Superrelaxation and the rate of convergence in minimizing quadratic functions subject to bound constraints *

Zdeněk Dostál, Marta Domorádová, and Marie Sadowská †

Abstract

The paper resolves the problem concerning the rate of convergence of the working set based MPRGP (modified proportioning with reduced gradient projection) algorithm with a long steplength of the reduced projected gradient step. The main results of this paper are the formula for the R-linear rate of convergence of MPRGP in terms of the spectral condition number of the Hessian matrix and the proof of the finite termination property for the problems whose solution does not satisfy the strict complementarity condition. The bound on the R-linear rate of convergence of the projected gradient is also included. For shorter steplengths these results were proved earlier by Dostál and Schöberl. The efficiency of the longer steplength is illustrated by numerical experiments. The result is an important ingredient in development of scalable algorithms for numerical solution of elliptic variational inequalities and substantiates the choice of parameters that turned out to be effective in numerical experiments.

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: Superrelaxation in minimizing quadratic functions

*This research has been supported by the grants GA CR 201/07/0294, AS CR 1ET400300415, and the Ministry of Education of the Czech Republic No. MSM6198910027.

†FEI VŠB-Technical University Ostrava, Tř 17 listopadu, CZ-70833 Ostrava, Czech Republic, E-mail: zdenek.dostal@vsb.cz, Phone: +420 596 995 227, Fax +420 596 919 597

1 Introduction

We are concerned with the problem to find

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

with $\Omega = \{x : x \geq \ell\}$, $f(x) = \frac{1}{2}x^T Ax - x^T b$, ℓ and b given column n-vectors, and A an $n \times n$ symmetric positive definite matrix. We are 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 applications of the duality based domain decomposition methods to the solution of discretized variational inequalities [18, 23, 42, 43] or in solving auxiliary problems in the augmented Lagrangian type algorithms for minimization of convex quadratic functions subject to more general constraints [16, 17].

The most popular algorithms for the solution of (1.1) are based on the active set strategy or the interior point methods; basic ideas of these methods are described, e.g., in the book by Nocedal and Wright [39]. Closely related to the active set method is the Newton semi-smooth method; see, e.g., Hintermüller, Ito, and Kunisch [33] or Hüeber, Stadler, and Wohlmuth [34]. Here we restrict our attention to the active set based MPRGP (modified proportioning with reduced gradient projection) algorithm proposed by Dostál and Schöberl [26]. The algorithm uses the conjugate gradient method to explore the face of the feasible region defined by the current iterate and invokes the reduced gradient projection with the fixed steplength to expand the active set, combining the minimization in the face of the Polyak algorithm [40] with the gradient projections pioneered by Calamai and Moré [6] and Moré and Toraldo [36]. The precision of approximate solutions of the auxiliary unconstrained problems is controlled by the norm of violation of the Karush–Kuhn–Tucker conditions as in the proportioning based algorithms studied by Friedlander and Martínez with their collaborators and Dostál [28, 30, 29, 4, 10, 11, 8]. For the steplength $\overline{\alpha} \in (0, \|A\|^{-1}]$, Dostál and Schöberl [26] gave the formula for the bound on the R-linear rate of convergence in terms of the spectral condition number of the Hessian matrix A and provided the proof of the finite termination property even for the problems whose solution does not satisfy the strict complementarity condition. Later Dostál [13] found also a bound on the R-linear rate of convergence of the projected gradient. More comprehensive discussion of the development of algorithms for the solution of bound constrained quadratic programming problems can be found, e.g., in Dostál and Schöberl [26] or Hager and Zhang [32]. Important ideas for the analysis of the rate of convergence of the gradient projection method were introduced by Luo and Tseng [35].

The MPRGP algorithm was a key ingredient in the development of scalable algorithms for numerical solution of variational inequalities [23, 25, 21, 24]. Though the algorithm turned out to be effective not only in theory, but also in practice, it was observed [38, 27] that the best performance was obtained for the values of the steplength slightly less then $2\|A\|^{-1}$, near the steplength minimizing the the bound on the Euclidean contraction of the projected gradient method [3] and beyond the range covered by the theory based on the analysis by Schöberl [42, 43, 26]. The point of this paper is to apply a recent result by Dostál [15] to improve the analysis of MPRGP [26, 13] so that it covers application of the steplength $\alpha \in (0, 2\|A\|^{-1}]$.

Our analysis extends also the results of Dostál [13] on the R-linear convergence of the projected gradient. This result is important when MPRGP is used in the inner loop of some other algorithm, such as the inexact augmented Lagrangians proposed by Dostál, Fiedlander, and Santos [17] or SMALBE (semimonotonic augmented Lagrangians for bound and equality constrained problems) algorithm proposed by Dostál [14]. Recall that the R-linear convergence of the projected gradient does not follow from the R-linear convergence of the iterates as the projected gradient is not a continuous function of the iterates.

We extend also the results of Dostál and Schöberl [26] on the finite termination property of our algorithm, including the problems with a dual degenerate solution. We believe that there are at least two reasons why to consider the finite termination results important. First the algorithm with the finite termination property is less likely to suffer from the oscillations often attributed to the active set based algorithms as it removes the indices from the active set only when there is some ground to do it. The second reason is that such algorithm is more likely to generate longer sequences of the conjugate gradient iterations and finally switches to the conjugate gradient method, so that it can better exploit its nice self-acceleration property [44]. It seems very difficult to enhance these characteristics of the algorithm into the rate of convergence.

The paper is organized as follows. After the introduction, we review relevant results and give the proof of the rate of convergence of the iterations and of the projected gradient. Then we extend the results of Dostál and Schöberl on the finite termination property. Finally we give results of numerical experiments that illustrate the performance of the algorithm for large steplength of the reduced gradient projection.

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

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

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

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

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

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

Its complement

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

and subset

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

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* $g^P(x) = \varphi(x) + \beta(x)$ is equal to zero.

The Euclidean norm and the $A-$energy norm of x will be denoted by $\|x\|$ and $\|x\|_A$, respectively. Thus $\|x\|^2 = x^T x$ and $\|x\|_A^2 = x^T Ax$. 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

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

where y^+ denotes for any n-vector y the vector with entries y_i^+ =max $\{y_i, 0\}$.

If $M \in I\!R^{n\times n}$, $v \in I\!R^n$, and $\mathcal{S} \subseteq \{1, \ldots, n\}$, then $A_\mathcal{S}$ and $v_\mathcal{S}$ denote the submatrix of A and the subvector of v with the row indices $i \in \mathcal{S}$.

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. To generate a sequence of iterates $\{x^k\}$ that approximate the solution of (1.1), it exploits a given constant $\Gamma > 0$, a test to decide about leaving the face and three types of steps.

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 $\overline{\alpha} \in (0, 2\|A\|^{-1}]$. This step may expand the current active set. To describe it without P_Ω, let us introduce, for any $x \in \Omega$ and $\alpha > 0$, the *reduced free gradient* $\widetilde{\varphi}_\alpha(x)$ with the entries

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

so that

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

Using the new notation, we can write for any $x \in \Omega$

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

If the inequality

$$||\beta(x^k)||^2 \leq \Gamma^2 \widetilde{\varphi}_{\overline{\alpha}}(x^k)^T \varphi(x^k) \tag{3.4}$$

holds then we call the iterate x^k *strictly proportional.* The test (3.4) is used to decide which component of the projected gradient $g^P(x^k)$ 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 [28, 30, 29, 4, 10, 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, 31] that α_{cg} that minimizes $f(x - \alpha d)$ for given d and x may be evaluated by the formula

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

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, 31] 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, 31]

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

The basic property of the conjugate directions $p^s, \dots, p^k$ that are generated by the recurrence (3.8) from the restart p^s is their mutual A-orthogonality, i.e. $(p^i)^T A p^j = 0$ for $i, j \in \{s, \dots, k\},\ i \neq j$. It follows easily [1, 31] that

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

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

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

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 $x^0 \in \Omega$, $\overline{\alpha} \in (0, 2\|A\|^{-1}]$, and $\Gamma > 0$ be given. For $k \geq 0$ and x^k known, choose x^{k+1} by the following rules:
(i) If $g^P(x^k) = 0$, set $x^{k+1} = x^k$.
(ii) If x^k is strictly proportional and $g^P(x^k) \neq 0$, 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 concerning the implementation of the algorithm (except the bound on the steplength) may be found in Dostál and Schöberl [26].

4 Auxiliary results

Let us assume that $x \in \Omega$ is arbitrary but fixed, so that we can define for each $\alpha \in I\!R$ a quadratic function

$$F_\alpha(y) = \alpha f(y) + \frac{1}{2}(y - x)^T (I - \alpha A)(y - x). \tag{4.1}$$

Then

$$F_\alpha(x) = \alpha f(x), \quad \nabla F_\alpha(x) = \alpha \nabla f(x), \quad \nabla^2 F_\alpha(x) = I, \tag{4.2}$$

and for $\delta \leq \|A\|^{-1}$

$$\delta f(y) \leq F_\delta(y) \quad \text{for any} \quad y \in I\!R^n. \tag{4.3}$$

Moreover, if $0 \leq \delta \leq \|A\| = 1$, then

$$\delta F_1(y) \leq \delta F_1(y) + \frac{1 - \delta}{2}\|y - x\|^2 = F_\delta(y). \tag{4.4}$$

It is easy to check that F_α is separable, i.e. there is a vector $c \in I\!R^n$ and $d_i \in I\!R, i-1,\ldots,n$, such that

$$F_\alpha(y) = \sum_{i=1}^{n} F_{\alpha i}(y_i), \quad F_{\alpha i}(y_i) = \frac{1}{2}y_i^2 - c_i y_i + d_i, \quad y = [y_i], \tag{4.5}$$

We shall use some other relations from [26].

Lemma 4.1 *Let $\widehat{x}$ denote a unique solution of (1.1), let $\lambda_{\min}$ denote the smallest eigenvalue of A, $\eta_f = 1 - \delta\lambda_{\min}$, $\delta \in (0, \|A\|^{-1}]$, $x \in \Omega$, and $g = Ax - b$. Let Then*

$$f(P_\Omega\,(x - \delta g)) - f(\widehat{x}) \le \eta_f(f(x) - f(\widehat{x})) \tag{4.6}$$

and

$$F_\delta(P_\Omega(x - (2-\delta)\varphi(x))) \le F_\delta(P_\Omega(x - \delta\varphi(x))). \tag{4.7}$$

Proof: The inequality (4.6) has been proved by Schöberl; see Theorem 4.1 of [26]. The inequality (4.7) for $n = 1$ can be proved by a straightforward analysis of all possibilities; see [15] for the details.

To prove the general case $n \ge 1$, first recall that P_Ω is separable and can be defined componentwise by

$$P_i(y) = \max\{y, \ell_i\},\ i = 1,\ldots,n,\ y \in I\!R.$$

Denoting $\mathcal{F}$, $\mathcal{A}$, and g_i the free set of x, the active set of x, and the components of the gradient $g(x)$, respectively, we can use the separable representation (4.5) of F_δ to reduce (4.7) to the case $n = 1$. Thus we have

$$\begin{aligned} F_\delta(P_\Omega(x - (2-\delta)\varphi(x))) &= \sum_{i=1}^{n} F_{\delta i}([P_\Omega(x - (2-\delta)\varphi(x))]_i) \\ &= \sum_{i\in\mathcal{F}} F_{\delta i}(P_i(x_i - (2-\delta)g_i)) + \sum_{i\in\mathcal{A}} F_{\delta i}(P_i(x_i)) \\ &\le \sum_{i\in\mathcal{F}} F_{\delta i}(P_i(x_i - \delta g_i)) + \sum_{i\in\mathcal{A}} F_{\delta i}(P_i(x_i)) \\ &= F_\delta(P_\Omega(x - \delta\varphi(x))). \end{aligned}$$

This proves (4.7). $\square$

5 Rate of convergence

Now we are ready to prove the R-linear rate of convergence of MPRGP in terms of bounds on the spectrum of the Hessian A for $\overline{\alpha} \in (0, 2\|A\|^{-1})$.

Theorem 5.1. *Let $\{x^k\}$ denote the sequence generated by Algorithm 3.1 with $\overline{\alpha} \in (0, 2\|A\|^{-1}]$ and $\Gamma > 0$. Then for any $k \geq 0$*

$$f(x^{k+1}) - f(\widehat{x}) \leq \eta_\Gamma \left(f(x^k) - f(\widehat{x}) \right), \tag{5.1}$$

where $\widehat{x}$ denotes a unique solution of (1.1),

$$\eta_\Gamma = 1 - \frac{\widehat{\alpha}\lambda_{\min}}{\vartheta + \vartheta\widehat{\Gamma}^2}, \qquad \widehat{\Gamma} = \max\{\Gamma, \Gamma^{-1}\}, \tag{5.2}$$

$$\vartheta = 2\max\{\overline{\alpha}\|A\|, 1\}, \quad \widehat{\alpha} = \min\{\overline{\alpha}, 2\|A\|^{-1} - \overline{\alpha}\}, \tag{5.3}$$

and $\lambda_{\min}$ denotes the smallest eigenvalue of A. The error in the A-norm is bounded by

$$\|x^k - \widehat{x}\|_A^2 \leq 2\eta_\Gamma^k \left(f(x^0) - f(\widehat{x}) \right). \tag{5.4}$$

Proof: Our main tools are the auxiliary results of Sect. 4 and the inequality

$$f\Big(P_\Omega\left(x^k - \alpha g(x^k)\right)\Big) \geq f(x^k) - \alpha(\widetilde{\varphi}_\alpha(x^k)^T\varphi(x^k) + \|\beta(x^k)\|^2), \tag{5.5}$$

which can be obtained for any $\alpha \geq 0$ by the Taylor expansion and (3.3).

Let us first assume that $\|A\| = 1$ and let x^{k+1} be generated by the expansion step (3.1). Using in sequence the definition of the dominating function associated with $x = x^k$ which satisfies (4.3), Lemma 4.1, the assumptions $\|A\| = 1$ and $\widehat{\alpha} \leq 1$ with (4.4), the Taylor expansion with (3.2) and (4.2), $\|\widetilde{\varphi}_{\widehat{\alpha}}(x^k)\|^2 \leq \widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T\varphi(x^k)$, and simple manipulations, we get

$$\begin{aligned}
\widehat{\alpha} f(x^{k+1}) &\leq \widehat{\alpha} F_1(x^{k+1}) = \widehat{\alpha} F_1\Big(P_\Omega\left(x^k - \overline{\alpha}\varphi(x^k)\right)\Big) \\
&\leq \widehat{\alpha} F_1\Big(P_\Omega\left(x^k - \widehat{\alpha}\varphi(x^k)\right)\Big) \leq F_{\widehat{\alpha}}\Big(P_\Omega\left(x^k - \widehat{\alpha}\varphi(x^k)\right)\Big) \\
&= F_{\widehat{\alpha}}\left(x^k\right) - \widehat{\alpha}^2\widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T\varphi(x^k) + \frac{\widehat{\alpha}^2}{2}\|\widetilde{\varphi}_{\widehat{\alpha}}(x^k)\|^2 \\
&\leq F_{\widehat{\alpha}}\left(x^k\right) - \frac{\widehat{\alpha}^2}{2}\widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T\varphi(x^k) = \widehat{\alpha} f(x^k) - \frac{\widehat{\alpha}^2}{2}\widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T\varphi(x^k).
\end{aligned}$$

Thus

$$f(x^{k+1}) \leq f(x^k) - \frac{\widehat{\alpha}}{2}\widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T\varphi(x^k). \tag{5.6}$$

The expansion step is used only when x^k is strictly proportional, i.e.

$$\|\beta(x^k)\|^2 \leq \Gamma^2\widetilde{\varphi}_{\overline{\alpha}}(x^k)^T\varphi(x^k).$$

Since $\widehat{\alpha} \le \overline{\alpha}$ by the definition, it follows that

$$\widetilde{\varphi}_{\overline{\alpha}}(x^k)^T \varphi(x^k) \le \widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T \varphi(x^k)$$

and

$$\|\beta(x^k)\|^2 \le \Gamma^2 \widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T \varphi(x^k). \tag{5.7}$$

After substituting (5.7) into (5.5) with $\alpha = \widehat{\alpha}$, we get

$$f(P_\Omega\left(x^k - \widehat{\alpha} g(x^k)\right)) \ge f(x^k) - \widehat{\alpha}(1+\Gamma^2)\widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T \varphi(x^k). \tag{5.8}$$

Thus for x^{k+1} generated by the expansion step, we get by elementary algebra and application of (5.6) that

$$\begin{aligned} f(x^{k+1}) &\le f(x^k) - \frac{\widehat{\alpha}}{2} \widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T \varphi(x^k) \\ &= \frac{1}{2+2\Gamma^2}\Big(f(x^k) - \widehat{\alpha}(1+\Gamma^2)\widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T\varphi(x^k) + (1+2\Gamma^2)f(x^k)\Big) \\ &\le \frac{1}{2+2\Gamma^2}\Big(f\left(P_\Omega\left(x^k - \widehat{\alpha} g(x^k)\right)\right) + (1+2\Gamma^2)f(x^k)\Big). \end{aligned}$$

Inserting $-f(\widehat{x}) + f(\widehat{x})$ into the last term and substituting (4.6) with $x = x^k$ and $\overline{\alpha} = \widehat{\alpha}$ into the resulting expression, we get

$$\begin{aligned} f(x^{k+1}) &\le \frac{\eta_f + 1 + 2\Gamma^2}{2+2\Gamma^2} f(x^k) + \frac{1-\eta_f}{2+2\Gamma^2} f(\widehat{x}) \\ &= \frac{\eta_f + 1 + 2\Gamma^2}{2+2\Gamma^2}\Big(f(x^k) - f(\widehat{x})\Big) + f(\widehat{x}). \end{aligned} \tag{5.9}$$

The proof of (5.1) for $\|A\| = 1$ is completed by

$$\frac{\eta_f + 1 + 2\Gamma^2}{2+2\Gamma^2} = \frac{\eta_f - 1 + 2 + 2\Gamma^2}{2+2\Gamma^2} = 1 - \frac{1-\eta_f}{2+2\Gamma^2} = 1 - \frac{\widehat{\alpha}\lambda_{\min}}{2+2\Gamma^2} \le \eta_\Gamma.$$

To prove the general case, it is enough to apply the theorem to $h = \|A\|^{-1} f$.

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

$$\begin{aligned} f(x^{k+1}) &\le f\left(x^k - \alpha_{cg}\varphi(x^k)\right) = f(x^k) - \frac{1}{2}\frac{\|\varphi(x^k)\|^4}{\varphi(x^k)^T A \varphi(x^k)} \\ &\le f(x^k) - \frac{1}{2}\|A\|^{-1}\|\varphi(x^k)\|^2. \end{aligned}$$

Taking into account $\widehat{\alpha} \le \|A\|^{-1}$ and $\widetilde{\varphi}_i \varphi_i \le \varphi_i^2$, $i = 1, \ldots, n$, we get

$$f(x^{k+1}) \le f(x^k) - \frac{1}{2}\|A\|^{-1}\|\varphi(x^k)\|^2 \le f(x^k) - \frac{\widehat{\alpha}}{2}\widetilde{\varphi}_{\widehat{\alpha}}(x^k)^T \varphi(x^k). \qquad (5.10)$$

Since the conjugate gradient step is carried out only if x^k is proportional, we can use the same reasoning as above to prove the same estimate for the conjugate gradient as we have proved for the expansion step.

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

$$\|\beta(x^k)\|^2 > \Gamma^2 \widetilde{\varphi}_{\overline{\alpha}}(x^k)^T \varphi(x^k) \qquad (5.11)$$

and

$$\begin{aligned} f(x^{k+1}) &= f\left(x^k - \alpha_{cg}\beta(x^k)\right) = f(x^k) - \frac{1}{2}\frac{\|\beta(x^k)\|^4}{\beta(x^k)^T A \beta(x^k)} \\ &\le f(x^k) - \frac{1}{2}\|A\|^{-1}\|\beta(x^k)\|^2. \end{aligned}$$

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

$$\overline{\alpha}/\vartheta \le \|A\|^{-1}/2$$

and

$$f(x^{k+1}) \le f(x^k) - \frac{\overline{\alpha}}{\vartheta}\|\beta(x^k)\|^2, \qquad (5.12)$$

where the right-hand side may be rewritten in the form

$$\begin{aligned} f(x^k) - \frac{\overline{\alpha}}{\vartheta}\|\beta(x^k)\|^2 &= \frac{1}{\vartheta(1+\Gamma^{-2})}\left(f(x^k) - \overline{\alpha}(1+\Gamma^{-2})\|\beta(x^k)\|^2\right) \\ &+ \frac{\vartheta + \vartheta\Gamma^{-2} - 1}{\vartheta(1+\Gamma^{-2})} f(x^k). \end{aligned} \qquad (5.13)$$

We can also substitute (5.11) into (5.5) to get

$$f\left(P_\Omega\left(x^k - \overline{\alpha} g(x^k)\right)\right) > f(x^k) - \overline{\alpha}(1+\Gamma^{-2})\|\beta(x^k)\|^2. \qquad (5.14)$$

After substituting (5.14) into (5.13), using (5.12), (4.6) with $x = x^k$, and simple manipulations, we get

$$\begin{aligned} f(x^{k+1}) &< \frac{1}{\vartheta + \vartheta\Gamma^{-2}} f\Big(P_\Omega(x^k - \overline{\alpha} g(x^k))\Big) + \frac{\vartheta + \vartheta\Gamma^{-2} - 1}{\vartheta + \vartheta\Gamma^{-2}} f(x^k) \\ &= \frac{1}{\vartheta + \vartheta\Gamma^{-2}}\Big(f\Big(P_\Omega\left(x^k - \overline{\alpha} g(x^k)\right)\Big) - f(\widehat{x})\Big) \end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{1}{\vartheta+\vartheta\Gamma^{-2}} f(\widehat{x}) + \frac{\vartheta+\vartheta\Gamma^{-2}-1}{\vartheta+\vartheta\Gamma^{-2}} f(x^k) \\
&\leq \quad \frac{\eta_f}{\vartheta+\vartheta\Gamma^{-2}}\Big(f(x^k)-f(\widehat{x})\Big) + \frac{1}{\vartheta+\vartheta\Gamma^{-2}} f(\widehat{x}) + \frac{\vartheta+\vartheta\Gamma^{-2}-1}{\vartheta+\vartheta\Gamma^{-2}} f(x^k) \\
&= \quad \frac{\eta_f+\vartheta+\vartheta\Gamma^{-2}-1}{\vartheta+\vartheta\Gamma^{-2}}\Big(f(x^k)-f(\widehat{x})\Big) + f(\widehat{x}).
\end{aligned}
$$

Comparing the last inequality with (5.9) and taking into account that by the definition $\Gamma \leq \widehat{\Gamma}$, $\Gamma^{-1} \leq \widehat{\Gamma}$, and $\vartheta \geq 2$, we obtain that the estimate

$$
f(x^{k+1}) - f(\widehat{x}) \leq \frac{\eta_f+\vartheta+\vartheta\Gamma^{-2}-1}{\vartheta+\vartheta\Gamma^{-2}}\Big(f(x^k)-f(\widehat{x})\Big)
$$

is valid for the all three steps. The proof of (5.1) is completed by

$$
\eta_\Gamma \quad = \quad \frac{\eta_f+\vartheta+\vartheta\Gamma^{-2}-1}{\vartheta+\vartheta\Gamma^{-2}} = 1 - \frac{1-\eta_f}{\vartheta+\vartheta\Gamma^{-2}} = 1 - \frac{\widehat{\alpha}\lambda_{\min}}{\vartheta+\vartheta\widehat{\Gamma}^2}. \qquad \square
$$

An inspection of the proof reveals that the factor ϑ appears only in the analysis of the proportioning step, which depends on $\overline{\alpha}$ only via the proportioning test.

6 Rate of convergence of projected gradient

To use the MPRGP algorithm in the inner loops of other algorithms, it is important to *recognize* when we are near the solution. There is a catch – the projected gradient is not a continuous function of the iterates! The R-linear convergence of the projected gradient is treated by the following theorem.

Theorem 6.1 *Let $\{x^k\}$ be generated by Algorithm 3.1 with $x^0 \in \Omega$, $\Gamma > 0$, and $\overline{\alpha} \in (0, 2\|A\|^{-1}]$. Let $\widehat{\alpha}$, $\widehat{\Gamma}$, ϑ, and η_Γ be those of Theorem 5.1. Let $\widehat{x}$ denote the unique solution of* (1.1).

Then for any $k \geq 0$

$$
\|g^P(x^{k+1})\|^2 \leq a_1 \eta_\Gamma^k \left(f(x^0) - f(\widehat{x})\right) \tag{6.1}
$$

and

$$
a_1 = \frac{38}{\widehat{\alpha}(1-\eta_\Gamma)} = \frac{38\vartheta(1+\widehat{\Gamma}^2)}{\widehat{\alpha}^2\lambda_{\min}}. \tag{6.2}
$$

Proof: First notice that it is enough to estimate separately $\beta(x^k)$ and $\varphi(x^k)$ as

$$
\|g^P(x^k)\|^2 = \|\beta(x^k)\|^2 + \|\varphi(x^k)\|^2.
$$

In particular, since $\widehat{\alpha} \leq \|A\|^{-1}$, we have for any vector d such that $d^T g(x) \geq \|d\|^2$

$$f(x) - f(x - \widehat{\alpha} d) = \widehat{\alpha} d^T g(x) - \frac{1}{2}\widehat{\alpha}^2 d^T A d \geq \frac{\widehat{\alpha}}{2}\|d\|^2. \tag{6.3}$$

It follows that we can combine (6.3) with $x^k - \widehat{\alpha}\beta(x^k) \geq \ell$ to estimate $\|\beta(x^k)\|$ by

$$\begin{aligned} f(x^k) - f(\widehat{x}) &= (f(x^k) - f(x^k - \widehat{\alpha}\beta(x^k)) + \left(f\left(x^k - \widehat{\alpha}\beta(x^k)\right) - f(\widehat{x})\right) \\ &\geq f(x^k) - f(x^k - \widehat{\alpha}\beta(x^k)) \geq \frac{\widehat{\alpha}}{2}\|\beta(x^k)\|^2. \end{aligned} \tag{6.4}$$

Applying (5.1), we get

$$\|\beta(x^k)\|^2 \leq \frac{2}{\widehat{\alpha}}\left(f(x^k) - f(\widehat{x})\right) \leq \frac{2\eta_\Gamma^k}{\widehat{\alpha}}(f(x^0) - f(\widehat{x})). \tag{6.5}$$

To estimate $\|\varphi(x^k)\|$, notice that the algorithm "does not know" about the components of the constraint vector ℓ when it generates x^{k+1} unless their indices belong to $\mathcal{A}(x^k)$ or $\mathcal{A}(x^{k+1})$. It follows that x^{k+1} may be considered also as an iterate generated by Algorithm 3.1 from x^k for the problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x_i \geq \ell_i \text{ for } i \in \mathcal{A}(x^k) \cup \mathcal{A}(x^{k+1}). \tag{6.6}$$

If we denote

$$\overline{f}^k = \min\{f(x) : \ x_i \geq \ell_i \text{ for } i \in \mathcal{A}(x^k) \cup \mathcal{A}(x^{k+1})\} \leq f(\widehat{x})$$

and $\overline{\delta}_k = f(\widehat{x}) - \overline{f}^k \geq 0$, we can use (5.1) to get

$$\begin{aligned} \overline{\delta}_k &= f(\widehat{x}) - \overline{f}^k \leq f(x^{k+1}) - \overline{f}^k \leq \eta_\Gamma\left(f(x^k) - \overline{f}^k\right) \\ &= \eta_\Gamma\left(f(x^k) - f(\widehat{x})\right) + \eta_\Gamma\overline{\delta}_k, \end{aligned}$$

so that

$$\overline{\delta}_k \leq \frac{\eta_\Gamma}{1-\eta_\Gamma}\left(f(x^k) - f(\widehat{x})\right) \leq \frac{\eta_\Gamma^{k+1}}{1-\eta_\Gamma}\left(f(x^0) - f(\widehat{x})\right). \tag{6.7}$$

Now observe that the indices of the unconstrained components of the minimization problem (6.6) are those belonging to $\mathcal{I}^k = \mathcal{F}(x^k) \cap \mathcal{F}(x^{k+1})$ as

$$\begin{aligned} \mathcal{I}^k = \mathcal{F}(x^k) \cap \mathcal{F}(x^{k+1}) &= \left(\mathcal{N} \setminus \mathcal{A}(x^k)\right) \cap \left(\mathcal{N} \setminus \mathcal{A}(x^{k+1})\right) \\ &= \mathcal{N} \setminus \left(\mathcal{A}(x^k) \cup \mathcal{A}(x^{k+1})\right). \end{aligned}$$

It follows that if $\mathcal{I}^k$ is nonempty, then by the definition of $\overline{\delta}_k$ and (6.3)

$$\overline{\delta}_k \geq f(\widehat{x}) - f(\widehat{x} - \widehat{\alpha} g_{\mathcal{I}^k}(\widehat{x})) \geq \frac{\widehat{\alpha}}{2} \|g_{\mathcal{I}^k}(\widehat{x})\|^2. \tag{6.8}$$

For convenience, let us define $g_{\mathcal{I}}(x) = o$ for any $x \in I\!R^n$ and empty set $\mathcal{I} = \emptyset$. Then (6.8) remains valid for $\mathcal{I}^k = \emptyset$, so that we can combine it with (6.7) to get

$$\|g_{\mathcal{I}^k}(\widehat{x})\|^2 \leq \frac{2}{\widehat{\alpha}} \overline{\delta}_k \leq \frac{2\eta_\Gamma^{k+1}}{\widehat{\alpha}(1-\eta_\Gamma)} (f(x^0) - f(\widehat{x})). \tag{6.9}$$

Since our algorithm is defined so that either $\mathcal{I}^k = \mathcal{F}(x^k) \subseteq \mathcal{F}(x^{k+1})$ or $\mathcal{I}^k = \mathcal{F}(x^{k+1}) \subseteq \mathcal{F}(x^k)$, it follows that either

$$\begin{aligned} \|g_{\mathcal{F}(x^k)}(\widehat{x})\|^2 = \|g_{\mathcal{I}^k}(\widehat{x})\|^2 \quad &\leq \quad \frac{2\eta_\Gamma^{k+1}}{\widehat{\alpha}(1-\eta_\Gamma)} (f(x^0) - f(\widehat{x})) \\ &\leq \quad \frac{2\eta_\Gamma^{k}}{\widehat{\alpha}(1-\eta_\Gamma)} (f(x^0) - f(\widehat{x})) \end{aligned} \tag{6.10}$$

or

$$\|g_{\mathcal{F}(x^{k+1})}(\widehat{x})\|^2 = \|g_{\mathcal{I}^k}(\widehat{x})\|^2 \leq \frac{2\eta_\Gamma^{k+1}}{\widehat{\alpha}(1-\eta_\Gamma)} (f(x^0) - f(\widehat{x})).$$

Using the same reasoning for x^{k-1} and x^k, we conclude that the estimate (6.10) is valid for any x^k such that

$$\mathcal{F}(x^{k-1}) \supseteq \mathcal{F}(x^k) \quad \text{or} \quad \mathcal{F}(x^k) \subseteq \mathcal{F}(x^{k+1}). \tag{6.11}$$

Let us now recall that using simple manipulations and (5.1), we get

$$\|g(x^k) - g(\widehat{x})\|^2 = \|A(x^k - \widehat{x})\|^2 \leq \|A\| \|x^k - \widehat{x}\|_A^2 \leq \frac{2}{\widehat{\alpha}} \eta_\Gamma^k \left(f(x^0) - f(\widehat{x})\right). \tag{6.12}$$

Thus for any k satisfying the relations (6.11), we get

$$\begin{aligned} \|\varphi(x^k)\| \quad &= \quad \|g_{\mathcal{F}(x^k)}(x^k)\| \leq \|g_{\mathcal{F}(x^k)}(x^k) - g_{\mathcal{F}(x^k)}(\widehat{x})\| + \|g_{\mathcal{F}(x^k)}(\widehat{x})\| \\ &\leq \quad \sqrt{\frac{2}{\widehat{\alpha}} \eta_\Gamma^k (f(x^0) - f(\widehat{x}))} + \sqrt{\frac{2}{\widehat{\alpha}(1-\eta_\Gamma)} \eta_\Gamma^k (f(x^0) - f(\widehat{x}))} \\ &\leq \quad 2\sqrt{\frac{2}{\widehat{\alpha}(1-\eta_\Gamma)} \eta_\Gamma^k (f(x^0) - f(\widehat{x}))}. \end{aligned}$$

Combining the last inequality with (6.5), we get for any k satisfying the relations (6.11) that

$$\|g^P(x^k)\|^2 = \|\beta(x^k)\|^2 + \|\varphi(x^k)\|^2 \leq \frac{10}{\widehat{\alpha}(1-\eta_\Gamma)} \eta_\Gamma^k (f(x^0) - f(\widehat{x})). \tag{6.13}$$

Now notice that the estimate (6.13) is valid for any iterate x^k which satisfies $\mathcal{F}(x^{k-1}) \supseteq \mathcal{F}(x^k)$, i.e. when x^k is generated by the conjugate gradient step or the expansion step. Thus it remains to estimate the projected gradient of the iterate x^k generated by the proportioning step. In this case $\mathcal{F}(x^{k-1}) \subseteq \mathcal{F}(x^k)$, so that we can use the estimate (6.13) to get

$$\|g^P(x^{k-1})\| \leq \sqrt{\frac{10}{\widehat{\alpha}(1-\eta_\Gamma)}\eta_\Gamma^{k-1}(f(x^0) - f(\widehat{x}))}. \tag{6.14}$$

Since the proportioning step is defined by $x^k = x^{k-1} - \alpha_{cg}\beta(x^{k-1})$, it follows that

$$\|g_{\mathcal{F}(x^k)}(x^{k-1})\| = \|g^P(x^{k-1})\|.$$

Moreover, using the basic properties of the norm, we get

$$\begin{aligned}\|\varphi(x^k)\| &= \|g_{\mathcal{F}(x^k)}(x^k)\| \leq \|g_{\mathcal{F}(x^k)}(x^k) - g_{\mathcal{F}(x^k)}(x^{k-1})\| + \|g_{\mathcal{F}(x^k)}(x^{k-1})\| \\ &\leq \|g(x^k) - g(\widehat{x})\| + \|g(\widehat{x}) - g(x^{k-1})\| + \|g^P(x^{k-1})\|,\end{aligned}$$

and by (6.12) and (6.14)

$$\begin{aligned}\|\varphi(x^k)\| &\leq \sqrt{\frac{2}{\widehat{\alpha}}\eta_\Gamma^k(f(x^0) - f(\widehat{x}))} + \sqrt{\frac{2}{\widehat{\alpha}}\eta_\Gamma^{k-1}(f(x^0) - f(\widehat{x}))} \\ &+ \sqrt{\frac{10}{\widehat{\alpha}(1-\eta_\Gamma)}\eta_\Gamma^{k-1}(f(x^0) - f(\widehat{x}))} \\ &\leq (\sqrt{5}+2)\sqrt{\frac{2}{\widehat{\alpha}(1-\eta_\Gamma)}\eta_\Gamma^{k-1}(f(x^0) - f(\widehat{x}))}.\end{aligned}$$

Combining the last inequality with (6.5), we get by simple computation that

$$\|g^P(x^k)\|^2 = \|\varphi(x^k)\|^2 + \|\beta(x^k)\|^2 \leq \frac{38}{\widehat{\alpha}(1-\eta_\Gamma)}\eta_\Gamma^{k-1}\left(f(x^0) - f(\widehat{x})\right).$$

Since the last estimate is obviously weaker than (6.13), it follows that (6.1) is valid for all indices k. $\square$

7 Finite termination

We shall start our exposition with the following identification lemma.

Lemma 7.1. *Let $\{x^k\}$ denote the sequence generated by Algorithm 3.1 with a given $\Gamma > 0$ and $\overline{\alpha} \in (0, 2\|A\|^{-1})$. Then there is k_0 such that for $k \geq k_0$*

$$\mathcal{F}(\widehat{x}) \subseteq \mathcal{F}(x^k), \quad \mathcal{F}(\widehat{x}) \subseteq \mathcal{F}(x^k - \overline{\alpha}\widetilde{\varphi}(x^k)), \quad \textit{and} \quad \mathcal{B}(\widehat{x}) \subseteq \mathcal{B}(x^k). \tag{7.1}$$

Proof: Since (7.1) is trivially satisfied when there is $k = k_0$ such that $x^k = \widehat{x}$, we shall assume in what follows that $x^k \neq \widehat{x}$ for any $k \geq 0$.

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

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

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

$$g_i(x^k) \leq \frac{\epsilon}{4\overline{\alpha}} \text{ for i} \in \mathcal{F}(\widehat{\mathrm{x}}) \tag{7.2}$$

$$x_i^k \geq \ell_i + \frac{\epsilon}{2} \text{ for i} \in \mathcal{F}(\widehat{\mathrm{x}}) \tag{7.3}$$

$$x_i^k \leq \ell_i + \frac{\overline{\alpha}\delta}{8} \text{ for i} \in \mathcal{B}(\widehat{\mathrm{x}}) \tag{7.4}$$

$$g_i(x^k) \geq \frac{\delta}{2} \text{ for i} \in \mathcal{B}(\widehat{\mathrm{x}}). \tag{7.5}$$

In particular, for $k \geq k_0$, the first inclusion of (7.1) follows from (7.3), while the second inclusion follows from (7.2) and (7.3) as for $i \in \mathcal{F}(\widehat{x})$

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

Let $k \geq k_0$ and observe that, by (7.4) and (7.5), for any $i \in \mathcal{B}(\widehat{x})$

$$x_i^k - \overline{\alpha}g_i(x^k) \leq \ell_i + \frac{\overline{\alpha}\delta}{8} - \frac{\overline{\alpha}\delta}{2} < \ell_i,$$

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

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

and $\mathcal{B}(x^{k+1}) \supseteq \mathcal{B}(\widehat{x})$. Moreover, using (7.5) and the definition of Algorithm 3.1, we can directly verify that if $\mathcal{B}(x^k) \supseteq \mathcal{B}(\widehat{x})$, then also $\mathcal{B}(x^{k+1}) \supseteq \mathcal{B}(\widehat{x})$.

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

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

so that for $i \in \mathcal{F}(x^k) \bigcap \mathcal{B}(\widehat{x})$, by (7.4) and (7.5),

$$x_i^k - \alpha_{cg}p_i^k = x_i^k - \alpha_{cg}g_i(x^k) \leq x_i^k - \frac{\overline{\alpha}}{2}g_i(x^k) \leq \ell_i + \frac{\overline{\alpha}\delta}{8} - \frac{\overline{\alpha}\delta}{4} < \ell_i. \tag{7.6}$$

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 $x^{k_0+l} = \widehat{x}$ and by the definition of the step (i) of Algorithm 3.1 $\mathcal{B}(x^k) = \mathcal{B}(\widehat{x})$ 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 $\mathcal{F}(\widehat{x}) = \emptyset$ and $\mathcal{B}(\widehat{x}) = \emptyset$ may be easily proved by the specialization of the above arguments. □

Corollary 7.2. *Let $\{x^k\}$ denote the sequence generated by Algorithm 3.1, and let the solution $\widehat{x}$ satisfies the condition of strict complementarity, i.e. $\widehat{x}_i = \ell_i$ implies $g_i(\widehat{x}) \neq 0$. Then there is $k \geq 0$ such that $x^k = \widehat{x}$.*
Proof: If $\widehat{x}$ satisfies the condition of strict complementarity, then $\mathcal{A}(\widehat{x}) = \mathcal{B}(\widehat{x})$, and by assumptions and Lemma 7.1, there is $k_0 \geq 0$ such that $\mathcal{F}(x^k) = F(\widehat{x})$ and $\mathcal{B}(x^k) = \mathcal{B}(\widehat{x})$. Thus all $x^k, k \geq k_0$ that satisfy $\widehat{x} \neq x^{k-1}$ 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 $x^k = \widehat{x}$. □

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

$$\Gamma\|\varphi(x)\| < \|\beta(x)\|. \tag{7.7}$$

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

$$f(y) < \min\{f(z) : z \in \mathcal{W}_{\mathcal{I}}\} \tag{7.8}$$

where $\mathcal{W}_{\mathcal{I}}$ is defined in (3.10).
Proof: See Dostál [11]. □

Lemma 7.4. *Let $\overline{\alpha} \in (0, 2\|A\|^{-1}]$, $x \in \Omega$, and $y = x - \overline{\alpha}\widetilde{\varphi}(x)$. Then*

$$\|\varphi(y)\|^2 \leq 9\widetilde{\varphi}(x)^T\varphi(x) \quad \textit{and} \quad \|\beta(y)\| \geq \|\beta(x)\| - 4\|\widetilde{\varphi}(x)\|. \tag{7.9}$$

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

$$g(y) = g(x) - \overline{\alpha} A\widetilde{\varphi}(x) \quad \text{and} \quad \widetilde{\varphi}_{\mathcal{F}}(x) = \varphi_{\mathcal{F}}(x) = g_{\mathcal{F}}(x), \tag{7.10}$$

we get

$$\|\varphi(y)\| = \|g_{\mathcal{F}}(y)\| = \|g_{\mathcal{F}}(x) - \overline{\alpha} A_{\mathcal{F}}\widetilde{\varphi}(x)\| \leq \|\widetilde{\varphi}_{\mathcal{F}}(x)\| + \overline{\alpha}\|A_{\mathcal{F}}\widetilde{\varphi}(x)\| \leq 3\|\widetilde{\varphi}(x)\|. \tag{7.11}$$

Using (7.11) and the definition of $\widetilde{\varphi}(x)$, we get

$$\|\varphi(y)\|^2 \leq 9\|\widetilde{\varphi}(x)\|^2 \leq 9\widetilde{\varphi}(x)^T\varphi(x). \tag{7.12}$$

To prove the second inequality of (7.9), denote $\mathcal{B} = \{i \in \mathcal{A}(x) : g_i(x) \leq 0\}$ and notice that

$$\mathcal{A}(y) \supseteq \mathcal{A}(x) \supseteq \mathcal{B}, \tag{7.13}$$

so that

$$\begin{aligned} \|\beta(y)\| = \|g_{\mathcal{A}(y)}(y)^-\| &\geq \|g_{\mathcal{B}}(y)^-\| = \|(g_{\mathcal{B}}(x) - \overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x))^-\| \\ &= \|(\beta_{\mathcal{B}}(x) - \overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x))^-\|. \end{aligned} \tag{7.14}$$

Using in sequence $\|\beta_{\mathcal{B}}(x)\| = \|\beta(x)\|$, $\|\overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x)\| \leq 2\|\widetilde{\varphi}(x)\|$, (7.14), properties of the norm, $\beta^-(x) = \beta(x)$, and $\|z - z^-\| \leq \|z - t\|$ for any t with non-positive entries, we get

$$\begin{aligned} \|\beta(x)\| &- \|\widetilde{\varphi}(x)\| - \|\beta(y)\| \leq \|\beta_{\mathcal{B}}(x)\| - \frac{1}{2}\|\overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x)\| - \|(\beta_{\mathcal{B}}(x) - \overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x))^-)\| \\ &\leq \|(\beta_{\mathcal{B}}(x) - \frac{1}{2}\overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x))\| - \|(\beta_{\mathcal{B}}(x) - \overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x))^-\| \\ &\leq \|(\beta_{\mathcal{B}}(x) - \overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x)) - (\beta_{\mathcal{B}}(x) - \overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x))^-\| + \frac{1}{2}\overline{\alpha}\|A_{\mathcal{B}}\widetilde{\varphi}(x))\| \\ &\leq \|\beta_{\mathcal{B}}(x) - \overline{\alpha} A_{\mathcal{B}}\widetilde{\varphi}(x) - \beta_{\mathcal{B}}(x)\| + \|\widetilde{\varphi}(x)\| \leq 3\|\widetilde{\varphi}(x)\|. \end{aligned}$$

This proves the second inequality of (7.9). □

Corollary 7.5. *Let* $\Gamma \geq 4$, $\overline{\alpha} \in (0, 2\|A\|^{-1}]$, $x \in \Omega$ *and*

$$\Gamma^2\widetilde{\varphi}(x)^T\varphi(x) < \|\beta(x)\|^2. \tag{7.15}$$

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

$$\frac{\Gamma - 4}{3}\|\varphi(y)\| < \|\beta(y)\|. \tag{7.16}$$

Proof: The inequality (7.16) holds trivially for $\Gamma = 4$. For $\Gamma > 4$, using in sequence (7.9), $\|\widetilde{\varphi}(x)\|^2 \leq \widetilde{\varphi}(x)^T\varphi(x)$, twice (7.15), and (7.9), we get

$$\begin{aligned} \|\beta(y)\| &\geq \|\beta(x)\| - 4\|\widetilde{\varphi}(x)\| \geq \|\beta(x)\| - 4\sqrt{\widetilde{\varphi}^T(x)\varphi(x)} \\ &> (\Gamma - 4)\sqrt{\widetilde{\varphi}^T(x)\varphi(x)} \geq \frac{\Gamma - 4}{3}\|\varphi(y)\|. \quad \square \end{aligned} \tag{7.17}$$

Theorem 7.6. *Let $\{x^k\}$ denote the sequence generated by Algorithm 3.1. with $\overline{\alpha} \in (0, 2\|A\|^{-1}]$ and*

$$\Gamma \geq 3\sqrt{\kappa(A)} + 4. \tag{7.18}$$

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

Proof: Let x^k be generated by Algorithm 3.1 and let Γ satisfy (7.18). Let k_0 be that of Lemma 7.1 and let $k \geq k_0$ be such that x^k is not strictly proportional, so that $\Gamma^2\widetilde{\varphi}(x^k)^T\varphi(x^k) < \|\beta(x^k)\|^2$. Then by Corollary 7.5 the vector $y = x^k - \overline{\alpha}\widetilde{\varphi}(x^k)$ satisfies

$$\Gamma_1\|\varphi(y)\| < \|\beta(y)\| \tag{7.19}$$

with

$$\Gamma_1 = \frac{\Gamma - 4}{3} \geq \sqrt{\kappa(A)}.$$

Moreover, $y \in \Omega$ and by Lemma 7.1 and definition of y

$$\mathcal{A}(\widehat{x}) \supseteq \mathcal{A}(y) \supseteq \mathcal{A}(x^k) \supseteq \mathcal{B}(x^k) \supseteq \mathcal{B}(\widehat{x}), \tag{7.20}$$

so that by Theorem 7.3 the vector

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

satisfies

$$f(z) < \min\{f(x) : x \in \mathcal{W}_I\} \tag{7.21}$$

with $I = \mathcal{A}(y)$. Since I satisfies by (7.20) $A(\widehat{x}) \supseteq I \supseteq \mathcal{B}(\widehat{x})$, we have also

$$f(\widehat{x}) = \min\{f(x) : x \in \Omega\} = \min\{f(x) : x \in \mathcal{W}_I\}. \tag{7.22}$$

However, $z \in \Omega$, so that (7.22) contradicts (7.21). Thus all x^k are strictly proportional for $k \geq k_0$ so that

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

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

8 Numerical experiments

The key ingredients of the algorithm have already proved to be useful in the development of scalable algorithms for numerical solution variational inequalities [42, 43, 21, 23, 20]. The experiments presented here were carried out by the codes developed originally for the research in the preconditioning of variational inequalities by M. Domorádová [9] and for the development of BETI based scalable algorithms for variational inequalities by M. Sadowská [5]. To illustrate the effect of the steplength in the expansion step, we give here only two examples, a 2D inner obstacle problem discretized by the finite element method and a 3D contact problem of elasticity discretized by the boundary element method in combination with the BETI domain decomposition method. We used the stopping criterion $\|g^P(x)\| \leq 10^{-4}\|b\|$ and $\Gamma = 1$.

The first 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 $\Omega = [0,1] \times [0,1]$ and

$$\mathcal{K} = \{u \in H^1(\Omega) : -0.1 \leq u \text{ on } \Omega,\ u(x,0) = u(0,y) = 0 \text{ for } x, y \in [0,1]\}.$$

The solution of our model problem can be interpreted as a vertical displacement of a quarter of the membrane subject to the vertical traction with the unit density. See Fig. 1.

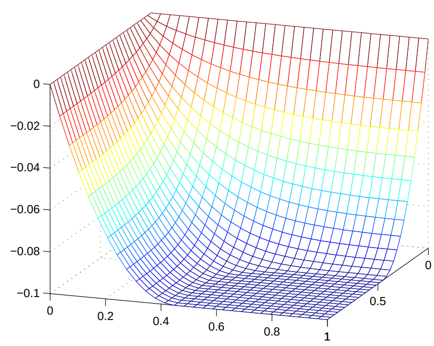

Figure 1: Solution of an inner obstacle model problem

The problem was discretized using the linear finite elements on a regular grid with 100 nodal variables in each direction, 10000 altogether. No preconditioning was used in order to isolate the effect of the steplength as much as possible. The number of the gradient and MPRGP iterations for the varying steplength $\overline{\alpha}$ are in Table 1.

Table 1: Performance of the algorithm on a 2D scalar problem

$\overline{\alpha}\|A\|$	gradient steps	MPRGP steps
0.2	545	871
0.4	445	761
0.6	338	689
0.8	293	625
1.0	247	557
1.2	221	530
1.4	204	504
1.6	187	529
1.8	169	495
2.0	156	488

The second problem arises from the application of the TBETI (total boundary element tearing and interconnecting) domain decomposition method [5] to the solution of a 3D contact problem of elasticity. The TBETI method proved to be an efficient scalable algorithm for the solution of variational inequalities. We applied it to evaluate the displacement of the elastic cube with a side equal to 10mm fixed on one vertical face above the plane obstacle and subjected to the volume forces -2100N/mm^3. The cube is 3mm over the obstacle and is decomposed into $5 \times 5 \times 5 = 125$ subdomains discretized by the piecewise constant boundary elements on a regular grid so that the primal and dual dimension of the discretized problem are 183000 and 92992, respectively. The Young modulus $E = 114000$MPa and the Poisson ratio $\nu = 0.24$. The solution is in Fig. 2.

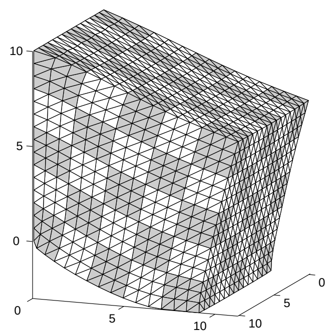

Figure 2: Cantilever cube over obstacle problem solution

The nonpenetration is described by 2500 inequalities, 550 being active in the solution. Observe that the unknowns are on the surface of the subdomains only. More details are in the Ph.D. Thesis by Marie Sadowská [41]. The number of the gradient and MPRGP iterations for the varying steplength $\overline{\alpha}$ are in Table 2.

Table 2: Performance of the algorithm on a 3D TBETI problem

$\overline{\alpha}\|A\|$	gradient steps	MPRGP steps
0.6	85	173
0.8	62	149
1.0	53	139
1.2	47	135
1.4	38	130
1.6	31	123
1.8	34	127
2.0	22	115

The results of numerical experiments are not surprising and indicate that the performance of the MPRPG algorithm is better for the longer steps. However, this result seems to contradict the estimate (5.1), which guarantees the best bound for $\overline{\alpha} = \|A\|^{-1}$. Our explanation is that the performance of the algorithm which combines several types of steps can hardly be captured by the worst one step bound. For example, the classical Euclidean contraction estimate [3] gives the best bound for

$$\overline{\alpha}_{\text{opt}} = \frac{2}{\lambda_{\min} + \lambda_{\max}},$$

so that it is natural to assume that the chain of consecutive expansion steps is effective for $\overline{\alpha} \approx \overline{\alpha}_{\text{opt}} \leq 2\|A\|^{-1}$. Moreover, for ill-conditioned problems obviously $\overline{\alpha}_{\text{opt}} \approx 2\|A\|^{-1}$. The estimate does not take into account the effect of the fast expansion of the active set.

9 Comments and conclusions

We extended the convergence theory of our MPRGP algorithm [26] so that it covers the longer expansion steps. The improved performance of the expansion steps with a longer steplength was observed a few years ago by M. Lesoinne [38], but no theory has been developed until now. Though the estimates do not explain the faster convergence of MPRGP for longer steplengths, they do substantiate the scalability of our FETI and BETI based algorithms for numerical solutions of variational inequalities [5, 23, 22, 25].

The result requires different tools of analysis as compared with the proofs of similar statements for $\overline{\alpha} \leq \|A\|^{-1}$. We extended also the results concerning the finite termination property. The proof shows that for a sufficiently large balancing parameter Γ the algorithm switches to the conjugate gradient method and enjoys a kind of superlinear convergence.

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] Bertsekas, D.P.: Nonlinear Optimization. Athena Scientific, Belmont 1999.

[4] 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, 67-102, 1997.

[5] J. Bouchala, Z. Dostál, and M. Sadowská, *Theoretically Supported Scalable BETI Method for Variational Inequalities*, Computing 82, 53-75, 2008.

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

[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, 321-338, 2001.

[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, 45-66, 2001.

[9] M. Domorádová and Z. Dostál, *Projector Preconditioning for Partially Bound Constrained Quadratic Optimization*, Numerical Linear Algebra with Applications 14,10, 791 - 806, 2007.

[10] Z. Dostál, *Box constrained quadratic programming with proportioning and projections*, SIAM J. Optimization 7, 3, 871-887, 1997.

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

[12] Z. Dostál, *A proportioning based algorithm for bound constrained quadratic programming with the rate of convergence*, Numerical Algorithms 34, 2-4, 293-302, 2003.

[13] Z. Dostál, *An optimal algorithm for bound and equality constrained quadratic programming problems with bounded spectrum*, Computing, 78, 311-328, 2006.

[14] Z. Dostál, *Inexact semi-monotonic Augmented Lagrangians with optimal feasibility convergence for quadratic programming with simple bounds and equality constraints*, SIAM J. Numer. Analysis, 45, 2, 500-513, 2007.

[15] Dostál, Z.: On the decrease of a quadratic function along the projected–gradient path, accepted in ETNA.

[16] 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, 161-173, 1998.

[17] Z. Dostál, A. Friedlander, and S.A. Santos, *Augmented Lagrangians with adaptive precision control for quadratic programming with simple bounds and equality constraints*, SIAM J. Optimization, 13,4, 1120-1140, 2003.

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

[19] 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, 1611-1627, 2000.

[20] Z. Dostál and D. Horák, *Scalability and FETI based algorithm for large discretized variational inequalities*, Math. and Comp. in Simulation 61, 3-6, 347 - 357, 2003.

[21] Z. Dostál and D. Horák, *Scalable FETI with Optimal Dual Penalty for Variational Inequalities*, Num. Lin. Algebra and Applications, 11, 5-6, 455 - 472, 2004.

[22] Z. Dostál and D. Horák *Theoretically supported scalable FETI for numerical solution of variational inequalities*, SIAM Journal on Numerical Analysis, 45, 2, 500513, 2007.

[23] Z. Dostál, D. Horák, and D. Stefanica, *A scalable FETI–DP algorithm for coercive variational inequalities,* IMACS Journal Applied Numerical Mathematics, 54, 3-4, 378-390, 2005.

[24] Z. Dostál, D. Horák, and D. Stefanica, *A Scalable FETI–DP algorithm with non–penetration mortar conditions on contact interface*, Submitted 2007.

[25] Z. Dostál, D. Horák, and D. Stefanica, *A Scalable FETI–DP Algorithm for Semi-coercive Variational Inequalities.* Computer Methods in Applied Mechanics and Engineering, 196, 8, 1369-1379, 2007.

[26] Z. Dostál and J. Schöberl, *Minimizing quadratic functions over non-negative cone with the rate of convergence and finite termination*, Computational Optimization and Applications, 30, 1, 23-44, 2005.

[27] Z. Dostál, V. Vondrák, D. Horák, C. Farhat, and P. Avery, *Scalable FETI algorithms for frictionless contact problems.* In: U. Langer et al., (eds.) Domain Methods in Science and Engineering XVII. Springer, Lecture Notes in Computational Science and Engineering (LNCSE) 60, Berlin 263-270, 2008.

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

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

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

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

[32] W.W. Hager, and H. Zhang, *A new active set algorithm for box constrained optimization*, 17, 2, 526-557, 2006.

[33] M. Hintermüller, K. Ito, and K. Kunisch, The primal–dual active set strategy as a semismooth Newton method. SIAM J. Optim., **13**, 865–888 (2003)

[34] S. Hüeber, G. Stadler, and B.I. Wohlmuth, A Primal-Dual Active Set Algorithm for Three-Dimensional Contact Problems with Coulomb Friction. SIAM J. Sci. Comput. **30**, 572–596, (2008)

[35] Z.-Q. Luo and P. Tseng, *Error bounds and convergence analysis of feasible descent methods: a general approach*, Annals of Operations Research, 46, 157-178, 1993.

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

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

[38] M. Lesoinne and Z. Dostál, *Accelerating Contact Area Detection in FETI-DPC Methods*, lecture at 16th Int. Conference on Domain Decomposition Methods, New York City, 2005.

[39] J. Nocedal and S.F.Wright, Numerical Optimization. Springer-Verlag, New York, 2000.

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

[41] M. Sadowská, *Scalable Total Beti for 2D and 3D Contact Problems*, FEI VŠB-TU Ostrava, Ostrava (2008).

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

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

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