ISSN 0965-5425, Computational Mathematics and Mathematical Physics, 2011, Vol. 51, No. 6, pp. 919–941. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © A.F. Izmailov, A.L. Pogosyan, 2011, published in Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 2011, Vol. 51, No. 6, pp. 983–1006.

A Semismooth Sequential Quadratic Programming Method for Lifted Mathematical Programs with Vanishing Constraints

A. F. Izmailov and A. L. Pogosyan

Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119991 Russia
e-mail: izmaf@ccas.ru, pogosyan85@gmail.com

Received November 9, 2010

Abstract—Mathematical programs with vanishing constraints are a difficult class of optimization problems with important applications to optimal topology design problems of mechanical structures. Recently, they have attracted increasingly more attention of experts. The basic difficulty in the analysis and numerical solution of such problems is that their constraints are usually nonregular at the solution. In this paper, a new approach to the numerical solution of these problems is proposed. It is based on their reduction to the so-called lifted mathematical programs with conventional equality and inequality constraints. Special versions of the sequential quadratic programming method are proposed for solving lifted problems. Preliminary numerical results indicate the competitiveness of this approach.

DOI: 10.1134/S0965542511060108

Keywords: mathematical program with vanishing constraints, lifted problem, mathematical program with complementarity constraints, constraint qualifications, optimality conditions, sequential quadratic programming.

1. MATHEMATICAL PROGRAM WITH VANISHING CONSTRAINTS

Consider the *mathematical program with vanishing constraints* (MPVC)

$$\begin{gathered} f(x) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \\ H_i(x) \ge 0, \quad G_i(x)H_i(x) \le 0, \quad i = 1, 2, \dots, s, \end{gathered} \tag{1.1}$$

where $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ is a twice differentiable function, while $h: \mathbb{R}^n \longrightarrow \mathbb{R}^l$, $g: \mathbb{R}^n \longrightarrow \mathbb{R}^m$, and $G, H: \mathbb{R}^n \longrightarrow \mathbb{R}^s$ are twice differentiable mappings. The name introduced in [1] for this class of problems is related to the following fact. Suppose that, for a point $x \in \mathbb{R}^n$, we have $H_i(x) > 0$ for some index $i \in \{1, 2, ..., s\}$. Then, one of the conditions for x to be a feasible point of problem (1.1) is that the constraint $G_i(x) \le 0$ must be satisfied. However, if $H_i(x) = 0$, then the inequality $G_i(x)H_i(x) \le 0$ is automatically fulfilled. Thus, the constraint $G_i(x) \le 0$ vanishes at such a point x.

Examples of applying MPVCs optimal topology design problems of mechanical structures can be found in [1, 2].

Let $\bar{x} \in \mathbb{R}^n$ be a feasible point of problem (1.1). Define the index sets

$$I_g = I_g(\bar{x}) = \{i = 1, 2, \dots, m \mid g_i(\bar{x}) = 0\},$$

$$I_+ = I_+(\bar{x}) = \{i = 1, 2, \dots, s \mid H_i(\bar{x}) > 0\},$$

$$I_0 = I_0(\bar{x}) = \{i = 1, 2, \dots, s \mid H_i(\bar{x}) = 0\},$$

as well as the further partition of I_+ into the subsets

$$I_{+0} = I_{+0}(\bar{x}) = \{i \in I_+ \mid G_i(\bar{x}) = 0\},$$

$$I_{+-} = I_{+-}(\bar{x}) = \{i \in I_+ \mid G_i(\bar{x}) < 0\},$$

and the partition of I_0 into the subsets

$$I_{0+} = I_{0+}(\bar{x}) = \{i \in I_0 \mid G_i(\bar{x}) > 0\},$$

$$I_{00} = I_{00}(\bar{x}) = \{i \in I_0 | G_i(\bar{x}) = 0\},$$

$$I_{0-} = I_{0-}(\bar{x}) = \{i \in I_0 | G_i(\bar{x}) < 0\}.$$

If $I_{00} = \varnothing$, then we say that the *lower-level strict complementarity condition* holds at $\bar{x}$. It was shown in [1] that, if this (rather stringent) condition is violated, then the constraints of problem (1.1) do not satisfy the Mangasarian–Fromovitz constraint qualification at the point $\bar{x}$. This makes MPVCs hard to analyze and solve numerically. Certain results concerning optimality conditions for MPVCs, their sensitivity, and numerical methods exploiting their special structure were obtained in [1, 3–9].

Note that, by introducing an extra variable $u \in \mathbb{R}^s$, problem (1.1) can be reduced to the *mathematical program with complementarity constraints* (MPCC)

$$\begin{gathered} f(x) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \quad G(x) - u \le 0, \\ H(x) \ge 0, \quad u \ge 0, \quad H_i(x)u_i = 0, \quad i = 1, 2, \dots, s. \end{gathered} \tag{1.2}$$

The MPCCs are a relatively well studied class of problems (e.g., see [10–12; 13, Section 4.3]). However, this reduction increases the dimension of the problem and has additionally the following serious drawback: for a given (local) solution $\bar{x}$ to problem (1.1), the corresponding optimal value of the extra variable u is not determined uniquely. Moreover, (local) solutions to problem (1.2) cannot be strict; consequently, no reasonable sufficient optimality conditions can be fulfilled for these solutions, which causes difficulties for the analysis and numerical solution of such problems. Below, we indicate how this drawback can be overcome by a special modification of the objective function of problem (1.2).

Let us recall certain concepts and facts from [1, 6, 7]. With a feasible point $\bar{x}$ of problem (1.1), we associate two auxiliary "conventional" mathematical programming problems, namely, the *relaxed nonlinear program* (RNLP), which has the form

$$f(x) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \quad H_{I_{0+}}(x) = 0, \quad H_{I_{00} \cup I_{0-}}(x) \ge 0, \quad G_{I_{+0}}(x) \le 0, \tag{1.3}$$

and the *tightened nonlinear program* (TNLP), which has the form

$$f(x) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \quad H_{I_{0+} \cup I_{00}}(x) = 0, \quad H_{I_{0-}}(x) \ge 0, \quad G_{I_{+0} \cup I_{00}}(x) \le 0. \tag{1.4}$$

Here, the symbol y_I, where I is a finite set, denotes the subvector of y with the components y_i, $i \in I$.

Now, we define the *MPVC-Lagrangian* of problem (1.1) as

$$\mathcal{L}(x, \mu) = f(x) + \langle \mu^h, h(x) \rangle + \langle \mu^g, g(x) \rangle - \langle \mu^H, H(x) \rangle + \langle \mu^G, G(x) \rangle,$$

where $x \in \mathbb{R}^n$ and $\mu = (\mu^h, \mu^g, \mu^H, \mu^G) \in \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s$.

A feasible point $\bar{x}$ of problem (1.1) is called a *strongly* (*weakly*) *stationary point* of this problem if it is a conventional stationary point of problem (1.3) (respectively, of problem (1.4)). Thus, the weak stationarity implies the existence of $\bar{\mu} = (\bar{\mu}^h, \bar{\mu}^g, \bar{\mu}^H, \bar{\mu}^G) \in \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s$ such that

$$\frac{\partial \mathcal{L}}{\partial x}(\bar{x}, \bar{\mu}) = 0, \tag{1.5}$$

$$\bar{\mu}^g_{I_g} \ge 0, \quad \bar{\mu}^g_{\{1, \dots, m\} \setminus I_g} = 0, \quad \bar{\mu}^H_{I_{0-}} \ge 0, \quad \bar{\mu}^H_{I_+} = 0, \quad \bar{\mu}^G_{I_{+0} \cup I_{00}} \ge 0, \quad \bar{\mu}^G_{I_{+-} \cup I_{0+} \cup I_{0-}} = 0. \tag{1.6}$$

It is such $\bar{\mu}$ that are called Lagrange multipliers of problem (1.4). The strong stationarity implies that, in addition,

$$\bar{\mu}^H_{I_{00}} \ge 0, \quad \bar{\mu}^G_{I_{00}} = 0. \tag{1.7}$$

In this case, $\bar{\mu}$ is called a *MPVC-multiplier* corresponding to the strongly stationary point $\bar{x}$.

We say that the *MPVC-linear independence constraint qualification* is fulfilled at a feasible point $\bar{x}$ of problem (1.1) if the gradients

$$h_i'(\bar{x}), \quad i = 1, 2, \dots, l, \quad g_i'(\bar{x}), \quad i \in I_g, \quad H_i'(\bar{x}), \quad i \in I_0, \quad G_i'(\bar{x}), \quad i \in I_{+0} \cup I_{00}, \tag{1.8}$$

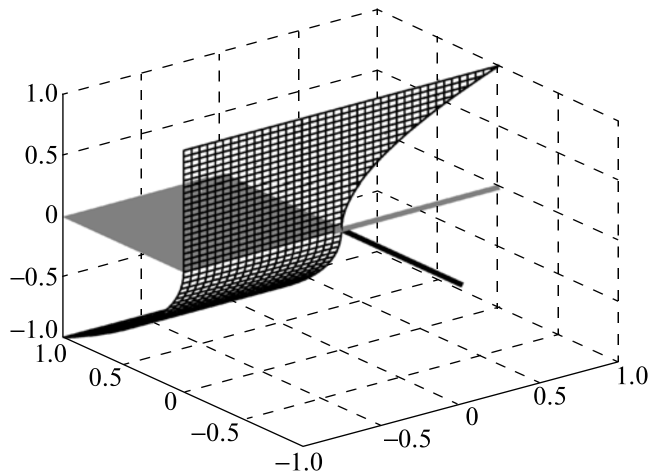

Fig. 1.

are linearly independent. Note that this is nothing else than the conventional linear independence constraint qualification stated for problem (1.4) at the point $\bar{x}$. If this condition is fulfilled, then the local solution $\bar{x}$ to problem (1.1) is a strongly stationary point of this problem.

In what follows, we also need a slightly weaker constraint qualification for problem (1.1). We say that the *MPVC-strict Mangasarian–Fromovitz constraint qualification* is fulfilled at a weakly stationary point $\bar{x}$ of this problem if the traditional strict Mangasarian–Fromovitz constraint qualification holds at $\bar{x}$ for problem (1.4). This implies the uniqueness of the Lagrange multiplier $\bar{\mu}$ that corresponds to $\bar{x}$ in problem (1.4). An equivalent form of this condition expressed in terms of derivatives is given in [7].

Below, we use the piecewise second-order sufficient optimality condition

$$\left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C_2 \backslash \{0\}, \tag{1.9}$$

where

$$C_2 = C_2(\bar{x}) = \{\xi \in \mathbb{R}^n \,|\, h'(\bar{x})\xi = 0, g'_{I_g}(\bar{x})\xi \le 0, H'_{I_{0+}}(\bar{x})\xi = 0, H'_{I_{00} \cup I_{0-}}(\bar{x})\xi \ge 0,$$
$$G'_{I_{+0}}(\bar{x})\xi \le 0, \langle G'_i(\bar{x}), \xi\rangle \langle H'_i(\bar{x}), \xi\rangle \le 0, i \in I_{00}, \langle f'(\bar{x}), \xi\rangle \le 0\}.$$

This sufficient condition is in a natural way associated with the piecewise second-order necessary optimality condition. The index 2 symbolizes the fact that, unlike the conventional critical cone of problem (1.1) at the point $\bar{x}$, the cone C_2 takes into account the second-order information about the last constraint in (1.1).

2. THE LIFTED MPVC

The reformulation of the MPVC discussed in this section is based on the same idea as the reformulation of the MPCC proposed in [14] and used in [15].

Namely, consider the set

$$D = \{(a, b) \in \mathbb{R}^2 \,|\, b \ge 0, ab \le 0\},$$

which is the union of an orthant and a ray. We introduce an extra variable c and, using smooth constraints, specify a set S in the space $\mathbb{R}^3$ of the variables (a, b, c) so that the projection of S on the plane (a, b) coincides with D. For instance, this can be done as follows:

$$S = \{(a, b, c) \in \mathbb{R}^3 \,|\, (\min\{0, c\})^2 = b, a \le (\max\{0, c\})^2\}.$$

Figure 1 shows such a set S (as a mesh) and the corresponding set D.

For $i = 1, 2, \ldots, s$, the pair of constraints $H_i(x) \ge 0$, $G_i(x)H_i(x) \le 0$ can be written as $(H_i(x), G_i(x)) \in D$. The idea is that the latter constraint can be replaced by the inclusion $(H_i(x), G_i(x), y_i) \in S$ with an extra variable y_i. The appearance of this variable explains the name a "lifted MPVC" used for the resulting problem

$$\begin{gathered} f(x) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \\ (\min\{0, y\})^2 - H(x) = 0, \quad G(x) - (\max\{0, y\})^2 \le 0, \end{gathered} \tag{2.1}$$

Here, $y \in \mathbb{R}^s$ is the extra variable. Hereinafter, taking a maximum or a minimum and raising to a power are understood as component-wise operations.

For a feasible point $\bar{x}$ of problem (1.1), the constraints of problem (2.1) do not, in general, uniquely determine the extra variable y. All of its values satisfying the relations

$$y_i \begin{cases} = -(H_i(\bar{x}))^{1/2}, & i \in I_+, \\ \ge (G_i(\bar{x}))^{1/2}, & i \in I_{0+}, \\ \ge 0, & i \in I_{00} \cup I_{0-}. \end{cases} \tag{2.2}$$

are suitable. Moreover, since y does not appear in the objective function of problem (2.1), relations (2.2) imply the following: if $I_0 \ne \varnothing$, then $(\bar{x}, \bar{y})$ cannot be a strict local solution to (2.1) whatever $\bar{y}$ is. This causes difficulties if the lifted problem (2.1) is straightforwardly used for the numerical solution of problem (1.1).

The reason why the solutions to problem (2.1) cannot be strong may be interpreted in one more way. Following [14, 15], we construct the lifted problem for MPCC (1.2):

$$\begin{gathered} f(x) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \\ G(x) - u \le 0, \quad (\min\{0, y\})^2 - H(x) = 0, \quad (\max\{0, y\})^2 - u = 0. \end{gathered} \tag{2.3}$$

The last constraint yields an explicit expression for u; namely,

$$u = (\max\{0, y\})^2. \tag{2.4}$$

If we eliminate u using this expression, then problem (2.3) converts into (2.1). Then, the fact that the local solutions to MPCC (1.2) are unavoidably non-strict ones implies that the local solutions to problem (2.1) cannot be strict as well.

To overcome this difficulty, we can associate (1.2) with the problem

$$\begin{gathered} f(x) + \sum_{i=1}^{s} (u_i^2 - cu_i) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \quad G(x) - u \le 0, \\ H(x) \ge 0, \quad u \ge 0, \quad H_i(x)u_i = 0, \quad i = 1, 2, \ldots, s, \end{gathered} \tag{2.5}$$

where $c > 0$ is a penalty parameter that takes into account unwanted variations of the variable u.

Proposition 1. *Let $\bar{x}$ be a local solution to the original problem* (1.1). *For an arbitrary scalar c that satisfies the inequality*

$$c > \max\{0, 2\max\{G_i(\bar{x}) \mid i \in I_{0+}\}\}, \tag{2.6}$$

define $\bar{u} \in \mathbb{R}^s$ as

$$\bar{u} = \begin{cases} 0, & i \in I_+, \\ c/2, & i \in I_0. \end{cases} \tag{2.7}$$

Then, the point $(\bar{x}, \bar{u})$ is a local solution to problem (2.5).

Proof. Suppose that inequality (2.7) is fulfilled and delete from problem (2.5) the constraints that are known to be inactive at $(\bar{x}, \bar{u})$. Then, using (2.7), we obtain the following "local" version of problem (2.5):

$$f(x) + \sum_{i \in I_0} (u_i^2 - cu_i) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \quad H_{I_0}(x) = 0, \quad G_{I_{+0}}(x) \le 0. \tag{2.8}$$

(For every feasible point (x, u) of problem (2.5) that is close to $(\bar{x}, \bar{u})$, the relation $u_{I_+} = 0$ holds automatically.) Since the constraints of problem (2.8) do not involve u, the minimization over u can be done independently of x, and the minimum of the corresponding summand in the objective function is attained at $u_i = \bar{u}_i = c/2$, $i \in I_0$. This yields the required result. The proposition is proved.

The above proof gives a good reason to expect that, if $\bar{x}$ is a strict local solution to problem (1.1), then the element $\bar{u}$ defined by (2.7) corresponds to the strict local solution $(\bar{x}, \bar{u})$ to problem (2.5).

In view of (2.4), modification (2.5) of MPCC (1.2) corresponds to the following modification of the lifted problem (2.1):

$$\begin{gathered} f_c(x, y) = f(x) + \sum_{i=1}^{s} ((\max\{0, y_i\})^4 - c(\max\{0, y_i\})^2) \longrightarrow \min, \\ h(x) = 0, \quad g(x) \le 0, \quad (\min\{0, y\})^2 - H(x) = 0, \quad G(x) - (\max\{0, y\})^2 \le 0. \end{gathered} \tag{2.9}$$

The following result is an analogue of Proposition 1 for problem (2.9).

Proposition 2. *Let $\bar{x}$ be a local solution to the original problem* (1.1). *For an arbitrary scalar c that satisfies inequality* (2.6), *define $\bar{y} \in \mathbb{R}^s$ as*

$$\bar{y}_i = \begin{cases} -(H_i(\bar{x}))^{1/2}, & i \in I_+, \\ (c/2)^{1/2}, & i \in I_0. \end{cases} \tag{2.10}$$

Then, $(\bar{x}, \bar{y})$ *is a local solution to problem* (2.9).

Proof. The argument is entirely analogous to that used in the proof of Proposition 1, the only distinction being that, instead of (2.8), the following "local" version of problem (2.9) is considered:

$$f(x) + \sum_{i \in I_0} (y_i^4 - cy_i^2) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \quad -H_{I_0}(x) = 0, \quad G_{I_{+0}}(x) \le 0. \tag{2.11}$$

(The components y_i $(i \in I_+)$ are uniquely determined; they are negative and do not affect the value of the objective function of problem (2.9).) The proposition is proved.

The following example demonstrates the main potential drawback of the approach under discussion; namely, even if $\bar{x}$ is a global solution to problem (1.1), the local solution $(\bar{x}, \bar{y})$ to problem (2.9) determined in accordance with (2.10) may not be a global solution no matter how large the chosen $c > 0$ is. Moreover, in this example, the component $\tilde{x}$ of the global solution $(\tilde{x}, \tilde{y})$ to problem (2.9) is not even a local solution to problem (1.1). The MPVC-linear independence constraint qualification is fulfilled at the point $\tilde{x}$, and it is a weakly stationary point; however, it is not strongly stationary.

Example 1. Let $n = 2$, $l = m = 0$, $s = 1$, $f(x) = x_1^2 + (x_2 - 1)^2$, $G(x) = x_1$, and $H(x) = x_2$. The only global solution to the corresponding problem (1.1) is $\bar{x} = (0, 1)$, where $I_{+0} = \{1\}$. Formula (2.10) yields $\bar{y} = -(H(\bar{x}))^{1/2} = -1$; consequently, $f_c(\bar{x}, \bar{y}) = 0$ for all c. On the other hand, problem (2.9) has exactly one additional stationary point $(\tilde{x}, \tilde{y})$, where $\tilde{x} = (0, 0)$ and $\tilde{y} = (c/2)^{1/2}$; moreover, this point is a local solution for all $c > 0$. Furthermore, for $c > 2$, we have $f_c(\tilde{x}, \tilde{y}) = 1 + (c/2)^2 - c^2/2 < 0 = f_c(\bar{x}, \bar{y})$. Thus, it is a point $(\tilde{x}, \tilde{y})$ that is a global solution to problem (2.9).

Nevertheless, the numerical results presented in Section 5 testify that the approach under discussion is promising.

To continue the analysis of the links between problems (1.1) and (2.9), we write the Lagrangian function of the latter problem:

$$L_c(x, y, \lambda) = f(x) + \sum_{i=1}^{s}((\max\{0, y_i\})^4 - c(\max\{0, y_i\})^2)$$
$$+ \langle \lambda^h, h(x)\rangle + \langle \lambda^g, g(x)\rangle + \langle \lambda^H, (\min\{0, y\})^2 - H(x)\rangle + \langle \lambda^G, G(x) - (\max\{0, y\})^2\rangle.$$

Here, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^s$, and $\lambda = (\lambda^h, \lambda^g, \lambda^H, \lambda^G) \in \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s$. Now, we have

$$\frac{\partial L_c}{\partial x}(x, y, \lambda) = \frac{\partial \mathcal{L}}{\partial x}(x, \lambda), \tag{2.12}$$

$$\frac{\partial L_c}{\partial y_i}(x, y, \lambda) = 4(\max\{0, y_i\})^3 - 2c\max\{0, y_i\} + 2\lambda_i^H\min\{0, y_i\} - 2\lambda_i^G\max\{0, y_i\}, \quad i = 1, 2, \ldots, s. \tag{2.13}$$

Proposition 3. *Let $\bar{x}$ be a feasible point of problem* (1.1).

Then, the following assertions are valid for an arbitrary scalar c that satisfies inequality (2.6):

(1) *Let $\bar{x}$ be a weakly stationary point of problem* (1.1), *and let $\bar{\mu}$ be a corresponding Lagrange multiplier of problem* (1.4). *Let $J \subset I_{00} \cup I_{0-}$ be an arbitrary index set such that*

$$J \supset \{i \in I_{00} | \bar{\mu}_i^G > 0\}. \tag{2.14}$$

Then, the point $(\bar{x}, y^J)$, where the vector $y^J \in \mathbb{R}^s$ has the components

$$y_i^J = \begin{cases} -(H_i(\bar{x}))^{1/2}, & i \in I_+, \\ (c/2)^{1/2}, & i \in I_{0+} \cup ((I_{00} \cup I_{0-})\backslash J), \\ 0, & i \in J, \end{cases} \tag{2.15}$$

is a stationary point of problem (2.9) *with $\bar{\lambda} = \bar{\mu}$ as a corresponding Lagrange multiplier.*

(2) *Let $\bar{x}$ be a strongly stationary point of problem* (1.1), *and let $\bar{\mu}$ be a corresponding MPVC-multiplier. Then*, $(\bar{x}, y)$ *is a stationary point of problem* (2.9) *if and only if $y = y^J$ for some set $J \subset I_{00} \cup I_{0-}$. Moreover, for all of these stationary points, a corresponding Lagrange multiplier is $\bar{\lambda} = \bar{\mu}$.*

(3) *Let $(\bar{x}, y)$ be a stationary point of problem* (2.9), *and let $\bar{\lambda}$ be a corresponding Lagrange multiplier. Then, $y = y^J$ for some set $J \subset I_{00} \cup I_{0-}$, and $\bar{\mu} = \bar{\lambda}$ satisfies relation* (1.5), *as well as relations* (1.6) *with the possible exception of the inequality $\bar{\mu}^H_{I_{0-}} \geq 0$.*

(4) *If $J \subset I_{00} \cup I_{0-}$ and $J \neq \varnothing$, then $(\bar{x}, y^J)$ is not even a local solution to problem* (2.9).

Proof. It follows from (2.12) and (2.13) that the Karush–Kuhn–Tucker system (the KKT system), which characterizes the stationary points of problem (2.9) and the corresponding Lagrange multipliers, has the form

$$\begin{gathered} \frac{\partial \mathcal{L}}{\partial x}(x, \lambda) = 0, \\ 2(\max\{0, y_i\})^3 - c(\max\{0, y_i\}) + \lambda_i^H\min\{0, y_i\} - \lambda_i^G\max\{0, y_i\} = 0, \quad i = 1, 2, \ldots, s, \\ h(x) = 0, \\ \lambda^g \geq 0, \quad g(x) \leq 0, \quad \langle \lambda^g, g(x)\rangle = 0, \\ (\min\{0, y\})^2 - H(x) = 0, \\ \lambda^G \geq 0, \quad G(x) - (\max\{0, y\})^2 \leq 0, \quad \langle \lambda^G, G(x) - (\max\{0, y\})^2\rangle = 0. \end{gathered} \tag{2.16}$$

In view of (2.6), for $x = \bar{x}$, this system converts into the system

$$\begin{gathered}
\frac{\partial \mathcal{L}}{\partial x}(\bar{x}, \lambda) = 0, \\
y_{I_+} = -(H_{I_+}(\bar{x}))^{1/2}, \\
y_i = (c/2)^{1/2}, \quad i \in I_{0+}, \quad y_i(2y_i^2 - c - \lambda_i^G) = 0, \quad i \in I_{00}, \quad y_i(2y_i^2 - c) = 0, \quad i \in I_{0-}, \\
\lambda_{I_g}^g \geq 0, \quad \lambda_{\{1,\,\ldots,\,m\}\setminus I_g}^g = 0, \\
\lambda_{I_+}^H = 0, \quad \lambda_{I_{+0}}^G \geq 0, \quad \lambda_{I_{+-} \cup I_{0+} \cup I_{0-}}^G = 0, \\
\lambda_{I_{00}}^G \geq 0, \quad y_{I_{00}} \geq 0, \quad \langle \lambda_{I_{00}}^G, y_{I_{00}} \rangle = 0.
\end{gathered} \tag{2.17}$$

If $J \subset I_{00} \cup I_{0-}$ satisfies condition (2.14), then relations (1.5) and (1.6) imply that, for y^J specified by (2.15), the pair $(y, \lambda) = (y^J, \bar{\mu})$ satisfies system (2.17). This proves assertion (1).

Now, assume that $\bar{x}$ is a strongly stationary point of problem (1.1) and $\bar{\mu}$ is a corresponding MPVC-multiplier. Using additionally relations (1.7), we conclude that, for y^J specified by (2.15), the pair $(y, \lambda) = (y^J, \bar{\mu})$ satisfies system (2.17) for all sets $J \subset I_{00} \cup I_{0-}$. On the other hand, assume that a pair $(y, \lambda) \in \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s)$ satisfies (2.17). Then, setting

$$J = \{i \in I_{00} \cup I_{0-} | y_i = 0\},$$

we obtain $y = y^J$. This proves assertions (2) and (3).

Finally, let $j \in J \subset I_{00} \cup I_{0-}$. For every $t > 0$, define the element $y(t) \in \mathbb{R}^s$ as follows: $y_i(t) = y_i^J$ ($i \in \{1, 2, \ldots, s\}\setminus\{j\}$), $y_j(t) = t > 0 = y_j^J$. Then, $(\bar{x}, y(t))$ is a feasible point of problem (2.9); moreover,

$$f_c(\bar{x}, y(t)) - f_c(\bar{x}, y^J) = (y_j(t))^4 - c(y_j(t))^2 = t^4 - ct^2 < 0$$

for all sufficiently small $t > 0$. Thus, $(\bar{x}, y^J)$ cannot be a local solution to problem (2.9), which justifies assertion (4). The proposition is proved.

In Example 1, the point $\tilde{x} = (0, 0)$, which corresponds to a global solution to problem (2.9), is at least a weakly (though not strongly) stationary point of problem (1.1). The following example shows that, for a stationary point $(\bar{x}, \bar{y})$ of problem (2.9), its component $\bar{x}$ may even not be a weakly stationary point for problem (1.1).

Example 2. Let $n = 2$, $l = m = 0$, $s = 1$, $f(x) = (x_1 + 1)^2 + (x_2 - 1)^2$, $G(x) = x_1$, and $H(x) = x_2$. Consider the point $\bar{x} = (-1, 0)$ for which $I_{0-} = \{1\}$. This is not a weakly stationary point for problem (1.1) because relation (1.5) is satisfied only by $\bar{\mu}^H = -2 < 0$ and $\bar{\mu}^G = 0$; thus, the condition $\bar{\mu}_{I_{0-}}^H \geq 0$ in (1.6) is violated. However, the corresponding lifted problem (2.9), that is, the problem

$$\begin{gathered}
f_c(x, y) = (x_1 + 1)^2 + (x_2 - 1)^2 + (\max\{0, y\})^4 - c(\max\{0, y\})^2 \longrightarrow \min, \\
(\min\{0, y\})^2 - x_2 = 0, \quad x_1 - (\max\{0, y\})^2 \leq 0
\end{gathered}$$

has the stationary points $(\bar{x}, \bar{y}^1)$ and $(\bar{x}, \bar{y}^2)$, where $\bar{y}^1 = 0$ and $\bar{y}^2 = (c/2)^{1/2}$; moreover, $\bar{\lambda}^H = \bar{\mu}^H$, and $\bar{\lambda}^G = \bar{\mu}^G$. Furthermore, for $c \geq 2$, the point $(\bar{x}, \bar{y}^2)$ is a global solution to problem (2.9). (For $0 < c \leq 2$, this problem has a different global solution $(\tilde{x}, \tilde{y})$, where $\tilde{x} = (-1, 1)$ is a global solution to problem (1.1) and $\tilde{y} = -1$.)

Thus, problem (2.9) can have parasitic stationary points $(\bar{x}, \bar{y})$ some of which may be its local (or even global) solutions. Note, however, that, by assertion (3) of Proposition 3, the corresponding points $\bar{x}$ satisfy a condition that is only slightly weaker than weak stationarity (only the inequality $\bar{\mu}_{I_{0-}}^H \geq 0$ in (1.6) is violated). In fact, weak stationarity is here a fairly strong concept of stationarity (see [9]). It is also clear that one cannot hope for more by transiting to a lifted problem.

3. SEMISMOOTH SEQUENTIAL QUADRATIC PROGRAMMING METHOD FOR THE LIFTED MPVC

To solve problem (2.9), we propose the following variant of the sequential quadratic programming (SQP) method. Let (x^k, y^k, λ^k) be the current approximation. Here, $x^k \in \mathbb{R}^n$, $y^k \in \mathbb{R}^s$, and $\lambda^k = ((\lambda^h)^k, (\lambda^g)^k, (\lambda^H)^k, (\lambda^G)^k) \in \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s$. The next primal approximation $(x^{k+1}, y^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^s$ is sought as a stationary point of the quadratic program

$$\begin{gathered}
\left\langle f_c'(x^k, y^k), \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix} \right\rangle + \frac{1}{2}\left\langle \mathcal{H}_k \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix}, \begin{pmatrix} x - x^k \\ y - y^k \end{pmatrix} \right\rangle \longrightarrow \min, \\
h(x^k) + h'(x^k)(x - x^k) = 0, \quad g(x^k) + g'(x^k)(x - x^k) \le 0, \\
(\min\{0, y^k\})^2 - H(x^k) - H'(x^k)(x - x^k) + 2B_{\min}(y^k)(y - y^k) = 0, \\
G(x^k) - (\max\{0, y^k\})^2 + G'(x^k)(x - x^k) - 2B_{\max}(y^k)(y - y^k) \le 0,
\end{gathered} \tag{3.1}$$

while the next dual approximation $\lambda^{k+1} \in \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s$ is sought as the Lagrange multiplier corresponding to this stationary point. Here, $\mathcal{H}_k$ is a symmetric $(n + s) \times (n + s)$ matrix, and, for $y \in \mathbb{R}^s$, we set

$$B_{\min}(y) = \operatorname{diag}(\min\{0, y\}), \quad B_{\max}(y) = \operatorname{diag}(\max\{0, y\}),$$

where, for every $z \in \mathbb{R}^s$, the symbol $\operatorname{diag}(z)$ denotes the diagonal $s \times s$ matrix with z on the principal diagonal. Observe that

$$f_c'(x^k, y^k) = (f'(x^k), 4(\max\{0, y^k\})^3 - 2c\max\{0, y^k\}).$$

In the basic sequential quadratic programming method (see [16, Section 4.4]), $\mathcal{H}_k$ is chosen as the Hessian matrix of the Lagrangian function of a problem to be solved. However, in the case under discussion, such a choice is impossible because both the objective function and the constraints of problem (2.9) can only once be differentiated. On the other hand, for this problem, the derivatives of the objective function and the constraints are semismooth functions. Consequently, instead of the basic choice of $\mathcal{H}_k$, we use the rule

$$\mathcal{H}_k \in (\partial_B)_{(x,y)} \frac{\partial L_c}{\partial(x, y)}(x^k, y^k, \lambda^k),$$

where the expression on the right-hand side is the so-called B-differential of the gradient mapping $(x, y) \longrightarrow \dfrac{\partial L_c}{\partial(x, y)}(x, y, \lambda^k) : \mathbb{R}^n \times \mathbb{R}^s \longrightarrow \mathbb{R}^n \times \mathbb{R}^s$ (the definitions of the semismoothness and B-differential can be found, for instance, in [16, Section 4.5]). Methods of this type as applied to mathematical programs with equality constraints (in particular, to lifted mathematical programs with complementary constraints) were studied in [17, 18], as well as in a recent report [19]. In the case under discussion, the B-differential can be calculated explicitly using (2.12) and (2.13). It consists of the matrices

$$\mathcal{H}_k = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial x^2}(x^k, \lambda^k) & 0 \\ 0 & 2\operatorname{diag}(a_c(y^k, \lambda^k)) \end{pmatrix}, \tag{3.2}$$

where, for $y \in \mathbb{R}^s$ and $\lambda \in \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s$, we set

$$a_c(y, \lambda) = 6(\max\{0, y\})^2 + b_c(y, \lambda), \tag{3.3}$$

$$(b_c(y, \lambda))_i = \begin{cases} \lambda_i^H, & y_i < 0, \\ \lambda_i^H \text{ или } -\lambda_i^G - c, & y_i = 0, \\ -\lambda_i^G - c, & y_i > 0, \end{cases} \quad i = 1, 2, \dots, s. \tag{3.4}$$

The situation is simpler when the local behavior of the above semismooth SQP method is analyzed because it reduces to the conventional SQP method. Let $\bar{x}$ be a strongly stationary point of problem (1.1), and let $\bar{\mu} = (\bar{\mu}^h, \bar{\mu}^g, \bar{\mu}^H, \bar{\mu}^G)$ be the corresponding MPVC-multiplier. Define $\bar{y}$ according to (2.10). If the element $y^k \in \mathbb{R}^s$ is close to $\bar{y}$, then (3.4) implies the equality

$$(b_c(y^k, \lambda^k))_i = \begin{cases} (\lambda_i^H)^k, & i \in I_+, \\ -(\lambda_i^G)^k - c, & i \in I_0. \end{cases} \tag{3.5}$$

In view of (3.2) and (3.3), iterative subproblem (3.1) is identical to an iterative subproblem of the basic SQP method as applied to the problem

$$\begin{gathered} f(x) + \sum_{i \in I_0} (y_i^4 - cy_i^2) \longrightarrow \min, \quad h(x) = 0, \quad g(x) \le 0, \\ y_{I_+}^2 - H_{I_+}(x) = 0, \quad -H_{I_0}(x) = 0, \quad G_{I_+}(x) \le 0, \quad G_{I_0}(x) - y_{I_0}^2 \le 0, \end{gathered} \tag{3.6}$$

where both the objective function and the constraints are twice differentiable. The constraint $y_{I_+}^2 - H_{I_+}(x) = 0$ is only needed to determine the components y_i ($i \in I_+$), which do not appear at any other place in problem (3.6). Consequently, this constraint can be dropped, as is the case with the corresponding linearized constraint in a subproblem of the SQP method. If, in addition, we drop the constraints of problem (3.6) that are known to be inactive at the point $(\bar{x}, \bar{y})$, then problem (2.11) used in the proof of Proposition 2 is obtained. The above analysis implies the following: for the method under discussion, the local superlinear convergence to the stationary point $(\bar{x}, \bar{y})$ of problem (2.9) and the corresponding Lagrange multiplier $\bar{\mu}$ (see assertion (2) in Proposition 3) can be justified under the same assumptions as the local superlinear convergence of the basic SQP method to the stationary point $(\bar{x}, \bar{y}_{I_0})$ of problem (2.11) and the corresponding Lagrange multiplier $(\bar{\mu}^h, \bar{\mu}^g, \bar{\mu}_{I_0}^H, \bar{\mu}_{I_{+0}}^G)$.

The weakest conditions of this sort were obtained in [20]. In addition to the continuity of the second derivatives of the objective function and the constraints at the desired stationary point, they include the strict Mangasarian–Fromovitz constraint qualification and the second-order sufficient optimality condition. For problem (2.11), the first of these conditions amounts to the uniqueness of the Lagrange multiplier $(\bar{\mu}^h, \bar{\mu}^g, \bar{\mu}_{I_0}^H, \bar{\mu}_{I_{+0}}^G)$ that corresponds to the stationary point $(\bar{x}, \bar{y}_{I_0})$. This condition is certainly fulfilled automatically if the gradients

$$h_i'(\bar{x}), \quad i = 1, 2, \dots, l, \quad g_i'(\bar{x}), \quad i \in I_g, \quad H_i'(\bar{x}), \quad i \in I_0, \quad G_i'(\bar{x}), \quad i \in I_{+0} \tag{3.7}$$

are linearly independent.

Proposition 4. *Let $\bar{x}$ be a strongly stationary point of problem* (1.1). *Assume that the scalar c satisfies inequality* (2.6), *while $\bar{y}$ is specified in accordance with* (2.10).

Then, the following assertions are valid:

(1) *Suppose that the MPVC-linear independence constraint qualification* (1.8) *is fulfilled at $\bar{x}$. Then, the linear independence constraint qualification for the gradients in* (3.7) *is also fulfilled. Consequently, the strict Mangasarian–Fromovitz constraint qualification holds at the stationary point $(\bar{x}, \bar{y}_{I_0})$ of problem* (2.11).

(2) *Suppose that the strict Mangasarian–Fromovitz constraint qualification is fulfilled at the stationary point* $(\bar{x}, \bar{y}_{I_0})$ *of problem* (2.11). *Then, the MPVC-strict Mangasarian–Fromovitz constraint qualification holds at* $\bar{x}$.

Proof. Assertion (1) is obvious because the collection of gradients (1.8) appearing in the definition of the MPVC-linear independence constraint qualification contains the gradients of system (3.7).

To prove assertion (2), we first observe that, if $\bar{\lambda}$ is the Lagrange multiplier corresponding to the stationary point $(\bar{x}, \bar{y})$ of problem (2.9), then, by assertion (3) of Proposition 3, we have $\bar{\lambda}^H_{I_+} = 0$. The immediate consequence is that the uniqueness of the Lagrange multiplier $(\bar{\lambda}^h, \bar{\lambda}^g, \bar{\lambda}^H_{I_0}, \bar{\lambda}^G_{I_{+0}})$ corresponding to the stationary point $(\bar{x}, \bar{y}_{I_0})$ of problem (2.11) implies the uniqueness of the Lagrange multiplier $\bar{\lambda} = (\bar{\lambda}^h, \bar{\lambda}^g, \bar{\lambda}^H, \bar{\lambda}^G)$ that corresponds to the stationary point $(\bar{x}, \bar{y})$ of problem (2.9). Indeed, for the latter multiplier, it must hold $\bar{\lambda}^H_{I_+} = 0$ and $\bar{\lambda}^G_{I_{+,-} \cup I_0} = 0$, while the other components are identical to the corresponding components of the multiplier for problem (2.11). Finally, by assertion (1) of Proposition 3, the uniqueness of the multiplier corresponding to the stationary point $(\bar{x}, \bar{y})$ of problem (2.9) entails the uniqueness of the Lagrange multiplier for the stationary point $\bar{x}$ of TNLP (1.4). This is exactly the MPVC-strict Mangasarian–Fromovitz constraint qualification. The proposition is proved.

The following example demonstrates that, in general, the reverse implication to that in assertion (2) of Proposition 4 is false.

Example 3. Let $n = 2$, $l = m = 0$, $s = 2$, $f(x) = (x_1 + 1)^2 + x_2^2$, $G(\cdot) \equiv (-1, -1)$, and $H(x) = (x_2, x_2)$. The only solution to problem (1.1), which is a local and a global solution at the same time, is the point $\bar{x} = (-1, 0)$, for which $I_{0-} = \{1, 2\}$. The corresponding TNLP is

$$f(x) = (x_1 + 1)^2 + x_2^2 \longrightarrow \min, \quad x_2 \geq 0, \quad x_2 \geq 0$$

(the constraint is repeated). The Lagrange multipliers corresponding to $\bar{x} = (-1, 0)$ in this TNLP are specified by the system

$$\mu_1^H + \mu_2^H = 0, \quad \mu_1^H \geq 0, \quad \mu_2^H \geq 0,$$

which has the unique solution $\bar{\mu}^H = 0$. Thus, the MPVC-strict Mangasarian–Fromovitz constraint qualification holds at the point $\bar{x} = (-1, 0)$.

On the other hand, consider the corresponding lifted problem (2.9):

$$(x_1 + 1)^2 + x_2^2 + (\max\{0, y_1\})^4 + (\max\{0, y_2\})^4 - c((\max\{0, y_1\})^2 + (\max\{0, y_2\})^2) \longrightarrow \min,$$

$$(\min\{0, y_1\})^2 - x_2 = 0, \quad (\min\{0, y_2\})^2 - x_2 = 0,$$

$$-1 - (\max\{0, y_1\})^2 \leq 0, \quad -1 - (\max\{0, y_2\})^2 \leq 0.$$

For every $c > 0$, the Lagrange multipliers at the point $(\bar{x}, \bar{y})$, where $\bar{y} = ((c/2)^{1/2}, (c/2)^{1/2})$ is calculated in accordance with (2.10), are specified by the system

$$\lambda_1^H + \lambda_2^H = 0, \quad \lambda_1^G = 0, \quad \lambda_2^G = 0,$$

which is nonuniquely solvable. Consequently, the MPVC-strict Mangasarian–Fromovitz constraint qualification does not hold for the lifted problem.

In view of (2.10), the second-order sufficient optimality condition for problem (2.11) has the form

$$0 < \left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})\xi, \xi \right\rangle + \sum_{i \in I_0} (12\bar{y}_i^2 - 2c)\eta_i^2 = \left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})\xi, \xi \right\rangle + 4c \sum_{i \in I_0} \eta_i^2 \quad \forall (\xi, \eta) \in K \backslash \{0\}, \tag{3.8}$$

where

$$K = K(\bar{x}) = \left\{(\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^s \,\middle|\, h'(\bar{x})\xi = 0, g'_{I_g}(\bar{x})\xi \le 0, H'_{I_0}(\bar{x})\xi = 0, G'_{I_{+0}}(\bar{x})\xi \le 0, \right.$$
$$\left. \langle f'(\bar{x}), \xi\rangle + \sum_{i \in I_0} (4\bar{y}_i^3 - 2c\bar{y}_i)\eta_i \le 0 \right\} = \tilde{C} \times \mathbb{R}^s, \tag{3.9}$$

$$\tilde{C} = \tilde{C}(\bar{x}) = \{\xi \in \mathbb{R}^n \,|\, h'(\bar{x})\xi = 0, g'_{I_g}(\bar{x})\xi \le 0, H'_{I_0}(\bar{x})\xi = 0, G'_{I_{+0}}(\bar{x})\xi \le 0, \langle f'(\bar{x}), \xi\rangle \le 0\} \subset C_2. \tag{3.10}$$

It is obvious that this sufficient condition is equivalent to the relation

$$\left\langle \frac{\partial^2 \mathcal{L}}{\partial x^2}(\bar{x}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in \tilde{C}\backslash\{0\}.$$

This relation, combined with the inclusion in (3.10), immediately yields the following result.

Proposition 5. *Let* $\bar{x}$ *be a strongly stationary point of problem* (1.1), *and let* $\bar{\mu}$ *be a corresponding MPVC-multiplier. Moreover, let the piecewise second-order sufficient optimality condition* (1.9) *be fulfilled. Assume that the scalar c satisfies inequality* (2.6), *while* $\bar{y}$ *is specified in accordance with* (2.10).

Then, the Lagrange multiplier $\bar{\mu}$ *corresponding to the stationary point* $(\bar{x}, \bar{y})$ *of problem* (2.9) *satisfies the second-order sufficient optimality condition* (3.8).

Combining assertion (1) of Proposition 4, Proposition 5, and the results of paper [20], we obtain the following characterization of the local superlinear convergence of the semismooth SQP method for problem (2.9). (We recall once more that the locally semismooth SQP method for problem (2.9) is identical to the basic SQP method for problem (2.11).)

Theorem 1. *Let the function f and the mappings h, g, H, and G be twice differentiable in some neighborhood of a strongly stationary point* $\bar{x}$ *of problem* (1.1), *and let their second derivatives be continuous at this point. Assume that the MPVC-linear independence constraint qualification* (1.8) *and the piecewise second-order sufficient optimality condition* (1.9) *are fulfilled at* $\bar{x}$ *for the* (*unique*) *MPVC-multiplier* $\bar{\mu} = (\bar{\mu}^h, \bar{\mu}^g, \bar{\mu}^H, \bar{\mu}^G)$ *that corresponds to* $\bar{x}$. *Assume that the scalar c satisfies inequality* (2.6), *while* $\bar{y}$ *is specified in accordance with* (2.10).

Then, there is a scalar $\delta > 0$ *such that, for an arbitrary choice of the matrices* $\mathcal{H}_k$ *satisfying* (3.2)–(3.4) *and for every initial approximation* $(x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s)$ *sufficiently close to* $(\bar{x}, \bar{y}, \bar{\mu})$, *there exists a sequence* $\{(x^k, y^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s)$ *with the following properties: for every k, the point* (x^{k+1}, y^{k+1}) *is a stationary point of problem* (3.1), *while* λ^{k+1} *is the corresponding Lagrange multiplier. Moreover, this sequence satisfies the inequality*

$$\|(x^{k+1} - x^k, y^{k+1} - y^k, \lambda^{k+1} - \lambda^k)\| < \delta ,$$

and every such sequence converges to $(\bar{x}, \bar{y}, \bar{\mu})$ *at a superlinear rate. If, in addition, the second derivatives of f and the mappings h, g, H, and G are locally Lipschitzian with respect to* $\bar{x}$, *then the convergence rate is quadratic.*

Note that, as was said above, the MPVC-linear independence constraint qualification (1.8) can be somewhat relaxed for this theorem; namely, it can be replaced by the strict Mangasarian–Fromovitz constraint qualification at the stationary point $(\bar{x}, \bar{y}_{I_0})$ of problem (2.11).

In [7], the authors proposed the so-called piecewise SQP method for solving problem (1.1). Its local superlinear convergence was justified under the assumption that the MPVC-strict Mangasarian–Fromovitz constraint qualification and the piecewise second-order sufficient optimality condition (1.9) are fulfilled. From assertion (2) of Proposition 4 and Example 3, we see that these conditions are somewhat weaker than what we need for the semismooth SQP method under discussion. For the active set methods proposed in the same paper [7], the local superlinear convergence was proved using again the MPVC-strict Mangasarian–Fromovitz constraint qualification; however, the piecewise second-order sufficient optimality condition (1.9) was replaced by a slightly stronger assumption, namely, by the conventional second-order sufficient optimality condition. Moreover, the piecewise SQP method and the active set methods lack an obvious ready-made globalization strategy, whereas a reasonable and natural globalization strategy for the semismooth SQP method as applied to problem (2.9) is described in Section 4.

The globalization strategy under discussion is based on the linesearch as applied to the exact penalty function for problem (2.9). However, the direction $p^k = (x^{k+1} - x^k, y^{k+1} - y^k)$, where (x^{k+1}, y^{k+1}) is a stationary point of problem (3.1), is guaranteed to be a descent direction for this penalty function at (x^k, y^k) only in the case where the matrix $\mathcal{H}_k$ is positive definite (e.g., see [16, Lemma 5.4.1]). The matrices $\mathcal{H}_k$ calculated according to (3.2)–(3.4) cannot be positive definite. These matrices can be replaced by their true quasi-Newton approximations, which, however, destroys the useful diagonal structure inherent in $\mathcal{H}_k$. Below, we present a different method for obtaining positive definite modifications of $\mathcal{H}_k$. It uses the structure of these matrices in much the same way as this was done in [19] for MPCCs.

The idea is to use the conventional quasi-Newton formulas only for the upper left block $\frac{\partial^2 \mathcal{L}}{\partial x^2}(x^k, \lambda^k)$ in (3.2). The diagonal entries of the lower right block $a_c(y^k, \lambda^k)$ are replaced (if required) by positive scalars so that the perturbation of the principal diagonal in $a_c(y^k, \lambda^k)$ asymptotically vanishes. Namely, $\mathcal{H}_k$ is now constructed as

$$\mathcal{H}_k = \begin{pmatrix} H_k & 0 \\ 0 & 2\operatorname{diag}(a^k) \end{pmatrix}, \tag{3.11}$$

where H_k is a symmetric positive definite $n \times n$ matrix provided by the quasi-Newton approximation of $\frac{\partial^2 \mathcal{L}}{\partial x^2}(x^k, \lambda^k)$, while the vector $a^k \in \mathbb{R}^s$ is calculated using the formula

$$a_i^k = \min\{\max\{(a_c(y^k, \lambda^k))_i, \rho(\sigma_c(x^k, y^k, \lambda^k))\}, M\}, \quad i = 1, 2, \ldots, s. \tag{3.12}$$

Here, the function $\rho : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is such that $\rho(t)$ is separated from zero by a positive constant if t is separated from zero and $\rho(t) \longrightarrow 0$ as $t \longrightarrow 0+$; $\sigma_c : \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m_+ \times \mathbb{R}^s \times \mathbb{R}^s_+) \longrightarrow \mathbb{R}$ is a certain residual of the KKT system (2.16) for problem (2.9) (that is, the function σ_c vanishes at the points that satisfy (2.16) and is positive at the other points); and $M > 0$ is an upper bound for the components of a^k (a reasonable choice is some "large" number). (The existence of such a bound is required in Theorem 3 on the global convergence, while the condition $M > 2c$ is sufficient to preserve the superlinear convergence in Theorem 2.)

Define the mapping $\Psi_c : \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s) \longrightarrow \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^l \times \mathbb{R}^s$ by the formula

$$\Psi_c(x, y, \lambda) = \begin{pmatrix} \frac{\partial \mathcal{L}}{\partial x}(x, \lambda) \\ \frac{\partial L_c}{\partial y}(x, y, \lambda) \\ h(x) \\ (\min\{0, y\})^2 - H(x) \end{pmatrix}.$$

For the KKT system (2.16), we use the residuals given below (there are certainly a number of other choices for the function σ_c). Let $x \in \mathbb{R}^n$, $y \in \mathbb{R}^s$, and $\lambda \in \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s$. Define

$$\bar{\sigma}_c(x, y, \lambda) = \left\| \begin{pmatrix} \Psi_c(x, y, \lambda) \\ \max\{0, g(x)\} \\ \max\{0, G(x) - (\max\{0, y\})^2\} \\ \langle \lambda^g, g(x) \rangle + \langle \lambda^G, G(x) - (\max\{0, y\})^2 \rangle \end{pmatrix} \right\|,$$

$$\tilde{\sigma}_c(x, y, \lambda) = \left\|\begin{pmatrix} \Psi_c(x, y, \lambda) \\ \max\{0, g(x)\} \\ (\lambda_1^g g_1(x), \dots, \lambda_m^g g_m(x)) \\ \max\{0, G(x) - (\max\{0, y\})^2\} \\ (\lambda_1^G(G_1(x) - (\max\{0, y_1\})^2), \dots, \lambda_s^G(G_s(x) - (\max\{0, y_s\})^2)) \end{pmatrix}\right\|,$$

$$\sigma_c^{\omega}(x, y, \lambda) = \left\|\begin{pmatrix} \Psi_c(x, y, \lambda) \\ \omega(\lambda^g, -g(x)) \\ \omega(\lambda^G, (\max\{0, y\})^2 - G(x)) \end{pmatrix}\right\|,$$

where $\omega : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ is a complementarity function applied componentwise; that is, the equality $\omega(a, b) = 0$ holds for this function only if $a \ge b$, $b \ge 0$, and $ab = 0$. Two most frequently used complementarity functions are the natural residual $\omega(a, b) = \min\{a, b\}$ and the Fischer–Burmeister function $\omega(a, b) = a + b - \sqrt{a^2 + b^2}$. The symbols σ_c^{NR} (from "Natural Residual") and σ_c^{FB} are used for these two choices of the function σ_c^{ω}.

For the numerical experiments discussed in Section 5, we used the residual σ_c^{FB}. The matrices H_k were calculated using the quasi-Newton Broyden–Fletcher–Goldfarb–Shanno (BFGS) formula with the modification proposed by Powell (see [21, pp. 536, 537]).

Suppose that H_k is a positive definite matrix and $\sigma_c(x^k, y^k, \lambda^k) > 0$; that is, the current point does not satisfy the KKT system (2.16) for problem (2.9). It is then obvious that the matrix $\mathcal{H}_k$ determined by formulas (3.11) and (3.12) is positive definite. At the same time, the following theorem shows that the above modification preserves the superlinear convergence with respect to the primal variables. By π_S, we denote the projection operator on a closed convex set S.

Theorem 2. *Let the function f and the mappings h, g, H, and G be twice differentiable in some neighborhood of a strongly stationary point* $\bar{x}$ *of problem* (1.1), *and let their second derivatives be continuous at this point. Let* $\bar{\mu}$ *be an MPVC-multiplier that corresponds to* $\bar{x}$. *Assume that the scalar c satisfies inequality* (2.6), $M > 2c$, *and* $\bar{y}$ *is specified in accordance with* (2.10). *Let the function* $\rho\colon \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ *be such that* $\rho(t)$ *is separated from zero by a positive constant if t is separated from zero and* $\rho(t) \longrightarrow 0$ *as* $t \longrightarrow 0+$, *and let* $\sigma_c : \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}_+^m \times \mathbb{R}^s \times \mathbb{R}_+^s) \longrightarrow \mathbb{R}$ *be a certain residual of the KKT system* (2.16) *for problem* (2.9). *Assume that the sequence* $\{H_k\}$ *consists of symmetric* $n \times n$ *matrices and the sequence* $\{(x^k, y^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s)$ *converges to* $(\bar{x}, \bar{y}, \bar{\mu})$. *Let* $\mathcal{H}_k$ *be the matrix calculated by formulas* (3.11) *and* (3.12) *for all sufficiently large k. Assume that* (x^{k+1}, y^{k+1}) *is a stationary point of problem* (3.1) *and* λ^{k+1} *is a Lagrange multiplier corresponding to this stationary point.*

If the sequence $\{(x^k, y^k)\}$ *converges at a superlinear rate, then the following analogue of the Dennis–More condition holds*:

$$\left\|\pi_{\tilde{C}}\left(\left(H_k - \frac{\partial^2 \mathcal{L}}{\partial x^2}(x^k, \lambda^k)\right)(x^{k+1} - x^k)\right)\right\| = o\left(\left\|\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix}\right\|\right). \tag{3.13}$$

Conversely, assume that the piecewise second-order sufficient optimality condition (1.9) *and the analog of the Dennis–More condition given by* (3.13) *are valid. Then, the sequence* $\{(x^k, y^k)\}$ *converges at a superlinear rate.*

Proof. Since the sequence $\{y^k\}$ converges to $\bar{y}$, formulas (2.10) imply that (3.5) is fulfilled for all sufficiently large k. Using this fact, the convergence of $\{\lambda^k\}$ to $\bar{\mu}$, and relations (1.6) and (1.7), we find that

$$\{(b_c(y^k, \lambda^k))_{I_+}\} \longrightarrow \bar{\mu}_{I_+}^H = 0, \quad \{(b_c(y^k, \lambda^k))_{I_0}\} \longrightarrow -\bar{\mu}_{I_0}^G - c = -c \quad \text{as} \quad k \longrightarrow \infty.$$

Hence, by (3.3), we have

$$\{(a_c(y^k, \lambda^k))_{I_+}\} \longrightarrow 0, \quad \{(a_c(y^k, \lambda^k))_{I_0}\} \longrightarrow 2c \quad \text{as} \quad k \longrightarrow \infty, \tag{3.14}$$

It follows that

$$\liminf_{k \to \infty}(a_c(y^k, \lambda^k))_i \geq 0 \quad \forall i = 1, 2, \ldots, s. \tag{3.15}$$

Furthermore, using (3.14), the limit relation $\rho(\sigma_c(x^k, y^k, \lambda^k)) \longrightarrow 0$ as $k \longrightarrow \infty$, and the inequality $M > 2c$, we conclude that, for all sufficiently large k,

$$\max\{(a_c(y^k, \lambda^k))_i, \rho(\sigma_c(x^k, y^k, \lambda^k))\} < M \quad \forall i = 1, 2, \ldots, s.$$

In view of (3.12), this yields

$$a_i^k = \max\{(a_c(y^k, \lambda^k))_i, \rho(\sigma_c(x^k, y^k, \lambda^k))\}, \quad i = 1, 2, \ldots, s. \tag{3.16}$$

Now, we show that

$$\lim_{k \to \infty}(a_i^k - (a_c(y^k, \lambda^k))_i) = 0 \quad \forall i = 1, 2, \ldots, s. \tag{3.17}$$

Fix $i \in \{1, 2, ..., s\}$. If $a_i^k = (a_c(y^k, \lambda^k))_i$ for all sufficiently large k, then (3.17) is obviously fulfilled. Otherwise, we define a subsequence $\{k_j\}$ such that $a_i^{k_j} \neq (a_c(y_j^k, \lambda^{k_j}))_i$. From (3.16), we derive the inequality

$$\liminf_{k \to \infty}(a_i^k - (a_c(y^k, \lambda^k))_i) \geq 0,$$

as well as the relations

$$\limsup_{k \to \infty}(a_i^k - (a_c(y^k, \lambda^k))_i) = \limsup_{j \to \infty}(a_i^{k_j} - (a_c(y^{k_j}, \lambda^{k_j}))_i) =$$

$$= \limsup_{j \to \infty}(\rho(\sigma_c(x^{k_j}, y^{k_j}, \lambda^{k_j})) - (a_c(y^{k_j}, \lambda^{k_j}))_i) = \limsup_{j \to \infty}(-(a_c(y^{k_j}, \lambda^{k_j}))_i) =$$

$$= -\liminf_{j \to \infty}(a_c(y^{k_j}, \lambda^{k_j}))_i \leq 0.$$

Here, the third equality is justified by the limit relation $\sigma_c(x^k, y^k, \lambda^k) \longrightarrow 0$ as $k \longrightarrow \infty$ and by the properties of the function ρ, while the inequality follows from (3.15). Thus, relation (3.17) is proved.

According to (3.2), (3.9), (3.11), and (3.17), condition (3.13) is equivalent to the relation

$$\left\|\pi_K\left((\mathcal{H}_k - \mathcal{W}_k)\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix}\right)\right\| = o\left(\left\|\begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix}\right\|\right). \tag{3.18}$$

where

$$\mathcal{W}_k = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial x^2}(x^k, \lambda^k) & 0 \\ 0 & 2\operatorname{diag}(a_c(y^k, \lambda^k)) \end{pmatrix},$$

and, for all sufficiently large k, the vector $a_c(y^k, \lambda^k)$ is uniquely determined by formulas (3.3) and (3.4).

The above discussion shows that the quasi-Newton semismooth SQP method for problem (2.9) analyzed in this section can be interpreted as a conventional quasi-Newton SQP method for problem (3.6); moreover, (3.18) is the Dennis–More condition for the latter method. Now, the desired result follows from [22, Theorem 4.1] and Proposition 5.

4. GLOBALIZATION OF CONVERGENCE

Define the l_1-exact penalty $\psi : \mathbb{R}^n \times \mathbb{R}^s \longrightarrow \mathbb{R}$ for the constraints of problem (2.9) by the formula

$$\psi(x, y) = \|h(x)\|_1 + \|(\min\{0, y\})^2 - H(x)\|_1$$
$$+ \sum_{j=1}^{m} \max\{0, g_j(x)\} + \sum_{i=1}^{s} \max\{0, G_i(x) - (\max\{0, y_i\})^2\}.$$

With this penalty, we associate the family of penalty functions $\varphi_{c,\beta} : \mathbb{R}^n \times \mathbb{R}^s \longrightarrow \mathbb{R}$ given by the formula

$$\varphi_{c,\beta}(x, y) = f_c(x, y) + \beta\psi(x, y),$$

where $\beta > 0$ is the penalty parameter. The following algorithm combines the special quasi-Newton SQP method proposed in Section 3 with the linesearch applied to the penalty functions from the above family. (This is a traditional technique for globalizing the convergence of SQP methods; see [16, Algorithm 5.4.1].)

Algorithm 1

Preliminary step. Choose the parameters $c > 0$, $M > 2c$, $\bar{\beta} > 0$, and $\varepsilon, \theta \in (0, 1)$. Choose the function $\rho : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ such that $\rho(t)$ is separated from zero when t is separated from zero and $\rho(t) \longrightarrow 0$ as $t \longrightarrow 0+$, and the residual $\sigma_c : \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s) \longrightarrow \mathbb{R}$ for the KKT system (2.16) of problem (2.9). Choose the initial point $(x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s)$ and set $k = 0$.

SQP step. If $\sigma_c(x^k, y^k, \lambda^k) = 0$, stop the process. Otherwise, choose a symmetric positive definite $n \times n$ matrix H_k and define the matrix $\mathcal{H}_k$ in accordance with (3.3), (3.4), (3.11), and (3.12). Compute $(\tilde{x}^{k+1}, \tilde{y}^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^s$ as a stationary point of problem (3.1) and $\lambda^{k+1} \in \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s$ as the Lagrange multiplier corresponding to $(\tilde{x}^{k+1}, \tilde{y}^{k+1})$. Set $\xi^k = \tilde{x}^{k+1} - x^k$, $\eta^k = \tilde{y}^{k+1} - y^k$, and $p^k = (\xi^k, \eta^k)$.

Linesearch step. Choose

$$\beta_k \geq \|\lambda^{k+1}\|_\infty + \bar{\beta}$$

and calculate

$$\Delta_k = \langle f'(x^k), \xi^k \rangle + \langle 4(\max\{0, y^k\})^3 - 2c\max\{0, y^k\}, \eta^k \rangle - \beta_k\psi(x^k, y^k).$$

Set $\alpha = 1$. If the inequality

$$\varphi_{c,\beta_k}((x^k, y^k) + \alpha p^k) \leq \varphi_{c,\beta_k}(x^k, y^k) + \varepsilon\alpha\Delta_k \tag{4.1}$$

is fulfilled, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$ and again check (4.1), etc. until (4.1) is fulfilled.

Set

$$(x^{k+1}, y^{k+1}) = (x^k, y^k) + \alpha_k p^k, \tag{4.2}$$

increase k by one, and go to the SQP step.

In order to justify the global convergence of Algorithm 1, we need the following auxiliary facts.

Lemma 1. *Assume that a bounded sequence* $\{a^k\} \subset \mathbb{R}^q_+$ *and a sequence* $\{b^k\} \subset \mathbb{R}^q$ *are such that*

$$\{\min\{0, b^k\}\} \longrightarrow 0, \quad \langle a^k, b^k \rangle \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty. \tag{4.3}$$

Then,

$$a_i^k b_i^k \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty \quad \forall i = 1, 2, \ldots, q.$$

Proof. Since the members of the sequence $\{a^k\}$ are nonnegative and the sequence is bounded, we conclude from the first relation in (4.3) that

$$a_i^k b_i^k \ge a_i^k \min\{0, b_i^k\} \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty \tag{4.4}$$

for each $i = 1, 2, \ldots, q$. It follows that

$$\liminf_{k \to \infty} a_i^k b_i^k \ge 0. \tag{4.5}$$

On the other hand, for each $i = 1, 2, \ldots, q$, the second condition in (4.3) and the limit relation (4.4) imply that

$$0 = \lim_{k \to \infty} \sum_{i=1}^{q} a_i^k b_i^k = \lim_{k \to \infty} \sum_{i=1}^{q} a_i^k (\min\{0, b_i^k\} + \max\{0, b_i^k\}) = \lim_{k \to \infty} \sum_{i=1}^{q} a_i^k \max\{0, b_i^k\}.$$

All the terms in the sum on the right-hand side are nonnegative because the members of $\{a^k\}$ are nonnegative. Consequently, this limit relation holds if and only if

$$a_i^k \max\{0, b_i^k\} \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty.$$

Using again the nonnegativity of the sequence $\{a^k\}$, we have

$$a_i^k b_i^k \le a_i^k \max\{0, b_i^k\} \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty.$$

It follows that

$$\limsup_{k \to \infty} a_i^k b_i^k \le 0. \tag{4.6}$$

Combining (4.5) and (4.6), we obtain the desired result. The lemma is proved.

Lemma 2. *Let $\{a_k\}$ and $\{b_k\}$ be given scalar sequences, and let $a_k \ge 0$ for all k.*

If

$$\min\{0, b_k\} \longrightarrow 0, \quad a_k b_k \longrightarrow 0 \quad as \quad k \longrightarrow \infty, \tag{4.7}$$

then each of the equivalent conditions

$$\min\{a_k, b_k\} \longrightarrow 0 \quad as \quad k \longrightarrow \infty \tag{4.8}$$

and

$$a_k + b_k - \sqrt{a_k^2 + b_k^2} \longrightarrow 0 \quad as \quad k \longrightarrow \infty. \tag{4.9}$$

is fulfilled.

Proof. It is easily verified that the inequality

$$|\min\{a, b\}| \le |\min\{0, b\}| + |ab|,$$

holds for all $a \ge 0$ and $b \le 0$, while the inequality

$$|\min\{a, b\}| \le \sqrt{|\min\{0, b\}| + |ab|}.$$

holds for all $a \ge 0$ and $b > 0$. Consequently, (4.7) implies (4.8). The equivalence of conditions (4.8) and (4.9) follows from [23, Lemma 3.1]. The lemma is proved.

Note that, in general, (4.8) (and, hence, the equivalent condition (4.9)) does not imply (4.7). To see this, it suffices to put $a_k = 1/k$ and $b_k = k$ ($k = 1, 2, \ldots$).

From Lemmas 1 and 2, we derive

Proposition 6. *Let the sequence $\{(x^k, y^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s)$ be such that $\{\lambda^k\}$ is a bounded sequence and, for all k, $(\lambda^g)^k \ge 0$ and $(\lambda^G)^k \ge 0$.*

Then, for every scalar c, any of the equivalent conditions

$$\bar{\sigma}_c(x^k, y^k, \lambda^k) \longrightarrow 0 \quad as \quad k \longrightarrow \infty$$

and

$$\tilde{\sigma}_c(x^k, y^k, \lambda^k) \longrightarrow 0 \quad as \quad k \longrightarrow \infty$$

implies each of the equivalent conditions

$$\sigma_c^{\mathrm{NR}}(x^k, y^k, \lambda^k) \longrightarrow 0 \quad as \quad k \longrightarrow \infty$$

and

$$\sigma_c^{\mathrm{FB}}(x^k, y^k, \lambda^k) \longrightarrow 0 \quad as \quad k \longrightarrow \infty.$$

In the following theorem, we establish the global convergence properties of Algorithm 1. Although the structure of this algorithm is fairly traditional, it does not suffice to refer to the well-known results concerning the global convergence of the SQP methods (stated, for instance, in [16, Theorem 5.4.1; 24, Theorem 17.2]). The reason is that, in general, the matrices $\mathcal{H}_k$ in Algorithm 1 are not uniformly positive definite, which is one of the basic assumptions in the traditional analysis of the global convergence of the SQP methods.

Theorem 3. *Let the function f and the mappings h, g, H, and G be differentiable on $\mathbb{R}^n$, and let their derivatives satisfy the Lipschitz condition on $\mathbb{R}^n$. Assume that, in Algorithm* 1, *the residual σ_c is chosen as $\bar{\sigma}_c$, $\tilde{\sigma}_c$, σ_c^{NR}, or σ_c^{FB}, while the matrices H_k are chosen so that the sequence $\{H_k\}$ is bounded and there exists $\gamma > 0$ such that*

$$\langle H_k\xi, \xi\rangle \geq \gamma\|\xi\|^2 \quad \forall \xi \in \mathbb{R}^n \tag{4.10}$$

for all k.

Then, for every initial approximation $(x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^s \times (\mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^s \times \mathbb{R}^s)$, either Algorithm 1 *terminates after a finite number of steps at a solution to the KKT system* (2.16) *for problem* (2.9) *or, at some iteration step, the constraints of subproblem* (3.1) *become inconsistent or the algorithm generates an infinite sequence $\{(x^k, y^k, \lambda^k)\}$. In the last case, if*

$$\beta_k = \beta > 0 \tag{4.11}$$

for all sufficiently large k, then at least one of the following assertions is true:

(1) *It holds that*

$$\varphi_{c,\beta}(x^k, y^k) \longrightarrow -\infty \quad as \quad k \longrightarrow \infty; \tag{4.12}$$

(2) *There exists a subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ such that*

$$\sigma_c(x^{k_j}, y^{k_j}, \lambda^{k_j}) \longrightarrow 0 \quad as \quad j \longrightarrow \infty, \tag{4.13}$$

In particular, every limit point of the subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ satisfies the KKT system (2.16) *for problem* (2.9). *Furthermore, every subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ such that*

$$\liminf_{j \to \infty} \sigma_c(x^{k_j}, y^{k_j}, \lambda^{k_j}) > 0, \tag{4.14}$$

satisfies the limit relations

$$\{p^{k_j}\} \longrightarrow 0, \quad \sigma_c(x^{k_j}, y^{k_j}, \lambda^{k_j+1}) \longrightarrow 0 \quad as \quad j \longrightarrow \infty.$$

In particular, let $(\bar{x}, \bar{y}, \bar{\lambda})$ be a limit point of the subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$. Then, the subsequence (x^{k_j+1}, y^{k_j+1}) converges to $(\bar{x}, \bar{y})$, and every limit point of the subsequence $\{(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1})\}$ satisfies the KKT system (2.16) *for problem* (2.9).

Proof. Suppose that $\sigma_c(x^k, y^k, \lambda^k) \neq 0$ for all k. Then, combining (3.11), (3.12), (4.10), and the properties of the function ρ used in Algorithm 1, we conclude that the matrix $\mathcal{H}_k$ is positive definite. Thus, subproblem (3.1) is a quadratic program with a strongly convex objective function. If the constraints of this problem are consistent, then the next SQP step of Algorithm 1 determines a point $(\tilde{x}^{k+1}, \tilde{y}^{k+1}, \lambda^{k+1})$. (Note that $\lambda^{k+1} = ((\lambda^h)^{k+1}, (\lambda^g)^{k+1}, (\lambda^H)^{k+1}, (\lambda^G)^{k+1})$ may be nonuniquely determined.) Moreover, the differences $\xi^k = \tilde{x}^{k+1} - x^k$ and $\eta^k = \tilde{y}^{k+1} - y^k$ satisfy the relations of the KKT system for problem (3.1):

$$\begin{gathered}
f'(x^k) + H_k\xi^k + (h'(x^k))^{\mathrm{T}}(\lambda^h)^{k+1} + (g'(x^k))^{\mathrm{T}}(\lambda^g)^{k+1} - (H'(x^k))^{\mathrm{T}}(\lambda^H)^{k+1} + (G'(x^k))^{\mathrm{T}}(\lambda^G)^{k+1} = 0,\\
2(\max\{0, y^k\})^3 - c\max\{0, y^k\} + \operatorname{diag}(a_c(y^k, \lambda^k))\eta^k + B_{\min}(y^k)(\lambda^H)^{k+1} - B_{\max}(y^k)(\lambda^G)^{k+1} = 0,\\
h(x^k) + h'(x^k)\xi^k = 0,\\
(\lambda^g)^{k+1} \ge 0, \quad g(x^k) + g'(x^k)\xi^k \le 0, \quad \langle(\lambda^g)^{k+1}, g(x^k) + g'(x^k)\xi^k\rangle = 0, \qquad (4.15)\\
(\min\{0, y^k\})^2 - H(x^k) - H'(x^k)\xi^k + 2B_{\min}(y^k)\eta^k = 0,\\
(\lambda^G)^{k+1} \ge 0, \quad G(x^k) - (\max\{0, y^k\})^2 + G'(x^k)\xi^k - 2B_{\max}(y^k)\eta^k \le 0,\\
\langle(\lambda^G)^{k+1}, G(x^k) - (\max\{0, y^k\})^2 + G'(x^k)\xi^k - 2B_{\max}(y^k)\eta^k\rangle = 0.
\end{gathered}$$

If $\xi^k = 0$ and $\eta^k = 0$, then (4.15) converts into the KKT system (2.16) for problem (2.9), where $x = x^k$, $y = y^k$, and $\lambda = \lambda^{k+1}$; therefore, $\sigma_c(x^k, y^k, \lambda^{k+1}) = 0$. Furthermore, $p^k = (\xi^k, \eta^k) = 0$, and the result of this step in Algorithm 1 is the point $(x^{k+1}, y^{k+1}, \lambda^{k+1})$, where $x^{k+1} = x^k$ and $y^{k+1} = y^k$. Consequently, $\sigma_c(x^{k+1}, y^{k+1}, \lambda^{k+1}) = 0$, which contradicts the assumption made above.

Thus, $p^k \ne 0$, which implies that

$$\varphi'_{c,\beta_k}((x^k, y^k); p^k) \le \Delta_k \le -\langle \mathcal{H}_k p^k, p^k\rangle - \bar{\beta}\psi(x^k, y^k) < 0$$

(see [16, Lemma 5.4.1]). Then, the linesearch performed in the algorithm terminates after a finite number of reductions of the initial tentative value $\alpha = 1$ by adopting some value $\alpha_k > 0$. Thus, the next approximation $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ is well determined.

If there is no subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ that satisfies (4.13), then there exists $\delta > 0$ such that

$$\sigma_c(x^k, y^k, \lambda^k) \ge \delta \qquad (4.16)$$

for all k. Due to the properties of the function ρ, there exists $\tilde{\delta} > 0$ such that $\rho(\sigma_c(x^k, y^k, \lambda^k)) \ge \tilde{\delta}$ for all k. It follows from (3.11), (3.12), and (4.10) that

$$\langle \mathcal{H}_k(\xi, \eta), (\xi, \eta)\rangle \ge \langle H_k\xi, \xi\rangle + 2\rho(\sigma_c(x^k, y^k, \lambda^k))\|\eta\|^2 \ge \min\{\gamma, 2\tilde{\delta}\}\|(\xi, \eta)\|^2 \quad \forall (\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^s.$$

Thus, the matrices $\mathcal{H}_k$ are uniformly positive definite. Moreover, in view of (3.12) and the fact that $\{H_k\}$ is a bounded sequence, the sequence $\{\mathcal{H}_k\}$ is also bounded. Now, the standard results concerning the global convergence of SQP methods with linesearch (e.g., see [16, Theorem 5.4.1]) imply that either (4.12) is fulfilled or

$$\{p^k\} \longrightarrow 0, \quad \bar{\sigma}_c(x^k, y^k, \lambda^{k+1}) \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty. \qquad (4.17)$$

Recall that $\bar{\sigma}_c(\cdot, \cdot, \lambda^k)$, $\tilde{\sigma}_c(\cdot, \cdot, \lambda^k)$, $\sigma_c^{\mathrm{NR}}(\cdot, \cdot, \lambda^k)$, and $\sigma_c^{\mathrm{FB}}(\cdot, \cdot, \lambda^k)$ are Lipschitzian functions. Using (4.2) and Proposition 6, we derive from (4.17) the limit relations

$$\|x^{k+1} - x^k\| \longrightarrow 0, \quad \|y^{k+1} - y^k\| \longrightarrow 0, \quad \sigma_c(x^{k+1}, y^{k+1}, \lambda^{k+1}) \longrightarrow 0 \quad \text{as} \quad k \longrightarrow \infty,$$

The last of these relations contradicts (4.16).

Thus, we have shown that, if (4.12) is not fulfilled, there exists a subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ that satisfies (4.13).

Finally, let $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}$ be a subsequence that satisfies (4.14). Then, the desired assertion is obtained by repeating the above argument for this subsequence. The theorem is proved.

Theorem 3 gives a good reason to expect that, if $(\bar{x}, \bar{y}, \bar{\lambda})$ is a limit point of the sequence $\{(x^k, y^k, \lambda^k)\}$ generated by Algorithm 1, then $(\bar{x}, \bar{y})$ is a stationary point of problem (3.1) and $\bar{\lambda}$ is the corresponding Lagrange multiplier. Note, however, that, in general, the latter does not even guarantee that $\bar{x}$ is a weakly stationary point of problem (1.1), but then the only possibly lacking ingredient of weak stationarity is the condition $\bar{\lambda}^H_{I_{0-}} \ge 0$ (see assertion (3) of Proposition 3). Moreover, as noted above, the weak stationarity of a MPVC is actually a fairly strong concept of stationarity.

5. NUMERICAL RESULTS

In this section, we present the results obtained by a numerical comparison of Algorithm 1 with certain alternative methods. To this end, 23 MPVCs were taken from all publications concerning this problem class and available to the authors.

For brevity, the number 1 is attributed to Algorithm 1 in which the matrices H_k are calculated as explained above, that is, by using the BFGS formula with the Powell modification. Number 2 is attributed to an analogue of Algorithm 1 in which the matrices $\mathcal{H}_k$ are solely calculated on the basis of the BFGS formula with the Powell modification (rather than using the special method adopted in Algorithm 1). Number 3 is used for the conventional quasi-Newton SQP method (which means the BFGS formula with the Powell modification) combined with the linesearch performed for an l_1-exact penalty function; this method is directly applied to problem (1.1) (with no "lifting" or any other modifications). All the three methods were implemented without invoking any ways of overcoming the possible inconsistency of constraints in the subproblems, as well as ways of overcoming the Maratos effect (see [16, p. 230]).

For the parameters in Algorithm 1 and the corresponding parameters in the other algorithms, we chose the values $c = 200$, $\bar{\beta} = 1$, $\varepsilon = 10^{-4}$, and $\theta = 0.5$. The parameter M, which was introduced for theoretical justification of the global convergence, should in practice be chosen so that, under normal conditions, it does not affect the computational process. Since there is no obvious reasonable rule for the choice of M, it was set equal to $+\infty$ in our calculations.

The function ρ was defined by the formula

$$\rho(t) = \begin{cases} t, & \text{if} \quad t < 0.1, \\ 0.1, & \text{if} \quad t \ge 0.1. \end{cases}$$

At each iteration step, the penalty parameter β_k was calculated as

$$\beta_k = \|\lambda^{k+1}\|_\infty + \bar{\beta}.$$

Note that this choice may not satisfy condition (4.11) in Theorem 3 on the global convergence; however, it has an appealing simplicity and works well in practice because it allows one to reduce large values of the penalty parameter possible at the early steps of the process. In more complicated rules for the choice of β_k (e.g., see [21, Section 18.3; 24, Section 17.1]), the above property and the validity of (4.11) can certainly be combined.

Our calculations were performed in the Matlab environment. Quadratic subproblems were solved using quadprog, the Matlab built-in solver. The stopping criterion used for Algorithms 1 and 2 was

$$\sigma_c^{\mathrm{FB}}(x^k, y^k, \lambda^k) < 10^{-6}.$$

Algorithm 3 was terminated when the analogous residual of the KKT system for the original problem (1.1) was less than 10^{-6}.

If the required accuracy was not attained during 500 iteration steps or, for some reason, the method was not able to execute the current step, the corresponding run was regarded as unsuccessful.

For each test example, we performed 100 runs of each algorithm from randomly generated initial points (which were the same for all the algorithms). The primal initial points x^0 were chosen from a cube

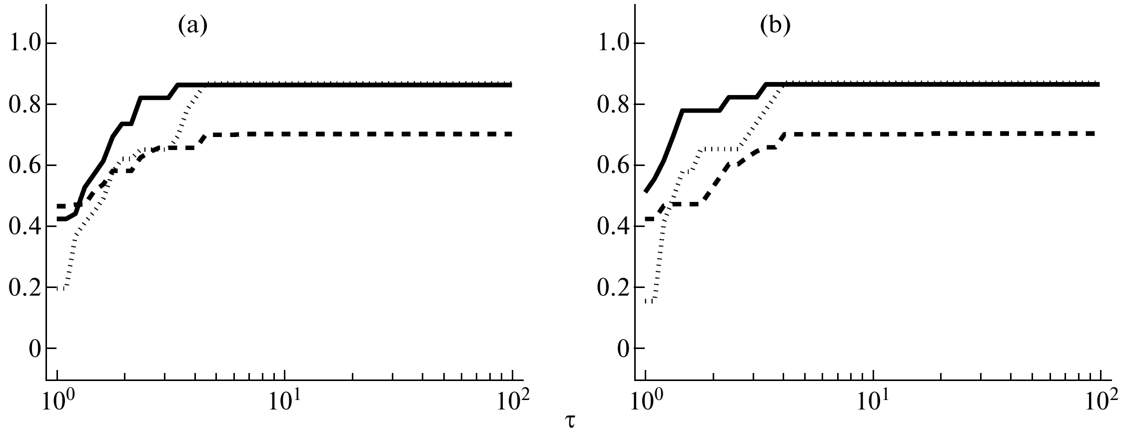

Fig. 2.

centered at the solution $\bar{x}$, which was known for each test, the edges of the cube being equal to 20. For Algorithms 1 and 2, the initial value of the auxiliary variable was determined by the rule

$$y_i^0 = \begin{cases} -\sqrt{H_i(x^0)}, & \text{if} \quad H_i(x^0) > 0, \\ \sqrt{c/2}, & \text{if} \quad H_i(x^0) \le 0, \end{cases}$$
$$i = 1, 2, \ldots, s.$$

For all the algorithms, the initial values of the dual variables were generated in a similar way, but the cube was centered at the zero and additional inequality constraints were used for the components corresponding to the nonnegativity requirements.

In the case of a successful run, convergence to the solution was declared when the distance from the final primal approximation x^k to $\bar{x}$ was less than 10^{-3}.

For all of the three methods, an idea of the relative average number of the outer and inner iterations can be gained from Fig. 2. Inner iteration steps are understood as the steps of quadprog. The results are shown in the form of the so-called "performance profiles". This aggregated method for representing the results of numerical experiments was first proposed in [25]. For each algorithm, we present the graph of the function defined as follows. (Hereinafter, the graphs corresponding to Algorithms 1–3 are depicted in solid line, dotted line, or dashed line, respectively.) Its value at $\tau \in [1, +\infty]$ is the relative sum (with respect to the total number of problems) of the fractions of successful runs taken over the problems for which the result (in this case, the average number of iteration steps) of the corresponding algorithm was at most τ times worse (in this case, greater) than the best result (over the three algorithms). In a sense, this value can be interpreted as the probability of the event that, for a problem in the given set, the result of a run of the corresponding algorithm is at most τ times worse than the best result. It is assumed that the result of an unsuccessful run is indefinitely worse than that of any successful run. The exact formulas for the functions whose graphs are represented in the performance profiles of the type we use here can be found in [26]. The value of such a function at $\tau = 1$ can be interpreted as a measure of the "pure" efficiency of the corresponding algorithm, that is, as the probability that this algorithm will be the best, while its value at $\tau = +\infty$ can be interpreted as a measure of the "pure" robustness, that is, as the probability that a run is successful.

It can be seen that Algorithm 3 is slightly more efficient than Algorithm 1 in terms of the number of outer iteration steps (see part (a) of Fig. 2), whereas the situation is opposite in terms of the number of inner iteration steps (see part (b)). In either term, Algorithm 2 is less efficient than the two other algorithms. On the other hand, Algorithm 3 clearly yields to the two other algorithms in terms of robustness.

A similar form of representation is used in Fig. 3 for the data concerning the relative average number of evaluations of the functions specifying the constraints (see part (a)) and the analogous number for their derivatives (see part (b)) per successful run. Figure 4a shows the relative average number of evaluations of the objective function. In all the cases, the results are similar to those in Fig. 2a.

In addition to efficiency and robustness, an important feature of an algorithm is the quality of its final approximations, that is, the portion of the cases when the algorithm converges to the genuine solution

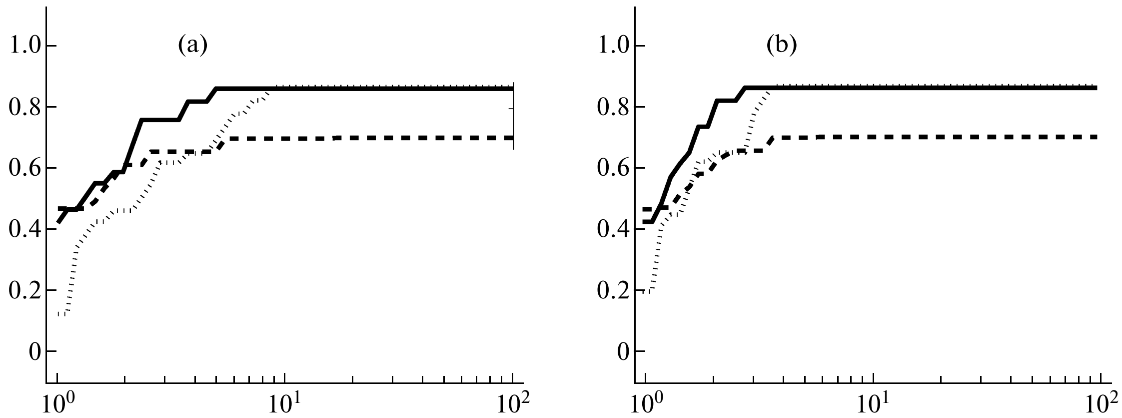

Fig. 3.

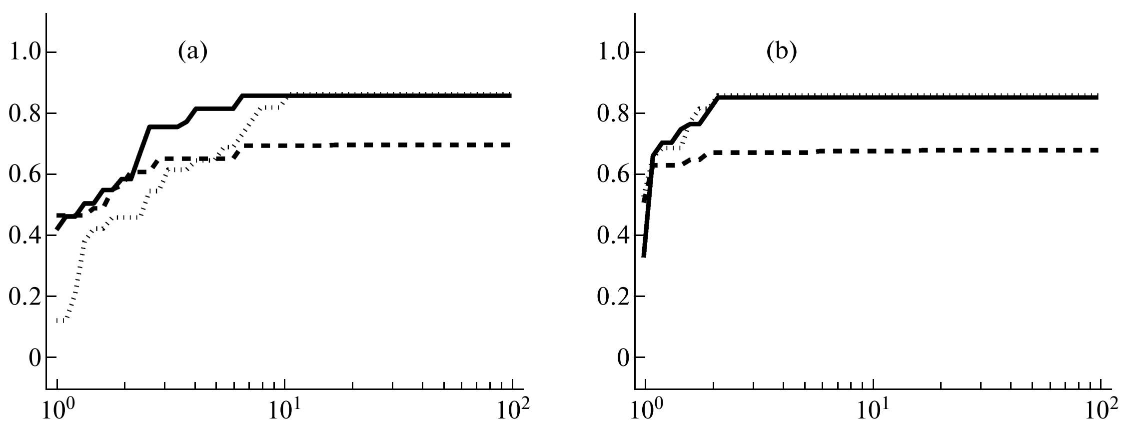

Fig. 4.

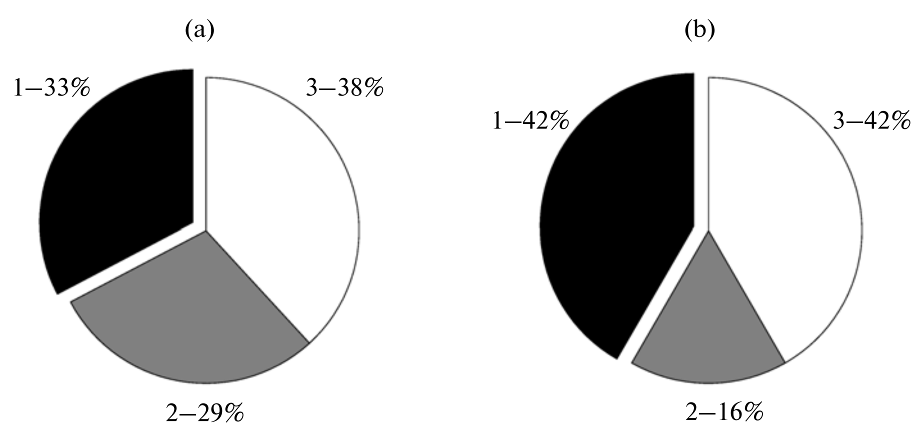

Fig. 5.

rather than to a nonoptimal stationary point. For the algorithms under discussion, this characteristic is illustrated by Fig. 4b. The result of an algorithm is understood here as the inverse of the number of convergencies to the solution. Note that the result was set equal to $+\infty$ if the algorithm never converged to the solution. Such occasions make an additional contribution to the total number of unsuccessful runs, which somewhat reduces the robustness performance of the algorithms in Fig. 4b compared to the previous performance profiles.

Figure 4b shows that, in terms of the convergence to the solution, Algorithms 1 and 2 demonstrate better performance than Algorithm 3. This somewhat surprising result is a consequence of the fact that Algorithm 3 has a lower general robustness.

The pie diagrams in Fig. 5 are aimed at giving an indication of the ability of an algorithm to achieve smaller values of the objective function (compared to the other algorithms) in the case of successful runs. In these cases, the resulting primal approximation is feasible or almost feasible; consequently, the attained value of the objective function can be considered as another reasonable measure of the behavior of the corresponding algorithm. The diagrams in Fig. 5 were obtained as follows. For each algorithm and for each problem, the average attained value of the objective function per successful run was calculated. If this value is minimal (over the three algorithms), then the corresponding algorithm falls into the category "best" (see Fig. 5a); if this value is maximal, then it falls into the category "worst" (see Fig. 5b). The average attained value of the objective function was regarded as equal to the minimal (maximal) value if it differed from the latter by less than 10^{-3}. Note that, for some problems, an algorithm could fall into both above categories if the average attained values of the objective function were identical for all the three algorithms. After the number of occurrencies of each algorithm in each category was determined, these numbers were summed inside each category. The fraction of each algorithm is shown in Fig. 5. In terms of the above characteristics, all the three algorithms are comparable, although Algorithm 2 somewhat yields to the two other algorithms.

We sum up as follows. Unlike the conventional methods as applied directly to the original problem (1.1), the proposed approach has justified global and local superlinear convergence. Moreover, the numerical results indicate that this approach is also quite competitive from a practical viewpoint.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 10-01-00251.

REFERENCES

1. W. Achtziger and C. Kanzow, "Mathematical Programs with Vanishing Constraints: Optimality Conditions and Constraint Qualifications," Math. Program. **114** (1), 69–99 (2007).
2. W. Achtziger, T. Hoheisel, and C. Kanzow, "A Smoothing-Regularization Approach to Mathematical Programs with Vanishing Constraints," 10.1080/10556788.2010.535170. Preprint No. 284 (Inst. Math. Univ. Würzburg, Würzburg, 2008).
3. T. Hoheisel and C. Kanzow, "First- and Second-Order Optimality Conditions for Mathematical Programs with Vanishing Constraints," Appl. Math. **52**, 495–514 (2007).
4. T. Hoheisel and C. Kanzow, "Stationarity Conditions for Mathematical Programs with Vanishing Constraints Using Weak Constraint Qualifications," J. Math. Anal. Appl. **337**, 292–310 (2008).
5. T. Hoheisel and C. Kanzow, "On the Abadie and Guignard Constraint Qualifications for Mathematical Programs with Vanishing Constraints," Optimization **58**, 431–448 (2009).
6. A. F. Izmailov and M. V. Solodov, "Mathematical Programs with Vanishing Constraints: Optimality Conditions, Sensitivity, and a Relaxation Method," J. Optim. Theory Appl. **142**, 501–532 (2009).
7. A. F. Izmailov and A. L. Pogosyan, "Optimality Conditions and Newton-Type Methods for Mathematical Programs with Vanishing Constraints," Zh. Vychisl. Mat. Mat. Fiz. **49**, 1184–1196 (2009) [Comput. Math. Math. Phys. **49**, 1128–1140 (2009)].
8. A. F. Izmailov and A. L. Pogosyan, "On Active-Set Methods for Mathematical Programs with Vanishing Constraints," in *Theoretical and Applied Problems in Nonlinear Analysis* (Vychisl. Tsentr Ross. Akad. Nauk, Moscow, 2009), pp. 18–49 [in Russian].
9. T. Hoheisel, C. Kanzow, and A. Schwartz, "Convergence of a Local Regularization Approach for Mathematical Programs with Complementarity or Vanishing Constraints," Optim. Methods Software. DOI: 10.1080/10556788.2010.535170.
10. Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints* (Cambridge Univ. Press, Cambridge, 1996).
11. J. V. Outrata, M. Kocvara, and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications, and Numerical Results* (Kluwer, Boston, 1998).
12. A. F. Izmailov, "Mathematical Programs with Complementarity Constraints: Regularity, Optimality Conditions, and Sensitivity," Zh. Vychisl. Mat. Mat. Fiz. **44**, 1209–1228 (2004) [Comput. Math. Math. Phys. **44**, 1145–1164 (2004)].
13. A. F. Izmailov, *Sensitivity in Optimization* (Fizmatlit, Moscow, 2006) [in Russian].

14. O. Stein, "Lifting Mathematical Programs with Complementarity Constraints," Math. Program. (2010). DOI 10/1007/s10107-010-0345-y..
15. A. F. Izmailov, A. L. Pogosyan, and M. V. Solodov, "Semismooth Newton Method for the Lifted Reformulation of Mathematical Programs with Complementarity Constraints," Comput. Optim. Appl. DOI 10.1007/s10589-010-9341-7.
16. A. F. Izmailov and M. V. Solodov, *Numerical Optimization Methods,* 2nd. ed. (Fizmatlit, Moscow, 2008) [in Russian].
17. L. Qi, "Superlinearly Convergent Approximate Newton Methods for LC^1 Optimization Problems," Math. Program. **64**, 277–294 (1994).
18. J. Han and D. Sun, "Superlinear Convergence of Approximate Newton Methods for LC^1 Optimization Problems without Strict Complementarity," *Recent Advances in Nonsmooth Optimization* (World Scientific, Singapore, 1993), pp. 353–367.
19. A. F. Izmailov, A. L. Pogosyan, and M. V. Solodov, Preprint A 675/2010/IMPA (Rio de Janeiro, 2010) (available at http://www.preprint.impa.br:80/Shadows/SERIE A/2010/675.html).
20. J. F. Bonnans, "Local Analysis of Newton-Type Methods for Variational Inequalities and Nonlinear Programming," Appl. Math. Optim. **29**, 161–186 (1994).
21. J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer-Verlag, New York, 2006).
22. D. Fernández, A. F. Izmailov, and M. V. Solodov, "Sharp Primal Superlinear Convergence Results for Some Newtonian Methods for Constrained Optimization," SIAM J. Optim. **20**, 3312–3334 (2010).
23. P. Tseng, "Growth Behavior of a Class of Merit Functions for the Nonlinear Complementarity Problem," J. Optim. Theory Appl. **89**, 17–37 (1996).
24. J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects* (Springer-Verlag, Berlin, 2006).
25. E. Dolan and J. Moré, "Benchmarking Optimization Software with Performance Profiles," Math. Program. **91**, 201–213 (2002).
26. A. N. Dar'ina and A. F. Izmailov, "Semismooth Newton Method for Quadratic Programs with Bound Constraints," Zh. Vychisl. Mat. Mat. Fiz. **49**, 1785–1795 (2009) [Comput. Math. Math. Phys. **49**, 1706–1716 (2009)].