J. Math. Anal. Appl. 360 (2009) 681–688

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Metric subregularity and the proximal point method

D. Leventhal

School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 24 February 2009
Available online 30 July 2009
Submitted by A. Daniilidis

Keywords:
Monotone operator
Firmly non-expansive mapping
Proximal point
Resolvent
Metric regularity
Metric subregularity
Randomization

ABSTRACT

We examine the linear convergence rates of variants of the proximal point method for finding zeros of maximal monotone operators. We begin by showing how metric subregularity is sufficient for local linear convergence to a zero of a maximal monotone operator. This result is then generalized to obtain convergence rates for the problem of finding a common zero of multiple monotone operators by considering randomized and averaged proximal methods.

1. Introduction

Let $\mathcal{H}$ be a real Hilbert space and let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-valued mapping. Two common problems that arise in several branches of applied mathematics are to

$$\text{Find } x \in \mathcal{H} \quad \text{such that} \quad 0 \in T(x) \tag{1.1}$$

and, more generally

$$\text{Find } x \in \mathcal{H} \quad \text{such that} \quad 0 \in \bigcap_{i \in I} T_i(x), \tag{1.2}$$

where I is some index set. Specifically, these problems correspond to finding a zero of an operator and, more generally, a common zero of multiple operators.

Suppose that the operators under consideration are *monotone*, meaning that

$$\langle x_1 - x_0, y_1 - y_0 \rangle \geqslant 0 \quad \text{for all } x_0, x_1 \in \mathcal{H},\ \ y_0 \in T(x_0),\ \ y_1 \in T(x_1).$$

For $\lambda > 0$, the mappings $J_{\lambda T} := (I + \lambda T)^{-1}$ are the *resolvents* of T, which were shown to be at most single-valued in [27]. One proposed method for solving problem (1.1) is the *proximal point algorithm*, considered originally in [26] and more thoroughly explored by [32], given by, for $k = 0, 1, 2, \ldots,$

$$x_{k+1} = J_{\lambda T}(x_k). \tag{1.3}$$

Our goal is to examine how appropriate regularity assumptions on the operators T (or $T_1, \ldots, T_m$, respectively) affect the speed of convergence of variants of the proximal point algorithm. In order to do so, the remainder of this paper is organized

E-mail address: leventhal@orie.cornell.edu.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2009.07.012

as follows. In Section 2, we provide notation and basic facts about monotone operators, metric regularity and subregularity, and the geometry of convex sets. Then, in Section 3, we show how assumptions of metric subregularity can be used to demonstrate linear convergence of both the proximal point algorithm for problem (1.1) and a randomized proximal point algorithm for problem (1.2).

2. Background and notation

A single-valued operator U is *firmly non-expansive* if

$$\big\|U(x)-U(y)\big\|^2+\big\|(I-U)(x)-(I-U)(y)\big\|^2\leqslant\|x-y\|^2\quad\forall x,y\in\mathcal{H}.\tag{2.1}$$

It was shown in [13,32] that an operator T is maximal monotone, meaning the graph of T is not strictly contained in the graph of another monotone operator, if and only if its resolvents are firmly non-expansive. The domain of T is $\{x\in\mathcal{H}\colon\ T(x)\neq\emptyset\}$ and the inverse operator, T^{-1}, is defined by $T^{-1}(y)=\{x\colon\ y\in T(x)\}$. It is known that (see [33], for example) T is maximal monotone if and only if T^{-1} is maximal monotone, in which case both T and T^{-1} are closed and convex-valued.

We are interested in how certain regularity conditions affect *local* rates of convergence. One prominent condition is the idea of metric regularity of set-valued mappings. We say the set-valued mapping Φ is *metrically regular* at $\bar{x}$ for $\bar{b}\in\Phi(\bar{x})$ if there exists $\gamma>0$ such that

$$d\big(x,\Phi^{-1}(b)\big)\leqslant\gamma d\big(b,\Phi(x)\big)\quad\text{for all }(x,b)\text{ near }(\bar{x},\bar{b}).\tag{2.2}$$

Further, the *modulus of regularity* is the infimum of all constants γ such that inequality (2.2) holds.

A slightly weaker condition is that of metric subregularity. We say the set-valued mapping Φ is *metrically subregular*[1] at $\bar{x}$ for $\bar{b}\in\Phi(\bar{x})$ if there exists $\gamma>0$ such that

$$d\big(x,\Phi^{-1}(\bar{b})\big)\leqslant\gamma d\big(\bar{b},\Phi(x)\big)\quad\text{for all }x\text{ near }\bar{x}.\tag{2.3}$$

Further, the *modulus of subregularity* is the infimum of all constants γ such that inequality (2.3) holds. Note that for metric subregularity, the reference vector $\bar{b}$ is fixed in inequality (2.3) but not in inequality (2.2). It is clear from the definitions that metric regularity implies metric subregularity; hence, the modulus of subregularity is no larger than the modulus of regularity, using the convention that the modulus of (sub)regularity is infinite if the mapping fails to be metrically (sub)regular.

The property of metric regularity is connected with other ideas in variational analysis. The simplest connection, as shown in [11, Ex. 1.1], is that metric regularity generalizes the Banach open mapping principle, essentially saying that a bounded and linear mapping is metrically regular if and only if it is surjective; in such a case, the modulus of regularity is simply $\sup_{y\in B}\{d(0,A^{-1}(y))\}$ where B is the unit ball. If the mapping Φ has a closed-convex graph, the Robinson–Ursescu Theorem says that Φ is metrically regular at $\bar{x}$ for $\bar{y}$ if and only if $\bar{y}$ is in the interior of the range of Φ. Metric regularity is also known to be equivalent to several others in variational analysis, namely the Aubin property of Φ^{-1} and the openness at linear rate of Φ. Additionally, metric regularity has been shown to be a generalization of the Eckart–Young result from matrix analysis on the distance to singularity of a matrix. Further, a result originating with Lyusternik and Graves [14,25] and extended by others (for example, [10,11,16]) shows that metric regularity is determined by the first-order behavior of a mapping and is preserved by sufficiently small first-order perturbations. Additional information about metric regularity and its relationship to other concepts in variational analysis can be found in [11,12,16], among others.

A central tool frequently appearing in variational analysis is that of the *normal cone* of a closed, convex set S. Specifically, the normal cone of S at $\bar{x}\in S$ can be defined as

$$N_S(\bar{x}):=\big\{x^*\in\mathcal{H}\colon\ \langle x^*,s-\bar{x}\rangle\leqslant 0\ \forall s\in S\big\}\tag{2.4}$$

and $N_S(\bar{x})=\emptyset$ if $\bar{x}\notin S$. Let $d(x,S)$ denote the distance from x to S, given by $d(x,S):=\inf_{s\in S}\|x-s\|$. Further, let $P_S(x)$ be the projection operator onto S, i.e., the set of such minimizers. If S is closed, convex and non-empty, then P_S is single-valued everywhere. Further, the projection operator is firmly non-expansive [9, Thm. 5.5] and can be characterized by

$$z=P_S(x)\quad\leftrightarrow\quad z\in S\text{ and }x-z\in N_S(z).\tag{2.5}$$

A method of characterizing regularity of closed sets $S_1,\dots,S_m$ is by considering regularity properties of a related set-valued mapping. Given a Hilbert space, $\mathcal{H}$, consider the product space $\mathcal{H}^m$ with the induced inner product defined by

$$\big\langle(x_1,x_2,\dots,x_m),(y_1,y_2,\dots,y_m)\big\rangle=\sum_{i=1}^m\langle x_i,y_i\rangle$$

[1] The definitions of metric subregularity (and strong subregularity, as discussed later) in [11] contain, in the right-hand side, the term $d(\bar{b},\Phi(x)\cap V)$ where V is a neighborhood of $\bar{b}$. However, as noted in [17], it can be easily verified that this condition is equivalent with our definition that removes V (though possibly with a different modulus γ).

and consider the set-valued mapping given by $\Phi(x) = [S_1 - x, \dots, S_m - x]^T$. Note that $0 \in \Phi(x)$ if and only if $x \in \bigcap_i S_i$. Using metric regularity as a starting point, suppose $\Phi(x)$ is metrically regular at $\bar{x}$ for 0. From the definition, metric regularity of Φ at $\bar{x}$ for 0 is equivalent to the *strong metric inequality*, examined in [19] and [20], among others, defined by the existence of $\beta, \delta > 0$ such that, for $i = 1, \dots, m$,

$$d\left(x, \bigcap_i (S_i - z_i)\right) \leqslant \beta \max_{1 \leqslant i \leqslant m} d(x + z_i, S_i) \quad \text{for all } x \in \bar{x} + \delta B,\ z_i \in \delta B. \tag{2.6}$$

Characterizing this in terms of normal cones, it was shown in [20, Thm. 1, Prop. 10, Cor. 2] that this is equivalent to the existence of a constant $k > 0$ such that

$$z_i \in \delta B,\ y_i \in N_{S_i}(\bar{x} + z_i)\ (i = 1, \dots, m) \quad \Rightarrow \quad \sum_i \|y_i\|^2 \leqslant k^2 \left\| \sum_i y_i \right\|^2. \tag{2.7}$$

By using the formula in [33, Thm. 9.43] for expressing the modulus of regularity in terms of coderivatives, it was shown in [23] that the modulus of regularity of Φ at $\bar{x}$ for 0 equals

$$\lim_{\delta \downarrow 0} \left\{ \inf \left\{ k\colon \text{ inequality (2.7) holds} \right\} \right\},$$

with this value being infinite being equivalent to a lack of metric regularity of Φ.

Consider a relaxed variant of the strong metric inequality, known simply as the *metric inequality* as studied in [16,20,28] among others, defined to hold at $\bar{x}$ if there exists $\beta > 0$ such that

$$d\left(x, \bigcap_i S_i\right) \leqslant \beta \max_{1 \leqslant i \leqslant m} d(x, S_i) \quad \text{for all } x \in \bar{x} + \delta B. \tag{2.8}$$

If inequality (2.8) is valid for $\delta = \infty$, we obtain the property of linear regularity and if it holds for all $\delta > 0$, it is equivalent to the property of bounded linear regularity, as studied in [3–7] and others, often in an algorithmic context. It is easy to show that the existence of a $\delta > 0$ such that inequality (2.8) holds is equivalent to the previously defined mapping Φ being metrically subregular at $\bar{x}$ for 0.

Our focus for the remainder of this paper will involve metric subregularity. Unfortunately, several of the stability properties and some of the geometric intuition that accompanies metric regularity—especially that relating to normal cones of sets—fail to have a natural equivalent for metric subregularity; some examples of this phenomenon are given in [12]. However, since metric regularity implies metric subregularity, the intuition provided by metric regularity can be applied to the following results when that property does, in fact, hold. Additionally, if the monotone operators under consideration are actually subdifferentials of convex functions, characterization of both metric regularity and subregularity in terms of the underlying function was shown in [2], providing additional intuition.

3. Metric regularity and linear convergence

We now return to problem (1.1), the problem of finding a zero of a maximal monotone operator. Variants of proximal point algorithms for solving this and related problems have been considered by a wide variety of authors, including [1,24, 29,32,34] and others.

Many authors consider an algorithmic framework much more general than the one considered in this paper. Some of the better-studied variants allow for a varying proximal parameter λ, allow approximate computation of the proximal iteration, allow over- or under-relaxation in the proximal step or incorporate an additional projective framework. These ideas have often proven worthwhile both for designing a computationally practical and efficient algorithm as well as for improving the convergence analysis. However, in this paper, we will only consider algorithms in their "classical" form, assuming exact computation of the resolvent with a fixed proximal parameter. Our particular interest is in exploring how naturally occurring constants—for example, the modulus of subregularity of the mappings themselves and of the mapping associated with the solution sets—govern the local rate of convergence and, further, how randomization as an analytical tool can emphasize this connection. To begin, consider the basic proximal point algorithm given by (1.3), where $x_{k+1} = J_{\lambda T}(x_k)$. Under an assumption of metric subregularity, we obtain the following initial result.

Theorem 3.1. *Suppose T is maximal monotone and metrically subregular at $\bar{x} \in T^{-1}(0)$ for 0 with subregularity modulus γ. Let $\bar{\gamma} > \gamma$ and suppose x_0 is sufficiently near $\bar{x}$. Then the iterates given by algorithm (1.3) are linearly convergent to $T^{-1}(0)$, the zero-set of T, satisfying*

$$d\big(x_{k+1}, T^{-1}(0)\big)^2 \leqslant \frac{\bar{\gamma}^2}{\lambda^2 + \bar{\gamma}^2} d\big(x_k, T^{-1}(0)\big)^2.$$

Proof. Let $\hat{x} \in T^{-1}(0)$ and note that $J_{\lambda T}(\hat{x}) = \hat{x}$. Since the resolvent of a maximal monotone operator is firmly non-expansive, it follows from inequality (2.1) that, for any x,

$$\left\| J_{\lambda T}(x) - J_{\lambda T}(\hat{x}) \right\|^2 \leqslant \|x - \hat{x}\|^2 - \left\| (I - J_{\lambda T})(x) - (I - J_{\lambda T})(\hat{x}) \right\|^2,$$

which can be equivalently written as

$$\left\| J_{\lambda T}(x) - \hat{x} \right\|^2 \leqslant \|x - \hat{x}\|^2 - \left\| x - J_{\lambda T}(x) \right\|^2. \tag{3.1}$$

However, by definition of $J_{\lambda T}$,

$$x - J_{\lambda T}(x) \in \lambda T\big(J_{\lambda T}(x)\big).$$

In particular,

$$\left\| x - J_{\lambda T}(x) \right\| \geqslant \lambda \min\big\{ \|z\|\colon z \in T\big(J_{\lambda T}(x)\big) \big\} = \lambda\, d\big(0, T\big(J_{\lambda T}(x)\big)\big). \tag{3.2}$$

Now, note that since the resolvents and projection operators are firmly non-expansive, if x_0 has the property of being sufficiently close to $\bar{x}$ such that inequality (2.3) holds with constant $\bar{\gamma}$, then x_j and $P_{T^{-1}(0)}(x_j)$ do as well for each $j \geqslant 0$. Therefore, it follows that

$$\begin{aligned}
d\big(x_{k+1}, T^{-1}(0)\big)^2 &\leqslant \left\| x_{k+1} - P_{T^{-1}(0)}(x_k) \right\|^2 \\
&\leqslant \left\| x_k - P_{T^{-1}(0)}(x_k) \right\|^2 - \left\| x_k - J_{\lambda T}(x_k) \right\|^2 \quad \big(\text{inequality (3.1)}\big) \\
&\leqslant d\big(x_k, T^{-1}(0)\big)^2 - \lambda^2 d\big(0, T\big(J_{\lambda T}(x_k)\big)\big)^2 \quad \big(\text{inequality (3.2)}\big) \\
&\leqslant d\big(x_k, T^{-1}(0)\big)^2 - \frac{\lambda^2}{\bar{\gamma}^2} d\big(J_{\lambda T}(x_k), T^{-1}(0)\big)^2 \quad \big(\text{inequality (2.3)}\big) \\
&= d\big(x_k, T^{-1}(0)\big)^2 - \frac{\lambda^2}{\bar{\gamma}^2} d\big(x_{k+1}, T^{-1}(0)\big)^2.
\end{aligned}$$

This implies that

$$\left(1 + \frac{\lambda^2}{\bar{\gamma}^2}\right) d\big(x_{k+1}, T^{-1}(0)\big)^2 \leqslant d\big(x_k, T^{-1}(0)\big)^2,$$

from which the result follows. $\square$

Further observe that by considering a sequence $\{\lambda_k\}$ such that $\lambda_k \to \infty$ instead of a fixed λ in the above algorithm, we obtain superlinear convergence.

Our primary interest in Theorem 3.1 is as a tool in proving the following result, Theorem 3.2. However, we note that Theorem 3.1 is similar to some previously known results. For example, in the paper by Rockafellar in [32], local linear convergence was shown under a framework that permitted error in evaluating the resolvent. In particular, [32, Thm. 2] showed that if $\bar{x}$ is the unique point satisfying $0 \in T(\bar{x})$ and if there exist $\gamma, \delta > 0$ such that

$$\|w\| \leqslant \delta,\ x \in T^{-1}(w) \quad \Rightarrow \quad \|x - \bar{x}\| \leqslant \gamma \|w\|,$$

then the algorithm satisfies $\|x_{k+1} - \bar{x}\|^2 \leqslant \frac{\gamma^2}{\lambda^2 + \gamma^2} \|x_k - \bar{x}\|^2$ in the special case of exact resolvent evaluation, similar to Theorem 3.1, though this assumption is stronger that metric subregularity. A result by Solodov and Svaiter in [34] uses a similar regularity assumption to show linear convergence in norm for a hybrid proximal-projection algorithm, with a different convergence rate.

From another perspective, Artacho, Dontchev and Geoffroy in [1] considered a highly generalized proximal iteration for metrically regular mappings, without the requirement that the mappings be monotone. By appealing to an appropriate fixed point result, they demonstrate the existence of a local linearly convergent sequence generated by the algorithm under conditions involving the Lipschitz constants of certain functions associated with the algorithm (these Lipschitz constants are, essentially, a generalized form of the proximal parameter λ). The authors then proceed to consider mappings, Φ, that are strongly subregular, essentially defined as metric subregularity at $\bar{x}$ for $\bar{b}$ where $\bar{x}$ is an isolated point of $\Phi^{-1}(\bar{b})$ (see [12] for additional information). In particular, they show that if the mapping T is strongly subregular at $\bar{x}$ for 0 (though not necessarily monotone), then the (generalized) proximal point algorithm is linearly convergent in norm, again under conditions involving the Lipschitz constants of certain associated functions. Additionally, under the assumptions of Theorem 3.1 with the condition of strong subregularity in place of metric subregularity, it can be verified that the conclusions of the theorem still hold with linear convergence in norm (to the unique zero, $\bar{x}$) in place of weak linear convergence.

We wish to generalize Theorem 3.1 to that of problem (1.2), finding a common zero among a group of maximal monotone operators, $T_1, \ldots, T_m$. Variants of proximal point algorithms for this problem have been considered by a variety of authors, including [8,15,18,21], among others. In what follows, consider the following randomized variant of a proximal point algorithm: for $k = 0, 1, 2, \dots,$

$$x_{k+1} = J_{\lambda T_i}(x_k) \quad \text{with probability } \frac{1}{m},\ i = 1, \dots, m. \tag{3.3}$$

Then we obtain the following result.

Theorem 3.2. *Suppose the following assumptions hold*:

1. *The maximal monotone operators* $\{T_i\colon i = 1, \dots, m\}$, *are metrically subregular at* $\bar{x} \in \bigcap_j T_j^{-1}(0)$ *for 0 with respective moduli* γ_i.
2. *The mapping* $\Phi(x) = [T_1^{-1}(0) - x, \dots, T_m^{-1}(0) - x]^T$ *is metrically subregular at* $\bar{x}$ *for* 0 *with modulus* κ.
3. $\bar{\gamma} > \max\{\gamma_i\colon i = 1, \dots, m\}$ *and* $\bar{\kappa} > \kappa$.

Then for x_0 *sufficiently close to* $\bar{x}$, *the sequence of iterates generated by algorithm* (3.3) *is linearly convergent in expectation to the common zero set,* $\bigcap_j T_j^{-1}(0)$, *satisfying*

$$\mathbf{E}\left[d\left(x_{k+1}, \bigcap_j T_j^{-1}(0)\right)^2 \,\Big|\, x_k\right] \leqslant \left(1 - \frac{1}{m\bar{\kappa}^2}\left[1 - \left(\frac{\bar{\gamma}^2}{\lambda^2 + \bar{\gamma}^2}\right)^{\frac{1}{2}}\right]^2\right) d\left(x_k, \bigcap_j T_j^{-1}(0)\right)^2.$$

Proof. If x_0 is sufficiently close to $\bar{x}$ such that inequality (2.3) holds with constant $\bar{\gamma}$ for each mapping T_i, it follows from the firm non-expansivity of the resolvents and the projection operator that each iterate x_k and the projection of each iterate onto the common zero set, $P_{\bigcap_j T_j^{-1}(0)}(x_k)$, are sufficiently close to $\bar{x}$ as well. Additionally, this implies the first conclusion of the theorem.

Suppose that at iteration k, the resolvent $J_{\lambda T_i}$ is chosen by the algorithm. Then it follows that

$$\begin{aligned}
d\left(J_{\lambda T_i}(x_k), \bigcap_j T_j^{-1}(0)\right)^2 &= \left\| J_{\lambda T_i}(x_k) - P_{\bigcap_j T_j^{-1}(0)}\big(J_{\lambda T_i}(x_k)\big)\right\|^2 \\
&\leqslant \left\| J_{\lambda T_i}(x_k) - P_{\bigcap_j T_j^{-1}(0)}(x_k)\right\|^2 \quad \text{(definition of projection)} \\
&\leqslant d\left(x_k, \bigcap_j T_j^{-1}(0)\right)^2 - \left\|x_k - J_{\lambda T_i}(x_k)\right\|^2 \quad \big(\text{inequality (2.3)}\big) \\
&= d\left(x_k, \bigcap_j T_j^{-1}(0)\right)^2 - \left\|\big[x_k - P_{T_i^{-1}(0)}(x_k)\big] + \big[P_{T_i^{-1}(0)}(x_k) - J_{\lambda T_i}(x_k)\big]\right\|^2 \\
&\leqslant d\left(x_k, \bigcap_j T_j^{-1}(0)\right)^2 - d\big(x_k, T_i^{-1}(0)\big)^2 - \left\|P_{T_i^{-1}(0)}(x_k) - J_{\lambda T_i}(x_k)\right\|^2 \\
&\quad - 2\big\langle x_k - P_{T_i^{-1}(0)}(x_k), P_{T_i^{-1}(0)}(x_k) - J_{\lambda T_i}(x_k)\big\rangle.
\end{aligned}$$

Note that

$$\begin{aligned}
-2\big\langle x_k - P_{T_i^{-1}(0)}(x_k), P_{T_i^{-1}(0)}(x_k) - J_{\lambda T_i}(x_k)\big\rangle &= 2\big\langle x_k - P_{T_i^{-1}(0)}(x_k), J_{\lambda T_i}(x_k) - P_{T_i^{-1}(0)}\big(J_{\lambda T_i}(x_k)\big)\big\rangle \\
&\quad + 2\big\langle x_k - P_{T_i^{-1}(0)}(x_k), P_{T_i^{-1}(0)}\big(J_{\lambda T_i}(x_k)\big) - P_{T_i^{-1}(0)}(x_k)\big\rangle \\
&\leqslant 2\big\langle x_k - P_{T_i^{-1}(0)}(x_k), J_{\lambda T_i}(x_k) - P_{T_i^{-1}(0)}\big(J_{\lambda T_i}(x_k)\big)\big\rangle \\
&\leqslant 2\left\|x_k - P_{T_i^{-1}(0)}(x_k)\right\| \left\| J_{\lambda T_i}(x_k) - P_{T_i^{-1}(0)}\big(J_{\lambda T_i}(x_k)\big)\right\| \\
&= 2d\big(x_k, T_i^{-1}(0)\big) d\big(J_{\lambda T_i}(x_k), T_i^{-1}(0)\big).
\end{aligned}$$

The first inequality comes from the fact that $x_k - P_{T_i^{-1}(0)}(x_k) \in N_{T_i^{-1}(0)}(P_{T_i^{-1}(0)}(x_k))$ and $P_{T^{-1}(0)}(J_{\lambda T_i}(x_k)) \in T_i^{-1}(0)$; therefore, inequality (2.4) can be applied from the definition of the normal cone. The second inequality is an application of the Cauchy–Schwartz inequality. The last equality follows from the definition of the projection operator. From this, we then obtain

$$\begin{aligned}d\Big(J_{\lambda T_i}(x_k),\bigcap_j T_j^{-1}(0)\Big)^2 &\leqslant d\Big(x_k,\bigcap_j T_j^{-1}(0)\Big)^2 - d\big(x_k,T_i^{-1}(0)\big)^2 - d\big(J_{\lambda T_i}(x_k),T_i^{-1}(0)\big)^2\\ &\quad + 2\,d\big(x_k,T_i^{-1}(0)\big)\,d\big(J_{\lambda T_i}(x_k),T_i^{-1}(0)\big)\\ &= d\Big(x_k,\bigcap_j T_j^{-1}(0)\Big)^2 - \big(d\big(x_k,T_i^{-1}(0)\big) - d\big(J_{\lambda T_i}(x_k),T_i^{-1}(0)\big)\big)^2.\end{aligned}$$

Noting that $d(x_k,T_i^{-1}(0)) - d(J_{\lambda T_i}(x_k),T_i^{-1}(0)) \geqslant 0$, it follows from an application of Theorem 3.1 that

$$d\Big(J_{\lambda T_i}(x_k),\bigcap_j T_j^{-1}(0)\Big)^2 \leqslant d\Big(x_k,\bigcap_j T_j^{-1}(0)\Big)^2 - \bigg[1-\Big(\frac{\bar{\gamma}^2}{\lambda^2+\bar{\gamma}^2}\Big)^{\frac12}\bigg]^2 d\big(x_k,T_i^{-1}(0)\big)^2.$$

Taking the expected value, we obtain

$$\begin{aligned}\mathbf{E}\bigg[d\Big(x_{k+1},\bigcap_j T_j^{-1}(0)\Big)^2 \mid x_k\bigg] - d\Big(x_k,\bigcap_j T_j^{-1}(0)\Big)^2 &\leqslant -\frac{1}{m}\bigg[1-\Big(\frac{\bar{\gamma}^2}{\lambda^2+\bar{\gamma}^2}\Big)^{\frac12}\bigg]^2 \sum_{i=1}^m d\big(x_k,T_i^{-1}(0)\big)^2\\ &= -\frac{1}{m}\bigg[1-\Big(\frac{\bar{\gamma}^2}{\lambda^2+\bar{\gamma}^2}\Big)^{\frac12}\bigg]^2 d\big(0,\Phi(x_k)\big)^2\\ &\leqslant -\frac{1}{m\bar{\kappa}^2}\bigg[1-\Big(\frac{\bar{\gamma}^2}{\lambda^2+\bar{\gamma}^2}\Big)^{\frac12}\bigg]^2 d\Big(x_k,\bigcap_j T_j^{-1}(0)\Big)^2,\end{aligned}$$

where the last inequality follows from the metric subregularity of the mapping $\Phi(x) = [T_1^{-1}(0)-x,\dots,T_m^{-1}(0)-x]^T$. □

It is worth noting that this type of convergence result implies that $d(x_k,\bigcap_j T_j^{-1}(0)) \to 0$ almost surely (cf. [22]).

The second assumption in the statement of Theorem 3.2 is a regularity condition on the zero sets of the individual maximal monotone operators. As discussed in Section 2, it is essentially equivalent to a local version of bounded linear regularity, a regularity condition which has been studied frequently in the context of projection algorithms for solving convex feasibility problems (see, for example, [3–7]). More generally, bounded linear regularity was also examined in [18] for demonstrating linear convergence of a projection-based algorithm for finding fixed points of firmly non-expansive mappings.

Regularity conditions on the mapping Φ itself have been previously studied in the context of projection algorithms (equivalently, in the case where the maximal monotone mappings are normal cone mappings). For example, the assumption regarding the metric subregularity of $\Phi(x)$ was examined in [22] in the context of a randomized projection algorithm. Further, if Φ is in fact metrically regular, then linear convergence of an averaged projections algorithm for certain classes of non-convex sets was demonstrated in [23].

One particularly simple way of de-randomizing algorithm (3.3) is by considering averaged resolvents or, in the terminology of [21], the *barycentric proximal method*. Specifically, given maximal monotone operators T_i, $i=1,\dots,m$ with respective resolvents $J_{\lambda T_i}$, $i=1,\dots,m$, consider the algorithm described such that, for $k=0,1,2,\dots,$

$$x_{k+1} = \frac{1}{m}\sum_{i=1}^m J_{\lambda T_i}(x_k) \tag{3.4}$$

and the associated fixed-point problem

$$\text{Find } x\in\mathcal{H} \quad \text{such that} \quad x = \frac{1}{m}\sum_{i=1}^m J_{\lambda T_i}(x). \tag{3.5}$$

The following proposition, found in [21], provides the necessary connection.

Proposition 3.3. *(See [21].) If $\bar{x} \in \bigcap_i T_i^{-1}(0)$, then $\bar{x}$ is a solution to problem* (3.5)*. Further, if $\bigcap_i T_i^{-1}(0) \neq \emptyset$, the fixed points of problem* (3.5) *are common fixed points of all the T_i's.*

Considering the example where each operator T_i is the normal cone mapping for some closed, convex set, it follows that algorithm (3.4) is simply the *averaged projections algorithm* studied by [3,22,23,30,31], among others. More generally, we can use the result of Theorem 3.2 to generalize a result on averaged projections found in [22, Thm. 5.8] to the barycentric proximal method.

Theorem 3.4. *Suppose the assumptions of Theorem* 3.2 *hold. Then the conclusions of Theorem* 3.2 *hold for algorithm* (3.4) *as well.*

Proof. Let x_k be the current iterate, x_{k+1}^{BP} be the new iterate in the barycentric proximal method, algorithm (3.4), and let x_{k+1}^{RP} be the new iterate in the randomized proximal point method, algorithm (3.3). First, note that since each set $T_i^{-1}(0)$ is convex, the distance function $d(\cdot, \bigcap_j T_j^{-1}(0))$ is as well, and

$$d\big(J_{\lambda T_i}(x_k), \bigcap_j T_j^{-1}(0)\big) \leqslant d\bigg(x_k, \bigcap_j T_j^{-1}(0)\bigg) \quad \text{for } i = 1, \dots, m,$$

from which it follows that

$$d\bigg(x_{k+1}^{BP}, \bigcap_j T_j^{-1}(0)\bigg) \leqslant d\bigg(x_k, \bigcap_j T_j^{-1}(0)\bigg).$$

Let $\alpha = 1 - \frac{1}{m\bar{\kappa}^2}[1 - (\frac{\bar{\gamma}^2}{\lambda^2+\bar{\gamma}^2})^{\frac{1}{2}}]^2$ and observe that the function $d(\cdot, \bigcap_j T_j^{-1}(0))^2$ is also convex. Noting that

$$x_{k+1}^{BP} = \frac{1}{m}\sum_{j=1}^{m} J_{\lambda T_j}(x_k) = \mathbf{E}\big[x_{k+1}^{RP} \bigm| x_k\big],$$

it follows that

$$\begin{aligned} d\bigg(x_{k+1}^{BP}, \bigcap_j T_j^{-1}(0)\bigg)^2 &= d\bigg(\mathbf{E}\big[x_{k+1}^{RP} \bigm| x_k\big], \bigcap_j T_j^{-1}(0)\bigg)^2 \\ &\leqslant \mathbf{E}\bigg[d\bigg(x_{k+1}^{RP}, \bigcap_j T_j^{-1}(0)\bigg)^2 \biggm| x_k\bigg] \\ &\leqslant \alpha d\bigg(x_k, \bigcap_j T_j^{-1}(0)\bigg)^2, \end{aligned}$$

from an application of Jensen's inequality. $\square$

In particular, the barycentric proximal method converges at least as quickly as the randomized proximal point method.

Acknowledgments

The author would like to thank Adrian Lewis and an anonymous referee for their helpful comments and suggestions on improving the presentation of this paper. Additionally, the author would like to thank Michael Todd for his comment that lead to an improvement of Theorem 3.2.

References

[1] F.J. Aragón Artacho, A.L. Dontchev, M.H. Geoffroy, Convergence of the proximal point method for metrically regular mappings, in: CSVAA 2004—Control Set-Valued Analysis and Applications, in: ESAIM Proc., vol. 17, EDP Sci., Les Ulis, 2007, pp. 1–8.
[2] F.J.A. Artacho, M.H. Geoffroy, Characterization of metric regularity of subdifferentials, J. Convex Anal. 15 (2) (2008) 365.
[3] H.H. Bauschke, J.M. Borwein, On the convergence of von Neumann's alternating projection algorithm for two sets, Set-Valued Anal. 1 (1993) 185–212.
[4] H.H. Bauschke, J.M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Rev. 38 (3) (1996) 367–426.
[5] H.H. Bauschke, J.M. Borwein, A.S. Lewis, The method of cyclic projections for closed convex sets in Hilbert space, in: Recent Developments in Optimization Theory and Nonlinear Analysis: Ams/Imu Special Session on Optimization and Nonlinear Analysis, vol. 2, May 24–26, 1995, Jerusalem, Israel, American Mathematical Society, 1997.
[6] H.H. Bauschke, J.M. Borwein, P. Tseng, Bounded linear regularity strong CHIP and CHIP are distinct properties, J. Convex Anal. 7 (2) (2000) 395–412.
[7] A. Beck, M. Teboulle, Convergence rate analysis error bounds for projection algorithms in convex feasibility problems, Optim. Methods Softw. 18 (4) (2003) 377–394.
[8] P.L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, Optimization 53 (5) (2004) 475–504.
[9] F. Deutsch, Best Approximation in Inner Product Spaces, Springer-Verlag, New York, 2001.
[10] A.V. Dmitruk, A.A. Milyutin, N.P. Osmolovskii, Lyusternik's theorem the theory of extrema, Russian Math. Surveys 35 (6) (1980) 11–51.
[11] A.L. Dontchev, A.S. Lewis, R.T. Rockafellar, The radius of metric regularity, Trans. Amer. Math. Soc. 355 (2) (2003) 493–517.
[12] A.L. Dontchev, R.T. Rockafellar, Regularity conditioning of solution mappings in variational analysis, Set-Valued Anal. 12 (2004) 79–109.
[13] J. Eckstein, Splitting methods for monotone operators with applications to parallel optimization, PhD thesis, Massachusetts Institute of Technology, 1989.
[14] L.M. Graves, Some mapping theorems, Duke Math. J. 17 (1) (1950) 111–114.
[15] S.A. Hirstoaga, Iterative selection methods for common fixed point problems, J. Math. Anal. Appl. 324 (2) (2006) 1020–1035.
[16] A.D. Ioffe, Metric regularity subdifferential calculus, Russian Math. Surveys 55 (3) (2000) 501–558.
[17] A.D. Ioffe, J.V. Outrata, On metric calmness qualification conditions in subdifferential calculus, Set-Valued Anal. 16 (2) (2008) 199–227.

[18] K.C. Kiwiel, B. Lopuch, Surrogate projection methods for finding fixed points of firmly nonexpansive mappings, SIAM J. Optim. 7 (4) (1997) 1084–1102.
[19] A. Kruger, Stationarity regularity of set systems, Pac. J. Optim. 1 (2005) 101–126.
[20] A.Y. Kruger, About regularity of collections of sets, Set-Valued Anal. 14 (2) (2006) 187–206.
[21] N. Lehdili, B. Lemaire, The barycentric proximal method, Comm. Appl. Nonlinear Anal. 6 (1999) 29–47.
[22] D. Leventhal, A.S. Lewis, Randomized methods for linear constraints: Convergence rates and conditioning, submitted for publication; available online at: http://arxiv.org/pdf/0806.3015.
[23] A.S. Lewis, D.R. Luke, J. Malick, Local convergence for alternating averaged nonconvex projections, Found. Comput. Math. 9 (4) (August 2009).
[24] F.J. Luque, Asymptotic convergence analysis of the proximal point algorithm, SIAM J. Control Optim. 22 (1984) 277.
[25] L.A. Lyusternik, Conditional extrema of functionals, Mat. Sb. 41 (3) (1934) 390–401.
[26] B. Martinet, Regularisation d'inequations variationnelles par approximations successives, Rev. Fr. Inform. Rech. Oper. 4 (1970) 154–159.
[27] G.J. Minty, Monotone (nonlinear) operators in Hilbert space, Duke Math. J. 29 (3) (1962) 341–346.
[28] H.V. Ngai, M. Théra, Metric inequality subdifferential calculus and applications, Set-Valued Anal. 9 (1) (2001) 187–216.
[29] T. Pennanen, Local convergence of the proximal point algorithm and multiplier methods without monotonicity, Math. Oper. Res. 27 (1) (2002) 170.
[30] G. Pierra, Decomposition through formalization in a product space, Math. Program. 28 (1984) 96–115.
[31] S. Reich, A limit theorem for projections, Linear Multilinear Algebra 13 (3) (1983) 281–290.
[32] R.T. Rockafellar, Monotone operators the proximal point algorithm, SIAM J. Control Optim. 14 (5) (1976) 877–898.
[33] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer-Verlag, Berlin, 1998.
[34] M.V. Solodov, B.F. Svaiter, A hybrid projection-proximal point algorithm, J. Convex Anal. 6 (1999) 59–70.