Journal of Nonlinear and Convex Analysis 22 (2021)

CHARACTERIZING FIRM NONEXPANSIVENESS OF PROX MAPPINGS BOTH LOCALLY AND GLOBALLY

R. Tyrrell Rockafellar[1]

Abstract

The prox mappings for convex functions that were introduced by Moreau have achieved wide importance as a tool in optimization theory and numerical methodology, but what can be said about prox mappings for nonconvex functions? Can the key property of being firmly nonexpansive carry over to some degree in the absence of convexity, if not globally at least locally? Such questions are answered in a framework of finite-dimensional variational analysis in which proximal subgradients can be utilized. On the global level convexity turns out to be essential, while on the local level variational convexity is sufficient but not necessary unless the function is prox-regular.

Keywords: *prox mappings, firm nonexpansiveness, maximal monotonicity, variational analysis, nonconvex functions, variational convexity, proximal subgradients, prox-regularity*

Version of 28 January 2020

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

1 Introduction

Let $f : I\!R^n \to (-\infty, \infty]$ be a lower semicontinous (lsc) function, $f \not\equiv \infty$, and let prox_f be the associated *prox mapping*, defined by

$$\text{prox}_f(z) = \text{argmin}_x \{\, f(x) + \tfrac{1}{2}|x - z|^2\}, \tag{1.1}$$

where $|\cdot|$ is the Euclidean norm. When f is convex, prox_f is single-valued and has the important property of being "firmly nonexpansive," about which more will be explained shortly. When f is not convex, prox_f generally has to be treated as a set-valued mapping. However, there could be examples of prox_f being single-valued despite f not being convex. Might prox_f then even be firmly nonexpansive as well? Such questions can be raised also for *localized* prox mappings as defined by

$$\text{prox}_f^B(z) = \underset{x \in B}{\text{argmin}} \{\, f(x) + \tfrac{1}{2}|x - z|^2\} \text{ for a subset } B \subset I\!R^n \text{ and } z \in B \tag{1.2}$$

Prox mappings in the global sense of (1.1) in the case of convex functions f were introduced in 1962 by Moreau [10, 11, 12]. They have since become a workhorse in convex analysis and a key ingredient of various numerical methods for solving problems of optimization, starting with the proximal point algorithm in [16, 17] and continuing into decomposition algorithms like those in [22, 20, 5, 6] and recently [19].

Characterizing the special properties of prox mappings is important for these reasons and motivates our undertaking here. The proximal point algorithm offers a particularly clear illustration. In simplest form, it aims to solve the global problem of minimizing f over $I\!R^n$ by solving a sequence of better-behaved subproblems. It generates a sequence of points x^ν for $\nu = 1, 2, \ldots$ by taking

$$x^{\nu+1} \in \text{argmin}_x \{\, f(x) + \tfrac{1}{2}|x - x^\nu|^2\} = \text{prox}_f(x^\nu). \tag{1.3}$$

This can be viewed as a fixed-point type of iteration through the fact that

$$\bar{x} \in \text{argmin}_x\, f(x) \quad \Longrightarrow \quad \bar{x} \in \text{prox}_f(\bar{x}). \tag{1.4}$$

When f is convex and prox_f is accordingly single-valued and "firmly nonexpansive," the implication in (1.4) becomes equivalence, and the sequence of points x^ν is sure always to converge to a particular global minimizer of f [16, 17]. But there are echoes also in employing the same scheme *locally* in pursuit of *local* minimizers of f when it might not be convex. Then the iterations take the form

$$x^{\nu+1} \in \underset{x \in B}{\text{argmin}} \{\, f(x) + \tfrac{1}{2}|x - x^\nu|^2\} = \text{prox}_f^B(x^\nu) \tag{1.5}$$

in a neighborhood B of a local minimizer $\bar{x}$. Building on the work of Pennanen [13], we showed in [18], with improvements in [19], that convergence to a local minimizer is assured then as long as f is "variationally convex" there, which is a weaker requirement than local convexity. This feature was promoted in [19] as an engine for decomposition methods capable of solving structured problems of nonconvex optimization by way of augmented Lagrangian-type subproblems.

In such applications the iterations in (1.5) can executed with the $\frac{1}{2}$ replaced by $\frac{1}{2\lambda}$ for some $\lambda > 0$, and that suggests looking at this modification in (1.1) and (1.2) as well. No really new kinds of mappings are obtained that way, because

$$\text{argmin}_x \Big\{\, f(x) + \frac{1}{2\lambda}|x - z|^2\Big\} = \text{prox}_{\lambda f}(z), \tag{1.6}$$

and likewise in (1.2). However, these λ-parameterized mappings will anyway turn out to be useful in technical analysis. With this in mind, we make the restriction that

$$\text{henceforth the function } f \text{ is prox-bounded,} \tag{1.7}$$

which corresponds among other things to the existence of a quadratic function q on $I\!R^n$ such that $q \leq f$, cf. [21, p. 21]. Otherwise all the mappings in (1.6) would have empty graph, because the infimum would be $-\infty$..

In this paper we won't be occupied directly with algorithmic procedures but rather with pinning down facts about prox mappings that enable the success of such procedures. Results will come on both the global and local levels. In short, we will establish for the first time that prox_f is firmly nonexpansive *only* when f is convex. For the local version of this with respect to mappings prox_f^B, we will show however that variational convexity of f is sufficient without being necessary. But particular circumstances will be identified in which neccessity does hold nonetheless.

The global result for prox_f has in its background a famous issue in approximation theory. A subset C of a metric space is called a Chebyshev set if the nearest-point projection mapping P_C is everywhere single-valued. When the metric space is $I\!R^n$ with the usual Euclidean norm, the Chebyshev sets are precisely the nonempty closed convex sets, although for infinite-dimensional Hilbert spaces the necessity of convexity remains a nagging open problem; see [7]. The connection with our topic in this paper is that

$$f = \delta_C \;\; \text{[indicator function]} \quad \Longrightarrow \quad \operatorname{prox}_f = P_C. \tag{1.8}$$

This might raise the idea of extending the approximation context from sets to functions by calling f a Chebyshev function if prox_f is a single-valued mapping. Such a function need not be convex, though, as seen from elementary examples like $f(x) = -\frac{1}{4}||x||^2$. It is the additional property of prox_f being "firmly nonexpansive" that turns out to make the convexity of f essential.

Much of what we uncover here might be conjectured to hold in any Hilbert space, not just $I\!R^n$, but the technical challenges involved in such an extension are daunting. The open problem about Chebyshev sets is one indication, but another is a lack of a supporting theory of "proximal subgradients," which are inherently tied to prox mappings, as will soon be our theme. For convex functions, subgradients are naturally always proximal, and in finite dimensions the subgradients of nonconvex functions can be anyway still be characterized as limits of proximal subgradients. In infinite dimensions such a characterization is unavailable, and the development of subgradients takes a different track [9].

2 Basic properties and their relationships

Several properties associated with mappings $T : I\!R^n \rightrightarrows I\!R^n$ will play a big role in what follows. The notation $\rightrightarrows$ indicates that T may be set-valued in general, with single-valuedness as a special case; $\operatorname{gph} T$ designates the graph of T within $I\!R^n \times I\!R^n$, and the domain and range of T are the sets $\operatorname{dom} T$ and $\operatorname{rge} T$ obtained by projecting $\operatorname{gph} T$ in its separate arguments.

The first property to recall is the *monotonicity* of T relative to a set $W \subset I\!R^n \times I\!R^n$, namely

$$\langle x_1 - x_2, y_1 - y_2 \rangle \geq 0 \;\text{ for all }\; (x_1, y_1), (x_2, y_2) \in W \cap \operatorname{gph} T. \tag{2.1}$$

The monotonicity is *maximal* relative to W if there is no mapping T' with $W \cap \operatorname{gph} T' \supset W \cap \operatorname{gph} T$, but $W \cap \operatorname{gph} T' \neq W \cap \operatorname{gph} T$, such that T' is monotone relative to W. These are simply referred to as monotonicity and maximal monotonicity when W is all of $I\!R^n \times I\!R^n$. Similarly, *strong* monotonicity and maximal monotonicity with modulus $\mu > 0$ correspond to replacing $\langle x_1 - x_2, y_1 - y_2 \rangle \geq 0$ in (2.1) by $\langle x_1 - x_2, y_1 - y_2\rangle \geq \mu|x_1 - x_2|^2$. For history and more details on these concepts, which date from the 1960s, see [21, Chapter 12].

Another property is the *nonexpansivity* of a mapping $T : I\!\!R^n \rightrightarrows I\!\!R^n$ relative to a set $B \subset I\!\!R^n$, which means

$$T \text{ is single-valued on } B \text{ with } \ |T(z_1) - T(z_2)| \leq |z_1 - z_2|. \tag{2.2}$$

It is *firmly nonexpansive* relative to B if it satisfies the stronger condition there that

$$|T(z_1) - T(z_2)|^2 + |(I-T)(z_1) - (I-T)(z_2)|^2 \leq |z_1 - z_2|^2. \tag{2.3}$$

Again, the "relative" falls away when B is all of $I\!\!R^n$. Browder [2] may have been the first one to focus on the latter property, although he called it "firmly contractive"; the terminology soon changed, however; see [3]. The recent book [1] of Bauschke and Combettes offers a comprehensive treatment of monotonicity and firm nonexpansiveness in an infinite-dimensional framework of convex analysis and fixed-point iterations.

The great importance of the perhaps obscure-looking condition in (2.3) is the fact [4] that

$$\begin{aligned} &\text{a mapping } T \text{ is firmly nonexpansive from } I\!\!R^n \text{ into itself} \quad \Longleftrightarrow \\ &T = (I+A)^{-1} \ \text{ for a maximal monotone mapping } A : I\!\!R^n \rightrightarrows I\!\!R^n, \end{aligned} \tag{2.4}$$

For our purposes a local version of this will also be useful, and we articulate it as follows.

Proposition 1 (localized firm nonexpansiveness)**.** *A mapping T is firmly nonexpansive relative to a set $B \subset \operatorname{dom} T$ if and only if $T = (I+A)^{-1}$ for a mapping $A : I\!\!R^n \rightrightarrows I\!\!R^n$ that is maximal monotone relative to the set $W = \{\,(x,y)\,|\,x+y \in B\}$.*

Proof. There is no loss of generality in truncating the graph of T so as to have $B = \operatorname{dom} T = \{\,z\,|\,T(z) \neq \emptyset\}$. The heart of the matter then is the fact that, for a pair of mappings $A : I\!\!R^n \rightrightarrows I\!\!R^n$ and $T : I\!\!R^n \rightrightarrows I\!\!R^n$ related by

$$T = (I+A)^{-1}, \ \text{ or equivalently } \ A = T^{-1} - I,$$

the monotonicity of A corresponds to T having the property that

$$|x_1 - x_2|^2 + |(z_1 - x_1) - (z_2 - x_2)|^2 \leq |z_1 - z_2|^2 \ \text{ for } \ (z_1, x_1),\, (z_2, x_2) \in \operatorname{gph} T. \tag{2.5}$$

This is a relationship, already known to underly (2.4), which can be combined with the observation that (2.5) implies single-valuedness: if (z, x_1), $(z, x_2) \in \operatorname{gph} T$, then $x_1 = x_2$. Thus (2.5) is the same as T being firmly nonexpansive relative to $\operatorname{dom} T$ and indicates moreover that

$$\operatorname{gph} A = \{\,(T(z), z - T(z))\,|\,z \in \operatorname{dom} T\}. \tag{2.6}$$

Only the automatic maximality of A in this situation with respect to the set W remains to be verified. Any monotone mapping can be extended be maximal monotone. Let $\overline{A}$ be such an extension of A. Then, by (2.4), the mapping $\overline{T} = (I + \overline{A})^{-1}$, the graph of which extends $\operatorname{gph} T$, is firmly nonexpansive on all of $I\!\!R^n$. The analog of (2.6) then holds for $\operatorname{gph} \overline{A}$ in terms of $\overline{T}$, but then the restriction of $\overline{T}$ to $\operatorname{dom} T$ must agree with T. Since $\overline{A}$ could have been any maximal extension of A, this confirms that no outside pair (x,y) with $x + y \in \operatorname{dom} T$ could have been added to $\operatorname{gph} A$ without upsetting the monotonicity. In other words, A is maximal monotone relative to the indicated set W. $\square$

Some of the mappings for which these properties will be of interest are subgradient mappings associated with lsc functions f that might be nonconvex. Their definitions will be recalled next, following [21]. At a point $x \in \operatorname{dom} f$, a vector y is a regular subgradient, denoted by $y \in \widehat{\partial} f(x)$, if

$$f(x') \geq f(x) + \langle y, x' - x \rangle + o(|x' - x|). \tag{2.7}$$

It is a general subgradient, denoted by $y \in \partial f(x)$, if

$$\exists\, y^\nu \in \widehat{\partial} f(x^\nu) \;\text{ such that }\; y^\nu \to y,\; x^\nu \to x,\; f(x^\nu) \to f(x). \tag{2.8}$$

Under convexity the error term in (2.7) is superfluous, and $\partial f(x) = \widehat{\partial} f(x)$.

Furthermore when f is convex, the subgradient mapping ∂f is maximal monotone. An easy argument for this, using the prox mappings for both f and its conjugate f^*, was given by Moreau in [12]. The converse fact, that if the mapping ∂f as defined above is maximal monotone, then f is convex, was proved much later by Poliquin [14]. Similarly, maximal strong monotonicity of ∂f is equivalent to strong convexity of f. More recently the question of local monotonicity of ∂f came up in the following way in our paper [18]. What does the following property,

$$\begin{array}{l} \exists\, \varepsilon > 0 \text{ and an open convex neighborhood } U \times V \text{ of } (\bar{x}, \bar{y}) \in \operatorname{gph} \partial f \text{ such that} \\ \partial f \text{ is maximal monotone relative to } W = \{\, (x, y) \in U \times V \mid f(x) < f(\bar{x}) + \varepsilon \}, \end{array} \tag{2.9}$$

say about the nature of f itself? The answer depended on a concept called the *variational convexity* of f with respect to $(\bar{x}, \bar{y})$, which means in terms of some ε, U, V and W as in (2.9) that

$$\begin{array}{l} \exists \text{ convex lsc function } h \text{ such that } W \cap \operatorname{gph} \partial h = W \cap \operatorname{gph} \partial f \\ \text{with } h \leq f \text{ on } U \text{ and } h(x) = f(x) \text{ when } (x, y) \in W \cap \operatorname{gph} \partial f. \end{array} \tag{2.10}$$

We showed in [18] that the variational convexity property in (2.10) is always sufficient for the local monotonicity in (2.9) (although with a possibly smaller choice of ε, U and V). Moreover it is necessary as long as $\bar{y} \in \widehat{\partial} f(\bar{x})$ instead of just $\bar{y} \in \partial f(\bar{x})$. (Whether that extra assumption of regularity is truly needed is an open question.)

Our work on coordinating these facts with properties of prox mappings, global and local, will be greatly aided by an appeal to a subclass of regular subgradients, called *proximal* subgradients. A vector y is a proximal subgradient of f at x, denoted by $y \in \partial^p f(x)$, if

$$\exists\, \lambda > 0,\; \delta > 0, \;\text{ such that }\; f(x') \geq f(x) + \langle y, x' - x \rangle - \frac{1}{2\lambda}|x' - x|^2 \;\text{ when }\; |x' - x| < \delta. \tag{2.11}$$

A valuable feature of proximal subgradients is their ability to replace regular subgradients in the definition of general subgradients:

$$\begin{array}{l} \text{the formula in (2.8) still describes } \partial f(x) \text{ if the} \\ \text{subgradients } y^\nu \text{ are restricted to being proximal.} \end{array} \tag{2.12}$$

See [21, Section 8I] for this and other background on proximal subgradients.

Due to our prox-boundedness assumption (1.7) on f, the local inequality in (2.11) can always be extended to a global inequality on $\mathbb{R}^n$ by taking λ smaller. Thus, in terms of the mappings $\partial^p_\lambda f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined for $\lambda > 0$ by

$$\partial^p_\lambda f(x) = \Big\{\, y \;\Big|\; \forall x',\; f(x') \geq f(x) + \langle y, x' - x \rangle - \frac{1}{2\lambda}|x' - x|^2 \Big\}, \tag{2.13}$$

we therefore have

$$\operatorname{gph} \partial^p f = \bigcup_{0<\lambda<\lambda_0} \operatorname{gph} \partial^p_\lambda f \ \text{ for any } \lambda_0 > 0. \tag{2.14}$$

The close connection between proximal subgradients and prox mappings can be seen by writing the inequality in (2.13) as a condition about x minimizing something:

$$f(x) = \min_{x'} \Big\{ f(x') - \langle y, x' - x\rangle + \frac{1}{2\lambda}|x' - x|^2 \Big\}. \tag{2.15}$$

A bit of algebra on quadratic terms, with z intervening as $x + \lambda y$, reveals that

$$y \in \partial^p_\lambda f(x) \quad \Longleftrightarrow \quad x \in \operatorname{prox}_{\lambda f}(z) \ \text{ for } \ z = x + \lambda y, \tag{2.16}$$

which yields

$$\partial^p_\lambda f = \lambda^{-1}[\operatorname{prox}^{-1}_{\lambda f} - I], \ \text{ or equivalently, } \ \operatorname{prox}_{\lambda f} = [I + \lambda \partial^p_\lambda f]^{-1}. \tag{2.17}$$

Proposition 2 (prox monotonicity and parameterization)**.** *The mappings* $\operatorname{prox}_{\lambda f} : I\!R^n \to I\!R^n$ *are monotone and related to each other graphically in the parameter* $\lambda > 0$ *by*

$$\lambda_2 < \lambda_1 \quad \Longrightarrow \quad \operatorname{prox}_{\lambda_2 f} \supset [(1-\lambda)I + \lambda \operatorname{prox}^{-1}_{\lambda_1 f}]^{-1} \ \textit{ for } \ \lambda = \lambda_2/\lambda_1 \in (0,1), \tag{2.18}$$

where in fact

$$x \in \operatorname{prox}_{\lambda_1 f}(z_1) \quad \Longrightarrow \quad \operatorname{prox}_{\lambda_2 f}(z_2) = \{x\} \ \textit{ for } \ z_2 = (1-\lambda)x + \lambda z_1. \tag{2.19}$$

Proof. The monotonicity has already been recorded in [21, 12.19], but we confirm it here with a simple argument. Suppose $x_i \in \operatorname{prox}_{\lambda f}(z_i)$ for $i = 1, 2$. Then

$$\lambda f(x_1) + \tfrac{1}{2}|x_1 - z_1|^2 \leq \lambda f(x_2) + \tfrac{1}{2}|x_2 - z_1|^2 \ \text{ and } \ \lambda f(x_2) + \tfrac{1}{2}|x_2 - z_2|^2 \leq \lambda f(x_1) + \tfrac{1}{2}|x_1 - z_2|^2,$$

from which it follows on adding the inequalities together that

$$0 \leq \tfrac{1}{2}|x_2 - z_1|^2 - \tfrac{1}{2}|x_1 - z_1|^2 + \tfrac{1}{2}|x_1 - z_2|^2 - \tfrac{1}{2}|x_2 - z_2|^2 = \langle x_1 - x_2, z_1 - z_2\rangle,$$

hence monotonicity. Next observe from the definition in (2.13) that having $\lambda_2 < \lambda_1$ implies $\partial^p_{\lambda_2} f \supset \partial^p_{\lambda_1} f$ in the graphical sense. Then from (2.18) we have $\lambda_2^{-1}[\operatorname{prox}^{-1}_{\lambda_2 f} - I] \supset \lambda_1^{-1}[\operatorname{prox}^{-1}_{\lambda_1 f} - I]$, which can be written with $\lambda = \lambda_2/\lambda_1$ as $\operatorname{prox}^{-1}_{\lambda_2 f} \supset \lambda[\operatorname{prox}^{-1}_{\lambda_1 f} - I] + I$. On inverting both sides, we get (2.18).

To verify the refinement in (2.19), we appeal to the interpretation of $y \in \partial^p_\lambda f(x)$ as meaning that x gives the minimum in (2.15). If that holds for λ_1 and z_1, then x is the unique minimizer of the expression on the right side of (2.15) for all smaller values λ_2. Passing that through (2.16) we get

$$x \in \operatorname{prox}_{\lambda_1 f}(z_1) \ \text{ for } \ z_1 = x + \lambda_1 y \quad \Longrightarrow \quad \{x\} = \operatorname{prox}_{\lambda_2 f}(z_2) \ \text{ for } \ z_2 = x + \lambda_2 y.$$

In this situation $\lambda_1^{-1}[z_1 - x] = y = \lambda_2^{-1}[z_2 - x]$, so that $z_2 = x + \lambda[z_1 - x]$ for $\lambda = \lambda_2/\lambda_1$. $\square$

3 Main results

We move on now to the issues raised in the introduction about the degree to which convexity properties of f are essential to the firm nonexpansiveness of prox_f or the local firm nonexpansiveness of the localized versions prox_f^B.

Theorem 1 (global characterization). *The convexity of f is not only sufficient for prox_f to be a firmly nonexpansive mapping, but also necessary.*

Proof. The sufficiency has long been known through the fact that the convexity of f entails the maximal monotonicity of ∂f and, through the optimality condition associated with the argmin in (1.1), makes $\operatorname{prox} f = (I + \partial f)^{-1}$.

In the other direction, under the assumption that prox_f is firmly nonexpansive, we know from the theory of firmly expansive mappings, as reviewed at the beginning of Section 2, that $\operatorname{prox}_f = (I+A)^{-1}$ for a maximal monotone mapping A. Then $(1-\lambda)I + \lambda \operatorname{prox}_f^{-1}$ is the mapping λA, likewise maximal monotone. The same then for its inverse, which as the case of (2.18) with $\lambda_1 = 1$ and $\lambda_2 = \lambda$ has its graph inside that of $\operatorname{prox}_{\lambda f}$. Since $\operatorname{prox}_{\lambda f}$ is itself monotone according to Proposition 2, the maximality prevents the graphical inclusion from being strict. Thus, $\operatorname{prox}_{\lambda f} = (1-\lambda)I + \lambda \operatorname{prox}_f^{-1}$ and in consequence $\lambda^{-1}[\operatorname{prox}_{\lambda f}^{-1} - I] = \operatorname{prox}_f^{-1} - I = A$ for all $\lambda \in (0,1)$.

This tells us through (2.17) that the mappings $\partial_\lambda^p f$ likewise all coincide with the maximal monotone mapping A, and the same then for $\partial^p f$ by (2.14). Hence $\partial^p f$ is maximal monotone. The limit process in the expression of ∂f in (2.8), which can be restricted to proximal subgradients y^ν as noted in (2.12), preserves monotonicity, so it follows now that ∂f is maximal monotone (and the limit process wasn't actually neeeded). Then, according to Poliquin [14], f must be convex. □

The corresponding result in the case of general f for the localized mappings prox_f^B provides less than an equivalence until extra assumptions in terms of $\operatorname{prox}_{\lambda f}^B$ for $\lambda \in (0,1)$ are supplied. But it is definitive nonetheless. Since prox_f^B is primarily of interest in the setting of the proximal point algorithm in finding a local minimizer of f, or at least a local critical point $\bar{x}$ in the sense that $0 \in \partial f(\bar{x})$, we focus on that case.

Theorem 2 (local characterization in general). *If f is variationally convex at $(\bar{x}, 0) \in \operatorname{gph} \partial f$ as described in (2.10) with a convex function h and a set $W = \{ (x,y) \in U \times V \mid f(x) \leq f(\bar{x}) + \varepsilon \}$, then $0 \in \widehat{\partial} f(\bar{x})$ and there is an open ball $B \subset U$ centered at $\bar{x}$ on which prox_f^B is firmly nonexpansive. Moreover the same holds then for all the mappings $\operatorname{prox}_{\lambda f}^B$ with $\lambda \in (0,1)$.*

Conversely, if $0 \in \widehat{\partial} f(\bar{x})$ and there is an open ball B centered at $\bar{x}$ on which prox_f^B and all the mappings $\operatorname{prox}_{\lambda f}^B$ for $\lambda \in (0,1)$ are firmly nonexpansive, then f is must be variationallly convex at $(\bar{x}, 0)$ (with respect to some choice of elements in the definition).

However, having just prox_f^B itself be firmly nonexpansive around $\bar{x}$ is not enough, in general, to necessitate that variational convexity.

Proof. Suppose f is variationally convex at $(\bar{x}, 0) \in \operatorname{gph} \partial f$ and let h, U, V and ε be as in the description (2.10) of that property. Then $0 \in \partial h(\bar{x})$, hence $\bar{x} \in \operatorname{argmin} h$ by convexity, in which case $\bar{x}$ must minimize f on U and in particular have $0 \in \widehat{\partial} f(\bar{x})$.

Note next that for any open ball $B \subset U$ centered at $\bar{x}$, having $x \in \operatorname{prox}_f^B(z)$ for $z \in B$ entails the first-order optimality condition that $z - x \in \partial f(x)$. By taking B small enough, we can ensure in this that $(x, z-x) \in U \times V$. Furthermore, through the fact that

$$x \in \operatorname{prox}_f^B(z) \text{ for } z \in B \implies f(x) + \tfrac{1}{2}|x - z|^2 \leq f(\bar{x}) + \tfrac{1}{2}|\bar{x} - z|^2,$$

we can guarantee, by taking B to be a smaller ball at $\bar{x}$ if necessary, that $f(x) \le f(\bar{x}) + \varepsilon$, so that $(x, z-x) \in W$. Then, in drawing on the property (2.10) of variational convexity, we also have $z-x \in \partial h(x)$, moreover with $h(x) = f(x)$. Since $z-x \in \partial h(x)$ is the necessary and sufficient condition for having $x \in \mathrm{prox}_h(z)$, we can conclude that prox_f^B reduces on B to prox_h. Because h is convex with $0 \in \partial h(\bar{x})$, that mapping is single-valued and firmly nonexpansive globally with $\mathrm{prox}_h(\bar{x}) = \bar{x}$ and takes B into itself. Hence prox_f^B itself is single-valued and firmly nonexpansive relative to B.

For any $\lambda \in (0,1)$, the function λf inherits variational convexity from f for the same $U \times V$ with respect to $\lambda\varepsilon$ and λh in place of ε and h. Then, by the same argument, $\mathrm{prox}_{\lambda f}$ is firmly nonexpansive relative to the same B.

Suppose now, for the establishing converse, that B is an open ball centered at $\bar{x}$ relative to which $\mathrm{prox}_{\lambda f}^B$ is firmly nonexpansive for all $\lambda \in (0,1]$. That property involves no points $x \notin B$, so we can simplify matters by redefining $f(x)$ to be ∞ outside the closure of B. The redefined function will still be lsc, and the global mappings $\mathrm{prox}_{\lambda f}$ for $\lambda \in (0,1]$ will be nonempty-valued with range in $\mathrm{cl}\, B$. For $z \in B$, having $x \in \mathrm{prox}_{\lambda f}^B(z)$ is equivalent to having $x \in B$ with

$$\lambda f(x') + \tfrac{1}{2}|x'-z|^2 \ge \lambda f(x) + \tfrac{1}{2}|x-z|^2 \ \text{ for all } \ x' \in B.$$

Then through lower semicontinuity the same inequality must hold on $\mathrm{cl}\, B$ and hence on all of $I\!\!R^n$, because $\mathrm{dom}\, f \subset \mathrm{cl}\, B$, so that $x \in \mathrm{prox}_{\lambda f}(z)$. This shows that

$$\mathrm{gph}[\mathrm{prox}_{\lambda f}^B] = [B \times B] \cap \mathrm{gph}[\mathrm{prox}_{\lambda f}]. \tag{3.1}$$

Having $T_\lambda = \mathrm{prox}_{\lambda f}^B$ be firmly nonexpansive on B is equivalent by Proposition 1 to having the mapping $A_\lambda = T_\lambda^{-1} - I$ be maximal monotone relative to $W = \{\,(x,y) \,|\, x+y \in B\}$. Then $\lambda^{-1}A_\lambda$ is maximal monotone relative to $\{\,(x,y) \,|\, x+\lambda y \in B\}$, which entails maximal monotonicity with respect to

$$W_\lambda^B = \{\,(x,y) \,|\, x \in B,\ x+\lambda y \in B\}, \tag{3.2}$$

inasmuch as $x \in B$ holds automatically for $x \in \mathrm{dom}\, A_\lambda$. On the other hand, because of (3.1), we have

$$y \in \lambda^{-1}A_\lambda(x) \iff x \in T_\lambda(x+\lambda y) \iff x \in \mathrm{prox}_{\lambda f}(x+\lambda y) \ \text{ with } \ (x, x+\lambda y) \in B \times B,$$

and therefore

$$\mathrm{gph}[\lambda^{-1}A_\lambda] = W_\lambda^B \cap \Big(\lambda^{-1}[\mathrm{prox}_{\lambda f}^{-1} - I]\Big).$$

But $\lambda^{-1}[\mathrm{prox}_{\lambda f}^{-1} - I] = \partial_\lambda^p f$, as noted in (2.17), so this tells us on the basis of $\lambda^{-1}A_\lambda$ being maximal monotone relative to W_λ that

$$\partial_\lambda^p f \ \text{ is maximal monotone relative to } W_\lambda^B. \tag{3.3}$$

Recall now that, when $0 < \lambda_2 < \lambda_1$, the graph of $\partial_{\lambda_2}^p f$ includes the graph of $\partial_{\lambda_1}^p f$, while noting that that $W_{\lambda_2}^B \supset W_{\lambda_1}^B$. The maximal monotonicity in (3.3) ensures than that

$$\mathrm{gph}[\partial_{\lambda_2}^p f] \ \text{ agrees with } \ \mathrm{gph}[\partial_{\lambda_1}^p f] \ \text{ in } \ W_{\lambda_1}^B. \tag{3.4}$$

Observing that W_λ^B increases to all of $B \times I\!\!R^n$ as $\lambda \to 0$, we see through (2.14) that $\mathrm{gph}\, \partial^p f$ must be maximal monotone relative to $B \times I\!\!R^n$. (If some pair $(x,y) \in B \in I\!\!R^n$ could be still be added to $\mathrm{gph}\, \partial^p f$ without interfering with monotonicity, then the same must be true for every $\mathrm{gph}\, \partial_\lambda^p f$. But for some λ small enough, $(x,y) \in W_\lambda^B$, so that would run into conflict with (3.3).)

Therefore ∂f must be maximal monotone relative to $B \times I\!R^n$ and agree there with $\partial^p f$, inasmuch as the limit process in (2.8) for constructing ∂f from $\partial^p f$ in accordance with (2.12) maintains monotonicity. This local monotonicity around $(\bar{x}, 0)$, together with having $0 \in \widehat{\partial} f(\bar{x})$, enables us to invoke the main result in [18] to confirm that f must be variationally convex for $(\bar{x}, 0)$.

The counterexample laid out below will justify the final claim in the theorem about firm nonexpansiveness of just prox_f falling short of implying variational convexity. □

Example. *There is a nonconvex lsc function $f : I\!R^2 \to (-\infty, \infty]$ having $(0,0) \in \widehat{\partial} f(0,0)$ without variational convexity there, and yet such that prox_f is firmly nonexpansive around $(0,0)$.*

Detail. To construct f, we start with the concave function $\theta(t) = 1 + 2t - t^2$ on $I\!R$ and fix a value $\tau \in (1,2)$, noting that $\theta'(1) = 0$ and hence θ has its maximum at 1. We define f on $I\!R^2$ by

$$f(x_1, x_2) = \begin{cases} x_1^2 \theta(|t|) \ \text{ when } \ x_1 \neq 0, \ |t| \leq \tau \ \text{ for } \ t = x_2/x_1^2, \\ 0 \ \text{ when } \ (x_1, x_2) = (0,0), \\ \infty \ \text{ otherwise.} \end{cases} \tag{3.5}$$

It is easy to see that f is lsc on $I\!R^2$ and even continuous relative to $\mathrm{dom}\, f$. Moreover it attains its minimum at $(0,0)$; that implies $(0,0) \in \widehat{\partial} f(0,0)$.

On the open subset of $I\!R^2$ where $x_1 \neq 0$ and $0 < |x_2|/x_1^2 < \tau$, f is twice continuously differentiable, but in view of the symmetry in (3.5), it will suffice to examine the partial derivatives when $x_1 > 0$ and $0 < x_2/x_1^2 < \tau$. Utilizing $dt/dx_1 = -2t/x_1$ and $dt/dx_2 = 1/x_1^2$, we calculate there that

$$\begin{array}{l} (df/dx_1)(x_1,x_2) = 2x_1\theta(t)dt/dx_1 + x_1^2\theta'(t) = 2x_1[\theta(t) - t\theta'(t)] = 2x_1[2 + t^2], \\ (df/dx_2)(x_1,x_2) = x_1^2\theta'(t)dt/dx_2 = \theta'(t) = 2 - 2t, \\ (d^2f/dx_1^2)(x_1,x_2) = 2[2+t^2] + 2x_1[2tdt/dx_1] = 4 - 2t^2, \\ (d^2f/dx_1\, dx_2)(x_1,x_2) = -2dt/dx_1 = 4tx_1^{-1}, \\ (d^2f/dx_2^2)(x_1,x_2) = -2dt/dx_2 = -2x_1^{-2}. \end{array} \tag{3.6}$$

In focusing on $\tau = 1$ we can investigate the behavior of f as the origin is approached along the parabola where $x_2 = x_1^2$. There f has gradient $(2x_1, 0)$ and a hessian matrix that fails to be monotone because of the negative second derivative in x_2 in (3.6). As $(x_1, x_2) \to (0,0)$ with $x_2 = x_1^2$, the gradient goes to $(0,0)$, but the nonmonotone hessians preclude the existence of a set W of the kind in (2.9) for $((0,0),(0,0))$ in which $\mathrm{gph}\, \partial f$ is monotone. Hence by, [18, Theorem 1], variational convexity must be lacking at this location.

Next we verify the claim about the firm nonexpansiveness of the mapping

$$\mathrm{prox}_f(z_1, z_2) = \underset{x_1,\, x_2}{\mathrm{argmin}} \{ f(x_1,x_2) + \tfrac{1}{2}|x_1 - z_1|^2 + \tfrac{1}{2}|x_2 - z_2|^2 \}. \tag{3.7}$$

The minimization behind (3.7) can first be carried out with respect to x_2, after which the residual can be minimized with respect to x_1. Symmetry allows us to concentrate on $x_1 > 0$ and $z_2 > 0$. The minimization in x_2 for fixed x_1 concerns the function

$$x_2 \mapsto f(x_1, x_2) + \tfrac{1}{2}|x_2 - z_2|^2 \ \text{ for } \ |x_2| \leq \tau x_1^2. \tag{3.8}$$

On the open intervals $(0, \tau x_1^2)$ and $(-\tau x_1^2, 0)$ the function (3.8) has second derivative $1 - 2x_1^{-2}$, as seen from the expression for $(d^2f/dx_2^2)(x_1,x_2)$ in (3.6). Thus, as long as

$$1 - 2x_1^{-1} < 0, \ \text{ or in other words } \ x_1^2 < 2, \tag{3.9}$$

it is concave on the closed intervals $[0,\tau x_1^2]$ and $[-\tau x_1^2,0]$. The minimum therefore has to be at one of the boundary points, and we can identify it by comparing them. It's obvious with $z_2>0$ that the value at $-\tau x_1^2$ is higher than at τx_1^2, so only the values at τx_1^2 and 0 are in contention. The issue then revolves around the sign of the difference, between those values, which we calulate as

$$\begin{aligned}&[f(x_1,\tau x_1^2)+\tfrac{1}{2}|\tau x_1^2-z_2|^2]-[f(x_1,0)+\tfrac{1}{2}|0-z_2|^2]\\&\quad=x_1^2\theta(\tau)-x_1^2\theta(0)+\tfrac{1}{2}|\tau x_1^2|^2-\tau x_1^2 z_2=x_1^2[\theta(\tau)-\theta(0)+\tau^2x_1^2-\tau z_2].\end{aligned}\tag{3.10}$$

When this is positive, the minumum is attained uniquely at 0, and that is sure to hold in particular if $\theta(\tau)-\theta(0)-\tau z_2>0$. Since $[\theta(\tau)-\theta(0)]/\tau=2-\tau>0$, that can be guaranteed by taking $z_2<2-\tau$. In summary with an appeal to symmetry,

$$\begin{array}{l}\text{under (3.9) and } |z_2|<2-\tau, \text{ the minimization in (3.7) comes down}\\ \text{to setting } x_2=0 \text{ and then minimizing } f(x_1,0)+\tfrac{1}{2}|x_1-z_1|^2 \text{ in } x_1.\end{array}\tag{3.11}$$

Here $f(x_1,0)=2x_1^2$, so the expression being minimized is convex with derivative $4x_2+[x_1-z_1]$ and vanishes when $x_1=z_1/5$. That x_1 satisfies (3.9) when $|z_1|$ is small enough. It follows that, for (z_1,z_2) in a small enough neighborhood of $(0,0)$,

$$\operatorname{prox}_f(z_1,z_2)=\operatorname{prox}_g(z_1,z_2)\ \text{ for }\ g(x_1,x_2)=\begin{cases}2x_1^2 & \text{if } x_2=0,\\ 0 & \text{if } x_2\neq 0.\end{cases}\tag{3.12}$$

Because g is convex, its prox mapping is firmly nonexpansive on $I\!R^2$. The claim that prox_f is firmly nonexpansive on a neighborhood of $(0,0)$ is thereby confirmed. □

Although this example demonstrates the need for more than just the firm nonexpansiveness of prox_f itself in Theorem 2, more can be said under an assumption of prox-regularity. The function f is *prox-regular* at $\bar{x}$ for the subgradient $0\in\partial f(\bar{x})$ if the subgradients $y\in\partial f(x)$ with (x,y) near enough to $(\bar{x},0)$ are all proximal subgradients as in (2.11) and that holds in a uniform sense [21, 13F]. Under our prox-boundedness assumption (1.7), this can be expressed conveniently as follows:

$$\begin{array}{l} f \text{ being prox-regular at } \bar{x} \text{ for } 0\in\partial f(\bar{x}) \text{ means that, for some } \lambda\in(0,1],\\ \operatorname{gph}\partial f \text{ agrees with } \operatorname{gph}\partial_\lambda^p f \text{ around } (\bar{x},0) \text{ in a set } W \text{ of the kind in (2.9).}\end{array}\tag{3.13}$$

Prox-regularity is a significant restriction, but many applications are well covered by it. For instance, a function $f=f_0+\delta_C$ is covered by it at $\bar{x}$ when f_0 is $\mathcal{C}^2$, or the max of a finite collection of $\mathcal{C}^2$ functions, and C is specified by a system of finitely many inequality and equation constraints for which a basic constraint qualification holds at $\bar{x}$, cf. [21, 13.32].

Theorem 3 (local characterization under prox-regularity). *Let f be prox-regular for $(\bar{x},0)\in\operatorname{gph}\partial f$. Then the variational convexity of f for $(\bar{x},0)$ is not just sufficient but also necessary for the existence of an open ball B centered at $\bar{x}$ on which prox_f^B is firmly nonexpansive.*

Proof. This builds on the proof of Theorem 2, from which we already know the sufficiency. As there, we take an open ball B centered at $\bar{x}$ relative to which $\operatorname{prox} f$ is firmly nonexpansive and, without losing generality, modify f so that $\operatorname{dom} f\subset\operatorname{cl} B$, thereby obtaining (3.1) in the case of $\lambda=1$:

$$\operatorname{gph}[\operatorname{prox}_f^B]=[B\times B]\cap\operatorname{gph}[\operatorname{prox}_f].\tag{3.14}$$

In denoting this truncated mapping by T, we get from its firm nonexpansiveness via Proposition 1 that $T=(!+A)^{-1}$ for a mapping A that is maximal monotone with respect to the open set $W=\{(x,y)\,|\,x+y\in B\}$, with $\operatorname{gph} A$ lying actually in the smaller open set $W^B=\{(x,y)\,|\,x\in B,\ x+y\in B\}$.

Drawing on the assumption of prox-regularity, which in particular makes $\partial f(\bar{x})$ coincide with $\widehat{\partial} f(\bar{x})$, let λ be as in (3.13). The key to our argument now will be the property in (3.19), which has the consequence that, for $\lambda \in (0,1)$,

$$\text{prox}_{\lambda f} \text{ is single-valued on the range } R_\lambda \text{ of the mapping } (1-\lambda)T + \lambda I. \tag{3.15}$$

That single-valuedness forces the inclusion in (3.18) in the case of $\lambda_1 = 1$ and $\lambda_2 = \lambda$ to hold as an equation on R_λ:

$$\text{prox}_{\lambda f} = [(1-\lambda)I + \lambda T^{-1}]^{-1} = [I + \lambda A]^{-1} \text{ when restricted to } R_\lambda.$$

Then the mapping $\lambda^{-1}[\text{prox}_{\lambda f}^{-1} - I]$, which is $\partial_\lambda^p f$ by (2.17), agrees graphically with A in the set $\{\, (x,y) \,|\, x + y \in R_\lambda \}$.

We will demonstrate next that R_λ is a neighborhood of $(\bar{x}, 0)$. This will show via (3.13) that ∂f agrees graphically with A in a small enough neighborhood of $(\bar{x}, 0)$ and thus is maximal monotone there. That will confirm the variational convexity of f for $(\bar{x}, 0)$ by virtue of [18, Theorem 1].

Let $\overline{A}$ be a maximal monotone extension of A and let $\overline{T} = (I + \overline{A})^{-1}$. Then $\overline{T}$ is firmly expansive globally and extends T beyond B. Likewise the mapping $(1-\lambda)\overline{T} + \lambda I$ is a global extension of $(1-\lambda)T + \lambda I$ that serves in particular as a nonexpansive mapping from $\mathbb{R}^n$ into itself with $\bar{x}$ as fixed point. At the same time, because $\overline{T}$ is monotone, the mapping $(1-\lambda)\overline{T} + \lambda I$ is strongly monotone. That implies its inverse is Lipschitz continuous from $\mathbb{R}^n$ into itself. Thus, the mapping $(1-\lambda)\overline{T} + \lambda I$ and its inverse both map $\mathbb{R}^n$ continuously onto itself. The image of the open ball B under $(1-\lambda)\overline{T} + \lambda I$ must therefore be an open set. But this is the same as the image of B under $(1-\lambda)T + \lambda I$, which by definition is R_λ. □

An earlier result relating variational convexity with prox-regularity appeared in [15], but without any connection to prox mappings.

References

[1] Bauschke H. H., and Combettes, P. L., *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Second edition, Springer, 2019.

[2] Browder, F. E., "Convergence theorems for sequences of nonlinear operators in Banach spaces." *Math. Zeitschrift* **100** (1967), 201–225.

[3] Bruck, R. E., "Nonexpansive projections on subsets of Banach spaces." *Pacific J. Math.* **47** (1973), 341–355.

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

[5] Chen, G., and Teboulle, M., "A proximal-based decomposition method for convex minimization problems." *Mathematical Programming* **64** (1994), 81–101.

[6] Eckstein, J., Bertsekas, D. P., "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators." *Mathematical Programming* **55** (1992), 293–318.

[7] Fletcher, J., and Moors, W. R., "Chebyshev sets." *J. Australian Math. Soc.* **98** (2015), 161–231.

[8] Minty, G. J., "Monotone (nonlinear) operators in Hilbert space." *Duke Mathematical J.* **29** (1962), 341–346.

[9] Mordukhovich, B. S., *Variational Analysis and Generalized Differentiation I: Basic Theory*. Springer-Verlag, 2006.

[10] Moreau, J.-J, "Fonctions convexes duales et points proximaux dans un espace hilbertien." *Compte Rendus Acad. Sci.* **255** (1962), 2897–2899.

[11] Moreau, J.-J, "Propriétés des applications 'prox'." *Compte Rendus Acad. Sci.* **256** (1963), 1069–1071.

[12] Moreau, J.-J, "Proximité et dualité dans un espace hilbertien." *Bulletin Soc. Math. de France* **93** (1965), 273–299.

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

[14] Poliquin, R. A., "Subgradient monotonicity and convex functions." *Nonlinear Analysis: Theory, Methods and Applications* **14** (1990), 385–398

[15] Poliquin, R. A., and Rockafellar, R. T., "Tilt stability of a local minimum." *SIAM J. Optimization* **8** (1998), 287–289.

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

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

[18] Rockafellar, R. T., "Variational convexity and local monotonicity of subgradient mappings," *Vietnam Journal of Mathematics* **47** (2019), 547–561.

[19] Rockafellar, R. T., "Decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity," *Set-Valued and Variational Analysis* **27** (2019), dx.doi.org/10.1007/s19913-019-00339-5.

[20] Rockafellar, R. T., and Wets, R. J-B, "Scenarios and policy aggregation in optimization under uncertainty." *Mathematics of Operations Research* **16** (1991), 119–147.

[21] Rockafellar, R. T., and Wets, R. J-B, *Variational Analysis*. Springer-Verlag, 1998.

[22] Spingarn, J., "Applications of the method of partial inverses to convex programming: decomposition," *Mathematical Programming* **32** (1985), 199-121.