MATHEMATICS OF OPERATIONS RESEARCH
Vol. 25, No. 2, May 2000, pp. 214–230
Printed in U.S.A.

AN INEXACT HYBRID GENERALIZED PROXIMAL POINT ALGORITHM AND SOME NEW RESULTS ON THE THEORY OF BREGMAN FUNCTIONS

M. V. SOLODOV AND B. F. SVAITER

We present a new Bregman-function-based algorithm which is a modification of the generalized proximal point method for solving the variational inequality problem with a maximal monotone operator. The principal advantage of the presented algorithm is that it allows a more constructive error tolerance criterion in solving the proximal point subproblems. Furthermore, we eliminate the assumption of pseudomonotonicity which was, until now, standard in proving convergence for paramonotone operators. Thus we obtain a convergence result which is new even for exact generalized proximal point methods. Finally, we present some new results on the theory of Bregman functions. For example, we show that the standard assumption of convergence consistency is a consequence of the other properties of Bregman functions, and is therefore superfluous.

1. Introduction. In this paper, we are concerned with proximal point algorithms for solving the variational inequality problem. Specifically, we consider the methods which are based on Bregman distance regularization. Our objective is two-fold. First of all, we develop a hybrid algorithm based on inexact solution of proximal subproblems. The important new feature of the proposed method is that the error tolerance criterion imposed on inexact subproblem solution is constructive and easily implementable for a wide range of applications. Second, we obtain a number of new results on the theory of Bregman functions and on the convergence of related proximal point methods. In particular, we show that one of the standard assumptions on the Bregman function (convergence consistency), as well as one of the standard assumptions on the operator defining the problem (pseudomonotonicity, in the paramonotone operator case), are extraneous.

Given an operator T on R^n (point-to-set, in general) and a closed convex subset C of R^n the associated *variational inequality problem* (Cottle, Giannessi and Lions 1980), from now on VIP(T,C), is to find a pair x^* and v^* such that

$$x^* \in C, \qquad v^* \in T(x^*), \qquad \langle x - x^*, v^* \rangle \geq 0 \qquad \forall x \in C, \tag{1}$$

where $\langle \cdot,\cdot \rangle$ stands for the usual inner product in R^n. The operator $T : R^n \to \mathscr{P}(R^n)$, where $\mathscr{P}(R^n)$ stands for the family of subsets of R^n, is *monotone* if

$$\langle u - v, x - y \rangle \geq 0$$

for any $x, y \in R^n$ and any $u \in T(x)$, $v \in T(y)$. T is *maximal monotone* if it is monotone and its graph $G(T) = \{(x,u) \in R^n \times R^n \mid u \in T(x)\}$ is not contained in the graph of any other monotone operator. Throughout this paper we assume that T is maximal monotone.

It is well known that VIP(T,C) is closely related to the problem of finding a zero of a maximal monotone operator $\hat{T}$:

$$0 \in \hat{T}(z), \qquad z \in R^n. \tag{2}$$

Received May 27, 1998; revised March 23, 1999 and July 8, 1999.
AMS subject classification. Primary: 90C25; secondary: 90C30.
OR/MS subject classification. Primary: Programming/nonlinear.
Key words. Proximal point method, Bregman function, maximal monotone operator, variational inequality.

0364-765X/00/2502/0214/$05.00
1526-5471 electronic ISSN,

Recall that we assume that T is maximal monotone. Therefore, (2) is a particular case of VIP(T, C) for $C = R^n$. On the other hand, define N_C as the normal cone operator; that is, $N_C : R^n \to \mathcal{P}(R^n)$ is given by

$$N_C(x) = \begin{cases} \{v \in R^n \mid \langle v, y - x \rangle \leq 0 \ \ \forall y \in C\} & \text{if } x \in C, \\ \emptyset, & \text{otherwise.} \end{cases}$$

The operator $T + N_C$ is monotone and x^* solves VIP(T, C) (with some $v^* \in T(x^*)$) if and only if

$$0 \in (T + N_C)(x^*).$$

Additionally, if the relative interiors of C and of the domain of T intersect, then $T + N_C$ is maximal monotone (Rockafellar 1970a), and the above inclusion is a particular case of (2), i.e., the problem of finding a zero of a maximal monotone operator. Hence, in this case, VIP(T, C) can be solved using the classical proximal point method for finding a zero of the operator $\hat{T} = T + N_C$. The proximal point method was introduced by Martinet (1970) and further developed by Rockafellar (1976b). Some other relevant papers on this method, its applications and modifications, are Moreau (1965), Rockafellar (1976), Brézis and Lions (1978), Passty (1979), Luque (1984), Ferris (1991), Güler (1992), and Eckstein and Bertsekas (1992); see Lemaire (1989) for a survey. The classical proximal point algorithm generates a sequence $\{x^k\}$ by solving a sequence of *proximal subproblems*. The iterate x^{k+1} is the solution of

$$0 \in c_k \hat{T}(x) + x - x^k,$$

where $c_k > 0$ is a regularization parameter. For the method to be implementable, it is important to handle approximate solutions of subproblems. This consideration gives rise to the inexact version of the method (Rockafellar 1976b), which can be written as

$$e^{k+1} + x^k \in c_k \hat{T}(x^{k+1}) + x^{k+1},$$

where e^{k+1} is the associated error term. To guarantee convergence, it is typically assumed that (see, for example, Rockafellar 1976b and Burke and Qian 1998),

$$\sum_{k=0}^{\infty} \|e^k\| < \infty.$$

Note that even though the proximal subproblems are better conditioned than the original problem, structurally they are as difficult to solve. This observation motivates the development of the "nonlinear" or "generalized" proximal point method; see Eggermont (1990), Eckstein (1993), Chen and Teboulle (1993), Iusem (1995), Kiwiel (1997), Iusem and Solodov (1997), Iusem (1998a), and Burachik and Iusem (1998).

In the generalized proximal point method, x^{k+1} is obtained solving the *generalized proximal point subproblem*

$$0 \in c_k T(x) + \nabla f(x) - \nabla f(x^k).$$

The function f is the Bregman (1967) function; namely it is strictly convex, differentiable in the interior of C and its gradient is divergent on the boundary of C (f also has to satisfy some additional technical conditions, which we shall discuss in §2). All information about the feasible set C is embedded in the function f, which is both a regularization and a penalization term. Properties of f (discussed in §2) ensure that solutions of subproblems belong to the interior of C without any explicit consideration of constraints. The advantage of the generalized proximal point method is that the subproblems are essentially unconstrained. For example, if VIP(T, C) is the classical nonlinear complementarity problem (Pang 1995), then a reasonable choice of f gives proximal subproblems which are (unconstrained!) systems of nonlinear equations. By contrast, subproblems given by the classical proximal algorithm are themselves nonlinear complementarity problems, which are structurally considerably more difficult to solve than systems of equations. We refer the reader to Burachik and Iusem (1998) for a detailed example.

As in the case of the classical method, implementable versions of the generalized proximal point algorithm must take into consideration inexact solution of subproblems:

$$e^{k+1} + \nabla f(x^k) \in c_k T(x^{k+1}) + \nabla f(x^{k+1}).$$

In Eckstein (1998), it was established that if

$$(3) \qquad \sum_{k=0}^{\infty} \|e^k\| < \infty \quad \text{and} \quad \sum_{k=0}^{\infty} \langle e^k, x^k \rangle \quad \text{exists and is finite,}$$

then the generated sequence converges to a solution (provided it exists) under basically the same assumptions that are needed for the convergence of the exact method. Other inexact generalized proximal algorithms are Burachik, Iusem and Svaiter (1997), Kiwiel (1997) and Teboulle (1997). However, the approach of Eckstein (1998) is the simplest and the easiest to use in practical computation (see the discussion in Eckstein 1998). Still, the error criterion given by (3) is not totally satisfactory. Obviously, there exist many error sequences $\{e^k\}$ that satisfy the first relation in (3), and it is not very clear which e^k should be considered acceptable for each specific iteration k. In this sense, criterion (3) is not quite constructive. The second relation in (3) is even somewhat more problematic.

In this paper, we present a hybrid generalized proximal-based algorithm which employs a more constructive error criterion than (3). Our method is completely implementable when the gradient of f is easily invertible, which is a common case for many important applications. The inexact solution is used to obtain the new iterate in a way very similar to Bregman generalized projections. When the error is zero, our algorithm coincides with the generalized proximal point method. However, for nonzero error, it is different from the inexact method of Eckstein (1998) described above. Our new method is motivated by Solodov and Svaiter (1999b, c), where a constructive error tolerance was introduced for the classical proximal point method. This approach has already proved to be very useful in a number of applications (Solodov and Svaiter 1999a, 2000a, b, c).

Besides the algorithm, we also present a theoretical result which is new even for exact methods. In particular, we prove convergence of the method for paramonotone operators, without the previously used assumption of pseudomonotonicity (paramonotone operators were introduced in Bruck (1975,1976); see also Censor and Zenios (1992) and Iusem (1998b); we shall state this definition in §3, together with the definition of pseudomonotonicity). It is important to note that the subgradient of a proper closed convex function is paramonotone, but need not be pseudomonotone. Hence, among other things, our result unifies the proof of convergence for paramonotone operators and for minimization.

We also remove the condition of convergence consistency which has been used to characterize Bregman functions, proving it to be a consequence of the other properties.

This work is organized as follows. In §2, we discuss Bregman functions and derive some new results on their properties. In §3, the error tolerance to be used is formally defined, the new algorithm is described and the convergence result is stated. Section 4 contains convergence analysis.

A few words about our notation are in order. Given a (convex) set A, ri(A) will denote the relative interior, $\bar{A}$ will denote the closure, int(A) will denote the interior, and bdry(A) will denote the boundary of A. For an operator T, Dom(T) stands for its domain, i.e., all points $x \in R^n$ such that $T(x) \neq \emptyset$.

2. Bregman function and Bregman distance. Given a convex function f on R^n, finite at $x, y \in R^n$ and differentiable at y, the *Bregman distance* (1967) between x and y, determined by f, is

$$D_f(x, y) := f(x) - f(y) - \langle \nabla f(y), x - y \rangle.$$

Note that, by the convexity of f, the Bregman distance is always nonnegative. We mention here the recent article (Bauschke and Borwein 1997) as one good reference on Bregman functions and their properties.

DEFINITION 2.1. Given S, a convex open subset of R^n, we say that $f : \bar{S} \to R$ is a *Bregman function with zone S* if

(1) f is strictly convex and continuous in $\bar{S}$,
(2) f is continuously differentiable in S,
(3) for any $x \in \bar{S}$ and $\alpha \in R$, the right partial level set

$$L(x, \alpha) = \{ y \in S \mid D_f(x, y) \leq \alpha \}$$

is bounded,
(4) If $\{y^k\}$ is a sequence in S converging to y, then

$$\lim_{k \to \infty} D_f(y, y^k) = 0.$$

Some remarks are in order regarding this definition. In addition to the above four items, there is one more standard requirement for Bregman function, namely *Convergence Consistency*:

If $\{x^k\} \subset \bar{S}$ is bounded, $\{y^k\} \subset S$ converges to y, and
$\lim_{k \to \infty} D_f(x^k, y^k) = 0$, then $\{x^k\}$ also converges to y.

This requirement has been imposed in all previous studies of Bregman functions and related algorithms (Censor and Zenios 1992, Chen and Teboulle 1993, Eckstein 1993, Iusem 1995, Censor, Iusem and Zenios 1998, Eckstein 1998, Bauschke and Borwein 1997, Iusem and Solodov 1997, Iusem 1998a, and Burachik and Isuem 1998). In what follows, we shall establish that convergence consistency holds automatically as a consequence of Definition 2.1 (we shall actually prove a stronger result).

The original definition of a Bregman function also requires the left partial level sets

$$L'(\alpha, y) = \{ x \in \bar{S} \mid D_f(x, y) \leq \alpha \}$$

to be bounded for any $y \in S$. However, it has already been observed that this condition is not needed to prove convergence of proximal methods (e.g., Eckstein 1998). And it is known that this boundedness condition is extraneous regardless, since it is also a consequence of Definition 2.1 (e.g., see Bauschke and Borwein 1997). Indeed, observe that for any y, the level set $L'(0, y) = \{y\}$, so it is nonempty and bounded. Also Definition 2.1 implies that $D_f(\cdot, y)$ is a proper closed convex function. Because this function has one level set which in nonempty and bounded, it follows that all of its level sets are bounded (i.e., $L'(\alpha, y)$ is bounded for every α) (Rockafellar 1970b, Corollary 8.7.1).

To prove convergence consistency using the properties given in Definition 2.1, we start with the following results.

LEMMA 2.2 (THE RESTRICTED TRIANGULAR INEQUALITY). *Let f be a convex function satisfying items* (1) *and* (2) *of Definition* 2.1. *If $x \in \bar{S}$, $y \in S$ and w is a proper convex* *combination of* x *and* y, *i.e.*, $w=(1-\theta)x+\theta y$ *with* $\theta\in(0,1)$, *then*

$$D_f(x,w)+D_f(w,y)\le D_f(x,y).$$

PROOF. We have that $w=y+(1-\theta)(x-y)$ with $0<\theta<1$. Clearly, $w\in S$. Since ∇f is monotone,

$$\langle\nabla f(w),\, w-y\rangle\ge\langle\nabla f(y),\, w-y\rangle.$$

Taking into account that $w-y=(1-\theta)(x-y)$, the latter relation yields

$$\langle\nabla f(w),\, x-y\rangle\ge\langle\nabla f(y),\, x-y\rangle.$$

Therefore

$$\begin{aligned}
&D_f(x,w)+D_f(w,y)\\
&\quad=[f(x)-f(w)-\langle\nabla f(w),\, x-w\rangle]\\
&\qquad+[f(w)-f(y)-\langle\nabla f(y),\, w-y\rangle]\\
&\quad= f(x)-f(y)-\theta\langle\nabla f(w),\, x-y\rangle-(1-\theta)\langle\nabla f(y),\, x-y\rangle\\
&\quad\le f(x)-f(y)-\theta\langle\nabla f(y),\, x-y\rangle-(1-\theta)\langle\nabla f(y),\, x-y\rangle\\
&\quad= f(x)-f(y)-\langle\nabla f(y),\, x-y\rangle\\
&\quad= D_f(x,y). \quad\square
\end{aligned}$$

LEMMA 2.3. *Let* f *be a convex function satisfying items* (1) *and* (2) *of Definition* 2.1. *If* $\{x^k\}$ *is a sequence in* S *converging to* x, $\{y^k\}$ *is a sequence in* S *converging to* y *and* $y\neq x$, *then*

$$\liminf_{k\to\infty} D_f(x^k,y^k)>0.$$

PROOF. Define

$$z^k=(1/2)(x^k+y^k).$$

Clearly $\{z^k\}$ is a sequence in S converging to $z=(1/2)(x+y)\in\bar{S}$. By the convexity of f, it follows that for all k:

$$\begin{aligned}
f(z^k)&\ge f(y^k)+\langle\nabla f(y^k),\, z^k-y^k\rangle\\
&=f(y^k)+(1/2)\langle\nabla f(y^k),\, x^k-y^k\rangle.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\frac{f(x^k)+f(y^k)}{2}-f(z^k)\\
&\quad\le\frac{f(x^k)+f(y^k)}{2}-f(y^k)-\frac{1}{2}\langle\nabla f(y^k),\, x^k-y^k\rangle\\
&\quad= D_f(x^k,y^k)/2.
\end{aligned}$$

Letting $k\to\infty$ we obtain

$$\frac{f(x)+f(y)}{2}-f\left(\frac{x+y}{2}\right)\le\frac{1}{2}\liminf_{k\to\infty} D_f(x^k,y^k).$$

Using the strict convexity of f and the hypothesis $x\neq y$, the desired result follows. $\square$

We are now ready to prove a result which is actually stronger than the property of convergence consistency discussed above. This result will be crucial for strengthening convergence properties of proximal point methods, carried out in this paper.

THEOREM 2.4. *Let f be a convex function satisfying items* (1) *and* (2) *of Definition* 2.1. *If $\{x^k\}$ is a sequence in $\bar{S}$, $\{y^k\}$ is a sequence in S,*

$$\lim_{k\to\infty} D_f(x^k, y^k) = 0,$$

and one of the sequences ($\{x^k\}$ or $\{y^k\}$) converges, then the other also converges to the same limit.

PROOF. Suppose, by contradiction, that one of the sequences converges and the other does not converge or does not converge to the same limit. Then there exist some $\varepsilon > 0$ and a subsequence of indices $\{k_j\}$ satisfying

$$\|x^{k_j} - y^{k_j}\| > \varepsilon.$$

Suppose first that $\{y^k\}$ converges and

$$\lim_{k\to\infty} y^k = y.$$

Define

$$\tilde{x}^j = y^{k_j} + \frac{\varepsilon}{\|x^{k_j} - y^{k_j}\|}(x^{k_j} - y^{k_j}),$$

i.e., $\tilde{x}^j$ is a proper convex combination of x^{k_j} and y^{k_j}. Using Lemma 2.2 we conclude that $D_f(\tilde{x}^j, y^{k_j}) \le D_f(x^{k_j}, y^{k_j})$, which implies that

$$\lim_{j\to\infty} D_f(\tilde{x}^j, y^{k_j}) = 0.$$

Since $\|\tilde{x}^j - y^{k_j}\| = \varepsilon$ and $\{y^{k_j}\}$ converges, it follows that $\{\tilde{x}^j\}$ is bounded and there exists a subsequence $\{\tilde{x}^{j_i}\}$ converging to some $\tilde{x}$. Therefore we have the following set of relations

$$\begin{aligned} \lim_{i\to\infty} D_f(\tilde{x}^{j_i}, y^{k_{j_i}}) &= 0,\\ \lim_{i\to\infty} y^{k_{j_i}} &= y,\\ \lim_{i\to\infty} \tilde{x}^{j_i} &= \tilde{x},\\ \|\tilde{x} - y\| &= \varepsilon > 0, \end{aligned}$$

which is in contradiction with Lemma 2.3.

If we assume that the sequence $\{x^k\}$ converges, then reversing the roles of $\{x^k\}$ and $\{y^k\}$ in the argument above, we reach a contradiction with Lemma 2.3 in exactly the same manner. □

It is easy to see that convergence consistency is an immediate consequence of Theorem 2.4.

We next state a well-known result which is widely used in the analysis of generalized proximal point methods.

LEMMA 2.5 (THREE-POINT LEMMA) (CHEN AND TEBOULLE 1993). *Let f be a Bregman function with zone S as in Definition* 2.1. *For any $x, z \in S$ and $y \in \bar{S}$, it holds that*

$$D_f(y,x) = D_f(z,x) + D_f(y,z) + \langle \nabla f(x) - \nabla f(z), z - y \rangle.$$

In the sequel, we shall use the following consequence of Lemma 2.5, which can be obtained by subtracting the three-point inequalities written with y, x, z and s, x, z.

COROLLARY 2.6 (FOUR-POINT LEMMA). *Let f be a Bregman function with zone S as in Definition* 2.1. *For any $x,z \in S$ and $y,s \in \bar{S}$, it holds that*

$$D_f(s,z) = D_f(s,x) + \langle \nabla f(x) - \nabla f(z), s - y\rangle + D_f(y,z) - D_f(y,x).$$

3. The inexact generalized proximal point method. We start with some assumptions which are standard in the study and development of Bregman-function-based algorithms.

Suppose C, the feasible set of VIP(T,C), has nonempty interior, and we have chosen f, an associated Bregman function with zone int(C). We also assume that

$$\mathrm{int}(C) \cap \mathrm{Dom}(T) \neq \emptyset,$$

so that $T + N_C$ is maximal monotone (Rockafellar 1970a). The solution set of VIP(T,C) is

$$X^* := \{s \in C \mid \exists v_s \in T(s),\ \forall y \in C\ \langle v_s, y - s\rangle \geq 0\}.$$

We assume this set to be nonempty, since this is the more interesting case. In principle, following standard analysis, results regarding unboundedness of the iterates can be obtained for the case when no solution exists.

Additionally, we need the assumptions which guarantee that proximal subproblem solutions exist and belong to the interior of C.

H1. For any $x \in \mathrm{int}(C)$ and $c > 0$, the generalized proximal subproblem

$$0 \in cT(\cdot) + \nabla f(\cdot) - \nabla f(x).$$

has a solution.

H2. For any $x \in \mathrm{int}(C)$, if $\{y^k\}$ is a sequence in int(C) and

$$\lim_{k\to\infty} y^k = y \in \mathrm{bdry}(C)$$

then

$$\lim_{k\to\infty} \langle \nabla f(y^k), y^k - x\rangle = +\infty.$$

A simple sufficient condition for **H1** is that the image of ∇f is the whole space R^n (see Burachik and Iusem 1998c, Proposition 3). Assumption **H2** is called *boundary coerciveness* and it is the key concept in the context of proximal point methods for constrained problems for the following reason. It is clear from Definition 2.1 that if f is a Bregman function with zone int(C) and P is any open subset of int(C), then f is also a Bregman function with zone P, which means that one cannot recover C from f. Therefore in order to use the Bregman distance D_f for penalization purposes, f has to possess an additional property. In particular, f should contain information about C. This is precisely the role of **H2** because it implies divergence of ∇f on bdry(C), which makes C defined by f:

$$\mathrm{Dom}(\nabla f) = \mathrm{int}(C). \tag{4}$$

Divergence of ∇f also implies that the proximal subproblems cannot have solutions on the boundary of C. We refer the readers to Censor, Iusem and Zenios (1998) and Burachik and Iusem (1998) for further details on boundary coercive Bregman functions. Note also that boundary coerciveness is equivalent to f being *essentially smooth* on int(C) (Bausche and Borwein (1997), Theorem 4.5(i)).

It is further worth noting that if the domain of ∇f is the interior of C, and the image of ∇f is R^n, then **H1** and **H2** hold automatically (see Burachik and Iusem 1998, Proposition 3 and Censor, Iusem and Zenios 1998, Proposition 7).

We are now ready to describe our error tolerance criterion. Take any $x \in \operatorname{int}(C)$ and $c > 0$, and consider the proximal subproblem

$$0 \in cT(\cdot) + \nabla f(\cdot) - \nabla f(x), \tag{5}$$

which is to find a pair (y, v) satisfying the *proximal system*

$$v \in T(y), \qquad cv + \nabla f(y) - \nabla f(x) = 0. \tag{6}$$

The latter is in turn equivalent to

$$v \in T(y), \qquad y = (\nabla f)^{-1}(\nabla f(x) - cv). \tag{7}$$

Therefore, an approximate solution of (5) (or (6) or (7)) should satisfy

$$v \in T(y), \qquad y \approx (\nabla f)^{-1}(\nabla f(x) - cv). \tag{8}$$

We next formally define the concept of inexact solutions of (6), taking the approach of (8).

Definition 3.1. Let $x \in \operatorname{int}(C)$, $c > 0$ and $\sigma \in [0, 1)$. We say that a pair (y, v) is an inexact solution with tolerance σ of the proximal subproblem (6) if

$$v \in T(y)$$

and z, the solution of equation

$$cv + \nabla f(z) - \nabla f(x) = 0,$$

satisfies

$$D_f(y, z) \leq \sigma^2 D_f(y, x).$$

From (4) (which is a consequence of **H2**), it follows that

$$z \in \operatorname{int}(C).$$

Note that equivalently z is given by

$$z = \nabla f^{-1}(\nabla f(x) - cv).$$

Therefore z, and hence $D_f(y, z)$, are easily computable from x, y and v whenever ∇f is explicitly invertible. In that case it is trivial to check whether a given pair (y, v) is an admissible approximate solution in the sense of Definition 3.1: it is enough to obtain $z = \nabla f^{-1}(\nabla f(x) - cv)$ and verify if $D_f(y, z) \leq \sigma^2 D_f(y, x)$. Since our algorithm is based on this test, it is most easy to implement when ∇f is explicitly invertible. We point out that this case covers a wide range of important applications. For example, Bregman functions with this property are readily available when the feasible set C is an orthant, a polyhedron, a box, or a ball (see Censor, Iusem and Zenios 1998).

Another important observation is that for $\sigma = 0$, we have that $y = z$. Hence, the only point which satisfies Definition 3.1 for $\sigma = 0$, is precisely the exact solution of the proximal subproblem. Therefore our view of inexact solution of generalized proximal subproblems is quite natural. We note, in passing, that it is motivated by the approach developed in Solodov and Svaiter (1999b, c) for the classical ("linear") proximal point method.

In that case, Definition 3.1 (albeit slightly modified) is equivalent to saying that the subproblems are solved within *fixed relative error* tolerance (see also Solodov and Svaiter 2000c). Such an approach seems to be computationally more realistic/constructive than the common summable-error-type requirements.

Regarding the existence of inexact solutions, the situation is clearly even easier than for exact methods. Since we are supposing that the generalized proximal problem (5) always has an exact solution in $\text{int}(C)$, this problem will certainly always have (possibly many) *inexact* solutions (y, v) satisfying also $y \in C$.

Now we can formally state our inexact generalized proximal method.

Algorithm 1. Inexact Generalized Proximal Method.

Initialization: *Choose some* $c > 0$, *and the error tolerance parameter* $\sigma \in [0, 1)$. *Choose some* $x^0 \in \text{int}(C)$. *Set* $k := 0$.

Iteration k**:** *Choose the regularization parameter* $c_k \geq c$, *and find* (y^k, v^k), *an inexact solution with tolerance* σ *of*

$$0 \in c_k T(\cdot) + \nabla f(\cdot) - \nabla f(x^k), \tag{9}$$

satisfying

$$y^k \in C. \tag{10}$$

Define

$$x^{k+1} = \nabla f^{-1}(\nabla f(x^k) - c_k v^k). \tag{11}$$

Set $k := k + 1$; *and repeat.*

We have already discussed the possibility of solving inexactly (9) with condition (10). Another important observation is that since for $\sigma = 0$ inexact subproblem solution coincides with the exact one, in that case Algorithm 1 produces the same iterates as the standard exact generalized proximal method. Hence, all our convergence results (some of them are new!) apply also to the exact method. For $\sigma \neq 0$ however, there is no direct relation between the iterates of Algorithm 1 and

$$e^{k+1} + \nabla f(x^k) \in c_k T(x^{k+1}) + \nabla f(x^{k+1}),$$

considered in Eckstein (1998). The advantage of our approach is that it allows an attractive constructive stopping criterion (given by Definition 3.1) for approximate solution of subproblems (at least, when ∇f is invertible).

Under our hypothesis, Algorithm 1 is well defined. From now on, $\{x^k\}$ and $\{(y^k, v^k)\}$ are sequences generated by Algorithm 1. Therefore, by the construction of Algorithm 1 and by Definition 3.1, for all k it holds that

$$y^k \in C, \qquad x^k \in \text{int}(C), \tag{12}$$

$$v^k \in T(y^k), \tag{13}$$

$$c_k v^k + \nabla f(x^{k+1}) - \nabla f(x^k) = 0, \tag{14}$$

$$D_f(y^k, x^{k+1}) \leq \sigma^2 D_f(y^k, x^k). \tag{15}$$

We now state our main convergence result. First, recall that a maximal monotone operator T is *paramonotone* (Bruck 1975, 1976; see also Censor, Iusem and Zenios 1998, Iusem 1998b) if $u \in T(x), v \in T(y)$ and $\langle u - v, x - y \rangle = 0 \Rightarrow u \in T(y)$ and $v \in T(x)$.

Some examples of paramonotone operators are subdifferentials of proper closed convex functions, and strictly monotone maximal monotone operators.

THEOREM 3.2. *Suppose that $VIP(T,C)$ has solutions and one of the following two conditions holds.*
(1) $X^* \cap \operatorname{int}(C) \neq \emptyset$.
(2) *T is paramonotone.*
Then the sequence $\{x^k\}$ converges to a solution of $VIP(T,C)$.

Thus we establish convergence of our inexact algorithm under assumptions which are even weaker than the ones that have been used, until now, for exact algorithms. Specifically, in the paramonotone case, we get rid of the "pseudomonotonicity" assumption on T (Burachik and Iusem 1998) which can be stated as follows:

> *Take any sequence $\{y^k\} \subset Dom(T)$ converging to y and any sequence $\{v^k\}$, $v^k \in T(y^k)$. Then for each $x \in Dom(T)$ there exists an element $v \in T(y)$ such that*

$$\langle v, y-x\rangle \leq \liminf_{k\to\infty} \langle v^k, y^k - x\rangle.$$

Until now, this (or some other, related) technical assumption was employed in the analysis of all generalized proximal methods (e.g., Eckstein 1998, Burachik and Iusem 1998, Burachik, Iusem and Svaiter 1997). Among other things, this resulted in splitting the proof of convergence for the case of minimization and for paramonotone operators (the subdifferential of a convex function is paramonotone, but it need not satisfy the above condition). And of course, the additional requirement of pseudomonotonicity makes the convergence result for paramonotone operators weaker. Since for the tolerance parameter $\sigma = 0$ our Algorithm 1 reduces to the exact generalized proximal method, Theorem 3.2 also constitutes a new convergence result for the standard setting of exact proximal algorithms. We note that the stronger than convergence consistency property of Bregman functions established in this paper is crucial for obtaining this new result.

To obtain this stronger result, the proof will be somewhat more involved than the usual, and some auxiliary analysis will be needed. However, we think that this is worthwhile since it allows us to remove some (rather awkward) additional assumptions.

4. Convergence analysis. Given sequences $\{x^k\}$, $\{y^k\}$ and $\{v^k\}$ generated by Algorithm 1, define $\tilde{X}$ as all points $x \in C$ for which the index set

$$\{k \in N \mid \langle v^k, y^k - x\rangle < 0\}$$

is finite. For $x \in \tilde{X}$, define $k(x)$ as the smallest integer such that

$$k \geq k(x) \Rightarrow \langle v^k, y^k - x\rangle \geq 0.$$

Of course, the set $\tilde{X}$ and the application $k(\cdot)$ depend on the particular sequences generated by the algorithm. These definitions will facilitate the subsequent analysis. Note that, by monotonicity of T,

$$X^* \subseteq \tilde{X};$$

and in fact,

$$k(x) = 0 \quad \text{for any } x \in X^*.$$

LEMMA 4.1. *For any* $s \in \tilde{X}$ *and* $k \geq k(s)$, *it holds that*

$$D_f(s,x^{k+1}) \leq D_f(s,x^k) - c^k\langle v^k, y^k - s\rangle - (1-\sigma^2)D_f(y^k,x^k) \leq D_f(s,x^k). \tag{16}$$

PROOF. Take $s \in \tilde{X}$ and $k \geq k(s)$. Using Lemma 2.6, we get

$$D_f(s,x^{k+1}) = D_f(s,x^k) + \langle \nabla f(x^k) - \nabla f(x^{k+1}), s - y^k\rangle + D_f(y^k,x^{k+1}) - D_f(y^k,x^k).$$

By (14) and (15), we further obtain

$$D_f(s,x^{k+1}) \leq D_f(s,x^k) - c^k\langle v^k, y^k - s\rangle - (1-\sigma^2)D_f(y^k,x^k),$$

which proves the first inequality in (16). Since the Bregman distance is always nonnegative and $\sigma \in [0,1)$, we have

$$D_f(s,x^{k+1}) \leq D_f(s,x^k) - c^k\langle v^k, y^k - s\rangle.$$

The last inequality in (16) follows directly from the hypothesis $s \in \tilde{X}$, $k \geq k(s)$ and the respective definitions. □

As an immediate consequence, we obtain that the sequence $\{D_f(\bar{x},x^k)\}$ is decreasing for any $\bar{x} \in X^*$.

COROLLARY 4.2. *If the sequence* $\{x^k\}$ *has an accumulation point* $\bar{x} \in X^*$ *then the whole sequence converges to* $\bar{x}$.

PROOF. Suppose that some subsequence $\{x^{k_j}\}$ converges to $\bar{x} \in X^*$. Using Definition 2.1 (item (4)), we conclude that

$$\lim_{j\to\infty} D_f(\bar{x},x^{k_j}) = 0.$$

Since the whole sequence $\{D_f(\bar{x},x^k)\}$ is decreasing and it has a convergent subsequence, it follows that it converges:

$$\lim_{k\to\infty} D_f(\bar{x},x^k) = 0.$$

Now the desired result follows from Theorem 2.4. □

COROLLARY 4.3. *Suppose that* $\tilde{X} \neq \emptyset$. *Then the following statements hold*:

(1) *The sequence* $\{x^k\}$ *is bounded*;
(2) $\sum_{k=0}^{\infty} D_f(y^k,x^k) < \infty$;
(3) *For any* $s \in \tilde{X}$, $\sum_{k=0}^{\infty}\langle v^k, y^k - s\rangle < \infty$;
(4) *The sequence* $\{y^k\}$ *is bounded.*

PROOF. Take some $s \in \tilde{X}$. From Lemma 4.1 it follows that for all k greater than $k(s)$, $D_f(s,x^k) \leq D_f(s,x^{k(s)})$. Therefore, $D_f(s,x^k)$ is bounded and from Definition 2.1 (item (3)), it follows that $\{x^k\}$ is bounded.

By Lemma 4.1, it follows that for any $r \in N$,

$$D_f(s,x^{k(s)+r+1}) \leq D_f(s,x^{k(s)}) - \sum_{k=k(s)}^{k(s)+r} (c_k\langle v^k, y^k - s\rangle + (1-\sigma^2)D_f(y^k,x^k)).$$

Therefore

$$\sum_{k=k(s)}^{k(s)+r} c_k \langle v^k, y^k - s\rangle + (1-\sigma^2) \sum_{k=k(s)}^{k(s)+r} D_f(y^k, x^k) \le D_f(s, x^{k(s)}).$$

Since r is arbitrary and the terms of both summations are nonnegative (recall the definition of $k(s)$), it follows that we can take the limit as $r \to \infty$ in both sides of the latter relation. Taking further into account that $\{c_k\}$ is bounded away from zero, the second and third assertions of the Corollary easily follow. As consequences, we also obtain that

$$\lim_{k\to\infty} D_f(y^k, x^k) = 0 \tag{17}$$

and

$$\lim_{k\to\infty} \langle v^k, y^k - s\rangle = 0, \qquad \forall s \in \tilde{X}. \tag{18}$$

Suppose now that $\{y^k\}$ is unbounded. Then there exists a pair of subsequences $\{x^{k_j}\}$ and $\{y^{k_j}\}$ such that $\{x^{k_j}\}$ converges but $\{y^{k_j}\}$ diverges. However, by (17) and Theorem 2.4, $\{y^{k_j}\}$ must converge (to the same limit as $\{x^{k_j}\}$), which contradicts the assumption. Hence, $\{y^k\}$ is bounded. □

The next proposition establishes the first part of Theorem 3.2, namely the convergence of the inexact generalized proximal algorithm in the case when $X^* \cap \operatorname{int}(C) \neq \emptyset$.

Proposition 4.4. *If $\tilde{X} \cap int(C) \neq \emptyset$ then $\{x^k\}$ converges to some $\bar{x} \in int(C)$ which is a solution of $VIP(T, C)$.*

Proof. By Corollary 4.3, it follows that $\{x^k\}$ is bounded, so it has some accumulation point $\bar{x} \in C$, and for some subsequence $\{x^{k_j}\}$,

$$\lim_{j\to\infty} x^{k_j} = \bar{x}.$$

Take any $\hat{x} \in \tilde{X} \cap \operatorname{int}(C)$. Suppose that $\bar{x} \in bdry(C)$. Since

$$D_f(\hat{x}, x^{k_j}) = f(\hat{x}) - f(x^{k_j}) - \langle \nabla f(x^{k_j}), \hat{x} - x^{k_j}\rangle$$

and, by **H2**,

$$\lim_{j\to\infty} \langle \nabla f(x^{k_j}), x^{k_j} - \hat{x}\rangle = +\infty,$$

it follows that

$$\lim_{j\to\infty} D_f(\hat{x}, x^{k_j}) = \infty.$$

But the latter is impossible because $D_f(\hat{x}, x^k)$ is a decreasing sequence, at least for $k \ge k(\hat{x})$ (by Lemma 4.1). Hence,

$$\bar{x} \in \operatorname{int}(C).$$

Next, we prove that $\bar{x}$ is a solution of VIP(T, C). By (17), we have that

$$\lim_{j\to\infty} D_f(y^{k_j}, x^{k_j}) = 0. \tag{19}$$

Because, by (15), $D_f(y^{k_j}, x^{k_j+1}) \le \sigma^2 D_f(y^{k_j}, x^{k_j})$, it follows that

$$\lim_{j\to\infty} D_f(y^{k_j}, x^{k_j+1}) = 0. \tag{20}$$

Since $\{x^{k_j}\}$ converges to $\bar{x}$, Theorem 2.4 and (19) imply that $\{y^{k_j}\}$ also converges to $\bar{x}$. Applying Theorem 2.4 once again, this time with (20), we conclude that $\{x^{k_j+1}\}$ also converges to $\bar{x}$. Since $\bar{x}\in \text{int}(C)$ and ∇f is continuous in $\text{int}(C)$, we therefore conclude that

$$\lim_{j\to\infty} \nabla f(x^{k_j+1}) - \nabla f(x^{k_j}) = 0.$$

Since $c^k \geq c > 0$, using (14) we get

$$\lim_{j\to\infty} v^{k_j} = 0,$$

where $v^{k_j}\in T(y^{k_j})$. Now the fact that $\{y^{k_j}\}\to\bar{x}$, together with the maximality of T, implies that $0\in T(\bar{x})$. Thus we have a subsequence $\{x^{k_j}\}$ converging to $\bar{x}\in X^*$. By Corollary 4.2, the whole sequence $\{x^k\}$ converges to $\bar{x}$. □

We proceed to analyze the case when T is paramonotone. By (18), we already know that if $s\in X^*$ then $\langle v^k, y^k - s\rangle \to 0$. *If* we could pass onto the limit with respect to v^k (for example, using the technical assumption of pseudomonotonicity stated above), then we could conclude that $0\geq\langle\bar{v},\bar{x}-s\rangle\geq\langle v_s,\bar{x}-s\rangle\geq 0$, where $\bar{x}$ is an accumulation point of $\{x^k\}$ (hence also of $\{y^k\}$), and $\bar{v}\in T(\bar{x})$, $v_s\in T(s)$. By paramonotonicity, it follows that $v_s\in T(\bar{x})$. Now by monotonicity, we further obtain that for any $x\in C, \langle v_s, x-\bar{x}\rangle = \langle v_s, x-s\rangle + \langle v_s, s-\bar{x}\rangle = \langle v_s, x-s\rangle \geq 0$, which means that $\bar{x}\in X^*$. However, in the absence of the assumption of pseudomonotonicity one cannot use this well-established line of argument. To overcome the difficulty resulting from the impossibility of directly passing onto the limit as was done above, we shall need some auxiliary constructions.

Let A be the affine hull of the domain of T. Then there exists some V, a subspace of R^n, such that

$$A = V + x$$

for any $x\in \text{Dom}(T)$. Denote by P the orthogonal projection onto V, and for each k define

$$u^k = P_V(v^k).$$

The idea is to show the following key facts:

$$u^k \in T(y^k),$$

$$\langle u^k, y^k - x\rangle = \langle v^k, y^k - x\rangle \qquad \forall x\in \text{Dom}(T),$$

$$\{u^k\} \text{ has an accumulation point.}$$

With these facts in hand, we could pass onto the limit in a manner similar to the above, and complete the proof.

First, note that

$$\forall x\in R^n, \qquad P\circ T(x) = T(x)\cap V. \tag{21}$$

This can be verified rather easily: if $x\notin \text{Dom}(T)$ then both sets in (21) are empty, so it is enough to consider $x\in\text{Dom}(T)$. Clearly, if some $u\in T(x)\cap V$ then $u\in P\circ T(x)$ because $u\in T(x)$ and also $u\in V$. Now take some $u\in P\circ T(x)$, so that $u = P_V(v)$ and $v\in T(x)$. By monotonicity of T, for any $z\in\text{Dom}(T)$ and any $w\in T(z)$, it holds that

$$\langle v - w, x - z\rangle \geq 0.$$

Since $x - z\in V$ and $u = P_V(v)$, it holds that

$$\langle v, x - z\rangle = \langle u, x - z\rangle.$$

Therefore

$$\langle u - w, x - z \rangle \geq 0,$$

which implies that $u \in T(x)$ by the maximality of T. Since $u \in V$ also, it follows that $u \in T(x) \cap V$, which establishes (21).

Lemma 4.5. *If $\tilde{X} \cap int(C) = \emptyset$ and $\tilde{X} \neq \emptyset$, then some subsequence of $\{u^k\}$ is bounded.*

Proof. We assumed that $\mathrm{Dom}(T) \cap \mathrm{int}(C) \neq \emptyset$. Therefore $\mathrm{ri}(\mathrm{Dom}(T)) \cap \mathrm{int}(C) \neq \emptyset$. Take some $\hat{x} \in \mathrm{ri}(\mathrm{Dom}(T)) \cap \mathrm{int}(C)$. Then $A = V + \hat{x}$, and let P be the projection operator onto V discussed above. In particular, $u^k = P_V(v^k) \in P \circ T(y^k) = T(y^k) \cap V$, so that $u^k \in T(y^k)$. Furthermore, the operator $\hat{T} : V \to \mathscr{P}(V)$ defined by

$$\hat{T}(\xi) = P(T(\xi + \hat{x}))$$

is maximal monotone as an operator on the space V (this can be easily verified using the maximal monotonicity of T on R^n). We also have that $0 \in \mathrm{int}(\mathrm{dom}(\hat{T}))$. Therefore $\hat{T}$ is bounded around zero (Rockafellar 1969). So, $P \circ T$ is bounded around $\hat{x}$, i.e., there exist some $r > 0$ and $M \geq 0$ such that

$$x - \hat{x} \in V, \qquad \|x - \hat{x}\| \leq r \Rightarrow \forall w \in P \circ T(x), \qquad \|w\| \leq M.$$

But $P \circ T(x) = T(x) \cap V$. Hence,

$$x - \hat{x} \in V, \qquad \|x - \hat{x}\| \leq r \Rightarrow \exists v \in T(x), \qquad \|v\| \leq M.$$

Since $\tilde{X} \cap \mathrm{int}(C) = \emptyset$, it follows that $\hat{x} \notin \tilde{X}$. Therefore, by the definition of $\tilde{X}$, there exists an infinite subsequence of indices $\{k_j\}$ such that

$$\langle v^{k_j}, y^{k_j} - \hat{x} \rangle < 0.$$

Note that $u^{k_j} \in T(y^{k_j})$ and, since $y^{k_j} - \hat{x} \in V$, it holds that

$$\langle u^{k_j}, y^{k_j} - \hat{x} \rangle = \langle v^{k_j}, y^{k_j} - \hat{x} \rangle < 0. \tag{22}$$

Define, for each j,

$$\hat{x}^j = \hat{x} + (r/\|u^{k_j}\|) u^{k_j}.$$

Then for each j there exists $\hat{v}^j \in T(\hat{x}^j)$ such that

$$\|\hat{v}^j\| \leq M.$$

Furthermore,

$$\begin{aligned} \langle \hat{v}^j, \hat{x}^j - y^{k_j} \rangle &\geq \langle u^{k_j}, \hat{x}^j - y^{k_j} \rangle \\ &= \langle u^{k_j}, \hat{x}^j - \hat{x} \rangle + \langle u^{k_j}, \hat{x} - y^{k_j} \rangle \\ &> \langle u^{k_j}, \hat{x}^j - \hat{x} \rangle \\ &= r\|u^{k_j}\|, \end{aligned}$$

where the first inequality is by the monotonicity of T, and the second is by (22). Further, using the Cauchy-Schwarz and triangular inequalities, we obtain

$$r\|u^{k_j}\| \leq \|\hat{v}^j\|(\|\hat{x}^j - \hat{x}\| + \|\hat{x} - y^{k_j}\|) \leq M(r + \|\hat{x} - y^{k_j}\|).$$

Since the sequence $\{y^k\}$ is bounded (Corollary 4.3, item (4)), it follows that $\{||u^{k_j}||\}$ is bounded. □

We conclude the analysis by establishing the second part of Theorem 3.2.

PROPOSITION 4.6. *Suppose* $X^* \neq \emptyset$ *and* T *is paramonotone. Then* $\{x^k\}$ *converges to some* $\bar{x} \in X^*$.

PROOF. If $\tilde{X} \cap \text{int}(C) \neq \emptyset$, then the conclusion follows from Proposition 4.4. Suppose now that

$$\tilde{X} \cap \text{int}(C) = \emptyset.$$

By Lemma 4.5, it follows that some subsequence of $\{u^k\}$ is bounded. Since $X^* \subseteq \tilde{X}$, from Corollary 4.3 it follows that the whole sequence $\{x^k\}$ is bounded. Hence, there exist two subsequences $\{x^{k_j}\}$, $\{u^{k_j}\}$ which both converge:

$$\lim_{j\to\infty} x^{k_j} = \bar{x}, \qquad \lim_{j\to\infty} u^{k_j} = \bar{u}.$$

Recall from the proof of Lemma 4.5 that $u^{k_j} \in T(y^{k_j})$. By Corollary 4.3 (item (2)), we have that

$$\lim_{j\to\infty} D_f(y^{k_j}, x^{k_j}) = 0.$$

Therefore, by Theorem 2.4,

$$\lim_{j\to\infty} y^{k_j} = \bar{x},$$

and

$$\bar{u} \in T(\bar{x}),$$

by the maximality of T. Take now some $s \in X^*$. There exists some $v_s \in T(s)$ such that

$$\langle v_s, x - s\rangle \geq 0$$

for all $x \in C$. Therefore, using also the monotonicity of T,

$$0 \leq \langle v_s, \bar{x} - s\rangle \leq \langle \bar{u}, \bar{x} - s\rangle. \tag{23}$$

Note that for any $x \in \text{Dom}(T)$

$$\langle v^{k_j}, y^{k_j} - x\rangle = \langle u^{k_j}, y^{k_j} - x\rangle.$$

Taking $x = s$, and passing onto the limit as $j \to \infty$, (18) implies that

$$\langle \bar{u}, \bar{x} - s\rangle = 0.$$

Together with (23), this implies that

$$\langle v_s, \bar{x} - s\rangle = \langle \bar{u}, \bar{x} - s\rangle = 0.$$

Using now the paramonotonicity of T, we conclude that

$$v_s \in T(\bar{x}).$$

Finally, for any $x \in C$, we obtain

$$\begin{aligned}\langle v_s, x - \bar{x}\rangle &= \langle v_s, x - s\rangle + \langle v_s, s - \bar{x}\rangle \\ &= \langle v_s, x - s\rangle \\ &\geq 0.\end{aligned}$$

Therefore $\bar{x} \in X^*$. Since we have a subsequence $\{x^{k_j}\}$ converging to $\bar{x} \in X^*$, from Corollary 4.2 it follows that the whole sequence $\{x^k\}$ converges to $\bar{x}$. □

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

References

Bauschke, H. H., J. M. Borwein. 1997. Legendre functions and the method of random Bregman projections. *J. Convex Anal.* **4** 27–67.

Bregman, L. M. 1967. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. and Math. Phys.* **7**(3) 200–217.

Brézis, H., P. L. Lions. 1978. Produits infinis de résolvantes. *Israel J. Math.* **29** 329–345.

Bruck, R. E. 1975. An iterative solution of a variational inequality for certain monotone operators in a Hilbert space. *Bull. Amer. Math. Soc.* **81** 890–892.

——. 1976. Corrigendum to [4]. *Bull. Amer. Math. Soc.* **82** 353.

Burachik, R. S., A. N. Iusem. 1998. A generalized proximal point algorithm for the variational inequality problem in a Hilbert space. *SIAM J. Optim.* **8** 197–216.

——, ——, B. F. Svaiter. 1997. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.* **5** 159–180.

Burke, J. V., M. Qian. 1998. A variable metric proximal point algorithm for monotone operators. *SIAM J. Control Optim.* **37** 353–375.

Censor, Y., A. N. Iusem, S. A. Zenios. 1998. An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. *Math. Programming* **81** 373–400.

——, S. Zenios. 1992. The proximal minimization algorithm with *D*-functions. *J. Optim. Theory Appl.* **73** 451–464.

Chen, G., M. Teboulle. 1993. Convergence analysis of proximal-like optimization algorithm using Bregman functions. *SIAM J. Optim.* **3** 538–543.

Cottle, R. W., F. Giannessi, J.-L. Lions. 1980. *Variational Inequalities and Complementarity Problems*: *Theory and Applications.* Wiley, New York.

Eckstein, J. 1993. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Math. Oper. Res.* **18** 202–226.

——. 1998. Approximate iterations in Bregman-function-based proximal algorithms. *Math. Programming* **83** 113–123.

——, D. P. Bertsekas. 1992. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathe. Programming* **55** 293–318.

Eggermont, P. P. B. 1990. Multiplicative iterative algorithms for convex programming. *Linear Algebra Appl.* **130** 25–32.

Ferris, M. C. 1991. Finite termination of the proximal point algorithm. *Math. Programming* **50** 359–366.

Güler, O. 1992. New proximal point algorithms for convex minimization. *SIAM J. Optim.* **2** 649–664.

Iusem, A. N. 1995. On some properties of generalized proximal point methods for quadratic and linear programming. *J. Optim. Theory Appl.* **85** 596–612.

——. 1998a. On some properties of generalized proximal point methods for the variational inequality problem. *J. Optim. Theory Appl.* **96** 337–362.

——. 1998b. On some properties of paramonotone operators. *J. Convex Anal.* **5** 269–278.

——, M. V. Solodov. 1997. Newton-type methods with generalized distances for constrained optimization. *Optim.* **41** 257–278.

Kiwiel, K. C. 1997. Proximal minimization methods with generalized Bregman functions. *SIAM J. Control Optim.* **35** 1142–1168.

Lemaire, B. 1989. The proximal algorithm. J. P. Penot, ed. *New Methods of Optimization and Their Industrial Use. International Series of Numerical Mathematics* **87** Birkhauser, Basel, Switzerland. 73–87.

Luque, F. J. 1984. Asymptotic convergence analysis of the proximal point algorithm. *SIAM J. Control Optim.* **22** 277–293.

Martinet, B. 1970. Regularisation d'inequations variationelles par approximations successives. *Rev. Française Inform. Rech. Opér.* **4** 154–159.

Moreau, J.-J. 1965. Proximité et dualité dans un espace Hilbertien. *Bull. Soc. Math. France* **93** 273–299.

Pang, J.-S. 1995. Complementarity problems. R. Horst and P. Pardalos, eds. *Handbook of Global Optimization.* Kluwer Academic Publishers, Boston, MA 271–338.

Passty, G. B. 1979. Weak convergence theorems for nonexpansive mappings in Banach spaces. *J. Math. Anal. Appl.* **67** 274–276.

Rockafellar, R. T. 1969. Local boundedness of nonlinear monotone operators. *Michigan Math. J.* **16** 397–407.

——. 1970a. On the maximality of sums of nonlinear monotone operators. *Trans. Amer. Math. Soc.* **149** 75–88.

——. 1970b. *Convex Analysis.* Princeton University Press, Princeton, NJ.

——. 1976a. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1** 97–116.

——. 1976b. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14** 877–898.

Solodov, M. V., B. F. Svaiter. 2000a. A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. *SIAM J. Optim.* **10** 605–625.

——, ——. 2000b. A comparison of rates of convergence of two inexact proximal point algorithms. G. Di Pillo and F. Giannessi, eds. *Nonlinear Optimization and Related Topics*. Kluwer Academic Publishers. Forthcoming.

——, ——. 2000c. Forcing strong convergence of proximal point iterations in a Hilbert space. *Math. Programming.* **87** 189–202.

——, ——. 1999a. A globally convergent inexact Newton method for systems of monotone equations. M. Fukushima and L. Qi, eds. *Reformulation—Nonsmooth*, *Piecewise Smooth*, *Semismooth and Smoothing Methods*. Kluwer Academic Publishers, Dordrecht, The Netherlands. 355–369.

——, ——. 1999c. A hybrid approximate extragradient—proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.* **7** 323–345.

——, ——. 1999b. A hybrid projection—proximal point algorithm. *J. Convex Anal.* **6**(1) 59–70.

Teboulle, M. 1997. Convergence of proximal-like algorithms. *SIAM J. Optim.* **7** 1069–1083.

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

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