A PROXIMAL AVERAGE FOR NONCONVEX FUNCTIONS: A PROXIMAL STABILITY PERSPECTIVE*

W. L. HARE†

Abstract. Bauschke, Lucet, and Trienis [SIAM Rev., 50 (2008), pp. 115–132] developed the concept of the proximal average of two convex functions. In this work we show the relationship between the proximal average and the Moreau envelope and exploit this relationship to develop stability theory for a generalized proximal average function. This approach allows us to extend the concept of the proximal average to include many nonconvex functions. The most basic theory requires only that the functions of interest be prox-bounded, while the most powerful results hold for prox-regular functions.

Key words. proximal average, Moreau envelope, proximal point mapping, nonconvex, prox-regular (compatible parameterization), sensitivity analysis

AMS subject classifications. Primary: 26B25, 90C31; Secondary: 90C26, 58C06, 58C20

DOI. 10.1137/07070913X

1. Introduction. Throughout this work we shall only consider proper functions that act on the Euclidean space $\mathbb{R}^n$, that is, functions $f : \mathbb{R}^n \to (-\infty, \infty]$ such that $f$ is finite-valued at least at one point.

The question of how to continuously transform one convex function into another was approached by Bauschke, Lucet, and Trienis in [3]. The transforming of the function $f_0$ into a new function $f_1$ can be thought of as the creation of a new parameterized function

$$F : \mathbb{R}^n \times \mathbb{R} \to (-\infty, \infty]$$

such that $F(x, 0) = f_0(x)$ and $F(x, 1) = f_1(x)$ for all $x \in \mathbb{R}^n$, and $F$ behaves “well” with respect to the parameter $\lambda$. In [3] the authors show that the “naive” approach to creating $F$ contains certain flaws in its application. In particular, if one defines $F(x, \lambda) := (1 - \lambda)f_0(x) + \lambda f_1(x)$, then $F$ will take on infinite values whenever either $f_0$ or $f_1$ does. As such, unless the domains of $f_0$ and $f_1$ agree, the approach will “fail.”

One method to avoid this difficulty is proposed in the form of the proximal average. To define the proximal average, Bauschke, Lucet, and Trienis make use of the Fenchel conjugate of a function. Recall that for a proper function $f$ the Fenchel conjugate $f^*$ is defined as

$$f^*(y) := \sup_x \{ (x, y) - f(y) \}.$$

With this in mind, the proximal average of two convex functions $f_0$ and $f_1$ is defined by

$$\text{(PA)}_{f_0, f_1} = \mathcal{PA} : \mathbb{R}^n \times \mathbb{R} \to (-\infty, \infty]$$

$$(x, \lambda) \mapsto \left( (1 - \lambda) \left( f_0 + \frac{1}{2}\eta \right)^* + \lambda \left( f_1 + \frac{1}{2}\eta \right)^* \right)(x) - \frac{1}{2}\eta(x),$$

*Received by the editors November 13, 2007; accepted for publication (in revised form) February 13, 2009; published electronically DATE.
http://www.siam.org/journals/siopt/x-x/70913.html
†IRMACS, Simon Fraser University, Burnaby V5A 1S6, BC, Canada (whare@irmacs.sfu.ca).
where \( q \) is the “norm-squared” function \( q(x) = |x|^2 \) (using the usual Euclidean norm).
(Note that, where appropriate, we shall retain the \( |x|^2 \) notation for ease of reading.) Using this formula, Bauschke, Lucet, and Trienis proceeded to show that the proximal average could be used to continuously transform one convex function into another [3].

**Theorem 1.1 (see [3]).** Let \( f_0 \) and \( f_1 \) be proper convex lower semicontinuous (lsc) functions. Then for any \( \lambda \in [0, 1] \), the proximal average \( PA \) is a well-defined convex function such that

1. \( PA(x, 0) = f_0(x) \) for all \( x \in \mathbb{R}^n \),
2. \( PA(x, 1) = f_1(x) \) for all \( x \in \mathbb{R}^n \), and
3. the function \( PA \) is epi-continuous in the sense that as \( \lambda \to \bar{\lambda} \), the epi-graph of \( F(\cdot, \lambda) \) converges setwise to the epi-graph of \( F(\cdot, \bar{\lambda}) \).

(Recall the epi-graph of \( F(\cdot, \lambda) \) is \( \{ (x, \alpha) : \alpha \geq F(x, \lambda) \} \).

The terminology “proximal average” dates back to 2004, when (1) was used in setting of Banach spaces to study situations when the proximal point algorithm converged weakly but not in norm [5, eq. (30)]. In particular, the authors made use of a relationship of the function in (1) to the average of the proximal point maps of \( f_0 \) and \( f_1 \) (the proximal point mapping is defined and discussed below). Since its initial conception, the proximal average has appeared in at least four other papers [4, 2, 13, 6]. In [4] the proximal average is used to reexamine how to generate “antiderivatives” of cyclically monotone operators. In [13], the authors examine computer algorithms for computing conjugates of functions and use the proximal average as a key example and numerical challenge. In [2] the notion of the proximal average is extended to take averages of finitely many convex functions, and the basic analytic properties of the proximal average is redeveloped for this setting. Finally, [6] studies the kernel average of two convex functions. (The kernel average replaces the \( \frac{1}{2}q \) in the proximal average with a more general kernel function.)

As mentioned, although Bauschke, Lucet, and Trienis develop and study the proximal average through the use of conjugate functions, one can easily show that the proximal average bares a close relationship to the Moreau envelope and its related proximal point mapping. The Moreau envelope (a.k.a. the proximal envelope) is, loosely speaking, the result of a smoothing technique applied to a given objective function. Developed by Moreau in the early 1960s (see [17] and [18]) the Moreau envelope has become useful in a diverse collection of areas. Most notably, the Moreau envelope has led to research in variational analysis (see [18], [1], [8], [19], [10], and [9] amongst many others) and algorithm design (see [14], [20], [7], [15], and [11] amongst many others).

We denote the Moreau envelope of the function \( f \) with the parameter \( r \) by \( e_r f \) and define it as follows:

\[
e_r f(x) := \inf_y \left\{ f(y) + \frac{r}{2} |y - x|^2 \right\}.
\]

The associated proximal point mapping, denoted by \( P_r f \), is the (possibly empty) set of all points where this infimum is achieved:

\[
P_r f(x) := \arg\min_y \left\{ f(y) + \frac{r}{2} |y - x|^2 \right\}.
\]

We refer to the parameter \( r \) as the prox-parameter and the point \( x \) as the prox-center. We call a function \( f \) prox-bounded if there exists \( r > 0 \) and a point \( \bar{x} \) such that \( e_r f(\bar{x}) > -\infty \). The infimum of the set of all such \( r \) is the threshold of prox-boundedness or simply threshold when the context is clear. There are several manners
to see a function is prox-bounded. For example, \( f \) is prox-bounded; if and only if \( f \) majorizes a quadratic; if and only if \( f + \frac{r}{2}q \) is bounded below for some \( r \); if and only if \( \lim \inf_{|x| \to \infty} \frac{f(x)}{|x|^2} > -\infty \) [21, Ex. 1.24].

Remark 1.2. When dealing with functions \( f \) which include both a function variable \( x \) and a parameter \( \lambda \), we will always assume that the Moreau envelope and proximal point mapping are computed only with respect to the variable and not the parameter. That is, throughout this work we will write \( e_r f \) to mean \( e_r f_\lambda(x) \), where \( f_\lambda(x) = f(x, \lambda) \).

It is a simple exercise to show that the Fenchel conjugate of a function is closely related to the Moreau envelope of the function. Indeed, in [21, Ex. 1.24] it is shown that for any \( x \in \mathbb{R}^n \) and \( r > 0 \),

\[
(f + \frac{r}{2}q)^*(rx) = (-e_r f + \frac{r}{2}q)(x).
\]

Using this we easily derive the following alternate definition for the proximal average:

\[
P_A(x, \lambda) = \left( (1 - \lambda) \left( f_0 + \frac{r}{2}q \right)^* + \lambda \left( f_1 + \frac{r}{2}q \right)^* \right)^*(x) - \frac{1}{2}q(x)
\]

\[
= \left( (1 - \lambda) \left( -e_1 f_0 + \frac{r}{2}q \right) + \lambda \left( -e_1 f_1 + \frac{r}{2}q \right) \right)^*(x) - \frac{1}{2}q(x)
\]

\[
= \left( - (1 - \lambda)e_1 f_0 - \lambda e_1 f_1 + \frac{r}{2}q \right)^*(x) - \frac{1}{2}q(x)
\]

\[
= -e_1 (-(1 - \lambda)e_1 f_0 - \lambda e_1 f_1)(x) + \frac{1}{2}q(x) - \frac{1}{2}q(x).
\]

This provides us with the final relation

\[
P_A(x, \lambda) = -e_1 (-(1 - \lambda)e_1 f_0 - \lambda e_1 f_1)(x).
\]

(A similar formula appears in the proof of Theorem 6.1 in [5], however, the final Moreau envelope notation is not used.)

In this work we reexamine the proximal average in terms of the Moreau envelope function. In doing so we will find that the framework of the Moreau envelope provides us with the opportunity to strengthen the properties of the proximal average, but only under a revised definition. We shall call this revised definition the NC-proximal average (for nonconvex proximal average) and show that it provides a good answer to the question of how to transform between two nonconvex functions.

The remainder of this work is organized as follows. In section 2 we define the NC-proximal average function and discuss when it is well defined. Basically, the NC-proximal average replaces the prox-parameter 1 in (4) with a general prox-parameter \( r \) plus a small perturbation term. In Proposition 2.5 we see the NC-proximal average is well defined whenever \( f_0 \) and \( f_1 \) are prox-bounded and \( r \) is sufficiently large. In order to examine the NC-proximal average further, we need several definitions from modern variational analysis. These definitions are reviewed in section 3. The goal of this paper is to examine the stability properties of the NC-proximal average and is accomplished in section 4. Interestingly, in order to achieve good results regarding the NC-proximal average, we do not require the strong condition that \( f_0 + \frac{r}{2}q \) be convex but instead use only a weaker local condition called prox-regularity (see section 3) and a limiting condition on the proximal point mapping (see condition 11). Finally in section 5 we provide some conclusions and open research questions regarding the NC-proximal average.

2. A generalized proximal average function. Equation (4) provides us with some insight into how the proximal average works. To begin, examining (4) at the
In Example 11.26 of [21] it is shown that a proper function $f$ satisfies $-e_r(-e_r f) = f$ if and only if $f + \frac{r}{2}q$ is convex and lsc. This provides us with some insight towards the basic workings of the proximal average. In particular, it provides us with an alternate explanation to parts i. and ii. of Theorem 1.1 and informs us that these results hold in a greater generality than was shown in [3]. Specifically, we see that parts i. and ii. of Theorem 1.1 hold if the functions $f_0 + \frac{1}{2}q$ and $f_1 + \frac{1}{2}q$ are convex (as opposed to $f_0$ and $f_1$ being convex).

This also provides us with some insight towards how one might generalize the proximal average to include some nonconvex functions. In particular, it suggests that one promising generalization of the proximal average would be the inclusion of a prox-parameter to create a definition of the form

$$-e_r(- (1 - \lambda) e_r f_0 - \lambda e_r f_1).$$

However, we will find that this particular formula is insufficient to develop strong results for nonconvex functions. In particular, we will find that the framework of Moreau envelope will provide us with more powerful results than the framework of conjugate functions, but only under the condition that the outer Moreau envelope uses a prox-parameter that is strictly larger than the prox-parameter used in the inner Moreau envelopes. That is, to invoke the full power of the Moreau envelope we will need a form

$$-e_{r_1}(-(1 - \lambda) e_{r_2} f_0 - \lambda e_{r_2} f_1),$$

with $r_1$ strictly greater than $r_2$. This is akin to the classical results of [21, Prop. 12.62], where Rockafellar and Wets show that the smoothness properties of such “double envelopes” are strong when the outer envelope has a strictly greater prox-parameter than the inner envelope. However, in order to apply [21, Ex. 11.26] we will require that $r_1 = r_2$, at least when $\lambda \in \{0, 1\}$. This leads to the key definition of the present work.

**Definition 2.1 (NC-proximal average).** Let the functions $f_0$ and $f_1$ be proper and prox-bounded. Let the prox-parameter $r > 0$ be greater than the threshold of prox-boundedness for both $f_0$ and $f_1$. Then the NC-proximal average of $f_0$ and $f_1$ is the parameterized function $\mathcal{PA}_r$ defined by

$$\mathcal{PA}_r : \mathbb{R}^n \times \mathbb{R} \mapsto (-\infty, \infty]$$

$$\quad (x, \lambda) \mapsto -e_{r + \lambda(1 - \lambda)}((-1 + \lambda)e_r f_0 - \lambda e_r f_1)(x).$$

To simplify notation, if $r$ is fixed, we will often use $F_\lambda$ to denote the parametrized function defined by

$$F_\lambda(x) := -(1 - \lambda)e_r f_0(x) - \lambda e_r f_1(x).$$

This simplifies (5) to $\mathcal{PA}_r(x, \lambda) = -e_{r + \lambda(1 - \lambda)} F_\lambda(x)$.

In the remainder of this paper we explore the generalized proximal average defined by (5). The remainder of this section is dedicated to showing the NC-proximal average is well defined in the sense that $\mathcal{PA}_r$ is a proper function in $x$ for each parameter.
\( \lambda \in [0,1] \). Indeed, in Proposition 2.5 we will show that \( \mathcal{PA}_r \) is a lower-\( C^2 \) function, defined next, in \( x \) whenever \( \lambda \in (0,1) \).

**Definition 2.2** (lower-\( C^2 \) function). A function \( f \) is lower-\( C^2 \) on an open set \( V \) if it is finite-valued on \( V \) and at any point \( x \in V \) the function appended with a quadratic term is convex on some open convex neighborhood \( V' \) of \( x \). The function is said to be lower-\( C^2 \) (with no mention of \( V \)) if the corresponding neighborhood \( V \) is the entire space \( \mathbb{R}^n \).

It is worth noting that to say a function \( f \) is lower-\( C^2 \) is to set the neighborhood \( V \) to be the entire space, and not the neighborhood \( V' \). The neighborhood \( V' \) is dependent on the point of interest \( x \) and is generally smaller than \( V \). In particular, all \( C^2 \) functions are lower-\( C^2 \), which would not be the case if the neighborhood \( V' \) was forced to be \( \mathbb{R}^n \) (consider the function defined by \( f(x) := -x^4 \)).

**Remark 2.3.** The original definition of lower-\( C^2 \) functions states the following:

The function \( f \) is lower-\( C^2 \) on an open set \( V \) if for each \( \bar{x} \in V \) there is a neighborhood \( V' \) of \( \bar{x} \) upon which a representation \( f(x) = \max_{t \in T} f_t(x) \) holds, where \( T \) is a compact set and the functions \( f_t \) are of class \( C^2 \) on \( V \) such that \( f_t, \nabla f_t, \) and \( \nabla^2 f_t \) depend continuously not just on \( x \in V \) but jointly on \( (t, x) \in T \times V \).

The definition we provide is shown equivalent to the original one in [21, Thm. 10.33].

In order to show that the NC-proximal average is well defined we require the following lemma regarding prox-boundedness. (In the following lemma we use \( \text{dom} f \) to represent the domain of the function \( f \).)

**Lemma 2.4** (properties of prox-boundedness). Let \( f \) and \( g \) be proper, lsc, prox-bounded functions, with thresholds \( \bar{r}_f \) and \( \bar{r}_g \) such that the intersect of their domains is nonempty, \( \text{dom} f \cap \text{dom} g \neq \emptyset \). Then \( f + g \) is lsc, proper, and prox-bounded with threshold less than or equal to \( \bar{r}_f + \bar{r}_g \). Also, for any \( \lambda \geq 0 \) the function \( \lambda f \) is lsc, proper, and prox-bounded, with threshold of prox-boundedness equal to \( \lambda \bar{r}_f \).

Furthermore, if \( r \) is greater than the threshold of prox-boundedness for \( f \), then 
\[ -e_rf \] is continuous, finite-valued, and prox-bounded. Moreover, \( -e_rf \) is a lower-\( C^2 \) function with threshold of prox-boundedness less than or equal to \( r \). Indeed, one has that \( e_r(-e_rf) \) is a proper function.

**Proof.** The operations in the first paragraph are easily confirmed from the definitions of lsc and proper (note, properness is preserved as \( \text{dom} f \cap \text{dom} g \neq \emptyset \)). To see \( f + g \) is prox-bounded, note that if \( f + r_\frac{1}{2}g \) and \( g + r_\frac{1}{2}g \) are bounded below, then \( f + r_\frac{1}{2}g + g + r_\frac{1}{2}g \) and \( \lambda f + \lambda r_\frac{1}{2}g \) are bounded below as well.

That \( e_r(-e_rf) \) is a proper function is shown in [21, Ex. 1.44], which therefore provides that the threshold of prox-boundedness is less than or equal to \( r \). That \( -e_rf \) is lower-\( C^2 \) whenever \( r \) is greater than the threshold of prox-boundedness is shown in [21, Ex. 10.32], which also shows \( -e_rf \) is continuous and finite-valued for such \( r \).

**Proposition 2.5** (\( \mathcal{PA}_r \) well defined). Let \( f_0 \) and \( f_1 \) be lsc, proper, prox-bounded functions and \( r \) be greater than the threshold of prox-boundedness for both functions. Then for all \( \lambda \in [0,1] \) the NC-proximal average is a proper function in \( x \). Furthermore, for all \( \lambda \in (0,1) \) the NC-proximal average defines a lower-\( C^2 \) function in \( x \).

Finally, if for either \( i = 0 \) or \( i = 1 \) one has that \( f_i + \frac{r}{2}q \) is convex, then \( \mathcal{PA}_r(x, i) = f_i(x) \) for all \( x \in \mathbb{R}^n \).

**Proof.** Since \( r \) is greater than the threshold of prox-boundedness for both functions, we know that both \( -e_rf_0 \) and \( -e_rf_1 \) are well defined. By Lemma 2.4 we know that \( -(1-\lambda)e_rf_0 + \lambda e_rf_1 \) is proper, lower-\( C^2 \), and prox-bounded, with threshold less than or equal to \( \lambda r + (1-\lambda)r = r \). Thus the Moreau envelope of \( -(1-\lambda)e_rf_0 + \lambda e_rf_1 \)
is well defined and proper whenever the prox-parameter is greater or equal to $r$ (as is the case when $\lambda \in [0, 1]$). Moreover, the Moreau envelope of $-\lambda e_r f_0 - (1 - \lambda) e_r f_1$ is lower-$C^2$ whenever the prox-parameter is greater than $r$ (as is the case when $\lambda \in (0, 1]$).

The second statement results from applying [21, Ex. 11.26(d)] to $PA(x, 0) = -e_r(-e_r f_0)(x)$ or $PA(x, 1) = -e_r(-e_r f_1)(x)$, as discussed earlier.

Proposition 2.5 shows that $PA_r$ is well defined and behaves nicely in the variable $x$ whenever $\lambda \in (0, 1)$. In order to examine the behavior of $PA_r$ in the parameter $\lambda$ we will require several definitions from modern variational analysis.

3. Prox-regularity and parametric prox-regularity. To continue our exploration of the NC-proximal average we must formalize our notation and recall some standard definitions of variational analysis. In general, we shall use the notation and definitions of [21, Chap. 8] and will appeal to [21] for many of the standard results of variational analysis.

We define the regular subdifferential of a function $f$ at a point $\bar{x} \in \text{dom } f$ via

$$\hat{\partial}f(\bar{x}) := \left\{ v : \lim_{x \to \bar{x}, x \neq \bar{x}} \inf f(x) - f(\bar{x}) - \langle v, x - \bar{x} \rangle \geq 0 \right\}$$

(the regular subdifferential being empty at any point where $f$ is infinite), and the (limiting) subdifferential

$$\partial f(\bar{x}) := \limsup_{x \to \bar{x}, f(x) \to f(\bar{x})} \hat{\partial}f(x).$$

For parameterized functions we shall often only examine the subdifferential with respect to the variable $x$ (and assume the parameter $\lambda$ to be fixed). Accordingly, we shall use $\partial_x f(x, \lambda)$ to denote $\partial f_{\lambda}(x)$, where $f_{\lambda}(x) = f(x, \lambda)$.

For a set $S$ we define the indicator function $\delta_S$ as

$$\delta_S(x) :=\begin{cases} 0 & x \in S, \\ \infty & x \notin S. \end{cases}$$

We define the regular normal cone to a set $S$ at a point $\bar{x} \in S$ by $\hat{N}_S(\bar{x}) := \hat{\partial}\delta_S(\bar{x})$ and the (limiting) normal cone by $N_S(\bar{x}) := \partial \delta_S(\bar{x})$. From the normal cone we then create the Clarke normal cone: $\hat{N}_S(\bar{x}) := \text{conv} N_S(\bar{x})$. All normal cones are defined to be empty for any $\bar{x}$ outside of the closure of $S$.

We say the set $S$ is regular at the point $\bar{x} \in S$ if it is locally closed at $\bar{x}$ and all definitions of the normal cones agree. We say the function $f$ is regular at the point $\bar{x} \in \text{dom } f$ if its epi-graph is regular at $(\bar{x}, f(\bar{x}))$: in this case, $\hat{\partial}f(\bar{x}) = \partial f(\bar{x})$. Finally we use the term regular (disregarding a point) to mean the object is regular at all possible points.

We shall refer to set-valued functions $F : \mathbb{R}^m \mapsto 2^{\mathbb{R}^m}$ as maps. For a map $F$ we define the coderivative at a point $\bar{x}$ relative to a point $\bar{v} \in F(\bar{x})$ evaluated at a point $v'$ as

$$x' \in D^*F(\bar{x}|\bar{v})(v') \iff (x', -v') \in N_{\text{gph } F}(\bar{x}, \bar{v}),$$

where gph$F$ is the graph of the function $F$ (gph$F := \{(x, v) : v \in F(x)\}$).

Next we introduce the notion of prox-regularity.

Definition 3.1 (prox-regularity). A function $f$ is prox-regular at a point $\bar{x}$ for a subgradient $\bar{v} \in \partial f(\bar{x})$ if $f$ is locally lsc at $\bar{x}$ and there exist $\varepsilon > 0$ and $\rho > 0$ such that

$$f(x') \geq f(x) + \langle v, x' - x \rangle - \frac{\rho}{2} |x' - x|^2$$

whenever $\lambda e_r(\bar{v}) + (1 - \lambda) e_r x(\bar{x}) \in N_{\text{gph } F}(\bar{x}, \bar{v})$, as discussed earlier.
is said to be prox-regular at \( \bar{x} \) for \( v \) if, in addition, \( f \) is continuous as a function of \((x, v) \in \text{gph}\partial f \) at \((\bar{x}, \bar{v})\). The function is said to be prox-regular at \( \bar{x} \) (with no mention of \( \bar{v} \)) if it is prox-regular at \( \bar{x} \) for every \( v \in \partial f(\bar{x}) \), and the function is said to be prox-regular if it is prox-regular at every \( \bar{x} \in \mathbb{R}^n \). In this case, the prox-regularity parameter may either be pointwise (i.e., dependent on \( x \in \mathbb{R}^n \)) or global (independent of \( x \in \mathbb{R}^n \)).

The concepts of convex, lower-C^2, prox-regular, and regular share a simple hierarchy. To begin, all convex functions are clearly lower-C^2. Furthermore, by [21, Prop. 13.33] lower-C^2 functions are prox-regular with a pointwise prox-regularity parameter. Finally, if a function is prox-regular at a point, then it is regular at that point [21, p. 610].

Prox-regularity and the proximal point mapping share an intimate connection. Specifically, in [9, Thm. 2.3] it was shown that a function \( f \) is prox-regular at \( \bar{x} \) for \( \bar{v} \in \partial f(\bar{x}) \) if and only if there exists an \( f \)-attentive \( \varepsilon \)-localization \( T \) of \( \partial f \) around \((\bar{x}, \bar{v})\) such that for all \( r \) sufficiently large \( P_r f = (I + (1/r)T)^{-1} \) near \( \bar{x} + (1/r)\bar{v} \), with \( P_r f(\bar{x} + (1/r)\bar{v}) = \bar{x} \). (The result was shown for the notion of prox-regularity in Hilbert spaces but is clearly applicable to our current setting.) In particular, [9, Thm. 2.3] implies that without prox-regularity, the first order optimality conditions convey no control over the Moreau envelope and its associated proximal point mapping. This suggests that when studying the Moreau envelope, it is not unreasonable to work under the assumption of prox-regularity. Therefore, a reasonable starting point for the analysis of the NC-proximal average is the setting of prox-regularity.

However, examining the NC-proximal average, we note that it contains a parameter \( \lambda \). To explore the use of this parameter we shall need the notion of parametric prox-regularity. Parametric prox-regularity was first introduced in [12] to explore stability properties of local optimal solutions for parameterized prox-regular functions and later used in [9] to explore the stability properties of the proximal point mapping. The formal definition of parametric prox-regularity follows.

**Definition 3.2** (parametric prox-regularity). A proper lsc function

\[
\begin{align*}
    f : \mathbb{R}^n \times \mathbb{R}^s & \to (-\infty, \infty) \\
    (x, \lambda) & \mapsto f(x, \lambda)
\end{align*}
\]

is prox-regular in \( x \) at the point \( \bar{x} \), with compatible parameterization in \( \lambda \) at \( \bar{\lambda} \in \text{dom} f(\bar{x}, \cdot) \) for the subgradient \( \bar{v} \in \partial_x f(\bar{x}, \bar{\lambda}) \), with parameters \( \rho \) and \( \varepsilon \) if

\[
(8) \quad f(x', \lambda) \geq f(x, \lambda) + \langle v, x' - x \rangle - \frac{\rho}{2}|x' - x|^2 \quad \text{for all } x', \text{ with } |x' - x| < \varepsilon
\]

whenever \( |x - \bar{x}| < \varepsilon \), \( |\lambda - \bar{\lambda}| < \varepsilon \), \( v \in \partial_x f(x, \lambda) \), with \( |v - \bar{v}| < \varepsilon \), and \( |f(x, \lambda) - f(\bar{x}, \bar{\lambda})| < \varepsilon \). It is continuously prox-regular in \( x \) at \( \bar{x} \) for \( \bar{v} \) with compatible parameterization by \( \lambda \) at \( \bar{\lambda} \) if, in addition, \( f \) is continuous as a function of \((x, \lambda, v) \in \text{gph} \partial_x f \) at \((\bar{x}, \bar{\lambda}, \bar{v})\).

If the subgradient \( \bar{v} \), the parameter \( \bar{\lambda} \), or point \( \bar{x} \) is omitted, prox-regularity with compatible parameterization is meant for all appropriate subgradients \((v \in \partial_x f(\bar{x}, \bar{\lambda}))\), parameters \((\lambda \in \text{dom} f(\bar{x}, \cdot))\), or points \((x \in \text{dom} f)\).

Parametric prox-regularity captures the idea that a function with a parameter is prox-regular with regards to its variables and maintains that prox-regularity as the parameter changes. It also includes the constraint that the parameters of prox-regularity do not fluctuate wildly with small changes in the function’s parameter. The
next lemma provides an example of when parametric prox-regularity exists and will
be useful in section 4.

**Lemma 3.3.** Let the functions $f$ and $g$ be lower-$C^2$ on the open set $V$ and the
point $\bar{x}$ be in $V$. Consider the parameterized function

$$F(x, \lambda) := \lambda f(x) + (1 - \lambda)g(x), \quad x \in \mathbb{R}^n, \quad \lambda \in [0, 1]$$

and an arbitrary fixed parameter $\lambda \in [0, 1]$. Then $F$ is continuously prox-regular in $x$
at the point $\bar{x}$ with compatible parameterization in $\lambda$ at $\bar{\lambda}$.

**Proof.** By fixing the domain of the parameter $\lambda$ we need only show that parametric
prox-regularity holds for $\lambda \in [0, 1]$ (if $\lambda \notin [0, 1]$ we can consider the condition $|F(x, \lambda) - F(\bar{x}, \bar{\lambda})| < \varepsilon$ to never hold, and the definition is trivially true). Since $f$ and $g$ are
lower-$C^2$, there exists $\rho_f > 0$ and $\rho_g > 0$ and an open convex neighborhood $V'$
of $\bar{x}$ such that $f + \rho_f \frac{1}{2}g$ and $g + \rho_g \frac{1}{2}g$ are both convex on $V'$. Letting $\rho = \max\{\rho_f, \rho_g\}$ we have that $f + \rho \frac{1}{2}q$ and $g + \rho \frac{1}{2}q$ are both convex on $V'$. Therefore, for each $\lambda \in [0, 1]$, the function $H_\lambda$ defined by

$$H_\lambda(x) := F(x, \lambda) + \rho \frac{1}{2}q(x) + \delta_{V'}(x)$$

is the sum of two convex functions and therefore convex (recall, $\delta_{V'}$ represents the
indicator function of $V'$). Thus, for any $x, x' \in V'$ and $w \in \partial_2H_\lambda(x)$ we have

$$H_\lambda(x') \geq H_\lambda(x) + \langle w, x' - x \rangle,$$

$$F(x', \lambda) + \rho \frac{1}{2}|x'|^2 + \delta_{V'}(x') \geq F(x, \lambda) + \rho \frac{1}{2}|x|^2 + \delta_{V'}(x) + \langle w, x' - x \rangle,$$

$$F(x', \lambda) + \rho \frac{1}{2}|x'|^2 \geq F(x, \lambda) + \langle w, x' - x \rangle + \rho \frac{1}{2}|x|^2,$$

$$F(x', \lambda) \geq F(x, \lambda) + \langle w, x' - x \rangle + \rho \frac{1}{2}|x|^2 - \rho \frac{1}{2}|x'|^2.$$

Applying [21, Cor. 10.9] and the fact that all lower-$C^2$ functions are regular we have that

$$\partial_2H_\lambda(x) = \partial_2F(x, \lambda) + \partial_2\rho \frac{1}{2}|x|^2 = \partial_2F(x, \lambda) + \rho x.$$}

Therefore, there is $v \in \partial_2F(x, \lambda)$, with $w = v + \rho x$ and

$$F(x', \lambda) \geq F(x, \lambda) + \langle v + \rho x, x' - x \rangle + \rho \frac{1}{2}|x|^2 - \rho \frac{1}{2}|x'|^2,$$

$$F(x', \lambda) \geq F(x, \lambda) + \langle v, x' - x \rangle - \rho \left(\frac{1}{2}|x|^2 - \langle x, x' \rangle + \frac{1}{2}|x'|^2\right),$$

$$F(x', \lambda) \geq F(x, \lambda) + \langle v, x' - x \rangle - \rho \frac{1}{2}|x - x'|^2.$$}

This is stronger than the necessary conditions for parametric prox-regularity. Since $f$ and $g$ are continuous in $x$ (all lower-$C^2$ functions are continuous; see Remark 2.3) and $\lambda f + (1 - \lambda)g$ is continuous in $\lambda$, we have that $F$ is continuously prox-regular in $x$ at the point $\bar{x}$ with compatible parameterization in $\lambda$ at $\bar{\lambda}$. \qed

**4. Stability of the NC-proximal average.** From Lemma 3.3 we gain the
following corollary.

**Corollary 4.1.** Let $f_0$ and $f_1$ be lsc, proper, prox-bounded functions and $r$ be
greater than the threshold of prox-boundedness for both functions. Then the function
$F$ defined by

$$F(x, \lambda) := -\lambda e_r f_0(x) - (1 - \lambda)e_r f_1(x), \quad x \in \mathbb{R}^n, \quad \lambda \in [0, 1]$$
is continuously prox-regular in \( x \) at the point \( \bar{x} \) with compatible parameterization in \( \lambda \) at \( \bar{\lambda} \) for any \( \lambda \in [0, 1] \).

Proof. Example 10.32 of [21] shows that \( -e_r f \) is lower-C^2 whenever \( f \) is lsc, proper, prox-bounded, and \( r \) is greater than the threshold of prox-boundedness. Lemma 3.3 with \( f = -e_r f_0 \) and \( g = -e_r f_1 \) completes the proof. \( \square \)

Corollary 4.1 essentially shows that the proximal average of two lsc, proper, prox-bounded functions is the Moreau envelope of a parametrically prox-regular function. In [9] such functions were studied and their stability properties analyzed. In Theorem 4.2, below, we reiterate the main result of [9] in slightly greater detail.

**Theorem 4.2** (see [9, Cor. 4.7]). Let the proper lsc function

\[
F : \mathbb{R}^n \times \mathbb{R}^n \mapsto (-\infty, \infty]
\]

be continuously prox-regular at \( \bar{x} \), with compatible parameterization in \( \lambda \) at \( \bar{\lambda} \) for \( \bar{v} \in \partial_x F(\bar{x}, \bar{\lambda}) \). Assume further that \( F \) is prox-bounded with threshold \( \rho \), that \( F \) satisfies the constraint qualification

\[
(0, y) \in \partial^\infty F(\bar{x}, \bar{\lambda}) \Rightarrow y = 0,
\]

that

\[
(0, \lambda') \in D^\infty(\partial_x F)(\bar{x}, \bar{\lambda}|\bar{v}) (0) \Rightarrow \lambda' = 0,
\]

that for some \( \rho > 0 \)

\[
(x', \lambda') \in D^\infty(\partial_x F)(\bar{x}, \bar{\lambda}|\bar{v})(v') \Rightarrow \langle x', v' \rangle > -\rho |v'|^2,
\]

and that the set-valued mapping \( \partial_x F(\bar{x}, \cdot) \) has a continuous selection \( g \) near \( \bar{\lambda} \), with \( g(\bar{\lambda}) = \bar{v} \) (i.e., \( g(\lambda) \in \partial_x F(\bar{x}, \lambda) \) and \( g \) is continuous at \( \lambda \)).

If \( \bar{r} \) is sufficiently large that

(a) \( P_r F(x, \bar{\lambda}) \) is single-valued for \( x \in \mathbb{R}^n \),
(b) \( \bar{r} \) is greater than the \( \rho \) in (9), and
(c) \( \bar{r} \) is greater than the \( \rho \) in the definition of parametric prox-regularity ((8)),

then there exists \( K > 0 \) and a neighborhood of \( (\bar{x} + (1/\bar{r})\bar{v}, \bar{\lambda}, \bar{r}) \) such that for all \( x, x', \lambda, \lambda', r, r' \), in this neighborhood we have \( P_r F_\lambda(x) \) and \( P_r F_\lambda'(x') \) are single-valued, with

\[
|P_r F_\lambda(x) - P_r F_\lambda'(x')| \leq K |(r(x - \bar{x}) - r'(x' - \bar{x}), \lambda - \lambda', r - r')|
\]

where \( F_\lambda(x) = F(x, \lambda) \).

Proof. In Corollary 4.7 of [9] the above is stated with the phrase \( \bar{r} \) sufficiently large \( \) instead of the three conditions listed here. These conditions are easily extracted from the proof of [9, Cor. 4.7]. (The proof of [9, Cor. 4.7] also uses that \( \bar{r} \) is greater than the threshold of prox-boundedness, which is subsumed in condition (a) of Theorem 4.2.) \( \square \)

To study the stability properties of the NC-proximal average we will apply Theorem 4.2. In order to do this we need to show that for any \( \bar{\lambda} \in [0, 1] \), the following six properties of \( F(x, \lambda) := -\lambda e_r f_0(x) - (1 - \lambda)e_r f_1(x) \) hold:

1. The function \( F \) is continuously prox-regular at \( \bar{x} \) with compatible parameterization in \( \lambda \) at \( \bar{\lambda} \) for \( \bar{v} \in \partial_x f(\bar{x}, \bar{\lambda}) \);
2. The function $F$ is prox-bounded;
3. $(0, y) \in \partial^\infty F(\bar{x}, \bar{\lambda}) \Rightarrow y = 0$;
4. $(0, \lambda') \in D^*(\partial_x F)(\bar{x}, \bar{\lambda}; v)(0) \Rightarrow \lambda' = 0 (\bar{v} \in \partial f(\bar{x}, \bar{\lambda})$;
5. For some $\rho > 0$ we have $(x', \lambda') \in D^*(\partial_x F)(\bar{x}, \bar{\lambda}; v')(v') \neq 0 \Rightarrow \langle x', v' \rangle > -\rho|v'|^2$;
6. The set valued mapping $\partial_x F(\bar{x}, \cdot)$ has a continuous selection $g$ near $\bar{\lambda}$.

In the remainder of this work we show these six results hold and apply them to viewing the NC-proximal average in light of Theorem 4.2. The first three of these properties follow easily from the analysis already completed in this work and require nothing further than the functions $f_0$ and $f_1$ be proper, lsc, and prox-bounded.

**Proposition 4.3** (properties 1, 2, and 3). Let $f_0$ and $f_1$ be proper, lsc, prox-bounded functions and $r$ be greater than the threshold of prox-boundedness for both functions. Then properties 1, 2, and 3 hold.

**Proof.** Property 1 is given in Corollary 4.1. Property 2 is shown in Proposition 2.5. Property 3 follows from the fact that $F$ is lower-$C^2$ in $x$ and linear in $\lambda$. In particular, this implies $F$ is strictly continuous in $x$ and $\lambda$ (see [21, Thm. 10.31]), and therefore $\partial^\infty F(x, \lambda) = \{(0, 0)\}$ for all $(x, \lambda)$ [21, Thm. 9.13].

The remaining properties are more difficult to show and require slightly stronger conditions. Next we see that if the functions $f_0$ and $f_1$ are prox-regular, then properties 4, 5, and 6 hold. In order to simplify the proof of this we begin with the following lemma regarding properties 4 and 5.

In the lemma below, note that the notation $D^*(H)(\bar{x}, \bar{\lambda}|H(\bar{x}, \bar{\lambda}))$ is correct, as $H$ single-valued implies that only $H(\bar{x}, \bar{\lambda}) \in H(\bar{x}, \bar{\lambda})$.

**Lemma 4.4** (sufficient condition for properties 4 and 5). Suppose the function

$$H : \mathbb{R}^n \times \mathbb{R}^s \rightarrow (-\infty, \infty]$$

$$(x, \lambda) \mapsto H(x, \lambda)$$

is finite, single-valued, and Lipschitz continuous in $(x, \lambda)$ near the point $(\bar{x}, \bar{\lambda})$ with local Lipschitz constant $\text{lip}H$. Then

$$(0, \lambda') \in D^*H(\bar{x}, \bar{\lambda}|H(\bar{x}, \bar{\lambda}))(0) \Rightarrow \lambda' = 0,$$

and for $\rho > \text{lip}H$, one has

$$(x', \lambda') \in D^*H(\bar{x}, \bar{\lambda}|H(\bar{x}, \bar{\lambda}))(v'), \ v' \neq 0 \Rightarrow \langle x', v' \rangle > -\rho|v'|^2.$$  

**Proof.** Since $H$ is single-valued Lipschitz continuous near the point $(\bar{x}, \bar{\lambda})$, it satisfies the Mordukhovich criterion

$$D^*(H)(\bar{x}, \bar{\lambda}|H(\bar{x}, \bar{\lambda}))(0) = \{(0, 0)\}$$

(see [16, Prop. 2.8]). This is stronger than the desired property.

Next note that if $H$ is finite, single-valued, and Lipschitz continuous, then its coderivative map is nonempty-valued such that

$$(x', \lambda') \in D^*H(\bar{x}, \bar{\lambda}|H(\bar{x}, \bar{\lambda}))(v') \Rightarrow |\langle x', \lambda' \rangle| \leq (\text{lip}H)|v'|,$$

where $\text{lip}H$ is the local Lipschitz constant of $H$ [21, Prop. 9.24]. Since $|x'| \leq |\langle x', \lambda' \rangle|$, this implies that for any $(x', \lambda') \in D^*H(\bar{x}, \bar{\lambda}|H(\bar{x}, \bar{\lambda}))(v')$, $v' \neq 0$, we have

$$\langle x', v' \rangle \geq -|x'||v'| \geq -(\text{lip}H)|v'|^2 > -\rho|v'|^2,$$

where $\rho > \text{lip}H$.  


PROPOSITION 4.5 (properties 4, 5 and 6). Let $f_0$ and $f_1$ be lsc, proper, prox-bounded functions and $r$ be greater than the threshold of prox-boundedness for both functions. If $P_r f_0$ and $P_r f_1$ are single-valued Lipschitz continuous functions, then properties 4, 5, and 6 hold.

In particular, if the functions $f_0$ and $f_1$ are prox-regular, then properties 4, 5, and 6 hold.

Proof. We begin by applying [9, Thm. 2.4] to note that $P_r f_i$ is single-valued Lipschitz continuous if and only if $e_i f_i \in C^{1+}$, with

$$\nabla e_i f_i = r(I - P_r f_i),$$

where $i = 0, 1$ (and $I$ represents the identity function). This implies that

$$\partial_x F(\bar{x}, \lambda) = \nabla_x (-\lambda e_r f_0 - (1 - \lambda)e_r f_1)(\bar{x}, \lambda)$$

$$= -\lambda r(\bar{x} - P_r f_0(\bar{x})) - (1 - \lambda)r(\bar{x} - P_r f_1(\bar{x}))$$

$$= \lambda r P_r f_0(\bar{x}) + (1 - \lambda)r P_r f_1(\bar{x}) - r \bar{x},$$

$$= r(\lambda P_r f_0(\bar{x}) + (1 - \lambda)P_r f_1(\bar{x}) - \bar{x}),$$

which shows property 6 (as $\partial_x F(\bar{x}, \lambda)$ is linear in $\lambda$). Properties 4 and 5 now follow from the fact that the function $H$ defined by $H(x, \lambda) := r(\lambda P_r f_0(x) + (1 - \lambda)P_r f_1(x) - x)$ is single-valued Lipschitz continuous, and Lemma 4.4.

The final statement follows from [9, Thm. 2.4] or more directly, from [19, Thm. 4.4].

We are now in a position to apply Theorem 4.2 to the setting of the NC-proximal average.

THEOREM 4.6 (stability of $\mathcal{P} A_r$). Let $f_0$ and $f_1$ be lsc, proper, prox-bounded functions and $r$ be greater than the threshold of prox-boundedness for both functions. Suppose $P_r f_0$ and $P_r f_1$ are single-valued Lipschitz continuous functions, as is the case if the functions $f_0$ and $f_1$ are prox-regular. Then the NC-proximal average is well defined and proper for any $\lambda \in [0, 1]$.

Furthermore, suppose that when $r$ is sufficiently large the Lipschitz constant of $\lambda P_r f_0 + (1 - \lambda)P_r f_1 - I$ is bounded by 1, so

$$\text{lip}(r(\lambda P_r f_0 + (1 - \lambda)P_r f_1 - I)) \leq r,$$

then, for $r$ sufficiently large and $\lambda \in (0, 1)$ we have

i. $\mathcal{P} A_r$ is $C^{1+}$ in $x$,

ii. $\mathcal{P} A_r$ is locally Lipschitz continuous in $\lambda$, and

iii. $\nabla_x \mathcal{P} A_r$ is locally Lipschitz continuous in $\lambda$.

If for either $i = 0$ or $i = 1$ one has that $f_i + \frac{r}{2} q$ is convex, then $\mathcal{P} A_r(x, i) = f_i(x)$ for all $x \in \mathbb{R}^n$.

Remark 4.7. Note that ii. of Theorem 4.6 does not follow from iii. of Theorem 4.6, as the derivative is taken in the $x$ variable and continuity is in the $\lambda$ parameter.

Proof. The NC-proximal average is well defined and finite-valued by Proposition 2.5.

As usual, let the function $F : \mathbb{R}^n \times \mathbb{R} \mapsto (-\infty, \infty]$ be defined by $F(x, \lambda) := -\lambda e_r f_0(x) - (1 - \lambda)e_r f_1(x)$. In order to apply Theorem 4.2 to $F$ we need to show, for $r$ sufficiently large and any $\epsilon > 0$, that $r + \epsilon$ satisfies the three conditions laid out for $\bar{r}$ in Theorem 4.2, and the properties 1 through 6 hold.

First we examine the three conditions laid out for $\bar{r}$ in Theorem 4.2. To see that $r + \epsilon$ satisfies condition (b), $r$ is greater than the $\rho$ in (9), we note that, by Lemma
\(4.4\), the \(\rho\) in (9) is any value greater than the local Lipschitz constant of \(\partial_x F\). Recall that
\[
\partial_x F = r(\lambda P_r f_0 + (1 - \lambda)P_r f_1 - I).
\]

By condition (11) we assume that for \(r\) sufficiently large, the local Lipschitz constant of \(\partial_x F\) is bounded by \(r\) and therefore less than \(r + \epsilon\) for any \(\epsilon\). Condition (c) of Theorem 4.2, that \(r\) is greater than the \(\rho\) in (8), also follows, as we note that the \(\rho\) in (8) is always less than the local Lipschitz constant of \(\partial_x F\) (see the proof of [21, Prop. 13.34], for example). Finally, to see that \(r + \epsilon\) satisfies condition (a) of Theorem 4.2, \(P_r F(x, \lambda)\) is single-valued for all \(x \in \mathbb{R}^n\), note that by [21, Prop. 13.37] \(P_r F\) is monotone, single-valued, and Lipschitz continuous whenever \(r\) is greater than the threshold of prox-boundedness for \(F\) (details appear in the proof of [21, Prop. 13.37]). By condition (c) of Theorem 4.2, we have that \(r + \epsilon\) is greater than the \(\rho\) is the definition of prox-regular, and by Lemma 2.4, we have that \(r + \epsilon\) is greater than the threshold of prox-boundedness for \(F\).

To see that properties 1 through 6 hold we appeal to Propositions 4.3 and 4.5. We may therefore apply Theorem 4.2.

Before doing this, let us note that the function \(g\) defined by \(g(\lambda) := \lambda(1 - \lambda)\) is Lipschitz continuous on \((0, 1)\) with Lipschitz constant 1:
\[
|\lambda(1 - \lambda) - \lambda'(1 - \lambda')| \leq |\lambda - \lambda'|.
\]
(We will use this inequality repeatedly throughout the remainder of this proof.)

Fix \(\bar{x}\) and \(\bar{\lambda} \in (0, 1)\). By Theorem 4.2 we have that there exists \(K > 0\) such that for any \(x, x'\) near \(\bar{x}\), and any \(\lambda, \lambda'\) near \(\bar{\lambda}\):
\[
|P_{\lambda + (1 - \lambda)}(x, \lambda) - \lambda'(1 - \lambda')F(x', \lambda')| 
\leq K\left|((r + \lambda(1 - \lambda))(x - \bar{x}) - (r + \lambda' (1 - \lambda'))(x' - \bar{x}),
\right.
\]
\[
\lambda - \lambda', r + \lambda(1 - \lambda) - r - \lambda'(1 - \lambda')|\right|.
\]
where \(F(x, \lambda) = -\lambda e_r f_0(x) - (1 - \lambda)e_r f_1(x)\). Simplifying we see that
\[
|P_{\lambda + (1 - \lambda)}(x, \lambda) - \lambda'(1 - \lambda')F(x', \lambda')| 
\leq K|\lambda - \lambda'| + K|\lambda(1 - \lambda) x - \lambda'(1 - \lambda') x' + K|\bar{x} + 1||\lambda(1 - \lambda) - \lambda'(1 - \lambda')| + K|\lambda - \lambda'|.
\]

Which shows,
\[
|P_{\lambda + (1 - \lambda)}(x, \lambda) - \lambda'(1 - \lambda')F(x', \lambda')| 
\leq K|\lambda(1 - \lambda) x - \lambda'(1 - \lambda') x'| + K(2 + |\bar{x}||\lambda - \lambda'|).
\]
As \(P_r F\) is single-valued, [9, Thm. 2.4] shows that \(P_r F = I - \frac{1}{r}\nabla x e_r F\). This yields
\[
|((r + \lambda(1 - \lambda) - (r + \lambda'(1 - \lambda')) - \frac{1}{r}\nabla x e_r F(x', \lambda'))| 
\leq (Kr + 1)|x - x'| + K|\lambda(1 - \lambda) x - \lambda'(1 - \lambda') x'| + 2(K + |\bar{x}|)|\lambda - \lambda'|,
\]
which is
\[
|((r + \lambda(1 - \lambda)) - (r + \lambda'(1 - \lambda')) - \frac{1}{r}\nabla x P A(x, \lambda) - \lambda'(1 - \lambda') x'| + 2(K + |\bar{x}|)|\lambda - \lambda'|.
\]

Setting $\lambda = \lambda'$ shows result i. of Theorem 4.6. Indeed,
\[
|\nabla_x \mathcal{P}A(x, \lambda) - \nabla_x \mathcal{P}A(x', \lambda)| \leq (Kr + 1 + K(\lambda(1 - \lambda))(r + \lambda(1 - \lambda)))|x - x'| \\
\leq (Kr + 1 + K)(r + 1)|x - x'|.
\]

Returning to (14) and setting $x = x'$ we find that
\[
|P_{r+\lambda(1-\lambda)}F(x, \lambda) - P_{r+\lambda'(1-\lambda')}F(x, \lambda')| \leq K(\lambda(1 - \lambda) - \lambda'(1 - \lambda'))|x| \\
+ K(2 + |\bar{x}|)|\lambda - \lambda'| \\
\leq K(|x| + 2 + |\bar{x}|)|\lambda - \lambda'|.
\]

Since $x$ is near $\bar{x}$, we may assume without loss of generality $|x| \leq |\bar{x}| + 1$ to conclude
\[
|P_{r+\lambda(1-\lambda)}F(x, \lambda) - P_{r+\lambda'(1-\lambda')}F(x, \lambda')| \leq \bar{K}|\lambda - \lambda'|,
\]
where $\bar{K} = K(3 + 2|\bar{x}|)$. Thus, if $y \in P_{r+\lambda(1-\lambda)}F(x, \lambda)$ and $y' \in P_{r+\lambda'(1-\lambda')}F(x, \lambda')$, then
\[
\mathcal{P}A_r(x, \lambda) = e_{r+\lambda(1-\lambda)}(-(1 - \lambda)e_r f_0 - \lambda e_r f_1) = (1 - \lambda)e_r f_0(y) + \lambda e_r f_1(y),
\]
\[
\mathcal{P}A_r(x, \lambda') = e_{r+\lambda'(1-\lambda')}(-(1 - \lambda')e_r f_0 - \lambda' e_r f_1) = (1 - \lambda')e_r f_0(y') + \lambda' e_r f_1(y'),
\]
and $|y - y'| \leq \bar{K}|\lambda - \lambda'|$. This implies
\[
|\mathcal{P}A_r(x, \lambda) - \mathcal{P}A_r(x, \lambda')| \\
= |(1 - \lambda)e_r f_0(y) + \lambda e_r f_1(y) - (1 - \lambda')e_r f_0(y') - \lambda' e_r f_1(y')| \\
\leq |e_r f_0(y') - e_r f_0(y)| + |\lambda'| e_r f_0(y') + \lambda |e_r f_1(y') - e_r f_1(y)| \\
\leq |e_r f_0(y') - e_r f_0(y)| + |\lambda||e_r f_0(y') - e_r f_0(y')| \\
+ |\lambda - \lambda'||e_r f_1(y') - e_r f_1(y')| + |\lambda||e_r f_1(y) - e_r f_1(y')| \\
\leq 2|e_r f_0(y') - e_r f_0(y')| + |e_r f_1(y) - e_r f_1(y')| \\
+ |e_r f_0(y')| + |e_r f_1(y')|)|\lambda - \lambda'|.
\]

Since $e_r f_0$ and $e_r f_1$ are both $C^1$ (see [9, Thm. 2.4]), we have the existence of $K_i$ such that $|e_r f_i(y) - e_r f_i(y')| \leq K_i|y - y'| \leq \bar{K}K_i|\lambda - \lambda'| (i = 0, 1), and the existence of $M_i$ such that $|e_r f_i(z)| \leq M_i$ for $z$ near $y$ ($i = 0, 1$). Thus we conclude that
\[
|\mathcal{P}A_r(x, \lambda) - \mathcal{P}A_r(x, \lambda')| \leq \bar{K}2K_i|\lambda - \lambda'| + \bar{K}K_i|\lambda - \lambda'| + (M_0 + M_1)|\lambda - \lambda'|,
\]
which shows $\mathcal{P}A_r$ is locally Lipschitz continuous in $\lambda$ (result ii. of Theorem 4.6).

If we set $x = x'$ in (15), we see that
\[
|(r + \lambda(1 - \lambda))^{-1}\nabla_x \mathcal{P}A(x, \lambda) - (r + \lambda'(1 - \lambda'))^{-1}\nabla_x \mathcal{P}A(x, \lambda')| \\
\leq K(\lambda(1 - \lambda)|x - x'| + \lambda'(1 - \lambda')|x - x'|) \\
\leq K|x - x'|.
\]

where $\bar{K} = K(3 + 2|\bar{x}|)$ as before (again applying the assumption that $|x| \leq |\bar{x}| + 1$). This shows that the function $G$ defined by $G(\lambda) := (r + \lambda(1 - \lambda))^{-1}\nabla_x \mathcal{P}A(x, \lambda)$ is Lipschitz continuous in $\lambda$. Since $(r + \lambda(1 - \lambda))$ is Lipschitz continuous in $\lambda$, we have
that \((r + \lambda(1 - \lambda))G(\lambda) = \nabla_x P_A(x, \lambda)\) is Lipschitz continuous in \(\lambda\) [21, Ex. 9.8(c)]. This is result iii. of Theorem 4.6.

The final remark regarding when \(f_i + \frac{i}{2}q\) is convex is a repeat of Proposition 2.5.

Condition (11) is applied twice in the proof of Theorem 4.6. First, to show condition (b) of Theorem 4.2 holds and then to show condition (c) holds. It is easy to see that for any prox-bounded function \(f\), the proximal point mapping \(P_r f\) converges to the identity (pointwise) as \(r \to \infty\):

\[
\text{Let } f + \frac{i}{2}q \text{ be bounded below by } M, \text{ then}
\]

\[
P_r f(x) \subseteq \{ f(y) + \frac{i}{2}|y - x|^2 \leq f(x) \} \subseteq \{ M + \frac{r}{2}|y - x|^2 \leq f(x) \} \to \{ x \} \text{ as } r \to \infty.
\]

Moreover, under prox-regularity \(P_r f\) is monotone for \(r\) sufficiently large [21, Prop. 13.37]. As \(I\) is maximally monotone and Lipschitz with constant 1, it seems extremely likely that condition (11) holds in some general sense. In Corollary 4.8 we provide one such case. However, a proof that this is always the case appears to be nontrivial, as Examples 4.9 and 4.10 demonstrate.

**Corollary 4.8** (preconvex functions). Let \(f_0\) and \(f_1\) be lsc, proper functions such that for some \(r > 0\), both \(f_0 + \frac{i}{2}q\) and \(f_1 + \frac{i}{2}q\) are convex. Then \(f_0\) and \(f_1\) are prox-regular and prox-bounded, and condition (11) holds. In particular, all results of Theorem 4.6 hold.

**Proof.** If both \(f_i + \frac{i}{2}q\) are convex \((i = 0, 1)\), then \(f_i\) is prox-bounded and lower-C^2 and therefore prox-regular. Next, note that

\[
P_i \left( f_i + \frac{r}{2}q \right) = P_{r+1} f_i.
\]

By [21, Prop. 12.19] \(I - P_i(f_i + \frac{i}{2}q)\) is nonexpansive, that is, Lipschitz with constant less or equal to 1. Thus

\[
lip \{ \lambda P_{r+1} f_0 + (1 - \lambda) P_{r+1} f_1 - I \} = lip \{ \lambda (I - P_{r+1} f_0) + (1 - \lambda) (I - P_{r+1} f_1) \} \\
\leq \lambda + (1 - \lambda) = 1.
\]

This provides condition (11).

In the following examples, we demonstrate that condition (11) is not trivially satisfied by some of the basic conditions generated by proximal point mappings: Lipschitz continuity and monotonicity. If the function \(f\) is prox-regular, then the proximal point mapping \(P_r f\) is Lipschitz continuous and monotone (when \(r\) is sufficiently large). One might conjecture that if a function \(g\) is Lipschitz continuous and monotone, then

\[
lip (g - I) \leq \lip (g) - \lip (I) = \lip (g) - 1.
\]

In this first example we show that this is not the case.

**Example 4.9** (difference of Lipschitz functions). Let the function \(g\) be defined by

\[
g(x) := \begin{cases} 
  x & \text{if } x < 0, \\
  0 & \text{if } x \in [0, 1], \\
  x - 1 & \text{if } x > 1.
\end{cases}
\]

Then \(\lip (g) = 1\) and \(\lip (g - I) = 1\).

A better hypothesis might be that \(\lip (g - I) \leq \max \{ \lip (g), \lip (I) \}\). However, this alone will not provide condition (11). In particular, although \(P_r f \to I\) (pointwise) as \(r \to \infty\), this is not equivalent to stating \(\lip P_r f \to 1\), as example 4.10 demonstrates.
Example 4.10 (Lipschitz behavior in limits). Let the functions \( g_k \) be defined by

\[
g_k(x) := \begin{cases} 
  x - 1/k + 1/k^2 & \text{if } x < -1/k^2, \\
  kx & \text{if } x \in [-1/k^2, 1/k^2], \\
  x + 1/k - 1/k^2 & \text{if } x > 1/k^2.
\end{cases}
\]

Then each \( g_k \) is maximally monotone and Lipschitz, and \( g_k \to I \) (pointwise) as \( k \to \infty \). However, the Lipschitz constant of \( g_k \) is \( k \), therefore unbounded as \( k \to \infty \).

It is not clear if \( \text{lip}_P r f \) must converge to 1 as \( r \to \infty \).

5. Conclusions. This work demonstrates several interesting results. First, the proximal average as defined in [3] is generalizable to include several nonconvex functions. In particular, if \( f_i + \frac{r}{2}q \) is convex for both \( i = 0 \) and \( i = 1 \), then the functions \( f_i \) are proper, lsc, prox-bounded, and prox-regular, and all of the results of Theorem 4.6 hold.

More interestingly, if one ignores the end points \( \lambda = 0 \) and \( \lambda = 1 \), then the properties required to maintain a well-behaved NC-proximal average become considerably less. For these interior parameter selections, only prox-regularity and condition (11) are required to maintain \( P_A^r \) to be \( C^{1+} \) in \( x \) and Lipschitz continuous in \( \lambda \). Moreover, although prox-regularity is sufficient for these properties to hold, it is stronger than necessary. For example, [21, p. 618] contains an example of a function that is not prox-regular but does maintain that \( P_r f \) is single-valued Lipschitz continuous. It is also possible that even this is not a necessary condition for the Lipschitz continuity of \( P_A^r \) in \( \lambda \). Furthermore, arguments are presented that demonstrate the condition (11) is likely to hold in some general sense. Gaining a better understanding of this provides inspiration for future work.

If we relax our concerns even further, one requires only the functions \( f_0 \) and \( f_1 \) to be lsc, proper, and prox-bounded to create a well-defined NC-proximal average. Although in this case \( P_A^r \) may not be well behaved in \( \lambda \); one retains \( P_A^r \) being lower-\( C^2 \).

Before discussing some open questions resulting from this work, we remark on the use and need for the perturbation \( \lambda(1 - \lambda) \) that is applied to the outer prox-parameter in defining this generalized proximal average. The need for this perturbation arises from the property that the Moreau envelope becomes well behaved only when the prox-parameter is strictly greater than the threshold of prox-boundedness. For example, if \( f_0 \) and \( f_1 \) are both the absolute value function \( |\cdot| \), then

\[
-\epsilon_r(\lambda(-\epsilon_r f_0) + (1 - \lambda)(-\epsilon_r f_1)) = -\epsilon_r(-\epsilon_r |\cdot|) = |\cdot|
\]

for all \( r > 0 \) by [21, Ex. 11.26]. Although this is a well-behaved function in the sense that it is convex, continuous, and finite-valued, it is not differentiable at 0, so the results of Theorem 4.6 do not hold (when the perturbation term is not added). However, if the outer prox-parameter is increased slightly, the resulting average, in some sense, becomes better behaved.

The perturbation function \( \lambda(1 - \lambda) \) is also somewhat arbitrary in its nature. We chose this function as the easiest example of a function that would suit the needs to complete the proof. Examining the proofs in this work, it would appear that we require the perturbation function to be Lipschitz continuous on \( (0, 1) \), strictly positive on \( (0, 1) \), and equal to 0 at 0 and 1. Clearly there are many functions that satisfy these requirements. However, none show any clear advantage over \( \lambda(1 - \lambda) \) at this time. (Although an advantage may appear in later research, if one examines algorithms to efficiently compute various generalized proximal averages.)
In Lemma 3.3 we provide an example of one method to create a parametrically prox-regular function. It is unknown if this example remains true if \( f \) and \( g \) are simply prox-regular (as opposed to lower-\( C^2 \)). The answer to this question is not necessary in this work but may prove useful in some future research. In particular, it may remove the need for condition (11) to show condition (b) of Theorem 4.2 holds. More generally, if

\[
F : \mathbb{R}^n \times \mathbb{R}^s \rightarrow (-\infty, \infty) \quad (x, \lambda) \mapsto \lambda_1^{\alpha_1} f_1(x) + \lambda_2^{\alpha_2} f_2(x) + \cdots + \lambda_s^{\alpha_s} f_s(x)
\]

is a polynomial sum in \( \lambda \) of well-behaved functions of \( x \), then is \( F \) parametrically prox-regular?

In Theorem 4.6 we make use of condition (11) twice. At the end of section 4 we provide some insights into this condition and some examples showing that it is not trivially obtained using the basic structure of proximal point mappings. However, it still appears likely that condition (11) occurs under some basic conditions. Understanding this better would lead to advancements in both the theory of generalized proximal averages and in proximal points in general. In particular, condition (11) provides insight into the behavior of the derivative of the Moreau envelope.

On a final note, it is worth remarking that many of the results of [3] may be expandable to a nonconvex setting in a much simpler manner than presented in this work. Specifically, suppose that \( f + \frac{r}{2} q \) and \( g + \frac{r}{2} q \) are convex functions. Then one could define a generalized proximal average of \( f \) and \( g \) by taking the proximal average of the functions \( f_0 := f + \frac{r}{2} q \) and \( f_1 := g + \frac{r}{2} q \) using the Moreau envelope description in (4). This would result in weaker results than those found in Theorem 4.6, as demonstrated by \( f_0 = f_1 = |\cdot| \) above. Moreover, the condition that \( f + \frac{r}{2} q \) and \( g + \frac{r}{2} q \) are convex functions is stronger than the conditions presented in Theorem 4.6.

Acknowledgments. The author is deeply indebted to the two anonymous referees for this work; their vigilance considerably altered the quality and scope of research within this paper, including insight that resulted in (16).

REFERENCES