Complexity of the relaxed Hybrid Proximal-Extragradient method under the large-step condition

Benar F. Svaiter[*] [†]

April 5, 2015

Abstract

In this note we review the iteration-complexity of a relaxed Hybrid-Proximal Extragradient Method under the large step condition. We also derive some useful proprieties of this method.

2000 Mathematics Subject Classification: 90C60, 90C25, 47H05.

Key words: monotone inclusion, hybrid proximal extragradient method, large step condition, complexity, ergodic convergence, relaxation, maximal monotone operator.

Introduction

In this note we review Rockafellar's Proximal Point Method and the Relaxed Hybrid Proximal-Extragradient (r-HPE) Method. We also some useful properties of the (r-HPE) and analyze its complexity under the large-step condition. All the presented results pertaining the r-HPE were essentially proved in [9]. The unique exception are the first inequalities in item 5 of Lemma 2.1 and in item 4 of Proposition 2.2.

1 Maximal monotone operators, the monotone inclusion problem, and Rockafellar's Proximal Point Method

Let H be a Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and associated norm $\|\cdot\|$. A *point-to-set operator* in H, $T : H \rightrightarrows H$, is a relation $T \subset H \times H$ and

$$T(z) = \{v \mid (z, v) \in T\}, \qquad z \in H.$$

The *inverse* of T is $T^{-1} : H \rightrightarrows H$, $T^{-1} = \{(v, z) \mid (z, v) \in T\}$. The *domain* and the range of T are, respectively,

$$D(T) = \{z \mid T(z) \neq \emptyset\}, \quad R(T) = \{v \mid \exists z \in H,\ v \in T(z)\}.$$

[*]IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil (benar@impa.br).

[†]This work was Partially supported by CNPq grants 302962/2011-5, 474996/2013-1, FAPERJ grant 201.584/2014 (Cientista d0 Nosso Estado) and by PRONEX-Optimization.

When $T(z)$ is a singleton for all $z \in D(T)$ it is usual to identify T with the map $D(T) \ni z \mapsto v \in H$ where $T(z) = \{v\}$. If $T_1, T_2 : H \rightrightarrows H$ and $\lambda \in \mathbb{R}$, then $T_1 + T_2 : H \rightrightarrows H$ and $\lambda T_1 : H \rightrightarrows H$ are defined as

$$(T_1 + T_2)(z) = \{v_1 + v_2 \mid v_1 \in T_1(z),\ v_2 \in T_2(z)\}, \qquad (\lambda T_1)(z) = \{\lambda v \mid v \in T_1(z)\}.$$

A point-to-set operator $T : H \rightrightarrows H$ is *monotone* if

$$\langle z - z', v - v' \rangle \geq 0 \qquad \forall (z, v), (z', v') \in T$$

and it is *maximal monotone* if it is a maximal element on the family of monotone point-to-set operators in H with respect to the partial order of set inclusion. Minty's theorem [3] states that if T is maximal monotone and $\lambda > 0$, then the *proximal map* $(\lambda T + I)^{-1}$ is a point-to-point non-expansive operator with domain H.

The *monotone inclusion problem* is: given $T : H \rightrightarrows H$ maximal monotone, find z such that

$$0 \in T(z). \tag{1}$$

Rockafellar's Proximal Point Method [8] (hereafter PPM) generates, for any starting $z_0 \in H$, a sequence (z_k) by the approximate rule

$$z_k \approx (\lambda_k T + I)^{-1} z_{k-1},$$

where (λ_k) is a sequence of strictly positive *stepsizes*. Rockafellar proved [8] that if (1) has a solution and

$$\left\| z_k - (\lambda_k T + I)^{-1}(z_{k-1}) \right\| \leq e_k,\ \sum_{k=1}^{\infty} e_k < \infty, \quad \inf \lambda_k > 0, \tag{2}$$

then (z_k) converges to a solution of (1).

In each step of the PPM, computation of the proximal map $(\lambda T + I)^{-1} z$ amounts to solving the *proximal (sub) problem*

$$0 \in \lambda T(z_+) + z_+ - z,$$

a regularized inclusion problem which, although well posed, is almost as hard as (1). From this fact stems the necessity of using approximations of the proximal map, for example, as prescribed in (2). Moreover, since each new iterate is, hopefully, just a better approximation to the solution than the old one, if it were compute with high accuracy, then the computational cost of each iteration would be too high (or even prohibitive) and this would impair the overall performance of the method (or even make it infeasible).

Unfortunately, prescription (2) neither tells which is the convenient error tolerance e_k to be used in the k-th iteration, nor it guarantees that the next iterate will be a better approximation than the current one.

2 Enlargements of maximal monotone operators and the Hybrid Proximal Extragradient Method

The Hybrid-Proximal Extragradient Method [10, 11] (hereafter HPE) is a modification of the PPM wherein

(a) the proximal subproblem, in each iteration, is to be solved within a *relative* error tolerance and (b) the update rule is modified so as to guarantee that the next iterate is closer to the solution set by a quantifiable amount.
An additional feature of (a) is that, in some sense, errors in the inclusion on the proximal subproblems are allowed. Recall that the ε-*enlargement* [1] of a maximal monotone operator $T : H \rightrightarrows H$ is

$$T^{[\varepsilon]}(z) = \{v \mid \langle z - z', v - v'\rangle \geq -\varepsilon \ \forall (z', v') \in T\}, \qquad x \in H,\ \varepsilon \geq 0. \tag{3}$$

From now on, $T : H \rightrightarrows H$ is a maximal monotone operator. The relaxed HPE (r-HPE) method [14] for the monotone inclusion problem (1) proceed as follows: choose $z_0 \in H$ and $\sigma \in (0, 1)$; for $k = 1, 2, \ldots$

$$\begin{array}{ll} \text{compute } \tilde{z}_k, v_k, \varepsilon_k, \lambda_k > 0 \text{ such that} & v_k \in T^{[\varepsilon_k]}(\tilde{z}_k),\quad \|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k\varepsilon_k \leq \sigma^2\|\tilde{z}_k - z_{k-1}\|^2, \\ \text{choose } t_k \in (0, 1] \text{ and set} & z_k = z_{k-1} - t_k\lambda_k v_k. \end{array} \tag{4}$$

In practical applications, each problem has a particular structure which may render feasible the computation of λ_k, $\tilde{z}_k$, v_k, and ε_k as above prescribed. For example, T may be Lipschitz continuous, it may be differentiable, or it may be a sum of an operator which has a proximal map easily computable with others with some of these properties. Prescription for computing λ_k, $\tilde{z}_k$, v_k, and ε_k under each one of these assumptions were presented in [10, 16, 13, 15, 5, 4, 6, 7].

An *exact* PPM iteration for (1) is $z_+ = (\lambda T + I)^{-1}(z)$, where z is the current iterate, z_+ is the new iterate, $\lambda > 0$ is the stepsize, and I is the identity map. Computation of such a point z_+ is equivalent to solving, in the variables v, z_+, the *proximal inclusion-equation system*:

$$v \in T(z_+),\ \lambda v + z_+ - z = 0.$$

Whence, the error criterion in (4) relaxed both the inclusion and the equality in the above inclusion-equation system. The next lemma shows that an approximate solution of the proximal inclusion-equation system satisfying that error criterion still conveys useful information for solving (1).

Lemma 2.1. *Take $z \in H$; suppose that $\lambda > 0$, $\sigma \in [0, 1)$, $t \in [0, 1]$, $\varepsilon > 0$*

$$v \in T^{[\varepsilon]}(\tilde{z}),\quad \|\lambda v + \tilde{z} - z\|^2 + 2\lambda\varepsilon \leq \sigma^2\|\tilde{z} - z\|^2; \tag{5}$$

and define $z_+ = z - t\lambda v$, $\gamma : H \to \mathbb{R}$, $\gamma(z') = \langle z' - \tilde{z}, v\rangle - \varepsilon$. Then

1. $(1 - \sigma)\|\tilde{z} - z\| \leq \|\lambda v\| \leq (1 + \sigma)\|\tilde{z} - z\|$ *and* $2\lambda\varepsilon \leq \sigma^2\|\tilde{z} - z\|$*;*
2. $z_+ = \operatorname{argmin}_{z' \in H} t\lambda\gamma(z') + \|z' - z\|^2/2$*;*
3. $\min_{z' \in H} t\lambda\gamma(z') + \dfrac{1}{2}\|z' - z\|^2 \geq \dfrac{1}{2}\Big((1 - \sigma^2)t\|\tilde{z} - z\|^2 + t(1 - t)\|\lambda v\|^2\Big)$*;*
4. *for any $z^* \in T^{-1}(0)$, $\gamma(z^*) \leq 0$ and $\|z - z^*\|^2 \geq \|z_+ - z^*\|^2 + (1 - \sigma^2)t\|\tilde{z} - z\|^2 + t(1 - t)\|\lambda v\|^2$;*
5. *for any $z^* \in T^{-1}(0)$, $\|z^* - \tilde{z}\| \leq \|z^* - z\|/\sqrt{1 - \sigma^2}$ and $\|\tilde{z} - z\| \leq \|z^* - z\|/\sqrt{1 - \sigma^2}$.*

Proof of Lemma 2.1. Since $\lambda > 0$ and $\varepsilon \geq 0$, $\|\lambda v + \tilde{z} - z\| \leq \sigma\|\tilde{z} - z\|$. Combining this inequality with triangle inequality we conclude that the two first inequalities in item 1 hold. The last inequality in item 1 follows trivially from the assumptions (5). Item 2 follows trivially from the definitions of γ and z_+.

Direct use of item 2 and of the definitions of z_+ and γ yields

$$\min_{z' \in \mathbb{R}^p} t\lambda\gamma(z') + \frac{1}{2}\|z' - z\|^2 = \frac{1}{2}\Big(t\left[\|\tilde{z} - z\|^2 - \left(\|\lambda v + \tilde{z} - z\|^2 + 2\lambda\varepsilon\right)\right] + t(1-t)\|\lambda v\|^2\Big),$$

which, combined with the inequality in (5) proves item 3.

The first inequality in item 4 follows from the inclusion $v \in T^{[\varepsilon]}(\tilde{z})$, the definition of γ, and the definition of $T^{[\varepsilon]}$ (3), with $z' = z^*$ and $v' = 0$. Since $\lambda > 0$, $t \geq 0$, and γ is affine, it follows from the first inequality in item 4, item 2 and item 3 that

$$\frac{1}{2}\|z^* - z\|^2 \geq t\lambda\gamma(z^*) + \frac{1}{2}\|z^* - z\|^2 = \frac{1}{2}\|z^* - z_+\|^2 + t\gamma(z_+) + \frac{1}{2}\|z_+ - z\|^2 \tag{6}$$

which, combined with items 2 and 3 proves the second inequality in item 4.

To prove the last item, define $\hat{z} = z - \lambda v$. Using item 4 with $t = 1$, $z' = \hat{z}$ and the inequality (5) we conclude that

$$\|z^* - z\|^2 \geq \|z^* - \hat{z}\|^2 + (1 - \sigma^2)\|\tilde{z} - z\|^2, \quad \sigma\|\tilde{z} - z\| \geq \|\hat{z} - \tilde{z}\|$$

Therefore,

$$\begin{aligned}
\|z^* - \tilde{z}\| &\leq \|z^* - \hat{z}\| + \|\hat{z} - \tilde{z}\| \leq \|z^* - \hat{z}\| + \sigma\|\tilde{z} - z\| \\
&\leq \sqrt{\|z^* - \hat{z}\|^2 + (1 - \sigma^2)\|\tilde{z} - z\|^2}\sqrt{1 + \frac{\sigma^2}{1 - \sigma^2}} \leq \frac{\|z^* - z\|}{\sqrt{1 - \sigma^2}}
\end{aligned}$$

where the fist inequality follow from triangle inequality and the third from Cauchy-Schwarz inequality. □

In the next proposition we show that z_k is closer than z_{k-1} to the solution set, with respect to the norm square, by a quantifiable amount and derive some useful estimations.

Proposition 2.2. *For any $k \geq 1$ and $x^* \in T^{-1}(0)$,*

1. $(1 - \sigma)\|\tilde{z}_k - z_{k-1}\| \leq \|\lambda_k v_k\| \leq (1 + \sigma)\|\tilde{z}_k - z_{k-1}\|$ *and* $2\lambda_k\varepsilon_k \leq \sigma^2\|\tilde{z}_k - z_{k-1}\|^2$;
2. $\|z^* - z_{k-1}\|^2 \geq \|z^* - z_k\|^2 + t_k(1 - \sigma^2)\|\tilde{z}_k - z_{k-1}\|^2 \geq \|z^* - z_{k-1}\|^2$;
3. $\|z^* - z_0\|^2 \geq \|z^* - z_k\|^2 + (1 - \sigma^2)\sum_{j=1}^{k} t_j\|\tilde{z}_j - z_{j-1}\|^2$;
4. $\|z^* - \tilde{z}_k\| \leq \|z^* - z_{k-1}\|/\sqrt{1 - \sigma^2}$ *and* $\|\tilde{z}_k - z_{k-1}\| \leq \|z^* - z_{k-1}\|/\sqrt{1 - \sigma^2}$.

Proof. Items 1 and 2 follow trivially from Lemma 2.1, items 1 and 4, and the assumption $\sigma \in [0, 1)$. Item 3 follows from item 2. Item 4 follows from Lemma 2.1, item 5. □

The aggregate stepsize Λ_k and the ergodic sequences $(\tilde{z}_k^a)$, $(\tilde{v}_k^a)$, and (ε_k^a) associated with the sequences (λ_k), $(\tilde{z}_k)$, (v_k), and (ε_k) are, respectively,

$$\begin{aligned}&\Lambda_k := \sum_{i=1}^{k} t_i\lambda_i,\\&\tilde{z}_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^{k} t_i\lambda_i\tilde{z}_i, \quad v_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^{k} t_i\lambda_i v_i, \quad \varepsilon_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^{k} t_i\lambda_i(\varepsilon_i + \langle \tilde{z}_i - \tilde{z}_k^a, v_i - v_k^a\rangle).\end{aligned} \tag{7}$$

The relevance of these ergodic sequences rests on the following theorem.

Theorem 2.3. *For any $k \geq 1$, $v_k^a \in T^{[\varepsilon_k^a]}(\tilde{z}_k^a)$. Moreover, if d_0 is the distance from z_0 to $T^{-1}(0) \neq \emptyset$, then*

$$\|v_k^a\| \leq \frac{2d_0}{\Lambda_k}, \qquad \varepsilon_k^a \leq \frac{2d_0^2}{\Lambda_k\sqrt{1-\sigma^2}}$$

for any $k \geq 1$

Proof of Theorem 2.3. The first part of the theorem follows from definitions (7), the inclusion in (4), and the transportation formula for the $T^{[\varepsilon]}$ [2, Theorem 3.11].

To prove the second part of the Theorem, let z^* be the projection of z_0 onto $T^{-1}(0)$. It follows from Proposition 2.2 item 2 that $\|z^* - z_k\| \leq \|z^* - z_0\| = d_0$ for any k. Theretofore,

$$\|z_k - z_0\| \leq 2d_0, \qquad \forall k \in \mathbb{N}. \tag{8}$$

Direct use of the update rule for z_k in (4) and of the definition of Λ_k and v_k^a in (7) yields

$$z_0 - z_k = \sum_{j=1}^{k} t_j\lambda_j v_k = \Lambda_k v_k^a. \tag{9}$$

The first inequality follows from the above equation and (8).

Define, for $k = 1, \ldots,$ the affine functions $\gamma_k, \Gamma_k : \mathbb{R}^p \to \mathbb{R}$,

$$\gamma_k(z) = \langle z - \tilde{z}_k, v_k\rangle - \varepsilon_k, \quad \Gamma_k(z) = \sum_{j=1}^{k} t_j\lambda_j\gamma_j(z). \tag{10}$$

We claim that for $k = 1, 2, \ldots$

$$z_k = \operatorname*{argmin}_{z\in\mathbb{R}^p} \Gamma_k(z) + \frac{1}{2}\|z - z_0\|^2, \quad \min_{z\in\mathbb{R}^p} \Gamma_k(z) + \frac{1}{2}\|z - z_0\|^2 \geq 0. \tag{11}$$

The first above relation follow trivially from (9). It follows from Lemma 2.1, items 2 and 3 and the assumption $0 \leq \sigma \leq 1$ that the second relation in (11) holds for $k = 1$. If the inequality in (11) holds for k, as $\Gamma_{k+1} = \Gamma_k + t_{k+1}\lambda_{k+1}\gamma_{k+1}$, it follows again from Lemma 2.1 items 2 and 3, the assumption $0 \leq \sigma \leq 1$, and the first relation in (11) that this inequality holds for $k + 1$.

It follows from (11) and definitions (10) that

$$\Gamma_k(\tilde{z}_k^a) + \frac{1}{2}\|\tilde{z}_k^a - z_0\|^2 = \frac{1}{2}\|\tilde{z}_k^a - z_k\|^2 + \Gamma_k(z_k) + \frac{1}{2}\|z_k - z_0\|^2 \geq \frac{1}{2}\|\tilde{z}_k^a - z_k\|^2$$

Direct use of the transportation formula for the $T^{[\varepsilon]}$ [2, Theorem 3.11], (7), and (10) shows that $-\Gamma_k(\tilde{z}_k^a) = \Lambda_k \varepsilon_k^a$. Therefore

$$2\Lambda_k\varepsilon_k \leq \|\tilde{z}_k^a - z_0\|^2 - \|\tilde{z}_k^a - z_k\|^2 = 2\langle \tilde{z}_k^a - z_0, z_k - z_0\rangle - \|z_k - z_0\|^2$$
$$\leq 2d_0\left(1 + \frac{1}{\sqrt{1-\sigma^2}}\right)\|z_k - z_0\| - \|z_k - z_0\|^2 \leq \frac{4d_0^2}{\sqrt{1-\sigma^2}}$$

□

Next we analyze the pointwise and ergodic complexities of the r-HPE when the large-step condition, introduced in [5, 6], is satisfied and the relaxation parameters t_k are bounded away from zero.

Theorem 2.4. *Let d_0 be the distance from z_0 to $T^{-1}(0) \neq \emptyset$. If for any $k \geq 1$,*

$$\lambda_k\|\tilde{z}_k - z_{k-1}\| \geq \eta > 0, \qquad t_k \geq \tau > 0 \tag{12}$$

then, for any $k \geq 1$,

1. *there exists i, $1 \leq i \leq k$, such that*

$$\|v_i\| \leq \frac{d_0^2}{\eta(1-\sigma)k\tau}, \qquad \varepsilon_i \leq \frac{\sigma^2}{2\eta}\frac{d_0^3}{\left((1-\sigma^2)k\tau\right)^{3/2}};$$

2. *$v_k^a \in T^{[\varepsilon_k^a]}(\tilde{z}_k^a)$,*

$$\|v_k^a\| \leq \frac{2d_0^2}{(\tau k)^{3/2}\eta\sqrt{1-\sigma^2}}, \qquad \varepsilon_k^a \leq \frac{2d_0^3}{(\tau k)^{3/2}\eta(1-\sigma^2)}.$$

Proof. It follows from Proposition 2.2, item 3, that there exists $1 \leq i \leq k$ such that

$$\|\tilde{z}_i - z_{i-1}\| \leq \frac{d_0}{\sqrt{(1-\sigma^2)\tau k}}$$

It follows from the first part of Proposition 2.2 and (12) that, in particular for such an i,

$$\|v_i\| \leq \frac{(1+\sigma)\|\tilde{z}_i - z_{i-1}\|}{\lambda_i}, \quad \varepsilon_i \leq \frac{\sigma^2\|\tilde{z}_i - z_{i-1}\|^2}{2\lambda_i}, \quad \frac{1}{\lambda_i} \leq \frac{\|\tilde{z}_i - z_{i-1}\|}{\eta}.$$

Item 1 follows from the above inequalities.

It follows from (12) and Proposition 2.2 item 3 that

$$\sum_{j=1}^{k} \tau\frac{\eta^2}{\lambda_j^2} \leq \sum_{j=1}^{k} t_j\|\tilde{z}_j - z_{j-1}\|^2 \leq \frac{d_0^2}{1-\sigma^2}.$$

Using this result and Lemma A.1 we conclude that

$$\sum_{j=1}^{k}\lambda_j \geq k^{3/2}\left(\frac{d_0^2}{\tau\eta^2(1-\sigma^2)}\right)^{-1/2} \quad \text{and} \quad \Lambda_k \geq (\tau k)^{3/2}\frac{\eta\sqrt{1-\sigma^2}}{d_0},$$

where the second inequality follows from (7) and the assumption $t_j \geq \tau$ for all j. Item 2 follows from the second above inequality and Theorem 2.3. □

A Auxiliary results

Lemma A.1. *If $\alpha_i > 0$ for $i = 1, \ldots, m$ and $\sum_{i=1}^m \alpha_i^{-2} \leq C$ then $\sum_{i=1}^m \alpha_i \geq m^{3/2}/C^{1/2}$.*

Proof. Take $\alpha \in \mathbb{R}^m_{++}$ such that $\sum_{i=1}^m \alpha_i^{-2} \leq C$ and let $\bar{\alpha} = \sum_{i=1}^m \alpha_i/m$. As t^{-2} is convex for $t > 0$,

$$\frac{1}{\bar{\alpha}^2} \leq \frac{1}{m} \sum_{i=1}^k \frac{1}{\alpha_i^2} \leq \frac{C}{m};$$

therefore, $\sqrt{m/C} \leq \bar{\alpha}$. To end the proof, use the definition of $\bar{\alpha}$. □

The next result was proved in [12, Corollary 1]

Lemma A.2. *If $T : H \rightrightarrows H$ is maximal monotone, $z \in H$ and $\tilde{v} \in T^{[\varepsilon]}(\tilde{z})$, then*

$$\|\lambda\tilde{v} + \tilde{z} - z\|^2 + 2\lambda\varepsilon \geq \left\|\tilde{z} - (\lambda T + I)^{-1}z\right\|^2 + \left\|\lambda^{-1}\left((\lambda T + I)^{-1}z - z\right)\right\|^2.$$

References

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

[2] R. S. Burachik and B. F. Svaiter. ϵ-enlargements of maximal monotone operators in Banach spaces. *Set-Valued Anal.*, 7(2):117–132, 1999.

[3] George J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.*, 29:341–346, 1962.

[4] R. D. C. Monteiro and B. F. Svaiter. Complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle point and convex optimization problems. *SIAM Journal on Optimization*, 21:1688–1720, 2010.

[5] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20:2755–2787, 2010.

[6] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM J. Optim.*, 22(3):914–935, 2012.

[7] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.*, 23(1):475–507, 2013.

[8] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.

[9] Mauricio R. Sicre and Benar F. Svaiter. An $\mathcal{O}(1/k^{3/2})$ hybrid proximal extragradient primal-dual interior point method for non-linear monotone complementarity problems. Preprint A735/2013, IMPA - Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ Brasil 22460-320, 2013.

[10] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.

[11] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.

[12] M. V. Solodov and B. F. Svaiter. Error bounds for proximal point subproblems and associated inexact proximal point algorithms. *Math. Program.*, 88(2, Ser. B):371–389, 2000. Error bounds in mathematical programming (Kowloon, 1998).

[13] M. V. Solodov and B. F. Svaiter. A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem. *SIAM J. Optim.*, 10(2):605–625 (electronic), 2000.

[14] M. V. Solodov and B. F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, 22(7-8):1013–1035, 2001.

[15] M. V. Solodov and B. F. Svaiter. A new proximal-based globalization strategy for the Josephy-Newton method for variational inequalities. *Optim. Methods Softw.*, 17(5):965–983, 2002.

[16] Michael V. Solodov and Benav F. Svaiter. A globally convergent inexact Newton method for systems of monotone equations. In *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997)*, volume 22 of *Appl. Optim.*, pages 355–369. Kluwer Acad. Publ., Dordrecht, 1999.