Nonlinear Optimization and Related Topics
G. Di Pillo and F. Giannessi, Editors
Applied Optimization 36

A Comparison of Rates of Convergence of Two Inexact Proximal Point Algorithms

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

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

Abstract

We compare the linear rate of convergence estimates for two inexact proximal point methods. The first one is the classical inexact scheme introduced by Rockafellar, for which we obtain a slightly better estimate than the one given in [16]. The second one is the hybrid inexact proximal point approach introduced in [25, 22]. The advantage of the hybrid methods is that they use more constructive and less restrictive tolerance criteria in inexact solution of subproblems, while preserving all the favorable properties of the classical method, including global convergence and local linear rate of convergence under standard assumptions. In this paper, we obtain a linear convergence estimate for the hybrid algorithm [22], which is better than the one for the classical method [16], even if our improved estimate is used for the latter.

Keywords: Maximal monotone operator, proximal point method, linear convergence.

1 Introduction

We consider the classical problem of finding a zero of a maximal monotone operator (or a multifunction) T on a real Hilbert space $\mathcal{H}$:

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in T(x)\,. \tag{1}$$

This is a well known problem which includes, as special cases, optimization and min-max problems, complementarity problems, and variational inequalities.

One of the fundamental approaches to solving (1) is the proximal point method, which generates the next iterate x^{k+1} by solving the subproblem

$$0 \in c_k T(x) + (x - x^k)\,, \tag{2}$$

where x^k is the current iterate, and $c_k > 0$ is a regularization parameter. The literature on this subject is vast and includes, but is not limited to, [15, 16, 14, 13, 12, 8, 9, 7] (see [11] for a survey).

For developing implementable computational techniques, of particular importance is treating the case when subproblems are solved only approximately. This leads to the following scheme:

$$\begin{array}{l}\text{find } x^{k+1} \in \mathcal{H} \text{ such that}\\ v^{k+1} \in T(x^{k+1}), \quad 0 = c_k v^{k+1} + (x^{k+1} - x^k) - r^k,\end{array} \tag{3}$$

where $r^k \in \mathcal{H}$ is an error associated with inexact solution of subproblem (2). The following tolerance criteria imposed on error terms r^k were first introduced in [16], and they (or their variations) are standard in the literature (see, for example, [1, 26, 4, 6, 5]):

$$\|r^k\| \le \delta_k\,, \quad \sum_{k=0}^{\infty} \delta_k < \infty\,, \tag{4}$$

and

$$\|r^k\| \le \delta_k \|x^{k+1} - x^k\|\,, \quad \sum_{k=0}^{\infty} \delta_k < \infty\,. \tag{5}$$

The first condition is typically needed to establish global convergence of the proximal point algorithm, while the second is required for local linear rate of convergence result under some additional natural assumptions.

Recently, some new proximal-based algorithms have been proposed [25, 22], which use improved error criteria. The method in [25] is based on projection ideas (related, for example, to [18]). It works as follows. A pair y^k, v^k is considered an admissible approximate solution of proximal point subproblem (2) if

$$v^k \in T(y^k)\,, \quad 0 = c_k v^k + (y^k - x^k) - r^k\,,$$

and the error r^k satisfies

$$\|r^k\| \le \delta \max\{c_k\|v^k\|, \|y^k - x^k\|\}, \tag{6}$$

where δ is a *fixed* scalar in $[0, 1)$. The next iterate x^{k+1} is obtained by projecting x^k onto the hyperplane

$$H_k = \{z \in \mathcal{H} \mid \langle v^k, z - y^k\rangle = 0\},$$

which can be shown to strictly separate x^k from any zero of T, whenever (6) is satisfied. This projection step is the key for ensuring global convergence of the method, i.e., without it, the inexact proximal method managed by (6) may fail (see an example in [25]). It is important to note that projection onto a hyperplane is explicit, and thus its cost is computationally negligible.

Another method, introduced in [22], is based on the extragradient ideas (related, for example, to [10, 19]). It works as follows:

$$x^{k+1} = x^k - c_k v^k,$$

where

$$v^k \in T^{\varepsilon_k}(y^k), \quad 0 = c_k v^k + (y^k - x^k) - r^k.$$

Here, T^{ε_k} is the ε_k-enlargement (outer approximation) of T introduced by Burachik, Iusem and Svaiter in [2] (see also [3]), and the error tolerance criterion is

$$\|r^k\|^2 + 2c_k\varepsilon_k \le \delta^2\|y^k - x^k\|^2, \tag{7}$$

where again the tolerance parameter $\delta \in [0, 1)$ can be fixed. In this method, inexact solutions of subproblems are treated both, through the explicit error term r^k, and through T^{ε_k}, a current approximation of T. If $v^k \in T(y^k)$ then $\varepsilon_k = 0$ and the above error tolerance criterion reduces to

$$\|r^k\| \le \delta\|y^k - x^k\|,$$

which is somewhat stronger than (6).

Both hybrid methods outlined above preserve all the desirable convergence properties of the exact proximal algorithm (or its inexact version given by (3) with error criteria (4), (5)), including global convergence and local linear rate under standard assumptions. The advantage of the new error criteria (6) or (7) is that they are constructive, and the relaxation parameter δ need not tend to zero to ensure convergence (for example, it can be fixed). Another way of looking at it, is to observe that (6) is equivalent to saying that the bound for the *relative error* in solving the proximal subproblem can be fixed (and need not tend to zero), which is computationally realistic. This compares favorably with the conditions in (4), (5). This hybrid strategy has proved to be useful in many applications (see [20, 21, 17, 23, 24]), among which is the design of truly globally convergent inexact Newton methods.

In the case when the proximal subproblems are solved without error, the hybrid (projection or extragradient) methods coincide with the classical exact proximal point method.

While it has been established in [25, 22] that the hybrid methods possess the linear convergence property under the same assumptions as the classical exact/inexact schemes of [16], the issue of which of the methods converges faster has not been addressed yet. This question has been posed to the authors several times since the methods were first proposed. In this paper, we give the answer by comparing the linear rate of convergence estimates of the method from [16] with a similar estimate of the hybrid extragradient-proximal point method from [22] described above. We note that this comparison is not obvious if one is simply to look at the results in [16] and [22]. The paper is organized as follows. In Section 2, we give a linear convergence rate result for the classical method. In fact, by slightly modifying the argument, we obtain a somewhat improved linear rate estimate than the one given in [16]. In Section 3, we refine the previous analysis in [22] for the hybrid extragradient-proximal point method, and demonstrate that its linear rate of convergence is actually faster than for the classical method, even if the improved result from Section 2 is used for the latter.

2 Linear convergence of the classical method

For the purposes of this section, we assume that the sequence generated by the inexact proximal point method converges (strongly), and that T^{-1} is Lipschitz continuous around zero, i.e., problem (1) has the unique solution, say x^*, and there exist some constants $a > 0$ and $\tau > 0$ such that

$$\|v\| \leq \tau, v \in T(y) \Rightarrow \|y - x^*\| \leq a\|v\|. \tag{8}$$

Specifically, we consider a sequence satisfying the following relations:

$$\begin{array}{c} 0 = c_k v^{k+1} + (x^{k+1} - x^k) - r^k\,, \quad v^{k+1} \in T(x^{k+1})\,, \\ \|r^k\| \leq \delta_k \|x^{k+1} - x^k\|\,, \end{array} \tag{9}$$

where the regularization parameters sequence $\{c_k\}$ is bounded away from zero. As it has been shown in [16, Theorem 2], the following linear convergence rate estimate holds (for iteration indices k sufficiently large):

$$\|x^{k+1} - x^*\| \leq \theta_k \|x^k - x^*\|\,, \tag{10}$$

where

$$\theta_k = \frac{\mu_k + \delta_k}{1 - \delta_k}\,, \tag{11}$$

with

$$1 > \mu_k = \frac{a}{\sqrt{a^2 + c_k^2}}\,. \tag{12}$$

Clearly, once δ_k is sufficiently small, it holds that $\theta_k < 1$, and so (10) implies linear rate of convergence. Note also that since for the purposes of this study we assume that $\{x^k\}$ converges to the solution, the requirement that $\sum_{k=0}^{\infty} \delta_k < \infty$ is no longer needed, and we can relax it by merely saying that δ_k is small enough to ensure that $\theta_k < 1$.

In what follows, by slightly modifying the analysis in [16], we obtain a better linear convergence estimate, where θ_k in (10) is replaced by

$$\hat{\theta}_k = \mu_k + \frac{\delta_k}{1-\delta_k} \quad (\leq \theta_k). \tag{13}$$

Observe that $\hat{\theta}_k < \theta_k$ whenever $\delta_k \neq 0$, i.e., the two estimates are the same only for the exact method.

Theorem 1 *Suppose a sequence $\{x^k\}$ defined by (9) converges to x^* satisfying condition (8). Then for all k sufficinetly large, it holds that*

$$\|x^{k+1} - x^*\| \leq \hat{\theta}_k \|x^k - x^*\|,$$

where $\hat{\theta}_k$ is given by (13).

Proof. Let s^k, $w^k \in T(s^k)$ be the exact solution of the k-th proximal point subproblem, that is

$$c_k w_k + s^k - x^k = 0.$$

Recall ([16, Proposition 3]) that the condition $\|r^k\| \leq \delta_k \|x^{k+1} - x^k\|$ implies that

$$\|x^{k+1} - s^k\| \leq \delta_k \|x^{k+1} - x^k\|. \tag{14}$$

Therefore

$$\|s^k - x^k\| \leq \|s^k - x^{k+1}\| + \|x^{k+1} - x^k\| \leq (1+\delta_k)\|x^{k+1} - x^k\|,$$

which, using the convergence of $\{x^k\}$, implies that $\|x^k - s^k\| \to 0$. Because $\{c_k\}$ is bounded away from zero, we further conclude that $w_k = (1/c_k)(x^k - s^k) \to 0$. Using (8), we have that for indices k sufficiently large

$$\|s^k - x^*\| \leq a\|w^k\| = (a/c_k)\|s^k - x^k\|. \tag{15}$$

We further obtain

$$\begin{aligned} \|x^k - x^*\|^2 &= \|x^k - s^k\|^2 + \|s^k - x^*\|^2 + 2\langle x^k - s^k, s^k - x^* \rangle \\ &= \|x^k - s^k\|^2 + \|s^k - x^*\|^2 + 2c_k\langle w^k, s^k - x^* \rangle \\ &\geq \|x^k - s^k\|^2 + \|s^k - x^*\|^2 \\ &\geq (1 + (c_k/a)^2)\|s^k - x^*\|^2, \end{aligned} \tag{16}$$

where the first inequality follows from the monotonicity of T, and the last follows from (15). Hence,

$$\|s^k - x^*\| \leq \mu_k \|x^k - x^*\|,$$

where μ_k is defined by (12). Using the latter relation and (14), we further obtain

$$\begin{aligned} \|x^{k+1} - x^*\| &\leq \|x^{k+1} - s^k\| + \|s^k - x^*\| \\ &\leq \delta_k \|x^{k+1} - x^k\| + \mu_k \|x^k - x^*\|. \end{aligned} \tag{17}$$

Similarly,

$$\begin{aligned} \|x^{k+1} - x^k\| &\leq \|x^{k+1} - s^k\| + \|s^k - x^k\| \\ &\leq \delta_k \|x^{k+1} - x^k\| + \|x^k - x^*\|, \end{aligned}$$

where also (16) was used in the last inequality. Therefore,

$$\|x^{k+1} - x^k\| \leq (1 - \delta_k)^{-1} \|x^k - x^*\|.$$

Now, combining the latter relation with (17), we obtain the desired result:

$$\|x^{k+1} - x^*\| \leq \left(\mu_k + \frac{\delta_k}{1 - \delta_k}\right) \|x^k - x^*\|.$$

The proof is complete. □

Remark. The difference between the analysis given above and [16, Theorem 2] is basically in the bound for $\|x^{k+1} - x^k\|$ once (17) is obtained. In [16, Theorem 2], the bound

$$\|x^{k+1} - x^k\| \leq \|x^{k+1} - x^*\| + \|x^* - x^k\|$$

is used, which in the end results in a weaker linear rate estimate.

3 Linear convergence of the hybrid method

In this section, we consider the extragradient-proximal point method of [22]. We derive a convergence rate estimate which is strictly better than the one for the classical inexact proximal algorithm.

In [2], Burachik, Iusem and Svaiter defined $T^\varepsilon(x)$, an ε-enlargement of a monotone operator T, as

$$T^\varepsilon(x) := \{v \in \mathcal{H} \mid \langle u - v, y - x \rangle \geq -\varepsilon \quad \forall\, y, u \in T(y)\}, \tag{18}$$

where $\varepsilon \geq 0$. Since T is assumed to be maximal monotone,

$$T^0(x) = T(x)$$

for any x. Furthermore, directly from the definition it follows that

$$0 \leq \varepsilon_1 \leq \varepsilon_2 \Rightarrow T^{\varepsilon_1}(x) \subseteq T^{\varepsilon_2}(x).$$

Thus T^ε is an *enlargement* of T. The use of elements in T^ε instead of T allows an extra degree of freedom, which is very usefull in various applications (see [2, 3, 24]). On the other hand, setting $\varepsilon = 0$ one retrieves the original operator T, so that the classical methods can be also treated.

We now turn our attention to the hybrid proximal-based method of [22]:

$$\begin{gathered} x^{k+1} = x^k - c_k v^k \,, \\ v^k \in T^{\varepsilon_k}(y^k)\,, \quad 0 = c_k v^k + (y^k - x^k) - r^k \,, \\ \sqrt{\|r^k\|^2 + 2c_k\varepsilon_k} \leq \delta_k \|y^k - x^k\| \,. \end{gathered} \tag{19}$$

As in Section 2, we study the case when the sequence $\{x^k\}$ converges to the unique zero of T, with T^{-1} being Lipschitz continuous around zero, i.e., (8) holds. In [22], the following linear rate estimate was given :

$$\|x^{k+1} - x^*\| \leq \sqrt{1 - \frac{1 - \delta_k}{(1 + a/c_k)^2(1 + \delta_k)}} \|x^k - x^*\|.$$

This estimate appears to be not better than the one for the classical method. In this section, by refining the analysis we obtain a new sharper estimate which suggests that the hybrid algorithm attains a faster linear rate of convergence than the classical method. Specifically, we show that the factor of linear convergence for the former is

$$\tilde{\theta}_k := \left(1 + \frac{1 - \delta_k^2}{\left(a/c_k + \delta_k \sqrt{1 + (a/c_k)^2} \right)^2} \right)^{-1/2}, \tag{20}$$

which is strictly (unless the method is exact) smaller than $\hat{\theta}_k$ in Theorem 1.

Theorem 2 *Suppose a sequence $\{x^k\}$ defined by (19) converges to x^* satisfying condition (8). Then for all k sufficiently large, it holds that*

$$\|x^{k+1} - x^*\| \leq \tilde{\theta}_k \|x^k - x^*\| \,,$$

where $\tilde{\theta}_k$ is given by (20).

Proof. First observe that by the construction of the algorithm,

$$r^k = c_k v^k + y^k - x^k = y^k - x^{k+1}. \tag{21}$$

By direct computation, we obtain

$$\begin{aligned}\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 &= \|y^k - x^*\|^2 + 2\langle y^k - x^*, x^k - y^k\rangle + \|y^k - x^k\|^2 \\ &\quad - \left(\|y^k - x^*\|^2 + 2\langle y^k - x^*, x^{k+1} - y^k\rangle + \|x^{k+1} - y^k\|^2\right) \\ &= 2\langle y^k - x^*, x^k - x^{k+1}\rangle + \|y^k - x^k\|^2 - \|x^{k+1} - y^k\|^2 \\ &= 2c_k\langle v^k, y^k - x^*\rangle + \|y^k - x^k\|^2 - \|r^k\|^2.\end{aligned}$$

Since $v_k \in T^{\varepsilon_k}(y^k)$ and $0 \in T(x^*)$, from (18) it follows that

$$\langle v^k, y^k - x^*\rangle \geq -\varepsilon_k \ ,$$

so

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \geq \|y^k - x^k\| - \left(\|r^k\|^2 + 2c_k\varepsilon_k\right) \ .$$

Using further the error tolerance criterion, we obtain

$$\|x^k - x^*\|^2 \geq \|x^{k+1} - x^*\|^2 + (1 - \delta_k^2)\|y^k - x^k\|^2. \tag{22}$$

Since $\{x^k\}$ converges to x^* (strongly), and $1 > \limsup_{k\to\infty} \delta_k$, it follows that

$$0 = \lim_{k\to\infty} \|y^k - x^k\|. \tag{23}$$

Now from the error tolerance criterion we conclude that

$$0 = \lim_{k\to\infty} \|r^k\| = \lim_{k\to\infty} c_k\varepsilon_k. \tag{24}$$

As before, let s^k, $w^k \in T(s^k)$ be the exact solution of the k-th proximal point subproblem, that is

$$c_k w_k + s^k - x^k = 0.$$

Observe that

$$c_k(v^k - w^k) = s^k - y^k + r^k.$$

Since $v^k \in T^{\varepsilon_k}(y^k)$ and $w^k \in T(s^k)$, using (18) we get

$$\begin{aligned}-c_k\varepsilon_k &\leq c_k\langle v^k - w^k, y^k - s^k\rangle \\ &= \langle s^k - y^k + r^k, y^k - s^k\rangle \\ &= -\|s^k - y^k\|^2 - \langle r^k, s^k - y^k\rangle \\ &= -\|s^k - y^k + r^k/2\|^2 + \|r^k\|^2/4.\end{aligned} \tag{25}$$

Define q^k as the middle point between y^k and x^{k+1}, i.e., $q^k = (y^k + x^{k+1})/2$. In view of (21), we have

$$\begin{aligned}q^k &= y^k - r^k/2 \\ &= x^{k+1} + r^k/2.\end{aligned} \tag{26}$$

Using (25) it now follows that

$$\|s^k - q^k\|^2 \le c_k\varepsilon_k + \|r^k\|^2/4. \tag{27}$$

By the triangle inequality and (26), (27), we have that

$$\begin{aligned}
\|s^k - x^k\| &\le \|y^k - x^k\| + \|s^k - q^k\| + \|q^k - y^k\| \\
&\le \|y^k - x^k\| + \sqrt{c_k\varepsilon_k + \|r^k\|^2/4} + \|r^k\|/2.
\end{aligned}$$

Using (23) and (24), we conclude that

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

Since c_k is bounded away from zero, it also follows from $w^k = (1/c_k)(x^k - s^k)$ that

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

Thus, using (8), for k large enough it holds that

$$\|s^k - x^*\| \le a\|w^k\| = (a/c_k)\|s^k - x^k\| \ .$$

Using this inequality and the triangle inequality, we obtain

$$\begin{aligned}
\|x^{k+1} - x^*\| &\le \|x^{k+1} - s^k\| + \|s^k - x^*\| \\
&\le \|x^{k+1} - s^k\| + (a/c_k)\|s^k - x^k\| \\
&\le \|x^{k+1} - s^k\| + (a/c_k)\|s^k - y^k\| + (a/c_k)\|y^k - x^k\|. \qquad (28)
\end{aligned}$$

Note that by the Cauchy-Schwarz inequality it holds that

$$\|x^{k+1} - s^k\| + (a/c_k)\|s^k - y^k\| \le \sqrt{\|x^{k+1} - s^k\|^2 + \|s^k - y^k\|^2}\ \sqrt{1 + (a/c_k)^2}\ .$$

By (26),

$$\begin{aligned}
s^k - y^k &= (s^k - q^k) - (1/2)r^k, \\
s^k - x^{k+1} &= (s^k - q^k) + (1/2)r^k.
\end{aligned}$$

Hence

$$\begin{aligned}
\|x^{k+1} - s^k\|^2 + \|s^k - y^k\|^2 &= 2\|s^k - q^k\|^2 + (1/2)\|r^k\|^2 \\
&\le 2c_k\varepsilon_k + \|r^k\|^2 \\
&\le \delta_k^2\|y^k - x^k\|^2,
\end{aligned}$$

where (27) was used for deriving the first inequality, and the error criterion was used in the second inequality. Therefore

$$\|x^{k+1} - s^k\| + (a/c_k)\|s^k - y^k\| \le \delta_k\sqrt{1 + (a/c_k)^2}\ \ \|y^k - x^k\|,$$

which combined with (28) gives

$$\|x^{k+1} - x^*\| \leq \left[(a/c_k) + \delta_k \sqrt{1 + (a/c_k)^2}\, \right] \|y^k - x^k\|.$$

Using further (22), we get

$$\|x^k - x^*\|^2 \geq \left[1 + \frac{1 - \delta_k^2}{\left((a/c_k) + \delta_k \sqrt{1 + (a/c_k)^2}\, \right)^2} \right] \|x^{k+1} - x^*\|^2.$$

So, for k large enough,

$$\|x^{k+1} - x^*\| \leq \tilde{\theta}_k \|x^k - x^*\|,$$

where $\tilde{\theta}_k$ is defined in (20). It remains to compare $\tilde{\theta}_k$ with $\hat{\theta}_k$ given by (13). Denote

$$t_k := c_k/a\ .$$

Then we have

$$\tilde{\theta}_k = \left(1 + \frac{t_k^2 (1 - \delta_k^2)}{\left(1 + \delta_k \sqrt{1 + t_k^2}\, \right)^2} \right)^{-1/2}.$$

By direct computations,

$$1 + \frac{t_k^2 (1 - \delta_k^2)}{(1 + \delta_k \sqrt{1 + t_k^2})^2} = \left(\frac{\delta_k + \sqrt{1 + t_k^2}}{1 + \delta_k \sqrt{1 + t_k^2}} \right)^2.$$

Recall that

$$\mu_k = 1/\sqrt{1 + t_k^2} < 1.$$

Hence,

$$\begin{aligned}
\tilde{\theta}_k &= \frac{1 + \delta_k \sqrt{1 + t_k^2}}{\delta_k + \sqrt{1 + t_k^2}} \\
&= \frac{(1/\sqrt{1 + t_k^2}) + \delta_k}{(\delta_k/\sqrt{1 + t_k^2}) + 1} \\
&= \frac{\mu_k + \delta_k}{\mu_k \delta_k + 1} \\
&\leq \mu_k + \delta_k \\
&\leq \mu_k + \frac{\delta_k}{1 - \delta_k} \\
&= \hat{\theta}_k \quad (\leq \theta_k\ ).
\end{aligned}$$

□

Remark. Observe that if $\delta_k = 0$ then $\tilde{\theta}_k = \hat{\theta}_k = \theta_k$ $(= \mu_k)$. This is natural because in that case all methods considered here reduce to the classical exact proximal algorithm, and hence have the same convergence rate. However, all inequalities above are strict, i.e., $\tilde{\theta}_k < \hat{\theta}_k < \theta_k$, whenever $\delta_k \neq 0$, which is the more interesting and important case corresponding to inexact methods.

4 Concluding remarks

We obtained sharper estimates for the linear convergence rates for both the classical inexact proximal point method, and the hybrid extragradient-proximal point algorithm. According to these results, the rate for the latter is faster than for the former. However, it is important to keep in mind that these results are only theoretical estimates. As of now, we have a better estimate for the hybrid algorithm. Since it is not known whether or not the estimates are tight, strictly speaking, our results do not automatically imply that the hybrid methods converge faster locally. However, this theoretical comparison is still useful. And of course, it is possible that one may further improve the rate of convergence estimates for either method through a more intricate analysis.

Acknowledgements. We are grateful to Professor R.T. Rockafellar who posed to us the question of comparison of the rates of convergence discussed in this paper.

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

[1] A. Auslender. Numerical methods for nondifferentiable convex optimization. *Mathematical Programming Study*, 30:102–126, 1987.

[2] R.S. Burachik, A.N. Iusem, and B.F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis*, 5:159–180, 1997.

[3] R.S. Burachik, C.A. Sagastizábal, and B.F. Svaiter. ε-Enlargements of maximal monotone operators: Theory and applications. In M. Fukushima and L. Qi, edi-

tors, *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pages 25–44. Kluwer Academic Publishers, 1998.

[4] J.V. Burke and M. Qian. A variable metric proximal point algorithm for monotone operators. *SIAM Journal on Control and Optimization*, 37:353–375, 1998.

[5] R. Cominetti. Coupling the proximal point algorithm with approximation methods. *Journal of Optimization Theory and Applications*, 95:581–600, 1997.

[6] J. Eckstein. Approximate iterations in Bregman-function-based proximal algorithms. *Mathematical Programming*, 83:113–123, 1998.

[7] J. Eckstein and D.P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293–318, 1992.

[8] M.C. Ferris. Finite termination of the proximal point algorithm. *Mathematical Programming*, 50:359–366, 1991.

[9] O. Güler. New proximal point algorithms for convex minimization. *SIAM Journal on Optimization*, 2:649–664, 1992.

[10] G.M. Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

[11] B. Lemaire. The proximal algorithm. In J.P. Penot, editor, *New Methods of Optimization and Their Industrial Use. International Series of Numerical Mathematics 87*, pages 73–87. Birkhauser, Basel, 1989.

[12] F.J. Luque. Asymptotic convergence analysis of the proximal point algorithm. *SIAM Journal on Control and Optimization*, 22:277–293, 1984.

[13] B. Martinet. Regularisation d'inequations variationelles par approximations successives. *Revue Française d'Informatique et de Recherche Opérationelle*, 4:154–159, 1970.

[14] S. Reich. Weak convergence theorems for nonexpansive mappings in Banach spaces. *Journal of Mathematical Analysis and Applications*, 67:274–276, 1979.

[15] R.T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.

[16] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14:877–898, 1976.

[17] M. V. Solodov and B. F. Svaiter. A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem, 1998. *SIAM Journal on Optimization*, submitted.

[18] M. V. Solodov and B. F. Svaiter. A new projection method for variational inequality problems. *SIAM Journal on Control and Optimization*, 37:765–776, 1999.

[19] M. V. Solodov and P. Tseng. Modified projection-type methods for monotone variational inequalities. *SIAM Journal on Control and Optimization*, 34:1814–1830, 1996.

[20] M.V. Solodov and B.F. Svaiter. Forcing strong convergence of proximal point iterations in a Hilbert space, 1997. *Mathematical Programming*, submitted.

[21] M.V. Solodov and B.F. Svaiter. A globally convergent inexact Newton method for systems of monotone equations. In M. Fukushima and L. Qi, editors, *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, pages 355–369. Kluwer Academic Publishers, 1998.

[22] M.V. Solodov and B.F. Svaiter. A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator, 1998. *Set-Valued Analysis*, submitted.

[23] M.V. Solodov and B.F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, 1998. *Mathematics of Operations Research*, submitted.

[24] M.V. Solodov and B.F. Svaiter. Error bounds for proximal point subproblems and associated inexact proximal point algorithms, 1999. *Mathematical Programming*, submitted.

[25] M.V. Solodov and B.F. Svaiter. A hybrid projection – proximal point algorithm. *Journal of Convex Analysis*, 6(1), 1999.

[26] P. Tossings. The perturbed proximal point algorithm and some of its applications. *Applied Mathematics and Optimization*, 29:125–159, 1994.