JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Journal of Computational and Applied Mathematics 110 (1999) 205–223

www.elsevier.nl/locate/cam

A unified approach to parallel space decomposition methods

Andreas Frommer[a,*], Rosemary A. Renaut[b]

[a] *Department of Mathematics, University of Wuppertal, D-42097 Wuppertal, Germany*

[b] *Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804, USA*

Received 15 February 1999

Abstract

We consider (relaxed) additive and multiplicative iterative space decomposition methods for the minimization of sufficiently smooth functionals without constraints. We develop a general framework which unites existing approaches from both parallel optimization and finite elements. Specifically this work unifies earlier research on the parallel variable distribution method in minimization, space decomposition methods for convex functionals, algebraic Schwarz methods for linear systems and splitting methods for linear least squares. We develop a general convergence theory within this framework, which provides several new results as well as including known convergence results. © 1999 Elsevier Science B.V. All rights reserved.

MSC: 65H10

Keywords: Space decomposition methods; Block Jacobi; Block SOR; Multisplittings; Parallel computation; Parallel variable distribution; Minimization; Coordinate descent; Finite elements

1. Introduction

In this paper we consider parallel iterative methods for the solution of the unconstrained minimization problem

$$\min_{x \in \mathbf{R}^n} f(x), \tag{1.1}$$

where $f(x)$ has Lipschitz continuous first partial derivatives in $\mathbf{R}^n$, $f \in C^1(\mathbf{R}^n)$. We also consider the application of the methods for *convex* $f(x)$, and where $f(x)$ is a *quadratic functional* of the form

$$f(x) = \tfrac{1}{2}x^{\mathrm{T}}Ax - b^{\mathrm{T}}x,\ A \in \mathbf{R}^{n\times n},\ b \in \mathbf{R}^n \tag{1.2}$$

* Corresponding author.

E-mail addresses: frommer@math.uni-wuppertal.de (A. Frommer), renaut@asu.edu (R.A. Renaut)

PII: S0377-0427(99)00235-6

with the matrix A being *symmetric and positive definite* (SPD). The functional of (1.2) has a unique minimizer x^* which solves the linear system $Ax=b$. Equivalently, these methods are appropriate for solution of the full rank linear least-squares problem

$$\min_{x\in\mathbf{R}^n} \|Lx-b\|_2,\ b\in\mathbf{R}^p,\ L\in\mathbf{R}^{p\times n},\ rk(L)=n\leqslant p. \tag{1.3}$$

Space decomposition methods have a long history, particularly as a means of preconditioning linear systems arising from discretizations of elliptic boundary value problems, in which context they are commonly called *additive* or *multiplicative* Schwarz preconditioners, see the monographs [7,11,17], e.g. In the present paper we apply the same space decomposition principle in order to obtain (parallel) iterative methods for minimizing f. Our work is based on [15,17], where such iterative space decomposition methods have been studied for general *strongly convex* functionals, and on [4] where the *parallel variable distribution* (PVD) algorithm was introduced. We show that for quadratic functionals the convergence results from [15] can be extended to allow for a certain amount of over-relaxation. As a special case, our results will show that an *overlapping* block SOR method converges whenever the relaxation factor ω satisfies $\omega\in(0,2)$. Our results also improve earlier investigations on parallel methods for linear least squares problems considered in [10], and they contribute a new convergence theorem to iterative multisplitting methods for solving SPD linear systems. We also give some results on the (linear) rate of convergence of these methods. Note that these results do not consider the possible benefits of a ‘coarse grid correction’ so that in this respect they are weaker than some of standard results for the additive Schwarz method for (linear) elliptic boundary value problems (see [7], e.g.).

The remainder of this paper is organized as follows: Section 2 develops the space decomposition principle and states the two basic algorithms, additive and multiplicative. Sections 3 and 4 detail the convergence theory for the additive and multiplicative variants, respectively, while Section 5 is devoted to a discussion of the relationship of our results to previous work, particularly linear least-squares and multisplitting methods. Auxiliary required results are presented in the appendix. In a future paper we will describe the extension of this theory to include results for constrained problems, cf. [3,4,13].

2. Space decomposition

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

$$V=\sum_{i=1}^{m} V_i. \tag{2.1}$$

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 $P_i: V_i\to V$ which maps x as an element of V_i on x as an element of V, and the corresponding surjective restriction $R_i=P_i^{\mathrm{T}}: V\to V_i$. The following lemma collects some useful properties of these operators:

Lemma 2.1. *For $i=1,\dots,m$ we have*
(i) $\|P_i\|=1$,
(ii) $R_iP_ix=x$ *whenever* $x\in V_i$,
(iii) *If* $R_ix=0$ *for* $i=1,\dots,m$ *then* $x=0$.
(iv) $\langle P_iR_ix, x-P_iR_ix\rangle=0$ *for all* $x\in V$.
(v) *There exists a constant* $c>0$ *such that*

$$\sum_{i=1}^{m}\|R_ix\|^2\geqslant c\|x\|^2 \text{ for all } x\in V.$$

(vi) *Let* $c(V_1,\dots,V_m)$ *be the largest possible constant* c *in* (v) *for a given space decomposition* $V_1,\dots,V_m$. *Then, if* $W_1,\dots,W_{m'}$ *is another space decomposition such that* $W_i\leqslant V_i$, $i=1,\dots,m'(\leqslant m)$, *we have*

$$c(W_1,\dots,W_{m'})\leqslant c(V_1,\dots,V_m). \tag{2.2}$$

Moreover, if the spaces W_i *are mutually orthogonal, then* $c(W_1,\dots,W_{m'})=1$.

Proof. (i) and (ii) are truly trivial. To show (iii) observe that since $\ker(R_i)=\operatorname{range}(P_i)^\perp$ we have $x\perp\operatorname{range}(P_i)$, $i=1,\dots,m$. Due to (2.1) this implies $x=0$. For (iv) we use (ii) to see that $\langle P_iR_ix,P_iR_ix\rangle=\langle P_iR_iP_iR_ix,x\rangle=\langle P_iR_ix,x\rangle$. Part (v) follows by taking c as the minimum of the continuous and positive mapping $x\rightarrow\sum_{i=1}^{m}\|R_ix\|^2$ on the n-dimensional unit sphere. Finally, for (vi) suppose that S_i and Q_i are the restriction and embedding operators for W_i, and note that Q_iS_i and P_iR_i are the orthogonal projections on W_i and V_i, respectively. Then, by $W_i\leqslant V_i$, $\|S_ix\|=\|Q_iS_ix\|\leqslant\|P_iR_ix\|=\|R_ix\|$ for all $x\in V$, and (2.2) follows. For the W_i mutually orthogonal we have $\sum_{i=1}^{m}\|R_ix\|^2=\|x\|^2$ by Pythagoras' theorem. □

Given a sequence of space decompositions $V=\sum_{i=1}^{m}V_i^k$, $k=0,1,\dots$, with corresponding prolongations P_i^k we formulate m unconstrained local minimization problems for the functions $f_i: V_i^k\rightarrow\mathbf{R}$, $i=1,\dots,m$, depending on a given vector $x\in V$,

$$\min_{y_i\in V_i^k} f_i(y_i)=\min_{y_i\in V_i^k} f(x+P_i^ky_i),\quad i=1,\dots,m. \tag{2.3}$$

(For notational convenience we do not explicitly denote the dependence of f_i and y_i on k.)

Two iterative algorithms, one *additive* and the other *multiplicative*, for the solution of the unconstrained global minimization problem will be based on the solution of these local minimizations. Note that throughout the superscript k on any variable indicates the value at the kth iteration, while index i is associated with the corresponding subspace V_i^k. We always assume that the local minimization problems (2.3) admit a solution.

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 $i=1,\dots,m$
    Compute $y_i^k\in V_i^k$ such that

$$f(x^k+P_i^ky_i^k)=\min_{y_i\in V_i^k} f(x^k+P_i^ky_i)\qquad \text{(! local minimization)} \tag{2.4}$$

$x^{i,k} = x^k + P_i^k y_i^k$

`End`

`Determine` $\alpha_i^k, i = 1, \ldots, m$, `form update` (`! synchronization`)

$$x^{k+1} = x^k + \sum_{i=1}^{m} \alpha_i^k P_i^k y_i^k \tag{2.5}$$

`such that`

$$f(x^{k+1}) \leqslant \sum_{i=1}^{m} \beta_i f(x^{i,k}) \tag{2.6}$$

`End.`

The descent condition (2.6) imposed in the synchronization step will turn out to be crucial to the convergence of the algorithm.

We consider three approaches for fulfilling (2.6) which provide three possible ways of determining suitable parameters α_i^k in (2.5).

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

$$\min_{(\alpha_1,\ldots,\alpha_m) \in \mathbf{R}^m} f(x^k + \sum_i \alpha_i P_i^k y_i^k). \tag{2.7}$$

(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 $\alpha_i^k = 0$ for $i \neq t$,

$$x^{k+1} = x^{t,k}. \tag{2.8}$$

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

$$x^{k+1} = \sum_{i=1}^{m} \beta_i x^{i,k} = x^k + \sum_{i=1}^{m} \beta_i P_i^k y_i^k \tag{2.9}$$

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

Note that with the last two strategies the synchronization step becomes particularly cheap. This may be important on parallel computers where synchronization steps can cause a bottleneck.

Algorithm 1 is closely related to many space decomposition methods in the literature which take the same space decomposition for all k. The simplest of these are the coordinate descent methods [1,9], which may be implemented in additive (block Jacobi) or multiplicative (Gauss–Seidel) form. The former is not guaranteed to converge because the descent condition (2.6) need not be satisfied. The space decomposition methods for convex functionals considered in [14,15] use the convex combination strategy (2.9). If f is restricted to be the quadratic functional (1.2), the above algorithm may be interpreted as the classical additive Schwarz iterative solver for the linear system $Ax = b$, see, e.g. [7].

The PVD method of [4] which includes a "forget-me-not" term in the local minimizations can also be formulated as a special case of the above algorithm. Both the optimal and the selection strategy were considered in [4]. Note that the synchronization step in [4] requires $f(x^{k+1}) \leqslant \min_{i=1}^{m} f(x^{i,k})$ which is less general than (2.6). The PVD formulation restricts the V_i^k to the form

$$V_i^k = W_i + W^k, \tag{2.10}$$

where the W_i form a nonoverlapping orthogonal space decomposition of V with each W_i being independent of k and spanned by Cartesian unit vectors. The spaces W^k (which depend on k but not on i) are of dimension m and spanned by m vectors, one from each W_i. The motivation behind this structure is to incorporate the forget-me-not term which provides global information from step $k-1$. For example the global steepest descent direction s^k is incorporated into the subspaces V_i^k by taking W^k as the space spanned by the orthogonal projections of s^k onto the W_i, e.g.,

$$W^k = \text{span}\{R_1 s^k, \ldots, R_m s^k\}. \tag{2.11}$$

Note that our theory does not preclude the W_i forming an overlapping space decomposition of V. Moreover, algorithm PVD0 of [4] which does not add the forget-me-not term is covered by taking $V_i^k = W_i$. The application of PVD0 for the solution of the linear least squares problem with update strategies (2.7)–(2.9) was considered by Renaut [10]. For this same problem, Dennis and Steihaug [2] motivate the choice $s^k = x^k - x^{k-1}$ for the forget-me-not-term in PVD as being close to optimal. This last paper also proposes a line search strategy as yet another way of satisfying (2.6).

If, instead of working in parallel, we perform the individual minimization steps one after the other, we arrive at the following multiplicative algorithm.

ALGORITHM 2. (*Multiplicative variant*)

Choose $x^0 \in V$,

For $k = 0, 1, \ldots,$ (! until $\nabla f(x^k) = 0$)

  Choose a space decomposition $V = \sum_i V_i^k$.

  For $i = 1, \ldots, m$

    Compute $y_i^k \in V_i^k$ such that

$$f\left(x^k + \omega \sum_{j=1}^{i-1} P_j^k y_j^k + P_i^k y_i^k\right) = \min_{y_i \in V_i} f\left(x^k + \omega \sum_{j=1}^{i-1} P_j^k y_j^k + P_i^k y_i\right) \tag{2.12}$$

  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. The more general term "multiplicative" follows from the terminology for the Schwarz method in preconditioning PDE solvers. For $\omega \neq 1$ it is of 'SOR-type', whereas the additive variant can be considered to be of "JOR" (relaxed Jacobi) type. The multiplicative variant is not, in general, amenable to parallel execution, unless a certain degree of independence exists and hence permits a *coloring* (see Section 3.3) to be applied across the spatial decomposition.

3. Convergence theory for the additive variant

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

$$\|\nabla f(y)-\nabla f(x)\|\leqslant K\|y-x\|,\quad \forall x,y\in \mathbf{R}^n. \tag{3.1}$$

We write this as $f\in LC^1_K(V)$. 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 $\nabla f(x)^{\mathrm{T}}$.

3.1. The additive algorithm

We start with the following general result.

Theorem 3.1 (Convergence of the additive algorithm). *Let $f\in LC^1_K(V)$ be bounded from below and let $V=\sum_{i=1}^m V_i^k$ be a sequence of space decompositions such that*

$$\sum_{i=1}^m \|R_i^k x\|^2\geqslant c\cdot\|x\|^2 \quad \textit{for all } x\in V \textit{ and } k=0,1,\ldots \tag{3.2}$$

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

Proof. The gradient of the auxilliary function f_i, defined by (2.3), is related to that of f via

$$\nabla f_i(y_i)=\left(\nabla f(x^k+P_iy_i)\right)P_i. \tag{3.3}$$

Now since f has a Lipschitz continuous gradient with constant K we get

$$\begin{aligned}\|\nabla f_i(y_i)-\nabla f_i(\bar y_i)\| &= \|(\nabla f(x^k+P_iy_i)-\nabla f(x^k+P_i\bar y_i))P_i\|\\ &\leqslant \|P_i\|\cdot K\cdot\|P_i(y_i-\bar y_i)\|\\ &\leqslant \|P_i\|^2\cdot K\cdot\|y_i-\bar y_i\|\end{aligned}$$

which shows that ∇f_i is Lipschitz-continuous, again with constant K, since $\|P_i\|=1$ by Lemma 2.1 (i). For $i=1,\ldots,m$ let z_i^k be the point $z_i^k=-(1/K)\cdot\nabla f_i(0)^{\mathrm{T}}$. Then by the second part of the Quadratic Bound Lemma A.6 (see the appendix)

$$f_i(0)-f_i(z_i^k)\geqslant\frac{1}{2K}\|\nabla f_i(0)^{\mathrm{T}}\|^2.$$

By the minimization, $f_i(z_i^k)\geqslant f_i(y_i^k)$ so that

$$f(x^k)-f(x^{i,k})=f_i(0)-f_i(y_i^k)\geqslant f_i(0)-f_i(z_i^k)\geqslant\frac{1}{2K}\|\nabla f_i(0)^{\mathrm{T}}\|^2,$$

which, using (3.3), results in

$$f(x^k) - f(x^{i,k}) \geqslant \frac{1}{2K}\|(\nabla f(x^k)P_i)^{\mathrm{T}}\|^2 = \frac{1}{2K}\|R_i \nabla f(x^k)^{\mathrm{T}}\|^2. \tag{3.4}$$

Multiplying with β_i, summing up and using (3.2) we obtain

$$f(x^k) - \sum_{i=1}^{m} \beta_i f(x^{i,k}) \geqslant \sum_{i=1}^{m} \frac{\beta_i}{2K}\|R_i \nabla f(x^k)^{\mathrm{T}}\|^2 \geqslant \frac{c\beta}{2K}\|\nabla f(x^k)^{\mathrm{T}}\|^2,$$

where $\beta = \min_{i=1}^{m} \beta_i > 0$. Now, by part (2.6) of the synchronization

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

yielding

$$f(x^k) - f(x^{k+1}) \geqslant \frac{c\beta}{2K}\|\nabla f(x^k)^{\mathrm{T}}\|^2.$$

Hence, by the Linear Convergence Lemma A.5 every accumulation point $\bar{x}$ of $\{x^k\}$ is stationary, $\nabla f(\bar{x}) = 0$. □

From the Linear Convergence Lemma A.5 we directly obtain the following corollary.

Corollary 3.2. *Assume that, in addition to the hypothesis of Theorem* 3.1, *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*

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

This result contains, in particular, the convergence results from [14] for the space decomposition methods considered there, i.e. when the space decomposition is independent from k and the convex combination strategy (2.9) is used. Note that in this case, by Lemma 2.1 (v), condition (3.2) is trivially fulfilled.

Within the PVD framework of [4], i.e. where we have the special space decompositions (2.10), condition (3.2) is again fulfilled, see Lemma 2.1 (vi). This yields the following additional corollary.

Corollary 3.3. *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* $\lim_{k\to\infty} \nabla f(x^k) = 0$. *Moreover, if f is strongly convex,* $\lim_{k\to\infty} x^k = x^*$, *where x^* is the unique minimizer of f.*

Corollary 3.3 improves upon the convergence results given in [2,4,12] by showing that we can entirely dispense with an additional assumption present there, namely that the vectors $R_i s^k$, $k=0,1,\ldots$ from (2.11) be bounded for $i=1,\ldots,m$. Moreover, it also immediately provides for the convergence result for the quadratic functional (1.2).

3.2. Inexact local solutions

In practical situations it might be difficult to solve the local minimization problems in (2.4) exactly. There might also be situations where exact local solutions will not yield the most rapidly convergent overall algorithm. Therefore results which do not assume exact local solutions are of interest. This was addressed by Solodov [12], for the PVD algorithm with a specific choice of search direction. There, local solutions are required to belong to an ε-stationary set of the subproblem. Here we relax this condition and observe that in the proof of Theorem 3.1 we showed that exact solutions $x^{i,k}$ of the local minimization problems satisfy (3.4), i.e.

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

which is a local sufficient descent condition. This is the only property of the solutions used in that proof and, in addition, this is also the only place where we made use of the Lipschitz-continuity of the gradient. Thus, making (3.4) part of the assumption, we obtain the following new theorem.

Theorem 3.4 (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*

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

for some $c > 0$. Let $\alpha > 0$ and assume that in Algorithm 1, instead of step (2.4) we accept "inexact" solutions $x^{i,k}$ to the local minimization problem whenever

$$f(x^k) - f(x^{i,k}) \geqslant \alpha \|R_i^k \nabla f(x^k)^{\mathrm{T}}\|^2. \tag{3.5}$$

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

We note that under assumption (3.5) Corollary 3.2 still applies. This is weaker than the rate of convergence result, Theorem 1 in [12], because the dependence on the number of processors is still present in the β term but at the same time the product $c\beta \leqslant 1$ for $\beta = 1/m$, because $c \leqslant m$. On the other hand the result is less restrictive because the sufficient descent condition (3.5) is all that is required for the local solutions.

As an interesting application of this theorem we are now able to state the following convergence result with *under* relaxation of *exact* local solutions for strongly convex functionals (see [14]), and with *over* relaxation for quadratic functionals.

Corollary 3.5. *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 3.4.*

(i) *Let* $\gamma_1,\dots,\gamma_m$ *be positive numbers such that* $\gamma := \sum_{i=1}^m \gamma_i \leqslant 1$ *and assume that for all* k *the synchronization step* (2.5) *in Algorithm* 1 *is replaced by*

$$x^{k+1} = x^k + \sum_{i=1}^m \gamma_i P_i^k y_i^k. \tag{3.6}$$

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

(ii) *In the special case of* f *being the quadratic functional* (1.2) *part* (i) *holds with* $\gamma \leqslant 1$ *replaced by* $\gamma < 2$.

Proof. We view the new synchronization step (3.6) as a convex combination of inexact local solutions via

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

where $\beta_i = \gamma_i/\gamma > 0,\ i = 1,\dots,m$, and $x^k + \gamma P_i^k y_i^k$ is an inexact local solution. Therefore, all we have to show is the validity of (3.5).

In case (i) let us first observe that for any convex functional, and any $x, y \in V,\ \gamma \in [0,1]$ we have

$$f(x+\gamma y) = f((1-\gamma)x + \gamma(x+y)) \leqslant (1-\gamma)f(x) + \gamma f(x+y),$$

which implies

$$f(x) - f(x+\gamma y) \geqslant \gamma(f(x) - f(x+y)).$$

Applying this for $x = x^k,\ y = P_i^k y_i^k$ and using (3.4) from the proof of Theorem 3.1 we get

$$f(x^k) - f(x^k + \gamma P_i^k y_i^k) \geqslant \frac{\gamma}{2K} \|R_i^k \nabla f(x^k)^{\mathrm{T}}\|^2.$$

This proves (3.5) (with $\alpha = \gamma/K$).

In case (ii) we use Lemma A.3 which directly shows

$$f(x^k) - f(x^k + \gamma P_i^k y_i^k) \geqslant \frac{\gamma(2-\gamma)}{2\lambda_{\max}} \|R_i^k \nabla f(x^k)^{\mathrm{T}}\|^2.$$

Therefore, (3.5) again holds (with $\alpha = \gamma(2-\gamma)/(2\lambda_{\max})$). □

3.3. Colorings

Part (i) of Corollary 3.5 above proves convergence in particular for the choice $\gamma_i = \omega \in (0, 1/m]$, $i = 1,\dots,m$ ($\omega \in (0, 2/m)$ in part (ii)). As is well known from the additive Schwarz theory for linear systems (see [7]) the admissible interval $(0, 2/m)$ can be substantially enlarged if the subspaces V_i^k are chosen commensurately with the sparsity structure of the coefficient matrix. We will formulate a corresponding result for the general nonlinear case using the concept of a coloring.

Definition 3.6. Let $\{W_i, i = 1,\dots,r\}$ be a set of subspaces of V with embedding operators Q_i and let $W_0 = \sum_{i=1}^r W_i$ with embedding operator Q_0. Then this set of subspaces is said to *represent a color* (with respect to f), if the following implication holds for all $x \in V$:

$$y_i \text{ solves } \min_{z_i \in W_i} f(x + Q_i z_i) \quad \textit{for } i = 1,\dots,r, \Rightarrow y_0 = Q_0^{\mathrm{T}} \sum_{i=1}^{r} Q_i y_i \text{ solves } \min_{z_0 \in W_0} f(x + Q_0 z_0).$$

A *coloring with* m_c *colors* of a space decomposition $V = \sum_{i=1}^{m} V_i$ is a partitioning of the set $\{V_1,\dots,V_m\}$ into m_c nonempty and pairwise disjoint subsets C_j such that for each $j = 1,\dots,m_c$ the subspaces from C_j represent a color (with respect to f).

Taking $C_j = \{V_j\}$, $j = 1,\dots,m$ shows that, trivially, a space decomposition with m subspaces admits a coloring with m colors. In discretizations of partial differential equations one usually obtains colorings with a fixed small number of colors, independently of m; see [7]. This comes from the fact that the Hessian of f is then sparsely structured. More precisely, we have the following general result.

Lemma 3.7. *Let* $f \in C^2(V)$ *be strongly convex and denote* $H(x)$ *its Hessian at the point* x. *Let* $W_1,\dots,W_r$ *be subspaces of* V *with corresponding embeddings* Q_i *such that for all* $x \in V$ *we have*

$$Q_i^{\mathrm{T}} H(x) Q_j = 0, \quad i,j = 1,\dots,r,\ i \neq j, \tag{3.7}$$

i.e. the corresponding "off-diagonal blocks" of $H(x)$ *are zero. Then* $\{W_1,\dots,W_r\}$ *represents a color (with respect to* f*).*

Proof. Assume that for $i = 1,\dots,r$ the point $y_i \in W_i$ solves $\min_{z_i \in W_i} f(x + Q_i z_i)$. Since f is strongly convex, y_i is the unique point satisfying

$$\nabla f(x + Q_i y_i) Q_i = 0, \quad i = 1,\dots,r. \tag{3.8}$$

Define $y_0 = Q_0^{\mathrm{T}} \sum_{i=1}^{r} Q_i y_i$ with Q_0 the embedding for $W_0 = \sum_{i=1}^{r} W_i$. Moreover, for i fixed, let us write $x' = x + Q_i y_i$ and $\sum{}' Q_j y_j = \sum_{j=1, j\neq i}^{r} Q_j y_j$. Since $Q_0 y_0 = \sum_{i=1}^{r} Q_i y_i$ we have

$$\begin{aligned}\nabla f(x + Q_0 y_0) &= \nabla f\left(x' + \sum{}' Q_j y_j\right)\\ &= \nabla f(x') + \int_0^1 \left(\sum{}' Q_j y_j\right)^{\mathrm{T}} H\left(x' + t \sum{}' Q_j y_j\right) \mathrm{d}t.\end{aligned}$$

This implies

$$\nabla f(x + Q_0 y_0) Q_i = \nabla f(x') Q_i + \int_0^1 \left(\sum{}' Q_j y_j\right)^{\mathrm{T}} H\left(x' + t \sum{}' Q_j y_j\right) Q_i \, \mathrm{d}t.$$

Since the subspaces represent a color, we have $Q_j^{\mathrm{T}} H(x' + t \sum{}' Q_j y_j) Q_i = 0$ for $t \in [0,1]$. Therefore, the term inside the integral vanishes completely so that we have

$$\nabla f(x + Q_0 y_0) Q_i = 0, \quad i = 1,\dots,r.$$

Due to $W_0=\sum_{i=1}^{r} W_i$ this yields $\nabla f(x+Q_0y_0)Q_0=0$ which, since f is strongly convex, shows that y_0 is the unique solution to $\min_{z_0\in W_0} f(x+Q_0z_0)$. □

Note that for the quadratic functional (1.2) we have $H(x)=A$ for all x, so that (3.7) is just a sparsity condition on A.

Now, assume that each space decomposition $V=\sum_{i=1}^{m} V_i^k$ in Algorithm 1 admits a coloring $C_1^k,\ldots,C_{m_c}^k$ with m_c colors (m_c independent of k). Let $\bar{P}_i^k$ denote the trivial embedding for the subspace

$$\bar{V}_i^k=\sum_{V_j^k\in C_i^k} V_j^k.$$

In light of Definition 3.6 an iterate

$$x^{k+1}=x^k+\sum_{i=1}^{m}\omega P_i^k y_i^k \quad \text{where } y_i^k \text{ solves } \min_{z_i\in V_i^k} f(x+P_iz_i)$$

(ω fixed) can alternatively be written as

$$x^{k+1}=x^k+\sum_{i=1}^{r}\omega \bar{P}_i^k \bar{y}_i^k \quad \text{where } \bar{y}_i^k \text{ solves } \min_{\bar{z}_i\in \bar{V}_i^k} f(x+\bar{P}_i^k\bar{z}_i).$$

Also note that $c(\bar{V}_1^k,\ldots,\bar{V}_{m_c}^k)\geqslant c(V_1^k,\ldots,V_m^k)$ by Lemma 2.1 (vi). Therefore, if we take $\omega\in(0,1/m_c]$ we are in the situation of Corollary 3.5 (i) with the space decompositions $V=\sum_{i=1}^{m_c}\bar{V}_i^k$. This observation thus yields the following theorem.

Theorem 3.8. *Assume that $f\in LC_K^1(V)$ is strongly convex and that each space decomposition $V=\sum_{i=1}^{m} V_i^k$ admits a coloring with m_c colors and that*

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

with c independent of k. Let $\omega\in(0,1/m_c]$ and assume further that we take

$$x^{k+1}=x^k+\sum_{i=1}^{k}\omega P_i^k y_i^k$$

in the synchronization step of Algorithm 1. *Then the iterates x^k converge to x^*, the unique minimizer of f.*

The following corollary for quadratic functionals follows in a similar manner as part (ii) of Corollary 3.5. We thus refrain from reproducing a proof for this result which can also be found in [7].

Corollary 3.9. *In the case of the quadratic functional* (1.2), *Theorem* 3.8 *remains valid if we take ω from the larger interval* $(0,2/m_c)$.

4. Convergence theory for the multiplicative variant

We now turn to study the convergence of the multiplicative variant (Algorithm 2). We immediately consider the case of a strongly convex functional, since we cannot expect an analogue of Theorem 3.1 to hold.

Theorem 4.1 (Convergence of the multiplicative algorithm). *Let $f \in LC^1_K(V)$ be strongly convex. Assume that $V = \sum_{i=1}^m V_i^k$ is a sequence of space decompositions such that*

$$\sum_{i=1}^{m} \|R_i^k x\|^2 \geqslant c\|x\|^2 \quad \textit{for all } x \in V \textit{ and } k = 0, 1, \ldots$$

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.

Proof. Let $x^{i,k}$ denote $x^k + \sum_{j=1}^{i} P_j^k y_j^k$, $i = 0, \ldots, m$, so that, in particular, $x^{m,k} = x^{k+1} = x^{0,k+1}$. We will show the following:
(i) $f(x^{i-1,k}) \geqslant f(x^{i,k})$ for $i = 1, \ldots, m, k = 0, 1, \ldots.$
(ii) $\lim_{k\to\infty} (x^{i-1,k} - x^{i,k}) = 0$ and $\lim_{k\to\infty} (x^{i,k} - x^k) = 0$, $i = 1, \ldots, m.$
(iii) $\lim_{k\to\infty} \nabla f(x^k) = 0.$
(iv) $\lim_{k\to\infty} x^k = x^*.$

Part (i) is trivial by the definition of $x^{i,k}$ in Algorithm 2. To show (ii) we notice that by (i) we have $f(x^0) = f(x^{0,0}) \geqslant f(x^{1,0}) \geqslant \cdots \geqslant f(x^{m,0}) = f(x^{0,1}) \geqslant \cdots \geqslant f(x^{m,1}) = f(x^{0,2}) \geqslant \cdots$, and this monotonically decreasing sequence is bounded from below by $f(x^*)$. Therefore we have, in particular,

$$\lim_{k\to\infty} (f(x^{i-1,k}) - f(x^{i,k})) = 0, \quad i = 1, \ldots, m. \tag{4.1}$$

By strong convexity (see Definition A.1)

$$f(x^{i-1,k}) - f(x^{i,k}) \geqslant -\nabla f(x^{i,k}) P_i^k y_i^k + \frac{C}{2}\|P_i^k y_i^k\|^2.$$

Since y_i^k minimizes $f_i^k(z_i^k) := f(x^{i-1,k} + P_i^k z_i^k)$ we have

$$0 = \nabla f_i^k(y_i^k) = \nabla f(x^{i,k}) P_i^k. \tag{4.2}$$

Therefore,

$$f(x^{i-1,k}) - f(x^{i,k}) \geqslant \frac{C}{2}\|P_i^k y_i^k\|^2,$$

and by (4.1) this yields

$$\lim_{k\to\infty} \|P_i^k y_i^k\| = 0, \quad i = 1, \ldots, m$$

which shows (ii). To prove (iii) observe that in

$$R_i^k \nabla f(x^k)^{\mathrm{T}} = R_i^k(\nabla f(x^k)^{\mathrm{T}} - \nabla f(x^{i,k})^{\mathrm{T}}) + R_i^k \nabla f(x^{i,k})^{\mathrm{T}},$$

we have $\|R_i^k\| = 1$, and $R_i^k \nabla f(x^{i,k})^{\mathrm{T}} = 0$ by (4.2). By continuity of ∇f, and using (ii), we obtain

$$\lim_{k\to\infty} R_i^k \nabla f(x^k)^{\mathrm{T}} = 0.$$

But, by assumption,

$$\sum_{i=1}^{m} \|R_i^k \nabla f(x^k)^{\mathrm{T}}\|^2 \geqslant c \cdot \|\nabla f(x^k)\|^2.$$

Hence $\lim_{k\to\infty} \nabla f(x^k) = 0$, i.e. (iii). Finally, for (iv) observe first that $\nabla f(x^*) = 0$, so that by strong convexity (second part of Definition A.1) we have

$$(\nabla f(x^k) - \nabla f(x^*))(x^k - x^*) = \nabla f(x^k)(x^k - x^*) \geqslant C \cdot \|x^k - x^*\|^2.$$

The Cauchy–Schwarz inequality then yields

$$\|\nabla f(x^k)\| \cdot \|x^k - x^*\| \geqslant C \cdot \|x^k - x^*\|^2$$

from which, using (iii), we get (iv). □

The above theorem has been given in [14] 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 (2.10), we get convergence for the multiplicative variant of the PVD method.

As for the additive variant we can also formulate a convergence result for the multiplicative method with *inexact* local solutions.

Theorem 4.2 (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*

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

for some $c > 0$. Let $\omega = 1$ and assume that in Algorithm 2, *instead of step* (2.12) *we accept "inexact" solutions y_i^k to the minimization problem whenever*

$$f(x^{i-1,k}) - f(x^{i,k}) \geqslant \alpha \cdot \|R_i^k \nabla f(x^{i-1,k})^{\mathrm{T}}\|^2, \quad i = 1, \ldots, m, \tag{4.3}$$

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

Proof. We show that assertions (i)–(iv) in the proof of Theorem 4.1 still hold. Note that (i) is again trivial by (4.3) and that, consequently, we again have

$$\lim_{k\to\infty} (f(x^{i-1,k}) - f(x^{i,k})) = 0, \quad i = 1, \ldots, m.$$

From (4.3) we thus get $\lim_{k\to\infty} R_i^k \nabla f(x^{i-1,k})^{\mathrm{T}} = 0$ for $i = 1, \ldots, m$. By strong convexity we have

$$f(x^{i,k}) - f(x^{i-1,k}) \geqslant \nabla f(x^{i-1,k}) P_i^k y_i^k + \frac{C}{2} \|P_i^k y_i^k\|^2. \tag{4.4}$$

Now we use $P_i^k = P_i^k R_i^k P_i^k$, the Cauchy–Schwarz inequality, and the fact that $\|P_i^k\| = 1$ to obtain

$$\begin{aligned} |\nabla f(x^{i-1,k}) P_i^k y_i^k| &= |\nabla f(x^{i-1,k}) P_i^k R_i^k P_i^k y_i^k| \\ &\leqslant \|\nabla f(x^{i-1,k}) P_i^k R_i^k\| \cdot \|P_i^k y_i^k\| \\ &= \|P_i^k R_i^k \nabla f(x^{i-1,k})^{\mathrm{T}}\| \cdot \|P_i^k y_i^k\| \\ &\leqslant \|R_i^k \nabla f(x^{i-1,k})^{\mathrm{T}}\| \cdot \|P_i^k y_i^k\|. \end{aligned}$$

Therefore, (4.4) yields

$$f(x^{i,k}) - f(x^{i-1,k}) \geqslant \left(-\|R_i^k \nabla f(x^{i-1,k})^{\mathrm{T}}\| + \frac{C}{2}\|P_i^k y_i^k\| \right) \cdot \|P_i^k y_i^k\|.$$

Since $\lim_{k\to\infty}(f(x^{i-1,k}) - f(x^{i,k})) = 0$ and $\lim_{k\to\infty} \|R_i^k \nabla f(x^{i-1,k})^{\mathrm{T}}\| = 0$ this implies $\lim_{k\to\infty} \|P_i^k y_i^k\| = 0$, i.e. assertion (ii) of the proof of Theorem 4.1. Assertion (iii) follows as in that proof with the exception that instead of $R_i^k \nabla f(x^{i,k})^{\mathrm{T}} = 0$ for all k we now use $\lim_{k\to\infty} R_i^k \nabla f(x^{i,k})^{\mathrm{T}} = 0$. Finally, (iv) follows in exactly the same way as before. □

If we interpret relaxation of exact solutions as inexact solutions, the above theorem yields the following corollary. We omit its proof since it is completely analogous to that for Corollary 3.5.

Corollary 4.3. *Assume that f and the space decomposition $V = \sum_{i=1}^m V_i^k$ are as in Theorem* 4.1. *Consider the multiplicative algorithm with relaxation factor ω. Then $\lim_{k\to\infty} x^k = x^*$ in the following two cases*:
(i) $\omega \in (0, 1]$.
(ii) *f is the quadratic functional* (1.2) *and* $\omega \in (0, 2)$.

Part (i) of this corollary extends the results from [14] to under relaxation. Part (ii) can be found in [7], e.g., stated in terms of the linear systems $Ax = b$.

5. Quadratic functionals

In this section we discuss our previous results for the special case that f is a (strongly convex) quadratic functional and relate them to those known from the literature. We consider two different iterations: Solutions of linear systems and the full rank least-squares problem.

5.1. Linear systems

Assume that we are given the linear system

$$Ax = b$$

with $A \in \mathbf{R}^{n\times n}$ SPD and unique solution $x^* = A^{-1}b$. Then the A-norm (energy norm) of an approximate solution is given by the quadratic functional defined in (1.2) for which $-\nabla f(x) = b - Ax$ is the residual of f. If W is a subspace of $V=\mathbf{R}^n$ with embedding operator P, the solution $y\in W$ to the minimization problem $\min_{z\in W} f(x+Pz)$ is the solution of the projected linear system (see Lemma A.3(i))

$$(P^{\mathrm{T}}AP)y = P^{\mathrm{T}}(b-Ax).$$

Consequently, the iterates of the additive algorithm, Algorithm 1, are given by

$$x^{k+1} = x^k + \sum_{i=1}^{m} \alpha_i^k P_i^k \left((P_i^k)^{\mathrm{T}} A P_i^k\right)^{-1} (P_i^k)^{\mathrm{T}}(b-Ax^k). \tag{5.1}$$

Let us assume that the space decomposition does not depend on k (so that we can write P_i instead of P_i^k etc.) If, in addition, we take all α_i^k equal to some $\omega > 0$, (5.1) describes the standard damped additive Schwarz method (see [7], for example). The convergence result from Corollary 3.5 (ii) then reduces to [7, Lemma 11.2.9.b].

Similarly, Corollary 3.9 for the case of a coloring can be found as the "strengthened estimate", Lemma 11.2.4 in [7]. Note that by Theorem 3.1 we also get convergence for additive Schwarz in the case we compute the α_i^k via the optimal strategy (2.7). In this case, the vector $\alpha^k=(\alpha_1^k,\ldots,\alpha_m^k)^{\mathrm{T}}$ solves

$$B^k\alpha^k = c^k,$$

where

$$B_{ij}^k = (P_i y_i^k)^{\mathrm{T}} A (P_j y_j^k), \quad c_i^k = b^{\mathrm{T}} P_i y_i^k, \quad i,j=1,\ldots,m,$$

as developed for PVD0 in [10].

To specialize further, assume that we fix a basis for $V=\mathbf{R}^n$ and that, furthermore, the subspaces V_i are all spanned by some subset of this basis. In such a situation we call the spaces V_i *coordinate subspaces*. Then $P_i^{\mathrm{T}}AP_i$ represents a block of the matrix A belonging to this basis, so that (5.1) becomes a damped block Jacobi iteration with overlapping blocks. Here, overlap occurs for all those components where the corresponding basis vector appears in more than one of the subspaces V_i. So, again, Corollaries 3.5 and 3.9 show the convergence of the damped overlapping block Jacobi iteration with damping factor in $(0,2/m)$ or $(0,2/m_{\mathrm{c}})$, respectively. Let us also note that overlapping block Jacobi methods can alternatively also be described through *multisplittings* [8,5]. For a detailed discussion see [5,10]. Corollaries 3.5 and 3.9 are therefore to be interpreted as completing known convergence results for multisplittings of SPD matrices, see [6,8,16].

It is now evident that the multiplicative algorithm, Algorithm 2, can be specialized in a similar manner, to the block SOR method with, possibly, overlapping blocks. Corollary 4.3(ii) proves convergence if ω, the relaxation factor, satisfies $\omega\in(0,2)$. It is worthwhile to notice that the crucial condition (3.2) is of course satisfied, if the set $\{V_i^k,\ldots,V_m^k\}$ of subspaces does not depend on k, whereas the numbering by the subscript i does. For example, if we have $V_i^k = V_{m+1-i}^{k+1}$, $i=1,\ldots,m$ for every other k, we find the resulting multiplicative method becomes block SSOR (symmetric SOR) with overlapping blocks, and we have again convergence for $\omega\in(0,2)$. We also have convergence in the more general case where we chose some numbering anew for each iterative step k.

5.2. Linear least squares

Assume that $L \in \mathbf{R}^{p\times n}$ has full rank $rk(L) = n \leqslant p$ and consider the linear least-squares problem defined by (1.3),

$$f(x) = \|Lx - b\|_2^2 = x^{\mathrm{T}}L^{\mathrm{T}}Lx - 2(b^{\mathrm{T}}L)x.$$

As before, let W be a subspace with embedding P. The local minimization problem $\min_{z\in W} f(x+Pz)$ now reads

$$\min_{z\in W} \|(LP)z - (b - Lx)\|_2^2. \tag{5.2}$$

which represents a full rank least-squares problem on the subspace W with matrix LP, $rk(LP) = \dim(W)$. The local least-squares problems (5.2) will usually be solved using a standard approach such as QR-decomposition on LP.

For this class of problems the additive algorithm was considered in [10] for the special case of coordinate subspaces which do not depend on k. In [10] both the optimal strategy and a damped convex combination strategy were considered. Our Corollaries 3.5 and 3.9 improve upon the convergence analysis for the convex combination strategy presented in [10], which assumed $\sum_{i=1}^{m} \gamma_i \leqslant 1$ (terminology of Corollary 3.5).

The very recent paper [2] also deals with the additive algorithm for linear least-squares problems. There, the space decompositions are obtained as in the PVD method. Theoretical and practical results are presented motivating the choice $x^k - x^{k-1}$ for the "forget-me-not" term s^k in (2.11). This reference also considers a line search strategy for the synchronization step (2.6).

Acknowledgements

We are thankful to Daniel B. Szyld for his helpful comments on an earlier version of this paper. The research of the second author was supported under grant NSF DMS9402943.

Appendix A. Auxiliary results and definitions

We collect the auxiliary results and necessary definitions in this appendix. As always $V = \mathbf{R}^n$.

Definition A.1 (Strong convexity). A differentiable function $f : V \to \mathbf{R}$ is strongly convex on V with constant $C > 0$ if

$$f(y) - f(x) - \nabla f(x)(y - x) \geqslant \frac{C}{2}\|y - x\|^2, \quad \forall x, y \in V \tag{A.1}$$

or equivalently

$$(\nabla f(y) - \nabla f(x))(y - x) \geqslant C\|y - x\|^2, \quad \forall x, y \in V.$$

Lemma A.2 (Rayleigh–Ritz inequalities). *Let $\lambda_{\min}, \lambda_{\max} > 0$ denote the smallest and largest eigenvalues of the symmetric matrix A, respectively. Then,*

$$\lambda_{\min} = \min_{x \neq 0} \frac{x^{\mathrm{T}}Ax}{x^{\mathrm{T}}x} \leqslant \max_{x \neq 0} \frac{x^{\mathrm{T}}Ax}{x^{\mathrm{T}}x} = \lambda_{\max}.$$

Specifically, for all $x \in V$, we have

$$\lambda_{\min}\|x\|^2 \leqslant x^{\mathrm{T}}Ax = \|x\|_A^2 \leqslant \lambda_{\max}\|x\|^2.$$

Lemma A.3. *Let f be the quadratic functional defined by* (1.2) *with A symmetric positive definite. Let W be a subspace of V and $P : W \to V$ the corresponding embedding operator. For x fixed let $y \in W$ be the solution of the local minimization problem*

$$f(x + Py) = \min_{z \in W} f(x + Pz).$$

Then
(i) $y = (P^{\mathrm{T}}AP)^{-1}P^{\mathrm{T}}(b - Ax)$,
(ii) *for $t \in [\gamma, 2 - \gamma]$ with $0 < \gamma < 1$ we have*

$$f(x) - f(x + tPy) \geqslant \frac{\gamma(2-\gamma)}{2\lambda_{\max}} \|\nabla f(x)P\|^2,$$

where $\lambda_{\max} > 0$ is the largest eigenvalue of A.

Proof. The minimizer y of the quadratic functional $g(z) = f(x + Pz)$ on W satisfies $0 = \nabla g(y) = \nabla f(x + Py)P = ((x + Py)^{\mathrm{T}}A - b^{\mathrm{T}})P$. This shows that y is given by

$$y = (P^{\mathrm{T}}AP)^{-1}P^{\mathrm{T}}(b - Ax),$$

i.e. (i). Consequently,

$$\begin{aligned} f(x + Py) &= f(x) + x^{\mathrm{T}}A(Py) + \tfrac{1}{2}(Py)^{\mathrm{T}}A(Py) - b^{\mathrm{T}}Py \\ &= f(x) - \tfrac{1}{2}y^{\mathrm{T}}P^{\mathrm{T}}APy \\ &= f(x) - \tfrac{1}{2}(b - Ax)^{\mathrm{T}}P(P^{\mathrm{T}}AP)^{-1}(P^{\mathrm{T}}AP)(P^{\mathrm{T}}AP)^{-1}P^{\mathrm{T}}(b - Ax) \\ &= f(x) - \tfrac{1}{2}r^{\mathrm{T}}(P^{\mathrm{T}}AP)^{-1}r, \end{aligned} \tag{A.2}$$

with $r = P^{\mathrm{T}}(b - Ax) = -(\nabla f(x)P)^{\mathrm{T}}$. By Lemma A.2, for $w \neq 0$, the Rayleigh quotient $w^{\mathrm{T}}(P^{\mathrm{T}}AP)^{-1}w / w^{\mathrm{T}}w$ is not smaller than the smallest eigenvalue of $(P^{\mathrm{T}}AP)^{-1}$, which in turn is the inverse of the largest eigenvalue $\lambda_{\max}(P^{\mathrm{T}}AP)$ of $P^{\mathrm{T}}AP$. Moreover, again by Lemma A.2

$$\lambda_{\max}(P^{\mathrm{T}}AP) = \max_{w \in W} \frac{w^{\mathrm{T}}(P^{\mathrm{T}}AP)w}{w^{\mathrm{T}}w} = \max_{w \in W} \frac{(Pw)^{\mathrm{T}}A(Pw)}{(Pw)^{\mathrm{T}}(Pw)} \leqslant \max_{v \in V} \frac{v^{\mathrm{T}}Av}{v^{\mathrm{T}}v} = \lambda_{\max}.$$

This shows

$$r^{\mathrm{T}}(P^{\mathrm{T}}AP)^{-1}r \geqslant \frac{1}{\lambda_{\max}}\|r\|^2,$$

so that from (A.2) we get

$$\sigma := f(x^k) - f(x+Py) \geqslant \frac{1}{2\lambda_{\max}} \|r\|^2. \tag{A.3}$$

Now, for $t \in \mathbf{R}$ the function $q(t) = f(x+tPy)$ is quadratic, having its minimum at $t=1$, so that it can be expressed as $q(t) = \sigma \cdot (t-1)^2 + f(x+Py)$. This shows $q(0) - q(t) = \sigma t(2-t)$, so that we have $f(x) - f(x+tPy) = q(0) - q(t) \geqslant \sigma\gamma(2-\gamma)$ for $t \in [\gamma, 2-\gamma]$ which, by (A.3) finally yields

$$f(x) - f(x+tPy) \geqslant \frac{\gamma(2-\gamma)}{2\lambda_{\max}} \|r\|^2. \qquad \square$$

Lemma A.4 (Uniqueness of minimum for strongly convex $f(x)$). *Suppose that $f(x) \in C^1(V)$ is strongly convex; then there exists a unique vector x^* which minimizes $f(x)$ over V.*

Proof. For arbitrary $y \in V$ the set $\{z \in V \mid f(z) \leqslant f(y)\}$ is compact. This shows that f has at least one minimizer. Assume that x^* and $\bar{x}$ are both minimizers. Then $\nabla f(x^*) = \nabla f(\bar{x}) = 0$, so that $(\nabla f(x^*) - \nabla f(\bar{x}))(\bar{x} - x^*) = 0$. By strong convexity this is only possible if $x^* = \bar{x}$. $\square$

Lemma A.5 (Ferris and Mangasarian [4, Linear convergence]). *Let $f: V \to \mathbf{R}$ and let $\{x^k\} \subseteq V$. If $f \in C^1(V)$ is bounded from below and*

$$f(x^k) - f(x^{k+1}) \geqslant \alpha \|\nabla f(x^k)\|^2 \tag{A.4}$$

for some $\alpha > 0$, $k = 0, 1, 2, \ldots$, then every accumulation point $\bar{x}$ of $\{x^k\}$ is stationary, that is $\nabla f(\bar{x}) = 0$.

If in addition, $f \in LC^1_K(V)$ is strongly convex with parameter $C > 0$, then $\{x^k\}$ converges to the unique solution x^* of $\min_{x \in V} f(x)$ at the linear root rate:

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

Lemma A.6 (Ferris and Mangasarian [4, Quadratic bound lemma], [9]). *Let $f \in LC^1_K(V)$. Then for all $x, y \in V$*

$$f(y) - f(x) - \nabla f(x)(y-x) \leqslant |f(y) - f(x) - \nabla f(x)(y-x)| \leqslant \frac{K}{2} \|y-x\|^2.$$

In particular, taking $y = x - (1/K)\nabla f(x)^{\mathrm{T}}$ we have

$$f(x) - f(y) \geqslant -\frac{1}{2K} \|\nabla f(x)^{\mathrm{T}}\|^2 + \frac{1}{K} \|\nabla f(x)^{\mathrm{T}}\|^2 = \frac{1}{2K} \|\nabla f(x)^{\mathrm{T}}\|^2.$$

References

[1] D.P. Bertsekas, J.N. Tsitsiklis, Parallel and Distributed Computation, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[2] J.E. Dennis, 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, O.L. Mangasarian, Parallel constraint distribution, SIAM J. Optim. 1 (1991) 487–500.

[4] M.C. Ferris, O.L. Mangasarian, Parallel variable distribution, SIAM J. Optim. 4 (1994) 815–832.
[5] A. Frommer, G. Mayer, Convergence of relaxed parallel multisplitting methods, Linear Algebra Appl. 119 (1989) 141–152.
[6] A. Frommer, G. Mayer, On the theory and practice of multisplitting methods in parallel computation, Computing 49 (1992) 63–74.
[7] W. Hackbusch, Iterative Methods for Large Sparse Linear Systems, Springer, Heidelberg, 1993.
[8] D.P. O'Leary, R.E. White, Multi-splittings of matrices and parallel solution of linear systems, SIAM J. Algeb. Discr. Meth. 6 (1985) 630–640.
[9] J.M. Ortega, W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.
[10] R.A. Renaut, A Parallel Multisplitting Solution of the Least Squares Problem, Numer. Linear Algebra with Appl. 5 (1998) 11–31.
[11] B.F. Smith, P.E. Bjørstad, W.D. Gropp, Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations, Cambridge University Press, Cambridge, 1996.
[12] M.V. Solodov, New inexact parallel variable distribution algorithms, Comput. Optim. Appl. 7 (1997) 165–182.
[13] M.V. Solodov, On the convergence of constrained parallel variable distribution algorithms, SIAM J. Optim. 8 (1) (1998) 187–196.
[14] X.C. Tai, Parallel function and space decomposition methods–Part II, Space decomposition, Beijing Math. 1 (Part 2) (1995) 135–152.
[15] X.C. Tai, M. Espedal, Rate of convergence of some space decomposition methods for linear and nonlinear problems, SIAM J. Numer. Anal. 35 (1998) 1558–1570.
[16] R.E. White, Multisplittings of a symmetric positive definite matrix, SIAM J. Matrix Anal. Appl. 11 (1990) 69–82.
[17] J. Xu, Iterative methods by space decomposition and subspace correction, SIAM Rev. 34 (1992) 581–613.