Schwarz Methods for the Time-Parallel Solution of Parabolic Control Problems

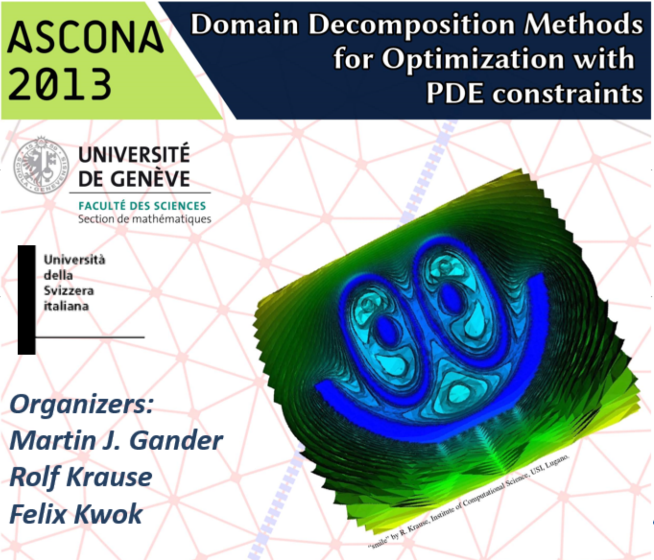
Felix Kwok

Hong Kong Baptist University

DD23, Jeju, South Korea
July 8, 2015

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

Two weeks before DD22 . . .




SWITZERLAND
© 2009 Ezilon.com All Right Reserved
Geneva
Ascona
Workshop
Lugano
DD22
FRANCE
GERMANY
AUSTRIA
LIECHTENSTEIN
ITALY

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

Collaborators

- Martin J. Gander (Geneva)
- Laurence Halpern (Paris-Nord)
- Bankim C. Mandal (Michigan Technological Univ.)
- Julien Salomon (Paris-Dauphine)

What is control?

According to Glowinski & Lions[1],

> *"At* **a given time horizon** *we want the system under study to behave* **exactly** *as we wish (or in a manner arbitrarily close to it)."*

[1] R. Glowinski & J.L. Lions, Exact and approximate controllability for distributed parameter systems, *Acta Numerica*, 1994.

Optimal Control

- Ingredients:
  1. A *system* governed by an ODE/PDE (or a system thereof),
  2. A control function that is an input to the system,
  3. A target state at the end of the time horizon, and/or
  4. A cost functional (e.g., energy of the input, deviation from expected trajectory, etc.)
- Goal:
  - Find the control of **minimal cost** such that the system reaches the desired state at the end of the horizon.

Example 1: Contaminant Tracking

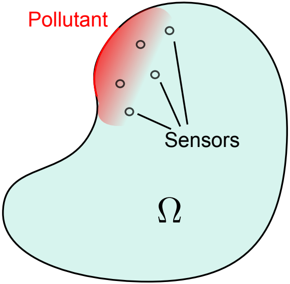
Problem: Calculate the rate $u(x,t)$ of pollution seepage

$$\min_{u\in\mathscr{U}} \frac{1}{2}\int_0^T \|u\|^2\,dt + \frac{1}{2}\int_0^T \|Cy(t;u) - \hat{y}\|^2\,dt$$

subject to

$$\dot{y} - \nabla\cdot(c\nabla y + \mathbf{b}y) = Bu \qquad \text{in } \Omega$$

+ initial and boundary conditions.




BAD NEWS

France : une usine autorisée à rejeter ses produits chimiques toxiques en mer

Publié Le 10 Septembre 2014 à 17h07

Le parc national des Calanques a autorisé ce lundi 8 septembre l'usine Altéo située non loin de Marseille à continuer à rejeter en mer ses eaux industrielles toxiques, chargées d'aluminium, de fer et d'arsenic.

© svedoliver

France : l'usine Alteo autorisée à rejeter ses "boues rouges" chimiques en mer

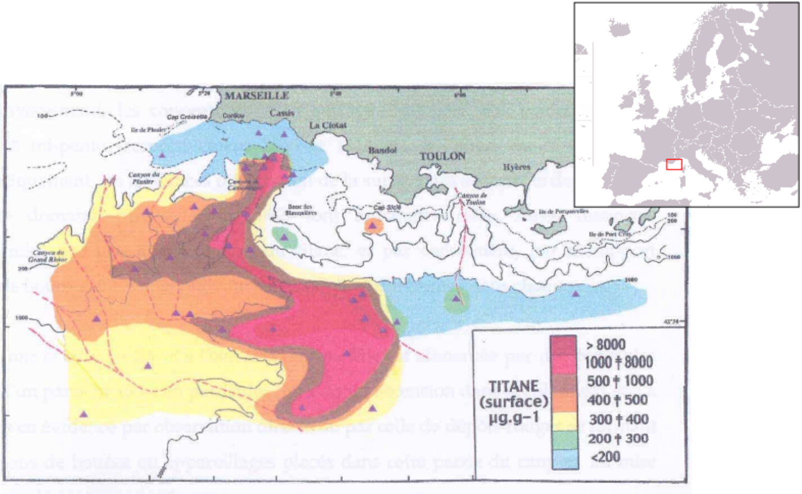
SUR LE MÊME THÈME

Une usine autorisée à polluer... un Parc naturel protégé

Non vous ne rêvez pas ! Le Conseil d'administration du Parc national des Calanques a autorisé (30 voix pour, 16 contre, 2 absentions) l'usine Alteo (ex-Péchiney), qui produit à Gardanne (Bouches-du-Rhône) de l'alumine à partir de bauxite, l'autorisation de continuer à rejeter en mer des "eaux de procédé" chargées de métaux lourds dont l'aluminium, le "fer total" et l'arsenic - qui dépassent en plus les seuils légaux de toxicité - pour encore 30 ans !

Ces rejets s'accompagneront de "meilleurs contrôles et d'un meilleur suivi des eaux rejetées", a essayé de rassurer le président du Parc...

Il faut noter que le Parc national des Calanques a été créé en avril

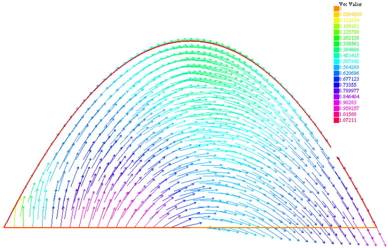


(Source: `http://www.robindesbois.org/dossiers/boues_rouges/alcan-gardanne.jpg`)

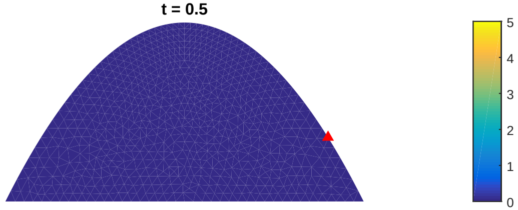
Example 1: Contaminant Tracking

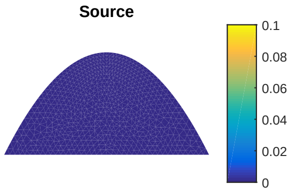
Find source term u that best match observation subject to the advection-diffusion equation

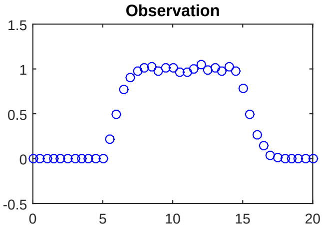
$$\frac{\partial c}{\partial t} + \nabla \cdot (\mathbf{v}c - \nu \nabla c) = u$$



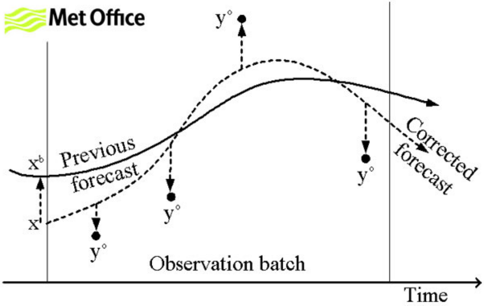
Example 1: Contaminant Tracking







Example 2: Data Assimilation in Weather Forecasting



(Source: UK Met Office)

- Forecast models contain errors and uncertainties
- Must correct initial conditions based on new measurements

Example 2: Data Assimilation in Weather Forecasting

- 4D-Var Model: minimize

$$J(y_0) = \|y_0 - y_0^{\text{model}}\|_D^2 + \sum_{k=1}^{K} \|Gy(t_k) - z_k\|_R^2$$

  subject to

$$F(y, \dot{y}) = 0, \qquad y(0) = y_0,$$

  where $z_k =$ observations, $y_0^{\text{model}} =$ initial conditions before data assimilation

- Used in weather models in ECMWF, France, UK, Japan, Canada,. . .

Example 2: Data Assimilation in Weather Forecasting

From the UK Met Office website[2]:

> *4D-Var is used at many operational centres as well as the Met Office. However, future computers will have an* **increasingly parallel architecture***, and ensemble methods, which can fully exploit this, will become more attractive. It is therefore necessary to establish the full potential of 4D-Var in order to see whether it will remain the preferred operational method in the future.*

[2] http://www.metoffice.gov.uk/research/areas/data-assimilation-and-ensembles/4d-var-research

Other Applications

- Aeronautics: Aircraft design for reduction of noise due to boundary layer separation (He-Glowinski-Metcalfe-Periaux 1998, Dandois 2007, Borel-Halpern-Ryan 2010, . . .)
- Bio-medicine: Drug administration in chemotherapy (Jackson & Byrne 2000, Rockne et al. 2010, Corwin et al. 2013,. . .)
- Oil & Gas: Oil field management optimization, data assimilation, history matching (Smart Field Consortium at Stanford, . . .)
- . . .

Model Problem

- We want to solve the semi-discretized *linear quadratic optimal control problem*

$$\min \frac{1}{2}\int_0^T \|Cy(t;u)-\hat{y}\|^2 + \frac{\gamma}{2}\|Dy(T;u)-y_T\|^2 + \frac{\nu}{2}\int_0^T \|u\|^2$$

subject to the **linear** parabolic PDE constraint

$$\dot{y} + Ay = Bu, \qquad y(x,0) = y_0$$

with $\langle Ay, y\rangle \geq 0$ for all $y \in V$.

- Existence and uniqueness results: Lions (1968), Glowinski-Lions (2004/05), ...

Optimization

- We seek
$$\min J(y,u)$$
subject to the PDE constraint
$$\dot{y} + Ay = Bu, \qquad y(x,0) = y_0.$$
- Derive first-order optimality conditions formally using Lagrange multipliers λ:
$$L(y,\lambda,u) = J(y,u) + \langle \lambda, \dot{y} + Ay - Bu \rangle.$$
- We choose the inner product $\langle u, v \rangle = \int_0^T u^T v \, dt$.

Optimality System

- Since the optimal solution is a stationary point of $L(y, \lambda, u)$, we have

$$\frac{\partial}{\partial \epsilon} L(y + \epsilon z, \lambda, u) = 0 \qquad \text{for all } z \in V,$$

which gives

$$0 = \langle Cy - \hat{y}, Cz \rangle + \gamma (Dy(T) - y_T, Dz(T)) + \int_0^T (\lambda, \dot{z} + Az)\, dt.$$

- Integration by parts gives

$$0 = \langle C^T(Cy - \hat{y}), z \rangle + \gamma (D^T(Dy(T) - y_T), z(T))$$
$$+ (\lambda(T), z(T)) - (\lambda(0), z(0)) + \int_0^T (-\dot{\lambda} + A^T \lambda, z)\, dt.$$

Optimality System

$$0 = \langle C^T(Cy - \hat{y}), z\rangle + \gamma(D^T(Dy(T) - y_T), z(T)) + (\lambda(T), z(T)) - \underbrace{(\lambda(0), z(0))}_{=0} + \int_0^T (-\dot{\lambda} + A^T\lambda, z)\,dt.$$

- This equation must be satisfied for all z with $z(0) = 0$, so we get the *adjoint problem*

$$\dot{\lambda} - A^T\lambda = C^T(Cy - \hat{y}) \qquad \text{on } (0, T),$$
$$\lambda(T) = -\gamma D^T(Dy(T) - y_T).$$

- Taking a variation with respect to u gives the algebraic constraint $u = \nu^{-1}B^T\lambda$.

Optimality System

- First order optimality system (using Lagrange multipliers):

$$\begin{cases} \dot{y} + Ay = \nu^{-1} B^T \lambda, \\ y(0) = y_0, \end{cases} \qquad \begin{cases} \dot{\lambda} - A^T \lambda = C^T (Cy - \hat{y}), \\ \lambda(T) = -\gamma D^T (Dy(T) - \hat{y}(T)), \end{cases}$$

Forward problem — Adjoint problem

- "Optimize-then-discretize" approach
- Coupled two-point boundary value problem!

Discretization

- Discretize-then-Optimize:
  - Use finite volumes/FEM/etc. to discretize in space
  - Discretize state equation and cost function in time
- Solution satisfies a KKT system of the form

$$\begin{bmatrix} \mathcal{M}_y & & -\mathcal{K}^T \\ & \mathcal{M}_u & \mathcal{N}^T \\ -\mathcal{K} & \mathcal{N} & 0 \end{bmatrix} \begin{pmatrix} y \\ u \\ \lambda \end{pmatrix} = \begin{pmatrix} f_1 \\ 0 \\ g \end{pmatrix}$$

- Huge linear system, $N_x \times N_t$ unknowns!
- Note: Discretization and optimization do not always commute, see Dontchev, Hager & Veliov (2000)

Algorithms for time-dependent control

1. Conjugate Gradient and descent methods:
   - Use u as primary variables and solve with PCG

$$(\mathcal{M}_u + \mathcal{N}^T \mathcal{K}^{-T} \mathcal{M}_y \mathcal{K}^{-1} \mathcal{N})u = \tilde{f}.$$

   - 4D-Var uses physics-based preconditioning
   - One matrix-vector multipliciation requires one forward solve and one backward solve
   - Equivalent to a shooting method

Algorithms for time-dependent control

2. All-at-once approach: directly precondition

$$\begin{bmatrix} \mathcal{M}_y & & -\mathcal{K}^T \\ & \mathcal{M}_u & \mathcal{N}^T \\ -\mathcal{K} & \mathcal{N} & 0 \end{bmatrix} \quad \text{by} \quad \begin{bmatrix} \hat{\mathcal{M}}_y & & \\ & \hat{\mathcal{M}}_u & \\ & & \hat{\mathcal{S}} \end{bmatrix}$$

where $\hat{\mathcal{S}}$ approximates the Schur complement

- Effective for heat equation, time-dependent Stokes, . . .
- Rees, Stoll & Wathen (2010), Pearson, Stoll & Wathen (2012)

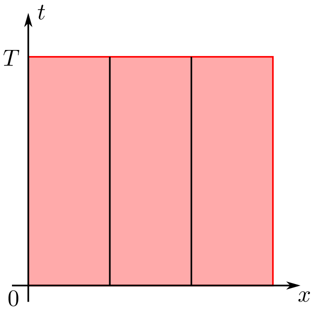
Dealing with storage

- Reduced basis techniques (Noor and Peters 1980, Quarteroni et al. 1997, Dedè 2008, Simoncini 2012, ...)
- Snapshot and windowing techniques (Griewank 1992, Berggren 1998, Restrepo et al. 1998)
- Distributed/parallel algorithms

Parallelization

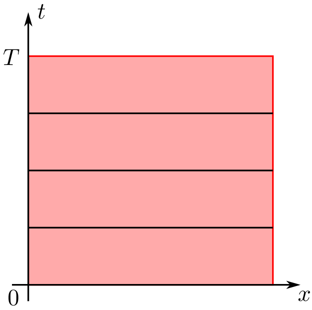
- Multigrid approaches, e.g. Borzì (2003) for parabolic problems
- Parallelize solution of forward and adjoint problems
  - In space
  - In time (Parareal, Lions and Maday (2001))
  - Waveform relaxation (Gander & Stuart 1998, Giladi & Keller 2002)
  - But: does not take the structure of control problems into account

Decomposition in space



- Heinkenschloss & Herty (2007): NNWR for parabolic control problems
- Gander and Mandal (2014)

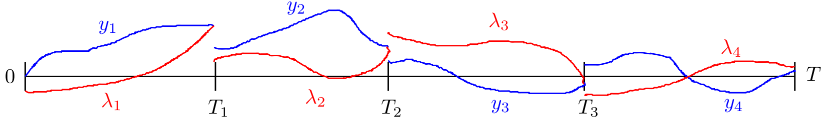
Decomposition in time



- Lagnese & Leugering (2003): Wave equation
- Heinkenschloss (2005): Parabolic problems
- Barker & Stoll (2013): Heat and Stokes equations

Multiple Shooting-based Preconditioning (Heinkenschloss (2005)

- Decompose $(0, T)$ into non-overlapping subintervals
- Define intermediate states $\tilde{y}_i$ and $\tilde{\lambda}_i$ at interfaces
- Minimize the sum of local objective functions, with locally optimal paths
- Path is globally optimal if there are no jumps in y



Multiple Shooting

For the block partition $v_i^T = (y_i^T, u_i^T, \lambda_i^T)$, the matrix becomes

$$
\mathscr{M} = \left(\begin{array}{cc|ccc|ccc|ccc|ccc|c}
-\rho\bar{B}_0^* & \bar{S}_0^\rho & 0 & 0 & \bar{B}_0^* & & & & & & & & & & \\
-I & \bar{B}_0 & 0 & 0 & 0 & & & & & & & & & & \\
\hline
\bar{Q}_1^{\rho,\delta} & 0 & -\rho\bar{A}_1^* & \bar{R}_1^\rho & -I & 0 & 0 & \bar{A}_1^* & & & & & & & \\
(\bar{R}_1^\rho)^* & 0 & -\rho\bar{B}_1^* & \bar{S}_1^\rho & 0 & 0 & 0 & \bar{B}_1^* & & & & & & & \\
\bar{A}_1 & 0 & -I & \bar{B}_1 & 0 & 0 & 0 & 0 & & & & & & & \\
\hline
-\delta\bar{A}_1 & 0 & \bar{Q}_2^{\rho,\delta} & -\delta\bar{B}_1 & 0 & -\rho\bar{A}_2^* & \bar{R}_2^\rho & -I & 0 & 0 & \bar{A}_2^* & & & & \\
0 & 0 & (\bar{R}_2^\rho)^* & 0 & 0 & -\rho\bar{B}_2^* & \bar{S}_2^\rho & 0 & 0 & 0 & \bar{B}_2^* & & & & \\
0 & 0 & \bar{A}_2 & 0 & 0 & -I & \bar{B}_2 & 0 & 0 & 0 & 0 & & & & \\
\hline
& & & & & & & & \ddots & & & & & & \\
& & & & & & & & & \ddots & & & & & \\
& & & & & & & & & & \ddots & & & & \\
\hline
& & & & & -\delta\bar{A}_{N_t-2} & 0 & 0 & \bar{Q}_{N_t-1}^{\rho,\delta} & -\delta\bar{B}_{N_t-2} & 0 & -\rho\bar{A}_{N_t-1}^* & \bar{R}_{N_t-1}^\rho & -I & \bar{A}_{N_t-1}^* \\
& & & & & 0 & 0 & 0 & (\bar{R}_{N_t-1}^\rho)^* & 0 & 0 & -\rho\bar{A}_{N_t-1}^* & \bar{S}_{N_t-1}^\rho & 0 & \bar{B}_{N_t-1}^* \\
& & & & & 0 & 0 & 0 & \bar{A}_{N_t-1} & 0 & 0 & -I & \bar{B}_{N_t-1} & 0 & 0 \\
\hline
& & & & & & & & -\delta\bar{A}_{N_t-1} & 0 & 0 & \bar{Q}_{N_t} + \delta I & -\delta\bar{B}_{N_t-1} & 0 & -I
\end{array}\right).
$$

Multiple Shooting

Heinkenschloss proposed using block Symmetric Gauss–Seidel + GMRES

$$
\mathscr{M} = \left(\begin{array}{cc|ccc|ccc|ccc|ccc|c}
-\rho\bar{B}_0^* & \bar{S}_0^\rho & 0 & 0 & \bar{B}_0^* & & & & & & & & & & \\
-I & \bar{B}_0 & 0 & 0 & 0 & & & & & & & & & & \\
\hline
\bar{Q}_1^{\rho,\delta} & 0 & -\rho\bar{A}_1^* & \bar{R}_1^\rho & -I & 0 & 0 & \bar{A}_1^* & & & & & & & \\
(\bar{R}_1^\rho)^* & 0 & -\rho\bar{B}_1^* & \bar{S}_1^\rho & 0 & 0 & 0 & \bar{B}_1^* & & & & & & & \\
\bar{A}_1 & 0 & -I & \bar{B}_1 & 0 & 0 & 0 & 0 & & & & & & & \\
\hline
-\delta\bar{A}_1 & 0 & \bar{Q}_2^{\rho,\delta} & -\delta\bar{B}_1 & 0 & -\rho\bar{A}_2^* & \bar{R}_2^\rho & -I & 0 & 0 & \bar{A}_2^* & & & & \\
0 & 0 & (\bar{R}_2^\rho)^* & 0 & 0 & -\rho\bar{B}_2^* & \bar{S}_2^\rho & 0 & 0 & 0 & \bar{B}_2^* & & & & \\
0 & 0 & \bar{A}_2 & 0 & 0 & -I & \bar{B}_2 & 0 & 0 & 0 & 0 & & & & \\
\hline
 & & & & & & & & \ddots & & & & & & \\
 & & & & & & & & & \ddots & & & & & \\
 & & & & & & & & & & \ddots & & & & \\
\hline
 & & & & & -\delta\bar{A}_{N_t-2} & 0 & 0 & \bar{Q}_{N_t-1}^{\rho,\delta} & -\delta\bar{B}_{N_t-2} & 0 & -\rho\bar{A}_{N_t-1}^* & \bar{R}_{N_t-1}^\rho & -I & \bar{A}_{N_t-1}^* \\
 & & & & & 0 & 0 & 0 & (\bar{R}_{N_t-1}^\rho)^* & 0 & 0 & -\rho\bar{A}_{N_t-1}^* & \bar{S}_{N_t-1}^\rho & 0 & \bar{B}_{N_t-1}^* \\
 & & & & & 0 & 0 & 0 & \bar{A}_{N_t-1} & 0 & 0 & -I & \bar{B}_{N_t-1} & 0 & 0 \\
\hline
 & & & & & & & & -\delta\bar{A}_{N_t-1} & 0 & 0 & \bar{Q}_{N_t} + \delta I & -\delta\bar{B}_{N_t-1} & 0 & -I
\end{array}\right).
$$

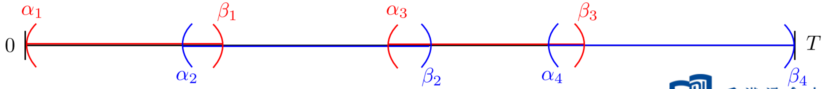
Overlapping Schwarz (Barker & Stoll (2013))

- Use overlapping subintervals (α_j, β_j)
- Solve the coupled forward-backward PDE on each subinterval in parallel

$$\dot{y}_j^k + A y_j^k = \nu^{-1}\lambda_j^k, \qquad \dot{\lambda}_j^k - A^T\lambda_j^k = y_j^k - \hat{y}$$

- Initial and final conditions from neighbours at previous iterate:

$$y_j^k(\alpha_j) = y_{j-1}^{k-1}(\alpha_j), \qquad \lambda_j^k(\beta_j) = \lambda_{j+1}^{k-1}(\beta_j).$$



Overlapping Schwarz (Barker & Stoll (2013))

They observe experimentally that:

- Fast convergence for Dirichlet problems
- For fixed overlap size, convergence is nearly independent of the spatial and temporal grid size
- Convergence may slow down when we increase the number of subintervals

Can we understand this behaviour?

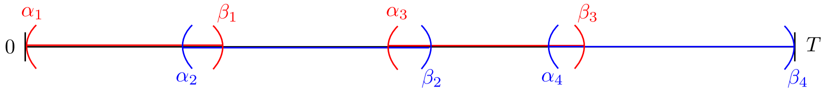
Optimized Schwarz Method (Gander & K., DD22 proceedings)

For $k = 1, 2, \ldots$, solve on each (α_j, β_j)

$$\begin{cases} \dot{y}_j^k + A y_j^k = \nu^{-1} \lambda_j^k & \text{on } (\alpha_j, \beta_j), \\ \dot{\lambda}_j^k - A^T \lambda_j^k = y_j^k - \hat{y}_j, & \end{cases}$$

with boundary conditions

$$y_j^k(\alpha_j) - q_j \lambda_j^k(\alpha_j) = y_{j-1}^{k-1}(\alpha_j) - q_j \lambda_{j-1}^{k-1}(\alpha_j),$$
$$\lambda_j^k(\beta_j) + p_j y_j^k(\beta_j) = \lambda_{j+1}^{k-1}(\beta_j) + p_j y_{j+1}^{k-1}(\beta_j).$$



Optimized Schwarz Method (Gander & K., DD22 proceedings)

For $p, q \neq 0$, this is equivalent to

$$\min \frac{1}{2}\int_{\alpha_j}^{\beta_j} \|y(t;u) - \hat{y}\|^2 + \frac{\nu}{2}\int_{\alpha_j}^{\beta_j} \|u\|^2 + \frac{p_j}{2}\|y(\beta_j;u) - p_j^{-1} g_{j+1}^{k-1}\|^2 + \frac{1}{2q_j}\|y(\alpha_j;u) - h_{j-1}^{k-1}\|^2$$

where

$$g_{j+1}^{k-1} = \lambda_{j+1}^{k-1}(\beta_j) + p_j y_{j+1}^{k-1}(\beta_j), \qquad h_{j-1}^{k-1} = y_{j-1}^{k-1}(\alpha_j) - q_j \lambda_{j-1}^{k-1}(\alpha_j)$$

For $p = q = 0$, this reduces to Dirichlet transmission conditions

Optimized Schwarz Method (Gander & K., DD22 proceedings)

For $p, q \neq 0$, this is equivalent to

$$\min \frac{1}{2}\int_{\alpha_j}^{\beta_j} \|y(t;u)-\hat{y}\|^2 + \frac{\nu}{2}\int_{\alpha_j}^{\beta_j}\|u\|^2 + \frac{p_j}{2}\|y(\beta_j;u)-p_j^{-1}g_{j+1}^{k-1}\|^2 + \frac{1}{2q_j}\|y(\alpha_j;u)-h_{j-1}^{k-1}\|^2$$

- Minimization problem with small changes in boundary conditions $\implies$ solvers available!

Subdomain solves

A shooting method: for a given initial condition y_0 and control, consider the mapping $F(y_0, u)$ as follows:

1. Integrate $\dot{y} + Ay = Bu$, $y(0) = y_0$ forwards to $t = T$
2. Let $\lambda(T) = h - py(T)$
3. Integrate $\dot{\lambda} - A^T y = C^T(Cy - \hat{y})$ backwards to $t = 0$.
4. $F(y_0, u) = (y_0 - q\lambda(0) - g, \nu u - B^T\lambda)$

Then

$$F(y_0, u) = F(0, 0) + K \begin{pmatrix} y_0 \\ u \end{pmatrix}$$

is an affine mapping, so we can solve $F(y_0, u) = 0$ using e.g. GMRES

Subdomain solves

A shooting method: for a given initial condition y_0 and control, consider the mapping $F(y_0, u)$ as follows:

1. Integrate $\dot{y} + Ay = Bu$, $y(0) = y_0$ forwards to $t = T$
2. Let $\lambda(T) = h - py(T)$
3. Integrate $\dot{\lambda} - A^T y = C^T(Cy - \hat{y})$ backwards to $t = 0$.
4. $F(y_0, u) = (y_0 - q\lambda(0) - g, \nu u - B^T\lambda)$

Alternatively, use an all-at-once approach, or any other solver for a single time interval.

Optimized Schwarz Method (Gander & K., DD22 proceedings)

For $k = 1, 2, \ldots$, solve on each (α_j, β_j)

$$\begin{cases} \dot{y}_j^k + A y_j^k = \nu^{-1} \lambda_j^k & \text{on } (\alpha_j, \beta_j), \\ \dot{\lambda}_j^k - A^T \lambda_j^k = y_j^k - \hat{y}_j, & \end{cases}$$

with boundary conditions

$$y_j^k(\alpha_j) - q_j \lambda_j^k(\alpha_j) = y_{j-1}^{k-1}(\alpha_j) - q_j \lambda_{j-1}^{k-1}(\alpha_j),$$
$$\lambda_j^k(\beta_j) + p_j y_j^k(\beta_j) = \lambda_{j+1}^{k-1}(\beta_j) + p_j y_{j+1}^{k-1}(\beta_j).$$

- Convergence for which values of p_j and q_j?
- How to choose p_j and q_j to optimize convergence?

香港浸會大學 HONG KONG BAPTIST UNIVERSITY

Convergence Analysis

- Diagonalization
  - + Explicit formula for contraction rate
  - + With or without overlap
  - − Assumes $A = A^T$
- Energy estimates
  - Integration by parts
  - + General setting ($A \neq A^T$, boundary control, etc.)
  - + Multiple subdomains
  - − No overlap

Convergence Analysis

- Diagonalization
  - $+$ Explicit formula for contraction rate
  - $+$ With or without overlap
  - $-$ Assumes $A = A^T$
- Energy estimates
  - Integration by parts
  - $+$ General setting ($A \neq A^T$, boundary control, etc.)
  - $+$ Multiple subdomains
  - $-$ No overlap

Analysis for two subdomains

- Subdomain problems:

$$\begin{cases} \begin{bmatrix} \dot{y}_1^k \\ \dot{\lambda}_1^k \end{bmatrix} + \begin{bmatrix} A & -\nu^{-1} I \\ -I & -A^T \end{bmatrix} \begin{bmatrix} y_1^k \\ \lambda_1^k \end{bmatrix} = \begin{bmatrix} 0 \\ -\hat{y} \end{bmatrix} & \text{on } I_1 = (0, \beta), \\ y_1^k(0) = y_0, \\ \lambda_1^k(\beta) + p y_1^k(\beta) = \lambda_2^{k-1}(\beta) + p y_2^{k-1}(\beta), \end{cases}$$

$$\begin{cases} \begin{bmatrix} \dot{y}_2^k \\ \dot{\lambda}_2^k \end{bmatrix} + \begin{bmatrix} A & -\nu^{-1} I \\ -I & -A^T \end{bmatrix} \begin{bmatrix} y_2^k \\ \lambda_2^k \end{bmatrix} = \begin{bmatrix} 0 \\ -\hat{y} \end{bmatrix} & \text{on } I_2 = (\alpha, T), \\ y_2^k(\alpha) - q\lambda_2^k(\alpha) = y_1^{k-1}(\alpha) - q\lambda_1^{k-1}(\alpha), \\ \lambda_2^k(T) = -\gamma(y_2^k(T) - \hat{y}(T)). \end{cases}$$

Analysis for two subdomains

- Assume $A = A^T$ and diagonalize: $y \to z$, $\lambda \to \mu$

$$\begin{cases} \begin{bmatrix} \dot{z}_1^k \\ \dot{\mu}_1^k \end{bmatrix} + \begin{bmatrix} D & -\nu^{-1} I \\ -I & -D \end{bmatrix} \begin{bmatrix} z_1^k \\ \mu_1^k \end{bmatrix} = \begin{bmatrix} 0 \\ -\hat{z} \end{bmatrix} & \text{on } I_1 = (0, \beta), \\ z_1^k(0) = z_0, & \\ \mu_1^k(\beta) + p z_1^k(\beta) = \mu_2^{k-1}(\beta) + p z_2^{k-1}(\beta), & \end{cases}$$

$$\begin{cases} \begin{bmatrix} \dot{z}_2^k \\ \dot{\mu}_2^k \end{bmatrix} + \begin{bmatrix} D & -\nu^{-1} I \\ -I & -D \end{bmatrix} \begin{bmatrix} z_2^k \\ \mu_2^k \end{bmatrix} = \begin{bmatrix} 0 \\ -\hat{z} \end{bmatrix} & \text{on } I_2 = (\alpha, T), \\ z_2^k(\alpha) - q\mu_2^k(\alpha) = z_1^{k-1}(\alpha) - q\mu_1^{k-1}(\alpha), & \\ \mu_2^k(T) = -\gamma(z_2^k(T) - \hat{z}(T)). & \end{cases}$$

Analysis for two subdomains

- Eliminating μ gives

$$\ddot{z}_1^{(i),k} - (d_i^2 + \nu^{-1}) z_1^{(i),k} = -\nu^{-1} \hat{z}^{(i)},$$

with boundary conditions

$$z_1^{(i),k}(0) = z_0^{(i)}(0)$$
$$\dot{z}_1^{(i),k} + (d_i + p\nu^{-1}) z_1^{(i),k} \Big|_{t=\beta} = \dot{z}_2^{(i),k-1} + (d_i + p\nu^{-1}) z_2^{(i),k-1} \Big|_{t=\beta}.$$

- Even for $p = 0$, this corresponds to Robin conditions!

Theorem (Gander & K., 2014)

The parallel Schwarz method converges whenever $\rho < 1$, where

$$\rho^2 = \max_{d_i \in \lambda(A)} \left| \frac{\sigma_i q \cosh(\sigma_i \alpha) + (q d_i - \nu^{-1}) \sinh(\sigma_i \alpha)}{\sigma_i \cosh(\sigma_i \beta) + (d_i + p\nu^{-1}) \sinh(\sigma_i \beta)} \right.$$

$$\cdot \frac{\nu^{-1/2} [p \cosh(\sigma_i(T-\beta)+\theta_i) - \gamma \cosh(\sigma_i(T-\beta)-\theta_i)] - (1-\nu^{-1} p\gamma) \sinh(\sigma_i(T}{\nu^{-1/2} [\cosh(\sigma_i(T-\alpha)+\theta_i) + q\gamma \cosh(\sigma_i(T-\alpha)-\theta_i)] + (q+\nu^{-1}\gamma) \sinh(\sigma_i(T}$$

with

- $d_i = i$th eigenvalue of A,
- $\sigma_i = \sqrt{d_i^2 + \nu^{-1}} > d_i \geq 0$,
- $\theta_i = \tanh^{-1}(d_i/\sigma_i)$.

Dirichlet Case ($p = q = 0$)

The convergence rate simplifies to

$$\rho^2 = \max_i \left(\frac{\sinh(\sigma_i \alpha)}{\cosh(\sigma_i \beta + \theta_i)} \cdot \frac{\nu^{1/2} \sinh(\sigma_i (T - \beta)) + \gamma \cosh(\sigma_i (T - \beta) - \theta_i)}{\gamma \sinh(\sigma_i (T - \alpha)) + \nu^{1/2} \cosh(\sigma_i (T - \alpha) + \theta_i)} \right).$$

Dirichlet Case ($p = q = 0$)

The convergence rate simplifies to

$$\rho^2 = \max_i \left(\frac{\sinh(\sigma_i \alpha)}{\cosh(\sigma_i \beta + \theta_i)} \cdot \frac{\nu^{1/2} \sinh(\sigma_i(T-\beta)) + \gamma \cosh(\sigma_i(T-\beta) - \theta_i)}{\gamma \sinh(\sigma_i(T-\alpha)) + \nu^{1/2} \cosh(\sigma_i(T-\alpha) + \theta_i)} \right).$$

Theorem ($\gamma = 0$, no target state)

For two subdomains with overlap $L \geq 0$, the parallel Schwarz method for two subdomains converges with the estimate

$$\rho \leq \frac{e^{-L\sqrt{d_{\min}^2 + \nu^{-1}}}}{\sqrt{1 + \nu d_{\min}^2} + \nu^{1/2} d_{\min}},$$

where $d_{\min} > 0$ is the smallest eigenvalue of A.

Dirichlet Case ($p = q = 0$)

Theorem ($\gamma = 0$, no target state)

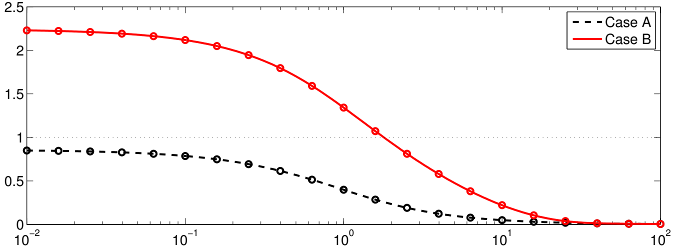
For two subdomains with overlap $L \geq 0$, the parallel Schwarz method for two subdomains converges with the estimate

$$\rho \leq \frac{e^{-L\sqrt{d_{\min}^2 + \nu^{-1}}}}{\sqrt{1 + \nu d_{\min}^2} + \nu^{1/2} d_{\min}},$$

where $d_{\min} > 0$ is the smallest eigenvalue of A.

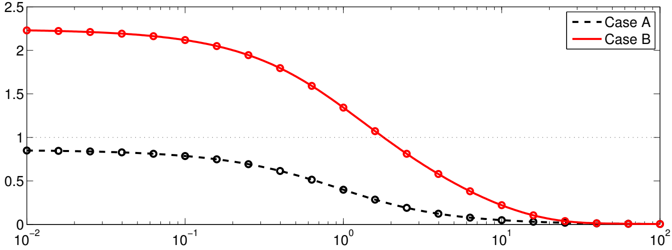
- Method converges even without overlap
- Convergence independent of the spatial mesh parameter!

Dirichlet Case ($p = q = 0$)



- Case A: $\Omega_1 = (0, 1)$, $\Omega_2 = (1, 3)$, $\gamma = 0$
- Case B: $\Omega_1 = (0, 2.9)$, $\Omega_2 = (2.9, 3)$, $\gamma = 10$

Dirichlet Case ($p = q = 0$)



- Case A converges for all positive definite matrices
- Convergence slow if $d_{\min} \ll 1$
- Case B diverges if $d_{\min} \lesssim 2$ (e.g. Neumann boundary)

Optimized case, $p = q$

- If $\gamma = 0$, the expression simplifies to

$$\rho^2 = \max_{d_i \in \lambda(A)} \left| \frac{\sigma_i p \cosh(\sigma_i \alpha) + (p d_i - \nu^{-1}) \sinh(\sigma_i \alpha)}{\sigma_i \cosh(\sigma_i \beta) + (d_i + p \nu^{-1}) \sinh(\sigma_i \beta)} \cdot \frac{p \sigma_i \cosh(\sigma_i (T - \beta)) + (p d_i - 1) \sinh(\sigma_i (T - \beta))}{\sigma_i \cosh(\sigma_i (T - \alpha)) + (p + d_i) \sinh(\sigma_i (T - \alpha))} \right| .$$

Optimized case, $p = q$

- If $\gamma = 0$, the expression simplifies to

$$\rho^2 = \max_{d_i \in \lambda(A)} \left| \frac{\sigma_i p \cosh(\sigma_i \alpha) + (p d_i - \nu^{-1}) \sinh(\sigma_i \alpha)}{\sigma_i \cosh(\sigma_i \beta) + (d_i + p\nu^{-1}) \sinh(\sigma_i \beta)} \cdot \frac{p \sigma_i \cosh(\sigma_i (T - \beta)) + (p d_i - 1) \sinh(\sigma_i (T - \beta))}{\sigma_i \cosh(\sigma_i (T - \alpha)) + (p + d_i) \sinh(\sigma_i (T - \alpha))} \right|.$$

- For high frequencies and no overlap, we have

$$\rho \longrightarrow p \cdot \underbrace{\lim_{d_i \to \infty} \left(\frac{\cosh(\sigma_i \alpha + \theta_i) \cosh(\sigma_i (T - \alpha) + \theta_i)}{\cosh(\sigma_i \alpha + \theta_i) \cosh(\sigma_i (T - \alpha) + \theta_i)} \right)^{1/2}}_{=1}.$$

So convergence cannot occur unless $p \in [0, 1)$.

Optimized case, $p = q$

- If $\gamma = 0$, the expression simplifies to

$$\rho^2 = \max_{d_i \in \lambda(A)} \left| \frac{\sigma_i p \cosh(\sigma_i \alpha) + (p d_i - \nu^{-1}) \sinh(\sigma_i \alpha)}{\sigma_i \cosh(\sigma_i \beta) + (d_i + p\nu^{-1}) \sinh(\sigma_i \beta)} \cdot \frac{p \sigma_i \cosh(\sigma_i (T - \beta)) + (p d_i - 1) \sinh(\sigma_i (T - \beta))}{\sigma_i \cosh(\sigma_i (T - \alpha)) + (p + d_i) \sinh(\sigma_i (T - \alpha))} \right| .$$

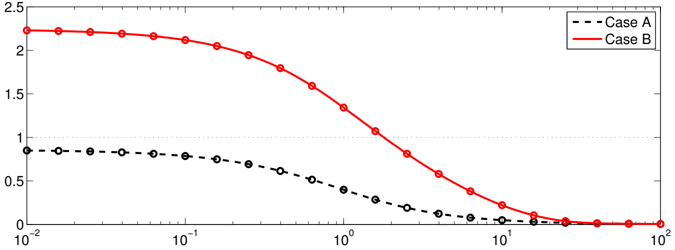
- For high frequencies and no overlap, we have

$$\rho \longrightarrow p \cdot \underbrace{\lim_{d_i \to \infty} \left(\frac{\cosh(\sigma_i \alpha + \theta_i) \cosh(\sigma_i (T - \alpha) + \theta_i)}{\cosh(\sigma_i \alpha + \theta_i) \cosh(\sigma_i (T - \alpha) + \theta_i)} \right)^{1/2}}_{=1} .$$

- Optimal p obtained by equioscillation: find p^* such that

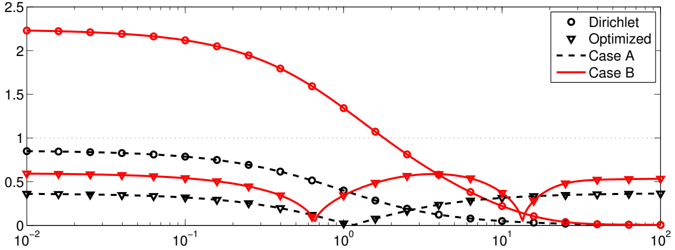
$$\lim_{d_i \to 0} \rho(p^*) = \lim_{d_i \to \infty} = p^* .$$

Dirichlet Case ($p = q = 0$)



- Case A: $\Omega_1 = (0, 1)$, $\Omega_2 = (1, 3)$, $\gamma = 0$
- Case B: $\Omega_1 = (0, 2.9)$, $\Omega_2 = (2.9, 3)$, $\gamma = 10$

Optimized case, $p = q$



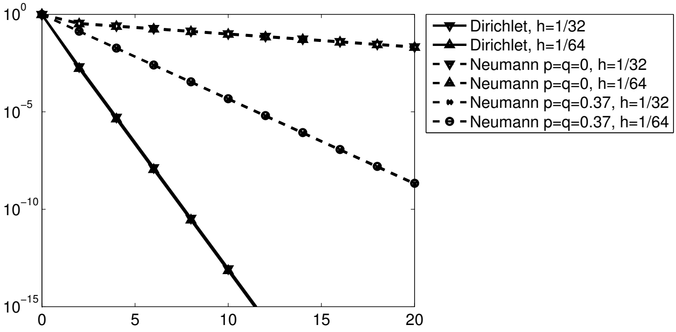
- Case A: $\Omega_1 = (0, 1)$, $\Omega_2 = (1, 3)$, $\gamma = 0$
- Case B: $\Omega_1 = (0, 2.9)$, $\Omega_2 = (2.9, 3)$, $\gamma = 10$
- Convergence for all frequencies

Numerical Example 1

- Governing PDE: $u_t = u_{xx}$ in $(x, t) \in (0, 1) \times (0, 3)$
- Discretization: Crank–Nicolson with $h = 1/32$ and $h = 1/64$
- Dirichlet or Neumann boundary conditions in space
- Two temporal subdomains: $\Omega_1 = (0, 1)$, $\Omega_2 = (1, 3)$

香港浸會大學 HONG KONG BAPTIST UNIVERSITY

Numerical Example 1



- Mesh independent convergence
- Optimized conditions beneficial for Neumann case

Analysis, Part II

- Diagonalization
  - + Explicit formula for contraction rate
  - + With or without overlap
  - – Assumes $A = A^T$
- Energy estimates
  - Integration by parts
  - + General setting ($A \neq A^T$, boundary control, etc.)
  - + Multiple subdomains
  - – No overlap

香港浸會大學 HONG KONG BAPTIST UNIVERSITY

Energy Estimates

- By linearity, subtract the exact solution to obtain the error equations

$$\dot{y} + Ay = \nu^{-1}\lambda, \qquad \dot{\lambda} - A^T\lambda = y.$$

- We want to prove that $(y_j^k, \lambda_j^k) \to 0$ as $k \to \infty$
- Consider the change of variables

$$\begin{pmatrix} z \\ \mu \end{pmatrix} = \underbrace{\begin{bmatrix} 1 & r \\ -s & 1 \end{bmatrix}}_{=B} \begin{pmatrix} y \\ \lambda \end{pmatrix} \iff \begin{pmatrix} y \\ \lambda \end{pmatrix} = \frac{1}{1+rs} \begin{bmatrix} 1 & -r \\ s & 1 \end{bmatrix} \begin{pmatrix} z \\ \mu \end{pmatrix},$$

where $r, s > 0$ are to be chosen as a function of p and q.

Energy Estimates

If we multiply the transformed system by (μ^T, z^T) and integrate, we obtain

$$0 = \mu(\alpha_j)^T z(\alpha_j) - \mu(\alpha_{j-1})^T z(\alpha_{j-1}) + \frac{1}{1+rs}\int_{\alpha_{j-1}}^{\alpha_j} \mu^T(r^2 - 2rH - \nu^{-1})\mu$$
$$+ \frac{1}{1+rs}\int_{\alpha_{j-1}}^{\alpha_j} z^T(s^2\nu^{-1} - 2sH - 1)z$$

with $H = \frac{1}{2}(A + A^T) \geq 0$. We want to choose r and s such that

- $r, s > 0$,
- $r^2 - 2rH - \nu^{-1}$ and $s^2\nu^{-1} - 2sH - 1$ are *negative definite*,
- $\mu^T z = (\lambda - sy)^T(y + r\lambda) = c_1|\lambda + py|^2 - c_2|y - q\lambda|^2$.

Energy Estimates

With this choice, we obtain the relation

$$c_1|\lambda(\alpha_{j-1}) + py(\alpha_{j-1})|^2 + c_2|y(\alpha_j) - q\lambda(\alpha_j)|^2$$
$$= c_1|\lambda(\alpha_j) + py(\alpha_j)|^2 + c_2|y(\alpha_{j-1}) - q\lambda(\alpha_{j-1})|^2 - \frac{1}{1+rs}\int_{\alpha_{j-1}}^{\alpha_j} \langle\text{pos. terms}$$

Thus, at the kth iteration, we have

$$c_1|\lambda_j^k(\alpha_{j-1}) + py_j^k(\alpha_{j-1})|^2 + c_2|y_j^k(\alpha_j) - q\lambda_j^k(\alpha_j)|^2$$
$$\leq c_1|\lambda_j^k(\alpha_j) + py_j^k(\alpha_j)|^2 + c_2|y_j^k(\alpha_{j-1}) - q\lambda_j^k(\alpha_{j-1})|^2$$

Energy Estimates

With this choice, we obtain the relation

$$c_1|\lambda(\alpha_{j-1}) + py(\alpha_{j-1})|^2 + c_2|y(\alpha_j) - q\lambda(\alpha_j)|^2$$
$$= c_1|\lambda(\alpha_j) + py(\alpha_j)|^2 + c_2|y(\alpha_{j-1}) - q\lambda(\alpha_{j-1})|^2 - \frac{1}{1+rs}\int_{\alpha_{j-1}}^{\alpha_j} \langle \text{pos. terms}$$

Thus, at the kth iteration, we have

$$c_1|\lambda_j^k(\alpha_{j-1}) + py_j^k(\alpha_{j-1})|^2 + c_2|y_j^k(\alpha_j) - q\lambda_j^k(\alpha_j)|^2$$
$$\leq c_1|\lambda_{j+1}^{k-1}(\alpha_j) + py_{j+1}^{k-1}(\alpha_j)|^2 + c_2|y_{j-1}^{k-1}(\alpha_{j-1}) - q\lambda_{j-1}^{k-1}(\alpha_{j-1})|^2$$

Energy Estimates, Two Subdomains

Theorem

Let $\gamma = 0$ (no target state). If $p > 0$ and $q > 0$ are such that

$$0 \leq 2\nu^{1/2} q \leq 1 - pq \leq 2p\nu^{-1/2},$$

then the two-subdomain OSM converges with

$$\rho \leq \frac{p(1 - \nu^{1/2} q)}{p + \nu^{1/2}} < 1.$$

- If $\nu = 1$, then we get $p = q \approx 0.414$, $\rho \leq p^2 \approx 0.1716$.

Multiple Subdomains

Choose p and q as follows:

1. Choose r and s small enough so that $r^2 - 2rH - \nu^{-1}$ and $s^2\nu^{-1} - 2sH - 1$ are negative definite.
2. Calculate p, q (and c_1, c_2) such that

$$\mu^T z = (\lambda - sy)^T(y + r\lambda) = c_1|\lambda + py|^2 - c_2|y - q\lambda|^2.$$

Theorem

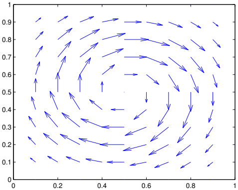
Let $\gamma = 0$ (no target state). Then there exists $p, q > 0$ such that $pq < 1$ and OSM with N subdomains converges.

Numerical Example 2

- 2D advection-diffusion equation on $\Omega = (0,1) \times (0,1)$

$$y_t - \nabla \cdot (\nabla y + \mathbf{b} y) = 0$$

$$\mathbf{b} = \sin \pi x \sin \pi y \begin{pmatrix} y - 0.5 \\ 0.5 - x \end{pmatrix}$$



- $T = 3$, split into two subdomains at $\alpha = 1$
- Neumann conditions, no target state
- Upwind discretization, $h = 1/16$ and $h = 1/32$
- Transmission conditions: $p = q = \sqrt{2} - 1$

Numerical Example 2

- Predicted convergence factor: 0.1716

Its	$h = 1/16$		$h = 1/32$	
	Error	Ratio	Error	Ratio
1	9.9908e-001		9.9977e-001	
2	1.3762e-001	0.1378	1.3810e-001	0.1381
3	2.0115e-002	0.1462	2.0266e-002	0.1468
4	3.0901e-003	0.1536	3.1234e-003	0.1541
5	4.9302e-004	0.1595	4.9936e-004	0.1599
6	8.0785e-005	0.1639	8.1899e-005	0.1640
7	1.3474e-005	0.1668	1.3659e-005	0.1668
8	2.2729e-006	0.1687	2.3023e-006	0.1686
9	3.8599e-007	0.1698	3.9046e-007	0.1696
10	6.5653e-008	0.1701	6.6306e-008	0.1698

Control and Observation over Subsets of Ω

- If the control is only defined on the boundary, then the PDE system becomes

$$\begin{pmatrix} \dot{y} \\ \dot{\lambda} \end{pmatrix} + \begin{bmatrix} A & -BB^T \\ -C^T C & -A^T \end{bmatrix} \begin{pmatrix} y \\ \lambda \end{pmatrix} = 0.$$

Control and Observation over Subsets of Ω

- Using the same calculation as before, we see that convergence occurs if

$$z^T(s^2BB^T - 2sH - C^TC)z \leq 0, \qquad \mu^T(r^2C^TC - 2rH - BB^T)\mu \leq 0$$

- This leads to the constraint

$$0 < s \leq \min_{\substack{z\in\mathrm{range}(B)\\ z\neq 0}} \frac{z^THz}{\|B^Tz\|^2} + \sqrt{\left(\frac{z^THz}{\|B^Tz\|^2}\right)^2 + \frac{\|Cz\|^2}{\|B^Tz\|^2}}$$

  and an analogous one for r.
- This minimum is non-zero whenever

$$\ker(HB) \cap \ker(CB) = \{0\}.$$

Control and Observation over Subsets of Ω

Theorem

Let $\gamma = 0$ (no target state). Suppose that

$$\ker(HB) \cap \ker(CB) = \{0\}$$

and

$$\ker(CH) \cap \ker(CB) = \{0\}$$

Then there exists $p, q > 0$ such that OSM with N subdomains converges.

- A good choice of s (and similarly for r) is given by twice the smallest eigenvalue of the GEVP

$$B^T HBv = \lambda(B^T B)^2 v.$$

Scalability

- The energy argument above does not give us a contraction factor ρ.
- To obtain a contraction estimate, we look for a constant $0 < C < 1$ such that

$$C(c_1|h(\alpha_j)|^2 + c_2|g(\alpha_{j-1})|^2) \leq \frac{1}{1+rs}\int_{\alpha_{j-1}}^{\alpha_j} \langle z_j, M_1 z_j\rangle + \langle \mu_j, M_2\mu_j\rangle.$$

- This constant must exist because z_j and μ_j solves the linear homogeneous PDE system, and thus belongs to a *finite-dimensional* vector space parameterized by the Robin traces.

Scalability

- This leads to a contraction factor of $\rho = 1 - C$, where C depends on the sub-interval length $H = \alpha_j - \alpha_{j-1}$.
- A scaling argument shows that as H decreases, the contraction factor behaves in the worst case like

$$\rho \approx 1 - cH,$$

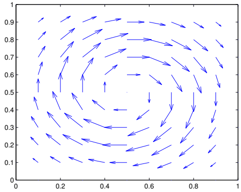
so a coarse grid is needed in general.

Numerical Example 3

- 2D advection-diffusion equation on $\Omega = (0,1) \times (0,1)$

$$y_t - \nabla \cdot (\nabla y + \mathbf{b} y) = 0$$

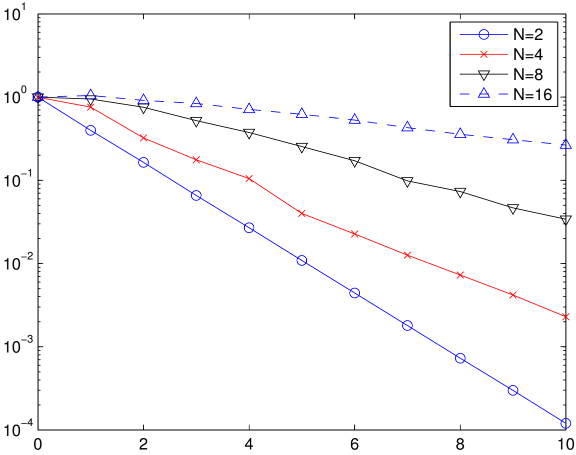
$$\mathbf{b} = \sin \pi x \sin \pi y \begin{pmatrix} y - 0.5 \\ 0.5 - x \end{pmatrix}$$



- $T = 4$, split into 2, 4, 8, 16 equal subdomains
- Neumann conditions, no target state
- Upwind discretization, $h = 1/16$
- Transmission conditions: $p = q = \sqrt{2} - 1$

香港浸會大學 HONG KONG BAPTIST UNIVERSITY

Numerical Example 3



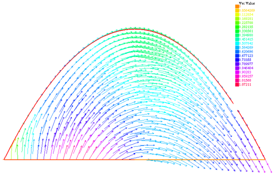
Numerical Example 3

We expect $\rho = 1 - CH$:

H	ρ	$1-\rho$	$H(1-\rho)$
1/2	0.4063	0.5937	1.1864
1/4	0.5659	0.4341	1.7364
1/8	0.6653	0.3347	2.6776
1/16	0.8409	0.1591	2.5456

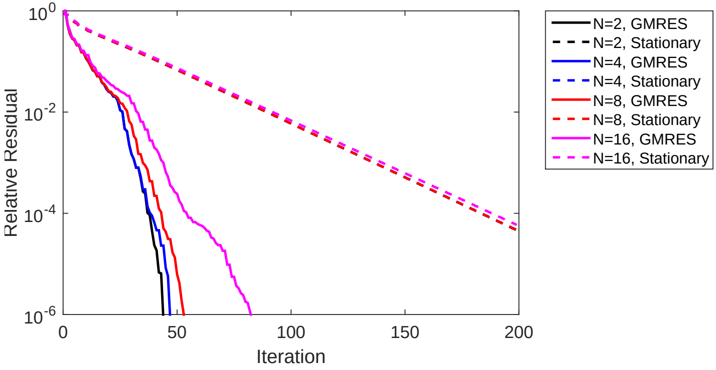
Numerical Example 4

- 2D advection-diffusion equation
- Flow field obtained by Stokes equation
- Finite volume method as in Bermúdez et al (1998)



- Source (control) at centre of domain, observation at one point on boundary
- 736 dof in space, 64 time steps
- $T = 32$, split into 2, 4, 8, 16 equal subdomains
- Transmission conditions: $p = q = 0.8563$

Numerical Example 4



Numerical Example 4

Timing obtained on SciBlade cluster at HKBU (2048 cores, Dell PowerEdge M600 blade server, Intel Xeon E5450 2.66GHz Quad-Core Processors, Peak Performance: 21.79 TFlops):

N	Global GMRES Iter.	Max local GMRES Iter.	Time per global Iter.	Total time
2	44	27	3.74	164.37
4	47	21	1.89	88.76
8	53	18	1.17	61.80
16	83	17	0.80	66.32

Summary

- Schwarz methods for parabolic control problems:
  - Inherent use of Robin conditions
  - Mesh independent convergence, even without overlap
  - Use of additional Robin parameters enhances convergence
  - Energy estimates

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

Ongoing work

- Design of coarse grid correction (with T. Wihler, U. Bern)
- Preconditioning for shooting method (with J. Salomon)
- Better eigenvalue distribution for GMRES?
- Experiments for other time-dependent problems (e.g. Stokes, nonlinear problems)
- Control constraints

THANK YOU!