SIAM J. OPTIM.
Vol. 20, No. 6, pp. 2913–2938

AN OPTIMAL ALGORITHM FOR MINIMIZATION OF QUADRATIC FUNCTIONS WITH BOUNDED SPECTRUM SUBJECT TO SEPARABLE CONVEX INEQUALITY AND LINEAR EQUALITY CONSTRAINTS*

ZDENĚK DOSTÁL† AND RADEK KUČERA‡

Abstract. An, in a sense, optimal algorithm for minimization of quadratic functions subject to separable convex inequality and linear equality constraints is presented. Its unique feature is an error bound in terms of bounds on the spectrum of the Hessian of the cost function. If applied to a class of problems with the spectrum of the Hessians in a given positive interval, the algorithm can find approximate solutions in a uniformly bounded number of simple iterations, such as matrix-vector multiplications. Moreover, if the class of problems admits a sparse representation of the Hessian, it simply follows that the cost of the solution is proportional to the number of unknowns. Theoretical results are illustrated by numerical experiments.

Key words. quadratic function, separable convex constraints, active set, augmented Lagrangian, gradient projections, convergence rate, optimality

AMS subject classifications. 65K05, 90C25, 90C06

DOI. 10.1137/090751414

1. Introduction.

We are concerned with the problem of efficiently finding the minimizer of a strictly convex quadratic function subject to separable convex inequality and linear equality constraints, that is,

$$\text{minimize } q(x) \quad \text{subject to} \quad x \in \Omega \tag{1.1}$$

with $q(x) = \frac{1}{2}\, x^\top A x - x^\top b$, $A \in \mathbb{R}^{n\times n}$ symmetric positive definite, $b \in \mathbb{R}^n$, and

$$\Omega = \{x \in \Omega_S : \ Cx = 0\}, \quad \Omega_S = \{x \in \mathbb{R}^n : \ f_1(\boldsymbol{x}_1) \le 0, \dots, f_{m_I}(\boldsymbol{x}_{m_I}) \le 0\}, \tag{1.2}$$

where $C \in \mathbb{R}^{m_E\times n}$, $f_i : \mathbb{R}^{n_i} \mapsto \mathbb{R}$ are continuously differentiable convex functions, and $\boldsymbol{x}_i \in \mathbb{R}^{n_i}$ denotes the ith segment of $x \in \mathbb{R}^n$, so that $x = (\boldsymbol{x}_1^\top, \dots, \boldsymbol{x}_{m_I}^\top)^\top$, $\sum_{i=1}^{m_I} n_i = n$. We suppose that Ω is nonempty, and, in order to avoid unnecessary complications, we assume, without loss of generality, that $f_i(\boldsymbol{x}_i) = 0$ implies $\nabla f_i(\boldsymbol{x}_i) \ne 0$. To enable an application of the stopping criterion, we assume $b \ne 0$ that is natural in many applications. We admit dependent rows of C.

Our research has been motivated by an effort to generalize the recent results of the first author and his collaborators in the development of scalable algorithms for the solution of multibody frictionless contact problems of linear elasticity [3, 14]. In this case, the dual equilibrium conditions for a system of elastic bodies in frictionless contact result in the bound constraints $\ell_i \le \boldsymbol{x}_i$ defined by $f_i(\boldsymbol{x}_i) = \ell_i - \boldsymbol{x}_i$ with $n_i = 1$. The same constraints appear in contact problems with Coulomb's friction in

*Received by the editors March 3, 2009; accepted for publication (in revised form) June 21, 2010; published electronically October 7, 2010. This research is supported by project MSM6198910027 provided by the Ministry of Education of the Czech Republic and by project GAČR 201/07/0294 provided by the Grant Agency of the Czech Republic.

http://www.siam.org/journals/siopt/20-6/75141.html

†Department of Applied Mathematics, FEECS VŠB-Technical University of Ostrava, Ostrava CZ-70833, Czech Republic (zdenek.dostal@vsb.cz).

‡Department of Mathematics and Descriptive Geometry, VŠB-Technical University of Ostrava, Ostrava CZ-70833, Czech Republic (radek.kucera@vsb.cz).

two dimensions [11, 12]. When $n_i = 2$, $\boldsymbol{x}_i = (x_j, x_{j+1})^\top$, and $f_i(\boldsymbol{x}_i) = x_j^2 + x_{j+1}^2 - r_i^2$, we get the circular constraint $x_j^2 + x_{j+1}^2 \le r_i^2$ corresponding to the dual representation of an isotropic friction law in three dimensions [25, 13]. If an orthotropic friction law is considered [26], then the circular constraints are replaced by the ellipsoidal ones $(x_j/a_j)^2 + (x_{j+1}/a_{j+1})^2 \le r_i^2$, for which the definition of f_i is a trivial modification of the previous case. Finally, if $\boldsymbol{x}_i$ is not subjected to any inequality constraint, we can take $f_i(\boldsymbol{x}_i) = -1$. A general form of (1.1) appears in the implementation of a sequential quadratically constrained quadratic programming method presented by Fukushima, Luo, and Tseng [21].

Another structure of (1.1) arises from application of the duality based domain decomposition methods for the parallel solution of discretized elliptic partial differential equations, such as variants of the FETI or BETI (finite/boundary element tearing and interconnecting) method introduced by Farhat and Roux [19] or by Langer and Steinbach [32], respectively. Since the Hessian A of the cost function associated with these methods is typically a well-conditioned matrix defined as a product of several large sparse matrices, it is natural to consider iterative solvers which do not require the matrix A to be explicitly assembled.

Efficient algorithms that are relevant for our research are based on the augmented Lagrangian method [27, 38]. This method generates the Lagrange multipliers for the equality constraints in the outer loop, while the inequality constrained auxiliary problems are solved in the inner loop. The auxiliary problems are of the type

$$\text{minimize } L(x, \mu^k, \rho) \quad \text{subject to} \quad x \in \Omega_S, \tag{1.3}$$

where

$$L(x, \mu, \rho) = q(x) + \mu^\top Cx + \frac{\rho}{2}\|Cx\|^2 \tag{1.4}$$

is the augmented Lagrangian to (1.1), $\mu \in \mathbb{R}^{m_E}$ denotes the vector of Lagrange multipliers for the equality constraints, and $\rho > 0$ is the penalty parameter. Such an approach was successfully applied by Conn, Gould, and Toint [4] in their code LANCELOT. The problems in the inner loop can be solved by a general method such as the trust region [5], conic programming [2], the gradient projection methods proposed by Rosen [39, 40], or, if the constraints are quadratic, by some algorithms developed for the solution of convex QCQP (quadratic constraints quadratic program) problems; see, e.g., Ecker and Niemi [17], Martínez [33], Anitescu [1], or Mehrotra and Sun [34]. However, we have not found any attempt to develop in this way a specialized algorithm with an error bounded independently of the constraints.

The aim of this paper is to develop an algorithm for the solution (1.1) with the rate of convergence in terms of bounds on the spectrum of the Hessian A of q. We call such algorithms *optimal for a fixed tolerance level* or briefly *optimal* because of their role in the development of optimal (scalable) algorithms for the solution of the optimization problems arising from the discretization of elliptic variational equalities or inequalities. Indeed, using the multigrid [23] or domain decomposition methods [18, 41, 3], it is possible to reduce such problems to the class of quadratic programming problems with the spectrum confined to a given interval, so that any algorithm with the rate of convergence in bounds on the spectrum can find an approximate solution with $O(1)$ matrix-vector multiplications, regardless of the discretization parameter. Since the Hessians of the discretized problems are sparse, it means that the cost of the solution increases asymptotically linearly with the number of variables, obviously the best that can be achieved.

So far, it seems that such algorithms were developed only for special cases of minimization of q subject to $x \in \Omega_S$, namely, the algorithms for simple bounds [15], for separable quadratic constraints [29], and, recently, for separable convex constraints [30]. Moreover, these results were also extended to bound and equality constraints [7, 8]. In particular, it was proved that if a variant of the augmented Lagrangian method is applied to a class of bound and equality constrained problems with uniformly bounded spectrum of the Hessian A, then it can find an approximate solution with $O(1)$ matrix-vector multiplications. The proof also required that the zero vector $0 \in \mathbb{R}^n$ is feasible. Here we prove similar results for (1.1) using a variant of the augmented Lagrangian method called SMALBE-M (semi-monotonic augmented Lagrangians for bounds and equality constraints) [9] which does not require an increase of the regularization parameter. Since the negative gradient can point into the direction which is not feasible, a special analysis had to be designed for adaptive precision control in the inner loop. Moreover, we relax the condition $0 \in \Omega$ in order to extend the applications of our results to the development of scalable algorithms for semicoercive contact problems with a given friction.

Let us introduce some conventions that we use throughout the whole paper. If $v \in \mathbb{R}^n$ is a vector, then $v_i \in \mathbb{R}$ denotes its ith entry, $1 \leq i \leq n$, $\boldsymbol{v}_i \in \mathbb{R}^{n_i}$ is its ith segment, $1 \leq i \leq m_I$, and $v = (\boldsymbol{v}_1^\top, \ldots, \boldsymbol{v}_{m_I}^\top)^\top$; the integers n_i are given by the formulation of problem (1.1). The Euclidean norm of $v \in \mathbb{R}^p$ is denoted by

$$\|v\| = (v_1^2 + \cdots + v_p^2)^{1/2},$$

and the same notation is used for the induced matrix norm. The eigenvalues of the Hessian A of q in (1.1) are denoted by $\lambda_i(A)$, and

$$\lambda_{\min}(A) = \lambda_1(A) \leq \cdots \leq \lambda_n(A) = \lambda_{\max}(A) = \|A\|.$$

The set of all eigenvalues is denoted by $\sigma(A)$, and it is called the spectrum of A. The spectral condition number $\kappa(A)$ of A is given by

$$\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}.$$

For any nonempty set of indices $\mathcal{I}$ and a vector $x \in \mathbb{R}^n$, we denote by $x_{\mathcal{I}}$ a subvector of x with the entries given by $\mathcal{I}$. In particular, $\boldsymbol{v}_i = v_{\mathcal{I}_i}$ with $\mathcal{I}_i = \{\sum_{k=1}^{i-1} n_k + 1, \ldots, \sum_{k=1}^{i} n_k\}$.

This paper is organized as follows. In section 2, two optimality criteria for the auxiliary problems (1.3) are discussed, i.e., the K-gradient that is the vector of the Karush–Kuhn–Tucker (KKT) conditions and the reduced gradient that is defined by the projection. The algorithm for solving (1.3) based on the reduced gradient is presented in section 3. We cite its rate of convergence proved in [30] that is the key point for our analysis of the augmented Lagrangian method. In section 4, we focus on the rate of convergence of the K-gradient. As the K-gradient is a discontinuous function, its convergence does not follow immediately from the convergence of iterates. Although the problem (1.1) is a generalization of problems investigated early in [7, 8], i.e., simple (lower) bounds are replaced by separable inequality constraints, it is not easy to adapt the proofs. The reason for this is that the boundary of the feasible set in (1.1) may be curved. To overcome this difficulty we linearize the constraints in the proof of Theorem 4.2 and then we use the three-step analysis based on Lemma 4.1. Section 5 presents our variant of the augmented Lagrangian method for solving (1.1).

In sections 6 and 7 we prove basic properties of the augmented Lagrangian originally derived for simple bound problems. The main results are summarized in sections 8 and 9, where we give upper bounds on the number of outer (augmented Lagrangian) iterations and on the number of matrix-vector multiplications, both independent of the size of the problem. Finally, section 10 presents results of numerical experiments that are in agreement with the theoretical analysis.

2. KKT and gradient splitting. This section deals with the problem

$$\text{minimize } q(x) \quad \text{subject to} \quad x \in \Omega_S. \tag{2.1}$$

Let us note that (2.1) comprises (1.3). We shall introduce two different optimality conditions to (2.1) and relations between them that are necessary for the next analysis.

As (2.1) is the minimization of the strictly convex function in the nonempty convex set, its solution $x^* \in \Omega_S$ exists and is necessarily unique [9, 36]. For an arbitrary $x \in \mathbb{R}^n$, let us denote the gradient $g = g(x)$ of q by

$$g = g(x) = \nabla q(x) = Ax - b.$$

It is well known that x^* is fully determined by the KKT conditions [36]:

$$f_i(\boldsymbol{x}_i^*) < 0 \quad \text{implies} \quad \boldsymbol{g}_i^* = \mathbf{0}, \tag{2.2}$$

$$f_i(\boldsymbol{x}_i^*) = 0 \quad \text{implies} \quad \boldsymbol{g}_i^* + \frac{\|\boldsymbol{g}_i^*\|}{\|\nabla f_i(\boldsymbol{x}_i^*)\|}\nabla f_i(\boldsymbol{x}_i^*) = \mathbf{0}, \tag{2.3}$$

where $\boldsymbol{g}_i^*$ is the ith segment of $g^* = g(x^*)$ and ∇f_i denotes the gradient of f_i, $i = 1, \ldots, m_I$.

Let

$$\mathcal{M} = \{1, 2, \ldots, m_I\}$$

denote the set of the indices of the separable constraints. Its subset $\mathcal{A}(x)$ comprising the indices for which $f_i(\boldsymbol{x}_i) = 0$ is called the *active set* of $x \in \Omega_S$, and the complement $\mathcal{F}(x)$ of $\mathcal{A}(x)$ in $\mathcal{M}$ is called the *free set* of $x \in \Omega_S$. Thus

$$\mathcal{A}(x) = \{i \in \mathcal{M} : f_i(\boldsymbol{x}_i) = 0\} \quad \text{and} \quad \mathcal{F}(x) = \{i \in \mathcal{M} : f_i(\boldsymbol{x}_i) < 0\}. \tag{2.4}$$

The following definitions introduced by Kučera [29] enable us to express alternatively the KKT conditions. Let us define the *free gradient* $\varphi = \varphi(x)$ and the *K-boundary gradient* $\beta = \beta(x)$ at $x \in \Omega_S$ by

$$\boldsymbol{\varphi}_i = \boldsymbol{g}_i \quad \text{for } i \in \mathcal{F}(x), \quad \boldsymbol{\varphi}_i = \mathbf{0} \quad \text{for } i \in \mathcal{A}(x), \tag{2.5}$$

$$\boldsymbol{\beta}_i = \mathbf{0} \quad \text{for } i \in \mathcal{F}(x), \quad \boldsymbol{\beta}_i = \boldsymbol{g}_i + \frac{\|\boldsymbol{g}_i\|}{\|\nabla f_i(\boldsymbol{x}_i)\|}\nabla f_i(\boldsymbol{x}_i) \quad \text{for } i \in \mathcal{A}(x). \tag{2.6}$$

It easily follows that KKT conditions (2.2), (2.3) are satisfied at $x^* \in \Omega_S$ if and only if the *K-gradient* $g^K = \varphi + \beta$ at x^* satisfies $g^K(x^*) = 0$. The function g^K is discontinuous at the boundary of Ω_S. Let us examine the properties of the components of g^K.

LEMMA 2.1. *If $x \in \Omega_S$, then*

$$\|\varphi(x)\|^2 = g(x)^\top \varphi(x), \tag{2.7}$$

$$\|\beta(x)\|^2 = 2g(x)^\top \beta(x). \tag{2.8}$$

Proof. Statement (2.7) is obvious. To prove (2.8), notice that (2.6) implies

$$\boldsymbol{\beta}_i^\top \boldsymbol{\beta}_i = 2\boldsymbol{g}_i^\top \left(\boldsymbol{g}_i + \frac{\|\boldsymbol{g}_i\|}{\|\nabla f_i(\boldsymbol{x}_i)\|} \nabla f_i(\boldsymbol{x}_i) \right) = 2\boldsymbol{g}_i^\top \boldsymbol{\beta}_i, \quad i \in \mathcal{A}(x). \qquad \square$$

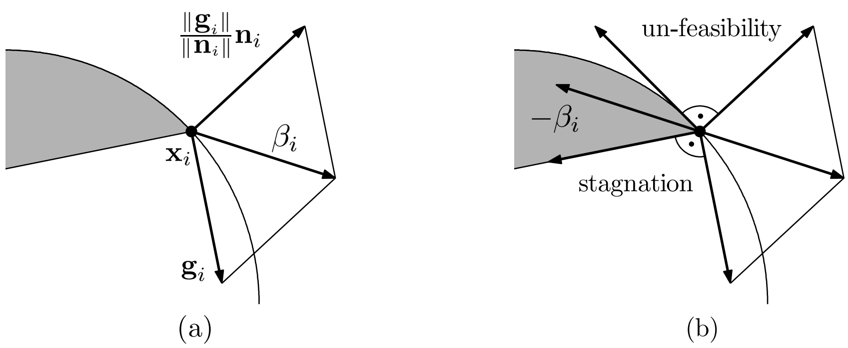

FIG. 2.1. *K-boundary gradient.*

Remark 2.1. Notice that $\boldsymbol{n}_i = \nabla f_i(\boldsymbol{x}_i)$ in the definition of $\boldsymbol{\beta}_i$ is the outer normal vector to the boundary of the feasible region for the ith segment of x. Thus $\boldsymbol{\beta}_i$ is the sum of two vectors of the same length as is shown in Figure 2.1(a). Moreover, examining Figure 2.1(b), one can deduce that $-\boldsymbol{\beta}_i$ is the feasible descent direction bisecting the angle between the stagnation direction and the direction which is not feasible. Therefore the negative K-boundary gradient $-\beta$ could be used when it is necessary to perform a descent step with the active segments of x. An algorithm based on this observation was developed in [29]. It is important that $-g^K$ is also a feasible decrease direction.

Remark 2.2. If the functions f_i define the bound constraints $x_i \geq \ell_i$, i.e., $f_i(x_i) = \ell_i - x_i$, $i = 1, \ldots, n$, then it is natural to assess the violation of the KKT conditions at $x \in \mathbb{R}^n$ by the *projected gradient* $g^P = g^P(x)$ with the components

$$g_i^P = g_i \quad \text{for} \quad i \in \mathcal{F}(x), \qquad g_i^P = \min\{g_i, 0\} \quad \text{for} \quad i \in \mathcal{A}(x).$$

See, e.g., Dostál [9]. A natural generalization of the projected gradient to the separable convex constraints reads as $g^P = \varphi^P + \beta^P$, where the components φ^P and β^P are defined by

$$\text{(2.9)} \quad \boldsymbol{\varphi}_i^P = \boldsymbol{g}_i \quad \text{for } i \in \mathcal{F}(x), \quad \boldsymbol{\varphi}_i^P = \mathbf{0} \quad \text{for } i \in \mathcal{A}(x),$$

$$\text{(2.10)} \quad \boldsymbol{\beta}_i^P = \mathbf{0} \quad \text{for } i \in \mathcal{F}(x), \quad \boldsymbol{\beta}_i^P = \frac{1}{2}\Big(\boldsymbol{g}_i + \frac{|\boldsymbol{g}_i^\top \nabla f_i(\boldsymbol{x}_i)|}{\|\nabla f_i(\boldsymbol{x}_i)\|^2} \nabla f_i(\boldsymbol{x}_i) \Big) \quad \text{for } i \in \mathcal{A}(x).$$

However, such generalization is not suitable for the development of algorithms as $-g^P$ *need not be a feasible decrease direction.* If f_i defines bound constraints, then $g_i^K = 2g_i^P$ or $g_i^K = g_i^P$, so that

$$\text{(2.11)} \qquad \|g^K(x)\| \geq \|g^P(x)\| \geq \frac{1}{2}\|g^K(x)\|.$$

Since the constraints that define Ω_S are separable, we can write $\Omega_S = \Omega_1 \times \cdots \times \Omega_{m_I}$, where $\Omega_i = \{\boldsymbol{x}_i \in \mathbb{R}^{n_i} : f_i(\boldsymbol{x}_i) \le 0\}$, $i \in \mathcal{M}$. Therefore, the orthogonal projection $P_{\Omega_S} : \mathbb{R}^n \mapsto \Omega_S$ is also separable and is defined by the orthogonal projections $P_{\Omega_i} : \mathbb{R}^{n_i} \mapsto \Omega_i$, $i \in \mathcal{M}$, so that

$$P_{\Omega_S}(x) = \begin{pmatrix} P_{\Omega_1}(\boldsymbol{x}_1) \\ \vdots \\ P_{\Omega_{m_I}}(\boldsymbol{x}_{m_I}) \end{pmatrix}.$$

It is easy to check that

$$P_{\Omega_i}(\boldsymbol{x}_i) = \begin{cases} \boldsymbol{x}_i & \text{for } \boldsymbol{x}_i \in \Omega_i, \\ \boldsymbol{y}_i & \text{for } \boldsymbol{x}_i \notin \Omega_i, \end{cases}$$

where $\boldsymbol{y}_i$ satisfies

$$f_i(\boldsymbol{y}_i) = 0, \qquad \boldsymbol{y}_i + \frac{\|\boldsymbol{x}_i - \boldsymbol{y}_i\|}{\|\nabla f_i(\boldsymbol{y}_i)\|} \nabla f_i(\boldsymbol{y}_i) = \boldsymbol{x}_i. \tag{2.12}$$

Let us define the *reduced gradient* $\widetilde{g} = \widetilde{g}(x)$ at $x \in \Omega_S$ for fixed $\overline{\alpha} > 0$ by

$$\widetilde{g}(x) = \frac{1}{\overline{\alpha}} \left(x - P_{\Omega_S} \left(x - \overline{\alpha} g(x) \right) \right)$$

or, equivalently, by

$$P_{\Omega_S}(x - \overline{\alpha} g(x)) = x - \overline{\alpha} \widetilde{g}(x).$$

It is easily seen that $\widetilde{g}$ is a continuous function and that KKT conditions (2.2) and (2.3) are satisfied at $x^* \in \Omega_S$ if and only if $\widetilde{g}(x^*) = 0$ [30]. Analogously to our decomposition of g^K, we also split $\widetilde{g}$ into the *reduced free gradient* $\widetilde{\varphi} = \widetilde{\varphi}(x)$ and the *reduced boundary gradient* $\widetilde{\beta} = \widetilde{\beta}(x)$ that are defined by

$$\widetilde{\boldsymbol{\varphi}}_i = \widetilde{\boldsymbol{g}}_i \quad \text{for } i \in \mathcal{F}(x), \quad \widetilde{\boldsymbol{\varphi}}_i = \boldsymbol{0} \quad \text{for } i \in \mathcal{A}(x),$$

$$\widetilde{\boldsymbol{\beta}}_i = \boldsymbol{0} \quad \text{for } i \in \mathcal{F}(x), \quad \widetilde{\boldsymbol{\beta}}_i = \widetilde{\boldsymbol{g}}_i \quad \text{for } i \in \mathcal{A}(x).$$

Observe that $\widetilde{\boldsymbol{\beta}}_i$ is defined so that

$$P_{\Omega_i}(\boldsymbol{x}_i - \overline{\alpha} \boldsymbol{g}_i) = \boldsymbol{x}_i - \overline{\alpha} \widetilde{\boldsymbol{\beta}}_i, \quad i \in \mathcal{A}(x). \tag{2.13}$$

Remark 2.3. We implicitly assume in this paper that the projection P_{Ω_i} can be computed at a negligible cost. For instance, the bound and circular constraints lead to simple explicit formulas for evaluating P_{Ω_i}. In the case of the ellipsoidal constraints, one can apply the Newton method in $\mathbb{R}^1$ in order to compute the polar (angle) coordinate of the projection [26].

The next lemma summarizes relevant relations including the components of $\widetilde{g}$.

LEMMA 2.2. *If $x \in \Omega_S$, then*

$$\|\widetilde{\varphi}(x)\|^2 \le \widetilde{\varphi}(x)^\top g(x), \tag{2.14}$$

$$\|\widetilde{\beta}(x)\|^2 \le \widetilde{\beta}(x)^\top g(x), \tag{2.15}$$

$$\widetilde{\varphi}(x)^\top g(x) \le \varphi(x)^\top g(x). \tag{2.16}$$

Proof. See [30] for the proof. ∎

COROLLARY 2.3. *Let us denote $\alpha_d = d^\top g(x)/d^\top Ad$ for $x \in \Omega_S$ and $d \in \mathbb{R}^n$, $d \neq 0$. Then*

$$\alpha_{\widetilde{\varphi}} \geq \|A\|^{-1}, \quad \alpha_{\widetilde{\beta}} \geq \|A\|^{-1}, \quad \text{and} \quad \alpha_{\varphi} \geq \|A\|^{-1}.$$

Proof. Use the previous lemma and (2.7) for the proof. ∎

There is a useful relation between β and $\widetilde{\beta}$ in the case that the active constraints are defined by linear functions.

LEMMA 2.4. *Let $\bar{x} \in \Omega_S$ be given, and let f_i be nonconstant linear functions, $i \in \mathcal{A}(\bar{x})$. Then*

$$\beta(\bar{x})^\top g(\bar{x}) \leq 2\widetilde{\beta}(\bar{x})^\top g(\bar{x}). \tag{2.17}$$

Proof. First observe that the vectors

$$\boldsymbol{n}_i = \nabla f_i(\bar{\boldsymbol{x}}_i)$$

are nonzero by the assumption and $f_i(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_i)^\top \boldsymbol{n}_i$, $i \in \mathcal{A}(\bar{x})$. We shall distinguish two cases.

(a) If $\bar{\boldsymbol{x}}_i - \overline{\alpha}\boldsymbol{g}_i \in \Omega_i$, then

$$\widetilde{\boldsymbol{\beta}}_i = \widetilde{\boldsymbol{g}}_i = \frac{1}{\overline{\alpha}}\big(\bar{\boldsymbol{x}}_i - P_{\Omega_i}(\bar{\boldsymbol{x}}_i - \overline{\alpha}\boldsymbol{g}_i)\big) = \boldsymbol{g}_i$$

and

$$0 \leq \boldsymbol{\beta}_i^\top \boldsymbol{g}_i = \boldsymbol{g}_i^\top \boldsymbol{g}_i + \frac{\|\boldsymbol{g}_i\|}{\|\boldsymbol{n}_i\|}\,\boldsymbol{n}_i^\top \boldsymbol{g}_i \leq 2\boldsymbol{g}_i^\top \boldsymbol{g}_i = 2\widetilde{\boldsymbol{\beta}}_i^\top \boldsymbol{g}_i.$$

(b) Let $\bar{\boldsymbol{x}}_i - \overline{\alpha}\boldsymbol{g}_i \notin \Omega_i$. Then

$$0 < f_i(\bar{\boldsymbol{x}}_i - \overline{\alpha}\boldsymbol{g}_i) = ((\bar{\boldsymbol{x}}_i - \overline{\alpha}\boldsymbol{g}_i) - \bar{\boldsymbol{x}}_i)^\top \boldsymbol{n}_i = -\overline{\alpha}\boldsymbol{g}_i^\top \boldsymbol{n}_i,$$

so that $\boldsymbol{n}_i^\top \boldsymbol{g}_i < 0$. Our next step is the proof that $\widetilde{\boldsymbol{\beta}}_i$ is determined by the formula

$$\widetilde{\boldsymbol{\beta}}_i = \boldsymbol{g}_i - \frac{\boldsymbol{n}_i^\top \boldsymbol{g}_i}{\|\boldsymbol{n}_i\|^2}\,\boldsymbol{n}_i. \tag{2.18}$$

As $\widetilde{\boldsymbol{\beta}}_i$ should satisfy the definition (2.13) containing the projection P_{Ω_i}, it is enough to verify (2.12) for $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ replaced by $\bar{\boldsymbol{x}}_i - \overline{\alpha}\boldsymbol{g}_i$ and $\bar{\boldsymbol{x}}_i - \overline{\alpha}\widetilde{\boldsymbol{\beta}}_i$, respectively, with $\widetilde{\boldsymbol{\beta}}_i$ given by (2.18). We get easily

$$f_i(\bar{\boldsymbol{x}}_i - \overline{\alpha}\widetilde{\boldsymbol{\beta}}_i) = -\overline{\alpha}\widetilde{\boldsymbol{\beta}}_i^\top \boldsymbol{n}_i = -\overline{\alpha}\left(\boldsymbol{g}_i^\top \boldsymbol{n}_i - \frac{\boldsymbol{n}_i^\top \boldsymbol{g}_i}{\|\boldsymbol{n}_i\|^2}\,\boldsymbol{n}_i^\top \boldsymbol{n}_i\right) = 0,$$

and, using simple manipulations, substituting (2.18), and $\boldsymbol{n}_i^\top \boldsymbol{g}_i < 0$,

$$\begin{aligned}
\bar{\boldsymbol{x}}_i - \overline{\alpha}\widetilde{\boldsymbol{\beta}}_i + \frac{\|\bar{\boldsymbol{x}}_i - \overline{\alpha}\boldsymbol{g}_i - \bar{\boldsymbol{x}}_i + \overline{\alpha}\widetilde{\boldsymbol{\beta}}_i\|}{\|\boldsymbol{n}_i\|}\,\boldsymbol{n}_i &= \bar{\boldsymbol{x}}_i - \overline{\alpha}\widetilde{\boldsymbol{\beta}}_i + \frac{\overline{\alpha}\|\boldsymbol{g}_i - \widetilde{\boldsymbol{\beta}}_i\|}{\|\boldsymbol{n}_i\|}\,\boldsymbol{n}_i \\
&= \bar{\boldsymbol{x}}_i - \overline{\alpha}\boldsymbol{g}_i + \frac{\overline{\alpha}}{\|\boldsymbol{n}_i\|^2}\,\big(\boldsymbol{n}_i^\top \boldsymbol{g}_i + |\boldsymbol{n}_i^\top \boldsymbol{g}_i|\big)\,\boldsymbol{n}_i \\
&= \bar{\boldsymbol{x}}_i - \overline{\alpha}\boldsymbol{g}_i.
\end{aligned}$$

Now we have proved (2.18). Using (2.18) and $-\|\boldsymbol{g}_i\| \leq \boldsymbol{g}_i^\top \boldsymbol{n}_i/\|\boldsymbol{n}_i\|$, we get

$$0 \leq \boldsymbol{\beta}_i^\top \boldsymbol{g}_i = \boldsymbol{g}_i^\top \boldsymbol{g}_i - \frac{\|\boldsymbol{g}_i\|}{\|\boldsymbol{n}_i\|} |\boldsymbol{n}_i^\top \boldsymbol{g}_i| \leq \boldsymbol{g}_i^\top \boldsymbol{g}_i + \frac{\boldsymbol{g}_i^\top \boldsymbol{n}_i}{\|\boldsymbol{n}_i\|^2} |\boldsymbol{n}_i^\top \boldsymbol{g}_i| = \widetilde{\boldsymbol{\beta}}_i^\top \boldsymbol{g}_i \leq 2\widetilde{\boldsymbol{\beta}}_i^\top \boldsymbol{g}_i.$$

We have shown $\boldsymbol{\beta}_i^\top \boldsymbol{g}_i \leq 2\widetilde{\boldsymbol{\beta}}_i^\top \boldsymbol{g}_i$ for all $i \in \mathcal{A}(\bar{x})$. As $\boldsymbol{\beta}_i = \widetilde{\boldsymbol{\beta}}_i = \mathbf{0}$ for all $i \in \mathcal{F}(\bar{x})$, the lemma is proved. □

To simplify the general formulation of our optimality results, let us introduce a more formal definition of optimal feasible algorithms.

Definition 2.5. *Let $\mathcal{T}$ be an index set. For each $t \in \mathcal{T}$, let q_t and Ω_S^t be a quadratic function and a nonempty closed convex set defined as in* (1.1) *and* (1.2), *respectively. Denote*

$$x_t^* = \arg\min_{x \in \Omega_S^t} q_t(x). \tag{2.19}$$

An algorithm is called R-linearly K-convergent *(or briefly K-convergent) for the solution of* (2.19) *if it generates, for each $t \in \mathcal{T}$ and each $x_t^0 \in \Omega_S^t$, the iterates $x_t^k \in \Omega_S^t$ converging to x_t^* such that there are $\bar{\kappa} > 0$ and $\nu \in (0,1)$ independent on t that satisfy*

$$\|g_t^K(x_t^k)\|^2 \leq \bar{\kappa}\nu^k \big(q_t(x_t^0) - q_t(x_t^*)\big), \quad k = 1, 2, \ldots,$$

where g_t^K is the K-gradient corresponding to q_t.

3. Generalized KPRGP algorithm for separable constraints.

The algorithm KPRGP (K-proportioning with reduced gradient projections) for solving (2.1) that we consider here is a modification of the modified proportioning with reduced gradient projection (MPRGP) algorithm for the solution of strictly convex bound constrained problems [15, 9]. The MPRGP algorithm is a result of the development starting from the early paper on adaptation of the conjugate gradient method to the solution of bound constrained problems by Polyak [37]. The algorithm also uses the gradient projections introduced by Moré and Toraldo [35] and an adaptive precision control introduced independently by Friedlander and Martínez [20] and by Dostál [6]. A unique feature of MPRGP is the error bound in terms of bounds on the spectrum [15, 9]. See Dostál [9] for more details and Kučera [30] for the extension to KPRGP.

The algorithm KPRGP generates a sequence of iterates $\{x^k\}$ that approximate the solution x^*. It exploits a given constant $\Gamma > 0$, a test to decide about leaving the face, a fixed steplength $\overline{\alpha} \in (0, \|A\|^{-1}]$ defining the components of the reduced gradient, and three types of steps:

- the *expansion step*: $x^{k+1} = x^k - \overline{\alpha}\widetilde{\varphi}(x^k)$,
- the *proportioning step*: $x^{k+1} = x^k - \overline{\alpha}\widetilde{\beta}(x^k)$,
- the *conjugate gradient step*: $x^{k+1} = x^k - \alpha_{cg}^k p^k$, where α_{cg}^k and the conjugate gradient directions p^k are computed recurrently [22]; the recurrence starts from x^s generated by the last expansion or proportioning step and satisfies $\mathcal{A}(x^{k+1}) = \mathcal{A}(x^s)$.

The expansion step may add indices to the current active set, while the proportioning step may release indices from the current active set. Notice that both steps are always feasible, since $\widetilde{\varphi}$ and $\widetilde{\beta}$ are defined via projections. The conjugate gradient steps are used to carry out efficiently the minimization of the objective q in the interior of the face

$$\mathcal{W}(x^s) = \{x \in \Omega_S : \ x_{\mathcal{A}} = x_{\mathcal{A}}^s, \ \mathcal{A} = \mathcal{A}(x^s)\}.$$

Moreover, the *releasing criterion*

$$\widetilde{\beta}(x^k)^\top g(x^k) \le \Gamma\ \widetilde{\varphi}(x^k)^\top g(x^k) \tag{3.1}$$

decides which of the steps will be performed. Notice that (3.1) assesses the relation between components of the optimality criterion $\widetilde{\varphi}(x^k)$ and $\widetilde{\beta}(x^k)$ in the projection to the gradient $g(x^k)$. Since $\widetilde{\beta}(x^k)$ is the normal direction to $\mathcal{W}(x^s)$ at x^k, it is natural to leave the face when $\widetilde{\beta}(x^k)$ dominates. The resulting algorithm reads as follows.

ALGORITHM 3.1 (KPRGP). *Let $x^0 \in \Omega_S$, $\Gamma > 0$, and $\overline{\alpha} \in (0, \|A\|^{-1}]$ be given. For x^k, x^s known, $0 \le s \le k$, where x^s is computed by the last expansion or proportioning step, choose x^{k+1} by the following rules:*

(i) *If $\widetilde{g}(x^k) = 0$, set $x^{k+1} = x^k$.*

(ii) *If x^k satisfies* (3.1)*, then try to generate x^{k+1} by the conjugate gradient step. If $x^{k+1} \in \mathit{Int}\,\mathcal{W}(x^s)$, then accept it; otherwise generate x^{k+1} by the expansion step.*

(iii) *If x^k does not satisfy* (3.1)*, then generate x^{k+1} by the proportioning step.*

Algorithm 3.1 enjoys essentially the same rate of convergence as MPRGP. The main result reads as follows.

THEOREM 3.1. *Let Γ, $a_{\min}$ be given constants, $0 < \Gamma$, $0 < a_{\min} \le \lambda_{\min}(A)$, and $\widehat{\Gamma} = \max\{\Gamma, \Gamma^{-1}\}$. Let $x^* \in \Omega_S$ be the solution to* (2.1)*, and let $\{x^k\}$ be the sequence generated by Algorithm* 3.1 *with $\overline{\alpha} \in (0, \|A\|^{-1}]$. Then*

$$q(x^{k+1}) - q(x^*) \le \nu \left(q(x^k) - q(x^*) \right), \tag{3.2}$$

where

$$\nu = 1 - \frac{\overline{\alpha} a_{min}}{2 + 2\widehat{\Gamma}}. \tag{3.3}$$

The solution error is bounded by

$$\|x^k - x^*\|^2 \le \frac{2\nu^k}{a_{\min}} \left(q(x^0) - q(x^*) \right).$$

Moreover, if x^{k+1} is generated by the conjugate gradient step, then

$$q(x^{k+1}) \le q(x^k - \alpha\varphi(x^k)) \quad \forall \alpha \in \mathbb{R}. \tag{3.4}$$

Proof. See [30] for the proof. ∎

Remark 3.1. The statement of Theorem 3.1 is independent of the choice of the functions f_i which define the constraints defining Ω_S.

Remark 3.2. The analysis of Algorithm 3.1 performed in [30] does not necessarily require defining the free and active set by (2.4). It needs only a complementary decomposition of the index set $\mathcal{M}$ at the kth iterate x^k on $\hat{\mathcal{F}}(x^k)$ and $\hat{\mathcal{A}}(x^k)$ so that $\mathcal{M} = \hat{\mathcal{F}}(x^k) \cup \hat{\mathcal{A}}(x^k)$ and $\hat{\mathcal{F}}(x^k) \cap \hat{\mathcal{A}}(x^k) = \emptyset$. Replacing in the kth iterate of Algorithm 3.1 $\mathcal{F}(x^k)$ and $\mathcal{A}(x^k)$ by an arbitrary choice of $\hat{\mathcal{F}}(x^k)$ and $\hat{\mathcal{A}}(x^k)$, respectively, we obtain the modified algorithm which exhibits the same convergence properties.

4. Rate of convergence of the K-gradient. In this section we present new results comprising the convergence of the K-gradient during Algorithm 3.1. Let us note that it is not enough to have the rate of convergence in terms of bounds on the spectrum of A, but it is also necessary to be able to recognize effectively when we are near the solution. In particular, the analysis of our variant of the augmented Lagrangian method for solving (1.1) also requires the convergence rate of the K-gradient g^K. This result is not an easy corollary of Theorem 3.1 as $g^K(x)$ is not continuous. Let us begin with the following auxiliary observations.

LEMMA 4.1. *Let the assumptions of Theorem* 3.1 *be satisfied. Then*

$$q(x^k) - q(x^{k+1}) \leq \frac{1+\nu}{1-\nu}\nu \left(q(x^{k-1}) - q(x^k)\right),$$

where ν *is defined by* (3.3).

Proof. We shall repeatedly apply (3.2). As

$$\begin{aligned} q(x^{k-1}) - q(x^k) &= \left(q(x^{k-1}) - q(x^*)\right) - \left(q(x^k) - q(x^*)\right) \\ &\geq (1-\nu)\left(q(x^{k-1}) - q(x^*)\right) \\ &\geq \frac{1-\nu}{\nu}\left(q(x^k) - q(x^*)\right) \end{aligned}$$

and

$$\begin{aligned} q(x^k) - q(x^{k+1}) &= 2\left(\frac{q(x^k)+q(x^{k+1})}{2} - q(x^{k+1})\right) \\ &\leq 2\left(\frac{q(x^k)+q(x^{k+1})}{2} - q(x^*)\right) \\ &\leq (1+\nu)\left(q(x^k) - q(x^*)\right), \end{aligned}$$

we get

$$\frac{1}{1+\nu}\left(q(x^k) - q(x^{k+1})\right) \leq q(x^k) - q(x^*) \leq \frac{\nu}{1-\nu}\left(q(x^{k-1}) - q(x^k)\right).$$

This completes the proof. ∎

THEOREM 4.2. *Let the assumptions of Theorem* 3.1 *be satisfied. Then the K-gradient is bounded by*

$$\|g^K(x^k)\|^2 \leq \bar{\kappa}\nu^k\left(q(x^0) - q(x^*)\right), \quad k \geq 1, \tag{4.1}$$

where ν *is defined by* (3.3) *and*

$$\bar{\kappa} = \widehat{\Gamma}\,\frac{10(1+\nu)}{\overline{\alpha}(1-\nu)}.$$

Proof. Notice that it is enough to estimate separately $\varphi(x^k)$ and $\beta(x^k)$ as

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

We start with $\beta(x^k)$. Let us introduce the linearized problem associated with x^k:

$$\text{minimize } q(x) \quad \text{subject to} \quad x \in \hat{\Omega}_1^k \times \cdots \times \hat{\Omega}_{m_I}^k, \tag{4.2}$$

where $\hat{\Omega}_i^k = \{\boldsymbol{x}_i \in \mathbb{R}^{n_i} : \hat{f}_i(\boldsymbol{x}_i) \leq 0\}$ and

$$\begin{aligned} \hat{f}_i(\boldsymbol{x}_i) &= (\boldsymbol{x}_i - \boldsymbol{x}_i^k)^\top \nabla f_i(\boldsymbol{x}_i^k) \qquad &&\text{for } i \in \mathcal{A}(x^k), \\ \hat{f}_i(\boldsymbol{x}_i) &= -1 &&\text{for } i \in \mathcal{F}(x^k). \end{aligned}$$

In other words, the constraints on $\boldsymbol{x}_i$, $i \in \mathcal{F}(x^k)$, are omitted in (4.2), while the ones on $\boldsymbol{x}_i$, $i \in \mathcal{A}(x^k)$, are represented by the linear functions tangential to Ω_i at $\boldsymbol{x}_i^k$. Although x^k is the iterate to problem (2.1), it may be interpreted also as an iterate to problem (4.2) but with the original active set $\mathcal{A}(x^{k-1})$; see Figure 4.1. This observation is the crucial point of our proof since, taking into account Remarks 3.1 and 3.2, we can apply Lemma 4.1 to three iterates x^{k-1}, x^k, and $\hat{x}^{k+1}$, where $\hat{x}^{k+1}$ is the new iterate performed in the linearized problem (4.2) at x^k. Before doing that, we shall separately analyze three possible steps generating $\hat{x}^{k+1}$.

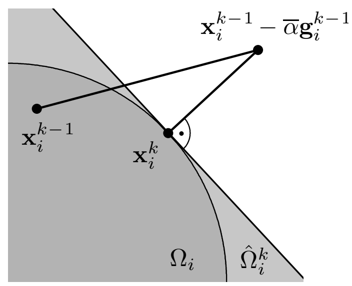

FIG. 4.1. *The same $\boldsymbol{x}_i^k$ is determined by Ω_i and $\hat{\Omega}_i^k$.*

(i) First assume that $\hat{x}^{k+1}$ is generated by the proportioning step. Using

$$\overline{\alpha} \leq \alpha_{\widetilde{\beta}} = \widetilde{\beta}(x^k)^\top g(x^k)/\widetilde{\beta}(x^k)^\top A\widetilde{\beta}(x^k)$$

(see Corollary 2.3) and (2.17), we obtain

$$\begin{aligned} q(x^k) - q(\hat{x}^{k+1}) &= \overline{\alpha}\widetilde{\beta}(x^k)^\top g(x^k) - \frac{\overline{\alpha}^2}{2}\,\widetilde{\beta}(x^k)^\top A\widetilde{\beta}(x^k) \\ &\geq \overline{\alpha}\widetilde{\beta}(x^k)^\top g(x^k) - \frac{\overline{\alpha}}{2}\,\alpha_{\widetilde{\beta}}\,\widetilde{\beta}(x^k)^\top A\widetilde{\beta}(x^k) \\ &= \frac{\overline{\alpha}}{2}\,\widetilde{\beta}(x^k)^\top g(x^k) \\ &\geq \frac{\overline{\alpha}}{4}\,\beta(x^k)^\top g(x^k). \end{aligned}$$

Combining the first and last terms with (2.8), we get

$$\|\beta(x^k)\|^2 \leq \frac{8}{\overline{\alpha}}\left(q(x^k) - q(\hat{x}^{k+1})\right). \tag{4.3}$$

(ii) Let $\hat{x}^{k+1}$ be generated by the expansion step. Analogously to the proportioning step, one can derive

$$q(x^k) - q(\hat{x}^{k+1}) \geq \frac{\overline{\alpha}}{2}\,\widetilde{\varphi}(x^k)^\top g(x^k). \tag{4.4}$$

As the expansion step is performed if (3.1) holds, we obtain by (2.17) that

$$q(x^k) - q(\hat{x}^{k+1}) \geq \Gamma^{-1}\,\frac{\overline{\alpha}}{2}\,\widetilde{\beta}(x^k)^\top g(x^k) \geq \Gamma^{-1}\,\frac{\overline{\alpha}}{4}\,\beta(x^k)^\top g(x^k).$$

Combining the latter inequality again with (2.8), we get

$$\|\beta(x^k)\|^2 \leq \Gamma \frac{8}{\overline{\alpha}} \left(q(x^k) - q(\hat{x}^{k+1}) \right). \tag{4.5}$$

(iii) Finally assume that $\hat{x}^{k+1}$ is generated by the conjugate gradient step. Using (3.4), Corollary 2.3, and (2.16), we get

$$q(x^k) - q(\hat{x}^{k+1}) \geq q(x^k) - q(x^k - \overline{\alpha}\varphi(x^k)) \geq \frac{\overline{\alpha}}{2} \varphi(x^k)^\top g(x^k) \geq \frac{\overline{\alpha}}{2} \widetilde{\varphi}(x^k)^\top g(x^k).$$

Hence we got again (4.4). Since the conjugate gradient step is performed when (3.1) holds, we get again (4.5). Summarizing (4.3) and (4.5), it follows that

$$\|\beta(x^k)\|^2 \leq \max\{1, \Gamma\} \frac{8}{\overline{\alpha}} \left(q(x^k) - q(\hat{x}^{k+1}) \right) \leq \widehat{\Gamma} \frac{8}{\overline{\alpha}} \left(q(x^k) - q(\hat{x}^{k+1}) \right) \tag{4.6}$$

for all three steps of Algorithm 3.1. Applying Lemma 4.1 to $\hat{x}^{k+1}$, x^k, and x^{k-1}, we obtain

$$\|\beta(x^k)\|^2 \leq \widehat{\Gamma} \frac{8(1+\nu)}{\overline{\alpha}(1-\nu)} \nu \left(q(x^{k-1}) - q(x^k) \right),$$

and using (3.2), we get the final bound

$$\begin{aligned} \|\beta(x^k)\|^2 &\leq \widehat{\Gamma} \frac{8(1+\nu)}{\overline{\alpha}(1-\nu)} \nu \left(q(x^{k-1}) - q(x^*) \right) \\ &\leq \widehat{\Gamma} \frac{8(1+\nu)}{\overline{\alpha}(1-\nu)} \nu^k \left(q(x^0) - q(x^*) \right). \end{aligned} \tag{4.7}$$

Performing the analogous analysis with $\varphi(x^k)$, we get the bound for $\|\varphi(x^k)\|^2$. We shall analyze again three possible steps generating $\hat{x}^{k+1}$ to the linearized problem (4.2) at x^k. As the proportioning step is performed when (3.1) does not hold, we obtain using the same arguments as in the previous case (i) that

$$q(x^k) - q(\hat{x}^{k+1}) \geq \frac{\overline{\alpha}}{2} \widetilde{\beta}(x^k)^\top g(x^k) > \Gamma \frac{\overline{\alpha}}{2} \widetilde{\varphi}(x^k)^\top g(x^k). \tag{4.8}$$

Notice that now $\widetilde{\varphi}(x^k) = \varphi(x^k)$, since $\widetilde{\varphi}(x^k)$ and $\varphi(x^k)$ are considered in the problem (4.2) in which the constraints corresponding to the free set $\mathcal{F}(x^k)$ are omitted. Therefore (4.8) and (2.7) give

$$q(x^k) - q(\hat{x}^{k+1}) \geq \Gamma \frac{\overline{\alpha}}{2} \varphi(x^k)^\top g(x^k) = \Gamma \frac{\overline{\alpha}}{2} \|\varphi(x^k)\|^2. \tag{4.9}$$

As both expansion and conjugate gradient steps lead to (4.4), for these two steps we get by (2.7) that

$$q(x^k) - q(\hat{x}^{k+1}) \geq \frac{\overline{\alpha}}{2} \varphi(x^k)^\top g(x^k) = \frac{\overline{\alpha}}{2} \|\varphi(x^k)\|^2. \tag{4.10}$$

Combining (4.9) and (4.10), it follows that

$$\|\varphi(x^k)\|^2 \leq \max\{1, \Gamma^{-1}\} \frac{2}{\overline{\alpha}} \left(q(x^k) - q(\hat{x}^{k+1}) \right) \leq \widehat{\Gamma} \frac{2}{\overline{\alpha}} \left(q(x^k) - q(\hat{x}^{k+1}) \right).$$

Applying Lemma 4.1 as in deriving the previous estimate, we arrive at

$$\|\varphi(x^k)\|^2 \le \widehat{\Gamma}\,\frac{2(1+\nu)}{\overline{\alpha}(1-\nu)}\,\nu^k\left(q(x^0)-q(x^*)\right). \tag{4.11}$$

We complete the proof by summing up (4.7) and (4.11). ▯

There is an immediate corollary of Theorem 4.2.

COROLLARY 4.3. *Let each* Ω_S^t *in* (2.19) *be defined as in* (1.2), *let* $a_{\min}$, $a_{\max}$, *and* $0 < a_{\min} \le a_{\max}$ *be such that the spectrum* $\sigma(A_t)$ *of the Hessian of each* q_t *satisfies* $\sigma(A_t) \subseteq [a_{\min}, a_{\max}]$.

Then Algorithm 3.1 *with parameters* $\Gamma > 0$ *and* $\overline{\alpha} \in (0, a_{\max}^{-1}]$ *is a K-convergent solver for the class of problems* (2.19).

5. Algorithm for equality and separable convex inequality constraints. Having an, in a sense, optimal algorithm for the solution of auxiliary problem (2.1), we are ready to address problem (1.1). In what follows, we denote its (unique) solution by $\widehat{x} \in \Omega$.

For $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^{m_E}$, and $\rho > 0$, we denote the gradient of augmented Lagrangian (1.4) with respect to the first variable by

$$g = g(x, \mu, \rho) = Ax - b + C^\top \mu + \rho C^\top C x.$$

Moreover, for $x \in \Omega_S$ we introduce the K-gradient $g^K = g^K(x, \mu, \rho)$ by (2.5) and (2.6). The solution $\widehat{x}$ is fully determined by the KKT conditions for the augmented Lagrangian [36]

$$g^K(\widehat{x}, \lambda, \rho) = 0, \quad C\widehat{x} = 0, \tag{5.1}$$

where $\lambda \in \mathbb{R}^{m_E}$ is not necessarily unique.

The following SMALSE-M (semimonotonic augmented Lagrangians for separable and equality constraints) algorithm is a modification of the SMALBE and SMALBE-M algorithms of [7, 8, 9] for the minimization of a convex quadratic function subject to bound and equality constraints. Unlike SMALBE, SMALSE-M admits more general constraints, keeps the penalty parameter ρ fixed, and updates the balancing parameter M_k in the same way as SMALBE-M. The complete SMALSE-M algorithm, without a stopping criterion, reads as follows.

ALGORITHM 5.1 (SMALSE-M). *Given a symmetric positive definite matrix* $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{m_E\times n}$, $b \in \mathbb{R}^n$, *and functions* f_i, $i = 1, \dots, m_I$, *defining separable convex inequality constraints* (1.2).

Initialization. Choose $\eta > 0$, $\beta > 1$, $M_0 > 0$, $\rho > 0$, $\lambda^0 \in \mathbb{R}^{m_E}$. Set $k := 0$.

Step 1 (inner iteration with adaptive precision control). Find $x^k \in \Omega_S$ such that

$$\|g^K(x^k, \lambda^k, \rho)\| \le \min\{M_k\|Cx^k\|, \eta\}. \tag{5.2}$$

Step 2 (updating the Lagrange multipliers).

$$\lambda^{k+1} = \lambda^k + \rho C x^k. \tag{5.3}$$

Step 3 (update M provided the increase of the Lagrangian is not sufficient). If $k > 0$ and

$$L(x^k, \lambda^k, \rho) < L(x^{k-1}, \lambda^{k-1}, \rho) + \frac{\rho}{2}\|Cx^k\|^2, \tag{5.4}$$

then $M_{k+1} = M_k/\beta$; else $M_{k+1} = M_k$.

Step 4. Set $k := k + 1$ and go to Step 1.

In Step 1 we can use any algorithm for minimizing the strictly convex quadratic function subject to the separable convex inequality constraints as long as it guarantees the convergence of the K-gradient g^K to zero, such as Algorithm 3.1 of section 3. The next lemma shows that Algorithm 5.1 is well defined, that is, any algorithm for the solution of the auxiliary problems required in Step 1 which guarantees the convergence of g^K to zero generates either x^k, which satisfies (5.2) in a finite number of steps, or approximations which converge to the solution of (1.1).

LEMMA 5.1. *Let $M > 0$, $\lambda \in \mathbb{R}^{m_E}$, $\eta > 0$, and $\rho \geq 0$ be given. Let $\{y^k\}$ be a sequence generated by a K-convergent algorithm to the problem*

$$\min_{x\in\Omega_S} L(x, \lambda, \rho). \tag{5.5}$$

Then $\{y^k\}$ either converges to the solution $\widehat{x}$ of (1.1)*, or there is an index k such that*

$$\|g^K(y^k, \lambda, \rho)\| \leq \min\{M\|Cy^k\|, \eta\}. \tag{5.6}$$

Proof. First notice that $\{y^k\}$ converges to $\widehat{y}$, which satisfies $g^K(\widehat{y}, \lambda, \rho) = 0$. The rest of the proof is based on the fact that $g^K(y^k, \lambda, \rho)$ converges to the zero vector by assumption. If (5.6) does not hold for any k, then $\|g^K(y^k, \lambda, \rho)\| > M\|Cy^k\|$ for any k and $\|Cy^k\|$ also converges to the zero vector. Thus $C\widehat{y} = 0$ and $\widehat{y}$ satisfies KKT conditions (5.1). Therefore $\widehat{y} = \widehat{x}$. ◻

Lemma 5.1 shows that it is necessary to include the stopping criterion in the procedure which implements the inner loop. It is also clear that there is no hidden enforcement of exact solution in (5.2), and typically inexact solutions of the auxiliary unconstrained problems are obtained in Step 1.

6. Inequalities involving augmented Lagrangians. In this section we establish basic inequalities which relate the bound on the norm of the K-gradient g^K of the augmented Lagrangian L to the values of L. These inequalities will be the key ingredients in the proof of convergence and other analysis concerning Algorithm 5.1. Similar inequalities were proved in [7, 9] for the projected gradient and bound constraints. Our generalization is possible due to the following lemma.

LEMMA 6.1. *Let $x, y \in \Omega_S$, $\lambda \in \mathbb{R}^{m_E}$, and $\rho > 0$. It holds that*

$$(y - x)^\top g(x, \lambda, \rho) \geq (y - x)^\top g^K(x, \lambda, \rho). \tag{6.1}$$

Proof. Let $\boldsymbol{g}_i$ and $\boldsymbol{g}_i^K$ be the ith segments of $g = g(x, \lambda, \rho)$ and $g^K = g^K(x, \lambda, \rho)$, respectively. First notice that for any $i \in \mathcal{F}(x)$, we have by definition $\boldsymbol{g}_i^K = \boldsymbol{g}_i$. If $i \in \mathcal{A}(x)$, then $\boldsymbol{g}_i^K = \boldsymbol{\beta}_i$, $f_i(\boldsymbol{x}_i) = 0$, $f_i(\boldsymbol{y}_i) \leq 0$, and by convexity of f_i

$$(\nabla f_i(\boldsymbol{x}_i))^\top (\boldsymbol{y}_i - \boldsymbol{x}_i) \leq f_i(\boldsymbol{y}_i) - f_i(\boldsymbol{x}_i).$$

Thus we get

$$\begin{aligned}
(\boldsymbol{y}_i - \boldsymbol{x}_i)^\top \boldsymbol{g}_i^K &= (\boldsymbol{y}_i - \boldsymbol{x}_i)^\top \left(\boldsymbol{g}_i + \frac{\|\boldsymbol{g}_i\|}{\|\nabla f_i(\boldsymbol{x}_i)\|}\nabla f_i(\boldsymbol{x}_i)\right) \\
&= (\boldsymbol{y}_i - \boldsymbol{x}_i)^\top \boldsymbol{g}_i + \frac{\|\boldsymbol{g}_i\|}{\|\nabla f_i(\boldsymbol{x}_i)\|} (\nabla f_i(\boldsymbol{x}_i))^\top (\boldsymbol{y}_i - \boldsymbol{x}_i) \\
&\leq (\boldsymbol{y}_i - \boldsymbol{x}_i)^\top \boldsymbol{g}_i + \frac{\|\boldsymbol{g}_i\|}{\|\nabla f_i(\boldsymbol{x}_i)\|}\big(f_i(\boldsymbol{y}_i) - f_i(\boldsymbol{x}_i)\big) \\
&\leq (\boldsymbol{y}_i - \boldsymbol{x}_i)^\top \boldsymbol{g}_i.
\end{aligned}$$

Finally,

$$(y-x)^\top g(x,\lambda,\rho) = \sum_{i=1}^{m_I} (\boldsymbol{y}_i - \boldsymbol{x}_i)^\top \boldsymbol{g}_i \geq \sum_{i=1}^{m_I} (\boldsymbol{y}_i - \boldsymbol{x}_i)^\top \boldsymbol{g}_i^K = (y-x)^\top g^K(x,\lambda,\rho). \qquad \square$$

Now we are ready to prove the relations that are relevant in what follows.

LEMMA 6.2. *Let $x, y \in \Omega_S$, $\lambda \in \mathbb{R}^{m_E}$, $\rho > 0$, $\eta > 0$, and $M > 0$. Let $\lambda_{\min}$ denote the smallest eigenvalue of A and $\widetilde{\lambda} = \lambda + \rho C x$.*

(i) *If*

$$\|g^K(x,\lambda,\rho)\| \leq M\|Cx\|, \tag{6.2}$$

then

$$L(y,\widetilde{\lambda},\rho) \geq L(x,\lambda,\rho) + \frac{1}{2}\left(\rho - \frac{M^2}{\lambda_{\min}}\right)\|Cx\|^2 + \frac{\rho}{2}\|Cy\|^2. \tag{6.3}$$

(ii) *If*

$$\|g^K(x,\lambda,\rho)\| \leq \eta, \tag{6.4}$$

then

$$L(y,\widetilde{\lambda},\rho) \geq L(x,\lambda,\rho) + \frac{\rho}{2}\|Cx\|^2 + \frac{\rho}{2}\|Cy\|^2 - \frac{\eta^2}{2\lambda_{\min}}. \tag{6.5}$$

(iii) *If* (6.4) *holds and $z_0 \in \Omega$, then*

$$L(x,\lambda,\rho) \leq q(z_0) + \frac{\eta^2}{2\lambda_{\min}}. \tag{6.6}$$

Proof. Let us denote $\delta = y - x$ and $A_\rho = A + \rho C^\top C$. Using (6.1),

$$L(x,\widetilde{\lambda},\rho) = L(x,\lambda,\rho) + \rho\|Cx\|^2, \quad \text{and} \quad g(x,\widetilde{\lambda},\rho) = g(x,\lambda,\rho) + \rho C^\top C x,$$

we get

$$\begin{aligned} L(y,\widetilde{\lambda},\rho) &= L(x,\widetilde{\lambda},\rho) + \delta^\top g(x,\widetilde{\lambda},\rho) + \frac{1}{2}\delta^\top A_\rho \delta \\ &= L(x,\lambda,\rho) + \delta^\top g(x,\lambda,\rho) + \frac{1}{2}\delta^\top A_\rho\delta + \rho\delta^\top C^\top C x + \rho\|Cx\|^2 \\ &\geq L(x,\lambda,\rho) + \delta^\top g^K(x,\lambda,\rho) + \frac{1}{2}\delta^\top A_\rho\delta + \rho\delta^\top C^\top C x + \rho\|Cx\|^2 \\ &\geq L(x,\lambda,\rho) + \delta^\top g^K(x,\lambda,\rho) + \frac{\lambda_{\min}}{2}\|\delta\|^2 + \frac{\rho}{2}\|C\delta\|^2 + \rho\delta^\top C^\top C x + \rho\|Cx\|^2. \end{aligned}$$

Noticing that

$$\frac{\rho}{2}\|Cy\|^2 = \frac{\rho}{2}\|C(\delta + x)\|^2 = \rho\delta^\top C^\top C x + \frac{\rho}{2}\|C\delta\|^2 + \frac{\rho}{2}\|Cx\|^2,$$

we get

$$L(y,\widetilde{\lambda},\rho) \geq L(x,\lambda,\rho) + \delta^\top g^K(x,\lambda,\rho) + \frac{\lambda_{\min}}{2}\|\delta\|^2 + \frac{\rho}{2}\|Cx\|^2 + \frac{\rho}{2}\|Cy\|^2. \tag{6.7}$$

Using (6.2) and simple manipulations, we obtain

$$
\begin{aligned}
L(y,\widetilde{\lambda},\rho) &\geq L(x,\lambda,\rho) - M\|\delta\|\|Cx\| + \frac{\lambda_{\min}}{2}\|\delta\|^2 + \frac{\rho}{2}\|Cx\|^2 + \frac{\rho}{2}\|Cy\|^2 \\
&= L(x,\lambda,\rho) + \left(\frac{\lambda_{\min}}{2}\|\delta\|^2 - M\|\delta\|\|Cx\| + \frac{M^2\|Cx\|^2}{2\lambda_{\min}}\right) \\
&\quad - \frac{M^2\|Cx\|^2}{2\lambda_{\min}} + \frac{\rho}{2}\|Cx\|^2 + \frac{\rho}{2}\|Cy\|^2 \\
&\geq L(x,\lambda,\rho) + \frac{1}{2}\left(\rho - \frac{M^2}{\lambda_{\min}}\right)\|Cx\|^2 + \frac{\rho}{2}\|Cy\|^2.
\end{aligned}
$$

This proves (i).

If we assume that (6.4) holds, then by (6.7)

$$
\begin{aligned}
L(y,\widetilde{\lambda},\rho) &\geq L(x,\lambda,\rho) - \|\delta\|\eta + \frac{\lambda_{\min}}{2}\|\delta\|^2 + \frac{\rho}{2}\|Cx\|^2 + \frac{\rho}{2}\|Cy\|^2 \\
&\geq L(x,\lambda,\rho) + \frac{\rho}{2}\|Cx\|^2 + \frac{\rho}{2}\|Cy\|^2 \; - \frac{\eta^2}{2\lambda_{\min}}.
\end{aligned}
$$

This proves (ii).

Finally, let $\widehat{z}$ denote the solution of the auxiliary problem

$$
\text{minimize} \quad L(z,\lambda,\rho) \quad \text{subject to} \quad z \in \Omega_S, \tag{6.8}
$$

let $z_0 \in \Omega$ so that $Cz_0 = 0$, and let $\widehat{\delta} = \widehat{z} - x$. If (6.4) holds, then we derive using (6.1) that

$$
\begin{aligned}
0 &\geq L(\widehat{z},\lambda,\rho) - L(x,\lambda,\rho) = \widehat{\delta}^\top g(x,\lambda,\rho) + \frac{1}{2}\widehat{\delta}^\top A_\rho \widehat{\delta} \\
&\geq \widehat{\delta}^\top g^K(x,\lambda,\rho) + \frac{1}{2}\widehat{\delta}^\top A_\rho\widehat{\delta} \geq -\|\widehat{\delta}\|\eta + \frac{1}{2}\lambda_{\min}\|\widehat{\delta}\|^2 \geq -\frac{\eta^2}{2\lambda_{\min}}.
\end{aligned}
$$

Since $L(\widehat{z},\lambda,\rho) \leq L(z_0,\lambda,\rho) = q(z_0)$, we conclude that

$$
L(x,\lambda,\rho) \leq L(x,\lambda,\rho) - L(\widehat{z},\lambda,\rho) + q(z_0) \leq q(z_0) + \frac{\eta^2}{2\lambda_{\min}}. \qquad \square
$$

7. Monotonicity and feasibility. Now we shall translate the results on the relations that are satisfied by the augmented Lagrangian into the relations concerning the iterates generated by Algorithm 5.1 (SMALSE-M).

LEMMA 7.1. *Let* $\{x^k\}, \{\lambda^k\}$*, and* $\{M_k\}$ *be generated by Algorithm* 5.1 *for the solution of* (1.1) *with* $\eta > 0,\ \beta > 1,\ M_0 > 0,\ \rho > 0,$ and $\lambda^0 \in \mathbb{R}^{m_E}$*. Let* $\lambda_{\min}$ *denote the smallest eigenvalue of the Hessian* A *of the quadratic function* q*.*

(i) *If* $k \geq 0$ *and*

$$
M_k^2 \leq \rho\lambda_{\min}, \tag{7.1}
$$

then

$$
L(x^{k+1},\lambda^{k+1},\rho) \geq L(x^k,\lambda^k,\rho) + \frac{\rho}{2}\|Cx^{k+1}\|^2. \tag{7.2}
$$

(ii) *For any* $k \geq 0$

$$L(x^{k+1}, \lambda^{k+1}, \rho) \geq L(x^k, \lambda^k, \rho) + \frac{\rho}{2}\|Cx^k\|^2 + \frac{\rho}{2}\|Cx^{k+1}\|^2 - \frac{\eta^2}{2\lambda_{\min}}. \tag{7.3}$$

(iii) *For any* $k \geq 0$ *and* $z_0 \in \Omega$

$$L(x^k, \lambda^k, \rho) \leq q(z_0) + \frac{\eta^2}{2\lambda_{\min}}. \tag{7.4}$$

Proof. In Lemma 6.2, let us substitute $x = x^k, \lambda = \lambda^k, M = M_k$, and $y = x^{k+1}$, so that inequality (6.2) holds by (5.2), and by (5.3) $\widetilde{\lambda} = \lambda^{k+1}$.

If (7.1) holds, we get by (6.3) that

$$L(x^{k+1}, \lambda^{k+1}, \rho) \geq L(x^k, \lambda^k, \rho) + \frac{\rho}{2}\|Cx^{k+1}\|^2, \tag{7.5}$$

so that (7.2) holds.

Since by the definition of Step 1 of Algorithm 5.1

$$\|g^K(x^k, \lambda^k, \rho)\| \leq \eta,$$

we can apply the same substitution as above to Lemma 6.2(ii) to get

$$L(x^{k+1}, \lambda^{k+1}, \rho) \geq L(x^k, \lambda^k, \rho) + \frac{\rho}{2}\|Cx^k\|^2 + \frac{\rho}{2}\|Cx^{k+1}\|^2 - \frac{\eta^2}{2\lambda_{\min}}. \tag{7.6}$$

This proves (7.3). Similarly, inequality (7.4) results from application of the substitution to Lemma 6.2(iii). ∎

THEOREM 7.2. *Let* $\{x^k\}, \{\lambda^k\}$, *and* $\{M_k\}$ *be generated by Algorithm* 5.1 *for the solution of* (1.1) *with* $\eta > 0$, $\beta > 1$, $M_0 > 0$, $\rho > 0$, *and* $\lambda^0 \in \mathbb{R}^{m_E}$. *Let* $\lambda_{\min}$ *denote the smallest eigenvalue of the Hessian* A *of the cost function* q, *and let* $s \geq 0$ *denote the smallest integer such that*

$$M_0^2 \leq \rho \lambda_{\min} \beta^{2s}. \tag{7.7}$$

Then the following statements hold.

(i) *The sequence* $\{M_k\}$ *satisfies*

$$M_0/\beta^s \leq M_k. \tag{7.8}$$

(ii) *If* $z_0 \in \Omega$, *then*

$$\sum_{k=1}^{\infty} \frac{\rho}{2}\|Cx^k\|^2 \leq q(z_0) - L(x^0, \lambda^0, \rho) + (1+s)\frac{\eta^2}{2\lambda_{\min}}. \tag{7.9}$$

Proof. Let $s \geq 0$ denote the smallest integer which satisfies (7.7), and let $\mathcal{I} \subseteq \{1, 2, \dots\}$ denote the possibly empty set of the indices k_i such that $M_{k_i} < M_{k_i - 1}$. Using Lemma 7.1(i), $M_{k_i} = M_{k_i-1}/\beta = M_0/\beta^i$ for $k_i \in \mathcal{I}$, and (5.4), we conclude that there is no k such that $M_k < M_0/\beta^s$. Thus $\mathcal{I}$ has at most s elements and (7.8) holds.

By the definition of Step 3 of Algoritm 5.1, if $k > 0$, then either $k \notin \mathcal{I}$ and

$$\frac{\rho}{2}\|Cx^k\|^2 \leq L(x^k, \lambda^k, \rho) - L(x^{k-1}, \lambda^{k-1}, \rho),$$

or $k \in \mathcal{I}$ and by (7.3)

$$\frac{\rho}{2}\|Cx^k\|^2 \le \frac{\rho}{2}\|Cx^{k-1}\|^2 + \frac{\rho}{2}\|Cx^k\|^2 \le L(x^k,\lambda^k,\rho) - L(x^{k-1},\lambda^{k-1},\rho) + \frac{\eta^2}{2\lambda_{\min}}. \tag{7.10}$$

Summing up the appropriate cases of the last two inequalities for $k = 1,\dots,j$ and taking into account that $\mathcal{I}$ has at most s elements, we get

$$\sum_{k=1}^{j} \frac{\rho}{2}\|Cx^k\|^2 \le L(x^j,\lambda^j,\rho) - L(x^0,\lambda^0,\rho) + s\frac{\eta^2}{2\lambda_{\min}}. \tag{7.11}$$

To get (7.9), it is enough to replace $L(x^j,\lambda^j,\rho)$ by the upper bound (7.4). ◻

8. Optimality of the outer loop. The aim of this section is to give an upper bound on the number of outer iterations of Algorithm 5.1 for a fixed tolerance level $\varepsilon > 0$ that does not depend (under certain assumptions) on the size of the problem. Theorem 7.2 suggests that this bound may also be independent of C. To present this feature explicitly, let $\mathcal{T}$ denote any set of indices and let for any $t \in \mathcal{T}$ be defined a problem

$$\text{minimize} \quad q_t(x) \quad \text{subject to} \quad x \in \Omega^t \tag{8.1}$$

with $q_t(x) = 1/2x^\top A_t x - b_t^\top x$, $A_t \in \mathbb{R}^{n_t\times n_t}$ symmetric positive definite, $b_t \in \mathbb{R}^{n_t}$, and

$$\Omega^t = \{x \in \mathbb{R}^{n_t} : C_t x = 0,\ f_1^t(\boldsymbol{x}_1) \le 0, \dots,\ f^t_{m_I^t}(\boldsymbol{x}_{m_I^t}) \le 0\},$$

where $C_t \in \mathbb{R}^{m_E^t \times n_t}$, $f_i^t : \mathbb{R}^{n_i^t} \mapsto \mathbb{R}$ are continuously differentiable convex functions, and $\boldsymbol{x}_i \in \mathbb{R}^{n_i^t}$ denotes the ith segment of $x_t \in \mathbb{R}^{n_t}$ so that $x_t = x = (\boldsymbol{x}_1^\top, \dots, \boldsymbol{x}_{m_I^t}^\top)^\top$, $\sum_{i=1}^{m_I^t} n_i^t = n_t$. Finally, we assume that there is a positive constant c_0 and that for any $t \in \mathcal{T}$ we are able to find

$$x_t^{-1} \in \Omega_S^t, \quad \Omega_S^t = \{x \in \mathbb{R}^{n_t} :\ f_1^t(\boldsymbol{x}_1) \le 0, \dots,\ f^t_{m_I^t}(\boldsymbol{x}_{m_I^t}) \le 0\},$$

such that

$$\|x_t^{-1}\| \le c_0\|b_t\|. \tag{8.2}$$

Let us point out that the last assumption is essential, as we must have for each problem an initial approximation which is not too far from the solution. For example, for bound and equality constrained problems, where the constraints $x_i \ge \ell_i^t$, $i = 1, \dots n_t$, are defined by $f_i^t(x_i) = \ell_i^t - x_i$, our assumption is satisfied when

$$\|\max\{\ell^t, 0\}\| \le c_0\|b_t\|, \tag{8.3}$$

where $\max\{\ell^t, 0\}$ denotes the vector with the entries $\max\{\ell_i^t, 0\}$. Notice that for unconstrained problems, we can always start from the zero vector, so that similar assumptions are not necessary. Our optimality result reads as follows.

THEOREM 8.1. *Let $\{x_t^k\}, \{\lambda_t^k\}$, and $\{M_{t,k}\}$ be generated by Algorithm* 5.1 *for* (8.1) *with $\|b_t\| \ge \eta_t > 0$, $\beta > 1$, $\rho > 0$, $M_{t,0} = M_0 > 0$, and $\lambda_t^0 = 0$. Let there be*

$a_{\min} > 0$ *such that the smallest eigenvalue* $\lambda_{\min}(A_t)$ *of the Hessian* A_t *of the quadratic function* q_t *satisfies*

$$\lambda_{\min}(A_t) \geq a_{\min},$$

let $s \geq 0$ *denote the smallest integer such that*

$$M_0^2 \leq \beta^{2s} a_{\min}\rho,$$

and denote

$$a = \frac{c+2+s}{a_{\min}\rho}, \quad c = a_{\min}(a_{\max}c_0^2 + 2c_0),$$

where $a_{\max} \geq \lambda_{\max}(A_t)$.

Then for each $\varepsilon > 0$ *there are indices* k_t, $t \in \mathcal{T}$, *such that*

$$k_t \leq a/\varepsilon^2 + 1 \tag{8.4}$$

and $x_t^{k_t}$ *is an approximate solution of* (1.1) *satisfying*

$$\|g_t^K(x_t^{k_t}, \lambda_t^{k_t}, \rho)\| \leq M_0\varepsilon\|b_t\| \quad \textit{and} \quad \|C_t x_t^{k_t}\| \leq \varepsilon\|b_t\|. \tag{8.5}$$

Moreover,

$$\sum_{i=1}^{\infty} \|C_t x_t^i\|^2 \leq \frac{c+2+s}{\rho a_{\min}}\|b_t\|^2 \tag{8.6}$$

and

$$M_0/\beta^s \leq M_{t,k}. \tag{8.7}$$

Proof. First notice that for any index j

$$\frac{\rho j}{2}\min\{\|C_t x_t^i\|^2 : \ i = 1, \dots, j\} \leq \sum_{i=1}^{j} \frac{\rho}{2}\|C_t x_t^i\|^2 \leq \sum_{i=1}^{\infty} \frac{\rho}{2}\|C_t x_t^i\|^2. \tag{8.8}$$

Denoting by $L_t(x, \lambda, \rho)$ the augmented Lagrangian for problem (8.1), we get for any $x \in \mathbb{R}^{n_t}$ and $\rho > 0$

$$L_t(x, 0, \rho) = \frac{1}{2}x^\top(A_t + \rho C_t^\top C_t)x - b_t^\top x \geq \frac{1}{2}a_{\min}\|x\|^2 - \|b_t\|\|x\| \geq -\frac{\|b_t\|^2}{2a_{\min}}. \tag{8.9}$$

Similarly, using (8.2) and the definition of c, we get

$$q_t(x_t^{-1}) \leq \frac{1}{2}a_{\max}\|x_t^{-1}\|^2 + \|b_t\|\|x_t^{-1}\| \leq \left(\frac{a_{\max}}{2}c_0^2 + c_0\right)\|b_t\|^2 = \frac{c}{2a_{\min}}\|b_t\|^2. \tag{8.10}$$

If we substitute $z_0 = x_t^{-1}$ into (7.9) and use (8.9), (8.10), and the assumption $\eta_t \leq \|b_t\|$, we get

$$\sum_{i=1}^{\infty} \frac{\rho}{2}\|C_t x_t^i\|^2 \leq \frac{c}{2a_{\min}}\|b_t\|^2 + \frac{1}{2a_{\min}}\|b_t\|^2 + (1+s)\frac{\eta_t^2}{2a_{\min}} \leq \frac{c+2+s}{2a_{\min}}\|b_t\|^2.$$

This proves (8.6). Using (8.8), the latter inequality, and the definition of a, we get

$$\frac{\rho j}{2}\min\{\|C_t x_t^i\|^2 : i = 1, \ldots, j\} \leq \frac{c+2+s}{2a_{\min}\varepsilon^2}\varepsilon^2\|b_t\|^2 = \frac{\rho a}{2\varepsilon^2}\varepsilon^2\|b_t\|^2.$$

Taking for j the smallest integer which satisfies $a/j \leq \varepsilon^2$, so that

$$a/\varepsilon^2 \leq j \leq a/\varepsilon^2 + 1,$$

and denoting for any $t \in \mathcal{T}$ by $k_t \in \{1, \ldots, j\}$ the index which minimizes $\{\|C_t x_t^i\| : i = 1, \ldots, j\}$, we can use the last inequality with simple manipulations to obtain

$$\|C_t x_t^{k_t}\|^2 = \min\{\|C_t x_t^i\|^2 : i = 1, \ldots, j\} \leq \frac{a}{j\varepsilon^2}\varepsilon^2\|b_t\|^2 \leq \varepsilon^2\|b_t\|^2.$$

The inequality

$$M_0^{-1}\|g_t^K(x_t^{k_t}, \lambda_t^{k_t}, \rho)\| \leq \|C_t x_t^{k_t}\| \leq \varepsilon\|b_t\|$$

results easily from the definition of Step 1 of Algorithm 5.1 and $M_{t,k_t} \leq M_0$. Inequality (8.7) follows by Theorem 7.2(i). ∎

Having proved that there is a bound on the number of outer iterations of SMALSE-M that is necessary to get an approximate solution, it remains to bound the number of inner iterations. In the next section, we consider implementation of the inner loop by Algorithm 3.1 and give sufficient conditions which guarantee that the number of inner iterations is bounded.

9. Optimality of the inner loop. In this section we shall prove that also the total number of inner iterations of Algorithm 5.1, i.e., the number of the matrix-vector multiplications, do not depend on the size of the problem. We need the following simple lemma to prove optimality of the inner loop implemented by the generalized KPRGP algorithm (Algorithm 3.1).

LEMMA 9.1. *Let* $\{x^k\}$, $\{\lambda^k\}$, *and* $\{M_k\}$ *be generated by Algorithm* 5.1 *for the solution of* (1.1) *with* $\eta > 0$, $\beta > 1$, $M_0 > 0$, $\rho > 0$, *and* $\lambda^0 \in \mathbb{R}^{m_E}$. *Let* $0 < a_{\min} \leq \lambda_{\min}$, *where* $\lambda_{\min}$ *denotes the smallest eigenvalue of* A. *Then for any* $k \geq 0$

$$L(x^k, \lambda^{k+1}, \rho) - L(x^{k+1}, \lambda^{k+1}, \rho) \leq \frac{\eta^2}{2a_{\min}} + \frac{\rho}{2}\|Cx^k\|^2. \tag{9.1}$$

Proof. Notice that by the definition of the Lagrangian function

$$L(x^k, \lambda^{k+1}, \rho) = L(x^k, \lambda^k, \rho) + \rho\|Cx^k\|^2,$$

so that by (7.3)

$$\begin{aligned} L(x^k, \lambda^{k+1}, \rho) - L(x^{k+1}, \lambda^{k+1}, \rho) &= L(x^k, \lambda^k, \rho) - L(x^{k+1}, \lambda^{k+1}, \rho) + \rho\|Cx^k\|^2 \\ &\leq \frac{\eta^2}{2a_{\min}} + \frac{\rho}{2}\|Cx^k\|^2. \qquad \square \end{aligned}$$

Now we are ready to prove the main result of this section, the optimality of Algorithm 5.1 (SMALSE-M) in terms of matrix-vector multiplications, provided Step 1 of Algorithm 5.1 is implemented by Algorithm 3.1 (generalized KPRGP).

THEOREM 9.2. *Let*

$$0 < \varepsilon < 1, \quad 0 < a_{\min} < a_{\max}, \quad \textit{and} \quad 0 < c_{\max}$$

be given constants, and let the class of problems (8.1) *satisfy*

$$a_{\min} \le \lambda_{\min}(A_t) \le \lambda_{\max}(A_t) \le a_{\max} \quad \textit{and} \quad \|C_t\| \le c_{\max}. \tag{9.2}$$

Let $\{x_t^k\}, \{\lambda_t^k\}$, *and* $\{M_{t,k}\}$ *be generated by Algorithm* 5.1 *(SMALSE-M) for* (8.1) *with*

$$\|b_t\| \ge \eta_t \ge \varepsilon\|b_t\|, \ \beta > 1, \ \rho > 0, \ M_{t,0} = M_0 > 0, \quad \textit{and} \quad \lambda_t^0 = 0.$$

Let $s \ge 0$ *denote the smallest integer such that*

$$M_0^2 \le \beta^{2s} a_{\min}\rho,$$

and let Step 1 *of Algorithm* 5.1 *be implemented by a* K*-convergent algorithm in order to generate the iterates* $x_t^{k,0}, x_t^{k,1}, \dots, x_t^{k,l} = x_t^k$ *for the solution of* (8.1) *starting from* $x_t^{k,0} = x_t^{k-1}$ *with* x_t^{-1} *which satisfies* (8.2), *where* $l = l_{t,k}$ *is the first index satisfying*

$$\|g_t^K(x_t^{k,l}, \lambda_t^k, \rho)\| \le \min\{M_{t,k}\|C_t x_t^{k,l}\|, \eta_t\}. \tag{9.3}$$

Then Algorithm 5.1 *generates an approximate solution* $x_t^{k_t}$ *of any problem* (8.1) *which satisfies*

$$\|g_t^K(x_t^{k_t}, \lambda_t^{k_t}, \rho)\| \le M_0\varepsilon\|b_t\| \quad \textit{and} \quad \|C_t x_t^{k_t}\| \le \varepsilon\|b_t\| \tag{9.4}$$

at $O(1)$ *inner iterations.*

Proof. Let $t \in \mathcal{T}$ be fixed, and let us denote by $L_t(x, \lambda, \rho)$ the augmented Lagrangian for problem (8.1). Using Theorem 8.1, we get

$$\|C_t x_t^k\|^2 \le \sum_{i=1}^{\infty} \|C_t x_t^i\|^2 \le \frac{c+2+s}{\rho a_{\min}} \|b_t\|^2$$

for any $k \ge 0$. Thus by (9.1)

$$\begin{aligned} L_t(x_t^{k-1}, \lambda_t^k, \rho) - L_t(x_t^k, \lambda_t^k, \rho) &\le \frac{\eta_t^2}{2a_{\min}} + \frac{\rho}{2}\|C_t x_t^{k-1}\|^2 \\ &\le \frac{c+3+s}{2a_{\min}}\|b_t\|^2, \end{aligned}$$

and since the minimizer $\overline{x}_t^k$ of $L_t(x, \lambda_t^k, \rho)$ subject to $x \in \Omega_S^t$ satisfies (5.2) and is a possible choice for x_t^k, it follows that

$$L_t(x_t^{k-1}, \lambda_t^k, \rho) - L_t(\overline{x}_t^k, \lambda_t^k, \rho) \le \frac{c+3+s}{2a_{\min}}\|b_t\|^2. \tag{9.5}$$

Now recall that any K-convergent algorithm starting from $x_t^{k,0} = x_t^{k-1}$ generates $x_t^{k,l}$ satisfying

$$\|g_t^K(x_t^{k,l}, \lambda_t^k, \rho)\|^2 \le \bar{\kappa}\nu^l \big(L_t(x_t^{k-1}, \lambda_t^k, \rho) - L_t(\overline{x}_t^k, \lambda_t^k, \rho)\big) \le \bar{\kappa}\nu^l \frac{c+3+s}{2a_{\min}}\|b_t\|^2, \tag{9.6}$$

where $\bar{\kappa} > 0$ and $\nu \in (0,1)$ are independent of t, and let $l_{\max}$ denote any index which satisfies

$$\nu^{l_{\max}} \bar{\kappa} \frac{c+3+s}{2a_{\min}} \le \min\{\varepsilon^2, M_0^2 \beta^{-2s} \varepsilon^2\}. \tag{9.7}$$

Let us assume that the inner loop is not completed in $l_{\max}$ steps, so that

$$\|g_t^K(x_t^{k,l_{\max}}, \lambda_t^k, \rho)\| > \min\{M_{t,k} \|C_t x_t^{k,l_{\max}}\|, \eta_t\}, \tag{9.8}$$

and notice that by (9.6), (9.7), and the assumptions of Theorem 9.2

$$\|g_t^K(x_t^{k,l_{\max}}, \lambda_t^k, \rho)\| \le \varepsilon \|b_t\| \le \eta_t \quad \text{and} \quad \|g_t^K(x_t^{k,l_{\max}}, \lambda_t^k, \rho)\| \le \varepsilon M_0 \beta^{-s} \|b_t\|. \tag{9.9}$$

Using (8.7), (9.8), and (9.9), we get

$$M_0 \beta^{-s} \|C_t x_t^{k,l_{\max}}\| \le M_{t,k} \|C_t x_t^{k,l_{\max}}\| < \|g_t^K(x_t^{k,l_{\max}}, \lambda_t^k, \rho)\| \le \varepsilon M_0 \beta^{-s} \|b_t\|, \tag{9.10}$$

i.e.,

$$\|C_t x_t^{k,l_{\max}}\| \le \varepsilon \|b_t\|,$$

and using the second inequality in (9.9), we get

$$\|g_t^K(x_t^{k,l_{\max}}, \lambda_t^k, \rho)\| \le \varepsilon M_0 \|b_t\|.$$

We conclude that the inner loop is either completed fewer than $l_{\max}$ steps, or $x_t^{k,l_{\max}}$ satisfies (9.4). If we combine the inequality (9.7) with Theorem 8.1, we get that the total number of inner iterations of Algorithm 5.1 for the solution of (8.1) for any $t \in \mathcal{T}$ is bounded by $(a/\varepsilon^2 + 1) l_{\max}$. ∎

The assumption that $\|C_t\|$ is bounded is essential; it guarantees that the spectrum of the Hessian of the Lagrangian is uniformly bounded. There is a simple result following from Theorem 9.2 and Corollary 4.3, which is a key to developing scalable algorithms for contact problems with the Tresca friction.

COROLLARY 9.3. *Under the assumptions of Theorem* 9.2*, let Step* 1 *of Algorithm* 5.1 *be implemented by Algorithm* 3.1 *(generalized KPRGP) with the parameters* $\Gamma > 0$ *and*

$$\overline{\alpha} \in \left(0, (a_{\max} + \rho c_{\max}^2)^{-1}\right]$$

in order to generate the iterates $x_t^{k,0}, x_t^{k,1}, \ldots, x_t^{k,l} = x_t^k$ *for the solution of* (8.1) *starting from* $x_t^{k,0} = x_t^{k-1}$ *with* x_t^{-1} *which satisfies* (8.2)*, where* $l = l_{t,k}$ *is the first index satisfying* (9.3)*.*

Then Algorithm 5.1 *generates an approximate solution* $x_t^{k_t}$ *of any problem* (8.1) *which satisfies* (9.4) *at* $O(1)$ *matrix-vector multiplications by the Hessian of the augmented Lagrangian* L_t *for* (8.1)*.*

Proof. Notice that each step of Algorithm 3.1 (KPRGP) requires at most two matrix-vector multiplications. The rest follows by Theorem 8.1. ∎

We conclude our analysis with the specialization of Theorem 9.2, which is useful for the development of scalable algorithms for frictionless contact problems.

COROLLARY 9.4. *Let the separable constraints in* (8.1) *be bound constraints. Under the assumptions of Theorem* 9.2*, let Step* 1 *of Algorithm* 5.1 *be implemented by the*

MPRGP algorithm of [15, 10] *with the parameters* $\Gamma > 0$ *and* $\overline{\alpha} \in \big(0, 2(a_{\max} + \rho c_{\max}^2)^{-1}\big)$ *in order to generate the iterates* $x_t^{k,0}, x_t^{k,1}, \ldots, x_t^{k,l} = x_t^k$ *for the solution of* (8.1) *starting from* $x_t^{k,0} = x_t^{k-1}$ *with* x_t^{-1} *which satisfies* (8.2), *where* $l = l_{t,k}$ *is the first index satisfying* (9.3).

Then Algorithm 5.1 *generates an approximate solution* $x_t^{k_t}$ *of any problem* (8.1), *which satisfies* (9.4) *at* $O(1)$ *matrix-vector multiplications by the Hessian of the augmented Lagrangian* L_t *for* (8.1).

Proof. It is proved in [8] that the MPRGP algorithm enjoys R-linear convergence of the projected gradient for $\overline{\alpha} \in \big(0, (a_{\max} + \rho c_{\max}^2)^{-1}\big)$ and $\Gamma > 0$. The generalization to

$$\overline{\alpha} \in \big(0, 2(a_{\max} + \rho c_{\max}^2)^{-1}\big)$$

is in [10]. It simply follows by equivalence relations (2.11) that MPRGP is K-convergent with the choice of parameters specified in our lemma. The statement follows by Theorem 9.2. ▯

10. Numerical evidence of optimality. Various versions of the algorithm presented here have already been widely tested on the solution of the quadratic programming problems with bound and equality constraints arising from the application of the FETI methods to the solution of boundary elliptic variational inequalities. The first optimality results were obtained for scalar variational inequalities; see Dostál [7]. Application to frictionless contact problems of elasticity can be found in Dostál et al. [14] or Dostál et al. [16]. Application to the problems discretized by the BETI methods is in Bouchala, Dostál, and Sadowská [3]. The results concerning the solution of two-dimensional and three-dimensional contact problems with friction by the algorithms which reduce these problems to the bound and equality constrained quadratic programming problems are in Dostál, Haslinger, and Kučera [11] and Haslinger, Kučera, and Dostál [25], respectively.

Here we consider the class of problems with the quadratic functions q_t, $t \in \mathcal{T} := \{11, \ldots, 20\}$, defined by the tridiagonal Hessians $A_t = (a_{ij}^t) \in \mathbb{R}^{n_t \times n_t}$, $a_{ii}^t = 4$, $a_{i,i+1}^t = a_{i-1,i}^t = -1$, $a_{ij}^t = 0$ for $|i-j| > 1$, and by $b_t = A_t y_t$, where $y_t = (y_i^t) \in \mathbb{R}^{n_t}$ is the auxiliary vector with $y_i^t = -5\tau_i^2 \sin\tau_i$, $y_{i+\frac{1}{2}n_t}^t = -\tau_i \sin\tau_i$, $\tau_i = (i-1)h_t$, $h_t = 2\pi/(\frac{1}{2}n_t - 1)$, $i = 1, \ldots, \frac{1}{2}n_t$, $n_t = 2^t$. The equality constraint matrix $C_t = (c_{ij}^t) \in \mathbb{R}^{\frac{1}{2}n_t \times n_t}$ has two nonzero entries in the ith row that are $c_{i,2i-1}^t = -1$ and $c_{i,2i-1+\frac{1}{2}n}^t = 1$. Finally, we consider the following inequality constraints:

$$x_{i+\frac{1}{2}n_t} \geq -0.7, \quad x_{i+\frac{1}{4}n_t}^2 + x_{i+\frac{3}{2}n_t}^2 \leq 10^2, \quad i = 1, \ldots, \frac{1}{4}n_t.$$

It is easy to see that a permutation of unknowns transforms these problems in the form of (1.1). Moreover, using the Gershgorin theorem, it can be checked that the assumptions for our analysis performed in the previous sections are satisfied, namely, $a_{\min} = 2$, $a_{\max} = 6$, and $c_{\max} = 2$. The matrix A is well conditioned to mimic the well-conditioned problems arising from the FETI or BETI method.

We test Algorithm 5.1 (SMALSE-M) that uses in the inner loop Algorithm 3.1 (KPRGP). The implementation is carried out in MATLAB 7 as the part of the MatSol system [28]. We use the following choice of the parameters: $\overline{\alpha} = 2$, $\rho = 50$, $\Gamma = 1$, $M_0 = 100$, $\eta = 0.01$, $\beta = 0.1$. The terminating tolerance is given by $\varepsilon = 10^{-6}$.

In Table 10.1 we report results of our experiments for which we check the numbers of the active/free simple bounds $n_{\mathcal{A}_b}/n_{\mathcal{F}_b}$ and the active/free quadratic constraints $n_{\mathcal{A}_q}/n_{\mathcal{F}_q}$ at the solution. Further, n_{exp}, n_{prop}, and n_{cg} are the numbers of the expansion steps, the proportioning steps, and the conjugate gradient steps, respectively, n_M is the number of updates of balancing ratio $M_{t,k}$, n_A is the total number of the matrix-vector multiplications, n_{it} is the number of outer (SMALSE-M) iterations, and *rel_eff* denotes the relative efficiency defined by

$$rel\_eff = n_A/n.$$

TABLE 10.1
Experiments with simple bounds and quadratic constraints.

n_t	$n_{\mathcal{A}_b}/n_{\mathcal{F}_b} : n_{\mathcal{A}_q}/n_{\mathcal{F}_q}$	n_{exp}	n_{prop}	n_{cg}	n_M	n_A	n_{it}	*rel_eff*
2048	94/418:220/292	65	156	63	2	366	11	0.1787
4096	187/837:439/585	101	159	73	2	451	11	0.1101
8192	374/1674:880/1168	94	110	37	2	352	13	0.0430
16384	747/3349:1760/2336	199	144	66	2	625	11	0.0381
32768	1495/6697:3521/4671	232	134	60	2	657	12	0.0201
65536	2995/13389:7041/9343	122	103	37	2	401	12	0.0061
131072	5980/26788:14082/18686	341	137	39	2	875	12	0.0067
262144	11959/53577:28164/37372	401	136	31	2	986	12	0.0038
524288	23923/107149:56329/74743	423	132	31	2	1026	12	0.0020
1048576	47837/214307:112658/149486	492	132	20	2	1153	12	0.0011

We can see that the number of outer iterations is almost invariable, while the number of matrix-vector multiplications in the column labeled n_A slowly increases. We conclude that the optimal performance predicted by our theory can be observed at least partly in practice. However, comparing the results with those obtained for bound and equality problems mentioned above, we can observe that the number of iterations in the above experiment is considerably higher. We consider it natural, as the more general separable constraints naturally complicate the solution. Some results of numerical solutions of quadratic programming problems with separable constraints arising in the solution of contact problems with friction can be found in [25] and [26, 24].

11. Comments and conclusions. We have presented a new algorithm for the minimization of a strictly convex quadratic function subject to convex separable inequality and linear equality constraints. The unique feature of this algorithm is an error bound in terms of bounds on the spectrum of the Hessian to the cost function, independent of the conditioning of the constraints. If applied to a class of problems with the cost functions whose Hessian matrices have the spectrum confined to a given positive interval, the algorithm can find an approximate solution in a uniformly bounded number of basic operations, such as the matrix-vector multiplications. Moreover, if the class of problems admits a sparse representation of the Hessian, it simply follows that the cost of the solution is proportional to the number of unknowns. We checked the performance of the algorithm on the solution of problems with up to over two million unknowns and more than 160000 active constraints. The results of numerical experiments indicate that the optimality can be observed in practice and that the algorithm can be efficient for the solution of real world problems.

The results of the paper in combination with the classical optimality results for the FETI and BETI domain decomposition methods open the way to the development of scalable algorithms for the solution of contact problems of linear elasticity with Tresca friction in three dimensions. Let us recall that these techniques provide an effective preconditioning that does not transform the separable constraints into more general constraints, so that they are able to precondition both linear and nonlinear steps [3, 9]. We shall give the details elsewhere.

Acknowledgments. We are grateful to the two anonymous referees for their careful reading of early versions of the paper and for many valuable comments and corrections that resulted in the improved paper.

REFERENCES

[1] M. Anitescu, *A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming*, SIAM J. Optim., 12 (2002), pp. 949–978.

[2] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, SIAM, Philadelphia, 2001.

[3] J. Bouchala, Z. Dostál, and M. Sadowská, *Scalable total BETI based algorithm for 3D coercive contact problems of linear elastostatics*, Computing, 85 (2009), pp. 189–217.

[4] A. Conn, N. I. M. Gould, and P. L. Toint, *Testing a class of methods for solving minimization problems with simple bounds on the variables*, Math. Comp., 50 (1988), pp. 399–430.

[5] A. Conn, N. I. M. Gould, and P. L. Toint, *Trust Region Methods*, SIAM, Philadelphia, 2000.

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

[7] Z. Dostál, *Inexact semimonotonic augmented Lagrangians with optimal feasibility convergence for convex bound equality constrained quadratic programming*, SIAM J. Numer. Anal., 43 (2005), pp. 96–115.

[8] Z. Dostál, *An optimal algorithm for bound and equality constrained quadratic programming problems with bounded spectrum*, Computing, 78 (2006), pp. 311–328.

[9] Z. Dostál, *Optimal Quadratic Programming Algorithms. With Applications to Variational Inequalities*, Springer Optim. Appl. 23, Springer, New York, 2009.

[10] Z. Dostál, M. Domorádová, and M. Sadowská, *Superrelaxation in minimizing quadratic functions subject to bound constraints*, Comput. Optim. Appl., to appear.

[11] Z. Dostál, J. Haslinger, and R. Kučera, *Implementation of the fixed point method in contact problems with Coulomb friction based on a dual splitting type technique*, J. Comput. Appl. Math., 140 (2002), pp. 245–256.

[12] Z. Dostál, T. Kozubek, P. Horyl, T. Brzobohatý, and A. Markopoulos, *Scalable TFETI algorithm for two dimensional multibody contact problems with friction*, J. Comput. Appl. Math., 235 (2010), pp. 403–418.

[13] Z. Dostál, T. Kozubek, A. Markopoulos, T. Brzobohatý, V. Vondrák, and P. Horyl, *Theoretically supported scalable TFETI algorithm for the solution of multibody 3D contact problems with friction*, Comput. Methods Appl. Mech. Engrg., submitted.

[14] Z. Dostál, T. Kozubek, V. Vondrák, T. Brzobohatý, and A. Markopoulos, *Scalable TFETI algorithm for the solution of coercive multibody contact problems of elasticity*, Internat. J. Numer. Methods Engrg., 82 (2010), pp. 1384–1405.

[15] Z. Dostál and J. Schöberl, *Minimizing quadratic functions subject to bound constraints with the rate of convergence and finite termination*, Comput. Optim. Appl., 30 (2005), pp. 23–45.

[16] Z. Dostál, V. Vondrák, D. Horák, C. Farhat, and P. Avery, *Scalable FETI algorithms for frictionless contact problems*, in Domain Methods in Science and Engineering XVII, U. Langer et al., eds., Lect. Notes Comput. Sci. Eng. 60, Springer, Berlin, 2008, pp. 263–270.

[17] J. G. Ecker and R. D. Niemi, *A dual method for quadratic programs with quadratic constraints*, SIAM J. Appl. Math., 28 (1975), pp. 568–576.

[18] C. Farhat, J. Mandel, and F.-X. Roux, *Optimal convergence properties of the FETI domain decomposition method*, Comput. Methods Appl. Mech. Engrg., 115 (1994), pp. 365–385.

[19] C. Farhat and F.-X. Roux, *An unconventional domain decomposition method for an efficient parallel solution of large-scale finite element systems*, SIAM J. Sci. Statist. Comput., 13 (1992), pp. 379–396.

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

[21] M. Fukushima, Z.-Q. Luo, and P. Tseng, *A sequential quadratically constrained quadratic programming method for differentiable convex minimization*, SIAM J. Optim., 13 (2003), pp. 1098–1119.

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.
[23] W. Hackbusch, *Multigrid Methods and Applications*, Springer-Verlag, Berlin, 1985.
[24] J. Haslinger and R. Kučera, *Scalable T-FETI based algorithm for 3D contact problems with orthotropic friction*, Lect. Notes Appl. Comput. Mech., accepted.
[25] J. Haslinger, R. Kučera, and Z. Dostál, *An algorithm for the numerical realization of 3D contact problems with Coulomb friction*, J. Comput. Appl. Math., 164/165 (2004), pp. 387–408.
[26] J. Haslinger, R. Kučera, and T. Kozubek, *Discretization and numerical realization of contact problems with orthotropic Coulomb friction*, submitted, 2009.
[27] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.
[28] T. Kozubek, A. Markopoulos, T. Brzobohatý, R. Kučera, V. Vondrák, and Z. Dostál, *MatSol—MATLAB Efficient Solvers for Problems in Engineering*, available online at http://www.am.vsb.cz/matsol.
[29] R. Kučera, *Minimizing quadratic functions with separable quadratic constraints*, Optim. Methods Softw., 22 (2007), pp. 453–467.
[30] R. Kučera, *Convergence rate of an optimization algorithm for minimizing quadratic functions with separable convex constraints*, SIAM J. Optim., 19 (2008), pp. 846–862.
[31] R. Kučera, J. Haslinger, and Z. Dostál, *A new FETI-based algorithm for solving 3D contact problems with Coulomb friction*, in Domain Decomposition Methods in Science and Engineering XVI, Lect. Notes Comput. Sci. Eng. 55, Springer, Berlin, 2007, pp. 643–650.
[32] U. Langer and O. Steinbach, *Boundary element tearing and interconnecting method*, Computing, 71 (2003), pp. 205–228.
[33] J. M. Martínez, *Local minimizers of quadratic functions on Euclidean balls and spheres*, SIAM J. Optim., 4 (1994), pp. 159–176
[34] S. Mehrotra and J. Sun, *A method of analytic centers for quadratically constrained convex quadratic programs*, SIAM J. Numer. Anal., 28 (1991), pp. 529–544.
[35] 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.
[36] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.
[37] B. T. Polyak, *The conjugate gradient method in extremal problems*, USSR Comput. Math. Math. Phys., 9 (1969), pp. 94–112.
[38] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., Academic Press, London, 1969, pp. 283–298.
[39] J. B. Rosen, *The gradient projection method for nonlinear programming. Part* I. *Linear constraints*, J. SIAM, 8 (1960), pp. 181–217.
[40] J. B. Rosen, *The gradient projection method for nonlinear programming. Part* II. *Nonlinear constraints*, J. SIAM, 9 (1961), pp. 514–532.
[41] A. Toselli and O. B. Widlund, *Domain Decomposition Methods—Algorithms and Theory*, Springer Ser. Comput. Math. 34, Springer-Verlag, Berlin, 2005.