SIAM J. CONTROL OPTIM.
Vol. 46, No. 5, pp. 1683–1704

ASYMPTOTIC CONVERGENCE ANALYSIS OF A NEW CLASS OF PROXIMAL POINT METHODS*

WILLIAM W. HAGER† AND HONGCHAO ZHANG‡

Abstract. Finite dimensional local convergence results for self-adaptive proximal point methods and nonlinear functions with multiple minimizers are generalized and extended to a Hilbert space setting. The principle assumption is a local error bound condition which relates the growth in the function to the distance to the set of minimizers. A local convergence result is established for almost exact iterates. Less restrictive acceptance criteria for the proximal iterates are also analyzed. These criteria are expressed in terms of a subdifferential of the proximal function and either a subdifferential of the original function or an iteration difference. If the proximal regularization parameter $\mu(\mathbf{x})$ is sufficiently small and bounded away from zero and f is sufficiently smooth, then there is local linear convergence to the set of minimizers. For a locally convex function, a convergence result similar to that for almost exact iterates is established. For a locally convex solution set and smooth functions, it is shown that if the proximal regularization parameter has the form $\mu(\mathbf{x}) = \beta\|f'[\mathbf{x}]\|^{\eta}$, where $\eta \in (0, 2)$, then the convergence is at least superlinear if $\eta \in (0, 1)$ and at least quadratic if $\eta \in [1, 2)$.

Key words. proximal point, degenerate optimization, multiple minima, self-adaptive method

AMS subject classifications. 90C06, 90C26, 65Y20

DOI. 10.1137/060666627

1. Introduction.

In this paper, we consider an optimization problem:

$$\min\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{H}\}, \tag{1.1}$$

where $\mathcal{H}$ is a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and $f : \mathcal{H} \mapsto \mathcal{R}$. It is assumed that the set of minimizers for (1.1), denoted $\mathbf{X}$, is nonempty and closed. We establish new convergence rate results for proximal point methods for solving (1.1).

Literature connected with the analysis and development of proximal point methods includes [1, 2, 3, 4, 6, 8, 9, 10, 11, 12, 15, 16, 18, 19, 20, 21, 24, 25, 26, 27]. In the proximal point method, iterates $\mathbf{x}_k$, $k \ge 1$, are generated by the following rule:

$$\mathbf{x}_{k+1} \in \arg\ \min\ \{F_k(\mathbf{x}) : \mathbf{x} \in \mathcal{H}\}, \tag{1.2}$$

where

$$F_k(\mathbf{x}) = f(\mathbf{x}) + \frac{1}{2}\mu_k\|\mathbf{x} - \mathbf{x}_k\|^2.$$

Here $\mathbf{x}_0 \in \mathcal{H}$ is an initial guess for a minimizer, the parameters μ_k, $k \ge 0$, are positive scalars, and $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$ is the usual Hilbert space norm. When f is twice continuously differentiable, the eigenvalues of the second derivative operator F_k'' are bounded from below by μ_k at a local minimizer; consequently, the regularization term $\mu_k\|\mathbf{x} - \mathbf{x}_k\|^2$ improves the conditioning of (1.1).

*Received by the editors August 1, 2006; accepted for publication (in revised form) May 9, 2007; published electronically November 2, 2007. This material is based upon work supported by the National Science Foundation under grants 0203270, 0619080, and 0620286.
http://www.siam.org/journals/sicon/46-5/66662.html
†Department of Mathematics, University of Florida, P.O. Box 118105, Gainesville, FL 32611-8105 (hager@math.ufl.edu, http://www.math.ufl.edu/~hager).
‡Institute for Mathematics and Its Applications (IMA), University of Minnesota, 400 Lind Hall, 207 Church Street S.E., Minneapolis, MN 55455-0436 (hozhang@ima.umn.edu).

In [21] Rockafellar shows that if f is convex, then the proximal point method converges linearly when μ_k is bounded away from zero and superlinearly when μ_k tends to zero. For these convergence results, $\mathbf{X}$ is a singleton. Luque [14] studied the case when $\mathbf{X}$ may contain more than one element. By assuming some growth properties for the (multivalued) inverse of the derivative, results analogous to those of Rockafellar were obtained. Kaplan and Tichatschke [11] consider the case where f is convex, μ_k is constant, and $\mathbf{X}$ may contain more than one element. A linear convergence result for the iterates is established under a growth condition for the function which is similar to the growth condition used in our paper (see Assumption 14.4 and Theorem 14.5 in [11]).

In another research direction, Combettes and Pennanen [2], Iusem, Pennanen, and Svaiter [10], and Pennanen [19] replace the monotonicity assumptions appearing in earlier work by a weaker hypomonotonicity condition for the inverse of the derivative, that is, the inverse of the derivative is monotone when a multiple of the identity is added. Additional assumptions, however, enter into the analysis which imply the solution set $\mathbf{X}$ is a singleton.

In [7] we present a new class of self-adaptive proximal point methods for finite dimensional optimization problems. Our analysis employs the following local error bound condition at $\hat{\mathbf{x}} \in \mathbf{X}$: There exist positive constants α and ρ such that

$$f(\mathbf{x}) - f^* \geq \alpha D(\mathbf{x}, \mathbf{X})^2 \quad \text{whenever } \|\mathbf{x} - \hat{\mathbf{x}}\| \leq \rho, \tag{1.3}$$

where f^* is the minimum value in (1.1) and

$$D(\mathbf{x}, \mathbf{X}) = \inf_{\mathbf{y} \in \mathbf{X}} \|\mathbf{x} - \mathbf{y}\|.$$

In other words, $D(\mathbf{x}, \mathbf{X})$ measures the distance to the solution set $\mathbf{X}$. If (1.3) is satisfied, then we say that f provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$. For an exact proximal iterate $\mathbf{x}_{k+1}$ satisfying (1.2), we show in [7] that for any starting guess $\mathbf{x}_0$ in a neighborhood of the solution set, the iterates converge to a solution $\mathbf{x}^*$ of (1.1) and the following estimate holds:

$$D(\mathbf{x}_{k+1}, \mathbf{X}) \leq C\mu_k D(\mathbf{x}_k, \mathbf{X}), \tag{1.4}$$

where $C = 2/(2\alpha - \mu_k)$.

In a Hilbert space setting, the exact proximal iterate (1.2) may not exist. In this paper, we establish a similar convergence result using the following acceptance criterion: $\mathbf{x}_{k+1}$ is acceptable when

$$\begin{aligned} F_k(\mathbf{x}_{k+1}) &\leq \inf_{\mathbf{x} \in \mathbf{X}} \left\{ F_k(\mathbf{x}) + \frac{\mu_k^2}{2} \|\mathbf{x} - \mathbf{x}_k\|^2 \right\} \\ &= \inf_{\mathbf{x} \in \mathbf{X}} \left\{ f(\mathbf{x}) + \left(\frac{\mu_k + \mu_k^2}{2} \right) \|\mathbf{x} - \mathbf{x}_k\|^2 \right\}. \end{aligned} \tag{C0}$$

In section 3 we show that there always exists an iterate satisfying (C0), and a convergence result of the form (1.4) holds.

Although (C0) leads to an elegant convergence theory, which can be applied to any function whose set of minimizers is nonempty and closed, the acceptance criterion is not easily implemented since it is expressed in terms of the solution set (which we are trying to compute). Consequently, we now introduce implementable acceptance criteria which are expressed in terms of the (basic) subdifferential of f (see [17, p. 82] and [22]) denoted $\partial f(\mathbf{x})$. If f is Fréchet differentiable, then $\partial f(\mathbf{x}) = f'[\mathbf{x}]$. If f is convex, then $\partial f(\mathbf{x})$ is the usual subdifferential of convex analysis. The acceptance criteria for (1.2) are

(C1) $F_k(\mathbf{x}_{k+1}) \le f(\mathbf{x}_k)$ and $\|\partial F_k(\mathbf{x}_{k+1})\|_{\inf} \le \mu_k \|\partial f(\mathbf{x}_k)\|_{\inf}$, and

(C2) $F_k(\mathbf{x}_{k+1}) \le f(\mathbf{x}_k)$ and $\|\partial F_k(\mathbf{x}_{k+1})\|_{\inf} \le \theta\mu_k \|\mathbf{x}_{k+1} - \mathbf{x}_k\|$.

Here θ is a positive constant smaller than $1/\sqrt{2}$, and $\|\cdot\|_{\inf}$ denotes the distance to the origin; that is, for any set $\mathcal{S} \subset \mathcal{H}$,

$$\|\mathcal{S}\|_{\inf} = \inf_{s \in \mathcal{S}} \|\mathbf{s}\|.$$

If $\mathcal{S} = \emptyset$, then we set $\|\mathcal{S}\|_{\inf} = \infty$.

In [14] and [21], the authors considered the acceptance condition

$$\|\partial F_k(\mathbf{x}_{k+1})\| \le \epsilon_k \mu_k \|\mathbf{x}_{k+1} - \mathbf{x}_k\|,$$

where $\sum_k \epsilon_k < \infty$. Our criterion (C2) corresponds to the case $\sum_k \epsilon_k = \infty$. In [14] and [21], the authors consider convex functionals, while here we obtain local convergence rates for general nonlinear functionals. Note that our acceptance criteria employ a subdifferential rather than the derivative used in our earlier work.

Slightly different versions of the proximal point method for maximal monotone operators are developed by Solodov and Svaiter in the series of papers [24, 25, 26]. They develop both a hybrid proximal point algorithm where an approximate proximal step is followed by a projection and a hybrid extragradient version in which the original operator is replaced by an ϵ enlargement. In order to compare their analysis to the results in our paper, we focus on the special case where the operator is the subdifferential of a convex function f. In each iteration of the Solodov/Svaiter scheme, they first compute an approximate proximal iterate $\mathbf{y}_k$ satisfying a relaxed version of (C2); they then update the iterate along the negative gradient:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - s_k \mathbf{g}_k, \tag{1.5}$$

where $\mathbf{g}_k \in \partial f(\mathbf{y}_k)$, s_k is the scalar stepsize, and $\theta < 1$. In [24], $s_k = 1/\mu_k$ (the reciprocal of the proximal regularization parameter), while in [25], s_k is chosen so that $\mathbf{x}_{k+1}$ is the projection of $\mathbf{x}_k$ onto the half-space

$$\{\mathbf{x} \in \mathcal{H} : \langle \mathbf{g}_k, \mathbf{x} - \mathbf{y}_k \rangle \ge 0\}.$$

Since the Solodov/Svaiter update (1.5) amounts to an extragradient step, their convergence theory for fixed θ (see [26, Thm. 8]) yields linear convergence, even when μ_k tends to 0, unless the accuracy criterion (C2) for the approximate proximal iterate $\mathbf{y}_k$ is strengthened. Ways to improve the accuracy of $\mathbf{y}_k$ so as to obtain superlinear convergence with the Solodov/Svaiter hybrid schemes are the following: (a) Replace θ by θ_k in (C2) and let θ_k tend to 0 (see [26, Rem. 9]). (b) In the case $\mathcal{H} = \mathbb{R}^n$, compute $\mathbf{y}_k$ by a Newton iteration applied to the proximal problem (1.2), with μ_k on the order of $\|\nabla f(\mathbf{x}_k)\|^{1/2}$ (see [26, sect. 5.2]).

In this paper, we obtain superlinear convergence with (C2) by letting μ_k approach 0, and we analyze how the convergence speed depends on the decay rate of μ_k. We consider both a convex cost function analogous to the maximal monotone operator in [24, 25, 26] and the more general case where the solution set $\mathbf{X}$ is locally convex and f is sufficiently smooth. We allow multiple solutions satisfying the local error bound condition (1.3), while in [24, 25, 26] the solution set is unique since the inverse operator is required to be Lipschitz continuous at zero [26, eq. (28)].

We will show that for either (C1) or (C2), for μ_k sufficiently small, and for either f locally convex or $\mathbf{X}$ locally convex and f sufficiently smooth, an estimate of the form (1.4) holds. For μ_k sufficiently small and bounded away from zero, and for smooth functions, there is at least local linear convergence to the set of minimizers. For $\mathbf{X}$ locally convex and f sufficiently smooth, and for $\mu_k = \beta\|f'[\mathbf{x}_k]\|^\eta$, where $\eta \in (0, 2)$, the convergence is superlinear when $\eta \in (0, 1)$ and at least quadratic when $\eta \in [1, 2)$.

Our paper is organized as follows: In section 2 we establish the equivalence, when f is twice continuously differentiable, of our local error bound condition and a gradient-based local error bound condition used in [5, 13, 14, 28, 29, 30]. Note, though, that our local error bound condition can be applied even when f has no derivative. In section 3 we analyze proximal iterates which satisfy (C0). Section 4 studies the criteria (C1) and (C2).

1.1. Notation. Throughout this paper, we use the following notation. If $A : \mathcal{H} \mapsto \mathcal{H}$ is a bounded linear operator, then $\|A\|$ is the operator norm induced by the Hilbert space norm $\|\cdot\|$. The empty set is denoted $\emptyset$. The complement of a set $\mathcal{S} \subset \mathcal{H}$ is denoted $\mathcal{S}^c$. If $\mathbf{x}$ and $\mathbf{y} \in \mathcal{H}$, then $[\mathbf{x}, \mathbf{y}]$ is the line segment connecting $\mathbf{x}$ and $\mathbf{y}$. $\mathcal{B}_\rho(\mathbf{x})$ is the ball with center $\mathbf{x}$ and radius ρ. $f'[\mathbf{x}]$ and $f''[\mathbf{x}]$ are the first- and second-order Fréchet derivatives of f at $\mathbf{x}$ when they exist. The derivatives are operators defined on either $\mathcal{H}$ or $\mathcal{H} \times \mathcal{H}$. We also view $f'[\mathbf{x}]$ as an element of $\mathcal{H}$ and write $f'[\mathbf{x}](\mathbf{y}) = \langle f'[\mathbf{x}], \mathbf{y}\rangle$. Similarly, we view $f''[\mathbf{x}]$ as a bounded linear map from $\mathcal{H}$ to itself and write

$$f''[\mathbf{x}](\mathbf{y}, \mathbf{z}) = \langle f''[\mathbf{x}]\mathbf{y}, \mathbf{z}\rangle.$$

2. Local error bound based on derivative. In this paper, we utilize the local error bound condition (1.3) based on function value. Earlier work [5, 13, 14, 28, 29, 30] has exploited a local error bound condition based on the derivative. Namely, f' provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$ if there exist positive constants α and ρ such that

$$\|f'[\mathbf{x}]\| \ge \alpha D(\mathbf{x}, \mathbf{X}) \quad \text{whenever } \|\mathbf{x} - \hat{\mathbf{x}}\| \le \rho. \tag{2.1}$$

We now show that when f is smooth enough, these two conditions are equivalent.

LEMMA 2.1. *If f is twice continuously Fréchet differentiable in a neighborhood of $\hat{\mathbf{x}} \in \mathbf{X}$, then f provides a local error bound at $\hat{\mathbf{x}}$ in the sense of* (1.3) *if and only if f' provides a local error bound at $\hat{\mathbf{x}}$ in the sense of* (2.1).

Proof. Suppose f provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$ with positive scalars α and ρ satisfying (1.3). Choose ρ smaller, if necessary, so that f is twice continuously Fréchet differentiable in $\mathcal{B}_\rho(\hat{\mathbf{x}})$ and

$$\|f''[\mathbf{x}] - f''[\mathbf{y}]\| \le \alpha/3 \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{B}_\rho(\hat{\mathbf{x}}). \tag{2.2}$$

Define $r = \rho/2$. Given $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$, let $\bar{\mathbf{x}}$ be any element of $\mathbf{X} \cap \mathcal{B}_r(\mathbf{x})$. Since $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$, we have $\hat{\mathbf{x}} \in \mathbf{X} \cap \mathcal{B}_r(\mathbf{x})$, which shows that $\mathbf{X} \cap \mathcal{B}_r(\mathbf{x})$ is nonempty. The triangle inequality implies that

$$\|\bar{\mathbf{x}} - \hat{\mathbf{x}}\| \le \|\bar{\mathbf{x}} - \mathbf{x}\| + \|\mathbf{x} - \hat{\mathbf{x}}\| \le 2r = \rho. \tag{2.3}$$

Since both $\mathbf{x}$ and $\bar{\mathbf{x}} \in \mathcal{B}_\rho(\hat{\mathbf{x}})$, f is twice continuously Fréchet differentiable in $\mathcal{B}_\rho(\hat{\mathbf{x}})$, and $f'[\bar{\mathbf{x}}] = \mathbf{0}$, we have

$$f(\mathbf{x}) - f^* = f(\mathbf{x}) - f(\bar{\mathbf{x}}) = \frac{1}{2}\langle \mathbf{x} - \bar{\mathbf{x}}, f''[\bar{\mathbf{x}}](\mathbf{x} - \bar{\mathbf{x}})\rangle + R_2(\mathbf{x}, \bar{\mathbf{x}}), \tag{2.4}$$

where R_2 is the remainder term. The bound (2.2) gives

$$|R_2(\mathbf{x}, \bar{\mathbf{x}})| \le \frac{\alpha}{3}\|\mathbf{x} - \bar{\mathbf{x}}\|^2$$

whenever $\mathbf{x}$ and $\bar{\mathbf{x}} \in \mathcal{B}_\rho(\hat{\mathbf{x}})$. In this case, (2.4) and the local error bound condition (1.3) give

$$\alpha D(\mathbf{x}, \mathbf{X})^2 \le f(\mathbf{x}) - f^* \le \frac{1}{2}\|\mathbf{x} - \bar{\mathbf{x}}\|\|f''[\bar{\mathbf{x}}](\mathbf{x} - \bar{\mathbf{x}})\| + \frac{\alpha}{3}\|\mathbf{x} - \bar{\mathbf{x}}\|^2. \tag{2.5}$$

Again, since f is twice continuously Fréchet differentiable in $\mathcal{B}_\rho(\hat{\mathbf{x}})$ and $f'[\bar{\mathbf{x}}] = \mathbf{0}$, we have

$$f'[\mathbf{x}] = f'[\mathbf{x}] - f'[\bar{\mathbf{x}}] = f''[\bar{\mathbf{x}}](\mathbf{x} - \bar{\mathbf{x}}) + \mathbf{R}_1(\mathbf{x}, \bar{\mathbf{x}}), \tag{2.6}$$

where $\mathbf{R}_1$ is the remainder term. The bound (2.2) gives

$$\|R_1(\mathbf{x}, \bar{\mathbf{x}})\| \le \frac{\alpha}{3}\|\mathbf{x} - \bar{\mathbf{x}}\| \tag{2.7}$$

whenever $\mathbf{x}$ and $\bar{\mathbf{x}} \in \mathcal{B}_\rho(\hat{\mathbf{x}})$. Combining (2.5)–(2.7) yields

$$\alpha D(\mathbf{x}, \mathbf{X})^2 \le \frac{1}{2}\left(\|\mathbf{x} - \bar{\mathbf{x}}\|\|f'[\mathbf{x}]\| + \alpha\|\mathbf{x} - \bar{\mathbf{x}}\|^2\right). \tag{2.8}$$

Since $\mathbf{X} \cap \mathcal{B}_r(\mathbf{x})$ is nonempty, we have

$$D(\mathbf{x}, \mathbf{X}) = \inf_{\bar{\mathbf{x}} \in \mathbf{X}} \|\mathbf{x} - \bar{\mathbf{x}}\| = \inf\{\|\mathbf{x} - \bar{\mathbf{x}}\| : \bar{\mathbf{x}} \in \mathbf{X} \cap \mathcal{B}_r(\mathbf{x})\}.$$

Minimizing the right-hand side of (2.8) over $\bar{\mathbf{x}} \in \mathbf{X} \cap \mathcal{B}_r(\mathbf{x})$ gives

$$\alpha D(\mathbf{x}, \mathbf{X})^2 \le \frac{1}{2}\left(D(\mathbf{x}, \mathbf{X})\|f'[\mathbf{x}]\| + \alpha D(\mathbf{x}, \mathbf{X})^2\right).$$

Rearranging this yields

$$\|f'[\mathbf{x}]\| \ge \alpha D(\mathbf{x}, \mathbf{X}).$$

Hence, $\partial f = f'$ provides a local error bound at $\hat{\mathbf{x}}$ with constants α and r.

Conversely, suppose f' provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$ with positive scalars α and ρ satisfying (2.1). Let ρ be as in the first half of the proof. Choose ρ smaller, if necessary, so that

$$\|f''[\mathbf{x}] - f''[\mathbf{y}]\| \le \frac{7\alpha^2}{18(\lambda + 1)} \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{B}_\rho(\hat{\mathbf{x}}), \tag{2.9}$$

where

$$\lambda = \sup\{\|f''[\mathbf{x}]\| : \mathbf{x} \in \mathcal{B}_\rho(\hat{\mathbf{x}})\}. \tag{2.10}$$

Let $r = \rho/2$, let $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$, and let $\bar{\mathbf{x}} \in \mathbf{X} \cap \mathcal{B}_r(\hat{\mathbf{x}})$.

Since f achieves a minimum at $\bar{\mathbf{x}} \in \mathbf{X}$, $f''[\bar{\mathbf{x}}]$ is positive. Thus, there exists a unique, positive self-adjoint bounded linear operator B, the square root of $f''[\bar{\mathbf{x}}]$, satisfying $f''[\bar{\mathbf{x}}] = B^2$ [23, Thm. 13.31]. By (2.6), (2.7), and the local error bound condition (2.1), we have

$$\begin{aligned}\alpha D(\mathbf{x},\mathbf{X}) &\le \|f'[\mathbf{x}]\| \\ &\le \|f''[\bar{\mathbf{x}}](\mathbf{x}-\bar{\mathbf{x}})\| + \frac{\alpha}{3}\|\mathbf{x}-\bar{\mathbf{x}}\| \\ &= \|B^2(\mathbf{x}-\bar{\mathbf{x}})\| + \frac{\alpha}{3}\|\mathbf{x}-\bar{\mathbf{x}}\| \\ &\le \|B\|\|B(\mathbf{x}-\bar{\mathbf{x}})\| + \frac{\alpha}{3}\|\mathbf{x}-\bar{\mathbf{x}}\|.\end{aligned}$$

Squaring both sides yields

$$\begin{aligned}\alpha^2 D(\mathbf{x},\mathbf{X})^2 &\le 2\|B\|^2\|B(\mathbf{x}-\bar{\mathbf{x}})\|^2 + \frac{2\alpha^2}{9}\|\mathbf{x}-\bar{\mathbf{x}}\|^2 \\ &= 2\langle \mathbf{x}-\bar{\mathbf{x}}, f''[\bar{\mathbf{x}}](\mathbf{x}-\bar{\mathbf{x}})\rangle \|f''[\bar{\mathbf{x}}]\| + \frac{2\alpha^2}{9}\|\mathbf{x}-\bar{\mathbf{x}}\|^2 \\ &\le 2\langle \mathbf{x}-\bar{\mathbf{x}}, f''[\bar{\mathbf{x}}](\mathbf{x}-\bar{\mathbf{x}})\rangle \lambda + \frac{2\alpha^2}{9}\|\mathbf{x}-\bar{\mathbf{x}}\|^2,\end{aligned}$$

where λ is defined in (2.10). It follows that

$$\langle \mathbf{x}-\bar{\mathbf{x}}, f''[\bar{\mathbf{x}}](\mathbf{x}-\bar{\mathbf{x}})\rangle \ge \frac{\alpha^2[9D(\mathbf{x},\mathbf{X})^2 - 2\|\mathbf{x}-\bar{\mathbf{x}}\|^2]}{18(\lambda+1)}.$$

(1 is added to the denominator to allow for the possibility that $\lambda = 0$.) Using this in (2.4) yields

$$\begin{aligned}f(\mathbf{x}) - f^* &= \frac{1}{2}\langle(\mathbf{x}-\bar{\mathbf{x}}), f''[\bar{\mathbf{x}}](\mathbf{x}-\bar{\mathbf{x}})\rangle + R_2(\mathbf{x},\bar{\mathbf{x}}) \\ &\ge \frac{\alpha^2[9D(\mathbf{x},\mathbf{X})^2 - 2\|\mathbf{x}-\bar{\mathbf{x}}\|^2]}{18(\lambda+1)} + R_2(\mathbf{x},\bar{\mathbf{x}}).\end{aligned} \tag{2.11}$$

By the choice of ρ in (2.9), we have

$$|R_2(\mathbf{x},\bar{\mathbf{x}})| \le \frac{\beta}{2}\|\mathbf{x}-\bar{\mathbf{x}}\|^2, \quad \beta = \frac{7\alpha^2}{18(\lambda+1)},$$

whenever $\mathbf{x}$ and $\bar{\mathbf{x}} \in \mathcal{B}_\rho(\hat{\mathbf{x}})$. By (2.11),

$$f(\mathbf{x}) - f^* \ge \frac{\alpha^2[9D(\mathbf{x},\mathbf{X})^2 - 2\|\mathbf{x}-\bar{\mathbf{x}}\|^2]}{18(\lambda+1)} - \frac{\beta}{2}\|\mathbf{x}-\bar{\mathbf{x}}\|^2.$$

Minimizing $\|\mathbf{x}-\bar{\mathbf{x}}\|$ over $\bar{\mathbf{x}} \in \mathbf{X} \cap \mathcal{B}_r(\mathbf{x})$ gives

$$f(\mathbf{x}) - f^* \ge \left(\frac{\beta}{2}\right) D(\mathbf{x},\mathbf{X})^2,$$

which completes the proof. ∎

3. Convergence analysis for almost exact minimization. We first show that (C0) can always be satisfied.

LEMMA 3.1. *If $\mu_k > 0$, then there exists $\mathbf{x}_{k+1} \in \mathcal{H}$ satisfying* (C0)*; moreover, for any $\mathbf{x}_{k+1}$ satisfying* (C0)*, we have*

$$\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \le \sqrt{1+\mu_k}\, D(\mathbf{x}_k, \mathbf{X}). \tag{3.1}$$

Proof. If $\mathbf{x}_k \in \mathbf{X}$, then the lemma holds trivially since $\mathbf{x}_{k+1} = \mathbf{x}_k$. Hence, assume that $D(\mathbf{x}_k, \mathbf{X}) > 0$. Since $\mu_k > 0$, we have

$$\begin{aligned}\inf_{\mathbf{x}\in\mathbf{X}} \left\{ F_k(\mathbf{x}) + \frac{\mu_k^2}{2}\|\mathbf{x}-\mathbf{x}_k\|^2 \right\} &= f^* + \left(\frac{\mu_k+\mu_k^2}{2}\right) \inf_{\mathbf{x}\in\mathbf{X}}\{\|\mathbf{x}-\mathbf{x}_k\|\} \\ &= f^* + \left(\frac{\mu_k+\mu_k^2}{2}\right) D(\mathbf{x}_k, \mathbf{X}) \\ &> f^* + \frac{\mu_k}{2} D(\mathbf{x}_k, \mathbf{X}) \\ &= \inf_{\mathbf{x}\in\mathbf{X}} F_k(\mathbf{x}) \ge \inf_{\mathbf{x}\in\mathcal{H}} F_k(\mathbf{x}).\end{aligned}$$

Since one of these inequalities is strict, there exists $\mathbf{x}_{k+1} \in \mathcal{H}$ satisfying (C0). Moreover, for all $\mathbf{x} \in \mathbf{X}$, (C0) yields

$$\begin{aligned}F_k(\mathbf{x}_{k+1}) = f(\mathbf{x}_{k+1}) + \frac{\mu_k}{2}\|\mathbf{x}_{k+1}-\mathbf{x}_k\|^2 &\le F_k(\mathbf{x}) + \frac{\mu_k^2}{2}\|\mathbf{x}-\mathbf{x}_k\|^2 \\ &= f^* + \frac{\mu_k+\mu_k^2}{2}\|\mathbf{x}-\mathbf{x}_k\|^2.\end{aligned}$$

Since $f^* \le f(\mathbf{x}_{k+1})$, we conclude that

$$\|\mathbf{x}_{k+1}-\mathbf{x}_k\| \le \sqrt{1+\mu_k}\|\mathbf{x}-\mathbf{x}_k\| \tag{3.2}$$

for all $\mathbf{x} \in \mathbf{X}$. Taking the infimum over $\mathbf{x} \in \mathbf{X}$ gives (3.1). ◻

Iterates which satisfy the criterion (C0) are now analyzed.

THEOREM 3.2. *Assume the following conditions are satisfied:*

(E0) *f provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$ with positive scalars α and ρ satisfying* (1.3).

(E1) *$\beta > 0$ is small enough that the following inequalities hold:*

$$\frac{\beta+2\beta^2}{2} \le \frac{\alpha}{3} \quad \textit{and} \quad \gamma := \frac{\beta\sqrt{3(1+\beta)(3+4\alpha)}}{2\alpha} < 1.$$

(E2) $\mu_k \in (0, \beta]$.

(E3) *$\mathbf{x}_0$ is close enough to $\hat{\mathbf{x}}$ that*

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}\|\left(1 + \frac{\sqrt{1+\beta}}{1-\gamma}\right) \le \rho.$$

Then any proximal iterates $\{\mathbf{x}_k\}$ satisfying (C0) *have the property that $\mathbf{x}_k \in \mathcal{B}_\rho(\hat{\mathbf{x}})$ for each k, and they approach a minimizer $\mathbf{x}^* \in \mathbf{X}$; moreover, for each k, we have*

$$\|\mathbf{x}_k - \mathbf{x}^*\| \le c_1\gamma^k D(\mathbf{x}_0, \mathbf{X}) \quad \textit{and} \quad D(\mathbf{x}_{k+1}, \mathbf{X}) \le c_2\mu_k D(\mathbf{x}_k, \mathbf{X}), \tag{3.3}$$

where

$$c_1 = \frac{\sqrt{1+\beta}}{1-\gamma} \quad \textit{and} \quad c_2 = \gamma/\beta. \tag{3.4}$$

Proof. For $j = 0$, (E3) implies that

$$\|\mathbf{x}_j - \hat{\mathbf{x}}\| \le \rho \quad \text{and} \quad D(\mathbf{x}_j, \mathbf{X}) \le \gamma^j D(\mathbf{x}_0, \mathbf{X}). \tag{3.5}$$

Proceeding by induction, suppose that (3.5) holds for all $j \in [0, k]$ and for some $k \geq 0$. We show that (3.5) also holds for $j = k + 1$. By the triangle inequality, Lemma 3.1, (E2), and the induction hypothesis, it follows that

$$\begin{aligned}\|\mathbf{x}_{k+1} - \mathbf{x}_0\| &\leq \sum_{j=0}^{k} \|\mathbf{x}_{j+1} - \mathbf{x}_j\| \leq \sum_{j=0}^{k} \sqrt{1+\mu_j} D(\mathbf{x}_j, \mathbf{X}) \\ &\leq \sqrt{1+\beta} \sum_{j=0}^{k} \gamma^j D(\mathbf{x}_0, \mathbf{X}) \leq \frac{\sqrt{1+\beta}}{1-\gamma} D(\mathbf{x}_0, \mathbf{X}) \leq \frac{\sqrt{1+\beta}}{1-\gamma} \|\mathbf{x}_0 - \hat{\mathbf{x}}\|.\end{aligned}$$

Again, by the triangle inequality and (E3),

$$\|\mathbf{x}_{k+1} - \hat{\mathbf{x}}\| \leq \|\mathbf{x}_{k+1} - \mathbf{x}_0\| + \|\mathbf{x}_0 - \hat{\mathbf{x}}\| \leq \left(1 + \frac{\sqrt{1+\beta}}{1-\gamma}\right) \|\mathbf{x}_0 - \hat{\mathbf{x}}\| \leq \rho. \tag{3.6}$$

For any $\mathbf{x} \in \mathcal{H}$, observe that

$$\begin{aligned}&\|\mathbf{x} - \mathbf{x}_k\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \\ &\quad = \langle \mathbf{x} + \mathbf{x}_{k+1} - 2\mathbf{x}_k, \mathbf{x} - \mathbf{x}_{k+1} \rangle \\ &\quad \leq (\|\mathbf{x} - \mathbf{x}_{k+1}\| + 2\|\mathbf{x}_{k+1} - \mathbf{x}_k\|)\|\mathbf{x} - \mathbf{x}_{k+1}\|.\end{aligned} \tag{3.7}$$

Rearranging (C0) and utilizing (3.7) gives, for all $\mathbf{x} \in \mathbf{X}$,

$$\begin{aligned}&f(\mathbf{x}_{k+1}) - f^* \\ &\quad \leq \frac{\mu_k}{2}(\|\mathbf{x} - \mathbf{x}_k\|^2 - \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2) + \frac{\mu_k^2}{2}\|\mathbf{x} - \mathbf{x}_k\|^2 \\ &\quad \leq \frac{\mu_k}{2}(\|\mathbf{x} - \mathbf{x}_{k+1}\| + 2\|\mathbf{x}_{k+1} - \mathbf{x}_k\|)\|\mathbf{x} - \mathbf{x}_{k+1}\| + \frac{\mu_k^2}{2}\|\mathbf{x} - \mathbf{x}_k\|^2 \\ &\quad \leq \frac{\mu_k}{2}(\|\mathbf{x} - \mathbf{x}_{k+1}\| + 2\|\mathbf{x}_{k+1} - \mathbf{x}_k\|)\|\mathbf{x} - \mathbf{x}_{k+1}\| + \frac{\mu_k^2}{2}(\|\mathbf{x} - \mathbf{x}_{k+1}\| + \|\mathbf{x}_{k+1} - \mathbf{x}_k\|)^2 \\ &\quad \leq \frac{\mu_k}{2}\|\mathbf{x} - \mathbf{x}_{k+1}\|^2 + \mu_k\|\mathbf{x}_{k+1} - \mathbf{x}_k\|\|\mathbf{x} - \mathbf{x}_{k+1}\| \\ &\qquad\qquad + \mu_k^2(\|\mathbf{x} - \mathbf{x}_{k+1}\|^2 + \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2).\end{aligned}$$

Utilizing the inequalities

$$\mu_k\|\mathbf{x}_{k+1} - \mathbf{x}_k\|\|\mathbf{x} - \mathbf{x}_{k+1}\| \leq \frac{\alpha}{3}\|\mathbf{x} - \mathbf{x}_{k+1}\|^2 + \frac{3\mu_k^2}{4\alpha}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2$$

and $\mu_k \leq \beta$, we obtain

$$f(\mathbf{x}_{k+1}) - f^* \leq \left(\frac{\beta + 2\beta^2}{2} + \frac{\alpha}{3}\right)\|\mathbf{x} - \mathbf{x}_{k+1}\|^2 + \left(\frac{3 + 4\alpha}{4\alpha}\right)\mu_k^2\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2. \tag{3.8}$$

Taking the infimum over $\mathbf{x} \in \mathbf{X}$ on the right-hand side of (3.8) gives

$$f(\mathbf{x}_{k+1}) - f^* \leq \left(\frac{\beta + 2\beta^2}{2} + \frac{\alpha}{3}\right) D(\mathbf{x}_{k+1}, \mathbf{X})^2 + \left(\frac{3 + 4\alpha}{4\alpha}\right)\mu_k^2\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2. \tag{3.9}$$

By (3.6), $\mathbf{x}_{k+1} \in \mathcal{B}_\rho(\hat{\mathbf{x}})$. Since f provides a local error bound at $\hat{\mathbf{x}}$,

$$\alpha D(\mathbf{x}_{k+1}, \mathbf{X})^2 \leq f(\mathbf{x}_{k+1}) - f^*. \tag{3.10}$$

Combining this with (3.9) gives

$$\left(\frac{2\alpha}{3} - \frac{\beta + 2\beta^2}{2}\right) D(\mathbf{x}_{k+1}, \mathbf{X})^2 \le \left(\frac{3 + 4\alpha}{4\alpha}\right) \mu_k^2 \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2. \tag{3.11}$$

By (E1), the coefficient of D in (3.11) is bounded from below by $\alpha/3$. Hence, (3.11), Lemma 3.1, and (3.5), with $j = k$, yield

$$\begin{aligned} D(\mathbf{x}_{k+1}, \mathbf{X}) &\le \mu_k \frac{\sqrt{3(3+4\alpha)}}{2\alpha} \|\mathbf{x}_{k+1} - \mathbf{x}_k\| \\ &\le \mu_k \sqrt{1 + \mu_k} \frac{\sqrt{3(3+4\alpha)}}{2\alpha} D(\mathbf{x}_k, \mathbf{X}) \\ &\le \mu_k \frac{\sqrt{3(1+\beta)(3+4\alpha)}}{2\alpha} D(\mathbf{x}_k, \mathbf{X}) \qquad (3.12) \\ &\le \gamma D(\mathbf{x}_k, \mathbf{X}) \le \gamma^{k+1} D(\mathbf{x}_0, \mathbf{X}). \qquad (3.13) \end{aligned}$$

Relations (3.6) and (3.13) complete the proof of the induction step. Relations (3.12) and (3.13) give the estimate (3.3).

By Lemma 3.1 and (3.5), the proximal iterates $\mathbf{x}_k$ form a Cauchy sequence in $\mathcal{H}$, which has a limit denoted $\mathbf{x}^*$. By (3.5), Lemma 3.1, and the bound $\mu_k \le \beta$, we have

$$\begin{aligned} \|\mathbf{x}_k - \mathbf{x}^*\| &\le \sum_{j=k}^{\infty} \|\mathbf{x}_{j+1} - \mathbf{x}_j\| \le \sum_{j=k}^{\infty} \sqrt{1 + \mu_k} D(\mathbf{x}_j, \mathbf{X}) \\ &\le \sqrt{1+\beta} \sum_{j=k}^{\infty} \gamma^j D(\mathbf{x}_0, \mathbf{X}) = \gamma^j \frac{\sqrt{1+\beta}}{1-\gamma} D(\mathbf{x}_0, \mathbf{X}). \qquad (3.14) \end{aligned}$$

By (3.5) and (3.14), for any $k \ge 0$ we have

$$D(\mathbf{x}^*, \mathbf{X}) \le D(\mathbf{x}_k, \mathbf{X}) + \|\mathbf{x}_k - \mathbf{x}^*\| \le \left(\gamma^k + \gamma^k \frac{\sqrt{1+\beta}}{1-\gamma}\right) D(\mathbf{x}_0, \mathbf{X}).$$

Thus $D(\mathbf{x}^*, \mathbf{X}) = 0$. Since $\mathbf{X}$ is closed, the limit $\mathbf{x}^* \in \mathbf{X}$. □

We now give a choice for μ_k which leads to a quadratic convergence rate for the proximal point iteration.

COROLLARY 3.3. *Assume that conditions* (E0), (E1), *and* (E3) *of Theorem* 3.2 *are satisfied. In addition, let* $e : \mathcal{H} \mapsto \mathbb{R}$ *be any nonnegative function with the property that*

$$e(\mathbf{x}) \le \beta \quad \text{and} \quad e(\mathbf{x}) \le L D(\mathbf{x}, \mathbf{X}) \tag{3.15}$$

for all $\mathbf{x} \in \mathcal{B}_\rho(\hat{\mathbf{x}})$ *and for some* $L \in \mathbb{R}$. *Then for the choice* $\mu_k = e(\mathbf{x}_k)$, *any proximal iterates* $\{\mathbf{x}_k\}$ *satisfying* (C0) *have the property that* $\mathbf{x}_k \in \mathcal{B}_\rho(\hat{\mathbf{x}})$ *for each* k, *the iterates approach a minimizer* $\mathbf{x}^* \in \mathbf{X}$, *and for each* k, *we have*

$$D(\mathbf{x}_{k+1}, \mathbf{X}) \le c_2 L D(\mathbf{x}_k, \mathbf{X})^2, \tag{3.16}$$

where c_2 *is given in* (3.4).

Proof. This follows directly from the proof of Theorem 3.2; simply append the condition $\mu_j \le \beta$ for each $j \in [0, k]$ to the induction hypothesis (3.5):

$$\|\mathbf{x}_j - \hat{\mathbf{x}}\| \le \rho, \quad D(\mathbf{x}_j, \mathbf{X}) \le \gamma^j D(\mathbf{x}_0, \mathbf{X}), \quad \text{and} \quad \mu_j \le \beta. \tag{3.17}$$

Since $\mathbf{x}_0 \in \mathcal{B}_\rho(\hat{\mathbf{x}})$, (3.15) implies that $\mu_0 = e(\mathbf{x}_0) \le \beta$. Hence, (3.17) is satisfied for $j = 0$. In the proof of Theorem 3.2, we show that if $\mu_j \le \beta$ for $j \in [0, k]$, then the first two conditions in (3.17) hold for $j = k+1$. In (3.6), we show that $\mathbf{x}_{k+1} \in \mathcal{B}_\rho(\hat{\mathbf{x}})$. Consequently, $\mu_{k+1} = e(\mathbf{x}_{k+1}) \le \beta$, and (3.17) holds for $j = k+1$. Replacing μ_k by $e(\mathbf{x}_k) \le LD(\mathbf{x}_k, \mathbf{X})$ in (3.3) gives (3.16). ☐

If f is Lipschitz continuously differentiable, then the function $e(\mathbf{x}) = \|f'[\mathbf{x}]\|$ satisfies the hypotheses of Corollary 3.3 when ρ is sufficiently small since $f'[\bar{\mathbf{x}}] = \mathbf{0}$ for all $\bar{\mathbf{x}} \in \mathbf{X}$.

4. Convergence analysis for approximate minimization.

We now analyze the situation where the proximal point iteration (1.2) need only satisfy (C1) or (C2). The following property of a convex function is used in the analysis.

PROPOSITION 4.1. *If $\mathbf{x}^*$ is a local minimizer of F_k and f is convex in $\mathcal{B}_\rho(\mathbf{x}^*)$ for some $\rho > 0$, then*

$$F_k(\mathbf{x}) \le F_k(\mathbf{x}^*) + \frac{\|\partial F_k(\mathbf{x})\|_{\inf}^2}{\mu_k} \tag{4.1}$$

for all $\mathbf{x} \in \mathcal{B}_\rho(\mathbf{x}^*)$.

Proof. If $\partial F_k(\mathbf{x})$ is empty, then $\|\partial F_k(\mathbf{x})\|_{\inf} = \infty$, and there is nothing to prove. Hence, we assume that $\partial F_k(\mathbf{x}) \ne \emptyset$. Since f is convex in $\mathcal{B}_\rho(\mathbf{x}^*)$, we have

$$F_k(\mathbf{x}^*) \ge F_k(\mathbf{x}) + \langle \mathbf{y}, \mathbf{x}^* - \mathbf{x}\rangle \quad \text{for all } \mathbf{y} \in \partial F_k(\mathbf{x}). \tag{4.2}$$

For a convex functional, the subdifferentials satisfy the monotonicity condition [17, Thm. 3.56]

$$\langle \bar{\mathbf{y}} - \mathbf{y}^*, \mathbf{x} - \mathbf{x}^*\rangle \ge 0 \quad \text{for all } \bar{\mathbf{y}} \in \partial f(\mathbf{x}) \quad \text{and} \quad \mathbf{y}^* \in \partial f(\mathbf{x}^*). \tag{4.3}$$

Given $\bar{\mathbf{y}} \in \partial f(\mathbf{x})$, define

$$\mathbf{y} = \bar{\mathbf{y}} + \mu_k(\mathbf{x} - \mathbf{x}_k) \in \partial F_k(\mathbf{x}). \tag{4.4}$$

Since $\mathbf{x}^*$ is a local minimizer of F_k, $\mathbf{0} \in \partial F_k(\mathbf{x}^*)$, or equivalently, there exists $\mathbf{y}^* \in \partial f(\mathbf{x}^*)$ such that

$$\mathbf{0} = \mathbf{y}^* + \mu_k(\mathbf{x}^* - \mathbf{x}_k). \tag{4.5}$$

By (4.3), (4.4), and (4.5), we have

$$\begin{aligned} \langle \mathbf{y}, (\mathbf{x} - \mathbf{x}^*)\rangle &= \langle \bar{\mathbf{y}} + \mu_k(\mathbf{x} - \mathbf{x}_k) - (\mathbf{y}^* + \mu_k(\mathbf{x}^* - \mathbf{x}_k)), \mathbf{x} - \mathbf{x}^*\rangle \\ &= \langle \bar{\mathbf{y}} - \mathbf{y}^*, \mathbf{x} - \mathbf{x}^*\rangle + \mu_k\|\mathbf{x} - \mathbf{x}^*\|^2 \\ &\ge \mu_k\|\mathbf{x} - \mathbf{x}^*\|^2 \end{aligned} \tag{4.6}$$

for any $\mathbf{y} \in \partial F_k(\mathbf{x})$. The Schwarz inequality yields

$$\|\mathbf{x} - \mathbf{x}^*\| \le \frac{\|\mathbf{y}\|}{\mu_k}. \tag{4.7}$$

Thus, it follows from (4.2) and (4.7) that for any $\mathbf{y} \in \partial F_k(\mathbf{x})$,

$$F_k(\mathbf{x}) \le F_k(\mathbf{x}^*) + \frac{\|\mathbf{y}\|^2}{\mu_k}.$$

Minimizing over $\mathbf{y} \in \partial F_k(\mathbf{x})$ gives (4.1). $\square$

In our first convergence result, we focus on the case where f is convex over the level set defined by the starting guess. We also employ a subdifferential generalization of the gradient-based local error bound condition (2.1): For some $\hat{\mathbf{x}} \in \mathbf{X}$, there exist positive constant α and ρ such that

$$\|\partial f(\mathbf{x})\|_{\inf} \geq \alpha D(\mathbf{x}, \mathbf{X}) \quad \text{whenever } \|\mathbf{x} - \hat{\mathbf{x}}\| \leq \rho. \tag{4.8}$$

If (4.8) is satisfied, then we say that ∂f provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$.

THEOREM 4.2. *Assume that the following conditions are satisfied:*

(A0) *∂f provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$ with positive scalars α and ρ satisfying* (4.8).

(A1) *If $\mathcal{L}$ is the level set $\{\mathbf{x} \in \mathcal{H} : f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$, then f is convex and lower semicontinuous on $\mathcal{L}$, and there exists a constant L such that $\|\partial f(\mathbf{x})\|_{\inf} \leq LD(\mathbf{x}, \mathbf{X})$ for all $\mathbf{x} \in \mathcal{L}$.*

(A2) *Define the parameters*

$$\Lambda = L + \tau \quad \textit{and} \quad \tau^2 = 1 + 2L^2 \ \textit{ if acceptance criterion (C1) is used,}$$

while

$$\Lambda = \tau(1 + \theta) \quad \textit{and} \quad \tau^2 = \frac{1}{1 - 2\theta^2} \ \textit{ if acceptance criterion (C2) is used.}$$

$\beta > 0$ is small enough that the following inequality holds:

$$\gamma := \frac{\beta \Lambda}{\alpha} < 1. \tag{4.9}$$

(A3) *$\mu_k \in (0, \beta]$ and $\theta < 1/\sqrt{2}$.*

(A4) *$\mathbf{x}_0$ is close enough to $\hat{\mathbf{x}}$ that*

$$\|\mathbf{x}_0 - \hat{\mathbf{x}}\| \left(1 + \frac{\tau}{1 - \gamma}\right) \leq \rho.$$

If the approximate proximal iterates $\mathbf{x}_k$ satisfy either (C1) *or* (C2), *then the iterates are all contained in $\mathcal{B}_\rho(\hat{\mathbf{x}})$, and they approach a minimizer $\mathbf{x}^* \in \mathbf{X}$; moreover, for each k, we have*

$$\|\mathbf{x}_k - \mathbf{x}^*\| \leq c_1 \gamma^k D(\mathbf{x}_0, \mathbf{X}) \quad \textit{and} \quad D(\mathbf{x}_{k+1}, \mathbf{X}) \leq c_2 \mu_k D(\mathbf{x}_k, \mathbf{X}), \tag{4.10}$$

where

$$c_1 = \frac{\tau}{1 - \gamma} \quad \textit{and} \quad c_2 = \gamma/\beta.$$

Proof. For $j = 0$, we have

$$\|\mathbf{x}_j - \hat{\mathbf{x}}\| \leq \rho, \quad \mathbf{x}_j \in \mathcal{L}, \quad \text{and} \quad D(\mathbf{x}_j, \mathbf{X}) \leq \gamma^j D(\mathbf{x}_0, \mathbf{X}). \tag{4.11}$$

Proceeding by induction, suppose that (4.11) holds for all $j \in [0, k]$ and for some $k \geq 0$. We show that (4.11) also holds for $j = k + 1$.

Due to the convexity and lower semicontinuity of f on $\mathcal{L}$, this level set is closed and convex. Suppose $j \in [0, k]$. By (C1) or (C2), we have

$$f(\mathbf{x}_{j+1}) \leq F_j(\mathbf{x}_{j+1}) \leq f(\mathbf{x}_j). \tag{4.12}$$

We conclude that $f(\mathbf{x}_j) \le f(\mathbf{x}_0)$ for each j. Since $F_j(\mathbf{x}_j) = f(\mathbf{x}_j) \le f(\mathbf{x}_0)$, minimizing F_j over $\mathcal{H}$ is equivalent to minimizing F_j over $\mathcal{L}$. Since F_j is strongly convex and lower semicontinuous on $\mathcal{L}$, F_j is weakly lower semicontinuous on $\mathcal{L}$, and there exists an exact proximal point iterate $\mathbf{x}_j^*$ defined by

$$\mathbf{x}_j^* \in \arg\ \min\ \{F_j(\mathbf{x}) : \mathbf{x} \in \mathcal{H}\}.$$

Moreover, $f(\mathbf{x}_j^*) \le f(\mathbf{x}_j) \le f(\mathbf{x}_0)$. Combining this with (4.12), both $\mathbf{x}_{j+1}$ and $\mathbf{x}_j^*$ lie in $\mathcal{L}$. By (A2) and Proposition 4.1, we have

$$\begin{aligned} F_j(\mathbf{x}_{j+1}) &= F_j(\mathbf{x}_j^*) + (F_j(\mathbf{x}_{j+1}) - F_j(\mathbf{x}_j^*)) \\ &\le F_j(\mathbf{x}_j^*) + \frac{\|\partial F_j(\mathbf{x}_{j+1})\|_{\inf}^2}{\mu_j} \\ &\le f^* + \frac{\mu_j}{2} D(\mathbf{x}_j, \mathbf{X})^2 + \frac{\|\partial F_j(\mathbf{x}_{j+1})\|_{\inf}^2}{\mu_j}. \end{aligned}$$

Since $f^* \le f(\mathbf{x}_{j+1})$, it follows that

$$\text{(4.13)} \qquad \frac{\mu_j}{2}\|\mathbf{x}_{j+1} - \mathbf{x}_j\|^2 \le \frac{\mu_j}{2} D(\mathbf{x}_j, \mathbf{X})^2 + \frac{\|\partial F_j(\mathbf{x}_{j+1})\|_{\inf}^2}{\mu_j}.$$

By (C1) and (A1),

$$\text{(4.14)} \qquad \|\partial F_j(\mathbf{x}_{j+1})\|_{\inf} \le \mu_j \|\partial f(\mathbf{x}_j)\|_{\inf} \le \mu_j L D(\mathbf{x}_j, \mathbf{X}).$$

Combining this with (4.13), we have

$$\text{(4.15)} \qquad \|\mathbf{x}_{j+1} - \mathbf{x}_j\|^2 \le (1 + 2L^2) D(\mathbf{x}_j, \mathbf{X})^2.$$

Similarly, if criterion (C2) is used, then $\|\partial F_j(\mathbf{x}_{j+1})\|_{\inf} \le \theta\mu_j \|\mathbf{x}_{j+1} - \mathbf{x}_j\|$, and by (4.13), we have

$$\text{(4.16)} \qquad \|\mathbf{x}_{j+1} - \mathbf{x}_j\|^2 \le \frac{1}{1 - 2\theta^2} D(\mathbf{x}_j, \mathbf{X})^2.$$

Together, (4.15) and (4.16) yield

$$\text{(4.17)} \qquad \|\mathbf{x}_{j+1} - \mathbf{x}_j\| \le \tau D(\mathbf{x}_j, \mathbf{X}),$$

where τ is defined in (A2); this holds for any $j \in [0, k]$.

By (4.11), we have

$$\begin{aligned} \|\mathbf{x}_{k+1} - \mathbf{x}_0\| &\le \sum_{j=0}^{k} \|\mathbf{x}_{j+1} - \mathbf{x}_j\| \le \sum_{j=0}^{k} \tau D(\mathbf{x}_j, \mathbf{X}) \\ &\le \tau \sum_{j=0}^{k} \gamma^j D(\mathbf{x}_0, \mathbf{X}) \le \frac{\tau}{1-\gamma} D(\mathbf{x}_0, \mathbf{X}) \le \frac{\tau}{1-\gamma} \|\mathbf{x}_0 - \hat{\mathbf{x}}\|. \end{aligned}$$

Again, by the triangle inequality and (A4),

$$\|\mathbf{x}_{k+1} - \hat{\mathbf{x}}\| \le \|\mathbf{x}_{k+1} - \mathbf{x}_0\| + \|\mathbf{x}_0 - \hat{\mathbf{x}}\| \le \left(1 + \frac{\tau}{1-\gamma}\right) \|\mathbf{x}_0 - \hat{\mathbf{x}}\| \le \rho.$$

Hence, $\mathbf{x}_{k+1} \in \mathcal{B}_\rho(\hat{\mathbf{x}})$, which establishes the first relation in (4.11).

By (A0), we have

$$\alpha D(\mathbf{x}_{k+1}, \mathbf{X}) \leq \|\partial f(\mathbf{x}_{k+1})\|_{\text{inf}} \leq \|\partial F_k(\mathbf{x}_{k+1})\|_{\text{inf}} + \mu_k \|\mathbf{x}_{k+1} - \mathbf{x}_k\|. \tag{4.18}$$

If (C1) is used, then $\|\partial F_k(\mathbf{x}_{k+1})\|_{\text{inf}} \leq \mu_k \|\partial f(\mathbf{x}_k)\|_{\text{inf}}$; hence, (A1) and (4.18) imply that

$$\begin{aligned} \alpha D(\mathbf{x}_{k+1}, \mathbf{X}) &\leq \mu_k (\|\partial f(\mathbf{x}_k)\|_{\text{inf}} + \|\mathbf{x}_{k+1} - \mathbf{x}_k\|) \\ &\leq \mu_k (L D(\mathbf{x}_k, \mathbf{X}) + \|\mathbf{x}_{k+1} - \mathbf{x}_k\|). \end{aligned} \tag{4.19}$$

If (C2) is used, then $\|\partial F_k(\mathbf{x}_{k+1})\|_{\text{inf}} \leq \theta \mu_k \|\mathbf{x}_{k+1} - \mathbf{x}_k\|$, and by (4.18), we have

$$\alpha D(\mathbf{x}_{k+1}, \mathbf{X}) \leq \mu_k (1 + \theta) \|\mathbf{x}_{k+1} - \mathbf{x}_k\|. \tag{4.20}$$

Inserting the bound (4.17) into (4.19) or (4.20) yields the second half of (4.10). By the second half of (4.10) and (A3), we have

$$D(\mathbf{x}_{k+1}, \mathbf{X}) \leq \left(\frac{\Lambda \mu_k}{\alpha} \right) D(\mathbf{x}_k, \mathbf{X}) \leq \gamma D(\mathbf{x}_k, \mathbf{X}) \leq \gamma^{k+1} D(\mathbf{x}_0, \mathbf{X}).$$

This establishes the last relation in (4.11) for $j = k+1$, and the proof of the induction step is complete. The proof that the $\mathbf{x}_k$ form a Cauchy sequence converging to a limit $\mathbf{x}^* \in \mathbf{X}$ and the first part of (4.10) are exactly as in Theorem 3.2. ∎

Suppose that $\mathbf{x}^*$ is a local minimizer of F_k, f is convex in $\mathcal{B}_\rho(\mathbf{x}^*)$ for some $\rho > 0$, and the following inequality holds:

$$f(\mathbf{x}_1) \geq f(\mathbf{x}_2) + \frac{1}{2} \langle \mathbf{y}_1 + \mathbf{y}_2, \mathbf{x}_1 - \mathbf{x}_2 \rangle \tag{4.21}$$

whenever $\mathbf{y_i} \in \partial f(\mathbf{x}_i)$, $i = 1, 2$. For example, when f is a quadratic, (4.21) is satisfied with equality. When (4.21) holds, Proposition 4.1 can be strengthened to

$$F_k(\mathbf{x}) \leq F_k(\mathbf{x}^*) + \frac{\|\partial F_k(\mathbf{x})\|_{\text{inf}}^2}{2\mu_k} \tag{4.22}$$

for all $\mathbf{x} \in \mathcal{B}_\rho(\mathbf{x}^*)$. In Theorem 4.2, we require that $\theta < 1/\sqrt{2}$ in (A3); this requirement arises at inequality (4.16) since we need to ensure that $1 - 2\theta^2 > 0$. If f satisfies (4.21), then by exploiting the stronger inequality (4.22), the restriction on θ for stopping criterion (C2) can be relaxed to $\theta < 1$.

We now relax the convexity requirement for f while strengthening the smoothness condition. We require only that $\mathbf{X}$ is locally convex, while f is locally, twice continuously differentiable. If the set $\mathcal{B}_\rho(\hat{\mathbf{x}}) \cap \mathbf{X}$ is convex for some $\rho > 0$, then the projection $\bar{\mathbf{x}}$ of $\mathbf{x}$ onto $\mathcal{B}_\rho(\hat{\mathbf{x}}) \cap \mathbf{X}$ exists. For $\mathbf{x} \in \mathcal{B}_{\rho/2}(\hat{\mathbf{x}})$, it follows that $\|\mathbf{x} - \mathbf{y}\| \geq \rho/2$ when $\mathbf{y} \in \mathcal{B}_\rho(\hat{\mathbf{x}})^c$, where c denotes complement. Hence, the distance from $\mathbf{x} \in \mathcal{B}_{\rho/2}(\hat{\mathbf{x}})$ to $\mathbf{X}$ is the same as the distance from $\mathbf{x}$ to $\mathbf{X} \cap \mathcal{B}_\rho(\hat{\mathbf{x}})$:

$$\|\mathbf{x} - \bar{\mathbf{x}}\| = \min\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in \mathcal{B}_\rho(\hat{\mathbf{x}}) \cap \mathbf{X}\} = \min\ \{\|\mathbf{x} - \mathbf{y}\| : \mathbf{y} \in \mathbf{X}\} = D(\mathbf{x}, \mathbf{X}).$$

The following lemma plays the role of Proposition 4.1 when we remove the convexity requirement for f.

LEMMA 4.3. *Suppose f provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$ with positive scalars α and ρ satisfying* (1.3)*, $\mathbf{X} \cap B_\rho(\hat{\mathbf{x}})$ is convex, and f is twice Lipschitz continuously*

Fréchet differentiable on $\mathcal{B}_\rho(\hat{\mathbf{x}})$. If $\mu_k = \beta\|f'[\mathbf{x}_k]\|^\eta$, where $\eta \ge 0$ and $\beta > 0$, then there exist $r \in (0, \rho/2]$ and positive constants C_1 and C_2 with the following property: For each $\mathbf{x}_k \in \mathcal{B}_r(\hat{\mathbf{x}})$, we have

$$F_k(\mathbf{x}) - F_k(\bar{\mathbf{x}}) \le \frac{C_1}{\mu_k}\|F_k'(\mathbf{x})\|^2 \tag{4.23}$$

whenever $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$ and $\|\mathbf{x} - \bar{\mathbf{x}}\| \le C_2\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta$.

Proof. If $\mathbf{x}_k = \bar{\mathbf{x}}_k$, then (4.23) is trivial. Hence, we assume that $\mathbf{x}_k \ne \bar{\mathbf{x}}_k$. By Lemma 2.1, f' provides a local error bound with constants α and $r \in (0, \rho/2]$. Hence, for any $\mathbf{x}_k \in \mathcal{B}_r(\hat{\mathbf{x}})$, we have

$$\alpha\|\mathbf{x}_k - \bar{\mathbf{x}}_k\| = \alpha D(\mathbf{x}_k, \mathbf{X}) \le \|f'[\mathbf{x}_k]\|.$$

Raising this inequality to the η power and utilizing the definition of μ_k gives

$$\mu_k \ge \beta\alpha^\eta\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta. \tag{4.24}$$

By the second-order necessary optimality condition, the second derivative operator $f''[\mathbf{x}]$ is positive for any $\mathbf{x} \in \mathbf{X}$. Hence, given any $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$ and $\mathbf{y} \in \mathcal{H}$, we deduce from (4.24) that

$$\begin{aligned}\langle \mathbf{y}, F_k''(\mathbf{x})\mathbf{y}\rangle &= \langle \mathbf{y}, (f''[\bar{\mathbf{x}}] + \mu_k\mathbf{I} + f''[\mathbf{x}] - f''[\bar{\mathbf{x}}])\mathbf{y}\rangle \\ &\ge \mu_k\|\mathbf{y}\|^2 + \langle \mathbf{y}, (f''[\mathbf{x}] - f''[\bar{\mathbf{x}}])\mathbf{y}\rangle \\ &\ge \left(\beta\alpha^\eta\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta - L_2\|\mathbf{x} - \bar{\mathbf{x}}\|\right)\|\mathbf{y}\|^2,\end{aligned}$$

where L_2 is a Lipschitz constant for f'' on $\mathcal{B}_\rho(\hat{\mathbf{x}})$. Hence, if $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$ satisfies

$$\|\mathbf{x} - \bar{\mathbf{x}}\| \le C_2\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta, \quad C_2 = \frac{\beta\alpha^\eta}{2L_2}, \tag{4.25}$$

then we have

$$\langle \mathbf{y}, F_k''(\mathbf{x})\mathbf{y}\rangle \ge \frac{\beta\alpha^\eta}{2}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta\|\mathbf{y}\|^2. \tag{4.26}$$

Let $\mathcal{A}$ be the collection of $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$ which satisfies (4.25):

$$\mathcal{A} = \{\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}}) : \|\mathbf{x} - \bar{\mathbf{x}}\| \le C_2\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta\}.$$

We now show that $\mathcal{A}$ is closed and convex. Since $r \le \rho/2$, it follows from the discussion preceding the lemma that for each $\mathbf{y}$ and $\mathbf{z} \in \mathcal{B}_r(\hat{\mathbf{x}})$, the projections $\bar{\mathbf{y}}$ and $\bar{\mathbf{z}}$ onto $\mathbf{X}$ exist in $\mathcal{B}_\rho(\hat{\mathbf{x}})$. By the convexity of $\mathcal{B}_\rho(\hat{\mathbf{x}}) \cap \mathbf{X}$, the line segment $[\bar{\mathbf{y}}, \bar{\mathbf{z}}]$ is contained in $\mathcal{B}_\rho(\hat{\mathbf{x}}) \cap \mathbf{X}$. Thus, if $\mathbf{y}$ and $\mathbf{z} \in \mathcal{A}$, then each $\mathbf{x} \in [\mathbf{y}, \mathbf{z}]$ lies in $\mathcal{A}$. $\mathcal{A}$ is closed since the projection onto a convex set is Lipschitz continuous.

By (4.26), F_k is strongly convex over the closed, convex set $\mathcal{A}$. Consequently, there exists a unique minimizer $\mathbf{x}_k^*$:

$$\mathbf{x}_k^* = \arg\min\{F_k(\mathbf{x}) : \mathbf{x} \in \mathcal{A}\}. \tag{4.27}$$

Given $\mathbf{x} \in \mathcal{A}$ and $t \in [0, 1]$, we define $\mathbf{x}(t) = \mathbf{x}_k^* + t(\mathbf{x} - \mathbf{x}_k^*)$. Since $\mathcal{A}$ is convex and both $\mathbf{x}$ and $\mathbf{x}_k^* \in \mathcal{A}$, it follows that $\mathbf{x}(t) \in \mathcal{A}$ for all $t \in [0, 1]$. Since $\mathbf{x}(0) = \mathbf{x}_k^*$

achieves the minimum in (4.27), we have $F_k(\mathbf{x}_k^*) \le F_k(\mathbf{x}(t))$ for all $t \in [0,1]$. Thus for $\phi(t) := F_k(\mathbf{x}(t))$, we have $\phi'(0) \ge 0$, and by (4.26),

$$\begin{aligned}\|F_k'(\mathbf{x})\|\|\mathbf{x}-\mathbf{x}_k^*\| &\ge F_k'(\mathbf{x})(\mathbf{x}-\mathbf{x}_k^*) = \phi'(1) \ge \phi'(1) - \phi'(0) = \phi''(s) \\ &= \langle F_k''(\mathbf{x}(s))(\mathbf{x}-\mathbf{x}_k^*), \mathbf{x}-\mathbf{x}_k^*\rangle \\ &\ge \frac{\beta\alpha^\eta}{2}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta \|\mathbf{x}-\mathbf{x}_k^*\|^2\end{aligned}$$

for some $s \in [0,1]$. Hence, we have

$$\|\mathbf{x}-\mathbf{x}_k^*\| \le \frac{2}{\beta\alpha^\eta\|\mathbf{x}_k-\bar{\mathbf{x}}_k\|^\eta}\|F_k'(\mathbf{x})\|. \tag{4.28}$$

By (4.26), $F_k(\mathbf{x}(t))$ is convex as a function of $t \in [0,1]$. This implies that $\phi'(t)$ is an increasing function of $t \in [0,1]$. We combine this observation with (4.27) and (4.28) to obtain

$$\begin{aligned}F_k(\mathbf{x}) - F_k(\bar{\mathbf{x}}_k) &\le F_k(\mathbf{x}) - F_k(\mathbf{x}_k^*) \\ &= F_k(\mathbf{x}(1)) - F_k(\mathbf{x}(0)) \\ &= \int_0^1 \phi'(t)dt \le \phi'(1) = F_k'(\mathbf{x})(\mathbf{x}-\mathbf{x}_k^*) \\ &\le \frac{2}{\beta\alpha^\eta\|\mathbf{x}_k-\bar{\mathbf{x}}_k\|^\eta}\|F_k'(\mathbf{x})\|^2 \qquad (4.29)\end{aligned}$$

whenever $\mathbf{x} \in \mathcal{A}$ and $\mathbf{x}_k \in \mathcal{B}_r(\hat{\mathbf{x}})$.

If L_1 is a Lipschitz constant for f' over $\mathcal{B}_\rho(\hat{\mathbf{x}})$, then we have

$$\|f'[\mathbf{x}_k]\| = \|f'[\mathbf{x}_k] - f'[\bar{\mathbf{x}}_k]\| \le L_1\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|$$

whenever $\mathbf{x}_k \in \mathcal{B}_r(\hat{\mathbf{x}})$. By the definition of μ_k, it follows that

$$\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta \ge \mu_k/(\beta L_1^\eta).$$

Combining this with (4.29) gives (4.23) with $C_1 = 2L_1^\eta/\alpha^\eta$. ◻

THEOREM 4.4. *Assume that the following conditions are satisfied:*

(B0) *f provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$ with positive scalars α and ρ satisfying* (1.3).

(B1) *The set $\mathcal{B}_\rho(\hat{\mathbf{x}}) \cap \mathbf{X}$ is convex.*

(B2) *f is twice Lipschitz continuously Fréchet differentiable in $\mathcal{B}_\rho(\hat{\mathbf{x}})$.*

(B3) *The proximal iterates $\mathbf{x}_k$ satisfy either* (C1) *or* (C2) *with $\mu_k = \beta\|f'[\mathbf{x}_k]\|^\eta$, where $\eta \in (0,2)$ and β is positive. If* (C2) *is used, then θ is small enough that $2C_1\theta^2 < 1$, where C_1 is the constant in* (4.23).

Then for ϵ sufficiently small and for each $\mathbf{x}_0 \in \mathcal{B}_\epsilon(\hat{\mathbf{x}})$, the proximal iterates $\mathbf{x}_k$ are all contained in $\mathcal{B}_\rho(\hat{\mathbf{x}})$, and they approach a minimizer $\mathbf{x}^* \in \mathbf{X}$; moreover, for each k, we have

$$\begin{aligned}&\|\mathbf{x}_k - \mathbf{x}^*\| \le c_1\gamma^k D(\mathbf{x}_0, \mathbf{X}) \quad \textit{and} \\ &D(\mathbf{x}_{k+1}, \mathbf{X}) \le c_2\mu_k D(\mathbf{x}_k, \mathbf{X}) \le \beta c_2 L_1^\eta D(\mathbf{x}_k, \mathbf{X})^{1+\eta}, \qquad (4.30)\end{aligned}$$

where $\gamma < 1$, c_1, and c_2 are constants independent of k.

Proof. We start by explaining how to choose ϵ so that the theorem holds. Define the parameters

$$\Lambda = L_1 + \tau \quad \text{and} \quad \tau^2 = 1 + 2C_1 L_1^2 \quad \text{if acceptance criterion (C1) is used,}$$

while

$$\Lambda = \tau(1+\theta) \quad \text{and} \quad \tau^2 = \frac{1}{1-2C_1\theta^2} \quad \text{if acceptance criterion (C2) is used,}$$

where C_1 is the constant in (4.23) and L_1 is a Lipschitz constant for f' over $\mathcal{B}_\rho(\hat{\mathbf{x}})$. Notice that the hypotheses of the theorem are satisfied if ρ is decreased. Choose ρ small enough that

$$\gamma := \sup_{\mathbf{x}\in\mathcal{B}_\rho(\hat{\mathbf{x}})} \frac{\beta\|f'[\mathbf{x}]\|^\eta \Lambda}{\alpha} < 1. \tag{4.31}$$

By Lemma 2.1, we can choose ρ smaller, if necessary, so that f' provides a local error bound with constants α and $\rho/2$. By Lemma 4.3, we can choose ρ smaller, if necessary, so that (4.23) holds whenever $\mathbf{x} \in \mathcal{B}_\rho(\hat{\mathbf{x}})$ and $\|\mathbf{x} - \bar{\mathbf{x}}\| \le C_2\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta$. Choose $\epsilon > 0$ small enough that

$$\left(\epsilon + \frac{E\epsilon^{1-\eta/2}}{1-\gamma^{1-\eta/2}}\right) \le \frac{\rho}{2}, \quad \text{where} \quad E = \sqrt{\frac{L_1}{2\beta\alpha^\eta}}, \tag{4.32}$$

and

$$\epsilon(L_1 + E\epsilon^{-\eta/2})\left(\frac{\beta L_1^\eta}{\alpha}\right) \le C_2 \quad \text{if stopping criterion (C1) is used,} \tag{4.33}$$

$$E\epsilon^{1-\eta/2}(1+\theta)\left(\frac{\beta L_1^\eta}{\alpha}\right) \le C_2 \quad \text{if stopping criterion (C2) is used.} \tag{4.34}$$

Since $\eta \in (0,2)$, (4.33) and (4.34) are satisfied for ϵ sufficiently small.

We now prove the theorem. Again, let $\bar{\mathbf{x}}$ be the projection of $\mathbf{x}$ onto $\mathbf{X}$. For $j = 0$, we have

$$\|\mathbf{x}_j - \hat{\mathbf{x}}\| \le \rho/2 \quad \text{and} \quad \|\mathbf{x}_j - \bar{\mathbf{x}}_j\| \le \gamma^j\|\mathbf{x}_0 - \bar{\mathbf{x}}_0\| \tag{4.35}$$

since $\mathbf{x}_0 \in \mathcal{B}_\epsilon(\hat{\mathbf{x}}) \subset \mathcal{B}_{\rho/2}(\hat{\mathbf{x}})$. Proceeding by induction, suppose that (4.35) holds for all $j \in [0,k]$ and for some $k \ge 0$.

For any $j \in [0,k]$, the condition $F_j(\mathbf{x}_{j+1}) \le f(\mathbf{x}_j)$ in (C1) or (C2) implies that

$$\mu_j\|\mathbf{x}_{j+1} - \mathbf{x}_j\|^2 \le f(\mathbf{x}_j) - f(\mathbf{x}_{j+1}) \le f(\mathbf{x}_j) - f(\bar{\mathbf{x}}_j). \tag{4.36}$$

By the induction hypothesis, $\mathbf{x}_j \in \mathcal{B}_{\rho/2}(\hat{\mathbf{x}})$, and by the triangle inequality, we have

$$\|\bar{\mathbf{x}}_j - \hat{\mathbf{x}}\| \le \|\bar{\mathbf{x}}_j - \mathbf{x}_j\| + \|\mathbf{x}_j - \hat{\mathbf{x}}\| \le \|\bar{\mathbf{x}}_0 - \mathbf{x}_0\| + \frac{\rho}{2} \le \rho. \tag{4.37}$$

Hence, $\bar{\mathbf{x}}_j \in \mathcal{B}_\rho(\hat{\mathbf{x}})$. We expand f in (4.36) in a Taylor series around $\bar{\mathbf{x}}_j$ and use the fact that $f'[\bar{\mathbf{x}}_j] = \mathbf{0}$ to obtain

$$\mu_j\|\mathbf{x}_{j+1} - \mathbf{x}_j\|^2 \le \frac{L_1}{2}\|\mathbf{x}_j - \bar{\mathbf{x}}_j\|^2, \tag{4.38}$$

where L_1 is a Lipschitz constant for f'. Since f' provides a local error bound with constants α and $\rho/2$ and $\mathbf{x}_j \in \mathcal{B}_{\rho/2}(\hat{\mathbf{x}})$, it follows that

$$\mu_j = \beta\|f'[\mathbf{x}_j]\|^\eta \geq \beta\alpha^\eta\|\mathbf{x}_j - \bar{\mathbf{x}}_j\|^\eta.$$

Combining this with (4.38) gives

$$\|\mathbf{x}_{j+1} - \mathbf{x}_j\| \leq E\|\mathbf{x}_j - \bar{\mathbf{x}}_j\|^{1-\eta/2} \tag{4.39}$$

for $j \in [0, k]$, where E is defined in (4.32). By the triangle inequality, (4.32), (4.35), and (4.39), we have

$$\begin{aligned}
\|\mathbf{x}_{k+1} - \hat{\mathbf{x}}\| &\leq \|\mathbf{x}_0 - \hat{\mathbf{x}}\| + \sum_{j=0}^{k} \|\mathbf{x}_{j+1} - \mathbf{x}_j\| \\
&\leq \|\mathbf{x}_0 - \hat{\mathbf{x}}\| + E\sum_{j=0}^{k} \|\mathbf{x}_j - \bar{\mathbf{x}}_j\|^{1-\eta/2} \\
&\leq \|\mathbf{x}_0 - \hat{\mathbf{x}}\| + E\|\mathbf{x}_0 - \hat{\mathbf{x}}\|^{1-\eta/2}\sum_{j=0}^{k} (\gamma^{1-\eta/2})^j \\
&\leq \|\mathbf{x}_0 - \hat{\mathbf{x}}\| + \frac{E\|\mathbf{x}_0 - \hat{\mathbf{x}}\|^{1-\eta/2}}{1-\gamma^{1-\eta/2}} \\
&\leq \epsilon + \frac{E\epsilon^{1-\eta/2}}{1-\gamma^{1-\eta/2}} \leq \rho/2.
\end{aligned}$$

This establishes the first half of (4.35) for $j = k + 1$.

To establish the second half of (4.35), we will apply Lemma 4.3 to $\mathbf{x} = \mathbf{x}_{k+1}$. Since $\mathbf{x}_{k+1} \in \mathcal{B}_{\rho/2}(\hat{\mathbf{x}})$, we need only show that $\mathbf{x} = \mathbf{x}_{k+1}$ satisfies the qualification $\|\mathbf{x} - \bar{\mathbf{x}}\| \leq C_2\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta$ for (4.23). Since $\|\mathbf{x}_{k+1} - \hat{\mathbf{x}}\| \leq \rho/2$ and since f' provides a local error bound at $\hat{\mathbf{x}}$ with constants α and $\rho/2$,

$$\alpha\|\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}\| \leq \|f'[\mathbf{x}_{k+1}]\| \leq \|F_k'(\mathbf{x}_{k+1})\| + \mu_k\|\mathbf{x}_{k+1} - \mathbf{x}_k\|. \tag{4.40}$$

Since f' is Lipschitz continuous over $\mathcal{B}_\rho(\hat{\mathbf{x}})$, it follows from (4.39), (4.40), and the definition of μ_k that for stopping criterion (C1),

$$\begin{aligned}
\|\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}\| &\leq \frac{\mu_k}{\alpha}(\|\nabla f(\mathbf{x}_k)\| + \|\mathbf{x}_{k+1} - \mathbf{x}_k\|) \\
&\leq \left(\frac{\beta L_1^\eta}{\alpha}(L_1\|\mathbf{x}_k - \bar{\mathbf{x}}_k\| + \|\mathbf{x}_{k+1} - \mathbf{x}_k\|)\right)\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta \\
&\leq \left(\frac{\beta L_1^\eta}{\alpha}(L_1\|\mathbf{x}_0 - \bar{\mathbf{x}}_0\| + E\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^{1-\eta/2})\right)\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta \\
&\leq \left(\frac{\beta L_1^\eta}{\alpha}(L_1\|\mathbf{x}_0 - \hat{\mathbf{x}}\| + E\|\mathbf{x}_0 - \hat{\mathbf{x}}\|^{1-\eta/2})\right)\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta \\
&\leq C_2\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta. \qquad (4.41)
\end{aligned}$$

The first inequality is due to (4.40) and (C1); the second inequality is based on the definition of μ_k and the Lipschitz continuity of f'; the third inequality utilizes the induction hypothesis (4.35), the bound (4.39) for $j = k$, and the fact that $\mathbf{x}_k$ and

$\bar{\mathbf{x}}_k \in \mathcal{B}_\rho(\hat{\mathbf{x}})$ (see (4.37)); and the fourth inequality is a consequence of the induction hypothesis and the fact that ϵ satisfies (4.33).

If stopping criterion (C2) is used, then in the same fashion, (4.34) gives

$$\begin{aligned}
\|\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}\| &\le \frac{\mu_k}{\alpha}(1+\theta)\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \\
&\le \left(\frac{\beta L_1^\eta}{\alpha}(1+\theta)\|\mathbf{x}_{k+1} - \mathbf{x}_k\|\right)\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta \\
&\le \left(\frac{\beta L_1^\eta}{\alpha}(1+\theta)E\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^{1-\eta/2}\right)\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta \\
&\le \left(\frac{\beta L_1^\eta}{\alpha}(1+\theta)E\|\mathbf{x}_0 - \hat{\mathbf{x}}\|^{1-\eta/2}\right)\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta \\
&\le C_2\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^\eta. \qquad (4.42)
\end{aligned}$$

Thus, if either stopping criterion (C1) or (C2) is used, then $\mathbf{x} = \mathbf{x}_{k+1}$ satisfies the qualifications of Lemma 4.3.

We now give another bound for the change $\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$. Lemma 4.3 yields

$$F_k(\mathbf{x}_{k+1}) \le F_k(\bar{\mathbf{x}}_k) + \frac{C_1}{\mu_k}\|F_k'(\mathbf{x}_{k+1})\|^2. \qquad (4.43)$$

Since $f(\bar{\mathbf{x}}_k) \le f(\mathbf{x}_{k+1})$, we conclude that

$$\frac{\mu_k}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \le \frac{\mu_k}{2}\|\bar{\mathbf{x}}_k - \mathbf{x}_k\|^2 + \frac{C_1}{\mu_k}\|F_k'(\mathbf{x}_{k+1})\|^2. \qquad (4.44)$$

If (C1) is used, then we have

$$\begin{aligned}
\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 &\le \|\bar{\mathbf{x}}_k - \mathbf{x}_k\|^2 + 2C_1\|f'[\mathbf{x}_k]\|^2 \\
&\le \|\bar{\mathbf{x}}_k - \mathbf{x}_k\|^2 + 2C_1L_1^2\|\bar{\mathbf{x}}_k - \mathbf{x}_k\|^2.
\end{aligned}$$

If (C2) is used, then (4.44) gives

$$\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \le \frac{1}{1 - 2C_1\theta^2}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^2.$$

In either case,

$$\|\mathbf{x}_{k+1} - \mathbf{x}_k\| \le \tau\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|, \qquad (4.45)$$

where τ is defined at the start of the proof.

If (C1) is used, then

$$\|F_k'(\mathbf{x}_{k+1})\| \le \mu_k\|f'[\mathbf{x}_k]\| = \mu_k\|f'[\mathbf{x}_k] - f'[\bar{\mathbf{x}}_k]\| \le \mu_k L_1\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|.$$

By (4.40) and (4.45), we have

$$\|\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}\| \le \frac{\mu_k}{\alpha}(L_1 + \tau)\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|. \qquad (4.46)$$

If (C2) is used, then $\|F_k'(\mathbf{x}_{k+1})\| \le \theta\mu_k\|\mathbf{x}_{k+1} - \mathbf{x}_k\|$, and by (4.40) and (4.45), we have

$$\|\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}\| \le \frac{\tau\mu_k(1+\theta)}{\alpha}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|. \qquad (4.47)$$

In either case, since $\mathbf{x}_k \in \mathcal{B}_{\rho/2}(\hat{\mathbf{x}})$, it follows from the definition of γ and μ_k that

$$\|\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}\| \le \gamma \|\mathbf{x}_k - \bar{\mathbf{x}}_k\|.$$

This establishes the second half of (4.35) for $j = k + 1$. Hence, the proof of the induction step is complete. The proof that the $\mathbf{x}_k$ form a Cauchy sequence converging to a limit $\mathbf{x}^* \in \mathbf{X}$ is exactly as in Theorem 3.2. The first half of (4.30) follows from (4.46) and (4.47). In the second half of (4.30), we replace μ_k by $\beta\|f'[\mathbf{x}_k]\|^\eta$ and exploit the Lipschitz continuity of f'. ◻

Remark. During the proof of Theorem 4.4, in (4.41) and (4.42), we show that each iterate $\mathbf{x}_{k+1}$ lies in the region where F_k is convex (4.26).

When $\eta = 0$, we can drop the requirement of Theorem 4.4 that $\mathbf{X}$ is locally convex. This convexity requirement arose since we use Lemma 4.3, which assumes that $\mathbf{X}$ is locally convex. We now show that Lemma 4.3 can be established without local convexity when μ_k is bounded away from 0.

LEMMA 4.5. *Let $\beta > 0$ and suppose that $\mu_k \ge \beta$ for each k. Given $\hat{\mathbf{x}} \in \mathbf{X}$ and $\delta \in (0, 1)$, suppose f is twice Lipschitz continuously Fréchet differentiable on $\mathcal{B}_\rho(\hat{\mathbf{x}})$, let L_2 be a Lipschitz constant for f'' on $\mathcal{B}_\rho(\hat{\mathbf{x}})$, and let $r = \min\{\rho, \delta\beta/L_2\}$. Then we have*

$$F_k(\mathbf{x}) - F_k^* \le \left(\frac{1}{(1-\delta)\mu_k}\right) \|F_k'(\mathbf{x})\|^2 \tag{4.48}$$

for all $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$, where

$$F_k^* = \min_{\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})} F_k(\mathbf{x}).$$

Proof. Suppose $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$. We start with the identity

$$\langle F_k''[\mathbf{x}]\mathbf{y}, \mathbf{y}\rangle = \langle F_k''[\hat{\mathbf{x}}]\mathbf{y}, \mathbf{y}\rangle + \langle (F_k''[\mathbf{x}] - F_k''[\hat{\mathbf{x}}])\mathbf{y}, \mathbf{y}\rangle.$$

By the second-order optimality condition, $f''[\hat{\mathbf{x}}]$ is positive. Consequently, we have

$$\langle F_k''[\hat{\mathbf{x}}]\mathbf{y}, \mathbf{y}\rangle \ge \mu_k \|\mathbf{y}\|^2.$$

Since $F_k''[\mathbf{x}] - F_k''[\hat{\mathbf{x}}] = f''[\mathbf{x}] - f''[\hat{\mathbf{x}}]$, it follows from the Lipschitz continuity of f'' that

$$\langle (F_k''[\mathbf{x}] - F_k''[\hat{\mathbf{x}}])\mathbf{y}, \mathbf{y}\rangle \le L_2 \|\mathbf{x} - \hat{\mathbf{x}}\| \, \|\mathbf{y}\|^2 \le \delta\beta \|\mathbf{y}\|^2 \le \delta\mu_k \|\mathbf{y}\|^2$$

since $r \le \delta\beta/L_2$. Hence, if $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$, we have

$$\langle F_k''[\mathbf{x}]\mathbf{y}, \mathbf{y}\rangle \ge (1-\delta)\mu_k \|\mathbf{y}\|^2. \tag{4.49}$$

By (4.49), F_k is convex on $\mathcal{B}_r(\hat{\mathbf{x}})$. Consequently, the minimizer $\mathbf{x}_k^*$ over $\mathcal{B}_r(\hat{\mathbf{x}})$ exists:

$$\mathbf{x}_k^* = \arg\min\{F_k(\mathbf{x}) : \mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})\}.$$

Since $\mathcal{B}_r(\hat{\mathbf{x}})$ is a convex set, the first-order optimality condition

$$\langle F_k'[\mathbf{x}_k^*], \mathbf{x} - \mathbf{x}_k^*\rangle \ge 0$$

holds for all $\mathbf{x} \in \mathcal{B}_r(\hat{\mathbf{x}})$. It follows that

$$\langle F_k'[\mathbf{x}], \mathbf{x} - \mathbf{x}_k^* \rangle \geq \langle F_k'[\mathbf{x}] - F_k'[\mathbf{x}_k^*], \mathbf{x} - \mathbf{x}_k^* \rangle.$$

We utilize the strong convexity property (4.49) to obtain

$$\langle F_k'[\mathbf{x}], \mathbf{x} - \mathbf{x}_k^* \rangle \geq (1 - \delta)\mu_k \|\mathbf{x} - \mathbf{x}_k^*\|^2,$$

which gives

$$\|\mathbf{x} - \mathbf{x}_k^*\| \leq \frac{1}{(1-\delta)\mu_k} \|F_k'[\mathbf{x}]\|. \tag{4.50}$$

The convexity of F_k on $\mathcal{B}_r(\hat{\mathbf{x}})$ implies that

$$F_k(\mathbf{x}^*) \geq F_k(\mathbf{x}) + F_k'[\mathbf{x}](\mathbf{x}_k^* - \mathbf{x}).$$

Combining this with (4.50) completes the proof. □

Theorem 4.4 holds with the following modifications: (i) The assumption (B0) that $\mathcal{B}_\rho(\hat{\mathbf{x}}) \cap \mathbf{X}$ is convex is dropped; and (ii) $\mu_k \in [\beta_0, \beta_1]$, where β_1 is chosen small enough that the constant $\gamma = \beta_2\Lambda/\alpha$ is less than 1. For completeness, we state the modified result.

THEOREM 4.6. *Assume that the following conditions are satisfied:*

(b0) *f provides a local error bound at $\hat{\mathbf{x}} \in \mathbf{X}$ with positive scalars α and ρ satisfying* (1.3).

(b1) *f is twice Lipschitz continuously Fréchet differentiable in $\mathcal{B}_\rho(\hat{\mathbf{x}})$.*

(b2) *The proximal iterates $\mathbf{x}_k$ satisfy either* (C1) *or* (C2) *with $\mu_k \in [\beta_0, \beta_1]$, where $\beta_0 > 0$. If* (C2) *is used, then $\theta < 1/\sqrt{2}$. $\delta \in (0, 1)$ is chosen small enough that $\theta^2 < (1 - \delta)/2$.*

(b3) *Define the parameters*

$$\Lambda = L_1 + \tau \quad \textit{and} \quad \tau^2 = 1 + 2L_1^2/(1-\delta) \ \textit{ if acceptance criterion (C1) is used,}$$

while

$$\Lambda = \tau(1+\theta) \quad \textit{and} \quad \tau^2 = \frac{1}{1 - 2\theta^2/(1-\delta)} \ \textit{ if acceptance criterion (C2) is used,}$$

where L_1 is a Lipschitz constant for f' over $\mathcal{B}_\rho(\hat{\mathbf{x}})$. β_1 is small enough that

$$\gamma := \frac{\beta_1 \Lambda}{\alpha} < 1.$$

Then for ϵ sufficiently small and for each $\mathbf{x}_0 \in \mathcal{B}_\epsilon(\hat{\mathbf{x}})$, the proximal iterates $\mathbf{x}_k$ are all contained in $\mathcal{B}_\rho(\hat{\mathbf{x}})$, and they approach a minimizer $\mathbf{x}^* \in \mathbf{X}$; moreover, for each k, we have

$$\|\mathbf{x}_k - \mathbf{x}^*\| \leq c_1 \gamma^k D(\mathbf{x}_0, \mathbf{X}),$$

where c_1 is a constant independent of k.

The proof of Theorem 4.6 is the same as the proof of Theorem 4.4, except that we use Lemma 4.5 instead of Lemma 4.3 and we replace expressions like $\|\mathbf{x} - \bar{\mathbf{x}}\|$ by $D(\mathbf{x}, \mathbf{X})$.

5. Final discussion. The local convergence results obtained in [7] for a new class of self-adaptive proximal point methods have been extended from a finite dimensional setting to a Hilbert space setting. In particular the local convergence estimates obtained for exact iterates in [7] are now established for approximate iterates satisfying (C0). Our analysis, which permits multiple minimizers, employs a local error bound condition (1.3) relating the growth in f to the distance from the set of minimizers. The gradient-based acceptance criteria in [7] have been replaced by subdifferential-based criteria (C1) and (C2). The local convergence results for the subdifferential-based stopping criteria are similar to the convergence results for iterates satisfying (C0). Three types of assumptions were considered in our analysis connected with (C1) and (C2): (a) f is convex and lower semicontinuous on a level set; (b) the set $\mathbf{X} \cap \mathcal{B}_\rho(\hat{\mathbf{x}})$ is convex for some $\rho > 0$, and f is twice continuously differentiable on $\mathcal{B}_\rho(\hat{\mathbf{x}})$; and (c) $\mu_k \in [\beta_0, \beta_1]$, with β_1 sufficiently small, $\beta_0 > 0$, and f twice continuously differentiable on $\mathcal{B}_\rho(\hat{\mathbf{x}})$. The conditions (b) and (c) are weaker than the local convexity requirement for f in [7].

The analysis in this paper has focused on local convergence. Global convergence issues are studied in section 6 of [7], where we also present computational results which show that for a class of ill-conditioned nonlinear optimization problem, a proximal point approach could reduce the computing time.

REFERENCES

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

[2] P. L. Combettes and T. Pennanen, *Proximal methods for cohypomonotone operators*, SIAM J. Control Optim., 43 (2004), pp. 731–742.

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

[4] J. Eckstein, *Approximate iterations in Bregman-function-based proximal algorithms*, Math. Programming, 83 (1998), pp. 113–123.

[5] J. Fan and Y. Yuan, *On the convergence of the Levenberg-Marquardt method without nonsingularity assumption*, Computing, 74 (2005), pp. 23–39.

[6] C. D. Ha, *A generalization of the proximal point algorithm*, SIAM J. Control Optim., 28 (1990), pp. 503–512.

[7] W. W. Hager and H. Zhang, *Self-adaptive inexact proximal point methods*, Comput. Optim. Appl., to appear.

[8] C. Humes and P. Silva, *Inexact proximal point algorithms and descent methods in optimization*, Optim. Eng., 6 (2005), pp. 257–271.

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

[10] 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.

[11] A. Kaplan and R. Tichatschke, *Stable Methods for Ill-Posed Variational Problems*, Akademie Verlag, Berlin, 1994.

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

[13] D. Li, M. Fukushima, L. Qi, and N. Yamashita, *Regularized Newton methods for convex minimization problems with singular solutions*, Comput. Optim. Appl., 28 (2004), pp. 131–147.

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

[15] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, 4 (1970), pp. 154–158.

[16] B. Martinet, *Determination approachée d'un point fixe d'une application pseudo-contractante*, C. R. Acad. Sci. Paris Sér. A-B, 274 (1972), pp. 163–165.

[17] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation* I, Springer-Verlag,

Berlin, 2006.
[18] J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.
[19] T. Pennanen, *Local convergence of the proximal point algorithm and multiplier methods without monotonicity*, Math. Oper. Res., 27 (2002), pp. 170–191.
[20] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 2 (1976), pp. 97–116.
[21] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.
[22] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
[23] W. Rudin, *Functional Analysis*, McGraw–Hill, New York, 1973.
[24] M. V. Solodov and B. F. Svaiter, *A hybrid extragradient-proximal point algorithm using the enlargement of a maximal monotone operator*, Set-Valued Anal., 7 (1999), pp. 323–345.
[25] M. V. Solodov and B. F. Svaiter, *A hybrid projection-proximal point algorithm*, J. Convex Anal., 6 (1999), pp. 59–70.
[26] M. V. Solodov and B. F. Svaiter, *A unified framework for some inexact proximal point algorithms*, Numer. Funct. Anal. Optim., 22 (2001), pp. 1013–1035.
[27] J. E. Spingarn, *Submonotone mappings and the proximal point algorithm*, Numer. Funct. Anal. Optim., 4 (1981), pp. 123–150.
[28] P. Tseng, *Error bounds and superlinear convergence analysis of some Newton-type methods in optimization*, in Nonlinear Optimization and Related Topics, G. Di Pillo and F. Giannessi, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000, pp. 445–462.
[29] N. Yamashita and M. Fukushima, *The proximal point algorithm with genuine superlinear convergence for the monotone complementarity problem*, SIAM J. Optim., 11 (2000), pp. 364–379.
[30] N. Yamashita and M. Fukushima, *On the rate of convergence of the Levenberg-Marquardt method*, in Topics in Numerical Analysis, Comput. Suppl. 15, Springer-Verlag, Vienna, 2001, pp. 239–249.