Minimizing quadratic functions over non-negative cone with the rate of convergence and finite termination. *

Zdeněk Dostál † Joachim Schöberl ‡

April 12, 2002

Abstract

A new active set based algorithm is proposed that uses the conjugate gradient method to explore the face of the feasible region defined by the current iterate and the reduced gradient projection with the fixed steplength to expand the active set. The precision of approximate solutions of the auxiliary unconstrained problems is controlled by the norm of violation of the Karush-Kuhn-Tucker conditions at active constraints and the scalar product of the reduced gradient and the reduced gradient projection. The modifications were exploited to find the rate of convergence in terms of the spectral condition number of the Hessian matrix, to preserve its finite termination property even for problems whose solution does not satisfy the strict complementarity condition, and to avoid any backtracking at the cost of evaluation of an upper bound for the spectral radius of the Hessian matrix. The performance of the algorithm is illustrated on solution of the inner obstacle problems. The result is an important ingredient in development of scalable algorithms for numerical solution of elliptic variational inequalities.

AMS classification: Primary 65K05; Secondary 90C20

Key words and phrases: quadratic programming, bound constraints, inexact active set strategy, rate of convergence, finite termination.

Short title: Minimizing quadratic with the rate of convergence

*This research has been supported by CEZ:J17/98:272400019 and by grant GAČR 101/00/0538.

†FEI VŠB-Technical University Ostrava, Tř 17 listopadu, CZ-70833 Ostrava, Czech Republic E-mail: zdenek.dostal@vsb.cz

‡Department of Mathematics, Johannes Kepler University, A-4040 Linz, Austria

1 Introduction

We shall be concerned with the problem to find

$$\min_{x \in \Omega} f(x) \tag{1.1}$$

with $\Omega = \{x : 0 \leq x\}$, $f(x) = \frac{1}{2}x^\top Ax - x^\top b$, b a given column n-vector, and A an $n \times n$ symmetric positive definite matrix. We shall be interested especially in problems with n large and A reasonably conditioned or preconditioned [1], so that the application of the conjugate gradient based methods is suitable. Problems of this type arise e.g. in application of the duality based domain decomposition methods to the solution of the discretized variational inequalities [13, 24, 25] or in the solution of auxiliary problems in the augmented Lagrangian type algorithms for minimization of convex quadratic functions subject to more general constraints [12].

A class of efficient algorithms for the solution of (1.1) that is relevant for our research is based on the active set method [16]. The method dates back at least to Polyak [23] who proposed to use the conjugate gradient method to minimize the cost function on the face of the region defined by the current iterate until either the minimum on the current face was reached or an unfeasible iteration was generated. In the first case the constraints that violated the Karush-Kuhn-Tucker conditions were released, while in the second case the conjugate gradient step was shortened to generate a feasible iterate at the boundary of Ω which typically resulted in adding one index to the active set. Since the Polyak algorithm decreases the cost function in each step, the faces with minimizers can never reappear. The number of all faces being finite, the algorithm necessarily finds the face with the solution of (1.1) and then the solution itself in a finite number of steps.

The Polyak algorithm suffers from several drawbacks. The first one is related to the trial conjugate iterates that are not feasible. An unpleasant consequence of the Polyak strategy is a lower bound on the number of iterations in terms of the difference between the numbers of the active constraints in the initial approximation and the solution. To relieve the problem, Dembo and Tulowitzski [6] and Yang and Tolle [27] proposed to use projections with backtracking.

Another drawback concerns the basic approach combining the conjugate gradient method, which is now understood as an efficient iterative method for approximate solution of linear systems [1], and the finite termination strategy based on combinatorial reasoning that gives extremely poor bound on the number of iterations that are necessary to find the solution of (1.1). On the basis of results of Calamai and Moré [4], Moré and Toraldo [21] proposed an algorithm that also exploits the conjugate gradients, but its convergence is driven by the gradient projections with the steplength satisfying the sufficient decrease condition. The steplength is found, as in earlier algorithms, by possibly expensive backtracking. In spite of purely iterative basis of their algorithm, the authors proved that their algorithm preserved the finite termination property of the original algorithm provided the solution satisfies the strict complementarity condition.

The last serious drawback of the Polyak algorithm that is relevant for our work concerns precision of the solution of the auxiliary minimization problems. The exact minimization assumed exact arithmetics, so that the original theory did not support standard computer implementation of the algorithm, and there were doubts about its efficiency, as it has been observed by O'Leary [22] that the number of iterations may be reduced to about a half with the algorithm that refined the accuracy of the conjugate gradient minimization during the course of iterations. The result does not seem surprising as the conjugate gradient iterations reduce directly only the norm of violation of the Karush-Kuhn-Tucker conditions at free variables even at the later stage of the minimization when it is relatively small as compared to that at the active variables. The effective theoretically supported strategies for adaptive precision control were presented independently by Friedlander and Martínez with their collaborators and Dostál [17, 18, 19, 3, 9, 10, 8]. The basic idea was to control the precision of the solution of the auxiliary problems by the ratio of norms of violation of the Karush-Kuhn-Tucker conditions at free and active variables. These results offered exploitation of the conjugate gradient method as an effective approximate iterative minimizer in faces, relaxed conditions for implementation of the gradient projections, and gave guidance for development of effective algorithms applied to the solution of complex engineering problems with up to millions of variables [13, 14, 15, 24, 25]. Moreover, it was shown that the finite termination property of the original Polyak algorithm may be preserved with inexact solution of the auxiliary problems even for problems whose solution does not satisfy the strict complementarity condition [9, 10]. By our opinion, the last result is important because it indicates that the algorithm does not suffer from oscillations often attributed to the active set based algorithms.

The common drawbacks of all the above mentioned strategies were possible backtracking in search of the gradient projection step and the lack of results on the rate of convergence. A key to further progress was the result of Schöberl [24, 25] who gave the rate of convergence of the gradient projection method for (1.1) in terms of the spectral condition number $\kappa(A)$ of A. Recently, it has been observed that this nice result may be plugged into the proportioning algorithm [11] to get a similar result. Moreover, the algorithm did not require any backtracking and it turned out that it was possible to preserve the finite termination property for the problems whose solution satisfies the strict complementarity condition.

In this paper, we give a new proof of the improved rate of convergence of the projection method. Then we introduce the precision control that takes into account also the part of the reduced gradient that can be used in the reduced gradient projection step. Finally a modification of the proportioning algorithm is presented that enables us to prove both the rate of convergence and the finite termination property for problems with solution that does not satisfy the strict complementarity condition. The algorithm works with the fixed steplength of the reduced gradient projection step so that it does not use any backtracking, but requires an upper bound on the spectral radius of the Hessian A of the cost function f. The performance of the algorithm is demonstrated on a numerical example.

2 Notations and preliminaries

It is well known that the solution to the problem (1.1) always exists, and it is necessarily unique [2]. For arbitrary n-vector x, let us define the gradient $g = g(x)$ of f by

$$g = g(x) = Ax - b. \tag{2.1}$$

Then the unique solution $\overline{x}$ of (1.1) is fully determined by the Karush-Kuhn-Tucker optimality conditions [2] so that for $i = 1, \ldots, n$,

$$\overline{x}_i = 0 \text{ implies } \overline{g}_i \geq 0 \text{ and } \overline{x}_i > 0 \text{ implies } \overline{g}_i = 0. \tag{2.2}$$

Let $\mathcal{N}$ denote the set of all indices so that

$$\mathcal{N} = \{1, 2, \ldots, n\}.$$

The set of all indices for which $x_i = 0$ is called an *active set* of x. We shall denote it by $\mathcal{A}(x)$ so that

$$\mathcal{A}(x) = \{i \in \mathcal{N} : x_i = 0\}.$$

Its complement

$$\mathcal{F}(x) = \{i \in \mathcal{N} : x_i \neq 0\}$$

and subset

$$\mathcal{B}(x) = \{i \in \mathcal{N} : x_i = 0 \text{ and } g_i > 0\}$$

are called a *free set* and a *binding set*, respectively.

To enable an alternative reference to the Karush-Kuhn-Tucker conditions (2.2), we shall introduce a notation for the *free gradient* φ and the *chopped gradient* β that are defined by

$$\varphi_i(x) = g_i(x) \text{ for } i \in \mathcal{F}(x), \varphi_i(x) = 0 \text{ for } i \in \mathcal{A}(x)$$

$$\beta_i(x) = 0 \text{ for } i \in \mathcal{F}(x), \beta_i(x) = g_i^-(x) \text{ for } i \in \mathcal{A}(x)$$

where we have used the notation $g_i^- = \min\{g_i, 0\}$. Thus the Karush- Kuhn-Tucker conditions (2.2) are satisfied iff the *projected gradient* $\nu(x) = \varphi(x) + \beta(x)$ is equal to zero.

The Euclidean norm and the $A-$energy norm of x will be denoted by $\|x\|$ and $\|x\|_A$, respectively. Thus $\|x\|^2 = x^\top x$ and $\|x\|_A^2 = x^\top Ax$. Analogous notation will be used for the induced matrix norm, so that the spectral condition number $\kappa(A)$ of the matrix A is defined by

$$\kappa(A) = \|A\|\|A^{-1}\|.$$

The projection P_Ω to Ω is defined for any n-vector x by

$$P_\Omega(x) = x^+$$

where x^+ is the vector with entries $x_i^+ = \max\{x_i, 0\}$.

3 Algorithm with proportioning and gradient projections

The algorithm for the solution of (1.1) that we propose here combines the proportioning algorithm mentioned above with the gradient projections. It exploits a given constant $\Gamma > 0$, a test to decide about leaving the face and three types of steps to generate a sequence of iterates $\{x^k\}$ that approximate the solution of (1.1).

The *expansion step* is defined by

$$x^{k+1} = P_\Omega\left(x^k - \overline{\alpha}\varphi(x^k)\right) \tag{3.1}$$

with the fixed steplength $\overline{\alpha} \in (0, \|A\|^{-1}]$. This step may expand the current active set. To describe it without P_Ω, let us introduce, for any $x \in \Omega$, the *reduced free gradient* $\widetilde{\varphi}(x)$ with the entries

$$\widetilde{\varphi}_i = \widetilde{\varphi}_i(x) = \min\{x_i/\overline{\alpha}, \varphi_i\},$$

so that

$$P_\Omega\left(x - \overline{\alpha}\varphi(x)\right) = x - \overline{\alpha}\widetilde{\varphi}(x). \tag{3.2}$$

Using the new notation, we can write also

$$P_\Omega\left(x - \overline{\alpha}g(x)\right) = x - \overline{\alpha}\left(\widetilde{\varphi}(x) + \beta(x)\right). \tag{3.3}$$

If the inequality

$$||\beta(x^k)||^2 \leq \Gamma^2\widetilde{\varphi}(x^k)^\top\varphi(x^k) \tag{3.4}$$

holds then we call the iterate x^k *strictly proportional.* The test (3.4) is used to decide which component of the projected gradient $\nu(x^k)$ will be reduced in the next step. Notice that the right-hand side of (3.4) blends the information about the current free gradient and its part that can be used in the expansion step, while the related relations in [17, 18, 19, 3, 9, 8, 7] consider only the norm of the free gradient.

The *proportioning step* is defined by

$$x^{k+1} = x^k - \alpha_{cg}\beta(x^k) \tag{3.5}$$

with the steplength α_{cg} that minimizes $f\left(x^k - \alpha\beta(x^k)\right)$. It is easy to check [1, 20] that α_{cg} that minimizes $f(x - \alpha d)$ for given d and x may be evaluated by the formula

$$\alpha_{cg} = \alpha_{cg}(d) = \frac{d^\top g(x)}{d^T A d}. \tag{3.6}$$

The purpose of the proportioning step is to remove indices from the active set. Note that if $x^k \in \Omega$, then $x^{k+1} = x^k - \alpha_{cg}\beta(x^k) \in \Omega$.

The *conjugate gradient step* is defined by

$$x^{k+1} = x^k - \alpha_{cg} p^k \tag{3.7}$$

where p^k is the conjugate gradient direction [1, 20] which is constructed recurrently. The recurrence starts (or restarts) from $p^s = \varphi(x^s)$ whenever x^s is generated by the expansion step or the proportioning step. If p^k is known, then p^{k+1} is given by the formulae [1, 20]

$$p^{k+1} = \varphi(x^k) - \gamma p^k, \quad \gamma = \frac{\varphi(x^k)^\top A p^k}{(p^k)^\top A p^k}. \tag{3.8}$$

The basic property of the conjugate directions $p^s, \ldots, p^k$ that are generated by the recurrence (3.8) from the restart p^s is their mutual A-orthogonality, i. e. $(p^i)^\top A p^j = 0$ for $i, j \in \{s, \ldots, k\}$, $i \neq j$. It follows easily [1, 20] that

$$f(x^{k+1}) = \min\{f(x^s + y) : y \in \operatorname{Span}\{p^s, \ldots, p^k\}\} \tag{3.9}$$

where $\operatorname{Span}\{p^s, \ldots, p^k\}$ denotes the vector space of all linear combinations of the vectors $p^s, \ldots, p^k$. The conjugate gradient steps are used to carry out the minimization in the face

$$\mathcal{W}_I = \{x : x_i = 0 \text{ for } i \in I\} \tag{3.10}$$

given by $I = \mathcal{A}(x^s)$ efficiently.

Let us define the algorithm that we propose in the form that is convenient for analysis.

Algorithm 3.1. Modified proportioning with reduced gradient projections (MPRGP).
Let $x^0 \in \Omega$, $\overline{\alpha} \in (0, \|A\|^{-1}]$, and $\Gamma > 0$ be given. For $k \geq 0$ and x^k known, choose x^{k+1} by the following rules:
(i) If $\nu(x^k) = 0$, set $x^{k+1} = x^k$.
(ii) If x^k is strictly proportional and $\nu(x^k) \neq 0$, try to generate x^{k+1} by the conjugate gradient step. If $x^{k+1} \in \Omega$, then accept it, else generate x^{k+1} by the expansion step.
(iii) If x^k is not strictly proportional, define x^{k+1} by proportioning.
More details about implementation of the algorithm may be found in Section 7.

4 Estimate for the gradient projection

We shall first present a new simplified proof of the strengthening of the above mentioned result by Schöberl [24].

Theorem 4.1. *Let $\overline{x}$ denote the unique solution of (1.1), let λ_1 denote the smallest eigenvalue of A, $\|A\| \leq 1$, and $x \in \Omega$. Then*

$$f\left(P_\Omega\left(x - g(x)\right)\right) - f(\overline{x}) \leq (1 - \lambda_1)\left(f(x) - f(\overline{x})\right). \tag{4.1}$$

Proof: Let $x \in \Omega$ be arbitrary but fixed, so that we can define a quadratic function

$$F(y) = f(y) + \frac{1}{2}(y - x)^\top (I - A)(y - x).$$

It is defined so that

$$F(y) \geq f(y) \text{ for any } y \in I\!R, \quad F(x) = f(x), \text{ and } \nabla F(x) = \nabla f(x).$$

Moreover, the Hessian matrix of F is equal to I, so that

$$\nabla F(y) = \nabla f(y) + (I - A)(y - x) = y - x + g(x).$$

We shall define also the *reduced gradient* $\widetilde{g}(x)$ with the entries

$$\widetilde{g}_i = \widetilde{g}_i(x) = \min\{x_i, g_i\},$$

so that

$$P_\Omega\left(x - g(x)\right) = x - \widetilde{g}, \quad 0 \leq g_i(x) - \widetilde{g}_i(x), \quad 0 \leq x_i - \widetilde{g}_i(x).$$

Since any $y \in \Omega$ may be written in the form $y = x - \widetilde{g}(x) + d$ with $d_i \geq \widetilde{g}_i(x) - x_i$, we get

$$\begin{aligned}
F(y) &= F(P_\Omega\left(x - g(x)\right) + d) = F\left(P_\Omega(x - g(x))\right) + d^\top \nabla F(x - \widetilde{g}(x)) + \frac{1}{2}\|d\|^2 \\
&= F\left(P_\Omega(x - g(x))\right) + d^\top \left(g(x) - \widetilde{g}(x)\right) + \frac{1}{2}\|d\|^2 \\
&\geq F\left(P_\Omega(x - g(x))\right) + (\widetilde{g}_i(x) - x_i)^\top \left(g(x) - \widetilde{g}(x)\right) + \frac{1}{2}\|d\|^2 \\
&= F\left(P_\Omega(x - g(x))\right) + \frac{1}{2}\|d\|^2 \geq F(P_\Omega\left(x - g(x)\right)).
\end{aligned} \tag{4.2}$$

Let us now denote by $[\overline{x}, x]$ the convex hull of $\{\overline{x}, x\}$ and let $d = \overline{x} - x$. Using (4.2), $[\overline{x}, x] = \{x + td : t \in [0, 1]\} \subseteq \Omega$ and $\lambda_1\|d\|^2 \leq d^\top Ad$, we get

$$f(P_\Omega\left(x - g(x)\right)) - f(\overline{x}) \leq F(P_\Omega\left(x - g(x)\right)) - f(\overline{x}) = \min\{F(y) - f(\overline{x}) : y \in \Omega\}$$

$$
\begin{aligned}
&\leq \min\{F(y) - f(\overline{x}) : y \in [\overline{x}, x]\} \\
&= \min\{F(x + td) - f(x + d) : t \in [0, 1]\} \\
&= \min\{td^\top g(x) + \frac{1}{2}t^2\|d\|^2 - d^\top g(x) - \frac{1}{2}d^\top Ad : t \in [0, 1]\} \\
&\leq \lambda_1 d^\top g(x) + \frac{1}{2}\lambda_1^2\|d\|^2 - d^\top g(x) - \frac{1}{2}d^\top Ad \\
&\leq \lambda_1 d^\top g(x) + \frac{1}{2}\lambda_1 d^\top Ad - d^\top g(x) - \frac{1}{2}d^\top Ad \\
&= (\lambda_1 - 1)(d^\top g(x) + \frac{1}{2}d^\top Ad) = (\lambda_1 - 1)\left(f(x + d) - f(x)\right) \\
&= (1 - \lambda_1)\left(f(x) - f(\overline{x})\right). \qquad \square \qquad (4.3)
\end{aligned}
$$

Corollary 4.2. *Let $\overline{x}$ denote the unique solution of (1.1), let λ_1 denote the smallest eigenvalue of A, $\overline{\alpha} \in (0, \|A\|^{-1}]$, and let $x \in \Omega$. Then*

$$
f\left(P_\Omega\left(x - \overline{\alpha}g(x)\right)\right) - f(\overline{x}) \leq \rho\left(f(x) - f(\overline{x})\right) \tag{4.4}
$$

where

$$
\rho = (1 - \overline{\alpha}\lambda_1) < 1. \tag{4.5}
$$

Proof: Apply Theorem 4.1 to the function $\overline{\alpha}f(x)$. $\square$

5 Rate of convergence

Theorem 5.1. *Let $\Gamma > 0$ be a given constant, let λ_1 denote the smallest eigenvalue of A, $\widehat{\Gamma} = \max\{\Gamma, \Gamma^{-1}\}$, let $\overline{x}$ denote the unique solution of (1.1), and let $\{x^k\}$ denote the sequence generated by Algorithm 3.1 with $\overline{\alpha} \in (0, \|A\|^{-1}]$. Then*

$$
f(x^{k+1}) - f(\overline{x}) \leq \eta\left(f(x^k) - f(\overline{x})\right) \tag{5.1}
$$

where $\overline{x}$ denotes the unique solution of (1.1) and

$$
\eta = 1 - \frac{\overline{\alpha}\lambda_1}{2 + 2\widehat{\Gamma}^2}. \tag{5.2}
$$

The error in the A-energy norm is bounded by

$$
\|x^k - \overline{x}\|_A^2 \leq 2\eta^k\left(f(x^0) - f(\overline{x})\right). \tag{5.3}
$$

Proof: We shall compare separately all three possible steps with the gradient projection estimate (4.4), using that for any vectors x and d

$$f(x+d) = f(x) + d^\top g(x) + \frac{1}{2} d^\top A d \geq f(x) + d^\top g(x). \tag{5.4}$$

In particular, combining (5.4) with (3.3), we get the estimate

$$f\left(P_\Omega\left(x^k - \overline{\alpha} g(x^k)\right)\right) \geq f(x^k) - \overline{\alpha}(\widetilde{\varphi}(x^k)^\top \varphi(x^k) + \|\beta(x^k)\|^2). \tag{5.5}$$

If x^{k+1} is generated by the expansion step ((3.1)), then after using

$$\overline{\alpha}^2 \widetilde{\varphi}(x^k)^\top A \widetilde{\varphi}(x^k) \leq \overline{\alpha}^2 \|A\| \|\widetilde{\varphi}(x^k)\|^2 \leq \overline{\alpha} \|\widetilde{\varphi}(x^k)\|^2 \leq \overline{\alpha} \widetilde{\varphi}(x^k)^\top \varphi(x^k),$$

(5.4) and a little manipulation, we get

$$\begin{aligned} f(x^{k+1}) &= f\left(x^k - \overline{\alpha}\widetilde{\varphi}(x^k))\right) = f(x^k) - \overline{\alpha}\widetilde{\varphi}(x^k)^\top \varphi(x^k) + \frac{\overline{\alpha}^2}{2} \widetilde{\varphi}(x^k)^\top A \widetilde{\varphi}(x^k) \\ &\leq f(x^k) - \frac{\overline{\alpha}}{2} \widetilde{\varphi}(x^k)^\top \varphi(x^k). \end{aligned} \tag{5.6}$$

If x^{k+1} is generated by the conjugate gradient step (3.7), then by (3.9) and (3.6)

$$f(x^{k+1}) \leq f(x^k - \alpha_{cg}\varphi(x^k)) = f(x^k) - \frac{1}{2} \frac{\|\varphi(x^k)\|^4}{\varphi(x^k)^\top A \varphi(x^k)}.$$

Taking into account the definition of $\overline{\alpha}$ and $\widetilde{\varphi}(x^k)$, we get

$$f(x^{k+1}) \leq f(x^k) - \frac{\overline{\alpha}}{2} \|\varphi(x^k)\|^2 \leq f(x^k) - \frac{\overline{\alpha}}{2} \widetilde{\varphi}(x^k)^\top \varphi(x^k). \tag{5.7}$$

Comparing (5.6) and (5.7), we may observe that we got the same estimates for the expansion and conjugate gradient steps. These steps are taken only when x^k is strictly proportional, so that

$$\|\beta(x^k)\|^2 \leq \Gamma^2 \widetilde{\varphi}(x^k)^\top \varphi(x^k). \tag{5.8}$$

After substituting (5.8) into (5.7), we get

$$f\left(P_\Omega\left(x^k - \overline{\alpha} g(x^k)\right)\right) \geq f(x^k) - \overline{\alpha}(1+\Gamma^2) \widetilde{\varphi}(x^k)^\top \varphi(x^k). \tag{5.9}$$

Thus for x^{k+1} generated by the expansion or conjugate gradient steps, we get after a little manipulation and application of (5.9) that

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) - \frac{\overline{\alpha}}{2} \widetilde{\varphi}(x^k)^\top \varphi(x^k) \\ &= \frac{1}{2(1+\Gamma^2)} \left(f(x^k) - \overline{\alpha}(1+\Gamma^2) \widetilde{\varphi}(x^k)^\top \varphi(x^k) + (1+2\Gamma^2) f(x^k) \right) \\ &\leq \frac{1}{2(1+\Gamma^2)} \left(f\left(P_\Omega\left(x^k - \overline{\alpha} g(x^k)\right)\right) + (1+2\Gamma^2) f(x^k) \right). \end{aligned} \tag{5.10}$$

After using (4.4) with $x = x^k$, (5.6) and a little manipulation, we get

$$f(x^{k+1}) \le \frac{\rho + 1 + 2\Gamma^2}{2 + 2\Gamma^2} f(x^k) + \frac{1-\rho}{2+2\Gamma^2} f(\overline{x}). \tag{5.11}$$

Let us finally assume that x^{k+1} is generated by the proportioning step (3.5), so that

$$\|\beta(x^k)\|^2 > \Gamma^2 \widetilde{\varphi}(x^k)^\top \varphi(x^k) \tag{5.12}$$

and

$$f(x^{k+1}) = f\left(x^k - \alpha_{cg}\beta(x^k)\right) = f(x^k) - \frac{1}{2}\frac{\|\beta(x^k)\|^4}{\beta(x^k)^\top A \beta(x^k)}.$$

Taking into account the definition of $\overline{\alpha}$, we get

$$f(x^{k+1}) \le f(x^k) - \frac{\overline{\alpha}}{2}\|\beta(x^k)\|^2, \tag{5.13}$$

where the right hand side may be rewritten in the form

$$f(x^k) - \frac{\overline{\alpha}}{2}\|\beta(x^k)\|^2 = \frac{1}{2(1+\Gamma^{-2})}\left(f(x^k) - \overline{\alpha}(1+\Gamma^{-2})\|\beta(x^k)\|^2 + (1+2\Gamma^{-2})f(x^k)\right). \tag{5.14}$$

We can also substitute (5.12) into (5.5) to get

$$f\left(P_\Omega\left(x^k - \overline{\alpha} g(x^k)\right)\right) > f(x^k) - \overline{\alpha}(1+\Gamma^{-2})\|\beta(x^k)\|^2. \tag{5.15}$$

Substituting (5.15) into (5.14) yields

$$f(x^k) - \frac{\overline{\alpha}}{2}\|\beta(x^k)\|^2 < \frac{1}{2(1+\Gamma^{-2})}\left(f\left(P_\Omega\left(x^k - \overline{\alpha} g(x^k)\right)\right) + (1+2\Gamma^{-2})f(x^k)\right). \tag{5.16}$$

After using (4.4) with $x = x^k$, (5.13) and a little manipulation, we get

$$f(x^{k+1}) \le \frac{\rho + 1 + 2\Gamma^{-2}}{2 + 2\Gamma^{-2}} f(x^k) + \frac{1-\rho}{2+2\Gamma^{-2}} f(\overline{x}). \tag{5.17}$$

Comparing our result with (5.11) we conclude that we have the same estimate for all three steps of Algorthm 3.1. The proof of (5.1) may be completed by direct computation.

To get the error bound (5.3), notice that the Karush-Kuhn-Tucker conditions (2.2) imply that $(x^k - \overline{x})^\top g(\overline{x}) \ge 0$, so that by (5.4) applied to $d = x^k - \overline{x}, x = \overline{x}$ and by (5.1)

$$\|x^k - \overline{x}\|_A^2 = 2\left(f(x^k) - f(\overline{x}) - (x^k - \overline{x})^\top g(\overline{x})\right) \le 2\left(f(x^k) - f(\overline{x})\right) \le 2\eta^k\left(f(x^0) - f(\overline{x})\right).$$

□

Theorem 5.1 yields the best estimate for $\Gamma = \widehat{\Gamma} = 1$ in agreement with heuristics that we should leave the face when the chopped gradient dominates the violation of the Karush-Kuhn-Tucker conditions. The formula for η than reads

$$\eta = 1 - \frac{\overline{\alpha}\lambda_1}{4}.$$

6 Finite termination

Lemma 6.1. *Let $\{x^k\}$ denote the sequence generated by Algorithm 3.1 with a given $\Gamma > 0$. Then there is k_0 such that for $k \geq k_0$*

$$\mathcal{F}(\overline{x}) \subseteq \mathcal{F}(x^k), \quad \mathcal{F}(\overline{x}) \subseteq \mathcal{F}(x^k - \overline{\alpha}\widetilde{\varphi}(x^k)) \quad \textit{and} \quad \mathcal{B}(\overline{x}) \subseteq \mathcal{B}(x^k). \tag{6.1}$$

Proof: Since (6.1) is trivially satisfied when there is $k = k_0$ such that $x^k = \overline{x}$, we shall assume in what follows that $x^k \neq \overline{x}$ for any $k \geq 0$.

Let us first assume that $\mathcal{F}(\overline{x}) \neq \emptyset$ and $\mathcal{B}(\overline{x}) \neq \emptyset$, so that

$$\epsilon = \min\{\mathrm{\overline{x}_i : i \in \mathcal{F}(\overline{x})}\} > 0 \quad \text{and} \quad \delta = \min\{\mathrm{g_i(\overline{x}) : i \in \mathcal{B}(\overline{x})}\} > 0.$$

Since by Theorem 5.1 $\{x^i\}$ converges to $\overline{x}$, there is k_0 such that for any $k \geq k_0$

$$g_i(x^k) \leq \frac{\epsilon}{4} \text{ for } \mathrm{i} \in \mathcal{F}(\overline{\mathrm{x}}) \tag{6.2}$$
$$x_i^k \geq \frac{\overline{\alpha}\epsilon}{2} \text{ for } \mathrm{i} \in \mathcal{F}(\overline{\mathrm{x}}) \tag{6.3}$$
$$x_i^k \leq \frac{\overline{\alpha}\delta}{4} \text{ for } \mathrm{i} \in \mathcal{B}(\overline{\mathrm{x}}) \tag{6.4}$$
$$g_i(x^k) \geq \frac{\delta}{2} \text{ for } \mathrm{i} \in \mathcal{B}(\overline{\mathrm{x}}). \tag{6.5}$$

In particular, for $k \geq k_0$, the first inclusion of (6.1) follows from (6.3), while the second inclusion follows from (6.2) and (6.3) as for $i \in \mathcal{F}(\overline{x})$

$$x_i^k - \overline{\alpha}\varphi_i(x^k) = x_i^k - \overline{\alpha}g_i(x^k) \geq \frac{\overline{\alpha}\epsilon}{2} - \frac{\overline{\alpha}\epsilon}{4} > 0$$

.

Let $k \geq k_0$ and observe that, by (6.4) and (6.5), for any $i \in \mathcal{B}(\overline{x})$

$$x_i^k - \overline{\alpha}g_i(x^k) \leq \frac{\overline{\alpha}\delta}{4} - \frac{\overline{\alpha}\delta}{2} < 0,$$

so that if some $x^{k+1}, k \geq k_0$ is generated by the expansion step and $i \in \mathcal{B}(\overline{x})$, then

$$x_i^{k+1} = (x_i^k - \overline{\alpha}g_i(x^k))^+ = 0$$

and $\mathcal{B}(x^{k+1}) \supseteq \mathcal{B}(\overline{x})$. Moreover, using (6.5) and definition of Algorithm 3.1, we can directly verify that if $\mathcal{B}(x^k) \supseteq \mathcal{B}(\overline{x})$, then also $\mathcal{B}(x^{k+1}) \supseteq \mathcal{B}(\overline{x})$. Thus it remains to prove that there must be some $s \geq k_0$ such that x^s is generated for by the expansion step.

Let us examine what may happen for $k \geq k_0$. First observe that if $x_i > 0$ for some $i \in \mathcal{B}(\overline{x})$, then we can never take the full conjugate direction step in the direction $p^k = \varphi(x^k)$. The reason is that

$$\alpha_{cg}(p^k) = \frac{\varphi(x^k)g(x^k)}{\varphi(x^k)^\top A\varphi(x^k)} = \frac{\|\varphi(x^k)\|^2}{\varphi(x^k)^\top A\varphi(x^k)} \geq \|A\|^{-1} \geq \overline{\alpha},$$

so that for $i \in \mathcal{F}(x^k) \bigcap \mathcal{B}(\overline{x})$, by (6.4) and (6.5),

$$x_i^k - \alpha_{cg} p_i^k = x_i^k - \alpha_{cg} g_i(x^k) \leq x_i^k - \overline{\alpha} g_i(x^k) \leq \frac{\overline{\alpha}\delta}{4} - \frac{\overline{\alpha}\delta}{2} < 0. \tag{6.6}$$

It follows by definition of Algorithm 3.1 that if $x^k, k \geq k_0$ is generated by the proportioning step, then the following trial conjugate gradient step is not feasible and x^{k+1} is necessarily generated by the expansion step.

To complete the proof, observe that Algorithm 3.1 can generate only a finite sequence of consecutive iterates by the conjugate gradient steps. In particular, it follows by the finite termination property of the conjugate gradient method [1] that if there is neither proportioning step nor the expansion step for $k \geq k_0$, then there is $l \leq n$ such that $\varphi(x^{k_0+l}) = 0$. Thus either $x^{k_0+l} = \overline{x}$ and by the definition of the step (i) of Algorithm 3.1 $\mathcal{B}(x^k) = \mathcal{B}(\overline{x})$ for $k \geq k_0 + l$, or x^{k_0+l} is not strictly proportional and the next iterate is generated by the proportioning step followed by the expansion step. This completes the proof, as the cases $\mathcal{F}(\overline{x}) = \emptyset$ and $\mathcal{B}(\overline{x}) = \emptyset$ may be easily proved by the specialization of the above arguments. $\square$

Corollary 6.2. *Let $\{x^k\}$ denote the sequence generated by Algorithm 3.1, and let the solution $\overline{x}$ satisfies the condition of strict complementarity, i.e. $\overline{x}_i = 0$ implies $g_i(\overline{x}) \neq 0$. Then there is $k \geq 0$ such that $x^k = \overline{x}$.*

Proof: If $\overline{x}$ satisfies the condition of strict complementarity, then $\mathcal{A}(\overline{x}) = \mathcal{B}(\overline{x})$, and by assumptions and Lemma 6.1, there is $k_0 \geq 0$ such that $\mathcal{F}(x^k) = F(\overline{x})$ and $\mathcal{B}(x^k) = \mathcal{B}(\overline{x})$. Thus all $x^k, k \geq k_0$ that satisfy $\overline{x} \neq x^{k-1}$ are generated by the conjugate gradient steps and by the finite termination property of the conjugate gradient method there is $k \leq k_0 + n$ such that $x^k = \overline{x}$. $\square$

Our final goal in this section is to obtain the result on finite termination of Algorithm 3.1 for the solution of (1.1) in case that it does not satisfy the condition of strict complementarity. We shall base our analysis on our earlier result on proportioning.

Theorem 6.3. *Let $x \in \Omega$ and $\kappa(A)^{1/2} \leq \Gamma$. Denote $I = \mathcal{A}(x)$, and suppose that*

$$\Gamma\|\varphi(x)\| < \|\beta(x)\|. \tag{6.7}$$

Then the vector $y = x - \|A\|^{-1}\beta(x)$ *satisfies*

$$j(y) < \min\{f(z) : z \in \mathcal{W}_I\} \tag{6.8}$$

where $\mathcal{W}_I$ *is defined in (3.10).*
Proof: See [10]. $\square$

Lemma 6.4. *Let* $\overline{\alpha} \in (0, \|A\|^{-1}]$, $x \in \Omega$ *and* $y = x - \overline{\alpha}\widetilde{\varphi}(x)$. *Then*

$$\|\varphi(y)\|^2 \le 4\widetilde{\varphi}^\top(x)\varphi(x) \quad \textit{and} \quad \|\beta(y)\| \ge \|\beta(x)\| - 2\|\widetilde{\varphi}(x)\|. \tag{6.9}$$

Proof: Let us denote $F = \mathcal{F}(y)$ and notice that $\mathcal{F}(y) \subseteq \mathcal{F}(x)$. Since

$$g(y) = g(x) - \overline{\alpha}A\widetilde{\varphi}(x) \quad \text{and} \quad \widetilde{\varphi}_F(x) = \varphi_F(x) = g_F(x), \tag{6.10}$$

we get

$$\|\varphi(y)\| = \|g_F(y)\| = \|g_F(x) - \overline{\alpha}A_F\widetilde{\varphi}(x)\| \le \|\widetilde{\varphi}_F(x)\| + \overline{\alpha}\|A_F\widetilde{\varphi}(x)\| \le 2\|\widetilde{\varphi}(x)\|. \tag{6.11}$$

Using (6.11) and the definition of $\widetilde{\varphi}(x)$, we get

$$\|\varphi(y)\|^2 \le 4\|\widetilde{\varphi}(x)\|^2 \le 4\widetilde{\varphi}^T(x)\varphi(x). \tag{6.12}$$

To prove the second inequality of (6.9), denote $B = \{i \in \mathcal{A}(x) : g_i(x) \le 0\}$ and notice that

$$\mathcal{A}(y) \supseteq \mathcal{A}(x) \supseteq B, \tag{6.13}$$

so that

$$\begin{aligned} \|\beta(y)\| = \|g_{\mathcal{A}(y)}(y)^-\| &\ge \|g_B(y)^-\| = \|(g_B(x) - \overline{\alpha}A_B\widetilde{\varphi}(x))^-\| \\ &= \|(\beta_B(x) - \overline{\alpha}A_B\widetilde{\varphi}(x))^-\|. \end{aligned} \tag{6.14}$$

Using in sequence $\|\beta_B(x)\| = \|\beta(x)\|$, $\|\overline{\alpha}A_B\widetilde{\varphi}(x)\| \le \|\widetilde{\varphi}(x)\|$, (6.14), properties of the norm, $\beta^-(x) = \beta(x)$, and $\|z - z^-\| \le \|z - t\|$ for any t with non-positive entries, we get

$$\begin{aligned} \|\beta(x)\| - \|\widetilde{\varphi}(x)\| - \|\beta(y)\| &\le \|\beta_B(x)\| - \|\overline{\alpha}A_B\widetilde{\varphi}(x)\| - \|(\beta_B(x) - \overline{\alpha}A_B\widetilde{\varphi}(x))^-)\| \\ &\le \|(\beta_B(x) - \overline{\alpha}A_B\widetilde{\varphi}(x))\| - \|(\beta_B(x) - \overline{\alpha}A_B\widetilde{\varphi}(x))^-\| \\ &\le \|(\beta_B(x) - \overline{\alpha}A_B\widetilde{\varphi}(x)) - (\beta_B(x) - \overline{\alpha}A_B\widetilde{\varphi}(x))^-\| \\ &\le \|(\beta_B(x) - \overline{\alpha}A_B\widetilde{\varphi}(x)) - \beta_B(x)\| \le \|\widetilde{\varphi}(x)\|. \end{aligned}$$

This proves the second inequality of (6.9). $\square$

Corollary 6.5. *Let* $\Gamma \geq 2$, $\overline{\alpha} \in (0, \|A\|^{-1}]$, $x \in \Omega$ *and*

$$\Gamma^2 \widetilde{\varphi}^\top(x)\varphi(x) < \|\beta(x)\|^2. \tag{6.15}$$

Then the vector $y = x - \overline{\alpha}\widetilde{\varphi}(x)$ *satisfies*

$$\frac{\Gamma - 2}{2}\|\varphi(y)\| < \|\beta(y)\|. \tag{6.16}$$

Proof: The inequality (6.16) holds trivially for $\Gamma = 2$. For $\Gamma > 2$, using (6.9), $\|\widetilde{\varphi}(x)\|^2 \leq \widetilde{\varphi}^\top(x)\varphi(x)$ and (6.15), we get

$$\|\beta(y)\| \geq \|\beta(x)\| - 2\|\widetilde{\varphi}(x)\| \geq \|\beta(x)\| - 2\sqrt{\widetilde{\varphi}^\top(x)\varphi(x)} > (1 - 2\Gamma^{-1})\|\beta(x)\|. \tag{6.17}$$

The inequality (6.16) follows from (6.15), (6.17) and (6.9). $\square$

Theorem 6.6. *Let* $\{x^k\}$ *denote the sequence generated by Algorithm 3.1. with*

$$\Gamma \geq 2\left(\sqrt{\kappa(A)} + 1\right). \tag{6.18}$$

Then there is $k \geq 0$ *such that* $x^k = \overline{x}$.

Proof: Let x^k be generated by Algorithm 3.1 and let Γ satisfy (6.18). Let k_0 be that of Lemma 6.1 and let $k \geq k_0$ be such that x^k is not strictly proportional, so that $\Gamma^2\widetilde{\varphi}(x^k)^\top\varphi(x^k) < \|\beta(x^k)\|^2$. Then by Corollary 6.5 the vector $y = x^k - \overline{\alpha}\widetilde{\varphi}(x^k)$ satisfies

$$\Gamma_1\|\varphi(y)\| < \|\beta(y)\| \tag{6.19}$$

with

$$\Gamma_1 = \frac{\Gamma - 2}{2} \geq \sqrt{\kappa(A)}.$$

Moreover, $y \in \Omega$ and by Lemma 6.1 and definition of y

$$\mathcal{A}(\overline{x}) \supseteq \mathcal{A}(y) \supseteq \mathcal{A}(x^k) \supseteq \mathcal{B}(x^k) \supseteq \mathcal{B}(\overline{x}), \tag{6.20}$$

so that by Theorem 6.3 the vector

$$z = y - \|A\|^{-1}\beta(y)$$

satisfies

$$f(z) < \min\{f(x) : x \in \mathcal{W}_I\} \tag{6.21}$$

with $I = \mathcal{A}(y)$. Since I satisfies by (6.20) $A(\overline{x}) \supseteq I \supseteq \mathcal{B}(\overline{x})$, we have also

$$f(\overline{x}) = \min\{f(x) : x \in \Omega\} = \min\{f(x) : x \in \mathcal{W}_I\}. \tag{6.22}$$

However, $z \in \Omega$, so that (6.22) contradicts (6.21). Thus all x^k are strictly proportional for $k \geq k_0$ so that

$$\mathcal{A}(x^{k_0}) \subseteq \mathcal{A}(x^{k_0+1}) \subseteq \dots.$$

Using the finite termination property of the conjugate gradient method, we conclude that there must be k such that $\overline{x} = x^k$. $\square$

7 Implementation of the algorithm

In this section, we shall describe in more details implementation of Algorithm 3.1 that we use in our numerical experiments. The implementation differs from Algorithm 3.1 in that it exploits the current conjugate direction to generate an intermediate iteration $x^{k+\frac{1}{2}}$ before the expansion step that generates x^k from $x^{k+\frac{1}{2}}$. Such modification does not require any additional matrix-vector multiplication and the estimate (5.1) remains valid as $f(x^{k+\frac{1}{2}}) - f(x^k) \leq 0$ and

$$f(x^{k+1}) - f(\overline{x}) \leq \eta(f(x^{k+\frac{1}{2}}) - f(\overline{x})) = \eta((f(x^{k+\frac{1}{2}}) - f(x^k)) + f(x^k) - f(\overline{x})) \leq \eta(f(x^k) - f(\overline{x})).$$

We also adapted the algorithm so that it can carry out the minimization over more general set $\Omega = \{x : l \leq x\}$ defined by a given n-vector l. If $l \neq 0$ then the reduced gradient is defined by

$$\widetilde{\varphi}_i = \widetilde{\varphi}_i(x) = \min\{(x_i - l_i)/\overline{\alpha}, \varphi_i\}.$$

We describe our algorithm in an easily understandable variant of the Matlab-like language. To preserve readability, we do not distinguish generations of variables by indices unless it is convenient for further reference.

Algorithm 7.1 (MODIFIED PROPORTIONING WITH REDUCED GRADIENT PROJECTIONS (MPRGP)). *Given a symmetric positive definite matrix A of the order n and n-vectors b, l, $\Omega = \{x : l \leq x\}, x^0 \in \Omega,\ \Gamma > 0,\ \overline{\alpha} \in (0, \|A\|^{-1}]$ and $\epsilon > 0$.*

Step 0. {*Initialization.*}
Set $k = 0,\ r = Ax^0 - b,\ p = \varphi(x^0)$
while $\|\nu(x^k)\| > \epsilon$
if $\|\beta(x^k)\|^2 \leq \Gamma\ \widetilde{\varphi}(x^k)^\top \varphi(x^k)$

Step 1. {*Proportional x^k. Trial conjugate gradient step.*}
$\alpha_{cg} = r^\top p / p^\top A p,\ y = x^k - \alpha_{cg} p$

$\alpha_f = \max\{\alpha : x^k - \alpha p \in \Omega\}$.

if $\alpha_{cg} \le \alpha_f$

Step 2. {*Conjugate gradient step.*}

$x^{k+1} = y,\ r = r - \alpha_{cg} Ap,$

$\gamma = \varphi(y)^\top Ap / p^\top Ap,\ p = \varphi(y) - \gamma p$

end *if*

else

Step 3. {*Expansion step.*}

$x^{k+\frac{1}{2}} = x^k - \alpha_f p;\ r = r - \alpha_f Ap;$

$x^{k+1} = P_\Omega(x^{k+\frac{1}{2}} - \overline{\alpha}\varphi(x^{k+\frac{1}{2}})),$

$r = Ax^{k+1} - b,\ p = \varphi(x^{k+1})$

end *else*

end *if*

else

Step 4. { *Proportioning step.*}

$d = \beta(x^k),\ \alpha_{cg} = r^\top d / d^\top Ad,$

$x^{k+1} = x^k - \alpha_{cg} d,\ r = r - \alpha_{cg} Ad,\ p = \varphi(x^{k+1})$

end *else*

$k = k + 1$

end *while*

Step 5. {*Return solution.*}

$x = x^k$.

8 Numerical experiments

We have implemented our algorithm in Matlab and compared its performance with that of two other algorithms on 1D and 2D inner obstacle problems. We used the stopping criterion $\|\nu(x)\| \le 10^{-5}\|b\|$, $\overline{\alpha} = \|A\|_\infty$ and $\Gamma = 1$.

The first problem arises from the discretization of the problem to find the minimum of

$$f(u) = \frac{1}{2}\int_0^{0.5} \|u'(x)\|^2 dx + \int_0^{0.5} u dx$$

subject to $u \in \mathcal{K}$, where

$$\mathcal{K} = \{u \in H^1[0, 0.5]\ :\ l \le u \text{ on } (0, 0.5) \text{ and } u(0) = u'(0.5) = 0\}.$$

The problem was discretized by linear finite elements on a regular grid with 50 nodal variables.

Table 1: Performance of the algorithm on 1D problems

Support	*Algorithm*	n^A	*Active constraints*
l	$MPRGP$	116	23
l	MP	112	23
l	$Polyak$	360	23
c	$MPRGP$	153	19
c	MP	179	19
c	$Polyak$	537	19

We have used two supports in our tests. The circular support c was defined by the lower part of the circle of radius $R = 2.03$ with center $S = (.5, 1.943)$. The line support l was defined by the horizontal line $y = -.04$.

To solve the problem, we used the algorithm $MPRPG$ of the previous section and, for comparison, also the monotone proportioning MP [9] and the Polyak algorithm $Polyak$ [23]. We started our computations always from $x^0 = 0$. The parameters of the circular support are such that if we start the solution with $x^0 = 0$, the first constraints that are activated in the process of solution are later released. We use the line support to examine the performance of the algorithm on problems with expanding active set. The performance of the MPRGP algorithm on problems with shrinking active set is similar to that of any other proportioning algorithm with the same Γ.

The results of computations are in Table 1. The computational cost is measured by the number n^A of the multiplications by the matrix A.

The second problem is the minimization of

$$f(u) = \frac{1}{2}\int_\Omega \|\nabla u(x)\|^2 d\Omega + \int_\Omega u d\Omega$$

subject to $u \in \mathcal{K}$, where $\Omega = [0, 0.5] \times [0, 0.5]$ and

$$\mathcal{K} = \{u \in H^1(\Omega) : l \leq u \text{ on } \Omega,\ u(x, .5) = u_y(x, .5) = u(.5, y) = u_x(.5, y) = 0 \text{ for } (x, y) \in \Omega\}.$$

The problem was discretized using the linear finite elements on a regular grid with 50 nodal variables in each direction. The results for the sphere obstacle s defined by the radius $R = 5$ and center $S = (0, 0, 4.94)$ and for the plane obstacle p defined by $z = -.05$ are in Table 2.

The tables indicate that the performance of the $MPRPG$ algorithm is for the problems considered comparable to the efficient proportioning algorithms. Let us mention that the

Table 2: Performance of the algorithm on 2D problems

Support	*Algorithm*	n^A	*Active constraints*
p	$MPRGP$	260	283
p	MP	283	283
p	*Polyak*	1897	283
s	$MPRGP$	484	245
s	MP	491	245
s	*Polyak*	1207	245

solution of the related unconstrained problems in D1 and D2 required 51 and 147 conjugate gradient iterations, respectively.

9 Comments and conclusions

We presented a new variant of the active set based algorithm for the solution of strictly convex quadratic programming problems with inexact solution of auxiliary linear problems. The unique feature of the new algorithm is balanced treatment of the chopped and free gradients that enables to prove both its rate of convergence and its finite termination property even for problems whose solution does not satisfy the strict complementary condition. The algorithm avoids any backtracking at the cost of using the upper bound on the spectral norm of the Hessian matrix. The experiments show that its performance is comparable with that of some other efficient methods.

The rate of convergence of the new algorithm seems to be a bit pessimistic as it is even worse than that of the gradient projection (4.4). The reason is that it is based on analysis of particular steps and does not take into account better estimates of the rate of convergence for the conjugate gradient method [1, 20] that yield much more efficient performance of the algorithm in consecutive conjugate gradient steps. In spite of that our estimate seems to be a considerable improvement over qualitative convergence results of related algorithms based on sufficient decrease condition that were discussed in the introduction.

Due to the finite termination property of the algorithm, the algorithm switches at certain stage to the conjugate gradient method, so that the nice self-accelerating properties of the latter [26] necessarily apply in the final stage of computation. Let us recall that there are algorithms with even superlinear convergence (e. g. Coleman and Hulbert [5]), but their typical step requires solution of related linear problem which may be comparable with solution of the whole problem by the algorithm presented here. The algorithm presented here is an important ingredient in development of scalable algorithms for numerical solution of elliptic variational inequalities based on the FETI methods [13, 14, 15, 24, 25].

References

[1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge 1994.

[2] M. S. Bazaraa and C. M. Shetty, *Nonlinear Programming*, J. Wiley, New York, 1979.

[3] R.H. Bielschowski, A. Friedlander, F.A.M. Gomes, J.M. Martínez, and M. Raydan, *An adaptive algorithm for bound constrained quadratic minimization*, Investigacion Operativa, 7 (1997), pp. 67-102.

[4] P. H. Calamai and J. J. Moré, *Projected gradient methods for linearly constrained problems*, Math. Programming 39 (1987) pp. 93-116.

[5] T. F. Coleman and L. A. Hulbert, *A globally and superlinearly convergent algorithm for convex quadratic programs with simple bounds*, SIAM J. Optim. 3,2 (1993) 298-321.

[6] R. S. Dembo and U. Tulowitzski, *On minimization of a quadratic function subject to box constraints*, Working Paper No. 71, Series B, School of organization and Management, Yale University, New Haven, CT, 1983.

[7] M. A. Diniz-Ehrhardt, Z. Dostál, M. A. Gomes-Ruggiero, J. M. Martínez, and S. A. Santos, *Nonmonotone strategy for minimization of quadratics with simple constraints*, Appl. Math. 46,5 (2001), pp. 321-338.

[8] M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero, and S. A. Santos, *Numerical analysis of the leaving-face criterion in bound-constrained quadratic minimization*, Optimization Methods and Software 15(2001) pp.45-66.

[9] Z. Dostál, *Box constrained quadratic programming with proportioning and projections*, SIAM J. Optim. 7, 3 (1997), pp. 871-887.

[10] Z. Dostál *Inexact solution of auxiliary problems in Polyak type algorithms*, Acta Univ. Palacki. Olomuc., Fac. rer. nat., Mathematica 38 (1999) 25-30.

[11] Z. Dostál, *A proportioning based algorithm for bound constrained quadratic programming with the rate of convergence*, submitted to Numerical Algorithms.

[12] Z. Dostál, A. Friedlander, and S.A. Santos, *Adaptive precision control in quadratic programming with simple bounds and/or equalities*, High Performance Software for Non-linear Optimization, eds. R. De Leone, A. Murli, P.M.Pardalos and G. Toraldo, Kluwer, Applied Optimization 24 (1998) pp. 161-173.

[13] Z. Dostál, F.A.M.Gomes, S. A. Santos, *Duality based domain decomposition with natural coarse space for variational inequalities*, J. Comput. Appl. Math. 126,1-2(2000) 397-415.

[14] Z. Dostál, F.A.M.Gomes, and S. A. Santos, *Solution of Contact problems by FETI domain decomposition with natural coarse space projection*, Computer Meth. in Appl. Mech. and Engineering 190,13-14 (2000) 1611-1627.

[15] Z. Dostál and D. Horák, *Scalability and FETI based algorithm for large discretized variational inequalities*, submitted to Math. and Comp. in Simulation.

[16] R. Fletcher, *Practical Methods of Optimization*, John Wiley&Sons, Chichester 1997.

[17] A. Friedlander and M. Martínez, *On the maximization of a concave quadratic function with box constraints*, SIAM J. Optimization 4 (1994), pp. 177-192.

[18] A. Friedlander, J.M. Martínez and S.A. Santos, *A new trust region algorithm for bound constrained minimization*, Applied Mathematics & Optimization 30 (1994), pp. 235-266.

[19] A. Friedlander, J.M. Martínez and M. Raydan, *A new method for large scale box constrained quadratic minimization problems*, *Optimization Methods and Software* **5** (1995), pp. 57-74.

[20] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, Philaledelphia 1997.

[21] J. J. Moré and G. Toraldo, *On the solution of large quadratic programming problems with bound constraints*, SIAM J. Optim. 1 (1991), pp. 93-113.

[22] D. P. O'Leary, *A generalised conjugate gradient algorithm for solving a class of quadratic programming problems*, Lin. Alg. Appl. 34 (1980), pp. 371-399.

[23] B. T. Polyak, *The conjugate gradient method in extremal problems*, USSR Comput. Math. and Math. Phys. 9 (1969), pp. 94-112

[24] J. Schöberl, *Solving the Signorini problem on the basis of domain decomposition techniques*, Computing 60,4(1998)323-344.

[25] J. Schöberl, *Efficient contact solvers based on domain decomposition techniques*, Int. J. Computers & Mathematics with Applications 42(2001)1217-1228.

[26] A. van der Sluis amd H. A. van der Vorst *The rate of convergence of the conjugate gradients*, Numer. Math. 48 (1986)543-560.

[27] E. K. Yang and J. W. Tolle, *A class of methods for solving large, convex quadratic programs subject to box constraints*, Mathematical Programming 51(1991)223-228.