Mathematics of Operations Research

Generalized Monotonicity and the Proximal Point Algorithm

D. Russell Luke, Matthew K. Tam

To cite this article:
D. Russell Luke, Matthew K. Tam (2025) Generalized Monotonicity and the Proximal Point Algorithm. Mathematics of Operations Research

Published online in Articles in Advance 21 Nov 2025

. https://doi.org/10.1287/moor.2025.0863

MATHEMATICS OF OPERATIONS RESEARCH
Articles in Advance, pp. 1–13
ISSN 0364-765X (print), ISSN 1526-5471 (online)
https://pubsonline.informs.org/journal/moor

Generalized Monotonicity and the Proximal Point Algorithm

D. Russell Luke,[a,*] Matthew K. Tam[b]

[a] Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, 370873 Göttingen, Germany; [b] School of Mathematics & Statistics, The University of Melbourne, Parkville, Victoria 3010, Australia

*Corresponding author

Contact: r.luke@math.uni-goettingen.de, https://orcid.org/0000-0002-4508-7360 (DRL); matthew.tam@unimelb.edu.au, https://orcid.org/0000-0002-3654-6553 (MKT)

Received: January 6, 2025
Revised: July 18, 2025
Accepted: August 31, 2025
Published Online in Articles in Advance: November 21, 2025

MSC2020 Subject Classifications: Primary: 90C30, 49J52; secondary: 90C56, 49J53

https://doi.org/10.1287/moor.2025.0863

Abstract. We study the proximal point algorithm when the operator of interest is metrically subregular and satisfies a submonotonicity property. The latter property can be viewed as a quantified weakening of the standard definition of a monotone operator. Our main result gives a condition under which locally, the proximal point algorithm generates sequences that are linearly convergent to a zero of the underlying operator. General properties of our notion of submonotonicity are also explored as well as connections to other concepts in the literature.

Funding: D. R. Luke was supported in part by the Deutsche Forschungsgemeinschaft [Grants 541767835 and SFB1456/1-432680300]. M. K. Tam was supported in part by the Australian Research Council [Grant DP230101749].

Keywords: generalized monotonicity • proximal point algorithm • metric subregularity • almost α-firmly nonexpansive • submonotone

1. Introduction

We consider the problem of finding zeros of possibly set-valued mappings in a Euclidean space, focusing in particular on the generalized implicit Euler method, otherwise known as the *proximal point algorithm* (Martinet [28], Rockafellar [35]). This algorithm can be interpreted as the fixed point iteration corresponding to the *resolvent* of the mapping. Minty [29] showed that when the original mapping has full domain and is *maximal monotone*, its resolvent is *firmly nonexpansive*.

In the absence of monotonicity, the proximal point algorithm has been studied for the family of so-called *hypomonotone* mappings and their variants by Combettes and Pennanen [9], Iusem et al. [18], and Pennanen [31] (see also Burachik and Iusem [7, chapter 6.9]). Roughly speaking, such mappings can be made "locally" monotone through the addition of a regularization term. Here, one of the key insights is that hypomonotonicity of the inverse is equivalent to monotonicity of the *Yosida regularization* of the original operator, a correspondence that allows for a great deal of structural properties of the original (nonmonotone) operator to be deduced and upon which the convergence analysis relies. The same idea can be found in the control literature under the name of solution mappings of *semiconvex functions* (Lasry and Lions [21], p. 259). In the contemporary continuous optimization literature, these functions are called *weakly convex functions*, and they have been used to achieve complexity estimates for first-order methods for weakly convex problems.

Following a different direction that avoids notions of monotonicity completely, Artacho et al. [2] showed that either *metric regularity* or *strong metric subregularity* alone suffices to prove that locally, the proximal point algorithm generates at least one convergent sequence that does so with linear rate (and moreover, that for appropriately chosen algorithm parameters, the sequence can be made to converge superlinearly) without the need for monotonicity or nonexpansivity of surrogates. Their techniques, however, seem not to apply to maps for which only *metric subregularity* as opposed to strong metric subregularity holds.

Our main contribution is to provide conditions that together with metric subregularity, guarantee local linear convergence of any sequence appropriately selected from the set-valued proximal point mapping. Our results bridge the gap between the two aforementioned approaches to analyzing the proximal point algorithm with nonmonotonity in the sense that the underlying mapping satisfies two properties, each related to the two different approaches: the first being a weaker condition than hypomonotonicity called *submonotonicity* here and the second being metric subregularity. Our main result, Theorem 2, considers metrically subregular operators that satisfy this submonotonicity property, which is shown to be equivalent to operators whose resolvents are *almost α-firmly nonexpansive* (3) (first proposed in Luke et al. [26]). This is a detailed and improved version of results sketched in the extended abstract by Luke and Tam [24]. Aside from the more in-depth treatment, Theorem 2 improves upon its analog in Luke and Tam [24], where only the existence of convergent sequences was established; in Theorem 2, we provide sufficient conditions for convergence of *any* sequence from the set-valued proximal point mapping.

In Section 2, we set notation and collect various preliminary results for use in the proof of the main result. Section 3 focuses on the study of our new generalized monotonicity property, which we call *submonotonicity*.[1] This machinery is then used in Section 4 to analyze the proximal point algorithm.

2. Preliminaries

The setting is restricted to a Euclidean space denoted by $\mathbb{E}$. The central problem is that of finding a zero of the multivalued mapping $F: \mathbb{E} \rightrightarrows \mathbb{E}$.

$$\text{Find } x \in \mathbb{E} \text{ such that } 0 \in F(x). \tag{1}$$

For brevity "such that" is henceforth abbreviated by "s.t.". Given $\lambda > 0$, the *resolvent* of F is the multivalued map $J_{\lambda F} := (\mathrm{Id} + \lambda F)^{-1}$. A fundamental numerical method to solve (1) and the focus of this study is the multivalued generalization of the implicit Euler method: the *proximal point algorithm*; given an initial point $\overline{x} \in \mathbb{E}$, choose a sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ of positive real numbers and a sequence $\{x^n\}_{n\in\mathbb{N}}$ such that

$$x^{n+1} \in J_{\lambda_n F}(x^n) \ \ \forall n \in \mathbb{N}. \tag{2}$$

When F is *maximal monotone*, the resolvent operators $\{J_{\lambda_n F}\}_{n\in\mathbb{N}}$ are single valued and *firmly nonexpansive* with full domain; that is, they satisfy

$$\|x - J_{\lambda_n F}(\overline{x})\|^2 \le \|x - \overline{x}\|^2 - \|(J_{\lambda_n F}(x) - x) - (J_{\lambda_n F}(\overline{x}) - \overline{x})\|^2 \quad \forall x, \overline{x} \in \mathbb{E}.$$

In this case, for any initial point $x^0 \in \mathbb{E}$ and choice of positive sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ bounded away from zero, there exists a unique sequence of proximal point iterates $\{x^n\}_{n\in\mathbb{N}}$ with $x^{n+1} = J_{\lambda_n F}(x^n)$, and moreover, the sequence converges whenever $F^{-1}(0) \neq \emptyset$ (Rolewicz [37, theorem 1]). It is worth emphasizing that without maximal monotonicity of F, the sequence generated by the proximal point algorithm need not even exist, let alone converge. Linear convergence of the *iterates* requires an additional property to firm nonexpansiveness, namely *metric subregularity* for set-valued maps.

Definition 1 (Metric Subregularity). A set-valued mapping $\Phi: \mathbb{E} \rightrightarrows \mathbb{E}$ is *metrically subregular* on $U \subset \mathbb{E}$ for $\bar{y} \in \mathbb{E}$ relative to $\Lambda \subset \mathbb{E}$ if there exists $\rho > 0$ such that

$$\mathrm{dist}(x, \Phi^{-1}(\bar{y}) \cap \Lambda) \le \rho\, \mathrm{dist}(\bar{y}, \Phi(x)) \quad \forall x \in U \cap \Lambda.$$

What we are calling metric subregularity is referred to as *linear* metric subregularity in Luke et al. [26] as a special case of *gauge* metric subregularity.

Convergence rates for mappings that are only *almost* firmly nonexpansive were established in Luke et al. [26]. On a closed subset $D \subset \mathbb{E}$, a general self-mapping $T: D \rightrightarrows D$ is said to be *point-wise almost α-firmly nonexpansive at* $\overline{x} \in D$ *on* D, abbreviated *point-wise aα-fne*, whenever $\alpha \in (0,1)$ and there exists $\epsilon \in [0,1]$ such that

$$(\forall x \in D)(\forall x^+ \in Tx)(\forall \overline{x}^+ \in T\overline{x}) : \|x^+ - \overline{x}^+\|^2 \le (1+\epsilon)\|x - \overline{x}\|^2 - \frac{1-\alpha}{\alpha}\|(x - x^+) - (\overline{x} - \overline{x}^+)\|^2. \tag{3}$$

When the above inequality holds for all $\overline{x} \in D$, then T is said to be *almost α-firmly nonexpansive (aα-fne) on D*. The *violation* is a value of ϵ for which (3) holds. The qualifier "almost" refers to the fact that the violation is positive but no greater than one. When the violation is zero, the qualifier "almost" is dropped, and the abbreviation *α-fne* is used. It is immediate from the definition that if T is point-wise aα-fne at y, then it is single valued at y. Note, however, that this does not imply that the mapping T is even continuous on a neighborhood of y. The projector in $\mathbb{R}^2$ onto the set $(\mathbb{R} \times \{0\}) \cup (\{0\} \times \mathbb{R})$, for instance, is single valued and point-wise α-fne at $(0,0)$ with $\alpha = 1/2$, but it is not continuous on any ball around the origin.

The definition of point-wise aα-fne mappings in Euclidean spaces appeared first in this form in Luke et al. [26] as a tool in the analysis of splitting algorithms in nonconvex and nonsmooth optimization. In normed vector spaces, these mappings, without violation, were first called *averaged* mappings (Baillon et al. [3], Bruck and Reich [6], Edelstein [12], Krasnoselski [19], Mann [27]). These notions have been extended to nonlinear spaces (Bërdëllima et al. [5]) without the addition operation, hence the change in terminology.

Metric subregularity and the point-wise aα-fne property are two of three central regularity properties in the convergence theory for fixed point iterations. The assumptions are on the behaviors of the mapping T at its fixed points, which for a multivalued mapping, are defined by $\operatorname{Fix} T := \{x \mid x \in Tx\}$.

Assumption 1 (Regularity Assumptions I). *Let $D \subset \mathbb{E}$. The following assumptions hold.*

a. (*Existence*) *The set* $\operatorname{Fix} T \cap D$ *is nonempty and closed.*

b. (*Self-mapping*) *The mapping T is a self-mapping on D: that is, $T : D \rightrightarrows D$.*

c. (*Almost quasicontractivity*) *The mapping T is point-wise aα-fne at all $y \in \operatorname{Fix} T \cap D$ with constant $\alpha \in (0,1)$ and violation $\epsilon \in [0,1]$:*

$$(\forall x \in D)(\forall x^+ \in Tx)(\forall y \in \operatorname{Fix} T \cap D) : \|x^+ - y\|^2 \le (1+\epsilon)\|x-y\|^2 - \frac{1-\alpha}{\alpha}\|x - x^+\|^2. \tag{4}$$

d. (*Stability*) *The mapping* $\operatorname{Id} - T$ *is metrically subregular on D relative to D for zero with constant $\rho > 0$:*

$$d(x, \operatorname{Fix} T \cap D) \le \rho d((\operatorname{Id} - T)x, 0) = \inf_{x^+ \in Tx} \rho \|x - x^+\|. \tag{5}$$

The next result shows the interplay between these properties.

Proposition 1 (Luke et al. [26, Corollary 2.3]). *Under Assumption 1, for any $x^0 \in D$, any fixed point sequence $x^{n+1} \in Tx^n$ satisfies*

$$d(x^{n+1}, \operatorname{Fix} T \cap D) \le \theta_{\alpha,\epsilon} d(x^n, \operatorname{Fix} T \cap D) \quad \forall n \in \mathbb{N}, \tag{6}$$

where

$$\theta_{\alpha,\epsilon} = \left((1+\epsilon) - \frac{1-\alpha}{\rho^2\alpha}\right)^{1/2} \tag{7}$$

for the constants α, ϵ, and ρ as in Assumption 1. If, in addition,

$$\sqrt{\frac{1-\alpha}{(1+\epsilon)\alpha}} \le \rho \le \sqrt{\frac{1-\alpha}{\epsilon\alpha}}, \tag{8}$$

then $\theta_{\alpha,\epsilon} \in [0,1)$ and $x^n \to \overline{x} \in \operatorname{Fix} T \cap D$ as $n \to \infty$ R linearly with rate $\theta_{\alpha,\epsilon}$.

It is worth noting that metric subregularity in Assumption 1(d) is a central requirement with connections to many other notions of stability and error bounds (Dontchev and Rockafellar [11], Ioffe [16], Ioffe [17], Kruger et al. [20]). It was shown in Luke et al. [25, corollary 1] to be *necessary* for linear convergence of fixed point mappings generated by mappings that are point-wise aα-fne at their fixed points. This carries over to fixed point iterations on uniformly convex metric spaces, with curvature bounded above (Lauster and Luke [22, theorem 16(a)]), and to stochastic fixed point iterations of mappings that are α-fne in expectation and have common fixed points (Hermer et al. [15, theorem 3.15]). For the proximal point algorithm under local monotonicity, the crucial role of metric subregularity in achieving linear convergence when starting near enough to a solution has also been detailed in Rockafellar [36]. Iterations of aα-fne mappings with error have been interpreted in Hermer et al. [15] as random function iterations. Here, convergence of randomly selected mappings, each satisfying Assumption 1(b)–(c), is obtained in the space of probability distributions with respect to the Wasserstein metric under the assumption of metric subregularity of the Markov operator with respect to the Wasserstein metric (Hermer et al. [14, theorem 2.6]). We do not delve into this theory here and assume exact evaluation throughout.

The goal of this study is to establish linear convergence guarantees for mappings F whose resolvents are only point-wise aα-fne.

3. Submonotone Mappings

In this section, we introduce a generalized monotonicity property of set-valued maps, which for lack of better terminology, we call *submonotonicity*. This name was given by Spingarn [38] to mappings that are strictly maximal *hypomonotone* (see Definition 3 below); subsequent work (Daniilidis and Georgiev [10]) studied this in relation to *approximate convexity* (Ngai et al. [30]). We repurpose this term for a differently defined object whose properties we investigate, exploring connections to other generalized monotonicity properties in the literature.

Point-wise aα-fne mappings discussed above lead to our notion of submonotone mappings in the following way. If F has a resolvent J_F that is point-wise aα-fne on U at all $y \in S$ with $\alpha = 1/2$ and violation ϵ, then F satisfies (Luke et al. [26, proposition 2.3])

$$(\forall (u,z) \in \operatorname{gph} F \text{ s.t. } u+z \in U)(\forall (v,w) \in \operatorname{gph} F \text{ s.t. } v+w \in S): -\frac{\epsilon}{2}\|(u+z)-(v+w)\|^2 \le \langle u-v,\, z-w\rangle. \tag{9}$$

Conversely, if F satisfies (9), then the resolvent is aα-fne at all $y \in S$ on U with constant $\alpha = 1/2$ and violation ϵ. The celebrated *Minty characterization* shows the correspondence between *(maximal) monotonicity* of an operator and firm nonexpansivity of its resolvent with full domain (for a modern treatment, see Bauschke and Combettes [4]). With this in mind, we define the following.

Definition 2 (Submonotone Mappings). Let U and W be subsets of $\mathbb{E}$, and let $\tau \ge 0$. A mapping $F : \mathbb{E} \rightrightarrows \mathbb{E}$ is said to be *submonotone* on U in W with violation τ if

$$(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap W)(\forall v^+ \in Fv \cap W): -\tau\|(u+u^+)-(v+v^+)\|^2 \le \langle u-v, u^+-v^+\rangle. \tag{10}$$

The mapping F is said to be *maximal submonotone* on U in W with violation τ if $F|_U \cap W = \widetilde{F}|_U \cap W$ for any operator $\widetilde{F} : \mathbb{E} \rightrightarrows \mathbb{E}$, which is submonotone on U in W with violation τ and has $\operatorname{gph} F \subseteq \operatorname{gph} \widetilde{F}$.

In terms of the graph of the operators involved, the definition of maximal submonotonicity of F (on U in W with violation τ) can be expressed as follows. If an operator $\widetilde{F} : \mathbb{E} \rightrightarrows \mathbb{E}$ is submonotone on U in W with violation τ and $\operatorname{gph} F \subseteq \operatorname{gph} \widetilde{F}$, then $\operatorname{gph} F \cap (U \times W) = \operatorname{gph} \widetilde{F} \cap (U \times W)$.

The existence of a maximal extension of a submonotone operator follows from the usual Zorn's lemma argument.

Proposition 2 (Maximal Submonotonicity). *Suppose $F : \mathbb{E} \rightrightarrows \mathbb{E}$ is submonotone on U in W with violation τ. Then, there exists an operator $\widetilde{F}$ with $\operatorname{gph} F \subseteq \operatorname{gph} \widetilde{F}$: that is, maximal submonotone on U in W with violation τ.*

Proof. Define a collection, $\mathcal{M}$, of set-valued operators from $\mathbb{E}$ to $\mathbb{E}$ by

$$\mathcal{M} := \{G : G \text{ is submonotone on } U \text{ in } W \text{ with violation } \tau,\ \operatorname{gph} F \subseteq \operatorname{gph} G\}.$$

This family is nonempty because it contains F and is partially ordered by the relation induced by $\operatorname{gph} G_1 \subseteq \operatorname{gph} G_2$. Now, take a family $\{G_i\}_{i\in I} \subseteq \mathcal{M}$ such that for all $i, j \in I$, either $\operatorname{gph} G_i \subseteq \operatorname{gph} G_j$ or $\operatorname{gph} G_i \supseteq \operatorname{gph} G_j$. Define an operator G according to $\operatorname{gph} G := \cup_{i\in I} \operatorname{gph} G_i$. Then, for any two elements $(x, x^+), (y, y^+) \in \operatorname{gph} G \cap (U \times W)$, there exist indices $i, j \in I$ such that $(x, x^+) \in \operatorname{gph} G_i \cap (U \times W)$ and $(y, y^+) \in \operatorname{gph} G_j \cap (U \times W)$. Because $\{G_i\}_{i\in I}$ is totally ordered, we may assume without loss of generality that $\operatorname{gph} G_j \subseteq \operatorname{gph} G_i$ so that we have both (x, x^+) and (y, y^+) in $\operatorname{gph} G_i \cap (U \times W)$. The fact that G_i is submonotone on U in W with violation τ then implies the same for G. To complete the proof, apply Zorn's lemma to deduce the existence of a maximal element $\widetilde{F} \in \mathcal{M}$. □

The following gives some equivalent forms of submonotonicity.

Proposition 3 (Characterizations of Submonotonicity). *Let U and W be nonempty subsets of $\mathbb{E}$, let $\tau > 0$, and let $F : \mathbb{E} \rightrightarrows \mathbb{E}$. The following assertions are equivalent.*

a. *The mapping F is submonotone on U in W with violation τ.*

b. $(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap W)(\forall v^+ \in Fv \cap W)$:
$$\|u-v\|^2 + \|u^+-v^+\|^2 \le (1+2\tau)\|(u+u^+)-(v+v^+)\|^2.$$

c. $(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap W)(\forall v^+ \in Fv \cap W)$:
$$-\tau(\|u-v\|^2 + \|u^+-v^+\|^2) \le (1+2\tau)\langle u-v, u^+-v^+\rangle.$$

d. $(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap W)(\forall v^+ \in Fv \cap W)$:
$$0 \le \|u-u^+\|^2 + \|v-v^+\|^2 + \frac{2\tau}{(1+2\tau)}(\|u-v\|^2 + \|u^+-v^+\|^2) - \|v-u^+\|^2 - \|u-v^+\|^2.$$

Proof. The equivalences follow from rearrangements of the definition of submonotonicity. □

As with monotone mappings, there is a close relationship between submonotonicity and Lipschitz continuity of resolvents. In the classical literature, nonexpansiveness is implied by firm nonexpansiveness, so the former appears to be the more fundamental property. However, the correspondence between Lipschitz continuity and the aα-fne property defined in (3) is a feature of the geometry of inner product spaces. In geodesic metric spaces with positive curvature, for instance, aα-fne mappings are well defined (Bërdëllima et al. [5, definition 1]) and key to the convergence of fixed point iterations, but these need not be Lipschitz (see Ariza-Ruiz et al. [1] and the discussion after Bërdëllima et al. [5, proposition 2]). Nevertheless, in the present setting, the correspondence is tight and fully resolved in the following result.

Proposition 4 (Resolvents of Submonotone Mappings). *Let $F: U' \rightrightarrows \mathbb{E}$ for $U' \subset \mathbb{E}$, and for the subsets $U \subset U'$, $W \subset \mathbb{E}$, define*

$$F_W(u) := \begin{cases} F(u) \cap W & \text{if } u \in U \\ \emptyset & \text{else,} \end{cases}$$

and $D := (\mathrm{Id} + F_W)(U)$. The mapping F is submonotone on U in W with violation $\tau < 1/2$ if and only if the resolvent $J_{F_W} := (\mathrm{Id} + F_W)^{-1}$ is aα-fne with constant $\alpha = 1/2$ and violation $\epsilon = 2\tau < 1$ on D. Equivalently, $\tilde{T} := 2J_{F_W} - \mathrm{Id}$ is Lipschitz continuous with constant $(1+4\tau)^{1/2}$ on D.

Proof. *The mapping F* being submonotone on U in W with violation $\tau < 1/2$ is equivalent to

$$(\forall (u,u^+) \in \mathrm{gph}\, F \cap (U \times W) \ \text{ s.t. } \ u + u^+ \in D)$$
$$(\forall (v,v^+) \in \mathrm{gph}\, F \cap (U \times W) \ \text{ s.t. } \ v + v^+ \in D): -\tau \|(u+u^+) - (v+v^+)\|^2 \le \langle u - v,\ u^+ - v^+ \rangle,$$

where $D = U + F_W(U)$. By Luke et al. [26, proposition 2.3(iv)], this is equivalent to the mapping J_{F_W} being pointwise aα-fne (hence, single valued) at all $y \in D$ on D with violation at most $\epsilon = 2\tau$ and constant $\underline{\alpha} = 1/2$. The restriction $\tau \le 1/2$ ensures that the violation stays in the interval [0, 1]. By Luke et al. [26, proposition 2.1(ii)] (with $\alpha = 1/2$), this is equivalent to $\tilde{T} := 2J_{F_W} - \mathrm{Id}$ satisfying

$$(\forall x, y \in D) \qquad \|\tilde{T}(x) - \tilde{T}(y)\| \le \sqrt{1 + 2\epsilon}\|x - y\|.$$

In other words, the mapping $\tilde{T}$ is Lipschitz continuous with constant $\sqrt{1+4\tau}$ on D. □

Proposition 5 (Submonotonicity and Inverses). *Let U and W be nonempty subsets of $\mathbb{E}$, let $\tau > 0$, and let $F: \mathbb{E} \rightrightarrows \mathbb{E}$. Then, F is (maximal) submonotone on U in W with violation τ if and only if F^{-1} is (maximal) submonotone on W in U with violation τ.*

Proof. Observe that by expressing the definition of submonotonicity in terms of the graph of F, we have

$$(\forall (u,u^+) \in \mathrm{gph}\, F \cap (U \times W))(\forall (v,v^+) \in \mathrm{gph}\, F \cap (U \times W)) : -\tau \|(u+u^+) - (v+v^+)\|^2 \le \langle u - v, u^+ - v^+ \rangle.$$

Here, we note that u and u^+ (v and v^+, respectively) can be interchanged without changing the validity of the inequality and furthermore, that

$$(u,u^+) \in \mathrm{gph}\, F \cap (U \times W) \Longleftrightarrow (u^+, u) \in \mathrm{gph}\, F^{-1} \cap (W \times U).$$

Thus, F is submonotone on U in W with violation τ if and only if F^{-1} is submonotone on W in U with violation τ. The preservation of maximality follows from the definition. □

Next, we turn our attention to the structure of the range of a submonotone operator. In order to give a useful description, we recall that an extended real-valued function f is said to be *weakly convex* with constant ρ if the function $f + \rho\|\cdot\|^2$ is convex (Vial [39]). In particular, a weakly convex function with $\rho = 0$ is convex. This same concept is elsewhere called *paraconvexity* (Rolewicz [37]) or *semiconvexity* (Lasry and Lions [21]).

Theorem 1. *Let U and W be subsets of $\mathbb{E}$, let $\tau \ge 0$, and suppose that $F: \mathbb{E} \rightrightarrows \mathbb{E}$ is maximal submonotone on U in W with violation τ. Then, for all $u \in U$, $Fu \cap W$ can be expressed as the intersection of W and the lower-level set of a proper, lower semicontinuous (lsc), weakly convex function with constant τ. Consequently, for every closed subset O of W, the set $Fu \cap O$ is closed. For any $v \in W$, the analogous statement holds for $F^{-1}(v) \cap U$.*

Proof. Let $v \in U$. We claim that

$$Fv \cap W = \bigcap_{u \in U,\ u^+ \in Fu \cap W} \{w \in W : 0 \le \tau \|(u+u^+) - (v+w)\|^2 + \langle u - v, u^+ - w \rangle\}. \tag{11}$$

The inclusion "⊆" follows immediately from the definition submonotonicity of F. For the reverse inclusion, let $w \in W$ be a point such that for all $u \in U$ and $u^+ \in Fu \cap W$,

$$0 \le \tau \|(u+u^+) - (v+w)\|^2 + \langle u - v, u^+ - w\rangle.$$

Define an operator $\widetilde{F} : \operatorname{dom} F \rightrightarrows \mathbb{E}$ by

$$\widetilde{F}u := \begin{cases} Fu & \text{if } u \in \operatorname{dom} F \setminus \{v\}, \\ Fv \cup \{w\} & \text{if } u = v. \end{cases}$$

Then, $\widetilde{F}$ is submonotone on U in W with violation τ and $\operatorname{gph} F \subseteq \operatorname{gph} \widetilde{F}$. But, $w \in \widetilde{F}v \cap W = Fv \cap W$ because F is maximal submonotone on U in W with violation τ. This establishes the "⊇" inclusion and completes the proof of equality.

Now, fix points $u, v \in U$, and $u^+ \in Fu$, and consider the set

$$\Omega := \{w \in \mathbb{E} : 0 \le \tau \|(u+u^+) - (v+w)\|^2 + \langle u - v, u^+ - w\rangle\}.$$

For any $w \in \Omega$, we have

$$\begin{aligned} 0 &\le \tau \|(u+u^+) - (v+w)\|^2 + \langle u - v, u^+ - w\rangle \\ &= \tau(\|u+u^+ - v\|^2 + \|w\|^2 - 2\langle u + u^+ - v, w\rangle) + \langle u - v, u^+\rangle - \langle u - v, w\rangle \\ &= [\tau\|u+u^+-v\|^2 + \langle u - v, u^+\rangle] - \langle (2\tau+1)(u-v) + 2\tau u^+, w\rangle + \tau\|w\|^2, \end{aligned}$$

which by denoting $f := \langle (1+2\tau)(u-v) + 2\tau u^+, \cdot\,\rangle - \tau\|\cdot\|^2$, may be expressed in the form

$$f(w) \le \tau\|u+u^+-v\|^2 + \langle u - v, u^+\rangle.$$

In other words, Ω is a lower-level set of the function f, which is a proper, lsc, weakly convex function with constant 2τ. Because the max of proper, lsc, weakly convex functions with constant τ is also proper, lsc, and weakly convex with the same constant τ,[2] (11) implies that $Fv \cap W$ is equal to the intersection of a lower-level set of a proper, lsc, weakly convex function (with constant τ) and W. Moreover, for any closed subset O of W, $Fv \cap O$ is equal to the intersection of a lower-level set of a proper, lsc, weakly convex function and the closed set O and hence, is closed itself.

The claimed result for the inverse of F follows from the result for F and the fact that by Proposition 5, F^{-1} is submonotone on W in U with violation τ. □

3.1. Relation to Other Notions of Generalized Monotonicity

In this section, we compare our newly introduced submonotonicity property with other weaker-than-monotonicity properties in the literature. The first such property that we discuss is that of *hypomonotonicity*, which has its origins in Poliquin and Rockafellar [33], Poliquin et al. [34], and Spingarn [38].

Definition 3 (Hypomonotonicity). Let U and W be subsets of $\mathbb{E}$ and $\sigma > 0$. A mapping $F : \mathbb{E} \rightrightarrows \mathbb{E}$ is said to be *hypomonotone* in U on W with violation σ if

$$(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap W)(\forall v^+ \in Fv \cap W) : -\sigma\|u - v\|^2 \le \langle u - v, u^+ - v^+\rangle. \tag{12}$$

With regard to Definition 3, observe that by rearranging of (12), it can be seen that hypomonotonicity of F on U for W is equivalent to monotonicity of $F + \sigma\,\mathrm{Id}$ in U for W.

Remark 1 (Monotone Operators). Recall that a set-valued map $F : \mathbb{E} \rightrightarrows \mathbb{E}$ is said to be *monotone* on U for W if

$$(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap W)(\forall v^+ \in Fv \cap W) : 0 \le \langle u - v, u^+ - v^+\rangle. \tag{13}$$

In the case that $U = W = \mathbb{E}$, this is just the usual definition of a monotone operator.

Monotonicity is preserved by taking inverses in the sense that (13) is equivalent to monotonicity of F^{-1} on W in U. As we have already seen in Proposition 5, the analogous statement is true of the inverse of a submonotone operator. It is, however, clear that one cannot expect the same to be true for hypomonotone operators in general. Indeed, the term *co-hypomonotonicity* has been used to refer to an operator whose inverse is hypomonotone.

On the other hand, hypomonotonicity of an operator is preserved under positive scalar multiplication, a property that also holds for monotone operators. More precisely, if F is (hypo-)monotone and $\lambda > 0$, then λF is also (hypo-)monotone. As we shall soon show, the same for this property is generally not satisfied by submonotone mappings.

Proposition 6 (Hypo-/Submonotonicity). *Let U and W be subsets of $\mathbb{E}$, and consider a set-valued mapping $F : \mathbb{E} \rightrightarrows \mathbb{E}$. The following assertions hold.*

a. *If F is hypomonotone on U in W with violation $\sigma \in [0, 1/2)$, then F is submonotone on U in W with violation $\tau := \sigma/(1-2\sigma) \ge 0$, and F^{-1} is submonotone on W in U with violation τ.*

b. *Let F be hypomonotone on U in W with violation $\sigma_1 \ge 0$, and let F^{-1} be hypomonotone on W in U with violation $\sigma_2 \ge 0$ and $\sigma \in [0, 1/2)$, where*

$$\sigma := \begin{cases} \sigma_1\sigma_2/(\sigma_1+\sigma_2) & \sigma_1 \ne 0 \text{ and } \sigma_2 \ne 0, \\ 0 & \text{otherwise}. \end{cases}$$

Then, F is submonotone on U in W with violation $\tau := 2\sigma/(1-\sigma)$.

c. *Let F be submonotone on U in W with violation τ, and suppose there exists a Lipschitz localization with constant $L \ge 0$; that is, F is single valued on U with $F(U) \subset W$ and*

$$(\forall u \in U)(\forall v \in U): \quad \|Fu - Fv\| \le L\|u - v\|. \tag{14}$$

Then, F is hypomonotone on U in W with violation $\sigma := \tau(1+L^2)/(1+2\tau) \ge 0$.

Proof.

a. Suppose that F is hypomonotone on U in W with violation $\sigma \in [0, 1/2)$. That is,

$$(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap W)(\forall v^+ \in Fv \cap W): -\sigma\|u - v\|^2 \le \langle u - v, u^+ - v^+\rangle. \tag{15}$$

Set $\tau := \sigma/(1-2\sigma)$ or equivalently, $\sigma = \tau/(1+2\tau)$. Then, $\tau > 0$ because $\sigma \in [0, 1/2)$. Using (15), we observe that

$$-\tau\left(\|u - v\|^2 + \|u^+ - v^+\|^2\right) \le \tau\|u - v\|^2 \le (1+2\tau)\langle u^+ - v^+, u - v\rangle.$$

The fact that F is submonotone on U in W with violation τ now follows from Proposition 3(c); submonotonicity of F^{-1} on W in U with violation τ follows from Proposition 5.

b. Let $(u, u^+) \in \operatorname{gph} F \cap (U \times W)$ and $(v, v^+) \in \operatorname{gph} F \cap (U \times W)$. By the assumed hypomonotonicity of F, we have

$$-\sigma_1\|u - v\|^2 \le \langle u - v, u^+ - v^+\rangle,$$

and by the assumed hypomonotonicity of F^{-1}, we have

$$-\sigma_2\|u^+ - v^+\|^2 \le \langle u - v, u^+ - v^+\rangle.$$

We distinguish two cases. First, suppose that either $\sigma_1 = 0$ or $\sigma_2 = 0$. Then, either F or F^{-1} is a monotone operator. But, because F is a monotone operator if and only if F^{-1} is a monotone operator, we have $\sigma_1 = \sigma_2 = 0$, from which the result follows. Suppose next that $\sigma_1\sigma_2 \ne 0$. Then, by combining the two previous equations, we deduce that

$$-\frac{\sigma_1\sigma_2}{\sigma_1+\sigma_2}\left(\|u - v\|^2 + \|u^+ - v^+\|^2\right) \le \langle u - v, u^+ - v^+\rangle.$$

The fact that F is submonotone on U in W with violation τ as defined here now follows from Proposition 3(c), and submonotonicity F^{-1} on W in U with violation τ follows from Proposition 5.

c. Suppose that F satisfies the assumed submonotonicity property as well as (14). Then, for all $(u, Fu) \in \operatorname{gph} F \cap (U \times W)$ and $(v, Fv) \in \operatorname{gph} F \cap (U \times W)$, we have

$$-\frac{\tau}{1+2\tau}(1+L^2)\|u - v\|^2 \le -\frac{\tau}{1+2\tau}(\|u - v\|^2 + \|Fu - Fv\|^2) \le \langle Fu - Fv, u - v\rangle.$$

This completes the proof. □

The next two examples show that the conditions of Proposition 6 cannot be weakened in general. In particular, Example 1 shows hypomonotonicity need not imply submonotonicity when the hypomonotonicity violation is greater than 1/2, and Example 2 gives an example of a (single-valued) non-Lipschitzian submonotone map that is not hypomonotone.

Example 1 (Small Hypomonotone Violation Implies Submonotone). Let $\tau > 0$; let $U := [0, \tau]$, $W := \mathbb{R}$; and consider the function $F(x) := -x^2$. Then, for all $(u, u^+) \in \operatorname{gph} F \cap (U \times W)$ and $(v, v^+) \in \operatorname{gph} F \cap (U \times W)$, we have

$$\langle u - v, Fu - Fv\rangle = (u - v)(-u^2 + v^2) = -(u + v)\|u - v\|^2 \ge -2\tau\|u - v\|^2, \tag{16}$$

which shows that F is hypomonotone on U in W with violation 2τ.

We claim, however, that F is submonotone only if $\tau < 1/2$. To see this, note that

$$\|(u+u^+)-(v+v^+)\|^2 = \|(u-v)-(u^2-v^2)\|^2 = (1-(u+v))^2\|u-v\|^2. \tag{17}$$

Whenever $u+v \neq 1$, combining (16) and (17) yields

$$-\frac{(u+v)}{(1-(u+v))^2}\|(u+u^+)-(v+v^+)\|^2 = \langle u-v, u^+-v^+\rangle. \tag{18}$$

On one hand, if $\tau < 1/2$, then $u+v \leq 2\tau$ and $(1-2\tau)^2 \leq (1-(u+v))^2$. We then have that

$$\tau := \frac{2\tau}{(1-2\tau)^2} \geq \frac{(u+v)}{(1-(u+v))^2}$$

and hence, that F is submonotone on U in W with violation τ. Consider next the case $\tau \geq 1/2$, and suppose F is submonotone on U in W with violation $\overline{\tau} \geq 0$. Let $\{u_n\}, \{v_n\}$ be two sequences contained in $[0,1/2)$, which both converge to $1/2$. Then, using (18), we have

$$\overline{\tau} \geq \frac{(u_n+v_n)}{(1-(u_n+v_n))^2} \to +\infty,$$

which contradictions the finiteness of $\overline{\tau}$. It follows that F cannot be submonotone on U for W in this case, proving the claim.

Example 2 (Submonotone But Not Hypomonotone). Let $U := [0,1/16]$; let $W := \mathbb{R}$; and consider the function $F(x) = -\sqrt{x}$ defined for $x \geq 0$, which is not Lipschitz at zero. Then, for all $(u,u^+) \in \operatorname{gph} F \cap (U \times W)$ and $(v,v^+) \in \operatorname{gph} F \cap (U \times W)$ such that $\sqrt{u}+\sqrt{v} \neq 0$, we have

$$\langle u-v, Fu-Fv\rangle = (u-v)(-\sqrt{u}+\sqrt{v}) = -\frac{1}{\sqrt{u}+\sqrt{v}}\|u-v\|^2. \tag{19}$$

Observe that $1/(\sqrt{u}+\sqrt{v}) \to +\infty$ as $u,v \to 0$, and hence, F cannot be hypomonotone on U in W for any violation constant. However, we claim that F is submonotone on U in W with violation $\tau := 2$. To see this, first observe that for $\sqrt{u}+\sqrt{v} \neq 0$, we have

$$\begin{aligned}\|(u+u^+)-(v+v^+)\|^2 &= \|(u-v)-(\sqrt{u}-\sqrt{v})\|^2\\ &= \left\|(u-v)-\frac{1}{\sqrt{u}+\sqrt{v}}(u-v)\right\|^2\\ &= \left(\frac{(\sqrt{u}+\sqrt{v})-1}{\sqrt{u}+\sqrt{v}}\right)^2\|u-v\|^2.\end{aligned} \tag{20}$$

Because $\sqrt{u}+\sqrt{v} \leq 1/2$, it follows that $\sqrt{u}+\sqrt{v} \leq 1/2$ and $1/4 \leq (\sqrt{u}+\sqrt{v}-1)^2$. Consequently, we have

$$\tau := 2 = \frac{1/2}{1/4} \geq \frac{\sqrt{u}+\sqrt{v}}{(\sqrt{u}+\sqrt{v}-1)^2} = \frac{1}{\sqrt{u}+\sqrt{v}}\left(\frac{\sqrt{u}+\sqrt{v}}{\sqrt{u}+\sqrt{v}-1}\right)^2.$$

By combining (19) and (20), whenever $\sqrt{u}+\sqrt{v} \neq 0$, we deduce that

$$-\tau\|(u+Fu)-(v+Fv)\|^2 \leq \langle u-v, Fu-Fv\rangle. \tag{21}$$

Moreover, if $\sqrt{u}+\sqrt{v} = 0$, then $u = v = 0$, and so, (21) remains true in this case. Altogether, this shows that F is submonotone on U in W violation $\tau = 2$ as claimed.

4. The Proximal Point Algorithm

To begin, we establish the implication of metric subregularity of a scaled, multivalued mapping λF for the resolvent residual mapping Φ defined in Proposition 1.

Lemma 1. *Let $F : \mathbb{E} \rightrightarrows \mathbb{E}$, $F^{-1}(0) \cap D \neq \emptyset$ closed for $D \subset \mathbb{E}$, where $J_{\lambda F} : D \to D$ for some $\lambda > 0$. If F is metrically subregular for zero on U relative to D with constant ρ and $U' := \{x \in D \mid J_{\lambda F}(x) \subset U \cap D\} \neq \emptyset$, then $\Phi_\lambda := \mathrm{Id} - J_{\lambda F}$ is metrically subregular for zero on U' relative to D with constant $(\lambda+\rho)/\lambda$.*

Proof. The definition of $J_{\lambda F}$ implies that $x - x^+ \in \lambda F(x^+)$ for any $x^+ \in J_{\lambda F}(x)$ and x. Hence, for all x and any $x^+ \in J_{\lambda F}(x)$ we have

$$\|x - x^+\| \geq \inf_{z \in \lambda F(x^+)} \|z\| = \lambda \left(\inf_{z \in F(x^+)} \|z\| \right) = \lambda \operatorname{dist}(0, F(x^+)). \tag{22}$$

Because $F^{-1}(0) \cap D$ is a closed, nonempty subset of a finite-dimensional space, there exists a point $\overline{x}^+ \in P_{F^{-1}(0)\cap D}(x^+)$ so that

$$\operatorname{dist}(x, F^{-1}(0) \cap D) \leq \|x - \overline{x}^+\| \leq \|x - x^+\| + \operatorname{dist}(x^+, F^{-1}(0) \cap D). \tag{23}$$

Because F is metrically subregular for zero on U relative to D with constant ρ, we have

$$\operatorname{dist}(x^+, F^{-1}(0) \cap D) \leq \rho \operatorname{dist}(0, F(x^+)), \quad \forall x^+ \in U \cap D. \tag{24}$$

By assumption, there exist points x with $J_{\lambda F}(x) \subset U \cap D$ and

$$\begin{aligned}\operatorname{dist}(x, F^{-1}(0) \cap D) &\overset{(24)}{\leq} \|x - x^+\| + \rho \operatorname{dist}(0, F(x^+)) \qquad \forall x \in U', \ \forall x^+ \in J_{\lambda F}(x) \\ &\overset{(22)}{\leq} \|x - x^+\| + \frac{\rho}{\lambda}\|x - x^+\| \\ &= \left(\frac{\lambda + \rho}{\lambda}\right)\|x - x^+\|. \end{aligned} \tag{25}$$

Because (25) holds for all $x^+ \in J_{\lambda F}(x)$, recalling that $F^{-1}(0) = \operatorname{Fix} J_{\lambda F} = \Phi_\lambda^{-1}(0)$, this yields

$$\operatorname{dist}(x, \Phi_\lambda^{-1}(0) \cap D) \leq \left(\frac{\lambda + \rho}{\lambda}\right) \operatorname{dist}(0, \Phi_\lambda(x)), \quad \forall x \in U' \cap D$$

as claimed. □

Remark 2. The assumption $\{x \in D \mid J_{\lambda F}(x) \subset U \cap D\} \neq \emptyset$ is satisfied in particular if U is a neighborhood including $\operatorname{Fix} J_{\lambda F}$, but of course, the statement above is only interesting for points that are not fixed points.

The next result is a generalization of the classical property of convergence for *Fejér monotone sequences*. A sequence of points $\{x^n\}_{n\in\mathbb{N}}$ is said to be *linearly monotone with respect to S with rate* $\kappa \in [0, 1]$ if

$$(\forall n \in \mathbb{N}) \quad d(x^{n+1}, S) \leq \kappa d(x^n, S). \tag{26}$$

This was introduced in Luke et al. [25] for more general gauges. Fejér monotone sequences, in contrast, satisfy

$$d(x^{n+1}, x) \leq d(x^n, x) \quad \forall x \in S, \ \forall n \in \mathbb{N}.$$

It is easy to see that any Fejér monotone sequence is linearly monotone, but the converse is not true (see Luke et al. [25, example 1]).

Lemma 2 (Convergence of Linearly Monotone Sequences). *Let $\{x^n\}_{n\in\mathbb{N}}$ be a sequence on $\mathbb{E}$. Suppose that for some closed subset $S \subset \mathbb{E}$ and some $\delta > 0$, we have that*

a. $\|x^{n+1} - x^n\| \leq \delta \operatorname{dist}(x^n, S)$ *for all $n \in \mathbb{N}$ and*
b. $\{x^n\}_{n\in\mathbb{N}}$ *is linearly monotone relative to S with rate $\kappa < 1$.*

Then, $\{x^n\}$ converges gauge monotonically to a point $\overline{x} \in S$ with rate $O\left(\frac{\kappa^n}{1-\kappa}\right)$.

Proof. The first assumption and iterative application of linear monotonicity yield

$$\|x^{n+1} - x^n\| \leq \delta \kappa^n \operatorname{dist}(x^0, S), \quad \forall n \in \mathbb{N}.$$

Let $t_0 = \operatorname{dist}(x^0, S)$. For any natural numbers n, l with $n < l$, the triangle inequality yields the upper estimate

$$\begin{aligned}\|x^n - x^l\| &\leq \|x^n - x^{n+1}\| + \|x^{n+1} - x^{n+2}\| + \cdots + \|x^{l-1} - x^l\| \\ &\leq \delta(\kappa^n t_0 + \kappa^{(n+1)} t_0 + \cdots + \kappa^{(l-1)} t_0) \\ &< \delta s_n t_0, \end{aligned}$$

where $s_n := \sum_{j=n}^{\infty} \kappa^j = \kappa^n/(1-\kappa)$. The sequence of partial sums $\{s_n\}$ converges to zero monotonically as $n \to \infty$, and hence, $\{x^n\}_{n\in\mathbb{N}}$ is a Cauchy sequence. By completeness of $\mathbb{E}$, this implies that the sequence converges to some

$x^* \in \mathbb{E}$. Letting $l \longrightarrow +\infty$ yields

$$\lim_{l\to+\infty} \|x^n - x^l\| = \|x^n - x^*\| \le \delta s_n t_0.$$

Therefore, $\{x^n\}_{n\in\mathbb{N}}$ converges gauge monotonically to x^* with rate $O(s_n)$.
It remains to show that $x^* \in S$. Because S is closed and $\mathbb{E}$ is boundedly compact, for every n, there exists a projection $\overline{x}^n \in P_S(x^n)$. The linear monotonicity assumption then yields

$$\|x^n - \overline{x}^n\| = \operatorname{dist}(x^n, S) \le \kappa \operatorname{dist}(x^{n-1}, S) \le \kappa^{(n-1)} t_0,$$

and hence, $\lim_{n\to\infty} \|x^n - \overline{x}^n\| = 0$. But, by the triangle inequality,

$$\|\overline{x}^n - x^*\| \le \|x^n - \overline{x}^n\| + \|x^n - x^*\|,$$

so $\lim_{n\to\infty} \|\overline{x}^n - x^*\| = 0$. By construction, $\{\overline{x}^n\}_{n\in\mathbb{N}} \subseteq S$, and S is closed; hence, $x^* \in S$. □

In the context of the above result, it is worth mentioning a similar result in Phan [32, proposition 2.11].

Assumption 2 (Regularity Assumptions II). *Let $F : \mathbb{E} \rightrightarrows \mathbb{E}$, $D \subset \mathbb{E}$, and fix the scaling $\lambda > 0$. The following assumptions hold.*

a. (*Existence*) *The set* zer $F \cap D \ne \emptyset$ *and is closed.*
b. (*Self-mapping*) *The scaled resolvent is self-mapping on D, that is, $J_{\lambda F} : D \rightrightarrows D$.*
c. (*Stability*) *Around every $\overline{x} \in F^{-1}(0) \cap D$, there is a neighborhood $\mathcal{N}$ such that $\mathcal{N}' := \{x \in D \mid J_{\lambda F}(x) \subset \mathcal{N} \cap D\} \ne \emptyset$ and F is metrically subregular for zero on $\mathcal{N}$ relative to D with constant $\overline{\rho}$:*

$$d(x, \text{zer } F \cap D) \le \overline{\rho} d(Fx, 0) \quad \forall x \in \mathcal{N} \cap D. \tag{27}$$

d. (*Maximal submonotonicity*) *For every $\overline{x} \in F^{-1}(0) \cap D$, there are neighborhoods U and U' of $\overline{x}$ and a neighborhood W of zero such that*
   i. *λF is maximal submonotone on U in W with violation τ, and*
   ii. *$U' \subseteq (I + \lambda\overline{F})(U)$, where $\overline{F}$ maps $x \longmapsto F(x) \cap (W/\lambda)$ and the constants satisfy $\tau(1 + \overline{\rho}/\lambda)^2 < 1/2$ for $\overline{\rho}$, the constant of metric subregularity in (27).*

We are now ready to state the following local convergence result for the proximal point algorithm.

Theorem 2 (Local Convergence). *For $F : \mathbb{E} \rightrightarrows \mathbb{E}$, $D \subset \mathbb{E}$, let Assumption 2 hold for $\overline{x} \in F^{-1}(0) \cap D$ and the fixed scaling $\lambda > 0$. Then, there exists a $\delta > 0$ such that $J_{\lambda F}(x) \cap \mathbb{B}_\delta(\overline{x}) \ne \emptyset$ whenever $x \in \mathbb{B}_\delta(\overline{x}) \cap D$, and any sequence $\{x^n\}_{n\in\mathbb{N}}$ generated by $x^{n+1} \in J_{\lambda F}(x^n) \cap \mathbb{B}_\delta(\overline{x})$ $(n \in \mathbb{N})$ converges R linearly to a point $\widehat{x} \in F^{-1}(0) \cap \mathbb{B}_\delta(\overline{x}) \cap D$. Moreover,*

$$\|x^n - \widehat{x}\| \le \frac{4\delta\sqrt{1+2\tau}\, \operatorname{dist}(x_0, F^{-1}(0) \cap \mathbb{B}_\delta(\overline{x}) \cap D)}{1-\kappa} \kappa^n, \text{ where } \kappa := \sqrt{1 + 2\tau - \left(\frac{\lambda}{\lambda+\rho}\right)^2} < 1.$$

Proof. Let $\rho > \overline{\rho}$ such that $\tau(1 + \rho/\lambda)^2 < 1/2$ (possible by Assumption 2(d)). Note that by Assumption 2(a), $(\lambda F)^{-1}(0) = F^{-1}(0) \ne \emptyset$ and that by Assumption 2(c) and Lemma 1, Φ_λ is metrically subregular for zero on $\mathcal{N}'$, a neighborhood of $\overline{x}$, relative to D with constant $\overline{\rho}$; hence,

$$(\forall x \in \mathcal{N}' \cap D)(\forall x^+ \in J_{\lambda F}(x)) : \frac{\lambda}{\lambda+\rho} \operatorname{dist}(x, F^{-1}(0) \cap D) \le \|x - x^+\|. \tag{28}$$

The submonotonicity of λF, together with Proposition 3, implies that

$$(\forall u \in U)(\forall u^+ \in \lambda F u \cap W)(\forall p \in F^{-1}(0) \cap U) : \|u - p\|^2 + \|u^+\|^2 \le (1 + 2\tau)\|(u + u^+) - p\|^2. \tag{29}$$

We claim that Assumption 2(d) remains valid upon replacing U and U' by subneighborhoods of $\overline{x}$. This is immediately clear for Assumption 2(d, i). To see that it is true for Assumption 2(d, ii), we argue as follows. For any point $x \in U'$, there exists $u \in U$ and $u^+ \in \lambda F u \cap W$ such that $x = u + u^+$. Let $\epsilon \ge 0$ such that $\mathbb{B}_\epsilon(\overline{x}) \subseteq U'$. Letting $\sigma := \sqrt{1 + 2\tau}\epsilon$, Equation (29) implies that

$$\|u - \overline{x}\| \le \sqrt{1+2\tau}\|x - \overline{x}\| \le \sqrt{1+2\tau}\epsilon = \sigma.$$

In other words, $\mathbb{B}_\sigma(\overline{x}) \subset U$ and $\mathbb{B}_\epsilon(\overline{x}) \subseteq (I + \lambda\overline{F})(\mathbb{B}_\sigma(\overline{x}))$, proving the claim.

We are now justified to assume, without loss of generality, that $U \subseteq \mathcal{N}$ and $U' - U \subseteq W$, replacing U and U' with subneighborhoods of $\overline{x}$ if necessary. By assumption, there exists a neighborhood U' of $\overline{x}$ such that $U' \subseteq (I + \lambda\overline{F})(U)$.

Let $x \in U' \cap D$. Then, by Assumption 2(b) and Assumption 2(d, ii), $J_{\lambda F}(x) \cap U$ is a nonempty subset of D. Hence, there exists $u^+ \in \lambda F u$ such that $u^+ = x - u \in U' - U \subseteq W$ whenever $u \in J_{\lambda F}(x) \cap U$. Thus, $u^+ \in \lambda F u \cap W$. Using (29) and setting $x^+ := u$, we deduce that

$$(\forall x \in U' \cap D)(\forall x^+ \in J_{\lambda F}(x) \cap U \subset D)(\forall p \in F^{-1}(0) \cap U) : \|x^+ - p\|^2 + \|x - x^+\|^2 \le (1 + 2\tau)\|x - p\|^2. \tag{30}$$

Choose $\delta > 0$ such that $\mathbb{B}_r(\overline{x}) \subseteq U' \cap U$, where $r := 2\delta\sqrt{1 + 2\tau}$. Appealing to (30), for all $x \in \mathbb{B}_\delta(\overline{x}) \cap D \subseteq \mathbb{B}_r(\overline{x}) \cap D \subseteq U' \cap D$ and all $x^+ \in J_{\lambda F}(x) \cap U \subset D$, we have

$$\|x^+ - \overline{x}\|^2 \le \|x^+ - \overline{x}\|^2 + \|x - x^+\|^2 \le (1 + 2\tau)\|x - \overline{x}\|^2 \le (1 + 2\tau)\delta^2 \le r^2,$$

from which we conclude that $x^+ \in \mathbb{B}_r(\overline{x}) \cap D$.

Because F is maximal submonotone and $\mathbb{B}_r \cap D$ is a closed subset of U, Theorem 1 implies that $F^{-1}(0) \cap \mathbb{B}_r(\overline{x}) \cap D$ is a nonempty, closed set. The projector $P_{F^{-1}(0)\cap \mathbb{B}_r(\overline{x})\cap D}$ is, therefore, everywhere nonempty, and moreover, we claim that $P_{F^{-1}(0)\cap \mathbb{B}_r(\overline{x})\cap D}(x) \subseteq P_{F^{-1}(0)}(x)$. To see this, suppose that there exists $q \in F^{-1}(0) \setminus \mathbb{B}_r(\overline{x})$ such that $\|q - x\| < \operatorname{dist}(x, F^{-1}(0) \cap \mathbb{B}_r(\overline{x}))$. Then

$$\|q - \overline{x}\| \le \|q - x\| + \|x - \overline{x}\| < \operatorname{dist}(x, F^{-1}(0) \cap \mathbb{B}_r(\overline{x})) + \|x - \overline{x}\| \le 2\|x - \overline{x}\| \le 2\delta \le r,$$

which shows that $q \in \mathbb{B}_r(\overline{x})$. This is a contradiction, which proves the claim.

Altogether, for any $x \in \mathbb{B}_\delta(\overline{x}) \cap D$ and any $x^+ \in J_{\lambda F}(x) \cap \mathbb{B}_r(\overline{x})$, the set $F^{-1}(0) \cap \mathbb{B}_r(\overline{x}) \cap D$ is nonempty, and for any $p \in P_{F^{-1}(0)\cap \mathbb{B}_r(\overline{x})\cap D}(x) \subseteq P_{F^{-1}(0)}(x)$, we have

$$\begin{aligned}
\operatorname{dist}(x^+, F^{-1}(0) \cap \mathbb{B}_r(\overline{x}) \cap D)^2 \le \|x^+ - p\|^2 &\overset{(30)}{\le} (1 + 2\tau)\|x - p\|^2 - \|x - x^+\|^2 \\
&\overset{(28)}{\le} (1 + 2\tau)\|x - p\|^2 - \left(\frac{\lambda}{\lambda + \rho}\right)^2 \operatorname{dist}(x, F^{-1}(0))^2 \\
&\le \left(1 + 2\tau - \left(\frac{\lambda}{\lambda + \rho}\right)^2\right)\|x - p\|^2 \\
&= \kappa^2 \operatorname{dist}(x, F^{-1}(0) \cap \mathbb{B}_r(\overline{x}) \cap D)^2
\end{aligned}$$

for $\kappa := \sqrt{(1 + 2\tau - (\lambda/(\lambda + \rho))^2)}$. That is, we have $\operatorname{dist}(x^+, F^{-1}(0) \cap \mathbb{B}_r(\overline{x}) \cap D) \le \kappa \operatorname{dist}(x, F^{-1}(0) \cap \mathbb{B}_r(\overline{x}) \cap D)$. Again, by Assumption 2(d), the constants satisfy $\tau(1 + \overline{\rho}/\lambda)^2 < 1/2$ for $\overline{\rho}$, the constant of metric subregularity in (27), so $\kappa < 1$. We conclude, therefore, that for all $x^0 \in \mathbb{B}_r(\overline{x}) \cap D$, any sequence $\{x^n\}_{n\in\mathbb{N}}$ with $x^{n+1} \in J_{\lambda F}(x^n) \cap \mathbb{B}_r(\overline{x})$ satisfies assumption (a) of Lemma 2 for all n (with δ in that context replaced by $1 + 2\tau$ because of (30)) as well as being linearly monotone relative to $F^{-1}(0) \cap \mathbb{B}_r(\overline{x}) \cap D$ with rate κ. The result then follows from Lemma 2. □

Corollary 1 (Global Convergence – Maximal Monotonicity). *Let $F : \mathbb{E} \rightrightarrows \mathbb{E}$ with $\overline{x} \in F^{-1}(0)$, and let $\lambda > 0$. Suppose the following assumptions hold.*

a. *The mapping F is maximal monotone, and*
b. *the mapping F is metrically subregular at $\overline{x}$ for zero with modulus $\overline{\rho}$.*

Then, there exists $\delta > 0$ such that for any $x_0 \in \mathbb{B}_\delta(\overline{x})$, the sequence $\{x^n\}_{n\in\mathbb{N}}$, given by $x^{n+1} = J_{\lambda F}(x^n)$ for all $n \in \mathbb{N}$, converges R linearly to a point $\widehat{x} \in F^{-1}(0) \cap \mathbb{B}_\delta(\overline{x})$. Furthermore,

$$\|x^n - \widehat{x}\| \le \frac{\kappa^n \|x_0 - \overline{x}\|}{1 - \kappa}, \text{ where } \kappa := \sqrt{1 - \left(\frac{\lambda}{\lambda + \rho}\right)^2} < 1.$$

Proof. Maximal monotonicity is precisely maximal submonotone on $\mathbb{E}$ in $\mathbb{E}$ with violation $\tau = 0$. We, therefore, have that $\tau(1 + \overline{\rho}/\lambda) = 0 < 1/2$ and that Assumption 2(d, i) is satisfied globally. Assumption 2(d, ii) holds globally because of the Minty correspondence. The rest follows from Theorem 2. □

Remark 3. In the setting of Corollary 1, Leventhal [23, theorem 3.1] showed that sequence of distances to the set of zeros satisfies

$$\operatorname{dist}(x^{n+1}, F^{-1}(0)) \le r \operatorname{dist}(x^n, F^{-1}(0)), \text{ where } r := \sqrt{\frac{\lambda^2}{\lambda^2 + \rho^2}}. \tag{31}$$

In other words, $\{\mathrm{dist}(x^{n+1}, F^{-1}(0))\}_{n\in\mathbb{N}}$ is R linearly convergent to zero with rate r. In this setting, the sequence $\{x^n\}$ is Fejér monotone (hence, linearly monotone) relative to $F^{-1}(0)$ with rate r (assumption (b) of Lemma 2). In this setting, assumption (a) of Lemma 2 holds with $\delta = 1$, and so, Lemma 2 applies. As shown in the proof of this statement,

$$\|x^n - \widehat{x}\| \le \frac{r^n}{1-r} d(x_0, F^{-1}(0)),$$

where $\widehat{x}$ denotes the limit of the sequence $\{x^n\}$. When $r < 1/2$, which corresponds to $\lambda < \rho/\sqrt{3}$, this yields a sharper constant than the estimate of two given by Bauschke [4, theorem 5.12]; for $r > 1/2$, the opposite holds. We, therefore, have the improved estimate

$$\|x^n - \widehat{x}\| \le \min\left\{2, \frac{1}{1-r}\right\} r^n d(x_0, F^{-1}(0)).$$

5. Concluding Remarks

Example 2 shows that submonotonicty is densely distinguishable from hypomonotonicity in the sense that the submonotone mapping given there is submonotone but not hypomonotone on neighborhoods of the domain. This example also points to one application in particular that we think is promising: sparse image recovery using the difference of convex functions (Esser et al. [13], Yin et al. [40]). The least-squares version of the problem is to minimize $\|Ax - b\|^2/2 + \lambda(\|x\|_1 - \|x\|_2)$. Here, we propose replacing the $-\|x\|_2$ portion of the objective with $-\|x\|_{3/2}$. This is beyond the scope of the present study. Another application of submonotonicity is to non-Euclidean proximal algorithms (more precisely, Bregman-type proximal algorithms for solving nonconvex and nonsmooth block optimization problems, where the smooth coupling part of the objective does not satisfy a global/partial Lipschitz gradient continuity assumption). The regularity assumption of the main local convergence result in Cohen [8, theorem 4.2] is exactly point-wise submonotonicity.

The analysis under noise/inexact evaluation of the proximal mapping could be approached in the usual way with vanishing errors. This issue is approached in Hermer et al. [15] by viewing such iterations as a Markov chain converging to an invariant distribution. More precisely, iterations of aα-fne mappings with error are interpreted as random function iterations. Here, convergence of randomly selected mappings, each satisfying Assumption 1(b) (Equation (4)), is obtained in the space of probability distributions with respect to the Wasserstein metric under the assumption of metric subregularity of the Markov operator with respect to the Wasserstein metric. In this way, there is no need for vanishing errors, and the deterministic convergence rates lift to the same convergence rates in the Wasserstein metric of the corresponding probability distributions. This has been done for aα-fne mappings in Hermer et al. [14, theorem 2.6] and so, directly applies to proximal point iterations of submonotone mappings with error/noise.

Endnotes

[1] The term has already been used in Daniilidis and Georgiev [10] and Spingarn [38], each for different notions of regularity.

[2] If the functions f_i are weakly convex, each with constant ρ, then the functions $f_i + \rho\|\cdot\|^2$ are convex. The maximum of convex functions is again convex, and so, $\max_i f_i = \max_i\{f_i + \rho\|\cdot\|^2\} - \rho\|\cdot\|^2$ is weakly convex with constant ρ.

References

[1] Ariza-Ruiz D, Leuştean L, López-Acedo G (2014) Firmly nonexpansive mappings in classes of geodesic spaces. *Trans. Amer. Math. Soc.* 366(8):4299–4322.

[2] Artacho FA, Dontchev A, Geoffroy M (2007) Convergence of the proximal point method for metrically regular mappings. *ESAIM Proc.* 17:1–8.

[3] Baillon JB, Bruck RE, Reich S (1978) On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces. *Houston J. Math.* 4(1):1–9.

[4] Bauschke HH, Combettes PL (2017) *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. (Springer, Cham, Switzerland).

[5] Bërdëllima A, Lauster F, Luke DR (2022) α-Firmly nonexpansive operators on metric spaces. *J. Fixed Point Theory Appl.* 24:14.

[6] Bruck RE, Reich S (1977) Nonexpansive projections and resolvents of accretive operators in Banach spaces. *Houston J. Math.* 3(4):459–470.

[7] Burachik RS, Iusem AN (2008) *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer Optimization and Its Applications, vol. 8 (Springer, New York).

[8] Cohen E, Luke DR, Pinta T, Sabach S, Teboulle M (2024) A semi-Bregman proximal alternating method for a class of nonconvex problems: Local and global convergence analysis. *J. Global Optim.* 89(1):33–55.

[9] Combettes PL, Pennanen T (2004) Proximal methods for cohypomonotone operators. *SIAM J. Control Optim.* 43(2):731–742.

[10] Daniilidis A, Georgiev P (2004) Approximate convexity and submonotonicity. *J. Math. Anal. Appl.* 291(1):292–301.

[11] Dontchev AL, Rockafellar RT (2014) *Implicit Functions and Solution Mappings*, 2nd ed. (Springer-Verlag, Dordrecht, Netherlands).
[12] Edelstein M (1966) A remark on a theorem of M. A. Krasnoselski. *Amer. Math. Monthly* 73(5):509–510.
[13] Esser E, Lou Y, Xin J (2013) A method for finding structured sparse solutions to nonnegative least squares problems with applications. *SIAM J. Imaging Sci.* 6(4):2010–2046.
[14] Hermer N, Luke DR, Sturm A (2019) Random function iterations for consistent stochastic feasibility. *Numer. Functional Anal. Optim.* 40(4):386–420.
[15] Hermer N, Luke DR, Sturm A (2023) Rates of convergence for chains of expansive Markov operators. *Trans. Math. Its Appl.* 7(1):tnad001.
[16] Ioffe AD (2011) Regularity on a fixed set. *SIAM J. Optim.* 21(4):1345–1370.
[17] Ioffe AD (2013) Nonlinear regularity models. *Math. Programming* 139(1–2):223–242.
[18] Iusem AN, Pennanen T, Svaiter BF (2003) Inexact variants of the proximal point algorithm without monotonicity. *SIAM J. Optim.* 13(4):1080–1097.
[19] Krasnoselski MA (1955) Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk* 63(1):123–127.
[20] Kruger AY, Luke DR, Thao NH (2018) Set regularities and feasibility problems. *Math. Programming* 168(1–2):279–311.
[21] Lasry JM, Lions PL (1986) A remark on regularization in Hilbert spaces. *Israel J. Math.* 55:257–266.
[22] Lauster F, Luke DR (2021) Convergence of proximal splitting algorithms in CAT(κ) spaces and beyond. *Fixed Point Theory Algorithms Sci. Engrg.* 2021:13.
[23] Leventhal D (2009) Metric subregularity and the proximal point method. *J. Math. Anal. Appl.* 360(2):681–688.
[24] Luke DR, Tam MK (2023) Generalized monotonicity and the proximal point algorithm. *Proc. 35th RAMP Sympos.* (Research Association of Mathematical Programming, Tokyo), 39–48.
[25] Luke DR, Teboulle M, Thao NH (2020) Necessary conditions for linear convergence of iterated expansive, set-valued mappings. *Math. Programming* 180:1–31.
[26] Luke DR, Thao NH, Tam MK (2018) Quantitative convergence analysis of iterated expansive, set-valued mappings. *Math. Oper. Res.* 43(4):1143–1176.
[27] Mann WR (1953) Mean value methods in iterations. *Proc. Amer. Math. Soc.* 4(3):506–510.
[28] Martinet B (1970) Régularisation d'inéquations variationnelles par approximations successives. *Revue Française D'automatique Informatique Recherche Opérationnelle* 3:154–158.
[29] Minty GJ (1962) Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.* 29(3):341–346.
[30] Ngai HV, Luc DT, Théra M (2000) Approximate convex functions. *J. Nonlinear Convex Anal.* 1(2):155–176.
[31] Pennanen T (2002) Local convergence of the proximal point algorithm and multiplier methods without monotonicity. *Math. Oper. Res.* 27(1):170–191.
[32] Phan H (2016) Linear convergence of the Douglas–Rachford method for two closed sets. *Optimization* 65(2):369–385.
[33] Poliquin RA, Rockafellar RT (1996) Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.* 348(5):1805–1838.
[34] Poliquin RA, Rockafellar RT, Thibault L (2000) Local differentiability of distance functions. *Trans. Amer. Math. Soc.* 352(11):5231–5249.
[35] Rockafellar RT (1976) Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* 14(5):877–898.
[36] Rockafellar RT (2023) Generic linear convergence through metric subregularity in a variable-metric extension of the proximal point algorithm. *Comput. Optim. Appl.* 86(3):1327–1346.
[37] Rolewicz S (1999) On $\alpha(\cdot)$-monotone multifunctions and differentiability of γ-paraconvex functions. *Stud. Math.* 133(1):29–37.
[38] Spingarn J (1981) Submonotone subdifferentials of Lipschitz functions. *Trans. Amer. Math. Soc.* 264(1):77–89.
[39] Vial J-P (1983) Strong and weak convexity of sets and functions. *Math. Oper. Res.* 8(2):231–259.
[40] Yin P, Lou Y, He Q, Xin J (2015) Minimization of ℓ_{1-2} for compressed sensing. *SIAM J. Sci. Comput.* 37(1):A536–A563.