

Parallel Space Decomposition for Minimization of Nonlinear Functionals

Andreas Frommer
University of Wuppertal
Department of Mathematics
D-42097 Wuppertal, Germany,
frommer@math.uni-wuppertal.de

Rosemary A. Renaut
Arizona State University
Department of Mathematics
Tempe, AZ 85287-1804
renaut@asu.edu

Abstract

We present an approach for parallel space decomposition which facilitates minimization of sufficiently smooth non-linear functionals with or without constraints on the variables. The framework for the spatial decomposition unites existing approaches from parallel optimization, **parallel variable distribution**, and finite elements, **Schwarz methods**. Additive and multiplicative algorithms based on the spatial decomposition are described. Convergence theorems are also presented, from which convergence for the case of convex functionals, and hence linear least squares problems, follows immediately.

1. Introduction

For our research we are concerned with the development of parallel algorithms for the solution of the problem

$$\min_{x \in X} f(x), \quad g(x) \geq 0, \quad (1)$$

where f and g are smooth nonlinear functionals on \mathbf{R}^n , and X is a closed nonempty set in \mathbf{R}^n . The work presented here provides **additive** and **multiplicative** algorithms for the solution of (1). A convergence theory provides for convergence when f is *convex* and the problem is constrained only by the requirement that $x \in X$, where X is also *convex*.

Approaches for parallel solution of (1) typically address the problem in one of two ways. There is a vast literature on appropriate sequential methods for the solution of optimization problems, with the format of the algorithms heavily dependent on the precise formulation and characteristic properties of the defining functionals. Intrinsic to most of these strategies are efficient kernels for iterative linear solvers. Thus, one approach

for parallelization is to focus on the most computationally intensive aspects of the algorithm and seek to parallelize these kernels. This has the advantage of maintaining all convergence characteristics of the underlying methods, which have been developed through many years of experience and research. On the other hand, some of the advantages of parallelization may be lost if significant portions of the code are still required to run in a sequential fashion. The alternative is to seek to design new algorithms for optimization which are particularly suited to parallelism and stand to offer high parallel efficiency. In this framework there is potential for the development of completely new strategies which could improve on current understanding of sequential formulations. At the same time, there is also the need to develop a theory to describe the convergence characteristics of any new method. Thus the second approach for parallelism does present a potentially greater improvement in parallel efficiency, but with the disadvantage that the significant progress that has been made in the understanding of sequential optimization may not carry over.

The work presented in this paper seeks to develop an approach for parallelism which is based on spatial decomposition and minimization on subdomains utilizing any method of choice for the subdomain minimizations. A key component of the resulting parallel algorithm also relies on an effective recombination, at minimal sequential cost, of local solutions to generate an iterative update of the global solution. Three methods for the update will be described. This work extends research by [9] and [8], on the use of iterative space decomposition methods for *strongly convex* functionals and provides a unifying framework which connects the approach to the ideas of [3] on the *parallel variable distribution* (PVD) algorithm.

In the next section we describe the spatial decompo-

sition and present the algorithms. Convergence results for the unconstrained case, both for exact and inexact local solutions, are given in Section 3. The constrained case is reviewed in Section 4. Proofs of all results, further discussion, and some numerical computations are discussed in [4].

2. Parallel Space decomposition

Throughout the whole paper we denote by $V_i, i = 1, \dots, m$, a collection of m (non-trivial) subspaces of $V = \mathbf{R}^n$ which span the whole of V , i.e.

$$V = \sum_{i=1}^m V_i. \quad (2)$$

We do not assume that this sum is direct, so a vector in V may have several different representations as a sum of components from the V_i . The $V_i, i = 1, \dots, m$ are termed a *space decomposition* of V . We consider all spaces V_i , as well as V , as being equipped with the Euclidean norm $\|\cdot\|$ from V .

With each of the spaces V_i we associate the linear and injective embedding operator $P_i : V_i \rightarrow V$ which maps x as an element of V_i on x as an element of V , and the corresponding surjective restriction $R_i = P_i^T : V \rightarrow V_i$. Properties of these operators are described in [4].

2.1. Algorithms

For the unconstrained problem (1) we formulate m unconstrained local minimization problems for the auxiliary functions $f_i : V_i \rightarrow \mathbf{R}, i = 1, \dots, m$,

$$\min_{y_i^k \in V_i} f_i(y_i) = \min_{y_i^k \in V_i} f(x + P_i^k y_i), \quad i = 1, \dots, m, \quad (3)$$

where $V = \sum_{i=1}^m V_i^k$ is a given sequence of space decompositions with corresponding prolongations P_i^k . Two iterative algorithms, one *additive* and the other *multiplicative*, for the solution of the unconstrained global minimization problem are based on the solution of these local minimizations. Here, and throughout, the superscript k on any variable indicates the value at the k th iteration, while index i is associated with the corresponding subspace V_i^k .

Algorithm 1 (*Additive variant for m processors*)

choose $x^0 \in V$ and $\beta_1, \dots, \beta_m > 0$ with $\sum_i \beta_i = 1$
 For $k = 0, 1, \dots$ (! until $\nabla f(x^k) = 0$)
 choose a space decomposition $V = \sum_{i=1}^m V_i^k$
 For $i = 1, \dots, m$

$$\text{Compute } y_i^k \in V_i^k \text{ such that} \\ f(x^k + P_i^k y_i^k) = \min_{y_i \in V_i^k} f(x^k + P_i^k y_i) \quad (4)$$

(! local minimization)
 $x^{i,k} = x^k + P_i^k y_i^k$
 Determine $\alpha_i^k, i = 1, \dots, m$, form update
 (! synchronization)

$$x^{k+1} = x^k + \sum_{i=1}^m \alpha_i^k P_i^k y_i^k \quad (5)$$

such that

$$f(x^{k+1}) \leq \sum_{i=1}^m \beta_i f(x^{i,k}) \quad (6)$$

End

End.

We consider three approaches for fulfilling (1) at each global update.

i) The *optimal* strategy. Determine $\alpha_1^k, \dots, \alpha_m^k$ by solving the m -dimensional minimization problem

$$\min_{(\alpha_1, \dots, \alpha_m) \in \mathbf{R}^m} f(x^k + \sum_i \alpha_i P_i^k y_i^k). \quad (7)$$

ii) The *selection* strategy. Determine t such that

$$f(x^{t,k}) = \min_{i=1}^m f(x^{i,k}).$$

Then, with $\alpha_t^k = 1$ and $\alpha_i^k = 1, i \neq t$,

$$x^{k+1} = x^{t,k}. \quad (8)$$

iii) The *convex combination* strategy. For convex f form the convex update

$$x^{k+1} = \sum_{i=1}^m \beta_i x^{i,k} = x^k + \sum_{i=1}^m \beta_i P_i^k y_i^k, \quad (9)$$

with $\alpha_i^k = \beta_i, i = 1, \dots, m$.

Note that this algorithm may also be seen as a coordinate descent method [1] if the choice $\alpha_i^k = 1$ is made for all choices of i and k . We do not consider this strategy for global update because the descent condition (1) need not be satisfied and hence convergence is not guaranteed. In addition, when (9) is used, the algorithm relates to the methods of [7], and [8], except that here the space decomposition is not required to be direct. When f is restricted to be a quadratic functional the above algorithm may also be interpreted as the classical additive Schwarz iterative solver for the equivalent linear system, e.g. [5].

A multiplicative version of the algorithm, which does not facilitate complete parallelism unless subproblem independence provides a suitable coloring, is given as follows:

Algorithm 2 (*Multiplicative variant*)

Choose $x^0 \in \mathbf{R}^n$,
For $k = 0, 1, \dots$, (! until $\nabla f(x^k) = 0$)
Choose a space decomposition $V = \sum_i V_i^k$.
For $i = 1, \dots, m$
Compute $y_i^k \in V_i^k$ such that

$$f\left(x^k + \omega \sum_{j=1}^{i-1} P_j^k y_j^k + P_i^k y_i^k\right)$$

$$= \min_{y_i \in V_i} f\left(x^k + \omega \sum_{j=1}^{i-1} P_j^k y_j^k + P_i^k y_i\right) \quad (10)$$
End

$$x^{k+1} = x^k + \omega \sum_{j=1}^m P_j^k y_j^k$$
End

Here $\omega \in \mathbf{R}$ is an *a priori* relaxation factor. The choice $\omega = 1$ corresponds to a Gauss-Seidel variant, whereas $\omega \neq 1$ gives the SOR variant of the algorithm.

3. Convergence Theory

3.1. The additive algorithm

We will now give convergence results for the additive algorithm and certain variants in the case that the functional f has a Lipschitz-continuous gradient, i.e. there exists $K > 0$ such that

$$\|\nabla f(y) - \nabla f(x)\| \leq K\|y - x\|, \quad \forall x, y \in \mathbf{R}^n. \quad (11)$$

We write this as $f \in LC_K^1(V)$. Note that we view the gradient primarily as a linear mapping from V to \mathbf{R}^n , i.e. as a ‘row vector’. Whenever we need to identify $\nabla f(x)$ with its dual from V we write $\nabla f(x)^T$.

We start with the following general result.

Theorem 1 (Convergence of the additive algorithm) *Let $f \in LC_K^1(V)$ be bounded from below and let $V = \sum_{i=1}^m V_i^k$ be a sequence of space decompositions such that*

$$\sum_{i=1}^m \|R_i^k x\|^2 \geq c \cdot \|x\|^2 \quad \text{for all } x \in V \text{ and } k = 0, 1, \dots \quad (12)$$

for some $c > 0$. Then every accumulation point of the sequence $\{x^k\}$ generated by Algorithm 1 is stationary and $\lim_{k \rightarrow \infty} \nabla f(x^k) = 0$.

A result on the rate of convergence for strongly convex f is then immediate from the definition of strong convexity and a result on linear convergence given in [3].

Corollary 1 *Assume that, in addition to the hypothesis of Theorem 1, the functional f is strongly convex with constant C . Then the sequence of iterates $\{x^k\}$*

converges to x^ , the unique minimizer of f , at the linear root rate*

$$\|x^k - x^*\| \leq \left(\frac{2}{C} (f(x^0) - f(x^k))\right)^{1/2} \left(1 - \frac{c\beta C^2}{K^2}\right)^{k/2}.$$

The PVD framework of [3] assumes a forget-me-not term in the local minimization by restricting each V_i^k to the form

$$V_i^k = W_i + W^k, \quad (13)$$

where the W_i form a non-overlapping orthogonal space decomposition of V with each W_i independent of k and spanned by Cartesian unit vectors. The spaces W^k are of dimension m and are spanned by vectors, one from each W_i . The following corollary then applies, and improves on the results given in [2], [3] and [6] by showing that the restrictions on the forget-me-not terms can be dispensed with.

Corollary 2 *Let $f \in LC_K^1(V)$ be bounded from below. Then every accumulation point of the sequence $\{x^k\}$ generated by the PVD algorithm is stationary and $\lim_{k \rightarrow \infty} \nabla f(x^k) = 0$. Moreover, if f is strongly convex, $\lim_{k \rightarrow \infty} x^k = x^*$, where x^* is the unique minimizer of f .*

3.2. Inexact local solutions

The results can also be relaxed to permit acceptance of inexact local solutions in (1) by observing that for the proof of Theorem 1 we need the local solutions to satisfy

$$f(x^k) - f(x^{i,k}) \geq \frac{1}{2K} \|R_i \nabla f(x^k)^T\|^2,$$

which is a local sufficient descent condition.

Theorem 2 (Convergence of the inexact additive algorithm) *Let $f \in LC_K^1(V)$ be bounded from below. Let $V = \sum_{i=1}^m V_i^k$ be a sequence of space decompositions such that*

$$\sum_{i=1}^m \|R_i^k x\|^2 \geq c \cdot \|x\|^2 \quad \text{for all } x \in V \text{ and } k = 0, 1, \dots,$$

for some $c > 0$. Let $\alpha > 0$ and assume that in Algorithm 1, instead of step (1) we accept ‘inexact’ solutions $x^{i,k}$ to the local minimization problem whenever

$$f(x^k) - f(x^{i,k}) \geq \alpha \cdot \|R_i^k \nabla f(x^k)^T\|^2. \quad (14)$$

Then every accumulation point of the sequence $\{x^k\}$ generated by the modified algorithm is stationary and $\lim_{k \rightarrow \infty} \nabla f(x^k) = 0$. Moreover, if f is strongly convex and its gradient is Lipschitz-continuous, then $\lim_{k \rightarrow \infty} x^k = x^*$, the unique minimizer of f .

This result then also provides the convergence result with *under* relaxation of *exact* local solutions for strongly convex functionals (see [7]), and with *over* relaxation for quadratic functionals.

Corollary 3 Assume that $f \in LC_K^1(V)$ is strongly convex and let the space decompositions $V = \sum_{i=1}^m V_i^k$ be as in Theorem 2.

- (i) Let $\gamma_1, \dots, \gamma_m$ be positive numbers such that $\gamma := \sum_{i=1}^m \gamma_i \leq 1$ and assume that for all k the synchronization step (1) in Algorithm 1 is replaced by

$$x^{k+1} = x^k + \sum_{i=1}^m \gamma_i P_i^k y_i^k. \quad (15)$$

Then $\lim_{k \rightarrow \infty} x^k = x^*$, the unique minimizer of f .

- (ii) In the special case when f is a quadratic functional part (i) holds with $\gamma \leq 1$ replaced by $\gamma < 2$.

3.3. The multiplicative algorithm

The following convergence results, largely equivalent to those for the additive variant, can also be obtained, but with in each case the assumption that f is strongly convex.

Theorem 3 (Convergence of the multiplicative algorithm) Let $f \in LC_K^1(V)$ be strongly convex. Assume that $V = \sum_{i=1}^m V_i^k$ is a sequence of space decompositions such that

$$\sum_{i=1}^m \|R_i^k x\|^2 \geq c \cdot \|x\|^2 \text{ for all } x \in V \text{ and } k = 0, 1, \dots,$$

for some $c > 0$. Then the sequence $\{x^k\}$ of iterates produced by Algorithm 2 with $\omega = 1$ converges to x^* , the unique minimizer of f in V .

The above theorem has been given in [7] for the case of minimization subject to block separable constraints on a closed convex set in V , and for which the space decomposition is independent of k . To illustrate just one new result covered by our general theorem, let us note that, due to (13), we get convergence for the multiplicative variant of the PVD method.

Theorem 4 (Convergence of inexact multiplicative algorithm) Let $f \in LC_K^1(V)$ be strongly convex. Let $V = \sum_{i=1}^m V_i^k$ be a sequence of space decompositions such that

$$\sum_{i=1}^m \|R_i^k x\|^2 \geq c \cdot \|x\|^2 \text{ for all } x \in V \text{ and } k = 0, 1, \dots,$$

for some $c > 0$. Let $\omega = 1$ and assume that in Algorithm 2, instead of step (2) we accept ‘inexact’ solutions y_i^k to the minimization problem whenever

$$f(x^{i-1,k}) - f(x^{i,k}) \geq \alpha \cdot \|R_i^k \nabla f(x^{i-1,k})^T\|^2, \quad (16)$$

$$i = 1, \dots, m,$$

where $x^{i,k} = x^k + \sum_{j=1}^i P_j^k y_j^k$, $i = 0, \dots, m$ and $\alpha > 0$ is fixed. Then $\lim_{k \rightarrow \infty} x^k = x^*$, the unique minimizer of f in V .

4. The Constrained PVD Algorithm

The PVD theory presented in [3] provides convergence of Algorithm 1 with synchronization steps (7) or (8) for the case of the block separable constrained problem, i.e., for which X in (1) is assumed to be a Cartesian product of closed convex sets, [3, Theorem 3.7]. In order to solve the constrained problem it is necessary to introduce modifications to the algorithms that incorporate the appropriate constraint conditions. For now we consider the general convex-constrained case, where X is a closed convex set, from which the result for the block separable case can be deduced. Then in Algorithm 1 the local minimizations (1) are replaced by

$$\begin{aligned} \min_{y_i \in V_i} f_i(y_i) &= \min_{y_i \in V_i} f(x^k + P_i y_i) \\ x^k + P_i y_i &\in X \quad x^k + P_i y_i \in X. \end{aligned} \quad (17)$$

Convergence for the constrained problem is defined in terms of the convergence to a stationary point \bar{x} at which the minimum principle necessary optimality condition

$$\bar{x} \in X \quad \text{and} \quad \nabla f(\bar{x})(y - \bar{x}) \geq 0 \quad \forall y \in X \quad (18)$$

is satisfied. Equivalently,

$$r(\bar{x}) = 0, \quad (19)$$

where r is the projected gradient residual function defined by

$$r(x) := x - [x - \nabla f(x)^T]_+, \quad (20)$$

and $[x]_+$ represents the orthogonal projection map for element $x \in V$ onto the set X .

We thus obtain convergence for general convex X .

Theorem 5 (Convergence for general convex X) Let $f \in LC_K^1(V)$ be strongly convex. Then the sequence of iterates $\{x^k\}$ produced by Algorithm 1 with local minimizations (17) converges to x^* , the unique minimizer of $f(x)$ over $X \subset V$.

Extensions for inexact local solutions, and to give rate of convergence results, may also be derived as in the unconstrained case.

4.1. Acknowledgments

The research of the second author was supported under grant NSF DMS9402943.

References

- [1] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation*, Prentice Hall, Englewood Cliffs, New Jersey, 1989.
- [2] J.E. Dennis and T. Steihaug, *A Ferris-Mangasarian Technique Applied to Linear Least Squares Problems*, CRPC-TR 98740, Center for Research on Parallel Computation, Rice University, 1998.
- [3] M.C. Ferris and O.L. Mangasarian, *Parallel Variable Distribution*, SIAM J. Optim. **4** (1994), 815–832.
- [4] A. Frommer and R. A. Renaut, *A Unified Approach to Parallel Space Decomposition Methods*, 1999.
- [5] W. Hackbusch, *Iterative Methods for Large Sparse Linear Systems*, Springer, Heidelberg, 1993.
- [6] M.V. Solodov, *New Inexact Parallel Variable Distribution Algorithms*, Comp. Optim. and Appl. **7** (1997), 165–182.
- [7] X.C. Tai, *Parallel function and space decomposition methods—Part II, Space decomposition*, Beijing Mathematics, 1, Part 2, (1995), 135–152.
- [8] X.C. Tai and M. Espedal, *Rate of Convergence of Some Space Decomposition Methods for Linear and Nonlinear Problems*, SIAM J. Numer. Anal. **35** (1998), 1558-1570 .
- [9] J. Xu, *Iterative Methods by Space Decomposition and Subspace Correction*, SIAM Review **34** (1992), 581–613.