SIAM J. OPTIM.
Vol. 19, No. 1, pp. 240–260

A CLASS OF INEXACT VARIABLE METRIC PROXIMAL POINT ALGORITHMS*

L. A. PARENTE†, P. A. LOTITO‡, AND M. V. SOLODOV§

Abstract. For the problem of solving maximal monotone inclusions, we present a rather general class of algorithms, which contains hybrid inexact proximal point methods as a special case and allows for the use of a variable metric in subproblems. The global convergence and local linear rate of convergence are established under standard assumptions. We demonstrate the advantage of variable metric implementation in the case of solving systems of smooth monotone equations by the proximal Newton method.

Key words. proximal point methods, variable metric, maximal monotone operators, approximation

AMS subject classifications. 90C30, 90C33

DOI. 10.1137/070688146

1. Introduction.

Given a maximal monotone operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, we consider the classical problem of finding a zero of T, i.e., $z \in \mathbb{R}^n$ such that

$$0 \in T(z). \tag{1.1}$$

As is well known, many important problems can be cast in this framework. Some examples are convex optimization, min-max problems, and monotone variational inequalities over convex sets; see, e.g., [23].

Given some $z^k \in \mathbb{R}^n$, the current approximation to a solution of (1.1), the proximal point iteration [19, 22] generates z^{k+1} as the solution of the regularized subproblem

$$0 \in c_k T(z) + z - z^k, \tag{1.2}$$

where $c_k > 0$ is the regularization parameter. As is well known, the proximal point method serves as a basis for developing and analyzing various useful computational techniques, such as splitting methods for variational problems (e.g., [18, 31, 13, 33, 34, 24]), the methods of multipliers (e.g., [21, 15]), and bundle methods for nonsmooth optimization (see, e.g., [16, 3]), to name a few. In computational context, it is important to handle approximate solutions of subproblems; this will be discussed a little further. Also, it is attractive to allow for the use of a variable metric (or preconditioning). Motivated by the latter issue, we shall consider the following *generalized* proximal subproblem:

$$0 \in c_k M_k T(z) + z - z^k, \tag{1.3}$$

*Received by the editors April 12, 2007; accepted for publication (in revised form) October 5, 2007; published electronically March 19, 2008.
http://www.siam.org/journals/siopt/19-1/68814.html

†CONICET, Universidad Nacional de Rosario, Argentina (lparente@fceia.unr.edu.ar).

‡CONICET, Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina (plotito@exa.unicen.edu.ar).

§IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil (solodov@impa.br). The research of this author is supported in part by CNPq grants 301508/2005-4, 490200/2005-2, and 550317/2005-8, by PRONEX–Optimization, and by FAPERJ.

where M_k is a symmetric positive definite matrix. The case of the classical (exact) iteration (1.2) corresponds to taking $M_k = I$ (the identity matrix) in (1.3). Given the presence of the matrix M_k, we could in principle dispense with the scalar parameter c_k in (1.3). We prefer, however, to keep it because this appears convenient in some parts of the convergence analysis and in our application to solving systems of monotone equations, discussed in section 5.

To handle approximate solutions, we shall use an extension to the variable metric setting of the rules proposed in [27, 26] and unified in [30]. In those algorithms, the *relative* error in the approximation needs only to be bounded (above, by one), which is a numerically sound requirement, and inexact values of the operator T are allowed, which is useful in various applications [29, 28, 24]. Specifically, to solve (1.3) approximately, the task would be to compute a triplet $(\hat{z}^k, \hat{v}^k, \varepsilon_k) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ such that

$$\begin{cases} \hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k), \\ c_k M_k \hat{v}^k + \hat{z}^k - z^k = \delta^k, \end{cases}$$
$$\|\delta^k\|^2_{M_k^{-1}} + 2c_k\varepsilon_k \le \sigma_k^2 \left(\|c_k M_k \hat{v}^k\|^2_{M_k^{-1}} + \|\hat{z}^k - z^k\|^2_{M_k^{-1}} \right),$$

where $\sigma_k \in [0,1)$ is the error tolerance (relaxation) parameter, by $\|\cdot\|_M$ we denote the norm induced by a symmetric positive definite matrix $M \in \mathcal{M}^n_{++}$, i.e.,

$$\|z\|_M = \sqrt{\langle z, Mz\rangle},$$

and $T^\varepsilon : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is the ε-enlargement of a maximal monotone operator T [5, 6], defined as

$$T^\varepsilon(z) := \{v \in \mathbb{R}^n \mid \langle w - v, y - z\rangle \ge -\varepsilon \ \forall y \in \mathbb{R}^n, \forall w \in T(y)\}, \ \varepsilon \ge 0.$$

We note that, to check the above criterion, one does not need to invert the matrix M_k, as will be explained in what follows. The presented approximation rule is constructive and has advantages in some situations, when compared to the original [22] (and its variations, e.g., [32, 11, 7]), where essentially one has $\varepsilon_k = 0$ and $\sum_{k=0}^\infty \|\delta^k\| < \infty$ (in the setting of $M_k = I$). We refer the reader to [26, 29, 28, 24] for some applications where the relative-error criterion appears useful. It will also play a central role in the method discussed in section 5.

Most proximal-related schemes in the literature that use variable metrics typically deal only with the special case of optimization, i.e., the case where the operator T is the subdifferential of a convex function [2, 20, 17, 10]. To our knowledge, the exception is [7] and some of the subsequent results [8, 9]. We note that our use of a variable metric is different from [7], where (exact) iteration is of the form

$$z^{k+1} = z^k + M_k((I + c_k T)^{-1} - I)z^k.$$

The exact iteration of solving (1.3) can be written as

$$z^{k+1} = (I + c_k M_k T)^{-1} z^k,$$

and the two are the same only when $M_k = I$. It should be noted, however, that [7] does not require M_k to be symmetric, and in this respect our development can be more restrictive for some applications. On the other hand, global convergence of the method of [7] requires a rather technical assumption about the matrices M_k.

Specifically, the assumption of [7, Hypothesis (H2)] is that there exists a nonempty bounded subset Ω of $T^{-1}(0)$ such that

$$\|(M_k - I)D_k(z^k)\| \leq \gamma_k \|D_k(z^k)\| \quad \text{for all } k,$$

where

$$D_k = (I + c_k T)^{-1} - I,$$
$$\gamma_k = \frac{\|D_k(z^k)\|}{2t_k + 3\|D_k(z^k)\|}, \quad \text{with } t_k = \sup_{z\in\Omega} \|z - z^k\|.$$

This assumption essentially means that matrices M_k should not deviate from the identity too much, in the given sense, and it is in general unverifiable (unless one takes $M_k = I$) and globally quite restrictive. The only assumption we make in our global convergence analysis is, by comparison, rather mild:

$$\frac{1}{1+\eta_k} M_k \preceq M_{k+1}, \ \eta_k > 0 \quad \text{for all } k, \quad \sum_{k=0}^{\infty} \eta_k < \infty, \tag{1.4}$$

where, for $A, B \in \mathcal{M}^n_{++}$, by $A \preceq B$ we mean that $B - A$ is a positive semidefinite matrix. This condition does not introduce any essential restrictions on the choice of the matrix M_{k+1} for a given k (for a particular k, the choice of η_k is rather flexible), and it is always satisfied if we take $M_k \preceq M_{k+1}$. Also, [7] does not allow approximations of the operator T and requires error terms to be summable, basically following [22]. In the aspect of inexact solution of subproblems, our conditions (already mentioned above) are more flexible and constructive.

A few more words about our notation are in order. By $\mathcal{M}^n_{++}$ we denote the space of symmetric positive definite matrices. For $M \in \mathcal{M}^n_{++}$, $\lambda_{min}(M)$ and $\lambda_{max}(M)$ stand for the minimal and the maximal eigenvalues of M, respectively. For any $A \preceq B$, it holds that $\|z\|_A \leq \|z\|_B$. In particular, if

$$0 < \lambda_l \leq \lambda_{min}(M) \leq \lambda_{max}(M) \leq \lambda_u,$$

then for any $x \in \mathbb{R}^n$ it holds that

$$\lambda_l \|x\|^2 \leq \|x\|^2_M \leq \lambda_u \|x\|^2, \quad \frac{1}{\lambda_u}\|x\|^2 \leq \|x\|^2_{M^{-1}} \leq \frac{1}{\lambda_l}\|x\|^2. \tag{1.5}$$

By $\langle x, y\rangle$ we denote the usual inner product between $x, y \in \mathbb{R}^n$. For a matrix $M \in \mathcal{M}^n_{++}$, we denote $\langle x, y\rangle_M = \langle Mx, y\rangle$. For a closed convex set $\Omega \subseteq \mathbb{R}^n$ and a matrix $M \in \mathcal{M}^n_{++}$, the "skewed" projection operator onto Ω under the matrix M is given by

$$P_{\Omega,M}(z) = \arg\min_{x\in\Omega} \frac{1}{2}\langle x - z, M(x - z)\rangle = \arg\min_{x\in\Omega} \frac{1}{2}\|x - z\|^2_M;$$

i.e., it is the projection operator with respect to the norm $\|\cdot\|_M$. Then the associated distance from $z \in \mathbb{R}^n$ to Ω is defined as $\mathrm{dist}(z, \Omega)_M = \|z - P_{\Omega,M}(z)\|$.

2. Approximate solutions of the generalized proximal subproblem. Given a maximal monotone operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, $z \in \mathbb{R}^n$, $c > 0$, and $M \in \mathcal{M}^n_{++}$, consider the generalized proximal point subproblem

$$0 \in cMT(y) + y - z, \tag{2.1}$$

with respect to $y \in \mathbb{R}^n$. This is clearly equivalent to

$$0 \in cT(y) + M^{-1}(y - z),$$

and the fact that the inclusion above has a solution follows, e.g., from [4, Proposition 3].

We next define the notion of approximate solutions of generalized proximal subproblems. Consider the system

$$\begin{cases} v \in T(y), \\ 0 = cMv + y - z, \end{cases} \tag{2.2}$$

which is equivalent to (2.1).

DEFINITION 2.1. *We say that a triplet* $(y, v, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ *is an approximate solution of the proximal system* (2.2) *with error tolerance* $\sigma \in [0, 1)$ *if*

$$v \in T^{\varepsilon}(y)$$

and

$$\|cMv + y - z\|_{M^{-1}}^2 + 2c\varepsilon \leq \sigma^2(\|cMv\|_{M^{-1}}^2 + \|y - z\|_{M^{-1}}^2). \tag{2.3}$$

Note that the exact solution of (2.2) corresponds to taking $\varepsilon = 0 = \sigma$ in the definition above. We next establish some properties of approximate solutions of generalized proximal systems.

LEMMA 2.2. *Let* $z \in \mathbb{R}^n, c > 0$, *and* $M \in \mathcal{M}_{++}^n$. *A triplet* $(y, v, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ *being an approximate solution of the proximal system* (2.2) *with error tolerance* $\sigma \in [0, 1)$ *is equivalent to the conditions*

$$v \in T^{\varepsilon}(y), \quad \langle v, z - y \rangle - \varepsilon \geq \frac{1 - \sigma^2}{2c}\left(\|cMv\|_{M^{-1}}^2 + \|y - z\|_{M^{-1}}^2\right). \tag{2.4}$$

In addition, it holds that

$$\frac{c(1 - \rho)}{1 - \sigma^2}\|Mv\|_{M^{-1}} \leq \|y - z\|_{M^{-1}} \leq \frac{c(1 + \rho)}{1 - \sigma^2}\|Mv\|_{M^{-1}}, \tag{2.5}$$

where $\rho = \sqrt{1 - (1 - \sigma^2)^2}$.

Furthermore, the three conditions

1. $0 \in T(z)$,
2. $v = 0$,
3. $y = z$

are equivalent and imply that $\varepsilon = 0$.

Proof. Rearranging terms in (2.3), we have

$$\begin{aligned} \sigma^2(\|cMv\|_{M^{-1}}^2 + \|y - z\|_{M^{-1}}^2) &\geq 2c\varepsilon + \|cMv\|_{M^{-1}}^2 + \|y - z\|_{M^{-1}}^2 + 2\langle cMv, y - z \rangle_{M^{-1}} \\ &= 2c\varepsilon + \|cMv\|_{M^{-1}}^2 + \|y - z\|_{M^{-1}}^2 - 2c\langle v, z - y \rangle, \end{aligned}$$

which gives the inequality in (2.4).

By using $\varepsilon \geq 0$ and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} \frac{1 - \sigma^2}{2c}(\|cMv\|_{M^{-1}}^2 + \|y - z\|_{M^{-1}}^2) &\leq \langle v, z - y \rangle - \varepsilon \\ &\leq \langle Mv, z - y \rangle_{M^{-1}} \leq \|Mv\|_{M^{-1}}\|y - z\|_{M^{-1}}. \end{aligned}$$

By denoting $t = \|y - z\|_{M^{-1}}$ and resolving the quadratic inequality

$$t^2 - \frac{2\|cMv\|_{M^{-1}}}{1-\sigma^2} t + \|cMv\|^2_{M^{-1}} \leq 0$$

with respect to t, we obtain (2.5).

Finally, suppose that $0 \in T(z)$. Since $v \in T^{\varepsilon}(y)$, we have

$$\langle v - 0, y - z \rangle \geq -\varepsilon \quad \Rightarrow \quad \langle v, z - y \rangle - \varepsilon \leq 0.$$

By using now (2.4), we have $cMv = 0$ (so that $v = 0$) and $y - z = 0$.

If we assume that $v = 0$, then (2.4) implies that $y = z$ and vice versa. In either case, $0 \in T(z)$. From (2.4) it is also clear that all of these conditions imply that $\varepsilon = 0$. ∎

The next result shows how to make progress towards a solution of the original problem (1.1), by using the obtained approximate solution of the generalized proximal subproblem.

LEMMA 2.3. *Let $z \in \mathbb{R}^n, y \in \mathbb{R}^n, \varepsilon \geq 0$, and $v \in T^{\varepsilon}(y)$. Suppose that*

$$\langle v, z - y \rangle - \varepsilon > 0.$$

Then, for any $z^* \in T^{-1}(0)$, any $M \in \mathcal{M}^n_{++}$, and any $\tau \geq 0$, it holds that

$$\|z^* - z^+\|^2_{M^{-1}} \leq \|z^* - z\|^2_{M^{-1}} - (1 - (1-\tau)^2) a^2 \|Mv\|^2_{M^{-1}},$$

where

$$z^+ := z - \tau a M v$$

and

$$a := \frac{\langle v, z - y \rangle - \varepsilon}{\|Mv\|^2_{M^{-1}}}.$$

Proof. Define the closed half-space

$$H = \{ w \in \mathbb{R}^n \mid \langle v, w - y \rangle - \varepsilon \leq 0 \}.$$

By the assumption, $z \notin H$. Let $\bar{z}$ be the skewed projection of z onto H, under the matrix M^{-1}. As is easily seen,

$$\bar{z} = P_{H, M^{-1}}(z) = z - aMv.$$

For any $x \in H$, it holds that

$$\langle x - \bar{z}, v \rangle = \langle x - z + aMv, v \rangle = \langle x - z, v \rangle + \frac{\langle v, z - y \rangle - \varepsilon}{\|Mv\|^2_{M^{-1}}} \langle Mv, v \rangle = \langle x - y, v \rangle - \varepsilon \leq 0.$$

Hence,

$$\langle x - \bar{z}, z^+ - z \rangle_{M^{-1}} = \langle x - \bar{z}, M^{-1}(-\tau a M v) \rangle = -\tau a \langle x - \bar{z}, v \rangle \geq 0.$$

Observe that $\bar z - z^+ = (\tau - 1)aMv$. We then obtain

$$\begin{aligned}
\|x - z\|^2_{M^{-1}} &= \|x - z^+\|^2_{M^{-1}} + \|z^+ - z\|^2_{M^{-1}} + 2\langle x - z^+, z^+ - z\rangle_{M^{-1}}\\
&= \|x - z^+\|^2_{M^{-1}} + \|z^+ - z\|^2_{M^{-1}} + 2\langle \bar z - z^+, z^+ - z\rangle_{M^{-1}}\\
&\quad +2\langle x - \bar z, z^+ - z\rangle_{M^{-1}}\\
&\geq \|x - z^+\|^2_{M^{-1}} + \|z^+ - z\|^2_{M^{-1}} + 2\langle \bar z - z^+, z^+ - z\rangle_{M^{-1}}\\
&= \|x - z^+\|^2_{M^{-1}} + (\tau a)^2\|Mv\|^2_{M^{-1}} + 2(\tau - 1)a(-\tau a)\|Mv\|^2_{M^{-1}}\\
&= \|x - z^+\|^2_{M^{-1}} + (1 - (1 - \tau)^2)a^2\|Mv\|^2_{M^{-1}}.
\end{aligned}$$

Suppose that $z^* \in T^{-1}(0)$. Since $v \in T^\varepsilon(y)$, we have

$$\langle v - 0, y - z^*\rangle \geq -\varepsilon.$$

This shows that $z^* \in H$. We can then set $x = z^*$ in the chain of inequalities above to complete the proof. $\square$

3. The algorithm. Lemma 2.3 shows that, with a proper choice of parameters, a step in the direction obtained from an approximate solution of the generalized proximal system, scaled by the chosen metric, brings us closer to the solution set of the original problem. This suggests the following scheme, which we shall call the variable metric hybrid inexact proximal point method.

ALGORITHM 3.1. **Initialization:** *Choose* $z^0 \in \mathbb{R}^n$, $c > 0$, $\bar\sigma \in (0, 1)$, $\theta \in (0, 1)$, *and* $0 < \lambda_l < \lambda_u$. *Set* $k := 0$.

Inexact proximal step: *Choose* $c_k \geq c$, *a symmetric positive definite matrix* M_k *satisfying* $\lambda_l \leq \lambda_{min}(M_k) \leq \lambda_{max}(M_k) \leq \lambda_u$, *and the error tolerance parameter* $\sigma_k \in [0, \bar\sigma)$. *Find* $\hat z^k \in \mathbb{R}^n, \hat v^k \in \mathbb{R}^n$, *and* $\varepsilon_k \geq 0$ *such that*

$$\text{(3.1)} \qquad \begin{cases} \hat v^k \in T^{\varepsilon_k}(\hat z^k),\\ \delta^k = c_k M_k \hat v^k + \hat z^k - z^k \end{cases}$$

and

$$\text{(3.2)} \qquad \|\delta^k\|^2_{M_k^{-1}} + 2c_k\varepsilon_k \leq \sigma_k^2\left(\|c_kM_k\hat v^k\|^2_{M_k^{-1}} + \|\hat z^k - z^k\|^2_{M_k^{-1}}\right).$$

Iterates update: *If* $\hat z^k = z^k$, *stop. Otherwise, choose* $\tau_k \in [1 - \theta, 1 + \theta]$, *and set*

$$z^{k+1} = z^k - \tau_k a_k M_k \hat v^k, \quad a_k = \frac{\langle \hat v^k, z^k - \hat z^k\rangle - \varepsilon_k}{\|M_k\hat v^k\|^2_{M_k^{-1}}}.$$

Set $k := k + 1$, *and go to the inexact proximal step.*

We note that it is not necessary to calculate the inverse of M_k in order to implement Algorithm 3.1 (in particular, for checking the condition (3.2) and computing a_k). Indeed, by (1.5), the condition (3.2) is satisfied if

$$\|\delta^k\|^2 + 2\lambda_u c_k\varepsilon_k \leq \frac{\lambda_u\sigma_k^2}{\lambda_l}\left(\|c_kM_k\hat v^k\|^2 + \|\hat z^k - z^k\|^2\right).$$

Alternatively, in the latter relation, instead of λ_l and λ_u one can use any other (in particular, tighter) lower and upper bounds for the eigenvalues of M_k. Also, the scalar a_k can be calculated as

$$a_k = \frac{\langle \hat v^k, z^k - \hat z^k\rangle - \varepsilon_k}{\langle M_k\hat v^k, \hat v^k\rangle}.$$

The next result shows that some specific realizations of Algorithm 3.1 allow for the simple update

$$z^{k+1} = z^k - c_k M_k \hat{v}^k.$$

This is the update that we shall use for our application in section 5. Specifically, we have the following.

PROPOSITION 3.1. *If the inequality in* (3.2) *is replaced by the stronger condition* $\|\delta^k\|^2_{M_k^{-1}} + 2c_k\varepsilon_k \le \sigma_k^2 \|\hat{z}^k - z^k\|^2_{M_k^{-1}}$, *and we choose* $\sigma_k \le \theta$, *then there exists* $\tau_k \in (1-\sigma_k, 1+\sigma_k) \subset (0,2)$ *such that* $\tau_k a_k = c_k$.

Proof. In the case of interest, $\hat{v}^k \neq 0$ and $\hat{z}^k \neq z^k$. By using the triangle inequality, from $\|\delta^k\|_{M_k^{-1}} \le \sigma_k \|\hat{z}^k - z^k\|_{M_k^{-1}}$ we obtain

$$\|\hat{z}^k - z^k\|_{M_k^{-1}} - c_k \|M_k \hat{v}^k\|_{M_k^{-1}} \le \sigma_k \|\hat{z}^k - z^k\|_{M_k^{-1}}$$

and

$$c_k \|M_k \hat{v}^k\|_{M_k^{-1}} - \|\hat{z}^k - z^k\|_{M_k^{-1}} \le \sigma_k \|\hat{z}^k - z^k\|_{M_k^{-1}},$$

implying that

$$(1-\sigma_k)\frac{\|\hat{z}^k - z^k\|_{M_k^{-1}}}{\|M_k \hat{v}^k\|_{M_k^{-1}}} \le c_k \le (1+\sigma_k)\frac{\|\hat{z}^k - z^k\|_{M_k^{-1}}}{\|M_k \hat{v}^k\|_{M_k^{-1}}}. \tag{3.3}$$

Furthermore, by the Cauchy–Schwarz inequality, since $\varepsilon_k \ge 0$ we have

$$a_k = \frac{\langle \hat{v}^k, z^k - \hat{z}^k\rangle - \varepsilon_k}{\|M_k \hat{v}^k\|^2_{M_k^{-1}}} \le \frac{\langle M_k \hat{v}^k, z^k - \hat{z}^k\rangle_{M_k^{-1}}}{\|M_k \hat{v}^k\|^2_{M_k^{-1}}} \le \frac{\|\hat{z}^k - z^k\|_{M_k^{-1}}}{\|M_k \hat{v}^k\|_{M_k^{-1}}}.$$

Finally, since

$$\begin{aligned}\langle \hat{v}^k, \hat{z}^k - z^k\rangle &= \langle M_k \hat{v}^k, \hat{z}^k - z^k\rangle_{M_k^{-1}} \\ &= \frac{\|c_k M_k \hat{v}^k + \hat{z}^k - z^k\|^2_{M_k^{-1}} - \|\hat{z}^k - z^k\|^2_{M_k^{-1}} - \|c_k M_k \hat{v}^k\|^2_{M_k^{-1}}}{2c_k},\end{aligned}$$

by using (2.5) and (3.3), we obtain

$$\begin{aligned}a_k &= \frac{\|\hat{z}^k - z^k\|^2_{M_k^{-1}} + \|c_k M_k \hat{v}^k\|^2_{M_k^{-1}} - \left(\|c_k M_k \hat{v}^k + \hat{z}^k - z^k\|^2_{M_k^{-1}} + 2c_k\varepsilon_k\right)}{2c_k \|M_k \hat{v}^k\|^2_{M_k^{-1}}} \\ &\ge \frac{c_k}{2} + (1-\sigma_k^2)\frac{\|\hat{z}^k - z^k\|^2_{M_k^{-1}}}{\|c_k M_k \hat{v}^k\|^2_{M_k^{-1}}} \ge \frac{c_k}{2}\left(1 + \frac{1-\sigma_k^2}{(1+\sigma_k)^2}\right) = \frac{c_k}{1+\sigma_k}.\end{aligned}$$

Hence,

$$(1-\sigma_k)a_k \le c_k \le (1+\sigma_k)a_k,$$

which establishes the claim. ∎

4. Convergence analysis. If Algorithm 3.1 terminates at some iteration k, then $z^k = \hat{z}^k$, and, by Lemma 2.2, z^k is a solution. We next consider the case when infinite sequences $\{z^k\}, \{\hat{z}^k\}, \{\hat{v}^k\}$, and $\{\varepsilon_k\}$ are generated. For any k, we have $\hat{v}^k \neq 0$, $\hat{z}^k \neq z^k$, and by Lemma 2.2,

$$\langle \hat{v}^k, z^k - \hat{z}^k \rangle - \varepsilon_k \geq \frac{1-\sigma_k^2}{2c_k}\left(\|c_k M_k \hat{v}^k\|^2_{M_k^{-1}} + \|\hat{z}^k - z^k\|^2_{M_k^{-1}}\right) > 0.$$

By the definition of a_k, we then conclude that

$$a_k \geq \frac{1-\sigma_k^2}{2c_k}\left(\frac{\|c_k M_k \hat{v}^k\|^2_{M_k^{-1}} + \|\hat{z}^k - z^k\|^2_{M_k^{-1}}}{\|M_k \hat{v}^k\|^2_{M_k^{-1}}}\right). \tag{4.1}$$

By the Cauchy–Schwarz inequality,

$$\|c_k M_k \hat{v}^k\|^2_{M_k^{-1}} + \|\hat{z}^k - z^k\|^2_{M_k^{-1}} \geq 2c_k \|M_k \hat{v}^k\|_{M_k^{-1}} \|\hat{z}^k - z^k\|_{M_k^{-1}}.$$

By combining this relation with (4.1), we obtain

$$a_k \|M_k \hat{v}^k\|_{M_k^{-1}} \geq (1-\sigma_k^2)\|\hat{z}^k - z^k\|_{M_k^{-1}}. \tag{4.2}$$

Combining (4.1) and (2.5), and using the definition of ρ_k, gives the following lower bound for a_k:

$$\begin{aligned}
a_k &\geq \frac{(1-\sigma_k^2)c_k}{2}\left(1 + \frac{\|\hat{z}^k - z^k\|^2_{M_k^{-1}}}{\|c_k M_k \hat{v}^k\|^2_{M_k^{-1}}}\right) \\
&\geq \frac{(1-\sigma_k^2)c_k}{2}\left(1 + \left(\frac{1-\rho_k}{1-\sigma_k^2}\right)^2\right) \\
&= \frac{c_k\left(\left(1-\sigma_k^2\right)^2 + \left(1 - \sqrt{1-(1-\sigma_k^2)^2}\right)^2\right)}{2\,(1-\sigma_k^2)} \\
&= \frac{c_k\left(1 - \sqrt{1-(1-\sigma_k^2)^2}\right)}{1-\sigma_k^2} \\
&= \frac{(1-\sigma_k^2)c_k}{1+\sqrt{1-(1-\sigma_k^2)^2}}.
\end{aligned} \tag{4.3}$$

Hence, the parameter a_k is bounded away from zero:

$$a_k \geq \frac{(1-\bar{\sigma}^2)c}{1+\sqrt{1-(1-\bar{\sigma}^2)^2}} > 0. \tag{4.4}$$

We proceed to establish the global convergence of Algorithm 3.1.

PROPOSITION 4.1. *Suppose that* $T^{-1}(0) \neq \emptyset$ *and condition* (1.4) *holds. Then any sequences generated by Algorithm* 3.1 *have the following properties:*

1. $\{z^k\}$ *is bounded.*
2. $\sum_{k=0}^{\infty} \|a_k M_k \hat{v}^k\|^2 < \infty$.
3. $\lim_{k\to\infty} \|\hat{z}^k - z^k\| = \lim_{k\to\infty} \|\hat{v}^k\| = \lim_{k\to\infty} \|\varepsilon_k\| = 0$.

Proof. By condition (1.4), it holds that

$$\prod_{k=0}^{\infty}(1+\eta_k) = p < \infty,$$

and, for all k,

$$M_{k+1}^{-1} \preceq (1+\eta_k)M_k^{-1}.$$

By (1.5), for all k we have

$$\lambda_u^{-1}\|z\|^2 \le \lambda_{min}(M_k^{-1})\|z\|^2 \le \|z\|^2_{M_k^{-1}} \le \lambda_{max}(M_k^{-1})\|z\|^2 \le \lambda_l^{-1}\|z\|^2 \quad \forall\, z \in \mathbb{R}^n.$$

By using (4.1) and Lemma 2.3, we have that for any $z^* \in T^{-1}(0)$ it holds that

$$\begin{aligned}\|z^* - z^{k+1}\|^2_{M_k^{-1}} &\le \|z^* - z^k\|^2_{M_k^{-1}} - (1-(1-\tau_k)^2)a_k^2\|M_k\hat{v}^k\|^2_{M_k^{-1}} \\ &\le \|z^* - z^k\|^2_{M_k^{-1}} - (1-\theta^2)\|a_k M_k \hat{v}^k\|^2_{M_k^{-1}}.\end{aligned}$$

Hence,

$$\begin{aligned}\lambda_u^{-1}\|z^* - z^{k+1}\|^2 &\le \|z^* - z^{k+1}\|^2_{M_{k+1}^{-1}} \\ &\le (1+\eta_k)\|z^* - z^{k+1}\|^2_{M_k^{-1}} \\ &\le (1+\eta_k)\left(\|z^* - z^k\|^2_{M_k^{-1}} - (1-\theta^2)\|a_k M_k \hat{v}^k\|^2_{M_k^{-1}}\right) \\ &\le (1+\eta_k)\|z^* - z^k\|^2_{M_k^{-1}} - (1-\theta^2)\|a_k M_k \hat{v}^k\|^2_{M_k^{-1}}.\end{aligned}$$

By applying this inequality consecutively, we obtain

$$\lambda_u^{-1}\|z^* - z^{k+1}\|^2 \le \prod_{i=0}^{k}(1+\eta_i)\|z^* - z^0\|^2_{M_0^{-1}} - (1-\theta^2)\sum_{i=0}^{k}\|a_i M_i \hat{v}^i\|^2_{M_i^{-1}}. \tag{4.5}$$

We therefore have, for any k,

$$\|z^* - z^k\|^2 \le \lambda_u \prod_{i=0}^{k-1}(1+\eta_i)\|z^* - z^0\|^2_{M_0^{-1}} \le \frac{p\lambda_u}{\lambda_l}\|z^* - z^0\|^2, \tag{4.6}$$

which shows that the sequence $\{z^k\}$ is bounded. From (4.5), we also have

$$(1-\theta^2)\sum_{i=0}^{k}\|a_i M_i \hat{v}^i\|^2_{M_i^{-1}} \le \prod_{i=0}^{k}(1+\eta_i)\|z^* - z^0\|^2_{M_0^{-1}}.$$

By passing onto the limit when $k \to \infty$ in this relation, we obtain

$$\sum_{k=0}^{\infty}\|a_k M_k \hat{v}^k\|^2 \le \lambda_u \sum_{k=0}^{\infty}\|a_k M_k \hat{v}^k\|^2_{M_k^{-1}} \le \frac{p\lambda_u}{1-\theta^2}\|z^* - z^0\|^2_{M_0^{-1}} < \infty.$$

This proves the second item in the assertion and, as a consequence, that

$$\lim_{k\to\infty}\|a_k M_k \hat{v}^k\| = 0.$$

From (4.2) and Lemma 2.2, we then conclude that

$$\lim_{k\to\infty} \|M_k \hat{v}^k\| = 0 \quad \text{and} \quad \lim_{k\to\infty} \|\hat{z}^k - z^k\| = 0.$$

Since the matrices M_k are uniformly positive definite, we also have $\lim_{k\to\infty} \hat{v}^k = 0$. Also, since $\varepsilon_k \le \langle \hat{v}^k, z^k - \hat{z}^k\rangle$, it follows that $\lim_{k\to\infty} \varepsilon_k = 0$. ∎

We are now in a position to complete the proof of global convergence of Algorithm 3.1. Given the properties established in Proposition 4.1, the argument is close to standard; we include it mainly for completeness.

THEOREM 4.2. *Suppose that* $T^{-1}(0) \ne \emptyset$ *and condition* (1.4) *holds. Then any sequence* $\{z^k\}$ *generated by Algorithm* 3.1 *converges to an element of* $T^{-1}(0)$.

Proof. By Proposition 4.1, the sequence $\{z^k\}$ is bounded. Therefore, it has some accumulation point, say, $\bar{z} \in \mathbb{R}^n$. Let $\{z^{k_j}\}$ be any subsequence converging to this accumulation point: $\lim_{j\to\infty} z^{k_j} = \bar{z}$. Since $\lim_{k\to\infty} \|\hat{z}^k - z^k\| = 0$, we have $\hat{z}^{k_j} \to \bar{z}$. For any $z \in \mathbb{R}^n$ and any $u \in T(z)$, $\langle u - v^{k_j}, z - \hat{z}^{k_j}\rangle \ge -\varepsilon_{k_j}$. Hence,

$$\langle u - 0, z - \hat{z}^{k_j}\rangle \ge \langle v^{k_j}, z - \hat{z}^{k_j}\rangle - \varepsilon_{k_j}.$$

Since $v^{k_j} \to 0$, $\varepsilon_{k_j} \to 0$, and $\hat{z}^{k_j} \to \bar{z}$, by passing onto the limit when $j \to \infty$ we obtain

$$\langle u - 0, z - \bar{z}\rangle \ge 0.$$

As $z \in \mathbb{R}^n$ and $u \in T(z)$ were arbitrarily chosen, and T is maximal monotone, the above relation shows that $0 \in T(\bar{z})$; i.e., $\bar{z}$ is a solution.

Suppose that there exists another subsequence $\{z^{t_i}\}$ converging to $\tilde{z} \ne \bar{z}$. Fix some $d \in (0, \|\tilde{z} - \bar{z}\|)$. Since $\tilde{z}$ and $\bar{z}$ are limits of corresponding subsequences, there exists an index i_0 such that for all $i \ge i_0$

$$\|z^{t_i} - \tilde{z}\| < \frac{d}{2}\sqrt{\frac{\lambda_l}{p\lambda_u}},$$

where $p = \prod_{k=0}^{\infty}(1+\eta_k)$, and there exists an index j_0 such that for all $j \ge j_0$

$$k_j > t_{i_0} \text{ and } \|z^{k_j} - \bar{z}\| < \frac{d}{2}.$$

Therefore,

$$\|z^{k_j} - \tilde{z}\| > \frac{d}{2} \quad \forall j \ge j_0.$$

Since, as already established above, $\tilde{z} \in T^{-1}(0)$, the same reasoning used to obtain (4.6) gives, for any $j \ge j_0$,

$$\frac{d}{2} < \|z^{k_j} - \tilde{z}\| \le \sqrt{\frac{p\lambda_u}{\lambda_l}}\|z^{t_{i_0}} - \tilde{z}\| < \frac{d}{2},$$

which is a contradiction.

Hence, $\{z^k\}$ has the unique accumulation point, which is a solution. ∎

We proceed with a convergence rate analysis of Algorithm 3.1. To this end, we first establish an *error bound* for the exact solution of the generalized proximal system

$$\begin{cases} v \in T(y), \\ 0 = cMv + y - z. \end{cases} \tag{4.7}$$

We note that the obtained bound is for the distance both in terms of y and in terms of v, and it does not involve any unknown constants. Specifically, we have the following.

LEMMA 4.3. *Let y^*, v^* be the exact solution of the proximal system* (4.7), *with some $c > 0$, $z \in \mathbb{R}^n$, and $M \in \mathcal{M}^n_{++}$. Then for any $y \in \mathbb{R}^n$ and any $v \in T^{\varepsilon}(y)$, it holds that*

$$\|y - y^*\|^2_{M^{-1}} + c^2\|Mv - Mv^*\|^2_{M^{-1}} \leq \|cMv + y - z\|^2_{M^{-1}} + 2c\varepsilon.$$

Proof. By using $cMv^* + y^* - z = 0$, we obtain

$$\begin{aligned}\|cMv + y - z\|^2_{M^{-1}} &= \|cMv + y - z - (cMv^* + y^* - z)\|^2_{M^{-1}} \\ &= \|cMv - cMv^* + y - y^*\|^2_{M^{-1}} \\ &= c^2\|Mv - Mv^*\|^2_{M^{-1}} + \|y - y^*\|^2_{M^{-1}} + 2c\langle v - v^*, y - y^*\rangle \\ &\geq c^2\|Mv - Mv^*\|^2_{M^{-1}} + \|y - y^*\|^2_{M^{-1}} - 2c\varepsilon. \qquad \square\end{aligned}$$

We shall show linear convergence of Algorithm 3.1 under the assumption that T^{-1} has the following Lipschitzian property at zero: There exist some $L_1 > 0$ and $L_2 > 0$ such that

$$T^{-1}(v) \subset T^{-1}(0) + L_1\|v\|\mathcal{B} \quad \forall v \in L_2\mathcal{B},$$

where $\mathcal{B} = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}$. Note that this condition does not imply that the solution set $T^{-1}(0)$ is a singleton. The equivalent form of this local Lipschitzian property, used below, is

$$\text{(4.8)} \qquad \operatorname{dist}(z, T^{-1}(0)) \leq L_1 \min_{v \in T(z)} \|v\| \quad \forall\, z \in \{z' \in \operatorname{dom} T \mid \min_{v \in T(z')} \|v\| \leq L_2\}.$$

We shall prove the linear convergence rate under one of the following two alternative assumptions on algorithm parameters. One is that $\bar{\sigma}$ is sufficiently small, while c is sufficiently large (note that those are user-chosen parameters). The other is that

$$\text{(4.9)} \qquad \frac{1}{1+\eta_k} M_k \preceq M_{k+1} \preceq (1+\eta_k)M_k,\ \eta_k > 0\,\forall\, k, \quad \sum_{k=0}^{\infty} \eta_k < \infty,$$

which is a strengthening of the condition (1.4) used for global convergence. Asymptotically, (4.9) means that the matrices should not differ too much on subsequent iterations (a natural requirement in a neighborhood of a solution).

THEOREM 4.4. *In addition to the assumptions of Theorem* 4.2, *suppose that condition* (4.8) *is satisfied.*

Then, for sufficiently small choices of σ_k and sufficiently large choices of c_k, the sequence $\{z^k\}$ generated by Algorithm 3.1 *converges to an element of $T^{-1}(0)$ at a linear rate. If $c_k \to \infty$ and $\sigma_k \to 0$, the rate of convergence is superlinear.*

If condition (4.9) *holds, then for any choice of parameters $\bar{\sigma}$ and c, there exists $k_0 \in \mathbb{N}$ such that the sequence $\{z^k\}$ converges at the linear rate in the norm induced by $M^{-1}_{k_0}$.*

Proof. For each k, let τ_k, a_k, z^k be as defined in Algorithm 3.1, and let $x^k, w^k \in T(x^k)$ be the exact solution of the proximal system

$$\begin{cases} w \in T(x), \\ 0 = b_k M_k w + x - z^k, \end{cases}$$

where $b_k = \tau_k a_k$. Since $\hat{v}^k \in T^{\varepsilon_k}(\hat{z}^k)$, by Lemma 4.3 and the definition of a_k, it follows that

$$\begin{aligned} &\|x^k - \hat{z}^k\|^2_{M_k^{-1}} + b_k^2 \|M_k \hat{v}^k - M_k w^k\|^2_{M_k^{-1}} \\ &\le \|b_k M_k \hat{v}^k + \hat{z}^k - z^k\|^2_{M_k^{-1}} + 2b_k \varepsilon_k \\ &= \|b_k M_k \hat{v}^k + \hat{z}^k - z^k\|^2_{M_k^{-1}} \\ &\quad -2b_k \left(a_k \|M_k \hat{v}^k\|^2_{M_k^{-1}} + \langle M_k \hat{v}^k, z^k - \hat{z}^k \rangle_{M_k^{-1}} \right) \\ &= \|\hat{z}^k - z^k\|^2_{M_k^{-1}} + (\tau_k^2 - 2\tau_k) \|a_k M_k \hat{v}^k\|^2_{M_k^{-1}}. \end{aligned}$$

By using (4.2), we then obtain

$$\|x^k - \hat{z}^k\|^2_{M_k^{-1}} + b_k^2 \|M_k \hat{v}^k - M_k w^k\|^2_{M_k^{-1}} \le \left(\tau_k^2 - 2\tau_k + \frac{1}{(1-\sigma_k^2)^2} \right) \|a_k M_k \hat{v}^k\|^2_{M_k^{-1}}. \tag{4.10}$$

By using further the definitions of w^k and $\hat{v}^k$, this gives

$$\|x^k - \hat{z}^k\|^2_{M_k^{-1}} + \|x^k - z^{k+1}\|^2_{M_k^{-1}} \le \left(\tau_k^2 - 2\tau_k + \frac{1}{(1-\sigma_k^2)^2} \right) \|a_k M_k \hat{v}^k\|^2_{M_k^{-1}}. \tag{4.11}$$

From (4.10), we also have

$$\|M_k \hat{v}^k - M_k w^k\|^2_{M_k^{-1}} \le \left(1 - \frac{2}{\tau_k} + \frac{1}{\tau_k^2 (1-\sigma_k^2)^2} \right) \|M_k \hat{v}^k\|^2_{M_k^{-1}}.$$

Since $\hat{v}^k \to 0$ (see Proposition 4.1), the last relation implies that $w^k \to 0$. Hence, there exists $k_1 \in \mathbb{N}$ such that $\|w^k\| < L_2$ for all $k > k_1$. By (4.8), we then have

$$\operatorname{dist}(x^k, T^{-1}(0)) \le L_1 \|w^k\| \quad \forall k > k_1.$$

Therefore, for $k > k_1$,

$$\begin{aligned} \operatorname{dist}(x^k, T^{-1}(0))^2_{M_k^{-1}} &\le \frac{1}{\lambda_l} \operatorname{dist}(x^k, T^{-1}(0))^2 \le \frac{L_1^2}{\lambda_l} \|w^k\|^2 \\ &\le \frac{L_1^2}{\lambda_l^2} \|w^k\|^2_{M_k} = \frac{L_1^2}{\lambda_l^2} \|M_k w^k\|^2_{M_k^{-1}} \\ &= \frac{L_1^2}{\lambda_l^2 b_k^2} \|z^k - x^k\|^2_{M_k^{-1}}. \end{aligned} \tag{4.12}$$

Let $\bar{x}^k$ be the skewed projection of x^k onto $T^{-1}(0)$ under the norm induced by M_k^{-1}, i.e.,

$$\bar{x}^k := P_{T^{-1}(0), M_k^{-1}}(x^k).$$

Then, for $k > k_1$, we have

$$\begin{aligned} \operatorname{dist}(z^{k+1}, T^{-1}(0))_{M_k^{-1}} &\le \|z^{k+1} - \bar{x}^k\|_{M_k^{-1}} \\ &\le \|z^{k+1} - x^k\|_{M_k^{-1}} + \operatorname{dist}(x^k, T^{-1}(0))_{M_k^{-1}} \\ &\le \|z^{k+1} - x^k\|_{M_k^{-1}} + \frac{L_1}{\lambda_l b_k} \|z^k - x^k\|_{M_k^{-1}} \\ &\le \|z^{k+1} - x^k\|_{M_k^{-1}} + \frac{L_1}{\lambda_l b_k} \|x^k - \hat{z}^k\|_{M_k^{-1}} + \frac{L_1}{\lambda_l b_k} \|\hat{z}^k - z^k\|_{M_k^{-1}}, \end{aligned}$$

where the third inequality is by (4.12). By the Cauchy–Schwarz inequality, it holds that

$$\begin{aligned}
&\frac{L_1}{\lambda_l b_k}\|x^k - \hat{z}^k\|_{M_k^{-1}} + \|x^k - z^{k+1}\|_{M_k^{-1}} \\
&\le \sqrt{1 + \frac{L_1^2}{\lambda_l^2 b_k^2}}\sqrt{\|x^k - \hat{z}^k\|_{M_k^{-1}}^2 + \|x^k - z^{k+1}\|_{M_k^{-1}}^2} \\
&\le \sqrt{\left(1 + \frac{L_1^2}{\lambda_l^2 b_k^2}\right)\left(\tau_k^2 - 2\tau_k + \frac{1}{(1-\sigma_k^2)^2}\right)}\,\|a_k M_k \hat{v}^k\|_{M_k^{-1}},
\end{aligned}$$

where the second inequality follows from (4.11). By combining the latter relation with (4.13) and using also (4.2), we obtain

$$\begin{aligned}
&\operatorname{dist}(z^{k+1}, T^{-1}(0))_{M_k^{-1}} \\
&\le \left(\sqrt{\left(1 + \frac{L_1^2}{\lambda_l^2 b_k^2}\right)\left(\tau_k^2 - 2\tau_k + \frac{1}{(1-\sigma_k^2)^2}\right)} + \frac{L_1}{\lambda_l b_k (1-\sigma_k^2)}\right)\|a_k M_k \hat{v}^k\|_{M_k^{-1}}.
\end{aligned} \tag{4.13}$$

Define

$$\mu_k := \sqrt{\alpha_k^2 + 1}\sqrt{\beta_k^2 - 1} + \alpha_k \beta_k, \tag{4.14}$$

where

$$\alpha_k := \frac{L_1\left(1 + \sqrt{1 - (1-\sigma_k^2)^2}\right)}{\lambda_l c_k (1-\sigma_k^2)(1-\theta)} \le \frac{L_1\left(1 + \sqrt{1 - (1-\bar{\sigma}^2)^2}\right)}{\lambda_l c (1-\bar{\sigma}^2)(1-\theta)} =: \alpha, \tag{4.15}$$

$$\beta_k := \frac{1}{1-\sigma_k^2} \le \frac{1}{1-\bar{\sigma}^2} =: \beta. \tag{4.16}$$

With those definitions, by using (4.13) and (4.3), we conclude that

$$\operatorname{dist}(z^{k+1}, T^{-1}(0))_{M_k^{-1}} \le \mu_k \|a_k M_k \hat{v}^k\|_{M_k^{-1}}. \tag{4.17}$$

Let $\bar{z}^k := P_{T^{-1}(0), M_k^{-1}}(z^k)$. By Lemma 2.3, it holds that

$$\operatorname{dist}(z^k, T^{-1}(0))_{M_k^{-1}}^2 \ge \|\bar{z}^k - z^{k+1}\|_{M_k^{-1}}^2 + (1 - (1-\tau_k)^2) a_k^2 \|M_k \hat{v}^k\|_{M_k^{-1}}^2 \tag{4.18}$$

$$\ge \operatorname{dist}(z^{k+1}, T^{-1}(0))_{M_k^{-1}}^2 + (1-\theta^2) a_k^2 \|M_k \hat{v}^k\|_{M_k^{-1}}^2. \tag{4.19}$$

By using (4.17), we then conclude that

$$\operatorname{dist}(z^k, T^{-1}(0))_{M_k^{-1}}^2 \ge \left(1 + \frac{1-\theta^2}{\mu_k^2}\right)\operatorname{dist}(z^{k+1}, T^{-1}(0))_{M_k^{-1}}^2. \tag{4.20}$$

Therefore,

$$\operatorname{dist}(z^{k+1}, T^{-1}(0)) \le \frac{\mu_k \sqrt{\lambda_u}}{\sqrt{\lambda_l(\mu_k^2 + 1 - \theta^2)}}\operatorname{dist}(z^k, T^{-1}(0)). \tag{4.21}$$

By the definitions (4.15) and (4.16), by taking c_k sufficiently large we can make α_k arbitrarily small, and by taking σ_k sufficiently small we can make β_k arbitrarily close to one. By the definition (4.14), this means that we can make μ_k arbitrarily small, so that the scalar in the right-hand side of (4.21) is less than one. Then (4.21) shows that the sequence $\{\operatorname{dist}(z^k, T^{-1}(0))\}$ converges linearly to zero. Also, the inequality (4.19) shows that this sequence is Fejér-monotone with respect to the set $T^{-1}(0)$ (for the given norm). For Fejér-monotone sequences, linear convergence of $\{\operatorname{dist}(z^k, T^{-1}(0))\}$ is equivalent to the linear convergence rate of $\{z^k\}$ to its limit (see, e.g., [1]).

By the same argument as above, if $c_k \to \infty$ and $\sigma_k \to 0$, then $\mu_k \to 0$, and (4.21) shows a superlinear convergence rate.

Assume now that the condition (4.9) holds. Then

$$\begin{aligned}
\frac{1}{(1+\eta_k)} \operatorname{dist}(z, T^{-1}(0))^2_{M_k^{-1}} &= \inf_{y \in T^{-1}(0)} \frac{1}{(1+\eta_k)} \|z-y\|^2_{M_k^{-1}} \\
&\leq \inf_{y \in T^{-1}(0)} \|z-y\|^2_{M_{k+1}^{-1}} \\
&= (1+\eta_k) \operatorname{dist}(z, T^{-1}(0))^2_{M_{k+1}^{-1}} \\
&\leq \inf_{y \in T^{-1}(0)} (1+\eta_k) \|z-y\|^2_{M_k^{-1}} \\
&= (1+\eta_k) \operatorname{dist}(z, T^{-1}(0))^2_{M_k^{-1}}. \qquad (4.22)
\end{aligned}$$

Define

$$\mu = \sqrt{\alpha^2+1}\sqrt{\beta^2-1} + \alpha\beta,$$

with α and β given by (4.15) and (4.16), respectively. Note that $\mu > \mu_k$ for all k.

Since $\prod_{i=0}^{\infty}(1+\eta_i) < \infty$, there exists $k_2 \in \mathbb{N}$ such that

$$\prod_{i=k_2}^{\infty} (1+\eta_i) < \frac{\sqrt{\mu^2+1-\theta^2}}{\mu}.$$

From (4.20), by applying (4.22) consecutively, for any $k \geq k_0 := \max\{k_1, k_2\}$, we have

$$\begin{aligned}
&\left(\prod_{i=k_0}^{\infty} \frac{1}{(1+\eta_i)}\right) \operatorname{dist}(z^{k+1}, T^{-1}(0))^2_{M_{k_0}^{-1}} \\
&\leq \left(\prod_{i=k_0}^{k-1} \frac{1}{(1+\eta_i)}\right) \operatorname{dist}(z^{k+1}, T^{-1}(0))^2_{M_{k_0}^{-1}} \\
&\leq \operatorname{dist}(z^{k+1}, T^{-1}(0))^2_{M_k^{-1}} \\
&\leq \frac{\mu^2}{\mu^2+1-\theta^2} \operatorname{dist}(z^k, T^{-1}(0))^2_{M_k^{-1}} \\
&\leq \left(\prod_{i=k_0}^{k-1} (1+\eta_i)\right) \frac{\mu^2}{\mu^2+1-\theta^2} \operatorname{dist}(z^k, T^{-1}(0))^2_{M_{k_0}^{-1}} \\
&\leq \left(\prod_{i=k_0}^{\infty} (1+\eta_i)\right) \frac{\mu^2}{\mu^2+1-\theta^2} \operatorname{dist}(z^k, T^{-1}(0))^2_{M_{k_0}^{-1}}.
\end{aligned}$$

In particular,

$$\operatorname{dist}(z^{k+1}, T^{-1}(0))_{M_{k_0}^{-1}} \leq \nu \operatorname{dist}(z^k, T^{-1}(0))_{M_{k_0}^{-1}},$$

where

$$\nu := \frac{\mu}{\sqrt{\mu^2 + 1 - \theta^2}} \prod_{i=k_0}^{\infty} (1 + \eta_i) < 1,$$

as claimed. □

5. A variable metric proximal Newton method. In this section, we show how the proposed variable metric approach can be used to obtain a computational advantage when solving a system of monotone differentiable equations

$$F(x) = 0, \tag{5.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$. Problems of this type appear, for example, in smooth multiplier methods for monotone complementarity problems [14]. We start with describing the method and giving its theoretical justification and then report on our numerical experiments.

5.1. Description and justification of the method. In [25, 29], it has been shown that hybrid inexact proximal point schemes (with a fixed metric) can be used to construct Newton methods for monotone problems with a very attractive combination of global and local convergence properties. In particular, global convergence from any starting point to a solution is guaranteed, regardless of any degeneracy along the trajectory, which is not true in the case of more standard merit function-based globalizations that can get stuck at stationary points of the function that are not global minimizers. Fast local convergence for nondegenerate problems is also preserved, in a natural way. We refer the reader to [25, 29] for more detailed discussion.

When the Newton step is computed for the proximal subproblem (with the fixed metric $M_k = I$)

$$c_k F(z) + (z - z^k) = 0,$$

as in [25], one needs to solve the system of linear equations

$$c_k F(z^k) + (c_k \nabla F(z^k) + I)d = 0, \tag{5.2}$$

with respect to $d \in \mathbb{R}^n$. The crucial point is that, under natural assumptions, this single Newton step is enough to obtain an acceptable approximate solution of the proximal subproblem. Note that the above system is, in general, *asymmetric*. For future comparison, note that to compute LU factorization of the matrix $c_k \nabla F(z^k) + I$ and then the solution d^k, the number of arithmetic operations required is $2(n^3/3 + n^2)$. If to solve the linear system one uses instead of matrix factorization the conjugate gradient method, calculation of $(\nabla F(z^k))^\top \nabla F(z^k)$ is needed. Apart from extra computational cost (which is not negligeable when n is large), the latter is in general a dense matrix even when $\nabla F(z^k)$ is sparse. In what follows, we show how choosing a special variable metric can reduce the number of calculations in the case of using matrix factorizations and can preserve sparsity if the conjugate gradient method is used.

The idea is to choose a metric in such a way that, instead of solving a general asymmetric linear system, we will have to solve one triangular system and one symmetric system (with a positive definite matrix). As we shall see, this has a number of advantages.

Consider the proximal subproblem

$$0 = c_k M_k F(z) + (z - z^k). \tag{5.3}$$

We shall compute the Newton step for its equivalent formulation

$$0 = c_k F(z) + A_k(z - z^k),$$

where A_k plays the role of the inverse of M_k (no matrices are actually inverted, of course; we simply choose A_k and work with it throughout, as explained next). The Newton step for the latter equation is given by

$$c_k F(z^k) + (c_k \nabla F(z^k) + A_k)d^k = 0. \tag{5.4}$$

In what follows, we shall show that, with proper choices of parameters, the point

$$y^k = z^k + d^k \tag{5.5}$$

is an acceptable approximate solution of (5.3), in the sense of Algorithm 3.1 (even more specifically, in the sense of Proposition 3.1). Then the next iterate is given by

$$z^{k+1} = z^k - c_k M_k F(y^k),$$

which can be implemented as solving the system of linear equations

$$c_k F(y^k) + A_k s = 0, \tag{5.6}$$

with respect to $s \in \mathbb{R}^n$, and setting

$$z^{k+1} = z^k + s^k. \tag{5.7}$$

As A_k we suggest to take the symmetrization of the upper triangular part of the matrix $-c_k \nabla F(z^k)$ with appropriate diagonal elements, so that it is positive definite. One choice is

$$(A_k)_{i,j} := \begin{cases} -c_k(\nabla F(z^k))_{i,j} & \text{for} \quad i < j, \\ (A_k)_{j,i} & \text{for} \quad i > j, \\ 1 + \sum_{i \neq j} |(A_k)_{i,j}| & \text{for} \quad i = j. \end{cases} \tag{5.8}$$

Since A_k is symmetric and strictly diagonally dominant, it is positive definite by the Gerschgorin theorem [12, Theorem 3.5.9], and

$$\lambda_{min}(A_k) \geq 1. \tag{5.9}$$

The proposed implementation, therefore, consists of solving the linear system (5.4) with the triangular matrix $c_k \nabla F(z^k) + A_k$ and the linear system (5.6) with the symmetric positive definite matrix A_k. If the Cholesky factorization is used for the latter, the total cost of the iteration is $n^3/3 + 2n^2 + n^2/2$ arithmetic operations. The savings compared to the fixed metric (asymmetric) implementation discussed above amounts to $n^2(n/3 - 1/2)$, which is significant for large n. If instead of matrix factorization the conjugate gradient method is used to solve (5.6), it is important that it works directly with the symmetric matrix A_k, which is sparse if $\nabla F(z^k)$ is also. Recall that, in the case of solving the asymmetric system, the method has to work with $(c_k \nabla F(z^k) + I)^\top (c_k \nabla F(z^k) + I)$, which is in general dense even when $\nabla F(z^k)$ is sparse.

To validate our proposal, it remains to show that the single Newton step defined by (5.4) produces a point acceptable by the approximation criterion of Algorithm 3.1 and that this strategy does not increase too much the overall number of iterations of the method as compared to the asymmetric fixed metric implementation. We deal with the first issue next and then present some numerical experiments to address the second.

Let $M_k = A_k^{-1}$. By (5.9), we have

$$\lambda_{max}(M_k) \le 1. \tag{5.10}$$

In particular, we can take $\lambda_u = 1$ in Algorithm 3.1. Since d^k is the solution of the linear system (5.4), we have

$$d^k = y^k - z^k = -c_k M_k F(z^k) - c_k M_k \nabla F(z^k) d^k. \tag{5.11}$$

To prove the claim that this Newton step is sufficient to solve the proximal subproblem within the required tolerance, we have to show that

$$\|c_k M_k F(y^k) + d^k\|^2_{M_k^{-1}} \le \sigma_k^2 \left(\|c_k M_k F(y^k)\|^2_{M_k^{-1}} + \|d^k\|^2_{M_k^{-1}} \right). \tag{5.12}$$

Let ∇F be Lipschitz-continuous with modulus $\gamma > 0$ (on the bounded set containing the sequences $\{z^k\}$ and $\{y^k\}$, whose boundedness has been already established). It then holds that

$$\|F(y^k) - F(z^k) - \nabla F(z^k) d^k\| \le \frac{\gamma}{2} \|d^k\|^2.$$

Since it follows from (5.11) that

$$-c_k F(z^k) - c_k \nabla F(z^k) d^k = M_k^{-1} d^k,$$

we obtain

$$\|c_k F(y^k) + M_k^{-1} d^k\| \le \frac{\gamma c_k}{2} \|d^k\|^2. \tag{5.13}$$

Furthermore, by recalling (5.10), we have

$$\|c_k F(y^k) + M_k^{-1} d^k\|^2 \ge \|c_k F(y^k) + M_k^{-1} d^k\|^2_{M_k} = \|c_k M_k F(y^k) + d^k\|^2_{M_k^{-1}}. \tag{5.14}$$

Also, by using (5.10) and (5.11), we obtain

$$\begin{aligned} \|d^k\|^2 &\le \|d^k\|^2_{M_k^{-1}} \\ &= \langle d^k, M_k^{-1}(-c_k M_k F(z^k) - c_k M_k \nabla F(z^k) d^k) \rangle \\ &= -c_k \langle d^k, F(z^k) \rangle - c_k \langle d^k, \nabla F(z^k) d^k \rangle \\ &\le c_k \|d^k\| \|F(z^k)\|, \end{aligned}$$

where we have used the fact that $\nabla F(z^k)$ is positive semidefinite (by the monotonicity of F). Hence,

$$\|d^k\| \le c_k \|F(z^k)\|.$$

By combining this relation with (5.13) and (5.14), we conclude that

$$\|c_k M_k F(y^k) + d^k\|_{M_k^{-1}} \le \frac{\gamma c_k^2 \|F(z^k)\|}{2} \|d^k\| \le \frac{\gamma c_k^2 \|F(z^k)\|}{2} \|d^k\|_{M_k^{-1}},$$

where (5.10) was also taken into account. Therefore, by choosing the regularization parameter

$$0 < c_k \le \frac{\sqrt{2}\sigma_k}{\sqrt{\gamma \|F(z^k)\|}},$$

we obtain (5.12). This analysis also shows that we are in the setting of Proposition 3.1, so that the step $z^{k+1} = z^k + c_k M_k F(y^k)$ is admissible (implemented above as solving the linear system (5.6)).

If an estimate for the Lipschitz constant γ of ∇F is not available, c_k can be obtained by an Armijo-type line-search procedure. Alternatively, instead of making one Newton step for each subproblem, we can make several, until the relative error approximation criterion is satisfied. In our computational experience, however, one Newton step was always enough. Moreover, by assuming the nonsingularity of ∇F at the solution, for k large enough one can take $c_k = \frac{\sqrt{2}\sigma_k}{\sqrt{\|F(z^k)\|}}$, without any line search, and make a single Newton step. The superlinear rate of convergence can be established by analysis analogous to [25].

5.2. Numerical experiments. We have compared the proximal Newton methods, with a fixed metric and a variable metric, on the following examples.

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be given by

$$F(z) = \tilde{F}(z) + Hz,$$

where

$$\tilde{F}_i(z) = \frac{1 + (-1)^{i+1}}{2} f(z_i),$$

$f : \mathbb{R} \to \mathbb{R}$ is a monotone function with a Lipschitz-continuous derivative, and H is the $n \times n$ matrix given by

$$(H)_{i,j} = \begin{cases} n/2 & \text{for } i = j = 1, \\ 5n & \text{for } i = 1 \text{ and } j = n, \\ -5n & \text{for } i = n \text{ and } j = 1, \\ n + i - 1 & \text{for } i = j \text{ and } i \notin \{1, n\}, \\ 1 & \text{for } j = n \text{ and } i \notin \{1, n\}, \\ 1 & \text{for } j < i \text{ and } i \neq n, \\ -1 & \text{for } i = n \text{ and } j \notin \{1, n\}, \\ 0 & \text{elsewhere.} \end{cases}$$

It can be seen that H is positive semidefinite (because $(H + H^\top)/2$ is diagonally dominant), but it is not positive definite (because $e_n^\top H e_n = 0$, where e_n is the nth vector of the canonical basis). This fact and the monotonicity of f imply that F is monotone. Note that, for $n = 2k$ with $k \in \mathbb{N}$, F is not strictly monotone, even if f is strictly monotone.

It can be seen that its Jacobian $\nabla F(z)$ is Lipschitz-continuous with the same Lipschitz constant as f', and, for any $z \in \mathbb{R}^n$, $\nabla F(z)$ is a nonsymmetric matrix, with a sparse upper triangular part.

We have coded both the Newton proximal method (NPM) and the variable metric Newton proximal method (VMNPM) by using Scilab 4.0 (INRIA-ENPC, see www.scilab.org). An iteration of NPM consists of solving the system of equations (5.2), while VMNPM is the procedure given by (5.4)–(5.7), with A_k defined in (5.8). For both methods, the regularization parameter is taken as $c_k = \sqrt{2/\|F(z^k)\|}$.

In the case of solving linear systems by matrix factorization, the comparison is exactly as predicted by the arithmetic operations counts, mentioned above. The variable metric approach requires more iterations, but already for moderate dimensions

TABLE 5.1

$f(x) = x + \exp(-x^2)$

Dim	NPM			VMNPM			T_1/T_2
	Iter	T_1	$\|F\|$	Iter	T_2	$\|F\|$	
100	4	0.16	3.98e-008	20	0.20	8.12e-008	0.77
300	4	1.34	3.40e-008	22	1.11	3.57e-008	1.21
500	4	4.16	4.04e-008	22	3.59	4.13e-008	1.16
700	4	9.06	4.32e-008	22	7.28	6.74e-008	1.24
900	4	16.22	4.88e-008	23	12.72	4.62e-008	1.28
1100	4	26.16	5.37e-008	23	19.06	5.10e-008	1.37
1300	4	39.00	6.81e-008	23	26.38	5.05e-008	1.48
1500	4	55.45	6.94e-008	23	35.13	4.65e-008	1.58
1700	4	75.94	9.14e-008	23	44.84	4.39e-008	1.69
1900	4	100.70	9.59e-008	23	55.91	5.12e-008	1.80

$f(x) = 2\arctan(x+1)$

Dim	NPM			VMNPM			T_1/T_2
	Iter	T_1	$\|F\|$	Iter	T_2	$\|F\|$	
100	4	0.13	7.42e-008	20	0.17	6.63e-008	0.73
300	4	1.36	4.80e-008	22	1.16	6.50e-008	1.18
500	4	4.38	5.77e-008	23	3.78	1.52e-008	1.16
700	4	9.22	6.42e-008	23	7.91	2.25e-008	1.17
900	4	16.45	6.51e-008	23	13.22	5.09e-008	1.24
1100	4	26.38	6.71e-008	23	19.66	9.95e-008	1.34
1300	4	39.27	7.14e-008	24	28.55	3.95e-008	1.38
1500	4	55.78	8.37e-008	24	37.89	3.18e-008	1.47
1700	4	76.92	9.02e-008	24	49.11	2.79e-008	1.57
1900	4	101.33	1.15e-007	24	60.64	3.97e-008	1.67

$f(x) = \frac{1}{2}x\sqrt{x^2+5} + \frac{5}{2}\ln(x + \sqrt{x^2+5})$

Dim	NPM			VMNPM			T_1/T_2
	Iter	T_1	$\|F\|$	Iter	T_2	$\|F\|$	
100	4	0.14	8.04e-008	20	0.20	9.22e-008	0.69
300	4	1.36	5.71e-008	23	1.19	1.78e-008	1.14
500	4	4.22	6.93e-008	23	3.80	5.05e-008	1.11
700	4	9.13	7.99e-008	24	8.08	5.35e-008	1.13
900	4	16.38	8.47e-008	24	13.30	4.05e-008	1.23
1100	4	26.31	8.51e-008	24	19.78	4.18e-008	1.33
1300	4	39.25	8.98e-008	24	27.94	9.05e-008	1.40
1500	4	55.73	9.75e-008	25	38.22	5.00e-008	1.46
1700	4	76.27	1.08e-007	25	48.89	3.49e-008	1.56
1900	4	101.08	1.18e-007	25	60.77	2.64e-008	1.66

(say, $n = 500$) the cheaper iteration cost starts to pay off, with the advantage growing with n, as predicted by the operations counts. We shall not report this comparison here, for the sake of brevity.

Instead, we shall report results for solving the linear systems by the conjugate gradient method. The Scilab `sparse` utility is used to take advantage of structure. As already pointed out, the matrix $(c_k \nabla F(z^k) + I)^\top (c_k \nabla F(z^k) + I)$ in the fixed metric approach is essentially dense, while the matrix A_k in the variable metric approach preserves structure.

The comparison of the respective performances, for three different choices of f, on a 1.66 GHz, 512 MB RAM Intel Centrino processor PC is shown in Table 5.1. The first column shows the dimension, then the number of iterations, the computation time in seconds, and the norm of the residual at termination. The last column shows the ratio between the computational times.

Figure 5.1 compares the computational time evolution for both methods. The performance of the NPM is almost the same for the three functions involved, and

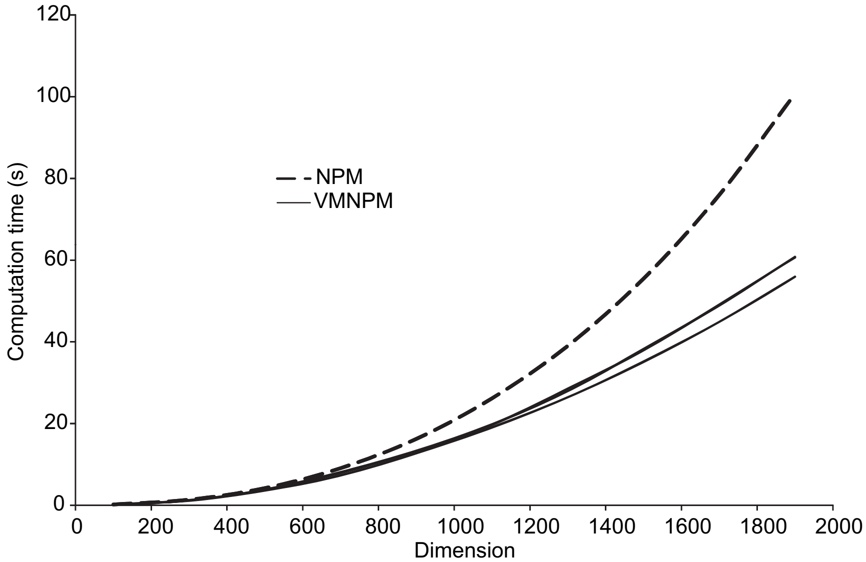

Fig. 5.1. *Performance comparison.*

it is not distinguishable in the graphic scale, while the performance of the VMNPM presents little variations for the three examples. As we have anticipated, the variable metric proximal Newton method outperforms the Newton proximal method already for moderate dimensions, with the advantage becoming more and more significant as n grows.

REFERENCES

[1] H. H. Bauschke, *Projection algorithms: Results and open problems*, in Inherently Parallel Algorithms in Feasibility and Optimization and their Applications, Stud. Comput. Math. 8, D. Butnariu, Y. Censor, and S. Reich, eds., Elsevier Science B. V., Amsterdam, 2001, pp. 11–22.

[2] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal, *A family of variable metric proximal point methods*, Math. Program., 68 (1995), pp. 15–47.

[3] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects*, Springer-Verlag, Berlin, 2003.

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

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

[6] R. S. Burachik, C. A. Sagastizábal, and B. F. Svaiter, *ε-Enlargements of maximal monotone operators: Theory and applications*, in Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Norwell, MA, 1999, pp. 25–44.

[7] J. V. Burke and M. Qian, *A variable metric proximal point algorithm for monotone operators*, SIAM J. Control Optim., 37 (1999), pp. 353–375.

[8] J. V. Burke and M. Qian, *On the local super-linear convergence of a matrix secant implementation of the variable metric proximal point algorithm for monotone operators*, in Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Norwell, MA, 1999, pp. 317–334.

[9] J. V. Burke and M. Qian, *On the superlinear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating*, Math. Program., 88 (2000), pp. 157–181.

[10] X. Chen and M. Fukushima, *Proximal quasi-Newton methods for nondifferentiable convex optimization*, Math. Program., 85 (1999), pp. 313–334.
[11] R. Cominetti, *Coupling the proximal point algorithm with approximation methods*, J. Optim. Theory Appl., 95 (1997), pp. 581–600.
[12] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
[13] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.
[14] J. Eckstein and M. C. Ferris, *Smooth methods of multipliers for complementarity problems*, Math. Program., 86 (1999), pp. 65–90.
[15] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., North–Holland, Amsterdam, 1983, pp. 299–331.
[16] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, 1993.
[17] C. Lemaréchal and C. Sagastizábal, *Variable metric bundle methods: From conceptual to implementable forms*, Math. Program., 76 (1997), pp. 393–410.
[18] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.
[19] B. Martinet, *Regularisation d'inequations variationelles par approximations successives*, Revue Française d'Informatique et de Recherche Opérationelle, 4 (1970), pp. 154–159.
[20] L. Qi and X. Chen, *A preconditioning proximal Newton's method for nondifferentiable convex optimization*, Math. Program., 76 (1995), pp. 411–430.
[21] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.
[22] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.
[23] R. T. Rockafellar and J.-B. Wets, *Variational Analysis*, Springer-Verlag, New York, 1997.
[24] M. V. Solodov, *A class of decomposition methods for convex optimization and monotone variational inclusions via the hybrid inexact proximal point framework*, Optim. Methods Softw., 19 (2004), pp. 557–575.
[25] M. V. Solodov and B. F. Svaiter, *A globally convergent inexact Newton method for systems of monotone equations*, in Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, M. Fukushima and L. Qi, eds., Kluwer Academic Publishers, Norwell, MA, 1999, pp. 355–369.
[26] 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 (1999), pp. 323–345.
[27] M. V. Solodov and B. F. Svaiter, *A hybrid projection–proximal point algorithm*, J. Convex Anal., 6 (1999), pp. 59–70.
[28] M. V. Solodov and B. F. Svaiter, *Error bounds for proximal point subproblems and associated inexact proximal point algorithms*, Math. Program., 88 (2000), pp. 371–389.
[29] M. V. Solodov and B. F. Svaiter, *A truly globally convergent Newton-type method for the monotone nonlinear complementarity problem*, SIAM J. Optim., 10 (2000), pp. 605–625.
[30] M. V. Solodov and B. F. Svaiter, *A unified framework for some inexact proximal point algorithms*, Numer. Funct. Anal. Optim., 22 (2001), pp. 1013–1035.
[31] J. E. Spingarn, *Applications of the method of partial inverses to convex programming*, Math. Program., 32 (1985), pp. 199–223.
[32] P. Tossings, *The perturbed proximal point algorithm and some of its applications*, Appl. Math. Optim., 29 (1994), pp. 125–159.
[33] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim., 29 (1991), pp. 119–138.
[34] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446.