Proceedings of the Thirty-Fifth RAMP Symposium
Tokyo Institute of Technology, Tokyo, November 20–21, 2023

# The proximal point algorithm without monotonicity

D. Russell Luke[1*] and Matthew K. Tam[2†]

**Abstract** We study the proximal point algorithm in the setting in which the operator of interest is metrically subregular and satisfies a submonoticity property. The latter 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 at least one sequence which is 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.

**Keywords** submonotone, proximal point algorithm, metric subregularity, almost $\alpha$-firmly nonexpansive, monotone, hypomonotone

## 1. Introduction

When cast abstractly, many problems amount to finding the zero of a set-valued map. Indeed, minimization problems are most often not solved directly, rather one seeks to satisfy an appropriate *optimality condition* which is typically phrased in terms of finding a zero of some *generalized derivative* operator. One method for finding such zeros is the *proximal point algorithm*. When the underlying set-valued map is *maximal monotone* the algorithm can be interpreted as the fixed point iteration corresponding map's *resolvent*; the latter being *firmly nonexpansive* with full domain due to a celebrated result originating from the work of Minty [16]. A great deal of the literature studying the algorithm, therefore, heavily relies on the aforementioned monotonicity and nonexpansive properties in their analyses. The interested reader is referred to [3, 6, 22] for further details.

In the absence of monotonicity, the proximal point algorithm has been studied for the family of so-called *hypomonotone* mappings and their variants by Pennanen and coauthors [7, 10, 18] (see also [6, Ch. 6.9]). Roughly speaking, such mappings can be made "locally" monotone through 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 which allows for a great deal of structural properties of the original (nonmonotone) operator to be deduced and upon which the convergence analysis relies. Following a different direction, dispatching with generalizations of monotonicity completely, Aragón Artacho, Dontchev and Geoffroy [1] 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 which does so with linear rate (and, moreover, that

1 Georg-August-University of Göttingen, Göttingen, Germany
2 School of Mathematics & Statistics, The University of Melbourne, Parkville, VIC 3010, Australia
∗ E-mail address: r.luke@math.uni-goettingen.de
† E-mail address: matthew.tam@unimelb.edu.au

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 which are merely *metrically subregular*.

The goal of this work is therefore to provide conditions for (local) convergence of the proximal point algorithm when the underlying mapping is metrically subregular but its inverse is not necessarily hypomonotone, thus bridging the gap between the two aforementioned approaches to analyzing the proximal point algorithm with nonmonotonity. Our main result (Theorem 12) considers metrically subregular operators that satisfy a new *submonotonicity* property for operators whose resolvents are *almost $\alpha$-firmly nonexpansive* (3).

The remainder of this work is organized as follows. In Section 2, we set notation and collect various preliminary results for use in the sequel. Section 3 focuses on the study of our new generalized monotonicity property which we call *submonotonicity*. The developed 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 multi-valued mapping $F : \mathbb{E} \rightrightarrows \mathbb{E}$;

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

Given $\lambda > 0$, the *resolvent* of $F$ is the multi-valued map $J_{\lambda F} := (\operatorname{Id} + \lambda F)^{-1}$. A fundamental numerical method to solve (1), and the focus of this study, is the multi-valued generalization of the implicit Euler method, the *proximal point algorithm*: given an initial point $x_0 \in \mathbb{E}$, 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 non-expansive* with full domain; that is, they satisfy

$$\|x - J_{\lambda_n F}(x_0)\|^2 \le \|x - x_0\| - \|(J_{\lambda_n F}(x) - x) - (J_{\lambda_n F}(x_0) - x_0)\|^2 \quad \forall x, x_0 \in \mathbb{E}.$$

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 $F^{-1}(0) \neq \emptyset$ [22, 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** (linear metric subregularity). *A set-valued mapping $\Phi : \mathbb{E} \rightrightarrows \mathbb{E}$ is* linearly 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*

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

Convergence rates for mappings that are only *almost* firmly nonexpansive was established in [14]. On a closed subset $D \subset \mathbb{E}$, a general self-mapping $T : D \rightrightarrows D$ is said to be *pointwise almost $\alpha$-firmly nonexpansive at $x_0 \in D$ on $D$*, abbreviated *pointwise a$\alpha$-fne*, whenever $\alpha \in (0,1)$ and there exists $\epsilon \in [0,1]$ such that

$$\begin{aligned}&(\forall x \in D)(\forall x^+ \in Tx)(\forall x_0^+ \in Tx_0):\\ &\|x^+ - x_0^+\|^2 \le (1+\epsilon)\|x - x_0\|^2 - \tfrac{1-\alpha}{\alpha}\left\|(x - x^+) - (x_0 - x_0^+)\right\|^2.\end{aligned} \tag{3}$$

When the above inequality holds for all $x_0 \in D$ then $T$ is said to be *almost $\alpha$-firmly nonexpansive (a$\alpha$-fne) on $D$*. The violation is a value of $\epsilon$ for which (3) holds. When the violation is 0, the qualifier "almost" is dropped and the abbreviation *$\alpha$-fne* is used. The definition of pointwise a$\alpha$-fne mappings in Euclidean spaces appeared first in this form in [14] 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 [2, 5, 9, 11, 15]. These notions have been extended to nonlinear spaces [4] where addition doesn't always exist, hence the change in terminology.

**Proposition 1** (Corollary 2.3 of [14])**.** *Let $T : D \rightrightarrows D$ for $D \subset \mathbb{E}$ with $\mathsf{Fix}\, T \cap D$ nonempty and closed. Denote $(\mathsf{Fix}\, T + \delta\mathbb{B}) \cap D$ by $S_\delta$ for a nonnegative real number $\delta$ and define $\Phi := \mathrm{Id} - T$. Suppose that, for all $\delta > 0$ small enough, there exist $\varepsilon > 0$ and $\alpha \in (0,1)$, such that*

*(a) $T$ is pointwise a$\alpha$-fne at all $y \in \mathsf{Fix}\, T \cap D$ with constant $\alpha$ and violation $\varepsilon$ on $S_\delta$, and*

*(b) $\Phi$ is linearly metrically subregular for 0 on $S_\delta$ relative to $D$ with constant $\rho > 0$ satisfying*

$$\sqrt{\frac{1-\alpha}{\alpha(1+\varepsilon)}} < \rho < \sqrt{\frac{1-\alpha}{\varepsilon\alpha}}. \tag{4}$$

*Then, for any $x_0 \in D$ close enough to $\mathsf{Fix}\, T \cap D$, the iterates $x_{n+1} \in Tx_n$ satisfy*

$$\mathrm{dist}\,(x_n, \mathsf{Fix}\, T \cap D) \to 0$$

*and*

$$\mathrm{dist}\,(x_{n+1}, \mathsf{Fix}\, T \cap D) \le c\, \mathrm{dist}\,(x_n, \mathsf{Fix}\, T \cap D) \quad \forall n \in \mathbb{N}, \tag{5}$$

*where $c := \sqrt{1 + \varepsilon - \left(\frac{1-\alpha}{\rho^2\alpha}\right)} < 1$.*

The goal of this study is to establish linear convergence guarantees for mappings $F$ whose resolvents are only pointwise a$\alpha$-fne.

## 3. Submonotone mappings

In this section, we introduce a generalized monotonicity property of set-valued maps which, for lack of a better terminology, we call *submonotonicity*. This name was given by Spingarn [23] to mappings that are maximally strictly *hypomonotone* (see Definition 3 below); subsequent work [8] studied this in relation to *approximate convexity* [17]. We repurpose this term for a differently defined object whose properties we investigate, exploring connections to other generalized monotonicity properties in the literature.

Pointwise a$\alpha$-fne mappings discussed above lead to our notion of submonotone mappings in the follow way. If $F$ has a resolvent $J_F$ that is pointwise a$\alpha$-fne on $U$ at all $y \in S$ with $\alpha = 1/2$ and violation $\epsilon$, then $F$ satisfies [14, Proposition 2.3]:

$$\begin{aligned}
&(\forall x \in U)(\forall (u, z) \in \operatorname{gph} F \text{ with } z = x - u \text{ for } u \in J_F(x)) \\
&(\forall y \in S)(\forall (v, w) \in \operatorname{gph} F \text{ with } w = y - v \text{ for } v \in J_F(y)) : \\
&\quad -\tfrac{\epsilon}{2} \|(u + z) - (v + w)\|^2 \le \langle u - v,\ z - w\rangle .
\end{aligned} \tag{6}$$

Conversely, if $F$ satisfies (6), then the resolvent is a$\alpha$-fne at all $y \in S$ on $U$ with constant $\alpha = 1/2$ and violation $\epsilon$. 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 [3]). With this in mind, we define the following.

**Definition 2** (submonotonicity)**.** *Let $U$ and $V$ 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 $V$ with violation $\tau$ if*

$$\begin{aligned}
&(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap V)(\forall v^+ \in Fv \cap V) : \\
&\quad -\tau\|(u + u^+) - (v + v^+)\|^2 \le \langle u - v, u^+ - v^+\rangle.
\end{aligned} \tag{7}$$

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

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

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 $V$ with violation $\tau$. Then there exists an operator $\widetilde{F}$ with* $\operatorname{gph} F \subseteq \operatorname{gph} \widetilde{F}$*, that is maximal submonotone on $U$ in $V$ with violation $\tau$.*

The following gives some equivalent forms of submonotonicity. The proof is omitted since these are just rearrangements of the definition.

**Proposition 3** (characterizations of submonotonicity)**.** *Let $U$ and $V$ be nonempty subsets of $\mathbb{E}$, let $\tau > 0$ and let $F : \mathbb{E} \to \mathbb{E}$. The following assertions are equivalent.*

*(a)* *$F$ is submonotone on $U$ in $V$ with violation $\tau$.*

*(b)* $(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap V)(\forall v^+ \in Fv \cap V)$ :

$$\|u - v\|^2 + \|u^+ - v^+\|^2 \le (1 + 2\tau)\|(u + u^+) - (v + v^+)\|^2.$$

*(c)* $(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap V)(\forall v^+ \in Fv \cap V)$ :

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

*(d)* $(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap V)(\forall v^+ \in Fv \cap V)$ :

$$0 \le \|u-u^+\|^2 + \|v-v^+\|^2 + \frac{2\tau}{(1+2\tau)}\left(\|u-v\|^2 + \|u^+ - v^+\|^2\right) - \|v-u^+\|^2 - \|u-v^+\|^2.$$

**Proposition 4** (submonotonicity and inverses)**.** *Let $U$ and $V$ 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 $V$ with violation $\tau$ if and only if $F^{-1}$ is (maximal) submonotone on $V$ in $U$ with violation $\tau$.*

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 *$\rho$-weakly convex* if the function $f + \rho\|\cdot\|^2$ is convex. In particular, a 0-weakly convex function is convex.

**Proposition 5.** *Let $U$ and $V$ be subsets of $\mathbb{E}$, let $\tau \ge 0$, and suppose that $F : \mathbb{E} \rightrightarrows \mathbb{E}$ is maximal submonotone on $U$ in $V$ with violation $\tau$. Then, for all $u \in U$, $F(u) \cap V$ can be expressed as intersection of the lower-level set of a proper, lsc, $\tau$-weakly convex function*[*1] *and $V$. Consequently, for every closed subset $O$ of $V$, the set $F(u) \cap O$ is closed. For any $v \in V$, the analogous statement holds for $F^{-1}(v) \cap U$.*

### 3.1. Relation to other notions of generalized monotonicity

In this section we compare our newly introduced submonotonicity property to other weakening of monotonicity in the literature. The first such property which we discuss is that of *hypomonotonicity* which has its origins in [20, 21, 23].

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

$$\begin{aligned} &(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap V)(\forall v^+ \in Fv \cap V) : \\ &-\kappa\|u-v\|^2 \le \langle u-v, u^+ - v^+\rangle. \end{aligned} \tag{8}$$

With regard to Definition 3, observe that, by rearranging of (8), it can be seen that hypomonotonicity $F$ on $U$ for $V$ is equivalent to monotonicity of $F + \kappa\,\mathrm{Id}$ in $U$ for $V$.

**Remark 6** (monotone operators)**.** *Recall that a set-valued map $F : \mathbb{E} \rightrightarrows \mathbb{E}$ is said to be* monotone *on $U$ for $V$ if*

$$\begin{aligned} &(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap V)(\forall v^+ \in Fv \cap V) : \\ &0 \le \langle u-v, u^+ - v^+\rangle. \end{aligned} \tag{9}$$

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

---

*1 If $\{f_i\}$ are $\rho$-weakly convex then $\{f_i + \rho\|\cdot\|^2\}$ are convex. The max of convex functions is again convex, and so $\max_i f_i = \max_i\{f_i + \rho\|\cdot\|^2\} - \rho\|\cdot\|^2$ is $\rho$-weakly convex.

*Monotonicity is preserved by taking inverses in the sense that* (9) *is equivalent to monotonicity of $F^{-1}$ on $V$ in $U$. As we have already seen in Proposition 4, the analogous statement is true of the inverse of a submonotone operator. It is, however, clear from* (9)*, that one can not expect the same to be true for hypomonotone operators in general. The term* co-hypomonotonicity *has been used to refer to an operator whose inverse is hypomonotone [7].*

*On the other hand, hypomonotonicity of an operator is preserved under positive scalar multiplication; a property which also holds for monotone operator. More precisely, if $F$ is (hypo-)monotone and $\lambda > 0$ then $\lambda F$ is also (hypo-)monotone. As we show in Example 1, the same this property is generally not satisfied by submonotone mappings.*

**Proposition 7** (hypo-/sub-monotonicity)**.** *Let $U$ and $V$ 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 $V$ with violation $\sigma \in [0, 1/2)$, then $F$ is submonotone on $U$ in $V$ with violation $\tau := \sigma/(1 - 2\sigma) \geq 0$, and $F^{-1}$ is submonotone on $V$ in $U$ with violation $\tau$.*

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

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

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

*(c) If $F$ is submonotone on $U$ in $V$ with violation $\tau$ and there exists a $\kappa \geq 0$ such that*

$$\begin{gathered}(\forall u \in U)(\forall v \in U)(\forall u^+ \in Fu \cap V)(\forall v^+ \in Fv \cap V) : \\ \|u^+ - v^+\| \leq \kappa \|u - v\|;\end{gathered} \tag{10}$$

*then $F$ is hypomonotone on $U$ in $V$ with violation $\sigma := \tau(1 + \kappa^2)/(1 + 2\tau) \geq 0$.*

**Remark 8.** *Regarding condition* (10) *of Proposition 7, in the case where $F(\cdot) \cap V$ is at most single-valued on $U$, it is clear that condition amounts to Lipschitz continuity of $F(\cdot)$ on $V$. For a general set-valued map, the condition is stronger than Lipschitz continuity of $F(\cdot)$ on $V$.*

We now give two examples to show that the conditions of Proposition 7 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 $\alpha > 0$, let $U := [0, \alpha]$, $V := \mathbb{R}$ and consider the function $F(x) := -x^2$. Then, for all $(u, u^+) \in \operatorname{gph} F \cap (U \times V)$ and $(v, v^+) \in \operatorname{gph} F \cap (U \times V)$, we have*

$$\langle u - v, u^+ - v^+ \rangle = (u - v)(-u^2 + v^2) = -(u + v)\|u - v\|^2 \geq -2\alpha \|u - v\|^2, \tag{11}$$

*which shows that $F$ is hypomonotone on $U$ in $V$ with violation $2\alpha$.*

*We claim, however, that $F$ is submonotone only if $\alpha < 1/2$. To this end, observe that*

$$\|(u+u^+)-(v+v^+)\|^2 = \|(u-v)-(u^2-v^2)\|^2 = (1-(u+v))^2\|u-v\|^2. \tag{12}$$

*Whenever $u+v \neq 1$, combining (11) and (12) yields*

$$-\frac{(u+v)}{(1-(u+v))^2}\|(u+u^+)-(v+v^+)\|^2 = \langle u-v, u^+-v^+\rangle. \tag{13}$$

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

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

*and hence that $F$ is submonotone on $U$ in $V$ with violation $\tau$.*

*On the other hand, that $\alpha \geq 1/2$ and there exists a $\tau \geq 0$ such that $F$ is submonotone on $U$ in $V$ with violation $\tau$. Let $\{u_n\}, \{v_n\}$ be two sequences contained in $[0, 1/2)$ which both converge to $1/2$. Then, using (13), we have*

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

*which contradictions the finiteness of $\tau$, and we conclude that $F$ is not submonotone on $U$ for $V$ for any violation constant.*

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

$$\langle u-v, u^+-v^+\rangle = (u-v)(-\sqrt{u}+\sqrt{v}) = -\frac{1}{\sqrt{u}+\sqrt{v}}\|u-v\|^2. \tag{14}$$

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

$$\begin{aligned} \|(u+u^+)-(v+v^+)\|^2 &= \|(u-v)-(\sqrt{u}-\sqrt{v})\|^2 \\ &= \left\|(u-v) - \frac{1}{\sqrt{u}+\sqrt{v}}(u-v)\right\|^2 \\ &= \left(\frac{(\sqrt{u}+\sqrt{v})-1}{\sqrt{u}+\sqrt{v}}\right)^2\|u-v\|^2. \end{aligned} \tag{15}$$

*Since $\sqrt{u}+\sqrt{v} \leq 1/2$, it holds 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 (14) and (15), whenever $\sqrt{u}+\sqrt{v} \neq 0$, we deduce that*

$$-\tau\|(u+u^+)-(v+v^+)\|^2 \leq \langle u-v, u^+-v^+\rangle. \tag{16}$$

*Moreover, if $\sqrt{u}+\sqrt{v} = 0$ then $u = v = 0$, and so (16) remains true in this case. Altogether, this shows that $F$ is submonotone on $U$ in $V$ violation $\tau = 2$, as was claimed.*

## 4. The proximal point algorithm

To begin, we establish the implication of metric subregularity of a scaled multi-valued mapping $\lambda F$ for the resolvent residual mapping $\Phi$ defined in Proposition 1.

**Lemma 9.** *Let $F : \mathbb{E} \rightrightarrows \mathbb{E}$, $F^{-1}(0) \cap D \neq \emptyset$ closed for $D \subset \mathbb{E}$ with $J_{\lambda F} : D \to D$ and $\lambda > 0$. If $F$ is metrically subregular for $0$ on $U$ relative to $D$ with constant $\rho$, 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 $0$ on $U'$ relative to $D$ with constant $\frac{\lambda+\rho}{\lambda}$.*

**Remark 10.** *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 $\mathsf{Fix}\, J_{\lambda F}$, 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

$$(\forall n \in \mathbb{N}) \quad \mathrm{dist}(x_{n+1}, S) \leq \kappa \, \mathrm{dist}(x_n, S). \tag{17}$$

This was introduced in [13] for more general gauges. Fejér monotone sequences, in contrast, satisfy

$$\mathrm{dist}(x_{n+1}, x) \leq \mathrm{dist}(x_n, x) \quad \forall x \in S, \forall n \in \mathbb{N}.$$

It is easy to see that any Fejér monotone sequence is linearly monotone, but the converse is not true (see [13, Example 1]).

**Lemma 11** (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*

*(a) $\|x_{n+1} - x_n\| \leq \delta \, \mathrm{dist}(x_n, S)$ 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 linearly monotonically to a point $\overline{x} \in S$ with rate $O(\frac{\kappa^n}{1-\kappa})$.*

In the context of the above result, it is worth mentioning a similar result in [19, Prop. 2.11].

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

**Theorem 12** (local convergence)**.** *Let $F : \mathbb{E} \rightrightarrows \mathbb{E}$, $D \subset \mathbb{E}$, with $\bar{x} \in F^{-1}(0) \cap D$, and $J_{\lambda F} : D \rightrightarrows D$ for $\lambda > 0$. Suppose the following assumptions hold.*

*(a) There exists a neighborhood $W$ of $\overline{x}$ such that $W' := \{x \in D \mid J_{\lambda F}(x) \subset W \cap D\} \neq \emptyset$ and $F$ is metrically subregular for $0$ on $W$ relative to $D$ with constant $\overline{\rho}$.*

*(b) There exists neighborhoods $U$ and $U'$ of $\bar{x}$, and $V$ of $0$ such that*

*(i) $\lambda F$ is maximal submonotone on $U$ in $V$ with violation $\tau$, and*

*(ii)* $U' \subseteq (I + \lambda\overline{F})(U)$ *where* $\overline{F}$ *maps* $x \mapsto F(x) \cap (V/\lambda)$*;*

*where the constants satisfy* $\tau\,(1+\overline{\rho}/\lambda)^2 < 1/2$.

*Then, there exist* $\delta > 0$ *and* $\rho > \overline{\rho}$ *such that, for any* $x_0 \in \mathbb{B}_\delta(\bar{x}) \cap D$*, there exists a sequence* $\{x_n\}_{n\in\mathbb{N}}$*, given by* $x_{n+1} \in J_{\lambda F}(x_n)$ *for all* $n \in \mathbb{N}$*, that converges R-linearly to a point* $\widehat{x} \in F^{-1}(0) \cap \mathbb{B}_\delta(\bar{x}) \cap D$*. Moreover,*

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

**Corollary 13** (maximal monotonicity)**.** *Let* $F : \mathbb{E} \rightrightarrows \mathbb{E}$ *with* $\bar{x} \in F^{-1}(0)$ *and let* $\lambda > 0$*. Suppose the following assumptions hold.*

*(a)* $F$ *is metrically subregular at* $\bar{x}$ *for* $0$ *with modulus* $\overline{\rho}$*, and*

*(b)* $F$ *is maximal monotone.*

*Then, there exist* $\delta > 0$ *and* $\rho > \overline{\rho}$ *such that, for any* $x_0 \in \mathbb{B}_\delta(\bar{x})$*, 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* $\widehat{x} \in F^{-1}(0) \cap \mathbb{B}_\delta(\bar{x})$*. Furthermore, it holds that*

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

**Remark 14.** *In the setting of Corollary 13, Leventhal [12, Th. 3.1] showed that sequence of distances to the zeros set satisfies*

$$\mathrm{dist}(x_{n+1}, F^{-1}(0)) \le r\ \mathrm{dist}(x_n, F^{-1}(0)) \text{ where } r := \sqrt{\frac{\lambda^2}{\lambda^2+\rho^2}}. \tag{18}$$

*In light of our development, the sequence* $\{x_n\}$ *is therefore linearly monotone with respect to* $F^{-1}(0)$ *with rate* $r$*, and so by Lemma 11* $x_n \to \widehat{x}$ *at least R-linearly with rate* $r$*.*

Acknowledgements.

DRL was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 432680300 - SFB 1456. MKT was supported in part by Australian Research Council grants DE200100063 and DP230101749.

## References

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

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

[3] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces. 2nd edition.* Cham: Springer, 2nd edition edition, 2017.

[4] A. Bërdëllima, F. Lauster, and D. R. Luke. $\alpha$-firmly nonexpansive operators on metric spaces. *J. Fixed Point Theory Appl.*, 24, 2022.

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

[6] R. S. Burachik and A. N. Iusem. *Set-Valued Mappings and Enlargements of Monotone Operators*, volume 8 of *Optimization and Its Applications*. Springer, 2008.

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

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

[9] M. Edelstein. A remark on a theorem of M. A. Krasnoselski. *Amer. Math. Monthly*, 73(5):509–510, May 1966.

[10] A. N. Iusem, T. Pennanen, and B. F. Svaiter. Inexact variants of the proximal point algorithm without monotonicity. *SIAM J. Optim.*, 13(4):1080–1097, 2003.

[11] M. A. Krasnoselski. Two remarks on the method of successive approximations. *Math. Nauk. (N.S.)*, 63(1):123–127, 1955. (Russian).

[12] D. Leventhal. Metric subregularity and the proximal point method. *Journal of Mathematical Analysis and Applications*, 360(2):681–688, 2009.

[13] D. R. Luke, M. Teboulle, and N. H. Thao. Necessary conditions for linear convergence of iterated expansive, set-valued mappings. *Math. Program.*, 180:1–31, 2018.

[14] D. R. Luke, N. H. Thao, and M. K. Tam. Quantitative convergence analysis of iterated expansive, set-valued mappings. *Math. Oper. Res.*, 43(4):1143–1176, 2018.

[15] W. R. Mann. Mean value methods in iterations. *Proc. Amer. Math. Soc.*, 4:506–510, 1953.

[16] G. J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke. Math. J.*, 29(3):341–346, 1962.

[17] H. V. Ngai, D. T. Luc, and M. Théra. Approximate convex functions. *J. Nonlinear Convex Anal.*, 1:155–176, 2000.

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

[19] H. Phan. Linear convergence of the Douglas–Rachford method for two closed sets. *Optimization*, 65(2):369–385, 2016.

[20] R. A. Poliquin and R. T. Rockafellar. Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.*, 348:1805–1838, 1996.

[21] R. A. Poliquin, R. T. Rockafellar, and L. Thibault. Local differentiability of distance functions. *Trans. Amer. Math. Soc.*, 352(11):5231–5249, 2000.

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

[23] J. Spingarn. Submonotone subdifferentials of lipschitz functions. *Trans. Amer. Math. Soc.*, 264:77–89, 1981.