SIAM J. OPTIM.
Vol. 30, No. 2, pp. 1366–1390

A PROXIMAL AVERAGE FOR PROX-BOUNDED FUNCTIONS*

J. CHEN†, X. WANG‡, AND C. PLANIDEN§

Abstract. We construct a proximal average for two prox-bounded functions, which recovers the classical proximal average for two convex functions. The new proximal average transforms continuously in epi-topology from one proximal hull to the other. When one of the functions is differentiable, the new proximal average is differentiable. We give characterizations for Lipschitz and single-valued proximal mappings and we show that the convex combination of convexified proximal mappings is always a proximal mapping. Subdifferentiability and behaviors of infimal values and minimizers are also studied.

Key words. almost differentiable function, arithmetic average, convex hull, epi-average, epi-convergence, Moreau envelope, Lasry–Lions envelope, prox-bounded function, proximal average, proximal hull, proximal mapping, resolvent, subdifferential operator

AMS subject classifications. 49J53, 26A51, 47H05, 26E60, 90C30

DOI. 10.1137/19M1287419

1. Introduction.

The proximal average provides a novel technique for averaging convex functions; see [5, 6]. The proximal average has been used widely in applications such as machine learning [24, 30], optimization [4, 14, 22, 23, 31], matrix analysis [17, 19], and modern monotone operator theory [27]. The proximal mapping of the proximal average is precisely the average of proximal mappings of the convex functions involved. Averages of proximal mappings are important in convex and nonconvex optimization algorithms [5, 18]. A proximal average for possible nonconvex functions has long been sought.

In this work, we propose a proximal average for prox-bounded functions, which enjoy a rich theory in variational analysis and optimization. This proximal average significantly extends the works of [6] from convex functions to nonconvex functions. The new average function provides an epi-continuous transformation between proximal hulls of functions and reverts to the convex proximal average definition in the case of convex functions. When studying the proximal average of nonconvex functions, two fundamental issues arise. The first is when the proximal mapping is convex valued; the second is when the function can be recovered from its proximal mapping. It turns out that resolving both difficulties requires the "proximal" condition in variational analysis.

*Received by the editors September 16, 2019; accepted for publication (in revised form) February 26, 2020; published electronically May 12, 2020.
https://doi.org/10.1137/19M1287419
Funding: The first author was partially supported by National Natural Science Foundation of China grant 11401487, and by the Fundamental Research Funds for the Central Universities, grant XDJK2020B048. The second author was partially supported by the Natural Sciences and Engineering Research Council of Canada.
†School of Mathematics and Statistics, Southwest University, Chongqing 400715, People's Republic of China (J.W.Chen713@163.com).
‡Corresponding author. School of Mathematics and Statistics, Southwest University, Chongqing 400715, People's Republic of China; and Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada (shawn.wang@ubc.ca).
§School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW, 2500, Australia (chayne@uow.edu.au).

1.1. Outline. The plan of the paper is as follows. In the remainder of this section, we give basic concepts from variational analysis, review related work in the literature, and state the blanket assumptions of the paper. In section 2, we prove some interesting and new properties of proximal functions, proximal mappings, and envelopes. In section 3, we give an explicit relationship between the convexified proximal mapping and the Clarke subdifferential of the Moreau envelope. In section 4, we provide characterizations of Lipschitz and single-valued proximal mappings. In section 5, we define the proximal average for prox-bounded functions and give a systematic study of its properties. Relationships to arithmetic average and epi-average and full epi-continuity of the proximal average are studied in section 6. We devote section 7 to optimal value and minimizers and convergence in minimization of the proximal average. In section 8, we investigate the subdifferentiability and differentiability of the proximal average. Finally, section 9 illustrates the difficulty when the proximal mapping is not convex valued.

Two distinguished features of our proximal average deserve to be singled out: whenever one of the function is differentiable, the new proximal average is differentiable; and the convex combinations of convexified proximal mappings is always a proximal mapping. While epi-convergence [1, 8] plays a dominant role in our analysis of convergence in minimization, the class of proximal functions, which is significantly broader than the class of convex functions, is indispensable for studying the proximal average. In carrying out the proofs later, we often cite results from the standard reference Rockafellar and Wets [26].

1.2. Constructs from variational analysis. In order to define the proximal average of nonconvex functions, we utilize the Moreau envelope and proximal hull. Henceforth, $\mathbb{R}^n$ is n-dimensional Euclidean space with Euclidean norm $\|x\| = \sqrt{\langle x, x\rangle}$ and inner product $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ for $x, y \in \mathbb{R}^n$.

DEFINITION 1.1. *For a proper function $f : \mathbb{R}^n \to\ ]-\infty, +\infty]$ and parameters $0 < \mu < \lambda$, the* Moreau envelope *function $e_\lambda f$ and* proximal mapping *are defined, respectively, by $e_\lambda f(x) = \inf_w \left\{f(w) + \frac{1}{2\lambda}\|w - x\|^2\right\}$ and*

$$\operatorname{Prox}_\lambda f(x) = \operatorname*{argmin}_w \left\{f(w) + \frac{1}{2\lambda}\|w - x\|^2\right\}.$$

The proximal hull *function $h_\lambda f$ is defined by $h_\lambda f(x) = \sup_w \left\{e_\lambda f(w) - \frac{1}{2\lambda}\|x - w\|^2\right\}$; and the* Lasry–Lions double envelope *$e_{\lambda,\mu} f$ is defined by*

$$e_{\lambda,\mu} f(x) = \sup_w \left\{e_\lambda f(w) - \frac{1}{2\mu}\|x - w\|^2\right\}.$$

DEFINITION 1.2. *The function $f : \mathbb{R}^n \to\ ]-\infty, +\infty]$ is* prox-bounded *if there exist $\lambda > 0$ and $x \in \mathbb{R}^n$ such that $e_\lambda f(x) > -\infty$. The supremum of the set of all such λ is the* threshold *λ_f of prox-boundedness for f.*

Any function $f : \mathbb{R}^n \to\ ]-\infty, +\infty]$ that is bounded below by an affine function has a threshold of prox-boundedness $\lambda_f = \infty$; cf. [26, Example 3.28]. A differentiable function f with a Lipschitz continuous gradient has $\lambda_f > 0$.

Our notation is standard. For every nonempty set $S \subset \mathbb{R}^n$, $\operatorname{conv} S$, $\operatorname{cl} S$, and ι_S denote the *convex hull*, *closure*, and *indicator function* of S, respectively. For a proper, lower semicontinuous (lsc) function $f : \mathbb{R}^n \to\ ]-\infty, +\infty]$, $\operatorname{conv} f$ is its convex hull and f^* is its *Fenchel conjugate*. We let $\inf f$ and $\operatorname{argmin} f$ denote the infimum and the set of minimizers of f on $\mathbb{R}^n$, respectively. We call f *level coercive* if $\liminf_{\|x\|\to\infty} \frac{f(x)}{\|x\|} > 0$ and *coercive* if $\liminf_{\|x\|\to\infty} \frac{f(x)}{\|x\|} = \infty$. We use $\partial f, \hat{\partial} f, \partial_L f, \partial_C f$ for the Fenchel subdifferential, Fréchet subdifferential, Mordukhovich limiting subdifferential, and Clarke subdifferential of f, respectively. At a point $x \in \operatorname{dom} f$, the *Fenchel subdifferential* of f at x is the set

$$\partial f(x) = \{s \in \mathbb{R}^n : \ f(y) \geq f(x) + \langle s, y - x\rangle \text{ for all } y \in \mathbb{R}^n\};$$

the *Fréchet subdifferential* of f at x is the set

$$\hat{\partial} f(x) = \{s \in \mathbb{R}^n : \ f(y) \geq f(x) + \langle s, y - x\rangle + o(\|y - x\|)\};$$

the *Mordukhovich limiting subdifferential* of f at x is

$$\partial_L f(x) = \{v \in \mathbb{R}^n : \ \exists \text{ sequences } x_k \xrightarrow{\text{f}} x \text{ and } s_k \in \hat{\partial} f(x_k) \text{ with } s_k \to v\},$$

where $x_k \xrightarrow{\text{f}} x$ means $x_k \to x$ and $f(x_k) \to f(x)$. We let $\operatorname{Id} : \mathbb{R}^n \to \mathbb{R} : x \mapsto x$ be the identity mapping and $\mathfrak{q} = \frac{1}{2}\|\cdot\|^2$. The mapping $J_{\mu\partial_L f} = (\operatorname{Id} + \mu\partial_L f)^{-1}$ is called the *resolvent* of $\mu\partial_L f$; cf. [26, page 539]. When f is locally Lipschitz at x, the *Clarke subdifferential* $\partial_C f$ at x is $\partial_C f(x) = \operatorname{conv} \partial_L f(x)$. For further details on subdifferentials, see [11, 21, 26]. For $f_1, f_2 : \mathbb{R}^n \to \,]-\infty, +\infty]$, the *infimal convolution* (or epi-sum) of f_1, f_2 is defined by $(\forall x \in \mathbb{R}^n)\ f_1 \Box f_2(x) = \inf_w \{f_1(x - w) + f_2(w)\}$, and it is exact at x if $\exists\ w \in \mathbb{R}^n$ such that $f_1 \Box f_2(x) = f_1(x - w) + f_2(w)$; $f_1 \Box f_2$ is exact if it is exact at every point of its domain.

1.3. Related work. A comparison to known work in the literature is in order. In [32], Zhang and in [33], Zhang, Crooks, and Orlando defined a lower compensated convex transform for $0 < \mu < +\infty$ by $C^l_\mu(f) = \operatorname{conv}(2\mu\,\mathfrak{q} + f) - 2\mu\,\mathfrak{q}$. The lower compensated convex transform is the proximal hull. In [33], Zhang, Crooks, and Orlando gave a comprehensive study on the average compensated convex approximation, which is an arithmetic average of the proximal hull and the upper proximal hull. While the proximal hull is a common ingredient, our work and theirs are completely different. By nature, the proximal mapping of the proximal average for convex functions is exactly the convex combination of proximal mappings of individual convex functions [6]. In [15], Hare proposed a proximal average by $\mathcal{PA}_{1/\mu} = -e_{1/(\mu+\alpha(1-\alpha))}(-\alpha e_{1/\mu} f - (1-\alpha) e_{1/\mu} g)$. For this average, $x \mapsto \mathcal{PA}_{1/\mu}(x)$ is $\mathcal{C}^{1+}$ for every $\alpha \in]0, 1[$ and enjoys other nice stabilities with respect to α [15, Theorem 4.6]. However, this average definition has two disadvantages. (i) When f, g are convex, it does not recover the proximal average for convex functions $-e_{1/\mu}(-\alpha e_{1/\mu} f - (1-\alpha) e_{1/\mu} g)$. (ii) Neither $\operatorname{Prox}_{1/(\mu+\alpha(1-\alpha))} \mathcal{PA}_{1/\mu}$ nor $\operatorname{Prox}_{1/\mu} \mathcal{PA}_{1/\mu}$ is the average of the proximal mappings $\operatorname{Prox}_{1/\mu} f$ and $\operatorname{Prox}_{1/\mu} g$. In [12], Goebel introduced a proximal average for saddle functions by using extremal convolutions. Some nice results about self-duality with respect to saddle function conjugacy and partial conjugacy are put forth and proved by Goebel [12]. Although Goebel's average becomes the proximal average for convex functions, in general, the proximal mapping of his average is not the convex combination of proximal mappings of saddle functions involved.

1.4. Blanket assumptions. Throughout the paper, the functions $f, g : \mathbb{R}^n \to \,]-\infty, +\infty]$ are proper, lsc, and prox-bounded with thresholds $\lambda_f, \lambda_g > 0$, respectively, $\bar{\lambda} = \min\{\lambda_f, \lambda_g\}$, $\lambda > 0$, $\mu > 0$, and $\alpha \in [0, 1]$.

2. Preliminaries. In this section, we collect several facts and present some auxiliary results on proximal mappings of proximal functions, Moreau envelopes, and proximal hulls, which will be used in what follows.

2.1. Relationship among three regularizations: $e_\lambda f$, $h_\lambda f$, and $e_{\lambda,\mu} f$. Some key properties about these regularizations are the following.

FACT 2.1 (see [26, Example 11.26]). *Let* $0 < \lambda < \lambda_f$.

(a) *The Moreau envelope* $e_\lambda f = -\left(f + \frac{1}{2\lambda}\|\cdot\|^2\right)^*\left(\frac{\cdot}{\lambda}\right) + \frac{1}{2\lambda}\|\cdot\|^2$ *is locally Lipschitz.*

(b) *The proximal hull satisfies* $h_\lambda f + \frac{1}{2\lambda}\|\cdot\|^2 = \left(f + \frac{1}{2\lambda}\|\cdot\|^2\right)^{**}$.

FACT 2.2 (see [26, Examples 1.44, 1.46, Exercise 1.29]). *Let* $0 < \mu < \lambda < \lambda_f$. *One has*

(a) $h_\lambda f = -e_\lambda(-e_\lambda f)$;

(b) $e_\lambda f = e_\lambda(h_\lambda f)$;

(c) $h_\lambda(h_\lambda f) = h_\lambda f$;

(d) $e_{\lambda,\mu} f = -e_\mu(-e_\lambda f) = h_\mu(e_{\lambda-\mu} f) = e_{\lambda-\mu}(h_\lambda f)$;

(e) $e_{\lambda_1}(e_{\lambda_2} f) = e_{\lambda_1+\lambda_2} f$ *for* $\lambda_1, \lambda_2 > 0$.

For more details about these regularizations, we refer the reader to [2, 3, 10, 16] and [26, Chapter 1].

2.2. Proximal functions. The concept of λ-proximal functions will play an important role. This subsection is dedicated to properties of λ-proximal functions.

DEFINITION 2.3. *We say that* f *is* λ-proximal *if* $f + \frac{1}{2\lambda}\|\cdot\|^2$ *is convex.*

The λ-*proximal* function is also called λ-*hypoconvex* [29, Definition 3.10].

LEMMA 2.4.

(a) *The negative Moreau envelope* $-e_\lambda f$ *is always* λ-*proximal.*

(b) *If* $e_\lambda f$ *is* $\mathcal{C}^1$, *then* $f + \frac{1}{2\lambda}\|\cdot\|^2$ *is convex, i.e.,* f *is* λ-*proximal.*

Proof. By Fact 2.1,

$$(\forall x \in \mathbb{R}^n)\ \frac{1}{2\lambda}\|x\|^2 - e_\lambda f(x) = \left(f + \frac{1}{2\lambda}\|\cdot\|^2\right)^*\left(\frac{x}{\lambda}\right). \tag{2.1}$$

(a) This is clear from (2.1).

(b) By (2.1), the assumption ensures that $\left(f + \frac{1}{2\lambda}\|\cdot\|^2\right)^*\left(\frac{x}{\lambda}\right)$ is differentiable. It follows from Soloviov's theorem [28] that $f + \frac{1}{2\lambda}\|\cdot\|^2$ is convex. ∎

While for convex functions, proximal mappings and resolvents are the same, they differ for nonconvex functions in general.

FACT 2.5 (see [26, Example 10.2]). *For any proper, lsc function* $f : \mathbb{R}^n \to\ ]-\infty, +\infty]$ *and any* $\mu > 0$, *one has* $(\forall x \in \mathbb{R}^n)\ P_\mu f(x) \subseteq J_{\mu\partial_L f}(x)$. *When* f *is convex, the inclusion holds as an equation.*

However, proximal functions have surprising properties.

PROPOSITION 2.6. *Let* $0 < \mu < \lambda_f$. *Then the following are equivalent:*

(a) $\operatorname{Prox}_\mu f = J_{\mu\partial_L f}$;

(b) f *is* μ-*proximal;*

(c) $\operatorname{Prox}_\mu f$ *is maximally monotone;*

(d) $\operatorname{Prox}_\mu f$ *is convex valued.*

Proof.

(b)⇒(a) See [26, Proposition 12.19] and [26, Example 11.26].

(a)⇒(b) As $\operatorname{Prox}_\mu f$ is always monotone, $(\operatorname{Prox}_\mu f)^{-1} = (\operatorname{Id} + \mu\partial_L f)$ is monotone and it suffices to apply [26, Proposition 12.19(c)⇒(b)].

(b)⇔(c) See [26, Proposition 12.19].

(c)⇒(d) This is clear; cf. [5, Proposition 20.31].

(d)⇒(c) By [26, Example 1.25], $\operatorname{Prox}_\mu f$ is nonempty, compact valued, and monotone with full domain. As $\operatorname{Prox}_\mu f$ is convex valued, it suffices to apply [20]. ∎

LEMMA 2.7. *Let f be λ-proximal and $0 < \mu < \lambda$. Then*
(a) $\operatorname{Prox}_\lambda f$ *is convex valued;*
(b) $\operatorname{Prox}_\mu f$ *is single valued.*
Consequently, $\operatorname{Prox}_\mu f$ *is maximally monotone if* $0 < \mu \leq \lambda$.

Proof.
(a) Observe that $e_\lambda f(x) = \inf_y \left\{ f(y) + \frac{1}{2\lambda}\|y\|^2 - \langle \frac{x}{\lambda}, y\rangle \right\} + \frac{1}{2\lambda}\|x\|^2$. Since $f + \frac{1}{2\lambda}\|\cdot\|^2 - \langle \frac{x}{\lambda}, \cdot\rangle$ is convex, $\operatorname{Prox}_\lambda f(x)$ is convex.

(b) This follows from the fact that $f + \frac{1}{2\mu}\|\cdot\|^2 - \langle \frac{x}{\mu}, \cdot\rangle$ is strictly convex and coercive. When $0 < \mu < \lambda$, $\operatorname{Prox}_\mu f$ is continuous and monotone, so maximally monotone by [26, Example 12.7]. For the maximal monotonicity of $\operatorname{Prox}_\lambda f$, apply (a) and [20] or Lemma 3.1. ∎

The set of proximal functions is a convex cone. It is easy to show the following.

PROPOSITION 2.8. *Let f_1 be λ_1-proximal and f_2 be λ_2-proximal. Then for any $\alpha, \beta > 0$, the function $\alpha f_1 + \beta f_2$ is $\frac{\lambda_1\lambda_2}{\beta\lambda_1 + \alpha\lambda_2}$-proximal.*

2.3. The proximal mapping of the proximal hull.

LEMMA 2.9. *Let $0 < \lambda < \lambda_f$. One has* $\operatorname{Prox}_\lambda(h_\lambda f) = \operatorname{conv} \operatorname{Prox}_\lambda f$.

Proof. Applying [26, Example 10.32] to $-e_\lambda f = -e_\lambda(h_\lambda f)$ yields

$$\operatorname{conv} \operatorname{Prox}_\lambda(h_\lambda f) = \operatorname{conv} \operatorname{Prox}_\lambda f.$$

Since h_λ is λ-proximal, by Lemma 2.7 we have $\operatorname{conv} \operatorname{Prox}_\lambda(h_\lambda f) = \operatorname{Prox}_\lambda(h_\lambda f)$. Hence the result follows. ∎

LEMMA 2.10. *Let $0 < \lambda < \lambda_f$. The following are equivalent:*
(a) $\operatorname{Prox}_\lambda(h_\lambda f) = \operatorname{Prox}_\lambda f$*;*
(b) f *is λ-proximal.*

Proof.
(a)⇒(b) Since $\operatorname{Prox}_\lambda(h_\lambda f) = \operatorname{conv} \operatorname{Prox}_\lambda(h_\lambda f)$, $\operatorname{Prox}_\lambda f$ is upper semicontinuous, convex, compact, and monotone with full domain, so maximally monotone in view of [20] or Lemma 3.1. By [26, Proposition 12.19], $f + \frac{1}{2\lambda}\|\cdot\|^2$ is convex, equivalently, f is λ-proximal by [26, Example 11.26].

(b)⇒(a) As f is λ-proximal, $\operatorname{Prox}_\lambda f$ is convex valued by Lemma 2.7. Then by Lemma 2.9, we have $\operatorname{Prox}_\lambda(h_\lambda f) = \operatorname{conv} \operatorname{Prox}_\lambda f = \operatorname{Prox}_\lambda f$. ∎

2.4. Proximal mappings and envelopes.

LEMMA 2.11. *Let $0 < \mu < \lambda < \bar{\lambda}$. The following are equivalent:*
(a) $e_\lambda f = e_\lambda g$*;*
(b) $h_\lambda f = h_\lambda g$*;*
(c) $\operatorname{conv}\left(f + \frac{1}{2\lambda}\|\cdot\|^2\right) = \operatorname{conv}\left(g + \frac{1}{2\lambda}\|\cdot\|^2\right)$*;*
(d) $e_{\lambda,\mu} f = e_{\lambda,\mu} g$*;*
(e) $\operatorname{conv} \operatorname{Prox}_\lambda f = \operatorname{conv} \operatorname{Prox}_\lambda g$*, and for some $x_0 \in \mathbb{R}^n$, $e_\lambda f(x_0) = e_\lambda g(x_0)$.*
Under any one of the conditions (a)–(e)*, one has $\overline{\operatorname{conv}} f = \overline{\operatorname{conv}} g$.*

Proof.
(a)⇒(b) We find $-e_\lambda f = -e_\lambda g \Rightarrow -e_\lambda(-e_\lambda f) = -e_\lambda(-e_\lambda g)$, which is (b).
(b)⇒(a) This follows from $e_\lambda f = e_\lambda(h_\lambda f) = e_\lambda(h_\lambda g) = e_\lambda g$.
(b)⇔(c) Since $\lambda < \bar{\lambda}$, we have that $f + \frac{1}{2\lambda}\|\cdot\|^2$ and $g + \frac{1}{2\lambda}\|\cdot\|^2$ are coercive, so $\operatorname{conv}\left(f + \frac{1}{2\lambda}\|\cdot\|^2\right)$ and $\operatorname{conv}\left(f + \frac{1}{2\lambda}\|\cdot\|^2\right)$ are lsc. Use $h_\lambda f = \operatorname{conv}\left(f + \frac{1}{2\lambda}\|\cdot\|^2\right) - \frac{1}{2\lambda}\|\cdot\|^2$ and $h_\lambda g = \operatorname{conv}\left(g + \frac{1}{2\lambda}\|\cdot\|^2\right) - \frac{1}{2\lambda}\|\cdot\|^2$, which follow from Fact 2.1.
(d)⇔(a) Invoking Fact 2.2, we have

$$\begin{aligned} e_{\lambda,\mu} f = e_{\lambda,\mu} g &\Leftrightarrow h_\mu(e_{\lambda-\mu} f) = h_\mu(e_{\lambda-\mu} g) \Leftrightarrow e_\mu(h_\mu(e_{\lambda-\mu} f)) = e_\mu(h_\mu(e_{\lambda-\mu} g)) \\ &\Leftrightarrow e_\mu(e_{\lambda-\mu} f) = e_\mu(e_{\lambda-\mu} g) \Leftrightarrow e_\lambda f = e_\lambda g. \end{aligned}$$

(a)⇒(e) The Moreau envelope $e_\lambda f(x) = e_\lambda g(x)$ for every $x \in \mathbb{R}^n$. Apply [26, Example 10.32] to $-e_\lambda f = -e_\lambda g$ to get $(\forall x \in \mathbb{R}^n)$ $(\operatorname{conv} \operatorname{Prox}_\lambda f(x) - x)\lambda^{-1} = (\operatorname{conv} \operatorname{Prox}_\lambda g(x) - x)\lambda^{-1}$, which gives (e) after simplifications.

(e)⇒(a) Since both $e_\lambda f$ and $e_\lambda g$ are locally Lipschitz and Clarke regular, $\operatorname{conv} \operatorname{Prox}_\lambda f = \operatorname{conv} \operatorname{Prox}_\lambda g$ implies $-e_\lambda f = -e_\lambda g + \text{constant}$ by [26, Example 10.32]. (As $-e_\lambda f + \frac{1}{2\lambda}\|\cdot\|^2$ is convex, this also follows from the Rockafellar integration for convex functions.) The constant has to be zero by $e_\lambda f(x_0) = e_\lambda g(x_0)$. Thus, (a) holds. The final claim follows from the equivalence of (a)–(d) and taking the Fenchel conjugate to $e_\lambda f = e_\lambda g$, followed by cancellation of terms and taking the Fenchel conjugate again. ∎

The notion of "proximal" is instrumental.

COROLLARY 2.12. *Let $0 < \mu \le \lambda < \bar{\lambda}$, and let f, g be λ-proximal. Then $e_\mu f = e_\mu g$ if and only if $f = g$.*

Proof. Since $\mu \le \lambda$, both f, g are also μ-proximal, so $f = h_\mu f, g = h_\mu g$ and Lemma 2.11 (a)⇔(b) applies. ∎

PROPOSITION 2.13. *Let $0 < \mu < \bar{\lambda}$, and let $\operatorname{Prox}_\mu f = \operatorname{Prox}_\mu g$. If f, g are μ-proximal, then $f - g \equiv$ constant.*

Proof. As $\operatorname{Prox}_\mu f = \operatorname{Prox}_\mu g$, by [26, Example 10.32], $\partial(-e_\mu f) = \partial(-e_\mu g)$. Since both $-e_\mu f, -e_\mu g$ are locally Lipschitz and Clarke regular, we obtain that there exists $-c \in \mathbb{R}$ such that $-e_\mu f = -e_\mu g - c$. Because f, g are μ-proximal, we have $f = -e_\mu(-e_\mu f) = -e_\mu(-e_\mu g - c) = -e_\mu(-e_\mu g) + c = g + c$, as needed. ∎

2.5. An example. The following example shows that one cannot remove the assumption of f, g being μ-proximal in Proposition 2.6, Corollary 2.12, and Proposition 2.13.

EXAMPLE 2.14. *Consider the function $f_k(x) = \max\{0, (1+\varepsilon_k)(1-x^2)\}$, where $\varepsilon_k > 0$. Then f_k is $1/(2(1+\varepsilon_k))$-proximal, but not $1/2$-proximal.*

Claim 1: *the functions f_k have the same proximal mappings and Moreau envelopes for all $k \in \mathbb{N}$. However, whenever $\varepsilon_{k_1} \ne \varepsilon_{k_2}$, $f_{k_1} - f_{k_2} \ne$ constant.* Indeed, simple calculus gives that for every $\varepsilon_k > 0$ one has

$$\operatorname{Prox}_{1/2} f_k(x) = \begin{cases} x & \text{if } x \ge 1, \\ 1 & \text{if } x \in (0,1), \\ \{-1, 1\} & \text{if } x = 0, \\ -1 & \text{if } x \in (-1, 0), \\ x & \text{if } x \le -1, \end{cases} \quad e_{1/2} f_k(x) = \begin{cases} 0 & \text{if } x \ge 1, \\ (x-1)^2 & \text{if } x \in (0,1), \\ (x+1)^2 & \text{if } x \in (-1,0), \\ 0 & \text{if } x \le -1. \end{cases}$$

Claim 2: $\operatorname{Prox}_{1/2} f_k \neq J_{1/2\partial_L f_k}$, *i.e., the proximal mapping differs from the resolvent.* Since $J_{1/2\partial_L f_k} = (\operatorname{Id} + 1/2\partial_L f_k)^{-1}$ and

$$\partial_L f_k(x) = \begin{cases} 0 & \text{if } x < -1, \\ [0, 2(1+\varepsilon_k)] & \text{if } x = -1, \\ -2(1+\varepsilon_k)x & \text{if } -1 < x < 1, \\ [-2(1+\varepsilon_k), 0] & \text{if } x = 1, \\ 0 & \text{if } x > 1, \end{cases}$$

we obtain

$$J_{1/2\partial_L f_k}(x) = \begin{cases} x & \text{if } x > 1, \\ 1 & \text{if } \varepsilon_k < x \leq 1, \\ \left\{-1, -\frac{x}{\varepsilon_k}, 1\right\} & \text{if } -\varepsilon_k \leq x \leq \varepsilon_k, \\ -1 & \text{if } -1 \leq x < -\varepsilon_k, \\ x & \text{if } x < -1, \end{cases}$$

which does not equal $\operatorname{Prox}_{1/2} f_k$ given above.

3. The convexified proximal mapping and Clarke subdifferential of the Moreau envelope.

The following result gives the relationship between the Clarke subdifferential of the Moreau envelope and the convexified proximal mapping.

LEMMA 3.1. *For $0 < \mu < \lambda_f$, the following hold.*

(a) *The convex hull* $\operatorname{conv} \operatorname{Prox}_\mu f = \partial \left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*$. *In particular,* $\operatorname{conv} \operatorname{Prox}_\mu f$ *is maximally monotone.*

(b) *The limiting subdifferential* $-\partial_L \left(-\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*\right) \subseteq \operatorname{Prox}_\mu f$.

(c) *The Clarke subdifferential*

$$\partial_C(e_\mu f) = -\partial_L(-e_\mu f) = \frac{\operatorname{Id} - \operatorname{conv} \operatorname{Prox}_\mu f}{\mu}. \tag{3.1}$$

If, in addition, f is μ-proximal, then

$$\partial_C(e_\mu f) = \frac{\operatorname{Id} - \operatorname{Prox}_\mu f}{\mu}. \tag{3.2}$$

Proof.

(a) By Fact 2.1,

$$-e_\mu f(x) = -\frac{1}{2\mu}\|x\|^2 + \left(f + \frac{1}{2\mu}\|\cdot\|^2\right)^* \left(\frac{x}{\mu}\right). \tag{3.3}$$

Using [26, Example 10.32] and the subdifferential sum rule [26, Corollary 10.9], we get

$$\frac{\operatorname{conv} \operatorname{Prox}_\mu f(x) - x}{\mu} = \partial_L(-e_\mu f)(x) = -\frac{x}{\mu} + \partial \left(f + \frac{1}{2\mu}\|\cdot\|^2\right)^* \left(\frac{x}{\mu}\right).$$

Simplification gives $\operatorname{conv} \operatorname{Prox}_\mu f(x) = \partial \left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*(x)$. Since $\mu f + \frac{1}{2}\|\cdot\|^2$ is coercive, we conclude that $\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*$ is a continuous convex function, so $\operatorname{conv} \operatorname{Prox}_\mu f$ is maximally monotone [26, Theorem 12.17].

(b) By (3.3), $-\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*(x) = \mu e_\mu f(x) - \frac{1}{2}\|x\|^2$. From [26, Example 10.32] we obtain $\partial_L(-(\mu f+\frac{1}{2}\|\cdot\|^2)^*)(x) = \partial_L(\mu e_\mu f)(x) - x \subseteq \mu\frac{x-\operatorname{Prox}_\mu f(x)}{\mu} - x = -\operatorname{Prox}_\mu f(x)$. Therefore, $-\partial_L(-(\mu f + \frac{1}{2}\|\cdot\|^2)^*)(x) \subseteq \operatorname{Prox}_\mu f(x)$.

(c) As $-e_\mu f$ is Clarke regular, using [26, Example 10.32] we obtain

$$\partial_C e_\mu f(x) = -\partial_C(-e_\mu f)(x) = -\partial_L(-e_\mu f)(x) = \frac{x - \operatorname{conv}\operatorname{Prox}_\mu f(x)}{\mu}.$$

If f is μ-proximal, then $\operatorname{Prox}_\mu f(x)$ is convex for all x, so (3.2) follows from (3.1). ◻

Remark 3.2. Lemma 3.1(a) and (c) extend [26, Exercise 11.27] and [26, Theorem 2.26], respectively, from convex functions to nonconvex functions.

It is tempting to ask whether $\partial_L(e_\mu f) = \frac{\operatorname{Id} - \operatorname{Prox}_\mu f}{\mu}$ holds. This is answered in the negative below.

Proposition 3.3. *Let $0 < \lambda < \lambda_f$ and $\psi = h_\lambda f$. Suppose that there exists $x_0 \in \mathbb{R}^n$ such that $\operatorname{Prox}_\lambda f(x_0)$ is not convex. Then $\partial_L e_\lambda \psi(x_0) \neq \frac{x_0 - \operatorname{Prox}_\lambda \psi(x_0)}{\lambda}$.*

Proof. We prove by contrapositive. Suppose the result fails, i.e.,

$$\partial_L e_\lambda \psi(x_0) = \frac{x_0 - \operatorname{Prox}_\lambda \psi(x_0)}{\lambda}. \tag{3.4}$$

In view of $e_\lambda \psi = e_\lambda f$ and [26, Example 10.32], we have

$$\partial_L e_\lambda \psi(x_0) = \partial_L e_\lambda f(x_0) \subseteq \frac{x_0 - \operatorname{Prox}_\lambda f(x_0)}{\lambda}. \tag{3.5}$$

Since $\operatorname{Prox}_\lambda \psi = \operatorname{conv}\operatorname{Prox}_\lambda f$ by Lemma 2.9, (3.4) and (3.5) give

$$\frac{x_0 - \operatorname{conv}\operatorname{Prox}_\lambda f(x_0)}{\lambda} \subseteq \frac{x_0 - \operatorname{Prox}_\lambda f(x_0)}{\lambda},$$

which implies that $\operatorname{Prox}_\lambda f(x_0)$ is a convex set. This is a contradiction. ◻

We give an example illustrating the conditions of Proposition 3.3.

Example 3.4. *Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = -|x|$. Then $e_\lambda f = f - \frac{\lambda}{2}$; and $\operatorname{Prox}_\lambda f(x) = x + \lambda$ if $x > 0$, $\{-\lambda, \lambda\}$ if $x = 0$, and $x - \lambda$ if $x < 0$. In particular, $\operatorname{Prox}_\lambda f(0)$ is not convex. Lemma* 2.9 *gives* $\operatorname{Prox}_\lambda \psi(0) = \operatorname{conv}\operatorname{Prox}_\lambda f(0) = [-\lambda, \lambda]$. *Thus, $\partial_L e_\lambda \psi(0) = \partial_L e_\lambda f(0) = \{-1, 1\} \neq [-1, 1] = \frac{0 - \operatorname{Prox}_\lambda \psi(0)}{\lambda}$.*

4. Characterizations of Lipschitz and single-valued proximal mappings.

Simple examples show that proximal mappings can be wild, although always monotone.

Fact 4.1 (see [26, Example 7.44]). *Let $f : \mathbb{R}^n \to \,]-\infty, +\infty]$ be proper, lsc, and prox-bounded with threshold λ_f, and $0 < \mu < \lambda_f$. Then $\operatorname{Prox}_\mu f$ is always upper semicontinuous and locally bounded.*

The following characterizations of the proximal mapping are of independent interest.

Proposition 4.2 (Lipschitz proximal mapping). *Let $0 < \mu < \lambda_f$. Then the following are equivalent.*

(a) *The proximal mapping $\operatorname{Prox}_\mu f$ is single valued and Lipschitz continuous with constant $\kappa > 0$;*

(b) *the function $f + \frac{\kappa - 1}{2\mu\kappa}\|\cdot\|^2$ is convex.*

Proof.

(a)⇒(b) By Lemma 3.1(a), $\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*$ is differentiable and its gradient is Lipschitz continuous with constant κ. By Soloviov's theorem [28], $\mu f + \frac{1}{2}\|\cdot\|^2$ is convex. It follows from [26, Proposition 12.60] that $\mu f + \frac{1}{2}\|\cdot\|^2$ is $\frac{1}{\kappa}$-strongly convex, i.e., $\mu f + \frac{1}{2}\|\cdot\|^2 - \frac{1}{\kappa}\frac{1}{2}\|\cdot\|^2$ is convex. Equivalently, $f + \frac{\kappa-1}{2\mu\kappa}\|\cdot\|^2$ is convex.

(b)⇒(a) We have that $\mu f + \frac{1}{2}\|\cdot\|^2 - \frac{1}{\kappa}\frac{1}{2}\|\cdot\|^2$ is convex, i.e., $\mu f + \frac{1}{2}\|\cdot\|^2$ is strongly convex with constant $\frac{1}{\kappa}$. Then [26, Proposition 12.60] implies that $\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*$ is differentiable and its gradient is Lipschitz continuous with constant κ. In view of Lemmas 2.7(b), and 3.1(a), $\operatorname{Prox}_\mu f$ is single valued and Lipschitz continuous with constant κ. ∎

COROLLARY 4.3. *Let $0 < \mu < \lambda_f$. Then the following are equivalent.*

(a) *The proximal mapping* $\operatorname{Prox}_\mu f$ *is Lipschitz continuous with constant* 1*, i.e., nonexpansive;*

(b) *the function f is convex.*

DEFINITION 4.4 (see [25, section 26] or [26, page 483]). *A proper, lsc, convex function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is*

(a) essentially strictly convex *if f is strictly convex on every convex subset of* $\operatorname{dom} \partial f$*;*

(b) essentially differentiable *if $\partial f(x)$ is a singleton whenever $\partial f(x) \neq \varnothing$.*

PROPOSITION 4.5 (single-valued proximal mapping). *Let $0 < \mu < \lambda_f$. Then the following are equivalent.*

(a) *The proximal mapping* $\operatorname{Prox}_\mu f$ *is a singleton for every $x \in \mathbb{R}^n$;*

(b) *the function $f + \frac{1}{2\mu}\|\cdot\|^2$ is essentially strictly convex and coercive.*

Proof.

(a)⇒(b) By Lemma 3.1(a), $\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*$ is differentiable. Soloviov's theorem [28] gives us that $\mu f + \frac{1}{2}\|\cdot\|^2$ is convex. It follows from [26, Proposition 11.13] that $\mu f + \frac{1}{2}\|\cdot\|^2$ is essentially strictly convex. Since $\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*$ has full domain and $\mu f + \frac{1}{2}\|\cdot\|^2$ is convex, the function $\mu f + \frac{1}{2}\|\cdot\|^2$ is coercive by [26, Theorem 11.8].

(b)⇒(a) Since $\mu f + \frac{1}{2}\|\cdot\|^2$ is essentially strictly convex, $\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*$ is essentially differentiable by [26, Theorem 11.13]. Because $\mu f + \frac{1}{2}\|\cdot\|^2$ is coercive, $\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*$ has full domain. Then $\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*$ is differentiable on $\mathbb{R}^n$. By Lemma 3.1(a), $\operatorname{Prox}_\mu f(x)$ is single valued for every $x \in \mathbb{R}^n$. ∎

Recall that for a nonempty, closed set $S \subseteq \mathbb{R}^n$ and every $x \in \mathbb{R}^n$, the projection $P_S(x)$ consists of the points in S nearest to x, so $P_S = \operatorname{Prox}_1 \iota_S$. Combining Corollary 4.3 and Proposition 4.5, we can derive the following result due to Rockafellar and Wets [26, Corollary 12.20].

COROLLARY 4.6. *Let S be a nonempty, closed set in $\mathbb{R}^n$. Then the following are equivalent:*

(a) *P_S is single valued;*

(b) *P_S is nonexpansive;*

(c) *S is convex.*

5. The proximal average for prox-bounded functions. The goal of this section is to establish a proximal average function that works for any two prox-bounded functions. Our framework will generalize the convex proximal average of [7] to include nonconvex functions, in a manner that recovers the original definition in the convex case. Remembering the standing assumptions in subsection 1.4, we define the *proximal average* of f, g associated with parameters μ, α by

$$\varphi_\mu^\alpha = -e_\mu(-\alpha e_\mu f - (1-\alpha)e_\mu g), \tag{5.1}$$

which essentially relies on the Moreau envelopes.

THEOREM 5.1 (basic properties of the proximal average). *Let $0 < \mu < \bar\lambda$, and let φ_μ^α be defined as in* (5.1). *Then the following hold.*

(a) *The Moreau envelope $e_\mu(\varphi_\mu^\alpha) = \alpha e_\mu f + (1-\alpha)e_\mu g$.*

(b) *The proximal average φ_μ^α is proper, lsc, prox-bounded with threshold $\lambda_{\varphi_\mu^\alpha} \geq \bar\lambda$.*

(c) *The proximal average $\varphi_\mu^\alpha(x) =$*

$$\left[\alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha)\operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right](x) - \frac{1}{2\mu}\|x\|^2, \tag{5.2}$$

where the inf-convolution $\Box$ is exact; consequently,

$$\begin{aligned}&\operatorname{epi}(\varphi_\mu^\alpha + 1/2\mu\|\cdot\|^2)\\ &= \alpha \operatorname{epi}\operatorname{conv}(f + 1/2\mu\|\cdot\|^2) + (1-\alpha)\operatorname{epi}\operatorname{conv}(g + 1/2\mu\|\cdot\|^2).\end{aligned} \tag{5.3}$$

(d) *The domain $\operatorname{dom}\varphi_\mu^\alpha = \alpha \operatorname{conv}\operatorname{dom} f + (1-\alpha)\operatorname{conv}\operatorname{dom} g$. In particular, $\operatorname{dom}\varphi_\mu^\alpha = \mathbb{R}^n$ if either one of $\operatorname{conv}\operatorname{dom} f$ or $\operatorname{conv}\operatorname{dom} g$ is $\mathbb{R}^n$.*

(e) *The proximal average of f and g is the same as the proximal average of proximal hulls $h_\mu f$ and $h_\mu g$, respectively.*

(f) *When $\alpha = 0$, $\varphi_\mu^0 = h_\mu g$; when $\alpha = 1$, $\varphi_\mu^1 = h_\mu g$.*

(g) *Each φ_μ^α is μ-proximal or, equivalently, μ-hypoconvex.*

(h) *When $f = g$, $\varphi_\mu^\alpha = h_\mu f$; consequently, $\varphi_\mu^\alpha = f$ when $f = g$ is μ-proximal.*

(i) *When $g \equiv c \in \mathbb{R}$, $\varphi_\mu^\alpha = e_{\mu/\alpha,\mu}(\alpha f + (1-\alpha)c)$, the Lasry–Lions envelope of $\alpha f + (1-\alpha)c$.*

Proof.

(a) Since $-\alpha e_\mu f - (1-\alpha)e_\mu g$ is μ-proximal by Lemma 2.4(a) and Proposition 2.8, we have $-e_\mu(\varphi_\mu^\alpha) = -e_\mu(-e_\mu(-\alpha e_\mu f - (1-\alpha)e_\mu g)) = h_\mu(-\alpha e_\mu f - (1-\alpha)e_\mu g) = -\alpha e_\mu f - (1-\alpha)e_\mu g$.

(b) Because $0 < \mu < \bar\lambda$, both $e_\mu f$ and $e_\mu g$ are continuous; see, e.g., [26, Theorem 1.25]. By (a), $e_\mu(\varphi_\mu^\alpha)$ is real valued and continuous. If φ_μ^α is not proper, then $e_\mu(\varphi_\mu^\alpha) \equiv -\infty$ or $e_\mu(\varphi_\mu^\alpha) \equiv \infty$, which is a contradiction. Hence, φ_μ^α must be proper. Lower semicontinuity follows from the definition of the Moreau envelope. To show that $\lambda_{\varphi_\mu^\alpha} \geq \bar\lambda$, take any $\delta \in]0, \bar\lambda - \mu[$. By [26, Exercise 1.29(c)] and (a), we have $e_{\delta+\mu}(\varphi_\mu^\alpha) = e_\delta(e_\mu(\varphi_\mu^\alpha)) = e_\delta(\alpha e_\mu f + (1-\alpha)e_\mu g) \geq \alpha e_\delta(e_\mu f) + (1-\alpha)e_\delta(e_\mu g) = \alpha e_{\delta+\mu} f + (1-\alpha)e_{\delta+\mu} g > -\infty$. Since $\delta \in]0, \bar\lambda - \mu[$ was arbitrary, φ_μ^α has prox-bound $\lambda_{\varphi_\mu^\alpha} \geq \bar\lambda$.

(c) Since $\mu < \bar\lambda$, both $e_\mu f$ and $e_\mu g$ are locally Lipschitz with full domain by Fact 2.1(a), so $\operatorname{dom}(f + \frac{1}{2\mu}\|\cdot\|^2)^* = \operatorname{dom}(g + \frac{1}{2\mu}\|\cdot\|^2)^* = \mathbb{R}^n$. It follows from [26, Theorem 11.23(a)] that

$$\left(\alpha\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)^*\right)^* \Box \left((1-\alpha)\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)^*\right)^*,$$

where the $\Box$ is exact; see, e.g., [25, Theorem 16.4]. By Fact 2.1, $-\alpha e_\mu f - (1-\alpha)e_\mu g = \alpha(f + \frac{1}{2\mu}\|\cdot\|^2)^*(\frac{x}{\mu}) + (1-\alpha)(g + \frac{1}{2\mu}\|\cdot\|^2)^*(\frac{x}{\mu}) - \frac{1}{2\mu}\|\cdot\|^2$. Substitute this into the definition of φ_μ^α and use Fact 2.1 again to obtain $\varphi_\mu^\alpha(x) =$

$$\begin{aligned}
&\left[\alpha\left(f+\frac{1}{2\mu}\|\cdot\|^2\right)^*\left(\frac{\cdot}{\mu}\right)+(1-\alpha)\left(g+\frac{1}{2\mu}\|\cdot\|^2\right)^*\left(\frac{\cdot}{\mu}\right)\right]^*\left(\frac{x}{\mu}\right)-\frac{1}{2\mu}\|x\|^2\\
&=\left[\alpha\left(f+\frac{1}{2\mu}\|\cdot\|^2\right)^*+(1-\alpha)\left(g+\frac{1}{2\mu}\|\cdot\|^2\right)^*\right]^*\left(\mu\frac{x}{\mu}\right)-\frac{1}{2\mu}\|x\|^2\\
&=\left[\alpha\left(f+\frac{1}{2\mu}\|\cdot\|^2\right)^{**}\left(\frac{\cdot}{\alpha}\right)\Box(1-\alpha)\left(g+\frac{1}{2\mu}\|\cdot\|^2\right)^{**}\left(\frac{\cdot}{1-\alpha}\right)\right](x)-\frac{1}{2\mu}\|x\|^2
\end{aligned}$$

$$(5.4)\quad =\left[\alpha\operatorname{conv}\left(f+\frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right)\Box(1-\alpha)\operatorname{conv}\left(g+\frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right](x)-\frac{1}{2\mu}\|x\|^2,$$

where $(f+\frac{1}{2\mu}\|\cdot\|^2)^{**}=\operatorname{conv}(f+\frac{1}{2\mu}\|\cdot\|^2)$, $(g+\frac{1}{2\mu}\|\cdot\|^2)^{**}=\operatorname{conv}(g+\frac{1}{2\mu}\|\cdot\|^2)$ because $f+\frac{1}{2\mu}\|\cdot\|^2$ and $g+\frac{1}{2\mu}\|\cdot\|^2$ are coercive; see, e.g., [26, Example 11.26(c)]. Also, in (5.4), the infimal convolution is exact because $(f+\frac{1}{2\mu}\|\cdot\|^2)^*$ and $(g+\frac{1}{2\mu}\|\cdot\|^2)^*$ have full domain and [25, Theorem 16.4] or [26, Theorem 11.23(a)]. Equation (5.3) follows from (5.2) and [5, Proposition 12.8(ii)] or [26, Exercise 1.28].

(d) This is immediate from (c) and [5, Proposition 12.6(ii)].

(e) Use (5.1), and the fact that $e_\mu(h_u f)=e_\mu f$ and $e_\mu(h_u g)=e_\mu g$.

(i) This follows from $\varphi_\mu^\alpha=-e_\mu(-\alpha e_\mu f-(1-\alpha)c)=-e_\mu(-e_{\mu/\alpha}(\alpha f)-(1-\alpha)c)=-e_\mu[-e_{\mu/\alpha}(\alpha f+(1-\alpha)c)]$, and Fact 2.2(d). ∎

PROPOSITION 5.2.
(a) *The proximal average φ_μ^α is always Clarke regular, prox-regular, and strongly amenable on $\mathbb{R}^n$.*
(b) *If one of the sets* $\operatorname{conv}\operatorname{dom} f$ *or* $\operatorname{conv}\operatorname{dom} g$ *is $\mathbb{R}^n$, then φ_μ^α is locally Lipschitz on $\mathbb{R}^n$.*
(c) *When f,g are both μ-proximal, φ_μ^α is the proximal average for convex functions.*

Proof. One always has $\varphi_\mu^\alpha=(\varphi_\mu^\alpha+\frac{1}{2\mu}\|\cdot\|^2)-\frac{1}{2\mu}\|\cdot\|^2$, where $\varphi_\mu^\alpha+\frac{1}{2\mu}\|\cdot\|^2$ is convex by Theorem 5.1(g).

(a) Use [26, Example 11.30] and [26, Exercise 13.35] to conclude that φ_μ^α is prox-regular. [26, Example 10.24(g)] shows that φ_μ^α is strongly amenable. Also, being a sum of a convex function and a $\mathcal{C}^2$ function, φ_μ^α is Clarke regular.

(b) By Theorem 5.1(d), $\operatorname{dom}\varphi_\mu^\alpha=\mathbb{R}^n$, then $\varphi_\mu^\alpha+\frac{1}{2\mu}\|\cdot\|^2$ is a finite-valued convex function on $\mathbb{R}^n$, so it is locally Lipschitz, hence, φ_μ^α.

(c) Since both $f+\frac{1}{2\mu}\|\cdot\|^2$ and $g+\frac{1}{2\mu}\|\cdot\|^2$ are convex, the result follows from Theorem 5.1(c) and [6, Definition 4.1]. ∎

COROLLARY 5.3. *Let $0<\mu<\bar\lambda$ and let φ_μ^α be defined as in* (5.1). *Then*

$$-\partial_L\left[-\left(\mu\varphi_\mu^\alpha+\frac{1}{2}\|\cdot\|^2\right)^*\right]\subseteq\alpha\operatorname{Prox}_\mu f+(1-\alpha)\operatorname{Prox}_\mu g.$$

Proof. By Theorem 5.1(a), $e_\mu\varphi_\mu^\alpha=\alpha e_\mu f+(1-\alpha)e_\mu g$. Since both $e_\mu f, e_\mu g$ are locally Lipschitz, the sum rule for ∂_L [26, Corollary 10.9] gives

$$\begin{aligned}
\partial_L e_\mu\varphi_\mu^\alpha(x)&\subseteq\alpha\partial_L e_\mu f(x)+(1-\alpha)\partial_L e_\mu g(x)\\
&\subseteq\alpha\frac{x-\operatorname{Prox}_\mu f(x)}{\mu}+(1-\alpha)\frac{x-\operatorname{Prox}_\mu g(x)}{\mu}
\end{aligned}$$

from which $\partial_L \left(e_\mu \varphi_\mu^\alpha - \frac{1}{2\mu}\|x\|^2 \right) \subseteq -\frac{\alpha \operatorname{Prox}_\mu f(x) + (1-\alpha)\operatorname{Prox}_\mu g(x)}{\mu}$. As

$$e_\mu \varphi_\mu^\alpha(x) - \frac{1}{2}\|x\|^2 = -\left(\varphi_\mu^\alpha + \frac{1}{2\mu}\|\cdot\|^2 \right)^* \left(\frac{x}{\mu} \right) = -\frac{\left(\mu\varphi_\mu^\alpha + \frac{1}{2}\|\cdot\|^2\right)^*(x)}{\mu},$$

we have $-\partial_L(-(\mu\varphi_\mu^\alpha + \frac{1}{2}\|\cdot\|^2)^*)(x) \subseteq \alpha \operatorname{Prox}_\mu f(x) + (1-\alpha)\operatorname{Prox}_\mu g(x)$, as needed. ∎

A natural question to ask is whether $\alpha \operatorname{Prox}_\mu f + (1-\alpha)\operatorname{Prox}_\mu g$ is still a proximal mapping. Although this is not clear in general, we have the following.

THEOREM 5.4 (the proximal mapping of the proximal average). *Let $0 < \mu < \bar{\lambda}$ and let φ_μ^α be defined as in* (5.1). *Then*

$$\operatorname{Prox}_\mu \varphi_\mu^\alpha = \alpha \operatorname{conv} \operatorname{Prox}_\mu f + (1-\alpha) \operatorname{conv} \operatorname{Prox}_\mu g. \tag{5.5}$$

(a) *When both f and g are μ-proximal, one has*

$$\operatorname{Prox}_\mu \varphi_\mu^\alpha = \alpha \operatorname{Prox}_\mu f + (1-\alpha)\operatorname{Prox}_\mu g.$$

(b) *Suppose that on an open subset $U \subset \mathbb{R}^n$ both $\operatorname{Prox}_\mu f, \operatorname{Prox}_\mu g$ are single valued (e.g., when $e_\mu f$ and $e_\mu g$ are continuously differentiable). Then the proximal mapping $\operatorname{Prox}_\mu \varphi_\mu^\alpha$ is single valued, and*

$$\operatorname{Prox}_\mu \varphi_\mu^\alpha = \alpha \operatorname{Prox}_\mu f + (1-\alpha)\operatorname{Prox}_\mu g \text{ on } U.$$

(c) *Suppose that on an open subset $U \subset \mathbb{R}^n$, both $\operatorname{Prox}_\mu f, \operatorname{Prox}_\mu g$ are single valued and Lipschitz continuous (e.g., when f and g are prox-regular). Then $\operatorname{Prox}_\mu \varphi_\mu^\alpha$ is single valued and Lipschitz continuous and*

$$\operatorname{Prox}_\mu \varphi_\mu^\alpha = \alpha \operatorname{Prox}_\mu f + (1-\alpha)\operatorname{Prox}_\mu g \text{ on } U.$$

Proof. By Theorem 5.1, $-e_\mu(\varphi_\mu^\alpha) = -\alpha e_\mu f - (1-\alpha)e_\mu g$. Since both $-e_\mu f, -e_\mu g$ are Clarke regular, the sum rule [26, Corollary 10.9] gives

$$\partial_L(-e_\mu(\varphi_\mu^\alpha)) = \alpha \partial_L(-e_\mu f) + (1-\alpha)\partial_L(-e_\mu g).$$

Apply [26, Example 10.32] to get

$$\frac{\operatorname{conv}\operatorname{Prox}_\mu \varphi_\mu^\alpha(x) - x}{\mu} = \alpha \frac{\operatorname{conv}\operatorname{Prox}_\mu f(x) - x}{\mu} + (1-\alpha)\frac{\operatorname{conv}\operatorname{Prox}_\mu g(x) - x}{\mu}$$

from which $\operatorname{conv}\operatorname{Prox}_\mu \varphi_\mu^\alpha = \alpha \operatorname{conv}\operatorname{Prox}_\mu f + (1-\alpha)\operatorname{conv}\operatorname{Prox}_\mu g$. Since φ_μ^α is μ-proximal, $\operatorname{conv}\operatorname{Prox}_\mu \varphi_\mu^\alpha = \operatorname{Prox}_\mu \varphi_\mu^\alpha$, therefore, (5.5) follows.

(a) Since f, g are μ-proximal, $\operatorname{Prox}_\mu f$ and $\operatorname{Prox}_\mu g$ are convex by Proposition 2.6.

(b) When $e_\mu f$ and $e_\mu g$ are continuously differentiable, both $\operatorname{Prox}_\mu f, \operatorname{Prox}_\mu g$ are single valued on U by [10, Proposition 5.1].

(c) When f and g are prox-regular on U, both $\operatorname{Prox}_\mu f, \operatorname{Prox}_\mu g$ are single valued and Lipschitz continuous on U by [10, Proposition 5.3] or [26, Proposition 13.37]. ∎

COROLLARY 5.5. *Let $0 < \mu < \bar{\lambda}$ and let φ_μ^α be defined as in* (5.1). *Then*

$$\operatorname{Prox}_\mu \varphi_\mu^\alpha = \alpha \operatorname{Prox}_\mu(h_\mu f) + (1-\alpha)\operatorname{Prox}_\mu(h_\mu g).$$

Proof. Combine Theorem 5.4 and Lemma 2.9. ∎

COROLLARY 5.6. *Let $\mu > 0$. The following set of proximal mappings*

$$\{\operatorname{Prox}_\mu f |\ f \text{ is } \mu\text{-proximal and } \mu < \lambda_f\}$$

is a convex set. Moreover, for every μ-proximal function, $\operatorname{Prox}_\mu f = (\operatorname{Id} + \mu \partial_L f)^{-1}$.

Proof. Apply Theorem 5.4(a), Theorem 5.1(b) and (g), and Proposition 2.6. ∎

6. Relationships to the arithmetic average and epi-average. We show in this section the connections of the proximal average with the arithmetic average and epi-average, and full epi-continuity of the proximal average.

DEFINITION 6.1 (epi-convergence and epi-topology). (See [26, Chapter 6].) *Let f and $(f_k)_{k\in\mathbb{N}}$ be functions from $\mathbb{R}^n$ to $]-\infty,+\infty]$. Then $(f_k)_{k\in\mathbb{N}}$* epi-converges *to f; in symbols $f_k \xrightarrow{e} f$ if for every $x\in\mathbb{R}^n$ the following hold:*
(a) $\big(\forall\,(x_k)_{k\in\mathbb{N}}\big)$ $x_k\to x \Rightarrow f(x)\le \liminf f_k(x_k)$;
(b) $\big(\exists (y_k)_{k\in\mathbb{N}}\big)$ $y_k\to x$ *and* $\limsup f_k(y_k)\le f(x)$.
We write e-$\lim_{k\to\infty} f_k = f$. *The* epi-topology *is induced by the epi-distance metric defined on epigraphs of all lsc functions, and it characterizes the epi-convergence; see* [26, Chapter 7, Theorem 7.58].

Let us remark that the threshold $\bar\lambda=+\infty$ whenever both f,g are bounded from below by an affine function.

THEOREM 6.2. *Let $0<\mu<\bar\lambda$. One has the following.*
(a) *For every fixed $x\in\mathbb{R}^n$, the function $\mu\mapsto\varphi_\mu^\alpha(x)$ is monotonically decreasing and left continuous on $]0,\bar\lambda]$.*
(b) *The pointwise limit* $\lim_{\mu\uparrow\bar\lambda}\varphi_\mu^\alpha=\inf_{\bar\lambda>\mu>0}\varphi_\mu^\alpha=$

$$\alpha\,\mathrm{conv}\left(f+\frac{1}{2\,\bar\lambda}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right)\Box(1-\alpha)\,\mathrm{conv}\left(g+\frac{1}{2\,\bar\lambda}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)-\frac{1}{2\,\bar\lambda}\|\cdot\|^2.$$

(c) *When $\bar\lambda=\infty$, the pointwise limit is*

$$\lim_{\mu\uparrow\infty}\varphi_\mu^\alpha=\inf_{\mu>0}\varphi_\mu^\alpha=\alpha\,\mathrm{conv}\,f\left(\frac{\cdot}{\alpha}\right)\Box(1-\alpha)\,\mathrm{conv}\,g\left(\frac{\cdot}{1-\alpha}\right), \tag{6.1}$$

and the epigraphical limit is

$$\operatorname*{e-lim}_{\mu\uparrow\infty}\varphi_\mu^\alpha=\mathrm{cl}\left[\alpha\,\mathrm{conv}\,f\left(\frac{\cdot}{\alpha}\right)\Box(1-\alpha)\,\mathrm{conv}\,g\left(\frac{\cdot}{1-\alpha}\right)\right]. \tag{6.2}$$

Proof. (a) We have $\varphi_\mu^\alpha(x)=$

$$\inf_{u+v=x}\left[\alpha\inf_{\substack{\sum_i\alpha_i x_i=\frac{u}{\alpha}\\ \sum_i\alpha_i=1,\alpha_i\ge0}}\sum_i\alpha_i\left(f(x_i)+\frac{1}{2\mu}\|x_i\|^2\right)+(1-\alpha)\inf_{\substack{\sum_j\beta_j y_j=\frac{v}{1-\alpha}\\ \sum_j\beta_j=1,\beta_j\ge0}}\sum_j\beta_j\left(g(y_j)+\frac{1}{2\mu}\|y_j\|^2\right)\right]$$
$$-\frac{1}{2\mu}\|x\|^2$$
$$=\inf_{\substack{\alpha\sum_i\alpha_i x_i+(1-\alpha)\sum_j\beta_j y_j=x\\ \sum_i\alpha_i=1,\sum_j\beta_j=1,\alpha_i\ge0,\beta_j\ge0}}\left[\alpha\sum_i\alpha_i f(x_i)+(1-\alpha)\sum_j\beta_j g(y_j)\right.$$
$$\left.+\frac{1}{2\mu}\underbrace{\left(\alpha\sum_i\alpha_i\|x_i\|^2+(1-\alpha)\sum_j\beta_j\|y_j\|^2-\left\|\alpha\sum_i\alpha_i x_i+(1-\alpha)\sum_j\beta_j y_j\right\|^2\right)}\right].$$

The underbraced part is nonnegative because $\|\cdot\|^2$ is convex, $\sum_i\alpha_i=1,\sum_j\beta_j=1$. It follows that $\mu\mapsto\varphi_\mu^\alpha$ is a monotonically decreasing function on $]0,+\infty[$. Let $\bar\mu\in]0,\bar\lambda]$. Then $\lim_{\mu\uparrow\bar\mu}\varphi_\mu^\alpha(x)=\inf_{\bar\mu>\mu>0}\varphi_\mu^\alpha(x)=$

$$\begin{aligned}
&\inf_{\bar\mu>\mu>0}\inf_{\substack{\alpha\sum_i\alpha_ix_i+(1-\alpha)\sum_j\beta_jy_j=x\\ \sum_i\alpha_i=1,\sum_j\beta_j=1,\alpha_i\ge0,\beta_j\ge0}}\Bigg[\alpha\sum_i\alpha_if(x_i)+(1-\alpha)\sum_j\beta_jg(y_j)\\
&\quad+\frac{1}{2\mu}\left(\alpha\sum_i\alpha_i\|x_i\|^2+(1-\alpha)\sum_j\beta_j\|y_j\|^2-\left\|\alpha\sum_i\alpha_ix_i+(1-\alpha)\sum_j\beta_jy_j\right\|^2\right)\Bigg]
\end{aligned}\tag{6.3}$$

$$\begin{aligned}
&=\inf_{\substack{\alpha\sum_i\alpha_ix_i+(1-\alpha)\sum_j\beta_jy_j=x\\ \sum_i\alpha_i=1,\sum_j\beta_j=1,\alpha_i\ge0,\beta_j\ge0}}\inf_{\bar\mu>\mu>0}\Bigg[\alpha\sum_i\alpha_if(x_i)+(1-\alpha)\sum_j\beta_jg(y_j)\\
&\quad+\frac{1}{2\mu}\left(\alpha\sum_i\alpha_i\|x_i\|^2+(1-\alpha)\sum_j\beta_j\|y_j\|^2-\left\|\alpha\sum_i\alpha_ix_i+(1-\alpha)\sum_j\beta_jy_j\right\|^2\right)\Bigg]
\end{aligned}\tag{6.4}$$

$$\begin{aligned}
&=\inf_{\substack{\alpha\sum_i\alpha_ix_i+(1-\alpha)\sum_j\beta_jy_j=x\\ \sum_i\alpha_i=1,\sum_j\beta_j=1,\alpha_i\ge0,\beta_j\ge0}}\Bigg[\alpha\sum_i\alpha_if(x_i)+(1-\alpha)\sum_j\beta_jg(y_j)\\
&\quad+\frac{1}{2\,\bar\mu}\left(\alpha\sum_i\alpha_i\|x_i\|^2+(1-\alpha)\sum_j\beta_j\|y_j\|^2-\left\|\alpha\sum_i\alpha_ix_i+(1-\alpha)\sum_j\beta_jy_j\right\|^2\right)\Bigg]
\end{aligned}\tag{6.5}$$

$$=\left[\alpha\,\mathrm{conv}\left(f+\frac{1}{2\,\bar\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right)\Box(1-\alpha)\,\mathrm{conv}\left(g+\frac{1}{2\,\bar\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right](x)-\frac{1}{2\,\bar\mu}\|x\|^2.$$

(b) This follows from (a).

(c) By (a), we have $\lim_{\mu\to\infty}\varphi_\mu^\alpha=\inf_{\mu>0}\varphi_\mu^\alpha$. Using similar arguments as (6.3)–(6.5), we obtain $\inf_{\mu>0}\varphi_\mu^\alpha(x)=$

$$\begin{aligned}
&\inf_{\substack{\alpha\sum_i\alpha_ix_i+(1-\alpha)\sum_j\beta_jy_j=x\\ \sum_i\alpha_i=1,\sum_j\beta_j=1,\alpha_i\ge0,\beta_j\ge0}}\left(\alpha\sum_i\alpha_if(x_i)+(1-\alpha)\sum_j\beta_jg(y_j)\right)\\
&=\inf_{u+v=x}\left(\alpha\inf_{\substack{\sum_i\alpha_ix_i=u/\alpha\\ \sum_i\alpha_i=1,\alpha_i\ge0}}\sum_i\alpha_if(x_i)+(1-\alpha)\inf_{\substack{\sum_j\beta_jy_j=v/(1-\alpha)\\ \sum_j\beta_j=1,\beta_j\ge0}}\sum_j\beta_jg(y_j)\right)\\
&=\inf_{u+v=x}\Big(\alpha(\mathrm{conv}\,f)(u/\alpha)+(1-\alpha)(\mathrm{conv}\,g)(v/(1-\alpha))\Big),
\end{aligned}$$

as required. To get (6.2), we combine (6.1) and [26, Proposition 7.4(c)]. ∎

In order to study the limit behavior when $\mu\downarrow0$, the following lemma is useful. We omit its simple proof.

LEMMA 6.3. *The Moreau envelope function respects the inequality*

$$e_\mu(\alpha f_1+(1-\alpha)f_2)\ge\alpha e_\mu f_1+(1-\alpha)e_\mu f_2.$$

THEOREM 6.4. *Let $0<\mu<\bar\lambda$. One has*
(a) $\alpha e_\mu f+(1-\alpha)e_\mu g\le\varphi_\mu^\alpha\le\alpha h_\mu f+(1-\alpha)h_\mu g\le\alpha f+(1-\alpha)g$, *and*
(b) *when $\mu\downarrow0$, the pointwise limit and epigraphical limit agree with*

$$\lim_{\mu\downarrow0}\varphi_\mu^\alpha=\sup_{\mu>0}\varphi_\mu^\alpha=\alpha f+(1-\alpha)g.\tag{6.6}$$

Furthermore, the convergence in (6.6) is uniform on compact subsets of $\mathbb{R}^n$ when f,g are continuous.

Proof. Apply Lemma 6.3 with $f_1 = -e_\mu f, f_2 = -e_\mu g$ to obtain

$$e_\mu(\alpha(-e_\mu f) + (1-\alpha)(-e_\mu g)) \geq \alpha e_\mu(-e_\mu f) + (1-\alpha)e_\mu(-e_\mu g).$$

Then

$$\varphi_\mu^\alpha \leq \alpha(-e_\mu(-e_\mu f)) + (1-\alpha)(-e_\mu(-e_\mu g)) = \alpha h_\mu f + (1-\alpha)h_\mu g. \tag{6.7}$$

On the other hand, $e_\mu(\alpha(-e_\mu f) + (1-\alpha)(-e_\mu g)) \leq \alpha(-e_\mu f) + (1-\alpha)(-e_\mu g)$, so

$$\varphi_\mu^\alpha \geq \alpha e_\mu f + (1-\alpha)e_\mu g. \tag{6.8}$$

Combining (6.7) and (6.8) gives (a). Equation (6.6) follows from (a) by sending $\mu \downarrow 0$. The pointwise and epigraphical limits agree because of [26, Proposition 7.4(d)]. Now assume that f, g are continuous. Since both $e_\mu f$ and f are continuous, and $e_\mu f \uparrow f$. Dini's theorem says that $e_\mu f \uparrow f$ uniformly on compact subsets of $\mathbb{R}^n$. The same can be said about $e_\mu g \uparrow g$. Hence, the convergence in (6.6) is uniform on compact subsets of $\mathbb{R}^n$ by (a). ∎

To study the epi-continuity of the proximal average, we recall the following two standard notions.

Definition 6.5. *A sequence of functions $(f_k)_{k\in\mathbb{N}}$ is eventually prox-bounded if there exists $\lambda > 0$ such that $\liminf_{k\to\infty} e_\lambda f_k(x) > -\infty$ for some x. The supremum of all such λ is then the threshold of eventual prox-boundedness of the sequence.*

Definition 6.6. *A sequence of functions $(f_k)_{k\in\mathbb{N}}$ converges continuously to f if $f_k(x_k) \to f(x)$ whenever $x_k \to x$.*

The following key result is implicit in the proof of [26, Theorem 7.37]. We provide its proof for completeness. Define $\mathcal{N}_\infty = \{N \subset \mathbb{N} \mid \mathbb{N}\backslash N \text{ is finite}\}$.

Lemma 6.7. *Let $(f_k)_{k\in\mathbb{N}}$ and f be proper, lsc functions on $\mathbb{R}^n$. Suppose that $(f_k)_{k\in\mathbb{N}}$ is eventually prox-bounded, $\bar{\lambda}$ is the threshhold of eventual prox-boundedness, and $f_k \xrightarrow{e} f$. Suppose also that $\mu_k, \mu \in]0, \bar{\lambda}[$, and $\mu_k \to \mu$. Then f is prox-bounded with threshold $\lambda_f \geq \bar{\lambda}$ and $e_{\mu_k} f_k$ converges continuously to $e_\mu f$. In particular, $e_{\mu_k} f_k \xrightarrow{e} e_\mu f$, and $e_{\mu_k} f_k \xrightarrow{p} e_\mu f$.*

Proof. Let $\varepsilon \in]0, \bar{\lambda}[$. The eventual prox-boundness of $(f_k)_{k\in\mathbb{N}}$ means that there exist $b \in \mathbb{R}^n$, $\beta \in \mathbb{R}$, and $N \in \mathcal{N}_\infty$ such that $(\forall k \in N)(\forall w \in \mathbb{R}^n)$ $f_k(w) \geq \beta - \frac{1}{2\varepsilon}\|b - w\|^2$. Let $\mu \in]0, \varepsilon[$. Consider any $x \in \mathbb{R}^n$ and any sequence $x_k \to x$ in $\mathbb{R}^n$, any sequence $\mu_k \to \mu$ in $(0, \bar{\lambda})$. Since $f_k \xrightarrow{e} f$, the functions $f_k + (1/2\mu_k)\| \cdot - x_k\|^2$ epi-converge to $f + (1/2\mu)\| \cdot - x\|^2$. Take $\delta \in]\mu, \varepsilon[$. Because $\mu_k \to \mu$, there exists $N' \subseteq N$, $N' \in \mathcal{N}_\infty$ such that $\mu_k \in (0, \delta)$ when $k \in N'$. Then $\forall k \in N'$,

$$\begin{aligned} f_k(w) + \frac{1}{2\mu_k}\|x_k - w\|^2 &\geq \beta - \frac{1}{2\varepsilon}\|b - w\|^2 + \frac{1}{2\delta}\|x_k - w\|^2 \\ &\geq \beta + \left(\frac{1}{2\delta} - \frac{1}{2\varepsilon}\right)\|b - w\|^2 - \frac{1}{\delta}\|x_k - b\|\|b - w\|. \end{aligned}$$

In view of $x_k \to x$, the sequence $(\|x_k - b\|)_{k\in\mathbb{N}}$ is bounded, say by $\rho > 0$. We have

$$(\forall k \in N')\ f_k(w) + \frac{1}{2\mu_k}\|x_k - w\|^2 \geq h(w) := \beta + \left(\frac{1}{2\delta} - \frac{1}{2\varepsilon}\right)\|b - w\|^2 - \frac{\rho}{\delta}\|b - w\|.$$

The function h is level bounded because $\delta < \varepsilon$. Hence, by [26, Theorem 7.33],

$$\lim_{k\to\infty} \inf_w \left(f_k(w) + \frac{1}{2\mu_k}\|x_k - w\|^2 \right) = \inf_w \left(f(w) + \frac{1}{2\mu}\|x - w\|^2 \right),$$

i.e., $e_{\mu_k} f_k(x_k) \to e_\mu f(x)$. Also, $e_\mu f(x)$ is finite, so $\lambda_f \geq \mu$. Since $\varepsilon \in]0, \bar{\lambda}[$ and $\mu \in]0, \varepsilon[$ were arbitrary, the result holds whenever $\mu \in]0, \bar{\lambda}[$. This in turn implies $\lambda_f \geq \bar{\lambda}$. ∎

For the convenience of analyzing the full epi-continuity, below we write the proximal average φ_μ^α explicitly in the form $\varphi_{f,g,\alpha,\mu}$.

Theorem 6.8 (full epi-continuity of the proximal average). *Let the sequences of functions $(f_k)_{k\in\mathbb{N}}$, $(g_k)_{k\in\mathbb{N}}$ on $\mathbb{R}^n$ be eventually prox-bounded with threshold of eventual prox-boundedness $\bar{\lambda} > 0$. Let $(\mu_k)_{k\in\mathbb{N}}$ be a sequence and μ in $]0, \bar{\lambda}[$ and let $(\alpha_k)_{k\in\mathbb{N}}$ be a sequence and α in $[0,1]$. Suppose that $f_k \xrightarrow{\text{e}} f$, $g_k \xrightarrow{\text{e}} g$, $\mu_k \to \mu$, and $\alpha_k \to \alpha$. Then $\varphi_{f_k,g_k,\alpha_k,\mu_k} \xrightarrow{\text{e}} \varphi_{f,g,\alpha,\mu}$.*

Proof. Consider any $x \in \mathbb{R}^n$ and any sequence $x_k \to x$. By [26, Example 11.26],

$$e_{\mu_k} f_k(\mu_k x_k) = \frac{\mu_k \|x_k\|^2}{2} - \left(f_k + \frac{1}{2\mu_k}\|\cdot\|^2 \right)^* (x_k).$$

Lemma 6.7 shows that $\lim_{k\to\infty} (f_k + \frac{1}{2\mu_k}\|\cdot\|^2)^*(x_k) =$

$$\lim_{k\to\infty} \frac{\mu_k\|x_k\|^2}{2} - e_{\mu_k} f_k(\mu_k x_k) = \frac{\mu\|x\|^2}{2} - e_\mu f(\mu x) = \left(f + \frac{1}{2\mu}\|\cdot\|^2 \right)^* (x).$$

Therefore, the functions $(f_k + \frac{1}{2\mu_k}\|\cdot\|^2)^*$ converge continuously to $(f + \frac{1}{2\mu}\|\cdot\|^2)^*$. It follows that $\alpha_k(f_k + \frac{1}{2\mu_k}\|\cdot\|^2)^* + (1-\alpha_k)(g_k + \frac{1}{2\mu_k}\|\cdot\|^2)^*$ converges continuously to $\alpha(f + \frac{1}{2\mu}\|\cdot\|^2)^* + (1-\alpha)(g + \frac{1}{2\mu}\|\cdot\|^2)^*$, so epi-converges. Wijsman's theorem [26, Theorem 11.34] implies that $[\alpha_k(f_k + \frac{1}{2\mu_k}\|\cdot\|^2)^* + (1-\alpha_k)(g_k + \frac{1}{2\mu_k}\|\cdot\|^2)^*]^*$ epi-converges to $[\alpha(f + \frac{1}{2\mu}\|\cdot\|^2)^* + (1-\alpha)(g + \frac{1}{2\mu}\|\cdot\|^2)^*]^*$. As $(\mu, x) \mapsto \frac{1}{2\mu}\|x\|^2$ is continuous on $]0, +\infty[\times \mathbb{R}^n$, we have that

$$\varphi_{f_k,g_k,\alpha_k,\mu_k} = \left[\alpha_k \left(f_k + \frac{1}{2\mu_k}\|\cdot\|^2 \right)^* + (1-\alpha_k)\left(g_k + \frac{1}{2\mu_k}\|\cdot\|^2 \right)^* \right]^* - \frac{1}{2\mu_k}\|\cdot\|^2$$

epi-converges to $\left[\alpha\left(f + \frac{1}{2\mu}\|\cdot\|^2 \right)^* + (1-\alpha)\left(g + \frac{1}{2\mu}\|\cdot\|^2 \right)^* \right]^* - \frac{1}{2\mu}\|\cdot\|^2.$ ∎

Corollary 6.9 (epi-continuity of the proximal average). *Let $0 < \mu < \bar{\lambda}$. Then the function $\alpha \mapsto \varphi_\mu^\alpha$ is continuous with respect to the epi-topology. That is, $\forall (\alpha_k)_{k\in\mathbb{N}}$ and α in $[0,1]$, one has $\alpha_k \to \alpha \Rightarrow \varphi_\mu^{\alpha_k} \xrightarrow{\text{e}} \varphi_\mu^\alpha$. In particular, $\varphi_\mu^\alpha \xrightarrow{\text{e}} h_\mu g$ when $\alpha \downarrow 0$, and $\varphi_\mu^\alpha \xrightarrow{\text{e}} h_\mu f$ when $\alpha \uparrow 1$.*

7. Optimal value and minimizers of the proximal average.

In this section, using results of section 6 we investigate optimization properties of the proximal average.

7.1. Relationship of infimum and minimizers among φ_μ^α, f, and g.

Proposition 7.1. *Let $0 < \mu < \bar{\lambda}$. One has*
(a) $\inf \varphi_\mu^\alpha = \inf[\alpha e_\mu f + (1-\alpha) e_\mu g]$ and $\operatorname{argmin} \varphi_\mu^\alpha = \operatorname{argmin}[\alpha e_\mu f + (1-\alpha) e_\mu g]$;
(b) $\alpha \inf f + (1-\alpha) \inf g \leq \inf \varphi_\mu^\alpha \leq \inf[\alpha h_\mu f + (1-\alpha) h_\mu g] \leq \inf[\alpha f + (1-\alpha) g]$.

Proof. For (a), apply Theorem 5.1(a) and $\operatorname{argmin}\varphi_\mu^\alpha = \operatorname{argmin} e_\mu\varphi_\mu^\alpha$. For (b), apply Theorem 6.4(a) and $\inf e_\mu f = \inf f$, and $\inf e_\mu g = \inf g$. ∎

THEOREM 7.2. *Suppose that* $\operatorname{argmin} f \cap \operatorname{argmin} g \neq \varnothing$ *and* $\alpha \in]0,1[$. *Then the following hold:*

(a) $\min(\alpha f + (1-\alpha)g) = \alpha \min f + (1-\alpha)\min g$ *and* $\operatorname{argmin}(\alpha f + (1-\alpha)g) = \operatorname{argmin} f \cap \operatorname{argmin} g$;

(b)
$$\begin{aligned} \min \varphi_\mu^\alpha &= \alpha \min f + (1-\alpha)\min g, \text{ and} \\ \operatorname{argmin}\varphi_\mu^\alpha &= \operatorname{argmin} f \cap \operatorname{argmin} g. \end{aligned} \tag{7.1}$$

Proof.

(a) This is simple to verify.

(b) Equation (7.1) follows from Proposition 7.1, Theorem 6.4(a) and (a). This also gives $(\operatorname{argmin} f \cap \operatorname{argmin} g) \subseteq \operatorname{argmin}\varphi_\mu^\alpha$. To show $(\operatorname{argmin} f \cap \operatorname{argmin} g) \supseteq \operatorname{argmin}\varphi_\mu^\alpha$, take any $x \in \operatorname{argmin}\varphi_\mu^\alpha$. By (7.1) and Theorem 6.4(a), we have

$$\alpha \min f + (1-\alpha)\min g = \varphi_\mu^\alpha(x) \geq \alpha e_\mu f(x) + (1-\alpha)e_\mu g(x)$$

from which $\alpha(e_\mu f(x) - \min f) + (1-\alpha)(e_\mu g(x) - \min g) \leq 0$. Since $\min f = \min e_\mu f$ and $\min g = \min e_\mu g$, it follows that $e_\mu f(x) = \min e_\mu f$ and $e_\mu g(x) = \min e_\mu g$, so $x \in (\operatorname{argmin} e_\mu f \cap \operatorname{argmin} e_\mu g) = (\operatorname{argmin} f \cap \operatorname{argmin} g)$ because of $\operatorname{argmin} e_\mu f = \operatorname{argmin} f$ and $\operatorname{argmin} e_\mu g = \operatorname{argmin} g$. ∎

To explore further optimization properties of φ_μ^α, we need the following two auxiliary results.

LEMMA 7.3. *Suppose that* $f_1, f_2 : \mathbb{R}^n \to]-\infty, +\infty]$ *are proper and lsc, and that* $f_1 \Box f_2$ *is exact. Then*

(a) $\inf(f_1 \Box f_2) = \inf f_1 + \inf f_2$ *and*

(b) $\operatorname{argmin}(f_1 \Box f_2) = \operatorname{argmin} f_1 + \operatorname{argmin} f_2$.

Proof. Use the fact that the epigraph of $f_1 \Box f_2$ is the sum of epigraphs of f_1 and f_2. ∎

LEMMA 7.4. *Let* $f_1 : \mathbb{R}^n \to]-\infty, +\infty]$ *be proper and lsc. Then*

(a) $\inf(\operatorname{conv} f_1) = \inf f_1$;

(b) *if, in addition,* f_1 *is coercive, then* $\operatorname{argmin}(\operatorname{conv} f_1) = \operatorname{conv}(\operatorname{argmin} f_1)$, *and* $\operatorname{argmin}(\operatorname{conv} f_1) \neq \varnothing$.

Proof. Combine [9, Comment 3.7(4)] and [26, Corollary 3.47]. ∎

We are now ready for the main result of this section.

THEOREM 7.5. *Let* $0 < \mu < \bar{\lambda}$, *and let* φ_μ^α *be defined as in* (5.1). *Then the following hold:*

(a) $\inf\left(\varphi_\mu^\alpha + \frac{1}{2\mu}\|\cdot\|^2\right) = \alpha \inf\left(f + \frac{1}{2\mu}\|\cdot\|^2\right) + (1-\alpha)\inf\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)$;

(b) $\operatorname{argmin}\left(\varphi_\mu^\alpha + \frac{1}{2\mu}\|\cdot\|^2\right) =$
$$\alpha \operatorname{conv}\left[\operatorname{argmin}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\right] + (1-\alpha)\operatorname{conv}\left[\operatorname{argmin}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\right] \neq \varnothing.$$

Proof. Theorem 5.1(c) gives $\varphi_\mu^\alpha + \frac{1}{2\mu}\|\cdot\|^2 =$

$$\left[\alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right)\right] \Box \left[(1-\alpha)\operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right],$$

in which the inf-convolution $\Box$ is exact.

(a) Using Lemmas 7.3(a) and 7.4(a), we deduce $\inf\left(\varphi_\mu^\alpha + \frac{1}{2\mu}\|\cdot\|^2\right) =$

$$\begin{aligned}
&\inf\left[\alpha \operatorname{conv}\left(f+\frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right)\right] + \inf\left[(1-\alpha)\operatorname{conv}\left(g+\frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right]\\
&= \alpha \inf\left[\operatorname{conv}\left(f+\frac{1}{2\mu}\|\cdot\|^2\right)\right] + (1-\alpha)\inf\left[\operatorname{conv}\left(g+\frac{1}{2\mu}\|\cdot\|^2\right)\right]\\
&= \alpha \inf\left(f+\frac{1}{2\mu}\|\cdot\|^2\right) + (1-\alpha)\inf\left(g+\frac{1}{2\mu}\|\cdot\|^2\right).
\end{aligned}$$

(b) Note that $f + \frac{1}{2\mu}\|\cdot\|^2$ and $g + \frac{1}{2\mu}\|\cdot\|^2$ are coercive because of $0 < \mu < \bar{\lambda}$. Using Lemmas 7.3(b) and 7.4(b), we deduce $\operatorname{argmin}\left(\varphi_\mu^\alpha + \frac{1}{2\mu}\|\cdot\|^2\right) =$

$$\begin{aligned}
&\operatorname{argmin}\left[\alpha \operatorname{conv}\left(f+\frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right)\right] + \operatorname{argmin}\left[(1-\alpha)\operatorname{conv}\left(g+\frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right]\\
&= \alpha \operatorname{argmin}\left[\operatorname{conv}\left(f+\frac{1}{2\mu}\|\cdot\|^2\right)\right] + (1-\alpha)\operatorname{argmin}\left[\operatorname{conv}\left(g+\frac{1}{2\mu}\|\cdot\|^2\right)\right]\\
&= \alpha \operatorname{conv}\left[\operatorname{argmin}\left(f+\frac{1}{2\mu}\|\cdot\|^2\right)\right] + (1-\alpha)\operatorname{conv}\left[\operatorname{argmin}\left(g+\frac{1}{2\mu}\|\cdot\|^2\right)\right].
\end{aligned}$$

Finally, these three sets of minimizers are nonempty by Lemma 7.4(b). ∎

In view of Theorem 6.2(c), when $\bar{\lambda} = \infty$, as $\mu \to \infty$ the pointwise limit is

$$\varphi_\mu^\alpha \xrightarrow{\text{p}} \left[\alpha \operatorname{conv} f\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha)\operatorname{conv} g\left(\frac{\cdot}{1-\alpha}\right)\right],$$

and the epi-limit is $\varphi_\mu^\alpha \xrightarrow{\text{e}} \operatorname{cl}[\alpha \operatorname{conv} f(\frac{\cdot}{\alpha}) \Box (1-\alpha)\operatorname{conv} g(\frac{\cdot}{1-\alpha})]$. We conclude this section with a result on minimization of this limit.

PROPOSITION 7.6. *Suppose that both f and g are coercive. Then the following hold:*

(a) $\alpha \operatorname{conv} f(\frac{\cdot}{\alpha}) \Box (1-\alpha)\operatorname{conv} g(\frac{\cdot}{1-\alpha})$ *is proper, lsc, and convex;*
(b) $\min[\alpha \operatorname{conv} f(\frac{\cdot}{\alpha}) \Box (1-\alpha)\operatorname{conv} g(\frac{\cdot}{1-\alpha})] = \alpha \min f + (1-\alpha)\min g$;
(c) $\operatorname{argmin}[\alpha \operatorname{conv} f(\frac{\cdot}{\alpha}) \Box (1-\alpha)\operatorname{conv} g(\frac{\cdot}{1-\alpha})] =$

$$\alpha \operatorname{conv} \operatorname{argmin} f + (1-\alpha)\operatorname{conv}\operatorname{argmin} g \neq \varnothing.$$

Proof. Since both f and g are coercive, by [26, Corollary 3.47], $\operatorname{conv} f$ and $\operatorname{conv} g$ are lsc, convex, and coercive. As

$$(\alpha f^* + (1-\alpha)g^*)^* = \operatorname{cl}\left[\alpha \operatorname{conv} f\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha)\operatorname{conv} g\left(\frac{\cdot}{1-\alpha}\right)\right]$$

and $\operatorname{dom} f^* = \mathbb{R}^n = \operatorname{dom} g^*$, the closure operation on the right-hand side is superfluous. This establishes (a). Moreover, the infimal convolution

$$\alpha \operatorname{conv} f\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha)\operatorname{conv} g\left(\frac{\cdot}{1-\alpha}\right) \tag{7.2}$$

is exact. For (b), (c), it suffices to apply Lemma 7.3 to (7.2) for functions $\alpha \operatorname{conv} f\left(\frac{\cdot}{\alpha}\right)$ and $\alpha \operatorname{conv} g\left(\frac{\cdot}{\alpha}\right)$, followed by invoking Lemma 7.4. ∎

7.2. Convergence in minimization. We need the following result on coercivity.

LEMMA 7.7. *Let* $0 < \mu < \bar{\lambda}$ *and let* $\psi : \mathbb{R}^n \to \mathbb{R}$ *be a convex function. If* $f \geq \psi, g \geq \psi$, *then* $\varphi_\mu^\alpha \geq \psi$.

Proof. Recall $\varphi_\mu^\alpha(x) =$

$$\left[\alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha) \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right](x) - \frac{1}{2\mu}\|x\|^2.$$

As $f + \frac{1}{2\mu}\|\cdot\|^2 \geq \psi + \frac{1}{2\mu}\|\cdot\|^2$ and the latter is convex, we have $\operatorname{conv}(f + \frac{1}{2\mu}\|\cdot\|^2) \geq \psi + \frac{1}{2\mu}\|\cdot\|^2$; similarly, $\operatorname{conv}(g + \frac{1}{2\mu}\|\cdot\|^2) \geq \psi + \frac{1}{2\mu}\|\cdot\|^2$. Then

$$\begin{aligned}&\alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha) \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\\ &\geq \alpha\left(\psi + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha)\left(\psi + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right) = \psi + \frac{1}{2\mu}\|\cdot\|^2\end{aligned}$$

in which we have used the coercivity and convexity of $\psi + \frac{1}{2\mu}\|\cdot\|^2$. The result follows. ∎

THEOREM 7.8. *Let* $0 < \mu < \bar{\lambda}$. *One has the following.*

(a) *If* f, g *are bounded from below, then* φ_μ^α *is bounded from below.*

(b) *If* f, g *are level coercive, then* φ_μ^α *is level coercive.*

(c) *If* f, g *are coercive, then* φ_μ^α *is coercive.*

Proof.

(a) Put $\psi = \min\{\inf f, \inf g\}$ and apply Lemma 7.7.

(b) By [26, Theorem 3.26(a)], there exist $\gamma \in (0, \infty)$, and $\beta \in \mathbb{R}$ such that $f \geq \psi, g \geq \psi$ with $\psi = \gamma\|\cdot\| + \beta$. Apply Lemma 7.7.

(c) By [26, Theorem 3.26(b)], for every $\gamma \in (0, \infty)$, there exists $\beta \in \mathbb{R}$ such that $f \geq \psi, g \geq \psi$ with $\psi = \gamma\|\cdot\| + \beta$. Apply Lemma 7.7. ∎

THEOREM 7.9. *Suppose that the proper, lsc functions* f, g *are level coercive. Then for every* $\bar{\alpha} \in [0,1]$, *we have* $\lim_{\alpha \to \bar{\alpha}} \inf \varphi_\mu^\alpha = \inf \varphi_\mu^{\bar{\alpha}}$ *(finite), and*

$$\limsup_{\alpha \to \bar{\alpha}} \operatorname{argmin} \varphi_\mu^\alpha \subseteq \operatorname{argmin} \varphi_\mu^{\bar{\alpha}}.$$

Moreover, $(\operatorname{argmin} \varphi_\mu^\alpha)_{\alpha \in [0,1]}$ *lies in a bounded set. Consequently,*

$$\lim_{\alpha \downarrow 0} \inf \varphi_\mu^\alpha = \inf g \text{ and } \limsup_{\alpha \downarrow 0} \operatorname{argmin} \varphi_\mu^\alpha \subseteq \operatorname{argmin} g;$$

$$\lim_{\alpha \uparrow 1} \inf \varphi_\mu^\alpha = \inf f \text{ and } \limsup_{\alpha \uparrow 1} \operatorname{argmin} \varphi_\mu^\alpha \subseteq \operatorname{argmin} f.$$

Proof. By assumption, there exist $\gamma > 0$ and $\beta \in \mathbb{R}$ such that $f \geq \gamma\|\cdot\| + \beta, g \geq \gamma\|\cdot\| + \beta$. By Lemma 7.7, we have that $\varphi_\mu^\alpha \geq \gamma\|\cdot\| + \beta$ for every $\alpha \in [0,1]$. Since $\gamma\|\cdot\| + \beta$ is level bounded, $(\varphi_\mu^\alpha)_{\alpha \in [0,1]}$ is uniformly level bounded (so eventually level bounded). By Corollary 6.9, we have that $\alpha \mapsto \varphi_\mu^\alpha$ is epi-continuous on $[0,1]$. As $\lambda_f = \lambda_g = \infty$, φ_μ^α and $\varphi_\mu^{\bar{\alpha}}$ are proper and lsc for every $\mu > 0$. Hence [26, Theorem 7.33] applies. ∎

THEOREM 7.10. *Suppose that the proper, lsc functions f, g are level coercive and* $\operatorname{dom} f \cap \operatorname{dom} g \neq \varnothing$. *Then* $\lim_{\mu\downarrow 0} \inf \varphi_\mu^\alpha = \inf(\alpha f + (1-\alpha)g)$, *and*

$$\limsup_{\mu\downarrow 0} \operatorname{argmin} \varphi_\mu^\alpha \subseteq \operatorname{argmin}(\alpha f + (1-\alpha)g).$$

Moreover, $(\operatorname{argmin} \varphi_\mu^\alpha)_{\mu>0}$ *lies in a bounded set.*

Proof. Note that each φ_μ^α is proper and lsc, and $\alpha f + (1-\alpha)g$ is proper and lsc. By Theorem 6.4, when $\mu \downarrow 0$, φ_μ^α epi-converges to $\alpha f + (1-\alpha)g$. By assumption, there exist $\gamma > 0$ and $\beta \in \mathbb{R}$ such that $f \geq \gamma\|\cdot\| + \beta, g \geq \gamma\|\cdot\| + \beta$. By Lemma 7.7, we have that $\varphi_\mu^\alpha \geq \gamma\|\cdot\| + \beta$ for every $\mu \in]0, \infty[$. Since $\gamma\|\cdot\| + \beta$ is level bounded, $(\varphi_\mu^\alpha)_{\mu\in]0,\infty[}$ is uniformly level bounded (so eventually level bounded). It remains to apply [26, Theorem 7.33]. ∎

THEOREM 7.11. *Suppose that the proper and lsc functions f, g are coercive. Then for every $\bar{\mu} \in]0, \infty]$, we have* $\lim_{\mu\uparrow\bar{\mu}} \inf \varphi_\mu^\alpha = \inf \varphi_{f,g,\alpha,\bar{\mu}}$ *(finite), and*

$$\limsup_{\mu\uparrow\bar{\mu}} \operatorname{argmin} \varphi_\mu^\alpha \subseteq \operatorname{argmin} \varphi_{f,g,\alpha,\bar{\mu}}. \tag{7.3}$$

Moreover, $(\operatorname{argmin} \varphi_\mu^\alpha)_{\mu>0}$ *lies in a bounded set. Consequently,* $\lim_{\mu\uparrow\infty} \inf \varphi_\mu^\alpha = \alpha \min f + (1-\alpha) \min g$, *and*

$$\limsup_{\mu\uparrow\infty} \operatorname{argmin} \varphi_\mu^\alpha \subseteq (\alpha \operatorname{conv} \operatorname{argmin} f + (1-\alpha) \operatorname{conv} \operatorname{argmin} g). \tag{7.4}$$

Proof. Note that each φ_μ^α is proper and lsc for $\mu \in]0, \infty[$. When $\mu = \infty$, By Proposition 7.6, we have that the epi-limit is proper, lsc, and convex. By Theorem 6.2(a), when $\mu \uparrow \bar{\mu}$, φ_μ^α monotonically decrease to $\varphi_{f,g,\alpha,\bar{\mu}}$. Since $\varphi_{f,g,\alpha,\bar{\mu}}$ is lsc, so φ_μ^α epi-converges to $\varphi_{f,g,\alpha,\bar{\mu}}$. By assumption, for every $\gamma > 0$ there exists $\beta \in \mathbb{R}$ such that $f \geq \gamma\|\cdot\| + \beta, g \geq \gamma\|\cdot\| + \beta$. By Lemma 7.7, we have that $\varphi_\mu^\alpha \geq \gamma\|\cdot\| + \beta$ for every $\mu \in]0, \infty[$. Since $\gamma\|\cdot\| + \beta$ is level bounded, $(\varphi_\mu^\alpha)_{\mu\in]0,\infty[}$ is uniformly level bounded (so eventually level bounded). Hence (7.3) follows from [26, Theorem 7.33]. Combining (7.3), Theorem 6.2, and Proposition 7.6 yields (7.4). ∎

8. Subdifferentiability of the proximal average. In this section, we focus on the subdifferentiability and differentiability of the proximal average. Following Benoist and Hiriart-Urruty [9] we say that a family of points $\{x_1, \ldots, x_m\}$ in $\operatorname{dom} f$ is called by $x \in \operatorname{dom} \operatorname{conv} f$ if $x = \sum_{i=1}^m \alpha_i x_i$, and $\operatorname{conv} f(x) = \sum_{i=1}^m \alpha_i f(x_i)$, where $\sum_{i=1}^m \alpha_i = 1$ and $(\forall i)\ \alpha_i > 0$. The following result is the central one of this section.

THEOREM 8.1 (subdifferentiability of the proximal average). *Let $0 < \mu < \bar{\lambda}$, let $x \in \operatorname{dom} \varphi_\mu^\alpha$ and $x = y + z$. Suppose the following conditions hold:*

(a)

$$\left[\alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha) \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right](x)$$
$$= \alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{y}{\alpha}\right) + (1-\alpha) \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{z}{1-\alpha}\right),$$

(b) $\{y_1, \ldots, y_l\}$ *are called by y/α in* $\operatorname{conv}(f + 1/2\mu\|\cdot\|^2)$, *and*

(c) $\{z_1, \ldots, z_m\}$ *are called by $z/(1-\alpha)$ in* $\operatorname{conv}(g + 1/2\mu\|\cdot\|^2)$.

Then $\hat{\partial}\varphi_\mu^\alpha(x) = \partial_L \varphi_\mu^\alpha(x) = \partial_C \varphi_\mu^\alpha(x) =$

$$\left[\cap_{i=1}^l \partial\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i)\right] \cap \left[\cap_{j=1}^m \partial\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)(z_j)\right] - \frac{x}{\mu}.$$

Proof. By Theorem 5.1(c), the Clarke regularity of φ_μ^α and the sum rule of limiting subdifferentials, we have $\hat{\partial}\varphi_\mu^\alpha(x) = \partial_C\varphi_\mu^\alpha(x) = \partial_L\varphi_\mu^\alpha(x) =$

$$\begin{aligned}
&\partial_L \left[\alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha) \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right](x) - \frac{x}{\mu} \\
&= \partial \left[\alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha) \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right](x) - \frac{x}{\mu}.
\end{aligned} \tag{8.1}$$

Using the subdifferential formula for infimal convolution [5, Proposition 16.61] or [34, Corollary 2.4.7], we obtain

$$\begin{aligned}
&\partial \left[\alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{\alpha}\right) \Box (1-\alpha) \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{\cdot}{1-\alpha}\right)\right](x) \\
&= \partial \left[\alpha \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{y}{\alpha}\right)\right] \cap \partial \left[(1-\alpha) \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\frac{z}{1-\alpha}\right)\right] \\
&= \partial \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(\bar{y}) \cap \partial \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)(\bar{z}),
\end{aligned} \tag{8.2}$$

where $\bar{y} = \frac{y}{\alpha}$, $\bar{z} = \frac{z}{1-\alpha}$. The subdifferential formula for the convex hull of a coercive function [9, Corollary 4.9] or [13, Theorem 3.2] gives

$$\partial \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(\bar{y}) = \cap_{i=1}^l \partial\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i), \tag{8.3}$$

$$\partial \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)(\bar{z}) = \cap_{j=1}^m \partial\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)(z_j). \tag{8.4}$$

Therefore, the result follows by combining (8.1)–(8.4). ∎

COROLLARY 8.2. *Let $0 < \mu < \bar{\lambda}$, let $\alpha_i > 0, \beta_j > 0$ with $\sum_{i=1}^l \alpha_i = 1, \sum_{j=1}^m \beta_j = 1$, and let $\alpha \in]0,1[$. Suppose that $x = \alpha \sum_{i=1}^l \alpha_i y_i + (1-\alpha)\sum_{j=1}^m \beta_j z_j$, and*

$$\left[\cap_{i=1}^l \partial\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i)\right] \cap \left[\cap_{j=1}^m \partial\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)(z_j)\right] \neq \varnothing. \tag{8.5}$$

Then $\hat{\partial}\varphi_\mu^\alpha(x) = \partial_L\varphi_\mu^\alpha(x) = \partial_C\varphi_\mu^\alpha(x) =$

$$\left[\cap_{i=1}^l \partial\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i)\right] \cap \left[\cap_{j=1}^m \partial\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)(z_j)\right] - \frac{x}{\mu}.$$

Proof. We will show that

$$\operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\sum_{i=1}^l \alpha_i y_i\right) = \sum_i^l \alpha_i \left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i). \tag{8.6}$$

By (8.5), there exists $y^* \in [\cap_{i=1}^l \partial(f + \frac{1}{2\mu}\|\cdot\|^2)(y_i)] \cap [\cap_{j=1}^m \partial(g + \frac{1}{2\mu}\|\cdot\|^2)(z_j)]$. For every y_i, we have

$$(\forall u \in \mathbb{R}^n) \quad \left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(u) \geq \left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i) + \langle y^*, u - y_i\rangle. \tag{8.7}$$

Multiplying each inequality by α_i, followed by summing them up, gives

$$(\forall u \in \mathbb{R}^n) \ \left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(u) \geq \sum_{i=1}^{l} \alpha_i \left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i) + \left\langle y^*, u - \sum_{i=1}^{l} \alpha_i y_i \right\rangle.$$

Then, $\forall u \in \mathbb{R}^n$

$$\operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(u) \geq \sum_{i=1}^{l} \alpha_i \left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i) + \left\langle y^*, u - \sum_{i=1}^{l} \alpha_i y_i \right\rangle, \tag{8.8}$$

from which

$$\operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\sum_{i=1}^{l} \alpha_i y_i\right) \geq \sum_{i=1}^{l} \alpha_i \left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i). \tag{8.9}$$

Since $\operatorname{conv}(f + \frac{1}{2\mu}\|\cdot\|^2)(\sum_{i=1}^{l} \alpha_i y_i) \leq \sum_{i=1}^{l} \alpha_i (f + \frac{1}{2\mu}\|\cdot\|^2)(y_i)$ is always true, (8.6) is established. Moreover, (8.6) and (8.8) imply

$$y^* \in \partial \operatorname{conv}\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\sum_{i=1}^{l} \alpha_i y_i\right). \tag{8.10}$$

Similar arguments give

$$\operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\sum_{j=1}^{m} \beta_j z_j\right) = \sum_{j}^{m} \beta_j \left(g + \frac{1}{2\mu}\|\cdot\|^2\right)(z_j), \text{ and} \tag{8.11}$$

$$y^* \in \partial \operatorname{conv}\left(g + \frac{1}{2\mu}\|\cdot\|^2\right)\left(\sum_{j=1}^{m} \beta_j z_j\right). \tag{8.12}$$

Put $x = y + z$ with $y = \alpha \sum_{i=1}^{l} \alpha_i y_i$ and $z = (1-\alpha)\sum_{j=1}^{m} \beta_j z_j$. Equations (8.10) and (8.12) guarantee assumption (a) of Theorem 8.1; (8.6) and (8.11) guarantee assumptions (b) and (c) of Theorem 8.1, respectively. Hence, Theorem 8.1 applies. ∎

Armed with Theorem 8.1, we now turn to the differentiability of φ_μ^α.

DEFINITION 8.3. *A function $f_1 : \mathbb{R}^n \to \,]-\infty, +\infty]$ is almost differentiable if $\hat{\partial} f_1(x)$ is a singleton for every $x \in \operatorname{int}(\operatorname{dom} f_1)$, and $\hat{\partial} f_1(x) = \varnothing$ for every $x \in \operatorname{dom} f_1 \setminus \operatorname{int}(\operatorname{dom} f_1)$, if any.*

LEMMA 8.4. *Let $f_1, f_2 : \mathbb{R}^n \to \,]-\infty, +\infty]$ be proper, lsc functions and let $x \in \operatorname{dom} f_1 \cap \operatorname{dom} f_2$. If f_2 is continuously differentiable at x, then*

$$\partial(f_1 + f_2)(x) \subseteq \hat{\partial}(f_1 + f_2)(x) = \hat{\partial} f_1(x) + \nabla f_2(x).$$

Proof. The "$\subseteq$" is immediate from the definition of ∂ and $\hat{\partial}$. The "$=$" is from [26, Exercise 8.8(c)]. ∎

LEMMA 8.5. *Let $f_1 : \mathbb{R}^n \to \,]-\infty, +\infty]$ be proper, lsc, and μ-proximal, and let $x \in \operatorname{int} \operatorname{dom} f_1$. If $\hat{\partial} f_1(x)$ is a singleton, then f_1 is differentiable at x.*

Proof. Observe that $f_2 = f_1 + \frac{1}{2\mu}\|\cdot\|^2$ is convex, and $\partial f_2(x) = \hat{\partial} f_2(x) = \hat{\partial} f_1(x) + \frac{x}{\mu}$. When $\hat{\partial} f_1(x)$ is a singleton, $\partial f_2(x)$ is a singleton. This implies that f_2 is differentiable at x because f_2 is convex and $x \in \operatorname{int}\operatorname{dom} f_2$. Hence, f_1 is differentiable at x. ∎

COROLLARY 8.6 (differentiability of the proximal average). *Let $0 < \mu < \bar{\lambda}$. Suppose that either f or g is almost differentiable (in particular, if f or g is differentiable at every point of its domain). Then φ_μ^α is almost differentiable. In particular, φ_μ^α is differentiable on the interior of its domain.*

Proof. Without loss of generality, assume that f is almost differentiable. By Lemma 8.4,

$$\partial\left(f + \frac{1}{2\mu}\|\cdot\|^2\right)(y_i) \subseteq \hat{\partial} f(y_i) + \frac{y_i}{\mu}. \tag{8.13}$$

It follows that $\partial(f + \frac{1}{2\mu}\|\cdot\|^2)(y_i)$ is at most single valued whenever $\hat{\partial} f(y_i)$ is single valued. With the same notation as in Theorem 8.1, we consider two cases.

Case 1: $x \in \operatorname{bdry}\operatorname{dom}\varphi_\mu^\alpha$. As $x = \alpha(y/\alpha) + (1-\alpha)(z/(1-\alpha))$, we must have $y/\alpha \in (\operatorname{bdry}\operatorname{conv}\operatorname{dom} f)$ and $z/(1-\alpha) \in \operatorname{bdry}(\operatorname{conv}\operatorname{dom} g)$; otherwise

$$x \in \operatorname{int}(\alpha \operatorname{conv}\operatorname{dom} f + (1-\alpha)\operatorname{conv}\operatorname{dom} g) = \operatorname{int}\operatorname{dom}\varphi_\mu^\alpha,$$

which is a contradiction. Then the family of $\{y_1, \ldots, y_m\}$ called by y/α must be from $\operatorname{bdry}\operatorname{dom} f$. As f is almost differentiable, $\hat{\partial} f(y_i) = \varnothing$, then $\hat{\partial}\varphi_\mu^\alpha(x) = \varnothing$ by Theorem 8.1 and (8.13).

Case 2: $x \in \operatorname{int}(\operatorname{dom}\varphi_\mu^\alpha)$. As φ_μ^α is μ-proximal, $\hat{\partial}\varphi_\mu^\alpha(x) \neq \varnothing$. We claim that the family of $\{y_1, \ldots, y_m\}$ called by y/α in Theorem 8.1 is necessarily from $\operatorname{int}\operatorname{dom} f$. If not, then $\partial(f + \frac{1}{2\mu}\|\cdot\|^2)(y_i) = \varnothing$ because of (8.13) and $\hat{\partial} f(y_i) = \varnothing$ for $y_i \in \operatorname{bdry}(\operatorname{dom} f)$. Then Theorem 8.1 implies $\hat{\partial}\varphi_\mu^\alpha(x) = \varnothing$, which is a contradiction. Now $\{y_1, \ldots, y_m\}$ are from $\operatorname{int}\operatorname{dom} f$ and f is almost differentiable, so $(\forall i)$ $\hat{\partial} f(y_i)$ is a singleton. Using (8.13) again and $\hat{\partial}\varphi_\mu^\alpha(x) \neq \varnothing$, we see that $(\forall i)$ $\partial(f + \frac{1}{2\mu}\|\cdot\|^2)(y_i)$ is a singleton. Hence, $\hat{\partial}\varphi_\mu^\alpha(x)$ is a singleton by Theorem 8.1.

Case 1 and Case 2 together show that φ_μ^α is almost differentiable. Finally, φ_μ^α is differentiable on $\operatorname{int}\operatorname{dom}\varphi_\mu^\alpha$ by Lemma 8.5. ∎

COROLLARY 8.7. *Let $0 < \mu < \bar{\lambda}$. Suppose that either f or g is almost differentiable and that either* $\operatorname{conv}\operatorname{dom} f = \mathbb{R}^n$ *or* $\operatorname{conv}\operatorname{dom} g = \mathbb{R}^n$. *Then φ_μ^α is differentiable on $\mathbb{R}^n$.*

Proof. By Theorem 5.1(d), $\operatorname{dom}\varphi_\mu^\alpha = \mathbb{R}^n$. It suffices to apply Corollary 8.6. ∎

We end this section with a result on Lipschitz continuity of the gradient of φ_μ^α.

PROPOSITION 8.8. *Suppose that f (or g) is differentiable with a Lipschtiz continuous gradient and μ-proximal. Then, for every $\alpha \in]0,1[$, the function φ_μ^α is differentiable with a Lipschitz continuous gradient.*

Proof. The function $f + \frac{1}{2\mu}\|\cdot\|^2$ (or $g + \frac{1}{2\mu}\|\cdot\|^2$) is convex and differentiable with a Lipschitz continuous gradient. Apply [26, Proposition 12.60] twice. ∎

9. The general question is still unanswered. According to Theorem 5.4, suppose that $0 < \mu < \bar{\lambda}$, $0 < \alpha < 1$, and $\operatorname{Prox}_\mu f$ and $\operatorname{Prox}_\mu g$ are convex valued. Then there exists a proper, lsc function φ_μ^α such that $\operatorname{Prox}_\mu \varphi_\mu^\alpha = \alpha \operatorname{Prox}_\mu f +$

$(1-\alpha)\operatorname{Prox}_\mu g$. When the proximal mapping is not convex valued, the situation is subtle. We illustrate this by revisiting Example 2.14. Recall that for $\varepsilon_k > 0$, the function $f_k(x) = \max\{0, (1+\varepsilon_k)(1-x^2)\}$ has

$$\operatorname{Prox}_{1/2} f_k(x) = \begin{cases} x & \text{if } x \geq 1, \\ 1 & \text{if } 0 < x < 1, \\ \{-1, 1\} & \text{if } x = 0, \\ -1 & \text{if } -1 < x < 0, \\ x & \text{if } x \leq -1. \end{cases}$$

With $\alpha = 1/2$, we have

$$(\alpha \operatorname{Prox}_{1/2} f_1 + (1-\alpha)\operatorname{Prox}_{1/2} f_2)(x) = \begin{cases} x & \text{if } x \geq 1, \\ 1 & \text{if } 0 < x < 1, \\ \{-1, 0, 1\} & \text{if } x = 0, \\ -1 & \text{if } -1 < x < 0, \\ x & \text{if } x \leq -1. \end{cases} \tag{9.1}$$

Because $\operatorname{Prox}_{1/2} f_i(0)$ is not convex valued, $(\alpha \operatorname{Prox}_{1/2} f_1 + (1-\alpha)\operatorname{Prox}_{1/2} f_2)(0)$ is neither $\operatorname{Prox}_{1/2} f_1(0)$ nor $\operatorname{Prox}_{1/2} f_2(0)$. One can verify that (9.1) is indeed $\operatorname{Prox}_{1/2} g(x)$, where

$$g(x) = \begin{cases} 0 & \text{if } x > 1, \\ -x(x-1) - x^2 + 1 & \text{if } 0 < x \leq 1, \\ -x(x+1) - x^2 + 1 & \text{if } -1 < x \leq 0, \\ 0 & \text{if } x \leq -1. \end{cases}$$

Regretfully, we do not have a systematic way to find g when $\operatorname{Prox}_\mu g$ is not convex valued. The challenging question is still open: *Is a convex combination of proximal mappings of possibly nonconvex functions always a proximal mapping?*

Acknowledgment. The authors thank two anonymous referees for insightful comments.

REFERENCES

[1] H. Attouch, *Variational Convergence for Functions and Operators*, Appl. Math., Pitman, Boston, MA, 1984.

[2] H. Attouch and D. Azé, *Approximation and regularization of arbitrary functions in Hilbert spaces by the Lasry-Lions method*, Ann. Inst. Henri Poincaré Anal. Non Linéaire, 10 (1993), pp. 289–312.

[3] M. Bačák, J. Borwein, A. Eberhard, and B. Mordukhovich, *Infimal convolutions and Lipschitzian properties of subdifferentials for prox-regular functions in Hilbert spaces*, J. Convex Anal., 17 (2010), pp. 737–763.

[4] S. Bartz, H. H. Bauschke, S. M. Moffat, and X. Wang, *The resolvent average of monotone operators: Dominant and recessive properties*, SIAM J. Optim., 26 (2016), pp. 602–634.

[5] H. H. Bauschke and P. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.

[6] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, *The proximal average: Basic theory*, SIAM J. Optim., 19 (2008), pp. 766–785.

[7] H. Bauschke, E. Matoušková, and S. Reich, *Projection and proximal point methods: Convergence results and counterexamples*, Nonlinear Anal., 56 (2004), pp. 715–738.

[8] G. Beer, *Topologies on Closed and Closed Convex Sets*, Math. Appl., Kluwer Academic, Dordrecht, The Netherlands, 1993.

[9] J. Benoist and J.-B. Hiriart-Urruty, *What is the subdifferential of the closed convex hull of a function?*, SIAM J. Math. Anal., 27 (1996), pp. 1661–1679.

[10] F. Bernard and L. Thibault, *Prox-regular functions in Hilbert spaces*, J. Math. Anal. Appl., 303 (2005), pp. 1–14.

[11] F. Clarke, *Optimization and Nonsmooth Analysis*, 2nd ed., Classics in Appl. Math. 5, SIAM, Philadelphia, 1990.

[12] R. Goebel, *The proximal average for saddle functions and its symmetry properties with respect to partial and saddle conjugacy*, J. Nonlinear Convex Anal, 11 (2010), pp. 1–11.

[13] A. Griewank and P. Rabier, *On the smoothness of convex envelopes*, Trans. Amer. Math. Soc., 322 (1990), pp. 691–709.

[14] A. Guevara and P. Wolenski, *Convergence results for a self-dual regularization of convex problems*, Optimization, 61 (2012), pp. 699–716.

[15] W. L. Hare, *A proximal average for nonconvex functions: A proximal stability perspective*, SIAM J. Optim., 20 (2009), pp. 650–666.

[16] A. Jourani, L. Thibault, and D. Zagrodny, *Differential properties of the Moreau envelope*, J. Funct. Anal., 266 (2014), pp. 1185–1237.

[17] S. Kim, J. Lawson, and Y. Lim, *The matrix geometric mean of parameterized, weighted arithmetic and harmonic means*, Linear Algebra Appl., 435 (2011), pp. 2114–2131.

[18] A. Lewis, D. Luke, and J. Malick, *Local linear convergence for alternating and averaged nonconvex projections*, Found. Comput. Math., 9 (2009), pp. 485–513.

[19] Y. Lim, *Contractive barycentric maps and l^1 ergodic theorems on the cone of positive definite matrices*, J. Math. Anal. Appl., 459 (2018), pp. 291–306.

[20] A. Lohne, *A characterization of maximal monotone operators*, Set-Valued Anal., 16 (2008), pp. 693–700.

[21] B. Mordukhovich, *Variational Analysis and Generalized Differentiation* I*: Basic Theory*, Grundlehren Math. Wiss. 330, Springer, Berlin, 2006.

[22] N. Parikh and S. Boyd, *Proximal algorithms*, Found. Trends Optim., 1 (2014), pp. 127–239.

[23] C. Planiden and X. Wang, *Strongly convex functions, Moreau envelopes and the generic nature of convex functions with strong minimzers*, SIAM J. Optim., 26 (2016), pp. 1341–1364.

[24] M. Reid and R. Williamson, *Information, divergence and risk for binary experiments*, J. Mach. Learn. Res., 12 (2011), pp. 731–817.

[25] R. Rockafellar, *Convex Analysis*, Princeton Landmarks Math., Princeton University Press, Princeton, New Jersey, 1997.

[26] R. Rockafellar and J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[27] S. Simons, *From Hahn-Banach to Monotonicity*, 2nd ed., Lecture Notes in Math. 1693, Springer, New York, 2008.

[28] V. Soloviov, *Duality for nonconvex optimization and its applications*, Anal. Math., 19 (1993), pp. 297–315.

[29] X. Wang, *On Chebyshev functions and Klee functions*, J. Math. Anal. Appl., 368 (2010), pp. 293–310.

[30] Y. Yu, *Better approximation and faster algorithm using the proximal average*, in Advances in Neural Information Processing Systems 27 (NIPS), Curran Associates, Red Hook, NY, 2013.

[31] A. Zaslavski, *A proximal point algorithm for finding a common zero of a finite family of maximal monotone operators in the presence of computational errors*, Nonlinear Anal., 75 (2012), pp. 6071–6087.

[32] K. Zhang, *Compensated convexity and its applications*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 25 (2008), pp. 743–771.

[33] K. Zhang, E. Crooks, and A. Orlando, *Compensated convexity methods for approximations and interpolations of sampled functions in Euclidean spaces: Theoretical foundations*, SIAM J. Math. Anal., 48 (2016), pp. 4126–4154.

[34] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, River Edge, NJ, 2002.