The kernel average for two convex functions and its application to the extension and representation of monotone operators

Heinz H. Bauschke[*] and Xianfu Wang[†]

May 9, 2007

Abstract

We provide and analyze a based average for two convex functions, based on a kernel function. It covers several known averages such as the arithmetic average, epigraphical average, and the proximal average. When applied to the Fitzpatrick function and the conjugate of Fitzpatrick function associated with a monotone operator, our average produces an autoconjugate (also known as selfdual Lagrangian) which can be used for finding an explicit maximal monotone extension of the given monotone operator. This completely settles one of the open problems posed by Fitzpatrick in the setting of reflexive Banach spaces.

2000 Mathematics Subject Classification: Primary 52A41, 47N10; Secondary 47H05.

Keywords: Arithmetic average, autoconjugate, convex function, epigraphical average, Fenchel conjugate, Fitzpatrick function, maximal monotone operator, monotone operator, proximal average, proximal mapping, selfdual Lagrangian.

1 Introduction

In the first part of this paper, we use X for $\mathbb{R}^n$ with a given norm $\|\cdot\|$ (not necessarily the Euclidean norm), and X^* for the dual of X. The class of proper lower semi-continuous convex functions on X is denoted by $\Gamma(X)$. We follow standard convex-analytical notation as in e.g., [18, 24]. Thus, for a convex function $h\colon X \to\ ]-\infty, +\infty]$, the (effective) domain is $\operatorname{dom} h := \{x \in X : \ h(x) < \infty\}$, and we use $\operatorname{ri}\operatorname{dom} h$ for the relative interior of the domain. The Fenchel conjugate h^* of h is the function defined on X^* by $h^*(x^*) := \sup\{\langle x^*, x\rangle - h(x) : \ x \in X\}$. For $x \in \operatorname{dom} h$, the set of subgradients

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

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

of h at x is $\partial h(x) := \{x^* \in X^* : \ h(y) - h(x) \geq \langle x^*, y - x\rangle \ \forall y \in X\}$, whereas $\partial h(x) := \varnothing$ if $x \notin \operatorname{dom} h$. By $\operatorname{cl} h$ we mean the lower semi-continuous hull of h. If $S \subset X$, then ι_S stands for the corresponding indicator function, i.e., $\iota_S(x) = 0$ for $x \in S$ and $+\infty$ for $x \notin S$.

Definition 1.1 (kernel average) *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$. Define $P(\lambda_1, f_1, \lambda_2, f_2, g)\colon X \to [-\infty, +\infty]$ at $x \in X$ by*

$$P(\lambda_1, f_1, \lambda_2, f_2, g)(x) := \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} \{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2)\} \tag{1}$$

$$= \inf_{x = z_1 + z_2} \left\{\lambda_1 f_1\left(\frac{z_1}{\lambda_1}\right) + \lambda_2 f\left(\frac{z_2}{\lambda_2}\right) + \lambda_1\lambda_2 g\left(\frac{z_1}{\lambda_1} - \frac{z_2}{\lambda_2}\right)\right\}.$$

We call this the average of f_1 and f_2 with respect to the kernel g *or the* g-average of f_1 and f_2*.*

The main objective of this paper is to study the kernel average and to provide its basic properties. The kernel average, which can be interpreted as an epigraphical average modified by the kernel g, is quite flexible as the following examples show.

Example 1.2 (arithmetic average) Let $g = \iota_0$ and set $h = P(\lambda_1, f_1, \lambda_2, f_2, g)$. Then

$$h(x) = \inf_{\substack{\lambda_1 y_1 + \lambda_2 y_2 = x \\ y_1 = y_2}} \{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2)\} = \lambda_1 f_1(x) + \lambda_2 f_2(x)$$

is the arithmetic average.

In the following, $\lambda \star f$ denotes the epi-product $\lambda f(\cdot/\lambda)$ if $\lambda > 0$ and $f \in \Gamma(X)$ (see [19]).

Example 1.3 (epigraphical average) Let $g = \iota_X$ and set $h = P(\lambda_1, f_1, \lambda_2, f_2, g)$. Then

$$h(x) = \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} \{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2)\} = \inf_{x = z_1 + z_2} \left\{\lambda_1 f_1\Big(\frac{z_1}{\lambda_1}\Big) + \lambda_2 f_2\Big(\frac{z_2}{\lambda_2}\Big)\right\}$$
$$= \big((\lambda_1 \star f_1) \Box (\lambda_2 \star f_2)\big)(x),$$

is the epigraphical average. Using $\frac{1}{\lambda_1} \star f_1$ and $\frac{1}{\lambda_2} \star f_2$ instead of f_1 and f_2 respectively, we obtain

$$P\big(\lambda_1, \tfrac{1}{\lambda_1} \star f_1, \lambda_2, \tfrac{1}{\lambda_2} \star f_2, \iota_X\big) = f_1 \Box f_2,$$

the usual infimal convolution.

Example 1.4 (proximal average) Let $g = \frac{1}{2}\|\cdot\|^2$, where $\|\cdot\|$ is the Euclidean norm, and set $h = P(\lambda_1, f_1, \lambda_2, f_2, g)$. Then

$$h(x) = \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} \left\{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \tfrac{1}{2}\lambda_1\lambda_2\|y_1 - y_2\|^2\right\}$$
$$= \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} \left\{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \tfrac{1}{2}\lambda_1\|y_1\|^2 + \tfrac{1}{2}\lambda_2\|y_2\|^2 - \tfrac{1}{2}\|x\|^2\right\}$$

is the proximal average; see [4, Proposition 4.3] and also [5].

As another illustration, let us note that several envelopes can be interpreted as kernel averages.

Example 1.5 (Attouch-Wets envelope) Let $f_2 = 0$, let $\lambda_1, \lambda_2 > 0$, and let $g = \lambda_2^{p-1}\frac{1}{\mu}\frac{1}{p}\|\cdot\|^p$ with $\mu > 0$ and $p \geq 1$. We have

$$f_1 \square \frac{1}{\mu}\frac{1}{p}\|\cdot\|^p = \frac{1}{\lambda_1} P\Big(\lambda_1, f_1, \lambda_2, 0, \lambda_2^{p-1}\frac{1}{\mu}\frac{1}{p}\|\cdot\|^p\Big).$$

Indeed, if $x = \lambda_1 y_1 + \lambda_2 y_2$, then $y_1 - y_2 = (y_1 - x)/\lambda_2$ and hence

$$\begin{aligned}
&P\Big(\lambda_1, f_1, \lambda_2, 0, \lambda_2^{p-1}\frac{1}{\mu}\frac{1}{p}\|\cdot\|^p\Big)(x)\\
&= \inf_{x=\lambda_1 y_1+\lambda_2 y_2} \Big(\lambda_1 f_1(y_1) + \lambda_2 0 + \lambda_1\lambda_2\lambda_2^{p-1}\frac{1}{\mu}\frac{1}{p}\|y_1 - y_2\|^p\Big)\\
&= \inf_{y_1} \Big(\lambda_1 f_1(y_1) + \lambda_1\lambda_2^p\frac{1}{\mu}\frac{1}{p}\|\frac{y_1 - x}{\lambda_2}\|^p\Big)\\
&= \inf_{y_1} \Big(\lambda_1 f_1(y_1) + \lambda_1\frac{1}{\mu}\frac{1}{p}\|y_1 - x\|^p\Big)\\
&= \lambda_1 \inf_{y_1} \Big(f_1(y_1) + \frac{1}{\mu}\frac{1}{p}\|x - y_1\|^p\Big).
\end{aligned}$$

While $p = 2$ gives the Moreau envelope, the assignment $p = 1$ leads to the Pasch-Hausdorff envelope. See also [19, page 296] for some history of these envelopes and further references.

It is interesting to ask what kind of nice properties does the kernel average given by (1) have. The purpose of this paper is to study our kernel average from three perspectives: conjugacy, subdifferentiability, and applications in optimization and monotone operator theory.

The paper is organized as follows. In Section 2 we study conjugacy and subdifferentiability properties of g-averages. It turns out that if g has full domain, then the Fenchel conjugate of the g-average is essentially (up to a minus sign) the g^*-average of f_1^* and f_2^* and the domain of g-average is the convex combination (in the sense of Minkowski sums) of $\operatorname{dom} f_1$ and $\operatorname{dom} f_2$. Moreover, if g is differentiable, the g-average is Legendre type if f_1 or f_2 is. The relationship between the minimizers of g-average and the minimizers of f_1 and f_2 is analyzed in Section 3. Section 4 introduces a corresponding g-proximal mapping of the g-average and provides an expression in terms of the g-proximal mappings of f_1 and f_2. Finally, in Section 5, we supply a kernel average in reflexive Banach spaces. We utilize it to construct autoconjugates, which can be used to find an explicit maximal extension of a monotone operator with an attractive duality property. Theorem 5.7 provides a complete solution to one of Fitzpatrick's open problems [10] in the reflexive setting.

2 Main Results

It is well known that the conjugate of the epigraphical average of f_1 and f_2 is the arithmetic average of conjugates f_1^* and f_2^* (see also Corollary 2.4). Our goal is to calculate the conjugate of the kernel average of f_1 and f_2. It turns out that this is essentially the kernel average of the conjugate functions with respect to the conjugate of the original kernel. We start with a simple result which plays an important role in the proof of our main theorem. Before we state and prove this auxiliary result, we introduce the difference operator $D : X \times X \to X$, defined by

$$D(x, y) := x - y, \tag{2}$$

and its adjoint $D^* : X^* \to X^* \times X^*$, which satisfies

$$D^* z = (z, -z).$$

Lemma 2.1 *Let $f : X \times X \to \,]-\infty, +\infty]$ be given by $f(x, y) = g(x - y)$, where $g \in \Gamma(X)$. Then*

$$f^*(x^*, y^*) = \begin{cases} g^*(x^*), & \text{if } x^* + y^* = 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{3}$$

Consequently, if $\lambda > 0$, then

$$(\lambda f)^*(x^*, y^*) = \lambda g^*(x^*/\lambda) + \iota_{\{0\}}(x^* + y^*) = \begin{cases} \lambda g^*(x^*/\lambda), & \text{if } x^* + y^* = 0; \\ +\infty, & \text{otherwise.} \end{cases} \tag{4}$$

Proof. Since $f = g \circ D$, [24, Theorem 2.3.1(ix)] implies that

$$f^* = (g \circ D)^* = (D^* g^*)^{**} = \mathrm{cl}(D^* g^*) \quad \text{on } X^* \times X^*.$$

For each (x^*, y^*), we have

$$\begin{aligned} \big(D^* g^*\big)(x^*, y^*) &= \inf \big\{ g^*(z) : \ D^* z = (x^*, y^*) \big\} \\ &= \inf \big\{ g^*(z) : \ (z, -z) = (x^*, y^*) \big\} \\ &= \begin{cases} g^*(x^*), & \text{if } x^* + y^* = 0; \\ +\infty, & \text{otherwise} \end{cases} \\ &= g^*(x^*) + \iota_{\{0\}}(x^* + y^*). \end{aligned}$$

Hence $D^* g^*$ is lower semicontinuous and thus $\mathrm{cl}(D^* g^*) = D^* g^*$. Therefore, $f^* = D^* g^*$ and (3) holds. Now (4) follows from (3) by [24, Theorem 2.3.1(v)]. ■

For convenience, we denote the g-average of f_1 and f_2 by h, and we recall (see Definition 1.1) that

$$h(x) = \inf \big\{ \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1 \lambda_2 g(y_1 - y_2) : \ \lambda_1 y_1 + \lambda_2 y_2 = x \big\}. \tag{5}$$

Define $A : X \times X \to X$ by

$$A(y_1, y_2) = \lambda_1 y_1 + \lambda_2 y_2. \tag{6}$$

Then $A^* : X^* \to X^* \times X^*$ is given by

$$A^* x^* = (\lambda_1 x^*, \lambda_2 x^*). \tag{7}$$

Define $F : X \times X \to \,]-\infty, +\infty]$ by

$$F(y_1, y_2) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2). \tag{8}$$

Then (5) becomes

$$h = AF. \tag{9}$$

We are now in a position to determine the conjugate of the kernel average. To formulate the result, we define, for a given function $f \in \Gamma(X)$, the function $f^\vee \in \Gamma(X)$ by $f^\vee(x) = f(-x)$.

Theorem 2.2 (Fenchel conjugate) *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$. Then for every $x^* \in X$,*

$$\big(P(\lambda_1, f_1, \lambda_2, f_2, g)\big)^*(x^*) = (\operatorname{cl}\varphi)(\lambda_1 x^*, \lambda_2 x^*), \tag{10}$$

where

$$\varphi(u, v) = \inf_{u+v=\lambda_1 z_1+\lambda_2 z_2} \left\{ \lambda_1 f_1^*(z_1) + \lambda_2 f_2^*(z_2) + \frac{\lambda_1\lambda_2}{2}\left(g^*\Big(\frac{u}{\lambda_1\lambda_2} - \frac{z_1}{\lambda_2}\Big) + g^*\Big(\frac{-v}{\lambda_1\lambda_2} + \frac{z_2}{\lambda_1}\Big)\right)\right\}.$$

If

$$(\operatorname{ri}\operatorname{dom} f_1 - \operatorname{ri}\operatorname{dom} f_2) \cap \operatorname{ri}\operatorname{dom} g \neq \varnothing, \tag{11}$$

then the closure operation in (10) can be omitted so that not only

$$\Big(P\big(\lambda_1, f_1, \lambda_2, f_2, g\big)\Big)^* = P\big(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee}\big), \tag{12}$$

but also the infimum in the definition of the kernel average is attained, i.e.,

$$P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee})(x^*) = \min_{x^*=\lambda_1 z_1+\lambda_2 z_2} \{\lambda_1 f_1^*(z_1) + \lambda_2 f_2^*(z_2) + \lambda_1\lambda_2 g^*(z_2 - z_1)\}. \tag{13}$$

Proof. Using (9) and [18, Theorem 16.3], we have

$$h^* = (AF)^* = F^* \circ A^* \quad \text{ on } X^*. \tag{14}$$

Since

$$F = g_1 + g_2,$$

where

$$g_1(y_1, y_2) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2), \quad g_2(y_1, y_2) = \lambda_1\lambda_2 g(y_1 - y_2),$$

it follows from [18, Theorem 16.4] that

$$F^* = (g_1 + g_2)^* = \operatorname{cl}(g_1^* \Box g_2^*). \tag{15}$$

Using Lemma 2.1, we see that for every $(u,v) \in X^* \times X^*$,

$$\begin{aligned}
&\left(g_1^* \Box g_2^*\right)(u,v) \\
&= \inf_{y_1,y_2} \left\{ \lambda_1 f_1^* \left(\frac{y_1}{\lambda_1} \right) + \lambda_2 f_2^* \left(\frac{y_2}{\lambda_2} \right) + g_2^*(u-y_1, v-y_2) \right\} \\
&= \inf_{u-y_1+v-y_2=0} \left\{ \lambda_1 f_1^* \left(\frac{y_1}{\lambda_1} \right) + \lambda_2 f_2^* \left(\frac{y_2}{\lambda_2} \right) + \lambda_1\lambda_2 g^* \left(\frac{u-y_1}{\lambda_1\lambda_2} \right) \right\} \\
&= \inf_{u+v=y_1+y_2} \left\{ \lambda_1 f_1^* \left(\frac{y_1}{\lambda_1} \right) + \lambda_2 f_2^* \left(\frac{y_2}{\lambda_2} \right) + \lambda_1\lambda_2 g^* \left(\frac{u}{\lambda_1\lambda_2} - \frac{y_1}{\lambda_1\lambda_2} \right) \right\} \qquad (16) \\
&= \inf_{u+v=\lambda_1 z_1+\lambda_2 z_2} \left\{ \lambda_1 f_1^*(z_1) + \lambda_2 f_2^*(z_2) + \frac{\lambda_1\lambda_2}{2} \left[g^* \left(\frac{u}{\lambda_1\lambda_2} - \frac{z_1}{\lambda_2} \right) + g^* \left(\frac{-v}{\lambda_1\lambda_2} + \frac{z_2}{\lambda_1} \right) \right] \right\}.
\end{aligned}$$

Hence (10) holds.

Now assume that

$$\operatorname{ri}\operatorname{dom} g_1 \cap \operatorname{ri}\operatorname{dom} g_2 \neq \varnothing. \qquad (17)$$

Then, by [18, Theorem 16.4], the closure in (15) is superfluous so that

$$F^* = g_1^* \Box g_2^* \qquad (18)$$

and the infimum in the definition of the last infimal convolution is attained. Since $g_2 = \lambda_1\lambda_2(g \circ D)$ and $\operatorname{ri}\operatorname{dom} g \neq \varnothing$, [18, Theorem 6.7] implies that

$$\operatorname{ri}\operatorname{dom}(g \circ D) = \operatorname{ri}(D^{-1}\operatorname{dom} g) = D^{-1}\operatorname{ri}\operatorname{dom} g.$$

Hence

$$\operatorname{ri}\operatorname{dom} g_1 = \operatorname{ri}(\operatorname{dom} f_1 \times \operatorname{dom} f_2) = \operatorname{ri}\operatorname{dom} f_1 \times \operatorname{ri}\operatorname{dom} f_2$$

and

$$\operatorname{ri}\operatorname{dom} g_2 = D^{-1}\operatorname{ri}\operatorname{dom} g = \{(y_1,y_2) :\ y_1 - y_2 \in \operatorname{ri}\operatorname{dom} g\}.$$

Therefore, (17) is equivalent to

$$(\operatorname{ri}\operatorname{dom} f_1 - \operatorname{ri}\operatorname{dom} f_2) \cap \operatorname{ri}\operatorname{dom} g \neq \varnothing,$$

i.e., to (11). Using (14), (18), (16), and the equivalence

$$\frac{x^*}{\lambda_2} = \frac{\lambda_1 z_1}{\lambda_2} + z_2 \quad \Leftrightarrow \quad \frac{x^*}{\lambda_2} - \frac{z_1}{\lambda_2} = z_2 - z_1,$$

we obtain

$$\begin{aligned}
h^*(x^*) &= \left(g_1^* \Box g_2^*\right)(\lambda_1 x^*, \lambda_2 x^*) \\
&= \inf_{x^*=y_1+y_2} \{\lambda_1 f_1^*(\frac{y_1}{\lambda_1}) + \lambda_2 f_2^*(\frac{y_2}{\lambda_2}) + \lambda_1\lambda_2 g(\frac{x^*}{\lambda_2} - \frac{y_1}{\lambda_1\lambda_2})\} \\
&= \inf_{x^*=\lambda_1 z_1+\lambda_2 z_2} \{\lambda_1 f_1^*(z_1) + \lambda_2 f_2^*(z_2) + \lambda_1\lambda_2 g^*(\frac{x^*}{\lambda_2} - \frac{z_1}{\lambda_2})\} \\
&= \inf_{x^*=\lambda_1 z_1+\lambda_2 z_2} \{\lambda_1 f_1^*(z_1) + \lambda_2 f_2^*(z_2) + \lambda_1\lambda_2 g^*(z_2 - z_1)\}.
\end{aligned}$$

Therefore,

$$\big(P(\lambda_1, f_1, \lambda_2, f_2, g)\big)^* = P\big(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee}\big),$$

i.e., (12) holds. The exactness of (13) follows from the exactness of (18), which in turn is guaranteed by (11). ■

Corollary 2.3 *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$. Assume that both g and g^* have full domain. Then both $P(\lambda_1, f_1, \lambda_2, f_2, g)$ and $P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee})$ are convex, lower semicontinuous, proper, and*

$$\big(P(\lambda_1, f_1, \lambda_2, f_2, g)\big)^* = P\big(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee}\big).$$

In particular, for $g = \frac{1}{p}\|\cdot\|^p$ with $p > 1$, we have

$$\Big(P\big(\lambda_1, f_1, \lambda_2, f_2, \tfrac{1}{p}\|\cdot\|^p\big)\Big)^* = P\big(\lambda_1, f_1^*, \lambda_2, f_2^*, \tfrac{1}{q}\|\cdot\|^q\big),$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

Corollary 2.4 (epi-average vs arithmetic average) *Let $f_1, f_2 \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$, with $\lambda_1, \lambda_2 > 0$. The epigraphical average of f_1, f_2, which at $x \in X$ is defined by*

$$h(x) := \inf\left\{\lambda_1 f_1\Big(\frac{x_1}{\lambda_1}\Big) + \lambda_2 f_2\Big(\frac{x_2}{\lambda_2}\Big) : x_1 + x_2 = x\right\}, \tag{19}$$

has the conjugate

$$h^* = \lambda_1 f_1^* + \lambda_2 f_2^*. \tag{20}$$

That is, the conjugate of the epi-average is the arithmetic average of the conjugates. If $f_2 = f_1^*$ and $\lambda_1 = \lambda_2 = \frac{1}{2}$, then

$$h \le \tfrac{1}{2} f_1 + \tfrac{1}{2} f_1^* = h^*.$$

Proof. The conjugation formula (20) follows from Theorem 2.2 with $g = 0$. Putting $x_i = \lambda_i x$ for $i \in \{1, 2\}$ in (19), we see that

$$h \le \lambda_1 f_1 + \lambda_2 f_2.$$

In turn, if $f_2 = f_1^*$ and $\lambda_1 = \lambda_2 = \frac{1}{2}$, the result follows. ■

The following result places the kernel average between the convex hull and the arithmetic average.

Proposition 2.5 *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$. Assume that $g \ge 0$ and $g(0) = 0$, and set*

$$h = P(\lambda_1, f_1, \lambda_2, f_2, g).$$

Then

$$\operatorname{conv}\{f_1, f_2\} \le (\lambda_1 \star f_1) \Box (\lambda_2 \star f_2) \le h \le \lambda_1 f_1 + \lambda_2 f_2 \le \max\{f_1, f_2\}.$$

Proof. Fix $x \in X$. On the one hand,

$$h(x) \le \lambda_1 f_1(x) + \lambda_2 f_2(x) + \lambda_1\lambda_2 g(x-x) = \lambda_1 f_1(x) + \lambda_2 f_2(x) \le \big(\max\{f_1, f_2\}\big)(x).$$

On the other hand, using [18, Theorem 5.6], we estimate

$$\begin{aligned}
h(x) &= \inf_{x=\lambda_1 y_1+\lambda_2 y_2} \{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + g(y_1 - y_2)\} \\
&\ge \inf_{x=\lambda_1 y_1+\lambda_2 y_2} \{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2)\} \\
&= \inf_{x=x_1+x_2} \{\lambda_1 f_1(x_1/\lambda_1) + \lambda_2 f_2(x_2/\lambda_2)\} \\
&= \big((\lambda_1 \star f_1)\Box(\lambda_2 \star f_2)\big)(x) \\
&\ge \inf_{\substack{x=\mu_1 y_1+\mu_2 y_2 \\ \mu_1+\mu_2=1, \mu_1,\mu_2\ge 0}} \{\mu_1 f_1(y_1) + \mu_2 f_2(y_2)\} \\
&= \big(\operatorname{conv}\{f_1, f_2\}\big)(x).
\end{aligned}$$

Altogether, the proof is complete. ■

The next result localizes the domain of the kernel average. It shall be convenient to denote the diagonal in $X \times X$ by

$$\Delta := \{(x, x) : \ x \in X\}.$$

Theorem 2.6 (domain) *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$, and set*

$$h = P(\lambda_1, f_1, \lambda_2, f_2, g).$$

Then $\operatorname{dom} h = A \operatorname{dom} F = A[\big(((\operatorname{dom} g) \times \{0\}) + \Delta\big) \cap (\operatorname{dom} f_1 \times \operatorname{dom} f_2)]$ *and*

$$\inf h = \inf AF = \inf F = \inf_{y_1,y_2\in X} \{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2)\},$$

where A and F are as in (6) *and* (8), *respectively. If* $\operatorname{dom} g = X$, *then*

$$\operatorname{dom} h = \lambda_1 \operatorname{dom} f_1 + \lambda_2 \operatorname{dom} f_2.$$

Proof. The (set-valued) inverse of the difference operator D defined in (2) is

$$D^{-1}z = (z, 0) + \Delta \quad \text{for } z \in X.$$

Let g_1 and g_2 be as in the proof of Theorem 2.2. Then $F = g_1 + g_2$, $\operatorname{dom} F = \operatorname{dom} g_1 \cap \operatorname{dom} g_2$, $\operatorname{dom} g_1 = \operatorname{dom} f_1 \times \operatorname{dom} f_2$, and $\operatorname{dom} g_2 = D^{-1} \operatorname{dom} g = \big((\operatorname{dom} g) \times \{0\}\big) + \Delta$. The conclusion now follows from [24, Theorem 2.1.3(viii)]. If $\operatorname{dom} g = X$, then $\big((\operatorname{dom} g) \times \{0\}\big) + \Delta = X \times X$. ■

To study subdifferentiability properties of the kernel average, we need the following result.

Lemma 2.7 *Let $g \in \Gamma(X)$, $(x, y) \in X \times X$, and $(x^*, y^*) \in X^* \times X^*$. Then*

$$(x^*, y^*) \in \partial(g \circ D)(x, y) \qquad \Leftrightarrow \qquad x^* = -y^*, \ x^* \in \partial g(x - y).$$

Proof. "$\Rightarrow$": Since $(g \circ D)(x,y) = g(x-y)$, we have

$$g(x+u-(y+v)) \geq g(x-y) + \langle x^*, u\rangle + \langle y^*, v\rangle \quad \forall (u,v) \in X \times X. \tag{21}$$

Setting $v = u$ in (21), we obtain

$$g(x-y) \geq g(x-y) + \langle x^* + y^*, u\rangle \quad \forall u \in X,$$

and hence $x^* = -y^*$. Thus (21) becomes

$$g(x+u-(y+v)) \geq g(x-y) + \langle x^*, u\rangle + \langle -x^*, v\rangle, \text{ i.e.,}$$

$$g(x-y+u-v) \geq g(x-y) + \langle x^*, u-v\rangle \quad \forall u, v \in X.$$

Therefore, $x^* \in \partial g(x-y)$. "$\Leftarrow$": Reverse the above arguments. ■

Theorem 2.8 (subdifferential) *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$. Define*

$$f := P(\lambda_1, f_1, \lambda_2, f_2, g), \tag{22}$$

assume that

$$(\operatorname{ri}\operatorname{dom} f_1 - \operatorname{ri}\operatorname{dom} f_2) \cap \operatorname{ri}\operatorname{dom} g \neq \varnothing, \tag{23}$$

and that $x, y_1, y_2 \in X$ satisfy $x = \lambda_1 y_1 + \lambda_2 y_2$.

(i) *If*

$$f(x) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2), \tag{24}$$

then

$$\partial f(x) = \big\{(\partial f_1(y_1) + \lambda_2 y^*) \cap (\partial f_2(y_2) - \lambda_1 y^*) : \ y^* \in \partial g(y_1 - y_2)\big\}. \tag{25}$$

(ii) *Conversely, if there exists $y^* \in \partial g(y_1 - y_2)$ such that*

$$(\partial f_1(y_1) + \lambda_2 y^*) \cap (\partial f_2(y_2) - \lambda_1 y^*) \neq \varnothing, \tag{26}$$

then

$$f(x) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2) \tag{27}$$

and hence the infimal convolution implicit in the definition of f is exact at x.

Proof. Using (9), we recall that $f = AF$ where A and F are given by (6) and (8), respectively. Now assume (24). By [24, Corollary 2.4.6],

$$\partial f(x) = \partial(AF)(x) = (A^*)^{-1}\partial F(y),$$

where $Ay = x$ and $y = (y_1, y_2) \in X \times X$. Using (7), we obtain

$$x^* \in \partial f(x) \quad \Leftrightarrow \quad (\lambda_1 x^*, \lambda_2 x^*) \in \partial F(y_1, y_2).$$

Since (23) holds, the sum rule [18, Theorem 23.8] and Lemma 2.7 imply

$$\begin{aligned}\partial F(y_1,y_2) &= \big(\lambda_1\partial f_1(y_1)\times\lambda_2\partial f_2(y_2)\big)+\lambda_1\lambda_2\partial(g\circ D)(y_1,y_2)\\ &= \big(\lambda_1\partial f_1(y_1)\times\lambda_2\partial f_2(y_2)\big)+\lambda_1\lambda_2\big\{(y^*,-y^*):\ y^*\in\partial g(y_1-y_2)\big\}.\end{aligned}$$

Consequently, $x^*\in\partial f(x)$ if and only if

$$(\lambda_1x^*,\lambda_2x^*)\in(\lambda_1\partial f_1(y_1)+\lambda_1\lambda_2y^*)\times(\lambda_2\partial f_2(y_2)-\lambda_1\lambda_2y^*),$$

for some $y^*\in\partial g(y_1-y_2)$. This is equivalent to the existence of $y^*\in\partial g(y_1-y_2)$ such that

$$x^*\in\partial f_1(y_1)+\lambda_2y^*,\quad x^*\in\partial f_2(y_2)-\lambda_1y^*,$$

i.e.,

$$x^*\in(\partial f_1(y_1)+\lambda_2y^*)\cap(\partial f_2(y_2)-\lambda_1y^*).$$

Hence (25) holds and (i) is verified.

Now assume that $y^*\in\partial g(y_1-y_2)$ satisfies (26) so that

$$x^*-\lambda_2y^*\in\partial f_1(y_1),\quad x^*+\lambda_1y^*\in\partial f_2(y_2),$$

for some $x^*\in X$. Let $z_1,z_2\in X$ be arbitrary. Then

$$f_1(z_1)\ge f_1(y_1)+\langle x^*-\lambda_2y^*,z_1-y_1\rangle, \tag{28}$$

$$f_2(z_2)\ge f_2(y_2)+\langle x^*+\lambda_1y^*,z_2-y_2\rangle. \tag{29}$$

Multiplying (28) by λ_1 and (29) by λ_2 yields

$$\lambda_1f_1(z_1)\ge\lambda_1f_1(y_1)+\langle x^*,\lambda_1(z_1-y_1)\rangle-\lambda_1\lambda_2\langle y^*,z_1-y_1\rangle,$$

$$\lambda_2f_2(z_2)\ge\lambda_2f_2(y_2)+\langle x^*,\lambda_2(z_2-y_2)\rangle+\lambda_1\lambda_2\langle y^*,z_2-y_2\rangle.$$

Adding these two inequalities, followed by adding $\lambda_1\lambda_2g(z_1-z_2)$ to both sides, we obtain

$$\begin{aligned}\lambda_1f_1(z_1)+\lambda_2f_2(z_2)+\lambda_1\lambda_2g(z_1-z_2)\ge{}&\lambda_1f_1(y_1)+\lambda_2f_2(y_2)+\lambda_1\lambda_2g(z_1-z_2)\\ &-\lambda_1\lambda_2\langle y^*,(z_1-z_2)-(y_1-y_2)\rangle\\ &+\langle x^*,\lambda_1(z_1-y_1)+\lambda_2(z_2-y_2)\rangle.\end{aligned}$$

Since $y^*\in\partial g(y_1-y_2)$, it follows that

$$\begin{aligned}\lambda_1f_1(z_1)+\lambda_2f_2(z_2)+\lambda_1\lambda_2g(z_1-z_2)\ge{}&\lambda_1f_1(y_1)+\lambda_2f_2(y_2)+\lambda_1\lambda_2g(y_1-y_2)\\ &+\langle x^*,(\lambda_1z_1+\lambda_2z_2)-(\lambda_1y_1+\lambda_2y_2)\rangle.\end{aligned}$$

Taking the infimum on both sides of the last inequality over all $(z_1,z_2)\in X\times X$ such that $\lambda_1z_1+\lambda_2z_2$ is equal to some fixed but arbitrary $z\in X$, we deduce that

$$f(z)\ge\lambda_1f_1(y_1)+\lambda_2f_2(y_2)+\lambda_1\lambda_2g(y_1-y_2)+\langle x^*,z-x\rangle;$$

the choice $z = x$ yields

$$f(x) \geq \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2).$$

On the other hand, by definition of f,

$$f(x) \leq \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2).$$

Altogether, $f(x) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2)$, i.e., (27) holds and (ii) is verified. ■

Corollary 2.9 *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$, and set $f = P(\lambda_1, f_1, \lambda_2, f_2, g)$. Assume that g is differentiable everywhere and that g^* has full domain. Then for every $x \in \operatorname{dom} f$, there exist $y_1, y_2 \in X$ such that*

$$x = \lambda_1 y_1 + \lambda_2 y_2, \quad f(x) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2), \tag{30}$$

and

$$\partial f(x) = \big(\partial f_1(y_1) + \lambda_2 \nabla g(y_1 - y_2)\big) \cap \big(\partial f_2(y_2) - \lambda_1 \nabla g(y_1 - y_2)\big). \tag{31}$$

In particular, this result holds when $g = \frac{1}{p}\|\cdot\|^p$ with $p > 1$.

Proof. Since $\operatorname{dom} g^{*\vee} = X$ and $g^{*\vee*\vee} = g$, Theorem 2.2 implies that

$$\big(P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee})\big)^* = P(\lambda_1, f_1, \lambda_2, f_2, g),$$

and is exact. This gives (30). Now (31) follows from (30) and Theorem 2.8(i). ■

The notions of essentially smooth, essentially strictly convex, and Legendre type are carefully studied in [18, Section 26].

Corollary 2.10 (Legendre functions) *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$, and set $f = P(\lambda_1, f_1, \lambda_2, f_2, g)$. Assume that both g and g^* are differentiable everywhere.*

(i) *If f_1 or f_2 is essentially smooth, then f is essentially smooth.*

(ii) *If f_1 or f_2 is essentially strictly convex, then f is essentially strictly convex.*

(iii) *If f_1 or f_2 is essentially strictly convex, and f_1 or f_2 is essentially smooth, then f is both essentially strictly convex and essentially smooth, i.e., Legendre type.*

Proof. (i): Recall that a function in $\Gamma(X)$ is essentially smooth if and only if its subdifferential operator is at most single-valued; see [18, Theorem 26.1]. Assume that f_1 is essentially smooth, so that ∂f_1 is at most single-valued. By Corollary 2.9, ∂f is at most single-valued. Therefore, f is essentially smooth. (ii): Recall that a function in $\Gamma(X)$ is essentially strictly convex if and only if its conjugate is essentially smooth; see [18, Theorem 26.3]. Assume that f_1 is essentially strictly convex. Then f_1^* is essentially smooth and so is $P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee})$ by (i). The conjugate of the last function is not only essentially strictly convex but also equal to f by Theorem 2.2. (iii): By definition, a function in $\Gamma(X)$ is of Legendre type if it is both essentially smooth and essentially strictly convex. Hence the result follows by combining (i) and (ii). ■

Our next result concerns antiderivatives of cyclically monotone operators; see, e.g., [18, Section 24] and [20] for background material.

Corollary 2.11 *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$, and set $f = P(\lambda_1, f_1, \lambda_2, f_2, g)$. Assume that g has full domain, $g \geq 0$, and $g(0) = 0$. Let $x \in X$. Then*

$$\partial f_1(x) \cap \partial f_2(x) \subset \partial f(x). \tag{32}$$

Proof. If $\partial f_1(x) \cap \partial f_2(x) = \varnothing$, then (32) clearly holds. Thus assume that $\partial f_1(x) \cap \partial f_2(x) \neq \varnothing$ and set $y_1 := y_2 := x$ and $y^* := 0 \in \partial g(0) = g(y_1 - y_2)$. Then (26) is true and so Theorem 2.8(ii) yields $f(x) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2)$. In turn, Theorem 2.8(i) implies that $\partial f(x) \supset \big(\partial f_1(y_1) + \lambda_2 y^*\big) \cap \big(\partial f_2(y_2) - \lambda_1 y^*\big) = \partial f_1(x) \cap \partial f_2(x)$. ■

Remark 2.12 Consider the setting of Corollary 2.11 and assume that $A : X \rightrightarrows X^*$ is cyclically monotone and $A(x) \subset \partial f_1(x) \cap \partial f_2(x)$, $\forall x \in X$. Then Corollary 2.11 guarantees that $f = P(\lambda_1, f_1, \lambda_2, f_2, g)$ is an antiderivative of A and hence Theorem 2.2 implies that $f^* = P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee})$ is an antiderivative of A^{-1}. Primal-dual symmetric methods for generating antiderivatives were recently investigated in [5].

3 Indicator Functions and Minimizers

We now turn to the kernel average of two indicator functions. When $g = \frac{1}{2}\|\cdot\|^2$ and $\|\cdot\|$ is the Euclidean norm, then the following result becomes [6, formula (32)].

Example 3.1 (indicator functions) Let C_1 and C_2 be two nonempty closed convex subsets of X, let $g \in \Gamma(X)$, and set $f = P(\lambda_1, \iota_{C_1}, \lambda_2, \iota_{C_2}, g)$. Then

$$f(x) = \lambda_1\lambda_2 \inf_{y_1 \in C_1, y_2 \in C_2} g(y_1 - y_2) = \lambda_1\lambda_2 \inf_{y \in \lambda_1(x - C_1) \cap \lambda_2(C_2 - x)} g\Big(-\frac{y}{\lambda_1\lambda_2}\Big). \tag{33}$$

If $g = \frac{1}{p}\|\cdot\|^p$ and $p > 1$, then $\operatorname{dom} f = \lambda_1 C_1 + \lambda_2 C_2$ and

$$f(x) = \frac{1}{(\lambda_1\lambda_2)^{p-1}} \inf_{y \in \lambda_1(x - C_1) \cap \lambda_2(C_2 - x)} \frac{1}{p}\|0 - y\|^p = \frac{1}{(\lambda_1\lambda_2)^{p-1}} \frac{1}{p} d^p_{\lambda_1(x - C_1) \cap \lambda_2(C_2 - x)}(0). \tag{34}$$

Proof. Fix $x \in X$. By definition of the kernel average, we have

$$f(x) = \inf_{\substack{x = \lambda_1 y_1 + \lambda_2 y_2 \\ y_1 \in C_1, y_2 \in C_2}} \lambda_1\lambda_2 g(y_1 - y_2).$$

If $x = \lambda_1 y_1 + \lambda_2 y_2$, then $y_2 - y_1 = (\lambda_1 x - \lambda_1 y_1)/(\lambda_1\lambda_2)$, and $(y_1, y_2) \in C_1 \times C_2$ if and only if $\lambda_1 y_1 \in \lambda_1 C_1 \cap (x - \lambda_2 C_2)$. Hence

$$f(x) = \inf_{\lambda_1 y_1 \in \lambda_1 C_1 \cap (x-\lambda_2 C_2)} \lambda_1\lambda_2 g\Big(-\frac{\lambda_1 x - \lambda_1 y_1}{\lambda_1\lambda_2}\Big) = \lambda_1\lambda_2 \inf_{z \in \lambda_1 C_1 \cap (x-\lambda_2 C_2)} g\Big(-\frac{\lambda_1 x - z}{\lambda_1\lambda_2}\Big).$$

Changing variables to $y = \lambda_1 x - z$ in the last infimum and observing that $z \in \lambda_1 C_1 \cap (x - \lambda_2 C_2)$ if and only if $y \in \lambda_1(x - C_1) \cap \lambda_2(C_2 - x)$, we obtain the more symmetric formula

$$f(x) = \lambda_1\lambda_2 \inf_{y \in \lambda_1(x-C_1)\cap\lambda_2(C_2-x)} g\Big(-\frac{y}{\lambda_1\lambda_2}\Big),$$

i.e., (33). Now suppose that $g = \frac{1}{p}\|\cdot\|^p$, where $p > 1$. Then $\operatorname{dom} g = X$ and Theorem 2.6 yields that $\operatorname{dom} f = \lambda_1 \operatorname{dom} \iota_{C_1} + \lambda_2 \operatorname{dom} \iota_{C_2} = \lambda_1 C_1 + \lambda_2 C_2$. Finally, (34) is an immediate consequence of (33). ■

Let $f \in \Gamma(X)$. Recall that $\inf f = \inf f(X)$, that $\operatorname{argmin} f = \{x \in X : f(x) = \inf f\}$, and that $\min f = \inf f$ provided that $\operatorname{argmin} f \neq \varnothing$.

Proposition 3.2 (minimizers) *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$, $\lambda_1, \lambda_2 > 0$, and set $f = P(\lambda_1, f_1, \lambda_2, f_2, g)$. Assume that $g \geq 0$. Then the following hold.*

(i) $\inf f \geq \lambda_1 \inf_1 + \lambda_2 \inf f_2$.

(ii) *If $g(0) = 0$ and $\operatorname{argmin} f_1 \cap \operatorname{argmin} f_2 \neq \varnothing$, then $\operatorname{argmin} f_1 \cap \operatorname{argmin} f_2 \subset \operatorname{argmin} f$ and $\min f = \lambda_1 \min f_1 + \lambda_2 \min f_2$.*

(iii) *If g^* has full domain, $\{x \in X : g(x) = 0\} = \{0\}$, and $\min f = \lambda_1 \min f_1 + \lambda_2 \min f_2$, then $\operatorname{argmin} f \subset \operatorname{argmin} f_1 \cap \operatorname{argmin} f_2$.*

Proof. (i): Indeed, for every $z \in X$, we have

$$f(z) \geq \inf_{z=\lambda_1 y_1 + \lambda_2 y_2} (\lambda_1 \inf f_1 + \lambda_2 \inf f_2 + g(y_1 - y_2) \geq \lambda_1 \inf f_1 + \lambda_2 \inf f_2.$$

The desired inequality now follows by infimizing over $z \in X$. (ii): Take $x \in \operatorname{argmin} f_1 \cap \operatorname{argmin} f_2$. Then

$$f(x) \leq \lambda_1 f_1(x) + \lambda_2 f_2(x) + \lambda_1\lambda_2 g(x - x) = \lambda_1 \min f_1 + \lambda_2 \min f_2.$$

On the other hand, (i) implies that $f(x) \geq \inf f \geq \lambda_1 \min f_1 + \lambda_2 \min f_2$. Altogether, we obtain $f(x) = \min f = \lambda_1 \min f_1 + \lambda_2 \min f_2$. (iii): Take $x \in \operatorname{argmin} f$. On the one hand,

$$f(x) = \lambda_1 \min f_1 + \lambda_2 \min f_2. \tag{35}$$

On the other hand, Theorem 2.2 (applied to $f_1^*, f_2^*, g^{*\vee}$) shows that there exist $y_1, y_2 \in X$ such that

$$x = \lambda_1 y_1 + \lambda_2 y_2 \quad \text{and} \quad f(x) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1\lambda_2 g(y_1 - y_2). \tag{36}$$

Combining (35) and (36), we deduce that $f_1(y_1) = \min f_1$, $f_2(y_2) = \min f_2(y_2)$ and $g(y_1 - y_2) = 0$. Hence $y_1 - y_2 = 0$. Thus $y_1 = y_2 = x$, $f_1(x) = \min f_1$ and $f_2(x) = \min f_2$. ■

Corollary 3.3 *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$, $\lambda_1, \lambda_2 > 0$, and set $f = P(\lambda_1, f_1, \lambda_2, f_2, g)$. Assume that $g \geq 0$, $\{x \in X : g(x) = 0\} = \{0\}$, and that g^* has full domain. Then*

$$\min f = \lambda_1 \min f_1 + \lambda_2 \min f_2 \quad \Leftrightarrow \quad \operatorname{argmin} f_1 \cap \operatorname{argmin} f_2 \neq \varnothing,$$

in which case $\operatorname{argmin} f = \operatorname{argmin} f_1 \cap \operatorname{argmin} f_2$. *Consequently, if f, f_1, f_2 all have minimizers but* $\operatorname{argmin} f_1 \cap \operatorname{argmin} f_2 = \varnothing$, *then* $\min f > \lambda_1 \min f_1 + \lambda_2 \min f_2$.

4 Proximal Mappings

In this section, we assume that the kernel function $g \in \Gamma(X)$ is differentiable everywhere, that g is uniformly convex on bounded convex subsets of X, and that g is supercoercive, i.e., $\lim_{\|x\| \to +\infty} g(x)/\|x\| = +\infty$. See [24] for further information on these notions. Since X is finite-dimensional, [24, Proposition 3.6.6(i) and Lemma 3.6.1] imply the following equivalent requirement:

$g \in \Gamma(X)$ is differentiable and strictly convex everywhere, and g^* has full domain.

Definition 4.1 (proximal mapping) *Let $f \in \Gamma(X)$. The g-*proximal mapping *of f is*

$$P_g f = (\partial f + \nabla g)^{-1} : X^* \rightrightarrows X.$$

Lemma 4.2 *Let $f \in \Gamma(X)$. Then* $\operatorname{dom} P_g f = \operatorname{ran}(\partial f + \nabla g) = X^*$, *$P_g f$ is single-valued, and $x^* \in \partial f(x)$ if and only if $x = P_g f(x^* + \nabla g(x))$.*

Proof. On the one hand, $f + g$ is uniformly convex on bounded convex subsets of X. On the other hand, since f possesses a continuous affine minorant, $f + g$ is supercoercive. Altogether, [24, Corollary 3.5.9] implies that $\operatorname{ran}(\partial(f + g)) = X^*$. It follows that $\operatorname{dom} P_g f = \operatorname{ran}(\partial f + \nabla g) = \operatorname{ran} \partial(f + g) = X^*$. Now fix $x^* \in X^*$ and take $x_1, x_2 \in P_g f(x^*)$. Then each $x_i \in (\partial f + \nabla g)^{-1}(x^*)$ $\Leftrightarrow x^* \in (\partial f + \nabla g)(x_i) \Rightarrow x^* - \nabla g(x_i) \in \partial f(x_i)$. Since ∂f is monotone, we get

$$\langle x^* - \nabla g(x_1) - (x^* - \nabla g(x_2)), x_1 - x_2 \rangle \geq 0,$$

i.e., $\langle \nabla g(x_1) - \nabla g(x_2), x_1 - x_2 \rangle \leq 0$. Since g is strictly convex, the gradient operator $\nabla g : X \to X^*$ is strictly monotone (see [24, Theorem 2.4.4(ii)]), and we thus conclude that $x_1 = x_2$. Hence $P_g f$ is single-valued. Therefore, $x^* \in \partial f(x) \Leftrightarrow x^* + \nabla g(x) \in \partial f(x) + \nabla g(x) = (\partial f + \nabla g)(x) \Leftrightarrow x = (\partial f + \nabla g)^{-1}(x^* + \nabla g(x)) = P_g f(x^* + \nabla g(x))$. ■

The following result relates the proximal mapping $P_g f$ to $P_g f_1$ and $P_g f_2$.

Theorem 4.3 *Let $f_1, f_2 \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$, and set $f = P(\lambda_1, f_1, \lambda_2, f_2, g)$. Then for every $x \in \operatorname{dom} \partial f$, there exist $y_1 \in \operatorname{dom} \partial f_1$ and $y_2 \in \operatorname{dom} \partial f_2$ such that for every $x^* \in \partial f(x)$, we have*

$$x = P_g f\big(x^* + \nabla g(x)\big) \tag{37}$$

$$= \lambda_1 P_g f_1\big(x^* - \lambda_2 \nabla g(y_1 - y_2) + \nabla g(y_1)\big) + \lambda_2 P_g f_2\big(x^* + \lambda_1 \nabla g(y_1 - y_2) + \nabla g(y_2)\big). \tag{38}$$

Moreover, for every $z^* \in X^*$*, there exist* $y_1 \in \operatorname{dom} \partial f_1$ *and* $y_2 \in \operatorname{dom} \partial f_2$ *such that*

$$P_g f(z^*) = \lambda_1 P_g f_1\big(z^* - \nabla g(x) - \lambda_2 \nabla g(y) + \nabla g(y_1)\big) + \lambda_2 P_g f_2\big(z^* - \nabla g(x) + \lambda_1 \nabla g(y) + \nabla g(y_2)\big), \quad (39)$$

where $x := \lambda_1 y_1 + \lambda_2 y_2 = P_g f(z^*)$ *and* $y := y_1 - y_2$*.*

Proof. Take $x \in \operatorname{dom} \partial f$. By Corollary 2.9, there exist $y_1 \in \operatorname{dom} f_1$ and $y_2 \in \operatorname{dom} f_2$ such that

$$x = \lambda_1 y_1 + \lambda_2 y_2, \quad (40)$$

and

$$\partial f(x) = (\partial f_1(y_1) + \lambda_2 \nabla g(y)) \cap (\partial f_2(y_2) - \lambda_1 \nabla g(y)),$$

where $y := y_1 - y_2$. Now take $x^* \in \partial f(x)$. Then $x^* - \lambda_2 \nabla g(y) \in \partial f_1(y_1)$ and $x^* + \lambda_1 \nabla g(y) \in \partial f_2(y_2)$. Hence $y_1 \in \operatorname{dom} \partial f$ and $y_2 \in \operatorname{dom} \partial f_2$. By Lemma 4.2,

$$y_1 = P_g f_1(x^* - \lambda_2 \nabla g(y) + \nabla g(y_1)) \quad \text{and} \quad y_2 = P_g f_2(x^* + \lambda_1 \nabla g(y) + \nabla g(y_2)). \quad (41)$$

On the other hand, Lemma 4.2 also implies that

$$x = P_g f(x^* + \nabla g(x)). \quad (42)$$

Altogether, we obtain (37)–(38) by combining (42), (40), and (41). Now take $z^* \in X^*$. Since $\operatorname{ran}(\partial f + \nabla g) = X^*$ by Lemma 4.2, there exists $x \in X$ such that $z^* \in \partial f(x) + \nabla g(x)$, i.e., $z^* - \nabla g(x) \in \partial f(x)$. Lemma 4.2 yields $x = P_g f(z^*)$. Therefore, (39) follows from (37)–(38). ■

We are now able to recover a known result about the proximal average; see [6, Theorem 6.1].

Corollary 4.4 *Let* $f_1, f_2 \in \Gamma(X)$*,* $\lambda_1 + \lambda_2 = 1$ *with* $\lambda_1, \lambda_2 > 0$*, and set* $f = P(\lambda_1, f_1, \lambda_2, f_2, g)$*, where* $g = \frac{1}{2}\|\cdot\|^2$ *and* $\|\cdot\|$ *is the Euclidean norm. Then*

$$P_g f = \lambda_1 P_g f_1 + \lambda_2 P_g f_2. \quad (43)$$

Proof. Take $z^* \in X^* = X$ and let y_1, y_2, x, y be as in Theorem 4.3. Since $\nabla g = \mathrm{Id}$, we obtain

$$z^* - \nabla g(x) - \lambda_2 \nabla g(y) + \nabla g(y_1) = z^* - (\lambda_1 y_1 + \lambda_2 y_2) - \lambda_2 (y_1 - y_2) + y_1 = z^*$$

and

$$z^* - \nabla g(x) + \lambda_1 \nabla g(y) + \nabla g(y_2) = z^* - (\lambda_1 y_1 + \lambda_2 y_2) + \lambda_1 (y_1 - y_2) + y_2 = z^*.$$

Hence (39) transpires to $P_g f(z^*) = \lambda_1 P_g f_1(z^*) + \lambda_2 P_g f_2(z^*)$. ■

5 Application to Monotone Operators

From now on, we assume that

X is a reflexive real Banach space, with norm $\|\cdot\|$, dual space X^*, and dual norm $\|\cdot\|_*$.

We shall study the kernel average — defined as in Definition 1.1 — in this setting and then apply it in the product space $X \times X^*$ to explicitly describe a maximal monotone extension of any given monotone operator. The kernel average can also be used to represent maximal monotone operators. Representations with attractive duality properties were studied previously by Svaiter (see [23]), by Penot (see [14, 15]), by Penot and Zălinescu ("autoconjugates", see [16]), and by Ghoussoub ("selfdual Lagrangians", see [11]). The works by Svaiter, by Penot, and by Ghoussoub were not explicit in the sense that either Zorn's Lemma or transfinite induction was utilized. Ghoussoub also imposed separability on the underlying space. Although explicit, the construction by Penot and Zălinescu required a constraint qualification. In this section, we shall provide an explicit construction without any constraint qualification, in the present setting of reflexive real Banach spaces.

We start with a variant of Theorem 2.2 adapted to the present setting. The proof, while similar to the one of Theorem 2.2, is included for completeness and for the reader's convenience. Given $f \in \Gamma(X)$, we denote by $\operatorname{cont} f$ the set of points at which f is finite and continuous.

Theorem 5.1 *Let $f_1, f_2, g \in \Gamma(X)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$. Then the following hold.*

(i) *If $(\operatorname{dom} f_1 - \operatorname{dom} f_2) \cap \operatorname{cont} g \neq \varnothing$, then*

$$\big(P(\lambda_1, f_1, \lambda_2, f_2, g)\big)^* = P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee}) \tag{44}$$

belongs to $\Gamma(X^*)$ and the infimum in the definition of $P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee})$ is attained.

(ii) *If $(\operatorname{dom} f_1^* - \operatorname{dom} f_2^*) \cap \operatorname{cont} g^{*\vee} \neq \varnothing$, then*

$$\big(P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee})\big)^* = P(\lambda_1, f_1, \lambda_2, f_2, g) \tag{45}$$

belongs to $\Gamma(X)$ and the infimum in the definition of $P(\lambda_1, f_1, \lambda_2, f_2, g)$ is attained.

Proof. (i): As in Section 2, we define $A\colon X \times X \to X\colon (x_1, x_2) \mapsto x_1 - x_2$, $F\colon X \times X \to\,]-\infty, +\infty]\colon (x_1, x_2) \mapsto \lambda_1 f_1(x_1) + \lambda_2 f_2(x_2) + \lambda_1\lambda_2 g(x_1 - x_2)$ so that $A^*\colon X^* \to X^* \times X^*\colon x^* \mapsto (\lambda_1 x^*, \lambda_2 x^*)$ and

$$P(\lambda_1, f_1, \lambda_2, f_2, g) = AF. \tag{46}$$

Furthermore,

$$F = g_1 + g_2,$$

where we define g_1 and g_2 on $X \times X$ by

$$g_1(x_1, x_2) = \lambda_1 f_1(x_1) + \lambda_2 f_2(x_2), \quad g_2(x_1, x_2) = \lambda_1\lambda_2 g(x_1 - x_2).$$

Assume that $(\operatorname{dom} f_1 - \operatorname{dom} f_2) \cap \operatorname{cont} g \neq \varnothing$, which implies that $\operatorname{dom} g_1 \cap \operatorname{cont} g_2 \neq \varnothing$. Using [24, Theorem 2.8.7(iii)], we see that for every $(x_1^*, x_2^*) \in X^* \times X^*$,

$$F^*(x_1^*, x_2^*) = \min_{(y_1^*, y_2^*) \in X^* \times X^*} g_1^*(y_1^*, y_2^*) + g_2^*(x_1^* - y_1^*, x_2^* - y_2^*). \tag{47}$$

Now for every $(x_1^*, x_2^*) \in X^* \times X^*$, we clearly have $g_1^*(x_1^*, x_2^*) = \lambda_1 f_1^*(x_1^*/\lambda_1) + \lambda_2 f_2^*(x_2^*/\lambda_2)$, and also $g_2^*(x_1^*, x_2^*) = \lambda_1\lambda_2 g^*(x_1^*/(\lambda_1\lambda_2)) + \iota_{\{0\}}(x_1^* + x_2^*)$ by Lemma 2.1. Using (46), [24, Theorem 2.3.1(ix)], and (47), we thus obtain for every $x^* \in X^*$,

$$\begin{aligned}
\big(P(\lambda_1, f_1, \lambda_2, f_2, g)\big)^*(x^*) &= (AF)^*(x^*) \\
&= F^*(A^*x^*) \\
&= \min_{y_1^*, y_2^*} g_1^*(y_1^*, y_2^*) + g_2^*(\lambda_1 x^* - y_1^*, \lambda_2 x^* - y_2^*) \\
&= \min_{y_1^* + y_2^* = x^*} \lambda_1 f_1^*(y_1^*/\lambda_1) + \lambda_2 f_2^*(y_2^*/\lambda_2) + \lambda_1\lambda_2 g^*\Big(\frac{\lambda_1 x^* - y_1^*}{\lambda_1\lambda_2}\Big) \\
&= \min_{\lambda_1 z_1^* + \lambda_2 z_2^* = x^*} \lambda_1 f_1^*(z_1^*) + \lambda_2 f_2^*(z_2^*) + \lambda_1\lambda_2 g^*\Big(\frac{x^* - z_1^*}{\lambda_2}\Big) \\
&= \min_{\lambda_1 z_1^* + \lambda_2 z_2^* = x^*} \lambda_1 f_1^*(z_1^*) + \lambda_2 f_2^*(z_2^*) + \lambda_1\lambda_2 g^{*\vee}(z_1^* - z_2^*) \\
&= P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee})(x^*).
\end{aligned}$$

(ii): Apply (i) to $f_1^*, f_2^*, g^{*\vee}$. ■

We now record a special case of Theorem 5.1 in the product space $X \times X^*$, the dual of which is $X^* \times X$.

Corollary 5.2 *Let $f_1, f_2, g \in \Gamma(X \times X^*)$, $\lambda_1 + \lambda_2 = 1$ with $\lambda_1, \lambda_2 > 0$, and assume that both g and g^* have full domain. Then $P(\lambda_1, f_1, \lambda_2, f_2, g) \in \Gamma(X \times X^*)$, $P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee}) \in \Gamma(X^* \times X)$, and*

$$\big(P(\lambda_1, f_1, \lambda_2, f_2, g)\big)^* = P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{*\vee}).$$

In particular, if $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\Big(P\big(\lambda_1, f_1, \lambda_2, f_2, \tfrac{1}{p}\|\cdot\|^p \oplus \tfrac{1}{q}\|\cdot\|_*^q\big)\Big)^* = P\big(\lambda_1, f_1^*, \lambda_2, f_2^*, \tfrac{1}{q}\|\cdot\|_*^q \oplus \tfrac{1}{p}\|\cdot\|^p\big).$$

From now on, we assume that

$$A\colon X \rightrightarrows X^* \quad \text{is a monotone operator,}$$

i.e., for all $(x, x^*), (y, y^*) \in \operatorname{gra} A := \{(z, z^*) \in X \times X^* :\ z^* \in Az\}$, we have $\langle x^* - y^*, x - y\rangle \geq 0$, and that $\operatorname{gra} A \neq \varnothing$. The operator A is said to be maximal monotone, if it is not possible to enlarge A without destroying monotonicity. Monotone operators play an important role in modern optimization and analysis; see, e.g., [19, 20]. The *Fitzpatrick function* [10] associated with A has recently turned out to be a very effective tool for analyzing A; see, e.g., [1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 21, 22, 23].

Definition 5.3 *The* Fitzpatrick function *associated with A is the function $F_A \in \Gamma(X \times X^*)$ defined at $(x, x^*) \in X \times X^*$ by*

$$F_A(x, x^*) := \sup_{(a,a^*) \in \operatorname{gra} A} \big(\langle a^*, x \rangle + \langle x^*, a \rangle - \langle a^*, a \rangle \big). \tag{48}$$

Given $f \in \Gamma(X^* \times X)$, it will be convenient to define its "transpose" function $f^\intercal \in \Gamma(X, X^*)$ by

$$f^\intercal(x, x^*) = f(x^*, x),$$

and similarly for $f \in \Gamma(X \times X^*)$. We can now simply express the Fitzpatrick function of the inverse of A by

$$F_{A^{-1}} = F_A^\intercal. \tag{49}$$

As in [14, 15, 16], we say that

$$f \text{ is autoconjugate} \quad \Leftrightarrow \quad f^\intercal = f^*;$$

in [11], f is then called a "selfdual Lagrangian".

Fact 5.4 *Let $(x, x^*) \in X \times X^*$. Then the following hold.*

(i) $F_A \leq F_A^{*\intercal}$.

(ii) *If $(x, x^*) \in \operatorname{gra} A$, then $F_A(x, x^*) = F_A^*(x^*, x) = \langle x^*, x \rangle$ and $(x^*, x) \in \partial F_A(x, x^*)$.*

(iii) *If A is maximal monotone and $(x, x^*) \notin \operatorname{gra} A$, then $F_A(x, x^*) > \langle x^*, x \rangle$.*

Proof. (i): See [10, Proposition 4.2]. (ii): See [10, Theorem 3.4 and Proposition 4.2]. (iii): See [10, Theorem 3.8]. ■

If A is maximal monotone, then (by Fact 5.4(ii)–(iii)) it is possible to represent A by F_A in the sense that $\operatorname{gra} A = \{(x, x^*) \in X \times X^* : F_A(x, x^*) = \langle x^*, x \rangle\}$. We now show how the kernel average can be utilized to construct extensions and autoconjugate representations. To this end, we let the kernel function in $\Gamma(X \times X^*)$ be given by

$$g_p(x, x^*) := \tfrac{1}{p}\|x\|^p + \tfrac{1}{q}\|x^*\|_*^q,$$

where $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Observe that $g_p(x, x^*) = g_q(x^*, x)$, when $g_q \in \Gamma(X^* \times X)$. Our analysis will carry through with more general kernels, but this particular choice of the kernel is sufficient for the construction of autoconjugate representations. Note that

$$g_p^{*\vee} = g_p^* = g_p^\intercal, \tag{50}$$

which itself is autoconjugate and which yields the following recipe for constructing autoconjugates.

Lemma 5.5 *Let $f \in \Gamma(X \times X^*)$. Then*

$$\Big(P\big(\tfrac{1}{2}, f, \tfrac{1}{2}, f^{*\intercal}, g_p\big)\Big)^* = P\big(\tfrac{1}{2}, f^*, \tfrac{1}{2}, f^\intercal, g_p^\intercal\big)$$

and $P\big(\tfrac{1}{2}, f, \tfrac{1}{2}, f^{*\intercal}, g_p\big)$ is autoconjugate.

Proof. Using Corollary 5.2 and (50), we have

$$\begin{aligned}\big(P(\tfrac{1}{2}, f, \tfrac{1}{2}, f^{*\intercal}, g_p)\big)^* &= P(\tfrac{1}{2}, f^*, \tfrac{1}{2}, f^{*\intercal *}, g_p^{*\vee}) = P(\tfrac{1}{2}, f^*, \tfrac{1}{2}, f^\intercal, g_p^\intercal)\\ &= P(\tfrac{1}{2}, f^\intercal, \tfrac{1}{2}, f^*, g_p^\intercal) = P(\tfrac{1}{2}, f^\intercal, \tfrac{1}{2}, f^{*\intercal\intercal}, g_p^\intercal)\\ &= \big(P(\tfrac{1}{2}, f, \tfrac{1}{2}, f^{*\intercal}, g_p)\big)^\intercal.\end{aligned}$$

The proof is complete. ■

Let $f \in \Gamma(X \times X^*)$. Define $G(f)\colon X \rightrightarrows X^*$ by

$$x^* \in G(f)x \quad \text{if and only if} \quad (x^*, x) \in \partial f(x, x^*). \tag{51}$$

This operator was first considered by Fitzpatrick [10, Section 2]. The following result shows that $G(f)$ has very nice properties if f is autoconjugate.

Fact 5.6 *Let $f \in \Gamma(X \times X^*)$ be autoconjugate. Then $G(f)$ is maximal monotone, and $x^* \in G(f)x$ if and only if $f(x, x^*) = \langle x^*, x\rangle$.*

Proof. (See [11, 16].) Let $(x, x^*) \in X \times X^*$. Since f is autoconjugate, we have the equivalences $x^* \in G(f)x \Leftrightarrow (x^*, x) \in \partial f(x, x^*) \Leftrightarrow f(x, x^*) + f^*(x^*, x) = \langle (x^*, x), (x, x^*)\rangle \Leftrightarrow f(x, x^*) = \langle x^*, x\rangle$. The Fenchel-Young inequality implies $f(x, x^*) + f^*(x^*, x) \geq 2\langle x^*, x\rangle$; hence, using again that $f^* = f^\intercal$, we see that $f(x, x^*) = f^*(x^*, x) \geq \langle x^*, x\rangle$. Now apply [16, Proposition 2.3], [22, Theorem 1.4], or [11, Proposition 2.2.1]. ■

Using Theorem 5.1 and Corollary 5.2, we now define $R_A \in \Gamma(X \times X^*)$ by

$$\begin{aligned} R_A(x, x^*) &:= P\big(\tfrac{1}{2}, F_A, \tfrac{1}{2}, F_A^{*\intercal}, g_p\big)(x, x^*) && (52)\\ &= \min_{(x,x^*)=\frac{1}{2}(x_1,x_1^*)+\frac{1}{2}(x_2,x_2^*)} \tfrac{1}{2}F_A(x_1, x_1^*) + \tfrac{1}{2}F_A^*(x_2^*, x_2) + \tfrac{1}{4}g_p(x_1 - x_2, x_1^* - x_2^*). && (53)\end{aligned}$$

We shall also write $R_{A,p}$ for R_A if we want to emphasize the dependence on p.

Theorem 5.7 (extension and representation) *The function R_A is autoconjugate,*

$$F_A \leq R_A \leq F_A^{*\intercal}, \tag{54}$$

and $G(R_A)$ is a maximal monotone extension of A. Moreover, this construction is primal-dual symmetric *in the sense that*

$$R_{A^{-1},q} = R_{A,p}^* = R_{A,p}^\intercal \tag{55}$$

and

$$(G(R_{A,p}))^{-1} = G(R_{A,p}^*) = G(R_{A,p}^\intercal) = G(R_{A^{-1},q}). \tag{56}$$

Consequently, if A is maximal monotone, then $A = G(R_A)$ and $A^{-1} = G(R_A^\intercal) = G(R_A^*)$.

Proof. Lemma 5.5 implies that R_A is autoconjugate. Proposition 2.5 and Fact 5.4(i) imply that $R_A \leq \max\{F_A, F_A^{*\intercal}\} = F_A^{*\intercal}$. Taking the Fenchel conjugate now yields $R_A^\intercal = R_A^* \geq F_A^{*\intercal *} = F_A^\intercal$. Altogether, this implies (54). By Fact 5.6, the operator $G(R_A)$ is maximal monotone. Take $(a, a^*) \in \operatorname{gra} A$. By Fact 5.4(ii), $F_A(a, a^*) = F_A^{*\intercal}(a, a^*) = \langle a^*, a\rangle$. Thus, by (54), $R_A(a, a^*) = \langle a^*, a\rangle$. Now Fact 5.6 yields $a^* \in G(R_A)a$, i.e., $(a, a^*) \in \operatorname{gra} G(R_A)$. Hence $G(R_A)$ extends A. Since R_A is autoconjugate, we have $R_A^* = R_A^\intercal$ and hence

$$G(R_A^*) = G(R_A^\intercal). \tag{57}$$

Since $F_{A^{-1}} = F_A^\intercal$ (see (49)), it follows that $R_{A^{-1},q} = R_{A,p}^\intercal$. Thus (55) holds and we see that

$$G(R_{A^{-1},q}) = G(R_{A,p}^\intercal). \tag{58}$$

Let $(x, x^*) \in X \times X^*$. Using Fact 5.6 and once again that $R_A^* = R_A^\intercal$, we obtain the equivalences $x \in (G(R_A))^{-1}x^* \Leftrightarrow x^* \in G(R_A)x \Leftrightarrow R_A(x, x^*) = \langle x^*, x\rangle \Leftrightarrow R_A^*(x^*, x) = \langle x, x^*\rangle \Leftrightarrow x \in G(R_A^*)x^*$. Hence

$$(G(R_A))^{-1} = G(R_A^*). \tag{59}$$

Therefore, (56) follows by combining (57), (58), and (59). ■

Remark 5.8 A closer look at the proof of Theorem 5.7 reveals that its conclusion remains valid if R_A is replaced by $P(\frac{1}{2}, f, \frac{1}{2}, f^{*\intercal}, g_p)$ provided that $f \in \Gamma(X \times X^*)$ satisfies $F_A \leq f \leq F_A^{*\intercal}$.

Remark 5.9 Fitzpatrick's [10, Problem 5.5] asks to find an autoconjugate function f such that $G(f)$ extends A. Theorem 5.7 completely settles this question in the present reflexive Banach space setting: simply take $f = R_{A,p}$. In fact, when $p = 2$ the result is particularly appealing since $G(R_{A,2})$ is a maximal monotone extension of A with $(G(R_{A,2}))^{-1} = G(R_{A^{-1},2})$. A comparison to previous works is in order. The works by Penot, Svaiter, and Ghoussoub (see [14, 23, 15, 11]) are nonconstructive. Ghoussoub also requires that X be separable. Penot and Zălinescu [16] provide a constructive representation, but not for arbitrary monotone operators (A is assumed to satisfy a constraint qualification). To the best of our knowledge, the nonreflexive case remains open.

We conclude with a concrete example.

Example 5.10 Suppose that X is a Hilbert space and that $\operatorname{gra} A \subset \{(x, x) : \ x \in X\}$. Then $R_{A,2}\colon X \times X \to \mathbb{R}\colon (x, x^*) \mapsto \frac{1}{2}\|x\|^2 + \frac{1}{2}\|x^*\|^2$ and $G(R_{A,2}) = \mathrm{Id}$.

Proof. Since $A = A^{-1}$, we deduce from (55) that $R_{A,2} = R_{A^{-1},2} = R_{A,2}^*$. However, the only function equal to its conjugate on the Hilbert space $X \times X$ is $(x, x^*) \mapsto \frac{1}{2}\|x\|^2 + \frac{1}{2}\|x^*\|^2$. Hence $R_{A,2}$ is as claimed. Since $\nabla R_{A,2} = \mathrm{Id}$, it follows that $G(R_{A,2}) = \mathrm{Id}$. ■

Acknowledgment

Heinz Bauschke was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. Xianfu Wang was partially supported by the Natural Sciences and Engineering Research Council of Canada. We thank Dr. Abbas Moameni for sending us [11].

References

[1] 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.

[2] H. H. Bauschke, "Fenchel duality, Fitzpatrick functions and the extension of firmly nonexpansive mappings", *Proceedings of the American Mathematical Society*, vol. 135, pp. 135–139, 2007.

[3] H. H. Bauschke, J. M. Borwein, and X. Wang, "Fitzpatrick functions and continuous linear monotone operators", *SIAM Journal on Optimization*, to appear.

[4] H. H. Bauschke, R. Goebel, Y. Lucet, and X. Wang, "The proximal average: basic theory", submitted.

[5] H. H. Bauschke, Y. Lucet, and X. Wang, "Primal-dual symmetric antiderivatives for cyclically monotone operators", submitted.

[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, D. A. McLaren, and H. S. Sendov, "Fitzpatrick functions: inequalities, examples and remarks on a problem by S. Fitzpatrick", *Journal of Convex Analysis*, vol. 13, pp. 499–523, 2006.

[8] J. M. Borwein, "Maximal monotonicity via convex analysis", *Journal of Convex Analysis*, vol. 13, pp. 561–586, 2006.

[9] R. S. Burachik and S. Fitzpatrick, "On the Fitzpatrick family associated to some subdifferentials", *Journal of Nonlinear and Convex Analysis*, vol. 6, pp. 165–171, 2005.

[10] S. Fitzpatrick, Representing monotone operators by convex functions, *Workshop/Miniconference on Functional Analysis and Optimization (Canberra 1988)*, Proceedings of the Centre for Mathematical Analysis, Australian National University vol. 20, Canberra, Australia, pp. 59–65, 1988.

[11] N. Ghoussoub, "Maximal monotone operators are selfdual vector fields and vice-versa", *Proceedings of the American Mathematical Society*, to appear.

[12] J.-E. Martínez-Legaz and B. F. Svaiter, "Monotone operators representable by l.s.c. convex functions," *Set-Valued Analysis,* vol. 13, pp. 21–46, 2005.

[13] J.-E. Martínez-Legaz and M. Théra, "A convex representation of maximal monotone operators," *Journal of Nonlinear and Convex Analysis,* vol. 2, pp. 243–247, 2001.

[14] J.-P. Penot, "Autoconjugate functions and representations of monotone operators", *Bulletin of the Australian Mathematical Society*, vol. 67, pp. 277–284, 2003.

[15] J.-P. Penot, "The relevance of convex analysis for the study of monotonicity", *Nonlinear Analysis*, vol. 58, pp. 855–871, 2004.

[16] J.-P. Penot and C. Zălinescu, "Some problems about the representation of monotone operators by convex functions", *The Australian & New Zealand Industrial and Applied Mathematics Journal*, vol. 47, pp. 1–20, 2005.

[17] S. Reich and S. Simons, "Fenchel duality, Fitzpatrick functions and the Kirszbraun-Valentine extension theorem", *Proceedings of the American Mathematical Society* vol. 133, pp. 2657–2660, 2005.

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

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

[20] S. Simons, *Minimax and Monotonicity*, Springer-Verlag, 1998.

[21] S. Simons and C. Zălinescu, "A new proof for Rockafellar's characterization of maximal monotone operators", *Proceedings of the American Mathematical Society*, vol. 132, pp. 2969–2972, 2004.

[22] S. Simons and C. Zălinescu, "Fenchel duality, Fitzpatrick functions and maximal monotonicity", *Journal of Nonlinear and Convex Analysis*, vol. 6, pp. 1–22, 2005.

[23] B. F. Svaiter, "Fixed points in the family of convex representations of a maximal monotone operator", *Proceedings of the American Mathematical Society*, vol. 131, pp. 3851–3859, 2003.

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