_i = \begin{cases} a, & x \in \Omega_i \setminus G, \\ a/2, & x \in G, \end{cases}$$

and

$$l_i(v) = \int_{\Omega_i} \alpha_i f v \, \mathrm{d}x$$

with

$$\alpha_i = \begin{cases} 1, & x \in \Omega_i \setminus G, \\ 1/2, & x \in G, \end{cases}$$

$i = 1, 2$. Then, let us define two new bilinear and linear forms by

$$\begin{aligned} \hat{a}(\bar{u},\, \bar{v}) &= a_1(u_1,\, v_1) + a_2(u_2,\, v_2), \\ b(\lambda,\, \bar{v}) &= \int_G \nabla\lambda \cdot \nabla(v_1 - v_2)\, \mathrm{d}x, \\ \hat{l}(\bar{v}) &= l_1(v_1) + l_2(v_2) \end{aligned} \tag{18}$$

where

$$\bar{v} = \begin{pmatrix} v_1, \\ v_2 \end{pmatrix}, \quad v_i \in V_i = \left\{ v\colon\ v \in H^1(\Omega_i),\ v = 0 \ \text{ on } \ \partial\Omega \cap \partial\Omega_i \right\}, \quad i = 1, 2,$$

and

$$\lambda \in V_\lambda = \left\{ \lambda\colon\ \lambda \in H^1(G),\ \lambda = 0 \ \text{ on } \ \partial\Omega \cap \partial G \right\}.$$

Then, the weak formulation of (1) based on the above overlapping decomposition with distributed Lagrange multipliers can be given by: find $\bar{u} \in \hat{V} = V_1 \times V_2$, $\lambda \in V_\lambda$ such that

$$\begin{aligned} \hat{a}(\bar{u},\, \bar{v}) + b(\lambda,\, \bar{v}) &= \hat{l}(v), \\ b(\bar{u},\, \mu) &= 0 \end{aligned} \tag{19}$$

$\forall \bar{v} \in \hat{V}$, $\mu \in V_\lambda$.

The finite element discretization of (19) can be suggested with the same formulae by replacing $\hat{V}$ and V_λ by $\hat{V}_h$ and $V_{\lambda,h}$ which are the traces of the finite element space V_h onto $\Omega_{1,h}$, $\Omega_{2,h}$ and G_h, respectively. The finite element discretization of (19) results in the system of algebraic equations

$$\mathcal{A} \begin{pmatrix} \bar{u} \\ \bar{\lambda} \end{pmatrix} \equiv \begin{pmatrix} A_1 & 0 & B_1^T \\ 0 & A_2 & B_2^T \\ B_1 & B_2 & 0 \end{pmatrix} \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \bar{\lambda} \end{pmatrix} = \begin{pmatrix} \bar{f}_1 \\ \bar{f}_2 \\ 0 \end{pmatrix}, \tag{20}$$

where

$$A_1 = \begin{pmatrix} A_{11} & A_{1G} \\ A_{G1} & A_{GG}^{(1)} \end{pmatrix} \qquad A_2 = \begin{pmatrix} A_{GG}^{(2)} & A_{G2} \\ A_{2G} & A_{22} \end{pmatrix},$$

$$B_1^T = \begin{pmatrix} 0 \\ B_G \end{pmatrix}, \qquad B_2^T = \begin{pmatrix} B_G \\ 0 \end{pmatrix}.$$

Here B_G is defined by

$$(B_G\bar{\lambda},\, \bar{\mu}) = \int_G \nabla\lambda_h \cdot \nabla\mu_h\, \mathrm{d}x, \qquad \forall\lambda_h, \mu_h \in V_{\lambda,h}, \tag{21}$$

i.e. $-B_G$ is the stiffness matrix for the Laplacian in the subdomain G_h.

We introduce a preconditioner $\mathcal{H}$ for $\mathcal{A}$ in the form of a block diagonal matrix:

$$\mathcal{H} = \begin{pmatrix} H_1 & 0 & 0 \\ 0 & H_2 & 0 \\ 0 & 0 & H_\lambda \end{pmatrix}, \tag{22}$$

where H_i is spectrally equivalent to A_i^{-1}, $i = 1, 2$, and H_λ^{-1} is spectrally equivalent to the Schur complement matrix

$$S_\lambda = B_1 A_1^{-1} B_1^T + B_2 A_2^{-1} B_2^T. \tag{23}$$

We have plenty of choices for H_1 and H_2, for instance, multigrid preconditioner. The question is only about a choice for H_λ.

The assumption about the norm preserving finite element extension results (in the context of the above method) is equivalent to the assumption that the matrix B_G is spectrally equivalent to matrices

$$S_G^{(i)} = A_G^{(i)} - A_{Gi} A_{ii}^{-1} A_{iG}, \qquad i = 1, 2.$$

In this case simple transformations show that the matrix S_λ is spectrally equivalent to the matrix B_G. The conclusion is obvious: we have to choose

$$H_\lambda = B_G^{-1}.$$

Implementation procedure for H_λ is very simple due to the formulae

$$\mathcal{HA} = \begin{pmatrix} H_1 & 0 & 0 \\ 0 & H_2 & 0 \\ 0 & 0 & I_\lambda \end{pmatrix} \begin{pmatrix} A_1 & 0 & B_1^T \\ 0 & A_2 & B_2^T \\ \tilde{B}_1 & \tilde{B}_2 & 0 \end{pmatrix},$$

where

$$\tilde{B}_1 = (0\ I_\lambda) \quad \text{and} \quad \tilde{B}_2 = (I_\lambda\ 0)\,.$$

Proposition 2 *Under the assumptions made, the eigenvalues of the matrix $\mathcal{HA}$ belong to the union of two segments $[d_1;\ d_2]$, $[d_3;\ d_4]$ with the end points $d_1 \le d_2 < 0 < d_3 \le d_4$ independent of the mesh Ω_h.*

Remark 2 *The values of d_1, d_2, d_3 and d_4 from Proposition 2 depend on the constants of spectral equivalence H_i and A_i, as well as B_G and $S_G^{(i)}$, $i = 1, 2$.*

Acknowledgments: This work was partially supported by NSF (grant CCR-9902035) and by Los Alamos Computer Science Institute (LASCI). The author thanks K.Lipnikov for providing the numerical experiments and technical assistance.

References

[BPX90]James H. Bramble, Joseph E. Pasciak, and Jinchao Xu. Parallel multilevel preconditioners. *Math. Comp.*, 55:1–22, 1990.

[DGH+92]Q. V. Dihn, R. Glowinski, J. He, V. Kwock, T. W. Pan, and J. Périaux. Lagrange multiplier approach to fictitious domain methods: Application to fluid dynamics and electromagnetics. In David E. Keyes, Tony F. Chan, Gérard A. Meurant, Jeffrey S. Scroggs, and Robert G. Voigt, editors, *Fifth International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 151–194, Philadelphia, PA, 1992. SIAM.

[GHJ+97]R. Glowinski, T. I. Hesla, D.D. Joseph, T.W. Pan, and J. Periaux. Distributed Lagrange multiplier methods for particulate flows. In M.O. Bristeau, G.J. Etgen, W. Fitzgibbon, J.L. Lions, J. Periaux, and M.F. Wheeler, editors, *Computational Science for the 21st Century*, pages 270–279, Chichester, 1997. Wiley.

[GK98]Roland Glowinski and Yuri Kuznetsov. On the solution of the Dirichlet problem for linear elliptic operators by a distributed Lagrange multiplier method. *C. R. Acad. Sci. Paris Sér. I Math.*, 327(7):693–698, 1998.

[Kuz97]Yuri. A. Kuznetsov. Overlapping domain decomposition with non matching grids. In Petter E. Bjørstad, Magne Espedal, and David Keyes, editors, *Domain Decomposition Methods in Sciences and Engineering*. J. Wiley, 1997. Proceedings from the Ninth International Conference, June 1996, Bergen, Norway.

[Kuz00]Yuri A. Kuznetsov. New iterative methods for singular perturbed positive definite matrices. *Russian J. Numer. Anal. Math. Modelling*, 15:65–71, 2000.

[Sau63]Valerij K. Saul'ev. On solution of some boundary value problems on high performance computers by fictitious domain method. *Siberian Math. J.*, 4(4):912–925, 1963. (in Russian).

[Wid87]Olof B. Widlund. An extension theorem for finite element spaces with three applications. In Wolfgang Hackbusch and Kristian Witsch, editors, *Numerical Techniques in Continuum Mechanics*, pages 110–122, Braunschweig/Wiesbaden, 1987. Notes on Numerical Fluid Mechanics, v. 16, Friedr. Vieweg und Sohn. Proceedings of the Second GAMM-Seminar, Kiel, January, 1986.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

7 Overlapping preconditioners for discontinuous Galerkin approximations of second order problems

C. Lasser [1], A. Toselli [2]

Introduction

The purpose of this paper is to present a two–level overlapping preconditioner for discontinuous Galerkin finite element discretization of advection-diffusion problems in two or three dimensions. Our problem is discretized using a discontinuous Galerkin (DG) finite element method. The original domain is then subdivided into overlapping subdomains in order to introduce a number of local problems. We propose two different coarse problems. The first one is an advection–diffusion problem discretized using a continuous finite element space on a coarse triangulation. The second employs a smoothed aggregation technique and does not require the introduction of a coarse mesh. The performance of the corresponding two methods is illustrated for two test problems in two dimensions discretized with linear finite elements.

Discontinuous Galerkin approximations have been used since the early 1970s and are recently becoming more and more popular for the approximation of a large class of problems; we refer to [CKS00] for a comprehensive review of these methods. Here, we consider a discontinuous hp-finite element method proposed in [HSS00]. As for many DG methods, the approximate solution belongs to a space of discontinuous finite element functions, i.e., it is piecewise polynomial of a certain degree on a given triangulation, being in general discontinuous across the elements. Suitable bilinear forms, which also contain interface contributions, are then employed, in order to ensure consistency.

We know of only two previous works on DD preconditioners for DG approximations; see [FK00, LT00].

Continuous and discrete problems

We consider the following scalar advection-diffusion problem with Dirichlet conditions

$$\begin{aligned} \mathcal{L}u \;=\; -\nabla\cdot(a\nabla u) + b\cdot\nabla u + cu &= f, \quad \text{in } \Omega\,, \\ u &= 0, \quad \text{on } \Gamma\,, \end{aligned} \tag{1}$$

where Ω is a bounded open polyhedral domain in $\mathbb{R}^d$, $d = 2, 3$, and Γ its boundary. Problem (1) describes a large class of diffusion-transport-reaction processes.

We consider problem (1) and make some further hypotheses. We assume that $a = \{a_{i,j}\}_{i,j=1}^{d}$ is a symmetric positive–semidefinite matrix,

$$\xi^T a(x)\xi \geq 0, \quad \xi \in \mathbb{R}^d, \quad x \in \Omega,$$

[1] Technische Universität München, classer@mathematik.tu-muenchen.de
[2] ETH Zürich, toselli@sam.math.ethz.ch

b and c are a vector field in $W^{1,\infty}(\Omega)$ and a function in $L^\infty(\Omega)$, respectively, such that

$$(c - \frac{1}{2}\nabla \cdot b)(x) \ \geq \ 0\,, \quad x \in \Omega\,, \tag{2}$$

and the right-hand side f is a function in $L^2(\Omega)$. The existence of a unique solution of (1) is shown in [HSS00]. We note that we have considered only the case of strongly–imposed homogeneous Dirichlet boundary conditions for simplicity, but that more general ones can be employed, such as Neumann, Robin, or weakly–imposed Dirichlet conditions. Our methods can be extended to these cases. We also recall that in case a does not have full rank, Dirichlet conditions can only be imposed on a part of the boundary; see [HSS00].

We next introduce $\mathcal{T}_h$, a conforming, shape–regular triangulation of Ω consisting of open simplices κ with diameter $O(h)$. We denote by $\mathcal{P}_k(\kappa)$ the space of polynomials on κ of total degree $k \in \mathbb{N}_0$ and define the vector of local polynomial degrees $\mathbf{p} = (p_\kappa : \kappa \in \mathcal{T}_h)$. We consider the finite element space

$$S^{\mathbf{p}}(\Omega, \mathcal{T}_h) = \{u \in L^2(\Omega) : u|_\kappa \in \mathcal{P}_{p_\kappa}(\kappa)\}\,.$$

and define $S_0^{\mathbf{p}}(\Omega, \mathcal{T}_h)$ as the subspace of functions in $S^{\mathbf{p}}(\Omega, \mathcal{T}_h)$ vanishing on Γ. Our FE approximation space is chosen as $V^h = S_0^{\mathbf{p}}(\Omega, \mathcal{T}_h)$.

We define $\mathcal{E}_{int}$ as the set of edges that are intersections of the element boundaries and Γ_{int} as the union of the edges in $\mathcal{E}_{int}$. For $\kappa \in \mathcal{T}_h$, we then denote the unit outward normal to $\partial\kappa$ at $x \in \partial\kappa$ by $\mu_\kappa(x)$ and partition the part of its boundary that is also contained in Γ_{int} into two sets:

$$\begin{aligned} \partial_-\kappa &= \{x \in \partial\kappa \cap \Gamma_{int} : b(x) \cdot \mu_\kappa(x) < 0\} \qquad \text{(inflow part)}\,, \\ \partial_+\kappa &= \{x \in \partial\kappa \cap \Gamma_{int} : b(x) \cdot \mu_\kappa(x) \geq 0\} \qquad \text{(outflow part)}\,. \end{aligned}$$

Given $v \in S^{\mathbf{p}}(\Omega, \mathcal{T}_h)$, its restriction to $\bar{D} \subset \bar{\Omega}$ is denoted by $v_D = v|_{\bar{D}}$. Then, for $x \in \partial_-\kappa$ there exists a unique neighbor κ' with $x \in \partial\kappa'$ and set

$$v_\kappa^+(x) = v_\kappa(x)\,, \quad v_\kappa^-(x) = v_{\kappa'}(x)\,, \quad \lfloor v \rfloor_\kappa = v_\kappa^+ - v_\kappa^-\,.$$

Given an interior edge $e \in \mathcal{E}_{int}$, there are two elements κ_i, κ_j, with, e.g., $i > j$, that share this edge. We define

$$[v]_e = v|_{\partial\kappa_i \cap e} - v|_{\partial\kappa_j \cap e}, \quad < v >_e = \frac{1}{2}(v|_{\partial\kappa_i \cap e} + v|_{\partial\kappa_j \cap e})\,,$$

and ν as the unit normal which points from κ_i to κ_j. We note, that μ and ν point in different directions in general and that $\lfloor \cdot \rfloor$ and $[\cdot]$ are distinct. Similarly, for $e = \partial\kappa \cap \Gamma$, we set

$$[v]_e = v|_e.$$

Finally, we introduce a discontinuity-penalization function σ defined on Γ_{int}:
for an edge $e \in \mathcal{E}_{int}$, we denote the diameter of e by h_e and define

$$\sigma_e \ = \ \sigma_0 \cdot \frac{< \bar{a}p^2 >_e}{h_e}\,,$$

where $\bar{a} = ||a||$ and σ_0 is a suitably chosen positive constant.

For $u, v \in V^h$, we consider the bilinear form

$$\begin{aligned} B(u,v) &= \sum_{\kappa\in\mathcal{T}_h}\int_\kappa a\nabla u\cdot\nabla v dx + \sum_{\kappa\in\mathcal{T}_h}\int_\kappa (b\cdot\nabla u + cu)v dx \\ &- \sum_{\kappa\in\mathcal{T}_h}\int_{\partial_-\kappa\cap\Gamma_{int}} (b\cdot\mu)\lfloor u\rfloor v^+ ds + \int_{\Gamma_{int}} \sigma[u][v] ds \\ &+ \int_{\Gamma_{int}} ([u] < (a\nabla v)\cdot\nu > - < (a\nabla u)\cdot\nu > [v])\, ds\,, \end{aligned}$$

which has been proposed in [HSS00]. Our DG approximation of (1) is then defined as the unique $u \in V^h$ such that

$$B(u,v) = (f,v)_{L^2(\Omega)}\,, \quad v \in V^h\,. \tag{3}$$

Problem (3) can be written in matrix form as

$$Bu = f, \tag{4}$$

where we have used the same notation for a function $u \in V^h$ and the corresponding vector of degrees of freedom, and a bilinear form, e.g., $B(\cdot,\cdot)$, and its matrix representation in the space V^h. Similarly, in the following we use the same notation for functional spaces and the corresponding spaces of vectors of degrees of freedom.

We consider the following scalar product in $S_0^{\mathbf{P}}(\Omega,\mathcal{T}_h)$:

$$A(u,v) = \sum_{\kappa\in\mathcal{T}_h}\int_\kappa a\nabla u\cdot\nabla v dx + \int_{\Gamma_{int}}\sigma[u][v]ds\,,$$

Two–level overlapping preconditioners

We consider preconditioners of B of the form

$$\hat{B}^{-1} = \sum_{i=1}^N R_i^T B_i^{-1} R_i + R_0^T B_0^{-1} R_0,$$

where the $\{B_i\}$ are local operators associated to a partition of Ω into subdomains and B_0 is associated to a global, low–dimensional problem. More precisely, we consider a non–overlapping partition of Ω into subdomains

$$\mathcal{F}_H = \{\Omega_i\}_{1\le i\le N},$$

of diameter $H > h$. We next extend each Ω_i to a larger region $\Omega_i' \subset \Omega$, in such a way that Ω_i' is the union of some elements in $\mathcal{T}_h$.

The first problem we need to address is the choice of the local solvers associated to the $\{\Omega_i'\}$. Here, we exploit the fact that we work with discontinuous FE functions and define our local spaces by

$$V_i = \{\, u \in V^h : \quad u_{|_\kappa} = 0,\ \kappa \subset \Omega\setminus\Omega_i'\,\}, \quad 1 \le i \le N\,. \tag{5}$$

We note that a function in V_i is discontinuous and, as opposed to the case of conforming approximations, in general does not vanish on $\partial\Omega_i'$. Then, $R_i^T : V_i \to V^h$ is the natural interpolation operator from the subspace V_i into V_h and the restriction $R_i : V^h \to V_i$ puts to zero the degrees of freedom outside $\bar{\Omega}_i'$.

We showed in [LT00] that, in the pure hyperbolic case $a = 0$, the local operator $B_i = R_i B R_i^T : V_i \to V_i$, $1 \leq i \leq N$, is the approximation of a Dirichlet problem with *weakly* imposed boundary conditions on the inflow part of the boundary $\partial\Omega_i'$, which is therefore well–posed. If some diffusion is present, B_i, although having contributions from bilinear forms defined on the boundary, is not the approximation of a Dirichlet problem with weakly imposed boundary conditions on $\partial\Omega_i'$. However, it is positive–definite and the corresponding local problem in Ω_i' is well–posed.

We also note that, thanks to the choice of the local spaces, the case of zero overlap,

$$\Omega_i' = \Omega_i, \quad 1 \leq i \leq N,$$

can be considered, as was already noted in [FK00].

The first coarse solver that we consider was already introduced in [LT00]. It requires that the partition $\mathcal{F}_H$ is a coarse mesh $\mathcal{T}_H$. The matrix B_0 is then the approximation of our advection–diffusion problem on the *continuous, piecewise linear* FE space

$$V_0 = S^1(\Omega, \mathcal{T}_H) \cap H_0^1(\Omega) \subset V^h.$$

If $R_0^T : V_0 \to V^h$ is the natural interpolation operator from the subspace V_0 into V_h, then our coarse solver is

$$B_0 = R_0 B R_0^T,$$

and it can be easily shown to be positive–definite. In [LT00], we proved that this choice of coarse space leads to an optimal, scalable preconditioner for GMRES. The second coarse solver is introduced in the next section.

Smoothed aggregation techniques

The use of smoothed aggregation (SA) techniques allows to build coarser spaces without the need of introducing coarser triangulations for multi–level and two–level preconditioners, and is particularly advantageous when dealing with problems on unstructured grids; see [VMB96, BV99, JKMK00]. The use of such techniques also appears to be promising for DG approximations, thanks to the possibility for the smoothed coarse basis functions to 'follow' the direction of the flow b.

We first suppose that the reaction term c is identically zero. We start by associating a vector to each subdomain. Let $\tilde{\phi}_i \in V_h$ be the characteristic function of Ω_i. The functions $\{\tilde{\phi}_i\}$ span a subspace of dimension N and are good candidates for building a coarse space since they are able to reproduce constant functions. Unfortunately, they have a high energy. We note that we are working with discontinuous functions and that the term in the energy bilinear form $A(\cdot,\cdot)$ responsible for this high energy is the penalization term. Indeed, we have

$$A(\tilde{\phi}_i, \tilde{\phi}_i) \approx \frac{H^{d-1}}{h}.$$

The idea of SA techniques is then to smooth these functions out by increasing their support using the stencil of suitable polynomials in B. The property of being able to reproduce constants relies on the kernel of the operator $\mathcal{L}$.

$$\mathcal{L}1 = 0.$$

For $q \in \mathbb{N}_0$, we define

$$\phi_i = S^q \tilde{\phi}_i := (I - \omega D^{-1} B)^q \tilde{\phi}_i, \quad 1 \leq i \leq N, \quad \omega \in (0, 1],$$

where D is a suitable diagonal matrix that can be chosen, for instance, as the diagonal part of B. We note that the smoothed functions are still able to reproduce the constants.

We first consider the pure hyperbolic case $a = 0$. Given an element $\kappa \in \mathcal{T}_h$ and a neighboring element κ' that share an edge e with κ, the degrees of freedom of κ on e are only coupled with the corresponding degrees of freedom of κ' on e through the upwinding term of the bilinear form. Due to this fact:

- we need two applications of S, in order to extend the supports of the $\{\tilde{\phi}_i\}$ of one layer;
- the support only increases along the streamlines, in the **positive** direction of the flow b.

This second property appears to be extremely favorable since the exchange of information produced by the coarse solve follows the same pattern as that of the original problem.

If the diffusion is not zero, the degrees of freedom of the element κ on e are coupled to all the degrees of freedom of κ'. In this case:

- we need one application of S, in order to extend the supports of the $\{\tilde{\phi}_i\}$ of one layer and we expect the entries of the smoothed functions to be higher in the direction of the flow for convection–dominated problems;
- their support is extended in all directions.

If the reaction coefficient is not zero, constant functions are not in general reproduced, but we expect this to be balanced by the better conditioning of B.

Our coarse space is then defined as

$$V_0 := \operatorname{span}\{\phi_i\}.$$

We remark that a SA technique provides now *all* the components of our preconditioner:

- the coarse space, through the matrix R_0^T, the columns of which are the vectors $\{\phi_i\}$;
- the local solvers, since the overlapping subdomains can be chosen as the supports of the functions $\{\phi_i\}$.

For the last property, we remark that, in the diffusive case the overlap between the subdomains is $O(qh)$.

Numerical results

In this section, we show some test cases for two simple problems in two dimensions. They are for uniform meshes on the unit square, consisting of $2(n \times n)$ triangles, and linear finite elements. We impose Dirichlet conditions weakly. We have employed $GMRES$ without restart. We have stopped our iterations once a reduction of the residual norm of 10^{-6} is achieved or after 100 iterations. We note that, for coarse spaces built with SA techniques, the overlap $\delta = qh$ is determined by the degree q of the smoothing operator.

We first consider the Poisson equation with inhomogeneous Dirichlet conditions:

$$-\Delta u = xe^y \quad \text{in } \Omega\,, \quad u = -xe^y \quad \text{on } \Gamma\,.$$

and partitions into $nc \times nc$ squares ($H = 1/nc$), with $nc = 2, 4, 8, 16, 32$. Table 1 shows the iteration counts for the two algorithms, as functions of h and the inverse of the relative overlap. We have also considered the case of zero overlap, denoted by $H/\delta = \infty$.

Both methods appear to be rather insensitive to the size of the original problem and the number of subdomains, when the relative overlap δ/H is positive and fixed. The two algorithms appear to be optimal and scalable. We also note that the iteration numbers decrease when the relative overlap increases, and that the iterations for method (II) (smoothed aggregation) are roughly double those of method (I) (standard coarse space). We remark that for symmetric positive–definite problems, the condition number grows linearly with H/δ for method (I), while in general we expect a quadratic growth for method (II); see [JKMK00].

The case of zero overlap requires a special discussion. Our results show that the number of iterations obtained are generally comparable to, but slightly higher than, those obtained in the case of $\delta > 0$ for both algorithms. From our numerical results for algorithm (I), we are unable to deduce whether it is optimal or non–optimal, with the number of iterations growing as a power of H/h. We refer to [FK00] for a method with the same local solvers but a different coarse space, which exhibits a rate of convergence that appears to grow linearly with H/h. On the other hand, we note that for algorithm (II), comparable numbers of iterations are obtained when the ratio H/h is fixed.

However, we believe that due to the minimal communication between the subdomains and the relatively small iteration counts that we have obtained, the algorithms with zero overlap might be competitive in practice.

We next consider the advection-diffusion equation

$$-\Delta u + b \cdot \nabla u + cu = f \quad \text{in } \Omega\,, \quad u = 0 \quad \text{on } \Gamma\,,$$

with constant coefficients and weakly–imposed zero Dirichlet boundary conditions. We consider the case

$$b = -(3\pi, 3\pi), \quad c = 3\pi^2.$$

The right-hand side f is always chosen such that the exact solution is

$$u = xe^{xy} \sin(\pi x) \sin(\pi y).$$

The numbers of iterations are shown in Table 2.

As for the Poisson problem with nonvanishing overlap, the iteration counts decrease when the overlap increases and are independent of the number of subdomains and the problem size.

two level (I)						
			H/δ			
n	nc	∞	16	8	4	2
16	4	13	-	-	13	14
32	4	15	-	13	12	13
32	8	13	-	-	13	15
64	4	19	15	14	13	13
64	8	16	-	13	13	14
64	16	13	-	-	13	15
128	4	25	18	16	14	13
128	8	35	15	14	13	14
128	16	15	-	13	13	15
128	32	12	-	-	13	15

two level (II), $\omega = 2/3$						
			H/δ			
n	nc	∞	16	8	4	2
16	4	21	-	-	22	18
32	4	28	-	30	25	20
32	8	25	-	-	26	22
64	4	37	40	33	27	21
64	8	33	-	35	31	25
64	16	26	-	-	27	25
128	4	48	53	44	35	28
128	8	43	47	42	34	27
128	16	34	-	37	35	28
128	32	26	-	-	27	25

Figure 1: Poisson problem with standard coarse space (left) and a coarse space built using a smooth aggregation technique (right).

two level (I): $b = -(3\pi, 3\pi)$, $c = 3\pi^2$						
			H/δ			
n	nc	∞	16	8	4	2
16	4	14	-	-	14	15
32	4	16	-	14	13	14
32	8	12	-	-	13	15
64	4	19	15	14	13	14
64	8	13	-	13	13	15
64	16	10	-	-	12	15
128	4	24	19	16	14	14
128	8	35	13	13	13	15
128	16	11	-	11	11	15

two level (II), $\omega = 2/3$						
			H/δ			
n	nc	∞	16	8	4	2
16	4	22	-	-	17	15
32	4	28	-	24	20	16
32	8	28	-	-	22	19
64	4	38	33	27	22	16
64	8	39	-	32	28	21
64	16	31	-	-	25	23
128	4	50	45	37	30	23
128	8	53	45	40	32	23
128	16	43	-	37	34	26

Figure 2: Advection–diffusion problem with standard coarse space (left) and a coarse space built using a smooth aggregation technique (right).

SA techniques also give numbers of iterations that are double those for a standard coarse space.

The behavior for zero overlap appears to be more regular when a standard coarse space is employed. The number of iterations appears to grow like H/h, *when* h *is fixed*. For a fixed value of H/h, slower convergence rates are obtained for h larger. We can then conclude that, for the case of zero overlap, the iteration counts are indeed bounded by a $C(H/h)$, with C a suitable constant; see also [FK00].

On the other hand, with a SA technique comparable numbers of iterations are obtained for fixed h, regardless the value of H, but the method does not appear to be optimal in this case.

References

[BV99]Marian Brezina and Petr Vaněk. A black–box iterative solver based on a two–level Schwarz method. *Computing*, 63:233–263, 1999.

[CKS00]Bernardo Cockburn, George E. Karniadakis, and Chi-Wang Shu, editors. *Discontinuous Galerkin Methods*, volume 11 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, 2000.

[FK00]Xiaobing Feng and Ohannes A. Karakashan. Two–level non–overlapping Schwarz methods for a discontinuous Galerkin method. Submitted to Siam J. on Numer. Anal., 2000.

[HSS00]Paul Houston, Endre Süli, and Christoph Schwab. Discontinuous hp–finite element methods for advection–diffusion problems. Technical Report 00–07, Seminar für Angewandte Mathematik, ETH, Zürich, 2000. Submitted to Math. Comp.

[JKMK00]Lea Jenkins, Tim Kelley, Cass T. Miller, and Christopher E. Kees. An aggregation-based domain decomposition preconditioner for groundwater flow. Technical Report TR00–13, Department of Mathematics, North Carolina State University, 2000.

[LT00]Caroline Lasser and Andrea Toselli. An overlapping domain decomposition preconditioner for a class of discontinuous Galerkin approximations of advection-diffusion problems. Technical Report 810, Dept. of Computer Science, Courant Institute, 2000. submitted to Math. Comp.

[VMB96]Petr Vaněk, Jan Mandel, and Marian Brezina. Algebraic multigrid by smooth aggregation for second and fourth order elliptic problems. *Computing*, 56:179–196, 1996.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

8 On Polynomial Reproduction of Dual FE Bases

Peter Oswald[1], Barbara Wohlmuth[2]

Introduction and abstract algebraic condition

We construct local piecewise polynomial dual bases for standard Lagrange finite element spaces which themselves provide maximal polynomial reproduction. By means of such dual bases for the Lagrange multiplier, extremely efficient realization of mortar methods on non-matching triangulations can be obtained without losing the optimality of the discretization errors. In contrast to the standard mortar approach, the locality of the constrained basis functions is preserved. The construction of dual bases and quasi-interpolants for univariate spline spaces is well-understood (see, e.g., [dB76, dB90, dBF73, Sch81]). However, the dual space is usually of a more complicated structure, and cannot be fixed beforehand, see also [DKU99, DS97, Ste00] for related research in the context of biorthogonal multiresolution analysis.

We start with an abstract framework. Let $P, V, W \subset X$ be subspaces of a real Hilbert space H. Furthermore, we assume that $q := \dim P \le n := \dim V = \dim W \le m := \dim X < \infty$. Let $\mathbf{P}$, Φ, Ψ and Θ be bases of P, V, W and X, respectively. All function systems are written as row vectors, the matrix notation used below will be consistent with this assumption. We also frequently use the notation $G_{\Psi_1,\Psi_2} := (\Psi_1^T, \Psi_2)_H$ for the Gram matrix associated with two finite systems $\Psi_1, \Psi_2 \subset H$. Note that $G_{\Psi_2,\Psi_1} = G_{\Psi_1,\Psi_2}^T$, and that $G_{\Psi,\Psi}^{-1}$ exists whenever the elements of Ψ are linearly independent. By our assumptions, there exist matrices A, $B \in \mathbb{R}^{m\times n}$ and $D \in \mathbb{R}^{m\times q}$, such that $\Phi = \Theta A$, $\Psi = \Theta B$ and $\mathbf{P} = \Theta D$. The sets of basis functions Φ and Ψ are called biorthogonal (or, equivalently, Ψ is dual to Φ) if

$$\mathrm{Id}_n = G_{\Psi,\Phi} \ . \tag{1}$$

The components of the function systems Φ and Ψ are denoted by ϕ_k and ψ_k, respectively. We introduce the dual operators $Q_W \nu = \sum_{k=1}^n (\phi_k, \nu)_H \, \psi_k$ and $Q_V \nu = \sum_{k=1}^n (\psi_k, \nu)_H \, \phi_k$, i.e., $(\nu, Q_W \mu)_H = (Q_V \nu, \mu)_H$. Assuming that Ψ and Φ are biorthogonal, we find that Q_W reproduces the subspace P, i.e.,

$$Q_W p = p \qquad \forall\, p \in P \ , \tag{2}$$

if and only if $P \subset W$.

In the rest of this section, we establish algebraic conditions on Ψ such that biorthogonality and subspace reproduction are satisfied for given choices of $\mathbf{P}$, Φ, and Θ.

Lemma 1 *Under the above assumptions, (1) and (2) hold if and only if*

$$\mathrm{Id}_n = A^T G_{\Theta,\Theta} B \ , \tag{3}$$

$$G_{\Theta,\mathbf{P}} = G_{\Theta,\Theta} B G_{\Phi,\mathbf{P}} \ . \tag{4}$$

[1]Bell Laboratories, Lucent Technologies, poswald@research.bell-labs.com
[2]Math. Inst. A, University of Stuttgart, wohlmuth@mathematik.uni-stuttgart.de
This work was supported in part by MSRI, UC Berkeley

Proof Equation (3) is equivalent to (1) since

$$G_{\Phi,\Psi} = (\Phi^T, \Psi)_H = A^T(\Theta^T, \Theta)_H B = A^T G_{\Theta,\Theta} B \, .$$

In a second step, we establish (4). If $P \subset W$, then there exists a $C \in \mathbb{R}^{n\times q}$ such that $\mathbf{P} = \Psi C$. Assuming (1), we find $C = G_{\Phi,\mathbf{P}}$ and $G_{\Theta,\mathbf{P}} = G_{\Theta,\Psi} C = G_{\Theta,\Theta} B G_{\Phi,\mathbf{P}}$. On the other hand, since Θ is a basis in $X \subset H$, $G^{-1}_{\Theta,\Theta}$ exists. Thus, if (4) is satisfied, $P \subset W$ follows from $\mathbf{P} = \Theta G^{-1}_{\Theta,\Theta} G_{\Theta,\mathbf{P}} = \Theta B G_{\Phi,\mathbf{P}} = \Psi G_{\Phi,\mathbf{P}}$. ■

Proposition 1 *For given subspaces $P, V \subset X \subset H$ and their bases $\mathbf{P}, \Phi, \Theta$ satisfying the above assumptions, there exists a subspace $W \subset X$ and its basis Ψ such that (1) and (2) are satisfied if and only if $G_{\mathbf{P},\Phi}$ has maximal rank $q = \dim P$.*

Proof We will use the result of Lemma 1. The necessity is obvious from (4). To proof the existence of W and Ψ we will find B from the system (3)-(4) using the SVD of A. Let $A = U(\Sigma, 0)^T Y^T$, where $U \in \mathbb{R}^{m\times m}$, $Y \in \mathbb{R}^{n\times n}$ are orthogonal, and $\Sigma \in \mathbb{R}^{n\times n}$ is diagonal and nonsingular. Obviously, B satisfies (3) if and only if it is of the form

$$B = G^{-1}_{\Theta,\Theta} U(\Sigma^{-1}, Z)^T Y^T \, , \qquad Z \in \mathbb{R}^{n\times(m-n)} . \tag{5}$$

It remains that the arbitrary matrix Z can be chosen such that (4) will be satisfied, too. Substituting the known factorizations for A and B, we obtain

$$G_{\mathbf{P},\Theta} U = G_{\mathbf{P},\Phi} B^T G_{\Theta,\Theta} U = G_{\mathbf{P},\Theta} A Y^T (\Sigma^{-1}, Z) = G_{\mathbf{P},\Theta} U \begin{pmatrix} \mathrm{Id}_n & \Sigma Z \\ 0 & 0 \end{pmatrix} .$$

Thus, (4) holds if and only if the matrix equation $0 = G_{\mathbf{P},\Theta} U (Z^T\Sigma, -\mathrm{Id}_{m-n})^T$ is satisfied. Using the factorization of $G_{\mathbf{P},\Phi}$, the latter can be rewritten as

$$G_{\mathbf{P},\Phi} Y Z = G_{\mathbf{P},\Theta} U (0, \mathrm{Id}_{m-n})^T \, . \tag{6}$$

Since the $q \times n$ matrix $G_{\mathbf{P},\Phi}$ has maximal rank q, solutions Z exist. ■

The above criterion does not depend on the particular choice of the bases Φ and $\mathbf{P}$ but rather on the choice of the spaces V and P themselves (it is equivalent to requiring that $(p, v)_H = 0$ for all $v \in V$ implies $p = 0$ for any $p \in P$). Equations (5)-(6) allow us to find all $\Psi \subset X$ dual to Φ, and such that the associated Q_W reproduces P. Whether this procedure is effective depends on the factorization of A, and the structure of $G_{\mathbf{P},\Phi}$ and $G^{-1}_{\Theta,\Theta}$. This is the place where the specific choices for $X \supset V, P$ and for the bases come in. In the subsequent sections, we specialize to the case $H = L_2(\Omega)$, to Lagrange finite element spaces V and corresponding spaces X of piecewise polynomials on a partition of $\Omega \subset \mathbb{R}^d$, $d = 1, 2$, and to P coinciding with the space P_r of all polynomials of degree $\leq r$. As we will see, the resulting matrices have then a simple, sparse structure, and can easily be computed, which in turn enables us to achieve additional properties of Ψ such as local support of all ψ_k.

1D construction

We consider the univariate finite element case. Let $\mathcal{T} = \{\Delta^n : n = 1, \ldots, N\}$ be the partition of a univariate interval $I = [a, b]$ into consecutive intervals Δ^n of length h_n. On the interval $[-1, 1]$, we define a special basis for P_r by

$$\Pi_r = [p_0, p_2, \ldots, p_r, p_1] \, ,$$

where $p_0(t) = (1-t)/2$, $p_1(t) = (1+t)/2$. The remaining polynomials $p_k(t)$ of degree $k = 2, \dots, r$ are supposed to vanish at $t = \pm 1$ and form an orthonormal system on $[-1,1]$ with respect to the $L^2(-1,1)$-scalar product. We note that $\dim P_r = r+1$ in 1D. Obviously, a basis in $X := \{v \in L^2(\Omega) \mid v_{|_{\Delta^n}} \in P_s,\ n = 1, \dots, N\}$, $s \geq r$, is given by

$$\Theta = [\Theta_{\Delta^1}, \dots, \Theta_{\Delta^N}] ,$$

where $\Theta_{\Delta^n} = [\theta_{0,n}, \dots, \theta_{s,n}]$ is the unscaled transformation of the system Π_s from $[-1,1]$ to Δ^n. For further reference, let $\Theta'_{\Delta^n} = [\theta_{1,n}, \dots, \theta_{s-1,n}]$. We note that Θ'_{Δ^n} is empty if $s = 1$. We restrict ourselves to the case that the conforming finite element space $V := X \cap H_0^1(I)$ satisfies homogeneous Dirichlet boundary condition. This is the interesting case for mortar finite elements. To obtain optimal results, the Lagrange multiplier space W which has by construction the same dimension as V has to be associated with the interior nodes on the interface. We found it convenient to use

$$\Phi = [\Theta'_{\Delta^1}, \theta_{s,1} + \theta_{0,2}, \Theta'_{\Delta^2}, \dots, \theta_{s,N-1} + \theta_{0,N}, \Theta'_{\Delta^N}]$$

as basis in V. In this form, it is sometimes called hierarchical finite element basis. Using the notation of the previous section, we can write $\Phi = \Theta\ U(\Sigma\ ,\ 0)^T$, where $\Sigma = \mathrm{diag}(\mathrm{Id}_{s-1}, \sqrt{2}, \mathrm{Id}_{s-1}, \sqrt{2}, \dots, \sqrt{2}, \mathrm{Id}_{s-1})$ and

$$U^T = \begin{pmatrix}
0 & \mathrm{Id}_{s-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & \dots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \mathrm{Id}_{s-1} & 0 & & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & & 0 & 0 & 0 \\
 & & \vdots & & & & \ddots & & & \\
0 & 0 & 0 & 0 & 0 & 0 & & \frac{1}{\sqrt{2}} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \mathrm{Id}_{s-1} & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 \\
0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \dots & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & & -\frac{1}{\sqrt{2}} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 1
\end{pmatrix} = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} . \quad (7)$$

We note that U is a $(s+1)N \times (s+1)N$ matrix, $U_1 \in \mathbb{R}^{sN-1 \times (s+1)N}$, $U_2 \in \mathbb{R}^{N+1 \times (s+1)N}$ whereas Σ is a $(sN-1) \times (sN-1)$ diagonal matrix. The dimension of X and V is $m = (s+1)N$ and $n = sN-1$, respectively. If $s = 1$, Id_{s-1} formally stands for a matrix of zero size. Having in mind Proposition 1, it is sufficient to find a subsystem $\Phi_r \subset \Phi$ of size $r+1$ such that $\det G_{\mathbf{P}_r, \Phi_r} \neq 0$ to guarantee the existence of a dual system Ψ satisfying $P_r \subset W$.

The remaining part of this section is devoted to the construction of suitable subsystems Φ_s for the case $r = s$ of maximal possible degree of polynomial reproduction, i.e., $P_s \subset W$. This is a little bit more than required in the mortar context, where optimal *a priori* error estimates can be obtained with a Lagrange multiplier space satisfying $P_{s-1} \subset W$. We assume $N \geq 3$ (the case $N \leq 2$, $s \geq 2$, can be dealt with as a simple linear algebra problem). Then, automatically, $\dim V \geq \dim P_s$. Consider any three consecutive intervals $\Delta^{n-1}, \Delta^n, \Delta^{n+1}$. Without loss of generality, after a suitable linear coordinate transform we can assume that Δ^n coincides with $[-1,1]$ and that the new intervals left and right to $[-1,1]$ have lengths $h = 2h_{n-1}/h_n$ and $h' = 2h_{n+1}/h_n$, respectively. We choose the system

$$\Phi_s = [\theta_{s,n-1} + \theta_{0,n}, \theta_{1,n}, \dots, \theta_{s-1,n}, \theta_{s,n} + \theta_{0,n+1}];, \quad (8)$$

and denote its transformation to $[-h-1,-1]\cup[-1,1]\cup[1,1+h']$ again by Φ_s. By construction, the transformations of $\theta_{1,n},\ldots,\theta_{s-1,n}$ have support in $[-1,1]$ and coincide with the functions $p_2,\ldots,p_s$ from Π_s while the remaining two functions are scaled piecewise linear hat functions.

In order to prove $\det G_{\mathbf{P}_s,\Phi_s}\neq 0$, we have a free choice of the basis in P_s. We will choose $\mathbf{P}_s=[\tilde{p}_0,\tilde{p}_1,\tilde{p}_2,\ldots,\tilde{p}_s]$ such that for $x\in[-1,1]$

$$\tilde{p}_0(x)=1-\sum_{k=2}^{s}\int_{-1}^{1}p_k(t)\,dt\cdot p_k(x)\ ,\quad \tilde{p}_1(x)=x-\sum_{k=2}^{s}\int_{-1}^{1}t\,p_k(t)\,dt\cdot p_k(x)\ ,$$

and $\tilde{p}_k(x)=p_k(x),\ k=2,\ldots,s$. We note the following properties:

(P1) Restricted to $[-1,1]$, the basis $\mathbf{P}_s$ is orthogonal, $\tilde{p}_0$ and $\tilde{p}_1$ are not normalized.

(P2) $\tilde{p}_0,\tilde{p}_2,\tilde{p}_4,\ldots$ are even, $\tilde{p}_1,\tilde{p}_3,\tilde{p}_5,\ldots$ are odd.

(P3) All zeros of $\tilde{p}_0,\tilde{p}_1$ are in $(-1,1)$, i.e., $\tilde{p}_0(t)>0$ for $t\in\mathbb{R}\backslash[-1,1]$ and $\tilde{p}_1(t)<0$ for $t<-1$ and $\tilde{p}_1(t)>0$ for $t>1$.

(P1) and (P2) follow from the definition. To prove (P3) for $\tilde{p}_0$, assume that it has zeros outside $[-1,1]$. Since $\tilde{p}_0$ is even and of even degree $2q'\leq s$, its zeros are symmetrically located with respect to the origin. Let $0<x_1<\ldots<x_k<1$ be the zeros of odd multiplicity inside $(0,1)$. By assumption at least one pair of zeros is outside $[-1,1]$ and thus $k\leq q'-1$. Recall also that by construction $\tilde{p}_0(-1)=\tilde{p}_0(1)=1$. Now, we define a polynomial

$$p(x):=(1-t^2)(t^2-x_1^2)\ldots(t^2-x_k^2)\ ,\quad t\in[-1,1]\ .$$

which has the same sign as $\tilde{p}_0$ everywhere in $(-1,1)$, with the exception of zeros of even multiplicity. Thus, $\int_{-1}^{1}p(t)\tilde{p}_0(t)\,dt>0$. This contradicts the orthogonality property since from $\deg(p)\leq 2q'\leq s$ and $p(-1)=p(1)=0$ we conclude that $p\in\operatorname{span}[p_2,\ldots,p_s]$. The same reasoning goes through for $\tilde{p}_1$, we leave this upon the reader. We are now in the position to show that $\det G_{\mathbf{P}_s,\Phi_s}\neq 0$ for the above Φ_s and $\mathbf{P}_s$. Equivalently, we show that orthogonality of

$$p=a_0\tilde{p}_0+a_1\tilde{p}_1+\sum_{k=2}^{s}a_k\tilde{p}_k$$

to all functions from Φ_s yields $a_k=0$ for all $k=0,\ldots,s$. Testing with the translates of $\theta_{l-1,n}$ (which coincide with $\tilde{p}_l|_{[-1,1]}$) and using (P1) immediately gives $a_l=0$ for all $l=2,\ldots,s$. Thus, only a_0 and a_1 can be different from zero. Now, we test $\tilde{p}_0$ and $\tilde{p}_1$ with the two remaining hat functions which will be denoted by $\hat{p}_1$ and $\hat{p}_2$. Let $\hat{p}_1$ be supported in $[-1-h,1]$ and $\hat{p}_2$ in $[-1,1+h']$. We recall that $\hat{p}_1$ and $\hat{p}_2$ are positive in $(-1-h,-1)$ and $(1,1+h')$, respectively. Moreover, in $[-1,1]$ they can be written as $\hat{p}_1=0.5(\tilde{p}_0-\tilde{p}_1)+\hat{q}_1$ and $\hat{p}_2=0.5(\tilde{p}_0+\tilde{p}_1)+\hat{q}_2$, where $\hat{q}_1,\hat{q}_2\in\operatorname{span}[p_2,\ldots p_s]$. Then, due to (P1) and (P3), the 2×2 determinant

$$\det G_{[\tilde{p}_0,\tilde{p}_1],[\hat{p}_1,\hat{p}_2]}=\begin{vmatrix}>0 & >0\\ <0 & >0\end{vmatrix}>0\ , \tag{9}$$

is positive. This shows $a_0=a_1=0$, and concludes the verification of $\det G_{\mathbf{P}_s,\Phi_s}\neq 0$. As a by-product, we see that the inverse $G^{-1}_{\mathbf{P}_s,\Phi_s}$ continuously depends on h,h'.

Proposition 2 *For the above defined basis Φ in the space V of C^0 finite elements of degree $s \geq 1$ on a partition $\mathcal{T}$ of an interval, there exists a locally supported dual basis Ψ consisting of piecewise polynomial functions of degree s on the same $\mathcal{T}$ such that $P_s \subset W$, and the associated projections Q_W and Q_V possess L_p-norm bounds ($1 \leq p \leq \infty$) which depend only on s, and on the local meshsize ratio $\gamma(\mathcal{T}) := \max_{|i-k|=1} h_i/h_k$.*

Proof The existence of a dual basis with $P_s \subset W$ has already been established. To construct a locally supported basis, we specify Z as follows: Obviously, the k-th column in Z is naturally associated with the vertex $x_k := a + \sum_{l=1}^{k-1} h_l$, $1 \leq k \leq N+1$, of $\mathcal{T}$, see (6) and the explicit form of U given in (7). For each k we chose three consecutive intervals Δ^{n_k}, Δ^{n_k+1} and Δ^{n_k+2} such that $k \in \{n_k - 1, n_k, n_k + 1, n_k + 2\}$. Then, we take the special choice of $\mathbf{P}_s$ and Φ_s given by (8) associated with these three consecutive intervals and define $\xi_k \in \mathbb{R}^{s+1}$, $1 \leq k \leq N+1$, by

$$G_{\mathbf{P}_s,\Phi_s}\xi_k = G_{\mathbf{P}_s,\Theta}(U_2)_k^T \ , \tag{10}$$

where $(U_2)_k$ is the k-th row of the matrix U_2. In a next step, we set the k-th column in Z by associating the components of ξ_k to the position in this column by correspondence to the functions in the chosen Φ_s and leaving zeros in positions corresponding to ϕ_l not in Φ_s. Note that we work with different subsystems Φ_s for the vertices x_k. Each column of Z has thus $\leq \dim P_s$ nonzero entries, associated with ϕ_l whose support is close to x_k by construction. Since (10) implies (6), we conclude that

$$B = G_{\Theta,\Theta}^{-1} U (\Sigma^{-1}, Z)^T \tag{11}$$

indeed defines a locally supported dual basis reproducing polynomials, with the supports of the ψ_l close to the supports of ϕ_l for all $l = 1, \ldots, \dim V$.

Since all steps in the construction depend only on the local meshsize ratio, the uniform L_p-stability bounds for the projections Q_W and Q_V (as well as local L_p-error estimates for smooth functions) can be derived. Since this is standard, we will not go into details. ■

Dual systems with basis functions of small support have been constructed in [Woh01] for $s \leq 2$ and $r = s - 1$. As was mentioned in the introduction, for the mortar finite element applications polynomial reproduction of degree $r = s - 1$ would suffice. Our above proof implies that for this case the construction of an adequate dual basis can be based on $\Phi_{s-1} = [\theta_{s-1,n-1} + \theta_{0,n}, \theta_{1,n}, \ldots, \theta_{s-2,n}, \theta_{s-1,n} + \theta_{0,n+1}]$ and $\mathbf{P}_{s-1}$, $s \geq 2$.

Higher order mortar finite elements

In this section, we establish optimal a priori estimates for the discretization error of nonconforming mortar finite element methods. These domain decomposition techniques provide a more flexible approach than standard conforming formulations, and are of special interest for time dependent problems, rotating geometries, diffusion coefficients with jumps, problems with local anisotropies, corner singularities and when different terms dominate in different regions of the simulation domain. To obtain optimal a priori estimates, the interface between the different regions has to be handled appropriately. Very often suitable matching conditions at the interfaces can be formulated as weak continuity conditions. We assume that the bounded polygonal subdomain $\Omega \subset \mathbb{R}^2$ is geometrically conforming decomposed in non-overlapping

polygonal subdomains Ω_k, $1 \le k \le K$. In particular, the situation with many crosspoints is included. Each subdomain is associated with a locally quasi-uniform simplicial or quadrilateral triangulation $\mathcal{T}_k$, and the discrete space of conforming finite elements of order s satisfying homogeneous Dirichlet boundary conditions on $\partial\Omega \cap \partial\Omega_k$ is denoted by $V_s(\Omega_k)$. On each interface $\gamma_{lk} := \partial\Omega_k \cap \partial\Omega_l$, we use the one-dimensional mesh inherited either from $\mathcal{T}_k$ or $\mathcal{T}_l$. The choice is arbitrary but fixed. Now, we replace the standard Lagrange multiplier space, see [BMP94], by our dual space. The basis functions of the Lagrange multiplier space $V_{s-1}(\gamma_{lk})$ on the interface γ_{lk} are defined as the scaled transformed dual basis functions $\psi_k \in \Psi$. Here Ψ is our locally supported basis consisting of piecewise polynomial functions of degree $s-1$ or s and reproducing polynomials of order $s-1$. Then, the constrained nonconforming mortar finite element space $Y_s \subset L^2(\Omega)$ is defined by

$$Y_s := \{v \in L^2(\Omega) \mid v_{|\Omega_k} \in V_s(\Omega_k), \int_{\gamma_{lk}} [v]\psi \, ds = 0, \, \psi \in V_{s-1}(\gamma_{lk}), \, 1 \le k < l \le K\}.$$

The analysis of the resulting jump terms across the interfaces plays an essential role for the a priori estimates of the discretization schemes. It is sufficient to analyze the jump term on the reference interface $I = [a, b]$. In particular, optimal methods can only be obtained if the consistency error is small enough compared with the best approximation error on the different subdomain. Indeed, in [Woh01, Conditions (Sa)–(Sd)], sufficient conditions for abstract Lagrange multiplier spaces are given to obtain a discretization error of order h^s and h^{s+1} in the H^1- and L^2-norm, respectively. For convenience, we briefly review the conditions. In a short form they read for dual spaces as: Locality of the support of the dual basis functions, polynomial reproduction of degree $s-1$, stability of the projections Q_V, Q_W and the existence of a well-defined stable operator $Q_{\hat{V}} : L^2(I) \to \hat{V}$. The projection $Q_{\hat{V}}$ will be defined by

$$\int_I Q_{\hat{V}} v\psi \, ds = \int_I v\psi \, ds, \quad \psi \in W.$$

Here, $\hat{V}$ is a subspace of $X \cap H^1(I)$ having the same dimension as V and satisfying a low order approximation property for all $v \in H^1(I)$. We point out that the required approximation property of $\hat{V}$ does not depend on the order s. For a more detailed discussion on the properties of $\hat{V}$, we refer to [Woh01]. We note that in the case of our locally supported dual basis Ψ, the best approximation property of the nonconforming space Y_s is automatically guaranteed. Since Q_V is by construction H_{00}-stable, no problem at the crosspoints occurs. The analysis of the consistency error requires the polynomial reproduction of degree $s-1$ and the existence of such a $Q_{\hat{V}}$. Of crucial importance is the weighted L^2-norm of the jump, $1/\sqrt{h}\|[v]\|_{0;\gamma_{lk}}$, $v \in Y_s$ across the interfaces γ_{lk}.

Lemma 2 *Replacing in the mortar finite element approach the standard Lagrange multiplier space on each interface γ_{lk} by our locally constructed dual space $V_{s-1}(\gamma_{lk})$ yields optimal a priori estimates for the discretization error in the L^2-norm (order h^{s+1}) and in the H^1-norm (order h^s). Moreover, the error in the Lagrange multiplier measured in a weighted L^2-norm, $\sqrt{h}\|[\cdot]\|_{0;\gamma_{lk}}$, is of order h^s.*

Proof Almost all required conditions, as locality, polynomial reproduction of degree $s-1$, and stability of Q_V, Q_W are satisfied by our above construction. Thus to establish the optimality, it is sufficient to define a suitable $\hat{V}$ and show that the corresponding projection $Q_{\hat{V}}$ is uniformly stable. The low order approximation property is, e.g., satisfied if $\tilde{V} :=$

$\operatorname{span}\{\theta_{0,1}+\theta_{s,1}+\theta_{0,2}, \theta_{s,2}+\theta_{0,3}, \ldots, \theta_{s,N-2}+\theta_{0,N-1}, \theta_{s,N-1}+\theta_{0,N}+\theta_{s,N}\}$ is a subspace of $\hat{V}$. Considering $\Phi = \Theta\, U(\Sigma, 0)^T$ and adding $\theta_{0,1}$ to the first basis function and $\theta_{s,N}$ to the last one provides a new set Φ_1 of linear independent functions. The associated space $\hat{V}$ satisfies $\tilde{V} \subset \hat{V}$, has the same dimension as V and a locally supported basis. In algebraic notation, we can write Φ_1 as $\Phi_1 := \Theta\, U(\Sigma, \hat{Z})^T$ where $\hat{Z} \in \mathbb{R}^{n\times m-n}$ has only two nonzero entries, $\hat{z}_{1,1} = \hat{z}_{n,m-n} = 1$. To show that $Q_{\hat{V}}$ is well defined, it is sufficient to prove that G_{Ψ,Φ_1} is non-singular. Using (11), we find $G_{\Psi,\Phi_1} = \mathrm{Id}_n + Z\hat{Z}^T$. The special structure of $\hat{Z}$ yields that the first and last column of $Z\hat{Z}^T$ is the first and last column of Z, respectively, all other columns are zero. Our construction of Z shows that the first column of Z depends on $\mathbf{P}_s$ or $\mathbf{P}_{s-1}$ which is associated with Δ^1, Δ^2, Δ^3. Since we have assumed $N \geq 3$, we find $z_{n,1} = z_{1,m-n} = 0$. Therefore it is sufficient to show that $1+z_{1,1}$ and $1+z_{n,m-n}$ are nonzero. Using the same notation as before $\tilde{p}_0$ and $\tilde{p}_1$ are orthonormal polynomials on Δ^2 and extended to I. The coefficient $z_{1,1}$ is the first component of the solution x of $G_{[\tilde{p}_0,\tilde{p}_1],[\hat{p}_1,\hat{p}_2]}x = y$, where $y_1 = \int_I \tilde{p}_0\theta_{0,1}dx$ and $y_2 = \int_I \tilde{p}_1\theta_{0,1}dx$. By means of (P3), we find $y_1 > 0$ and $y_2 < 0$ which together with (9) yields that $z_{1,1} > 0$. The same reasoning holds for $z_{n,m-n}$, and we obtain $\det G_{\Psi,\Phi_1} \neq 0$. ■

1D examples for $s \leq 3$

The aim of this section is to illustrate the above theoretical result, and to provide explicit formulas for $s \leq 3$, at least for the case of uniform partitions. We base the construction of our dual bases on (8). From now on we will assume that $\Omega = [0, N]$, $\Delta^n = [n-1, n]$, $n = 1, \ldots, N$, and $s \leq 3$. We refer to [OW00] for a detailed discussion and for an explicit representation of the matrices Z and B. For our convenience, we will fix the basis for Θ_{Δ^1} on $\Delta^1 = [0, 1]$ and obtain the bases Θ_{Δ^n} by translation:

$$\Theta_{\Delta^1} = \begin{cases} [\theta_1^1, \theta_1^2] = [1-t, t]\,, & s = 1\,, \\ [\theta_1^1, \theta_1^3, \theta_1^2] = [1-t, 6t(1-t), t]\,, & s = 2\,, \\ [\theta_1^1, \theta_1^3, \theta_1^4, \theta_1^2] = [1-t, 6t(1-t), 10t(1-t)(2t-1), t]\,, & s = 3\,. \end{cases}$$

This will lead to the following n-independent formulas for the diagonal blocks $G^{-1}_{\Theta_{\Delta^n},\Theta_{\Delta^n}}$ of $G^{-1}_{\Theta,\Theta}$:

$$G^{-1}_{\Theta_{\Delta^n},\Theta_{\Delta^n}} = \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix}, \quad \begin{pmatrix} 9 & -5 & 3 \\ -5 & 5 & -5 \\ 3 & -5 & 9 \end{pmatrix}, \quad \begin{pmatrix} 16 & -5 & 7 & -4 \\ -5 & 5 & 0 & -5 \\ 7 & 0 & 7 & -7 \\ -4 & -5 & -7 & 16 \end{pmatrix},$$

for $s = 1, 2, 3$, respectively. The explicit formulas for Σ and U only differ in the sizes of the identity matrices for different s. Thus, we have all ingredients ready for using (11), with the exception of the matrix Z. The construction of Z is described in the proof of Proposition 2, and depends on s and the desired degree of polynomial reproduction $r \leq s$. We refer to [OW00] for the calculation of the entries of Z. Although we provide explicit results only for uniform partitions, the construction can be used for non-uniform partitions without essential changes. For $s = 1$, the dual basis functions ψ_n obtained along the lines of the previous

section are as follows. Away from the endpoints of Ω, we get

$$\psi_n(x) := \begin{cases} -\frac{2}{3}\theta^1_n + \frac{4}{3}\theta^2_n, & x \in [n-1, n), \\ \frac{7}{3}\theta^1_{n+1} - \frac{2}{3}\theta^2_{n+1}, & x \in [n, n+1], \\ -\frac{2}{3}\theta^1_{n+2} + \frac{1}{3}\theta^2_{n+2}, & x \in (n+1, n+2], \end{cases} \qquad 3 \le n \le N-3\,,$$

for the *interior dual basis functions* (here and below, we only show formula for the intervals in the support of ψ_n). For the *boundary dual basis functions* near the left and right endpoint of Ω, we obtain modified expressions:

$$\psi_1(x) := \begin{cases} 2\theta^1_1 + \theta^2_1, & x \in [0,1), \\ \theta^1_2, & x \in [1,2], \\ -\frac{2}{3}\theta^1_3 + \frac{1}{3}\theta^2_3, & x \in (2,3), \end{cases} \quad \psi_2(x) := \begin{cases} -\theta^1_1, & x \in [0,1), \\ \theta^2_2, & x \in [1,2], \\ \frac{7}{3}\theta^1_3 - \frac{2}{3}\theta^2_3, & x \in (2,3], \\ -\frac{2}{3}\theta^1_4 + \frac{1}{3}\theta^2_4, & x \in (3,4], \end{cases}$$

ψ_{N-2}, ψ_{N-1} are defined in a similar way. Now, it is easy to verify that the locally supported basis functions ψ_n are biorthogonal with the standard hat functions. Furthermore, $\sum_{n=1}^{N-1} \psi_n = 1$ and $\sum_{n=1}^{N-1} n\, \psi_n = x$ and thus $P_1 \subset W$. We note that in [Woh00, Woh01], a dual basis with smaller support but only $P_0 \subset W$ has been constructed.

For $s = 2$, we introduce a dual basis satisfying $P_2 \subset W$. We distinguish between two different types of dual basis functions ψ^b_n and ψ^h_n which are associated with the bubble and hat functions of the finite element basis functions, respectively. The interior dual basis functions ψ^b_n with support on $\Delta_{n-1} \cup \Delta_n \cup \Delta_{n+1}$ ($n = 3, \ldots, N-2$) and ψ^h_n with support on $\Delta_n \cup \Delta_{n+1}$ ($n = 2, \ldots, N-2$) are defined by the corresponding θ^i_k, $1 \le i \le 3$, as follows:

$$\psi^b_n(x) := \begin{cases} -\frac{3}{2}\theta^1_{n-1} + \frac{5}{6}\theta^3_{n-1} - \frac{1}{2}\theta^2_{n-1} \\ -3\theta^1_n + \frac{10}{3}\theta^3_n - 3\theta^2_n \\ -\frac{3}{2}\theta^1_{n+1} + \frac{5}{6}\theta^3_{n+1} - \frac{1}{2}\theta^2_{n+1} \end{cases}, \quad \psi^h_n(x) := \begin{cases} \frac{3}{2}\theta^1_n - \frac{5}{2}\theta^3_n + \frac{9}{2}\theta^2_n \\ \frac{9}{2}\theta^1_{n+1} - \frac{5}{2}\theta^3_{n+1} \frac{3}{2}\theta^2_{n+1} \end{cases}.$$

The modifications for the boundary dual basis functions ψ^b_1, ψ^b_2, and ψ^h_1 concern only their values on Δ_1, otherwise the above formula apply correspondingly. We define on Δ_1: $\psi^b_1 := 33/8\theta^1_1 - 5/8\theta^3_1 - 5/8\theta^2_1$, $\psi^b_2 := 17/8\theta^1_1 - 5/8\theta^3_1 - 5/8\theta^2_1$, and $\psi^h_1 := -21/4\theta^1_1 - 5/4\theta^3_1 - 9/4\theta^2_1$. The modifications for ψ^b_{N-1}, ψ^b_N, and ψ^h_{N-1} on Δ_N are analogous. Note that the support of the dual basis functions ψ^b_n, ψ^h_n is contained in ≤ 3 neighboring Δ^n close to the support of the corresponding finite element basis functions. Moreover, we find by construction $P_2 \subset W$.

For $s = 3$, we do not specify the explicit formulas for the basis functions and refer to [OW00] for details. Figure 1 illustrates the interior dual basis functions, $s = 3$, for $P_2 \subset W$ and $P_3 \subset W$, respectively. We have three different types associated with hat functions, quadratic, and cubic bubbles, and supports consisting of two, three and one/two consecutive intervals, respectively.

2D results

The above approach generalizes to higher dimensions, as we demonstrate with the following example. We consider the space V of quadratic Lagrange C^0-elements, i.e., $s = 2$, with homogeneous Dirichlet boundary conditions on a triangulation $\mathcal{T}$ of a bounded polygonal domain $\Omega \subset \mathbb{R}^2$, and show the existence of a dual basis of locally supported piecewise quadratics on $\mathcal{T}$ such that W reproduces linear polynomials locally, i.e., $P_1 \subset W$, under a certain

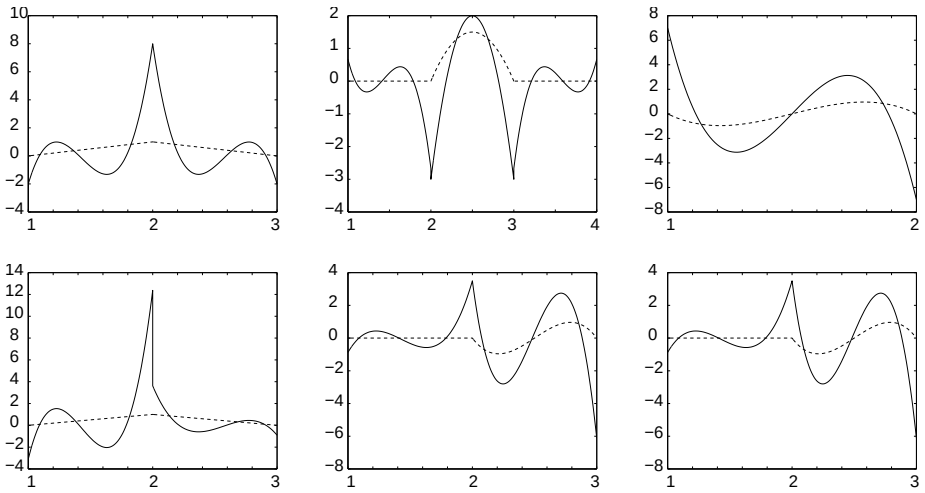


Figure 1: Interior dual basis functions $P_r \subset W$, $r = 2$ (above) and $r = 3$ (below)

regularity condition on $\mathcal{T}$. For lowest order finite elements, dual basis functions satisfying $P_0 \subset W$ have been constructed in [KLPV01, WK01, Woh01].

The basis Θ in X which is set to be the space of discontinuous piecewise quadratics is conveniently given by the collection of all elemental nodal shape functions $\theta_{\Delta,P}$, piecewise linear barycentric coordinate function for vertex P of triangle Δ, and $\theta_{\Delta,e}$, quadratic tent function associated with triangle Δ and its edge e. Each such function is supported on a single triangle, and there are 6 of them for each Δ. As before, an explicit, sparse factorization of A can be found (see [OW00]), and Proposition 1 can be applied. Following the considerations of Section 8, it is sufficient to find a locally defined subsystem Φ_1 such that $\det G_{\Phi_1,\mathbf{P}_1} \neq 0$. Let $\Delta \in \mathcal{T}$ be any triangle all edges of which are interior to Ω. We specify a basis $\mathbf{P}_1$ of P_1 by setting $\mathbf{P}_1 := [p_1, p_2, p_3]$ where p_k denotes the extension of the barycentric coordinate function associated with the vertex P_k of Δ to all of $\mathbb{R}^2$ which is defined by requiring $p_k \in P_1$ and $p_k(P_l) = \delta_{kl},\ k, l = 1, 2, 3$. The subsystem Φ_1 is defined by $[\phi_{e_1}, \phi_{e_2}, \phi_{e_3}]$ where ϕ_{e_k} denote the conforming quadratic bubble functions associated with the edges e_k of the triangle Δ. Using the affine invariance of both Φ_1 and $\mathbf{P}_1$ we can without loss of generality assume that Δ is equilateral, with area $A = 1$. All the other notation can be found in the left of Figure 2. The area of the triangle Δ_k, attached to Δ along e_k, is denoted by A_k.

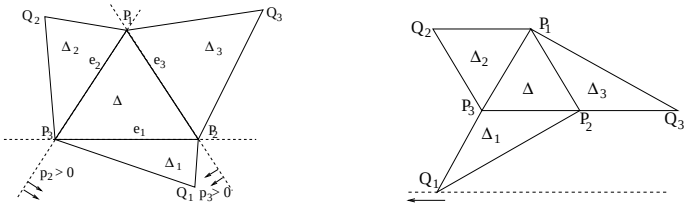


Figure 2: Notation for Lemma 3 (left) and counterexample (right)

Lemma 3 *Let the triangles in the left part of Figure 2 satisfy the following condition: For each $k = 1, 2, 3$, the diagonal $P_k Q_k$ belongs to the closure of the corresponding quadrilateral $\Delta \cup \Delta_k$. Then the determinant of $G_{\mathbf{P}_1,\Phi_1}$ is positive and depends continuously on the location of the Q_k. If the additional geometric assumption is dropped, the matrix $G_{\mathbf{P}_1,\Phi_1}$ may become singular.*

Proof The proof is based on elementary calculations. We start by stating the formula

$$\int_\Delta \theta_{\Delta,e_1} \cdot \sum_{k=1}^{3} \alpha_k p_k \, dx = \frac{A}{15}(\alpha_1 + 2(\alpha_2 + \alpha_3)) \ ,$$

which holds, due to affine invariance of all functions involved, for all triangles. This allows us to compute all scalar products necessary for $G_{\Phi_1,\mathbf{P}_1}$. E.g.,

$$\int_\Omega \theta_{\Delta,e_1} \cdot p_1 \, dx = \int_\Delta \theta_{\Delta,e_1} \cdot p_1 \, dx + \int_{\Delta_1} \theta_{\Delta,e_1} \cdot p_1 \, dx = \frac{A}{15} + \frac{A_1}{15} p_1(Q_1) = \frac{1 - A_1^2}{15}$$

since $p_1(Q_1) = -A_1/A = -A_1$. Since $p_1 + p_2 + p_3 \equiv 1$, we have $p_2(Q_1) + p_3(Q_1) = 1 - p_1(Q_1) = 1 + A_1$ which leads to the ansatz

$$p_2(Q_1) = \frac{1 + A_1 - \epsilon_1}{2} \ , \quad p_3(Q_1) = \frac{1 + A_1 + \epsilon_1}{2} \ ,$$

where our geometric assumption implies that $\min(p_2(Q_1), p_3(Q_1)) \geq 0$ or, equivalently, $|\epsilon_1| \leq 1 + A_1$. With this at hand, we compute

$$\int_\Omega \theta_{\Delta,e_1} \cdot p_2 \, dx = \frac{2A}{15} + \frac{A_1}{15}(2 + p_2(Q_1)) = \frac{4 + 5A_1 + A_1^2 - A_1\epsilon_1}{30} \ ,$$

and, analogously,

$$\int_\Omega \theta_{\Delta,e_1} \cdot p_3 \, dx = \frac{4 + 5A_1 + A_1^2 + A_1\epsilon_1}{30} \ .$$

Applying the same analysis to the other rows of $G_{\Phi_1,\mathbf{P}_1}$ and observing that the rows almost completely divide by $(1 + A_k)$, we get the following explicit formula

$$\frac{30^3 \ \det G_{\Phi_1,\mathbf{P}_1}}{(1 + A_1)(1 + A_2)(1 + A_3)} = D \equiv \begin{vmatrix} 2 - 2A_1 & 4 + A_1 - \epsilon_1' & 4 + A_1 + \epsilon_1' \\ 4 + A_2 + \epsilon_2' & 2 - 2A_2 & 4 + A_2 - \epsilon_2' \\ 4 + A_3 - \epsilon_3' & 4 + A_3 + \epsilon_3' & 2 - 2A_3 \end{vmatrix} \ ,$$

where $|\epsilon_k'| = A_k|\epsilon_k|/(1 + A_k) \leq A_k$, $k = 1, 2, 3$, follows from our assumption.

A straightforward calculation reveals that

$$D = 10(3s_2(\mathbf{A}) + 4s_1(\mathbf{A}) + 4 + f(\mathbf{A}, \epsilon')) \ ,$$

where $s_1(\mathbf{x}) = x_1 + x_2 + x_3$, $s_2(\mathbf{x}) = x_1x_2 + x_2x_3 + x_3x_1$ for any $\mathbf{x} \in \mathbb{R}^3$, and

$$f(\mathbf{A}, \epsilon') = s_2(\epsilon') + \epsilon_1'(A_2 - A_3) + \epsilon_2'(A_3 - A_1) + \epsilon_3'(A_1 - A_2) \ .$$

The global minimum of f with respect to the cube $\epsilon'_k \in [-A_k, A_k]$, $k = 1, 2, 3$, is attained on the boundary of this cube, and can be determined easily:

$$f(\mathbf{A}, \epsilon') \geq s_2(\mathbf{A}) - 4\max(A_1A_2, A_2A_3, A_3A_1) \geq -3s_2(\mathbf{A})$$

holds for all ϵ'_k of interest. Substitution gives

$$D \geq 40(s_1(\mathbf{A}) + 1) > 40\ . \tag{12}$$

since $A_k > 0$, $k = 1, 2, 3$. This shows the assertions of Lemma 3 under the geometric assumptions made. The continuous dependence of the determinant and thus the inverse of $G_{\mathbf{P}_1,\Phi_1}$ on the local topology is obvious.

It remains to provide a counterexample that shows that the above choice for Φ_1 may fail to guarantee the invertibility of $G_{\mathbf{P}_1,\Phi_1}$. The right part of Figure 2 contains the counterexample. We claim that if Q_1 is moved to the left, the determinant of $G_{\mathbf{P}_1,\Phi_1}$ will vanish at some point. Indeed, the specification of the example is such that $A = A_1 = A_2 = A_3 = 1$, both Δ and Δ_2 are equilateral (thus, $\epsilon'_2 = 0$), and $\epsilon'_3 = 1$ since Q_3 belongs to the extension of e_1. Thus, according to the above formula, $D = \alpha\epsilon'_1 + \beta$ is a linear function with respect to ϵ'_1, with slope $\alpha = 10(\epsilon'_2 + \epsilon'_3 + A_2 - A_3) = 10$ and $\beta = 250$ (since for $\epsilon'_1 = 0$ the geometric assumption is satisfied and therefore (12) is valid). Thus, moving Q_1 sufficiently far to the left or, equivalently, decreasing ϵ'_1, we finally hit a zero value for D. This proves our claim. ■

Remark 1 *One possible modification is to start the construction of dual bases with a finite element space X corresponding to a refined partition $\mathcal{T}'$ rather than with the space of non-smooth piecewise polynomials on the same $\mathcal{T}$. This could make the resulting W suitable for applications, where higher smoothness of the functions in the dual system is required. However, for use as Lagrange multiplier subspaces of $H^{-1/2}$ in the mortar finite element method this is not essential.*

Remark 2 *In contrast to constructions of biorthogonal wavelet systems [DKU99, DS97, Ste00], the spaces W obtained here are not refinable, i.e., if $\mathcal{T}'$ is a proper refinement of $\mathcal{T}$, we cannot expect to have $W' \supset W$. However, as suggested in a similar problem in [Osw99], we still have refinability $X' \supset X$ for the container spaces of piecewise polynomials which enables the use of our systems in a multilevel setup.*

Acknowledgement: The authors would like to thank C. de Boor and R. Verfürth for their interest and fruitful comments.

References

[BMP94]Christine Bernardi, Yvon Maday, and Anthony T. Patera. A new non conforming approach to domain decomposition: The mortar element method. In Haim Brezis and Jacques-Louis Lions, editors, *Collège de France Seminar*. Pitman, 1994. This paper appeared as a technical report about five years earlier.

[dB76]Carl de Boor. On local linear functionals which vanish at all b-splines but one. In A. G. Law and B. N. Sahney, editors, *Theory of Approximation with Applications*, pages 120–145, New York, 1976. Academic Press.

[dB90]Carl de Boor. *Splinefunktionen*. Birkhäuser, Basel, 1990.

[dBF73]Carl de Boor and George Fix. Spline approximation by quasi-interpolants. *J. Approx. Th.*, 8:19–45, 1973.

[DKU99]Wolfgang Dahmen, Angela Kunoth, and Karsten Urban. Biorthogonal wavelets on the interval - stability and moment conditions. *Appl. Comp. Harmon. Anal.*, 6:132–196, 1999.

[DS97]Wolfgang Dahmen and Robert Stevenson. Element-by-element construction of wavelets satisfying stability and moment conditions. Technical Report 145, IGPM, RWTH Aachen, 1997.

[KLPV01]Chisup Kim, Raytcho Lazarov, Joseph Pasciak, and Panayot Vassilevski. Multiplier spaces for the mortar finite element method in three dimensions. *SIAM J. Numer. Anal.*, 39:519–538, 2001.

[Osw99]Peter Oswald. Interface preconditioners and multilevel extension operators. In C.-H. Lai et al., editor, *Proc. 11th Intern. Conf. on Domain Decomposition Methods in Science and Engineering*, pages 96–103. Domain Decomposition Press, 1999.

[OW00]Peter Oswald and Barbara Wohlmuth. On polynomial reproduction of dual fe bases. Technical Report 10009640-000512-07, TechBell Laboratories, Lucent Technologies, 2000. extended version.

[Sch81]Larry L. Schumaker. *Spline Functions: Basic Theory*. Wiley, New York, 1981.

[Ste00]Rob Stevenson. Locally supported, piecewise polynomial biorthogonal wavelets on non-uniform meshes. Technical report, Uni. of Utrecht, 2000.

[WK01]Barbara Wohlmuth and Rolf Krause. Multigrid methods based on the unconstrained product space arising from mortar finite element discretizations. *SIAM J. Numer. Anal.*, 39:192–213, 2001.

[Woh00]Barbara Wohlmuth. A mortar finite element method using dual spaces for the Lagrange multiplier. *SIAM J. Numer. Anal.*, 38:989–1012, 2000.

[Woh01]Barbara Wohlmuth. *Discretization Methods and Iterative Solvers Based on Domain Decomposition*, volume 17 of *Lecture Notes in Computational Science and Engineering*. Springer, 2001.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

9 Decomposition Algorithms for DDM

Olivier Pironneau[1] and Stéphane Del Pino[2] and Jacques-Louis Lions[3]

Control and DDM

We present here decomposition algorithms similar to Schwarz' for the numerical solution of the elliptic and parabolic problems in complex domain. They are well suited to domains described by Constructive Solid Geometry (CSG), set operations on simple shapes, a data structure often used in image synthesis and Virtual Reality[BC94].
We introduce briefly also the decomposition of evolution problems into subproblems on overlapping and nonoverlapping subdomains obtained with Lagrange multipliers without referring to an optimization problem so as to avoid two point boundary value problems.

This paper summarizes several earlier ones, parts of a long term project aimed at solving PDEs with the data structures of VR [LP99b][LP99c] [LP99a] [LP98b] [LP98a] [GLP99] [HLP99] [BLP01]. It is being implemented into freefem3d, a user-friendly, language driven PDE solver. Freefem3d takes VRML[HW96] data and POV-Ray input (http://www.povray.org); it uses the fictitious domain method with finite element discretization and it is well suited to DDM[BW86] at the algorithmic level.

Consider the problem of adjusting v so that $y(v)$ be nearest to $y_0 \in L^2(\mathcal{O})$ and subject to

$$y \in V, \quad a(y,\hat{y}) = s(v,\hat{y}) \quad \forall \hat{y} \in V. \tag{1}$$

where

$$s(v,\hat{y}) = \int_\Sigma v\hat{y}d\Gamma, \quad v \in L^2(\Sigma) \tag{2}$$

Naturally it can be solved by minimizing

$$J(v) = \frac{\alpha}{2}s(v,v) + \frac{1}{2}\int_{\mathcal{O}}(y-y_0)^2dx \tag{3}$$

and the method is feasible for any non empty $\mathcal{O} \subset \Omega$.

When $\Omega = \cup\Omega_i$ and $\Omega_i \cap \Omega_j \neq \emptyset$ (see Figure 4 for the notations) we can combine optimal control and domain decomposition.

Consider the solutions of

$$a_i(y_i,\hat{y}_i) = s_i(v_i,\hat{y}_i) + \sum_j \int_{\Gamma_{ij}} \lambda_i\hat{y}_i d\Gamma \tag{4}$$

[1]UP6: pironneau@ann.jussieu.fr
[2]UP6: Université Paris VI
[3]Collège de France

Let

$$\text{supp } \tau_i \subset \mathcal{O} \cap \Omega_i, \sum \tau_i = 1 \qquad c_i(f_i) = \int_{\mathcal{O}} \tau_i f_i^2 dx. \tag{5}$$

$$K(v,\lambda) = \frac{\alpha}{2}\sum_i s_i(v_i) + \frac{1}{2}\sum_i c_i(y_i - y_{oi}) + \frac{1}{2}\sum_{i,j} \|\lambda_i\|^2_{L^2(\Gamma_{ij})}, \tag{6}$$

where $v = \{v_i\} \in \Pi L^2(\Sigma \cap \bar{\Omega}_i)$. We solve

$$\inf_{v,\lambda} K(v,\lambda), \quad \text{subject to } r_{ji} y_i = r_{ij} y_j \tag{7}$$

The key point is to observe that this problem decouples when solved by a gradient method so that the overhead is small when control is added to DDM.

Note that the method works also when the problem is only to solve a PDE like

$$y \in V, \quad a(y,\hat{y}) = (f,\hat{y}) \quad \forall \hat{y} \in V. \tag{8}$$

Then the (virtual) control is an artefax to convert the problem into an optimization problem which decouples on each sub-domain.
Some numerical results are shown on Figure 3 for a Laplace equation (see [LP99c] for more details). It shows that the cost and convergence is comparable to Schwarz' (see [Lio78]).
The same is true of time dependent partial Differential equations (see [LP99a]) .

Virtual Control

The previous exercise leads us to investigate "virtual controls" in a more general framework. To solve

$$a_\Omega(u,\hat{u}) = (f,\hat{u}) \quad \forall \hat{u} \in V = H_0^1(\Omega). \tag{9}$$

we introduce the *virtual controls* $\lambda_1, \lambda_2 \in L^2(\mathcal{O})$ with $\lambda_1 + \lambda_2 = 0$ (see figure 4). Let a_i be the same operator as a but with integrals on Ω_i. Then solve by a conjugate gradient algorithm for example:

$$\min_{\lambda_i} \mathcal{J}(\lambda) = \frac{\varepsilon}{2}\int_{\mathcal{O}} (\lambda_1^2 + \lambda_2^2) dx + \frac{1}{2}\sum_i \int_{\Gamma_{ij}} u_i^2 \tag{10}$$

subject to

$$a_{\Omega_1}(u_1,\hat{u}_1) = (f_1,\hat{u}_1) + (\lambda_1,\hat{u}_1)_{\mathcal{O}} \quad \forall \hat{u}_1 \in V_1, \tag{11}$$

$$a_{\Omega_2}(u_2,\hat{u}_2) = (f_2,\hat{u}_2) + (\lambda_2,\hat{u}_2)_{\mathcal{O}} \quad \forall \hat{u}_2 \in V_2 \tag{12}$$

with $V_i = H^1(\Omega_i) \cap V$. Notice that the solution to (9) is the sum of functions each having its support in Ω_i (see figure 5): $y = u_1 + u_2$. At the solution $\partial u_1/\partial n = u_1 = 0$ on $\Omega_2 \cap \partial\Omega_1$

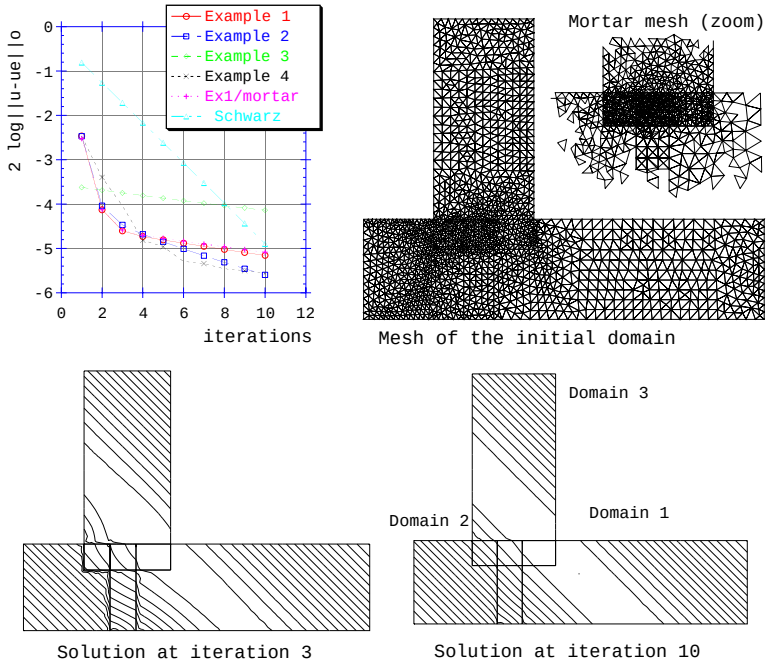


Figure 1: *A Laplace equation whose solution is $y = x_1 + x_2$ is solved on a domain decomposed into 3 subdomains by virtual controls on their common boundaries. Several formulations are compared with the Schwarz algorithm. Each domain has its own mesh, so the method is non-conforming*

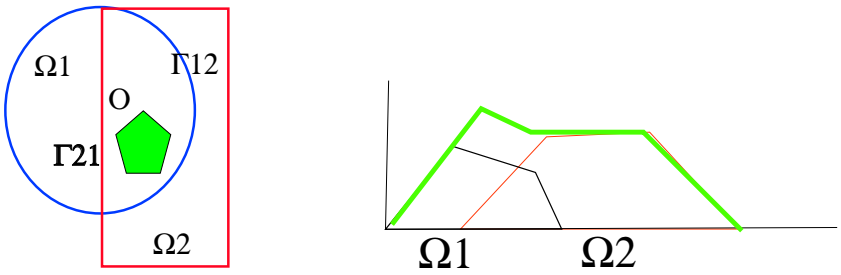


Figure 2: **Left** *The virtual control set $\mathcal{O}$ is in $\Omega_1 \cap \Omega_2$.* **Right** *Decomposition of a function (thick line) into two functions with support each in Ω_i, $i = 1, 2$*

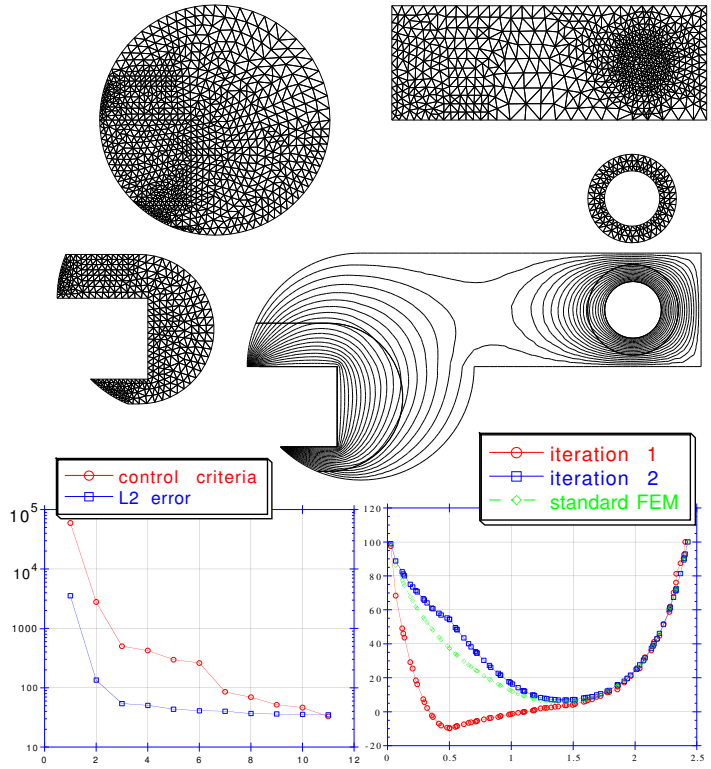


Figure 3: *Computation of a Laplace equation on a four piece domain by the method of virtual control and comparison with Schwarz' algorithm*

Decomposition of the Space

While introducing the concept of virtual control there was another important idea when we wrote that the solution is the sum of functions with support in each sub-domain; in fact a decomposition of the variational space can be used:

$$H_0^1(\Omega) = H_0^1(\Omega_1) + H_0^1(\Omega_2) \tag{13}$$

Consider again a simple elliptic problem like

$$u|_\Gamma = 0 \qquad -\Delta u = f \text{ in } \Omega = \Omega_1 \cup \Omega_2 \quad \Omega_1 \cap \Omega_2 \neq 0 \tag{14}$$

$u = u_1 + u_2$, $u_i^n|_{\Omega_i} \in H_0^1(\Omega_i)$.

Optimal control is not the only tool to apply the decomposition of the space; the fixed point algorithm for instance works too. Let u_i^n be defined recursively by

$$\begin{array}{l} \beta(u_1^{n+1} - u_1^n) - \Delta(u_1^{n+1} + u_2^n) = f \text{ in } \Omega_1 \\ \beta(u_2^{n+1} - u_2^n) - \Delta(u_1^n + u_2^{n+1}) = f \text{ in } \Omega_2 \end{array} \tag{15}$$

Such an iterative scheme is a sort of regularized Schwarz algorithm (it is Schwarz when $\beta = 0$); converges is shown in [HLP99] for the continuous case and in [BLP01] for the discrete case, the numerical difficulty being the evaluation of mixed integrals like

$$\int_{\Omega_1 \cap \Omega_2} \nabla u_1 \cdot \nabla \hat{u}_2 \tag{16}$$

Convergence

More precisely to evaluate

$$a_h(u_1 + u_2, w_1 + w_2) = a(u_1, w_1) + a(u_2, w_2) + a_h(u_1, w_2) + a_h(u_2, w_1) \tag{17}$$

we use the following quadrature with quadrature points $\{\xi_{jk}^i\}_{j=1..J}$, $i = 1, 2$ in triangle T_k:

$$a_h(u, v) = \frac{1}{2} \sum_k |T_k^1| \left(\nabla u(\xi_{jk}^1) \cdot \nabla v(\xi_{jk}^1) \right) + \frac{1}{2} \text{ idem on } T_k^2. \tag{18}$$

Proposition(F. Brezzi) *When the quadrature points are the vertices of both triangulations (18) is an admissible quadrature for a and a coercive bilinear forms Furthermore the fixed point algorithm converges when discretized with P^1 elements and the error is optimal.*

Chimera

Chimera as introduced by Steger[SB87] is a Schwarz algorithm with $\Omega = \mathcal{C} \backslash \mathcal{O} = \Omega_1 \cup \Omega_2$. For instance, potential flow around an airfoil involves solving Laplace's equation in a domain outside the airfoil[Pir87].

$$-\Delta \psi_i^{n+1} = f \qquad \psi_1^{n+1}|_{\Gamma_{12}} = \psi_2^n \quad \psi_2^{n+1}|_{\Gamma_{21}} = \psi_1^n \tag{19}$$

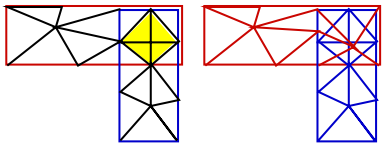


Figure 4: *With non matching grid one must compute integrals of products of functions piecewise linear on each grid.*

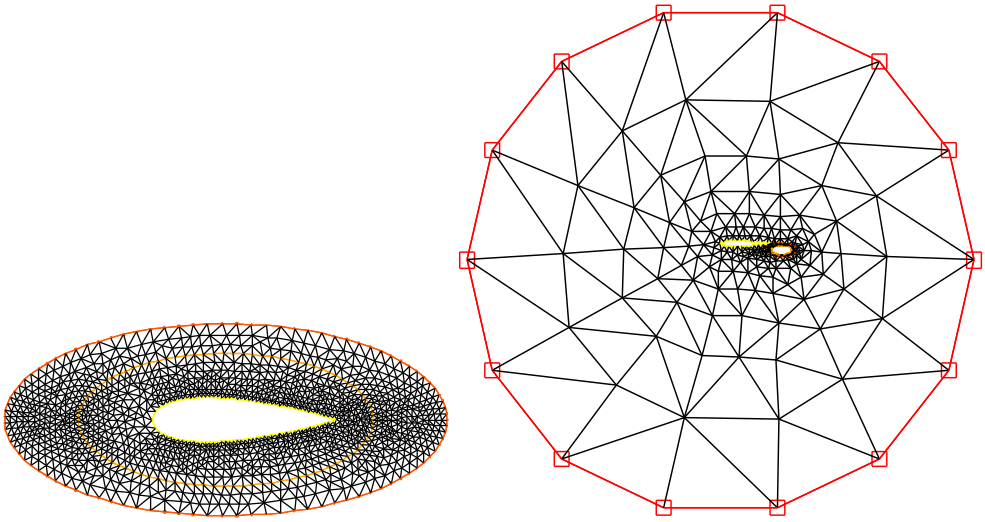


Figure 5: *Meshes and domain decomposition to compute the stream function around a two-pieces airfoil, namely the solution of* $\Delta\psi = 0$ *with Dirichlet data by the Chimera method. A finer mesh is built around the smaller airfoil (on the left) and a coarse mesh for the rest of the domain, with an elliptic hole in place of the small airfoil (the scale for both domains is not the same on this picture). The whole domain is the union of the fine and coarse domains.*

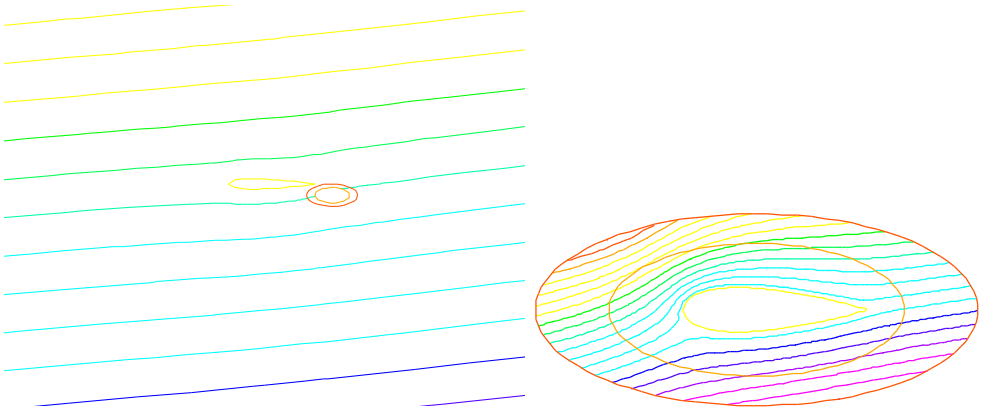


Figure 6: *Stream function around a two-pieces airfoil (namely solution of* $\Delta\psi = 0$ *with Dirichlet data) by the Chimera method (i.e. Schwarz algorithm). The convergence is obtained after 4 iterations.*

Here too we can use a decomposition of the variational space and therefore prove convergence of the Chimera method for arbitrary meshes. In our numerical test the domain is the region outside two airfoils and within a circle which approximates infinity. The finite element method of order one on triangles has been used. The domain is divided in two: a domain near the airfoil which is triangulated with small triangles and the rest of the domain which uses bigger triangles. Here the domain has two airfoils, a large one and a small one. The decomposition must be such that the physical domain is the union of both domain, and the domains must overlap. Then Schwarz algorithm is used with translation and quadratures at the vertices as explained above. Four iterations are sufficient for convergence to machine accuracy (figures 6 and 7.)

Decomposition of Operators

Our last idea is in the family of operator splitting methods. Consider

$$
\begin{aligned}
a_i(u,\hat{u}) &= \sum \int_{\Omega_i} \rho_i a_{kl} \frac{\partial u}{\partial x_l} \frac{\partial \hat{u}}{\partial x_k}, \\
c_i(u,\hat{u}) &= \int_{\Omega_i} \sigma_i u \hat{u} dx, \\
\rho_1 + \rho_2 &= 1, \quad \sigma_1 + \sigma_2 = 1 \text{ in } \Omega
\end{aligned}
$$

($\rho_i \geq 0,\ \sigma_i \geq 0$ are extended by 0 outside Ω_i).

Let $\epsilon_1, \epsilon_2, \eta_1, \eta_2$ be positive and small; we now introduce the system

$$
\begin{aligned}
c_i(\frac{\partial u_i}{\partial t}, \hat{u}_i) + a_i(u_i, \hat{u}_i) + (\hat{u}_i, \lambda_i - \lambda_j)_{H_{ij}} &= c_i(f, \hat{u}_i) \\
\epsilon_i(\frac{\partial \lambda_i}{\partial t}, \hat{\lambda}_i)_{H_{ij}} + \eta_i b_i(\lambda_i, \hat{\lambda}_i) - (\hat{\lambda}_i, u_i - u_j)_{H_{ij}} &= 0
\end{aligned}
$$

This method works with/without overlapping.

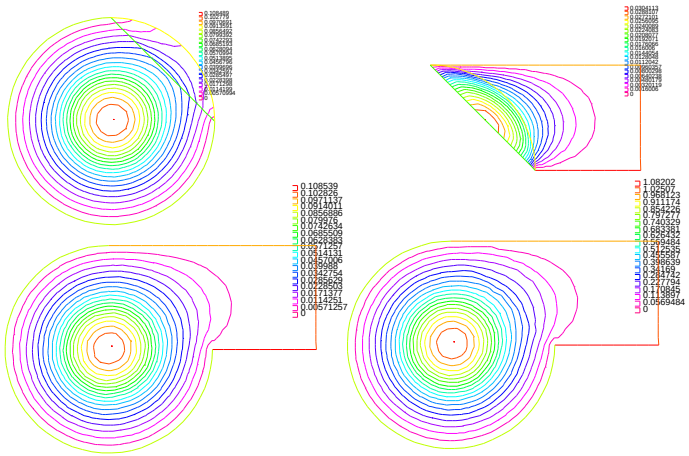


Figure 7: *Decomposition of operator: a Laplace equation solved on a composite domain with overlapping.*

Numerical Example

We consider the heat equation

$$\partial_t u - \Delta u = 1_D \text{ in } \Omega \times (0,T) \tag{20}$$

with zero initial and boundary conditions.

The domain $\Omega = \Omega_1 \cup \Omega_2$ is made of the unit circle centered at the origin and a rectangle $(0,2) \times (0,1)$.

The source term is in a disk centered at the origin and of radius 0.4.

The algorithm is

$$\begin{aligned} &\frac{1}{\delta t}(u_i^{m+1} - u_i^m) - \Delta u_i^{m+1} = 1_D + (-1)^i 1_C \lambda, \text{ in } \Omega_i \\ &\frac{1}{\delta t}(\lambda^{m+1} - \lambda^m) - \Delta \lambda^{m+1} = \frac{1}{\epsilon}(u_1 - u_2) \text{ in } C \end{aligned} \tag{21}$$

with Dirichlet conditions on $\partial\Omega$ and Neumann conditions on ∂C, where $C = \Omega_1 \cap \Omega_2$.

The parameters of the computations are $\delta t = 0.015$, $\epsilon = 0.01$ and the results are shown on figure 7 For problems with discontinuous coefficients the method without overlapping is more attractive. We consider the convection-diffusion equation

$$\partial_t u + \mathbf{v} \cdot \nabla u - \nabla \cdot (\nu \nabla u) = 0 \text{ in } \Omega \times (0,T) \tag{22}$$

with initial and boundary conditions.

The problem is in $\Omega = (0,1) \times (0,2)$. It is an academic example of the dissipation of a pollutant from an enclosure C into a medium Ω_1 (rock) with low diffusion but cracked (boundary Σ). Furthermore below in Ω_1 in another medium Ω_2 (sand) with large diffusion constant ν_2 , the pollutant is also convected (water in sand) at velocity $\mathbf{v}$. The velocity derives from a potential ϕ solution of

$$-\nabla \cdot (\mu \nabla \phi) = 0 \text{ in } \Omega_2 \quad \phi|_{x=0} = 0 \quad \phi|_{x=1} = 1 \tag{23}$$

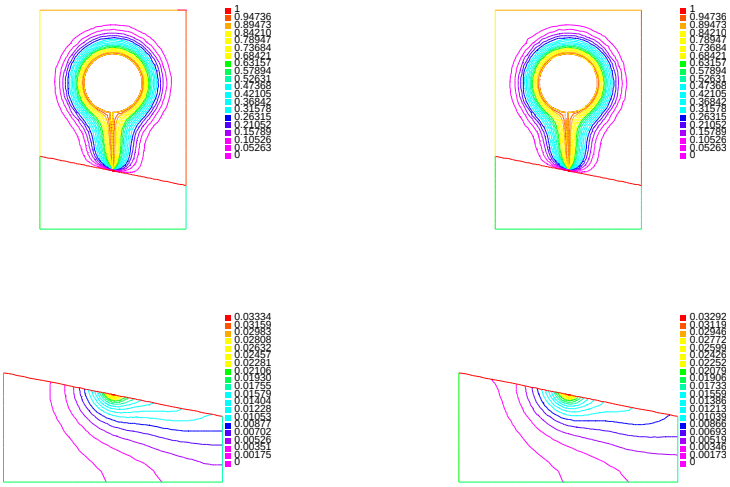


Figure 8: **Top**: *reconstructed solution at the two instants* $t = 0.3$ *and* $t = 0.6$. **Bottom**: *solution in each subdomain at* $t = 0.6$

and $\mathbf{v} = -\mu\nabla\phi$.

Equation (22) is discretized in time by an implicit Euler scheme and in space by the finite element method of degree one on triangles. The convection term is treated by the Galerkin-Characteristic method. Equation (23) is also discretized by the same finite element method.

We have chosen the following Domain Decomposition Method:

$$\frac{1}{\delta t_i}(u_i^{m+1} - u_i^m \circ X^m, \hat{u}_i) + (\nu\nabla u_i^{m+1}, \nabla\hat{u}_i) + b_i(u_i - u_j, \hat{u}_i) = 0 \tag{24}$$

$$\forall \hat{u}_i \in V_i \quad i, j = 1, 2,\ j \neq i \tag{25}$$

where $u^m \circ X^m(x) \approx u^m(x - \mathbf{v}^m(x)\delta t)$, V_i is the finite element space on Ω_i and

$$b_i(u, v) = \int_S (\alpha_i uv + \beta_i \frac{\partial u}{\partial s}\frac{\partial v}{\partial s}) \quad S = \overline{\Omega}_1 \cap \overline{\Omega}_2 \tag{26}$$

The parameters chosen are:

$$\mu = 10,\ \nu_1 = 0.001,\ \nu_2 = 0.05,\ \delta t_1 = 0.005,\ \delta t_2 = 0.02,\ T = 0.15 \tag{27}$$

$$\alpha_1 = 0,\ \beta_1 = 0.01,\ \alpha_2 = 100,\ \beta_2 = 0 \tag{28}$$

The mesh of Ω_1 is 1.5 times finer than the mesh of Ω_2. The method is not unconditionally stable; we have tried several values for the operator b and not all of them work; but the fact that the coefficients of the PDE are constant in each sub-domain and the inherent parallelism are the two major advantages. The results are on figure 8.

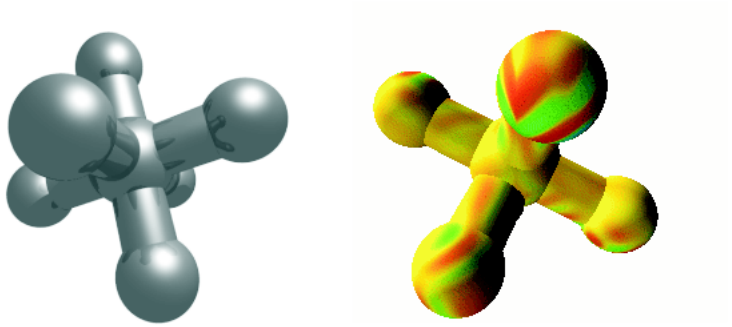


Figure 9: **Left***A scene displayed by* `POV-Ray`*. The objects are never intersected, it is the graphic rendering that takes care of the problem.* **Right** *The trace of the real part of the scattered acoustic field on the surface of the geometry [PHPT00]*

Computation in Virtual Realities

Parallel computing with non-conforming meshes is easy to implement once a fast and robust interpolator is available to compute functions on all the meshes. Such is the case of `freefem+`[BHOP99], a public domain software written by one of the authors. The previous numerical results of this chapter were obtained with it.

Freefem3d

DDM is potentially useful to speed-up computation of virtual scenes created by Constructive Solid Geometry. However all of them are in 3d, so we are currently developing a 3d version of `freefem`. It has a language which is interpreted using `bison`, it reads geometries created with `POV-Ray` (the file `cross.pov` below for instance) and it uses the Fictitious Domain Embedding Method (FDEM). Results are displayed with IBM's Data Explorer (cf http://www.dx.com) or even Pov-Ray (Suzuki's path in http://www.public.usit.net/ rsuzuki/ e/ povray/iso/index.html)

```
vector a = (0,0,0);
vector b = (1,1,1);
vector n = (100,100,100);
structmesh Mesh(n,a,b);
scene S("cross.pov",Mesh);
array mu(Mesh) = 1;
w = 5;
solve(u) {
u * w^2 + div(mu*grad(u)) =0;
dnu(u)-I*w*u=I*w *(Nx-1) on<1,0,0>
        };
plot(u);
```

Operator Overloading in `C++` makes it easy to program the vector case by using `templates` over the scalar case (*generic programming*). Thus the following is possible with any number

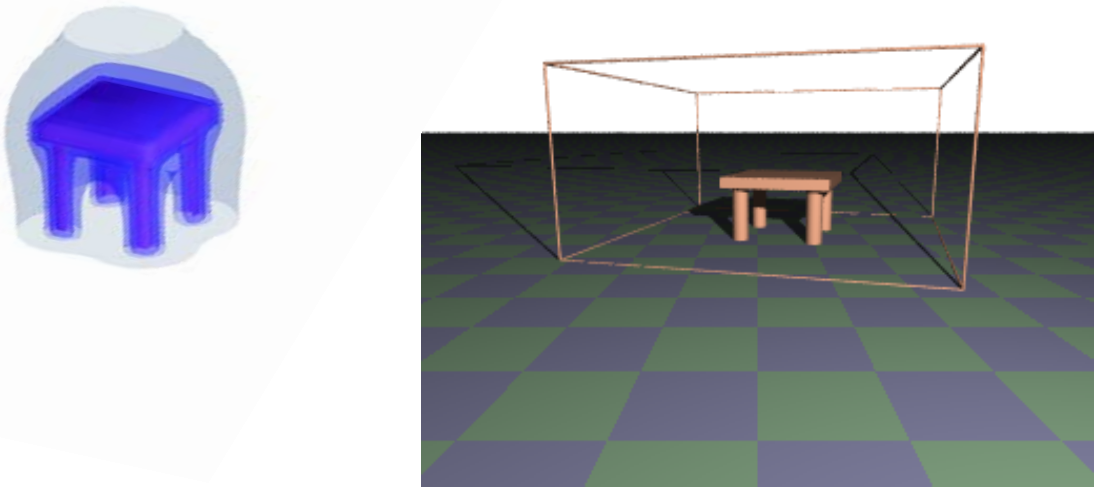


Figure 10: *Displayed are 4 iso-temperature surfaces (0.95, 0.5, 0.25, 0.05) for a transient solution of the heat equation at time 0.1, around a table-shaped object at temperature 1 with Neumann conditions on the boundary of the computational domain, (shown on the right with POV-Ray) and initial temperature zero; the program in* `freefem3d` *language is given above.*

of unknowns (2 here):

```
 solve(u,v){
  pde(u) - laplace(u) = f1;
     on(a) dnu(u) + v = g;
  pde(v) u - laplace(v) = f2;
     on(a) u + 10*v = h };
```

However it will be quite a challenge to find a general preconditioner for iterative solutions of the linear systems.

We conclude with a last example with the heat equation

```
double i=0; double dt=0.1; do{
  solve(u) {  u - div (dt * grad(u)) = u;
              u=1  on <1,0,0>;
              dnu(u) = 0 on Mesh;
            };
  i=i+1;
  dxplot("u.dat",u,Mesh);
}while(i<=5);
```

The results are shown on figure 10

References

[BC94]Gordon Burden and Philippe Coiffe. *Virtual Reality Technology*. Wiley, Chichester, 1994.

[BHOP99]Dominique Bernardi, Frederic Hecht, Kohji Otsuka, and Olivier Pironneau. freefem+, a finite element software to handle several meshes. *http://www.freefem.org*, 1999.

[BLP01]Franco Brezzi, Jacques-Louis Lions, and Olivier Pironneau. Analysis of a chimera method. *C.R.A.S. I*, 332(7):655–664, 2001.

[BW86]Petter E. Bjørstad and Olof B. Widlund. Iterative methods for the solution of elliptic problems on regions partitioned into substructures. *SIAM J. Numer. Anal.*, 23(6):1093–1120, 1986.

[GLP99]R. Glowinski, J.L. Lions, and O. Pironneau. Decomposition of energy spaces and applications. C.R.A.S., Paris, 1999.

[HLP99]F. Hecht, J.L. Lions, and O. Pironneau. Domain decomposition algorithm for computed aided design. In A. Sequeira et al, editor, *Applied Nonlinear Analysis*, pages 185–198. Kluwer Academic-Plenum Publishers, New York, 1999.

[HW96]John Hartman and John Wernecke. *The VRML 2.0 Handbook*. Addison Weisley, New-York, 1996.

[Lio78]Pierre Louis Lions. Interprétation stochastique de la méthode alternée de Schwarz. *C. R. Acad. Sci. Paris*, 268:325–328, 1978.

[LP98a]J.L. Lions and O. Pironneau. Algorithmes parallèles pour la solution des problémes aux limites. *C.R.A.S., Paris*, 327:947–952, 1998.

[LP98b]J.L. Lions and O. Pironneau. Sur le contrôle parallèle des systèmes distribués. *C.R.A.S., Paris*, 327:993–998, 1998.

[LP99a]J.L. Lions and O. Pironneau. Contrôle virtuel, répliques et décomposition d'opérateurs. *C.R.A.S.*, 1999.

[LP99b]J.L. Lions and O. Pironneau. Domain decomposition method for CAD. *C.R.A.S., Paris*, 328:73–80, 1999.

[LP99c]J.L. Lions and O. Pironneau. Domain decomposition methods for cad. *C.R. Acad. Sci. Paris*, 328:73–80, 1999.

[PHPT00]Stephan Del Pino, Errki Heikkola, Olivier Pironneau, and Jari Toivanen. A finite element method for virtual reality data. *C.R.A.S. I*, 330(12):1107–1115, 2000.

[Pir87]Olivier Pironneau. *Finite Element Methods for Fluids*. Wiley, Chichester, 1987.

[SB87]J. Steger and J. Benek. On the use of composite grid schemes in computational aerodynamics. *Comp. Meth. Appl. Mech. Eng.*, 64:301–320, 1987.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

10 Cartesian and Curvilinear Grid Methods for Multi-domain, Moving Boundary Problems

W. Shyy[1] M. Francois[2] H.S. Udaykumar[3]

Introduction

A variety of physical phenomena involve the coupling of evolution of multiple materials with boundaries that move, deform or evolve in time. Examples include the deformation of drops, bubbles, liquid free surfaces, phase boundaries in solidification and vaporization, fluid-structure interaction problems at the large scale such as in aeroelasticity and in the small scale such in biomechanics, and a whole host of other interesting phenomena. These problems are challenging due to the complexity associated with the often severely deformed boundaries, multiple time and length scales, and the nonlinearity resulting from the coupling of the interface dynamics with the dynamics of the material. Ideally one would like to track the moving boundary as a sharp front (allowing discontinuities in quantities such as stress and energy across the interface) without smearing the information at the front. Also, one would like to solve the field equations within each region separated by the interfaces with satisfactory accuracy. If the interfaces become multiply-connected, it is desirable to follow the evolution of the interfaces through such topological changes. Numerous techniques exist for tracking arbitrarily shaped moving interfaces, each with its own strengths and weaknesses [Cra84, FR89, SURS96]. These techniques may be classified under two main categories: (a) surface tracking or predominantly Lagrangian methods [FZP$^+$93, SS95, SS96] and (b) volume tracking or Eulerian methods [HN81, AP91]. The main features of the two types are presented in Figure 1. We offer the following comments to contrast the relative characteristics among different approaches.

a. Interface Definition

The Lagrangian methods maintain the interface as a discontinuity and explicitly track its evolution. If detailed information regarding the interface location is desired, Eulerian methods may need elaborate procedures to deduce the interface location based on the volume fraction information, and uncertainty corresponding to one grid cell is unavoidable [AP91, HN81, SZ99]. In the Lagrangian case, the interface can be tracked as a (n-1)-dimensional entity for a n-dimensional space [DS85, GGL$^+$88, WM86]. No modeling is necessary to define the interface or its effect on the flow field. In the case of Eulerian schemes, modeling or solution of additional equations is required to obtain information regarding phase fractions or other functions yielding information in the two-phase regions.

[1]Department of Aerospace Engineering, Mechanics and Engineering Science, University of Florida, wss@aero.ufl.edu
[2]Department of Aerospace Engineering, Mechanics and Engineering Science, University of Florida, francoim@aero.ufl.edu
[3]Department of Mechanical Engineering, University of Iowa.

b. Interfacial Boundary Conditions

In the Lagrangian methods, boundary conditions can be applied at the exact location of the interface since the interface position is explicitly known at each instant. In the Eulerian methods, the boundary conditions are manipulated to appear in the governing transport equations [BKZ92]. This leads to the smearing of boundary information.

c. Discretization of the Domain

In the Lagrangian methods, the grid adapts to the interface and hence grid rearrangement and motion terms have to be incorporated. When the interface begins to distort, the grid needs to be regenerated each time. The resulting grid on which the field variables are computed may be skewed and unevenly distributed, thus influencing the accuracy of the field solver. The Eulerian methods have an advantage in this regard since the computations are performed on a fixed grid, hence obviating the need for grid rearrangement. However, when the interface is arbitrarily shaped, improved resolution in desired regions is difficult to obtain, unless complicated local refinements are adopted. In the Lagrangian method a set of governing equations needs to be solved for each different material and region, whereas in an Eulerian method only a single set of equations with appropriate source terms is solved for the entire domain.

d. Movement and Deformation of the Interface

Lagrangian methods have so far experienced difficulty in handling topological changes, mainly due to the breakdown of the structured grid arrangement and the need for redistribution of field information in the vicinity of the interface for unstructured grid methods [WM86]. On the other hand, in Eulerian methods mergers and fragmentations are taken care of automatically, merely by updating the values of the phase fraction. However, the detailed physical features involved during such events may not be fully resolved due to the smearing of information as mentioned above. The choice of moving boundary method from the general categories above depends to a large extent on its appropriateness of the physical problem chosen. In the following, we highlight recent efforts in developing computational techniques for treating moving boundary problems. Both moving and fixed grid methods will be considered. To aid the discussion, for the fixed grid method, we use the impact dynamics, and for the moving grid method, we use the soft contact lens dynamics to highlight the solution characteristics.

A Fixed-grid, Sharp-interface Method for Multiple Moving Boundaries: Impact Dynamics

The dynamics of impact between materials is characterized by large deformation and short time scales. Wave propagation in the impacting media is highly nonlinear, and involves localized phenomena such as shear bands, crack propagation, and wave refraction [Mey94]. These problems are typically challenging to solve because, in contrast to conventional structural dynamics problems, the deviatoric and pressure terms in the stress tensors are both important and need to be modeled separately. In contrast to conventional fluid dynamics problems, the stress and strain fields are related through nonlinear elasto-plastic yield surfaces, the models for which must be included in the governing equations. Furthermore,

the interface between materials experiences not only fast motion, but also large variations in shape. In this section we summarize a numerical solution technique progressively developed in [SURS96, USR96, UKSTST97, UMS99, UTS+00, YMUS99] for the simulation of high-speed multi-material impact. Of particular interest is the interaction of solid impactors with targets. This problem is important in applications such as munitions-target interactions, automobile collision assessment, geological impact dynamics, and shock processing [Mey94]. Such interactions present the following challenges to numerical simulation techniques:

1. High velocities of impact leading to large deformations of the impactor as well as targets.

2. Nonlinear wave-propagation and the development of shocks in the systems.

3. Modeling of the constitutive properties of materials under intense impact conditions and accurate numerical calculation of the elasto-plastic behaviour described by the models.

4. Phenomena at multiple interfaces (such as impactor-target, target-ambient and impactor-ambient), i.e. both free surface and surface-surface dynamics.

The method adopted falls under the class of combined Eulerian-Lagrangian method. It operates on a fixed Cartesian mesh (the Eulerian part) while the interfaces move through the mesh (the Lagrangian part). The method treats the interfaces as discontinuities without smearing on the mesh, therefore it is a sharp interface method. The advantage of the fixed grid approach is obviously that grid topology remains simple while large distortions of the interface take place. This allows an extension of highly accurate shock-capturing methods (Essentially Non-Oscillatory or ENO [HEOC97, SO88, SO89] in the present case) developed for scalar conservation laws in fixed grid settings to solve moving boundary problems with arbitrarily distorted interfaces. A Cartesian grid ENO formulation suffers little change when applied to the present problem. We now proceed to describe the method in detail.

Interface Tracking Algorithm

The interface is described by interfacial markers defined by the coordinates $X(s)$. The spacing between the markers is maintained at some fraction of the grid spacing h, $0.5h < ds < 1.5h$. The convention adopted is that as one traverses the interface along the arc-length, the material enclosed by the interface lies to the right. This is illustrated in Figure 2. The functions $x(s) = a_x s^2 + b_x s + c_x$ and $y(s) = a_y s^2 + b_y s + c_y$ are generated. The coefficients $a_{x/y}$, $b_{x/y}$ and $c_{x/y}$ at any interfacial point i are obtained by fitting polynomials through the coordinates (x_{i-1}, y_{i-1}), (x_i, y_i) and (x_{i+1}, y_{i+1}). The coefficients $a_{x/y}$, $b_{x/y}$ and $c_{x/y}$ are stored for each marker point. Once the interface has been defined, the information on its relationship with the grid has to be established. There may be several interfaces (henceforth called objects) immersed in the domain. Each of the objects may enclose material with different transport properties. Therefore it is necessary to identify which phase each computational point (i.e. cell center point) lies in. An illustration is shown in Figure 3. The end result of the procedures is the following pieces of information which are required to set up the discretization scheme for the present method: (i) The interfacial cell in which each interface marker lies. (ii) The interfacial marker, which is closest in distance to a computational point. (iii) The material in which each computational point in the mesh lies. (iv) Several geometric details such as the

shape of the resulting cut-cell, the locations where the interface cuts the cell faces and where it intersects the cell center lines (the dotted lines shown in Figure 3). These details of a cell are used in constructing the stencil for each interfacial cell. (v) A list of all interfacial cells. These pieces of information regarding the interface and its relationship to the underlying grid are computed only in a lower-dimensional set of interface cells. In summary, the computational formulation tracks moving boundaries on a fixed underlying grid while striving to achieve the following objectives:

1. The interface is tracked as a discontinuity and boundary conditions of the Dirichlet/Neumann type are applied on the tracked fronts.

2. The discretization to include the embedded boundaries involves simple measures in the vicinity of the interface. Such points are few compared to the overall grid size.

3. Based on truncation error analysis the discretization can be performed so that global second-order accuracy in the field variable can be maintained.

4. The problem of stiffness of the interface evolution in curvature-driven dynamics [HLS94] is surmounted by using an implicit formulation to couple the interface evolution with the field equation.

5. The issue of change of material of a grid point when the boundary crosses over it is dealt with by a simple analogy with purely Lagrangian methods.

This involves redefinition of the stencils in the points adjoining the interface to account for the grid points that have changed phase. The various components of the solution algorithm can easily be extended to 3D. It is demonstrated by [UMS99, YMUS99] that the field calculation is second-order accurate while the position of the phase front is calculated to first-order accuracy. Furthermore, the accuracy estimates hold for the cases where there are property jumps across the interface.

Results and Discussion

2D computations were performed in a square domain of size 1mx1m as illustrated in Figure 4. As shown there the objects were placed some distance apart on the mesh and impact was initiated by prescribing a velocity to one or both interfaces. Initially there is a region of void between the two interfaces. This void disappears at the material- material interface. In Figure 5, we show the impact of a cylinder with a plane surface. Both surfaces are copper and the material properties in the model correspond to that metal. In the figure, we show on the left the contours of velocity magnitude in the impactor and the target along with the velocity vectors in the flow domain. On the right we show contours of equivalent stress. Also shown in each of the figures is the shape of the boundaries of the two materials. As can be seen in these figures there is an abrupt transition in the corners from a material-material interface to a material-void interface for each material. Appropriate governing laws and boundary conditions are discussed in [UTS$^+$00]. Zero-gradient conditions are applied at the sides of the domain assuming that the target has infinite extent in all except the +y direction. Figures 5 (a), (b) and (c) correspond to time instants $2.5\mu s$, $50\mu s$ and $100\mu s$ after impact respectively. The progression of the elasto-plastic waves and the formation of large gradients in the velocity as well stress fields is evident from the figure. At the rim of the impactor, the interfaces are

constantly in collision since the material-void interfaces are being pushed against each other to form material-material interfaces. Thefore the rim of the impact region registers large stress and correspondingly, strain values. Stress waves are propagated into the materials from this point. In Figure 5(c) it can be seen that the velocity field is such as to continuously push the impactor into the target leading to the production of an upswell in the target material around the rim. This is also indicated clearly by the velocity vectors shown. Regions of compression and tension are seen from the contours of stress. The current method has the following capabilities:

1. The interface can be tracked through large distortions.

2. Accurate shock-capturing schemes can be implemented for Cartesian grids and extended in a straightforward manner to incorporate the presence of the moving interfaces.

3. Boundary conditions are developed for the 1D uniaxial strain case and 2D plane strain case and these are applied at the exact locations of the boundaries.

4. Different regions of the boundaries can have different boundary conditions, i.e. the material-material and material-void boundary conditions. These are applied at the interface points identified to lie in regions where the interfaces are in contact and where the interface is exposed to void respectively. These boundary conditions are physically dictated or numerical boundary conditions. The suitability of the set of boundary conditions is determined based on numerical experimentation. The singularity resulting from an abrupt transition from a material-material to material- void boundary condition at the interfaces is handled well.

A Moving-grid Method for Fluid-structure Interaction: Soft Contact Lens

A soft contact lens is spherical in projected shape and has a diameter around 12 to 14 mm. The optical power of the lens determines the lens posterior central radius (base curve radius) and its thickness. Commonly used base curve radii range from 7.5 to 9.0 mm. Soft contact lenses are made of hydrophilic polymers that have an elastic modulus varying with water content. For example the "1-Day Acuvue" lens by Vistakon contains 58% of water and has an elastic modulus of 0.36 Mpa [WSB98]. A typical lens weighs about 10 mg. When placed on the eye, a contact lens is separated from the eye surface by a thin tear film. The thickness of the tear film beneath the lens is around $10\mu m$. So, the aspect ratio between the lens diameter and tear film height is very large, of the order of 1000. Figure 6(a) illustrates the lens-eye profile. An eye-blink creates a force on the contact lens that causes the contact lens to move and deform. Since the lens material makes the soft contact lens very flexible, the lens can exhibit complex shape deformation characteristics.

Francois et al. [FSU99] have presented a computational capability to simultaneously model the dynamics of a soft contact lens and fluid dynamics of the tear film flow. In the present model, the deformable contact lens is considered to be an elastic membrane. A schematic of the computational model is presented in Figure 6(b). Specifically, the main features of their model can be summarized as follows: (1) the tear film is considered to be a single layer, Newtonian fluid governed by the Navier-Stokes equations; (2) the soft contact

lens is modeled as an elastic membrane whose tension is regulated by the membrane thickness; (3) the lens is fixed at the edge; (4) the ambient pressure variation is responsible for the lens movement and deformation; (5) the lens structural dynamics and the tear film are modeled as a coupled system so that both the lens and the tear film characteristics, and their interaction, can be investigated simultaneously; (6) the lens thickness is of variable profile based on typical commercial design.

Governing Equations of Tear Fluid and Contact Lens

The governing equation of the soft contact lens considered is the equilibrium equation of an elastic massless membrane in 2D [SS95, SURS96]. Figure 7 illustrates an elastic membrane restrained at its both extremities. Here only the equilibrium lens equation in the normal direction is considered:

$$-\frac{\Delta P}{\gamma} = \frac{P - P_a}{\gamma} = \left[\frac{d^2y}{dx^2} \left(1 + \left(\frac{dy}{dx} \right)^2 \right)^{-\frac{3}{2}} \right] \tag{1}$$

where P_a is the outside or applied pressure, P is the pressure in the tear film beneath the lens, γ is the lens tension, which is taken to be proportional to the product of the lens elastic modulus and the lens thickness and (x, y) are the space coordinates. The fluid flow computation is based on a well-established pressure-correction type finite volume solver of the Navier-Stokes equations, in body-fitted curvilinear coordinates, as detailed in [Shy94, SURS96, SS97]. Since the initial structure configuration and associated body-fitted grid do not correspond to an equilibrium configuration a moving grid procedure is employed [SURS96, SS95, SS96] wherein the grid is continuously updated during the course of computation, in response to the shape change of the lens. Three key information items are required to facilitate the moving grid technique, namely,

1. Kinematics conditions apply at the interface (moving boundaries).
2. The geometric conservation law is invoked [SURS96] to estimate the Jacobian of term to enforce volume conservation.
3. The contravariant velocity components and Cartesian velocity components at the boundary are computed to enforce mass conservation.

Results and Discussion

Results of the computations for three configurations with tear film aspect ratios (ratio of the horizontal projected length between the center and the pinned end point to the tear film height at the pinned location) of 10, 100, and 1000 are presented. It should be noted that, while under practical wearing conditions the aspect ratio of the tear film is around 1000, we have treated this aspect as a parameter to gain a more comprehensive understanding of the physics. The computational domain and boundary conditions used are presented in Figure 6(b). The overall geometry including lens, tear film, and cornea and the variable lens thickness profile is illustrated in Figure 8. An externally imposed time-dependent pressure, modeling eye blinking process, is represented in Figure 9. Figure 10 shows the maximum lens deflections

normalized by the initial tear film height at the lens center, for the three cases with variable thickness, in response to the imposed pressure variation. Figure 11 presents the maximum pressure difference inside the tear film normalized by the applied pressure oscillation versus time. From Figures 10 and 11, it is clear that the pressure variations and the lens deformation are correlated with each other. Accordingly, as shown in Figure 12, the tear flow rate is influenced by such a correlation.

Francois et al. [FSU99] have also discussed the fluid physics of soft contact lens. First, the tear fluid velocity, responding to the lens deformation, increases from the central region toward the end of the lens. Second, the pressure gradient does not develop only along the direction of the lens, as suggested by a typical thin film approximation. The reason is that the lens movement is not slow and the tear film is not a simple parallel viscous flow. A straightforward application of the thin film theory without due consideration of the lens movement can introduce large error in the analysis. Under the present condition, as illustrated in Figure 13, the Reynolds number increases as the aspect ratio increases, from negligibly small to about 100. These observations indicate that common practices in the literature, with either a straightforward application of the thin film theory, without an explicit consideration of the lens movement, or a direct account of the structure dynamics with an assumed pressure field are unsatisfactory.

Concluding Remarks

We have described the development of numerical techniques based on both fixed and moving grid to treat sharp interface for the simulation of moving boundary problems. Examples arising from fast transient multi-material impact dynamics and soft contact lens are used to illustrate the main features of each method. For the fixed grid method, computations of the deformation process are carried to large distortions while the interfaces travel through the mesh in a stable and robust manner. Such a technique has also been successfully applied to treat problems arising from crystal growth [UMS99]. On the other hand, if the detail of the interface characteristics can be smeared out, then a simpler treatment involving the immersed boundary treatment [Pes77, UT92, UKSTST97, KUSTST98] can be highly effective. This approach has tackled a variety of problems, especially for those related to multiphase and cellular dynamics. For the moving grid method, the implications of the lens tension variations on the lens response and the nonlinear interaction between the fluid flow and the soft contact lens are demonstrated. When the interface does not exhibit substantial deformation, the moving grid method can be highly effective. It can also be robust in terms of the size of the time steps. This approach has been successfully applied to handle fluid-structure interaction problems with high Reynolds numbers, including the effect of turbulence and laminar-to-turbulent transition [SS95, SS96, SJS97, HFS$^+$00]. No method is universally superior for treating moving boundary problems. Depending on the nature of the problem and the goal of the computation, an intelligent selection of an appropriate technique can help successfully address the physical and numerical challenges.

Acknowledgements

The research reported has been supported by AFOSR, Eglin AFB, and Johnson & Johnson.

References

[AP91]Nasser Ashgriz and J.Y. Poo. Flair - fux line-segment model for advection and interface reconstruction. *Journal of Computational Physics*, 93(2):449–468, April 1991.

[BKZ92]J.U. Brackbill, Douglas B. Kothe, and C. Zemach. A continuum method for modeling surface tension. *Journal of Computational Physics*, 100(2):335–354, June 1992.

[Cra84]John Crank. *Free and Moving Boundary Problems*. Oxford University Press, New York, 1984.

[DS85]A.J. DeGregoria and L.W. Schwartz. Finger breakup in hele-shaw cells. *Physics of Fluids*, 28:2313–2314, 1985.

[FR89]J.M. Floryan and Henning Rasmussen. Numerical methods for viscous flows with moving boundaries. *Applied Mechanics Reviews*, 42(12):323–340, 1989.

[FSU99]Marianne Francois, Wei Shyy, and HS Udaykumar. Computational mechanics of soft contact lenses. Bulletin of the American Physical Society, 52nd Annual Meeting of the Division of Fluid Dynamics, November 1999.

[FZP+93]J. Fukai, Z. Zhao, Dimos Poulikakos, Constantine M. Megaridis, and O. Miyatake. Modeling of the deformation of a liquid droplet impinging upon a flat surface. *Physics of Fluids*, 5:2588–2599, 1993.

[GGL+88]James Glimm, J Grove, B Lindquist, OA McBryan, and G Tryggvason. The bifurcation of tracked scalar waves. *SIAM J. Sci. Stat. Comput.*, 9:61–79, 1988.

[HEOC97]Ami Harten, Bjorn Engquist, Stanley Osher, and Sukumar R. Chakravarthy. Uniformly high-order accurate essentially non-oscillatory schemes, iii. *Journal of Computational Physics*, 131:3–47, 1997.

[HFS+00]Xiong He, Carlos Fuentes, Wei Shyy, Yongsheng Lian, and Bruce Carroll. Computation of transitional flows around an airfoil with a movable flap. AIAA Fluids 2000 and Exhibit, Paper no. AIAA-2000-2240, June 2000.

[HLS94]Thomas Y. Hou, John S. Lowengrub, and Michael J. Shelley. Removing the stiffness from interfacial flows with surface tension. *Journal of Computational Physics*, 114:312–338, 1994.

[HN81]CW Hirt and BD Nichols. Volume of fluid (vof) method for the dynamics of free boundaries. *Journal of Computational Physics*, 39:201–225, 1981.

[KUSTST98]Heng-Chuan Kan, HS Udaykumar, Wei Shyy, and Roger Tran-Son-Tay. Hydrodynamics of a compound drop with application to leukocyte modeling. *Physics of Fluids*, 10(4):760–774, April 1998.

[Mey94]Marc A. Meyers. *Dynamics Behavior of Materials*. John Wiley & Sons Inc., New York, 1994.

[Pes77]Charles S Peskin. Numerical analysis of blood flow in the heart. *Journal of Computational Physics*, 25:220–252, 1977.

[Shy94]Wei Shyy. *Computational Modeling for Fluid Flow and Interfacial Transport*. Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1994.

[SJS97]Wei Shyy, David A. Jenkins, and Richard W. Smith. Study of adaptive shape airfoils at low reynolds number in oscillatory flows. *AIAA Journal*, 35:1545–1548, 1997.

[SO88]Chi-Wang Shu and Stanley Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77:439–471, 1988.

[SO89]CW Shu and S Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes ii. *Journal of Computational Physics*, 83:32–78, 1989.

[SS95]Richard W. Smith and Wei Shyy. Computation of unsteady laminar flow over a flexible two-dimensional membrane wing. *Physics of Fluids*, 7:2175–2184, 1995.

[SS96]Richard W. Smith and Wei Shyy. Computation of aerodynamic coefficients for a flexible membrane airfoil in turbulent flow: A comparison with classical theory. *Physics of Fluids*, 8:3346–3353, 1996.

[SS97]Richard W. Smith and Wei Shyy. Incremental potential flow based membrane wing element. *AIAA Journal*, 35(5):782–788, May 1997.

[SURS96]Wei Shyy, HS Udaykumar, MM Rao, and Richard W. Smith. *Computational Fluid Dynamics with Moving Boundaries*. Taylor and Francis, 1996.

[SZ99]Ruben Scardovelli and Stephane Zaleski. Direct numerical simulation of free-surface and interfacial flow. *Annual Review of Fluid Mechanics*, 31:567–603, 1999.

[UKSTST97]HS Udaykumar, Heng-Chuan Kan, Wei Shyy, and Roger Tran-Son-Tay. Mutiphase dynamics in arbitrary geometries on fixed cartesian grids. *Journal of Computational Physics*, 137:366–405, 1997.

[UMS99]HS Udaykumar, Rajat Mittal, and Wei Shyy. Computation of solid-liquid phase fronts in the sharp interface limit on fixed grids. *Journal of Computational Physics*, 153:535–574, 1999.

[USR96]HS Udaykumar, Wei Shyy, and MM Rao. Elafint - a mixed eulerian-lagrangian method for fluid flows with complex and moving boundaries. *Int. J. Numer. Meths. Fluids*, 22:691–704, 1996.

[UT92]Salih O. Unverdi and Gretar Tryggvason. A front tracking method for viscous, incompressible, multi-fluid flows. *Journal of Computational Physics*, 100:25–37, 1992.

[UTS$^+$00]HS Udaykumar, L Tran, Wei Shyy, K Vanden, and DM Belk. A combined immersed interface and eno shock capturing method for impact dynamics. AIAA Fluids 2000 and Exhibit, Paper no. AIAA-2000-2664, June 2000.

[WM86]HP Wang and RT McLay. Automatic remeshing scheme for modeling hot forming process. *J. Fluids Engrg.*, 108:465–469, 1986.

[WSB98]G Wilson, JD Schwallie, and RE Bauman. Comparison by contact lens cytology and clinical tests of three contact lens types. *Optometry and Vision Science*, 75(5):323–329, 1998.

[YMUS99]Tao Ye, Rajat Mittal, HS Udaykumar, and Wei Shyy. A cartesian grid method for simulation of viscous incompressible flow with complex immersed boundaries. *Journal of Computational Physics*, 156:209–240, 1999.

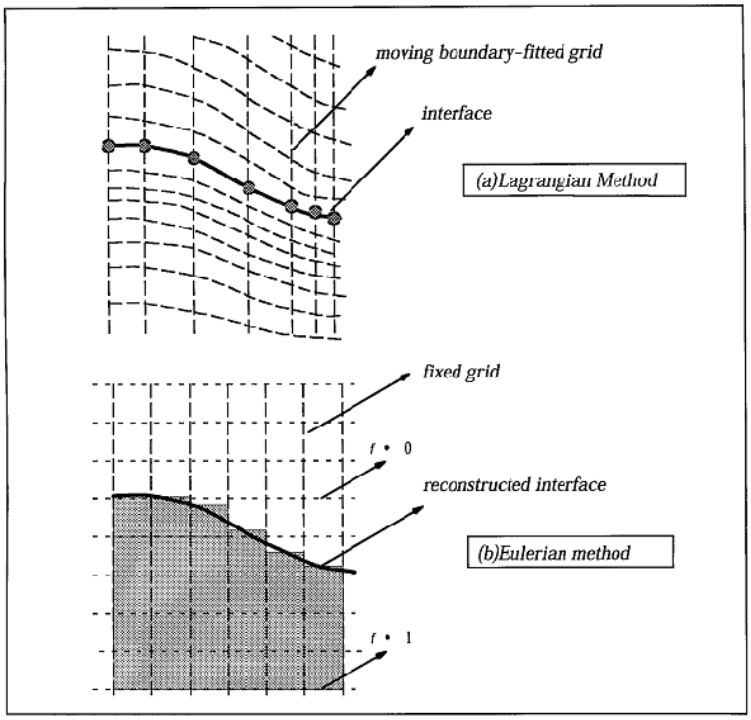


Figure 1: Comparison of Lagrangian and Eulerian methods for interface tracking. (a) Purely Lagrangian method with a moving, boundary conforming grid. (b) Fixed grid Eulerian method with a phase fraction definition of the interface.

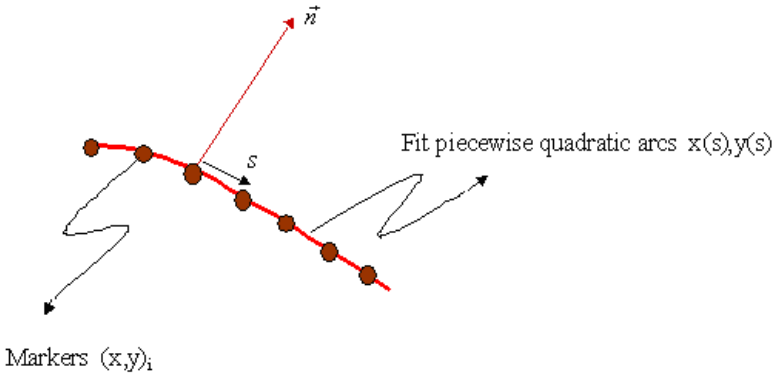


Figure 2: Illustration of interface properties. The normal to the interface and arclength coordinate are shown.

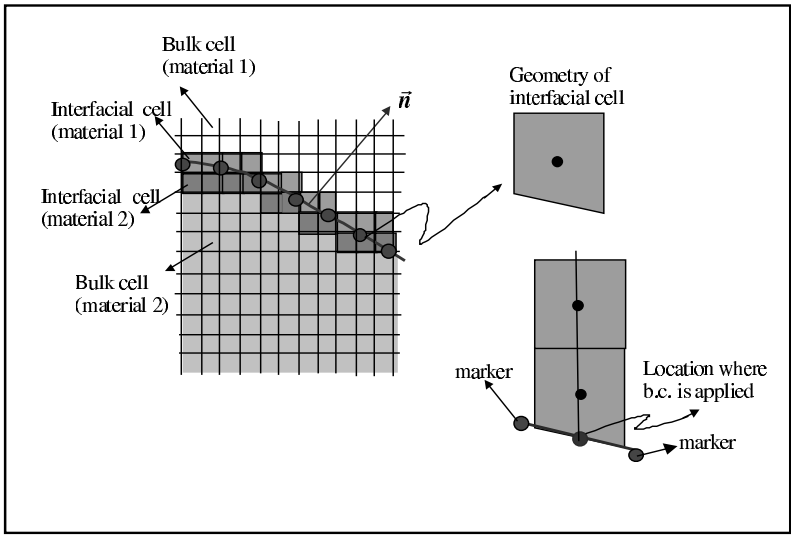


Figure 3: Interfacial cell and bulk cell classification on a grid. With interface passing through it. Also shown are interfacial cell properties.

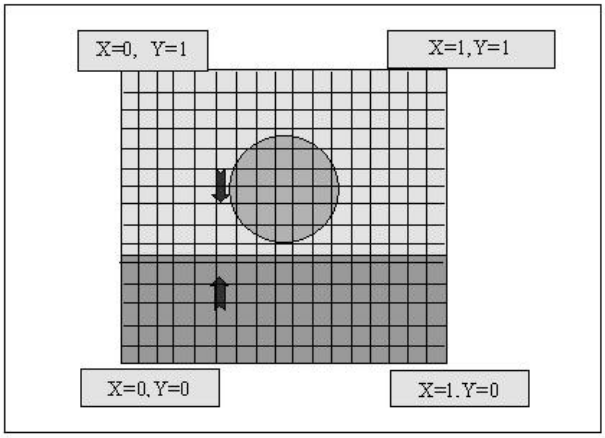


Figure 4: Impact of two objects in the 2D case. A plane strain problem is solved.

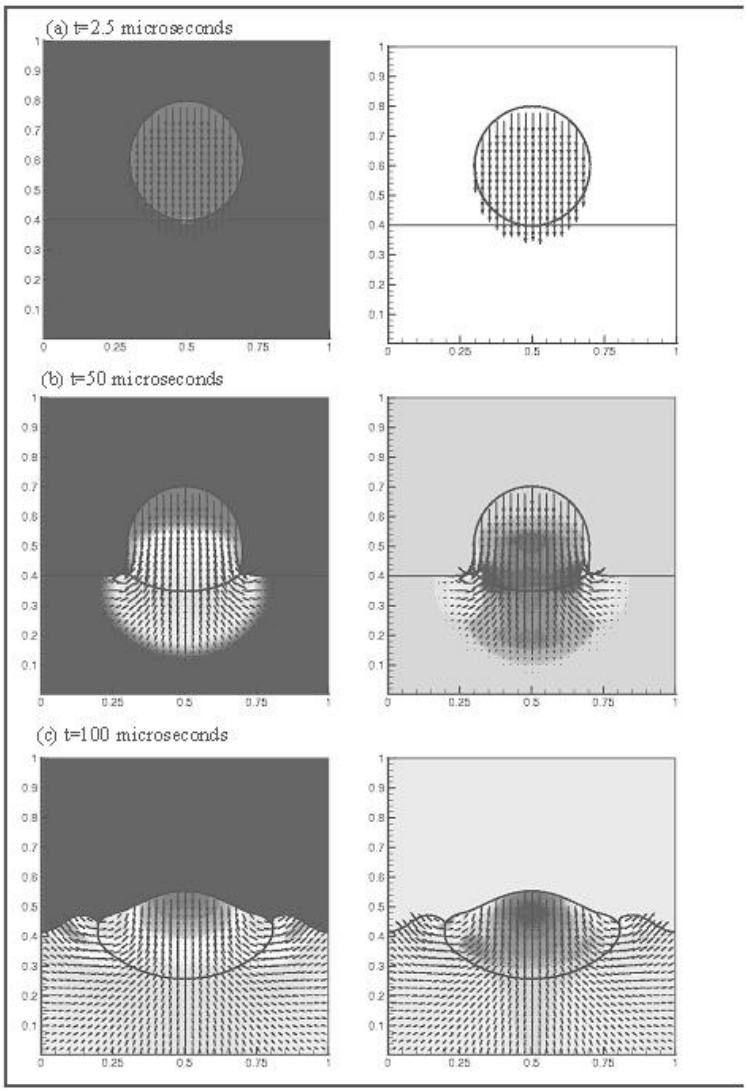
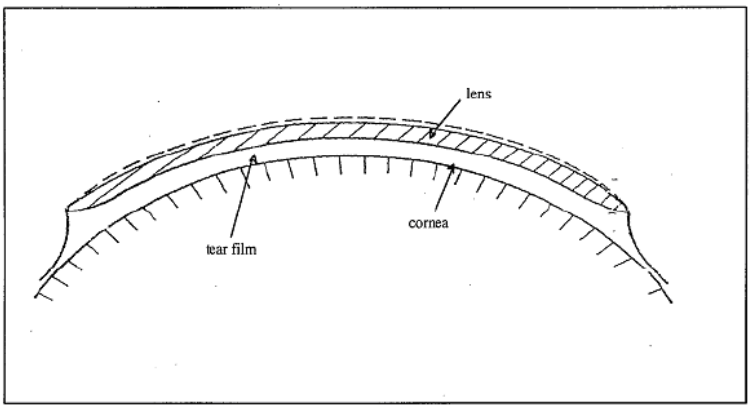
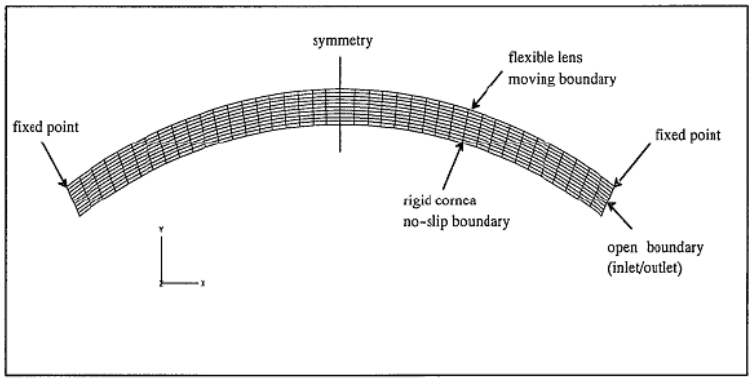


Figure 5: Impact of a cylinder with a planar surface. The cylinder impacts the target with a velocity of $2000 m/s$ directed downward. The figures on left show velocity contours and vectors along with the interface shapes. The time after impact are indicated alongside the figures. The figures on the right show stress contours.



(a) Experimental observation (adopted after Hayashi 1977)



(b) Grid and boundary conditions of the computational model

Figure 6: Schematic of contact lens and computational model.

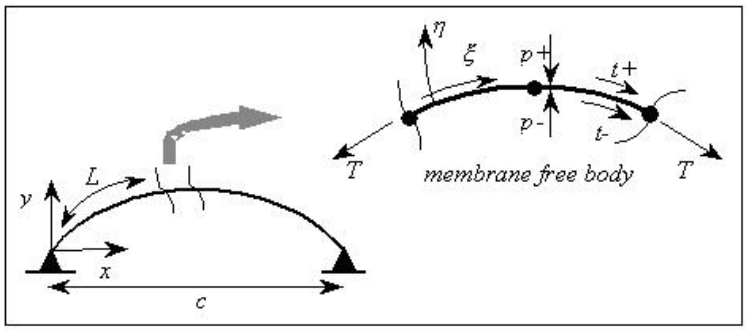
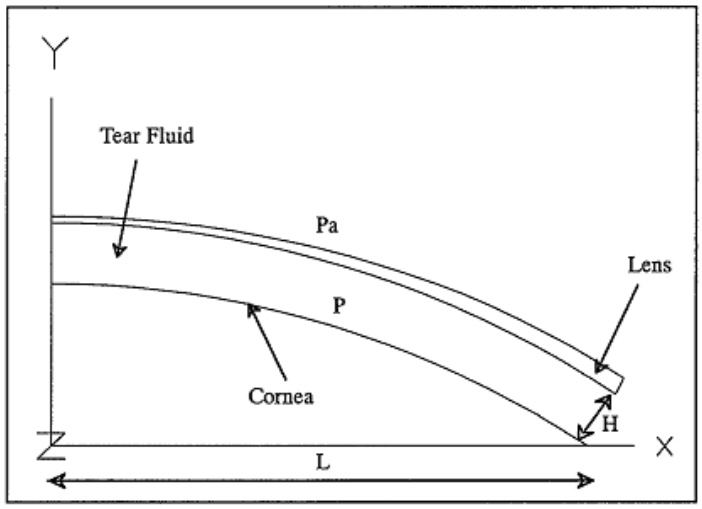
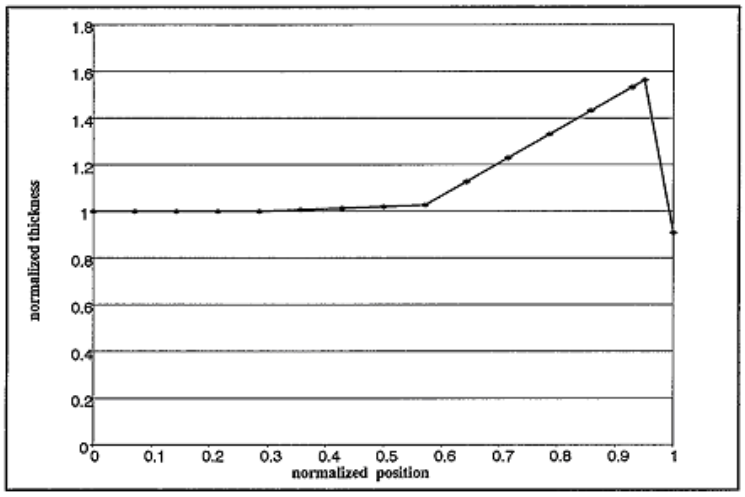


Figure 7: End constrained elastic structure.



(a) Variable thickness lens (γ variable)



(b) Profile of the variable thickness lens
(0 and 1 indicate, respectively, the center and the end of the lens)

Figure 8: Schematic of the contact lens model and thickness profile.

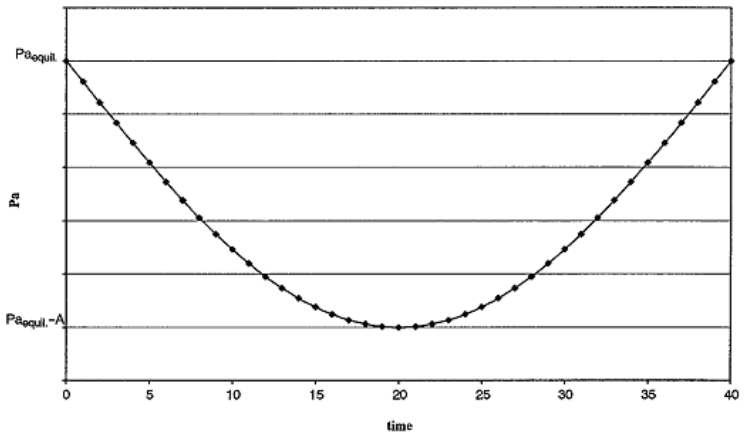


Figure 9: Time History of the applied external pressure $P_a = P_{a_{equil}} - Asin(\frac{\pi}{40}t)$.

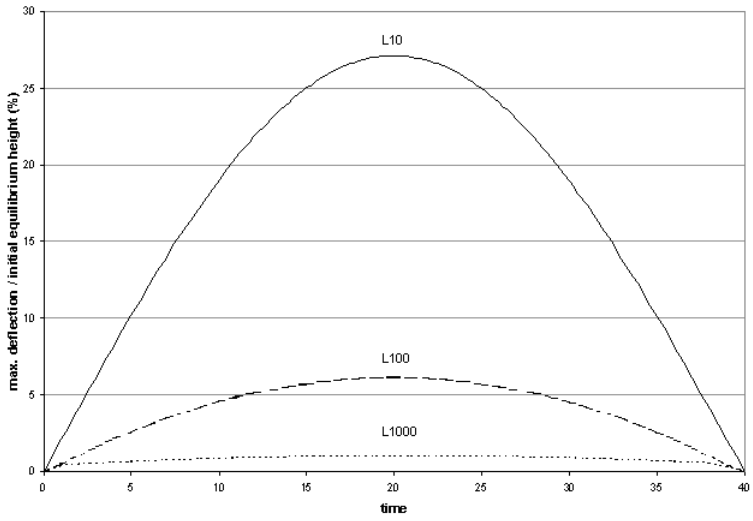


Figure 10: Maximum deflection versus time for three different aspect ratios 10, 100, 1000 with variable thickness.

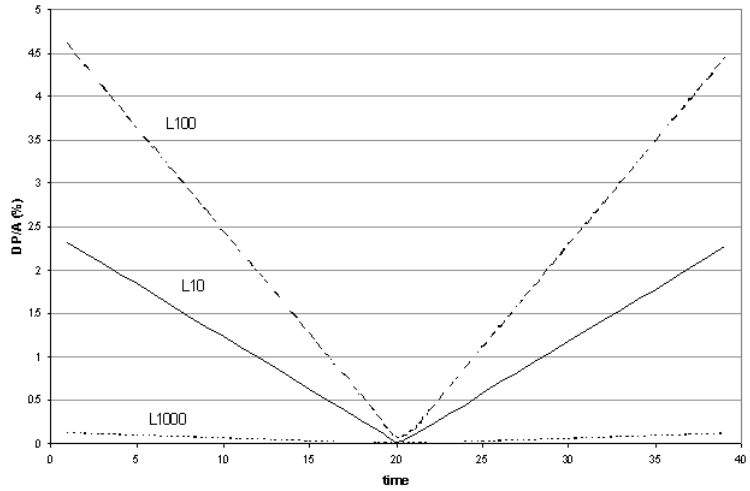


Figure 11: Maximum pressure difference inside the tear film.

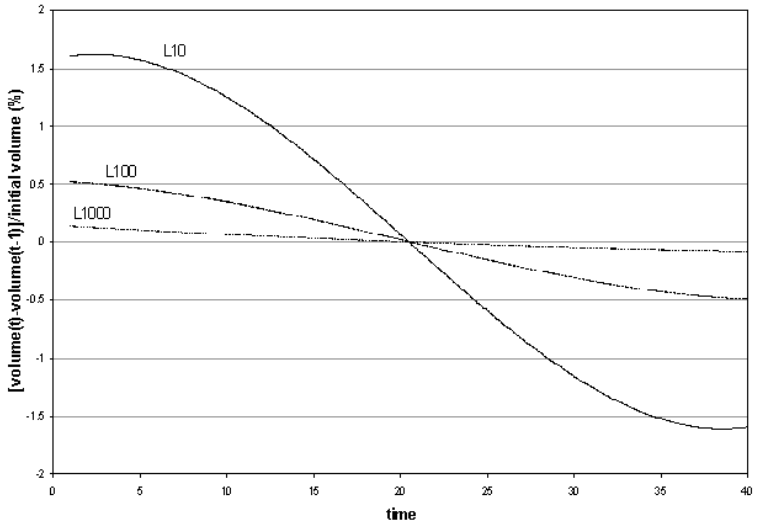


Figure 12: Variation of tear fluid volume going in/out of domain with time.

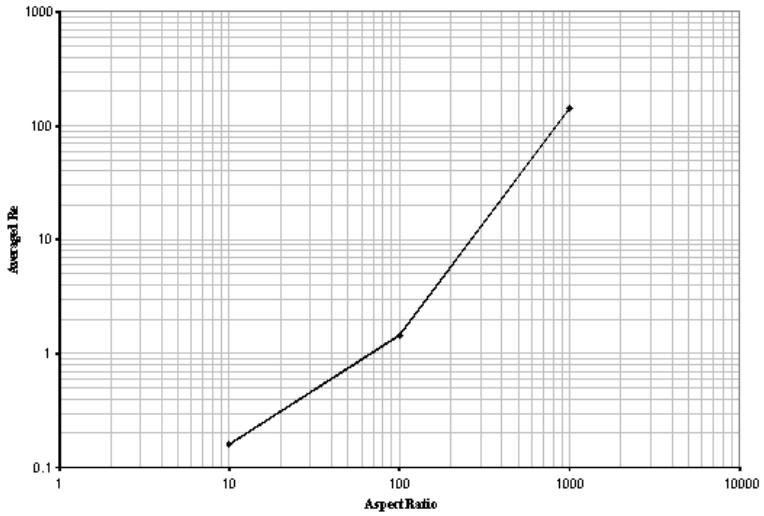


Figure 13: Averaged Reynolds number versus aspect ratio.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

11 Rate of Convergence for Parallel Subspace Correction Methods for nonlinear variational inequalities

X.-C. Tai[1], B. Heimsund[2] and J. Xu[3]

Introduction

Given a reflexive Banach space V and a convex functional $F : V \mapsto R$, we shall consider the following nonlinear optimization problem

$$\min_{v \in K} F(v), \quad K \subset V \ . \tag{1}$$

The nonempty convex subset K is assumed to be closed in the strong topology of V. We are interested in the case that the space V can be decomposed into a sum of subspaces V_i, i.e.

$$V = V_1 + V_2 + \cdots + V_m = \sum_{i=1}^{m} V_i \ . \tag{2}$$

This means that for any v, there exists $v_i \in V_i$ such that $v = \sum_{i=1}^{m} v_i$.

After the decomposition of the space as in (2), there are two different ways to solve the nonlinear problem (1). The first alternative is to decompose K into a sum of $K_i \subset V_i$, $i = 1, 2, \cdots, m$, i.e.

$$K = K_1 + K_2 + \cdots + K_m = \sum_{i=1}^{m} K_i,$$

and then solve a minimization problem over each subset K_i in parallel or sequentially. The convergence analysis and numerical experiments have been done in [Tai00].

For the second alternative, we only need to decompose the space V as in (2), but we do not need to decompose the constraint set K, see the next section for the detailed algorithms. Uniform linear convergence rate analysis for these algorithms is still missing in the literature. The contribution of this work to give a mesh independent linear convergence rate estimate for domain decomposition and multigrid methods for these algorithms. The techniques used in the analysis are extensions of the techniques used in [TE98, Tai00, TT98, TX01].

We will find the proper assumptions on the decomposed subspaces to guarantee that the algorithms will have a uniform linear convergence rate and then we verify that these assumptions are really valid for domain decomposition and multigrid methods.

[1]Department of Mathematics, University of Bergen, Johannes Brunsgate 12, 5007, Bergen, Norway. Email: Tai@mi.uib.no and URL: http://www.mi.uib.no/~tai.

[2]Department of Mathematics, University of Bergen, Johannes Brunsgate 12, 5007, Bergen, Norway. Email: bjornoh@mi.uib.no and URL: http://www.mi.uib.no/~bjornoh.

[3]Department of Mathematics, Pennsylvania State University, University Park, Pennsylvania 16802. Email: xu@math.psu.edu.

The algorithms and some assumptions

Algorithm 1 For a given $u^n \in K$ and $\alpha \in (0, 1/m)$, compute $e_i^{n+1} \in V_i$ in parallel for $i = 1, 2, \cdots, m$ such that

$$e_i^{n+1} = \arg \min_{\substack{v_i + u^n \in K \\ v_i \in V_i}} G(v_i) \quad \text{with} \quad G(v_i) = F(u^n + v_i). \tag{3}$$

and then update

$$u^{n+1} := u^n + \alpha \sum_{i=1}^{m} e_i^{n+1}$$

■

Algorithm 2 For a given $u^n \in K$ and $\alpha \in (0, 1)$, compute $e_i^{n+1} \in V_i$ sequentially for $i = 1, 2, \cdots, m$ such that

$$e_i^{n+1} = \arg \min_{\substack{v_i + u^{n+\frac{i-1}{m}} \in K \\ v_i \in V_i}} G(v_i) \quad \text{with} \quad G(v_i) = F(u^{n+\frac{i-1}{m}} + v_i). \tag{4}$$

and update

$$u^{n+\frac{i}{m}} := u^{n+\frac{i-1}{m}} + \alpha e_i^{n+1}$$

■

For the minimization functional F, we only need to assume that F is Gâteaux differentiable (see [ET76]) and that there exists a constant $\kappa > 0$ such that

$$\langle F'(w) - F'(v), w - v \rangle \geq \kappa \|w - v\|_V^2, \quad \forall w, v \in V \ . \tag{5}$$

Here $\langle \cdot, \cdot \rangle$ is the duality pairing between V and its dual space V', i.e. the value of a linear function at an element of V. Under the assumption (5), problem (1) has a unique solution, see [ET76, p. 35]. For some nonlinear problems, the constant κ may depend on v and w.

As in [TE98, TT98, TX01], we shall use two constants in the estimation of the rate of the convergence of the algorithms. First, we assume that there exists a constant $C_1 > 0$ and this constant is only related to the decomposition (2). With the constant C_1 and the decomposition (2), it is assumed that for any $v, w \in K$, one can find $z_i \in V_i$ to satisfy

$$v - w = \sum_{i=1}^{m} z_i, \quad z_i + w \in K, \quad \text{and} \quad \left(\sum_{i=1}^{m} \|z_i\|_V^2 \right)^{\frac{1}{2}} \leq C_1 \|v - w\|_V \ . \tag{6}$$

In addition to the assumption of the existence of such a constant C_1, we also assume that there is a $C_2 > 0$ which is the least constant satisfying the following inequality for any $w_{ij} \in V, u_i \in V_i$ and $v_j \in V_j$:

$$\sum_{i,j=1}^{m} \left| \langle F'(w_{ij} + u_i) - F'(w_{ij}), v_j \rangle \right| \leq C_2 \left(\sum_{i=1}^{m} \|u_i\|_V^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{m} \|v_j\|_V^2 \right)^{\frac{1}{2}}. \tag{7}$$

Later we shall show that these assumptions are valid for domain decomposition and multigrid methods. Moreover, the constants C_1 and C_2 are mesh independent.

The convergence of the parallel subspace correction method

We shall only do the convergence analysis for Algorithm 1. For notation simplicity, we define u to be the unique solution of (1) and for any $n \geq 0$ we define

$$u^n = \sum_{i=1}^{m} u_i^n, \;\; \hat{u}^{n+1} = u^n + \sum_{i=1}^{m} e_i^{n+1}, \; d_n = F(u^n) - F(u). \tag{8}$$

Theorem 1 *Assuming that the space decomposition satisfies (6), (7) and that the functional F satisfies (5). Then for Algorithms 1, we have*

$$\frac{F(u^{n+1}) - F(u)}{F(u^n) - F(u)} \leq 1 - \frac{\alpha}{(\sqrt{1+C^*} + \sqrt{C^*})^2}, \tag{9}$$

with

$$C^* = \left(C_2 + \frac{(C_1 C_2)^2}{2\kappa} \right) \frac{2}{\kappa}. \tag{10}$$

Proof. Since e_i^{n+1} minimizes (3), it satisfies (see [ET76])

$$\langle F'(u^n + e_i^{n+1}), v_i - e_i^{n+1} \rangle \geq 0, \quad \forall v_i \in V_i \quad \text{satisfying} \quad v_i + u^n \in K. \tag{11}$$

Under the assumption of (5), it is known that (see [TE98, Lemma 3.2])

$$F(w) - F(v) \geq \langle F'(v), w - v \rangle + \frac{\kappa}{2} \|w - v\|_V^2 \,, \quad \forall v, w \in V. \tag{12}$$

Using these results, we get that

$$\begin{aligned}
F(u^n) - F(u^{n+1}) &= F(u^n) - F\left(u^n + \alpha \sum_{i=1}^{m} e_i^{n+1} \right) \\
&= F(u^n) - F\left(\sum_{i=1}^{m} \alpha (u^n + e_i^{n+1}) + (1 - \alpha m) u^n \right) \\
&\geq F(u^n) - \alpha \sum_{i=1}^{m} F(u^n + e_i^{n+1}) - (1 - \alpha m) F(u^n) \\
&= \alpha \sum_{i=1}^{m} (F(u^n) - F(u^n + e_i^{n+1})) \\
&\geq \frac{\alpha \kappa}{2} \sum_{i=1}^{m} \|e_i^{n+1}\|_V^2 \qquad \text{(using (11) and (12))}.
\end{aligned} \tag{13}$$

The argument used to get the above estimates is the same as the unconstrained case, see [TX01]. For notational simplicity, we introduce for a given i

$$\sigma_j^n = \begin{cases} u^n + \displaystyle\sum_{k=i}^{j+i-1} e_k^{n+1} \,, & \forall j \in [1, m-i+1] \,; \\ u^n + \displaystyle\sum_{k=i}^{m} e_k^{n+1} + \sum_{k=1}^{j-m+i-1} e_k^{n+1} \,, & \forall j \in [m-i+2, m] \,. \end{cases} \tag{14}$$

It is clear that σ_j^n depends on i. Moreover, we see that

$$\begin{aligned}
\sigma_1^n &= u^n + e_i^{n+1}, \\
\sigma_2^n &= u^n + e_i^{n+1} + e_{i+1}^{n+1}, \\
&\vdots \\
\sigma_m^n &= u^n + \sum_{k=1}^{m} e_k^{n+1}.
\end{aligned}$$

It is easy to see that

$$F'\Big(u^n + \sum_{j=1}^{m} e_j^{n+1}\Big) - F'(u^n + e_i^{n+1}) = \sum_{j=2}^{m} \left(F'(\sigma_j^n) - F'(\sigma_{j-1}^n)\right) \ . \tag{15}$$

From assumption (6), there exists $z_i^n \in V_i$ such that

$$u - u^n = \sum_{i=1}^{m} z_i^n, \qquad z_i^n + u^n \in K, \qquad \Big(\sum_{i=1}^{m} \|z_i^n\|^2\Big)^{\frac{1}{2}} \le C_1 \|u - u^n\|. \tag{16}$$

We shall now use all of the above to estimate

$$\begin{aligned}
\langle F'(\hat{u}^{n+1}), \hat{u}^{n+1} - u\rangle &= \sum_{i=1}^{m} \Big\langle F'(\hat{u}^{n+1}), e_i^{n+1} - z_i^n \Big\rangle \\
&\le \sum_{i=1}^{m} \Big\langle F'(\hat{u}^{n+1}) - F'(u + e_i^{n+1}), e_i^{n+1} - z_i^n \Big\rangle \qquad \text{(using (11) and (16))} \\
&= \sum_{i=1}^{m}\sum_{j=2}^{m} \Big\langle F'(\sigma_j^n) - F'(\sigma_{j-1}^n), e_i^{n+1} - z_i^n \Big\rangle \qquad \text{(using (15))} \\
&\le C_2 \Big(\sum_{i=1}^{m} \|e_i^{n+1}\|^2\Big)^{\frac{1}{2}} \Big(\sum_{i=1}^{m} \|e_i^{n+1} - z_i^n\|^2\Big)^{\frac{1}{2}} \qquad \text{(using (7))} \qquad (17) \\
&\le C_2 \Big(\sum_{i=1}^{m} \|e_i^{n+1}\|^2\Big)^{\frac{1}{2}} \Big(\Big(\sum_{i=1}^{m} \|e_i^{n+1}\|^2\Big)^{\frac{1}{2}} + C_1\|u - u^n\|\Big) \ \text{(using (6), (8) and (16))} \\
&= C_2 \sum_{i=1}^{m} \|e_i^{n+1}\|^2 + C_1 C_2 \Big(\sum_{i=1}^{m} \|e_i^{n+1}\|^2\Big)^{\frac{1}{2}} \|u - u^n\|.
\end{aligned}$$

The rest of the proof is the same as in [Tai00]. ■

The general theory developed for (1) will be applied to the following obstacle problem in connection with finite element approximations:

$$\text{Find } u \in K, \quad \text{such that} \quad a(u, v-u) \ge f(v-u), \quad \forall v \in K, \tag{18}$$

with $a(v,w) = \int_\Omega \nabla v \cdot \nabla w \, dx$, $K = \{v \in H_0^1(\Omega) |\ v(x) \ge \psi(x) \text{ a.e. in } \Omega\}$. For the analysis, it can be assumed without loss of any generality that

$$\psi = 0. \tag{19}$$

It is well known that the above problem is equivalent to the following minimization problem

$$\min_{v\in K} F(v), \quad F(v) = \frac{1}{2}a(v,v) - f(v), \tag{20}$$

assuming that $f(v)$ is a linear functional on $H_0^1(\Omega)$. For the obstacle problem (18), the minimization space $V = H_0^1(\Omega)$. Correspondingly, we have $\kappa = 1$ for assumption (5).

Overlapping domain decomposition

In this section we apply our algorithms to the overlapping domain decomposition method. For the domain Ω, we first partition it into a coarse mesh division $\{\mathcal{T}_H\}$ with a mesh size H and then refine it into a fine mesh partition $\{\mathcal{T}_h\}$ with a mesh size $h < H$. We assume that both the coarse mesh and the fine mesh are shape-regular. Let $\{\Omega_i\}_{i=1}^M$ be a nonoverlapping domain decomposition for Ω and each Ω_i is the union of some coarse mesh elements. Let $S^H \subset W^{1,\infty}(\Omega)$ and $S^h \subset W^{1,\infty}(\Omega)$ be the continuous, piecewise linear finite element spaces over the H-level and h-level subdivisions of Ω respectively. More specifically,

$$S^H = \left\{ v \in W^{1,\infty}(\Omega_H) \middle| \quad v|_{\Omega_i} \in P_1(\Omega_i), \forall i \right\},$$

$$S^h = \left\{ v \in W^{1,\infty}(\Omega_h) \middle| \quad v|_{\mathcal{T}} \in P_1(\mathcal{T}), \forall \mathcal{T} \in \mathcal{T}_h \right\}.$$

For each Ω_i, we consider an enlarged subdomain Ω_i^δ consisting of elements $\mathcal{T} \in \mathcal{T}_h$ with $\text{dist}(\mathcal{T}, \Omega_i) \leq \delta$. The union of Ω_i^δ covers $\bar{\Omega}_h$ with overlaps of size δ. Let us denote the piecewise linear finite element space with zero traces on the boundaries $\partial\Omega_i^\delta \backslash \partial\Omega$ as $S^h(\Omega_i^\delta)$. Then one can show that

$$S^h = \sum_{i=1}^M S^h(\Omega_i^\delta) \quad \text{and} \quad S^h = S^H + \sum_{i=1}^M S^h(\Omega_i^\delta)\,. \tag{21}$$

For the overlapping subdomains, assume that there exist m colors such that each subdomain Ω_i^δ can be marked with one color, and the subdomains with the same color will not intersect with each other. For suitable overlaps, one can always choose $m = 2$ if $d = 1; m \leq 4$ if $d = 2; m \leq 8$ if $d = 3$. Let Ω_i^c be the union of the subdomains with the i^{th} color, and

$$V_i = \{ v \in S^h | \quad v(x) = 0, \quad x \notin \Omega_i^c \}\ i = 1, 2, \cdots, m.$$

By denoting subspaces $V_0 = S_H$, $V = S_h$, we get from (21) that

$$a). \quad V = \sum_{i=1}^m V_i \qquad \text{and} \qquad b). \quad V = V_0 + \sum_{i=1}^m V_i. \tag{22}$$

Note that the summation index is from 0 to m instead of from 1 to m when the coarse mesh is added.

Let $\{\theta_i\}_{i=1}^m$ be a partition of unity with respect to $\{\Omega_i^c\}_{i=1}^m$, i.e. $\theta_i \in V_i$, $\theta_i \geq 0$ and $\sum_{i=1}^m \theta_i = 1$. It can be chosen so that

$$|\nabla\theta_i| \leq C/\delta, \quad \theta_i(x) = \begin{cases} 1 & \text{if } x \in \tau, \text{ distance } (\tau, \partial\Omega_i^c) \geq \delta \text{ and } \tau \subset \Omega_i^c, \\ 0 & \text{on } \overline{\Omega \backslash \Omega_i^c}. \end{cases} \tag{23}$$

In the following, we shall give the definition of a nonlinear interpolation operator $I_H^{\ominus}$: $S_h \mapsto S_H$ which was introduced in [Tai00]. Denote by $\mathcal{N}_H = \{x_0^i\}_{i=1}^{n_0}$ all the interior nodes for $\mathcal{T}_H$. For a given x_0^i, let ω_i be the union of the mesh elements of $\mathcal{T}_H$ having x_0^i as one of its vertices, i.e. $\omega_i := \cup\{\tau \in \mathcal{T}_H, x_0^i \in \bar{\tau}\}$. Let $\{\phi_0^i\}_{i=1}^{n_0}$ be the associated nodal basis functions of S_H satisfying $\phi_0^i(x_0^k) = \delta_{ik}$, $\phi_0^i \geq 0$, $\forall i$ and $\sum_i \phi_0^i(x) = 1$. It is clear that ω_i is the support of ϕ_0^i. Given a nodal point $x_0^i \in \mathcal{N}_H$ and a $v \in S_h$, let $I_i v = \min_{\omega_i} v(x)$. The interpolated function is then defined as

$$I_H^{\ominus} v := \sum_{x_0^i \in \mathcal{N}_H} (I_i v)\phi_0^i(x).$$

From the definition, it is easy to see that

$$I_H^{\ominus} v \leq v, \quad \forall v \in S_h, \tag{24}$$
$$I_H^{\ominus} v \geq 0, \quad \forall v \geq 0, v \in S_h. \tag{25}$$

Moreover, the interpolation for a given $v \in S_h$ on a finer mesh is always bigger than the corresponding interpolation on a coarser mesh due to the fact that each coarser mesh element contains several finer mesh elements, i.e.

$$I_{h_1}^{\ominus} v \leq I_{h_2}^{\ominus} v, \quad \forall h_1 \geq h_2 \geq h, \quad \forall v \in S_h. \tag{26}$$

In addition, the interpolation operator also has the following approximation properties (c.f. [Tai00])

$$\|I_H^{\ominus} v - I_H^{\ominus} w - (v - w)\|_0 \leq c_d H |v - w|_1, \forall v, w \in S_h \tag{27}$$
$$\|I_H^{\ominus} v - v\|_0 \leq c_d H |v|_1, \qquad \|I_H^{\ominus} v - I_H^{\ominus} w\|_1 \leq c_d \|v - w\|_1, \forall v, w \in S_h, \tag{28}$$

where $c_d = C$ if $d = 1$; $c_d = C\left(1 + \left|\log \frac{H}{h}\right|^{\frac{1}{2}}\right)$ if $d = 2$ and $c_d = C\left(\frac{H}{h}\right)^{\frac{1}{2}}$ if $d = 3$.

We first use decomposition (22.a) to decompose the finite element space $V = S_h$, i.e. the coarse mesh is not used in the computations. Let I_h be the Lagrangian interpolation operator which uses the function values at the h-level nodes. In order to estimate constant C_1, we take $z_i = I_h(\theta_i(v - w))$ for any $v, w \in K$. As $\sum_{i=1}^m \theta_i = 1$, thus $\sum_{i=1}^m z_i = v - w$ using the linearity of I_h. Moreover,

$$z_i + w = I_h(\theta_i(v - w) + w).$$

Under assumption (19), it follows from the fact that $\theta_i \in (0, 1)$ and the convexity of K that $z_i \in K$. It is easy to prove that the following estimate is correct and the proof is exactly the same as for the linear unconstrained case [SBG96, Xu92]:

$$C_1 \leq C(1 + \delta^{-1}), \quad C_2 = \sqrt{m}.$$

If we shall use the coarse mesh, then the decomposition is as given in (22.b). The estimation for C_2 is the same as for linear problems, we just need to find the biggest constant which satisfies (6).

In order to show that assumption (6) is valid for decomposition (22.b), we first decompose $v - w$ into

$$v - w = \sigma^{\oplus} - \sigma^{\ominus}, \quad \sigma^{\oplus} = \max(0, v - w),\ \sigma^{\ominus} = \max(0, w - v), \tag{29}$$

and then define $z_0 \in V_0$ as

$$z_0 = I_H^{\ominus} I_h \sigma^{\oplus} - I_H^{\ominus} I_h \sigma^{\ominus}.$$

Under assumption (19), we see that $v, w \geq 0$. From (24) and (25), it is true that

$$0 \leq I_H^{\ominus} I_h \sigma^{\oplus} \leq I_h \sigma^{\oplus} \leq v, \quad 0 \leq I_H^{\ominus} I_h \sigma^{\ominus} \leq I_h \sigma^{\ominus} \leq w \text{ and so } -w \leq z_0 \leq v. \tag{30}$$

Due to the special structure of the functions $\sigma^{\oplus}$ and $\sigma^{\ominus}$, it is in fact easy to prove that

$$|I_h \sigma^{\oplus}|_1 \leq C|v-w|_1, \quad |I_h \sigma^{\ominus}|_1 \leq C|v-w|_1, \tag{31}$$

where the constant C only depends on the minimal angle conditions. From the above inequalities and estimate (27), it is easy to see that

$$\|z_0 - (v-w)\|_l \leq c_d H^{1-l}|v-w|_1, \quad l = 0, 1.$$

Taking

$$z_i = I_h(\theta_i(v-w-z_0)),\ i = 1, 2, \cdots, m,$$

we get by using the linearity of I_h, the equality $I_h w = w$ and (30) that

$$z_0 + \sum_{i=1}^{m} z_i = v - w, \quad z_0 + w \geq 0, \quad \text{and} \quad z_i + w = I_h(\theta_i(v-z_0) + (1-\theta_i)w) \geq 0.$$

Using the approximation properties (27)-(28), the following estimate is correct and the proof is the same as for the linear unconstrained case, see [TX01]:

$$\left(\|z_0\|_1^2 + \sum_{i=1}^{m} \|z_i\|_1^2 \right)^{\frac{1}{2}} \leq (m+1)^{\frac{1}{2}} c_d \left(1 + \left(\frac{H}{\delta} \right)^{\frac{1}{2}} \right) |v-w|_1. \tag{32}$$

Thus it is shown that assumption (6) is valid for decomposition (22.b) with

$$C_1 = (m+1)^{\frac{1}{2}} c_d \left(1 + \left(\frac{H}{\delta} \right)^{\frac{1}{2}} \right).$$

Assumption (7) has been shown to be correct for the decomposition (22.b) with $C_2 = \sqrt{m+1}$ and m being the number of colors, see [TX01], see also [SBG96, DW94, Xu92].

Multigrid decomposition

A multigrid algorithm is built upon the subspaces that are defined on a nested sequence of finite element partitions. We assume that the finite element partition $\mathcal{T}$ is constructed by a successive refinement process. More precisely, $\mathcal{T} = \mathcal{T}_J$ for some $J > 1$, and $\mathcal{T}j$ for, $j \leq J$ is a nested sequence of quasi-uniform finite element partitions, i.e. $\mathcal{T}j$ consist of finite elements $\mathcal{T}j = \{\tau_j^i\}$ of size h_j such that $\Omega = \cup_i \tau_j^i$ for which the quasi-uniformity constants are independent of j and τ_{j-1}^l is a union of elements of $\{\tau_j^i\}$. We further assume that there is a constant $\gamma < 1$, independent of j, such that h_j is proportional to γ^{2j}.

As an example, in the two dimensional case, a finer grid is obtained by connecting the midpoints of the edges of the triangles of the coarser grid, with $\mathcal{T}_1$ being the given coarsest

initial triangulation, which is quasi-uniform. In this example, $\gamma = 1/\sqrt{2}$. We can use much smaller γ in constructing the meshes, but the constant C_1 is getting larger when γ is becoming smaller, see (35).

Corresponding to each finite element partition $\mathcal{T}j$, a finite element space $\mathcal{M}_j$ can be defined by

$$\mathcal{M}_j = \{v \in W^{1,\infty}(\Omega) : v|_\tau \in \mathcal{P}_1(\tau), \quad \forall\, \tau \in \mathcal{T}j\}.$$

Each finite element space $\mathcal{M}_j$ is associated with a nodal basis, denoted by $\{\phi_j^i\}_{i=1}^{n_j}$ satisfying

$$\phi_j^i(x_j^k) = \delta_{ik}$$

where $\{x_j^k\}_{k=1}^{n_j}$ is the set of all nodes of the elements of $\mathcal{T}_j$. Associated with each such a nodal basis function, we define a one dimensional subspace as follows

$$\mathcal{M}_j^i = \text{span } (\phi_j^i).$$

It is easy to see that

$$\mathcal{M}_J = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \mathcal{M}_j^i. \tag{33}$$

Similar as for the two-level decomposition, we first decompose $v - w$ for any $v, w \geq 0$ as in (29). We then define $\sigma_j^{\oplus} = I_{h_j}^{\ominus} I_h \sigma^{\oplus}$, $\sigma_j^{\ominus} = I_{h_j}^{\ominus} I_h \sigma^{\ominus}$ for $j = 1, 2, \cdots, J$ and $\sigma_0^{\oplus} = 0,\ \sigma_0^{\ominus} = 0$. From properties (24)–(28) and the fact that $v, w \geq 0$, it is true that

$$0 \leq \sigma_j^{\oplus} \leq I_h \sigma^{\oplus} \leq v, \quad \|\sigma_j^{\oplus} - \sigma^{\oplus}\|_l \leq \tilde{c}_d h_j^{1-l} |v - w|_1,\ l = 0, 1.$$

$$0 \leq \sigma_j^{\ominus} \leq I_h \sigma^{\ominus} \leq w, \quad \|\sigma_j^{\ominus} - \sigma^{\ominus}\|_l \leq \tilde{c}_d h_j^{1-l} |v - w|_1,\ l = 0, 1,$$

where

$$\tilde{c}_d = \begin{cases} C, & \text{if} \quad d = 1; \\ C(1 + |\log h|^{\frac{1}{2}}), & \text{if} \quad d = 2; \\ C h^{-\frac{1}{2}}, & \text{if} \quad d = 3. \end{cases}$$

Define

$$z_J = v - w - \sigma_{J-1}^{\oplus} + \sigma_{J-1}^{\ominus}, \quad z_j = \sigma_j^{\oplus} - \sigma_{j-1}^{\oplus} - (\sigma_j^{\ominus} - \sigma_{j-1}^{\ominus}),\ j = 1, 2, \cdots J - 1.$$

From (25) and (26), we see that $\sigma_j^{\oplus} - \sigma_{j-1}^{\oplus} \geq 0$ and $\sigma_{j-1}^{\ominus} \geq 0$, we thus get

$$z_j + w \geq w - \sigma_j^{\ominus} \geq 0, \quad j = 1, 2, \cdots, J - 1. \tag{34}$$

Similar argument shows that $z_J + w = v - \sigma_{J-1}^{\oplus} + \sigma_{J-1}^{\ominus} \geq 0$. The fact that $\sum_{j=1}^{J} z_j = v - w$ is an easy consequence of the definitions of z_j. A further decomposition of z_j is given by

$$z_j = \sum_{i=1}^{n_j} z_j^i \quad \text{with} \quad z_j^i = z_j(x_j^i)\phi_j^i.$$

It is easy to see that

$$v - w = \sum_{j=1}^{J} z_j = \sum_{j=1}^{J} \sum_{i=1}^{n_j} z_j^i.$$

From (34) and the fact that $w \geq 0$, it is true that

$$z_j^i + w \geq 0 \quad \forall i, j \quad \text{which means that} \quad z_j^i + w \in K.$$

Using the approximation properties (27)–(28), the following estimate is correct, see [TX01, Tai00]:

$$\sum_{j=1}^{J} \sum_{i=1}^{n_j} \left|z_j^i\right|_1^2 = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \left|z_j(x_j^i)\right|^2 \left|\phi_j^i\right|_1^2 \leq C \sum_{j=1}^{J} h_j^{d-2} \sum_{i=1}^{n_j} \left|z_j(x_j^i)\right|^2$$
$$\leq C \sum_{j=1}^{J} h_j^{-2} \left|z_j\right|_0^2 \leq \tilde{c}_d \sum_{j=1}^{J} h_j^{-2} h_{j-1}^2 \left|v - w\right|_1^2 \leq \tilde{c}_d \gamma^{-2} J |v - w|_1^2.$$

The estimation for C_2 is the same as for the unconstrained case [Tai00]. Thus for the multigrid decomposition (33) we have

$$C_1 = \tilde{c}_d \gamma^{-1} J^{\frac{1}{2}} = \tilde{c}_d \gamma^{-1} |\log h|^{\frac{1}{2}}, \quad C_2 = C(1 - \gamma^d)^{-1}. \tag{35}$$

In the above, γ is the mesh ratio for the multigrid method and d is the dimension for $\Omega \subset R^d$. Thus the assumptions (6)–(7) are valid for the multigrid decomposition. Using Theorem 1, we see that the convergence rate for Algorithm 1 is:

$$\frac{F(u^{n+1}) - F(u)}{F(u^n) - F(u)} \leq 1 - \frac{\alpha}{1 + \tilde{c}_d \gamma^{-2} J}.$$

Some numerical tests

Numerical tests shall be done both for Algorithm 1 and Algorithm 2. However, we shall only explain some of the implementation details for Algorithm 1. The implementation for Algorithm 2 follows the similar techniques.

Define $u^{n+\frac{i}{m}} = u^n + e_i^{n+1}$ for Algorithm 1. When decomposition (22.b) is used for the finite element method, it can be seen that the subproblems we need to solve over each of the subdomains is:

$$a). \begin{cases} -\Delta u^{n+\frac{i}{m}} \geq f & \text{in } \Omega_i^c, \\ u^{n+\frac{i}{m}} = u^n & \text{on } \partial\Omega_i^c, \\ u^{n+\frac{i}{m}} \geq \psi & \text{in } \Omega_i^c. \end{cases} \quad \text{or} \quad b). \begin{cases} -\Delta e_i^{n+1} \geq f + \Delta u^n & \text{in } \Omega_i^c, \\ e_i^{n+1} = 0 & \text{on } \partial\Omega_i^c, \\ e_i^{n+1} \geq \psi - u^n & \text{in } \Omega_i^c. \end{cases} \tag{36}$$

It is better to solve (36.a) then get $e_i^{n+1} = u^{n+\frac{i}{m}} - u^n$. If we use (36.b) to get e_i^{n+1}, then we must compute the residual $f + \Delta u^n$ over each subdomain. This does not require extra cost for the parallel algorithm 1 as the residual is needed for the coarse mesh subproblem anyway.

However, it requires extra cost for the sequential algorithm 2 and for the case when the coarse mesh is not used. If the coarse mesh is used, we need to solve

$$e_0^{n+1} = \arg \min_{\substack{v_0 \in V_0 \\ v_0 \geq \psi - u^n}} G(v_0) \quad \text{with} \quad G(v_0) = F(u^n + v_0).$$

The unknowns for the minimization problem is the coarse mesh nodal values, but the constraint $v_0 \geq \psi - u^n$ is imposed over all the fine mesh nodes. This is not an easy problem to solve. In our implementation, we have used the Augmented Lagrangian method to minimize the functional and at the same time to impose the constraint over all the fine mesh nodes.

For the multigrid decomposition (33), each subproblem (3) is one dimensional. We just need to solve

$$\tilde{e}_{i,j}^{n+1} = \arg \min_{v_j^i \in \mathcal{M}_j^i} G(v_j^i) \quad \text{with} \quad G(v_j^i) = F(u^n + v_j^i) \tag{37}$$

and then project the value above the one dimensional constraint, i.e.

$$e_{i,j}^{n+1}(x_j^i) = \max_{x \in \text{support}(\phi_j^i)} \left(\tilde{e}_{i,j}^{n+1}(x_j^i),\ \frac{\psi(x) - u^n(x)}{\phi_j^i(x)} \right). \tag{38}$$

The solving of (37) is the same as the unconstrained case. The only extra thing we need to do is the projection given in (38).

For the test results, we shall solve the obstacle problem on $\Omega = [-2, -2] \times [-2, 2]$ with $f = 0$. The obstacle is $\psi(x, y) = \sqrt{x^2 + y^2}$ when $x^2 + y^2 \leq 1$ and $\psi(x, y) = -1$ elsewhere. This problem has an analytical solution [Tai00]. Note that the obstacle function ψ is not even in $H^1(\Omega)$ due to the discontinuity. Even for such a difficult problem, uniform linear convergence has been observed in our experiments. In the implementations, the non-zero obstacle can be shifted to the right hand side.

We will try both sequential and parallel domain decomposition. In the plots, en is the error between the computed solution and the true FEM solution in the energy norm. In all the computations, u_0 is taken to be $\psi + 100$. The domain decomposition solvers all use a two-element overlap and the subproblems are solved by an augmented Lagrangian iterative method. The mesh is discretized by letting $h = 4/64$ and $H = 4/8$. So there are 64 subdomains. The convergence-results are shown in Figure 1. In the figure, we also compare the convergence with the two corresponding algorithms of [Tai00]. It seems that the algorithms here has the same convergence rate as the one of [Tai00]. However, the B-sequential algorithm, which refers to Algorithm 2, seems to be slightly faster than the C-sequential algorithm which refers to that of [Tai00].

For the multigrid method, we have only tested the sequential algorithms. We use a V-cycle method. This is equivalent to repeat the one dimensional subspace once more in the decomposition (33) and order them properly. The convergence for 5, 6 and 7 levels are shown in Figure 2. The convergence rate is about 0.6 for all the three different levels.

References

[DW94]Maksymilian Dryja and Olof B. Widlund. Domain decomposition algorithms with small overlap. *SIAM J. Sci.Comput.*, 15(3):604–620, May 1994.

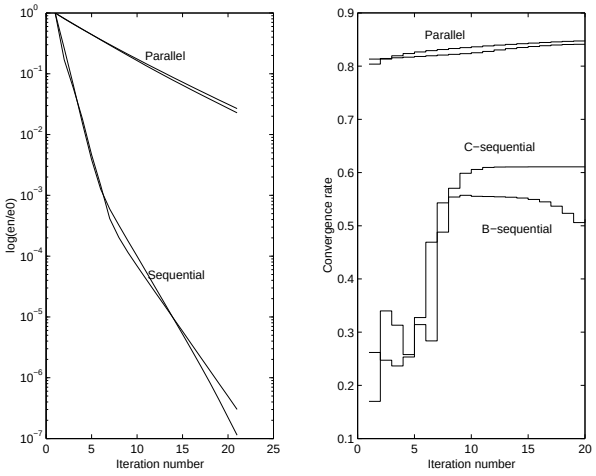


Figure 1: Domain decomposition. The B solver is Algorithm 2 and the C solver is the corresponding algorithm of [Tai00].

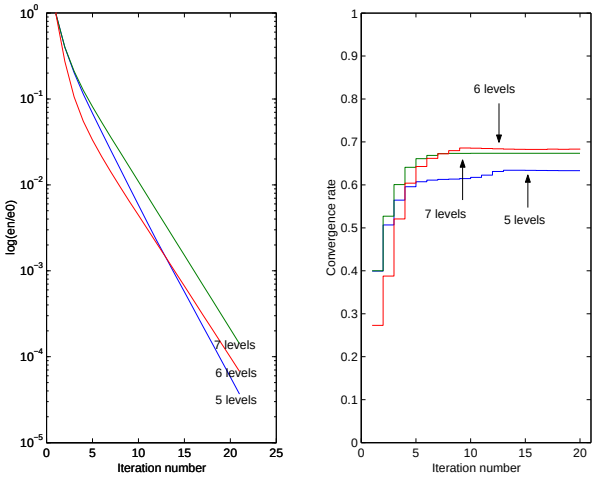


Figure 2: Convergence rate of the multigrid solver with several different levels

[ET76]I. Ekeland and R. Temam. *Convex analysis and variational problems.* North-Holland, Amsterdam, 1976.

[SBG96]Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[Tai00]Xue-Cheng Tai. Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities. Technical Report 150, Department of Mathematics, University of Bergen, November 2000.

[TE98]X.-C. Tai and M. S. Espedal. Rate of convergence of some space decomposition method for linear and non-linear elliptic problems. *SIAM J. Numer. Anal.*, 35:1558–1570, 1998.

[TT98]Xue-Cheng Tai and Paul Tseng. Convergence rate analysis of an asynchronous space decomposition method for convex minimization. Technical Report 120, Department of Mathematics, University of Bergen, August 1998.

[TX01]X. C. Tai and J. Xu. Global convergence of subspace correction methods for convex optimization problems. *Math. Comput.*, 2001.

[Xu92]Jinchao Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, December 1992.

Part II

Theory

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

13 Discontinuous Hybrid Formulation turned to Domain Decomposition

A. Agouzal[1], N. Debit[2]

Introduction

We consider a macro hybrid primal finite element formulation turned to domain decomposition which produces a completely discontinuous approximation. The key point of the framework is an analogous of an argument already used in stabilization techniques for DDM with non matching grids, [BFMR97]. The resulting approximation is conforming and the convergence is established with no inspection of consistency error, nor inf-sup condition.

The finite element approximation of the second order elliptic equations has been investigated using several different approaches (see e.g. [Cia78] and the references therein). Previous analysis in primal formulation of these problems has been done for three types of approximation schemes : one which produces a continuous piecewise polynomial approximation, one which produces a piecewise polynomial approximation with a fixed number of continuous moments accross interelement edges (nonconforming approximation) and one which produces completely discontinuous polynomial approximation (interior penalty methods) [Arn82]. All these finite element methods have optimal order of convergence, assuming sufficient regularity. More recently, there has been growing interest in methods which can produce a completely discontinuous approximation for diffusion problems [JO98]. The motivation for developing these methods was the flexiblity afforded by discontinuous finite element spaces. Another advantage that has recently become apparent is the application of domain decomposition algorithms for the solution of the discrete solution.

1 Macro hybrid formulation for the model problem

Let Ω be a simply connected polygonal domain of $\mathbb{R}^d$, $d = 2$ or 3, and Γ its boundary. Let us perform a non overlaping domain decomposition on Ω,

$$\overline{\Omega} = \cup_{i=1}^{I} \overline{\Omega}_i$$

$$\Omega_i \cap \Omega_j = \emptyset, \;\; 1 \leq i \neq j \leq I.$$

We assume that each subdomain Ω_i is polygonal and set the following notations

$$\Gamma_{ij} = \partial\Omega_i \cap \partial\Omega_j, \;\; \text{for } \; 1 \leq i \neq j \leq I,$$

$$\forall i \in \{1, \ldots, I\}, \Gamma_i = \partial\Omega_i \setminus \Gamma$$

[1]University Lyon 1, , 69622 Villeurbanne, France, agouzal@numerix.univ-lyon1.fr
[2]MCS-ISTIL - University Lyon 1, 69622 Villeurbanne, France, debit@mcs.univ-lyon1.fr

and

$$\mathcal{M} = \{m = (i,j),\ 1 \le i \ne j \le I, \text{ such that } \ \text{meas}(\partial\Omega_i \cap \partial\Omega_j) \ne 0\}.$$

We consider, for simplicity, the Dirichlet problem for the Laplace equation :

$$-\Delta u = f \text{ in } \Omega, \qquad u = 0 \text{ on } \partial\Omega = \Gamma. \tag{1}$$

where $f \in L^2(\Omega)$.
First, we introduce the following functional spaces

$$W_i = \{v \in H^1(\Omega_i);\ \frac{\partial u}{\partial n_i} \in L^2(\Gamma_i) \ \text{ and } \ v_{|\partial\Omega_i \cap \Gamma} = 0 \text{ if meas}(\partial\Omega_i \cap \Gamma) \ne 0\}$$

where $\dfrac{\partial u}{\partial n_i}$ is the outward normal derivative of u to the boundary $\Gamma_i, i = 1, \ldots, I$.

$$\dot{W} = \prod_{i=1}^{I} W_i,$$

$$S = \{\phi = (\phi_i = v_{|\Gamma_i})_{1 \le i \le I} \quad \text{with } v \in H_0^1(\Omega)\},$$

For $((\dot{u}, \phi), (\dot{v}, \psi)) \in (\dot{W} \times S)^2$, we define the product bilinear form

$$\dot{B}((\dot{u}, \phi), (\dot{v}, \psi)) = \sum_{i=1}^{I} \int_{\Omega_i} \nabla u_i \nabla v_i \, dx - < \frac{\partial u_i}{\partial n_i}, v_i - \psi_i >_{0,\Gamma_i} \tag{2}$$

$$+ < \frac{\partial v_i}{\partial n_i}, u_i - \phi_i >_{0,\Gamma_i} + \delta_i (u_i - \phi_i, v_i - \psi_i)_{0,\Gamma_i} \tag{3}$$

$$\tag{4}$$

For $i = 1, \ldots, I$, let $\mathcal{T}_{h_i}$ be a regular triangulation of the subdomain Ω_i with triangular $(d = 2)$ or tetrahedral $(d = 3)$ finite elements whose diameters are less or equal than h_i and k_i be a positive integer. We assume that the triangulation is uniformily regular near Γ_i. We introduce the standard finite element space

$$V_{h_i} = \{v_{h_i} \in \mathcal{C}^0(\overline{\Omega}_i); \forall T \in \mathcal{T}_{h_i}, v_{h_i|T} \in P_{k_i}(T),\ v_{|\partial\Omega_i \cap \Gamma} = 0 \text{ if meas}(\partial\Omega_i \cap \Gamma) \ne 0\}$$

and we set

$$\dot{W}_h = \prod_{i=1}^{I} V_{h_i}.$$

Remark that V_{h_i} is a subspace of W_i, and so $\dot{W}_h$ is a subspace of $\dot{W}$.
Let us now proceed with the squeleton; For all $m = (i,j) \in \mathcal{M}$, let $\mathcal{T}_{h_m}$ be a regular subdivision $(d = 2)$ or triangulation $(d = 3)$ of $\Gamma_{i,j}$ by finite elements whose diameters are less or equal than h_m and k_m be a positive integer. We introduce related finite element space

$$S_{h_m} = \{\psi_{h_m} \in \mathcal{C}^0(\Gamma_{ij}) \text{ such that } \forall T \in \mathcal{T}_{h_m},\ \psi_{h_m} \in P_{k_m}(T)\}$$

and we set the global related space

$$S_h = \{\phi_h = (\phi_i)_{1 \le i \le I} \in S; \forall m = (i,j) \in \mathcal{M}, \phi_{i|\Gamma_{ij}} \in S_{h_m},$$

$$\text{and } \phi_{i\,|\Gamma_i\cap\Gamma} = 0 \text{ if meas}(\Gamma_i \cap \Gamma) \neq 0\}$$

The discrete problem states then as,

$$\text{Find } (\dot{u}_h, \phi_h) \in \dot{W}_h \times S_h \text{ such that} \tag{5}$$

$$\dot{B}((\dot{u}_h, \phi_h), (\dot{v}_h, \psi_h)) = \sum_{i=1}^{I} \int_{\Omega_i} f v_{h_i} dx, \quad \forall (\dot{v}_h, \psi_h) \in \dot{W}_h \times S_h. \tag{6}$$

$$\tag{7}$$

The functional space $\dot{W} \times S$ is equipped with the norm

$$\forall (\dot{v}, \psi) \in \dot{W} \times S, \quad \|(\dot{v}, \psi)\|^2 = \sum_{i=1}^{I} (|v_i|^2_{1,\Omega_i} + \delta_i \|v_i - \psi_i\|^2_{0,\Gamma_i}).$$

In the sequel, C is a generic constant independent of $\dot{h} = (h_i)_{i=1}^I$ and $\dot{\delta} = (\delta_i)_{i=1}^I$.

Lemma *The bilinear form $\dot{B}$ is continuous and coercive with respect to the $\dot{W} \times S$ norm in the following sense,*

$$\forall (\dot{v}_h, \psi_h) \in \dot{W}_h \times S_h, \quad \dot{B}((\dot{v}_h, \psi_h), (\dot{v}_h, \psi_h)) = \|(\dot{v}_h, \psi_h)\|^2;$$

$\forall ((\dot{v}, \psi), (\dot{w}_h, \zeta_h)) \in (\dot{W} \times S) \times (\dot{W}_h \times S_h)$,

$$\dot{B}(((\dot{v}, \psi), (\dot{w}_h, \zeta_h)) \leq C \|(\dot{w}_h, \zeta_h)\| \{\|((\dot{v}, \psi)\|^2 + \sum_{i=1}^{I} \frac{1}{\delta_i} \|\frac{\partial v_i}{\partial n_i}\|^2_{0,\Gamma_i} + \frac{1}{h_i} \|v_i - \psi_i\|^2_{0,\Gamma_i}\}^{\frac{1}{2}}.$$

♢

If u is the weak solution of the model problem (1) and $\phi = (\phi_i = u_{|\partial\Omega_i})_{i=1}^I$. such that $\dot{u} = (u_i := u_{|\Omega_i})_{i=1}^I \in \dot{W}$, then,

$$\forall (\dot{w}_h, \psi_h) \in \dot{W}_h \times S_h, \;\; \dot{B}((\dot{u}, \phi), (\dot{w}_h, \psi_h)) = \sum_{i=1}^{I} \int_{\Omega_i} f w_h dx.$$

It is a trivial consequence of Lax-Milgram lemma that the discrete problem (5) has the unique solution $(\dot{u}_h, \phi_h) \in \dot{W}_h \times S_h$. Moreover by standard arguments and for $\delta_i = \dfrac{1}{h_i}, i = 1, \ldots, I$,

$$\|(\dot{u} - \dot{u}_h, \phi - \phi_h)\| \leq C \inf_{(\dot{v}_h, \zeta_h) \in \dot{W}_h \times S_h} \{\|((\dot{u} - \dot{v}_h, \psi - \zeta_h)\|^2 + \sum_{i=1}^{I} h_i \|\frac{\partial (u_i - v_{h_i})}{\partial n_i}\|^2_{0,\Gamma_i}\}^{\frac{1}{2}}.$$

The main result states then as,

Theorem *Let $(\dot{u}_h, \phi_h) \in \dot{W}_h \times S_h$ be the solution of discrete problem (2), u be the weak solution of the model problem and $\phi = (\phi_i := u_{|\partial\Omega_i})_{1 \leq i \leq I} \in S$. We assume that $\dot{u} = (u_i :=$*

$u_{|\Omega_i})_{1\le i\le I} \in \prod_{i=1}^{I} H^{\sigma_i}(\Omega_i)$, $\dfrac{3}{2} < \sigma_i \le k_i + 1,\ i = 1, ..., I$, *and for all* $m = (i,j) \in \mathcal{M}$, $\phi_m := \phi_{i|\Gamma_{ij}} \in H^{\sigma_m}(\Gamma_{ij})$, $\frac{1}{2} \le \sigma_m \le k_m + 1$. *Assume moreover that*

$$\forall m = (i,j) \in \mathcal{M}, \quad h_m \le C\ min\,(h_i, h_j),$$

and

$$\forall i = 1, ..., I, \ \ \delta_i = \frac{C}{h_i}.$$

Then the following estimate holds,

$$||\dot{u} - \dot{u}_h, \phi - \phi_h||^2 \le C\{\sum_{i=1}^{I} h_i^{2(\sigma_i - 1)} \|u_i\|^2_{\sigma_i,\Omega_i} + \sum_{m=(i,j)\in\mathcal{M}} h_m^{2\sigma_m - 1} \|\phi_m\|^2_{\sigma_m,\Gamma_{ij}}\}.$$

♢

The proof of the theorem requires the following technical lemma, given as an appendix.

Lemma *Let* T *be a regular triangle (tetrahedron), and* e *an edge (face) of* T. *For all* $v \in H^{1+\sigma}(T)$ *with* $1/2 < \sigma \le 1$, *we have*

$$h_T^{\frac{1}{2}} \|\frac{\partial v}{\partial \nu_e}\|_{0,e} \le C\,(h_T^{\sigma} |v|_{1+\sigma,T} + |v|_{1,T}).$$

Proof As usually, let $\hat{T}$ be a reference triangle (tetrahedron), and $F(\hat{x}) = B\hat{x} + b$, the affine application defined from $\hat{T}$ onto T such that $F(\hat{T}) = T$. First, we have

$$\|\frac{\partial v}{\partial \nu_e}\|_{0,e} \le \|\nabla v\|_{0,e},$$

then

$$\|\frac{\partial v}{\partial \nu_e}\|_{0,e} \le (\frac{\text{meas}(\hat{e})}{\text{meas}(e)})^{\frac{1}{2}} \|B^{-1}\|\ \ \|\nabla \hat{v}\|_{0,\hat{e}}$$

with $\hat{v} = v \circ F$ and $\hat{e} = F^{-1}(e)$.
Using the trace theorem applied to $\nabla \hat{v}$ on $\hat{e}$, we obtain

$$\text{meas}(e)^{\frac{1}{2}} \|\frac{\partial v}{\partial \nu_e}\|_{0,e} \le C\|B^{-1}\|(|\hat{v}|_{1,\hat{T}} + |\hat{v}|_{1+\sigma,\hat{T}}).$$

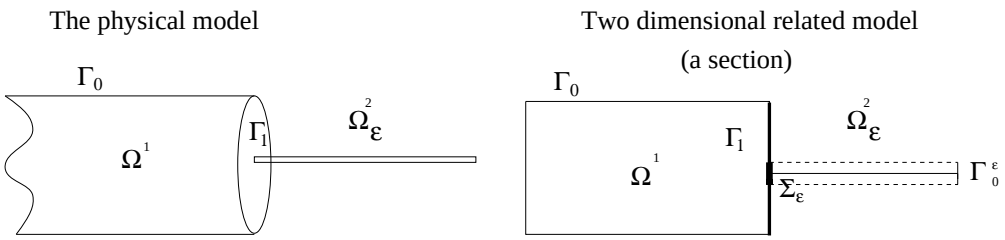
Then

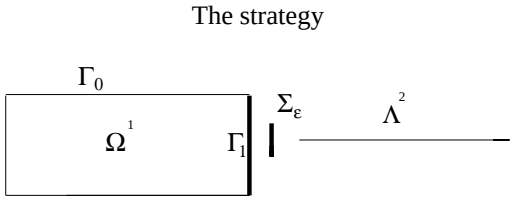
$$\text{meas}(e)^{\frac{1}{2}} \|\frac{\partial v}{\partial \nu_e}\|_{0,e} \le C\|B^{-1}\|\ \ \|B\|\ \ |detB|^{-\frac{1}{2}} (|v|_{1,T} + \|B\|^{\sigma} (\|B\|^{d}\ |detB|^{-1})^{s} |v|_{1+\sigma,T})$$

with $s = 0$ if $\sigma = 1$ and $s = \dfrac{1}{2}$ otherwise.
Since T is regular, we obtain the required inequality,

$$h_T^{\frac{1}{2}} \|\frac{\partial v}{\partial \nu_e}\|_{0,e} \le C\,(h_T^{\sigma} |v|_{1+\sigma,T} + |v|_{1,T}).$$

♢





2 Application to heterogeneous domain decomposition

Let us turn now to the motivation of this study. This methodology has been first investigated for the treatmant of models of elastic multi-structures. Consider the junction of two elastic bodies Ω^1 and Ω^2_ϵ, with $\epsilon << 1$. In Ω^2_ϵ the model is a 2-dimensional model derived from a thin 3-dimensional linearly elastic plate using variational methods [SA99]. In 2D the related model reduces to a formulation on the union of a macro-element Ω^1, a patch element Σ_ϵ and a one-dimensional element Λ^2. Internal domain decomposition can be performed on each element.

Since the methodology is intended for PDEs arising from general elastic multi-structures models, we present it here for the Laplace equation.

We consider the two dimensional (a section) global model problem,

$$\left\{\begin{array}{rcll} -\Delta u^\epsilon & = & f & \text{in } \Omega^1 \cup \Omega^2_\epsilon \\ u^\epsilon & = & 0 & \text{on } \Gamma_0 \\ \dfrac{\partial u^\epsilon}{\partial n} & = & 0 & \text{on } \Gamma_1 \setminus \Sigma_\epsilon \\ u & = & 0 & \text{on } \Gamma^\epsilon_0 \\ \dfrac{\partial u^\epsilon}{\partial n} & = & 0 & \text{on } \partial\Omega^2_\epsilon \setminus \{\Sigma_\epsilon \cup \Gamma^\epsilon_0\} \end{array}\right.$$

where f is no longer dependent on the y-variable in the domain Ω^2_ϵ. The asymptotic problem (the strategy) states as

$$\left\{\begin{array}{rcll} -\Delta u & = & f & \text{in } \Omega^1 \\ u & = & 0 & \text{on } \Gamma_0 \\ \dfrac{\partial u}{\partial n} & = & 0 & \text{on } \Gamma_1 \end{array}\right. \qquad \left\{\begin{array}{rcll} -w^{"} & = & g & \text{on } \Lambda^2 = (0,1) \\ w(1) & = & 0 & \end{array}\right.$$

$$\left\{ \begin{array}{rcl} u_{|_{\Sigma_\epsilon}} & = & w(0) \\ w'(0) & = & \dfrac{1}{\epsilon} \displaystyle\int_{\Sigma_\epsilon} \frac{\partial u}{\partial n}\, d\sigma \end{array} \right.$$

Let us set the adapted hybrid functional framework : the space
$\dot{W} = \{(u,w) \in\ H^1(\Omega^1) \times H^1(\Lambda^2)\ ;\ u_{|_{\Gamma_0}} = 0,\ w(1) = 0,\ \dfrac{\partial u}{\partial n}_{|_{\Sigma_\epsilon}} \ \in\ L^2(\Sigma_\epsilon)\}$,
equipped with the norm $\|(u,w)\|_\delta^2 = (|u|_{1,\Omega^1}^2 + \epsilon|w|_{1,\Lambda^2}^2 + \delta\|u - w(0)\|_{0,\Sigma_\epsilon}^2$
and the bilinear form $B_\delta((u,w),(v,z)) = \displaystyle\int_{\Omega^1} \nabla u \nabla v dx\ +\ \epsilon \int_{\Lambda^2} w'\ z'\ dx\ -$
$\displaystyle\int_{\Sigma_\epsilon} \frac{\partial u}{\partial n}(v - z(0))\ d\sigma\ +\ \int_{\Sigma_\epsilon} \frac{\partial v}{\partial n}(u - w(0))\ d\sigma\ +\ \delta \int_{\Sigma_\epsilon} (u - w(0))(v - z(0))\ d\sigma.$
And the adapted hybrid formulation states then as

$$\text{Find } (u,w) \in \dot{W} \text{ such that} \tag{8}$$

$$B_\delta((u,w),(v,z)) = \int_{\Omega^1} f\ v\ dx\ +\ \epsilon \int_{\Lambda^2} g\ z\ d\sigma \quad \forall (v,z) \in \dot{W} \tag{9}$$

Since $w \in H^2(\Lambda^2)$, if $u \in H^{\sigma+1}(\Omega^1),\ 0 < \sigma \leq 1$, the analysis carried in this context with minor adaptation for a standard P_1- finite element discretization as performed in the previous section and $\delta = \dfrac{1}{h_1}$, gives the following error estimate,

$$\|(u - u_h, w - w_h)\|_\delta \leq\ C\ ({h_1}^\sigma\ |u|_{1+\sigma,\Omega^1}\ +\ h_2\ |w|_{2,\Lambda^2})$$

with the constant C independent of ϵ.
Since w is the opproximation of u^ϵ on Ω_ϵ^2, it is clear that if the error
$\|w(0) - u^\epsilon(0,.))\|_{1/2,\Sigma^\epsilon}$ is small, then the error $|u - u^\epsilon|_{1,\Omega^1}$ is also small. This is due to the fact that $e = u - u^\epsilon$ is the weak solution of the following elliptic equation

$$\left\{ \begin{array}{rcll} -\Delta e & = & 0 & \text{in } \Omega^1 \\ e & = & 0 & \text{on } \Gamma_0 \\ \dfrac{\partial e}{\partial n} & = & 0 & \text{on } \Gamma_1 \setminus \Sigma_\epsilon \\ e & = & w(0) - u^\epsilon(0,.) & \text{on } \Sigma_\epsilon \end{array} \right.$$

More precisely, we have

$$|u - u^\epsilon|_{1,\Omega^1} \leq C\|w(0) - u^\epsilon(0,.))\|_{1/2,\Sigma_\epsilon}$$

where C is a constant independent of ϵ.
The following plots of the solution on the one dimensional subdomain Λ^2 illustrate this remark.

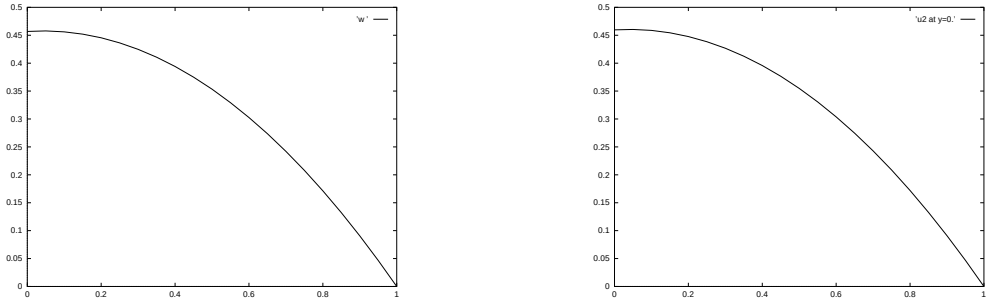


Figure 1: Plots of the solution w and $u_\epsilon(.,0)$for different values of ϵ, on subdomain Λ^2.

References

[Arn82]Douglas N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19:742–760, 1982.

[BFMR97]Franco Brezzi, Leopoldo P. Franca, Luisa D. Marini, and Alessandro Russo. Stabilization techniques for domain decomposition methods with nonmatching grids. In Petter E. Bjørstad, Magne Espedal, and David Keyes, editors, *Domain Decomposition Methods in Sciences and Engineering*. John Wiley & Sons, 1997. Proceedings from the Ninth International Conference, June 1996, Bergen, Norway.

[Cia78]Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[JO98]I. Babuška & C.E. Baumann J.T. Oden. A discontinuous hp finite element method for diffusion problems. *J. of Comp. Phys.*, 146:491 – 519, 1998.

[SA99]R.S. Falk & A.L. Madureira S.M. Alessandrini, D.N. Arnold. *Derivation and Justification of Plate Models by Variational Methods*, volume 21 of *CRM Proc. Lecture Notes*. A.M.S. Providence, 1999.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

14 The Singular Complement Method

f. Assous[1], P. Ciarlet, Jr.[2], S. Labrunie[3], and S. Lohrengel[4]

Introduction

In this paper, we propose a method, called the *Singular Complement Method* (latter referred to as the SCM), which allows to solve PDEs, such as the Laplace problem, Maxwell's equations, etc., in a non-smooth and non-convex domain. In order to define the SCM, let us recall first some basic ingredients of Domain Decomposition Methods (or DDM).

Consider the variational problem (with obvious notations)
find $u \in V$ *such that*

$$a(u,v) = l(v), \quad \forall v \in V. \tag{1}$$

In order to solve it, one can use a DDM, which generally consists in splitting the Hilbert space V into the sum of K subspaces

$$V = V_1 + V_2 + \cdots + V_K, \tag{2}$$

and then getting the solution u of (1), *via* some solves of subproblems such as *find* $u_i \in V_i$ *such that*

$$a_i(u_i,v_i) = l_i(v_i), \quad \forall v_i \in V_i, \quad 1 \leq i \leq K. \tag{3}$$

This can be achieved iteratively or not. The aim is primarily to reduce the overall amount of work, necessary to compute a good numerical approximation of the solution. When the discretization of the problem is achieved by a Finite Element Method (or FEM), one usually obtains the splitting (2) with the help of a partition of the mesh.

The philosophy of the SCM is different, although the tools are similar: the idea is still to split the space V, but with respect to *regularity*.
Indeed, elements of V belong to the scale of Sobolev spaces $H^\alpha(\Omega)$, or $H^\alpha(\Omega)^n$, where $\Omega \subset \mathbb{R}^n$ is the computational domain, and $\alpha \in \mathbb{R}^+$. Interestingly, for a given space V, the supremum $\alpha_{\min}$ of all possible values of the exponent α, depends on the convexity of the domain and on the smoothness of its boundary. Let α_0 be the supremum when the domain is convex, or smooth. When the domain is non-convex and non-smooth, $\alpha_{\min} < \alpha_0$ usually holds.
Then, let $V_R = V \cap H^{\alpha_0}(\Omega)$ (or $V_R = V \cap H^{\alpha_0}(\Omega)^n$ for vector fields) be the space of *regular* elements. Assume that V_R is *closed* in V, and let

$$V = V_R \oplus V_S \tag{4}$$

[1]CEA-DAM/DIF, BP12 - 91680 Bruyères-le-Châtel, France.
[2]ENSTA, 32, boulevard Victor, 75739 Paris Cedex 15, France.
[3]Univ. Henri Poincaré Nancy I, 54506 Vandœuvre-lès-Nancy Cedex, France.
[4]Univ. de Nice Sophia-Antipolis, Parc Valrose, 06108 Nice Cedex 2, France.

with V_S the space of *singular* elements. The sum is *direct*; in addition, it can be *orthogonal*. When the domain is convex or smooth, one has $V_S = \{0\}$ by definition. Then, regular elements are approximated by a *Lagrange FEM*, whereas elements of V_S are computed in a manner, which depends on the problem to solve: in other words, the idea behind the SCM is to *enlarge* the space of test-functions. Basically, it is designed to achieve the following results:

- Improve the *convergence rate* (for the Laplace problem),
- Capture numerically the *real* solution (for Maxwell's equations).

In what follows, we shall introduce, in Section 1, the SCM for the Laplace problem and for Maxwell's equations in a polyhedron. We describe the main theoretical results that are required to solve electromagnetic problems and, in particular, we emphasize the strong links between the singular elements for both problems. In Sections 2 and 3, we present the theory, and the numerical tools, which we have developed, to solve the static, time-harmonic and time-dependent Maxwell equations in a polygon of $\mathbb{R}^2$, or in an axisymmetric domain of $\mathbb{R}^3$.

1 The problems in a polyhedron

Let Ω be a bounded, simply connected, Lipschitz polyhedron, Γ its connected boundary, $(\Gamma_k)_{1\le k\le F}$ the set of faces, and $\mathbf{n}$ the unit outward normal to Γ.
The L^2-scalar product is denoted by $(f,g)_0$, the associated norm by $\|f\|_0$. We shall use the differential operators div, $\mathbf{curl}$ and the related 'non-standard' Sobolev spaces and norms

$$
\begin{aligned}
&\mathbf{L}^2(\Omega) := \{\mathbf{v} = (v_1, v_2, v_3)^T : \ v_i \in L^2(\Omega),\ 1 \le i \le 3\},\\
&\|\mathbf{v}\|_0 := \left(\|v_1\|_0^2 + \|v_2\|_0^2 + \|v_3\|_0^2\right)^{1/2};\\
&\mathbf{H}(\mathrm{div}\,,\Omega) := \{\mathbf{v} \in \mathbf{L}^2(\Omega) : \ \mathrm{div}\,\mathbf{v} \in L^2(\Omega)\},\\
&\|\mathbf{v}\|_{0,\mathrm{div}} := \left(\|\mathbf{v}\|_0^2 + \|\mathrm{div}\,\mathbf{v}\|_0^2\right)^{1/2};\\
&\mathbf{H}(\mathbf{curl}\,,\Omega) := \{\mathbf{v} \in \mathbf{L}^2(\Omega) : \ \mathbf{curl}\,\mathbf{v} \in \mathbf{L}^2(\Omega)\},\\
&\|\mathbf{v}\|_{0,\mathbf{curl}} := \left(\|\mathbf{v}\|_0^2 + \|\mathbf{curl}\,\mathbf{v}\|_0^2\right)^{1/2};\\
&\mathbf{H}(\mathbf{curl}\,,\mathrm{div}\,,\Omega) := \mathbf{H}(\mathbf{curl}\,,\Omega) \cap \mathbf{H}(\mathrm{div}\,,\Omega),\\
&\|\mathbf{v}\|_{0,\mathbf{curl}\,,\mathrm{div}} := \left(\|\mathbf{v}\|_0^2 + \|\mathbf{curl}\,\mathbf{v}\|_0^2 + \|\mathrm{div}\,\mathbf{v}\|_0^2\right)^{1/2}, \text{ and}
\end{aligned}
$$

$$
|\mathbf{v}|_{\mathbf{curl}\,,\mathrm{div}} := \left(\|\mathbf{curl}\,\mathbf{v}\|_0^2 + \|\mathrm{div}\,\mathbf{v}\|_0^2\right)^{1/2}.
$$

In addition, the usual Sobolev spaces for vector fields shall be written $\mathbf{H}^s(\Omega)$, and $\mathbf{H}^s(\Gamma)$. Then, let us recall that fields of $\mathbf{H}(\mathrm{div}\,,\Omega)$ (resp. $\mathbf{H}(\mathbf{curl}\,,\Omega)$) have a normal trace (resp. tangential components) on Γ, which belongs to $H^{-1/2}(\Gamma)$ (resp. $\mathbf{H}^{-1/2}(\Gamma)$); this allows to define the subspaces with the vanishing corresponding trace, and

$$
\mathcal{X} := \mathbf{H}_0(\mathbf{curl}\,,\Omega) \cap \mathbf{H}(\mathrm{div}\,,\Omega), \quad \mathcal{Y} := \mathbf{H}(\mathbf{curl}\,,\Omega) \cap \mathbf{H}_0(\mathrm{div}\,,\Omega).
$$

Let us state the WEBER inequality, which stems from the compact embedding results of Weber [Web80].

Proposition 1 *In $\mathcal{X}$ and $\mathcal{Y}$, the semi-norm $|\cdot|_{\mathbf{curl}\,,\mathrm{div}}$ is a norm, which is equivalent to the full norm.*

Last, let us mention that one can generalize what we state below, to the case of a Lipschitz *curvilinear* polyhedron, by using the work of Costabel *et al.* [CDN99].

1.1 The Laplace problem

The model problem is, given $f \in L^2(\Omega)$, solve
find $\phi \in H_0^1(\Omega)$ *such that*

$$-\Delta\phi = f \text{ in } \Omega. \tag{5}$$

The regularity of the solution depends on the geometry of the domain [Gri85, Dau88]. Let us call *minimal regularity* of the solution the supremum of the set

$$\{\alpha \in \mathbb{R} : \forall f \in L^2(\Omega),\ \phi \text{ solution of (5) belongs to } H^\alpha(\Omega)\}.$$

Theorem 1 *If Ω is convex, the minimal regularity is $\alpha_0 = 2$.*
If Ω is non-convex, the minimal regularity is $\alpha_{\min} = 3/2 + \sigma$, with $0 < \sigma < 1/2$ depending on the geometry, i.e. conical angles at reentrant vertices, dihedral angles at reentrant edges.

In the non-convex case, by minimal regularity, we mean that all solutions ϕ belong to $H^{3/2+\sigma-\varepsilon}(\Omega)$, for any $\varepsilon > 0$, and that some do not belong to $H^{3/2+\sigma}(\Omega)$.

If one discretizes (5) with the P_1 Lagrange FEM, with h as the meshsize, there holds by standard analysis

Corollary 1 *If Ω is convex, the convergence rate in H^1-norm is in $O(h)$.*
If Ω is non-convex, the convergence rate in H^1-norm is in $O(h^{1/2+\sigma-\varepsilon})$, $\forall\varepsilon$.

Remark 1 *Here, it is crucial to impose $f \in L^2(\Omega)$. If f is only in $H^{-1}(\Omega)$, the regularity of ϕ can be as low as $\phi \in H^1(\Omega)$, for Ω convex or not: the convergence rate is undetermined, and there are no methods that allow to improve it.*

To improve the convergence rate, one can think of: mesh refinement, the (Dual) Singular Function Method, multigrid methods [Bre98], the SCM, etc.

The mesh refinement techniques are well-kwown [Wah91]. So are the (Dual) Singular Function Methods (or (D)SFM), which work in 2d domains, see for instance [Gri85, Gri92]. They are based on the adjunction of test-functions, the (dual) singular functions, to the space of FE.
The SCM is based on the same idea, as mentioned in the Introduction. Its origin (2d case) can be traced back to Moussaoui [Mou84]. More precisely, let

$$\Phi := \{\phi \in H_0^1(\Omega) :\ \Delta\phi \in L^2(\Omega)\}, \text{ and } \Phi_R := \Phi \cap H^2(\Omega).$$

One has the

Theorem 2 *In Φ, $\|\phi\|_\Phi := \|\Delta\phi\|_0$ is a norm, which is equivalent to the graph norm $\phi \mapsto \{\|\phi\|_1^2 + \|\Delta\phi\|_0^2\}^{1/2}$. As a consequence, $\|\cdot\|_\Phi$ is equivalent to $\|\cdot\|_2$ in Φ_R.*

Proof : Thanks to the POINCARÉ inequality, the graph norm is equivalent to $\{|\phi|_1^2+\|\Delta\phi\|_0^2\}^{1/2} = \|\mathbf{grad}\,\phi\|_{0,\mathbf{curl},\mathrm{div}}$, with $\mathbf{grad}\,\phi$ in $\mathcal{X}$. Now, one infers from the WEBER inequality that it is also equivalent to $\|\Delta\phi\|_0$.
To prove the other half, let $\phi \in \Phi_R$. There holds

$$\|\phi\|_\Phi^2 = |\mathbf{grad}\,\phi|_{\mathbf{curl},\mathrm{div}}^2 = \|\mathbf{grad}\,(\mathbf{grad}\,\phi)\|_0^2 = |\phi|_2^2,$$

where the second equality has been obtained by Costabel *et al* [Cos91, CDN99], as $\mathbf{grad}\,\phi$ belongs to $\mathbf{H}^1(\Omega)$ and has vanishing tangential components on Γ. Finally, one can use the first part of the Theorem to conclude. ■

Corollary 2 *$\Delta\Phi_R$ is closed in $L^2(\Omega)$.*

Starting from this result, one can first define its orthogonal, called N:

$$L^2(\Omega) = \Delta\Phi_R \overset{\perp}{\oplus} N, \tag{6}$$

and then Φ_S, the inverse image of N. By construction, both Φ_R and Φ_S are closed in Φ and so $\Phi = \Phi_R \overset{\perp}{\oplus} \Phi_S$, i.e. (4). Now, following [AC97], it is possible to characterize elements of N, and, as consequence, elements of Φ_S.

Theorem 3 *An element p of $L^2(\Omega)$ belongs to N if and only if*

$$\Delta p = 0 \text{ in } \Omega, \quad p_{|\Gamma_k} = 0 \text{ in } H_{00}^{-1/2}(\Gamma_k),\ 1 \le k \le F.$$

(Recall that $H_{00}^{1/2}(\Gamma_k) := \{f \in H^{1/2}(\Gamma_k) : \rho_k^{-1/2} f \in L^2(\Gamma_k)\}$, where ρ_k denotes the distance to the boundary of Γ_k; $H_{00}^{-1/2}(\Gamma_k)$ is the dual space of $H_{00}^{1/2}(\Gamma_k)$.)
As for the numerical computation of elements of N and Φ_S, see the next Section for problems in axisymmetric domains and §3.2 for problems in 2d. Let us mention that in the 2d case (see [Mou84]), one gets results similar to those of the DSFM, that is, the recovery of an overall convergence rate in $O(h)$ in H^1-norm.

Let us conclude this Subsection by some extensions.

The first one is the *homogeneous Neumann problem*, for which the same theory can be developed in $\Psi/\mathbb{R}$, where
$\Psi := \{\psi \in H^1(\Omega) : \partial_n\psi_{|\Gamma} = 0 \text{ on } \Gamma,\ \Delta\psi \in L^2(\Omega)\}$.

Another one is about the scalar *wave equation* which, given $T > 0$, reads
find $\phi(t) \in H_0^1(\Omega)$ such that

$$\frac{\partial^2\phi}{\partial t^2} - \Delta\phi = f \text{ in } \Omega\times]0,T[, \quad \phi(0) = \phi_0. \tag{7}$$

The theory of Lions and Magenes [LM72] leads to

Theorem 4 *Assume that $f \in L^2(0,T;H_0^1(\Omega))$ and $\phi_0 \in \Phi$. Then, there exists one and only one solution of the wave equation (7), with regularity*

$$\phi \in \mathcal{C}^0(0,T;\Phi) \cap \mathcal{C}^1(0,T;H_0^1(\Omega)).$$

Next, from (4) applied to Φ, there comes the *continuous decomposition in time* of the solution, that is the

Corollary 3 *One can write $\phi(t) = \phi_R(t) + \phi_S(t)$ for all t, with*

$$(\phi_R, \phi_S) \in \mathcal{C}^0(0,T;\Phi_R \times \Phi_S).$$

Finally, one could use the same kind of idea for a *non-homogeneous boundary condition*, provided that the data is smooth enough on Γ, or for problems with jumps.

1.2 Mathematical tools for Maxwell's equations

We consider the electromagnetic fields in vacuum, enclosed by a perfectly conducting material. The electric permittivity and magnetic permeability are set to one. The electromagnetic field is denoted by $(\mathcal{E}, \mathcal{B})$. The sets of equations are :
The *time-dependent* Maxwell equations in $(\mathcal{E}, \mathcal{B})$:

$$\begin{cases} \partial_t \mathcal{E} - \mathbf{curl}\,\mathcal{B} = -\mathcal{J},\ \partial_t \mathcal{B} + \mathbf{curl}\,\mathcal{E} = 0 \text{ in } \Omega\times]0,T[, \\ \operatorname{div}\mathcal{E} = \rho,\ \operatorname{div}\mathcal{B} = 0 \text{ in } \Omega\times]0,T[, \\ \mathcal{E}\times\mathbf{n} = 0 \text{ on } \Gamma\times]0,T[, \\ \mathcal{E}(0) = \mathcal{E}_0,\ \mathcal{B}(0) = \mathcal{B}_0. \end{cases} \tag{8}$$

The *time-harmonic* Maxwell equations on $\mathcal{E}$, a *complex-valued* field:

$$\begin{cases} \mathbf{curl}\,\mathbf{curl}\,\mathcal{E} - k^2\mathcal{E} = \mathcal{J} \text{ in } \Omega, \\ \operatorname{div}\mathcal{E} = 0 \text{ in } \Omega, \\ \mathcal{E}\times\mathbf{n} = 0 \text{ on } \Gamma. \end{cases} \tag{9}$$

The *static* Maxwell equations, with $\mathcal{U}$ being either the electrostatic or the magnetostatic field:

$$\begin{cases} \mathbf{curl}\,\mathcal{U} = \mathcal{F} \text{ in } \Omega, \\ \operatorname{div}\mathcal{U} = \mathcal{G} \text{ in } \Omega, \\ \mathcal{U}\times\mathbf{n} = 0 \text{ on } \Gamma, \text{ or } \mathcal{U}\cdot\mathbf{n} = 0 \text{ on } \Gamma. \end{cases} \tag{10}$$

Unless otherwise specified, we consider that (10) is the electrostatic problem.

Let us say a few words on the existence and uniqueness of the solution of each problem (cf. [AC00, BHL99, CZ97], in this order).

Theorem 5 The time-dependent problem.
Assume that $\mathcal{J} \in \mathcal{C}^0(0,T;\mathbf{H}(\operatorname{div},\Omega)) \cap H^1(0,T;\mathbf{L}^2(\Omega))$ and $\rho \in \mathcal{C}^1(0,T;L^2(\Omega))$. Then, there exists one and only one solution $(\mathcal{E},\mathcal{B})$ of (8), with

$$(\mathcal{E},\mathcal{B}) \in \mathcal{C}^0(0,T;\mathcal{X}\times\mathcal{Y}) \cap \mathcal{C}^1(0,T;\mathbf{H}(\operatorname{div},\Omega)\times\mathbf{H}(\operatorname{div},\Omega)).$$

The time-harmonic problem.
Assume that $\mathcal{J}$ belongs to $\mathbf{H}(\operatorname{div},\Omega)$, $\operatorname{div}\mathcal{J} = 0$, and $Im(k) \neq 0$. Then, there exists one and only one solution $\mathcal{E}$ solution of (9) in the 'complexified' $\mathcal{X}$.

The static problem.
Assume that $\mathcal{F}$ belongs to $\mathbf{H}_0(\operatorname{div},\Omega)$, with $\operatorname{div}\mathcal{F} = 0$, and that $\mathcal{G}$ is in $L^2(\Omega)$. Then, there exists one and only one solution $\mathcal{E}$ solution of (10) in $\mathcal{X}$.

Here, we considered that the data is L^2-regular. Actually, this is *equivalent* to the assumption that we made previously for the Laplace problem, i.e. that the Laplacian of the solution is in $L^2(\Omega)$.

As for the regularity of the solution, one finds again that it depends on the geometry of the domain [ABDG98]: let us consider, for instance, the static field $\mathcal{U}$.

Theorem 6 *If Ω is convex, the minimal regularity is $\alpha_0 = 1$.*
If Ω is non-convex, the minimal regularity is $\alpha_{\min} = 1/2 + \sigma$.

In the case of Maxwell's equations, we thus let

$$\mathcal{X}_R := \mathcal{X} \cap \mathbf{H}^1(\Omega), \text{ and } \mathcal{Y}_R := \mathcal{Y} \cap \mathbf{H}^1(\Omega).$$

The original idea was to take advantage of the H^1-regularity of the field, when the domain is *convex* [ADH$^+$93], to discretize it by the P_1 Lagrange FEM, instead of the 'usual' edge FEM [Néd80, Néd86]. As a matter of fact, the former includes two key ingredients, which the latter lacks:

- For the time-dependent Maxwell equations, the mass matrix can be lumped, with no loss in precision, thus leading to very inexpensive numerical schemes.

- The numerical electromagnetic field is continuous, so the method can be used in conjunction with a particle-pushing scheme, to solve the coupled Vlasov-Maxwell system of equations.

The question to be answered is: what happens when Ω is a *non-convex domain*? For that, let us begin with the

Theorem 7 *$\mathcal{X}_R$ (resp. $\mathcal{Y}_R$) is closed in $\mathcal{X}$ (resp. $\mathcal{Y}$).*
Therefore, when Ω is non-convex, $\mathcal{X}_R$ (resp. $\mathcal{Y}_R$) is not dense in $\mathcal{X}$ (resp. $\mathcal{Y}$).

Proof : The norm in $\mathcal{X}$ is $|\cdot|_{\mathbf{curl},\mathrm{div}}$. With the help of the formula [Cos91, CDN99]:

$$(\mathbf{grad}\,\mathbf{u}, \mathbf{grad}\,\mathbf{v})_0 = (\mathbf{curl}\,\mathbf{u}, \mathbf{curl}\,\mathbf{v})_0 + (\mathrm{div}\,\mathbf{u}, \mathrm{div}\,\mathbf{v})_0, \ \forall \mathbf{u}, \mathbf{v} \in \mathcal{X}_R \times \mathcal{X}_R, \qquad (11)$$

one gets that the norm in $\mathcal{X}_R$ is equivalent to the $\mathbf{H}^1$-norm, and thus $\mathcal{X}_R$ is closed in $\mathcal{X}$. As a consequence, $\mathcal{X}_R$ is dense in $\mathcal{X}$ iff $\mathcal{X}_R = \mathcal{X}$. According to Theorem 6, this is not the case when Ω is non-convex.
(The proof for $\mathcal{Y}$ and its regular subspace is identical.) ∎

The immediate consequence is that one *can not capture numerically* the solution of the above problems, with the help of the Lagrange FEM only, if the solution is not in the regular space. In particular, mesh refinement techniques *do not work*.
As a matter of fact, let us split $\mathcal{X}$ *à la* (4), $\mathcal{X} = \mathcal{X}_R \overset{\perp}{\oplus} \mathcal{X}_S$, with $\mathcal{X}_S = \mathcal{X}_R^{\perp}$. Is is clear, from the definition of $\mathcal{X}_R$, that any subspace of $\mathcal{X}$ generated by the P_1 Lagrange FEM is actually a subspace of $\mathcal{X}_R$. Thus, with self-explanatory notations, (4) leads to

$$\|\mathcal{E} - \mathcal{E}^h\|_{\mathcal{X}}^2 = \|\mathcal{E}_R - \mathcal{E}^h\|_{\mathcal{X}}^2 + \|\mathcal{E}_S\|_{\mathcal{X}}^2 \geq \|\mathcal{E}_S\|_{\mathcal{X}}^2.$$

Is there a hope of finding an intermediate solution, between the edge FEM, and the P_1 Lagrange FEM? The answer is clearly 'no', if one looks for a piecewise smooth FE (i.e. a FE, whose restriction to each element of the triangulation is smooth), like the edge or Lagrange FEMs. Indeed, it has been remarked by Hazard and Lenoir [HL96] that any $\mathbf{H}(\mathbf{curl}\,, \mathrm{div}\,)$-conforming FEM, with a piecewise smooth FE, is actually $\mathbf{H}^1$-conforming.

Therefore, it is required that one adds the SCM (or the SFM) to be able to compute an approximation of the solution[5]. One discretizes the regular part with the P_1 Lagrange FEM, which means a P_1 approximation component by component, and taking into account the boundary condition. Evidently, this method can be applied to all three Maxwell problems: time-dependent (8), time-harmonic (9) or static (10).
Now, how can one *approximate the singular part*? One possible idea, that we develop further in the other Sections, is to relate the singular electric fields to singular elements of the Laplace operator, i.e. to elements of Φ_S.
Let us conclude this Subsection by displaying this relationship. For that, we need a result, obtained by Birman and Solomyak [BS87].

Theorem 8 *Let Ω be a bounded Lipschitz domain. Then, for all $\mathbf{u}$ in $\mathcal{X}$, there exist $\mathbf{u}_0$ in $\mathcal{X}_R$ and $\phi \in \Phi$ such that*

$$\mathbf{u} = \mathbf{u}_0 + \mathbf{grad}\,\phi, \quad ||\mathbf{u}_0||_1^2 + ||\Delta\phi||_0^2 \leq C\,||\mathbf{u}||_{0,\mathbf{curl}\,,\mathrm{div}\,}^2. \tag{12}$$

Here, C denotes a nonnegative constant, which is independent of $\mathbf{u}$.

In $\mathcal{Y}$, they proved the same result, provided that the domain has a piecewise-smooth boundary [BS87, BS90]. As a consequence, one can prove the

Theorem 9 *The following decomposition is direct and continuous*

$$\mathcal{X} = \mathcal{X}_R \overset{c}{\oplus} \mathbf{grad}\,\Phi_S.$$

Proof : From (12), it is clear that $\mathcal{X} = \mathcal{X}_R + \mathbf{grad}\,\Phi_S$.
Then, let $\mathbf{v} \in \mathcal{X}_R \cap \mathbf{grad}\,\Phi_S$: by construction, $\mathbf{v} \in \mathbf{H}^1(\Omega) \cap \mathbf{grad}\,\Phi$, i.e. $\mathbf{v} \in \mathbf{grad}\,\Phi_R$. Also, one infers from (4) applied to Φ that $\mathbf{grad}\,\Phi$ can be split (in $\mathcal{X}$) into $\mathbf{grad}\,\Phi_R \oplus \mathbf{grad}\,\Phi_S$. So, $\mathbf{v} = 0$, and the sum is direct.
Last, the application

$$\begin{aligned} &\mathcal{X}_R \times \mathbf{grad}\,\Phi_S \to \mathcal{X} \\ &(\mathbf{v}_R, \mathbf{grad}\,\phi_S) \mapsto \mathbf{v} = \mathbf{v}_R + \mathbf{grad}\,\phi_S \end{aligned}$$

is linear, continuous and bijective. Now, as $\mathcal{X}_R \times \mathbf{grad}\,\Phi_S$ and $\mathcal{X}$ are Banach spaces, the open mapping Theorem allows to conclude that the inverse of the application is also continuous. ■

Again, one can prove the same type of result on $\mathcal{Y}$. In other words, the singular electric or magnetic fields are *one-to-one* with the gradients of the singular elements of the Laplacian.

[5] Another alternative is to use the edge FEM, possibly with a specifically designed SCM.

2 Maxwell's equations in an axisymmetric domain

Let Ω be the domain limited by a surface of *revolution* Γ; ω and γ_b their intersections with a meridian half-plane. One has $\gamma := \partial\omega = \gamma_a \cup \gamma_b$, where γ_a is the segment of the axis lying between the extremities of γ_b. ν is its unit outward normal, and τ the unit tangential vector such that (τ, ν) is direct.

Moreover, it is assumed that γ_b is a polygonal line with edges $(\gamma_k)_{1\leq k\leq F}$. The Γ_k are the corresponding faces of Γ; the off-axis corners of γ_b generate circular edges in Γ, whereas the extremities are conical vertices of Γ.

The natural coordinates for this domain are the cylindrical coordinates (r, θ, z), with the basis vectors $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_z)$. A meridian half-plane is defined by the equation $\theta = cst$, and (r, z) are cartesian coordinates in this half-plane.

Definition 10 *For any vector field, the meridian and azimuthal components of* $\mathbf{u}$ *are resp.* $\mathbf{u}_m := \varpi_m(\mathbf{u}) := u_r\,\mathbf{e}_r + u_z\,\mathbf{e}_z$ *and* $\mathbf{u}_\theta := \varpi_\theta(\mathbf{u}) := u_\theta\,\mathbf{e}_\theta$.

We are interested in the case where the sources of the electromagnetic fields, and hence the fields themselves, possess a *symmetry of revolution*. This fact means that the scalar (resp. vector) fields are entirely characterized by their "trace" in ω, i.e. the datum of their value in a meridian half-plane (resp. by the trace of their cylindrical components). Obviously, this is equivalent to the vanishing of all derivatives with respect to θ of these fields or components. In this Section, it is thus assumed that $\partial_\theta \cdot = 0$.

Proposition 2 *For any axisymmetric vector field* $\mathbf{u}$, *the following identities hold:* $\mathbf{curl}\,\mathbf{u}_m = \varpi_\theta(\mathbf{curl}\,\mathbf{u})$, $\mathbf{curl}\,\mathbf{u}_\theta = \varpi_m(\mathbf{curl}\,\mathbf{u})$, $\operatorname{div}\mathbf{u}_m = \operatorname{div}\mathbf{u}$, $\Delta\mathbf{u}_m = \varpi_m(\Delta\mathbf{u})$, $\Delta\mathbf{u}_\theta = \varpi_\theta(\Delta\mathbf{u})$. *Hence, if* $\mathbf{u}$ *is meridian (*$\varpi_\theta(\mathbf{u}) = 0$*),* $\mathbf{curl}\,\mathbf{u}$ *is azimuthal and* $\Delta\mathbf{u}$ *is meridian; if* $\mathbf{u}$ *is azimuthal (*$\varpi_m(\mathbf{u}) = 0$*),* $\mathbf{curl}\,\mathbf{u}$ *is meridian,* $\Delta\mathbf{u}$ *is azimuthal and* $\operatorname{div}\mathbf{u} = 0$,

A similar property holds for the Jacobian of an axisymmetric vector field: there is a decoupling of the meridian and azimuthal components.

Finally, as the meridian and azimuthal components of vector fields are mutually orthogonal pointwise, the same is true in the sense of the $\mathbf{L}^2(\Omega)$ scalar product: for $(\mathbf{u}, \mathbf{v}) \in [\mathbf{L}^2(\Omega)]^2$, there holds $(\mathbf{u}_\theta, \mathbf{v}_m)_{0,\Omega} = 0$. This property is also true for the curl and the vector Laplace operators, or the Jacobian of a field, provided that they belong to $\mathbf{L}^2(\Omega)$.

2.1 Reduction to two-dimensional problems.

Thanks to Proposition 2, it is possible to decouple each of the Maxwell systems (8, 9, 10) into a couple of problems set in $\Omega\times]0,T[$, involving different components of the fields $\mathcal{E}$ and $\mathcal{B}$. Given the expression of differential operators in cylindrical coordinates, these problems read as follows in $\omega\times]0,T[$.

The *time-dependent equations* (8), split into a system of unknowns $(\mathbf{E}_m, B_\theta)$:

$$\left\{\begin{array}{l} \partial_t\mathbf{E}_m - r^{-1}\,\mathbf{curl}\,(r\,B_\theta) = -\mathbf{J}_m,\ \partial_t B_\theta + \operatorname{curl}\mathbf{E}_m = 0 \text{ in } \omega\times]0,T[, \\ r^{-1}\operatorname{div}(r\,\mathbf{E}_m) = \rho \text{ in } \omega\times]0,T[, \quad \mathbf{E}_m\cdot\tau = 0 \text{ on } \gamma_b\times]0,T[, \\ \mathbf{E}_m(0) = \mathbf{E}_{m0},\ B_\theta(0) = B_{\theta 0}. \end{array}\right. \tag{13}$$

and a system of unknowns $(E_\theta, \mathbf{B}_m)$:

$$\left\{\begin{array}{l} \partial_t E_\theta - \operatorname{curl} \mathbf{B}_m = -J_\theta,\ \partial_t \mathbf{B}_m + r^{-1}\,\mathbf{curl}\,(r\,E_\theta) = 0 \text{ in } \omega\times]0,T[, \\ r^{-1}\operatorname{div}(r\,\mathbf{B}_m) = 0 \text{ in } \omega\times]0,T[, \quad \mathbf{B}_m\cdot\nu = 0,\ E_\theta = 0 \text{ on } \gamma_b\times]0,T[, \\ E_\theta(0) = E_{\theta 0},\ \mathbf{B}_m(0) = \mathbf{B}_{m0}. \end{array}\right. \tag{14}$$

The *static equations* (10), split into a system of unknown $\mathbf{U}_m$:

$$\left\{\begin{array}{l} \operatorname{curl}\mathbf{U}_m = F_\theta \text{ in } \omega, \\ r^{-1}\operatorname{div}(r\,\mathbf{U}_m) = \mathcal{G} \text{ in } \omega, \\ \mathbf{U}_m\cdot\tau = 0 \text{ on } \gamma_b, \text{ or } \mathbf{U}_m\cdot\nu = 0 \text{ on } \gamma_b. \end{array}\right. \tag{15}$$

and a system of unknown U_θ:

$$\left\{\begin{array}{l} r^{-1}\,\mathbf{curl}\,(r\,U_\theta) = \mathbf{F}_m \text{ in } \omega, \\ U_\theta = 0 \text{ on } \gamma_b, \text{ for the electrostatic problem only.} \end{array}\right. \tag{16}$$

2.2 Sobolev spaces

We denote by a $\breve{}$ the respective subspaces of axisymmetric vector fields in the various Sobolev spaces, e.g. $\breve{\mathbf{L}}^2(\Omega)$, $\breve{\mathbf{H}}^1(\Omega)$, $\breve{\mathbf{H}}(\mathbf{curl}\,,\operatorname{div}\,,\Omega)$, $\breve{\mathcal{X}}$, $\breve{\mathcal{Y}}_R$; by $\|\cdot\|_{s,\Omega}$ the H^s-norm, by $\|\cdot\|_{0,\mathbf{curl}\,,\operatorname{div}\,,\Omega}$ the $H(\mathbf{curl}\,,\operatorname{div}\,)$-norm.

As pointed out earlier, elements of these spaces are characterized by their trace in ω. We refer to [BDM99] for their full description. For now, we only need the

Definition 11 *For $\alpha \in \mathbb{R}$, let $L^2_\alpha(\omega)$ be the space of square-integrable functions in ω with respect to the measure $r^\alpha\,dr\,dz$, and $H^s_\alpha(\omega)$, for $s \in \mathbb{R}$, the related scale of Sobolev spaces, with the canonical norms $||\cdot||_{s,\alpha,\omega}$.*

2.3 Closedness results.

The aim of this Subsection is to prove the analogue of Theorem 7. Because of the technicalities induced by the geometry [ACL00], it is necessary to distinguish between the *inductive* proof for the electric field and the *constructive* proof for the magnetic field.

We shall now sketch these proofs; details are found in [ACL00].

Lemma 1 *The following inequalities are equivalent:*

$$\exists C_1,\quad \forall \mathbf{u}\in\mathcal{X}_R,\quad ||\mathbf{u}||_{1,\Omega} \leq C_1\,||\mathbf{u}||_{0,\mathbf{curl}\,,\operatorname{div}\,,\Omega}, \tag{17}$$

$$\exists C_2,\quad \forall \phi\in\Phi_R,\quad ||\phi||_{2,\Omega} \leq C_2\,||\Delta\phi||_{0,\Omega}. \tag{18}$$

Proof : For $\mathbf{u} = \mathbf{grad}\,\phi$, (17) implies (18) by POINCARÉ's and WEBER's inequalities. Conversely, (17) stems from (18) and Theorem 8. ∎

Theorem 12 *(17) is satisfied in Ω if and only if all the conical angles at the vertices are different from a prescribed value $\pi/\delta_- \simeq 130°$. This case excluded, $\mathcal{X}_R$ is closed within $\mathcal{X}$.*

Proof : (17) is equivalent to (18). The necessary and sufficient condition for (18) to hold has been established by Dauge [Dau88]. ■

In the following, when considering the *electric* case, we suppose that all conical angles are different from π/δ_-.

Theorem 13 *In $\breve{\mathcal{Y}}_R$, the following estimate holds:*

$$\exists K, \quad \forall \mathbf{u} \in \breve{\mathcal{Y}}_R, \quad \|\nabla \mathbf{u}\|^2_{0,\Omega} \leq K \left(\|\mathbf{curl}\, \mathbf{u}\|^2_{0,\Omega} + \|\mathrm{div}\, \mathbf{u}\|^2_{0,\Omega} \right). \tag{19}$$

Hence, by POINCARÉ *and* WEBER*'s inequalities, the $\|\cdot\|_{1,\Omega}$ and $\|\cdot\|_{0,\mathbf{curl},\mathrm{div},\Omega}$ norms are equivalent on this space, and $\breve{\mathcal{Y}}_R$ is closed within $\breve{\mathcal{Y}}$.*

The equivalent of (11) in Ω reads (cf. [CD99]): for any $(\mathbf{u}, \mathbf{v}) \in \left[\mathbf{H}^2(\Omega)\right]^2$,

$$(\nabla \mathbf{u}, \nabla \mathbf{v})_{0,\Omega} = (\mathbf{curl}\, \mathbf{u}, \mathbf{curl}\, \mathbf{v})_{0,\Omega} + (\mathrm{div}\, \mathbf{u}, \mathrm{div}\, \mathbf{v})_{0,\Omega} - b(\mathbf{u}, \mathbf{v}) + d(\mathbf{u}, \mathbf{v}), \tag{20}$$

where $b(\cdot,\cdot)$ and $d(\cdot,\cdot)$ are bilinear forms defined on the boundary. The term $d(\mathbf{u}, \mathbf{v})$ vanishes when $(\mathbf{u}, \mathbf{v}) \in [\mathcal{Y} \cap \mathbf{H}^2(\Omega)]^2$. It is proven in [ACL00] that this space is dense within $\mathcal{Y}_R$. So one can extend d by 0 to $\mathcal{Y}_R$.

All other terms in (20) are meaningful for $(\mathbf{u}, \mathbf{v}) \in \left[\mathbf{H}^1(\Omega)\right]^2$: for an axisymmetric domain Ω, the bilinear form $b(\mathbf{u}, \mathbf{v})$ is $\displaystyle\int_\Gamma \frac{\nu_r}{r} \left(u_\theta\, v_\theta + u_\nu\, v_\nu\right) d\Gamma$. Hence (20) is valid for $(\mathbf{u}, \mathbf{v}) \in [\mathcal{Y}_R]^2$, with $u_\nu = v_\nu = 0$ on the boundary.

The inequality (19) is now equivalent, thanks to (20), to

$$\exists k < 1, \quad -b(\mathbf{u}, \mathbf{u}) \leq k\, \|\nabla \mathbf{u}\|^2_{0,\Omega}. \tag{21}$$

Since $\nabla \mathbf{u}_m$ and $\nabla \mathbf{u}_\theta$ are $\mathbf{L}^2$-orthogonal, and $b(\mathbf{u}, \mathbf{u})$ depends only on $\mathbf{u}_\theta$, it is sufficient to check (21)—or (19)—for $\mathbf{u} \in E^1_\theta = \left\{ \mathbf{u} \in \breve{\mathbf{H}}^1(\Omega) : \mathbf{u} \parallel \mathbf{e}_\theta \right\}$.

Lemma 2 *The space $H^1_{-1}(\omega)$ is continuously imbedded into $L^2_{-3}(\omega)$, i.e.*

$$\exists K_1, \quad \forall u \in H^1_{-1}(\omega), \quad |||u|||^2_{0,-3,\omega} \leq K_1\, ||\mathbf{grad}\, u||^2_{0,-1,\omega}.$$

This 2d Hardy inequality is obtained by localization and Fubini Theorem.

Proposition 3 *The inequality (19) is satisfied for all $\mathbf{u} \in \breve{\mathcal{Y}}_R$.*

Proof : Let $\mathbf{u} \in E^1_\theta$ and $v = r\, u_\theta$. From the expressions of $\mathbf{curl}\, \mathbf{u}$, $\mathrm{div}\, \mathbf{u}$ and $\nabla \mathbf{u}$ in cylindrical coordinates, it follows: $\|\mathbf{curl}\, \mathbf{u}\|^2_{0,\Omega} + \|\mathrm{div}\, \mathbf{u}\|^2_{0,\Omega} = 2\pi\, \|\mathbf{grad}\, v\|^2_{0,-1,\omega}$ and $\|\nabla \mathbf{u}\|^2_{0,\Omega} = 2\pi \left[\|\mathbf{grad}\, u_\theta\|^2_{0,1,\omega} + \|u_\theta\|^2_{0,-1,\omega} \right]$. The latter norm is equivalent to $\|\mathbf{grad}\, v\|^2_{0,-1,\omega} + \|v\|^2_{0,-3,\omega}$. Thus (19) stems from the above Lemma; it also proves that *any* azimuthal vector field in $\breve{\mathbf{H}}(\mathbf{curl}, \mathrm{div}, \Omega)$ is in $\breve{\mathbf{H}}^1(\Omega)$. ■

2.4 A characterization of singular fields

This Subsection describes the relationship between the singular electric and magnetic fields and scalar singularities of Laplace-like operators.

Electric case. Let $\breve{\mathcal{X}}$ be the natural space of electric fields, and $\breve{\mathcal{X}}_R$ its regular subspace. We derive from Theorem 12 the direct and continuous decomposition

$$\breve{\mathcal{X}} = \breve{\mathcal{X}}_R \overset{c}{\oplus} \mathbf{grad}\, \breve{\Phi}_S. \tag{22}$$

As the elements of $\breve{\Phi}_S$ are characterized by their Laplacian, we will study $\breve{N} = \Delta\breve{\Phi}_S$. For this purpose, we shall adapt the method of [ACS98, ACRS99] and the references therein, with a specific treatment for the conical vertices. To that end, on any face Γ_k, let $\breve{H}(\Gamma_k)$ be the axisymmetric subspace of $H_{00}^{1/2}(\Gamma_k)$.

Lemma 3 *The application γ_1^k, which is the trace on Γ_k of the normal derivative, is continuous and surjective from $\breve{\Phi}_R$ onto $\breve{H}(\Gamma_k)$, and there exists a continuous lifting operator from $\breve{H}(\Gamma_k)$ to $\breve{\Phi}_R$.*

This result allows to prove an integration by parts formula, between elements of $\breve{\Phi}_R$ and elements of the space $D(\Delta,\Omega) := \{g \in L^2(\Omega) :\ \Delta g \in L^2(\Omega)\}$.

Lemma 4 *Let $p \in D(\Delta,\Omega)$ and $u \in \breve{\Phi}_R$. There holds*

$$\int_\Omega (p\,\Delta u - u\,\Delta p)\,d\Omega = \sum_{1\le k\le F} {}_{\breve{H}(\Gamma_k)'}\left\langle p, \gamma_1^k u\right\rangle_{\breve{H}(\Gamma_k)}.$$

The first characterization of $\breve{N}$ follows from the above Lemmas.

Theorem 14 *Let $p \in \breve{L}^2(\Omega)$: p belongs to $\breve{N}$ if and only if*

$$\Delta p = 0 \text{ in } \Omega, \quad p_{|\Gamma_k} = 0 \text{ in } \breve{H}(\Gamma_k)',\ 1 \le k \le F.$$

In a meridian half-plane, the second characterization of elements of $\breve{N}$ is then

Corollary 4 *Let $p \in L^2_1(\omega)$: p belongs to $\breve{N}$ if and only if*

$$\begin{aligned}
&\Delta^+ p := \frac{\partial^2 p}{\partial r^2} + \frac{1}{r}\frac{\partial p}{\partial r} + \frac{\partial^2 p}{\partial z^2} = 0 \text{ in } \omega,\\
&p_{|\gamma_k} = 0,\ 1 \le k \le F,\\
&p \in \mathcal{C}^\infty(\overline{\omega} \setminus \mathcal{V}_b),\ \text{for any neighborhood } \mathcal{V}_b \text{ of } \gamma_b.
\end{aligned}$$

The study of the Laplace-like operator Δ^+ is performed in [ACL01]. It extends Grisvard's work [Gri92] to the axisymmetric case:

Theorem 15 *The space $\breve{N}$, and consequently $\breve{\mathcal{X}}_S$, is of finite dimension, equal to $K_e + K_v$, with K_e the number of reentrant edges, et K_v the number of vertices with conical angle larger than π/δ_-.*

Magnetic case. The *natural space* of axisymmetric magnetic fields is

$$\breve{\mathcal{W}} = \{\mathbf{v} \in \breve{\mathcal{Y}} :\ \mathrm{div}\,\mathbf{v} = 0\},\ \text{with norm } ||\mathbf{curl}\,\mathbf{v}||_{0,\Omega}. \tag{23}$$

Then, if $\breve{\mathcal{W}}_R = \breve{\mathcal{W}} \cap \breve{\mathbf{H}}^1(\Omega)$ is the space of regular fields, we infer from Theorem 13 that $\breve{\mathcal{W}}_R$ is closed in $\breve{\mathcal{W}}$. Let $\breve{\mathcal{W}}_S$ be its orthogonal, i.e.

$$\breve{\mathcal{W}} = \breve{\mathcal{W}}_R \overset{\perp}{\oplus} \breve{\mathcal{W}}_S. \tag{24}$$

We had remarked in the proof of Proposition 3 that an azimuthal field is always regular; hence, the singular fields are meridian. Moreover, elements of $\breve{\mathcal{W}}$ are determined *via* their curl. So, given $\mathcal{B}_S \in \breve{\mathcal{W}}_S$, define $\mathcal{P} = \mathbf{curl}\, \mathcal{B}_S$: $\mathcal{B}_S$ is meridian, therefore $\mathcal{P}$ is azimuthal.

Now, let $\breve{\mathcal{M}}_R$ be the space $\mathbf{curl}^{-1}\breve{\mathcal{W}}_R$ of potentials of elements of $\breve{\mathcal{W}}_R$. The orthogonality, in the sense of (23), of $\mathcal{B}_S$ and elements of $\breve{\mathcal{W}}_R$ becomes

$$(\mathcal{P}, \Delta\mathcal{A})_{0,\Omega} = 0, \quad \forall \mathcal{A} \in \breve{\mathcal{M}}_R$$

as $\Delta = -\mathbf{curl}\,\mathbf{curl} + \mathbf{grad}\,\mathrm{div}$. As $\mathcal{P}$ is azimuthal, it is enough to consider only elements of $\breve{\mathcal{M}}_{\theta R} = \{\mathcal{A} \in \breve{\mathcal{M}}_R : \mathcal{A} \parallel \mathbf{e}_\theta\}$. So, we are left with a scalar problem, similar to the electric case, and we obtain the two characterizations of $\mathbf{curl}\,\breve{\mathcal{W}}_S$.

Theorem 16 *Let $\mathcal{P} \in \breve{\mathbf{L}}^2(\Omega)$, $\mathcal{P} \parallel \mathbf{e}_\theta$; $\mathcal{P} = P_\theta\, \mathbf{e}_\theta$ belongs to $\mathbf{curl}\,\breve{\mathcal{W}}_S$ iff*

$$\Delta\mathcal{P} = 0 \text{ in } \Omega, \quad P_{\theta\,|\Gamma_k} = 0 \text{ in } \breve{H}(\Gamma_k)', \ 1 \le k \le F.$$

Corollary 5 *Let $P_\theta = p/r$: $p \in L^2_{-1}(\omega)$ is characterized as a solution of*

$$\begin{aligned} &\Delta^- p := \frac{\partial^2 p}{\partial r^2} - \frac{1}{r}\frac{\partial p}{\partial r} + \frac{\partial^2 p}{\partial z^2} = 0 \text{ in } \omega, \\ &p_{|\gamma_k} = 0, \ 1 \le k \le F, \\ &p/r \in \mathcal{C}^\infty(\overline{\omega} \setminus \mathcal{V}_b), \text{ for any neighborhood } \mathcal{V}_b \text{ of } \gamma_b. \end{aligned}$$

Again, the study of the operator Δ^- (cf. [ACL01]) gives the equivalent of Grisvard's result [Gri92] in this case:

Theorem 17 *The space defined by Corollary 5, and consequently $\breve{\mathcal{W}}_S$, is of finite dimension, equal to K_e, the number of reentrant edges.*

Now, it is more convenient for numerical computations to use the variable $P = P_\theta$. It satisfies:

$$P \in L^2_1(\omega), \ \Delta' P := \frac{\partial^2 P}{\partial r^2} + \frac{\partial^2 P}{\partial z^2} + \frac{1}{r}\frac{\partial P}{\partial r} - \frac{P}{r^2} = 0 \text{ in } \omega, \ P = 0 \text{ on } \gamma. \tag{25}$$

2.5 Existence and uniqueness results.

If the data and initial conditions are axisymmetric, so are the solutions of (8) and (10), and, under the hypotheses of Theorem 5

$$\mathcal{E} \in \mathcal{C}^0(0,T;\breve{\mathcal{X}}), \quad \mathcal{B} \in \mathcal{C}^0(0,T;\breve{\mathcal{W}}).$$

Then it follows from (22) and (24) that the electromagnetic field can be decomposed into regular and singular parts continuously with respect to time:

$$\begin{aligned} &\mathcal{E}(t) = \mathcal{E}_R(t) + \mathcal{E}_S(t), \quad (\mathcal{E}_R, \mathcal{E}_S) \in \mathcal{C}^0(0,T;\breve{\mathcal{X}}_R \times \breve{\mathcal{X}}_S), \\ &\mathcal{B}(t) = \mathcal{B}_R(t) + \mathcal{B}_S(t), \quad (\mathcal{B}_R, \mathcal{B}_S) \in \mathcal{C}^0(0,T;\breve{\mathcal{W}}_R \times \breve{\mathcal{W}}_S). \end{aligned}$$

Moreover, as the projections ϖ_m and ϖ_θ are smooth, each of the systems (13–16) admits a unique solution in the relevant space; that of (13) and (14) depend continuously on time. As a consequence, the decomposition (4) can be refined by using three subspaces: meridian regular, meridian singular, azimuthal. (Recall that azimuthal implies regular.) Each of the problems (3) then admits a unique and continuous solution.

2.6 Principle of the numerical method.

The SCM follows from the above decomposition. As the singular parts span a finite-dimensional space, it is sufficient to find an approximation of a basis. The problems (3) amount to a classical FE formulation, for the regular parts, and a low-dimensional linear system, for the singular parts.

Computation of bases of $\mathcal{N}$ and $\breve{N}$. We look for a basis of the spaces $\breve{N}$ and $\mathcal{N} := \{P \text{ satisfying } (25)\}$, whose dimensions are given by Theorems 15 and 17. We have at hand an approximate knowledge of these bases [ACL01].

- There is one basis function $p_j^- \in \mathcal{N}$ or $p_j^+ \in \breve{N}$ associated to each relevant geometric singularity A_j as follows. In a neighborhood ω_j of A_j, there holds $p_j^\pm|_{\omega_j} = p_j^S + q_j^\pm$, where the *principal part* p_j^S is just in $L_1^2(\omega_j)$, and the remainder $q_j^\pm$ is of H^1-style regularity in ω_j. In $\omega_j' = \omega \setminus \overline{\omega}_j$, $p_j^\pm$ is of H^1-style regularity.

- In ω_j, define local polar coordinates (ρ_j, ϕ_j) centered at A_j.

- If A_j is a reentrant edge of opening $\beta_j = \pi/\alpha_j$, $1/2 < \alpha_j < 1$, one has $p_j^S = \rho_j^{-\alpha_j} \sin(\alpha_j \phi_j)$, for the electric and magnetic cases.

- (For the electric case only.) If A_j is a conical vertex of opening π/δ_j, $1/2 < \delta_j < \delta_-$, one finds $p_j^S = \rho_j^{-1-\nu_j} P_{\nu_j}(\cos\phi_j)$, where P_ν denotes the Legendre function and $\nu_j \in]0, 1/2[$ is given by $P_{\nu_j}(\cos(\pi/\delta_j)) = 0$.

In the whole of ω the function $q_j^\pm = p_j^\pm - p_j^S$ satisfies

$$-\Delta^+ q_j^+ = \Delta^+ p_j^S, \text{ resp. } -\Delta' q_j^- = \Delta' p_j^S \text{ in } \omega, \quad q_j^\pm\big|_\gamma = -p_j^S\big|_\gamma \text{ on } \gamma, \tag{26}$$

but, unlike in the cartesian geometry, it is not possible to compute it variationally: if A_j is an edge, neither p_j^S nor $q_j^\pm$ is of H^1-style regularity near the axis, and the problem (26) is ill-posed. This hindrance can be overcome:

- either by multiplying p_j^S by the 'not-too-noisy' cut-off function $\eta(r) = r/r(A_j)$, i.e. defining $\widehat{q}_j = p_j - \eta\, p_j^S$ which is regular in the whole of ω;

- or by domain decomposition, computing q_j in ω_j and p_j in ω_j', and enforcing standard transmission conditions between ω_j and ω_j' (*à la* §3.2.2).

Computation of bases of $\breve{\mathcal{W}}_S$ and $\mathbf{grad}\,\breve{\Phi}_S$. Our task is now to compute $\mathcal{B}_j = \mathbf{curl}\,(-\Delta')^{-1} p_j^-$, which is in $\breve{\mathcal{W}}_S$ since $\mathbf{curl}\,\mathcal{B}_j = p_j^-\,\mathbf{e}_\theta$, and $\mathcal{E}_j = \mathbf{grad}\,(-\Delta^+)^{-1} p_j^+$. First, one solves variationally:

$$-\Delta'\psi_j = p_j^- \text{ in } \omega, \quad \psi_j = 0 \text{ on } \gamma,$$
$$-\Delta^+\varphi_j = p_j^+ \text{ in } \omega, \quad \varphi_j = 0 \text{ on } \gamma_b.$$

One has: $\psi_j = \psi_j^R + \sum_{1\le i\le K_e} c_j^i\,\psi_i^S$, $\varphi_j = \varphi_j^R + \sum_{1\le i\le K_e+K_v} d_j^i\,\varphi_i^S$, where:
- the ψ_j^R and φ_j^R are of H^2-style regularity,
- $\psi_i^S = \varphi_i^S = \rho_i^{\alpha_i}\,\sin(\alpha_i\,\phi_i)$ near a reentrant edge,
- $\varphi_i^S = \rho_i^{\nu_i}\,P_{\nu_i}(\cos\phi_i)$ near a vertex of conical angle larger than π/δ_-.

The singularity coefficients c_j^i, d_j^i can be extracted by quadrature formulas [BDM99] or spectral methods. In $\omega_0 = \omega \setminus \bigcup\overline{\omega}_i$, ψ_j and φ_j are regular. The corresponding decompositions of $\mathcal{B}_j$ and $\mathcal{E}_j$ are:

$$\mathcal{B}_j = \mathbf{curl}\,\psi_j^R + \sum_{1\le i\le K_e} c_j^i\,\mathbf{curl}\,\psi_i^S, \quad \mathcal{E}_j = \mathbf{grad}\,\varphi_j^R + \sum_{1\le i\le K_e+K_v} d_j^i\,\mathbf{grad}\,\varphi_i^S.$$

$\mathbf{curl}\,\psi_j^R$ and $\mathbf{grad}\,\varphi_j^R$ are of H^1-style regularity and can be computed variationally, while $\mathbf{curl}\,\psi_i^S$ and $\mathbf{grad}\,\varphi_i^S$ are analytically known. In ω_0, the whole of $\mathcal{B}$ and $\mathcal{E}$ can be computed variationally.

Finally, it is possible to orthogonalize the decomposition (22) by subtracting to $\mathcal{E}_j$ its orthogonal projection on $\breve{\mathcal{X}}_R$. This is no difficulty.

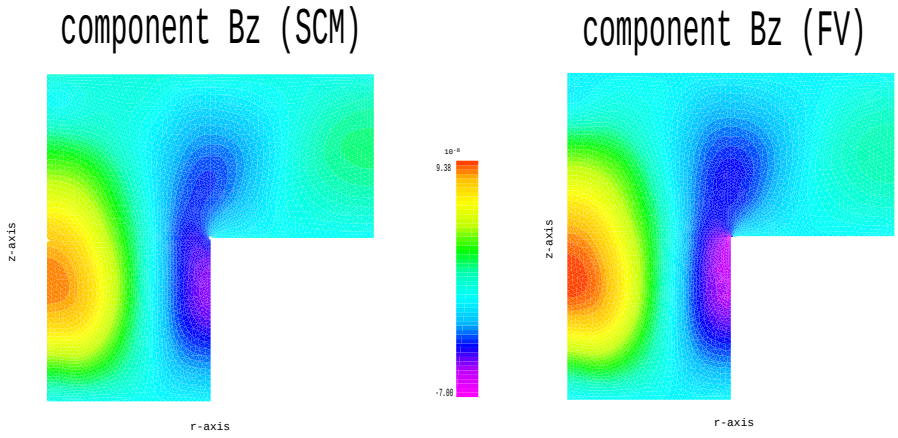


Figure 1: Computed magnetic field: The SCM and Finite Volume techniques.

2.7 Numerical Results

As an illustration of the SCM in the axisymmetric case, one can compute the electromagnetic field generated by a current. A top hat domain Ω (ω L-shaped) is considered, and a perfectly conducting boundary condition is imposed. The initial conditions are set to zero. The electromagnetic wave is generated by a current $\mathcal{J}(\mathbf{x},t) = J_\theta \mathbf{e}_\theta$, $J_\theta = 10\sin(\lambda t)$, with a frequency $\lambda/2\pi = 2,5.10^9$ Hertz. The support of this current is a little disc centered at the middle of the domain. Because it is impossible to provide an analytical solution, we compare our results to the computations made by another code, based on Finite Volume (FV) techniques

à la Delaunay [Her93]. The space and time discretizations of the SCM are detailed later on, in Section 3.2.3. Figure 1 shows the isovalues of the magnetic field (B_z component after 1000 time steps), which have been computed by the two methods. The results obtained by both methods are comparable, which shows the feasability of the SCM. Moreover, the SCM provides a numerical solution which is less noisy.

3 Maxwell's equations in a polygon

In what follows, it is assumed that both the data and the initial conditions *do not depend* on the transverse variable z. Then the original problem can be identified with a problem posed in a section of an infinite cylinder, which is a 2d polygon ω, with boundary γ, a set of edges $(\gamma_k)_{1\leq k\leq E}$, and a unit outward normal ν. The notations are those of Section 1, except that the Sobolev spaces are based on the scalar curl; also the 2d calligraphic spaces ($\mathcal{X}$, $\mathcal{Y}$) and unknowns ($\mathcal{E}$, etc.) are written in *boldface*, i.e. $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{E} = (E_1, E_2)^T$, etc.

3.1 The time-harmonic Maxwell equations

This Section summarizes the results obtained in [BHL99] and [HL00], and we refer to these papers for any detail.

We are looking for a numerical approximation of the solution $\mathbf{E}$ to

$$\left\{ \begin{array}{l} \mathbf{curl}\,\mathrm{curl}\,\mathbf{E} - k^2\mathbf{E} = \mathbf{J} \text{ in } \omega, \\ \mathbf{E}\cdot\tau = 0 \text{ on } \gamma. \end{array} \right. \tag{27}$$

For sake of simplicity regarding existence and uniqueness questions, we suppose that k is a complex number with nonzero imaginary part (which means that we are concerned with the electromagnetic problem in a homogeneous and dissipative medium) or, in order to include stationnary problems, that $k = 0$. The vector field $\mathbf{J}$ is a datum that represents the impressed current density. We assume that

$$\mathrm{div}\,\mathbf{J} = 0 \text{ in } \omega,$$

which amounts to saying that the electric charge density vanishes in the whole domain. The *singular field method* is based on the fact that the solution of (27) can be found by solving an equivalent *regularized* problem similar to the vector Helmholtz equation. Formally, the regularized problem is given by

$$\left\{ \begin{array}{l} -\Delta\mathbf{E} - k^2\mathbf{E} = \mathbf{J} \text{ in } \omega, \\ \mathbf{E}\cdot\tau = 0 \text{ on } \gamma, \\ \mathrm{div}\,\mathbf{E} = 0 \text{ on } \gamma. \end{array} \right. \tag{28}$$

Indeed, a solution of (27) is clearly divergence free and, thus, satisfies (28). Conversely, let $\mathbf{E}$ be a solution of (28). Its divergence $\varphi = \mathrm{div}\,\mathbf{E}$ satisfies

$$\left\{ \begin{array}{l} -\Delta\varphi - k^2\varphi = 0 \text{ in } \omega, \\ \varphi = 0 \text{ on } \gamma, \end{array} \right.$$

which yields $\varphi = 0$ (provided φ is assumed regular enough).

The Section is organized as follows: in a first part (§3.1.1) we make precise the functional setting and give the corresponding variational formulation. In particular, we address the question of equivalence between the classical and the regularized formulations of the problem. We show that the latter can be set in two 'neighboring' functional spaces whenever the domain ω has at least one reentrant corner. Of course, only one of them leads to the equivalence with the classical formulation. The key of the method lies in the fact that the 'right' functional space can be written as the direct sum of a space of regular fields completed by a (finite-dimensional) space of singular fields. We give two possible decompositions which lead to the *singular field method* (SFM) and its orthogonal variant (OSFM) described in §3.1.2. In §3.1.3, the analysis of the convergence of these methods is addressed. It turns out that both numerical schemes have the same rate of convergence but the numerical applications presented in §3.1.4 clearly show that OSFM yields far better results: we shall try to explain why.

3.1.1 Classical and regularized formulations

The variational interpretation of the classical problem (27) leads us naturally to seek $\mathbf{E}$ in the space $\mathbf{H}_0(\mathrm{curl}\,)$. If we assume the datum $\mathbf{J}$ to belong to $\mathbf{L}^2(\omega)$, the weak form of (27) is given by

$$\mathcal{P}_0(\mathrm{curl}\,) \qquad \left\{ \begin{array}{l} \textit{Find } \mathbf{E} \in \mathbf{H}_0(\mathrm{curl}\,) \textit{ such that} \\ (\mathrm{curl}\,\mathbf{E}, \mathrm{curl}\,\mathbf{E}')_0 - k^2(\mathbf{E}, \mathbf{E}')_0 = (\mathbf{J}, \mathbf{E}')_0 \ \forall \mathbf{E}' \in \mathbf{H}_0(\mathrm{curl}\,). \end{array} \right.$$

The sesquilinear form $(\mathrm{curl}\,\mathbf{E}, \mathrm{curl}\,\mathbf{E}')_0 - k^2(\mathbf{E}, \mathbf{E}')_0$ being coercive on $\mathbf{H}_0(\mathrm{curl}\,)$ (due to condition $Im(k) \neq 0$), we infer the existence and uniqueness of the solution of $\mathcal{P}_0(\mathrm{curl}\,)$ from Lax-Milgram's theorem.

Let us now consider the regularized problem (28). Its variational formulation involves the functional space $\mathbf{X}$ and thus amounts simply to adding $(\mathrm{div}\,\mathbf{E}, \mathrm{div}\,\mathbf{E}')_0$ in $\mathcal{P}_0(\mathrm{curl}\,)$. We therefore consider the problem

$$\mathcal{P}_0(\mathrm{curl}\,, \mathrm{div}\,) \qquad \left\{ \begin{array}{l} \textit{Find } \mathbf{E} \in \mathbf{X} \textit{ such that} \\ a(\mathbf{E}, \mathbf{E}') = (\mathbf{J}, \mathbf{E}')_0 \ \forall \mathbf{E}' \in \mathbf{X}, \end{array} \right.$$

where $a(\mathbf{E}, \mathbf{E}') := (\mathrm{curl}\,\mathbf{E}, \mathrm{curl}\,\mathbf{E}')_0 + (\mathrm{div}\,\mathbf{E}, \mathrm{div}\,\mathbf{E}')_0 - k^2(\mathbf{E}, \mathbf{E}')_0$.
For the same reason as above, $\mathcal{P}_0(\mathrm{curl}\,, \mathrm{div}\,)$ has a unique solution which coincides with that of $\mathcal{P}_0(\mathrm{curl}\,)$ provided $\mathrm{div}\,\mathbf{J} = 0$. Indeed, choosing $\mathbf{E}' = \mathbf{grad}\,\varphi'$ with $\varphi' \in \mathcal{D}(\omega)$ in $\mathcal{P}_0(\mathrm{curl}\,)$ yields that the solution of $\mathcal{P}_0(\mathrm{curl}\,)$ is divergence-free. It thus belongs to $\mathbf{X}$ and satisfies the variational equation of $\mathcal{P}_0(\mathrm{curl}\,, \mathrm{div}\,)$, in other words it does coincide with the solution of $\mathcal{P}_0(\mathrm{curl}\,, \mathrm{div}\,)$.

We thus deduce that $\mathbf{X}$ is the appropriate functional frame for the regularized problem. But the situation becomes more involved if we consider the following problem given on the subspace of $\mathbf{X}$ of *regular* fields:

$$\mathcal{P}_0(\mathbf{grad}\,) \qquad \left\{ \begin{array}{l} \textit{Find } \mathbf{E} \in \mathbf{X}_R \textit{ such that} \\ a(\mathbf{E}, \mathbf{E}') = (\mathbf{J}, \mathbf{E}')_0 \ \forall \mathbf{E}' \in \mathbf{X}_R. \end{array} \right.$$

As mentioned in Section 1, $\mathbf{X}_R$ is a closed subspace of $\mathbf{X}$ and hence, the form $a(\cdot, \cdot)$ is still coercive on $\mathbf{X}_R$. Whenever ω has at least one reentrant corner, $\mathbf{X}_R$ is strictly contained in $\mathbf{X}$, and the respective solutions to $\mathcal{P}_0(\mathrm{curl}\,, \mathrm{div}\,)$ and $\mathcal{P}_0(\mathbf{grad}\,)$ are in general different. In particular, an H^1-conforming FE discretization can only provide an approximation of the

non-physical problem $\mathcal{P}_0(\mathbf{grad}\,)$.

In order to perform a method based on nodal (Lagrange) FE, which is able to capture the singular behavior (and thus solves problem $\mathcal{P}_0(\mathrm{curl}\,,\mathrm{div}\,)$), we decompose $\mathbf{X}$ into a regular and a singular part,

$$\mathbf{X} = \mathbf{X}_R \oplus \mathbf{X}_S.$$

Of course, $\mathbf{X}_S$ is not uniquely determined. Hereafter, we will give two possible choices, leading to two different methods.

Notice that the above existence and equivalence results keep true in the stationnary case corresponding to $k = 0$. In order to simplify the presentation, we will consider this case only, and we thus set from now on

$$a(\mathbf{E},\mathbf{E}') := (\mathrm{curl}\,\mathbf{E},\mathrm{curl}\,\mathbf{E}')_0 + (\mathrm{div}\,\mathbf{E},\mathrm{div}\,\mathbf{E}')_0.$$

Let us set some notations. Without loss of generality, assume that ω has exactly one reentrant corner of measure $\beta = \pi/\alpha$, $1/2 < \alpha < 1$, at the vertex S. We use the local polar coordinates (r,θ), and we fix a regular cut-off function $\eta = \eta(r)$ such that $\eta \equiv 1$ near S and $\eta \equiv 0$ near the other vertices. The function

$$s(r,\theta) = r^\alpha \sin(\alpha\theta)$$

belongs to $H^1(\omega) \setminus H^2(\omega)$ as $\alpha < 1$ and is called *singular function* at S. We finally introduce the subspace of $H_0^1(\omega)$ given by

$$\mathcal{S} = \mathrm{span}\{\eta s\}.$$

Owing to Theorem 8 (see also [BHL99]), we have the direct decomposition:

Theorem 18

$$\mathbf{X} = \mathbf{X}_R \overset{c}{\oplus} \mathbf{grad}\,\mathcal{S}. \tag{29}$$

An *orthogonal* decomposition can be deduced from (29) solving an auxiliary inhomogeneous variational problem, which is similar to $\mathcal{P}_0(\mathbf{grad}\,)$:

Theorem 19

$$\mathbf{X} = \mathbf{X}_R \overset{\perp}{\oplus} \mathrm{span}\{\mathbf{grad}\,(s) + \mathbf{F}\} \tag{30}$$

where $\mathbf{F}$ *is the solution of the problem*
find $\mathbf{F} \in \mathbf{H}^1(\omega)$ such that

$$a(\mathbf{F},\mathbf{E}') = 0,\ \forall \mathbf{E}' \in \mathbf{X}_R, \quad \mathbf{F}\cdot\tau = -\mathbf{grad}\,s\cdot\tau \text{ on } \gamma. \tag{31}$$

Remark 2 *Decomposition (30) is orthogonal in the sense that*

$$a(\mathbf{grad}\,(s) + \mathbf{F}, {\mathbf{E}'}_R) = 0\ \forall {\mathbf{E}'}_R \in \mathbf{X}_R.$$

Notice, however, that the above relation fails whenever $k \neq 0$. Nevertheless, in this case, the remaining terms are of lower order and involve only the L^2-scalar product of the sesquilinear form.

3.1.2 Description of the method

We give the algorithms of both SFM and OSFM which are based respectively on the decompositions (29) and (30). To this end, let $(\mathcal{T}_h)_{0<h<h_0}$, be a family of regular triangulations of the domain ω. We consider the P_1 Lagrange FEM:

$$\mathbf{Y}_h := \{\mathbf{E}_h \in \mathbf{H}^1(\omega) :\ \mathbf{E}_{h|T_h} \text{ is affine } \forall T_h \in \mathcal{T}_h\}.$$

Let $\{M_I\}$ be the set of nodal points of the triangulation and

$$\mathbf{V}_h := \{\mathbf{E}_h \in \mathbf{Y}_h :\ (\mathbf{E}_h \cdot \tau)(M_I) = 0,\ \forall M_I \in \gamma\}$$

the discretization space of $\mathbf{X}_R$. Let $N_h = \dim \mathbf{V}_h$ and $(\mathbf{w}_I)_{I=1,\dots,N_h}$ be the basis functions. Note that the discrete boundary condition $(\mathbf{E}_h \cdot \tau)(M_I) = 0$ is ambiguous if M_I is a vertex of the polygon; in this case it should be understood as $\mathbf{E}_h(M_I) = 0$ (i.e. both components of $\mathbf{E}_h(M_I)$ vanish.)

The singular field method (SFM) Owing to (29), the discretization space is given by

$$\mathbf{X}_h = \mathbf{V}_h \oplus \mathbf{grad}\, \mathcal{S}.$$

The matrix form of the discrete problem then reads as follows:

$$\begin{pmatrix} \mathbb{A} & C \\ C^T & \mathbb{A}_s \end{pmatrix} \begin{pmatrix} \vec{\mathsf{E}}_R \\ e_s \end{pmatrix} = \begin{pmatrix} \vec{\mathsf{J}} \\ j_s \end{pmatrix}, \tag{32}$$

where

- $\mathbb{A}$ and $\vec{\mathsf{J}}$ respectively denote the stiffness matrix and the right-hand side corresponding to the FE space $\mathbf{V}_h$,

$$\mathbb{A}_{IJ} = a(\mathbf{w}_J, \mathbf{w}_I),\ I, J = 1, \dots, N_h \text{ and } \mathsf{J}_J = (\mathbf{J}, \mathbf{w}_J)_0,\ J = 1, \dots, N_h.$$

- $\mathbb{A}_s$ and j_s denote the matrix and the right-hand side of order 1 corresponding to the singular field,

$$\mathbb{A}_s = a(\mathbf{grad}\,(\eta s), \mathbf{grad}\,(\eta s)) \text{ and } j_s = (\mathbf{J}, \mathbf{grad}\,(\eta s))_0.$$

- C is a matrix of order $N_h \times 1$ coupling the basis functions of FE-type to the singular field,

$$C_{I1} = a(\mathbf{grad}\,(\eta s), \mathbf{w}_I),\ 1 \le I \le N_h.$$

In order to preserve the advantages of the sparse matrix A_{FE} in the resolution of (32), the SFM consists in solving separately the two linear systems

$$\mathbb{A}\vec{\mathsf{E}}^\star = \vec{\mathsf{J}},\ \mathbb{A}S = C,$$

and taking into account that (32) may be written as

$$\mathbb{A}\vec{\mathsf{E}}_R = \mathbb{A}\left(\vec{\mathsf{E}}^\star - e_s S\right), \quad \mathbb{A}_s\, e_s = j_s - C^T \vec{\mathsf{E}}_R. \tag{33}$$

The left equality clearly implies that $\vec{\mathsf{E}}_R = \vec{\mathsf{E}}^\star - e_s S$. Substituting this identity into the right one thus yields the singular coefficient e_s. Thus,

$$\mathbf{E}_h = \sum_{1 \le l \le N_h} (\mathsf{E}_I^\star - e_s S_I)\, \mathbf{w}_I + e_s \mathbf{grad}\,(\eta s).$$

Notice the similarity of (33) with (3).

The orthogonal singular field method (OSFM) This time, the discretization space is given by

$$\widetilde{\mathbf{X}_h} = \mathbf{V}_h \oplus \mathrm{span}\{\mathbf{grad}\,(s) + \mathbf{F}_h\},$$

where $\mathbf{F}_h$ denotes the FE-approximation of problem (31).

In consequence, the matrix form of the discrete problem is the following:

$$\begin{pmatrix} \mathbb{A} & 0 \\ 0 & \widetilde{\mathbb{A}_s} \end{pmatrix} \begin{pmatrix} \vec{\mathsf{E}}'_R \\ e'_s \end{pmatrix} = \begin{pmatrix} \vec{\mathsf{J}} \\ \widetilde{j_s} \end{pmatrix},$$

where $\mathbb{A}$ and $\vec{\mathsf{J}}$ take the same significance as before, and $\widetilde{\mathbb{A}_s}$ and $\widetilde{j_s}$ are respectively given by

$$\widetilde{\mathbb{A}_s} = a(\mathbf{F}_h, \mathbf{F}_h) \text{ and } \widetilde{j_s} = (\mathbf{J}, \mathbf{grad}\,(s) + \mathbf{F}_h)_0,$$

taking into account that $\mathbf{grad}\,(s)$ is curl- and divergence-free.

The algorithm of the OSFM is then straightforward.

Remark 3 *Both methods can be extended to the case of K_c reentrant corners: $\vec{\mathsf{e}}_s$ is a vector of $\mathbb{R}^{K_c}$, and C and S (for the SFM) are matrices of order $N_h \times K_c$. See also §3.2.3, in which algorithms are given in this case.*

3.1.3 Error analysis

We state in this Section the main convergence results. All proofs may be found in [HL00].

Theorem 20 *Let $\mathbf{E}$ be the solution of $\mathcal{P}_0(\mathrm{curl}\,, \mathrm{div}\,)$ and $\mathbf{E}_h$ its approximation by the SFM. Assume that the regular part of $\mathbf{E}$ belongs to $\mathbf{H}^{s+1}(\omega)$ with s in $]0, 1]$. Then, we have*

$$\|\mathbf{E} - \mathbf{E}_h\|_{0,\mathrm{curl},\mathrm{div}} = \mathcal{O}(h^s)$$

for the error in the energy norm, and

$$\|\mathbf{E} - \mathbf{E}_h\|_0 = \mathcal{O}(h^\lambda),\ \forall \lambda < s + 2\alpha - 1,$$

in the L^2-norm. Moreover, the error of the OSFM is of the same order as the one of the SFM.

3.1.4 Numerical results

In this Section, we present numerical tests of both methods in the case where the exact solutions are known. The domain is formed by three quarters of a circle with center $\mathbf{0}$ and radius 2, the only reentrant corner being of measure $\beta = 3\pi/2$ ($\alpha = 2/3$). We consider two families of solutions, for $n \in \mathbb{N}$:

$$\mathbf{G}_n(r, \theta) = \mathbf{grad}\,(\eta(r) r^{n\alpha} \sin(n\alpha\theta)) \text{ and } \mathbf{H}_n(r, \theta) = \mathbf{curl}\,(\eta(r) r^{n\alpha} \cos(n\alpha\theta)),$$

$\mathbf{G}_n$ and $\mathbf{H}_n$ have the same regularity depending on n: $\mathbf{G}_n \in \mathbf{H}^s(\omega), \forall s < n\alpha$. In particular, $\mathbf{G}_n$ is of class $\mathbf{H}^1$ for $n > 1$, whereas $\mathbf{G}_1$ has a *non-zero* component in any complementary space of $\mathbf{X}_R$.

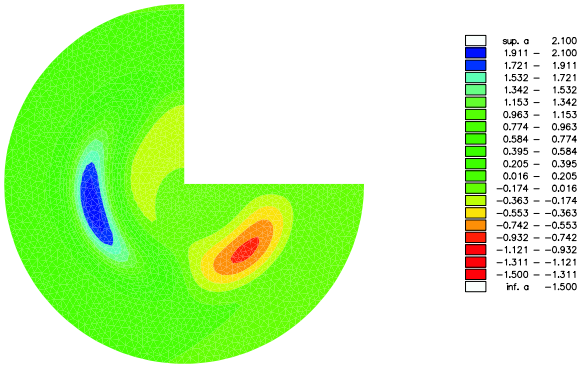


Figure 2: FE-approximation of $\mathbf{G}_1$.

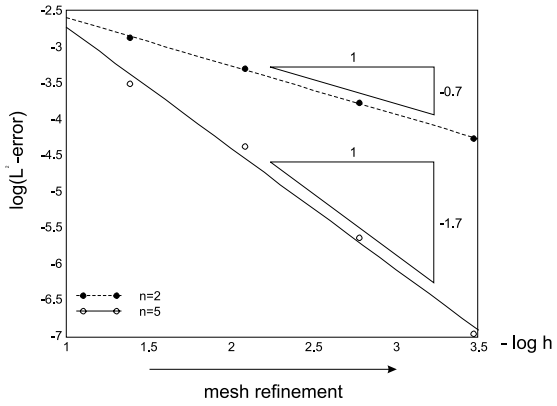


Figure 3: The SFM for $\mathbf{H}_n$, $n = 2, n = 5$.

Both methods have been tested on four unstructured grids. The mesh parameter h varies from $h = 2^{-1}$ to $h = 2^{-4}$, the latter corresponding to roughly 2×8.500 degrees of freedom. Notice that no particular mesh refinement has been done near the corner. The cut-off function η is a piecewise polynomial function of class $\mathcal{C}^3$. The coefficients of the terms $\mathbb{A}$, $\vec{\mathrm{J}}$ and C are calculated using a 7-point-quadrature formula (exact for polynomials up to order 5). The coefficients $\mathbb{A}_s$ and j_s are calculated analytically. The implementation of the boundary condition is realized *via* a rotation which maps the canonical basis on a local basis of the normal and tangential vectors; in the latter basis the vector boundary condition is decoupled and standard techniques apply. The linear systems occuring in the algorithms are solved by a direct method based on Cholesky factorization. All tests have been realized with the FE-code MELINA[6].

It may be clearly seen on Figure 2 that the standard FEM fails for a singular solution field

[6]developed by O. Debayser (ENSTA, Paris, France) and D. Martin (IRMAR, University of Rennes 1, France) at SMP, ENSTA, see [Mar97].

(here, we represent the x-component of the FE-approximation of $\mathbf{G}_1$). Indeed, the condition $\mathbf{E}_h \cdot \tau_{|\gamma} = 0$ forces the FE-approximation to vanish at $\mathbf{0}$ whereas the exact solution tends to ∞ at the corner. Hence, we are not faced with an accuracy problem (which could be handled alternatively by a *local* mesh refinement), but with the choice of the appropriated functional frame: the FE-aproximation converges to the solution of $\mathcal{P}_0(\mathbf{grad})$ which is *globally* different from the physical solution.

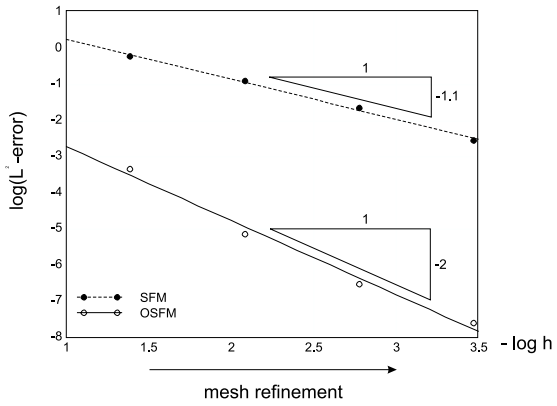


Figure 4: the SFM/OSFM for $\mathbf{H}_1$.

Figure 3 shows the discrete L^2-error of the SFM,

$$\|\mathbf{E} - \mathbf{E}_h\|_h := \left(\frac{1}{N_h} \sum_{I \in \overset{\circ}{\omega}} |\mathbf{E}(M_I) - \mathbf{E}_h(M_I)|^2 \right)^{1/2},$$

in logarithmic scale for the regular fields $\mathbf{H}_2$ and $\mathbf{H}_5$.

The numerical values are in good accordance with the theory of §3.1.3. Figure 4 compares the SFM- and OSFM-approximations of the singular field $\mathbf{H}_1$. It turns out that the OSFM yields the better results. This is probably due to the cut-off function η involved in the implementation of SFM. Indeed, this numerical instability is known for *singular function methods* (see for example [BD82]) and leads to high values of the constant in the error estimates, and thus to poor accuracy.

3.2 The time-dependent Maxwell equations

We are looking now for a numerical approximation of the 2d time-dependent Maxwell equations, (8) being rewritten as two decoupled sets of second order in time equations. In this paper, we focus on the first one (the second one could be written in the same way [ACS98]).

It can be written as follows:

$$\begin{cases} \dfrac{\partial^2 \mathbf{E}}{\partial t^2} + \mathbf{curl}\,\mathrm{curl}\,\mathbf{E} = -\dfrac{\partial \mathbf{J}}{\partial t}, \quad \dfrac{\partial^2 B_z}{\partial t^2} - \Delta B_z = \mathrm{curl}\,\mathbf{J} \text{ in } \omega\times]0,T[, \\ \mathrm{div}\,\mathbf{E} = \rho \text{ in } \omega\times]0,T[, \\ \mathbf{E}\cdot\tau = 0, \quad \dfrac{\partial B_z}{\partial \nu} - \mathbf{J}\cdot\tau = 0 \text{ on } \gamma\times]0,T[, \\ \mathbf{E}(0) = \mathbf{E}_0, \quad B_z(0) = B_{z0}, \\ \dfrac{\partial \mathbf{E}}{\partial t}(0) = \mathbf{curl}\, B_{z0} - \mathbf{J}(\cdot,0), \quad \dfrac{\partial B_z}{\partial t}(0) = -\mathrm{curl}\,\mathbf{E}_0. \end{cases}$$

(The second order in time system of equations is closed with the help of initial conditions on $\partial_t \mathbf{E}$ and $\partial_t B_z$.)

As mentioned in Section 1, the B_z component, as the solution of a wave equation, always belongs to $H^1(\omega)$, even in a non-convex domain. As a consequence, we consider below only the computation of the field $\mathbf{E}$.

Remark 4 *For the sake of simplicity, the problem will be written in the absence of charges:* $\mathrm{div}\,\mathbf{E} = 0$. *The space of solutions becomes*

$$\mathbf{V} = \{\mathbf{v} \in \mathbf{X} : \mathrm{div}\,\mathbf{v} = 0\}$$

By using the Helmholtz decomposition, it can be proved that the singular space $\mathbf{X}_S$ *of* $\mathbf{X}$ *is a (strict) subspace of* $\mathbf{curl}\,\Phi_S + \mathbf{grad}\,\Psi_S$, *where* Φ_S *is the space introduced in Section 1, and* Ψ_S *its counterpart for the homogeneous Neumann problem. Hence, the method described here for* $\mathbf{V}$ *can be adapted to* $\mathbf{X}$.

3.2.1 Description of the method

We first introduce a variational form of the equations, i.e.
find $\mathbf{E}(t) \in \mathbf{V}$ *such that*

$$\frac{d^2}{dt^2}(\mathbf{E},\mathbf{F})_0 + (\mathrm{curl}\,\mathbf{E}, \mathrm{curl}\,\mathbf{F})_0 = -\frac{d}{dt}(\mathbf{J},\mathbf{F})_0 \quad \forall \mathbf{F} \in \mathbf{V},$$

with the same initial conditions. As in the 3d case (see Theorem 5), there exists one and only one solution of this problem. Moreover, we have the following orthogonal decomposition of $\mathbf{V}$, analogous to the one previously obtained in $\mathcal{X}$.

Theorem 21 *The space* $\mathbf{V}$ *can be split in the orthogonal sum* $\mathbf{V} = \mathbf{V}_R \overset{\perp}{\oplus} \mathbf{V}_S$.

From this splitting, we obtain a continuous (orthogonal) decomposition in time of the electric field, that is

$$\mathbf{E}(t) = \mathbf{E}_R(t) + \mathbf{E}_S(t)\ .$$

By using again the relation between the singular solutions of Maxwell's equations and those of the Laplace problem, we obtain that the vector space $\mathbf{V}_S$ is finite dimensional, of dimension

K_c, the number of reentrant corners, defined by $\mathrm{curl}\,\mathbf{V}_S = N$ (N introduced at (6)). For $(\mathbf{v}_S^j)_{1\leq j\leq K_c}$ a basis of $\mathbf{V}_S$, we have

$$\mathbf{E}(t) = \mathbf{E}_R(t) + \sum_{1\leq j\leq K_c} \kappa_j(t)\,\mathbf{v}_S^j,$$

where $(\kappa_j)_{1\leq j\leq K_c}$ are K_c functions at least continuous in time. With this decomposition, the variational formulation becomes:
find $\mathbf{E}_R \in \mathbf{V}_R$ *such that*

$$\frac{d^2}{dt^2}(\mathbf{E}_R, \mathbf{F}_R)_0 \quad + \quad (\mathrm{curl}\,\mathbf{E}_R, \mathrm{curl}\,\mathbf{F}_R)_0 = -\frac{d}{dt}(\mathbf{J}, \mathbf{F}_R)_0 - \sum_{1\leq j\leq K_c} \kappa_j''(t)\,(\mathbf{v}_S^j, \mathbf{F}_R)_0, \quad \forall \mathbf{F}_R \in \mathbf{V}_R\,, \tag{34}$$

completed with K_c equations, derived by using $(\mathbf{v}_S^i)_{1\leq i\leq K_c}$ as K_c test functions. Thanks to the orthogonality of regular and singular fields, one gets:

$$\frac{d^2}{dt^2}(\mathbf{E}_R, \mathbf{v}_S^i)_0 + \sum_{1\leq j\leq K_c} \kappa_j''(t)(\mathbf{v}_S^j, \mathbf{v}_S^i)_0 \quad + \quad \kappa_i(t)(\mathrm{curl}\,\mathbf{v}_S^j, \mathrm{curl}\,\mathbf{v}_S^i)_0 = -\frac{d}{dt}(\mathbf{J}, \mathbf{v}_S^i)_0\,, 1 \leq i \leq K_c. \tag{35}$$

In order to compute numerically the solution, we have first to determine a basis of $\mathbf{V}_S$, and then to solve the time-dependent formulation.

3.2.2 Determination of a basis of $\mathbf{V}_S$

For the sake of simplicity, let us assume that K_c is equal to 1. To compute $\mathbf{v}_S$, a basis of $\mathbf{V}_S$, the isomorphism between $\mathbf{V}_S$ and N is used. The framework of the algorithm is then:

- Compute a basis of N, i.e. a non vanishing element p_S of $L_0^2(\omega)$, such that

$$\Delta p_S = 0 \text{ in } \omega, \quad \frac{\partial p_S}{\partial \nu} = 0 \text{ on } \gamma_k,\ 1 \leq k \leq E.$$

- Compute $\mathbf{v}_S \in \mathbf{V}$, the solution of

$$\mathrm{curl}\,\mathbf{v}_S = p_S \text{ in } \omega, \quad \mathrm{div}\,\mathbf{v}_S = 0 \text{ in } \omega, \quad \mathbf{v}_S \cdot \tau = 0 \text{ on } \gamma. \tag{36}$$

Instead of solving (36), it is more practical to make use of another isomorphism, in the same spirit as in Section 1: to $\mathbf{v}_S \in \mathbf{V}_S$, there corresponds one and only one scalar potential $\phi_S \in H^1(\omega)/\mathbb{R}$ such that

$$-\Delta\phi_S = p_S \text{ in } \omega, \quad \frac{\partial \phi_S}{\partial \nu} = 0 \text{ on } \gamma.$$

Now, as ϕ_S is sufficiently smooth (i.e. with regularity H^1), one can easily solve this problem with the help of a variational formulation. The computation of $\mathbf{v}_S \in \mathbf{V}_S$ then stems from the

identity $\mathbf{v}_S = \mathbf{curl}\,\phi_S$.

Computation of p_S (ϕ_S, $\mathbf{v}_S$): first method
A partition of ω into ω_c and ω_e is introduced, where ω_c stands for an open angular sector of radius R centered at the reentrant corner, with an angle $\beta = \pi/\alpha$, $1/2 < \alpha < 1$, and where ω_e is the open domain such that $\omega_c \cap \omega_e = \emptyset$ and $\overline{\omega}_c \cup \overline{\omega}_e = \overline{\omega}$. Last, Let γ_c (resp. γ_e) denote the boundary of ω_c (resp. ω_e), which is split in $\mathcal{B} \cup \tilde{\gamma_c}$ (resp. $\mathcal{B} \cup \tilde{\gamma_e}$), with the interface $\mathcal{B} = \gamma_c \cap \gamma_e$.

The computation of p_S (for instance) can be divided in three substeps (cf. [ACS98]).

1. The restriction of p_S to ω_c, p_S^c, can be written using the polar coordinates,
$$p_S^c(r,\theta) = \sum_{n \geq -1} A_n r^{n\alpha} \cos(n\alpha\theta), \text{ with } A_{-1} \neq 0\ .$$
Every A_n can be written as an integral of $p^c_{S|\mathcal{B}}$ over $\mathcal{B}$.

2. Let ν^c denote the unit outward normal to ω_c. One then defines the capacitance operator $T : p^c_{S|\mathcal{B}} \mapsto \dfrac{\partial p_S^c}{\partial \nu^c}_{|\mathcal{B}}$, by
$$T(p_S^c) = T_1(p_S^c) - 2\alpha \frac{A_{-1}}{R^{\alpha+1}} \cos(\alpha\theta),$$
$$\text{where} \qquad T_1(p_S^c) = \frac{2\alpha^2}{\pi R} \sum_{n \geq 1} n \left\{ \int_0^\beta p_S^c(R,\theta') \cos(n\alpha\theta')\, d\theta' \right\} \cos(n\alpha\theta).$$

3. With the help of the transmission conditions: $p_S^e = p_S^c$ and $\partial_{\nu^c} p_S^e = \partial_{\nu^c} p_S^c$ on $\mathcal{B}$, one gets the missing boundary condition for the exterior problem (on the interface). Let ν^e denote the unit outward normal to ω_e, the exterior problem, written in a variational form, reads
find $p_S^e \in H^1(\omega_e)/\mathbb{R}$ such that
$$\int_{\omega_e} \nabla p_S^e \cdot \nabla q\, d\omega + \int_{\mathcal{B}} T_1(p_S^e) q\, d\sigma = \frac{2\alpha A_{-1}}{R^{\alpha+1}} \int_{\mathcal{B}} \cos(\alpha\theta) q\, d\sigma,\ \forall q \in H^1(\omega_e)/\mathbb{R}.$$
Clearly, the bilinear form $(p,q) \mapsto \displaystyle\int_{\mathcal{B}} T_1(p) q\, d\sigma$ is symmetric positive. Thus, for a given A_{-1}, the above exterior problem is well-posed.

The computation of ϕ_S and $\mathbf{v}_S$ can be carried out in the same way.

Computation of p_S (ϕ_S, $\mathbf{v}_S$): second method
Instead of partitioning ω into ω_c and ω_e, one can split p_S the basis of N into
$$p_S = p_S^{reg} + s^\star(r,\theta)$$
where $s^\star(r,\theta) = r^{-\alpha}\cos(\alpha\theta)$ is the *dual singular function* (see Section 3.1) for the Neumann problem, and p_S^{reg} the regular part of the solution (that belongs here in $H^1(\omega)$). To compute p_S, we have only to solve the problem in p_S^{reg}
$$-\Delta p_S^{reg} = \Delta s^\star (= 0) \text{ in } \omega,$$
$$\frac{\partial p_S^{reg}}{\partial \nu} = 0 \text{ on } \tilde{\gamma_c}, \quad \frac{\partial p_S^{reg}}{\partial \nu} = -\frac{\partial s^\star}{\partial \nu} \text{ on } \tilde{\gamma_e}.$$

Remark 5 *This second one only requires the knowledge of the dual singular function, which is easier to get than the complete local solution, and should carry out to 3d problems. Moreover its implementation is simpler in the case of several reentrant corners.*

3.2.3 Solution to the time-dependent problem

We consider here the case of K_c reentrant corners. One proceeds first a *semi-discretization in space*, by using the P_1 Lagrange FEM. Let $(\mathbf{w}_I)_{I,1,\cdots,N_h}$ be a basis of $\mathbf{V}_R^h$, the FE approximation space of $\mathbf{V}_R$. The formulation (34) can be written equivalently as a linear system, where $'$ stands for the derivative in time:

$$\mathbb{M}_\omega\ \vec{\mathsf{E}}_R''\ +\ \mathbb{R}_\omega\ \vec{\mathsf{E}}_R\ = -\mathbb{M}_\omega\ \vec{\mathsf{J}}'\ -\ \sum_{1\le j\le K_c} \kappa_j''(t)\ \vec{\Lambda}_j\ , \tag{37}$$

where $\mathbb{M}_\omega$ is the mass matrix, $\mathbb{R}_\omega$ is the curl matrix, and $\vec{\Lambda}_j$ (for a fixed j) the vector whose components are $(\Lambda_j)_I = (\mathbf{v}_S^j, \mathbf{w}_I)_0$, $1 \le I \le N_h$.
We denote by $\vec{\mathsf{k}}(t)$ the vector of $\mathbb{R}^{K_c}$ whose components are $\kappa_j(t)$. Starting from (35), we obtain

$$(\vec{\mathsf{e}}^s)'' + \mathbb{V}_s\vec{\mathsf{k}}'' + \mathbb{P}_s\vec{\mathsf{k}} = -(\vec{\mathsf{j}}^s)'$$

where $\vec{\mathsf{e}}^s$ and $\vec{\mathsf{j}}^s$ are vectors of $\mathbb{R}^{K_c}$, with components

$$\mathsf{e}_j^s = (\mathbf{E}_R, \mathbf{v}_S^j)_0 = (\vec{\mathsf{E}}_R|\vec{\Lambda}_j) \text{ and } \mathsf{j}_j^s = (\mathbf{J}, \mathbf{v}_S^j)_0 = (\vec{\mathsf{J}}|\vec{\Lambda}_j).$$

$\mathbb{V}_s$ and $\mathbb{P}_s$ are $K_c \times K_c$ matrices, defined by $(\mathbb{V}_s)_{ij} = (\mathbf{v}_S^i, \mathbf{v}_S^j)_0$ and $(\mathbb{P}_s)_{ij} = (p_S^i, p_S^j)_0$. By plugging this expression in (37), one obtains

$$\mathbb{M}_\omega\ \vec{\mathsf{E}}_R''\ +\ \mathbb{R}_\omega\ \vec{\mathsf{E}}_R\ = -\mathbb{M}_\omega\ \vec{\mathsf{J}}'\ +\ \sum_{1\le j\le K_c} \{{\mathbb{V}_s}^{-1}((\vec{\mathsf{j}}^s)' + \mathbb{P}_s\vec{\mathsf{k}} + (\vec{\mathsf{e}}^s)'')\}_j\vec{\Lambda}_j\ ,$$

which is implicit in $\vec{\mathsf{E}}_R''$. After a *time discretization* involving a second-order explicit (leap-frog) scheme, the scheme reads

$$\mathbb{M}_\omega\ \vec{\mathsf{E}}_R^{n+1}\ -\ \sum_{1\le j\le K_c} \{{\mathbb{V}_s}^{-1}(\vec{\mathsf{e}}^s)^{n+1}\}_j\vec{\Lambda}_j\ = \vec{\mathsf{G}}^n\ .$$

Here the superscript ^{n} (resp. $^{n+1}$) stands for a variable at time $t^n = n\Delta t$ (resp. t^{n+1}), and $\vec{\mathsf{G}}^n$ is a set of quantities known at time t^n. After a few elementary algebraic manipulations, this expression can be written as

$$(\mathbb{M}_\omega\ -\ \sum_{1\le j\le K_c} \vec{\mathsf{U}}_j\vec{\Lambda}_j^T)\ \vec{\mathsf{E}}_R^{n+1}\ = \vec{\mathsf{G}}^n\ , \tag{38}$$

where $\vec{\mathsf{U}}_j$ is a linear combination of the $(\vec{\Lambda}_k)_{1\le k\le K_c}$: $\vec{\mathsf{U}}_j = \sum_{1\le k\le K_c}({\mathbb{V}_s}^{-1})_{kj}\vec{\Lambda}_k$. It can be solved (for instance) with the help of the following formula (see [Hag89] for a review), $\mathbb{A}$ $N \times N$, $\mathbb{U}$ and $\mathbb{W}$ $N \times K_c$

$$(\mathbb{A} - \mathbb{U}\mathbb{W}^T)^{-1} = \mathbb{A}^{-1} + \mathbb{A}^{-1}\mathbb{U}(\mathbb{I} - \mathbb{W}^T\mathbb{A}^{-1}\mathbb{U})^{-1}\mathbb{W}^T\mathbb{A}^{-1}$$

that only requires (compared to the unmodified system, that is $\mathbb{M}_\omega\ \vec{\mathsf{E}}_R^{n+1} = \vec{\mathsf{G}}^n$) the additional computation of the small $K_c \times K_c$ matrix $(\mathbb{I} - \mathbb{W}^T \mathbb{A}^{-1} \mathbb{U})^{-1}$. This formula is applied here for $\mathbb{A} = \mathbb{M}_\omega$. Recall that the mass matrix $\mathbb{M}_\omega$ is diagonalized thanks to a quadrature formula (see [ADH$^+$93]), which preserves the accuracy. In this way, the linear system to solve (38) appears as a slight modification compared to the one obtained without the SCM.

3.2.4 Numerical results

Results of the computation of a basis of $\mathbf{V}_S$ are similar to those shown in §3.1.4 and will not be presented here. We refer the reader to [ACS00] for more detailed numerical examples.

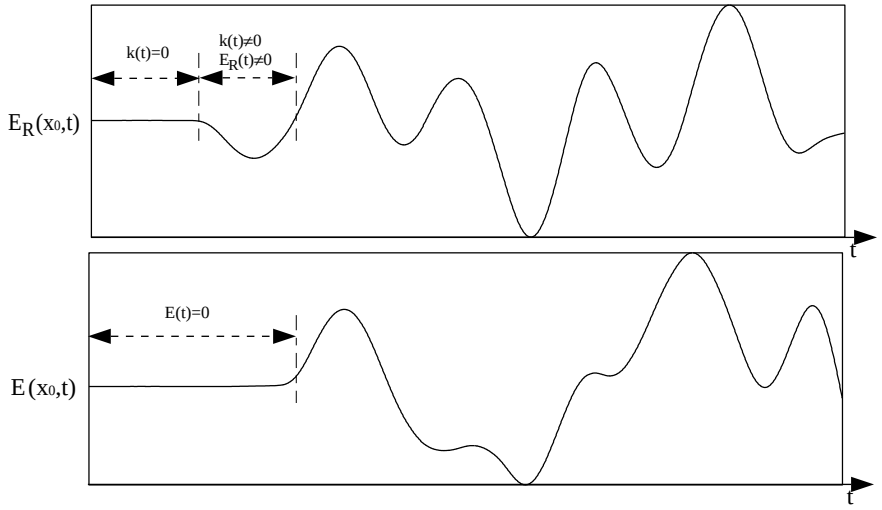


Figure 5: At a given point $\mathbf{x}_0$, comparison of $\mathbf{E}_R(\mathbf{x}_0, t)$ (top) and $\mathbf{E}(\mathbf{x}_0, t)$ (bottom) with t varying.

For the first case, one computes the electromagnetic field generated by a current, the space and time characteristics of which are similar to those of a bunched beam of particles. An L-shaped domain ω is considered where perfectly conducting boundary condition is imposed. The initial conditions are set to zero. The electromagnetic wave is generated by a current $\mathbf{J}(\mathbf{x}, t) = (J_1, J_2)^T$, the support of which is bounded, with $J_1 = 0$, $J_2 = 10 \sin(\lambda t)$, for λ associated to a frequency of $2,5.10^9$ Hertz. This current generates a wave which propagates both on the left and on the right. Physically, as long as the wave has not reached the reentrant corner, the field is smooth. Let t_s be the impact time, then, if one writes $\mathbf{E}(\mathbf{x}, t) = \mathbf{E}_R(\mathbf{x}, t) + \kappa(t)\mathbf{v}_S(\mathbf{x})$, $\kappa(t)$ is equal to zero for all t lower than t_s, and so $\mathbf{E}_R(\mathbf{x}, t)$ and $\mathbf{E}(\mathbf{x}, t)$ coincide. Now, on the one hand, for t greater than t_s , $\kappa(t) \neq 0$, and the support of $\mathbf{v}_S$ being non local (in fact, the support of $\mathbf{v}_S$ spans the whole of ω), one has $\kappa(t)\mathbf{v}_S(\mathbf{x}) \neq 0$, for all $\mathbf{x} \in \omega$ and $t > t_s$. On the other hand, however, one wishes to reproduce the obvious physical behavior, which is that for any point $\mathbf{x}$ and time t, $\mathbf{E}(\mathbf{x}, t) = 0$ if $t < t_{\mathbf{x}}$, where $t_{\mathbf{x}}$ denotes the time at which the electromagnetic wave reaches $\mathbf{x}$. One can check (see Figure 5)

that $\mathbf{E}_R(\mathbf{x},t)$ takes non-zero values, and therefore that it 'compensates' for $\kappa(t)\ \mathbf{v}_S(\mathbf{x})$, i.e. $\mathbf{E}_R(\mathbf{x},t) = -\kappa(t)\ \mathbf{v}_S(\mathbf{x})$. Thus, $\mathbf{E}(\mathbf{x},t)$ remains equal to zero while $t_s < t < t_{\mathbf{x}}$.

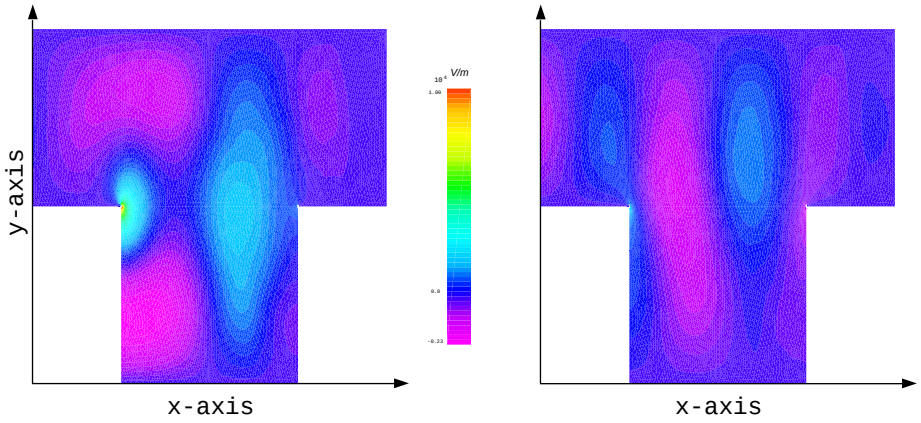


Figure 6: Computed electric field: with and without the SCM.

The second example is a guided wave which propagates in a standard *singular* geometry, as commonly studied devices such as hyperfrequency systems often include waveguides. This case illustrates of the possibilities of the method, when it is used on a more 'complete' formulation, that is with different types of boundary conditions and several reentrant corners. An incident wave enters in a step waveguide through the left boundary, and exits through the right boundary. At the initial time, the electromagnetic field is equal to zero all over the guide.

The Figure 6 depicts the isovalues of the first component of the electric field after 1000 time-steps. the SCM provides a numerical solution which is precise especially in the neighborhood of the corners. The result obtained *via* the classical nodal FE code (without the SCM) shows a most unlikely approximation of the true solution (no singular behavior).

Conclusion

We proposed a method, called the *Singular Complement Method*, to solve PDEs in a non-smooth and a non-convex domain. It is based on a splitting of the space of solutions V with respect to regularity (cf. (4)), in a subspace V_R made of regular elements, which is equal to V when the domain is smooth or convex, and a subspace of singular elements V_S. Regular elements are approximated by the P_1 Lagrange FEM, and test-functions are added to capture numerically the singular part of the solution.

In 3d domains, for the Laplace problem as well as for Maxwell's equations, the theoretical aspects are under control, but there still remains to provide an effective approximation of the singular part of the solution. Basically, two problems have to be overcome:

- The dimension of V_S is infinite.
- The edge and vertex singularities are linked (cf. [Gri85]).

These difficulties are not really equivalent. Indeed, on the one hand, one usually deals with infinite dimensional vector spaces: for instance, the space V_R is efficiently approximated with the help of the P_1 Lagrange FEM. On the other hand, finding an approximation, which takes into account the links between the two types of geometrical singularities, is much more challenging.

In 2d (or in axisymmetric domains), the situation is well understood theoretically, and numerical experiments are well under way and partial results are satisfactory. Moreover, the SCM is easy to implement, as it can be included in already existing codes, without having to rewrite them in their entirety; also, it generates a reasonable overhead (low additional memory requirements, small cpu costs). So, all's well that ends well, cf. [Sha98].

References

[ABDG98]C. Amrouche, C. Bernardi, M. Dauge, and V. Girault. Vector potentials in three-dimensional non-smooth domains. *Math. Meth. Appl. Sci.*, 21:823–864, 1998.

[AC97]F. Assous and P. Ciarlet Jr. A characterization of the orthogonal of $\Delta(H^2(\Omega)\cap H_0^1(\Omega))$ in $L^2(\Omega)$. *C. R. Acad. Sc. Paris, Série I*, t. 325:605–610, 1997.

[AC00]F. Assous and P. Ciarlet Jr. *Models and methods for Maxwell's equations (in French)*. ENSTA (Paris, France), 2000. Graduate course of the Univ. of Versailles and St-Quentin-en-Yvelines.

[ACL00]F. Assous, P. Ciarlet Jr., and S. Labrunie. Theoretical tools for solving the axisymmetric Maxwell equations. Technical Report 343, ENSTA (Paris, France), 2000.

[ACL01]Franck Assous, Patrick Ciarlet Jr., and Simon Labrunie. Resolution of axisymmetric maxwell equations. Technical Report 50, Institut Elie Cartan, Nancy, France, 2001.

[ACRS99]F. Assous, P. Ciarlet Jr., P.A. Raviart, and E. Sonnendrücker. Characterization of the singular part of the solution of Maxwell's equations in a polyhedral domain. *Math. Meth. Appl. Sci.*, 22:485–499, 1999.

[ACS98]F. Assous, P. Ciarlet Jr., and E. Sonnendrücker. Resolution of the Maxwell equations in a domain with reentrant corners. *Math. Mod. Num. Anal.*, 32:359–389, 1998.

[ACS00]F. Assous, P. Ciarlet Jr., and J. Segré. Numerical solution to the time dependent Maxwell equations in two dimensional singular domains: the Singular Complement Method. *J. Comput. Phys.*, 161:218–249, 2000.

[ADH+93]F. Assous, P. Degond, E. Heintzé, P.A. Raviart, and J. Segré. On a finite element method for solving the three-dimensional Maxwell equations. *J. Comput. Phys.*, 109:222–237, 1993.

[BD82]H. Blum and M. Dobrowolski. On finite element methods for elliptic equations on domains with corners. *Computing*, 28:53–63, 1982.

[BDM99]C. Bernardi, M. Dauge, and Y. Maday. *Spectral methods for axisymmetric domains*. Series in Applied Mathematics. Gauthier-Villars, Paris, and North Holland, Amsterdam, 1999.

[BHL99]A.S. Bonnet-Ben Dhia, C. Hazard, and S. Lohrengel. A singular field method for the solution of Maxwell's equations in polyhedral domains. *SIAM J. Appl. Math.*, 59:2028–2044, 1999.

[Bre98]S. Brenner. Overcoming corner singularities using multigrid methods. *SIAM J. Numer. Anal.*, 35:1883–1892, 1998.

[BS87]M.Sh. Birman and M.Z. Solomyak. Maxwell operator in regions with nonsmooth boundaries. *Sib. Math. J.*, 28:12–24, 1987.

[BS90]M.Sh. Birman and M.Z. Solomyak. Construction in a piecewise smooth domain of a function of the class H^2 from the value of the conormal derivative. *Sib. Math. J.*, 49:1128–1136, 1990.

[CD99]M. Costabel and M. Dauge. Maxwell and Lamé eigenvalues on polyhedra. *Math. Meth. Appl. Sci.*, 22:243–258, 1999.

[CDN99]M. Costabel, M. Dauge, and S. Nicaise. Singularities of Maxwell interface problems. *Math. Mod. Num. Anal.*, 33:627–649, 1999.

[Cos91]M. Costabel. A coercive bilinear form for Maxwell's equations. *J. Math. Anal. and Appl.*, 157:527–541, 1991.

[CZ97]P. Ciarlet Jr. and J. Zou. Finite element convergence for the Darwin model to Maxwell's equations. *Math. Mod. Num. Anal.*, 31:213–250, 1997.

[Dau88]M. Dauge. *Elliptic boundary value problems on corner domains*. Lecture Notes in Mathematics. Springer Verlag, Berlin, 1988.

[Gri85]P. Grisvard. *Elliptic problems in nonsmooth domains*. Pitman Publishing, Boston, 1985.

[Gri92]P. Grisvard. *Singularities in boundary value problems*, volume 22 of *RMA*. Masson, Paris, 1992.

[Hag89]W. W. Hager. Updating the inverse of a matrix. *SIAM Review*, 31:221–239, 1989.

[Her93]Francois Hermeline. Two coupled particle-finite volume methods using delaunay-voronoi meshes for the approximation of vlasov-poisson and vlasov-maxwell equations. *J. Comput. Phys.*, 106, 1993.

[HL96]C. Hazard and M. Lenoir. On the solution of time-harmonic scattering problems for Maxwell's equations. *SIAM J. Math. Anal.*, 27:1597–1630, 1996.

[HL00]C. Hazard and S. Lohrengel. A singular field method for Maxwell's equations: numerical aspects in two dimensions. Technical Report 595, Laboratoire J. A. Dieudonné, University of Nice-Sophia Antipolis (Nice, France), 2000.

[LM72]Jacques-Louis Lions and Enrico Magenes. *Nonhomogeneous Boundary Value Problems and Applications*, volume I. Springer, New York, Heidelberg, Berlin, 1972.

[Mar97]D. Martin. Documentation MELINA, 1997.

[Mou84]M. Moussaoui. Sur l'approximation des solutions du problème de Dirichlet dans un ouvert avec coins. In P. Grisvard, W. Wendland, and J.R. Whiteman, editors, *Singularities and constructive methods for their treatment*, pages 199–206. Springer Verlag, 1121, 1984.

[Néd80]J.-C. Nédélec. Mixed finite elements in R^3. *Numer. Math.*, 35:315–341, 1980.

[Néd86]J.-C. Nédélec. A new family of mixed finite elements in R^3. *Numer. Math.*, 50:57–81, 1986.

[Sha98]W. Shakespeare. *Much ado about nothing*. Unknown, 1598.

[Wah91]L. B. Wahlbin. Local behavior in finite element methods. In P. G. Ciarlet and J.-L. Lions, editors, *Handbook of numerical analysis. Volume II*, pages 353–522. North Holland, 1991.

[Web80]C. Weber. A local compactness theorem for Maxwell's equations. *Math. Meth. Appl. Sci.*, 2:12–25, 1980.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

15 Mortar spectral element discretization of Darcy's equations

Mejdi Azaïez[1], Faker Ben Belgacem[2] Christine Bernardi[3]

Introduction

Darcy's equations model the filtration of an incompressible viscous fluid in porous media. However, exactly the same equations are involved in the mixed formulation of the Laplace equation with Neumann boundary conditions and also in the projection–diffusion algorithm of Chorin [Cho68] and Temam [Tem68] for solving the time-dependent Navier–Stokes equations. So proposing discretizations of this problem which are both accurate and efficient, seems rather important. We first write its variational formulation, which involves the domain of the divergence operator, and prove that it is well-posed. We describe a spectral discretization of the problem that relies on the mortar domain decomposition technique introduced by Bernardi, Maday and Patera [BMP94], since it combines the accuracy of standard spectral methods with the advantage of handling complex geometries via the mortar algorithm. We prove the convergence of the discrete solution towards the exact one and derive error estimates.

Detailed proofs of the results presented in this paper can be found in [ABB03], and numerical experiments are under consideration.

Darcy's equations and their variational formulation

Let Ω be a bounded connected domain in $\mathbb{R}^d$, $d = 2$ or 3, with a Lipschitz–continuous boundary, and let $\mathbf{n}$ denote the unit normal vector outward to Ω. Darcy's equations in this domain write

$$\begin{aligned} \mathbf{u} + \mathbf{grad}\, p &= \mathbf{f} && \text{in } \Omega, \\ \operatorname{div} \mathbf{u} &= 0 && \text{in } \Omega, \\ \mathbf{u} \cdot \mathbf{n} &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{1}$$

where the unknowns are the velocity $\mathbf{u}$ and the pressure p. In order to write the variational formulation of problem (1), we first consider the space

$$H(\operatorname{div}, \Omega) = \left\{ \mathbf{v} \in L^2(\Omega)^d;\ \operatorname{div} \mathbf{v} \in L^2(\Omega) \right\}, \tag{2}$$

[1]Institut de Mécanique des Fluides de Toulouse (UMR C.N.R.S. 5502), Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France.
[2]Mathématiques pour l'Industrie et la Physique (UMR C.N.R.S. 5640),
Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France.
[3]Analyse Numérique, C.N.R.S. & Université Pierre et Marie Curie,
B.C. 187, 4 place Jussieu, 75252 Paris Cedex 05, France.

provided with the natural norm

$$\|\mathbf{v}\|_{H(\mathrm{div},\Omega)} = \big(\|\mathbf{v}\|^2_{L^2(\Omega)^d} + \|\mathrm{div}\,\mathbf{v}\|^2_{L^2(\Omega)}\big)^{\frac{1}{2}}. \tag{3}$$

We note that $H(\mathrm{div},\Omega)$ is a Hilbert space and we recall from [GR86](Chap. I, Thm 2.4) that the space $\mathcal{D}(\overline{\Omega})^d$ of restrictions of infinitely differentiable functions on $\mathbb{R}^d$ to $\overline{\Omega}$ is dense in $H(\mathrm{div},\Omega)$. As a consequence, the trace operator: $\mathbf{v} \mapsto \mathbf{v}\cdot\mathbf{n}$, defined from the formula

$$\forall \varphi \in H^1(\Omega), \quad \langle \mathbf{v}\cdot\mathbf{n}, \varphi\rangle = \int_\Omega \big(\mathbf{v}\cdot\mathbf{grad}\,\varphi + (\mathrm{div}\,\mathbf{v})\varphi\big)(\mathbf{x})\,d\mathbf{x} \tag{4}$$

is continuous from $H(\mathrm{div},\Omega)$ onto the dual space $H^{-\frac{1}{2}}(\partial\Omega)$ of $H^{\frac{1}{2}}(\partial\Omega)$. So, we can now define the subspace

$$H_0(\mathrm{div},\Omega) = \big\{\mathbf{v} \in H(\mathrm{div},\Omega);\ \mathbf{v}\cdot\mathbf{n} = 0 \text{ on } \partial\Omega\big\}, \tag{5}$$

which is also a Hilbert space and is the closure for the norm defined in (3) of the space $\mathcal{D}(\Omega)^d$ of functions in $\mathcal{D}(\overline{\Omega})^d$ with a compact support in Ω. Finally, we introduce the space

$$L^2_0(\Omega) = \big\{q \in L^2(\Omega);\ \int_\Omega q(\mathbf{x})\,d\mathbf{x} = 0\big\}. \tag{6}$$

The variational formulation of problem (1) now reads
Find $(\mathbf{u},p)$ *in* $H_0(\mathrm{div},\Omega)\times L^2_0(\Omega)$ *such that*

$$\begin{aligned} &\forall \mathbf{v} \in H_0(\mathrm{div},\Omega), \quad a(\mathbf{u},\mathbf{v}) + b(\mathbf{v},p) = \int_\Omega \mathbf{f}(\mathbf{x})\cdot\mathbf{v}(\mathbf{x})\,d\mathbf{x},\\ &\forall q \in L^2_0(\Omega), \quad b(\mathbf{u},q) = 0, \end{aligned} \tag{7}$$

where the bilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ are defined by

$$a(\mathbf{u},\mathbf{v}) = \int_\Omega \mathbf{u}(\mathbf{x})\cdot\mathbf{v}(\mathbf{x})\,d\mathbf{x}, \qquad b(\mathbf{v},q) = -\int_\Omega (\mathrm{div}\,\mathbf{v})(\mathbf{x})q(\mathbf{x})\,d\mathbf{x}. \tag{8}$$

From the density of $\mathcal{D}(\Omega)^d$ in $H_0(\mathrm{div},\Omega)$, it is readily checked that problem (7) is equivalent to problem (1). Problem (7) is of saddle-point type, and the arguments for proving its well-posedness are given in [GR86] (Chap. I, Thm 4.1). First, the bilinear forms $a(\cdot,\cdot)$ and $b(\cdot,\cdot)$ are continuous on $H_0(\mathrm{div},\Omega)\times H_0(\mathrm{div},\Omega)$ and $H_0(\mathrm{div},\Omega)\times L^2_0(\Omega)$, respectively. Second, let V stand for the kernel

$$V = \big\{\mathbf{v} \in H_0(\mathrm{div},\Omega);\ \forall q \in L^2_0(\Omega),\ b(\mathbf{v},q) = 0\big\}, \tag{9}$$

or, equivalently,

$$V = \big\{\mathbf{v} \in H_0(\mathrm{div},\Omega);\ \mathrm{div}\,\mathbf{v} = 0 \text{ in } \Omega\big\}. \tag{10}$$

The following ellipticity property is then obvious

$$\forall \mathbf{v} \in V, \quad a(\mathbf{v},\mathbf{v}) = \|\mathbf{v}\|^2_{H(\mathrm{div},\Omega)}. \tag{11}$$

Third, the following inf-sup condition, for a constant $\beta > 0$,

$$\forall q \in L_0^2(\Omega), \qquad \sup_{\mathbf{v} \in H_0(\mathrm{div},\Omega)} \frac{b(\mathbf{v},q)}{\|\mathbf{v}\|_{H(\mathrm{div},\Omega)}} \geq \beta \, \|q\|_{L^2(\Omega)}, \tag{12}$$

is derived by taking $\mathbf{v}$ equal to $\mathbf{grad}\, \varphi$, where φ is the solution of the Laplace equation with data q and homogeneous Neumann boundary conditions. Combining all this leads to the following statement.

Proposition 1. *For any data* $\mathbf{f}$ *in* $L^2(\Omega)^d$*, problem* (7) *has a unique solution* $(\mathbf{u}, p)$ *in* $H_0(\mathrm{div},\Omega) \times L_0^2(\Omega)$.

Unfortunately, even for smooth data, the solution of problem (7) is not very regular. For any data $\mathbf{f}$ in $L^2(\Omega)^d$ such that $\mathbf{curl}\, \mathbf{f}$ belongs to $L^2(\Omega)^{2d-3}$, the solution $(\mathbf{u}, p)$ belongs to $H^s(\Omega)^d$ for $s = \frac{1}{2}$ in the general case, $s = 1$ if Ω is convex and some $s > \frac{1}{2}$ if Ω is a polygon or polyhedron (we refer to [Cos90], [Dau92] and [ABDG98] for these results).

Remark: Another variational formulation of problem (1) exists, where the spaces $H_0(\mathrm{div},\Omega)$ and $L_0^2(\Omega)$ are replaced by $L^2(\Omega)^d$ and $H^1(\Omega) \cap L_0^2(\Omega)$, respectively. Then, the boundary conditions in (1) are enforced in a variational way. However this second formulation does not seem appropriate when Darcy's system appears in the discretization of the Stokes problem, since the pressure in this problem does not belong to $H^1(\Omega)$ in most cases when Ω is a non convex polygon or polyhedron.

The mortar spectral element discrete problem

From now on, in view of applying the mortar element method to our problem, we assume that Ω admits a disjoint decomposition into a finite number of (open) rectangles in dimension $d = 2$, rectangular parallelepipeds in dimension $d = 3$:

$$\overline{\Omega} = \bigcup_{k=1}^{K} \overline{\Omega}_k \qquad \text{and} \qquad \Omega_k \cap \Omega_{k'} = \emptyset, \quad 1 \leq k \neq k' \leq K. \tag{13}$$

We make the further assumption that the intersection of each $\partial\Omega_k$ with $\partial\Omega$, if not empty, is a corner, a whole edge or a whole face of Ω_k. We denote by $\mathbf{n}_k$, $1 \leq k \leq K$, the unit normal vectors outward to Ω_k. We introduce the skeleton $\mathcal{S}$ of the decomposition, $\mathcal{S} = \cup_{k=1}^{K} \partial\Omega_k \setminus \partial\Omega$. According to the ideas in [BMP94], we choose a disjoint decomposition of this skeleton into mortars:

$$\overline{\mathcal{S}} = \bigcup_{m=1}^{M} \overline{\gamma}_m \qquad \text{and} \qquad \gamma_m \cap \gamma_{m'} = \emptyset, \quad 1 \leq m \neq m' \leq M, \tag{14}$$

where each γ_m is a whole edge in dimension $d = 2$, face in dimension $d = 3$, of a subdomain Ω_k, denoted by $\Omega_{k(m)}$. To describe the discrete spaces, for each nonnegative integer n, we define on each Ω_k, resp. on each edge or face Γ of Ω_k, the space $I\!P_n(\Omega_k)$, resp. $I\!P_n(\Gamma)$, of restrictions to Ω_k, resp. Γ, of polynomials with d, resp. $d-1$, variables and degree $\leq n$ with

respect to each variable. The discretization parameter δ is then a K–tuple $(N_1, \ldots, N_K)$ of integers $N_k \geq 2$. We first introduce the space $M_\delta(\Omega)$ of discrete pressures:

$$M_\delta(\Omega) = \big\{ q_\delta \in L^2_0(\Omega);\ q_{\delta\,|\Omega_k} \in I\!P_{N_k-2}(\Omega_k),\ 1 \leq k \leq K \big\}. \tag{15}$$

Next, in analogy with the standard definition of the mortar approximation of $H^1(\Omega)$ [BMP94], we define the discrete space $X_\delta(\Omega)$ which approximates $H_0(\mathrm{div}, \Omega)$. It is the space of functions $\mathbf{v}_\delta$ such that:

- their restrictions $\mathbf{v}_{\delta\,|\Omega_k}$ to each Ω_k, $1 \leq k \leq K$, belong to $I\!P_{N_k}(\Omega_k)^d$,
- their normal traces $\mathbf{v}_\delta \cdot \mathbf{n}$ vanish on $\partial\Omega$,
- the mortar function φ being defined on each γ_m, $1 \leq m \leq M$, by

$$\varphi_{|\gamma_m} = \mathbf{v}_{\delta\,|\Omega_{k(m)}} \cdot \mathbf{n}_{k(m)}, \tag{16}$$

the following matching condition holds on each edge or face Γ of Ω_k, $1 \leq k \leq K$, which is not a mortar:

$$\forall \chi \in I\!P_{N_k-2}(\Gamma), \quad \int_\Gamma (\mathbf{v}_{\delta\,|\Omega_k} \cdot \mathbf{n}_k + \varphi)(\tau)\, \chi(\tau)\, d\tau = 0. \tag{17}$$

Remark: The space X_δ is not contained in $H(\mathrm{div}, \Omega)$ since the matching conditions on the normal derivative through the interfaces are only enforced in a weak way. So the discretization is nonconforming. Starting from the standard Gauss–Lobatto formula on $]-1, 1[$, we define on each Ω_k and in each direction:

- the nodes x_i^k and y_i^k, and the weights $\rho_i^{x,k}$ and $\rho_i^{y,k}$, $0 \leq i \leq N_k$, in the case of dimension $d = 2$,
- the nodes x_i^k, y_i^k and z_i^k, and the weights $\rho_i^{x,k}$, $\rho_i^{y,k}$ and $\rho_i^{z,k}$, $0 \leq i \leq N_k$, in the case of dimension $d = 3$.

A discrete product is then introduced on each Ω_k, according if $d = 2$ or 3, by

$$(u_\delta, v_\delta)_\delta^k = \begin{cases} \sum_{i=0}^{N_k} \sum_{j=0}^{N_k} u_\delta(x_i^k, y_j^k) v_\delta(x_i^k, y_j^k)\, \rho_i^{x,k} \rho_j^{y,k} \\ \\ \sum_{i=0}^{N_k} \sum_{j=0}^{N_k} \sum_{p=0}^{N_k} u_\delta(x_i^k, y_j^k, z_p^k) v_\delta(x_i^k, y_j^k, z_p^k)\, \rho_i^{x,k} \rho_j^{y,k} \rho_p^{z,k}. \end{cases} \tag{18}$$

The global discrete product on Ω:

$$(u_\delta, v_\delta)_\delta = \sum_{k=1}^{K} (u_\delta, v_\delta)_\delta^k, \tag{19}$$

coincides with the scalar product of $L^2(\Omega)$ for all functions u_δ and v_δ such that each product $(u_\delta v_\delta)_{|\Omega_k}$, $1 \leq k \leq K$, belongs to $I\!P_{2N_k-1}(\Omega_k)$. The discrete problem is now built from the variational formulation (7). For any continuous data $\mathbf{f}$ on $\overline{\Omega}$, it reads

Find $(\mathbf{u}_\delta, p_\delta)$ *in* $X_\delta(\Omega) \times M_\delta(\Omega)$ *such that*

$$\begin{aligned} &\forall \mathbf{v}_\delta \in X_\delta(\Omega), \quad a_\delta(\mathbf{u}_\delta, \mathbf{v}_\delta) + b_\delta(\mathbf{v}_\delta, p_\delta) = (\mathbf{f}, \mathbf{v}_\delta)_\delta, \\ &\forall q_\delta \in M_\delta(\Omega), \quad b_\delta(\mathbf{u}_\delta, q_\delta) = 0, \end{aligned} \tag{20}$$

where the bilinear forms $a_\delta(\cdot, \cdot)$ and $b_\delta(\cdot, \cdot)$ are defined by

$$a_\delta(\mathbf{u}_\delta, \mathbf{v}_\delta) = (\mathbf{u}_\delta, \mathbf{v}_\delta)_\delta, \qquad b_\delta(\mathbf{v}_\delta, q_\delta) = -(\mathrm{div}\, \mathbf{v}_\delta, q_\delta)_\delta. \tag{21}$$

Note however that, thanks to the exactness property of the quadrature formula, we have

$$\forall \mathbf{v}_\delta \in X_\delta(\Omega), \forall q_\delta \in M_\delta(\Omega), \quad b_\delta(\mathbf{v}_\delta, q_\delta) = b(\mathbf{v}_\delta, q_\delta). \tag{22}$$

To check the wellposedness of problem (20), we first state the discrete analogue of the inf-sup condition in (12), its proof combines the arguments in [ABG94] and [BBCM00]. It involves the "broken" norm

$$\|\mathbf{v}\|_{H(\mathrm{div},\cup\Omega_k)} = \Big(\sum_{k=1}^{K} \|\mathbf{v}\|^2_{H(\mathrm{div},\Omega_k)}\Big)^{\frac{1}{2}}. \tag{23}$$

Lemma 2. *There exists an integer N_D and a positive constant β_D, both depending on the decomposition of Ω but independent of δ, such that, if all the N_k are $\geq N_D$, the following inf-sup condition holds*

$$\forall q_\delta \in M_\delta(\Omega), \qquad \sup_{\mathbf{v}_\delta \in X_\delta(\Omega)} \frac{b(\mathbf{v}_\delta, q_\delta)}{\|\mathbf{v}_\delta\|_{H(\mathrm{div},\cup\Omega_k)}} \geq \beta_D \, \|q_\delta\|_{L^2(\Omega)}, \tag{24}$$

Proposition 3. *For any continuous data $\mathbf{f}$ on $\overline{\Omega}$ and if all the N_k are $\geq N_D$, problem (20) has a unique solution $(\mathbf{u}_\delta, p_\delta)$ in $X_\delta(\Omega) \times M_\delta(\Omega)$.*

Proof: Problem (20) results into a square linear system, so that it has a unique solution if and only if the only solution for $\mathbf{f} = \mathbf{0}$ is $(\mathbf{0}, 0)$. So we take $\mathbf{f}$ equal to $\mathbf{0}$. Choosing $\mathbf{v}_\delta$ equal to $\mathbf{u}_\delta$ in (20) yields that $a_\delta(\mathbf{u}_\delta, \mathbf{u}_\delta)$ is zero and, since the weights of the Gauss–Lobatto formula are positive, this imples that $\mathbf{u}_\delta$ vanishes in the $(N_k+1)^d$ nodes of a tensorized grid on each Ω_k, hence is zero. Then, $b_\delta(\mathbf{v}_\delta, p_\delta)$ is equal to zero for all $\mathbf{v}_\delta$ in X_δ, hence p_δ is zero due to (24).

A priori analysis

The main difficulty for evaluating the error on the velocity comes from the fact that the form $a_\delta(\cdot,\cdot)$ is no longer uniformly elliptic with respect to the norm $\|\cdot\|_{H(\mathrm{div},\cup\Omega_k)}$ on the discrete kernel

$$V_\delta = \big\{\mathbf{v} \in X_\delta(\Omega);\ \forall q_\delta \in M_\delta(\Omega),\ b_\delta(\mathbf{v}_\delta, q_\delta) = 0\big\}, \tag{25}$$

since V_δ is not made of exactly divergence-free functions. So the usual arguments for bounding the error does not hold, and we must evaluate "by hand" the quantity $\|\mathbf{u} - \mathbf{u}_\delta\|_{L^2(\Omega)^d}$. It involves three terms:

- the approximation error, which is easy to evaluate in dimension $d = 2$ but requires some further conformity assumptions in dimension $d = 3$,
- the error issued from numerical integration,
- the consistency error, which gives rise to a term of type (here, $[\cdot]$ denotes the jump through $\mathcal{S}$ with the appropriate sign)

$$\inf_{q_\delta \in M_\delta(\Omega)} \sup_{\mathbf{w}_\delta \in X_\delta(\Omega)} \frac{\int_{\mathcal{S}} (\mathbf{w}_\delta \cdot \mathbf{n})(\tau)\, [p - q_\delta](\tau)\, d\tau}{\|\mathbf{w}_\delta\|_{H(\mathrm{div},\cup\Omega_k)}}, \tag{26}$$

and it seems that using an inverse inequality is unavoidable to bound this term, which leads to non optimal estimates. Once the error on the velocity is evaluated, the error on the pressure is

derived from the inf-sup condition (24). Let μ_δ denote the maximal ratio $N_k/N_{k'}$ for all pairs of subdomains Ω_k and $\Omega_{k'}$, $1 \le k \ne k' \le K$, such that $\partial\Omega_k \cap \partial\Omega_{k'}$ has a positive measure in $\mathcal{S}$.

Theorem 4. *In dimension $d = 2$, assume the data $\mathbf{f}$ such that each $\mathbf{f}_{|\Omega_k}$, $1 \le k \le K$, belongs to $H^{\sigma_k}(\Omega_k)^2$, $\sigma_k > 1$, and the solution $(\mathbf{u}, p)$ of problem (1) such that each $(\mathbf{u}_{|\Omega_k}, p_{|\Omega_k})$, $1 \le k \le K$, belongs to $H^{s_k}(\Omega_k)^2 \times H^{s_k+1}(\Omega_k)$, $s_k > 0$. If all the N_k are $\ge N_D$, the following error estimate holds between this solution $(\mathbf{u}, p)$ and the solution $(\mathbf{u}_\delta, p_\delta)$ of problem (20):*

$$\begin{aligned}\|\mathbf{u} - \mathbf{u}_\delta\|_{L^2(\Omega)^2} + \|p - p_\delta\|_{L^2(\Omega)} \\ \le c \sum_{k=1}^{K} \big(\mu_\delta\, N_k^{\frac{1}{2}-s_k} \left(\|\mathbf{u}\|_{H^{s_k}(\Omega_k)^2} + \|p\|_{H^{s_k+1}(\Omega_k)}\right) \\ + N_k^{-\sigma_k} \|\mathbf{f}\|_{H^{\sigma_k}(\Omega_k)^2}\big).\end{aligned} \tag{27}$$

This estimate is not optimal, however the same lack of optimality appears in several finite element discretizations of Darcy's equations (for instance, when Crouzeix–Raviart finite elements are employed for the approximation of the velocity). Moreover, if the parameter μ_δ is bounded independently of δ, the convergence of the method can be derived form this estimate in all polygons, thanks to the regularity results stated in Section 1.

To conclude, we recall that the decomposition (13) of Ω is said to be conforming if the intersection of all $\partial\Omega_k$ and $\partial\Omega_{k'}$, $1 \le k < k' \le K$, if not empty, is a whole edge in dimension $d = 2$, a whole face in dimension $d = 3$, of both Ω_k and $\Omega_{k'}$. The mortar element method does not require the conformity of the decomposition. However, if the decomposition is conforming, an approximation q_δ of the pressure p can be constructed in $M_\delta(\Omega) \cap H^1(\Omega)$, which means that the quantity in (26) vanishes for this q_δ. So the error estimate is optimal in this case.

Corollary 5. *If all assumptions of Theorem 4 hold and if, moreover,*
(i) the decomposition (13) of Ω is conforming,
(ii) in dimension $d = 3$, for $1 \le m \le M$, $N_{k(m)}$ is $\ge N_k$, where γ_m is the intersection of this $\overline{\Omega}_k$ and $\overline{\Omega}_{k(m)}$,
the following error estimate holds between the solution $(\mathbf{u}, p)$ of problem (1) and the solution $(\mathbf{u}_\delta, p_\delta)$ of problem (20):

$$\begin{aligned}&\|\mathbf{u} - \mathbf{u}_\delta\|_{L^2(\Omega)^d} + \|p - p_\delta\|_{L^2(\Omega)} \\ &\le c \sum_{k=1}^{K} \left(N_k^{-s_k} \left(\|\mathbf{u}\|_{H^{s_k}(\Omega_k)^d} + \|p\|_{H^{s_k+1}(\Omega_k)}\right) + N_k^{-\sigma_k} \|\mathbf{f}\|_{H^{\sigma_k}(\Omega_k)^d}\right).\end{aligned} \tag{28}$$

Conclusion

As a conclusion, the mortar spectral element discretization of problem (1) is fully optimal in the case of a conforming decomposition. It is not for a nonconforming decomposition,

however estimate (27) can be improved in this case by taking into account the local properties of conformity. It can also be noted that, for smooth data, the solution of problem (1) is regular outside a neighbourhood of the corners and edges of Ω, so that enforcing the conformity of the decomposition is more important in a neighbourhood of $\partial\Omega$ than elsewhere.

References

[ABB03]M. Azaiez, F. Ben Belgacem, and C. Bernardi. The mortar spectral element method in $h(\mathrm{div},\omega)$ and $h(\mathbf{curl},\omega)$. 2003.

[ABDG98]C. Amrouche, C. Bernardi, M. Dauge, and V. Girault. Vector potentials in three-dimensional non-smooth domains. *Math. Meth. in the Applied Sciences*, 21:823 – 864, 1998.

[ABG94]M. Azaiez, C. Bernardi, and M. Grundmann. Spectral methods applied to porous media. *East-West Journal on Numerical Analysis*, 2:91 – 105, 1994.

[BBCM00]F. Ben Belgacem, C. Bernardi, N. Chorfi, and Y. Maday. Inf-sup conditions for the mortar spectral element discretization of the Stokes problem. *Numer. Math.*, 85:257 – 281, 2000.

[BMP94]C. Bernardi, Y. Maday, and A.T. Patera. A new nonconforming approach to domain decomposition: the mortar element method. In H. Brezis and J.-L. Lions, editors, *Collège de France Seminar, Vol. XI*. Pitman, 1994.

[Cho68]A.J. Chorin. Numerical solution of the Navier–Stokes equations. *Math. Comput.*, 22:745 – 762, 1968.

[Cos90]M. Costabel. A remark on the regularity of solutions of Maxwell's equations on Lipschitz domains. *Math. Meth. in the Applied Sciences*, 12:365 – 368, 1990.

[Dau92]M. Dauge. Neumann and mixed problems on curvilinear polyhedra. *Integr. Equat. Oper. Th.*, 15:227 – 261, 1992.

[GR86]V. Girault and P.-A. Raviart. *Finite Element Methods for Navier–Stokes Equations, Theory and Algorithms*. Springer-Verlag, New York, 1986.

[Tem68]R. Temam. Une méthode d'approximation de la solution des équations de Navier–Stokes. *Bull. Soc. Math. France*, 98:115 – 152, 1968.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

16 Substructuring techniques and Wavelets for Domain Decomposition

Silvia Bertoluzza [1]

Introduction

We consider in this paper a substructuring approach for preconditioning the linear system arising from the reduction to the interface unknown of the discrete three fields formulation of domain decomposition. In particular we concentrate on choosing the stabilization technique, needed to circumvent the otherwise very restrictive inf-sup conditions required for stability and convergence, in such a way that the stabilized method falls in the range for which the estimate on the preconditioner holds. For such preconditioner to work, it is in fact necessary that the stabilized bilinear form verifies continuity and coercivity *with respect to the same norm*. This leads us to choose a stabilization technique based on adding a residual term on the subdomain boundaries, measured in the natural norm of type $H^{1/2}$. The $H^{1/2}$ type scalar product can be cheaply realized in terms of a wavelet decomposition. Remark that wavelets are employed here as a tool for implementing stabilization and they do not need to be employed as discretization space.

A substructuring preconditioner for the three fields domain decomposition method

Let $\Omega \subset \mathbb{R}^2$ be a convex polygonal domain. We will consider the following simple model problem: given $f \in L^2(\Omega)$, find u satisfying

$$-\Delta u = f \text{ in } \Omega, \qquad u = 0 \text{ on } \partial\Omega. \tag{1}$$

In this paper we consider the *three fields domain decomposition* formulation of such a problem [BM94]. More precisely, considering for simplicity a geometrically conforming decomposition $\Omega = \cup_k \Omega_k$, with Ω_k quadrangles regular in shape, $\Gamma_k = \partial\Omega_k$, and letting $\Sigma = \cup_k \Gamma_k$, we introduce the following functional spaces

$$V = \prod_k H^1(\Omega_k), \qquad \Lambda = \prod_k H^{-1/2}(\Gamma_k),$$
$$\Phi = \{\varphi \in L^2(\Sigma) : \text{there exists } u \in H_0^1(\Omega),\ u = \varphi \text{ on } \Sigma\} = H_0^1(\Omega)|_\Sigma,$$

respectively equipped with the norms:

$$\|u\|_V^2 = \sum_k \|u^k\|_{H^1(\Omega_k)}^2, \qquad \|\lambda\|_\Lambda^2 = \sum_k \|\lambda^k\|_{H^{-1/2}(\Gamma_k)}^2,$$

[1] Istituto di Analisi Numerica del C.N.R. di Pavia, aivlis@ian.pv.cnr.it

and (see [Ber00a])

$$\|\varphi\|_{\Phi}^2 = \inf_{u\in H_0^1(\Omega): u=\varphi \text{ on } \Sigma} \|u\|_{H^1(\Omega)}^2 \simeq \sum_k |\varphi|_{H^{1/2}(\Gamma_k)}^2.$$

We remark that here and in the following we will use the notation c and C to indicate several positive constants independent of any relevant parameter, like the mesh size or the size of the subdomains. The expression $A \simeq B$ will stand for $cA \leq B \leq CA$.

Let $a^k : H^1(\Omega_k) \times H^1(\Omega_k) \to \mathbb{R}$ denote the bilinear form corresponding to the Laplace operator:

$$a^k(w,v) = \int_{\Omega_k} \nabla w \nabla v.$$

The continuous three fields formulation of equation (1) is the following ([BM94]): find $(u,\lambda,\varphi) \in V \times \Lambda \times \Phi$ such that

$$\begin{cases} \forall k,\ \forall v^k \in H^1(\Omega_k),\ \forall \mu^k \in H^{-1/2}(\Gamma_k): \\ \quad a^k(u^k,v^k) - \int_{\Gamma_k} v^k\lambda^k = \int_{\Omega_k} f v^k, \\ -\int_{\Gamma_k} u^k\mu^k + \int_{\Gamma_k} \mu^k\varphi = 0, \\ \text{and } \forall \psi \in \Phi: \\ \quad \sum_k \int_{\Gamma_k} \lambda^k\psi = 0. \end{cases} \tag{2}$$

It is known that this problem admits a unique solution (u,λ,φ), where u is indeed the solution of (1) and such that $\lambda^k = \partial u^k/\partial \nu^k$ on Γ_k, and $\varphi = u$ on Σ, where ν^k denotes the outer normal derivative to the subdomain Ω_k. After choosing discretization spaces $V_h = \prod_k V_h^k \subset \prod_k H^1(\Omega_k)$, $\Lambda_h = \prod_k \Lambda_h^k \subset \prod_k H^{-1/2}(\Gamma_k)$ and $\Phi_h \subset \Phi$, equation (2) can be discretized by a Galerkin scheme. The linear system stemming from such an approximation takes the form

$$\begin{pmatrix} A & B^T & 0 \\ B & 0 & C^T \\ 0 & C & 0 \end{pmatrix} \cdot \begin{pmatrix} \underline{u}_h \\ \underline{\lambda}_h \\ \underline{\varphi}_h \end{pmatrix} = \begin{pmatrix} \underline{f} \\ 0 \\ 0 \end{pmatrix}, \tag{3}$$

($\underline{u}_h$, $\underline{\lambda}_h$, and $\underline{\varphi}_h$ being the vectors of the coefficients of u_h, λ_h and φ_h in the bases chosen for V_h, Λ_h and Φ_h respectively). By a Schur complement argument the solution of (3) can be reduced to a system in the unknown $\underline{\varphi}_h$, which takes the form

$$\mathbf{C}\mathbf{A}^{-1}\mathbf{C}^T \underline{\varphi}_h = -\mathbf{C}\mathbf{A}^{-1} \begin{pmatrix} \underline{f} \\ 0 \end{pmatrix}, \quad \mathbf{C} = [\ 0 \quad C\ ], \quad \mathbf{A} = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}. \tag{4}$$

The matrix $S = \mathbf{C}\mathbf{A}^{-1}\mathbf{C}^T$ does not need to be assembled. The system (4) can rather be solved by an iterative technique (like for instance a conjugate gradient method) and therefore only the action of S on a given vector needs to be implemented. Multiplying by S implies the need for solving a linear system with matrix $\mathbf{A}$. This reduces, by a proper reordering of the

unknowns, to independently solving a discrete Dirichlet problem with Lagrange multipliers in each subdomain.

Existence, uniqueness and stability of the solution of the discretized problem rely on the validity of two *inf-sup* conditions,

$$\inf_{\lambda_h\in\Lambda_h}\sup_{u_h\in V_h}\frac{\sum_k\int_{\Gamma_k}\lambda_h^k u_h^k}{\|u_h\|_V\|\lambda_h\|_\Lambda}\geq\beta_1>0,\qquad \inf_{\varphi_h\in\Phi_h}\sup_{\lambda_h\in\Lambda_h}\frac{\sum_k\int_{\Gamma_k}\lambda_h^k\varphi_h}{\|\varphi_h\|_\Phi\|\lambda_h\|_\Lambda}\geq\beta_2>0 \tag{5}$$

respectively coupling V_h with Λ_h, and Λ_h with Φ_h. Provided (5) holds, it is possible to prove that the bilinear form $s:\Phi_h\times\Phi_h\to\mathbb{R}$ corresponding to the Schur complement matrix S and defined by

$$s(u_h,v_h)=\underline{v}_h^T S\underline{u}_h,$$

is continuous and coercive with respect to the Φ norm:

$$s(\varphi_h,\psi_h)\leq M_0\|\varphi_h\|_\Phi\|\psi_h\|_\Phi,\qquad s(\varphi_h,\varphi_h)\quad\geq\quad\alpha_0\|\varphi_h\|_\Phi^2, \tag{6}$$

(M_0 and α_0 positive constants).

The problem arises then to precondition the Schur complement matrix S. This can be done by a substructuring approach ([BPS86, Ber00b]). To this end we introduce a decomposition of the skeleton $\Sigma\setminus\partial\Omega=\cup_i e_i$ as the disjoint union of M macro-edges e_i, (each being the edge of two adjacent subdomains), and we split the discrete space Φ_h as the direct sum of a coarse space $\mathcal{L}_H$ of functions linear on each macro-edge of Σ,

$$\mathcal{L}_H=\{\varphi\in C^0(\Sigma):\quad\forall i=1,\dots,M,\ \ \varphi|_{e_i}\in\mathbb{P}_1(e_i),\ \varphi=0\text{ on }\partial\Omega\},$$

($\mathbb{P}_1$ denoting the space of polynomials of degree at most one) plus some local spaces (one per macro-edge) $\Phi_h^{0,i}$,

$$\Phi_h^{0,i}=\{\varphi_h\in\Phi_h:\ \ \varphi_h|_{\Sigma\setminus e_i}=0\},$$

consisting in those functions in Φ_h vanishing outside the macro-edge e_i. Corresponding to such a decomposition we will consider a block-Jacobi type preconditioner. More precisely, it is possible to prove the following theorem.

Theorem 1 *Let $\hat{s}_H:\mathcal{L}_H\times\mathcal{L}_H\to\mathbb{R}$ and $\hat{s}_i:\Phi_h^{0,i}\times\Phi_h^{0,i}\to\mathbb{R}$ be symmetric bilinear forms satisfying*

$$\hat{s}_H(\varphi_H,\varphi_H)\simeq\|\varphi_H\|_\Phi^2\ \ \forall\varphi_H\in\mathcal{L}_H,\qquad\text{and}\qquad\hat{s}_i(\varphi_h,\varphi_h)\simeq\|\varphi_h\|_\Phi^2,\ \ \forall\varphi_h\in\Phi_h^{0,i},$$

and let $\hat{s}:\Phi_h\times\Phi_h\to\mathbb{R}$ be the bilinear form which, for $\varphi_h=\varphi_H+\sum_{i=1}^M\varphi_h^{0,i}$ and $\psi_h=\psi_H+\sum_{i=1}^M\psi_h^{0,i}$, is defined by

$$\hat{s}(\varphi_h,\psi_h)=\hat{s}_H(\varphi_H,\psi_H)+\sum_{i=1}^M\hat{s}_i(\varphi_h^{0,i},\psi_h^{0,i}).$$

Then for all $\varphi_h\in\Phi_h$ it holds

$$c\|\varphi_h\|_\Phi^2\leq\hat{s}(\varphi_h,\varphi_h)\lesssim\max_k\left(1+\log\frac{H_k}{h_k}\right)^2\|\varphi_h\|_\Phi^2,$$

where h_k and H_k are respectively the smallest mesh size of $\Phi_h|_{\Gamma_k}$ and the diameter of the subdomain Ω_k.

Thanks to (6), by a well known argument, Theorem 1 implies that we can derive the following corollary, where we denote by $\hat{S}$ the matrix corresponding to the Galerkin discretization of the bilinear forms $\hat{s}$, which has a block diagonal structure.

Corollary 1 *If (5) holds, then*

$$cond(\hat{S}^{-1}S) \lesssim \max_k \left(1 + \log \frac{H_k}{h_k}\right)^2.$$

Wavelet stabilization

The need for the two inf-sup conditions (5) to hold, leads to discard several otherwise desirable choices for the three discretization spaces V_h, Λ_h and Φ_h. A possible remedy in this direction is to advocate a suitable *stabilization technique*, allowing to circumvent one or both inf-sup conditions. Several proposals have been made in this respect (see for instance [BFMR97]). In this particular context, we want however to choose the stabilization technique in such a way that the substructuring preconditioner briefly described in the previous section still applies. Therefore, the bilinear form corresponding to the Schur complement matrix deriving from the stabilized method needs to satisfy (6). A choice that fulfills such requirement is the *wavelet stabilization* proposed in [BK00]. This consists in introducing symmetric bilinear forms $[\cdot,\cdot]_{1/2,k} : H^{1/2}(\Gamma_k) \times H^{1/2}(\Gamma_k) \to \mathbb{R}$ satisfying the following bounds for all $\varphi_h, \psi_h \in \Phi_h|_{\Gamma_k}$ and for two suitable positive constants C_1 and c_1:

$$[\varphi_h, \psi_h]_{1/2,k} \leq C_1 |\varphi_h|_{H^{1/2}(\Gamma_k)} |\psi_h|_{H^{1/2}(\Gamma_k)}, \qquad [\varphi_h, \varphi_h]_{1/2,k} \geq c_1 |\varphi_h|^2_{H^{1/2}(\Gamma_k)}. \quad (7)$$

The *stabilized three fields formulation* of problem (1) reads: find u_h, λ_h and φ_h such that

$$\begin{cases} \forall k,\ \forall v_h^k \in V_h^k,\ \forall \mu_h^k \in \Lambda_h^k : \\ a^k(u_h^k, v_h^k) + \gamma[u_h^k, v_h^k]_{1/2,k} - \int_{\Gamma_k} v_h^k \lambda_h^k \quad - \gamma[\varphi_h, v_h^k]_{1/2,k} = \int_{\Omega_k} f v_h^k, \\ -\int_{\Gamma_k} u_h^k \mu_h^k \quad + \int_{\Gamma_k} \mu_h^k \varphi_h \quad = 0, \\ \text{and } \forall \psi_h \in \Phi_h : \\ -\sum_k \gamma[u_h^k, \psi_h]_{1/2,k} + \sum_k \int_{\Gamma_k} \lambda_h^k \psi_h + \sum_k \gamma[\varphi_h, \psi_h]_{1/2,k} \quad = 0, \end{cases} \quad (8)$$

where $\gamma > 0$ is a parameter independent of the choice of the discretization spaces. Such formulation is consistent with the original continuous problem, that is by substituting in (8) the solution (u, λ, φ) of (2) at the place of $(u_h, \lambda_h, \varphi_h)$ we obtain an identity. The linear system stemming from such a problem takes this time the following form:

$$\begin{pmatrix} \check{A} & B^T & -\gamma D^T \\ B & 0 & C^T \\ -\gamma D & C & \gamma E \end{pmatrix} \cdot \begin{pmatrix} \underline{u}_h \\ \underline{\lambda}_h \\ \underline{\varphi}_h \end{pmatrix} = \begin{pmatrix} \underline{f} \\ 0 \\ 0 \end{pmatrix}, \quad (9)$$

with $\check{A} = A + \gamma F$, the matrices D, E and F deriving from the stabilizing terms. Again, the solution of (9) can be reduced to a system in the unknown $\underline{\varphi}_h$, this time taking the form

$$\check{S}\underline{\varphi}_h := \left(\mathbf{D}\check{\mathbf{A}}^{-1}\mathbf{D}^T + \gamma E\right)\underline{\varphi}_h = -\mathbf{D}\check{\mathbf{A}}^{-1} \begin{pmatrix} \underline{f} \\ 0 \end{pmatrix}$$

with

$$\check{\mathbf{A}} = \begin{pmatrix} \check{A} & B^T \\ B & 0 \end{pmatrix}, \qquad \mathbf{D} = [\ -\gamma D \quad C\ ].$$

Once again we let $\check{s} : \Phi_h \times \Phi_h \to \mathbb{R}$ be the bilinear form corresponding to the Schur complement matrix $\check{S}$

$$\check{s}(\varphi_h, \psi_h) = \underline{\psi}_h{}^T \check{S} \underline{\varphi}_h,$$

and, if the space V_h and Λ_h satisfy the first of the two inf-sup conditions (5), also the bilinear form $\check{s}$ is continuous and coercive with respect to the Φ norm:

$$\check{s}(\varphi_h, \psi_h) \le M_1 \|\varphi_h\|_\Phi \|\psi_h\|_\Phi, \qquad \check{s}(\varphi_h, \varphi_h) \ge \alpha_1 \|\varphi_h\|_\Phi^2.$$

Also for the bilinear form $\check{s}$, Theorem 1 yields then the corollary

Corollary 2 *It holds*

$$cond(\hat{S}^{-1}\check{S}) \lesssim \max_k \left(1 + \log \frac{H_k}{h_k}\right)^2.$$

We need at this point to provide bilinear forms $[\cdot,\cdot]_{1/2,k}$ with the required characteristics. Following the proposal of [BK00], these are designed by means of a wavelet decomposition. For simplicity, let us assume that the subdomains are squares (otherwise we would need to map them onto a square). Since the $H^{1/2}(\Gamma_k)$ seminorm is invariant under changes of scale, we can rescale the subdomain in such a way that $|\Gamma_k| = 1$ (that is $H = 1/4$). For simplicity, let us concentrate on the case in which the skeleton Σ is discretized by means of P1 finite elements, and let us assume that on each macro-edge e_i the grid is uniform, with L_i elements, L_i being a power of two:

$$L_i = 2^{j_i} \qquad \text{for some } j_i \ge 1,$$

so that for all k, $\Phi_h|_{\Gamma_k} \subset V_{j_k+2}$, with $j_k = \max_{i:|e_i \cup \Gamma_k|>0} j_i$, where, for $j > 0$, V_j denotes the space of 1-periodic P1 finite elements on the uniform grid with mesh size $1/2^j$.

The sequence $\{V_j\}_{j\ge 0}$ forms a so called *multiresolution analysis* of $L^2(\Gamma_k)$ and it is well known (see for example [CDF92]) that there exists several wavelet bases associated with such a multiresolution analysis. More precisely there exist several P1 compactly supported functions $\theta \in C^0(\mathbb{R})$ defined on the uniform grid of mesh size 1 and integer nodes, such that, if we define *wavelets* $\theta_{m,\ell}$ by $\theta_{m,\ell} = \sum_{n=-\infty}^{+\infty} 2^{m/2}\theta(2^m(x-n)-\ell)$, all functions $\eta \in V_j$ can be written as

$$\eta = \eta_0 + \sum_{m=0}^{j-1} \sum_{\ell=1}^{2^m} \eta_{m,\ell}\theta_{m,\ell}, \qquad \eta_0 \text{ constant},$$

and such that

$$\eta \in V_j \quad \Longrightarrow \quad |\eta|^2_{H^{1/2}(\Gamma_k)} \simeq \sum_{m=0}^{j-1} \sum_{\ell=1}^{2^m} 2^m |\eta_{m,\ell}|^2.$$

If, for $\zeta, \xi \in L^2(\Gamma_k)$, we express in terms of the wavelet basis $\{\theta_{m,\ell}\}$ the respective $L^2(\Gamma_k)$ projections $\Pi_k(\zeta)$ and $\Pi_k(\xi)$ onto V_{j_k+2},

$$\Pi_k(\zeta) = \zeta_0 + \sum_{m=0}^{j_k+1} \sum_{\ell=1}^{2^m} \zeta_{m,\ell}\theta_{m,\ell}, \qquad \Pi_k(\xi) = \xi_0 + \sum_{m=0}^{j_k+1} \sum_{\ell=1}^{2^m} \xi_{m,\ell}\theta_{m,\ell},$$

we can define the bilinear form $[\cdot,\cdot]_{1/2,k}$ as

$$[\zeta, \xi]_{1/2,k} = \sum_{m=0}^{j_k+1} \sum_{\ell=1}^{2^m} 2^m \zeta_{m,\ell}\xi_{m,\ell}.$$

It is possible to prove ([BK00]) that the bilinear forms thus defined satisfies (7).

With this definition, the computation of $[u_h^k - \varphi_h, v_h^k - \psi_h]_{1/2,k}$ essentially reduces to first computing the nodal values of $\Pi_{j_k}(u_h^k)$, $\Pi_{j_k}(v_h^k)$, $\Pi_{j_k}(\varphi_h)$ and $\Pi_{j_k}(\psi_h)$ respectively and then applying a *Fast Wavelet Transform*.

Numerical results

We will consider problem (1) with $f = 1$ and $\Omega =]0,1[^2$. We consider an uniform decomposition of Ω in $K = 4 \times 4$ equal square subdomains of size $H \times H$, $H = 1/4$. In each subdomain Ω_k we take an uniform mesh composed by $N_k \times N_k$ equal square elements of size $\delta_k \times \delta_k$, $\delta_k = H/N_k = 1/(4N_k)$. We then define V_h^k to be the corresponding space of Q1 finite elements. The value of N_k is randomly assigned in such a way that for about one third of the subdomains $N_k = 5$, for about another third $N_k = 10$, and for the remaining subdomains $N_k = 15$. The multiplier space Λ_h^k is then defined as the trace on Γ_k of V_h^k. With such a choice it is possible to prove that the spaces Λ_h and V_h satisfy the first of the two inf-sup conditions needed for stability. The space Φ_h is chosen to be a P1 finite element space corresponding to a uniform grid on Σ with mesh size $1/(4 \cdot 2^J)$. As J increases, the second inf-sup condition – coupling Φ_h and Λ_h – fails. The consequent instability clearly appears in Figure 1, where on top we plot the solution φ_h obtained by the unstabilized formulation (2) for $J = 3$ (on the left) and $J = 5$ (on the right). On the bottom, we plot the solution φ_h obtained by the stabilized formulation (8) for the same values of J and for $\gamma = .05$. The stabilizing effect of the correction is evident. We next show, for different values of the stabilization parameter γ, the performance of the block Jacobi type preconditioner introduced in Section **??**, where the bilinear forms $\hat{s}_H$ and $\hat{s}_i$ are chosen according to [BPS86, Ber00b]. While the stabilized system is better preconditioned then the unstabilized one (first column in the table), apparently the stabilization parameter influences its performance, so its correct choice is important.

References

[Ber00a]Silvia Bertoluzza. Analysis of a stabilized domain decomposition method. Technical report, I.A.N.-C.N.R. Pavia, 2000.

[Ber00b]Silvia Bertoluzza. Substructuring preconditioners for the three fields domain decomposition method. Technical report, I.A.N.-C.N.R., 2000.

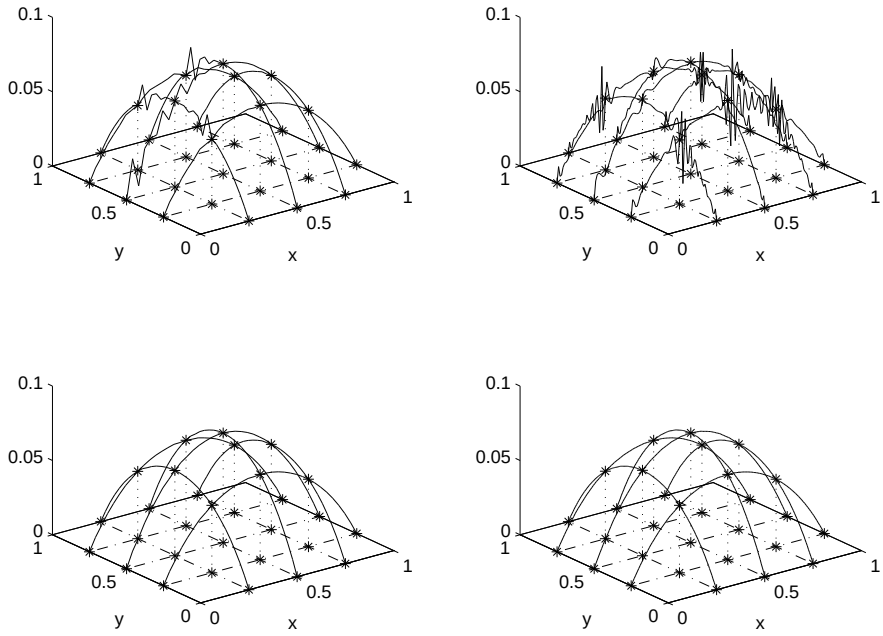


Figure 1: Effect of the stabilization: on top we display the results of the plain formulation and at the bottom the ones obtained by adding the stabilization term

2^J	$\gamma = 0$	$\gamma = .00125$	$\gamma = .05$	$\gamma = .25$
4	11	11	11	11
8	40	44	15	16
16	—	57	17	25
32	—	59	21	41

Table 1: Number of CG iterations needed to reduce the residual of a factor 10^{-5}. For $\gamma = 0$ and $J \geq 4$ the conjugate gradient procedure did not converge in the maximum number of iteration (which was set to 100).

[BFMR97]Franco Brezzi, Leopoldo P. Franca, Luisa D. Marini, and Alessandro Russo. Stabilization techniques for domain decomposition methods with nonmatching grids. In Petter E. Bjørstad, Magne Espedal, and David Keyes, editors, *Domain Decomposition Methods in Sciences and Engineering*. John Wiley & Sons, 1997. Proceedings from the Ninth International Conference, June 1996, Bergen, Norway.

[BK00]Silvia Bertoluzza and Angela Kunoth. Wavelet stabilization and preconditioning for domain decomposition. *IMA J. Numer. Anal.*, 20:533–559, 2000.

[BM94]Franco Brezzi and Luisa D. Marini. A three fields domain decomposition method. *Contemp. Math.*, 157:27–34, 1994.

[BPS86]James H. Bramble, Joseph E. Pasciak, and Alfred H. Schatz. The construction of preconditioners for elliptic problems by substructuring, I. *Math. Comp.*, 47(175):103–134, 1986.

[CDF92]Albert Cohen, Ingrid Daubechies, and Jean-Christophe Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

17 A Neumann-Neumann method using a finite volume discretization

R. Cautrès [1], T. Gallouët [2], S. Gerbi [3], R. Herbin [4]

Introduction

In this work, we present a non-overlapping domain decomposition method which is well adapted to the discretization of convection-diffusion equations by the finite volume scheme. The method which we shall consider is closely related to the so-called Neumann-Neumann relaxation operator which was studied in the finite element framework by several researchers among whom Glowinski et al. [BGLTV89] Dryja and Widlund [DW95] and Quarteroni and Marini [MQ89].
The algorithm is first written in the continuous case and then its discrete counterpart is presented in the framework of a finite volume discretization.
For the sake of simplicity, we shall only consider here the classical Laplace equation:

$$\begin{cases} -\Delta u = f \ \text{ on } \ \Omega, \\ u = 0 \ \text{ on } \ \partial\Omega, \end{cases} \tag{1}$$

where Ω is a bounded open subset of $\mathbb{R}^d$, $d = 2$ or 3, whose boundary $\partial\Omega$ is Lipschitz-continuous, $f \in L^2(\Omega)$. The generalization of the method to convection-diffusion equations seems possible since the convection term is easily handled in the finite volume scheme. This is the object of on-going work.

The Neumann-Neumann method

For the sake of simplicity, we shall consider here a non-overlapping domain decomposition which is defined by two subdomains Ω_1 and Ω_2 of Ω, which are bounded open subsets of $\mathbb{R}^d$ with Lipschitz-continuous boundaries such that $\overline{\Omega} = \overline{\Omega_1} \cup \overline{\Omega_2}$, and the interface $\gamma = \overline{\Omega_1} \cap \overline{\Omega_2}$ has a non zero $(d-1)$-dimensional measure.
For $i = 1, 2$, we denote by $\Gamma_i = \partial\Omega \bigcap \partial\Omega_i$, and $f_i = f_{|\Omega_i}$. We consider for $i = 1, 2$ the Hilbert spaces $V_i = \{\varphi_i \in H^1(\Omega_i), \quad \exists\ \varphi \in H_0^1(\Omega), \quad \varphi_i = \varphi_{|\Omega_i}\}$ equipped with the L^2 norm of the gradient. Let $H_{00}^{\frac{1}{2}}(\gamma)$ be the space of traces of elements of $H_0^1(\Omega)$ (or V_i) on γ. This space may be endowed with the norms of the harmonic lift in V_i on Ω_i , $\|\cdot\|_{i,H_{00}^{\frac{1}{2}}}$, for $i = 1$ or $i = 2$. It is well known that these norms are equivalent (see [QV99]). Hence there exist α and $\beta \in \mathbb{R}_+^*$ which only depend on Ω, Ω_1 and Ω_2, such that

$$\alpha \, \|\lambda\|_{1,H_{00}^{\frac{1}{2}}} \leq \|\lambda\|_{2,H_{00}^{\frac{1}{2}}} \leq \beta \, \|\lambda\|_{1,H_{00}^{\frac{1}{2}}}. \tag{2}$$

[1]University of Marseille, France, Rene.Cautres@cmi.univ-mrs.fr
[2]University of Marseille, France, Thierry.Gallouet@cmi.univ-mrs.fr
[3]University of Chambéry, France, Stephane.Gerbi@univ-savoie.fr
[4]University of Marseille, France, Raphaele.Herbin@cmi.univ-mrs.fr

Definition 1
For any $\lambda \in H_{00}^{\frac{1}{2}}(\gamma)$, let $u_i^{(\lambda)}$, for $i = 1, 2$ be the unique weak solution to the following problem

$$\left\{ \begin{array}{c} -\Delta u_i = f_i \ \text{ on } \ \Omega_i, \\ u_i = 0 \text{ on } \Gamma_i, \\ u_i = \lambda \ \text{ on } \ \gamma, \end{array} \right. \tag{3}$$

and $\Phi(\lambda)$ be the jump between the normal fluxes of the solutions $u_1^{(\lambda)}$ and $u_2^{(\lambda)}$, namely

$$\Phi(\lambda) = \sum_{i=1}^{2} \nabla u_i^{(\lambda)} \cdot n_i, \tag{4}$$

where n_i denotes the unit normal vector to the interface γ outward to Ω_i. We then define $v_1^{(\lambda)}$ as the unique weak solution to the following problem

$$\left\{ \begin{array}{c} -\Delta v_1 = 0 \ \text{ on } \ \Omega_1, \\ v_1 = 0 \ \text{ on } \ \Gamma_1, \\ \nabla v_1 \cdot n_1 = \Phi(\lambda) \ \text{ on } \ \gamma, \end{array} \right. \tag{5}$$

and $S_1(\lambda)$ as the trace of $v_1^{(\lambda)}$ on γ. Finally, for $\rho > 0$, let $T_{1,\rho}$ be defined from $H_{00}^{\frac{1}{2}}(\gamma)$ to $H_{00}^{\frac{1}{2}}(\gamma)$ by $T_{1,\rho}(\lambda) = \lambda - \rho S_1(\lambda)$. Let us now present the Neumann-Neumann type domain decomposition method. Let $\lambda^{(0)}$ be a given function of $H_{00}^{\frac{1}{2}}(\gamma)$. Assume that $\lambda^{(i)} \in H_{00}^{\frac{1}{2}}$ is known for $i \leq n$. Then iteration n consists in :

$$\left\{ \begin{array}{ll} (1)_n & \text{Let } u_i^{(\lambda^{(n)})} \in H_0^1(\Omega_i) \text{ for } i = 1, 2 \text{ be the solution of (3) with } \lambda = \\ & \lambda^{(n)}, \text{ and let } \Phi(\lambda^{(n)}) \in (H_{00}^{\frac{1}{2}}(\gamma))' \text{ be defined by Formula (4).} \\ (2)_n & \text{Let } v_1^{(\lambda^{(n)})} \text{ be the solution to (5) with } \lambda = \lambda^{(n)} \text{ and let} \\ & S_1(\lambda^{(n)}) \in H_{00}^{\frac{1}{2}}(\gamma) \text{ be the trace of } v_1^{(\lambda^{(n)})} \text{ on } \gamma. \\ (3)_n & \text{Set } \lambda^{(n+1)} = T_{1,\rho}(\lambda^{(n)}). \end{array} \right. \tag{6}$$

The following convergence result holds (the proof of which can be performed by a fixed point theorem applied to the operator $T_{1,\rho}$) :

Theorem 1 *There exists $\rho_0 > 0$ such that if $0 < \rho < \rho_0$ then the sequence $(\lambda^{(n)})_{n\in\mathbb{N}}$ converges in $H_{00}^{\frac{1}{2}}(\gamma)$ towards $\lambda \in H_{00}^{\frac{1}{2}}(\gamma)$ as n tends to infinity, where λ is the trace of the unique weak solution u to Problem (1) on the interface γ.*
Furthermore if $u^{(n)}$ denotes the element of $H_0^1(\Omega)$ such that $u^{(n)}_{|\Omega_i} = u_i^{(\lambda^{(n)})}$, for $i = 1, 2$, the sequence $(u^{(n)})_{n\in\mathbb{N}}$ converges to u in $H_0^1(\Omega)$ as n tends to infinity.

The cell centered finite volume scheme

We now assume that Ω_1 and Ω_2 are polygonal bounded open subsets of $\mathbb{R}^d$, $d = 2, 3$ and the interface $\gamma = \overline{\Omega_1} \cap \overline{\Omega_2}$ is polygonal. The basic principle of the finite volume method is to write

the balance equation associated with (1) over each discretization cell (or "control volume") of the mesh, and use the Stokes formula to obtain: $\int_{\partial K} -\nabla u.\mathbf{n}(s)d\gamma(s) = \int_K f(x)dx$ for any cell K ($\mathbf{n}$ denotes the outward normal unit vector to ∂K).

The finite volume method is known to be well adapted to the discretization of partial differential equations under conservative form. It yields a good approximation of the diffusion fluxes on the cell boundaries and it is quite easy to write and implement, thanks to the balanced form of the equations which is used. Moreover it is well adapted for convection-diffusion equations since the discrete solution satisfies the maximum principle with no condition on the mesh size, see [TGV00]. Since we use here a cell centered scheme, we want to approximate the fluxes $-\nabla u.\mathbf{n}$ on each edge (or face in 3D) of the mesh using the discrete unknowns $(u_K)_{K\in\mathcal{T}}$ associated to the cells. Let $\mathcal{T}$ be a finite volume admissible mesh of Ω in the sense of [REH00], that is roughly speaking (see [REH00] for a precise definition), a set of non intersecting convex polygonals $\{K \in \mathcal{T}\}$ which is such that there exists an associate family of points $\{x_K, K \in \mathcal{T}\}$ such that for any two neighbours K and L the edge (or face) between K and L is orthogonal to the line segment $x_K x_L$. This condition is needed in order to define a consistent approximation of the normal flux $-\nabla u.\mathbf{n}$ through any edge. Meshes satisfying this condition include rectangular and triangular meshes satisfying the Delaunay condition, Voronoï meshes, and mixed meshes with triangular and rectangular cells of this type (see [REH00] or [TGV00]).

Let $\mathcal{T}_i$ be an admissible mesh of Ω_i, for $i = 1, 2$, such that $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$. We denote by $\mathcal{E}_i$ the edges of control volumes of $\mathcal{T}_i$, for $i = 1, 2$. $\Gamma_i = \partial\Omega \bigcap \partial\Omega_i$. We denote by $\mathcal{E}_{i,D}$ the Dirichlet edges of $\mathcal{E}_i$ which are included in $\Gamma_i = \partial\Omega \cap \partial\Omega_i$, for $i = 1, 2$. Since $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$ is an admissible mesh of Ω, one has $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2$. We denote by $\mathcal{E}_\gamma$ the edges of $\mathcal{E}$ which are included in γ and by $\mathcal{E}_{i,int}$ the edges of control volumes of $\mathcal{T}_i$ which are not included in $\partial\Omega_i$, for $i = 1, 2$.

For any $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$ we denote by $d_{K,\sigma}$ the Euclidean distance between x_K and σ. For any $\sigma \in \mathcal{E}$, we define $d_\sigma = d_{K,\sigma} + d_{L,\sigma}$ if $\sigma = K|L \in \mathcal{E}_{int}$ (in which case d_σ is the Euclidean distance between x_K and x_L) and $d_\sigma = d_{K,\sigma}$ if $\sigma \in \mathcal{E}_{ext} \cap \mathcal{E}_K$. For any $\sigma \in \mathcal{E}$, let $\tau_\sigma = meas(\sigma)/d_\sigma$ if $d_\sigma \neq 0$ and $\tau_\sigma = 0$ if $d_\sigma = 0$.

Let $X(\mathcal{T})$ be the set of functions from Ω to $\mathbb{R}$ which are a.e. constant over each control volume of the mesh, and $Y(\gamma)$ the set of functions from γ to $\mathbb{R}$ which are a.e. constant over each edge of the interface γ. Let us denote by λ_σ the value on the edge σ of $\mathcal{E}_\gamma$ of an element λ of $Y(\gamma)$. For a given set of values $(u_K)_{K\in\mathcal{T}}$, we shall denote by $u_\mathcal{T}$ the corresponding piecewise constant function of $X(\mathcal{T})$ defined a. e. by $u_\mathcal{T}(x) = u_K$ if $x \in K$. For all $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$, we introduce some auxiliary unknowns, namely the numerical fluxes, $F_{K,\sigma}(u_\mathcal{T})$ and for all $\sigma \in \mathcal{E}$ some approximation of u on edge σ, denoted by u_σ. The cell centered finite volume scheme for the approximation of Problem (1) writes:

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(u_\mathcal{T}) = meas(K)f_K, \ \forall K \in \mathcal{T}, \tag{7}$$

where the discrete fluxes $F_{K,\sigma}$ are defined with respect to the discrete unknowns as follows:

$$F_{K,\sigma}(u_\mathcal{T}) = -F_{L,\sigma}(u_\mathcal{T}), \ \forall\sigma \in \mathcal{E}_{int}, \text{ if } \sigma = K|L, \tag{8}$$

$$F_{K,\sigma}(u_{calT})\ d_{K,\sigma} = -meas(\sigma)(u_\sigma - u_K), \ \forall\sigma \in \mathcal{E}_K, \ \forall K \in \mathcal{T}, \tag{9}$$

$$u_\sigma = 0, \forall\sigma \in \mathcal{E}_{ext}, \tag{10}$$

and or all $K \in \mathcal{T}$, $f_K = \frac{1}{meas(K)} \int_K f(x)dx$.

For $i = 1, 2$, let $Z(\mathcal{T}_i, \gamma)$ be the space of functions defined a.e. on $\Omega_i \cup \gamma$ which are constant on the control volumes of $\mathcal{T}_i$ and on the edges of $\mathcal{E}_\gamma$. We then define on $Z(\mathcal{T}_i, \gamma)$ the following bilinear form

$$\begin{array}{rl} (z^{(\lambda)}_{i,\mathcal{T}_i}, z^{(\mu)}_{i,\mathcal{T}_i})_{Z(\mathcal{T}_i,\gamma)} = & \sum_{\sigma\in\mathcal{E}_{i,int},\, \sigma=K|L} \tau_\sigma (w^{(\lambda)}_{i,K} - w^{(\lambda)}_{i,L})(w^{(\mu)}_{i,K} - w^{(\mu)}_{i,L}) \\ & + \sum_{\sigma\in\mathcal{E}_{i,D}} \tau_\sigma w^{(\lambda)}_{i,K_\sigma}\, w^{(\mu)}_{i,K_\sigma} + \\ & \sum_{\sigma\in\mathcal{E}_\gamma} \tau_{i,\sigma} (\lambda_\sigma - w^{(\lambda)}_{i,K_{i,\sigma}})(\mu_\sigma - w^{(\mu)}_{i,K_{i,\sigma}}), \end{array} \tag{11}$$

where $z^{(\lambda)}_{i,\mathcal{T}_i}$ (respectively $z^{(\mu)}_{i,\mathcal{T}_i}$) is the element of $Z(\mathcal{T}_i, \gamma)$ defined a.e. on each $K \in \mathcal{T}_i$ and each $\sigma \in \mathcal{E}_\gamma$ by $z^{(\lambda)}_{i,\mathcal{T}_i}(x) = w^{(\lambda)}_{i,K}$ if $x \in K$, $z^{(\lambda)}_{i,\mathcal{T}_i}(x) = \lambda_\sigma$(resp. μ_σ) if $x \in \sigma$.

The space $Y(\gamma)$ is then endowed with the following inner products:

$$(\lambda, \mu)_{i,Y(\gamma)} = (z^{(\lambda)}_{i,\mathcal{T}_i}, z^{(\mu)}_{i,\mathcal{T}_i})_{Z(\mathcal{T}_i,\gamma)},\ \forall \lambda \in Y(\gamma),\ \forall \mu \in Y(\gamma),\ \text{for } i = 1, 2, \tag{12}$$

where $z^{(\lambda)}_{i,\mathcal{T}_i}$ (respectively $z^{(\mu)}_{i,\mathcal{T}_i}$) is the element of $Z(\mathcal{T}_i, \gamma)$ such that $z^{(\lambda)}_{i,\mathcal{T}_i}(x) = w^{(\lambda)}_{i,K}$ if $x \in K$, $z^{(\lambda)}_{i,\mathcal{T}_i}(x) = \lambda_\sigma$(resp. μ_σ) if $x \in \sigma$, and $(w^{(\lambda)}_{i,K})_{K\in\mathcal{T}_i}$, $(w^{(\lambda)}_{i,\sigma})_{\sigma\in\mathcal{E}_\gamma}$ is the unique solution of the following problem :

$$w_{i,\sigma} = 0\ \forall \sigma \in \mathcal{E}_{i,D}. \tag{13}$$

$$w_{i,\sigma} = \lambda_\sigma\ (\text{resp. } w_{i,\sigma} = \mu_\sigma),\ \forall \sigma \in \mathcal{E}_\gamma. \tag{14}$$

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(w_{\mathcal{T}_i}) = 0,\ \forall K \in \mathcal{T}, \tag{15}$$

where the numerical fluxes $F_{K,\sigma}(w_{\mathcal{T}_i})$ are defined as in (8)-(9).

The Euclidean norms $\|\cdot\|_{1,Y(\gamma)}$ and $\|\cdot\|_{2,Y(\gamma)}$ are equivalent on the finite dimensional space $Y(\gamma)$. Hence there exist $\alpha_\mathcal{T}$ and $\beta_\mathcal{T} \in \mathbb{R}^*_+$ depending on the open bounded subsets Ω, Ω_i and on the meshes $\mathcal{T}_i$, for $i = 1, 2$ such that $\alpha_\mathcal{T}\|\lambda\|_{1,Y(\gamma)} \le \|\lambda\|_{2,Y(\gamma)} \le \beta_\mathcal{T}\|\lambda\|_{1,Y(\gamma)}$ for all λ in $Y(\gamma)$.

Let us now define the discrete counterparts of the continuous operators of Definition 1.

Definition 2

For any $\lambda \in Y(\gamma)$, let us define $u^{(\lambda)}_{i,\mathcal{T}_i} \in Z(\mathcal{T}_i, \gamma)$ for $i = 1, 2$ such that $u^{(\lambda)}_{i,K})_{K\in\mathcal{T}_i}$ $(u^{(\lambda)}_{i,\sigma})_{\sigma\in\mathcal{E}_\gamma}$ is the unique solution to the following problem:

$$u_{i,\sigma} = 0,\ \forall \sigma \in \mathcal{E}_{i,D}. \tag{16}$$

$$u_{i,\sigma} = \lambda_\sigma,\ \forall \sigma \in \mathcal{E}_\gamma. \tag{17}$$

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(u_{\mathcal{T}_i}) = meas(K) f_K,\ \forall K \in \mathcal{T}_i, \tag{18}$$

Let us denote by $\Phi_\mathcal{T}(\lambda) \in Y(\gamma)$ the jump between the numerical normal fluxes of the discrete solutions $u^{(\lambda)}_{\mathcal{T}_1}$ and $u^{(\lambda)}_{\mathcal{T}_2}$, that is

$$\Phi_{\mathcal{T}}(\lambda)_\sigma = \sum_{i=1}^{2} \tau_{i,\sigma}(\lambda_\sigma - u^{(\lambda)}_{i,K_{i,\sigma}}),\ \forall \sigma \in \mathcal{E}_\gamma. \tag{19}$$

Let $v^{(\lambda)}_{\mathcal{T}_1} \in Z(\mathcal{T}_i,\gamma)$ such that $(v^{(\lambda)}_{1,K})_{K\in\mathcal{T}_1}$, $(v^{(\lambda)}_{1,\sigma})_{\sigma\in\mathcal{E}_\gamma}$ is the unique solution of the following problem

$$v_{1,\sigma} = 0,\ \forall \sigma \in \mathcal{E}_{1,D}. \tag{20}$$

$$F_{K,\sigma}(v_{\mathcal{T}_1}) = -\Phi_{\mathcal{T}}(\lambda)_\sigma,\ \forall \sigma \in \mathcal{E}_\gamma. \tag{21}$$

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(v_{\mathcal{T}_1}) = 0,\ \forall K \in \mathcal{T}_1. \tag{22}$$

We define the discrete trace of $v^{(\lambda)}_{\mathcal{T}_1}$ a.e. on the interface γ by

$$v^{(\lambda)}_1|_\gamma(x) = v^{(\lambda)}_{1,\sigma},\ \forall x \in \sigma,\ \forall \sigma \in \mathcal{E}_\gamma, \tag{23}$$

where $v^{(\lambda)}_{1,\sigma}$ is defined by the equations

$$\tau_{1,\sigma}(v^{(\lambda)}_{1,\sigma} - v^{(\lambda)}_{1,K_{1,\sigma}}) = -F_{K,\sigma}(v^{(\lambda)}_{\mathcal{T}_1}) = \Phi_{\mathcal{T}}(\lambda)_\sigma,\ \forall \sigma \in \mathcal{E}_\gamma. \tag{24}$$

We define the function $S_{\mathcal{T}_1}$ from $Y(\gamma)$ to $Y(\gamma)$ by

$$S_{\mathcal{T}_1}(\lambda) = v^{(\lambda)}_{\mathcal{T}_1|\sigma}. \tag{25}$$

Finally, for $\rho > 0$, let $T_{\mathcal{T}_1,\rho}$ be defined from $Y(\gamma)$ to $Y(\gamma)$ by

$$T_{\mathcal{T}_1,\rho}(\lambda) = \lambda - \rho S_{\mathcal{T}_1}(\lambda). \tag{26}$$

Let us now describe the discrete counterpart of the domain decomposition iteration (6). Let $\lambda^{(0)}$ be a given function of $Y(\gamma)$. Assume that $\lambda^{(i)} \in Y(\gamma)$ is known for $i \le n$.

$$\left\{\begin{array}{ll} (1)_n & \text{Let } u^{(\lambda^{(n)})}_{\mathcal{T}_i} \in Z(\mathcal{T}_i,\gamma) \text{ for } i = 1,2 \text{ solution of (16)-(18) with } \lambda_\sigma = \\ & \lambda^{(n)}_\sigma, \sigma \in \mathcal{E} \text{ and let } \Phi_{\mathcal{T}}(\lambda^{(n)}) \in Y(\gamma) \text{ be defined by Formula (19).} \\ (2)_n & \text{Let } v^{(\lambda^{(n)})}_1 \text{ be the solution to (20)-22) with } \lambda = \lambda^{(n)} \text{ and let} \\ & S_1(\lambda^{(n)}) \in H^{\frac{1}{2}}_{00}(\gamma) \text{ be the trace of } v^{(\lambda^{(n)})}_1 \text{ on } \gamma. \\ (3)_n & \text{Set } \lambda^{(n+1)} = T_{\mathcal{T}_1,\rho}(\lambda^{(n)}). \end{array}\right. \tag{27}$$

Theorem 2 *Let $\mathcal{T}$ be any admissible mesh of Ω, and $\mathcal{T}_i, i = 1,2$ be an admissible mesh of $\Omega_i, i = 1,2$ such that $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_1$. Let $(\lambda^{(n)})_{n\in\mathbb{N}}$ and $(u^{(\lambda^{(n)})}_{\mathcal{T}_i})_{n\in\mathbb{N}}$ be the sequences defined by (27). Let $u_{\mathcal{T}} \in X(\mathcal{T})$ be such that $(u_K)_{K\in\mathcal{T}}$,$(u_\sigma)_{\sigma\in\mathcal{E}}$ is the unique solution to Problem (8)-(7). Let $u_{\mathcal{T}}|_\gamma \in Y(\gamma)$ be defined a.e on γ by $u_{\mathcal{T}}|_\gamma(x) = u_\sigma, \forall \sigma \in \mathcal{E}_\gamma$. Let $u^{(n)}_{\mathcal{T}} \in X(\mathcal{T})$ be defined a.e. on Ω by $u^{(n)}_{\mathcal{T}}(x) = u^{(\lambda^{(n)})}_{i,\mathcal{T}_i}(x)$ if $x \in \Omega_i$ for $i = 1,2$.*

There exists $\rho^{(\mathcal{T})}_0 > 0$ such that if $0 < \rho < \rho^{(\mathcal{T})}_0$, the sequence $(\lambda^{(n)})_{n\in\mathbb{N}}$ converges in $Y(\gamma)$ with either of the norms defined by (12), towards $u_{\mathcal{T}|\gamma}$ as $n \longrightarrow \infty$ and the sequence $(u^{(n)}_{\mathcal{T}})_{n\in\mathbb{N}}$ converges to $u_{\mathcal{T}}$ in $L^2(\Omega)$ as $n \longrightarrow \infty$.

We do not give here the details of the proof of this theorem for reasons of space limitation; we only mention that it is a discrete adaptation of the proof of Theorem 1 and refer to a forthcoming paper for the details.
Let us mention that in a finite element discretisation, by using the inverse inequality, one can prove that the convergence rate does not depend of the mesh size. This result is known as the *finite element uniform extension theorem* (see, for instance [QV99] pp 105-106 and the references therein). Such a result in the finite volume framework is not yet known. Nevertheless, numerical results show that the convergence rate is still independent of the mesh size. It is the goal of on-going work to prove this fact.

Numerical results

Let $\Omega = (-1,1) \times (0,1) = \Omega_1 \bigcup \Omega_2$, $\Omega_1 = (-1,0.5) \times (0,1)$. We consider a 20×20 rectangular regular mesh. We choose the right hand side of Problem (1) so that the exact solution is: $u(x,y) = sin(\pi\, x) \sin(\pi\, y)$. Let:

$$\mathcal{K}_\rho(n) = \frac{\|\lambda^{n+1} - \lambda^n\|}{\|\lambda^n - \lambda^{n-1}\|} \forall n = 1, n_{it}, \text{ and } \mathcal{K}(\rho) = \left(\prod_{n=1}^{nit} \mathcal{K}_\rho(n) \right)^{\frac{1}{nit}}$$

where n_{it} is the total number of iterations performed. The optimal parameter ρ minimizes the Lipschitz constant $\kappa(\rho)$ of the discrete operator $T_{\mathcal{T}_1,\rho}$ defined by (26). In the proof of the Theorem 2, we show that this Lipschitz constant is a polynomial of degree 2 in the variable ρ. In order to automatically compute the optimal parameter, we use the golden section method and approximate $\kappa(\rho)$ by $\mathcal{K}_\rho(n)$ at iteration n.

We present a comparison between this "relaxation" procedure and the method consisting in solving the trace equation by a conjuguate gradient method presented in [LT94] which we shall call "the Schur complement method" int the sequel. Since our relaxation method has a computational cost by iteration greater than the Schur complement method (because of the number of unknowns), we present the error versus the CPU time in seconds, rather than the number of iterations.

The plotted error is defined by the discrete L^2 norm of the difference between λ^n and λ_{exact} . First, one can observe from Figure 1 that the relaxation method behaves as well as the Schur complement method. Moreover, we can remark from Figure 2 that as in the finite element discretization, the convergence rate does not depend on the mesh size.

We now consider a 80×80 rectangular regular mesh and in Table 1 we present results on a decomposition featuring more that 2 subdomains.

It is quite clear from Table 1 that even for this sequential experiment, the CPU time decreases very fast with respect to the mesh size, thanks to the fact that the local systems to be solved decrease in size.

We finally present some results of a parallel implementation which was set up on a Sun Ultra using up to 32 processors, using the PVM communication protocol. One processor is assigned to one subdomain.

We give in Table 2 the parallel efficiency, i.e. the ratio of the CPU time for n processors over the CPU time using 1 processor, using in both cases the n subdomains decomposition method. One may observe a decrease of the efficiency due to fact that the communication cost between processors increases faster than the CPU time decreases with the number of subdomains.

Conclusion

We have shown that the Neumann-Neumann method works well as a relaxation method in the finite volume setting. It would then also be interesting to apply it as a preconditioner in a conjugate gradient iteration. There also remains to prove the independence of the convergence rate of the method with respect to the mesh, and to adapt the proof of convergence for the case of a convection-diffusion equation.

References

[BGLTV89]Jean-François Bourgat, Roland Glowinski, Patrick Le Tallec, and Marina Vidrascu. Variational formulation and algorithm for trace operator in domain decomposition calculations. In Tony Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors, *Domain Decomposition Methods*, pages 3–16, Philadelphia, PA, 1989. SIAM.

[DW95]Maksymilian Dryja and Olof B. Widlund. Schwarz methods of Neumann-Neumann type for three-dimensional elliptic finite element problems. *Comm. Pure Appl. Math.*, 48(2):121–155, February 1995.

[LT94]Patrick Le Tallec. Domain decomposition methods in computational mechanics. In J. Tinsley Oden, editor, *Computational Mechanics Advances*, volume 1 (2), pages 121–220. North-Holland, 1994.

[MQ89]Luisa D. Marini and Alfio Quarteroni. A relaxation procedure for domain decomposition methods using finite elements. *Numer. Math*, (5):575–598, 1989.

[QV99]Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

[REH00]Thierry Gallouët Robert Eymard and Raphaèle Herbin. The finite volume method. In Philippe Ciarlet and Jean-Louis Lions, editors, *Handbook of Numerical Analysis*, pages 713–1020. North Holland, 2000. This paper appeared as a technical report four years ago.

[TGV00]Raphaèle Herbin Thierry Gallouët and Marie-Hélène Vignal. Error estimates for the approximate finite volume solution of convection diffusion equations with general boundary conditions. *SIAM J. Numer. Anal.*, 37(6):1935 – 1972, 2000.

# subdomains	2	4	8	16	32
mesh by subdomains	80×40	40×40	40×20	20×20	10×20
cpu (s)	359.2	221.4	89.4	29.8	22.4

Table 1: CPU time for different numbers of subdomains

# of processors	2	4	8	16	32
mesh by sub-domain	80×40	40×40	40×20	20×20	10×20
CPU	185.64	60.48	16.77	3.18	1.31
Speed-up	1.9348	3.66	5.3334	9.3572	17.102
Efficiency, E_p (%)	96.74	91.50	66.66	58.48	53.44

Table 2: CPU time and efficiency for several decompositions

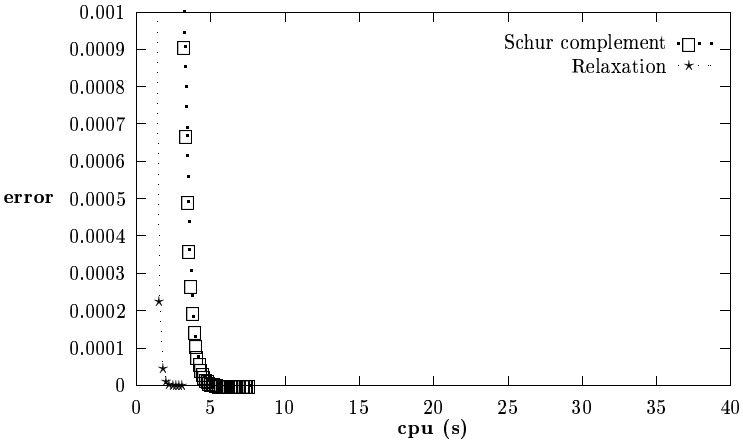


Figure 1: Comparison of the Neumann-Neumann method and the Schur complement method

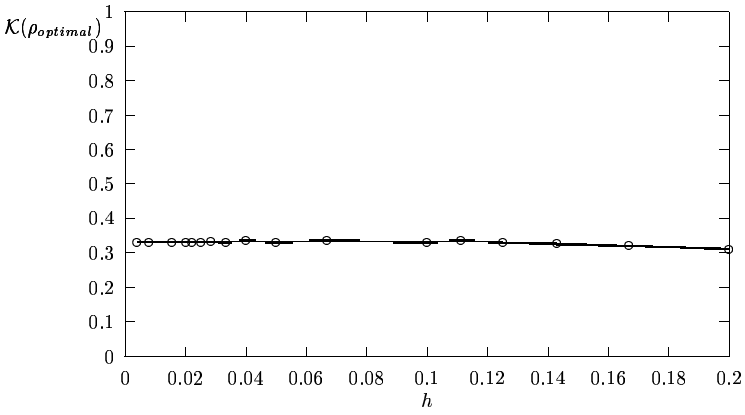


Figure 2: Convergence rate $\mathcal{K}(\rho_{optimal})$ as a function of the discretization step h

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

18 Nonmatching finite volume grids and the nonoverlapping Schwarz algorithm

R. Cautrès[1], R. Herbin[2], F. Hubert[3]

Introduction

We consider the following diffusion-convection problem :

$$\begin{cases} -\Delta u + \mathrm{div}(\mathbf{v}u) + bu = f \ \text{ on } \ \Omega, \\ u = 0 \ \text{ on } \ \partial\Omega, \end{cases} \tag{1}$$

where Ω is an open bounded polygonal subset of $\mathbb{R}^d$, $d = 2, 3$, $\mathbf{v} \in C^1(\Omega, \mathbb{R}^d)$, $b \in L^\infty(\Omega)$, and $f \in L^2(,\Omega)$. The domain Ω is discretized with a grid which may feature some non-matching cells, such as described in Figure 18. Our purpose is first to study a finite volume scheme for Problem (1) on such a mesh and prove an error estimate under adequate assumptions on the unique weak solution to Problem (1). We only study here the case of homogeneous Dirichlet boundary conditions, but Neumann and Robin conditions may also be considered with the technical tools developed in [TGV00].

We then consider the decomposition of Ω in two nonoverlapping domains Ω_1 and Ω_2 and use a discrete version of the Lions adaptation [Lio90] of the Schwarz algorithm in order to solve Problem (1): for a given $\alpha \in \mathbb{R}_+$, choose $u^0 \in H_0^1(\Omega)$, and solve for each $n \geq 0$ and for $i = 1, 2$:

$$\begin{cases} -\Delta u_i^{(n+1)} + \mathrm{div}(\mathbf{v}u_i^{(n+1)}) + bu_i^{(n+1)} = f_i \ \text{ on } \ \Omega_i, \\ u_i^{(n+1)} = 0 \ \text{ on } \ \Gamma_i, \\ -\dfrac{\partial u_i}{\partial n_i}^{(n+1)} + \alpha u_i^{(n+1)} = \dfrac{\partial u_j}{\partial n_j}^{(n)} + \alpha u_j^{(n)} \ \text{ on } \ \gamma, j = 1, 2, i \neq j, \end{cases} \tag{2}$$

where $\Gamma_i = \partial\Omega_i \cap \partial\Omega$, n_i is the normal unit vector to the interface $\gamma = \overline{\Omega_1} \cap \overline{\Omega_2}$ outward to Ω_i and $f_i = f_{|\Omega_i}$.
We present a finite volume version of this algorithm to which the proof of convergence of P.L. Lions may be adapted.

The finite volume scheme

The finite volume method is known to be well adapted to the discretization of partial differential equations under conservative form. It yields a good approximation of the diffusive fluxes on the cell interfaces and it is easy to implement. Our aim here is to study how the method behaves in the presence of non-matching cells such as presented in Figure 18.

[1]University of Marseille, France, Rene.Cautres@cmi.univ-mrs.fr
[2]University of Marseille, France, Raphaele.Herbin@cmi.univ-mrs.fr
[3]University of Marseille, France, Florence.Hubert@cmi.univ-mrs.fr

Let us consider a family $\mathcal{T}$ of grid cells or "control volumes" K, which are open polygonal convex subsets of Ω such that the closure of the union of all the control volumes is $\overline{\Omega}$. In [REH00], it is assumed that there exists a family $(x_K)_{K\in\mathcal{T}}$ (see Figure 18) such that for any two neighbouring cells K and L with common interface $K|L$, the line segment $x_K x_L$ is orthogonal to $K|L$. Here we shall relax this assumption on a number of "atypical cells", the set of which is denoted by $\mathcal{T}_a$. In the sequel, we shall use the following notations:

• for any $K \in \mathcal{T}$, the set of the edges of K is denoted by $\mathcal{E}_K$. The set of the edges of the control volumes of $\mathcal{T}$ is denoted by $\mathcal{E}$, and the set of "interior" (resp. "exterior") edges by $\mathcal{E}_{\rm int} = \{\sigma \in \mathcal{E}; \sigma \notin \partial\Omega\}$ (resp. $\mathcal{E}_{\rm ext} = \{\sigma \in \mathcal{E}; \sigma \subset \partial\Omega\}$).

• for any $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}$, $\mathrm{m}(K)$ is the area (or volume in 3D) of K and $\mathrm{m}(\sigma)$ the length (or area in 3D) of σ. For any $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$ we denote by $d_{K,\sigma}$ the Euclidean distance between x_K and σ.

• for any $\sigma \in \mathcal{E}$, we define $d_\sigma = d_{K,\sigma} + d_{L,\sigma}$ if $\sigma = K|L \in \mathcal{E}_{\rm int}$ and $d_\sigma = d_{K,\sigma}$ if $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}_K$.

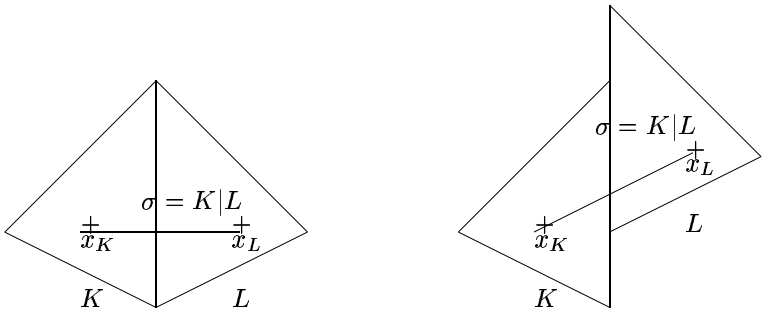


Figure 1: Example of "standard"(left) and 'atypical" (right) control volumes in the 2D triangular case.

Let $X(\mathcal{T})$ be the set of functions from Ω to $\mathbb{R}$ which are constant over each control volume of the mesh. We define a "discrete' H_0^1" norm on $X(\mathcal{T})$ by:

$$\|u\|_{1,\mathcal{T}} = \left(\sum_{\sigma\in\mathcal{E}} \mathrm{m}(\sigma) d_\sigma \left(\frac{D_\sigma u}{d_\sigma}\right)^2 \right)^{\frac{1}{2}}, \tag{3}$$

where, for any $\sigma \in \mathcal{T}$, $D_\sigma u = |u_K - u_L|$ if $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = K|L$, $D_\sigma u = |u_K|$ if $\sigma \in \mathcal{E}_{ext} \cap \mathcal{E}_K$, where u_K denotes the value taken by u on the control volume K.

Let $(u_K)_{K\in\mathcal{T}}$ be the discrete unknowns and let $(u_\sigma)_{\sigma\in\mathcal{E}}$ be a set of values which are expected to be approximations of u on edge σ, for all $\sigma \in \mathcal{E}$. The values u_σ are auxiliary since they may be eliminated from the resulting linear system.

The finite volume scheme is obtained by discretizing the balance equation associated to (1), which writes :

$$-\sum_{\sigma\in\mathcal{E}_K} \int_\sigma \nabla u \cdot \mathbf{n}_{\mathbf{K},\sigma} ds + \sum_{\sigma\in\mathcal{E}_K} \int_\sigma u\mathbf{v} \cdot \mathbf{n}_{\mathbf{K},\sigma} ds + \int_K b\, u dx = \int_K f dx$$

where $\mathbf{n}_{\mathbf{K},\sigma}$ denotes the unit normal vector to $\partial\Omega$ outward to Ω. Let us introduce a set of discrete unknowns $(u_k)_{K\in\mathcal{T}}$, and discrete fluxes $(F_{K,\sigma})_{K\in\mathcal{T}}$ which are the numerical ap-

proximations of $\int_\sigma -\nabla u \cdot \mathbf{n_{K,\sigma}} ds$ by a finite difference approximation. In order to discretize the convection term $\mathrm{div}(\mathbf{v}(x)u(x))$ in a stable way, let us define the upstream choice $u_{\sigma+}$ of u on an edge σ with respect to $\mathbf{v}$ in the following way. For $K \in \mathcal{T}$ and $\sigma \in \mathcal{E}_K$, let $\mathbf{n}_{K,\sigma}$ denote the normal unit vector to σ outward to K and $\mathbf{v}_{K,\sigma} = \int_\sigma \mathbf{v} \cdot \mathbf{n}_{K,\sigma} ds$.

If $\mathbf{v}_{K,\sigma} \geq 0$ and $\sigma \in \mathcal{E}_K$ then $u_{\sigma+} = u_K$. If $\mathbf{v}_{K,\sigma} < 0$, $\sigma \in \mathcal{E}_{\rm int}$ and $\sigma = K|L$ then $u_{\sigma+} = u_L$. If $\mathbf{v}_{K,\sigma} < 0$ and $\sigma \in \mathcal{E}_{ext}$, then $u_{\sigma+} = u_\sigma$.

Let $f_K = \frac{1}{\mathrm{m}(K)} \int_K f dx$ and $b_K = \frac{1}{\mathrm{m}(K)} \int_K b dx$. Then with the notations defined above, a discretization by a cell centered finite volume method yields the following scheme:

$$\sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma} + \sum_{\sigma \in \mathcal{E}_K} \mathbf{v}_{K,\sigma} u_{\sigma+} + b_K \, \mathrm{m}(K) u_K = \mathrm{m}(K) f_K, \; \forall K \in \mathcal{T}, \tag{4}$$

where:

$$F_{K,\sigma} \, d_{K,\sigma} = -\mathrm{m}(\sigma)(u_\sigma - u_K), \; \forall \sigma \in \mathcal{E}_K, \; \forall K \in \mathcal{T} \tag{5}$$

$$F_{K,\sigma} = -F_{L,\sigma}, \; \forall \sigma \in \mathcal{E}_{\rm int}, \;\; \text{if } \sigma = K|L, \tag{6}$$

and

$$u_\sigma = 0, \forall \sigma \in \mathcal{E}_{ext}. \tag{7}$$

Note that the unknowns $(u_\sigma)_{\sigma \in \mathcal{E}}$ may be eliminated by using (6) and (5).

Error estimate

We now present error estimates in the discrete H_0^1 norm under some regularity assumptions on the solution to Problem (1). Some similar results are also in [ELV91] for rectangular meshes and some recent work of F. Nataf *et al* with a different computation of the diffusion fluxes on the atypical interfaces (see these proceedings). The analysis of the scheme is carried out under the following assumptions:

$$\begin{cases} \Omega \text{ is a polygonal open bounded subset of } \mathbb{R}^d. \\ f \in L^2(\Omega). \\ \mathbf{v} \in C^1(\Omega, \mathbb{R}^d). \\ b \in L^\infty(\Omega). \\ \frac{1}{2}\mathrm{div}\mathbf{v}(x) + b(x) \geq 0, \text{ a.e. } x \in \Omega. \end{cases} \tag{8}$$

Theorem 1 *Under Assumptions (8) , let $(u_K)_{K \in \mathcal{T}}$ be the solution to (6)-(4). Assume that the unique variational solution u of Problem (1) satisfies $u \in C^2(\overline{\Omega})$. Let $e_\mathcal{T} \in X(\mathcal{T})$ be defined by $e_\mathcal{T}(x) = e_K = u(x_K) - u_K$ a.e. $x \in K$, $K \in \mathcal{T}$. Then, there exists $C > 0$, only depending on u, $\mathbf{v}$, b, d and Ω, such that*

$$\|e_\mathcal{T}\|_{1,\mathcal{T}} \leq C \left(\mathrm{size}(\mathcal{T}) + \left(\sum_{K \in \mathcal{T}_a} \mathrm{m}(K) \right)^{\frac{1}{2}} \right), \tag{9}$$

where $\mathrm{size}(\mathcal{T}) = \sup\{\mathrm{diam}(K),\ K \in \mathcal{T}\}$. *Furthermore:*

$$\|e_{\mathcal{T}}\|_{L^2(\Omega)} \leq C \left(\mathrm{size}(\mathcal{T}) + \left(\sum_{K \in \mathcal{T}_a} \mathrm{m}(K) \right)^{\frac{1}{2}} \right). \tag{10}$$

If we now assume that the unique variational solution u *to (1) only belongs to* $H^2(\Omega)$ *then (9) and (10) still hold with* C *only depending on* u, $\mathbf{v}$, b, Ω, d *and* $\zeta = \min_{K \in \mathcal{T}} \min_{\sigma \in \mathcal{E}_K} \dfrac{d_{K,\sigma}}{\mathrm{diam}(K)}$.

The proof of this theorem is an adaptation of the techniques used in the case of an admissible mesh [Her95] (see also [REH00]) and will be presented in a forthcoming paper.

The main ingredients in the proof of convergence are the conservativity of the fluxes, i.e. $F_{K,\sigma} = -F_{L,\sigma}$ for two neighbouring cells (K, L), and the consistency of the approximation of the fluxes by finite differences. In the case of an atypical edge, the conservativity holds, but the consistency is lost on the diffusion flux because of the missing orthogonality condition. However, the approximation of the convective flux is still consistent.

If the number of "atypical" control volumes of $\mathcal{T}_a$ is of order $card(\mathcal{T})^{1/2}$ (this is the case for instance if the atypical cells neighbour a the interface between the subdomains of a given domain decomposition), then Inequality (9) (resp. (10)) yields an estimate of order $\frac{1}{2}$ for the discrete H_0^1 norm (resp. L^2 norm) of the error on the solution; numerical results (see section 5) seem to show that this estimate is not sharp. In fact for special examples of atypical cells, we have been able to obtain an order 1.

The discrete algorithm

We shall consider here a nonoverlapping domain decomposition of Ω, under the following assumptions

$$\begin{cases} \Omega_1 \text{ and } \Omega_2 \text{ are polygonal bounded connected open subsets of } \mathbb{R}^d. \\ \overline{\Omega} = \overline{\Omega_1} \cup \overline{\Omega_2}. \\ \text{The interface } \gamma = \overline{\Omega_1} \cap \overline{\Omega_2} \text{ is polygonal and has a non zero} \\ \text{measure in } \mathbb{R}^{d-1}. \\ \Gamma_i = \partial\Omega \cap \partial\Omega_i, \text{ for } i = 1, 2. \\ \text{For } i = 1, 2, \text{ the mesh } \mathcal{T}_i \text{ is an admissible mesh of } \Omega_i \text{ which is the} \\ \text{restriction of the mesh } \mathcal{T} \text{ to } \Omega_i. \end{cases} \tag{11}$$

For $i = 1, 2$, the set of edges (resp. interior edges, resp. exterior edges) of the mesh $\mathcal{T}_i$ is denoted by $\mathcal{E}_i$ (resp. $\mathcal{E}_{i,ext}$, resp. $\mathcal{E}_{i,ext}$). We define $\mathcal{E}_{i,D} = \{\sigma \in \mathcal{E}_i,\ \sigma \subset \Gamma_i\}$ (Dirichlet edges) and $\mathcal{E}_\gamma = \{\sigma \in \mathcal{E}_i,\ \sigma \subset \gamma\}$ (interface edges), with $\mathcal{E}_{i,ext} = \mathcal{E}_{i,D} \cup \mathcal{E}_\gamma$.

The discrete version of the algorithm defined by equations (2) is then:
Given $u_{\mathcal{T}}^{(0)} \in X(\mathcal{T})$ and assuming $u_{\mathcal{T}}^{(k)} \in X(\mathcal{T})$ for $1 \leq k \leq n$ to be known, let $u_{\mathcal{T}_i}^{(k)}$ be the element of $X(\mathcal{T}_i)$, defined for $i = 1, 2$ by: $u_{\mathcal{T}_i}^{(k)}(x) = u_{\mathcal{T}}^{(k)}|_{\Omega_i}(x)$, a.e. $x \in \Omega_i$ and $u_{i,K}^{(k)} = u_{\mathcal{T}_i}^{(k)}|_K$ for a.e. $x \in K$, for $K \in \mathcal{T}_i$. We compute $u_{\mathcal{T}}^{(n+1)} \in X(\mathcal{T})$ defined by $u_{\mathcal{T}}^{(n+1)}(x) = u_{i,K}^{(n+1)}$, for a.e. $x \in K$, for any $K \in \mathcal{T}_i$, $i = 1, 2$, where $\left(u_{i,K}^{(n+1)}\right)_{K \in \mathcal{T}_i}$ is

the unique solution to the following problem:

$$\sum_{\sigma\in\mathcal{E}_K} F_{i,K,\sigma}^{(n+1)} + \sum_{\sigma\in\mathcal{E}_K} \mathbf{v}_{K,\sigma} u_{i,\sigma+}^{(n+1)} + b_K \mathrm{m}(K) u_{i,K}^{(n+1)} = \mathrm{m}(K) f_K, \, \forall K \in \mathcal{T}_i, \tag{12}$$

with

$$F_{i,K,\sigma}^{(n+1)} d_{K,\sigma} = -\mathrm{m}(\sigma)(u_{i,\sigma}^{(n+1)} - u_{i,K}^{(n+1)}), \, \forall \sigma \in \mathcal{E}_K, \, \forall K \in \mathcal{T}_i, \tag{13}$$

$$F_{i,K,\sigma}^{(n+1)} = -F_{i,L,\sigma}^{(n+1)}, \, \forall \sigma \in \mathcal{E}_{i,int}, \text{ if } \sigma = K|L, \tag{14}$$

$$u_{i,\sigma}^{(n+1)} = 0, \, \forall \sigma \in \mathcal{E}_{i,D}, \tag{15}$$

and

$$-F_{i,K,\sigma}^{(n+1)} + \alpha \, u_{i,K}^{(n+1)} = F_{j,L,\sigma}^{(n)} + \alpha \, u_{j,L}^{(n)}, \, \forall \sigma \in \mathcal{E}_\gamma, \text{ for } j = 1, 2, j \neq i. \tag{16}$$

where for $K \in \mathcal{T}_i$ and $\sigma \in \mathcal{E}_K$, $F_{i,K,\sigma}^{(n)} = \mathrm{m}(\sigma) \dfrac{u_\sigma^{(n)} - u_K^{(n)}}{d_{K,\sigma}}$, and where $u_{i,\sigma+}^{(n+1)}$ is defined by the usual upstream scheme if σ is an interior edge, and by the following upstream choice if $\sigma \in \mathcal{E}_\gamma$ lies on the interface:

if $\sigma = K|L$ with $K \in \mathcal{T}_i$ and $L \in \mathcal{T}_j$, $j = 1, 2$, $j \neq i$, we choose $u_{i,\sigma+}^{(n+1)} = u_{i,K}^{(n+1)}$ if $\mathbf{v}_{K,\sigma} \geq 0$ and $u_{i,\sigma+}^{(n+1)} = u_{j,L}^{(n)}$ if $\mathbf{v}_{K,\sigma} < 0$.

Theorem 2 *Under Assumptions (8) and (11), the sequence* $\left(u_{\mathcal{T}}^{(n)}\right)_{n\in\mathbb{N}}$ *defined by the discrete algorithm (14)-(12) converges in* $L^2(\Omega)$ *towards* $u_{\mathcal{T}}$*, the unique solution to Problem (6)-(4)*

The proof of this theorem is an adaptation of the proof of Lions [Lio90] in a discrete finite volume setting.

Numerical results

Let us first study the convergence of the finite volume discretization for a set featuring some atypical cells. In Figure 2, the domain Ω is meshed with a coarse rectangular mesh on the left and a fine rectangular mesh on the right; the set $\mathcal{T}_a$ of atypical edges is such that $\sum_{K\in\mathcal{T}_a} m(K) \leq Ch$, thanks to assumptions on the mesh; hence for this case the result of Theorem 2 is an estimate of order $h^{1/2}$ where h is the maximum step size of the mesh. However, the numerical results show that when the mesh step decreases (with constant ratio between coarse and fine mesh), then the order of convergence behaves like 2 in the L^2 norm and 1 in the discrete H^1 norm; this shows that the error estimate is non optimal.

In Figure 3, we show the influence of the parameter α on the convergence of the Lions algorithm (12)-(15). The optimal parameter is roughly .85 and numerical results which are not shown here because of space limitations show that it is independent of the mesh size.

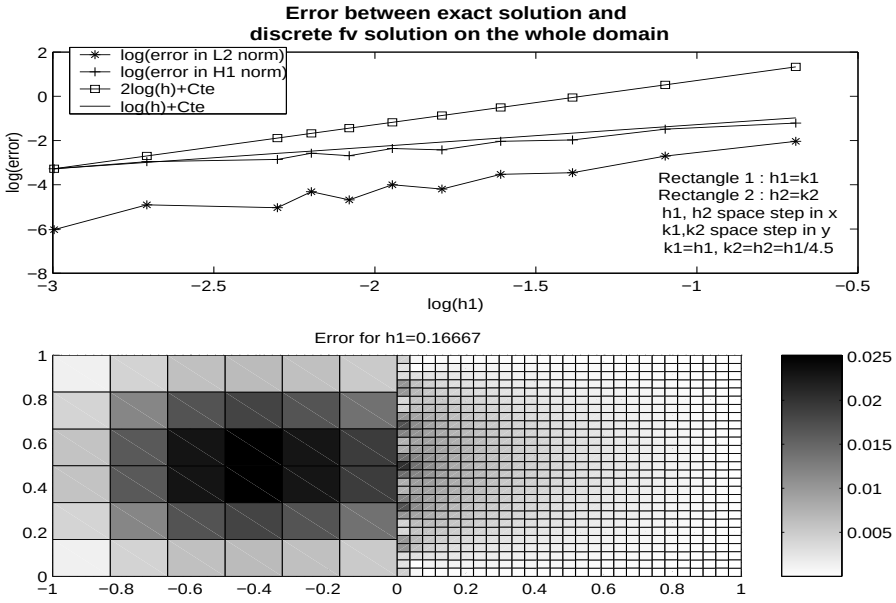


Figure 2: Convergence rate of the finite volume discretization

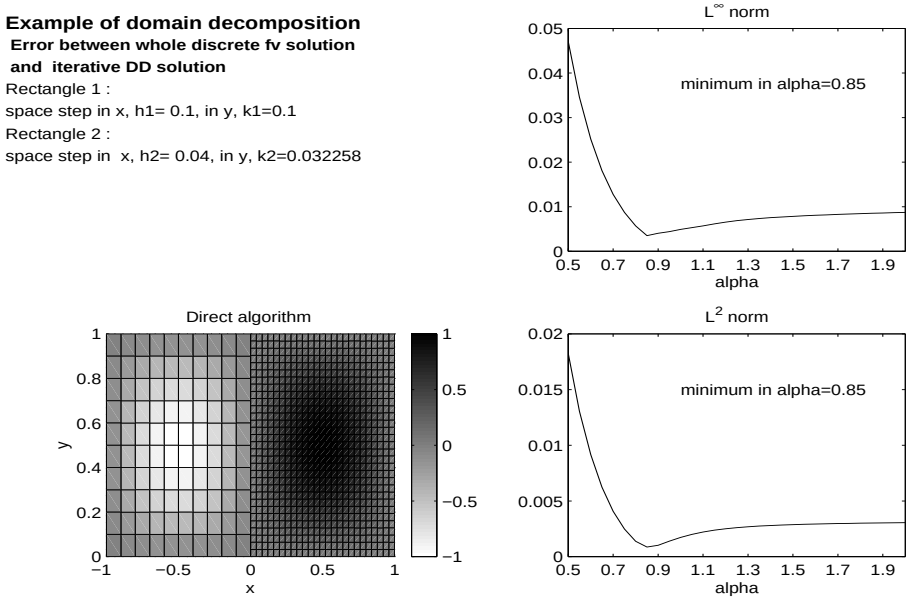


Figure 3: Optimal coefficient α

The solution by a direct solve of the finite volume system (4)-(7) is presented on the grid.

Finally, Figure 4 shows the difference between the solution of the direct solve of system refered to as "Direct Algorithm" and the solution by the domain decomposition algorithm for a relative maximum error of 10^{-3}. It is clear that the error is concentrated at the interface where the atypical meshes are located.

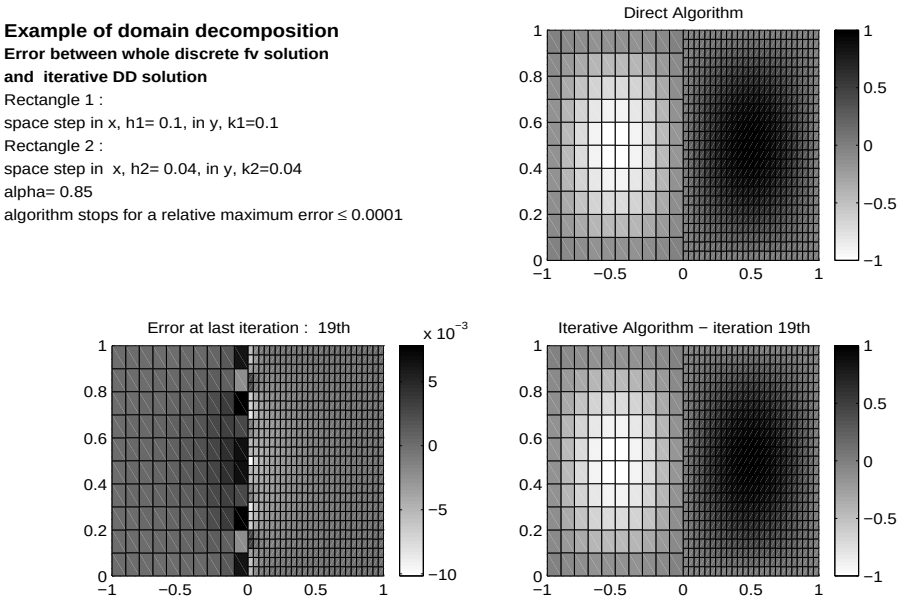


Figure 4: Error between domain decomposition and direct solve

References

[ELV91]R.E. Ewing, R.D. Lazarov, and P.S. Vassilevski. Local refinement techniques for elliptic problems on cell-centered grids. i: Error analysis. *Math. Comput.*, 56(194):437–461, 1991.

[Her95]Raphaèle Herbin. An error estimate for a finite volume scheme for a diffusion-convection problem on a triangular mesh. *Numer. Methods Partial Differ. Equations*, 11(2):165–173, 1995.

[Lio90]Pierre-Louis Lions. On the Schwarz alternating method. III: a variant for nonoverlapping subdomains. In Tony F. Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors, *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations , held in Houston, Texas, March 20-22, 1989*, Philadelphia, PA, 1990. SIAM.

[REH00]Thierry Gallouët Robert Eymard and Raphaèle Herbin. The finite volume method. In Philippe Ciarlet and Jean-Louis Lions, editors, *Handbook of Numerical Analysis*, pages 713–1020. North Holland, 2000. This paper appeared as a technical report four years ago.

[TGV00]Raphaèle Herbin Thierry Gallouët and Marie-Hélène Vignal. Error estimates for the approximate finite volume solution of convection diffusion equations with general boundary conditions. *SIAM J. Numer. Anal.*, 37(6):1935 – 1972, 2000.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

19 The Mortar Element Method for the Rotated $Q1$ Element

Jinru Chen[1], Xuejun Xu [2]

Introduction

Many authors have made significant contributions to the so-called mortar element method (see [4] [5] [7] [8] [10] [11], and references therein). The mortar element method is a non-conforming domain decomposition method with non-overlapping subdomains. The meshes on different subdomains need not align across subdomain interfaces, and the matching of discretizations on adjacent subdomains is only enforced weakly. This offers the advantages of freely choosing highly varying mesh sizes on different subdomains and is very promising to approximate the problems with abruptly changing diffision coefficients or local anisotropies.

The rotated $Q1$ element is an important nonconforming element. It was first proposed and analysed in [12] for numerically solving the Stokes problem. The rotated $Q1$ element provides the simplest example of discretely divergence-free nonconforming element on quadrilaterals. Due to its simplicity, the rotated $Q1$ element is used to simulate the deformation of martensitic crystals with microstructure in [9]. Independently, it also was derived within the framwork of mixed element method (see [2]). In [2] it was proven that Raviart-Thomas mixed rectangle element method is equivalent to rotated $Q1$ nonconforming element method.

The purpose of this paper is to study the rotated $Q1$ mortar element method. A mortar element version for rotated Q1 element is proposed. By constructing some relations between rotated $Q1$ mortar element and bilinear element, the optimal error estimate for rotated $Q1$ mortar element method is proven.

For convenience, the symbols $\preceq$, $\succeq$, and $\asymp$ will be used in this paper, and $x_1 \preceq y_1$, $x_2 \succeq y_2$, and $x_3 \asymp y_3$ mean that $x_1 \leq C_1 y_1$, $x_2 \geq c_2 y_2$, and $c_3 x_3 \leq y_3 \leq C_3 x_3$ for some constants C_1, c_2, c_3, and C_3 that are independent of mesh parameters. For any subdomain $D \subset \Omega$, we use usual L^2 inner product $(\cdot,\cdot)_D$, Sobolev space $H^s(D)$ with usual Sobolev norm $\|\cdot\|_{H^s(D)}$ and seminorm $|\cdot|_{H^s(D)}$. If $D = \Omega$, we denote the usual L^2 inner product by $(\cdot,\cdot)$, the Sobolev norm by $\|\cdot\|_s$ and seminorm by $|\cdot|_s$, where s may be fractional (for details see [1]).

Preliminaries

Consider the following model problem: find $u \in H_0^1(\Omega)$ such that

$$a(u,v) = f(v), \quad \forall v \in H_0^1(\Omega), \tag{1}$$

[1]Department of Mathematics, Nanjing Normal University, Nanjing, 210097, P.R. China, e-mail: jrchen@pine.njnu.edu.cn. This work was supported by the national natural science foundation of China under grant 19901014.

[2]Institute of Computational Mathematics, Academy of Mathematices and System Sciences, Chinese Academy of Sciences, P.O.Box 2719, Beijing 100080, P.R. China, e-mail: xxj@lsec.cc.ac.cn. This work was subsidized by the special funds for major state basic research projects.

where

$$a(u,v)=(\nabla u,\nabla v),\quad f(v)=(f,v),$$

$f\in L^2(\Omega)$, Ω is a rectangular or L-shape bounded domain.

Divide Ω into geometrically conforming rectangular substructures, i.e., $\bar{\Omega}=\bigcup_{k=1}^{N}\bar{\Omega}_k$ with $\bar{\Omega}_k\cap\bar{\Omega}_l$ being empty set or a vertex or an edge for $k\neq l$. With each Ω_k we associate a quasi-uniform triangulation $\mathcal{T}_h(\Omega_k)$ made of elements that are rectangles whose edges are parallel to x-axis or y-axis. The mesh parameter h_k is the diameter of the largest element in $\mathcal{T}_h(\Omega_k)$. Let Γ_{kl} denote the open edge that is common to Ω_k and Ω_l. Denote by Γ the set of all interfaces between the subdomains, i.e., $\Gamma=\bigcup\partial\Omega_k\backslash\partial\Omega$. Each edge inherits two triangulations made of segments that are edges of elements of the triangulations of Ω_k and Ω_l respectively. In this way each Γ_{kl} is provided with two independent and different one dimensional meshes, which are denoted by $\mathcal{T}_h^k(\Gamma_{kl})$ and $\mathcal{T}_h^l(\Gamma_{kl})$ respectively. Let $\Omega_{k,h}$ and $\partial\Omega_{k,h}$ be the sets of vertices of the triangulation $\mathcal{T}_h(\Omega_k)$ that are in $\overline{\Omega}_k$ and $\partial\Omega_k$ respectively.

For each triangulation $\mathcal{T}_h(\Omega_k)$, the rotated $Q1$ element space is defined by

$$\begin{aligned}X_h(\Omega_k)=\{v\in L^2(\Omega_k)\quad|\quad & v|_E=a_E^1+a_E^2x+a_E^3y+a_E^4(x^2-y^2),\\ & a_E^i\in\mathcal{R},\ \int_{\partial E\cap\partial\Omega}v|_{\partial\Omega}ds=0,\ \ \forall E\in\mathcal{T}_h(\Omega_k);\\ & \text{for }\ E_1,E_2\in\mathcal{T}_h(\Omega_k),\ \text{ if }\ \partial E_1\cap\partial E_2=e,\ \text{ then}\\ & \int_e v|_{\partial E_1}ds=\int_e v|_{\partial E_2}ds\},\end{aligned}$$

with norm and seminorm

$$\|v\|_{H_h^1(\Omega_k)}=(\sum_{E\in\mathcal{T}_h(\Omega_k)}\|v\|^2_{H^1(E)})^{1/2},\quad |v|_{H_h^1(\Omega_k)}=(\sum_{E\in\mathcal{T}_h(\Omega_k)}|v|^2_{H^1(E)})^{1/2}.$$

Introduce the global discrete space

$$X_h(\Omega)=\prod_{k=1}^{N}X_h(\Omega_k),$$

with norm $\|v\|_{1,h}=(\sum_{k=1}^{N}\|v\|^2_{H_h^1(\Omega_k)})^{1/2}$ and seminorm $|v|_{1,h}=(\sum_{k=1}^{N}|v|^2_{H_h^1(\Omega_k)})^{1/2}$.

Define one of the sides of Γ_{kl} as mortar denoted by $\gamma_{m(k)}$ and the other as nonmortar denoted by $\delta_{m(l)}$. Assume that the mortar for $\gamma_{m(k)}=\delta_{m(l)}=\Gamma_{kl}$ is chosen by the condition $h_k\leq h_l$, i.e., the fine side is chosen as mortar. Based on this assumption, the two elements of the slave triangulation $\mathcal{T}_h^l(\delta_{m(l)})$ that touch the ends of $\delta_{m(l)}$ are longer than the respective elements of the mortar triangulation $\mathcal{T}_h^k(\gamma_{m(k)})$. Define an auxiliary test space $M^{h_l}(\delta_{m(l)})$ to be a subspace of the space $L^2(\Gamma_{kl})$ such that its functions are piecewise constants on $\mathcal{T}_h^l(\delta_{m(l)})$. The dimension of $M^{h_l}(\delta_{m(l)})$ is equal to the number of elements on the $\delta_{m(l)}$. For each nonmortar $\delta_{m(l)}=\Gamma_{kl}$, we define an L^2-orthogonal projection $Q_m: L^2(\Gamma_{kl})\to M^{h_l}(\delta_{m(l)})$ by

$$(Q_mv,w)_{L^2(\delta_{m(l)})}=(v,w)_{L^2(\delta_{m(l)})},\quad\forall w\in M^{h_l}(\delta_{m(l)}).\tag{2}$$

Now we define rotated $Q1$ mortar element space

$$V_h = \{v \in X_h(\Omega) \quad | \quad Q_m v_l = Q_m v_k, \quad \forall \delta_{m(l)} = \gamma_{m(k)} \subset \Gamma\},$$

where $v_k = v|_{\gamma_{m(k)}}$ and $v_l = v|_{\delta_{m(l)}}$. The condition of the equality of the L^2-orthogonal projection of traces onto the test space for each interface is called the mortar condition. The rotated $Q1$ mortar element approximation of problem (1) is: find $u_h \in V_h$ such that

$$a_h(u_h, v_h) = (f, v_h), \quad \forall v_h \in V_h, \tag{3}$$

where

$$a_h(u_h, v_h) = \sum_{k=1}^{N} a_{h,k}(u_h, v_h), \quad a_{h,k}(u_h, v_h) = \sum_{E \in \mathcal{T}_h(\Omega_k)} (\nabla u_h, \nabla v_h)_E.$$

Some Technical Lemmas

In this section we present some auxiliary technical lemmas that are necessary to prove our results.

Let $\mathcal{T}_{h/2}(\Omega_k)$ be the partition which is constructed by connecting midpoints of the opposite edges of elements of $\mathcal{T}_h(\Omega_k)$, $\tilde{V}^{h/2}(\Omega_k)$ be piecewise bilinear conforming element space defined on $\mathcal{T}_{h/2}(\Omega_k)$, and $\tilde{V}_0^{h/2}(\Omega_k)$ be the subspace of $\tilde{V}^{h/2}(\Omega_k)$ consisting of functions with zero traces on $\partial\Omega_k$. Define operator $\mathcal{M}_k : X_h(\Omega_k) \to \tilde{V}^{h/2}(\Omega_k)$ as follows:

Definition 1 Given $v \in X_h(\Omega_k)$, we define $\mathcal{M}_k v \in \tilde{V}^{h/2}(\Omega_k)$ by the values of $\mathcal{M}_k v$ at the vertices of the partition $\mathcal{T}_{h/2}(\Omega_k)$. The vertices are divided into four sets of points:

• If P is a central point of E, $E \in \mathcal{T}_h(\Omega_k)$, then

$$(\mathcal{M}_k v)(P) = \frac{1}{4} \sum_{e_i \in \partial E} \frac{1}{|e_i|} \int_{e_i} v ds;$$

• If P is a midpoint of one dege $e \in \partial E$, $E \in \mathcal{T}_h(\Omega_k)$, then

$$(\mathcal{M}_k v)(P) = \frac{1}{|e|} \int_e v ds;$$

• If $P \in \Omega_{k,h} \backslash \partial\Omega_{k,h}$, then

$$(\mathcal{M}_k v)(P) = \frac{1}{4} \sum_{e_i} \frac{1}{|e_i|} \int_{e_i} v ds,$$

where the sum is taken over all edges e_i with the common vertex P, $e_i \in \partial E_i$, $E_i \in \mathcal{T}_h(\Omega_k)$;

• If $P \in \partial\Omega_{k,h}$, then

$$(M_k v)(P) = \frac{|e_l|}{|e_l| + |e_r|} (\frac{1}{|e_l|} \int_{e_l} v ds) + \frac{|e_r|}{|e_l| + |e_r|} (\frac{1}{|e_r|} \int_{e_r} v ds),$$

where $e_l \in \partial E_1 \cap \partial\Omega_k$ and $e_r \in \partial E_2 \cap \partial\Omega_k$ are the left and right neighbor edges of P, E_1, $E_2 \in \mathcal{T}_h(\Omega_k)$. If P is a vertex of Ω_k, then $E_1 = E_2$.

The above operator $\mathcal{M}_k$ has the following properties.

Lemma 1 *For any $v \in X_h(\Omega_k)$, we have*

$$
\begin{aligned}
&|\mathcal{M}_k v|_{H^1(\Omega_k)} \asymp |v|_{H_h^1(\Omega_k)},\\
&\|\mathcal{M}_k v\|_{L^2(\Omega_k)} \asymp \|v\|_{L^2(\Omega_k)},\\
&\int_{\partial\Omega_k} \mathcal{M}_k v ds = \int_{\partial\Omega_k} v ds,\\
&\|\mathcal{M}_k v - v\|_{L^2(\Omega_k)} \preceq h_k |v|_{H_h^1(\Omega_k)},\\
&\|\mathcal{M}_k v - v\|_{L^2(\varepsilon)} \preceq h_k^{1/2} |v|_{H_h^1(\Omega_k)},
\end{aligned}
$$

where ε is an edge of Ω_k.

We now introduce a subspace $X_h^\varepsilon(\Omega_k)$ of $X_h(\Omega_k)$ for each open edge ε of Ω_k as follows:

$$X_h^\varepsilon(\Omega_k) = \{v \in X_h(\Omega_k) \ | \ \int_e v ds = 0, \quad \forall e \in \partial\Omega_k \backslash \varepsilon\}.$$

Define an operator $\mathcal{M}_k^\varepsilon : X_h^\varepsilon(\Omega_k) \to \tilde{V}^{h/2}(\Omega_k)$ by

Definition 2 Given $v \in X_h^\varepsilon(\Omega_k)$, we define $\mathcal{M}_k^\varepsilon v \in \tilde{V}^{h/2}(\Omega_k)$ by the values of $\mathcal{M}_k^\varepsilon v$ at the vertices of the partition $\mathcal{T}_{h/2}(\Omega_k)$.

- If P is a central point of E or a midpoint of one edge of E, $E \in \mathcal{T}_h(\Omega_k)$, or $P \in \Omega_{k,h} \backslash \partial\Omega_{k,h}$, then $(\mathcal{M}_k^\varepsilon v)(P) = (\mathcal{M}_k v)(P)$;
- If $P \in \partial\Omega_{k,h} \backslash \varepsilon$, then $(\mathcal{M}_k^\varepsilon v)(P) = 0$;
- If $P \in \partial\Omega_{k,h} \cap \varepsilon$, then

$$(\mathcal{M}_k^\varepsilon v)(P) = \frac{|e_r|}{|e_l| + |e_r|}\left(\frac{1}{|e_l|}\int_{e_l} v ds\right) + \frac{|e_l|}{|e_l| + |e_r|}\left(\frac{1}{|e_r|}\int_{e_r} v ds\right),$$

where $e_l \in \partial E_1 \cap \partial\Omega_k$ and $e_r \in \partial E_2 \cap \partial\Omega_k$ are the left and right neighbor edges of P, $E_1, E_2 \in \mathcal{T}_h(\Omega_k)$. If P is a vertex of Ω_k, $E_1 = E_2$.

Define the pseudo-inverse map $(\mathcal{M}_k)^+ : \tilde{V}^{h/2}(\Omega_k) \to X_h(\Omega_k)$ by

$$\frac{1}{|e|}\int_e (\mathcal{M}_k)^+ v ds = v(P), \quad \forall v \in \tilde{V}^{h/2}(\Omega_k),$$

where $e \in \partial E$, $E \in \mathcal{T}_h(\Omega_k)$, P is the midpoint of e. Obviously, we have

$$(\mathcal{M}_k)^+ \mathcal{M}_k v = v, \quad (\mathcal{M}_k)^+ \mathcal{M}_k^\varepsilon w = w, \quad \forall v \in X_h(\Omega_k), \ \ \forall w \in X_h^\varepsilon(\Omega_k).$$

Using the discrete norms, we can prove the following Lemma holds.

Lemma 2 *For any $v \in \tilde{V}^{h/2}(\Omega_k)$, we have*

$$|(\mathcal{M}_k)^+ v|_{H_h^1(\Omega_k)} \preceq |v|_{H_h^1(\Omega_k)}, \quad \|(\mathcal{M}_k)^+ v\|_{L^2(\Omega_k)} \preceq \|v\|_{L^2(\Omega_k)}.$$

Let $\mathcal{A}_k$ be a special set of edges which belong to $\partial\Omega_k$ or are the edges of rectangles which have one side on a mortar $\gamma_{m(k)}$. We introduce a special subspace $X_h^k(\Omega_k) \subset X_h(\Omega_k)$ as follows:

$$X_h^k(\Omega_k) = \{v \in X_h(\Omega_k) \ | \ \int_e v ds = 0, \quad \forall e \in \mathcal{A}_k\}.$$

Define a discrete harmonic part $H_k v$ of $v \in X_h(\Omega_k)$ by

$$a_{h,k}(H_k v, w) = 0, \quad \forall w \in X_h^k(\Omega_k),$$
$$\int_e H_k v ds = \int_e v ds, \quad \forall e \in \mathcal{A}_k.$$

Also we define a projection operator $P_k : X_h(\Omega_k) \to X_h^k(\Omega_k)$ by

$$a_{h,k}(P_k v, w) = a_{h,k}(v, w), \quad \forall w \in X_h^k(\Omega_k).$$

Lemma 3 *Let $\varepsilon = \delta_{m(k)}$ be a nonmortar edge of Ω_k, and v be discrete harmonic in Ω_k with $\int_e v ds = 0$ for any $e \in \mathcal{A}_k \backslash \delta_{m(k)}$. Then*

$$|v|_{H_h^1(\Omega_k)} \preceq \|\mathcal{M}_k^{\varepsilon} v\|_{H_{00}^{1/2}(\delta_{m(k)})}.$$

Let $\delta_{m(l)}$ be a nonmortar edge of Ω_l, $W_0^{h_l}(\delta_{m(l)})$ be the continuous function space whose elements are piecewise linear over all segments that have the midpoints of edges belonging to $\delta_{m(l)}$ as their nodals and equal zero at the ends of $\delta_{m(l)}$. Let $\delta_{m(l)}^m$ be the set of midpoints of edges in $\mathcal{T}_h^l(\delta_{m(l)})$. Define an auxiliary operator $\Pi_m : L^2(\delta_{m(l)}) \to W_0^{h_l}(\delta_{m(l)})$ as follows:

$$(\Pi_m v)(P) = (Q_m v)(P), \quad \forall P \in \delta_{m(l)}^m.$$

Lemma 4 $\|\Pi_m v\|_{L^2(\delta_{m(l)})} \preceq \|v\|_{L^2(\delta_{m(l)})}, \forall v \in L^2(\delta_{m(l)})$.

By interpolation estimate [6] and operator interpolation theory in Chapter 12 in [3], we can derive the following result.

Lemma 5 $\|v - Q_m v\|_{L^2(\delta_{m(l)})} \preceq h_l^{1/2} |v|_{H^{1/2}(\delta_{m(l)})}, \forall v \in H^{1/2}(\delta_{m(l)})$.

Error Estimate

The following result is the well-known second Strang Lemma.

Lemma 6 *Let u and u_h be the solutions of (1) and (3) respectively, if $\frac{\partial u}{\partial n} \in L^2(\partial E)$, then*

$$|u - u_h|_{H_h^1(\Omega)} \preceq \inf_{v \in V_h} |u - v|_{H_h^1(\Omega)} + \sup_{w \in V_h} |\sum_{k=1}^{N} \sum_{E \in \mathcal{T}_h(\Omega_k)} \frac{\int_{\partial E} \frac{\partial u}{\partial n} w ds}{|w|_{H_h^1(\Omega)}}|. \tag{4}$$

The first term in (4) is known as the approximation error, while the second term is called the consistency error.

Using Lemmas 1-5, arguing as in [11], we can prove the following two Lemmas.

Lemma 7 *Let u and u_h be the solution of (1) and (3) respectively. Assume $u|_{\Omega_k} \in H^2(\Omega_k)$, then we have*

$$|\sum_{k=1}^{N} \sum_{E \in \mathcal{T}_h(\Omega_k)} \int_{\partial E} \frac{\partial u}{\partial n} w ds| \preceq (\sum_{k=1}^{N} h_k^2 |u|_{H^2(\Omega_k)}^2)^{1/2} |w|_{H_h^1(\Omega)}, \quad \forall w \in V_h.$$

Lemma 8 *For any $u \in H_0^1(\Omega)$ with $u|_{\Omega_k} \in H^2(\Omega_k)$, we have*

$$\inf_{v \in V_h} |u - v|_{H_h^1(\Omega)} \preceq (\sum_{k=1}^{N} h_k^2 |u|^2_{H^2(\Omega_k)})^{1/2}.$$

From Lemmas 6-8 we obtain the following optimal error estimate.

Theorem 1 *Let u and u_h be the solution of (1) and (3) respectively, $u|_{\Omega_k} \in H^2(\Omega_k)$, then*

$$|u - u_h|_{H_h^1(\Omega)} \preceq (\sum_{k=1}^{N} h_k^2 |u|^2_{H^1(\Omega_k)})^{1/2}.$$

References

[1]R.A.Adams, Sobolev Space, Academic Press, New York, 1975.
[2]T.Arbogast and Z.X.Chen, On the implementation of mixed methods as nonconforming methods for second order elliptic problems, Math. Comp. 64 (1995), 943-971.
[3]B.C.Brenner and L.R.Scott, The Mathematical Theory of Finite Element Methods, Springer-Verlag, 1996.
[4]F.B.Belgacem and Y.Maday, The mortar element method for three dimensional finite elements, RAIRO Numer. Anal. 31 (1997), 289-309.
[5]C.Bernardi, Y.Maday and A.Patera, A new nonconforming approach to domain decomposition: the mortar element method, In Nonlinear Partial Differential Equations and their Applications, College de France Seminar, Vol. XI, H.Brezis and I.L.Lions, eds., Pitman, 1994, 13-51.
[6]P.G.Ciarlet, The Finite Element Method for Elliptic Problems, North-Holland, New York, 1978.
[7]M.A.Casarin and O.B.Widlund, A hierarchical preconditioner for the mortar finite element method, ETNA, 4 (1996), 75-88.
[8]J.Gopalakrishnan and J.E.Pasciak, Multigrid for the mortar finite element method, SIAM J. Numer. Anal. 37 (2000), 1029-1052.
[9]P.Kloucek, B.Li, and M.Luskin, Analysis of a class of nonconforming finite elements for crystalline microstructure, Math. Comp. 65 (1996), 1111-1135.
[10]L.Marcinkowski, Mortar element method for quasilinear elliptic boundary value problems, East-West J. Numer. Math., 4 (1996), 293-309.
[11]L.Marcinkowski, The mortar element method with locally nonconforming elements, BIT, 39 (1999), 716-739.
[12]R.Rannacher and S.Turek, Simple nonconforming quadrilateral Stokes element, Numer. Meth. Partial Diff. Equations 8 (1992), 97-111.
[13]J.Xu and J.Zou, Some nonoverlapping domain decomposition methods, SIAM Rev. 40 (1998), 857-914.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

20 Overlapping Schwarz Waveform Relaxation for Convection Reaction Diffusion Problems

D.S. Daoud [1], M.J. Gander [2]

Introduction

Overlapping Schwarz waveform relaxation is a long name for an algorithm which simply solves evolution problems in parallel. It got its name as follows: the distribution of the computation is achieved by partitioning the spatial domain into overlapping subdomains, like in the classical Schwarz method. However on subdomains, time dependent problems are solved in the iteration and thus the algorithm is also of waveform relaxation type. Hence the name overlapping Schwarz waveform relaxation. These algorithms have been introduced in [GK97] and independently in [GZ97] for the solution of evolution problems in a parallel environment with slow communication links, since they permit to solve over several time steps before communicating information to the neighboring subdomains. They are ideal when one wants to use large existing networks of PC's with a high latency network but reasonable throughput as a super-computer. An earlier analysis for first order hyperbolic problems of the same type of algorithm can be found in [Bjø95].

These algorithms stand in contrast to the classical approach in domain decomposition for evolution problems, where time is first discretized uniformly using an implicit discretization and then at each time step a problem in space only is solved using domain decomposition, see for example [Meu91] and [Cai91, Cai94]. The main disadvantage of the classical approach is that one is forced to use the same time step in all subdomains and thus looses one of the main features of domain decomposition, namely to treat subdomains numerically differently. A second disadvantage is that one needs to exchange information at each time step. Overlapping Schwarz waveform relaxation is a remedy for both problems.

In this paper we study overlapping Schwarz waveform relaxation for space decompositions in all generality for the linear convection reaction diffusion equation in n dimensions. We prove linear convergence of the algorithm on unbounded time intervals and state a theorem about superlinear convergence on bounded time intervals. Both results hold at the continuous level, which leads to algorithms that converge independently of the mesh size if the overlap is held constant.

Problem Description

We are interested to solve parabolic partial differential equations in n dimensions on a parallel computer with slow communication links. We consider as our guiding example the convection

[1]Dept. of Mathematics, Eastern Mediterranean University, Famagusta, North Cyprus Via mersin 10, TURKEY.
[2]Dept. of Mathematics and Statistics, McGill University, Montreal, QC H3A 2K6, CANADA.

reaction diffusion equation on a bounded domain $\Omega \subset R^n$ with a smooth boundary $\partial\Omega$,

$$\begin{aligned} \mathcal{L}(u) := -\frac{\partial u}{\partial t} + \nu\Delta u + \mathbf{a}\cdot\nabla u + cu &= f(\mathbf{x},t) && \mathbf{x}\in\Omega, \quad 0<t<T, \\ u(\mathbf{x},t) &= g(\mathbf{x},t) && \mathbf{x}\in\partial\Omega, \quad 0<t<T, \\ u(\mathbf{x},0) &= u_0(\mathbf{x}) && \mathbf{x}\in\Omega. \end{aligned} \tag{1}$$

We assume that the initial condition $u_0(\mathbf{x})$ and the boundary condition $g(\mathbf{x},t)$ are bounded piecewise continuous and $f(\mathbf{x},t)$ is continuous. This gives existence and uniqueness of a solution to (1). In our analysis we will use the maximum principle satisfied by the solution $u(\mathbf{x},t)$ of (1):

Theorem 1 (Maximum Principle) *Assume that $\mathcal{L}(u) \geq 0$ ($\mathcal{L}(u) \leq 0$). Let $M = \sup_\Omega u$ ($\inf_\Omega u$). Assume that $u = M$ at some interior point $(\mathbf{x}_0, t_0) \in \Omega$ and that one of the following holds:*

1. *$c = 0$ and M is arbitrary.*
2. *$c \leq 0$ and $M \geq 0$ ($M \leq 0$).*
3. *$M = 0$ and c is arbitrary.*

Then $u = M$ on $\bar{\Omega}\times[0,t_0]$.

Proof The proof can be found in [Lie96]. ■

To distribute the computation, we partition the domain Ω into overlapping subdomains. Such a partition can be obtained by first partitioning Ω into N non-overlapping subdomains $\widetilde{\Omega}_j$ with boundaries $\partial\widetilde{\Omega}_j$, $j = 1, 2, \ldots, N$. We denote the boundaries of the subdomain $\widetilde{\Omega}_j$ interior to the domain Ω by $\widetilde{\Gamma}_j$. Then we construct an overlapping decomposition Ω_j with boundary $\partial\Omega_j$ by enlarging each $\widetilde{\Omega}_j$ so that the boundaries of the new subdomains Γ_j interior to Ω are at least a distance δ away from $\widetilde{\Gamma}_j$. To solve the parabolic problem (1), the overlapping Schwarz waveform relaxation iteration constructs iteratively u_j^{k+1} on each subdomain Ω_j using as the boundary condition the values from the neighboring subdomains u_l^k at the previous iteration. To pass the boundary information, the boundary of Ω_j is decomposed into disjoint subsets Γ_{jl}, $l = 1, \ldots, N$ such that the Euclidean distance of $\mathbf{x} \in \Gamma_{jl}$ from the boundary of Ω_l is at least δ. This is possible because of the way the overlapping decomposition was constructed: we simply use the solutions obtained in Ω_l only within the smaller region $\widetilde{\Omega}_l$. Doing this for each subdomain, we define a complete approximation to the solution at step k on the whole of Ω which can be used at step $k+1$ as boundary condition for the next subdomain solves. We denote also by Γ_{j0} the part of the boundary that subdomain Ω_j shares with the original domain Ω.

Linear Convergence for Unbounded Time Domains

For the convergence analysis, it suffices by linearity to consider the homogeneous problem, $f(\mathbf{x},t) = g(\mathbf{x},t) = u_0(\mathbf{x}) = 0$ in (1) and to analyze convergence to zero. We first consider the case where $T = \infty$ and hence restrict $c \leq 0$ to have bounded solutions. On each subdomain

Ω_j we solve at each step $k+1$ of the overlapping Schwarz waveform relaxation iteration the subproblem

$$\begin{array}{rcll} \mathcal{L}(u^{k+1}) & = & 0 & \mathbf{x} \in \Omega_j, \quad 0 < t < T, \\ u_j^{k+1}(\mathbf{x},t) & = & u_l^k(\mathbf{x},t) & \mathbf{x} \in \Gamma_{jl}, \quad 0 < t < T, \\ u_j^{k+1}(\mathbf{x},t) & = & 0 & \mathbf{x} \in \Gamma_{j0}, \quad 0 < t < T, \\ u_j^{k+1}(\mathbf{x},0) & = & 0 & \mathbf{x} \in \Omega_j, \end{array} \tag{2}$$

for $j = 1, 2, \ldots, N$, using the boundary information from the neighboring subdomains at step k. This corresponds to an additive Schwarz or Jacobi iteration which can be done in parallel. One can also consider a multiplicative Schwarz or Gauss Seidel iteration which would need a special coloring of subdomains to remain a parallel algorithm.

We define the integer distance quantity m_j for each subdomain Ω_j to be the least number of subdomains one has to pass through to touch the boundary $\partial\Omega$, and also the maximum $m := \max_j m_j$. We further define the index sets $I_l := \{j : m_j = l\}$ so that the index set I_l contains the indices of all the subdomains which are within distance l of the boundary. Defining for bounded functions $g(\mathbf{x},t) : \Omega \times [0,\infty) \to \mathbb{R}$ the norm

$$||g(\cdot,\cdot)||_\infty := \sup_{\mathbf{x}\in\Omega, t>0} |g(\mathbf{x},t)|$$

we have the following

Lemma 1 *The iterates of (2) satisfy for $T = \infty$ and $c \leq 0$ the estimate*

$$\max_j ||u_j^{k+m+2}(\cdot,\cdot)||_\infty \leq \gamma(m,\delta) \max_j ||u_j^k(\cdot,\cdot)||_\infty \tag{3}$$

where $\gamma(m,\delta)$ is a number strictly less than one and independent of k.

Proof The idea of the proof is to construct a sequence of elliptic upper bounds on the iterates and then to apply the convergence analysis based on the maximum principle for the elliptic upper bounds in Lions [Lio88]. For k fixed we define $U^k := \max_j ||u_j^k(\cdot,\cdot)||_\infty$ and note that on each subdomain the solution $\tilde{u}_j^{k+1}$ of the elliptic problem

$$\begin{array}{rcll} \nu\Delta\tilde{u}_j^{k+1} + \mathbf{a}\cdot\nabla\tilde{u}_j^{k+1} + c\tilde{u}_j^{k+1} & = & 0 & \mathbf{x} \in \Omega_j, \\ \tilde{u}_j^{k+1}(\mathbf{x}) & = & U^k & \mathbf{x} \in \Gamma_{jl}, \\ \tilde{u}_j^{k+1}(\mathbf{x}) & = & 0 & \mathbf{x} \in \Gamma_{j0} \end{array} \tag{4}$$

is an upper bound on the modulus of u_j^{k+1}. Now $\tilde{u}_j^{k+1}$ satisfies a maximum principle and for $j \in I_0$ we have $\tilde{u}_j^{k+1} < U^k$ in the interior of $\widetilde{\Omega}_j$, since $\tilde{u}_j^{k+1}$ satisfies on part of the boundary of Ω_j a homogeneous boundary condition. Note that for $j \notin I_0$ we have $\tilde{u}_j^{k+1}$ not necessarily strictly less than U^k since $\tilde{u}_j^{k+1}$ might have the value U^k on all its boundaries and thus by the maximum principle $\tilde{u}_j^{k+1} \equiv U^k$. Define

$$U^{k+1} := \sup_{\mathbf{x}\in\widetilde{\Omega}_l, l\in I_0} \tilde{u}_l^{k+1} \leq \gamma_1(\delta) U^k$$

for some constant $\gamma_1(\delta) < 1$. Note that γ_1 depends on the size of the overlap, but not on k since $\tilde{u}_j^{k+1}$ is a linear function of the boundary condition. Now for the next iteration by

definition part of the boundary of subdomains Ω_j with $j \in I_1$ lie strictly within $\widetilde{\Omega}_l$ with $l \in I_0$ and therefore for $j \in I_1$ the solution $\tilde{u}_j^{k+2}$ of the elliptic problem

$$\begin{aligned} \nu\Delta\tilde{u}_j^{k+2} + \mathbf{a}\cdot\nabla\tilde{u}_j^{k+2} + c\tilde{u}_j^{k+2} &= 0 \qquad \mathbf{x} \in \Omega_j, \\ \tilde{u}_j^{k+2}(\mathbf{x}) &= U^k \qquad \mathbf{x} \in \Gamma_{jl},\ l \notin I_0, \\ \tilde{u}_j^{k+2}(\mathbf{x}) &= U^{k+1} \qquad \mathbf{x} \in \Gamma_{jl},\ l \in I_0 \end{aligned} \tag{5}$$

is an upper bound on the modulus of u_j^{k+2}. Since $U^{k+1} \leq \gamma_1(\delta)U^k$ we have by the maximum principle $\tilde{u}_j^{k+2} < U^k$ in $\widetilde{\Omega}_j$ and defining U^{k+2} similarly to U^{k+1} before, we find $U^{k+2} \leq \gamma_2(\delta)U^k$ for some constant $\gamma_1(\delta) \leq \gamma_2(\delta) < 1$ independent of k. By induction we find at step $k+m+1$ for the iterate in the subdomains Ω_j with $j \in I_m$ the elliptic upper bound

$$\begin{aligned} \nu\Delta\tilde{u}_j^{k+m+1} + \mathbf{a}\cdot\nabla\tilde{u}_j^{k+m+1} + c\tilde{u}_j^{k+m+1} &= 0 \qquad \mathbf{x} \in \Omega_j, \\ \tilde{u}_j^{k+m+1}(\mathbf{x}) &= U^k \qquad \mathbf{x} \in \Gamma_{jl},\ l \notin I_{m-1}, \\ \tilde{u}_j^{k+m+1}(\mathbf{x}) &= U^{k+m} \qquad \mathbf{x} \in \Gamma_{jl},\ l \in I_{m-1} \end{aligned} \tag{6}$$

and $\tilde{u}_j^{k+m+1} < U^k$ in $\widetilde{\Omega}_j$. Defining U^{k+m+1} as before we find $U^{k+m+1} \leq \gamma_{m+1}(\delta)U^k$ for some constant $\gamma_1(\delta) \leq \gamma_2(\delta) \leq \ldots \leq \gamma_{m+1}(\delta) < 1$ independent of k. Now for the next iteration step $k+m+2$ all the u_j^{k+m+2} have boundary values less than or equal to $U^{k+m+1} \leq \gamma_{m+1}(\delta)U^k$, since they come from iteration step $k+m+1$ in the interior of neighboring subdomains. Defining $\gamma(m,\delta) := \gamma_{m+1}(\delta)$ the result follows. ■

Theorem 2 (Linear Convergence) *For $c \leq 0$ the overlapping Schwarz waveform relaxation algorithm (2) converges on unbounded time intervals $t \in [0, T = \infty)$ at least at the linear rate*

$$\max_j ||u_j^{k(m+2)}(\cdot,\cdot)||_\infty \leq (\gamma(m,\delta))^k \max_j ||u_j^0(\cdot,\cdot)||_\infty \tag{7}$$

where $\gamma(m,\delta) < 1$ as in Lemma 1.

Proof The proof follows by induction from Lemma 1. ■

The convergence result we derived on unbounded time domains depends on the number of subdomains, as one can see explicitly from the dependence of γ on m. The more subdomains one uses, the longer it takes for information to propagate from the outer boundary of Ω to the inner subdomains. This is because the steady state solution is limiting the convergence rate, and the steady state solution does not see the zero initial condition. This is different if the algorithm is analyzed over a bounded time interval. This analysis, which is beyond the scope of this short paper, leads to a superlinear convergence result for the algorithm. Defining for bounded functions $g(\mathbf{x},t) : \Omega \times [0,T) \to \mathbb{R}$ the norm

$$||g(\cdot,\cdot)||_T := \sup_{\mathbf{x}\in\Omega, 0<t<T} |g(\mathbf{x},t)|$$

we have the following

Theorem 3 (Superlinear Convergence) *For $c \leq 0$ the overlapping Schwarz waveform relaxation algorithm converges superlinearly on bounded time intervals $t \in [0, T < \infty)$ in the infinity norm,*

$$\max_j ||u_j^k(\cdot,\cdot)||_T \leq \left(2n\cosh(\delta\bar{a}/(2\nu\sqrt{n}))\right)^k \mathit{erfc}\left(\frac{k\delta}{2\sqrt{nT}}\right) \max_j ||u_j^0(\cdot,\cdot)||_T. \tag{8}$$

There are two interesting facts to note about this theorem: first the convergence rate is independent of the number of subdomains, there is no dependence on a parameter m related to the number of subdomains as in Theorem 2. second the superlinear convergence rate is faster than the superlinear convergence rate found for classical waveform relaxation algorithms. The classical result gives a contraction governed by a factorial [MN87] with asymptotic expansion

$$\frac{(CT)^k}{k!} = \left(\frac{1}{\sqrt{2\pi}} + O(k^{-1})\right) e^{-k\ln k + (1+\ln(CT))k - \frac{1}{2}\ln k} \sim e^{-k\ln k}$$

whereas the new result (8) gives a contraction with asymptotic expansion

$$C_1^k \mathrm{erfc}(\frac{C_2 k}{\sqrt{T}}) = \left(\frac{\sqrt{T}}{C_2\sqrt{\pi}} + O(k^{-2})\right) e^{-\frac{C_2^2}{T}k^2 + \ln(C_1)k - \ln k} \sim e^{-k^2}.$$

Numerical Experiments

We perform all our experiments on the two dimensional model problem

$$\frac{\partial u}{\partial t} = \nu\Delta u + \mathbf{a}\cdot\nabla u + cu, \quad (x_1, x_2) \in [0,1]\times[0,1],\ t \in [0,T]. \tag{9}$$

The convection is chosen to be diagonal, $\mathbf{a} := (1,1)$ and the other parameters are $c = 0$ and $\nu = 1/10$. We decompose the domain into smaller squares with equal size and overlap both in the x_1 and x_2 direction and simulate directly the error equations. In space we discretize using central finite differences and in time using backward Euler. To see linear convergence the problem is integrated over a relatively long time interval $t \in [0,10]$ and to see superlinear convergence the problem is integrated over a shorter time interval $t \in [0,0.5]$. Our analysis showed that for both the linear and superlinear convergence the convergence rate depends on the size of the overlap as usual. Increasing the overlap, the error decays faster, as shown in Figure 1 on the left for a long time interval and on the right for a short time interval. We used $\delta = 0.1$ and $\delta = 0.06$ for the overlap parameter and 2×2 subdomains.

Theorem 2 shows for the linear convergence regime that the decay of the error depends on the number of subdomains; the parameter m appears in equation (7), which is similar to the results found for the heat equation in [GS98]. Thus for a long time interval, the overlapping Schwarz waveform relaxation algorithm does not scale with respect to the number of subdomains. This is illustrated in Figure 2 on the left for $\mathbf{a} = (1,1)$, $c = 0$, $\nu = 1/20$, overlap parameter $\delta = 0.04$ and $t \in [0,5]$. Note how initially the algorithm does not exhibit convergence, the information needs to be propagated first from the domains connected to the boundary towards the interior, as we saw in the analysis. In the superlinear convergence regime however for the same problem parameters and $t \in [0,0.1]$ the convergence rate is independent of the number of subdomains, as stated in Theorem 3. This is confirmed in the numerical experiments shown in Figure 2 on the right and corresponds to the result found earlier for the heat equation in [GZ97]. Note how the error reduction in the superlinear convergence regime is considerably faster than the one in the linear convergence regime. Note also that the error reduction in the superlinear convergence regime is considerably faster than the one in the linear convergence regime.

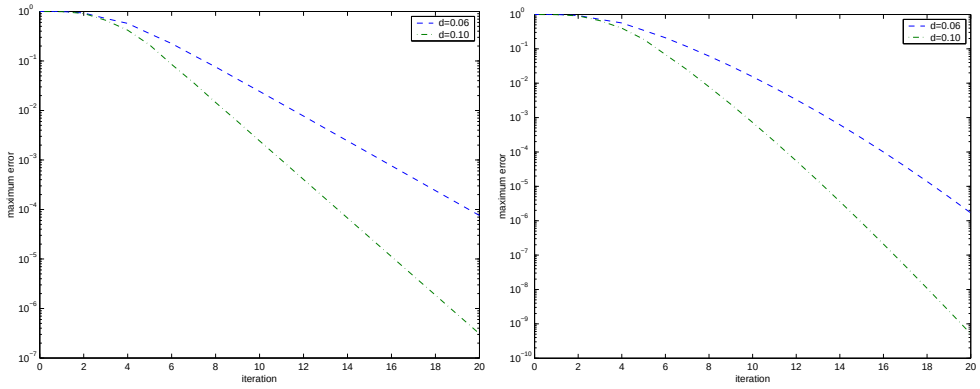


Figure 1: Two dimensional problem with four subdomains and different size of overlap for a long time interval on the left, where the algorithm is in the linear convergence regime and for a short time interval on the right, where the algorithm is in the superlinear convergence regime.

Conclusions

We have shown that the overlapping Schwarz waveform relaxation algorithm for general linear convection reaction diffusion equations with very general domain decomposition exhibits two different types of convergence regimes: on unbounded time intervals the algorithm converges at least at a linear rate depending on the size of the overlap, the problem parameters and the number of subdomains. On bounded time intervals however the convergence is superlinear. The convergence rate depends on the overlap and the diffusion coefficient, but is independent of the number of subdomains and the other problem parameters.

The main interest of the algorithm are the following three points:

1. The original problem is solved on subdomains in space-time and thus one can refine both in space and time independently on each subdomain.

2. Communication is not necessary at each time step, each processor continues to solve over a whole time window before it needs to communicate.

3. Theorem 3 shows that algorithm converges superlinearly and independently of the number of subdomains, so there is no coarse grid needed for scalability.

For a given hardware configuration, it remains to find the best length of time windows so that the convergence speed of the algorithm is balanced with the communication cost. Longer time windows lead to slower convergence, but they require less often communication which makes them faster.

References

[Bjø95]Morten Bjørhus. *On Domain Decomposition, Subdomain Iteration and Waveform Relaxation*. PhD thesis, University of Trondheim, Norway, 1995.

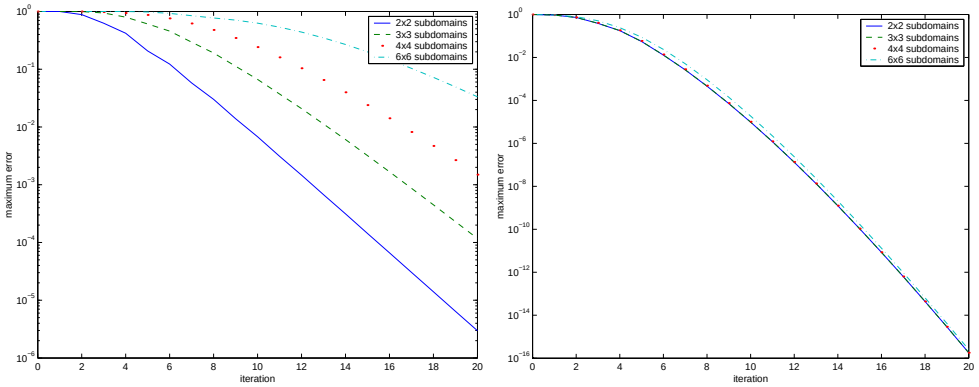


Figure 2: The effect of the number of subdomains on the algorithm. On the left for long time intervals where the linear convergence rate depends on the number of subdomains and on the right for short time intervals where the superlinear convergence rate is independent of the number of subdomains.

[Cai91]Xiao-Chuan Cai. Additive Schwarz algorithms for parabolic convection-diffusion equations. *Numer. Math.*, 60(1):41–61, 1991.

[Cai94]Xiao-Chuan Cai. Multiplicative Schwarz methods for parabolic problems. *SIAM J. Sci Comput.*, 15(3):587–603, 1994.

[GK97]Eldar Giladi and Herbert Keller. Space time domain decomposition for parabolic problems. Technical Report 97-4, Center for research on parallel computation CRPC, Caltech, 1997.

[GS98]Martin J. Gander and Andrew M. Stuart. Space-time continuous analysis of waveform relaxation for the heat equation. *SIAM Journal for Scientific Computing*, 19(6):2014–2031, 1998.

[GZ97]Martin J. Gander and Hongkai Zhao. Overlapping Schwarz waveform relaxation for parabolic problems in higher dimension. In A. Handlovičová, Magda Komorníkova, and Karol Mikula, editors, *Proceedings of Algoritmy 14*, pages 42–51. Slovak Technical University, September 1997.

[Lie96]Gary M. Lieberman. *Second Order Parabolic Differential Equations*. World Scientific, December 1996.

[Lio88]Pierre-Louis Lions. On the Schwarz alternating method. I. In Roland Glowinski, Gene H. Golub, Gérard A. Meurant, and Jacques Périaux, editors, *First International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 1–42, Philadelphia, PA, 1988. SIAM.

[Meu91]Gérard A. Meurant. Numerical experiments with a domain decomposition method for parabolic problems on parallel computers. In Roland Glowinski, Yuri A. Kuznetsov, Gérard A. Meurant, Jacques Périaux, and Olof Widlund, editors, *Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations*, Philadelphia, PA, 1991. SIAM.

[MN87]Ulla Miekkala and Olavi Nevanlinna. Convergence of dynamic iteration methods for initial value problems. *SIAM J. Sci. Stat. Comput.*, 8:459–482, 1987.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

21 Analysis of Two-Level Overlapping Additive Schwarz Preconditioners for a Discontinuous Galerkin Method

Xiaobing Feng[1], Ohannes A. Karakashian[2]

Introduction

The Schwarz method refers to a general methodology, based on the idea of divide-and-conquer, for solving the systems of linear algebraic equations resulting from numerical discretizations of partial differential equations. In the past fifteen years extensive research has been done on the method to solve different types of algebraic systems which arise from various discretizations of partial differential equations such as finite difference/element/volume methods, spectral methods and mortar finite element methods (cf. [SBG96, Xu92] and references therein). On the other hand, very few results on the Schwarz method have been known in the literature for discontinuous Galerkin methods (cf. [FK01a, LT00, RVW96]). Discontinuous Galerkin methods use piecewise, totally discontinuous polynomial trial and test function spaces, that is, no continuity constraints are explicitly imposed on the functions across the element interfaces. As a consequence, weak formulations must include jump terms across interfaces and typically penalty terms are (artificially) added to control the jump terms (cf. [Arn82, DD76, Whe78]).

Discontinuous Galerkin methods have several advantages over other types of finite element methods. For example, the trial and test spaces are very easy to construct; they can naturally handle inhomogeneous boundary conditions and curved boundaries; they also allow the use of highly nonuniform and unstructured meshes. In addition, the fact that the mass matrices are block diagonal is an attractive feature in the context of time-dependent problems, especially if explicit time discretizations are used. On the other hand, discontinuous Galerkin methods would seem to be at a disadvantage in view of a relatively larger number of degrees of freedom per element. Therefore, to offset this disadvantage, effective remedies must be found at the level of solution of the systems of algebraic equations.

The objective of this paper is to develop some two-level *overlapping* additive Schwarz preconditioners for a discontinuous Galerkin method for solving second order elliptic problems. In Section 2, the discontinuous Galerkin and some known facts about the method, as well as a trace inequality and a generalized Poincaré inequality for discontinuous, piecewise H^1 functions are recalled. In Section 3, some two-level overlapping additive Schwarz preconditioners are proposed and analyzed for the discontinuous Galerkin method. The main result is to show that the condition numbers of the preconditioned systems are of the order $O(\frac{H}{\delta})$, where H and δ stand for the coarse mesh size and the size of overlaps between subdomains.

This paper is the second in a sequel devoted to developing Schwarz methods for discontinuous Galerkin methods. [FK01a] contains non-overlapping Schwarz methods for discontinuous Galerkin methods. The condition number estimates of the order $O(\frac{H}{h})$ are established and numerical experiments are presented. In [FK01b], Schwarz methods are developed for the discontinuous Galerkin method of Baker [Bak77] for the biharmonic problems.

[1]Department of Mathematics, The University of Tennessee, Knoxville, TN 37996, U.S.A. xfeng@math.utk.edu
[2]Department of Mathematics, The University of Tennessee, Knoxville, TN 37996, U.S.A. ohannes@math.utk.edu

Preliminaries

Let $\Omega \subset \mathbf{R}^d,\ d = 2, 3$ be a bounded domain. For the sake of simplicity, we restrict ourselves to the following model problem:

$$-\Delta u = f \quad \text{in } \Omega, \tag{1}$$
$$u = g \quad \text{on } \partial\Omega. \tag{2}$$

We remark that although we only consider the above model problem, extension of our construction and analysis of this paper to more general second order elliptic problems can be easily carried out.

The discontinuous Galerkin method to be considered in this paper for discretizing problem (1)–(2) is the one proposed in [Bak77, BJK90]. In this paper, we shall adopt the same notation as that of [BJK90].

Let $\mathcal{T}_h = \{K_i :\ i = 1, 2, \cdots, m_h\}$ be a family of star-like partitions (triangulations) of the domain Ω parametrized by $0 < h \leq 1$. Note that $\mathcal{T}_h$ does not have to be geometrically conforming. We define

$\partial K_i =$ the boundary of K_i, $\quad \partial K_{ij} = \partial K_i \cap \partial K_j$, $\quad \partial K_i^e = \partial K_i \cap \partial\Omega$,
$\mathcal{N}_i = \{\ell;\ \text{meas}(\partial K_{i\ell}) > 0\}$, $\quad h_i = \text{diam}(K_i)$, $\quad h_{ij} = \text{diam}(\partial K_{ij})$,
$\tau_{ij} = 1,\ \text{if } i > j; \tau_{ij} = 0,\ \text{if } i \leq j.$

We shall refer to $\mathcal{T}_h$ as the "fine" mesh and assume that it satisfies the following assumptions:

(i) The elements of $\mathcal{T}_h$ satisfy the minimal angle condition

(ii) $\mathcal{T}_h$ is locally quasi-uniform, that is if K_j and K_ℓ are adjacent and meas$(\partial K_{j\ell}) > 0)$, then $h_j \approx h_\ell$.

Now define the "energy space" E by $E = H^2(K_1) \times H^2(K_2) \times \cdots \times H^2(K_{m_h})$ and the bilinear form $a_h(\cdot,\cdot)$ on $E \times E$ as follows: For $u, v \in E$,

$$\begin{aligned} a_h(u,v) = \sum_{K_j \subset \Omega} \Bigg\{ & (\nabla u^{(j)}, \nabla v^{(j)})_{K_j} - \sum_{\ell \in \mathcal{N}_j} \tau_{j\ell} \Bigg[\left\langle \frac{\partial u^{(j)}}{\partial n}, v^{(j)} - v^{(\ell)} \right\rangle_{\partial K_{j\ell}} \\ & + \left\langle \frac{\partial v^{(j)}}{\partial n}, u^{(j)} - u^{(\ell)} \right\rangle_{\partial K_{j\ell}} - \gamma h_{j\ell}^{-1} \langle u^{(j)} - u^{(\ell)}, v^{(j)} - v^{(\ell)} \rangle_{\partial K_{j\ell}} \Bigg] \\ & - \left\langle \frac{\partial u^{(j)}}{\partial n}, v^{(j)} \right\rangle_{\partial K_j^e} - \left\langle \frac{\partial v^{(j)}}{\partial n}, u^{(j)} \right\rangle_{\partial K_j^e} + \gamma h_j^{-1} \langle u^{(j)}, v^{(j)} \rangle_{\partial K_j^e} \Bigg\} \end{aligned} \tag{3}$$

Here $u^{(j)}$ denotes the restriction of u to the element K_j and $(\cdot,\cdot)_{K_j}$ the L^2 integral over K_j; $\langle u^{(j)}, u^{(\ell)} \rangle_{\partial K_{j\ell}}$ is the L^2 integral over the interface $\partial K_{j\ell}$ of the traces of $u^{(j)}$ and $u^{(\ell)}$. The terms including γ are the so-called penalty terms.

The bilinear form $a_h(\cdot,\cdot)$ induces the following norm on the space E

$$\|v\|_{1,h,\Omega} = \left(\sum_{K_j \subset \Omega} \left\{ ||\nabla v^{(j)}||^2_{0,K_j} + \sum_{\ell \in \mathcal{N}_j} \tau_{j\ell} \left[h_{j\ell} \left| \frac{\partial v^{(j)}}{\partial n} \right|^2_{0,\partial K_{j\ell}} \right.\right.\right.$$
$$\left.\left.\left. + h_{j\ell}^{-1} |v^{(j)} - v^{(\ell)}|^2_{0,\partial K_{j\ell}} \right] + h_j \left| \frac{\partial v^{(j)}}{\partial n} \right|^2_{0,\partial K_j^e} + h_j^{-1} |v^{(j)}|^2_{0,\partial K_j^e} \right\} \right)^{\frac{1}{2}}. \quad (4)$$

The weak formulation of (1)–(2) is defined as seeking $u \in E \cap H^1(\Omega) \cap H^2_{\text{loc}}(\Omega)$ such that

$$a_h(u,v) = F(v), \qquad \forall v \in E \cap H^1(\Omega) \cap H^2_{\text{loc}}(\Omega), \quad (5)$$

where

$$F(v) = (f,v) - \sum_{\partial K_j^e \subset \partial\Omega} \left\langle g, \frac{\partial v^{(j)}}{\partial n} - \gamma h_j^{-1} v^{(j)} \right\rangle_{\partial K_j^e}.$$

For any integer $r \geq 2$, let $P_{r-1}(D)$ denote the set of all polynomials of degree less than or equal to $r-1$ on D. Define the finite element space V^h as

$$V^h = P_{r-1}(K_1) \times P_{r-1}(K_2) \times \cdots \times P_{r-1}(K_{m_h}).$$

Clearly, $V^h \subset E \subset L^2(\Omega)$. But $V^h \not\subset H^1(\Omega)$. The coercivity and continuity of $a_h(\cdot,\cdot)$ with respect to $\|\cdot\|_{1,h,\Omega}$ norm is summarized in the following lemma.

Lemma 1 (cf. [BJK90]) *There exists $\gamma_0 > 0$, which only depends on r, such that for $\gamma \geq \gamma_0$*

$$|a_h(u,v)| \leq (1+\gamma)\|u\|_{1,h,\Omega}\|v\|_{1,h,\Omega}, \qquad \forall u,v \in E. \quad (6)$$
$$a_h(v,v) \geq C\|v\|^2_{1,h,\Omega}, \qquad \forall v \in V^h. \quad (7)$$

The discontinuous Galerkin method based on the weak formulation (6) is defined as follows: find $u_h \in V^h$ such that

$$a_h(u_h, v_h) = F(v_h), \qquad \forall v_h \in V^h. \quad (8)$$

We refer to [BJK90] for a detailed exposition on this particular method.

We conclude this section by introducing two inequalities, a trace inequality and a generalized Poincaré inequality, for totally discontinuous piecewise H^1 functions generalizing two well-known inequalities for H^1 functions. These inequalities play a key role for the convergence analysis in the next section.

Let D be a bounded, simply connected star-like domain with diameter H in $\mathbf{R}^d$, $d = 2, 3$ (cf. [BJK90, FK01a]), and $\mathcal{T}_D$ be a family of partitions (triangulations) of D parameterized by $0 < h \leq H$. Let V_D be the space of all piecewise, totally discontinuous H^1 functions over $\mathcal{T}_D$. For a given number $h \leq \delta << H$, let D_δ denote the boundary layer of D with the width δ. That is, $D_\delta = \{x \in D;\ \text{dist}(x, \partial D) \leq \delta\}$. For simplicity, we assume that the boundary ∂D_δ of D_δ is aligned with $\mathcal{T}_D$.

Lemma 2 (cf. [FK01a]) *For any $u \in V_D$, there holds the following trace inequality*

$$|u|^2_{0,\partial D} \leq C[H^{-1}\|u\|^2_{0,D} + H|u|^2_{1,h,D}], \tag{9}$$

where

$$|u|^2_{1,h,D} = \sum_{K\in\mathcal{T}_D} \|\nabla u\|^2_{0,D} + \sum_{\substack{\partial K_{ij}\subset\Omega \\ i>j}} h^{-1}|u^{(i)} - u^{(j)}|^2_{\partial K_{ij}}. \tag{10}$$

Lemma 3 (cf. [FK01a]) *For any $u \in V_D$, the following generalized Poincaré inequality holds.*

$$\|u\|^2_{0,D_\delta} \leq C[\delta H^{-1}\|u\|^2_{0,D} + \delta(\delta + H)|u|^2_{1,h,D}]. \tag{11}$$

The overlapping Schwarz method

Formulation of the additive Schwarz preconditioners

Let $\mathcal{T}_H$ denote a coarse partition (triangulation) of Ω with the mesh size $H > 0$ and V^H denote the discontinuous Galerkin finite element space of order $r-1$ associated with the mesh $\mathcal{T}_H$. Suppose that $\mathcal{T}_h$ is obtained as a refinement of $\mathcal{T}_H$ and its members are star-like. Let $\Omega = \cup_{j=1}^J \Omega_j$ be an overlapping decomposition of Ω, where each Ω_j satisfying $\mathrm{diam}(\Omega_j) \approx H$ is a star-like open subdomain of Ω and is aligned with $\mathcal{T}_h$. Moreover, we assume there exist nonnegative C^∞-functions $\{\theta_j\}_{j=1}^J$ such that

$$\sum_{j=1}^J \theta_j = 1 \quad \text{in } \overline{\Omega}, \quad \theta_j = 0 \quad \text{in } \Omega\setminus\Omega_j, \quad \|\nabla\theta_j\|_{L^\infty} \leq \frac{1}{\delta}. \tag{12}$$

We also assume that there exist two positive constants C_0 and C_1 such that $C_0 h \leq \delta \leq C_1 H$. Let $N(x)$ denote the number of subdomains which contain x. We assume that $N_c \equiv \max_{x\in\Omega} N(x)$ is a constant which is independent of h, H, J and δ. Recall that the parameter δ measures the amount of overlaps among the subdomains $\{\Omega_j\}$. For the construction of Ω_j, we refer to [SBG96] and the references therein.

Introduce the notation

$$\begin{array}{ll}
\Gamma_j = \partial\Omega_j \cap \partial\Omega, & \Gamma_{jk} = \partial\Omega_j \cap \Omega_k, \\
\mathcal{N}_j^i = \{\ell \in \mathcal{N}_j;\ \partial K_{j\ell} \subset \Omega_i\}, & \overline{\mathcal{N}}_j^i = \{\ell \in \mathcal{N}_j;\ \partial K_{j\ell} \subset \Gamma_{ik} \text{for some } \Gamma_{ik}\}, \\
\Omega_j^I = \{x \in \Omega_j;\ x \notin \Omega_k \text{ for all } k \neq j\}, & \Omega_j^\delta = \Omega_j \setminus \Omega_j^I.
\end{array}$$

It is well-known (cf. [SBG96, Xu92]) that the first step towards constructing the additive Schwarz preconditioners is to have a valid subspace decomposition of the finite element space V^h. For the discontinuous Galerkin method considered in this paper, since $V^h \subset L^2(\Omega)$ and no continuity constrain is imposed for the functions in V^h, it is easy to construct such a space decomposition.

We define the subspace $\{V_j^h\}_{j=1}^J$ associated with the subdomain $\{\Omega_j\}_{j=1}^J$ by

$$V_j^h = \{v_h \in V^h;\ v_h = 0 \text{ in } \Omega\setminus\overline{\Omega}_j\}, \quad j = 1, 2, \cdots, J.$$

In addition to $V_1^h, \cdots, V_J^h$, we now introduce a coarse subspace V_0^h corresponding to $\mathcal{T}_H$. Let the integer r_H be chosen satisfying $2 \leq r_H \leq r$. Let

$$V_0^h = \prod_{D \in \mathcal{T}_H} P_{r_H - 1}(D). \tag{13}$$

It is easy to see that V_0^h is a subspace of V^h. Also, our (theoretical) estimates are valid independent of the choice of r_H. Clearly, $V_0^h = V^H$ when $r_H = r$.

It is easy to check that the following space decomposition holds.

$$V^h = V_0^h + V_1^h + V_2^h + \cdots + V_J^h. \tag{14}$$

Having obtained the above space decomposition, the second step requires the construction of a subdomain bilinear form (or a subdomain solver) on each subdomain. To this end, we define $a_i(\cdot,\cdot)$ on $V_i^h \times V_i^h$ to be the restriction of $a_h(\cdot,\cdot)$ on Ω_j for $i = 1, 2, \cdots, J$, and $a_0(\cdot,\cdot) = a_h(\cdot,\cdot)$. Notice that, $a_0(\cdot,\cdot)$ differs from $a_H(\cdot,\cdot)$ only in the choice of the penalty parameter γ on $V_0^h \times V_0^h$.

Now we are ready to define the additive operator

$$T = T_0 + T_1 + \cdots + T_J, \tag{15}$$

where T_j is a projection operator from V^h to V_j^h which is defined by

$$a_j(T_j u, v) = a_h(u, v) \qquad \forall v \in V_j^h,\ j = 0, 1, 2, \cdots, J. \tag{16}$$

The additive Schwarz method is defined by replacing the discrete problem (9) by the equation (cf. [SBG96])

$$Tu = g, \qquad\qquad g = \sum_{j=0}^{J} g_j, \tag{17}$$

where $g_j = T_j u$ is defined as the solution of

$$a_j(g_j, v) = F(v) \qquad \forall v \in V_j^h,\ j = 0, 1, 2, \cdots, J. \tag{18}$$

Condition number estimate for the additive Schwarz method

To estimate the condition number of T, we will use the abstract convergence framework of Schwarz methods given in [SBG96]. To this end, we need some preliminary lemmas, including the decomposition lemma (see Lemma 7).

Let $W_0^H \subset H_0^1(\Omega)$ be the standard P_1 conforming finite element space associated with the coarse mesh $\mathcal{T}_H$. Trivially, $W_0^H \subset V_0^h$. We recall the following approximation property of the finite element space W_0^H.

Lemma 4 (cf. [Cia78]) *For any $\psi \in H^2(\Omega) \cap H_0^1(\Omega)$, the following estimate holds*

$$\inf_{v \in W_0^H} a_h(\psi - v, \psi - v)^{\frac{1}{2}} = \inf_{v \in W_0^H} \|\nabla(\psi - v)\|_{0,\Omega} \leq CH\|\psi\|_{2,\Omega}. \tag{19}$$

Next, for any function $u \in V^h$, we define $P_H u$ to be the projection of u into W_0^H with respect to $a_h(\cdot,\cdot)$, that is,

$$a_h(P_H u, v) = a_h(u, v) \qquad \forall v \in W_0^H. \tag{20}$$

The operator P_H satisfies the following stability and approximation properties.

Lemma 5 *There exists a positive constant C, which is independent of h, J, δ and H, such that*

$$a_h(P_H u, P_H u) \leq a_h(u, u). \tag{21}$$

$$\|u - P_H u\|_{0,\Omega} \leq CH a_h(u, u)^{\frac{1}{2}}. \tag{22}$$

To save space, we omit the proof and refer to [FK01a] for a proof of similar type.

For each $K \in \mathcal{T}_h$, let Π_K denote the usual interpolation operator to the polynomial space $P_{r-1}(K)$ as defined in the conforming finite element methods. Define the interpolation operator $\Pi_h : \prod_{j=1}^{m_h} C^0(\overline{K_j}) \longrightarrow V^h$ by

$$\Pi_h \phi = \Pi_K \phi \quad \text{in } K, \quad \forall K \in \mathcal{T}_h, \ \forall \phi \in \prod_{j=1}^{m_h} C^0(\overline{K_j}) \tag{23}$$

For any $u \in V^h$, we introduce the following decomposition of u

$$u = u_0 + u_1 + \cdots + u_J, \quad u_j \in V_j^h, \tag{24}$$

where

$$u_0 = P_H u, \quad u_j = \Pi_h[\theta_j(u - P_H u)], \ j = 1, 2, \cdots, J. \tag{25}$$

We emphasize that the operator P_H is only needed in the analysis and does not contribute to the construction of the computational method.

Lemma 6 *For any $u \in V^h$, let $u_j \in V_j^h$ be defined as above. Then there is a positive constant C which is independent of h, J, δ and H such that*

$$a_i(u_i, u_i) \leq C \left(\|u - P_H u\|_{1,h,\Omega_i}^2 + \frac{1}{\delta^2} \|u - P_H u\|_{0,\Omega_i^\delta}^2 \right), \ i = 1, 2, \cdots, J, \tag{26}$$

Proof: Let $w = u - u_0$. Since $\theta_i = 0$ in $\Omega \setminus \Omega_i$, $u_i = 0$ on every $\partial K_{j\ell} \subset \partial\Omega_i \setminus \partial\Omega$. By the definition of $a_i(\cdot,\cdot)$ and the Schwarz inequality we have

$$a_i(u_i, u_i) \leq C \sum_{K_j \subset \Omega_i} \left\{ \|\nabla u_i^{(j)}\|_{0,K_j}^2 + \sum_{\ell \in \mathcal{N}_j^i} \tau_{j\ell} \left[h_{j\ell} \left| \frac{\partial u_i^{(j)}}{\partial n} \right|_{0,\partial K_{j\ell}}^2 \right.\right.$$
$$\left.\left. + h_{j\ell}^{-1} |u_i^{(j)} - u_i^{(\ell)}|_{0,\partial K_{j\ell}}^2 \right] + h_j \left| \frac{\partial u_i^{(j)}}{\partial n} \right|_{0,\partial K_j^e}^2 + h_j^{-1} |u_i^{(j)}|_{0,\partial K_j^e}^2 \right\}. \tag{27}$$

Let $\overline{\theta}_{ij}$ be the average of θ_i over an element $K_j \subset \Omega_i$. It is known that (cf. [SBG96])

$$\|\theta_i - \overline{\theta}_{ij}\|_{L^\infty(K_j)} \leq \begin{cases} Ch_j\delta^{-1} & \forall K_j \in \Omega_i^\delta, \\ 0 & \forall K_j \in \Omega_i^I. \end{cases} \tag{28}$$

For each term on the right hand side of (27) we have the following estimate.

$$\begin{aligned} \|\nabla u_i^{(j)}\|_{0,K_j}^2 &\leq \|\nabla w^{(j)}]\|_{0,K_j}^2 + Ch_j^{-2}\|\Pi_{K_j}[(\theta_i - \overline{\theta}_{ij})w^{(j)}]\|_{0,K_j}^2 \\ &\leq \|\nabla w^{(j)}]\|_{0,K_j}^2 + \begin{cases} C\delta^{-2}\|w^{(j)}\|_{0,K_j}^2 & \text{if } K_j \in \Omega_i^\delta, \\ 0 & \text{if } K_j \in \Omega_i^I. \end{cases} \end{aligned} \tag{29}$$

$$|u_i^{(j)} - u_i^{(\ell)}|_{0,\partial K_{j\ell}}^2 = |\Pi_{K_{j\ell}}[\theta_i(w^{(j)} - w^{(\ell)})]|_{0,\partial K_{j\ell}}^2 \leq C|w^{(j)} - w^{(\ell)}|_{0,\partial K_{j\ell}}^2. \tag{30}$$

$$\begin{aligned} \left|\frac{\partial u_i^{(j)}}{\partial n}\right|_{0,\partial K_{j\ell}}^2 &\leq \left|\frac{\partial w_i^{(j)}}{\partial n}\right|_{0,\partial K_{j\ell}}^2 + Ch_j^{-1}\|\nabla\Pi_{K_j}[(\theta_i - \overline{\theta}_{ij})w^{(j)}]\|_{0,K_j}^2 \\ &\leq \left|\frac{\partial w_i^{(j)}}{\partial n}\right|_{0,\partial K_{j\ell}}^2 + \begin{cases} Ch_j^{-1}\delta^{-2}\|w^{(j)}\|_{0,K_j}^2 & \text{if } K_j \in \Omega_i^\delta, \\ 0 & \text{if } K_j \in \Omega_i^I. \end{cases} \end{aligned} \tag{31}$$

$$\left|\frac{\partial u_i^{(j)}}{\partial n}\right|_{0,\partial K_j^e}^2 \leq C\left(\left|\frac{\partial w_i^{(j)}}{\partial n}\right|_{0,\partial K_j^e}^2 + h_j^{-1}\delta^{-2}\|w^{(j)}\|_{0,K_j}^2\right). \tag{32}$$

$$\left|u_i^{(j)}\right|_{0,\partial K_j^e}^2 = \left|\Pi_{K_j^e}[\theta_i w^{(j)}]\right|_{0,\partial K_j^e}^2 \leq C\left|w^{(j)}\right|_{0,\partial K_j^e}^2 \tag{33}$$

Finally, the estimate (26) follows from (27), (29)–(33) and the definition of $\|w\|_{1,h,\Omega_i}$.
The following lemma follows directly from applying Lemma 6 and Lemma 3 on each Ω_j.

Lemma 7 *For any $u \in V^h$, let $u_j \in V_j^h$ be as in (25). There is a positive constant C which is independent of h, J, δ and H such that*

$$\sum_{j=0}^{J} a_i(u_j, u_j) \leq C\frac{H}{\delta}a_h(u, u). \tag{34}$$

It is trivial to show the next lemma (cf. [FK01a]).

Lemma 8 *There holds the following identity.*

$$a_h(v_j, v_j) = a_j(v_j, v_j), \quad \forall v_j \in V_j^h, \quad j = 0, 1, \cdots, J. \tag{35}$$

Using a coloring argument (cf. [SBG96]), it is easy to show the following lemma.

Lemma 9 *Let u and u_i be as in (25). Let $0 \leq \mathcal{E}_{ij} \leq 1$ to be the minimal values such that*

$$|a_h(u_i, u_j)| \leq \mathcal{E}_{ij} a_h(u_i, u_i)^{\frac{1}{2}} a_h(u_j, u_j)^{\frac{1}{2}}, \quad i, j = 1, 2, \cdots, J. \tag{36}$$

Then there holds the following estimate

$$\rho(\mathcal{E}) \leq N_c + 1. \tag{37}$$

We are now ready to establish the main theorem of this paper.

Theorem 1 *There exists a positive constant C which is independent of h, J, δ and H such that there holds the estimate*

$$cond(T) \leq C(2 + N_c) H \delta^{-1}. \tag{38}$$

Proof: The estimate (38) follows immediately from Lemma 7–9 and Lemma 3 of Chapter 5 of [SBG96] with $C_0^2 = O(H\delta^{-1})$, $\omega = 1$ and $\rho(\mathcal{E}) = 1 + N_c$.

References

[Arn82]Douglas N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19:742–760, 1982.

[Bak77]Garth Baker. Finite element methods for elliptic equations using nonconforming elements. *Math. Comp.*, 31:44–59, 1977.

[BJK90]Garth A. Baker, Wadi N. Jureidini, and Ohannes A. Karakashian. Piecewise solenoidal vector fields and the stokes problems. *SIAM J. Numer. Anal.*, 27:1466–1485, 1990.

[Cia78]Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[DD76]Jim Douglas and Todd Dupont. *Interior penalty procedures for elliptic and parabolic Galerkin methods*, pages 207–216. Lecture Notes In Physics 58. Springer–Verlag, Berlin, 1976.

[FK01a]Xiaobing Feng and Ohannes A. Karakashian. Two-level additive schwarz methods for a discontinuous galerkin approximation of second order elliptic problems. *SIAM J. Numer. Anal.*, 2001. to appear.

[FK01b]Xiaobing Feng and Ohannes A. Karakashian. Two-level schwarz methods for a discontinuous galerkin approximation of the biharmonic equation. in preparation, 2001.

[LT00]Caroline Lasser and Andrea Toselli. An overlapping domain decomposition preconditioner for a class of discontinuous Galerkin approximations of advection-diffusion problems. Technical Report 810, Dept. of Computer Science, Courant Institute, 2000. submitted to Math. Comp.

[RVW96]Torgeir Rusten, Panayot S. Vassilevski, and Ragnar Winther. Interior penalty preconditioners for mixed finite element approximations of elliptic problems. *Math. Comp.*, 65:447–466, 1996.

[SBG96]Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[Whe78]Mary F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15:152–161, 1978.

[Xu92]Jinchao Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, December 1992.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

22 Optimized Schwarz Methods for Helmholtz Problems

M. J. Gander [1]

Introduction

The classical Schwarz algorithm has a long history. It was invented by Schwarz more than a century ago to prove existence and uniqueness of solutions to Laplace's equation on irregular domains. It gained popularity with the advent of parallel computers and was analyzed in depth both at the continuous level and as a preconditioner for discretized problems (see the books by Quarteroni and Valli [QV99] and Smith, Bjørstad and Gropp [SBG96] and references therein). The classical Schwarz algorithm is however not effective for Helmholtz problems, because the convergence mechanism of the Schwarz algorithm works only for the evanescent modes, not for the propagative ones. Nevertheless the Schwarz algorithm has been applied to Helmholtz problems by adding a relatively fine coarse mesh in [CW92] and changing the transmission conditions from Dirichlet in the classical Schwarz case to Robin, as done in [DJR92], [BD97], [Gha97], [dLBFM$^+$98], [MSRKA98] and [CCEW98]. The influence of the transmission conditions on the Schwarz algorithm for the Helmholtz equation has first been studied for a nonoverlapping version of the Schwarz algorithm in [CN98] and for the overlapping case in [GHN00]. We begin this paper by recalling the optimal transmission conditions which lead to the best possible convergence of the Schwarz algorithm and which even work without overlap. These optimal transmission conditions are however non local in nature and thus not ideal for implementations. One therefore approximates the optimal transmission conditions locally. A first result we present is that no matter how one approximates, the new optimized Schwarz method has a better convergence rate than the classical Schwarz method. Then we present a new second order optimized transmission condition for a nonoverlapping variant of the optimized Schwarz method with better asymptotic performance than the one presented in [GMN01]. If h denotes the mesh parameter, then the new method has a convergence rate of $1 - O(h^{1/4})$ whereas the best optimized Schwarz method so far for the Helmholtz equation had a convergence rate of $1 - O(h^{1/2})$, as given in [GMN01].

Classical Schwarz for the Helmholtz Equation

We consider the Helmholtz equation in two dimensions,

$$(\Delta + \omega^2)(u) = f, \quad \text{in } \Omega = \mathbb{R}^2, \tag{1}$$

with Sommerfeld radiation conditions at infinity. We apply the Schwarz algorithm with two overlapping subdomains $\Omega_1 = (-\infty, L] \times \mathbb{R}$, $L > 0$ and $\Omega_2 = [0, \infty) \times \mathbb{R}$ which leads to the Schwarz iteration

$$\begin{array}{rcll} \Delta v^{n+1} + \omega^2 v^{n+1} & = & f, & \text{in } \Omega_1, \\ v^{n+1}(L, y) & = & w^n(L, y), & y \in \mathbb{R} \end{array} \tag{2}$$

[1] Dept. of Mathematics and Statistics, McGill University, Montreal, QC H3A 2K6, CANADA.

and

$$\begin{aligned} \Delta w^{n+1} + \omega^2 w^{n+1} &= f, && \text{in } \Omega_2, \\ w^{n+1}(0,y) &= v^n(0,y), && y \in \mathbb{R}. \end{aligned} \tag{3}$$

To analyze the convergence rate of this algorithm, we use Fourier analysis. By linearity it suffices to analyze the homogeneous problem, $f(x,y) = 0$, and show convergence to the zero solution. Applying a Fourier transform in the y variable with Fourier parameter k leads to the ordinary differential equations

$$\begin{aligned} \frac{\partial^2 \hat{v}^{n+1}}{\partial x^2} + (\omega^2 - k^2)\hat{v}^{n+1} &= 0, && x < L,\ k \in \mathbb{R}, \\ \hat{v}^{n+1}(L,k) &= \hat{w}^n(L,k), && k \in \mathbb{R}, \\ \frac{\partial^2 \hat{w}^{n+1}}{\partial x^2} + (\omega^2 - k^2)\hat{w}^{n+1} &= 0, && x > 0,\ k \in \mathbb{R}, \\ \hat{w}^{n+1}(0,k) &= \hat{v}^n(0,k), && k \in \mathbb{R}. \end{aligned}$$

Solving the second equation at step n and inserting the result into the first one we find after evaluating at $x = 0$

$$\hat{v}^{n+1}(0,k) = e^{-2\sqrt{k^2-\omega^2}L}\hat{v}^{n-1}(0,k).$$

Hence the convergence rate of the classical Schwarz method is

$$\rho_{cla} := e^{-2\sqrt{k^2-\omega^2}L}. \tag{4}$$

This shows the main problem of the classical Schwarz method when applied to a Helmholtz problem: while evanescent or high frequency modes, $k^2 > \omega^2$, converge as in the case of Laplace's equation, the propagating or low frequency modes, $k^2 < \omega^2$, do not converge at all, $|\rho_{cla}| = 1$ for those modes. Figure 1 shows the error in a numerical experiment for an example on a domain $\Omega = [0,2] \times [0,1]$ split into two subdomains in the x-direction and $\omega = 10$. The error on the left subdomain is shown as the iteration progresses and one can see that the classical Schwarz algorithm has problems converging because of the low frequency modes, whereas the high frequency modes introduced at the interface by the initial guess are reduced effectively. Figure 2 shows on the left the corresponding convergence rate (4) for this example as a function of the frequency parameter k.

Optimized Schwarz for the Helmholtz Equation

We consider again the Helmholtz equation (1) in two dimensions and we apply a Schwarz algorithm with the same overlapping subdomains $\Omega_1 = (-\infty, L] \times \mathbb{R}$, $L > 0$ and $\Omega_2 = [0,\infty) \times \mathbb{R}$ as before. But this time we do not use Dirichlet transmission conditions, but more general ones,

$$\begin{aligned} \Delta v^{n+1} + \omega^2 v^{n+1} &= f, && \text{in } \Omega_1, \\ (\partial_x + \Lambda_v)(v^{n+1}(L,y)) &= (\partial_x + \Lambda_v)(w^n(L,y)), && y \in \mathbb{R} \end{aligned} \tag{5}$$

and

$$\begin{aligned} \Delta w^{n+1} + \omega^2 w^{n+1} &= f, && \text{in } \Omega_2, \\ (\partial_x + \Lambda_w)(w^{n+1}(0,y)) &= (\partial_x + \Lambda_w)(v^n(0,y)), && y \in \mathbb{R}. \end{aligned} \tag{6}$$

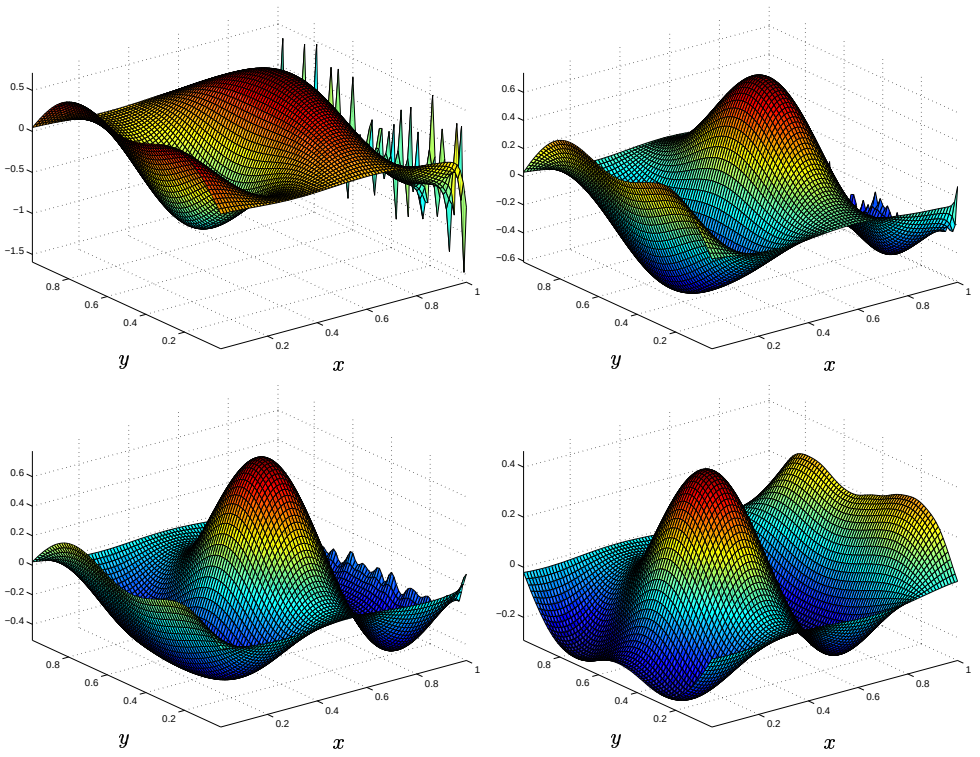


Figure 1: Error in iterations 1, 2, 3 and 8 on the left of the two subdomains of the classical Schwarz algorithm applied to a Helmholtz equation. Clearly the low frequency modes are not effectively reduced by the method.

The operators Λ_v in (5) and Λ_w in (6) are linear operators in the y direction along the interface which we will try to determine to obtain optimal performance of the Schwarz algorithm. Using Fourier analysis like in the case of the classical Schwarz algorithm and setting $f(x,y) = 0$, we obtain the iteration in the Fourier transformed domain

$$\begin{aligned}
\frac{\partial^2 \hat{v}^{n+1}}{\partial x^2} + (\omega^2 - k^2)\hat{v}^{n+1} &= 0, & x < L,\ k \in \mathbb{R},\\
(\partial_x + \lambda_v(k))(\hat{v}^{n+1}(L,k)) &= (\partial_x + \lambda_v(k))(\hat{w}^n(L,k)), & k \in \mathbb{R},\\
\frac{\partial^2 \hat{w}^{n+1}}{\partial x^2} + (\omega^2 - k^2)\hat{w}^{n+1} &= 0, & x > 0,\ k \in R,\\
(\partial_x + \lambda_w(k))(\hat{w}^{n+1}(0,k)) &= (\partial_x + \lambda_w(k))(\hat{v}^n(0,k)), & k \in \mathbb{R}.
\end{aligned}$$

Solving the second equation at step n and inserting the result into the first equation we find after evaluating at $x = 0$

$$\hat{v}^{n+1}(0,k) = \frac{\lambda_v(k) - \sqrt{k^2 - w^2}}{\lambda_v(k) + \sqrt{k^2 - w^2}} \cdot \frac{\lambda_w(k) + \sqrt{k^2 - w^2}}{\lambda_w(k) - \sqrt{k^2 - w^2}} e^{-2\sqrt{k^2-\omega^2}L} \hat{v}^{n-1}(0,k)$$

and hence the convergence rate of the new Schwarz method is

$$\rho_{opt} := \frac{\lambda_v(k)-\sqrt{k^2-w^2}}{\lambda_v(k)+\sqrt{k^2-w^2}} \cdot \frac{\lambda_w(k)+\sqrt{k^2-w^2}}{\lambda_w(k)-\sqrt{k^2-w^2}} e^{-2\sqrt{k^2-\omega^2}L}, \tag{7}$$

where we can choose the symbols $\lambda_v(k)$ and $\lambda_w(k)$ of the linear operators Λ_v and Λ_w along the interface to influence the performance of the new Schwarz method.

An Optimal Schwarz Method

There is a best choice for the free parameters in the convergence rate (7) of the new Schwarz method: choosing $\lambda_v(k) = \sqrt{k^2 - w^2}$ and $\lambda_w(k) = -\sqrt{k^2 - w^2}$, the convergence rate becomes zero for all values of the frequency parameter k and hence the method converges in 2 iterations. In addition for this choice the convergence rate is independent of the overlap, the exponential factor in (7) is irrelevant and hence the Schwarz method can be used without overlap as well. One can show that this result generalizes to convergence in N iterations if N subdomains in strips are employed [NR95]. But for real computations, we do not want to depend on Fourier transforms, we want to do the computations as usual on a given finite element or finite difference mesh. Hence we need the inverse transform of the optimal transmission conditions, $\lambda_{vw}(k) = \pm\sqrt{k^2 - w^2}$. Unfortunately, this inverse transform leads to nonlocal operators Λ_{vw} in the y variable, because of the square root in their symbol. Even though such non-local operators can be implemented by using a convolution on the boundary, it is much more cumbersome than to implement local transmission conditions. If the symbol of the optimal transmission conditions was a polynomial in k however, then the operator in real space would be local, because a polynomial in k transforms into derivatives in real space, and derivatives are local operators. Therefore, instead of using the best possible transmission conditions, we introduce local approximations to those conditions which are easy to implement. One can either choose a Taylor expansion about a low frequency to improve the low frequency behavior of the algorithm or, even better, optimize the approximation for the performance of the algorithm by making ρ_{opt}, the natural measure of performance, as small as possible. This leads to the new class of optimized Schwarz methods.

Optimized Schwarz Methods

We introduce local approximations of the best transmission operator,

$$\Lambda_v = +(\alpha_1 + \beta_1 k^2), \text{ and } \Lambda_w = -(\alpha_2 + \beta_2 k^2),$$

where $\alpha_j, \beta_j \in \mathbb{C}$, $j = 1, 2$. Note that we do not include a first order term because the Helmholtz operator is symmetric. For non-symmetric problems one would include the first order term as well. The case $\beta_j = 0$ leads to Robin transmission conditions and gives us four coefficients to optimize the performance (two complex numbers α_1 and α_2). If $\beta_j \neq 0$ we obtain transmission conditions including second order tangential derivatives which gives us eight coefficients to optimize the performance of the algorithm. In the sequel we restrict our analysis for simplicity to the special case where $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 = \beta_2 = \beta$, for which the convergence rate of the optimized Schwarz method can be simplified to

$$\rho_{opt} = \left(\frac{\alpha + \beta k^2 - \sqrt{k^2 - w^2}}{\alpha + \beta k^2 + \sqrt{k^2 - w^2}} \right)^2 e^{-2\sqrt{k^2-\omega^2}L}. \tag{8}$$

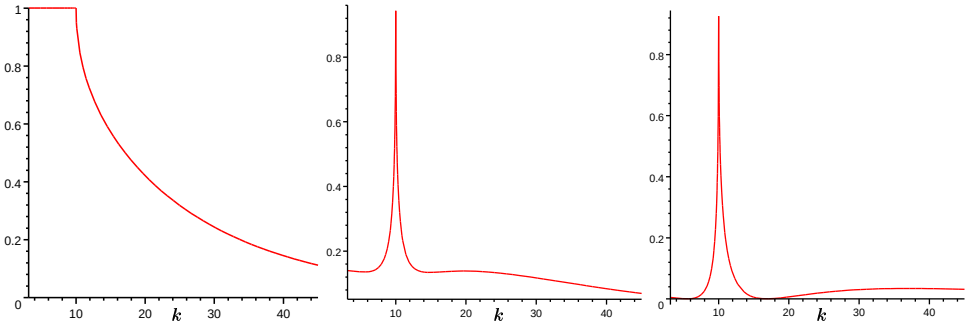


Figure 2: Comparison of the convergence rate of classical Schwarz on the left with optimized Schwarz using Robin transmission conditions in the middle and second order optimized transmission conditions on the right.

This cuts the number of optimization parameters in half and simplifies the optimization, at the cost of not finding the best possible second order transmission conditions. For symmetric positive definite problems the difference is investigated in [Gan00] and is found to be significant.

Theorem 1 *If* $\Re(\alpha), \Im(\alpha), \Re(\beta), \Im(\beta) \geq 0$ *then the optimized Schwarz method always converges faster than the classical Schwarz method.*

Proof We have to show under the conditions of the theorem that $\rho_{opt}(k)$ given in (8) is smaller or equal to $\rho_{cla}(k)$ given in (4) for all k. The only difference between the two convergence rates is the additional factor in front of the exponential in (8). But the modulus of this factor is

$$\frac{(\Re(\alpha)+\Re(\beta)k^2-\Re(\sqrt{k^2-w^2}))^2+(\Im(\alpha)+\Im(\beta)k^2-\Im(\sqrt{k^2-w^2}))^2}{(\Re(\alpha)+\Re(\beta)k^2+\Re(\sqrt{k^2-w^2}))^2+(\Im(\alpha)+\Im(\beta)k^2+\Im(\sqrt{k^2-w^2}))^2} \leq 1$$

if the real and imaginary parts of α and β are non-negative, which completes the proof. ■

This indicates that one should not use the classical Schwarz method any longer, whatever one does to the coefficients in the transmission conditions, the optimized Schwarz method will work better than the classical Schwarz method. Figure 2 shows a comparison of the convergence rates of the classical Schwarz method and optimized Schwarz methods with Robin and second order transmission conditions as a function of the frequency parameter k. Note that for optimized Schwarz methods the low frequency modes converge as well, not just the high frequency ones. Only at the resonance frequency $k^2 = \omega^2$ the convergence rate equals one for optimized Schwarz methods. This is however not a problem when optimized Schwarz is used as a preconditioner for a Krylov subspace method, since such a method easily corrects one bad mode in the spectrum.

An Optimized Schwarz Method without Overlap

We optimize now the coefficients α and β in (8) for the case of no overlap, $L = 0$. For the continuous problem we would need to optimize for all frequency parameters $k \in \mathbb{R}$ which

would lead to convergence problems as $k \longrightarrow \pm\infty$. But in a numerical computation, the frequency range is bounded, from below by the smallest frequency $k_{\min}$ relevant to the subdomain and from above by the largest frequency $k_{\max}$ supported by the numerical grid. The largest frequency $k_{\max}$ is of order π/h. We therefore have to solve the optimization problem

$$\min_{\alpha,\beta\in\mathbb{C}} \left(\max_{k\in(k_{\min},\omega_-)\cup(\omega_+,k_{\max})} \left| \frac{\alpha+\beta k^2-\sqrt{k^2-w^2}}{\alpha+\beta k^2+\sqrt{k^2-w^2}} \right|^2 \right) \tag{9}$$

where ω_- and ω_+ are parameters to be chosen to exclude the single mode with convergence rate one at the resonance frequency $k^2 = \omega^2$. We have the following asymptotic convergence result

Theorem 2 *There exist parameters $\alpha, \beta \in \mathbb{C}$ such that the asymptotic convergence rate of the optimized Schwarz method is*

$$\rho_{opt} = 1 - 4\left(\frac{\pi}{\sqrt{d\omega(2\omega - d\omega)}}\right)^{1/4} h^{1/4} + O(h^{1/2})$$

where $d\omega := \omega_+ - \omega = \omega - \omega_-$.

The proof of this result is beyond the scope of this short paper, since it involves the asymptotic solution of the min-max problem (9). But it is important to notice that the classical Schwarz method does not converge without overlap, not even in the symmetric positive definite case. If the overlap is of order h, then the convergence rate of classical Schwarz is $1 - O(h)$ in the symmetric positive definite case. The optimized Schwarz method without overlap converges even for the indefinite case at the much better rate of $1 - O(h^{1/4})$ except for the resonance mode. The numerical results in the next section show that the optimized Schwarz method used as a preconditioner for a Krylov methods exhibits a convergence rate of nearly $1 - O(h^{1/8})$, gaining almost the expected square-root from Krylov acceleration.

Numerical Results

We chose the model problem of a tube,

$$\begin{aligned}
\Delta u + \omega^2 u &= f & 0 < x, y < 1,\\
u &= 0 & 0 < x < 1,\ y = 0, 1,\\
\tfrac{\partial u}{\partial x} - i\omega u &= 0 & x = 0,\ 0 < y < 1,\\
-\tfrac{\partial u}{\partial x} - i\omega u &= 0 & x = 1,\ 0 < y < 1.
\end{aligned}$$

and two nonoverlapping subdomains $\Omega_1 = [0, 1/2] \times [0, 1]$, $\Omega_2 = [1/2, 1] \times [0, 1]$. For experiments with overlap, see [GHN00]. Table 1 shows the number of iterations required to converge to a desired tolerance $10e-6$ using optimized Schwarz as a preconditioner for GMRES and compares this to a non-optimized local approximation of the optimal transmission conditions using a Taylor expansion for low frequencies.

Figure 3 shows the asymptotic convergence rate in h achieved by the optimized Schwarz method. Note how Krylov acceleration gives almost the additional square-root, $\rho_{opt} = 1 - O(h^{1/8})$ as one can expect in ideal situations. It would have been interesting to do the experiment for $h = 1/1600$, but the case $h = 1/800$ shown constitutes already a complex linear system with 640'000 unknowns and is at the limit of current workstation capacities.

h	1/50	1/100	1/200	1/400	1/800
Taylor Order 2	25	32	38	46	57
Optimized Order 2	10	10	10	11	13

Table 1: Optimized Schwarz second order transmission conditions compared to a simpler second order Taylor approximation of the optimal transmission conditions for low frequencies.

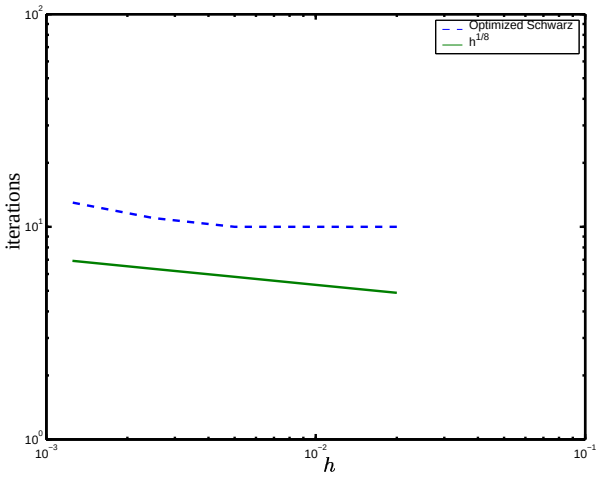


Figure 3: Asymptotic convergence rate of the second order optimized Schwarz method without overlap.

References

[BD97]Jean-David Benamou and Bruno Deprés. A domain decomposition method for the helmholtz equation and related optimal control problems. *J. of Comp. Physics*, 136:68–82, 1997.

[CCEW98]Xiao-Chuan Cai, Mario A. Casarin, Frank W. Elliott Jr., and Olof B. Widlund. Overlapping Schwarz algorithms for solving Helmholtz's equation. In *Domain decomposition methods, 10 (Boulder, CO, 1997)*, pages 391–399. Amer. Math. Soc., Providence, RI, 1998.

[CN98]Philippe Chevalier and Frédéric Nataf. Symmetrized method with optimized second-order conditions for the Helmholtz equation. In *Domain decomposition methods, 10 (Boulder, CO, 1997)*, pages 400–407. Amer. Math. Soc., Providence, RI, 1998.

[CW92]Xiao-Chuan Cai and Olof B. Widlund. Domain decomposition algorithms for indefinite elliptic problems. *SIAM J. Sci. Statist. Comput.*, 13(1):243–258, January 1992.

[DJR92]Bruno Després, Patrick Joly, and Jean E. Roberts. A domain decomposition method for the harmonic Maxwell equations. In *Iterative methods in linear algebra (Brussels, 1991)*, pages 475–484, Amsterdam, 1992. North-Holland.

[dLBFM$^+$98]Armel de La Bourdonnaye, Charbel Farhat, Antonini Macedo, Frédéric Magoulès, and François-Xavier Roux. A non-overlapping domain decomposition method for exterior Helmholtz problems. In *Domain decomposition methods, 10 (Boulder, CO,*

1997), pages 42–66, Providence, RI, 1998. Amer. Math. Soc.

[Gan00]Martin J. Gander. Optimized Schwarz methods for symmetric positive definite problems. in preparation, 2000.

[Gha97]Souad Ghanemi. A domain decomposition method for Helmholtz scattering problems. In P. E. Bjørstad, M. Espedal, and D. Keyes, editors, *Ninth International Conference on Domain Decomposition Methods*, pages 105–112. ddm.org, 1997.

[GHN00]Martin J. Gander, Laurence Halpern, and Frédéric Nataf. Optimized Schwarz methods. In *12th international conference on domain decomposition methods*, 2000.

[GMN01]Martin J. Gander, Frédéric Magoulès, and Frédéric Nataf. Optimized schwarz methods without overlap for the helmholtz equation. *SIAM J. Numer. Anal.*, 2001. to appear.

[MSRKA98]Lois C. McInnes, Romeo F. Susan-Resigna, David E. Keyes, and Hafiz M. Atassi. Additive Schwarz methods with nonreflecting boundary conditions for the parallel computation of Helmholtz problems. In *Domain decomposition methods, 10 (Boulder, CO, 1997)*, pages 325–333. Amer. Math. Soc., 1998.

[NR95]Frédéric Nataf and Francois Rogier. Factorization of the convection-diffusion operator and the Schwarz algorithm. M^3AS, 5(1):67–93, 1995.

[QV99]Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

[SBG96]Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

23 Optimized Schwarz Algorithms for Coupling Convection and Convection-Diffusion Problems

M.J. Gander[1], L. Halpern[2], C. Japhet[3]

Introduction

When solving the compressible Navier-Stokes equations in an exterior domain, it is of interest in the computation to select regions where the viscosity is small and to solve the Euler equations instead in these regions, since the Euler equations are less costly computationally. In recent years, fundamental work has been done to study the range of applicability of this approach. Error estimates have been developed for small viscosity, coupled problems have been formulated and more recently iterative algorithms have been developed to solve the coupled problems (see [Dub93], [GQL90]).

For problems in fluid mechanics new domain decomposition methods with optimized transmission conditions based on artificial boundary conditions [Hal86] have been introduced [CQ95, NR95]. In particular, it was proposed for the convection-diffusion equation to use transmission conditions such that the rate of convergence can be optimized [Jap98]. These transmission conditions lead to very fast convergence, and the convergence rate is nearly independent of both the physical and the discretization parameters.

Here we extend these transmission conditions to the case of the coupled convection and convection-diffusion problem. We consider the convection-diffusion equation

$$\begin{array}{rcll} \mathcal{L}_{cd}(u) \equiv -\nu\Delta u + \mathrm{div}(\mathbf{a}u) + cu & = & f & \text{in } \Omega, \\ \mathcal{C}(u) & = & g & \text{on } \partial\Omega, \end{array} \tag{1}$$

where Ω is a bounded open set of $\mathbb{R}^2$, and $\mathcal{C}$ is a linear operator such as the identity or the normal derivative. Here $\nu > 0$ is the viscosity, $c > 0$ is a constant and $\mathbf{a} = (a, b) \in (L^\infty(\Omega))^2$ is the velocity field with $\mathrm{div}\mathbf{a} \in L^\infty(\Omega)$ and $\mathrm{div}\mathbf{a} + c \geq \delta > 0$. This ensures that the problem is well-posed, because it can be associated with a continuous and coercive bilinear form.

We suppose that the diffusion process is only physically relevant in a subregion Ω_- of Ω. Let $\overline{\Omega} = \overline{\Omega}_- \cup \overline{\Omega}_+$ with $\Omega_- \cap \Omega_+ = \emptyset$. We denote by Γ the common interface between Ω_- and Ω_+ and by $\mathbf{n}$ the unit outward normal for Ω_-. To solve the original problem (1), we want to use the fact that the diffusion is only relevant in Ω_-. We therefore couple the convection-diffusion equation

$$\mathcal{L}_{cd}(v) \equiv -\nu\Delta v + \mathrm{div}(\mathbf{a}v) + cv = f \quad \text{in } \Omega_-$$

with the convection equation

$$\mathcal{L}_c(w) \equiv \mathrm{div}(\mathbf{a}w) + cw = f \quad \text{in } \Omega_+$$

with $\mathcal{C}$ imposed on $\partial\Omega_- \cap \partial\Omega$ and $\partial\Omega_+ \cap \partial\Omega$ and with suitable transmission conditions on Γ.

[1]Department of Mathematics and Statistics, McGill University, Montreal, QC, CANADA
[2]LAGA, Université Paris XIII, Avenue J-B Clément, 93430 Villetaneuse, France
[3]LAGA, Université Paris XIII, Avenue J-B Clément, 93430 Villetaneuse, France

We first present the optimized Schwarz algorithm for $\Omega = \mathbb{R}^2$ to show the link between transmission conditions and artificial boundary conditions. We consider both inflow into the purely convective region, $\mathbf{a} \cdot \mathbf{n} > \mathbf{0}$, and outflow of the purely convective region, $\mathbf{a} \cdot \mathbf{n} < \mathbf{0}$. Then we present the optimized Schwarz algorithm for an arbitrary velocity field. We recall error estimates for small viscosity and compare in numerical experiments the new optimized Schwarz method for coupled problems to an earlier coupling algorithm in [GQL90].

Inflow into the Purely Convective Region

Let $\Omega = \mathbb{R}^2$, $\Omega_- = \mathbb{R}^- \times \mathbb{R}$, $\Omega_+ = \mathbb{R}^+ \times \mathbb{R}$ and $\Gamma = \{(x, y),\ y \in \mathbb{R},\ x = 0\}$. In the case of inflow into the purely convective region, the coupling on Γ needs both a condition on v and a condition on w. Let Λ_+ be the Dirichlet to Neumann operator of the left half plane defined by

$$\Lambda_+(g) = \frac{\partial u}{\partial x} \quad \text{where } u \text{ solves} \quad \begin{cases} \mathcal{L}_{cd}(u) = 0 & \text{in } \Omega_-, \\ u = g & \text{on } \Gamma, \\ u & \text{bounded at infinity.} \end{cases}$$

If the coefficients of $\mathcal{L}_{cd}$ are constants, we can compute the symbol of Λ_+ using a Fourier transform in the y direction. The symbol is given by the root with positive real part of the characteristic polynomial

$$-\nu\lambda^2 + a\lambda + (\nu k^2 + ibk + c) = 0.$$

Then $(\frac{\partial}{\partial x} - \Lambda_+)$ is the transparent operator on Γ for the convection diffusion problem in Ω_- (see [Hal86]). If we consider the Schwarz algorithm

$$\begin{cases} \mathcal{L}_{cd}(v^n) = f & \text{in } \Omega_- \\ \mathcal{L}_c(v^n) = \mathcal{L}_c(w^{n-1}) & \text{on } \Gamma \end{cases} \qquad \begin{cases} \mathcal{L}_c(w^n) = f & \text{in } \Omega_+ \\ \left(\frac{\partial}{\partial x} - \Lambda_+\right)(w^n) = \left(\frac{\partial}{\partial x} - \Lambda_+\right)(v^{n-1}) & \text{on } \Gamma \end{cases} \tag{2}$$

then, because the transmission operators are the transparent operators for Ω_- and Ω_+, we have the following optimal convergence result.

Theorem 1 *The algorithm (2) converges in 2 iterations to the solution of the coupled problem*

$$\begin{cases} \mathcal{L}_{cd}(v) = f & \text{in } \Omega_-, \\ \mathcal{L}_c(w) = f & \text{on } \Omega_+, \end{cases} \qquad \begin{aligned} v &= w && \text{on } \Gamma, \\ \tfrac{\partial v}{\partial x} &= \tfrac{\partial w}{\partial x} && \text{on } \Gamma. \end{aligned} \tag{3}$$

Note that the coupling conditions satisfied at convergence in (3) are the coupling conditions satisfied by the original viscous problem in both subdomains. The continuity of both the values and normal derivatives in the coupled solution seems to be important, since we neglect the diffusion term only for computational purposes, not because the diffusion is physically zero in one subdomain. This is an important distinction from the transmission conditions derived in [GQL90] from a mathematical point of view, which led to a coupled solution with jumps in the the normal derivatives across the artificial interface.

The transmission condition for Ω_+ in the optimal Schwarz algorithm (2) involves a non-local operator, which requires a convolution along the interface in a numerical implementation. To avoid this, we replace the non-local operator Λ_+ by a local approximation given by a differential operator in the y variable, which leads to the new transmission condition

$$\mathcal{B}_+ = \frac{\partial}{\partial x} - \alpha - \beta\frac{\partial}{\partial y} - \gamma\frac{\partial^2}{\partial y^2}$$

where $\alpha > 0$, $\gamma \geq 0$ and the coefficients α, β and γ are chosen to optimize the convergence rate of the Schwarz algorithm as it was done for convection-diffusion problems in [Jap98]. The optimized Schwarz algorithm for the coupled problem is therefore given by

$$\left\{\begin{array}{ll} \mathcal{L}_{cd}(v^n) = f & \text{in } \Omega_- \\ \mathcal{L}_c(v^n) = \mathcal{L}_c(w^{n-1}) & \text{on } \Gamma \end{array}\right. \qquad \left\{\begin{array}{ll} \mathcal{L}_c(w^n) = f & \text{in } \Omega_+ \\ \mathcal{B}_+(w^n) = \mathcal{B}_+(v^{n-1}) & \text{on } \Gamma \end{array}\right. \tag{4}$$

Remark 1 *Note that on the interface, $\frac{\partial w^n}{\partial x}$ can be replaced by $\frac{1}{a}(f - cw^n - b\frac{\partial w^n}{\partial y})$ using the convection equation in Ω_+.*

A priori estimates show the well-posedness as in [GQL90] and [NR95].

Theorem 2 *Let $H^{1,1}(\Omega_\pm) = \{v \in H^1(\Omega_\pm), v|_\Gamma \in H^1(\Gamma)\}$. Then the algorithm (4) has a unique solution (v^n, w^n) in $H^{1,1}(\Omega_-) \times H^{1,1}(\Omega_+)$.*

Because the transmission condition for Ω_- is still transparent for w^2 we have

Theorem 3 *The algorithm (4) converges in 3 iterations to the solution of the coupled problem (3). More precisely we have $v^2 = v$, $w^3 = w$.*

Outflow of the Purely Convective Region

In this case only one transmission condition can be imposed and we choose here to impose the continuity of the function values, $v = w$, on Γ. Note that one could also choose continuity of the normal derivatives or a linear combination. The boundary conditions imposed on the purely convective region $\partial\Omega \cap \partial\Omega_+$ are such that w is uniquely determined by

$$\mathcal{L}_c(w) = f \quad \text{in } \Omega_+$$

without information required from the other subdomain (see [GQL90]). With the solution w, the solution v on the other subdomain is then defined by

$$\left\{\begin{array}{rcll} \mathcal{L}_{cd}(v) & = & f & \text{in } \Omega_-, \\ v & = & w & \text{on } \Gamma \end{array}\right.$$

and there is no need to iterate.

The Case of Mixed Inflow and Outflow

We define $\Gamma_{out} = \{x \in \Gamma, \mathbf{a} \cdot \mathbf{n} < \mathbf{0}\}$ and $\Gamma_{in} = \{x \in \Gamma, \mathbf{a} \cdot \mathbf{n} > \mathbf{0}\}$ with $\Gamma_{in} \cap \Gamma_{out} = \emptyset$ and $\overline{\Gamma}_{in} \cup \overline{\Gamma}_{out} = \Gamma$ as shown in Figure 1.
We propose the optimized Schwarz algorithm

$$\left\{\begin{array}{rcll} \mathcal{L}_{cd}(v^n) & = & f & \text{in } \Omega_- \\ \mathcal{L}_c(v^n) & = & \mathcal{L}_c(w^{n-1}) & \text{on } \Gamma_{in} \\ v^n & = & w^{n-1} & \text{on } \Gamma_{out} \end{array}\right. \qquad \left\{\begin{array}{rcll} \mathcal{L}_c(w^n) & = & f & \text{in } \Omega_+ \\ \mathcal{B}_+(w^n) & = & \mathcal{B}_+ v^{n-1} & \text{on } \Gamma_{in} \end{array}\right. \tag{5}$$

Again, a priori estimates lead to the following

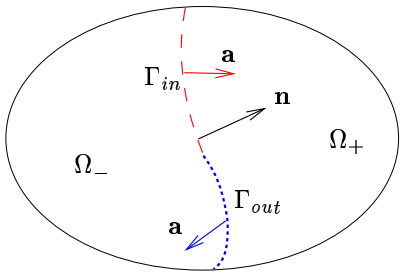


Figure 1: A problem with both inflow and outflow along the artificial interface.

Theorem 4 *The algorithm (5) is well-posed in* $H^{1,1}(\Omega_-) \times H^{1,1}(\Omega_+)$.

We do not yet have a convergence proof of algorithm (5), but numerical experiments show that the iterates (v^n, w^n) of the optimized Schwarz method (5) converge to the solution (v, w) of the coupled problem

$$\left\{ \begin{array}{rcll} \mathcal{L}_{cd}(v) & = & f & \text{in } \Omega_-, \\ \mathcal{L}_{c}(w) & = & f & \text{on } \Omega_+, \end{array} \right. \qquad \left\{ \begin{array}{rcll} v & = & w & \text{on } \Gamma, \\ \frac{\partial v}{\partial n} & = & \frac{\partial w}{\partial n} & \text{on } \Gamma_{in}. \end{array} \right.$$

Estimates for Small Viscosity

Let $\Omega = \mathbb{R}^2$, $\Omega_- = \mathbb{R}^- \times \mathbb{R}$, $\Omega_+ = \mathbb{R}^+ \times \mathbb{R}$ and $\Gamma = \{(x, y),\ y \in \mathbb{R},\ x = 0\}$. Let U be the solution of the convection-diffusion equation in $\mathbb{R}^2$. Dubach [Dub93] obtained for $\mathbf{a} \cdot \mathbf{n} = a > 0$ and the problem

$$\left\{ \begin{array}{rcll} a\partial_x v & = & f, & x < 0, \\ a\partial_x w - \nu \Delta w & = & f, & x > 0, \\ \left(\partial_x - \frac{a}{\nu}\right) w & = & -\frac{a}{\nu} v, & x = 0, \end{array} \right.$$

the estimates

$$\|(v - U)_x\|_{L^2(\mathbb{R}^2_-)} = 0(\frac{\nu}{a}) \quad \text{and} \quad \|(w - U)_x\|_{L^2(\mathbb{R}^2_+)} = 0((\frac{\nu}{a})^{\frac{3}{2}}).$$

For the problem

$$\left\{ \begin{array}{rcll} a\partial_x v - \nu \Delta v & = & f, & x < 0, \\ a\partial_x w & = & f, & x > 0, \end{array} \right. \qquad \left\{ \begin{array}{rcll} v & = & w, & \text{at } x = 0, \\ \partial_x v & = & \partial_x w, & \text{at } x = 0, \end{array} \right.$$

he obtained the estimates

$$\|(v - U)_x\|_{L^2(\mathbb{R}^2_-)} = 0((\frac{\nu}{a})^{\frac{3}{2}}) \quad \text{and} \quad \|(w - U)_x\|_{L^2(\mathbb{R}^2_+)} = 0((\frac{\nu}{a})^{\frac{1}{2}})$$

which will be verified by our numerical experiments.

Numerical Experiments

We discretize the global convection-diffusion problem and the subproblems in the optimized Schwarz method by upwind finite difference schemes. We call the solution of the global convection-diffusion problem the viscous solution. We use the mesh size $h = 1/200$ and both the viscous solution as well as the subdomain solutions are obtained by a direct solver. We first consider an inflow problem into the purely convective region of the domain and then a rotating velocity. We compare the results obtained with the optimized Schwarz method to the results obtained with the algorithm from [GQL90].

Inflow into the Purely Convective Region

We solve the coupled problem on the unit square using the optimized Schwarz method (4) with $\nu = 10^{-3}$, $\mathbf{a} = (0.1, 0.1)$ and $c = 10^{-6}$. The boundary conditions we use are given in Figure 2 and the interface is located at $x = 0.2$.

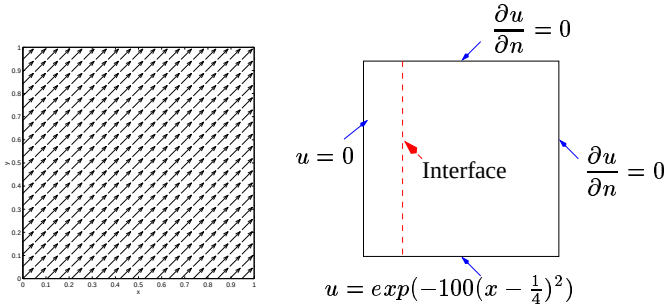


Figure 2: Convection field and boundary conditions for inflow into the purely convective region.

On the left in Figure 3 we compare the viscous solution and the solution obtained by the optimized Schwarz method for the coupled problem on the line $y = 0.01$ after 2 iterations. On the right of Figure 3 we show the results obtained using the algorithm and transmission conditions from [GQL90] obtained by letting ν go to zero. They are given by

$$\begin{cases} \left(-\nu \frac{\partial}{\partial \mathbf{n}} + \mathbf{a} \cdot \mathbf{n}\right) v^n &= -(\mathbf{a} \cdot \mathbf{n}) w^{n-1} & \text{on } \Gamma, \\ w^n &= v^{n-1} & \text{on } \Gamma_{in} \end{cases} \tag{6}$$

and do therefore not satisfy continuity of the derivatives across the interface, as one can see in Figure 3.

A Rotating Velocity

We use the optimized Schwarz method to solve the problem with $\nu = 10^{-2}$, $\mathbf{a} = (0.5 - y, 0.5)$ and $c = 10^{-6}$ and boundary conditions as given in Figure 4 on the unit square. The interface is again located at $x = 0.2$. We compare the solution obtained by the optimized Schwarz method after 3 iterations to the viscous solution. Figure 5 shows both solutions on the line $y = 0.15$ (inflow into the purely convective region) and on the line $y = 0.8$ (outflow of the purely convective region). The computed solution is continuous on the interface and also its

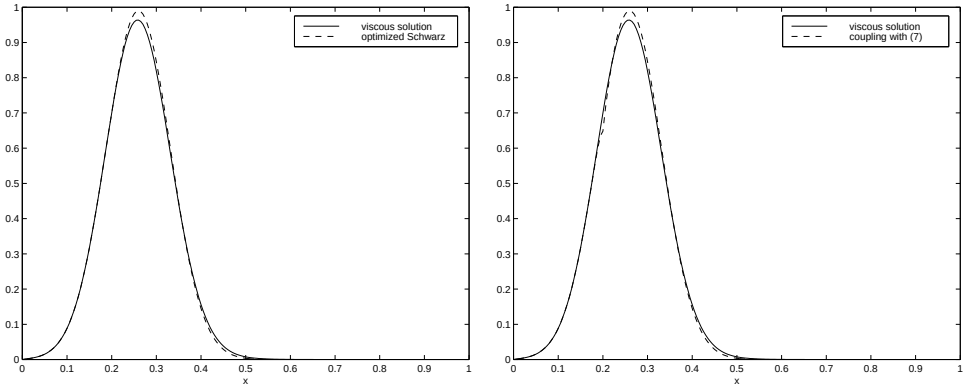


Figure 3: Result for constant velocity, the solid line denotes the viscous solution, the dashed line on the left the coupled solution obtained by the optimized Schwarz method and the dashed line on the right the solution from the algorithm proposed in [GQL90]. Note the discontinuity in the derivative on the right at $x = 0.2$.

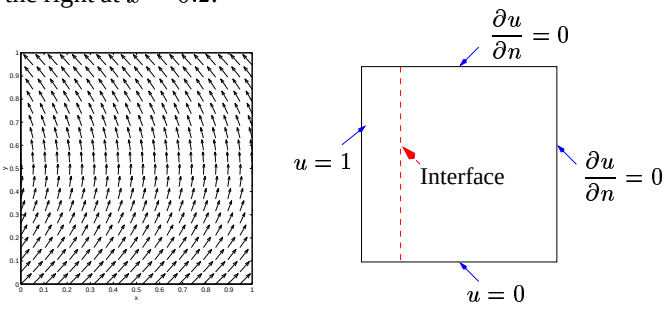


Figure 4: Convective field and boundary conditions for the rotation velocity case.

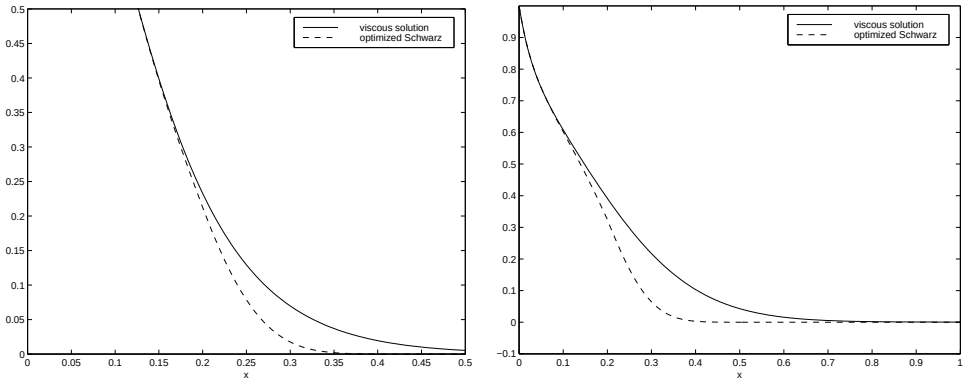


Figure 5: Result for rotating velocity on the left at $y = 0.15$ with a zoom on $[0, 0.5] \times [0, 0.5]$, on the right at $y = 0.8$. The solid line denotes the viscous solution, the dashed line the optimized Schwarz solution.

derivative is continuous on Γ_{in}. On Γ_{out} there is a small jump in the normal derivative of the solution because only one condition can be satisfied, as we have seen in the analysis. In Figure 6 we show the results obtained with the transmission conditions (6) from [GQL90] after

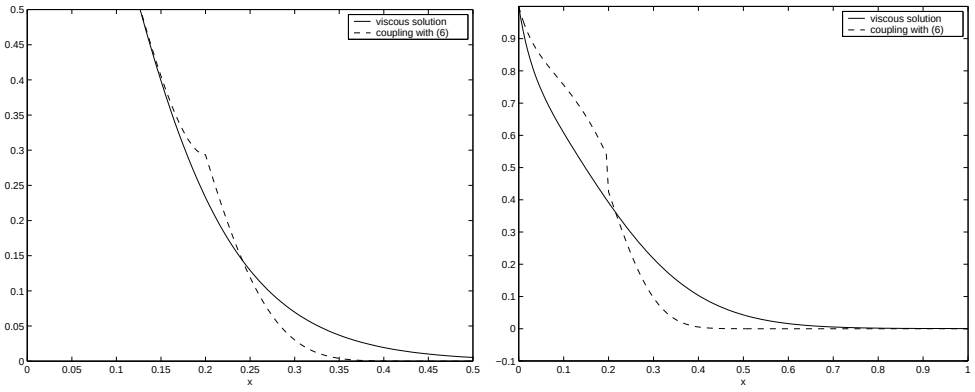


Figure 6: Same graphs as in Figure 5 but for the algorithm with transmission conditions (6) from [GQL90].

3 iterations. Note that on Γ_{in} there is a jump in the normal derivative, whereas on Γ_{out} there is a jump in the function value and in the normal derivative. The size of these discontinuities diminishes however with diminishing viscosity. Nevertheless as a physical solution to the original viscous problem, the solutions obtained by the optimized Schwarz methods seem to be preferable.

Finally we show in Figure 7 the error on the interface as a function of decreasing viscosity. The results confirm the asymptotic results from [Dub93].

References

[CQ95]Claudio Carlenzoli and Alfio Quarteroni. Adaptive domain decomposition methods for advection-diffusion problems. *The IMA Volumes in Mathematics and its Applications, Springer Verlag*, 75:165–186, 1995.

[Dub93]Eric Dubach. *Contribution a la Resolution des Equations fluides en domaine non borne*. PhD thesis, Universite Paris 13, fevrier 1993.

[GQL90]Fabio Gastaldi, Alfio Quarteroni, and Giovanni Sacchi Landriani. On the coupling of two dimensional hyperbolic and elliptic equations : Analytical and numerical approach. In T. Chan et al eds., editor, *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 22–63, Philadelphia, 1990. SIAM.

[Hal86]Laurence Halpern. Artificial boundary conditions for the advection-diffusion equations. *Math. Comp.*, 174:425–438, 1986.

[Jap98]Caroline Japhet. Optimized Krylov-Ventcell method. Application to convection-diffusion problems. In Petter E. Bjørstad, Magne S. Espedal, and David E. Keyes, editors, *Proceedings of the 9th international conference on domain decomposition methods*, pages 382–389. ddm.org, 1998.

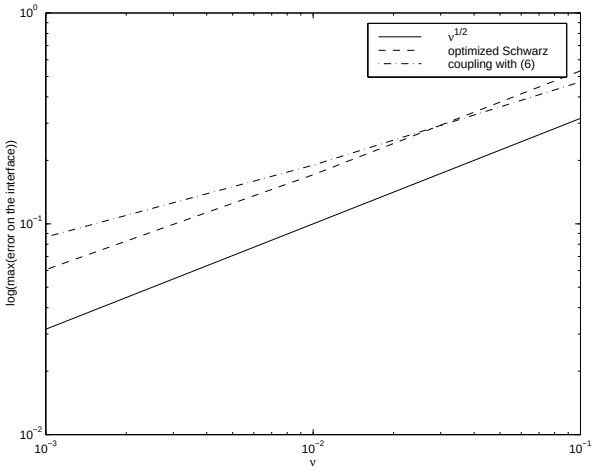


Figure 7: Asymptotic results in ν for the rotating velocity, the solid line is the reference for $\nu^{\frac{1}{2}}$, the dashed line denotes the optimized Schwarz solution, and the dotted line the solution of the algorithm with transmission conditions (6).

[NR95]Frédéric Nataf and Francois Rogier. Factorization of the convection-diffusion operator and the Schwarz algorithm. M^3AS, 5(1):67–93, 1995.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

24 Domain decomposition and virtual control for fourth order problems

P. Gervasio[1], J.-L. Lions[2], A. Quarteroni[3]

Introduction

In this paper we consider domain decomposition strategies for fourth order operators featuring a dominant second order component. More specifically, given an open and bounded domain $\Omega \subset \mathbb{R}^2$ with continuous and Lipschitz boundary $\partial\Omega$, the fourth order problem we consider reads:

$$\begin{cases} \sigma^2\Delta^2 u - \Delta u = f & \text{in } \Omega \\ u = g, \qquad \mathbf{n}\cdot\nabla u = h & \text{on } \partial\Omega \end{cases} \tag{1}$$

where $\sigma = const.$ and the functions f, g and h are assigned with sufficient regularity, while $\mathbf{n}$ is the unit outward normal vector on $\partial\Omega$.

We will partition Ω into several subdomains (overlapping or not) and consider different ways to formulate (1) at the subdomain level. In particular, we are looking for suitable control problems, the control variables being faced on the subdomain interfaces. Furthermore, we address the so-called heterogeneous case, i.e. a situation in which the coefficient σ is set to zero on a subregion of Ω. Our control approach is then devised in order to handle the coupling between the original fourth order problem and the second order one that is obtained when dropping σ out. A similar heterogeneous coupling has been previously investigated for a second-order advection diffusion problem with dominant advection (see [GLQ00]).

An outline of the paper is as follows. First the overlapping decomposition and the heterogeneous coupling are considered: a natural choice for the cost functional is introduced and it has been proved that its minimization leads to a unique solution for the coupled problem. After, the non overlapping decomposition is taken into account and both homogeneous and heterogeneous coupling are considered. Numerical results are shown for both overlapping and non-overlapping decompositions.

The overlapping situation

For the sake of exposition we consider the case of decompositions by two subdomains Ω_1 and Ω_2, which satisfy

$$\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2, \quad \Omega_1 \cap \Omega_2 \neq \emptyset, \quad \Gamma = \partial\Omega.$$

We define $\Gamma_i = \partial\Omega_i \cap \Gamma$ and $S_i = \partial\Omega_i \setminus \Gamma_i$, for $i = 1, 2$. Then $\Gamma = \Gamma_1 \cup \Gamma_2$. Further we define the differential operators

$$L_1 := -\Delta, \quad L_2 := \sigma^2\Delta^2 - \Delta.$$

[1]Department of Mathematics, University of Brescia, via Valotti 9, 25133 Brescia (Italy), Paola.Gervasio@unibs.it
[2]Collège de France, 3, rue d'Ulm, 75231 Paris Cedex 05 (France)
[3]Department of Mathematics, EPFL, CH-1015 Lausanne (Switzerland) and Department of Mathematics, Politecnico di Milano, via Bonardi 9, 20133 Milano (Italy), Alfio.Quarteroni@epfl.ch

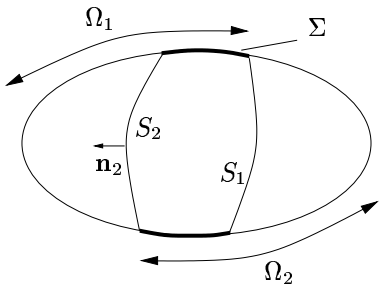


Figure 1: An overlapping decomposition of Ω in two subdomains.

The heterogeneous coupling by means of virtual control is formulated as follows:

$$\left\{\begin{array}{ll} L_1 u_1 = f & \text{in } \Omega_1 \\ u_1 = g & \text{on } \Gamma_1 \\ u_1 = \lambda_1 & \text{on } S_1 \end{array}\right. \qquad \left\{\begin{array}{lll} L_2 u_2 = f & & \text{in } \Omega_2 \\ u_2 = g, & \mathbf{n}\cdot\nabla u_2 = h & \text{on } \Gamma_2 \\ u_2 = \lambda_2, & \mathbf{n}_2\cdot\nabla u_2 = \mu_2 & \text{on } S_2 \end{array}\right. \tag{2}$$

where $\partial\Omega_i = S_i \cup \Gamma_i$, for $i = 1, 2$ (see Figure 1) and $\mathbf{n}_2$ is the unit outward normal vector on S_2.

The functions $\lambda_1,\ \lambda_2$ and μ_2 are the *virtual controls*. They are chosen in such a way that u_1 and u_2 "adjust" in the best possible way on the overlap $\Omega_1 \cap \Omega_2$. To this aim we introduce the cost functional

$$J(\lambda_1, \lambda_2, \mu_2) = \frac{1}{2}\int_{\Omega_1\cap\Omega_2} (u_1(\lambda_1) - u_2(\lambda_2, \mu_2))^2 \, d\Omega,$$

and consider the minimization problem:

$$\inf_{\lambda_1,\lambda_2,\mu_2} J(\lambda_1, \lambda_2, \mu_2). \tag{3}$$

This problem has a unique solution. Indeed, let us rewrite the solutions u_1 and u_2 of (2) as

$$u_1 = u_1^0 + v_1, \quad u_2 = u_2^0 + v_2,$$

where u_1^0 depends on the data f and g, u_2^0 depends on $f,\ g$ and h, v_1 depends on λ_1, v_2 depends on λ_2 and μ_2, and satisfy:

$$\begin{array}{llllll} L_1 u_1^0 = f & \text{in } \Omega_1, & u_1^0 = g & \text{on } \Gamma_1, & u_1^0 = 0 & \text{on } S_1, \\ L_1 v_1 = 0 & \text{in } \Omega_1, & v_1 = 0 & \text{on } \Gamma_1, & v_1 = \lambda_1 & \text{on } S_1, \end{array} \tag{4}$$

and

$$\begin{array}{llllll} L_2 u_2^0 = f & \text{in } \Omega_2, & u_2^0 = g, & \mathbf{n}\cdot\nabla u_2^0 = h & & \text{on } \Gamma_2, \\ & & u_2^0 = 0, & \mathbf{n}_2\cdot\nabla u_2^0 = 0 & & \text{on } S_2, \\ L_2 v_2 = 0 & \text{in } \Omega_2, & v_2 = 0, & \mathbf{n}\cdot\nabla v_2 = 0 & & \text{on } \Gamma_2, \\ & & v_2 = \lambda_2 & \mathbf{n}_2\cdot\nabla v_2 = \mu_2 & & \text{on } S_2. \end{array} \tag{5}$$

Then

$$J(\lambda_1, \lambda_2, \mu_2) = \frac{1}{2} Q(\lambda_1, \lambda_2, \mu_2) + \mathcal{L}(\lambda_1, \lambda_2, \mu_2),$$

where the quadratic functional Q is given by

$$Q(\lambda_1, \lambda_2, \mu_2) = \int_{\Omega_1 \cap \Omega_2} (v_1 - v_2)^2 \, d\Omega,$$

while $\mathcal{L}$ is an affine functional. Consequently, if the functions λ_i and μ_2 are smooth enough, one can define a semi-norm

$$|||\{\lambda_1, \lambda_2, \mu_2\}||| = (Q(\lambda_1, \lambda_2, \mu_2))^{1/2}, \tag{6}$$

on the space of $\{\lambda_1, \lambda_2, \mu_2\}$.
Actually, this is a *norm*. Indeed if $Q(\lambda_1, \lambda_2, \mu_2) = 0$, then $v_1 = v_2 = v$ in $\Omega_1 \cap \Omega_2$. ¿From (4) we know that $\Delta v = 0$ in $\Omega_1 \cap \Omega_2$, and $v = 0$ on $\Sigma = \partial(\Omega_1 \cap \Omega_2) \cap \partial\Omega$. Moreover, from (5) we obtain that $\mathbf{n} \cdot \nabla v = 0$ on Σ too. Thus by the continuation theorem it follows that $v \equiv 0$ in $\Omega_1 \cap \Omega_2$. Then $\lambda_1 = \lambda_2 = \mu_2 = 0$ which leads to the conclusion that (6) is a norm.

Therefore, if all data are smooth enough, $\inf J(\lambda_1, \lambda_2, \mu_2)$ admits a solution in the space of $\{\lambda_1, \lambda_2, \mu_2\}$ obtained by completion for the norm (6).

Numerical results for the overlapping heterogeneous decomposition

In order to approximate the fourth order problem by Galerkin method with Lagrangian polynomials, we consider a mixed formulation of problem (1). For the sake of simplicity we consider homogeneous boundary data, that is $g \equiv 0$ and $h \equiv 0$. The mixed formulation we have adopted reads as follows. Given $f \in L^2(\Omega)$, find $(u, w) \in V := H_0^1(\Omega) \times H^1(\Omega)$:

$$\begin{cases} (\nabla u, \nabla z)_\Omega - \sigma(\nabla w, \nabla z)_\Omega = (f, z)_\Omega & \forall z \in H_0^1(\Omega) \\ \sigma(\nabla u, \nabla v)_\Omega + (w, v)_\Omega = 0 & \forall v \in H^1(\Omega), \end{cases} \tag{7}$$

where $(\cdot, \cdot)_\Omega$ denotes the L_2 inner product in Ω.

Remark 1 *Let us set*

$$\mathcal{A}(u, w; z, v) = (\nabla u, \nabla z)_\Omega - \sigma(\nabla w, \nabla z)_\Omega + \sigma(\nabla u, \nabla v)_\Omega + (w, v)_\Omega.$$

$\mathcal{A}$ is continuous over the space V and is positive over the space $H_0^1(\Omega) \times L^2(\Omega)$. In fact $\mathcal{A}(u, w; u, w) = \|\nabla u\|^2_{L^2(\Omega)} + \|w\|^2_{L^2(\Omega)}$. Then, if the solution of (7) exists, it is unique. On the other hand, the weak form of problem (1) reads: find $u \in H_0^2(\Omega)$ such that:

$$\sigma^2(\Delta u, \Delta v)_\Omega + (\nabla u, \nabla v)_\Omega = (f, v)_\Omega \quad \forall v \in H_0^2(\Omega).$$

Existence and uniqueness of u follows by Lax-Milgram Lemma. Moreover, $u \in H^4(\Omega)$ (if Ω is regular enough) and the couple $(u, w = \sigma\Delta u)$ is a solution to problem (7).

In order to formulate the mixed heterogeneous problem we define:
$\mathring{V}_2 = H_0^1(\Omega_2) \times H^1(\Omega_2)$, $\mathring{W}_1 = H_0^1(\Omega_1)$, $V_2 = H^1_{\Gamma_2}(\Omega_2) \times H^1(\Omega_2)$, $W_1 = H^1_{\Gamma_1}(\Omega_1)$ where

$H^1_{\Gamma_i}(\Omega_i) = \{v \in H^1(\Omega_i) : \ v_{|\Gamma_i} = 0\}$. Then we solve the minimization problem (3) where $u_1 \in W_1$, $(u_2, w_2) \in V_2$ are the solutions to the following problem

$$\begin{cases} (\nabla u_2, \nabla z)_{\Omega_2} - \sigma(\nabla w_2, \nabla z)_{\Omega_2} = (f, z)_{\Omega_2} & \forall z \in H^1_0(\Omega_2) \\ \sigma(\nabla u_2, \nabla v)_{\Omega_2} + (w_2, v)_{\Omega_2} = \sigma \int_{S_2} \mu_2 v ds & \forall v \in H^1(\Omega_2) \\ (\nabla u_1, \nabla z)_{\Omega_1} = (f, z)_{\Omega_1} & \forall z \in H^1_0(\Omega_1) \\ u_1 = \lambda_1 \ \text{ on } S_1, \qquad u_2 = \lambda_2 \ \text{ on } S_2 \end{cases} \tag{8}$$

The minimization problem (3) is solved by the BFGS Quasi-Newton method with a mixed quadratic and cubic line search procedure ([JS96]), while we use a Galerkin approximation by conformal spectral elements to solve the associated problem (8).

We have considered the following domain and its decomposition:

$$\Omega = (-1,1)^2, \ \Omega_1 = (-1,.5) \times (-1,1), \ \Omega_2 = (0,1) \times (-1,1).$$

The right-hand side and the boundary data are chosen so that the analytical solution is $u(x,y) = (x^2-1)e^y + (y^2-1)e^x$.
In Ω_1 we have considered 3×2 equal spectral elements, while in Ω_2 2×2 equal spectral elements. If not otherwise specified, the polynomial degree has been set $N = 4$.
In order to assess numerically the above theory, we consider the following error terms, that we show in Table 1. The minimum value attained by the functional $J(\lambda_1, \lambda_2, \mu_2)$: $\hat{J}$; the maximum interface errors and the H^2-norm errors for $i = 1, 2$:

$$s_i := \|u_1 - u_2\|_{L^\infty(S_i)}, \quad \mathcal{E}(u)_i = \frac{\|u_i - u\|_{H^2(\Omega_i)}}{\|u\|_{H^2(\Omega_i)}}, \quad \mathcal{E}(u_N)_i = \frac{\|u_i - u_N\|_{H^2(\Omega_i)}}{\|u_N\|_{H^2(\Omega_i)}}, \tag{9}$$

where u is the analytical solution of the global fourth-order problem (1), u_i are the numerical solutions of the virtual control problem (3) and u_N is the spectral element solution of the discretized global fourth order problem (1).

σ	s_1	s_2	$\hat{J}$	$\mathcal{E}(u)_1$	$\mathcal{E}(u)_2$	$\mathcal{E}(u_N)_1$	$\mathcal{E}(u_N)_2$
1	1.90e-1	9.92e-2	6.58e-4	1.95	2.02	1.96	2.02
10^{-2}	3.64e-4	2.97e-3	1.47e-7	1.04e-3	3.08e-2	2.74e-4	3.08e-2
10^{-4}	1.28e-6	1.23e-6	6.36e-14	1.02e-3	6.96e-4	3.33e-6	3.62e-6
10^{-6}	1.25e-6	1.13e-6	6.18e-14	1.02e-3	6.96e-4	3.33e-6	1.06e-6

Table 1: Numerical results for the heterogeneous coupling with overlap.

We note that the minimum value attained by the functional J_1 tends to zero when the coefficient σ tends to zero, as well as the jumps of the solution across the interfaces. The H^2-norm errors are bounded from below by the discretization error, which depends on the spectral polynomial degree N.

The non overlapping situation

We consider now a decomposition by two disjoint subdomains Ω_1 and Ω_2 and a unique interface $S = \partial\Omega_1 \cap \partial\Omega_2$. Again, $\Gamma_i = \partial\Omega_i \cap \partial\Omega$ for $i = 1, 2$ (see Figure 2).

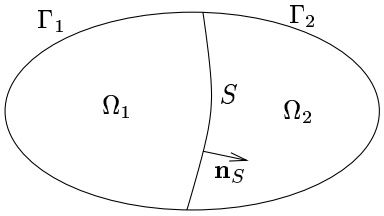


Figure 2: A partition of Ω in two disjoint subdomains.

The *homogeneous coupling* for the fourth order problem (1) would read as follows: we look for λ, μ on S which solve the minimization problem

$$\inf_{\lambda,\mu} J(u_1(\lambda,\mu), u_2(\lambda,\mu)) \tag{10}$$

where u_1 and u_2 satisfy:

$$\left\{\begin{array}{llll} L_2 u_1 = f & & & \text{in } \Omega_1 \\ u_1 = g, & \mathbf{n}\cdot\nabla u_1 = h & & \text{on } \Gamma_1 \\ u_1 = \lambda, & \mathbf{n}_S\cdot\nabla u_1 = \mu & & \text{on } S \end{array}\right. \quad \left\{\begin{array}{llll} L_2 u_2 = f & & & \text{in } \Omega_2 \\ u_2 = g, & \mathbf{n}\cdot\nabla u_2 = h & & \text{on } \Gamma_2 \\ u_2 = \lambda, & \mathbf{n}_S\cdot\nabla u_2 = \mu & & \text{on } S, \end{array}\right. \tag{11}$$

and $\mathbf{n}_S$ is the unit normal vector on S directed from Ω_1 to Ω_2.

The most natural choice of the cost functional is

$$\begin{aligned} J_1(\lambda,\mu) = \quad & \frac{1}{2}\int_S \Big[(u_1-u_2)^2 + \left(\frac{\partial u_1}{\partial n_S} - \frac{\partial u_2}{\partial n_S}\right)^2 \\ & +(\Delta u_1 - \Delta u_2)^2 + \left(\frac{\partial \Delta u_1}{\partial n_S} - \frac{\partial \Delta u_2}{\partial n_S}\right)^2\Big] ds \end{aligned} \tag{12}$$

where both u_1 and u_2 depend on the virtual controls λ and μ and $\partial/\partial n_S$ stands for $\mathbf{n}_S\cdot\nabla$.

Remark 2 *The choice of the functional J_1 is justified by the fact that the global solution of problem (7), which annihilates the right hand side of (12), is looked for in $H_0^1(\Omega)\times H^1(\Omega)$.*

Another possible choice for the cost functional is obtained by looking at the mixed formulation of problem (11) that we are going to introduce. For $i=1,2$ we define $\overset{\circ}{V}_i = H_0^1(\Omega_i)\times H^1(\Omega_i)$ and $V_i = H^1_{\Gamma_i}(\Omega_i)\times H^1(\Omega_i)$. The mixed approach for the homogeneous coupled problem (11) reads: find $(u_i, w_i)\in V_i$ for $i=1,2$ such that:

$$(\nabla u_1, \nabla z_1)_{\Omega_1} - \sigma(\nabla w_1, \nabla z_1)_{\Omega_1} = (f, z_1)_{\Omega_1} \qquad \forall z_1 \in H_0^1(\Omega_1) \tag{13}$$

$$\sigma(\nabla u_1, \nabla v_1)_{\Omega_1} + (w_1, v_1)_{\Omega_1} = \sigma\int_S \mu v_1 \qquad \forall v_1 \in H^1(\Omega_1) \tag{14}$$

$$(\nabla u_2, \nabla z_2)_{\Omega_2} - \sigma(\nabla w_2, \nabla z_2)_{\Omega_2} = (f, z_2)_{\Omega_2} \qquad \forall z_2 \in H_0^1(\Omega_2) \tag{15}$$

$$\sigma(\nabla u_2, \nabla v_2)_{\Omega_2} + (w_2, v_2)_{\Omega_2} = -\sigma\int_S \mu v_2 \qquad \forall v_2 \in H^1(\Omega_2) \tag{16}$$

$$u_1 = u_2 = \lambda \qquad \text{on } S, \tag{17}$$

and the virtual controls λ and μ are determined by the minimization problem (10).
The choice of the functional is made based on the following observation. Taking z and $v \in C_0^\infty(\Omega)$ in (7) we obtain by integration by parts

$$-\Delta u + \sigma \Delta w = f \quad x - \text{a.e. in } \Omega \tag{18}$$

$$-\sigma \Delta u + w = 0 \quad x - \text{a.e. in } \Omega. \tag{19}$$

To be more general, let us assume that σ takes two different values σ_1 in Ω_1 and σ_2 in Ω_2. Then let $\varphi \in H_{00}^{1/2}(S)$ and denote by $\tilde{\varphi}_i$ an extension of φ in Ω_i such that $\tilde{\varphi}_i \in H^1(\Omega_i)$, $\tilde{\varphi}_{i|\Gamma_i} = 0$, $\tilde{\varphi}_{i|S} = \varphi$, $i = 1, 2$. Then, taking

$$z = \begin{cases} \tilde{\varphi}_1 & \text{in } \Omega_1 \\ \tilde{\varphi}_2 & \text{in } \Omega_2 \end{cases}$$

in (7) and using (18) we deduce that

$$\int_S \left(\frac{\partial u_1}{\partial n_S} - \sigma_1 \frac{\partial w_1}{\partial n_S} \right) \varphi - \int_S \left(\frac{\partial u_2}{\partial n_S} - \sigma_2 \frac{\partial w_2}{\partial n_S} \right) \varphi = 0 \quad \forall \varphi \in H_{00}^{1/2}(S). \tag{20}$$

Proceeding in a similar way in the second equation of (7), this time using (19), we obtain that

$$\int_S \left(\sigma_1 \frac{\partial u_1}{\partial n_S} - \sigma_2 \frac{\partial u_2}{\partial n_S} \right) \varphi = 0 \quad \forall \varphi \in H_{00}^{1/2}(S). \tag{21}$$

This latter condition is implicitly guaranteed by having chosen the same multiplier μ in (14) and (16). On the other hand, since problem (13)-(17) guarantees neither the continuity of w across the interface nor the transmission condition (20), we look for these properties by choosing the following cost functional

$$J_2(\lambda, \mu) = \frac{1}{2} \int_S \left[(w_1 - w_2)^2 + \left(\left(\frac{\partial u_1}{\partial n_S} - \sigma \frac{\partial w_1}{\partial n_S} \right) - \left(\frac{\partial u_2}{\partial n_S} - \sigma \frac{\partial w_2}{\partial n_S} \right) \right)^2 \right]$$

In Table 2 we show the numerical results obtained by the minimization of functional J_1, versus the coefficient σ. The quantities s_{du}, s_w and s_{dw} stand for the maximum norm of the jumps of $\partial u/\partial n_S$, w and $\partial w/\partial n_S$ on S, respectively, while $\hat{J}_1$ is the minimum value achieved by the cost functional J_1. Moreover $\mathcal{E}(u)_i$ and $\mathcal{E}(u_N)_i$ (for $i = 1, 2$) are the errors defined in (9). The jump of u on S is not reported since it is always of the same order of the machine precision.

σ	s_{du}	s_w	s_{dw}	$\hat{J}_1$	$\mathcal{E}(u)_1$	$\mathcal{E}(u)_2$	$\mathcal{E}(u_N)_1$	$\mathcal{E}(u_N)_2$
1.	3.71e-5	2.85e-5	4.73e-4	5.88e-08	9.80e-4	6.97e-4	6.21e-6	3.80e-6
10^{-2}	9.81e-7	2.72e-5	1.54e-5	2.22e-10	9.79e-4	6.96e-4	1.55e-7	1.20e-7
10^{-4}	2.03e-8	5.12e-7	2.26e-6	1.01e-12	9.79e-4	6.96e-4	1.20e-6	8.44e-7
10^{-6}	8.00e-5	1.47e-7	2.26e-8	3.41e-09	9.79e-4	6.96e-4	1.17e-6	8.23e-7

Table 2: Numerical results for the homogeneous coupling without overlap. Minimization of the functional J_1.

σ	s_{du}	s_w	s_{dw}	$\hat{J}_2$	$\mathcal{E}(u)_1$	$\mathcal{E}(u)_2$	$\mathcal{E}(u_N)_1$	$\mathcal{E}(u_N)_2$
1.	2.15e-5	2.83e-5	2.10e-4	1.20e-08	9.79e-4	6.96e-4	2.49e-6	2.47e-6
10^{-2}	3.31e-6	3.98e-6	2.07e-4	7.51e-12	9.79e-4	6.96e-4	1.31e-6	9.18e-7
10^{-4}	4.62e-6	5.32e-6	2.23e-6	1.12e-11	9.79e-4	6.96e-4	1.07e-6	7.49e-7
10^{-6}	8.00e-5	2.92e-7	2.27e-8	3.41e-09	9.79e-4	6.96e-4	1.18e-6	8.27e-7

Table 3: Numerical results for the homogeneous coupling without overlap. Minimization of the functional J_2.

σ	s_{du}	s_ϕ	$\hat{J}_3$	$\mathcal{E}(u)_1$	$\mathcal{E}(u)_2$	$\mathcal{E}(u_N)_1$	$\mathcal{E}(u_N)_2$
1.	8.28	2.47e-4	1.76e-02	1.82	1.78e-1	1.82	1.78e-1
10^{-2}	1.50e-1	8.62e-5	5.89e-05	1.13e-3	3.08e-2	5.63e-4	3.08e-2
10^{-4}	1.60e-5	3.08e-7	6.83e-09	9.79e-4	6.96e-4	1.13e-6	3.41e-6
10^{-6}	2.68e-7	2.69e-7	7.05e-13	9.79e-4	6.96e-4	1.19e-6	8.31e-7

Table 4: Numerical results for the heterogeneous coupling without overlap. Minimization of the functional J_3.

Remark 3 *When the functional J_1 is replaced by a simpler functional in which the terms depending on u_i are dropped, similar results to those of Table 2 are obtained.*

In Table 3 we show the numerical results obtained by the minimization of functional J_2.

The *heterogeneous coupling* for non overlapping situations reads as (8), where we use the virtual controls μ instead of μ_2 and a single control λ instead of λ_1 and λ_2 and then we solve the minimization problem (10). In this case we choose the following cost functional:

$$J_3(\lambda,\mu) = \frac{1}{2}\int_S \Big[\left(\frac{\partial u_1}{\partial n_S} - \frac{\partial u_2}{\partial n_S} + \sigma\frac{\partial w_2}{\partial n_S}\right)^2 + \left(\sigma\frac{\partial u_2}{\partial n_S}\right)^2 \Big] ds.$$

Note that through the minimization of J_3 we are enforcing the fulfillment of the matching conditions (20) and (21) where, this time, we have taken $\sigma_1 = 0$.

In Table 4 we show the numerical results obtained by the minimization of functional J_3 on the heterogeneous coupling without overlap. In particular we define $s_\phi = \|\partial u_1/\partial n_S - \partial u_2/\partial n_S + \sigma\partial w_2/\partial n_S\|_{L^\infty(S)}$.

As for the overlapping case, we note that the minimum value attained by the functional J_3 tends to zero when the coefficient σ tends to zero, as well as the jump of the normal derivative of u across the interface S. Again, the H^2-norm errors are bounded from below by the discretization error, which depends on the spectral polynomial degree N.

References

[GLQ00]P. Gervasio, J.-L. Lions, and A. Quarteroni. Heterogeneous coupling by virtual control methods. Technical Report 10/2000, EPFL-DMA, Lausanne, CH, 2000. Accepted for publication in Numerische Mathematik.

[JS96]J.E. Dennis Jr. and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

25 Building preconditioners for incompressible Stokes equations from saddle point solvers of smaller dimensions

L. F. Pavarino[1], O. B. Widlund[2]

Introduction

Balancing Neumann-Neumann methods are introduced and studied for incompressible Stokes equations discretized with mixed finite or spectral elements with discontinuous pressures. After decomposing the original domain of the problem into nonoverlapping subdomains, the interior unknowns, which are the interior velocity component and all except the constant pressure component, of each subdomain problem are implicitly eliminated. The resulting saddle point Schur complement is solved with a Krylov space method with a balancing Neumann-Neumann preconditioner based on the solution of a coarse Stokes problem with a few degrees of freedom per subdomain and on the solution of local Stokes problems with natural and essential boundary conditions on the subdomains. This preconditioner is of hybrid form in which the coarse problem is treated multiplicatively while the local problems are treated additively. The condition number of the preconditioned operator is independent of the number of subdomains and is bounded from above by the product of the square of the logarithm of the local number of unknowns in each subdomain and a factor that depends on the inverse of the inf-sup constants of the discrete problem and of the coarse subproblem. Numerical results show that the method is quite fast; they are also fully consistent with the theory. This work is described in much more detail in [PW02], which contains a full proof of our result as well as many more references to the literature.

The Stokes System and Discretizations

We consider the incompressible Stokes equations on a polyhedral domain $\Omega \subset \mathbf{R}^d, d = 2, 3$,

$$\begin{cases} \nu \displaystyle\int_\Omega \nabla \mathbf{u} : \nabla \mathbf{v} dx - \int_\Omega \mathrm{div}\mathbf{v}\, p dx &= \displaystyle\int_\Omega \mathbf{f}\cdot\mathbf{v} dx \quad \forall \mathbf{v} \in (H_0^1(\Omega))^d, \\ -\displaystyle\int_\Omega \mathrm{div}\mathbf{u}\, q dx &= 0 \qquad\quad \forall q \in L_0^2(\Omega), \\ \mathbf{u}_{|\partial\Omega} = \mathbf{g}, \end{cases} \tag{1}$$

where $\mathbf{f} \in (H^{-1}(\Omega))^d$, $\mathbf{g} \in (H^{1/2}(\partial\Omega))^d$, and $\int_{\partial\Omega} \mathbf{g}\cdot\mathbf{n} ds = 0$. We discretize this system with any pair of stable Stokes elements with discontinuous pressures, such as $Q_2(h) - Q_0(h)$ and $Q_2(h) - P_1(h)$ mixed finite elements (see Brezzi and Fortin [BF91]), or $Q_n - Q_{n-2}$ mixed spectral elements (see Maday, Meiron, Patera, and Rønquist [MMPR93]). This last

[1]Università di Milano, pavarino@mat.unimi.it
[2]Courant Institute of Mathematical Sciences, widlund@cims.nyu.edu

choice is not uniformly stable, since the inf-sup constant of the discrete problem decays as $\beta_n = Cn^{-(\frac{d-1}{2})}$.

The discrete system obtained has the form

$$K \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ p \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix}. \tag{2}$$

Substructuring for Saddle Point Problems

The domain Ω is decomposed into open, nonoverlapping quadrilateral (hexahedral) subdomains Ω_i, of characteristic size H, and the interface Γ, i.e.,

$$\Omega = \cup_{i=1}^N \Omega_i \cup \Gamma.$$

Here $\Gamma = \left(\bigcup_{i=1}^N \partial\Omega_i\right) \setminus \partial\Omega$. Each Ω_i typically consists of one, or a few, spectral elements of degree n or of many finite elements. We denote by Γ_h and $\partial\Omega_h$ the set of nodes belonging to the interface Γ and $\partial\Omega$, respectively. The starting point of our algorithm is the implicit elimination (static condensation) of the interior degrees of freedom, i.e., the velocity component that is supported in the open subdomains and the interior pressure components with zero average over the individual subdomains. This process is carried out by solving decoupled local Stokes problems on each subdomain Ω_i with Dirichlet boundary conditions for the velocities given on $\partial\Omega_i$. We then obtain a saddle point Schur complement problem for the interface velocities and a constant pressure in each subdomain. This reduced problem will be solved by a preconditioned Krylov space iteration normally the preconditioned conjugate gradient method.

In order to eliminate the interior degrees of freedom, we reorder the vector of unknowns as

$$\begin{bmatrix} \mathbf{u}_I \\ p_I \\ \mathbf{u}_\Gamma \\ p_0 \end{bmatrix} \begin{array}{l} \text{interior velocities} \\ \text{interior pressures with zero average} \\ \text{interface velocities} \\ \text{constant pressures in each } \Omega_i. \end{array}$$

Then, after using the same permutation, the discrete Stokes system matrix can be written as

$$\begin{bmatrix} K_{II} & K_{\Gamma I}^T \\ K_{\Gamma I} & K_{\Gamma\Gamma} \end{bmatrix} = \left[\begin{array}{cc|cc} A_{II} & B_{II}^T & A_{\Gamma I}^T & 0 \\ B_{II} & 0 & B_{I\Gamma} & 0 \\ \hline A_{\Gamma I} & B_{I\Gamma}^T & A_{\Gamma\Gamma} & B_0^T \\ 0 & 0 & B_0 & 0 \end{array}\right].$$

Eliminating the interior unknowns $\mathbf{u}_I$ and p_I by static condensation, we obtain the saddle point Schur complement system

$$S \begin{bmatrix} \mathbf{u}_\Gamma \\ p_0 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{b}} \\ 0 \end{bmatrix}$$

where

$$S = K_{\Gamma\Gamma} - K_{\Gamma I} K_{II}^{-1} K_{\Gamma I}^T =$$

$$= \begin{bmatrix} A_{\Gamma\Gamma} & B_0^T \\ B_0 & 0 \end{bmatrix} - \begin{bmatrix} A_{\Gamma I} & B_{I\Gamma}^T \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_{II} & B_{II}^T \\ B_{II} & 0 \end{bmatrix}^{-1} \begin{bmatrix} A_{\Gamma I}^T & 0 \\ B_{I\Gamma} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} S_\Gamma & B_0^T \\ B_0 & 0 \end{bmatrix},$$

and

$$\begin{bmatrix} \tilde{\mathbf{b}} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_\Gamma \\ 0 \end{bmatrix} - \begin{bmatrix} A_{\Gamma I} & B_{I\Gamma}^T \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_{II} & B_{II}^T \\ B_{II} & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{b}_I \\ 0 \end{bmatrix}.$$

One can show that S_Γ is positive definite. By using a second permutation that reorders the interior velocities and pressures subdomain by subdomain, we note that K_{II}^{-1} represents the solution of N decoupled Stokes problems, one for each subdomain and all uniquely solvable, in parallel, with Dirichlet data given on $\partial\Omega_i$, i.e., $K_{II}^{-1} = diag({K_{II}^{(i)}}^{-1})$.

The Schur complement S does not need to be explicitly assembled since only its action Sv on a vector v is needed in a Krylov iteration. This operation essentially only requires the action of K_{II}^{-1} on a vector, i.e., the solution of N decoupled Stokes problems. In other words, Sv is computed by subassembling the actions of the subdomain Schur complements $S^{(i)}$ defined for Ω_i, given by

$$S^{(i)} = K_{\Gamma\Gamma}^{(i)} - K_{\Gamma I}^{(i)} (K_{II}^{(i)})^{-1} {K_{\Gamma I}^{(i)}}^T =$$

$$= \begin{bmatrix} S_\Gamma^{(i)} & {B_0^{(i)}}^T \\ B_0^{(i)} & 0 \end{bmatrix}.$$

Once $\begin{bmatrix} \mathbf{u}_\Gamma \\ p_0 \end{bmatrix}$ is known, $\begin{bmatrix} \mathbf{u}_I \\ p_I \end{bmatrix}$ can be found by back-substitution.

This substructuring procedure can be described in terms of a space decomposition of the discrete spaces, in the spirit of the standard Schwarz framework; cf. [PW02].

A Neumann-Neumann Preconditioner

We will solve the saddle point Schur complement problem

$$S \begin{bmatrix} \mathbf{u}_\Gamma \\ p_0 \end{bmatrix} = \begin{bmatrix} S_\Gamma & B_0^T \\ B_0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_\Gamma \\ p_0 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{b}} \\ 0 \end{bmatrix}$$

by a preconditioned Krylov space method such as GMRES or PCG. We note that this Schur complement problem is positive definite on the benign subspace where the constraints hold.

We are therefore able to use the PCG because we will start and keep the iterates in the benign subspace $\mathbf{V}_{\Gamma,B} = Ker(B_0)$.

Our balancing Neumann-Neumann preconditioner is based on the solution of a coarse Stokes problem with a few degrees of freedom per subdomain and of local Stokes problems with natural and essential boundary conditions on each subdomain. This preconditioner is of hybrid form in which the coarse problem is treated multiplicatively while the local problems are treated additively; cf. [SBG96, p. 152]. It is closely analogous to the balancing Neumann-Neumann preconditioner for the positive definite case, except that the coarse and local problems are saddle point problems. For previous work on Neumann-Neumann methods for elliptic problems, see [Man93], [MB96], [LT94], [TMV98], [DW95], and the references in [PW02].

The matrix form of the preconditioner is

$$Q = Q_H + (I - Q_H S) \sum_{i=1}^{N} Q_i (I - S Q_H),$$

where the coarse operator Q_H and local operators Q_i are defined below. The preconditioned operator is then

$$T = QS = T_0 + (I - T_0) \sum_{i=1}^{N} T_i (I - T_0),$$

where $T_0 = Q_H S$ and $T_i = Q_i S$.

Coarse solver: Given a residual vector r, the coarse term $Q_H r$ is the solution of a coarse, global Stokes problem with a few velocity degrees of freedom and one constant pressure per subdomain Ω_i:

$$Q_H = R_H^T S_0^{-1} R_H,$$

where

$$R_H = \begin{bmatrix} L_0^T & 0 \\ 0 & I \end{bmatrix},$$

and

$$S_0 = R_H S R_H^T = \begin{bmatrix} L_0^T S_\Gamma L_0 & L_0^T B_0^T \\ B_0 L_0 & 0 \end{bmatrix}.$$

We will consider three choices for the matrix L_0, resulting in the coarse velocity spaces $\mathbf{V}_0^0, \mathbf{V}_0^1$, and $\mathbf{V}_0^2$, respectively. Some of the columns of L_0 are always defined in terms of the Neumann-Neumann counting functions μ_i associated with each subdomain Ω_i: μ_i is zero at the interface nodes outside $\partial\Omega_i$ while its value at any node on $\partial\Omega_i$ equals the number of subdomains shared by that node. Its pseudo inverse $\mu_i^\dagger$ is the function $1/\mu_i(x)$ for all nodes where $\mu_i(x) \neq 0$, and it vanishes at all other points of $\Gamma_h \cup \partial\Omega_h$. We note that we use the function $\mu_i^\dagger$ in all or almost all of the subdomains and for each velocity component. Then the

columns of L_0 are defined by one of the following three choices:

0) the inverse counting functions $\mu_i^\dagger$,

1) the $\mu_i^\dagger$ and the continuous coarse piecewise bi- or tri-linear functions,

2) the $\mu_i^\dagger$ and the continuous coarse piecewise bi- or tri-quadratic functions.

In order to avoid linearly dependent $\mu_i^\dagger$ functions, and hence a singular coarse space problem, we might have to drop all of the components of these functions for one subdomain, depending on the coarse triangulation.

Local solver: The local operators Q_i will only be applied to residuals of velocity fields in the benign subspace $\mathbf{V}_{\Gamma,B}$ and thus the second residual component will vanish. Each local operator Q_i is based on the solution of a local Stokes problem on Ω_i with natural boundary condition. These local problems are nonsingular for all subdomains Ω_i the boundaries of which intersect $\partial\Omega$, but they are singular otherwise, i.e., for the *floating* subdomains. To avoid possible complications with singular problems, we modify the local Stokes problems on the floating subdomains, by adding ϵ times the velocity mass matrix to the local stiffness matrix $K^{(i)}$. It can be shown that after the coarse correction, that follows the local solvers, the iterates will be independent of the pressure field computed locally.

Given a residual vector with a first component r_Γ and a zero second component, $Q_i r$ is the weighted solutions of a local Stokes problem on subdomain Ω_i with a natural boundary condition on $\partial\Omega_i \setminus \partial\Omega$:

$$Q_i r = \begin{bmatrix} R_i^T D_i^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} S_{\Gamma,\epsilon}^{(i)} & B_0^{(i)^T} \\ B_0^{(i)} & 0 \end{bmatrix}^{-1} \begin{bmatrix} D_i^{-1} R_i & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} r_\Gamma \\ 0 \end{bmatrix}.$$

Here R_i are $0,1$ restriction matrices mapping r_Γ into r_{Γ_i} and D_i are diagonal matrices representing multiplication by the counting functions μ_i. Moreover,

$$S_\epsilon^{(i)} = \begin{bmatrix} S_{\Gamma,\epsilon}^{(i)} & B_0^{(i)^T} \\ B_0^{(i)} & 0 \end{bmatrix}$$

is the local saddle point Schur complement, associated with subdomain Ω_i, of the regularized local stiffness matrix

$$K_\epsilon^{(i)} = \begin{bmatrix} A_{II,\epsilon}^{(i)} & B_{II}^{(i)^T} & A_{\Gamma I,\epsilon}^{(i)^T} & 0 \\ B_{II}^{(i)} & 0 & B_{I\Gamma}^{(i)} & 0 \\ A_{\Gamma I,\epsilon}^{(i)} & B_{I\Gamma}^{(i)^T} & A_{\Gamma\Gamma,\epsilon}^{(i)} & B_0^{(i)^T} \\ 0 & 0 & B_0^{(i)} & 0 \end{bmatrix},$$

where

$$A_\epsilon^{(i)} = A^{(i)} + \epsilon M^{(i)}.$$

ϵ is a positive parameter, and $M^{(i)}$ is the local velocity mass matrix.

This balancing Neumann-Neumann preconditioner can be associated with a subspace decomposition of the interface space; cf. [PW02]. Our main result in that paper is:

Theorem 1 *On the benign subspace $\mathbf{V}_{\Gamma,B} \times U_0$ the balancing Neumann-Neumann operator T is symmetric positive definite with respect to the S bilinear form and*

$$cond(T) \leq C(1 + \frac{1}{\beta_0})\frac{1}{\beta^2}\,\alpha,$$

where

$$\alpha = \begin{cases} (1+\log(H/h))^2 & \text{for finite elements} \\ (1+\log n)^2 & \text{for spectral elements,} \end{cases}$$

β_0 and β are the inf-sup constants of the coarse problem and the original discrete Stokes problem respectively.

We note that $\mathbf{V}_0^0$ results in a poor constant β_0, while we can prove that $\mathbf{V}_0^2$ results in a constant β_0 uniformly bounded away from 0. $\mathbf{V}_0^1$ also gives satisfactory results.

Numerical Results with $Q_n - Q_{n-2}$ Spectral Elements in the Plane

We report, in this last section, results of some numerical experiments, carried out in Matlab 5.3 on Unix workstations, for a model Stokes problem on the unit square and with homogeneous Dirichlet boundary conditions. The problem was discretized with $Q_n - Q_{n-2}$ spectral elements and the domain Ω divided into $\sqrt{N} \times \sqrt{N}$ square subdomains. After the implicit elimination of the interior unknowns, the saddle point Schur complement is solved iteratively by PCG, starting and keeping the iterations in the benign subspace $\mathbf{V}_{\Gamma,B} \times U_0$. The initial guess is always zero, the right hand side is random and uniformly distributed, and the stopping criterion is $\|r_k\|_2/\|r_0\| \leq 10^{-6}$, where r_k is the residual at the $k-$th iterate. The singularity of the local Neumann solves for the floating subdomains is avoided by shifting the diagonal of the local velocity stiffness matrices by $\epsilon = 10^{-5}$.

The iteration counts are reported in Figure 1. These results show that PCG with our balancing Neumann-Neumann preconditioner is quasi-optimal and scalable, except with the first choice of coarse space $\mathbf{V}_0^0$. In fact, we have found the $\mathbf{V}_0^0$ coarse space not to be inf-sup stable and the iteration counts of PCG seem to grow linearly with N in that case.

The maximum eigenvalue of T is reported in Figure 2 (it can be established that the minimum eigenvalue is always close to 1). The left panel of Figure 2 shows the corresponding results for Poisson equation. We note that the iteration counts for the Stokes case are just slightly worse than for the Poisson case; see [PW02] for more complete results.

References

[BF91]F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, New York – Berlin – Heidelberg, 1991.

[DW95]Maksymilian Dryja and Olof B. Widlund. Schwarz methods of Neumann-Neumann type for three-dimensional elliptic finite element problems. *Comm. Pure Appl. Math.*, 48(2):121–155, February 1995.

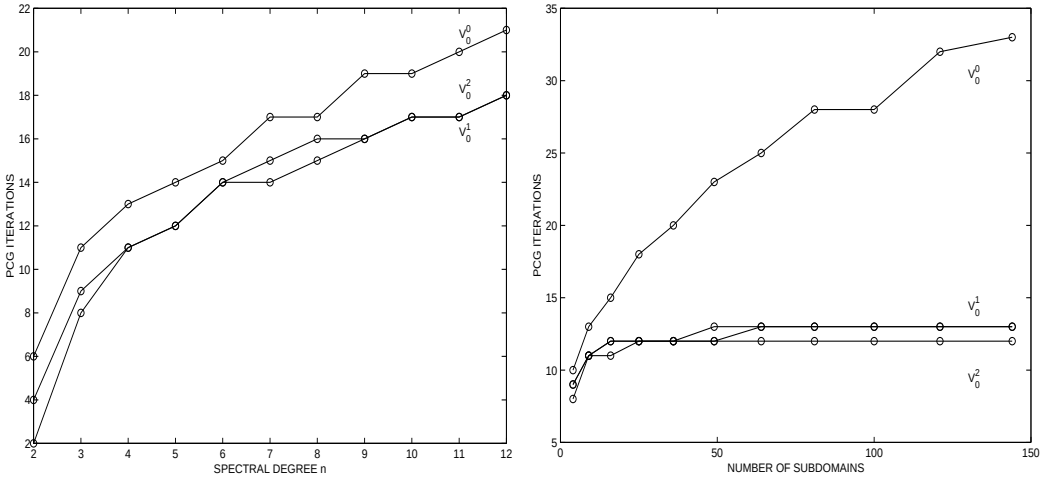


Figure 1: PCG iteration counts for the Stokes solver vs. spectral degree n when $N = 3 \times 3$ (left) and number of subdomains N when $n = 4$ (right)

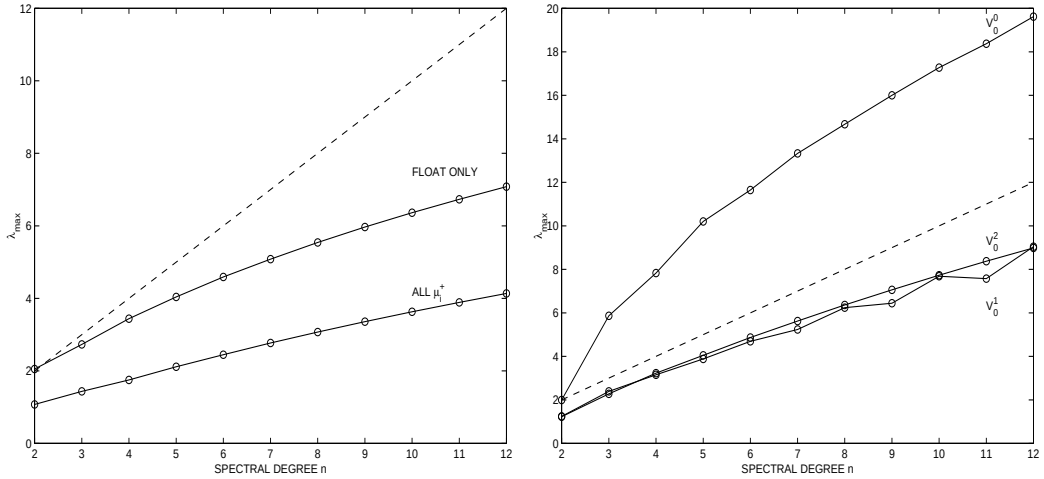


Figure 2: Maximum eigenvalue of the preconditioned operator vs. spectral degree n: Laplace solver (left) and Stokes solver (right)

[LT94]Patrick Le Tallec. Domain decomposition methods in computational mechanics. In J. Tinsley Oden, editor, *Computational Mechanics Advances*, volume 1 (2), pages 121–220. North-Holland, 1994.

[Man93]Jan Mandel. Balancing domain decomposition. *Comm. Numer. Meth. Engrg.*, 9:233–241, 1993.

[MB96]J. Mandel and M. Bresina. Balancing domain decomposition for problems with large jumps in coefficients. *Math.Comp.*, 65(216):1387–1401, 1996.

[MMPR93]Yvon Maday, Dan Meiron, Anthony T. Patera, and Einar M. Rønquist. Analysis of iterative methods for the steady and unsteady Stokes problem: Application to spectral element discretizations. *SIAM J. Sci. Comp.*, 14(2):310–337, 1993.

[PW02]Luca F. Pavarino and Olof B. Widlund. Balancing Neumann-Neumann methods for incompressible Stokes equations. *Communication on Pure and Applied Mathematics*, 55(3):302–335, 2002.

[SBG96]Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[TMV98]P. Le Tallec, J. Mandel, and M. Vidrascu. A Neumann-Neumann domain decomposition algorithm for solving plate and shell problems. *SIAM J. Numer. Math.*, 35:836–867, 1998.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

26 Multigrid for the Mortar-type Nonconforming Element Method for Nonsymmetric and Indefinite Problems

Zhong-Ci Shi [1], Xuejun Xu [2], Jinru Chen[3]

Introduction

The mortar finite element method has been used to deal with non-overlapping domain decompositions. It can handle the situation where the mesh on different subdomains need not align across interfaces, and the matching of discretizations on adjacent subdomains is only enforced weakly. In [2], Bernardi, Maday and Patera introduced basic concepts of general mortar elements, including the coupling of spectral elements with finite elements. Recently, many works have been done in constructing efficient iteration solvers for the discrete system resulting from the mortar element method. In [4], Gopalakrishnan and Pasciak presented a variable V-cycle preceonditioner, while Braess, Dahmen and Wieners [3] established another kind of W-cycle multigrid based on a hybrid formulation which gives rise to a saddle point problem. However, there are only few papers that are concerned with nonconforming elements, e.g. Marcinkowski [5] presented a P_1 nonconforming mortar element, but only for symmetric and definite problem. Meanwhile, an optimal multigrid for this method was given in [7].

The purpose of this paper is twofold. First, a mortar-type nonconforming element method is suggested for nonsymmetric and indefinite problems together with optimal error estimates. Second, a multigrid algorithm is proposed for the mortar element method which gives an optimal convergence rate, independent of the mesh size and mesh level. Finally, we describe the construction of the basis of the mortar-type nonconforming element space.

A model problem and the mortar element method

Consider the following model problem

$$\begin{cases} \mathcal{L}u = -\nabla \cdot (a\nabla u + b \cdot \nabla u) + du &= f \quad \text{in} \quad \Omega, \\ u &= 0 \quad \text{on} \quad \partial\Omega, \end{cases} \tag{1}$$

where $\Omega \subset R^2$ is a bounded polygonal domain, $a(x) = (a_{ij})$ is a uniformly symmetric positive definite tensor on $\bar{\Omega}$, $a_{ij}(x) \in C^1(\bar{\Omega})$, $b(x) \in (C^1(\bar{\Omega}))^2$, $d(x) \in C^0(\bar{\Omega})$, and $f \in L^2(\Omega)$.

[1]Institute of Computational Mathematics, Academy of Mathematices and System Sciences, Chinese Academy of Sciences, P.O.Box 2719, Beijing 100080, P.R. China, e-mail: xxj@lsec.cc.ac.cn. This work was subsidized by the special funds for major state basic research projects.

[2]Institute of Computational Mathematics, Academy of Mathematices and System Sciences, Chinese Academy of Sciences, P.O.Box 2719, Beijing 100080, P.R. China, e-mail: xxj@lsec.cc.ac.cn.

[3]Department of Mathematics, Nanjing Normal University, Nanjing, 210097, P.R. China, e-mail: jrchen@pine.njnu.edu.cn. This work was supported by the national natural science foundation of china under grant 19901014.

The variational form of (1) is to find $u \in H_0^1(\Omega)$ such that

$$a(u, v) = (f, v) \quad \forall v \in H_0^1(\Omega), \tag{2}$$

where the bilinear form

$$a(u, v) = (a\nabla u, \nabla v) - (b \cdot \nabla u, v) + (\tilde{d}u, v),$$

where $\tilde{d} = d - \nabla \cdot b$.

Assume problem (1) has the following regularity.

(H1). *For any $f \in L^2(\Omega)$, it holds that*

$$\|u\|_2 \leq C\|f\|_0.$$

We now introduce a mortar finite element method for solving (1). First, we partition Ω into nonoverlapping polygonal subdomains such that

$$\overline{\Omega} = \bigcup_{i=1}^{N} \overline{\Omega}_i \quad \text{and} \quad \Omega_i \cap \Omega_j = \emptyset, \quad i \neq j.$$

They are arranged so that the intersection of $\Omega_i \cap \Omega_j$ for $i \neq j$ is either an empty set, an edge or a vertex, i.e., the partition is geometrically conforming. The interface

$$\Gamma = \bigcup_{i=1}^{N} \partial\Omega_i \backslash \partial\Omega$$

is broken into a set of disjoint open straight segments $\gamma_m (1 \leq m \leq M)$ (that are the edges of subdomains) called mortars, i.e.

$$\Gamma = \bigcup_{m=1}^{M} \bar{\gamma}_m, \quad \gamma_m \cap \gamma_n = \emptyset, \ \text{if} \ m \neq n.$$

We denote the common open edge to Ω_i and Ω_j by γ_m. By $\gamma_{m(i)}$ we denote an edge of Ω_i which is a mortar and by $\delta_{m(j)}$ an edge of Ω_j that geometrically occupies the same place called nonmortar.

Let $\mathcal{T}_1^i$ be the coarsest triangulation of Ω_i with the mesh size h_1. The triangulation generally does not align at the subdomain interface. Denote the global mesh $\cup_i \mathcal{T}_1^i$ by $\mathcal{T}_1$. We refine the triangulation $\mathcal{T}_1$ to produce $\mathcal{T}_2$ by joining the mid-points of the edges of the triangles in $\mathcal{T}_1$. Obviously, the mesh size h_2 in $\mathcal{T}_2$ is $h_2 = h_1/2$. Repeating this process, we get the l-time refined triangulation $\mathcal{T}_l$ with mesh size $h_l = h_1 2^{-l+1}$ $(l = 1, ..., L)$. Let CR *nodal points* denote the nonconforming nodal points, i.e. the midpoints of the edges of the elements in $\mathcal{T}_l$. Moreover, on each level l, the sets of CR *nodal points* belonging to $\overline{\Omega}_i$, $\partial\Omega_i$ and $\partial\Omega$ are denoted by $\Omega_{l,i}^{CR}$, $\partial\Omega_{l,i}^{CR}$ and $\partial\Omega_l^{CR}$, respectively.

Define

$$Z = \{v | v|_{\Omega_i} \in H^1(\Omega_i), \ \forall i = 1, ..., N, \ v = 0 \text{ on } \partial\Omega\}.$$

On each level l, we define the P1 nonconforming element space locally and introduce the space $V_{l,i}(\Omega_i)$ whose functions are piecewise linear on each triangle of $\mathcal{T}_l^i$ and are continuous at the CR nodes of $\Omega_{l,i}^{CR} \backslash \partial\Omega_{l,i}^{CR}$, and equal zero at the CR nodes of $\partial\Omega_l^{CR}$.

Let

$$\tilde{V}_l = \prod_{i=1}^{N} V_{l,i} = \{v_l | v_l|_{\Omega_i} = v_{l,i} \in V_{l,i}\}.$$

Of course, we have

$$\tilde{V}_1 \not\subseteq \cdots \not\subseteq \tilde{V}_L.$$

Moreover, the P1 linear continuous finite element space over the triangulation $\mathcal{T}_l^i$ is denoted by $W_{l,i}$, whose functions have zero trace on $\partial\Omega$. Let

$$\tilde{W}_l = \prod_{i=1}^{N} W_{l,i},$$

for all $l = 1, ..., L$.

Obviously,

$$\tilde{W}_1 \subseteq \cdots \subseteq \tilde{W}_L.$$

and

$$\tilde{W}_l \subseteq \tilde{V}_l.$$

For any interface $\gamma_m = \gamma_{m(i)} = \delta_{m(j)}$, $1 \le m \le M$, there are two different and independent 1D triangulations $\mathcal{T}_l(\gamma_{m(i)})$ and $\mathcal{T}_l(\delta_{m(j)})$. Moreover, there are two sets of CR nodes belonging to γ_m: the midpoints of the elements belonging to $\mathcal{T}_l(\gamma_{m(i)})$ and to $\mathcal{T}_l(\delta_{m(j)})$ denoted by $\gamma_{l,m(i)}^{CR}$ and $\delta_{l,m(j)}^{CR}$ respectively. Additionly, we need an auxiliary test space $S_l(\delta_{m(j)})$ which is defined by

$$S_l(\delta_{m(j)}) \hat{=} \{v | v \in L^2(\delta_{m(j)}) \text{ and } v \text{ is piecewise constant} \\ \text{on the element of the nonmortar triangulation } \mathcal{T}_l(\delta_{m(j)})\}.$$

The dimension of $S_l(\delta_{m(j)})$ is equal to the number of midpoints on the $\delta_{m(j)}$, i.e. the number of elements on $\delta_{m(j)}$.

For each nonmortar $\delta_{m(j)}$, define an L^2-projection operator $Q_{l,\delta_{m(j)}} : L^2(\gamma_m) \to S_l(\delta_{m(j)})$ by

$$(Q_{l,\delta_{m(j)}} v, w)_{L^2(\delta_{m(j)})} = (v, w)_{L^2(\delta_{m(j)})} \quad \forall w \in S_l(\delta_{m(j)}),$$

where $(\cdot,\cdot)_{L^2(\delta_{m(j)})}$ denotes the L^2 inner product over the space $L^2(\delta_{m(j)})$.

Now we can introduce the following mortar finite element space for P1 nonconforming element on each level l:

$$V_l = \{v_l | v_l \in \tilde{V}_l,\ Q_{l,\delta_{m(j)}}(v_l|_{\delta_{m(j)}}) = Q_{l,\delta_{m(j)}}(v_l|_{\gamma_{m(i)}}),\ \text{for } \forall \gamma_m = \gamma_{m(i)} = \delta_{m(j)} \in \Gamma\}.$$

Define

$$\|v\|_{l,i} \hat{=} \sum_{K \in \mathcal{T}_l^i} \int_K \nabla v \cdot \nabla v dx \ \ \forall v \in V_{l,i},$$

and let

$$\|v\|_l^2 \hat{=} \sum_{i=1}^{N} \|v\|_{l,i}^2.$$

We know that $\|\cdot\|_l$ is a norm over the space V_l (see [5] for details).

Then the mortar element approximation of the problem (2) is to find $u_l \in V_l$ such that

$$a_l(u_l, v_l) = (f, v_l) \quad \forall v_l \in V_l, \tag{3}$$

where

$$\begin{aligned} a_l(u_l, v_l) &= a_l^s(u_l, v_l) + b_l(u_l, v_l) \\ a_l^s(u_l, v_l) &= \sum_{i=1}^{N} \sum_{K \in \mathcal{T}_l^i} (a\nabla u_l, \nabla v_l)_K \\ b_l(u_l, v_l) &= \sum_{i=1}^{N} \sum_{K \in \mathcal{T}_l^i} -(b \cdot \nabla u_l, v_l)_K + (\tilde{d} u_l, v_l). \end{aligned}$$

we can prove the following result.

Theorem 1 *Assume that u is the solution of (2), and $u_l \in V_l$ is the solution of (3). Then if h_l is sufficiently small, we have*

$$\|u - u_l\|_l \le C(\sum_{i=1}^{N} h_{l,i}^2 \|u\|_{2,\Omega_i}^2)^{\frac{1}{2}}.$$

Proof. We only give a brief sketch. First we can prove

$$\begin{aligned} \|u - u_l\|_l \quad \le \quad & C\{\|u - u_l\|_0 + \inf_{v_l \in V_l} \{\|u - v_l\|_0 + \|u - v_l\|_l\} \\ & + \sup_{w_l \in V_l} \frac{|a_l(u, w_l) - (f, w_l)|}{\|w_l\|_{1,l}}\}. \end{aligned}$$

Then we can show that there exists an element $v_l \in V_l$ such that

$$\begin{aligned} \|u - v_l\|_0 &\le C(\sum_{i=1}^{N} h_{l,i}^4 \|u\|_{2,\Omega_i}^2)^{1/2}, \\ \|u - v_l\|_l &\le C(\sum_{i=1}^{N} h_{l,i}^2 \|u\|_{2,\Omega_i}^2)^{1/2}, \\ \sup_{w_l \in V_l} \frac{|a_l(u, w_l) - (f, w_l)|}{\|w_l\|_{1,l}} &\le C(\sum_{i=1}^{N} h_{l,i}^2 \|u\|_{2,\Omega_i}^2)^{1/2}. \end{aligned}$$

Finally, using the idea of Schatz in [6], we can complete the proof.

Multigrid method

Due to the nonnestedness of the mesh spaces, we first introduce an intergrid transfer operator in this section. Based on this operator, a multigrid iterative method is suggested for solving (3). Some preliminary results are given in this section, which will be used to derive the convergence results of the multigrid. In the following, we always assume that the mesh sizes

$h_{l,i}$ for all i are comparable. The reason is that the convergence of multigrid always requires similar mesh parameters.

Define the operator $A_l : V_l \to V_l$ as:

$$(A_l v, w) = a_l(v, w) \quad \forall v, w \in V_l.$$

and

$$(\hat{A}_l v, w) = a_l^s(v, w) \quad \forall v, w \in V_l.$$
$$(B_l v, w) = b_l(v, w) \quad \forall v, w \in V_l.$$

It is easy to check that

$$A_l = \hat{A}_l + B_l.$$

Then (3) can be written as

$$A_l u_l = f_l,$$

where $(f_l, v) = f(v), \quad \forall v \in V_l$.

Before describing the algorithm, we must define a suitable intergrid transfer operator for the nonnested mesh space V_l. First, we give an operator $J_l^i : V_{l-1,i} \to W_{l,i}$ (see [7] for details) as follows:

- Case 1. If $p \in \Omega_{l-1,i}^{CR}$,
$$(J_l^i v)(p) = v(p).$$
- Case 2. If $p \in \Omega_{l,i}^N \backslash \Omega_{l-1,i}^{CR}$ and $p \notin \partial\Omega$,
$$(J_l^i v)(p) = \frac{1}{N(p)} \sum_{K_i} v|_{K_i}(p)$$
where $\Omega_{l,i}^N$ is the set of the vertices of the triangulation $\mathcal{T}_l^i$ that are in $\bar{\Omega}_i$ and the sum is taken over all triangles $K_i \in \mathcal{T}_l^i$ with the common vertex p and $N(p)$ is the number of those triangles.
- Case 3. If $p \in \partial\Omega \cap \partial\Omega_{l,i}^N$, then
$$(J_l^i v)(p) = 0,$$
where $\partial\Omega_{l,i}^N$ is the set of the vertices of the triangulation $\mathcal{T}_l^i$ that are in $\partial\Omega_i$.

Remark 1 Note that for different p, the value of $N(p)$ may be different. For example, if p is the vertex of triangular substructure Ω_i (see Fig.1 in [7]), then $N(p) = 1$ and if $p \in \Omega_{l,i}^N \backslash \partial\Omega_{l-1,i}^{CR}$, but $p \notin \partial\Omega_{l,i}^N$, then $N(p) = 6$, and if $p \in \partial\Omega_{l,i}^N$, but is not the vertex of substructure Ω_i and $p \notin \partial\Omega$, then $N(p) = 3$ (see Fig. 1. in [7] for details).

For the operator J_l^i, we have[7]

Lemma 1 *For $v \in V_{l-1,i}$, it holds that*

$$(1). \|J_l^i v\|_{l,i} \le C\|v\|_{l-1,i}.$$
$$(2). \|J_l^i v - v\|_0 \le Ch_l\|v\|_{l-1,i}.$$
$$(3). \|J_l^i v - v\|_{0,\gamma_m} \le Ch_l^{1/2}\|v\|_{l-1,i}.$$

where γ_m is an edge of Ω_i.

Proof. Please refer to [7] for details.

Based on the operator J_l^i, we define an intergrid transfer operator $J_l : \tilde{V}_{l-1} \to \tilde{V}_l$ as follows:

For any $v = (v_1, ..., v_N) \in \tilde{V}_{l-1}$,

$$J_l v = (J_l^1 v_1, ..., J_l^N v_N) \in \tilde{V}_l.$$

Moreover, the operator $\Xi_{l,\delta_{m(j)}} : \tilde{V}_l \to \tilde{V}_l$ is defined by

$$(\Xi_{l,\delta_{m(j)}}(v))(m_l) = \begin{cases} (Q_{l,\delta_{m(j)}}(v|_{\gamma_{m(i)}} - v|_{\delta_{m(j)}}))(m_l) & m_l \in \delta_{l,m(j)}^{CR}, \\ 0 & \text{otherwise.} \end{cases}$$

Based on above preparation, we now define an intergrid transfer operator $I_l : \tilde{V}_{l-1} \to V_l$ which will appear in the following multigrid algorithm. For any $v \in \tilde{V}_{l-1}$,

$$I_l v = J_l v + \sum_{m=1}^{M} \Xi_{l,\delta_{m(j)}}(J_l v) \in V_l.$$

Lemma 2 *For the operator I_l, we have*

$$(1). \|I_l v\|_l \le C\|v\|_{l-1};$$
$$(2). \|v - I_l v\|_0 \le C h_l \|v\|_{l-1}, \quad \forall v \in V_l.$$

Proof. Please refer to [7] for the proof.

Similar as in [1], we now describe an l-level scheme. The l-level iteration with initial guess z_0 yields $MG(l, z_0, G)$ as an approximation solution to the following problem:

Find $z \in V_l$, such that

$$a_l(z, v) = G(v) \quad \forall v \in V_l, \text{ where } G \in V_l'.$$

For $l = 1$, $MG(1, z_0, G)$ is the solution obtained by a direct method. For $l > 1$, $MG(l, z_0, G) = z_n + I_l q_p$, where $z_n \in V_l$ is constructed recursively from z_0 and the equations

$$z_i = z_{i-1} - \lambda_l^{-1}(G - A_l z_{i-1}) \quad 1 \le i \le n,$$

where λ_l is the largest eigenvalue of the operator $\hat{A}_l$. The coarse grid correction $q_p \in V_{l-1}$ is obtained by applying the $l-1$-level iteration p times ($p \ge 2$)

$$q_0 = 0, \ \ q_i = MG(l-1, q_{i-1}, \bar{G}), \quad 1 \le i \le p,$$

where $\bar{G} \in V_{l-1}'$ is defined by

$$\bar{G}(v) \hat{=} G(I_l v) - a_l(z_n, I_l v) \quad \forall v \in V_{l-1}.$$

Note that $q_p \in V_{l-1}$ is the approximation of $\bar{q}_{l-1} \in V_{l-1}$ which satisfies

$$a_{l-1}(\bar{q}_{l-1}, v) = \bar{G}(v), \quad \forall v \in V_{l-1}.$$

The main result of this paper is the following theorem

Theorem 2 *Let $p \geq 2$. If the number of the smoothing steps is large enough, and the coarsest mesh size h_1 is sufficiently small, then there exists $\delta \in (0,1)$, independent of l, such that if*

$$\|\bar{q}_{l-1} - q_p\|_{l-1} \leq C\delta^p \|\bar{q}_{l-1}\|_{l-1},$$

then

$$\|z - MG(l, z_0, G)\|_l \leq \delta \|z - z_0\|_l.$$

Proof. Here we also only provide a brief sketch. First we introduce a projection $P_{l-1} : V_l \to V_{l-1}$ defined by

$$a_{l-1}(P_{l-1}v, w) = a_l(v, I_l w), \ \ \forall v \in V_l, \ w \in W_{l-1}.$$

Then we can prove

$$\|v - I_l P_{l-1} v\|_{1,l} \leq C h_l \|A_l v\|_0, \ \ \forall v \in V_l, \tag{4}$$

$$\|P_{l-1} v\|_{1,l-1} \leq C \|v\|_{1,l}, \ \ \forall v \in V_l. \tag{5}$$

Note that $e_{n+1} = e_n - I_l q_p$, we have

$$\|e_{n+1}\|_{1,l} \leq \|e_n - I_l \bar{q}_{l-1}\|_{1,l} + \|I_l(\bar{q}_{l-1} - q_p)\|_{1,l} \equiv E_1 + E_2,$$

where $\bar{q}_{l-1} = P_{l-1} e_n$.

Finally, using Lemmas 1-2 and (4)-(5) we can get

$$E_1 \leq C \frac{1}{n^{1/2}} (1 + Ch_1)^{2n} \|e_0\|_l,$$

$$E_2 \leq C\delta^p (1 + Ch_1)^2 \|e_0\|_l.$$

Therefore, we can choose $\delta \in (0,1)$, and obtain the desired result for sufficiently small h_1.

Construction of the basis

Let $\{y_l^i\}$ denote the CR nodes of $\mathcal{T}_l$. Define operator $\varepsilon_{l,\gamma} : Z \to \tilde{V}_l$ by

$$\varepsilon_{l,\gamma} \tilde{v}(y_l^i) = \begin{cases} (Q_{l,\delta_{m(j)}}(\tilde{v}_\gamma^M - \tilde{v}_\gamma^{NM}))(y_l^i), & \text{if} \quad y_l^i \in \delta_{m(j)} = \gamma, \\ 0, & \text{otherwise}, \end{cases}$$

where $\tilde{v}_\gamma^M$ and $\tilde{v}_\gamma^{NM}$ denote the restriction of $\tilde{v} \in Z$ on mortar $\gamma_{m(i)} = \gamma$ and nonmortar $\delta_{m(j)} = \gamma$ respectively. It is easy to see that if $\tilde{v}$ is in $\tilde{V}_l$ then $v = \tilde{v} + \sum_{\gamma \in \Gamma} \varepsilon_{l,\gamma} \tilde{v}$ is an element of V_l.

Let $\{\tilde{\phi}_l^i : i = 1, \cdots, \tilde{N}_l\}$ be the basis of $V_{l,i}$. Then the basis of V_l consists of functions of the form

$$\phi_l^i = \tilde{\phi}_l^i + \sum_{\gamma \in \Gamma} \varepsilon_{l,\gamma}(\tilde{\phi}_l^i).$$

References

[1]R.E. Bank, A comparison of two multilevel iterative methods for nonsymmetric and indefinite elliptic finite element equations, SIAM J. Numer. Math., **18**(1981), 724-743.

[2]C. Bernardi, Y. Maday, and A. T. Patera, A new nonconforming approach to domain decomposition: the mortar element method. In H. Brezis and J.L. Lions, editors,*Nonlinear partial differential equations and their applications, College de France Seminar*, Volume XI, number 299 in Pitman Research Notes in Mathematics. Longman Scientific Technical, 1994.

[3]D. Braess, W. Dahmen, C. Wieners, A multigrid algorithm for the mortar finite element method, SIAM J. Numer. Anal. **37**(2000), 48-69.

[4]J. Gopalakrishnan, J. P. Pasciak, Multigrid for the mortar finite element method, SIAM J. Numer. Anal. **37**(2000), 1029-1052.

[5]L. Marcinkowski, The Mortar element method with locally nonconforming elements, BIT, **39**(1999), 716-739.

[6]A.H.Schatz, An observation concerning Ritz-Garlerkin methods with indefinite bilinear forms, **28**(1974), 956-962.

[7]X. Xu, J. Chen, Multigrid for the mortar finite element method for nonconforming P1 element, Numer. Math. **88**(2001), 381-398.

Part III

Algorithms

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

27 Recent development on Aitken-Schwarz method

J. Baranger, M. Garbey, F. Oudin-Dardun [1]

Introduction

The idea of using Aitken acceleration [Hen64] [SB80], on the classical Schwarz additive domain decomposition method has been introduced in [GTD99]. For an elliptic operator with constant coefficient on a regular grid, this method is called Aitken-Schwarz procedure, and is a direct solver. This method has shown very good numerical performances, and has been used in more complex situations [GTD01].

In this work, we extend Aitken-Schwarz procedure to the case of a 2-D cartesian grid, **not necessarily regular**. The key idea is the replacement of the 1-D Fourier transform used on the regular space step discretization of the artificial interface grid by a transform using the eigenvectors of a suitable 1-D operator. For simplicity, this presentation is limited here to the Laplacian operator and to two subdomains. However, our method can be applied to the Helmholtz operator for example and one-dimensional domain decomposition with an arbitrary number of subdomains.

In section 2, we recall the basic idea of Aitken-Schwarz method on a regular grid. In section 3, we describe two extensions of the method on a general cartesian grid : one using all the eigenvectors, the other a limited number of them. Numerical experiments are described and analyzed in section 4.

Aitken-Schwarz method on a regular grid

We first recall the basic ideas of Aitken-Schwarz method as described in [GTD01].

Let us consider a linear problem

$$L[U] = f \text{ in } \Omega,\ U_{|\partial\Omega} = 0. \tag{1}$$

We partition the domain Ω into two overlapping strips : $\Omega = \Omega_1 \cup \Omega_2$; Γ_1 (resp.Γ_2) denotes the part of the boundary of Ω_1 (resp. Ω_2) which is not included in $\partial\Omega$ (boundary of Ω). The additive Schwarz algorithm is :

$$L[u_1^{n+1}] = f \text{ in } \Omega_1,\ u_{1|\Gamma_1}^{n+1} = u_{2|\Gamma_1}^{n}, \tag{2}$$

$$L[u_2^{n+1}] = f \text{ in } \Omega_2,\ u_{2|\Gamma_2}^{n+1} = u_{1|\Gamma_2}^{n}. \tag{3}$$

We observe that the operator T, defined by :

$$T : (u_{1|\Gamma_2}^{n} - U_{|\Gamma_2}, u_{2|\Gamma_1}^{n} - U_{|\Gamma_1}) \rightarrow (u_{1|\Gamma_2}^{n+1} - U_{|\Gamma_2}, u_{2|\Gamma_1}^{n+1} - U_{|\Gamma_1})$$

[1] MCS-ISTIL - University Lyon 1, 69622 Villeurbanne, France
{baranger, garbey, foudin}@mcs.univ-lyon1.fr

is linear.

Let us consider first the one-dimensional case $\Omega = (0,1)$; then, we have the following linear (affine) relation :

$$\left\{ \begin{array}{l} u^{n+1}_{1|\Gamma_2} - U_{|\Gamma_2} = \delta_1 (u^n_{2|\Gamma_1} - U_{|\Gamma_1}), \\ u^{n+1}_{2|\Gamma_1} - U_{|\Gamma_1} = \delta_2 (u^n_{1|\Gamma_2} - U_{|\Gamma_2}), \end{array} \right. \tag{4}$$

where δ_1 (resp. δ_2) is the amplification factor associated to the operator L in subdomain Ω_1 (resp. Ω_2). Consequently :

$$\left\{ \begin{array}{l} u^{2}_{1|\Gamma_2} - u^{1}_{1|\Gamma_2} = \delta_1 (u^1_{2|\Gamma_1} - u^0_{2|\Gamma_1}), \\ u^{2}_{2|\Gamma_1} - u^{1}_{2|\Gamma_1} = \delta_2 (u^1_{1|\Gamma_2} - u^0_{1|\Gamma_2}). \end{array} \right. \tag{5}$$

So, except if the initial boundary conditions $u^0_{1|\Gamma_2}$ or $u^0_{2|\Gamma_1}$ matches with the exact solution U at the interfaces Γ_i, the amplification factors δ_1 and δ_2 can be computed from (5). Then, if $\delta_1 \delta_2 \neq 1$, the limit $U_{|\Gamma_i}$, $i = 1, 2$, is obtained as the solution of the linear system (4).

The Aitken acceleration procedure gives the exact limit of the sequence on the interface Γ_i based on two successive Schwarz iterates $u^n_{i|\Gamma_i}$, $n = 1, 2$, and the initial condition $u^0_{i|\Gamma_i}$. An additional solve of each subproblem (2), (3) with boundary conditions $u^\infty_{i|\Gamma_i}$, gives the solution of (4). The Aitken acceleration thus transforms the additive Schwarz procedure into an exact solver regardless of the speed of convergence of the original Schwarz method.

Let us consider now the 2-D Poisson problem $u_{xx} + u_{yy} = f$ in the square $\Omega = (0,1)^2$, with Dirichlet boundary conditions.

We introduce the regular discretization in the y-direction : $y_i = (i-1)h$, $h = \dfrac{1}{N-1}$, and central second-order finite differences approximation of the u_{yy} derivative.

Let us denote by $\hat{u}_i$ (resp. $\hat{f}_i$) the coefficient of the sine expansion of u (resp. f). The Poisson problem decomposes then into N independent semi-discretised equations corresponding to sinus waves $\sin(iy)$, $i = 1, ..., N$:

$$\hat{u}_{i,xx} - 4/h^2 \sin^2(ih\pi/2)\hat{u}_i = \hat{f}_i. \tag{6}$$

These N problems are linear, 1-D and independent. We can apply to each of them the Aitken acceleration procedure described above. Then, the algorithm becomes :

- step 1 : compute two iterates of Schwarz algorithm with solver of 2-dimensional subdomain problem of choice;
- step 2 : compute the sine expansion of the traces on the artificial interfaces, for the two iterates and the initial condition ; apply Aitken acceleration separately to each wave coefficients ; re-compose the trace ;
- step 3 : compute an other iteration of Schwarz algorithm.

This method gives satisfactory results. Its parallel implementation share the same communication pattern than the parallel CFD test case [ETL99]. The analysis of its parallel efficiency is therefore well known. However, this method is limited to grid with constant space step in y-direction. We proceed then with a generalization of this method to arbitrary space step grid in y direction.

Method on a general cartesian grid

Let us consider the Poisson problem $u_{xx} + u_{yy} = f$ in the square $\Omega = (0,1)^2$, with Dirichlet boundary conditions. We consider a P_1 finite element approximation on the triangles obtained by cutting each element of the cartesian mesh of Ω :

$$\Omega = \bigcup_{i,j} K_{i,j},$$

where $K_{i,j} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$, and $h_i = x_i - x_{i-1}$, $k_j = y_j - y_{j-1}$.

Let φ_{ij} be the P_1 basis function associated to the node (x_i, y_j) ; then, the finite element approximation of u is defined by :

$$u_h = \sum_{i,j} u_{ij}\varphi_{ij},$$

where the unknowns u_{ij} satisfied the following equations :

$$\begin{array}{l} \left\{ -\frac{1}{h_i}u_{i-1j} + (\frac{1}{h_i} + \frac{1}{h_{i+1}})u_{ij} - \frac{1}{h_{i+1}}u_{i+1j} \right\} \frac{k_j + k_{j+1}}{2} + \\ \left\{ -\frac{1}{k_j}u_{ij-1} + (\frac{1}{k_j} + \frac{1}{k_{j+1}})u_{ij} - \frac{1}{k_{j+1}}u_{ij+1} \right\} \frac{h_i + h_{i+1}}{2} = f_{ij}, \end{array} \tag{7}$$

for $i = 1, ..., N$ and $j = 1, ..., M$, with boundary conditions. Let us define the vector $u_i = (u_{ij})_j$; these equations appear in matrix form:

$$-\frac{1}{h_i}Ku_{i-1} + ((\frac{1}{h_i} + \frac{1}{h_{i+1}})K + \frac{h_i + h_{i+1}}{2}\overline{K})u_i - \frac{1}{h_{i+1}}Ku_{i+1} = f_i,$$

where the matrix K and $\overline{K}$ are defined by : $K = diag(\frac{1}{k_j} + \frac{1}{k_{j+1}})$ and $\overline{K} = tridiag(-\frac{1}{k_j}, (\frac{1}{k_j} + \frac{1}{k_{j+1}}), -\frac{1}{k_{j+1}})$.

Remark 1 : In the case of a uniform mesh : $h_i = k_j = h$, the equations (7) becomes:

$$-u_{i+1j} - u_{i-1j} + 4u_{ij} - u_{ij-1} - u_{ij+1} = f_{ij}.$$

If we consider the sine expansion of u_{ij} :

$$u_{ij} = \sum_{k=1}^{M} \hat{u}_{ik} \sin(kjh),$$

and if we notice that : $\sum_{j=1}^{M} \sin(kjh)\sin(ljh) = \frac{M+1}{2}\delta_{kl}$, then we obtain the discrete analogue of (6) :

$$-\hat{u}_{i+1l} - \hat{u}_{i-1l} + d_l\hat{u}_{il} = \hat{f}_{il}, \tag{8}$$

for $i = 1, ..., N$, $l = 1, ..., M$ and with $d_l = 2(1 + 2\sin^2(lh/2))$. So the Aitken procedure must be applied separately on each mode •

For a general cartesian grid, coming back to (7), we define in place of the Fourier transform as in Remark 1, a new transformation based on vector Φ_l to be chosen later:

$$u_{ij} = \sum_{l=1}^{M} \hat{u}_{il} \Phi_{lj}.$$

For technical reasons, which will be clear later, we introduce weights d_i ; by multiplication of each line of equations (7) by $(\frac{2}{k_j + k_{j+1}} d_j^2 \Phi_{mj})$ and summation on the parameter j, we obtain :

$$\begin{aligned}
&\sum_{l=1}^{M} \left\{ \left(-\frac{1}{h_i}\hat{u}_{i-1l} - \frac{1}{h_{i+1}}\hat{u}_{i+1l} + (\frac{1}{h_i} + \frac{1}{h_{i+1}})\hat{u}_{il} \right) \sum_{j=1}^{M} d_j \Phi_{lj} d_j \Phi_{mj} \right. \\
&\left. + \hat{u}_{il} \frac{h_i + h_{i+1}}{2} \sum_{j=1}^{M} d_j \left(-\frac{1}{k_j}\Phi_{lj-1} - \frac{1}{k_{j+1}}\Phi_{lj+1} + (\frac{1}{k_j} + \frac{1}{k_{j+1}})\Phi_{lj} \right) \frac{2}{k_j + k_{j+1}} d_j \Phi_{mj} \right\} \\
&= \sum_{j=1}^{M} \frac{2}{k_j + k_{j+1}} f_{ij} d_j^2 \Phi_{mj}.
\end{aligned} \tag{9}$$

We use the following notations (with $D = diag(d_j)$) :

- $\tilde{c}_{lm} = \sum_{j=1}^{M} d_j \Phi_{lj} d_j \Phi_{mj} = (D\Phi_l, D\Phi_m)$;
- $\tilde{d}_{lm} = \sum_{j=1}^{M} d_j \left(-\frac{1}{k_j}\Phi_{lj-1} - \frac{1}{k_{j+1}}\Phi_{lj+1} + (\frac{1}{k_j} + \frac{1}{k_{j+1}})\Phi_{lj} \right) \frac{2}{k_j + k_{j+1}} d_j \Phi_{mj}$
  $= (D\tilde{K}\Phi_l, D\Phi_m)$,
  where $(.,.)$ denotes the discrete scalar product, $\Phi_l = (\Phi_{lj})_j$ and
  $\tilde{K} = tridiag(\frac{-2}{k_j(k_j + k_{j+1})}, \frac{2}{k_j k_{j+1}}, \frac{-2}{k_{j+1}(k_j + k_{j+1})})$.

In order to obtain an uncoupled system for the unknowns vectors $\hat{u}_l = (\hat{u}_{il})_{i=1,\ldots,M}$, we need to diagonalise simultaneously the matrices $(\tilde{c}_{lm})$ and $(\tilde{d}_{lm})$. The matrix $\tilde{K}$ is non symmetric, but the choice

$$d_j = \sqrt{\frac{k_{j+1} + k_j}{k_2 + k_1}} \tag{10}$$

implies that the matrix $D\tilde{K}D^{-1}$ is symmetric.

We now choose the vectors $D\Phi_l$ as the orthogonal family of eigenvectors of the matrix $D\tilde{K}D^{-1}$, and denote by λ_l the eigenvalue associated to $D\Phi_l$:

$$D\tilde{K}D^{-1}(D\Phi_l) = \lambda_l D\Phi_l. \tag{11}$$

Then :

$$\tilde{c}_{lm} = |D\Phi_l|^2 \delta_{lm} \tag{12}$$

and :

$$\tilde{d}_{lm} = \lambda_l |D\Phi_l|^2 \delta_{lm}. \tag{13}$$

We can now state :

Theorem 1 *With d_j defined by (10) and (λ_l, Φ_l) defined by (11), the system (9) is uncoupled in $\hat{u}_l$:*

$$\left\{ -\frac{1}{h_i}\hat{u}_{i-1l} - \frac{1}{h_{i+1}}\hat{u}_{i+1l} + ((\frac{1}{h_i} + \frac{1}{h_{i+1}}) + \frac{h_i + h_{i+1}}{2}\lambda_l)\hat{u}_{il} \right\} |D\Phi_l|^2 = \sum_{j=1}^{M} \frac{2}{k_j + k_{j+1}} f_{ij} d_j^2 \Phi_{lj},$$

for $i = 1, ..., M$.

Each small system in $\hat{u}_l$ is the analog for the general grid of the discrete version of (6) for the regular grid that is (8).

We denote by $\hat{u}_{n_1 l}$ (resp. $\hat{u}_{n_2 l}$) the transformed unknowns corresponding to the boundary Γ_1 (resp. Γ_2) ; $(\hat{u}_{n_1 l})_l$ and $(\hat{u}_{n_2 l})_l$ satisfy a relation analogous to (4). Aitken acceleration procedure is applied to each of them.

Finally, we remark that :

$$(Du_i, D\Phi_l) = \sum_m \hat{u}_{im}(D\Phi_m, D\Phi_l) = \hat{u}_{il}|D\Phi_l|^2,$$

which allows to compute the l-component of u_i for the decomposition into the vectors basis Φ_l. We can now summarize the Aitken-Schwarz algorithm extended to grid in y direction with arbitrary space step:

Algorithm 1

- Step 1 : computation of the matrix D and $D\tilde{K}D^{-1}$;
- Step 2 : research of an orthogonal family of eigenvectors of $D\tilde{K}D^{-1}$, noted $(U_l)_l$;
- Step 3 : decomposition in the base $(\Phi_l)_l$ (with $\Phi_l = D^{-1}U_l$) of u_{n_1} and u_{n_2} (which compute vectors $(\hat{u}_{n_1 m})_m$ and $(\hat{u}_{n_2 m})_m$) ;
- Step 4 : acceleration of all modes, using (3) and (4) ;
- Step 5 : recomposition of the trace, using $u_{n_i} = \sum_l \hat{u}_{n_i l}\Phi_l$ $(i = 1, 2)$.

We notice that in case of Poisson solve with multiple right-hand sides as in Pressure solve for the time integration of the unsteady Navier Stokes equation with the projection method, Step 1 and 2 can be done once for all. The arithmetic complexity of the method is then dominated by the subdomain problem solves. Otherwise the arithmetic complexity of step 1 and 2 is of order N^3, and therefore slightly higher than a fast Poisson solver. But such fast Poisson solver does not work anyway with tensorial product grid exhibiting an arbitrary space step in one of the spatial direction.

We can improve the efficiency of our method by accelerating only the P first modes ($P < M$) corresponding to the P smallest eigenvalues. The efficiency of this method depends then on how small is the damping factor for the remaining higher order modes, and how good is the truncated representation of the trace on the interface. All these question are well known for the Fourier case, but less clear for grid with arbitrary space steps.

This second algorithm writes:

Algorithm 2

- Step 1 : computation of the matrix D and $D\tilde{K}D^{-1}$;
- Step 2 : research of an orthogonal family of P eigenvectors (corresponding to the P first eigenvalues) of $D\tilde{K}D^{-1}$, noted $(U_l)_{l=1,\dots,P}$;
- Step 3 : decomposition in the base $(\Phi_l)_{l=1,\dots,P}$ (with $\Phi_l = D^{-1}U_l$) of u_{n_1} and u_{n_2} (which compute vectors $(\hat{u}_{n_1 m})_m$ and $(\hat{u}_{n_2 m})_m$) ;
- Step 4 : acceleration of the P first modes, using (3) and (4) ;
- Step 5 : recomposition of the trace, using :

$$u_{n_i} = \sum_{l=1}^{P} \hat{u}^{\infty}_{n_i l}\Phi_l + \sum_{l=P+1}^{M} \hat{u}^{2}_{n_i l}\Phi_l, \quad i = 1, 2,$$

(where $(\hat{u}^{\infty}_{n_i l})_l$ denotes the accelerated vector, and $(\hat{u}^{2}_{n_i l})_l$, the last iterated vector).

The level of truncation P should be decided adaptively, comparing for example acceleration with P modes and $P+1$ modes. We proceed now with few numerical experiments illustrating the method.

Numerical results

We consider, on the domain $\Omega =]0,1[\times]0,1[$, the Poisson problem : $-(u_{xx} + u_{yy}) = f$, with $u = g$ on $\partial\Omega$, such that the exact solution is : $u(x,y) = 150x(x-1)y(y-1)(y-1/2)$. We use a cartesian grid of Ω with 31×31 elements, uniform in x, randomize in y, (see Figure 1) and an overlap on one element.

The algorithm using all the modes gives the error and residue shown in Figure 2. The error is then of order 10^{-6}, after one Aitken acceleration. However, one can apply the method twice to reduce this error to machine precision level. Figure 3 show compares the performance of Algorithm 2 depending on the number of modes that are accelerated. Several cycles of Schwarz Aitken acceleration are applied. We are using a direct solver for the subdomains problems and P has then a marginal impact on the number of flops. For $P \leq 20$, further cycles of Aitken Schwarz acceleration does not improve the situation because the higher modes left out from the acceleration process are the limiting factor of convergence.

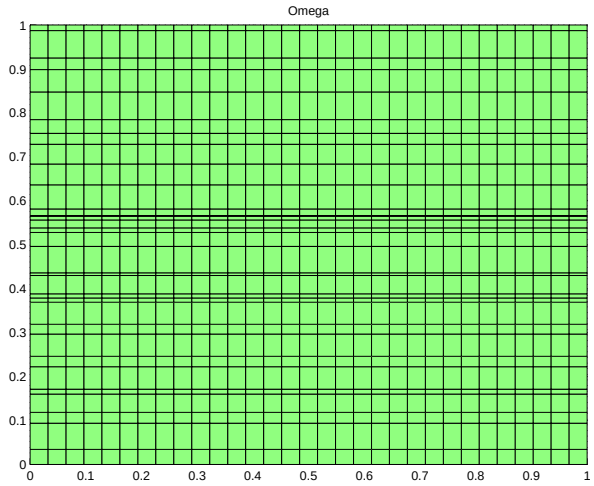


Figure 1: Mesh

Conclusion

We have shown a generalization of the so-called Aitken-Schwarz algorithm to the Poisson problem discretised on tensorial product grid with arbitrary space step in each direction. The arithmetic complexity of the method is slightly more expensive than in the case of constant space step. We expect that the Steffensen-Schwarz analogue of this method will be numerically efficient for nonlinear problem that are perturbation of the Laplace operator as the Bratu problem. However the generalization of this method to fully unstructured grid remains the interesting challenge.

References

[ETL99]A. Ecer, I. Tarkan, and E. Lemoine. Communication cost evaluations for the paecfd test case. In D. Keyes and al editors, editors, *Proc. Parallel CFD99*, 1999.

[GTD99]M. Garbey and D. Tromeur-Dervout. Operator splitting and domain decomposition for multiclusters. In D. Keyes and al editors, editors, *Proc. Parallel CFD99*, 1999. to appear.

[GTD01]M. Garbey and D. Tromeur-Dervout. Two level domain decomposition for multiclusters. In T. Chan and all editors, editors, *Domain Decomposition in Sciences and Engineering*, pages 325–339. DDM.org, 2001.

[Hen64]P. Henrici. *Elements of Numerical Analysis*. John Wiley & Sons Inc, New York-London-Sydney, 1964.

[SB80]J. Stoer and R. Burlish. *Introduction to numerical analysis*. TAM 12 Springer, New York, 1980.

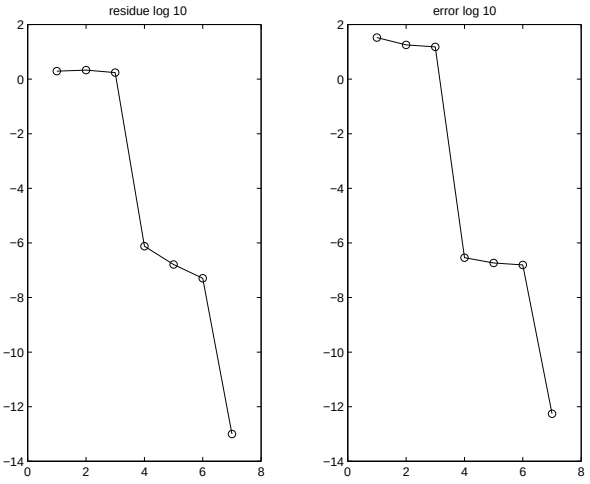


Figure 2: Error and residue - all modes

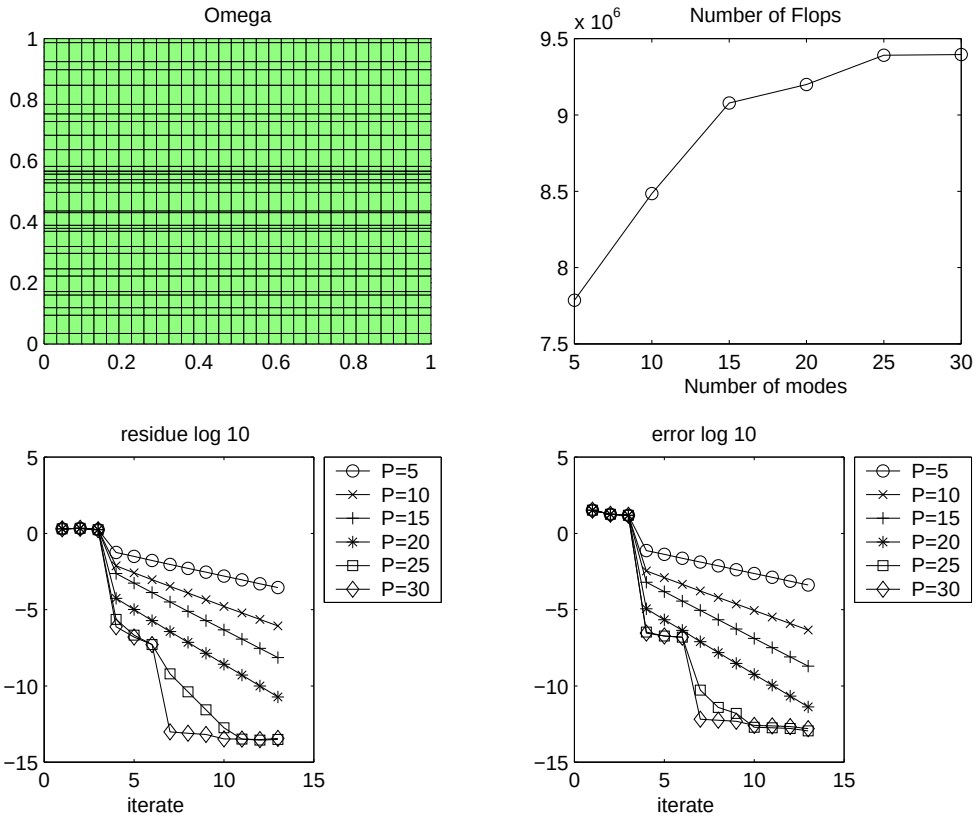


Figure 3: Error and residue - first modes

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

28 Ahpik: A Parallel Multithreaded Framework Using Adaptivity and Domain Decomposition Methods for Solving PDE Problems

A. Ben-Abdallah[1], A.S. Charão[2], I. Charpentier[3], B. Plateau[4]

Introduction

Domain decomposition methods are a valuable approach when solving partial differential equation (PDE) problems on parallel computers. In this paper, we focus our attention onto parallelization strategies for these numerical methods when dealing with irregular applications, more specifically when adaptive refinement techniques [Ver96] are applied to PDE problems involving unstructured meshes. A parallel object-oriented framework called AHPIK [CCP00] has been developed to cope with such irregular behaviors of simulations relying on domain decomposition. It provides general abstractions that are suitable for solving PDE problems on distributed memory machines using finite difference or adaptive finite element discretizations, along with overlapping or nonoverlapping, synchronous or asynchronous domain decomposition methods. One of the main features of AHPIK is the use of multithreading techniques on distributed memory machines (thus scalable) together with a message passing library (MPI). This offers a degree of freedom for traditional parallel solvers, where subdomain computations are scheduled in the context of heavyweight processes which are assigned to a given processor once for all. The use of multiple threads leads to programs that are flexible in terms of data exchange, facilitating a task scheduling with potential for masking communication overhead. Moreover the object-oriented techniques used in AHPIK make reusability, flexibility and expressiveness of source code easy.

Our goal in this paper is to show the efficiency of AHPIK concepts by comparing an AHPIK implementation with an original MPI code which solves an unsteady incompressible Navier-Stokes problem. The impact of our parallel strategy is investigated in two situations : we first consider a well-balanced distribution of the subdomains, then we induce an irregular parallel behavior by adding adaptive mesh refinement to the original code. This paper is organized in three sections: in the first one, domain decomposition methods and adaptivity techniques are discussed from a parallelism point of view. The second section presents the multithreaded framework AHPIK, while the third one provides performance results and analysis after a brief introduction to our trial application.

[1]Laboratoire de Modélisation et Calcul, IMAG, Grenoble, Adnene.Ben-Abdallah@imag.fr
[2]Laboratoire Informatique et Distribution, IMAG, Grenoble, Andrea.Charao@imag.fr
[3]Laboratoire de Modélisation et Calcul, IMAG, Grenoble, Isabelle.Charpentier@imag.fr
[4]Laboratoire Informatique et Distribution, IMAG, Grenoble, Brigitte.Plateau@imag.fr

Mathematical Methods

Domain decomposition methods (DD)

AHPIK can be used for a large variety of domain decomposition methods. In this paper, we describe its basic design ideas for the resolution of the Laplace equation by a dual Schur complement method.

In a bounded two-dimensional polygonal domain Ω, we consider the Laplace problem (1):
Find $u \in H_0^1(\Omega)$ *such that*

$$\left\{ \begin{array}{rcll} -\Delta u & = & f & \text{in } \Omega, \\ u & = & 0 & \text{on } \Gamma. \end{array} \right. \tag{1}$$

We denote by Γ the boundary of the domain and f is assumed to be square integrable. In the sequel, functions are supposed to belong to well chosen spaces, that is, the PDE problems have a unique solution.

Let choose a nonoverlapping domain decomposition $\{\Omega_k\}_{k=1,..,K}$ of Ω such that

$$\left\{ \begin{array}{l} \overline{\Omega} = \cup_{k=1,..,K} \overline{\Omega_k}, \\ \Omega_k \cap \Omega_l = \emptyset, \quad \forall (k,l) \in \{1,..,K\}^2,\ k \neq l, \end{array} \right. \tag{2}$$

We denote by Γ_k and γ_{kl} the boundaries of Ω_k that are respectively included in Γ and interfaces with other subdomains ($l = 1,..,K,\ l \neq k$) such that

$$\forall (k,l) \in \{1,..,K\}^2,\ k \neq l, \left\{ \begin{array}{l} \gamma_{kl} = \partial\Omega_k \cap \partial\Omega_l, \\ \Gamma^k = \partial\Omega_k \cap \partial\Omega. \end{array} \right. \tag{3}$$

A PDE problem is then defined on each subdomain and boundary conditions are prescribed on γ_{kl} ($(k,l) \in \{1,..,K\}^2,\ k \neq l$) to satisfy continuity constraints. A Lagrange multiplier (4) allows to write the variational formulation of the local PDE problem as:
Find u *in the appropriate space such that*

$$\left\{ \begin{array}{l} \displaystyle\sum_{k=1}^{K} \int_{\Omega_k} \nabla u_k . \nabla v_k dx = \int_{\Omega_k} f v_k dx,\ \forall v_k \text{ in the appropriate space}, \\ \displaystyle\int_{\gamma_{kl}} \mu (u_k - u_l) ds = 0 \quad \forall\, \mu \in H^{1/2}(\gamma_{kl}),\ \forall (k,l) \in \{1,..,K\}^2,\ k \neq l. \end{array} \right. \tag{4}$$

Problem (4) may be solved using Uzawa's method. Let f_k be the restrictions of function f to domains Ω_k, ν_k be the outer normals to Ω_k and λ_{kl}^0 ($(k,l) \in \{1,..,K\}^2$) be initial data. The knowledge of λ_{kl}^n at iterate n allows to compute u_k^n and λ_{kl}^{n+1} as solutions of (5) and (6):

$$\forall\, k \in \{1,..,K\}, \left\{ \begin{array}{rcll} -\Delta u_k^n & = & f_k & \text{in } \Omega_k, \\ \dfrac{\partial u_k^n}{\partial \nu_k} & = & \lambda_{kl}^n & \text{on } \gamma_{kl} \quad \forall l \in \{1,..,K\},\ l \neq k, \\ u_k^n & = & 0 & \text{on } \Gamma^k, \end{array} \right. \tag{5}$$

$$\lambda_{kl}^{n+1} = -\lambda_{lk}^{n+1} = \lambda_{kl}^n + \rho (u_k^n - u_l^n)_{|\gamma kl},\ \forall\ (k,l) \in \{1,..,K\}^2,\ k < l,$$
(parameter ρ has to be determined). (6)

The trace operator appearing in (6) is defined with respect to the domain decomposition method. The previous description then applies to both the dual Schur method when the global mesh is conform and the mortar method [BMP94] when meshes differ from one side of the interface to the other.

This iterative process may be seen as a composition of computational tasks T_{Ω_k} ($k \in \{1, .., K\}$) that solve local PDE problems (5) in domains Ω_k and computational tasks $T_{\gamma_{kl}}$ ($(k, l) \in \{1, .., K\},\ k \neq l$) that update the Lagrange multiplier (6) corresponding to interfaces γ_{kl}. The decomposition of DD algorithms into separate tasks is generic. It is already coded in AHPIK for the Schwarz overlapping method [Sch90][Lio88] (note that interface computations $T_{\gamma_{kl}}$ are empty in this case), the Schur and dual Schur complement methods and the mortar method. This approach is clearly extensible to the Dirichlet-Neumann method [MQ89]. This also applies when coding asynchronous algorithm (see for example [BT89]).

Adaptation of the discrete space

Standard *a priori* error estimates are sufficient to choose a discrete space convenient with respect to a desired accuracy. Nevertheless, in some cases, solutions may contain singularities, for which *a priori* estimates induce the refinement of all the domain for computing accurate solutions. Adaptivity is an alternate solution. It basically consists in an iterative method that computes local *a posteriori* estimates [Ver96] related to the solution at an iteration. They indicate the part of the mesh that need to be refined, thus allowing to compute a more accurate solution at a lower cost than if global refinement was used.

When solving a PDE problem in parallel via domain decomposition methods, the use of adaptive mesh refinement techniques leads to load imbalances among cooperating processors. The result is an important loss of efficiency since processors solving local PDE problem on coarse meshes may be idle, waiting for processors working on refined meshes. This is all the more true as soon as the chosen domain decomposition method is implemented with synchronous process. An interesting way to cope with this problem consists in assigning several subdomains to each processor, and let the computations be scheduled upon availability of the data they depend on. Doing so, idle times due to communications may be masked with computations. This approach can be coupled with load balancing strategies which allow to perform a new repartition of the subdomains over the processors. One can eventually "move" subdomains from one processor to another during the simulation if needed. As the management of such dynamic parallel behavior is usually cheaper with lightweight processes (threads) than with classical operating system processes, we propose to use the first ones. For a thorough discussion on the advantages of using threads for parallel irregular applications see, for example, [Chr96] and [BT98].

Overview of AHPIK

The AHPIK framework is basically composed of C++ classes that provide abstractions for developing PDE solvers based on domain decomposition methods. Two key abstractions in AHPIK are *internal* tasks and *interface* tasks. Internal tasks perform local computations, i.e., computations that require only local data within a subdomain. Tasks T_{Ω_k} identified in the previous section are examples of internal tasks: they solve local PDE problems, what usually needs solving the sparse linear equation system associated to each subdomain. Interface tasks,

on the other hand, carry out computations or updates over interface degrees of freedom. They require data from neighboring subdomains, as well as results of local computations performed by internal tasks. Tasks $T_{\gamma_{kl}}$ identified in the previous section are examples of interface tasks. Most domain decomposition methods can be described as an iterative process composed by interactions between these two types of tasks. The methods differ in terms of actual operations performed by internal and interface tasks and in the manner these tasks communicate and synchronize their execution.

Based on these ideas, AHPIK programming interface offers C++ classes which encapsulate various communication and synchronization patterns for internal and interface tasks. This includes synchronous and asynchronous algorithms, combined with different convergence control mechanisms. Writing a new domain decomposition solver then involves "filling in" the internal and interface tasks with computations, as well as specifying interface data objects that must be exchanged between processors solving neighboring subdomains. Actual communications are thus hidden from the user. Such high-level approach is achieved through object-oriented programming, which is employed in AHPIK as a means of providing strong separation between programming interface and parallel, multithreaded implementation.

Internally in AHPIK, each task is performed by a specialized thread. Additional sender/receiver threads are employed to carry out communication of boundary data needed for solving each interface problem. Threads are scheduled by the operating system upon availability of data. When a subdomain has more than one interface, interface computations can be performed in parallel by different threads as soon as their input data are available. Several subdomains can be assigned to each processor by multiplexing the set of threads performing internal and interface tasks. One can also solve uncoupled problems in parallel over the same subdomain. This is particularly interesting to efficiently exploit symmetric multiprocessor (SMP) architectures that are widely available nowadays. On such platforms, the different threads composing a parallel program can run simultaneously on different processors. Without multithreading, solving different problems at the same time over the same subdomain usually implies replicating some data. Multithreading techniques are currently used by other frameworks addressing the development of parallel PDE solvers, for example in [RHC+96] and [BBD+98]. Among these, AHPIK is distinguishable by combining multithreading with message-passing on distributed memory machines, and by being specially targeted to domain decomposition methods. The reader will find in [Cha01] a detailed description of the AHPIK framework as well as the basic ideas that have oriented its design and implementation.

Numerical Experiments

Our trial simulation model is a nonstationary incompressible flow around a cylinder with a circular cross section at Reynolds number $Re = 100$. This case corresponds to the 2D case of Schäfer and Turek's benchmark [ST96]. The flow is governed by the Navier–Stokes equations. The problem is solved using a parallel projection scheme based on mortar decomposition method, and a conjugate gradient method is used to solve the interface problem. Details are given in [Abd98].

The domain is divided into $K = 92$ nonoverlapping subdomains. A regular P_1-iso-P_2/P_1 mixed finite element triangulation is defined on each subdomain. One notices that we do not require the grids of each subdomain to match; the weak continuity through the subdomain interfaces is enforced by mortar functions. One of the particular points of this application

is that viscosity and incompressibility of the fluid are treated within two separate steps, and components of the velocity field can be computed in parallel.

Our first experiment consists in comparing the original implementation with AHPIK implementation in a case where the workload is well distributed over the subdomains. Such comparison is carried out over two different platforms : a PC cluster composed of uniprocessor nodes, and a SMP PC cluster comprising 2-processor nodes. Both clusters are homogeneous, but we notice that processors in the SMP PC cluster have higher clock speeds than the cluster of uniprocessor PCs. We use 22 nodes for each parallel execution and most nodes have 4 subdomains to solve. Figure 1 shows the duration of one iteration for both MPI-based original code and AHPIK implementation.

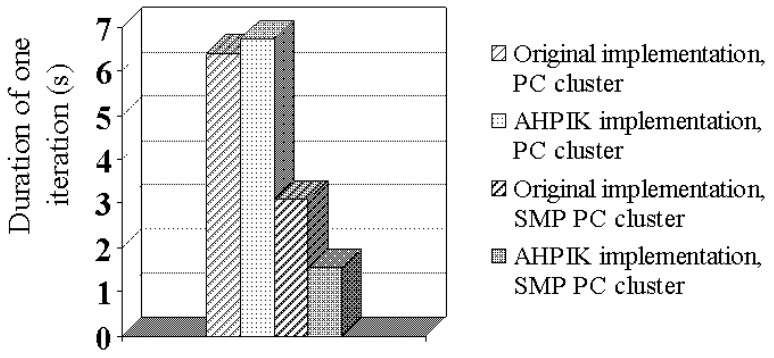


Figure 1: Results for a well-balanced distribution of the subdomains.

We see that the AHPIK version slightly decrease the performance as compared to the original implementation on uniprocessor. This can be explained by the good workload distribution that characterize this experiment. Indeed, when processor utilisation rate is high, using threads introduce overhead. In a multi-processor node, the AHPIK version produces better performance than the original implementation for identical execution parameters. We see that AHPIK implementation mixing threads and message passing automatically adapts to the multi-processor machine, while the original code keeps using only one processor.

Our second experiment consists in adding adaptive mesh refinement to either MPI-based original application and the AHPIK application. This introduces load imbalances as the number of degrees of freedom vary from one processor to another during the time iterative execution. To simplify implementation, we always refine a whole subdomain, thus we achieve the final mesh configuration within few adaptations. Figure 2 show results obtained when running the adaptive codes on each PC cluster platform. While results on the SMP cluster reproduce the behavior observed in the first experiment, results on the PC cluster composed of uniprocessor nodes show that, as long as the workload is unbalanced, the AHPIK implementation can reach or slightly surpass the performance of the MPI-based adaptive code. One can notice that this experiment reproduces a worst case situation as subdomain computations are coarse-grain and the interface solution scheme requires frequent global synchronizations. We expect better performance of the multithreaded version if synchronisation could be relaxed.

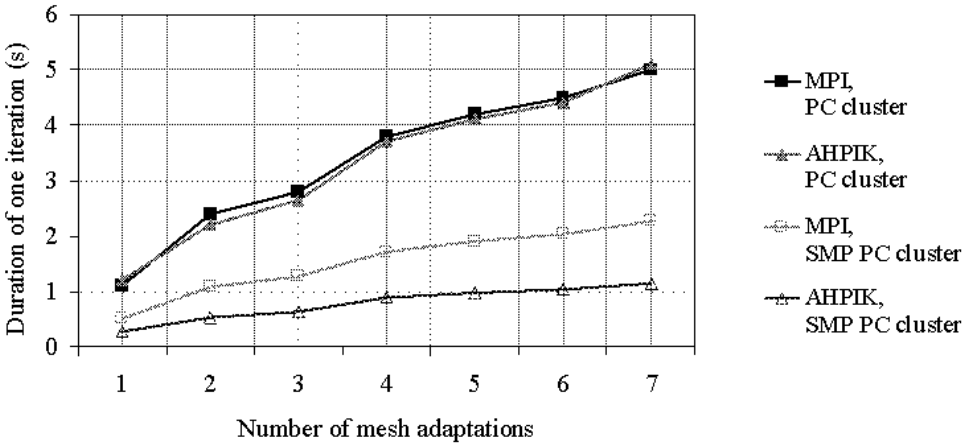


Figure 2: Results for the adaptive case.

1 Conclusion

In this paper we have introduced an object-oriented framework which uses multithreading combined with message-passing as a parallel implementation strategy for domain decomposition methods. The object-oriented approach provides general abstractions that are suitable for a variety of domain decomposition methods, including overlapping and nonoverlapping, synchronous and asynchronous methods. Such abstractions compose a programming interface where communication and synchronization details do not need to be hand-coded as in MPI-based applications.

We have investigated the performance of AHPIK compared to MPI-only domain decomposition implementation for an unsteady incompressible Navier-Stokes problem. Results show that multithreading associated to message-passing introduces more flexibility in parallel PDE solvers relying on domain decomposition, as subdomain computations are dynamically scheduled upon availability of data, and the resulting codes automatically adapt to different parallel architectures. This approach offers a potential for overlapping communication with computations when dealing with irregular applications, however the benefits of such technique are limited by the globally synchronous behavior of some numerical methods. In this sense, one of the important contributions of AHPIK rely on its support to multiple synchronization schemes that can be easily manipulated in the parallel code. This allows for an easy experimental evaluation of different numerical algorithms with different synchronization behaviors for solving a given problem.

These considerations lead us to conclude that the AHPIK approach offers a good compromise between performance and flexibility for implementing parallel PDE solvers based on domain decomposition. In a near future, we plan to use multithreading combined with message-passing to implement and evaluate dynamic load balancing strategies for adaptive PDE computations.

References

[Abd98]Adnene Ben Abdallah. *Méthode de projection pour la simulation de grandes structures turbulentes sur calculateurs parallèles*. PhD thesis, Université Pierre et Marie Curie, Paris, 1998.

[BBD+98]Federico Bassetti, David Brown, Kei Davis, William Henshaw, and Dan Quinlan. OVERTURE: An object-oriented framework for high-performance scientific computing. In *Proceedings of Supercomputing'98 (CD-ROM)*. ACM SIGARCH and IEEE, nov 1998.

[BMP94]Christine Bernardi, Yvon Maday, and Anthony T. Patera. A new non conforming approach to domain decomposition: The mortar element method. In Haim Brezis and Jacques-Louis Lions, editors, *Collège de France Seminar*. Pitman, 1994. This paper appeared as a technical report about five years earlier.

[BT89]Dimitri P. Bertsekas and John N. Tsitsiklis. *Parallel and Distributed Computation*. Prentice-Hall Inc., 1989.

[BT98]Pierre Eric Bernard and Denis Trystram. Report on a parallel molecular dynamics implementation. In E. H. D'Hollander, G. R. Joubert, F. J. Peters, and U. Trottenberg, editors, *Advances in Parallel Computing*, pages 217–220. North Holland, 1998.

[CCP00]Andréa S. Charão, Isabelle Charpentier, and Brigitte Plateau. A framework for parallel multithreaded implementation of domain decomposition methods. In E. H. D'Hollander, G. R. Joubert, F. J. Peters, and H. J. Sips, editors, *Parallel Computing: Fundamentals and Applications*, pages 95–102. Imperial College Press, 2000.

[Cha01]Andréa S. Charão. *Multiprogrammation parallèle générique des méthodes de décomposition de domaine*. PhD thesis, Institut National Polytechnique de Grenoble, 2001.

[Chr96]Nikos Chrisochoides. Multithreaded model for dynamic load balancing parallel adaptive PDE computations. *Applied Numerical Mathematics Journal*, 6:1–17, 1996.

[Lio88]Pierre-Louis Lions. On the Schwarz alternating method. I. In Roland Glowinski, Gene H. Golub, Gérard A. Meurant, and Jacques Périaux, editors, *First International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 1–42, Philadelphia, PA, 1988. SIAM.

[MQ89]Luisa D. Marini and Alfio Quarteroni. A relaxation procedure for domain decomposition methods using finite elements. *Numer. Math*, (5):575–598, 1989.

[RHC+96]John V. W. Reynders, Paul J. Hinker, Julian C. Cummings, Susan R. Atlas, Subhankar Banerjee, William F. Humphrey, Steve R. Karmesin, Katarzyna Keahey, Marikani Srikant, and Mary Dell Tholburn. POOMA: A Framework for Scientific Simulations of Paralllel Architectures. In Gregory V. Wilson and Paul Lu, editors, *Parallel Programming in C++*, chapter 14, pages 547–588. MIT Press, 1996.

[Sch90]H. A. Schwarz. *Gesammelte Mathematische Abhandlungen*, volume 2, pages 133–143. Springer, Berlin, 1890. First published in Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, volume 15, 1870, pp. 272–286.

[ST96]Michael Shäfer and Stefan Turek. Benchmark computations of laminar flow around cylinder. In *Flow Simulation with High-Performance Computers II*. Vieweg, 1996.

[Ver96]Rüdiger Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Wiley and Teubner, 1996.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

29 Efficient Schwarz Methods for Elliptic Mortar Finite Element Problems

P.E. Bjørstad[1], M. Dryja[2], T. Rahman[3]

Introduction

In this paper we investigate an additive and a hybrid Schwarz method for solving systems of algebraic equations resulting from the approximation of second order elliptic boundary value problems with (highly) discontinuous coefficients. The discretization is obtained by using the mortar finite element method on nonmatching meshes, a technique which was first introduced by Bernardi-Maday-Patera [BMP94]. Several efficient iterative methods have thereafter been developed for the mortar element, see for example [CW96, Dry96, Dry97, AMW99, CDS99, BDR00, BDW99, GP00, WK01], and the references therein. The work of this paper is a continuation of the work done in [BDR00], where two variants of the additive Schwarz methods were proposed, the average method and the coarse reformulated average method. The reformulated variant is obtained from the average variant by simply replacing its coarse space by the sum of two special coarse spaces, one associated with the subdomains and the other one defined on the skeleton of the partition of the domain. This results in an algorithm which is very well suited for parallel computation and at the same time retains the necessary convergence behavior of a good scalable additive type Schwarz method. In this paper we improve its parallel feature a step further by splitting the skeleton coarse space into two subspaces, associated with the set of vertices and the set of mortar nodes, respectively. Experiments show that this modification does not change the convergence behavior. In this connection, we also introduce a hybrid version of the method for the problem. Both methods are insensitive to jumps in the coefficients.

The remainder of this paper is organized as follows. In the next section we recall the mortar finite element method for the elliptic problem. Then, in the following two sections, we present our Schwarz methods, and in the last section, we show some preliminary numerical examples.

The Discrete Problem

Let $\overline{\Omega} = \cup_{i=1}^{N} \overline{\Omega}_i$ be the partition of the computational domain in two dimensions, where each Ω_i is a polygonal subregion (subdomain), and the subregions are nonoverlapping. We consider the following differential problem: Find $u^* \in H_0^1(\Omega)$ such that

$$a(u^*, v) = f(v),\ v \in H_0^1(\Omega), \tag{1}$$

[1]Institute for Informatics, University of Bergen, petter@ii.uib.no
[2]Department of Mathematics, University of Warsawa, dryja@mimuw.edu.pl
[3]Institute for Mathematics, University of Augsburg, talal.rahman@math.uni-augsburg.de

where

$$a(u,v) = \sum_{i=1}^{N} a_i(u,v) = \sum_{i=1}^{N} \rho_i (\nabla u, \nabla v)_{L^2(\Omega_i)},$$

and

$$f(v) = \int_\Omega fv\, dx = \sum_{i=1}^{N} \int_{\Omega_i} fv\, dx,$$

with ρ_i being positive and constant in each subregion. We remark that the proposed methods can be used as preconditioners for the problem when the coefficients ρ_i depend on x and are discontinuous only across the boundary of Ω_i. In which case, the constant ρ_i can be taken as an average of $\rho_i(x)$ over Ω_i.

We consider only the geometrically conforming case, i.e., the intersection between the closure of two different subdomains is either empty, a vertex, or a whole edge. The subdomains together form a coarse triangulation of the whole domain Ω with the mesh parameter $H = \max_i H_i$, where H_i is the diameter of Ω_i. In each subdomain Ω_i, we use triangular elements. We assume that the triangles touching the subdomain boundary $\partial\Omega_i$ are quasi-uniform, having a mesh size of order h_i. We do not put such restriction on the interior triangles. We also assume that the coarse triangulation of Ω and the fine triangulation in each Ω_i are shape regular in the sense of [Cia78]. The resulting triangulation can be nonmatching across subdomain interfaces.

Let $X_i(\Omega_i)$ be the finite element space of piecewise linear continuous functions defined on the triangulation of Ω_i and vanishing on $\partial\Omega_i \cap \partial\Omega$, and let

$$X^h(\Omega) = X_1(\Omega_1) \times X_2(\Omega_2) \cdots \times X_N(\Omega_N).$$

In order to describe the discrete problem, we need the following auxiliary notations and finite element spaces. Let Γ_{ij} be an open edge common to Ω_i and Ω_j, i.e., $\overline{\Gamma}_{ij} = \overline{\Omega}_i \cap \overline{\Omega}_j$, and let $W^{h_i}(\Gamma_{ij})$ and $W^{h_j}(\Gamma_{ij})$ be the restrictions of $X_i(\Omega_i)$ and $X_j(\Omega_j)$ onto Γ_{ij}, respectively. Note that each interface Γ_{ij} inherits two different discretizations from its two sides. We select one side of Γ_{ij} as the master side, called the mortar, and the other side as the slave side, called the nonmortar. Define the skeleton $\mathcal{S} = (\cup\partial\Omega_i) \setminus \partial\Omega$ as follows:

$$\overline{\mathcal{S}} = \cup_m \overline{\gamma}_m, \text{ and } \gamma_m \cap \gamma_n = \emptyset \text{ if } m \neq n,$$

where each γ_m denotes an open mortar edge. We write γ_m as $\gamma_{m(i)}$ if it is an edge of Ω_i, i.e., $\gamma_{m(i)} \subset \partial\Omega_i$. Let $\delta_m = \delta_{m(j)} \subset \partial\Omega_j$ be the corresponding open nonmortar edge of Ω_j that occupies the same geometrical space as $\gamma_{m(i)}$, i.e., $\gamma_{m(i)} = \Gamma_{ij} = \delta_{m(j)}$. See Fig. 1 for illustration, where a thick line is drawn on the mortar side of an interface. The thick dots are used to represent the end points of a mortar or a nonmortar. We say that a function on a mortar is nonzero if the corresponding thick line is black and zero if the edge is light gray. The same applies to the end points.

As a general rule for choosing the mortars and the nonmortars, we let $\gamma_{m(i)}$ be the mortar and $\delta_{m(j)}$ the corresponding nonmortar if $\rho_i \geq \rho_j$. This is necessary for our Schwarz methods to have a rate of convergence which is independent of the jump of the coefficients. We define by ν_i and γ_i respectively the set of vertices and the set of mortar nodes (nodes on open mortar edges) of Ω_i.

Since the triangulations on Ω_i and Ω_j may not match on their interface Γ_{ij}, the functions in $X^h(\Omega)$ can be discontinuous across the interface Γ_{ij}. A weak continuity is therefore imposed across the interface using a condition called the mortar condition. Let $u_h \in X^h$, where $u_h = \{u_i\}_{i=1}^N$. A function $u_h \in X^h$ satisfies the mortar condition on $\delta_{m(j)}$, if, for all functions $\psi \in M^{h_j}(\delta_{m(j)})$ ($\gamma_{m(i)} = \delta_{m(j)} = \Gamma_{ij}$),

$$\int_{\delta_{m(j)}} (u_{i|\gamma_{m(i)}} - u_{j|\delta_{m(j)}})\psi \, ds = 0. \tag{2}$$

Here the space $M^{h_j}(\delta_{m(j)})$ is a subspace of $W^{h_j}(\delta_{m(j)})$, with functions being constants on elements touching $\partial\delta_{m(j)}$. V^h is a subspace of X^h of functions which satisfy the mortar condition for all $\delta_m \subset \mathcal{S}$. The discrete problem has the form: Find $u_h^* = \{u_i\}_{i=1}^N \in V^h$ such that

$$a(u_h^*, v_h) = f(v_h), \quad \forall v_h \in V^h \tag{3}$$

where

$$a(u_h, v_h) = \sum_{i=1}^N a_i(u_i, v_i) = \sum_{i=1}^N \rho_i(\nabla u_i, \nabla v_i)_{L^2(\Omega_i)},$$

and $v_h = \{v_i\}_{i=1}^N \in V^h$. V^h is a Hilbert space with an inner product defined by $a(u_h, v_h)$. This problem has a unique solution and its error bound is known, see [BMP94].

Let $\{\phi_k\}$ be the set of basis functions of V^h so that $V^h = span\{\phi_k\}$. These basis functions are associated with the subdomain interior nodes (Ω_{ih}), the vertices (ν_i) and the mortar nodes ($\gamma_{m(i)h}$, $\gamma_{m(i)} \subset \partial\Omega_i$), which are not on the boundary $\partial\Omega$. The values on the nonmortar nodes are determined by the mortar condition. We use $\Pi_m(u_i, u_j)$ to denote the values on the nonmortar side $\delta_{m(j)}$, where the values of u_i on the corresponding mortar side and the values of $u_{j|\partial\delta_{m(j)}}$ are given.

For the rest of the paper we use the following notations. $x_k^{(i)}$ is the local representation of the node x_k, indicating that the node belongs to $\overline{\Omega}_i$. $\varphi_k^{(i)}$ denotes the standard nodal basis function associated with the node $x_k^{(i)}$.

The Additive Schwarz Method

In this section we introduce the additive Schwarz method for the problem (3). The method is defined using the general framework for the additive Schwarz methods, see [SBG96], i.e., in terms of a decomposition of the global space V^h into subspaces and the bilinear forms defined on these subspaces.

The decomposition of the finite element space V^h takes the form

$$V^h = V^{(-2)} + V^{(-1)} + V^{(0)} + \sum_{i=1}^N V^{(i)}, \tag{4}$$

where $V^{(i)}, i = 1, \cdots, N$, is a subspace of V^h restricted to the subdomain Ω_i with zero values on $\partial\Omega_i$ and the remaining subdomains. The subspaces $V^{(-2)}$, associated with the vertices,

and $V^{(-1)}$, associated with the mortar nodes, are defined as follows.

$$\begin{aligned} V^{(-1)} &= \left\{v \in V^h : v(x) = 0,\ x \in \cup_i(\gamma_i \cup \Omega_{ih})\right\}, \\ V^{(-2)} &= \left\{v \in V^h : v(x) = 0,\ x \in \cup_i(\nu_i \cup \Omega_{ih})\right\}. \end{aligned}$$

The sum $V^{(-2)} + V^{(-1)}$ equals the skeleton coarse space of the reformulated variant (cf. [BDR00]). Note that the basis functions on an interface have nonlocal supports on the non-mortar side, which results in a very dense coupling between the vertices and the mortar nodes in the skeleton coarse stiffness matrix. The idea of the above splitting of the skeleton coarse space is to eliminate the effect of such coupling in the algorithm, and, thereby, improving the computational complexity and the parallel property of the algorithm. The space $V^{(0)}$ is the same as the space $V^{(0)}$ of the reformulated variant. We restate its definition here, but first, some definitions and notations.

Let χ_i, associated with the subdomain Ω_i, be the piecewise linear continuous function on the triangulation of Ω_i, defined by its nodal values at $x \in \overline{\Omega}_{ih}$. For each such node x,

$$\chi_i(x) = \frac{1}{\sum_j \rho_j(x)},$$

where the sum is taken over the subdomains that x is connected to. We say that a node x_k is connected to the subdomain Ω_i if $x_k \in \overline{\Omega}_{ih}$. If the node $x_k \in \overline{\gamma}_{m(i)h}$ ($x_k \in \overline{\delta}_{m(i)h}$) then x_k is said to be connected to both Ω_i and Ω_j if $\gamma_{m(i)} = \delta_{m(j)}$ ($\delta_{m(i)} = \gamma_{m(j)}$). Note that for $\rho_i = \rho_j = 1$, χ_i is 1 at $x \in \Omega_{ih}$, $\frac{1}{2}$ at $x \in (\partial\Omega_{ih} \setminus \nu_i)$ and $\frac{1}{3}$ at $x \in \nu_i$.

We associate with each subdomain Ω_i the sets G_i and Q_i containing the indices of its neighboring subdomains defined as follows. G_i contains the index of a neighbor Ω_j if it shares an edge Γ_{ij} ($\overline{\Gamma}_{ij} = \overline{\Omega}_i \cap \overline{\Omega}_j$) with Ω_i. Q_i contains the index of a neighbor Ω_j if $\overline{\Omega}_i \cap \overline{\Omega}_j$ is a crosspoint, there is a subdomain Ω_k such that Γ_{ki} ($\overline{\Gamma}_{ki} = \overline{\Omega}_k \cap \overline{\Omega}_i$) and Γ_{jk} ($\overline{\Gamma}_{jk} = \overline{\Omega}_j \cap \overline{\Omega}_k$) are the two edges of Ω_k which intersect at that crosspoint, and Γ_{ki} is a mortar in Ω_k, cf. Fig. 1(c).

We are now ready to define the coarse space $V^{(0)}$ which is given as the span of its basis functions, $\Phi_i, i = 1, \cdots, N$, i.e.,

$$V^{(0)} = span\left\{\Phi_i : i = 1, \cdots, N\right\}. \tag{5}$$

Each function Φ_i, associated with the subdomain Ω_i, is a function in the finite element space V^h.

For an interior subdomain Ω_i ($\partial\Omega_i \cap \partial\Omega = \emptyset$), the function Φ_i is constructed in three steps. We define Φ_i first (*i*) on $\overline{\Omega}_i$, then (*ii*) on $\overline{\Omega}_j$ for $i \in G_j$, and then (*iii*) on $\overline{\Omega}_j$ for $i \in Q_j$.

(*i*) Φ_i on $\overline{\Omega}_i$ is given as

$$\Phi_i(x) = \begin{cases} 1, & x \in \Omega_{ih}, \\ \rho_i\chi_i(x), & x \in \gamma_{m(i)h} \cup \nu_i, \\ \rho_i\Pi_m(\chi_j, \chi_i)(x), & x \in \delta_{m(i)h}, \delta_{m(i)} = \gamma_{m(j)}. \end{cases} \tag{6}$$

(*ii*) Φ_i on $\overline{\Omega}_j$, where $i \in G_j$, we have two cases to consider. For the first case, let $\Gamma_{ij} = \delta_{m(j)} = \gamma_{m(i)}$, see Fig. 1(a). Then, on $\overline{\Omega}_j$,

$$\Phi_i(x) = \begin{cases} \rho_i\Pi_m(\chi_i, \chi_j)(x), & x \in \overline{\delta}_{m(j)h},\ \delta_{m(j)} = \gamma_{m(i)}, \\ \Phi(x), & x \in \overline{\delta}_{n(j)h},\ \partial\delta_{n(j)} \cap \partial\delta_{m(j)} \neq \emptyset, \\ 0, & \text{at all other } x \text{ in } \overline{\Omega}_{jh}. \end{cases} \tag{7}$$

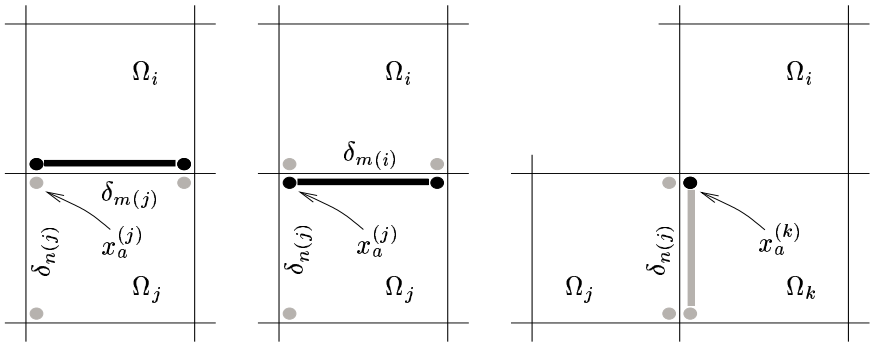


Figure 1: Illustrating Φ_i on $\overline{\Omega}_j$, where $i \in G_j$ ((a) and (b)) and $i \in Q_j$ ((c)). Here Φ_i is the basis function associated with the interior subdomain Ω_i.

For the second case, let $\Gamma_{ij} = \gamma_{m(j)} = \delta_{m(i)}$, see Fig. 1(b). Φ_i on $\overline{\Omega}_j$ is then given as

$$\Phi_i(x) = \begin{cases} \rho_i \chi_j(x), & x \in \overline{\gamma}_{m(j)h},\ \gamma_{m(j)} = \delta_{m(i)}, \\ \Phi(x), & x \in \overline{\delta}_{n(j)h},\ \partial\delta_{n(j)} \cap \partial\gamma_{m(j)} \neq \emptyset, \\ 0, & \text{at all other } x \text{ in } \overline{\Omega}_{jh}. \end{cases} \tag{8}$$

For the function $\Phi(x)$ in (7) and (8), we assume there is no vertex which is a cross point of exactly three subdomains. $\Phi(x)$ is then given as

$$\Phi(x) = \rho_i \chi_j(x_a^{(j)}) \Pi_n(0, \varphi_a^{(j)}), \tag{9}$$

where $x_a^{(j)} \in \nu_j$ (cf. figures 1a-1b).

(*iii*) Φ_i on $\overline{\Omega}_j$, where $i \in Q_j$, is given as follows. Let Γ_{ki} and Γ_{jk} be the two edges such that $\Gamma_{jk} = \delta_{n(j)} = \gamma_{n(k)}$ and $x_a^{(k)} = \partial\Gamma_{ki} \cap \partial\Gamma_{jk} \in \nu_k$, (cf. Fig. 1(c)). We have then

$$\Phi_i(x) = \begin{cases} \rho_i \chi_k(x_a^{(k)}) \Pi_n(\varphi_a^{(k)}, 0), & x \in \overline{\delta}_{n(j)h}, \\ 0, & \text{at all other } x \text{ in } \overline{\Omega}_{jh}. \end{cases} \tag{10}$$

On the remaining subdomains, $\Phi_i = 0$. This completes the definition of Φ_i for an interior subdomain Ω_i.

If Ω_i is a boundary subdomain ($\partial\Omega_i \cap \partial\Omega \neq \emptyset$) then the function Φ_i is defined as above but by imposing $\chi_j(x) = 0$ at $x \in \partial\Omega_{jh} \cap \partial\Omega_h$ for all $\Omega_j \in N_B$. The values of Φ_i on some nonmortar edges touching $\partial\Omega$ will be different, for details see [BDR00].

A somewhat similar but simpler coarse space defined in terms of discrete harmonic functions in the context of substructuring algorithms for mortar finite element problems can be found in [Dry97].

We use the exact bilinear form for all subproblems, i.e., for $i = -2, \cdots, N$ and $u, v \in V^{(i)}$, we define $b^{(i)}(.,.) : V^{(i)} \times V^{(i)} \to \mathbb{R}$ as $b^{(i)}(u,v) = a(u,v)$. The projection like operators $T^{(i)} : V^h \to V^{(i)}$ are defined in the standard way, i.e., for $i = -2, \cdots, N$ and $u \in V^h$, $T^{(i)}u \in V^{(i)}$ is the solution of

$$b^{(i)}(T^{(i)}u, v) = a(u,v),\ v \in V^{(i)}.$$

The additive Schwarz operator is then given as $T = \sum_{i=-2}^{N} T^{(i)}$, which can be written implicitly as BA, where B is the additive preconditioner. If we define $B^{(i)}$ as $T^{(i)} = B^{(i)}A$, then the action of B on a function r can be calculated as $v \leftarrow \sum_{i=-2}^{N} B^{(i)}r$. We have the following estimate for $T = BA$, the proof follows from [BDR00].

Theorem 1 *For $u \in V^h$,*

$$c_0 \frac{h}{H} a(u,u) \leq a(Tu,u) \leq c_1 a(u,u), \tag{11}$$

where both c_0 and c_1 are positive constants independent of the mesh parameters $h = \inf_i h_i$ and $H = \max_i H_i$ and the jumps of the coefficients ρ_i.

The Hybrid Schwarz Method

We introduce the hybrid method by replacing the additive preconditioner by the following hybrid preconditioner B. The action of B on r is now calculated in three steps as

$$\begin{aligned} v &\leftarrow (B^{(-2)} + B^{(-1)} + B^{(0)})r \\ v &\leftarrow v + B^{(i)}(r - Av),\ i = 1, \cdots, N \\ v &\leftarrow v + (B^{(-2)} + B^{(-1)} + B^{(0)})(r - Av). \end{aligned}$$

The last step is necessary for symmetrizing the preconditioner. Note that the subdomain solves in the second line can be done completely in parallel since we only have nonoverlapping subdomains. Basically, for this method, in each iteration, we need two extra calculations of the residual, and one extra solving of each coarse problem as compared to the additive method. The residual updates are, however, not expensive since we only need nearest neighbor communication among the subdomains (processors or virtual processors). Due to the special coarse spaces, it is very cheap to calculate the first residual update, and also, in this case, it is possible to avoid communication among the subdomains as only the values at the subdomain interior nodes are needed in the subdomain solves. The analysis of this method can be done using the general theory for Schwarz methods, see [SBG96], resulting in Theorem 1 for $T = BA$ where B is now the hybrid preconditioner.

Numerical Examples

We now present some numerical results using the Schwarz methods of this paper, as preconditioners for the conjugate gradient method. We compare the results with those of the reformulated average method introduced in [BDR00].

For simplicity, we let our model elliptic problem have zero boundary values. The force function f has the form $f(x) = 2\pi^2 \sin(\pi x_1) \sin(\pi x_2)$, and the domain is the unit square. The coefficients ρ_i are picked uniformly from the interval $[10^{-1}, 10^3]$ and then distributed randomly among the subdomains.

The test results are presented in Table 1. Each column of the table corresponds to a method, showing the iteration counts and the condition number estimates (in parentheses) for different partitions of the domain. The ratio $\frac{H}{h}$ remains fixed in all tests.

Subdomains	Additive method		Hybrid method
	Reform. variant	Modified reform.	
4×4	28 (13.36)	31 (16.07)	15 (4.18)
8×8	32 (13.75)	35 (16.19)	17 (4.21)

Table 1: *The number of iterations required to reduce the residual norm by* 10^{-6} *and a condition number estimate for each test.*

The additive Schwarz method of this paper ("Modified reform.") shows condition number estimates (iteration counts) which are close to those of the original reformulated variant ("Reform. variant"). The former method, however, needs less computation per iteration than the latter one. This is due to the splitting of the skeleton coarse space, which, in addition, makes the modified variant simpler and more suitable for parallel computation.

In the third column, we see a very substantial reduction in the condition number for the hybrid method. Thus, the hybrid method needs approximately half the number of iterations compared with the additive methods, but this is partially offset by more computation per iteration. So far, we have not made any comparison between these two methods considering a more detailed model of their computational complexity and parallel performance, this remains to be checked. The results show that the methods are all insensitive to jumps of the coefficient ρ_i across the subdomain boundaries.

We believe that this work extends and complements the work in [BDR00] and that a detailed computational study as well as experiments with realistic applications should follow in the future.

Acknowledgement

The second author was supported in part by the National Science Foundation under the Grant NSF-CCR-9732208 and in part by the Polish Science Foundation under the Grant 2 P03A 021 16.

References

[AMW99]Yves Achdou, Yvon Maday, and Olof B. Widlund. Iterative substructuring preconditioners for mortar element methods in two dimensions. *SIAM J. Numer. Anal.*, 36(2):551–580, 1999.

[BDR00]Petter E. Bjørstad, Maksymilian Dryja, and Talal Rahman. Additive average Schwarz methods for elliptic mortar finite element problems. Reports in Informatics 197, Institute for Informatics, University of Bergen, 2000. Submitted to Numer. Math.

[BDW99]Dietrich Braess, Wolfgang Dahmen, and Christian Wieners. A multigrid algorithm for the mortar finite element method. *SIAM J. Numer. Anal.*, 37:48–69, 1999.

[BMP94]Christine Bernardi, Yvon Maday, and Anthony T. Patera. A new non conforming approach to domain decomposition: The mortar element method. In Haim Brezis and Jacques-Louis Lions, editors, *Collège de France Seminar*. Pitman, 1994. This paper appeared as a technical report about five years earlier.

[CDS99]X.-C. Cai, M. Dryja, and M. V. Sarkis. Overlapping nonmatching grid mortar element methods for elliptic problems. *SIAM J. Numer. Anal.*, 36:581–606, 1999.

[Cia78]Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[CW96]Mario A. Casarin and Olof B. Widlund. A hierarchical preconditioner for the mortar finite element method. *ETNA*, 4:75–88, June 1996.

[Dry96]Maksymilian Dryja. Additive Schwarz methods for elliptic mortar finite element problems. In K. Malanowski, Z. Nahorski, and M. Peszynska, editors, *Modeling and Optimization of Distributed Parameter Systems with Applications to Engineering*. IFIP, Chapman & Hall, London, 1996. To appear.

[Dry97]Maksymilian Dryja. An iterative substructuring method for elliptic mortar finite element problems with a new coarse space. *East-West J. Numer. Math.*, 5(2):79–98, 1997.

[GP00]Jayadeep Gopalakrishnan and Joseph E. Pasciak. Multigrid for the mortar finite element method. *SIAM J. Numer. Anal.*, 37(3):1029–1052, 2000.

[SBG96]Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[WK01]Barbara Wohlmuth and Rolf Krause. Multigrid methods based on the unconstrained product space arising from mortar finite element discretizations. *SIAM J. Numer. Anal.*, 39:192–213, 2001.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

30 Uniform Domain Decomposition for a Convection-Diffusion Problem

I. Boglaev[1]

Introduction

In this paper, for solving a singularly perturbed parabolic problem with a convection-dominated term, we present a finite difference domain decomposition algorithm based on a classical upwind difference approximation in a spatial variable and on the piecewise equidistant mesh of Shishkin-type [MOS96]. These meshes allow us to decompose a computational domain into subdomains outside boundary layers and inside them as well, and possess load balancing. This property is very important for implementation of iterative algorithms on parallel computers, since it avoids loss of efficiency due to one processor being idle. Our purpose is to construct and analyse a domain decomposition algorithm based on decomposition of boundary layers. We use a modification of the Schwarz alternating method proposed in [DDD91], in which the computational domain is partitioned into many nonoverlapping subdomains with interface Γ. Small interfacial subdomains are introduced near the interface Γ, and approximate boundary values computed on Γ are used for solving problems on the nonoverlapping subdomains. Thus, this approach may be considered as a variant of a block Gauss-Seidel iteration (or in the parallel context as a multicoloured algorithm) for the subdomains with a Dirichlet-Dirichlet coupling through the interface variables. This modification of the Schwarz method has been applied in [Bog98] for solving singularly perturbed reaction-diffusion problems.

In [Mat98], for singularly perturbed parabolic problems with convection-dominated terms, uniform convergent properties of some Schwarz-type methods based on continuous multidomain decomposition (i.e. without resort to discretization in the subdomains) have been studied. Here, we construct more accurate estimations of a contraction factor for the multidomain decomposition algorithm in a discrete form and additionally investigate this algorithm when the subdomains located inside the boundary layer.

We consider the following singularly perturbed parabolic problem:

$$\varepsilon u_{xx} + b(x,t)u_x - u_t = f(x,t,u), \quad (x,t) \in Q = \Omega \times (0,T], \tag{1}$$

$$\Omega = \{x : 0 < x < 1\}, \quad u(0,t) = u(1,t) = 0, \quad u(x,0) = u^0(x),\ x \in \Omega,$$

where ε is a positive parameter, functions $b(x,t),\ f(x,t,u)$ and $u^0(x)$ are sufficiently smooth. We assume that

$$b(x,t) \geq \beta_* = \text{const} > 0,\ \partial f/\partial u \geq 0,\ (x,t,u) \in Q \times (-\infty, +\infty).$$

Under suitable continuity and compatibility conditions on the data a unique solution $u(x,t)$ of (1) exists. For $\varepsilon \ll 1$ problem (1) is singularly perturbed and characterized by an exponential layer at $x = 0$.

[1]Institute of Fundamental Sciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand, i.boglaev@massey.ac.nz. This work was supported in part by Marsden Fund MAU809 of the Royal Society of New Zealand.

Undecomposed Algorithm

Consider an implicit two-level time difference scheme which possesses an uniform in the perturbation parameter ε convergence.

On set $\bar{Q}$ introduce a rectangular mesh $\bar{\Omega}^h \times \bar{\Omega}^\tau$, where

$$\bar{\Omega}^h = \{x_i, i = 0, 1, \dots, N, x_0 = 0, x_N = 1, h_i = x_{i+1} - x_i\},$$

$$\bar{\Omega}^\tau = \{t_k = k\tau, k = 0, 1, \dots, N_\tau, N_\tau \tau = T\}.$$

For a mesh function $U(x,t)$ we use the following classical implicit difference scheme

$$\Lambda U(x,t) - \tau^{-1}[U(x,t) - U(x,t-\tau)] = f(x,t,U),\ (x,t) \in \Omega^h \times \Omega^\tau, \tag{2}$$

$$U(0,t) = U(1,t) = 0,\ t \in \bar{\Omega}^\tau, \quad U(x,0) = u^0(x),\ x \in \bar{\Omega}^h,$$

where $\Lambda U(x,t)$ is defined by

$$\Lambda U(x,t) = \varepsilon D_+ D_- U(x,t) + b(x,t) D_+ U(x,t),$$

$D_+ D_- U(x,t)$ and $D_+ U(x,t)$ are the central and forward difference approximations to the second and first derivatives in the x-direction, respectively.

The piecewise equidistant mesh of Shishkin-type from [MOS96] is formed by dividing interval $\bar{\Omega}$ into two parts $[0,\sigma],\ [\sigma,1]$, and in each part we use a uniform grid with $N/2+1$ mesh points. The step sizes of the mesh are defined by

$$h_i = h_\varepsilon = 2\sigma N^{-1}, i = 0, 1, \dots, N/2 - 1, \tag{3}$$

$$h_i = h = 2(1-\sigma) N^{-1}, i = N/2, \dots, N-1.$$

The transition point σ from [MOS96] is determined by $\sigma = \min\{2^{-1}, 2\varepsilon\beta_*^{-1} \ln N\}$. If $\sigma = 1/2$, then N^{-1} is very small relative to ε. This is unlikely in practice, and in this case the difference scheme (2) can be analyzed using standard techniques. We therefore assume that

$$\sigma = 2\varepsilon\beta_*^{-1} \ln N, \quad h_\varepsilon = 4\varepsilon\beta_*^{-1} N^{-1} \ln N, \quad N^{-1} < h < 2N^{-1}. \tag{4}$$

We note here that the size of the boundary layer is of order $O(\varepsilon|\ln\varepsilon|)$. Thus, for $\varepsilon \le N^{-1}$, the transition from the layer to the outside region is determined by the transition point σ which is located inside the boundary layer.

Theorem 1 *Let u(x,t) be the solution to problem (1). Then the solution of the difference scheme (2) on the mesh (3), (4) converges ε-uniformly to $u(x,t)$:*

$$\max_{(x,t) \in \bar{\Omega}^h \times \bar{\Omega}^\tau} |u(x,t) - U(x,t)| \le C(N^{-1} \ln N + \tau),$$

where N is the number of mesh points in the space direction, τ is the time step-size and constant C is independent of ε, N and τ.

Domain Decomposition Algorithm

We consider decomposition of domain $\bar{\Omega}$ into M nonoverlapping (adjoining) subdomains $\bar{\Omega}_m, m = 1, \dots, M$:

$$\Omega_m = (x_{m-1}, x_m), \quad \bar{\Omega}_m \cap \bar{\Omega}_{m+1} = x_m, \quad x_0 = 0, x_M = 1.$$

Additionally, we consider $M-1$ interfacial subdomains $\omega_m, m = 1, \dots, M-1$:

$$\omega_m = (x_m^b, x_m^e), \quad \omega_{m-1} \cap \omega_m = \emptyset, \quad x_m^b < x_m < x_m^e.$$

On $\bar{\Omega}_m, m = 1, \dots, M$ and $\bar{\omega}_m, m = 1, \dots, M-1$ we introduce meshes $\bar{\Omega}_m^h$, and $\bar{\omega}_m^h$, respectively, where

$$\bar{\Omega}_m^h = \{x_{mi}, i = 0, 1, \dots, N_m, x_{m0} = x_{m-1}, x_{N_m} = x_m, h_{mi} = x_{m,i+1} - x_{mi}\}, \quad (5)$$

$$\bar{\omega}_m^h = \{X_{mi}, i = 0, 1, \dots, N_{m\omega}, X_{m0} = x_m^b, X_{N_{m\omega}} = x_m^e, H_{mi} = X_{m,i+1} - X_{mi}\},$$

and suppose that $\bar{\Omega}^h = \bigcup \bar{\Omega}_m^h$, and the mesh points in $\bar{\omega}_m^h, m = 1, \dots, M-1$ coincide with the mesh points in $\bar{\Omega}^h$.

On each time-level t_k, we shall implement n_0 iterative steps of a domain decomposition algorithm. On each iterative step, firstly, we solve problems on the nonoverlapping subdomains $\bar{\Omega}_m^h, m = 1, \dots, M$ with Dirichlet boundary conditions passed from the previous iterate. Then Dirichlet data are passed from these subdomains to the interfacial subdomains $\bar{\omega}_m^h, m = 1, \dots, M-1$, and problems on the interfacial subdomains are computed. Finally, we impose continuity for piecing the solutions on the subdomains together.

On subdomains $\bar{\Omega}_m^h, m = 1, \dots, M$, introduce mesh functions $v_m^{(n)}(x, t_k), m = 1, \dots, M$ (here the index n stands for a number of iterative steps, and $n = 1, \dots, n_0$) satisfying the following implicit difference schemes

$$\Lambda v_m^{(n)}(x, t_k) - (1/\tau)[v_m^{(n)}(x, t_k) - V(x, t_{k-1})] = f(x, t_k, v_m^{(n)}(x, t_k)), \ x \in \Omega_m^h, \quad (6)$$

$$v_m^{(n)}(x, t_k) = V^{(n-1)}(x, t_k), \ x = x_{m-1}, x_m, \quad v_1^{(n)}(0, t_k) = 0, \quad v_M^{(n)}(1, t_k) = 0.$$

On the interfacial subdomains $\bar{\omega}_m^h, m = 1, \dots, M-1$, we solve the difference problems

$$\Lambda z_m^{(n)}(x, t_k) - (1/\tau)[z_m^{(n)}(x, t_k) - V(x, t_{k-1})] = f(x, t_k, z_m^{(n)}(x, t_k)), \ x \in \omega_m^h, \quad (7)$$

$$z_m^{(n)}(x_m^b, t_k) = v_m^{(n)}(x_m^b, t_k), \quad z_m^{(n)}(x_m^e, t_k) = v_{m+1}^{(n)}(x_m^e, t_k).$$

The mesh function $V^{(n)}(x, t_k)$ is determined in the form

$$V^{(n)}(x, t_k) = \begin{cases} v_m^{(n)}(x, t_k), & x \in \overline{\Omega_m^h \setminus (\omega_{m-1}^h \cup \omega_m^h)}, m = 1, \dots, M; \\ z_m^{(n)}(x, t_k), & x \in \overline{\omega}_m^h, m = 1, \dots, M-1, \end{cases} \quad (8)$$

where we introduce the following notations

$$V(x, t_k) = V^{(n_0)}(x, t_k), \quad V^{(0)}(x, t_k) = V(x, t_{k-1}), \ k \geq 1,$$

$$V(x, 0) = u^0(x), \ x \in \bar{\Omega}^h.$$

Algorithm (6)-(8) can be carried out by parallel processing, since on each iterative step n the M problems (6) for $v_m^{(n)}(x,t_k), m = 1,\dots,M$ and the $M-1$ problems (7) for $z_m^{(n)}(x,t_k), m = 1,\dots,M-1$ can be implemented concurrently.

On a mesh $\bar{\Omega}_*^h = \{x_i, i = 0,1,\dots,N_*; x_0 = x_a, x_{N_*} = x_b\}$, consider the difference problems

$$\Lambda\Phi^s(x) - \tau^{-1}\Phi^s(x) = 0,\ x \in \Omega_*^h,\ s = 1,2, \tag{9}$$

$$\Phi^1(x_0) = 1,\ \Phi^1(x_{N_*}) = 0, \quad \Phi^2(x_0) = 0,\ \Phi^2(x_{N_*}) = 1.$$

Introduce the notations

$$q_m^b = \Phi_m^1(x_m^b) + \Phi_m^2(x_m^b),\ q_m^e = \Phi_{m+1}^1(x_m^e) + \Phi_{m+1}^2(x_m^e),\ m = 1,\dots,M-1,$$

$$q_m = q_m^b\Phi_{m,\omega}^1(x_m) + q_m^e\Phi_{m,\omega}^2(x_m),\ m = 1,\dots,M-1,$$

where $\Phi_m^{1,2}(x)$ and $\Phi_{m,\omega}^{1,2}(x)$ are the solutions to (9) on $\bar{\Omega}_m^h$ and $\bar{\omega}_m^h$, respectively.

Theorem 2 *Algorithm (6)- (8) on mesh (3), (4) converges to the solution of (1) with the following rate:*

$$\max_{(x,t)\in\bar{\Omega}^h\times\bar{\Omega}^\tau} |u(x,t) - V(x,t)| \le C((N\tau)^{-1}\ln N + T)(N^{-1}\ln N + \tau + q^{n_0}),$$

$$q = \max_{1\le m\le M-1} q_m, \tag{10}$$

where the contraction coefficient $q \in (0,1)$ and constant C is independent of ε, h, τ and q.

Theorem 2 guarantees that the domain decomposition algorithm (6)- (8) converges for any initial guesses. From Theorem 2, it follows that asymptotically one would expect to choose the number of mesh points N in the space direction such that $N \approx N_\tau$. If $N \approx N_\tau$, then we conclude the following estimate

$$\max_{(x,t)\in\bar{\Omega}^h\times\bar{\Omega}^\tau} |u(x,t) - V(x,t)| \le C(N^{-1}\ln N + \tau + q^{n_0}),$$

where constant C is independent of ε, N, τ and q.

Estimates on Rate of Convergence

The interfacial subdomains outside the boundary layer. Consider algorithm (6)- (8) with the interfacial subdomains $\omega_m^h,\ m = 1,\dots,M-1$, located outside the boundary layer. Suppose for simplicity that the centre of the discrete interval $\bar{\omega}_m^h$ is located at x_m, i.e. in (5) $x_m = I_{m\omega}h,\ N_{m\omega} = 2I_{m\omega}$. For sufficiently small values of ε, we can approximate q in (10) uniformly in ε by

$$q \approx \exp[-I_\omega \ln(1 + h(\beta_{**}\tau)^{-1})],\ \beta_{**} = \max_{(x,t)\in\bar{Q}} b(x,t).$$

We compare this estimate with the convergence rate of the Schwarz alternating method obtained in [Mat98]:

$$\max_{(x,t)}|u^{(n+1)}-u|\le\hat{q}\max_{(x,t)}|u^{(n)}-u|,\ \hat{q}=\exp(-\alpha d\tau^{-1/2}),\tag{11}$$

where $u^{(n)}$ is the Schwarz iterate, $d>0$ measures the overlap between two subdomains and $\alpha>0$ is independent of τ. Outside the boundary layer $d=O(I_\omega h)$, the contraction factor $\hat{q}$ is approximated by

$$\hat{q}\approx\exp(-\bar{\alpha}I_\omega h\tau^{-1/2}),$$

where $\bar{\alpha}>0$. From Theorem 2, one would expect to choose $\tau\approx h$, and asymptotically we get

$$q\approx\exp[-I_\omega\ln(1+\beta_{**}^{-1})],\ \hat{q}\approx\exp(-\bar{\alpha}I_\omega\tau^{1/2})\to 1,\tau\to 0,\ I_\omega\le N(2M)^{-1}.$$

It follows that the estimate of the convergence rate from [Mat98] is impractical.

The interfacial subdomains inside the boundary layer. Suppose that N is divisible by $2M$ and M is even, we decompose the boundary layer $[0,\sigma]$ and the region outside the layer $[\sigma,1]$ into $M/2$ equal subdomains, respectively, where σ from (4). We note that each subdomain $\bar{\Omega}_m^h$ contains the same number of mesh points $2I+1, I=N/(2M)$. From (5), we have

$$\begin{aligned}\bar{\Omega}_m^h&=\{x_{mi},x_{mi}=x_{m-1}+ih_\varepsilon,i=0,1,\dots,2I\},\\ x_{m-1}&=2(m-1)Ih_\varepsilon,\ m=1,\dots,M/2,\\ \bar{\Omega}_m^h&=\{x_{mi},x_{mi}=x_{m-1}+ih,i=0,1,\dots,2I\},\\ x_{m-1}&=\sigma+2(m-M/2-1)Ih,m=M/2+1,\dots,M,\end{aligned}\tag{12}$$

where h, h_ε are the uniform step sizes outside and inside the boundary layer. We choose the interfacial subdomains in the following forms:

$$\begin{aligned}\bar{\omega}_m^h&=\{X_{mi},X_{mi}=x_m^b+ih_\varepsilon,i=0,1,\dots,2I_\omega\},\\ x_m^b&=x_m-I_\omega h_\varepsilon,\ m=1,\dots,M/2-1,\\ \bar{\omega}_{M/2}^h&=\{X_{M/2,i},X_{M/2,i}=x_{M/2}^b+ih_\varepsilon,i=0,1,\dots,I_\omega;\\ &\quad X_{M/2,i}=\sigma+ih,i=I_\omega+1,\dots,2I_\omega\},\\ x_{M/2}^b&=\sigma-I_\omega h_\varepsilon,\\ \bar{\omega}_m^h&=\{X_{mi},X_{mi}=x_m^b+ih,i=0,1,\dots,2I_\omega\},\\ x_m^b&=x_m-I_\omega h,\ m=M/2+1,\dots,M-1.\end{aligned}$$

Here the interfacial subdomains $\bar{\omega}_m^h, m=1,\dots,M-1$ contain the same number of mesh points $2I_\omega+1$, and the centre of the discrete interval $\bar{\omega}_m^h$ is located at x_m. We suppose $1\le I_\omega\le I$, such that $\omega_{m-1}^h\cap\omega_m^h=\emptyset, m=2,\dots,M-1$. On this domain decomposition, we can approximate the contraction factor q in (10) by

$$q\approx\varepsilon\tau h^{-2}+[(2+\sqrt{2})/2]^{-\frac{N}{2M}}+\exp[-I_\omega\ln(1+\beta_{**}^{-1})],\ I_\omega\le N(2M)^{-1}.$$

If in (11) $d=O(I_\omega h_\varepsilon)$, then $\hat{q}$ is approximated by $\hat{q}\approx\exp(-\bar{\alpha}I_\omega h_\varepsilon\tau^{-1/2})$. In the case of the maximal size of the interfacial subdomains $I_\omega=N(2M)^{-1}$, we get

$$\hat{q}\approx\exp(-\bar{\alpha}\varepsilon\ln N(M\sqrt{\tau})^{-1}).$$

Again, we conclude that the estimate of the convergence rate from (11) is impractical for the proposed domain decomposition.

Numerical Results

As a test problem, consider the following problem

$$\varepsilon u_{xx} + u_x - u_t = 0,\ (x,t) \in (0,1) \times (0,T],$$

$$u(0,t) = 1,\ u(1,t) = 0,\ u(x,0) = 0.$$

with $b(x,t) = 1$. Note that in the new variable $\tilde{u}(x,t) = u(x,t) + (x-1)$, this problem becomes (1) with $f(x,t,u) = 1$ and $u^0(x) = x - 1$.

On each time-level, we implement n_0 iterates of algorithm (6)-(8) to satisfy the stopping criterion

$$\max_{x \in \bar{\Omega}^h} |V^{(n_0)}(x,t_k) - U(x,t_k)| \le \delta,\ \delta = \max(N^{-1} \ln N, \tau),$$

where $U(x,t_k)$ is the solution of the undecomposed algorithm (2) at time-level t_k.

Consider the domain decomposition (12) with the interfacial subdomains inside the boundary layer. In Table 1, for $\tau = 10^{-2}, 5 \cdot 10^{-3}, 10^{-3}$ and various values of ε, M, we give the average (over ten time-levels) number of iterations n_0 with $N = 64$ and the maximal size of the interfacial subdomains $I_\omega = N(2M)^{-1}$. From the data, it follows that for M fixed, n_0 is a monotone increasing function with respect to the time mesh spacing τ, and for $\varepsilon \le 10^{-3}$, n_0 is independent of the perturbation parameter. We notice that the number of iterations approaches 1 as $\tau \to 0$. These results substantiate the theoretical convergent estimates.

M	n_0		
2	2; 2; 2	1.4; 1.4; 1	1; 1; 1
4	2; 2; 2	1.4; 1.4; 1	1; 1; 1
8	2; 2; 2	1.4; 1.4; 1	1; 1; 1
16	2.4; 2; 2	1.4; 1.4; 1	1; 1; 1
32	8.2; 5; 2	1.4; 1.4; 1	1; 1; 1
ε	0.1	0.01	0.001

Table 1: Average number of iterations n_0 for $N = 64$, $\tau = 10^{-2}, 5 \cdot 10^{-3}, 10^{-3}$.

M	n_0				
2	2	2	2	2	2
4	2	2	2	2	2
8	7.2	3.7	2.5	2	2
16	11.2	5.6	3.8	3	3
I_ω	1	2	3	4	$N(2M)^{-1}$

Table 2: Average numbers of iterations n_0 for $N = 128, \tau = 10^{-2}, \varepsilon = 10^{-1}$.

In Table 2, for various numbers M and sizes I_ω of the interfacial subdomains, we represent the average number of iterations with $N = 128, \tau = 10^{-2}, \varepsilon = 10^{-1}$. Note that the last column in the table corresponds to the interfacial subdomains with the maximal size. The average number of iterations as a function of the size of the interfacial subdomains is a monotone

decreasing function, and this is in agreement with our theoretical estimates. Another notable feature is that this function varies very quickly for small values of I_ω, and relatively small sizes of the interfacial subdomains are needed to essentially reduce the number of iterations.

Conclusion

We summarise our discussion concerning the theoretical results and numerical experiments.

1. We emphasise here the domain decomposition algorithm (6)-(8) on the piecewise uniform mesh (3), (4) possesses uniform in the perturbation parameter convergence. Thus, the proposed algorithm keeps the main property of the most effective undecomposed algorithms for singular perturbation problems.

2. In the context of parallel computing, the proposed uniform decomposition (12) guarantees us load balancing of a multiprocessor computer.

3. The numerical experiments confirm effectiveness of the proposed domain decomposition algorithm. Algorithm (6)-(8) requires few iterations on each time-level and sufficiently small sizes of the interfacial subdomains and still maintains stable approximation.

References

[Bog98]Igor Boglaev. On a domain decomposition algorithm for a singularly perturbed reaction-diffusion problem. *J. of Comput. Appl. Math.*, 98, 1998.

[DDD91]Clint N. Dawson, Qiang Du, and Todd F. Dupont. A finite difference domain decomposition algorithm for numerical solution of the heat equation. *Math. Comput.*, 57:63–71, 1991.

[Mat98]Tarek P. Mathew. Uniform convergence of the schwarz alternating method for solving singularly perturbed advection-diffusion equations. *SIAM J. Numer. Anal.*, 35:1663–1683, 1998.

[MOS96]John J.H. Miller, Eugene O'Riordan, and Grigorii I. Shishkin. *Fitted numerical methods for singular perturbation problems.* World Scientific, Singapore, 1996.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

31 Domain decomposition methods for solving scattering problems by a boundary element method

Y. Boubendir[1], A. Bendali[2]

Introduction

Integral equation methods are widely used for the numerical solution of scattering problems. Among their advantages, we mention direct and simple dealing with the radiation condition, accuracy and reduction of the mesh only to the boundary. As a counterpart, this method generates large dense complex matrices and in the case of dielectric layers may need some extra auxiliary unknowns, namely the equivalent magnetic currents. Also, the repetition of some geometrical patterns can drastically increase the size of the final system to be solved in an artificial way. The aim of this paper is to show how these difficulties can be overcome by a suitable use of a nonoverlapping domain decomposition method while however keeping the advantages of the boundary integral equations solutions.

The main technique used to decompose the solution domain into smaller domains consists in expressing the usual matching of the Cauchy data of the problem (the equivalent currents as they are generally refered to in computational electromagnetics) in terms of some equivalent boundary conditions of impedance (also called Robin) type.

The method also applies to a conductor covered by a dielectric layer with now two advantages. First, at each step, the problem to be solved has for unknown the electric current only whereas the direct solution also involves the magnetic current as a supplemental unknown. Moreover, at each step, unknown interior and exterior currents are completely uncoupled.

Another interesting aspect of this method is to couple a finite element and a boundary element method. This approach has been investigated by several authors (e.g., [JN80], [Cos87], [dLB95], [Lan94], [HW92]). However, the resulting final system is generally large and difficult to solve because it involves equations coming both from the FEM and BEM formulations. On the contrary, the method proposed in this paper uncouples completely the two solution procedures.

[1]UMR MIP INSA-CNRS-UPS, Cerfacs, France, boubendi@cerfacs.fr
[2]UMR MIP INSA-CNRS-UPS, Cerfacs, France, bendali@gmm.insa-tlse.fr

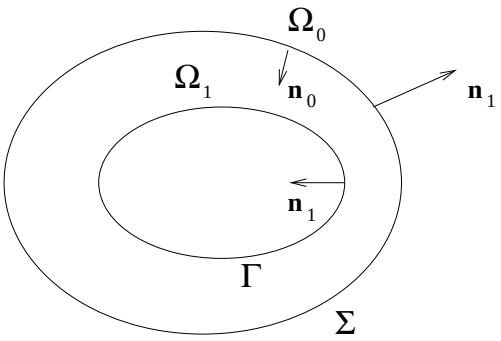


Figure 1: A typical geometry

Nonoverlapping domain decomposition method

To be specific, we consider the following problem related to the scattering of an TE wave by a coated perfectly conducting cylinder

$$\left\{\begin{array}{l} \text{find a sufficiently smooth } u \text{ such that} \\ \nabla\cdot(\frac{1}{\varepsilon}\nabla u)+k^2\frac{n^2}{\varepsilon}u=0 \quad \text{in } \Omega_1, \\ \Delta u+k^2u=0 \quad \text{in } \Omega_0, \\ \partial_{\mathbf{n_1}}u=0 \quad \text{on } \Gamma, \\ u_1=u_0, \quad \varepsilon^{-1}\partial_{\mathbf{n_1}}u_1=\partial_{\mathbf{n_1}}u_0 \quad \text{on } \Sigma, \\ \lim\limits_{|x|\to+\infty}|x|^{1/2}\left(\nabla(u-u^{\rm inc})\cdot\dfrac{x}{|x|}-ik(u-u^{\rm inc})\right)=0, \end{array}\right. \tag{1}$$

where $\mathbf{n_1}$ and $\mathbf{n_0}$ are respectively the unit normal to Σ outwardly directed to Ω_1 and to Ω_0 (fig. 1), k is the wave number, n and ε, respectively the index and the relative permittivity of the dielectric medium filling Ω_1. Superscript 1 and 0 indicate respective limits on Σ from within Ω_1 and Ω_0.

To uncouple the exterior problem solution in Ω_0 and the interior one in Ω_1, we use the methods initiated by P.-L. Lions [Lio90] and later developed for wave propagation problems by B. Després [Dep91] to write the transmission conditions on Σ in the following equivalent form

$$\left\{\begin{array}{l} \varepsilon^{-1}\partial_{\mathbf{n}_1}u_1+\eta Lu_1=-\partial_{\mathbf{n}_0}u_0+\eta Lu_0 \quad \text{on } \Sigma, \\ \partial_{\mathbf{n}_0}u_0+\eta Lu_0=-\varepsilon^{-1}\partial_{\mathbf{n}_1}u_1+\eta Lu_1 \quad \text{on } \Sigma. \end{array}\right. \tag{2}$$

where L is positive self-adjoint inversible operator, $\eta=-ik(\mathcal{R}+i\mathcal{X})$ with $\mathcal{R}>0$ and $\mathcal{X}\geq 0$. Therefore, the computation of the solution consists in solving the following two problems separately at each step n

$$\left\{\begin{array}{l} \nabla\cdot(\frac{1}{\varepsilon}\nabla u_1^{(n+1)})+k^2\frac{n^2}{\varepsilon}u_1^{(n+1)}=0 \quad \text{in } \Omega_1, \\ \partial_{\mathbf{n}_1}u_1^{(n+1)}=0 \quad \text{on } \Gamma, \end{array}\right. \tag{3a}$$

$$\frac{1}{\varepsilon}\partial_{\mathbf{n}_1}u_1^{(n+1)}+\eta Lu_1^{(n+1)}=-\partial_{\mathbf{n}_0}u_0^{(n)}+\eta Lu_0^{(n)} \quad \text{on } \Sigma, \tag{3b}$$

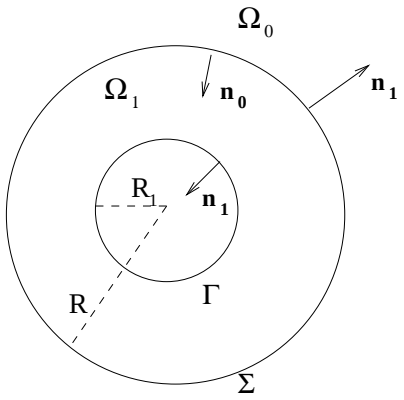


Figure 2: A circular geometry

$$\begin{cases} \Delta u_0^{(n+1)} + k^2 u_0^{(n+1)} = 0 \quad \text{in } \Omega_0, \\ \lim\limits_{|x|\to+\infty} |x|^{1/2} \left(\nabla(u_0^{(n+1)} - u^{\text{inc}}) \cdot \dfrac{x}{|x|} - ik(u_0^{(n+1)} - u^{\text{inc}}) \right) = 0, \end{cases} \tag{4a}$$

$$\partial_{\mathbf{n}_0} u_0^{(n+1)} + \eta L u_0^{(n+1)} = -\frac{1}{\varepsilon} \partial_{\mathbf{n}_1} u_1^{(n)} + \eta L u_0^{(n)} \quad \text{on } \Sigma. \tag{4b}$$

It is well-known (e.g., [CZ92]) that both problems (1), (3) and (4) are well-posed in an appropriate functional setting. Observe that the direct solution of problem (1) requires the determination of the following coupled Cauchy data $\lambda_\Sigma = u_0 = u_1$ on Σ, $p_\Sigma = \varepsilon^{-1}\partial_{\mathbf{n}_1} u_1 = \partial_{\mathbf{n}_1} u_0$ on Σ and $\lambda_\Gamma = u_1$ on Γ (e.g., [BS94]), whereas, the solution of problem (3) requires the determination of $\lambda_1^{(n+1)}$ and $\lambda_\Gamma^{(n+1)}$ and that of problem (4), the determination of $\lambda_0^{(n+1)}$ only.

Convergence of the domain decomposition method

For $\Re(\eta) = 0$, i.e. $\mathcal{X} = 0$, the theoretical convergence of the algorithm (3) and (4) is well known [Dep91], [CGJ00]. However, plots of the residual in figure 3 clearly indicates that the discrete version of the algorithm converges for $\mathcal{X} > 0$ only. It seems that only variational schemes like finite element methods can keep the convergence properties of the algorithm at the discrete level (e.g., [Dep91], [dLBFM$^+$98]). Boundary element method is not based on such a principle and thus results in a non convergent scheme for $\mathcal{X} = 0$.

For $\mathcal{X} > 0$, the proof of convergence seems to be out of reach for the general case. This is probably due to a lack of a suitable way to handle propagative and evanescent parts of the solution separately. However, for all cases when a decomposition of the solution in propagative and evanescent modes can be done, we are able to prove that the algorithm with $\mathcal{X} > 0$ has a better behaviour than with $\mathcal{X} = 0$. The following example rather strikingly illustrates this claim.

For a circular geometry (fig. 2) with $\Omega_0 = \{x \in \mathbb{R}^2; \quad |x| > R\}$, $\Omega_1 = \{x \in \mathbb{R}^2; \quad R_1 < |x| < R\}$, we can decompose the error in modes from a Fourier-Hankel series

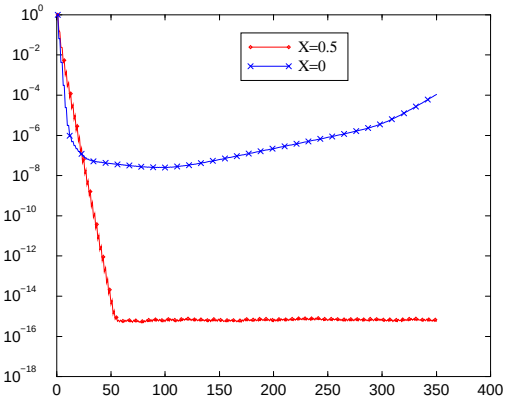


Figure 3: Behaviour of the residuals

expansion and analyze separately the convergence of the propagative and evanescent parts of the wave. Setting

$$u_0(r,\theta) = \sum_{m=-\infty}^{+\infty} u_0^{(m)}(r)e^{im\theta}, \qquad u_1(r,\theta) = \sum_{m=-\infty}^{+\infty} u_1^{(m)}(r)e^{im\theta}, \tag{5}$$

problems (3), (4) are reduced to the following one-dimensional problems

$$\begin{cases} \dfrac{1}{r}\partial_r(r\partial_r u_0^{(m)}) - \dfrac{m^2}{r^2}u_0^{(m)} + k^2 u_0^{(m)} = 0, \quad r > R, \\ \lim\limits_{r\to+\infty} r^{1/2}\left(\partial_r u_0^{(m)} - iku_0^{(m)}\right) = 0, \end{cases} \tag{6a}$$

$$-\partial_r u_0^{(m)} + \eta L_m u_0^{(m)} = g_0^{(m)}, \quad r = R, \tag{6b}$$

$$\begin{cases} \dfrac{1}{r}\partial_r(r\partial_r u_1^{(m)}) - \dfrac{m^2}{r^2}u_1^{(m)} + k^2 u_1^{(m)} = 0, \quad R_1 < r < R, \\ -\partial_r u_1^{(m)} = 0, \quad r = R_1, \end{cases} \tag{7a}$$

$$\varepsilon^{-1}\partial_r u_1^{(m)} + \eta L_m u_1^{(m)} = g_1^{(m)}, \quad r = R. \tag{7b}$$

We have assumed that the operator L is diagonal relatively to the Fourier series expansion. Solutions to problems (6), (7) are respectively obtained by $u_0^{(m)} = \alpha_m H_m^{(1)}(kr)$ and $u_1^{(m)} = \beta_m N_m(knr)$ where $H_m^{(1)}$ represents the Hankel function of the first kind and $N_m(knr)$ is a solution of the Bessel equation of order m which can be expressed by a linear combination of the Bessel J_m and Neumann Y_m functions of order m such that $N_m^{'}(knR_1) = 0$. The iteration operator is characterized by the matrices

$$\mathcal{S} = \begin{pmatrix} 0 & \mathcal{S}_m^{(0)} \\ \mathcal{S}_m^{(1)} & 0 \end{pmatrix}, \tag{8}$$

where $\mathcal{S}_m^{(0)}$ et $\mathcal{S}_m^{(1)}$ are defined by

$$\mathcal{S}_m^{(0)} g_0^{(m)} = -1 + 2\eta L_m u_0^{(m)}(R), \qquad \mathcal{S}_m^{(1)} g_1^{(m)} = -1 + 2\eta L_m u_1^{(m)}(R). \tag{9}$$

First, we give a criterion characterizing the convergence of the algorithm.

Theorem 1 *The domain decomposition algorithm converges if and only if for all* m $\rho(\mathcal{S}_m) < 1$, $\rho(\mathcal{S}_m)$ *being the spectral radius of matrix* S_m.

Proof Let $g = (g_0, g_1)^T$. One possible definition of the norm in $H^{-s}(\Sigma) \times H^{-s}(\Sigma)$ is given by $\|g\|_{-s}^2 = \sum_{-\infty}^{+\infty} (1+m^2)^{-s} \left|g^{(m)}\right|^2$, where $g^{(m)}$ is defined by $g = \sum_{-\infty}^{+\infty} g^{(m)} e^{im\theta}$ and $g^{(m)} = (g_0^{(m)}, g_1^{(m)})^T$. The convergence of the method will be established if we can show that $\lim_{n\to+\infty} \|\mathcal{S}^n g\|_{-s} = 0$ with

$$\|\mathcal{S}^n g\|_{-s}^2 = \sum_{-\infty}^{+\infty} (1+m^2)^{-s} \left|(\mathcal{S}_m)^n g^{(m)}\right|^2. \tag{10}$$

If it exists m_0 such that $\rho(\mathcal{S}_{m_0}) > 1$, clearly the method does not converge. So, we can restrict the discussion to the case where $\rho(\mathcal{S}_m) < 1$ for all m. The matrix $\mathcal{S}_m$ has two distinct eigenvalues $\lambda_m = \pm\sqrt{\mathcal{S}_m^{(0)}\mathcal{S}_m^{(1)}}$. So, it can be put in a diagonal form by $\mathcal{S}_m = P_m D_m P_m^{-1}$, D_m being a diagonal matrix. Therefore, we obtain

$$\left\|\mathcal{S}_m^n g^{(m)}\right\| \leq \|P_m\| \, \|P_m^{-1}\| \, \|g^{(m)}\| \, \left(\rho(\mathcal{S}_m)\right)^n.$$

The most important point in the proof is that the condition number $\|P_m\| \, \|P_m^{-1}\|$ of matrix S_m remains uniformly bounded. Elementary arguments then permit to end the proof. ■

The previous characterization establishes that the method converges if $|\mathcal{S}_m^{(0)}\mathcal{S}_m^{(1)}| < 1$ for each m to obtain the convergence of the method. Solving problems (6), (7), we get

$$\mathcal{S}_m^{(0)} = \frac{-\mathcal{Z}_0 + i(\mathcal{R} + i\mathcal{X})}{\mathcal{Z}_0 + i(\mathcal{R} + i\mathcal{X})}, \quad \mathcal{S}_m^{(1)} = \frac{-\mathcal{Z}_1 + i(\mathcal{R} + i\mathcal{X})}{\mathcal{Z}_1 + i(\mathcal{R} + i\mathcal{X})}, \tag{11}$$

where $\mathcal{Z}_0 = H_m^{(1)'}(kR)/L_m H_m^{(1)}(kR)$, $\quad \mathcal{Z}_1 = -N_m'(knR)/L_m N_m(knR)$, which are well defined because both the two problems are well posed.

Proposition 1

- *For both evanescent and propagative modes,* $|\mathcal{S}_m^{(0)}| < 1$.
- *If* m *corresponds to an evanescent mode, that is,* $m \geq m_0$, *for* m_0 *large enough, then* $|\mathcal{S}_m^{(1)}| < 1$.

Proof Let $\mathcal{Z}_0 = -x_m + iy_m$. Clearly, it is enough to show that x_m and y_m are both > 0 to prove that $|\mathcal{S}_m^{(0)}| < 1$. Signs of x_m and y_m are respectively that of $\Im(H_m^{(1)'}(kR)\overline{H_m^{(1)}(kR)})$ and $-\Re(H_m^{(1)'}(kR)\overline{H_m^{(1)}(kR)})$. From [CK92], it is well-known that $\Im(H_m^{(1)'}(kR)\overline{H_m^{(1)}(kR)})$ is equal to the Wronskian $W(J_m(kR), Y_m(kR)) = 2/\pi kR$. Since $kR > 0$, from [CK92] we get that $y_m > 0$. The property $x_m > 0$ uses a more difficult argument. First, we remark that

$\Re(H_m^{(1)'}(t)\overline{H_m^{(1)}(t)})_{t=kR} = \frac{1}{2}(|H_m^{(1)}(t)|^2)'_{t=kR}$. Using Nicholson's formula [Wat22], we get that function $|H_m^{(1)}(kR)|^2$ is a strictly decreasing function, so the quantity $\Re(H_m^{(1)'}(kR)\overline{H_m^{(1)}(kR)})$ is negative and then $x_m > 0$. We conclude that for all $\mathcal{R} > 0$ and $\mathcal{X} > 0$, $|\mathcal{S}_m^{(0)}| < 1$.

For the problem in the bounded domain, the previous sign determination can be more easily obtained from coerciveness estimates. Let $\mathcal{Z}_1 = -x_m + iy_m$. The variational formulation of problem (7) gives

$$\Re\left(Ru_1^{(m)'}(R)\overline{u_1^{(m)}}(R)\right) = \int_{R_1}^{R}\left\{r|u_1^{(m)'}|^2 + (\frac{m^2}{r^2} - k^2n^2)|u_1^{(m)}|^2 r\right\}dr,$$

and then if m is large enough, using coerciveness property, we get that

$$\Re(Ru_1^{(m)'}(R)\overline{u_1^{(m)}}(R)) > 0. \tag{12}$$

Definition of $u_1^{(m)}$ and $\mathcal{Z}_1$ yields $x_m > 0$. Since we have considered that the material filling Ω_1 is without losses ($\Im(\eta) = 0$) and perfectly reflecting boundary condition on Γ, we are led the most severe case $y_m = 0$. Indeed, in this case

$$\mathcal{S}_m^{(1)} = \frac{(\mathcal{X} - x_m) + i\mathcal{R}}{(\mathcal{X} + x_m) + i\mathcal{R}}.$$

For $\mathcal{X} = 0$ (Despré's algorithm [Dep91]) $|\mathcal{S}_m^{(1)}| = 1$ and so $|\mathcal{S}_m^{(0)}\mathcal{S}_m^{(1)}| < 1$. The algorithm converges as expected from the study for the general case [Dep91]. Observe however that parameter $\mathcal{S}_m^{(1)}$ has no influence upon the convergence of the algorithm and $\mathcal{S}_m^{(0)}$ gives a less effective damping of the evanescent modes. The interesting point is that taking $\mathcal{X} > 0$ also gives $|\mathcal{S}_m^{(1)}| < 1$ for all m except a finite number generally corresponding to propagative modes. But since for $\mathcal{X} = 0$ $|\mathcal{S}_m^{(0)}\mathcal{S}_m^{(1)}| < 1$, it is sufficient to tune $\mathcal{X}$ for each of these exceptional mode to obtain a maximal value for $\mathcal{X}$ insuring the convergence of the algorithm. ■

Numerical results

At each step, problem (4) in the unbounded domain Ω_0 has been solved by a boundary element method [BBC00] and problem (3) in the bounded domain Ω_1 by an usual nodal finite element method. The exterior problem in Ω_0 is solved by a BEM following the approach introduced in [Ver99]. The solution is represented as a superposition of a single- and a double-layer potentials

$$u_0(x) = u^{\text{inc}} + \int_\Sigma G(x,y)p(y)d\Sigma(y) - \int_\Sigma \partial_{\mathbf{n}_y}G(x,y)\lambda(y)d\Sigma(y), \tag{13}$$

where the unknown densities p and λ are linked by the following relation induced by the impedance condition

$$p + \eta L = 0. \tag{14}$$

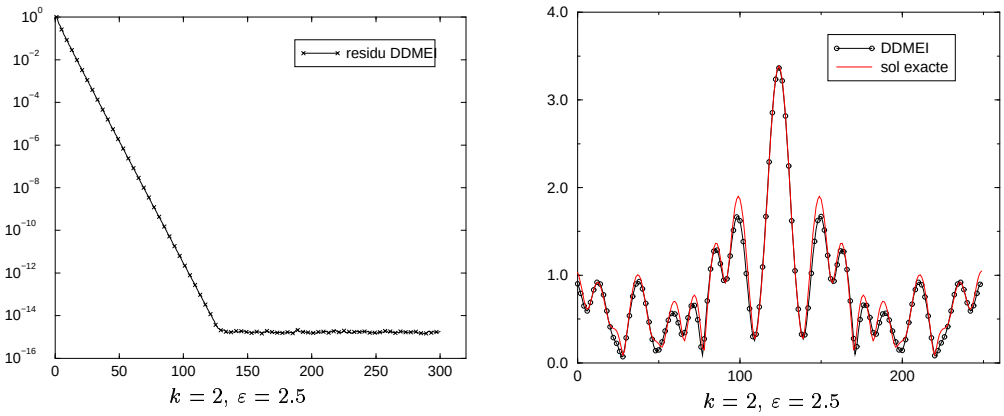


Figure 4: Coupling FEM and BEM

The boundary condition can then be expressed variationnally as

$$\int_\Sigma \left(\partial_{\mathbf{n}_0} u_0 \lambda' - u_0 p'\right) d\Sigma = \int_\Sigma g_0 \lambda' d\Sigma, \tag{15}$$

with p' and λ' are linked by the same relation that p and λ. Formulating these constrains through a Lagrange multiplier, both the latter and the magnetic currents p and p' can be eliminated at the element level when all the unknowns are approximated by a $\mathbb{P}_1$-continuous BEM, (see [Ver99] for more details).

Plots in figure 4 give the residual and comparison between exact and computed electric current on Σ. The incident wave is a plane wave propagating along the x-axis.

The interesting point is that now, with $\mathcal{X} > 0$, the discrete algorithm converges using either a nodal finite element or a boundary element method.

References

[BBC00]Y. Boubendir, A. Bendali, and F. Collino. Domain decomposition methods and integral equations for solving helmholtz diffraction problem. In *Fifth International Conference on Mathematical and Numerical Aspects of Wave Propagation*, pages 760–764, Philadelphia, july 2000. SIAM.

[BS94]A. Bendali and M. Souilah. Consistency estimates for a double-layer potential and application to the numerical analysis of the boundary-element approximation of acoustic scattering by a penetrable object. *Mathematics of computation*, 62(205):65–91, 1994.

[CGJ00]F. Collino, S. Ghanemi, and P. Joly. Domain decomposition method for harmonic wave propagation: a general presentation. *Computer methods in applied mechanics and engineering*, 184:171–211, 2000.

[CK92]D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*, volume 93. Springer-Verlag, 1992.

[Cos87]M Costabel. Symmetric methods for the coupling of finite elements and boundary elements. In *Boundary Elements IV, Vol.1, Comput. Mech.*, pages 441–420. Brebier, Southampton, 1987.

[CZ92]G. Chen and J. Zhou. *Boundary element methods*. Academic Press London, 1992.

[Dep91]B. Deprés. *Méthodes de décomposition de domains pour les problèms de propagation d'ondes en régime harmonique*. PhD thesis, Université Paris IX Dauphine, 1991.

[dLB95]Armel de La Bourdonnaye. Some formulations coupling finite element and integral equation methods for helmholtz equation and electromagnetism. *Numerische Mathematik*, 69(3):257–268, 1995.

[dLBFM+98]Armel de La Bourdonnaye, Charbel Farhat, Antonini Macedo, Frédéric Magoulès, and François-Xavier Roux. A non-overlapping domain decomposition method for exterior Helmholtz problems. In *Domain decomposition methods, 10 (Boulder, CO, 1997)*, pages 42–66, Providence, RI, 1998. Amer. Math. Soc.

[HW92]G. C. Hsiao and W. L. Wendland. Domain decomposition via boundary element methods. Technical Report 92-14, Mathematics Institut A, Universität Stuttgart, 1992.

[JN80]C. Johnson and J. Nedelec. On the coupling of boundary integral and finite element methods. *Math. Comp.*, 35:1063–1079, 1980.

[Lan94]Ulrich Langer. Parallel iterative solution of symmetric coupled BE/FE–equations via domain decomposition. *Contemp. Math.*, 157:335–344, 1994.

[Lio90]Pierre-Louis Lions. On the Schwarz alternating method. III: a variant for nonoverlapping subdomains. In Tony F. Chan, Roland Glowinski, Jacques Périaux, and Olof Widlund, editors, *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations , held in Houston, Texas, March 20-22, 1989*, Philadelphia, PA, 1990. SIAM.

[Ver99]Laurent Vernhet. Boundary element solution of a scattering problem involving a generalized impedance boundary condition. *Math. Meth. Appl.*, 22(7):587–603, 1999.

[Wat22]G. N. Watson. *A treatise on the theory of Bessel functions*. Cambridge University Press, 1922.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

32 On the use of iterative Schwarz algorithms in the solution of an optimal control problem

A. Bounaïm[1]

Introduction

We present two methods for solving an optimal control problem governed by a partial differential equation. Our methods combine optimal control techniques and Schwarz algorithms using an overlapping domain decomposition at each step of the minimization process. We design parallel algorithms based on the iterative Schwarz methods used either as solver or as preconditioner. Numerical results are presented to show the behavior of the optimization solver with respect to some parameters related to domain decomposition.

As a model problem, we consider a boundary control problem of which the state variable is the solution of an elliptic partial differential equation:

$$\begin{cases} -\Delta y(v) &= f & \text{in} & \Omega, \\ y(v) &= 0 & \text{on} & \Gamma_N \cup \Gamma_S, \\ \dfrac{\partial y(v)}{\partial n} &= v & \text{on} & \Gamma_E \cup \Gamma_W. \end{cases} \tag{1}$$

The control v is taken on the east and west boundaries of a rectangular 2D domain Ω whereas the observations y_d are distributed over the whole domain Ω. The solution of such a problem involves the techniques of a cost function J that minimizes, in a least-square formulation, the quadratic distance between the solution of the state equation (1) and given observations:

$$J(v) = \frac{1}{2}\Big(\int_\Omega |y(v) - y_d|^2 dx + \nu \int_{\Gamma_E \cup \Gamma_W} |v|^2 d\sigma\Big).$$

And we set the optimal control problem as:

$$(\mathcal{P}) \qquad \inf_{v \in \mathcal{U}_{ad}} J(v) = J(u), \qquad u \in \mathcal{U}_{ad},$$

where $\mathcal{U}_{ad}$ is a set of admissible controls. The solution of $(\mathcal{P})$ is commonly based on descent methods [Lio68]: At the nth iteration, from the known u^n, we compute successively the direct state $y(u^n)$ and the adjoint state $p(u^n)$. We then get the value of $J(u^n)$ and the gradient $\nabla J(u^n)$ which is an expression of $p(u^n)$ and u^n (See [Lio68]). A minimization step is shown in Figure 1.

Discretization and numerical framework

The domain Ω is meshed by a uniform grid $\Delta x = \Delta y = h = \frac{1}{N+1}$ (N is the number of points in the y-direction). A finite difference scheme is used to discretize the direct state $y(u_n)$. The discrete adjoint state $p(u_n)$ is then deduced from the transpose system of $y(u_n)$ with the appropriate right hand side.

[1]Department of Informatics, University of Oslo, Norway, aichab@ifi.uio.no

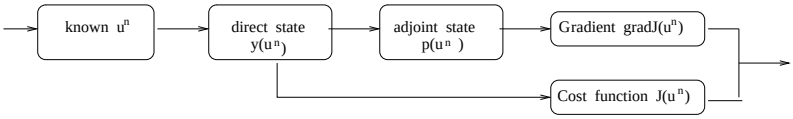


Figure 1: One minimization step: calculation of the cost function and its gradient.

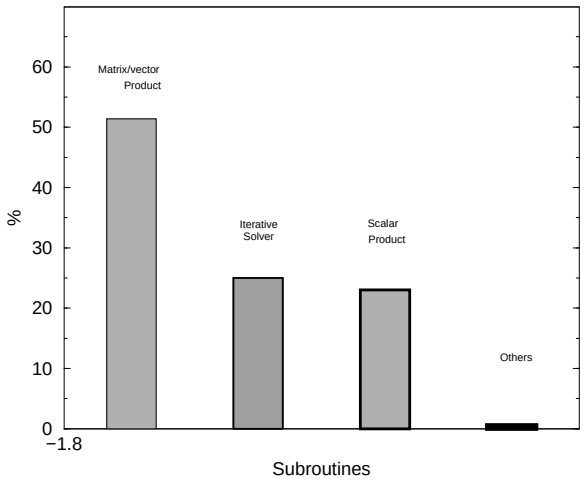


Figure 2: Profiling of the sequential code on the whole domain, $N = 64$.

The resulting linear systems are solved by a Krylov solver: `Bicgstab` (Stabilized bi-conjugate gradient). The minimization phase is carried out by the quasi-Newton method with the BFGS formula [2][GL89].
For the numerical tests, we have the following:
$\Omega =]0, 4[\times]0, 1[$ is the domain of computation, and for (x, y) in Ω:

$$\begin{aligned} f(x,y) &= 2(-x^2 - y^2 + 4x + y), \\ y_d(x,y) &= (x^2 - 4x)(y - y^2) - 8\nu. \end{aligned}$$

Remark. The computation of the discrete gradient of the cost function is the main step of the minimization process, since the precision of the descent method depends on the precision of the discrete gradient calculation.

Motivation

When we solve sequentially the optimal control problem $(\mathcal{P})$ on the whole domain, we find that most of the CPU time required for the minimization process is related to the scalar and

[2]The M1QN3 code is developped in the MODULOPT project of INRIA by J.-C. Gilbert and C. Lemaréchal. We have used its double precision version: N1QN3.

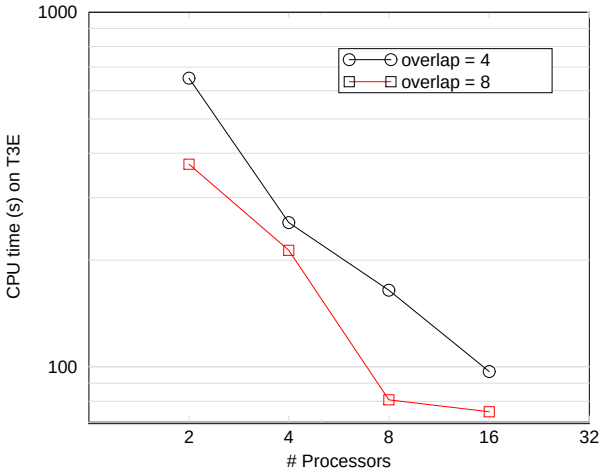


Figure 3: CPU time versus the number of processors. Effect of the overlap size on the behavior of N1QN3 with multiplicative Schwarz method. $N = 64$.

matrix-vector products that are the base in the calculation of the cost function and its gradient. Figure 2 shows the time percentage of each code part. We need 10 iterations to achieve the given precision $\epsilon = \frac{\|\nabla J(u^n)\|}{\|\nabla J(u^0)\|} = 10^{-6}$ in N1QN3.
So, we propose to implement efficient algorithms for parallel architectures using a load allocation of the solvers of both the discrete direct and adjoint states.

Domain decomposition techniques

The main idea of the proposed domain decomposition method consists in using iterative Schwarz methods either as solver or preconditioner for the direct and adjoint linear systems required at each step of the minimization algorithm. In contarst, the minimization instead remains global over the domain of calculation, i.e., the control in N1QN3 is not decomposed. All the results are given for the parallel machine CRAY-T3E using the message passage interface library MPI.

The Schwarz algorithm as a solver

Description

We consider an overlapping decomposition of the domain Ω and using the multiplicative version of the Schwarz algorithm with Dirichlet boundary conditions to solve the direct state so that we get on each subdomain Ω_i:

$$A_i y_i^{n+1} = F_i(f_i, u_i^{n+1}, y_j^n|_{\partial\Omega_i \cap \Omega_j}). \tag{2}$$

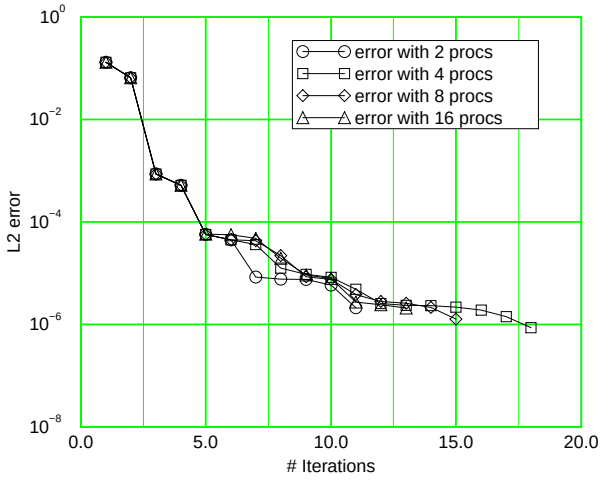


Figure 4: L^2 error, $N = 64$, $\delta_i = 4$. Behavior of N1QN3 with multiplicative Schwarz method as solver.

The discrete adjoint state is then computed by transposing the y_i^{n+1}-local system (2) such that we get formally the p_i^{n+1} system:

$$A_i^t p_i^{n+1} = G_i(y_i^{n+1} - y_{d,i}, p_j^n|_{\partial\Omega_i \cap \Omega_j}) \tag{3}$$

Analysis of numerical results

The tests were carried out on a mesh of $4N \times N$ nodes where $N = 64$ and for a stopping criterion $\epsilon = \frac{\|\nabla J(u^n)\|}{\|\nabla J(u^0)\|} = 10^{-6}$ in the minimization method N1QN3. The local linear systems are solved by the `Bicgstab` method.
We study the behavior of the N1QN3 "minimizer" with respect to different parameters such as the overlap size, the type of the decomposition and the number of processors. Furthermore, to make the implementation possible on the parallel machine, we have used a coloring technique such that neighbouring subdomains have different colors.

¿From Figure 3, it is shown that the CPU time drops when the overlap gets large (in fact, we consider in the figures the relative overlap δ_i which is linked to the real overlap δ between two subdomains by $\delta = 2\delta_i h$). This reflects one of the properties of the multiplicative Schwarz method [SBG96].

In addition, to show the effect of the multiplicative Schwarz method mixed with the N1QN3 otpimizer, we present in Figure 4 the L^2-error between the computed solution (the direct state associated with the computed optimal control) and the analytical solution against the number of iterations in N1QN3. For different numbers of processors and with a relative overlap equal to 4, it is shown that it is only from the 5th iteration that the precision deteriorates.
The number of iterations in the optimiser also varies slightly (in fact, N1QN3 needs only 10 iterations for the whole domain). Thus, the best result is obtained with 2 processors but for

more processors the precision is almost lost. One can conclude that this is due to the "oscillations" of the precision quantity of N1QN3 within the communications between subdomains.

The Schwarz algorithm as a preconditioner

Let $\mathbf{x}_0$, ϵ, ϵ_{stop} be given
refresh_news($\mathbf{x}_0$)
$\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$, $\mathbf{v} = \mathbf{p} = 0$, $\rho_0 = \alpha = \omega = 1$
While $\dfrac{\text{Global\_}\|\mathbf{r}_k\|_2}{\text{Global\_}\|\mathbf{r}_0\|_2} > \epsilon$ and Global_ $\|\mathbf{r}_k\|_2 > \epsilon_{stop}$ Do
$\rho = \text{Global\_}({\mathbf{r}_0}^T\mathbf{r})$
$\beta = \alpha\rho/\rho_0\omega$, $\rho_0 = \rho$
$\mathbf{p} = \mathbf{r} + \beta(\mathbf{p} - \omega\mathbf{v})$
Solve $\mathbf{C}\widehat{\mathbf{p}} = \mathbf{p}$
refresh_news($\widehat{\mathbf{p}}$)
$\mathbf{v} = \mathbf{A}\widehat{\mathbf{p}}$
$\alpha = \rho_1/\text{Global\_}({\mathbf{r}_0}^T\mathbf{v})$
$\mathbf{s} = \mathbf{r} - \alpha\mathbf{v}$
Solve $\mathbf{C}\mathbf{z} = \mathbf{s}$
refresh_news($\mathbf{z}$)
$\mathbf{t} = \mathbf{A}\mathbf{z}$
$\omega = \dfrac{\text{Global\_}(\mathbf{t}^T\mathbf{s})}{\text{Global\_}(\mathbf{t}^T\mathbf{t})}$
$\mathbf{x} = \mathbf{x} + \alpha\widehat{\mathbf{p}} + \omega\mathbf{z}$
$\mathbf{r} = \mathbf{s} - \omega\mathbf{t}$
Endwhile

Figure 5: Partitioned Bicgstab algorithm

Description

It is well known that when used as a preconditioner of a parallel Krylov solver, the overlapping domain decomposition methods allow us to improve the convergence rate of such iterative linear system solvers and to limit the time of communications needed for their implementation on parallel architectures. In the preconditioning step of the distributed `Bicgstab` (Figure 5), we first extend the local contribution of `s` or `p` to the subdomain enlarged by the overlap in the four cardinal directions. Then, on each subdomain, we solve exactly the local problem with Dirichlet boundary condiditons. And finally, the global solution $\hat{\mathbf{p}}$ or `z` is deduced from the projections of the solution of each local problem. We have used the same notations as in [KA98] (see also [KST95]).

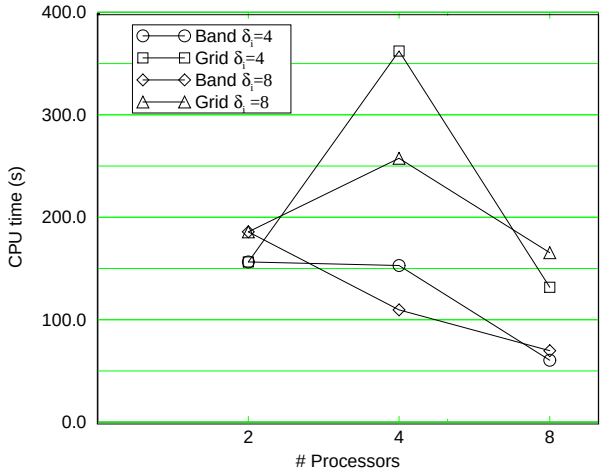


Figure 6: CPU time(s) on Cray-T3E versus the number of processors. $N = 64$. Effect of the decomposition type on the CPU time of N1QN3 algorithm used with preconditioned `Bicgstab`.

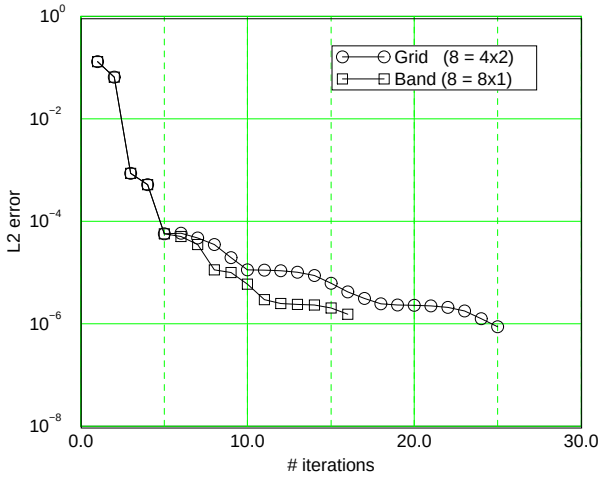


Figure 7: L^2 error versus the iteration number of N1QN3. $N = 64$, $\delta_i = 4$. Effect of the decomposition type on the behavior of N1QN3 with preconditioned `Bicgstab`.

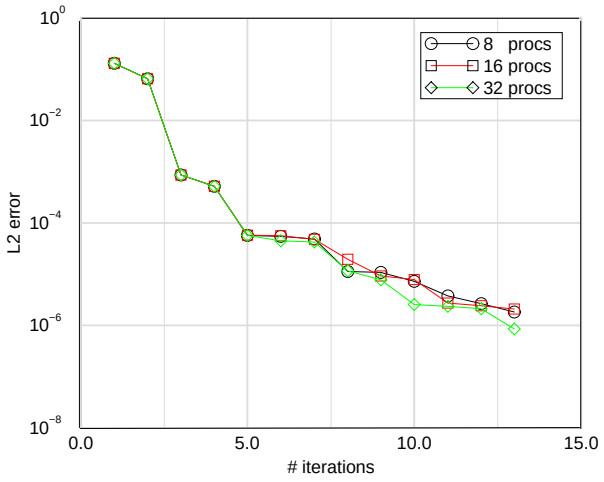


Figure 8: L^2 error versus the iterations number of N1QN3. $N = 64$, $\delta_i = 4$. Behavior of N1QN3 with `Bicgstab` preconditioned versus the number of processors.

Analysis of numerical results

We first remark that the cpu times are better than those obtained with the multiplicative Schwarz method as a solver. On Figure 6, we observe that for $\delta_i = 8$, the cpu time is halved when the number of processors goes from 4 to 8. From Figure 7, we observe that the cpu times are small when $\delta_i = 4$ and the computations are done on 16 processors. We observe in the same figure the important effect of the decomposition type on the behavior of the optimiser N1QN3: with 8 band-disposed processors, we need more iterations than the grid disposition of the processors (in this case, we have 4 processors in x-direction and 2 in y-direction) to reach the given precision in the optimiser. Obviously, in the case of a small size of the problem, a grid decomposition involves more communication than a band one.
From this test series, the relative overlap $\delta_i = 4$ should be an optimal one for the preconditioned distributed `Bicgstab` since the behavior of N1QN3 is not affected by the number of processors (Figure 8).

Conclusion

The methods presented mix minimization algorithms and iterative Schwarz methods (solver or preconditioner). In both cases, the optimal control is computed for a given stopping criterion and the influence of the decomposition parameters on the behavior of the minimization method is shown.
The multiplicative Schwarz method used in the solution of an optimal control problem yields a robust but time consuming method, whereas the additive method used as a preconditioner at each step of the minimization process is less time consuming. The best results are obtainded for a relative overlapping of 25%. Moreover, we have compared this method with the direct

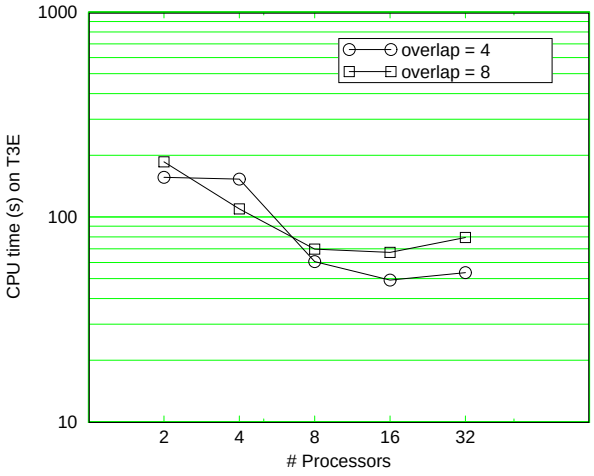


Figure 9: CPU time (s) on Cray-T3E versus the number of processors, $N = 64$. Effect of the overlapping on the behavior of N1QN3 with `Bicgstab` preconditioned by the additive Schwarz method.

parallelization of `Bicgstab` [Bou99] and it is expected (from the curved look of the Figure 9) to be more competitive for a large number of degrees of freedom since the minimizer is only slightly affected by the second method.

References

[Bou99]Aïcha Bounaim. *Domain Decomposition Methods: Application to the Solution of Optimal Control Problems*. PhD thesis, Joseph Fourier University, Grenoble, France, June 1999. (in french).

[GL89]Jean-Charles Gilbert and Claude Lemaréchal. Some numerical experiments with variable storage quasi-Newton algorithms. *Mathematical Programming*, 45:407–435, 1989.

[KA98]Samuel Kortas and Philippe Angot. Parallel preconditioners for a fourth-order discretization of the viscous Bürgers equation. In Petter E. Bjørstad, Magne Espedal, and David Keyes, editors, *Proceedings of the 9th international conference on domain decomposition method*, pages 387–405, 1998.

[KST95]David E. Keyes, Youcef Saad, and Donald G. Truhlar, editors. *Domain-Based Parallelism and Problem Decomposition Methods in Computational Sciences and Engineering*. SIAM, 1995.

[Lio68]J.L. Lions. *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles*. Paris. Dunod, Gauthier Villars, 1968.

[SBG96]Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

33 RASHO: A Restricted Additive Schwarz Preconditioner with Harmonic Overlap

X.-C. Cai[1], M. Dryja[2], M. Sarkis[3]

Introduction

A restricted additive Schwarz (RAS) preconditioning technique was introduced recently for solving general nonsymmetric sparse linear systems [BGMS97, CFS98, CS99, FS01, GKK+00, LSHF01, SK00, QV99]. The RAS preconditioner improves the classical additive Schwarz preconditioner (AS), [SBG96], in the sense that it reduces the number of iterations of the iterative method, such as GMRES, and also reduces the communication cost per iteration when implemented on distributed memory computers. However, RAS in its original form is a nonsymmetric preconditioner and therefore can not be used with the Conjugate Gradient method (CG). In this paper, we provide an extension of RAS for symmetric positive definite problems using the so-called harmonic overlaps (RASHO). Both RAS and RASHO outperform their counterparts of the classical additive Schwarz variants. Roughly speaking, the design of RASHO is based on a deeper understanding of the behavior of Schwarz type methods in the overlapping regions, and in the construction of the overlap. Under RASHO, the overlap is obtained by extending the nonoverlapping subdomains only in the directions that do not cut the boundaries of other subdomains, and all functions are made harmonic in the overlapping regions. As a result, the subdomain problems in RASHO are smaller than those of AS, and the communication cost is also smaller when implemented on distributed memory computers, since the right-hand sides of discrete harmonic systems are always zero which does not need to be communicated. We will show numerically that the RASHO preconditioned CG takes fewer iterations than the corresponding AS preconditioned CG. An almost optimal convergence theory will be presented for the RASHO for elliptic problems discretized with a finite element method.

Recall that the basic building blocks of classical Schwarz type algorithms are the operations of the form $(R_i^\delta)^T (A_i^\delta)^{-1} R_i^\delta$, where A_i^δ is the subdomain matrix and R_i^δ is the restriction operator for the extended subdomain (formal definitions will be given later in the paper). The multiplication of the such an operator with a vector, v, is realized by solving the linear system

$$A_i^\delta w = R_i^\delta v \tag{1}$$

on each extended subdomain. The key idea of RAS is that equation (1) is replaced by

$$A_i^\delta w = \begin{cases} v & \text{inside the unextended subdomain} \\ 0 & \text{in the overlapping part of the subdomain.} \end{cases} \tag{2}$$

[1]Department of Computer Science, University of Colorado, Boulder, CO 80309, (*cai@cs.colorado.edu*). The work was supported in part by the NSF grants ASC-9457534, ECS-9725504, and ACI-0072089

[2]Faculty of Mathematics, Informatics and Mechanics, Warsaw University, Warsaw, (*dryja@mimuw.edu.pl*). This work was supported in part by the NSF grant CCR-9732208 and in part by the Polish Science Foundation grant 2 P03A 021 16

[3]Mathematical Sciences Department, Worcester Polytechnic Institute, Worcester, MA 01609, (*msarkis@wpi.edu*). The work was supported in part by the NSF grant CCR-9984404

Note that the solution of (2) is discrete harmonic in the overlapping part of the subdomain, and therefore carries minimum energy in some sense. In this paper, we further explore the idea of "harmonic overlap" and at the same time keep the symmetry of the preconditioner.

The algorithm to be discussed below is applicable for symmetric positive definite problems. In order to provide a complete mathematical analysis, we restrict ourselves to the Poisson problem discretized with a finite element method. We consider a simple variational problem: Find $u \in H_0^1(\Omega)$, such that

$$a(u,v) = f(v), \ \ \forall\, v \in H_0^1(\Omega), \tag{3}$$

where

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v \, dx \ \text{ and } \ f(v) = \int_\Omega fv \, dx \ \text{ for } \ f \in L^2(\Omega).$$

For simplicity, let Ω be a bounded polygonal region in $\Re^2$ with a diameter of size $O(1)$. The extension of the algorithm and results to $\Re^3$ can be carried out easily. Let $\mathcal{T}^h(\Omega)$ be a shape regular, quasi-uniform triangulation, of size $O(h)$, of Ω and $\mathcal{V}(\Omega) \subset H_0^1(\Omega)$ the finite element space consisting of continuous piecewise linear functions associated with the triangulation. We are interested in solving the following discrete problem associated with (3): Find $u^* \in \mathcal{V}$ such that

$$a(u^*,v) = f(v), \ \ \forall\, v \in \mathcal{V}. \tag{4}$$

Using the standard basis functions, (4) can be rewritten as a linear system of equations

$$Au^* = f. \tag{5}$$

For simplicity, we understand u^* and f both as functions and vectors depending on the situation.

Notations

Let n be the total number of interior nodes of $\mathcal{T}^h(\Omega)$ and W the set of nodes. We assume that a node-based partitioning has been applied and resulted in N nonoverlapping subsets $W_i^0, i = 1,\ldots,N$, whose union is W. For each W_i^0, we define a region Ω_i^R as the union of all elements of $\mathcal{T}^h(\Omega)$ that have all three vertices on $W_i^0 \cup \partial\Omega$. We denote H as the representative size of the subregion Ω_i^R. We define the overlapping partition of W as follows. Let $\{W_i^1\}$ be the one-overlap partition of W, where $W_i^1 \supset W_i^0$ is obtained by including all the immediate neighboring vertices of the vertices in W_i^0. Using the idea recursively, we can define a δ-overlap partition $W = \bigcup_{i=1}^N W_i^\delta$. δh is approximately the extend of the extension.

We next define a subregion of Ω induced by a set of nodes of $\mathcal{T}^h(\Omega)$ as follows. Let Z be a subset of W. The induced subregion, denoted as $\Omega(Z)$, is defined as the union of: (1) the set Z itself; (2) the union all the open elements (triangles) of $\mathcal{T}^h(\Omega)$ that have at least one vertex in Z; and (3) the union of the open edges of these triangles that have at least one endpoint as a vertex of Z. Note that $\Omega(Z)$ is always an open region. The extended region Ω_i^δ is defined as $\Omega(W_i^\delta)$. We introduce the subspace

$$\mathcal{V}_i^\delta \equiv \mathcal{V} \cap H_0^1(\Omega_i^\delta) \text{ extended by zero to } \Omega\backslash\Omega_i^\delta.$$

It is easy to check that

$$\mathcal{V} = \mathcal{V}_1^\delta + \mathcal{V}_2^\delta + \cdots + \mathcal{V}_N^\delta.$$

This decomposition is used in defining the classical one-level additive Schwarz algorithm without a coarse space [SBG96]. Let us define $P_i^\delta : \mathcal{V} \to \mathcal{V}_i^\delta$ by

$$a(P_i^\delta u, v) = a(u, v), \ \ \forall u \in \mathcal{V}, \quad \forall v \in \mathcal{V}_i^\delta. \tag{6}$$

Then, the classical one-level additive Schwarz operator has the form

$$P^\delta = P_1^\delta + \cdots + P_N^\delta.$$

Let $\Gamma_i^\delta = \partial\Omega_i^\delta \backslash \partial\Omega$; i.e., the part of the boundary of Ω_i^δ that does belong to the Dirichlet part of the boundary. We define the interface overlapping boundary Γ^δ as the union of all Γ_i^δ; i.e., $\Gamma^\delta = \cup_{i=1}^N \Gamma_i^\delta$. We then define the following subsets of W:

- $W^{\Gamma^\delta} \equiv W \cap \Gamma^\delta$ (interface nodes)
- $W_i^{\Gamma^\delta} \equiv W^{\Gamma^\delta} \cap W_i^\delta$ (local interface nodes)
- $W_{i,in}^{\Gamma^\delta} \equiv W^{\Gamma^\delta} \cap W_i^0$ (local internal interface nodes)
- $W_{i,cut}^{\Gamma^\delta} \equiv W_i^{\Gamma^\delta} \backslash W_{i,in}^{\Gamma^\delta}$ (local cut interface nodes)
- $W_{i,ovl}^\delta \equiv (W_i^\delta \backslash W_i^{\Gamma^\delta}) \cap (\bigcup_{j \neq i} W_j^\delta)$ (local overlapping nodes)
- $W_{i,non}^\delta \equiv W_i^\delta \backslash (W_i^{\Gamma^\delta} \cup W_{i,ovl}^\delta)$ (local nonoverlapping nodes)
- $W_{i,in}^\delta \equiv W_{i,non}^\delta \cup W_{i,in}^{\Gamma^\delta}$ (internal nodes)

We note that the notions of subdomains, harmonic overlaps, the classification of nodal points can all be defined in terms of the graph of the sparse matrix.

We frequently use functions that are discrete harmonic at certain nodes. Let $x_k \in W$ be a mesh point and $\phi_{x_k}(x) \in \mathcal{V}$ the finite element basis function associated with x_k; i.e., $\phi_{x_k}(x_k) = 1$, and $\phi_{x_k}(x_j) = 0, j \neq k$. We say $u \in \mathcal{V}$ is discrete harmonic at x_k if $a(u, \phi_{x_k}) = 0$. If u is discrete harmonic at a set of nodal points Z, we say u is discrete harmonic in $\Omega(Z)$.

Our new algorithm will be built on $\widetilde{\mathcal{V}}_i^\delta$ defined as a subspace of $\mathcal{V}_i^\delta$. $\widetilde{\mathcal{V}}_i^\delta$ consists of all functions in $\mathcal{V}_i^\delta$ that vanish on $W_{i,cut}^{\Gamma^\delta}$ and discrete harmonic at the nodes $W_{i,ovl}^\delta$. Note that the support of the subspace $\widetilde{\mathcal{V}}_i^\delta$ is

$$\widetilde{W}_i^\delta \equiv W_i^\delta \backslash W_{i,cut}^{\Gamma^\delta}$$

and, since the values at the harmonic nodes are not independent, they can not be counted toward the degree of freedoms. The dimension of $\widetilde{\mathcal{V}}_i^\delta$ is $dim(\widetilde{\mathcal{V}}_i^\delta) = |W_{i,in}^\delta|$. Let $\widetilde{\Omega}_i^\delta \equiv \Omega(\widetilde{W}_i^\delta)$ be the induced domain. It is easy to see that $\widetilde{\Omega}_i^\delta$ is the same as Ω_i^δ but with cuts. We have then $\widetilde{\mathcal{V}}_i^\delta = \mathcal{V} \cap H_0^1(\widetilde{\Omega}_i^\delta)$ and discrete harmonic on $\Omega_{i,ovl}^\delta \equiv \Omega(W_{i,ovl}^\delta)$. We define $\widetilde{\mathcal{V}}^\delta \subset \mathcal{V}^\delta$ as

$$\widetilde{\mathcal{V}}^\delta \equiv \widetilde{\mathcal{V}}_1^\delta \oplus \cdots \oplus \widetilde{\mathcal{V}}_N^\delta,$$

which is a direct sum. We remark that functions in $\widetilde{\mathcal{V}}^\delta$ are, by definition, the sum of functions $u_i \in \widetilde{\mathcal{V}}_i^\delta$, $i = 1, \cdots, N$. Functions in $\widetilde{\mathcal{V}}^\delta$ can, in fact, be characterized easily as in the following lemma.

Lemma 1 *[CDS01] If $u \in \mathcal{V}$ and u is discrete harmonic at all the overlapping nodes, i.e., on $\cup_{i=1}^N W_{i,ovl}^\delta$, then $u \in \widetilde{\mathcal{V}}^\delta$.*

RAS with Harmonic Overlap

Let $\widetilde{P}_i^\delta : \widetilde{\mathcal{V}}^\delta \to \widetilde{\mathcal{V}}_i^\delta$ be a projection operator satisfying

$$a(\widetilde{P}_i^\delta u, v) = a(u, v), \quad \forall u \in \widetilde{\mathcal{V}}^\delta, \quad \forall v \in \widetilde{\mathcal{V}}_i^\delta. \tag{7}$$

The RASHO operator can be defined as

$$\widetilde{P}^\delta = \widetilde{P}_1^\delta + \cdots + \widetilde{P}_N^\delta. \tag{8}$$

Note that the solution u^* of (5) is not, generally speaking, in the subspace $\widetilde{\mathcal{V}}^\delta$, therefore, the operator $\widetilde{P}^\delta$ can not be used to solve the linear system (5) directly. We will need to modify the right-hand side of the system; see Lemma 2. We will also show that the elimination of the variables associated with the overlapping nodes is not needed in order to apply $\widetilde{P}^\delta$ to a vector $v \in \widetilde{\mathcal{V}}^\delta$.

We now introduce the matrix form of (8). We define the restriction operator, or a matrix, $\widetilde{R}_i^\delta$ as follows. Let $v = (v_1, \ldots, v_n)^T$ be a vector corresponding to the nodal values of a function $u \in \mathcal{V}$; namely for any node $x_i \in W$, $v_i = u(x_i)$. For convenience, we say "v is defined on W". Its restriction on $\widetilde{W}_i^\delta$, $\widetilde{R}_i^\delta v$, is defined as

$$\left(\widetilde{R}_i^\delta v\right)(x_i) = \begin{cases} v_i & \text{if } x_i \in \widetilde{W}_i^\delta \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The matrix representation of $\widetilde{R}_i^\delta$ is given by a diagonal matrix with 1 for nodal points in $\widetilde{W}_i^\delta$ and zero for the remaining nodal points. We remark that, by way of definition, the operator $\widetilde{R}_i^\delta$ is symmetric; i.e., $(\widetilde{R}_i^\delta)^T = \widetilde{R}_i^\delta$. Use this restriction operator, we define the subdomain stiffness matrix as

$$\widetilde{A}_i^\delta = \widetilde{R}_i^\delta \, A \, (\widetilde{R}_i^\delta)^T,$$

which can also be obtained by the discretization of the original problem on $\widetilde{W}_i^\delta$ with zero Dirichlet data on nodes $W \setminus \widetilde{W}_i^\delta$. The matrix $\widetilde{A}_i^\delta$ is block diagonal with blocks corresponding to the structure of $\widetilde{R}_i^\delta$ and its inverse is understood as an inverse of the nonzero block. A matrix representation of $\widetilde{P}_i^\delta$ denoted also by $\widetilde{P}_i^\delta$ is equal to

$$\widetilde{P}_i^\delta = \left(\widetilde{A}_i^\delta\right)^{-1} \, A$$

and

$$\widetilde{P}^{\delta} = \left((\widetilde{A}_1^{\delta})^{-1} + \cdots + (\widetilde{A}_N^{\delta})^{-1} \right) \; A. \tag{10}$$

The next lemma tell us how to modify the system (5) so that its solution belongs to $\widetilde{\mathcal{V}}^{\delta}$.

Lemma 2 *[CDS01] Let u^* and f be the exact solution and the right-hand side of (5), and*

$$w = \sum_{i=1}^{N} (\widetilde{A}_i^{\delta})^{-1} \widetilde{R}_i^0 f, \tag{11}$$

where $\widetilde{R}_i^0$ is defined by (9) with $\delta = 0$. Then, we have $\widetilde{u}^* = u^* - w \in \widetilde{\mathcal{V}}^{\delta}$ is the solution of the modified linear system of equations

$$A\widetilde{u}^* = f - Aw = \widetilde{f}.$$

We remark that RASHO has several advantages over the classical AS. Let us recall AS briefly. Let

$$\left(R_i^{\delta} v\right)(x_i) = \begin{cases} v_i & \text{if } x_i \in W_i^{\delta} \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Then the AS operator takes the following matrix form

$$P^{\delta} = \left((A_1^{\delta})^{-1} R_1^{\delta} + \cdots + (A_N^{\delta})^{-1} R_N^{\delta} \right) \; A, \tag{13}$$

where $A_i^{\delta} = R_i^{\delta} A (R_i^{\delta})^T$. We remark that the size of the matrix A_i^{δ} is $|W_i^{\delta}|$, which is bigger than the size of the matrix $\widetilde{A}_i^{\delta}$, which is $|\widetilde{W}_i^{\delta}|$. In a distributed memory implementation, the operation $R_i^{\delta} v$ involves moving data from one processor to another, but the operation $\widetilde{R}_i^{\delta} v$ does not involve any communication. In RASHO, if $u \in \widetilde{\mathcal{V}}^{\delta}$, then it is easy to see that

$$\widetilde{R}_i^{\delta} A u = \widetilde{R}_{i,in}^{\delta} A u, \tag{14}$$

where $\widetilde{R}_{i,in}^{\delta}$ is defined as

$$\left(\widetilde{R}_{i,in}^{\delta} v\right)(x_i) = \begin{cases} v_i & \text{if } x_i \in W_{i,in}^{\delta} \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Therefore, for functions in $\widetilde{\mathcal{V}}^{\delta}$, we can rewrite $\widetilde{P}^{\delta}$, as in (10), in the following form

$$\widetilde{P}^{\delta} = \left((\widetilde{A}_1^{\delta})^{-1} \widetilde{R}_{1,in}^{\delta} + \cdots + (\widetilde{A}_N^{\delta})^{-1} \widetilde{R}_{N,in}^{\delta} \right) \; A. \tag{16}$$

Although the operator (16) does not look like a symmetric operator, it is indeed symmetric when applying to functions in the subspace $\widetilde{\mathcal{V}}^{\delta}$. The form (14) takes the advantage of the fact that the operator $\widetilde{R}_{i,in}^{\delta}$ is communication-free in the sense that it needs only the residual associated with nodes in $W_{i,in}^{\Gamma^{\delta}} \subset \Omega_i^0$.

We make some further comments on how the residual Au can be calculated in a distributed memory environment, for a given vector $u \in \widetilde{\mathcal{V}}^\delta$. In a typical implementation, the matrix A is constructed and stored in the form of $\{\widetilde{A}_i^\delta\}$, each processor has one or several of the subdomain matrix $\widetilde{A}_i^\delta$. Similarly u is stored in the form of $\{u_i\}$, where $u_i \in \widetilde{\mathcal{V}}_i^\delta$. We note, however, that to compute the residual at nodes $W_{i,in}^{\Gamma^\delta}$ some communications are required. The processor associated with subdomain Ω_i^δ needs to obtain the local solution from the neighboring subdomains at nodes connected to $W_{i,in}^{\Gamma^\delta}$. It is important to note that the amount of communications does not depend on the size of the overlap since only one layer of nodes is required. This shows that, in terms of the communication cost, RASHO is superior to AS and RAS.

Main Results

The algorithm presented in the previous section is applicable for general sparse, symmetric positive definite linear systems. The notions of subdomains, harmonic overlaps, the classification of nodal points, etc, can all be defined in terms of the graph of the sparse matrix. The following theorem provides a nearly optimal estimate of the condition number of the RASHO operator $\widetilde{P}^\delta$ in terms of the fine mesh size h, the subdomain size H, and the overlapping factor δ for a Poisson equation discretized with a piecewise linear finite element method. We note that because we do not include a coarse space, the constant will depend on the subdomain size H.

Theorem 1 *[CDS01] The RASHO operator $\widetilde{P}^\delta$ is symmetric in the inner product $a(\cdot,\cdot)$, nonsingular, and bounded in the following sense*

$$C_0^{-2} a(u,u) \le a(\widetilde{P}^\delta u, u) \le C_1\ a(u,u) \quad \forall u \in \widetilde{\mathcal{V}}^\delta. \tag{17}$$

Here

$$C_0^2 = C\left((1+\log(\delta+1))\left(1+\log\left(\frac{H}{h}\right)\right) + \frac{1}{H^2}\left(1+\log(\delta+1)+\frac{H}{(2\delta+1)h}\right)\right).$$

The constants $C, C_1 > 0$ are independent of h, H, and δ.

We remark that the corresponding convergence rate estimate for the regular one-level AS [DW94], in terms of the constant C_0, is

$$C_0^2 = C\left(1+\frac{1}{H(2\delta+1)h}\right).$$

The lower bound C_0^2 of RASHO is theoretically slightly worse than the lower bound of AS in the case of large overlap, but roughly the same for the case of small overlap. On the other hand, the upper bound C_1 of RASHO is better, since the overlap bewteen subspaces $\widetilde{\mathcal{V}}_k^\delta$ is generally smaller than the overlap between subspaces $\mathcal{V}_k^\delta$. Because of the smaller upper bound, the numerical performance of RASHO presented in the next section is better than that of AS. It is interesting to point out that, for the case of generous overlap, our estimate is equivalent to the estimate for the iterative substructuring algorithms [DSW94] without a coarse space. We also remark that the results of the paper is for only one-level Schwarz algorithms. Because of the "harmonic overlap" requirement, the extension of the algorithm to multiply levels is not as trivial as the multilevel AS.

Table 1: RASHO and AS preconditioned CG for solving the Poisson equation on a 128×128 mesh decomposed into $2 \times 2 = 4$ subdomains with overlap = $ovlp$. The AS/CG results are shown in ( ). The "+1" is for the preprocessing step needed for RASHO.

$ovlp$	iter	cond	max	min
0	42 (42)	129.(129.)	1.98 (1.98)	0.0154 (0.0154)
1	24+1 (28)	48.4 (86.3)	1.94 (4.00)	0.0402 (0.0464)
2	20+1 (23)	33.3 (51.8)	1.91 (4.00)	0.0574 (0.0773)
3	18+1 (20)	27.2 (37.0)	1.89 (4.00)	0.0694 (0.1081)

Table 2: RASHO and AS preconditioned CG for solving the Poisson equation on a $32 * DOM \times 32 * DOM$ mesh decomposed into $DOM \times DOM$ subdomains with overlap = 1.

$DOM \times DOM$	iter	cond	max	min
2×2	19+1 (20)	26.8 (43.7)	1.89 (4.00)	0.0708 (0.0916)
4×4	39+1 (42)	86.9 (145.)	1.95 (4.00)	0.0225 (0.0276)
8×8	75+1 (78)	328. (550.)	1.97 (4.00)	0.0060 (0.0073)
16×16	147+1(156)	1295(2168.)	1.98 (4.00)	0.0015 (0.0018)

Numerical Experiments

We present some numerical results for solving the Poisson equation on the unit square with zero Dirichlet boundary conditions. We compare the performance of RASHO/CG and AS/CG in terms of the number of iterations and the condition numbers. We pay particular attention to the dependence on the number of subdomains and the size of the overlap.

In order to use RASHO/CG, we need to modify the linear system by forcing its modified solution to belong to $\widetilde{\mathcal{V}}^{\delta}$. To do so, we use (11). The stopping condition for CG is to reduce the energy norm of the initial residual by a factor of 10^{-6}. The exact solution of the equation is taken to be $u(x,y) = e^{5(x+y)} \sin(\pi x) \sin(\pi y)$. All subdomain problems are solved exactly. The iteration count (iter), the condition number (cond), the maximum (max) and minimum (min) eigenvalues of the preconditioned matrix are summarized in Table 1, and Table 2. It is clear that the newly introduced RASHO/CG is always better than the classical AS/CG in terms of the iteration counts and the condition numbers. Although we do not have any parallel results to report at this point, we are confident that RASHO/CG would be even better than AS/CG on a parallel computers with distributed memory since much less communication is required.

References

[BGMS97]S. Balay, W. Gropp, L.C. McInnes, and B. Smith. *PETSc 2.0 User Manual*. Argonne National Laboratory, http://www.mcs.anl.gov/petsc/, 2001.

[CDS01]X.-C. CAI, M. DRYJA, AND M. SARKIS. Restricted additive Schwarz preconditioners with harmonic overlap for symmetric positive definite linear systems. *Tech Report*, Dept. of Computer Science, University of Colorado at Boulder, 2001.

[CFS98]X.-C. Cai, C. Farhat, and M. Sarkis. A minimum overlap restricted additive Schwarz preconditioner and applications to 3D flow simulations. *Contemporary Mathematics*, 218:479–485, 1998.

[CS99]X.-C. Cai and M. Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM J. Sci. Comput.*, 21:239–247, 1999.

[DSW94]M. Dryja, B. Smith, and O. Widlund. Schwarz analysis of iterative substructuring algorithms for elliptic problems in three dimensions. *SIAM J. Numer. Anal.*, 31(6):1662–1694, December 1994.

[DW94]M. Dryja and O. Widlund. Domain decomposition algorithms with small overlap. *SIAM J. Sci.Comput.*, 15(3):604–620, May 1994.

[FS01]A. Frommer and D. Szyld. An algebraic convergence theory for restricted additive Schwarz methods using weighted max norms. *SIAM Journal on Numerical Analysis*, 39:463–479, 2001.

[GKK+00]W. GROPP, D. KAUSHIK, D. KEYES, AND B. SMITH, *Performance modeling and tuning of an unstructured mesh CFD application*, Proceedings of SC2000, IEEE Computer Society, 2000.

[LSHF01]M. Lesoinne, M. Sarkis, U. Hetmaniu, and C. Farhat. A linearized method for the frequency analysis of three-dimensional fluid/structure interaction problem in all flow regimes. *Comp. Meth. Appl. Mech. and Eng.*, 190:3121–3146, 2001.

[QV99]A. Quarteroni and A. Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

[SBG96]B. Smith, P. Bjørstad, and W. Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[SK00]M. Sarkis and B. Koobus. A scaled and minimun overlap restricted additive Schwarz method with application on aerodynamics. *Comp. Meth. Appl. Mech. and Eng.*, 184:391–400, 2000.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

34 A Nonlinear Additive Schwarz Preconditioned Inexact Newton Method for Shocked Duct Flows

Xiao-Chuan Cai[1], David E. Keyes[2], David P. Young[3]

Introduction

A nonlinearly preconditioned inexact Newton algorithm (PIN) was recently introduced, in [CK00], for solving large sparse nonlinear system of equations arising from the discretization of nonlinear partial differential equations. In PIN the nonlinear system $F(u) = 0$ is transformed into a new nonlinear system $\mathcal{F}(u) = 0$, which has the same solution as the original system. For certain applications the nonlinearities of the new function $\mathcal{F}(u)$ are more balanced and, as a result, the inexact Newton method converges more rapidly. In this paper, we shall use the nonlinear additive Schwarz algorithm as the preconditioner and focus on the performance of PIN for a compressible shock tube problem, which is known to be a difficult test case for inexact Newton type algorithms.

A motivating problem

We consider a one-dimensional compressible flow problem described by the full potential equation in a variable-area duct [BBH+93]. The problem is to determine the solution potential $u(x)$ satisfying

$$(A\rho u_x)_x = 0, \tag{1}$$

for $0 < x < 2$ and $u(0) = 0$ and $u(2) = u_R$ given. The duct area

$$A = A(x) = 0.4 + 0.6(x-1)^2,$$

and the density ρ is given by

$$\rho = \rho(v) = (c^2)^{1/(\gamma-1)} = \left(1 + \frac{\gamma-1}{2}(1-v^2)\right)^{1/(\gamma-1)}.$$

Here $v = u_x$ is the velocity, $\gamma = 1.4$ is the ratio of specific heat and c is the speed of sound. The flow is supersonic at each point of the interval $(0, 2)$ where the Mach number $M = |v|/c$ exceeds 1. We use a standard finite difference method to discretize (1) on a uniform mesh

$$0 = x_0 < x_1 < \cdots < x_n < x_{n+1} = 2.$$

[1]Department of Computer Science, University of Colorado, Boulder, CO 80309-0430 (*cai@cs.colorado.edu*). The work was supported in part by the NSF grants ASC-9457534, ECS-9725504, and ACI-0072089, and by Lawrence Livermore National Laboratory under subcontract B509471.

[2]Department of Mathematics & Statistics, Old Dominion University, Norfolk, VA 23529-0077; ISCR, Lawrence Livermore National Laboratory, Livermore, CA 94551-9989; and ICASE, NASA Langley Research Center, Hampton, VA 23681-2199 (*keyes@icase.edu*). This work was supported in part by NASA under contract NAS1-19480 and by Lawrence Livermore National Laboratory under subcontract B347882.

[3]The Boeing Company, Seattle, WA 98124. (*dpy6629@cfdd51.cfd.ca.boeing.com*).

Let $u^h = (u_1^h, \ldots, u_n^h)^T$ be the solution vector of the finite difference problem, and

$$v_i = (u_{i+1}^h - u_i^h)/(x_{i+1} - x_i).$$

The discrete nonlinear problem is of the form:

$$A_j \tilde{\rho}_j v_j = A_{j+1} \tilde{\rho}_{j+1} v_{j+1}, \quad j = 1, \ldots, n, \tag{2}$$

where A_j denotes the midpoint value $A((x_j + x_{j+1})/2)$, and $\tilde{\rho}_j$ is an approximation of $\rho_j = \rho(x_j)$ defined using the so-called first order density biasing [BBH+93, HMS78],

$$\tilde{\rho}_j = \rho_j - \mu_j \Delta_- \rho_j,$$

where Δ_- denotes the undivided upwind difference operator, i.e., $\Delta_- \rho_j = \rho_j - \rho_{j-1}$, and where the switching function μ_j is defined as

$$\mu_j = \max_{j-k<m<j+k} \max \left\{ 0, 1 - \frac{M_c^2}{M_j^2} \right\}. \tag{3}$$

In (3), M_j is the local Mach number at $(x_j + x_{j+1})/2$ and M_c is a given cutoff Mach number taken to be 0.95 in this paper. k is the level of the switching function, which is taken to be 2 in our numerical experiments. This means that μ_j is replaced by the maximum of the 5 values centered around x_j. The switching function μ_j controls the amount of artificial viscosity. At points where $M_j < M_c$, no upwinding is applied therefore $\tilde{\rho}_j = \rho_j$. As M_j increases above M_c, $\tilde{\rho}_j$ provides an increasing amount of upwinding. In the following discussion, we denote the nonlinear system (2) in the form of a standard equation:

$$F(u^*) = 0, \tag{4}$$

where $F = (F_1, \ldots, F_n)^T$, $F_i = F_i(u_1, \ldots, u_n)$, and we drop the superscript h and simply use $u = (u_1, \ldots, u_n)^T$ to denote vectors in the space R^n. The problem looks rather simple; however, it is quite a challenging equation for the inexact Newton algorithm (IN), which is commonly used for solving such systems ([DS83, DES82, EW94]), and can briefly be described here. Suppose $u^{(k)}$ is the current approximate solution; a new approximate solution $u^{(k+1)}$ can be computed through the following steps: Find the inexact Newton direction $p^{(k)}$ such that

$$\|F(u^{(k)}) - F'(u^{(k)}) p^{(k)}\| \leq \eta_k \|F(u^{(k)})\|, \tag{5}$$

and then the new approximate solution

$$u^{(k+1)} = u^{(k)} - \lambda^{(k)} p^{(k)}.$$

Here η_k is a scalar that determines how accurately the Jacobian system needs to be solved using, for example, Krylov subspace methods [BS90, BS94, EW94, EW96]. $\lambda^{(k)}$ is another scalar that determines how far one should go in the selected inexact Newton direction [DS83]. IN has two well-known properties. First, if the initial guess is close enough to the desired solution then the convergence is very fast. Second, such a good initial guess is generally very difficult to obtain, especially for nonlinear equations that have unbalanced nonlinearities [LRW96]. The step length $\lambda^{(k)}$ is often determined by the components with the worst nonlinearities, and this may lead to an extended period of stagnation in the nonlinear residual curve; see Fig.2 for a typical picture and more in the references [CGK+98, GKM+00, JF95, PCS+99, YMB+90, YMB+91].

Descriptions of algorithms

Let us recall the nonlinearly preconditioned inexact Newton algorithms [CK00]: Find the solution $u^* \in R^n$ of (4) by solving a preconditioned system

$$\mathcal{F}(u^*) = 0. \tag{6}$$

Note that $\mathcal{F}$ and F may have different forms, but we require that they have the same solution. In general, $\mathcal{F}$ is a function of both F and u, and we do not expect to know explicitly how $\mathcal{F}$ depends on F or u. As an example, $\mathcal{F}$ may take the form of a composite function

$$\mathcal{F}(u^*) \equiv G(F(u^*)),$$

which makes G look like a preconditioner and some desirable properties of G include:

1. If $G(x) = 0$, then $x = 0$.
2. $G \approx F^{-1}$ in some sense.
3. $G(F(w))$ is easily computable for $w \in R^n$.
4. If a Newton-Krylov type method is used for solving (6), then the matrix-vector product $(G(F(w)))'v$ should also be easily computable for $w, v \in R^n$.

As in the linear equation case, the definition of a preconditioner can not be given precisely, nor is it necessary. Also as in the linear equation case, preconditioning can greatly improve the robustness of the iterative methods, since the preconditioner is designed so that the new system (6) has more uniform nonlinearities. Note that the Jacobian of the preconditioned function can be computed, at least in theory, using the chain rule; i.e.,

$$\mathcal{F}'(u) = \frac{\partial G}{\partial v}\frac{\partial F}{\partial u}. \tag{7}$$

If G is close to F^{-1} in the sense that $G(F(u)) \approx u$, then $\frac{\partial G}{\partial v}\frac{\partial F}{\partial u} \approx I$, i.e., $\mathcal{F}'(u) \approx I$. In this case, the algorithm converges in one iteration, or few iterations, depending on how close is G to F^{-1}. Most of the current research has been on the case of linear G; see, for example, [CGK+98, GKM+00, PW98]. In this paper, we shall focus on the case when G is the single-level nonlinear additive Schwarz method [CD94, DH97].

Let $S = (1, \ldots, n)$ be an index set; i.e., one integer for each unknown u_i and F_i. We assume that $S_1, \ldots, S_N$ is a partition of S in the sense that

$$\cup_{i=1}^{N} S_i = S, \text{ and } S_i \subset S.$$

Here we allow the subsets to have overlap. Let n_i be the dimension of S_i; then, in general, $\sum_{i=1}^{N} n_i \geq n$. Using the partition of S, we introduce subspaces of R^n and the corresponding restriction and extension matrices. For each S_i we define $V_i \subset R^n$ as

$$V_i = \{v | v = (v_1, \ldots, v_n)^T \in R^n, v_k = 0, \text{ if } k \notin S_i\}$$

and a $n \times n$ restriction (also extension) matrix I_{S_i} whose kth column is either the kth column of the $n \times n$ identity matrix $I_{n\times n}$ if $k \in S_i$ or zero if $k \notin S_i$. Similarly, let s be a subset

of S; we denote by I_s the restriction on s. Note that the matrix I_s is always symmetric and the same matrix can be used as both restriction and extension operator. Many other forms of restriction/extension are available in the literature; however, we only consider the simplest form in this paper.

Using the restriction operator, we define the subdomain nonlinear function as

$$F_{S_i} = I_{S_i} F.$$

We next define the major component of the algorithm, namely the nonlinearly preconditioned function. For any given $v \in R^n$, define $T_i(v) \in V_i$ as the solution of the following subspace nonlinear system

$$F_{S_i}(v - T_i(v)) = 0,$$

for $i = 1, \ldots, N$. We introduce a new function

$$\mathcal{F}(u) = \sum_{i=1}^{N} T_i(u), \tag{8}$$

which we will refer to as the nonlinearly preconditioned $F(u)$ and the corresponding algorithm additive Schwarz preconditioned inexact Newton method (ASPIN).

We remark that the evaluation of the function $\mathcal{F}(v)$, for a given v, involves the calculation of T_i, which in turn involves the solution of nonlinear systems on S_i. If the overlap is zero, then this is simply a block nonlinear Jacobi preconditioner. Assuming that all the subdomain problems are uniquely solvable, it is proved in [CK00] that the nonlinear systems (4) and (6) are equivalent in the sense that they have the same solution.

If (6) is solved using a Newton type algorithm, then the Jacobian is needed in one form or another. Let

$$J = F' = \left(\frac{\partial F_i}{\partial u_j}\right)_{n \times n} \text{ and } J_{S_i} = (I_{S_i} J I_{S_i})_{n \times n}$$

be the Jacobians of the original nonlinear system and subdomain nonlinear system, respectively. Then, as shown in [CK00], the Jacobian of the preconditioned nonlinear system can be approximated by

$$\mathcal{J} \approx \sum_{i=1}^{N} J_{S_i}^{-1} J. \tag{9}$$

(9) is an extremely interesting formula since it corresponds exactly to the additive Schwarz preconditioned linear Jacobian system of the original un-preconditioned equation. This fact implies that, first of all, we know how to solve the Jacobian system of the preconditioned nonlinear system, and second, the Jacobian itself is already well-conditioned. In other words, nonlinear preconditioning automatically offers a linear preconditioning for the corresponding Jacobian system.

Numerical experiments

We show a few numerical experiments in this section using ASPIN. In all the experiments, the subdomain Jacobian matrices J_{S_i} are formed using a finite difference scheme. The implementation is done using PETSc [BGM+01] on a cluster of workstations. In the tests, we always set $u_R = 1.15$ and the corresponding Mach distribution of the solution is given in Fig.1. The level number k in the switching function is set to 2.

We stop the global ASPIN iterations if

$$\|\mathcal{F}(u^{(k)})\| \leq 10^{-10}\|\mathcal{F}(u^{(0)})\|.$$

The global linear iteration for solving the global Jacobian system is stopped if

$$\|\mathcal{F}(u^{(k)}) - \mathcal{F}'(u^{(k)})p^{(k)}\| \leq 10^{-3}\|\mathcal{F}(u^{(k)})\|.$$

At the kth global nonlinear iteration, nonlinear subsystems

$$F_{S_i}(g_i^{(k)}) = 0,$$

have to be solved. We use the standard inexact Newton with a cubic line search for such systems with initial guess $g_{i,0}^{(k)} = 0$. The local nonlinear iteration in subdomain S_i is stopped if $\|F_{S_i}(g_{i,l}^{(k)})\| \leq 10^{-2}\|F_{S_i}(g_{i,0}^{(k)})\|$.

For comparison purposes, we first solve the problem using the regular inexact Newton's method. The Jacobian problems are solved with GMRES, and the nonlinear residual history are shown in Fig.2 for two mesh sizes $h = 1/128$ and $h = 1/256$. It can be seen clearly the convergence degenerates as the mesh is refined. In general, The finer the mesh, the longer the plateau period lasts. This happens no matter how accurately one solves the Jacobian problems. We next solve the same discrete nonlinear systems using ASPIN. We use 8 subdomains with the overlapping size equals to $5h$. The numbers of ASPIN iterations are shown in Fig.1. The iteration numbers are much smaller than that of the regular inexact Newton's method (Fig.2), and the nonlinear iteration numbers do not change that much as we refine the mesh from $h = 1/128$ to $h = 1/256$ to get a better resolution of the shock wave.

References

[BGM+01]S. Balay, W. Gropp, L. McInnes, and B. Smith. *The Portable, Extensible Toolkit for Scientific Computing, version 2.1.0*. `http://www.mcs.anl.gov/petsc`, code and documentation, 2001.

[BBH+93]M. Bieterman, J. Bussoletti, C. Hilmes, F. Johnson, R. Melvin, and D. Young. Second order upwinding for full potential aerodynamics problems. BCSTECH-93-013, The Boeing Company, 1993.

[BS90]P. N. Brown and Y. Saad. Hybrid Krylov methods for nonlinear systems of equations. *SIAM J. Sci. Stat. Comput.*, 11 (1990), pp. 59–71.

[BS94]P. N. Brown and Y. Saad. Convergence theory of nonlinear Newton-Krylov algorithms. *SIAM J. Optimization*, 4 (1994), pp. 297–330.

[CD94]X.-C. Cai and M. Dryja. Domain decomposition methods for monotone nonlinear elliptic problems. *Contemporary Math.*, 180 (1994), pp. 21–27.

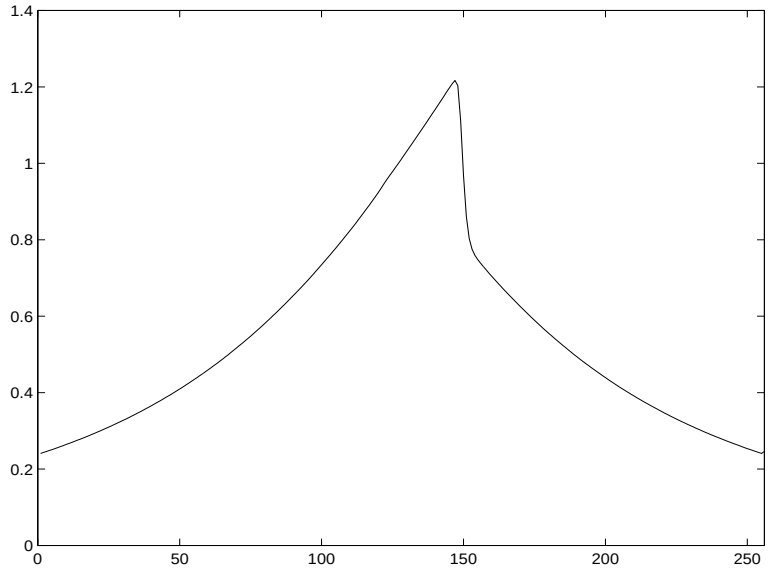


Figure 1: Mach distribution and the shock location.

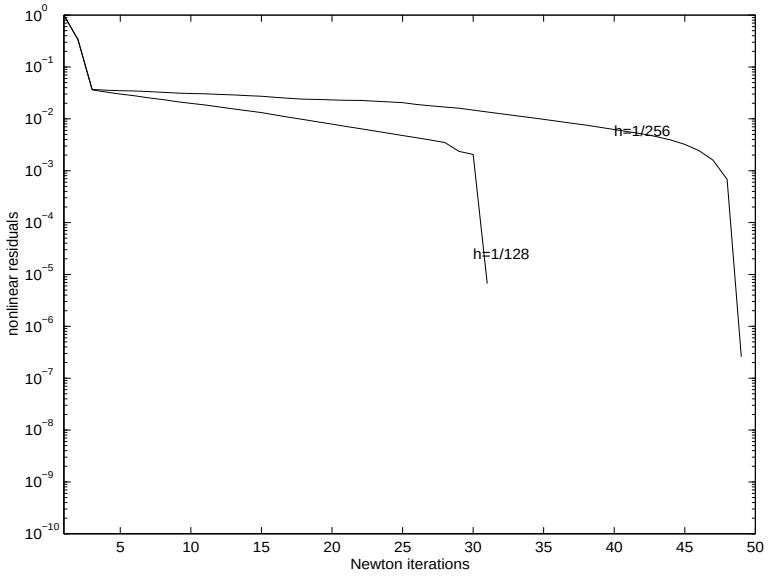


Figure 2: Nonlinear residual history of the inexact Newton's algorithm for the flow problem with mesh sizes $h = 1/128$ and $h = 1/256$.

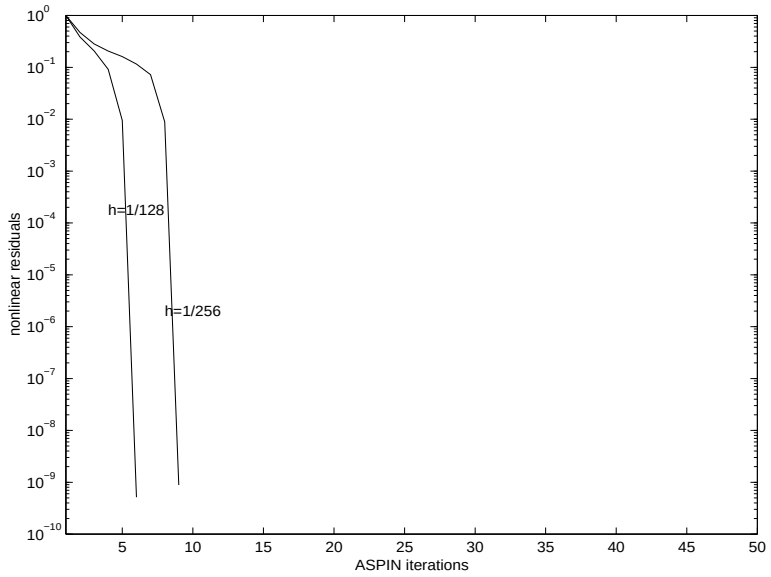


Figure 3: Nonlinear residual history of the additive Schwarz preconditioned inexact Newton's algorithm for the flow problem with mesh sizes $h = 1/128$ and $h = 1/256$.

[CGK+98]X.-C. Cai, W. D. Gropp, D. E. Keyes, R. G. Melvin, and D. P. Young. Parallel Newton–Krylov–Schwarz algorithms for the transonic full potential equation. *SIAM J. Sci. Comput.*, 19 (1998), pp. 246–265.

[CK00]X.-C. Cai and D. E. Keyes. Nonlinearly preconditioned inexact Newton algorithms. *SIAM J. Sci. Comput.*, 2000. (submitted)

[DS83]J. E. Dennis and R. B. Schnabel. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, NJ, 1983.

[DES82]R. S. Dembo, S. C. Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM J. Numer. Anal.*, 19 (1982), pp. 400–408.

[DH97]M. Dryja and W. Hackbusch. On the nonlinear domain decomposition method. *BIT*, (1997), pp. 296-311.

[EW94]S. C. Eisenstat and H. F. Walker. Globally convergent inexact Newton methods. *SIAM J. Optimization*, 4 (1994), pp. 393–422.

[EW96]S. C. Eisenstat and H. F. Walker. Choosing the forcing terms in an inexact Newton method. *SIAM J. Sci. Comput.*, 17 (1996), pp. 16-32.

[GKM+00]W. D. Gropp, D. E. Keyes, L. C. McInnes and M. D. Tidriri. Globalized Newton-Krylov-Schwarz algorithms and software for parallel implicit CFD. *Int. J. High Performance Computing Applications*, 14 (2000), pp. 102-136.

[HMS78]M. Hafez, E. Murman, and J. South. Artificial compressibility methods for numerical solution of the transonic full potential equation. *AIAA Paper* 78-1148, 1978.

[JF95]H. Jiang and P. A. Forsyth. Robust linear and nonlinear strategies for solution of the transonic Euler equations. *Computer and Fluids*, 24 (1995), pp. 753-770.

[LRW96]P. J. Lanzkron, D. J. Rose, and J. T. Wilkes. An analysis of approximate nonlinear elimination. *SIAM J. Sci. Comput.*, 17 (1996), pp. 538–559.

[PCS+99]M. Paraschivoiu, X.-C. Cai, M. Sarkis, D. P. Young, and D. Keyes. Multi-domain multimodel formulation for compressible flows: Conservative interface coupling and parallel implicit solvers for 3D unstructured meshes. *AIAA Paper* 99-0784, 1999.

[PW98]M. Pernice and H. Walker. NITSOL: A Newton iterative solver for nonlinear systems. *SIAM J. Sci. Comput.*, 19 (1998), pp. 302–318.

[YMB+90]D. P. Young, R. G. Melvin, M. B. Bieterman, F. T. Johnson, and S. S. Samant. Global convergence of inexact Newton methods for transonic flow. *Int. J. Numer. Meths. Fluids*, 11 (1990), pp. 1075-1095.

[YMB+91]D. P. Young, R. G. Mervin, M. B. Bieterman, F. T. Johnson, S. S. Samant and J. E. Bussoletti. A locally refined rectangular grid finite element method: Application to computational fluid dynamics and computational physics. *J. Comput. Phys.*, 92 (1991), pp. 1-66.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

35 Fictitious domain based solvers for particulate flows

D. Dashevski[1], R. Glowinski[1], Yu. Kuznetsov[1], K. Lipnikov[1]

Introduction

In this article we discuss the application of fictitious domain methods to the numerical simulation of incompressible viscous flow with suspended moving particles. The model coupling the Navier-Stokes equations from fluid dynamics with the Newton equations for the particle motion has been extensively studied in the literature (see e.g. [GPH$^+$98, GPH$^+$00]). Among the problems for its practical application are fluidized beds, sedimention, a blood flow around artificial heart valve, etc.

The solution method discussed here combines finite element discretizations in space, time discretization by a projection scheme and the method of characteristics [GP92] for the treatment of the convection term. The key points of our method are locally refined locally adapted grids for space discretization and efficient iterative solvers based on fictitious domain methods. The methodologies we follow in this paper were proposed and studied in [Ast78, GK98, MKM86]. We shall show in Section 4, that the concrete choice of the optimal domain embedding is strongly governed by the computational domain topology. Therefore, we focus in our research on simulations with a few solid particles to investigate in details the behavior of iterative solvers for the case of particle collisions.

Formulation of the particulate flow problem

Let $B(t)$, $t > 0$, be the union of a few solid particles suspended in an incompressible viscous fluid occupying the fixed domain Π. The fluid velocity $\boldsymbol{u}$ and pressure p are solutions of the Navier-Stokes equations

$$\begin{aligned} \frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} - \nu\Delta\boldsymbol{u} + \frac{1}{\rho}\nabla p &= \boldsymbol{g} \quad \text{in } \Omega(t), \\ \nabla\cdot\boldsymbol{u} &= 0 \quad \text{in } \Omega(t), \end{aligned} \tag{1}$$

with initial and boundary conditions

$$\begin{aligned} \boldsymbol{u} &= \boldsymbol{u}_0 \quad \text{in } \Omega(0), \\ \boldsymbol{u} &= \boldsymbol{g}_0 \quad \text{on } \partial\Omega(t). \end{aligned} \tag{2}$$

Here $\Omega(t) = \Pi \setminus \bar{B}(t)$ is the domain occupied by the fluid, ρ is the fluid density, ν is the kinematic viscosity and $\boldsymbol{g}$ is the gravity. Without loss of generalization, we assume that the fluid-particle system is at rest at $t = 0$, i.e. $\boldsymbol{u}_0 = \boldsymbol{0}$ and $\boldsymbol{g}_0 = \boldsymbol{0}$. For $t > 0$, the particle motion

[1]Department of Mathematics, University of Houston, Houston, TX, 77204, USA, e-mail: ddl@math.uh.edu, roland@math.uh.edu, kuz@math.uh.edu, lipnikov@math.uh.edu

satisfies Newton's law:

$$\begin{aligned} m_i \frac{d\boldsymbol{V}_i}{dt} &= \int\limits_{\partial B_i(t)} \boldsymbol{\sigma}\cdot\boldsymbol{n}\,ds + m_i\boldsymbol{g}, \\ \frac{d}{dt}(\boldsymbol{I}_i\boldsymbol{\omega}_i) &= \int\limits_{\partial B_i(t)} (\boldsymbol{x}-\boldsymbol{O}_i)\times(\boldsymbol{\sigma}\cdot\boldsymbol{n})\,ds, \qquad i=1,\ldots,N, \end{aligned} \tag{3}$$

where $\boldsymbol{V}_i$, $\boldsymbol{\omega}_i$, m_i, $\boldsymbol{I}_i$ and $\boldsymbol{O}_i$ are translational and angular velocities, mass, inertia tensor and centroid of the i-th particle, respectively, $\boldsymbol{\sigma} = \nu\rho(\nabla\boldsymbol{u}+\nabla^T\boldsymbol{u}) - p\boldsymbol{I}$ is the stress tensor and $\boldsymbol{n}$ is the unit normal vector on the particle boundary $\partial B_i(t)$ pointing outward. We assume that the no-slip boundary condition on $\partial B_i(t)$ holds, namely:

$$\boldsymbol{u}(\boldsymbol{x},t) = \boldsymbol{V}_i + \boldsymbol{\omega}_i\times(\boldsymbol{x}-\boldsymbol{O}_i), \qquad i=1,\ldots,N. \tag{4}$$

Problem approximation

We use the Galerkin finite element formulation for the space discretization and finite differences for the time discretization.

Time discretization scheme

Let Δt denote the time step, $t^n = n\Delta t$, $\Omega^n = \Omega(t^n)$, and $\boldsymbol{u}^n$, p^n, $\boldsymbol{V}_i^n$, $\boldsymbol{\omega}_i^n$, $i=1,\ldots,N$, be approximations of the continuous solution at time t^n. The discretization scheme for problem (1)-(4) includes five steps:

Convection step. For any $\boldsymbol{x}\in\Omega^n$, we compute the characteristics $\psi(\boldsymbol{x},t)$, $t\in[t^{n-1},t^n]$, ending at $\boldsymbol{x}$ and set $\boldsymbol{u}^{n+1/2}(\boldsymbol{x}) = \boldsymbol{u}(\psi(\boldsymbol{x},t^{n-1}))$.

We use a first order Runge Kutta integration scheme. In the case when the characteristic leaves the domain Ω^n, a special numerical procedure is used to estimate $\psi(\boldsymbol{x},t^{n-1})$.

Diffusion step. Using the convected field $\boldsymbol{u}^{n+1/2}$, we approximate the total time derivative by the first order implicit Euler scheme:

$$\begin{aligned} \frac{\tilde{\boldsymbol{u}}^{n+1}-\boldsymbol{u}^{n+1/2}}{\Delta t} - \nu\Delta\boldsymbol{u}^{n+1} + \frac{1}{\rho}\nabla p^n &= \boldsymbol{g} \quad \text{in } \Omega^n, \\ \tilde{\boldsymbol{u}}^{n+1} &= \boldsymbol{u}^n \quad \text{on } \partial\Omega^n. \end{aligned} \tag{5}$$

Projection step. The computed field $\tilde{\boldsymbol{u}}^{n+1}$ is projected onto a space of divergent-free functions by solving a Poisson equation for $p^{n+1}-p^n$:

$$\begin{aligned} \boldsymbol{u}^{n+1} &= \tilde{\boldsymbol{u}}^{n+1} - \frac{\Delta t}{\rho}\nabla(p^{n+1}-p^n) \quad \text{in } \Omega^n, \\ \nabla\cdot\boldsymbol{u}^{n+1} &= 0 \quad \text{in } \Omega^n, \end{aligned}$$

with the Dirichlet boundary condition $\boldsymbol{u}^{n+1} = \boldsymbol{u}^{n+1/2}$ on $\partial\Omega^n$.

Particle motion step. Using the computed solutions $\boldsymbol{u}^{n+1}$ and p^{n+1} and the first order Euler scheme, we discretize the motion equations:

$$\begin{aligned} m_i \frac{\boldsymbol{V}_i^{n+1} - \boldsymbol{V}_i^n}{\Delta t} &= \int\limits_{\partial B_i(t^n)} \boldsymbol{\sigma}^n \cdot \boldsymbol{n} ds + m_i \boldsymbol{g}, \\ \boldsymbol{I}_i \frac{\boldsymbol{\omega}_i^{n+1} - \boldsymbol{\omega}_i^n}{\Delta t} &= \int\limits_{\partial B_i(t^n)} (\boldsymbol{x} - \boldsymbol{O}_i^n) \times (\boldsymbol{\sigma}^n \cdot \boldsymbol{n}) ds \;-\; \boldsymbol{\omega}_i^n \times \boldsymbol{I}_i \boldsymbol{\omega}_i^n, \\ \boldsymbol{O}_i^{n+1} &= \boldsymbol{O}_i^n + \boldsymbol{V}_i^n \Delta t \end{aligned}$$

In the case of a few moving particles, a collision strategy based on the physics of solid bodies is used [GPH+00].

Interpolation step. The new particle positions define the domain Ω^{n+1} for the next time step. We use the finite element interpolation to compute the fluid velocity $\boldsymbol{u}^{n+1}$ and pressure p^{n+1} in the new domain.

Local adaptive locally fitted grids

The time-discretized problem is approximated by a finite element method. It is quite clear that the meshes used for space discretization are as important as the time discretization schemes and iterative solvers. Indeed the mesh determines the size of the algebraic problem and accuracy of the approximation. Taking this into account leads strongly to choose structured Locally Refined Locally Adapted (LRLA) meshes (see Figure 1).

A LRLA grid is built in three steps. First, a locally refined fully hierarchical grid is constructed in the domain Π to satisfy requirements on the mesh size imposed by the geometry and the discretization. Second, the locally refined mesh is adapted to the particle boundaries to provide the second order of the discretization. Finally, the LRLA grid Π_h is restricted to the computational domain Ω.

The hierarchical structure of the grid allows the use of advanced preconditioners like multigrid methods and provides natural tree data structure which can be used for effective implementation of the interpolation step.

Space discretization

Let Ω_h^n be a triangulation of the domain Ω^n. A triangulation $\Omega_{h/2}^n$ is obtained from Ω_h^n by one level of uniform refinement, i.e. by splitting every tetrahedron in Ω_h^n into 8 smaller tetrahedra. Let $V_p \subset H^1(\Omega_h^n)$ be the space of piecewise linear functions with a zero mean value defined on triangulation Ω_h^n. Similarly, let $V_u \subset [H^1(\Omega_{h/2}^n)]^3$ be a space of vector piecewise linear functions defined on triangulation $\Omega_{h/2}^n$. We denote by V_{0u} a subspace of V_u of functions vanishing on $\partial\Omega_h^n$.

The Galerkin finite element formulation of the Diffusion step (5) is to find $\tilde{\boldsymbol{u}}_h^{n+1} \in V_u$, $\tilde{\boldsymbol{u}}_h^{n+1} = \boldsymbol{u}_h^n$ on $\partial\Omega_h^n$, such that

$$\int\limits_{\Omega_h^n} \left(\frac{1}{\nu\Delta t}\tilde{\boldsymbol{u}}_h^{n+1} \cdot \boldsymbol{v}_h + \nabla\tilde{\boldsymbol{u}}_h^{n+1} : \nabla\boldsymbol{v}_h\right) dx = \int\limits_{\Omega_h^n} \boldsymbol{f} \cdot \boldsymbol{v}_h dx \qquad \forall \boldsymbol{v}_h \in V_{0u}, \tag{6}$$

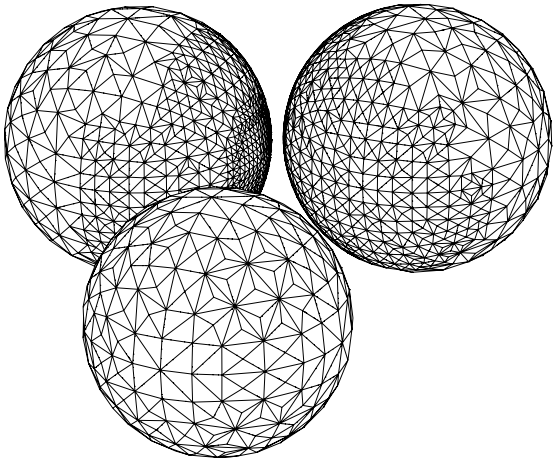


Figure 1: The trace of a LRLA grid on the particle boundaries

where $\boldsymbol{f} = \frac{1}{\nu}(\boldsymbol{g} + \frac{1}{\Delta t}\boldsymbol{u}_h^{n+1/2} - \frac{1}{\rho}\nabla p_h^n)$.

The weak formulation of the Projection step is to find a pair of functions $(\boldsymbol{u}_h^{n+1},\, p_h^{n+1}) \in V_u \times V_p$, $\boldsymbol{u}_h^{n+1} = \boldsymbol{u}_h^{n+1/2}$ on $\partial\Omega_h^n$, such that

$$\begin{aligned} &\int\limits_{\Omega_h^n} \boldsymbol{u}_h^{n+1} \cdot \boldsymbol{v}_h dx + \frac{\Delta t}{\rho} \int\limits_{\Omega_h^n} p_h^{n+1} \nabla \cdot \boldsymbol{v}_h dx = \int\limits_{\Omega_h^n} \boldsymbol{f} \cdot \boldsymbol{v}_h dx, \\ &\int\limits_{\Omega_h^n} q_h \nabla \cdot \boldsymbol{u}_h^{n+1} dx = 0 \qquad \forall (\boldsymbol{v}_h,\, q_h) \in V_{0u} \times V_p, \end{aligned} \tag{7}$$

where $\boldsymbol{f} = \tilde{\boldsymbol{u}}_h^{n+1} + \frac{\Delta t}{\rho}\nabla p_h^n$.

Fictitious domain method

Both problems (6) and (7) can be solved by fictitious domain methods (FDM). For simplicity of presentation we omit upper indices for unknown variables and computational domain, and assume that the flux satisfies the homogeneous Dirichlet boundary condition. Additionally, let the particles be spheres of the same radius R. We embed Ω into domain Π_δ in such a way that the triangulation Ω_h is a part of a triangulation $\Pi_{\delta h}$. Technically, both Ω_h and $\Pi_{\delta h}$ are traces of the fully hierarchical grid Π_h.

A concrete choice of the domain Π_δ depends on the topology of Ω. We assume that $\Omega \subset \Pi_\delta \subset \Pi$ and parameter δ characterizes the value of embedding. In other words it will be the thickness of a spherical layer $B_i(t^n) \cap \Pi_\delta$, i.e., $0 \le \delta \le R$. Thus $\Pi_0 = \Omega$ and $\Pi_R = \Pi$. Let $\gamma = \partial\Pi_\delta \setminus \partial\Pi$ be a part of the boundary $\partial\Pi_\delta$ living inside particles. Note that $\gamma = \emptyset$ when $\delta = R$.

FDM with distributed Lagrange multipliers

Following [GK98] we replace (6) by the equivalent saddle point problem with *distributed Lagrange multipliers*: find $(\boldsymbol{u}_h, \boldsymbol{\lambda}_h) \in W_u \times W_\lambda$:

$$\int_{\Pi_\delta} (\frac{1}{\nu\Delta t}\boldsymbol{u}_h \cdot \boldsymbol{v}_h + \nabla\boldsymbol{u}_h : \nabla\boldsymbol{v}_h)dx + \int_{\Pi_\delta \setminus \bar{\Omega}_h} (\frac{1}{\nu\Delta t}\boldsymbol{\lambda}_h \cdot \boldsymbol{v}_h + \nabla\boldsymbol{\lambda}_h : \nabla\boldsymbol{v}_h)dx = \int_{\Omega_h} \boldsymbol{f} \cdot \boldsymbol{v}_h dx,$$
$$\int_{\Pi_\delta \setminus \bar{\Omega}_h} (\frac{1}{\nu\Delta t}\boldsymbol{u}_h \cdot \boldsymbol{\mu}_h + \nabla\boldsymbol{u}_h : \nabla\boldsymbol{\mu}_h)dx = 0 \qquad \forall(\boldsymbol{v}_h, \boldsymbol{\mu}_h) \in W_u \times W_\lambda,$$

where $W_u \subset H^1(\Pi_\delta, \partial\Pi_\delta)$ and $W_\lambda \subset H^1(\Pi_\delta \setminus \bar{\Omega}_h, \gamma)$ are subspaces of piecewise linear functions vanishing on $\partial\Pi_\delta$ and γ, respectively. In algebraic form it reads:

$$\begin{pmatrix} A_u & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}. \tag{8}$$

By dividing the components of the solution u into separate groups denoted by u_1 and u_2 we obtain a useful block representation of the stiffness matrix. The subvectors u_1 and u_2 corresponds to the mesh nodes from the mesh domains Ω_h and $\Pi_{\delta h} \setminus \Omega_h$, respectively. By reordering the vector u according to this partitioning the linear system (8) can be written in the block form

$$\begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & B_2^T \\ 0 & B_2 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} f_1 \\ 0 \\ 0 \end{pmatrix}. \tag{9}$$

One of the main results of [GK98] is that the subvector u_1 is the solution to (6). Indeed, the matrix B_2 is symmetric and positive definite. Thus, eliminating $u_2 = 0$ from the first block equation, we end up with an algebraic problem equivalent to (6). It turns out that the linear system (9) can be solved much more easily than the reduced system. By introducing the new variable $\tilde{\lambda} = B_2\lambda$, we simplify (9):

$$A \begin{pmatrix} u_1 \\ u_2 \\ \tilde{\lambda} \end{pmatrix} \equiv \begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & I_2 \\ 0 & I_2 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \tilde{\lambda} \end{pmatrix} = \begin{pmatrix} f_1 \\ 0 \\ 0 \end{pmatrix}, \tag{10}$$

where I_2 is the identity matrix. This problem can be solved iteratively with a preconditioner H proposed in [GK98]:

$$H = \begin{pmatrix} H_{11} & H_{12} & 0 \\ H_{21} & H_{22} & 0 \\ 0 & 0 & B_2 \end{pmatrix} \equiv \begin{pmatrix} H_u & 0 \\ 0 & B_2 \end{pmatrix}.$$

Lemma 1 (Glowinski, Kuznetsov (1998)) *Let H_u be spectrally equivalent to A_u^{-1} ($H_u \sim A_u^{-1}$). Then $(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \sim B_2$ and $H \sim A^{-1}$.*

Note that A_u is a matrix form of the operator $-\Delta + \frac{1}{\nu\Delta t}$ in Ω_h with Dirichlet boundary conditions. One of the possible choices for H_u is a multigrid preconditioner, for example, the BPX preconditioner [BPX90]. It is most efficient when the grid is fully hierarchical. Recall that the construction of Ω_h is already based on a fully hierarchical grid Π_h. From this viewpoint we have to take $\Pi_\delta = \Pi$. On the other hand spectral properties of HA depend on δ and may be deteriorated for larger values of δ. The maximal embedding guaranteeing independence with respect to δ is determined by the particle radius R and the distance d between two neighboring particles. Let $|||\cdot|||_{1,\omega}$ be a norm given by

$$|||\boldsymbol{u}_h|||^2_{1,\omega} = \int_\omega (\frac{1}{\nu\Delta t}|\boldsymbol{u}_h|^2 + |\nabla \boldsymbol{u}_h|^2)\, dx.$$

Let $\boldsymbol{u}_{1h}$ and $\boldsymbol{u}_{2h}$ be finite element counterparts of subvectors u_1 and u_2, respectively. Given a function $\boldsymbol{u}_{2h}$ we define its norm-preserving extension $\boldsymbol{u}_{1h}$ in such a way that $\boldsymbol{u}_{1h} = \boldsymbol{u}_{2h}$ on γ. The following quantity plays the crucial role in the general theory of fictitious domain methods (see, for example, [MKM86]):

$$c = \max_{\boldsymbol{u}_{2h}\neq 0} \min_{\boldsymbol{u}_{1h}} \frac{|||\boldsymbol{u}_{1h}|||^2_{1,\Omega_h} + |||\boldsymbol{u}_{2h}|||^2_{1,\Pi_\delta\setminus\bar{\Omega}_h}}{|||\boldsymbol{u}_{2h}|||^2_{1,\Pi_\delta\setminus\bar{\Omega}_h}}. \tag{11}$$

Lemma 2 (Marchuk, Kuznetsov, Matsokin (1986)) *Let c be as above. Then* $\mathrm{cond}(H_u A_u) \lesssim c$.

Reasonable estimates of c can be obtained by analyzing collision of two particles. Let d be the distance between these particles.

Lemma 3 *The parameter c is independent of d, δ and R when either (a) $\sqrt{\nu\Delta t} \lesssim R$ and $\sqrt{\nu\Delta t} \lesssim d$ or (b) $\delta \lesssim d$.*

The proof is based on the norm preserving finite element extension theorem [Ast78, Nep91, Wid87] and scaling arguments.

An important corollary from Lemma 3 is that when particles are close to each other a severe restriction is imposed on δ. Therefore grid $\Pi_{\delta h}$ has only a few fully hierarchical levels. Fortunately, we are solving the singular perturbed elliptic problem for which the BPX preconditioner leads to a well conditioned coarse grid problem.

FDM for the Neumann boundary value problem

The saddle point problem (7) from the Projection step in the algebraic form reads:

$$A\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} M_u & B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}. \tag{12}$$

Consider a block diagonal matrix

$$H = \begin{pmatrix} H_u & 0 \\ 0 & H_p \end{pmatrix},$$

where

$$H_u \sim M_u^{-1} \quad \text{and} \quad H_p \sim (BM_u^{-1}B^T)^{-1}.$$

By using arguments similar to that in the proof of Lemma 1, we can show that the matrix H is spectrally equivalent to the stiffness matrix A. The simplest choice for H_u is the diagonal lumping of mass matrix M_u. The Schur complement $BM_u^{-1}B^T$ is spectrally equivalent to a discrete Laplace operator in Ω with Neumann boundary conditions. Such boundary conditions allow us to construct a very simple preconditioner proposed by Astrakhantsev [Ast78].

Let L be a matrix spectrally equivalent to the discrete Laplace operator on Π_h with Neumann boundary conditions. Let us divide mesh nodes into two separate groups. The first group includes nodes from the mesh domain Ω_h. The second group contains the remainder of the nodes. According to this partitioning matrix L can be written in the block form

$$L = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix}.$$

One of the main results of [Ast78] is that the Schur complement $S_{11} = L_{11} - L_{12}L_{22}^{-1}L_{21}$ is spectrally equivalent to the discrete Laplace operator in Ω_h with Neumann boundary conditions. Therefore we set $H_p = S_{11}^{-1}$. Obviously that action of H_p on a vector p is reduced to solving a linear system with matrix L and the right-hand side $(p,\ 0)^T$. One of the possible choices for L is a multigrid preconditioner, for example, the BPX preconditioner.

Numerical experiments

For the simulation we used 27 identical balls of radius $R = 0.01m$ centered at nodes of a 3x3x3 cubic grid with the mesh step size $5R/2$ (see Figure 2). The particles were placed in a parallelepiped with the square base $0.12m \times 0.12m$ and height $0.23m$ filled with a glycerin. We have chosen particles with density twice as large as the fluid density. The particle positions at times $t = 0s$, $t = 0.23s$ and $t = 0.31s$ are shown on Figure 2. The variable time step strategy was chosen to minimize the number of time steps. The total simulation required 64 time steps with about 520000 degrees of freedom for velocity and 65000 for the pressure.

A very interesting symmetric aggregation of particles in triples is observed after a few collisions between particles. Despite decreasing the distance between particles, we did not change the embedding domain Π. The number of iterations for solving the singular perturbed problem (6) with the BPX preconditioner has been changed from 18 to 30 when distance between particles has been changed from $5R/2$ to $R/5$ (the minimal allowed distance). We expect that the new embedding strategy described above will decrease the number of iterations and allow us a more detailed numerical analysis of particles collisions.

Aknowlegments. The author acknowledge the helpful discussions and suggestions of D. Joseph and T.-W. Pan. They also acknowledge the support of the NSF (grants DMS-9973318 and CCR-9902035).

References

[Ast78]G. P. Astrakhatsev. Method of fictitious domains for a second-order elliptic equation with natural boundary conditions. *U.S.S.R. Computational Math. and Math. Phys.*, 18:114–121, 1978.

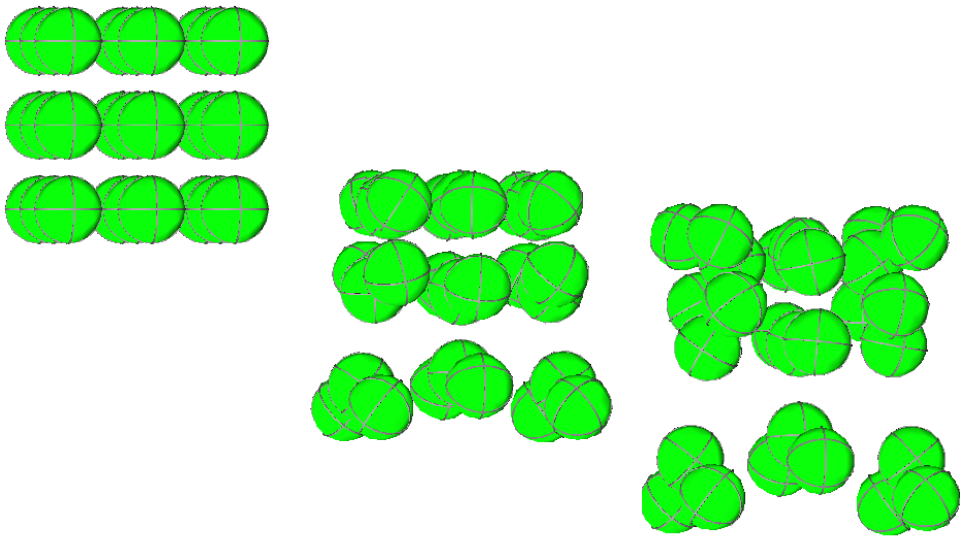


Figure 2: The motion of the cubic structure of spheres

[BPX90]James H. Bramble, Joseph E. Pasciak, and Jinchao Xu. Parallel multilevel preconditioners. *Math. Comp.*, 55:1–22, 1990.

[GK98]Roland Glowinski and Yuri Kuznetsov. On the solution of the Dirichlet problem for linear elliptic operators by a distributed Lagrange multiplier method. *C. R. Acad. Sci. Paris Sér. I Math.*, 327(7):693–698, 1998.

[GP92]Ronald Glowinski and Olivier Pironneau. Finite element methods for Navier-Stokes equations. *Annual Review of Fluid Mechanics*, 24:167–204, 1992.

[GPH+98]R. Glowinski, T.W. Pan, T.I. Hesla, D.D. Joseph, and J. Periaux. A fictitious domain method with distributed Lagrange multipliers for the numerical simulation of particulate flow. In J. Mandel, C. Farhat, , and X.C. Cai, editors, *Domain Decomposition Methods 10*, pages 121–137, Providence, RI, 1998. AMS.

[GPH+00]Ronald Glowinski, Tsorng-Whay Pan, Todd I. Hesla, Daniel D. Joseph, and Jacques Periaux. A distributed lagrange multiplier/fictitious domain method for the simulation of flows around moving rigid bodies: Application to particulate flow. *Comput. Meth. Appl. Mech. Engrg.*, 184:241–268, 2000.

[MKM86]G.I. Marchuk, Y.A. Kuznetsov, and A.M. Matsokin. Fictitious domain and domain decomposition methods. *Sov.J.Numer.Anal.Math.Modelling*, 1(1):3–35, 1986.

[Nep91]Sergey V. Nepomnyaschikh. Mesh theorems of traces, normalizations of function traces and their inversions. *Sov. J. Numer. Anal. Math. Modeling*, 6:1–25, 1991.

[Wid87]Olof B. Widlund. An extension theorem for finite element spaces with three applications. In Wolfgang Hackbusch and Kristian Witsch, editors, *Numerical Techniques in Continuum Mechanics*, pages 110–122, Braunschweig/Wiesbaden, 1987. Notes on Numerical Fluid Mechanics, v. 16, Friedr. Vieweg und Sohn. Proceedings of the Second GAMM-Seminar, Kiel, January, 1986.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

36 Scalabilities of FETI for variational inequalities and contact shape optimization

Zdeněk Dostál, David Horák, Jan Szweda and Vít Vondrák [1]

Introduction

We review our work on development of an efficient algorithm for numerical solution of variational inequalities and their application to the solution of multi-body contact shape optimization problems solved by the gradient methods. The method presented exploits optimal features of the linear FETI domain decomposition method with the natural coarse grid and a special structure of quadratic programming problems arising in dual formulation of the state problem. Results of numerical experiments are reported that document both numerical and parallel scalability of the algorithm for the solution of a model variational inequality and illustrate its efficiency in the solution of a contact shape optimization problem with the semi-analytic sensitivity analysis.

Following [DFS98, DGS00a, DGS00b], we start our exposition by describing the discretized variational inequality as a convex quadratic programming (QP) problem with a block diagonal stiffness matrix and general equality and inequality constraints. Then we show that the difficulties arising from general inequality constraints and possible semi-definiteness can be essentially reduced by the application of the duality theory. The matrix of the dual quadratic form turns out to be positive definite with a spectrum that is more favorably distributed for application of the conjugate gradient based methods than its primal counterpart. The performance of the method can be further improved by means of the natural coarse space projectors[FMR94]. The algorithm and the corresponding theoretical results are then reviewed in Section 36.

In Section 36, we show that the algorithm complies well with the semi-analytic method [HN96, DVR01] for evaluation of the gradients of the cost function that are necessary for implementation of the feasible direction method. In particular, it turns out that the gradient may be evaluated with only one decomposition of the stiffness matrix, regardless of the number of the design variables.

The algorithm has been implemented by means of PETSc [BGMS97] package on SP2 for the solution of a model problem. The results of numerical experiments indicate both numerical and parallel scalability of the algorithm. For solution of 2D contact and contact shape optimization problems, the algorithm has been implemented into the system ODESSY [RKO91] developed at the Institute of Mechanical Engineering of the Aalborg University. Reported numerical experiments indicate again high performance of the algorithm in the solution of the contact shape optimization problems. Let us recall that interesting results concerning numerical scalabity of a different algorithm for variational inequalities can be found in Schöberl [Sch98].

[1] VŠB-Technical University Ostrava, 17. listopadu, CZ-708 33 Ostrava-Poruba, Czech Republic, zdenek.dostal@vsb.cz, david.horak@vsb.cz, jan.szweda@vsb.cz, vit.vondrak@vsb.cz

Discretized variational inequality and duality

Let $I\!K$ denote a closed convex subset of a Sobolev space $I\!V$ defined on a domain Ω in $I\!R^d$, $d = 2, 3$ with sufficiently smooth boundary Γ, and consider a problem to find $u \in I\!K$ so that

$$a(u, v - u) \geq b(v - u) \text{ for all } v \in I\!K, \tag{1}$$

where a and b are a symmetric positive semidefinite bilinear form and a linear functional, respectively. We restrict our attention to problems (1) arising from discretization of free boundary elliptic problems [Glo83] with a spatial domain Ω comprising subdomains $\Omega^1, \ldots, \Omega^s$. An important special case is a problem to find an equilibrium of a system of elastic bodies in contact, possibly with auxiliary domain decomposition [DGS00b].

The finite element discretization of $\Omega = \Omega^1 \cup \ldots \cup \Omega^s$ with a suitable numbering of nodes results in the QP problem to find

$$\min \frac{1}{2} u^T K u - f^T u \text{ subject to } N_1 u \leq c_1, \quad N_2 u = c_2 \tag{2}$$

with a symmetric block-diagonal matrix $K = \text{diag}(K_1, \ldots, K_s)$ of order n, $f \in I\!R^n$, an $m \times n$, $m \leq n$ full rank matrix N comprising blocks N_1 and N_2, and similarly $c \in I\!R^m$ comprising subvectors c_1 and c_2. The diagonal blocks K_p that correspond to the subdomains $\Omega^p, p = 1, \ldots, s$ are positive definite or semidefinite sparse matrices. Moreover, we assume that the nodes are numbered in such a way that $K_1, \ldots, K_s$ are banded matrices that can be effectively decomposed by the Cholesky factorization. If a contact problem of elasticity is considered, then the vector f describes the nodal forces arising from the volume forces or some other tractions, the matrix N_1 and the vector c_1 describe the linearized incremental non-interpenetration conditions, and the matrix N_2 with $c_2 = 0$ describe the "gluing" conditions on auxiliary interfaces. More details may be found in [DFS98].

Even though (2) is a standard convex QP problem, its numerical solution may be expensive. The reasons are that K is typically ill-conditioned or singular, and that the feasible set is so complex that projections onto it can hardly be effectively computed, so that it would be very difficult to achieve fast identification of the contact interface and fast solution of auxiliary linear problems. These complications may be essentially reduced by applying the duality theory of convex programming (e.g. [Dos95, DFS98]).

Following [DGS00a, DGS00b], let us first assume that the matrix K has a nontrivial null space that may be used to define the natural coarse grid [FMR94]. The Lagrangian associated with problem (2) is

$$L(u, \lambda) = \frac{1}{2} u^T K u - f^T u + \lambda^T (N u - c), \tag{3}$$

where the vector of multipliers comprises subvectors λ_1, λ_2 that comply with the block structure of N, so that we can rewrite the problem (2) as the saddle point problem

$$\text{Find} \quad (\hat{u}, \hat{\lambda}) \quad \text{such that} \quad L(\hat{u}, \hat{\lambda}) = \sup_{\lambda_1 \geq 0} \inf_u L(u, \lambda). \tag{4}$$

If we eliminate u from (4), we shall get the minimization problem

$$\min \Theta(\lambda) \quad \text{s.t.} \quad \lambda_1 \geq 0 \quad \text{and} \quad R^T (f - N^T \lambda) = 0, \tag{5}$$

where R denotes a matrix whose columns span the null space of K, $K^\dagger$ denotes any matrix that satisfies $KK^\dagger K = K$, and

$$\Theta(\lambda) = \frac{1}{2}\lambda^T N K^\dagger N^T \lambda - \lambda^T (N K^\dagger f - c). \tag{6}$$

Once the solution $\hat{\lambda}$ of (5) is obtained, the vector u that solves (4) can be evaluated by means of explicit formulas that may be found in [Dos95, DFS98]. The Hessian of Θ is closely related to that of the basic FETI method by Farhat and Roux, so that its spectrum is relatively favorably distributed for application of the conjugate gradient method.

Even though problem (5) is much more suitable for computations than (2) and has been used for efficient solution of contact problems [DFS98], further improvement may be achieved by the natural coarse grid projectors of Farhat, Mandel and Roux [FMR94]. In this way, it is even possible to achieve that the effective spectral condition number of the Hessian of the Lagrangian involved in computations is bounded independently of both the penalty parameter and the number of subdomains [DGS00b]. It does not follow that the resulting algorithm is scalable as it is still necessary to find the active constraints of the solution.

If the stiffness matrix K is regular, than the same procedure leads to the dual problem

$$\min \Theta(\lambda) \quad \text{s.t.} \quad \lambda \geq 0. \tag{7}$$

Algorithm

The problem (7) comprises only bound constraints, so that efficient algorithms using projections and adaptive precision control [Dos97] may be used. To apply this algorithm also for the problem (5), we shall use a variant of the augmented Lagrangian type algorithm proposed by Conn, Gould and Toint [CGT91] for identification of stationary points for more general problems. However, the algorithm that we describe here is modified in order to exploit the specific structure of our problem. Main improvement is in a sense adaptive precision control in Step 1.

To simplify our notation, let us denote $F = NK^\dagger N^T$, $G = R^T N^T$, and $d = R^T f$, and let us introduce the augmented Lagrangian with the penalization parameter ρ and the multiplier μ for the equality constraints for problem (5) by

$$L(\lambda, \mu, \rho) = \frac{1}{2}\lambda^T F \lambda - \lambda^T f + \mu^T (G\lambda - d) + \frac{1}{2}\rho ||G\lambda - d||^2.$$

If we denote by $g = g(\lambda, \mu, \rho)$ the gradient of L with respect to λ, then the *projected gradient* $g^P = g^P(\lambda, \mu, \rho)$ of L at λ is given component-wise by

$$g_i^P = g_i \text{ for } \lambda_i > 0 \text{ or } i \notin I \text{ and } g_i^P = g_i^- \text{ for } \lambda_i = 0 \text{ and } i \in I$$

with $g_i^- = \min(g_i, 0)$, where I is the set of indices of constrained entries of λ.

All the parameters that must be defined prior to the application of the algorithm are listed in Step 0.

Algorithm 3.1. (Simple bound and equality constraints)

Step 0. Initialization of parameters

Set $0 < \alpha < 1,\ 1 < \beta,\ \rho_0 > 0,\ \eta_0 > 0,\ M > 0,\ \mu^0,\ \lambda^0$.

Step 1. Find λ^k so that $||g^P(\lambda^k, \mu^k, \rho_k)|| \leq M||G\lambda^k||$.

Step 2. If $||g^P(\lambda^k, \mu^k, \rho_k)||$ and $||G\lambda^k||$ are sufficiently small, then stop.

Step 3. $\mu^{k+1} = \mu^k + \rho_k G\lambda^k$

Step 4. $If ||G\lambda^k|| \leq \eta_k$

Step 4a. then $\rho_{k+1} = \rho_k, \eta_{k+1} = \alpha\eta_k$

Step 4b. else $\rho_{k+1} = \beta\rho_k, \eta_{k+1} = \eta_k$

end if.

Step 5. Increase k and return to Step 1.

An implementation of Step 1 is carried out by the minimization of the augmented Lagrangian L subject to $\lambda \geq 0$ by an efficient algorithm that can be found in [Dos97]. The proposed algorithm has been proved [DFS01] to converge for any set of parameters that satisfy the prescribed relations. Moreover, an estimate of the rate of convergence of the approximations of the Lagrange multipliers has been proved that does not have any term that accounts for inexact solution of the bound constrained problems that are solved in Step 1, and it was proved that the penalty parameter is uniformly bounded. These results give theoretical support to Algorithm 3.1.

Discretized contact shape optimization problem

Let us now consider a contact shape optimization problem assuming for simplicity that the bodies occupy in a reference configuration subdomains $\Omega^1, \ldots, \Omega^s$ and that the shape of the first region Ω^1 depends on a vector of design variables α, so that the energy functional will have the form

$$j(u, \alpha) = \frac{1}{2}u^T K(\alpha)u - f^T(\alpha)u, \tag{8}$$

where the stiffness matrix $K(\alpha)$ and possibly the vector of nodal forces $f(\alpha)$ depend on α. The matrix N and the vector c now describe the linearized incremental conditions of non-interpenetration so that they also depend on α and the solution $u(\alpha)$ of the contact problem with the region $\Omega^1 = \Omega^1(\alpha)$ satisfies

$$u(\alpha) = \arg\min\{j(u, \alpha) : u \in C(\alpha)\}, \tag{9}$$

where

$$C(\alpha) = \{u : N(\alpha)u \leq c(\alpha)\}.$$

We shall consider the contact shape optimization problem to find

$$\min\{J(\alpha) : \alpha \in D_{adm}\}, \tag{10}$$

where $J(\alpha)$ is the cost functional that defines the cost function for design of body $\Omega^1(\alpha)$. The set of admissible design variables D_{adm} defines all feasible designs. For example, we can consider the cost functional $J(\alpha) \equiv -j(u,\alpha)$ that defines the minimal compliance problem. The set of admissible design parameters will be defined by

$$D_{adm} = \{-l \leq \alpha \leq r : \text{vol}(\Omega(\alpha)) \leq \text{vol}(\Omega(0))\}, \tag{11}$$

where l, r are given vectors with non-negative entries that define bounds on the design variables, and vol(.) is a mapping that assigns to each domain its volume. It has been proved that the minimal compliance problem has at least one solution and that the functional $j(u,\alpha)$ considered as a function of α is differentiable under reasonable assumptions [HN96].

If we want to exploit differentiability of problem (10), we must evaluate effectively partial derivatives of u with respect to the design variables $\alpha_1, \ldots, \alpha_k$. Our experience shows that the semi-analytic sensitivity analysis [HN96] is a method of choice. Let us denote by $I = \{1, ..., m\}$ the set of indices of the Lagrange multipliers λ, $I_s = \{i \in I : N_{ij}(\alpha)u_j(\alpha) = d_i(\alpha) \wedge \lambda_i(\alpha) > 0\}$ the set of indices that correspond to couples of nodes in strong contact, and $I_w = \{i \in I : N_{ij}(\alpha)u_j(\alpha) = d_i(\alpha) \wedge \lambda_i(\alpha) = 0\}$ the set of indices that correspond to couples of nodes in weak contact. We have used the standard summing convention. Analysis of the Karush-Kuhn-Tucker conditions [HN96] enables to evaluate the directional derivative $u'(\alpha,\beta)$ in the direction β by solving the quadratic programming problem

$$\min_{\substack{N_w(\alpha)z \leq d_w(\alpha,\beta) \\ N_s(\alpha)z = d_s(\alpha,\beta)}} \frac{1}{2} z^T K(\alpha) z - z^T (f'(\alpha,\beta) - K'(\alpha,\beta)u(\alpha) + N'^T(\alpha,\beta)\lambda(\alpha)), \tag{12}$$

where $K'(\alpha,\beta)$, $f'(\alpha,\beta)$ and $N'(\alpha,\beta)$ denote computable directional derivatives of the stiffness matrix, traction vector and the constraint matrix, respectively. Matrices $N_w(\alpha)$ and $N_s(\alpha)$ are formed by the rows of the matrix $N(\alpha)$ with the indices that belong to I_w and I_s, respectively. Similarly, the vectors $d_w(\alpha,\beta)$ and $d_s(\alpha,\beta)$ are formed by the entries of $d'(\alpha,\beta) - N'(\alpha,\beta)u(\alpha)$ with indices in I_w and I_s, respectively. Solving (12) for $\beta = e_i,\ i = 1, ..., m$, where e_i are the standard unit vectors, we evaluate the gradient of the state problem. Denoting $\bar{f}(\alpha,\beta) = f'(\alpha,\beta) - K'(\alpha,\beta)u(\alpha) + N'^T(\alpha,\beta)\lambda(\alpha)$, we can see that the problem (12) has the same structure as the problem (2), so that we can rewrite (12) into the dual form.

It turns out that the semi-analytic sensitivity analysis based on the dual formulation requires only one assembly and decomposition of the stiffness matrix. More information may be found in [HN96, VDR99, DVR01].

Numerical experiments

We have tested our algorithm on the solution of a simple model problem

$$\begin{aligned} &\text{Minimize} \quad q(u_1,u_2) = \sum_{i=1}^{2} \left(\int_{\Omega^i} |\nabla u_i|^2 d\Omega - \int_{\Omega^i} f u_i d\Omega \right) \\ &\text{subject to} \quad u_1(0,y) \equiv 0 \text{ and } u_1(1,y) \leq u_2(1,y) \text{ for } y \in [0,1], \end{aligned}$$

where $\Omega^1 = (0,1)\times(0,1)$, $\Omega^2 = (1,2)\times(0,1)$, $f(x,y) = -3$ for $(x,y) \in (0,1)\times[0.75,1)$, $f(x,y) = 0$ for $(x,y) \in (0,1)\times(0,0.75)$, $f(x,y) = -1$ for $(x,y) \in (1,2)\times(0,0.25)$ and

$f(x,y) = 0$ for $(x,y) \in (1,2) \times (0.25,1)$. This problem is semicoercive due to the lack of Dirichlet data on the boundary of Ω^2.

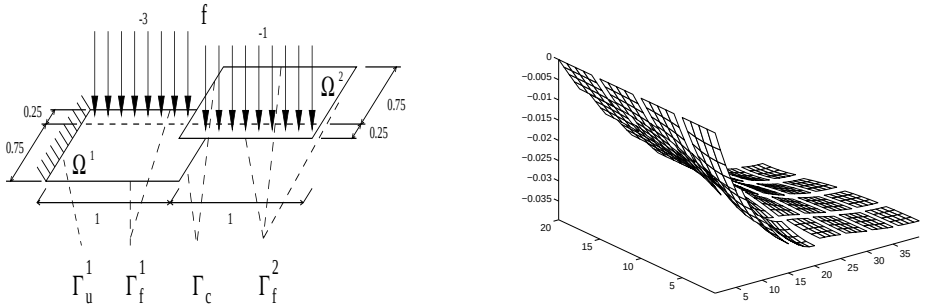
The solution of our model problem may be interpreted as the displacement of two membranes under the traction f. The left membrane is fixed on the left and the left edge of the right membrane is not allowed to penetrate below the edge of the left membrane as indicated in Figure 1a. The solution is unique because the right membrane is pressed down. More details about this model problem including some other results may be found in [DGS00a].

The model problem was discretized by regular grids defined by the stepsize $h = 1/n$ with $n+1$ nodes in each direction per subdomain Ω^i, $i = 1,2$. Each subdomain Ω^i was decomposed into $N \times N$ identical rectangles with dimensions $H = 1/N$. The solution of the model problem discretized by $H = 1/4$ and $h = 1/16$ can be seen in Figure 1b.

The model problem was solved for $h \in \{1/64, 1/128, 1/256, 1/512\}$, $H/h = 64$ with the stopping criterion

$$||g^P(\lambda,\mu,0)|| \le 10^{-4}||Nf|| \text{ and } ||G\lambda|| \le 10^{-4}||d||.$$

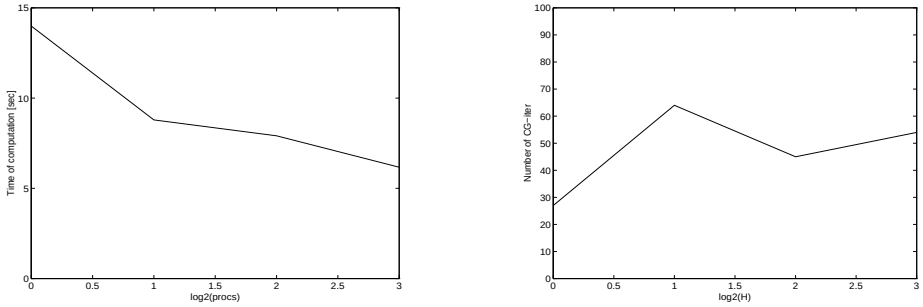
Both numerical and parallel scalabilities are demonstrated in Figure 2. Figure 2a demonstrates the dependence of elapsed time on the number of processors. Let us point out that the times were effected by the order and variety of used processors. Figure 2b then demonstrates high degree of numerical scalability of our algorithm for variational inequalities. In particular, the number of the conjugate gradient iterations ranged from 27 to 65 with only 54 iterations for the largest problem. The primal dimension ranged from 8450 to 540800. To solve the problem to the prescribed precision, it was necessary to identify about 350 active constraints on the contact interface comprising 520 couples of nodes that might have come into contact. The dual dimension was 14975.



(a) (b)

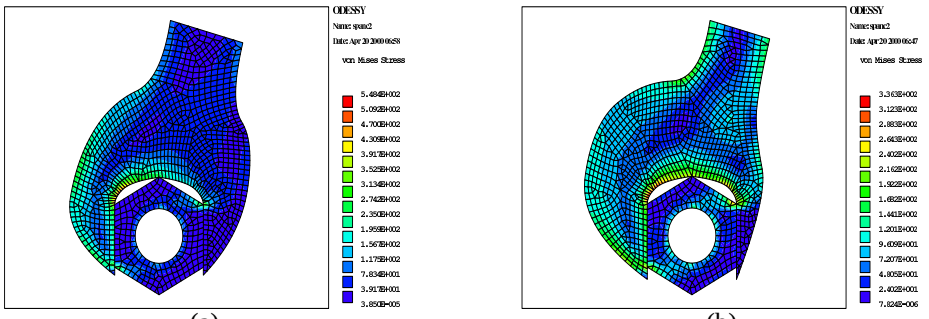
Figure 1: Model problem and its solution $h = 1/16, H = 1/4$

We have also tested our algorithm on the solution of a problem to find a shape of the spanner in Figure 3a that minimizes the maximum of von Mizes stress. To this end, we have implemented our algorithm into the system ODESSY developed at the Institute of Mechanical Engineering of the Aalborg University [RKO91]. The problem has been discretized by the finite element method using 2606 degrees of freedom with 46 couples of nodes that may get in contact. The admissible shape of the spanner was restricted by the box constraints on the design variables and by the upper bound on the volume. The initial and optimized designs are displayed in Figures 3a and 3b together with the values of the cost function. To get the results, we carried out 79 design steps.



(a) $h = 1/128, H = 1/4$ (b) $H/h = 64$

Figure 2: Parallel and numerical scalabilities



(a) (b)

Figure 3: Initial and optimized shape of the spanner

For comparison, we attempted to solve the problem also by the commercial software ANSYS. It turned out that the implementation of our algorithm in ODESSY was considerably more efficient. The analysis step in ODESSY required only 13 seconds, while it required 12 minutes to get a comparable result by ANSYS on the same computer. We were not able to carry out the optimization in ANSYS.

Comments and conclusions

The FETI-based domain decomposition algorithms for the solution of coercive and semicoercive variational inequalities has been reviewed and tested. Presented results of solution of a model variational inequality indicate both numerical and parallel scalability of the algorithm. Development of the theory is in progress. Theoretical results published so far [DGS00b] guarantee the convergence and robustness of the method. The method has been applied to optimization of a spanner and the efficiency of the method has been confirmed also by comparison with the commercial software. The salient feature of the algorithm in contact shape optimization is the reduction in the costs in preparing domain decomposition based solutions for related QP problems that appear in the dual formulation of the sensitivity analysis. In particular, it turns out that for each design step, it is necessary to carry out the preparation step only once regardless the number of the design variables. Further improvement may be achieved by the application of the mixed finite element discretization [DHK00, WK01].

Acknowledgements

This research has been supported by the grants GA ČR 101/01/0538 and 105/99/129 and by the project CEZ J:17/98:272400019 supported by the Ministry of Education of the Czech Republic.

References

[BGMS97]Satish Balay, William D. Gropp, Louis C. McInnes, and Barry F. Smith. *PETSc 2.0 User Manual*. Argonne National Laboratory, http://www.mcs.anl.gov/petsc/, 1997.

[CGT91]Andrew R. Conn, Nicholas I. M. Gould, and Philipe L. Toint. A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM J. Num. Anal.*, 28:545–572, 1991.

[DFS98]Zdeněk Dostál, Ana Friedlander, and Sandra A. Santos. Solution of coercive and semicoercive contact problems by FETI domain decomposition. *Contemporary Math.*, 218:82–93, 1998.

[DFS01]Zdeněk Dostál, Ana Friedlander, and Sandra A. Santos. Augmented lagrangians with adaptive precision control for quadratic programming with simple bounds and equality constraints. *SIAM J. Opt.*, 2001. submitted.

[DGS00a]Zdeněk Dostál, Francisco A. M. Gomes, and Sandra A. Santos. Duality based domain decomposition with natural coarse space for variational inequalities. *J. Comput. Appl. Math.*, 126:397–415, 2000.

[DGS00b]Zdeněk Dostál, Francisco A. M. Gomes, and Sandra A. Santos. Solution of contact problems by FETI domain decomposition. *Computer Meth. Appl. Mech. Engng.*, 190:1611–1627, 2000.

[DHK00]Zdeněk Dostál, Jaroslav Haslinger, and Radek Kučera. Implementation of fixed point method for duality based solution of contact problems with friction. *Int. J.Comput. Appl. Math.*, 2000. submitted.

[Dos95]Zdeněk Dostál. Duality based domain decomposition with proportioning for the solution of free boundary problems. *J. Comp. Appl. Math.*, 63:203–208, 1995.

[Dos97]Zdeněk Dostál. Box constrained quadratic programming with proportioning and projections. *SIAM J. Opt.*, 7:871–887, 1997.

[DVR01]Zdeněk Dostál, Vít Vondrák, and John Rasmussen. Efficient algorithms for contact shape optimization. In V. Schulz, editor, *Workshop Fast Solution of Discretized Optimization Problems*, pages 98–106, WIAS Berlin, 2001.

[FMR94]Charbel Farhat, Jan Mandel, and Francois-Xavier Roux. Optimal convergence properties of the FETI domain decomposition method. *Comput. Methods Appl. Mech. Engrg.*, 115:367–388, 1994.

[Glo83]Roland Glowinski. *Numerical Methods for Nonlinear Variational Problems*. Springer Verlag, New York, 1983.

[HN96]Jaroslav Haslinger and Pekka Neittaanmäki. *Finite element approximation for optimal shape, material and topology design*. John Wiley and Sons, London, 1996.

[RKO91]John Rasmussen, Oluf Krogh, and Niels Olhoff. *ODESSY System for Optimization*. IME Aalborg University, 1991.

[Sch98]Joachim Schöberl. Solving the signorini problem on the basis of domain decomposition techniques. *Computing*, 60:323–344, 1998.

[VDR99]Vít Vondrák, Zdeněk Dostál, and John Rasmussen. FETI domain decomposition algorithms for sensitivity analysis in contact shape optimization. In Choi-Hong Lai, Petter E. Bjørstad, Mark Cross, and Olof B. Widlund, editors, *11th International Conference on Domain Decomposition in Science and Engineering*, pages 561–567, Bergen, 1999. Domain Decomposition Press.

[WK01]Barbara Wohlmuth and Rolf Krause. Multigrid methods based on the unconstrained product space arising from mortar finite element discretizations. *SIAM J. Numer. Anal.*, 39:192–213, 2001.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

37 An Algebraic Convergence Theory for Restricted Additive and Multiplicative Schwarz Methods

A. Frommer[1], R. Nabben[2], D. B. Szyld[3]

Introduction

In this contribution we use the algebraic representation recently developed for the classical additive and multiplicative Schwarz methods in [FS99, BFNS01] to analyze the *restrictive additive Schwarz* (RAS) and *restrictive multiplicative Schwarz* (RMS) methods; see [CS96, CFS98, CS99, QV99].

RAS was introduced in [CS99] as an efficient alternative to the classical additive Schwarz preconditioner. Practical experiments have proven RAS to be particularly attractive, because it reduces communication time while maintaining the most desirable properties of the classical Schwarz methods [CFS98, CS99]. RAS preconditioners are widely used in practice and are the default preconditioner in the PETSc software package [BGMS97]. Similar savings in communication time can be expected in the case of RMS; see [CS96]. In fact, we announce here that we can prove that RMS is better than RAS, in the sense that the corresponding iteration matrix has a smaller norm, for a certain weighted max norm.

Our results provide the theoretical underpinnings for the behavior of the RAS preconditioners as observed in [CS99]. The theory we develop is not complete in the sense that we do not get quantitative results (like mesh independence in the presence of a coarse grid, for example). However, such results can be obtained indirectly by using some of the comparison results of [FS01] and classical results for the usual Schwarz method.

Our approach is purely algebraic, and therefore our results apply to discretization of differential equations as well as to algebraic additive Schwarz. We believe that the algebraic tools used here and in [FS99, BFNS01] complement the usual analytic tools used for the analysis of Schwarz methods; see, e.g., the books [SBG96, QV99] and the extensive bibliography therein.

One of the reasons why the algebraic approach presented here is a good alternative to the classical approach is that the operators defining RAS and RMS are not orthogonal projections (see [FS01]), and thus the usual theory as described, e.g., in [BM91] does not apply.

This paper is organized as follows. We start by giving algebraic representations of the usual and the restricted additive Schwarz methods and we introduce the splittings associated with each of the methods. Then, our convergence theorem for RAS, as well as results on the effect of overlap on the quality of the preconditioner are presented. Finally, convergence of RMS is shown, together with the comparison between RMS and RAS.

We note that using the same formulation described in this paper, several variants of RAS

[1]Fachbereich Mathematik, Bergische Universität GH Wuppertal, Gauss-Strasse 20, D–42097 Wuppertal, Germany, frommer@math.uni-wuppertal.de

[2]Fakultät für Mathematik, Universität Bielefeld, Postfach 10 01 31, 33501 Bielefeld, Germany, nabben@mathematik.uni-bielefeld.de

[3]Department of Mathematics, Temple University (038-16), 1805 N. Broad Street, Philadelphia, Pennsylvania 19122-6094, USA, szyld@math.temple.edu . Supported by the U.S. National Science Foundation grant DMS-9973219

and RMS preconditioners can be analyzed, including the cases of inexact local solutions and of weighted methods; see [FS01, NS01].

The algebraic representation

The $n \times n$ linear system is given as

$$Ax = b. \tag{1}$$

As in [CS99] we consider p nonoverlapping subspaces $W_{i,0},\ i = 1, \ldots, p$ which are spanned by columns of the $n \times n$ identity I and which are then augmented to produce overlap. For a precise definition, let $S = \{1, \ldots, n\}$ and let

$$S = \bigcup_{i=1}^{p} S_{i,0}$$

be a partition of S into p disjoint, non-empty subsets. For each of these sets $S_{i,0}$ we consider a nested sequence of larger sets $S_{i,\delta}$ with

$$S_{i,0} \subseteq S_{i,1} \subseteq S_{i,2} \ldots \subseteq S = \{1, \ldots, n\}, \tag{2}$$

so that we again have $S = \cup_{i=1}^{p} S_{i,\delta}$ for all values of δ, but for $\delta > 0$ the sets $S_{i,\delta}$ are not necessarily pairwise disjoint, i.e., we have introduced *overlap*. A common way to obtain the sets $S_{i,\delta}$ is to add those indices to $S_{i,0}$ which correspond to nodes lying at distance δ or less from those nodes corresponding to $S_{i,0}$ in the (undirected) graph of A.

Let $n_{i,\delta} = |S_{i,\delta}|$ denote the cardinality of the set $S_{i,\delta}$. For each nested sequence from (2) we can find a permutation π_i on $\{1, \ldots, n\}$ with the property that for all $\delta \geq 0$ we have $\pi_i(S_{i,\delta}) = \{1, \ldots, n_{i,\delta}\}$.

We now build $n_{i,\delta} \times n$ matrices whose rows are precisely those rows j of the identity for which $j \in S_{i,\delta}$. Formally, such a matrix $R_{i,\delta}$ can be expressed as

$$R_{i,\delta} = [I_{i,\delta}|O]\, \pi_i \tag{3}$$

with $I_{i,\delta}$ the identity on the $n_{i,\delta}$-space. Finally, we define the $n \times n$ weighting matrices

$$E_{i,\delta} = R_{i,\delta}^T R_{i,\delta} \quad \left(= \pi_i^T \begin{bmatrix} I_{i,\delta} & O \\ O & O \end{bmatrix} \pi_i \right)$$

and the subspaces

$$W_{i,\delta} = \mathrm{range}(E_{i,\delta}),\ \ i = 1, \ldots, p.$$

Note the inclusion $W_{i,\delta} \supseteq W_{i,\delta'}$ for $\delta \geq \delta'$, and in particular $W_{i,\delta} \supseteq W_{i,0}$ for all $\delta \geq 0$.

We view the matrices $R_{i,\delta}$ as restriction operators and $R_{i,\delta}^T$ as prolongations. We can identify the image of $R_{i,\delta}^T$ with the subspace $W_{i,\delta}$. For each subspace $W_{i,\delta}$ we define a restriction of the operator A on $W_{i,\delta}$ as

$$A_{i,\delta} = R_{i,\delta} A R_{i,\delta}^T.$$

The classical additive Schwarz method consists of using the following preconditioner in a Krylov subspace method for solving (1):

$$M_{AS,\delta}^{-1} = \sum_{i=1}^{p} R_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta}. \tag{4}$$

In order to describe the *restricted* additive Schwarz method we introduce 'restricted' $n_{i,\delta} \times n$ operators $\widetilde{R}_{i,\delta}$ as

$$\widetilde{R}_{i,\delta} = R_{i,\delta} E_{i,0} \tag{5}$$

The image of $\widetilde{R}_{i,\delta}^T = E_{i,0} R_{i,\delta}^T$ can be identified with $W_{i,0}$, so $\widetilde{R}_{i,\delta}^T$ 'restricts' $R_{i,\delta}^T$ in the sense that the image of the latter, $W_{i,\delta}$, is restricted to its subspace $W_{i,0}$, the space from the non-overlapping decomposition. The restricted additive Schwarz method from [CFS98, CS99] replaces the prolongation operator $R_{i,\delta}^T$ by $\widetilde{R}_{i,\delta}^T$ and thus uses

$$M_{RAS,\delta}^{-1} = \sum_{i=1}^{p} \widetilde{R}_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta} \tag{6}$$

instead of (4)[4]. For practical parallel implementations, replacing $R_{i,\delta}^T$ by $\widetilde{R}_{i,\delta}^T$ means that the corresponding part of the computation will not require any communication, since the images of the $\widetilde{R}_{i,\delta}^T$ do not overlap. In addition, the numerical results in [CS99] indicate that the restrictive additive Schwarz method is at least as fast (in terms of number of iterations and/or CPU time) as the classical one. Note that we lose symmetry, however, since if A is symmetric, $M_{AS,\delta}^{-1}$ will be symmetric as well, whereas $M_{RAS,\delta}^{-1}$ will usually be nonsymmetric.

For the convergence analysis of these Krylov methods, the relevant matrices are $M_{AS,\delta}^{-1} A$ and $M_{RAS,\delta}^{-1} A$. Alternatively, we can consider the iteration matrices $T_{AS,\delta} = I - M_{AS,\delta}^{-1} A$ and $T_{RAS,\delta} = I - M_{RAS,\delta}^{-1} A$. To analyze these matrices, we write the orthogonal projections

$$P_{i,\delta} = R_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta} A, \; i = 1, \ldots, p$$

and the oblique projections

$$Q_{i,\delta} = \tilde{R}_{i,\delta}^T A_{i,\delta}^{-1} R_{i,\delta} A, \; i = 1, \ldots, p,$$

and thus we have the representation

$$T_{AS,\delta} = I - \sum_{i=1}^{p} P_{i,\delta}, \;\; T_{RAS,\delta} = I - \sum_{i=1}^{p} Q_{i,\delta}. \tag{7}$$

With this notation, the iteration matrix corresponding to the classical multiplicative Schwarz method is

$$T_{MS} = (I - P_{p,\delta})(I - P_{p-1,\delta}) \cdots (I - P_{1,\delta}) \tag{8}$$

[4]We note that the representations (4) and (6) using rectangular matrices $R_{i,\delta}$ and matrices $A_{i,\delta}$ of smaller size is consistent with the standard literature [SBG96, QV99] and different than that of [CS99] where $n \times n$ matrices are used.

and the corresponding iteration matrix for the RMS method is

$$T_{RMS} = (I - Q_{p,\delta})(I - Q_{p-1,\delta}) \cdots (I - Q_{1,\delta}) \tag{9}$$

As in [FS99, BFNS01], the key to our analysis is the use of the nonsingular matrices $M_{i,\delta}$ defined as

$$M_{i,\delta} = \pi_i^T \begin{bmatrix} A_{i,\delta} & O \\ O & D_{\neg i,\delta} \end{bmatrix} \pi_i$$

where $D_{\neg i,\delta}$ is the diagonal part of the principal submatrix of A 'complementary' to $A_{i,\delta}$, i.e.,

$$D_{\neg i,\delta} = \operatorname{diag}\left([O|I_{\neg i,\delta}] \cdot \pi_i \cdot A \cdot \pi_i^T \cdot [O|I_{\neg i,\delta}]^T\right)$$

with $I_{\neg i,\delta}$ the $n - n_{i,\delta} \times n - n_{i,\delta}$ identity. Here, we assume that $A_{i,\delta}$ and $D_{\neg i,\delta}$ are nonsingular. With these matrices we can write

$$P_{i,\delta} = E_{i,\delta} M_{i,\delta}^{-1} A, \tag{10}$$

$$Q_{i,\delta} = E_{i,0} M_{i,\delta}^{-1} A, \tag{11}$$

and this provides a new representation of the matrices (7), (8), and (9); see [FS99, FS01, BFNS01, NS01]. The new representation of the additive Schwarz methods is very much in the spirit of multisplittings; see [OW85], or [BMPS95] and its bibliography.

We note that with the RAS preconditioning the corresponding weighting matrices satisfy

$$\sum_{i=1}^{p} E_{i,0} = I,$$

consistent with the traditional multisplitting theory, while for additive Schwarz we have

$$qI \geq \sum_{i=1}^{p} E_{i,\delta} \geq I,$$

where the inequalities are componentwise and

$$q = \max_{j=1,\ldots,n} |\{i : j \in S_{i,\delta}\}|. \tag{12}$$

In the p.d.e. setting, q is the maximum number of subdomains to which each node of the mesh belongs.

Convergence of RAS

We show in this section that for M-matrices the spectral radius $\rho(I - M_{RAS,\delta}^{-1} A)$ of the RAS iteration matrix is less than 1 for all values of $\delta \geq 0$. This implies in particular that the spectrum of the preconditioned system $\sigma(M_{RAS,\delta}^{-1} A)$ is located in the right half plane and contained in a disk of radius less than one around the point 1.

We start by recalling some basic terminology. The natural partial ordering $\leq$ between matrices $A = (a_{ij})$, $B = (b_{ij})$ of the same size is defined component-wise, i.e., $A \leq B$ iff

$a_{ij} \leq b_{ij}$ for all i, j. If $A \geq O$ we call A nonnegative. If all entries of A are positive, we say that A is positive and write $A > O$. This notation and terminology carries over to vectors as well. An $n \times n$ matrix A is called a (nonsingular) M-matrix if it has nonpositive off-diagonal elements and $A^{-1} \geq O$; see [Var62].

Consider the splitting $A = M - N$ with M nonsingular. This splitting is said to be *weak nonnegative of the first type* (also called *weak regular*) if

$$M^{-1} \geq O \text{ and } M^{-1}N \geq O. \tag{13}$$

Theorem 1 [OR70] *Let $A = M - N$ be a weak nonnegative splitting of the first type. Then $\rho(I - M^{-1}A) < 1$ iff A is nonsingular and $A^{-1} \geq 0$.*

We are now able to formulate the central convergence result of this section.

Theorem 2 *Let A be a nonsingular M-matrix. Then for each value of $\delta \geq 0$, the splitting $A = M_{RAS,\delta} - N_{RAS,\delta}$, corresponding to the RAS method, is weak nonnegative of the first type. In particular, the iteration matrix $M_{RAS,\delta}^{-1}N_{RAS,\delta} = I - M_{RAS,\delta}^{-1}A$ satisfies*

$$\rho(I - M_{RAS,\delta}^{-1}A) < 1. \tag{14}$$

The proof consists of showing that $M_{RAS,\delta}^{-1} \geq O$, and that $I - M_{RAS,\delta}^{-1}A \geq O$, as per (13), and then apply Theorem 1; see [FS01].

We point out that in general a convergence result such as (14) does not hold for the classical additive Schwarz preconditioner (4). To guarantee convergence, a damping (or relaxation) parameter $\theta > 0$ is introduced. It can be shown that if $\theta \leq 1/q$, then $\rho(I - \theta M_{AS,\delta}^{-1}A) < 1$, where q is defined in (12); see [FS99, BFNS01]. Thus, one of the attractive features of the RAS preconditioner is that no damping parameter is needed for convergence.

Using an appropriate norm, we study the effect of varying the overlap. More precisely, we prove comparison results on the spectral radii and/or on certain weighted max norms for the corresponding iteration matrices $T_{RAS,\delta}$ as defined in (7) for different values of $\delta \geq 0$.

We want to compare one RAS splitting, defined through the sets $S_{i,\delta'}$ with another one with more overlap defined through sets $S_{i,\delta}$ where $S_{i,\delta'} \subseteq S_{i,\delta}, i = 1, \ldots, p$. We show that the larger the overlap ($\delta \geq \delta'$), the faster RAS method converges as measured in certain weighted max norms. This is consistent with the experiments in Tables 1 and 2 of [CS99], where an increase of the overlap is associated with fewer iterations.

For a positive vector w we denote $\|x\|_w$ the weighted max norm in n-space given by

$$\|x\|_w = \max_{i=1,\ldots,n} |x_i|/w_i.$$

The resulting operator norm in $n \times n$-space is denoted similarly.

The following theorem from [FS01] is very similar to [FP95, Theorem 2.1].

Theorem 3 *Let A be a nonsingular M-matrix and let $w > 0$ be any positive vector such that $Aw > 0$, e.g., $w = A^{-1}v$ with $v > 0$. Then, if $\delta \geq \delta'$,*

$$\|T_{RAS,\delta}\|_w \leq \|T_{RAS,\delta'}\|_w. \tag{15}$$

Moreover, if the Perron vector $w_{\delta'}$ of $T_{RAS,\delta'}$ satisfies $w_{\delta'} > 0$ and $Aw_{\delta'} \geq 0$, then we also have

$$\rho(T_{RAS,\delta}) \leq \rho(T_{RAS,\delta'}). \tag{16}$$

In the case that (16) holds, Theorem 3 indicates that the spectrum of the preconditioned matrix is included in a possibly smaller disk when the overlap is increased.

We remark that (15) (as well as most results using the weighted max norms in the paper) holds for *any* positive vector w such that Aw is positive, so that one has a lot of freedom in choosing the norm. For example, if all row-sums of A are positive we can choose as w the vector of all ones, and thus the weighted max norm is simply the max norm. A commonly chosen vector w is the row-sums of A^{-1}, which is always positive.

For $\delta' = 0$, i.e., for the block Jacobi preconditioner we can always provide the comparison of the spectral radii (16), in addition to the comparison (15). The following theorem is in fact [FP95, Theorem 2.2].

Theorem 4 *Let A be a nonsingular M-matrix. Then, for any value of $\delta \geq 0$,*

$$\rho(T_{RAS,\delta}) \leq \rho(T_{RAS,0}) \, .$$

Convergence of RMS

Using the new algebraic representation (10), it was shown in [BFNS01] that for any $w = A^{-1}e > 0$ with $e > 0$, we have $\rho(T_{MS}) \leq \|T_{MS}\|_w < 1$. In a similar way, using the representation (11), we can prove the following result; see [NS01].

Theorem 5 *Let A be a nonsingular M-matrix. For any $w = A^{-1}e > 0$ with $e > 0$, we have $\rho(T_{RMS}) \leq \|T_{RMS}\|_w < 1$. Furthermore, there exists a unique splitting $A = B - C$ such that $T_{MRS} = B^{-1}C$, and this splitting is weak nonnegative of the first type.*

It is well known that bounds for the convergence using the standard multiplicative Schwarz preconditioner are better than those obtained for the standard additive Schwarz; see, e.g. [SBG96, QV99]. For the restrictive preconditioners we can actually show that the weighted max norm of the RMS iteration matrix is smaller than that of RAS.

Theorem 6 *Let A be a nonsingular M-matrix and let $w > 0$ be any positive vector such that $Aw > 0$, e.g., $w = A^{-1}e$ with $e > 0$. Then,*

$$\|T_{RMS,\delta}\|_w \leq \|T_{RAS,\delta}\|_w \, .$$

Moreover, if the Perron vector w_δ of $T_{RAS,\delta}$ satisfies $w_\delta > 0$ and $Aw_\delta \geq 0$, then we also have

$$\rho(T_{RMS,\delta}) \leq \rho(T_{RAS,\delta}) \, .$$

The proof consists of showing that $M^{-1}_{RMS,\delta} \geq M^{-1}_{RAS,\delta}$, where $M_{RMS,\delta} = (I - T_{RMS,\delta})^{-1}A$. This inequality together with theorems 2 and 5, and Theorem 4.1 of [FS99] provides the needed norm and spectral radii inequalities; see [NS01].

As is the case for RAS, one can also show that by increasing the overlap, the weighted max norm of the iteration matrix decreases, i.e., that if $\delta \geq \delta'$,

$$\|T_{RMS,\delta}\|_w \leq \|T_{RMS,\delta'}\|_w < 1$$

for any $w > 0$ such that $Aw > 0$. Furthermore, it can be shown that overlap is always better than no overlap, i.e., for any value of $\delta \geq 0$,

$$\rho(T_{RMS,\delta}) \leq \rho(T_{RMS,0}) \, .$$

References

[BFNS01]Michele Benzi, Andreas Frommer, Reinhard Nabben, and Daniel B. Szyld. Algebraic theory of multiplicative Schwarz methods. *Numerische Mathematik*, 89:605–639, 2001.

[BGMS97]Satish Balay, William D. Gropp, Louis C. McInnes, and Barry F. Smith. *PETSc 2.0 User Manual*. Argonne National Laboratory, http://www.mcs.anl.gov/petsc/, 1997.

[BM91]Petter E. Bjørstad and Jan Mandel. On the spectra of sums of orthogonal projections with applications to parallel computing. *BIT*, 31:76–88, 1991.

[BMPS95]Rafael Bru, Violeta Migallón, José Penadés, and Daniel B. Szyld. Parallel, synchronous and asynchronous two-stage multisplitting methods. *Electronic Transactions on Numerical Analysis*, 3:24–38, 1995.

[CFS98]X.-C. Cai, C. Farhat, and M. Sarkis. A minimum overlap restricted additive Schwarz preconditioner and applications to 3D flow simulations. *Contemporary Mathematics*, 218:479–485, 1998.

[CS96]Xiao-Chuan Cai and Youcef Saad. Overlapping domain decomposition algorithms for general sparse matrices. *Numerical Linear Algebra with Applications*, 3:221–237, 1996.

[CS99]Xiao-Chuan Cai and Marcus Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM Journal on Scientific Computing*, 21:239–247, 1999.

[FP95]Andreas Frommer and Bert Pohl. A comparison result for multisplittings and waveform relaxation methods. *Numerical Linear Algebra with Applications*, 2:335–346, 1995.

[FS99]Andreas Frommer and Daniel B. Szyld. Weighted max norms, splittings, and overlapping additive schwarz iterations. *Numerische Mathematik*, 83:259–278, 1999.

[FS01]Andreas Frommer and Daniel B. Szyld. An algebraic convergence theory for restricted additive Schwarz methods using weighted max norms. *SIAM Journal on Numerical Analysis*, 39:463–479, 2001.

[NS01]Reinhard Nabben and Daniel B. Szyld. Convergence theory of restrictive multiplicative Schwarz methods. Technical Report 01-05-17, Department of Mathematics, Temple University, Philadelphia, Pa., May 2001.

[OR70]James M. Ortega and Werner C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York and London, 1970.

[OW85]D. P. O'Leary and R.E. White. Multi-splittings of matrices and parallel solution of linear systems. *SIAM J. Alg. Disc. Meth.*, 6:630–640, 1985.

[QV99]Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

[SBG96]Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[Var62]Richard S. Varga. *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1962. Second Edition, Springer, Berlin, 2000.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

38 Some Remarks on Multilevel Method, Extrapolation and Code Verification

M. Garbey [1] [2]

1 Motivation

Large scale CFD is becoming an important tool in the industry, and its success is very much connected to the reliability of the output and the cost to produce the data. It is rare that the reliability of the answer can be based exclusively on a firm mathematical basis, because the models of interest for industry are often too complex for that. Code validation and verification are therefore becoming essential.

Code Validation (1) and Verification (2) decompose roughly into a search for: Physical Modeling Errors (1), Discretization Errors (2), Programming Errors (2) (i.e mistakes), and Computer Roundoff errors (2). We refer to [OBA95] for a proper taxonomy of errors. According to [R98], discretization error can be evaluated by grid refinement verification studies. The so-called grid convergence index [R98] can be computed to assess accuracy. Richardson extrapolation plays a central role, in this type of code verification but is limited to approximation method with a known order of convergence. This information for complex flow simulation is not available in general, either because the hypothesis of the approximation theory are not fully satisfied or simply because error estimates are essentially asymptotic relations and practical calculation have not meshes fine enough to be in the range of the asymptotic limit description. A better solution from the stand point of applied mathematics is to use an a posteriori estimator, and lead the grid refinement with this tool. However, we will stay away from this solution in order to deal with a CFD code as a black box, as it should be done in principle for code verification.

We would like further to point out that the verification of large scale computation code is difficult with modern parallel computing environment. Discretized errors might be affected because the type of grid that is distributed on parallel computers is not necessary the classical one that used on a sequential machine ; classical examples are overlapping grids , non matching grids, etc ... Programming errors on parallel systems with distributed memory as Beowulf systems are of main concern because parallel codes are more complex than sequential code. Also there is no shared resources on Multiple Instruction Multiple Data Architecture (MIMD), communication might be asynchronous, and result can depend slightly on the run, with modern iterative solvers [1]. Round off errors are critical, because the grids have very large number of nodes and therefore the condition numbers might be very high.

In this paper, we will discuss some methodology that tries to combine multilevel method, such as cascade of computation on refined grids, with some aspect of code verification such as order of accuracy verification. One of the key constraints that we would like to keep in the design of this methodology is the fact that one reuses an existing CFD code on different grids without rewriting the solvers, and one basically accumulates experience from the coarse grid

[1] COCS, University of Houston
[2] CDCSP/ISTIL - University Lyon 1, 69622 Villeurbanne, France
{garbey}@cdcsp.univ-lyon1.fr, http://cdcsp.univ-lyon1.fr

to the finest grid computation in order to save eventually cpu time and improve the solution quality. Our ultimate goal, in the design of the algorithm, is therefore to separate completely possible improvements of the CFD code done by the user from post processing/preprocessing methods that should increase the efficiency and trust in the overall numerical process.

2 Verification of Code and Cascade of Grids

We are going to present an attempt [EGH00] to make the additive Schwarz procedure more efficient while keeping as much as possible the simplicity of the original method and using cascade algorithm with three level of grids for acceleration purpose as well as verification. Most of the concepts describe thereafter can be applied in a straightforward way to FE or FV discretization with unstructured grid. But for simplicity, we will report on second-order FD solution of the well-known Bratu problem in a square. The problem is written :

$$N[u] = \Delta u + \lambda e^u = O \; in \; \Omega = (0,1)^2, \qquad u_{|\partial\Omega} = 0 \tag{1}$$

The discretized problem has the form,

$$N^h[U] = 0 \equiv \begin{cases} \frac{U_{i+1,j}-2U_{i,j}+U_{i-1,j}}{h_x^2} + \frac{U_{i,j+1}-2U_{i,j}+U_{i,j-1}}{h_y^2} + \lambda \, e^{U_{i,j}} = 0, \\ U_{1,j} = U_{N_x,j} = U_{i,1} = U_{i,N_y} = 0, \\ i = 1, \cdots, N_x - 1, \; j = 1, \cdots, N_y - 1, \end{cases}$$

where h_x (resp. h_y) denotes the space step in x variable (resp y variable). We consider a basic decomposition of the domain into nd-overlapping strips, $(x_A(k), x_B(k)) \times (0,1), \; k = 1 \cdots nd$ with arbitrary overlap of q intervals (meshes) that is : $x_B(k) - x_A(k+1) = q \cdot h_x, \qquad k = 1 \cdots nd-1$, q being a positive integer. At continuous level, the algorithm appears, for $k = 1, \cdots, nd$:

$$N[U^{k,n+1}] = 0$$

$$U^{k,n+1}(x_A(k), \bullet) = U^{k-1,n}(x_A(k), \bullet)$$

$$U^{k,n+1}(x_B(k), \bullet) = U^{k+1,n}(x_B(k), \bullet)$$

with, in our notation, formally $U^{0,n} \equiv U^{nd+1,n} \equiv 0$.
This process constructs a multi-valued piecewise solution because of the overlap and we define the global solution as follows. For $x \in (x_A(k), x_B(k))$:

$$u^{n+1}(x,y) = \sum_{k=1}^{nd} \aleph^k u^{k,n+1}(x, \bullet) + (1 - \aleph^k) \begin{cases} u^{k-1,n+1}(x, \bullet) & \text{if } x > \frac{1}{2}(x_A + x_B)(k) \\ u^{k+1,n+1}(x, \bullet) & \text{if } x \leq \frac{1}{2}(x_A + x_B)(k) \end{cases}$$

with $\aleph^k$ a smooth partition of unity that is 1 in $(x_B(k-1), x_A(k+1))$ and 0 outside $(x_A(k), x_B(k))$.

We know that the convergence of this algorithm is linear and very slow. However, there is obviously no need to solve exactly each nonlinear problem in each sub-block, since the domain decomposition is an iterative process. One can optimize then the stopping criterion of the subdomain iterative solver by using a prediction of some norm of the jump at artificial interfaces at the end of each Schwarz iterate [EGH00].

The block solver is actually a nonlinear solver based on the Newton algorithm and a Preconditioned Conjugate Gradient algorithm with Incomplete LU factorization (PCGILU) for the linear system. In our basic strategy to enhance the additive Schwarz algorithm, we introduce three levels of grids, G^m, $m = 1, 2, 3$. For simplicity, we restrict ourselves to embedded grids with discretization ratio 2, but as it will be seen later on, this simplification is not necessary. The classical idea of cascade algorithm is to provide as an initial guess for the iterative solution process on the grid level m an initial guess that is obtained from the discretized solution on the coarser grid $m - 1$. This very old idea is called nested loop when Successive Over Relaxation (SOR) is the block solver. It is known that in terms of arithmetic efficiency the nested loop method should be limited to two levels of grids. The implementation is very simple, since one always go from the coarse grid to the next finer grid level, as long as grid level can be defined properly. The main purpose of using three levels of grids instead of two will be that it will provide enough information in order to proceed to some code verification. We will denote in the following $U^{(m,h)}$ the discrete solution obtained by our iterative method on grid level m. So the algorithm that we propose is to solve, first, with additive Schwarz and iterative block solver, the discretized problem on the grid G^1. Then, we project the solution on grid G^2 and use an interpolation procedure to define the initial guess every where. Second-order linear interpolation seems at first sight a natural tool. The same solution procedure is then reproduced on the grid G^2. If we basically project the solution obtained on G^2 into G^3, we miss a very important property of our approximation method. The discretized solution of Bratu problem should converge to the exact solution with order two accuracy in space as $h = (h_x, h_y)$ goes to zero, that is $U - U^h = O(h_x^2 + h_y^2)$. Using the classical Richardson extrapolation principle reported as in Roache [R98] for code verification, we may therefore produce an initial guess for the iterative solution procedure on grid G^3 that is much better than a basic projection of U^2 onto G^3. We define the initial guess for the iterative solution on G^3 to be

$$U_1^{(3)} = 4/3U^{(2,h)} - 1/3U^{(1,h)}. \tag{2}$$

We encounter two difficulties. First of all, $U_1^{(3)}$ is defined only on grid $G^{(1)}$. We do need therefore to interpolate $U_1^{(3)}$ with an interpolation procedure that keeps the Richardson extrapolation procedure effective on $G^{(3)}$. If $U^{(2,h)}$ and $U^{(1,h)}$ are known at second-order on $G^{(1)}$, we can expect $U_1^{(3)}$ to be at least a third-order approximation of $U^{(3,h)}$ on $G^{(1)}$. Actually with regular grids of constant space steps, one obtains a fourth-order approximation. In order to preserve as much as possible the quality of this information, we should use a third-order interpolation method to extend our grid solution $U_1^{(3)}$ from $G^{(1)}$ to $G^{(3)}$. In our case, we have applied cubic bilinear interpolation as well as spline interpolation. A second difficulty is that $U^{(1,h)}$ and $U^{(2,h)}$ are computed with an iterative procedure. The Richardson extrapolation principle applies to exact discrete solution. The error is then given by a truncation error formula based on Taylor expansion of the discrete operator with respect to discrete parameter h for which the leading coefficients are a priori independent of h. As opposed to the iterative

solution procedure on a single grid, one therefore needs to compute an approximation of $U^{(j,h)}$, $j = 1, 2$ on the grids $G^{(j)}$, $j = 1, 2$, with a residual that is of the order the expected rate of convergence of Richardson extrapolation method, that is at least h^3. This procedure combining Richardson extrapolation and high order interpolation allows then to produce a good initial guess for the iterative solution on grid $G^{(3)}$, and the code can be verified by simply proceeding with the computation of $U^{(3,h)}$. In principle, the number of Schwarz iterates should be small if the construction of $U^{(3)}$ and the resolution of $U^{(j,h)}$, $j = 1, 2$ are correct. The solution process is however robust because $U^{(3)}$ is only used as an initial guess. In the mean time the order $B(x)$ of the method can be checked with formula:

$$B = log_{10}|U^{(1)} - U^{(2)}|/|U^{(2)} - U^{(3)}|/log_{10}(2), \tag{3}$$

at each point x.

This formula can be used in combination to the computation of the residual in order to verify the stopping criterion of the iterative computation of the solution on the finest grid $G^{(3)}$that in practice dominates the overall cost of the method.

In order to illustrate the method, we report thereafter on the numerical solution of the Bratu problem with $\lambda = 6$. Our computation, realized with matlab, is rather modest. The fine grid $G^{(3)}$ has 65×61 grid points. The main thrust of this method is not the arithmetic efficiency but its simplicity. However, we compare the number of floating point operations realized with our algorithm to the flops performance of PCGILU, no domain decomposition, one single fine grid and the trivial initial guess on $G^{(3)}$. Parallel efficiency is analogous to [PARCFD] and it is known that the additive Schwarz scales well even on MIMD parallel systems with slow network as long as the load per processors is high enough .

The Schwarz algorithm is applied with minimum overlap on grid $G^{(j)}$, $j = 1, 2$. The size of the overlap is chosen on $G^{(3)}$ in order to minimize the overall flops performance. The order of convergence of the method $B(x)$ is computed on the coarse grid; as expected, we observe that $B(x)$ is 2 within 5%. Table 1 shows the global efficiency of the iterative solver by listing the total number of Mflops used on cascade algorithm, and the number of Schwarz iterates for each grid level. It should be noticed that monitoring the order of the method with formula (3) may avoid premature iteration stops of the Schwarz iteration process. It is also interesting to notice that the overall number of flops to reach the correct solution is relatively insensitive to the number of subdomains, but that the number of Schwarz iterates grows as expected with the number of subdomains. This has an impact on the parallel efficiency of the method that is usually limited by the network of communications, i.e. the more messages i.e. Schwarz iterates, the less is the parallel efficiency. To conclude this section we observe that our Bratu problem's example is characterized by a very smooth solution and our results cannot be reproduced for non smooth problems. For example, the Richardson extrapolation methods break down for the cavity flow problem with a finite difference computation of $\omega - \psi$ formulation even for modest Reynolds number of order 100, because of the existence of a singularity in the flow field at the corner [SGAW01] We proceed therefore with a modification of the Richardson extrapolation method that may work with numerical method with varying order of approximation.

Number of subdomains	2	3	4	5	6
Flops ratio versus no DD method	1.69	1.06	0.95	0.95	0.99
Mflops used	89.5	142	120	120	115
Number of Schwarz iterates on G1	79	100	119	150	202
Number of Schwarz iterates on G2	88	131	149	168	196
Number of Schwarz iterates on G3	8	15	12	13	14

Table 1: *Cascade-Newton-Schwarz on 2D Bratu problem with 65×61 unknowns.*

3 A Generalized Extrapolation Method

In order to present the idea, we restrict ourselves to problems in one space dimensions with regular grids. We refer to the report [GS01] to provide more details on the analysis and application results in higher space dimensions. Let us consider two continuous functions $U(x,h_1)$ and $U(x,h_2)$ that are approximations of a continuous function $U(x),\ x \in (0,\pi)$. A general consistent linear extrapolation formula writes:

$$\alpha\ U^{(1,h)}\ +\ (1-\alpha)\ U^{(2,h)}.$$

If the approximation method to build $U(x,h)$ is order q, the extrapolation formula becomes

$$\tilde{U}(x,h) = \frac{h_2^q U(x,h_1) - h_1^q U(x,h_2)}{h_2^q - h_1^q}.$$

Let us consider now a linear differential problem $L[U] = 0$, and a sequence of consistent approximation L^h obtained with Finite Differences or Finite Volume for instance. The following discussion is fairly general and can be reproduced for variational formulations. Let us suppose that the consistency error in some norm is of order q, and can be expanded as follows:

$$L^h[U(x,h_j] = R_q(x)h_j^q + R_{q+1}(x)h_j^{q+1} + \cdot. \tag{4}$$

Then q is the constant that minimizes the *asymptotic* order of the residual $L[\alpha U(x,h_1) + (1-\alpha)U(x,h_2)]$, with $\alpha = \frac{h_2^q}{h_2^q - h_1^q}$. Using a stability estimate on L^h, one can then prove that the Richardson extrapolation $\tilde{U}$ is a better approximation of U than any of the approximation $U(x,h_j),\ j=1,2$.

The main difficulty in practice, is that h_1 and h_2 may not be small enough to produce residual for which the first-order term R_q in the expansion (4) is significantly greater than the next order term $R_{q+1}h$. Even worst, R_q and R_{q+1} are space dependent functions and the approximations can behave as a q-order approximation in some subdomain and a $q+1$-order approximation elsewhere. Classical Navier-Stokes provides a large collection of such examples when boundary layer or transition layer occur. In addition, lift and drag calculations may not require method with uniform order approximation.

We propose therefore to define the following problem:
find $\alpha(x)$ in some vector space to be defined later on, that minimizes the residual:

$$L^h[\alpha U(x,h_1) + (1-\alpha)U(x,h_2)], \tag{5}$$

in some norm to be defined.

$\tilde{U} = \alpha(x)U(x,h_1) + (1-\alpha(x))U(x,h_2)$ will be then our new extrapolation formula.

We have now several difficulties: first, we deal with grid functions $U^{(j,h)}$ instead of continuous approximation $U(x,h_j)$. Second, following the lines of Sect 2, the practical purpose of this optimization problem is to provide a good solution on the fine grid G^3. Therefore the computation of the residual on G^3 requires a high order interpolation of $U^{j,h}, j = 1,2$ that is robust with respect to differentiation. Third the optimization problem should be well posed and its solution should cost much less than the fine grid computation on G^3.

We propose in the following a basic algorithm that seems to be a good candidate to improve the basic Richardson extrapolation procedure. We look for α as the shifted Fourier expansion

$$\alpha_{N_F} \; = \; \alpha_0 + \alpha_1 \; cos(x) + \Sigma_{j=2..N_F} \alpha_j \; sin((j-1)x), \tag{6}$$

that is solution of the least square problem

$$L^h[\alpha U^{(1,h)} + (1-\alpha)U^{(2,h)}] = 0, \; on \; grid \; G^3. \tag{7}$$

We observe that the solution of the least square problem $\alpha_N = \alpha \; on \; G^3$ with $\alpha \; in \; C^2(0,\pi)$ is a second-order approximation in maximum norm of α on G^3, and a third-order approximation away from the end boundaries. We observe however that at location where $\delta U = U^{(1,h)} - U^{(2,h)}$ is close to zero, and $U^{(1,h)}, \; U^{(2,h)}$ are not close to the exact discrete solution $U^{(3,h)}$ within asymptotic order $O(\delta U)$, we have a singularity. In practice, there is not much improvement that one may expect from classical extrapolation formula in such situation. We therefore can detect local failure of our grid solutions G^1 and G^2 with (6, 7) but not fix it. We found in practice more robust to use three levels of grids $G^j, j = 1..3$ in order to predict the fine grid solution on G^4 with the following least square problem:
find α and β *with expansion similar to (6), that is solution of the least square problem*

$$L^h[\alpha U^{(1,h)} + \beta U^{(2,h)} + (1-\alpha-\beta)U^{(3,h)}] = 0, \; on \; grid \; G^4. \tag{8}$$

Some details on the analysis of this method and its application to Navier Stokes problem can be found in [GS01]. But in this proceeding paper, we would like to show with the following classical Burgers problem,

$$-\epsilon u_{xx} - (\frac{1}{2}u^2)_x = 0, \; on \; (0,\pi), u(0) = \pi/2 + \epsilon, \; u(\pi) = -\pi/2 + \epsilon,$$

that even if the Richardson extrapolation methods fails to improve the underlines grid solutions $U^{(j,h)}, \; j = 2,3$, our least square extrapolation method may give a much better solution. Obviously since Burgers is a nonlinear problem, we apply recursively a least square linear solve to the linearized problem with a Newton like loop. Fig 1 shows the error curves with respect to the exact discrete solution on G_4, with central finite differences. The number of grid points on G^1, (resp. $G^2, \; G^3$) is $N_1 = 11$, (resp. $N_2 = 19$, $N_3 = 25$) and $\epsilon = 0.1$. As an interpolant for the grid functions $U^{(j,h)}$, we have used spline interpolant (Fig 1) as well as shifted Fourier interpolant similar to (6). N_F in the shifted Fourier expansion of α is limited to 6 and the number of Newton iteration to 3. In each case, the Richardson extrapolation assuming a first-order method (R1 curve), or a second-order method (R2 curve) are less accurate than the solution on the fine grid G3. Nevertheless the least square extrapolation improves significantly the accuracy of this fine grid solution. These results are quiet good and have been extended to multidimensional problems [GS01].

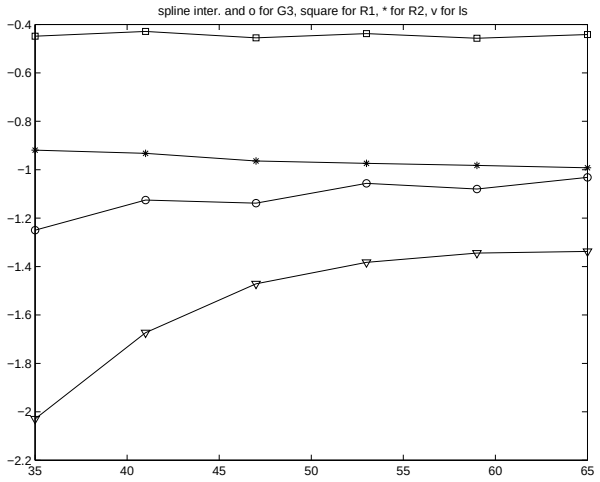


Figure 1: The x-axis is for the number of grid points on G_4. The y-axis gives the error in maximum norm in log_{10} scale on grid G_4.

4 Conclusions:

We have briefly presented, first in this paper some background on code verification, second a practical way of combining efficiency in the solution process and some code verification by using standard additive Schwarz algorithm with cascade method. In the development of our solution, we restrict ourselves carefully to a method that should be easily generalized to non-structured grid and FV or FE discretization. We have shown some practical limitation on the use of Richardson extrapolation and presented a least square extrapolation variant that looks promising for CFD application [S94]. Finally, we observe that multilevel methods give us the opportunity to provide solutions on several grids and that it should be an important tool used to understand better the convergence accuracy of a CFD code for which it is rare that all mathematical hypothesis are fulfilled correctly.

References

[BT89]D.P. Bertesekas and J. N. Tsitsiklis, *Parallel and Distributed Computation- Numerical Methods*, Prentice Hall 1989.

[EGH00]A. Ecer, M. Garbey and M. Hervin, Proceedings of 12th international conference Parallel CFD 2000, North-Holland publisher, Trondheim.

[GS01]M. Garbey and W. Shyy, *A Least Square Richardson Extrapolation Method for PDE's*, preprint CDCSP, 2001.

[OBA95]W.L. Oberkampf, F.G. Blottner and D.Aeshliman, Methodology for Computational Fluid Dynamics Code Verification and Validation, 26th AIAA Fluid Dynamic Conference, June 19-22, 1995/ San Diego, CA, AIAA Paper 95-2226.

[PARCFD]A. Ecer "et al.", *Parallel CFD test case*, Series of Parallel CFD Conferences, http//www.parcfd.org

[R98]P.J. Roache, *Verification and Validation in Computational Science and Engineering*, Hermosa Publishers, Albuquerque, New Mexico, 1998.

[S94]W. Shyy, *Computational Modeling for Fluid Flow and Interfacial Transport*, New York, Elsevier 1994.

[SGAW01]W.Shyy, M.Garbey, A.Appukuttan and J.Wu, *Evaluation of Richardson Extrapolation in Computational Fluid Dynamics*, to appear in Numerical Heat Transfer, PartB: Fundamentals.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

39 A Fast Solver for Systems of Reaction-Diffusion Equations

M. Garbey,[1] H. G. Kaper,[2] and N. Romanyukha[3]

1 Introduction

In this paper we present a fast algorithm for the numerical solution of systems of reaction-diffusion equations,

$$\partial_t u + a \cdot \nabla u = \Delta u + F(x,t,u), \quad x \in \Omega \subset \mathbf{R}^2,\ t > 0. \tag{1}$$

Here, u is a vector-valued function, $u \equiv u(x,t) \in \mathbf{R}^m$, m is large, and the corresponding system of ODEs, $\partial_t u = F(x,t,u)$, is stiff. Typical examples arise in air pollution studies, where a is the given wind field and the nonlinear function F models the atmospheric chemistry.

The time integration of Eq. (1) is well handled by the method of characteristics [P89]. The problem is thus reduced to designing for the reaction-diffusion part a fast solver that has good stability properties for the given time step and does not require the computation of the full Jacobi matrix.

An operator-splitting technique, even a high-order one, combining a fast nonlinear ODE solver with an efficient solver for the diffusion operator is less effective when the reaction term is stiff. In fact, the classical Strang splitting method may underperform a first-order source splitting method [VS98]. The algorithm we propose in this paper uses an *a posteriori* filtering technique to stabilize the computation of the diffusion term. The algorithm parallelizes well, because the solution of the large system of ODEs is done pointwise [FG01]; however, the integration of the chemistry may lead to load-balancing problems [DS96, E97]. The Tchebycheff acceleration technique proposed in [L00, D90] offers an alternative that complements the approach presented here.

To facilitate the presentation, we limit the discusssion to domains Ω that either admit a regular discretization grid or decompose into subdomains that admit regular discretization grids. We describe the algorithm for one-dimensional domains in Section 2 and for multidimensional domains in Section 3. Section 4 briefly outlines future work.

2 One-Dimensional Domains

Consider the scalar equation

$$\partial_t u = \partial_x^2 u + f(u), \quad x \in (0,\pi),\ t > 0. \tag{2}$$

[1]COCS, University of Houston and University Lyon 1

[2]Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA. Work supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, U.S. Department of Energy, under Contract W-31-109-Eng-38.

[3]Institute for Mathematical Modeling, Russian Academy of Science, 125047, Moscow, Russia. Supported by the Russian Foundation of Basic Research under Grant 00-01-0291.

We combine a backward Euler approximation in time with an explicit finite-difference approximation of the diffusive term,

$$\frac{3u^{n+1} - 4u^n + u^{n-1}}{2\Delta t} = 2D_{xx}u^n - D_{xx}u^{n-1} + f(u^{n+1}). \tag{3}$$

This scheme is second-order accurate in both space and time [P83, SVLCPDS97, VHB98]. To analyze its stability, we take the Fourier transform of the linear equation,

$$\frac{3\hat{u}^{n+1} - 4\hat{u}^n + \hat{u}^{n-1}}{2\Delta t} = \Lambda_k(2\hat{u}^n - \hat{u}^{n-1}), \tag{4}$$

where $\Lambda_k = 2h^{-2}(\cos(hk) - 1)$, from which we obtain the stability condition

$$2\frac{\Delta t}{h^2}\left|\cos\left(\frac{k\pi}{N}\right) - 1\right| < \frac{4}{3}, \quad h = \frac{\pi}{N}. \tag{5}$$

Thus we conclude that the time step must satisfy the constraint

$$\Delta t < \tfrac{1}{3}h^2. \tag{6}$$

However, this constraint is imposed by the high frequencies, which are poorly handled by second-order finite differences anyway. For example, with central differences, the relative error for high-frequency waves $\cos(kx)$ with $k \approx N$ can grow at a rate of up to 9%. The idea is therefore to relax the constraint on the time step by applying a filter after each time step, which removes the high frequencies but maintains second-order accuracy in space.

2.1 Filters

By a *filter of order p* we mean an even function $\sigma : \mathbf{R} \to \mathbf{R}$ that satisfies the conditions (i) $\sigma(0) = 1$, (ii) $\sigma^{(l)}(0) = 0$ for $l = 1, \ldots, p-1$, (iii) $\sigma(\eta) = 0$ for $|\eta| \geq 1$, and (iv) $\sigma \in C^{p-1}(\mathbf{R})$.

Theorem [GS97]. Let f be a piecewise C^p function with one point of discontinuity, ξ, and let σ be a filter of order p. For any point $y \in [0, 2\pi]$, let $d(y) = \min\{|y - \xi + 2k\pi| : k = -1, 0, 1\}$. If $f_N^\sigma = \sum_{k=-\infty}^{\infty} \hat{f}_k \sigma(k/N) \mathrm{e}^{iky}$, then

$$|f(y) - f_N^\sigma(y)| \leq CN^{1-p}(d(y))^{1-p}K(f) + CN^{1/2-p}\|f^{(p)}\|_{L^2},$$

where

$$K(f) = \sum_{l=0}^{p-1}(d(y))^l |f^{(l)}(\xi^+) - f^{(l)}(\xi^-)| \int_{-\infty}^{\infty} |G_l^{(p-l)}(\eta)| \, \mathrm{d}\eta,$$

and

$$G_l(\eta) = \frac{\sigma(\eta) - 1}{\eta^l}.$$

In other words, a discontinuity of f leads to a Fourier expansion with an error that is $O(1)$ near the discontinuity and $O(N^{-1})$ away from the discontinuity. We must therefore apply a shift and extend to $[0, 2\pi]$ before applying a filter.

2.2 The Algorithm

We now describe the postprocessing algorithm that is to be applied after each time step. (We do not explicitly indicate the dependence of u on the time step, and we use the abbreviations $u_0 = u(0)$ and $u_\pi = u(\pi)$.)

First, we apply a low-frequency shift,

$$v(x) = u(x) - (\alpha_1 + \alpha_2 \cos(x)), \quad \alpha_2 = \tfrac{1}{2}(u_0 - u_\pi),\ \alpha_1 = \tfrac{1}{2}(u_0 + u_\pi). \tag{7}$$

Then we extend v to $(0, 2\pi)$, using the definition

$$v(2\pi - x) = -v(x), \quad x \in (0, \pi). \tag{8}$$

Thus, v is a 2π-periodic function in $C^1(0, 2\pi)$. Let $\hat{v}_k$ be the kth coefficient of its Fourier expansion.

Next, we apply an eighth-order filter [GS97],

$$\sigma_N v(x) = \sum_k \sigma\left(\kappa \frac{k}{N}\right) \hat{v}_k e^{ikx}, \tag{9}$$

where

$$\sigma(\xi) = (35 - 84y + 70y^2 - 20y^3)y^4, \quad y \equiv y(\xi) = \tfrac{1}{2}(1 + \cos(\pi\xi)). \tag{10}$$

Here, κ is a stretching factor, $\kappa > 1$. The correct choice of κ follows from a Fourier analysis of Eq. (5),

$$\kappa > \kappa_c = \frac{\pi}{\cos(1 - 2h^2/(3\Delta t))}. \tag{11}$$

We observe that the filter still damps some of the high frequencies less than $\frac{N}{\kappa}$. The choice $\kappa = \frac{1}{2}\kappa_c$ can give satisfactory results, but in principle one can compute the optimum value of κ at each time step by monitoring the growth of the high-frequency waves that have not been completely filtered out.

Finally, we recover u from the inverse shift,

$$u(x) = \sigma_N v(x) + \alpha_1 + \alpha_2 \cos(x). \tag{12}$$

The theorem quoted in the preceding section shows that the filtering process may affect the spatial accuracy of the method. Since the filter is applied to a 2π-periodic function that is C^1 at the points $x_k = k\pi$, $k \in \mathbf{Z}$, and C^2 everywhere else, the error is of the order of N^{-2} in the neighborhood of x_k and N^{-3} away from x_k. In principle, we maintain therefore second-order accuracy in space as long as κ is of order one.

If the solution is in $C^3(0, \pi)$ at each time level, we can improve the algorithm by replacing the first-order shift (7) by a third-order shift,

$$v(x) = u(x) - \sum_{j=1}^{4} \alpha_j \cos((j-1)x), \tag{13}$$

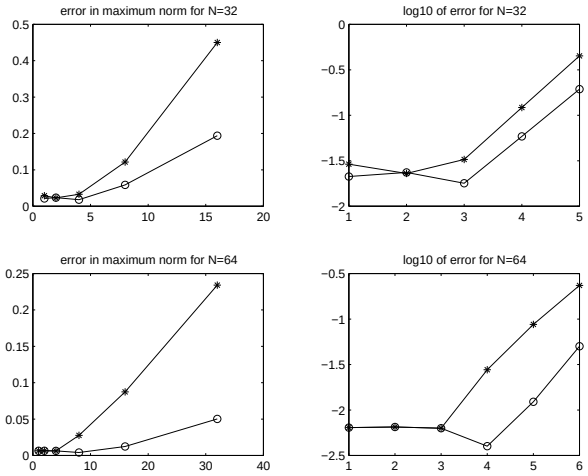


Figure 1: Accuracy of the stabilized explicit scheme for the heat equation. The horizontal coordinate is $3\Delta t/h^2$. * : first-order shift; o : third-order shift.

such that the extension of v to a 2π-periodic function is in $C^3(0, 2\pi)$. The first- and third-order derivatives of v are zero at the points x_k, and the second-order derivative is approximately given by

$$u_{xx}(x_k) \approx \frac{3u^{n+1}(x_k) - 4u^n(x_k) + u^{n-1}(x_k)}{2\Delta t} - f(u^{n+1}(x_k)). \tag{14}$$

The coefficients α_j are found by solving a linear system of equations,

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 &= u(0), \\ \alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 &= u(\pi), \\ -\alpha_2 - 4\alpha_3 - 9\alpha_4 &= u_{xx}(0), \\ \alpha_2 - 4\alpha_3 + 9\alpha_4 &= u_{xx}(\pi). \end{aligned}$$

The third-order shift improves the performance of the filter for large κ and allows for a larger time step.

2.3 Numerical Results

Figure 1 shows some accuracy results for Eq. (2), where

$$u(x,t) = \cos(t)((x/\pi)^4 + \cos(3x)), \quad x \in (0,\pi),\ t > 0. \tag{15}$$

We observe a plateau for small time steps, when the second-order spatial error dominates. The second-order error in time becomes dominant as the time step increases. The figure confirms the superior performance of the third-order shift (13) over the first-order shift (7) at large time steps.

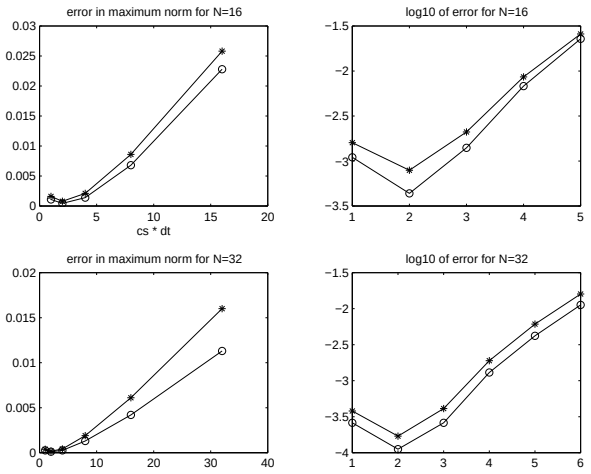


Figure 2: Accuracy of the stabilized explicit scheme for the predator-prey system. The horizontal coordinate is $3\Delta t/h^2$. * : first-order shift; o : third-order shift.

Although the algorithm is based only on linear stability considerations, it is still effective for systems of nonlinear reaction-diffusion equations. In Figure 2 we present some results for a predator-prey system,

$$\partial_t u = \partial_{xx} u + au - buv,\ \partial_t v = \partial_{xx} v - cu - duv, \quad x \in (0,\pi),\, t > 0, \tag{16}$$

with $a = 1.2$, $b = 1.0$, $c = 0.1$, and $d = 0.2$. At these parameter values, the ODE system (reactions only) has a limit cycle. However, when the boundary conditions are constant in time, the solution of the reaction-diffusion system goes to steady state. To build a relevant test case for the algorithm, we impose periodic excitations at both boundaries,

$$u(0/\pi, t) = u_{0/\pi}(1 + \cos(t)),\ v(0/\pi, t) = v_{0/\pi}(1 + \cos(t)), \quad t > 0.$$

Although the time step can still be limited by nonlinear instabilities, we never observed negative values of the unknowns u and v, which are commonly associated with such instabilities.

The algorithm (7)–(12) extends in a straightforward way when one uses a domain decomposition scheme with overlapping subdomains. One simply applies the same algorithm at each time step to each subdomain separately. However, the number of waves per subdomain is of the order of the total number of grid points, N, divided by the number of subdomains, N_d, so the balance between the order of accuracy of the filter— $(N/(\kappa N_d))^{-3}$ for a first-order filter or $(N/(\kappa N_d))^{-5}$ for a third-order filter—and the second-order accuracy N^{-2} of the spatial discretization of the underlying algorithm (3) deteriorates as N_d increases. The maximum time step for which the scheme remains stable may become even less than when no domain decomposition is used. Furthermore, the Gibbs phenomenon tends to destabilize the algorithm. This phenomenon is a consequence of the jump in the derivatives at the endpoints of the subdomains (second-order derivatives in the case of the first-order shift (7), fourth-order derivatives in the case of the third-order shift (13)). Since the Gibbs phenomenon arises at the artificial interface and is damped away from it, an increase of the overlap generally produces

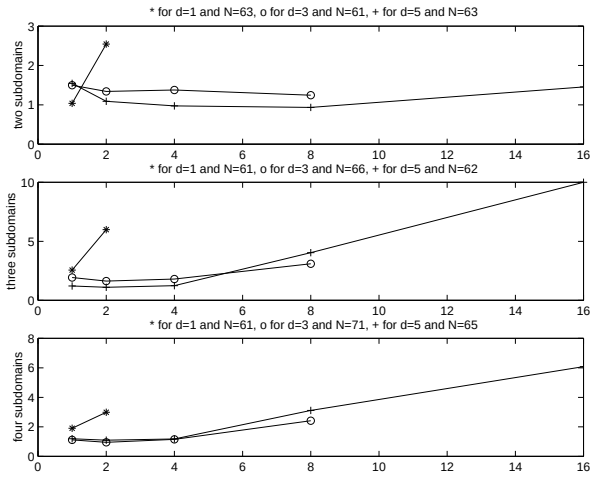


Figure 3: Relative Accuracy of the stabilized explicit scheme applied to the heat equation with overlapping subdomains.

a composite signal u that has fewer oscillations than each of the piecewise (overlapping) components. One can therefore obtain good results by adapting the size of the overlap. The larger the overlap, the larger the time step that can be taken; see Figure 3.

3 Two-Dimensional Domains

We now consider a Dirichlet problem in two dimensions,

$$\partial_t u = \Delta u + f(u), \quad (x,y) \in (0,\pi)^2,\ t > 0, \tag{17}$$

$$u(x,0/\pi) = g_{0/\pi}(x),\ u(0/\pi,y) = h_{0/\pi}(y), \quad x,y \in (0,\pi), \tag{18}$$

where the functions g satisfy the compatibility conditions $g_{0/\pi}(0) = h_0(0/\pi)$ and $g_{0/\pi}(\pi) = h_\pi(0/\pi)$. We consider a numerical scheme similar to Eq. (3), where the diffusive term is approximated, for example, by a five-point stencil. The postprocessing algorithm is essentially the same, except that we need an apropriate low-frequency shift so we can apply a filter to a smooth periodic function in two space dimensions. The shift is constructed in two steps. In the first step, we render the boundary condition in the x direction homogeneous,

$$v(x,y) = u(x,y) - (\alpha_1(y) + \alpha_2(y)\cos(x)), \tag{19}$$

$$\alpha_2(y) = \tfrac{1}{2}(g_0(y) - g_\pi(y)),\ \alpha_1(y) = \tfrac{1}{2}(g_0(y) + g_\pi(y)). \tag{20}$$

In the second step, we shift in the y direction,

$$w(x,y) = v(x,y) - (\beta_1(x) + \beta_2(x)\cos(y)), \tag{21}$$

$$\beta_2(x) = \tfrac{1}{2}(v(x,0) - v(x,\pi)),\ \beta_1(x) = \tfrac{1}{2}(v(x,0) + v(x,\pi)). \tag{22}$$

The final step is the reconstruction step,

$$u(x,y) = \sigma_N w(x,y) + \alpha_1(y) + \alpha_2(y)\cos(x) + \beta_1(x) + \beta_2(x)\cos(y). \tag{23}$$

To make sure that no high-frequency waves remain, we filter the high-frequency components from the boundary conditions g with a procedure similar to (7)–(12).

It is much more difficult to construct a high-order filter similar to (13) in two dimensions, because the second-order derivatives cannot be obtained from the PDE, as in the one-dimensional case (14). So far, we have used only the first-order shifts (19) and (21) in our numerical experiments. Nevertheless, the algorithm allows for a significant increase of the time step. We have also tested the domain-decomposition version of the algorithm, using strip subdomains with an adaptive overlap, with good results.

We note that the computation in each block can be done in parallel and that the Jacobi matrix does not depend on the spatial variables. The arithmetic complexity of the algorithm is therefore relatively small. Also, the algorithm is suitable for multicluster architectures. Each block can be assigned to a cluster, and parallel fast sine transforms can be used for the filtering process inside each cluster. The cost of communication between blocks is minimal, since the scheme is similar to the communication scheme of the additive Schwarz algorithm.

4 Conclusion

In this paper we have presented a postprocessing algorithm that stabilizes the time integration of systems of reaction-diffusion equations when the diffusion term is treated explicitly. The algorithm is easy to code and can be combined with domain-decomposition methods that use regular grids in each subblock. In future work, we will consider the performance of its parallel implementation and its robustness for large systems of reaction-diffusion equations with stiff chemistry, which arise in some air pollution models [FG01].

References

[DS96]D. Dabdub and J.H.Steinfeld, *Parallel Computation in Atmospheric Chemical Modeling,* Parallel Computing Vol22, 111-130, 1996.

[D90]J.J.Droux, *Simulation Numerique Bidimensionnelle et Tridimensionnelle de Processus de Solidification,* These No 901, L ausanne EPFL, 1990.

[L00]V.I.Lebedev, Explicit Difference Schemes for Solving Stiff Problems with a Complex or Separable Spectrum, Computational Mathematics and Mathematical Physics, Vol.40., No 12, 1801-1812, 2000.

[E97]H. Elbern, *Parallelization and Load Balancing of a Comprehensive Atmospheric Chemistry Transport Model,* Atmospheric Environment, Vol 31, No 21, 3561-3574, 1997.

[FG01]W.E.Fitzgibbon, M. Garbey, *Fast solver for Reaction-Diffusion-Convection Systems: application to air quality models* ECCOMAS 2001, 17pp, 2001.

[GS97]D. Gottlieb and Chi-Wang Shu, *On the Gibbs Phenomenon and its Resolution,* SIAM review, Vol 39, No 4, 644-668, 1997.

[P83]L. Petzold, *Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations,* SIAM J. Sci. Stat. Comput. Vol 4, No 1, March 1983.

[P89]O. Pironneau, *Finite Element Methods for Fluids,* Wiley, 1989.

[SVLCPDS97]A. Sandu, J.G. Verwer, M. Van Loon, G.R. Carmichael, F.A. Potra, D. Dadbud and J.H.Seinfeld, *Benchmarking Stiff ODE Solvers for Atmospheric Chemistry Problems I: implicit versus explicit,* Atm. Env. 31, 3151-3166, 1997.

[VS98]J.G.Verwer and B. Sportisse, *A Note on Operator Splitting in a Stiff Linear Case*, MAS-R9830, http://www.cwi.nl, Dec 98.

[VHB98]J.G.Verwer, W.H.Hundsdorfer and J.G.Blom, *Numerical Time Integration for Air Pollution Models*, MAS-R9825, http://www.cwi.nl, International Conference on Air Pollution Modelling and Simulation APMS'98.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

40 A Hierarchical Domain Decomposition Method with Low Communication Overhead

M. Israeli[1], E. Braverman[2], A. Averbuch[3]

1 Introduction

We present a low communication, non-iterative algorithm for a high order (spectral) solution of the Poisson equation. The domain is decomposed into nonoverlapping subdomains. Particular solutions are found in subdomains and subsequently hierarchically matched, such that only the solution in the adjacent subdomains are coupled at each matching step, then these joint subdomains are matched etc. If originally we had 2^{2k} subdomains, after k steps we obtain a smooth global solution.

Implicit discretization of time dependent problems in computational physics, semiconductor device simulation, electromigration and fluid dynamics often gives rise to equations of Poisson and modified Helmholtz type. Thus, fast and accurate methods for elliptic equations are important for such applications.

We solve the Poisson equation

$$\Delta u = f(x,y) \text{ in } \Omega \tag{1}$$

or the modified Helmholtz equation

$$\Delta u - \lambda^2 u = f(x,y) \text{ in } \Omega \tag{2}$$

in the rectangular/square domain $\Omega = [0,L] \times [0,K]$ with Dirichlet

$$u = \Phi(x,y) \text{ on } \partial\Omega \tag{3}$$

boundary conditions by the Domain Decomposition (DD) methods.

An algorithm for a fast solution of the Poisson equation by decomposition of the domain into square domains and the subsequent matching of these solutions by the fast multipole method was developed in [GL96]. Previously [ABI00] we adopted a DD method where the equation was solved in each subdomain with assumed boundary conditions, resulting in jumps in function or derivative on subdomain boundaries. The solution in each rectangular domain is fast and accurate and is based on the algorithm developed in [AIV97, AIV98]. The jumps at the interfaces were removed by the introduction of singularity layers. In order to account for the global effect of these layers we had to compute the influence of each layer on each subdomain boundary. In order to alleviate this heavy computational task we took into account the decay or smoothing out of the influence as a function of the distance from the layer. To

[1]Technion, Computer Science Dept., Haifa 32000, Israel, israeli@cs.technion.ac.il. The research of the first author was supported by the VPR fund for promotion of research at the Technion.

[2]Technion, Computer Science Dept., Haifa 32000, Israel, maelena@csd.technion.ac.il; now on leave in Yale University, Dept. of Math., 10 Hillhouse Ave., New Haven, CT 06520, USA, braverm@cyndra.cs.yale.edu

[3]Tel Aviv University, School of Math. Sciences, Tel Aviv 69978,Israel, amir@math.tau.ac.il

reduce the communication load, compression in a multiwavelet basis was applied. Nevertheless, this part of the procedure can become expensive as the number of subdomains grows considerably.

The algorithm developed in [ABI00] consists of the following steps:

1. In each subdomain a particular solution $u_1^{(s)}$ of the non-homogeneous equation with arbitrary Neumann (Dirichlet) boundary conditions is found.

2. The collection of particular solutions $u_1^{(s)},\ s = 1, \ldots, l$, usually have discontinuities (or discontinuities in the derivatives) on the boundaries of the subdomains. We introduce double (single) layers on the boundaries to match the solutions from different domains to have continuous global solution. The effect of these layers on other boundaries is calculated.

3. With the boundary conditions that were computed in the previous step, the solutions $u_1^{(s)}$ are patched by adding the solutions $u_2^{(s)},\ s = 1, \ldots, l$, of the Laplace equation.

4. An additional solution of the Laplace equation is added to satisfy the boundary conditions on $\partial\Omega$. Namely, for the Dirichlet case the solution u_3 of the homogeneous equation on the boundary $\partial\Omega$ is derived by

$$u_3(x, y) = \Phi(x, y) - u_1(x, y) - u_2(x, y) \tag{4}$$

(the case with Neumann boundary conditions is treated similarly). Thus $u = u_1 + u_2 + u_3$ is the solution of the non-homogeneous equation with the initial non-homogeneous boundary conditions.

The interface jump removal can become cheaper if only adjacent boxes are matched, which is a basis of the hierarchical approach which is proposed in the present paper. The present hierarchical approach matching only two adjacent boxes at each level requires only local corrections at the boundaries of these boxes. The result is a much more efficient computation.

2 Outline of the Algorithm

In the new hierarchical approach the domain is decomposed into k^2 subdomains; first (see Fig. 1) the smallest domains 1,2,3,4 are matched, then they are matched with larger blocks 5,6,7, and, finally, the resulting box is matched with 8,9,10.

The "elementary step" of the hierarchical algorithm is the following.

1. First, in each of four subdomains some smooth boundary conditions are defined. These conditions should not contradict the given right hand side, at the junctions. The Poisson equation is solved with these boundary conditions by a fast spectral algorithm which takes $O(N^2 \log N)$ operations (N is a number of points in each direction).

2. The solutions have a discontinuity in the first derivative. We match the subdomains by adding certain discontinuous functions. In fact we only evaluate these functions at the boundaries of four adjacent subdomains and then solve homogeneous equations in each subdomain with the cumulative boundary conditions.

3. The global homogeneous equation is solved in such a way that it satisfies the assumed conditions at the "global boundaries" of the merged subdomains.

This step is repeated $\log k$ times, for a smalled number of larger subdomains each time.

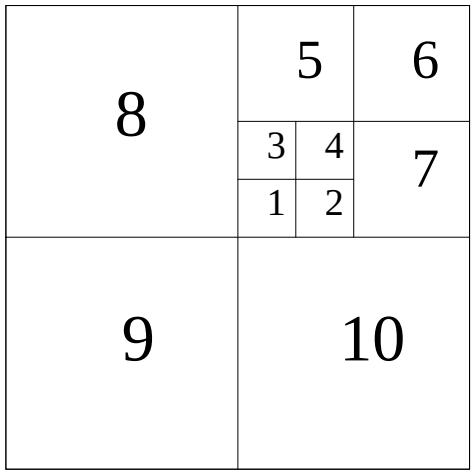


Figure 1: The domain is decomposed into k^2 subdomains; first the smallest domains 1,2,3,4 are matched, then they are matched with larger blocks 5,6,7, and, finally, the obtained box is matched with 8,9,10.

The algorithm can be also implemented on parallel multiprocessors. The parallelization of the serial algorithm is achieved by decomposition of the computational domain into smaller domains. Each domain is assigned to a processor. The information transmitted between the processors is the influences of a function/derivative jumps at the interfaces. of a function/derivative jumps at the interfaces. The low communication is achieved due to the fast decay of these influences and their efficient representation in multiwavelet bases.

For instance, when computing the influence of the derivative jump in the form of the sum of random Gaussians

$$\sum_{k=1}^{12} \exp\{-\alpha_k((x-x_k)^2+(y-y_k)^2)\}, \quad 0.2 \le \alpha_k \le 7,$$

at distance 3 we have only 4 (of 256) multiwavelet coefficients above 10^{-6}, 7 above 10^{-8}, 15 above 10^{-10}.

3 Matching step of the algorithm

The fast and efficient solution of the Poisson/modified Helmholtz and their homogeneous analogs was in detail described in [AIV97, AIV98]. Thus we focus here on the matching step of the algorithm.

Let us consider the simplest ("linear") geometry of the 2-D problem (see Fig. 2) and present the corresponding steps concerned either with the choice of the initial boundary conditions or with patching jumps between the subdomains.

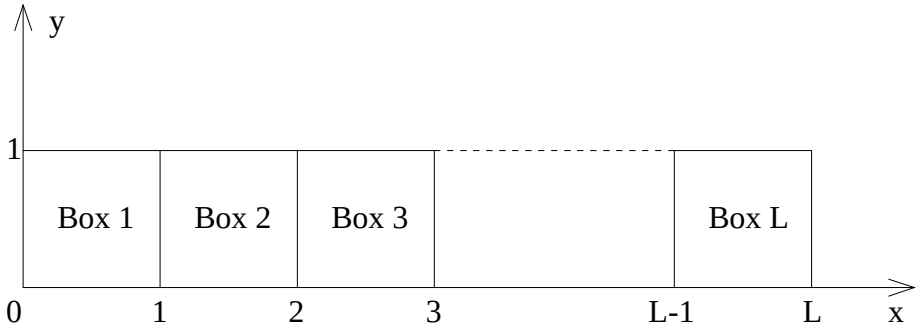


Figure 2: The domain is decomposed into L subdomains.

Step 1. At the boundary of the global domain we assume the original boundary conditions, to avoid singularities at the corners. At the interfaces $x = x_0,\ 0 < y < 1$ we assume

$$u(x_0, y) = u_1(x_0, y) + u_2(x_0, y), \tag{5}$$

where

$$u_1(x_0, y) = \Phi(x_0, 0)\frac{\sinh(\lambda(1-y))}{\sinh(\lambda)} + \Phi(x_0, 1)\frac{\sinh(\lambda y)}{\sinh(\lambda)} \tag{6}$$

matches the values at the interfaces with the value at the boundary and

$$u_2(x_0, y) = \frac{f(x_0, 0) - \Phi(x_0, 0)\lambda^2}{\lambda_1^2 - \lambda_2^2)}\left[\frac{\sinh(\lambda_1(1-y))}{\sinh(\lambda_1)} - \frac{\sinh(\lambda_2(1-y))}{\sinh(\lambda_2)}\right]$$

$$+\frac{f(x_0, 1) - \Phi(x_0, 1)\lambda^2}{\lambda_1^2 - \lambda_2^2)}\left[\frac{\sinh(\lambda_1 y)}{\sinh(\lambda_1)} - \frac{\sinh(\lambda_2 y)}{\sinh(\lambda_2)}\right] \tag{7}$$

to satisfy the Poisson equation at the corners of the interfaces. The latter function vanishes at $x = x_0, y = 0, 1$.

Step 2. We solve the Poisson equation with the prescribed (at the first step) boundary conditions. There is a jump of the first derivative at the interfaces

$$\frac{\partial u}{\partial x}(x_0+, y) - \frac{\partial u}{\partial x}(x_0-, y) = h(y). \tag{8}$$

Since the original boundary conditions are smooth, then $h(0) = h(1) = 0$. After subtracting a function

$$g(y) = \frac{h''(1)}{\lambda_1^2 - \lambda_2^2}\left[\frac{\sin(\lambda_1 y)}{\sin(\lambda_1)} - \frac{\sin(\lambda_2 y)}{\sin(\lambda_2)}\right] \tag{9}$$

$h(y)$ vanishes at $y = 1$ together with its second derivative. A similar function is subtracted for $y = 0$. Then the remaining part $\tilde{h}$ can be accurately expanded into the sine series (in fact, the fourth and higher even derivatives can be also eliminated by an analogous procedure)

$$\tilde{h}(y) = \sum a_k \sin(\pi k y) \tag{10}$$

Then, after adding to the solution to the left of $x = x_0$ the following function

$$\begin{aligned}
&\frac{h''(1)}{2(\lambda_1^2 - \lambda_2^2)} \left[\frac{\cosh(\lambda_1(x - x_0 + L))}{\lambda_1 \sinh(\lambda_1 L)} \frac{\sin(\lambda_1 y)}{\sin(\lambda_1)} - \frac{\sinh(\lambda_2(x - x_0 + L))}{\cosh(\lambda_2 L)} \frac{\sin(\lambda_1 y)}{\sin(\lambda_1)} \right] \\
&\quad + \frac{h''(0)}{2(\lambda_1^2 - \lambda_2^2)} \left[\frac{\cosh(\lambda_1(x - x_0 + L))}{\lambda_1 \sinh(\lambda_1 L)} \frac{\sin(\lambda_1(1 - y))}{\sin(\lambda_1)} \right. \\
&\quad \left. - \frac{\sinh(\lambda_2(x - x_0 + L))}{\cosh(\lambda_2 L)} \frac{\sin(\lambda_1(1 - y))}{\sin(\lambda_1)} \right] \\
&\quad + \frac{1}{2\pi k \cosh(\pi k L)} \sum a_k \sin(\pi k y) \cosh(\pi k(L - x_0 + x))
\end{aligned} \tag{11}$$

and a symmetric (with respect to axis $x = x_0$) to the right of this axis, we obtain a function which is smooth together with its first derivative. Besides, the function which we add decays exponentially with the growth of the distance from $x = x_0$.

4 Numerical Results

Assume that u is the exact solution and u' is the computed solution. In the examples we will use the following measures to estimate the errors:

$$\begin{aligned}
\varepsilon_{MAX} &= \max \|u'_i - u_i\| \\
\varepsilon_{MSQ} &= \sqrt{\frac{\sum_{i=1}^{N} (u'_i - u_i)^2}{n}} \\
\varepsilon_{\mathcal{L}^2} &= \sqrt{\frac{\sum_{i=1}^{N} (u'_i - u_i)^2}{\sum_{i=1}^{N} u_i^2}}
\end{aligned}$$

4.1 Linear geometry

We assume the geometry of Fig. 2 where the domain is decomposed in one dimension only.

Example 1. We solve the Poisson equation $\Delta u = -2 \cos x \cos y$ with the boundary conditions corresponding to the exact solution $u(x, y, z) = \cos x \cos y$ in the domain $[0, 3] \times [0, 1]$ which is divided into three equal boxes.

$N_x \times N_y$ in each box	ε_{MAX}	ε_{MSQ}	$\varepsilon_{\mathcal{L}^2}$
32×32	4.1e-7	1.2e-7	2.1e-7
64×64	2.9e-8	8.2e-9	1.4e-8
128×128	2.0e-9	5.4e-10	9.2e-10
256×256	1.3e-10	3.5e-11	5.9e-11
512×512	8.5e-12	2.2e-12	3.8e-12

Table 1: MAX, MSQ and $\mathcal{L}^2$ errors for the Poisson equation with the exact solution $u(x, y, z) = \cos x \cos y$ for three boxes

Example 2. We solve the Poisson equation with boundary conditions corresponding to the exact solution

$$u(x,y) = \exp\left\{-\alpha\left((x-x_0)^2 + (y-y_0)^2\right)\right\},$$

with $x_0 = 1.5,\ y_0 = 0.5, \alpha = 2$ in the domain $[0,3] \times [0,1]$ which is divided into three equal boxes.

$N_x \times N_y$ in each box	ε_{MAX}	ε_{MSQ}	$\varepsilon_{\mathcal{L}^2}$
32×32	6.8e-6	2.0e-6	4.4e-6
64×64	3.9e-7	1.1e-7	2.4e-7
128×128	2.3e-8	6.6e-9	1.4e-8
256×256	1.4e-9	4.0e-10	8.5e-10
512×512	8.7e-11	2.4e-11	5.2e-11

Table 2: MAX, MSQ and $\mathcal{L}^2$ errors for the Poisson equation with the exact solution $u(x,y) = \exp\{-2(x-1.5)^2 + 2(y-0.5)^2\}$.

4.2 Hierarchical subdomains matching

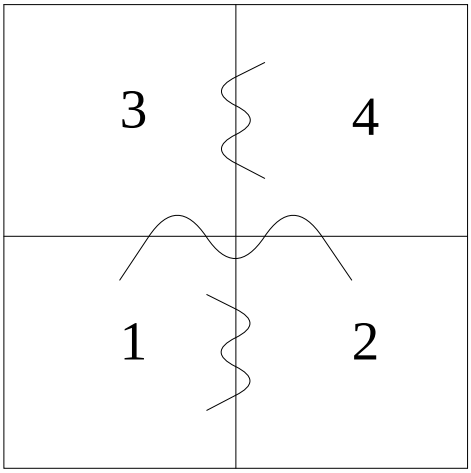


Figure 3: The domain is decomposed into four subdomains

In examples 3, 4 the global subdomain was decomposed into four subdomains (see Fig. 3). In the practical implementation first two pairs of adjacent subdomains were matched: box 1 and 2, box 3 and 4. Afterwards the two resulting boxes were patched. This is also valid for examples 5,6 with 16 subdomains, where each four subdomains were matched in the same way.

Example 3. We solve the Poisson equation with boundary conditions corresponding to the exact solution

$$u(x,y) = \exp\left\{-\alpha\left((x-x_0)^2 + (y-y_0)^2\right)\right\},$$

with $x_0 = 0.5,\ y_0 = 0.5, \alpha = 2$ in the domain $[0,1] \times [0,1]$ divided into four equal boxes.

$N_x \times N_y$ in each subdomain	ε_{MAX}	ε_{MSQ}	$\varepsilon_{\mathcal{L}^2}$
8×8	1.7e-4	5.1e-5	7.1e-5
16×16	1.3e-5	3.6e-6	4.9e-6
32×32	1.2e-6	3.5e-7	4.7e-7
64×64	1.0e-7	2.8e-8	3.7e-8
128×128	8.0e-9	2.0e-9	2.7e-9
256×256	6.1e-10	1.4e-10	1.9e-10

Table 3: MAX, MSQ and $\mathcal{L}^2$ errors for the Poisson equation with the exact solution $u(x,y) = \exp\{2((x-0.5)^2 + 2(y-0.5)^2)\}$. in the domain $[0,1] \times [0,1]$

Table 4 presents the results for the same exact solution when the Dirichlet problem is solved in the square $[0,2] \times [0,2]$.

$N_x \times N_y$ in each subdomain	ε_{MAX}	ε_{MSQ}	$\varepsilon_{\mathcal{L}^2}$
8×8	3.5e-3	1.2e-3	3.1e-3
16×16	1.7e-4	5.7e-5	1.4e-4
32×32	8.3e-6	2.5e-6	6.1e-6
64×64	4.4e-7	1.2e-7	2.9e-7
128×128	2.5e-8	6.5e-9	1.6e-8
256×256	1.5e-9	3.8e-10	9.4e-10

Table 4: MAX, MSQ and $\mathcal{L}^2$ errors for the Poisson equation with the exact solution $u(x,y) = \exp\{-2((x-0.5)^2 + 2(y-0.5)^2)\}$ in the domain $[0,2] \times [0,2]$

Example 4. The exact solution is a steep Gaussian

$$u(x,y) = \exp\left\{-10\left((x-0.2)^2 + (y-0.3)^2\right)\right\}.$$

Table 5 presents the numerical errors.

$N_x \times N_y$ in each subdomain	ε_{MAX}	ε_{MSQ}	$\varepsilon_{\mathcal{L}^2}$
8×8	1.9e-2	4.4e-3	2.4e-2
16×16	4.9e-4	9.1e-5	4.9e-4
32×32	1.1e-5	2.1e-6	1.1e-5
64×64	3.0e-7	6.0e-8	3.2e-7
128×128	1.1e-8	1.9e-9	1.0e-8
256×256	4.7e-10	6.4e-11	3.5e-10

Table 5: MAX, MSQ and $\mathcal{L}^2$ errors for the Poisson equation with the exact solution $u(x,y) = \exp\{-10((x-0.2)^2 + 2(y-0.3)^2)\}$. in the domain $[0,2] \times [0,2]$

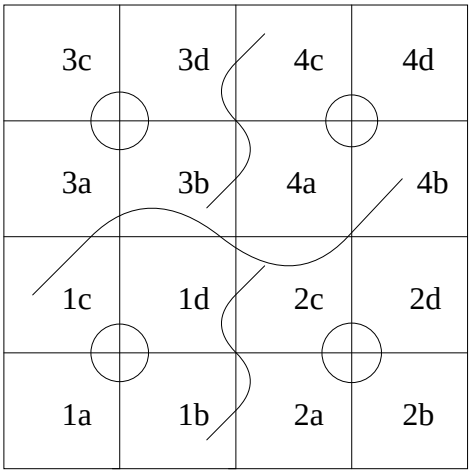


Figure 4: The domain is decomposed into sixteen subdomains

The domain is decomposed into sixteen subdomains which are hierarchically matched (see Fig. 4): first each four small boxes and then the resulting four "big" (joint) boxes. Such domain decomposition was implemented in Example 5.

Example 5. We solve the Poisson equation with the boundary conditions corresponding to the exact solution

$$u(x,y) = \exp\left\{-2\left((x-0.5)^2 + (y-0.5)^2\right)\right\}$$

in the domain $[0,1] \times [0,1]$ divided into sixteen equal boxes (Table 6).

$N_x \times N_y$ in each subdomain	ε_{MAX}	ε_{MSQ}	$\varepsilon_{\mathcal{L}^2}$
4×4	3.6e-4	1.6e-4	2.2e-4
8×8	2.4e-5	1.0e-5	1.4e-5
16×16	2.3e-6	9.6e-7	1.3e-6
32×32	1.9e-7	8.0e-8	1.1e-7
64×64	1.5e-8	6.1e-9	8.2e-9
128×128	1.1e-9	4.5e-10	6.0e-10
256×256	8.3e-11	3.3e-11	4.4e-11

Table 6: MAX, MSQ and $\mathcal{L}^2$ errors for the Poisson equation with the exact solution $u(x,y) = \exp\{-2((x-0.5)^2 + 2(y-0.5)^2)\}$ in the domain $[0,1] \times [0,1]$

5 Summary

1. The procedure developed here reduces drastically (by a factor of $O(k^2/\log k)$ times) the number of computations as compared to our previous algorithm where influences of the layers at interfaces are evaluated.

2. The algorithm has an adaptive DD version and also achieves high accuracy ($10^{-7} - 10^{-8}$ for 64×64 points in the smallest subdomains).

3. The algorithm is applicable for parallel implementation as its previous version developed in [ABI00].

4. This algorithm can be used as a preconditioner for the solution of elliptic equations with nonconstant coefficients by solving local constant coefficient problems in subdomains.

5. The present algorithm is close to a multigrid strategy, where the discretization points are replaced by the small boxes in where the equation is satisfied with spectral accuracy which is preserved in the final solution.

References

[ABI00]Amir Averbuch, Elena Braverman, and Moshe Israeli. Parallel adaptive solution of a Poisson equation with multiwavelets. *SIAM J. Sci. Comput.*, 22(3):1053–1086, 2000.

[AIV97]Amir Averbuch, Moshe Israeli, and Lev Vozovoi. On fast direct elliptic solver by modified Fourier method. *Numer. Algorithms*, 15:287–313, 1997.

[AIV98]Amir Averbuch, Moshe Israeli, and Lev Vozovoi. A fast Poisson solver of arbitrary order accuracy in rectangular regions. *SIAM J. Sci. Comput.*, 19:933–952, 1998.

[GL96]Leslie Greengard and June-Yub Lee. A direct adaptive Poisson solver of arbitrary order accuracy. *J. Comput. Phys.*, 125:415–424, 1996.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

41 FETI-DP Methods for Elliptic Problems with Discontinuous Coefficients in Three Dimensions

Axel Klawonn[1], Olof B. Widlund[2]

Introduction

Farhat, Lesoinne, Le Tallec, Pierson, and Rixen [FLLT$^+$01] have recently introduced a dual–primal FETI (FETI-DP) algorithm suitable for second order elliptic problems in the plane and for plate problems. A convergence analysis in the case of benign coefficients is given by Mandel and Tezaur [MT01]. Numerical experiments show a poor performance for this algorithm in three dimensions; cf. [FLLT$^+$01]. Recent experiments with alternative algorithms are reported in [FLP00, Pie00]. We give a brief description of our own recent work in the third section; see [KWD01] for many more details.

The remainder of this paper is organized as follows. In the next section , we introduce our elliptic problems and the basic geometry of the decomposition. In the third section, we present results on new dual–primal FETI methods for problems with discontinuous coefficient in three dimensions; see [KWD01].

Elliptic model problem, finite elements, and geometry

Let $\Omega \subset \mathbf{R}^3$, be a bounded, polyhedral region, let $\partial\Omega_D \subset \partial\Omega$ be a closed set of positive measure, and let $\partial\Omega_N := \partial\Omega \setminus \partial\Omega_D$ be its complement. We impose homogeneous Dirichlet and general Neumann boundary conditions, respectively, on these two subsets and introduce the Sobolev space $H^1_0(\Omega, \partial\Omega_D) := \{v \in H^1(\Omega) : v = 0 \text{ on } \partial\Omega_D\}$.

We decompose Ω into non-overlapping subdomains $\Omega_i, i = 1, \ldots, N$, also known as substructures, and each of which is the union of shape-regular elements with the finite element nodes on the boundaries of neighboring subdomains matching across the interface $\Gamma := \overline{\left(\bigcup_{i=1}^N \partial\Omega_i\right) \setminus \partial\Omega}$. The interface Γ is decomposed into subdomain faces, regarded as open sets, which are shared by two subregions, edges which are shared by more than two subregions and the vertices which form the endpoints of edges. We denote faces of Ω_i by $\mathcal{F}^{ij}$, edges by $\mathcal{E}^{ik}$, and vertices by $\mathcal{V}^{i\ell}$.

For simplicity, we will only consider a piecewise linear, conforming finite element approximation of the following scalar, second order model problem:

[1]Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, D–53754 Sankt Augustin, Germany. E-mail: klawonn@scai.fraunhofer.de, URL: http://www.scai.fraunhofer.de/people/klawonn.html. This work was supported in part by the National Science Foundation under Grants NSF-CCR-9732208.

[2]Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA. E-mail: widlund@cs.nyu.edu, URL: http://www.cs.nyu.edu/cs/faculty/widlund. This work was supported in part by the National Science Foundation under Grants NSF-CCR-9732208 and in part by the US Department of Energy under Contract DE-FG02-92ER25127.

Find $u \in H_0^1(\Omega, \partial\Omega_D)$, such that

$$a(u,v) = f(v) \quad \forall v \in H_0^1(\Omega, \partial\Omega_D), \tag{1}$$

where

$$a(u,v) = \sum_{i=1}^{N} \rho_i \int_{\Omega_i} \nabla u \cdot \nabla v dx, \quad f(v) = \sum_{i=1}^{N} \Big(\int_{\Omega_i} f v dx + \int_{\partial\Omega_i \cap \partial\Omega_N} g_N v ds \Big). \tag{2}$$

where g_N is the Neumann boundary data defined on $\partial\Omega_N$; it provides a contribution to the load vector of the finite element problem. We assume that ρ_i is a positive constant on each subregion Ω_i.

In our theoretical analysis, we assume that each subregion Ω_i is the union of a number of shape regular tetrahedral coarse elements and that the number of such tetrahedra is uniformly bounded for each subdomain. Thus, the subregions are not very thin and we can also easily show that the diameters of any pair of neighboring subdomains are comparable.

We also make a number of technical assumptions on the intersection of the boundary of the substructures and $\partial\Omega_D$; see [KWD01]. The sets of nodes in Ω_i, on $\partial\Omega_i$, and on Γ are denoted by $\Omega_{i,h}, \partial\Omega_{i,h}$, and Γ_h, respectively.

We denote the standard finite element space of continuous, piecewise linear functions on Ω_i by $W^h(\Omega_i)$. For simplicity, we assume that the triangulation of each subdomain is quasi uniform. The diameter of Ω_i is H_i, or generically, H. We denote the corresponding finite element trace spaces by $W_i := W^h(\partial\Omega_i), i = 1, \ldots, N$, and by $W := \prod_{i=1}^{N} W_i$ the associated product space. We will often consider elements of W which are discontinuous across the interface.

The finite element approximation of the elliptic problem is continuous across Γ and we denote the corresponding subspace of W by $\widehat{W}$. We note that while the stiffness matrix K and its Schur complement S, obtained from K by elimination of the interior subdomain variables, which both correspond to the product space W generally are singular those of $\widehat{W}$ are not.

For the dual–primal FETI methods, we will use additional, intermediate subspaces $\widetilde{W}$ of W for which a relatively small number of continuity constraints are enforced across the interface throughout the iteration. One of the benefits of working in $\widetilde{W}$, rather than in W, is that certain related Schur complements $\widetilde{S}$ and S_Δ are positive definite.

As in previous work on Neumann–Neumann and FETI algorithms, a crucial role is played by *the weighted counting functions* $\mu_i \in \widehat{W}$, which are associated with the individual subdomain boundaries $\partial\Omega_i$; cf., e.g., [DSW96, DW95]. In present context they will be used in the definition of certain diagonal scaling matrices. These functions are defined, for $\gamma \in [1/2, \infty)$, and for $x \in \Gamma_h \cup \partial\Omega_h$, by a sum of contributions from Ω_i, and its relevant next neighbors

$$\mu_i(x) = \begin{cases} \sum\limits_{j \in \mathcal{N}_x} \rho_j^\gamma(x) & x \in \partial\Omega_{i,h} \cap \partial\Omega_{j,h}, \\ \rho_i^\gamma(x) & x \in \partial\Omega_{i,h} \cap (\partial\Omega_h \setminus \Gamma_h), \\ 0 & x \in (\Gamma_h \cup \partial\Omega_h) \setminus \partial\Omega_{i,h}. \end{cases} \tag{3}$$

Here, $\mathcal{N}_x$ is the set of indices of the subregions which have x on its boundary. We note that any node of Γ_h belongs either to two faces, more than two edges, or to the vertices of several substructures.

The pseudo inverses $\mu_i^\dagger$ are defined, for $x \in \Gamma_h \cup \partial\Omega_h$, by

$$\mu_i^\dagger(x) = \begin{cases} \mu_i^{-1}(x) & \text{if } \mu_i(x) \neq 0, \\ 0 & \text{if } \mu_i(x) = 0. \end{cases}$$

New Dual–Primal FETI methods

In previous studies of dual–primal FETI methods for problems in two dimensions, see Farhat, Lesoinne, Le Tallec, Pierson, and Rixen [FLLT$^+$01] and Mandel and Tezaur [MT01], the constraints on the degrees of freedom associated with the vertices of the substructures are enforced, i.e., the corresponding degrees of freedom have been added to the primal set of variables, while all the constraints associated with the edge nodes are enforced only at the convergence of the iterative method. In each step of the iteration a fully assembled linear subsystem is solved. In a simple two–dimensional case, this subsystem corresponds to all the interior and cross point variables; these variables can be eliminated at a modest expense since we can first eliminate all the interior variables, in parallel across the subdomains, resulting in a Schur complement for the cross point variables which can be shown to be sparse. It has a dimension which equals the number of subdomain vertices which do not belong to $\partial\Omega_D$.

In their recent paper, Mandel and Tezaur [MT01] established a condition number bound of the form $C(1 + \log(H/h))^2$ for the resulting FETI method equipped with a Dirichlet preconditioner which is very similar to those used for the older FETI methods and which is built from local solvers on the subregions with zero Dirichlet conditions at the vertices of the subregions. They also established a corresponding result for a fourth-order elliptic problem in the plane.

The same algorithm is also defined for three dimensions but it does not perform well. This is undoubtedly related to the poor performance of many vertex-based iterative substructuring methods; see [DSW94, Section 6.1] and [KWD01]. Recently, Farhat, Lesoinne, and Pierson added edge and face constraints to this basic algorithm, see [FLP00], and improved the performance.

In the present study, as well as in others of FETI–DP methods, it is convenient to work in subspaces $\widetilde{W} \subset W$ for which sufficiently many constraints are enforced so that the resulting leading diagonal block matrix of the saddle point problem, though no longer block diagonal, is strictly positive definite. We will explain how this can be accomplished and also introduce two subspaces, $\widehat{W}_\Pi \subset \widehat{W}$ and $\widetilde{W}_\Delta$, corresponding to a primal and a dual part of the space $\widetilde{W}$. These subspaces will play an important role in the description and analysis of our iterative method. The direct sum of these spaces equals $\widetilde{W}$, i.e.,

$$\widetilde{W} = \widehat{W}_\Pi \oplus \widetilde{W}_\Delta. \tag{4}$$

The second subspace, $\widetilde{W}_\Delta$, is the direct sum of local subspaces $\widetilde{W}_{\Delta,i}$ of $\widetilde{W}$ where each subdomain Ω_i contributes a subspace $\widetilde{W}_{\Delta,i}$; only its $i-th$ component in the sense of the product space $\widetilde{W}$ is nontrivial.

In the description of our algorithms, we will need certain standard finite element cutoff functions $\theta_{\mathcal{E}^{ik}}$, $\theta_{\mathcal{F}^{ij}}$, and $\theta_{\mathcal{V}^{i\ell}}$. The first two are the discrete harmonic functions which equal 1 on $\mathcal{E}_h^{ik}$ and $\mathcal{F}_h^{ij}$, respectively, and which vanish elsewhere on Γ_h; $\theta_{\mathcal{V}^{i\ell}}$ denotes the piecewise discrete harmonic extension of the standard nodal basis function associated with the vertex $\mathcal{V}^{i\ell}$. These cutoff functions are also used in the analysis of the methods; see [KWD01].

We are now ready to describe our algorithms in terms of pairs of subspaces.

Algorithm A: The primal subspace, $\widehat{W}_\Pi$, is spanned by the nodal finite element basis functions $\theta_{\mathcal{V}^{i\ell}}$. The local subspace $\widetilde{W}_{\Delta,i}$ is defined in terms of the subspace of W_i of elements which vanish at the subdomain vertices, i.e., by

$$\widetilde{W}_{\Delta,i} := \{u \in W_i : u(\mathcal{V}^{i\ell}) = 0 \ \forall \mathcal{V}^{i\ell} \in \partial\Omega_i\}.$$

Hence, $\widetilde{W}$ is the subspace of W of functions that are continuous at the subdomain vertices.

Algorithm B: The primal subspace, $\widehat{W}_\Pi$, is spanned by the vertex nodal finite element basis functions $\theta_{\mathcal{V}^{i\ell}}$ and the cutoff functions $\theta_{\mathcal{E}^{ik}}$ and $\theta_{\mathcal{F}^{ij}}$ associated with all the individual edges and faces, respectively, of the interface. The local subspaces $\widetilde{W}_{\Delta,i}$ are defined as the subspaces of W_i where the values at the subdomain vertices vanish together with the averages $\overline{u}_{\mathcal{E}^{ik}}$ and $\overline{u}_{\mathcal{F}^{ij}}$, i.e., by

$$\widetilde{W}_{\Delta,i} := \{u \in W_i : u(\mathcal{V}^{i\ell}) = 0, \overline{u}_{\mathcal{E}^{ik}} = 0, \overline{u}_{\mathcal{F}^{ij}} = 0 \ \forall \mathcal{V}^{i\ell}, \mathcal{E}^{ik}, \mathcal{F}^{ij} \subset \partial\Omega_i\}.$$

Hence, $\widetilde{W}$ is the subspace of W of functions that are continuous at the subdomain vertices and have the same averages $\overline{u}_{\mathcal{E}^{ik}}$ and $\overline{u}_{\mathcal{F}^{ij}}$ independently of which component of $u \in \widetilde{W}$ is used in the the evaluation of these averages. Here the averages $\bar{u}_{\mathcal{E}^{ik}}$ and $\bar{u}_{\mathcal{F}^{ij}}$, which by assumption take on unique values $\forall u_h \in \widetilde{W}$, are defined by,

$$\bar{u}_{\mathcal{E}^{ik}} = \frac{\int_{\mathcal{E}^{ik}} u ds}{\int_{\mathcal{E}^{ik}} 1 ds} \quad \text{and} \quad \bar{u}_{\mathcal{F}^{ij}} = \frac{\int_{\mathcal{F}^{ij}} u dx}{\int_{\mathcal{F}^{ij}} 1 dx}. \tag{5}$$

Algorithm C: The primal subspace, $\widehat{W}_\Pi$, is spanned by the vertex nodal finite element basis functions $\theta_{\mathcal{V}^{i\ell}}$ and the cutoff functions $\theta_{\mathcal{E}^{ik}}$ defined on all the edges of Γ. The local subspaces $\widetilde{W}_{\Delta,i}$ are defined as the subspaces of W_i where the values at the subdomain vertices vanish together with the averages $\overline{u}_{\mathcal{E}^{ik}}$, i.e., by

$$\widetilde{W}_{\Delta,i} := \{u \in W_i : u(\mathcal{V}^{i\ell}) = 0, \overline{u}_{\mathcal{E}^{ik}} = 0, \ \forall \mathcal{V}^{i\ell}, \mathcal{E}^{ik} \subset \partial\Omega_i\}.$$

Hence, $\widetilde{W}$ is the subspace of W of functions that are continuous at the subdomain vertices and have common averages $\overline{u}_{\mathcal{E}^{ik}}$ for the individual edges. The number of degrees of freedom of the corresponding primal subspace $\widehat{W}_\Pi$ is therefore equal to the sum of the number of vertices and the number of edges; this $\widehat{W}_\Pi$ will be of lower dimension than the primal space of Algorithm B.

The number of constraints enforced in all the iterations of Algorithms B and C is substantially larger than when only the vertex constraints are satisfied as in Algorithm A, but we are still able to work with a uniformly bounded number of such constraints for each substructure. In order to put this in perspective, we consider Algorithms B and C in the very regular case of cubic substructures. There are then seven global variables for each interior substructure in the case of Algorithm B since there are eight vertices, each shared by eight cubes, twelve edges, each shared by four, and six faces each shared by a pair of substructures. The count for Algorithm C is four. We note that the counts would be different, relative to the number of substructures, in the case of tetrahedral subregions.

It is useful to distinguish between the continuity constraints at the vertices and the other constraints. The latter are sometimes called optional constraints since they are not needed to

guarantee solvability of the subproblems if there are enough vertex constraints. The vertex constraints are enforced in the subassembly process, for the primal problem, outlined above. The optional constraints could be similarly incorporated after a change of variables. Another possibility, which we advocate, is to introduce an additional set of Lagrange multipliers which are computed exactly in each iteration to enforce the required optional constraints of the primal subspace; see Farhat, Lesoinne, and Pierson [FLP00], where this approach is used. For a more detailed description of this approach, we refer to section 4.2, especially formulae (24)-(28), of that paper.

We are able to show as strong a result for Algorithm C as for Algorithm B. It is therefore natural to attempt to drop additional constraints, i.e., further decrease the primal subspace $\widetilde{W}_\Pi$ while attempting to preserve the fast convergence of the FETI-DP method. This leads to the introduction of our final algorithm.

Algorithm D: The primal subspace $\widehat{W}_\Pi$, is defined in terms of constraints associated with a subset of the edges and vertices of the interface. We first describe the requirements on a minimal set of primal constraints which we have found necessary to give a complete proof of a good bound for Algorithm D. For each face, we should have at least one designated, primal edge. Additionally, for all pairs of substructures Ω_i, Ω_j, which have an edge in common, we must have an acceptable *edge path* between the two subdomains. An acceptable edge path is a path from Ω_i to Ω_j, possibly via several other subdomains, Ω_k, which have the edge $\mathcal{E}^{ij}$ in common and such that their coefficients satisfy $TOL * \rho_k \geq \min(\rho_i, \rho_j)$ for some chosen tolerance TOL. The path can only pass from one subdomain to another through an edge designated as primal. Finally, we consider all pairs of substructures which have a vertex $\mathcal{V}^{i\ell}$ but not a face or an edge in common. Then, we assume that either $\mathcal{V}^{i\ell}$ is a primal vertex or that we have an acceptable edge path of the same nature as above, except that we can be more lenient and only insist on $TOL * \rho_k \geq (h_k/H_k)\min(\rho_i, \rho_j)$. A possible algorithm of selecting the set of primal constraints is given in [KWD01].

We can now formulate our FETI–DP algorithms. The primal part of the algorithm is based on the exact elimination of all unknowns of the primal subspace as well as the interior variables. The remaining system is written in terms of a Schur complement $\widetilde{S}$. Thus, for all the algorithms, we arrive at this reduced problem after eliminating the primal variables associated with the interior nodes, the vertex nodes designated as primal, as well as the Lagrange multipliers related to the optional constraints. This Schur complement $\widetilde{S}$ can also be defined in terms of a minimum property; cf. [KWD01]. Analogously, we get from the load vectors associated with each subdomain a reduced right hand side $\tilde{f}_\Delta$. We can now reformulate the original finite element problem, reduced to the degrees of freedom of the second subspace $\widetilde{W}_\Delta$, as a minimization problem with constraints given by the requirement of continuity across Γ_h:

Find $u_\Delta \in \widetilde{W}_\Delta$, such that

$$\left.\begin{array}{c} J(u_\Delta) := \frac{1}{2}\langle \widetilde{S}u_\Delta, u_\Delta\rangle - \langle \tilde{f}_\Delta, u_d\rangle \rightarrow \min \\ B_\Delta u_\Delta = 0 \end{array}\right\}. \tag{6}$$

The matrix B_Δ is constructed from $\{0, 1, -1\}$ such that the values of the solution u_Δ, associated with more than one subdomain, coincide when $B_\Delta u_\Delta = 0$. These constraints are very simple and just express that the nodal values coincide across the interface; in comparison with the one-level FETI method, see, e.g., [KW01], we can drop some of the constraints, in particular those associated with the vertex nodes of the primal space. However, we will otherwise

use all possible constraints and thus work with a fully redundant set of Lagrange multipliers as in [KW01, section 5].

By introducing a set of Lagrange multipliers $\lambda \in V := range\,(B_\Delta)$, to enforce the constraints $B_\Delta u_\Delta = 0$, we obtain a saddle point formulation of (6), which is similar to that of the one-level FETI method; see, e.g., Klawonn and Widlund [KW01]. We use that $\widetilde{S}$ is invertible and eliminate the subvector u_Δ, and obtain the following system for the dual variable:

$$F\lambda = d, \tag{7}$$

where

$$F := B_\Delta \widetilde{S}^{-1} B_\Delta^t$$

and the right hand side

$$d := B_\Delta \widetilde{S}^{-1} \tilde{f}_\Delta.$$

To define the FETI–DP Dirichlet preconditioner, we need to introduce an additional set of Schur complement matrices,

$$S_\Delta^{(i)} := K_{\Delta\Delta}^{(i)} - K_{\Delta I}^{(i)} (K_{II}^{(i)})^{-1} K_{I\Delta}^{(i)}, \quad i = 1, \ldots, N,$$

Here, $K_{\Delta\Delta}^{(i)}$ is the principal minor of the stiffness matrix after the change of variables and it is related to the variables of $\widetilde{W}_\Delta$. The associated block–diagonal matrix is denoted by

$$S_\Delta := diag_{i=1}^N (S_\Delta^{(i)}).$$

We can compute the action of S_Δ on a vector from the second subspace $\widetilde{W}_\Delta$ by solving local problems with solutions that are constrained to vanish or to have zero average at the designated, primal variables, as required by the algorithm in question; these constraints can be enforced by using Lagrange multipliers or a partial change of basis.

We also introduce diagonal scaling matrices $D_\Delta^{(i)}$ that operate on the Lagrange multiplier spaces. Each element on the main diagonal corresponds to a Lagrange multiplier which enforces continuity between the nodal values of some $w_i \in \widetilde{W}_i$ and $w_j \in \widetilde{W}_j$ at some point $x \in \Gamma_h$. This diagonal element is defined as $\rho_j^\gamma(x)\mu_j^\dagger(x)$. Finally, we define a scaled jump operator by

$$B_{D,\Delta} := [D_\Delta^{(1)} B_\Delta^{(1)}, \ldots, D_\Delta^{(N)} B_\Delta^{(N)}].$$

As in Klawonn and Widlund [KW01, section 5], we solve the dual system (7) using the preconditioned conjugate gradient algorithm with the preconditioner

$$M^{-1} := B_{D,\Delta} S_\Delta B_{D,\Delta}^t. \tag{8}$$

The dual–primal FETI method is now the standard preconditioned conjugate gradient algorithm for solving the preconditioned system

$$M^{-1} F\lambda = M^{-1} d.$$

This definition of M clearly depends on the choice of the subspaces $\widehat{W}_\Pi$ and $\widetilde{W}_\Delta$ for the different algorithms.

A proof of the following theorem can be found in Klawonn, Widlund, and Dryja [KWD01].

Theorem 1 *The condition numbers of the preconditioned FETI–DP Algorithms B and C satisfy*

$$\kappa(M^{-1}F) \leq C\,(1+\log(H/h))^2$$

and the condition number of Algorithm D satisfies

$$\kappa(M^{-1}F) \leq C\,\max(1,TOL)\,(1+\log(H/h))^2.$$

Here, C is independent of h, H, γ, and the values of the ρ_i.

Remark 1 *A weaker condition number estimate, with an additional factor H/h, can be given for Algorithm A; see [KWD01].*

References

[DSW94]Maksymilian Dryja, Barry F. Smith, and Olof B. Widlund. Schwarz analysis of iterative substructuring algorithms for elliptic problems in three dimensions. *SIAM J. Numer. Anal.*, 31(6):1662–1694, December 1994.

[DSW96]Maksymilian Dryja, Marcus V. Sarkis, and Olof B. Widlund. Multilevel Schwarz methods for elliptic problems with discontinuous coefficients in three dimensions. *Numer. Math.*, 72(3):313–348, 1996.

[DW95]Maksymilian Dryja and Olof B. Widlund. Schwarz methods of Neumann-Neumann type for three-dimensional elliptic finite element problems. *Comm. Pure Appl. Math.*, 48(2):121–155, February 1995.

[FLLT+01]Charbel Farhat, Michel Lesoinne, Patrick Le Tallec, Kendall Pierson, and Daniel Rixen. FETI-DP: A dual-primal unified FETI method – part I: A faster alternative to the two-level FETI method. *Int. J. Numer. Meth. Engng.*, 50:1523–1544, 2001.

[FLP00]Charbel Farhat, Michel Lesoinne, and Kendall Pierson. A scalable dual-primal domain decomposition method. *Numer. Lin. Alg. Appl.*, 7:687–714, 2000.

[KW01]Axel Klawonn and Olof B. Widlund. FETI and Neumann–Neumann Iterative Substructuring Methods: Connections and New Results. *Comm. Pure Appl. Math.*, 54:57–90, January 2001.

[KWD01]Axel Klawonn, Olof Widlund, and Maksymilian Dryja. Dual-primal FETI methods for three-dimensional elliptic problems with heterogeneous coefficients. Technical Report 815, Courant Institute of Mathematical Sciences, Department of Computer Science, April 2001.

[MT01]Jan Mandel and Radek Tezaur. On the convergence of a dual-primal substructuring method. *Numer. Math.*, 88:543–558, 2001.

[Pie00]Kendall H. Pierson. *A family of domain decomposition methods for the massively parallel solution of computational mechanics problems*. PhD thesis, University of Colorado at Boulder, Aerospace Engineering, 2000.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

42 Comparison of domain decomposition methods for solving continuous casting problem

E. Laitinen[1], J. Pieskä[2], J. Saranen[3], A. Lapin[4]

Introduction

Two different kind of domain decomposition methods and algorithms to solve the continuous casting problem are presented and analyzed. The multiplicative Schwarz method with overlapping subdomains, and splitting iterative method with nonoverlapping subdomains are studied. Results considering convergence for both of these methods are presented and studied via numerical example. The finite element method with rectangular elements was used to discretize the problem. Advantages and disadvantages for both of these methods for this problem are discussed and analyzed.

The continuous casting problem can be stated mathematically as follows. Let $\Omega = \{0 < x_1 < L_{x_1}, 0 < x_2 < L_{x_2}\}$ be the rectangular domain with the boundary $\Gamma = \partial\Omega$ consisting of two parts: $\Gamma_1 = \{x \in \partial\Omega : x_2 = 0 \vee x_2 = L_{x_2}\}$, $\Gamma_2 = \{x \in \partial\Omega \setminus \Gamma_1\}$. We assume that the domain $\Omega \subset \mathbb{R}^2$ is occupied by thermodynamically homogeneous and isotropic steel. We denote by $H(x,t)$ the enthalpy related to unit mass and by $u(x,t)$ the temperature for $(x,t) \in \Omega\times]0,T[$. We have constitutive law

$$H = H(u) = \rho \int_0^u c(\Theta)d\Theta + \rho L(1 - f_s(u)) \text{ in } \Omega\times]0,T[,$$

where ρ is density, $c(u)$ is specific heat, L is latent heat and $f_s(u)$ is solid fraction.

Graph $H(u)$ is a increasing function $\mathbb{R} \to \mathbb{R}$ involving near vertical segments corresponding to the phase transition states, namely, for $u \in [T_L, T_S]$ where $0 < T_L < T_S$ are melting and solidification temperatures, correspondingly.

We study the following boundary-value problem: find $u = u(x,t)$ such that

$$\text{(P)} \begin{cases} \dfrac{\partial H(u)}{\partial t} + v\dfrac{\partial H(u)}{\partial x_2} - \Delta u = 0 \text{ for } x \in \Omega, t > 0, \\ u = z(x_1,t) > 0 \text{ for } x \in \Gamma_1, t > 0, \\ \partial u/\partial n + au + b|u|^3 u = g, a \geq 0, b \geq 0, g \geq 0 \text{ for } x \in \Gamma_2, t > 0, \\ u = u_0(x) > 0 \text{ for } x \in \bar{\Omega}, t = 0. \end{cases}$$

The existence and uniqueness of the weak solution for the problem (P) are proved in [RY90].

[1]Department of Mathematical Sciences, University of Oulu, P.O. Box 3000, Oulu 90401, FINLAND, erkki.laitinen@oulu.fi
[2]Department of Mathematical Sciences, University of Oulu, P.O. Box 3000, Oulu 90401, FINLAND, jpieska@cc.oulu.fi
[3]Department of Mathematical Sciences, University of Oulu, P.O. Box 3000, Oulu 90401, FINLAND, jsaranen@cc.oulu.fi
[4]Department of Computing Mathematics and Cybernetics, Kazan State University, Kazan 4200008, RUSSIA, alapin@ksu.ru

To approximate the problem (P) we rewrite it as the integral equality for fixed $t > 0$. Let $V = H^1(\Omega), V^0 = \{u \in V : u(x) = 0 \text{ for } x \in \Gamma_1\}$ and $V^z = \{u \in V : u(x) = z \text{ for } x \in \Gamma_1\}$. The solution of the problem (P) for fixed $t > 0$ satisfies the following equality for all $\eta \in V^0, u(t) \in V^{z(t)}$:

$$\int_\Omega (\partial H/\partial t + v(t)\partial H/\partial x_2)\eta dx + \int_\Omega \nabla u \nabla \eta dx + \int_{\Gamma_2} (au + b|u|^3 u)\eta d\Gamma = \int_{\Gamma_2} g\eta d\Gamma$$

Let T_h be the triangulation of Ω in rectangular elements of dimensions $h_1 \times h_2$ and $V_h = \{u_h(x) \in H^1(\Omega) : u_h(x) \in Q_1 \text{ for all } \delta \in T_h\}$, where Q_1 is the space of bilinear functions. By Π_h we denote the local Q_1-interpolant. We also use the following notations: $V_h^0 = \{u_h(x) \in V_h : u_h(x) = 0, \text{ for all } x \in \Gamma_1\}$, $V_h^z = \{u_h(x) \in V_h : u_h(x) = z_h, \text{ for all } x \in \Gamma_1\}$ for the subsets of V_h. Here z_h is the V_h - interpolant of z on the boundary Γ_1. For any continuous function $v(x)$ we put

$$S_\delta(v) = \int_{\delta_h} \Pi_h(v)dx; S_\Omega(v) = \sum_{\delta \in T_h} S_\delta(v),$$

$$S_{\partial\delta}(v) = \int_{\partial\delta_h} \Pi_h(v)dx; S_{\Gamma_2}(v) = \sum_{\partial\delta_h \in T_h \cap \bar{\Gamma}_2} S_{\partial\delta}(v).$$

Let also $\omega_\tau = \{t_k = k\tau, 0 \le k \le M, M\tau = T\}$ be the uniform mesh in time on the segment $[0, T]$. To approximate the term $\left(\frac{\partial}{\partial t} + v(t)\frac{\partial}{\partial x_2}\right) H$ we use characteristics of this first order differential operator [Che91, JR82]. We use the notation

$$d_{\bar{t}} H = \frac{1}{\tau}(H(x,t) - \tilde{H}(x, t-\tau))$$

for the difference quotient approximating the term $\left(\frac{\partial}{\partial t} + v(t)\frac{\partial}{\partial x_2}\right) H$ in each mesh point on time level t by using characteristic method.

Then the approximation scheme can be written as follows: for all $t \in \omega_\tau, t > 0,$ find $u_h \in V_h^z$ such that

$$S_\Omega(d_{\bar{t}} H_h \eta_h) + S_\Omega(\nabla u_h \nabla \eta_h) + S_{\Gamma_2}((au_h + b|u_h|^3|u_h|)\eta_h) = S_{\Gamma_2}(g\eta_h) \text{ for all } \eta_h \in V_h^0. \tag{1}$$

Let $N_0 = \operatorname{card} V_h^0$ and $u \in \mathbb{R}^{N_0}$ be the vector of nodal values for $u_h \in V_h^0$. Below we use the writing $u_h \Leftrightarrow u$ for this bijection. For the matrices $N_0 \times N_0$ we have the relations: for all $u_h \in V_h^0 \Leftrightarrow u \in \mathbb{R}^{N_0}, \eta_h \in V_h^0 \Leftrightarrow \eta \in \mathbb{R}^{N_0}$

$$(Au, \eta) = S_\Omega(\nabla u_h \nabla \eta_h) + S_{\Gamma_2}(au_h\eta_h); (Bu, \eta) = S_\Omega(1/\tau u_h \eta_h).$$

$$(Cu, \eta) = S_{\Gamma_2}(b|u_h|^3|u_h|\eta_h);$$

Similarly we define the vector f: $(f, \eta) = S_{\Gamma_2}(g\eta_h) + S_\Omega(1/\tau \tilde{H}_h \eta_h)$. Let now $\tilde{z}_h(x) \in V_h$ be the function which is equal to z_h in $\bar{\Gamma}_1$ and 0 for all nodes in ω, then f_0 is defined by the equality: $(f_0, \eta) = S_\Omega(\nabla \tilde{z}_h, \nabla \eta_h)$ for all $\eta_h \in V_h^0$. Finally we get $F = f + f_0$. In these notations the algebraic form for the mesh scheme (1) at fixed time level can be written as follows:

$$Au + BH(u) + Cu = F. \tag{2}$$

Here A, B are symmetric, positive definite M-matrices (moreover B is diagonal one) and $H(u)$ is vector with components $(H(u))_i = H(u_i)$. The operator C has the diagonal form: $Cu = (c_1(u_1), c_2(u_2), ..., c_N(u_N))^T$, where c_i are continuous non-decreasing functions.

Schwarz alternating methods

We study the convergence of multiplicative Schwarz alternating method (MSAM) and additive Schwarz alternating method (ASAM) for (2).

For the simplicity but without loss of generality we suppose that the domain Ω is decomposed into two overlapping subdomains Ω_1 and Ω_2, consisiting of the elements of triangulation T_h; any internal node of the grid in Ω is the internal node of at least one of the subdomains. We arrange the internal nodes of the mesh as follows. First, we enumerate the internal nodes lying in Ω_1, then the nodes in $\overline{\Omega_1 \cap \Omega_2}$ and at last the nodes in Ω_2. The vector $u \in \mathbb{R}^N$ takes the form $u = (u_{11}, u_{12}, u_{22})^T$ with subvector u_{ii} corresponding to the values of the mesh function $V_h \ni u_h \Leftrightarrow u$ in the nodes $x \in \text{int } \Omega_i$ and subvector u_{12} corresponding to the values in $x \in \overline{\Omega_1 \cap \Omega_2}$.

This decomposition implies also the partitioning of the matricies and nonlinear operator C:

$$A = (A_{ij})_{ij=1}^3, B = (B_{ij})_{ij=1}^3, C = \text{diag}(C_1, C_2, C_3).$$

We need some more notations, namely:

$$A_0^1 = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, B_0^1 = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, A_1^1 = diag(0, A_{23}), B_1^1 = diag(0, B_{23});$$

$$A_0^2 = \begin{pmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{pmatrix}, B_0^2 = \begin{pmatrix} B_{22} & B_{23} \\ B_{32} & B_{33} \end{pmatrix}, A_1^2 = diag(A_{21}, 0), B_1^2 = diag(B_{21}, 0);$$

$$C^1 = diag(C_1, C_2), C^2 = diag(C_2, C_3),\ u_1 = (u_{11}, u_{12})^T, u_2 = (u_{12}, u_{22})^T$$

and similar for other vectors. (We note, that $A_{13}, A_{31}, B_{13}, B_{31}$ are zero matricies.)

Then MSAM can be written as follows:

$$\begin{cases} A_0^1 v_1^{k+1} + B_0^1 H(v_1^{k+1}) + C^1 v_1^{k+1} = f_1 - A_1^1 u_2^k - B_1^1 H(u_2^k) \\ v_{22}^{k+1} = u_{22}^k \\ \\ u_{11}^{k+1} = v_{11}^{k+1} \\ A_0^2 u_2^{k+1} + B_0^2 H(u_2^{k+1}) + C^2 u_2^{k+1} = f_2 - A_1^2 v_1^{k+1} - B_1^2 H(v_1^{k+1}) \end{cases} \tag{3}$$

and ASAM has the form:

$$\begin{cases} A_0^1 v_1^{k+1} + B_0^1 H(v_1^{k+1}) + C^1 v_1^{k+1} = f_1 - A_1^1 u_2^k - B_1^1 H(u_2^k) \\ A_0^2 w_2^{k+1} + B_0^2 H(w_2^{k+1}) + C^2 w_2^{k+1} = f_2 - A_1^2 u_1^k - B_1^2 H(u_1^k) \\ \\ u_{11}^{k+1} = v_{11}^{k+1}, u_{22}^{k+1} = w_{22}^{k+1}, u_{12}^{k+1} = \alpha v_{12}^{k+1} + (1-\alpha) w_{12}^{k+1} \end{cases} \tag{4}$$

Here $k = 0, 1, 2, ...$, initial guess $u^0 = (u_{11}^0, u_{12}^0, u_{22}^0)^T$ and $\alpha \in (0, 1)$.

Along with these methods we consider also the block variant of Jacoby method (BJM). Let $A^0 = \mathrm{diag}(A_{11}, A_{22}, A_{33})$ be the block diagonal submatrix of A, $A^1 = A - A^0$ and $B = B^0 - B^1$ with similar splitting. Then A^0, B^0 are M - matricies and $A^1 \gg 0, B^1 \gg 0$. Moreover the iterative method (BJM) can be written in the form:

$$A^0 u^{k+1} + B^0 H(u^{k+1}) + C u^{k+1} = f - A^1 u^k - B^1 H(u^k). \tag{5}$$

Theorem 1 *Let A, B are $M-$matrices, where A is weakly diagonally dominant in columns, B is strictly diagonally dominant and C has the diagonal form $Cu = (c_1(u_1), c_2(u_2), ..., c_N(u_N))^T$, where c_i are continuous non-decreasing functions. Let also there exist sub- and supersolutions for the problem (2). Then the iterative methods (3), (4) and (5) are correctly defined for any initial guess u^0 from ordered interval $< \underline{u}, \overline{u} >$. If the initial guess is supersolution then the sequences of iterations for all methods (3), (4) and (5) converge monotonically decreasing to the unique solution of the problem (2). Moreover, let the iterations of MSAM, ASAM and BJM be denoted by $u^k_{MSAM}, u^k_{ASAM}, u^k_{BJM}$. Then for any k the following inequalities hold:*

$$u^k_{MSAM} \ll u^k_{ASAM} \ll u^k_{BJM}.$$

If starting from subsolution, then the inequalities are vice versa and the iterative sequences converge monotonically increasing [LLP99].

Splitting iterative method

Let now Ω be divided into p nonoverlapping subdomains Ω_i with the interfaces $S_{ij} = \overline{\Omega}_i \cap \overline{\Omega}_j$. We suppose that all interfaces as well as $\overline{\partial_1 \Omega}$ consist of the sides of $\delta \in T_h$.

The restrictions of functions from V^0_h on subdomains Ω_i form the spaces $V^i_h, i = 1, 2, ..., p$. We also denote by $V_h = V^1_h \times V^2_h \times \cdots \times V^p_h$. It is easy to check that V^0_h is isomorphic to the subspace K_h of V_h: $K_h = \{u_h = (u^1_h, u^2_h, ..., u^p_h) \in V_h : u^i_h(x) = u^j_h(x) \text{ for } x \in S_{ij}, i, j = 1, 2, ..., p\}$.

Let us put in the correspondence to the function $u^i_h \in V^i_h$ and the vector $u^i \in \mathbb{R}^{N_i}$ of its nodal values for nodes from $\bar{\Omega}_i \setminus \overline{\partial_1 \Omega}$ and denote this bijection by $u^i \Leftrightarrow u^i_h$. To $u_h \in V_h$ corresponds the vector $u \in \mathbb{R}^N, N = N_1 + N_2 + \cdots + N_p$. The subspace K_h corresponds to subspace of $\mathbb{R}^N$ which we denote by K. We have the following relations for $N_i \times N_i$ matrices: for all $V^i_h \ni u_h \Leftrightarrow u \in \mathbb{R}^{N_i}, V^i_h \ni \eta_h \Leftrightarrow \eta \in \mathbb{R}^{N_i}$

$$(A_i u_i, \eta_i)_i = S_{\Omega_i}(\nabla u_h \nabla \eta_h) + S_{\Gamma_2 \cap \partial\Omega_i}(a u_h \eta_h); (B_i u_i, \eta_i) = S_{\Omega_i}(1/\tau u_h \eta_h) \text{ and}$$

$$(c_i u_i, \eta_i)_i = S_{\Gamma_2 \cap \partial\Omega_i}(b|u_h|^3 |u_h| \eta_h).$$

Similarly we define the vectors f_i, f_{0i}: $(f_i, \eta_i)_i = S_{\Gamma_2 \cap \partial\Omega_i}(g\eta_h) + S_{\Omega_i}(1/\tau \tilde{H}_h \eta_h)$ $(f_{0i}, \eta_i)_i = S_{\Omega_i}(\nabla \tilde{z}_h, \nabla \eta_h)$ for all $\eta_h \in V^i_h$. Finally we get $F_i = f_i + f_{0i}$.

Let further $A = diag(A_1, A_2, ..., A_p), B = diag(B_{01}, B_{02}, ..., B_{0p})$ and $F = (F_1, F_2, ..., F_p) \in \mathbb{R}^N$. Below we denote by $C(u) = BH(u) + cu + \partial I_K(u)$, where I_K is the indicator function of the subspace K. The operator A is bounded, hemicontinuous and uniformly monotone, C is maximal monotone operator. In these notations the algebraic form for the mesh scheme using DDM can be written (at fixed time level) as follows:

$$Au + Cu \ni F. \tag{6}$$

Due to the properties of A and C there exists unique solution u to the problem (6) [Bre73, Roc70].

We solve the inclusion (6) by splitting iterative method:

$$D_0^{-1}(u^{k+1/2} - u^k) + Au^k + Cu^{k+1/2} \ni F \qquad (7)$$
$$D_1(u^{k+1} - u^k) = u^{k+1/2} - u^k.$$

where D_0 and D_1 are some positive definite matrices. Due to the properties of D_0 and D_1 there exist the unique solutions $u^{k+1/2}$ and u^{k+1} for any k. For other examples of splitting methods see [Gab83, LS88, LM79].

For theoretical study of the convergence and rate of convergence for this splitting iterative method we can proof:

Theorem 2 *Let $V = V_1 \times V_2 \times ... \times V_p$, where V_i are Hilbert spaces with inner products $(.,.)_i$ and norms $||.||_i = (.,.)_i^{1/2}$ and let A be diagonal linear operator: $A = diag(A_1, A_2, ..., A_p)$ with $A_i : V_i \to V_i$ satisfying for all i the following assumptions: $m_i I_i \le A_i = A_i^* \le M_i I_i$ for all $i, m_i > 0$. Let also C be a maximal monotone operator and $z^k = u^k - u$, where u^k is the kth iteration and u is the exact solution.*

If $D_0 = diag(\lambda_1 I_1, \lambda_2 I_2, ..., \lambda_p I_p)$ and either $D_1 = I + D_0 A$ or $D_1 = 1/2(I + D_0 A)$ then the iterative method (7) converges for any $\lambda_i > 0$ and for the optimal choice of the iterative parameter $\lambda_i = 1/\sqrt{(m_i M_i)}$ the following estimate for rate of convergence is valid:

$$\|D_0^{-1/2}(I + D_0 A^{(n)})z^n\| \le q^n \|D_0^{-1/2}(I + D_0 A^{(0)})z^0\|, \qquad (8)$$

with $q = q_1 = \max\limits_{1 \le i \le p} \dfrac{\sqrt{M_i}}{\sqrt{M_i} + \sqrt{m_i}}$ for the first choice of D_1 (corresponds to Douglas-Rachford scheme) and with $q = q_2 = \max\limits_{1 \le i \le p} \dfrac{\sqrt{M_i} - \sqrt{m_i}}{\sqrt{M_i} + \sqrt{m_i}}$ for for the second choice of D_1 (corresponds to Peaceman-Rachford scheme).

The iterative method (7) with, for example, $D_1 = I + D_0 A$ for DDM mesh scheme (6) leads to algorithm

$$D_0^{-1}(u^{k+1/2} - u^k) + Au^k + Cu^{k+1/2} \ni f \qquad (9)$$
$$(I_i + \lambda_i A_i)(u^{i,k+1} - u^{i,k}) = u^{i,k+1/2} - u^{i,k}, i = 1, 2, ..., p, \qquad (10)$$

$u^k = (u^{1,k}, u^{2,k}, ..., u^{p,k})$.

Linear equations (10) may be solved independently for $i = 1, 2, ..., p$. As for (9) then for coordinates of $u^{k+1/2}$ corresponding to internal nodes $x \in \Omega_i$ operator C has diagonal form: $C = \partial\theta$. It means that the system of non-coupled scalar nonlinear equations corresponds to these points. For nodes lying on the interfaces S_{ij} system (9) contains subsystems of two (if it is the interior node of the interface) or several (if it is a cross-point of several interfaces) coupled equations. These subsystems can be also reformulated as problems to minimise convex differentiable functions of two or several variables. To solve these subproblems we can use one of standard optimization method.

The assumptions of Theorem 2 are satisfied with $m_i = O(1), M_i = O(\tau h^{-2})$. If we choose $\lambda_i = O(h/\tau^{1/2})$ in method (7) with either $D_1 = I + D_0 A$ or $D_1 = 1/2(I + D_0 A)$, $D_0 = diag(\lambda_1 I_1, \lambda_2 I_2, ..., \lambda_p I_p)$, then $q_1 = 1 - O(h/\tau^{1/2}), q_2 = 1 - O(h/\tau^{1/2})$ and the number of iterations to achieve accuracy ϵ is $n(\epsilon) = O(\tau^{1/2} h^{-1} \ln 1/\epsilon)$.

Numerical results

To validate the numerical schemes described in sections 42 and 42 the following numerical example was considered.

Let $\Omega =]0,1[\times]0,1[$ with the boundary Γ divided in two parts such that $\Gamma_D = \{x \in \partial\Omega : x_2 = 0 \vee x_2 = 1\}$ and $\Gamma_N = \Gamma \setminus \Gamma_D$, moreover let $T = 1$. Let us consider the case where the phase change temperature $u_{SL} = 1$ and the latent heat $L = 1$. Let the phase change interval be $[u_{SL} - \varepsilon, u_{SL} + \varepsilon]$, $\varepsilon = 0.01$, and the velocity is $v(t) = \frac{1}{5}$. Our numerical example is

$$\begin{array}{rcll} \frac{\partial H}{\partial t} - \Delta K + v(t)\frac{\partial H}{\partial x_2} & = & f(x;t) & \text{on } \Omega, \\ u(x_1,x_2;t) & = & (x_1 - \frac{1}{2})^2 - \frac{1}{2}e^{-4t} + \frac{5}{4} & \text{on } \Gamma_D, \\ \frac{\partial u}{\partial n} & = & 1 & \text{on } \Gamma_N, \\ u(x_1,x_2;0) & = & (x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2 + \frac{1}{2} & \text{on } \Omega, \end{array}$$

where

$$K(u) = \begin{cases} u & \text{if } u < u_{SL} - \varepsilon, \\ \frac{3}{2}u - \frac{1-\varepsilon}{2} & \text{if } u \in [u_{SL} - \varepsilon, u_{SL} + \varepsilon], \\ 2u - 1 & \text{if } u > u_{SL} + \varepsilon, \end{cases}$$

and

$$H(u) = \begin{cases} 2u & \text{if } u < u_{SL} - \varepsilon, \\ \left(\frac{1+8\varepsilon}{2\varepsilon}\right)(u-1) + \frac{5+4\varepsilon}{2} & \text{if } u \in [u_{SL} - \varepsilon, u_{SL} + \varepsilon], \\ 6u - 3 & \text{if } u > u_{SL} + \varepsilon. \end{cases}$$

Furthermore

$$f(x;t) = \begin{cases} 4e^{-4t} + \frac{1}{5}(4x_2 - 2) - 4 & \text{if } u < u_M, \\ 12e^{-4t} + \frac{1}{5}(12x_2 - 6) - 8 & \text{if } u > u_M. \end{cases}$$

The exact solution of our problem is $u(x_1,x_2;t) = (x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2 - \frac{1}{2}e^{-4t} + 1$.

We split the enthalpy function $H(u)$ as follows: $H(u) = \alpha u + H_0(u)$, where α is the minimal slope of the enthalpy function. In our numerical example $\alpha = 2$.

For splitting iterative method the optimal iterative parameter $\lambda_i = \frac{1}{\sqrt{m_i M_i}}$, where $m_i = \alpha + \tau\mu^i_{min}(A_{00})$ and $M_i = \alpha + \tau\mu^i_{max}(A_{00})$, where $\mu^i_{min}(A_{00})$ is the smallest eigenvalue of the matrix $(A_{00})_i$, which is the approximation of the Laplacian operator and correspondingly $\mu^i_{max}(A_{00})$ is the biggest eigenvalue.

The numerical test was done such away that everything for different methods would be optimal. Numerical test were run in the computer Cedar in CSC, Espoo Finland, (128 RISC processors); mainly 4 processors were used. The stopping criterion was the norm of residual $\|r\| \leq 10^{-4}$.

From the tables below **splitter** is splitting iterative method, **multi2** is multiplicative Schwarz with overlapping size $2h$ and **multi4** is multiplicative Schwarz with overlapping size $4h$. Moreover **proc** means the number of processors, **iter** the number of iterations and **S** is speedup.

Conclusions

Two different method was used to solve the problem (P). From Table 1 it can be seen that Splitting iterative method (SIM) is better (faster) than the Multiplicative Schwarz Alternating Method (MSAM) for the continuous casting problem. The speedups from the Table 1 show that (SIM) can be parallelized better than (MSAM).

	Splitter		**multi2**		**multi4**	
proc	Time [s]	**S**	Time [s]	**S**	Time [s]	**S**
1	466.4	–	259.4	–	259.4	–
2	166.8	2.8	212.6	1.22	177.5	1.46
4	124.6	3.74	174.9	1.48	157.4	1.65
6	106.7	4.37	140.3	1.85	131.9	1.97
8	85.4	5.46	119.3	2.17	109.6	2.37
10	70.9	6.58	95.4	2.72	92.7	2.80
12	59.3	7.87	85.6	3.03	85.2	3.04

Table 1: The comparison of calculation times and speedups when grid size is fixed to be 129×129 and 256 time steps. Number of processors are changed.

		Splitter		**multi2**		**multi4**	
grid	time steps	Time [s]	**iter**	Time [s]	**iter**	Time [s]	**iter**
17×17	32	0.45	24	0.68	6	0.49	4
33×33	65	1.75	25	1.44	7	1.31	4
65×65	128	12.3	26	14.2	8	12.6	5
129×129	256	124.6	29	174.9	9	157.4	6
161×161	320	188.2	29	391.8	9	350.1	6
257×257	512	1949.4	26	4425.2	9	3875.8	7

Table 2: The comparison of calculation times and number of iterations for different grid size and fixed number of processors; 4 processors.

From Table 2 it can be seen that when grid size increases the difference between calculation times for (MSAM) and (SIM) increases. Splitting iterative method is much more suitable for big continuous casting problems when we can use many processors and number of unknows are big, like in many real industrial application. For (SIM) we also know how to determine the optimal iterative parameter. The numerical experiments have shown that the theoretical optimal value for the iterative parameter is close to the practical optimal one.

References

[Bre73]Haim Brezis. *Operateurs Maximaux Monotones et Semigroups de Contractions dans les Espaces de Hilbert*. North-Holland Publishing Company, 1973.

[Che91]Zhiming Chen. Numerical solutions of a two-phase continuous casting problem. In P. Neittaanmaki, editor, *Numerical Methods for Free Boundary Problem*, pages 103–121, Basel, 1991. International Series of Numerical Mathematics, Birkhuser.

[Gab83]Daniel Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-value Problems*, Amsterdam, 1983. North-Holland Publishing Company.

[JR82]Jim Douglas Jr. and Thomas F. Russel. Numerical methods for convection-dominated

diffusion problem based on combining the method of characteristic with finite element or finite difference procedure. *Siam J. Numer Anal.*, 19:871–885, 1982.

[LLP99]Erkki Laitinen, Alexandr Lapin, and Jali Pieska. Mesh approximation and iterative solution of the continuous casting problem. In P. Neittaanmaki, T. Tiihinen, and P. Tarvainen, editors, *ENUMATH 99 - Proceedings of the 3rd European Conference on Numerical Mathematics and Advanced Applications*, pages 601–617. World Scientific, Singapore, 1999.

[LM79]Pierre L. Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. *Siam J. Numer. Anal.*, 16:964–979, 1979.

[LS88]A. Lapin and D. O. Solovyev. Splitting iterative methods for variational inequalities. Preprint 783, Center of Calcul., Novosibirsk, 1988.

[Roc70]Tyrrel R. Rockafellar. On the maximality of sums of nonlinear monotone operators. *Trans. Amer. Math. Soc.*, 149:75–88, 1970.

[RY90]José F. Rodrigues and Fahuai Yi. On a two-phase continuous casting stefan problem with nonlinear flux. *Euro. of Applied Mathematics*, 1:259–278, 1990.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

43 A Mesh Refinement Method for Optimization with DDM

Géraldine Lemarchand, Olivier Pironneau[1] and Elijah Polak[2]

Approximate Gradient

We apply here an idea developed in [PP02] whereby mesh refinement can be mixed with approximate gradients within an optimization loop. This is particularly useful for problems where the exact gradient is difficult to compute, which is the case of DDM problems[BW86]

Consider a generic optimization problem and its finite dimensional approximation

$$\min_{z\in Z} J(z) \qquad \min_{z\in Z_h} J_h(z). \tag{1}$$

The following is the method of Steepest descent with a Goldstein/Armijo rule for the step size:

Algorithm 1 *:*

while $\|\mathrm{grad}_z J_h(z^m)\| > \epsilon$ ***do***
{
$z^{m+1} = z^m - \rho \mathrm{grad}_z J_h(z^m)$ *where* ρ *is such that*

$$-\beta\rho\|w\|^2 < J_h(z^m - \rho w) - J_h(z^m) < -\alpha\rho\|w\|^2 \tag{2}$$

with $w = \mathrm{grad}_z J_h(z^m)$ *Set* $m := m + 1$*;*
}

Now consider the same algorithm with parameter refinement

Algorithm 2 *:*

while $h > h_{min}$ ***do***
{ ***while*** $\|\mathrm{grad}_z J_h(z^m)\| > \epsilon h^\gamma$ ***do***
{
$z^{m+1} = z^m - \rho \mathrm{grad}_z J_h(z^m)$ *where* ρ *such that,*

$$-\beta\rho\|w\|^2 < J_h(z^m - \rho w) - J_h(z^m) < -\alpha\rho\|w\|^2 \tag{3}$$

with $w = \mathrm{grad}_z J_h(z^m)$*. Set* $m := m + 1$*;*
}
$h := h/2$*;*
}

Convergence is straightforward to establish as it is either Steepest Descent or $\mathrm{grad} J_h \to 0$ by the fact that $h \to h/2$.

[1]Université Paris VI and IUF (pironneau@ann.jussieu.fr)
[2]EEC, University of California, Berkeley

Approximate Gradients

Another possible gain in speed arises from the observation that we may not need to compute the exact gradient $\mathrm{grad}_z J_h$!

Assume that N is an iteration parameter and that $J_{h,N}$ and $\mathrm{grad}_z J_{h,N}$ denote approximations of J_h and $\mathrm{grad}_z J_h$ in the sense that

$$\lim_{N\to\infty} J_{h,N}(z) = J_h(z) \qquad \lim_{N\to\infty} \mathrm{grad}_{z\,N} J_{h,N}(z) = \mathrm{grad}_z J_h(z). \tag{4}$$

Now consider the following algorithm with additional parameter K and $N(h)$ with $N(h) \to \infty$ when $h \to 0$:
The following is Steepest descent with Goldstein/Armijo rule, mesh refinement and approximate gradients:

Algorithm 3 *:*

while $h > h_{min}$
{
while $|\mathrm{grad}_{z\,N} J^m| > \epsilon h^\gamma$
{
try *to find a step size* ρ *with* $w = \mathrm{grad}_{z\,N} J(z^m)$

$$-\beta\rho\|w\|^2 < J(z^m - \rho w) - J(z^m) < -\alpha\rho\|w\|^2 \tag{5}$$

if *success* ***then***
$\{z^{m+1} = z^m - \rho \mathrm{grad}_{z\,N} J^m; m := m+1;\}$
else $N := N + K;$
}
$h := h/2; N := N(h);$
}

The convergence is established by observing that Goldstein's rule gives a bound on the step size:

$$-\beta\rho \mathrm{grad}_z J \cdot h < J(z+\rho h) - J(z) = \rho \mathrm{grad}_z J \cdot h + \frac{\rho^2}{2} J'' hh \tag{6}$$

$$\Rightarrow \quad \rho > 2(\beta - 1)\frac{\mathrm{grad}_z J \cdot h}{J''(\xi)hh} \tag{7}$$

so that

$$J^{m+1} - J^m < -2\frac{\alpha(1-\beta)}{\|J''\|}|\mathrm{grad}_z J|^2 \tag{8}$$

Thus at each grid level the number of gradient iterations is bounded by $O(h^{-2\gamma})$. Therefore the algorithm does not jam and as before the norm of the gradient decreases with h.

Applications

Distributed control and DDM

Let $S \subset \Gamma = \partial\Omega$

$$\min_{v \in L^2(S)} J(v) = \int_\Omega [(u-u_d)^2 + |\nabla(u-u_d)|^2] \tag{9}$$

subject to

$$u - \Delta u = 0 \text{ in } \Omega, \frac{\partial u}{\partial n}|_S = \xi v \quad u_{\Gamma - S} = u_d\} \tag{10}$$

Then the optimality conditions are

$$\delta J = \int_S \xi(u-u_d)\delta v \tag{11}$$

Let $\Omega = \Omega_1 \cup \Omega_2$, let $\Gamma = \partial\Omega$ and $\Gamma_{ij} = \partial\Omega_i \cap \Omega_j$. The multiplicative Schwarz algorithm for the Laplace equation starts from a guess u_1^0, u_2^0 and computes the solution of

$$u - \Delta u = f \text{ in } \Omega, \quad u|_\Gamma = u_\Gamma \tag{12}$$

as the limit in n of $u_i^n, i = 1, 2$ defined by

$$\begin{aligned}
&u_1^{n+1} - \Delta u_1^{n+1} = f \text{ in } \Omega_1, \\
&u^{n+1}|_{\Gamma \cap \overline{\Omega}_1 - S} = u_\Gamma \quad u_1^{n+1}|_{\Gamma_{12}} = u_2^n \quad \frac{\partial u_1^{n+1}}{\partial n}|_S = \xi v \\
&u_2^{n+1} - \Delta u_2^{n+1} = f \text{ in } \Omega_2, \\
&u^{n+1}|_{\Gamma \cap \overline{\Omega}_2 - S} = u_\Gamma \quad u^{n+1}|_{\Gamma_{21}} = u_1^n \quad \frac{\partial u_2^{n+1}}{\partial n}|_S = \xi v
\end{aligned}$$

The discretized problem is

$$\begin{aligned}
&\min_{v \in V_h} J_h^N(v) = \|u^N - u_d\|_\Omega^2 : \quad u_j^0 = 0, \quad n = 1..N \quad \forall w \in V_h \\
&u_j^n|_{\partial\Omega_{ij}} = u_j^{n-1}, \quad \int_{\Omega_j} [u_j^n w + \nabla u_j^n \nabla w] = \int_S \xi v w
\end{aligned}$$

where N is the number of Schwarz iterations. The exact discrete optimality conditions are difficult to implement because we may need to store all intermediate functions generated by the Schwarz algorithm (at least for the nonlinear cases) and integrate the system for the adjoint vectors in the reverse order. So here we will try to use the approximate gradient

$$\theta_{h,N} = \|u_{h,N} - u_d\|_S \tag{13}$$

where u_h is computed by N iterations of the Schwarz algorithm.

while $h > hmin$ {
while $\theta_{h,N} > \epsilon(h)$ {
if $(J_{h,N}^{m+1} - J_{h,N}^m < -\alpha\rho^m\theta_{h,N}^2)$
{ do a gradient iteration of step size ρ^m and m:=m+1 }
else N:=N+K
}
h:= h/2 }

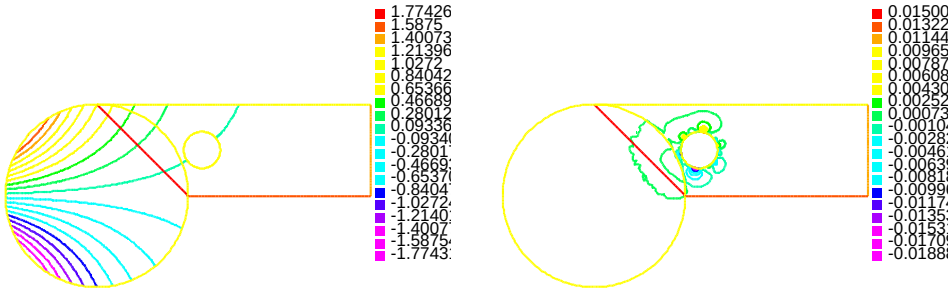


Figure 1: *The computed solution u (left) and the error $u - u_d$ (right).*

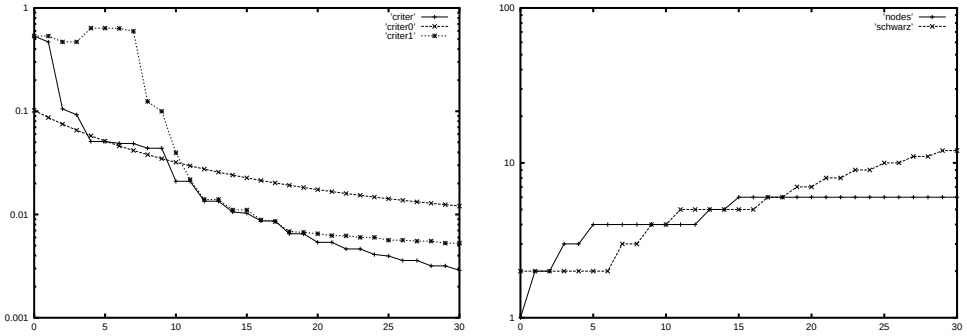


Figure 2: *After 30 iterations the gradient is 10^{-6} times its initial value, while without mesh refinment it has been divided by 100 only (embedded grid effect). On the* **left**, *is shown the cost function versus iteration number with and without mesh adaptation for Problem P_1. The smooth curve $(- + -)$ corresponds to standard steepest descent on the finest mesh with 500 Gauss-Seidel iterations for the linear systems. The broken curve $(- \times -)$ shows cost function decrease with Algorithm 1. Although the two curves are similar, there is an order of magnitude decrease in computing time using Algorithm 1. On the* **right** *is shown the history of the parameters in the algorithm, N and h.*

Numerical results

$u_d = e^{-x\sqrt{2}} sin(y)$. $\xi = \sin(30 * (x - 1.15)) + \sin(30 * (y - 0.5))$. The number of Schwarz iterations is initialized at 1. Results are shown in Figures 1 and 2.

Control in the coefficients

An absorbant coating of thickness α on an airfoil S is optimized to cancel the reflected acoustic wave in a sector σ. The Leontowitch conditions models the thin coating:

$$\min_{\alpha} \int_{\Sigma} |u|^2 \quad \text{subject to}$$

$$\omega^2 u + \Delta u = 0 \qquad \frac{\partial u}{\partial n} - i\omega u = 0 \text{ on } \Gamma_\infty, \quad \frac{\partial u}{\partial n} + \alpha\omega u = 0 \text{ on } S$$

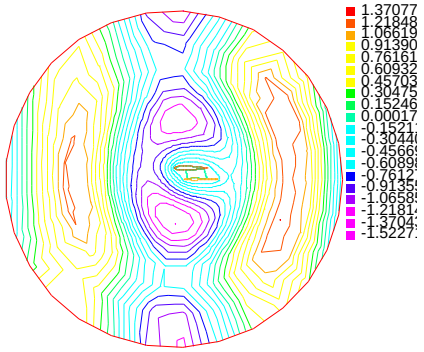


Figure 3: *Real part of the solution of Helmholtz equation*

The problem is discretized by the finite element method of degree 1 [Cia78] on triangles. The linear systems are solved with a Gauss factorization. The same gradient method with inexact gradients is applied (i.e. the gradient of the continuous problem discretized) with domain decomposition where one domain surrounds one of the airfoil. Figures 3 and 4 show the solution and Figure 5 shows the history of the convergence compared with a straight steepest descent method and a steepest descent with mesh refinement only and no DDM. The FEM software [BHOP99] has been used.

Optimal Shape Design

A transonic flow is computed by solving the Euler system of partial differential equation with NSC2KE[MP01] and the profile is optimized so as to minimize the pressure drag. The state equation is non-linear and the acceleration by approximate gradient is on the number of Newton iterations in the flow solver. There is no DDM here. The results are shown in Figures 6 and 7.

References

[BHOP99]Dominique Bernardi, Frederic Hecht, Kohji Otsuka, and Olivier Pironneau. freefem+, a finite element software to handle several meshes. *http://www.freefem.org*, 1999.

[BW86]Petter E. Bjørstad and Olof B. Widlund. Iterative methods for the solution of elliptic problems on regions partitioned into substructures. *SIAM J. Numer. Anal.*, 23(6):1093–1120, 1986.

[Cia78]Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[MP01]Bijan Mohammadi and Olivier Pironneau. *Applied OPtimal Shape Design*. Oxford University Press, Oxford, 2001.

[PP02]Olivier Pironneau and Elijah Polak. Consistent approximations and approximate functions and gradients in optimal control. *SIAM J. Numer. Anal.*, page to appear, 2002.

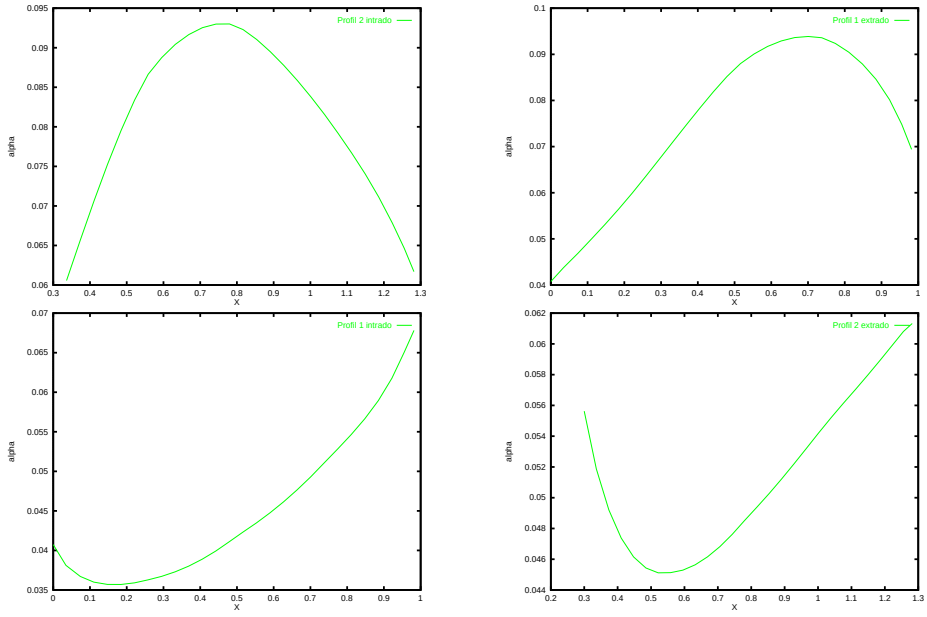


Figure 4: α *versus distance to the leading edge on the two sides of each airfoil.*

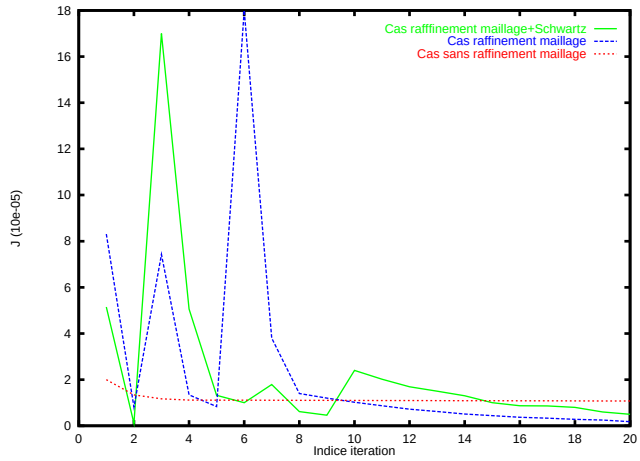


Figure 5: *History of the convergence of the cost function for the coating problem. The method with mesh refinement and adapted Schwarz iteration number (green curve) is compared with a straight steepest descent method (red curve) and a steepest descent with mesh refinement only and no DDM (blue curve).*

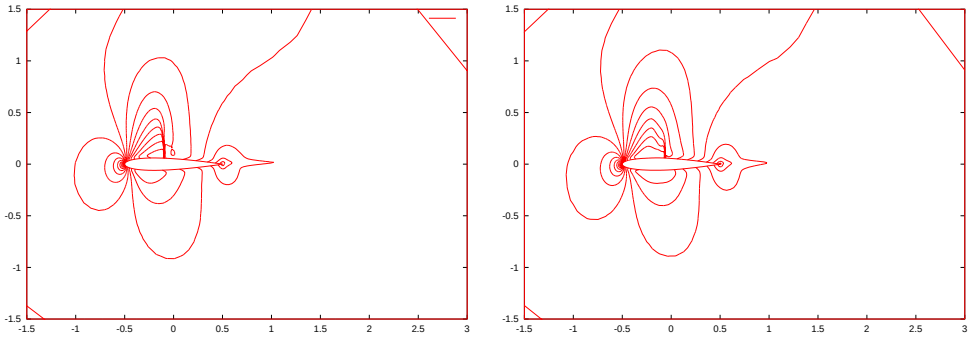


Figure 6: *Mach lines for the flow around the airfoil before shape optimization (left) and after. Notice that the shock tends to disappear, an expected result since the drag is a pressure drag.*

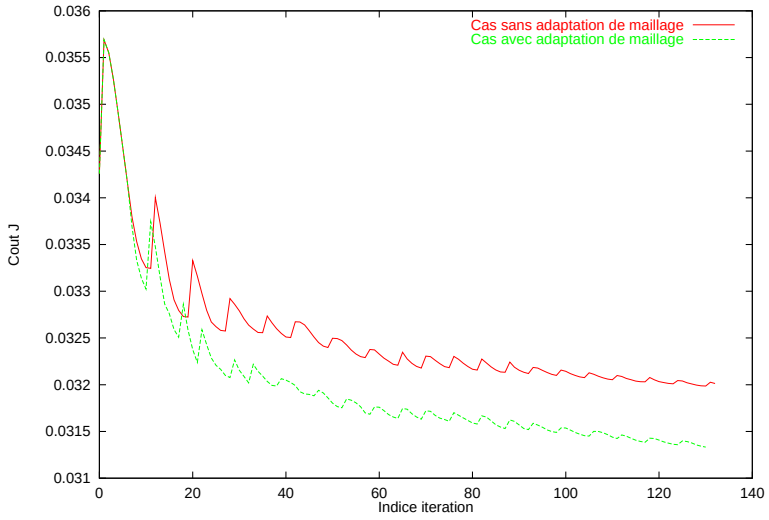


Figure 7: *History of the decrease of the cost function with and without mesh refinement and approximate gradient based on non converged flow solvers. The curve in red (top curve) is without mesh refinement but with control over the iteration number for the flow solver and the green curve is the same with mesh refinement.*

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

44 A Preconditioner for Linear Elasticity Problems

J. Martikainen[1], R.A.E. Mäkinen[2], T. Rossi[3], J. Toivanen[4]

Introduction

We consider the linear elasticity problem for homogeneous and isotropic material with mixed boundary conditions. The traditional formulation of the problem reads [NH80]

$$\begin{cases} -2\mu\nabla\cdot\varepsilon(\vec{u}) - \lambda\nabla(\nabla\cdot\vec{u}) = \vec{f} & \text{in } \Omega \subset \mathbb{R}^d, \\ \vec{u} = 0 & \text{on } \Gamma_0 \subset \partial\Omega, \\ [2\mu\varepsilon(\vec{u}) + \lambda(\nabla\cdot\vec{u})I]\cdot\vec{n} = 0 & \text{on } \Gamma_1 = \partial\Omega \setminus \Gamma_0, \end{cases} \tag{1}$$

where $\vec{n}$ is the outward unit normal vector, $\varepsilon(\vec{u})$ is the strain tensor and the Lamé coefficients μ and λ are defined by the Young modulus E and the Poisson ratio ν as follows

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} \quad \text{and} \quad \mu = \frac{E}{2(1+\nu)}.$$

Hereafter, it is assumed that the Poisson ratio ν satisfies $0 \le \nu \le 1/2$, although in theoretical considerations it is assumed that $\nu \le \hat{\nu} < 1/2$, where $\hat{\nu}$ is a constant. The measures of the boundaries Γ_0 and Γ_1 are assumed to be positive. The drawback of formulation (1) is that for (nearly) incompressible materials the parameter λ approaches infinity and the problem becomes ill conditioned. One remedy for this problem is to alter the formulation [BF91], [NH80]. We define a scalar function $p = -\lambda(\nabla\cdot\vec{u})$. This definition is added to the problem as a second equation. Then, we divide the equation by λ and get the following set of equations (see, for example, [Kob94] and references therein)

$$\begin{cases} -2\mu\nabla\cdot\varepsilon(\vec{u}) + \nabla p = \vec{f} & \text{in } \Omega, \\ -\nabla\cdot\vec{u} - \lambda^{-1}p = 0 & \text{in } \Omega, \\ \vec{u} = 0 & \text{on } \Gamma_0, \\ [2\mu\varepsilon(\vec{u}) - pI]\cdot\vec{n} = 0 & \text{on } \Gamma_1. \end{cases} \tag{2}$$

For the Poisson ratio $\nu = 1/2$ the latter equation of (2) is $-\nabla\cdot\vec{u} = 0$, which is exactly the incompressibility constraint. There are other possibilities to treat the elasticity problem for almost incompressible material such as hp-methods [SS96], nonconforming methods [Fal91] and reduced integration rules [ZT89].

Our purpose is to develop an efficient method for the numerical solution of discretized counterpart of the partial differential system (2). Our tools for this are a block diagonal preconditioner, a fictitious domain method and distributed Lagrange multipliers. The idea of the fictitious domain method is to extend the problem with complicated geometry to a larger, simple

[1]University of Jyväskylä, jamartik@mit.jyu.fi
[2]University of Jyväskylä, rainom@mit.jyu.fi
[3]University of Jyväskylä, tro@mit.jyu.fi
[4]University of Jyväskylä, tene@mit.jyu.fi

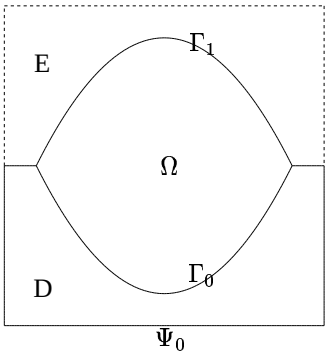


Figure 1: Rectangular domain Π with subdomains.

domain where an efficient solver can be used. This procedure can be justified with extension theorem for finite element functions [Wid87] on which the spectral optimality of the fictitious domain preconditioning is based [Ast78]. The incorporation of distributed Lagrange multipliers in fictitious domain method has been proposed in [GK98],[GPH$^+$99], for example. The advantage of the distributed Lagrange multipliers compared to the boundary Lagrange multipliers is the ease of preconditioning for both two-dimensional and three-dimensional problems.

Weak formulation of the elasticity problem

For the finite element discretization of the problem (2) we present a corresponding weak formulation. We define spaces

$$V = \{\vec{v} \in \left[H^1(\Omega)\right]^d \ : \ \vec{v}\,|_{\Gamma_0} = \vec{0}\} \quad \text{and} \quad Q = L^2(\Omega).$$

Then, the problem is to find $\vec{u} \in V$ and $p \in Q$, such that

$$\begin{cases} \displaystyle\int_\Omega 2\mu\varepsilon(\vec{u})\varepsilon(\vec{v}) - (\nabla\cdot\vec{v})p\,dx = \int_\Omega \vec{f}\cdot\vec{v}\,dx & \forall \vec{v} \in V \\ \displaystyle\int_\Omega -(\nabla\cdot\vec{u})q - \lambda^{-1}pq\,dx = 0 & \forall q \in Q. \end{cases} \tag{3}$$

In practise, we use a formulation which is equivalent to (3), but allows the application of the fictitious domain method. Therefore, we assume that there is a simple domain $\Pi \subset \mathbb{R}^d$, such that $\Omega \subset \Pi$ and a domain $D \subset \Pi$, such that $\Gamma_0 \subset \partial D$, $\Gamma_1 \cap \partial D = \emptyset$ and $D \cap \Omega = \emptyset$ as in Figure 1. We assume also that $\partial\Pi = \Psi_0 \cup \Psi_1$ and the measures of the boundary Ψ_0 and the domains Ω, D and $\Pi \setminus \overline{(\Omega \cup D)}$ are all positive. Then, we define spaces U and Ξ as follows

$$U = \{\vec{v} \in \left[H^1(\Omega \cup D)\right]^d \ : \ \vec{v}\,|_{\Psi_0} = \vec{0}\} \quad \text{and} \quad \Xi = \left[H^1(D)\right]^d.$$

A problem equivalent to (3) is to find $\vec{u} \in U$, $p \in Q$ and $\xi \in \Xi$, such that

$$\begin{cases} 2\mu \int_{\Omega \cup D} \varepsilon(\vec{u})\varepsilon(\vec{v})\, dx - \int_{\Omega} (\nabla \cdot \vec{v}) p\, dx + \langle \vec{\xi}, \vec{v} \rangle_{H^1(D)} = \int_{\Omega} \vec{f} \cdot \vec{v}\, dx & \forall \vec{v} \in U \\ \int_{\Omega} -(\nabla \cdot \vec{u}) q - \lambda^{-1} pq\, dx = 0 & \forall q \in Q \\ \langle \vec{u}, \vec{\eta} \rangle_{H^1(D)} = 0 & \forall \eta \in \Xi, \end{cases} \tag{4}$$

where $\langle \cdot, \cdot \rangle_{H^1(D)}$ is the $[H^1(D)]^d$ inner product. We use finite element method with quadratic triangular elements for the displacement components and distributed Lagrange multipliers and linear triangular elements for the pressure, also known as the Taylor-Hood element combination [BP79]. In this way, the elements have a one-to-one correspondence and the implementation of the discretization process is straightforward. If the element spaces are selected this way the distributed Lagrange multipliers tie the displacement components node by node and the system is equivalent to the system arising from (3) where the Dirichlet boundary conditions are treated by elimination. For this reason, the the triangulation must be compatible with the boundary Γ_0. The finite element mesh is assumed to be regular.
The form of the discretized linear system is

$$\begin{pmatrix} \mu\mathbf{A} & \mathbf{B}^T & \hat{\mathbf{C}}^T \\ \mathbf{B} & -\lambda^{-1}\mathbf{M} & \mathbf{0} \\ \hat{\mathbf{C}} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \\ \hat{\boldsymbol{\xi}} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \tag{5}$$

where $\hat{\mathbf{C}}$ has the form $\hat{\mathbf{C}} = (\hat{\mathbf{E}}\ \mathbf{0})$. By applying the change of variables $\boldsymbol{\xi} = \hat{\mathbf{E}}\hat{\boldsymbol{\xi}}$ and multiplying the last block row of (5) by $\hat{\mathbf{E}}^{-1}$, we get our final linear system

$$\begin{pmatrix} \mu\mathbf{A} & \mathbf{B}^T & \mathbf{C}^T \\ \mathbf{B} & -\lambda^{-1}\mathbf{M} & \mathbf{0} \\ \mathbf{C} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \\ \boldsymbol{\xi} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \tag{6}$$

where $\mathbf{C} = (\mathbf{I}\ \mathbf{0}) = \hat{\mathbf{E}}^{-1}\hat{\mathbf{C}}$. Here, the matrix $\mathbf{C}$ is defined by $\mathbf{C}\mathbf{v} = \mathbf{v}_D$ for all $\mathbf{v} = (\mathbf{v}_\Omega, \mathbf{v}_D)$. The system matrix of (6), which we denote by $\mathcal{A}$, is symmetric but indefinite.

The construction of the preconditioner

We would like to solve the linear system (6) using the preconditioned MINRES-method. We will show that a good preconditioner $\mathcal{P}$ for the discrete problem, given by its inverse is

$$\mathcal{P}^{-1} = \begin{pmatrix} \mu^{-1}\mathbf{R}\mathbf{A}_\Pi^{-1}\mathbf{R}^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \alpha^{-1}\mathbf{M}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu\mathbf{A}_D \end{pmatrix}. \tag{7}$$

The constant α is defined by $\alpha = \lambda^{-1} + \mu^{-1}$ and the matrix $\mathbf{R}$ by $\mathbf{R}\mathbf{v} = \mathbf{v}_{\Omega \cup D}$, where $\mathbf{v} = (\mathbf{v}_{\Omega \cup D}, \mathbf{v}_E)$. The matrices $\mathbf{A}_\Pi$ and $\mathbf{A}_D$ correspond to the elliptic part $2\int \varepsilon(\vec{u})\varepsilon(\vec{v})\, dx$ discretized in the domains Π and D with the element space for the displacement components, respectively and $\mathbf{M}$ is the mass matrix discretized in the domain Ω with the element space for the pressure.

We define the spectral equivalency of the matrices $\mathcal{A}$ and $\mathcal{P}$ as in [Kuz00]: Let λ_i, $i = 1, \ldots, m$ be the eigenvalues of the matrix $\mathcal{P}^{-1}\mathcal{A}$. If there exists positive constants c_1 and c_2 such that $c_1 \leq |\lambda_i| \leq c_2$ for all $i = 1, \ldots, m$ then the matrices $\mathcal{A}$ and $\mathcal{P}$ are spectrally equivalent with the constants c_1 and c_2. The condition number of the matrix $\mathcal{P}^{-1}\mathcal{A}$ is then bounded by $\kappa(\mathcal{P}^{-1}\mathcal{A}) \leq c_2/c_1$.
Since the convergence of the MINRES depends on the condition number of the preconditioned system we can guarantee the convergence rate by showing that the proposed preconditioner is spectrally equivalent to the system matrix.

Theorem 1 *Let us assume, that the Poisson ratio ν satisfies $0 < \nu \leq \hat{\nu} < 1/2$, where $\hat{\nu}$ is a constant. Then, the preconditioner $\mathcal{P}$ is spectrally equivalent to the system matrix $\mathcal{A}$ with constants c_1 and c_2 which are independent of the mesh step size h and the Poisson ratio ν.*

We begin to proof this result by showing that the matrix $\mathcal{A}$ is spectrally equivalent to a block diagonal matrix. Results for generalized eigenvalue problems, resembling Lemma 1 have been considered in articles [Kla95], [Kuz00] and [SW94]. We denote

$$\mathbf{G} = \begin{pmatrix} \mathbf{B} \\ \mathbf{C} \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} \lambda^{-1}\mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Notice, that the requirement $|\Gamma_0|, |\Gamma_1|, |\Psi_0| \geq \rho > 0$ is essential for the blocks $\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ and $\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T$ to be positive definite.

Lemma 1 *Let us assume that matrices $\mathbf{A}$ and $\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T$ are positive definite and matrix $\mathbf{H}$ is positive semidefinite. Then the eigenvalues θ of the generalized eigenvalue problem*

$$\begin{pmatrix} \mathbf{A} & \mathbf{G}^T \\ \mathbf{G} & -\mathbf{H} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \theta \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} + \mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \tag{8}$$

belong to the intervals $\left[-1, \frac{1-\sqrt{5}}{2}\right]$ and $\left[1, \frac{1+\sqrt{5}}{2}\right]$.

Proof The solution of the eigenvalue problem (8) satisfies the equations:

$$\mathbf{A}\mathbf{u} + \mathbf{G}^T\mathbf{v} = \theta\mathbf{A}\mathbf{u} \tag{9}$$

and

$$\mathbf{G}\mathbf{u} - \mathbf{H}\mathbf{v} = \theta\left(\mathbf{H} + \mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T\right)\mathbf{v}. \tag{10}$$

We assume that $\theta \neq 1$ and $\mathbf{v} \neq \mathbf{0}$. The vector $\mathbf{u}$ can be solved from (9) with respect to $\mathbf{v}$

$$\mathbf{u} = \mathbf{A}^{-1}\mathbf{G}^T\mathbf{v}/(\theta - 1).$$

Inserting this in the equation (10) gives us

$$\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T\mathbf{v}/(\theta - 1) - \mathbf{H}\mathbf{v} = \theta\mathbf{H}\mathbf{v} + \theta\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T\mathbf{v}. \tag{11}$$

We multiply the equation (11) from left by $\mathbf{v}^T$, collect the terms $\mathbf{v}^T\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T\mathbf{v}$ and divide the equation by it. Then, we have

$$\frac{1}{\theta - 1} - \theta - (\theta + 1)\frac{\mathbf{v}^T\mathbf{H}\mathbf{v}}{\mathbf{v}^T\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T\mathbf{v}} = 0.$$

We denote $\alpha(\mathbf{v}) = \frac{\mathbf{v}^T\mathbf{H}\mathbf{v}}{\mathbf{v}^T\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T\mathbf{v}}$. From the assumptions for the matrices it follows that $0 \leq \alpha(\mathbf{v}) < \infty$. Now, the eigenvalue θ can be solved from the equation $-(1+\alpha(\mathbf{v}))\theta^2 + \theta + (1+\alpha(\mathbf{v})) = 0$, and this gives the intervals $\left[-1, \frac{1-\sqrt{5}}{2}\right]$ and $\left]1, \frac{1+\sqrt{5}}{2}\right]$. By including the value $\theta = 1$, which was excluded during the calculations, the final intervals are obtained. ■

We continue the proof of the Theorem 1 by showing that the Schur complement matrix $\mathbf{S}_1 = \mathbf{H} + \mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T$ is again spectrally equivalent to a block diagonal matrix. First, we need to assume the following:

From here on, we assume that there exists positive constants c_1 and c_2 such that the inequality

$$c_1\mathbf{p}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T\mathbf{p} \leq \mathbf{p}^T\mathbf{M}\mathbf{p} \leq c_2\mathbf{p}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T\mathbf{p}$$

holds for any $\mathbf{p}$, where the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{M}$ are discretized with the formulation (4). Note that the assumption above holds if the element combination satisfies the LBB-condition with the given mixed boundary conditions [SEKW01].

Lemma 2 *For the Schur complements*

$$\mathbf{S}_1 = \lambda^{-1}\begin{pmatrix}\mathbf{M} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} + \mu^{-1}\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T \quad \textit{and}$$
$$\mathbf{S}_2 = \lambda^{-1}\begin{pmatrix}\mathbf{M} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} + \mu^{-1}\begin{pmatrix}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T & \mathbf{0}\\ \mathbf{0} & \mathbf{C}\mathbf{A}^{-1}\mathbf{C}^T\end{pmatrix}$$

there exists positive constants β and γ such that the inequality

$$\gamma\mathbf{s}^T\mathbf{S}_2\mathbf{s} \leq \mathbf{s}^T\mathbf{S}_1\mathbf{s} \leq \beta\mathbf{s}^T\mathbf{S}_2\mathbf{s}$$

holds for any $\mathbf{s} = (\mathbf{p}, \boldsymbol{\xi})$, when $0 < \nu \leq \hat{\nu} < 1/2$.

Proof We study the quadratic forms related to the matrices $\mathbf{S}_1$ and $\mathbf{S}_2$. We denote $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$, $\tilde{\mathbf{C}} = \mathbf{C}\mathbf{A}^{-1}\mathbf{C}^T$ and $\tilde{\mathbf{D}} = \mathbf{B}\mathbf{A}^{-1}\mathbf{C}^T$. It must be shown that the inequalities

$$\begin{aligned}&\beta\left(\mathbf{p}^T(\lambda^{-1}\mathbf{M} + \mu^{-1}\tilde{\mathbf{B}})\mathbf{p} + \mu^{-1}\boldsymbol{\xi}\tilde{\mathbf{C}}\boldsymbol{\xi}\right) \geq \\ &\quad \mathbf{p}^T(\lambda^{-1}\mathbf{M} + \mu^{-1}\tilde{\mathbf{B}})\mathbf{p} + \mu^{-1}\boldsymbol{\xi}\tilde{\mathbf{C}}\boldsymbol{\xi} + 2\mu^{-1}\mathbf{p}^T\tilde{\mathbf{D}}\boldsymbol{\xi},\end{aligned} \tag{12}$$

$$\begin{aligned}&\gamma\left(\mathbf{p}^T(\lambda^{-1}\mathbf{M} + \mu^{-1}\tilde{\mathbf{B}})\mathbf{p} + \mu^{-1}\boldsymbol{\xi}\tilde{\mathbf{C}}\boldsymbol{\xi}\right) \leq \\ &\quad \mathbf{p}^T(\lambda^{-1}\mathbf{M} + \mu^{-1}\tilde{\mathbf{B}})\mathbf{p} + \mu^{-1}\boldsymbol{\xi}\tilde{\mathbf{C}}\boldsymbol{\xi} + 2\mu^{-1}\mathbf{p}^T\tilde{\mathbf{D}}\boldsymbol{\xi}\end{aligned} \tag{13}$$

are satisfied. Clearly, (12) holds for any $\beta \geq 2$, since $\mathbf{A}$ is positive definite.

For every $\mathbf{p}$ and $\boldsymbol{\xi}$

$$\begin{aligned}0 &\leq \mu^{-1}\begin{pmatrix}(1+c_1\mu\lambda^{-1})^{1/4}\mathbf{p}\\ (1+c_1\mu\lambda^{-1})^{-1/4}\boldsymbol{\xi}\end{pmatrix}^T\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T\begin{pmatrix}(1+c_1\mu\lambda^{-1})^{1/4}\mathbf{p}\\ (1+c_1\mu\lambda^{-1})^{-1/4}\boldsymbol{\xi}\end{pmatrix}\\ &= \mu^{-1}(1+c_1\mu\lambda^{-1})^{-1/2}\left((1+c_1\mu\lambda^{-1})\mathbf{p}^T\tilde{\mathbf{B}}\mathbf{p} + \boldsymbol{\xi}^T\tilde{\mathbf{C}}\boldsymbol{\xi}\right) + 2\mu^{-1}\mathbf{p}^T\tilde{\mathbf{D}}\boldsymbol{\xi}\\ &\leq (1+c_1\mu\lambda^{-1})^{-1/2}\left(\lambda^{-1}\mathbf{p}^T\mathbf{M}\mathbf{p} + \mu^{-1}\mathbf{p}^T\tilde{\mathbf{B}}\mathbf{p} + \mu^{-1}\boldsymbol{\xi}^T\tilde{\mathbf{C}}\boldsymbol{\xi}\right) + 2\mu^{-1}\mathbf{p}^T\tilde{\mathbf{D}}\boldsymbol{\xi}\end{aligned}$$

holds. Therefore,

$$-(1+c_1\mu\lambda^{-1})^{-1/2}\left(\lambda^{-1}\mathbf{p}^T\mathbf{M}\mathbf{p}+\mu^{-1}\mathbf{p}^T\tilde{\mathbf{B}}\mathbf{p}+\mu^{-1}\boldsymbol{\xi}^T\tilde{\mathbf{C}}\boldsymbol{\xi}\right)\leq 2\mu^{-1}\mathbf{p}^T\tilde{\mathbf{D}}\boldsymbol{\xi}.$$

By adding the term $\lambda^{-1}\mathbf{p}^T\mathbf{M}\mathbf{p}+\mu^{-1}\mathbf{p}^T\tilde{\mathbf{B}}\mathbf{p}+\mu^{-1}\boldsymbol{\xi}^T\tilde{\mathbf{C}}\boldsymbol{\xi}$ to both sides of the inequality, we get the following lower bound

$$\gamma = 1-\frac{1}{\sqrt{1+\frac{\mu}{\lambda}c_1}} = 1-\frac{1}{\sqrt{1+\frac{1-2\nu}{2\nu}c_1}} \geq 1-\frac{1}{\sqrt{1+\frac{1-2\hat{\nu}}{2\hat{\nu}}c_1}} > 0.$$

■

Now, we have shown that the system matrix $\mathcal{A}$ is spectrally equivalent to the matrix

$$\begin{pmatrix} \mu\mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda^{-1}\mathbf{M}+\mu^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu^{-1}\mathbf{C}\mathbf{A}^{-1}\mathbf{C}^T \end{pmatrix}.$$

Since the domain is extended only over the boundary Γ_1 with natural boundary condition and due to the Korn inequality the matrix block $\mathbf{A}$ is spectrally equivalent to the discretized vector laplacian, the fictitious domain preconditioner $\mathbf{R}\mathbf{A}_\Pi^{-1}\mathbf{R}^T$ is optimal for $\mathbf{A}$ [Ast78]. Using the same principles, $\mathbf{A}_D$ is an optimal preconditioner for $\mathbf{C}\mathbf{A}^{-1}\mathbf{C}^T$ [GK98]. It follows from the assumption above that $\lambda^{-1}\mathbf{M}+\mu^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T$ and $(\lambda^{-1}+\mu^{-1})\mathbf{M}$ are spectrally equivalent. This concludes the proof of Theorem 1.

Numerical example

In this numerical example the operator $\mathbf{A}_\Pi^{-1}$ is approximated with the multigrid method using one symmetric multigrid V-cycle with one pre-smooth and one post-smooth with forward and backward Gauss-Seidel, respectively [Hac85]. This approximation is accurate and can be computed efficiently. The coarsest level problem is not solved exactly, but ten symmetric Gauss-Seidel sweeps are used instead. While this does not give the smallest possible number of outer iterations, it is very economical in the sense of the total computing time. The multigrid method is based fully on linear triangular elements. Systems with the mass matrix $\mathbf{M}$ are solved using the conjugate gradient method in machine precision with the lumped mass matrix as a preconditioner.

The example problem is a mixed boundary value problem in a domain Ω bounded by two parabole; see Figure 1. Homogeneous Dirichlet boundary condition is imposed on the lower half of the boundary and natural bondary condition is satisfied on the upper half of the boundary. The force term, which is pulling the structure left is distributed evenly over the computational domain. The displaced mesh can be seen in Figure 2. The numbers of iterations (it) and CPU times in seconds (time) with respect to the degrees of freedom (d.o.f.) and Poisson ratio (ν) are presented in Table 1. The numbers of iterations show that the convergence of the method is independent of both the mesh step size and the Poisson ratio and the method works well even for Poisson ratio 1/2. The numerical tests are performed on a PC with 400MHz Celeron processor running Linux operating system. The programs are compiled with the gcc compiler.

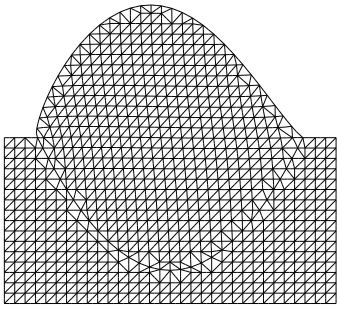


Figure 2: The finite element mesh with the displacements.

ν		0.1		0.2		0.3		0.4		0.5
d.o.f.	it	time	it	time	it	time	it	time	it	time
2253	108	2.7	105	2.6	102	2.6	99	2.5	99	2.5
8817	109	13.2	105	12.7	105	12.7	104	12.6	103	12.4
34889	108	57.0	105	55.6	103	54.6	102	54.2	103	54.8
138809	110	243.4	105	233.8	103	228.7	103	227.6	103	226.4

Table 1: The numbers of iterations and CPU times for the test problem.

Acknowledgments

The authors are grateful to Professor Y.A. Kuznetsov for fruitful discussions. This work was supported by the Academy of Finland, grants #43066 and #66407.

References

[Ast78]G. P. Astrakhatsev. Method of fictitious domains for a second-order elliptic equation with natural boundary conditions. *U.S.S.R. Computational Math. and Math. Phys.*, 18:114–121, 1978.

[BF91]F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, New-York, 1991.

[BP79]M. Bercovier and O. Pironneau. Error estimates for finite element method solution of the Stokes problem in the primitive variables. *Numer. Math.*, 33(2):211–224, 1979.

[Fal91]Richard S. Falk. Nonconforming finite element methods for the equations of linear elasticity. *Math. Comp.*, 57(196):529–550, 1991.

[GK98]Roland Glowinski and Yuri Kuznetsov. On the solution of the Dirichlet problem for linear elliptic operators by a distributed Lagrange multiplier method. *C. R. Acad. Sci. Paris Sér. I Math.*, 327(7):693–698, 1998.

[GPH+99]Roland Glowinski, Tsorng-Whay Pan, Todd I. Hesla, Daniel D. Joseph, and Jacques Priaux. A distributed Lagrange multiplier/fictitious domain method for flows around moving rigid bodies: Application to particulate flow. *Int. J. Num. Meth. Fluids*, 30(8):1043–1066, 1999.

[Hac85]Wolfgang Hackbusch. *Multigrid Methods and Applications*. Springer, Berlin, 1985.

[Kla95]Axel Klawonn. An optimal preconditioner for a class of saddle point problems with a penalty term, Part II: General theory. Technical Report 14/95, Westfälische Wilhelms-Universität Münster, Germany, April 1995. Also available as Technical Report 683 of the Courant Institute of Mathematical Sciences, New York University.

[Kob94]Georgij Kobelkov. On singular perturbation of the Stokes problem. In *Numerical analysis and mathematical modelling*, pages 79–83. Polish Acad. Sci., 1994.

[Kuz00]Yuri A. Kuznetsov. New iterative methods for singular perturbed positive definite matrices. *Russian J. Numer. Anal. Math. Modelling*, 15:65–71, 2000.

[NH80]Jindrich Necas and Ivan Hlavácek. *Mathematical theory of elastic and elasto-plastic bodies: an introduction*. Elsevier Scientific Publishing Co., 1980.

[SEKW01]David Silvester, Howard Elman, David Kay, and Andrew Wathen. Efficient preconditioning of the linearized Navier-Stokes equations for incompressible flows. *J. Comput. Appl. Math.*, (128):261–279, 2001.

[SS96]Rolf Stenberg and Manil Suri. Mixed hp finite element methods for problems in elasticity and Stokes flow. *Numer. Math.*, 72(3):367–389, 1996.

[SW94]David J. Silvester and Andrew J. Wathen. Fast iterative solution of stabilised Stokes systems part II: using general block preconditioners. *SIAM J. Numer. Anal.*, 31:1352–1367, 1994.

[Wid87]Olof B. Widlund. An extension theorem for finite element spaces with three applications. In Wolfgang Hackbusch and Kristian Witsch, editors, *Numerical Techniques in Continuum Mechanics*, pages 110–122, Braunschweig/Wiesbaden, 1987. Notes on Numerical Fluid Mechanics, v. 16, Friedr. Vieweg und Sohn. Proceedings of the Second GAMM-Seminar, Kiel, January, 1986.

[ZT89]O.C. Zienkiewicz and R. Taylor. *The finite element method*. McGraw-Hill, 1989.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

45 Comparison of two iterative substructuring methods for advection-diffusion problems

Gerd Rapin[1], Gert Lube[2]

Introduction

In this paper two different methods of domain decomposition for the advection-diffusion-reaction problem are considered. Both methods are analysed on the continuous level. The first approach is an additive nonoverlapping iteration-by-subdomains algorithm with Robin-type transmission conditions at the interface. This method has been well investigated in the last years (cf. [NR95], [LMO00], [Ott99]). We will give a short review of the results.
The second approach is a Schur complement method. The Schur complement is solved by a preconditioned Richardson iteration. As a preconditioner we use a generalised Neumann-Neumann preconditioner, the Robin-Robin preconditioner (cf. [ATNV00], [QV99], [BS00]). The application of the preconditioner requires the solution of a mixed problem with a Robin interface condition in each subdomain.
We apply a two-dimensional Fourier analysis to both methods in order to illustrate the different convergence behaviour of the methods. Finally we summarize the comparison of the two methods.

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with Lipschitz boundary $\partial\Omega$. We consider the following boundary value problem:

$$\left\{ \begin{array}{rcll} Lu := -\epsilon\triangle u + \mathbf{b}\cdot\nabla u + cu &=& f & \text{in } \Omega \\ u &=& 0 & \text{on } \partial\Omega \end{array} \right. \tag{1}$$

with the diffusion coefficient $\epsilon > 0$, a given flow $\mathbf{b} \in (W^{1,\infty}(\Omega))^2$, the source term $f \in L^2(\Omega)$ and the reaction coefficient $c \in L^\infty(\Omega)$. The variational formulation of (1) is given by

$$\text{Find } u \in V := H_0^1(\Omega) \ : \quad a_\Omega(u,v) = l_\Omega(v), \quad \forall v \in V \tag{2}$$

with

$$\begin{aligned} a_G(u,v) &:= \int_G \epsilon\nabla u\nabla v dx + \int_G (c - \frac{1}{2}\nabla\cdot\mathbf{b})uv dx + \frac{1}{2}\int_G (\mathbf{b}\cdot\nabla uv - \mathbf{b}\cdot\nabla vu)\, dx, \\ l_G(v) &:= \int_G fv dx \end{aligned}$$

for a domain $G \subseteq \Omega$. The weak formulation is obtained in the usual way using integration by parts on $\int_\Omega(-\epsilon\triangle u + \frac{1}{2}\mathbf{b}\cdot\nabla u)v dx$. We require the existence of a constant $\alpha > 0$ such that $c - \frac{1}{2}\nabla\cdot\mathbf{b} \geq \alpha > 0$ is satisfied almost everywhere in Ω. Thus we get by virtue of the Lax-Milgram lemma, that there exists a unique solution of (2).

[1]University of Göttingen, Math. Departm., NAM, D-37073 Göttingen,
E-mail: grapin@math.uni-goettingen.de
[2]University of Göttingen, Math. Departm., NAM, D-37073 Göttingen,
E-mail: lube@math.uni-goettingen.de

The Robin-Robin algorithm

Description of the algorithm

We begin by partitioning the domain Ω into N nonoverlapping subdomains $\Omega_1, \ldots, \Omega_N$, i.e. $\overline{\Omega} = \bigcup_{i=1}^{N} \overline{\Omega}_i$ and $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$, where each subdomain $\Omega_i,\ i \in \{1, \ldots, N\}$, is itself a Lipschitz domain with piecewise smooth boundary. We denote the interfaces by $\Gamma_i := \partial\Omega_i \setminus \partial\Omega$ and $\Gamma_{ij} := \partial\Omega_i \cap \partial\Omega_j$ for $i \neq j$. We identify Γ_{ij} with Γ_{ji}. For simplicity we assume that the decomposition is stripwise, i.e $\Gamma_{ij} \neq \Gamma_{kl}$ implies $\inf_{\substack{x\in\Gamma_{ij}\\ y\in\Gamma_{kl}}} |x-y| > 0$. Now the main idea of the algorithm is straight forward. In each subdomain Ω_i a local problem with an iterative coupling of the interface values has to be solved. Defining for $i = 1, \ldots N$

$$\phi_i(u) := \epsilon \frac{\partial u}{\partial \mathbf{n}_i} + \left(-\frac{1}{2}\mathbf{b}\cdot\mathbf{n}_i + z_i\right) u$$

with a strictly positive function $z_i \in L^\infty(\Omega_i)$ and the outward normal $\mathbf{n}_i$ we get

The Robin-Robin algorithm - version 1
Solve for $k \geq 1$ and for all $i = 1, \ldots, N$:

$$\begin{cases} Lu_i^k &= f & \text{in } \Omega_i \\ u_i^k &= 0 & \text{on } \partial\Omega \cap \partial\Omega_i \\ \phi_i(u_i^k) &= \theta\phi_i(u_j^{k-1}) + (1-\theta)\phi_i(u_i^{k-1}) & \text{on } \Gamma_{ij},\ j \neq i, \end{cases} \tag{3}$$

where $\theta \in (0, 1]$ is a relaxation parameter and u_i^0 is a given initial guess. An appropriate choice of the parameter z_i is a difficult problem. We will give some suggestions later.
To derive a variational formulation of the algorithm, we first define the spaces V_i by $V_i := V|_{\Omega_i}$ and W_{ij} by $W_{ij} := Tr_{\Gamma_{ij}} V$. Thus W_{ij} consists of traces on Γ_{ij} of functions belonging to V. Taking into account that the decomposition is stripwise, we obtain $W_{ij} = H_{00}^{\frac{1}{2}}(\Gamma_{ij})$, or $W_{ij} = H^{\frac{1}{2}}(\Gamma_{ij})$ if Γ_{ij} is closed.
Using the notation from above we can rewrite the DD-algorithm (3) starting with an initial guess $(\Lambda_{ij}^0)_{i\neq j}$ (cf. Robin-Robin alg. - vers. 2).

The variational algorithm is equivalent to algorithm (3). The well-posedness of the variational formulation has been shown (cf. [Ott99], Theorem 3.2):

Theorem 1 *If $z_i \in L^\infty(\Gamma_i)$ is a strictly positive function for all i then the Robin-Robin algorithm is well defined, i.e. at each iteration step k each subdomain boundary value problem has a unique solution in V_i.*
Moreover if the initial guess Λ_{ij}^0 belongs to $L^2(\Gamma_{ij})$ for all $i \neq j$, then $\Lambda_{ij}^k \in L^2(\Gamma_{ij})$ for all iteration steps k as well.

Convergence results and optimal choice of the parameter

For the remainder of this paper we restrict ourselves to the case of two subdomains. But keep in mind that most of the results remain valid in the multidomain case and can also be applied to the discrete stabilized case (cf. [LMO00]). In this setting we have the following convergence result (cf. [Ott99], Th. 3.3; [LMO00], Th. 3.1):

The Robin-Robin algorithm - version 2

1. given $(\Lambda_{ij}^n)_{i\neq j}$ solve for $i = 1, \ldots, N$

$$\begin{cases} \text{Find } u_i^{n+1} \in V_i \text{ with} \\ \qquad a_{\Omega_i}(u_i^{n+1}, v_i) \quad + \langle z_i u_i^{n+1}, v_i\rangle_{\Gamma_i} = \\ \qquad\qquad\qquad\quad l_{\Omega_i}(v_i) + \sum_{j\neq i}\langle \Lambda_{ji}^n, v_i\rangle_{\Gamma_{ij}}, \quad \forall v_i \in V_i \end{cases}$$

2. update $\Lambda_{ij}^{n+1} \in W_{ij}^*$ for $i \neq j$ by

$$\langle \Lambda_{ij}^{n+1}, \phi\rangle_{\Gamma_{ij}} = \theta\langle (z_i + z_j)u_i^{n+1}, \phi\rangle_{\Gamma_{ij}} - \theta\langle \Lambda_{ji}^n, \phi\rangle_{\Gamma_{ij}} + (1-\theta)\langle \Lambda_{ij}^n, \phi\rangle_{\Gamma_{ij}}$$

   for all $\phi \in W_{ij}$.

3. until convergence: $n \mapsto n+1$ and goto step 1.

Theorem 2 *Let be $c - \frac{1}{2}\nabla\cdot\mathbf{b} \geq \alpha > 0$. Furthermore, let be $z = z_1 = z_2$ on $\Gamma := \Gamma_{12}$ and $\Lambda^0 := \Lambda_{12}^0 \in L^2(\Gamma)$. Then the sequences $\{u_i^n\}_n$ for $i = 1, 2$ converge according to*

$$\|u_i^n - u_i\|_{V_i} \to 0, \quad n \to \infty,$$

where $u_i := u|_{\Omega_i}$ is the restriction of the global solution $u \in V$ of (2) onto Ω_i.

Remark 1 *Unfortunately the proof contains no indication about the speed of convergence. With help of Fourier analysis it can be shown that, even in a simple case, the convergence speed is not linear (cf. the section about Fourier analysis).*

If we require some further assumptions, it is possible to give an a-posteriori error estimate. The proof is similiar to [OLM01], [LMO00]. The local error is measured in the ϵ-dependent norm

$$|||u|||_i^2 := \epsilon|u|_{1,\Omega_i}^2 + \|\sqrt{c}u\|_{0,\Omega_i}^2, \qquad \forall u \in V_i, \quad i = 1, 2.$$

Theorem 3 *Let $\nabla\cdot\mathbf{b} = 0$, $c \geq 0$ and $\theta = 1$. Let both subdomains be connected with $\partial\Omega$, i.e. $\partial\Omega \cap \partial\Omega_i \neq \emptyset$ for $i = 1, 2$. Defining $\|\cdot\|_{\infty,\Gamma} := \|\cdot\|_{L^\infty(\Gamma)}$, $\|\cdot\|_{2,\Gamma} := \|\cdot\|_{L^2(\Gamma)}$ we get for $i = 1, 2$ and $j = 3 - i$*

$$|||u_i^{k+1} - u_i|||_i \leq C\epsilon^{-\frac{1}{2}}\|z_i - \frac{1}{2}\mathbf{b}\cdot\mathbf{n}_i\|_{\infty,\Gamma}\|u_2^k - u_1^{k+1}\|_{2,\Gamma} + CL_j\|u_2^k - u_1^{k+1}\|_{W_{12}} \quad (4)$$

where

$$L_j = \sqrt{\epsilon} + C_{F,j}\sqrt{C_{\infty,j}} + B_{\infty,j}\min\{\frac{1}{\sqrt{C_{0,j}}}, \frac{C_{F,j}}{\sqrt{\epsilon}}\},$$

$C_{\infty,i} := \|c\|_{\infty,\Omega_i}$, $B_{\infty,i} := \max_{l=1,2}\|b_l\|_{\infty,\Omega_i}$ and $C_{0,i} := ess\ \inf_{x\in\Omega_i} c(x)$. $C_{F,i}$ is the constant of the Friedrich's inequality and C is a constant independent of ϵ, $\mathbf{b}$ and c.

Remark 2 *The estimate (4) bounds the global error by interface terms, which can be computed without any knowledge about the global solution. Therefore we can simply control the convergence within a practical implementation.*

Now we derive a suitable choice of the parameter z_i from the a-posteriori error estimate using a method which was applied in [OLM01] to the Oseen equations. To be consistent with the case $\epsilon = 0$ we propose

$$z_i = \frac{1}{2}|\mathbf{b} \cdot \mathbf{n}_i| + R \quad \text{with} \quad \lim_{\epsilon \to 0} R = 0 \text{ on } \Gamma_i^0 \cup \Gamma_i^+$$

where $\Gamma_i^0 \cup \Gamma_i^+ := \{x \in \Gamma_i \mid (\mathbf{b} \cdot \mathbf{n}_i)(x) \geq 0\}$ is the non-inflow part. Then equilibration of the terms on the right hand side of (4) motivates for $z = z_1 = z_2$ the choice

$$\begin{aligned} z_i &= \frac{1}{2}|\mathbf{b} \cdot \mathbf{n}_i| + \max_{j=1,2} \frac{L_j}{\epsilon^{-\frac{1}{2}}} \\ &= \frac{1}{2}|\mathbf{b} \cdot \mathbf{n}_i| + \sqrt{\epsilon} \max_{j=1,2} \left(\sqrt{\epsilon} + C_{F,j}\sqrt{C_{\infty,j}} + B_{\infty,j} \min\{\frac{1}{\sqrt{C_{0,j}}}, \frac{C_{F,j}}{\sqrt{\epsilon}}\} \right). \end{aligned} \quad (5)$$

Remark 3 *Further suggestions and numerical experiments concerning an appropriate choice of z_i can be found in [LMO00]. Other approaches like the technique of absorbing boundary conditions [NR95] or asymptotic analysis for singularly perturbed problems [Ott99] yield similiar results. In the last section additional variants are derived from Fourier analysis.*

The Robin-Robin preconditioner for the Schur complement equation

The Schur complement equation

It is possible to reduce the original problem (1) on Ω to an interface problem on $\Gamma := \bigcup_{i=1}^N \Gamma_i$. This is a very natural way to transform the global problem into local problems on the subdomains Ω_i (cf. [QV99]).
First we introduce two operators, which extend functions from the interface Γ_i to the subdomain Ω_i. Defining the trace spaces $W := Tr_\Gamma V$ and $W_i := H_{00}^{\frac{1}{2}}(\Gamma_i)$ the $a_{\Omega_i}(\cdot,\cdot)$-extension $Tr_i^{-1}w_i$ for a $w_i \in W_i$ is given by $Tr_i^{-1}w_i = y_i \in V_i$ with

$$a_{\Omega_i}(y_i, v_i) = 0, \quad \forall v_i \in H_0^1(\Omega_i) \quad \text{and} \quad Tr_{\Gamma_i}(y_i) = w_i. \quad (6)$$

Remark 4 *If y_i is regular enough, this implies that y_i satisfies*

$$(Ly_i)(x) = 0 \text{ in } \Omega_i, \quad y_i = w_i \text{ on } \Gamma_i, \quad y_i = 0 \text{ on } \partial\Omega_i \cap \partial\Omega.$$

Analogously the extension operator Tr_i^{-*} for a $w_i \in W_i$ is defined by $Tr_i^{-*}w_i = y_i$ with

$$a_{\Omega_i}(v_i, y_i) = 0 \quad \forall v_i \in H_0^1(\Omega_i), \quad \text{and} \quad Tr_{\Gamma_i}(y_i) = w_i. \quad (7)$$

The extensions are well-posed (cf. [QV99], ch. 5.1):

Lemma 1 *Equations (6) and (7) have unique solutions. They satisfy the a-priori estimates*

$$\|Tr_i^{-1}w_i\|_{1,\Omega_i} \leq C\|w_i\|_{W_i} \quad \textit{and} \quad \|Tr_i^{-*}w_i\|_{1,\Omega_i} \leq C\|w_i\|_{W_i}$$

for all $w_i \in W_i$ and $i = 1, \ldots N$.

Decomposing $u, v \in V$ into $u = Tr^{-1}(Tr_\Gamma u) + \sum_{i=1}^N u_i^0$, $v = Tr^{-*}(Tr_\Gamma v) + \sum_{i=1}^N v_i^0$ with $u_i^0, v_i^0 \in H_0^1(\Omega_i)$ for all $i = 1, \dots N$ and $Tr^{-1} = Tr_i^{-1}$, $Tr^{-*} = Tr_i^{-*}$ on subdomain Ω_i, the Schur complement equation can be derived from (2). The *Schur complement equation* is given by

$$\text{Find } \bar{u} \in W \ : \quad \langle S\bar{u}, \bar{v}\rangle = \sum_{i=1}^N \langle S_i\bar{u}, \bar{v}\rangle = \langle F, \bar{v}\rangle, \quad \forall \bar{v} \in W \tag{8}$$

with

$$\langle S_i\bar{u}, \bar{v}\rangle = a_{\Omega_i}(Tr_{\Gamma_i}^{-1}\bar{u}, Tr_{\Gamma_i}^{-*}\bar{v}) \quad \text{and} \quad \langle F, \bar{v}\rangle = \sum_{i=1}^N l_{\Omega_i}(Tr_{\Gamma_i}^{-*}\bar{v}).$$

It can be proved, that the *Steklov-Poincaré operator* S is continuous and coercive on W (cf. [QV99], ch. 5.1). Thus we have proven the following:

Lemma 2 *There exists a unique solution $\bar{u} \in W$ of (8). Furthermore, if $u \in V$ is a solution of (2) then $u|_\Gamma$ is a solution of (8).*

The Robin-Robin Preconditioner

Here the Schur complement equation (8) is solved by a preconditioned Richardson iteration:

$$\lambda_{k+1} = \lambda_k + \theta T(F - S\lambda_k), \quad k \in \mathbb{N} \tag{9}$$

with an initial guess $\lambda_0 \in W$, a relaxation parameter $\theta > 0$ and a preconditioner T.
From now on we consider again the case of two subdomains. The Robin-Robin preconditioner is thus given by a sum of weighted inverses of local Steklov-Poincaré operators (cf. [QV99], [BS00], [ATNV00]):

$$T = \sigma_1 S_1^{-1} + \sigma_2 S_2^{-1}$$

with $\sigma_1, \sigma_2 \geq 0$. The operator T is continuous and coercive for $\sigma_1 + \sigma_2 > 0$. Thus, by virtue of the Lax-Milgram Lemma, T^{-1} exists and is also continuous and coercive (cf. [QV99], p. 108). Unfortunately up to now linear convergence can only be proven in the diffusion dominated case for two subdomains (cf. [QV99], ch. 5.1).
It is interesting that this method can also be interpreted as an iteration-by-subdomains method (cf. [BS00]). To see this note that (9) can be written equivalently as: given $\lambda_{k-1} \in W$, solve for $i = 1, 2$ the Dirichlet problems and the mixed Dirichlet-Robin problems

$$\left\{\begin{array}{rcll} Lw_i^k &=& f & \text{in } \Omega_i \\ w_i^k &=& 0 & \text{on } \partial\Omega_i \setminus \Gamma \\ w_i^k &=& \lambda^{k-1} & \text{on } \Gamma \end{array}\right. \qquad \left\{\begin{array}{rcll} Ly_i^k &=& 0 & \text{in } \Omega_i \\ y_i^k &=& 0 & \text{on } \partial\Omega_i \setminus \Gamma \\ \Phi_i(y_i^k) &=& \Phi_i(w_i^k) & \\ &+& \Phi_i(w_j^k) & \text{on } \Gamma \end{array}\right. \tag{10}$$

with $\Phi_i(y) := \epsilon\frac{\partial y}{\partial \mathbf{n}_i} - \frac{1}{2}\mathbf{b}\cdot\mathbf{n}_i y$ on Γ. Finally, update the interface function by

$$\lambda^k := \lambda^{k-1} + \theta\left(\sigma_1 y_1^k|_\Gamma + \sigma_2 y_2^k|_\Gamma\right).$$

Fourier Analysis

In this section we consider the special case that the flow $\mathbf{b}$ and the reaction term c are constants. The domain Ω is given by $(0, L) \times (0, 1)$ and is divided into $\Omega_1 = (0, A) \times (0, 1)$ and $\Omega_2 = (A, L) \times (0, 1)$ (cf. figure (a)). Now we carry out a Fourier analysis for both methods.

The Robin-Robin algorithm

Via separation of variables we obtain the following representation of the error $e_i^k := u_i^k - u|_{\Omega_i}$ of the k-th step:

$$\begin{aligned} e_1^k(x,y) &= \exp\left(\frac{\mathbf{b}\cdot(x,y)}{2\epsilon}\right)\sum_{l=1}^{\infty} F_{1,l}^k \sinh(\nu_l x)\sin(l\pi y), \\ e_2^k(x,y) &= \exp\left(\frac{\mathbf{b}\cdot(x,y)}{2\epsilon}\right)\sum_{l=1}^{\infty} F_{2,l}^k \sinh(\nu_l (L-x))\sin(l\pi y) \end{aligned}$$

where $\nu_l^2 := \frac{|\mathbf{b}|^2}{4\,\epsilon^2} + \frac{c}{\epsilon} + l^2\pi^2$. (cf. GASTALDI ET. AL. [GGQ96]). Inserting the boundary condition on $\Gamma = \Gamma_{12}$ yields in the case of $\theta = 1$ the recursion formulas

$$F_{i,l}^k = K_l^{RR} F_{i,l}^{k-2}$$

for $i = 1, 2$ where

$$K_l^{RR} = \frac{(-z + \epsilon\nu_l \coth(\nu_l A))\,(z - \epsilon\nu_l \coth(\nu_l (L-A)))}{(z + \epsilon\nu_l \coth(\nu_l A))\,(-z - \epsilon\nu_l \coth(\nu_l (L-A)))}.$$

F. NATAF and F. ROGIER [NR95] perform a similiar analysis for the case of infinite strips with Fourier transform techniques. They require exact boundary condition for the first Fourier mode to yield the following choice for the free parameter:

$$z = \frac{1}{2}\sqrt{b_1^2 + 4\epsilon c}. \tag{11}$$

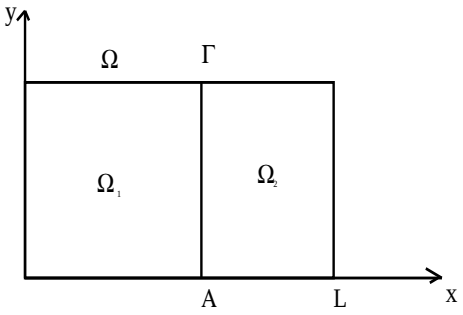
Analogously, assuming exact boundary conditon for the first Fourier mode, we get

$$z = \frac{1}{2}\sqrt{|\mathbf{b}|^2 + 4\epsilon c + 4\pi^2\epsilon^2}, \tag{12}$$

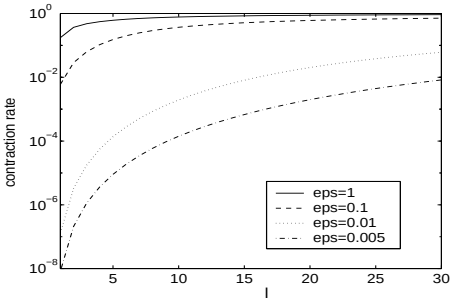
where the term $\coth(\nu_1 A) \approx 1$ is neglected. In [JNR01] the choice (11) is improved by adding additional interface terms in the tangential direction. Then the constants are determined by minimizing the convergence ratio for a certain range of wave numbers. Numerical experiments of the choice (5) resulting from the a-posteriori estimate show, that this choice also damps the lower wave numbers very well (cf. [LMO00]).

With help of the recursion formulas it can be shown directly, that the algorithm converges for this special domain decomposition and positive constant z. The convergence rate, however, is in general not linear.

We illustrate the contraction rates $|K_l^{RR}|$ in figure (b) for different ϵ, the choice (12) and the parameters $L = 1$, $A = 0.1$, $\mathbf{b} = (1, 1)^t$, $c = 1$. We observe that the contraction rates $|K_l^{RR}|$ tend to 1 for $l \to \infty$. Thus higher modes are reduced slower. Further we recognize that the algorithm works well for the case of small ϵ.



(a) The domain Ω



(b) The contraction rate $|K_l^{RR}|$

Robin-Robin preconditioner for the Schur complement equation

Next we examine the preconditioned Richardson iteration of the Schur complement equation. With help of the differential interpretation of the algorithm it is also possible to apply Fourier analysis.
Denoting the error at the k-th step by $\tilde{e}_i^k := w_i^k - u|_{\Omega_i}$, where w_i^k is the solution of the Dirichlet problem in (10), the following representation can be derived in a similiar manner described for the Robin-Robin algorithm:

$$\begin{aligned}\tilde{e}_1^k(x,y) &= \exp\left(\frac{\mathbf{b}\cdot(x,y)}{2\epsilon}\right)\sum_{l=1}^{\infty} C_{1,l}^k \sinh(\nu_l x)\sin(l\pi y)\\ \tilde{e}_2^k(x,y) &= \exp\left(\frac{\mathbf{b}\cdot(x,y)}{2\epsilon}\right)\sum_{l=1}^{\infty} C_{2,l}^k \sinh(\nu_l (L-x))\sin(l\pi y).\end{aligned}$$

Inserting again the boundary conditions on Γ yields

$$C_{2,l}^{k+1} = \frac{\sinh(\nu_l A)}{\sinh(\nu_l (L-A))} C_{1,l}^{k+1} \quad \text{and} \quad C_{1,l}^{k+1} = K_l C_{1,l}^k$$

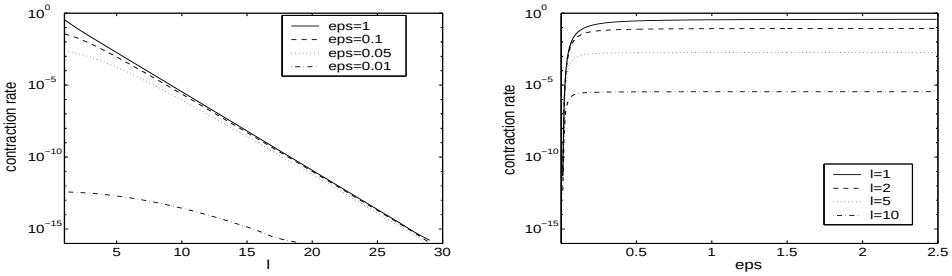
with

$$K_l = 1 - \theta\{\sigma_1 + \sigma_2 + \sigma_1 \coth(\nu_l (L-A))\tanh(\nu_l A) + \sigma_2 \coth(\nu_l A)\tanh(\nu_l (L-A))\}.$$

Thus, again, the convergence behaviour depends on the contraction rates $|K_l|$. In figure (c) and (d) the contraction rates are illustrated for the following choice of the parameters: $L = 1$, $A = 0.1$, $\mathbf{b} = (1,1)^t$, $c = 1$, $\sigma_1 = \sigma_2 = \frac{1}{4}$, $\theta = 1$. In contrast to the Robin-Robin algorithm we can state that the contraction rates $|K_l|$ tend to 0 for $l \to \infty$. This allows us to prove linear convergence in the H^1-norm for this special case. Further we observe that for $\epsilon \to 0$ the contraction rates K_l tend to 0.

Comparison of the two methods

First we consider the convergence behaviour. Numerical experiments indicate that convergence of the Schur complement method is linear, but up to now, it is not proved in the general



(c)+(d) Contraction rates $|K_l|$ for different ϵ resp. l

case. Conversely, it can be shown, that the Robin-Robin method converges, but the convergence is in general not linear. In the case of the Robin-Robin algorithm one need compute only one local problem in each subdomain per iteration step. The other method needs the computation of two local problems per iteration step. Thus the Robin-Robin method is easier to implement. A problem of the Robin-Robin algorithm is an appropriate choice of the free parameter z_i. Numerical experiments have shown that the algorithm is sensitive to the choice of z_i.

References

[ATNV00]Y. Achdou, P. Le Tallec, F. Nataf, and M. Vidrascu. A domain decoposition preconditioner for an advection-diffusion problem. *Comp. Meth. Appl. Mech. Engng*, 184:145–170, 2000.

[BS00]Luigi C. Berselli and Fausto Saleri. New substructuring domain decomposition methods for advection-diffusion equations. *J. Comput. Appl. Math.*, 116:201–220, 2000.

[GGQ96]Fabio Gastaldi, Lucia Gastaldi, and Alfio Quarteroni. Adaptive domain decomposition methods for advection dominated equations. *East-West J. Numer. Math.*, 4:165–206, 1996.

[JNR01]Caroline Japhet, Frederic Nataf, and Francois Rogier. The optimized order 2 method. application to convection-diffusion problems. *Future Generation Computer Systems FUTURE*, 18, 2001.

[LMO00]Gert Lube, Lars Müller, and Frank-Christian Otto. A non-overlapping domain decomposition method for the advection-diffusion problem. *Computing*, 64:49–68, 2000.

[NR95]Frédéric Nataf and Francois Rogier. Factorization of the convection-diffusion operator and the Schwarz algorithm. M^3AS, 5(1):67–93, 1995.

[OLM01]Frank-Christian Otto, Gert Lube, and Lars Müller. An iterative substructuring method for div-stable finite element approximations of the oseen problem. *Computing*, 67:91–117, 2001.

[Ott99]Frank-Christian Otto. *A non-overlapping Domain Decomposition Method for Elliptic Equations*. PhD thesis, Universität Göttingen, 1999.

[QV99]Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

Part IV

Applications

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

46 Analysis of a defect correction method for computational aeroacoustics

G. S. Djambazov, C.-H. Lai [1], K. A. Pericleous and Z-K Wang [2]

Introduction

Many problems of fundamental and practical importance are of multi-scale nature. As a typical example, the velocity field in turbulent transport problems fluctuates randomly and contains many scales depending on the Reynolds number of the flow. In another typical example, which is the main concern of this paper, sound waves are several orders of magnitude smaller than the pressure variations in the flow field that account for flow acceleration. These sound waves are manifested as pressure fluctuations which propagate at the speed of sound in the medium, not as a transported fluid quantity. As a result, numerical solutions of the Navier-Stokes equations which describe fluid motion do not resolve the small scale pressure fluctuations. The direct numerical simulation to include the above multiple scale problems is still an expensive tool for sound analysis [1].

In essence, there are at least three different scales embedded in the flow variables, namely (i) the mean flow, (ii) flow perturbations or aerodynamic sources of sound, and (iii) the acoustic perturbation. While flow perturbation or aerodynamic sources of sound may be easier to recover, it is not true for the acoustic perturbation because of its comparatively small magnitude. ¿From an engineering perspective, much of the larger scales behaviour may be resolved with the state-of-the-art CFD packages which implement various numerical methods of solving Navier-Stokes equations. This paper examines, in more detail, a defect correction method, first proposed in [2], for the recovery of smaller scales that have been left behind. The authors have demonstrated the accurate computation of (i) and (ii) in [3][4][5]. In the present study, a two-scale decomposition of flow variables is considered, i.e. the flow variable U is written as $\bar{u} + u$, where $\bar{u}$ denotes the mean flow and part of aerodynamic sources of sound and u denotes the remaining part of the aerodynamic sources of sound and the acoustic perturbation.

The concept of defect correction [6] has been used in various contexts since the early days. A typical example of defect correction is the computation of a refined approximation to the approximate solution $\bar{x}$ of the nonlinear equation $f(x) = 0$. Since $\bar{x}$ is an approximate solution, the defect may be computed as $-f(\bar{x})$. The idea of a defect correction method is to use a modified/derived version of the original problem such as the one defined by $\bar{f}(x) \equiv f'(\bar{x})(x - \bar{x}) + f(\bar{x}) = 0$. If one replaced $x - \bar{x}$ as v, then v is the correction computed by solving $f'(\bar{x})v = -f(\bar{x})$ and a refined approximation can be evaluated by using $x := \bar{x} + v$. More details in expanding the concept to discretised problems and multigrid methods can be found in [6]. Here, the authors would like to concentrate on using the defect correction concept at the level of the physical problem rather than the discretised problem. For a given mathematical problem and a given approximate solution, the residue or defect may be treated as a quantity to measure how well the problem has been solved. Such information may

[1] C.H.Lai@gre.ac.uk
[2] Same address for all authors: School of Computing and Mathematical Sciences, University of Greenwich, 30 Park Row, Greenwich, London SE10 9LS, UK

then be used in a modified/derived version of the original mathematical problem to provide an appropriate correction quantity. The correction can then be applied to correct the approximate solution in order to obtain a refined approximate solution to the original mathematical problem.

This paper follows the basic principle of the defect correction as discussed above and applies it to the recovery of the propagating acoustic perturbation. The method relies on the use of a lower order partial differential equation defined on the same computational domain where a residue exists such that the acoustic perturbation may be retrieved through a properly defined coarse mesh.

This paper is organised as follows. First, the derivation of a lower order partial differential equation resulting from the Navier-Stokes equations is given. Truncation errors due to the model reduction are examined. Second, accurate representation of residue on a coarse mesh is discussed. The coarse mesh is designed in such a way as to allow various frequencies of noise to be studied. Suitable interpolation operators are studied for the two different meshes. Third, numerical tests are performed for different mesh parameters to illustrate the concept. Finally, future work is discussed.

The defect correction method

The aim here is to solve the non-linear equation

$$\mathcal{L}\{U\}U := \mathcal{L}\{\bar{u}+u\}(\bar{u}+u) = 0 \tag{1}$$

where $\mathcal{L}\{U\}$ is a non-linear operator depending on U. For simplicity, U is considered to have two different scales of magnitudes as $\bar{u}+u$. Here $\bar{u}$ is the mean flow and u is the acoustic perturbation as described in Section 46. Note that $u \ll \bar{u}$ and that

$$\frac{1}{\delta t}\int_{t_0}^{t_0+\delta t} u dt \to 0$$

with δt much larger than any significant period of the perturbation velocity. The problem here is thus purely related to the scale of magnitude. In the case of sound generated by the motion of fluid, it is natural to imagine $\mathcal{L}$ as the Navier-Stokes operator. For a 2-D problem,

$$\bar{u} = \begin{bmatrix} \bar{\rho} \\ \bar{v}_1 \\ \bar{v}_2 \end{bmatrix} \quad u = \begin{bmatrix} \rho \\ v_1 \\ v_2 \end{bmatrix}$$

where ρ is the density of fluid and v_1 and v_2 are the velocity components along the two spatial axes. Using the summation notation of subscripts, the 2-D Navier-Stokes problem $\mathcal{L}\{u\}u = 0$ is written as

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho v_j)}{\partial x_j} = 0,$$

$$\frac{\partial v_i}{\partial t} + v_j\frac{\partial v_i}{\partial x_j} + \frac{1}{\rho}\frac{\partial P}{\partial x_i} - \frac{\mu}{\rho}\nabla^2 v_i = 0$$

where P is the pressure and $(\mu/\rho)\nabla^2 v_i$ is the viscous force along i-th axis.

Suppose (1) may be split and re-written as

$$\mathcal{L}\{\bar{u}+u\}(\bar{u}+u) \equiv \mathcal{L}\{\bar{u}\}\bar{u} + E\{\bar{u}\}u + K[\bar{u},u] \tag{2}$$

where $\mathcal{L}\{\bar{u}\}$ and $E\{\bar{u}\}$ are operators depending on the knowledge of $\bar{u}$ and $K[\bar{u},u]$ is a functional depending on the knowledge of both $\bar{u}$ and u. Following the concept of defect correction, $\bar{u}$ may be considered as an approximate solution to (1). Hence one can evaluate the residue of (1) as

$$R \equiv \mathcal{L}\{\bar{u}+u\}(\bar{u}+u) - \mathcal{L}\{\bar{u}\}\bar{u} = -\mathcal{L}\{\bar{u}\}\bar{u}$$

which may then be substituted into (2) to give

$$E\{\bar{u}\}u + K[\bar{u},u] = R \tag{3}$$

In many cases, $K[\bar{u},u]$ is small and can then be neglected. In those cases, the problem in (3) is a linear problem and may be solved more easily to obtain the acoustics fluctuation u. A non-linear iterative solver is required in order to obtain u for cases when $K[\bar{u},u]$ is not negligible. Finally, to obtain the approximate solution $\bar{u}$, one only needs to solve $\mathcal{L}\{\bar{u}\}\bar{u} = 0$. Expanding $\mathcal{L}\{\bar{u}+u\}(\bar{u}+u) = 0$ for $\mathcal{L}$ being the Navier-Stokes operator and re-arranging we obtain

$$\frac{\partial\rho}{\partial t} + \bar{v}_j\frac{\partial\rho}{\partial x_j} + \bar{\rho}\frac{\partial v_j}{\partial x_j} + [v_j\frac{\partial(\bar{\rho}+\rho)}{\partial x_j} + \rho\frac{\partial(\bar{v}_j+v_j)}{\partial x_j}] = -[\frac{\partial\bar{\rho}}{\partial t} + \bar{v}_j\frac{\partial\bar{\rho}}{\partial x_j} + \bar{\rho}\frac{\partial\bar{v}_j}{\partial x_j}]$$

and

$$\frac{\partial v_i}{\partial t} + \bar{v}_j\frac{\partial v_i}{\partial x_j} + \frac{1}{\bar{\rho}}\frac{\partial P}{\partial x_i} - \frac{\mu}{\bar{\rho}}\nabla^2 v_i \tag{4}$$

$$+[\frac{\rho}{\bar{\rho}}\frac{\partial(\bar{v}_i+v_i)}{\partial t} + (v_j + \frac{\rho}{\bar{\rho}}(\bar{v}_j+v_j))\frac{\partial(\bar{v}_i+v_i)}{\partial x_j}] = -[\frac{\partial\bar{v}_i}{\partial t} + \bar{v}_j\frac{\partial\bar{v}_i}{\partial x_j} + \frac{1}{\bar{\rho}}\frac{\partial\bar{P}}{\partial x_i} - \frac{\mu\nabla^2\bar{v}_i}{\bar{\rho}}]$$

It can be seen that (4) may be written in the form of (3) where

$$E\{\bar{u}\}u = \begin{bmatrix} \frac{\partial\rho}{\partial t} + \bar{v}_j\frac{\partial\rho}{\partial x_j} + \bar{\rho}\frac{\partial v_j}{\partial x_j} \\ \frac{\partial v_i}{\partial t} + \bar{v}_j\frac{\partial v_i}{\partial x_j} + \frac{1}{\bar{\rho}}\frac{\partial P}{\partial x_i} - \frac{\mu}{\bar{\rho}}\nabla^2 v_i \end{bmatrix} \tag{5}$$

$$K[\bar{u},u] = \begin{bmatrix} v_j\frac{\partial(\bar{\rho}+\rho)}{\partial x_j} + \rho\frac{\partial(\bar{v}_j+v_j)}{\partial x_j} \\ \frac{\rho}{\bar{\rho}}\frac{\partial(\bar{v}_i+v_i)}{\partial t} + (v_j + \frac{\rho}{\bar{\rho}}(\bar{v}_j+v_j))\frac{\partial(\bar{v}_i+v_i)}{\partial x_j} \end{bmatrix} \tag{6}$$

$$R = \begin{bmatrix} -[\frac{\partial\bar{\rho}}{\partial t} + \bar{v}_j\frac{\partial\bar{\rho}}{\partial x_j} + \bar{\rho}\frac{\partial\bar{v}_j}{\partial x_j}] \\ -[\frac{\partial\bar{v}_i}{\partial t} + \bar{v}_j\frac{\partial\bar{v}_i}{\partial x_j} + \frac{1}{\bar{\rho}}\frac{\partial\bar{P}}{\partial x_i} - \frac{\mu}{\bar{\rho}}\nabla^2\bar{v}_i] \end{bmatrix} \equiv -\mathcal{L}\{\bar{u}\}\bar{u} \tag{7}$$

¿From the knowledge of physics of fluids, the acoustic perturbations ρ and v_j are of very small magnitude (this is not true for their derivatives), therefore, K may be considered negligible due to the reason that any feedback from the propagating waves to the flow may be completely ignored, except in some cases of acoustic resonance, which we are not concerned with here.

Hence the equation $E\{\bar{u}\}u = R$, with E given by (5), which is known as the linearised Euler equation, can be solved in an easier way. The numerics and the techniques involved here are often referred to as Computational AeroAcoustics (CAA) methods.
The remaining question is to obtain the approximate solution $\bar{u}$ to the original problem (2). It is well known that CFD analysis packages provide excellent methods for the solution of $\mathcal{L}\{\bar{u}\}\bar{u} = 0$. Therefore one requires to use a Reynolds averaged Navier-Stokes package supplemented with turbulence models such as [7, 8] to provide a solution of $\bar{u}$. One requires $\bar{u}$ to be as accurate as possible to capture all the physics of interest, such as flow turbulence and the presence of vortices.
The use of a CFD analysis package effectively solves $\mathcal{L}\{\bar{u}\}\bar{u} = 0$ instead of $\mathcal{L}\{\bar{u}+u\}(\bar{u}+u) = 0$. Following the concept of truncation error in a finite difference method, the truncation error due to the removal of the perturbation part of the flow variable may be defined by

$$\tau = \mathcal{L}\{\bar{u}+u\}(\bar{u}+u) - \mathcal{L}\{\bar{u}\}(\bar{u}+u) \tag{8}$$

Using the relation $\mathcal{L}\{\bar{u}\}(\bar{u}+u) = \mathcal{L}\{\bar{u}\}\bar{u} + E\{\bar{u}\}u$, the truncation error in the present context is thus given by

$$\tau = K[\bar{u}, u] \tag{9}$$

Note that this truncation error is not related to the discretisation of continuous model.

A two-level multigrid method

In order to simulate accurately the approximate solution, $\bar{u}$, to the original problem, $\mathcal{L}U = 0$, the QUICK differencing scheme [9] is used which produces sufficiently accurate results of $\bar{u}$ for the purpose of evaluating the residue as defined in (7). A sufficiently fine mesh has to be used in order to preserve vorticity motion. However, much coarser mesh may be used for the numerical solutions of linearised Euler equations [3, 4, 5]. It certainly has to obey the Courant limit and also to account for the fact that the acoustic wavelength may be larger than a typical flow feature which needs to be resolved, e.g. a travelling vortex [10]. The present defect correction method requires to calculate the residue on the CFD mesh and to transfer these residuals onto the acoustic mesh. Physically, the residue is effectively the sound source that would have disappeared without the proper retrieval technique as discussed in this paper.
Let h denote the mesh to be used in the Reynolds averaged Navier-Stokes solver. Instead of evaluating $\bar{u}$, one would solve the discretised approximation $\mathcal{L}_h\bar{u}_h = 0$ to obtain $\bar{u}_h$. The residue on the fine mesh h can be computed as $\mathcal{L}\bar{u}_h$ by means of a higher order approximation [5]. Let H denote the mesh for the linearised Euler equations solver. Again instead of evaluating u, one would solve the discretised approximation $E_H\{\bar{u}_H\}u_H = R_H$ to obtain u_H. Here R_H is the projection of R onto the mesh H. Let $I_{\{h,H\}}$ be a restriction operator to restrict the residue computed on the fine mesh h to the coarser mesh H. The restricted residue can then be used in the numerical solutions of linearised Euler equations. Therefore the two-level numerical scheme is (for non-resonance problems) :

Solve $\mathcal{L}_h\bar{u}_h = 0$
$R_H := -I_{\{h,H\}}\mathcal{L}\bar{u}_h$
$\bar{u}_H := I_{\{h,H\}}\bar{u}_h$
Solve $E_H\{\bar{u}_H\}u_H = R_H$
$U_H := \bar{u}_H + u_H$

Here U_H denotes the discretised approximation of the resultant solution on mesh H. Note that R_H cannot be computed as $\mathcal{L}I_{\{h,H\}}\bar{u}_h$ because $\mathcal{L}$ is a non-linear operator.
In the actual implementation, a pressure-density relation which also defines the speed of sound c in air is used:

$$\frac{\partial P}{\partial \rho} = c^2 \approx 1.4\frac{\bar{P}}{\bar{\rho}} \tag{10}$$

and the first component of the linearised Euler equations in (5) becomes

$$\frac{\partial P}{\partial t} + \bar{v}_j\frac{\partial P}{\partial x_j} + \bar{\rho}c^2\frac{\partial v_j}{\partial x_j} = -c^2[\frac{\partial \bar{\rho}}{\partial t} + \bar{v}_j\frac{\partial \bar{\rho}}{\partial x_j} + \bar{\rho}\frac{\partial \bar{v}_j}{\partial x_j}] \tag{11}$$

The purpose of this substitution is to make sure that the new fluctuations P and v_i do not contain a hydrodynamic component, and hence can be resolved on regular Cartesian meshes [4] which is essential for the accurate representation of the acoustic waves or the fluctuation quantity u. On the other hand, an unstructured mesh may be used to obtain $\bar{u}_h$. The two different meshes overlap one another on the computational domain. The computational domain for the linearised Euler equations is not necessarily the same as the one for the CFD solutions. It must be large enough to contain at least the longest wavelength of a particular problem under consideration or a number of wavelengths where propagation is of interest. The numerical example as shown in Section 46 does not contain any complicating solid objects, the restriction operator $I_{\{h,H\}}$ may then be chosen as an arithmetic averaging process [10].

Numerical experiments with various grid parameters

The propagation of the following one-dimensional pulse is considered: an initial pressure distribution with a peak in the origin generates two opposite acoustic waves in both directions. The exact solution of this problem (12) can be verified by substitution in the linearised Euler equations.

$$\begin{array}{rcl} P & = & f(x-ct) + f(x+ct) \\ \bar{\rho}cv_1 & = & f(x-ct) - f(x+ct) \\ f(x) & = & \begin{cases} \frac{A}{2}(1+\cos 2\pi\frac{x}{\lambda}), |x| < \frac{\lambda}{2} \\ 0, |x| \geq \frac{\lambda}{2} \end{cases} \end{array} \tag{12}$$

Here A is the amplitude and λ is the wavelength of the two sound waves that start from the origin ($x = 0$) at $t = 0$. The example was reported in [2]. This paper provides a detailed numerical study on various aspects of the grid parameters being used in the two-level method. The CFD domain is of 12 wavelengths and the CAA domain is of 14 wavelengths.
The effects of the following parameters on the solution accuracy are studied. These parameters are (a) the ratio H:h, (b) number of points per wavelength, and (c) the restriction operator for residual transfer from fine grid to coarse grid. In all cases, the norm $\|P_H - P\|_\infty$ is compared. Here P_H is the approximation obtained on the coarse mesh (CAA) after correction and P is the exact solution of the pressure variable.
Let δt_h and δt_H be the step lengths in the temporal axis for the CFD mesh and the CAA mesh respectively. Figure 1 shows the effect on the accuracy for Case (a). Here δt_h and δt_H are

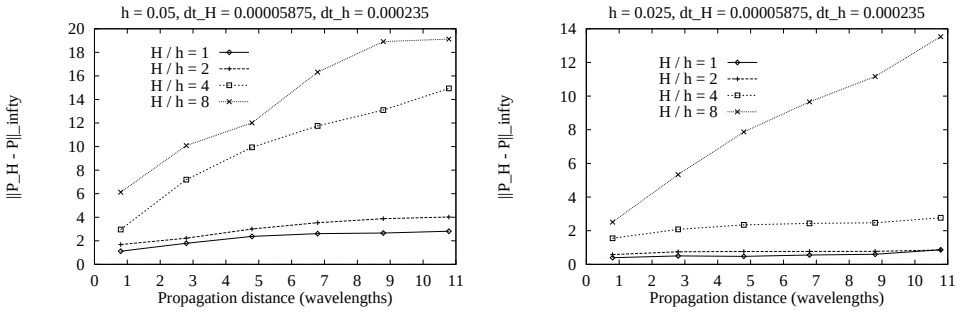


Figure 1: The effect of mesh ratio H:h on the accuracy.

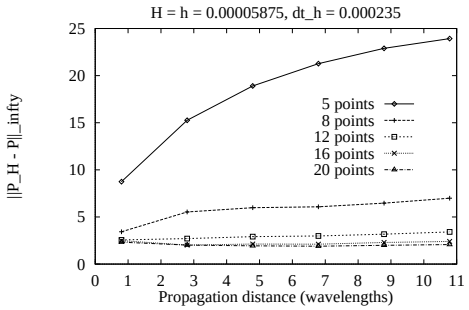


Figure 2: The effect of number of grid points per wavelength on the accuracy.

chosen to be 0.000235 and 0.00005875 respectively. Two different mesh sizes for the CFD are chosen and they are 0.05 and 0.025. It can be seen that when h is not fine enough, say h = 0.05, to resolve some of the physics, it is still possible to use the mesh $H = 2h$ or $H = h$ to recover the small scale signal. If a finer mesh was used, say $h = 0.025$, it is possible to use $H \leq 4h$. This property essentially links with the Courant number of the coarse mesh for CAA [5], i.e. H, and is also confirmed in the test performed for Case (b).

Figure 2 shows the effect on the accuracy for Case (b). The most accurate solution may be achieved with more than 12 grid points per wavelength, e.g. 16 or more grid points. This confirms the theoretical study based on Courant limits as discussed in [5]. For number of grid points per wavelength less than 12, the accuracy deteriorates very fast.

Figure 3 shows the effect on the accuracy for Case (c). The restriction operators being used in this test to transfer the function g_h onto the coarse mesh H includes

$$\text{3 point formula: } I_{\{h,2h\}} g_h = \frac{1}{4}(g_{i-1} + 2g_i + g_{i+1})$$

$$\text{5 point formula: } I_{\{h,4h\}} g_h = \frac{1}{12}(g_{i-2} + 2g_{i-2} + 6g_i + 2g_{i+1} + g_{i+2})$$

$$\text{7 point formula: } I_{\{h,6h\}} g_h = \frac{1}{16}(g_{i-3} + 2g_{i-2} + 3g_{i-1} + 4g_i + 3g_{i+1} + 2g_{i+2} + g_{i+3})$$

$$\text{9 point formula: } I_{\{h,8h\}} g_h = \frac{1}{48}(g_{i-4}+2g_{i-3}+6g_{i-2}+8g_{i-1}+14g_i+8g_{i+1}+6g_{i+2}+2g_{i+3}+g_{i+4})$$

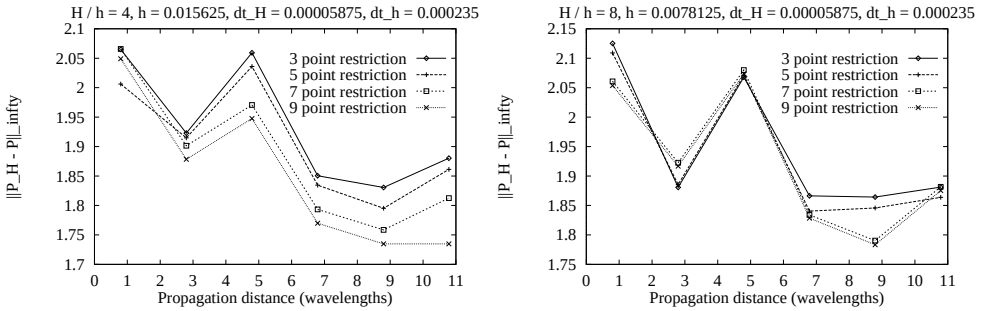


Figure 3: The effect of restriction operators on the accuracy.

For very fine CFD mesh, one can retrieve the small scale signal even on a relatively coarse mesh. In the present study, with $h = 0.0078125$ one can use $H \leq 8h$ while still maintaining the accuracy. The accuracy exhibited by using the coarse mesh $H = 8h = 0.0625$ is compatible with the result for Case (a) as depicted in Figure 1.

Conclusions

This paper provides a numerical method for the retrieval of sound signals using the defect correction method. A detailed numerical experiments to examine various grid parameters are provided. Truncation error of solving $\mathcal{L}\{\bar{u}\}\bar{u} = 0$ instead of $\mathcal{L}\{\bar{u} + u\}\bar{u} + u = 0$ is derived. The authors are currently applying the present method to sound propagation in vortex-vortex interactions.

References

[1] E.J. Avital, N.D. Sandham, and K.H. Luo. Mach wave radiation by time-developing mixing layers. Part II: Analysis of the source field. *Theoretical and Computational Fluid Dynamics*, 1998.

[2] G.S. Djambazov, C.-H. Lai, and K.A. Pericleous. A defect correction method for the retrieval of acoustic waves. In *Abstract: 12th Domain Decomposition Conference, Chiba, Japan, October 25 - 29, 1999*, page 93. 1999.

[3] G.S. Djambazov, C.-H. Lai, and K.A. Pericleous. Development of a domain decomposition method for computational aeroacoustics. In *Domain Decomposition Methods in Sciences and Engineering vol 9*. DDM.org, 1999.

[4] G.S. Djambazov, C.-H. Lai, and K.A. Pericleous. Efficient computation of aerodynamic noise. In *Contemporary Mathematics*, volume 218, pages 506–512. American Mathematical Society, 1998.

[5] G.S. Djambazov. *Numerical Techniques for Computational Aeroacoustics*. PhD thesis, University of Greenwich, 1998.

[6] K. Böhmer and H.J. Stetter. *Defect Correction Methods: Theory and Applications*. Springer-Verlag, 1984.

Figure 4: Hello

[7] N. Croft, K. Pericleous, and M. Cross. PHYSICA: A multiphysics environment for complex flow processes. In C. Taylor et al., editors, *Num. Meth. Laminar & Turbulent Flow '95*, volume 9, part 2, page 1269. Pineridge Press, U. K., 1995.

[8] CHAM Ltd, Wimbledon, UK. *PHOENICS, Version 2.1.3*, 1995.

[9] B.P. Leonard. A stable and accurate convective modelling procedure based on quadratic upstream interpolation. *Computer Methods in Applied Mechanics and Engineering*, 19:59–98, 1979.

[10] G.S. Djambazov, C.-H. Lai, and K.A. Pericleous. On the coupling of Navier-Stokes and linearised Euler equations for aeroacoustic simulation. *Comput Visual Sci*, 3:9–12, 2000.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

47 Nonoverlapping Domain Decomposition Algorithms for the System of Euler Equations

V. Dolean, D.Lanteri[1], F. Nataf[2]

Introduction

We report on our recent efforts concerning the construction of nonoverlapping additive Schwarz type algorithms for the solution of the system of Euler equations for compressible flows. We are specifically concerned with the construction of appropriate interface conditions that improve the convergence rate of the Schwarz algorithm. In Quarteroni and Stolcis[QS95], these transmission conditions are Dirichlet conditions for the characteristic variables corresponding to incoming waves. Such conditions can be qualified as "classical interface conditions" by opposition to more sophisticated formulations such as the "optimized interface conditions " studied in [JNR98] for an advection-diffusion equation. Here, we are interested in extending the principle of optimized interface conditions to the solution of the Euler equations. For this purpose, general type interface operators are introduced in the formulation of the additive Schwarz type algorithm. A convergence analysis is performed in the continuous case by considering the linearized Euler equations. An interface iteration is deduced from the formulation of the Schwarz algorithm in the Fourier space. In [DLN00]-[JNR01], such a convergence analysis has been performed by applying a classical diagonalization method to the operator matrix involved in the problem. In this study, we apply the Smith factorization theory[Gan66] in order to deduce a general form of the interface conditions. Then, the goal is to optimize the convergence rate with respect to certain parameters entering in the definition of these interface conditions. The analysis is limited to a two-subdomain decomposition in vertical strips.

Domain decomposition for the Euler equations

Mathematical model

The conservative form of the Euler equations is given by :

$$\frac{\partial W}{\partial t} + \frac{\partial F_1(W)}{\partial x} + \frac{\partial F_2(W)}{\partial y} = 0 \text{ with } W = \left(\rho \ , \ \rho\vec{V} \ , \ E\right)^T \tag{1}$$

where $W = W(\vec{x}, t)$ is the vector of conservative variables; $\vec{x}$ and t respectively denote the spatial and temporal variables while $\vec{\mathcal{F}}(W) = (F_1(W) \ , \ F_2(W))^T$ is the conservative flux whose components are given by :

[1]INRIA, 2004 Route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (FRANCE),E-Mail : Victorita.Dolean/Stephane.Lanteri@inria.fr

[2]CMAP, Ecole Polytechnique and CNRS, UMR7641,91128 Palaiseau Cedex (FRANCE), E-Mail : nataf@cmapx.polytechnique.fr

$$\begin{aligned} F_1(W) &= \left(\rho u \ , \ \rho u^2 + p \ , \ \rho uv \ , \ u(E+p)\right)^T \\ F_2(W) &= \left(\rho v \ , \ \rho uv \ , \ \rho v^2 + p \ , \ v(E+p)\right)^T \end{aligned}$$

In the above expressions, ρ is the density, $\vec{V} = (u, v)^T$ is the velocity vector, E is the total energy per unit of volume and p is the pressure. The pressure is deduced from the other variables using the state equation for a perfect gas $p = (\gamma - 1)(E - \frac{1}{2}\rho \parallel \vec{V} \parallel^2)$ where γ is the ratio of specific heats ($\gamma = 1.4$ for the air). Under the hypothesis that the solution is regular one can also write a nonconservative (or quasi-linear) equivalent form of Eq. (1) :

$$\frac{\partial W}{\partial t} + A_1(W)\frac{\partial W}{\partial x} + A_2(W)\frac{\partial W}{\partial y} = 0 \tag{2}$$

where the Jacobian matrices of the flux vectors $F_1(W)$ and $F_2(W)$ (see Dolean[Dol01] for more details). Suppose that we first proceed to an integration in time of (1) using a backward Euler implicit scheme involving a linearization of the flux functions. This operation results in the linearized system :

$$\mathcal{L}(U) \equiv \frac{\mathrm{Id}}{\Delta t}U + A_1\frac{\partial U}{\partial x} + A_2\frac{\partial U}{\partial y} = f \tag{3}$$

where $U \equiv W^{n+1} - W^n$ where $W^{n+1} = W(x, (n+1)\Delta t)$, and A_1 (respectively A_2) is a shorthand for $A_1(W^n)$ (respectively $A_2(W^n)$).
In the following we are interested in solving the problem (3), associated to a suitable set of boundary conditions, by a nonoverlapping additive Schwarz type algorithm. An algorithm based on transmission conditions at subdomain interfaces that consist in Dirichlet conditions for the characteristic variables corresponding to incoming waves (following a strategy already studied by Quarteroni and Stolcis[QS95]) has been considered in Dolean and Lanteri[DL99]. The main originality of this preliminary study is that in the discrete case the interface conditions are expressed in terms of upwind conservative normal fluxes computed using the approximate Riemann solver of Roe[Roe81]. This choice is before all motivated by the starting point of our study which was given by a flow solver based on a combined finite element/finite volume formulation on unstructured triangular meshes for the spatial discretization. Time integration of the resulting semi-discrete equations is obtained using a linearized backward Euler implicit scheme. As a result, each pseudo time step requires the solution of a sparse linear system for the flow variables, which is the discrete counterpart of (3).

The two-subdomain case

We consider the case of a two-subdomain decomposition with $\Omega_1 = \mathbb{R}_- \times \mathbb{R}$, $\Omega_2 = \mathbb{R}_+ \times \mathbb{R}$ separated by the interface $x = 0$; let $\vec{n} = (1 \ , \ 0)$ denote the normal vector at the interface $x = 0$, directed from Ω_1 to Ω_2. Let :

$$M_n = \frac{\vec{V}.\vec{n}}{c} = \frac{u}{c} \quad \text{and} \quad M_t = \frac{\vec{V}.\vec{t}}{c} = \frac{v}{c}$$

respectively denote the normal and the tangential Mach number at the interface $x = 0$. We also have that, at any point of $\Omega_1 \bigcup \Omega_2$, the Mach number can be expressed as $M =$

$\frac{\sqrt{u^2+v^2}}{c} = \sqrt{M_n^2 + M_t^2}$. Let $A_{\mathbf{n}} = n_x A_1 + n_y A_2$ for any vector $\vec{n} = (n_x\ ,\ n_y)^T$. Then, it is well known (from the hyperbolic nature of the system of Euler equations) that the matrix $A_{\mathbf{n}}$ is diagonalizable with real eigenvalues :

$$\begin{array}{c} A_{\mathbf{n}} W = T_{\mathbf{n}}(W)\Lambda_{\mathbf{n}}(W)T_{\mathbf{n}}^{-1}(W) \\ \text{with} \quad \Lambda_{\mathbf{n}}(W) = \text{diag}\left(\vec{V}.\vec{n} - c\ ,\ \vec{V}.\vec{n}\ ,\ \vec{V}.\vec{n}\ ,\ \vec{V}.\vec{n} + c\right) \end{array}$$

Let $U_i^{(0)}$ denote the initial appoximation of the solution in subdomain Ω_i. A general formulation of an additive Schwarz type algorithm for computing $U_i^{(p+1)}$ from $U_i^{(p)}$ (where p defines the iteration of the Schwarz algorithm) writes as :

$$\begin{array}{rcl} \Omega_1 & : & \left\{ \begin{array}{lcl} \mathcal{L}(U_1^{(p+1)}) & = & f_1 \quad \text{for } x < 0 \\ \mathcal{B}_1(U_1^{(p+1)}) & = & \mathcal{B}_1(U_2^{(p)}) \quad \text{for } x = 0 \end{array} \right. \\ \Omega_2 & : & \left\{ \begin{array}{lcl} \mathcal{L}(U_2^{(p+1)}) & = & f_2 \quad \text{for } x > 0 \\ \mathcal{B}_2(U_2^{(p+1)}) & = & \mathcal{B}_2(U_2^{(p)}) \quad \text{for } x = 0 \end{array} \right. \end{array} \tag{4}$$

where the $\mathcal{B}_{1,2}$'s are interface operators. Natural (also qualified as " classical ") interface conditions resulting from the variational formulation of the initial and boundary value problem associated to system (1) are given by :

$$\mathcal{B}_1 = A_{\mathbf{n}}^- = T_{\mathbf{n}}\Lambda_{\mathbf{n}}^- T_{\mathbf{n}}^{-1} \quad \text{and} \quad \mathcal{B}_1 = A_{\mathbf{n}}^+ = T_{\mathbf{n}}\Lambda_{\mathbf{n}}^+ T_{\mathbf{n}}^{-1} \tag{5}$$

In the particular case $\vec{n} = (1\ ,\ 0)$ we have that $T(W) \equiv T_{\mathbf{n}}(W)$ with :

$$T(W) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ u - c & 0 & u & u + c \\ v & c\sqrt{2} & v & v \\ \dfrac{u^2+v^2}{2} - cu + \dfrac{c^2}{\gamma - 1} & vc\sqrt{2} & \dfrac{u^2+v^2}{2} & \dfrac{u^2+v^2}{2} + cu + \dfrac{c^2}{\gamma - 1} \end{pmatrix}$$

By considering the approach adopted by Kroner[Kro91], we can use the matrix $T(W)$ to obtain a symmetrized form of the system (3) :

$$\widetilde{\mathcal{L}}(\widetilde{U}) \equiv \frac{\text{Id}}{\Delta t}\widetilde{U} + \widetilde{A}_1 \frac{\partial \widetilde{U}}{\partial x} + \widetilde{A}_2 \frac{\partial \widetilde{U}}{\partial y} = \widetilde{f} \tag{6}$$

where $\widetilde{U} = T^{-1}U$ and :

$$\begin{array}{rcll} \widetilde{A}_1(W) & = & T^{-1}(W)A_1(W)T(W) & = \quad \text{diag}(u - c\ ,\ u\ ,\ u\ ,\ u + c) \\ \widetilde{A}_2(W) & = & T^{-1}(W)A_2(W)T(W) & \quad \text{is a symmetric matix} \end{array}$$

Smith factorization

The first step consists in applying a Laplace transform in the x direction (the Laplace variable is denoted by λ) and a Fourier transform in the y direction (the Fourier variable is denoted by k) to system (6). The transformed system writes $A(\lambda,k)\widehat{W} = \hat{f}$. The expression of the transformed matrix $A(\lambda,k)$ is given in Dolean[Dol01]. An important result of the Smith factorization theory[Gan66] is that the polynomial matrix $A(\lambda,k)$ can be factorized as :

$$A(\lambda,k) = E(\lambda,k)D_s(\lambda,k)F(\lambda,k)$$

where $D_s(\lambda,k)$ represents the Smith diagonal form of $A(\lambda,k)$; $E(\lambda,k)$ (respectively $F(\lambda,k)$) is a permutation matrix that operates on the lines (respectively the columns) of $A(\lambda,k)$. In the present case, we obtain :

$$D_s(\lambda,k) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \mathcal{G}(\lambda,k) & 0 \\ 0 & 0 & 0 & \mathcal{G}(\lambda,k)\mathcal{L}(\lambda,k) \end{pmatrix} \tag{7}$$

where :

$$\begin{cases} \mathcal{L}(\lambda,k) &= -(c^2-u^2)\lambda^2 + 2u(\beta+ikv)\lambda + c^2k^2 + (\beta+ikv)^2 \\ \mathcal{G}(\lambda,k) &= \lambda u + (\beta+ikv) \end{cases} \tag{8}$$

and :

$$F(\lambda,k) = \begin{pmatrix} \beta & \lambda & ik & 0 \\ 0 & F_{22} & F_{23} & (\gamma-1)\beta^2 \\ 0 & F_{32} & F_{33} & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{9}$$

Smith form of the Schwarz algorithm

Let $W = (w_1\ ,\ w_2\ ,\ w_3\ ,\ w_4)^T$ denote the vector of conservative variables and $\overline{W} = FW$ the corresponding vector of Smith variables. The equations within each subdomain can be rewritten as :

$$D_s\overline{W} = \bar{f} \ \Leftrightarrow \ \begin{cases} \beta w_1 + \lambda w_2 + ikw_3 = \bar{f}_1 \\ F_{22}w_2 + F_{23}w_3 + (\gamma-1)\beta^2 w_4 = \bar{f}_2 \\ \mathcal{G}\bar{w}_s \equiv \mathcal{G}(F_{32}w_2 + F_{33}w_3) = \bar{f}_3 \\ \mathcal{G}\mathcal{L}\bar{w}_2 = \bar{f}_4 \end{cases} \tag{10}$$

Because of the structure of the matrix D_s it is sufficient to work with two Smith variables, w_2 and w_s, the other ones being obtained from the relations (11). Let $(E_i^{(p)})(x) = (W_i^{(p)} -$

$W_i)(x) = (e_1^i\ ,\ e_2^i\ ,\ e_3^i\ ,\ e_4^i)^T$ be the error vector in the subdomain Ω_i after the iteration p of the Schwarz algorithm. Using the change of variables $\overline{E} = FE$, the Schwarz algorithm is given by :

$$\begin{array}{lll} \Omega_1 & : & \left\{ \begin{array}{l} \mathcal{G}\left((\bar{e}_s^1)^{(p+1)}\right) \ \text{ and } \ \mathcal{GL}\left((\bar{e}_2^1)^{(p+1)}\right) \quad \text{for } \ x<0 \\ \mathcal{B}(\overline{E}_1^{(p+1)})_j = \mathcal{B}(\overline{E}_2^{(p)})_j \quad \text{for } \ x=0 \ \text{ and } \ \lambda_j(A_1)<0 \end{array} \right. \\ \Omega_2 & : & \left\{ \begin{array}{l} \mathcal{G}\left((\bar{e}_s^2)^{(p+1)}\right) \ \text{ and } \ \mathcal{GL}\left((\bar{e}_2^2)^{(p+1)}\right) \quad \text{for } \ x>0 \\ \mathcal{B}(\overline{E}_2^{(p+1)})_j = \mathcal{B}(\overline{E}_1^{(p)})_j \quad \text{for } \ x=0 \ \text{ and } \ \lambda_j(A_1)>0 \end{array} \right. \end{array} \tag{11}$$

where $\mathcal{B} = \mathcal{B}(\lambda, k)$ is a 4×2 matrix corresponding to the last two columns of the 4×4 matrix $T^{-1}(W)F^{-1}(\lambda,k)$. From now, we assume that the flow in subsonic i.e. $M<1$. By taking into account the sign of the eigenvalues we obtain :

$$\begin{array}{lll} \Omega_1 & : & \left\{ \ b_{11}(\bar{e}_2^1)^{(p+1)} \ = \ b_{11}(\bar{e}_2^2)^{(p)} + b_{12}(\bar{e}_s^2)^{(p)} \right. \\ \Omega_2 & : & \left\{ \begin{array}{rcl} b_{21}(\bar{e}_2^2)^{(p+1)} + b_{22}(\bar{e}_s^2)^{(p+1)} & = & b_{21}(\bar{e}_2^1)^{(p)} \\ b_{31}(\bar{e}_2^2)^{(p+1)} + b_{32}(\bar{e}_s^2)^{(p+1)} & = & b_{31}(\bar{e}_2^1)^{(p)} \\ b_{41}(\bar{e}_2^2)^{(p+1)} + b_{42}(\bar{e}_s^2)^{(p+1)} & = & b_{41}(\bar{e}_2^1)^{(p)} \end{array} \right. \end{array} \tag{12}$$

On the other hand, the local solutions are explicitly given by :

$$\bar{e}_2^1 = \alpha_1 e^{\lambda_{\mathcal{L}_1} x}\ ,\ \bar{e}_2^2 = \alpha_2 e^{\lambda_{\mathcal{G}} x} + \alpha_3 e^{\lambda_{\mathcal{L}_2} x}\ ,\ \bar{e}_s^2 = \alpha_4 e^{\lambda_{\mathcal{G}} x} \tag{13}$$

where $\lambda_{\mathcal{G}}$ and $\lambda_{\mathcal{L}_{1,2}}$ are the eigenvalues of the Fourier symbols $\Lambda_{\mathcal{G}}$ and $\Lambda_{\mathcal{L}_{1,2}}$ that factorize the operators $\mathcal{G}$ and $\mathcal{L}$ i.e. $\mathcal{G} = \partial_x - \Lambda_{\mathcal{G}}$ and $\mathcal{L} = (\partial_x - \Lambda_{\mathcal{L}_1})(\partial_x - \Lambda_{\mathcal{L}_2})$.

Generalized interface conditions

Using the relation $(\bar{e}_s^2)^{(p+1)} = \dfrac{b_{21}}{b_{22}}\left((\bar{e}_2^1)^{(p)} - (\bar{e}_2^2)^{(p+1)}\right)$ we can rewrite the interface iterations (12) as:

$$\begin{array}{lll} \Omega_1 & : & \left\{ \ b_{11}b_{22}(\bar{e}_2^1)^{(p+1)} \ = \ (b_{11}b_{22} - b_{21}b_{12})(\bar{e}_2^2)^{(p)} + b_{21}b_{12}(\bar{e}_2^1)^{(p-1)} \right. \\ \Omega_2 & : & \left\{ \begin{array}{l} (b_{31}b_{22} - b_{21}b_{32})(\bar{e}_2^2)^{(p+1)} = (b_{31}b_{22} - b_{21}b_{32})(\bar{e}_2^1)^{(p)} \\ (b_{41}b_{22} - b_{21}b_{42})(\bar{e}_2^2)^{(p+1)} = (b_{41}b_{22} - b_{21}b_{42})(\bar{e}_2^1)^{(p)} \end{array} \right. \end{array} \tag{14}$$

In order to obtain a general form of the iterations we introduce the operators $\mathcal{B}_i = p_i(k)\partial_x^2 + q_i(k)\partial_x + r_i(k)\ i = 1,4$ and we consider the Schwarz algorithm :

$$
\begin{array}{lll}
\Omega_1 & : & \left\{ \begin{array}{lcl} \mathcal{L}(e_1^{(p+1)}) & = & 0 \quad \text{for} \quad x < 0 \\ \mathcal{B}_1(e_1^{(p+1)}) & = & (\mathcal{B}_1 + \mathcal{B}_2)(e_2^{(p)}) - \mathcal{B}_2(e_1^{(p-1)}) \quad \text{for} \quad x = 0 \end{array} \right. \\
\Omega_2 & : & \left\{ \begin{array}{lcl} \mathcal{L}(e_2^{(p+1)}) & = & 0 \quad \text{for} \quad x > 0 \\ \mathcal{B}_{3,4}(e_2^{(p+1)}) & = & \mathcal{B}_{3,4}(e_1^{(p)}) \quad \text{for} \quad x = 0 \end{array} \right.
\end{array}
\tag{15}
$$

where $p_i(k)$, $q_i(k)$, $r_i(k)$ are polynomials in ik. Then, our strategy consists in several steps (see Dolean[Dol01] for more details). First, we derive a new form of the interface conditions by generalizing the expressions of $p_i(k)$, $q_i(k)$, $r_i(k)$. Second, we construct the interface operator $\mathcal{B}$. Finally, we retrieve the interface conditions in physical variables by using the matricial relation $SW = \mathcal{B}F^{-1}W$. The interface conditions in physical variables are obtain from the the matrix S that generalizes the matrix T^{-1} :

$$
S(W) = \begin{pmatrix}
\frac{\sigma+1}{2} & -\frac{1}{2}\frac{(\sigma-1)(u+c)}{c-u} & 0 & -\frac{1}{2}\frac{u(\sigma-1)}{c-u} \\
-\frac{\sigma+\delta-1}{2} & \frac{(\sigma+\delta-1)(c+u)+2(c-u)}{2(c-u)} & 0 & \frac{(\sigma+\delta-1)u}{2(c-u)} \\
s_{31} & s_{32} & \alpha & s_{34} \\
s_{41} & s_{42} & 0 & s_{44}
\end{pmatrix}
$$

with :

$$
\left\{ \begin{array}{lcl}
s_{31} & = & -\frac{1}{2\sqrt{2}}\frac{-(\alpha_1-1)u + c(\delta+\sigma-1))(ikv+\beta)}{ikcu} \\
s_{32} & = & \frac{1}{2\sqrt{2}}\frac{(\alpha_1-1)u(c-u) + c(c+u)(\delta+\sigma-1))(ikv+\beta)}{ikcu(c-u)} \\
s_{34} & = & \frac{1}{2\sqrt{2}}\frac{(\alpha_1-1)(c-u) + c(\delta+\sigma-1))(ikv+\beta)}{ikcu(c-u)} \\
s_{41} & = & \frac{c(\delta+\sigma-\alpha_2) + (\alpha_2-1)u}{u} \\
s_{42} & = & -\frac{(u+c)(c(\delta+\sigma-\alpha_2) + (\alpha_2-1)u)}{u(c-u)} \\
s_{44} & = & -\frac{c(\delta+\sigma-\alpha_2) + \alpha_2 u - c}{c-u}
\end{array} \right.
$$

In Dolean[Dol01] simple interface conditions (without derivatives) are derived from the expression of the convergence rate associated to the iterations (15). These conditions are obtained by setting $\alpha_2 = 1$ and $\delta = 1 - \sigma$. This results in interface conditions that depend on the parameter σ only.

Table 1: Nonoverlapping additive Schwarz type algorithm
Classical interface conditions versus generalized interface conditions

M_n	OPT0	OPT1	M_∞	OPT0	OPT1
0.1 and $M_t = 0.0$	20	20	0.3 and $M_t = 0.0$	24	19
0.6 and $M_t = 0.0$	27	17	0.1 and $M_t = 0.1$	24	21
0.3 and $M_t = 0.2$	24	28	0.6 and $M_t = 0.4$	32	18
0.6 and $M_t = 0.7$	25	21	0.8 and $M_t = 0.5$	42	21

Numerical results

Space and time discretization methods

The spatial discretization method adopted here combines the following elements (see Dolean and Lanteri[DL99] for more details) : (1) a finite volume formulation on triangular meshes together with upwind schemes for the discretization of the convective fluxes; (2) an extension to second order accuracy that relies on the MUSCL (Monotonic Upstream Schemes for Conservation Laws) introduced by van Leer[Lee79] and extended to unstructured triangular meshes by Fezoui and Stoufflet[FS89]. Time integration of the resulting semi-discrete equations is obtained using a linearized backward Euler implicit scheme[FS89]. As a result, each pseudo time step requires the solution of a sparse linear system for the flow variables. In this study, a nonoverlapping domain decomposition algorithm is used for advancing the solution at each implicit time step.

Numerical results

We present here a set of preliminary results of numerical experiments that are concerned with the evaluation of the influence of the interface conditions on the convergence of the nonoverlapping additive Schwarz type algorithm of the form (4). The computational domain is given by the rectangle $[0\ ,\ 2] \times [0\ ,\ 1]$. The numerical investigation is limited to the resolution of the linear system resulting from the first implicit time step using a Courant number CFL=100. A slipping condition ($\vec{V}.\vec{n} = 0$) is applied on the lower ($y = 0$) and upper ($y = 1$) walls; an inflow (respectively outflow) condition is applied on the left $x = 0$ (respectively right $x = 10$) boundary. Table 1 summarizes the number of Schwarz iterations required to reduce the initial linear residual by a factor 10^{-10} for different values of the reference Mach number. The underlying triangular mesh is a regular one deduced from a finite difference grid containing 4000 nodes (200×20). In this table, OPT0 stands for the classical interface conditions while OPT1 corresponds to the algorithm based on the generalized interface conditions.

Conclusions

In this work we were interested in the acceleration of the convergence of a nonoverlapping additive Schwarz type algorithm by modifying the transmission conditions applied to the subdomain interfaces. We built generalized zero order interface conditions using Smith theory of diagonalizing polynomial matrices. The numerical experiments confirmed at least qualita-

tively the behaviour in accordance with the theory even if from the discrete point of view we couldn't reproduce identically the results obtained in the continuous case. The preliminary results are very encouraging as the lead to a very good convergence rate for certain Mach numbers.

References

[DL99]Victorita Dolean and Stephane Lanteri. A domain decomposition approach to finite volume solutions of the Euler equations on triangular meshes. Technical Report 3751, INRIA, oct 1999.

[DLN00]Victorita Dolean, Stephane Lanteri, and Frederic Nataf. Convergence analysis of a Schwarz type domain decomposition method for the solution of the Euler equations. Technical Report 3916, INRIA, aopr 2000.

[Dol01]Victorita Dolean. *Algorithmes par décomposition de domaine et accélération multigrille pour le calcul d'écoulements compressibles*. PhD thesis, Université de Nice-Sophia Antipolis, 2001.

[FS89]Loula Fezoui and Bruno Stoufflet. A class of implicit upwind schemes for Euler simulations with unstructured meshes. *J. of Comp. Phys.*, 84:174–206, 1989.

[Gan66]Felix R. Gantmacher. *Theorie des matrices*. Dunod, 1966.

[JNR98]Caroline Japhet, Frederic Nataf, and Francois-Xavier Roux. The Optimized Order 2 Method with a coarse grid preconditioner. application to convection-diffusion problems. In P. Bjorstad, M. Espedal, and D. Keyes, editors, *Ninth International Conference on Domain Decompositon Mehods in Science and Engineering*, pages 382–389. John Wiley & Sons, 1998.

[JNR01]Caroline Japhet, Frederic Nataf, and Francois Rogier. The optimized order 2 method. application to convection-diffusion problems. *Future Generation Computer Systems FUTURE*, 18, 2001.

[Kro91]D. Kroner. Absorbing boundary conditions for the linearized Euler equations in 2D. *Math. Comp.*, 57:153–167, 1991.

[Lee79]Bram Van Leer. Towards the ultimate conservative difference scheme V : a second-order sequel to Godunov's method. *J. of Comp. Phys.*, 32:361–370, 1979.

[QS95]A. Quarteroni and L. Stolcis. Homogeneous and heterogeneous domain decomposition for compressible fluid flows at high reynolds numbers. *Numerical Methods for Fluid Dynamics*, 5:113–128, 1995.

[Roe81]Philipp L. Roe. Approximate Riemann solvers, parameter vectors and difference schemes. *J. Comput. Phys.*, 43:357–372, 1981.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

48 Optimized Interface Conditions for Sedimentary Basin Modeling

I. Faille[1], E. Flauraud[1], F. Nataf[2], F. Schneider[1], F. Willien[1]

Why DDM for basin modeling ?

Basin modeling aims at reconstructing the time evolution of a sedimentary basin in order to make quantitative predictions of geological phenomena leading to oil accumulations. It accounts for porous medium compaction, heat transfer, hydrocarbon formation and migration. Recent evolutions of basin simulators have contributed to improve the treatment of geological discontinuities such as faults and salt domes. Faults divide the basin into blocks which slide between themselves. They may be a preferential path or in opposite a barrier for hydrocarbons migration. A salt is an impervious medium and becomes a trap for hydrocarbon. CERES is an advanced prototype of $2D$ sedimentary basin tool that can handle non-vertical faults and salt or mud tectonics (figure 1). Domain decomposition methods provide a way to solve the

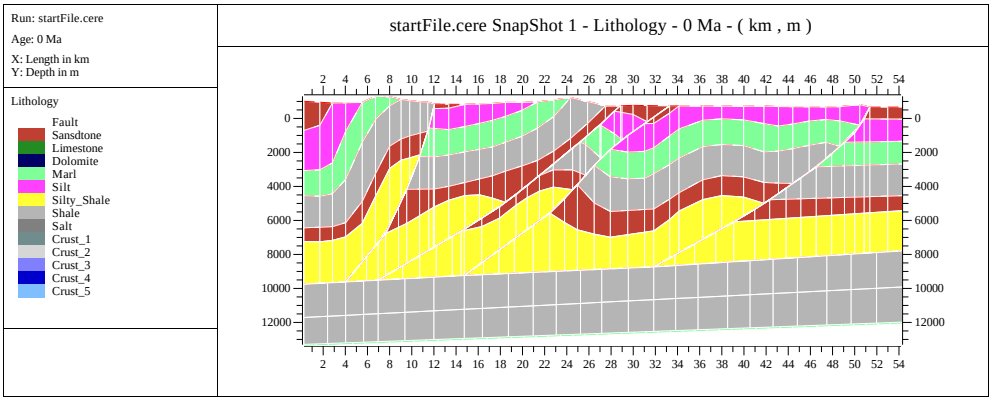


Figure 1: CERES 2D real basin

equations on the complex geometries considered, naturally defined as a set of adjacent sliding blocks and faults.

Following the work of [NR95], [NRdS94], [JNR01], we use nonoverlapping techniques and study several interface conditions, namely Robin type conditions.

The paper is organized as follows. First, the physical models and the governing equations taken into account are reviewed. Then the DDM are presented on a simpler equation in which we find the main characteristics of the problem. The optimized interface conditions are detailed for this equation. Finally, numerical results are shown.

[1]Institut Français du Pétrole, 1 et 4 avenue de Bois Préau, 92852 Rueil Malmaison cedex, FRANCE isabelle.faille@ifp.fr, eric.flauraud@ifp.fr, frederic.schneider@ifp.fr, francoise.willien@ifp.fr

[2]CNRS, UMR7641, CMAP, École Polytechnique, 91128 Palaiseau cedex, FRANCE,nataf@cmapx.polytechnique.fr;

Models and governing equations

In the blocks, the model accounts for the porous medium compaction, erosion, heat transfer, hydrocarbon formation and migration. The equations are mass conservation of solid and fluids (water,oil,gas) coupled with the Darcy's law and a compaction law. The faults have a constant porosity, but the permeability of the faults evolves in time. We consider only incompressible multiphase flows.
To present the DD method we consider a simplified basin model where geometry does not evolve in time. Using an IMPES (IMplicit Pressure, Explicit Saturation) scheme, we first solve a parabolic pressure equation and then update explicitly the phase saturations.
After time discretization, the pressure equation is then roughly written as follows :

$$\mathcal{L}(P) = \frac{\alpha}{\Delta t} P + div(-K\overrightarrow{grad}P) = f \tag{1}$$

where P is the pressure, α the compressibility of the porous medium and K the intrinsic permeability tensor of the porous medium divided by the fluid viscosity. The permeability depends heavily on the lithology under consideration. The contrast in the lithologies can induce a discontinuity of the permeability tensor of several orders of magnitude (up to six orders).
Moreover we have to deal with subdomains of various size, block width is about $10\ km$ while fault width is about $10\ m$.

The DDM for the pressure equation

Our goal is to find a domain decomposition method robust enough to deal with the strong discontinuities that can arise along and across an interface between two subdomains and whose behavior is not ruined by small subdomains. To cope with these difficulties, we introduce a Robin type interface condition whose coefficients are computed in order to optimize the convergence rate of the Additive Schwarz method (ASM for short).
We consider the parabolic linear equation (1) with strongly discontinuous coefficients K. We cut the domain into nonoverlapping subdomains Ω_i and solve the equation in each subdomain. In the framework of this paper, we weigh up only matching grid but the approach is extended to non-matching grid as see on figure 1.
Pressure and flux continuity between two subdomains Ω_1 and Ω_2 are expressed as Robin conditions on the interface Γ:

$$\alpha_2 P_1 + \beta_2 (K\overrightarrow{grad}P)_1.\vec{n_1} = \alpha_2 P_2 - \beta_2 (K\overrightarrow{grad}P)_2.\vec{n_2} \quad \text{on } \Gamma \tag{2}$$

$$\alpha_1 P_2 + \beta_1 (K\overrightarrow{grad}P)_2.\vec{n_2} = \alpha_1 P_1 - \beta_1 (K\overrightarrow{grad}P)_1.\vec{n_1} \quad \text{on } \Gamma \tag{3}$$

where α_i, β_i are real such that $\alpha_1\beta_2 + \alpha_2\beta_1 \neq 0$ and $\alpha_i\beta_i > 0$.
The idea is to find the coefficients (α_i, β_i) which allow a fast convergence of DD algorithm, namely ASM with the boundary condition (2) in Ω_1 and boundary condition (3) in Ω_2. This has been introduced by Nataf and co-author for convection-diffusion equation [NRdS94] [NR95].

To compute the Robin coefficients, we successively address two main difficulties. First, we consider the case of two subdomains with a jump of permeability across the interface. Secondly, we deal with two subdomains separated by a small fault.

Robin conditions for two unbounded subdomains

We consider two unbounded subdomains Ω_1, Ω_2 with an interface Γ. The subdomains have homogeneous permeability K_1 in subdomain Ω_1 and K_2 in Ω_2. The computation of Robin co-

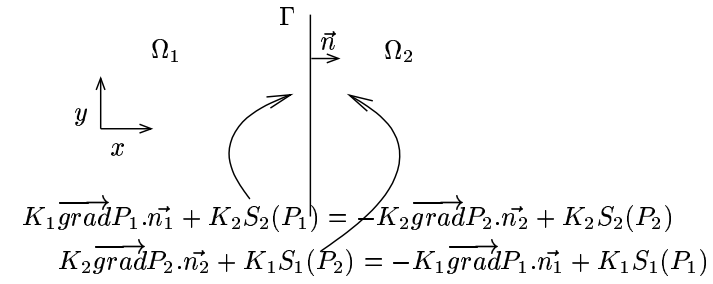


Figure 2: Optimal interface condition for 2 domains

efficients is based on the approximation of the Optimal Interface Condition. Optimal Interface Conditions are conditions which ensure convergence of ASM in 2 iterations for a decomposition into 2 subdomains. They extend the Artificial Boundary Condition and keep the idea of "packing the neighboring domain problem on the interface Γ". To do so, let us define the classical Steklov-Poincaré operator S_i associated to Ω_i :

$$S_i : P_{|\Gamma} \longrightarrow \overrightarrow{grad} P_i . \vec{n}_{i|\Gamma} \qquad \text{where } P_i \quad \begin{cases} \mathcal{L}_i(P_i) & = 0 \text{ in } \Omega_i \\ P_i & = P \text{ on } \Gamma \end{cases}$$

We can write the Optimal Interface Condition as follows, for Ω_1 :

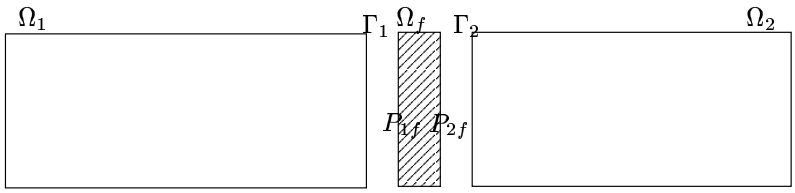
$$K_1 \overrightarrow{grad} P_1 . \vec{n_1} + K_2 S_2(P_1) = -K_2 \overrightarrow{grad} P_2 . \vec{n_2} + K_2 S_2(P_2)$$

For Ω_2, we have :

$$K_2 \overrightarrow{grad} P_2 . \vec{n_2} + K_1 S_1(P_2) = -K_1 \overrightarrow{grad} P_1 . \vec{n_1} + K_1 S_1(P_1)$$

The contribution of subdomain Ω_2 appears through S_2. The jump of permeability K is found in the two terms K_1, K_2.
To find the OIC, we need to make explicit the Steklov-Poincaré operator . Therefore we perform a Fourier transform with respect to y of equation (1) and we solve exactly the obtained equation which only depends on the x variable. The expression of Λ_i, symbol of S_i, are not polynomial in k, dual variable with respect to y. The Steklov-Poincaré operator and so the OIC are not local in space and must be approximated.
In order to obtain the Robin boundary condition, two constant approximations have been considered, giving the following coefficients:

- taking a Taylor expansion of Λ_i, we obtain: $(\alpha_i, \beta_i) = (K_i \omega_i, 1)$,
- we perform the minimization of the convergence rate on a frequency slot (k_{min}, k_{max}). This can be done only in the homogeneous case. We have the following expression:



$$K_f\overrightarrow{grad}P_{1f}.\vec{n_{1f}} + K_1S_1(P_{1f}) = -K_1\overrightarrow{grad}P_1.\vec{n_1} + K_1S_1(P_1)$$
$$K_f\overrightarrow{grad}P_{2f}.\vec{n_{2f}} + K_2S_2(P_{2f}) = -K_2\overrightarrow{grad}P_2.\vec{n_2} + K_2S_2(P_2)$$

Figure 3: OIC for the fault

$(\alpha_i\ ,\beta_i) = (K_i\sqrt{\sqrt{\omega_i^2 + k_{min}^2}\sqrt{\omega_i^2 + k_{max}^2}}\ , 1)$. This leads to the geometric average of the operator at the two extreme frequencies (k_{min}, k_{max}).

Although optimized in the homogeneous case, these conditions are efficient in the heterogeneous case. Indeed the coefficient α_i is a good approximation of S_i.
These last conditions have been implemented and give good results for subdomains with more or less similar size [WFS96]. However in the case of two blocks separated by a fault, the results need to be improved.

Robin conditions for two unbounded domains and one fault

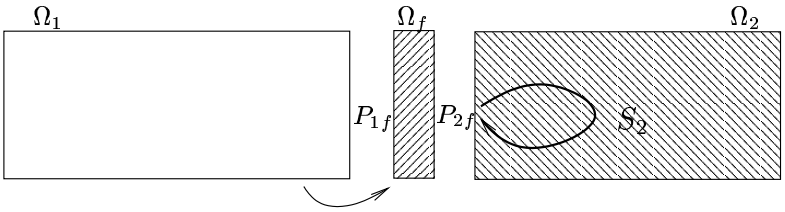
Now, we examine the case of three subdomains : we have two unbounded subdomains Ω_1, Ω_2 with a small fault Ω_f. The permeability of each subdomain is uniform : K_1 in Ω_1, K_2 in Ω_2 and K_f in Ω_f. We first set up the interface condition on the fault boundaries (see figure 3). As for a given boundary, the fault has only one neighbor, so we can define the Optimal Interface Condition as previously, where S_i is the Steklov-Poincaré operator associated to domain Ω_i.
Next we set up the OIC on the blocks boundaries. Assume, we are on the boundary of Ω_1 to find the OIC (see figure 4). The OIC uses S_{f+2} the Steklov-Poincaré operator associated to $\Omega - \Omega_1$:

$S_{f+2} : P_{1f} \longrightarrow \overrightarrow{grad}P.\vec{n_f}(P_{1f})$, where P is solution to

$$\begin{cases} \mathcal{L}_f(P) = 0 \text{ in } \Omega_f \quad ; \quad \mathcal{L}_2(P) = 0 \text{ in } \Omega_2 \\ P = P_{1f} \text{ on } \Gamma_1 \quad ; \quad P \text{ and } K\overrightarrow{grad}P.\vec{n} \text{ continuous on } \Gamma_2 \end{cases}$$

Although more complex, the Steklov-Poincaré operator is determined as previously performing a Fourier transform with respect to y. This operator S_{f+2} depends on the permeabilities K_2 in domain Ω_2, K_f in the fault and the fault width. As before the operators S_i, S_{i+f} are non-local in space, so we compute a polynomial approximation:

- we take the OIC for a frequency k_0,
- it is difficult to optimize the convergence rate, so we keep the idea to approximate the operator by the geometric average of two extreme values:



$$K_1\overrightarrow{grad}P_1.\vec{n_1} + K_f S_{f+2}(P_1) = \quad -K_f\overrightarrow{grad}P_f.\vec{n_f} + K_f S_{f+2}(P_f)$$

Figure 4: OIC for the subdomain Ω_1

Fault boundary condition:$(\alpha_{fi}\ ,\beta_{fi}) = (K_i\sqrt{\Lambda_i(k_{min})\Lambda_i(k_{max})}\ , 1)$;
Block boundary condition: $(\alpha_i\ ,\beta_i) = (K_f\sqrt{\Lambda_{f+i}(k_{min})\Lambda_{f+i}(k_{max})}\ , 1)$.

In the following, we denote by OIC, Robin conditions with these last (α_i, β_i) coefficients.

Numerical results

The approach is then extended to a real basin model (CERES 2D) which accounts for porous medium sedimentation, compaction, erosion and blocks displacements along faults. In CERES 2D, we have a discontinuity jump of the permeability K along the interface, so the Robin coefficient α_i is computed locally on each edge. The interface problem is solved with the GMRES algorithm. The unknowns are $H_1 = \alpha_2 P_1 + \beta_2 F_1$, $H_2 = \alpha_1 P_1 + \beta_1 F_2$, .

$$\mathcal{H}_{12} \ : \ H_1 \rightarrow \alpha_1 P_1 + \beta_1 F_1 \quad \text{with } F_1 = (-K_1\overrightarrow{grad}P_1.\vec{n_1}) \quad \text{and} \begin{cases} \mathcal{L}_1(P_1) & = f \text{ in } \Omega_1 \\ BC & \text{on } \partial\Omega_1/\Gamma \\ \alpha_2 P_1 + \beta_2 & (K_1\overrightarrow{grad}P_1.\vec{n_1}) = H_1 \text{ on } \Gamma \end{cases}$$

The equations are
Pressure and Flux continuity
$$\begin{cases} \mathcal{H}_{12}(H_1) - H_2 & = 0 \\ H_1 - \mathcal{H}_{21}(H_2) & = 0 \end{cases}$$

Numerical results show the good behavior of the Robin interface conditions. Comparisons with the Dirichlet-Neumann conditions illustrate the robustness and the good convergence rate of DD algorithms such as additive Schwarz method, possibly accelerated by GMRES.

Mesh refinement

We consider a synthetic basin (figure 5) composed of two heterogeneous blocks with K_1, K_2 permeability and a fault with K_f permeability. We want to study the influence of vertical mesh subdivision: each row is subdivided in 2 then 4, so the number of interface unknowns is growing. Moreover, the permeability of the fault is either: pervious, impervious or an average of the two neighbouring block cells (variable for short).
In Figure 6, we report the number of DDM iterations as a function of the number of interface unknowns for different fault permeability. The dotted lines show the Dirichlet-Neumann

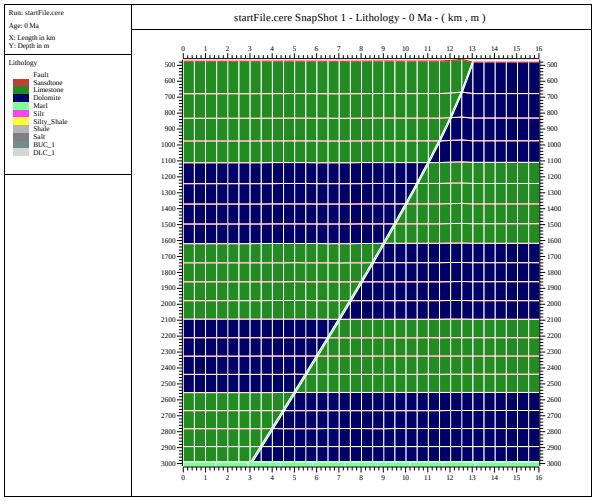


Figure 5: Synthetic basin

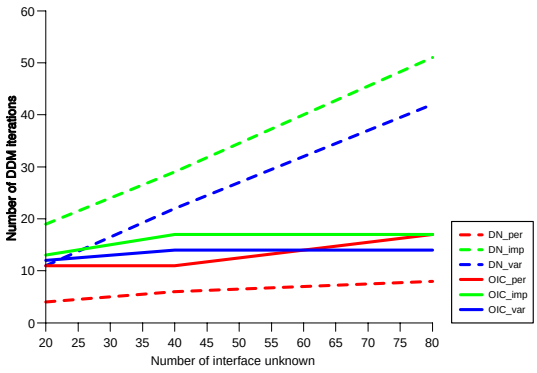


Figure 6: Number of DDM iterations

condition, Dirichlet on blocks, Neumann on fault,(DN for short) and the solid lines show the Robin condition. The Dirichlet-Neumann interface condition is very competitive for pervious fault but not for the other cases. The Robin conditions converge well; all curves correspond to 10 iterations of DDM. The number of DDM iterations does not increase too much with the number of interface unknowns for OIC. The behavior of OIC is regular for all spreading of fault permeability.

Time evolution

We can study the time evolution of DDM since the number of unknowns increases through time as new layers of sediments are deposited. Each layer corresponds to a row of homogeneous cells. On the following figure 7, a synthetic basin is composed of two heterogeneous

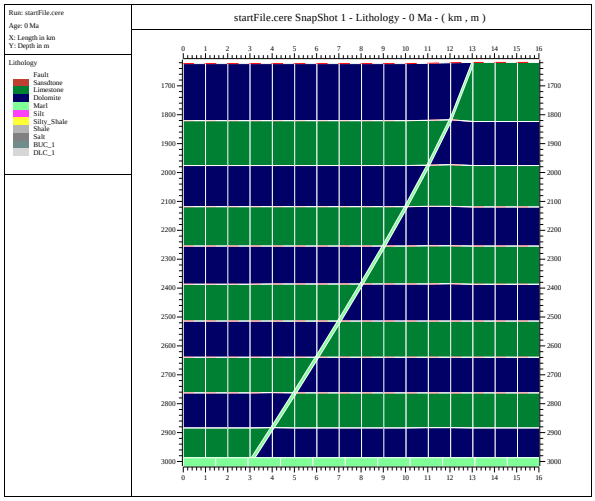


Figure 7: Synthetic basin

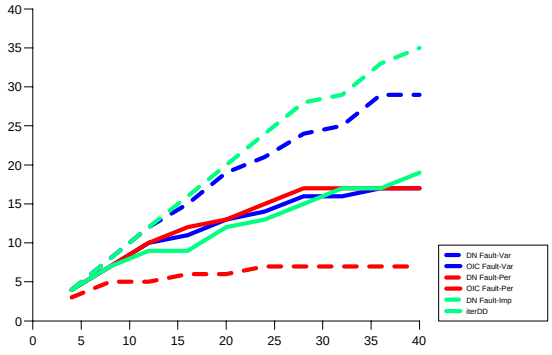


Figure 8: Number of DDM iterations

blocks and one fault. Each block has alternated and shift row of pervious-impervious medium. The fault permeability is successively: pervious, impervious or variable.
In Figure 8, we report the number of DDM iterations as a function of the number unknowns. The dotted lines show DN condition and the solid lines show OIC. The number of DDM iterations grows slowly with the number of unknowns for the OIC. Here again, OIC show robustness regarding subdomain heterogeneities.

Conclusion

We introduced a domain decomposition method applied to sedimentary basin modeling. The DDM is robust enough to overcome high jump of heterogeneity, up to 6 orders and various sizes of subdomains. To do this, we have chosen a nonoverlapping ASM with Optimal In-

terface condition, Robin type. The interface problem is solved with a GMRES algorithm. Despite the good results obtained, the fault is still a very small subdomain compared to the blocks. Therefore it seems promising to consider one dimensional fault [Fla01]. The fault model is then included in the interface condition between two blocks. We wish to improve the non-matching approach so as to win in flexibility and to have less interface unknowns. Another improvement is to extend the DDM to a "fully implicit" discretization scheme for multiphase flow. DDM will be applied to a system of pressure and saturation variables.

References

[Fla01]Eric Flauraud. *Méthode de décomposition de domaine pour des milieux poreux faillés*. PhD thesis, Paris VI, 2001. in preparation.

[JNR01]Caroline Japhet, Frederic Nataf, and Francois Rogier. The optimized order 2 method. application to convection-diffusion problems. *Future Generation Computer Systems FUTURE*, 18, 2001.

[NR95]Frédéric Nataf and Francois Rogier. Factorization of the convection-diffusion operator and the Schwarz algorithm. M^3AS, 5(1):67–93, 1995.

[NRdS94]Frédéric Nataf, Francois Rogier, and Eric de Sturler. Optimal interface conditions for domain decomposition methods. Technical report, CMAP (Ecole Polytechnique), 1994.

[WFS96]Françoise Willien, Isabelle Faille, and Frédéric Schneider. Domain decomposition methods for fluid flow in porous medium. In Petter Bjrstad, Magne Espedal, and David Keyes, editors, *Proceedings of the ninth domain decomposition conference*, pages 736–744, Bergen, Norway, 1996. Domain Decomposition Press.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

49 Domain decomposition methods in semiconductor device modeling

L. Giraud[1], J. Koster[2], A. Marrocco[3], J.-C. Rioual[4]

Introduction

In this paper, we present some parallel implementations of domain decomposition techniques for the solution of the drift diffusion equations involved in 2D semiconductor device modeling. The model describes the stationary state of a device when biases are applied to its bounds. The mixed dual formulation is retained. Therefore, we have to deal with a system of six totally coupled nonlinear partial differential equations. This system is decoupled and discretized in time by a semi-implicit nonlinear scheme using local time stepping. At each time step, we have to solve three systems of two nonlinear partial differential equations. The first system is associated with electrostatic potential, the second with the negative charges (electrons) and the third with the positive charges (holes). Each pair of equations is naturally discretized in space by mixed finite elements defined on 2D unstructured meshes and then solved by a Newton-Raphson method [HM94]. At each step of the Newton-Raphson method, a linear system of equations has to be solved. Depending upon which nonlinear system is being solved, these linear systems can be either symmetric positive definite or unsymmetric. These systems are sparse with a maximum of five nonzero entries per row due to the mixed finite element triangulation. A complete simulation is decomposed into two phases: first the solution of the equilibrium problem, then the solution of the static problem. The equilibrium problem consists of applying a zero potential to the bounds of the device and its numerical solution only involves the solution of symmetric positive definite linear systems. In this paper, we consider only the solution of the equilibrium problem.
Our objective is to obtain a fully parallel code in a distributed memory environment with MPI as message-passing library.
The formulation is mainly vectorial and is naturally parallelizable. The main difficulty consists of the efficient implementation of suitable linear solvers. This numerical kernel is the most time consuming part of the code. In this respect, we investigate substructuring approaches using either direct or iterative linear solvers.

Substructuring techniques

We assume that the domain Ω with boundary $\partial\Omega$ is partitioned into N non-overlapping subdomains $\Omega_1, \ldots, \Omega_N$ with boundaries $\partial\Omega_1, \ldots, \partial\Omega_N$. After discretization, we obtain a linear system $Au = f$, where the matrix A is sparse, unstructured, and symmetric positive definite. Let B be the set of all the indices of the discretized points which belong to the interfaces between the subdomains. Grouping the points corresponding to B in the vector u_B and the

[1]CERFACS, 42 Av. G. Coriolis, 31057 Toulouse, France. giraud@cerfacs.fr
[2]Parallab, University of Bergen, N-5020 Bergen, Norway. jak@ii.uib.no
[3]INRIA, Rocquencourt, 78153 Le Chesnay Cedex, France. Americo.Marrocco@inria.fr
[4]CERFACS, 42 Av. G. Coriolis, 31057 Toulouse, France. rioual@cerfacs.fr

points corresponding to the interior I of the subdomains in u_I, we get the reordered problem

$$\begin{pmatrix} A_{II} & A_{IB} \\ A_{IB}^T & A_{BB} \end{pmatrix} \cdot \begin{pmatrix} u_I \\ u_B \end{pmatrix} = \begin{pmatrix} f_I \\ f_B \end{pmatrix} . \tag{1}$$

The matrix A_{II} can be reordered to a block diagonal matrix in which the i-th diagonal block A_{ii} corresponds to the internal variables of subdomain Ω_i. Eliminating u_I from the second block row of Equation (1) leads to the reduced problem

$$S u_B = f_B - A_{IB}^T A_{II}^{-1} f_I , \quad \text{where} \quad S = A_{BB} - A_{IB}^T A_{II}^{-1} A_{IB} \tag{2}$$

is the Schur complement matrix of the matrix A_{II} in A. The matrix S is symmetric positive definite,

Let Γ be the interface between the subdomains. The local interface of a subdomain can be defined as $\Gamma_i = \partial\Omega_i \backslash \partial\Omega$. Let $R_i : \Gamma \to \Gamma_i$ be the canonical pointwise restriction which maps vectors defined on Γ onto vectors defined on Γ_i, and let $R_i^T : \Gamma_i \to \Gamma$ be its transpose. For a stiffness matrix A arising from a finite element discretization, the Schur complement matrix can be written as the sum of N (local) smaller Schur complement matrices

$$S = \sum_{i=1}^{N} R_i^T S^{(i)} R_i, \quad \text{where} \quad S^{(i)} = A_{\Gamma_i}^{(i)} - A_{i\Gamma_i}^T A_{ii}^{-1} A_{i\Gamma_i} , \tag{3}$$

where the local Schur complement matrix $S^{(i)}$ associated with subdomain Ω_i is computed from the subdomain stiffness matrix $A^{(i)}$, defined by

$$A^{(i)} = \begin{pmatrix} A_{ii} & A_{i\Gamma_i} \\ A_{i\Gamma_i}^T & A_{\Gamma_i}^{(i)} \end{pmatrix} . \tag{4}$$

In a parallel (multi-processor) computing environment, each subdomain matrix $A^{(i)}$ is assigned to one processor. In this way, the local Schur complement matrices $S^{(i)}$ can be computed simultaneously. The (global) Schur complement matrix S is available through Equation (3) and is never assembled explicitly.

Implementation

The implementation of substructuring methods usually consists of three steps. They are summarized in Algorithm 1.

Algorithm 1 : Substructuring algorithm	
Step 1 :	Factorize the matrix A_{II} and compute the Schur complement matrix S.
Step 2 :	Solve the Schur complement system $S u_B = f_B - A_{IB}^T A_{II}^{-1} f_I$, for u_B.
Step 3 :	Solve the system $A_{II} u_I = f_I - A_{IB} u_B$, for u_I.

Since the matrix A_{II} is a block diagonal matrix, steps 1 and 3 each consist of N independent smaller steps (one for each subdomain). Furthermore, since each subdomain matrix $A^{(i)}$ is assigned to a different processor, these N steps can be performed in parallel. When the number of subdomains increases, the amount of parallelism naturally increases in steps 1 and 3 (but the total amount of work decreases). The solution of the Schur complement problem in step 2 becomes more complex when the number of subdomains gets larger; the Schur complement matrix S is not assembled and therefore efficient parallel direct or iterative schemes are required for step 2.

Direct substructuring

The three steps of Algorithm 1 are implemented as follows.

Computation of the local Schur complement matrices

Since the discretization of the local problem is a sparse matrix, a direct (factorization) approach can be used in step 1 to compute the local Schur complement matrices $S^{(i)}$ explicitly. In fact, computing $S^{(i)}$ for subdomain Ω_i is equivalent to a partial Cholesky factorization of the matrix $A^{(i)}$ of Equation (4). In our implementation, we use the parallel multifrontal sparse direct solver MUMPS [ADLK01] that can compute a Cholesky factorization of the sparse submatrix A_{ii} and return the Schur complement matrix $S^{(i)}$. As mentioned before, the local Schur complement matrices are computed simultaneously (one on each processor). Therefore, we use one instance of the MUMPS solver on each processor.

Solution of the interface problem

The global Schur complement matrix S is the sum of the local Schur complements matrices $S^{(i)}$, see Equation (3). After computing the matrices $S^{(i)}$ explicitly, the matrix S is available as a sparse matrix that is distributed over the processors. MUMPS can accept such matrices in distributed form and can therefore be used to solve the interface problem directly and in parallel to obtain vector u_B. For this step, we use only one instance of MUMPS over all processors.

Solution of the interior problems

Step 3 consists of the solution of N independent systems $A_{ii}u_i = f_i - A_{i\Gamma_i}R_i u_B$ where $R_i u_B$ is the restriction of the interface solution vector u_B to interface Γ_i and f_i is the restriction of f_I to the internal variables of subdomain Ω_I. The systems can be solved by using the factorizations of the matrices A_{ii} that were computed in step 1.

Iterative substructuring

Iterative substructuring differs from direct substructuring in the solution of the interface problem (step 2 of Algorithm 1). Instead of a parallel Cholesky factorization, preconditioned conjugate gradient (PCG) iterations are used. Two important components of the preconditioned iteration are the matrix-vector product and the construction (and use) of the preconditioner.

Matrix-vector product

As the local Schur complements have been computed explicitly in the first step of the algorithm, it is straightforward to implement the matrix-vector product with the global Schur complement system.

For 2D problems, this implementation is more efficient than a classical implicit formulation where the local Schur complements are not known explicitly and where step 1 of Algorithm 2

Algorithm 2 : Computation of $y \rightarrow Sx$

Step 1 : $y_i = S^{(i)} R_i x$
Step 2 : $y = \sum_{i=1}^{N} R_i^T y_i$

requires a forward and back substitution and three sparse matrix-vector products (see Equation (3)). We display in Table 1 a comparison between the explicit and the implicit version of the Schur matrix-vector products for subdomains with 16000 unknowns each. For this example, the computation of the explicit local Schur complement is twice as expensive as the factorization of the local internal problem, but the subsequent gain for each Krylov iteration is very important (without preconditioning, the matrix-vector product remains the most expensive part of a Krylov iteration). For 2D problems, the extra storage cost for the local Schur complement is not prohibitive.

	implicit	explicit
Numerical factorization	14.2	27.2
Matrix-vector product	2.2	0.1
Time for factorization + 20 products	58.2	29.2

Table 1: Time comparison between implicit and explicit Schur matrix-vector product for subdomains of 16000 unknowns and with interfaces of size 1600. The computation has been performed on an SGI Origin 2000.

Preconditioners for the Schur complement

The rate of convergence of the conjugate gradient method depends on the condition number of the matrix of the linear system. In this section, we present two preconditioners for the Schur complement system.

Balanced Neumann-Neumann preconditioner

This two-level preconditioner was first introduced in [Man93]. It can be formulated as

$$M^{-1}S = P + (I - P)M_1 S(I - P), \quad \text{and} \quad M_1 = \sum_{i=1}^{N} R_i^T D_i^T (S^{(i)})^{-1} D_i R_i.$$

Here, M_1 is the one-level Neumann-Neumann preconditioner that was originally proposed in [DRLT91]. The weight matrices D_i form a decomposition of unity. P denotes the S-orthogonal projection onto the coarse space defined by $\{\sum_{i=1}^{N} R_i^T D_i^T Z_i v_i\}$ for some local subspaces Z_i and arbitrary vectors v_i.
In the experiments for this paper, we used a software package developed by Parallab (University of Bergen). One of the options that is available in that package is to construct the coarse space from local subspaces Z_i that are spanned by the eigenvectors associated with selected (small) eigenvalues of the local Schur complement matrices $S^{(i)}$. In particular, each subspace Z_i must contain the null space of $S^{(i)}$. We refer to [BKK00] for more details on this preconditioner.

Assembled Schur preconditioner

We consider another preconditioning technique for the solution of the Schur complement system. Let $\overline{S^{(i)}}$ be the local assembled Schur complement associated with a subdomain Ω_i. The assembled Schur preconditioner [CGM01] is defined by

$$\sum_{i=1}^{n} R_i^T (\overline{S^{(i)}})^{-1} R_i.$$

$\overline{S^{(i)}}$ corresponds to the restriction of the global Schur matrix to the boundary $\partial\Omega_i$ of the subdomain Ω_i and can be algebraically written as $\overline{S^{(i)}} = R_i S R_i^T$. A second level can be added to this preconditioner to define

$$M^{-1} = \sum_{i=1}^{n} R_i^T (\overline{S^{(i)}})^{-1} R_i + R_0^T (R_0 S R_0^T)^{-1} R_0,$$

where R_0 is a restriction operator from the global interface Γ onto a coarse space V_0. The second term of the equation defines a second level for the preconditioner. In our experiments, we consider a coarse space where we have one degree of freedom per edge of the decomposition, that is one degree of freedom per interface between two subdomains [CGT01].

Numerical experiments

We have performed experiments on various meshes, but here we only report results on a problem with 154892 unknowns. The size of the interface Γ varies from 440 unknowns for the 4-subdomain decomposition to 2446 unknowns for the 32-subdomain decomposition. The size of this problem is modest but it is representative for many problems that are solved in device modeling. Indeed, due to the complexity of the underlying physical problem, the complete code is memory consuming. We used a 32-processor SGI Origin 2000 for our experiments.

Iterative substructuring methods

In this section, we use the following notation:

1. AS : Assembled Schur preconditioner without coarse space.
2. NN : Neumann-Neumann preconditioner without coarse space.
3. AS-edge : Assembled Schur preconditioner with a coarse space defined by one degree of freedom per interface of the decomposition.
4. BNN1 : Balanced Neumann-Neumann preconditioner with a coarse space with one degree of freedom per subdomain.
5. BNN3 : Balanced Neumann-Neumann preconditioner with a coarse space with three degrees of freedom per subdomain.

For the sake of comparison between direct and iterative substructuring approaches, we stop the preconditioned conjugate gradient iterations only when the 2-norm of the residual of the current iterate normalized by the 2-norm of the right hand side is less than 10^{-15}. Note that such accuracy is often obtained by direct methods, without iterative refinement.
The condition number of the preconditioned Schur complement matrix varies from a few tens to a few hundreds. It is obtained from the eigenvalues of the tridiagonal matrix of the PCG coefficients although this may not be very reliable when the number of iterations is small.
We display in Table 2 the average number of conjugate gradient iterations that are needed to solve a linear problem during nonlinear solution of the equilibrium problem. The table shows

	Preconditioner					
No. of subdomains	none	AS	NN	AS-edge	BNN1	BNN3
4	67	4	15	6	15	15
8	87	20	22	23	21	20
16	117	31	29	33	27	25
32	125	37	33	40	32	29

Table 2: Average number of conjugate gradient iterations per linear system. 'none' denotes that no preconditioner is used. 19 linear systems were solved during the simulation.

that all the preconditioners improve the convergence of the conjugate gradient iteration. We observe that one-level preconditioner AS is more efficient than NN on a small number of subdomains. NN becomes more efficient when the number of subdomains increases. For all methods, the number of iterations of the preconditioned conjugate gradient increases moderately when the number of subdomains increases. The use of two-level preconditioners does not eliminate this increase. The AS-edge preconditioner performs worse than the AS preconditioner on this example.

Time measurements

In this section, we compare iterative substructuring with direct substructuring from an elapsed time point of view. We also make a comparison with MUMPS used as a black-box parallel sparse direct solver on the complete original problem, where the stiffness matrices are considered as in a distributed format (a feature of the MUMPS software).
In Table 3, we display the elapsed times to solve the complete equilibrium problem, by using direct substructuring (DSS), MUMPS as a black box direct parallel solver (MS) or iterative substructuring with preconditioned conjugate gradient (using AS, NN or BNN3),

	Algorithm				
no. of subdomains	AS	NN	BNN3	DSS	MS
4	49.65	49.50	53.01	57.47	63.97
8	26.11	31.04	28.53	31.59	44.24
16	14.79	16.74	17.89	22.81	43.66
32	10.65	11.87	16.31	16.98	44.07

Table 3: Times (in s) for solving the equilibrium problem (19 linear systems).

Table 3 shows that the substructuring algorithms are more efficient than the multifrontal parallel solver that does not scale very well because of the relative modest size of the linear

systems. We also observe that iterative substructuring is slightly more efficient than direct substructuring. Iterative substructuring will perform even better when we use a relaxed accuracy of the iterative linear solvers, while maintaining the same nonlinear path. However, this behaviour might not be true for other equations and to make a fair comparison, we required the same accuracy for all the linear solvers.

Table 3 shows that the one-level algorithms (AS and NN) perform better than the two-level algorithms, even though the latter require slightly fewer iterations for larger number of subdomains. This shows that a smaller number of iterations is not always an indication for better overall efficiency of a preconditioned Krylov solver.

Conclusion

We compared some efficient parallel linear solvers based on direct or iterative substructuring methods that enable us to solve problems that cannot be tackled on a single processor computer. For a set of linear systems that arise in semiconductor device modeling, we have shown that the two-level Balanced Neumann-Neumann preconditioner converges faster, but is less efficient in elapsed time, than the one-level preconditioners (Neumann-Neumann and Assembled Schur). The overall efficiency of the iterative substructuring approaches for these 2D simulations is due to the multifrontal direct sparse solver MUMPS that is for example able to compute the local Schur complement matrices efficiently.

Finally, we mention that domain decomposition techniques to parallelize the solution of the unsymmetric systems arising from the solution of the continuity equations associated with the electrons and the holes concentrations in the domain is still an ongoing work.

References

[ADLK01]Patrick R. Amestoy, Iain S. Duff, Jean-Yves L'Excellent, and Jacko Koster. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Analysis and Applications*, 23(1):15–41, 2001.

[BKK00]Petter E. Bjørstad, Jacko Koster, and Piotr Krzyżanowski. Domain decomposition solvers for large scale industrial finite element problems. In *PARA2000 Workshop on Applied Parallel Computing*. Lecture Notes in Computer Science 1947, Springer-Verlag, 2000.

[CGM01]Luiz M. Carvalho, Luc Giraud, and Gérard Meurant. Local preconditioners for two-level non-overlapping domain decomposition methods. *Numerical Linear Algebra with Applications*, 8(4):207–227, 2001.

[CGT01]Luiz M. Carvalho, Luc Giraud, and Patrick Le Tallec. Algebraic two-level preconditioners for the schur complement method. *SIAM J. Scientific Computing*, 22(6):1987–2005, 2001.

[DRLT91]Yann-Hervé De Roeck and Patrick Le Tallec. Analysis and test of a local domain decomposition preconditioner. In Roland Glowinski, Yuri Kuznetsov, Gérard Meurant, Jacques Périaux, and Olof Widlund, editors, *Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 112–128. SIAM, Philadelphia, PA, 1991.

[HM94]Frédéric Hecht and Americo Marrocco. Mixed finite element simulation of hetero-

junction structures including a boundary layer model for the quasi-fermi levels. *COMPEL*, 13(4):757–770, 1994.

[Man93]Jan Mandel. Balancing domain decomposition. *Comm. Numer. Meth. Engrg.*, 9:233–241, 1993.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

50 3D Structural Optimization in Electromagnetics

R.H.W. Hoppe[1], S. Petrova[2], V. Schulz[3]

Introduction

We consider the optimal design and layout of high power electronic devices that are based on the pulse width modulation technique such as DC-AC converter modules used in applications as electric drives for high power electromotors. The design objective is to minimize power losses caused by eddy currents that build up in the device due to fast switching times and steep current ramps (cf., e.g., [BFS99, BHM01, DGH98]).
In mathematical terms this leads to a topology optimization problem with the electric conductivity of the material as the design parameter and the electric and the magnetic field as the state variables that are supposed to satisfy the quasistationary limit of Maxwell's equations. With the optimal design of mechanical structures described by continuum mechanical models being by now a well established discipline (cf., e.g., [BEN95] and the references therein), not much work has been done with regard to the optimization of systems whose operational behavior is governed by Maxwell's equations. Moreover, the use of modern discretization and numerical solution techniques such as multigrid and domain decomposition methods for optimization problems with PDE constraints is still in its infancy (cf., e.g., [HEI00, HPS01, MAS00]).
In this paper, we focus on an approach relying on a primal-dual Newton interior-point method for the discretized optimization problem where the discretization of the eddy currents equations is taken care of by curl-conforming edge elements. Domain decomposition methods on nonmatching grids can be used for the numerical solution of the discretized field equations which is an integral part of the optimization routine featuring logarithmic barrier functions and simultaneous sequential quadratic programming.

The topology optimization problem

We consider a DC-AC converter module consisting of specific semiconductor devices such as IGBTs (Insulated Gate Bipolar Transistors) and GTOs (Gate Turn-Off Thyristors) that are interconnected and linked to the high power source as well as the load by copper made bus bars (cf. Figure 1).
Each bus bar contains a certain number of ports where currents are either supplied to or taken off the bar. The IGBTs and GTOs serve as valves for the currents which can be in the range of several kA. During operation of the module, electromagnetic fields $\mathbf{E}$ and $\mathbf{H}$ are generated that can be described by the eddy currents equations

$$\frac{\partial \mathbf{B}}{\partial t} + \mathbf{curl}\,\mathbf{E} = \mathbf{0}\ ,\ \operatorname{div}\mathbf{B} = 0\ ,\ \mathbf{curl}\,\mathbf{H} = \mathbf{J}\ , \tag{1}$$

[1]Institute of Mathematics , University of Augsburg, D-86159 Augsburg, Germany
[2]Institute of Mathematics , University of Augsburg, D-86159 Augsburg, Germany ; on leave from the Bulgarian Academy of Sciences, Sofia, Bulgaria
[3]Weierstrass Institute Berlin, D-10117 Berlin, Germany

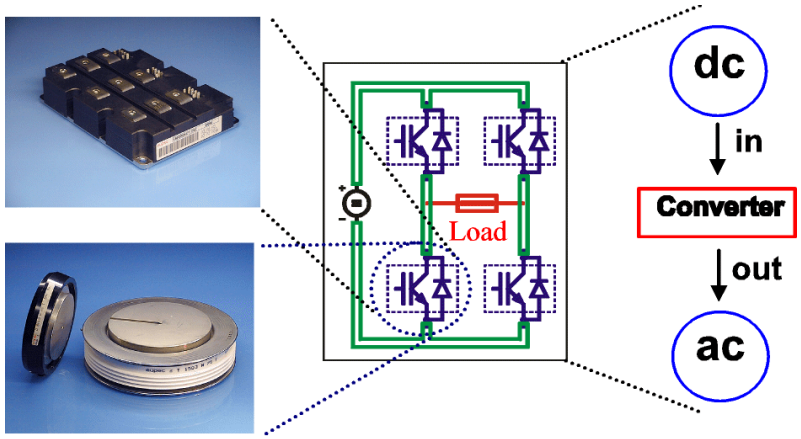


Figure 1: DC-AC converter module

$$\mathbf{B} \;=\; \mu\,\mathbf{H} \;\;,\;\; \mathbf{J} \;=\; \sigma\,\mathbf{E} \;\;. \tag{2}$$

where $\mathbf{B}$ and $\mathbf{J}$ stand for the magnetic induction and the current density, μ denotes the magnetic permeability, and σ is the electric conductivity.
Considering a module $\Omega = \cup_{\nu=1}^{N}\Omega_\nu$ with N bars Ω_ν , $1 \le \nu \le N$, each bar containing N_ν ports $\Gamma_{\nu\alpha}$, $1 \le \alpha \le N_\nu$, and introducing a scalar electric potential φ and a magnetic vector potential $\mathbf{A}$ according to

$$\mathbf{E} \;=\; -\text{grad}\,\varphi \;-\; \frac{\partial\,\mathbf{A}}{\partial\,t} \quad , \quad \mathbf{B} \;=\; \text{curl}\,\mathbf{A}$$

we are led to the following coupled system of PDEs

$$\text{div}\,(\sigma\;\text{grad}\,\varphi) \;=\; 0 \quad \text{in} \quad \Omega \;\;, \tag{3}$$

$$\sigma\;\mathbf{n}\cdot\text{grad}\,\varphi \;=\; \begin{cases} -I_{\nu\alpha}(t) & \text{on}\;\Gamma_{\nu\alpha} \\ 0 & \text{else} \end{cases} \tag{4}$$

$$\sigma\,\frac{\partial\,\mathbf{A}}{\partial\,t} \;+\; \text{curl}\;\mu^{-1}\,\text{curl}\;\mathbf{A} \;=\; \begin{cases} -\,\sigma\,\text{grad}\,\varphi & \text{in}\;\;\Omega \\ 0 & \text{in}\;\mathbf{R}^3\setminus\Omega \end{cases} \tag{5}$$

with appropriate initial and boundary conditions.
Note that in (4) we refer to $I_{\nu\alpha}$ as the fluxes associated with the ports $\Gamma_{\nu\alpha}$ satisfying $\sum_{\nu=1}^{N}\sum_{\alpha=1}^{N_\nu} I_{\nu\alpha} = 0$.
The total inductivity caused by the eddy currents can be described by the functional

$$L(\sigma,\varphi,\mathbf{A}) := \left(\sum_{\nu,\alpha}\sum_{\mu,\beta}\int_0^T |\,L_{\nu\alpha,\mu\beta}(t)\,|^2\,dt\right)^{1/2} . \tag{6}$$

Here, $L_{\nu\alpha,\mu\beta}(t)$ are the generalized transient inductivity coefficients

$$L_{\nu\alpha,\mu\beta}(t) := \sigma^{-1} \int_{\Omega_\nu} \mathbf{J}_{\nu\alpha}(x) \cdot \mathbf{S}(t)\mathbf{J}_{\mu\beta}(x)\, dx$$

where $\mathbf{J}_{\nu\alpha}$ denotes the current density generated by $I_{\nu\alpha}$ at the port $\Gamma_{\nu\alpha}$ of the bus bar Ω_ν and $\mathbf{S}(\cdot)$ is the solution operator associated with (5).
The design objective is to distribute the material in terms of the electric conductivity σ as the design parameter in such a way that the total inductivity is minimized

$$\inf_{\sigma,\varphi,\mathbf{A}} L(\sigma, \varphi, \mathbf{A}) \tag{7}$$

subject to the equality constraints

$$\varphi \text{ and } \mathbf{A} \text{ satisfy the state equations (3),(4),(5),} \tag{8}$$

$$\int_\Omega \sigma\, dx \;=\; C \tag{9}$$

and the inequality constraints

$$\sigma_{min} \;\leq\; \sigma \;\leq\; \sigma_{max} \tag{10}$$

where $0 < \sigma_{min} \ll 1$ and σ_{max} refers to the conductivity of copper.
Note that (10) represents relaxed constraints on the design parameter, since allowing only $\sigma = \sigma_{max}$ or $\sigma = \sigma_{min}$ would lead to an ill-posed optimization problem. In practice, we scale the conductivity by means of

$$\eta(\sigma) \;=\; \Big(\frac{\sigma - \sigma_{\min} + \varepsilon}{\sigma_{\max} - \sigma_{\min}}\Big)^m \;,\quad 0 < \varepsilon \ll 1 \tag{11}$$

with an appropriately chosen $m \geq 1$.

The primal-dual Newton interior-point method

The discretization of the state equations (3),(4),(5) is performed as follows: For the interior-exterior domain problem (5) we use a domain decomposition approach on nonmatching grids featuring individual edge element discretizations of the interior and exterior domain problems with respect to simplicial triangulations $\mathcal{T}_h^{(I)}$ and $\mathcal{T}_h^{(E)}$ whereas the discretization in time is done by the backward Euler scheme. Moreover, the elliptic boundary value problem (3),(4) is discretized by means of nonconforming Crouzeix-Raviart elements. The electric conductivity σ serving as the design parameter is discretized by elementwise constants, i.e., $\vec{\sigma}_h = (\sigma_h^{(1)}, ..., \sigma_h^{(m_h)})^T$, $m_h := \text{card}\, \mathcal{T}_h^{(I)}$. Comprising the discrete state variables $\vec{\varphi}_h$ and $\vec{\mathbf{A}}_h$ to a vector $\vec{\mathbf{u}}_h = (\vec{\varphi}_h, \vec{\mathbf{A}}_h)^T$, the discretized state equations can be stated in compact form

$$A_h(\vec{\sigma}_h)\, \vec{\mathbf{u}}_h \;=\; \vec{b}_h \;. \tag{12}$$

If we further denote by $L_h(\vec{\sigma}_h, \vec{\varphi}_h, \vec{\mathbf{A}}_h)$ the discretized objective functional, the topology optimization problem in the discrete regime reads as follows:

$$\min_{\vec{\sigma}_h, \vec{\varphi}_h, \vec{\mathbf{A}}_h} L_h(\vec{\sigma}_h, \vec{\varphi}_h, \vec{\mathbf{A}}_h) \tag{13}$$

subject to the constraints

$$\vec{\mathbf{u}}_h = (\vec{\varphi}_h, \vec{\mathbf{A}}_h)^T \text{ satisfies } (12) \ , \tag{14}$$

$$g_h(\vec{\sigma}_h) \ := \ \sum_{i=1}^{m_h} \mid K_i \mid \ \sigma_h^{(i)} \ = \ C \ , \tag{15}$$

$$\sigma_{\min} \, \vec{\mathbf{e}}_h \ \leq \ \vec{\sigma}_h \ \leq \ \sigma_{\max} \, \vec{\mathbf{e}}_h \ , \tag{16}$$

where $K_i \in \mathcal{T}_h^{(I)}$, $1 \leq i \leq m_h$, and $\vec{\mathbf{e}}_h := (1, ..., 1)^T$.
Among the most efficient numerical solution techniques for constrained optimization problems like (13)-(16) are primal-dual Newton interior-point methods (cf., e.g., [ETT96, FOG98, GOW98]). The idea is to take care of the inequality constraints (16) by parametrized logarithmic barrier functions

$$B_h^p(\vec{\sigma}_h, \vec{\mathbf{u}}_h) \ := \ L_h(\vec{\sigma}_h, \vec{\varphi}_h, \vec{\mathbf{A}}_h) \ - \ p \, [\log (\vec{\sigma}_h \ - \ \sigma_{min} \, \vec{\mathbf{e}}_h) \ + \ \log (\sigma_{max} \, \vec{\mathbf{e}}_h \ - \ \vec{\sigma}_h)]$$

and to couple the equality constraints (14),(15) by Lagrangian multipliers. This gives rise to the saddle point problem

$$\min_{\vec{\sigma}_h, \vec{\mathbf{u}}_h} \max_{\vec{\lambda}_h, \eta_h} \mathcal{L}_h^{(p)}(\vec{\sigma}_h, \vec{\mathbf{u}}_h, \vec{\lambda}_h, \eta_h) \tag{17}$$

for the Lagrangian

$$\mathcal{L}_h^{(p)}(\vec{\sigma}_h, \vec{\mathbf{u}}_h, \vec{\lambda}_h, \eta_h) \ := \ B_h^p(\vec{\sigma}_h, \vec{\mathbf{u}}_h) \ + \ \vec{\lambda}_h^T \, (A_h(\vec{\sigma}_h) \, \vec{\mathbf{u}}_h \ - \ \vec{\mathbf{b}}_h) \ + \ \eta_h \, (g_h(\vec{\sigma}_h) \ - \ C) \, .$$

For the solution of the above primal-dual interior-point approach we use simultaneous sequential quadratic programming in the sense that Newton's method is applied to the Karush-Kuhn-Tucker conditions associated with (13). Denoting the Newton increments by $\Delta \vec{\Psi}_h := (\Delta \vec{\mathbf{u}}_h, \Delta \vec{\lambda}_h, \Delta \vec{\sigma}_h, \Delta \eta_h)^T$, this gives rise to a linear system

$$\mathcal{K}_h \, \Delta \vec{\Psi}_h \ = \ \vec{\mathbf{d}}_h \tag{18}$$

which is solved iteratively by right transforming iterations

$$\Delta \vec{\Psi}_h^{\nu+1} \ = \ \Delta \vec{\Psi}_h^{\nu} \ + \ \mathcal{K}_h^R \, (\mathcal{M}_h^{(1)})^{-1} \, (\vec{\mathbf{d}}_h \ - \ \mathcal{K}_h \, \Delta \vec{\Psi}_h^{\nu}) \tag{19}$$

based on a regular splitting $\mathcal{K}_h \, \mathcal{K}_h^R = \mathcal{M}_h^{(1)} - \mathcal{M}_h^{(2)}$ involving an appropriately chosen right transform $\mathcal{K}_h^R$. The new iterate $\vec{\Psi}_h^{(\text{new})} := (\vec{\mathbf{u}}_h^{(\text{new})}, \vec{\lambda}_h^{(\text{new})}, \vec{\sigma}_h^{(\text{new})}, \eta_h^{(\text{new})})^T$ is then obtained by a line search

$$\vec{\Psi}_{h,i}^{(\text{new})} \ = \ \vec{\Psi}_{h,i}^{(\text{old})} \ + \ s_i \, (\Delta \vec{\Psi}_h)_i \quad , \quad 1 \leq i \leq 4 \ , \tag{20}$$

where the steplengths are tested by means of a hierarchy of merit functions. We refer to [HPS00, HPS01] for details.

Domain decomposition on nonmatching grids

The simultaneous sequential quadratic programming approach being integral part of the primal-dual Newton interior-point method, described in the previous section, requires an iterative solver of the discretized state equations. In this section, we briefly sketch a domain decomposition technique on nonmatching grids for the implicitly in time discretized equation (5) with respect to a nonoverlapping geometrically conforming decomposition $\bar{\Omega} = \bigcup_{i=1}^{n} \bar{\Omega}_i$ with skeleton $S = \bigcup_{i \neq j} \Gamma_{ij}$, $\Gamma_{ij} := \bar{\Omega}_i \cap \bar{\Omega}_j$. In particular, we consider individual simplicial triangulations $\mathcal{T}_h^{(i)}$ of the subdomains and discretize the subdomain problems by the lowest order curl-conforming edge elements $Nd_1(K) := \{ \mathbf{q} = \mathbf{a} + \mathbf{b} \wedge \mathbf{x} \mid \mathbf{a}, \mathbf{b} \in \mathbf{R^3} \}$, $K \in \mathcal{T}_h^{(i)}$ with the degrees of freedom given by the moments of the tangential components with respect to the edges of K (cf. [NED80]). Since nonconforming nodal points may occur on the interfaces $\Gamma_{ij} \subset S$, continuity of the tangential components across the interfaces is not guaranteed requiring weak continuity constraints on the skeleton in order to achieve consistency of the global approximation. This is taken care of by appropriately chosen Lagrangian multipliers living in multiplier spaces $\mathbf{M}_h(\Gamma_{ij})$, $\Gamma_{ij} \subset S$ (for the construction of $\mathbf{M}_h(\Gamma_{ij})$ we refer to [HOP99]). Introducing the product spaces

$$\mathbf{V}_h(\Omega) := \prod_{i=1}^{n} \mathrm{Nd}_1(\Omega_i, \mathcal{T}_h^{(i)}) \quad , \quad \mathbf{M}_h(S) := \prod_{\Gamma_{ij} \subset S} \mathbf{M}_h(\Gamma_{ij}) \ ,$$

where $\mathrm{Nd}_1(\Omega_i, \mathcal{T}_h^{(i)})$ are the edge element spaces associated with the subdomains, the domain decomposition approach leads to the discrete saddle point problem:
Find $(\mathbf{u}_h, \lambda_h) \in \mathbf{V}_h(\Omega) \times \mathbf{M}_h(S)$ such that

$$a_h(\mathbf{u}_h, \mathbf{v}_h) + b_h(\mathbf{v}_h, \lambda_h) = \ell(\mathbf{v}_h) \quad , \quad \mathbf{v}_h \in \mathbf{V}_h(\Omega) \ , \tag{21}$$

$$b_h(\mathbf{u}_h, \mu_h) = 0 \quad , \quad \mu_h \in \mathbf{M}_h(S) \ . \tag{22}$$

Here, the bilinear form $a_h : \mathbf{V}_h(\Omega) \times \mathbf{V}_h(\Omega) \to \mathbf{R}$ and the functional $\ell_h : \mathbf{V}_h(\Omega) \to \mathbf{R}$ are given by

$$a_h(\mathbf{u}_h, \mathbf{v}_h) := \sum_{i=1}^{n} \int_{\Omega_i} [\, \Delta t \mu^{-1} \, \mathbf{curl}\, \mathbf{u}_h \cdot \mathbf{curl}\, \mathbf{v}_h + \sigma \, \mathbf{u}_h \cdot \mathbf{v}_h \,] \, d\mathbf{x} \ ,$$

$$\ell_h(\mathbf{v}_h) := \int_{\Omega} \sigma \, [\, \mathbf{u}_h^{m-1} \cdot \mathbf{v}_h - \Delta t \, \mathbf{grad} \varphi_h^m \cdot \mathbf{v}_h \,] \, d\mathbf{x} \ ,$$

where $\mathbf{u}_h^{m-1}$ and φ_h^m refer to the FE approximations of the magnetic vector potential and the scalar electric potential at time t_{m-1} and t_m, respectively, and $\Delta t := t_m - t_{m-1}$.
Moreover, the bilinear form $b_h : \mathbf{V}_h(\Omega) \times \mathbf{M}_h(S) \to \mathbf{R}$ realizing the weak continuity of the tangential components across the interfaces is chosen as follows

$$b_h(\mathbf{v}_h, \vec{\mu}_h) := \sum_{\Gamma_{ij} \subset S} \int_{\Gamma_{ij}} \vec{\mu}_h \cdot [\mathbf{n} \wedge \mathbf{v}_h] \, |_{\Gamma_{ij}} \ ds$$

with $[\mathbf{n} \wedge \mathbf{v}_h] \, |_{\Gamma_{ij}}$ denoting the jump of $\mathbf{n} \wedge \mathbf{v}_h$ across the interface $\Gamma_{ij} \subset S$.
It can be shown that $a_h(\cdot, \cdot)$ is elliptic on the kernel of the operator associated with $b_h(\cdot, \cdot)$ and

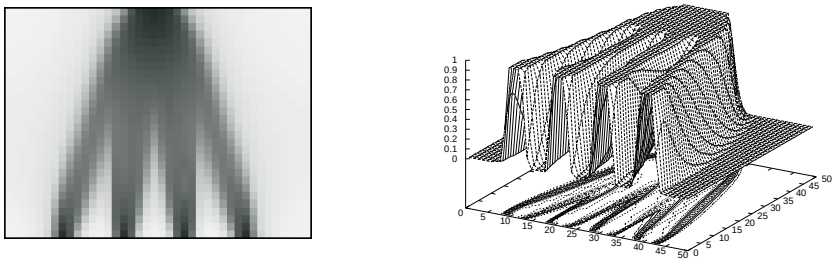


Figure 2: Material distribution (5 ports)

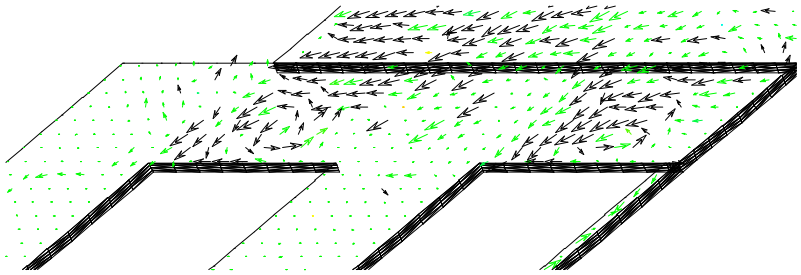


Figure 3: Magnetic induction between two ports (zoom)

that $b_h(\cdot,\cdot)$ satisfies an LBB-condition (cf. [HOP99]). The numerical solution of (21),(22) is done by preconditioned Richardson-type iterations with a multilevel preconditioner and features an additional defect correction in subspaces of irrotational vector fields that takes care of the nontrivial kernel of the discrete curl-operator. We refer to [HOP00] for details (cf. also [BBM99] for a related approach). Grid adaptation strategies based on efficient and reliable residual-type a posteriori error estimators can be performed along the lines of [BHH00]).

Numerical results

The primal-dual Newton interior-point method has been tested in 2D with the total amount of dissipated electric energy to be minimized and an optimal design has been computed in 3D for an individual bus bar by using the techniques described in the previous sections. The numerical simulation provides a material distribution that can be visualized by grey-scales ranging from black ($\sigma = \sigma_{max}$) to white ($\sigma = \sigma_{min}$) and by corresponding height profiles. Figure 2 displays the material distribution for a 2D test case (bus bar with 5 ports).
We observe a sharp resolution of the interface “material - no material”. The performance of the primal-dual Newton interior-point method depends on the number of ports and the parameter m in (11) (for details see [HPS00]).
For an individual 3D bus bar, Figure 3 shows a visualization of the computed magnetic induction $\mathbf{B}$ for the final design in a vicinity between two ports. One clearly recognizes the effect of the topology optimization (holes close to the ports) on the distribution of the magnetic

induction (for a more detailed documentation we refer to [BHM01]).

Acknowledgments. The work has been supported by grants from the Federal Ministry for Education and Research (BMBF) under Grant No. 03HO7AU1-8 and Grant No. 03HOM3A1 and from the German National Science Foundation (DFG) under Grant No. HO877/4-1 and Grant No. HO877/5-1. The second author is indebted to the Alexander-von-Humboldt Foundation for an AvH-Fellowship.

References

[BDH99]R. Beck, P. Deuflhard, R. Hiptmair, R.H.W. Hoppe, and B. Wohlmuth, "Adaptive multilevel methods for edge element discretizations of Maxwell's equations", Surveys of Math. in Industry, Vol. **8**, pp. 271–312, (1999).

[BHH00]R. Beck, R. Hiptmair, R.H.W. Hoppe, and B. Wohlmuth, "Residual based a posteriori error estimators for eddy current computation", to appear in M^2AN Math. Modelling and Numer. Anal., (2000).

[BBM99]F. Ben Belgacem, A. Buffa, and Y. Maday, "The mortar finite element method for 3D Maxwell equations: First results", Report 99023, Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie, Paris, (1999).

[BEN95]M.P. Bendsøe, "Optimization of Sructural Topology, Shape, and Material", Springer, Berlin-Heidelberg-New York, 1995.

[BFS99]P. Böhm, E. Falck, J. Sigg, and G. Wachutka, "Numerical analysis of distributed inductive parasitics in high power bus bars", In: High Performance Scientific and Engineering Computing. Proc. "Int. FORTWIHR-Symposium", Munich, March 1998 (Bungartz, H., Durst, F.,and Zenger, Chr.; eds.), pp. 397-404, Lecture Notes in Computational Science and Engineering, Vol. **8**, Springer, Berlin-Heidelberg-New York, 1999

[BHM01]P. Böhm, R.H.W. Hoppe, G. Mazurkevitch, S. Petrova, G. Wachutka, and E. Wolfgang, "Optimal design of high power electronic devices by topology optimization", to appear in: Mathematik - Schlüsseltechnologie für die Zukunft. Verbundprojekte zwischen Mathematik und Industrie, Springer, Berlin-Heidelberg-New York, (2001).

[DGH98]St. Dürndorfer, V. Gradinaru, R.H.W. Hoppe, E.-R. König, G. Schrag, and G. Wachutka, "Numerical simulation of microstructured semiconductor devices, transducers, and systems", In: High Performance Scientific and Engineering Computing. Proc. "Int. FORTWIHR-Symposium", Munich, March 1998 (Bungartz, H., Durst, F.,and Zenger, Chr.; eds.), pp. 309–323, Lecture Notes in Computational Science and Engineering, Vol. **8**, Springer, Berlin-Heidelberg-New York, (1999).

[ETT96]A.S. El-Bakry, R.A. Tapia, T. Tsuchiya, and Y. Zhang, "On the formulation of the Newton interior-point method for nonlinear programming", Journal of Optimization Theory and Applications, **89**, 507-541, (1996).

[FOG98]A. Forsgen and Ph. Gill,"Primal-dual interior methods for nonconvex nonlinear programming", SIAM J. Optimization **8**, 1132-1152, (1998).

[GOW98]D.M. Gay, M.L. Overton, and M.H. Wright, "A primal-dual interior method for nonconvex nonlinear programming", In: Advances in Nonlinear Programming (Yuan, Y.; ed.), pp. 31-56, Kluwer, Dordrecht, (1998)

[HEI00]M. Heinkenschloss, “Time-domain decomposition iterative methods for the solution of parabolic linear-quadratic optimal control problems”, Techn. Rep., Department of Comput. and Appl. Math., Rice University, Houston, (2000).

[HOP99]R.H.W. Hoppe, “Mortar edge elements in $\mathbf{R}^3$”, East-West J. Numer. Anal., Vol. **7**, 159–173, (1999).

[HOP00]R.H.W. Hoppe, “Adaptive mortar edge elements in the computation of eddy currents”, In: Proc. Conf. ”Analysis and Approximation of Boundary Value Problems”, Jyväskylä (Finland), October 1998 (Neittaanmäki, P. et al.; eds.), (2000).

[HPS00]R.H.W. Hoppe, S. Petrova, and V. Schulz, “A primal-dual Newton-type interior-point method for topology optimization”, to appear in Journal of Optimization: Theory and Applications, 2002

[HPS01]R.H.W. Hoppe, S. Petrova, and V. Schulz, “Topology optimization of high power electronic devices”, to appear in Proc. Oberwolfach Conf. ”Optimal Control of Complex Dynamical Structures”, June 4-10, 2000 (Hoffmann, K.-H. and Leugering, G.; eds.), Birkhäuser, Basel, 2001

[MAS00]B. Maar and V. Schulz, “Interior point multigrid methods for topology optimization”, Structural Optimization **19**, 214-224, (2000).

[NED80]J.-C. Nédélec, “Mixed finite elements in $\mathbf{R}^3$”, Numer. Math. **35**, 315-341, (1980).

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

51 Domain Decomposition Method Applied to a Coupling Vibration Problem between Shell and Acoustics

T. Kako[1], H. M. Nasir[2]

Introduction

We consider the numerical method for the structural-acoustic coupling vibration problem between a shell and acoustic fields by the finite element method. The structure is a shell S which encloses a bounded acoustic region Ω_1 and is surrounded by an unbounded acoustic region Ω_2. The structural-acoustic system is described by a coupled problem between the acoustic pressure perturbations of the inner and outer regions and the tangential and normal deformations of the shell. The problem can be regarded as a domain decomposition formulation for the acoustic fields with a generalized Lagrangian multiplier. The normal deformation of the shell acts as the Lagrangian multiplier which is in turn coupled with tangential deformation of the shell. The finite element approximation to the problem results in a block matrix equation. In order to solve this matrix equation by iterative methods, we consider two techniques: one is based on the Schur complement of the block matrix with appropriate preconditioners and the other is a direct iteration with some block preconditioners. We use a descretized version of fictitious domain method to construct the block matrices and use the Krylov subspace iteration method for solving the system of equations [HKNT98]. The fictitious domain is used to obtain preconditioners for the diagonal matrix blocks. The Schur complement technique requires a double iteration whereas the direct iteration techniques requires only a single iteration. We observe that the direct iteration technique with block preconditioners performs well compared to the Schur complement technique.

Let there be two acoustic regions Ω_1 and Ω_2 in $\mathbb{R}^d, d = 2, 3$, separated by a closed shell structure. The domain Ω_1 is bounded and enclosed by the shell and the domain Ω_2 is unbounded (see Fig. 1). Let $\hat{p}_i, \rho_i$ and c_i be the acoustic pressure perturbation, the density of acoustic material and the sound velocity in the domain $\Omega_i, i = 1, 2$ respectively and p^{inc} be the pressure perturbation of an incident wave from the outer region Ω_2. Then the governing equations for the vibration of the system are given by

$$
\begin{aligned}
\frac{\partial^2 \hat{p}_i}{\partial t^2} - c_i^2 \Delta \hat{p}_i &= 0 && \text{in } \Omega_i, \quad i = 1, 2, \\
\frac{\partial \hat{p}_i}{\partial n} &= -\rho_i \frac{\partial^2 u_n}{\partial t^2} && \text{on } S, \quad i = 1, 2, \\
\rho_0 \frac{\partial^2 \mathbf{u}}{\partial t^2} + \mathbf{A}\mathbf{u} &= (\hat{p}_1 - \hat{p}_2)|_S \mathbf{n} && \text{on } S, \\
\hat{p}_2 - p^{inc} & && \text{is outgoing,}
\end{aligned}
$$

where $\mathbf{n}$ is the outward unit normal to the shell surface; $\mathbf{u}$ is the vector of the shell deformations; $u_n = \mathbf{u} \cdot \mathbf{n}$ is the shell deformation along the normal $\mathbf{n}$; $\mathbf{A}$ is the shell force operator.

[1]Univ. of Elecro-Comm., Chofu, Tokyo, Japan, kako@im.uec.ac.jp
[2]University of Peradeniya, Sri Lanka, nasirhm@fedu.uec.ac.jp

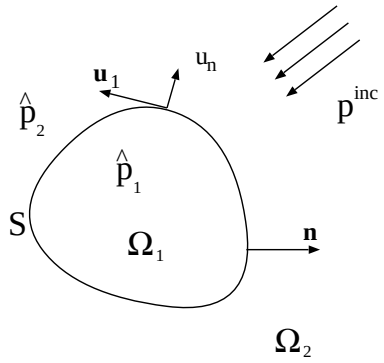


Figure 1: Structural-acoustic coupling

For a general shaped arc shell in two dimensions, the operator $\mathbf{A}$ consists of membrane and flexural components: $\mathbf{A} = D(\mathbf{A}^{memb} + \frac{e^2}{12}\mathbf{A}^{flex})$ with

$$\mathbf{A}^{memb} = \begin{bmatrix} -\partial_s^2 & \partial_s\kappa \\ -\kappa\partial_s & \kappa^2 \end{bmatrix} \text{ and } \mathbf{A}^{flex} = \begin{bmatrix} -A_1^2 & -A_1A_2 \\ A_2A_1 & A_2^2 \end{bmatrix}$$

where $A_1 = 2\kappa\partial_s + \kappa'$; $A_2 = \partial_s^2 - \kappa^2$; ∂_s and $'$ denote differentiation with respect to the arc length s, κ and e are the curvature and thickness of the shell respectively and $D = E/(1-\nu^2)$ is the flexural rigidity of the shell with Young's modulus E and Poisson ratio $\nu(0 < \nu \leq 1/2)$. Let $\hat{p}_1 = p_1$ and $\hat{p}_2 = p_2 + p^{inc}$. Then, p_1 and p_2 represent the pressures of scattered waves inside and outside respectively. We also assume that the incident wave, the scattering waves and the deformations of the shell are time-harmonic: $f(x,t) = f(x)e^{\mathbf{i}wt}, f = p_1, p_2, p^{inc}$ or $\mathbf{u}$. Then, the problem can be written as follows:

$$\begin{aligned}
-\Delta p_i - k_i^2 p_i &= 0, && \text{in } \Omega_i, i = 1,2, && \text{(1a)}\\
\frac{\partial p_1}{\partial n} &= \rho_1\omega^2 u_n && \text{on } S, && \text{(1b)}\\
\frac{\partial p_2}{\partial n} &= \rho_2\omega^2 u_n - \frac{\partial p^{inc}}{\partial n} && \text{on } S, && \text{(1c)}\\
\mathbf{A}\mathbf{u} - \rho_0\omega^2\mathbf{u} &= (p_1 - p_2 - p^{inc})|_S\mathbf{n} && \text{on } S, && \text{(1d)}\\
r^{\frac{d-1}{2}}\left(\frac{\partial p_2}{\partial n} - \mathbf{i}k_2 p_2\right) &\longrightarrow 0, && \text{as } r \to \infty, && \text{(1e)}
\end{aligned}$$

where $k_i = \omega/c_i, i = 1,2$ are the wave numbers corresponding to the inner and outer acoustic regions respectively. The last condition is the Sommerfeld radiation condition for the scattering wave p_2 which allows only the out-going waves in the solution for the outer region.

Approximate problem and weak formulation

For the numerical treatment of the problem, introducing an artificial boundary Γ_R, we restrict the unbounded domain Ω_2 into a bounded domain Ω_R and impose an artificial radiation

boundary condition on Γ_R. We choose Γ_R to be a circle or a sphere of radius R for the two or three dimensional problems respectively.
The Sommerfeld radiation condition is then replaced by the radiation boundary condition

$$\frac{\partial p_2}{\partial n} = Mp_2 \tag{2}$$

where M is a differential or pseudo-differential operator with respect to the tangent parameter of the boundary Γ_R.
Let us consider the function spaces $V_1 = H^1(\Omega_1), V_2 = H^1(\Omega_2), V_3 = H^1(S)$ and $V_4 = H^2(S)$ as the solution spaces for $p_1, p_2, \mathbf{u}_t$ and u_n respectively. Here, $\mathbf{u} = (\mathbf{u}_t, u_n)$ and $\mathbf{u}_t$ is the vector of tangential deformation.
The weak formulation of the problem (1) can be given as follows:
Find $(p_1, p_2, \mathbf{u}_t, u_n) \in V_1 \times V_2 \times V_3 \times V_4$ such that, for all $(q_1, q_2, \mathbf{v}_t, v_n) \in V_1 \times V_2 \times V_3 \times V_4$

$$\begin{aligned} a_1(p_1, q_1) + \rho_1\omega^2(u_n, q_1)_S &= 0, && \text{(3a)}\\ a_1(p_2, q_2) - m(p_2, q_2)_{\Gamma_R} - \rho_2\omega^2(u_n, q_2)_{\Omega_R} &= (\partial p^{inc}/\partial n, q_2)_S, && \text{(3b)}\\ b(\mathbf{u}, \mathbf{v}) - (p_1, v_n)_S + (p_2, v_n)_S &= -(p^{inc}, v_n)_S && \text{(3c)} \end{aligned}$$

where

$$\begin{aligned} a_1(p_1, q_1) &= (\nabla p_1, \nabla q_1)_{\Omega_1} - k_1(p_1, q_1)_{\Omega_1},\\ a_2(p_2, q_2) &= (\nabla p_2, \nabla q_2)_{\Omega_R} - k_2(p_2, q_2)_{\Omega_R},\\ m(p_2, q_2)_{\Gamma_R} &= \int_{\Gamma_R} (Mp_2)\bar{q}_2 ds \quad \text{and} \quad b(\mathbf{u}, \mathbf{v}) = \int_S (\mathbf{A} - \rho_0\omega^2)\mathbf{u}\bar{\mathbf{v}} d\sigma. \end{aligned}$$

We introduce finite dimensional subspaces V_{ih} of $V_i, i = 1, 2, 3, 4$ respectively and consider the approximate weak formulation, i.e., the finite element method:
Find $(p_{1h}, p_{2h}, \mathbf{u}_{th}, u_{nh}) \in V_{1h} \times V_{2h} \times V_{3h} \times V_{4h}$ such that for all
$(q_1, q_2, \mathbf{v}_t, v_n) \in V_{1h} \times V_{2h} \times V_{3h} \times V_{4h}$,

$$\begin{aligned} a_1(p_{1h}, q_1) + \rho_1\omega^2(u_{nh}, q_1)_S &= 0, && \text{(4a)}\\ a_1(p_{2h}, q_2) - m(p_{2h}, q_2)_{\Gamma_R} - \rho_2\omega^2(u_{nh}, q_2)_{\Omega_R} &= (\partial p^{inc}/\partial n, q_2)_S, && \text{(4b)}\\ b(\mathbf{u}_h, \mathbf{v}) - (p_{1h}, v_n)_S + (p_{2h}, v_n)_S &= -(p^{inc}, v_n)_S. && \text{(4c)} \end{aligned}$$

By choosing bases for the function spaces and writing $p_{1h}, p_{2h}, \mathbf{u}_{th}$ and u_{nh} with respect to these bases, we obtain the block matrix equation as follows:

$$\begin{bmatrix} M_1 & 0 & 0 & -\rho_1\omega^2 L_1^T \\ 0 & M_2 & 0 & \rho_2\omega^2 L_2^T \\ 0 & 0 & A & B^T \\ -L_1 & L_2 & B & C \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ U_t \\ U_n \end{bmatrix} = \begin{bmatrix} 0 \\ F \\ 0 \\ G \end{bmatrix} \tag{5}$$

where each block corresponds to the sesquilinear form and its entries are given with respect to the chosen base functions.
The matrices M_1 and M_2 are constructed by the finite element discretization of fictitious domains. For the inner bounded domain Ω_1, we consider a rectangular region such that Ω_1 is

included in it. We discretize the rectangular domain by a uniform orthogonal rectangular grid. Then, the nodes close to the boundary of Ω_1 are moved onto the boundary so that the new locally modified partition is topologically equivalent to the orthogonal grid partition. Then, the modified rectangles are triangulated such that the resulting triangles satisfy a regularity condition. The computational domain for the inner region is then obtained by discarding the extended portion in the rectangular fictitious domain.
Similarly, for the outer domain Ω_R, we enlarge the domain towards inside the inner boundary of Ω_R so that it makes an annulus including the inner boundary of Ω_R. We discretize the annulus by a uniform orthogonal polar grid. The nodes near the inner boundary of Ω_R are modified as in the case of inner domain (see Fig. 2).

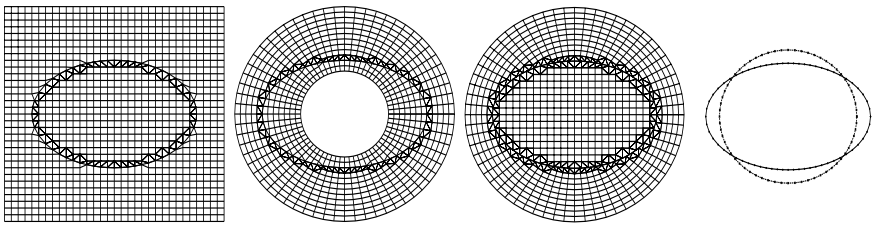


Figure 2: Fictitious domains and partitioning

Preconditioners for the matrices M_1 and M_2 that are constructed by the fictitious domain method are obtained by using the enlarged fictitious domain itself. We explain the construction of the preconditioner for the inner region. The one for the outer region follows analogously.
The unmodified orthogonal mesh is used to obtain a matrix by using the same weak formulation on the fictitious domain. This will give a matrix $\mathbf{N}$ which we write in a block form as follows:

$$\mathbf{N}_1 = \begin{bmatrix} N_{11} & N_{12} \\ N_{12}^T & N_{22} \end{bmatrix}$$

where the matrix N_{11} corresponds to the nodes on the inner region, but not moved; the matrix N_{22} corresponds to the nodes outside the inner region. The matrix $\mathbf{N}_1$ is obviously larger in size than the original matrix M_1 which corresponds to the moved nodes on the inner boundary. When we want to solve a matrix equation of the form

$$M_1 P_1 = F_1,$$

we enlarge the system as follows:

$$\mathbf{M}_i = \begin{bmatrix} M_i & N_{12} \\ 0 & N_{22} \end{bmatrix} \begin{bmatrix} P_1 \\ Q_1 \end{bmatrix} = \begin{bmatrix} F_1 \\ 0 \end{bmatrix}.$$

The two system of equations are equivalent in the sense that the solution P_1 is the same for both systems. Hence, we solve the enlarged system using the Krylov subspace iteration method with the matrix $\mathbf{N}_1$ as a preconditioner. For more details of the fictitious domain method see [HKNT98].

Schur Complement Method

The Schur complement of the block matrix with respect to its last block is obtained by solving the block matrix equation (5) for the vector component U_2:

$$\left[C - BA^{-1}B^T - \rho_1\omega^2 L_1 M_1^{-1} L_1^T - \rho_2\omega^2 L_2 M_2^{-1} L_2^T\right] U_2 = G - M_2^{-1}F \qquad (6)$$

This matrix can then be solved numerically by using the Krylov subspace method. The terms involving matrix inverses in this Schur complement are computed based on the fictitious domain method with preconditioners obtained from the fictitious domains.

Direct Iteration Method

In this method, we directly use the Krylov subspace iteration procedure to solve the block matrix equation (5). For this purpose, the block matrix equation is enlarged to the the one with the size corresponding to that of their fictitious domain preconditioners as follows:

$$\begin{bmatrix} \mathbf{M}_1 & 0 & 0 & -k^2\mathbf{L}_1^T \\ 0 & \mathbf{M}_2 & 0 & k^2\mathbf{L}_2^T \\ 0 & 0 & A & B^T \\ -\mathbf{L}_1 & \mathbf{L}_2 & B & C \end{bmatrix} \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \mathbf{U}_t \\ \mathbf{U}_n \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{F} \\ \mathbf{0} \\ \mathbf{G} \end{bmatrix}$$

where the matrices in bold symbols are the enlarged matrices of their counterparts in the block matrix equation (5).
The preconditioner used for this method is based on the preconditioning technique by Bramble and Pasciak [BP88] which is given as follows:

$$\begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ -\mathbf{L}_1 & \mathbf{L}_2 & B & -I \end{bmatrix} \begin{bmatrix} \mathbf{N}_1^{-1} & 0 & 0 & 0 \\ 0 & \mathbf{N}_2^{-1} & 0 & 0 \\ 0 & 0 & A_0^{-1} & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$$

where the matrix A_0 is the preconditioner for the matrix A based on the fictitious shell domain (see Fig. 2). The first matrix is an elementary pre-multiplication matrix which makes the preconditioned matrix symmetric.

Numerical Results

We present in this section the results of the implementation of the method. All computations were carried out on VT-Alpha5, 533Mhz, 512MB RAM with Linux operating system environment with double precision arithmetic using object oriented C++ codes.
We test the two iterative methods in the last two sections for a two dimensional shell-acoustic coupling problem. The shell is a circular arc of radius $r_0 = 1$. The densities of the acoustic material in both inner and outer acoustic regions are the same $\rho_1 = \rho_2 = 1$. The artificial boundary for the outer acoustic region is a circle of radius $R = 2$ is chosen. The incident wave is a plane wave with wave number $k = \pi$.
In the Schur complement methods, each iteration step requires the matrix inverses of M_1, M_2 and A_0. These are performed by an inner iteration using the fictitious domain method. Hence,

each iteration step of the Schur complement matrix equation involves other iterations. Table 1 shows the number of iterations and times for both Schur complement and direct iteration methods.

Method	Outer Iter.	Inv. Mat. Multiplications			time (sec.)
		M_1	M_2	A	
Schur Complement	23	1276	267	23	80.45
Direct iteration	35	35	35	35	3.20

Table 1: Performances of Schur complement and direct iteration methods

The Schur complement method has 23 outer iteration steps each of them has inner iterations. The total numbers of inner iterations are 1276 for M_1, 267 for M_2, and 23 for A and the total time for the iterations is 80.45sec. For the case of A, the preconditioner is A_0 the same as A, because the shell is circular. Hence, it has only one iteration per step.
For the direct preconditioning method, the total iterations required to achieve the same result is 35. Each iterative step requires one matrix multiplication with the preconditioned matrices for M_1, M_2 and A_0. Hence the total numbers of iterations is 35 for each matrices. The time required for the iterations is 3.20sec.
Figure 3 shows the real part of the scattering waves for circular and elliptic shell cases. The radius of the artificial boundary is $R = 2$. The incident wave is a plane wave coming from left along the x-axis direction with wave numbers $k = 2\pi$ and 3π. The radius of the shell is $r_0 = 1$ and the major and minor axes of the elliptic shell are $2a = 3.2$ and $2b = 2$.

Conclusion

The structural-acoustic coupling problem between a shell and inner and outer acoustic fields is considered by the finite element method. Fictitious domain method is used to discretize the acoustic domains and the resulting block matrix equation is solved by a Krylov subspace iteration methods.
Two schemes are used: A Schur complement method and a direct block iteration method. The Schur complement method requires a double iteration while the direct block iteration method needs a single iteration.
The block iteration method performs very well in terms of the numbers of matrix multiplications and computing time.

References

[Ber96]Michel Bernadou. *Finite Element Methods for Thin Shell Problems*. John Wiley & Sons, 1996. Translated by Claude Andrew James.

[BP88]James H. Bramble and Joseph E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Mathematics of Computation*, 50(181):1–17, 1988.

[BW90]Christoph Börgers and Olof B. Widlund. On finite element domain imbedding methods. *SIAM J. Numer. Anal.*, 27(4), August 1990.

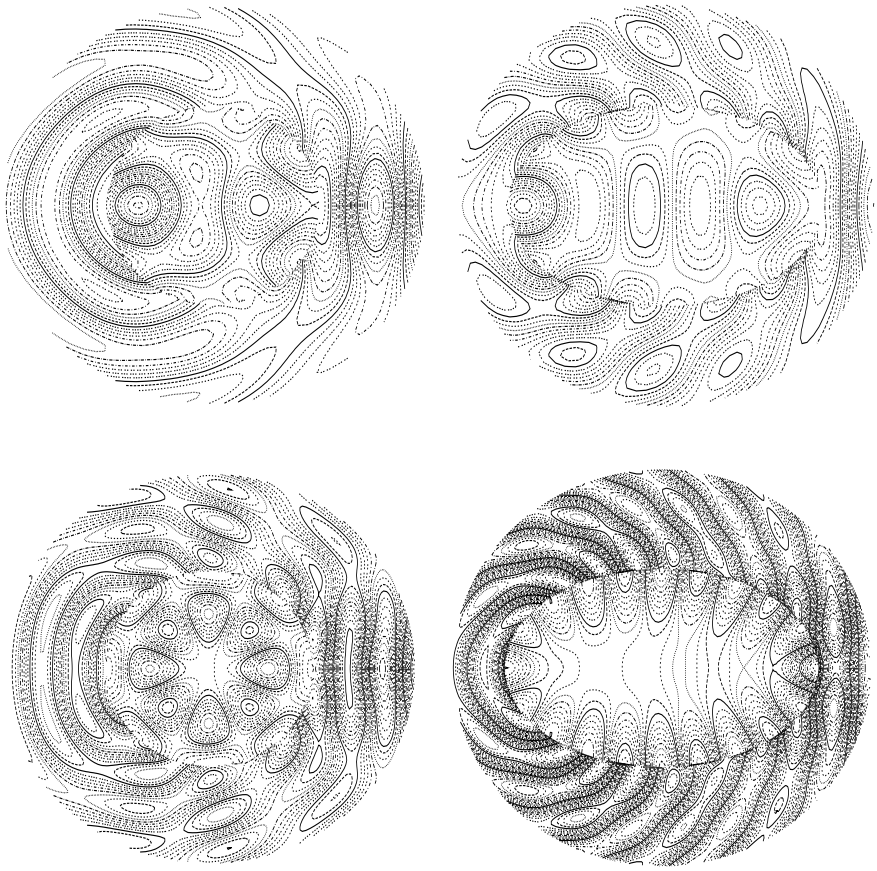


Figure 3: Scattering waves: circular and elliptic cases

[Cia78]Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[Ern96]Oliver G. Ernst. A finite-element capacitance matrix method for exterior helmholtz problems. *Numer. Math.*, 75(2):175–204, 1996.

[Fre93]R. W. Freund. A transpose-free quasi-minimal residual algorithm for non-hermitian linear systems. *SIAM J. Sci. Comput.*, 14(2):470–482, 1993.

[HKNT98]E. Heikkola, Y. A. Kuznetsov, P. Neittaanmaki, and J. Toivanen. Fictitious domain methods for the numerical solution of two-dimesional scattering problems. *J. Comput. Phys.*, 145:89–109, 1998.

[Li98]Deng Li. *Study on Finite Element Approximation and Perturbation Analysis of Eigenvalue Problems for Structural-Acoustic Coupled System*. PhD thesis, The University of Electro-Communications, Japan, March 1998.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

52 Iterative substructuring methods for incompressible and nonisothermal flows using the $k-\epsilon$ turbulence model

T. Knopp[1], G. Lube[2], H. Müller[3]

Introduction

We consider the parallel solution of the incompressible Navier-Stokes equations coupled with the energy equation. For turbulent flows, the k/ϵ model is used. The iterative process requires the fast solution of advection-diffusion-reaction and Oseen type problems. These linearized problems are discretized using stabilized FEM. We apply an iterative substructuring method which couples the subdomain problems via Robin-type interface conditions. Then we apply the approach to the simulation of indoor air flow problems.
The mathematical model under consideration is the incompressible, nonisothermal (Reynolds averaged) Navier-Stokes problem in a bounded polyhedral domain $\Omega \subset \mathbf{R}^d$. For turbulent flows we apply the $k-\epsilon$ model, cf. [CS99, MP94, Mue99]. Turbulent effects are modelled as additional turbulent viscosity $\nu_t = C_\mu \frac{k^2}{\epsilon}$ and thermal diffusivity $a_t = \frac{\nu_t}{Pr_t}$, using the turbulent kinetic energy k and turbulent dissipation ϵ. Bouyancy effects are taken into account using the Boussinesq approximation.
The velocity $\vec{u}$, the (reduced) pressure p, and the temperature θ, and in the turbulent case, the quantities k and ϵ are solutions of the coupled nonlinear system

$$\left\{\begin{array}{rcl} \partial_t\vec{u} - \vec{\nabla}\cdot(2\nu_e S(\vec{u})) + (\vec{u}\cdot\vec{\nabla})\vec{u} + \vec{\nabla}p &=& -\beta\theta\vec{g} \\ \vec{\nabla}\cdot\vec{u} &=& 0 \\ \partial_t\theta + (\vec{u}\cdot\vec{\nabla})\theta - \vec{\nabla}\cdot(a_e\vec{\nabla}\theta) &=& \dot{q}^V/c_p \\ \partial_t k + (\vec{u}\cdot\vec{\nabla})k - \vec{\nabla}\cdot(\nu_k\vec{\nabla}k) &=& P_k + G - \epsilon \\ \partial_t\epsilon + (\vec{u}\cdot\vec{\nabla})\epsilon - \vec{\nabla}\cdot(\nu_\epsilon\vec{\nabla}\epsilon) + C_2\epsilon^2 k^{-1} &=& C_1\epsilon k^{-1}(P_k + G) \end{array}\right. \quad (1)$$

with constants $C_1, C_2, C_\mu, C_t, Pr_t, Pr_k, Pr_\epsilon$, effective viscosities $\nu_e = \nu + \nu_t$, $a_e = a + a_t, \nu_k = \nu + \frac{\nu_t}{Pr_k}, \nu_\epsilon = \nu + \frac{\nu_t}{Pr_\epsilon}$, production and bouyancy terms

$$P_k := 2\nu_t|S(\vec{u})|^2, \quad G := C_t\beta\frac{\nu_t}{Pr_t}\vec{g}\cdot\vec{\nabla}\theta \quad \text{with } S(\vec{u}) := \frac{1}{2}(\vec{\nabla}\vec{u} + \vec{\nabla}\vec{u}^T).$$

In laminar flows we set $k \equiv 0$ and skip the $k-\epsilon$ equations in (1). The boundary is divided into inlet, outlet and wall zones Γ_-, Γ_+ and Γ_0 depending on the sign of $\vec{u}\cdot\vec{n}$. Using $\tau = 2\nu_e S(\vec{u})$, we set in *laminar* flows

$$(\tau - pI)\vec{n} = \tau_n\vec{n} \text{ on } \Gamma_N \subset \Gamma_- \cup \Gamma_+, \qquad \vec{u} = \vec{u}_b \text{ on } \Gamma_D \subset \Gamma_- \cup \Gamma_+, \quad (2)$$

[1]University of Göttingen, Math. Departm., D-37073 Göttingen, knopp@math.uni-goettingen.de
[2]University of Göttingen, Math. Departm., D-37073 Göttingen, lube@math.uni-goettingen.de
[3]Aerotec Engrg. GmbH Hamburg

with $\Gamma_D \cap \Gamma_N = \emptyset$ and $\overline{\Gamma}_D \cup \overline{\Gamma}_N = \overline{\Gamma}_- \cup \overline{\Gamma}_+$. On Γ_0, we prescribe either the tangential stresses and the normal velocity or a no-slip condition

$$(i)\ \ (I - \vec{n} \otimes \vec{n})\tau\vec{n} = \vec{\tau}_t, \quad \vec{u} \cdot \vec{n} = 0, \qquad \text{or } (ii)\ \ \vec{u} = \vec{0} \quad \text{on } \Gamma_0. \tag{3}$$

In this paper, we consider only case (i) with $\Gamma_D = \emptyset$. For θ we set

$$\theta = \theta_{in} \quad \text{on } \Gamma_-\ ,\ a_e\vec{\nabla}\theta \cdot \vec{n} = 0 \text{ on } \Gamma_+\ ,\ a_e\vec{\nabla}\theta \cdot \vec{n} = \dot{q}_0/c_p \ \text{ on } \Gamma_0. \tag{4}$$

For *turbulent* flows, we apply the concept of wall functions in a neighbourhood Ω_δ of Γ_0 containing at least the so-called viscous sub-layer. Firstly, as usual in wall law theory, the r.h.s. $\vec{\tau}_t$ in (3) and $\dot{q}_0$ in (4) are modified. We set $\vec{\tau}_t = U_*^2\vec{u}/||\vec{u}||$ and seek $(U_*, \dot{q}_0)$ as solutions of coupled nonlinear equations. Secondly, for the $k-\epsilon$ equations the computational domain is $\Omega \backslash \Omega_\delta$. Dirichlet data are prescribed on Γ_- and on the artificial boundary $\Gamma_\delta = \partial\Omega_\delta \cap \Omega$. A no-flow condition is specified on Γ_+. A computational algorithm has to control that Γ_δ, being discretized with mesh points with minimal distance to Γ_0, belongs to the so-called log-layer. For details see [KLGR00, Mue99].

Discretization, decoupling, and linearisation

Semidiscretization in time of system (1):
We are mainly interested in the long-term behaviour of the model. So we apply the backward Euler scheme on a partition $\{t_m\}_{m=0}^M$ of $[0,T]$ with $t_0 = 0,\ t_M = T$. We use the abbreviation $F^m = F(t_m) \equiv F(t_m, \cdot)$ for a function F. The time derivative $\partial_t F(t_m)$ is approximated by $\partial_t^m F = (F^m - F^{m-1})/\triangle_m$ with time-step $\triangle_m = t_m - t_{m-1}$. We arrive at the semidiscrete system

$$\begin{aligned}
\partial_t^m\vec{u} - \vec{\nabla} \cdot (2\nu_e^m S(\vec{u}^m)) + (\vec{u}^m \cdot \vec{\nabla})\vec{u}^m + \vec{\nabla}p^m &= -\beta\theta^m\vec{g} \\
\vec{\nabla} \cdot \vec{u}^m &= 0 \\
\partial_t^m\theta + (\vec{u}^m \cdot \vec{\nabla})\theta^m - \vec{\nabla} \cdot (a_e^m\vec{\nabla}\theta^m) &= (\dot{q})^{V^m}/c_p \\
\partial_t^m k + (\vec{u}^m \cdot \vec{\nabla})k^m - \vec{\nabla} \cdot (\nu_k^m\vec{\nabla}k^m) &= P_k^m + G^m - \epsilon^m \\
\partial_t^m\epsilon + (\vec{u}^m \cdot \vec{\nabla})\epsilon^m - \vec{\nabla} \cdot (\nu_\epsilon^m\vec{\nabla}\epsilon^m) + C_2\frac{(\epsilon^m)^2}{k^m} &= C_1\frac{\epsilon^m}{k^m}(P_k^m + G^m).
\end{aligned} \tag{5}$$

Decoupling and linearization:
We use a block Gauss-Seidel method for the iterative decoupling of (5). A second upper index denotes the iteration step. Furthermore we replace $\partial_t^m F$ by $\tilde{\partial}_t^m F := (F^{m,i} - F^{m-1})/\triangle_m$. Given $\vec{u}^{m,0}$, $p^{m,0}$, $\theta^{m,0}$, $k^{m,0}$, $\epsilon^{m,0}$ as the solutions of the previous time step, the algorithm reads:

(1) Initialization: Set $it_{dlc} \longleftarrow 1$.

(2) Set $i \longleftarrow it_{dlc}$ and update turbulent viscosity $\nu_t^m \longleftarrow \nu_t^m(k^{m,i-1}, \epsilon^{m,i-1})$. Update $U^*, \dot{q}_0$ according to (3),(4) using $\vec{u}^{m,i-1}$ and $\theta^{m,i-1}$.

(3) Update ν_e^m and solve the linearized Navier-Stokes-equation

$$\begin{aligned}
\tilde{\partial}_t^m\vec{u} + (\vec{u}^{m,i-1} \cdot \vec{\nabla})\vec{u}^{m,i} - \vec{\nabla} \cdot (2\nu_e^m S(\vec{u}^{m,i})) + \vec{\nabla}p^{m,i} &= -\beta\theta^{m,i-1}\vec{g} \\
\vec{\nabla} \cdot \vec{u}^{m,i} &= 0
\end{aligned}$$

(4) Update a_e^m and solve the θ-equation.

$$\tilde{\partial}_t^m \theta + (\vec{u}^{m,i} \cdot \vec{\nabla})\theta^{m,i} - \vec{\nabla} \cdot (a_e^m \vec{\nabla}\theta^{m,i}) \;=\; (\dot{q}^V)^m / c_p$$

(5) Update ν_k^m, P_k^m, G^m using $\vec{u}^{m,i}$, $\theta^{m,i}$ and solve the k-equation.

$$\tilde{\partial}_t^m k + (\vec{u}^{m,i} \cdot \vec{\nabla})k^{m,i} - \vec{\nabla} \cdot (\nu_k^m \vec{\nabla} k^{m,i}) = P_k^m + G^m - \epsilon^{m,i-1}$$

(6) Update P_k^m, G^m, ν_ϵ^m using $\vec{u}^{m,i}$, $\theta^{m,i}$, $k^{m,i}$ and solve the ϵ-equation.

$$\tilde{\partial}_t^m \epsilon + (\vec{u}^{m,i} \cdot \vec{\nabla})\epsilon^{m,i} - \vec{\nabla} \cdot (\nu_\epsilon^m \vec{\nabla}\epsilon^{m,i}) + C_2 \frac{\epsilon^{m,i-1}}{k^{m,i}} \epsilon^{m,i} = C_1 \frac{\epsilon^{m,i-1}}{k^{m,i}} (P_k^m + G^m)$$

(7) Stopping-criterion for linearization cycle : If $it_{dlc} < max_{dlc}$ and if stopping criteria for $\{\vec{u}^{m,i}\}_i$, $\{\theta^{m,i}\}_i$, $\{k^{m,i}\}_i$, $\{\epsilon^{m,i}\}_i$ are not yet fulfilled, then set $it_{dlc} \longleftarrow it_{dlc} + 1$ and goto (2). Otherwise goto next time step.

Linearized kernels:
The iterative scheme requires the solution of two basic model problems. First, the linearized equations for θ, k and ϵ are *advection-diffusion problems* with non-constant viscosity of the general form :

$$\left\{ \begin{array}{rcll} Lu \equiv -\vec{\nabla} \cdot (\nu \vec{\nabla} u) + (\vec{b} \cdot \vec{\nabla})u + cu &=& f & \text{in } \tilde{\Omega} \\ u &=& g & \text{on } \tilde{\Gamma}_D \\ \nu \vec{\nabla} u \cdot \vec{n} &=& h & \text{on } \tilde{\Gamma}_N . \end{array} \right. \tag{6}$$

For θ we set $\tilde{\Omega} = \Omega$, $\tilde{\Gamma}_D = \Gamma_-$, $\tilde{\Gamma}_N = \Gamma_0 \cup \Gamma_+$, $h|_{\Gamma_0} = \dot{q}_0 / c_p$, $h|_{\Gamma_+} = 0$. For k and ϵ set $\tilde{\Omega} = \Omega \backslash \Omega_\delta$, $\tilde{\Gamma}_D = (\Gamma_- \cap \partial\tilde{\Omega}) \cup \Gamma_\delta$ with appropriate g and $\tilde{\Gamma}_N = \Gamma_+$ with $h = 0$. The other data are given in the following table.

equation	u	ν	$\vec{b}$	cu	f
for θ	$\theta^{m,i}$	a_e^m	$\vec{u}^{m,i}$	$\theta^{m,i}/\triangle_m$	$\dot{q}^V/c_p + \theta^{m-1}/\triangle_m$
for k	$k^{m,i}$	ν_k^m	$\vec{u}^{m,i}$	$k^{m,i}/\triangle_m$	$(P_k^m + G^m) - \epsilon^{m,i-1}$ $+k^{m-1}/\triangle_m$
for ϵ	$\epsilon^{m,i}$	ν_ϵ^m	$\vec{u}^{m,i}$	$C_2 \frac{\epsilon^{m,i-1}}{k^{m,i}} \epsilon^{m,i}$ $+\epsilon^{m,i}/\triangle_m$	$C_1 \frac{\epsilon^{m,i-1}}{k^{m,i}} (P_k^m + G^m)$ $+\epsilon^{m-1}/\triangle_m$

Later on, we simply write Ω and omit the indices of viscosities and production terms.

The linearized Navier-Stokes-equation is an *Oseen*-type problem with a positive reaction term and non-constant viscosity:

$$\begin{array}{rcll} L_O(\vec{a}, \vec{u}, p) \equiv -\vec{\nabla} \cdot (2\nu S(\vec{u})) + (\vec{a} \cdot \vec{\nabla})\vec{u} + c\vec{u} + \vec{\nabla} p &=& \vec{f} & \text{in } \Omega \\ \vec{\nabla} \cdot \vec{u} &=& 0 & \text{in } \Omega \\ (\tau - pI)\vec{n} &=& \tau_n \vec{n} & \text{on } \Gamma_- \cup \Gamma_+ \\ (I - \vec{n} \otimes \vec{n})\tau\vec{n} = \vec{\tau}_t, \quad \vec{u} \cdot \vec{n} &=& 0 & \text{on } \Gamma_0 . \end{array} \tag{7}$$

Comparison with step (3) of the algorithm yields $\vec{u} = \vec{u}^{m,i}$, $\nu = \nu_e$, $\vec{a} = \vec{u}^{m,i-1}$, $c = \triangle_m^{-1}$, $p = p^{m,i}$, $\vec{f} = -\beta\theta^{m,i-1}\vec{g} + \triangle_m^{-1}\vec{u}^{m-1}$.

Stabilized finite element discretization of (6)-(7):
Assume an admissible triangulation $\mathcal{T}_h$ of the Lipschitz domain Ω and define finite element subspaces $X_h^l \equiv \{v \in C(\bar{\Omega}) \mid v|_K \in \Pi_l(K)\ \forall K \in \mathcal{T}_h\}$, $l \in \mathbf{N}$.
For the *advection-diffusion-reaction problem* (6), for simplicity with $g = 0$ on Γ_D, we apply the Galerkin-FEM with SUPG-stabilization:

$$\text{Find } u \in V_h = \{v \in X_h^l \mid v|_{\Gamma_D} = 0\} \text{ s.t.: } b^s(u,v) = l^s(v) \quad \forall v \in V_h\ , \tag{8}$$

$$\begin{aligned} b^s(u,v) &= \int_\Omega \Big(\nu\vec{\nabla}u\cdot\vec{\nabla}v + (\vec{b}\cdot\vec{\nabla})u\ v + cuv\Big)\,dx + \sum_{T\in\mathcal{T}_h}\int_T \delta_T Lu(\vec{b}\cdot\vec{\nabla})v\ dx \\ l^s(v) &= \int_\Omega fv\,dx + \int_{\Gamma_N} hv\,ds + \sum_{T\in\mathcal{T}_h}\int_T \delta_T f\ (\vec{b}\cdot\vec{\nabla})v\ dx \end{aligned}$$

with appropriate parameter set $\{\delta_T\}_T$, see [KLGR00]. The SUPG solutions may suffer from local crosswind oscillations in layers, hence negative values of k or ϵ can occur. As a remedy, we add in a consistent way crosswind diffusion thus leading to the (nonlinear) shock-capturing method, for details see [CS99].

For the *Oseen-problem* (7), we define the discrete spaces $\mathbf{V}_h \times Q_h = (X_h^r)^d \times X_h^s$ with $r, s \in \mathbf{N}$. The Galerkin FEM requires the (bi)linear forms

$$\mathcal{A}(U,V) = a(\vec{u},\vec{v}) + b(\vec{v},p) - b(\vec{u},q)\ , \qquad \mathcal{L}(V) = L(\vec{v}).$$

with $U = (\vec{u},p)$, $V = (\vec{v},q)$ and $b(\vec{v},p) = -\int_\Omega p(\vec{\nabla}\cdot\vec{v})\,dx$. Furthermore set

$$\begin{aligned} a(\vec{u},\vec{v}) &= \int_\Omega 2\nu S(\vec{u}) : \vec{\nabla}\vec{v} + ((\vec{a}\cdot\vec{\nabla})\vec{u} + c\vec{u})\cdot\vec{v}\,dx + \int_{\Gamma_0}(pI - \vec{n}\otimes\vec{n}\tau)\vec{n}\cdot\vec{v}\,ds \\ L(\vec{v}) &= \int_\Omega \vec{f}\cdot\vec{v}\,dx + \int_{\Gamma_-\cup\Gamma_+}\tau_n\vec{n}\cdot\vec{v}\,ds + \int_{\Gamma_0}\vec{\tau}_t\cdot\vec{v}\,ds. \end{aligned}$$

When using equal order ansatz functions $r = s$, the discrete Babuska-Brezzi condition is not satisfied. This problem is circumvented using a pressure (PSPG) stabilization. In addition, divergence and a SUPG stabilization is used to deal with dominating first order terms. More precisely, we set

$$\begin{aligned} \mathcal{A}^s(U,V) &= \mathcal{A}(U,V) + \sum_{T\in\mathcal{T}_h}\int_T \Big[L_O(\vec{a},\vec{u},p)\ \Big(\delta_{1u}^T\vec{a}\cdot\vec{\nabla})\vec{v} + \delta_{1p}^T\ \vec{\nabla}q\Big)\,dx \\ &\qquad + \int_T \delta_{2u}^T\ (\vec{\nabla}\cdot\vec{u})(\vec{\nabla}\cdot\vec{v})\ dx\Big] \\ \mathcal{L}^s(V) &= \mathcal{L}(V) + \sum_{T\in\mathcal{T}_h}\int_T \vec{f}\Big(\delta_{1u}^T(\vec{a}\cdot\vec{\nabla})\vec{v} + \delta_{1p}^T\vec{\nabla}q\Big)\ dx. \end{aligned}$$

Finally, the stabilized problem to the Oseen equation (7) reads

$$\text{Find } U = (\vec{u},p) \in \mathbf{V}_h \times Q_h, \text{ s.t. } \mathcal{A}^s(U,V) = \mathcal{L}^s(V)\ \forall V \in \mathbf{V}_h \times Q_h. \tag{9}$$

For the choice of the stabilization parameters $\delta_{1u}^T, \delta_{2u}^T$ and δ_{1p}^T see [KLGR00].

Domain decomposition of the linearized problems

Here we apply a nonoverlapping domain decomposition method with Robin interface conditions to the basic linearized problems (6), (7). Consider a nonoverlapping partition of Ω into convex, polyhedral subdomains being aligned with the finite element mesh, i.e.

$$\bar{\Omega} = \cup_{k=1}^{N} \bar{\Omega}_k, \quad \Omega_k \cap \Omega_j = \emptyset \quad \forall k \neq j\ , \quad \forall K \in \mathcal{T}_h\ \exists k\ :\ K \subset \Omega_k.$$

Furthermore, set $\Gamma_k := \partial\Omega_k \backslash \partial\Omega,\ \Gamma_{jk} := \partial\Omega_j \cap \partial\Omega_k,\ j \neq k$, where Γ_{kj} is identified with Γ_{jk}. Assume, for simplicity, that the partition is stripwise.

For the (continuous) *advection-diffusion-reaction* problem (6) the DDM reads:
for given u_k^n from iteration step n on each Ω_k, seek (in parallel) for u_k^{n+1}

$$\begin{cases} Lu_k^{n+1} = f & \text{in } \Omega_k \\ u_k^{n+1} = 0 & \text{on } \Gamma_D \cap \partial\Omega_k \\ \nu\vec{\nabla}u_k^{n+1} \cdot \vec{n}_k = h & \text{on } \Gamma_N \cap \partial\Omega_k \\ \Phi_k(u_k^{n+1}) = \theta\Phi_k(u_j^n) + (1-\theta)\Phi_k(u_k^n) & \text{on } \Gamma_{jk},\ j = 1, \ldots, N,\ j \neq k. \end{cases} \tag{10}$$

$\theta \in (0, 1]$ is a relaxation parameter. The interface function is specified as

$$\Phi_k(u) = \nu\vec{\nabla}u \cdot \vec{n}_k + (-\frac{1}{2}\vec{b} \cdot \vec{n}_k + z_k)u. \tag{11}$$

Let $V_{k,h}$, b_k^s and l_k^s denote the restrictions of V_h, b^s and l^s to Ω_k, respectively. $W_{kj,h}$ is the restriction of V_h to the interface part Γ_{kj}. Furthermore, $\langle\cdot,\cdot\rangle_{\Gamma_{kj}}$ is the inner product in $L^2(\Gamma_{kj})$ or, whenever needed, the dual product between $(W_{kj,h})^*$ and $W_{kj,h}$. The *fully discretized* DDM reads for $k = 1, \ldots, N$:

Parallel computation step : *Find $u_k^{n+1} \in V_{k,h}$ such that $\forall\ v_k \in V_{k,h}$*

$$b_k^s(u_k^{n+1}, v_k) + \langle(-\frac{1}{2}\vec{b} \cdot \vec{n}_k + z_k)u_k^{n+1}, v_k\rangle_{\Gamma_k} = l_k^s(v_k) + \sum_{j(\neq k)} \langle\Lambda_{jk}^n, v_k\rangle_{\Gamma_{kj}}.$$

Communication step : *For all $j \neq k$, update the Lagrangian multipliers*

$$\langle\Lambda_{kj}^{n+1}, \phi\rangle_{\Gamma_{kj}} = \langle\theta(z_k + z_j)u_k^{n+1} - \theta\Lambda_{jk}^n + (1-\theta)\Lambda_{kj}^n, \phi\rangle_{\Gamma_{kj}} \quad \forall\phi \in W_{kj,h}.$$

The analysis of the method, given in [LMO00], can be easily extended to the case of non-constant viscosity ν: The algorithm is well-posed if $z_k = z_j > 0$. The sequences $\{u_k^n\}_n$, $k = 1, ..., N$ converge strongly to the restrictions of the global discrete solution to Ω_k w.r.t. the stabilized energy norm induced by the symmetric part of $b_k^s(\cdot,\cdot)$.

Furthermore, an *a posteriori* estimate allows to control the convergence on subdomains via jumps of discrete DD solutions across the interface. Besides this estimate yields the following design of the interface function

$$z_k = \frac{1}{2}|\vec{b} \cdot \vec{n}_k| + R_k \tag{12}$$

with strictly positive $R_k = \mathcal{O}(\sqrt{\nu})$ depending on problem data. Formula (12) is compatible with the vanishing viscosity limit $\nu \to 0$. Moreover, it is shown in [LMO00] that (12) allows

a considerable acceleration of convergence. More precisely, the lower (and certain moderate) frequencies of the error are quickly damped. In this range, formula (12) is surprisingly sharp w.r.t. data. The convergence speed slows down when the level of the discretization error is reached. An acceleration of the method w.r.t. higher frequencies of the error is under consideration.

For the *Oseen problem* (7) we use the abbreviation $\pi_{t,k} := I - \vec{n}_k \otimes \vec{n}_k$. Then the DDM is defined as follows:
for given $(\vec{u}_k^n, p_k^n)$ *from step* n *on each* Ω_k, *seek (in parallel) for* $(\vec{u}_k^{n+1}, p_k^{n+1})$

$$\begin{cases} L_O(\vec{a}, \vec{u}_k^{n+1}, p_k^{n+1}) &= \vec{f} & \text{in } \Omega_k \\ \vec{\nabla} \cdot \vec{u}_k^{n+1} &= 0 & \text{in } \Omega_k \\ (\tau_k^{n+1} - p_k^{n+1} I)\vec{n}_k &= \tau_n \vec{n}_k & \text{on } \partial\Omega_k \cap (\Gamma_- \cup \Gamma_+) \\ \pi_{t,k}\tau_k^{n+1}\vec{n}_k &= \vec{\tau}_t, \quad -\vec{u}_k^{n+1} \cdot \vec{n}_k = 0 & \text{on } \partial\Omega_k \cap \Gamma_0 \\ \Phi_k(\vec{u}_k^{n+1}, p_k^{n+1}) &= \theta\Phi_k(\vec{u}_j^n, p_j^n) + (1-\theta)\Phi_k(\vec{u}_k^n, p_k^n) & \text{on } \Gamma_{jk}. \end{cases} \quad (13)$$

$\theta \in (0, 1]$ is again a relaxation parameter. The interface function is given by

$$\Phi_k(u, p) = \nu\vec{\nabla}\vec{u} \cdot \vec{n}_k - p\vec{n}_k + (-\frac{1}{2}\vec{a} \cdot \vec{n}_k + z_k)\vec{u} \quad (14)$$

with acceleration parameter z_k.
The corresponding parallel algorithm can be formulated (in weak form) similarly as for the scalar case. For this DD algorithm (and certain variants of it), a similar a-priori and a-posteriori analysis is available as briefly described for the scalar problem (6). In particular, the interface function z_k in (14) has the same structure as in (12). For details, we refer to [LMO01], [LMM00].

Application to room-air flow simulation

We applied our research code *Parallel NS* [Mue99] with piecewise linear ansatz functions for all unknowns ($l = r = s = 1$) on a triangular (resp. tetrahedral) mesh in 2D (resp 3D) to the numerical simulation of room-air flow.

Example 1. We present a stationary ventilated *laminar* flow with $Re = 66$, $Pr = 0.71$ through a cube $\Omega = (0, 1)^3$ with inlet zone $\Gamma_- = (0; 0.5) \times \{0\} \times (0; 0.5)$ and outlet zone $\Gamma_+ = (0.5; 1) \times \{1\} \times (0.5; 1)$, cf. Fig. 1. We impose $\tau_n = 0$, $\vec{u}|_{\Gamma_0} = \vec{0}$, $\theta(t = 0) = 293.15K$, $\theta_{in} = 283.15K$, $\theta|_{\Gamma_0} = 293.15K$. Furthermore, we used time step $\triangle_m = 1.0s$, a uniform mesh with 42^3 nodes and $max_{dlc} = 1$.

We studied the DDM on different macro partitions. Fig. 2 shows the reasonable convergence history (w.r.t. a mesh-dependent norm including H^1- and L^2-convergence of velocity and pressure, respectively) of the DD solution (with two subdomains) to the sequential discrete solution. ♣

Example 2. The application of the DDM to *turbulent* flows in 2D has been considered in [Mue99]. Here we present the natural convection for $Ra = 5.3 \cdot 10^{10}$, $Pr = 0.71$, $\dot{q}^V = 0$ in a cavity Ω of width $0.5m$ and height $H = 2.5m$. The flow is driven by a temperature difference of 45.8^o between the vertical walls and gravity. Further we impose $\vec{u}|_{\Gamma_0} = \vec{0}$ on $\Gamma_0 \equiv \partial\Omega$ and

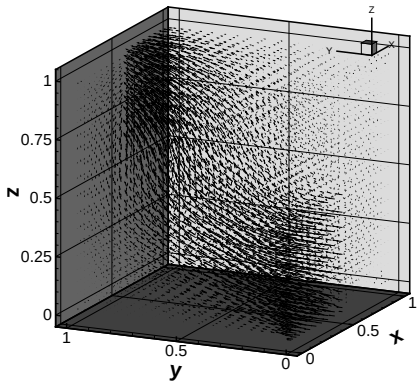


Figure 1: Laminar flow field in a ventilated room

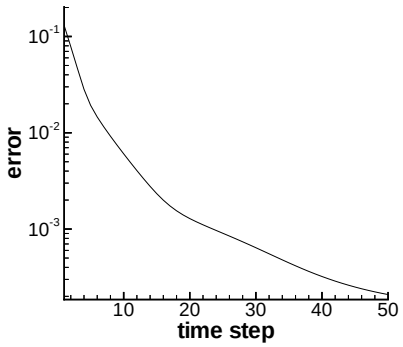


Figure 2: Convergence history of DD solution to discrete solution

adiabatic conditions for θ at the top and bottom of the flow domain Ω.

In Fig. 3 we compare the vertical velocity profiles at height $x_2 = 0.5H$ and $x_2 = 0.765H$

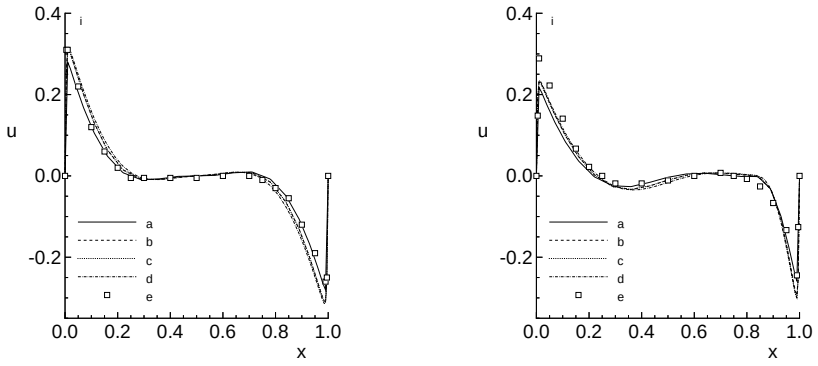


Figure 3: Vertical turbulent velocity at $x_2 = 0.5H$ and $x_2 = 0.765H$

for a DD solution with 10 subdomains and 3.432 finite elements (c) and the sequential solution on different grids with 2.728 (a), 3.432 (b) and 10.912 (d) elements. The results are in good agreement with measurements (e) by Cheesewright et.al (1986). ♣

The proposed method is currently applied at the Dresden University of Technology to the simulation of turbulent indoor air flows. Such calculations allow to predict certain parameters of the indoor-air climate over longer periods and to simulate different variants of ventilation or of heating systems. Results of this ongoing research will be presented elsewhere. Let us finally remark that the convergence of the method for the iteratively decoupled nonlinear problem (1) is rather sensitive w.r.t. different ingredients. A more robust implementation is probably given with an iterative substructuring method based on Dirichlet-Robin coupling, see [ATNV00], and with transformation to logarithmic variables in the $k - \epsilon$ equations.

References

[ATNV00]Y. Achdou, P. Le Tallec, F. Nataf, and M. Vidrascu. A domain decoposition preconditioner for an advection-diffusion problem. *Comp. Meth. Appl. Mech. Engng*, 184:145–170, 2000.

[CS99]Ramon Codina and Orlando Soto. Finite element implementations of two-equation and algebraic stress turbulence models for steady incompressible flow. *Intern. J. Numer. Meths. Fluids*, 90(3):309–334, 1999.

[KLGR00]Tobias Knopp, Gert Lube, Ralf Gritzki, and Markus Roesler. Simulation of indoor air movement using stabilized finite elements methods. Technical report, University of Goettingen, Math. Departm., 2000. accepted for Proc. Intern. Conf. FEM3D, Jyvaskyla 2000.

[LMM00]Gert Lube, Lars Mueller, and Hannes Mueller. A new non-overlapping domain decomposition method for stabilized finite element methods applied to the nonstationary Navier-Stokes equations. *Numer. Lin. Alg. Appl.*, 7:449–472, 2000.

[LMO00]Gert Lube, Lars Mueller, and Frank-Christian Otto. A non-overlapping domain decomposition method for the advection-diffusion problem. *Computing*, 64:49–68, 2000.

[LMO01]Gert Lube, Lars Mueller, and Frank-Christian Otto. A non-overlapping domain decomposition method for stabilized finite element approximations of the Oseen equations. *J. Comp. Appl. Math.*, 132:211–236, 2001.

[MP94]Bijan Mohammadi and Olivier Pironneau. *Analysis of the K-Epsilon Turbulence Model*. John Wiley and Sons, 1994.

[Mue99]Hannes Mueller. *A concept for the numerical simulation of incompressible flows using the discontinuous Galerkin method and a nonoverlapping DDM (in German)*. PhD thesis, University of Technology Dresden, 1999.

Thirteenth International Conference on Domain Decomposition Methods
Editors: N. Debit, M.Garbey, R. Hoppe, J. Périaux, D. Keyes, Y. Kuznetsov

53 Schur Complement Based Preconditioners for Compressible Flow Computations

Marzio Sala[1]

Introduction

The solution of linear systems arising from compressible flow computations is a great challenge in the field of scientific computing. Modern high-performance computers are very often organised as a distributed environment, and every efficient solver must account for their multiprocessor nature. Domain decomposition (DD) techniques provide a natural possibility to combine classical and well-tested single-processor algorithms with parallel new ones. The basic idea is to decompose the original computational domain Ω into M smaller parts, called subdomains $\Omega^{(i)}$, $i = 1, \ldots, M$, such that $\cup_{i=1}^{M} \bar{\Omega}^{(i)} = \bar{\Omega}$. Then we replace the global problem on Ω with M problems on $\Omega^{(i)}$. Of course, additional interface conditions must be provided.

The DD methods can roughly be classified into two groups [QV99, SBG96, CM94]. In the former, named after Schwarz, the computational domain is subdivided into overlapping subdomains, and local Dirichlet-type problems are then solved on each subdomain. The latter group, instead, uses non-overlapping subdomains. It is thus possible to decompose the unknowns into two sets: one formed by the unknowns on the interface between subdomains, and another formed by the unknowns associated to nodes internal to the subdomains. One may then compute a Schur complement (SC) matrix by "condensing" the unknowns in the second set. The system is then solved by first computing the interface unknowns and then solving the independent problems for the internal unknowns.

It can be shown [QV99] that the SC system is better conditioned than the global system. However, the solution of this system requires computing as many linear problems as the number of subdomains used. The dimension of these problems can be very large, unless the number of processors used is sufficiently high. A possible solution can be to solve the internal problems inexactly, using, for example, an incomplete factorisation [Saa96], or few steps of an iterative solver. The resulting approximate SC matrix can be seen as a preconditioner for the global system. Here we present some numerical results concerning the application of the SC matrix as a preconditioner for the global (unreduced) system. We have tested an elliptic problem, as well as a hyperbolic one. In the former the matrix arises from the Laplace operator, while in the latter from the compressible Euler equations.

This paper is organised as follows. Second section describes the SC system, showing two possible formulations, named element-oriented and vertex-oriented. Differences between the element-oriented and vertex-oriented SC matrix are here outlined. Third section describes the use of the SC system as a preconditioner for the global system. Numerical results for an elliptic test case and for the compressible Euler equations are reported in fourth section. The tests have been conducted on a distributed memory parallel machine. Conclusions are drawn in last section.

[1]Département de Mathématiques, EPF-Lausanne

The Schur complement Method

Let us consider the solution of the following linear system:

$$A\mathbf{u} = \mathbf{f}\,, \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ is a (sparse) real matrix, $\mathbf{u}$ and $\mathbf{f} \in \mathbb{R}^n$ two column vectors. In general, we can think of (1) as begin the algebraic counterpart of a variational boundary value problem which reads

$$\begin{aligned} &\text{find } u_h \in V_h \text{ such that:} \\ &a\left(u_h, v_h\right) = \left(f, v_h\right) \quad \text{for } \forall v_h \in V_h\,, \end{aligned}$$

where V_h is a finite dimensional space generated from finite element basis functions. For an elliptic problem $a(u_h, v_h)$ is bilinear form and u_h the discrete solution, while for the compressible Euler equations $a(u_h, v_h)$ should be regarded as the bilinear form expressing the Jacobian of the Euler system (after time and space discretisation), and u_h plays the role of the increment of the physical variables.

We now consider a partition of the domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, made in the following way. We first triangulate Ω and we indicate by $\mathcal{T}_h^{(\Omega)}$ the corresponding mesh. For the sake of simplicity we assume that the boundary of Ω coincides with the boundary of the triangulation. We then partition $\mathcal{T}_h^{(\Omega)}$ into 3 parts, namely $\mathcal{T}_h^{(1)}$, $\mathcal{T}_h^{(2)}$ and $\mathcal{T}_h^{(\Gamma)}$ such that $\mathcal{T}_h^{(1)} \cup \mathcal{T}_h^{(2)} \cup \mathcal{T}_h^{(\Gamma)} = \mathcal{T}_h^{(\Omega)}$. We may associate to $\mathcal{T}_h^{(1)}$ and $\mathcal{T}_h^{(2)}$ two disjoint subdomains $\Omega^{(1)}$ and $\Omega^{(2)}$ formed by the interior of the union of the elements of $\mathcal{T}_h^{(1)}$ and $\mathcal{T}_h^{(2)}$ respectively, while $\Gamma^{(1,2)}$ is formed by the "elements" contained on $\mathcal{T}_h^{(\Gamma)}$.

We will consider two cases:

- $\Gamma^{(1,2)}$ reduces to a finite number of disjoint measurable $d-1$ manifolds. This situation represents the common case where $\bar{\Omega}^{(1)} \cap \bar{\Omega}^{(2)} = \Gamma^{(1,2)}$, i.e. $\Gamma^{(1,2)}$ is the discretisation of the common part Γ of the boundary of $\Omega^{(1)}$ and $\Omega^{(2)}$. This type of decomposition will be called *element oriented* (EO) decomposition, because each element of $\mathcal{T}_h^{(i)}$, $i = 1, 2$ belongs exclusively to one of the two subdomains $\bar{\Omega}^{(1)}$ and $\bar{\Omega}^{(2)}$.

- $\Gamma^{(1,2)} \subset \mathbb{R}^d$ and it is formed by only one layer of elements. That is, each node of $\Gamma^{(1,2)}$ coincides with a node of either $\mathcal{T}_h^{(1)}$ or $\mathcal{T}_h^{(2)}$. The portion of space Γ of which $\Gamma^{(1,2)}$ is a triangulation, is now formed by the union of a finite number of "strips" laying between $\Omega^{(1)}$ and $\Omega^{(2)}$. It will be called *vertex oriented* (VO) decomposition, because each vertex belongs exclusively to one of the two subdomains $\bar{\Omega}^{(i)}, i = 1, 2$.

A node is said to be *internal* if it is not connected to any node of other subdomains, while a node that lies on $\Gamma^{(1,2)}$ is said to be a *border* node. In the following, we will consistently use the subscripts I and B to indicate internal and border nodes, respectively, while the superscript (i) will denote the domain which we are referring to.

Vertex Oriented Schur complement Matrix

Let us consider again problem (1). The block representation reads

$$A\mathbf{u} = \begin{pmatrix} A^{(1)} & \begin{matrix} 0 & 0 \\ 0 & E^{(1,2)} \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & E^{(2,1)} \end{matrix} & A^{(2)} \end{pmatrix} \begin{pmatrix} \mathbf{u}_I^{(1)} \\ \mathbf{u}_B^{(1)} \\ \mathbf{u}_I^{(2)} \\ \mathbf{u}_B^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_I^{(1)} \\ \mathbf{f}_B^{(1)} \\ \mathbf{f}_I^{(2)} \\ \mathbf{f}_B^{(2)} \end{pmatrix}, \tag{2}$$

where the submatrix $A^{(i)}$, relative to subdomain $\Omega^{(i)}$, can be written as

$$A^{(i)} = \begin{pmatrix} A_{II}^{(i)} & A_{IB}^{(i)} \\ A_{BI}^{(i)} & A_{BB}^{(i)} \end{pmatrix}.$$

In this partitioning the border nodes are subdivided into two sets: $B^{(1)}$ is the set of nodes of the triangulation of the strips $\Gamma^{(1,2)}$ which lay on the boundary of $\Omega^{(1)}$, while $B^{(2)}$ is that of nodes lying on the boundary of $\Omega^{(2)}$. Correspondingly, we have the blocks $\mathbf{u}_B^{(1)}$ and $\mathbf{u}_B^{(2)}$ in the vector of unknowns and $\mathbf{f}_B^{(1)}$ and $\mathbf{f}_B^{(2)}$ in the right hand side. $E^{(i,j)}$ represents the contribution to the equation associated to $B^{(i)}$ coming from the nodes in $B^{(i)}$. We call the nodes of $\Gamma^{(1,2)}$ contributing to $B^{(i)}$ *external nodes* of domain $\Omega^{(i)}$.
We can perform a LU elimination of internal nodes (which are coupled only to border nodes), obtaining the following Schur complement system:

$$S_{VO}\,\mathbf{u}_B = \begin{pmatrix} S^{(1)} & E^{(1,2)} \\ E^{(2,1)} & S^{(2)} \end{pmatrix} \begin{pmatrix} \mathbf{u}_B^{(1)} \\ \mathbf{u}_B^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{g}^{(1)} \\ \mathbf{g}^{(2)} \end{pmatrix}, \tag{3}$$

where

$$S^{(i)} = A_{BB}^{(i)} - A_{BI}^{(i)} {A_{II}^{(i)}}^{-1} A_{IB}^{(i)} \text{ and } \mathbf{g}^{(i)} = \mathbf{f}_B^{(i)} - \sum_{i=1}^{M} R_i^T A_{BI}^{(i)} {A_{II}^{(i)}}^{-1} \mathbf{f}_I^{(i)}, \tag{4}$$

with $i = 1, 2$. Note that S_{VO} is in general dense on the block diagonal, while the blocks $E^{(i,j)}$ are sparse. The technique just shown may be extended to any number of domains.
The Schur operators built on an element-oriented or a vertex-oriented DD are clearly different. The theoretical properties of the former are better known, since it has a more direct interpretation at differential level [QV99], while the latter is normally the result of a purely algebraic approach. Although the element-oriented decomposition has better theoretical foundations, the literature for complex CFD computations refers more frequently to vertex-oriented decomposition. In fact, the vertex-oriented approach is more simple derived by purely algebraic manipulations on the original system matrix.
Although the SC matrix is better conditioned that the unreduced matrix A, a suitable preconditioner has to be found. Many methods have been proposed in literature for the EO decomposition; see for example [QV99, SBG96] for an overview. Among them, we recall the balancing Neumann/Neumann, the wire-basket preconditioners FETI [FPL00] and others [CGT98]. These methods couple a local preconditioner with a coarse correction to avoid the degradation of the performance as the number of subdomains grows. Another possible way to derive a preconditioner for equation (3) is to use using the relation $S_{VO}^{-1} = R_B A^{-1} R_B^T$,

where R_B the restriction operator on the interface variables. It follows that form that from *any* preconditioner P_A for the matrix A one can obtain a preconditioner P_S for the matrix S_{VO}. The preconditioning operation for S_{VO} which is induced from P_A is defined by

$$P_S^{-1}\mathbf{v}_B = R_B P_A^{-1} \begin{pmatrix} 0 \\ \mathbf{v}_B \end{pmatrix} = R_B P_A^{-1} R_B^T \mathbf{v}_B \ .$$

For example, a Schwarz-type preconditioner can be used for the solution of the vertex-oriented SC matrix. In fact, we recall that the SC matrix obtained from a vertex-oriented decomposition is less dense than the one obtained from an element-oriented one. As one may note from equation (3), S_{VO} has dense diagonal blocks, while the non-diagonal blocks are sparse.

The Schur Complement System as a Preconditioner

The bottleneck of the SC system is the solution of the internal problems. This step can be done in parallel, however it can be very expensive for both memory requirement and time. Direct solvers can be used only with small problems, while iterative solvers need to be preconditioned.
Let us write the matrix A in the following block form, putting before all the internal nodes, followed by all border nodes:

$$A = \begin{pmatrix} A_{II} & 0 \\ A_{BI} & I \end{pmatrix} \begin{pmatrix} I & A_{II}^{-1} A_{IB} \\ 0 & S \end{pmatrix} \ . \tag{5}$$

A possible preconditioner is

$$P_{ASC} = \begin{pmatrix} \tilde{A}_{II} & 0 \\ A_{BI} & I \end{pmatrix} \begin{pmatrix} I & \tilde{A}_{II}^{-1} A_{IB} \\ 0 & \tilde{S} \end{pmatrix} \ , \tag{6}$$

where $\tilde{A}_{II}$ is, for example, an ILU decomposition of A_{II}, and $\tilde{S}$ is given, for instance, by few steps of an iterative method where in the *global* Schur Complement the internal Dirichlet problems are solved (approximately) using $\tilde{A}_{II}$. To apply P_{ASC} to a vector, we need to solve some local linear systems with the matrices A_{II} and a *global* linear system with $\tilde{S}$. In this case, the role of $\tilde{S}$ is to couple all the subdomains. In this way, we may avoid the definition of a coarse space. In fact, the definition of the coarse problem may be difficult when dealing with complex geometries or non-matching grids [CSZ96], especially for the choice of the boundary conditions. On the contrary, the definition of the ASC preconditioner is purely algebraic and it can be easily applied to any kind of matrices (provided that the incomplete factorization of A_{II} exists). This approach is similar to the one followed in [Zha00] and other papers with the same aim to construct the preconditioner without dealing with the geometrical data of the underline physical problem.

Numerical Implementation

In first subsection we present some numerical results concerning a Laplace operator, while in second subsection we apply the SC preconditioner to the solution of the compressible Euler

equations. All the numerical results here presented have been obtained on a SGI-Cray machine located at the EPFL, with 32 MIPS R14000 processors, each of them has 256Kbytes of local memory, 32 Kbytes of first level cache and 4 Mbytes of second level cache. For the solution of the linear system, we have used the AZTEC library, developed at the Sandia National Laboratories. The linear solver used is GMRESR, a variant of GMRES that allows the preconditioner to be different at each iteration. We have stopped the solver after a reduction of 10^{-5} of the initial residual. Each processor is given a single subdomain, and the MPI communicator has been used. About the Schwarz preconditioner, we have solved the local problem using an ILU decomposition. For the ASC preconditioner, in the solution of the linear system with $\tilde{S}$, we have used GMRES. The approximation is obtained replacing the exact LU of A_{II} decomposition the an incomplete factorisation $ILU(0)$.

An Elliptic Problem: the Laplace Operator

We have considered the following linear problem:

$$\left\{ \begin{array}{rcl} -\Delta u & = & f \text{ in } \Omega \\ u & = & g \text{ on } \partial\Omega \, , \end{array} \right. \tag{7}$$

where $\Omega = (0,1) \times (0,1) \times (0,1)$ is discretized by piece-linear finite elements on tetrahedra regular grids. For this simple test case we have partitioned the domain into slices, using a vertex-oriented decomposition as indicated in Section 53. In Table 1 we have reported the iterations to converge using 4 and 8 subdomains for different values of the numbers of the unknowns. We indicate with np the non preconditioned case, sw1 the Schwarz preconditioner with an overlap of 1 element, sw2 with an overlap of 2 elements. ASC-L represents the ASC preconditioner, with L steps of the nested iterative solver.
One may note that the ASC preconditioner behaves better than the 1-level Schwarz preconditioner for suitable values of L. This value can be increased to improve the efficacy of the ASC preconditioner. Moreover, as the number of subdomains grows, the ASC preconditioner requires less iterations to converge.

A Hyperbolic Problem: The Euler equations

Let us consider the Euler equations for compressible flows, that can be written in the following form:

$$\frac{\partial \mathbf{U}}{\partial t} + \sum_{j=1}^{d} \frac{\partial \mathbf{F}_j}{\partial x_j} = 0 \quad \text{in } \Omega \, , \, t > 0 \, , \tag{8}$$

(plus suitable boundary conditions on $\partial\Omega$), where $\mathbf{U}$ and $\mathbf{F_j}$ are, respectively, the vector of conservative variables and the flux vector:

$$\mathbf{U} = \begin{pmatrix} \rho \\ \rho u_i \\ \rho E \end{pmatrix} \quad , \quad \mathbf{F}_j = \begin{pmatrix} \rho u_j \\ \rho u_i u_j + p\delta_{ij} \\ \rho H u_j \end{pmatrix} \, ,$$

with $i = 1, \ldots, d$ ($\Omega \subset \mathbb{R}^d$), and $\mathbf{u}$ is the velocity vector, ρ the density, p the pressure, H the specific enthalpy and δ_{ij} the Kronecker symbol.
The spatial discretisation applied to the Euler equations leads eventually to a system of ODE in time, which may be written as $dU/dt = R(U)$, where $U = (U_1, U_2, \ldots, U_l, \ldots)^T$ is the

N_unks	np	sw1	sw2	ASC-2	ASC-5	ASC-10
4 processor						
64.000	116	38	33	35	28	24
125.000	154	46	39	43	36	32
216.000	227	54	45	50	42	39
512.000	253	66	58	66	56	51
1.000.000	454	91	66	81	67	64
8 processor						
64.000	116	43	35	32	22	14
125.000	154	51	40	38	29	20
216.000	227	56	48	45	37	27
512.000	253	68	59	60	49	40
1.000.000	454	105	67	73	64	54

Table 1: 3D Laplace problem. Number of iterations with 4 and 8 processors.

vector of unknown nodal states $U_i = U_i(t)$ and $R(U)$ the result of the spatial discretization of the Euler fluxes. Applying the backward Euler method to the semi-discrete equation yields

$$U^{n+1} - U^n = \Delta t R\left(U^{n+1}\right) \ , \tag{9}$$

where Δt here represents a *diagonal matrix* of local time steps, since we are interested only in steady-state solutions. We adopt the so-called "local time stepping" technique, where the degrees of freedom associated to each node evolve with their own time step. This is a rather common technique to accelerate convergence to steady state. In order to solve the nonlinear problem (9), the Newton method is used. We refer to the literature for more detailed explanations (see, for example [BCT98, SLW98, KKS98]).

For its spatial discretization, we have used the code THOR, developed at the von Karman Institute, that makes use of the multidimensional upwind finite element discretization, while for the vertex-oriented decomposition of the computational domain we have used the software METIS. This decomposition is unstructured and the subdomains have no particular shape, as one may appreciate from the picture on the left of Figure 1.

The first test case is represented by a NACA0012 airfoil with one degree of angle of attack. The free-stream Mach number is 0.85. 41 time iterations were required to reach the convergence to the steady state. The CFL number goes from 10 to 10^5, and it is multiplied on each iteration by 2.

Tables 2 reports a comparison between the Schwarz and the ASC preconditioners using from 4 to 32 processors. As we can observe, the gain in terms of number of iterations using the ASC preconditioner can be very high especially as the number of subdomains grows. Although, the time is (slightly) bigger that the one needed by the Schwarz preconditioners.

The second test case correspond to the solution of the compressible Euler equations around an ONERA M6 wing. The 3D unstructured grid has 94493 nodes and 555514 elements. The free-stream Mach number is 0.84, and the angle of attack is 3.06. The CFL number goes from 10 to 10^6, multiplied by 2 at each time iteration. In Table 3 we have reported the CPU time required to reach the steady state. As we can observe, few iterations in the solution of $\tilde{S}$ seems to be appropriate to reach the prescribed accuracy.

N_procs	sw1	sw2	ASC-2	ASC-5	sw1	sw2	ASC-2	ASC-5
	Iterations				CPU-time			
4	487	454	370	369	124.8	162.7	312.7	276.0
8	507	458	357	351	56.6	65.4	125.8	165.0
16	544	488	329	317	36.1	40.5	60.2	66.8
32	587	502	311	278	21.3	24.8	29.2	42.0

Table 2: Iterations and total time (in seconds) required to reach the convergence. NACA0012 airfoil, 9239 nodes.

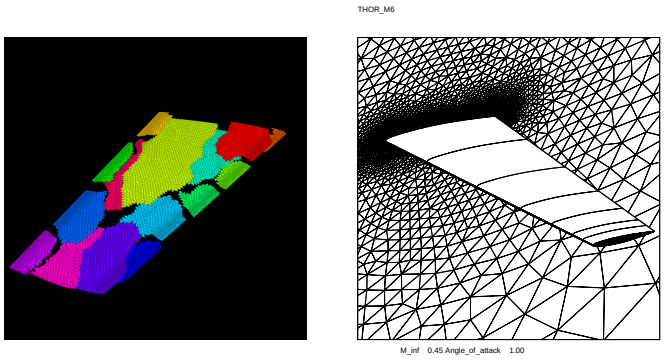


Figure 1: M6 Wing. Decomposition of the elements on the surface among subdomains (left) and particular of the unstructured 3D grid (right).

N_procs	ASC-2	ASC-4	ASC-8
8	**1538.4**	1600.4	1859.9
16	**544.8**	569.1	1330.5
32	**248.5**	286.0	358.9

Table 3: M6 wing, 94K nodes. CPU-time (in seconds) for ASC preconditioner, using different values of L.

Conclusion

In this paper, a preconditioner based on an approximation of the Schur complement system has been described. Numerical results have been presented for an elliptic problem and for the solution of the compressible Euler equations on 2D and 3D unstructured grids. The key idea is to use an approximate Schur complement matrix to precondition the unreduced matrix, exploiting the good parallel properties of the SC matrix. The approximate system is then solved by an iterative Krylov method, that couples all the subdomains.
The use of the SC system as a preconditioner for the global system seems promising, especially when the number of subdomains is large enough. The number of iterations to converge needed by the outer iterative solver is much lower than using a Schwarz preconditioner. Moreover, the effectiveness of the ASC preconditioner increases with the number of subdomains, even without a coarse operator, dislike the Schwarz method. Further numerical tests will be conducted to better investigate the parallel properties of this preconditioner.

References

[BCT98]T. J. Barth, T. F. Chan, and W. Tang. A parallel non-overlapping domain decomposition algorithm for compressible fluid flow problems on triangulated domains. In J. Mandel, C. Farhat, and X.-C. Cai, editors, *Tenth International Conference on Domain Decomposition Methods*, pages 23–41. AMS, Contemporary Mathematics 218, 1998.

[CGT98]Luiz Carvalho, Luc Giraud, and Patrick Le Tallec. Algebraic two-level preconditioners for the schur complement method. Technical report tr/pa/98/18, CERFACS, Toulouse, France, 1998. Preliminary version of the paper to appear in SIAM Journal on Scientific Computing.

[CM94]Tony F. Chan and Tarek P. Mathew. Domain decomposition algorithms. In *Acta Numerica 1994*, pages 61–143. Cambridge University Press, 1994.

[CSZ96]Tony F. Chan, Barry F. Smith, and Jun Zou. Overlapping Schwarz methods on unstructured meshes using non-matching coarse grids. *Numer. Math.*, 73(2):149–167, 1996.

[FPL00]C. Farhat, K. H. Pierson, and M. Lesoinne. The second generation of feti methods and their application to the parallel solution of large-scale linear and geometrically nonlinear structural analysis problems. *Computer Methods in Applied Mechanics and Engineering*, 184:333–374, 2000.

[KKS98]D. K. Kaushik, D. E. Keyes, and B. F. Smith. NKS methods for compressible and incompressible flows on unstructured grids. *Proceedings of the 11th Intl. Conf. on Domain Decomposition Methods*, pages 513–520, 1998.

[QV99]Alfio Quarteroni and Alberto Valli. *Domain Decomposition Methods for Partial Differential Equations*. Oxford Science Publications, 1999.

[Saa96]Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, 1996.

[SBG96]Barry F. Smith, Petter E. Bjørstad, and William Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

[SLW98]G. De Spiegeleer, A. Lerat, and Z. Wu. Implicit multidomain computation of compressible flows with a large number of subdomains. *Computational Fluid Dynamics Journal*, 7, 1998.

[Zha00]Jun Zhang. Preconditioned krylov subspace methods for solving nonsymmetric matrices. *Computer Methods in Applied Mechanics and Engineering*, 189(3):825–840, 2000.