SIAM J. CONTROL OPTIM.
Vol. 43, No. 2, pp. 731–742

PROXIMAL METHODS FOR COHYPOMONOTONE OPERATORS*

PATRICK L. COMBETTES† AND TEEMU PENNANEN‡

Abstract. Conditions are given for the viability and the weak convergence of an inexact, relaxed proximal point algorithm for finding a common zero of countably many cohypomonotone operators in a Hilbert space. In turn, new convergence results are obtained for an extended version of the proximal method of multipliers in nonlinear programming.

Key words. cohypomonotone operator, common zero problem, hypomonotone operator, method of multipliers, nonlinear programming, proximal point method, weak convergence

AMS subject classifications. 47H04, 65K10, 90C26, 90C30

DOI. 10.1137/S0363012903427336

1. Introduction.

Let $\mathcal{H}$ be a real Hilbert space with scalar product $\langle \cdot , \cdot \rangle$, norm $\| \cdot \|$, and distance d. A basic problem in applied mathematics and optimization is to find a zero of a maximal monotone operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$, that is, a point $x \in \mathcal{H}$ such that $0 \in Ax$ [22, 23, 29]. Assuming $0 \in \operatorname{ran} A$, since the resolvent $(\mathrm{Id} + A)^{-1}$ of A is a firmly nonexpansive operator with fixed point set $A^{-1}(0)$, a zero of A can be constructed iteratively through the recursion

$$(\forall n \in \mathbb{N}) \;\; x_{n+1} = (\mathrm{Id} + A)^{-1} x_n. \tag{1.1}$$

Indeed, since an operator T is nonexpansive if and only if its average $(T + \mathrm{Id})/2$ is firmly nonexpansive [28, Lemma 1.1], it follows from [19, Theorem 3] that, for any $x_0 \in \mathcal{H}$, the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the successive approximations (1.1) converges weakly to a zero of A (see also [17] for a special case). More generally, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $\inf_{n\in\mathbb{N}} \gamma_n > 0$ and let $(e_n)_{n\in\mathbb{N}}$ be an absolutely summable sequence in $\mathcal{H}$. Then, for every $x_0 \in \mathcal{H}$, the so-called proximal point iterations $x_{n+1} = (\mathrm{Id} + \gamma_n A)^{-1} x_n + e_n$ converge weakly to a zero of A [22, Theorem 1] (see also [4] for further analysis). This result was shown in [11, Theorem 3] to remain true for the relaxed proximal iterations

$$(\forall n \in \mathbb{N}) \;\; x_{n+1} = x_n + \lambda_n \big((\mathrm{Id} + \gamma_n A)^{-1} x_n + e_n - x_n \big), \tag{1.2}$$

where $(\lambda_n)_{n\in\mathbb{N}}$ lies in $[\varepsilon, 2 - \varepsilon]$ for some arbitrary $\varepsilon \in]0, 1[$. A further extension was proposed in [2, Corollary 6.1(i)] (for $e_n \equiv 0$) and then in [7, Theorem 6.9(i)], where weak convergence to a common zero of a countable family of maximal monotone operators $(A_i)_{i\in I}$ was established for the iterations

$$(\forall n \in \mathbb{N}) \;\; x_{n+1} = x_n + \lambda_n \big((\mathrm{Id} + \gamma_n A_{\mathrm{i}(n)})^{-1} x_n + e_n - x_n \big), \tag{1.3}$$

where $\mathrm{i}\colon \mathbb{N} \to I$ sweeps through the indices with some regularity. It will be convenient to cast this algorithm in the following more general framework.

*Received by the editors May 7, 2003; accepted for publication (in revised form) December 1, 2003; published electronically August 4, 2004.
http://www.siam.org/journals/sicon/43-2/42733.html

†Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France (plc@math.jussieu.fr).
‡Department of Management Science, Helsinki School of Economics, 00101 Helsinki, Finland (pennanen@hkkk.fi).

ALGORITHM 1.1. *Let $(A_i)_{i\in I}$ be a countable family of set-valued operators from $\mathcal{H}$ to $\mathcal{H}$, let $(\gamma_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ be sequences in $]0,+\infty[$, let $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{H}$, let i be a mapping from $\mathbb{N}$ to I, and let x_0 be a point in $\mathcal{H}$. A sequence $(x_n)_{n\in\mathbb{N}}$ is constructed according to the updating rule*

$$(\forall n \in \mathbb{N}) \;\; x_{n+1} = x_n + \lambda_n(x_{n+\frac{1}{2}} + u_n - x_n), \tag{1.4}$$

where $x_{n+\frac{1}{2}}$ is a solution to the inclusion

$$v_n \in x_{n+\frac{1}{2}} - x_n + \gamma_n A_{\mathrm{i}(n)} x_{n+\frac{1}{2}}. \tag{1.5}$$

In the case of maximal monotone operators, the weak convergence properties of Algorithm 1.1 are summarized in the next theorem, which is derived from a result of [7]. This theorem captures the weak convergence results of [2, 11, 17, 22] for the proximal point algorithm, as well as standard results on the weak convergence of sequential projection methods for convex feasibility problems, such as those of [5, 6, 13], when the operators are taken to be normal cones to closed convex sets.

THEOREM 1.2. *Suppose that in Algorithm 1.1 the following conditions are satisfied:*

- (i) (a) *For every $i \in I$, A_i is maximal monotone;*
  (b) $S = \bigcap_{i\in I} A_i^{-1}(0) \neq \varnothing$.
- (ii) $(\forall i \in I)(\exists\, M_i \in \mathbb{N} \smallsetminus \{0\})(\forall n \in \mathbb{N}) \;\; i \in \{\mathrm{i}(n), \ldots, \mathrm{i}(n + M_i - 1)\}$.
- (iii) $\inf_{n\in\mathbb{N}} \gamma_n > 0$.
- (iv) $(\exists\, \varepsilon \in]0,1[)(\forall n \in \mathbb{N}) \;\; \varepsilon \leq \lambda_n \leq 2 - \varepsilon$.
- (v) $\sum_{n\in\mathbb{N}} \|u_n\| < +\infty$ *and* $\sum_{n\in\mathbb{N}} \|v_n\| < +\infty$.

Then every orbit generated by Algorithm 1.1 converges weakly to a point in S.

Proof. For every $n \in \mathbb{N}$, set

$$e_n = u_n + (\mathrm{Id} + \gamma_n A_{\mathrm{i}(n)})^{-1}(x_n + v_n) - (\mathrm{Id} + \gamma_n A_{\mathrm{i}(n)})^{-1} x_n. \tag{1.6}$$

Then (1.4)–(1.5) coincides with (1.3), which is itself a special case of [7, Algorithm 6.7] (obtained by taking $I^{(1)} = I^{(2)} = \varnothing$ and $(I_n)_{n\in\mathbb{N}} = (\{\mathrm{i}(n)\})_{n\in\mathbb{N}}$ there). On the other hand, since the resolvents $\big((\mathrm{Id} + \gamma_n A_{\mathrm{i}(n)})^{-1}\big)_{n\in\mathbb{N}}$ are nonexpansive [1, Proposition 3.5.3], we obtain

$$(\forall n \in \mathbb{N}) \;\; \|e_n\| \leq \|u_n\| + \|v_n\|. \tag{1.7}$$

Hence, (v) implies that $\sum_{n\in\mathbb{N}} \|e_n\| < +\infty$ and the claim therefore follows at once from [7, Theorem 6.9(i)]. ◻

Remark 1.3. The sequences $(v_n)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ model errors at various steps of the execution of the iterations, thereby allowing for some tolerance in the numerical implementation of the algorithm. It is clear from the above proof that, in the presence of monotone operators, the errors $(v_n)_{n\in\mathbb{N}}$ can easily be absorbed in the errors $(u_n)_{n\in\mathbb{N}}$ and are, in this sense, redundant. However, since our ultimate goal is to investigate the behavior of Algorithm 1.1 with nonmonotone operators, the use of two error sequences is required to obtain a more flexible algorithmic model. An illustration of how condition (v) can be checked in practice is provided in section 4 (see Remark 4.2).

Extensions of the basic proximal iterations (1.1) have also been investigated in another direction, namely, by relaxing the monotonicity requirements on A. The motivation for this line of work stems from the fact that proximal iterations have been observed to converge to zeros of nonmonotone operators in certain numerical experiments, e.g., [12]. Attempts to explain this behavior in the case of general variational inclusions can be traced back to [26], where a convergence proof is given which does not assume monotonicity. However, the assumptions made in that early work are rather stringent as they impose, essentially, that the inverse of the operator be differentiable at the origin with a monotone derivative.

Relaxing the monotonicity property of an operator is equivalent to relaxing the monotonicity property of its inverse. In some applications, however, it is more natural to work directly with the inverse. For instance, since multiplier methods are based on applying the proximal algorithm to a dual formulation of the original problem, it is more pertinent to impose relaxed monotonicity conditions on the inverse of the operator. This observation was the starting point of the investigation proposed in [20], where local convergence is analyzed under the condition that the mapping be cohypomonotone, i.e., that its inverse be hypomonotone (see Definition 2.2). The analysis of [20] is incomplete, however, at least in the sense that it assumes that the proximal steps can be computed exactly. This is an unrealistic assumption in most practical applications. In [14], an effort was made to remove this assumption by investigating the convergence in the case of inexact computations under a so-called relative error criterion. The analysis of [14] requires that the values of the operator outside a certain neighborhood be discarded. However, since this neighborhood is usually unknown in concrete applications, the applicability of this conceptual analysis is limited.

The goal of this paper is to unify and extend various convergence results on proximal iterations, by investigating the asymptotic behavior of Algorithm 1.1 when applied to a family of cohypomonotone operators. Such operators are discussed in section 2. Our main result is presented in section 3, where local viability and weak convergence conditions are established for Algorithm 1.1. An application to nonlinear programming is presented in section 4, where local convergence of a relaxed inexact proximal method of multipliers is proven for a nonconvex problem.

Throughout, $B(x;\eta)$ denotes the closed ball of center $x \in \mathcal{H}$ and extended radius $\eta \in \,]0,+\infty]$; d_C the distance function to a nonempty set $C \subset \mathcal{H}$; P_C the projection operator onto a nonempty closed convex set $C \subset \mathcal{H}$; and N_C its normal cone map. $\operatorname{Fix} T$ the set of fixed points of an operator T, $\operatorname{dom} T$ its domain, $\operatorname{ran} T$ its range, and $\operatorname{gph} T$ its graph. The complement of a set C is denoted by $\complement C$.

2. Cohypomonotone operators. Our goal is to prove the local convergence of Algorithm 1.1 under a relaxed monotonicity assumption on the operators $(A_i)_{i\in I}$ that we now define.

DEFINITION 2.1. *Let U be a subset of $\mathcal{H}^2$. The U-localization of an operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ is the operator denoted by $A|^U\colon \mathcal{H} \rightrightarrows \mathcal{H}$ whose graph is* $\operatorname{gph}(A|^U) = U \cap \operatorname{gph} A$.

DEFINITION 2.2. *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$, $\rho \in [0,+\infty[$, and $U \subset \mathcal{H}^2$. Then A is* [maximal] ρ-hypomonotone *on U if there exists an operator $\bar{A}\colon \mathcal{H} \rightrightarrows \mathcal{H}$ such that $\bar{A} + \rho\,\mathrm{Id}$ is [maximal] monotone and $\bar{A}|^U = A|^U$. The operator A is* [maximal] ρ-cohypomonotone *on U if A^{-1} is [maximal] ρ-hypomonotone on U.*

The above definition of ρ-hypomonotonicity on a set is related to the pointwise notion of hypomonotonicity of [25, Example 12.28] and [8] as follows: If W is a neighborhood of $x \in \mathcal{H}$ and there exists $\rho \in [0,+\infty[$ such that A is ρ-hypomonotone on $W \times \mathcal{H}$, then A is hypomonotone at x in the sense of [8, 25].

Maximal hypomonotonicity has been studied extensively in the variational analysis literature. Thus, classes of functions with hypomonotone subdifferentials have been investigated in various settings [8, 24, 25, 27]. Interesting connections between hypomonotonicity and Aubin continuity, Lipschitz continuity, and strict graphical derivatives have also been found [15, 16]. On the other hand, maximal hypomonotonicity is a less stringent requirement than imposing the existence of Lipschitz localizations. The latter has been studied in the context of variational inequality and nonlinear programming problems, e.g., [9, 10, 15, 16]. For completeness, we provide a simple proof of this important fact.

LEMMA 2.3. *Suppose that $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ has a Lipschitz localization at a point $(x, y) \in \operatorname{gph} A$; that is, there exist open sets $X \ni x$ and $Y \ni y$ such that the mapping $z \mapsto A(z) \cap Y$ is single-valued and ρ-Lipschitz continuous on X. Then A is maximal ρ-hypomonotone on $X \times Y$.*

Proof. Set $\widetilde{A} = A|^{X\times Y} + \rho\operatorname{Id}$ and take $(u, v) \in X^2$. Then, by Cauchy–Schwarz,

$$\left\langle u - v\,, \widetilde{A}u - \widetilde{A}v\right\rangle \geq \|u - v\| \left(\rho\|u - v\| - \left\|A|^{X\times Y}(u) - A|^{X\times Y}(v)\right\|\right) \geq 0. \tag{2.1}$$

Hence, $\widetilde{A}$ is monotone. Let A' be a maximal monotone extension of $\widetilde{A}$ and set $\bar{A} = A' - \rho\operatorname{Id}$. Then $\bar{A} + \rho\operatorname{Id}$ is maximal monotone and, to complete the proof, it suffices to show that $\bar{A}|^{X\times Y} = A|^{X\times Y}$. By construction, $\operatorname{gph}(A|^{X\times Y}) \subset \operatorname{gph}(\bar{A}|^{X\times Y})$. Conversely, take $(\bar{x}, \bar{y}) \in \operatorname{gph} \bar{A}|^{X\times Y}$ and let $z = \bar{y} - A|^{X\times Y}(\bar{x})$. Then $(\bar{x}, \bar{y} + \rho\bar{x}) \in \operatorname{gph} A'$ and, since X is open, we have $\bar{x} + \varepsilon z \in X$ for $\varepsilon > 0$ sufficiently small. Since $\operatorname{gph}(A|^{X\times Y} + \rho\operatorname{Id}) \subset \operatorname{gph} A'$ and A' is monotone, we have

$$\begin{aligned} 0 &\leq \left\langle \bar{x} + \varepsilon z - \bar{x}\,, A|^{X\times Y}(\bar{x} + \varepsilon z) + \rho(\bar{x} + \varepsilon z) - (\bar{y} + \rho\bar{x})\right\rangle \\ &= \left\langle \varepsilon z\,, A|^{X\times Y}(\bar{x} + \varepsilon z) + \rho\varepsilon z - \bar{y}\right\rangle. \end{aligned} \tag{2.2}$$

Dividing by ε and letting $\varepsilon \downarrow 0^+$, the continuity of $A|^{X\times Y}$ gives $0 \leq -\left\|A|^{X\times Y}(\bar{x}) - \bar{y}\right\|^2$, whence $(\bar{x}, \bar{y}) \in \operatorname{gph} A|^{X\times Y}$. ∎

The relevance of cohypomonotonicity in proximal methods hinges on the following identity.

LEMMA 2.4. *Let $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ and let $(\gamma, \rho) \in \mathbb{R}^2$, where $\gamma \neq 0$. Then*

$$\operatorname{Id} + \left(1 - \frac{\rho}{\gamma}\right)\left(\left(\operatorname{Id} + \gamma A\right)^{-1} - \operatorname{Id}\right) = \left(\operatorname{Id} + (\gamma - \rho)\left(A^{-1} + \rho\operatorname{Id}\right)^{-1}\right)^{-1}. \tag{2.3}$$

Proof. If $\gamma = \rho$, the identity is clear. Otherwise, take $(x, u) \in \mathcal{H}^2$. Then

$$\begin{aligned} & u \in \left(\operatorname{Id} + (\gamma - \rho)\left(A^{-1} + \rho\operatorname{Id}\right)^{-1}\right)^{-1} x \\ &\Leftrightarrow \frac{x - u}{\gamma - \rho} \in \left(A^{-1} + \rho\operatorname{Id}\right)^{-1} u \\ &\Leftrightarrow u \in A^{-1}\left(\frac{x - u}{\gamma - \rho}\right) + \frac{\rho}{\gamma - \rho}(x - u) \\ &\Leftrightarrow \frac{x - u}{\gamma - \rho} \in A\left(\frac{\gamma u - \rho x}{\gamma - \rho}\right) \\ &\Leftrightarrow x \in \left(\operatorname{Id} + \gamma A\right)\left(\frac{\gamma u - \rho x}{\gamma - \rho}\right) \\ &\Leftrightarrow u \in \left(\operatorname{Id} + \left(1 - \frac{\rho}{\gamma}\right)\left(\left(\operatorname{Id} + \gamma A\right)^{-1} - \operatorname{Id}\right)\right) x. \end{aligned} \tag{2.4}$$

∎

The above lemma states that relaxing a proximal step for the original operator A amounts to computing a proximal step for the operator $(A^{-1} + \rho\,\mathrm{Id})^{-1}$. Clearly, when A is cohypomonotone, the latter behaves locally like a monotone operator. This observation will play a central role in our convergence analysis.

3. Proximal iterations with cohypomonotone operators. In this section, we establish our main convergence result for the inexact relaxed proximal point Algorithm 1.1 with cohypomonotone operators.

THEOREM 3.1. *Suppose that in Algorithm* 1.1 *the following conditions are satisfied:*

(i) *There exist a number* $\delta \in \,]0, +\infty[$*, a sequence* $(\rho_i)_{i\in I}$ *in* $[0, +\infty[$*, and open sets* V *and* $(X_i)_{i\in I}$ *in* $\mathcal{H}$ *for which the following hold:*
   (a) $0 \in V$*;*
   (b) *for every* $i \in I$*,* $S_i = X_i \cap A_i^{-1}(0)$ *is closed and* $S_i + B(0;\delta) \subset X_i$*;*
   (c) *for every* $i \in I$*,* A_i *is maximal* ρ_i*-cohypomonotone on* $V \times X_i$*;*
   (d) $S = \bigcap_{i\in I} S_i \neq \varnothing$.

(ii) $(\forall i \in I)(\exists\, M_i \in \mathbb{N} \smallsetminus \{0\})(\forall n \in \mathbb{N})\;\; i \in \{\mathrm{i}(n), \ldots, \mathrm{i}(n + M_i - 1)\}$.

(iii) $\inf_{n\in\mathbb{N}}(\gamma_n - \rho_{\mathrm{i}(n)}) > 0$.

(iv) $(\exists\, \varepsilon \in \,]0, 1[)(\forall n \in \mathbb{N})\;\; \varepsilon \le \frac{\lambda_n}{1-\rho_{\mathrm{i}(n)}/\gamma_n} \le 2 - \varepsilon$.

Then there exists a closed ball B *of radius* $\eta \in \,]0, +\infty]$ *centered at a point in* S *such that if the following conditions hold:*

(v) x_0 *is sufficiently close to* S*, say,*

$$d_S(x_0) < \nu = \frac{4 - 2\varepsilon}{5 - 2\varepsilon}\eta; \tag{3.1}$$

(vi) $\sum_{n\in\mathbb{N}}(\|u_n\| + 2\|v_n\|) < \frac{\nu - d_S(x_0)}{2-\varepsilon}$*;*

(vii) *for every* $n \in \mathbb{N}$*, one selects* $x_{n+\frac{1}{2}} \in B$ *in* (1.5)*,*

then there is one and only one orbit $(x_n)_{n\in\mathbb{N}}$ *of Algorithm* 1.1 *contained in* B *and, furthermore,* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $B \cap S$.

Proof. Take $i \in I$. By (i)(c), there exists an operator $\bar{A}_i \colon \mathcal{H} \rightrightarrows \mathcal{H}$ such that

$$(X_i \times V) \cap \operatorname{gph} \bar{A}_i = (X_i \times V) \cap \operatorname{gph} A_i \tag{3.2}$$

and $\bar{A}_i^{-1} + \rho_i\,\mathrm{Id}$ is maximal monotone. Consequently, it follows from (i)(a), (i)(b), and (i)(d) that

$$X_i \cap \bar{A}_i^{-1}(0) = X_i \cap A_i^{-1}(0) \neq \varnothing \tag{3.3}$$

is closed. Therefore, by maximal monotonicity, $\bar{A}_i^{-1}(0) = (\bar{A}_i^{-1} + \rho_i\,\mathrm{Id})(0)$ is closed and convex [1, Proposition 3.5.6]. Thus, the convex set $\bar{A}_i^{-1}(0)$ is the union of the two disjoint closed sets $X_i \cap \bar{A}_i^{-1}(0) \neq \varnothing$ and $(\complement X_i) \cap \bar{A}_i^{-1}(0)$, which forces the latter to be empty. Indeed, otherwise, by convexity of $\bar{A}_i^{-1}(0)$, we could find $a \in X_i \cap \bar{A}_i^{-1}(0)$ and $b \in (\complement X_i) \cap \bar{A}_i^{-1}(0)$ such that the closed segment $[a, b]$ is the union of the two disjoint closed sets $[a, b] \cap X_i \cap \bar{A}_i^{-1}(0)$ and $[a, b] \cap (\complement X_i) \cap \bar{A}_i^{-1}(0)$, which is impossible since $[a, b]$ is connected by [3, Théorème IV.2.5.4]. To sum up,

$$S_i = X_i \cap A_i^{-1}(0) = X_i \cap \bar{A}_i^{-1}(0) = \bar{A}_i^{-1}(0) \tag{3.4}$$

is closed and convex. It therefore follows from (i)(d) that the projection $P_S x_0$ of x_0 onto S is well defined. On the other hand, (i)(a) and (i)(b) yield $0 \in \operatorname{int} V$ and

$P_S x_0 \in \operatorname{int} \bigcap_{i \in I} X_i$, respectively. As a result, we can find $\eta \in \,]0, +\infty]$ such that

$$B(P_S x_0; \eta) \subset \bigcap_{i \in I} X_i \quad \text{and} \quad B\left(0; 2\eta / \inf_{n \in \mathbb{N}} \gamma_n\right) \subset V. \tag{3.5}$$

We now set

$$B = B(P_S x_0; \eta) \quad \text{and} \quad D = B(P_S x_0; \nu) \tag{3.6}$$

and observe that (3.1) forces

$$x_0 \in \operatorname{int} D. \tag{3.7}$$

Next, take $(x, u) \in B^2$ and $n \in \mathbb{N}$. Then it follows from (3.5) that $\big(u, (x-u)/\gamma_n\big) \in X_{\mathrm{i}(n)} \times V$. Consequently, by (3.2),

$$\begin{aligned} u \in \big(\operatorname{Id} + \gamma_n A_{\mathrm{i}(n)}\big)^{-1} x &\Leftrightarrow \left(u, \frac{x-u}{\gamma_n}\right) \in \operatorname{gph} A_{\mathrm{i}(n)} \\ &\Leftrightarrow \left(u, \frac{x-u}{\gamma_n}\right) \in \operatorname{gph} \bar{A}_{\mathrm{i}(n)} \\ &\Leftrightarrow u \in \big(\operatorname{Id} + \gamma_n \bar{A}_{\mathrm{i}(n)}\big)^{-1} x. \end{aligned} \tag{3.8}$$

Thus,

$$\big(\operatorname{Id} + \gamma_n A_{\mathrm{i}(n)}\big)^{-1}\big|^{B \times B} = \big(\operatorname{Id} + \gamma_n \bar{A}_{\mathrm{i}(n)}\big)^{-1}\big|^{B \times B}. \tag{3.9}$$

On the other hand, let

$$T_n = \operatorname{Id} + \left(1 - \frac{\rho_{\mathrm{i}(n)}}{\gamma_n}\right)\Big(\big(\operatorname{Id} + \gamma_n \bar{A}_{\mathrm{i}(n)}\big)^{-1} - \operatorname{Id}\Big). \tag{3.10}$$

Alternatively, using Lemma 2.4, we can write

$$T_n = \Big(\operatorname{Id} + \tau_n \big(\bar{A}_{\mathrm{i}(n)}^{-1} + \rho_{\mathrm{i}(n)} \operatorname{Id}\big)^{-1}\Big)^{-1}, \quad \text{where} \quad \tau_n = \gamma_n - \rho_{\mathrm{i}(n)}. \tag{3.11}$$

Since $\tau_n > 0$ by (iii), T_n is therefore the resolvent of the operator $\tau_n C_{\mathrm{i}(n)}$, where

$$C_{\mathrm{i}(n)} = \big(\bar{A}_{\mathrm{i}(n)}^{-1} + \rho_{\mathrm{i}(n)} \operatorname{Id}\big)^{-1}, \tag{3.12}$$

which is maximal monotone as the inverse of such an operator. Hence, it follows from [2, Proposition 2.3] that

$$T_n \colon \operatorname{dom} T_n = \mathcal{H} \to \mathcal{H} \quad \text{and} \quad (\forall (x, z) \in \mathcal{H} \times \operatorname{Fix} T_n) \quad \langle z - T_n x \,, x - T_n x \rangle \le 0, \tag{3.13}$$

which, by [7, Proposition 2.3(ii)], implies

$$(\forall \mu \in [0, 2])(\forall (x, z) \in \mathcal{H} \times \operatorname{Fix} T_n) \quad \|x + \mu(T_n x - x) - z\|^2 \le \|x - z\|^2 - \mu(2 - \mu)\|T_n x - x\|^2. \tag{3.14}$$

Now, let

$$\mu_n = \frac{\lambda_n}{1 - \rho_{\mathrm{i}(n)}/\gamma_n}. \tag{3.15}$$

Then we get from (iv) that

$$\mu_n \in [\varepsilon, 2-\varepsilon]. \tag{3.16}$$

We also obtain from (3.5), (3.3), (3.12), and (3.11) that

$$B \cap S_{\mathrm{i}(n)} = B \cap A^{-1}_{\mathrm{i}(n)}(0) = B \cap \bar{A}^{-1}_{\mathrm{i}(n)}(0) = B \cap C^{-1}_{\mathrm{i}(n)}(0) = B \cap \operatorname{Fix} T_n. \tag{3.17}$$

Hence, $P_S x_0 \in B \cap S \subset B \cap \operatorname{Fix} T_n$, and it results from (3.14) with $\mu = 1$ and $z = P_S x_0$ that

$$T_n(B) \subset B. \tag{3.18}$$

Let us now show that Algorithm 1.1 is viable, i.e., that the recursion (1.4)–(1.5) does generate an infinite sequence. To this end, we shall show that the sequence $(x_n)_{n\in\mathbb{N}}$ is well defined and that it lies in $\operatorname{int} D$, whereas the sequence $(x_n + v_n)_{n\in\mathbb{N}}$ lies in $\operatorname{int} B$. Since (vi) yields $\|v_0\| < \eta - d_S(x_0)$, it follows from (3.7) that $\|x_0 + v_0 - P_S x_0\| \leq \|v_0\| + d_S(x_0) < \eta$, whence $x_0 + v_0 \in \operatorname{int} B$. Now assume that, for some $n \in \mathbb{N}$, the points $(x_k)_{0\leq k\leq n}$ and $(x_k + v_k)_{0\leq k\leq n}$ lie in $\operatorname{int} D$ and $\operatorname{int} B$, respectively. Then it results from (vii) and (3.9) that (1.5) can be written as

$$x_{n+\frac{1}{2}} \in (\mathrm{Id} + \gamma_n A_{\mathrm{i}(n)})^{-1}\big|^{B\times B}(x_n + v_n) = (\mathrm{Id} + \gamma_n \bar{A}_{\mathrm{i}(n)})^{-1}\big|^{B\times B}(x_n + v_n). \tag{3.19}$$

In view of (3.15), (1.4) can now be written as

$$x_{n+1} \in x_n + \mu_n \left(1 - \frac{\rho_{\mathrm{i}(n)}}{\gamma_n}\right)\left((\mathrm{Id} + \gamma_n \bar{A}_{\mathrm{i}(n)})^{-1}\big|^{B\times B}(x_n + v_n) + u_n - x_n\right), \tag{3.20}$$

which, by virtue of (3.10), yields

$$x_{n+1} \in x_n + \mu_n \left(T_n\big|^{B\times B}(x_n + v_n) + w_n - x_n\right), \tag{3.21}$$

where

$$w_n = \left(1 - \frac{\rho_{\mathrm{i}(n)}}{\gamma_n}\right) u_n - \frac{\rho_{\mathrm{i}(n)}}{\gamma_n} v_n. \tag{3.22}$$

However, since $x_n + v_n \in B$, (3.18) yields $T_n\big|^{B\times B}(x_n + v_n) = T_n(x_n + v_n)$. Hence, since T_n is single-valued and defined everywhere on $\mathcal{H}$ (see (3.13)), we deduce from (3.21) that x_{n+1} is uniquely defined by

$$x_{n+1} = x_n + \mu_n\big(T_n(x_n + v_n) + w_n - x_n\big). \tag{3.23}$$

Now put

$$e_n = w_n + T_n(x_n + v_n) - T_n x_n. \tag{3.24}$$

Then we derive from (3.23) that

$$x_{n+1} = x_n + \mu_n(T_n x_n - x_n) + \mu_n e_n, \tag{3.25}$$

and it follows from (3.16) and (3.14) that

$$\begin{aligned} \|x_{n+1} - P_S x_0\| &\leq \|x_n - P_S x_0 + \mu_n(T_n x_n - x_n)\| + \mu_n\|e_n\| \\ &\leq \|x_n - P_S x_0\| + \mu_n\|e_n\|. \end{aligned} \tag{3.26}$$

Consequently, since $(x_k + v_k)_{0\leq k\leq n}$ lies in B, we have

$$\begin{aligned} \|x_{n+1} - P_S x_0\| &\leq \|x_0 - P_S x_0\| + \sum_{k=0}^{n} \mu_k\|e_k\| \\ &\leq d_S(x_0) + (2-\varepsilon)\sum_{k\in\mathbb{N}} \|e_k\|. \end{aligned} \tag{3.27}$$

On the other hand, it follows from (3.24), the nonexpansivity of T_n [1, Proposition 3.5.3], (3.22), and (iii) that

$$\|e_n\| \leq \|w_n\| + \|T_n(x_n + v_n) - T_n x_n\| \leq \|w_n\| + \|v_n\| \leq \|u_n\| + 2\|v_n\|. \tag{3.28}$$

Therefore, we derive from (vi) that

$$d_S(x_0) + (2-\varepsilon)\sum_{k\in\mathbb{N}} \|e_k\| < \nu, \tag{3.29}$$

and deduce from (3.27) that $\|x_{n+1} - P_S x_0\| < \nu$, i.e., $x_{n+1} \in \operatorname{int} D$. In turn, (vi) and (3.1) yield

$$\|x_{n+1} + v_{n+1} - P_S x_0\| < \nu + \|v_{n+1}\| < \nu + \frac{\nu}{2(2-\varepsilon)} = \eta, \tag{3.30}$$

i.e., $x_{n+1} + v_{n+1} \in \operatorname{int} B$. We have thus shown by induction that the entire sequence $(x_n + v_n)_{n\in\mathbb{N}}$ lies in B and that $(x_n)_{n\in\mathbb{N}}$ is a well-defined sequence which lies entirely in $\operatorname{int} D \subset \operatorname{int} B$. In view of (3.23), (3.11), and (3.12), the recursion governing the sequence $(x_n)_{n\in\mathbb{N}}$ can now be rewritten as

$$(\forall n \in \mathbb{N}) \;\; x_{n+1} = x_n + \mu_n(x_{n+\frac{1}{2}} + w_n - x_n), \tag{3.31}$$

where $x_{n+\frac{1}{2}}$ is the unique solution to the inclusion

$$v_n \in x_{n+\frac{1}{2}} - x_n + \tau_n C_{\mathrm{i}(n)} x_{n+\frac{1}{2}}, \tag{3.32}$$

namely, $x_{n+\frac{1}{2}} = \big(\mathrm{Id} + \tau_n C_{\mathrm{i}(n)}\big)^{-1}(x_n + v_n) = T_n(x_n + v_n) \in B$, where the last inclusion follows from $x_n + v_n \in B$ and (3.18). In summary, since the operators $(C_i)_{i\in I}$ are maximal monotone, $\inf_{n\in\mathbb{N}} \tau_n > 0$, $(\mu_n)_{n\in\mathbb{N}}$ lies in $[\varepsilon, 2-\varepsilon]$, $\sum_{n\in\mathbb{N}} \|w_n\| \leq \sum_{n\in\mathbb{N}} \|u_n\| + \sum_{n\in\mathbb{N}} \|v_n\| < +\infty$, and $\sum_{n\in\mathbb{N}} \|v_n\| < +\infty$, it follows from Theorem 1.2 that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point x in $\bigcap_{i\in I} C_i^{-1}(0)$. On the other hand, since $(x_n)_{n\in\mathbb{N}}$ lies in the weakly closed set B, $x \in B$. As a result, (3.17) yields $x \in B \cap \bigcap_{i\in I} C_i^{-1}(0) = B \cap S$. □

Remark 3.2. The above result unifies and extends several results found in the literature.

- If $I = \{1\}$ (a single operator is considered), $\lambda_n \equiv 1$, and $u_n \equiv 0 \equiv v_n$, then Theorem 3.1 is found in [20, Theorem 9].
- If $I = \{1\}$, $\rho_1 = 0$, and $V = \mathcal{H} = X_1$, Theorem 3.1 reduces to [11, Theorem 3], and to [22, Theorem 1] if we further assume $\lambda_n \equiv 1$.
- If $\rho_i \equiv 0$ and $V = \mathcal{H} \equiv X_i$, Theorem 3.1 reduces to Theorem 1.2.
- If $\rho_i \equiv 0$, $V = \mathcal{H} \equiv X_i$, and $u_n \equiv 0 \equiv v_n$, Theorem 3.1 corresponds to [2, Corollary 6.1(i)].
- If $I = \{1, \dots, m\}$, $(S_i)_{i \in I}$ is a family of closed convex sets in $\mathcal{H}$ with associated projection operators $(P_i)_{i \in I}$, $\mathrm{i}\colon n \mapsto n$ modulo $m+1$, for every $i \in I$, $A_i = N_{S_i}$ (hence $\rho_i \equiv 0$ and $V = \mathcal{H} \equiv X_i$), and $u_n \equiv 0 \equiv v_n$, then Algorithm 1.1 produces the method of cyclic projections

$$(3.33)\quad (\forall n \in \mathbb{N})\;\; x_{n+1} = x_n + \lambda_n (P_{n \text{ modulo } m+1} x_n - x_n), \quad \text{where} \quad \varepsilon \le \lambda_n \le 2 - \varepsilon,$$

  and Theorem 3.1 reduces to [13, Theorem 1].

4. Nonlinear programming application. Using the same arguments as in [20, section 5], one can derive multiplier methods for quite general variational inclusions by combining Theorem 3.1 with an abstract duality framework for set-valued mappings. Instead of going through all the steps and applications discussed in [20], we analyze the proximal method of multipliers for nonlinear (nonconvex) programming as an example. The proximal method of multipliers was introduced and analyzed in the convex case by Rockafellar [23] and in the nonconvex case in [20, section 7] with exact, unrelaxed iterates.

Consider the nonlinear programming problem

$$(4.1)\qquad \text{minimize} \quad f_0(x) \quad \text{subject to} \quad \begin{cases} f_i(x) = 0 & \text{for } 1 \le i \le r, \\ f_i(x) \le 0 & \text{for } r+1 \le i \le m, \end{cases}$$

where $(f_i)_{0 \le i \le m}$ are real-valued C^2-functions defined on the standard Euclidean space $\mathbb{R}^N$. Our aim is to find Karush–Kuhn–Tucker (KKT) points for (4.1). To this end, we introduce the closed convex cone $K = \{0\}^r \times \mathbb{R}_-^{m-r}$, let $F\colon x \mapsto \big(f_1(x), \dots, f_m(x)\big)$, and set $\mathcal{H} = \mathbb{R}^N \times \mathbb{R}^m$. We shall derive from Theorem 3.1 a local convergence result for the following proximal method of multipliers.

ALGORITHM 4.1. *Let $(x_0, y_0) \in \mathcal{H}$, let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence in $]0, +\infty[$, and let $(w_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{R}^N$. A sequence $\big((x_n, y_n)\big)_{n \in \mathbb{N}}$ is constructed according to the updating rule*

$$(4.2)\qquad (\forall n \in \mathbb{N})\;\; (x_{n+1}, y_{n+1}) = (x_n, y_n) + \lambda_n \big((x_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}) - (x_n, y_n)\big),$$

where $x_{n+\frac{1}{2}}$ minimizes approximately the function

$$(4.3)\qquad \varphi_n\colon x \mapsto f_0(x) + \frac{1}{2\gamma_n}\|x - x_n\|^2 + \frac{1}{2\gamma_n} d_K\big(y_n + \gamma_n F(x)\big)^2$$

in the sense that $\nabla \varphi_n(x_{n+\frac{1}{2}}) = w_n$, and

$$(4.4)\qquad \begin{cases} y^i_{n+\frac{1}{2}} = y^i_n + \gamma_n f_i(x_{n+\frac{1}{2}}) & \text{for } 1 \le i \le r, \\ y^i_{n+\frac{1}{2}} = \max\big\{y^i_n + \gamma_n f_i(x_{n+\frac{1}{2}}), 0\big\} & \text{for } r+1 \le i \le m. \end{cases}$$

Remark 4.2. It was shown in the proof of [20, Theorem 19] that, under condition (i) of Theorem 4.3 and for x near $\bar{x}$, the condition $\nabla\varphi_n(x) = 0$ implies that x is a local minimizer of φ_n. For this reason, the condition $\nabla\varphi_n(x_{n+\frac{1}{2}}) = w_n$ is interpreted in Algorithm 4.1 as an approximate minimization. In practice, the parameter γ_n is often chosen adaptively while the size of the vector $\|w_n\|$ can be made arbitrarily small by choosing the stopping criterion appropriately in the minimization of φ_n in (4.3).

Define a mapping $L\colon \mathcal{H} \rightrightarrows \mathcal{H}$ by

$$L\colon (x,y) \mapsto \big(\nabla f_0(x) + \langle y\,, \nabla F(x)\rangle\,, -F(x) + N_{K^*}(y)\big), \tag{4.5}$$

where $K^* = \mathbb{R}^r \times \mathbb{R}_+^{m-r}$ is the polar cone of K. Then the KKT system for (4.1) can be written as [25, Example 11.46]

$$(0,0) \in L(x,y). \tag{4.6}$$

Let $(\bar{x},\bar{y}) \in \mathcal{H}$ be a point satisfying the KKT conditions for (4.1) and define

$$I^+ = \{1,\ldots,r\} \cup \{r+1 \leq i \leq m \mid f_i(\bar{x}) = 0 \;\text{ and }\; \bar{y}_i > 0\}. \tag{4.7}$$

Now let $l\colon (x,y) \mapsto f_0(x) + \langle y\,, F(x)\rangle$. Recall that $(\bar{x},\bar{y})$ is said to satisfy the *strong second order sufficient condition* for (4.1) if [21]

$$(\forall y \in \mathbb{R}^m)\;\; \begin{cases} y \neq 0 \\ (\forall i \in I^+) \;\; \langle y\,, \nabla f_i(\bar{x})\rangle = 0 \end{cases} \quad\Rightarrow\quad \big\langle y\,, \nabla^2_{xx} l(\bar{x},\bar{y})y\big\rangle > 0. \tag{4.8}$$

THEOREM 4.3. *Suppose that in Algorithm* 4.1 *the following conditions are satisfied:*

(i) $(\bar{x},\bar{y}) \in \mathcal{H}$ *is a KKT point for* (4.1) *satisfying* (4.8) *and such that the gradients* $\big(\nabla f_i(\bar{x})\big)_{i\in I^+}$ *are linearly independent;*

(ii) $\inf_{n\in\mathbb{N}} \gamma_n$ *is large enough;*

(iii) $(\exists\, \varepsilon \in \,]0,1[)(\forall n \in \mathbb{N})\;\; \varepsilon \leq \lambda_n \leq 2-\varepsilon$.

Then there exists a closed ball B centered at $(\bar{x},\bar{y})$ such that if the following conditions hold:

(iv) (x_0,y_0) *is sufficiently close to* $(\bar{x},\bar{y})$*;*

(v) $\sum_{n\in\mathbb{N}} \gamma_n \|w_n\|$ *is small enough;*

(vi) *for every* $n \in \mathbb{N}$, $(x_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}) \in B$,

then there is one and only one orbit $\big((x_n,y_n)\big)_{n\in\mathbb{N}}$ *of Algorithm* 4.1 *contained in B and, furthermore,* $\big((x_n,y_n)\big)_{n\in\mathbb{N}}$ *converges to* $(\bar{x},\bar{y})$.

Proof. By [18, Exemple 4.b and Proposition 7.d],

$$\nabla d_K^2 = 2(\operatorname{Id} - P_K) = 2P_{K^*}. \tag{4.9}$$

Thus, it follows from (4.3) that

$$\nabla\varphi_n(x) = \nabla f_0(x) + \gamma_n^{-1}(x - x_n) + \big\langle P_{K^*}\big(y_n + \gamma_n F(x)\big)\,, \nabla F(x)\big\rangle\,. \tag{4.10}$$

Note also that $y_{n+\frac{1}{2}}$ in (4.4) can be expressed as

$$y_{n+\frac{1}{2}} = P_{K^*}\big(y_n + \gamma_n F(x_{n+\frac{1}{2}})\big) = (I + N_{K^*})^{-1}\big(y_n + \gamma_n F(x_{n+\frac{1}{2}})\big). \tag{4.11}$$

The update rules for $(x_{n+\frac{1}{2}}, y_{n+\frac{1}{2}})$ are thus equivalent to the system

$$\begin{cases} w_n = \nabla f_0(x_{n+\frac{1}{2}}) + \gamma_n^{-1}(x_{n+\frac{1}{2}} - x_n) + \left\langle y_{n+\frac{1}{2}}\,, \nabla F(x_{n+\frac{1}{2}}) \right\rangle, \\ y_n + \gamma_n F(x_{n+\frac{1}{2}}) \in y_{n+\frac{1}{2}} + N_{K^*}(y_{n+\frac{1}{2}}). \end{cases} \tag{4.12}$$

Alternatively, $(x_{n+\frac{1}{2}}, y_{n+\frac{1}{2}})$ is a solution to

$$(\gamma_n w_n, 0) \in (x_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}) - (x_n, y_n) + \gamma_n L(x_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}). \tag{4.13}$$

Now, set $A_1 = L$, $I = \{1\}$, and $\mathrm{i}(n) = 1$ for all $n \in \mathbb{N}$. Then, in view of (4.13), the iterations described by (4.2)–(4.4) are seen to conform to the format (1.4)–(1.5), and Algorithm 4.1 therefore fits the general framework of Algorithm 1.1. Accordingly, it suffices to verify the conditions of Theorem 3.1 to establish the claims. By [21, Theorem 4.1], condition (i) implies that L^{-1} has a Lipschitz localization at $\big((0,0),(\bar{x},\bar{y})\big)$. Therefore, by Lemma 2.3, condition (i) of Theorem 3.1 holds with $S = S_1 = \{(\bar{x},\bar{y})\}$ for some $\rho \in [0,+\infty[$. Now let $\gamma = \inf_{n\in\mathbb{N}} \gamma_n$. Then condition (iii) of Theorem 3.1 reads $\gamma > \rho$, and is trivially implied by condition (ii) above. Next, let us show that condition (iii) above implies condition (iv) of Theorem 3.1, i.e.,

$$(\exists\, \zeta \in\,]0,1[)(\forall n \in \mathbb{N})\ \ \zeta \leq \frac{\lambda_n}{1-\rho/\gamma_n} \leq 2 - \zeta. \tag{4.14}$$

It is readily checked that for $\gamma > 2\rho/\varepsilon$ (as is allowed by (ii) above), we have

$$\zeta = \frac{\varepsilon\gamma - 2\rho}{\gamma - \rho} \in\,]0,\varepsilon[\,. \tag{4.15}$$

Hence, it follows from (iii) above that

$$(\forall n \in \mathbb{N})\ \ \zeta < \varepsilon \leq \lambda_n < \frac{\lambda_n}{1-\rho/\gamma_n} \leq \frac{2-\varepsilon}{1-\rho/\gamma} = 2 - \zeta, \tag{4.16}$$

which establishes (4.14). Finally, it is clear that conditions (v)–(vii) of Theorem 3.1 are implied by conditions (iv)–(vi) above. □

Remark 4.4.

(i) In most concrete problems, it is not possible to obtain the value of ρ in the above proof [21]. As a result, condition (ii) and (v) in Theorem 4.3 are stated in qualitative terms rather than with hard bounds involving ρ.

(ii) The above result extends [20, Theorem 19] by allowing for relaxations and inexact computation of the iterates, thus making the algorithm more practical and flexible.

REFERENCES

[1] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis,* Birkhäuser, Boston, MA, 1990.

[2] H. H. Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces*, Math. Oper. Res., 26 (2001), pp. 248–264.

[3] N. Bourbaki, *Topologie Générale* 1–4, Masson, Paris, 1990.

[4] H. Brézis and P. L. Lions, *Produits infinis de résolvantes*, Israel J. Math., 29 (1978), pp. 329–345.

[5] F. E. Browder, *Convergence theorems for sequences of nonlinear operators in Banach spaces,* Math. Z., 100 (1967), pp. 201–225.

[6] P. L. Combettes, *Hilbertian convex feasibility problem: Convergence of projection methods*, Appl. Math. Optim., 35 (1997), pp. 311–330.

[7] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently Parallel Algorithms for Feasibility and Optimization, D. Butnariu, Y. Censor, and S. Reich, eds., Elsevier, New York, 2001, pp. 115–152.

[8] A. Daniilidis and P. Georgiev, *Cyclic hypomonotonicity, cyclic submonotonicity, and integration*, J. Optim. Theory Appl., 122 (2004), pp. 19–40.

[9] A. L. Dontchev and R. T. Rockafellar, *Characterizations of strong regularity for variational inequalities over polyhedral convex sets,* SIAM J. Optim., 6 (1996), pp. 1087–1105.

[10] A. L. Dontchev and R. T. Rockafellar, *Characterizations of Lipschitzian stability in nonlinear programming*, in Mathematical Programming with Data Perturbations, Lecture Notes in Pure and Appl. Math. 195, Dekker, New York, 1998, pp. 65–82.

[11] 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.

[12] J. Eckstein and M. C. Ferris, *Smooth methods of multipliers for complementarity problems,* Math. Program., 86 (1999), pp. 65–90.

[13] L. G. Gubin, B. T. Polyak, and E. V. Raik, *The method of projections for finding the common point of convex sets*, U.S.S.R. Comput. Math. and Math. Phys., 7 (1967), pp. 1–24.

[14] A. N. Iusem, T. Pennanen, and B. F. Svaiter, *Inexact variants of the proximal point algorithm without monotonicity,* SIAM J. Optim., 13 (2003), pp. 1080–1097.

[15] A. B. Levy, *Lipschitzian multifunctions and a Lipschitzian inverse mapping theorem,* Math. Oper. Res., 26 (2001), pp. 105–118.

[16] A. B. Levy and R. A. Poliquin, *Characterizing the single-valuedness of multifunctions*, Set-Valued Anal., 5 (1997), pp. 351–364.

[17] B. Martinet, *Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox,* C. R. Acad. Sci. Paris Sér. A Math., 274 (1972), pp. 163–165.

[18] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[19] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., 73 (1967), pp. 591–597.

[20] T. Pennanen, *Local convergence of the proximal point algorithm and multiplier methods without monotonicity,* Math. Oper. Res., 27 (2002), pp. 170–191.

[21] S. M. Robinson, *Strongly regular generalized equations,* Math. Oper. Res., 5 (1980), pp. 43–62.

[22] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.

[23] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.

[24] R. T. Rockafellar, *Favorable classes of Lipschitz-continuous functions in subgradient optimization,* in Progress in Nondifferentiable Optimization, E. A. Nurminski, ed., IIASA Collaborative Proc. Ser. CP-82, International Institute for Applied Systems Analysis, Laxenburg, 1982, pp. 125–143.

[25] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis,* Springer-Verlag, New York, 1998.

[26] J. E. Spingarn, *Submonotone mappings and the proximal point algorithm,* Numer. Funct. Anal. Optim., 4 (1981/82), pp. 123–150.

[27] J. P. Vial, *Strong and weak convexity of sets and functions,* Math. Oper. Res., 8 (1983), pp. 231–259.

[28] E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory,* in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York, 1971, pp. 237–424.

[29] E. Zeidler, *Nonlinear Functional Analysis and Its Applications II/B—Nonlinear Monotone Operators,* Springer-Verlag, New York, 1990.