The Proximal Average: Basic Theory

Heinz H. Bauschke[*], Rafal Goebel[†], Yves Lucet[‡] and Xianfu Wang[§]

February 14, 2008 (revised version; originally submitted on April 6, 2007)

Abstract

The recently introduced proximal average of two convex functions is a convex function with many useful properties. In this paper, we introduce and systematically study the proximal average for finitely many convex functions. The basic properties of the proximal average with respect to the standard convex-analytical notions (domain, Fenchel conjugate, subdifferential, proximal mapping, epi-continuity, and others) are provided and illustrated by several examples.

2000 Mathematics Subject Classification:
Primary 90C25; Secondary 26A51, 26B25, 26E60, 46C05, 47H05, 52A41.

Keywords: Arithmetic average, arithmetic mean, convex analysis, convex function, epi-convergence, epigraphical average, epi-topology, essential smoothness, essential strict convexity, Fenchel conjugate, harmonic mean, Legendre function, proximal average, proximal mapping, subdifferential operator.

1 Overview

Let f_1 and f_2 be two functions that are convex, lower semicontinuous, and proper, and let λ_1 and λ_2 be strictly positive real numbers adding up to 1. How can we average the two functions f_1 and f_2 with respect to the weights λ_1 and λ_2 in a useful way? Perhaps the first approach is to consider the *arithmetic average* $\lambda_1 f_1 + \lambda_2 f_2$. However, functions in convex analysis are allowed to take on

[*]Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: `heinz.bauschke@ubc.ca`.

[†]Mathematics, University of Washington, Box 354350, Seattle, Washington 98195-4350, USA. E-mail: `goebel@math.washington.edu`.

[‡]Computer Science, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: `yves.lucet@ubc.ca`.

[§]Mathematics, Irving K. Barber School, University of British Columbia Okanagan, Kelowna, British Columbia V1V 1V7, Canada. E-mail: `shawn.wang@ubc.ca`.

the value $+\infty$, for example to model constraints in optimization problems. Thus, the arithmetic average can turn out to be $+\infty$ everywhere and then carries little information about f_1 and f_2; this happens whenever f_1 and f_2 are nowhere both finite. How could we possibly average such functions? A second thought may suggest to construct the *epigraphical average* $\lambda_1 \star f_1 \oplus \lambda_2 \star f_2$ obtained by forming a convex combination of the epigraphs of f_1 and f_2. Unfortunately, if the functions f_1 and f_2 lack coercivity, then the epigraphical average fails to be helpful: for instance, if f_1 and f_2 are two distinct linear functions, then their epigraphical average is identically equal to $-\infty$, and hence of little use. The *proximal average*, first introduced in [6] in the context of fixed point theory and recently studied in [4, 5, 7, 10, 15] from various viewpoints, avoids the mentioned difficulties and possesses numerous properties that are attractive to Convex Analysts.

The aim of this paper is to provide the basic theory of the proximal average. In addition, we extend it to more than two functions and we allow for an additional positive parameter. For the reader's convenience and the sake of completeness, the presentation of the theory is largely self-contained. It is shown that the proximal average has many desirable properties in terms of its domain, Fenchel conjugate, Moreau envelope, proximal mapping, subdifferential, epi-continuity, and other convex-analytical notions. Moreover, the arithmetic and epigraphical averages turn out to be limits of the proximal average, as the parameter tends to 0 and $+\infty$, respectively. Various examples illustrate our results. An interesting topic for future research is the extension to series and integrals.

The rest of this paper is organized as follows. Section 2 collects the notation used throughout this paper, and in Section 3 we collect and present results that simplify later proofs. The proximal average is introduced in Section 4 where also its domain is characterized. In Section 5, we present one very useful result (Theorem 5.1) which states that the Fenchel conjugate of the proximal average is the proximal average of the Fenchel conjugates. An important consequence of this result is that the proximal average is convex, lower semicontinuous and proper. In Section 6 we consider the Moreau envelope and proximal mapping of the proximal average, in Section 7 its subdifferential operator as well as essential smoothness and essential strict convexity. In Section 8, it is shown that the arithmetic and epigraphical averages are pointwise limits of the proximal average. Epi-convergence properties are discussed in the final Section 9, where the arithmetic and epigraphical averages are shown to be limiting instances of the proximal average with respect to epi-convergence.

2 Standing Assumptions and Notation

Throughout this paper,

$$X \text{ is a real Hilbert space with inner product } \langle\cdot,\cdot\rangle \text{ and corresponding norm } \|\cdot\|. \tag{1}$$

Due to its repeated use, we abbreviate the quadratic energy function by

$$\mathsf{q} = \tfrac{1}{2}\|\cdot\|^2. \tag{2}$$

We set

$$\Gamma(X) = \big\{f\colon X \to \left]-\infty,+\infty\right] \mid f \text{ is convex, lower semicontinuous, and proper}\big\}. \tag{3}$$

We assume throughout that

$$n \in \{1, 2, 3, \ldots\}, \tag{4}$$

that

$$f_1, \ldots, f_n \text{ belong to } \Gamma(X), \tag{5}$$

that

$$\lambda_1, \ldots, \lambda_n \text{ are nonnegative real numbers such that } \lambda_1 + \cdots + \lambda_n = 1, \tag{6}$$

and that

$$\mu \text{ is a strictly positive real number.} \tag{7}$$

The Fenchel conjugate of a function f is denoted by f^*. It will be convenient to set

$$\boldsymbol{f} = (f_1, \ldots, f_n), \quad \boldsymbol{f}^* = (f_1^*, \ldots, f_n^*), \quad \text{and} \quad \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n). \tag{8}$$

Other notation not explicitly defined here or later is standard in Convex Analysis and as in, e.g., [21, 22, 24]. Let f be a convex function and S be a set. Then we write $\operatorname{dom} f$, $\operatorname{epi} f$, ∂f, $\operatorname{cl} f$, $\inf f$, $\min f$, $\operatorname{argmin} f$, d_S, $\operatorname{conv} S$, $\operatorname{int} S$, ι_S, and N_S to denote the (effective) domain, epigraph, subdifferential operator, lower closure, infimum value, minimum value if the infimum value is attained, the set of minimizers, distance function, convex hull, interior, indicator function, and normal cone operator, respectively. The identity operator is represented by Id.

3 Auxiliary Results

We start by reviewing the key notions of epi-multiplication and epi-addition, following the viewpoint taken in [22, Section 1.H]. Let $\alpha \geq 0$, $f \in \Gamma(X)$, $g \in \Gamma(X)$, and $h \in \Gamma(X)$. Then

$$\alpha \star f = \begin{cases} \alpha f(\cdot/\alpha), & \text{if } \alpha > 0; \\ \iota_{\{0\}}, & \text{if } \alpha = 0. \end{cases} \tag{9}$$

The term "epi-multiplication" stems from the fact that $\operatorname{epi}(\alpha \star f) = \alpha \operatorname{epi}(f)$ when $\alpha > 0$. Epi-addition or infimal convolution is defined by

$$f \mathbin{\dot{+}} g \colon X \to [-\infty, +\infty] : x \mapsto \inf_{y+z=x} \big(f(y) + g(z)\big); \tag{10}$$

and the term "epi-addition" stems from the fact that the strict epigraph of $f \mathbin{\dot{+}} g$ is the Minkowski sum of the strict epigraphs of f and g, i.e., $\big\{(x,r) \in X \times \mathbb{R} \mid (f \mathbin{\dot{+}} g)(x) < r\big\} = \big\{(y,s) \in X \times \mathbb{R} \mid f(y) < s\big\} + \big\{(z,t) \in X \times \mathbb{R} \mid g(z) < t\big\}$. The epi-sum of finitely many functions is defined analogously.

To avoid excessive usage of parentheses, epi-multiplication and regular multiplication are given precedence over epi- and regular addition, i.e., $\alpha \star f + g = (\alpha \star f) + g$, $\alpha \star f \mathbin{\dot{+}} g = (\alpha \star f) \mathbin{\dot{+}} g$, $\alpha f + g = (\alpha f) + g$, and $\alpha f \mathbin{\dot{+}} g = (\alpha f) \mathbin{\dot{+}} g$. It will also be convenient to give epi-addition a higher precedence than regular addition or subtraction, i.e., $f \mathbin{\dot{+}} g + h = (f \mathbin{\dot{+}} g) + h$ and $f \mathbin{\dot{+}} g - h = (f \mathbin{\dot{+}} g) - h$.

The next three propositions are elementary. Proofs for the finite-dimensional case are in [22]; they extend without difficulty to the present Hilbert space setting.

Proposition 3.1 *Let $f \in \Gamma(X)$, let $\alpha \geq 0$, and let $\beta \geq 0$. Then the following hold.*

(i) $\alpha > 0 \Rightarrow \operatorname{epi}(\alpha \star f) = \alpha(\operatorname{epi} f)$.

(ii) $\operatorname{dom}(\alpha \star f) = \alpha(\operatorname{dom} f)$.

(iii) $f \mathbin{\dot{+}} \iota_{\{0\}} = f$.

(iv) $\operatorname{dom}(f_1 \mathbin{\dot{+}} \cdots \mathbin{\dot{+}} f_n) = (\operatorname{dom} f_1) + \cdots + (\operatorname{dom} f_n)$.

(v) $\alpha \star (f_1 \mathbin{\dot{+}} \cdots \mathbin{\dot{+}} f_n) = \alpha \star f_1 \mathbin{\dot{+}} \cdots \mathbin{\dot{+}} \alpha \star f_n$.

(vi) $\alpha(f_1 \mathbin{\dot{+}} \cdots \mathbin{\dot{+}} f_n) = \alpha f_1 \mathbin{\dot{+}} \cdots \mathbin{\dot{+}} \alpha f_n$.

(vii) $\alpha \star (\beta \star f) = (\alpha\beta) \star f$.

(viii) $(\alpha + \beta) \star f = \alpha \star f \mathbin{\dot{+}} \beta \star f$.

(ix) $\alpha > 0 \Rightarrow \alpha(\beta \star (\alpha^{-1} f)) = \beta \star f$.

Proof. The conclusions all follow readily from the definitions; see also [22, Exercise 1.28(a)] for (i), [22, page 25] for (v) and (vii), and [22, Exercise 2.24(c)] for (viii). ■

Proposition 3.2 *Let $\alpha \geq 0$. Then the following hold.*

(i) $(\alpha f)^* = \alpha \star f^*$.

(ii) $(\alpha \star f)^* = \alpha f^*$.

(iii) $(f_1 \mathbin{\dot{+}} \cdots \mathbin{\dot{+}} f_n)^* = f_1^* + \cdots + f_n^*$.

Proof. The statements are simple consequences of the definitions; see also [22, page 475] for (i) and (ii), and [22, Theorem 11.23(a)] for (iii). ■

Proposition 3.3 *Let $f \in \Gamma(X)$ and let $\alpha \geq 0$. Then the following hold.*

(i) $\mathsf{q}^* = \mathsf{q}$*; in fact,* q *is the only function equal to its Fenchel conjugate.*

(ii) $\alpha > 0 \Rightarrow \alpha^{-1} \star \mathsf{q} = \alpha\, \mathsf{q}$.

(iii) $(\alpha \star \mathsf{q})^* = \alpha\, \mathsf{q}$.

(iv) $(\alpha\, \mathsf{q})^* = \alpha \star \mathsf{q}$.

(v) $(f \mathbin{\dot{+}} \mathsf{q}) + (f^* \mathbin{\dot{+}} \mathsf{q}) = \mathsf{q}$.

Proof. (i): See, e.g., [22, Example 11.11]. (ii): An immediate consequence of the definition of $\mathfrak{q}$. (iii): Combine Proposition 3.2(ii) with (i). (iv): Combine Proposition 3.2(i) with (i). (v): See [19] or [22, Example 11.26]. ■

The next result is deep and stated as a fact.

Fact 3.4 *The following hold.*

(i) *If* $\operatorname{int}\operatorname{dom} f_1 \cap \cdots \cap \operatorname{int}\operatorname{dom} f_{n-1} \cap \operatorname{dom} f_n \neq \varnothing$, *then* $(f_1 + \cdots + f_n)^* = f_1^* \mathbin{\Box} \cdots \mathbin{\Box} f_n^*$ *and the epi-sum is exact, i.e., the infimum in the definition of the epi-sum is attained.*

(ii) *If* $\operatorname{int}\operatorname{dom} f_1^* \cap \cdots \cap \operatorname{int}\operatorname{dom} f_{n-1}^* \cap \operatorname{dom} f_n^* \neq \varnothing$, *then* $f_1 \mathbin{\Box} \cdots \mathbin{\Box} f_n$ *is exact and* $\operatorname{epi}(f_1 \mathbin{\Box} \cdots \mathbin{\Box} f_n) = (\operatorname{epi} f_1) + \cdots + (\operatorname{epi} f_n)$.

Proof. This is a consequence of [24, Theorem 2.8.7]. ■

The following result on the conjugate of the difference will be useful.

Fact 3.5 *Let* $g \in \Gamma(X)$ *and let* $h \in \Gamma(X)$ *such that both* h *and* h^* *have full domain. Then*

$$\big(\forall x^* \in X\big) \quad (g - h)^*(x^*) = \sup_{y^* \in X} \big(g^*(y^*) - h^*(y^* - x^*)\big). \tag{11}$$

Proof. This is a consequence of [11, Theorem 2.2]. ■

Corollary 3.6 *Let* $g \in \Gamma(X)$. *Then*

$$(g - \mu \star \mathfrak{q})^* = \mu(\mathfrak{q} - \mu^{-1} g^*)^* - \mu^{-1} \star \mathfrak{q}. \tag{12}$$

Proof. Set $h = \mu \star \mathfrak{q}$. Then $h^* = \mu\,\mathfrak{q}$ by Proposition 3.3(iii) and hence both h and h^* have full domain. Using Fact 3.5, we deduce that for every $x^* \in X$

$$\begin{aligned}
(g - h)^*(x^*) &= \sup_{y^* \in X} \big(g^*(y^*) - \mu\,\mathfrak{q}(y^* - x^*)\big) \\
&= \sup_{y^* \in X} \big(g^*(y^*) - \mu\,\mathfrak{q}(y^*) - \mu\,\mathfrak{q}(x^*) + \mu\langle y^*, x^*\rangle\big) \\
&= -\mu\,\mathfrak{q}(x^*) + \sup_{y^* \in X} \big(\langle y^*, \mu x^*\rangle - \big(\mu\,\mathfrak{q}(y^*) - g^*(y^*)\big)\big) \\
&= -\mu\,\mathfrak{q}(x^*) + \mu \sup_{y^* \in X} \big(\langle y^*, x^*\rangle - \big(\mathfrak{q}(y^*) - \mu^{-1} g^*(y^*)\big)\big) \\
&= -(\mu^{-1} \star \mathfrak{q})(x^*) + \mu(\mathfrak{q} - \mu^{-1} g^*)^*(x^*).
\end{aligned} \tag{13}$$

The proof is complete. ■

Lemma 3.7 $\big(\lambda_1 \star (f_1 + \mu \star \mathfrak{q}) \mathbin{\Box} \cdots \mathbin{\Box} \lambda_n \star (f_n + \mu \star \mathfrak{q})\big)^* = \lambda_1(f_1^* \mathbin{\Box} \mu\,\mathfrak{q}) + \cdots + \lambda_n(f_n^* \mathbin{\Box} \mu\,\mathfrak{q})$.

Proof. Using Proposition 3.2(iii), Proposition 3.2(ii), Fact 3.4(i), and Proposition 3.3(iii), we compute that

$$\begin{aligned}\big(\lambda_1 \star (f_1 + \mu \star \mathsf{q}) \dotplus \cdots \dotplus \lambda_n \star (f_n + \mu \star \mathsf{q})\big)^* &= \big(\lambda_1 \star (f_1 + \mu \star \mathsf{q})\big)^* + \cdots + \big(\lambda_n \star (f_n + \mu \star \mathsf{q})\big)^* \\ &= \lambda_1 (f_1 + \mu \star \mathsf{q})^* + \cdots + \lambda_n (f_n + \mu \star \mathsf{q})^* \\ &= \lambda_1 \big(f_1^* \dotplus (\mu \star \mathsf{q})^*\big) + \cdots + \lambda_n \big(f_n^* \dotplus (\mu \star \mathsf{q})^*\big) \\ &= \lambda_1 (f_1^* \dotplus \mu\, \mathsf{q}) + \cdots + \lambda_n (f_n^* \dotplus \mu\, \mathsf{q}). \end{aligned} \tag{14}$$

This completes the proof. ■

Fact 3.8 *Let* $(\forall i)\ x_i \in \operatorname{dom} f_i$*, and set* $x = x_1 + \cdots + x_n$*. Then the following implications hold.*

(i) $(f_1 \dotplus \cdots \dotplus f_n)(x) = f_1(x_1) + \cdots + f_n(x_n) \Rightarrow \partial(f_1 \dotplus \cdots \dotplus f_n)(x) = \partial f_1(x_1) \cap \cdots \cap \partial f_n(x_n)$.

(ii) $\partial f_1(x_1) \cap \cdots \cap \partial f_n(x_n) \neq \varnothing \Rightarrow (f_1 \dotplus \cdots \dotplus f_n)(x) = f_1(x_1) + \cdots + f_n(x_n)$.

Proof. See [24, Corollary 2.4.7]. ■

Proposition 3.9 *Let* $f \in \Gamma(X)$ *and let* $\alpha > 0$*. Then* $\partial(0 \star f) = N_{\{0\}}$ *and* $\partial(\alpha \star f) = (\partial f) \circ (\alpha^{-1} \operatorname{Id})$.

Proof. Since $0 \star f = \iota_{\{0\}}$, we deduce that $\partial(0 \star f) = \partial \iota_{\{0\}} = N_{\{0\}}$. Also, $\partial(\alpha \star f) = \partial(\alpha f \circ (\alpha^{-1} \operatorname{Id}))$; the formula thus follows from Convex Calculus (see, e.g., [24, Theorem 2.8.3]). ■

4 Definition, Reformulations, Domain and Exactness

In Section 1, we have seen that the idea of computing the averaged Minkowski sum is doomed in general, due to the potential lack of coercivity properties of the terms. The proximal average can be interpreted as a three-step remedy of this idea: First, each function is "coercified" by epi-adding $\mu \star \mathsf{q}$. Second, the epi-average of the coercified terms is computed. The third step removes $\mu \star \mathsf{q}$ through subtraction. We are now ready to describe the proximal average.

Definition 4.1 (proximal average) *The* $\boldsymbol{\lambda}$-weighted proximal average of $\boldsymbol{f}$ with parameter μ *is*

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 \star (f_1 + \mu \star \mathsf{q}) \dotplus \cdots \dotplus \lambda_n \star (f_n + \mu \star \mathsf{q}) - \mu \star \mathsf{q}, \tag{15}$$

i.e., if $I = \big\{ i \in \{1, \ldots, n\} \mid \lambda_i > 0 \big\}$*, then*

$$(\forall x \in X) \quad p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = \frac{1}{\mu}\Big(-\tfrac{1}{2}\|x\|^2 + \inf_{\sum_{i \in I} x_i = x} \sum_{i \in I} \lambda_i \big(\mu f_i(x_i/\lambda_i) + \tfrac{1}{2}\|x_i/\lambda_i\|^2\big)\Big). \tag{16}$$

We also write $p(\boldsymbol{f}, \boldsymbol{\lambda})$ *if* $\mu = 1$*,* $p_\mu(\boldsymbol{f})$ *if all* λ_i *coincide, and* $p(\boldsymbol{f})$ *if* $\mu = 1$ *and all* λ_i *coincide.*

Remark 4.2 Some immediate consequences of the definition are the following.

(i) $p_\mu(f_1,1)=f_1$.

(ii) If $I=\big\{i\in\{1,\ldots,n\}\mid \lambda_i>0\big\}$, $\widetilde{\boldsymbol{f}}=(f_i)_{i\in I}$ and $\widetilde{\boldsymbol{\lambda}}=(\lambda_i)_{i\in I}$, then $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})=p_\mu(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\lambda}})$.

(iii) If π is a permutation of $I=\{1,\ldots,n\}$, $\widetilde{\boldsymbol{f}}=(f_{\pi(i)})_{i\in I}$ and $\widetilde{\boldsymbol{\lambda}}=(\lambda_{\pi(i)})_{i\in I}$, then $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})=p_\mu(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\lambda}})$.

(iv) $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})=\mu^{-1}p_1(\mu\boldsymbol{f},\boldsymbol{\lambda})$; equivalently, $p(\mu\boldsymbol{f},\boldsymbol{\lambda})=\mu p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$.

(v) If $\Lambda_{n-1}=\lambda_1+\cdots+\lambda_{n-1}>0$, then

$$\begin{aligned}p_1(\boldsymbol{f},\boldsymbol{\lambda})&=p_1\big((f_1,\ldots,f_n),(\lambda_1,\ldots,\lambda_n)\big)\\&=p_1\Bigg(\bigg(p_1\big((f_1,\ldots,f_{n-1}),\Lambda_{n-1}^{-1}(\lambda_1,\ldots,\lambda_{n-1})\big),f_n\bigg),(\Lambda_{n-1},\lambda_n)\Bigg).\end{aligned}\tag{17}$$

The identities in items (iv) and (v) may be useful if one wishes to develop the theory of results for a general $\mu>0$ and a general $n\geq 2$ from the simpler case $\mu=1$ and $n=2$; however, the direct approach favoured in this paper is not only self-contained but it also yields proofs that we found much more readable. Nonetheless, (iv) and (v) may be convenient for the numerical computation of the proximal average — especially when the simpler case is already implemented [15].

Proposition 4.3 (reformulations)

$$\begin{aligned}p_\mu(\boldsymbol{f},\boldsymbol{\lambda})&=\big(\lambda_1(f_1^*\,\square\,\mu\,\mathsf{q})+\cdots+\lambda_n(f_n^*\,\square\,\mu\,\mathsf{q})\big)^*-\mu^{-1}\,\mathsf{q} &(18)\\&=\big(\lambda_1(f_1+\mu^{-1}\,\mathsf{q})^*+\cdots+\lambda_n(f_n+\mu^{-1}\,\mathsf{q})^*\big)^*-\mu^{-1}\,\mathsf{q} &(19)\end{aligned}$$

and

$$\big(\forall x\in X\big)\quad p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)=\inf_{\sum\lambda_iy_i=x}\sum\lambda_if_i(y_i)+\frac{1}{\mu}\bigg(\Big(\sum\lambda_i\,\mathsf{q}(y_i)\Big)-\mathsf{q}(x)\bigg).\tag{20}$$

Proof. By Proposition 3.1(iv), $(\forall i)$ $\operatorname{dom}(f_i^*\,\square\,\mu\,\mathsf{q})=(\operatorname{dom}f_1^*)+(\operatorname{dom}\mu\,\mathsf{q})=X$. Fact 3.4(i), Proposition 3.2(i), Proposition 3.2(iii), and Proposition 3.3(iv) imply that

$$\begin{aligned}\big(\lambda_1(f_1^*\,\square\,\mu\,\mathsf{q})+\cdots+\lambda_n(f_n^*\,\square\,\mu\,\mathsf{q})\big)^*&=(\lambda_1(f_1^*\,\square\,\mu\,\mathsf{q}))^*\,\square\cdots\square\,(\lambda_n(f_n^*\,\square\,\mu\,\mathsf{q}))^*\\&=\lambda_1\star(f_1^*\,\square\,\mu\,\mathsf{q})^*\,\square\cdots\square\,\lambda_n\star(f_n^*\,\square\,\mu\,\mathsf{q})^*\\&=\lambda_1\star(f_1^{**}+(\mu\,\mathsf{q})^*)\,\square\cdots\square\,\lambda_n\star(f_n^{**}+(\mu\,\mathsf{q})^*)\\&=\lambda_1\star(f_1+\mu\star\mathsf{q})\,\square\cdots\square\,\lambda_n\star(f_n+\mu\star\mathsf{q}).\end{aligned}\tag{21}$$

This and Proposition 3.3(ii) yield (18). In turn, Fact 3.4(i) and Proposition 3.3(iv) imply (19). Changing variables, we see that (20) is equivalent to (16). ■

Remark 4.4 (some history) In [6], the proximal average was considered for $n=2$ and $\mu=1$, and written equivalently as

$$\big(\lambda_1(f_1^*\,\square\,\mathsf{q})+\lambda_2(f_2^*\,\square\,\mathsf{q})\big)^*-\mathsf{q};\tag{22}$$

see (18). The function (22) was utilized in [6] to explicitly illustrate Moreau's observation [19] that the set of proximal mappings is convex. More recently, the proximal average was considered in [4], again with $n = 2$ and $\mu = 1$, though it was written as (see (19))

$$\big(\lambda_1(f_1 + \mathsf{q})^* + \lambda_2(f_2 + \mathsf{q})^*\big)^* - \mathsf{q}. \tag{23}$$

Example 4.5 (connection to means of numbers) Let $\alpha_1, \ldots, \alpha_n$ be strictly positive numbers and suppose that $(\forall i)\ f_i = \alpha_i\,\mathsf{q}$. Using (19), we see that

$$p_{\mu^{-1}}(\boldsymbol{f}, \boldsymbol{\lambda}) = \Big(\sum_{i=1}^n \lambda_i(\alpha_i\,\mathsf{q} + \mu\,\mathsf{q})^*\Big)^* - \mu\,\mathsf{q} = \Big(\sum_{i=1}^n \frac{\lambda_i}{\alpha_i + \mu}\,\mathsf{q}\Big)^* - \mu\,\mathsf{q} = \Big(\sum_{i=1}^n \frac{\lambda_i}{\alpha_i + \mu}\Big)^{-1}\mathsf{q} - \mu\,\mathsf{q} \tag{24}$$

and thus

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \left(\Big(\sum_{i=1}^n \frac{\lambda_i}{\alpha_i + \mu^{-1}}\Big)^{-1} - \mu^{-1}\right)\mathsf{q}. \tag{25}$$

Denote the coefficient of q in (25) by δ. Since δ is the difference of the weighted harmonic mean of $\alpha_1 + \mu^{-1}, \ldots, \alpha_n + \mu^{-1}$ and μ^{-1}, the Harmonic-Arithmetic Mean Inequality implies that δ does not exceed the weighted arithmetic mean

$$\sum_{i=1}^n \lambda_i \alpha_i. \tag{26}$$

As $\mu \to +\infty$, we note that δ converges to the weighted harmonic mean

$$\Big(\sum_{i=1}^n \frac{\lambda_i}{\alpha_i}\Big)^{-1}, \tag{27}$$

while a calculus exercise shows that δ approaches, as $\mu \to 0^+$, the weighted arithmetic mean (26). In Remark 8.6, we revisit this example from a more general point of view.

The next result locates the domain of the proximal average exactly; moreover, it strengthens [4, Theorem 4.11], where equality was observed only for the closures and interiors.

Theorem 4.6 (domain) $\operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_n \operatorname{dom} f_n$.

Proof. Using Proposition 3.1(iv) and Proposition 3.1(ii), we obtain $\operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \operatorname{dom}(\lambda_1 \star (f_1 + \mu \star \mathsf{q})) + \cdots + \operatorname{dom}(\lambda_1 \star (f_1 + \mu \star \mathsf{q})) = \lambda_1 \operatorname{dom}(f_1 + \mu \star \mathsf{q}) + \cdots + \lambda_n \operatorname{dom}(f_n + \mu \star \mathsf{q}) = \lambda_1 \operatorname{dom}(f_1) + \cdots + \lambda_n \operatorname{dom}(f_n)$. ■

Corollary 4.7 *Suppose that at least one function f_i has full domain and that $\lambda_i > 0$. Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ has full domain.*

Example 4.8 Assume each $\lambda_i > 0$ and each $f_i = \iota_{C_i}$, where C_i is a nonempty closed convex subset of X. In X^n, set $H = \big\{(z_i) \mid \sum \sqrt{\lambda_i} z_i = 0\big\}$ and $(\forall x \in X)$ D_x is the Cartesian product $\times \big(\sqrt{\lambda_i} C_i - \sqrt{\lambda_i} x\big)$. Then

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\colon X \to \left]-\infty, +\infty\right] : x \mapsto \tfrac{1}{2\mu} d^2_{H \cap D_x}(0). \tag{28}$$

Proof. Fix $x \in X$. Using (16), we obtain

$$\begin{aligned} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) &= \mu^{-1}\Big(-\tfrac{1}{2}\|x\|^2 + \inf_{\sum_i x_i = x} \sum \lambda_i\big(\mu\, \iota_{C_i}(x_i/\lambda_i) + \tfrac{1}{2}\|x_i/\lambda_i\|^2\big)\Big) \\ &= \mu^{-1} \inf_{\substack{\text{each } c_i \in C_i \\ \sum \lambda_i c_i = x}} \sum \lambda_i\big(\tfrac{1}{2}\|c_i\|^2 - \tfrac{1}{2}\|x\|^2\big) \\ &= \mu^{-1} \inf_{z = (z_i) \in H \cap D_x} \sum \lambda_i\big(\tfrac{1}{2}\big\|x + z_i/\sqrt{\lambda_i}\big\|^2 - \tfrac{1}{2}\|x\|^2\big) \\ &= \mu^{-1} \inf_{z = (z_i) \in H \cap D_x} \sum \tfrac{1}{2}\|z_i\|^2, \end{aligned} \tag{29}$$

which completes the proof. ■

Remark 4.9 Consider Example 4.8 with $n = 2$, $\mu = 1$, $\lambda_1 > 0$, and $\lambda_2 > 0$. Then (28) simplifies to

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\colon X \to \left]-\infty, +\infty\right] : x \mapsto \frac{1}{2\lambda_1\lambda_2} d^2_{(\lambda_1(C_1 - x)) \cap (\lambda_2(x - C_2))}(0), \tag{30}$$

which is a formula first observed in [6, Theorem 6.1].

Theorem 4.10 (exactness) *For every $x \in \operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ there exist $y_i \in \lambda_i \operatorname{dom} f_i$ such that $x = y_1 + \cdots + y_n$ and $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) = (\lambda_1 \star (f_1 + \mu \star \mathsf{q}))(y_1) + \cdots + (\lambda_n \star (f_n + \mu \star \mathsf{q}))(y_n) - (\mu \star \mathsf{q})(x)$.*

Proof. Set $(\forall i)$ $g_i = \lambda_i \star (f_i + \mu \star \mathsf{q})$. If $\lambda_i = 0$, then $g_i = \iota_{\{0\}}$ and hence $g_i^* = \iota_X$ has full domain. If $\lambda_i > 0$, then using Proposition 3.2(i), Fact 3.4(i), and Proposition 3.3(iv), we see that

$$g_i^* = \big(\lambda_i \star (f_i + \mu \star \mathsf{q})\big)^* = \lambda_i (f_i + \mu \star \mathsf{q})^* = \lambda_i (f_i^* \mathbin{\text{⨥}} (\mu \star \mathsf{q})^*) = \lambda_i (f_i^* \mathbin{\text{⨥}} \mu\, \mathsf{q}); \tag{31}$$

thus, g_i^* also has full domain. Therefore, by Fact 3.4(ii), the epi-sum

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) + \mu \star \mathsf{q} = g_1 \mathbin{\text{⨥}} \cdots \mathbin{\text{⨥}} g_n \tag{32}$$

is *exact*. Since $\operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_n \operatorname{dom} f_n$ by Theorem 4.6, the existence of the y_i is now clear. ■

5 Fenchel Conjugate

In this section, we compute the Fenchel conjugate of the proximal average. The explicit form obtained has several interesting consequences. We begin with a reformulation of Lemma 3.7:

$$\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})+\mu\star\mathfrak{q}\big)^*=\lambda_1(f_1^*\oplus\mu\,\mathfrak{q})+\cdots+\lambda_n(f_n^*\oplus\mu\,\mathfrak{q}). \tag{33}$$

We are now ready for a useful generalization of [6, Theorem 6.1] where $n=2$ and $\mu=1$.

Theorem 5.1 (Fenchel conjugate) $\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big)^*=p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda})$.

Proof. Set

$$g=p_\mu(\boldsymbol{f},\boldsymbol{\lambda})+\mu\star\mathfrak{q}. \tag{34}$$

By (33), we have

$$g^*=\lambda_1(f_1^*\oplus\mu\,\mathfrak{q})+\cdots+\lambda_n(f_n^*\oplus\mu\,\mathfrak{q}). \tag{35}$$

In view of (6), (35), Proposition 3.1(vi), Proposition 3.3(v), and Proposition 3.2(i), we obtain that

$$\begin{aligned}\mathfrak{q}-\mu^{-1}g^*&=\lambda_1\big(\mathfrak{q}-\mu^{-1}(f_1^*\oplus\mu\,\mathfrak{q})\big)+\cdots+\lambda_n\big(\mathfrak{q}-\mu^{-1}(f_1^*\oplus\mu\,\mathfrak{q})\big)\\&=\lambda_1\big(\mathfrak{q}-(\mu^{-1}f_1^*\oplus\mathfrak{q})\big)+\cdots+\lambda_n\big(\mathfrak{q}-(\mu^{-1}f_1^*\oplus\mathfrak{q})\big)\\&=\lambda_1\big((\mu^{-1}f_1^*)^*\oplus\mathfrak{q}\big)+\cdots+\lambda_n\big((\mu^{-1}f_n^*)^*\oplus\mathfrak{q}\big)\\&=\lambda_1\big(\mu^{-1}\star f_1\oplus\mathfrak{q}\big)+\cdots+\lambda_n\big(\mu^{-1}\star f_n\oplus\mathfrak{q}\big).\end{aligned} \tag{36}$$

Consequently, using Fact 3.4(i), Proposition 3.2(i), Proposition 3.2(iii), Proposition 3.2(ii), Proposition 3.3(i), we see that

$$\begin{aligned}\big(\mathfrak{q}-\mu^{-1}g^*\big)^*&=\Big(\lambda_1\big(\mu^{-1}\star f_1\oplus\mathfrak{q}\big)+\cdots+\lambda_n\big(\mu^{-1}\star f_n\oplus\mathfrak{q}\big)\Big)^*\\&=\Big(\lambda_1\big(\mu^{-1}\star f_1\oplus\mathfrak{q}\big)\Big)^*\oplus\cdots\oplus\Big(\lambda_n\big(\mu^{-1}\star f_n\oplus\mathfrak{q}\big)\Big)^*\\&=\lambda_1\star\big(\mu^{-1}\star f_1\oplus\mathfrak{q}\big)^*\oplus\cdots\oplus\lambda_n\star\big(\mu^{-1}\star f_n\oplus\mathfrak{q}\big)^*\\&=\lambda_1\star\big((\mu^{-1}\star f_1)^*+\mathfrak{q}^*\big)\oplus\cdots\oplus\lambda_n\star\big((\mu^{-1}\star f_n)^*+\mathfrak{q}^*\big)\\&=\lambda_1\star\big(\mu^{-1}f_1^*+\mathfrak{q}\big)\oplus\cdots\oplus\lambda_n\star\big(\mu^{-1}f_n^*+\mathfrak{q}\big).\end{aligned} \tag{37}$$

Now Proposition 3.1(vi), Proposition 3.1(ix), and Proposition 3.3(ii) imply that

$$\begin{aligned}\mu\big(\mathfrak{q}-\mu^{-1}g^*\big)^*&=\mu\Big(\lambda_1\star\big(\mu^{-1}(f_1^*+\mu\,\mathfrak{q})\big)\oplus\cdots\oplus\lambda_n\star\big(\mu^{-1}(f_n^*+\mu\,\mathfrak{q})\big)\Big)\\&=\mu\Big(\lambda_1\star\big(\mu^{-1}(f_1^*+\mu\,\mathfrak{q})\big)\Big)\oplus\cdots\oplus\mu\Big(\lambda_n\star\big(\mu^{-1}(f_n^*+\mu\,\mathfrak{q})\big)\Big)\\&=\lambda_1\star(f_1^*+\mu\,\mathfrak{q})\oplus\cdots\oplus\lambda_n\star(f_n^*+\mu\,\mathfrak{q})\\&=\lambda_1\star(f_1^*+\mu^{-1}\star\mathfrak{q})\oplus\cdots\oplus\lambda_n\star(f_n^*+\mu^{-1}\star\mathfrak{q}).\end{aligned} \tag{38}$$

Combining (34), Corollary 3.6, and (38), we conclude that

$$\begin{aligned}\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big)^* &= (g - \mu \star \mathfrak{q})^* \\ &= \mu(\mathfrak{q} - \mu^{-1} g^*)^* - \mu^{-1} \star \mathfrak{q} \\ &= \lambda_1 \star (f_1^* + \mu^{-1} \star \mathfrak{q}) \oplus \cdots \oplus \lambda_n \star (f_n^* + \mu^{-1} \star \mathfrak{q}) - \mu^{-1} \star \mathfrak{q} \\ &= p_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda}),\end{aligned} \tag{39}$$

as claimed. ■

Corollary 5.2 (lower semicontinuity) $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$ *is convex, lower semicontinuous, and proper.*

Proof. Applying Theorem 5.1 twice, we deduce that $(p_\mu(\boldsymbol{f},\boldsymbol{\lambda}))^{**} = (p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda}))^* = p_{(\mu^{-1})^{-1}}(\boldsymbol{f}^{**},\boldsymbol{\lambda}) = p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$. ■

The next result refines the corresponding two-function version [4, Proposition 4.8].

Example 5.3 $p(\boldsymbol{f},\boldsymbol{f}^*) = \mathfrak{q}$.

Proof. Theorem 5.1 readily implies that the $p(\boldsymbol{f},\boldsymbol{f}^*)$ is equal to its conjugate; consequently, it must be equal to $\mathfrak{q}$ by Proposition 3.3(i). ■

Theorem 5.4 (inequalities) $(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* \le p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) \le \lambda_1 f_1 + \cdots + \lambda_n f_n$.

Proof. The right inequality follows from (20) (by setting $y_i = x$). Applying the right inequality to $\boldsymbol{f}^*$ and μ^{-1}, we learn that

$$p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda}) \le \lambda_1 f_1^* + \cdots + \lambda_n f_n^*. \tag{40}$$

Taking the Fenchel conjugate of (40) and utilizing Theorem 5.1, we deduce that $p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \big(p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda})\big)^* \ge \big(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*\big)^*$. ■

Corollary 5.5 (infimum value)

$$\lambda_1 \inf f_1 + \cdots + \lambda_n \inf f_n \le \inf p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) \le \inf(\lambda_1 f_1 + \cdots + \lambda_n f_n). \tag{41}$$

Corollary 5.6 (common minimizers) *Suppose that* $\bigcap_{i:\,\lambda_i>0} \operatorname{argmin}(f_i) \neq \varnothing$. *Then*

$$\min p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \sum_{i:\,\lambda_i>0} \lambda_i \min f_i \quad \textit{and} \quad \operatorname{argmin} p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \bigcap_{i:\,\lambda_i>0} \operatorname{argmin}(f_i). \tag{42}$$

Proof. Combine Theorem 5.4 and Corollary 5.5. ■

6 Moreau Envelope and Proximal Mapping

Definition 6.1 *Let $f \in \Gamma(X)$. The* Moreau envelope *of f with parameter μ is $e_\mu f = f \oplus \mu \star \mathfrak{q}$.*

Observe that

$$e_\mu f = (f^* + \mu\,\mathfrak{q})^*. \tag{43}$$

Theorem 6.2 (Moreau envelope and its Fenchel conjugate)

(i) $e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n$.

(ii) $\big(e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* = \lambda_1 \star (e_\mu f_1)^* \oplus \cdots \oplus \lambda_n \star (e_\mu f_n)^*$.

Proof. Fix $y \in X$ and set $I = \big\{i \in \{1, \ldots, n\} \mid \lambda_i > 0\big\}$. Using (16), we obtain

$$\begin{aligned}
\big(e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)(y) &= \inf_x p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) + \frac{1}{2\mu}\|y - x\|^2 \\
&= \inf_x \inf_{\sum_{i \in I} x_i = x} \sum_{i \in I} \lambda_i \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|x_i/\lambda_i\|^2\Big) + \frac{1}{2\mu}\|y\|^2 - \frac{1}{\mu}\langle x, y\rangle \\
&= \inf_x \inf_{\sum_{i \in I} x_i = x} \sum_{i \in I} \lambda_i \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|x_i/\lambda_i\|^2 + \frac{1}{2\mu}\|y\|^2 - \frac{1}{\mu}\langle x_i/\lambda_i, y\rangle\Big) \\
&= \inf_x \inf_{\sum_{i \in I} x_i = x} \sum_{i \in I} \lambda_i \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|y - x_i/\lambda_i\|^2\Big) \\
&= \inf_{x_i, i \in I} \sum_{i \in I} \lambda_i \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|y - x_i/\lambda_i\|^2\Big) \\
&= \sum_{i \in I} \lambda_i \inf_{x_i} \Big(f_i(x_i/\lambda_i) + \frac{1}{2\mu}\|y - x_i/\lambda_i\|^2\Big) \\
&= \sum_{i \in I} \lambda_i \big(e_\mu f_i\big)(y).
\end{aligned} \tag{44}$$

This implies (i), and (ii) follows by Fenchel conjugation. Alternatively, using Definition 6.1, Proposition 3.2(iii), Theorem 5.1, Proposition 3.3(iv), and Proposition 3.3(ii), one may prove (ii) via $\big(e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* = \big(p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \oplus \mu \star \mathfrak{q}\big)^* = \big(p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})\big)^* + \mu\,\mathfrak{q} = p_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda}) + \mu^{-1} \star \mathfrak{q} = \lambda_1 \star (f_1^* + \mu^{-1} \star \mathfrak{q}) \oplus \cdots \oplus \lambda_n \star (f_n^* + \mu^{-1} \star \mathfrak{q}) = \lambda_1 \star (f_1^* + \mu\,\mathfrak{q}) \oplus \cdots \oplus \lambda_n \star (f_n^* + \mu\,\mathfrak{q}) = \lambda_1 \star (e_\mu f_1)^* \oplus \cdots \oplus \lambda_n \star (e_\mu f_n)^*$, and then deduces (i) by Fenchel conjugation. ■

The following result is well known.

Proposition 6.3 *Let $f \in \Gamma(X)$. Then* $\operatorname{argmin} e_\mu f = \operatorname{argmin} f$.

Proof. $\operatorname{argmin} e_\mu f = \partial(e_\mu f)^*(0) = \partial(f^* + \mu\,\mathfrak{q})(0) = (\partial f^* + \mu\,\mathrm{Id})(0) = \partial f^*(0) = \operatorname{argmin} f$. ■

Corollary 6.4 (minimizers) $\operatorname{argmin} p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \operatorname{argmin}\big(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n\big)$.

Proof. Combine Proposition 6.3 and Theorem 6.2(i). ■

Example 6.5 (least squares solutions) Let $C_1,\ldots,C_n$ be nonempty closed convex subsets of X and suppose that $(\forall i)$ $f_i = \iota_{C_i}$. Then $\operatorname{argmin} p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) = \operatorname{argmin}(\lambda_1 d_{C_1}^2 + \cdots + \lambda_n d_{C_n}^2)$.

Proof. This is a consequence of Corollary 6.4 since $(\forall i)$ $e_\mu f_i = e_\mu \iota_{C_i} = \iota_{C_i} \mathbin{\text{⨦}} \mu \star \mathfrak{q} = \mu^{-1}\iota_{C_i} \mathbin{\text{⨦}} \mu^{-1}\mathfrak{q} = \mu^{-1}(\iota_{C_i} \mathbin{\text{⨦}} \mathfrak{q}) = \mu^{-1}\frac{1}{2}d_{C_i}^2$. ■

Definition 6.6 *Let* $f \in \Gamma(X)$. *The* proximal mapping *of* f *with parameter* μ *is* $P_\mu f = (\operatorname{Id} + \mu\partial f)^{-1}$.

Observe that

$$\mu^{-1}(P_\mu f)^{-1} = \partial f + \mu^{-1}\operatorname{Id}, \tag{45}$$

that

$$P_\mu f = \big(\nabla(f + \mu^{-1}\mathfrak{q})^*\big) \circ (\mu^{-1}\operatorname{Id}), \tag{46}$$

and that

$$(P_\mu f) \circ (\mu \operatorname{Id}) = \nabla(e_{\mu^{-1}}(f^*)). \tag{47}$$

We now show that the proximal mapping of the proximal average is simply the average of the individual proximal mappings. This result, which also explains how the proximal average got its name, was first proved in [6, Theorem 6.1] when $n = 2$ and $\mu = 1$.

Theorem 6.7 (proximal mapping) $P_\mu\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big) = \lambda_1 P_\mu f_1 + \cdots + \lambda_n P_\mu f_n$.

Proof. Theorem 5.1 and Theorem 6.2(i) (the latter applied to $\boldsymbol{f}^*$ and μ^{-1}) show that

$$e_{\mu^{-1}}\big((p_\mu(\boldsymbol{f},\boldsymbol{\lambda}))^*\big) = e_{\mu^{-1}}\big(p_{\mu^{-1}}(\boldsymbol{f}^*,\boldsymbol{\lambda})\big) = \lambda_1 e_{\mu^{-1}}(f_1^*) + \cdots + \lambda_n e_{\mu^{-1}}(f_n^*); \tag{48}$$

in turn, taking gradients yields

$$\nabla\big(e_{\mu^{-1}}\big((p_\mu(\boldsymbol{f},\boldsymbol{\lambda}))^*\big)\big) = \lambda_1\nabla(e_{\mu^{-1}}(f_1^*)) + \cdots + \lambda_n\nabla\big(e_{\mu^{-1}}(f_n^*)\big). \tag{49}$$

Using (47), we see that this is equivalent to

$$\big(P_\mu\big(p_\mu(\boldsymbol{f},\boldsymbol{\lambda})\big)\big) \circ (\mu\operatorname{Id}) = \lambda_1(P_\mu f_1) \circ (\mu\operatorname{Id}) + \cdots + \lambda_n(P_\mu f_n) \circ (\mu\operatorname{Id}). \tag{50}$$

The result follows. ■

7 Subdifferential

Theorem 7.1 (subdifferential) *Let* $(\forall i)\ x_i \in \operatorname{dom} f_i$ *and set* $x = \lambda_1 x_1 + \cdots + \lambda_n x_n$. *Then the following hold.*

(i) *If* $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = (\lambda_1 \star (f_1 + \mu \star \mathsf{q}))(\lambda_1 x_1) + \cdots + (\lambda_n \star (f_n + \mu \star \mathsf{q}))(\lambda_n x_n) - (\mu \star \mathsf{q})(x)$, *then*

$$\partial p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = -\mu^{-1}x + \bigcap_i \partial(\lambda_i \star (f_i + \mu \star \mathsf{q}))(\lambda_i x_i) \tag{51}$$

$$= -\mu^{-1}x + \bigcap_{i:\ \lambda_i>0} \big(\partial f_i(x_i) + \mu^{-1}x_i\big) \tag{52}$$

$$= -\mu^{-1}x + \bigcap_{i:\ \lambda_i>0} \big(\mu^{-1}(P_\mu f_i)^{-1}(x_i)\big). \tag{53}$$

(ii) *If* $\bigcap_{i:\ \lambda_i>0}(P_\mu f_i)^{-1}(x_i) \neq \varnothing$, *then*

$$p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = (\lambda_1 \star (f_1 + \mu \star \mathsf{q}))(\lambda_1 x_1) + \cdots + (\lambda_n \star (f_n + \mu \star j))(\lambda_n x_n) - (\mu \star \mathsf{q})(x). \tag{54}$$

Proof. Set $(\forall i)\ g_i = \lambda_i \star (f_i + \mu \star \mathsf{q})$. Theorem 4.6, Theorem 4.10, and Proposition 3.3(ii) imply that

$$g_1 \mathbin{\text{⨥}} \cdots \mathbin{\text{⨥}} g_n = p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) + \mu \star \mathsf{q} = p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) + \mu^{-1}\,\mathsf{q} \tag{55}$$

is exact on $\operatorname{dom}(g_1 \mathbin{\text{⨥}} \cdots \mathbin{\text{⨥}} g_n) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_n \operatorname{dom} f_n = \operatorname{dom} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$. (i): (51), (52), and (53) follow from Fact 3.8(i), Proposition 3.9, and (45), respectively. (ii): Use Fact 3.8(ii). ■

Corollary 7.2 $(\forall x \in X)\ \bigcap_{i:\ \lambda_i>0} \partial f_i(x) \subseteq \partial p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$.

Proof. Take $x^* \in \bigcap_{i:\ \lambda_i>0} \partial f_i(x)$. Then $(\forall i)\ \lambda_i > 0 \Rightarrow \mu x^* + x \in \mu\partial f_i(x) + x = (P_\mu f_i)^{-1}(x)$. By Theorem 7.1(ii), $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = (\lambda_1 \star (f_1 + \mu \star \mathsf{q}))(\lambda_1 x) + \cdots + (\lambda_n \star (f_n + \mu \star j))(\lambda_n x) - (\mu \star \mathsf{q})(x)$. Using Theorem 7.1(i), we deduce that $x^* = -\mu^{-1}x + \mu^{-1}(\mu x^* + x) \in \partial p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$. ■

For the following results, it will be convenient to write $x = x_1 \oplus \cdots \oplus x_n$ if $x = x_1 + \cdots + x_n$ and $x_i \perp x_j$ for $i \neq j$. We also write $K_1 \oplus \cdots \oplus K_n = \big\{x_1 \oplus \cdots \oplus x_n \mid \text{each } x_i \in K_i \text{ and } x_i \perp x_j \text{ for } i \neq j\big\}$.

Corollary 7.3 *Let* $K_1, \ldots, K_n$ *be nonempty closed convex cones and set* $(\forall i)\ P_i = P_{K_i}$, *the orthogonal projector onto* K_i. *Suppose that*

$$(\forall x = x_1 \oplus \cdots \oplus x_n \in K_1 \oplus \cdots \oplus K_n)(\forall i) \quad P_i x = x_i, \tag{56}$$

that

$$(\forall x \in X) \quad x = P_1 x \oplus \cdots \oplus P_n x, \tag{57}$$

and that $(\forall i)\ f_i = \iota_{K_i}$ *and* $\lambda_i > 0$. *Then*

$$(\forall x \in X) \quad p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \frac{1}{2\mu}\sum_i \frac{(1-\lambda_i)}{\lambda_i}\|P_i x\|^2. \tag{58}$$

Proof. Observe that $(\forall i)\ P_\mu f_i = (\operatorname{Id} + \mu\partial\iota_{K_i})^{-1} = (\operatorname{Id} + \partial\iota_{K_i})^{-1} = P_i$. Take $x \in X$ and set

$$(\forall i) \quad x_i = \tfrac{1}{\lambda_i} P_i x = P_i\big(\tfrac{1}{\lambda_i} x\big). \tag{59}$$

Using (57), we obtain that

$$x = \lambda_1 x_1 \oplus \cdots \oplus \lambda_n x_n. \tag{60}$$

Now set

$$z = x_1 \oplus \cdots \oplus x_n. \tag{61}$$

By (56), we have $(\forall i)\ P_i z = x_i$. Thus $z \in \bigcap_i (P_\mu f_i)^{-1}(x_i)$. Therefore, by (60) and Theorem 7.1(ii),

$$\begin{aligned}
p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) &= (\lambda_1 \star (f_1 + \mu \star \mathfrak{q}))(\lambda_1 x_1) + \cdots + (\lambda_n \star (f_n + \mu \star \mathfrak{q}))(\lambda_n x_n) - (\mu \star \mathfrak{q})(x) \\
&= \mu^{-1}\lambda_1\, \mathfrak{q}(x_1) + \cdots + \mu^{-1}\lambda_n\, \mathfrak{q}(x_n) - \mu^{-1}\, \mathfrak{q}(x) \\
&= \frac{1}{2\mu}\big(\lambda_1\|x_1\|^2 + \cdots + \lambda_n\|x_n\|^2 - \|\lambda_1 x_1 + \cdots + \lambda_n x_n\|^2\big) \\
&= \frac{1}{2\mu}\sum_i \lambda_i(1-\lambda_i)\|x_i\|^2.
\end{aligned} \tag{62}$$

The conclusion thus follows from (59). ■

The following two examples are special cases of Corollary 7.3.

Example 7.4 Let $K_1, \ldots, K_n$ be closed subspaces that are pairwise orthogonal and such that $K_1 \oplus \cdots \oplus K_n = X$ and suppose that $f_i = \iota_{K_i}$. Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \mu^{-1}\sum_i(\lambda_i^{-1} - 1)(\,\mathfrak{q} \circ P_{K_i})$.

Example 7.5 (See also [4, Example 4.9].) Let K be a nonempty closed convex cone in X and let $\lambda \in\,]0,1[$. Then

$$(\forall x \in X) \quad p\big((\iota_K, \iota_{K^\ominus}), (1-\lambda, \lambda)\big)(x) = \frac{1}{2(1-\lambda)\lambda}\big(\lambda^2\|P_K x\|^2 + (1-\lambda)^2\|P_{K^\ominus} x\|^2\big), \tag{63}$$

where $K^\ominus$ is the polar cone of K.

Remark 7.6 We are now in a position to show that the inequalities in Theorem 5.4 can be strict. Suppose that $n = 2$, that $f_1 = \iota_K$ that $f_2 = \iota_{K^\ominus}$, where K is a nonempty closed convex cone in X, and that $\lambda_2 = \lambda \in\,]0,1[$. Using Example 7.5, we see that Theorem 5.4 becomes

$$(\forall x \in X) \quad \iota_X(x) \leq \frac{1}{2(1-\lambda)\lambda}\big(\lambda^2\|P_K x\|^2 + (1-\lambda)^2\|P_{K^\ominus} x\|^2\big) \leq \iota_{\{0\}}(x). \tag{64}$$

The inequalities are strict for every $x \in X \smallsetminus \{0\}$.

Let $f \in \Gamma(X)$. Following [3, Section 5], we say that f is *essentially smooth* if ∂f is at most single-valued and $\operatorname{int}\operatorname{dom} f$ is nonempty, that f is *essentially strictly convex* if f^* is essentially smooth, and that f is *Legendre* if f is both essentially smooth and essentially strictly convex. These notions coincide in our (reflexive) Hilbert space setting with the well known notions of the same name in Euclidean space (see [21, Section 26]).

The next three results extend corresponding results in [4, Section 6] considerably.

Corollary 7.7 (essential smoothness) *Suppose that at least one function f_i is essentially smooth and that $\lambda_i > 0$. Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is essentially smooth.*

Proof. Since f_i is essentially smooth, the set $\operatorname{dom} f_i$ has nonempty interior. Thus $\lambda_i \operatorname{dom} f_i$ and $\operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 \operatorname{dom} f_1 + \cdots + \lambda_n \operatorname{dom} f_n$ (see Theorem 4.6) both have nonempty interiors as well. Now take $x \in \operatorname{dom} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ and let $y_1, \ldots, y_n$ be as in Theorem 4.10, say $(\forall i)\ y_i = \lambda_i x_i$, where $x_i \in \operatorname{dom} f_i$. By Theorem 7.1(i), $\partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) \subseteq -\mu^{-1}x + \partial f_i(x_i) + \mu^{-1}x_i$. Because f_i is essentially smooth, the set $\partial f_i(x_i)$ is either empty or singleton. Thus $\partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x)$ is either empty or singleton. Altogether, $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is essentially smooth. ■

Corollary 7.8 (essential strict convexity) *Suppose that at least one function f_i is essentially strictly convex and that $\lambda_i > 0$. Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is essentially strictly convex.*

Proof. Since f_i is essentially strictly convex, its conjugate f_i^* is essentially smooth. By Corollary 7.7, $p_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda})$ is essentially smooth. Hence $(p_{\mu^{-1}}(\boldsymbol{f}^*, \boldsymbol{\lambda}))^*$ is essentially strictly convex. This last function is equal to $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ (by Theorem 5.1) and the proof is thus complete. ■

Corollary 7.9 (Legendre function) *Suppose that at least one function f_i is essentially smooth and that $\lambda_i > 0$. Furthermore, suppose that at least one function f_j is essentially strictly convex and that $\lambda_j > 0$. (It does not matter whether j and i are identical or distinct.) Then $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is both essentially smooth and essentially strictly convex, i.e., Legendre.*

Proof. Combine Corollary 7.7 and Corollary 7.8. ■

Before we formulate and prove the last result in this section, we briefly return to the Moreau envelope and the proximal mapping. Let $f \in \Gamma(X)$. Applying Proposition 3.3(v) to μf, we readily deduce that (see also [22, Example 11.26(b)])

$$\mu(e_\mu f) + \mu \star (e_{\mu^{-1}}(f^*)) = \mathfrak{q}. \tag{65}$$

Taking gradients and recalling (47) yields $\operatorname{Id} = P_\mu f + \mu(P_{\mu^{-1}}(f^*)) \circ (\mu^{-1} \operatorname{Id})$; equivalently, $\mu \operatorname{Id} = (P_\mu f) \circ (\mu \operatorname{Id}) + \mu P_{\mu^{-1}}(f^*)$ or

$$\operatorname{Id} = \mu^{-1}(P_\mu f) \circ (\mu \operatorname{Id}) + P_{\mu^{-1}}(f^*). \tag{66}$$

The following result generalizes [5, Theorem 4.22], where $n = 2$, $\lambda_1 = \lambda_2 = \frac{1}{2}$, and $\mu = 1$.

Theorem 7.10 *Suppose that $(a, a^*) \in X \times X$ satisfies $a^* \in \partial f_1(a) \cap \cdots \cap \partial f_n(a)$ and that $\{1, 2, \ldots, n\}$ is the disjoint union of two sets of indices I and J. Set $\lambda_J = \sum_{j \in J} \lambda_j$ and suppose that $\lambda_J > 0$. Then for every $z \in a + \big(\bigcap_{i \in I} N_{\operatorname{dom} f_i}(a) \cap \bigcap_{j \in J} N_{\operatorname{dom} f_j^*}(a^*)\big)$, we have*

$$a^* + \mu^{-1}(\lambda_J^{-1} - 1)(z - a) \in \partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(z). \tag{67}$$

Consequently, $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is differentiable on $a + \operatorname{int}\big(\bigcap_{i \in I} N_{\operatorname{dom} f_i}(a) \cap \bigcap_{j \in J} N_{\operatorname{dom} f_j^*}(a^*)\big)$, with gradient $z \mapsto a^* + \mu^{-1}(\lambda_J^{-1} - 1)(z - a)$.

Proof. Let z be as in the conclusion and set $y = z - a$. Fix $i \in I$. Now $a^* \in \partial f_i(a)$ and $\lambda_J^{-1} y \in N_{\operatorname{dom} f_i}(a) = \partial \iota_{\operatorname{dom} f_i}(a) = \partial \iota_{\operatorname{dom} \mu f_i}(a)$. Hence $\mu a^* \in \mu \partial f_i(a) = \partial(\mu f_i)(a)$. Thus $\mu a^* + \lambda_J^{-1} y \in \partial(\mu f_i)(a) + \partial(\iota_{\operatorname{dom} \mu f_i})(a) \subseteq \partial(\mu f_i + \iota_{\operatorname{dom} \mu f_i})(a) = \partial(\mu f_i)(a)$. It follows that

$$(\forall i \in I) \quad a = (P_\mu f_i)(\mu a^* + \lambda_J^{-1} y + a). \tag{68}$$

Next, fix $j \in J$. Then $a + \lambda_J^{-1} y \in \partial f_j^*(a^*)$ and $\mu^{-1} a + \mu^{-1} \lambda_J^{-1} y \in \partial(\mu^{-1} f_j^*)(a^*)$. Using (66), we thus have $a^* = (P_{\mu^{-1}} f_j^*)(\mu^{-1} a + \mu^{-1} \lambda_J^{-1} y + a^*) = \mu^{-1} a + \mu^{-1} \lambda_J^{-1} y + a^* - \mu^{-1}(P_\mu f_j)(a + \lambda_J^{-1} y + \mu a^*)$. Hence

$$(\forall j \in J) \quad a + \lambda_J^{-1} y = (P_\mu f_j)(a + \lambda_J^{-1} y + \mu a^*). \tag{69}$$

Now (68), (69), and Theorem 6.7 imply that

$$a + y = (P_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}))(a + \lambda_J^{-1} y + \mu a^*); \tag{70}$$

equivalently,

$$a^* + \mu^{-1}(\lambda_J^{-1} - 1) y \in \partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(a + y). \tag{71}$$

This verifies (67). Denote the intersection of the n normal cones by N. On $a + \operatorname{int} N$, the mapping $z \mapsto a^* + \mu^{-1}(\lambda_J^{-1} - 1)(z - a)$ is thus a continuous selection of $\partial p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$; therefore, $\nabla p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(z) = a^* + \mu^{-1}(\lambda_J^{-1} - 1)(z - a)$ by [20, Proposition 2.8]. ■

8 Pointwise Limits of the Proximal Average

Proposition 8.1 *Let $f \in \Gamma(X)$. Then $e_{\mu^{-1}}\big(f \circ (\mu \operatorname{Id})\big) = (e_\mu f) \circ (\mu \operatorname{Id})$.*

Proof. For every $x \in X$, we have $e_{\mu^{-1}}\big(f \circ (\mu \operatorname{Id})\big)(x) = \inf_y \big(f(\mu y) + \mu\, \mathfrak{q}(x - y)\big) = \inf_y \big(f(\mu y) + \mu^{-1}\, \mathfrak{q}(\mu x - \mu y)\big) = \inf_z \big(f(z) + \mu^{-1}\, \mathfrak{q}(\mu x - z)\big) = e_\mu f(\mu x)$. ■

Proposition 8.2 [22, Example 11.26(c)] *Let $f\colon X \to [-\infty, +\infty]$. Then*

$$(f + \mu\, \mathfrak{q})^* = (\mu\, \mathfrak{q} - e_{\mu^{-1}} f) \circ (\mu^{-1} \operatorname{Id}). \tag{72}$$

Proof. For every $x^* \in X$, we obtain that

$$\begin{aligned}
(f + \mu\, \mathfrak{q})^*(x^*) &= \sup_x \big(\langle x, x^* \rangle - f(x) - \mu\, \mathfrak{q}(x)\big) \\
&= \sup_x \big(\langle x, x^* \rangle - f(x) - \mu\, \mathfrak{q}(x - \mu^{-1} x^*) + \mu^{-1}\, \mathfrak{q}(x^*) - \langle x, x^* \rangle\big) \\
&= \mu^{-1}\, \mathfrak{q}(x^*) + \sup_x \big(- f(x) - \mu\, \mathfrak{q}(x - \mu^{-1} x^*)\big) \\
&= \mu^{-1}\, \mathfrak{q}(x^*) - \inf_x \big(f(x) + \mu\, \mathfrak{q}(\mu^{-1} x^* - x)\big) \\
&= \mu^{-1}\, \mathfrak{q}(x^*) - (f \mathbin{\Box} \mu\, \mathfrak{q})(\mu^{-1} x^*) \\
&= \mu\, \mathfrak{q}(\mu^{-1} x^*) - (f \mathbin{\Box} \mu^{-1} \star \mathfrak{q})(\mu^{-1} x^*) \\
&= \big(\mu\, \mathfrak{q} - e_{\mu^{-1}} f\big)(\mu^{-1} x^*).
\end{aligned} \tag{73}$$

The result follows. ■

The following alternative expression of the proximal average was discovered by Warren Hare for the case when $n = 2$ and $\mu = 1$.

Theorem 8.3 [9] $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = -e_\mu\big(-(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)\big)$.

Proof. Set $g = -(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)$. Taking the Fenchel conjugate on both sides of (33) leads to $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \big(\lambda_1(f_1^* \,⨣\, \mu\,\mathfrak{q}) + \cdots + \lambda_n(f_n^* \,⨣\, \mu\,\mathfrak{q})\big)^* - \mu \star \mathfrak{q}$. On the other hand, $(\forall i)$ $f_i^* \,⨣\, \mu\,\mathfrak{q} = (f_i + \mu \star \mathfrak{q})^*$ by Fact 3.4(i) and Proposition 3.3(iii). Altogether,

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \big(\lambda_1(f_1 + \mu \star \mathfrak{q})^* + \cdots + \lambda_n(f_n + \mu \star \mathfrak{q})^*\big)^* - \mu \star \mathfrak{q}. \tag{74}$$

Using (74), Proposition 3.3(ii), Proposition 8.2, and Proposition 8.1 we deduce that

$$\begin{aligned}
p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) &= \big(\lambda_1(f_1 + \mu \star \mathfrak{q})^* + \cdots + \lambda_n(f_n + \mu \star \mathfrak{q})^*\big)^* - \mu \star \mathfrak{q} \\
&= \big(\lambda_1(f_1 + \mu^{-1}\,\mathfrak{q})^* + \cdots + \lambda_n(f_n + \mu^{-1}\,\mathfrak{q})^*\big)^* - \mu \star \mathfrak{q}. \\
&= \big(\lambda_1(\mu^{-1}\,\mathfrak{q} - e_\mu f_1) \circ (\mu\,\mathrm{Id}) + \cdots + \lambda_n(\mu^{-1}\,\mathfrak{q} - e_\mu f_n) \circ (\mu\,\mathrm{Id})\big)^* - \mu \star \mathfrak{q} \\
&= \big(\mu\,\mathfrak{q} + g \circ (\mu\,\mathrm{Id})\big)^* - \mu \star \mathfrak{q} \\
&= \big(\mu\,\mathfrak{q} - e_{\mu^{-1}}(g \circ (\mu\,\mathrm{Id}))\big) \circ (\mu^{-1}\,\mathrm{Id}) - \mu \star \mathfrak{q} \\
&= \mu^{-1}\,\mathfrak{q} - \big(e_{\mu^{-1}}(g \circ (\mu\,\mathrm{Id}))\big) \circ (\mu^{-1}\,\mathrm{Id}) - \mu \star \mathfrak{q} \\
&= -\big((e_\mu g) \circ (\mu\,\mathrm{Id})\big) \circ (\mu^{-1}\,\mathrm{Id}) \\
&= -e_\mu g.
\end{aligned} \tag{75}$$

This verifies the result. ■

The *μ-proximal hull* of a function g is defined by $h_\mu g = -e_\mu(-e_\mu g)$; it satisfies $e_\mu g \le h_\mu g \le g$ and $e_\mu(h_\mu g) = e_\mu g$ (see [22, Example 1.44]). Theorem 8.3 shows that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ can be interpreted as some sort of weighted proximal hull of the functions $f_1, \ldots, f_n$. We now turn to the proximal hull of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$.

Corollary 8.4 (proximal hull) $h_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$.

Proof. By Theorem 6.2(i), $e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n$. Hence, using Theorem 8.3, $h_\mu(p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})) = -e_\mu(-e_\mu p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})) = -e_\mu(-\lambda_1 e_\mu f_1 - \cdots - \lambda_n e_\mu f_n) = p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$. Since $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) + \mu \star \mathfrak{q}$ is clearly convex and lower semicontinuous (by Corollary 5.2), the result follows alternatively from [22, Example 11.26(d)]. ■

Let us now determine the pointwise behaviour of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$.

Theorem 8.5 (pointwise limits) *Let $x \in X$. Then the function*

$$]0, +\infty[\; \to \;]-\infty, +\infty] : \mu \mapsto p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x) \quad \textit{is decreasing.} \tag{76}$$

Consequently, $\lim_{\mu\to 0^+} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$ *and* $\lim_{\mu\to+\infty} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$ *exist. In fact,*

$$\lim_{\mu\to 0^+} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \sup_{\mu>0} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \big(\lambda_1 f_1 + \cdots + \lambda_n f_n\big)(x) \tag{77}$$

and

$$\lim_{\mu\to +\infty} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \inf_{\mu>0} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) = \big(\lambda_1 \star f_1 \,\square\!\!\!\!+ \cdots \,\square\!\!\!\!+ \lambda_n \star f_n\big)(x). \tag{78}$$

Proof. The fact that $\mu \mapsto p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x)$ is decreasing follows from (20); consequently, the two limits exist and the supremum/infimum descriptions are clear. Now $e_\mu\big(-(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)\big) \leq -(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)$. Thus, using Theorem 8.3, we deduce that $\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n \leq -e_\mu\big(-(\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n)\big) = p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$. On the other hand, Theorem 5.4 implies that $p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) \leq \lambda_1 f_1 + \cdots + \lambda_n f_n$. Altogether,

$$\lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n \leq p_\mu(\boldsymbol{f},\boldsymbol{\lambda}) \leq \lambda_1 f_1 + \cdots + \lambda_n f_n. \tag{79}$$

It is well known that Moreau envelopes converge pointwise to the underlying function as the parameter approaches 0; see, e.g., [1, Theorem 2.64] or [22, Theorem 1.25 and Theorem 2.26]. Thus $(\forall i)$ $\lim_{\mu\to 0^+} e_\mu f_i = f_i$ pointwise and (77) follows from taking the pointwise limit in (79) at x as $\mu \to 0^+$. Using (20), we deduce that

$$\begin{aligned}
\lim_{\mu\to +\infty} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) &= \inf_{\mu>0} p_\mu(\boldsymbol{f},\boldsymbol{\lambda})(x) \\
&= \inf_{\mu>0} \inf_{\sum \lambda_i y_i = x} \sum \lambda_i f_i(y_i) + \frac{1}{\mu}\Big(\Big(\sum \lambda_i\, \mathsf{q}(y_i)\Big) - \mathsf{q}(x)\Big) \\
&= \inf_{\sum \lambda_i y_i = x} \inf_{\mu>0} \sum \lambda_i f_i(y_i) + \frac{1}{\mu}\Big(\Big(\sum \lambda_i\, \mathsf{q}(y_i)\Big) - \mathsf{q}(x)\Big) \\
&= \inf_{\sum \lambda_i y_i = x} \sum \lambda_i f_i(y_i) \\
&= \inf_{\sum' x_i = x} \sum{}' \lambda_i f_i(x_i/\lambda_i) \\
&= \inf_{\sum' x_i = x} \sum{}' (\lambda_i \star f_i)(x_i) \\
&= \big(\lambda_1 \star f_1 \,\square\!\!\!\!+ \cdots \,\square\!\!\!\!+ \lambda_n \star f_n\big)(x),
\end{aligned} \tag{80}$$

where the indices in the $\sum'$ sums range over all i such that $\lambda_i > 0$. ■

The following nice observation, which is based on the comments of an anonymous referee, builds a bridge to [17].

Remark 8.6 (parallel sums) Suppose that $X = \mathbb{R}^N$, let $A_1, \ldots, A_n$ be positive definite $N \times N$ matrices, and suppose that $(\forall i)$ $f_i(x) = \frac{1}{2}\langle x, A_i x\rangle$, i.e., identify each A_i with its quadratic form. As $\mu \to 0^+$, $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$ converges pointwise to $\lambda_1 f_1 + \cdots + \lambda_n f_n$ and, as $\mu \to +\infty$, $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$ converges pointwise to $\lambda_1 \star f_1 \,\square\!\!\!\!+ \cdots \,\square\!\!\!\!+ \lambda_n \star f_n$. Using [17] (see also [12, Example IV.2.3.8], [14], and [16]), the matrices corresponding to the quadratic forms $\lambda_1 f_1 + \cdots + \lambda_n f_n$, $\lambda_1 \star f_1 \,\square\!\!\!\!+ \cdots \,\square\!\!\!\!+ \lambda_n \star f_n$, and $p_\mu(\boldsymbol{f},\boldsymbol{\lambda})$

are, respectively, the arithmetic average $\lambda_1 A_1 + \cdots + \lambda_n A_n$; the harmonic average $(\lambda_1 A_1^{-1} + \cdots + \lambda_n A_n^{-1})^{-1}$, i.e., the *parallel sum* of the matrices $\lambda_1^{-1} A_1, \ldots, \lambda_n^{-1} A_n$; and $(\lambda_1 (A_1 + \mu^{-1}\,\mathrm{Id})^{-1} + \cdots + \lambda_n (A_n + \mu^{-1}\,\mathrm{Id})^{-1})^{-1} - \mu^{-1}\,\mathrm{Id}$, i.e., a μ^{-1}-shifted version of the harmonic average (in accordance with the comment before Definition 4.1). Note that this provides another proof of Example 4.5 and that the theory for parallel sum extends to matrices that are only positive semidefinite.

9 Epi-Continuity and Epi-Limits of the Proximal Average

We now discuss the convergence behaviour of the proximal average with respect to the epi-topology. Analogously to [4, Section 5], we assume throughout this section that

$$X \text{ is finite-dimensional.} \tag{81}$$

Definition 9.1 (epi-convergence and epi-topology) (See [22, Chapter 6].) *Let g and $(g_k)_{k\in\mathbb{N}}$ be functions from X to $]-\infty, +\infty]$. Then $(g_k)_{k\in\mathbb{N}}$* epi-converges *to g, in symbols $g_k \xrightarrow{\mathrm{e}} g$, if the following hold for every $x \in X$.*

(i) $\big(\forall\, (x_k)_{k\in\mathbb{N}}\big)\ x_k \to x \Rightarrow g(x) \le \varliminf g_k(x_k)$.

(ii) $\big(\exists\, (y_k)_{k\in\mathbb{N}}\big)\ y_k \to x$ *and* $\varlimsup g_k(y_k) \le g(x)$.

The epi-topology *is the topology induced by epi-convergence.*

Fact 9.2 *Let g and $(g_k)_{k\in\mathbb{N}}$ be in $\Gamma(X)$ such that $g_k \xrightarrow{\mathrm{e}} g$, and let h and $(h_k)_{k\in\mathbb{N}}$ be in $\Gamma(X)$ such that $h_k \xrightarrow{\mathrm{e}} h$. Let ρ and $(\rho_k)_{k\in\mathbb{N}}$ be in $[0, +\infty[$ such that $\rho_k \to \rho$ and let $q\colon X \to \mathbb{R}$ be continuous. Then the following hold.*

(i) $g_k \pm q \xrightarrow{\mathrm{e}} g \pm q$.

(ii) $\rho > 0 \Rightarrow \rho_k g_k \xrightarrow{\mathrm{e}} \rho g$.

(iii) $\rho = 0$ *and* $\operatorname{dom} g = X \Rightarrow \rho_k g_k \xrightarrow{\mathrm{e}} \rho g$.

(iv) $g_k^* \xrightarrow{\mathrm{e}} g^*$.

(v) $0 \in \operatorname{int}(\operatorname{dom} g - \operatorname{dom} h) \Rightarrow g_k + h_k \xrightarrow{\mathrm{e}} g + h$.

Proof. (i): See [22, Exercise 7.8(a)]. (ii): See [22, Exercise 7.8(d)]. (iii): See [4] or verify this directly. (iv): See [22, Theorem 11.34]. (v): See [22, Exercise 7.47(b)]. ■

Lemma 9.3 *Let $g_1, \ldots, g_n, h$ be in $\Gamma(X)$ and let $(g_{1,k})_{k\in\mathbb{N}}, \ldots, (g_{n,k})_{k\in\mathbb{N}}, (h_k)_{k\in\mathbb{N}}$ be sequences in $\Gamma(X)$ such that $(\forall i)$ $g_{i,k} \xrightarrow{\mathrm{e}} g_i$ and $h_k \xrightarrow{\mathrm{e}} h$. Let ρ and $(\rho_k)_{k\in\mathbb{N}}$ be in $[0, +\infty[$ such that $\rho_k \to \rho$. Suppose that* $\operatorname{dom} g_1^* = \cdots = \operatorname{dom} g_{n-1}^* = \operatorname{dom} h^* = X$ *and that* $(\forall i \in \{1, \ldots, n-1\})(\forall k)$ $\operatorname{dom} g_{i,k}^* = X$*. Then the following hold.*

(i) $g_{1,k} \dotplus \cdots \dotplus g_{n,k} \xrightarrow{\mathrm{e}} g_1 \dotplus \cdots \dotplus g_n$.

(ii) $\rho_k \star h_k \xrightarrow{\mathrm{e}} \rho \star h$.

Proof. (i): Fact 9.2(iv)&(v) imply that $g^*_{1,k} + \cdots + g^*_{n,k} \xrightarrow{\mathrm{e}} g^*_1 + \cdots + g^*_n$. Using Fact 9.2(iv), we see that $(g^*_{1,k} + \cdots + g^*_{n,k})^* \xrightarrow{\mathrm{e}} (g^*_1 + \cdots + g^*_n)^*$, which is equivalent to $g_{1,k} \dotplus \cdots \dotplus g_{n,k} \xrightarrow{\mathrm{e}} g_1 \dotplus \cdots \dotplus g_n$ by Fact 3.4(i). (ii): Fact 9.2(ii)–(iv) imply that $\rho_k h^*_k \xrightarrow{\mathrm{e}} \rho h^*$. Using Fact 9.2(iv) once more, we deduce that $(\rho_k h^*_k)^* \xrightarrow{\mathrm{e}} (\rho h^*)^*$, which is the same as the conclusion in view of Proposition 3.2(i). ■

Remark 9.4 Using the horizon functions associated with $g_1, \ldots, g_n$ and [22, Proposition 7.56], one may obtain a stronger version of Lemma 9.3 where the assumption on the functions $g^*_{i,k}$ is less restrictive; however, this is not needed in the sequel.

The next result extends [4, Theorem 5.4].

Theorem 9.5 (epi-continuity of the proximal average) *Let $(f_{i,k})_{k\in\mathbb{N}}$ be sequences in $\Gamma(X)$ such that $(\forall i)$ $f_{i,k} \xrightarrow{\mathrm{e}} f_i$, let $(\lambda_{i,k})_{k\in\mathbb{N}}$ be sequences in $[0,1]$ such that $(\forall k)$ $\sum_i \lambda_{i,k} = 1$ and $(\forall i)$ $\lambda_{i,k} \to \lambda_i$, and let $(\mu_k)_{k\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $\mu_k \to \mu$. Then*

$$p_{\mu_k}\big((f_{1,k}, \ldots, f_{n,k}), (\lambda_{1,k}, \ldots, \lambda_{n,k})\big) \xrightarrow{\mathrm{e}} p_\mu\big((f_1, \ldots, f_n), (\lambda_1, \ldots, \lambda_n)\big) = p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}). \tag{82}$$

Proof. By Theorem 9.3(ii),

$$\mu_k \star \mathfrak{q} \xrightarrow{\mathrm{e}} \mu \star \mathfrak{q}. \tag{83}$$

Furthermore,

$$(\forall i) \quad f_{i,k} + \mu_k \star \mathfrak{q} \xrightarrow{\mathrm{e}} f_i + \mu \star \mathfrak{q} \tag{84}$$

by Fact 9.2(v) because $(\mu \star \mathfrak{q})^* = \mu\,\mathfrak{q}$ has full domain. Using (84), Lemma 9.3(ii), and the fact that $(\forall i)$ $(f_i + \mu \star \mathfrak{q})^* = (f^*_i \dotplus (\mu \star \mathfrak{q})^*)^{**} = (f^*_i \dotplus \mu\,\mathfrak{q})^{**}$ has full domain (and similarly for $(f_{i,k} + \mu_k \star \mathfrak{q})^*$), we deduce that

$$(\forall i) \quad \lambda_{i,k} \star (f_{i,k} + \mu_k \star \mathfrak{q}) \xrightarrow{\mathrm{e}} \lambda_i \star (f_i + \mu \star \mathfrak{q}). \tag{85}$$

Since $(\forall i)$ $\big(\lambda_i \star (f_i + \mu \star \mathfrak{q})\big)^* = \lambda_i (f_i + \mu \star \mathfrak{q})^* = \lambda_i (f^*_i \dotplus \mu \star \mathfrak{q})$ has full domain (and similarly for $(\lambda_{i,k} \star (f_{i,k} + \mu_k \star \mathfrak{q}))^*$), (85) and Lemma 9.3(i) yield

$$\lambda_{1,k} \star (f_{1,k} + \mu_k \star \mathfrak{q}) \dotplus \cdots \dotplus \lambda_{n,k} \star (f_{n,k} + \mu_k \star \mathfrak{q}) \xrightarrow{\mathrm{e}} \lambda_1 \star (f_1 + \mu \star \mathfrak{q}) \dotplus \cdots \dotplus \lambda_n \star (f_n + \mu \star \mathfrak{q}). \tag{86}$$

In turn, (83), (86) and Fact 9.2(i) imply (82). ■

We now describe the behaviour of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ when μ approaches either 0 or $+\infty$ while $\boldsymbol{f}$ and $\boldsymbol{\lambda}$ are fixed.

Corollary 9.6 *$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \xrightarrow{\mathrm{e}} \lambda_1 f_1 + \cdots + \lambda_n f_n$ as $\mu \to 0^+$, and $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \xrightarrow{\mathrm{e}} \mathrm{cl}(\lambda_1 \star f_1 \dotplus \cdots \dotplus \lambda_n \star f_n)$ as $\mu \to +\infty$.*

Proof. Theorem 8.5 shows that $\mu \mapsto p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ is pointwise increasing. In view of (77) and the lower semicontinuity of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ (see Corollary 5.2), an application of [22, Proposition 7.4(d)] yields that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \xrightarrow{e} \lambda_1 f_1 + \cdots + \lambda_n f_n$ as $\mu \to 0^+$. Combining (78) with [22, Proposition 7.4(e)], we deduce similarly that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \xrightarrow{e} \operatorname{cl}(\lambda_1 \star f_1 \mathbin{\dot{+}} \cdots \mathbin{\dot{+}} \lambda_n \star f_n)$ as $\mu \to +\infty$. ■

Corollary 9.6 and (77) show that as $\mu \to 0^+$, the pointwise and epigraphical limits of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ coincide. When $\mu \to +\infty$, the pointwise and epigraphical limits of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ may differ as we illustrate next.

Example 9.7 Suppose that $X = \mathbb{R}^2$, that $n = 2$, that $\lambda_1 > 0$, that $\lambda_2 > 0$, that $f_1 = \iota_{C_1}$, and that $f_2 = \iota_{C_2}$, where C_1 and C_2 are nonempty closed convex subsets of X such that $\lambda_1 C_1 + \lambda_2 C_2$ is not closed. Concretely, we may let C_1 and C_2 be the epigraphs of $x \mapsto \exp(x)$ and $x \mapsto \exp(-x)$, respectively. Then the pointwise limit (see (78))

$$\lim_{\mu \to +\infty} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) = \lambda_1 \star f_1 \mathbin{\dot{+}} \lambda_2 \star f_2 = \iota_{\lambda_1 C_1 + \lambda_2 C_2} \tag{87}$$

is not lower semicontinuous, and hence different from the epigraphical limit (see Corollary 9.6) $\operatorname{cl}(\lambda_1 \star f_1 \mathbin{\dot{+}} \lambda_2 \star f_2)$, which is the indicator function of the closure of $\lambda_1 C_1 + \lambda_2 C_2$.

We now show that the limiting behaviour as $\mu \to +\infty$ cannot be obtained by conjugation.

Example 9.8 Suppose that $X = \mathbb{R}^2$, that $n = 2$, that $f_1 \colon (x, y) \mapsto -x + \iota_{\{0\}}(y)$, that $f_2 \colon (x, y) \mapsto x + \iota_{\{0\}}(y)$, that $\lambda_1 > 0$, and that $\lambda_2 > 0$. Now fix $(x, y) \in \mathbb{R}^2$. Using (16) and some calculus, we calculate

$$p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x, y) = (\lambda_2 - \lambda_1)x + \iota_{\{0\}}(y) - 2\mu\lambda_1\lambda_2 = (\lambda_1 f_1 + \lambda_2 f_2)(x, y) - 2\mu\lambda_1\lambda_2. \tag{88}$$

Letting $\mu \to 0^+$ in (88) and in accordance with (77), we observe that $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda}) \to \lambda_1 f_1 + \lambda_2 f_2$ pointwise. Recalling (78) and letting $\mu \to +\infty$ in (88), we see that

$$(\lambda_1 \star f_1 \mathbin{\dot{+}} \lambda_2 \star f_2)(x, y) = \lim_{\mu \to +\infty} p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})(x, y) = \begin{cases} -\infty, & \text{if } y = 0; \\ +\infty, & \text{if } y \neq 0. \end{cases} \tag{89}$$

Since $f_1^*(x, y) = \iota_{\{-1\}}(x)$ and $f_2^*(x, y) = \iota_{\{1\}}(x)$, we have $\operatorname{dom}(f_1^*) \cap \operatorname{dom}(f_2^*) = \varnothing$ and thus $\lambda_1 f_1^* + \lambda_2 f_2^* \equiv +\infty$. Altogether,

$$\lambda_1 \star f_1 \mathbin{\dot{+}} \lambda_2 \star f_2 \neq \big(\lambda_1 f_1^* + \lambda_2 f_2^*\big)^* \equiv -\infty. \tag{90}$$

Therefore, due to the absence of a constraint qualification on f_1^* and f_2^*, the epigraphical convergence of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ to the epigraphical average of f_1 and f_2 as $\mu \to +\infty$ could not have been obtained by conjugating the epigraphical convergence of $p_{\mu^{-1}}(f_1^*, f_2^*, \lambda_1, \lambda_2)$ to $\lambda_1 f_1^* + \lambda_2 f_2^*$ as $\mu \to +\infty$.

In the presence of a constraint qualification, we can use the proximal average to construct a homotopic curve with very nice properties.

Remark 9.9 (epigraphical and arithmetic averages are homotopic) Suppose that $\operatorname{int}\operatorname{dom} f_1^* \cap \cdots \cap \operatorname{int}\operatorname{dom} f_{n-1}^* \cap \operatorname{dom} f_n^* \neq \varnothing$. By Fact 3.4(i) and Proposition 3.2(i), we have $(\lambda_1 f_1^* + \cdots + \lambda_n f_n^*)^* = \lambda_1 \star f_1 \,\hat{+}\, \cdots \,\hat{+}\, \lambda_n \star f_n$. Therefore,

$$\operatorname{cl}(\lambda_1 \star f_1 \,\hat{+}\, \cdots \,\hat{+}\, \lambda_n \star f_n) = \lambda_1 \star f_1 \,\hat{+}\, \cdots \,\hat{+}\, \lambda_n \star f_n \tag{91}$$

and hence the pointwise and epigraphical limits of $p_\mu(\boldsymbol{f}, \boldsymbol{\lambda})$ as either $\mu \to 0^+$ or $\mu \to +\infty$ coincide by Theorem 8.5 and Corollary 9.6. Now set

$$(\forall \rho \in [0,1]) \quad q_\rho \colon x \mapsto \begin{cases} (\lambda_1 f_1 + \cdots + \lambda_n f_n)(x), & \text{if } \rho = 0; \\ p_{\tan(\rho\pi/2)}(\boldsymbol{f}, \boldsymbol{\lambda})(x), & \text{if } 0 < \rho < 1; \\ (\lambda_1 \star f_1 \,\hat{+}\, \cdots \,\hat{+}\, \lambda_n \star f_n)(x), & \text{if } \rho = 1. \end{cases} \tag{92}$$

Then Theorem 8.5, Corollary 9.5, and Corollary 9.6 show that $(q_\rho)_{\rho \in [0,1]}$ is a decreasing, pointwise convergent, homotopic (with respect to the epi-topology) curve between the arithmetic average $\lambda_1 f_1 + \cdots + \lambda_n f_n$ and the epigraphical average $\lambda_1 \star f_1 \,\hat{+}\, \cdots \,\hat{+}\, \lambda_n \star f_n$.

Acknowledgment

We wish to thank Prof. J.-B. Hiriart-Urruty for lending us his copy of [16] and two anonymous referees for their comments. Heinz Bauschke was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. Yves Lucet was partially supported by the Natural Sciences and Engineering Research Council of Canada. Xianfu Wang was partially supported by the Natural Sciences and Engineering Research Council of Canada.

References

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

[2] S. Bartz, H. H. Bauschke, J. M. Borwein, S. Reich, and X. Wang, "Fitzpatrick functions, cyclic monotonicity, and Rockafellar's antiderivative", *Nonlinear Analysis*, vol. 66, pp. 1198–1223, 2007.

[3] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, "Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces", *Communications in Contemporary Mathematics*, vol. 3, pp. 615–647, 2001.

[4] H. H. Bauschke, Y. Lucet, and M. Trienis, "How to transform one convex function continuously into another", *SIAM Review*, vol. 50, pp. 115–132, 2008.

[5] H. H. Bauschke, Y. Lucet, and X. Wang, "Primal-dual symmetric antiderivatives for cyclically monotone operators", *SIAM Journal on Control and Optimization*, vol. 46, pp. 2031–2051, 2007.

[6] H. H. Bauschke, E. Matoušková, and S. Reich, "Projection and proximal point methods: Convergence results and counterexamples", *Nonlinear Analysis*, vol. 56, pp. 715–738, 2004.

[7] H. H. Bauschke and X. Wang, "The kernel average for two convex functions and its applications to the extension and representation of monotone operators", to appear in *Transactions of the American Mathematical Society*.

[8] H. Brézis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland, 1973.

[9] W. L. Hare, personal communication 2006.

[10] W. L. Hare, "The proximal average of nonconvex functions: a proximal stability perspective", preprint, November 2007.

[11] J.-B. Hiriart-Urruty, "A general formula on the conjugate of the difference of functions", *Canadian Mathematical Bulletin*, vol. 29, pp. 482–485, 1986.

[12] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*, Springer, New York, 1996.

[13] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms II*, Springer, New York, 1996.

[14] J.-B. Hiriart-Urruty and M.-L. Mazure, "Formulations variationnelles de l'addition parallèle et de la soustraction parallèlle d'opérateurs semi-définis positifs", *Comptes Rendus des Séances de l'Académie des Sciences. Série I. Mathématique*, vol. 302, pp. 527–530, 1986.

[15] Y. Lucet, H. H. Bauschke, and M. Trienis, "The piecewise linear-quadratic model for computational convex analysis", to appear in *Computational Optimization and Applications*.

[16] M.-L. Mazure, *Analyse variationnelle des formes quadratriques convexes*, Thèse de doctorat de l'Université Paul Sabatier, Toulouse, France, 1986.

[17] M.-L. Mazure, "L'addition parallèle d'opérateurs interprétée comme inf-convolution de formes quadratiques convexes", *RAIRO Modélisation Mathématique et Analyse Numérique*, vol. 20, pp. 497–515, 1986.

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

[19] J. J. Moreau, "Proximité et dualité dans un espace hilbertien", *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[20] R. R. Phelps, *Convex Functions, Monotone Operators, and Differentiability* (second edition), Lecture Notes in Mathematics, vol. 1364, Springer-Verlag, 1993.

[21] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

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

[23] S. Simons, *Minimax and Monotonicity*, Lecture Notes in Mathematics, vol. 1693, Springer-Verlag, 1998.

[24] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing, 2002.