PROXIMAL DOMAIN DECOMPOSITION ALGORITHMS AND APPLICATION TO ELLIPTIC PROBLEMS

by

Said Oualibouch and Noureddine El Mansouri

Institut d'Informatique et I.A., Université de Neuchâtel,
Rue Emile-Argand 11, CH-2007 Neuchâtel.
(Switzerland)
e-mail: said.oualibouch@info.unine.ch

Abstract: Many problems in mathematical economics, engineering and mechanics reduce to large, sparse, unstructured, and poorly conditioned optimization problems. Because of their size and lack of structure, these problems are hard to solve by classical and global methods which are not adapted to modern parallel computers. In this setting, decomposition methods are very attractive, because the solution of the global problem can be reduced to the iterative solution of many subproblems of smaller size. These techniques are designed for exploiting the full power of modern parallel computers, because of the built-in parallelism of the algorithms and the character of the associated data. In this paper, we are interested in linear and nonlinear elliptic problems, and we present two additive algorithms based on the proximal techniques. They yield two overlapping additive domain decomposition methods: *Proximal-Jacobi method* and *Proximal Schwarz methods*. These methods can also be viewed as regularized versions of Additive Schwarz methods. In order to validate these algorithms, some numerical results are given for the homogeneous Dirichlet problem.

Keywords: Domain Decomposition Methods, Additive Schwarz Methods, Proximal Techniques.

0.1 INTRODUCTION

The main characteristic of decomposition algorithms in convex programming is the splitting of a large-scale problem into a set of reduced size subproblems which may be solved either in parallel or in sequence. Very often, the structure of system induces a splitting of the variable set in disjoint subsystems yielding some coupling constraints. In our approach, the subsystems may overlap, but the advantage is that there are no artificial coupling constraints. The efficiency of the decomposition methods depends not only on the specific properties of the objective functions as convexity, smoothness or separability but mostly on the degree of the coupling between the subsystems.
Besides the motivation of reducing the size of the problems appears the possibility of decentralizing the optimal decision among the local subproblems as in an ideal hierarchical organization. It is well-known, see [1], that most classical approaches perform only a partial decentralization and need a heavy coordination upper level to build a solution from the local proposals. This is due to the lack of uniqueness of subproblems solutions which in turn is a

Domain Decomposition Methods in Sciences and Engineering, edited by R. Glowinski *et al.*

direct consequence of the non-smoothness of the coordination function. It is then natural to introduce regularizing terms in the decentralized process to cope with both non-uniqueness and non-smoothness as in the proximal point algorithm [11].

In this paper we are interested in solving the following convex problem:

$$\min_{v\in K} f(v), \qquad K \subset H \tag{0.1}$$

where the function f is proper convex lower-semi-continuous, H a finite dimensional Hilbert space H and K a closed convex subset of H such that $\mathrm{ri}\,\mathrm{dom}\, f \cap \mathrm{ri}\, K \neq \emptyset$. We suppose that $H = \sum_{i=0}^{m} H_i$ and $K = \sum_{i=0}^{m} K_i$, $K_i \subset H_i$. The problem (0.1) is equivalent to the following one:

$$\min_{v_i\in H_i,\ i=0,\ldots,m} F(\sum_{i=0}^{m} v_i) \tag{0.2}$$

where $F = f + \chi_K$ with χ_K is the $\{0,\infty\}$-indicator function of the subset K. In this case, we can use parallel methods (at least m processors) to solve the problem (0.2).

To solve a partial differential equation or a variational inequality often is equivalent to minimizing an energy function. In practice, there are different ways to decompose this energy function and to decompose the space of minimization, see [5], [13] and the references therein. In order to illustrate some of the possible ways, let us consider the following homogeneous Dirichlet problem:

$$\begin{cases} -\Delta u &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega \end{cases} \tag{0.3}$$

Where Ω is a bounded domain in $\mathbb{R}^n$ and $\partial\Omega$ its boundary. The problem (0.3) can be reduced to the following one

$$\min_{v\in H} F(v) \tag{0.4}$$

where the function F is defined by: $F(v) = \int_\Omega (\frac{1}{2}|\nabla v|^2 - fv)dx$ and $H = H_0^1(\Omega) \cap V_h$, V_h being the finite element space. Space decomposition (SD) can be done in different ways. For example, the finite element space itself is the linear span of the finite element basis, therefore it can be easily regarded as the sum of subspaces. The multilevel method is another way to decompose a finite element space. In the overlapping DDM, we decompose a domain Ω into overlapping subdomains Ω_i, $i = 1,\ldots,m$, this means that $\Omega = \cup_{i=1}^{m}\Omega_i$, and for each Ω_i, there exists a subdomain Ω_j such that $\Omega_i \cap \Omega_j \neq \emptyset$. If the subdomains overlap uniformly, it is known that, see [6],

$$H_0^1(\Omega) = H_0^1(\Omega_1) + H_0^1(\Omega_2) + \ldots + H_0^1(\Omega_m) \tag{0.5}$$

The central idea to construct parallel methods by SD derives from the observation that the space H can be decomposed into the sum of smaller

and simpler subspaces as in (0.5), then the minimization problem (0.4) can be replaced by a problem similar to (0.2) where $H_i = H_0^1(\Omega_i) \cap V_h$. For a general SD, the minimizer of (0.2) may not be unique, but it has been proved that several methods converge for (0.2), for example the sequential Gauss-Seidel method and the parallel Jacobi method, see [12]. In the overlapping DDM case, the Schwarz alternating method is nothing else but the Gauss-Seidel method. Applying the Gauss-Seidel and Jacobi methods to some linear and nonlinear problems, several sequential and parallel overlapping DDM's can be got, see [12], [13].

The rest of this paper is organized as follow: the section 0.2 will give an overview of proximal techniques. In section 0.3, we will present two versions of additive proximal domain decomposition methods. Finally, we will conclude with some numerical results and remarks.

0.2 PROXIMAL TECHNIQUES

Let denote the inner product on H by $\langle .,. \rangle$ and the induced norm by $\|.\|$. Let $T : H \rightrightarrows H$ be a set-valued map (operator) on H, we define its graph by: $\mathrm{Gr}(T) = \{(x, y) \in H \times H | w \in Tx\}$, the inverse T^{-1} of T is the operator defined by: $\mathrm{Gr}(T^{-1}) = \{(x, y) \in H \times H | (y, x) \in \mathrm{Gr}(T)\}$. The operator T is said to be monotone if $\langle x - x', y - y' \rangle \geq 0$ for all (x, y) and (x', y') in $\mathrm{Gr}(T)$. T is said to be maximal monotone if it is monotone and its graph is not properly contained in the graph of any other monotone operator. Several authors have extensively studied the theory of maximal monotone operators in Hilbert spaces and applications, among others Brezis [3], Dolezal [4] and Aubin and Ekeland [2].

Many problems from mathematical programming, complementarity, mathematical economics and other fields can be formulated in the way of the fundamental problem of finding an element $z \in H$ such that

$$0 \in Tz \tag{0.6}$$

For example, if T is the subdifferential operator ∂F of F ($F \not\equiv +\infty$), then Minty [9] have shown that T is maximal monotone, and the problem of minimizing the function F is equivalent to that of finding a zero of T, i.e. $F(z) = \min F(x)$ means that $0 \in Tz$. The problem (0.6) is equivalent to that of finding a fixed point of the resolvent operator $J_\lambda^T = (I + \lambda T)^{-1}$ (to simplify, we denote this operator by J_λ), $\lambda > 0$, i.e. find $z \in H$ such that: $z = J_\lambda z$.

As early as 1962, Minty [10] pointed out that, when the operator T is maximal monotone, its resolvent (the Moreau-Yosida resolvent) J_λ is single-valued on $\mathbb{R}^n$ and non-expansive. This result suggests that a solution to the inclusion $0 \in T(z)$ may be iteratively approximated using the classical iteration $z^{(k+1)} = J_\lambda z^{(k)}$. One could modify this scheme by iteratively varying the scalar λ and by choosing the iterator $z^{(k+1)}$ to be an approximate solution to the equation $(I + \lambda T)z = z^{(k)}$, i.e., $z^{(k+1)} \approx J_\lambda z^{(k)}$. The proximal point algorithm precisely

applies these ideas. The algorithm, starting from any point $z^{(0)}$, generates a sequence $\{z^{(k)}\}$ in H as follows:

$$x^{(k+1)} = J_{\lambda_k} x^{(k)} \tag{0.7}$$

where $\{\lambda_k\}$ is some sequence of positive real numbers. A wide variety of global convergence results for the proximal point algorithm can be found in literature. As early as 1970 and 1972, Martinet [7] [8] proved the convergence of the exact proximal point algorithm for certain special cases of the operator T with fixed $\lambda_k \equiv \lambda$. The first theorem on the convergence of the general proximal point algorithm was proved by Rockafellar in 1976 [11]. His theorem not only insures the global convergence under a mild approximating rule, but also describes the global behavior if the inclusion $0 \in T(z)$ has no solution. The convergence rate of the proximal point algorithm depends on the properties of the operator T, the choice of the sequence $\{\lambda_k\}$, and the accuracy of the approximation $z^{(k+1)} \approx J_{\lambda_k} z^{(k)}$.

Consider the general form of the convex optimization problem (0.4). One method to solve (0.4) is to regularize the objective function by using the proximal regularization, already introduced in this document. Given a real positive parameter λ, we recall that a proximal approximation (regularization) of F is defined by:

$$F_\lambda(x) = \inf_u \{F(u) + \frac{1}{2\lambda}\|u - x\|^2\} \tag{0.8}$$

This function is convex and differentiable and when minimized possesses the same set of minimizers and the same optimal value as problem (0.4). When we consider the right part of (0.8) and seek its optimality condition, we have

$$\begin{aligned} 0 \in \partial F(u^*) + (1/\lambda)(u^* - x) &\Leftrightarrow 0 \in \lambda \partial F(u^*) + (u^* - x) \\ &\Leftrightarrow x \in (I + \lambda \partial F)(u^*) \\ &\Leftrightarrow u^* = (I + \lambda \partial F)^{-1}(x) \end{aligned}$$

This is the motivation of the terminology "proximal regularization". In such case $T = \partial F$ and the proximal point iteration (0.7) is equivalent to:

$$x^{(k+1)} = \text{Argmin}\{F(x) + (1/2\lambda_k)\|x - x^{(k)}\|^2\} \tag{0.9}$$

If the function F is separable and consequently the variables $(x_i)_{i=0}^m$ are independent, the space H can be expressed as a product space, i.e. $H = \prod_{i=0}^m H_i$ and $T = \prod_{i=0}^m T_i$, where T_i is a maximal monotone operator on H_i, then we have $(I + \lambda T)^{-1} = \prod_{i=0}^m (I + \lambda T_i)^{-1}$. This property is fundamental when we are interested in a parallel implementation of a proximal decomposition algorithm. In this case, observe that we can substitute the proximal step (0.9) by $m + 1$ elementary proximal steps that can be executed simultaneously each on its own processor.

0.3 ADDITIVE PROXIMAL ALGORITHMS

In the case where the function F is not supposed to be separable but the variables are independent, Martinet [7] proposed the following sequential algorithm: at iteration k, we solve sequentially for $j = 0, ..., m$

$$x_j^{k+1} = \arg\min_{x_j \in H_j} \{F(\sum_{i=0}^{j-1} x_i^{k+1} + x_j + \sum_{i=j+1}^{m} x_i^k) + \frac{1}{2\lambda}\|x_j - x_j^k\|^2\} \qquad (0.10)$$

Martinet [7] has established the convergence of this regularized relaxation method when F has continuous partial derivatives. But in our application we don't assume the separability of the function nor the independence of the variables, i.e, the subspaces H_i, $i = 0, ..., m$ are not orthogonal to each other. In such case, Tai [12] proposed a parallel Jacobi-like iteration and established its convergence when F has Lipschitz continuous and coercive derivative. In this paper, we refer to this algorithm by Jacobi-Schwarz. This Jacobi-Schwarz is given by:

Do in parallel for $j = 0, ..., m$*:*

Step 0: Choose $u_j^0 \in H_j$ and $\alpha_j > 0$, such that $\sum_j \alpha_j \leq 1$.

Step 1: $u_j^{k+1/2} = \arg\min_{u_j \in H_j}\{F(\sum_{i=0}^{j-1} u_i^k + u_j + \sum_{i=j+1}^{m} u_i^k)\}$ (0.11)

Step 2: $u_j^{k+1} = u_j^k + \alpha_j(u_j^{k+1/2} - u_j^k)$ (0.12)

In our approach, we have no assumptions about the smoothness of the objective, we overcome this thanks to the proximal regularization of the functions to be minimized in the right side of (0.11).

0.3.1 Parallel Proximal-Jacobi Algorithm (PJ)

Do in parallel for $j = 0, ..., m$*:*

Step 0: Choose $u_j^0 \in H_j$ and α_j, such that $\sum_j \alpha_j \leq 1$.

Step 1: $u_j^{k+1/2} = \arg\min_{u_j \in H_j}\{F(\sum_{i=0}^{j-1} u_i^k + u_j + \sum_{i=j+1}^{m} u_i^k) + \frac{1}{2\lambda}\|u_j - u_j^k\|^2\}$ (0.13)

Step 2: $u_j^{k+1} = u_j^k + \alpha_j(u_j^{k+1/2} - u_j^k)$ (0.14)

The minimizer of (0.2) may not be unique, therefore in the convergence analysis of our algorithm, we will only prove that $u^{n+1} = \sum_{i=0}^{m} u_i^{n+1}$ converges to the minimizer of (0.2)

Before giving some convergence results, we recall some definitions in convex analysis.

Theorem 1 *Assume that F is coercive, convex lower semi-continuous on H and $\sum_j \alpha_j \leq 1$, $0 < \alpha_i < 1$. Then the sequence $\{u^n = \sum_{i=0}^{m} u_i^n\}_n$ has the convergence property.*

Proof: Let us begin by proving the convergence of the sequence $\{F(u^n)\}_n$,

$$\begin{aligned} F(u^{n+1}) &= F(\textstyle\sum_{i=0}^m u_i^{n+1}) \\ &= F(u^n + \textstyle\sum_{i=0}^m \alpha_i (u_i^{n+1/2} - u_i^n)) \\ &= F(\textstyle\sum_{i=0}^m \alpha_i (u^n + u_i^{n+1/2} - u_i^n) + (1 - \sum_{i=0}^m \alpha_i) u^n) \\ &\leq \textstyle\sum_{i=0}^m \alpha_i F(u^n + u_i^{n+1/2} - u_i^n) + (1 - \sum_{i=0}^m \alpha_i) F(u^n) \\ &\leq \textstyle\sum_{i=0}^m \alpha_i (F(u^n) - \frac{1}{2\lambda} \|u_i^n - u_i^{n+1/2}\|) + (1 - \sum_{i=0}^m \alpha_i) F(u^n) \end{aligned}$$

$$F(u^{n+1}) \leq F(u^n) - \frac{1}{2\lambda} \sum_{i=0}^m \alpha_i \|u_i^n - u_i^{n+1/2}\|^2 \tag{0.15}$$

Therefore, we proved that the sequence $\{F(u^n)\}_n$ is a decreasing sequence bounded below by $F(u)$, u is the minimizer of (0.2). So $\{F(u^n)\}_n$ is a convergent sequence. From the last inequality (0.15), we can easily deduce that the sequence $\{u^{n+1} - u^n\}$ vanishes as $n \to \infty$. F is coercive, so the sequences $\{u^n\}$ and $\{u^{n+1/2}\}$ have a limit point u^∞. If we consider the optimality condition of the minimization problem in the right side of (0.13), it can be seen that $y_i^{n+1/2} = \partial F_i^n(u_i^{n+1/2}) = (u_i^n - u_i^{n+1/2})/\lambda$, where $F_i^n(v_i) = F(u^n - u_i^n + v_i)$. And From (0.15), we can also deduce that for each $i = 0, ..., m$, $y_i^{n+1/2}$ vanishes as $n \to \infty$. Finally, the limit point of u^n is a minimizer of the problem (0.2).

Remark 1 *When we assume that the function F is differentiable and its gradient is uniformly continuous, the proof of convergence of the PJ algorithm is similar to that of Jacobi-Schwarz algorithm, proposed by X.-C. Tai in [12].*

0.3.2 Proximal Schwarz Algorithm (PS)

The basic iteration of the algorithm considered in this section is similar to that multiplicative one proposed by Martinet [7], as described in (0.10). The difference is that our algorithm is additive and the separability of the problem is not necessary for us, in other words, the subdomains may overlap. This parallel Proximal-Schwarz Algorithm is described bellow:
Do in parallel for $j = 0, ..., m$:

Step 0: Choose $u_j^0 \in H_j$.

Step 1: $u_j^{k+1/2} = \arg\min_{u_j \in H_j} \{F(\sum_{i=0}^{j-1} u_i^k + u_j + \sum_{i=j+1}^m u_i^k)\}$ (0.16)

The convergence of this algorithm is proven without using any assumption about the smoothness of the objective function.

0.4 NUMERICAL RESULTS AND CONCLUSIONS

We have implemented these algorithms for a homogeneous Dirichlet problem (0.3) on a 2D domain decomposed into two overlapped subdomains, i.e., two subspaces. We take $\alpha_1 = \alpha_2 = 0.45$ and for each value of λ, we count the number

of iterations until $\|u^n - \overline{u}\| < 10^{-12}$ ($\overline{u}$ is the exact solution). The obtained numerical results show that the parameter λ influences the rate of convergence of the Proximal-Jacobi (PJ) and Proximal-Schwarz (PS) algorithms. Figure 0.1 shows the behavior of these algorithms as well as our reference algorithm Jacobi-Schwarz (JS). We notice that, for our example, the algorithm PS is more efficient for a good choice of λ.

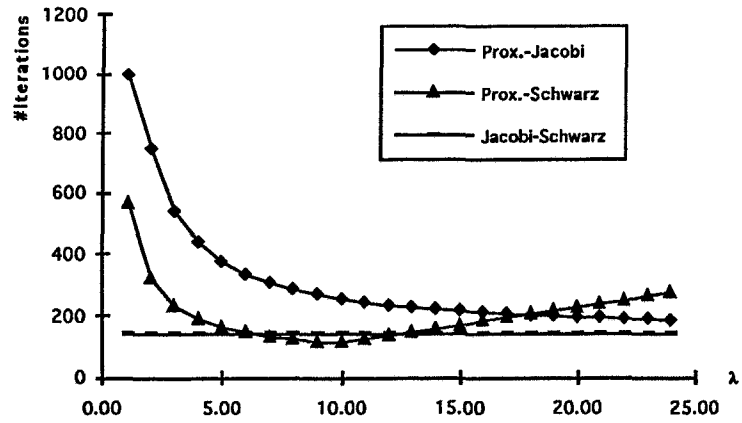

Figure 0.1 Behavior of the proposed algorithms in function of λ

To sum up, in this paper we have proposed some new methods that are useful for both linear and nonlinear elliptic problems. The numerical results show that for certain values of the parameter λ, the proximal regularization is a good way to accelerate the Jacobi-like domain decomposition methods. The corresponding theoretical results and a more detailed proof of theorem 1 and some applications to nonlinear elliptic problems will be submitted elsewhere.

REFERENCES

[1] Arrow K. and Hurwics L. (1960) Decentralization and computation in resource allocation. *Essays in Economics and Econometrics*, R. Phouts ed., Chapell Hill.

[2] Aubin J.P. and Ekeland I. (1984) Applied Nonlinear Analysis, *Wiley Interscience*, New York.

[3] Brezis H. (1973) Opérateurs Maximaux Monotones. *North Holland*, Amsterdam.

[4] Dolezal V. (1979) Monotone Operators and Applications in Control and Network Theory, *Elsevier*, Amsterdam.

[5] Le Tallec P. (1994) Domain decomposition methods in computational mechanics. *Computational Mechanics Advances, North-Holland*, J.T. ODEN ed., 1: 121–220.

[6] Lions P.L. (1988) On the schwarz alternating method. 1. *First Intenational Symposium on Domain Decomposition Methods for Partial Differential Equations*, R. Glowinski and G. H. Golub and G. A. Meurant and J. Periaux ed., SIAM.

[7] Martinet B. (1971) Minimisation d'une fonctionnelle dans un espace produit par une méthode de relaxation. *Revue francaise d'informatique et de recherche operationnelle*, R-3: 121–126.

[8] Martinet B. (1972) Determination approchée d'un point fixe d'une application pseudo-contractante. *Compte rendu de l'accademie des sciences, Paris*, Serie A 274: 163–165.

[9] Minty G.J. (1964) On the monotonicity of the gradient of a convex function. *Pacific J. Math.*, 14: 243–247.

[10] Minty G.J. (1962) Monotone nonlinear operators in Hilbert space. *Duke Mathematics Journal*, 29: 341–346.

[11] Rockafellar R.T. (1976) Monotone operators and proximal point algorithm.*SIAM J. Control*, 14:5: 877–898.

[12] Tai X.-C. (1994) Parallel function decomposition and space decomposition methods, Part 1: Theory. *Report 94, Bergen University.*

[13] Xu J.C. (1992) Iterative methods by space decomposition and subspace correction. SIAM Review, 34: 4: 581–613.