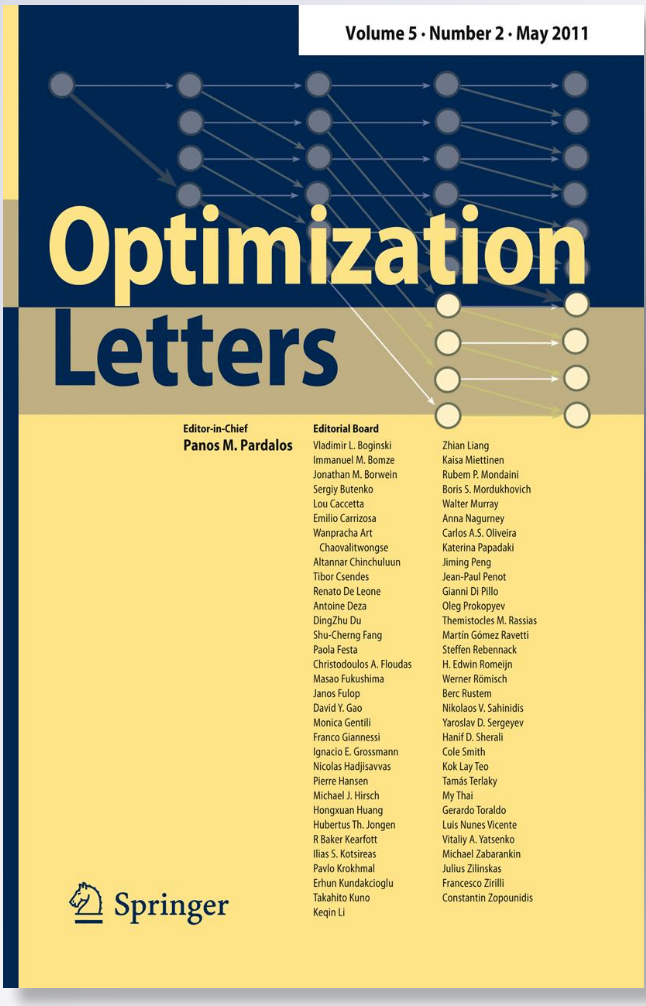
On the convergence of a Jacobi-type algorithm for singly linearly-constrained problems subject to simple bounds

Optimization Letters

ISSN 1862-4472
Volume 5
Number 2

Optim Lett (2011) 5:347-362
DOI 10.1007/s11590-010-0214-x



Springer

Optim Lett (2011) 5:347–362
DOI 10.1007/s11590-010-0214-x

ORIGINAL PAPER

On the convergence of a Jacobi-type algorithm for singly linearly-constrained problems subject to simple bounds

Giampaolo Liuzzi · Laura Palagi · Mauro Piacentini

Received: 15 January 2010 / Accepted: 25 June 2010 / Published online: 6 July 2010

Abstract In this work we define a block decomposition Jacobi-type method for nonlinear optimization problems with one linear constraint and bound constraints on the variables. We prove convergence of the method to stationary points of the problem under quite general assumptions.

Keywords Decomposition method · Jacobi-type iteration

1 Introduction

Let us consider the problem

$$\begin{aligned} \min \ & f(x) \\ & x \in \mathcal{F} \end{aligned} \tag{1}$$

where

$$\mathcal{F} = \{x \in \mathbb{R}^n : \ a^T x = b, \ l \le x \le u\}$$

G. Liuzzi
Consiglio Nazionale delle Ricerche, Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti" (CNR, IASI), Viale Manzoni 30, Rome, Italy
e-mail: liuzzi@iasi.cnr.it

L. Palagi (✉) · M. Piacentini
Dipartimento di Informatica e Sistemistica A. Ruberti (DIS), Sapienza Università di Roma, Via Ariosto 25, Rome, Italy
e-mail: palagi@dis.uniroma1.it

M. Piacentini
e-mail: piacentini@dis.uniroma1.it

and $a, l, u \in \mathbb{R}^n$, with $-\infty \leq l < u \leq +\infty$, $b \in \mathbb{R}$. We allow the possibility that some of the variables are unbounded by permitting both $l_i = -\infty$ and $u_i = \infty$ for some $i \in \{1, \dots, n\}$.

There are many problems that can be formulated as special cases of problem (1). In particular, training a support vector machine (SVM) (see e.g. [20]) leads to a problem of type (1) where $f(x)$ is a convex quadratic function, $b = 0$ and $a_i \in \{-1, 1\}$, $0 = l_i < u_i = C$ with $C > 0$ for $i = 1, \dots, n$. Another problem of type (1) is the so called standard quadratic programming problem (StQP) where $f(x)$ is a quadratic form, $b = 1$ and $a_i = 1, l_i = 0, u_i = \infty$ for all $i = 1, \dots, n$. StQP problems arises, for example, in portfolio optimization [16] and as continuous formulation of maximum clique problems (see e.g. [2,17]).

In this paper, we are interested in decomposition methods, which involve the solution of subproblems of smaller dimensions in place of the original one. In literature, decomposition methods for unconstrained problems can be roughly classified into two main classes: Gauss–Seidel methods (see e.g. [1,3,7,8]) and Jacobi methods (see e.g. [1,5,7]). The minimization version of the block nonlinear Gauss–Seidel method [1] defines the new iterate x^{k+1} by successive minimization with respect to each component vector of variables. On the other hand, in the nonlinear Jacobi method the minimizations with respect to each component vector of variables are carried out simultaneously in order to define the new iterate x^{k+1}. Convergence results for these two classes of methods have been proved for unconstrained problems in e.g. [1,5,7]. For constrained problems, convergence of the nonlinear block Gauss–Seidel method has been proved when each block component of variables belongs to a closed convex subset $X_i \subseteq \mathbb{R}^{n_i}$ and the feasible set is the Cartesian product of $X_i, = 1, \dots, m$ [8]. The presence of the linear constraint in problem (1) does not allow to use such results for defining convergent decomposition methods. However, in the context of SVM training, much effort has been devoted to the definition of convergent decomposition schemes that fit in the class of block Gauss–Seidel-type methods. In this decomposition framework, starting from a feasible point, at each iteration k a subset of indices $W^k \subset \{1, \dots, n\}$ is chosen and the new iterate x^{k+1} is defined by updating only the variables with indices belonging to W^k. The choice of set W^k at each iteration plays a crucial role in proving convergence of the sequence $\{x^k\}$. It is worth noting that different selection rules lead to different algorithms. In particular, in most decomposition methods for SVM problem, the indices in W^k are selected on the basis of the violation of the optimality conditions at x^k (see e.g. [6,9–13,15,18]). In [14] a convergent decomposition algorithm for Problem (1) has been defined that differs from the other ones in that the working set selection (WSS) rule does not require to apply any specific ordering procedure.

In this paper, taking inspiration from [14], we define a block Jacobi-type convergent decomposition method for problem (1). Up to our knowledge there is no proof of convergence for Jacobi-type decomposition methods for singly linearly constrained problems with bound constraints. We prove convergence of our algorithm towards stationary points under standard assumptions.

The paper is organized as follows. In Sect. 2, we introduce some notations and definitions. In Sect. 3 we define the block Jacobi-type decomposition method for Problem (1) which we call `CoJac`. The convergence properties of Algorithm `CoJac` are studied in Sect. 4 under some general conditions on the search directions used. Section 5 is devoted to the choice of directions that fulfill the conditions used in the convergence analysis.

2 Notation and definitions

In this section we introduce some useful definitions and the basic notation that will be used throughout the paper.

Given an index set $W \subset \{1, \dots, n\}$, we denote by $\overline{W}$ its complement with respect to $\{1, \dots, n\}$, that is, $\overline{W} = \{1, \dots, n\} \backslash W$. Given a vector $v \in \mathbb{R}^n$ and an index set W, we denote by $v_W \in \mathbb{R}^{|W|}$ the subvector of v made up of the component v_i with $i \in W$. Further, for the sake of simplicity we use the notation $\nabla_W f$ for $(\nabla f)_W$.

We denote by $\mathcal{W} = \{W^1, \dots, W^M\} \subseteq 2^{\{1,\dots,n\}}$ a family of index sets of cardinality $q^1, \dots, q^M$, with $M > 0$.

The set of feasible directions at a point $x \in \mathcal{F}$ is the cone

$$D(x) = \left\{ d \in \mathbb{R}^n \mid a^T d = 0,\ d_i \geq 0,\ \forall\, i : \ x_i = l_i, \text{ and } d_i \leq 0,\ \forall\, i : \ x_i = u_i \right\}.$$

Next we define a stationary point for Problem (1).

Definition 2.1 (*Stationary point*) A point $x^* \in \mathcal{F}$, is stationary for Problem (1) if

$$\nabla f(x^*)^T d \geq 0 \qquad \text{for all } d \in D(x^*).$$

Given a feasible point $\tilde{x}$, and a subset $W \subset \{1, \dots, n\}$, let us define the subproblem $\mathcal{P}_W(\tilde{x})$ as

$$\begin{aligned} \min\ & f(x_W, \tilde{x}_{\overline{W}}) \\ & x_W \in \mathcal{F}_W(\tilde{x}) \end{aligned} \tag{2}$$

where

$$\mathcal{F}_W(\tilde{x}) = \{x_W : (x_W, \tilde{x}_{\overline{W}}) \in \mathcal{F}\} = \{x_W \in \mathbb{R}^{|W|} : \ a_W^T x_W = b - a_{\overline{W}}^T \tilde{x}_{\overline{W}},\ l_W \leq x_W \leq u_W\}$$

At any feasible point $x \in \mathcal{F}$, let us denote with $D_W(x_W)$ the set of feasible directions at x_W with respect to $\mathcal{F}_W(x)$, that is

$$D_W(x_W) = \left\{ d \in \mathbb{R}^{|W|} : \ a_W^T d = 0,\ d_i \geq 0,\ \forall\, i \in W : \ x_i = l_i, \text{ and } d_i \leq 0,\ \forall\, i \in W : \ x_i = u_i \right\}.$$

We introduce the definition of stationary points for Problem (2).

Definition 2.2 (*Stationary point of* $\mathcal{P}_W(x)$) Given $x \in \mathcal{F}$ and the corresponding problem $\mathcal{P}_W(x)$, a point $x_W^* \in \mathcal{F}_W(x)$ is stationary for $\mathcal{P}_W(x)$ if

$$\nabla_W f \left(x_W^*, x_{\overline{W}}\right)^T d \geq 0 \qquad \text{for all } d \in D_W\left(x_W^*\right).$$

3 A constrained Jacobi-type algorithm

In this section we introduce a block Jacobi-type decomposition algorithm for Problem (1). The iterate x^{k+1} is generated by using information on the simultaneous (approximated) minimizations with respect to the components of the vector. To be more precise, let $\mathcal{W} = \{W^1, \ldots, W^M\}$ be a family of index sets that need not necessarily define a partition of $\{1, \ldots, n\}$, so that, differently from the block Jacobi-type method defined in [7], $W^i \cap W^j$, with $i \neq j$, may be nonempty.

We require that the sequence be generated in such a way to satisfy

$$f(x^{k+1}) \le \min_{W^i} f^k_{ref_{W^i}}$$

where $f^k_{ref_{W^i}}$ are reference values for subproblems $\mathcal{P}_{W^i}(x^k)$, $i = 1, \ldots, M$. In a standard Jacobi method, reference values are obtained by performing exact minimizations with respect to each x_{W^i}, namely setting

$$f^k_{ref_{W^i}} = \min_{x_{W^i} \in \mathcal{F}_{W_i}(x^k)} f\left(x_{W^i}.x^k_{\overline{W^i}}\right),$$

However, this could be a too strict requirement, particularly when $|W^i| > 2$ or the objective function is not convex. Hence, following the idea in [13,14] we relax this requirement by asking only for a "sufficient reduction" of the objective function, and the values $f^k_{ref_{W^i}}$ are obtained by means of an Armijo-type line search procedure along a feasible direction d_{W^i} at $x^k_{W^i}$ for $i = 1, \ldots, M$.

In particular, given a feasible point $x \in \mathcal{F}$, a set $W \in \mathcal{W}$, and a direction $d_W \in D_W(x_W)$, let $\beta_{\mathcal{F}}$ be the maximum feasible step length along direction d_W with respect to the bound constraints. Namely, $\beta_{\mathcal{F}} = \beta_{\mathcal{F}}(x, W, d_W)$ is such that

$$l_W \le x_W + \beta d_W \le u_W \qquad \text{for all } \beta \in [0, \beta_{\mathcal{F}}],$$

and (since $-\infty \le l < u \le \infty$) we have that either $\beta_{\mathcal{F}} = +\infty$ or at least an index $i \in W$ exists such that

$$x_i + \beta_{\mathcal{F}} d_i = l_i \quad \text{or} \quad x_i + \beta_{\mathcal{F}} d_i = u_i.$$

Further, let $\beta_u > 0$ be a preselected positive scalar.

We report below the `Step_length` procedure that returns the stepsize α_W along the direction d_W.

Procedure `Step_length`(x, W, d_W)

Parameter. $\gamma \in (0, 1)$, $\delta \in (0, 1)$, $\beta_u > 0$.
Data. $x \in \mathcal{F}$, $W \in \mathcal{W}$, $d_W \in D_W(x_W)$ and $\beta = \min\{\beta_{\mathcal{F}}(x, W, d_W), \beta_u\}$.
Inizialization. If d_W satisfies $\nabla_W f(x)^T d_W \geq 0$, **Return** $\alpha_W = 0$;
otherwise set $\alpha = \beta$.
While $\left(f(x_W + \alpha d_W, x_{\overline{W}}) > f(x) + \gamma\alpha\nabla_W f(x)^T d_W\right)$
Set $\alpha = \delta\alpha$
End While
Return $\alpha_W = \alpha$.

The stepsize α_W is zero if and only if d_W does not satisfy the descent condition $\nabla_W f(x)^T d_W < 0$ at x.

The following proposition shows that Procedure `Step_length` is well-defined.

Proposition 3.1 *Let $W \in \mathcal{W}$. Assume $x \in \mathcal{F}$ and $d_W \in D_W(x_W)$. Then Procedure* `Step_length` *determines, in a finite number of iterations, a scalar α_W such that*

$$f(x + \alpha_W d) \leq f(x) + \gamma\alpha_W \nabla_W f(x)^T d_W. \tag{3}$$

Proof The proof is quite standard and, for the sake of completeness, we report it in the appendix A. □

Now we are ready to define the decomposition algorithm `CoJac`.

Constrained Jacobi-type (`CoJac`) **Algorithm**

Data. A point $x^0 \in \mathcal{F}$, and $\mathcal{W} = \{W^1, \ldots, W^M\}$.
For $k = 0, 1, 2, \ldots$
1. **For each** $i \in \{1, \ldots, M\}$
   choose $d^k_{W^i} \in D_{W^i}(x^k)$;
   compute $\alpha^k_{W^i}$ by means of procedure `Step_length`$(x^k, W^i, d^k_{W^i})$;
   set $y^k_{W^i} = x^k_{W^i} + \alpha^k_{W^i} d^k_{W^i}$.
   End For each
2. **Choose** x^{k+1} such that the following improvement condition holds:

$$f(x^{k+1}) \leq \min_{1 \leq i \leq M} f\left(y^k_{W^i}, x^k_{\overline{W^i}}\right). \tag{4}$$

End For

We point out the degree of freedom on the choice for the next iterate x^{k+1} in Algorithm `CoJac`: only the improvement condition (4) is required.

We further observe that, for each $i \in \{1, \ldots, M\}$, the instructions of the **For each** statement at Step 1 do not depend on $j \in \{1, \ldots, M\} \setminus \{i\}$ and in this sense they are suitable to be performed in a distributed way.

Convergence properties of the algorithm `CoJac` will be analyzed in the next section, under quite standard assumptions on the feasible directions used in the procedure `Step_length` and on the rule for the selection of the family of index sets $\mathcal{W}$.

4 Convergence results

In order to simplify the exposition, we first introduce an assumption on the behavior in the limit of the directions used by the algorithm. This is a basic assumption that can be satisfied by employing suitable methods for the computation of the search directions d_W^k. Possible choices of directions satisfying Assumption 1 are reported and discussed in Sect. 5.

Assumption 1 Let $\{x^k\}$ be a sequence of feasible points such that, for $K \subseteq \{0, 1, 2, \ldots\}$,

$$\lim_{k\to\infty, k\in K} x^k = \hat{x}.$$

For a given working set $W \in \mathcal{W}$, let $\{d_W^k\}$ be a sequence such that $d_W^k \in D_W(x_W^k)$.

The sequence $\{d_W^k\}$ is said to satisfy the assumption if:

(i) $x_W^k + d_W^k \in \mathcal{F}_W(x_W^k)$;
(ii) a constant $U > 0$ exists such that $\|d_W^k\| \leq U$ for all $k \in K$;
(iii) if $\hat{x}_W$ is not a stationary point for $\mathcal{P}_W(\hat{x})$, then

$$\limsup_{k\to\infty, k\in K} \nabla_W f(x^k)^T d_W^k < 0.$$

From the definition of procedure `Step_length`, when the set $W \in \mathcal{W}$ is held fixed, we can state the following preliminary convergence result.

Proposition 4.1 *Let $\{x^k\} \subset \mathcal{F}$ be a sequence of feasible points converging to a given point $\hat{x}$. For a fixed $W \in \mathcal{W}$, and every k, let $d_W^k \in D_W(x_W^k)$ satisfy Assumption 1 and*

$$y_W^k = x_W^k + \alpha_W^k d_W^k, \tag{5}$$

where α_W^k is computed by means of the `Step_length` *Procedure. Then the sequence $\{y_W^k\}$, is such that:*

(i) $f(y_W^k, x_{\bar{W}}^k) \leq f(x^k)$*, for all* $k = 0, 1, 2, \ldots$

(ii) *if*

$$\lim_{k\to\infty}\left(f(x^k)-f\left(y^k_W,x^k_{\overline{W}}\right)\right)=0 \tag{6}$$

then $\hat{x}_W$ is a stationary point for $\mathcal{P}_W(\hat{x})$.

Proof The proof is quite technical and is therefore reported in the appendix A. □

We stress the fact that $\{x^k\}$ is a given sequence that may not depend on Procedure `Step_length` in the sense that x^{k+1} is not necessarily defined on the basis of y^k_W.

Now we report the main result regarding convergence of the CoJac Algorithm. As standard in decomposition algorithms, we must require that the family of index sets $\mathcal{W}=\{W^1,\dots,W^M\}$ satisfies a suitable rule. In particular we require the following condition.

Pairwise Inclusion Property (PIP)[1]

For each pair $\{i,j\}\subset\{1,\dots,n\}$, there exists at least an $\ell\in\{1,\dots,M\}$ such that, $\{i,j\}\subseteq W^\ell$.

We observe that the definition of a family of index sets $\mathcal{W}$ satisfying Condition PIP does not require any information about the current iterate and can thus be defined a priori.

In [14] a working set selection (WSS) condition has been introduced which essentially requires each pair of indices $\{i,j\}$, which identifies a feasible descent direction at iterate x^k, to be inserted in the working set within a maximum number M of successive iterations. We note that PIP is new to this paper and in particular it satisfies the WSS condition. Indeed, given a family $\mathcal{W}=\{W^1,\dots,W^M\}$ satisfying PIP, a possibile way to implement WSS condition consists in using the sets $W^1,\dots,W^M$ in a cyclic way (see Sect. 6 for a possible implementative choice). On the other hand, given a sequence $\{W^k\}$ of working sets satisfying WSS, it might be not possible to extract a family $\mathcal{W}$ of M sets satisfying PIP. Hence WSS condition, although more general than PIP, cannot be used to prove convergence of Algorithm `CoJac`, in the sense that the current proof of the proposition below is valid only under the stronger PIP condition.

Proposition 4.2 *Let the family of index sets $\mathcal{W}=\{W^1,\dots,W^M\}$ satisfy condition PIP. Let $\{x^k\}$ and $\{d^k_{W^i}\}$ for each $i=1,\dots,M$ be the sequences defined by Algorithm* `CoJac`. *Assume that, for each $i=1,\dots,M$, sequences $\{d^k_{W^i}\}$ satisfy Assumption 1.*

Then, every limit point of $\{x^k\}$ is a stationary point for Problem (1).

Proof The proof of the proposition depends on a number of technical results and is therefore reported in the appendix A. □

[1] As an example, for $n=6$ a possible family of index sets $\mathcal{W}$ with $q^i=4$ for all i satisfying condition PIP is $\mathcal{W}=\{W^1,\dots,W^3\}$ with $W^1=\{1,2,3,4\}$, $W^2=\{1,2,5,6\}$, $W^3=\{3,4,5,6\}$.

5 Search directions

In this section, for any fixed W^h, $h = 1, \dots, M$, we consider two well-known methods, Frank–Wolfe and Projected Gradient methods, for calculating a feasible direction $d^k_{W^h}$ and we show that such directions satisfy Assumption 1.

First we report a useful theoretical result (whose proof can be found in [14]) that will be used in the section.

Proposition 5.1 *Let $\{x^k\} \subset \mathcal{F}$ be a sequence of feasible points converging to a point $\tilde{x} \in \mathcal{F}$. Then, for sufficiently large values of k,*

$$D(\tilde{x}) \subseteq D(x^k).$$

5.1 Frank–Wolfe direction

For a given $W \in \mathcal{W}$ and a feasible point x^k, the Frank–Wolfe (F–W) direction is

$$d^k_W = \bar{x}^k_W - x^k_W,$$

where $\bar{x}^k_W$ is an optimal solution of the following linear programming problem

$$\min_{x_W \in \mathcal{F}_W(x^k)} \nabla_W f(x^k)^T \left(x_W - x^k_W\right) \tag{7}$$

In the next proposition we show that the F–W direction is well defined and that the desired properties stated in Assumption 1 hold.

Proposition 5.2 *Assume that the feasible set $\mathcal{F}$ is compact. Let $W \in \mathcal{W}$, $\{x^k\}$ be a sequence of feasible points and $\{d^k_W\}$ be the associated sequence of F–W directions. Then, sequence $\{d^k_W\}$ is well defined and the following conditions hold:*

- (i) *for any k, $d^k_W \in D_W(x^k_W)$ and $x^k_W + d^k_W \in \mathcal{F}_W(x^k)$;*
- (ii) *for any k, $\nabla_W f(x^k)^T d^k_W < 0$ if and only if x^k_W is not a stationary point for $\mathcal{P}_W(x^k)$;*
- (iii) *for any k, $\|d^k_W\| \le U$ for a given constant $U > 0$;*
- (iv) *assume that*

$$\lim_{k\to\infty, k\in K} x^k = \tilde{x},$$

where $K \subseteq \{0, 1, 2, \dots\}$ and $\tilde{x}_W$ is not a stationary point of $\mathcal{P}_W(\tilde{x})$, then

$$\limsup_{k\to\infty, k\in K} \nabla_W f(x^k)^T d^k_W < 0.$$

Proof For any k, compactness of $\mathcal{F}$ implies that problem (7) admits a solution $\bar{x}^k_W$ so that d^k_W is well defined. Point (i) follows by definition of d^k_W.

By definition of d_W^k, for all $d \in D_W(x_W^k) \setminus \{0\}$ there is a scalar $t_d > 0$ such that

$$\nabla_W f\left(x^k\right)^T d_W^k = \nabla_W f(x^k)^T \left(\bar{x}_W^k - x_W^k\right) \leq t_d \nabla_W f(x^k)^T d. \tag{8}$$

Hence, if x^k is not a stationary point of $\mathcal{P}_W(x^k)$ we have that $\nabla_W f(x^k)^T d_W^k < 0$ and vice versa, so that point (ii) holds. Further, we observe that $\|d_W^k\| = \|\bar{x}_W^k - x_W^k\| \leq \|\bar{x}_W^k\| + \|x_W^k\|$ so that by compactness of $\mathcal{F}$ we have point (iii).

Now, let us relabel as $\{x^k\}$ the subsequence converging to $\tilde{x} \in \mathcal{F}$ where $\tilde{x}_W$ is not a stationary point of $\mathcal{P}_W(\tilde{x})$. Let $\tilde{d}_W \in D_W(\tilde{x}_W)$ be a feasible direction such that

$$\nabla_W f(\tilde{x})^T \tilde{d}_W < 0.$$

For sufficiently large values of k, by Proposition (5.1), we have

$$\tilde{d}_W \in D_W\left(x_W^k\right),$$

and by continuity of the gradient

$$\nabla_W f(x^k)^T \tilde{d}_W < 0.$$

So that, using (8), we also get that for sufficiently large values of k

$$\nabla_W f(x^k)^T d_W^k \leq t_{\tilde{d}_W} \nabla_W f(x^k)^T \tilde{d}_W < 0.$$

By taking the limit, we obtain

$$\limsup_{k\to\infty} \nabla_W f(x^k)^T d_W^k \leq t_{\tilde{d}_W} \lim_{k\to\infty} \nabla_W f(x^k)^T \tilde{d}_W \leq t_{\tilde{d}_W} \nabla_W f(\tilde{x}) \tilde{d}_W < 0, \tag{9}$$

which finally proves point (iv). □

5.2 Projected gradient direction

In [4] and [19] methods to obtain the projected gradient (PG) direction have been proposed. We recall that the projection $P_S(x)$ of a point x onto a non-empty closed convex set S is the solution of the following problem

$$\min_{y\in S} \|x - y\|.$$

The projection operator enjoys the following properties (see Proposition 3.2 in [1]).

Proposition 5.3 *The projection operator is continuous and not expansive. Further, $P_S(x)$ is the projection of x onto S if and only if*

$$(x - P_S(x))^T (y - P_S(x)) \leq 0, \quad \forall\, y \in S. \tag{10}$$

For a fixed $W \in \mathcal{W}$ and a sequence of feasible points $\{x^k\}$, let us define $P_W^k := P_S$ with $S = \mathcal{F}_W(x^k)$ as the projection operator onto $\mathcal{F}_W(x^k)$. We consider the following PG direction

$$d_W^k = \bar{x}_W^k - x_W^k$$

where

$$\bar{x}_W^k = P_W^k(x_W^k - s\nabla_W f(x^k)) \tag{11}$$

and s is a positive scalar.

In the next proposition we show that the PG direction is well defined and that Assumption 1 holds.

Proposition 5.4 *Assume that $\mathcal{F}$ is compact. Let $W \in \mathcal{W}$, $\{x^k\}$ be a sequence of feasible points and $\{d_W^k\}$ be the associated sequence of PG directions. Then, sequence $\{d_W^k\}$ is well defined and the following conditions hold.*

(i) *for any k, $d_W^k \in D_W(x_W^k)$ and $x_W^k + d_W^k \in \mathcal{F}_W(x^k)$;*
(ii) *for any k, $\nabla_W f(x^k)^T d_W^k < 0$ if and only if x_W^k is not a stationary point for $\mathcal{P}_W(x^k)$;*
(iii) *for any k, $\|d_W^k\| \le U$ for a given constant $U > 0$;*
(iv) *assume that*

$$\lim_{k\to\infty, k\in K} x^k = \tilde{x},$$

where $K \subseteq \{0, 1, 2, \ldots\}$ and $\tilde{x}_W$ is not a stationary point of $\mathcal{P}_W(\tilde{x})$, then

$$\limsup_{k\to\infty, k\in K} \nabla_W f(x^k)^T d_W^k < 0.$$

Proof For any k, the point $\bar{x}_W^k$ as in (11) is always defined and hence also d_W^k. Furthermore, by definition $x_W^k + d_W^k = \bar{x}_W^k \in \mathcal{F}_W(x^k)$ so that we get point (i).

By (10), $\bar{x}_W^k$ satisfies

$$(x_W^k - s\nabla_W f(x^k) - \bar{x}_W^k)^T \left(y - \bar{x}_W^k\right) \le 0, \quad \forall\, y \in \mathcal{F}_W(x^k), \tag{12}$$

so that, choosing $y = x_W^k$, we obtain by simple manipulations

$$\nabla_W f(x^k)^T d_W^k \le -\frac{1}{s}\|d_W^k\|^2 \le 0. \tag{13}$$

Recalling that x^k is such that x_W^k a stationary point of Problem $\mathcal{P}_W(x^k)$ if and only if (see Proposition 3.3 in [1])

$$x_W^k = P_W^k(x_W^k - s\nabla_W f(x_W^k)),$$

for any scalar $s > 0$, we have that $d_W^k = 0$ if and only if x_W^k is a stationary point of Problem $\mathcal{P}_W(x^k)$. Hence point (ii) follows from (13).

Reasoning as in the proof of Proposition 5.2, we have also point (iii).

Now, let us relabel as $\{x^k\}$ the subsequence converging to $\tilde{x} \in \mathcal{F}$ where $\tilde{x}_W$ is not a stationary point of $\mathcal{P}_W(\tilde{x})$. Let $\tilde{d}_W \in D_W(\tilde{x}_W)$ be a feasible direction such that

$$\nabla_W f(\tilde{x})^T \tilde{d}_W < 0.$$

By the continuity of the gradient, we have,

$$\lim_{k\to\infty} \nabla_W f(x^k)^T \tilde{d}_W = \nabla_W f(\tilde{x})^T \tilde{d}_W < 0$$

then for sufficiently large values of k we have

$$\nabla_W f(x^k)^T \tilde{d}_W < 0. \tag{14}$$

Further, by Proposition 5.1 it holds that

$$\tilde{d}_W \in D_W(x_W^k). \tag{15}$$

Reasoning by contradiction, we suppose that point (iv) does not hold, namely

$$\limsup_{k\to\infty} \nabla_W f(x^k)^T d_W^k \geq 0,$$

so that as result of (13)

$$\limsup_{k\to\infty} \|d_W^k\|^2 = 0. \tag{16}$$

Let $\bar{t} = \frac{1}{2}\beta_F(\tilde{x}, W, \tilde{d}_W) > 0$ so that $l_i < \tilde{x}_i + \bar{t}\tilde{d}_i < u_i$ for all $i \in W$ and $a_W^T \tilde{d}_W = 0$. Hence by (15), for k sufficiently large, the point $y = x_W^k + \bar{t}\tilde{d}_W \in \mathcal{F}_W(x^k)$. Substituting in (12), we have that

$$\begin{aligned} 0 &\geq \left(x_W^k - s\nabla_W f(x^k) - \bar{x}_W^k\right)^T \left(x_W^k + \bar{t}\tilde{d}_W - \bar{x}_W^k\right) \\ &= \left(-d_W^k - s\nabla_W f(x^k)\right)^T \left(-d_W^k + \bar{t}\tilde{d}_W\right). \end{aligned}$$

Rearranging the above inequality

$$\begin{aligned} \nabla_W f(x^k)^T d_W^k &\leq -\frac{1}{s}\|d_W^k\|^2 + \frac{\bar{t}}{s}\tilde{d}_W^T d_W^k + \bar{t}\nabla_W f(x^k)^T \tilde{d}_W \\ &\leq -\frac{1}{s}\|d_W^k\|^2 + \frac{\bar{t}}{s}\|\tilde{d}_W\|\|d_W^k\| + \bar{t}\nabla_W f(x^k)^T \tilde{d}_W \end{aligned}$$

Taking the limit of the above inequality, using (16) and (14), we get the contradiction

$$\limsup_{k\to\infty} \nabla_W f(x^k)^T d_W^k \leq \bar{t} \limsup_{k\to\infty} \nabla_W f(x^k)^T \tilde{d}_W < 0.$$

□

6 Conclusions

In the paper, we define a nonlinear block Jacobi type minimization method `CoJac` for the solution of singly linearly constrained problem with simple bounds on the variable. Convergence of the method is proved under standard assumptions and by imposing a simple condition PIP on the family of working sets used. Note that the definition of a family of index sets $\mathcal{W}$ satisfying Condition PIP does not require any information about the current iterate and can thus be defined a priori. In [14] numerical testing with a decomposition method using the WSS condition has been carried out. In particular the WSS rules used amounts to define the following family $\mathcal{W} = \{(1, 2), (1, 3), \ldots, (1, n), (2, 3), \ldots, (2, n), \ldots, (n-1, n)\}$ satisfying PIP, and using each $W^i \in \mathcal{W}$ sequentially in a cyclic way. The results reported in [14] are promising and we believe that this particular choice of $\mathcal{W}$ can be exploited also in our framework.

Moreover, it is worth noting that the method requires only a sufficient descent condition of the objective function which is enforced by a simple Armijo type line search. Algorithm `CoJac` is suitable to be implemented on a distributed machine, since the computations related to each subproblem could be performed concurrently. This is subject of further research.

Acknowledgments The authors thank two anonymous Referees and the Associate Editor for their helpful comments and suggestions that greatly helped us to improve the presentation of the paper.

Appendix A: Proofs of the convergence results

Proof of Proposition 3.1 If d_W does not satisfy $\nabla_W f(x)^T d_W < 0$, then $\alpha_W = 0$ and the condition is obviously satisfied.

Assume now that d_W satisfies $\nabla_W f(x)^T d_W < 0$ and by contradiction that the algorithm does not terminate. Hence, let $d = (d_W, d_{\overline{W}})$ with $d_{\overline{W}} = 0$, we can write

$$f(x + \beta\delta^j d) > f(x) + \gamma\beta\delta^j \nabla f(x)^T d \qquad \text{for all } j.$$

By applying the Mean Value theorem we have

$$\nabla f(x + \theta_j \beta\delta^j d)^T d > \gamma \nabla f(x)^T d \qquad \text{for all } j, \tag{17}$$

with $\theta_j \in (0, 1)$. Taking limits in (17) for $j \to \infty$ we obtain

$$(1 - \gamma)\nabla f(x)^T d \geq 0,$$

which implies, together with the fact that $\gamma \in (0, 1)$, that $\nabla f(x)^T d \geq 0$, and this contradicts the descent assumption on d. □

Proof of Proposition 4.1 Point (i) easily follows from the definition of the Armijo-type scheme and of y_W^k in (5).

Now, let us consider point (ii). As a result of closedness of the feasible set $\mathcal{F}$, the limit point $\hat{x}$ is feasible. By definition of Armijo-type rule (3), we have

$$f(x^k) - f\left(y_W^k, x_{\overline{W}}^k\right) \geq |\gamma \alpha_W^k \nabla_W f(x^k)^T d_W^k| \geq 0,$$

which, by (6), yields

$$\lim_{k \to \infty} \alpha_W^k \nabla_W f(x^k)^T d_W^k = 0. \tag{18}$$

Reasoning by contradiction, let us suppose that $\hat{x}$ is not a stationary point. Then, by Assumption 1, we have that

$$\limsup_{k \to \infty} \nabla_W f(x^k)^T d_W^k < 0.$$

Let $K \subseteq \{0, 1, 2 \ldots\}$ be a subset of the iteration set such that

$$\lim_{k \to \infty, k \in K} \nabla_W f(x^k)^T d_W^k < 0, \tag{19}$$

so that, for $k \in K$ and sufficiently large, d_W^k satisfies $\nabla_W f(x^k)^T d_W^k < 0$. Then, it follows from (18) that

$$\lim_{k \to \infty, k \in K} \alpha_W^k = 0. \tag{20}$$

By definition of the Procedure `Step_length` and by point (i) of Assumption 1 we have that, for all $k \in K$, the initial stepsize β is such that $\beta \geq \min\{1, \beta_u\} > 0$. Hence, by virtue of (20), the initial stepsize will be reduced at least once, so that, for $k \in K$ and sufficiently large,

$$f(x^k) - f\left(x_W^k + \frac{\alpha_W^k}{\delta} d_W^k, x_{\overline{W}}^k\right) < -\gamma \frac{\alpha_W^k}{\delta} \nabla_W f(x^k)^T d_W^k.$$

By using the mean value theorem, we obtain from the above relation

$$- \nabla_W f\left(x^k + \eta^k \frac{\alpha_W^k}{\delta} d_W^k, x_{\overline{W}}^k\right)^T d_W^k < -\gamma \nabla_W f(x^k)^T d_W^k, \tag{21}$$

where $\eta^k \in (0, 1)$.

Since $\{d_W^k\}$ is bounded by Assumption 1, we can find a further set of indices, that we relabel again K, such that

$$\lim_{k\to\infty, k\in K} d_W^k = \hat{d}_W \quad \text{and} \quad \lim_{k\to\infty, k\in K} x^k = \hat{x}$$

Hence, taking the limit in (21) for $k \to \infty$ and $k \in K$, it follows that

$$0 \leq (1-\gamma)\nabla_W f(\hat{x})^T \hat{d}_W,$$

which, recalling that $\gamma \in (0, 1)$, yields

$$0 \leq \nabla_W f(\hat{x})^T \hat{d}_W.$$

The proof follows by noting that the above inequality contradicts (19), that is,

$$\lim_{k\to\infty, k\in K} \nabla_W f(x^k)^T d_W^k = \nabla_W f(\hat{x})^T \hat{d}_W < 0.$$

$\square$

To prove the main convergence results of the paper we need to introduce some more technical notation and preliminary results concerning problem (1).

Given a point $x \in \mathcal{F}$ let us define the following index sets, following [11,13]

$$\begin{aligned} R(x) &= \{i : (x_i < u_i \quad \text{and} \quad a_i > 0) \quad \text{or} \quad (x_i > l_i \quad \text{and} \quad a_i < 0)\}, \\ S(x) &= \{i : (x_i < u_i \quad \text{and} \quad a_i < 0) \quad \text{or} \quad (x_i > l_i \quad \text{and} \quad a_i > 0)\}. \end{aligned}$$

Further, given indices $i, j \in \{1, \ldots, n\}$, with $i \neq j$, we denote by $d^{i,j}$ a vector belonging to $\mathbb{R}^n$ such that

$$d_h^{i,j} = \begin{cases} 1/a_i, & \text{if } h = i \\ -1/a_j, & \text{if } h = j \\ 0, & \text{otherwise} \end{cases}$$

Then, we denote by $D_{RS}(x)$ the set of directions $d^{i,j}$ with $i \in R(x)$ and $j \in S(x)$, namely

$$D_{RS}(x) = \bigcup_{\substack{i\in R(x)\\ j\in S(x)\\ i\neq j}} d^{i,j}.$$

It is easy to see that the set $D_{RS}(x)$ is a subset of the set of feasible directions at x. In [13] it is proved that it contains the generators of $D(x)$. From this we can easily derive the following characterization of the stationary points of Problem (1) which uses only the directions in $D_{RS}(x)$ (see also [14]).

Proposition 6.1 *A feasible point $x^* \in \mathcal{F}$ is stationary for Problem* (1) *if and only if*

$$\nabla f(x^*)^T d^{i,j} \geq 0 \qquad \forall d^{i,j} \in D_{RS}(x^*).$$

Now, given a subset of indices W and a feasible point $x \in \mathcal{F}$, let us consider subproblem $\mathcal{P}_W(x)$. We denote by $D_W^{RS}(x)$ the subset of feasible directions of $D_W(x_W)$ with exactly two nonzero components, namely

$$D_W^{RS}(x) = \bigcup_{\substack{i \in R(x) \cap W \\ j \in S(x) \cap W \\ i \neq j}} d_W^{i,j}. \tag{22}$$

Now we are ready to prove Proposition 4.2.

Proof of Proposition 4.2 Let $\bar{x}$ be any limit point of a subsequence of $\{x^k\}$, i.e., there exists an infinite subset $K \subseteq \{0, 1, \ldots\}$ such that

$$\lim_{k \to \infty, k \in K} x^k = \bar{x}.$$

By contradiction, let us assume that $\bar{x}$ is not a stationary point for Problem (1). By Proposition 6.1 there exists at least a pair $(i, j) \in R(\bar{x}) \times S(\bar{x})$, and a direction $d^{i,j} \in D_{RS}(\bar{x})$ such that:

$$\nabla f(\bar{x})^T d^{i,j} < 0. \tag{23}$$

By Condition PIP on the family of working sets, we know that $W^h \in \mathcal{W}$, with $h \in \{1, \ldots, M\}$, exists such that $(i, j) \subseteq W^h$. Let us consider the subvector $d_{W^h}^{i,j} \in D_{W^h}^{RS}(\bar{x})$, so that we have

$$\nabla_{W^h} f(\bar{x})^T d_{W^h}^{i,j} < 0, \tag{24}$$

that is to say that $\bar{x}_{W^h}$ is not a stationary point of problem $\mathcal{P}_{W^h}(\bar{x})$ as well. By construction, we have

$$f\left(x^{k+1}\right) \leq f\left(y_{W^h}^k, x_{\overline{W^h}}^k\right) \leq f\left(x^k\right)$$

so that the sequence $\{f(x^k)\}_K$ is not increasing and thus converges to $f(\bar{x})$. Hence, we get that

$$\lim_{k \to \infty, k \in K} \left(f(x^k) - f\left(y_{W^h}^k, x_{\overline{W^h}}^k\right)\right) = 0. \tag{25}$$

Then, using Proposition 4.1, we have that

$$\nabla_{W^h} f(\bar{x})^T \bar{d}_{W^h} \geq 0, \quad \forall\ \bar{d}_{W^h} \in D^{RS}_{W^h}(\bar{x})$$

which, for $\bar{d}_{W^h} = d^{i,j}_{W^h}$, contradicts (24) and (23), thus concluding the proof. □

References

1. Bertsekas, D.P., Tsitsiklis, J.N.: Parallel and distributed computation. Prentice-Hall, Englewood Cliffs (1989)
2. Bomze, I.M.: Evolution towards the maximum clique. J. Global Optim. **10**, 143–164 (1997)
3. Cassioli, A., Sciandrone, M.: A convergent decomposition method for box-constrained optimization problems. Optim. Lett. **3**, 397–409 (2009)
4. Dai, Y.-H., Fletcher, R.: New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. Math. Program. **106**(3), 403–421 (2006)
5. Ferris, M.C., Mangasarian, O.L.: Parallel variable distribution. SIAM J. Optim. **4**, 815–832 (1994)
6. Glasmachers, T., Igel, C.: Maximum-gain working set selection for SVMs. J. Mach. Learn. Res. **7**, 1437–1466 (2006)
7. Grippo, L., Sciandrone, M.: Globally convergent block-coordinate techniques for unconstrained optimization. Optim. Methods Softw. **10**(4), 587–637 (1999)
8. Grippo, L., Sciandrone, M.: On the convergence of the block nonlinear Gauss–Seidel method under convex constraints. Oper. Res. Lett. **26**(3), 127–136 (2000)
9. Joachims, T.: Making large scale SVM learning practical. In: Schölkopf, C.B.B., Burges, C.J.C., Smola, A. (eds.) Advances in Kernel Methods—Support Vector Learning, MIT Press, Cambridge (1998)
10. Keerthi, S.S., Gilbert, E.G.: Convergence of a generalized SMO algorithm for SVM classifier design. Mach. Learn. **46**, 351–360 (2002)
11. Lin, C.-J.: On the convergence of the decomposition method for Support Vector Machines. IEEE Trans. Neural Netw. **12**, 1288–1298 (2001)
12. Lin, C.-J.: Asymptotic convergence of an SMO algorithm without any assumptions. IEEE Trans. Neural Netw. **13**, 248–250 (2002)
13. Lin, C.-J., Lucidi, S., Palagi, L., Risi, A., Sciandrone, M.: A decomposition algorithm model for singly linearly constrained problems subject to lower and upper bounds. J. Optim. Theory Appl. **141**(1), 107–126 (2009)
14. Lucidi, S., Palagi, L., Risi, A., Sciandrone, M.: A convergent decomposition algorithm for support vector machines. Comput. Optim. Appl. **38**, 217–234 (2007)
15. Lucidi, S., Palagi, L., Risi, A., Sciandrone, M.: A convergent hybrid decomposition algorithm model for SVM training. IEEE Trans. Neural Netw. **20**(6), 1055–1060 (2009)
16. Markowitz, H.: Portfolio selection. J. Finance **7**(1), 77–91 (1952)
17. Motzkin, T.S., Strauss, E.G.: Maxima for graphs and a new proof of a theorem of Turan. Can. J. Math. **17**, 533–540 (1965)
18. Palagi, L., Sciandrone, M.: On the convergence of a modified version of SVM^{light} algorithm. Optim. Methods Softw. **20**(2–3), 311–328 (2005)
19. Serafini, T., Zanghirati, G., Zanni, L.: Gradient projection methods for large quadratic programs and applications in training support vector machines. Optim. Methods Softw. **20**, 353–378 (2005)
20. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)