Generic Linear Convergence Through Metric Subregularity in a Variable-Metric Extension of the Proximal Point Algorithm

R. Tyrrell Rockafellar[1]

Abstract

The proximal point algorithm finds a zero of a maximal monotone mapping by iterations in which the mapping is made strongly monotone by the addition of a proximal term. Here it is articulated with the norm behind the proximal term possibly shifting from one iteration to the next, but under conditions that eventually make the metric settle down. Despite the varying geometry, the sequence generated by the algorithm is shown to converge to a particular solution. Although this is not the first variable-metric extension of proximal point algorithm, it is the first to retain the flexibility needed for applications to augmented Lagrangian methodology and progressive decoupling. Moreover, in a generic sense, the convergence it generates is Q-linear at a rate that depends in a simple way on the modulus of metric subregularity of the mapping at that solution. This is a tighter rate than previously identified and reveals for the first time the definitive role of metric subregularity in how the proximal point algorithm performs, even in fixed-metric mode.

Keywords: *maximal monotone mappings, proximal point algorithm, variable-metric implementation, localized executability, linear convergence guarantees, metric subregularity, convex optimization*

Version of 5 May 2023

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*, URL: sites.math.washington.edu/~rtr/mypage.html

1 Introduction

The proximal point algorithm, PPA, solves problems of optimization by way of optimality conditions posed as "generalized equations." In its original formulation in [6] for a Hilbert space $\mathcal{H}$ and a set-valued mapping $T : \mathcal{H} \rightrightarrows \mathcal{H}$, represented by the set $\operatorname{gph} T = \{\, (z, w) \,|\, w \in T(z)\}$, it seeks to

$$\text{determine } \bar{z} \in Z := T^{-1}(0), \text{ i.e., such that } T(\bar{z}) \ni 0. \tag{1.1}$$

It relies for this on T being *maximal monotone*, where monotonicity is the property that

$$\langle w_1 - w_0, z_1 - z_0 \rangle \geq 0 \text{ for all pairs } (z_0, w_0),\, (z_1, w_1) \text{ in } \operatorname{gph} T, \tag{1.2}$$

with $\langle \cdot, \cdot \rangle$ being the inner product of $\mathcal{H}$, and maximality refers to the nonexistence of a monotone mapping $T' : \mathcal{H} \rightrightarrows \mathcal{H}$ such that $\operatorname{gph} T'$ properly extends $\operatorname{gph} T$. A prime example of a maximal monotone mapping is $T = \partial f$ for a closed proper convex function f on $\mathcal{H}$, in which case a solution to problem (1.1) is a global minimizer of f. Other examples in convex optimization involve dual variables and lead to augmented Lagrangian techniques; see [6], [7], [13].

Especially motivating for the efforts we undertake here are the applications of the PPA to problem decomposition by way of the *progressive decoupling algorithm* in [9, 10]. Under a new sufficient condition for local optimality developed in [10, 12], even nonconvex problems of optimization can be decomposed by taking advantage of the observation of Pennanen [5] that the proximal point algorithm can operate in a neighborhood of a solution without using more than a local portion of the graph of the mapping T. However, the catch so far is that convergence of the progressive decoupling algorithm has only been validated for *error-free* iterations in which the proximal parameter has a value *fixed from the start* — because the geometry in $\mathcal{H}$ is anchored by it. This goes back to origins in Spingarn's *method of partial inverses* [17], where the proximal parameter actually had to equal 1. Flexibility in adjusting the parameter, or perhaps a multiplicity or proximal parameters tailored to different subproblems in a decomposition scheme, is important for understanding when linear convergence might be expected, and how the rate might influenced. Identification of stopping criteria that are implementable in an application is, of course, crucial to theory in exploring intrinsic robustness with respect to error.

The needs can be met by a variable-metric extension of the original proximal point algorithm that can fully accommodate the delicate circumstances in progressive decoupling while clearly being articulated in a manner that only draws on local monotonicity. Our aim is to provide that here for *finite-dimensional* $\mathcal{H}$ along with new insights into linear convergence analysis from a general perspective.

Variable-metric offshoots of the PPA in finite dimensions have already been devised, starting with Burke and Qian [1] (1999) under some seriously limiting assumptions, but later most interestingly by Parente, Lotito and Solodov in [4] (2008). Their method has significant advantages for some applications, both in executability and allowing for approximations of the mapping T. However, its provisions for inexactness in computations are so different from those in the original PPA that the two approaches coincide only when there is no inexactness at all! Moreover, those provisions in [4] take a form that is generally incompatible with producing implementable stopping criteria in progressive decoupling applications, in contrast to the provisions in [6], which do support that. Also, no attention was placed in [4] on whether just local maximality of T would be enough, which has now emerged as a key issue.

We made a start in [11] on a variable-metric PPA version that fulfills the listed requirements, but now add significant improvements. We definitively tie linear convergence to *metric subregularity* at the solution. The rate of linear convergence is revealed in fact to be a simple function of the modulus in that property (Theorem 2.2), and it is tighter than rates identified in the past, *even for fixed-metric PPA schemes*. Instead, the variable-metric PPA in [4], covering the classical iterations only in *exact* execution as already mentioned, relies for getting linear convergence on assuming a solution set property that is more restrictive, not just local, and unable to yield our tight rate. The convergence rate in [4, Theorem 4.4] comes out furthermore as an opaque function of several algorithmic parameters whose influence can only be gleaned from background arguments.[2] In addition, the claim in [4, Theorem 4.4] that linear convergence is obtained with respect to the limit point, rather than the distance to the solution set as in other works on the subject, seems to be in error.[3] Our results thus simultaneously complement, sharpen and correct the current convergence picture.

Properties of metric regularity, which furnish estimates useful in solving generalized equations for set-valued mappings, are explored from many angles in the book [2] of Asen Dontchev and the present writer.[4] Their relevance to proximal point theory has largely gone unrecognized as such, but will be forcefully brought out here. The crucial condition will be seen to be *metric subregularity* of T at $\bar{z}$ with respect to an element $\bar{w} \in T(\bar{z})$, according to which $\operatorname{dist}(z, T^{-1}(\bar{w})) \leq a \operatorname{dist}(\bar{w}, T(z))$ holds locally for z near $\bar{z}$ and some a, or as known to be equivalent and will be more convenient here,

$$\begin{array}{l} \text{for some neighborhood } \mathcal{N} \text{ of } (\bar{z},\bar{w}),\ \exists\, a \in (0,\infty) \ \text{ such that} \\ (z,w) \in \mathcal{N},\ w \in T(z) \quad\Longrightarrow\quad \operatorname{dist}(z, T^{-1}(\bar{w})) \leq a||w-\bar{w}||. \end{array} \tag{1.3}$$

The corresponding *modulus of subregularity* is

$$\operatorname{subreg}(T; \bar{z}\,|\,\bar{w}) := \liminf \text{ of } a \text{ values in (1.3) as the neighborhood shrinks}, \tag{1.4}$$

with the absence of subregularity being indicated then by $\operatorname{subreg}(T^{-1} : \bar{w}\,|\,\bar{z}) = \infty$. The property in (1.3) is called *strong* metric subregularity if, in addition, $\bar{z}$ is an isolated point of $T^{-1}(\bar{w})$, having a neighborhood that contains no other point of $T^{-1}(\bar{w})$. But under maximal monotonicity $T^{-1}(\bar{w})$ is convex, so that's the same as $T^{-1}(\bar{w}) = \{\bar{z}\}$.

The condition of metric subregularity can be understood from the angle that $T^{-1}(\bar{w})$ is the solution set for finding $\bar{z}$ with $T(\bar{z}) \ni \bar{w}$, and w is a perturbation away from $\bar{w}$, making $T^{-1}(w)$ a correspondingly perturbed solution set away from $T^{-1}(\bar{w})$. When w is near enough to $\bar{w}$, perturbed solutions $z \in T^{-1}(w)$ near enough to $\bar{z}$ are required by (1.3) to have their distance from being a true solution bounded in a linear way by the distance of w from $\bar{w}$. In connection with (1.1), the case where $\bar{w} = 0$ is at the fore. However, it will be important for later developments about genericity to keep in mind that we could be looking at that problem as *parameterized* by $\bar{w}$, with $\bar{w} = 0$ just a convenient "normalization" for handling a representententative instance of it.

The original result on linear convergence of the PPA, in [6], concerned Q-linear convergence of $||z^k - \bar{z}||$ to 0 at a rate $r \in [0,\infty)$, which means that[5]

$$\forall \varepsilon > 0,\ \exists k_\varepsilon \ \text{ such that } \ ||z^{k+1} - \bar{z}|| \leq (r+\varepsilon)||z^k - \bar{z}|| \ \text{ when } \ k > k_\varepsilon. \tag{1.5}$$

The extra condition in getting it was actually the *strong metric subregularity* of T at $\bar{z}$ with respect to having $0 \in T(\bar{z})$, although that concept and terminology didn't exist then. Its demand for $\bar{z}$ to be the

[2]This rate is given by the value ν at the top of page 254 of [4], which is indicated as depending on a value μ and a stepsize relaxation parameter $\theta \in (0,1)$. But $\mu = \sqrt{\alpha^2+1}\sqrt{\beta^2-1} + \alpha\beta$ where α already depends on θ and both α and β depend also on another parameter $\bar{\sigma}$.

[3]The proof says this follows from Fejér monotonicity, but offers no evidence for Fejér monotonicity, which is indeed doubtful for the situation at hand.

[4]This paper is dedicated to the memory of Asen Dontchev, who died 16 September 2021. Metric regularity was a subject very dear to him.

[5]By avoiding a ratio, this accommodates the possibility that $||z^k - \bar{z}||$ might sometimes be 0 but not stay at 0.

unique solution to (1.1) was perceived to be a limitation indicating the need for more work. Luque in [3] provided a substitute condition that relaxed the uniqueness, at the cost of getting, instead of Q-linear convergence of $||z^k - \bar{z}||$ to 0, that of $\mathrm{dist}(z^k, Z)$ to 0. But his condition had hidden shortcomings, noted in [11], because it depended on the entire (possibly even unbounded) convex solution set, not just a part of it near $\bar{z}$.[6] We were able, though, to demonstrate in [11] that a weaker form of Luque's condition, recognized here as amounting to plain *metric subregularity* (not strong), was already enough to ensure his linear convergence when the problem had more than one solution. The replacement of Luque's condition by one that is local to the graph of T is of course a crucial step toward confirming that PPA iterations can succeed on the basis of local information only.

Going further in [11], we came up with a condition on $\bar{z}$ that recovered the Q-linear convergence of $||z^k - \bar{z}||$ to 0 without the solution set needing to be a singleton, at least under tighter control of approximations through the stopping criterion that governs them. In contrast to metric subregularity, this development seems too fragile for translation to the variable-metric PPA format with inexactness, but that may not matter very much, in light of recent progress in [14].

There, in specializing the theory of metric regularity properties to maximal monotone mappings, it was shown that *not just metric subregularity, but even strong metric subregularity is in fact generic*, in a sense that will be clarified here in due course. The import is that, if we enlarge the picture (1.1) by replacing $T(\bar{z}) \ni 0$ by $T(\bar{z}) \ni \bar{w}$ with $\bar{w}$ as parameter, or equivalently replacing T by $T_{\bar{w}} : z \mapsto T(z) - \bar{w}$ with corresponding solution set $Z_{\bar{w}}$, and add a natural assumption such as $Z_{\bar{w}}$ being nonempty and bounded for at least one $\bar{w}$, then strong metric regularity will *typically* be on hand and Q-linear convergence of z^k to $\bar{z}$ will therefore prevail.

The ingredients of the variable-metric PPA under consideration in this paper, besides the maximally monotone mapping T, are the usual proximal parameters $c_k > 0$, assumed to satisfy

$$1 \le c_k \to c_\infty \le \infty, \tag{1.6}$$

and linear mappings

$$B_k : \mathcal{H} \to \mathcal{H} \ \text{ self-adjoint and positive-definite} \tag{1.7}$$

with their associated inner products and norms,

$$\langle w, z \rangle_{B_k} = \langle w, B_k z \rangle = \langle B_k w, z \rangle, \qquad ||z||^2_{B_k} = \langle z, B_k z \rangle. \tag{1.8}$$

The algorithm generates a sequence of points z^k from an initial choice of z^0 by iterations that in principle are of the form

$$z^{k+1} \approx P_k(z^k) := (I + c_k B_k^{-1} T)^{-1}(z^k), \tag{1.9}$$

but equivalently correspond to solving modifications of the generalized equation in (1.1), namely

$$\text{determine } \ z^{k+1} \ \text{ such that } \ S_k(z^{k+1}) \approx\ni 0 \ \text{ for } \ S_k(z) = T(z) + c_k^{-1} B_k[z - z^k]. \tag{1.10}$$

Here "≈" refers to approximation. The monotonicity of T in (1.2) translates to the monotonicity of $B_k^{-1} T$ in the B_k inner product in (1.8), with maximality preserved, and that makes the mappings P_k in (1.9) be not only single-valued but nonexpansive in the B_k norm:

$$||P_k(z') - P_k(z)||_{B_k} \le ||z' - z||_{B_k} \ \text{ for all } \ z,\, z'. \tag{1.11}$$

[6]For the set R that is the closure of the range of T, this being convex because of T being maximal monotone, the normal cone at 0 must contain the normal cones at all points of R in some neighborhood of 0, as holds for instance when R is polyhedral.

When $T = \partial f$ for a closed proper convex function f, the iterations in form (1.10) signify that

$$z^{k+1} \approx\in \operatorname{argmin}_z \left\{ f(z) + \frac{1}{2c_k} ||z - z^k||^2_{B_k} \right\} \tag{1.12}$$

and thus have the evident potential of tuning the proximal term to the second-order properties of f, at least to some degree, perhaps progressively in quasi-Newton fashion. But a bigger motivation for our variable-metric extension, as explained, is furnishing support for the approaches to problem decomposition in [9] and [10]; applications to that will be taken up in [15]. The convergence results obtained here will exceed the ones in [11] in meeting those needs of support.

In (1.9), approximation refers to the distance between z^{k+1} and $P_k(z^k)$, but it will be taken in the B_k-norm. In (1.10), approximation is based on the distance of 0 from the set $S_k(z^{k+1})$, again in the B_k norm. The second kind of approximation is usually easier in practice and provides a convenient estimate for the first kind, through

$$||z^{k+1} - P_k(z^k)||_{B_k} \;\leq\; c_k \operatorname{dist}_{B_k}(0, B_k^{-1} S_k(z^{k+1})) \;\leq\; c_k ||B_k^{-1}|| \operatorname{dist}(0, S_k(z^{k+1})), \tag{1.13}$$

see [11, (4.10)]. In the case of $T = \partial f$, for example, in getting the update z^{k+1} from z^k via (1.12), has $S_k(z^{k+1})$ being the set of subgradients at z^{k+1} of the objective in (1.12). The test for z^{k+1} being good enough as an approximate minimizer is then the existence of a subgradient at z^{k+1} that is close enough to 0.

Stopping criteria for these approximations furnish the standards for "close enough." They utilize error parameters ε_k, assumed to satisfy

$$\varepsilon_k \in (0,1) \;\text{ with }\; \sum\nolimits_{k=1}^{\infty} \varepsilon_k < \infty, \tag{1.14}$$

and they can be invoked at two levels of tightness:

$$||z^{k+1} - P_k(z^k)||_{B_k} \leq \begin{cases} \text{(a)} & \varepsilon_k \\ \text{(b)} & \varepsilon_k \min\{1, ||z^{k+1} - z^k||_{B_k}\}. \end{cases} \tag{1.15}$$

The estimate in (1.13) allows these to be replaced by

$$c_k \operatorname{dist}_{B_k^{-1}}(0, S_k(z^{k+1})) \leq \begin{cases} \text{(a)} & \varepsilon_k \\ \text{(b)} & \varepsilon_k \min\{1, ||z^{k+1} - z^k||_{B_k}\}, \end{cases} \tag{1.16}$$

where distance in the underlying norm could substitute for the distance in the B_k^{-1} norm through having

$$\operatorname{dist}_{B_k^{-1}}(z, S) \leq ||B_k^{-1}|| \operatorname{dist}(z, S). \tag{1.17}$$

These criteria in the fixed-metric case, $B_k \equiv I$, have been part of PPA theory from the start, with (a) adequate for simple convergence of z^k to some $\bar{z}$ and (b) entering in support of linear convergence. That general picture will remain here, but for the variable-metric PPA to behave something like the original fixed-metric PPA, the successive changes in norm must not be too wild. Here we improve the conditions about that in [11, Section 4] and the view of how the changes in metric may be generated.

Similar provisions for taming the switches in metric were figured out by Parente, Lotito and Solodov for their variable metric PPA in [4]; we overlap in that respect. But their approach to inexact iterations, motivated by applications other than the ones we have in mind, was distinctly different and only admits (1.15) and (1.16) when $\varepsilon_k = 0$. Instead of the update (1.9), it takes (in simplest implementation[7]) first $\hat{z}^k \approx P_k(z^k)$ and then $z^{k+1} = z^k - u^k$, where u^k is a variable-metric projection of $z^k - \hat{z}^k$ in the direction of a vector $\hat{v}^k \in T(\hat{z}^k)$. While that may be an excellent innovation for many reasons, it's incompatible with the maneuvers associated with the subspace decomposition behind algorithms evolved from Spingarn's partial inverse method. There, in optimization for instance, the PPA is essentially being applied to finding a local convex-concave saddle point in a subproblem, the dual component of which can be a known function of the primal component. However, the scheme in [4] just described can't take advantage of that and insists instead on a joint primal-dual approximation which doesn't fit with an implementable stopping criterion.

Our main results about convergence will be established in Section 2 after some preliminaries about B_k relationships. They refer to convergence from any starting point z^0 in $\mathcal{H}$. Section 3 takes up the localized alternative, which concerns starting from z^0 *sufficiently close* to the solution set Z. That's important because the algorithm is thereby shown capable of acting entirely on local information, and the mappings T in some attractive applications to nonconvex optimization exhibit maximal monotonicity only in a *local* form. For $T = \partial f$, that local form of maximal monotonicity corresponds to *variational convexity* of f with respect to a point-subgradient pair [8]. Moreover, it relates deeply to the linkage problems in [10] and the scheme there of solving them by "progressive decoupling."

2 Global convergence results

From the general theory of monotonicity, available for background in [16, Chapter 12], the maximal monotonicity of $B_k^{-1}T$ in the B_k norm, which carries over to $c_k B_k^{-1}T$, implies not just that the resolvent mapping P_k in (1.9) is B_k-nonexpansive as in (1.11), but that it is *firmly nonexpansive* in the B_k norm:

$$||P_k(z') - P_k(z)||_{B_k}^2 + ||Q_k(z') - Q_k(z)||_{B_k}^2 \le ||z' - z||_{B_k}^2, \text{ where } Q_k = I - P_k. \tag{2.1}$$

The solution set Z in (1.1) is the set of fixed points of P_k, so in implementing (1.9) exactly by taking $z^{k+1} = P_k(z^k)$, we would have $\operatorname{dist}_{B_k}(z^{k+1}, Z) < \operatorname{dist}_{B_k}(z^k, Z)$, unless $z^k \in Z$. This would seem to suggest that the z^k sequence will get ever closer to Z, but beside trouble that might come from approximations, there could be confusion because the standard for nearness keeps changing. To counter that, an assumption is needed that forces the distance standards to ultimately stabilize.

For that, we can start by taking advantage of the fact that, because $\mathcal{H}$ is finite-dimensional, the different norms are related to each other by estimates of the form

$$\alpha_k \ge \frac{||z||_{B_k}}{||z||_{B_{k-1}}} \ge \alpha_k^{-1}, \qquad \alpha_0 \ge \frac{||z||_{B_0}}{||z||} \ge \alpha_0^{-1}. \tag{2.2}$$

It is insightful, however, to look at this from the perspective of B_k being generated as computations proceed, rather than having been fixed in advance. *Instead of thinking of α_k in (2.2) as derived after B_k has been chosen to replace B_{k-1}, we can think of it as coming before and exercising some control over that choice.* In other words, α_k can be considered as a parameter provided in the implementation of the algorithm, like c_k in (1.6) and ε_k in (1.14). A condition placed on the α_k sequence can serve that way to keep the repeated changes in norm from getting out of hand. Here,

> the parameter values $\alpha_k \in [1, \infty)$ are assumed to converge to 1 fast enough that the increasing products $\beta_k := \alpha_k \alpha_{k-1} \cdots \alpha_1 \alpha_0$ stay bounded: $\beta_k \nearrow \beta_\infty := \Pi_{k=0}^\infty \alpha_k < \infty$. (2.3)

[7]With their parameters $\theta = 0$ and $\varepsilon_k = 0$; our B_k corresponds to their M_k^{-1}. However they actually require $\theta \in (0, 1)$!

To see it another way, the values $\log \alpha_k \in [0,\infty)$ are assumed to be summable, in a sort of echo of (1.14). The choice of the successor B_k to B_{k-1} in each iteration, being constrained by (2.2), will accordingly be forced to settle down.

In [11], we only looked at upper bounds of the form in (2.2) when imposing (2.3). Such upper bounds were utilized also by Parente, Lotito and Solodov in earlier work in a variable-metric setting [4] under a slightly stronger assumption than (2.3). We deployed separate conditions in [11] to make the inverses B_k^{-1} behave well enough. However, the symmetry in (2.2) has significant advantages, which we now lay out.

Proposition (helpful properties of the variable-metric scheme)**.** *The inequalities in (2.2) are symmetric with respect to taking inverses; they have the equivalent form*

$$\alpha_k \geq \frac{||z||_{B_k^{-1}}}{||z||_{B_{k-1}^{-1}}} \geq \alpha_k^{-1}, \qquad \alpha_0 \geq \frac{||z||_{B_0^{-1}}}{||z||} \geq \alpha_0^{-1}. \tag{2.4}$$

The inequalities in (2.2), chained together, also imply

$$\beta_k \geq \frac{||z||_{B_k}}{||z||} \geq \beta_k^{-1}, \quad \beta_k \geq \frac{||z||_{B_k^{-1}}}{||z||} \geq \beta_k^{-1}, \text{ hence } \max\{||B_k||, ||B_k^{-1}||\} \leq \beta_k^2 \leq \beta_\infty^2. \tag{2.5}$$

Moreover, the combination of (2.2) and (2.3) ensures that

$$B_k \to B_\infty \text{ self-adjoint and positive-definite,} \tag{2.6}$$

and then

$$\frac{\beta_\infty}{\beta_k} \geq \frac{||z||_{B_\infty}}{||z||_{B_k}} \geq \frac{\beta_k}{\beta_\infty}, \qquad \frac{\beta_\infty}{\beta_k} \geq \frac{||z||_{B_\infty^{-1}}}{||z||_{B_k^{-1}}} \geq \frac{\beta_k}{\beta_\infty}. \tag{2.7}$$

Proof. Because B_k is selfadjoint and positive-definite, there is a unique mapping $B_k^{1/2}$ that is likewise symmetric and positive-definite and has $(B_k^{1/2})^2 = B_k$ (as seen through eigenvalues and diagonalization). Thinking of (2.2) as requiring the range of the ratio $\langle z, B_k z\rangle / \langle z, B_{k-1} z\rangle$ to lie in the interval $[\alpha_k^{-2}, \alpha_k^2]$, we can translate that through the change of variables $u = B_{k-1}^{1/2}$ to requring that the range of the ratio $\langle z, B_{k-1}^{-1/2} B_k B_{k-1}^{-1/2} z\rangle / ||z||^2$ lies in that interval. For any self-adjoint positive-definite B, the range of $\langle z, Bz\rangle/||z||^2$ is the interval between $||B^{-1}||^{-1}$ and $||B||$, so the bounds in question are equivalent to

$$||B_{k-1}^{-1/2} B_k B_{k-1}^{-1/2}|| \leq \alpha_k \text{ and } ||(B_{k-1}^{-1})^{-1/2} B_k^{-1} (B_{k-1}^{-1})^{-1/2}|| \leq \alpha_k$$

and thus are symmetric with respect to the B_k sequence to the B_k^{-1} sequence. This proves (2.4).

The first part of (2.5) is evident from the definition of β_k in (2.3). The rest is seen from expressing the bounds as $\langle z, B_k z\rangle/||z||^2 \leq \beta_k^2$ and $||z||^2/\langle z, B_k z\rangle \leq \beta_k^2$, then through the change of variables $w = B_k^{1/2} z$ identifying the second ratio with $\langle w, B_k^{-1} w\rangle / ||w||^2$.

To demonstrate (2.6), we show that the combination of (2.2) and (2.3) forces $\{B_k\}_{k=0}^\infty$ be a Cauchy sequence, having norms $||B_k - B_j||$ smaller than any ε when j is high enough and $k > j$. From $||B_k - B_j||$ being the maximum of $\langle z, [B_k - B_j] z\rangle$ subject to $||z|| = 1$ and the representation

$$\langle z, [B_k - B_j] z\rangle = \left[\frac{\langle z, B_k z\rangle}{\langle z, B_{k-1} z\rangle} \frac{\langle z, B_{k-1} z\rangle}{\langle z, B_{k-2} z\rangle} \cdots \frac{\langle z, B_{j+1} z\rangle}{\langle z, B_j z\rangle} - 1 \right] \langle z, B_j z\rangle,$$

we get the estimate that $||B_k - B_j|| \le (\alpha_k^2\alpha_{k-1}^2 \cdots \alpha_{j+1}^2 - 1)\beta_\infty^2$ through (2.2) and (2.5). This upper bound can be made arbitrarily small by taking j and k high enough, due to the convergence assumed in (2.3).

With the existence of B_∞ in hand, the inequalities in (2.2) and (2.4) can be propagated to higher indices to get (2.7) through the observation that $\Pi_{j=k+1}^\infty \alpha_j = \beta_\infty/\beta_k$. □

Example 2.1 (variable metrics based on rescaling). *Suppose $\mathcal{H}$ is $I\!R^n$ with $z = (z_1, \ldots, z_n)$, and let the quadratic functions $\langle z, B_k z\rangle$ have the form $\lambda_{1,k}z_1^2 + \cdots + \lambda_{n,k}z_n^2$. Updating from B_{k-1} to B_k corresponds then to updating $\lambda_{j,k-1}$ to $\lambda_{j,k}$, and the constraint on this imposed by the algorithmic parameters α_k in (2.2) is just that*

$$\alpha_k^2 \ge \frac{\lambda_{j,k}}{\lambda_{j,k-1}} \ge \alpha_k^{-2} \text{ for } j = 1, \ldots, n.$$

This corresponds to requiring $|\log\lambda_{j,k} - \log\lambda_{j,k-1}| \le \mu_k$ for parameters $\mu_k \ge 0$ with $\sum_{k=1}^\infty \mu_k < \infty$.

Detail. This is seen by taking squares in (2.2) and calculating that the maximum over z is the highest ratio $\lambda_{j,k}/\lambda_{j,k-1}$ while the minimum is the lowest. □

The basic convergence result we now formulate is essentially a specialization of [11, Theorem 4.1]. That theorem, though, was aimed at establishing local convergence when T is only assumed to be maximal monotone locally, and that added extra complications to its statement and proof. We therefore provide for the global setting at hand a straightforward and leaner proof which mimics the local one, but also adapts to our improved assumptions on the mappings B_k. (This also gets around notational shifts and a small glitch after [11, (4.7)] that misidentified the bound obtained on $||B_k||$ as β_k instead of β_k^2 as in (2.5).)

Theorem 2.1 (global convergence to a solution). *For T maximal monotone globally with the solution set $Z = T^{-1}(0)$ nonempty,*[8] *let the proximal point algorithm (1.9) or equivalently (1.10), be initiated under (1.6), (1.7), (2.2), and (2.3), from any $z^0 \in \mathcal{H}$ under the stopping criterion (1.15a), or (1.16a). Then the sequence of iterates z^k is sure to converge to some particular $\bar{z} \in Z$.*

Proof. Let $\bar{z}^0$ be the projection of z^0 on the closed convex set Z, noting that $P_k(\bar{z}^0) = \bar{z}^0$ for the mapping P_k in (1.9), which is nonexpansive in the B_k norm by (1.11). Because $||z^{k+1} - \bar{z}^0||_{B_k} \le ||z^{k+1} - P_k(z^k)||_{B_k} + ||P_k(z^k) - \bar{z}^0||_{B_k}$ with $||z^{k+1} - P_k(z^k)||_{B_k} \le \varepsilon_k$ by the stopping criterion (1.15a) (as implied also by (1.16a)), and $||P_k(z^k) - \bar{z}^0||_{B_k} = ||P_k(z^k) - P_k(\bar{z}^0)||_{B_k} \le ||z^k - \bar{z}^0||_{B_k}$ by the nonexpansivity, we have

$$||z^{k+1} - \bar{z}^0||_{B_k} \le ||z^k - \bar{z}^0||_{B_k} + \varepsilon_k.$$

On the other hand $||z^{k+1} - \bar{z}^0||_{B_k} \ge \alpha_{k+1}^{-1}||z^{k+1} - \bar{z}^0||_{B_{k+1}}$ by (2.2), so that $\beta_k^{-1}a_{k+1}^{-1}||z^{k+1} - \bar{z}^0||_{B_{k+1}} \le \beta_k^{-1}[z^k - \bar{z}^0||_{B_k} + \varepsilon_k$. Since $\beta_{k+1} = \alpha_{k+1}\beta_k$ and $\beta_{k+1} \ge 1$, this yields

$$\beta_{k+1}^{-1}||z^{k+1} - \bar{z}^0||_{B_{k+1}} \le \beta_k^{-1}||z^k - \bar{z}^0||_{B_k} + \varepsilon_k, \tag{2.8}$$

which can be applied iteratively to get

$$\beta_k^{-1}||z^k - \bar{z}^0||_{B_k} \le \beta_0^{-1}||z^0 - \bar{z}^0||_{B_0} + \varepsilon_0 + \varepsilon_1 + \cdots \varepsilon_{k-1}. \tag{2.9}$$

[8]Plenty of criteria are available for this; see for instance [16, 12.51]. Classically it's known that emptiness is signaled by PPA iterates "converging to the horizon," but this isn't taken up here for the current algorithm.

Here $\beta_0 = \alpha_0$ by definition in (2.3) and $\alpha_0^{-1}||z^0 - \bar{z}^0||_{B_0} \leq ||z^0 - \bar{z}^0||$ by (2.2), with $||z^0 - \bar{z}^0|| = \operatorname{dist}(z^0, Z)$ inasmuch as $\bar{z}^0$ is the nearest point of Z to z^0. Hence (2.9) implies

$$||z^k - \bar{z}^0||_{B_k} \leq \beta_k[\operatorname{dist}(z^0, Z) + \sigma_k] \text{ for } \sigma_k = \sum\nolimits_{j=0}^{k-1} \varepsilon_j. \tag{2.10}$$

Also $||z^k - \bar{z}^0||_{B_k} \geq \beta_k^{-1}||z^k - \bar{z}^0||$ from (2.5), so we have

$$||z^k - \bar{z}^0|| \leq \beta_k^2[\operatorname{dist}(z^0, Z) + \sigma_k] \leq \beta_\infty^2[\operatorname{dist}(z^0, Z) + \sigma_\infty] \text{ for all } k, \tag{2.11}$$

where $\sigma_\infty = \sum_{k=0}^\infty \varepsilon_k$ is finite by (1.14).

This demonstrates that the z^k sequence is bounded and thus that its set of cluster points, which we denote now by Z^∞, is nonempty. To complete the proof of the theorem, we have to show that $Z^\infty \subset Z$ and that no more than a single $\bar{z}$ can belong to Z^∞. It will help in this that Z^∞ is also the set of all cluster points of the sequence $P_k(z^k)$ because $||z^{k+1} - P_k(z^k)|| \leq \beta_k||z^{k+1} - P_k(z^k)||_{B_k} \leq \beta_k\varepsilon_k \to 0$ under our stopping criterion, since $\beta_k \leq \beta_\infty < \infty$.

The derivation of (2.8) only utilized about $\bar{z}^0$ that, as an element of Z, it was a fixed point of P_k, so actually

$$\beta_{k+1}^{-1}||z^{k+1} - z||_{B_{k+1}} \leq \beta_k^{-1}||z^k - z||_{B_k} + \varepsilon_k \text{ for any } z \in Z. \tag{2.12}$$

This implies that the values $\gamma_k(z) = \beta_k^{-1}||z^k - z||_{B_k}$ approach a finite limit as $k \to \infty$, in view of our assumption (1.14) on ε_k and having $\gamma_{k+m}(z) \leq \gamma_k(z) + \sum_{j=k}^{m-1} \varepsilon_j$ for any $m > k$. Because at the same time $\beta_k \to \beta_\infty$, there follows the existence of a limit

$$\mu(z) = \lim_{k\to\infty} ||z^k - z||_{B_k}, \text{ which is also } \lim_{k\to\infty} ||P_k(z^k) - z||_{B_k}, \text{ for any } z \in Z. \tag{2.13}$$

From (2.1) in the case of $z' = z^k$, with $z \in Z$ implying through $P_k(z) = z$ that $Q_k(z) = 0$, we get

$$||Q_k(z^k)||_{B_k}^2 = ||z^k - z||^2 - ||P_k(z^k) - z||_{B_k}^2 \to \mu(z) - \mu(z) = 0.$$

In view of the bounds on the B_k norm in (2.5), this implies $Q_k(z^k) \to 0$ and also that $c_k^{-1}B_kQ_k(z^k) \to 0$, since $c_k^{-1} \leq 1$. The next thing to note is

$$(P_k(z^k), c_k^{-1}B_kQ_k(z^k)) \in \operatorname{gph} T \tag{2.14}$$

as a consequence of the definitions of P_k in (1.9) and Q_k in (2.1). By the monotonicity of T, we therefore have, for any z and $w \in T(z)$, that $\langle z - P_k(z^k), w - c_k^{-1}B_kQ_k(z^k)\rangle \geq 0$, and in the limit as $k \to \infty$ that

$$\langle z - \bar{z}, w\rangle \geq 0 \text{ for any cluster point } \bar{z} \text{ of } P_k(z^k), \text{ i.e., any } \bar{z} \in Z^\infty.$$

This being true for an arbitrary pair $(z, w) \in \operatorname{gph} T$, we must have $(\bar{z}, 0) \in \operatorname{gph} T$ because the monotonicity of T is maximal. Thus, $Z^\infty \subset Z$, hence from (2.13)

$$\lim_{k\to\infty} ||z^k - \bar{z}||_{B_k} = \mu(\bar{z}) \text{ for any } \bar{z} \in Z^\infty. \tag{2.15}$$

Having $\bar{z} \in Z^\infty$ means that some subsequence of $||z^k - \bar{z}||$ converges to 0, but $||z^k - \bar{z}||_{B_k} \leq \beta_\infty^2||z^k - \bar{z}||$ by (2.5). This implies that $\mu(\bar{z}) = 0$ in (2.15), and therefore, again through (2.4), that not just the particular subsequence of $||z^k - \bar{z}||$ in (2.15) converges to 0, but the entire sequence; we have $z^k \to \bar{z}$ and $Z^\infty = \{\bar{z}\}$. □

Moving on from simple convergence to linear convergence, we need to appeal to a condition on T at the solution point $\bar{z} \in Z$ reached in Theorem 2.1. The condition we employ is the metric subregularity of T at $\bar{z}$ with respect to $0 \in T(\bar{z})$, which is the case of (1.3) with $\bar{w} = 0$. It is equivalent to a property of T^{-1} called its *calmness* at 0 for $\bar{z} \in T^{-1}$, which we won't need to get into; see [2]. Metric subregularity of T in the given norm $||\cdot||$ carries over, through norm equivalence, to the same property of $B_k^{-1}T$ in the B_k-norm, but changing the norm can change the modulus of subregularity. Especially of interest here will be the expression of the metric subregularity in terms of the limit mapping B_∞ in (2.6). It comes out as the property that

$$\begin{array}{c}\text{for some neighborhood } \mathcal{N} \text{ of } (\bar{z},0),\ \exists\, a \in (0,\infty) \text{ such that} \\ (z,w) \in \mathcal{N},\ w \in T(z) \quad\Longrightarrow\quad \operatorname{dist}_{B_\infty}(z,Z) \le a||w||_{B_\infty^{-1}},\end{array} \tag{2.16}$$

with the associated B_∞-modulus of subregularity given by

$$\operatorname{subreg}_{B_\infty}(T;\bar{z}\,|\,\bar{w}) := \liminf \text{ of } a \text{ values in (2.16) as the neighborhood shrinks.} \tag{2.17}$$

The next result specializes [11, Theorem 4.2] to T being globally monotone while at the same time improving on the underlying assumptions and revealing the decisive influence of the subregularity modulus in (2.17). The proof of the earlier result depended on explicitly assuming the convergence property in (2.6) that we know here to be true on the basis of (2.2)–(2.3). Also, the earlier statement and proof involved an approach to subregularity that led to a Q-linear rate of convergence as in (2.18) below, but for a value a_∞ derived differently from the subregularity modulus (2.17) now incorporated in (2.18) below. The new proof succeeds furthermore in establishing that this subregularity modulus is actually the *lowest* of all the possible values derivable that earlier way.

Theorem 2.2 (linear convergence to the solution set). *If the stopping criterion for the proximal point iterations in Theorem 2.1 is tightened to (1.15b), or (1.16b), in getting $z^k \to \bar{z} \in Z$, and if the metric subregularity condition (1.3), expressed equivalently as (2.16) with modulus (2.17), holds at $\bar{z}$, then*

$$\begin{array}{c}\operatorname{dist}_{B_\infty}(z^k,Z) \text{ converges Q-linearly to 0 at the rate } r = a_\infty/\sqrt{a_\infty^2 + c_\infty^2}, \\ \text{where } a_\infty = \operatorname{subreg}_{B_\infty}(T;\bar{z}\,|\,0) < \infty.\end{array} \tag{2.18}$$

Proof. The argument for Theorem 2.1 showed not only that $z^k \to \bar{z} \in Z$, but also that the pairs (2.14) in $\operatorname{gph} T$ converge to $(\bar{z},0)$. Ultimately, then, for any $\delta > 0$, we will have both $||P_k(z^k) - \bar{z}|| < \delta$ and $||c_k^{-1}B_kQ_k(z^k)|| < \delta$, say when $k \ge \bar{k}(\delta)$. Let

$$a_k(\delta) = \inf\{\, a \ge 0 \,|\, \operatorname{dist}_{B_k}(z,Z) \le a||w||_{B_k^{-1}} \text{ if } (z,w) \in \operatorname{gph} T,\ ||z-\bar{z}|| \le \delta,\ ||w|| \le \delta\}, \tag{2.19}$$

so that, when $k \ge \bar{k}(\delta)$,

$$\operatorname{dist}_{B_k}(P_k(z^k),Z) \le a_k(\delta)||c_k^{-1}B_kQ_k(z^k)||_{B_k^{-1}} = a_k(\delta)c_k^{-1}||Q_k(z^k)||_{B_k}, \tag{2.20}$$

inasmuch as $||B_kz||_{B_k^{-1}} = ||z||_{B_k}$ in (1.8). The bounds in (2.7) say

$$\begin{array}{c}\frac{\beta_\infty}{\beta_k}\operatorname{dist}_{B_\infty}(z,Z) \ge \operatorname{dist}_{B_k}(z,Z) \ge \frac{\beta_k}{\beta_\infty}\operatorname{dist}_{B_\infty}(z,Z), \\ \frac{\beta_\infty}{\beta_k}||w||_{B_\infty^{-1}} \ge ||w||_{B_k^{-1}} \ge \frac{\beta_k}{\beta_\infty}||w||_{B_\infty^{-1}},\end{array} \tag{2.21}$$

and tell us that, for $w \ne 0$,

$$\frac{(\beta_\infty/\beta_k)\operatorname{dist}_{B_\infty}(z,Z)}{(\beta_k/\beta_\infty)||w||_{B_\infty^{-1}}} \ge \frac{\operatorname{dist}_{B_k}(z,Z)}{||w||_{B_k^{-1}}} \ge \frac{(\beta_k/\beta_\infty)\operatorname{dist}_{B_\infty}(z,Z)}{(\beta_\infty/\beta_k)||w||_{B_\infty^{-1}}}.$$

In applying this to (2.19), we get $(\beta_\infty/\beta_k)^2 a_\infty(\delta) \geq a_k(\delta) \geq (\beta_k/\beta_\infty)^2 a_\infty(\delta)$ so that

$$\lim_{k\to\infty} a_k(\delta) = a_\infty(\delta), \quad \text{whereas} \quad \lim_{\delta \searrow 0} a_\infty(\delta) = \operatorname{subreg}_{B_\infty}(T : \bar{z} \,|\, 0), \tag{2.22}$$

the latter by the definition of the modulus in (2.17).

Let $\bar{z}^k = \operatorname{proj}_{B_k}(z^k, Z)$, the nearest point of Z to z^k, so $||z^k - \bar{z}^k||_{B_k} = \operatorname{dist}(z^k, Z)$. The case of the firm nonexpansivity relation (2.1) with $z' = z^k$ and $z = \bar{z}^k$, therefore $P_k(\bar{z}^k) = \bar{z}^k$ and $Q_k(z') = 0$, yields

$$||P_k(z^k) - \bar{z}^k||_{B_k}^2 + ||Q_k(z^k)||_{B_k}^2 \leq ||z^k - \bar{z}^k||^2,$$

where $||P_k(z^k) - \bar{z}^k||_{B_k}^2 \geq \operatorname{dist}_{B_k}^2(P_k(z^k), Z)$, hence

$$||Q_k(z^k)||_{B_k}^2 \leq \operatorname{dist}_{B_k}^2(z^k, Z) - \operatorname{dist}_{B_k}^2(P_k(z^k), Z). \tag{2.23}$$

Juxtaposing this inequality with (2.20) leads to

$$\operatorname{dist}_{B_k}^2(P_k(z^k), Z) \leq a_k(\delta)^2 c_k^{-2} \left(\operatorname{dist}_{B_k}^2(z^k, Z) - \operatorname{dist}_{B_k}^2(P_k(z^k), Z) \right),$$

which comes out as

$$\operatorname{dist}_{B_k}(P_k(z^k), Z) \leq r_k(\delta) \operatorname{dist}_{B_k}(z^k, Z) \;\text{ for }\; r_k(\delta) = a_k(\delta)/\sqrt{a_k(\delta)^2 + c_k^2} < 1. \tag{2.24}$$

With the goal of (2.18) in mind, we need to build on this by relating $\operatorname{dist}_{B_k}(P_k(z^k), Z)$ to $\operatorname{dist}_{B_k}(z^{k+1}, Z)$ through the stopping criterion (1.15b) (satisfied in particular under (1.16b)). Let $p^k = \operatorname{proj}_{B_k}(P_k(z^k), Z)$ get $||z^{k+1} - p^k||_{B_k}$ as an upper bound on $\operatorname{dist}_{B_k}(z^{k+1}, Z)$ to work from. We have

$$\begin{aligned} ||z^{k+1} - p^k||_{B_k} &\leq ||z^{k+1} - P_k(z^k)||_{B_k} + ||P_k(z^k) - p^k||_{B_k} \\ &= ||z^{k+1} - P_k(z^k)||_{B_k} + \operatorname{dist}_{B_k}(P_k(z^k), Z) \\ &\leq ||z^{k+1} - P_k(z^k)||_{B_k} + r_k(\delta) \operatorname{dist}_{B_k}(z^k, Z) \end{aligned} \tag{2.25}$$

by (2.24), where the stopping criterion gives us

$$\begin{aligned} ||z^{k+1} - P_k(z^k)||_{B_k} &\leq \varepsilon_k ||z^{k+1} - z^k|| \leq \varepsilon_k \Big(||z^{k+1} - p^k||_{B_k} + ||p^k - \bar{z}^k||_{B_k} + ||\bar{z}^k - z^k||_{B_k} \Big) \\ &\leq \varepsilon_k ||z^{k+1} - p^k||_{B_k} + \varepsilon_k ||p^k - \bar{z}^k||_{B_k} + \varepsilon_k \operatorname{dist}_{B_k}(z^k, Z). \end{aligned}$$

Together, this estimate and (2.24) give us

$$(1 - \varepsilon_k) ||z^{k+1} - p^k||_{B_k} \leq \varepsilon_k ||p^k - \bar{z}^k||_{B_k} + [\varepsilon_k + r_k(\delta)] \operatorname{dist}_{B_k}(z^k, Z). \tag{2.26}$$

The B_k-nonexpansivity of the B_k-projection mapping onto Z, along with (2.23), provide

$$\begin{aligned} ||\bar{z}^k - p^k||_{B_k} &= ||\operatorname{proj}_{B_k}(z^k, Z) - \operatorname{proj}_{B_k}(P_k(z^k), Z)||_{B_k} \\ &\leq ||z^k - P_k(z^k)||_{B_k} = ||Q_k(z^k)||_{B_k} \leq \operatorname{dist}_{B_k}(z^k, Z). \end{aligned}$$

Putting this into (2.26) and recalling that $\operatorname{dist}_{B_k}(z^{k+1}, Z)_{B_k} \leq ||z^{k+1} - p^k||_{B_k}$, since $p^k \in Z$, we get

$$\operatorname{dist}_{B_k}(z^{k+1}, Z) \leq \left[\frac{r_k(\delta) + 2\varepsilon_k}{1 - \varepsilon_k} \right] \operatorname{dist}_{B_k}(z^k, Z).$$

Invoking the distance bounds in (2.21), we arrive at the convergence estimate

$$\operatorname{dist}_{B_\infty}(z^{k+1}, Z) \leq \bar{r}_k(\delta) \operatorname{dist}_{B_\infty}(z^k, Z) \;\text{ for }\; \bar{r}_k(\delta) = \frac{b_\infty^2}{b_k^2} \left[\frac{r_k(\delta) + 2\varepsilon_k}{1 - \varepsilon_k} \right]$$

As $k \to \infty$, not only $\varepsilon_k \to 0$ and $\beta_k \to \beta_\infty$, but also $r_k(\delta) \to r_\infty(\delta)$ in (2.24) because $a_k(\delta) \to a_\infty(\delta)$ in (2.22). Therefore, $\bar{r}_k(\delta) \to r_\infty(\delta)$ as well, so our convergence estimate indicates that $\mathrm{dist}_{B_\infty}(z^k, Z)$ converges Q-linearly to 0 at the rate $r_\infty(\delta)$. Here δ can be arbitrarily small, and as $\delta \to 0$, $r_\infty(\delta)$ decreases to the rate r in (2.18) by (2.22). This completes the proof. □

Whether or not the metric subregularity condition in Theorem 2.2 holds at $\bar{z}$, as presumed in the hypothesis, is ordinarily hard to know in advance of actually having determined $\bar{z}$. It will be seen below that it does at least hold "usually," from an interesting standpoint. However, here is an example of when it is sure to hold.

Example 2.2 (guaranteed linear convergence under piecewise polyhedrality)**.** *If the mapping T is piecewise polyhedral, which means that its graph is the union of a finite collection of polyhedral convex sets, then the metric subregularity condition at $\bar{z}$ in Theorem 2.2 is guaranteed to be satisfied, and with it the linear convergence behavior that is described there.*

Detail. Piecewise polyhedral mappings are know to be *calm* everwhere [16, 9.57]. When T is piecewise polyhedral, its inverse T^{-1} is piecewise polyhedral as well and therefore calm everywhere. But calmness is the inverse property paired with metric subregularity [2, 3H.3], so this implies that T is metrically subregular everywhere. □

In the case where $T = \partial f$, the maximal monotone mapping T is piecewise polyhedral if and only if the convex function f is piecewise linear-quadratic [16, 10E, 12.30]. More generally in convex optimization, T is piecewise polyhedral when it comes from a Lagrangian in piecewise linear-quadratic programming [16, pp. 506, 513], with traditional quadratic programming as a particular example.

The message of Theorem 2.2, as a follow-up to Theorem 2.1, is that linear convergence of the variable-metric PPA can be achieved by taking extra care in approximations through a tighter stopping criterion. But this linear convergence is that of the distance of z^k from the solution set Z, instead of the distance of z^k from its limit $\bar{z}$. As explained in Section 1, there is an extra condition on top of metric subregularity which, with the stopping criteria at level (b) upgraded to a level "(c)" in which the distance between z^{k+1} and z^k is replaced by its square, Q-linear convergence of z^k to $\bar{z}$ is assured for the *fixed-metric* PPA — as shown in [11]. It appears to be very difficult, if not impossible though, to carry that result forward into a variable-metric implementation, because of the additional stress on controlling the approximations.

Nevertheless, something important can be said about Q-linear convergence of z^k to $\bar{z}$ by looking at the issue from a wider perspective. That's the *parametric* view, in which problem (1.1) is seen as imbedded in the family of problems

$$\text{given } \bar{w}, \text{ find } \bar{z} \in Z_{\bar{w}} = T_{\bar{w}}^{-1}(0), \text{ where } T_{\bar{w}}(z) = T(z) - \bar{w}. \tag{2.27}$$

This replaces the generalized equation $T(z) \ni 0$ by $T(z) \ni \bar{w}$ with a potentially varying right side $\bar{w}$, but in the mode of a parameterized family of maximal monotone mappings $T_{\bar{w}}$ in which the problem we have been looking at is a single instance. That's a context in which we can ask about "typical" performance on the algorithm with respect to choices of $\bar{w}$.

Properties associated with maximal monotonicity that haven't yet been put to use will clarify how to pose this question in a most helpful manner. Certainly we only want to focus on the choices of $\bar{w}$ for which the solution set in (2.27) isn't just empty, and they comprise the *effective domain* of T^{-1}, denoted by $\mathrm{dom}\, T^{-1}$. Much is known about it. For instance, it is a *nearly convex* set [16, 12.41], which means that it differs from a closed convex set only by perhaps lacking some of the boundary points of that set relative to its affine hull. From that, $\mathrm{dom}\, T^{-1}$ is full dimensional if and only if its convex relative interior is a true interior. Having $\bar{w}$ belong to $\mathrm{int}(\mathrm{dom}\, T^{-1})$ corresponds however to having $T^{-1}(\bar{w})$ not just be nonempty, but also bounded [16, 12.38]. In contrast, if $\operatorname{dom} T^{-1}$ does have empty interior, the nonempty solution sets in (2.27) are all "degenerate" convex sets in the sense of having a nonzero subspace in their recession cones [16, 12.37]. They must that way take the form of *bundles of parallel lines, planes or even hyperplanes.*

It would be appropriate of course, for most applications, to suppose in solving problem (1.1) numerically, that the solution set Z is nonempty and not "degenerate" in the manner just described. From the background explained about domains, that means supposing $0 \in \operatorname{dom} T^{-1}$ with $\operatorname{int}(\operatorname{dom} T^{-1}) \neq \emptyset$. Making sure the solution set $Z = T^{-1}(0)$ in (1.1) is also bounded, which likewise is a generally good idea for computations, corresponds to the slightly stronger assumption that $0 \in \operatorname{int}(\operatorname{dom} T^{-1})$. But the alternative problems in (2.27) for $\bar{w} \neq 0$ are subject to the same considerations. The ones for $\bar{w} \in \operatorname{int}(\operatorname{dom} T^{-1})$ are thus of main interest and furnish the natural platform for asking whether some behavior is "typical." Here is an answer to that well posed question.

Theorem 2.3 (global linear convergence to a solution point achieved generically)**.** *For almost every $\bar{w}$ in the open convex set consisting of $\bar{w}$ for which the solution set $Z_{\bar{w}}$ in (2.27) is nonempty and bounded, $Z_{\bar{w}}$ is a singleton $\{\bar{z}\}$ and $T_{\bar{w}}$ exhibits metric subregularity at $\bar{z}$ for $0 \in T_{\bar{w}}(\bar{z})$. Therefore, in applying the variable metric PPA to solve the $\bar{w}$-problem in (2.27) in the manner of Theorem 2.2, the sequence z^k will converge Q-linearly to $\bar{z}$ at the indicated rate in (2.18), with the modulus for T in (2.17) being replaced by the corresponding one for $T_{\bar{w}}$, namely*

$$\operatorname{subreg}_{B_\infty}(T_{\bar{w}}; \bar{z} \,|\, 0) = \operatorname{subreg}_{B_\infty}(T; \bar{z} \,|\, \bar{w}), \tag{2.28}$$

Proof. Here "almost everywhere" refers to the set of exceptions being negligible with respect to Lebesgue measure. When there is a unique solution $\bar{z}$, linear convergence to the solution set is obviously linear convergence to that $\bar{z}$. From the background about $\operatorname{dom} T^{-1}$ ahead of this theorem's statement, it's clear then that confirming the validity of the claim comes down to confirming that strong metric subregularity of T holds for almost every $\bar{w} \in \operatorname{int}(\operatorname{dom} T^{-1})$. Exactly this has recently been established in [14, Theorem 1], and that lays the matter to rest. □

The generic property in Theorem 2.3 can be given a probabilistic interpretation. The "right side" parameter $\bar{w}$ in (2.27) may be comprised of variables subject to measurement error or other uncertainties that can be modeled by a positive probability density function. Then "almost every" translates to "with probability one" — there is zero probability of being presented with an instance of the generalized equation for which strong metric subregularity is lacking. But this interpretation also puts light on a shortcoming of the genericity. Some aspects of $\bar{w}$ may be fixed by the problem's technical formulation and not subject to any uncertainties.

Another thing to note is that the generic guarantee of linear convergence in Theorem 2.3, however interpreted, offers no insights into a "typical" *rate* of linear convergence being achieved. That rate depends by Theorem 2.2 on the modulus in (2.28), which might vary unpredictably even under small changes in $\bar{w}$. The only good exception seems to be for T being piecewise polyhedral as in Example 2.2. In that case there is a uniform upper bound in modulus values for the "calmness" of T^{-1} that's behind it [16, 9.57], and those values coincide with the values in (2.28) by [16, 3H.3].

3 Localized executability of the algorithm

Up to now, the mapping T has been maximal monotone in the usual sense — global. But monotonicity can also be considered relative to a subset $\mathcal{Z} \times \mathcal{W} \subset \mathcal{H} \times \mathcal{H}$:

$$\langle w_1 - w_0, z_1 - z_0 \rangle \geq 0 \ \text{ for all pairs } (z_0, w_0),\ (z_1, w_1) \text{ in } [\mathcal{Z} \times \mathcal{W}] \cap \operatorname{gph} T, \tag{3.1}$$

again with maximality signifying that no mapping T' with a properly larger graph within $\mathcal{Z}\times\mathcal{W}$ also has this property. For T maximal monotone this way relative to $\mathcal{Z}\times\mathcal{W}$, there always exists $\bar{T}:\mathcal{H}\rightrightarrows\mathcal{H}$ maximal monotone globally such that $[\mathcal{Z}\times\mathcal{W}]\cap\mathrm{gph}\,\bar{T}=[\mathcal{Z}\times\mathcal{W}]\cap\mathrm{gph}\,T$; see [16, 12.6]. Local maximal monotonicity around a pair $(\bar{z},\bar{w})\in\mathrm{gph}\,T$, in the sense of $\mathcal{Z}\times\mathcal{W}$ being a neighborhood of $(\bar{z},\bar{w})$ is important in nonconvex optimization. In the case of $T=\partial f$ for a lower semicontinuous, proper function on $\mathcal{H}$, not necessarily convex, corresponds to f being *variationally convex* at $\bar{z}$ for $\bar{w}$ [8], with its subgradients and associated function values in the primal-dual localization being indistinguishable from those of a convex function. This is a key in the door to applying the proximal point algorithm to problems of nonconvex optimization by way of PPA-based approaches to problem decomposition, such as the progressive decoupling algorithm in [10].

Here in the maximal monotonicity of T relative to $\mathcal{Z}\times\mathcal{W}$ we take

$$\mathcal{Z}\text{ and }\mathcal{W}\text{ open convex with }\mathcal{Z}\cap Z\neq\emptyset\text{ and }0\in\mathcal{W}\tag{3.2}$$

and address the question of whether the algorithm, as already articulated for T being maximal monotone globally, will keep within $\mathcal{Z}\times\mathcal{W}$, when initiated at a point z^0 that is near enough to Z in $\mathcal{Z}$, moreover without drawing on anything in $\mathrm{gph}\,T$ outside of $\mathcal{Z}\times\mathcal{W}$. A positive answer will be put together which includes a prescription for how near is enough. That will establish that the variable metric PPA can operate as a local procedure when T is known only to be maximal monotone relative to $\mathcal{Z}\times\mathcal{W}$, because T could replaced by a maximal monotone extension $\bar{T}$ as indicated above, and the algorithm wouldn't see the difference.

For getting a prescription for how near to Z a potential choice of z^0 with $(z^0,0)\in\mathcal{Z}\times\mathcal{W}$ ought to be, we introduce σ such that

$$\infty>\sigma\geq\sum\nolimits_{k=0}^{\infty}\varepsilon_k\tag{3.3}$$

and take

$$\rho>\beta_\infty^2[\mathrm{dist}(z^0,Z)+\sigma]\ \text{ such that }\ (z,w)\in\mathcal{Z}\times\mathcal{W}\ \text{ if }\ ||z-z^0||<2\rho,\ ||w||<2\rho.\tag{3.4}$$

Note that this puts a requirement on the size of σ as well as on the size of $\mathrm{dist}(z^0,Z)$. It indicates that the approximations allowed by the stopping criteria in (1.15) and (1.16) may need to be controlled with less error, depending on the narrowness of the scope of localization.

Since our intention is to make the algorithm applicable even when T is just maximal monotone locally, we need to guard against the fact that, in this setting, the presciption for updating to z^{k+1} from z^k might be ambiguous due the presence of extraneous elements from possibly nonmonotone parts of $\mathrm{gph}\,T$ outside of $\mathcal{Z}\times\mathcal{W}$ if monotonicity is only know locally. This will handled by asking also that the iterations leave out of consideration potential choices of z^{k+1} that fail to satisfy

$$||z^{k+1}-z^k||_{B_k}<\rho.\tag{3.5}$$

Theorem 3.1 (guarantee of local executability)**.** *Under (3.2), let the algorithm be initiated at a point z^0 satisfying (3.4), with updates from z^k to z^{k+1} subjected to (3.5). That update condition will not interfere with the feasibility of iterations, and the z^k sequence will then remain in $\mathcal{Z}$, as will the $P_k(z^k)$ sequence. Moreover, the procedure will exhibit the convergence behavior in Theorems 2.1 and 2.2 without involving in its execution anything about $\mathrm{gph}\,T$ outside of $\mathcal{Z}\times\mathcal{W}$.*

Proof. The graph of T enters the execution and the justification of convergence in Theorem 2.1 in two ways. First, the definition of P_k and the firm nonexpansiveness in (2.1) arise from having

$$(P_k(z^k),Q_k(z^k))\in\mathrm{gph}[c_kB_k^{-1}T],\ \text{ i.e., }\ (P_k(z^k),c_k^{-1}B_kQ_k(z^k))\in\mathrm{gph}\,T.\tag{3.6}$$

Second, the stopping criteria (1.16) bring in vectors $w \in S_k(z^{k+1})$ with $||c_k w||_{B_k^{-1}} \leq \varepsilon_k$, and in terms of the definition of S_k this comes down to

$$(z^{k+1}, w - c_k^{-1} B_k[z^{k+1} - z^k]) \in \operatorname{gph} T \ \text{ with } \ ||w||_{B_k^{-1}} \leq \varepsilon_k c_k^{-1}. \tag{3.7}$$

As far at Theorem 2.1 is concerned, therefore, and in view of the balls of radius 2ρ specified to lie in $\mathcal{Z}$ and $\mathcal{W}$ in (3.4), our task is verifying that

$$\begin{gathered} ||z^k - z^0|| < 2\rho, \quad ||P_k(z^k) - z^0|| < 2\rho, \quad ||c_k^{-1} B_k Q_k(z^k))|| < 2\rho, \\ \text{and if } \ ||w||_{B_k^{-1}} \leq \varepsilon_k c_k^{-1}, \ \text{ also } \ ||w - c_k^{-1} B_k[z^{k+1} - z^k]|| < 2\rho. \end{gathered} \tag{3.8}$$

From the firm nonexpansiveness in (2.1) as applied to $z' = z^k$ and $z = \bar{z}^0 \in Z$, for which $P_k(\bar{z}^0) = \bar{z}^0$ and $Q_k(\bar{z}^0) = 0$, we see that $||z^k - \bar{z}^0||_{B_k}$ is an upper bound to both $||P_k(z^k) - \bar{z}_0||_{B_k}$ and $||Q_k(z^k)||_{B_k}$. Therefore, by (2.10),

$$\left.\begin{array}{c} ||z^k - \bar{z}^0||_{B_k} \\ ||P_k(z^k) - \bar{z}_0||_{B_k} \\ ||Q_k(z^k)||_{B_k} \end{array}\right\} \leq \beta_k[\operatorname{dist}(z^0, Z) + \sigma_k] \tag{3.9}$$

and consequently, through the stopping criterion because $z^{k+1} - z^k = z^{k+1} - P_k(z^k) - Q_k(z^k)$,

$$\begin{aligned} ||z^{k+1} - z^k||_{B_k} &\leq ||z^{k+1} - P_k(z^k)||_{B_k} + ||Q_k(z^k)||_{B_k} \\ &\leq \varepsilon_k + \beta_k[\operatorname{dist}(z^0, Z) + \sigma_k] \leq \beta_k[\operatorname{dist}(z^0, Z) + \sigma_{k+1}] \\ &< \beta_\infty[\operatorname{dist}(z^0, Z) + \sigma] \leq \beta_\infty^2[\operatorname{dist}(z^0, Z) + \sigma] < \rho, \end{aligned} \tag{3.10}$$

inasmuch as $\beta_k \geq 1$ and $\beta_\infty \geq 1$. Thus, the extra condition imposed in (3.5) won't exclude any of the potential update choices associated with the algorithm in its global articulation, while allowing only aspects of T in its localization to $\mathcal{Z} \times \mathcal{W}$ to come into play.

Through (3.9), the bound in (2.5), and the fact that $||z - z^0|| \leq ||z - \bar{z}^0|| + ||\bar{z}^0 - z^0||$, where in particular we can take $z = z^k$ or $z = P_k(z^k)$, we get

$$\begin{aligned} \left.\begin{array}{c} ||z^k - z^0|| \\ ||P_k(z^k) - z^0|| \end{array}\right\} &\leq \beta_k(\beta_k[\operatorname{dist}(z^0, Z) + \sigma_k]) + \operatorname{dist}(z^0, Z) \\ &\leq \beta_\infty^2[\operatorname{dist}(z^0, Z) + \sigma] + \operatorname{dist}(z^0, Z) < 2\rho. \end{aligned} \tag{3.11}$$

Thus, through (3.4), z^k and $P_k(z^k)$ do obey (3.8). Noting now from (1.8) that

$$||B_k z||_{B_k^{-1}} = ||z||_{B_k}, \ \text{ because } \ ||B_k z||_{B_k^{-1}}^2 = \langle B_k z, B_k^{-1} B_k z \rangle, \tag{3.12}$$

and recalling the bound on the B_k^{-1}-norm in (2.5), we observe next from (3.9) that

$$\begin{aligned} &||c_k^{-1} B_k Q_k(z^k)|| \leq \beta_k ||c_k^{-1} B_k Q_k(z^k)||_{B_k^{-1}} = \beta_k c_k^{-1} ||Q_k(z^k)||_{B_k} \\ &\leq \beta_k(\beta_k[\operatorname{dist}(z^0, Z) + \sigma_k]) < \beta_\infty^2[\operatorname{dist}(z^0, Z) + \sigma] < \rho. \end{aligned} \tag{3.13}$$

This confirms the third bound in (3.8) via (3.4). For the last part of (3.8) we estimate

$$||w - c_k^{-1} B_k(z^{k+1} - z^k)||_{B_k^{-1}} \leq ||c_k^{-1} B_k(z^{k+1} - z^k)||_{B_k^{-1}} + ||w||_{B_k^{-1}}$$

where by (3.12) and (3.9)

$$||c_k^{-1} B_k(z^{k+1} - z^k)||_{B_k^{-1}} = c_k^{-1} ||z^{k+1} - z^k||_{B_k} \leq c_k^{-1}(\beta_k[\operatorname{dist}(z^0, Z) + \sigma_k])$$

and, by assumption, $||w||_{B_k^{-1}} \leq \varepsilon_k c_k^{-1}$. Utilizing again the bound on the B_k^{-1}-norm in (2.5), and recalling from (1.6) that $c_k^{-1} \leq 1$, we then get

$$\begin{aligned}||w - c_k^{-1}B_k(z^{k+1} - z^k)|| &\leq \beta_k(\beta_k[\mathrm{dist}(z^0, Z) + \sigma_{k+1}] + \varepsilon_k)\\ &\leq \beta_\infty^2[\mathrm{dist}(z^0, Z) + \sigma] + \beta_\infty\sigma \leq 2\beta_\infty^2[\mathrm{dist}(z^0, Z) + \sigma] < 2\rho.\end{aligned}$$

That finishes the validation of (3.10) and the claims concerning Theorem 2.1.

The only thing more that needs attention in connection with Theorem 2.2 is the employment in its proof of the $\bar{z}^k$ and p^k as the nearest points of z^k and $P_k(z^k)$ to Z in the B_k norm. They need to lie in $\mathcal{Z}$ as well — but only for sufficiently high k, since only such k are of importance for the linear convergence argument. We already know that z^k and $P_k(z^k)$ both converge to $\bar{z} \in Z$. Nearest points to them in Z therefore converge to $\bar{z}$ as well, so that eventually, through the convexity of Z in $\mathcal{Z}$, the behavior is entirely captured withing $\mathcal{Z}$, as required. □

Data availability: No data was generated or analyzed.
Competing interests: The author, a co-editor, reliquished any role in this paper's review.

References

[1] Burke, J.V., Qian, M., "A variable metric proximal point algorithm for monotone operators," *SIAM J. Control Optim.* **37** (1999), 353–375.

[2] Dontchev, A.D., Rockafellar, R.T., *Implicit Functions and Solution Mappings,* Springer Verlag, second edition: 2014.

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

[4] Parente, L.A., Lotito, P.A., Solodov, M.V., "A class of inexact variable metric proximal point algorithms," *SIAM J. Optimization* 19 (2008), 240–260.

[5] Pennanen, T., "Local convergence of the proximal point algorithm and multiplier methods without monotonicity." *Mathematics of Operations Research* **27** (2002), 170–191.

[6] Rockafellar, R. T., "Monotone operators and the proximal point algorithm." *SIAM J. Control Opt.* **14** (1976), 877–898.

[7] Rockafellar, R. T., "Augmented Lagrangians and applications of the proximal point algorithm in convex programming." *Math. of Operations Research* 1 (1976), 97–116.

[8] Rockafellar, R. T., "Variational convexity and the local monotonicity of subgradient mappings," *Vietnam J. Mathematics* (issue for A.D. Ioffe's 80th birthday) **47** (2019), 547–561.

[9] Rockafellar, R. T., "Progressive decoupling of linkages in monotone variational inequalities and convex optimization," *Proceedings of the 10th International Conference on Nonlinear Analysis and Convex Analysis* (Chitose, Japan, 2017), M. Hojo, M. Hoshino, W. Takahashi (eds.), Yokohama Publishers, 2019, 271–291.

[10] Rockafellar, R. T., "Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity," *Set-Valued and Variational Analysis* **27** (2019), 863–893.

[11] Rockafellar, R. T., "Advances in convergence and scope of the proximal point algorithm," *J. Nonlinear and Convex Analysis* **22** (2021), 2347–2375.

[12] Rockafellar, R. T., "Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality," *Mathematical Programming* 198 (2023), 159–194, published online June 2022.

[13] Rockafellar R. T., "Convergence of augmented Lagrangian methods in extensions beyond nonlinear programming," *Mathematical Programming*, published online June 2022, DOI 10.1007/s10107-022-01832-5

[14] Rockafellar R. T., "Metric regularity properties of monotone mappings," *Mathematica Serdica* (memorial issue for Asen Dontchev), accepted Febuary 2023.

[15] Rockafellar R. T., "Generaliations of the proximal method of multipliers in convex optimization," *Computational Optimization and Applications*, submitted February 2023.

[16] Rockafellar, R. T., and Wets, R. J-B, *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 1998.

[17] Spingarn, J., "Partial inverse of a monotone operator," *Applied Mathematics and Optimization* **10** (1983), 247–265.