MATHEMATICS OF OPERATIONS RESEARCH
Vol. 27, No. 1, February 2002, pp. 170–191
*Printed in U.S.A.*

# LOCAL CONVERGENCE OF THE PROXIMAL POINT ALGORITHM AND MULTIPLIER METHODS WITHOUT MONOTONICITY

TEEMU PENNANEN

This paper studies the convergence of the classical proximal point algorithm without assuming monotonicity of the underlying mapping. Practical conditions are given that guarantee the local convergence of the iterates to a solution of $T(x) \ni 0$, where $T$ is an arbitrary set-valued mapping from a Hilbert space to itself. In particular, when the problem is "nonsingular" in the sense that $T^{-1}$ has a Lipschitz localization around one of the solutions, local linear convergence is obtained. This kind of regularity property of variational inclusions has been extensively studied in the literature under the name of *strong regularity*, and it can be viewed as a natural generalization of classical constraint qualifications in nonlinear programming to more general problem classes. Combining the new convergence results with an abstract duality framework for variational inclusions, the author proves the local convergence of multiplier methods for a very general class of problems. This gives as special cases new convergence results for multiplier methods for nonmonotone variational inequalities and nonconvex nonlinear programming.

**1. Introduction.** Consider the problem of finding a solution to an inclusion of the form,

$$(P) \qquad T(x) \ni 0,$$

where $T : H \rightrightarrows H$ is a set-valued mapping on a Hilbert space $H$. It is well known that most problems of variational character, such as minimization or maximization (constrained or not) of a function, variational inequalities, minimax problems, etc., can be written in this form. In many (but not all) situations, a given problem can be written in the form $(P)$ with the property that $T$ is *monotone*, i.e.,

$$\langle x_1 - x_2, v_1 - v_2 \rangle \geq 0 \quad \forall (x_i, v_i) \in \operatorname{gph} T,$$

where $\operatorname{gph} T = \{(x, v) \mid v \in T(x)\}$ is the *graph* of $T$. This is the case, for example, in convex programming, monotone variational inequalities, and convex-concave minimax problems. See for example Brezis (1973), Zeidler (1990), or Rockafellar and Wets (1998, Chapter 12) for a reference on monotone mappings.

Rockafellar (1976a) gave a general convergence and rate of convergence analysis for an algorithm for solving the monotone case of $(P)$. This algorithm, named the *proximal point algorithm*, was a generalization of the algorithm of Martinet (1970) for convex minimization. It was proved that when the mapping $T$ is *maximal monotone*, i.e., monotone and not extendable to another monotone mapping, the sequence $\{x_k\}$, generated from any starting point $x_0$ by the rule

$$(1) \qquad x_{k+1} \approx P_k(x_k),$$

Received October 25, 2000; revised May 7, 2001, and October 8, 2001.
*MSC 2000 subject classification*. Primary: 90C30, 90C33, 90C48.
*OR/MS subject classification*. Primary: Programming/nonlinear.
*Key words*. Proximal point algorithm, local convergence, strong regularity.



0364-765X/02/2701/0170/$05.00
1526-5471 electronic ISSN, © 2002, INFORMS

converges weakly to a point in the solution set $T^{-1}(0)$ of $(P)$, provided the approximation is made sufficiently accurate as the iteration proceeds. Here $P_k$ is defined by $P_k(x) = (I + c_k T)^{-1}$ for a sequence $\{c_k\}$ of positive scalars that are bounded away from zero. Iteration (1) means that $x_{k+1}$ is an approximate (see Rockafellar 1976a, Proposition 3) solution to the inclusion

$$T(x) + c_k^{-1}(x - x_k) \ni 0. \tag{2}$$

At first sight, this looks almost as hard as the original problem. The key idea in Rockafellar (1976b), a follow-up paper to Rockafellar (1976a), was to use convex programming duality to reformulate (2) into a problem that is much easier than $(P)$. More precisely, it was shown in Rockafellar (1976b) how the proximal point algorithm can be combined with convex programming duality theory to give a general convergence analysis for the multiplier method (augmented Lagrangian method, Rockafellar 1973) in convex programming. Furthermore, a new algorithm that was named the *proximal method of multipliers* was obtained that had better theoretical convergence properties than the ordinary multiplier method.

Since the publication of Rockafellar (1976a, b), many generalizations have been made to the classical proximal point algorithm; see, for example, Eckstein and Bertsekas (1992), Eckstein (1993), Tossings (1994), Iusem et al. (1994), Solodov and Svaiter (1999), Auslender and Teboulle (2000), and the references therein. Combining such a generalized proximal point algorithm with the general strategy in Rockafellar (1976b), one obtains generalized multiplier methods that may have advantages over the classical methods. Also, using generalized duality frameworks beyond convex programming, the approach of Rockafellar (1976b) has been used to derive multiplier methods for variational inequalities in Rockafellar (1978), Gabay (1983), Eckstein and Ferris (1999), and Auslender and Teboulle (2000).

All the above-mentioned works rely heavily on monotonicity of the underlying mapping. It is known, however, that the classical multiplier method for minimization often converges even without convexity; see Bertsekas (1982), Ito and Kunisch (1990), Contesse-Becker (1993), Conn et al. (1996), and the discussion below. Moreover, Eckstein and Ferris (1999) successfully applied their multiplier method to nonmonotone variational inequalities, even though its convergence analysis was based on monotonicity. This suggests that monotonicity might not be needed for local convergence of the general proximal point algorithm for set-valued mappings. The first result in this direction was given in Spingarn (1981), where conditions were given that guarantee the local convergence of the proximal point algorithm without requiring monotonicity of $T$. However, these conditions that essentially require that $T^{-1}$ be differentiable at a solution and that the derivative be monotone (Spingarn 1981 Theorem 2) were rather stringent, and they exclude some important applications. In a recent paper, Kaplan and Tichatschke (1998) review the literature on proximal point algorithm for nonconvex minimization, and they give new conditions for convergence.

In this paper, we give new, more practical conditions on the operator $T$ that guarantee the local convergence of the proximal point algorithm, given that the parameters $c_k$ are kept sufficiently high. Except for the first section, which deals with the fully monotone case, we will not allow errors in the computation of the iterates, so that (1) becomes $x_{k+1} = P_k(x_k)$. Although this restriction will compromise the practicality of the algorithm, we have chosen to do so since the inclusion of an error term would considerably complicate the analysis and the presentation of our results. The study of an approximate version is the subject of an ongoing research.

Our analysis is based on the observation that the solution set of $(P)$ coincides with that of the inclusion

$$T_\rho(x) \ni 0,$$

where $T_\rho = (T^{-1} + \rho I)^{-1}$ is the *Yosida regularization* of $T$ with parameter $\rho > 0$. We apply an over-relaxed proximal point algorithm to $T_\rho$, and show that, with the right choice of relaxation parameters, this yields the classical proximal point algorithm for $T$ itself. The main point is that for large enough $\rho > 0$, *the mapping $T_\rho$ can be locally monotone even when $T$ is not, and that local maximal monotonicity is enough to guarantee the local convergence of the proximal point algorithm*.

An important case covered by our analysis is when $T^{-1}$ has a *Lipschitz localization* around the solution. The existence of such localizations has been studied extensively in the literature, and there are available specific conditions guaranteeing it for many classes of problems; see, for example, Robinson (1980), Klatte and Tammer (1990), Dontchev and Rockafellar (1996, 1998, 2001), Klatte and Kummer (1999), Levy et al. (2000), and the references therein. Most notably, Robinson (1980) gave a sufficient condition called *strong regularity*, which guarantees the existence of a Lipschitz localization for the solution mapping of a variational inequality. He also showed that, in the case of nonlinear programming, it is guaranteed by the *strong second-order sufficient condition*. In general, strong regularity can be seen as a natural generalization to the set-valued case of the classical nonsingularity condition for smooth mappings.

Using the new convergence results for the proximal point algorithm, we will prove the local convergence of multiplier methods for a very general class of inclusions. Our analysis is based in part on a generalized duality scheme, much like convex programming duality was used in Rockafellar (1976b) to analyze multiplier methods for minimization. These results are then specialized to variational inequalities, and we obtain what seems to be the first convergence result for a multiplier method for nonmonotone variational inequalities.

As a further application, we derive a convergence result for the proximal method of multipliers for nonlinear (nonconvex) programming. This result also seems to be new. In Bertsekas (1982, §2.2), the convergence of the classical method of multipliers for equality constrained problem was proved. These results were then used in Bertsekas (1982, §3.1), to prove the convergence of the multiplier method for inequality constrained problems under the assumption of *strict complementarity*. Stronger convergence results not requiring strict complementarity were later derived in Ito and Kunisch (1990), Contesse-Becker (1993), and Conn et al. (1996). Our convergence result for the proximal method of multipliers is very similar to the results in Ito and Kunisch (1990), Contesse-Becker (1993), and Conn et al. (1996), but its derivation based on the general proximal point algorithm is simpler. The same strategy would yield a simplified convergence proof also for the classical method of multipliers.

We begin by recalling the relaxed proximal point algorithm of Eckstein and Bertsekas (1992) for monotone mappings in §2, and we show that under a standard assumption, local linear convergence is obtained. In §3, we show that the global monotonicity of the underlying mapping can be replaced by "local monotonicity" without affecting the local behavior of the relaxed proximal point algorithm. The main results are given in §4. In §5, these results are combined with a duality scheme for variational inclusions, and we obtain multiplier methods for a very general class of problems. In §§6 and 7 these are specialized to variational inequalities and nonlinear programming, respectively.

**2. The relaxed proximal point algorithm.** Throughout this paper, $H$ will be a real Hilbert space, unless otherwise specified. In this section, we recall the relaxed proximal point algorithm from Eckstein and Bertsekas (1992), and give a convergence rate result for it. Although these results are based on monotonicity of the underlying mapping, they will eventually lead to our main results for the nonmonotone case.

THEOREM 1 (ECKSTEIN AND BERTSEKAS). *Let $T : H \rightrightarrows H$ be maximal monotone with $0 \in \operatorname{rge} T$ and let $\{\sigma_k\}$, $\{c_k\}$ and $\{\epsilon_k\}$ be scalar sequences such that*

$$\inf \sigma_k > 0, \quad \sup \sigma_k < 2, \quad \inf c_k > 0, \quad \sum \epsilon_k < \infty.$$

*Then for any* $x_0 \in H$, *the sequence* $\{x_k\}$ *generated by the rule*

$$x_{k+1} = \sigma_k \bar{x}_k + (1-\sigma_k)x_k,$$

*where* $\bar{x}_k$ *is such that*

$$\|\bar{x}_k - P_k(x_k)\| \leq \epsilon_k,$$

*converges weakly to a point in* $T^{-1}(0)$.

Under a local Lipschitz-type condition on the mapping $T^{-1}$ and a tightened error tolerance, a convergence rate analysis was provided for the ordinary proximal point algorithm in Rockafellar (1976a). This result has been strengthened by Luque (1984) and Robinson (1999). Under the same conditions as in Rockafellar (1976a), we can derive a convergence rate estimate for the over-relaxed algorithm of Theorem 1.

PROPOSITION 2. *In the situation of Theorem* 1, *assume that* $\sigma_k \geq 1$, $c_k \nearrow \bar{c} \leq \infty$, *and*

$$\epsilon_k \leq \delta_k \|\bar{x}_k - x_k\|, \tag{3}$$

*where* $\delta_k \to 0$. *If* $T(x) \ni 0$ *has a unique solution* $\bar{x}$, *and for some neighborhood* $U \ni 0$ *and a constant* $a \geq 0$

$$y \in U, \quad x \in T^{-1}(y) \implies \|x - \bar{x}\| \leq a\|y\|, \tag{4}$$

*then* $\{x_k\}$ *converges linearly to* $\bar{x}$ *with rate*

$$\sqrt{1 - \hat{\sigma}(2-\hat{\sigma})\frac{\bar{c}^2}{a^2+\bar{c}^2}} < 1,$$

*where* $\hat{\sigma} = \limsup_{k\to\infty} \sigma_k$.

PROOF. We have that $x_k - P_k(x_k) \to 0$ (see Eckstein and Bertsekas 1992, p. 300), so for $k$ large enough, $x_k - P_k(x_k) \in U$. This together with (4) implies (see Rockafellar 1976a, p. 886)

$$\|P_k(x_k) - \bar{x}\|^2 \leq \frac{a^2}{a^2 + c_k^2}\|x_k - \bar{x}\|^2. \tag{5}$$

Defining $\tilde{x}_{k+1} = \sigma_k P_k(x_k) + (1-\sigma_k)x_k$, we have

$$\begin{aligned}\|\tilde{x}_{k+1} - \bar{x}\|^2 &= \|\sigma_k(P_k(x_k) - \bar{x}) + (1-\sigma_k)(x_k - \bar{x})\|^2\\ &= \sigma_k^2\|P_k(x_k) - \bar{x}\|^2 + (1-\sigma_k)^2\|x_k - \bar{x}\|^2 + 2\sigma_k(1-\sigma_k)\langle P_k(x_k) - \bar{x}, x_k - \bar{x}\rangle,\end{aligned}$$

where $\langle P_k(x_k) - \bar{x}, x_k - \bar{x}\rangle = \langle P_k(x_k) - P_k(\bar{x}), x_k - \bar{x}\rangle \geq \|P_k(x_k) - P_k(\bar{x})\|^2 = \|P_k(x_k) - \bar{x}\|^2$ by strong monotonicity of $P_k^{-1}$ (this is the well-known firm nonexpansiveness property of $P_k$). Since $\sigma_k \geq 1$, we then get

$$\|\tilde{x}_{k+1} - \bar{x}\|^2 \leq \sigma_k(2-\sigma_k)\|P_k(x_k) - \bar{x}\|^2 + (1-\sigma_k)^2\|x_k - \bar{x}\|^2,$$

and using (5), $\|\tilde{x}_{k+1} - \bar{x}\| \leq \alpha_k\|x_k - \bar{x}\|$, where

$$\alpha_k = \sqrt{\sigma_k(2-\sigma_k)\frac{a^2}{a^2+c_k^2} + (1-\sigma_k)^2} = \sqrt{1 - \sigma_k(2-\sigma_k)\frac{c_k^2}{a^2+c_k^2}}.$$

From (3), and from the fact that $\sigma_k(\bar{x}_k - x_k) = x_{k+1} - x_k$, we get

$$\begin{aligned}\|x_{k+1}-\bar{x}\| &\le \|\tilde{x}_{k+1}-\bar{x}\| + \sigma_k\epsilon_k \\ &\le \|\tilde{x}_{k+1}-\bar{x}\| + \sigma_k\delta_k\|\bar{x}_k - x_k\| \\ &\le \|\tilde{x}_{k+1}-\bar{x}\| + \delta_k\|x_{k+1}-x_k\| \\ &\le \|\tilde{x}_{k+1}-\bar{x}\| + \delta_k\|x_{k+1}-\bar{x}\| + \delta_k\|x_k-\bar{x}\|,\end{aligned}$$

so

$$\|x_{k+1}-\bar{x}\| \le \frac{\alpha_k+\delta_k}{1-\delta_k}\|x_k-\bar{x}\|,$$

where

$$\limsup \frac{\alpha_k+\delta_k}{1-\delta_k} = \limsup \alpha_k = \sqrt{1-\hat{\sigma}(2-\hat{\sigma})\frac{\bar{c}^2}{a^2+\bar{c}^2}} < 1. \quad \square$$

Note that if $\hat{\sigma} > 1$, the convergence rate estimate is strictly positive even if $\bar{c} = \infty$. However, if we choose $\sigma_k$ so that $\hat{\sigma} = 1$ and let $c_k \nearrow \infty$, we obtain superlinear convergence under the conditions of Proposition 2. It is also possible to derive a convergence rate estimate for the case where $\sigma_k < 1$ is allowed but that will not be needed in what follows.

**3. Convergence under local monotonicity.** In this section, we study a localized version of maximal monotonicity, and we show that it is enough to guarantee the local convergence of the over-relaxed proximal point algorithm. Although this result is not itself interesting from the practical point of view when dealing with truly nonmonotone problems, it will be one of the key ingredients in deriving our main convergence result in the next section.

DEFINITION 3. A set-valued mapping $T : H \rightrightarrows H$ is *monotone in* $U \subset H \times H$ if $\operatorname{gph} T \cap U$ is the graph of a monotone mapping. It will be called *maximal monotone* in $U$ if there exists a maximal monotone mapping $\overline{T} : H \rightrightarrows H$ such that $\operatorname{gph} \overline{T} \cap U = \operatorname{gph} T \cap U$.

Recall that by Zorn's lemma, any monotone mapping has a (possibly nonunique) maximal monotone extension. A mapping $T$ is maximal monotone in $U$ if and only if at least one of the maximal monotone extensions of $\operatorname{gph} T \cap U$ coincides with $\operatorname{gph} T$ on $U$. Note that both local monotonicity and local maximal monotonicity are preserved under inversion, just like their global counterparts. Using the global results in Rockafellar (1970), it is easily checked that the (generalized) subdifferential mapping $\partial f : H \rightrightarrows H$ of a lower semicontinuous function $f$ is maximal monotone in $X \times H$ if $X$ is an open convex set on which $f$ is convex.

The following gives a local version of the well-known continuity criterion for maximal monotonicity. In what follows, singleton sets $\{v\}$ will be denoted simply by $v$.

PROPOSITION 4. *Let $T : H \rightrightarrows H$ be such that for some open $X, Y \subset H$ the mapping $x \mapsto T(x) \cap Y$ is single valued and weakly continuous on $X$. If $T$ is monotone in $X \times Y$, then it is also maximal monotone in $X \times Y$.*

PROOF. Let $\overline{T}$ be a maximal monotone extension of $\operatorname{gph} T \cap (X \times Y)$ and let $(\tilde{x}, \tilde{v}) \in \operatorname{gph} \overline{T} \cap (X \times Y)$. It suffices to show that $(\tilde{x}, \tilde{v}) \in \operatorname{gph} T$. For any $z \in H$, there is an $\epsilon > 0$ such that $\tilde{x} + \epsilon z \in X$, so by monotonicity of $\overline{T}$

$$\langle T(\tilde{x}+\epsilon z)\cap Y - \tilde{v}, (\tilde{x}+\epsilon z) - \tilde{x}\rangle = \epsilon\langle T(\tilde{x}+\epsilon z)\cap Y - \tilde{v}, z\rangle \ge 0,$$

and then by the assumed weak continuity

$$\langle T(\tilde{x}) \cap Y - \tilde{v}, z\rangle \ge 0.$$

Since $z \in H$ was arbitrary, we must have $\tilde{v} = T(\tilde{x}) \cap Y$, so that $(\tilde{x}, \tilde{v}) \in \operatorname{gph} T$. $\square$

We will need the following in analyzing the proximal point algorithm for locally monotone mappings.

LEMMA 5. *Let $T : H \rightrightarrows H$ be monotone. For any $c > 0$ and $\sigma \in (0, 2)$, the mapping*

$$\sigma(I + cT)^{-1} + (1 - \sigma)I$$

*is nonexpansive, and its fixed points are the zeros of $T$.*

PROOF. We have $(I + c\,T)^{-1} = \frac{1}{2}(I + N)$, where $N$ is the unique nonexpansive mapping corresponding to $T$ as defined by Lawrence and Spingarn (1987) (see also Eckstein and Ferris 1999). We then get

$$\sigma(I + cT)^{-1} + (1 - \sigma)I = \frac{2 - \sigma}{2} I + \frac{\sigma}{2} N,$$

which is nonexpansive as a convex combination of two nonexpansive mappings. □

In what follows, the open ball with center $x$ and radius $\epsilon$ will be denoted by $\mathbb{B}(x, \epsilon)$, and the distance from a point $x$ to a set $C$ will be denoted by $d_C(x)$. When $x = 0$, we will simply write $\mathbb{B}(\epsilon)$ in place of $\mathbb{B}(0, \epsilon)$.

PROPOSITION 6. *Let $T : H \rightrightarrows H$ be a set-valued mapping that is maximal monotone in $X \times Y$, where $X, Y \subset H$ are open sets such that $0 \in Y$, and $T^{-1}(0) \cap X$ is nonempty and closed with $T^{-1}(0) \cap X + \mathbb{B}(\delta) \subset X$ for some $\delta > 0$. Let $c_k, \sigma_k$ be such that $\inf c_k > 0$, $\inf \sigma_k \geq 1$ and $\sup \sigma_k < 2$. If $d_{T^{-1}(0) \cap X}(x_0)$ is small enough, then there exist $\epsilon > 0$ and $\bar{x} \in T^{-1}(0) \cap X$ such that $x_0 \in \mathbb{B}(\bar{x}, \epsilon)$, and the rule*

$$x_{k+1} = \mathbb{B}(\bar{x}, \epsilon) \cap [\sigma_k (I + c_k T)^{-1}(x_k) + (1 - \sigma_k) x_k]$$

*generates a unique sequence $\{x_k\}$ converging weakly and Fejér monotonically to a point in $T^{-1}(0) \cap X \cap \mathbb{B}(\bar{x}, \epsilon)$.*

*If, in addition, $c_k, \sigma_k$, and the mapping $y \mapsto T^{-1}(y) \cap X$ satisfy the assumptions of Proposition 2, then $\{x_k\}$ converges linearly to $T^{-1}(0) \cap X = \{\bar{x}\}$ with rate*

$$\sqrt{1 - \hat{\sigma}(2 - \hat{\sigma}) \frac{\overline{c}^2}{a^2 + \overline{c}^2}} < 1.$$

PROOF. Let $\overline{T}$ be maximal monotone such that $\operatorname{gph} \overline{T} \cap (X \times Y) = \operatorname{gph} T \cap (X \times Y)$. Since $\overline{T}^{-1}(0)$ is convex and $\overline{T}^{-1}(0) \cap X = T^{-1}(0) \cap X$, which is nonempty and closed with $X$ open, we must have $\overline{T}^{-1}(0) \subset X$ and $\overline{T}^{-1}(0) = T^{-1}(0) \cap X$. Indeed, writing $\overline{T}^{-1}(0) = (\overline{T}^{-1}(0) \cap X) \cup (\overline{T}^{-1}(0) \cap X^c)$, we have a splitting of $\overline{T}^{-1}(0)$ into two disjoint closed sets, which, by convexity of $\overline{T}^{-1}(0)$, is impossible unless one of the sets is empty. Thus, to prove the first claim, it suffices by Theorem 1 and Lemma 5 to show that if $d_{\overline{T}^{-1}(0)}(x_0)$ is small enough, then there exist $\epsilon > 0$ and $\bar{x} \in T^{-1}(0) \cap X$ such that $x_0 \in \mathbb{B}(\bar{x}, \epsilon)$, and for every $k$

$$\mathbb{B}(\bar{x}, \epsilon) \cap [\sigma_k (I + c_k T)^{-1}(x) + (1 - \sigma)x] = \sigma_k \big(I + c_k \overline{T}\big)^{-1}(x) + (1 - \sigma_k)x \tag{6}$$

for all $x \in \mathbb{B}(\bar{x}, \epsilon)$.

By our assumptions on $X$ and $Y$, we can find an $\epsilon > 0$ such that $\overline{T}^{-1}(0) + \mathbb{B}(\epsilon) \subset X$ and $\mathbb{B}(2\epsilon/\underline{c}) \subset Y$, where $\underline{c} = \inf c_k > 0$. Assume that $d_{\overline{T}^{-1}(0)}(x_0) \leq \epsilon$, and let $\bar{x}$ be the projection of $x_0$ on the closed convex set $\overline{T}^{-1}(0)$. We will first show that

$$\big(I + c_k \overline{T}\big)^{-1}(x) = \mathbb{B}(\bar{x}, \epsilon) \cap (I + c_k T)^{-1}(x), \quad \forall\, x \in \mathbb{B}(\bar{x}, \epsilon). \tag{7}$$

Let $x \in \mathbb{B}(\bar{x}, \epsilon)$ be arbitrary, and define $x' = (I + c_k\overline{T})^{-1}(x)$. Since $(I + c_k\overline{T})^{-1}$ is nonexpansive and $\bar{x} = (I + c_k\overline{T})^{-1}(\bar{x})$, we have

$$x' \in \mathbb{B}(\bar{x}, \epsilon) \subset X,$$
$$c_k^{-1}(x - x') \in c_k^{-1}[\mathbb{B}(\bar{x}, \epsilon) - \mathbb{B}(\bar{x}, \epsilon)] \subset \mathbb{B}(2\epsilon/\underline{c}) \subset Y.$$

Since $x' = (I + c_k\overline{T})^{-1}(x)$ means that $(x', c_k^{-1}(x - x')) \in \operatorname{gph} \overline{T}$, we then also have $(x', c_k^{-1}(x - x')) \in \operatorname{gph} T$, or equivalently, $x' \in (I + c_k T)^{-1}(x)$, and so

$$\big(I + c_k\overline{T}\big)^{-1}(x) \in \mathbb{B}(\bar{x}, \epsilon) \cap (I + c_k T)^{-1}(x).$$

On the other hand, if $x'' \in \mathbb{B}(\bar{x}, \epsilon) \cap (I + c_k T)^{-1}(x)$, then

$$(x'', c_k^{-1}(x - x'')) \in (X \times Y) \cap \operatorname{gph} T \subset \operatorname{gph} \overline{T},$$

and so $x'' = \big(I + c_k\overline{T}\big)^{-1}(x)$, which proves (7).

Now, let

$$x' = \sigma_k\big(I + c_k\overline{T}\big)^{-1}(x) + (1 - \sigma_k)x,$$

which is a point in $\mathbb{B}(\bar{x}, \epsilon)$, by Lemma 5. Reordering, and using (7), we get

$$\frac{x' + (\sigma_k - 1)x}{\sigma_k} = \big(I + c_k\overline{T}\big)^{-1}(x) = \mathbb{B}(\bar{x}, \epsilon) \cap (I + c_k T)^{-1}(x) \subset (I + c_k T)^{-1}(x),$$

so that

$$x' \subset \mathbb{B}(\bar{x}, \epsilon) \cap [\sigma_k(I + c_k T)^{-1}(x) + (1 - \sigma_k)x].$$

On the other hand, if

$$x'' \subset \mathbb{B}(\bar{x}, \epsilon) \cap [\sigma_k(I + c_k T)^{-1}(x) + (1 - \sigma_k)x],$$

then

$$\frac{x'' + (\sigma_k - 1)x}{\sigma_k} \subset (I + c_k T)^{-1}(x),$$

where

$$\frac{x'' + (\sigma_k - 1)x}{\sigma_k} \in \mathbb{B}(\bar{x}, \epsilon)$$

by convexity of $\mathbb{B}(\bar{x}, \epsilon)$ and the fact that $\sigma_k \in [1, 2]$. So by (7)

$$\frac{x'' + (\sigma_k - 1)x}{\sigma_k} = \big(I + c_k\overline{T}\big)^{-1}(x),$$

and then

$$x'' \in \sigma_k\big(I + c_k\overline{T}\big)^{-1}(x) + (1 - \sigma_k)x,$$

which completes the proof of (6).

Since the sequence $\{x_k\}$ conforms to the algorithm of Theorem 1 applied to $\overline{T}$, and since $\overline{T}^{-1}(y) \cap X = T^{-1}(y) \cap X$ on $Y$, the convergence rate result follows from Proposition 2. □

Note that the condition that $T^{-1}(0) \cap X + \mathbb{B}(\delta) \subset X$ for some $\delta > 0$ holds in particular when $T^{-1}(0) \cap X$ is compact.

**4. Local convergence without monotonicity.** For any $T : H \rightrightarrows H$ and $\rho \in \mathbb{R}$, we have that $T_\rho^{-1}(0) = T^{-1}(0) + \rho 0 = T^{-1}(0)$, where $T_\rho$ denotes the Yosida regularization of $T$. This means that the solution sets to the inclusions $T(x) \ni 0$ and $T_\rho(x) \ni 0$ coincide. What makes the latter one more attractive from our point of view is that $T_\rho$ may be locally monotone even when $T$ is not. This happens exactly when $T^{-1}$ is *locally hypomonotone*, that is, when $T^{-1} + \rho I$ is locally monotone for some $\rho \in \mathbb{R}$.

For a lower semicontinuous function $f$, there is a close connection between hypomonotonicity of its subdifferential and *prox-regularity* of $f$; see Rockafellar and Wets (1998, §13F). In particular, 13.33 and 13.36 of Rockafellar and Wets (1998) imply that if $f$ is *lower*-$C^2$ on a neighborhood of a point $x$, then for every $v \in \partial f(x)$, $\partial f$ is hypomonotone in a neighborhood of $(x, v)$. Recall that lower-$C^2$ functions are the ones that can be expressed locally as $g - h$, where $g$ is finite and convex and $h$ is $C^2$; see Rockafellar and Wets (1998, 10.33).

A particularly convenient situation occurs when $T^{-1}$ has a Lipschitz localization. Recall that a mapping $T : H \rightrightarrows H$ is said to have a *Lipschitz localization* at a point $(x, v) \in \operatorname{gph} T$ if there are neighborhoods $X \ni x$ and $Y \ni v$ such that the mapping $x \mapsto T(x) \cap Y$ is single-valued and Lipschitz continuous on $X$.

PROPOSITION 7. *If $T^{-1}$ has a Lipschitz localization at $(0, \bar{x})$ with Lipschitz constant $\kappa$, then for $\rho \geq \kappa$, $T_\rho$ is maximal monotone in a neighborhood of $(0, \bar{x})$.*

PROOF. Let $Y \ni 0$ and $X \ni \bar{x}$ be open sets such that $y \mapsto T^{-1}(y) \cap X$ is Lipschitz continuous and define $F : H \rightrightarrows H$ by $\operatorname{gph} F = \operatorname{gph} T^{-1} \cap (Y \times X)$. Since $F$ is Lipschitz continuous on $Y$, we have for any $y_1, y_2 \in Y$ that

$$\begin{aligned}\langle F(y_1) + \rho y_1 - (F(y_2) + \rho y_2), y_1 - y_2 \rangle &\geq -\|F(y_1) - F(y_2)\| \|y_1 - y_2\| + \rho \|y_1 - y_2\|^2 \\ &\geq (\rho - \kappa)\|y_1 - y_2\|^2,\end{aligned}$$

so for $\rho \geq \kappa$, $F + \rho I$ is monotone. Since for any $y \in \operatorname{dom}(F + \rho I) = Y$

$$(F + \rho I)(y) = T^{-1}(y) \cap X + \rho y = (T^{-1}(y) + \rho y) \cap (X + \rho y),$$

we have $\operatorname{gph}(F + \rho I) = \operatorname{gph}(T^{-1} + \rho I) \cap U$, where

$$U = \{(y, x) \mid y \in Y, \;\; x \in X + \rho y\},$$

which is a neighborhood of $(0, \bar{x})$. This proves the local monotonicity of $T_\rho^{-1}$ and $T_\rho$, and the local maximal monotonicity follows by Proposition 4 from the local continuity of $T_\rho^{-1}$. $\square$

When $T_\rho$ is locally maximal monotone, Proposition 6 says that if we apply the proximal point algorithm to it, we obtain a sequence converging to a point in $T_\rho^{-1}(0) = T^{-1}(0)$. The following identity shows how to get the resolvants of $T_\rho$ from those of $T$.

LEMMA 8. *For any $T : H \rightrightarrows H$, and $c, \rho \in \mathbb{R}$ such that $c + \rho$, $c \neq 0$, we have*

$$(I + cT_\rho)^{-1} = \frac{c}{c+\rho}[I + (c+\rho)T]^{-1} + \frac{\rho}{c+\rho} I.$$

PROOF.

$$\begin{aligned}x' \in (I + cT_\rho)^{-1}(x) &\iff \frac{x - x'}{c} \in T_\rho(x') \\ &\iff x' \in (T^{-1} + \rho I)\Big(\frac{x - x'}{c}\Big) \\ &\iff \frac{(c+\rho)x' - \rho x}{c} \in T^{-1}\Big(\frac{x - x'}{c}\Big) \\ &\iff \frac{x - x'}{c} \in t\Big[\frac{(c+\rho)x' - \rho x}{c}\Big] \\ &\iff \frac{1}{c+\rho}\Big[x - \frac{(c+\rho)x' - \rho x}{c}\Big] \in T\Big[\frac{(c+\rho)x' - \rho x}{c}\Big] \\ &\iff x \in [I + (c+\rho)T]\Big[\frac{(c+\rho)x' - \rho x}{c}\Big] \\ &\iff \frac{(c+\rho)x' - \rho x}{c} \in [I + (c+\rho)T]^{-1}(x) \\ &\iff x' \in \frac{c}{c+\rho}[I + (c+\rho)T]^{-1}(x) + \frac{\rho}{c+\rho} x. \quad \square\end{aligned}$$

Thus, for $c > 0$, $\rho \geq 0$, the resolvant of $T_\rho$ is obtained from that of $T$ by "under-relaxation." Applying the proximal point algorithm with an appropriate over-relaxation, we can compensate for this to obtain the ordinary proximal point algorithm.

THEOREM 9. *Let $T : H \rightrightarrows H$ be a set-valued mapping such that for some $\rho \geq 0$, $T_\rho$ is maximal monotone in $X \times Y$, where $X, Y \subset H$ are open sets such that $0 \in Y$ and $T^{-1}(0) \cap X$ is nonempty and closed with $T^{-1}(0) \cap X + \mathbb{B}(\delta) \subset X$ for some $\delta > 0$. If $d_{T^{-1}(0)\cap X}(x_0)$ is small enough and* $\inf c_k > 2\rho$, *then there exist $\epsilon > 0$ and $\bar{x} \in T^{-1}(0) \cap X$ such that $x_0 \in \mathbb{B}(\bar{x}, \epsilon)$, and the rule*

$$x_{k+1} = \mathbb{B}(\bar{x}, \epsilon) \cap (I + c_k T)^{-1}(x_k) \tag{8}$$

*defines a unique sequence $\{x_k\}$, converging weakly and Fejér monotonically to a point in $T^{-1}(0) \cap X$.*

*If, in addition, $c_k$ and the mapping $y \mapsto T^{-1}(y) \cap (X - \rho Y)$ satisfy the assumptions of Proposition* 2, *then $\{x_k\}$ converges linearly to $T^{-1}(0) \cap X = \{\bar{x}\}$ with rate*

$$\sqrt{1 - \frac{\overline{c}}{\overline{c} - \rho}\left(2 - \frac{\overline{c}}{\overline{c} - \rho}\right)\frac{\overline{c}^2}{(a + \rho)^2 + \overline{c}^2}} < 1.$$

PROOF. The condition $\inf c_k > 2\rho$ implies $\inf c_k' > 0$, $\inf \sigma_k \geq 1$, and $\sup \sigma_k < 2$, where

$$c_k' = c_k - \rho \quad \text{and} \quad \sigma_k = \frac{c_k}{c_k - \rho}.$$

Thus, by Proposition 6, there exist $\epsilon > 0$ and $\bar{x} \in T^{-1}(0) \cap X$ such that $x_0 \in \mathbb{B}(\bar{x}, \epsilon)$, and the rule

$$x_{k+1} = \mathbb{B}(\bar{x}, \epsilon) \cap [\sigma_k (I + c_k' T_\rho)^{-1}(x_k) + (1 - \sigma_k) x_k]$$

defines a sequence $\{x_k\}$ with the required properties. The first claim follows by noting that by Lemma 8,

$$\begin{aligned}
\sigma_k (I + c_k' T_\rho)^{-1} + (1 - \sigma_k) I &= \sigma_k \left[\frac{c_k'}{c_k' + \rho}(I + (c_k' + \rho)T)^{-1} + \frac{\rho}{c_k' + \rho} I\right] + (1 - \sigma_k) I \\
&= \frac{\sigma_k (c_k - \rho)}{c_k}(I + c_k T)^{-1} + \left(\frac{\sigma_k \rho}{c_k} + 1 - \sigma_k\right) I \\
&= (I + c_k T)^{-1}.
\end{aligned}$$

If the mapping $y \mapsto T^{-1}(y) \cap (X - \rho Y)$ satisfies Condition (4), then

$$\begin{aligned}
y \in U, \quad x \in T_\rho^{-1}(y) \cap X &\iff y \in U, \quad x \in T^{-1}(y) \cap (X - \rho y) + \rho y \\
&\implies y \in U, \quad x \in T^{-1}(y) \cap (X - \rho Y) + \rho y \\
&\implies y \in U, \quad x = x' + \rho y, \quad x' \in T^{-1}(y) \cap (X - \rho Y) \\
&\implies \|x\| \leq \|x'\| + \rho \|y\| \leq a\|y\| + \rho \|y\|,
\end{aligned}$$

so the mapping $y \mapsto T_\rho^{-1}(y) \cap X$ satisfies (4) with the constant $a$ replaced by $a + \rho$, so that Proposition 6 gives the rate

$$\sqrt{1 - \frac{\overline{c}}{\overline{c} - \rho}\left(2 - \frac{\overline{c}}{\overline{c} - \rho}\right)\frac{\overline{c}^2}{(a + \rho)^2 + \overline{c}^2}} < 1. \quad \square$$

Rule (8) means that $x_{k+1}$ is the unique solution in $\mathbb{B}(\bar{x}, \epsilon)$ to the inclusion

$$T(x) + c_k^{-1}(x - x_k) \ni 0. \tag{9}$$

Since the subproblems (9) are not in general monotone, it is necessary to restrict the region where the iterates are sought in order to make them well defined.

The condition $\inf c_k > 2\rho$ is reminiscent of the condition used in Bertsekas (1982, Proposition 2.7) and in Contesse-Becker (1993, Theorem 2.3), which concern the multiplier method for nonlinear programming. There, $\rho$ is the negative of the smallest eigenvalue of the Hessian of the associated value function. If, in Theorem 9, $T^{-1}$ is the subdifferential of the value function, then the condition that $T^{-1} + \rho I$ be locally maximal monotone is a natural generalization of the condition that $\rho$ be larger than the negative of the smallest eigenvalue of the Hessian. This new condition allows $T^{-1}$ to be nonsmooth and even set-valued, and it does not require $T^{-1}$ to be a subdifferential of a function.

Combining Theorem 9 with Proposition 7, we obtain the following result for the case where $T^{-1}$ has a Lipschitz localization.

COROLLARY 10. *Assume that $T^{-1}$ has a Lipschitz localization at $(0, \bar{x})$, with Lipschitz constant $\kappa$. Let $\{c_k\}$ be such that $\inf c_k > 2\kappa$, and $c_k \nearrow \bar{c} \le \infty$. If $\|x_0 - \bar{x}\|$ is small enough, then there exists an $\epsilon > 0$ such that $x_0 \in \mathbb{B}(\bar{x}, \epsilon)$ and the rule*

$$x_{k+1} = \mathbb{B}(\bar{x}, \epsilon) \cap (I + c_k T)^{-1}(x_k)$$

*generates a unique sequence $\{x_k\}$ converging linearly and Fejér monotonically to $\bar{x}$ with rate*

$$\sqrt{1 - \frac{\bar{c}}{\bar{c} - \kappa}\left(2 - \frac{\bar{c}}{\bar{c} - \kappa}\right)\frac{\bar{c}^2}{(2\kappa)^2 + \bar{c}^2}} < 1.$$

**5. General form of multiplier methods.** Let $H$ and $V$ be Hilbert spaces. For an arbitrary $S : H \rightrightarrows H$, maximal monotone $T : V \rightrightarrows V$, and a $C^2$ function $F : H \to V$, consider the problem of finding a solution to the inclusion

$$S(x) + \nabla F(x)^* T(F(x)) \ni 0, \tag{$P$}$$

where $\nabla F(x)^*$ is the adjoint of the Jacobian of $F$ at a point $x$. This provides a flexible model for various applications, and it is convenient for developing and analyzing multiplier methods based on the proximal point algorithm.

The structure of $(P)$ supports a natural duality theory, which adjoins to the *primal problem* $(P)$ dual and primal-dual problems much as in the duality theory of convex programming. Denote

$$F_0(x) = S(x) + \nabla F(x)^* T(F(x)),$$

so that $(P)$ can be written as $F_0(x) \ni 0$. Define the *Lagrangian* $L : H \times V \rightrightarrows H \times V$ by

$$L(x, y) = (\nabla F(x)^* y, -F(x)) + S(x) \times T^{-1}(y),$$

and consider the *primal-dual problem*

$$L(x, y) \ni (0, 0). \tag{$PD$}$$

The Lagrangian is related to $F_0$ by

$$\begin{aligned} F_0(x) &= \{v \in H \mid \exists y \in V : \ (v, 0) \in L(x, y)\} \\ &= \{S(x) + \nabla F(x)^* y \mid \exists y \in V : \ 0 \in -F(x) + T^{-1}(y)\} \\ &= \{S(x) + \nabla F(x)^* y \mid \exists y \in V : \ y \in T(F(x))\} \\ &= S(x) + \nabla F(x)^* T(F(x)). \end{aligned} \tag{10}$$

Expression (10) and a desire for symmetry suggest defining a mapping $G_0 : V \rightrightarrows V$ by

$$G_0(y) = \{u \in V \mid \exists x \in H: \ (0, u) \in L(x, y)\} \tag{11}$$

and to define the *dual problem*

$$G_0(y) \ni 0. \tag{$D$}$$

Before going into an example, we recall some notation and facts from convex analysis (Rockafellar 1970). The *indicator function* $\delta_C$ of a set $C \subset H$ is the function that has value 0 for $x \in C$ and $+\infty$ otherwise. The *normal cone operator* of $C$ is the subdifferential of $\delta_C$:

$$N_C(x) = \{v \in H \mid \langle v, y - x\rangle \leq 0 \, \forall \, y \in C\}.$$

If $C$ is closed and convex, then $\delta_C$ is a closed convex function and $N_C$ is maximal monotone. The *polar cone* of a cone $K \subset H$ is the closed convex cone

$$K^* = \{v \in H \mid \langle v, x\rangle \leq 0 \, \forall \, x \in K\}.$$

If $K$ is closed and convex, the functions $\delta_K$ and $\delta_{K^*}$ are conjugates to each other, so that $K^{**} = K$ and $N_{K^*} = N_K^{-1}$.

Example 11. Consider the nonlinear programming problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \begin{cases} = 0 & \text{for } i = 1, \ldots, r \\ \leq 0 & \text{for } i = r+1, \ldots, m \end{cases} \end{array} \tag{$P_{NLP}$}$$

for real-valued $C^2$ functions $f_i$. Defining $S = \nabla f_0$, $H = \mathbb{R}^n$, $F(x) = (f_1(x), \ldots, f_m(x))$ and $T = N_K$ for

$$K = \{0\}^r \times \mathbb{R}_-^{m-r},$$

the mapping $F_0$ becomes

$$F_0(x) = \nabla f_0(x) + \nabla F(x)^* N_K(F(x)).$$

At any point $\bar{x}$ satisfying the constraint qualification

$$y \in N_K(F(\bar{x})), \quad \nabla F(\bar{x})^* y = 0 \Longrightarrow y = 0,$$

$F_0(\bar{x})$ is equal to the subdifferential of the essential objective $f_0 + \delta_K \circ F$ of $(P_{NLP})$ at the point $\bar{x}$; see Rockafellar and Wets (1998, 10.8). In this case, $(P)$ is a necessary optimality condition for $(P_{NLP})$ (Rockafellar and Wets 1998, 10.1).

The Lagrangian now becomes

$$L(x, y) = (\nabla F(x)^* y, -F(x)) + \nabla f(x) \times N_{K^*}(y),$$

where we have used the relation $N_K^{-1} = N_{K^*}$. So now $(PD)$ is just the classical Karush-Kuhn-Tucker system for $(P_{NLP})$. Defining the *Lagrangian function* $l(x, y) = f_0(x) + \langle y, F(x)\rangle - \delta_{K^*}(y)$, $L$ can be expressed as

$$L(x, y) = \partial l(x, y) \times \partial[-l](x, y),$$

and one can define the dual objective $g(y) = \inf_x l(x, y)$, which is always a closed convex function. If $(P_{NLP})$ is a convex program, the solutions of $(PD)$ are exactly the saddle-points of $l$, and they can be expressed as $(x, y)$, where $x$ solves $(P_{NLP})$ and $y$ solves the dual problem of maximizing $g$; see, for example, Rockafellar (1974, §7). If, in addition, the dual problem is strongly consistent in the sense of Rockafellar (1970), the mapping $G_0$ is the subdifferential mapping of $g$, so that the problem of maximizing $g$ is equivalent to the dual inclusion $G_0(y) \ni 0$.

In the absence of convexity, the traditional duality theory collapses since the saddle-points of $l$ or the maximizers of $g$ may not have anything to do with the solutions of $(P_{NLP})$. However, for the inclusions $(P)$, $(PD)$, and $(D)$ we always have the following duality result.

THEOREM 12. *An $\bar{x}$ solves $(P)$ if and only if there is a $\bar{y}$ such that $(\bar{x}, \bar{y})$ solves $(PD)$, in which case $\bar{y}$ solves $(D)$. Dually, a $\bar{y}$ solves $(D)$ if and only if there is an $\bar{x}$ such that $(\bar{x}, \bar{y})$ solves $(PD)$, in which case $\bar{x}$ solves $(P)$. Moreover, the solutions $(\bar{x}, \bar{y})$ of $(PD)$ satisfy $\bar{y} \in T(F(\bar{x}))$.*

PROOF. Using Expression (10), we see that $\bar{x}$ solves $(P)$ if and only if there is a $\bar{y}$ such that $(0, 0) \in L(\bar{x}, \bar{y})$. This means that $(\bar{x}, \bar{y})$ solves $(PD)$, and by the definition of $G_0$, that $\bar{y}$ solves $(D)$. Dually, if $\bar{y}$ solves $(D)$, there exists, by the definition of $G_0$, an $\bar{x}$ such that $(0, 0) \in L(\bar{x}, \bar{x})$, which by Expression (10) means that $\bar{x}$ solves $(P)$. To prove the last claim, we note that by the definition of $L$, $(0, 0) \in L(\bar{x}, \bar{y})$ implies $0 \in -F(\bar{x}) + T^{-1}(\bar{y})$, which is the same as $\bar{y} \in T(F(\bar{x}))$. □

It is clear from the proof that monotonicity of $T$ and the continuity properties of $F$ are irrelevant in Theorem 12. These properties are needed in all the applications of this paper, so for simplicity, we have chosen to assume them from the start. The above duality framework is a special case of the more general construction introduced in Pennanen (2000), but we will refer to it as the *general duality framework* in what follows. For other duality frameworks for generalized equations, see Attouch and Théra (1996), Pennanen (1999), Auslender and Teboulle (2000), Robinson (1999), and the references therein.

**5.1. Method of multipliers.** Applying the proximal point algorithm to $(D)$, we get an iteration where $y_{k+1}$ is obtained from $y_k$ by solving (if possible) the modified dual problem

$$G_0(y) + c_k^{-1}(y - y_k) \ni 0.$$

It turns out that this is exactly the dual problem in the general duality framework with the mapping $T$ replaced by the mapping

$$T_k(u) = (I + c_k T^{-1})^{-1}(y_k + c_k u). \tag{12}$$

Indeed, since $T_k^{-1}(y) = T^{-1}(y) + c_k^{-1}(y - y_k)$, the Lagrangian $L_k$ corresponding to $T_k$ can be written as

$$\begin{aligned} L_k(x, y) &= (\nabla F(x)^* y, -F(x)) + S(x) \times T_k^{-1}(y) \\ &= (\nabla F(x)^* y, -F(x)) + S(x) \times [T^{-1}(y) + c_k^{-1}(y - y_k)] \\ &= L(x, y) + (0, c_k^{-1}(y - y_k)), \end{aligned}$$

so the dual problem corresponding to the primal-dual problem

$$L_k(x, y) \ni (0, 0) \tag{$PD_k$}$$

reads

$$G_0^k(y) \ni 0, \tag{$D_k$}$$

where

$$\begin{aligned} G_0^k(y) &= \{u \in V \mid \exists x \in H : (0,u) \in L_k(x,y)\} \\ &= \{u \in V \mid \exists x \in H : (0,u) \in L(x,y)\} + c_k^{-1}(y - y_k) \\ &= G_0(y) + c_k^{-1}(y - y_k). \end{aligned}$$

Just as in (10), we find the corresponding primal problem to be

$$(P_k) \qquad S(x) + \nabla F(x)^* T_k(F(x)) \ni 0.$$

Note that $T_k$ is single-valued and Lipschitz continuous since it is the composition of $u \mapsto y_k + c_k u$ and the resolvant of the maximal monotone mapping $T^{-1}$.

Since $(P_k)$, $(PD_k)$, and $(D_k)$ are in the format of the general duality framework, their solutions are related according to Theorem 12. In particular, we have the following.

LEMMA 13. *The solutions to the proximal point subproblem $(D_k)$ (if any) are precisely the points of the form $y_{k+1} = T_k(F(x_{k+1}))$, where $x_{k+1}$ is a solution of $(P_k)$. Moreover, the points $(x_{k+1}, y_{k+1})$ are the solutions of $(PD_k)$.*

In summary, we can write the proximal point algorithm for $(D)$ in the following form.

**Algorithm 1** (Method of Multipliers).
*Step* 0. Choose $y_0$ and set $k = 0$.
*Step* 1. Solve

$$S(x) + \nabla F(x)^* T_k(F(x)) \ni 0$$

for $x_{k+1}$.
*Step* 2. Set $y_{k+1} = T_k(F(x_{k+1}))$, $k = k+1$ and go to 1.

In practical applications of the method of multipliers, it is usually possible to write an explicit expression for the mapping $T_k$, so that Step 2 does not present any computational difficulties. For example, in Example 11, $T = N_K$, so that by Rockafellar and Wets (1998, 6.17),

$$T_k(u) = (I + c_k N_K^{-1})^{-1}(y_k + c_k u) = (I + c_k N_{K^*})^{-1}(y_k + c_k u) = P_{K^*}(y_k + c_k u), \tag{13}$$

which is just the projection of $y_k + c_k u$ on the cone

$$K^* = \{y \in \mathbb{R}^m \mid y_i \geq 0,\ i = r+1, \ldots, m\}.$$

The problems to be solved in Step 1 are of the same form as the original problem $(P)$, but the mapping $T$ has been replaced by $T_k$. The reason why the method of multipliers might be useful is that $T_k$ is guaranteed to be single-valued and Lipschitz continuous, whereas $T$ is typically set-valued with $\operatorname{dom} T \neq V$. In the case where $S$ is single-valued (e.g., in Example 11), $(P_k)$ is an *equation* instead of an inclusion.

THEOREM 14. *Assume that $L^{-1}$ has a Lipschitz localization at a point $(0, 0, \bar{x}, \bar{y})$, with Lipschitz constant $\kappa$. Let $\{c_k\}$ be such that $\inf c_k > 2\kappa$, and $c_k \nearrow \bar{c} \leq \infty$. If $\|y_0 - \bar{y}\|$ is small enough, then there exists a sequence $\{(x_k, y_k)\}$ conforming to Algorithm 1 along with open neighborhoods $X_k$, such that for each $k, x_{k+1}$ is the unique solution in $X_k$ to $(P_k)$. The sequence $\{y_k\}$ converges linearly and Fejér monotonically to $\bar{y}$ with rate*

$$\sqrt{1 - \frac{\bar{c}}{\bar{c} - \kappa}\left(2 - \frac{\bar{c}}{\bar{c} - \kappa}\right)\frac{\bar{c}^2}{(2\kappa)^2 + \bar{c}^2}} < 1.$$

*The sequence $\{x_k\}$ converges to $\bar{x}$, and there is an $M \in \mathbb{R}$ such that $\|x_k - \bar{x}\| \leq M\|y_k - \bar{y}\|$.*

PROOF. By the Lipschitz assumption, there exist open neighborhoods $X \ni \bar{x}$, $Y \ni \bar{y}$, $W \ni 0$, and $U \ni 0$ such that the mapping

$$(w, u) \mapsto (X \times Y) \cap L^{-1}(w, u) \tag{14}$$

is Lipschitz on $W \times U$. It follows that the mapping

$$u \mapsto P_V[(X \times Y) \cap L^{-1}(0, u)]$$

is Lipschitz on $U$. Denote by $\widetilde{G}_0$ the inverse of this mapping. Then by Corollary 10, there exists an $\epsilon > 0$ such that for any $y_0 \in \mathbb{B}(\bar{y}, \epsilon)$, the rule

$$y_{k+1} \in \mathbb{B}(\bar{y}, \epsilon), \quad \widetilde{G}_0(y_{k+1}) + c_k^{-1}(y_{k+1} - y_k) \ni 0 \tag{15}$$

generates a unique sequence converging linearly and Fejér monotonically to $\bar{y}$. Rule (15) can be written as

$$y_{k+1} \in \mathbb{B}(\bar{y}, \epsilon), \quad y_{k+1} \in \widetilde{G}_0^{-1}(c_k^{-1}(y_k - y_{k+1})),$$

which by definition of $\widetilde{G}_0$ means that there exists an $x_{k+1}$ such that

$$\begin{aligned} &(x_{k+1}, y_{k+1}) \in [X \times (Y \cap \mathbb{B}(\bar{y}, \epsilon))] \cap L^{-1}(0, c_k^{-1}(y_k - y_{k+1})) \\ &\iff (x_{k+1}, y_{k+1}) \in L^{-1}(0, c_k^{-1}(y_k - y_{k+1})), \quad x_{k+1} \in X, \quad y_{k+1} \in Y \cap \mathbb{B}(\bar{y}, \epsilon) \\ &\iff L_k(x_{k+1}, y_{k+1}) \ni (0, 0), \quad x_{k+1} \in X, \quad y_{k+1} \in Y \cap \mathbb{B}(\bar{y}, \epsilon). \end{aligned} \tag{16}$$

By Lemma 13, Rule (15) thus means that $y_{k+1} = T_k(F(x_{k+1}))$, where $x_{k+1}$ solves $(P_k)$ and $x_{k+1} \in X_k := X \cap (T_k \circ F)^{-1}(Y \cap \mathbb{B}(\bar{y}, \epsilon))$, which is an open set by the continuity of $T_k \circ F$. Relation (16) and the Lipschitz continuity of Mapping (14) imply the uniqueness of $x_{k+1}$ in $X_k$.

Corollary 10 also gives the rate of linear convergence for $\{y_k\}$, so there are constants $r_k$ converging to this rate, with $\|y_{k+1} - \bar{y}\| \le r_k \|y_k - \bar{y}\|$ for all $k$. Then Relation (16) and the Lipschitz continuity of Mapping (14) yield

$$\begin{aligned} \|x_{k+1} - \bar{x}\| &\le \|(x_{k+1}, y_{k+1}) - (\bar{x}, \bar{y})\| \\ &\le \kappa \|c_k^{-1}(y_k - y_{k+1})\| \\ &\le \kappa c_k^{-1}(\|y_k - \bar{y}\| + \|y_{k+1} - \bar{y}\|) \\ &\le \kappa c_k^{-1}(\|y_k - \bar{y}\| + r_k \|y_k - \bar{y}\|) \\ &= \kappa \frac{1 + r_k}{c_k} \|y_k - \bar{y}\| \le M \|y_k - \bar{y}\|, \end{aligned}$$

where $M = \sup \kappa((1 + r_k)/c_k) < \infty$. □

To obtain local convergence of $\{y_k\}$ to a solution of $(D)$, it would suffice by Theorem 9 to have $G_0^{-1}$ locally hypomonotone instead of the existence of a Lipschitz localization for $L^{-1}$. However, this would not give the convergence of $\{x_k\}$ to a solution of $(P)$, which is what we are really interested in. The situation is very similar to the convergence analysis in the monotone case: Assuming only maximal monotonicity of the underlying mappings, one cannot guarantee the convergence of the primal sequence, but only that all its accumulation points are solutions of the primal problem (Rockafellar 1976b, Eckstein 1993, Eckstein and Ferris 1999). The convergence properties obtained above are similar to the ones for nonlinear nonconvex programming in Bertsekas (1982), Ito and Kunisch (1990), Contesse-Becker (1993), and Conn et al. (1996), where certain regularity properties were assumed to hold at the solution; see also §7.

**5.2. Proximal method of multipliers.** If we apply the proximal point algorithm to $(PD)$, we get an iteration where $(x_{k+1}, y_{k+1})$ is obtained by solving

$$(PD_k) \qquad L(x,y) + c_k^{-1}(x - x_k, y - y_k) \ni (0,0).$$

This is the primal-dual problem in the general duality framework if we replace the mapping $S$ by $S_k(x) = S(x) + c_k^{-1}(x - x_k)$ and the mapping $T$ by

$$T_k(u) = (I + c_k T^{-1})^{-1}(y_k + c_k u),$$

like in the method of multipliers. Indeed, with these modifications, the Lagrangian becomes

$$\begin{aligned} L_k(x,y) &= (\nabla F(x)^* y, -F(x)) + S_k(x) \times T_k^{-1}(y) \\ &= (\nabla F(x)^* y, -F(x)) + [S(x) + c_k^{-1}(x - x_k)] \times [T^{-1}(y) + c_k^{-1}(y - y_k)] \\ &= L(x,y) + c_k^{-1}(x - x_k, y - y_k). \end{aligned}$$

The primal problem associated with the mappings $S_k$ and $T_k$ according to the general duality framework is

$$(P_k) \qquad S(x) + c_k^{-1}(x - x_k) + \nabla F(x)^* T_k(F(x)) \ni 0.$$

The corresponding dual problem could again be defined through (11), but it will not be used here.

Applying Theorem 12 to $(P_k)$ and $(PD_k)$, we get the following.

LEMMA 15. *The solutions to the proximal point subproblem $(PD_k)$ (if any) are precisely the points of the form $(x_{k+1}, T_k(F(x_{k+1})))$, where $x_{k+1}$ is a solution of $(P_k)$.*

We are thus lead to the following algorithm.

**Algorithm 2** (Proximal Method of Multipliers).
*Step* 0. Choose $(x_0, y_0)$ and set $k = 0$.
*Step* 1. Solve

$$(P_k) \qquad S(x) + c_k^{-1}(x - x_k) + \nabla F(x)^* T_k(F(x)) \ni 0$$

for $x_{k+1}$,
*Step* 2. Set $y_{k+1} = T_k(F(x_{k+1}))$, $k = k+1$ and go to 1.

The only difference from Algorithm 1 is that now we have an additional regularizing term $c_k^{-1}(x - x_k)$ in $F_k$. This small difference allows us to get the following improved convergence result.

THEOREM 16. *Assume that $L^{-1}$ has a Lipschitz localization at a point $(0,0,\bar{x},\bar{y})$ with Lipschitz constant $\kappa$. Let $\{c_k\}$ be such that* $\inf c_k > 2\kappa$ *and $c_k \nearrow \bar{c} \le \infty$. If $\|(x_0, y_0) - (\bar{x}, \bar{y})\|$ is small enough, then there exists a sequence $\{(x_k, y_k)\}$ conforming to Algorithm 2 along with open neighborhoods $X_k$ such that for each $k$, $x_{k+1}$ is the unique solution in $X_k$ to $(P_k)$. Furthermore, the sequence $\{(x_k, y_k)\}$ converges linearly and Fejér monotonically to $(\bar{x}, \bar{y})$ with rate*

$$\sqrt{1 - \frac{\bar{c}}{\bar{c} - \kappa}\left(2 - \frac{\bar{c}}{\bar{c} - \kappa}\right)\frac{\bar{c}^2}{(2\kappa)^2 + \bar{c}^2}} < 1.$$

Proof. Since Algorithm 2 is the proximal point algorithm applied to $(PD)$, Corollary 10 implies the existence of an $\epsilon > 0$ and a sequence $\{(x_k, y_k)\}$ with the claimed properties such that for each $k$, $(x_{k+1}, y_{k+1})$ is the unique solution in $\mathbb{B}((\bar{x}, \bar{y}), \epsilon)$ to $(PD_k)$. Let $X_k = H_k^{-1}(\mathbb{B}((\bar{x}, \bar{y}), \epsilon))$, where $H_k(x) = (x, T_k(F(x)))$. Then $X_k$ is open by the continuity of $H_k$, and $x_{k+1}$ is the unique solution of $(P_k)$ in $X_k$. Indeed, if $x'_{k+1}$ were any other such solution, then by Lemma 15, $H_k(x'_{k+1}) \in \mathbb{B}((\bar{x}, \bar{y}), \epsilon)$ would solve $(PD_k)$, but $H_k(x'_{k+1}) \neq (x_{k+1}, y_{k+1})$. $\square$

Note that, using Theorem 9 instead of Corollary 10 in the above proof, we would obtain a convergence result under the weaker condition of local hypomonotonicity of $L^{-1}$. However, the above version with Lipschitz localization is more directly related to the existing theory on variational problems, and it turns out to be directly applicable in familiar contexts. In the following two sections, we consider two special classes of problems that can be expressed in the form $(P)$ and for which more specific conditions are available that guarantee the existence of such localizations.

**6. Application to variational inequalities.** Consider problem $(P)$ in the case where $S = f + N_X$ and $T = N_K$ for a $C^1$ function $f : H \to H$, a closed convex set $X \subset H$, and a closed convex cone $K \subset V$:

$$(P_{VI}) \qquad f(x) + N_X(x) + \nabla F(x)^* N_K(F(x)) \ni 0.$$

To motivate this problem format, let $C = \{x \in X \mid F(x) \in K\}$, which is nonconvex in general. By Rockafellar and Wets (1998, Theorem 6.14), the (generalized) normal cone mapping $N_C$ of $C$ satisfies

$$N_C(x) \supset N_X(x) + \nabla F(x)^* N_K(F(x)),$$

where equality holds at any point $x$ satisfying the constraint qualification

$$(CQ) \qquad y \in N_K(F(x)), \quad -\nabla F(x)^* y \in N_X(x) \Longrightarrow y = 0.$$

So, any solution of $(P_{VI})$ solves

$$f(x) + N_C(x) \ni 0, \tag{17}$$

and vice versa at any point satisfying $(CQ)$. The inclusion (17) can be seen as a generalized variational inequality. Indeed, when $C$ is convex, (17) means that

$$x \in C, \quad \langle f(x), y - x\rangle \geq 0, \quad \forall\, y \in C.$$

The idea of applying multiplier methods to variational inequalities goes back to Rockafellar (1978), where the case of monotone variational inequalities with convex constraint functions was treated. Gabay (1983) derived a multiplier method for variational inequalities, which can be written in the form

$$S(x) + B^* \partial f(Bx) \ni 0,$$

where $S$ is maximal monotone, $B$ is linear and continuous, and $f$ is a closed convex function. Eckstein and Ferris (1999) introduced multiplier methods that solve a monotone complementarity problem by solving a sequence of smooth equations. This work was extended by Auslender and Teboulle (2000) to monotone variational inequalities with convex constraint functions. The convergence results in Rockafellar (1978), Gabay (1983), Eckstein and Ferris (1999), and Auslender and Teboulle (2000) were based on generalized dualization and on convergence of proximal point methods for maximal monotone mappings.

Specializing the results of the previous section, we can use a similar technique to obtain local convergence results for variational inequalities without assuming monotonicity.

In the current situation, the mapping $T_k$ in (12) can be written as in (13)

$$T_k(u) = (I + c_k N_K^{-1})^{-1}(y_k + c_k u) = P_{K^*}(y_k + c_k u),$$

so that Algorithm 2 becomes:

**Algorithm 3** (Proximal Method of Multipliers).
*Step* 0. Choose $(x_0, y_0)$ and set $k = 0$.
*Step* 1. Solve

$$(P_{VI}^k) \qquad f(x) + c_k^{-1}(x - x_k) + N_X(x) + \nabla F(x)^* P_{K^*}(y_k + c_k F(x)) \ni 0,$$

for $x_{k+1}$.
*Step* 2. Set $y_{k+1} = P_{K^*}(y_k + c_k F(x_{k+1}))$, $k = k + 1$ and go to 1.
Note that since by assumption $X$ is convex, problem $(P_{VI}^k)$ can be written as an ordinary variational inequality even though the original problem $(P_{VI})$ could not. The case where $X = H$ is especially interesting, since then $(P_{VI}^k)$ becomes an equation. Even though this equation is in general nonsmooth, there are numerical algorithms such as generalized Newton methods (see, for example, Robinson 1994) that could be used in solving it.

To apply Theorem 16, we need to establish the existence of a Lipschitz localization for the inverse of the Lagrangian, which in this case can be written as

$$L(x, y) = (f(x) + \nabla F(x)^* y, -F(x)) + N_X(x) \times N_{K^*}(y).$$

We see that now $(PD)$ is a variational inequality for the $C^1$ function

$$g(x, y) = (f(x) + \nabla F(x)^* y, -F(x))$$

and the closed convex set $X \times K^*$. According to Robinson (1980), $L^{-1}$ has a Lipschitz localization at a solution $(\bar{x}, \bar{y})$ of $(PD)$ if the inverse of the semilinearized mapping

$$\widetilde{L}(x, y) = g(\bar{x}, \bar{y}) + \nabla g(\bar{x}, \bar{y})(x - \bar{x}, y - \bar{y}) + N_X(x) \times N_{K^*}(y)$$

has one at $(0, 0)$. Following Robinson (1980), such a property of the variational inequality $(PD)$ will be called *strong regularity at* $(\bar{x}, \bar{y})$. We can now state Theorem 16 in the following form.

THEOREM 17. *Assume that* $(PD_{VI})$ *is strongly regular at* $(\bar{x}, \bar{y})$. *If* $\|(x_0, y_0) - (\bar{x}, \bar{y})\|$ *is small enough, and if* $c_k$ *are large enough with* $c_k \nearrow \bar{c} \leq \infty$, *then there exists a sequence* $\{(x_k, y_k)\}$ *conforming to Algorithm* 3 *along with open neighborhoods* $X_k$ *such that for each* $k$, $x_{k+1}$ *is the unique solution in* $X_k$ *to* $(P_k)$. *Then also, the sequence* $\{(x_k, y_k)\}$ *converges linearly and Fejér monotonically to* $(\bar{x}, \bar{y})$ *with rate* $r(\bar{c}) < 1$ *that is decreasing in* $\bar{c}$ *and* $r(\bar{c}) \searrow 0$ *as* $\bar{c} \nearrow \infty$.

In a similar fashion, one can use Algorithm 1 to derive a multiplier method for $(P_{VI})$ where the additional regularizing term $c_k^{-1}(x - x_k)$ would be missing from the subproblems $(P_{VI}^k)$. Theorem 14 would then imply the convergence of this method, just as in the case of the proximal method of multipliers above, where Theorem 16 was used.

Since the publication of Robinson (1980), strong regularity has been studied by many authors. Dontchev and Rockafellar (1996) studied the important special case of variational inequalities over polyhedral convex sets, and they gave sufficient as well as *necessary* condition, called the *critical face* condition, for strong regularity. In the light of their results, strong regularity can be seen as a generalized constraint qualification that is a generic property of a variational inequality, like the classical constraint qualifications in nonlinear programming. In Dontchev and Rockafellar (1996), stability results for nonlinear programming were obtained by specializing the more general results on variational inequalities. In the same vein, we obtain convergence results for multiplier methods of nonlinear programming in the next section.

**7. Application to nonlinear programming.** Consider again the nonlinear programming problem

$$(P_{NLP}) \qquad \begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \begin{cases} = 0 & \text{for } i = 1, \ldots, r \\ \leq 0 & \text{for } i = r+1, \ldots, m \end{cases} \end{array}$$

of Example 11. Recall that the primal-dual inclusion in Example 11 is the Karush-Kuhn-Tucker system,

$$(PD_{NLP}) \qquad \begin{aligned} \nabla f_0(x) + \nabla F(x)^* y &= 0 \\ -F(x) + N_{K^*}(y) &\ni 0 \end{aligned}$$

for $(P_{NLP})$.

We will use Theorem 17 to derive a local convergence result for a nonconvex case of the following proximal method of multipliers, which was introduced in Rockafellar (1976b) for convex programming.

**Algorithm 4** (Proximal Method of Multipliers).
*Step* 0. Choose $(x_0, y_0)$ and set $k = 0$.
*Step* 1. Solve

$$(P_{NLP}^k) \qquad \text{minimize} \quad \phi_k(x) := f_0(x) + \frac{1}{2c_k}\|x - x_k\|^2 + \frac{1}{2c_k} d_K(y_k + c_k F(x))^2$$

for $x_{k+1}$.
*Step* 2. Set

$$\begin{aligned} y_{k+1}^i &= y_k^i + c_k f_i(x_{k+1}) && \text{for } i = 1, \ldots, r, \\ y_{k+1}^i &= \max\{y_k^i + c_k f_i(x_{k+1}), 0\} && \text{for } i = r+1, \ldots, m, \end{aligned}$$

$k = k+1$ and go to 1.

From implementation point of view, this differs from the classical method of multipliers only in the additional regularizing term $\frac{1}{2c_k}\|x - x_k\|^2$ in $\phi_k$. As in the convex case in Rockafellar (1976b), it turns out that this term allows us to deduce better convergence properties than for the ordinary method of multipliers.

Let $(\bar{x}, \bar{y})$ be a point satisfying the KKT conditions for $(P_{NLP})$, and define the index sets

$$\begin{aligned} I_1 &= \{1, \ldots, r\} \cup \{i \in [r+1, m] \mid f_i(\bar{x}) = 0,\ \bar{y}_i > 0\}, \\ I_2 &= \{i \in [r+1, m] \mid f_i(\bar{x}) = 0,\ \bar{y}_i = 0\}, \\ I_3 &= \{i \in [r+1, m] \mid f_i(\bar{x}) < 0,\ \bar{y}_i = 0\}. \end{aligned}$$

Recall that $(\bar{x}, \bar{y})$ is said to satisfy the *strong second-order sufficient condition* (SSOSC) for $(P_{NLP})$ if

$$\langle w, \nabla_{xx}^2 l(\bar{x}, \bar{y}) w \rangle > 0 \quad \forall\, w \neq 0 : \langle \nabla f_i(\bar{x}), w \rangle = 0 \quad \forall\, i \in I_1.$$

Here the function $l$ is defined by $l(x,y)=f_0(x)+\langle y, F(x)\rangle$. Note that SSOSC holds if and only if there is a $c\in\mathbb{R}$ such that

$$\langle w, \nabla^2_{xx} l(\bar{x},\bar{y})w\rangle + c\sum_{i\in I_1}\langle \nabla f_i(\bar{x}), w_i\rangle^2 > 0 \quad \forall\, w\neq 0. \tag{18}$$

We will make use of the following identity.

LEMMA 18. *For any nonempty closed convex cone $K$*

$$\frac{1}{2}d_K(u)^2 = \sup_{y\in K^*}\left\{\langle u,y\rangle - \frac{1}{2}\|y\|^2\right\}.$$

PROOF. Computing the conjugate of the function $g(u)=\frac{1}{2}d_K(u)^2$

$$\begin{aligned} g^*(y) &= \sup_{u,v}\left\{\langle u,y\rangle - \frac{1}{2}\|v-u\|^2 - \delta_K(v)\right\}\\ &= \sup_{v}\left\{\langle v+y,y\rangle - \frac{1}{2}\|y\|^2 - \delta_K(v)\right\}\\ &= \sup_{v}\{\langle v,y\rangle - \delta_K(v)\} + \frac{1}{2}\|y\|^2 = \delta_{K^*}(y) + \frac{1}{2}\|y\|^2, \end{aligned}$$

we have by the continuity of $g$ that

$$\frac{1}{2}d_K(u)^2 = g^{**}(u) = \sup_{y\in K^*}\left\{\langle u,y\rangle - \frac{1}{2}\|y\|^2\right\};$$

see Rockafellar and Wets (1998, 11.1). □

THEOREM 19. *Let $(\bar{x},\bar{y})$ be a KKT pair for problem $(P_{NLP})$ satisfying the SSOSC and assume that the gradients $\nabla f_i(\bar{x})$ for $i\in I_1$ are linearly independent. If the $\{c_k\}$ are large enough with $c_k\nearrow\bar{c}\le\infty$ and if $\|(x_0,y_0)-(\bar{x},\bar{y})\|$ is small enough, then there exists a sequence $\{(x_k,y_k)\}$ conforming to Algorithm* 4 *along with open neighborhoods $X_k$ such that for each $k$, $x_{k+1}$ is the unique solution in $X_k$ to $(P^k_{NLP})$. Then also, the sequence $\{(x_k,y_k)\}$ converges linearly and Fejér monotonically to $(\bar{x},\bar{y})$ with rate $r(\bar{c})<1$ that is decreasing in $\bar{c}$ and $r(\bar{c})\searrow 0$ as $\bar{c}\nearrow\infty$.*

PROOF. By Robinson (1980, Theorem 4.1), the SSOSC and the linear independence condition imply that the KKT system $(PD_{NLP})$ is strongly regular at $(\bar{x},\bar{y})$. Thus, by Theorem 17, it suffices to show that Algorithm 4 is equivalent to Algorithm 3.

Applying Rockafellar and Wets (1998, 12.23) to the indicator function $\delta_K$ of $K$, we get

$$\nabla d_K(\cdot)^2 = 2(I+N_K^{-1})^{-1} = 2(I+N_{K^*})^{-1} = 2P_{K^*},$$

where the last equality follows from Rockafellar and Wets (1998, 6.17). Thus,

$$\nabla\phi_k(x) = \nabla f_0(x) + c_k^{-1}(x-x_k) + \nabla F(x)^* P_{K^*}(y_k + c_k F(x)).$$

Since $K^*=\mathbb{R}^r\times\mathbb{R}^{m-r}_+$, Step 2 can be written as $y_{k+1}=P_{K^*}(y_k+c_kF(x_{k+1}))$. It remains to show that for large enough $c_k$, the unique stationary point of $\phi_k$ in $X_k$ is in fact a minimizer of $\phi_k$. To this end, we apply the second-order sufficient condition in Rockafellar and Wets (1998, 13.26), which applies to composite-type minimization problems like the one we have in Step 1.

By Lemma 18,

$$\frac{1}{2c_k} d_K(y_k + c_k u)^2 = \sup_{y \in Y} \left\{ \langle y, u + c_k^{-1} y_k \rangle - \frac{1}{2} \langle y, By \rangle \right\},$$

where $Y = K^*$ and $B = c_k^{-1} I$. Thus, problem $(P_{NLP}^k)$ can be written in the format of Rockafellar and Wets (1998, 13.26) as

$$\text{minimize} \quad \tilde{f}_0(x) + \theta_{Y,B}\big(\widetilde{F}(x)\big) \quad \text{over all } x \in X,$$

where $X = \mathbb{R}^n$, $\tilde{f}_0(x) = f_0(x) + \frac{1}{2c_k}\|x - x_k\|^2$, and $\widetilde{F}(x) := F(x) + c_k^{-1} y_k$. Part (b) of Rockafellar and Wets (1998, 13.26) then says that a sufficient condition for $x_{k+1}$ to be a local minimizer of $\tilde{f}_0 + \theta_{Y,B} \circ \widetilde{F}$ is that

$$(19) \qquad \max_{y \in Y(x_{k+1})} \big\langle w, \nabla^2_{xx} \widetilde{L}(x_{k+1}, y) w \big\rangle + 2\theta_{Y'(x_{k+1}),B}\big(\nabla \tilde{F}(x_{k+1}) w\big) > 0 \quad \forall\, w \in X'(x_{k+1}) \backslash \{0\},$$

where

$$\begin{aligned}
\widetilde{L}(x, y) &= \tilde{f}_0(x) + \big\langle y, \widetilde{F}(x) \big\rangle - \frac{1}{2}\langle y, By \rangle, \\
Y(x_{k+1}) &= \big\{ y \mid -\nabla_x \widetilde{L}(x_{k+1}, y) \in N_X(x_{k+1}), \quad \nabla_y \widetilde{L}(x_{k+1}, y) \in N_Y(y) \big\}, \\
X'(x_{k+1}) &= T_X(x_{k+1}) \cap \nabla_x \widetilde{L}(x_{k+1}, y)^{\perp}, \\
Y'(y) &= T_Y(y) \cap \nabla_y \widetilde{L}(x_{k+1}, y)^{\perp}, \\
\theta_{Y'(x_{k+1}),B}(u) &= \sup_{y \in Y'(x_{k+1})} \left\{ \langle y, u \rangle - \frac{1}{2}\langle y, By \rangle \right\},
\end{aligned}$$

and $T_C$ denotes the tangent cone of a set $C$; see Rockafellar and Wets (1998, §6A). We see that $Y(x_{k+1})$ is exactly the set of $y$s satisfying the system

$$(20) \qquad \begin{aligned}
\nabla f_0(x) + \nabla F(x)^* y + c_k^{-1}(x - x_k) &= 0 \\
-F(x) + N_{K^*}(y) + c_k^{-1}(y - y_k) &\ni 0
\end{aligned}$$

together with $x_{k+1}$. Since this is the proximal point subproblem $(PD_k)$ in the current situation, we know from Lemma 15 that its solutions are of the form $\{(x, T_k(F(x))) \mid \nabla \phi_k(x) = 0\}$, so $Y(x_{k+1})$ is the singleton set $\{y_{k+1}\}$. Since $X = \mathbb{R}^n$, we have $X'(x_{k+1}) = \mathbb{R}^n$, so the sufficient condition (19) can be written as

$$(21) \qquad \langle w, (\nabla^2_{xx} l(x_{k+1}, y_{k+1}) + c_k^{-1} I) w \rangle + 2\theta_{Y'(x_{k+1}),B}(\nabla F(x_{k+1}) w)^2 > 0 \quad \forall\, w \neq 0,$$

where $Y'(x_{k+1}) = T_{K^*}(y_{k+1}) \cap [F(x_{k+1}) - c_k^{-1}(y_{k+1} - y_k)]^{\perp}$.

Define $\widetilde{Y} = \{y \in \mathbb{R}^m \mid y_i = 0\ \forall\, i \notin I_1\}$. By choosing $\|(x_0, y_0) - (\bar{x}, \bar{y})\|$ small enough, we can, by the Fejér monotonicity of the sequence $\{(x_k, y_k)\}$, guarantee that $y_{k+1} - \bar{y}$ is so small that $y^i_{k+1} > 0$ for all $i \in I_1 \backslash \{1, \ldots, r\}$, and then $\widetilde{Y} \subset T_{K^*}(y_{k+1})$. Since $(x_{k+1}, y_{k+1})$ satisfies (20), we see that the positivity of $y^i_{k+1}$ for $i \in I_1 \backslash \{1, \ldots, r\}$ also implies that $f_i(x_{k+1}) - c_k^{-1}(y^i_{k+1} - y^i_k) = 0$. Thus, $\widetilde{Y} \subset Y'(x_{k+1})$, and then we get from Lemma 18 that for every $u$,

$$\begin{aligned}
\theta_{Y'(x_{k+1}),B}(u) &\geq \sup_{y \in \widetilde{Y}} \left\{ \langle y, u \rangle - \frac{1}{2}\langle y, By \rangle \right\} = \sup_{y \in \widetilde{Y}} \left\{ \langle y, u \rangle - \frac{1}{2c_k}\|y\|^2 \right\} \\
&= \frac{c_k}{2} d_{\widetilde{Y}^*}(u)^2 = \frac{c_k}{2} \sum_{i \in I_1} |u_i|^2.
\end{aligned}$$

The sufficient condition (21) is thus guaranteed by the condition

$$\langle w, \nabla^2_{xx} l(x_{k+1}, y_{k+1})w\rangle + c_k \sum_{i\in I_1} \langle \nabla f_i(x_{k+1}), w_i\rangle^2 > 0 \quad \forall\, w \neq 0,$$

which holds by the continuity of $\nabla^2_{xx} l$ and $\nabla f_i$, and by (18), provided $c_k$ is large enough. □

An analogous derivation based on Theorem 14 would give a convergence result for the classical multiplier method for nonlinear programming. Such results have been obtained by completely different methods in Bertsekas (1982), Ito and Kunisch (1990), Contesse-Becker (1993), and Conn et al. (1996).

Let us stress that the problems treated in §§5–7 are not the most general ones that can be handled with the present approach. We have chosen them as illustrative examples on how to use the general results of §4 in combination with an abstract duality framework. In these examples, we have relied on the Lipschitz condition of Corollary 10 instead of the much more general condition of local hypomonotonicity in Theorem 9. This was done partially because of the role of the Lipschitz condition as a generalized constraint qualification in variational analysis. In light of Proposition 7, the hypomonotonicity condition could serve as a still more general constraint qualification. As already observed, this condition has indeed been related to regularity properties of functions; see Rockafellar and Wets (1998, §13F).

**Acknowledgments.** Most of this work was done in June–July 2000 during my visit at the Instituto de Matemática Pura e Aplicada in Rio de Janeiro. I would like to express my deepest gratitude to Alfredo Iusem for his invitation and hospitality. I gratefully acknowledge the financial support from CNPq and the Finnish Cultural Foundation. I am grateful to one of the referees for pointing out several errors and making suggestions that resulted in significant improvements in this paper.

## References

Attouch, H., M. Théra. 1996. A general duality principle for the sum of two operators. *J. Convex Anal.* **3** 1–44.

Auslender, A., M. Teboulle. 2000. Lagrangian duality and related multiplier methods for variational inequality problems. *SIAM J. Optim.* **10** 1097–1115.

Bertsekas, D. P. 1982. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press Inc., New York–London.

Brézis, H. 1973. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Mathematics Studies, No. 5. North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York.

Conn, A. R., N. Gould, A. Sartenaer, Ph. L. Toint. 1996. Convergence results of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. *SIAM J. Optim.* **6** 674–703.

Contesse-Becker, L. 1993. Extended convergence results for the method of multipliers for non-strictly binding inequality constraints. *J. Optim. Theory Appl.* **79** 273–310.

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

———, ———. 1998. Characterizations of Lipschitzian stability in nonlinear programming. Mathematical programming with data perturbations, Lecture Notes in Pure and Applied Math., 195, Dekker, New York, 65–82.

———, ———. 2001. Ample parameterization of variational inclusions. *SIAM J. Optim.* **12** 170–187.

Eckstein, J. 1993. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Math. Oper. Res.* **18** 203–226.

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

———, D. Bertsekas. 1992. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming Ser. A* **55**(3) 293–318.

———, M. C. Ferris. 1999. Smooth methods of multipliers for complementarity problems. *Math. Programming* **86** 65–90.

Gabay, D. 1983. Applications of the method of multipliers to variational inequalities. M. Fortin, R. Glowinski, eds. *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*. North-Holland, Amsterdam, The Netherlands, 299–331.

Ito, K., K. Kunisch. 1990. The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces. *Math. Programming* **46** 341–360.

Iusem, A., B. Svaiter, M. Teboulle. 1994. Entropy-like proximal methods in convex programming. *Math. Oper. Res.* **19**(4) 790–814.

Kaplan, A., R. Tichatschke. 1998. Proximal point methods and nonconvex optimization. *J. Global Optim.* **13**(4) 389–406.

Klatte, D., B. Kummer. 1999. Strong stability in nonlinear programming revisited. *J. Australian Math. Soc. Ser. B* **40** 336–352.

———, K. Tammer. 1990. Strong stability staility of stationary solutions and Karush-Kuhn-Tucker points in nonlinear optimization. *Ann. Oper. Res.* **27** 285–307.

Lawrence, J., J. E. Spingarn. 1987. On fixed points of nonexpansive piecewise isometric mappings. *Proc. London Math. Soc.* **55**(3) 605–624.

Levy, A. B., R. A. Poliquin, R. T. Rockafellar. 2000. Stability of locally optimal solutions. *SIAM J. Optim.* **10** 580–604.

Luque, F. J. 1984. Asymptotic convergence analysis of the proximal point algorithm. *SIAM J. Control Optim.* **22**(2) 277–293.

Martinet, B. 1970. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Francaise Inform. Rech. Oper. Ser. R*-3 **4** 154–158.

Pennanen, T. 1999. Dualization of monotone generalized equations. Ph.D. thesis, Department of Mathematics, University of Washington, Seattle, WA.

———. 2000. Dualization of generalized equations of maximal monotone type. *SIAM J. Optim.* **10** 809–835.

Robinson, S. M. 1980. Strong regularity of generalized equations. *Math. Oper. Res.* **5** 43–62.

———. 1994. Newton's method for a class of nonsmooth functions. *Set-Valued Anal.* **2** 291–305.

———. 1999a. Composition duality and maximal monotonicity. *Math. Programming* **85** 1–13.

———. 1999b. Linear convergence of epsilon-subgradient descent methods for a class of convex functions. *Math. Programming Ser. A* **86**(1) 41–50.

Rockafellar, R. T. 1970a. *Convex Analysis*. Princeton University Press, Princeton, NJ.

———. 1970b. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.* **33** 209–216.

———. 1973. Penalty methods and augmented Lagrangians in nonlinear programming. *5th Conf. Optim. Tech.,* Lecture Notes in Computer Science 3, Springer, Berlin, Germany, 418–425.

———. 1974. Conjugate Duality and Optimization. *SIAM*, Philadelphia, PA.

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

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

———. 1978. Monotone operators and augmented Lagrangian methods in nonlinear programming. *Nonlinear programming* **3** *Proc. Sympos., Special Interest Group Math. Programming.* University of Wisconsin-Madison, WI; Academic Press, New York-London, 1–25.

———. 1980. Lagrange multipliers and variational inequalities. R. W. Cottle, F. Giannessi, J-L. Lions, eds. *Variational Inequalities and Complementarity Problems*. John Wiley, New York, 303–322.

———, R. J-B. Wets. 1998. *Variational Analysis*. Springer-Verlag, Berlin, Germany.

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

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

Tossings, P. 1994. The perturbed proximal point algorithm and some of its applications. *Appl. Math. Optim.* **29** 125–159.

Zeidler, E. 1990. *Nonlinear Functional Analysis and Its Applications II*. Springer-Verlag, Berlin, Germany.

T. Pennanen: Department of Management Science, Helsinki School of Economics, PL 1210, 00101 Helsinki, Finland; e-mail: pennanen@hkkk.fi

Copyright 2002, by INFORMS, all rights reserved. Copyright of Mathematics of Operations Research is the property of INFORMS: Institute for Operations Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.