

M.V. Solodov · B.F. Svaiter

Error bounds for proximal point subproblems and associated inexact proximal point algorithms

Received: April 28, 1999 / Accepted: March 24, 2000
Published online July 20, 2000 – © Springer-Verlag 2000

Abstract. We study various error measures for approximate solution of proximal point regularizations of the variational inequality problem, and of the closely related problem of finding a zero of a maximal monotone operator. A new merit function is proposed for proximal point subproblems associated with the latter. This merit function is based on Burachik-Iusem-Svaiter's concept of ε -enlargement of a maximal monotone operator. For variational inequalities, we establish a precise relationship between the regularized gap function, which is a natural error measure in this context, and our new merit function. Some error bounds are derived using both merit functions for the corresponding formulations of the proximal subproblem. We further use the regularized gap function to devise a new inexact proximal point algorithm for solving monotone variational inequalities. This inexact proximal point method preserves all the desirable global and local convergence properties of the classical exact/inexact method, while providing a constructive error tolerance criterion, suitable for further practical applications. The use of other tolerance rules is also discussed.

Key words. variational inequalities – proximal point subproblems – error bound – (enlargement of) maximal monotone operator – regularized gap function

1. Introduction

One of the principal applications of error bounds in mathematical programming [24], is their use for constructing computational algorithms, and analyzing their convergence. This paper is devoted to the study of a new merit function for proximal point problems and associated error bounds. It also discusses the relation between this merit function and the regularized gap function [1, 13] in the context of variational inequalities. Finally, these theoretical results are applied to devise inexact proximal-based algorithms for the variational inequality problem which employ *constructive* approximation criteria.

Given a function $F : \mathcal{H} \mapsto \mathcal{H}$, where \mathcal{H} is a real Hilbert space, and a set C in \mathcal{H} , the associated variational inequality problem, $\text{VIP}(F, C)$ for short, is to find a point $x \in \mathcal{H}$ such that

$$x \in C, \quad \langle F(x), y - x \rangle \geq 0 \quad \forall y \in C, \quad (1)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in \mathcal{H} . Throughout this paper, we shall assume that the set C is closed and convex, and F is (maximal) monotone, i.e., $\langle F(x) - F(y), x - y \rangle \geq 0$ for all $x, y \in \mathcal{H}$. A problem closely related to $\text{VIP}(F, C)$ is that of finding a zero of a maximal monotone operator T : find an $x \in \mathcal{H}$ such that

$$0 \in T(x). \quad (2)$$

M.V. Solodov, B.F. Svaiter: Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil, e-mail: {solodov,benar}@impa.br

Mathematics Subject Classification (1991): 49M45, 90C25, 90C33

Recall that maximality means that the graph $G(T) = \{(x, v) \in \mathcal{H} \times \mathcal{H} \mid v \in T(x)\}$ is not properly contained in the graph of any other monotone operator. Clearly, (2) can be related to (1) by the choice of $C = \mathcal{H}$, $F = T$. Conversely, (1) can be reformulated as the inclusion problem (2) if we take $T = F + N_C$, where N_C is the normal operator for the set C . This operator T is monotone, and it is also maximal because C , the domain of N_C , intersects the interior of the domain of F (which we assume here to be \mathcal{H}) [26].

In this setting, the classical regularization technique is the proximal point method (see [27, 28, 25, 22, 20, 15, 12]; and [19] for a survey). This method consists of solving a sequence of regularized subproblems. Specifically, if applied to (1), the method consists of resolving subproblems of the form

$$\text{VIP}(F_k, C), \quad F_k(x) := \lambda_k F(x) + (x - x^k), \quad (3)$$

where x^k is the current approximation to a solution, and λ_k is a positive regularization parameter; and if applied to (2), subproblems are

$$0 \in T_k(x), \quad T_k(x) := \lambda_k T(x) + (x - x^k). \quad (4)$$

The devise of inexact proximal point methods, where subproblems are solved only approximately, is of special importance for obtaining *realistic* frameworks which can further serve as the basis for implementable/practical proximal-based computational techniques. As we shall see, the theory and methodology of *error bounds* [24] will be very helpful in this regard. We shall study inexact solutions for both formulations of proximal subproblems (i.e., (3) and (4)), paying a special attention to how the two are connected in the inexact framework.

We note that convergence of most inexact proximal methods in the literature requires *a priori* summability of certain tolerance parameters or error terms (e.g., [28, 2, 39, 8, 11, 9]; see also Sect. 2 below). This requirement is known to be not very constructive. Recently, methods with improved *constructive* error tolerance criteria have been introduced in [37, 36]. The key advantage of these criteria is that the condition of *a priori* summability is removed, and the tolerance parameters are required to be merely bounded away from one (as opposed to summable!) which gives, in principle, a more realistic computational model. To force convergence, an inexact proximal step is supplemented (followed) by a readily available projection step in [37], or extragradient step in [36]. Furthermore, to guarantee local (super)linear convergence, one does not have to change the error tolerance criterion used for global convergence. The relaxed constructive tolerance rules give a more flexible proximal framework which can be used, for example, for constructing Newton-type methods with improved global convergence properties [35, 33]. Another interesting application is the modified forward-backward splitting method of Tseng [40] (developed independently), which can be shown to belong to the framework of [36].

The methods in [37, 36] deal with the problem (2) of finding a zero of an operator, rather than $\text{VIP}(F, C)$ which is our problem here. Even though the two problems can be regarded as equivalent, it often appears advantageous to deal with $\text{VIP}(F, C)$ directly, i.e., to take full advantage of the special structure of the operator T (which is $F + N_C$). So in this paper, we propose an approach based on approximate proximal iterations which is designed specifically to solve the variational inequality problem. The approximation

criteria are managed by the regularized gap function [1, 13], which is a natural measure of the error in this setting. To facilitate this development, we first study the relationship between inexact solutions of proximal subproblems in the form (3) and in the form (4), as well as the associated merit functions and error bounds. This study is interesting by itself, but it also would allow us to adapt the convergence analysis in the general framework of [36] to the proposed new algorithm.

The rest of the paper is organized as follows. In Sect. 2, we formally introduce merit functions for various forms of proximal subproblems, and derive some error bounds based on them. We also establish the relationships between different error measures. In Sect. 3, we show how the introduced error measures can be used in the design of inexact proximal point methods. Furthermore, a new inexact proximal point method with a constructive error tolerance criterion based on the regularized gap function is proposed. Convergence results are also given in this section. Some concluding remarks can be found in Sect. 4.

2. Merit functions and error bounds for proximal subproblems

We start with the problem (2) of finding a zero of T (for now, an arbitrary maximal monotone operator). The associated proximal point problem at $x \in \mathcal{H}$ with some regularization parameter $\lambda > 0$, is to find $y \in \mathcal{H}$ such that

$$0 \in \lambda T(y) + y - x. \tag{5}$$

It is convenient to state this problem in the form of a “proximal system”: find $y, v \in \mathcal{H}$ such that

$$v \in T(y), \tag{6}$$

$$\lambda v + y - x = 0. \tag{7}$$

To handle approximate solutions, equation (7) is typically relaxed to

$$\|\lambda v + y - x\| \leq \sigma \quad \text{or} \quad \|\lambda v + y - x\| \leq \sigma \|y - x\|, \tag{8}$$

where $\sigma > 0$ is an error tolerance parameter. For convergence of most proximal-based algorithms, parameters σ have to be *summable* (and hence, tend to zero) over the course of iterations, e.g. [28, 2, 39, 8, 11, 9]. One drawback in this regard is that it is often not clear how to choose the sequence of σ ’s in a constructive way to ensure this summability condition. Recently, methods with constructive (and less restrictive) tolerance rules have been developed in [37, 36]. In particular, in [37] (8) was replaced by

$$\|\lambda v + y - x\| \leq \sigma \max\{\lambda \|v\|, \|y - x\|\},$$

where the tolerance parameters σ have to be merely uniformly less than one for convergence (i.e., they need not be summable, or tend to zero; they can even be fixed). Furthermore, in [36], the inclusion (6) was also relaxed, as we shall discuss next, to the

condition $v \in T^\varepsilon(y)$, where $T^\varepsilon(y)$ is a certain outer approximation of $T(y)$. Specifically, given $\varepsilon \geq 0$, the ε -enlargement of T is defined by

$$T^\varepsilon(x) = \{v \in \mathcal{H} \mid \forall y \in \mathcal{H}, u \in T(y), \langle v - u, x - y \rangle \geq -\varepsilon\}. \tag{9}$$

This enlargement of T has been introduced by Burachik, Iusem and Svaiter in [4]; its properties and some applications can be found in [4, 6, 36, 7, 5]. It is easy to observe that

$$\begin{aligned} T^0 &= T, \\ T(x) &\subseteq T^\varepsilon(x), \quad \forall \varepsilon \geq 0, x \in \mathcal{H}. \end{aligned}$$

These basic properties are immediate consequences of the definition (9), and of the maximal monotonicity of T . Thus T^ε can be regarded as an outer approximation of T .

An approximate solution of (6)–(7), as introduced in [36], is a pair (y, v) such that

$$\begin{aligned} v &\in T^\varepsilon(y), \\ \|\lambda v + y - x\| &\text{ and } \varepsilon \text{ are "small"}. \end{aligned} \tag{10}$$

This notion, and its use in a convergent proximal-based algorithm, would be made precise later. At first glance, the introduction of T^ε instead of T may seem in some sense artificial. However, it is not so for several reasons. First of all, the relations $v \in T(y)$, $\lambda v + y - x = \delta \neq 0$ do *not* imply the existence of some $\varepsilon \geq 0$ such that $0 \in \lambda T^\varepsilon(y) + y - x$, no matter how small is δ or how large ε is allowed to be (see an example in [36]). Thus (10) is a meaningful extension of the classical notion of inexact solution of the proximal point subproblems. Second, as we shall see, T^ε arises naturally when one considers inexact solutions of regularized VIP(F_k, C). Specifically, even if exact values of F are used in the latter, approximate resolution of VIP(F_k, C) gives rise to certain elements in $T_k^\varepsilon = (F_k + N_C)^\varepsilon$. The precise characterization of this relationship is one of the subjects of this paper.

We next proceed to formally study approximate solutions in the sense of (10), the associated error measure (merit function), and the resulting error bounds. Let us define a function $\varepsilon_T : \mathcal{H} \times \mathcal{H} \mapsto \Re \cup \{+\infty\}$ by

$$\varepsilon_T(x, v) := \inf\{\varepsilon \geq 0 \mid v \in T^\varepsilon(x)\}. \tag{11}$$

Observe that ε_T is always nonnegative. Note also that $\varepsilon_T(x, v) = 0$ if, and only if, $v \in T(x)$. It is also easy to see that the function ε_T is lower semi-continuous, and for any $\varepsilon \geq 0$ it holds that

$$\varepsilon \geq \varepsilon_T(x, v) \Leftrightarrow v \in T^\varepsilon(x).$$

Now we are ready to introduce our *merit function for the proximal system* (6)–(7). In particular, define $S_{T,\lambda,x} : \mathcal{H} \times \mathcal{H} \mapsto \Re \cup \{+\infty\}$ as

$$S_{T,\lambda,x}(y, v) := \|\lambda v + y - x\|^2 + 2\lambda \varepsilon_T(y, v). \tag{12}$$

This is indeed a merit function, because it is nonnegative, and

$$\begin{aligned} S_{T,\lambda,x}(y, v) = 0 &\Leftrightarrow v = (1/\lambda)(x - y), \quad v \in T(y) \\ &\Leftrightarrow (y, v) \text{ solves (6)–(7)}. \end{aligned}$$

Of course, in general the exact value of $\epsilon_T(y, v)$, and hence of $S_{T,\lambda,x}(y, v)$, is not computable. Observe however, that if some $\epsilon \geq 0$ is known such that $v \in T^\epsilon(y)$, then one immediately obtains an upper bound:

$$v \in T^\epsilon(y) \implies S_{T,\lambda,x}(y, v) \leq \|\lambda v + y - x\|^2 + 2\lambda\epsilon. \tag{13}$$

Computable bounds would be precisely what we shall use in our algorithm.

2.1. General properties of the proximal-system merit function

In this subsection, we primarily focus on properties of the merit function S in the general case, i.e., when no additional assumptions are imposed on the operator T , other than it being maximal monotone.

We start with a technical result which is a consequence of the “weak transportation formula” [6]. We shall prove it directly.

Lemma 1. Take $v_1 \in T^{\epsilon_1}(x_1)$, $v_2 \in T^{\epsilon_2}(x_2)$, and $p, q \geq 0$, $p + q = 1$. Define

$$\bar{x} := px_1 + qx_2, \quad \bar{v} := pv_1 + qv_2, \quad \bar{\epsilon} := p\epsilon_1 + q\epsilon_2 + pq\langle v_1 - v_2, x_1 - x_2 \rangle.$$

Then $\bar{\epsilon} \geq 0$ and $\bar{v} \in T^{\bar{\epsilon}}(\bar{x})$.

Proof. The identity

$$\langle w - \bar{v}, z - \bar{x} \rangle = p\langle w - v_1, z - x_1 \rangle + q\langle w - v_2, z - x_2 \rangle - pq\langle v_1 - v_2, x_1 - x_2 \rangle$$

holds for any $z, w \in \mathcal{H}$. In particular, if $w \in T(z)$, using Definition 9 and the above identity, we obtain that

$$\begin{aligned} \langle w - \bar{v}, z - \bar{x} \rangle &\geq -p\epsilon_1 - q\epsilon_2 - pq\langle v_1 - v_2, x_1 - x_2 \rangle \\ &= -\bar{\epsilon}. \end{aligned}$$

If $\bar{\epsilon} < 0$, then the maximality of T implies $\bar{v} \in T(\bar{x})$, and taking $z = \bar{x}$, $w = \bar{v}$, the above inequality yields a contradiction. Hence $\bar{\epsilon} \geq 0$, and by (9) it follows that $\bar{v} \in T^{\bar{\epsilon}}(\bar{x})$. □

Note that $\mathcal{H} \times \mathcal{H}$ is itself a Hilbert space, with the canonical inner product:

$$\langle (x_1, x_2), (y_1, y_2) \rangle := \langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle, \quad \forall (x_1, x_2), (y_1, y_2) \in \mathcal{H} \times \mathcal{H}.$$

We next prove a somewhat surprising property that $S_{T,\lambda,x}$ is a closed proper (strongly) convex function on $\mathcal{H} \times \mathcal{H}$.

Theorem 1. Take any $x \in \mathcal{H}$, $\lambda > 0$. The function $S_{T,\lambda,x}(\cdot, \cdot)$ is closed proper convex. Moreover, it is $(1, \lambda^2)$ -strongly convex, i.e., if $(a, b) \in \partial S_{T,\lambda,x}(z, w)$, then

$$\begin{aligned} S_{T,\lambda,x}(y, v) &\geq S_{T,\lambda,x}(z, w) + \langle y - z, a \rangle + \langle v - w, b \rangle \\ &\quad + \|y - z\|^2 + \lambda^2\|v - w\|^2 \end{aligned}$$

for any $(y, v) \in \mathcal{H} \times \mathcal{H}$.

Proof. Note that $S_{T,\lambda,x}$ is nonnegative, lower-semicontinuous, and not identically $+\infty$. Therefore, it is enough to show that for any

$$(x_1, v_1), (x_2, v_2) \in \mathcal{H} \times \mathcal{H}, \quad p, q \geq 0, \quad p + q = 1,$$

it holds that

$$S_{T,\lambda,x}(\bar{x}, \bar{v}) \leq pS_{T,\lambda,x}(x_1, v_1) + qS_{T,\lambda,x}(x_2, v_2) - pq(\|x_1 - x_2\|^2 + \lambda^2\|v_1 - v_2\|^2), \tag{14}$$

where

$$\bar{x} := px_1 + qx_2, \quad \bar{v} := pv_1 + qv_2.$$

Inequality (14) will then imply that $S_{T,\lambda,x}$ is (closed proper) convex and in fact, $(1, \lambda^2)$ -strongly convex on $\mathcal{H} \times \mathcal{H}$.

If $p = 0$ or $q = 0$ then (14) holds trivially. The same applies to the case $S_{T,\lambda,x}(x_1, v_1) = +\infty$ or $S_{T,\lambda,x}(x_2, v_2) = +\infty$. Therefore, let us suppose that

$$\varepsilon_1 := \varepsilon_T(x_1, v_1) < +\infty, \quad \varepsilon_2 := \varepsilon_T(x_2, v_2) < +\infty. \tag{15}$$

Then $v_1 \in T^{\varepsilon_1}(x_1)$, $v_2 \in T^{\varepsilon_2}(x_2)$, and from Lemma 1 we immediately get that

$$\varepsilon(\bar{x}, \bar{v}) \leq p\varepsilon_1 + q\varepsilon_2 + pq(v_1 - v_2, x_1 - x_2). \tag{16}$$

For any $\xi_1, \xi_2 \in \mathcal{H}$, it holds that

$$\|p\xi_1 + q\xi_2\|^2 = p\|\xi_1\|^2 + q\|\xi_2\|^2 - pq\|\xi_1 - \xi_2\|^2.$$

Taking $\xi_1 = \lambda v_1 + x_1 - x$, $\xi_2 = \lambda v_2 + x_2 - x$, and using the above identity, we further obtain

$$\|\lambda\bar{v} + \bar{x} - x\|^2 = p\|\lambda v_1 + x_1 - x\|^2 + q\|\lambda v_2 + x_2 - x\|^2 - pq\|\lambda(v_1 - v_2) + x_1 - x_2\|^2. \tag{17}$$

Multiplying inequality (16) by 2λ , adding it to inequality (17), and using (12), we have that

$$\begin{aligned} S_{T,\lambda,x}(\bar{x}, \bar{v}) &\leq p(\|\lambda v_1 + x_1 - x\|^2 + 2\lambda\varepsilon_1) + q(\|\lambda v_2 + x_2 - x\|^2 + 2\lambda\varepsilon_2) \\ &\quad + pq(2\lambda(v_1 - v_2, x_1 - x_2) - \|\lambda(v_1 - v_2) + x_1 - x_2\|^2) \\ &= p(\|\lambda v_1 + x_1 - x\|^2 + 2\lambda\varepsilon_1) + q(\|\lambda v_2 + x_2 - x\|^2 + 2\lambda\varepsilon_2) \\ &\quad - pq(\|x_1 - x_2\|^2 + \lambda^2\|v_1 - v_2\|^2). \end{aligned}$$

Using now (15) and again (12), the inequality (14) follows. □

The next result provides information on the structure of the subgradient of $S_{T,\lambda,x}(\cdot, \cdot)$.

Lemma 2. *Take any $x \in \mathcal{H}$, $\lambda > 0$. If $z \in \mathcal{H}$ and $w \in T(z)$, then*

$$2(u, \lambda u) \in \partial S_{T,\lambda,x}(z, w),$$

where $u := \lambda w + z - x$.

Proof. Since $w \in T(z)$, for any $y, v \in \mathcal{H}$ we have

$$\langle v - w, y - z \rangle \geq -\epsilon_T(y, v).$$

Multiplying the above inequality by -2λ and adding $\|\lambda v + y - x\|^2$ to both sides, we obtain that

$$\|\lambda v + y - x\|^2 - 2\lambda \langle v - w, y - z \rangle \leq S_{T,\lambda,x}(y, v).$$

Observe that $S_{T,\lambda,x}(z, w) = \|\lambda w + z - x\|^2$. Direct algebraic manipulations yield the following:

$$\begin{aligned} & \|\lambda v + y - x\|^2 - 2\lambda \langle v - w, y - z \rangle = \\ & = \|\lambda w + z - x\|^2 + 2\langle \lambda w + z - x, \lambda(v - w) + y - x \rangle \\ & \quad + \|\lambda(v - w) + y - z\|^2 - 2\lambda \langle v - w, y - z \rangle \\ & = S_{T,\lambda,x}(z, w) + 2\langle \lambda w + z - x, \lambda(v - w) + y - x \rangle \\ & \quad + \lambda^2 \|v - w\|^2 + \|y - z\|^2 \\ & \geq S_{T,\lambda,x}(z, w) + 2\langle \lambda w + z - x, \lambda(v - w) + y - x \rangle \\ & = S_{T,\lambda,x}(z, w) + 2\langle u, y - z \rangle + 2\langle \lambda u, v - w \rangle, \end{aligned}$$

which implies the desired result. □

As a direct consequence of the strong convexity of S , we obtain the following error bound.

Corollary 1. *Take any $x \in \mathcal{H}$ and $\lambda > 0$. Let y^*, v^* be the exact solution of the proximal system (6)–(7). Then for any $y, v \in \mathcal{H}$, it holds that*

$$\|y - y^*\|^2 + \lambda^2 \|v - v^*\|^2 \leq S_{T,\lambda,x}(y, v).$$

In particular,

$$\|y - y^*\| \leq \sqrt{S_{T,\lambda,x}(y, v)}.$$

Proof. It holds that $S_{T,\lambda,x}(y^*, v^*) = 0$. Since $S_{T,\lambda,x}$ is always nonnegative, (y^*, v^*) is its minimizer. By the classical optimality conditions for convex optimization, $(0, 0) \in \partial S_{T,\lambda,x}(y^*, v^*)$ (see also Lemma 2, where we take $u = 0$). Choosing now $z = y^*$ and $w = v^*$ in Theorem 1 implies the result. □

As an aside, we shall now briefly discuss the case where $T = \partial f$, the subdifferential of a proper lower semicontinuous convex function f , and state an additional property of our merit function in that case. The primary focus of the present paper is the variational inequality problem, not optimization, and so we do not study the case $T = \partial f$ in much detail. We note that an application to nondifferentiable optimization of the error tolerance criterion of [37] (which is conceptually related to the merit function introduced

above), can be found in [30,29] in the context of bundle methods. This development, however, is technically quite involved, and so we shall not describe it here.

If $T = \partial f$, then the proximal subproblem (5) is equivalent to the following optimization problem:

$$\min_{y \in \mathcal{H}} \Psi_{\lambda,x}(y) := \lambda f(y) + \frac{1}{2} \|y - x\|^2. \tag{18}$$

Let y^* be the solution of (18) (equivalently, the solution of (5)). By the convexity of $\Psi_{\lambda,x}$, for all $y \in \mathcal{H}$ and $r \in \partial \Psi_{\lambda,x}(y)$, it holds that

$$\Psi_{\lambda,x}(y) - \Psi_{\lambda,x}(y^*) \leq \langle r, y - y^* \rangle \leq \|r\| \|y - y^*\|.$$

Using Corollary 1, and the fact that $\partial \Psi_{\lambda,x}(y) \ni r = \lambda v + y - x$, where $v \in \partial f(y) = T$, we obtain that

$$\Psi_{\lambda,x}(y) - \Psi_{\lambda,x}(y^*) \leq S_{T,\lambda,x}(y, v).$$

Hence, S provides a bound for the optimal value of the proximal problem (18).

The next lemma shows how the proposed merit function S (or any appropriate computable upper bound, for example given by (13)), can be used to effectively manage the approximate solutions of proximal subproblems. Here, T is again an arbitrary maximal monotone operator.

Lemma 3. *Let y, v be an approximate solution of (6)–(7). Define*

$$x^+ := x - \lambda v.$$

Then for any solution of (2) x^ (i.e., $0 \in T(x^*)$), it holds that*

$$\|x^* - x^+\|^2 \leq \|x^* - x\|^2 + (S_{T,\lambda,x}(y, v) - \|y - x\|^2).$$

Proof. First note that $\langle x^* - y, 0 - v \rangle \geq -\epsilon_T(y, v)$. Therefore,

$$\langle x^* - y, x - x^+ \rangle \leq \lambda \epsilon_T(y, v).$$

Note also that $\|\lambda v + y - x\|^2 = \|y - x^+\|^2$. Combining these relations with the identity

$$\|x^* - x^+\|^2 - \|x^* - x\|^2 = \|y - x^+\|^2 + 2\langle x^* - y, x - x^+ \rangle - \|y - x\|^2,$$

and using (12), implies the desired result. □

Lemma 3 reveals that if we have any $y, v \in \mathcal{H}$ such that $S_{T,\lambda,x}(y, v)$ is small (in comparison with $\|y - x\|^2$), then $x^+ = x - \lambda v$ is closer to the solution set $T^{-1}(0)$ than x . This is the basis for general inexact extragradient-proximal framework discussed in Sect. 3.

2.2. The case of VIP

In this section, we study the precise relationship between inexact solutions of regularized proximal variational inequality subproblems (3), and inexact solutions of regularized inclusion subproblems (4). Specifically, we establish that the (computable!) regularized gap function measure for the former provides an upper bound for the proximal merit function S , defined by (12), for the latter.

The *regularized gap function* for $\text{VIP}(F, C)$, introduced independently by Fukushima [13] and Auchmuty [1], is defined as

$$\begin{aligned}
 f_{\alpha, F, C}(x) &:= \max_{y \in C} \langle F(x), x - y \rangle - \frac{1}{2\alpha} \|x - y\|^2 \\
 &= \langle F(x), x - P_C(x - \alpha F(x)) \rangle - \frac{1}{2\alpha} \|x - P_C(x - \alpha F(x))\|^2,
 \end{aligned}
 \tag{19}$$

where P_C denotes the orthogonal projection operator onto C , and $\alpha > 0$ is a parameter (we use F and C as subscripts in this definition, because in our applications F will change from one iteration to the next). This regularized gap function, and some other functions closely related to it (e.g., the implicit Lagrangian, the D-gap function), are fundamental in the theory, error bounds, and computational techniques for variational inequality and complementarity problems (see, for example, [13, 21, 41]; and [14, 31] for related surveys). Among other things, it holds that $f_{\alpha, F, C}(x) \geq 0$ for all $x \in C$ and all $\alpha > 0$, and $f_{\alpha, F, C}(x) = 0$ for some $x \in C$, $\alpha > 0$ if, and only if, x solves $\text{VIP}(F, C)$. Thus $f_{\alpha, F, C}$ is a *merit function* on the feasible set C for $\text{VIP}(F, C)$. Furthermore, $f_{\alpha, F, C}(\cdot)$ is continuous/differentiable whenever so is $F(\cdot)$ [13].

Having any $x \in C$, $\lambda > 0$, define

$$F_{\text{reg}}(y) := \lambda F(y) + y - x.$$

Then by $f_{1, F_{\text{reg}}, C}$ we shall denote the regularized gap function for $\text{VIP}(F_{\text{reg}}, C)$, where we set $\alpha = 1$ for simplicity. Our result is the following.

Lemma 4. *Take any $x \in C$, $\lambda > 0$. Then for any $y \in C$, it holds that*

$$S_{(F+N_C), \lambda, x}(y, v) \leq 2f_{1, F_{\text{reg}}, C}(y),$$

where

$$v := (1/\lambda) (x - P_C(x - \lambda F(y))).$$

Proof. Define

$$q := P_C(y - F_{\text{reg}}(y)), \quad v := (y - F_{\text{reg}}(y)) - q, \quad r := y - q, \quad \varepsilon := -(1/\lambda) \langle v, r \rangle.$$

Using the properties of the normal cone, it is easy to see that

$$v \in N_C(q), \quad \varepsilon \geq 0, \quad v + F_{\text{reg}}(y) = r.$$

Furthermore,

$$\begin{aligned}
 f_{1, F_{\text{reg}}, C}(y) &= \langle F_{\text{reg}}(y), y - q \rangle - (1/2)\|y - q\|^2 \\
 &= \langle r - v, r \rangle - (1/2)\|r\|^2 \\
 &= (1/2)\|r\|^2 - \langle v, r \rangle \\
 &= (1/2)\|r\|^2 + \lambda\varepsilon.
 \end{aligned}
 \tag{20}$$

Note also that

$$y - F_{\text{reg}}(y) = x - \lambda F(y),$$

and therefore

$$\begin{aligned}
 q &= P_C(x - \lambda F(y)), \\
 v &= x - \lambda F(y) - P_C(x - \lambda F(y)).
 \end{aligned}$$

Hence,

$$r = \lambda v + y - x, \tag{21}$$

$$v = F(y) + (1/\lambda)v. \tag{22}$$

Now we claim that

$$v \in (F + N_C)^\varepsilon(y). \tag{23}$$

Indeed, if $w \in (F + N_C)(z)$ then $z \in C$, and there exists some $\zeta \in N_C(z)$ such that $w = F(z) + \zeta$. Therefore

$$\begin{aligned}
 \langle v - w, y - z \rangle &= \langle F(y) - F(z), y - z \rangle + \langle (1/\lambda)v - \zeta, y - z \rangle \\
 &\geq \langle (1/\lambda)v - \zeta, y - z \rangle \\
 &= (1/\lambda)\langle v, y - z \rangle + \langle \zeta, z - y \rangle \\
 &\geq (1/\lambda)\langle v, y - z \rangle,
 \end{aligned}$$

where the first inequality follows from the monotonicity of F , and the second follows from $y \in C, \zeta \in N_C(z)$. Now observe that

$$\begin{aligned}
 \langle v, y - z \rangle &= \langle v, y - q \rangle + \langle v, q - z \rangle \\
 &\geq \langle v, y - q \rangle \\
 &= \langle v, r \rangle,
 \end{aligned}$$

where the inequality follows from $z \in C, v \in N_C(q)$. Combining the above results with the definition of ε , we conclude that for any z and any $w \in (F + N_C)(z)$, it holds that

$$\langle v - w, y - z \rangle \geq -\varepsilon,$$

which establishes (23). It immediately follows (see (13)) that

$$S_{(F+N_C), \lambda, x}(y, v) \leq \|\lambda v + y - x\|^2 + 2\lambda\varepsilon.$$

Now, using also (21) and (20), we obtain the desired result.

□

To conclude this section, we next establish an error bound on the distance to the solution of $VIP(F, C)$, assuming that the following condition is satisfied: $VIP(F, C)$ has the unique solution x^* , and there exists some $\mu > 0$ such that

$$\langle F(x), x - x^* \rangle \geq \mu \|x - x^*\|^2, \quad \forall x \in C. \tag{24}$$

Our error bound is an improvement of a bound based on the regularized gap function, and obtained in [38], in two respects. First, we obtain a tighter bound, i.e., the quantity bounding the distance to the solution set in our bound is smaller than the value of the regularized gap function. And secondly, our assumption (24) is weaker than the strong monotonicity of F used in [38]. Indeed, it is easy to see that (24) is weaker than the strong pseudo-monotonicity of F on C [18, 10], which is the property that for all $x, y \in C$ it holds that

$$\langle F(y), x - y \rangle \geq 0 \implies \langle F(x), x - y \rangle \geq \mu \|x - y\|^2.$$

It can be checked that if the strongly pseudo-monotone $VIP(F, C)$ has a solution, then this solution is unique. Condition (24) can be thought of as strong pseudo-monotonicity of F with respect to the solution x^* . The assumption of strong pseudo-monotonicity is in turn clearly weaker than strong monotonicity, which is

$$\langle F(x) - F(y), x - y \rangle \geq \mu \|x - y\|^2 \quad \forall x, y \in C.$$

Define

$$R_\alpha(x) := x - P_C(x - \alpha F(x)),$$

the *natural (projection) residual* for $VIP(F, C)$. Note that for $\|R_\alpha(x)\|$ to provide a global error bound for $VIP(F, C)$, F has to be strongly monotone and Lipschitz continuous [23]. Our assumptions here are weaker.

Theorem 2. *Suppose $VIP(F, C)$ has the unique solution x^* , and that (24) is satisfied. Then for any $\alpha \geq 1/\mu$ and every $x \in C$, it holds that*

$$\begin{aligned} \|x - x^*\| &\leq \frac{1}{2} \|R_\alpha(x)\| + \frac{1}{2} \sqrt{4\alpha \langle F(x), R_\alpha(x) \rangle - 3 \|R_\alpha(x)\|^2} \\ &\leq \sqrt{2\alpha f_{\alpha, F, C}(x)}. \end{aligned}$$

Proof. Take any $x \in C$. For $\alpha \geq 1/\mu$, using (24), we obtain

$$\alpha \langle F(x), x - x^* \rangle \geq \|x - x^*\|^2.$$

The latter relation can be further re-written as

$$\langle R_\alpha(x), x - x^* \rangle + \langle R_\alpha(x) - \alpha F(x), x^* - x \rangle \geq \|x - x^*\|^2. \tag{25}$$

Observe that since $x^* \in C$, by the properties of the projection operator,

$$\langle R_\alpha(x) - \alpha F(x), x^* - P_C(x - \alpha F(x)) \rangle \leq 0.$$

Hence,

$$\begin{aligned} \langle R_\alpha(x) - \alpha F(x), x^* - x \rangle &\leq \langle R_\alpha(x) - \alpha F(x), P_C(x - \alpha F(x)) - x \rangle \\ &= \alpha \langle F(x), R_\alpha(x) \rangle - \|R_\alpha(x)\|^2. \end{aligned}$$

Therefore, using (25) and the Cauchy-Schwarz inequality, we obtain

$$\|R_\alpha(x)\| \|x - x^*\| + \alpha \langle F(x), R_\alpha(x) \rangle - \|R_\alpha(x)\|^2 \geq \|x - x^*\|^2.$$

Denoting

$$t := \|x - x^*\|, \quad q := \|R_\alpha(x)\|, \quad s := \alpha \langle F(x), R_\alpha(x) \rangle - \|R_\alpha(x)\|^2,$$

and resolving the quadratic inequality in t

$$t^2 - qt - s \leq 0,$$

we conclude that

$$t \leq \frac{q + \sqrt{q^2 + 4s}}{2},$$

from which the first inequality in the assertion of the theorem follows immediately. It remains to show that the right-hand side in this inequality is no larger than $(2\alpha f_{\alpha, F, C}(x))^{1/2}$. Set

$$\eta := q^2 + 2s.$$

We further obtain

$$t \leq \frac{q + \sqrt{q^2 + 4s}}{2} = \frac{\sqrt{\eta - 2s} + \sqrt{\eta + 2s}}{2} \leq \sqrt{\eta},$$

where the last inequality follows from the concavity of the square root function (note that this inequality is in fact strict, unless $s = 0$). Observe now that

$$\begin{aligned} \eta &= q^2 + 2s \\ &= 2\alpha \langle F(x), R_\alpha(x) \rangle - \|R_\alpha(x)\|^2 \\ &= 2\alpha f_{\alpha, F, C}(x). \end{aligned}$$

The proof is complete. □

Note that since F_k in (3) is strongly monotone, Theorem 2 (with F replaced by F_k) provides a global error bound for $\text{VIP}(F_k, C)$.

3. Algorithmic applications

We start with a general algorithmic framework for solving the inclusion $0 \in T(x)$, based on the merit function S introduced in Sect. 2. This algorithm is a generalization of the Hybrid Extragradient–Proximal Point Algorithm discussed in [36]. In this form,

the algorithm is quite conceptual since it involves the values of the function S which, in general, may not be easy to compute. However, we believe it is meaningful to first develop a general convergent algorithm, and then pass to implementable methods based on it.

Algorithm 1. Choose any $x^0 \in \mathcal{H}$, and $\bar{\sigma} \in [0, 1)$, $\bar{\lambda} > 0$. Having x^k , choose a regularization parameter $\lambda_k \geq \bar{\lambda}$, and an error tolerance parameter $\sigma_k \in [0, \bar{\sigma}]$. Find $y^k, v^k \in \mathcal{H}$ satisfying

$$S_{T, \lambda_k, x^k}(y^k, v^k) \leq \sigma_k \|y^k - x^k\|^2. \tag{26}$$

Set

$$x^{k+1} := x^k - \lambda_k v^k,$$

and repeat for the next iteration $k := k + 1$.

It is quite clear that in the error tolerance criterion (26) we can actually use any upper bound for S . For example, due to the upper bound (13), the inexact proximal step in Algorithm 1 can be replaced by:

$$\begin{aligned} &\text{Find } y^k, v^k \in \mathcal{H}, \varepsilon_k \geq 0 \text{ such that} \\ &v^k \in T^{\varepsilon_k}(y^k), \\ &\|\lambda_k v^k + y^k - x^k\|^2 + 2\lambda_k \varepsilon_k \leq \sigma_k \|y^k - x^k\|^2. \end{aligned} \tag{27}$$

This special case of Algorithm 1 has been introduced in [36].

Algorithm 1 provides a very broad framework which can be used as a basis for dealing with a number of interesting applications depending on the structure of the operator T . When T is a sum of two operators one of which is single-valued and continuous, it is demonstrated in [36] that the recently proposed modified forward-backward splitting algorithm of Tseng [40] can be viewed as a specific implementation of Algorithm 1. An important feature of the method of Tseng is that the restrictive requirement that the continuous part of the operator must be strongly monotone can be relaxed to simple monotonicity, see [40]. Another application of Algorithm 1 discussed in [36] is a globally convergent regularized Newton method for the case when T is differentiable (see also [35, 33] for Newton-type methods based on the related approach of [37]). The key fact in this application is that under some natural assumptions, just *one* Newton step applied to the proximal subproblem $0 = \lambda_k T(\cdot) + (\cdot - x^k)$ is sufficient to solve it within the error tolerance (27), and hence also (26).

The main convergence properties of the algorithm are the following.

Theorem 3. *Let $\{x^k\}$ be any sequence generated by Algorithm 1. Then the sequence $\{x^k\}$ converges weakly to a solution of $0 \in T(x)$, provided it exists.*

Proof. The proof is similar to that in [36]. We shall supply some main points for completeness. Take any x^* such that $0 \in T(x^*)$. By Lemma 3 and (26), it follows that

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - (1 - \sigma_k) \|y^k - x^k\|^2. \tag{28}$$

Hence, the sequence $\{\|x^k - x^*\|\}$ converges, and since $\sigma_k \leq \bar{\sigma} < 1$ for all k , we have that

$$0 = \lim_{k \rightarrow \infty} \|y^k - x^k\|. \tag{29}$$

Now using (26), we conclude that

$$0 = \lim_{k \rightarrow \infty} S_{T, \lambda_k, x^k}(y^k, v^k),$$

or, equivalently,

$$\lim_{k \rightarrow \infty} \|\lambda_k v^k + y^k - x^k\| = 0 = \lim_{k \rightarrow \infty} \lambda_k \epsilon_T(y^k, v^k).$$

Using further the fact that $0 < \bar{\lambda} \leq \lambda_k$ and (29), the latter relations imply that

$$\lim_{k \rightarrow \infty} \|v^k\| = 0 = \lim_{k \rightarrow \infty} \epsilon_T(y^k, v^k). \tag{30}$$

By (28), the sequence $\{x^k\}$ is bounded, so it must have some weakly convergent subsequence. Let \hat{x} be any weak accumulation point of $\{x^k\}$, and let $\{x^{k_i}\}$ be some subsequence weakly converging to \hat{x} . By (29), \hat{x} is also the weak limit of $\{y^{k_i}\}$. By the definition of ϵ_T (see (11)), for any $q \in \mathcal{H}$ and $w \in T(q)$ it holds that

$$\langle w - v^{k_i}, q - y^{k_i} \rangle \geq -\epsilon_T(y^{k_i}, v^{k_i}).$$

Equivalently,

$$\langle w - 0, q - y^{k_i} \rangle \geq \langle \tilde{v}^{k_i}, q - y^{k_i} \rangle - \epsilon_T(y^{k_i}, v^{k_i}).$$

Because $\{y^{k_i}\}$ converges weakly to \hat{x} , taking the limit as $i \rightarrow \infty$, and using (30), we have that

$$\langle w - 0, q - \hat{x} \rangle \geq 0.$$

Since T is maximal monotone, it holds that $0 \in T(\hat{x})$, i.e., \hat{x} is a solution.

We have thus proved that every weak accumulation point of $\{x^k\}$ is a solution. The proof of uniqueness of the weak accumulation point in this setting is standard (e.g., see [28]).

□

We note that if no solution exists, it can be shown that the sequence $\{x^k\}$ is unbounded (see [36] for details).

Theorem 4. *Suppose, in addition, that there exist two positive constants L and δ such that*

$$v \in T(x), \|v\| \leq \delta \Rightarrow \|x - x^*\| \leq L\|v\|,$$

where x^* is the (unique) solution of $0 \in T(x)$. Then the sequence $\{x^k\}$ converges linearly to x^* .

It is interesting to note that the rate of linear convergence estimate for Algorithm 1 is actually strictly better than that for the classical inexact proximal point method of [28]; see [34] for this refined comparison.

We now turn our attention to the variational inequality problem (1). In this setting, the proximal point subproblem is $VIP(F_k, C)$ with $F_k(x) = \lambda_k F(x) + (x - x^k)$. Roughly speaking, the new algorithm we propose here computes $y^k \in C$ as an appropriate *approximate* solution of $VIP(F_k, C)$, and then defines $x^{k+1} := P_C(x^k - \lambda_k F(y^k))$. Clearly, the critical part in this framework is a constructive/implementable test to decide which y^k is acceptable. For this task, we propose to employ the regularized gap function for the subproblem $VIP(F_k, C)$:

$$\begin{aligned} f_k(x) &:= f_{1, F_k, C}(x) \\ &= \langle F_k(x), x - P_C(x - F_k(x)) \rangle - \frac{1}{2} \|x - P_C(x - F_k(x))\|^2. \end{aligned} \tag{31}$$

Specifically, we show that any $y^k \in C$ such that $f_k(y^k) \leq (\sigma_k/2) \|y^k - x^k\|^2$, is acceptable to obtain convergence.

Note that since f_k is a computable merit function, it can also be itself used for solving $VIP(F_k, C)$ via (approximately) resolving the problem

$$\min_{x \in C} f_k(x). \tag{32}$$

In fact, if F is differentiable, then f_k is also differentiable. And since F_k is strongly monotone, in that case problem (32) has the unique stationary point which is the solution of $VIP(F_k, C)$ [13]. So any reasonable algorithm employed to solve (32) would generate a sequence of iterates converging to the solution of $VIP(F_k, C)$. Of course, one might also choose any other method for solving $VIP(F_k, C)$, not necessarily related to f_k , and only use the latter for the error tolerance test.

We now state formally our algorithm.

Algorithm 2. Choose any $x^0 \in C$ and $\bar{\sigma} \in [0, 1)$, $\bar{\lambda} > 0$.

Having x^k , choose $\lambda_k \geq \bar{\lambda}$, $\sigma_k \in [0, \bar{\sigma}]$, and compute $y^k \in C$, an approximate solution of $VIP(F_k, C)$, satisfying

$$f_k(y^k) \leq \frac{\sigma_k}{2} \|y^k - x^k\|^2, \tag{33}$$

where f_k is defined in (31).

Set

$$x^{k+1} := P_C(x^k - \lambda_k F(y^k)), \tag{34}$$

and repeat for the next iteration $k := k + 1$.

Note that if y^k is the *exact* solution of $VIP(F_k, C)$ (for example, if the error tolerance parameter $\sigma_k = 0$, then necessarily $f_k(y^k) = 0$ which means $VIP(F_k, C)$ is solved exactly), then such k -th iteration of Algorithm 2 reduces to the step of the classical exact proximal point method. Of course, this is both natural and desirable. To see this, observe that if y^k solves $VIP(F_k, C)$, it then follows that there exists some $v^k \in N_C(y^k)$

such that $\lambda_k F(y^k) + v^k + y^k - x^k = 0$. Therefore, $x^k - \lambda_k F(y^k) = v^k + y^k$, and $x^{k+1} = P_C(v^k + y^k)$ must be equal to y^k , because $v^k \in N_C(y^k)$. In other words, in this case $x^{k+1} = y^k$, the exact solution of the proximal point subproblem. However, the inexact version of Algorithm 2, corresponding to $\sigma_k \neq 0$, is different from standard inexact schemes in the literature, which require the error terms to be *a priori* summable for convergence. It is worth to point out that the computational cost of the extragradient step (34) is negligible relative to the cost of solving $VIP(F_k, C)$.

Algorithm 2 preserves all the attractive convergence properties of the classical (exact or inexact) proximal point method. We shall establish this fact by showing that Algorithm 2 falls within the framework of the general scheme of Algorithm 1. Thus convergence results will follow from Theorems 3 and 4.

Theorem 5. *Let $\{x^k\}$ be any sequence generated by Algorithm 2. Then all the conclusions of Theorems 3 and 4 hold.*

Proof. Let $\{x^k\}$, $\{y^k\}$, $\{\lambda_k\}$, $\{\sigma_k\}$ be the sequences generated/employed by Algorithm 2. Define, for each k ,

$$v^k := (1/\lambda_k)[x^k - P_C(x^k - \lambda_k F(y^k))].$$

Applying Lemma 4 with $x = x^k$, $\lambda = \lambda_k$ and $y = y^k$, and taking into account the notation $F_{reg} = F_k$, $f_1, F_{reg, C} = f_k$, we conclude that

$$S_{T, \lambda_k, x^k}(y^k, v^k) \leq 2f_k(y^k),$$

where

$$T := F + N_C.$$

Using the error tolerance rule (33), it now follows that

$$S_{T, \lambda_k, x^k}(y^k, v^k) \leq \sigma_k \|y^k - x^k\|^2.$$

Note also that from the definition of v^k ,

$$\begin{aligned} x^k - \lambda_k v^k &= P_C(x^k - \lambda_k F(y^k)) \\ &= x^{k+1}. \end{aligned}$$

Therefore, the sequences $\{x^k\}$, $\{y^k\}$, $\{v^k\}$ are valid sequences for Algorithm 1, and the convergence results follow, taking into account the equivalence between the problem $0 \in T(x)$ and $VIP(F, C)$. □

As a future application of our new algorithm, we mention the development of a globalization strategy for the (Josephy-) Newton method for variational inequality problems (see [17,3,16] for its definition and local convergence properties). For the monotone nonlinear complementarity problem, a special case of VIP, a *truly* globally convergent Newton method based on the framework of [37] and [32] has recently been developed in [33]. This algorithm has the advantage that the *whole* sequence of iterates

globally converges to a solution of the problem without any (local) solution-uniqueness or regularity-type assumptions (in particular, the solution set can even be unbounded). In the context of Newton-type methods, this property is quite remarkable.

To further highlight the advantages of our Algorithm 2 when compared to typical (i.e., “summable error”) approximation criteria, we next observe that we could, in principle, also employ a criterion of this type. Specifically, we could replace our tolerance rule (33) with

$$f_k(y^k) \leq \frac{\delta_k}{2}, \quad \sum_{k=0}^{\infty} \sqrt{\delta_k} < \infty, \tag{35}$$

and accept such y^k as the next iterate:

$$x^{k+1} := y^k. \tag{36}$$

This scheme would be much in the spirit of classical inexact proximal methods (e.g., [28, 8]). However, as already discussed, such a rule is typically not constructive. It is also worth to note that if our rule (33) is used, the extragradient step in (34) is indispensable, i.e., if one accepts y^k as the next iterate x^{k+1} as in (36), the method may diverge (see an example in [36]). This also shows that (33) is less restrictive than (35).

To verify that the method based on (35)–(36) converges, we shall only supply the key steps. Define, for each k ,

$$v^{k+1} := (1/\lambda_k)[x^k - P_C(x^k - \lambda_k F(x^{k+1}))].$$

Then applying Lemma 4 and (35), we obtain

$$S_{(F+N_C), \lambda_k, x^k}(x^{k+1}, v^{k+1}) \leq \delta_k.$$

Now using the error bound of Corollary 1, we conclude that

$$\|x^{k+1} - (I + \lambda_k(F + N_C))^{-1} x^k\| \leq \sqrt{\delta_k}, \quad \sum_{k=0}^{\infty} \sqrt{\delta_k} < \infty,$$

and from here, the classical convergence analysis of [28] applies. For the linear convergence result, one further has to assume that

$$f_k(x^{k+1}) \leq \frac{\delta_k}{2} \|x^{k+1} - x^k\|^2, \quad \sum_{k=0}^{\infty} \sqrt{\delta_k} < \infty,$$

similar to [28]. Note that another advantage of Algorithm 2 is that one does not need to change the error tolerance criterion to obtain linear convergence. Furthermore, from the convergence of Algorithm 2 one obtains, *a posteriori*, that

$$\sum_{k=0}^{\infty} f_k(y^k) < \infty,$$

while (35) requires

$$\sum_{k=0}^{\infty} \sqrt{f_k(x^{k+1})} < \infty,$$

which is a stronger condition.

4. Concluding remarks

Various merit functions, error measures, and error bounds for the approximate solution of proximal point subproblems have been discussed. A new inexact proximal point method for solving the variational inequality problem has been proposed. The tolerance criterion in approximate solution of subproblems was based on the regularized gap function. This tolerance criterion is readily implementable, and its use for variational inequalities is quite natural. An important intended application of the presented algorithm is the construction of regularized Newton-type methods for the variational inequality problem with improved global convergence properties, along the lines of those for systems of equations [35] and nonlinear complementarity problems [33].

Acknowledgements. Research of the first author is supported by CNPq Grant 300734/95-6, by PRONEX–Optimization and by FAPERJ, research of the second author is supported by CNPq Grant 301200/93-9(RN), by PRONEX–Optimization and by FAPERJ.

References

1. Auchmuty, G. (1989): Variational principles for variational inequalities. *Numer. Funct. Anal. Optim.* **10**, 863–874
2. Auslender, A. (1987): Numerical methods for nondifferentiable convex optimization. *Math. Program. Study* **30**, 102–126
3. Bonnans, J.F. (1994): Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Appl. Math. Optim.* **29**, 161–186
4. Burachik, R.S., Iusem, A.N., Svaiter, B.F. (1997): Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.* **5**, 159–180
5. Burachik, R.S., Sagastizábal, C.A., Svaiter, B.F. (1999): Bundle methods for maximal monotone operators. In: Tichatschke, R., Théra, M., eds., *Ill-posed variational problems and regularization techniques. Lecture Notes in Economics and Mathematical Systems*, No. 477, pp. 49–64, Springer
6. Burachik, R.S., Sagastizábal, C.A., Svaiter, B.F. (1999): \mathcal{E} -Enlargements of maximal monotone operators: Theory and applications. In: Fukushima, M., Qi, L., eds., *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pp. 25–44, Kluwer Academic Publishers
7. Burachik, R.S., Svaiter, B.F. (1999): \mathcal{E} -Enlargements of maximal monotone operators in Banach spaces. *Set-Valued Anal.* **7**, 117–132
8. Burke, J.V., Qian, M. (1998): A variable metric proximal point algorithm for monotone operators. *SIAM J. Control Optim.* **37**, 353–375
9. Cominetti, R. (1997): Coupling the proximal point algorithm with approximation methods. *J. Optim. Theory Appl.* **95**, 581–600
10. Crouzeix, J.-P. (1998): Characterizations of generalized convexity and generalized monotonicity, a survey. In: Crouzeix, J.-P. et al., eds., *Generalized convexity, generalized monotonicity: Recent results*, pp. 237–256, Kluwer Academic Publishers
11. Eckstein, J. (1998): Approximate iterations in Bregman-function-based proximal algorithms. *Math. Program.* **83**, 113–123
12. Eckstein, J., Bertsekas, D.P. (1992): On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* **55**, 293–318
13. Fukushima, M. (1992): Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Math. Program.* **53**, 99–110
14. Fukushima, M. (1996): Merit functions for variational inequality and complementarity problems. In: Di Pillo, G., Giannessi, F., eds., *Nonlinear Optimization and Applications*, pp. 155–170, Plenum Publishing Corporation, New York, New York
15. Güler, O. (1992): New proximal point algorithms for convex minimization. *SIAM J. Optim.* **2**, 649–664
16. Harker, P.T., Pang, J.-S. (1990): Finite-dimensional variational inequality problems : A survey of theory, algorithms and applications. *Math. Program.* **48**, 161–220
17. Josephy, N.H. (1979): Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin

18. Karamardian, S., Schaible, S. (1990): Seven kinds of monotone maps. *J. Optim. Theory Appl.* **66**, 37–47
19. Lemaire, B. (1989): The proximal algorithm. In: Penot, J.P., ed., *New Methods of Optimization and Their Industrial Use. International Series of Numerical Mathematics* 87, pp. 73–87, Birkhauser, Basel
20. Luque, F.J. (1984): Asymptotic convergence analysis of the proximal point algorithm. *SIAM J. Control Optim.* **22**, 277–293
21. Mangasarian, O.L., Solodov, M.V. (1993): Nonlinear complementarity as unconstrained and constrained minimization. *Math. Program.* **62**, 277–297
22. Martinet, B. (1970): Regularisation d'inequations variationnelles par approximations successives. *Revue Française d'Informatique et de Recherche Opérationnelle* **4**, 154–159
23. Pang, J.-S. (1987): A posteriori error bounds for the linearly-constrained variational inequality problem. *Math. Oper. Res.* **12**, 474–484
24. Pang, J.-S. (1997): Error bounds in mathematical programming. *Math. Program.* **79**, 299–332
25. Reich, S. (1979): Weak convergence theorems for nonexpansive mappings in Banach spaces. *J. Math. Anal. Appl.* **67**, 274–276
26. Rockafellar, R.T. (1970): On the maximality of sums of nonlinear monotone operators. *Trans. Am. Math. Soc.* **149**, 75–88
27. Rockafellar, R.T. (1976): Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1**, 97–116
28. Rockafellar, R.T. (1976): Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**, 877–898
29. Sagastizábal, C. (1999): Variantes de métodos de haces para resolver problemas de talla grande. In: Montenegro, J. et al., eds., *CEDYA 99, Actas*, pp. 512–519, Universidad de Las Palmas
30. Sagastizábal, C. (2000): Bundle management: a hybrid method suitable after compression. Working paper
31. Solodov, M.V. (2000): Implicit Lagrangian. In: Floudas, C., Pardalos, P., eds., *Encyclopedia of Optimization*. Kluwer Academic Publishers
32. Solodov, M.V., Svaiter, B.F. (1999): A new projection method for variational inequality problems. *SIAM J. Control Optim.* **37**, 765–776
33. Solodov, M.V., Svaiter, B.F. (2000): A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. *SIAM J. Optim.* **10**, 605–625
34. Solodov, M.V., Svaiter, B.F. (2000): A comparison of rates of convergence of two inexact proximal point algorithms. In: Di Pillo, G., Giannessi, F., eds., *Nonlinear Optimization and Related Topics*, pp. 415–427. Kluwer Academic Publishers
35. Solodov, M.V., Svaiter, B.F. (1999): A globally convergent inexact Newton method for systems of monotone equations. In: Fukushima, M., Qi, L., eds., *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pp. 355–369, Kluwer Academic Publishers
36. Solodov, M.V., Svaiter, B.F. (1999): A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.* **7**, 323–345
37. Solodov, M.V., Svaiter, B.F. (1999): A hybrid projection – proximal point algorithm. *J. Convex Anal.* **6**, 59–70
38. Taji, K., Fukushima, M., Ibaraki, T. (1993): A globally convergent Newton method for solving strongly monotone variational inequalities. *Math. Program.* **58**, 369–383
39. Tossings, P. (1994): The perturbed proximal point algorithm and some of its applications. *Appl. Math. Optim.* **29**, 125–159
40. Tseng, P. (2000): A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.* **38**, 431–446
41. Tseng, P., Yamashita, N., Fukushima, M. (1996): Equivalence of complementarity problems to differentiable minimization : A unified approach. *SIAM J. Optim.* **6**, 446–460