The proximal average for saddle functions and its symmetry properties with respect to partial and saddle conjugacy

Rafal Goebel*

December 3, 2009

Abstract

The concept of the proximal average for convex functions is extended to saddle functions. Self-duality of the proximal average is established with respect to partial conjugacy, which pairs a convex function with a saddle function, and saddle function conjugacy, which pairs a saddle function with a saddle function.

1 Introduction

The proximal average of convex functions is an operation that produces a proper, lower semicontinuous, and convex functions from a collection of proper, lower semicontinuous, and convex functions. It was introduced in [4] and studied in detail, in a broader framework, in [2]. For more intuition behind the concept, see [3]. For an extension to non-convex but prox-bounded functions, see [6]. For further references, see [2].

The proximal average avoids the pitfalls present when the arithmetic average of convex functions with disjoint effective domains is considered, but does recover the arithmetic average in the limit, when a certain parameter is driven to 0. A striking feature of the proximal average of convex functions is its self-duality with respect to convex conjugacy: the convex conjugate of the proximal average of a collection of convex functions is the proximal average of the convex conjugates of the functions in the collection.

This paper extends the concept of the proximal average from the setting of convex functions to the setting of saddle functions, in a way that preserves the self-duality. More specifically, the proximal average turns out to be self-dual with respect to partial conjugacy, which pairs a convex function with a saddle function, and with respect to saddle function conjugacy, which pairs a saddle function with its conjugate saddle function. The self-duality should provide a tool for further study of the proximal average of saddle functions, by enabling translation of the properties of the proximal average of convex function via partial conjugacy.

2 Background

2.1 Proximal average for convex functions

A convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = [-\infty, \infty]$, is *proper* if $f(x) \neq -\infty$ for all $x \in \mathbb{R}^n$ and $f(x) < \infty$ for some $x \in \mathbb{R}^n$. Given a proper, lower semicontinuous (lsc), and convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$,

*Department of Mathematics and Statistics, Loyola University Chicago, Chicago, IL 60626, USA. Email: *rgoebel1@luc.edu*

its *conjugate function* $f^*: \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined by

$$f^*(p) = \sup_{x \in \mathbf{R}^n} \{p \cdot x - f(x)\}.$$

The function f^* is itself proper, lsc, and convex, and $(f^*)^* = f$. For more background on convex functions, consult [9].

Consider proper, lsc, and convex functions $f_1, f_2, \ldots, f_r : \mathbb{R}^n \to \overline{\mathbb{R}}$, numbers $\lambda_1, \lambda_2, \ldots, \lambda_r > 0$ such that $\sum_{i=1}^r \lambda_i = 1$, and $\mu > 0$. Let $\mathbf{f} = (f_1, f_2, \ldots, f_r)$, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)$. The *$\lambda$-weighted proximal average of* $\mathbf{f}$ *with parameter* μ is

$$\mathcal{P}_\mu(\lambda, \mathbf{f}) = \lambda_1 \star (f_1 + \mu \star j) \# \lambda_2 \star (f_2 + \mu \star j) \# \ldots \# \lambda_r \star (f_r + \mu \star j) - \mu j, \tag{1}$$

where $\star$ is the operation of *epi-multiplication*, i.e., given $\alpha > 0$, $\phi : \mathbb{R}^n \to \overline{\mathbb{R}}$,

$$\alpha \star \phi(x) = \alpha \phi(x/\alpha)$$

for all $x \in \mathbb{R}^n$; $\#$ is the operation of *epi-addition* or *inf-convolution*, i.e., given $\phi_1, \phi_2, \ldots, \phi_r : \mathbb{R}^n \to \overline{\mathbb{R}}$,

$$\phi_1 \# \phi_2 \# \ldots \#_r(x) = \inf \{\phi_1(x_1) + \phi_2(x_2) + \cdots + \phi_r(x_r) \mid x_1 + x_2 + \cdots + x_r = x\}$$

for all $x \in \mathbb{R}^n$; and $j(x) = \frac{1}{2}|x|^2$ for all $x \in \mathbb{R}^n$. For more background on the operations of epi-multiplication and inf-convolution, consult [11]. Regarding the proximal average, Theorem 5.1 and Corollary 5.2 in [2] state this:

Theorem 2.1 *For proper, lsc, and convex functions $f_1, f_2, \ldots, f_r : \mathbb{R}^n \to \overline{\mathbb{R}}$, numbers $\lambda_1, \lambda_2, \ldots, \lambda_r > 0$ such that $\sum_{i=1}^r \lambda_i = 1$, and $\mu > 0$, the function $\mathcal{P}_\mu(\lambda, \mathbf{f})$ is a proper, lsc, and convex function. Furthermore,*

$$(\mathcal{P}_\mu(\lambda, \mathbf{f}))^* = \mathcal{P}_{\mu^{-1}}(\lambda, \mathbf{f}^*),$$

where $\mathbf{f}^* = (f_1^*, f_2^*, \ldots, f_r^*)$.

Further properties, for example the convergence of $\mathcal{P}_\mu(\lambda, \mathbf{f})$ to the arithmetical average of functions f_i when $\mu \to 0$ and to the epi-graphical average when $\mu \to \infty$, can be found in [2].

2.2 Saddle functions

In this article, a *saddle function* is a function $h : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ such that $h(x, y)$ is convex in x for each fixed y and concave in y for each fixed x. A saddle function h is *proper* and *closed* if its *convex parent* $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ and its *concave parent* $g : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$, obtained from h via partial conjugacy formulas

$$f(x, q) = \sup_{y \in \mathbf{R}^n} \{h(x, y) + \langle q, y\rangle\}, \quad g(p, y) = \inf_{x \in \mathbf{R}^m} \{h(x, y) - \langle p, x\rangle\} \tag{2}$$

are such that f and $-g$ are proper convex functions conjugate to each other, that is,

$$-g(p, y) = \sup_{x \in \mathbf{R}^m, q \in \mathbf{R}^n} \{\langle p, x\rangle + \langle y, q\rangle - f(x, q)\}. \tag{3}$$

An alternative definition of a closed saddle function, one that relies on lower and upper semicontinuous closures, can be found in [11, Section 34]. The *equivalence class* $[h]$ containing a proper and closed saddle function h consists of all proper and closed saddle functions that have the same parents as h. This equivalence class can be also described as the set of all saddle functions h' such that $\underline{h} \le h' \le \overline{h}$, where the lowest element $\underline{h}$ and the highest elements $\overline{h}$ of the class are given, respectively, by

$$\underline{h}(x,y) = \sup_{p\in \mathbf{R}^m} \{g(p,y) + \langle x,p\rangle\}, \quad \overline{h}(x,y) = \inf_{q\in \mathbf{R}^n} \{f(x,q) - \langle y,q\rangle\}. \tag{4}$$

Each of the two formulas above yields a proper and closed saddle function for every proper and upper semicontinuous function g and every proper and lower semicontinuous function f, respectively. For a proper and closed saddle function h, its *domain* $\operatorname{dom} h$ is the product set $C \times D$, where $C = \{x \in I\!\!R^m \mid h(x,y) < \infty \ \forall y \in I\!\!R^n\}$, $D = \{y \in I\!\!R^n \mid h(x,y) > -\infty \ \forall x \in I\!\!R^m\}$. Proper and closed saddle functions from the same equivalence class have the same domains.

Given an equivalence class $[h]$ of proper and closed saddle functions, the class conjugate to it (in the saddle sense), denoted $[h^*]$, has the lowest and the greatest elements given by

$$\begin{aligned} \underline{h^*}(p,q) &= \sup_{x\in \mathbf{R}^m} \inf_{y\in \mathbf{R}^n} \{\langle p,x\rangle + \langle q,y\rangle - h(x,y)\}, \\ \overline{h^*}(p,q) &= \inf_{y\in \mathbf{R}^n} \sup_{x\in \mathbf{R}^m} \{\langle p,x\rangle + \langle q,y\rangle - h(x,y)\}, \end{aligned} \tag{5}$$

where h is any function in $[h]$. In other words, the class conjugate to $[h]$ comes from a convex parent $(p,y) \mapsto -g(p,-y)$ and a concave parent $(x,q) \mapsto -f(x,-q)$. For details, see [8], [9], and for the infinite-dimensional case, [10] and [1].

For a simple example of a saddle function, consider $h(x,y) = aj(x) - bj(y)$ for $a, b > 0$. The equivalence class $[h]$ consists of a single element h. Then the convex and concave parents of h are given, respectively, by $f(x,q) = aj(x) + b^{-1}j(q)$ and $g(p,y) = -a^{-1}j(p) - bj(y)$. The class conjugate to h consists of a single element as well, given by $h^*(p,q) = a^{-1}j(x) - b^{-1}j(q)$. All the calculations follow simply from the fact that, for convex conjugacy, $(aj)^* = a^{-1}j$. In particular, $h^* = h$ if $a = b = 1$. A more interesting example of a saddle function is given by $h(x,y) = x \cdot y$, in the case of $m = n$. The convex and concave parents of h are given, respectively, by $f(x,q) = \delta_0(x+q)$ and $g(p,y) = -\delta_0(y-p)$, where $\delta_0 : I\!\!R^n \to \overline{I\!\!R}$ is given by $\delta_0(0) = 0$, $\delta_0(x) = \infty$ if $x \neq 0$. The saddle conjugate of h is given by $h^*(p,q) = p \cdot q$, so $h = h^*$.

The operation that generalizes the epi-addition, or inf-convolution, of convex functions to the setting of saddle functions is that of extremal convolution. Following [7], we say that the *extremal convolution* of $[h_1], [h_2], \dots, [h_r]$ is well-defined if all the convex-concave functions of the form

$$(x,y) \mapsto \sup_{\sum y_i = y, y_i \in D_i} \ \inf_{\sum x_i = x, x_i \in \mathbf{R}^m} \sum h_i(x_i, y_i), \tag{6}$$

and of the form

$$(x,y) \mapsto \inf_{\sum x_i = x, x_i \in C_i} \ \sup_{\sum y_i = y, y_i \in \mathbf{R}^n} \sum h_i(x_i, y_i) \tag{7}$$

where $h_i \in [h_i]$ and $C_i \times D_i = \operatorname{dom} h_i$, $i = 1, 2, \dots, r$, belong to a single equivalence class. That equivalence class is then denoted $[h_1 \# h_2 \# \dots \# h_r]$. A sufficient condition for the extremal convolution of $[h_1], [h_2], \dots, [h_r]$ to be a well-defined class of proper and closed saddle functions is that $[h_1^*], [h_2^*], \dots, [h_r^*]$ be finite-valued; see [7, Theorem 2]. In such a case, replacing h_i by $\underline{h_i}$ in (6) and h_i by $\overline{h_i}$ in (7) yields, respectively, the least and the greatest elements of $[h_1 \# h_2 \# \dots \# h_r]$; see [7, Theorem 3].

3 Proximal average for saddle functions

Throughout this section, let $[h_1], [h_2], \ldots, [h_r]$ be equivalence classes of proper and closed saddle functions; let $\lambda_1, \lambda_2, \ldots, \lambda_r > 0$ be such that $\sum_{i=1}^r \lambda_i = 1$; and let $\mu, \eta > 0$. For notational convenience, we will let $j_x, j_y : I\!R^m \times I\!R^n \to \overline{I\!R}$ be given by $j_x(x,y) = j(x)$ and $j_y(x,y) = j(y)$. We will also write $\mathbf{h}$ for $([h_1], [h_2], \ldots, [h_r])$.

3.1 Definition and main results

Extending the proximal average to the setting of saddle functions requires replacing the inf-convolution in (1) by the extremal convolution. In (1), the convex functions are augmented by a quadratic term before the inf-convolution is taken. A similar procedure, in the saddle setting, ensures that the extremal convolution is well-defined.

Lemma 3.1 *Let $\widetilde{h}_i = h_i + \mu \star j_x - \eta \star j_y$, $i = 1, 2, \ldots, r$. Then the extremal convolution*

$$\left[\left(\lambda_1 \star \widetilde{h}_1\right) \# \left(\lambda_2 \star \widetilde{h}_2\right) \# \ldots \# \left(\lambda_r \star \widetilde{h}_r\right)\right]$$

is well-defined.

Proof. For $i = 1, 2, \ldots, r$, given any choice of $h_i \in [h_i]$, the function $\widetilde{h}_i = h_i + \mu \star j_x - \eta \star j_y$ is strongly convex in x, strongly concave is y, and hence $\widetilde{h}_i^*$ is finite-valued; see for example [9, Theorem 37.3]. The same conclusions apply to $\lambda_i \star \widetilde{h}_i$. Now, [7, Theorem 2] finishes the proof. □

The *proximal average* of $[h_1], [h_2], \ldots, [h_r]$, with coefficients $\lambda_1, \lambda_2, \ldots, \lambda_r$ and parameters $\mu, \eta > 0$, is *well-defined* if all convex-concave functions on $I\!R^m \times I\!R^n$ of the form

$$h - \mu \star j_x + \eta \star j_y, \qquad \text{where} \begin{cases} h \in \left[\left(\lambda_1 \star \widetilde{h}_1\right) \# \left(\lambda_2 \star \widetilde{h}_2\right) \# \ldots \# \left(\lambda_r \star \widetilde{h}_r\right)\right], \\ \widetilde{h}_i = h_i + \mu \star j_x - \eta \star j_y, \ h_i \in [h_i], \ i = 1, 2, \ldots, r, \end{cases} \tag{8}$$

belong to a single equivalence class. In such a case, the proximal average of $[h_1], [h_2], \ldots, [h_r]$ will be represented as

$$\left[\mathcal{P}^{\cup\cap}_{\mu,\eta}(\lambda, \mathbf{h})\right] = \left[\left(\lambda_1 \star \widetilde{h}_1\right) \# \left(\lambda_2 \star \widetilde{h}_2\right) \# \ldots \# \left(\lambda_r \star \widetilde{h}_r\right) - \mu \star j_x + \eta \star j_y\right].$$

Theorem 3.2 *The proximal average of $[h_1], [h_2], \ldots, [h_r]$, with coefficients $\lambda_1, \lambda_2, \ldots, \lambda_r$ and parameters $\mu, \eta > 0$, is well-defined.*

Theorem 3.2 is proved, along with Theorem 3.3, in the next section, by showing that the least and the greatest saddle functions of the form (8) are obtained via partial conjugacy, as in (4), from a pair of functions, one convex and one concave, related to one another as in (3). The general theory of saddle functions then ensures that all saddle functions of the form (8) are equivalent.

In fact, the pair of proper, lsc, and convex functions mentioned above, the concave and convex parents of $\left[\mathcal{P}^{\cup\cap}_{\mu,\eta}(\lambda, \mathbf{h})\right]$, turn out to be appropriately understood proximal averages of the convex and the concave parents of $[h_1], [h_2], \ldots, [h_r]$. More precisely, let f_i and g_i be, respectively, the convex and the concave parent of $[h_i]$, $i = 1, 2, \ldots, r$. Let $\mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda, \mathbf{f})$ be the proximal average of the convex functions $f_1, f_2, \ldots, f_r$ with coefficients $\lambda_1, \lambda_2, \ldots, \lambda_r$ and parameters μ, η^{-1}, defined by

$$\mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda, \mathbf{f}) = \left(\lambda_1 \star \widehat{f}_1\right) \# \left(\lambda_2 \star \widehat{f}_2\right) \# \ldots \# \left(\lambda_r \star \widehat{f}_r\right) - \mu \star j_x - \frac{1}{\eta} \star j_y \tag{9}$$

where $\widehat{f}_i = f_i + \mu \star j_x + \frac{1}{\eta} \star j_y$, $i = 1, 2, \ldots, r$. Let $\mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda, \mathbf{g})$ be the proximal average of the concave functions $g_1, g_2, \ldots, g_r$ with coefficients $\lambda_1, \lambda_2, \ldots, \lambda_r$ and parameters μ^{-1}, η, defined by

$$\mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda, \mathbf{g}) = -\left((\lambda_1 \star (-\widehat{g}_1)) \mathbin{\#} (\lambda_2 \star (-\widehat{g}_2)) \mathbin{\#} \ldots \mathbin{\#} (\lambda_r \star (-\widehat{g}_r)) - \frac{1}{\mu} \star j_x - \eta \star j_y \right)$$

where $\widehat{g}_i = g_i - \frac{1}{\mu} \star j_x - \eta \star j_y$, $i = 1, 2, \ldots, r$.

Theorem 3.3 *The convex and the concave parents of the proximal average of* $\left[\mathcal{P}^{\cup\cap}_{\mu,\eta}(\lambda, \mathbf{h})\right]$ *are, respectively, the functions* $\mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda, \mathbf{f})$ *and* $\mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda, \mathbf{g})$.

A straightforward consequence of Theorem 3.3 is that the conjugate function, in the saddle sense, to the proximal average of $[h_1], [h_2], \ldots, [h_r]$, is the proximal average of $[h_1^*], [h_2^*], \ldots, [h_r^*]$.

Corollary 3.4 *The equivalence classes* $\left[\mathcal{P}^{\cup\cap}_{\mu,\eta}(\lambda, \mathbf{h})\right]$ *and* $\left[\mathcal{P}^{\cup\cap}_{\mu^{-1},\eta^{-1}}(\lambda, \mathbf{h}^*)\right]$ *are conjugate to one another in the sense of saddle function conjugacy.*

Proof. Theorem 3.3 says that the concave parent of $\left[\mathcal{P}^{\cup\cap}_{\mu,\eta}(\lambda, \mathbf{h})\right]$ is $\mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda, \mathbf{g})$. The convex parent of the class conjugate to $\left[\mathcal{P}^{\cup\cap}_{\mu,\eta}(\lambda, \mathbf{h})\right]$ is the function $(p, y) \mapsto -\mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda, \mathbf{g})(p, -y)$; recall (5) and the comments following it. Now, $-\mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda, \mathbf{g})(p, -y) = \mathcal{P}^{\cup\cup}_{\mu^{-1},\eta}(\lambda, \mathbf{g}')(p, y)$ where $\mathbf{g}'(p, y) = -\mathbf{g}(p, -y)$ and the functions $(p, y) \mapsto -g_i(p, -y)$ are convex parents of $[h_i]^*$. Hence the class conjugate to $\left[\mathcal{P}^{\cup\cap}_{\mu,\eta}(\lambda, \mathbf{h})\right]$ has, as its convex parents, the proximal average of convex parents of $[h_i]^*$. Another use of Theorem 3.3 shows that the class conjugate to $\left[\mathcal{P}^{\cup\cap}_{\mu,\eta}(\lambda, \mathbf{h})\right]$ is exactly $\left[\mathcal{P}^{\cup\cap}_{\mu^{-1},\eta^{-1}}(\lambda, \mathbf{h}^*)\right]$. $\square$

When, for every $i = 1, 2, \ldots, r$, $h_i = a_i j_x - b_i j_y$ for some $a_i, b_i > 0$, the proximal average can be found explicitly, thanks to the formula for the convex proximal average of convex quadratic functions; see [2, Example 4.5]. Here, as an illustration, we find the proximal average of the saddle functions h_1, h_2 on $I\!R^n \times I\!R^n$ given by:

$$h_1(x, y) = \frac{1}{2}|x|^2 - \frac{1}{2}|y|^2, \qquad h_2(x, y) = x \cdot y,$$

with parameters $\mu = \eta = 1$. One has:

$$\widetilde{h}_1(x, y) = |x|^2 - |y|^2, \qquad \widetilde{h}_2 = x \cdot y + \frac{1}{2}|x|^2 - \frac{1}{2}|y|^2,$$

$$\lambda_1 \star \widetilde{h}_1(x, y) = \frac{1}{\lambda_1}|x|^2 - \frac{1}{\lambda_1}|y|^2, \qquad \lambda_2 \star \widetilde{h}_2(x, y) = \frac{1}{\lambda_2} x \cdot y + \frac{1}{2\lambda_2}|x|^2 - \frac{1}{2\lambda_2}|y|^2,$$

and the proximal average $\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h})(x, y)$ is given by

$$\inf_{x_1 + x_2 = x} \sup_{y_1 + y_2 = y} \left\{ \frac{1}{\lambda_1}|x_1|^2 + \frac{1}{2\lambda_2}|x_2|^2 + \frac{1}{\lambda_2} x_2 \cdot y_2 - \frac{1}{\lambda_1}|y_1|^2 - \frac{1}{2\lambda_2}|y_2|^2 \right\} - \frac{1}{2}|x|^2 + \frac{1}{2}|y|^2.$$

The computation involves miminizing and maximizing linear-quadratic functions and is straightforward. One obtains

$$\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h})(x, y) = \frac{1 - \lambda_2^2}{1 + \lambda_2^2} \left(\frac{1}{2}|x|^2 - \frac{1}{2}|y|^2 \right) + \frac{2\lambda_2}{1 + \lambda_2^2} x \cdot y.$$

Note that $\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h}) = h_1$ when $\lambda_1 = 1$ and $\lambda_2 = 0$, while $\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h}) = h_2$ when $\lambda_2 = 1$. An interesting feature of $\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h})$ is its self-duality with respect to saddle conjugacy: $\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h})^* = \mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h})$. This self-duality follows from $h_1^* = h_1$, $h_2^* = h_2$, and Corollary 3.4 but can be also verified directly. For the case of $n = 1$, the self-duality of such a function was noted in [5]. A geometric interpretation of $\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h})$, for the case of $n = 1$, is that its graph is obtained from the graph of h_1 through the rotation, about the line $x = y = 0$, by the angle α such that $\sin 2\alpha = 2\lambda_2/(1 + \lambda_2^2)$. When $\lambda_2 = 1$ and $\alpha = \pi/4$, one obtains the graph of h_2.

Theorem 3.3 suggests an alternative way to compute $\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h})$. The convex parents of h_1, h_2 are given by, respectively,

$$f_1(x, q) = \frac{1}{2}|x|^2 + \frac{1}{2}|q|^2, \qquad f_2(x, q) = \delta_0(x + q).$$

The proximal average of f_1, f_2, $\mathcal{P}^{\cup\cup}_{1,1}(\lambda, \mathbf{f})$ is given at (x, q) by

$$\inf_{x_1+x_2=x, q_1+q_2=q} \left\{ \frac{1}{\lambda_1}|x_1|^2 + \frac{1}{\lambda_1}|q_1|^2 + \delta_0(x_2 + q_2) + \frac{1}{2\lambda_2}|x_2|^2 + \frac{1}{\lambda_1}|q_1|^2 \right\} - \frac{1}{2}|x|^2 - \frac{1}{2}|q|^2,$$

which simplifies to

$$\inf_{x_2} \left\{ \frac{1}{\lambda_1}|x - x_2|^2 + \frac{1}{\lambda_1}|q + x_2|^2 + \frac{1}{\lambda_2}|x_2|^2 \right\} - \frac{1}{2}|x|^2 - \frac{1}{2}|q|^2.$$

A simpler than in the saddle case computation yields

$$\mathcal{P}^{\cup\cup}_{1,1}(\lambda, \mathbf{f})(x, q) = \frac{1 + \lambda_2^2}{1 - \lambda_2^2}\frac{1}{2}|x|^2 - \frac{2\lambda_2}{1 - \lambda_2^2} x \cdot q + \frac{1 + \lambda_2^2}{1 - \lambda_2^2}\frac{1}{2}|q|^2.$$

Then, the relationship $\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h})(x, y) = \inf_q \left\{ \mathcal{P}^{\cup\cup}_{1,1}(\lambda, \mathbf{f})(x, q) - y \cdot q \right\}$ yields the same formula for $\mathcal{P}^{\cup\cap}_{1,1}(\lambda, \mathbf{h})$ as obtained before.

3.2 A key proposition

Let $\overline{H} : I\!R^m \times I\!R^n \to \overline{I\!R}$ be given, at each $(x, y) \in I\!R^m \times I\!R^n$, by

$$\overline{H}(x, y) = h(x, y) - \mu \star j(x) + \eta \star j\,(y)$$

where $\widetilde{h}_i = \overline{h}_i + \mu \star j_x - \eta \star j_y$, $i = 1, 2, \ldots, r$, and

$$h(x, y) = \inf_{\sum x_i = x, x_i \in \lambda_i C_i} \ \sup_{\sum y_i = y, y_i \in I\!R^n} \sum \left(\lambda_i \star \widetilde{h}_i\right)(x_i, y_i).$$

Proposition 3.5 *For all $(x, y) \in I\!R^m \times I\!R^n$,*

$$\inf_{q \in I\!R^n} \left\{ \mathcal{P}^{\cup\cup}_{\mu, \eta^{-1}}(\lambda, \mathbf{f})(x, q) - y \cdot q \right\} = \overline{H}(x, y). \tag{10}$$

The remainder of this section is devoted to proving this result. The summation symbol $\sum$ in what follows always means $\sum_{i=1}^r$. The left hand side of (10) equals $\inf_q \inf_{\mathbf{x}\colon \sum x_i = x} \inf_{\mathbf{q}\colon \sum q_i = q} H_1(q, \mathbf{q}, \mathbf{x})$, where $\mathbf{q} = (q_1, q_2, \ldots, q_r)$, $\mathbf{x} = (x_1, x_2, \ldots, x_r)$ and

$$H_1(q, \mathbf{q}, \mathbf{x}) = \sum \left[\lambda_i f_i \left(\frac{x_i}{\lambda_i}, \frac{q_i}{\lambda_i} \right) + \frac{1}{\lambda_i \mu} j(x) + \frac{\eta}{\lambda_i} j(q_i) \right] - \frac{1}{\mu} j(x) - \eta j(q) - y \cdot q.$$

With no loss of generality, the infimum can be taken over $\mathbf{x}$ such that, for $i = 1, 2, \ldots, r$, x_i is such that there exists q_i with $\left(\frac{x_i}{\lambda_i}, \frac{q_i}{\lambda_i}\right) \in \operatorname{dom} f_i$, and thus, over $x_i \in \lambda_i C_i$, where $C_i \times D_i = \operatorname{dom} h_i$. This fact, switching the order of taking the infima, and then using the first equation in (2), with the greatest element $\overline{h}_i$ of $[h]$ in place of h, and with y_i/λ_i taking place of y, turns the left hand side of (10) to $\inf_{\mathbf{x}:\, \sum x_i = x, x_i \in \lambda_i C_i} \inf_q \inf_{\mathbf{q}:\, \sum q_i = q} \sup_{\mathbf{y}} H_2(q, \mathbf{x}, \mathbf{y})$, where $\mathbf{y} = (y_1, y_2, \ldots, y_r)$ and

$$\begin{aligned} H_2(q, \mathbf{q}, \mathbf{x}, \mathbf{y}) &= \sum \left[\lambda_i \overline{h}_i \left(\frac{x_i}{\lambda_i}, \frac{y_i}{\lambda_i} \right) + \frac{\eta}{\lambda_i} j(q_i) + \frac{y_i}{\lambda_i} q_i \right] \\ &\quad - \eta j(q) - y \cdot q + \sum \frac{1}{\lambda_i \mu} j(x_i) - \frac{1}{\mu} j(x). \end{aligned}$$

As $x_i \in \lambda_i C_i$, the function $\overline{h}_i \left(\frac{x_i}{\lambda_i}, \cdot \right)$ is upper semicontinuous (see the discussion above [9, Theorem 33.3]) and proper as a concave function, in the sense that it does not take on the value ∞ and is not identically $-\infty$ (see [9, Theorem 43.3]). Thus, the first line in the formula for H_2 above is a proper and closed saddle function of $(\mathbf{q}, \mathbf{y}) \in (\mathbb{R}^m)^r \times (\mathbb{R}^n)^r$. It is also strongly convex in $\mathbf{q}$ and, consequently, it has no directions of recession in $\mathbf{q}$; see [11, Theorem 8.7]. Then, [9, Theorem 37.3] ensures that switching the order of the infimum and supremum is possible. Thus, the left hand side of (10) becomes $\inf_{\mathbf{x}:\, \sum x_i = x, x_i \in \lambda_i C_i} \inf_q \sup_{\mathbf{y}} \inf_{\mathbf{q}:\, \sum q_i = q} H_2(q, \mathbf{q}, \mathbf{x}, \mathbf{y})$. Some simplification is now possible, thanks to the following lemma.

Lemma 3.6

$$\inf_{\mathbf{q}:\, \sum q_i = q} \sum \left[\frac{\eta}{\lambda_i} j(q_i) + \frac{y_i}{\lambda_i} q_i \right] = \frac{1}{\eta} j \left(\sum y_i \right) + \eta j(q) + q \cdot \sum y_i - \sum \frac{1}{\eta \lambda_i} j(y_i) \tag{11}$$

Proof. Let $\phi_i : \mathbb{R}^n \to \mathbb{R}$ be given by $\phi_i(q_i) = \frac{\eta}{\lambda_i} j(q_i) + \frac{y_i}{\lambda_i} q_i$, $i = 1, 2, \ldots, r$; let $\Phi : \mathbb{R}^{nr} \to \mathbb{R}$ be the convex and finite-valued function given by $\Phi(\mathbf{q}) = \sum \phi_i(q_i)$; let $A : \mathbb{R}^{nr} \to \mathbb{R}^n$ be given by $A\mathbf{q} = \sum q_i$; and let $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be the indicator function of q, i.e., $\Phi(y) = 0$ if $y = q$, $\Phi(y) = \infty$ if $y \neq q$, which makes Ψ proper, lsc, and convex. The left hand side of (11) is

$$\inf_{\mathbf{q} \in \mathbb{R}^{rn}} \left\{ \Phi(\mathbf{q}) + \Psi(A\mathbf{q}) \right\}.$$

Since Φ is finite-valued, Fenchel Duality (see, for example, [9, Corollary 31.2.1]) yields that the infimum above equals

$$\sup_{v \in \mathbb{R}^n} \left\{ -\Psi^*(v) - \Phi^*(-A^* v) \right\} = \sup_{v \in \mathbb{R}^n} \left\{ -q \cdot v - \sum \phi_i^*(-v) \right\} = \sup_{v \in \mathbb{R}^n} \left\{ q \cdot v - \sum \phi_i^*(v) \right\}.$$

Now,

$$\begin{aligned} \phi_i^*(v) &= \sup_{q \in \mathbb{R}^n} \left\{ v \cdot q - \frac{y_i}{\lambda_i} q - \frac{\eta}{\lambda_i} j(q) \right\} = \frac{\eta}{\lambda_i} \sup_{q \in \mathbb{R}^n} \left\{ \frac{\lambda_i}{\eta} \left(v - \frac{y_i}{\lambda_i} \right) q - j(q) \right\} \\ &= \frac{\eta}{\lambda_i} j \left(\frac{\lambda_i}{\eta} \left(v - \frac{y_i}{\lambda_i} \right) \right) = \frac{\lambda_i}{\eta} j \left(v - \frac{y_i}{\lambda_i} \right) \end{aligned}$$

and

$$\begin{aligned}
\sup_{v\in\boldsymbol{R}^n}\left\{q\cdot v-\sum\phi_i^*(v)\right\} &= \sup_{v\in\boldsymbol{R}^n}\left\{q\cdot v-\sum\frac{\lambda_i}{\eta}j\left(v-\frac{y_i}{\lambda_i}\right)\right\} \\
&= \sup_{v\in\boldsymbol{R}^n}\left\{q\cdot v-\frac{1}{\eta}j(v)+\frac{1}{\eta}v\cdot\sum y_i\right\}-\frac{1}{\eta}\sum\frac{1}{\lambda_i}j(y_i) \\
&= \frac{1}{\eta}\sup_{v\in\boldsymbol{R}^n}\left\{\left(\sum y_i+\eta q\right)\cdot v-j(v)\right\}-\frac{1}{\eta}\sum\frac{1}{\lambda_i}j(y_i) \\
&= \frac{1}{\eta}j\left(\sum y_i+\eta q\right)-\frac{1}{\eta}\sum\frac{1}{\lambda_i}j(y_i)
\end{aligned}$$

which expands to the right hand side of (11). □

Lemma 3.6 shows that the left hand side of (10) is $\inf_{\mathbf{x}:\sum x_i=x, x_i\in\lambda_i C_i}\inf_q\sup_{\mathbf{y}} H_3(q,\mathbf{x},\mathbf{y})$, where

$$\begin{aligned}
H_3(q,\mathbf{x},\mathbf{y}) &= q\cdot\left(\sum y_i-y\right)+\frac{1}{\eta}j\left(\sum y_i\right)+\sum\left[\lambda_i\overline{h}_i\left(\frac{x_i}{\lambda_i},\frac{y_i}{\lambda_i}\right)-\frac{1}{\eta\lambda_i}j(y_i)\right] \\
&\quad -\frac{1}{\mu}j(x)+\sum\frac{1}{\lambda_i\mu}j(x_i).
\end{aligned}$$

Lemma 3.7 below shows that, for every $x_i\in\lambda_i C_i$,

$$\inf_q\sup_{\mathbf{y}} H_3(q,\mathbf{x},\mathbf{y}) = \sup_{\mathbf{y}:\sum y_i=y} H_4(\mathbf{x},\mathbf{y})$$

where

$$H_4(\mathbf{x},\mathbf{y})=\sum\left[\lambda_i\overline{h}_i\left(\frac{x_i}{\lambda_i},\frac{y_i}{\lambda_i}\right)+\frac{1}{\lambda_i\mu}j(x_i)-\frac{1}{\eta\lambda_i}j(y_i)\right]-\frac{1}{\mu}j(x)+\frac{1}{\eta}j\left(y\right).$$

Consequently, the left hand side of (10) is

$$\inf_{\mathbf{x}:\sum x_i=x, x_i\in\lambda_i C_i}\ \sup_{\mathbf{y}:\sum y_i=y} H_4(\mathbf{x},\mathbf{y}).$$

This verifies (10).

Lemma 3.7 *Given $x_i\in\lambda_i C_i$, let $\phi:(\mathbb{R}^n)^r\to\overline{\mathbb{R}}$ be given by*

$$-\phi(\mathbf{y})=\frac{1}{\eta}j\left(\sum y_i\right)+\sum\left[\lambda_i\overline{h}_i\left(\frac{x_i}{\lambda_i},\frac{y_i}{\lambda_i}\right)-\frac{1}{\eta\lambda_i}j(y_i)\right].$$

Then

$$\inf_q\sup_{\mathbf{y}}\left\{q\cdot\left(\sum y_i-y\right)-\phi(\mathbf{y})\right\}=-\inf_{\mathbf{y}:\sum y_i=y}\phi(\mathbf{y}).$$

Proof. The function

$$\mathbf{y}\mapsto\sum\lambda_i j\left(\frac{y_i}{\lambda_i}\right)-j\left(\sum y_i\right)=\sum\lambda_i j\left(\frac{y_i}{\lambda_i}\right)-j\left(\sum\lambda_i\frac{y_i}{\lambda_i}\right)$$

is convex, finite-valued, and hence continuous. Indeed, it is quadratic and, by convexity of j, nonnegative. Furthermore, it is strongly convex in every direction except $y_1/\lambda_1=y_2/\lambda_2=\cdots=y_r/\lambda_r$. Consequently, the function ϕ is convex, proper, lower semicontinuous, and

$$\inf_q\sup_{\mathbf{y}}\left\{q\cdot\left(\sum y_i-y\right)-\phi(\mathbf{y})\right\}=\inf_q\left\{\phi^*\left(A^Tq\right)-q\cdot y\right\}=-\sup_q\left\{y\cdot q-\phi^*\left(A^Tq\right)\right\},$$

where $A : (\mathbb{R}^n)^r \to \mathbb{R}^n$ is given by $A\mathbf{y} = \sum y_i$, and so $A^T q = (q, q, \dots, q) \in (\mathbb{R}^n)^r$. The statement of the lemma amounts to $\psi^*(y) = \inf_{\mathbf{y}: \sum y_i = y} \phi(\mathbf{y})$ for the convex function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ given by $\psi(q) = \phi^*\left(A^T q\right)$. This holds, by [11, Theorem 11.23], if $0 \in \operatorname{int}\left(\operatorname{dom}\phi^* - \operatorname{rge} A^T\right)$. Suppose that this constraint qualification fails: there exist $\mathbf{v} = (v_1, v_2, \dots, v_r) \in (\mathbb{R}^n)^r$, $\mathbf{v} \neq 0$, such that $\mathbf{v}\cdot\mathbf{w} - \mathbf{v}\cdot A^T q \leq 0$ for all $\mathbf{w} \in \operatorname{dom}\phi^*$, $q \in \mathbb{R}^n$. In particular, $\mathbf{v}\cdot\mathbf{w} \leq 0$ for all $\mathbf{w} \in \operatorname{dom}\phi^*$ while $0 = \mathbf{v}\cdot A^T q$ for all $q \in \mathbb{R}^n$. Since ϕ is strongly convex in every direction except possibly $y_1/\lambda_1 = y_2/\lambda_2 = \cdots = y_r/\lambda_r$, the horizon function ϕ^∞ of ϕ is such that $\phi^\infty(\mathbf{y}) = \infty$ except possibly when $y_1/\lambda_1 = y_2/\lambda_2 = \cdots = y_r/\lambda_r$. By [11, Theorem 11.5], this horizon function is the support function of $\operatorname{dom}\phi^*$. Thus, $\mathbf{v}\cdot\mathbf{w} \leq 0$ for all $\mathbf{w} \in \operatorname{dom}\phi^*$ implies that $v_1/\lambda_1 = v_2/\lambda_2 = \cdots = v_r/\lambda_r$. But then $0 = A\mathbf{v}\cdot q = (1 + \lambda_2/\lambda_1 + \cdots + \lambda_r/\lambda_1)\, v_1 \cdot q$ for all $q \in \mathbb{R}^n$ implies $\mathbf{v} = 0$. This is a contradiction. ☐

3.3 Proof of Theorems 3.2 and 3.3

Since the functions $\widetilde{h}_i$ in the definition of $\overline{H}$ are the greatest elements of $[\widetilde{h}_i]$, $i = 1, 2, \dots, r$, the function $\overline{H}$ is the greatest function of the form (8). The function $\overline{H}$ turns out to be a partial conjugate, as in the second formula in (4), of a proper, lsc, and convex function $\mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda, \mathbf{f})$. Thus $\overline{H}$ is a proper and closed saddle function, and in particular, there exists a proper and closed saddle function of the form (8).

Let $\underline{H} : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ be given, at each $(x, y) \in \mathbb{R}^m \times \mathbb{R}^n$, by

$$\underline{H}(x, y) = h(x, y) - \mu \star j(x) + \eta \star j\,(y)$$

where $\widetilde{h}_i = \underline{h_i} + \mu \star j_x - \eta \star j_y$, $i = 1, 2, \dots, r$, and

$$h(x, y) = \sup_{\sum y_i = y, y_i \in D_i} \inf_{\sum x_i = x, x_i \in \mathbb{R}^m} \sum \widetilde{h}_i(x_i, y_i).$$

Since $\widetilde{h}_i$ are the least elements of $[\widetilde{h}_i]$, $i = 1, 2, \dots, r$, the function $\underline{H}$ is the least function of the form (8). The function $\underline{H}$ turns out to be a partial conjugate, as in the first formula in (4), of a proper, lsc, and convex function $\mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda, \mathbf{g})$. Indeed, Arguments symmetric to those showing Proposition 3.5 imply the following: for all $(x, y) \in \mathbb{R}^m \times \mathbb{R}^n$,

$$\sup_{p \in \mathbb{R}^m} \left\{ \mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda, \mathbf{g})(p, y) - x \cdot p \right\} = \underline{H}(x, y). \tag{12}$$

Thus $\underline{H}$ is a proper and closed saddle function.

The proper and lower semicontinuous convex functions $\mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda, \mathbf{f})$ and $-\mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda, \mathbf{g})$ turn out to be conjugate to one another. This is a consequence of f_i and $-g_i$ being conjugate to one another, $i = 1, 2, \dots, r$, and the self-duality of the proximal average in the convex setting, as summarized below.

Lemma 3.8 *For all* $(p, y) \in \mathbb{R}^m \times \mathbb{R}^n$,

$$\left(\mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda, \mathbf{f})\right)^*(p, y) := \sup_{(x,q) \in \mathbb{R}^m \times \mathbb{R}^n} \left\{ p \cdot x + y \cdot q - \mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda, \mathbf{f})(x, q) \right\} = \mathcal{P}^{\cup\cup}_{\mu^{-1},\eta}(\lambda, \mathbf{f}^*)(p, y).$$

Proof. For the functions $\widehat{f}_i$, $i = 1, 2, \dots, r$ in (9), one has, for all $(x, q) \in \mathbb{R}^m \times \mathbb{R}^n$,

$$\widehat{f}_i(x, q) \quad = \quad f_i(x, q) + \mu \star j_x(x, q) + \frac{1}{\eta} \star j_y(x, q) = f_i(x, q) + \mu \star j(x) + \frac{1}{\eta} \star j(q)$$

$$
\begin{aligned}
&= f_i\left(\sqrt{\mu}\frac{x}{\sqrt{\mu}}, \frac{1}{\sqrt{\eta}}\sqrt{\eta}q\right) + j\left(\frac{x}{\sqrt{\mu}}\right) + j\left(\sqrt{\eta}q\right) \\
&= g_i\left(A(x,q)\right) + j\left(A(x,q)\right)
\end{aligned}
$$

where $A(x,q) = \left(\frac{x}{\sqrt{\mu}}, \sqrt{\mu}q\right)$ and $g_i(x,q) = f_i\left(\sqrt{\mu}x, \frac{1}{\sqrt{\eta}}q\right) = f_i\left(A^{-1}(x,q)\right)$. Thus, for all $(x,q) \in I\!R^m \times I\!R^n$,

$$
\mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda,\mathbf{f})(x,q) = \mathcal{P}^{\cup\cup}_{1,1}(\lambda,\mathbf{g})\left(A(x,q)\right).
$$

Self-duality of the proximal average for convex functions, as stated in Theorem 2.1, and [9, Theorem 12.3] combined with the fact that the linear mapping A is invertible and self-adjoint,

$$
\left(\mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda,\mathbf{f})\right)^*(p,y) = \left(\mathcal{P}^{\cup\cup}_{1,1}(\lambda,\mathbf{g})\right)^*\left(A^{-1}(p,y)\right) = \mathcal{P}^{\cup\cup}_{1,1}(\lambda,\mathbf{g}^*)\left(A^{-1}(p,y)\right).
$$

[9, Theorem 12.3] also gives that $g_i^*(p,y) = f_i^*(A(p,y))$. A calculation, similar to the one carried out above for $\widehat{f}_i$, finishes the argument. ☐

Conjugacy of the functions $\mathcal{P}^{\cup\cup}_{\mu,\eta^{-1}}(\lambda,\mathbf{f})$ and $-\mathcal{P}^{\cap\cap}_{\mu^{-1},\eta}(\lambda,\mathbf{g})$, and their relationship to $\overline{H}$ and $\underline{H}$ captured by (10) and (12) implies that all saddle functions H such that $\underline{H} \le H \le \overline{H}$ are proper, closed, and form an equivalence class. Since $\underline{H}$ and $\overline{H}$ are, respectively, the least and the greatest saddle functions of the form (8), the proof of Theorem 3.2 and Theorem 3.3 is complete.

References

[1] H. Attouch, D. Azé, and R. J-B Wets. Convergence of convex-concave saddle functions: continuity properties of the Legendre-Fenchel transform and applications to convex programming and mechanics. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 5:537–572, 1988.

[2] H. Bauschke, R. Goebel, Y. Lucet, , and X. Wang. The proximal average: basic theory. *SIAM J. Optim.*, 19(2):766–785, 2008.

[3] H.H. Bauschke, Y. Lucet, and M. Trienis. How to transform one convex function continuously into another. *SIAM Rev.*, 50:115–132, 2008.

[4] H.H. Bauschke, E. Matoušková, and S. Reich. Projection and proximal point methods: convergence results and counterexamples. *Nonlinear Anal.*, 56(5):715–738, 2004.

[5] R. Goebel. *Convexity, Convergence and Feedback in Optimal Control.* PhD thesis, University of Washington, Seattle, 2000.

[6] W. L. Hare. A proximal average for nonconvex functions: a proximal stability perspective. *SIAM J. Optim.*, 20(2):650–666, 2009.

[7] L. McLinden. Dual operations on saddle functions. *Trans. Amer. Math. Soc.*, 179:363–381, 1973.

[8] R.T. Rockafellar. A general correspondence between dual minimax problems and convex programs. *Pacific J. Math.*, 25:597–611, 1968.

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

[10] R.T. Rockafellar. *Conjugate Duality and Optimization.* Number 16 in CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, 1974.

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