THE KERNEL AVERAGE FOR TWO CONVEX FUNCTIONS
AND ITS APPLICATION TO THE EXTENSION AND
REPRESENTATION OF MONOTONE OPERATORS

HEINZ H. BAUSCHKE AND XIANFU WANG

ABSTRACT. We provide and analyze an 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.

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 $h: X \to [-\infty, +\infty]$, the (effective) domain is $\text{dom } h := \{ x \in X : h(x) < \infty \}$, and we use $\text{ri } \text{dom } h$ 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 $x \in \text{dom } h$, the set of subgradients of $h$ at $x$ is $\partial h(x) := \{ x^* \in X^* : h(y) - h(x) \geq \langle x^*, y-x \rangle \text{ for all } y \in X \}$, whereas $\partial h(x) := \emptyset$ if $x \notin \text{dom } h$. By $\text{cl } h$ we mean the lower semi-continuous hull of $h$. If $S \subset X$, then $\iota_S$ 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 $P(\lambda_1, f_1, \lambda_2, f_2, g): X \to [-\infty, +\infty]$ at $x \in X$ by

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

\[
= \inf_{x = z_1 + z_2} \left\{ \lambda_1 f_1 \left( \frac{z_1}{\lambda_1} \right) + \lambda_2 f_2 \left( \frac{z_2}{\lambda_2} \right) + \lambda_1 \lambda_2 g \left( \frac{z_1}{\lambda_1} - \frac{z_2}{\lambda_2} \right) \right\}.
\]
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

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

is the arithmetic average.

In the following, \( \lambda \ast f \) 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

\[
h(x) = \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} \{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2)\} = \inf_{x = z_1 + z_2} \left\{\lambda_1 f_1\left(\frac{z_1}{\lambda_1}\right) + \lambda_2 f_2\left(\frac{z_2}{\lambda_2}\right)\right\}
\]

\[
= (\lambda_1 \ast f_1) \Box (\lambda_2 \ast f_2)(x)
\]

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

\[
P\left(\lambda_1, \frac{1}{\lambda_1} \ast f_1, \lambda_2, \frac{1}{\lambda_2} \ast f_2, \iota_X\right) = f_1 \Box f_2,
\]

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

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

\[
= \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} \left\{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \frac{1}{2} \lambda_1 \|y_1\|^2 + \frac{1}{2} \lambda_2 \|y_2\|^2 - \frac{1}{2} \|x\|^2\right\}
\]

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 \( g = \lambda_2^{p-1} \frac{1}{\mu} \| \cdot \|^p \) with \( \mu > 0 \) and \( p \geq 1 \). We have

\[
f_1 \Box \frac{1}{\mu} \| \cdot \|^p = \frac{1}{\lambda_1}P\left(\lambda_1, f_1, \lambda_2, 0, \lambda_2^{p-1} \frac{1}{\mu} \| \cdot \|^p\right).
\]
Indeed, if \( x = \lambda_1 y_1 + \lambda_2 y_2 \), then \( y_1 - y_2 = (y_1 - x)/\lambda_2 \), and hence
\[
P\left(\lambda_1, f_1, \lambda_2, 0, \lambda_2^{-1} \frac{1}{\mu} \parallel \cdot \parallel^p \right)(x)
\]
\[
= \inf_{x = \lambda_1 y_1 + \lambda_2 y_2} \left( \lambda_1 f_1(y_1) + \lambda_2 0 + \lambda_1 \lambda_2 \lambda_2^{-1} \frac{1}{\mu} \parallel y_1 - y_2 \parallel^p \right)
\]
\[
= \inf_{y_1} \left( \lambda_1 f_1(y_1) + \lambda_1 \lambda_2 \frac{1}{\mu} \parallel y_1 - x \parallel^p \right)
\]
\[
= \inf_{y_1} \left( \lambda_1 f_1(y_1) + \frac{1}{\mu} \parallel y_1 - x \parallel^p \right).
\]

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 the kernel average given by (1) has. 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 \( \text{dom} f_1 \) and \( \text{dom} f_2 \). Moreover, if \( g \) is differentiable, the \( g \)-average is of Legendre type if \( f_1 \) or \( f_2 \) is as well. 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
\[
D(x, y) := x - y,
\]
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

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

Consequently, if $\lambda > 0$, then

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

Proof. Since $f = g \circ D = \iota_{X \times X} + g \circ D$, [24, Theorem 2.8.3(vii)] implies that for each $(x^*, y^*)$, we have

\[ f^*(x^*, y^*) = (g \circ D)^*(x^*, y^*) = \min \left\{ \iota_{(0,0)}(x^* - u^*, y^* + u^*) + g^*(u^*): u^* \in X^* \right\} \]

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

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

\[ h(x) = \inf \left\{ \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1 \lambda_2 g(y_1 - y_2): \lambda_1 y_1 + \lambda_2 y_2 = x \right\}. \]

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

\[ A(y_1, y_2) = \lambda_1 y_1 + \lambda_2 y_2. \]

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

\[ A^* x^* = (\lambda_1 x_1^*, \lambda_2 x_2^*). \]

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

\[ F(y_1, y_2) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1 \lambda_2 g(y_1 - y_2). \]

Then (3) becomes

\[ h = AF. \]

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 $x^* \in X$,

\[ (P(\lambda_1, f_1, \lambda_2, f_2, g))^*(x^*) = (c_1 \varphi)(\lambda_1 x_1^*, \lambda_2 x_2^*), \]

where

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

If

\[ (\text{ri dom } f_1 - \text{ri dom } f_2) \cap \text{ri dom } g \neq \emptyset, \]
then the closure operation in (10) can be omitted so that not only
\[(12) \quad (P(\lambda_1, f_1, \lambda_2, f_2, g))^* = P(\lambda_1, f_1', \lambda_2, f_2', g^\vee'), \]
but also the infimum in the definition of the kernel average is attained, i.e.,
\[(13) \quad P(\lambda_1, f_1', \lambda_2, f_2', g^\vee')(x^*) = \min_{x^* = \lambda_1 z_1 + \lambda_2 z_2} \{\lambda_1 f_1'^*(z_1) + \lambda_2 f_2'^*(z_2) + \lambda_1 \lambda_2 g^*(z_2 - z_1)\}.
\]

Proof. Using (9) and [18, Theorem 16.3], we have
\[(14) \quad h^* = (AF)^* = F^* \circ A^* \quad \text{on } X^*.
\]
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
\[(15) \quad F^* = (g_1 + g_2)^* = \text{cl}(g_1^\sqcap g_2^*).
\]
Using Lemma [21] we see that for every \((u, v) \in X^* \times X^*\),
\[
\begin{align*}
(g_1^\sqcap g_2^*)(u, v) &= \inf_{y_1, y_2} \left\{ \lambda_1 f_1^*(y_1) + \lambda_2 f_2^*(y_2) + g_2^*(u - y_1, v - y_2) \right\} \\
&= \inf_{u - y_1 + v - y_2 = 0} \left\{ \lambda_1 f_1^* \left( \frac{y_1}{\lambda_1} \right) + \lambda_2 f_2^* \left( \frac{y_2}{\lambda_2} \right) + \lambda_1 \lambda_2 g^* \left( \frac{u - y_1}{\lambda_1} - \frac{y_1}{\lambda_1} \right) \right\} \\
&= \inf_{u + v = y_1 + y_2} \left\{ \lambda_1 f_1^* \left( \frac{y_1}{\lambda_1} \right) + \lambda_2 f_2^* \left( \frac{y_2}{\lambda_2} \right) + \lambda_1 \lambda_2 g^* \left( \frac{u}{\lambda_1 \lambda_2} - \frac{y_1}{\lambda_1 \lambda_2} \right) \right\} \\
&= \inf_{u + v = \lambda_1 z_1 + \lambda_2 z_2} \left\{ \lambda_1 f_1'^*(z_1) + \lambda_2 f_2'^*(z_2) + \frac{\lambda_1 \lambda_2}{2} \left[ g^* \left( \frac{u}{\lambda_1 \lambda_2} - \frac{z_1}{\lambda_2} \right) \right. \right. \\
&\quad \left. \left. + g^* \left( \frac{-v}{\lambda_1 \lambda_2} - \frac{z_2}{\lambda_2} \right) \right] \right\}.
\end{align*}
\]
Hence (10) holds.

Now assume that
\[(17) \quad \text{ri dom } g_1 \cap \text{ri dom } g_2 \neq \emptyset.
\]
Then, by [18, Theorem 16.4], the closure in (15) is superfluous so that
\[(18) \quad F^* = g_1^\sqcap g_2^*.
\]
and the infimum in the definition of the last infimal convolution is attained. Since
\[g_2 = \lambda_1 \lambda_2 (g \circ D) \quad \text{and} \quad \text{ri dom } g \neq \emptyset, \quad [18, \text{Theorem 6.7}] \quad \text{implies that}
\]
\[\text{ri dom}(g \circ D) = \text{ri}(D^{-1} \text{ dom } g) = D^{-1} \text{ ri dom } g.
\]
Hence
\[\text{ri dom } g_1 = \text{ri}(\text{dom } f_1 \times \text{ dom } f_2) = \text{ri dom } f_1 \times \text{ri dom } f_2
\]
and
\[\text{ri dom } g_2 = D^{-1} \text{ ri dom } g = \{(y_1, y_2) : y_1 - y_2 \in \text{ ri dom } g\}.
\]
Therefore, (17) is equivalent to
\[(\text{ri dom } f_1 - \text{ri dom } f_2) \cap \text{ri dom } g \neq \emptyset,\]
i.e., to (11). Using (14), (18), (16), and the equivalence
\[
x^*_{\lambda_2} = \frac{\lambda_1 z_1}{\lambda_2} + z_2 \iff \frac{x^*}{\lambda} - \frac{z_1}{\lambda_2} = z_2 - z_1,
\]
we obtain
\[
h^*(x^*) = (g_1^*, g_2^*)(\lambda_1 x^*, \lambda_2 x^*)
\]
\[
= \inf_{x^* = y_1 + y_2} \{ \lambda_1 f_1^*(\frac{y_1}{\lambda_1}) + \lambda_2 f_2^*(\frac{y_2}{\lambda_2}) + \lambda_1 \lambda_2 g(\frac{x^*}{\lambda_2} - \frac{y_1}{\lambda_1 \lambda_2}) \}
\]
\[
= \inf_{x^* = x_1 + x_2 = z_1 + z_2} \{ \lambda_1 f_1^*(z_1) + \lambda_2 f_2^*(z_2) + \lambda_1 \lambda_2 g(\frac{x^*}{\lambda_2} - \frac{z_1}{\lambda_2}) \}
\]
\[
= \inf_{x^* = x_1 + x_2 = z_1 + z_2} \{ \lambda_1 f_1^*(z_1) + \lambda_2 f_2^*(z_2) + \lambda_1 \lambda_2 g(z_2 - z_1) \}.
\]
Therefore,
\[
(P(\lambda_1, f_1, \lambda_2, f_2, g))^* = P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^*),
\]
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 \(P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^*)\) are convex, lower semi-continuous, proper, and
\[
(P(\lambda_1, f_1, \lambda_2, f_2, g))^* = P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^*).
\]
In particular, for \(g = \frac{1}{p} \| \cdot \|_p\) with \(p > 1\), we have
\[
(P(\lambda_1, f_1, \lambda_2, f_2, \frac{1}{p} \| \cdot \|_p))^* = P(\lambda_1, f_1^*, \lambda_2, f_2^*, \frac{1}{q} \| \cdot \|_q),
\]
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
\[
(19) \quad h(x) := \inf \left\{ \lambda_1 f_1 \left( \frac{x_1}{\lambda_1} \right) + \lambda_2 f_2 \left( \frac{x_2}{\lambda_2} \right) : x_1 + x_2 = x \right\},
\]
has the conjugate
\[
(20) \quad h^* = \lambda_1 f_1^* + \lambda_2 f_2^*.
\]
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
\[
h \leq \frac{1}{2} f_1 + \frac{1}{2} f_1^* = h^*.
\]
**Proof.** The conjugation formula (20) follows from Theorem 2.2 with \(g = 0\). Putting
\[
x_i = \lambda_i x \text{ for } i \in \{1, 2\} \text{ in (19), we see that }
\]
\[
h \leq \lambda_1 f_1 + \lambda_2 f_2.
\]
In turn, if \(f_2 = f_1^*\) and \(\lambda_1 = \lambda_2 = \frac{1}{2}\), the result follows. □
The following result, which is related to the observation that the kernel average is increasing with respect to its arguments $f_1$, $f_2$, and $g$, places the kernel average between the convex hull of $f_1$ and $f_2$ (i.e., the largest convex function below $f_1$ and $f_2$) and the arithmetic average of $f_1$ and $f_2$.

**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 $g \geq 0$ and $g(0) = 0$, and set

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

Then

$$\text{conv}\{f_1, f_2\} \leq (\lambda_1 \ast f_1) \square (\lambda_2 \ast f_2) \leq h \leq \lambda_1 f_1 + \lambda_2 f_2 \leq \max\{f_1, f_2\}.$$

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

$$h(x) = \inf_{x = \lambda_1 y_1 + \lambda_2 y_2} \{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + g(y_1 - y_2)\} \geq \inf_{x = \lambda_1 y_1 + \lambda_2 y_2} \{\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2)\} = \inf_{x = \lambda_1 x_1 + \lambda_2 x_2} \{\lambda_1 f_1(x_1/\lambda_1) + \lambda_2 f_2(x_2/\lambda_2)\} = \left((\lambda_1 \ast f_1) \square (\lambda_2 \ast f_2)\right)(x) \geq \inf_{x = \mu_1 y_1 + \mu_2 y_2} \{\mu_1 f_1(y_1) + \mu_2 f_2(y_2)\} = \left(\text{conv}\{f_1, f_2\}\right)(x).$$

The result follows. (Alternatively, we may observe that $\iota(x_1) \leq g \leq \iota(X)$ and use the observation that the kernel average is increasing with respect to the kernel to prove this result.)

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

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

where $A$ and $F$ are as in (8) and (9), respectively. If $\text{dom } g = X$, then

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

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

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

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

\[\square\]
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
\[
(x^*, y^*) \in \partial (g \circ D)(x, y) \quad \iff \quad y^* = -x^*, \; x^* \in \partial g(x - y).
\]

**Proof.** As in the proof of Lemma 2.1, we use [24, Theorem 2.8.3(vii)] to deduce that \( \partial (g \circ D) = D^* \circ (\partial g) \circ D \), from which the result follows. \( \square \)

**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
\[
f := P(\lambda_1, f_1, \lambda_2, f_2, g),
\]
and assume that
\[
\text{ri dom } f_1 - \text{ri dom } f_2 \cap \text{ri dom } g \neq \emptyset
\]
and that \( x, y_1, y_2 \in X \) satisfy \( x = \lambda_1 y_1 + \lambda_2 y_2 \).

(i) If
\[
f(x) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1 \lambda_2 g(y_1 - y_2),
\]
then
\[
\partial f(x) = \bigcup_{y^* \in \partial g(y_1 - y_2)} (\partial f_1(y_1) + \lambda_2 y^*) \cap (\partial f_2(y_2) - \lambda_1 y^*).
\]

(ii) Conversely, if there exists \( y^* \in \partial g(y_1 - y_2) \) such that
\[
(\partial f_1(y_1) + \lambda_2 y^*) \cap (\partial f_2(y_2) - \lambda_1 y^*) \neq \emptyset,
\]
then
\[
f(x) = \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1 \lambda_2 g(y_1 - y_2),
\]
and hence the infimal convolution implicit in the definition of \( f \) is exact at \( x \).

**Proof.** Using [4], we recall that \( f = AF \), where \( A \) and \( F \) are given by [6] and [8], respectively. Since [22] holds, the sum rule [18 Theorem 23.8] and Lemma 2.7 imply
\[
\partial F(y_1, y_2) = \left( \lambda_1 \partial f_1(y_1) \times \lambda_2 \partial f_2(y_2) \right) + \lambda_1 \lambda_2 \partial (g \circ D)(y_1, y_2)
\]
\[
= \left( \lambda_1 \partial f_1(y_1) \times \lambda_2 \partial f_2(y_2) \right) + \lambda_1 \lambda_2 \left\{ (y^*, -y^*) : y^* \in \partial g(y_1 - y_2) \right\}.
\]

Consequently,
\[
A^* x^* = (\lambda_1 x^*, \lambda_2 x^*) \in \partial F(y_1, y_2)
\]
is equivalent to the existence of \( y^* \in \partial g(y_1 - y_2) \) such that
\[
(\lambda_1 x^*, \lambda_2 x^*) \in (\lambda_1 \partial f_1(y_1) + \lambda_1 \lambda_2 y^*) \times (\lambda_2 \partial f_2(y_2) - \lambda_1 \lambda_2 y^*),
\]
which can be expressed as \( x^* \in \partial f_1(y_1) + \lambda_2 y^*, x^* \in \partial f_2(y_2) - \lambda_1 y^* \), and as
\[
x^* \in (\partial f_1(y_1) + \lambda_2 y^*) \cap (\partial f_2(y_2) - \lambda_1 y^*).
\]

Now assume \( \partial f(x) = \partial (AF)(x) = (A^*)^{-1} \partial F(y) \).
where \( Ay = x \) and \( y = (y_1, y_2) \in X \times X \). Using (7) and the equivalence of (27) and (28), we thus obtain
\[
x^\ast \in \partial f(x) \iff (\lambda_1 x^\ast, \lambda_2 x^\ast) \in \partial F(y_1, y_2)
\]
\[
\iff x^\ast \in \bigcup_{y^\ast \in \partial g(y_1 - y_2)} (\partial f_1(y_1) + \lambda_2 y^\ast) \cap (\partial f_2(y_2) - \lambda_1 y^\ast).
\]

(ii) Now assume that there exists \( y^\ast \in \partial g(y_1 - y_2) \) such that (25) holds. Next, take \( x^\ast \in (\partial f_1(y_1) + \lambda_2 y^\ast) \cap (\partial f_2(y_2) - \lambda_1 y^\ast) \). The equivalence of (24) and (28) shows that \( A^\ast x^\ast \in \partial F(y_1, y_2) \). Let \( (z_1, z_2) \in A^{-1}x \), i.e., \( A(z_1, z_2) = x = A(y_1, y_2) \). Then \( 0 = (x - x, x^\ast) = (A(z_1, z_2) - A(y_1, y_2), x^\ast) = (z_1, z_2, y_1, y_2), A^\ast x^\ast \leq F(z_1, z_2) - F(y_1, y_2) \). It follows that
\[
f(x) = (AF)(x) = \inf_{A(z_1, z_2) = x} F(z_1, z_2) \geq F(y_1, y_2)
\]
\[
= \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1 \lambda_2 g(y_1 - y_2)
\]
\[
\geq f(x).
\]
This verifies (28) and completes the proof. \( \Box \)

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 \( g^\ast \) has full domain. Then for every \( x \in \text{dom } f \), there exist \( y_1, y_2 \in X \) such that
\begin{equation}
(29) \quad 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),
\end{equation}
and
\begin{equation}
(30) \quad \partial f(x) = (\partial f_1(y_1) + \lambda_2 \nabla g(y_1 - y_2)) \cap (\partial f_2(y_2) - \lambda_1 \nabla g(y_1 - y_2)).
\end{equation}
In particular, this result holds when \( g = \frac{1}{\| \| p \|} \| \cdot \|_p \) with \( p > 1 \).

Proof. Since \( \text{dom } g^\ast = X \) and \( g^{\ast \ast} = g \), Theorem 2.2 implies that
\[
(P(\lambda_1 f_1, \lambda_2 f_2, g^\ast))^\ast = P(\lambda_1 f_1, \lambda_2 f_2, g).
\]
and is exact. This gives (29). Now (30) follows from (29) and Theorem 2.3(1) \( \Box \)

The notions of essentially smooth, essentially strictly convex, and Legendre type are carefully studied in [13; 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 \( g^\ast \) 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 [13; Theorem 26.1]. Assume that \( f_1 \) is essentially smooth, so that \( \partial f_1 \) is at most single-valued. By Corollary 2.4, \( \partial f \) 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 [13; Theorem 26.3]. Assume that \( f_1 \) is essentially strictly convex. Then \( f_1^\ast \) is essentially smooth and so is \( P(\lambda_1 f_1^\ast, \lambda_2 f_2^\ast, g^{\ast \ast}) \)
by (i). The conjugate of the last function is not only essentially strictly convex but also equal to $f$ by Theorem 2.2.

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

\[
\partial f_1(x) \cap \partial f_2(x) \subset \partial f(x).
\]

**Proof.** If $\partial f_1(x) \cap \partial f_2(x) = \emptyset$, then (31) clearly holds. Thus assume that $\partial f_1(x) \cap \partial f_2(x) \neq \emptyset$ and set $y_1 := y_2 := x$ and $y^* := 0 \in \partial g(0) = \partial g(y_1 - y_2)$. Then (29) is true and so Theorem 2.2(i) 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 2.2(i) implies that $\partial f(x) \supset (\partial f_1(y_1) + \lambda_2 y^*) \cap (\partial f_2(y_2) - \lambda_1 y^*) = \partial f_1(x) \cap \partial f_2(x)$. □

**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 $f^* = P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^*)$ is an antiderivative of $A^{-1}$. Primal-dual symmetric methods for generating antiderivatives were recently investigated in [13].

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 [13] 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

\[
f(x) = \lambda_1 \lambda_2 \inf_{y_1 \in C_1, y_2 \in C_2} \inf_{y_1 + \lambda_2 y_2} g(y_1 - y_2) = \lambda_1 \lambda_2 \inf_{y \in \lambda_1 (x - C_1) \cap \lambda_2 (C_2 - x)} g\left(- \frac{y}{\lambda_1 \lambda_2}\right).
\]

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

\[
f(x) = \frac{1}{\lambda_1 \lambda_2} \inf_{y \in \lambda_1 (x - C_1) \cap \lambda_2 (C_2 - x)} \frac{1}{p} \|0 - y\|^p = \frac{1}{\lambda_1 \lambda_2} \inf_{y \in \lambda_1 (x - C_1) \cap \lambda_2 (C_2 - x)} \frac{1}{p} d_{\lambda_1 (x - C_1) \cap \lambda_2 (C_2 - x)}(0).
\]

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

\[
f(x) = \inf_{y_1 + \lambda_2 y_2} \frac{1}{\lambda_1 \lambda_2} \lambda_1 \lambda_2 g(y_1 - y_2).
\]
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

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

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

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

i.e., (32). Now suppose that $g = \frac{1}{2} \| \cdot \|^p$, where $p > 1$. Then $\text{dom } g = X$ and Theorem 2.4 yields that $\text{dom } f = \lambda_1 \text{dom } \lambda C_1 + \lambda_2 \text{dom } \lambda C_2 = \lambda_1 C_1 + \lambda_2 C_2$. Finally, (33) is an immediate consequence of (32). \qed

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

**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) $\inf f \geq \lambda_1 \inf f_1 + \lambda_2 \inf f_2$.

(ii) If $g(0) = 0$ and $\text{argmin } f_1 \cap \text{argmin } f_2 \neq \emptyset$, then $\text{argmin } f_1 \cap \text{argmin } f_2 \subseteq \text{argmin } f$ 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 $\text{argmin } f \subseteq \text{argmin } f_1 \cap \text{argmin } f_2$.

Proof. \[1\] 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 $z \in X$.

Take $x \in \text{argmin } f_1 \cap \text{argmin } f_2$. 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, \[1\] implies that $f(x) \geq \inf f \leq \lambda_1 \min f_1 + \lambda_2 \min f_2$. Altogether, we obtain $f(x) = \min f = \lambda_1 \min f_1 + \lambda_2 \min f_2$.

Take $x \in \text{argmin } f$. On the one hand,

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

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

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

Combining \[34\] and \[35\], 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
\[
\min f = \lambda_1 \min f_1 + \lambda_2 \min f_2 \iff \ \text{argmin} f_1 \cap \text{argmin} f_2 \neq \emptyset,
\]
in which case \( \text{argmin} f = \text{argmin} f_1 \cap \text{argmin} f_2 \). Consequently, if \( f, f_1, f_2 \) all have minimizers but \( \text{argmin} f_1 \cap \text{argmin} f_2 = \emptyset \), 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:
\[
g \in \Gamma(X) \text{ is differentiable and strictly convex everywhere, and } g^* \text{ has full domain.}
\]

**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 \( \text{dom} P_g f = \text{ran}(\partial f + \nabla g) = X^* \), \( P_g f \) 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 \( \text{ran}(\partial f + \nabla g) = X^* \). It follows that \( \text{dom} P_g f = \text{ran}(\partial f + \nabla g) = \text{ran}(f + g) = X^* \). Since \( f + g \) is uniformly convex on bounded convex subsets of \( X \), it is also strictly convex. By [24, Theorem 2.4.4(ii)], \( \partial(f + g) = \partial f + \nabla g \) is strictly monotone; thus, \( P_g f \) 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 \( x \in \text{dom} \partial f \), there exist \( y_1 \in \text{dom} \partial f_1 \) and \( y_2 \in \text{dom} \partial f_2 \) such that for every \( x^* \in \partial f(x) \), we have
\[
x = P_g f(x^* + \nabla g(x)) \\
= \lambda_1 P_g f_1(x^* - \lambda_2 \nabla g(y_1 - y_2) + \nabla g(y_1)) \\
+ \lambda_2 P_g f_2(x^* + \lambda_1 \nabla g(y_1 - y_2) + \nabla g(y_2)).
\]

Moreover, for every \( z^* \in X^* \), there exist \( y_1 \in \text{dom} \partial f_1 \) and \( y_2 \in \text{dom} \partial f_2 \) such that
\[
P_g f(z^*) = \lambda_1 P_g f_1(z^* - \nabla g(x) + \nabla g(y_1)) \\
+ \lambda_2 P_g f_2(z^* - \nabla g(x) + \lambda_1 \nabla g(y_1) + \nabla g(y_2)),
\]
where \( x := \lambda_1 y_1 + \lambda_2 y_2 = P_g f(z^* ) \) and \( y := y_1 - y_2 \).

**Proof.** Take \( x \in \text{dom} \partial f \). By Corollary[24] there exist \( y_1 \in \text{dom} f_1 \) and \( y_2 \in \text{dom} f_2 \) such that
\[
x = \lambda_1 y_1 + \lambda_2 y_2,
\]
 Altogether, we obtain (36)–(37) by combining (41), (39), and (40). Now take 

\[ y := y_1 - y_2. \]

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 \( y_1 \in \text{dom} \partial f_1 \) and \( y_2 \in \text{dom} \partial f_2 \). By Lemma 4.2, \( y_1 = P_{g,f_1}(x^*-\lambda_2 \nabla g(y)+\nabla g(y_1)) \) and \( y_2 = P_{g,f_2}(x^*+\lambda_1 \nabla g(y)+\nabla g(y_2)) \).

On the other hand, Lemma 4.2 also implies that

\[ x = P_{g,f}(x^*+\nabla g(x)). \]

 Altogether, we obtain (36)–(37) by combining (41), (39), and (40). Now take \( z^* \in X^* \). Since \( \text{ran}(\partial f + \nabla g) = X^* \) 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, (38) follows from (36)–(37).

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

\[ P_{g,f} = \lambda_1 P_{g,f_1} + \lambda_2 P_{g,f_2}. \]

**Proof.** Take \( z^* \in X^* = X \) and let \( y_1, y_2, x, y \) be as in Theorem 4.3. Since \( \nabla g = \text{Id} \), 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 (38) 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 \( \text{cont} f \) the set of points at which \( f \) is finite and continuous.
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 $(\text{dom } f_1 - \text{dom } f_2) \cap \text{cont } g \neq \emptyset$, then
\begin{equation}
(P(\lambda_1, f_1, \lambda_2, f_2, g))^* = P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{\ast\ast})
\end{equation}
belongs to $\Gamma(X^*)$ and the infimum in the definition of $P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{\ast\ast})$ is attained.

(ii) If $(\text{dom } f_1^* - \text{dom } f_2^*) \cap \text{cont } g^{\ast\ast} \neq \emptyset$, then
\begin{equation}
(P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{\ast\ast}))^* = P(\lambda_1, f_1, \lambda_2, f_2, g)
\end{equation}
belongs to $\Gamma(X)$ and the infimum in the definition of $P(\lambda_1, f_1, \lambda_2, f_2, g)$ is attained.

**Proof.** [1] As in Section 2 we define $A: X \times X \rightarrow X$: $(x_1, x_2) \mapsto x_1 - x_2$, $F: X \times X \rightarrow [-\infty, +\infty]: (x_1, x_2) \mapsto \lambda_1 f_1(x_1) + \lambda_2 f_2(x_2) + \lambda_1 \lambda_2 g(x_1 - x_2)$ so that $A^*: X^* \rightarrow X^* \times X^*: x^* \mapsto (\lambda_1 x^*, \lambda_2 x^*)$ and
\begin{equation}
P(\lambda_1, f_1, \lambda_2, f_2, g) = AF.
\end{equation}
Furthermore,
\begin{equation}
F = g_1 + g_2,
\end{equation}
where we define $g_1$ and $g_2$ on $X \times X$ by
\begin{equation}
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).
\end{equation}
Assume that $(\text{dom } f_1 - \text{dom } f_2) \cap \text{cont } g \neq \emptyset$, which implies that $\text{dom } g_1 \cap \text{cont } g_2 \neq \emptyset$. Using [24] Theorem 2.3.1(iii), we see that for every $(x_1^*, x_2^*) \in X^* \times X^*$,
\begin{equation}
F^*(x_1^*, x_2^*) = \min_{(y_1^*, y_2^*) \in X^* \times X^*} g_1^*(y_1^* - y_1^*) + g_2^*(y_2^* - y_2^*).
\end{equation}
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 $g_2^*(x_1^* - x_2^*) = \lambda_1 \lambda_2 g^{\ast\ast}(x_1^*/\lambda_2) + \lambda_1 \lambda_2 g^{\ast\ast}(x_2^*/\lambda_1)$. By Lemma 2.4 Using [15], [24] Theorem 2.3.1(i), and [40], we thus obtain for every $x^* \in X^*$,
\begin{align*}
&P(\lambda_1, f_1, \lambda_2, f_2, g)^*(x^*) = (AF)^*(x^*) = F^*(A^* x^*) \\
&= \min_{y_1^*, y_2^*} g_1^*(y_1^* - y_1^*) + g_2^*(y_1^* - y_1^* - y_2^* - 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^{\ast\ast}(\frac{\lambda_1 x^* - y_1^*}{\lambda_1 \lambda_2}) \\
&= \lambda_1 \lambda_2 g^{\ast\ast}(z^*_1 - z^*_2) + \lambda_1 \lambda_2 g^{\ast\ast}(\frac{\lambda_1 x^* - y_1^*}{\lambda_1 \lambda_2}) \\
&= P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{\ast\ast})(x^*).
\end{align*}

[1] Apply [1] to $f_1^*, f_2^*, g^{\ast\ast}$.

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 $g^{\ast\ast}$ have full domain. Then $P(\lambda_1, f_1, \lambda_2, f_2, g) \in \Gamma(X \times X^*)$, $P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{\ast\ast}) \in \Gamma(X^* \times X)$, and
\begin{equation}
(P(\lambda_1, f_1, \lambda_2, f_2, g))^* = P(\lambda_1, f_1^*, \lambda_2, f_2^*, g^{\ast\ast}).
\end{equation}
In particular, if $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, then
\[
(P(\lambda_1, f_1, \lambda_2, f_2, \frac{1}{p} \parallel \cdot \parallel^p + \frac{1}{q} \parallel \cdot \parallel^q))^* = P(\lambda_1^*, f_1^*, \lambda_2^*, f_2^*, \frac{1}{q} \parallel \cdot \parallel^q + \frac{1}{p} \parallel \cdot \parallel^p).
\]

From now on, we assume that
\[A: X \rightrightarrows X^* \]
is a monotone operator,
i.e., for all $(x, x^*), (y, y^*) \in \text{gra} A := \{ (z, z^*) \in X \times X^*: z^* \in A z \}$, we have $\langle x^* - y^*, x - y \rangle \geq 0$, and that gra $A \neq \emptyset$. 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., [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
\[
(47) \quad F_A(x, x^*) := \sup_{(a, a^*) \in \text{gra} A} \left\{ \langle a^*, x \rangle + \langle x^*, a \rangle - \langle a^*, a \rangle \right\}.
\]

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

\[
f^\top(x, x^*) = f(x^*, x),
\]
and similarly for $f \in \Gamma(X \times X^*)$. We can now simply express the Fitzpatrick function of the inverse of $A$ by
\[
F_A^{-1} = F_A^\top.
\]

As in [11, 15, 16], we say that
\[
f \text{ is autoconjugate } \iff f^\top = f^*;
\]
in [11], $f$ is then called a “selfdual Lagrangian”.

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

(i) $F_A \leq F_A^\top$.

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

(iii) If $A$ is maximal monotone and $(x, x^*) \notin \text{gra} A$, 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)), it is possible to represent $A$ by $F_A$ in the sense that $\text{gra} A = \{ (x, x^*) \in X \times X^*: F_A(x, x^*) = \langle x^*, x \rangle \}$. 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
\[
g_p(x, x^*) := \frac{1}{p} \parallel x \parallel^p + \frac{1}{q} \parallel x^* \parallel^q,
\]
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
\[
g_p^\top = g_p^* = g_p^\top,
\]
which itself is autoconjugate and which yields the following recipe for constructing autoconjugates.

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

$$
\left( P\left( \frac{1}{2}, f, \frac{1}{2}, f^{**}, g_p \right) \right)^{\ast} = P\left( \frac{1}{2}, f^*, \frac{1}{2}, f^T, g_p^T \right)
$$

and $P\left( \frac{1}{2}, f, \frac{1}{2}, f^{**}, g_p \right)$ is autoconjugate.

**Proof.** Using Corollary 5.2 and (49), we have

$$
\left( P\left( \frac{1}{2}, f, \frac{1}{2}, f^{**}, g_p \right) \right)^{\ast} = P\left( \frac{1}{2}, f^*, \frac{1}{2}, f^{**T}, g_p^* \right) = P\left( \frac{1}{2}, f^*, \frac{1}{2}, f^T, g_p^T \right)
$$

$$
= P\left( \frac{1}{2}, f^T, \frac{1}{2}, f^*, g_p^T \right) = P\left( \frac{1}{2}, f^T, \frac{1}{2}, f^{**T}, g_p^* \right)
$$

$$
= \left( P\left( \frac{1}{2}, f, \frac{1}{2}, f^{**}, g_p \right) \right)^{\ast}.
$$

The proof is complete. \qed

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

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

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 $f(x, x^*) = (x^*, x)$.

**Proof.** (See [11, 16].) Let $(x, x^*) \in X \times X^*$. Since $f$ is autoconjugate, we have the equivalences $x^* \in G(f)x \Leftrightarrow (x^*, x) \in \partial f(x, x^*) \Leftrightarrow f(x, x^*) + f^*(x^*, x) = \langle (x^*, x), (x^*, x^*) \rangle \Leftrightarrow f(x, x^*) = (x^*, x)$. The Fenchel-Young inequality implies $f(x, x^*) + f^*(x^*, x) \geq 2\langle x^*, x \rangle$; hence, using again that $f^* = f^T$, we see that $f(x, x^*) = f^*(x^*, x) \geq (x^*, x)$. Now apply [10] Proposition 2.3, [22] Theorem 1.4, or [11] Proposition 2.2.1. \qed

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

$$
R_A(x, x^*) := P\left( \frac{1}{2}, F_A, \frac{1}{2}, F_A^T, g_p \right)(x, x^*)
$$

$$
= \min_{(x, x^*) = \frac{1}{2}(x_1, x_1^*) + \frac{1}{2}(x_2, x_2^*)} \frac{1}{4} F_A(x_1, x_1^*) + \frac{1}{2} F_A^T(x_2, x_2)
$$

$$
+ \frac{1}{4} g_p(x_1 - x_2, x_1^* - x_2^*).
$$

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,

$$
F_A \leq R_A \leq F_A^T,
$$

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

$$
R_{A^{-1}, p} = R^*_{A, p} = R^T_{A, p}
$$

and

$$
(G(R_{A, p}))^{-1} = G(R^*_{A, p}) = G(R^T_{A, p}) = G(R_{A^{-1}, p}).
$$

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

Proof. Lemma 5.5 implies that \( R_A \) is autoconjugate. Proposition 2.5 and Fact 5.4(1) imply that \( R_A \leq \max\{F_A, F_A^\top\} = F_A^\top. \) Taking the Fenchel conjugate now yields \( R_A^\top = R_A^\top \geq F_A^{\star \top} = F_A^\top. \) Altogether, this implies (55). By Fact 5.6, the operator \( G(R_A) \) is maximal monotone. Take \((a, a^\star) \in \text{gra } A. \) By Fact 5.4(ii), \( F_A(a, a^\star) = F_A^{\star \top}(a, a^\star) = (a^\star, a). \) Thus, by (55), \( R_A(a, a^\star) = (a^\star, a). \) Now Fact 5.6 yields \( a^\star \in G(R_A)a, \) i.e., \((a, a^\star) \in \text{gra } G(R_A). \) Hence \( G(R_A) \) extends \( A. \) Since \( R_A \) is autoconjugate, we have \( R_A^\star = R_A^\top \) and hence
\[
G(R_A^\star) = G(R_A^\top). \tag{56}
\]
Since \( F_{A^{-1}} = F_A^\top \) (see (48)), it follows that \( R_{A^{-1},q} = R_{A,p}^\top. \) Thus (54) holds and we see that
\[
G(R_{A^{-1},q}) = G(R_{A,p}^\top). \tag{57}
\]
Let \((x, x^\star) \in X \times X^\star. \) Using Fact 5.6 and once again that \( R_A^\star = R_A^\top, \) we obtain the equivalences \( x \in (G(R_A))^{-1} x^\star \iff x^\star \in G(R_A)x \iff R_A(x, x^\star) = (x^\star, x) \iff R_A^\star(x^\star, x) = (x, x^\star) \iff x \in G(R_A^\star)x^\star. \) Hence
\[
G(R_A)^{-1} = G(R_A^\star). \tag{58}
\]
Therefore, (55) follows by combining (55), (57), and (58). \qedsymbol

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 \( P(\frac{1}{2}, f, \frac{1}{2}, f^\top, g_p) \) provided that \( f \in \Gamma(X \times X^\star) \) satisfies \( F_A \leq f \leq F_A^\top. \)

Remark 5.9. Fitzpatrick’s [10, Problem 5.5] asks one 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 \( \text{gra } A \subset \{ (x, x) : x \in X \}. \) Then \( R_{A,2} : X \times X \to \mathbb{R} : (x, x^\star) \mapsto \frac{1}{2}\|x\|^2 + \frac{1}{2}\|x^\star\|^2 \) and \( G(R_{A,2}) = \text{Id.} \)

Proof. Since \( A = A^{-1}, \) we deduce from (54) that \( R_{A,2} = R_{A^{-1},2} = R_A^\star. \) However, the only function equal to its conjugate on the Hilbert space \( X \times X \) is \( (x, x^\star) \mapsto \frac{1}{2}\|x\|^2 + \frac{1}{2}\|x^\star\|^2. \) Hence \( R_{A,2} \) is as claimed. Since \( \nabla R_{A,2} = \text{Id.} \) it follows that \( G(R_{A,2}) = \text{Id.} \) \qedsymbol

Acknowledgment

The authors thank the referee for his very helpful comments which led to significantly shorter proofs of Lemma 2.1, Lemma 2.7, Theorem 2.8, and Lemma 4.2. The first author was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. The second author was partially supported by the Natural Sciences and Engineering Research Council of Canada. We thank Dr. Abbas Moameni for sending us [11].
References


Department of Mathematics, Irving K. Barber School, UBC Okanagan, Kelowna, British Columbia, Canada V1V 1V7

E-mail address: heinz.bauschke@ubc.ca

Department of Mathematics, Irving K. Barber School, UBC Okanagan, Kelowna, British Columbia, Canada V1V 1V7

E-mail address: shawn.wang@ubc.ca