SIAM J. OPTIM.
Vol. 10, No. 3, pp. 878–895

REFORMULATION OF VARIATIONAL INEQUALITIES ON A SIMPLEX AND COMPACTIFICATION OF COMPLEMENTARITY PROBLEMS*

ROBERTO ANDREANI† AND JOSÉ MARIO MARTÍNEZ‡

Abstract. Many variational inequality problems (VIPs) can be reduced, by a compactification procedure, to a VIP on the canonical simplex. Reformulations of this problem are studied, including smooth reformulations with simple constraints and unconstrained reformulations based on the penalized Fischer–Burmeister function. It is proved that bounded level set results hold for these reformulations under quite general assumptions on the operator. Therefore, it can be guaranteed that minimization algorithms generate bounded sequences and, under monotonicity conditions, these algorithms necessarily find solutions of the original problem. Some numerical experiments are presented.

Key words. variational inequalities, complementarity, minimization algorithms, reformulation

AMS subject classifications. 90C33, 90C30

PII. S1052623499352826

1. Introduction.

We are interested in reformulations of variational inequality problems (VIPs) where the domain is a simplex. The main motivation is that variational inequalities on generalized (perhaps unbounded) boxes can be reduced to the simplex case if one knows appropriate lower bounds for each variable and a bound for the sum of the variables. The reformulations of the VIP on a simplex do not have, in principle, bounded variables. However, we will be able to show, for some reformulations, that the objective function has bounded level sets. It is worth mentioning that reformulations of complementarity problems do not have, in general, bounded level sets, unless suitable restrictions are imposed on the problem. Therefore, when one applies a general solver to such a reformulation, the risk of divergence exists, even when one knows that stationary points are solutions of the VIP.

The following example will clarify the compactification strategy. Suppose that we want to solve the nonlinear system of equations

$$F(x) = 0, \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ has continuous first derivatives. Usually, globally convergent algorithms for solving (1.1) rely on the unconstrained minimization problem

$$\text{Minimize } \|F(x)\|_2^2. \tag{1.2}$$

See [12, 38, 41]. (Obviously, (1.1) is a variational inequality problem where the domain is $\mathbb{R}^n$.) Most algorithms (for example, globalizations of Newton's method) have the property that every limit point of the iterates is stationary, that is,

$$\nabla \|F(x)\|_2^2 \equiv 2F'(x)^T F(x) = 0. \tag{1.3}$$

*Received by the editors March 3, 1999; accepted for publication (in revised form) October 8, 1999; published electronically May 2, 2000.

http://www.siam.org/journals/siopt/10-3/35282.html

†Department of Computer Science and Statistics, University of the State of S. Paulo (UNESP), C.P. 136, CEP 15054-000, São José do Rio Preto-SP, Brazil (andreani@nimitz.dcce.ibilce.unesp.br). This author was supported by FAPESP (grant 99/08381-4) and CNPq (grant 301115/96-6).

‡Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil (martinez@ime.unicamp.br). This author was supported by PRONEX, FAPESP (grant 90-3724-6), CNPq, and FAEP-UNICAMP.

The most obvious drawbacks of this approach are

(1) The algorithm might converge to a stationary point that is not a solution ($F'(x)$ being singular in this case).

(2) Limit points of the generated sequence might not exist at all.

With the aim of guaranteeing the existence of limit points (and, in fact, avoiding possible overflows in the computer calculation), artificial bounds are frequently added to (1.2). In this case, the globalization procedure must use techniques of bound-constrained minimization [7, 10, 18, 19, 31, 36] and the limit points will be stationary points of the *bound-constrained* problem. Unfortunately, if an artificial constraint is active at the limit point, the limit point might not be a stationary point of (1.2) and, hence, also not a solution of (1.1). This usually happens when the sequence generated by the unconstrained algorithm applied to (1.2) tends to infinity.

In [5] it has been suggested that a better way in which the domain of (1.1) can be compactified is to consider the *variational inequality problem* defined by $F(x)$ on the domain defined by the artificial bounds. In that paper it was proved that, under suitable conditions, any *stationary* point of a smooth reformulation (with bounded level sets) of the variational inequality problem must be a *solution* of (1.1). Therefore, the results in [5] represent sufficient conditions under which neither of the two objections just exposed is problematic.

The discussion above can be repeated, with minor modifications, if the original problem is a complementarity problem instead of a nonlinear system. See [2, 3, 4, 5, 13, 24, 34, 35].

The main drawback of the approach of [5] is that the reformulation of the original problem requires $2n$ additional variables. In [4, 14] a different reformulation with the same triplicating property can be found. In the present research, we introduce reformulations with $n+3$ additional variables having similar properties as those proved in [5]. The idea, as we mentioned above, is to consider first the variational inequality problem on a smaller simplicial region.

This paper is organized as follows. In section 2 we explain how a variational inequality problem on a generalized box can be reduced to a VIP in which the domain is a simplex. In section 3 we define *smooth* reformulations of the VIP on the simplex, for which the level sets are bounded and, under suitable conditions, stationary points coincide with solutions of the variational inequality problem. In section 4 we repeat the work of section 3 with respect to an unconstrained reformulation that uses the penalized Fischer–Burmeister [8] function. See, also, [11, 14, 16, 17, 23, 27, 29, 30]. In section 5 we present numerical experiments. Conclusions are given in section 6.

2. Reduction to the simplex form. The fact that, under certain conditions, the solution of a restricted variational inequality problem is a solution of the original one seems to be known by many researchers, although the result is not easily found in the literature. The argument is as follows.

Consider the variational inequality problem $VIP(F, \Omega)$, which consists of finding $x \in \Omega$ such that

$$\langle F(x), z - x \rangle \geq 0 \quad \forall z \in \Omega, \tag{2.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ and Ω is closed and convex. Let $B \subset \mathbb{R}^n$ be closed and convex too. Denote by B' the set of interior points of B. Define $\Omega_{small} = \Omega \cap B$ and consider the variational inequality problem defined by

$$\langle F(x), z - x \rangle \geq 0 \quad \forall z \in \Omega_{small}. \tag{2.2}$$

Clearly, any solution of (2.1) that belongs to Ω_{small} will be a solution of (2.2). Let us show that, under certain conditions, every solution of (2.2) is a solution of (2.1). Denote by S_{small} the set of solutions of (2.2). Essentially, the proof of the following theorem is that given (under slightly stronger hypotheses) in [45].

THEOREM 2.1. *Assume that the set of solutions of* (2.1) *is closed. Assume, moreover, that for every solution x of* (2.2) *there exists a sequence $\{x^k\} \subset B' \cap S_{small}$ such that $\lim x^k = x$. Then, every solution of* (2.2) *solves* (2.1).

Proof. Let x be a solution of (2.2). Let $\{x^k\} \subset B' \cap S_{small}$ be the sequence (convergent to x) that is mentioned in the hypothesis. Let $z \in \Omega$. Since $x^k \in B'$ and Ω is convex, there exists $t > 0$ such that $x^k + t(z - x^k) \in \Omega_{small}$. Therefore, since x^k solves (2.2), $\langle F(x^k), t(z - x^k)\rangle \geq 0$. So, $\langle F(x^k), z - x^k\rangle \geq 0$. Since z and k are arbitrary, this means that x^k solves (2.1) for all k. But the set of solutions of (2.1) is closed, so x also solves (2.1). $\square$

Now, we consider the problem $VIP(F, \Omega_1)$, where

$$\Omega_1 = \{x \in \mathbb{R}^n \mid x_i \geq 0 \quad \forall\, i \in I\} \tag{2.3}$$

and $I \subset \{1, \ldots, n\}$. Nonlinear complementarity problems (NCPs) and nonlinear systems are particular cases of $VIP(F, \Omega_1)$, where $I = \{1, \ldots, n\}$ and $I = \emptyset$, respectively.

Define

$$\Omega_2 = \left\{ x \in \mathbb{R}^n \mid x \geq \ell \ \text{ and } \ \sum_{i=1}^{n} x_i \leq M \right\},$$

where $\ell \in \mathbb{R}^n$, $\ell_i = 0$ for all $i \in I$, and $\sum_{i=1}^n \ell_i < M$. Clearly, $\Omega_2 \subset \Omega_1$. We denote by S_2 the set of solutions of $VIP(F, \Omega_2)$. The application of Theorem 2.1 to $VIP(F, \Omega_1)$ is given in the following theorem.

THEOREM 2.2. *Suppose that F is continuous on Ω_1, S_2 is convex, and there exists $\bar{x} \in S_2$ such that $\sum_{i=1}^n \bar{x}_i < M$ and $\bar{x}_i > \ell_i$ for all $i \notin I$. Then, any solution of $VIP(F, \Omega_2)$ solves $VIP(F, \Omega_1)$.*

Proof. Since F is continuous, the set of solutions of $VIP(F, \Omega_1)$ is closed. Let $x \in S_2$. By the convexity of S_2, we have that $[\bar{x}, x) \subset S_2$. Moreover, for all $y \in [\bar{x}, x)$, we have that $\sum_{i=1}^n y_i < M$ and $y_i > \ell_i$, $i \notin I$. Therefore, the hypothesis of Theorem 2.1 holds for the sequence defined by $x^k = x + \frac{1}{k}(\bar{x} - x)$. This implies the desired result. $\square$

Defining $G_1 : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ by

$$G_1(y, x_{n+1}) = (F(y), 0) \quad \forall\, y \in \mathbb{R}^n, x_{n+1} \in \mathbb{R}, \tag{2.4}$$

$$\Omega_3 = \left\{ x \in \mathbb{R}^{n+1} \mid \sum_{i=1}^{n+1} x_i = M \ \text{ and } \ x_i \geq \ell_i, i = 1, \ldots, n+1 \right\},$$

and $\ell_{n+1} = 0$, it is easy to see that solving $VIP(G_1, \Omega_3)$ is equivalent to solving $VIP(F, \Omega_2)$. Finally, after a suitable change of variables, we can consider that $M = 1$ and $\ell_i = 0$ for all $i = 1, \ldots, n+1$, so that the original problem is reduced to a variational inequality problem on the canonical simplex.

3. Bounded smooth reformulations. The discussion in section 2 justifies the study of the problem $VIP(G, \mathcal{S})$, where $G : \mathbb{R}^m \to \mathbb{R}^m$ and

$$\mathcal{S} = \left\{ x \in \mathbb{R}^m \mid x \geq 0 \text{ and } \sum_{i=1}^{m} x_i = 1 \right\}.$$

According to [1, 22] (see, also, [2, 3, 4, 15, 20, 21, 32]) we define the following reformulation for $VIP(G, \mathcal{S})$:

$$\text{Minimize } \Phi_1(x, v, \lambda) \quad \text{subject to } x \geq 0, v \geq 0, \tag{3.1}$$

where

$$\Phi_1(x, v, \lambda) = \rho_0 \|G(x) + \mathbf{1}\lambda - v\|_2^2 + \rho_1 \left(\sum_{i=1}^{m} x_i - 1 \right)^2 + \langle x, v \rangle^2,$$

$x, v \in \mathbb{R}^m$, $\lambda \in \mathbb{R}$, $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^m$, $\rho_0, \rho_1 > 0$.

Let us prove that $\Phi_1(x, v, \lambda)$ has bounded level sets on the set $x \geq 0, v \geq 0$. From now on, we denote $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$.

THEOREM 3.1. *Assume that G is continuous on $\mathbb{R}^m_+$. Then, for all $\theta \in (0, \rho_1)$, the set*

$$L_1 = \{(x, v, \lambda) \in \mathbb{R}^{2m+1} \mid x \geq 0, v \geq 0, \ \Phi_1(x, v, \lambda) \leq \theta\}$$

is bounded.

Proof. Assume that

$$(x^k, v^k, \lambda_k) \in L_1 \quad \forall \ k = 0, 1, 2, \ldots.$$

Since $(\sum_{i=1}^m x_i^k - 1)^2 \leq \theta$ and $x^k \geq 0$ for all $k = 0, 1, 2, \ldots$, we have that the sequence $\{x^k\}$ is bounded. Therefore, by the continuity of G, the sequence $\{G(x^k)\}$ is also bounded.

Since $\rho_0 \|G(x^k) + \mathbf{1}\lambda_k - v^k\|_2^2 \leq \theta$ for all $k = 0, 1, 2, \ldots$ and $\{G(x^k)\}$ is bounded, we have that $\{\lambda_k - v_i^k\}$ is bounded for all $i = 1, \ldots, m$. Therefore, if $\{v_i^k\}$ is bounded for some $i \in \{1, \ldots, m\}$, λ_k is also bounded, implying that v_i^k is bounded for all $i = 1, \ldots, m$.

Now, if there exists $j \in \{1, \ldots, m\}$ such that $\{v_j^k\}$ is unbounded, we can extract a subsequence such that $v_j^k \to \infty$. For the same subsequence, $\lambda_k \to \infty$ and, so, $v_i^k \to \infty$ for all $i = 1, \ldots, m$. Let us call, for this subsequence, $v^k = (v_1^k, \ldots, v_m^k)$.

But $\langle x^k, v^k \rangle^2 \leq \theta$ for all $k = 0, 1, 2, \ldots$. Therefore, since $x^k \geq 0$ for all k, we have that

$$\lim_{k \to \infty} x^k = 0.$$

Hence,

$$\lim_{k \to \infty} \rho_1 \left(\sum_{i=1}^{m} x_i^k - 1 \right)^2 = \rho_1.$$

Thus, for k large enough, $\Phi_1(x^k, v^k, \lambda_k) > \theta$ and, so, $(x^k, v^k, \lambda_k) \notin L_1$. This means that the assumption on the unboundedness of v_i^k is not possible. This completes the proof. $\square$

It is easy to find (x^0, v^0, λ_0) such that $\Phi_1(x^0, v^0, \lambda_0) < \rho_1$. For example, take $x^0 \geq 0$ such that $\sum_{i=1}^m x_i^0 = 1$, arbitrarily choosing $\lambda_0 \in \mathbb{R}$ and $v^0 \geq 0$. Therefore,

$$\Phi_1(x^0, v^0, \lambda_0) = \rho_0 \|G(x^0) + \mathbf{1}\lambda_0 - v^0\|_2^2 + \langle x^0, v^0 \rangle^2$$

and, so, the condition $\Phi_1(x^0, v^0, \lambda_0) < \rho_1$ holds if we choose

$$\rho_1 > \rho_0 \|G(x^0) + \mathbf{1}\lambda_0 - v^0\|_2^2 + \langle x^0, v^0 \rangle^2.$$

Theorem 3.1 implies that, with the proper choice of x^0, ρ_0, and ρ_1, any reasonable iterative minimization algorithm for solving (3.1) necessarily produces a sequence that has limit points. In fact, the sequence generated by such an algorithm will satisfy

$$\Phi_1(x^k, v^k, \lambda^k) \leq \Phi_1(x^0, v^0, \lambda^0) \quad \forall\ k = 0, 1, 2, \ldots;$$

so, by Theorem 3.1, (x^k, v^k, λ^k) will be bounded. Moreover, for most iterative minimization algorithms, limit points are stationary (KKT) points of the minimization problem. This guarantees that stationary points of problem (3.1) will be necessarily found. It remains to relate the stationary points of (3.1) to the solutions of $VIP(G, \mathcal{S})$. This is done in the following theorem.

THEOREM 3.2. *If G is monotone and has continuous first derivatives, all the stationary points of* (3.1) *are solutions of* $VIP(G, \mathcal{S})$.

Proof. Since $\mathcal{S}$ is bounded, this result follows from Theorem 4 of [22]. See also [1, 20]. ☐

It is easy to see that G_1, defined by (2.4), is monotone if and only if F is monotone. Therefore, stationary points of (3.1) define (after changing variables) solutions of $VIP(F, \Omega_2)$. Under the interiority hypothesis of Theorem 2.2, these are also solutions of $VIP(F, \Omega_1)$.

An interesting consequence of the results of this section comes from analyzing the nonlinear system $F(x) = 0$, where F is monotone (see [43]) but $\|F(x)\|_2^2$ has stationary points or even local minimizers that are not solutions of the system. Essentially, in this section it has been proved that if one selects adequate artificial bounds ℓ and M and the reformulation (3.1) is applied, there is no risk of convergence to spurious stationary points of the squared norm of F.

We finish this section considering a different smooth reformulation of $VIP(G, \mathcal{S})$. See [37]. Consider the minimization problem

$$\text{Minimize } \Phi_2(x, v, \lambda) \quad \text{subject to} \quad x \geq 0, v \geq 0, \tag{3.2}$$

where

$$\Phi_2(x, v, \lambda) = \rho_0 \|G(x) + \mathbf{1}\lambda - v\|_2^2 + \rho_1 \left(\sum_{i=1}^m x_i - 1 \right)^2 + \sum_{i=1}^m (x_i v_i)^2.$$

As in the case of (3.1) it is easy to see that solutions of $VIP(G, \mathcal{S})$ correspond to global solutions of (3.2) for which the objective function vanishes. Moreover, the following results can be proved using the same techniques of Theorem 3.1 and Theorem 3.2. Finally, an initial bounded level set can be obtained choosing ρ_1 similarly to above.

THEOREM 3.3. *Assume that G is continuous on $\mathbb{R}_+^m$. Then, for all $\theta \in (0, \rho_1)$, the set*

$$L_2 \equiv \{(x, v, \lambda) \in \mathbb{R}^{2m+1} \mid x \geq 0, v \geq 0,\ \Phi_2(x, v, \lambda) \leq \theta\}$$

is bounded.

THEOREM 3.4. *If G is monotone and has continuous first derivatives, all the stationary points of* (3.2) *are solutions of* $VIP(G, \mathcal{S})$.

Remark. The compactification procedure is essential to guarantee that stationary points of smooth reformulations are solutions of the associated monotone NCPs. For example, consider the NCP defined by $F : \mathbb{R}^1 \to \mathbb{R}^1$, where

$$F(x) = \begin{cases} -1 & \text{if } x \leq 1, \\ -1 + \frac{2}{3}(x-1)^2 & \text{if } 1 \leq x \leq 2, \\ 1 - \frac{4}{3}e^{-x+2} & \text{if } x \geq 2. \end{cases} \tag{3.3}$$

Clearly, F is monotone and the unique solution of the associated NCP is $2 + \ln \frac{4}{3}$.

The optimization problem associated with the smooth reformulations (without compactification) is to minimize $\Phi(x, z) = (F(x) - z)^2 + (xz)^2$ subject to $x, z \geq 0$. This problem has, besides the global solution, the stationary point $(0, 0)$, which is not a solution of the NCP. Moreover, since $\Phi(k, \frac{1}{k^2}) \leq 1$ for all $k = 3, 4, \ldots$, the level set

$$\{(x, z) \mid \Phi(x, z) \leq 1,\ x \geq 0, z \geq 0\}$$

is not bounded.

4. Penalized Fischer–Burmeister reformulation. The Fischer–Burmeister function, defined by

$$\varphi(a, b) = a + b - \sqrt{a^2 + b^2} \quad \forall\ a, b \in \mathbb{R}, \tag{4.1}$$

has been used in many reformulations of complementarity and variational inequality problems [11, 14, 16, 17, 23, 27, 29, 30, 42]. Its main property is that $\varphi(a, b) = 0$ if and only if $a \geq 0, b \geq 0$, and $ab = 0$.

The penalized Fischer–Burmeister (PFB) function has been introduced recently in [8]. It is defined by

$$\psi_\mu(a, b) = \varphi(a, b) + \mu a_+ b_+, \tag{4.2}$$

where $\mu \geq 0$, $c_+ = \max\{c, 0\}$, and φ is the Fischer–Burmeister function (4.1). Related functions have been proposed in [30, 33].

Based on this function, Chen, Chen, and Kanzow [8] introduced a new method for solving NCPs for which an excellent practical performance has been reported. These authors proved a bounded level set result (if $\mu > 0$) under the condition that F is a monotone function with a strictly feasible point or that F is an R_0-function (see [9]).

Similarly to (3.1), we define the following reformulation of $VIP(G, \mathcal{S})$:

$$\text{Minimize } \Phi_3(x, v, \lambda), \tag{4.3}$$

where

$$\Phi_3(x, v, \lambda) = \rho_0 \|G(x) + \mathbf{1}\lambda - v\|_2^2 + \rho_1 \left(\sum_{i=1}^{m} x_i - 1 \right)^2 + \sum_{i=1}^{m} \psi_\mu(x_i, v_i)^2$$

and ρ_0, ρ_1, and $\mathbf{1}$ are as in (3.1). As in the previously defined reformulations, the objective function of (4.3) vanishes if and only if x is a solution of $VIP(G, \mathcal{S})$.

If G is differentiable, the objective function Φ_3 is once (but not twice) continuously differentiable. Boundedness of the level sets associated with Fischer–Burmeister ($\mu = 0$) reformulations of complementarity problems has been proved in [29] under restrictive conditions on F. Here we will prove bounded level set results that hold assuming only continuity of G.

THEOREM 4.1. *Assume that G is continuous on $\mathbb{R}^m$, $\mu = 0$, and $\theta \in (0, 1/m)$ is such that*

$$\rho_1(\sqrt{\theta m} - 1)^2 > \theta. \tag{4.4}$$

Then, the set

$$L_3 = \{(x, v, \lambda) \in \mathbb{R}^{2m+1} \mid \Phi_3(x, v, \lambda) \leq \theta\} \tag{4.5}$$

is bounded.

Proof. Suppose that $\{(x^k, v^k, \lambda_k)\} \in L_3$ for all $k = 0, 1, 2, \dots$. Let us suppose, by contradiction, that this sequence is not bounded.

Since $\Phi_3(x^k, v^k, \lambda_k) \leq \theta$, we have that

$$\varphi(x_i^k, v_i^k)^2 \leq \theta \quad \forall\ k = 0, 1, 2, \dots.$$

So,

$$-\varphi(x_i^k, v_i^k) \leq \sqrt{\theta} \quad \forall\ k = 0, 1, 2, \dots.$$

By an elementary property of the Fischer–Burmeister function, this implies that

$$x_i^k \geq -\sqrt{\theta} \quad \text{and} \quad v_i^k \geq -\sqrt{\theta} \quad \forall\ k = 0, 1, 2, \dots.$$

So, $x^k \geq (-\sqrt{\theta}, \dots, -\sqrt{\theta})$ and $(\sum_{i=1}^m x_i^k - 1)^2 \leq \theta/\rho_1$ for all $k = 0, 1, 2, \dots$. This implies that $\{x^k\}$ is bounded.

By the continuity of G, $\{G(x^k)\}$ is also bounded. Therefore, since $\{\|G(x^k) + \mathbf{1}\lambda_k - v^k\|_2^2\}$ is obviously bounded and $v_i^k \geq -\sqrt{\theta}$ $\forall\ k = 0, 1, 2, \dots$, the unboundedness of $\{(x^k, v^k, \lambda_k)\}$ implies that there exists a subsequence such that (after relabeling)

$$\lim_{k \to \infty} v_i^k = \infty \quad \forall\ i = 1, \dots, m. \tag{4.6}$$

But the sequence $\{x^k\}$ is bounded, so it has a convergent subsequence. Therefore, we can ensure that for a suitable subsequence (4.6) holds. So, after a new relabeling,

$$\lim_{k \to \infty} x_i^k = a_i \quad \forall\ i = 1, \dots, m. \tag{4.7}$$

By an elementary property of (4.1), (4.6) and (4.7) imply that

$$\lim_{k \to \infty} \varphi(x_i^k, v_i^k) = a_i$$

for some $a_i \geq -\sqrt{\theta}$, $i = 1, \dots, m$. Thus

$$\lim_{k \to \infty} \sum_{i=1}^m \varphi(x_i^k, v_i^k)^2 = \sum_{i=1}^m a_i^2.$$

Since $\sum_{i=1}^m \varphi(x_i^k, v_i^k)^2 \leq \theta$ for all k, this implies that

$$\sum_{i=1}^m a_i^2 \leq \theta.$$

Hence,

$$\sum_{i=1}^{m} a_i \le \sqrt{\theta m}.$$

But, by (4.4), $\sqrt{\theta m} < 1$, so

$$\rho_1 \left(\sum_{i=1}^{m} a_i - 1 \right)^2 \ge \rho_1(\sqrt{\theta m} - 1)^2.$$

Therefore, by (4.4),

$$\rho_1 \left(\sum_{i=1}^{m} a_i - 1 \right)^2 > \theta.$$

This implies that, for k large enough,

$$\rho_1 \left(\sum_{i=1}^{m} x_i^k - 1 \right)^2 > \theta,$$

and, so,

$$\Phi_3(x^k, v^k, \lambda_k) > \theta.$$

This contradicts the fact that $(x^k, v^k, \lambda_k) \in L_3$. ∎

As in the case of Φ_1 and Φ_2, with a suitable choice of ρ_0, we can ensure that

$$\Phi_3(x^0, v^0, \lambda_0) < \theta, \tag{4.8}$$

where θ satisfies (4.4). In fact, we take $x^0 \ge 0$ such that $\sum_{i=1}^m x_i^0 = 1$ and $v^0 = 0$. Then, $\Phi_3(x^0, v^0, \lambda_0) = \rho_0 \|G(x^0) + \mathbf{1}\lambda_0 - v^0\|_2^2$ and condition (4.8) holds if ρ_0 is chosen to be sufficiently small.

Remark. The classical Fischer–Burmeister reformulation of the NCP defined by (3.3) consists of minimizing $\Phi(x) \equiv (\sqrt{x^2 + F(x)^2} - x - F(x))^2$. Since $\Phi(k) \le 1$ for all $k = 3, 4, 5, \ldots$, this function fails to have bounded level sets. Of course, the level sets $F(x) \le \alpha$ are bounded if $\alpha > 0$ is small enough, but it is not possible to predict for which point x^0 the level set $\{x \in \mathbb{R} \mid \Phi(x) \le \Phi(x^0)\}$ is bounded. Of course, a rather trivial way to obtain examples where the level sets of all classical reformulations of the monotone NCP are not bounded is to consider problems with an unbounded solution set. Finally, in the absence of monotonicity, examples of unbounded level sets are easy to obtain for all the classical reformulations.

THEOREM 4.2. *Assume that G is continuous on $\mathbb{R}^m$. If $\mu > 0$, for all $\theta \in (0, \rho_1)$, the set L_3, defined in* (4.5)*, is bounded.*

Proof. Suppose that $(x^k, v^k, \lambda_k) \in L_3$ for all $k = 0, 1, 2, \ldots$. Therefore,

$$\psi_\mu(x_i^k, v_i^k)^2 \le \theta \quad \forall\ i = 1, \ldots, m,\ k = 0, 1, 2, \ldots.$$

So, by (4.2),

$$-\varphi(x_i^k, v_i^k) \le \sqrt{\theta} \quad \forall\ i = 1, \ldots, m,\ k = 0, 1, 2, \ldots.$$

This implies, as in Theorem 4.1, that

$$x_i^k \geq -\sqrt{\theta} \quad \forall\ i = 1, \ldots, m,\ k = 0, 1, 2, \ldots.$$

So, since $\rho_1(\sum_{i=1}^m x_i^k - 1)^2 \leq \theta$ for all $k = 0, 1, 2, \ldots$, we have that $\{x^k\}$ is bounded. By the continuity of G, $\{G(x^k)\}$ is bounded and, so, $\{\lambda_k - v_i^k\}$ is bounded for all $i = 1, \ldots, m$. As in previous boundedness theorems, we only need to prove that the assumption $v_i^k \to \infty$ for all $i = 1, \ldots, m$ leads to a contradiction. In fact, if $v_i^k \to \infty$ for all $i = 1, \ldots, m$ we have, as in Theorem 4.1, that, for a suitable subsequence, $\{x_i^k\}$ is convergent and $\{\varphi(x_i^k, v_i^k)\}$ is bounded. Assume, for a moment, that there exists $i \in \{1, \ldots, m\}$ and $\varepsilon > 0$ such that

$$x_i^k \geq \varepsilon$$

for an infinite set of indices. This implies that $x_i^k v_i^k \to \infty$ and, so, the sequence is not contained in L_3. Therefore,

$$\lim_{k\to\infty} x^k \leq 0.$$

This implies that

$$\lim_{k\to\infty} \rho_1 \left(\sum_{i=1}^m x_i^k - 1 \right)^2 \geq \rho_1.$$

This is impossible, since $(x^k, v^k, \lambda_k) \in L_3$. So, the proof is complete. □

Clearly, an initial estimate that belongs to a bounded level set can be chosen as we did in the smooth reformulations studied in section 3.

The following theorem is a sufficiency result for the reformulation (4.3) that corresponds to Theorem 3.2 and Theorem 3.4 of section 3.

THEOREM 4.3. *If G is monotone and has continuous first derivatives, all the stationary points of* (4.3) *are solutions of* $VIP(G, \mathcal{S})$.

Proof. The case $\mu = 0$ follows from a straightforward generalization of Theorem 2.4 of [28]. In the case $\mu > 0$, use Proposition 3.3 of [8] to generalize Theorem 2.4 of [28]. Then, generalize this result as in the case $\mu = 0$. □

5. Preliminary numerical experience. We solved some VIPs on the simplex $\mathcal{S}$ using the reformulations studied in this paper. Our objective here is to get a preliminary idea of the comparative behavior of different reformulations. The first problem considered was

$$\langle G(x), z - x \rangle \geq 0 \quad \forall\ z \in \mathcal{S},$$

where $G : \mathbb{R}^m \to \mathbb{R}^m$ was given by $G(x) = Ax - c$, A was the 10×10 Hilbert matrix ($[A]_{i,j} = \frac{1}{i+j-1}$), and the entries of c_i were chosen randomly in $[0, 2]$. In Table 5.1 we recall the different reformulations studied in this paper.

To solve the optimization problems associated with different reformulations, we used the general purpose algorithm SPG given in [6]. This is a very simple algorithm that generally outperforms conjugate gradient methods in the unconstrained case (see [40]) and is comparable to good large-scale bound-constrained solvers when simple constraints are present. Of course, this algorithm does not take into account the structure of the problems at all and, so, can be very inefficient in many cases, but it

TABLE 5.1
Reformulations and optimization problems.

Reformulation	Objective function	Complementarity term	Feasible region
Smooth 1	Φ_1	$\langle x, v\rangle^2$	$\mathbb{R}^m_+ \times \mathbb{R}^m_+ \times \mathbb{R}$
Smooth 2	$\Phi 2$	$\sum_{i=1}^m (x_i v_i)^2$	$\mathbb{R}^m_+ \times \mathbb{R}^m_+ \times \mathbb{R}$
PFB	Φ_3	$\sum_{i=1}^m \psi_\mu(x_i, v_i)^2$	$\mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$

TABLE 5.2
Comparison of reformulations.

Reformulation	Successful executions	Best in
Smooth 1 (3.1)	4	1 problem
Smooth 2 (3.2)	3	0 problems
PFB (4.3), $\mu = 0$	7	3 problems
PFB (4.3), $\mu = 0.1$	6	3 problems
PFB (4.3), $\mu = 1.0$	8	1 problem
PFB (4.3), $\mu = 10.$	4	0 problems

is useful when the goal is to compare reformulations as we do in this case. In this first set of experiments we used a modest computer environment (Pentium with 90 MHZ) and the code was written in (double precision) Fortran 77.

The convergence criterion used to terminate the execution of SPG was

$$\|P(z - \nabla\Phi_i(z)) - z\| \le 10^{-6},$$

where $z = (x, v, \lambda)$ and P is the projection on the feasible region. As an initial approximation we took x_i = uniformly random between 0 and 1 and then divided each coordinate by $\sum_{i=1}^m x_i$. We also took $v = 0$ and $\lambda = 0$. In order to ensure bounded level sets we chose $\rho_0 = 1$ and $\rho_1 = \max\{1, 1.1\|G(x^0)\|_2^2\}$.

We solved 10 problems with different random generations of c and the initial x. We considered that the execution was successful if the solution was obtained in less than 25 seconds. In general, successful executions used less than 5 seconds for all the formulations. In Table 5.2, we show the number of successful executions and the number of times each reformulation was the best, in terms of execution time. In all the successful cases, the solutions were obtained with the same precision. In two problems, all the reformulations failed. We considered that there was not a "best reformulation" in these two cases.

Both in the condensed Table 5.2, and looking in detail at the experiments, the behavior of "Smooth 2" appears to be similar to PFB with $\mu = 10$. This is not surprising, since a large μ in $\psi_\mu(a, b)$ gives more weight to the multiplicative term ab in the positive orthant and "Smooth 2" only uses this term.

The penalty parameter μ in the function $\psi_\mu(a, b)$ affects the measure of "lack of complementarity" in the positive orthant ($a \ge 0, b \ge 0$) in the following way: If $\mu \approx 0$, then $\psi_\mu(a, b) \approx \varphi(a, b)$ and, so, $\psi_\mu(\varepsilon, M)$ is "approximately independent" of M if $\varepsilon > 0$ is small and M is large. This comes from $\lim_{M\to\infty} \varphi(\varepsilon, M) = \varepsilon$. In other words, $\varphi(\varepsilon, M) \approx \min\{\varepsilon, M\}$. On the other hand, if μ is large or if we are using "Smooth 2," the measure of lack of complementarity tends quickly to ∞ if one of the variables tends to infinity and the other is kept fixed. Whether it is better to consider that (ε, M) is almost complementary or not is a problem-dependent question. However, at the beginning of iterative processes, it is dubious that the variable that corresponds to the smaller complementary variable will be zero at the solution and, so, it seems convenient to try to reduce both. This decision corresponds to "μ large" in the PFB reformulation.

The "Smooth 1" reformulation is "more global" in the sense that the influence of the lack of complementarity of the pair (x_j, v_j) depends on the lack of complementarity of the other pairs. In fact, since

$$\frac{\partial}{\partial v_j}\left(\sum_{i=1}^{m} x_i v_i\right)^2 = 2\left(\sum_{i=1}^{m} x_i v_i\right) x_j \quad \text{and} \quad \frac{\partial}{\partial x_j}\left(\sum_{i=1}^{m} x_i v_i\right)^2 = 2\left(\sum_{i=1}^{m} x_i v_i\right) v_j,$$

the contribution of the jth lack of complementarity to the gradient of the objective function grows with the deviation from complementarity of the remaining pairs. In other words, "Smooth 1" will try a large step toward zero on the variable v_j not only when x_j and $x_j v_j$ are large but also when some of the products $x_i v_i$ (for $i \neq j$) are large.

The small number of experiments described above encouraged us to define a Newton-type algorithm that uses Φ_3 as the objective function, with the aim of comparing more systematically different choices of μ. Observe that finding a zero value of Φ_3 is equivalent to solving the $(2m+1) \times (2m+1)$ nonlinear system

$$H(z) = 0,$$

where $z = (x, v, \lambda)$ and

$$H(z) = \left(G(x) - v + \lambda \mathbf{1},\ \sqrt{\rho_1}\left(\sum_{i=1}^{m} x_i - 1\right),\ \psi_\mu(x_1, v_1), \ldots, \psi_\mu(x_m, v_m)\right).$$

If G is smooth, H is smooth except when $x_i v_i = 0$ for some i. (However, Φ_3 is smooth for all z.) Therefore, the Newtonian direction

$$d(z) = -B(z)^{-1} H(z), \tag{5.1}$$

where $B(z) \in \partial_B G(z)$, is well defined whenever a nonsingular element of $\partial_B G(z)$ can be found. See [39]. This allows us to define a nonmonotone safeguarded Newton-gradient algorithm along the lines of [11, 25]. From now on, we write $\Phi(z) = \Phi_3(x, v, \lambda)$ for the sake of simplicity.

Assume that $\gamma \in (0,1), \beta_1, \beta_2 > 0, \alpha \in (0, 1/2), \nu \in \{0, 1, 2, \ldots\}$ are given independently of k. Suppose that the iterate z^k has been computed for some $k \geq 0$. Then, if $\nabla\Phi(z^k) \neq 0$, the iterate z^{k+1} is computed as follows.

ALGORITHM 5.1 (nonmonotone safeguarded Newton-gradient).

Step 1. *If $d(z^k)$ (given by* (5.1)*) exists and, in addition,*

$$\langle d(z^k), \nabla\Phi(z^k)\rangle \leq -\gamma \|d(z^k)\|_2 \|\nabla\Phi(z^k)\|_2 \tag{5.2}$$

and

$$\beta_1 \|\nabla\Phi(z^k)\|_2 \leq \|d(z^k)\|_2 \leq \beta_2 \|\nabla\Phi(z^k)\|_2, \tag{5.3}$$

define $d^k = d(z^k)$. Otherwise, define $d^k = -\nabla\Phi(z^k)$.

Step 2. *Starting with $t = 1$ and using classical safeguarded backtracking (see* [11, 12]*), compute $t_k > 0$ such that*

$$\Phi(z^k + t_k d^k) \le \tilde{\Phi}_k + \alpha t_k \langle d^k, \nabla\Phi(z^k)\rangle, \tag{5.4}$$

where

$$\tilde{\Phi}_k = \max\{\Phi(z^k), \ldots, \Phi(z^\tau)\}$$

and $\tau = \max\{0, k - \nu\}$ *(see* [25]*).*

Define $z^{k+1} = z^k + t_k d^k$. Using slight modifications of the results of [11] and [25] we can prove that every limit point of a sequence generated by this algorithm is stationary. Since we have proved that, choosing the appropriate initial point and ρ_1, the generated sequences are bounded, it turns out that stationary points are necessarily found, in the limit, by Algorithm 5.1.

We wrote a double precision Fortran code implementing this algorithm for the unconstrained minimization of Φ_3. We chose $\gamma = \beta_1 = 1/\beta_2 = \alpha = 10^{-4}$ and $\nu = 9$. As initial point we took $x^0 = (1, 1/2, \ldots, 1/m)/\sum_{i=1}^{m}(1/i)$, $v^0 = 0$, $\lambda_0 = 0$. We ran the algorithm for different choices of μ using problems defined by operators $G(x)$ taken from the nonlinear-system literature. Namely, we define the following problems.

Problem 1 (Hilbert). $m = 100$.

$G(x) = Ax - c$, where A is defined as the Hilbert matrix and $c = (1, 1/2, \ldots, 1/m)$.

Problem 2 (Broyden). $m = 100$.

$$[G(x)]_1 = (3 - 2x_1)x_1 - 2x_2 + 1,$$

$$[G(x)]_i = (3 - 2x_i)x_i - x_{i-1} - 2x_{i+1} + 1, \quad i = 2, \ldots, m-1,$$

$$[G(x)]_m = (3 - 2x_m)x_m - x_{m-1}.$$

Problem 3 (Rosenbrock). $m = 20$.

$$[G(x)]_i = 10(x_{i+1} - x_i^2) \quad \text{if } i \text{ is odd},$$

$$[G(x)]_i = 1 - x_{i-1} \quad \text{if } i \text{ is even}.$$

Problem 4 (Helical valley). $m = 99$.

For $i = 1, \ldots, m/3$,

$$[G(x)]_{3i} = 10x_{3i+2} - \frac{50}{\pi} \text{ atan } (x_{3i+1}/3i) - 50 \quad \text{if } x_{3i} < 0,$$

$$[G(x)]_{3i} = 10x_{3i+2} - \frac{50}{\pi} \text{ atan } (x_{3i+1}/3i) \quad \text{if } x_{3i} > 0,$$

$$[G(x)]_{3i+1} = \sqrt{x_{3i}^2 + x_{3i+1}^2},$$

and

$$[G(x)]_{3i+2} = x_{3i+2}.$$

Problem 5 (Watson). $m = 31$.

For $i = 1, \ldots, 29$,

TABLE 5.3
Comparison of different penalty parameters in PFB.

Problem	m	μ	$\|H(z)\|_\infty$	It	FE	Changes	Time
Hilbert	100	10^{-6}	$0.95E-10$	10	64	0	2.16
		0.10	$0.22E-10$	13	74	0	2.81
		1.00	$0.14E-09$	12	84	0	2.56
		10.00	$0.17E-11$	11	69	0	2.32
		100.0	$0.34E-10$	13	92	0	2.78
Broyden	100	10^{-6}	$0.42E-09$	4	5	0	0.82
		0.10	$0.42E-09$	4	5	0	0.82
		1.00	$0.42E-09$	4	5	0	0.82
		10.00	$0.42E-09$	4	5	0	0.82
		100.0	$0.42E-09$	4	5	0	0.82
Rosenbrock	20	10^{-6}	$0.48E-09$	3	4	0	< 0.1
		0.10	$0.48E-09$	3	4	0	< 0.1
		1.00	$0.48E-09$	3	4	0	< 0.1
		10.00	$0.48E-09$	3	4	0	< 0.1
		100.0	$0.48E-09$	3	4	0	< 0.1
Helical	99	10^{-6}	$0.36E-14$	122	2021	104	25.7
		0.10	$0.36E-14$	122	2025	105	26.3
		1.00	$0.13E-11$	122	1995	104	26.2
		10.00	$0.28E-11$	62	830	45	13.0
		100.0	$0.40E-09$	153	2594	134	32.6
Watson	31	10^{-6}	$0.10E-09$	35	197	1	0.1
		0.10	$0.10E-09$	35	197	1	0.1
		1.00	$0.10E-09$	35	197	1	0.1
		10.00	$0.10E-09$	35	197	1	0.1
		100.0	$0.10E-09$	35	197	1	0.1
Murty	100	10^{-6}	$0.44E-15$	176	561	1	25.6
		0.10	$0.77E-10$	150	397	1	31.3
		1.00	$0.26E-09$	173	491	1	35.9
		10.00	$0.80E-12$	166	511	1	34.4
		100.0	$0.16E-11$	165	493	1	34.2

$$[G(x)]_i = \sum_{j=1}^{m}(j-1)x_j(i/29)^{j-2} - \left[\sum_{j=1}^{m} x_j(i/29)(i/29)^{j-2}\right]^2 - 1,$$

$$[G(x)]_{30} = x_1,$$

$$[G(x)]_{31} = x_2 - x_1^2 - 1.$$

Problem 6 (Murty). $m = 100$.

$G(x) = Ax - c$, where A is upper-triangular, $[A]_{ij} = 2$ if $i < j$, $[A]_{ii} = 1$ for all $i = 1, \ldots, m$, and $c = (1, \ldots, 1)$.

The experiments that we report below were run in a SPARCstation Sun Ultra 1, with an UltraSPARC 64 bits processor, 167-MHz clock, and 128-MBytes of RAM memory. The stopping criterion was $\|H(z)\|_\infty \leq 10^{-8}$. Besides number of iterations (It), number of function evaluations (FE), and CPU time (in seconds), we report in Table 5.3 the number of times the Newton direction needed to be replaced by the gradient direction. In this preliminary implementation, the linear systems were solved by Gaussian elimination, without taking advantage of their sparsity. Obviously, the computer time must decrease dramatically if a sparse implementation is developed, but the other indicators would not change.

We observe that in three problems (Broyden, Rosenbrock, and Watson) the behavior of the five penalty parameters is the same. In Hilbert and Murty the smallest

TABLE 5.4
Algorithm 5.1 *with Problem Hilbert,* $\mu = 0.1$.

Iteration k	Evaluations	$\|H(z^k)\|_2$
0	1	1.1659909612460
1	7	1.1506050628261
2	19	1.1559111113737
3	31	1.1586923058790
4	39	1.1556372319450
5	46	1.1414773334813
6	51	1.1299367786945
7	56	1.1386723439549
8	62	1.1617009941393
9	66	1.0730683927287
10	70	0.98505951078780
11	72	0.67329707514733
12	73	$4.2472036158979E-03$
13	74	$2.6507263515338E-11$

μ was, marginally, the best. However, in Helical, "$\mu = 10$" clearly outperformed the other alternatives. Probably, very large values of μ should be discarded from practical implementations (at least in well-scaled problems), but the best choice among "small" values of μ seems to depend strongly on the problem characteristics.

The number of functional evaluations per iteration appears to be large in the problems Hilbert, Watson, and Murty. With the aim of understanding this phenomenon, we ran some problems choosing the gradient direction in the first (5 or 10) iterations. Running Hilbert with $\mu = 0.1$ and 5 (first) gradient iterations, the computer time decreased to 1.6 seconds and, with 10 (first) gradient iterations, to 1.1 seconds. We found it instructive to show the detailed behavior of the algorithm in the ordinary case and in the two modified cases. See Tables 5.4, 5.5, and 5.6. We observe that, in fact, the first Newton iterations are not worthwhile in terms of the progress they provide, whereas, of course, they are much more expensive than gradient iterations. The quadratic convergence of Newton is quite evident in the last two iterations. We also ran the algorithm using only gradient iterations, and we observed, as expected, an extremely slow convergence behavior. In fact, convergence did not occur after 1000 iterations in this case.

The qualitative behavior described for Hilbert is essentially the same in the Watson problem. In this case, with 10 initial gradient iterations, the computer time reduced to 0.18 seconds and even the number of iterations decreased. On the other hand, the modification of the algorithm in the Murty problem did not cause meaningful improvements. In this problem, the number of iterations increased moderately and the computer time remained more or less the same.

Problem Helical is instructive in a different sense. In this case, the Newton direction was rejected at most iterations and the algorithm behaves, essentially, as a steepest descent method. We decided to modify the algorithmic parameters in order to weaken the criterion of acceptance of the Newton direction at Step 1 of Algorithm 5.1. Consequently, we chose $\gamma = \beta_1 = 1/\beta_2 = 10^{-25}$ and ran the problem with these new parameters. The results were quite impressive, showing how sensitive this type of algorithm can be with respect to safeguarding constants. For $\mu = 0.1$ the Newton direction was never rejected, convergence occurred in 10 iterations with 18 function evaluations and 1.9 seconds of CPU time. Similar improvements were obtained for the other values of μ.

TABLE 5.5
Algorithm 5.1 *with Problem Hilbert,* $\mu = 0.1$ *First* 5 *are gradient iterations.*

Iteration k	Evaluations	$\Vert H(z^k)\Vert_2$
0	1	1.1659909612460
1	9	1.0818877741731
2	17	1.0192301426184
3	24	1.0804586556667
4	32	0.98801773901278
5	40	0.91336945465545
6	41	0.46240603597402
7	45	0.78454888864694
8	46	0.86704125514515
9	47	0.93591242869472
10	54	1.0832676503144
11	55	$1.2486098348917E-06$
12	56	$9.6170809859694E-14$

TABLE 5.6
Algorithm 5.1 *with Problem Hilbert,* $\mu = 0.1$. *First* 10 *are gradient iterations.*

Iteration k	Evaluations	$\Vert H(z^k)\Vert_2$
0	1	1.1659909612460
1	9	1.0818877741731
2	17	1.0192301426184
3	24	1.0804586556667
4	32	0.98801773901278
5	40	0.91336945465545
6	48	0.85190911080106
7	55	1.1099440799153
8	63	0.97638264851556
9	71	0.87222456838420
10	79	0.78914420687610
11	80	0.29364275115164
12	81	0.95227479134610
13	82	$1.0932851184041E-02$
14	83	$1.9355846720355E-09$

6. Final remarks. We believe that the results presented in this paper have a reasonably wide scope of applications. Consider the general variational inequality problem defined by F_1 on Ω, where F_1 is smooth,

$$\Omega = \{x \in \mathbb{R}^q \ \mid \ g(x) \leq 0\},$$

$g = (g_1, \ldots, g_p)$, and g_i smooth and convex for all $i = 1, \ldots, p$. Under a suitable constraint qualification [26], this problem is equivalent to

$$F_1(x) + \sum_{i=1}^{p} w_i \nabla g_i(x) = 0,$$

$$w \geq 0, \qquad g(x) \leq 0,$$

$$\sum_{i=1}^{p} g_i(x) w_i = 0.$$

Defining $n = p+q$, $z = (x, w)$, $F(z) = (F_1(x) + \sum_{i=1}^{p} w_i \nabla g_i(x), -g(x))$, and $I = \{p+1, \ldots, p+q\}$ we obtain a problem of type $VIP(F, \Omega_1)$ (2.3). So, after compactification, we obtain the VIP on the simplex.

In this research we proved that, using several potentially useful reformulations, the boundedness of the sequences generated by standard algorithms can be guaranteed, so that limit points exist and sufficiency results can be applied.

Sufficiency results of the type "stationarity implies solution" usually depend on "monotonicity-like" assumptions. However, one should not interpret that the reformulations must be tried *only when* the monotonicity assumption is guaranteed to hold. Optimization algorithms usually guarantee stationary points, but their practical efficiency is linked to their ability to find global minimizers in a substantial number of cases. This means that we can try to solve the reformulation in any situation, with the hope that using good global strategies we will probably find solutions of the original problem.

In [43, 44], Solodov and Svaiter presented Newton-like methods for solving monotone nonlinear systems and monotone NCPs, respectively. Their convergence results are very strong but, on the other hand, the monotonicity assumption seems to be more essential for their algorithms than it is for the different reformulations presented here. The conditions under which specific algorithms for reformulations enjoy the "true" convergence properties of [43, 44] should be investigated.

Acknowledgments. We are indebted to two anonymous referees, whose comments helped us to improve the paper.

REFERENCES

[1] R. Andreani, A. Friedlander, and J. M. Martínez, *On the solution of finite-dimensional variational inequalities using smooth optimization with simple bounds*, J. Optim. Theory Appl., 94 (1997), pp. 635–657.

[2] R. Andreani and J. M. Martínez, *Solving complementarity problems by means of a new smooth constrained nonlinear solver*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, pp. 1–24.

[3] R. Andreani and J. M. Martínez, *On the solution of the extended linear complementarity problem*, Linear Algebra Appl., 281 (1998), pp. 247–257.

[4] R. Andreani and J. M. Martínez, *On the reformulation of nonlinear complementarity problems using the Fischer–Burmeister function*, Appl. Math. Lett., 12 (1999), pp. 7–12.

[5] R. Andreani and J. M. Martínez, *On the solution of bounded and unbounded mixed complementarity problems*, Optimization, to appear.

[6] E. G. Birgin, J. M. Martínez, and M. Raydan, *Nonmonotone spectral projected gradient methods on convex sets*, SIAM J. Optim., to appear.

[7] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, *A limited memory algorithm for bound constrained optimization*, SIAM J. Sci. Comput., 16 (1995), pp. 1190–1208.

[8] B. Chen, X. Chen, and C. Kanzow, *A penalized Fischer–Burmeister NCP-function: Theoretical investigation and numerical results*, Math. Program., to appear.

[9] B. Chen and P. T. Harker, *Smooth approximations to nonlinear complementarity problems*, SIAM J. Optim., 7 (1997), pp. 403–420.

[10] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Global convergence of a class of trust region algorithms for optimization with simple bounds*, SIAM J. Numer. Anal., 25 (1988), pp. 433–460. See also SIAM J. Numer. Anal., 26 (1989) pp. 764–767.

[11] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problem*, Math. Programming, 75 (1996), pp. 407–439.

[12] J. E. Dennis, Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall Ser. Comput. Math., Prentice-Hall, Englewood Cliffs, NJ, 1983.

[13] S. P. Dirkse and M. C. Ferris, *A collection of nonlinear mixed complementarity problems*, Optim. Methods Softw., 5 (1995), pp. 319–345.

[14] F. Facchinei, A. Fischer, and C. Kanzow, *A semismooth Newton method for variational inequalities: The case of box constraints*, in Complementarity and Variational Problems:

State of the Art, M. C. Ferris and J.-S. Pang, eds., SIAM, Philadelphia, PA, 1997, pp. 76–90.

[15] L. Fernandes, A. Friedlander, M. Guedese, and J. Júdice, *Solution of generalized linear complementarity problems using smooth optimization and its application to bilinear programming and LCP*, Appl. Math. Optim., to appear.

[16] M. C. Ferris and C. Kanzow, *Complementarity and Related Problems: A Survey*, Tech. Report, Comput. Sci. Dept., University of Wisconsin, Madison, WI, 1998.

[17] M. C. Ferris, C. Kanzow, and T. S. Munson, *Feasible descent algorithms for mixed complementarity problems*, Math. Program., 86 (1999), pp. 475–497.

[18] A. Friedlander and J. M. Martínez, *On the numerical solution of bound constrained optimization problems*, RAIRO Rech. Opér., 23 (1989), pp. 319–341.

[19] A. Friedlander, J. M. Martínez, and S. A. Santos, *A new trust region algorithm for bound constrained minimization*, Appl. Math. Optim., 30 (1994), pp. 235–266.

[20] A. Friedlander, J. M. Martínez, and S. A. Santos, *On the resolution of linearly constrained convex minimization problems*, SIAM J. Optim., 4 (1994), pp. 331–339.

[21] A. Friedlander, J. M. Martínez, and S. A. Santos, *Solution of linear complementarity problems using minimization with simple bounds*, J. Global Optim., 6 (1995), pp. 253–267.

[22] A. Friedlander, J. M. Martínez, and S. A. Santos, *A new strategy for solving variational inequalities on bounded polytopes*, Numer. Funct. Anal. Optim., 16 (1995), pp. 653–668.

[23] C. Geiger and C. Kanzow, *On the resolution of monotone complementarity problems*, Comput. Optim. Appl., 5 (1996), pp. 155–173.

[24] M. S. Gowda, *On the extended linear complementarity problem*, Math. Programming, 72 (1996), pp. 33–50.

[25] L. Grippo, F. Lampariello, and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal., 23 (1986), pp. 707–716.

[26] P. T. Harker and J. S. Pang, *Finite-dimensional variational inequality and nonlinear complementarity problem: A survey of theory, algorithms and applications*, Math. Programming, 48 (1990), pp. 161–220.

[27] H. Jiang and L. Qi, *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM J. Control Optim., 35 (1997), pp. 178–193.

[28] C. Kanzow, *An unconstrained optimization technique for large-scale linearly constrained convex minimization problems*, Computing, 53 (1994), pp. 101–117.

[29] C. Kanzow, *A new approach to continuation methods for complementarity problems with uniform P-functions*, Oper. Res. Lett., 20 (1997), pp. 85–92.

[30] C. Kanzow, N. Yamashita, and M. Fukushima, *New NCP-functions and properties*, J. Optim. Theory Appl., 94 (1997), pp. 115–135.

[31] D. N. Kozakevich, J. M. Martínez, and S. A. Santos, *Solving nonlinear systems of equations with simple constraints*, Comput. Appl. Math., 16 (1997), pp. 215–235.

[32] J. J. Júdice, *Algorithms for linear complementarity problems*, in Algorithms for Continuous Optimization, E. Spedicato, ed., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994, pp. 435–474.

[33] Z.-Q. Luo and P. Tseng, *A new class of merit functions for the nonlinear complementarity problem*, in Complementarity and Variational Problems: State of the Art, M. C. Ferris and J.-S. Pang, eds., SIAM, Philadelphia, PA, 1997, pp. 204–225.

[34] O. L. Mangasarian and J. S. Pang, *The extended linear complementarity problem*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 359–368.

[35] O. L. Mangasarian and M. V. Solodov, *Nonlinear complementarity problem as unconstrained and constrained minimization*, Math. Programming, 62 (1993), pp. 277–297.

[36] J. M. Martínez, *Quasi-inexact-Newton methods with global convergence for solving constrained nonlinear systems*, Nonlinear Anal., 30 (1997), pp. 1–8.

[37] J. J. Moré, *Global methods for nonlinear complementarity problems*, Math. Oper. Res., 21 (1996), pp. 589–614.

[38] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

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

[40] M. Raydan, *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem*, SIAM J. Optim., 7 (1997), pp. 26–33.

[41] H. Schwetlick, *Numerische Lōsung nichtlinearer Gleinchungen*, Deutscher Verlag der Wissenschaften, Berlin, 1978.

[42] M. V. Solodov, *Some optimization reformulations of the extended linear complementarity problem*, Comput. Optim. Appl., 13 (1999), pp. 187–200.

[43] M. V. Solodov and B. F. Svaiter, *A globally convergent inexact Newton method for systems of monotone equations*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, pp. 355–369.

[44] M. V. Solodov and B. F. Svaiter, *A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem*, SIAM J. Optim., 10 (2000), pp. 605–625.

[45] M. V. Solodov and P. Tseng, *Two methods based on the D-gap function for solving monotone variational inequalities*, Comput. Optim. Appl., to appear.