A SIMPLY CONSTRAINED OPTIMIZATION REFORMULATION OF KKT SYSTEMS ARISING FROM VARIATIONAL INEQUALITIES*

Francisco Facchinei[1], Andreas Fischer[2], Christian Kanzow[3], and Ji-Ming Peng[4]

[1] Università di Roma "La Sapienza"
Dipartimento di Informatica e Sistemistica
Via Buonarroti 12, I-00185 Roma, Italy
e-mail: soler@dis.uniroma1.it

[2] Technical University of Dresden
Institute of Numerical Mathematics
D-01062 Dresden, Germany
e-mail: fischer@math.tu-dresden.de

[3] University of Hamburg
Institute of Applied Mathematics
Bundesstrasse 55, D-20146 Hamburg, Germany
e-mail: kanzow@math.uni-hamburg.de

[4] Academia Sinica
State Key Laboratory of Scientific and Engineering Computing
P.O. Box 2719, Beijing, 100080, P.R. China
e-mail: pjm@lsec.cc.ac.cn

October 14, 1996 (revised October 22, 1997)

*The work of the first three authors was partially supported by NATO under grant CRG 960137. The work of the last author was supported by the Natural Science Foundation for Youth of China.

Abstract. The Karush-Kuhn-Tucker (KKT) conditions can be regarded as optimality conditions for both variational inequalities and constrained optimization problems. In order to overcome some drawbacks of recently proposed reformulations of KKT systems, we propose to cast KKT systems as a minimization problem with nonnegativity constraints on some of the variables. We prove that, under fairly mild assumptions, every stationary point of this constrained minimization problem is a solution of the KKT conditions. Based on this reformulation, a new algorithm for the solution of the KKT conditions is suggested and shown to have some strong global and local convergence properties.

Key words. KKT conditions, variational inequalities, constrained optimization problems, global convergence, quadratic convergence, semismoothness, strong regularity.

Running title. Optimization Reformulation of KKT Systems.

AMS (MOS) subject classification. 90C33, 90C30.

1 Introduction

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be once and $h : \mathbb{R}^n \to \mathbb{R}^p, g : \mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable. Define the Lagrangian $L : \mathbb{R}^{n+p+m} \to \mathbb{R}^n$ by

$$L(x, y, z) := F(x) + \nabla h(x)y - \nabla g(x)z,$$

and consider the following Karush-Kuhn-Tucker (KKT) system:

$$\begin{array}{rcl} L(x, y, z) & = & 0, \\ h(x) & = & 0, \\ g(x) \geq 0, z \geq 0, z^T g(x) & = & 0. \end{array} \quad (1)$$

Systems of this type arise in several situations. For example, under any standard constraint qualification, system (1) represents the KKT necessary conditions for a vector $x^* \in X$ to be a solution of the variational inequality problem VIP(X, F)

$$F(x^*)^T (x - x^*) \geq 0 \quad \forall x \in X,$$

where

$$X := \{x \in \mathbb{R}^n | \, h(x) = 0, g(x) \geq 0\},$$

see [18]. In particular, if $F = \nabla f$ for a function $f : \mathbb{R}^n \to \mathbb{R}$, then the KKT conditions represent, again under a constraint qualification, the first order necessary optimality conditions for the minimization problem

$$\min f(x) \quad \text{subject to} \quad h(x) = 0, g(x) \geq 0,$$

see, e.g., [17].

In this paper we focus on the problem of finding a KKT point $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$, i.e., a triple satisfying the KKT system (1). This is actually the aim of most algorithms for the solution of variational inequality and nonlinear programming problems.

The method we will describe in this paper is related to the recent proposal [9], where system (1) is transformed into a differentiable unconstrained minimization problem. The reformulation considered in [9] is based on the simple convex function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\varphi(a, b) := \sqrt{a^2 + b^2} - a - b \quad (2)$$

and introduced in [14]. It is easy to check that

$$\varphi(a, b) = 0 \quad \Longleftrightarrow \quad a \geq 0, \; b \geq 0, \; ab = 0. \quad (3)$$

Hence we can reformulate system (1) as a nonlinear system of equations $\Phi(w) = 0$, where the nonsmooth mapping $\Phi : \mathbb{R}^{n+p+m} \to \mathbb{R}^{n+p+m}$ is defined by

$$\Phi(w) := \Phi(x, y, z) := \begin{pmatrix} L(x, y, z) \\ h(x) \\ \phi(g(x), z) \end{pmatrix}$$

and

$$\phi(g(x), z) := (\varphi(g_1(x), z_1), \ldots, \varphi(g_m(x), z_m))^T \in \mathbb{R}^m.$$

We can now associate to this system its natural merit function, i.e.

$$\Psi(w) := \frac{1}{2}\Phi(w)^T\Phi(w) = \frac{1}{2}\|\Phi(w)\|^2,$$

so that solving system (1) is equivalent to finding a global solution of the problem

$$\min \quad \Psi(w). \tag{4}$$

This approach was studied both from the theoretical and algorithmic point of view in [9, 10] to which we refer the interested reader for a detailed motivation and for a comparison to other methods. We remark that, in order to find a solution of system (1), one has to seek global solutions of the minimization problem (4), while usual unconstrained minimization algorithms can only provide stationary points of (4). One of the central questions dealt with in [9] is therefore the study of conditions implying that a stationary point of Ψ is a global minimum of Ψ. The conditions given in [9], although relatively weak, all include the assumption that the Jacobian of the Lagrangian with respect to the x-variables,

$$\nabla_x L(x, y, z) = \nabla F(x)^T + \sum_{j=1}^{p} y_j \nabla^2 h_j(x) - \sum_{i=1}^{m} z_i \nabla^2 g_i(x),$$

is positive semidefinite. This condition is satisfied, e.g., if F is monotone and the constraints are affine so that $\nabla F(x)$ is positive semidefinite and the Hessians $\nabla^2 h_j(x)$ and $\nabla^2 g_i(x)$ all vanish. However, if one considers the most natural extension of this case, i.e., F monotone, h linear, and g nonlinear and concave, it is easy to see that, since the matrices $\nabla^2 g_i(x)$ are negative semidefinite, if z_i is negative and large enough, $\nabla_x L(x, y, z)$ cannot be positive semidefinite. Note also that this conclusion is independent of the structure of F or h. We illustrate this point by the following example, taken from [26]. Let $n = m = 1$, and $p = 0$ and set

$$F(x) := \frac{1}{2}x - 5, \quad g(x) := -\frac{1}{2}x^2 + x,$$

so that F is strongly monotone and g is (strongly) concave. Then $\nabla\Psi(w) = \nabla\Psi(x, z) = 0$ both for $(x, z) = (0, -1)$ and for $(x, z) = (2, 4)$. But while the latter stationary point satisfies the KKT conditions (1), the first one does not. In fact, it is easy to check that $\Psi(0, -1) > \Psi(2, 4) = 0$, so that $(2, 4)$ is a global solution of (4) but $(0, -1)$ is not.

This feature is somewhat disturbing, since it implies that, even if we solve a strongly monotone variational inequality over a convex set defined by nonlinear inequalities, we cannot ensure convergence to the unique solution of the variational inequality. Since the problem clearly arises because of negative values of the variables z_i that we know *a priori* have to be nonnegative at a solution of system (1), we are naturally led to consider the following variant of problem (4):

$$\min \quad \Psi(w) \quad \text{s.t.} \quad z \geq 0. \tag{5}$$

Therefore, this paper is devoted to the study of this reformulation of the KKT system (1). In particular we shall give conditions which ensure that every stationary point of problem (5) is a solution of system (1). We shall also propose a specific algorithm for the solution of problem (5) which fully exploits its characteristics (note that, as we shall discuss more in detail in the next section, the operator Φ is not differentiable, while Ψ is not twice differentiable, so that standard methods are not appropriate). Some preliminary results on the issues dealt with in this paper can be found in [26]. A related approach was proposed in the context of nonlinear complementarity problems in [16].

This paper is structured in the following way. In the next section we recall some known facts about the functions Ψ and Φ and about a nonsmooth Newton method. Then, in Section 3, we give conditions ensuring that every stationary point of problem (5) is a solution of the KKT system (1). In Section 4, we introduce an algorithm for the solution of (1). This algorithm reduces the merit function Ψ in each step while maintaining the variable z nonnegative. We prove global and local convergence results for this algorithm in Section 5.

Notation. The index sets $\{1, \ldots, p\}$ and $\{1, \ldots, m\}$ will be abbreviated by the capital letters J and I, respectively. If $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ is a KKT point, we will denote by I_0 the set of active inequality constraints and by I_+ the set of strongly active inequality constraints, i.e.,

$$I_0 = \{i \in I |\, g_i(x^*) = 0\}, \quad I_+ = \{i \in I_0 |\, z_i^* > 0\}.$$

To denote the transposed Jacobian of a function we will use the symbol ∇. If the function is real-valued the transposed Jacobian coincides with the gradient, i.e., we view the gradient as a column vector. Moreover, ∇^2 is used as a symbol for the Hessian. All vector norms are Euclidean norms, whereas matrix norms are assumed to be consistent with the Euclidean norm.

2 Preliminaries

In this section we recall results on the differentiability of the functions Φ and Ψ which are at the heart of this paper.

By the differentiability assumptions we made on the functions F, h, and g, and by the convexity of φ, it is obvious that the mapping Φ is locally Lipschitzian and thus almost everywhere differentiable by Rademacher's theorem. Let us denote by D_Φ the set of points $w \in \mathbb{R}^{n+p+m}$ at which Φ is differentiable. Then, we can consider the *B-subdifferential* (or pre-generalized Jacobian matrix in Clarke's sense) of Φ at w,

$$\partial_B \Phi(w) := \{H |\, H = \lim_{w^k \to w, w^k \in D_\Phi} \nabla \Phi(w^k)^T\}$$

which is a nonempty and compact set whose convex hull

$$\partial \Phi(w) := \operatorname{conv} \partial_B \Phi(w)$$

is Clarke's [3] *generalized Jacobian* of Φ at w. Related to the notion of B-subdifferential is the BD-regularity condition, see [27], which will play an important role in the analysis of the convergence rate of our method.

Definition 2.1 *The vector w^* is called* BD-regular *for Φ if all elements $H \in \partial_B \Phi(w^*)$ are nonsingular.*

The following result gives an overestimate of the generalized Jacobian of Φ. It basically follows from known rules on the calculation of the generalized Jacobian [3]. For a precise proof, we refer to [9, Proposition 3.2].

Proposition 2.2 *Let $w = (x, y, z) \in \mathbb{R}^{n+p+m}$. Then*

$$\partial \Phi(w)^T \subseteq \begin{pmatrix} \nabla_x L(w) & \nabla h(x) & \nabla g(x) D_a(w) \\ \nabla h(x)^T & 0 & 0 \\ -\nabla g(x)^T & 0 & D_b(w) \end{pmatrix},$$

where $D_a(w) = diag\,(a_1(w), \dots, a_m(w))\,, D_b(w) = diag\,(b_1(w), \dots, b_m(w)) \in \mathbb{R}^{m \times m}$ are diagonal matrices whose ith diagonal elements are given by

$$a_i(w) = \frac{g_i(x)}{\sqrt{g_i(x)^2 + z_i^2}} - 1, \quad b_i(w) = \frac{z_i}{\sqrt{g_i(x)^2 + z_i^2}} - 1$$

if $(g_i(x), z_i) \neq 0$, and by

$$a_i(w) = \xi_i - 1, \quad b_i(w) = \zeta_i - 1 \quad \textit{for any } (\xi_i, \zeta_i) \textit{ with } \|(\xi_i, \zeta_i)\| \leq 1$$

if $(g_i(x), z_i) = 0$.

In the following result, we make use of Robinson's strong regularity condition without restating it's definition here. We refer the interested reader to Robinson [29] and Liu [21] for several characterizations of a strongly regular KKT point. In the discussion of Subsection 5.2 we will also give some more details about the relationship of Robinson's strong regularity condition to some other well known concepts in the optimization literature.

Proposition 2.3 *A solution $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ of system (1) is strongly regular if and only if all matrices in Clarke's generalized Jacobian $\partial \Phi(w^*)$ are nonsingular. In particular, the strong regularity of w^* is sufficient for w^* to be a BD-regular solution of the system $\Phi(w) = 0$.*

Proof. See [9, Corollary 4.7]. □

Besides the notion of BD-regularity, the concept of (strong) semismoothness [22, 28] will be of importance. (Strong) semismoothness can be used to establish superlinear (quadratic) convergence of a class of nonsmooth Newton-type methods [27, 28]. We shall not need here the exact definition of (strong) semismoothness, for which we refer to [22, 28]. However we note that it can be shown that both differentiable and convex functions are semismooth [22]. Moreover, it is known that the composition of (strongly) semismooth functions is again (strongly) semismooth [15, 22]. With regard to the differentiability assumptions on the functions F, g, and h as stated in Section 1, and to the fact that φ is strongly semismooth [15], we therefore get the following result [9, Proposition 3.1].

Proposition 2.4 *The following statements hold:*

(a) The mapping Φ *is semismooth.*

(b) If ∇F, $\nabla^2 h_j$ $(j \in J)$, *and* $\nabla^2 g_i$ $(i \in I)$ *are locally Lipschitzian, then* Φ *is strongly semismooth.*

As a direct consequence of the (strong) semismoothness of Φ and known results about (strongly) semismooth functions (see, in particular, [24, Proposition 1], [11, Proposition 2] and [15, Lemma 2.8]), we obtain the following proposition.

Proposition 2.5 *The following statements hold:*

(a) It holds that

$$\|\Phi(w+h) - \Phi(w) - Hh\| = o(\|h\|) \quad \text{for } h \to 0 \text{ and } H \in \partial\Phi(w+h).$$

(b) If Φ *is strongly semismooth, then*

$$\|\Phi(w+h) - \Phi(w) - Hh\| = O(\|h\|^2) \quad \text{for } h \to 0 \text{ and } H \in \partial\Phi(w+h).$$

The first part of the next result follows, basically, from the upper semicontinuity of the generalized Jacobian [3, Proposition 2.6.2 (c)] and the assumed BD-regularity; the second part is a standard result which is also a consequence of the BD-regularity assumption. For the precise proofs, we refer the interested reader to [27, Lemma 2.6] and [24, Proposition 3].

Proposition 2.6 *Let* w^* *be a BD-regular solution of* $\Phi(w) = 0$. *Then the following statements hold:*

(a) There exist numbers $c_1 > 0$ *and* $\delta_1 > 0$ *such that the matrices* $H \in \partial_B\Phi(w)$ *are nonsingular and satisfy*

$$\|H^{-1}\| \leq c_1$$

for all w *with* $\|w - w^*\| \leq \delta_1$.

(b) There exist numbers $c_2 > 0$ *and* $\delta_2 > 0$ *such that*

$$\|\Phi(w)\| \geq c_2\|w - w^*\|$$

for all w *with* $\|w - w^*\| \leq \delta_2$.

We conclude this section by recalling a not obvious, but simple result from [9] which will play a crucial role in the design and analysis of our algorithm.

Proposition 2.7 Ψ *is continuously differentiable, and* $\nabla\Psi(w) = H^T\Phi(w)$ *for every* H *in* $\partial\Phi(w)$.

3 A Simply Constrained Reformulation of KKT Systems

In this section we consider stationary points of problem (5) and its relation to the solutions of system (1). For convenience we restate here problem (5):

$$\min \quad \Psi(w) \quad \text{s.t.} \quad z \geq 0.$$

We recall that a point $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ with $z^* \geq 0$ is a stationary point of problem (5) if $\nabla_x \Psi(w^*) = 0$, $\nabla_y \Psi(w^*) = 0$, and

$$\begin{aligned} z_i^* > 0 \quad &\Longrightarrow \tfrac{\partial \Psi(w^*)}{\partial z_i} = 0, \\ z_i^* = 0 \quad &\Longrightarrow \tfrac{\partial \Psi(w^*)}{\partial z_i} \geq 0. \end{aligned}$$

In the sequel we shall indicate by $I_>$ the set of those indices for which $z_i^* > 0$.

In the next theorem we give conditions ensuring that a stationary point of problem (5) is a global solution and, therefore, a solution of system (1).

Theorem 3.1 *Let $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ be a stationary point of (5). Assume that*

(a) $\nabla_x L(w^*)$ is positive semidefinite on $\mathbb{R}^n$;

(b) $\nabla_x L(w^*)$ is positive definite on the cone

$$\mathcal{C}(x^*) := \{v \in \mathbb{R}^n |\, \nabla h(x^*)^T v = 0, \nabla g_i(x^*)^T v = 0 \,(i \in I_>), \nabla g_i(x^*)^T v \leq 0 \,(i \notin I_>)\};$$

and either of the following two conditions holds:

(c1) $\nabla h(x^*)$ has full column rank;

(c2) h is an affine function, and the system $h(x) = 0$ is consistent.

Then w^* is a solution of the KKT system (1).

Proof. Suppose that $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ is a stationary point of (5). Using Propositions 2.7 and 2.2, this can be written as

$$\nabla_x L(w^*) L + \nabla h(x^*) h + \nabla g(x^*) D_a \phi = 0, \tag{6}$$

$$\nabla h(x^*)^T L = 0, \tag{7}$$

$$z^* \geq 0, \quad (z^*)^T(-\nabla g(x^*)^T L + D_b \phi) = 0, \quad -\nabla g(x^*)^T L + D_b \phi \geq 0, \tag{8}$$

where L, h, ϕ, D_a, and D_b are used as abbreviations for $L(w^*)$, $h(x^*)$, $\phi(g(x^*), z^*)$, $D_a(w^*)$, and $D_b(w^*)$, respectively. Note that it follows immediately from Proposition 2.2 and property (3) of the function φ that both D_a and D_b are negative semidefinite and that a diagonal element can be 0 only if the corresponding element in ϕ is 0. Therefore and since these diagonal matrices are always postmultiplied by the vector ϕ in the system

(6)–(8), we can assume without loss of generality that D_a and D_b are negative definite diagonal matrices.

Multiplying (6) by L^T, and taking into account (7), we obtain

$$L^T \nabla_x L(w^*) L + L^T \nabla g(x^*) D_a \phi = 0. \tag{9}$$

Now, using (9) and (8), it is possible to show that

$$L^T \nabla_x L(w^*) L + \phi^T D_b D_a \phi \leq 0. \tag{10}$$

To this end we consider three cases.
1) $z_i^* > 0$. From (8) it follows that $-(\nabla g(x^*)^T L)_i + (D_b \phi)_i = 0$, so that

$$(\nabla g(x^*)^T L)_i = (D_b \phi)_i. \tag{11}$$

2) $z_i^* = 0$ and $g_i(x^*) \geq 0$. In this case, by the property (3) of the function φ, we have

$$\phi_i = 0. \tag{12}$$

3) $z_i^* = 0$ and $g_i(x^*) < 0$. From (8) it follows that $-(\nabla g(x^*)^T L)_i + (D_b \phi)_i \geq 0$, so that $(\nabla g(x^*)^T L)_i \leq (D_b \phi)_i$. Furthermore, since in this case $(D_a \phi)_i < 0$, we have that

$$(D_b \phi)_i (D_a \phi)_i \leq (L^T \nabla g(x^*))_i (D_a \phi)_i. \tag{13}$$

From (11), (12) and (13) we see that

$$\phi^T D_b D_a \phi \leq L^T \nabla g(x^*) D_a \phi,$$

which, in turn, recalling (9), implies (10). Assume now, by contradiction, that $\phi \neq 0$. Then, by Assumption (a) and the positive definiteness of the matrix $D_b D_a$, we get a contradiction to (10), so that we have

$$\phi = 0. \tag{14}$$

From (8) we therefore obtain

$$\nabla g_i(x^*)^T L \leq 0 \qquad (i \in I).$$

For $i \in I_>$, we actually have

$$\nabla g_i(x^*)^T L = 0,$$

cf. (11). Taking into account (7), we see that $L \in \mathcal{C}(x^*)$. Hence

$$L = 0 \tag{15}$$

follows from Assumption (b), (14), and (9). Using (6), (14) and (15), we can conclude that

$$\nabla h(x^*) h = 0.$$

Hence we immediately get

$$h = 0 \tag{16}$$

from Assumption (c1). If, instead, condition (c2) holds, we can reason as in the proof of Theorem 5.2 in [9] in order to see that (16) still holds. Hence we have, by (14), (15) and (16), that w^* satisfies the KKT conditions (1). □

The next corollary easily follows from Theorem 3.1.

Corollary 3.2 *Let $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ be a stationary point of (5). Assume that*

(a) F is monotone, h is affine and g is concave (i.e., each component function of g is concave);

(b) $\nabla_x L(w^*)$ is positive definite on the cone $\mathcal{C}(x^*)$

(c) the system $h(x) = 0$ is consistent.

Then w^* is a solution of the KKT system (1).

We note that condition (b) of Theorem 3.1 as well as of Corollary 3.2 weakens the assumption used in [26] for the case of monotone variational inequalities. There, $\nabla F(x^*)$ is required to be positive definite. Furthermore, we stress that assumption (a) of Corollary 3.2 is satisfied, in particular, for monotone variational inequalities as well as for convex optimization problems. We also note that the assumptions of Corollary 3.2 are obviously satisfied by the example of the introduction. Finally, it may be interesting to remark that condition (b) is certainly satisfied if F is strongly monotone.

4 Algorithm

In this section we describe an algorithm for the solution of problem (5) that takes into account its particular structure.

Before stating our algorithm formally, we begin with some motivational remarks. We first recall that we want to solve the constrained nonsmooth system of equations

$$\Phi(w) = 0, \quad z \geq 0. \tag{17}$$

One simple idea for solving (17) would be to iteratively solve the linearized system

$$H_k \Delta w = -\Phi(w^k), \quad z^k + \Delta z \geq 0, \tag{18}$$

where w^k is the current iteration vector, $H_k \in \partial_B \Phi(w^k)$ and $\Delta w = (\Delta x, \Delta y, \Delta z) \in \mathbb{R}^{n+p+m}$. However, even if the matrix H_k is nonsingular, the system (18) is usually not solvable. Hence it seems reasonable to solve (18) in a linear least squares sense, i.e., to replace the constrained linear system (18) by the following constrained linear least squares problem:

$$\min \frac{1}{2} \|\Phi(w^k) + H_k \Delta w\|^2 \quad \text{s.t.} \quad z^k + \Delta z \geq 0. \tag{19}$$

Now, taking into account Proposition 2.7, it is easy to see that the merit function of (19) can be rewritten as

$$\begin{aligned} \frac{1}{2}\|\Phi(w^k) + H_k \Delta w\|^2 &= \frac{1}{2}\left(\Phi(w^k) + H_k \Delta w\right)^T \left(\Phi(w^k) + H_k \Delta w\right) \\ &= \Psi(w^k) + \nabla\Psi(w^k)^T \Delta w + \frac{1}{2}\Delta w^T H_k^T H_k \Delta w. \end{aligned}$$

Since $\Psi(w^k)$ is just a constant, problem (19) is therefore equivalent to

$$\min \nabla\Psi(w^k)^T \Delta w + \frac{1}{2}\Delta w^T H_k^T H_k \Delta w \quad \text{s.t.} \quad z^k + \Delta z \geq 0. \tag{20}$$

If the system $\Phi(w) = 0$ were differentiable this would simply be a constrained version of the usual Gauss-Newton method, which is known to have some drawbacks [7]; in particular (20) might not have a unique solution. Then, on the basis of analogous results in the smooth case, it seems advisable to consider some kind of modification of the search direction subproblem (20), see e.g. [7]. In this paper we consider a Levenberg-Marquardt-type modification. To this end, let $\rho : \mathbb{R} \to \mathbb{R}_+$ be a *forcing function*, i.e., a continuous function which takes nonnegative values and is 0 if and only if its argument is 0. The subproblem actually used in our algorithm is the following regularized version of (20):

$$\min \nabla\Psi(w^k)^T \Delta w + \frac{1}{2}\Delta w^T \left(H_k^T H_k + \rho(\Psi(w^k))I\right) \Delta w \quad \text{s.t.} \quad z^k + \Delta z \geq 0. \tag{21}$$

Note that the matrix $H_k^T H_k + \rho(\Psi(w^k))I$ is positive definite as long as w^k is not a solution of (17). Since, on the other hand, the feasible set of the quadratic program (21) is obviously nonempty, problem (21) always admits a unique solution.

We can now give a formal description of our algorithm. It basically solves the KKT system by solving a sequence of problems (21). This procedure is globalized by using a simple line search procedure based on the merit function Ψ.

Algorithm 4.1 *(Nonsmooth QP-based Algorithm)*

(S.0) (Initial Data)
Choose $w^0 = (x^0, y^0, z^0) \in \mathbb{R}^{n+p+m}$ *with* $z^0 \geq 0, \sigma \in (0,1), \beta \in (0,1)$, *and set* $k := 0$.

(S.1) (Termination Criterion)
If w^k *is a stationary point of (5): Stop.*

(S.2) (Quadratic Programming Subproblem)
Select an element $H_k \in \partial_B \Phi(w^k)$. *Let* $\Delta w^k = (\Delta x^k, \Delta y^k, \Delta z^k) \in \mathbb{R}^{n+p+m}$ *be the unique solution of the quadratic programming problem* (QP_k)*:*

$$\min \ \nabla\Psi(w^k)^T \Delta w + \frac{1}{2}\Delta w^T \left(H_k^T H_k + \rho(\Psi(w^k))I\right) \Delta w \quad s.t. \quad z^k + \Delta z \geq 0.$$

(S.3) (Line Search)
Let $t_k := \max\{\beta^\ell | \, \ell = 0, 1, 2, \ldots\}$ *such that*

$$\Psi(w^k + t_k \Delta w^k) \leq (1 - \sigma t_k^2) \Psi(w^k) \tag{22}$$

(S.4) (Update)
Set $w^{k+1} := w^k + t_k \Delta w^k$, $k := k+1$, *and go to (S.1).*

It is easy to see that any sequence $\{w^k\} = \{(x^k, y^k, z^k)\} \subset \mathbb{R}^{n+p+m}$ generated by Algorithm 4.1 remains feasible for problem (5), i.e., $z^k \geq 0$ for all k. We also note that the stepsize t_k on the right-hand side of (22) is squared in contrast to usual stepsize selection rules.

The following theorem shows that the algorithm is well-defined.

Theorem 4.2 *Let* $w^k = (x^k, y^k, z^k) \in \mathbb{R}^{n+p+m}$ *with* $z^k \geq 0$ *be an arbitrary vector and* $\Delta w^k \in \mathbb{R}^{n+p+m}$ *be a solution of* (QP_k)*. Then we have*

$$\nabla \Psi(w^k)^T \Delta w^k \leq 0.$$

If w^k *is not a stationary point of problem (5), then*

$$\nabla \Psi(w^k)^T \Delta w^k < 0.$$

Moreover, Algorithm 4.1 is well-defined, in the sense that a positive t_k *can always be found at Step (S.3).*

Proof. Since $\Delta w^k \in \mathbb{R}^{n+p+m}$ is a solution of (QP_k) and $\Delta w = 0$ is feasible for (QP_k), we have

$$\nabla \Psi(w^k)^T \Delta w^k + \frac{1}{2} (\Delta w^k)^T \left(H_k^T H_k + \rho(\Psi(w^k)) I \right) \Delta w^k \leq 0. \tag{23}$$

Since the matrix $H_k^T H_k + \rho(\Psi(w^k)) I$ is always positive semidefinite, (23) implies

$$\nabla \Psi(w^k)^T \Delta w^k \leq 0.$$

Now assume that $\nabla \Psi(w^k)^T \Delta w^k = 0$ and that w^k is not a stationary point of problem (5). Then (23) becomes

$$\frac{1}{2} (\Delta w^k)^T \left(H_k^T H_k + \rho(\Psi(w^k)) I \right) \Delta w^k \leq 0$$

which, since $\rho(\Psi(w^k)) > 0$, is only possible if $\Delta w^k = 0$. Note that Δw^k, as a solution of (QP_k), satisfies in particular the stationary conditions of (QP_k). Writing down these conditions and taking into account the fact that $\Delta w^k = 0$, it can easily be seen that $w^k = (x^k, y^k, z^k)$ satisfies the following conditions:

$$\begin{aligned} \nabla_x \Psi(w^k) &= 0, \\ \nabla_y \Psi(w^k) &= 0, \\ \nabla_z \Psi(w^k) \geq 0, z^k \geq 0, \nabla_z \Psi(w^k)^T z^k &= 0, \end{aligned}$$

i.e., w^k is a stationary point of (5), a contradiction to our assumption. We therefore have

$$\nabla\Psi(w^k)^T\Delta w^k < 0. \tag{24}$$

Assume finally that an iterate w^k exists such that

$$\Psi(w^k + \beta^\ell \Delta w^k) > (1 - \sigma\beta^{2\ell})\Psi(w^k)$$

for all $\ell \geq 0$. Then

$$\frac{\Psi(w^k + \beta^\ell \Delta w^k) - \Psi(w^k)}{\beta^\ell} > -\sigma\beta^\ell\Psi(w^k)$$

follows. Hence, for $\ell \to \infty$, we obtain $\nabla\Psi(w^k)^T\Delta w^k \geq 0$ which contradicts (24). Therefore it is always possible to find a steplength $t_k > 0$ satisfying the line search condition (22), i.e., Algorithm 4.1 is well-defined. □

5 Convergence Analysis

In this section we first investigate the global convergence properties of the algorithm introduced in the previous section and then analyze its convergence rate.

In order to put these results in the right perspective, however, some preliminary considerations are in order. We reformulated the KKT system (1) as a nonsmooth system of equations, and this may seem unnecessarily cumbersome. In fact, it is not difficult to give smooth reformulations of (1). Nevertheless, recent research has clearly established that it is preferable to consider nonsmooth reformulations of systems like (1) since they are usually numerically more stable and guarantee superlinear convergence under weaker assumptions (see, e.g., [6, 14, 20, 23] and references therein). However, the use of nonsmooth reformulations is not without drawbacks: global convergence results are harder to establish and require the use of assumptions that are not needed in the analysis of similar algorithms for the solution of smooth systems of equations. In our view, it is remarkable that we can establish global convergence results under assumptions that exactly parallel those used in the smooth case. More in particular, we shall establish, without any assumption besides those already made, that the nonsmooth QP-based method introduced in the previous section generates a sequence such that every limit point is a stationary point of (5); these limit points will be solutions of (1) under the assumptions of Theorem 3.1. Furthermore, we shall establish a superlinear/quadratic convergence rate under a condition weaker that Robinson's strong regularity. A comparison with results for similar algorithms, see, e.g., [23, 25], show that our results are stronger than already known ones.

5.1 Global Convergence

The aim of this subsection is to prove a global convergence result for Algorithm 4.1 towards stationary points of (5).

The proof of our global convergence result is based on the following stability result for positive definite quadratic programs with lower bound constraints which easily follows from a more general theorem by Daniel [5, Theorem 4.4].

Lemma 5.1 *Consider the quadratic programs*

$$\min \ \frac{1}{2}x^TQx + q^Tx \quad s.t. \quad x_i \geq l_i, \quad i \in L \tag{25}$$

and

$$\min \ \frac{1}{2}x^T\tilde{Q}x + \tilde{q}^Tx \quad s.t. \quad x_i \geq \tilde{l}_i, \quad i \in L \tag{26}$$

where $Q, \tilde{Q} \in \mathbb{R}^{n\times n}$ *with* Q *positive definite,* $q, \tilde{q} \in \mathbb{R}^n$ *and* $L \subseteq I$*. Let us write*

$$\varepsilon := \max_{i\in L}\{\|Q - \tilde{Q}\|, \|q - \tilde{q}\|, |l_i - \tilde{l}_i|\}.$$

Then there exist constants $c > 0$ *and* $\bar{\varepsilon} > 0$ *such that*

$$\|x^* - \tilde{x}^*\| \leq \varepsilon c$$

whenever $\varepsilon \leq \bar{\varepsilon}$*, where* x^* *and* $\tilde{x}^*$ *denote solutions of (25) and (26), respectively.*

We are now in the position to state our global convergence result.

Theorem 5.2 *Every limit point of the sequence generated by Algorithm 4.1 is a stationary point of (5).*

Proof. The sequence $\{\Psi(w^k)\}$ is obviously decreasing and bounded from below by zero, so that it converges to a nonnegative value Ψ^*. If $\Psi^* = 0$, then, by continuity, every limit point is a global solution and hence a stationary point of (5). So consider the case $\Psi^* > 0$. The fact that $\lim_{k\to\infty}\left(\Psi(w^{k+1}) - \Psi(w^k)\right) = 0$ and (22) gives

$$\lim_{k\to\infty} t_k^2\Psi(w^k) = 0. \tag{27}$$

Assume now that w^* is an accumulation point of $\{w^k\}$ and that $\{w^k\}_{K_1}$ is a subsequence converging to w^*. In view of the upper semicontinuity of the B-subdifferential (see [3]), it follows that the sequence $\{H_k\}_{K_1}$ remains bounded. Hence there is a subsubsequence $\{H_k\}_{K_2}, K_2 \subseteq K_1$, such that $\{H_k\}_{K_2}$ converges to some matrix $H_* \in \partial_B\Phi(w^*)$. Let us denote by Δw^* the (unique) solution of the quadratic program

$$\min \ \nabla\Psi(w^*)^T\Delta w + \frac{1}{2}\Delta w^T(H_*^TH_* + \rho(\Psi^*)I)\Delta w \quad \text{s.t.} \quad z^* + \Delta z \geq 0. \tag{28}$$

Since Δw^k is a solution of (QP_k) and $w^k \to w^*, H_k \to H_*, \nabla\Psi(w^k) \to \nabla\Psi(w^*)$ and $\rho(\Psi(w^k)) \to \rho(\Psi^*) > 0$ by the continuity of the forcing function ρ (all limits being taken on a subsequence), it follows immediately from Lemma 5.1 that $\{\Delta w^k\}_{K_2} \to \Delta w^*$. We now show that (27) implies

$$\nabla\Psi(w^*)^T\Delta w^* = 0. \tag{29}$$

Since $\Psi^* > 0$, (27) and (22) yield $t_k \to 0$. Let ℓ_k be the unique integer taken in Step (S.3) of Algorithm 4.1 such that $t_k = \beta^{\ell_k}$. Then it follows from (22) that

$$\frac{\Psi(w^k + \beta^{\ell_k - 1}\Delta w^k) - \Psi(w^k)}{\beta^{\ell_k - 1}} > -\sigma\beta^{\ell_k - 1}\Psi(w^k). \tag{30}$$

In view of $\ell_k \to \infty$, we therefore obtain, taking the limit $k \to \infty$ $(k \in K_2)$ in (30):

$$\nabla\Psi(w^*)^T \Delta w^* \geq 0.$$

Hence, by Theorem 4.2, w^* is a stationary point of (5). □

Theorem 5.2 guarantees a subsequential convergence to stationary points of problem (5). Conditions for such a stationary point to be a solution of the original KKT system (1) were given in Section 3.

5.2 Local Convergence

In this section we want to show that Algorithm 4.1 is locally fast convergent under suitable assumptions. The probably most famous conditions for establishing local superlinear and quadratic convergence of an algorithm that locates a solution of (1) are the following three: A KKT triple $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ of (1) satisfies the

(A1): *nondegeneracy condition*, i.e., $g_i(x^*) + z_i^* > 0$ for all $i \in I$,

(A2): *linear independence constraint qualification*, i.e., the gradients $\nabla h_j(x^*)$ $(j \in J)$ and $\nabla g_i(x^*)$ $(i \in I_0)$ are linearly independent,

(A3): *second order condition*, i.e., the Jacobian $\nabla_x L(w^*)$ is positive definite on the subspace $\{v \in \mathbb{R}^n | \, \nabla h(x^*)^T v = 0, \, \nabla g_i(x^*)^T v = 0 \ (i \in I_0)\}$.

Note, however, that even a nondegenerate solution of (1) is in general a degenerate solution of our reformulation (5). This can be seen by observing first that, with regard to Proposition 2.7, $\nabla\Psi(w^*) = 0$ for any solution of (1). Then, the stationary conditions of problem (5) show that w^* is a degenerate solution of (5) if $z_i^* = 0$ for at least one index i.

In order to avoid the nondegeneracy condition (A1), Assumption (A3) is usually replaced by

(A3'): *strong second order condition*, i.e., the Jacobian $\nabla_x L(w^*)$ is positive definite on the subspace $\{v \in \mathbb{R}^n | \, \nabla h(x^*)^T v = 0, \, \nabla g_i(x^*)^T v = 0 \ (i \in I_+)\}$.

Obviously, if (A1) holds, then (A3) and (A3') are equivalent, so that (A1)–(A3) imply (A2) and (A3'), whereas in general (A3') is a stronger condition than (A3). On the other hand, it is known that (A2) and (A3') together imply Robinson's strong regularity condition, see [29]. As far as we are aware of, there are only two algorithms for the solution of constrained optimization or variational inequality problems which are known to be fast convergent under Robinson's strong regularity condition: Josephy's method [19] and Bonnans' one [2]. Besides beeing purely local, both this methods require, at each iteration, the solution of a (possibly) non symmetric linear complementarity problem. The only assumption we will use in this section is that a KKT triple $w^* = (x^*, y^*, z^*)$ of (1) is a BD-regular solution of the system $\Phi(w) = 0$. In view of Proposition 2.3, a strongly regular KKT point is, in particular, BD-regular.

In order to prove our local convergence theorem, we will need some lemmas.

Lemma 5.3 *Assume that* $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ *is a BD-regular solution of* $\Phi(w) = 0$*. Then there exists a constant* $c > 0$ *such that*

$$\|\Delta w^k\| \leq c\|\Phi(w^k)\|$$

for all $w^k = (x^k, y^k, z^k)$ *with* $z^k \geq 0$ *sufficiently close to* w^**, where* Δw^k *denotes a solution of* (QP_k)*.*

Proof. Since w^* is a BD-regular KKT point, the matrices $H_k \in \partial_B \Phi(w^k)$ are uniformly nonsingular for all w^k sufficiently close to w^* by Proposition 2.6 (a), i.e., there exists a constant $c_1 > 0$ such that

$$\|\Delta w^k\| \leq \|H_k^{-1}\| \, \|H_k \Delta w^k\| \leq c_1 \|H_k \Delta w^k\|. \tag{31}$$

On the other hand, since Δw^k is a solution of (QP_k) and $\Delta w = 0$ is feasible for (QP_k), it follows from Proposition 2.7 and the Cauchy-Schwarz inequality that

$$\begin{aligned} 0 &\geq \nabla \Psi(w^k)^T \Delta w^k + \frac{1}{2}(\Delta w^k)^T \left(H_k^T H_k + \rho(\Psi(w^k)) I \right) \Delta w^k \\ &\geq \nabla \Psi(w^k)^T \Delta w^k + \frac{1}{2}(\Delta w^k)^T H_k^T H_k \Delta w^k \\ &= \Phi(w^k)^T H_k \Delta w^k + \frac{1}{2}\|H_k \Delta w^k\|^2 \\ &\geq \frac{1}{2}\|H_k \Delta w^k\|^2 - \|\Phi(w^k)\| \, \|H_k \Delta w^k\|, \end{aligned}$$

so that

$$\|H_k \Delta w^k\| \leq 2\|\Phi(w^k)\|. \tag{32}$$

Combining (31) and (32) yields

$$\|\Delta w^k\| \leq c\|\Phi(w^k)\|$$

with $c := 2c_1$. □

Lemma 5.4 *Suppose that* $w^* = (x^*, y^*, z^*) \in \mathbb{R}^{n+p+m}$ *is a BD-regular solution of* $\Phi(w) = 0$*. Let* $\{w^k\} = \{(x^k, y^k, z^k)\} \subset \mathbb{R}^{n+p+m}$ *denote any sequence that converges to* w^* *and that satisfies* $z^k \geq 0$ *for all* k*. For each* w^k *let* Δw^k *denote a solution of* (QP_k)*. Then*

$$\|w^k + \Delta w^k - w^*\| = o(\|w^k - w^*\|).$$

Moreover, if ∇F*,* $\nabla^2 h_j$ $(j \in J)$*, and* $\nabla^2 g_i$ $(i \in I)$ *are locally Lipschitzian, and if* $\rho(\Psi(w^k)) = O(\Psi(w^k))$*, we have*

$$\|w^k + \Delta w^k - w^*\| = O(\|w^k - w^*\|^2).$$

Proof. By the BD-regularity of w^* we have for w^k sufficiently close to w^* and $H_k \in \partial_B\Phi(w^k)$:

$$\|w^k + \Delta w^k - w^*\| \leq \|H_k^{-1}\| \, \|H_k(w^k + \Delta w^k - w^*)\| \leq c_1\|H_k\Delta w^k + H_k(w^k - w^*)\|, \quad (33)$$

where c_1 denotes the constant from Proposition 2.6 (a). Since the mapping Φ is semismooth by Proposition 2.4 (a), we obtain

$$\|\Phi(w^k) - \Phi(w^*) - H_k(w^k - w^*)\| = o(\|w^k - w^*\|) \quad (34)$$

by Proposition 2.5 (a). Moreover, if ∇F, $\nabla^2 h$ and $\nabla^2 g$ are locally Lipschitzian, Φ is strongly semismooth (see Proposition 2.4 (b)), so that

$$\|\Phi(w^k) - \Phi(w^*) - H_k(w^k - w^*)\| = O(\|w^k - w^*\|^2) \quad (35)$$

by Proposition 2.5 (b) follows. Since Δw^k is a solution of (QP_k) and $\Delta\hat{w}^k := w^* - w^k$ is obviously feasible for (QP_k), we obtain, using Proposition 2.7,

$$\begin{aligned}
\frac{1}{2}\|\Phi(w^k) + H_k\Delta w^k\|^2 &= \Psi(w^k) + \nabla\Psi(w^k)^T\Delta w^k + \frac{1}{2}(\Delta w^k)^T H_k^T H_k \Delta w^k \\
&\leq \Psi(w^k) + \nabla\Psi(w^k)^T\Delta w^k + \\
&\quad \frac{1}{2}(\Delta w^k)^T \left(H_k^T H_k + \rho(\Psi(w^k))I\right) \Delta w^k \\
&\leq \Psi(w^k) + \nabla\Psi(w^k)^T\Delta\hat{w}^k + \\
&\quad \frac{1}{2}(\Delta\hat{w}^k)^T \left(H_k^T H_k + \rho(\Psi(w^k))I\right) \Delta\hat{w}^k \\
&= \frac{1}{2}\|\Phi(w^k) + H_k\Delta\hat{w}^k\|^2 + \frac{1}{2}\rho(\Psi(w^k))\|\Delta\hat{w}^k\|^2 \\
&= \frac{1}{2}\|\Phi(w^k) - H_k(w^k - w^*)\|^2 + \frac{1}{2}\rho(\Psi(w^k))\|w^k - w^*\|^2 \\
&= \frac{1}{2}\|\Phi(w^k) - \Phi(w^*) - H_k(w^k - w^*)\|^2 + \\
&\quad \frac{1}{2}\left(\sqrt{\rho(\Psi(w^k))}\|w^k - w^*\|\right)^2.
\end{aligned}$$

Hence, we obtain from (34) and from $\rho(\Psi(w^k)) \to 0$ that

$$\|\Phi(w^k) + H_k\Delta w^k\| = o(\|w^k - w^*\|). \quad (36)$$

Therefore, using (33), (34) and (36), we get

$$\begin{aligned}
\|w^k + \Delta w^k - w^*\| &\leq c_1\|H_k\Delta w^k + H_k(w^k - w^*)\| \\
&\leq c_1\left(\|\Phi(w^k) + H_k\Delta w^k\| + \|\Phi(w^k) - H_k(w^k - w^*)\|\right) \\
&= o(\|w^k - w^*\|),
\end{aligned}$$

so that the first statement of the lemma follows. In order to prove the second part, first note that since Φ is locally Lipschitzian there is a constant $L > 0$ such that

$$\|\Phi(w^k)\| = \|\Phi(w^k) - \Phi(w^*)\| \leq L\|w^k - w^*\|$$

for all k sufficiently large. Hence we have

$$\Psi(w^k) = O(\|w^k - w^*\|^2)$$

and therefore, by our assumption,

$$\sqrt{\rho(\Psi(w^k))} = O(\|w^k - w^*\|). \tag{37}$$

The second part can now be shown in a similar way as the first part by using (35) instead of (34) and by taking into account (37). □

An immediate consequence of the proof of Lemma 5.4 is the following corollary.

Corollary 5.5 *Suppose that w^* is a BD-regular solution of $\Phi(w) = 0$. Then, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that, whenever $\|w^k - w^*\| \leq \delta$ with $z^k \geq 0$,*

$$\|w^k + \Delta w^k - w^*\| \leq \varepsilon \|w^k - w^*\|.$$

We now state the local convergence result.

Theorem 5.6 *Let $\{w^k\}$ be an infinite sequence generated by Algorithm 4.1, and let w^* be an accumulation point of this sequence. If w^* is a BD-regular solution of the system $\Phi(w) = 0$, then the following statements hold:*

(a) The whole sequence $\{w^k\}$ converges to w^*.

(b) There is an index k_0 such that $t_k = 1$ for all $k \geq k_0$.

(c) The rate of convergence is Q-superlinear.

(d) The rate of convergence is Q-quadratic if, in addition, the assumptions of the second part of Lemma 5.4 are satisfied.

Proof. Let $\{w^k\}_K$ denote a subsequence converging to w^*. Choose $\varepsilon > 0$ such that

$$\varepsilon \leq \min\{1, \frac{c_2(1-\sigma)}{L}\}, \tag{38}$$

where $c_2 > 0$ is the constant from Proposition 2.6 (b) and $L > 0$ is the local Lipschitz-constant of Φ in the ball around w^* with radius δ_2, where δ_2 is also taken from Proposition 2.6 (b). For ε according to (38) we take $\delta > 0$ as given by Corollary 5.5. Then, using Corollary 5.5 and Proposition 2.6 (b), we get, for w^k with $\|w^k - w^*\| \leq \min\{\delta, \delta_2\}$ and $z^k \geq 0$:

$$\begin{aligned}
\sqrt{2}\left(\Psi(w^k + \Delta w^k)^{\frac{1}{2}} - (1-\sigma)^{\frac{1}{2}}\Psi(w^k)^{\frac{1}{2}}\right) &\leq \sqrt{2}\left(\Psi(w^k + \Delta w^k)^{\frac{1}{2}} - (1-\sigma)\Psi(w^k)^{\frac{1}{2}}\right) \\
&= \|\Phi(w^k + \Delta w^k) - \Phi(w^*)\| - \\
&\quad (1-\sigma)\|\Phi(w^k)\| \\
&\leq L\|w^k + \Delta w^k - w^*\| - \\
&\quad c_2(1-\sigma)\|w^k - w^*\| \\
&\leq (L\varepsilon - c_2(1-\sigma))\|w^k - w^*\| \\
&\leq 0.
\end{aligned}$$

Therefore,

$$\Psi(w^k + \Delta w^k) \leq (1-\sigma)\Psi(w^k) \tag{39}$$

follows and Algorithm 4.1 takes the stepsize $t_k = 1$, i.e., $w^{k+1} = w^k + \Delta w^k$. On the other hand, Corollary 5.5 yields

$$\|w^k + \Delta w^k - w^*\| \leq \varepsilon \|w^k - w^*\| \leq \varepsilon \min\{\delta, \delta_2\}.$$

Hence, by induction, we see that $t_k = 1$ and $w^{k+1} = w^k + \Delta w^k$ holds for all k sufficiently large. In particular, we have $\|w^k - w^*\| \leq \delta_2$ for all k sufficiently large. This together with $\|\Phi(w^k)\| \to 0$ (due to (39)) and with Proposition 2.6 implies that the whole sequence $\{w^k\}$ converges to w^*.

The rate of convergence now directly follows from Lemma 5.4. □

References

[1] Bertsekas DP (1982) Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York.

[2] Bonnans JF (1994) Local analysis of Newton type methods for variational inequalities and nonlinear programming. Appl Math Optim 29: 161–186.

[3] Clarke FH (1983) Optimization and Nonsmooth Analysis, John Wiley and Sons, New York.

[4] Cottle RW, Pang JS, Stone RE (1992) The Linear Complementarity Problem, Academic Press, Boston.

[5] Daniel JW (1973) Stability of the solution of definite quadratic programs. Math Programming 5: 41–53.

[6] De Luca T, Facchinei F, Kanzow C (1996) A semismooth equation approach to the solution of nonlinear complementarity problems. Math Programming 75: 407–439.

[7] Dennis JE, Schnabel RB (1983) Numerical methods for unconstrained optimization and nonlinear equations, Prentice-Hall, Englewood-Cliffs, New Jersey.

[8] Dirkse SP, Ferris MC (1995) MCPLIB: A collection of nonlinear mixed complementarity problems. Optim Methods Software 5: 123–156.

[9] Facchinei F, Fischer A, Kanzow C (1998) Regularity properties of a semismooth reformulation of variational inequalities. SIAM J Optim, to appear.

[10] Facchinei F, Fischer A, Kanzow C (1997) A semismooth Newton method for variational inequalities: The case of box constraints. In Ferris MC, Pang JS (eds.): Complementarity and Variational Problems: State of the Art. SIAM, Philadelphia, PA, 76–90.

[11] Facchinei F, Kanzow C (1997) A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems. Math Programming 76: 493–512.

[12] Facchinei F, Soares J (1997) A new merit function for nonlinear complementarity problems and a related algorithm. SIAM J Optim 7: 225–247.

[13] Ferris MC, Pang JS (1997) Engineering and economic applications of complementarity problems. SIAM Review, to appear.

[14] Fischer A (1992) A special Newton-type optimization method. Optim 24: 269–284.

[15] Fischer A (1995) Solution of monotone complementarity problems with locally Lipschitzian functions. Math Programming 76: 513-532.

[16] Fischer A (1998) A new constrained optimization reformulation for complementarity problems. J Optim Theory Appl, to appear.

[17] Fletcher R (1987) Practical Methods of Optimization, 2nd edn. John Wiley and Sons, Chichester.

[18] Harker PT, Pang JS (1990) Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. Math Programming 48: 161–220.

[19] Josephy NH (1979) Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, WI.

[20] Kanzow C, Kleinmichel H (1995) A class of Newton-type methods for equality and inequality constrained optimization. Optim Methods Software 5: 173–198.

[21] Liu J (1995) Strong stability in variational inequalities. SIAM J Control Optim 33: 725–749.

[22] Mifflin R (1977) Semismooth and semiconvex functions in constrained optimization. SIAM J Control Optim 15: 957–972.

[23] Pang JS, Gabriel SA (1993) NE/SQP: A robust algorithm for the nonlinear complementarity problem. Math Programming 60: 295–337.

[24] Pang JS, Qi L (1993) Nonsmooth equations: motivation and algorithms. SIAM J Optim 3: 443–465.

[25] Pang JS and Qi L (1995) A globally convergent Newton method for convex SC^1 minimization problems. J Optim Theory Appl 85: 633–648.

[26] Peng JM (1995) A global method for monotone variational inequality problems with inequality constraints. Technical Report, State Key Laboratory of Scientific and Engineering Computing, Academic Sinica, Beijing, China, 1995.

[27] Qi L (1993) A convergence analysis of some algorithms for solving nonsmooth equations. Math Oper Res 18: 227–244.

[28] Qi L, Sun J (1993) A nonsmooth version of Newton's method. Math Programming 58: 353–368.

[29] Robinson SM (1980) Strongly regular generalized equations. Math Oper Res 5: 43–62.