On Convergence of Inexact Augmented Lagrangians for Separable and Equality Convex QCQP Problems without Constraint Qualification

Zdenek DOSTAL, Petr BEREMLIJSKI

Department of Applied Mathematics, Faculty of Electrical Engineering and Computer Science, VSB–Technical University of Ostrava, 17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

zdenek.dostal@vsb.cz, petr.beremlijski@vsb.cz

DOI: 10.15598/aeee.v15i2.2219

Abstract. *The classical convergence theory of the augmented Lagrangian method has been developed under the assumption that the solutions satisfy a constraint qualification. The point of this note is to show that the constraint qualification can be limited to the constraints that are not enforced by the Lagrange multipliers. In particular, it follows that if the feasible set is non-empty and the inequality constraints are convex and separable, then the convergence of the algorithm is guaranteed without any additional assumptions. If the feasible set is empty and the projected gradients of the Lagrangians are forced to go to zero, then the iterates are shown to converge to the nearest well posed problem.*

Keywords

Augmented Langrangians, constraint qualification, KKT conditions, quadratically constrained quadratic program, SMALSE-M.

1. Introduction

We are interested in the minimization of a convex quadratic function subject to possibly nonlinear separable convex inequality constraints and linear equality constraints

$$\min_{\mathbf{x}\in\Omega_{SE}} f(\mathbf{x}), \qquad f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b}, \tag{1}$$

where

$$\begin{gathered} \Omega_E = \{\mathbf{x}\in\mathbb{R}^n : \mathbf{B}\mathbf{x} = \mathbf{o}\}, \\ \Omega_S = \{\mathbf{x}\in\mathbb{R}^n : h_i(\mathbf{x}_i) \le 0, \mathbf{x}_i \in \mathbb{R}^{n_i}, \\ i = 1,\ldots,s\}, \\ \Omega_{SE} = \Omega_E \cap \Omega_S, \end{gathered} \tag{2}$$

$\mathbf{b}\in\mathbb{R}^n$, h_i are differentiable and convex functions, $\mathbf{A}$ is an $n\times n$ Symmetric Positive Definite (SPD) matrix, and $\mathbf{B}\in\mathbb{R}^{m\times n}$, $\mathbf{B}\neq\mathbf{O}$. If not specified otherwise, we assume $\Omega_{SE}\neq\emptyset$ and admit dependent rows of $\mathbf{B}$, but we do not require that $\mathbf{B}$ is a full column rank matrix, so that Ker$\mathbf{B}\neq\{\mathbf{o}\}$. Observe that some more general Quadratic Programming (QP) or QCQP problems can be reduced to Eq. (1) by duality, a suitable shift of variables, or by a modification of f. The problem arises in the solution of many engineering problems, e.g., in the design of electromagnetic brakes [16] or contact mechanics [13].

Here we restrict our attention to the Augmented Lagrangian method (also called the method of multipliers) proposed by Hestenes (see [19]) and Powell (see [21]) for the solution of nonlinear optimization problems with equality constraints. The algorithms can be considered as a modification of the exterior penalty method which enables to reduce the original equality constrained problem to the series of unconstrained problems without increasing the penalty parameter to infinity. The algorithm is described, e.g., in the classical monographs [3] and [18].

It has been soon observed that the algorithm is useful also for the solution of more general problems, such as the problems with bound and equality constraints. An important instance of such approach was used by Conn, Gold, and Toint (see [6]) in their software LANCELOT (see [7]). An up to date description of the general algorithm can be found in [4]. We are interested in a modification of the algorithm called SMALSE-M (Semimonotonic Augmented Lagrangians for Separable and Equality Constraints). Unlike the LANCELOT, it uses the adaptive precision control introduced by [11] and keeps the regularization parameter constant (see [10], or [9]). The algorithm was proved to generate the iterates with the feasibility error and projected gradients converging to zero (see [12] or [15]). Though some kind of optimality has been proved for SMALSE-M, no proof of convergence of the iterates themselves was given.

Our main goal here is to show that the iterates generated by SMALSE-M converge to the solution *without assuming any constraint qualification* and to prove the convergence of the Lagrange multipliers under the assumption of the regularity of the solution. Let us recall that the optimality properties of SMALSE-M were exploited in the development of scalable algorithms for the solution of contact problems with friction [13] and [14]. If the feasible set is empty and the projected gradients of the Lagrangians are forced to go to zero, then the iterates of a slightly modified algorithm are shown to converge to the solution of a nearest well posed problem. Notice that a similar results were obtained by Gilbert and Chiche [5] or Gilbert and Joannopoulis [17]. The paper is an extension of the results that appeared in [14].

2. KKT Conditions

Since Ω_{SE} is closed, convex and nonempty and f is assumed to be strictly convex, the solution of problem Eq. (1) exists and is necessarily unique ([9], Proposition 2.5). If the intersection of Ω_{SE} with the interior of Ω_S is nonempty, then the solution satisfies the Slater constraint qualification and can be characterized by means of the augmented Lagrangian

$$\begin{aligned} L(\mathbf{x}, \lambda, \mu, \varrho) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b} + \mathbf{x}^T\mathbf{B}^T\lambda + \\ + \frac{\varrho}{2}\|\mathbf{B}\mathbf{x}\|^2 + \sum_{i=1}^{s}\mu_i h_i(\mathbf{x}_i), \\ \lambda \in \mathbb{R}^m, \quad \mu \in \mathbb{R}^s, \end{aligned} \tag{3}$$

whose gradient reads

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \mu, \varrho) = (\mathbf{A} + \varrho\mathbf{B}^T\mathbf{B})\mathbf{x} - \mathbf{b} + \mathbf{B}^T\lambda + \\ + \sum_{i=1}^{s}\mu_i \nabla h_i(\mathbf{x}_i). \end{aligned} \tag{4}$$

If the solution satisfies the Slater constraint qualification, then a feasible vector $\mathbf{x} \in \Omega_{SE}$ is a solution of Eq. (1) if and only if there is $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^s$ such that

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \mu, \varrho) = \mathbf{o}, \quad \mu_i \geq 0, \\ \text{and} \quad \mu_i h_i(\mathbf{x}_i) = 0, \quad i = 1, \ldots, s. \end{aligned} \tag{5}$$

Having effective algorithms for the solution of QCQP problems with separable constraints, it is convenient to use explicitly the Lagrange multipliers only for the equality constraints, i.e., to use

$$L(\mathbf{x}, \lambda, \varrho) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{b} + \mathbf{x}^T\mathbf{B}^T\lambda + \frac{\varrho}{2}\|\mathbf{B}\mathbf{x}\|^2, \tag{6}$$

and the gradient $\mathbf{g} = \nabla_{\mathbf{x}} L$ of the reduced augmented Lagrangian $\mathbf{g} = \mathbf{g}(\mathbf{x}, \lambda, \varrho)$ defined by

$$\mathbf{g}(\mathbf{x}, \lambda, \varrho) = (\mathbf{A} + \varrho\mathbf{B}^T\mathbf{B})\mathbf{x} - \mathbf{b} + \mathbf{B}^T\lambda. \tag{7}$$

To simplify the notation, we use

$$\mathbf{g} = \mathbf{g}(\mathbf{x}) = \mathbf{g}(\mathbf{x}, \lambda, \varrho), \tag{8}$$

when we can specify the remaining arguments from the context.

To rewrite the KKT conditions Eq. (5) in a form suitable for our exposition, we introduce some auxiliary notations. Let $\mathcal{S}$ denote the set of all indices of the constraints so that

$$\mathcal{S} = \{1, 2, \ldots, s\}. \tag{9}$$

For any $\mathbf{x} \in \mathbb{R}^n$, we define the *active set* of $\mathbf{x}$ by

$$\mathcal{A}(\mathbf{x}) = \{i \in \mathcal{S} : h_i(\mathbf{x}_i) = 0\}. \tag{10}$$

Its complement

$$\mathcal{F}(\mathbf{x}) = \{i \in \mathcal{S} : h_i(\mathbf{x}_i) \neq 0\}, \tag{11}$$

is called a *free set*. For $\mathbf{x} \in \Omega_S$, we define the outer unit normal $\mathbf{n}$ by

$$\begin{aligned} \mathbf{n}_i = \mathbf{n}_i(\mathbf{x}) = \\ = \begin{cases} \|\nabla h_i(\mathbf{x}_i)\|^{-1}\nabla h_i(\mathbf{x}_i) & \text{for } i \in \mathcal{A}(\mathbf{x}), \\ \mathbf{o} & \text{for } i \in \mathcal{F}(\mathbf{x}). \end{cases} \end{aligned} \tag{12}$$

The components of the gradient that violate the KKT conditions Eq. (5) in the free and active set are called the *free gradient* φ and the *chopped gradient* β, respectively. They are defined by

$$\varphi_i(\mathbf{x}) = \mathbf{g}_i(\mathbf{x}) \text{ for } i \in \mathcal{F}(\mathbf{x}), \tag{13}$$
$$\varphi_i(\mathbf{x}) = \mathbf{o} \text{ for } i \in \mathcal{A}(\mathbf{x}), \tag{14}$$
$$\beta_i(\mathbf{x}) = \mathbf{o} \text{ for } i \in \mathcal{F}(\mathbf{x}), \tag{15}$$
$$\beta_i(\mathbf{x}) = \mathbf{g}_i(\mathbf{x}) - (\mathbf{n}_i^T\mathbf{g}_i)^-\mathbf{n}_i \text{ for } i \in \mathcal{A}(\mathbf{x}), \tag{16}$$

where we use the notation $(\mathbf{n}_i^T\mathbf{g}_i)^- = \min\{\mathbf{n}_i^T\mathbf{g}_i, 0\}$.

If we define the *projected gradient* by

$$\mathbf{g}^P(\mathbf{x}) = \varphi(\mathbf{x}) + \beta(\mathbf{x}), \tag{17}$$

it is easy to check that $\mathbf{x} \in \Omega_{SE}$ is a solution of Eq. (1) if and only if there is $\lambda \in \mathbb{R}^m$ such that

$$\mathbf{g}^P(\mathbf{x}, \lambda, \varrho) = \mathbf{o}. \tag{18}$$

We shall need yet another simple property of the projected gradient.

Lemma 1. *Let* $\mathbf{x}, \mathbf{y} \in \Omega_S$ *and* $\mathbf{g} = \nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \varrho)$*. Then*

$$\mathbf{g}^T(\mathbf{y}-\mathbf{x}) \geq (\mathbf{g}^P)^T(\mathbf{y}-\mathbf{x}). \tag{19}$$

Proof: First observe that

$$\mathbf{g}^T(\mathbf{y}-\mathbf{x}) = (\mathbf{g}-\mathbf{g}^P)^T(\mathbf{y}-\mathbf{x}) + (\mathbf{g}^P)^T(\mathbf{y}-\mathbf{x}). \tag{20}$$

Using the definition of the projected gradient, we get

$$\begin{aligned} &(\mathbf{g}-\mathbf{g}^P)^T(\mathbf{y}-\mathbf{x}) = \\ &= \sum_{i\in\mathcal{S}} (\mathbf{g}_i - \mathbf{g}_i^P)^T(\mathbf{y}_i - \mathbf{x}_i) = \\ &= \sum_{i\in\mathcal{A}(x)} (\mathbf{n}_i^T\mathbf{g}_i)^- \mathbf{n}_i^T(\mathbf{y}_i - \mathbf{x}_i). \end{aligned} \tag{21}$$

To finish the proof, it is sufficient to observe that for $i \in \mathcal{A}(\mathbf{x})$

$$\mathbf{n}_i^T(\mathbf{y}_i - \mathbf{x}_i) \leq 0 \tag{22}$$

due to the convexity of h_i. □

3. SMALSE-M

The complete SMALSE-M algorithm reads as follows.

In Step 1 we can use any algorithm for minimizing a strictly convex quadratic function subject to separable constraints as long as it guarantees the convergence of the projected gradient to zero. The next lemma recalls that Alg. 1 is well defined. Typical choice of parameters is $\eta_k = 0.1\|\mathbf{b}\|$, $\beta = 0.1$, $M_0 = 10$, and $\varrho = \|\mathbf{A}\|$.

Lemma 2. *Let* $M > 0$, $\lambda \in \mathbb{R}^m$, $\eta > 0$, $\varrho \geq 0$, *and* $\alpha \in (0, 2(\|\mathbf{A} + \varrho\mathbf{B}^T\mathbf{B}\|^2)^{-1})$ *be given. Let* $\{\mathbf{y}^k\} \in \Omega_S$ *denote any sequence such that*

$$\widehat{\mathbf{y}} = \lim_{k\to\infty} \mathbf{y}^k = \arg\min_{\mathbf{y}\in\Omega_S} L(\mathbf{y}, \lambda, \varrho), \tag{26}$$

and $\mathbf{g}^P(\mathbf{y}^k, \lambda, \varrho)$ *converges to the zero vector. Then* $\{\mathbf{y}^k\}$ *either converges to the unique solution* $\widehat{\mathbf{x}}$ *of problem Eq. (1), or there is an index* k *such that*

$$\|\mathbf{g}^P(\mathbf{y}^k, \lambda, \varrho)\| \leq \min\{M\|\mathbf{B}\mathbf{y}^k\|, \eta\}. \tag{27}$$

Proof: If Eq. (27) does not hold for any k, then

$$\|\mathbf{g}^P(\mathbf{y}^k, \lambda, \varrho)\| > M\|\mathbf{B}\mathbf{y}^k\|, \tag{28}$$

for any k. Since $\mathbf{g}^P(\mathbf{y}^k, \lambda, \varrho)$ converges to the zero vector by the assumption, it follows that $\|\mathbf{B}\mathbf{y}^k\|$ converges to zero. Thus $\mathbf{B}\widehat{\mathbf{y}} = \mathbf{o}$ and

$$\mathbf{g}^P(\widehat{\mathbf{y}}, \lambda, \varrho) = \mathbf{o}. \tag{29}$$

Algorithm 1 Semimonotonic augmented Lagrangians for separable and equality constrained QCQP problems (SMALSE-M).

Given an SPD matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{m\times n}$, n-vector $\mathbf{b}$, constraints $\mathbf{h}$.

Step 0. {Initialization.}
Choose $\eta \geq \eta_k > 0$, $0 < \beta < 1$, $M_0 > 0$, $\varrho > 0$, $\lambda^0 \in \mathbb{R}^m$

for $k = 0, 1, 2, \ldots$ **do**
Step 1. {Inner iteration.}
Find $\mathbf{x}^k \in \Omega_S$ such that

$$\|\mathbf{g}^P(\mathbf{x}^k, \lambda^k, \varrho)\| \leq \min\{M_k\|\mathbf{B}\mathbf{x}^k\|, \eta_k\} \tag{23}$$

Step 2. {Updating the Lagrange multipliers.}

$$\lambda^{k+1} = \lambda^k + \varrho\mathbf{B}\mathbf{x}^k \tag{24}$$

Step 3. {Update of M.}
if $k > 0$ and $M_{k-1} = M_k$ and

$$L(\mathbf{x}^k, \lambda^k, \varrho) < L(\mathbf{x}^{k-1}, \lambda^{k-1}, \varrho) + \frac{\varrho}{2}\|\mathbf{B}\mathbf{x}^k\|^2 \tag{25}$$

then
$M_{k+1} = \beta M_k$
else
$M_{k+1} = M_k$
end if
end for

Thus $\widehat{\mathbf{y}}$ satisfies the KKT conditions Eq. (18) and $\widehat{\mathbf{y}} = \widehat{\mathbf{x}}$. □

The tools used for the analysis of SMALSE-M were sufficient to prove only the convergence of the feasibility error and the projected gradient [12]. The results relevant for our exposition read as follows.

Proposition 1. *Let* $\{\mathbf{x}^k\}, \{\lambda^k\}$, *and* $\{M_k\}$ *be generated by Alg. 1 for the solution of Eq. (1) with* $\eta > 0$, $0 < \beta < 1$, $M_0 > 0$, $\varrho > 0$, *and* $\lambda^0 \in \mathbb{R}^m$. *Let* $\lambda_{\min}$ *denote the smallest eigenvalue of the Hessian* $\mathbf{A}$ *of the cost function* f, *and let* $p \geq 0$ *denote the smallest integer such that* $\beta^p\varrho \geq M_0^2/\lambda_{\min}$. *Then the following statements hold:*
(i) M_k *satisfies*

$$M_k \geq \min\{M_0, \beta\sqrt{\varrho\lambda_{\min}}\}, \quad k = 0, 1, 2, \ldots \,. \tag{30}$$

(ii) If $\Omega_{SE} \neq \emptyset$, *then*

$$\lim_{k\to\infty} \mathbf{g}^P(\mathbf{x}^k, \lambda^k, \varrho) = \mathbf{o} \quad \text{and} \quad \lim_{k\to\infty} \mathbf{B}\mathbf{x}^k = \mathbf{o}. \tag{31}$$

4. Boundedness

The first step toward the proof of convergence of our SMALSE-M Alg. 1 is to show that the iterates $\mathbf{x}^k$ are bounded.

Proposition 2. *Let $\{\mathbf{x}^k\}$ and $\{\lambda^k\}$ be generated by Alg. 1 for the solution of Eq. (1) with $\eta > 0$, $0 < \beta < 1$, $M_0 > 0$, $\varrho > 0$, and $\lambda^0 \in \mathbb{R}^m$. For each $i \in \{1, \ldots, s\}$, let $\mathcal{I}_i$ denote the indices of the components of $\mathbf{x}_i$, where $\mathbf{x}_i$ is the argument of the constraint function h_i, and let us define*

$$\begin{aligned} \widetilde{\mathcal{A}}(\mathbf{x}) &= \cup_{i\in\mathcal{A}(\mathbf{x})}\mathcal{I}_i, \quad \widetilde{\mathcal{F}}(\mathbf{x}) = \mathcal{S} \setminus \widetilde{\mathcal{A}}(\mathbf{x}), \\ \mathbf{x}_{\mathcal{A}} &= \mathbf{x}_{\widetilde{\mathcal{A}}(\mathbf{x})}, \quad \mathbf{x}_{\mathcal{F}} = \mathbf{x}_{\widetilde{\mathcal{F}}(\mathbf{x})}. \end{aligned} \tag{32}$$

Assume that the boundary of Ω_S is bounded, i.e., there is $C > 0$ such that for any $\mathbf{x} \in \Omega_S$

$$\|\mathbf{x}_{\mathcal{A}(\mathbf{x})}\|^2 \leq C. \tag{33}$$

Then $\{\mathbf{x}^k\}$ is bounded. Moreover, if the solution of Eq. (1) is regular, then also $\{\lambda^k\}$ is bounded.

Proof: Since there is only a finite number of different subsets $\mathcal{F}$ of the set of all indices $\mathcal{S} = \{1, \ldots, s\}$, and $\{\mathbf{x}^k\}$ is bounded if and only if $\{\mathbf{x}^k_{\mathcal{F}}\}$ is bounded, we can restrict our attention to the analysis of infinite subsequences $\{\mathbf{x}^k_{\mathcal{F}} : \widetilde{\mathcal{F}}(\mathbf{x}^k) = \mathcal{F}\}$ that are defined by the nonempty subsets $\mathcal{F}$ of $\mathcal{S}$.

Let $\mathcal{F} \subseteq \mathcal{S}$, $\mathcal{F} \neq \emptyset$, let $\mathcal{K} = \{k : \widetilde{\mathcal{F}}(\mathbf{x}^k) = \mathcal{F}\}$ be infinite, and denote

$$\mathcal{A} = \mathcal{S} \setminus \mathcal{F}, \quad \mathbf{H} = \mathbf{A} + \varrho\mathbf{B}^T\mathbf{B}. \tag{34}$$

We get

$$\mathbf{g}^k = \mathbf{g}(\mathbf{x}^k, \lambda^k, \varrho) = \mathbf{H}\mathbf{x}^k + \mathbf{B}^T\lambda^k - \mathbf{b}, \tag{35}$$

and

$$\begin{aligned} &\begin{bmatrix} \mathbf{H}_{\mathcal{F}\mathcal{F}} & \mathbf{B}^T_{*\mathcal{F}} \\ \mathbf{B}_{*\mathcal{F}} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{x}^k_{\mathcal{F}} \\ \lambda^k \end{bmatrix} = \\ &= \begin{bmatrix} \mathbf{g}^k_{\mathcal{F}} + \mathbf{b}_{\mathcal{F}} - \mathbf{H}_{\mathcal{F}\mathcal{A}}\mathbf{x}^k_{\mathcal{A}} \\ \mathbf{B}_{*\mathcal{F}}\mathbf{x}^k_{\mathcal{F}} \end{bmatrix}. \end{aligned} \tag{36}$$

Since for $k \in \mathcal{K}$

$$\begin{aligned} \mathbf{B}_{*\mathcal{F}}\mathbf{x}^k_{\mathcal{F}} &= \mathbf{B}\mathbf{x}^k - \mathbf{B}_{*\mathcal{A}}\mathbf{x}^k_{\mathcal{A}}, \\ \|\mathbf{g}^k_{\mathcal{F}}\| &= \|\mathbf{g}_{\mathcal{F}}(\mathbf{x}^k, \lambda^k, \varrho)\| \leq \|\mathbf{g}^P(\mathbf{x}^k, \lambda^k, \varrho)\|, \end{aligned} \tag{37}$$

and both $\|\mathbf{g}^P(\mathbf{x}^k, \lambda^k, \varrho)\|$ and $\|\mathbf{B}\mathbf{x}^k\|$ converge to zero by the definition of $\mathbf{x}^k$ in Step 1 of Alg. 1 and Eq. (31), the right-hand side of Eq. (36) is bounded. Since $\mathbf{H}_{\mathcal{F}\mathcal{F}}$ is nonsingular, it is easy to check that the matrix of the system Eq. (36) is nonsingular when $\mathbf{B}_{*\mathcal{F}}$ is a full row rank matrix. It simply follows that both $\{\mathbf{x}^k\}$ and $\{\lambda^k\}$ are bounded provided the matrix of the system Eq. (36) is nonsingular.

If $\mathbf{B}_{*\mathcal{F}}$ is not a full row rank matrix, then its rank r satisfies $r < m$, and by the singular value decomposition formula there are orthogonal matrices

$$\begin{aligned} \mathbf{U} &= [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbb{R}^{m\times m}, \\ \mathbf{V} &= [\mathbf{v}_1, \ldots, \mathbf{v}_s] \in \mathbb{R}^{s\times s}, \end{aligned} \tag{38}$$

and the diagonal matrix

$$\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0), \quad \mathbf{\Sigma} \in \mathbb{R}^{m\times s}, \tag{39}$$

with the nonzero diagonal entries $\sigma_1 > 0, \ldots, \sigma_r > 0$ such that $\mathbf{B}_{*\mathcal{F}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Taking

$$\begin{aligned} \hat{\mathbf{U}} = [\mathbf{u}_1, \ldots, \mathbf{u}_r], \quad \hat{\mathbf{\Sigma}} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r), \\ \text{and} \quad \hat{\mathbf{V}} = [\mathbf{v}_1, \ldots, \mathbf{v}_r], \end{aligned} \tag{40}$$

we have $\mathbf{B}_{*\mathcal{F}} = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{V}}^T$ and we can define a full row rank matrix

$$\hat{\mathbf{B}}_{*\mathcal{F}} = \hat{\mathbf{D}}\hat{\mathbf{V}}^T = \hat{\mathbf{U}}^T\mathbf{B}_{*\mathcal{F}}, \tag{41}$$

that satisfies for any vector $\mathbf{x}$

$$\begin{aligned} \hat{\mathbf{B}}^T_{*\mathcal{F}}\hat{\mathbf{B}}_{*\mathcal{F}} = \mathbf{B}^T_{*\mathcal{F}}\mathbf{B}_{*\mathcal{F}} \quad \text{and} \\ \|\hat{\mathbf{B}}_{*\mathcal{F}}\mathbf{x}_{\mathcal{F}}\| = \|\mathbf{B}_{*\mathcal{F}}\mathbf{x}_{\mathcal{F}}\|. \end{aligned} \tag{42}$$

We shall assign to any $\lambda \in \mathbb{R}^m$ the vector $\hat{\lambda} = \hat{\mathbf{U}}^T\lambda$, so that $\hat{\mathbf{B}}^T_{\mathcal{F}}\hat{\lambda} = \mathbf{B}^T_{\mathcal{F}}\lambda$. Using the latter identity and Eq. (36), we get the system

$$\begin{aligned} &\begin{bmatrix} \mathbf{H}_{\mathcal{F}\mathcal{F}} & \hat{\mathbf{B}}^T_{*\mathcal{F}} \\ \hat{\mathbf{B}}_{*\mathcal{F}} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{x}^k_{\mathcal{F}} \\ \hat{\lambda}^k \end{bmatrix} = \\ &= \begin{bmatrix} \mathbf{g}^k_{\mathcal{F}} + \mathbf{b}_{\mathcal{F}} - \mathbf{H}_{\mathcal{F}\mathcal{A}}\mathbf{x}^k_{\mathcal{A}} \\ \hat{\mathbf{B}}_{*\mathcal{F}}\mathbf{x}^k_{\mathcal{F}} \end{bmatrix}, \end{aligned} \tag{43}$$

with a nonsingular matrix. The right-hand side of Eq. (43) being bounded due to $\|\hat{\mathbf{B}}_{*\mathcal{F}}\mathbf{x}^k_{\mathcal{F}}\| = \|\mathbf{B}_{*\mathcal{F}}\mathbf{x}^k_{\mathcal{F}}\|$, we conclude that the set $\{\mathbf{x}^k_{\mathcal{F}} : \mathcal{F}(\mathbf{x}^k) = \mathcal{F}\}$ is bounded. See also [15] or [12]. □

Remark 1. *Notice that the assumption on the boundary of Ω_S does not imply the compactness of Ω_{SE}.*

5. Convergence

Now we are ready to prove the main convergence results. To describe them, let $\mathcal{F} = \widetilde{\mathcal{F}}(\widehat{\mathbf{x}})$ denote the set of the indices of variables that are involved in the free constraints of the unique solution $\widehat{\mathbf{x}}$ and recall that $\widehat{\mathbf{x}}$ is a *regular solution* of Eq. (1) if $\mathbf{B}_{*\mathcal{F}}$ is a full row rank matrix ($*$ denotes the full set of indices).

Theorem 1. *Let $\{\mathbf{x}^k\}$ and $\{\lambda^k\}$ be generated by Alg. 1 for the solution of Eq. (1) with $\eta > 0$, $0 < \beta < 1$, $M_0 > 0$, $\varrho > 0$, and $\lambda^0 \in \mathbb{R}^m$. Then the following statements hold.*
(i) $\{\mathbf{x}^k\}$ converges to the solution $\widehat{\mathbf{x}}$ of (Eq. (1).
(ii) If the solution $\widehat{\mathbf{x}}$ of Eq. (1) is regular, then $\{\lambda^k\}$ converges to the uniquely determined vector $\widehat{\lambda}$ of Lagrange multipliers of Eq. (1).
(iii) Let us define for any $\mathbf{d} \in \mathbb{R}^m$

$$\overline{\Omega}_E(\mathbf{d}) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{B}\mathbf{x} = \mathbf{d}\}. \tag{44}$$

If the feasible set Ω_{SE} is empty and $\eta_k \to 0$, then $\{\mathbf{x}^k\}$ converges to the solution $\overline{\mathbf{x}}$ of the nearest feasible problem with shifted equality constraints,

$$\overline{\Omega}_E = \overline{\Omega}_E(\overline{\mathbf{d}}), \tag{45}$$

where

$$\overline{\mathbf{d}} = \arg\min \|\mathbf{d}\| \quad \text{s.t.} \quad \Omega_E(\mathbf{d}) \cap \Omega_S \neq \emptyset. \tag{46}$$

Proof: (i) Since the iterates $\mathbf{x}^k$ are bounded due to Prop. 2, it follows that there is a cluster point $\overline{\mathbf{x}}$ of $\{\mathbf{x}^k\}$ and there is $\mathcal{K} \subseteq \mathbb{N}$ such that

$$\lim_{k\to\infty} \{\mathbf{x}^k\}_{k\in\mathcal{K}} = \overline{\mathbf{x}}. \tag{47}$$

Moreover, since $\mathbf{x}^k \in \Omega_S$ and by Eq. (31)

$$\lim_{k\to\infty} \|\mathbf{B}\mathbf{x}^k\| = 0, \tag{48}$$

it follows that $\mathbf{B}\overline{\mathbf{x}} = \mathbf{o}$ and $\overline{\mathbf{x}} \in \Omega_{SE}$.

To show that $\overline{\mathbf{x}}$ solves Eq. (1), let $\mathbf{d} \in \mathbb{R}^n$ satisfy $\overline{\mathbf{x}}+\mathbf{d} \in \Omega_{SE}$, so that $\mathbf{B}\mathbf{d} = \mathbf{o}$, and set $\mathbf{d}^k = \mathbf{d}+(\overline{\mathbf{x}}-\mathbf{x}^k)$, so that $\mathbf{x}^k + \mathbf{d}^k \in \Omega_{SE}$. Thus we can use Lem. 1 to get for each $k \in \mathcal{K}$

$$\begin{aligned} & f(\mathbf{x}^k + \mathbf{d}^k) - f(\mathbf{x}^k) = \\ & = L(\mathbf{x}^k + \mathbf{d}^k, \lambda^k, \varrho) - L(\mathbf{x}^k, \lambda^k, \varrho) = \\ & = \mathbf{g}^T(\mathbf{x}^k, \lambda^k, \varrho)\mathbf{d}^k + \tfrac{\varrho}{2}\left(\mathbf{d}^k\right)^T \mathbf{A}\mathbf{d}^k \geq \\ & \geq \left(\mathbf{d}^k\right)^T \mathbf{g}^P(\mathbf{x}^k, \lambda^k, \varrho) + \tfrac{\varrho}{2}\left(\mathbf{d}^k\right)^T \mathbf{A}\mathbf{d}^k \geq \\ & \geq -M_0\|\mathbf{B}\mathbf{x}^k\| + \frac{\varrho\lambda_{\min}}{2}\|\mathbf{d}^k\|^2. \end{aligned} \tag{49}$$

We used the relation $\mathbf{g}^T\mathbf{d}^k \geq (\mathbf{g}^P)^T\mathbf{d}^k$ (see Lem. 1 of [12])

Taking the limits, we get

$$f(\overline{\mathbf{x}} + \mathbf{d}) - f(\overline{\mathbf{x}}) \geq \frac{\varrho\lambda_{\min}}{2}\|\mathbf{d}\|^2. \tag{50}$$

Thus $\overline{\mathbf{x}}$ solves Eq. (1). The solution $\widehat{\mathbf{x}}$ of Eq. (1) being unique, it follows that $\mathbf{x}^k$ converges to $\overline{\mathbf{x}} = \widehat{\mathbf{x}}$.

(ii) Let us recall that $\mathcal{F} = \widetilde{\mathcal{F}}(\widehat{\mathbf{x}})$ and denote $\mathbf{H} = \mathbf{A} + \varrho\mathbf{B}^T\mathbf{B}$. Since we have just proved that $\{\mathbf{x}^k\}$ converges to $\widehat{\mathbf{x}}$, there is k_1 such that $\mathcal{F} \subseteq \mathcal{F}\{\mathbf{x}^k\}$ for $k \geq k_1$ and

$$\mathbf{g}_{\mathcal{F}}(\mathbf{x}^k, \lambda^k, \varrho) = \mathbf{H}_{\mathcal{F}*}\mathbf{x}^k - \mathbf{b}_{\mathcal{F}} + \mathbf{B}^T_{*\mathcal{F}}\lambda^k, \tag{51}$$

converges to zero. It follows that the sequence

$$\mathbf{B}^T_{*\mathcal{F}}\lambda^k = \mathbf{b}_{\mathcal{F}} - \mathbf{H}_{\mathcal{F}*}\mathbf{x}^k + \mathbf{g}_{\mathcal{F}}(\mathbf{x}^k, \lambda^k, \varrho), \tag{52}$$

is bounded. Moreover, if $\overline{\lambda}$ is any vector of Lagrange multipliers, then

$$\mathbf{b} = \mathbf{H}\widehat{\mathbf{x}} + \mathbf{B}^T\overline{\lambda}, \tag{53}$$

and

$$\mathbf{B}^T_{*\mathcal{F}}(\lambda^k - \overline{\lambda}) = -\mathbf{H}_{\mathcal{F}*}(\mathbf{x}^k - \widehat{\mathbf{x}}) + \mathbf{g}_{\mathcal{F}}(\mathbf{x}^k, \lambda^k, \varrho), \tag{54}$$

converges to zero.

If the solution $\widehat{\mathbf{x}}$ of Eq. (1) is regular, then $\mathbf{B}^T_{*\mathcal{F}}$ is a full column rank matrix, i.e., $\mathbb{R}^m = \mathrm{Im}\mathbf{B}_{*\mathcal{F}}$, and there is the unique Lagrange multiplier $\widehat{\lambda}$ for problem Eq. (1). Moreover, since

$$\lambda^k - \widehat{\lambda} \in \mathrm{Im}\mathbf{B}_{*\mathcal{F}}, \tag{55}$$

it follows that

$$\|\mathbf{B}^T_{*\mathcal{F}}(\lambda^k - \widehat{\lambda})\| \geq \overline{\sigma}^{\mathcal{F}}_{\min}\|\lambda^k - \widehat{\lambda}\|, \tag{56}$$

where $\overline{\sigma}^{\mathcal{F}}_{\min}$ denotes the smallest nonzero singular value of $\mathbf{B}_{*\mathcal{F}}$. The convergence of the right-hand side of Eq. (56) to zero thus implies that λ^k converges to $\widehat{\lambda}$.

(iii) As in the proof of (i), there is a cluster point $\overline{\mathbf{x}}$ of $\mathbf{x}^k$. Using the other arguments of the proof of (i), we get that $\overline{\mathbf{x}}$ solves the problem to find $\min f(\mathbf{x})$ subject to $\mathbf{x} \in \Omega_S$ and $\mathbf{B}\mathbf{x} = \mathbf{B}\overline{\mathbf{x}}$. Moreover, if the feasible set Ω_{SE} is empty, then the multipliers λ^k are necessarily unbounded. After some simple manipulations, we get that the vector $\overline{\mathbf{n}}$ with the nonzero components

$$\overline{\mathbf{n}}_{\mathcal{I}_i} = \nabla h_i(\overline{\mathbf{x}}_i), \tag{57}$$

satisfies $\overline{\mathbf{n}} \in \mathrm{Im}\mathbf{B}^T$. Due to the convexity, it follows that $\overline{\mathbf{x}}$ is the nearest point of Ω_S to Ω_E. □

6. Numerical Results

Now we are ready to show some numerical examples which illustrate the preceding results. The feasible set in the first example contains only one point and the solution of the minimization problem in Exm. 1 is not regular. The feasible set in Exm. 2 is empty and Alg. 1 converges to the closest point of the set Ω_S to the set Ω_E (see Exm. 2 for formal definition). The sequence of the Lagrange multipliers for the equality constraint is unbounded in both cases.

Example 1. *In the first example the feasible set Ω_{SE} is restricted to the point. There is no point of the feasible set which satisfies any constraint qualification (Slater constraint qualification, regularity condition, etc.). The optimization problem reads as follows:*

$$\min_{\mathbf{x}\in\Omega_{SE}} f(\mathbf{x}), \qquad f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{x} - \mathbf{x}^T \left[\begin{array}{c} 1 \\ 1 \end{array} \right], \tag{58}$$

where

$$\begin{aligned} \Omega_E &= \{\mathbf{x} \in \mathbb{R}^2 : x_1 = 0\}, \\ \Omega_S &= \{\mathbf{x} \in \mathbb{R}^2 : (x_1 - 1)^2 + x_2^2 - 1 \leq 0\}, \\ \Omega_{SE} &= \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \in \Omega_E \ \text{ and } \mathbf{x} \in \Omega_S\}. \end{aligned} \tag{59}$$

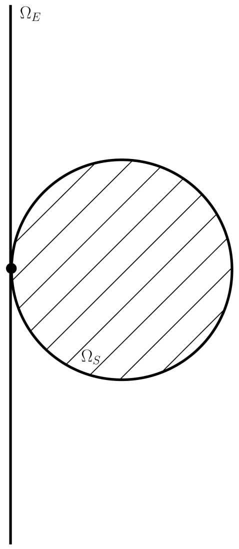

Fig. 1: Example 1, equality and inequality constraints.

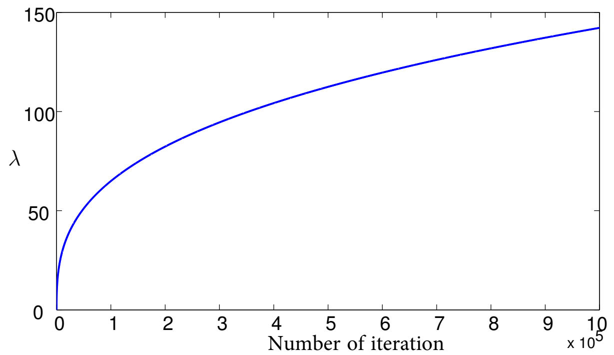

Fig. 2: Example 1, values of the Lagrange multipliers during the iteration process.

The solution of the problem is $\widehat{\mathbf{x}} = (0, 0)$.

Figure 1 shows the geometric interpretation of the equality and inequality constraints. The values of the

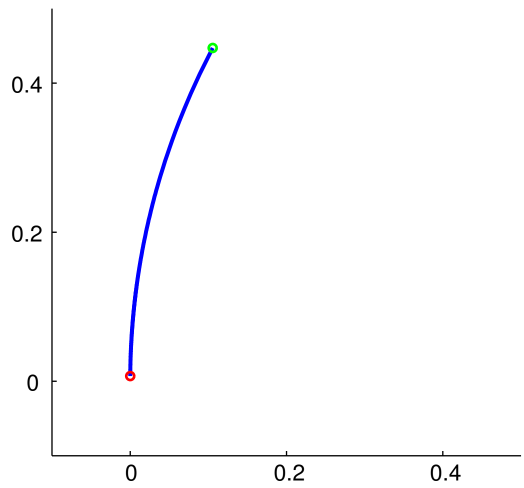

Fig. 3: Example 1, values of the vectors $\mathbf{x}$ during the iteration process.

Lagrange multipliers for the equality constraint during the iteration process of Alg. 1 are depicted in Fig. 2. The values of the vectors $\mathbf{x}$ during the iteration process of the same algorithm are depicted in Fig. 3. The green circle marks the initial iteration and the red circle indicates the last iteration.

Example 2. *In the second example the feasible set Ω_{SE} is empty. There is no point of the feasible set which satisfies any constraint qualification (Slater constraint qualification, regularity condition, etc.). The optimization problem reads as follows:*

$$\min_{\mathbf{x}\in\Omega_{SE}} f(\mathbf{x}), \qquad f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{x} - \mathbf{x}^T \left[\begin{array}{c} 1 \\ 1 \end{array} \right], \tag{60}$$

where

$$\begin{aligned} \Omega_E &= \{\mathbf{x} \in \mathbb{R}^2 : x_1 = 0\}, \\ \Omega_S &= \{\mathbf{x} \in \mathbb{R}^2 : (x_1 - 10)^2 + x_2^2 - 1 \leq 0\}, \\ \Omega_{SE} &= \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \in \Omega_E \ \text{ and } \mathbf{x} \in \Omega_S\}. \end{aligned} \tag{61}$$

If $\eta^k \rightarrow 0$, then Alg. 1 converges to the vector $\widehat{\mathbf{x}} = (9, 0)$ which is the closest point of Ω_S to Ω_E.

Figure 4 shows the geometric interpretation of the equality and inequality constraints. The values of the Lagrange multipliers for the equality constraint during the iteration process of Alg. 1 are depicted in Fig. 5. The values of the vectors $\mathbf{x}^k$ during the iteration process of the same algorithm are depicted in Fig. 6. The green circle marks the initial iteration and the red circle indicates the last iteration.

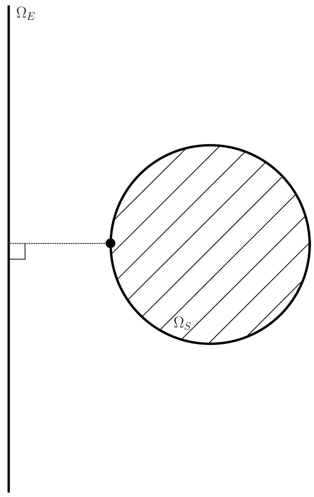

Fig. 4: Example 2, equality and inequality constraints.

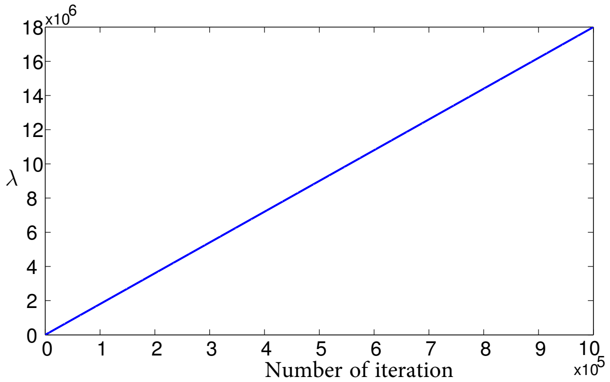

Fig. 5: Example 2, values of the Lagrange multipliers during the iteration process.

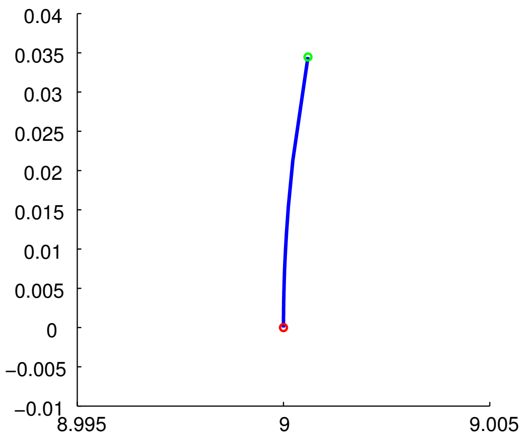

Fig. 6: Example 2, values of the vectors $\mathbf{x}$ during the iteration process.

Acknowledgment

The work is partially supported by Grant of SGS No. SP2017/122, VSB–Technical University of Ostrava, Czech Republic.

References

[1] ANDREANI, R., G. HAESER and J. M. MARTINEZ. On sequential optimality conditions for smooth constrained optimization. *Optimization*. 2011, vol. 60, iss. 5, pp. 627–641. ISSN 0233-1934. DOI: 10.1080/02331930903578700.

[2] ANDREANI, R., J. M. MARTINEZ and B. F. SVAITER. A New Sequential Optimality Condition for Constrained Optimization and Algorithmic Consequences. *SIAM Journal on Optimization*. 2010, vol. 20, no. 6, pp. 3533–3554. ISSN 1052-6234. DOI: 10.1137/090777189.

[3] BERTSEKAS, D. P. *Constrained Optimization and Lagrange Multiplier Methods*. 1st ed. London: Academic Press, 1982. ISBN 978-1-48-326047-1.

[4] BIRGIN, E. M. and J. M. MARTINEZ. *Practical Augmented Lagrangian Method for Constrained Optimization*. 1st ed. Philadelphia: SIAM, 2014. ISBN 978-1-611973-35-8.

[5] CHICHE, A. and J. C. GILBERT. How the augmented Lagrangian algorithm can deal with an infeasible convex quadratic optimization problem. *Journal of Convex Analysis*. 2016, vol. 23, iss. 2, pp. 425–459. ISSN 0944-6532.

[6] CONN, A. R., N. I. M. GOULD and P. L. TOINT. A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds. *SIAM Journal on Numerical Analysis*. 1991, vol. 28, no. 2, pp. 545–572. ISSN 0036-1429. DOI: 10.1137/0728030.

[7] CONN, A. R., N. I. M. GOULD and P. L. TOINT. *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization (Release A)*. 1st ed. New York: Springer, 1992. ISBN 978-3-540-55470-7.

[8] DOSTAL, Z. Inexact Semimonotonic Augmented Lagrangians with Optimal Feasibility Convergence for Convex Bound and Equality Constrained Quadratic Programming. *SIAM Journal on Numerical Analysis*. 2005, vol. 43, no. 1, pp. 96–115. ISSN 0036-1429. DOI: 10.1137/S0036142903436393.

[9] DOSTAL, Z. *Optimal Quadratic Programming Algorithms, with Applications to Variational Inequalities*. 1st ed. New York: Springer, 2009. ISBN 978-0-387-84806-8.

[10] DOSTAL, Z., T. BRZOBOHATY, D. HORAK, T. KOZUBEK and P. VODSTRCIL. On R-linear convergence of semi-monotonic inexact augmented Lagrangians for bound and equality constrained

quadratic programming problems with application. *Computers & Mathematics with Applications*. 2014, vol. 67, iss. 3, pp. 515–526. ISSN 0898-1221. DOI: 10.1016/j.camwa.2013.11.009.

[11] DOSTAL, Z., A. FRIEDLANDER and S. A. SANTOS. Augmented Lagrangians with adaptive precision control for quadratic programming with simple bounds and equality constraints. *SIAM Journal on Optimization*. 2003, vol. 13, no. 4, pp. 1120–1140. ISSN 1052-6234. DOI: 10.1137/S1052623499362573.

[12] DOSTAL, Z. and T. KOZUBEK. An optimal algorithm with superrelaxation for minimization of a quadratic function subject to separable constraints with applications. *Mathematical Programming*. 2012, vol. 135, iss. 1–2, pp. 195–220. ISSN 0025-5610. DOI: 10.1007/s10107-011-0454-2.

[13] DOSTAL, Z., T. KOZUBEK, A. MARKOPOULOS, T. BRZOBOHATY, V. VONDRAK and P. HORYL. A theoretically supported scalable TFETI algorithm for the solution of multibody 3D contact problems with friction. *Computer Methods in Applied Mechanics and Engineering*. 2012, vol. 205–208, iss. 1, pp. 110–120. ISSN 0045-7825. DOI: 10.1016/j.cma.2011.02.015.

[14] DOSTAL, Z., T. KOZUBEK, M. SADOWSKA and V. VONDRAK. *Scalable Algorithms for Contact Problems*. 1st ed. New York: Springer, 2017. ISBN 978-1-4939-6832-9.

[15] DOSTAL, Z. and R. KUCERA. An Optimal Algorithm for Minimization of Quadratic Functions with Bounded Spectrum Subject to Separable Convex Inequality and Linear Equality Constraints. *SIAM Journal on Optimization*. 2010, vol. 20, no. 6, pp. 2913–2938. ISSN 1052-6234. DOI: 10.1137/090751414.

[16] FLEMISCH, B., B. WOHLMUTH, F. RAPETTI and Y. MADAY. Scalar and vector potentials' coupling on nonmatching grids for the simulation of an electromagnetic brake. *The international journal for computation and mathematics in electrical and electronic engineering*. 2005, vol. 24, no. 3, pp. 1061–1070. ISSN 0332-1649. DOI: 10.1108/03321640510571246

[17] GILBERT, J. C. and E. JOANNOPOULOS. Convex quadratic optimization solvers using the augmented Lagrangian approach, with an appropriate behavior on infeasible or unbounded problems. HAL archives-ouvertes. *HAL-inria* [online]. 2014. Available at: `https://hal.inria.fr/hal-011103622014`.

[18] GLOWINSKI, R. and P. L. TALLEC. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. 1st ed. Philadelphia: SIAM, 1989. ISBN 978-0-89871-230-8.

[19] HESTENES, M. R. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*. 1969, vol. 4, iss. 5, pp. 303–320. ISSN 0022-3239. DOI: 10.1007/BF00927673.

[20] MARTINEZ J. M. and B. F. SVAITER. A Practical Optimality Condition Without Constraint Qualifications for Nonlinear Programming. *Journal of Optimization Theory and Applications*. 2003, vol. 118, iss. 1, pp. 117–133. ISSN 0022-3239. DOI: 10.1023/A:1024791525441.

[21] POWELL, M. J. D. A method for nonlinear constraints in minimization problems. In: *Optimization*. 1st ed. New York: Academic Press, 1969, pp. 283–298. ISBN 978-0-12260-650-2.

About Authors

Zdenek DOSTAL was born in Olomouc, Czech Republic. He received his M.Sc. from Computer Science in 1970. He finished Ph.D. in 1979 in Mathematical Institute of the Czechoslovak Academy of Science, Prague, associate professor since 1992 from Charles University, Prague, fulll professor at VSB–Technical University of Ostrava since 1994, ad DSc. from the Academy of Science of Czech Republic, Prague 2005. His research interests include numerical linear algebra, optimization, and computational mechanics.

Petr BEREMLIJSKI was born in Ostrava, Czech Republic. He received his M.Sc. from Computer Science in 1999. He finished Ph.D. in 2007 at VSB–Technical University of Ostrava, associate professor since 2016. His research interests include nonsmooth optimization, Clarke and Mordukhovich calculus and shape optimization for contact problems with friction.