# On the Kernel Average for $n$ Functions

by

Sarah Michelle Moffat

B.Sc., The University of British Columbia, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The College of Graduate Studies

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

December 2009

# Abstract

After an introduction to Hilbert spaces and convex analysis, the proximal average is studied and two smooth operators are provided. The first is a new version of an operator previously supplied by Goebel, while the second one is new and uses the proximal average of a function and a quadratic to find a smooth approximation of the function.

Then, the kernel average of two functions is studied and a reformulation of the proximal average is used to extend the definition of the kernel average to allow for any number of functions. The Fenchel conjugate of this new kernel average is then examined by calculating the conjugate for two specific kernel functions that represent two of the simplest cases that could be considered. A closed form solution was found for the conjugate of the first kernel function and it was rewritten in three equivalent forms. A solution was also found for the conjugate of the second kernel function, but the two solutions do not have the same form which suggests that a general solution for the conjugate of any kernel function will not be found.

# Table of Contents

# List of Figures

# Acknowledgements

First, I would like to thank my supervisors Heinz Bauschke and Shawn Wang for their excitement about mathematics and their flexibility when it came to fitting mathematics into my life. I'd also like to thank Liangjin Yao and Mason Macklem for all their help, and Richard Taylor who first inspired me to pursue graduate studies.

# Dedication

For my wonderful husband Jim and our daughter Zoe, who was born during the production of this thesis.

# Chapter 1

# Introduction

When averaging functions the most natural place to start is the arithmetic average, defined by

$$\mathcal{A} := \lambda f_1 + (1-\lambda) f_2, \tag{1.1}$$

where $0 \leq \lambda \leq 1$. This works well if both functions $f_1$ and $f_2$ are everywhere defined. But if the functions are not differentiable at some points or if their domains do not intersect, then the arithmetic average will not be differentiable or will be $+\infty$ everywhere. In [2] Bauschke, Lucet, and Trienis define a new average, the proximal average, and discuss the benefits of using the proximal average for these types of cases. In particular, the proximal average produces a continuous and differentiable function even if the original functions are non-smooth and their domains do not intersect, provided at least one of the functions is differentiable with a full domain.

From the definition of the proximal average, the more general kernel average [3] was defined for averaging two functions based on a kernel function. Both the arithmetic and proximal averages can be derived as special cases of the kernel average. This thesis extends the definition of the kernel average to an arbitrary number of functions and examines the convex conjugate of the kernel average for $n$ functions.

# Chapter 2

# Hilbert Spaces

In this chapter we give some background material on inner product spaces. The notion of vector spaces and their extensions are central to much of the following thesis, so a quick reminder of some concepts from linear algebra is also included to refresh the reader's memory. For more on vector spaces see [7, Chapters 1 and 7] or [11, Chapter 4].

## 2.1 General Vector Spaces

**Definition 2.1.1 (Vector Space)** *A vector space consists of a set $V$ with elements called vectors, along with two operations such that the following properties hold:*

*(1) Vector addition: Let $u, v \in V$, then there is a vector $u + v \in V$ and the following are satisfied.*

*(i) Commutativity: $u + v = v + u$, $\forall u, v \in V$.*

*(ii) Associativity: $u + (v + w) = (u + v) + w$, $\forall u, v, w \in V$.*

*(iii) Zero: there is a vector $\boldsymbol{0} \in V$ such that $\boldsymbol{0} + u = u = u + \boldsymbol{0}$, $\forall u \in V$.*

*(iv) Inverses: for each $u \in V$, there is a vector $-u$ such that $u + (-u) = \boldsymbol{0}$.*

*(2) Scalar multiplication: Let $u, v \in V$ and $r, s \in \mathbb{R}$, then the following are satisfied.*

*(i) Left distributivity: $(r + s)v = rv + sv$.*

*(ii) Associativity: $r(sv) = (rs)v$.*

*(iii) Right distributivity: $r(u + v) = ru + rv$.*

*(iv) Neutral element: $1v = v$.*

*(v) Absorbing element: $0v = \boldsymbol{0}$.*

*(vi) Inverse neutral element: $(-1)v = -v$.*

**Example 2.1.2** *The space $\mathbb{R}^n$ consists of vectors $v = (v_1, \cdots, v_n)$ with $v_i \in \mathbb{R}$ for $1 \leq i \leq n$ and operations defined by*

$$(u_1, \cdots, u_n) + (v_1, \cdots, v_n) := (u_1 + v_1, \cdots, u_n + v_n)$$
$$r(v_1, \cdots v_n) := (rv_1, \cdots, rv_n),$$

*where $r \in \mathbb{R}$.*

**Definition 2.1.3** *A* subspace *of a vector space $V$ is a subset $W$ of $V$ with $W \neq \varnothing$ and $W$ is a vector space using the operations of $V$.*

**Definition 2.1.4** *Let $S \subseteq V$, the* span *of $S$ is the smallest subspace containing $S$ and is denoted by $\operatorname{span} S$.*

**Fact 2.1.5** *The subspace spanned by a nonempty set $S \subseteq V$ consists of all linear combinations of the elements of $S$.*

**Definition 2.1.6** *A* linear transformation *$A$ from a vector space $V$ to a vector space $W$ is a function $A : V \to W$ satisfying*

$$A(r_1v_1 + r_2v_2) = r_1Av_1 + r_2Av_2,\ \forall v_1, v_2 \in V \text{ and } \forall r_1, r_2 \in \mathbb{R}.$$

## 2.2 Inner Product Spaces

We recall the definitions of a norm and an inner product.

**Definition 2.2.1** *A* norm *on a vector space $V$ is a function $\|\cdot\| : V \to [0, +\infty[$ with the following properties.*

*(i) Positive definite: $\|x\| = 0$ if, and only if, $x = 0$,*

*(ii) Homogeneous: $\|\alpha x\| = |\alpha|\|x\|$, $\forall x \in V$ and $\alpha \in \mathbb{R}$,*

*(iii) Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in V$.*

**Definition 2.2.2** *An* inner product *on a vector space $V$ is a function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ satisfying the following properties.*

*(i) Positive definite: $\langle x, x\rangle \geq 0$, $\forall x \in V$ and $\langle x, x\rangle = 0$ only if $x = 0$;*

*(ii) Symmetry: $\langle x, y\rangle = \langle y, x\rangle$, $\forall x, y \in V$;*

*(iii) Bilinearity $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$, $\forall x, y, z \in V$ and $\alpha, \beta \in \mathbb{R}$.*

We call a vector space paired with an inner product and norm induced by $\|x\| := \langle x, x\rangle^{1/2}$, an *inner product space.*

**Definition 2.2.3** *In a normed vector space $(V, \|\cdot\|)$, a sequence $(v_n)_{n=1}^{\infty}$* converges *to $v \in V$ if $\lim_{n\to\infty} \|v_n - v\| = 0$.*

**Definition 2.2.4** *A sequence* $(v_n)_{n=1}^{\infty}$ *is called a* Cauchy sequence *if for every* $\epsilon > 0$*, there is an integer* $N > 0$ *such that* $\|v_n - v_m\| < \epsilon$ *for all* $n, m \geq N$*.*

**Remark 2.2.5** *While every convergent sequence is a Cauchy sequence, the converse is not true. For example, consider the sequence*

$$1, \frac{14}{10}, \frac{141}{100}, \frac{1414}{1000}, \frac{14142}{10000}, \cdots$$

*in* $\mathbb{Q}$ *approaching* $\sqrt{2}$*. This sequence does not converge in* $\mathbb{Q}$*.*

**Definition 2.2.6** *An inner product space* $V$ *is* complete *if every Cauchy sequence in* $V$ *converges to some vector* $v \in V$*.*

**Definition 2.2.7** *A complete inner product space is called a* Hilbert space*.*

**Example 2.2.8** *The following inner product spaces are Hilbert spaces:*

*(i) [7, Theorem 4.2.5]* $\mathbb{R}^n$ *paired with the inner product* $\langle x, y\rangle = x_1y_1 + \cdots + x_ny_n$*.*

*(ii) [7, Theorem 7.5.8] The space* $\ell^2$*, consisting of all sequences* $x = (x_n)_{n=1}^{\infty}$ *such that* $\|x\|_2 := \left(\sum_{n=1}^{\infty} x_n^2\right)^{1/2}$ *is finite, with the inner product* $\langle x, y\rangle = \sum_{n=1}^{\infty} x_ny_n$*.*

*(iii) [10, Example 2.2-7] The space* $L^2[0,1]$*, consisting of all (equivalence classes of) functions* $f$ *on* $[0,1]$ *such that* $\|f\|_2 := (\int_0^1 |f(x)|^2dx)^{\frac{1}{2}} < +\infty$*, paired with the inner product* $\langle f, g\rangle = \int_0^1 f(x)g(x)dx$*.*

**Definition 2.2.9** *[4, Section 1.1] Let $X, V$ be vector spaces and let $A : X \to V$ be a linear operator. The corresponding* adjoint *linear transformation from $V$ to $X$ is the unique operator $A^*$ such that the following identity holds*

$$\langle Ax, y\rangle = \langle x, A^* y\rangle \tag{2.1}$$

*for all $x \in X$ and $y \in V$.*

**Example 2.2.10** *Let $X$ be a vector space and $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$. Let $A : X^n \to X$ be the linear operator defined by $x = (x_1, \cdots, x_n) \mapsto \lambda_1 x_1 + \cdots + \lambda_n x_n$. Then the adjoint of $A$ is the operator $A^* : X \to X^n$ such that $z \mapsto (\lambda_1 z, \cdots, \lambda_n z)$.*

*Proof.* Let $x \in X^n$ and $y \in X$, then

$$\langle Ax, y\rangle = \langle \lambda_1 x_1 + \cdots + \lambda_n x_n, y\rangle = \sum_{i=1}^{n} \lambda_i \langle x_i, y\rangle.$$

On the other hand,

$$\begin{aligned}\langle x, A^* y\rangle &= \langle (x_1, \cdots, x_n), (\lambda_1 y, \cdots, \lambda_n y)\rangle \\ &= \langle x_1, \lambda_1 y\rangle + \cdots + \langle x_n, \lambda_n y\rangle = \sum_{i=1}^{n} \lambda_i \langle x_i, y\rangle.\end{aligned}$$

Since $\langle Ax, y\rangle = \langle x, A^* y\rangle$ $\forall x \in X^n, y \in X$ and the adjoint is unique then $A^*$ is the adjoint of $A$. ■

## 2.3 Facts on Maximization and Minimization

In this section, we assume that $X$ is an inner product space.

**Fact 2.3.1** *Let $x \in X$ then*

$$\sup_{y \in X} \langle x, y \rangle = \begin{cases} 0, & \text{if } x = 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

**Fact 2.3.2** *Let $f$ and $g$ be functions from $X \to ]-\infty, +\infty]$, then*

*(i)* $\sup\limits_{x,y \in X} (f(x) + g(y)) = \sup\limits_{x \in X} (f(x)) + \sup\limits_{y \in X} (g(y))$

*(ii)* $\inf\limits_{x,y \in X} (f(x) + g(y)) = \inf\limits_{x \in X} (f(x)) + \inf\limits_{y \in X} (g(y))$.

**Fact 2.3.3** *Let $x, y \in X$ then*

*(i)* $\sup\limits_{x \in X} \sup\limits_{y \in X} f(x,y) = \sup\limits_{y \in X} \sup\limits_{x \in X} f(x,y)$

*(ii)* $\inf\limits_{x \in X} \inf\limits_{y \in X} f(x,y) = \inf\limits_{y \in X} \inf\limits_{x \in X} f(x,y)$

These will be used frequently later on.

# Chapter 3

# Convex Analysis

Let $\mathcal{H}$ denote a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$, norm $\| \cdot \|$, and identity mapping Id. We'll now introduce some necessary convex analysis, for a more in-depth look at convex analysis please see [12].

## 3.1 Convex Sets

**Definition 3.1.1** *A set $A \subseteq \mathcal{H}$ is affine if $x \in A$, $y \in A$, and $\theta \in \mathbb{R}$ imply that $\theta x + (1-\theta)y \in A$.*

**Definition 3.1.2** *A set $C \subseteq \mathcal{H}$ is convex if $x \in C$, $y \in C$, and $0 \leq \theta \leq 1$ imply that $\theta x + (1-\theta)y \in C$.*

This means that for any two points in a convex set $C$, the line segment joining the two points is also contained in $C$.

**Example 3.1.3** *The following sets are convex:*

*(i) Affine sets;*

*(ii) Halfspaces: A set $H$ is a halfspace if for some $b \in \mathcal{H}$ and $\beta \in \mathbb{R}$, $H = \{x \in \mathcal{H} : \langle x, b \rangle \leq \beta\}$;*

*(iii) Closed ball of radius* $r > 0$ *centered at a point* $x_c$*:* $B(x_c, r) := \{x \in \mathcal{H} : \|x_c - x\| \leq r\}$.

The following definitions describe some important properties of sets that are frequently used in convex analysis.

**Definition 3.1.4** *The* indicator function *of a set* $C \subseteq \mathcal{H}$ *is the function* $\iota_C : \mathcal{H} \to [0, +\infty]$ *defined by*

$$\iota_C(x) = \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise.} \end{cases}$$

**Definition 3.1.5** *The* support function *of a set* $C \subseteq \mathcal{H}$ *is the function* $\sigma_C : \mathcal{H} \to ]-\infty, +\infty]$ *defined by*

$$\sigma_C(x) = \sup_{u \in C} \langle x, u \rangle.$$

### 3.1.1 Cones

A set $C$ is called a *cone* if for every $x \in C$ and $\theta > 0$ we have $\theta x \in C$. A set $C$ is a *convex cone* if it is both convex and a cone. That is, for every $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$ then $\theta_1 x_1 + \theta_2 x_2 \in C$.

**Definition 3.1.6** *The* conical hull *of a set* $C \subseteq \mathcal{H}$ *is the set*

$$\operatorname{cone} C = \bigcup_{\lambda > 0} \{\lambda x : x \in C\}.$$

**Fact 3.1.7** *[5, Section 2.1.5] The conical hull of $C$ is the smallest convex cone that contains the set $C$.*

**Example 3.1.8** *Let $D = \{z = (x, y, 0) \in \mathbb{R}^3 : \|z\| \leq 1\}$ be a closed unit disc in $\mathbb{R}^3$, then* $\operatorname{cone} D = \mathbb{R}^2 \times \{0\}$.

**Definition 3.1.9** *Let $C$ be a nonempty convex set in $\mathcal{H}$. We say that $C$* recedes *in the direction of $y$, $y \neq 0$, if and only if $x + \lambda y \in C$ for every $\lambda \geq 0$ and $x \in C$. Directions in which $C$ recedes are also called the* directions of recession.

**Definition 3.1.10** *The* recession cone *of a set $C$ is the set of all vectors $y \in \mathcal{H}$ such that for each $x \in C$, $x + \lambda y \in C$ for all $\lambda \geq 0$. The recession cone of $C$ is denoted by $0^+C$.*

**Fact 3.1.11** *[12, Theorem 8.1] Let $C$ be a non-empty convex set. Then the recession cone, $0^+C$, is a convex cone containing the origin.*

### 3.1.2 Interiors of Sets

**Definition 3.1.12** *The* interior *of a set $C \subseteq \mathcal{H}$ is the set*

$$\operatorname{int} C = \{x \in \mathcal{H} : \exists \epsilon > 0 \text{ such that } B(x, \epsilon) \subseteq C\},$$

*where $B(x, \epsilon) = \{y : \|y - x\| < \epsilon\}$.*

**Definition 3.1.13** *The* relative interior *of a convex set $C \subseteq \mathcal{H}$ is the set*

$$\operatorname{ri} C = \{x \in \mathcal{H} : \operatorname{cone}(C - x) = \operatorname{span}(C - x)\}.$$

The following example illustrates the need to distinguish between the interior and the relative interior of a set.

**Example 3.1.14** *Consider again the closed disc* $D = \{z = (x, y, 0) \in \mathbb{R}^3 : \|z\| \leq 1\}$*. We get that* $\operatorname{int} D = \varnothing$ *since no ball in* $\mathbb{R}^3$ *can be contained in* $D$*, however* $\operatorname{ri} D = \{z = (x, y, 0) \in \mathbb{R}^3 : \|z\| < 1\}$*.*

**Definition 3.1.15** *A point* $x$ *is a* limit point *of a subset* $C \subseteq \mathcal{H}$ *if there is a sequence* $(x_n)_{n=1}^{\infty}$ *with* $x_n \in C$ *such that* $x = \lim_{n\to\infty} x_n$*. A set* $C \subseteq \mathcal{H}$ *is* closed *if it contains all of its limit points.*

**Definition 3.1.16** *The* closure *of a set* $C \subseteq \mathcal{H}$ *is the smallest closed set containing* $C$*, and is denoted by* $\operatorname{cl} C$*.*

## 3.2 Convex Functions

The effective domain of a function $f : \mathcal{H} \to ]-\infty, +\infty]$ is the set of points:

$$\operatorname{dom} f = \{x \in \mathcal{H} : f(x) < +\infty\}. \tag{3.1}$$

The set of global minimizers of $f$ is denoted by $\operatorname{argmin}_{x\in\mathcal{H}} f(x)$.

**Definition 3.2.1** *We call a function* $f$ proper *if it never takes on the value of* $-\infty$ *and is not identically equal to* $+\infty$*.*

**Definition 3.2.2** *A function* $f$ *is* lower semicontinuous *at a point* $x_0$ *if*

$$\liminf_{x\to x_0} f(x) \geq f(x_0),$$

*where* lim inf *is as defined in [13, Definition 1.5]. The function is said to be lower semicontinuous if it is lower semicontinuous at every point $x_0 \in \mathcal{H}$.*

**Definition 3.2.3** *A function $f$ is* coercive *if*

$$\lim_{\|x\| \to \infty} f(x) = \infty.$$

*A function is* supercoercive *if*

$$\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = \infty.$$

**Definition 3.2.4** *The epigraph of a function $f : \mathcal{H} \to ]-\infty, +\infty]$ is the set*

$$\operatorname{epi} f = \{(x, t) \in \mathcal{H} \times \mathbb{R} : f(x) \leq t\}.$$

*This is illustrated in Figure 3.1.*

**Definition 3.2.5** *A function $f : \mathcal{H} \to ]-\infty, +\infty]$ is convex if* $\operatorname{epi} f$ *is a convex set.*

We denote the set of proper, lower semicontinuous, convex functions in $\mathcal{H}$ by $\Gamma_0(\mathcal{H})$. While Definition 3.2.5 conveniently ties the notion of a convex function to that of a convex set, in practice a convex function is usually synonymous with the following result.

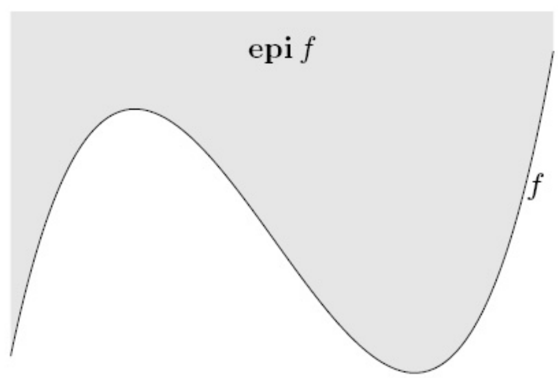



Figure 3.1: Epigraph of a function
[5, Figure 3.5]

**Theorem 3.2.6** *[12, Theorem 4.1] A function $f : \mathcal{H} \to ]\infty, +\infty]$ is convex if and only if for $x, y \in \mathcal{H}$ and $0 < \theta < 1$,*

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y).$$

A function is said to be *strictly convex* if the inequality in Theorem 3.2.6 is strict; that is, that $f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$ provided that $x \neq y$.

**Fact 3.2.7** *[14, Theorem 2.5.1(ii) and Proposition 2.5.6] If a function $f$ is both coercive and strictly convex then it has a unique minimizer, $\bar{x}$. That is,* $\operatorname{argmin}_{x\in\mathcal{H}} f(x) = \{\bar{x}\}$.

**Definition 3.2.8** *A function* $f : \mathcal{H} \to ]\infty, +\infty]$ *is* concave *if* $-f$ *is convex.*

**Definition 3.2.9** *A function* $f : \mathcal{H} \to ]\infty, +\infty]$ *is* affine *if* $f$ *is finite and both convex and concave.*

**Example 3.2.10** *Let* $f : \mathcal{H} \to ]\infty, +\infty]$ *be defined by* $x \mapsto \langle a, x\rangle - b$ *where* $a \in \mathcal{H}$ *and* $b \in \mathbb{R}$. *Then* $f$ *is an affine function.*

*Proof.* Let $x, y \in \mathcal{H}$ and $\theta \in \mathbb{R}$, then

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \langle a, \theta x + (1-\theta)y\rangle - b \\ &= \theta(\langle a, x\rangle - b) + (1-\theta)(\langle a, y\rangle - b) \\ &= \theta f(x) + (1-\theta) f(y). \end{aligned}$$

Therefore the conditions for convexity and concavity are both satisfied and $f$ is affine. ∎

Another method of determining if a function is convex is by checking the first and second order conditions for convexity.

**Fact 3.2.11 (First order condition)** *[5, Section 3.1.3] Let* $f : \mathbb{R}^n \to ]-\infty, +\infty]$ *be a differentiable function; that is, its gradient* $\nabla f$ *exists at each point of its open domain. Then* $f$ *is convex if and only if its domain is convex and*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle$$

*holds for all* $x, y \in \operatorname{dom} f$.

This condition says that for a convex function the first-order Taylor series approximation is a global underestimator of the function, and conversely if the Taylor approximation is a global underestimator then the function is convex.

**Fact 3.2.12 (Second order condition)** *[5, Section 3.1.4] Let $f : \mathbb{R}^n \to ]-\infty, +\infty]$ be a twice differentiable function; that is, its* Hessian *or second derivative $\nabla^2 f(x)$ exists at each point of it open domain. Then $f$ is convex if and only if* $\operatorname{dom} f$ *is convex and its Hessian is positive semidefinite:*

$$\nabla^2 f(x) \succeq 0.$$

**Remark 3.2.13** *A matrix $A \in \mathbb{R}^{m\times n}$ is positive semidefinite if $y^T Ay \geq 0$ for all $y \in \mathbb{R}^n$, and is denoted by $A \succeq 0$.*

**Fact 3.2.14 (Composition with affine mapping)** *[5, Section 3.2.2] Suppose $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathcal{H}^{m\times n}$, and $b \in \mathbb{R}^n$. Define $g : \mathbb{R}^n \to \mathbb{R}$ by*

$$g(x) = f(Ax + b);$$

*with* $\operatorname{dom} g = \{x : Ax + b \in \operatorname{dom} f\}$. *Then if $f$ is convex, so is $g$; if $f$ is concave, so is $g$.*

**Definition 3.2.15** *Given two functions $f, g$ from $\mathcal{H} \to ]-\infty, +\infty]$, $f$ is said to* majorize *$g$ if $f(x) \geq g(x)$ $\forall x \in \mathcal{H}$.*

**Definition 3.2.16** *Let $f$ be a function from $\mathcal{H} \to \mathbb{R}$. The* lower semi-continuous hull *of $f$ is the greatest lower semi-continuous function majorized by $f$, i.e, the function whose epigraph is the closure of the epigraph of $f$.*

**Definition 3.2.17** *Let $f \in \Gamma_0(\mathcal{H})$. The* recession function *of $f$ is the function $f^\infty : \mathcal{H} \to ]-\infty, +\infty]$ whose epigraph is the recession cone of the epigraph of $f$, $0^+(\operatorname{epi} f)$.*

**Fact 3.2.18** *[14, Theorem 2.1.5 (ii)] Let $f \in \Gamma_0(\mathcal{H})$ be such that $\operatorname{int}(\operatorname{dom} f) \neq \varnothing$ and let $t_0 \in \operatorname{dom} f$. The function $\varphi_{t_0} : \operatorname{dom} f \backslash \{t_0\} \to \mathbb{R}$ defined by*

$$\varphi_{t_0} := \frac{f(t) - f(t_0)}{t - t_0},$$

*is nondecreasing; if $f$ is strictly convex then $\varphi_{t_0}$ is increasing.*

**Proposition 3.2.19** *Let $f \in \Gamma_0(\mathcal{H})$ and $x_0 \in \operatorname{dom} f$, then the recession function of $f$ is*

$$f^\infty(u) = \lim_{t \to \infty} \frac{f(x_0 + tu) - f(x_0)}{t}$$

*for all $u \in \mathcal{H}$.*

*Proof.* Using Definition 3.2.17 and Definition 3.1.10 we have

$$\begin{aligned}(u, \lambda) \in 0^+(\operatorname{epi} f) &\Leftrightarrow \forall t > 0 : (x_0, f(x_0)) + t(u, \lambda) \in \operatorname{epi} f \\ &\Leftrightarrow \frac{f(x_0 + tu) - f(x_0)}{t} \leq \lambda \\ &\Leftrightarrow \sup_{t>0} \frac{f(x_0 + tu) - f(x_0)}{t} \leq \lambda.\end{aligned}$$

Fix $u \in \mathcal{H}$, the function $t \mapsto \frac{f(x_0+tu)-f(x_0)}{t}$ is nondecreasing on $]0, +\infty]$ due to the convexity of $f$ and Fact 3.2.18. Then,

$$(u, \lambda) \in 0^+(\operatorname{epi} f) \Leftrightarrow \lim_{t \to \infty} \frac{f(x_0 + tu) - f(x_0)}{t} = \sup_{t>0} \frac{f(x_0 + tu) - f(x_0)}{t} \leq \lambda.$$

Therefore for all $u \in \mathcal{H}$, $f^{\infty}(u) = \lim\limits_{t\to\infty} \frac{f(x_0+tu)-f(x_0)}{t}$. ■

### 3.2.1 Subgradients

In order to deal with nonsmooth convex functions, we will now introduce a concept that is analogous to a derivative of a differentiable function.

**Definition 3.2.20** *We say that $y \in \mathcal{H}$ is a* subgradient *of a convex function $f$ at the point $x$ if*

$$f(x) + \langle y, z - x\rangle \leq f(z) \ (\forall z \in \mathcal{H}). \tag{3.2}$$

The set of all subgradients of $f$ at $x$ is called the *subdifferential of $f$ at $x$* and is denoted by $\partial f(x)$. That is,

$$\partial f(x) := \{y \in \mathcal{H} : f(x) + \langle y, z - x\rangle \leq f(z) \ \forall z \in \mathcal{H}\}. \tag{3.3}$$

**Example 3.2.21** *Let $f(x) = |x|$, then*

$$\partial f(x) = \begin{cases} \{-1\}, & \text{if } x < 0; \\ [-1, 1], & \text{if } x = 0; \\ \{1\}, & \text{if } x > 0. \end{cases}$$

**Fact 3.2.22** *[12, Theorem 26.1] Let $f : \mathcal{H} \to\ ]-\infty, +\infty]$ be a differentiable function. Then $\partial f = \{\nabla f\}$.*

**Fact 3.2.23** *[14, Theorem 2.5.7] If $f$ is a proper convex function, then $x \in \operatorname{dom} f$ is a minimum point for $f$ if and only if $0 \in \partial f(x)$. In particular,* *if $f$ is differentiable at $x$ then $x$ is a minimum if $\nabla f(x) = 0$.*

**Fact 3.2.24 (Bronstead-Rockafellar)** *[14, Theorem 3.1.2] Let $\mathcal{H}$ be a Hilbert space and $f \in \Gamma_0(\mathcal{H})$. Then* $\operatorname{dom} f \subseteq \overline{\operatorname{dom} \partial f}$ *and* $\operatorname{dom} f^* \subseteq \overline{\operatorname{ran} \partial f}$.

### 3.2.2 Fenchel Conjugate

**Definition 3.2.25** *Let $f : \mathcal{H} \to \,]-\infty, +\infty]$. The* Fenchel conjugate, *or convex conjugate, of $f$ is defined as*

$$f^*(y) = \sup_{x \in \operatorname{dom} f} \left( \langle x, y \rangle - f(x) \right),$$

*for all $y \in \mathcal{H}$.*

It is interesting to note that since $\langle x, y \rangle - f(x)$ is affine with respect to $y$ for a fixed $x$, then $f^*$ is the supremum of a set of convex functions and is therefore convex regardless of whether $f$ is convex or not. In the case where $f$ is a finite, coercive and twice continuously differentiable function, the Fenchel conjugate is also referred to as the *Legendre transform* of $f$ [13, Example 11.9].

**Proposition 3.2.26 (Fenchel-Young Inequality)** *Let $f : \mathcal{H} \to \,]-\infty, +\infty]$. The following holds*

$$f(x) + f^*(x^*) \geq \langle x, x^* \rangle \tag{3.4}$$

*for all $x, x^* \in \mathcal{H}$.*

*Proof.* Follows directly from the definition of the conjugate function. ■

**Proposition 3.2.27** *If $f_1 \leq f_2$ then $f_1^* \geq f_2^*$.*

*Proof.* Let $y \in \mathcal{H}$, since $f_1 \leq f_2$ then $\langle x, y\rangle - f_1(x) \geq \langle x, y\rangle - f_2(x)$ for all $x \in \mathcal{H}$. Taking the supremum over $x$ of each side yields

$$\sup_{x\in\mathcal{H}} \left(\langle x, y\rangle - f_1(x)\right) \geq \sup_{x\in\mathcal{H}} \left(\langle x, y\rangle - f_2(x)\right).$$

Therefore $f_1^*(y) \geq f_2^*(y)$. ■

**Proposition 3.2.28** *Let* $f : \mathcal{H} \to ]-\infty, +\infty]$ *and* $c \in \mathbb{R}$ *be a constant. Then*

$$(f(\cdot - c))^*(x^*) = \langle x^*, c\rangle + f^*(x^*).$$

*Proof.* Let $x^* \in \mathcal{H}$, then $(f(\cdot - c))^*(x^*) = \sup_x \left(\langle x^*, x\rangle - f(x - c)\right)$. Letting $x - c = x'$, this becomes

$$\begin{aligned}(f(\cdot - c))^*(x^*) &= \sup_{x'} \big(\langle x^*, x' + c\rangle - f(x')\big) = \langle x^*, c\rangle + \sup_{x'} \big(\langle x^*, x'\rangle - f(x')\big)\\ &= \langle x^*, c\rangle + f^*(x^*).\end{aligned}$$

■

**Example 3.2.29** *Let* $\mathfrak{q} = \frac{1}{2}\|\cdot\|^2$, *then* $\mathfrak{q}^* = \mathfrak{q}$ *and this is the only self-conjugate function, i.e.* $f = f^*$.

*Proof.* From the definition of the conjugate we get that

$$\mathfrak{q}^*(y) = \sup_x \left(\langle x, y\rangle - \frac{1}{2}\|x\|^2\right)$$

Setting $h(x) = \langle x, y\rangle - \frac{1}{2}\|x\|^2$ and differentiating, we get $h'(x) = y - x$ and $h''(x) = -\operatorname{Id}$, which is negative definite: Hence, $h$ is strictly concave and therefore the critical point $x = y$ is the maximizer. Setting $x = y$ in the equation above yields $\mathsf{q}^*(y) = \frac{1}{2}\|y\|^2$.

On the other hand, suppose $f$ is a convex function satisfying $f = f^*$. Then $f$ is proper and using Proposition 3.2.26 we get

$$\langle x, x\rangle \leq f(x) + f^*(x) = 2f(x)$$
$$\Leftrightarrow \mathsf{q}(x) \leq f(x)$$

Then by Fact 3.2.27 $f^* \leq \mathsf{q}^* = \mathsf{q}$. Since $f^* = f$, we get $\mathsf{q} \leq f \leq \mathsf{q}$, therefore $f = \mathsf{q}$. ■

Due to its frequent use, from here on we will use the notation that

$$\mathsf{q} := \frac{1}{2}\|\cdot\|^2. \tag{3.5}$$

**Example 3.2.30** *Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \frac{1}{p}|x|^p$ then for $p = 1$*

$$f^*(x^*) = \begin{cases} 0, & \textit{if } -1 \leq x^* \leq 1; \\ +\infty, & \textit{otherwise.} \end{cases}$$

*And when $p > 1$*

$$f^*(x^*) = \frac{1}{q}|x^*|^q$$

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

*Proof.* When $p = 1$, $f(x) = |x|$ and $f^*(x^*) = \sup\limits_{x\in\mathbb{R}} (xx^* - |x|)$

If $x^* \geq 0$, $f^*(x^*) = \sup_{x \in \mathbb{R}_+} (xx^* - x) = \sup_{x \in \mathbb{R}_+} (x(x^* - 1)) = \begin{cases} +\infty, & \text{if } x^* > 1; \\ 0, & \text{if } x^* \leq 1. \end{cases}$

If $x^* < 0$, $f^*(x^*) = \sup_{x \in \mathbb{R}_-} (xx^* + x) = \sup_{x \in \mathbb{R}_-} (x(x^* + 1)) = \begin{cases} 0, & \text{if } x^* \geq -1; \\ +\infty, & \text{if } x^* < -1. \end{cases}$

Altogether,

$$f^*(x^*) = \begin{cases} 0, & \text{if } -1 \leq x^* \leq 1; \\ +\infty, & \textit{otherwise.} \end{cases}$$

When $p > 1$, $f(x) = \frac{1}{p}|x|^p$ and $f^*(x^*) = \sup_{x \in \mathbb{R}} \left( xx^* - \frac{1}{p}|x|^p \right)$
Differentiating to find the critical point,

$$\frac{d}{dx}(xx^* - \frac{1}{p}|x|^p) = x^* - |x|^{p-1} \operatorname{sgn}(x) = 0,$$

where $\operatorname{sgn}(x)$ denotes the sign of $x$. Then solving for $x$ yields

$$x^* = |x|^{p-1} \operatorname{sgn}(x) \Rightarrow |x^*| = |x|^{p-1} \Rightarrow |x| = |x^*|^{\frac{1}{p-1}}.$$

Substituting this back into the definition of the conjugate,

$$\begin{aligned} x|x|^{p-1} \operatorname{sgn}(x) - \frac{1}{p}|x|^p &= |x||x|^{p-1} - \frac{1}{p}|x|^p = |x|^p - \frac{1}{p}|x|^p \\ &= (\frac{p-1}{p})|x|^p = (\frac{p-1}{p})(|x^*|^{\frac{1}{p-1}})^p = (\frac{p-1}{p})|x^*|^{\frac{p}{p-1}}. \end{aligned}$$

Letting $q = \frac{p}{p-1}$, we get that $f^*(x^*) = \frac{1}{q}|x^*|^q$ where $\frac{1}{p} + \frac{1}{q} = 1$. ■

**Example 3.2.31** *Let $f(x) = \frac{1}{p}|x - c|^p$ for some constant $c$ with $p > 1$. Then $f$ is convex.*

*Proof.* Let

$$g(z) = (\frac{1}{q}|z|^q)^* = \frac{1}{p}|z|^p,$$

where $\frac{1}{p} + \frac{1}{q} = 1$. Then $g(z)$ is convex since it is a conjugate function and $f(x)$ is convex by Fact 3.2.14. ■

**Fact 3.2.32 (Biconjugate theorem)** *[14, Theorem 2.3.3] Assume that $f : \mathcal{H} \to ]-\infty, +\infty]$ is a proper function. Then $f^{**} := (f^*)^* = f$ if and only if $f \in \Gamma_0(\mathcal{H})$.*

**Fact 3.2.33 (Conjugate of the difference of functions)** *[8, Theorem 2.1] Let $g \in \Gamma_0(\mathcal{H})$ and let $h \in \Gamma_0(\mathcal{H})$ such that $h$ and $h^*$ both have full domain. Then*

$$(\forall x^* \in \mathcal{H})(g - h)^*(x^*) = \sup_{y^* \in \mathcal{H}} (g^*(y^*) - h^*(y^* - x^*)) \tag{3.6}$$

**Definition 3.2.34** *Let $f : \mathcal{H} \to ]-\infty, +\infty]$, and $A$ be a linear transformation from $\mathcal{H}$ to $\mathcal{H}$. Define*

$$Af(x) := \inf\{f(y) : Ay = x\}.$$

**Proposition 3.2.35** *Let $f : \mathcal{H} \to ]-\infty, +\infty]$, and $A$ be a linear transformation from $\mathcal{H}$ to $\mathcal{H}$. Then*

$$(Af)^* = f^* \circ A^*.$$

*Proof.* Let $x^* \in \mathcal{H}$, then

$$\begin{aligned}(Af)^*(x^*) &= \sup_{x\in\mathcal{H}} (\langle x, x^*\rangle - Af(x)) = \sup_{x\in\mathcal{H}} \left(\langle x, x^*\rangle - \inf_{\{y:Ay=x\}} f(y)\right)\\ &= \sup\left(\langle x, x^*\rangle - f(y) : (x,y) \in \mathcal{H}\times\mathcal{H}, Ay = x\right)\\ &= \sup\left(\langle Ay, x^*\rangle - f(y) : y \in \mathcal{H}\right)\\ &= \sup\left(\langle y, A^*x^*\rangle - f(y) : y \in \mathcal{H}\right)\\ &= f^*(A^*x^*) = (f^* \circ A^*)(x^*).\end{aligned}$$

■

**Fact 3.2.36 (Fenchel's Duality Theorem)** *[12, Theorem 31.1] Let $f$ and $g$ be a proper convex functions on $\mathcal{H}$. Suppose at least one of the following holds:*

*(i)* $\mathrm{ri}(\mathrm{dom}\, f) \cap \mathrm{ri}(\mathrm{dom}\, g) \neq \varnothing$,

*(ii) $f$ and $g$ are lower semicontinuous, and* $\mathrm{ri}(\mathrm{dom}\, f^*) \cap \mathrm{ri}(\mathrm{dom}\, g^*) \neq \varnothing$.

*Then*

$$\inf_x (f(x) + g(x)) = \sup_{x^*} \left(-g^*(-x^*) - f^*(x^*)\right).$$

**Fact 3.2.37** *[14, Theorem 2.3.1] Let $f : \mathcal{H} \to \,]-\infty, +\infty]$ and $\alpha, \beta \in \mathbb{R}$. If $\alpha > 0$ then $(\alpha f)^*(x^*) = \alpha f^*(\alpha^{-1}x^*)$ for every $x^* \in \mathcal{H}$; if $\beta \neq 0$ then $(f(\beta\cdot))^*(x^*) = f^*(\beta^{-1}x^*)$ for every $x^* \in \mathcal{H}$.*

### 3.2.3 Epi-addition and Epi-multiplication

**Definition 3.2.38** *Let $f_1, f_2 \in \Gamma_0(\mathcal{H})$. The infimal convolution, or epi-addition, is defined by*

$$(f \,\square\, g)(x) := \inf_{x_1+x_2=x} (f_1(x_1) + f_2(x_2)) \tag{3.7}$$

*for all $x \in \mathcal{H}$.*

The infimal convolution is said to be *exact* at a point $x$ if the infimum is attained.

**Definition 3.2.39** *Let $f \in \Gamma_0(\mathcal{H})$ and $\alpha \geq 0$. We define epi-multiplication by*

$$\alpha \star f = \begin{cases} \alpha f(\cdot/\alpha), & \text{if } \alpha > 0; \\ \iota_{\{0\}}, & \text{if } \alpha = 0. \end{cases} \tag{3.8}$$

**Proposition 3.2.40** *[1, Proposition 3.1] Let $f, f_1, \cdots, f_n \in \Gamma_0(\mathcal{H})$. Then the following properties hold:*

*(i)* $\operatorname{dom}(\alpha \star f) = \alpha(\operatorname{dom} f)$

*(ii)* $\operatorname{dom}(f_1 \,\square \cdots \square\, f_n) = (\operatorname{dom} f_1) + \cdots + (dom f_n)$

**Proposition 3.2.41** *Let $\alpha \geq 0$. Then the following hold:*

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

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

*(iii)* $(f_1 \,\square \cdots \square\, f_n)^* = f_1^* + \cdots + f_n^*$.

*Proof.* $(i)$ Let $x^* \in \mathcal{H}$, then we consider two cases.

(1) If $\alpha > 0$,

$$\begin{aligned}(\alpha f)^*(x^*) &= \sup_x (\langle x, x^* \rangle - \alpha f(x)) = \alpha \sup_x (\langle x, x^*/\alpha \rangle - f(x)) \\ &= \alpha f^*(x^*/\alpha).\end{aligned}$$

(2) If $\alpha = 0$

$$\begin{aligned}(\alpha f)^*(x^*) &= \sup_x (\langle x, x^* \rangle - 0]) = \begin{cases} 0, & \text{if } x^* = 0; \\ +\infty, & \text{otherwise} \end{cases} \qquad (3.9) \\ &= \iota_{\{0\}}(x^*) = (\alpha \star f^*)(x^*).\end{aligned}$$

Altogether, $(\alpha f)^*(x) = (\alpha \star f^*)(x^*)$.

$(ii)$

$$\begin{aligned}(\alpha \star f)^* &= \begin{cases} \sup_x (\langle x, x^* \rangle - \alpha f(x/\alpha)) & \text{if } \alpha > 0 \\ \sup_x \big(\langle x, x^* \rangle - \iota_{\{0\}}(x)\big) & \text{if } \alpha = 0 \end{cases} \\ &= \begin{cases} \alpha \sup_x (\langle x/\alpha, x^* \rangle - f(x/\alpha)) & \text{if } \alpha > 0 \\ 0 & \text{if } \alpha = 0 \end{cases} \\ &= \alpha f^*(x^*).\end{aligned}$$

(iii)

$$\begin{aligned}
(f_1 \,\square \cdots \square\, f_n)^*(x^*) &= \sup_x \left( \langle x, x^* \rangle - \inf_{x_1+\cdots+x_n=x} (f_1(x_1) + \cdots + f_n(x_n)) \right) \\
&= \sup_x \left( \langle x_1 + \cdots + x_n, x^* \rangle + \sup_{x_1+\cdots+x_n=x} (-f_1(x_1) - \cdots - f_n(x_n)) \right) \\
&= \sup_{x_1} (\langle x_1, x^* \rangle - f_1(x_1)) + \cdots + \sup_{x_1} (\langle x_1, x^* \rangle - f_1(x_1)) \\
&= (f_1^* + \cdots + f_n^*)(x^*).
\end{aligned}$$

■

**Fact 3.2.42** *[1, Fact 3.4] The following hold.*

*(i) If* $\operatorname{int}\operatorname{dom} f_1 \cap \cdots \operatorname{int}\operatorname{dom} f_{n-1} \cap \operatorname{dom} f_n \neq \varnothing$, *then* $(f_1 + \cdots + f_n)^* = f_1^* \,\square \cdots \square\, f_n^*$ *and the infimal convolution is exact.*

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

**Lemma 3.2.43** $(\lambda_1 \star (f_1 + \mathfrak{q}) \,\square \cdots \square\, \lambda_n \star (f_n + \mathfrak{q}))^* = \lambda_1(f_1^* \,\square\, \mathfrak{q}) + \cdots + \lambda_n(f_n^* \,\square\, \mathfrak{q})$

*Proof.* Using Proposition 3.2.41(iii), Proposition 3.2.41(i), Fact 3.2.42, and Example 3.2.29, we get that

$$\begin{aligned}
(\lambda_1 \star (f_1 + \mathfrak{q}) \,\square \cdots \square\, \lambda_n \star (f_n + \mathfrak{q}))^* &= (\lambda_1 \star (f_1 + \mathfrak{q}))^* + \cdots + (\lambda_n \star (f_n + \mathfrak{q}))^* \\
&= \lambda_1 (f_1 + \mathfrak{q})^* + \cdots + \lambda_n (f_n + \mathfrak{q})^* \\
&= \lambda_1 (f_1^* \,\square\, \mathfrak{q}^*) + \cdots + \lambda_n (f_n^* \,\square\, \mathfrak{q}^*) \\
&= \lambda_1 (f_1^* \,\square\, \mathfrak{q}) + \cdots + \lambda_n (f_n^* \,\square\, \mathfrak{q}).
\end{aligned} \tag{3.10}$$

■

The following lemma illustrates the beauty of the epi-multiplication notation and will be used for a couple of proofs in the following chapters.

**Lemma 3.2.44** *Let $f_i : \mathcal{H} \to ]-\infty, +\infty]$ for $1 \le i \le n$. Let $\mathbf{f} = (f_1, \cdots, f_n)$, $x = (x_1, \cdots, x_n)$ and $\tilde{f}(x) = \sum \lambda_i f_i(x_i)$: Then $\tilde{f}^*(x^*) = \sum_{i=1}^{n} \lambda_i \star f_i^*(x_i^*)$.*

*Proof.*

$$
\begin{aligned}
\tilde{f}^*(x^*) &= \sup_{x \in \mathcal{H}} \{\langle x, x^* \rangle - \tilde{f}(x)\} \\
&= \sup_{x=(x_1,\cdots,x_n) \in \mathcal{H}} \{\langle x_1, x_1^* \rangle + \cdots \langle x_n, x_n^* \rangle - \sum \lambda_i f_i(x_i)\} \\
&= \lambda_1 \sup_{x_1} \{\langle x_1, x_1^*/\lambda_1 \rangle - f_1(x_1)\} + \cdots + \lambda_n \sup_{x_n} \{\langle x_n, x_n^*/\lambda_n \rangle - f_n(x_n)\} \\
&= \lambda_1 f_1^*(\frac{x_1^*}{\lambda_1}) + \cdots + \lambda_n f_n^*(\frac{x_n^*}{\lambda_n}) = \lambda_1 \star f_1^*(x_1^*) + \cdots + \lambda_n \star f_n^*(x_n^*).
\end{aligned}
$$

■

## 3.3 Proximity Operators

**Definition 3.3.1** *The proximity operator, or proximal mapping, of a function $f \in \Gamma_0(\mathcal{H})$ is defined by*

$$
(\forall x \in \mathcal{H}) \ \ \operatorname{Prox}_f x = \operatorname{argmin}_{y \in \mathcal{H}} \left( f(y) + \mathfrak{q}(x-y) \right). \tag{3.11}
$$

**Fact 3.3.2** *[6, Section 2.2] For all $x \in \mathcal{H}$ and for all $p \in \mathcal{H}$*

$$
p = \operatorname{Prox}_f x \Leftrightarrow x - p \in \partial f(p),
$$

*and*

$$\operatorname{Prox}_f = (\operatorname{Id} + \partial f)^{-1}.$$

**Remark 3.3.3** *Note that since the function $y \mapsto f(y) + \mathfrak{q}(x - y)$ is strictly convex and supercoercive, it has a unique minimizer, $p = \operatorname{Prox}_f(x)$.*

## 3.4 Minimax Theory

Minimax problems are optimization problems that involve both minimization and maximization. Let $X$ and $Y$ be arbitray subsets of $\mathcal{H}$ with $X \neq \varnothing$ and $Y \neq \varnothing$, and let $F$ be a function from $X \times Y$ to $[-\infty, +\infty]$. Minimax theory deals with problems of the form $\sup\limits_{x \in X} \inf\limits_{y \in Y} F(x, y)$ or $\inf\limits_{y \in Y} \sup\limits_{x \in X} F(x, y)$. For more on minimax theory please see chapters 36 and 37 in [12].

**Definition 3.4.1** *If $\sup\limits_{x \in X} \inf\limits_{y \in Y} F(x, y) = \inf\limits_{y \in Y} \sup\limits_{x \in X} F(x, y)$ then the common value is called the* minimax *or the* saddle-value *of $F$.*

**Definition 3.4.2** *$F$ is a* concave-convex *function if $F(\cdot, y)$ is a concave function on $X$ for all $y \in Y$ and $F(x, \cdot)$ is convex on $Y$ for all $x \in X$. Similarly, $F$ is a* convex-concave *function if $F(\cdot, y)$ is convex on $X$ for all $y \in Y$ and $F(x, \cdot)$ is concave on $Y$ for all $x \in X$.*

The following fact gives us conditions for determining whether the saddle-value exists.

**Fact 3.4.3** *[12, Theorem 37.3] Let $F : \mathbb{R}^m \times \mathbb{R}^n \to ]-\infty, +\infty]$ be a proper concave-convex function with effective domain $X \times Y$. Then either of the* *following conditions implies that the saddle-value of $F$ exists. If both conditions hold, the saddle-value must be finite.*

*(a) The convex functions $F(x, \cdot)$ for $x \in \operatorname{ri} X$ have no common direction of recession.*

*(b) The convex functions $-F(\cdot, y)$ for $y \in \operatorname{ri} Y$ have no common direction of recession.*

# Chapter 4

# The Proximal Average

When averaging functions, the traditional arithmetic average

$$\lambda_1 f_1 + \cdots + \lambda_n f_n \tag{4.1}$$

is the natural place to begin. However, when the domains of the functions do not intersect then the result of (4.1) is a function that is everywhere infinity. The proximal average provides a useful method of averaging functions, even when their domains do not intersect.

In this chapter, we give a new proof to the self-duality of the proximal average:

$$(\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}))^* = \mathcal{P}(\mathbf{f}^*, \boldsymbol{\lambda}).$$

We also supply two self-dual smooth operators, $S_\beta f$ and $T_\beta f$.

For this chapter, let $f_1, \cdots, f_n \in \Gamma_0(\mathcal{H})$, and $\lambda_1, \cdots, \lambda_n$ be strictly positive real numbers such that $\sum_{i=1}^{n} \lambda_i = 1$.

## 4.1 Definitions

**Definition 4.1.1 (Proximal Average)** *Let* $\mathbf{f} = (f_1, \cdots, f_n)$ *and* $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_n)$. *The* $\boldsymbol{\lambda}$*-weighted proximal average of* $n$ *functions* $f_i$, $1 \leq i \leq n$, *is*

$$\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}) = \lambda_1 \star (f_1 + \mathfrak{q}) \,\Box \cdots \Box\, \lambda_n \star (f_n + \mathfrak{q}) - \mathfrak{q}. \tag{4.2}$$

*That is,*

$$(\forall x \in \mathcal{H})\ \mathcal{P}(\mathbf{f}, \boldsymbol{\lambda})(x) = \inf_{\sum_{i=1}^{n} x_i = x} \sum_{i=1}^{n} \Big( \lambda_i (f_i(x_i/\lambda_i) + \mathfrak{q}(x_i/\lambda_i)) \Big) - \mathfrak{q}(x).$$

This can be reformulated in two different ways.

**Proposition 4.1.2** *The proximal average can be equivalently defined by*

*(i)* $\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda})(x) = \inf_{\sum \lambda_i y_i = x} \Big( \sum_i \lambda_i f_i(y_i) + \sum_i \lambda_i \mathfrak{q}(y_i) \Big) - \mathfrak{q}(x)$

*(ii)* $\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}) = (\lambda_1 (f_1 + \mathfrak{q})^* + \cdots + \lambda_n (f_n + \mathfrak{q})^*)^* - \mathfrak{q}$.

*Proof.* Using the change of variables $y_i = x_i/\lambda_i$ in Definition 4.1.1 we immediately get (i).

For (ii), first note that by Proposition 3.2.40(ii),

$$(\forall i \in \mathbb{N})\, \mathrm{dom}(f_i^* \,\Box\, \mathfrak{q}) = (\mathrm{dom}\, f_i^*) + (\mathrm{dom}\, \mathfrak{q}) = \mathcal{H}.$$

Then Fact 3.2.42(i), Proposition 3.2.41(i), and Fact 3.2.32 yield

$$\begin{aligned}(\lambda_1(f_1+\mathfrak{q})^*+\cdots+\lambda_n(f_n+\mathfrak{q})^*)^* &= (\lambda_1(f_1+\mathfrak{q})^*)^*\,\Box\cdots\Box(\lambda_n(f_n+\mathfrak{q})^*)^*\\ &= \lambda_1\star(f_1+\mathfrak{q})^{**}\,\Box\cdots\Box\,\lambda_n\star(f_n+\mathfrak{q})^{**}\\ &= \lambda_1\star(f_1+\mathfrak{q})\,\Box\cdots\Box\,\lambda_n\star(f_n+\mathfrak{q}).\end{aligned}$$

■

## 4.2 Properties

**Theorem 4.2.1 (Domain)**

$$\operatorname{dom}\mathcal{P}(\mathbf{f},\boldsymbol{\lambda}) = \lambda_1\operatorname{dom}f_1+\cdots+\lambda_n\operatorname{dom}f_n$$

*Proof.* Using Proposition 3.2.40i and Proposition 3.2.40 ii

$$\begin{aligned}\operatorname{dom}\mathcal{P}(\mathbf{f},\boldsymbol{\lambda}) &= \operatorname{dom}(\lambda_1\star(f_1+\mathfrak{q})\,\Box\cdots\Box\,\lambda_n\star(f_n+\mathfrak{q}))\\ &= \operatorname{dom}(\lambda_1\star(f_1+\mathfrak{q}))+\cdots+\operatorname{dom}(\lambda_n\star(f_n+\mathfrak{q}))\\ &= \lambda_1\operatorname{dom}(f_1+\mathfrak{q})+\cdots+\lambda_n\operatorname{dom}(f_n+\mathfrak{q})\\ &= \lambda_1\operatorname{dom}f_1+\cdots+\lambda_n\operatorname{dom}f_n.\end{aligned} \tag{4.3}$$

■

**Corollary 4.2.2** *If at least one function $f_i$ has full domain and $\lambda_i > 0$ then $\mathcal{P}(\mathbf{f},\boldsymbol{\lambda})$ has full domain.*

### 4.2.1 Fenchel Conjugate

Next we examine the conjugate of the proximal average. The purpose of this section is to give a new proof for $(\mathcal{P}(\mathbf{f},\boldsymbol{\lambda}))^* = \mathcal{P}(\mathbf{f}^*,\boldsymbol{\lambda})$ without using Toland's Duality Theorem. First we must prove the following lemma, which will also be used later to reformulate the proximal average.

**Lemma 4.2.3** *The following identity holds*

$$\sum_{i=1}^{n} \lambda_i \mathfrak{q}(y_i) - \mathfrak{q}(\sum_{i=1}^{n} \lambda_i y_i) = \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j \|y_i - y_j\|^2. \tag{4.4}$$

*Proof.* Consider first the left hand side of (4.4),

$$\begin{aligned}
&\sum_{i=1}^{n} \lambda_i \mathfrak{q}(y_i) - \mathfrak{q}(\sum_{i=1}^{n} \lambda_i y_i) \\
&= \frac{1}{2}\sum_{i=1}^{n} \lambda_i \|y_i\|^2 - \frac{1}{2}\|\sum_{i=1}^{n} \lambda_i y_i\|^2 \\
&= \frac{1}{2}\sum_{i=1}^{n} \lambda_i \|y_i\|^2 - \frac{1}{2}\sum_{i=1}^{n} \lambda_i^2 \|y_i\|^2 - \sum_{i\neq j} \lambda_i\lambda_j \langle y_i, y_j\rangle. 
\end{aligned} \tag{4.5}$$

On the other hand, from the right hand side we get,

$$
\begin{aligned}
&\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\|y_i-y_j\|^2\\
&=\frac{1}{4}\sum_{i=1}^{n}\lambda_i\bigg(\sum_{j=1}^{n}\lambda_j\|y_i-y_j\|^2\bigg)\\
&=\frac{1}{4}\sum_{i=1}^{n}\lambda_i\bigg(\sum_{j=1}^{n}(\lambda_j\|y_i\|^2-2\lambda_j\langle y_i,y_j\rangle+\lambda_j\|y_j\|^2)\bigg)\\
&=\frac{1}{4}\sum_{i=1}^{n}\lambda_i\bigg(\sum_{j=1}^{n}\lambda_j\|y_i\|^2-2\sum_{j=1}^{n}\lambda_j\langle y_i,y_j\rangle+\sum_{j=1}^{n}\lambda_j\|y_j\|^2\bigg)\\
&=\frac{1}{4}\sum_{i=1}^{n}\lambda_i\bigg(\|y_i\|^2-2\sum_{j=1}^{n}\lambda_j\langle y_i,y_j\rangle+\sum_{j=1}^{n}\lambda_j\|y_j\|^2\bigg)\\
&=\frac{1}{4}\bigg(\sum_{i=1}^{n}\lambda_i\|y_i\|^2-2\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\langle y_i,y_j\rangle+\sum_{i=1}^{n}\lambda_i\bigg(\sum_{j=1}^{n}\lambda_j\|y_j\|^2\bigg)\bigg)\\
&=\frac{1}{2}\sum_{i=1}^{n}\lambda_i\|y_i\|^2-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\langle y_i,y_j\rangle\\
&=\frac{1}{2}\sum_{i=1}^{n}\lambda_i\|y_i\|^2-\frac{1}{2}\sum_{i=1}^{n}\lambda_i^2\|y_i\|^2-\sum_{i\neq j}\lambda_i\lambda_j\langle y_i,y_j\rangle. \qquad (4.6)
\end{aligned}
$$

Since (4.5) and (4.6) are equal, the proof is complete. ■

The following lemma is new.

**Lemma 4.2.4** *Let*

$$g(y_1,\cdots,y_n)=\lambda_1\mathfrak{q}(y_1)+\cdots+\lambda_n\mathfrak{q}(y_n)-\mathfrak{q}(\lambda_1y_1+\cdots+\lambda_ny_n),$$

*for* $(y_1, \cdots, y_n) \in \mathcal{H}^n$ *and* $\sum_{i=1}^{n} \lambda_n = 1$. *Then*

$$g^*(x_1^*, \cdots, x_n^*) = \begin{cases} \lambda_1 \star \mathfrak{q}(x_1^*) + \cdots + \lambda_n \star \mathfrak{q}(x_n^*), & \textit{if } x_1^* + \cdots + x_n^* = 0; \\ +\infty, & \textit{otherwise.} \end{cases}$$

*Proof.* For every $(x_1^*, \cdots, x_n^*) \in \mathcal{H}^n$ we have

$$g^*(x_1^*, \cdots, x_n^*) = \sup_{y_1, \cdots, y_n} \Big( \langle x_1^*, y_1 \rangle + \cdots + \langle x_n^*, y_n \rangle - \lambda_1 \mathfrak{q}(y_1) - \cdots - \lambda_n \mathfrak{q}(y_n) + \mathfrak{q}(\lambda_1 y_1 + \cdots + \lambda_n y_n) \Big). \tag{4.7}$$

In light of Lemma 4.2.3, the equation within the supremum is concave and therefore solving for critical points will yield the supremum. Taking the partial derivatives, with respect to $y_i$ for $i = 1...n$, and setting them equal to zero gives

$$\begin{aligned} x_1^* - \lambda_1 y_1 + (\lambda_1 y_1 + \cdots + \lambda_n y_n)\lambda_1 &= 0 \\ &\vdots \\ x_n^* - \lambda_n y_n + (\lambda_1 y_1 + \cdots + \lambda_n y_n)\lambda_n &= 0. \end{aligned} \tag{4.8}$$

Then taking the sum of all of the above equations yields

$$\begin{aligned} &\sum_{i=1}^{n} x_i^* - \sum_{i=1}^{n} \lambda_i y_i + \sum_{i=1}^{n} \lambda_i (\sum_{i=1}^{n} \lambda_i y_i) = 0 \\ \Leftrightarrow &\sum_{i=1}^{n} x_i^* - \sum_{i=1}^{n} \lambda_i y_i + \sum_{i=1}^{n} \lambda_i y_i \Leftrightarrow \sum_{i=1}^{n} x_i^* = 0. \end{aligned}$$

So $x_1^* + \cdots + x_n^* = 0$ and if we let $y_1 = x_1^*/\lambda_1, \cdots, y_n = x_n^*/\lambda_n$ then $(y_1, \cdots, y_n)$ is a solution to (4.8). Consequently, $\lambda_1 y_1 + \cdots + \lambda_n y_n = 0$ and $\langle x_i^*, x_i^*/\lambda_i\rangle = 2\lambda_i \mathfrak{q}(x_i^*/\lambda_i) = 2(\lambda_i \star \mathfrak{q})(x_i^*)$ for $i = 1 \cdots n$. This gives us that

$$\begin{aligned}
&\langle x_1^*, y_1\rangle + \cdots + \langle x_n^*, y_n\rangle - \lambda_1 \mathfrak{q}(y_1) - \cdots - \lambda_n \mathfrak{q}(y_n) \\
&= 2\sum_{i=1}^{n} \lambda_i \star \mathfrak{q}(x_i^*) - \lambda_1 \star \mathfrak{q}(x_1^*) - \cdots - \lambda_n \star \mathfrak{q}(x_n^*) \\
&= \lambda_1 \star \mathfrak{q}(x_1^*) + \cdots + \lambda_n \star \mathfrak{q}(x_n^*)
\end{aligned}$$

If $\sum_i x_i^* \neq 0$ then let $y_1 = y_2 = \cdots = y_n = y$ and then (4.7) becomes

$$\begin{aligned}
g^*(x_1^*, \cdots, x_n^*) &\geq \sup_y \left( \langle \sum_{i=1}^{n} x_i^*, y\rangle - \lambda_1 \mathfrak{q}(y) - \cdots - \lambda_n \mathfrak{q}(y) + \mathfrak{q}(y) \right) \\
&\geq \sup_y \left( \langle \sum_{i=1}^{n} x_i^*, y\rangle - \mathfrak{q}(y) + \mathfrak{q}(y) \right) \\
&\geq \sup_y \left( \langle \sum_{i=1}^{n} x_i^*, y\rangle \right) = +\infty.
\end{aligned}$$

Thus,

$$g^*(x_1^*, \cdots, x_n^*) = \begin{cases} \lambda_1 \star \mathfrak{q}(x_1^*) + \cdots + \lambda_n \star \mathfrak{q}(x_n^*), & \text{if } x_1^* + \cdots + x_n^* = 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

■

**Remark 4.2.5** *It can be noted that* $\frac{\partial g}{\partial y_i} = \lambda_i y_i - (\lambda_1 y_1 + \cdots + \lambda_n y_n)\lambda_i$ *for*

*$i = 1 \cdots n$ so that $\frac{\partial g}{\partial y_1} + \cdots + \frac{\partial g}{\partial y_n} = 0$. This means that*

$$\operatorname{ran} \partial g \subseteq \{(x_1^*, \cdots, x_n^*) : x_1^* + \cdots + x_n^* = 0\}.$$

*Conversely, if $x_1^* + \cdots + x_n^* = 0$ and we let $y_1 = x_1^*/\lambda_1, \cdots, y_n = x_n^*/\lambda_n$ then $\nabla g(y_1, \cdots, y_n) = (x_1^*, \cdots, x_n^*)$ and*

$$\{(x_1^*, \cdots, x_n^*) : x_1^* + \cdots + x_n^* = 0\} \subseteq \operatorname{ran} \partial g.$$

*Therefore $\operatorname{ran} \partial g = \{(x_1^*, \cdots, x_n^*) : x_1^* + \cdots + x_n^* = 0\}$. Now since $\operatorname{ran} \partial g \subseteq \operatorname{dom} g^* \subseteq \overline{\operatorname{ran} \partial g}$ by Fact 3.2.24 and we have $\operatorname{ran} \partial g = \overline{\operatorname{ran} \partial g}$ then $\operatorname{dom} g^* = \{(x_1^*, \cdots, x_n^*) : x_1^* + \cdots + x_n^* = 0\}$, as we saw in the previous lemma.*

**Theorem 4.2.6 (Fenchel Conjugate)** *[1, Theorem 5.1]*

$$(\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}))^* = \mathcal{P}(\mathbf{f}^*, \boldsymbol{\lambda}) \tag{4.9}$$

*Proof.* Let

$$f(x) = \mathcal{P}(\mathbf{f}, \boldsymbol{\lambda})(x) = \inf_{\sum_{i=1}^{n} \lambda_i y_i = x} \Big(\lambda_1 f_1(y_1) + \cdots \lambda_n f_n(y_n) + \lambda_1 \mathfrak{q}(y_1) + \cdots + \mathfrak{q}(y_n) \\ - \mathfrak{q}(\lambda_1 y_1 + \cdots + \lambda_n y_n)\Big),$$

and let $A : \mathcal{H}^n \to \mathcal{H}$ be a linear operator defined by $A = \begin{pmatrix} \lambda_1, & \cdots, & \lambda_n \end{pmatrix}$, i.e. $A(x_1, \cdots, x_n) = \sum_{i=1}^{n} \lambda_i x_i$. Then $A^* : \mathcal{H} \to \mathcal{H}^n$, the adjoint of $A$, is

$A^* = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$, i.e. $A^*(z) = (\lambda_1 z, \cdots, \lambda_n z)$. Then we can write $f$ as $f = AF$ where

$$F(y) = \lambda_1 f_1(y_1) + \cdots + \lambda_n f_n(y_n) + \lambda_1 \mathfrak{q}(y_1) + \cdots + \lambda_n \mathfrak{q}(y_n) - \mathfrak{q}(\lambda_1 y_1 + \cdots + \lambda_n y_n)$$

and

$$AF(y) := \inf_{\{x : Ax = y\}} F(x).$$

For ease of notation, say that $F = j + g$ where $j(y) = \lambda_1 f_1(y_1) + \cdots + \lambda_n f_n(y_n)$ and $g(y) = \lambda_1 \mathfrak{q}(y_1) + \cdots + \lambda_n \mathfrak{q}(y_n) - \mathfrak{q}(\lambda_1 y_1 + \cdots + \lambda_n y_n)$. By Proposition 3.2.35,

$$f^*(x^*) = (AF)^*(x^*) = (F^* \circ A^*)(x^*).$$

Since $j \in \Gamma_0(\mathcal{H})$ and $\operatorname{dom} g = \mathcal{H} \times \cdots \times \mathcal{H}$, then $\operatorname{int}(\operatorname{dom} f) \cap \operatorname{int}(\operatorname{dom} g) \neq \varnothing$ and we can use Fact 3.2.42(i) and Lemma 3.2.44 to get

$$\begin{aligned} f^*(x^*) &= (j^* \,\Box\, g^*) A^*(x^*) \\ &= \Big( (\lambda_1 \star f_1^* + \cdots + \lambda_n \star f_n^*) \,\Box\, g^* \Big) A^*(x^*). \end{aligned}$$

Then using Lemma 4.2.4 we get

$$\begin{aligned}
f^*(x^*) &= \Big((\lambda_1 \star f_1^* + \cdots + \lambda_n \star f_n^*) \,\square\, g^*\Big)(\lambda_1 x^*, \cdots, \lambda_n x^*) \\
&= \inf_{y_1,\cdots,y_n} \Big(\lambda_1 f_1^*(y_1/\lambda_1) + \cdots + \lambda_n f_n^*(y_n/\lambda_n) + g^*(\lambda_1 x^* - y_1, \cdots, \lambda_n x^* - y_n)\Big) \\
&= \inf_{\lambda_1 x^* - y_1 + \cdots + \lambda_n x^* - y_n = 0} \Big(\lambda_1 f_1^*(y_1/\lambda_1) + \cdots + \lambda_n f_n^*(y_n/\lambda_n) \\
&\qquad + \lambda_1 \mathfrak{q}(\frac{\lambda_1 x^* - y_1}{\lambda_1}) + \cdots + \lambda_n \mathfrak{q}(\frac{\lambda_n x^* - y_n}{\lambda_n})\Big) \\
&= \inf_{x^* = y_1 + \cdots + y_n} \Big(\lambda_1 f_1^*(y_1/\lambda_1) + \cdots + \lambda_n f_n^*(y_n/\lambda_n) + \lambda_1 \mathfrak{q}(x^* - \frac{y_1}{\lambda_1}) + \cdots \\
&\qquad + \lambda_n \mathfrak{q}(x^* - \frac{y_n}{\lambda_n})\Big).
\end{aligned}$$

Expanding the last set of terms in the above equation yields

$$\begin{aligned}
\lambda_i \mathfrak{q}(x^* - \frac{y_i}{\lambda_i}) = \frac{\lambda_i}{2}\|x^* - \frac{y_i}{\lambda_i}\|^2 &= \frac{\lambda_i}{2}\|x^*\|^2 - \langle x^*, y_i\rangle + \frac{\lambda_i}{2}\|\frac{y_i}{\lambda_i}\|^2 \\
&= \lambda_i \mathfrak{q}(x^*) - \langle x^*, y_i\rangle + \lambda_i \mathfrak{q}(\frac{y_i}{\lambda_i})
\end{aligned}$$

for all $i = 1 \cdots n$. Taking the sum of all of these terms and substituting back into the infimum equation produces

$$\begin{aligned}
f^*(x^*) &= \inf_{x^* = y_1 + \cdots + y_n} \Big(\lambda_1 f_1^*(y_1/\lambda_1) + \cdots + \lambda_n f_n^*(y_n/\lambda_n) + \mathfrak{q}(x^*) - \langle x^*, y_1 + \cdots + y_n\rangle \\
&\qquad + \lambda_1 \mathfrak{q}(\frac{y_1}{\lambda_1}) + \cdots + \lambda_n \mathfrak{q}(\frac{y_n}{\lambda_n})\Big) \\
&= \inf_{x^* = y_1 + \cdots + y_n} \Big(\lambda_1 f_1^*(y_1/\lambda_1) + \cdots + \lambda_n f_n^*(y_n/\lambda_n) + \lambda_1 \mathfrak{q}(\frac{y_1}{\lambda_1}) + \cdots + \lambda_n \mathfrak{q}(\frac{y_n}{\lambda_n}) \\
&\qquad - \mathfrak{q}(x^*)\Big).
\end{aligned}$$

Making the simple change of variable $z_i = y_i/\lambda_i$ for $i = 1 \cdots n$ generates

$$f^*(x^*) = \inf_{x^*=\lambda_1 z_1+\cdots+\lambda_n z_n} \Big(\lambda_1 f_1^*(z_1) + \cdots + \lambda_n f_n^*(z_n) + \lambda_1 \mathfrak{q}(z_1) + \cdots \lambda_n \mathfrak{q}(z_n) - \mathfrak{q}(x^*)\Big)$$
$$= \mathcal{P}(\mathbf{f}^*, \boldsymbol{\lambda}).$$

■

**Example 4.2.7** *Let* $\mathbf{f} = (f, f^*)$ *and* $\boldsymbol{\lambda} = \left(\frac{1}{2}, \frac{1}{2}\right)$, *then* $\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}) = \mathfrak{q}$.

*Proof.* By Fact 4.2.6,

$$(\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}))^* = \mathcal{P}(\mathbf{f}^*, \boldsymbol{\lambda}).$$

Since $\mathbf{f}^* = (f^*, f^{**}) = (f^*, f)$, then we get that $\mathcal{P}(\mathbf{f}^*, \boldsymbol{\lambda}) = \mathcal{P}(\mathbf{f}, \lambda)$. Therefore, using Example 3.2.29 we see that $(\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}))^* = \mathfrak{q}$. ■

**Corollary 4.2.8** $\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda})$ *is convex, lower semicontinuous, and proper.*

*Proof.* We can apply Fact 4.2.6 twice to see that

$$(\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}))^{**} = (\mathcal{P}(\mathbf{f}^*, \boldsymbol{\lambda}))^* = \mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}).$$

Therefore, in light of Fact 3.2.32 $\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}) \in \Gamma_0(\mathcal{H})$. ■

**Definition 4.2.9 (Moreau Envelope)** *The Moreau envelope of* $f \in \Gamma_0(\mathcal{H})$ *with parameter* $\mu > 0$ *is*

$$e_\mu f = f \,\square\, \mu \star \mathfrak{q}.$$

**Fact 4.2.10 (Moreau Envelope of the Proximal Average)** *[1, Theorem 6.2]* *(i)* $e_\mu \mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}) = \lambda_1 e_\mu f_1 + \cdots + \lambda_n e_\mu f_n$

*(ii)* $(e_\mu \mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}))^* = \lambda_1 \star (e_\mu f_1)^* \,\square\, \cdots \,\square\, \lambda_n \star (e_\mu f_n)^*$

**Fact 4.2.11 (Proximal Mapping)** *[1, Theorem 6.7]*

$$\operatorname{Prox}_{\mathcal{P}(\mathbf{f}, \boldsymbol{\lambda})} = \lambda_1 \operatorname{Prox}_{f_1} + \cdots + \lambda_n \operatorname{Prox}_{f_n}$$

## 4.3 Applications

### 4.3.1 Self-dual Smooth Approximations

A function $f$ in $\mathbb{R}^n$ is *smooth* if $f$ is finite and differentiable everywhere in $\mathbb{R}^n$. It can be helpful in cases of nondifferentiable convex functions to find a smooth approximation of the function. Here, two smooth approximations are defined using the proximal average. The first smooth operator, $S_\beta f$, has a simple expression in terms of the Moreau envelope which can be seen in [9]. The second smooth operator, $T_\beta f$, has a simple expression in terms of the proximal average and is new. Both approximations are "self-dual" in the sense that the conjugate of the approximation of $f$ is equal to the approximation of the conjugate of $f$.

Recall that,

$$\mathcal{P}(f_1, \lambda_1, \cdots, f_n, \lambda_n) := (\lambda_1 (f_1 + \mathsf{q})^* + \cdots + \lambda_n (f_n + \mathsf{q})^*)^* - \mathsf{q}.$$

For $0 \leq \beta \leq 1$ and a proper lower-semicontinuous function $f$, define $S_\beta f$ :

$\mathbb{R}^n \to ]-\infty, +\infty]$ by

$$S_\beta f(x) := (1+\beta)^2 \mathcal{P}(f, \frac{1-\beta}{1+\beta}, \mathfrak{q}, \frac{2\beta}{1+\beta})(\frac{x}{1+\beta}) \tag{4.10}$$

for all $x \in \mathbb{R}^n$.

**Theorem 4.3.1** *(i)* $(S_\beta f)^* = S_\beta f^*$

*(ii) When* $0 < \beta \leq 1$, *we have*

$$S_\beta f = (1+\beta)\mathcal{P}(1-\beta, f, \beta, 0) + \beta\mathfrak{q} = (1-\beta)^2(f \,\square\, \frac{1}{\beta}\mathfrak{q}) + \beta\mathfrak{q}. \tag{4.11}$$

*Therefore when* $\beta \to 0$, $S_\beta f \to \lim_{\beta \to 0+}(f \,\square\, \frac{1}{\beta}\mathfrak{q}) = f$.

*Proof.* (i) By Theorem 4.2.6, we have

$$(S_\beta f)^* = \left((1+\beta)^2 \mathcal{P}(f, \frac{1-\beta}{1+\beta}, \mathfrak{q}, \frac{2\beta}{1+\beta})(\frac{\cdot}{1+\beta})\right)^*$$

By Fact 3.2.37 and Proposition 3.2.41(i) we then get

$$\begin{aligned}
(S_\beta f)^* &= \left((1+\beta)^2 \mathcal{P}(f, \frac{1-\beta}{1+\beta}, \mathfrak{q}, \frac{2\beta}{1+\beta})\right)^*((1+\beta)\cdot) \\
&= (1+\beta)^2 \left(\mathcal{P}(f, \frac{1-\beta}{1+\beta}, \mathfrak{q}, \frac{2\beta}{1+\beta})\right)^*(\frac{(1+\beta)\cdot}{(1+\beta)^2}) \\
&= (1+\beta)^2 \mathcal{P}\left(f^*, \frac{1-\beta}{1+\beta}, \mathfrak{q}, \frac{2\beta}{1+\beta}\right)(\frac{\cdot}{(1+\beta)}) \\
&= S_\beta f^*.
\end{aligned}$$

(ii) For every $x$, by the definition of the proximal average, Proposition 3.2.41(i), and Example 3.2.29

$$\begin{aligned}
S_\beta f(x) &= (1+\beta)^2\left[\left(\frac{1-\beta}{1+\beta}(f+\mathfrak{q})^* + \frac{2\beta}{1+\beta}(\mathfrak{q}+\mathfrak{q})^*\right)^*\Big(\frac{x}{1+\beta}\Big) - \frac{1}{2}\frac{\|x\|^2}{(1+\beta)^2}\right] \\
&= (1+\beta)^2\left(\frac{1-\beta}{1+\beta}(f+\mathfrak{q})^* + \frac{\beta}{1+\beta}\mathfrak{q}\right)^*\Big(\frac{x}{1+\beta}\Big) - \mathfrak{q}(x) \\
&= (1+\beta)\Big((1-\beta)(f+\mathfrak{q})^* + \beta\mathfrak{q}\Big)^*(x) - \mathfrak{q}(x) \\
&= (1+\beta)\left[\Big((1-\beta)(f+\mathfrak{q})^* + \beta\mathfrak{q}\Big)^*(x) - \mathfrak{q}(x)\right] + \beta\mathfrak{q}(x).
\end{aligned}$$

This is the first equality in (4.11). To continue, apply Fact 3.2.42(i), Fact 3.2.32, Example 3.2.29, and Proposition 3.2.41(i) to

$$S_\beta f = (1+\beta)\Big((1-\beta)(f+\mathfrak{q})^* + \beta\mathfrak{q}\Big)^*(x) - \mathfrak{q}(x),$$

to get

$$\begin{aligned}
S_\beta f(x) &= (1+\beta)\left[((1-\beta)(f+\mathfrak{q})^*)^* \,\square\, (\beta\mathfrak{q})^*\right](x) - \mathfrak{q}(x) \\
&= (1+\beta)\left[((1-\beta)(f+\mathfrak{q})(\frac{\cdot}{1-\beta}))\,\square\,\frac{1}{\beta}\mathfrak{q}\right](x) - \mathfrak{q}(x).
\end{aligned}$$

Using Definition 3.2.38,

$$\begin{aligned}
S_\beta f(x) &= (1+\beta)\inf_u\left[(1-\beta)(f+\mathfrak{q})(\frac{u}{1-\beta}) + \frac{1}{\beta}\mathfrak{q}(x-u)\right] - \mathfrak{q}(x) \\
&= (1+\beta)\inf_u\left[(1-\beta)f(\frac{u}{1-\beta}) + (1-\beta)\mathfrak{q}(\frac{u}{1-\beta}) + \frac{1}{\beta}\mathfrak{q}(x-u)\right] - \mathfrak{q}(x) \\
&= (1-\beta^2)\inf_u\left[f(\frac{u}{1-\beta}) + \mathfrak{q}(\frac{u}{1-\beta}) + \frac{1}{\beta(1-\beta)}\mathfrak{q}(x-u) - \frac{1}{1-\beta^2}\mathfrak{q}(x)\right].
\end{aligned} \tag{4.12}$$

Simplifying the portion of (4.12) containing $\mathfrak{q}$ and using $\frac{1}{(1-\beta)^2}+\frac{1}{\beta(1-\beta)}=\frac{1}{\beta(1-\beta)^2}$ and $\frac{1}{\beta(1-\beta)}-\frac{1}{(1-\beta^2)}=\frac{1}{\beta(1-\beta^2)}$

$$
\begin{aligned}
&\mathfrak{q}(\frac{u}{1-\beta})+\frac{1}{\beta(1-\beta)}\mathfrak{q}(x-u)-\frac{1}{1-\beta^2}\mathfrak{q}(x)\\
&=\frac{1}{(1-\beta)^2}\frac{\|u\|^2}{2}+\frac{1}{\beta(1-\beta)}\frac{\|x\|^2}{2}-\frac{1}{\beta(1-\beta)}\langle x,u\rangle+\frac{1}{\beta(1-\beta)}\frac{\|u\|^2}{2}-\frac{1}{1-\beta^2}\frac{\|x\|^2}{2}\\
&=\frac{1}{\beta(1-\beta)^2}\frac{\|u\|^2}{2}-\frac{1}{\beta(1-\beta)}\langle x,u\rangle+\frac{1}{\beta(1-\beta^2)}\frac{\|x\|^2}{2}\\
&=\frac{1}{2\beta}(\frac{\|u\|^2}{(1-\beta)^2}-2\langle x,\frac{u}{1-\beta}\rangle+\|x\|^2)+(\frac{1}{\beta(1-\beta^2)}-\frac{1}{\beta})\frac{\|x\|^2}{2}\\
&=\frac{1}{2\beta}\|x-\frac{u}{1-\beta}\|^2+\frac{\beta}{1-\beta^2}\frac{\|x\|^2}{2}.
\end{aligned}
$$

Plugging this back into (4.12) gives

$$
\begin{aligned}
S_\beta f(x)&=(1-\beta^2)\inf_u\left(f(\frac{u}{1-\beta})+\frac{1}{\beta}\mathfrak{q}(x-\frac{u}{1-\beta})+\frac{\beta}{1-\beta^2}\mathfrak{q}(x)\right)\\
&=(1-\beta^2)\inf_w\left(f(w)+\frac{1}{\beta}\mathfrak{q}(x-w)\right)+\beta\mathfrak{q}(x)\\
&=(1-\beta^2)\left(f\,\Box\,\frac{1}{\beta}\mathfrak{q}\right)(x)+\beta\mathfrak{q}(x),
\end{aligned}
$$

which is the second equality of (4.11). The convergence result follows from [13, Theorem 1.25]. ■

Another smooth operator is defined by

$$T_\beta f:=\mathcal{P}(f,1-\beta,\mathfrak{q},\beta). \tag{4.13}$$

**Theorem 4.3.2** *(i)* $(T_\beta f)^*=T_\beta f^*$

*(ii) When $0 < \beta \leq 1$, we have*

$$T_\beta f = (1-\beta)\left(f \,\square\, \frac{2-\beta}{\beta}\mathfrak{q}\right)\left(\frac{2}{2-\beta}\cdot\right) + \frac{\beta}{2-\beta}\mathfrak{q}.$$

*Proof.* (i) This follows from Theorem 4.2.6 and Example 3.2.29

$$\begin{aligned}(T_\beta f)^* &= (\mathcal{P}(f, 1-\beta, \mathfrak{q}, \beta))^* = \mathcal{P}(f^*, 1-\beta, \mathfrak{q}^*, \beta) = \mathcal{P}(f^*, 1-\beta, \mathfrak{q}, \beta)\\ &= T_\beta f^*\end{aligned}$$

(ii) Applying Proposition 4.1.2(ii), Proposition 3.2.41(i), Example 3.2.29, Fact 3.2.42(i), and Definition 3.2.38,

$$\begin{aligned}T_\beta f(x) &= ((1-\beta)(f+\mathfrak{q})^* + \beta(2\mathfrak{q})^*)^*(x) - \mathfrak{q}(x)\\ &= \left((1-\beta)(f+\mathfrak{q})^* + \beta\frac{\mathfrak{q}}{2}\right)^*(x) - \mathfrak{q}(x)\\ &= \left((1-\beta)(f+\mathfrak{q})(\frac{\cdot}{1-\beta}) \,\square\, \frac{2}{\beta}\mathfrak{q}\right)(x) - \mathfrak{q}(x)\\ &= \inf_u \left((1-\beta)f(\frac{u}{1-\beta}) + (1-\beta)\mathfrak{q}(\frac{u}{1-\beta}) + \frac{2}{\beta}\mathfrak{q}(x-u)\right) - \mathfrak{q}(x).\end{aligned}$$

This is equivalent to,

$$T_\beta f(x) = (1-\beta)\inf_u \left(f(\frac{u}{1-\beta}) + \mathfrak{q}(\frac{u}{1-\beta}) + \frac{2}{\beta(1-\beta)}\mathfrak{q}(x-u) - \frac{1}{1-\beta}\mathfrak{q}(x)\right). \tag{4.14}$$

Note that

$$
\begin{aligned}
&\mathfrak{q}(\frac{u}{1-\beta}) + \frac{2}{\beta(1-\beta)}\mathfrak{q}(x-u) - \frac{1}{1-\beta}\mathfrak{q}(x) \\
&= \frac{1}{(1-\beta)^2}\frac{\|u\|^2}{2} + \frac{2}{\beta(1-\beta)}\frac{\|x\|^2 - 2\langle x, u\rangle + \|u\|^2}{2} - \frac{1}{1-\beta}\frac{\|x\|^2}{2} \\
&= \frac{2-\beta}{\beta(1-\beta)^2}\frac{\|u\|^2}{2} - \frac{2\langle x, u\rangle}{\beta(1-\beta)} + \frac{2-\beta}{\beta(1-\beta)}\frac{\|x\|^2}{2} \\
&= \frac{2-\beta}{2\beta}\left(\frac{\|u\|^2}{(1-\beta)^2} - 2\langle\frac{2x}{2-\beta}, \frac{u}{1-\beta}\rangle + \|\frac{2x}{2-\beta}\|^2\right) + \left(\frac{2-\beta}{\beta(1-\beta)} - \frac{4}{\beta(2-\beta)}\right)\frac{\|x\|^2}{2} \\
&= \frac{2-\beta}{2\beta}\|\frac{2x}{2-\beta} - \frac{u}{1-\beta}\|^2 + \frac{\beta}{(1-\beta)(2-\beta)}\frac{\|x\|^2}{2}.
\end{aligned}
$$

Substitute this back into (4.14) to get

$$
\begin{aligned}
T_\beta f(x) &= (1-\beta)\inf_u \left(f(\frac{u}{1-\beta}) + \frac{2-\beta}{2\beta}\|\frac{2x}{2-\beta} - \frac{u}{1-\beta}\|^2\right) + \frac{\beta}{2-\beta}\frac{\|x\|^2}{2} \\
&= (1-\beta)\inf_w \left(f(w) + \frac{2-\beta}{\beta}\mathfrak{q}(\frac{2x}{2-\beta} - w)\right) + \frac{\beta}{2-\beta}\mathfrak{q}(x) \\
&= (1-\beta)\left(f \,\Box\, \frac{2-\beta}{\beta}\mathfrak{q}\right)(\frac{2x}{2-\beta}) + \frac{\beta}{2-\beta}\mathfrak{q}(x),
\end{aligned}
$$

which proves the desired equality. ∎

# Chapter 5

# The Kernel Average of Two Functions

## 5.1 Definition

The kernel average for two functions is given by Bauschke and Wang in [3] as a generalization of the proximal average. A natural extension of the definition of the proximal average, the kernel average replaces the use of $\mathfrak{q}$ with any kernel function $g$. Using the same notation as the previous chapter, we assume $f_1, f_2$, and $g$ are functions in $\Gamma_0(\mathcal{H})$, and $\lambda_1, \lambda_2$ are strictly positive real numbers such that $\lambda_1 + \lambda_2 = 1$.

**Definition 5.1.1 (Kernel Average)** *Let* $\mathbf{f} = (f_1, f_2)$ *and* $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, *we define* $P(\mathbf{f}, \boldsymbol{\lambda}, g) : \mathcal{H} \to [-\infty, +\infty]$ *at* $x \in \mathcal{H}$ *by*

$$\begin{aligned} P(\mathbf{f}, \boldsymbol{\lambda}, g)(x) &:= \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} (\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1 \lambda_2 g(y_1 - y_2)) \\ &= \inf_{x = z_1 + z_2} \left( \lambda_1 f_1(\frac{z_1}{\lambda_1}) + \lambda_2 f_2(\frac{z_2}{\lambda_2}) + \lambda_1 \lambda_2 g(\frac{z_1}{\lambda_1} - \frac{z_2}{\lambda_2}) \right). \end{aligned} \tag{5.1}$$

*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 $f_2$.

**Example 5.1.2 (Arithmetic Average)** *Set* $g = \iota_{\{0\}}$*, then*

$$\begin{aligned} P(\mathbf{f}, \boldsymbol{\lambda}, g)(x) &= \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} (\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1 \lambda_2 \iota_0(y_1 - y_2)) \\ &= \inf_{\lambda_1 y_1 + \lambda_2 y_1 = x} (\lambda_1 f_1(y_1) + \lambda_2 f_2(y_1)) \\ &= \lambda_1 f_1(x) + \lambda_2 f_2(x) \end{aligned}$$

*is the arithmetic average.*

**Lemma 5.1.3** *The following equality holds when* $\lambda_1 + \lambda_2 = 1$*.*

$$\lambda_1 \lambda_2 \|y_1 - y_2\|^2 = \lambda_1 \|y_1\|^2 + \lambda_2 \|y_2\|^2 - \|\lambda_1 y_1 + \lambda_2 y_2\|^2.$$

*Proof.* Let $y_1, y_2 \in \mathcal{H}$ then by Lemma 4.2.3,

$$\begin{aligned} \lambda_1 \|y_1\|^2 + \lambda_2 \|y_2\|^2 - \|\lambda_1 y_1 + \lambda_2 y_2\|^2 &= 2\left(\frac{1}{4} \sum_{i=1}^{2} \sum_{j=1}^{2} \lambda_1 \lambda_j \|y_i - y_j\|^2\right) \\ &= 2\left(\frac{1}{4} \lambda_1 \lambda_2 \|y_1 - y_2\|^2 + \frac{1}{4} \lambda_2 \lambda_1 \|y_2 - y_1\|^2\right) \\ &= \lambda_1 \lambda_2 \|y_1 - y_2\|^2 \end{aligned}$$

■

**Example 5.1.4 (Proximal Average)** *If* $g = \mathfrak{q}$*, then*

$$P(\mathbf{f}, \boldsymbol{\lambda}, g)(x) = \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} \left(\lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \lambda_1 \lambda_2 \frac{1}{2} \|y_1 - y_2\|^2\right).$$

*Applying Lemma 5.1.3*

$$P(\mathbf{f}, \boldsymbol{\lambda}, g) = \inf_{\lambda_1 y_1 + \lambda_2 y_2 = x} \left( \lambda_1 f_1(y_1) + \lambda_2 f_2(y_2) + \frac{1}{2}\lambda_1 \|y_1\|^2 + \frac{1}{2}\lambda_2 \|y_2\|^2 - \frac{1}{2}\|x\|^2 \right),$$

*which is the proximal average with $n = 2$.*

**Example 5.1.5** *Let $f_1 = \iota_{\{a\}}$ and $f_2 = \iota_{\{b\}}$, with $a, b \in \mathbb{R}$. Then*

$$P(\mathbf{f}, \boldsymbol{\lambda}, g) = \begin{cases} \lambda_1 \lambda_2 g(a - b) & \textit{if } x = \lambda_1 a + \lambda_2 b \\ +\infty & \textit{otherwise.} \end{cases}$$

## 5.2 Properties

**Fact 5.2.1 (Fenchel Conjugate)** *[3, Theorem 2.2] Let $f_1, f_2, g \in \Gamma_0(\mathcal{H})$. For every $x^* \in \mathcal{H}$,*

$$(P(\mathbf{f}, \boldsymbol{\lambda}, g))^*(x^*) = (\operatorname{cl} \varphi)(\lambda_1 x^*, \lambda_2 x^*) \tag{5.2}$$

*where*

$$\varphi(u, v) = \inf_{\lambda_1 z_1 + \lambda_2 z_2 = u + v} \left( \lambda_1 f_1^*(z_1) + \lambda_2 f_2^*(z_2) + \frac{1}{2}\lambda_1 \lambda_2 (g^*(\frac{u}{\lambda_1 \lambda_2} - \frac{z_1}{\lambda_2}) + g^*(\frac{-v}{\lambda_1 \lambda_2} - \frac{z_2}{\lambda_1})) \right).$$

*Furthermore, if $(\operatorname{ri} \operatorname{dom} f_1 - \operatorname{ri} \operatorname{dom} f_2) \bigcap \operatorname{ri} \operatorname{dom} g \neq \varnothing$ then the closure operation in (5.2) can be omitted and we get that*

$$(P(\mathbf{f}, \boldsymbol{\lambda}, g))^*(x^*) = P(\mathbf{f}^*, \boldsymbol{\lambda}, (g^*)^\vee) \tag{5.3}$$

*where for a given function $g \in \Gamma_0(\mathcal{H})$, let $g^\vee(x) = g(-x)$, and the infimum in Definition 5.1.1 is attained, that is*

$$P(\mathbf{f}^*, \boldsymbol{\lambda}, (g^*)^\vee)(x^*) = \min_{x^* = \lambda_1 z_1 + \lambda_2 z_2} \left(\lambda_1 f_1^*(z_1) + \lambda_2 f_2^*(z_2) + \lambda_1 \lambda_2 g^*(z_2 - z_1)\right). \tag{5.4}$$

**Corollary 5.2.2** *Let $f_1, f_2, g \in \Gamma_0(\mathcal{H})$, and assume that both $g$ and $g^*$ have full domain. Then both $P(\mathbf{f}, \boldsymbol{\lambda}, g)$ and $P(\mathbf{f}^*, \boldsymbol{\lambda}, (g^*)^\vee)$ are in $\Gamma_0(\mathcal{H})$ and*

$$(P(\mathbf{f}, \boldsymbol{\lambda}, g))^* = P(\mathbf{f}^*, \boldsymbol{\lambda}, (g^*)^\vee). \tag{5.5}$$

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

$$(P(\mathbf{f}, \boldsymbol{\lambda}, \frac{1}{p}\|\cdot\|^p))^* = P(\mathbf{f}^*, \boldsymbol{\lambda}, \frac{1}{q}\|\cdot\|^q) \tag{5.6}$$

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

Can the definition of the kernel average be generalized for $n$ functions? What is the reformulation of Definition 5.1.1? This will be addressed in the next chapter.

# Chapter 6

# The Kernel Average of $n$ Functions

## 6.1 Motivation

In this chapter, we look at another reformulation of the proximal average and see how that reformulation can be used to extend the definition of the kernel average to $n$ functions. Similar to the previous chapters, we assume $f_1, \cdots, f_n$ and $g$ are functions in $\Gamma_0(\mathcal{H})$, and $\lambda_1, \cdots, \lambda_n$ are strictly positive real numbers such that $\sum_{i=1}^{n} \lambda_i = 1$. First we will prove a new reformulation of the proximal average.

**Theorem 6.1.1** *Let* $\mathbf{f} = (f_1, \cdots f_n)$ *with* $f_1, \cdots, f_n \in \Gamma_0(\mathcal{H})$*, and* $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_n)$ *with* $\lambda_1, \cdots, \lambda_n \geq 0$ *and* $\sum_{i=1}^{n} \lambda_i = 1$*. Define*

$$f := \mathcal{P}(\mathbf{f}, \boldsymbol{\lambda}) = \big(\lambda_1(f_1 + \mathfrak{q})^* + \cdots + \lambda_n(f_n + \mathfrak{q})^*\big)^* - \mathfrak{q}.$$

*Then for every* $x \in \mathcal{H}$*,*

$$f(x) = \inf_{\lambda_1 y_1 + \cdots \lambda_n y_n = x} \lambda_1 f_1(y_1) + \cdots + \lambda_n f_n(y_n) + \frac{1}{2} \sum_{i<j} \lambda_i \lambda_j \|y_i - y_j\|^2.$$

*Proof.* By Proposition 4.1.2(i) and Lemma 4.2.3

$$\begin{aligned}
f(x) &= \inf_{\sum_{i=1}^{n} \lambda_i y_i = x} \Big( \sum_{i=1}^{n} \lambda_i f_i(y_i) + \sum_{i=1}^{n} \lambda_i \mathfrak{q}(y_i) - \mathfrak{q}(x) \Big) \\
&= \inf_{\sum_{i=1}^{n} \lambda_i y_i = x} \Big( \sum_{i=1}^{n} \lambda_i f_i(y_i) + \sum_{i=1}^{n} \lambda_i \mathfrak{q}(y_i) - \mathfrak{q}(\sum_{i=1}^{n} \lambda_i y_i) \Big) \\
&= \inf_{\sum \lambda_i y_i = x} \Big( \sum \lambda_i f_i(y_i) + \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \|y_i - y_j\|^2 \Big) \\
&= \inf_{\sum \lambda_i y_i = x} \Big( \sum \lambda_i f_i(y_i) + \sum_{i<j} \lambda_i \lambda_j \frac{1}{2} \|y_i - y_j\|^2 \Big).
\end{aligned}$$

■

This reformulation of the proximal average suggests a generalization where $\frac{1}{2}\|\cdot\|^2$ is replaced by a function $g$. We'll call this generalization the kernel average of $n$ functions, defined by

$$Q_g(\mathbf{f}, \boldsymbol{\lambda})(x) := \inf_{\sum \lambda_i y_i = x} \Big( \sum \lambda_i f_i(y_i) + \sum_{i<j} \lambda_i \lambda_j g(y_i - y_j) \Big). \tag{6.1}$$

This definition is the same as the kernel average when $n = 2$, but extends the kernel average definition by allowing more than two functions. We'll now explore the kernel average for $n$ functions in a bit more detail, and from here on when we refer to the kernel average we mean $Q_g(\mathbf{f}, \boldsymbol{\lambda})$.

## 6.2 The Kernel Average Conjugate

We will now consider the kernel average as

$$Q_g(\mathbf{f}, \boldsymbol{\lambda})(x) = \inf_{Ay=x} \{h(y) := \tilde{f}(y) + \tilde{g}(y)\} = (Ah)(x),$$

where

$$\begin{aligned} A(x_1, \cdots, x_n) &= \sum_{i=1}^{n} \lambda_i x_i Ah = \inf\{Ay = x\}\{h(x)\}, \\ \tilde{f}(y) &= \sum \lambda_i f_i(y_i), \text{ and} \\ \tilde{g}(y) &= \sum_{i<j} \lambda_i \lambda_j g(y_i - y_j). \end{aligned}$$

In light of Proposition 3.2.35, we get

$$\begin{aligned} Q_g^*(x^*) &= (h^* \circ A^*)(x^*) \\ &= \Big( (\tilde{f} + \tilde{g})^* \circ A^* \Big)(x^*), \end{aligned}$$

so we can see that to get $Q_g^*$ we need to compute $h^* = (\tilde{f} + \tilde{g})^*$, which by Fact 3.2.42 we have $h^* = \tilde{f}^* \,\square\, \tilde{g}^*$. It is quite simple to compute $\tilde{f}^*$, and this was done in Lemma 3.2.44, but $\tilde{g}^*$ is more challenging.

To consider $\tilde{g}^* = \big(\sum_{i<j} \lambda_i \lambda_j g(y_i - y_j)\big)^*$, we'll first begin with the case which gives us the proximal average, where $g = \mathsf{q}$, with the hope that this will allow us to find a general formula for any $g$.

**Proposition 6.2.1** *Let $x = (x_1, \cdots, x_n)$ and $g_1(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2}\|x_i - x_j\|^2 =$*

$2\sum_{i<j} \frac{1}{2}\|x_i - x_j\|^2$ *then*

$$g_1^*(x^*) = \begin{cases} \frac{1}{8n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|x_i^* - x_j^*\|^2, & \textit{if } x_1^* + \cdots x_n^* = 0; \\ +\infty, & \textit{otherwise.} \end{cases}$$

*Proof.* Let $\lambda_i = \frac{1}{n}$ for $i = 1 \cdots n$, then $g_1$ can be rewritten as

$$g_1(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2}\|x_i - x_j\|^2 = (2n^2) \cdot \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} \|x_i - x_j\|^2.$$

Using Lemma 4.2.3,

$$\begin{aligned} g_1(x) &= 2n^2 \left( \sum_{i=1}^{n} \lambda_i \mathfrak{q}(x_i) - \mathfrak{q}(\sum_{i=1}^{n} \lambda_i x_i) \right) \\ &= 2n^2 \cdot g(x), \end{aligned}$$

where $g(x)$ is as defined in Lemma 4.2.4. Then by Proposition 3.2.41(i) and Lemma 4.2.4

$$\begin{aligned} g_1^*(x^*) &= 2n^2 g^*(\frac{x^*}{2n^2}) \\ &= \begin{cases} 2n^2 \left( \sum_{i=1}^{n} \lambda_i \star \mathfrak{q}(\frac{x_i^*}{2n^2}) \right), & \text{if } \frac{x_1^*}{2n^2} + \cdots + \frac{x_n^*}{2n^2} = 0; \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned}$$

Now, $\lambda_i \star \mathfrak{q}(\frac{x_i^*}{2n^2}) = \frac{1}{n}\mathfrak{q}(\frac{nx_i^*}{2n^2}) = \frac{1}{4n^3}\mathfrak{q}(x_i^*)$ so

$$\begin{aligned} 2n^2\Bigg(\sum_{i=1}^n \lambda_i \star \mathfrak{q}(\frac{x_i^*}{2n^2})\Bigg) &= \frac{1}{2n}\sum_{i=1}^n \mathfrak{q}(x_i^*) \\ &= \frac{1}{2}\Bigg(\sum_{i=1}^n \frac{1}{n}\mathfrak{q}(x_i^*)\Bigg) = \frac{1}{2}\Bigg(\sum_{i=1}^n \lambda_i \mathfrak{q}(x_i^*)\Bigg). \end{aligned}$$

Applying Lemma 4.2.3

$$2n^2\Bigg(\sum_{i=1}^n \lambda_i \star \mathfrak{q}(\frac{x_i^*}{2n^2})\Bigg) = \frac{1}{2}\Bigg(\mathfrak{q}(\sum_{i=1}^n \lambda_i x_i^*) + \frac{1}{4}\sum_{i=1}^n\sum_{j=1}^n \lambda_i\lambda_j \|x_i^* - x_j^*\|^2\Bigg).$$

Since $\frac{x_1^*}{2n^2} + \cdots + \frac{x_n^*}{2n^2} = 0 \Leftrightarrow \frac{1}{2n}\Bigg(\lambda_1 x_1^* + \cdots + \lambda_n x_n^*\Bigg) = 0 \Leftrightarrow \sum\limits_{i=1}^n \lambda_i x_i^* = 0$, we get

$$\begin{aligned} 2n^2\Bigg(\sum_{i=1}^n \lambda_i \star \mathfrak{q}(\frac{x_i^*}{2n^2})\Bigg) &= \frac{1}{2}\Bigg(\frac{1}{4}\sum_{i=1}^n\sum_{j=1}^n \lambda_i\lambda_j \|x_i^* - x_j^*\|^2\Bigg) \\ &= \frac{1}{8n^2}\sum_{i=1}^n\sum_{j=1}^n \|x_i^* - x_j^*\|^2. \end{aligned}$$

Altogether,

$$g_1^*(x^*) = \begin{cases} \frac{1}{8n^2}\sum\limits_{i=1}^n\sum\limits_{j=1}^n \|x_i^* - x_j^*\|^2, & \text{if } x_1^* + \cdots + x_n^* = 0; \\ \\ +\infty, & \text{otherwise.} \end{cases}$$

■

Proposition 6.2.1 can also be proven directly using critical points. *Proof.*

By definition,

$$g_1^*(x^*) = \sup_{x_1,\cdots,x_n} \left( \langle x_1^*, x_1\rangle + \cdots + \langle x_n^*, x_n\rangle - \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{2}\|x_i - x_j\|^2 \right).$$

Let $\hat{g}_1(x) = \langle x_1^*, x_1\rangle + \cdots + \langle x_n^*, x_n\rangle - \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} \frac{1}{2}\|x_i - x_j\|^2$. Since $\frac{\partial g_1}{\partial x_i} = 2\sum\limits_{j=1}^{n}(x_i - x_j)$, we get

$$\nabla \hat{g}_1(x) = (x_1^* - 2\sum_{l=1}^{n}(x_1 - x_l), \cdots, x_n^* - 2\sum_{l=1}^{n}(x_n - x_l)).$$

Setting this equal to zero to solve for the critical points, we see

$$(x_1^*, \cdots, x_n^*) = (2\sum_{l=1}^{n}(x_1 - x_l), \cdots, 2\sum_{l=1}^{n}(x_n - x_l)),$$

and therefore that $x_1^* + \cdots + x_n^* = 0$. And we also get

$$(x_i^* - x_j^*) = 2\sum_{l=1}^{n}(x_i - x_l) - (x_j - x_l) = 2\sum_{l=1}^{n}(x_i - x_j) = 2n(x_i - x_j),$$

which implies that $(x_i - x_j) = \frac{1}{2n}(x_i^* - x_j^*)$ for all $i, j$ with $1 \le i, j \le n$. Since $\hat{g}_1$ is a sum of linear and concave functions, then $\hat{g}_1$ is concave. Thus, the critical point is a maximum and we can substitute into $g_1^*$ to find the supremum.

Doing this, we find

$$\begin{aligned}
g_1^*(x^*) &= \langle x_1^*, x_1\rangle + \cdots + \langle x_{n-1}^*, x_{n-1}\rangle + \langle -x_1^* - \cdots - x_{n-1}^*, x_n\rangle \\
&\qquad - \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{2}\Big(\frac{1}{2n}\Big)^2 \|x_i^* - x_j^*\|^2 \\
&= \langle x_1^*, x_1 - x_n\rangle + \cdots + \langle x_{n-1}^*, x_{n-1} - x_n\rangle - \frac{1}{8n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i^* - x_j^*\|^2 \\
&= \langle x_1^*, \frac{1}{2n}(x_1^* - x_n^*)\rangle + \cdots + \langle x_{n-1}^*, \frac{1}{2n}(x_{n-1}^* - x_n^*)\rangle - \frac{1}{8n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i^* - x_j^*\|^2 \\
&= \frac{1}{2n}(\langle x_1^*, x_1^*\rangle + \cdots + \langle x_{n-1}^*, x_{n-1}^*\rangle) - \frac{1}{2n}\langle x_1^*, x_n^*\rangle - \cdots - \frac{1}{2n}\langle x_{n-1}^*, x_n^*\rangle \\
&\qquad - \frac{1}{8n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i^* - x_j^*\|^2 \\
&= \frac{1}{2n}(\|x_1^*\|^2 + \cdots + \|x_{n-1}^*\|^2 + \frac{1}{2n}\langle -x_1^* - \cdots - x_{n-1}^*, x_n^*\rangle - \frac{1}{8n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i^* - x_j^*\|^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n}\|x_i^*\|^2 - \frac{1}{8n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i^* - x_j^*\|^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n}\|x_i^*\|^2 - \frac{1}{2}\Big\|\frac{x_1^* + \cdots + x_n^*}{n}\Big\|^2 - \frac{1}{8n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i^* - x_j^*\|^2.
\end{aligned}$$

Then using (4.2.3), we get that

$$\begin{aligned}
g_1^*(x^*) &= \frac{1}{4n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i^* - x_j^*\|^2 - \frac{1}{8n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i^* - x_j^*\|^2 \\
&= \frac{1}{8n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i^* - x_j^*\|^2
\end{aligned}$$

when $x_1^* + \cdots + x_n^* = 0$. If $x_1^* + \cdots + x_n^* \neq 0$ then set $x_1 = \cdots = x_n = x$ and get that

$$g_1^*(x^*) \geq \sup_x \left( \langle \sum_{i=1}^n x_i^*, x \rangle \right) = +\infty.$$

■

This formula can be written in several equivalent forms using the fact that $\sum_{i=1}^n x_i^* = 0$. Using this, we can see that

$$\begin{aligned}
\sum_{i=1}^n \sum_{j=1}^n \|x_i^* - x_j^*\|^2 &= \sum_{i=1}^n \sum_{j=1}^n (\|x_i^*\|^2 + \|x_j^*\|^2 - 2\langle x_i^*, x_j^* \rangle) \\
&= \sum_{i=1}^n \left( \sum_{j=1}^n (\|x_i^*\|^2 + \|x_j^*\|^2) - 2\langle x_i^*, \sum_{j=1}^n x_j^* \rangle \right) \\
&= \sum_{i=1}^n \left( \sum_{j=1}^n (\|x_i^*\|^2 + \|x_j^*\|^2) - 0 \right) \\
&= \sum_{i=1}^n \left( \sum_{j=1}^n (\|x_i^*\|^2 + \|x_j^*\|^2) + 2\langle x_i^*, \sum_{j=1}^n x_j^* \rangle \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n \|x_i^* + x_j^*\|^2.
\end{aligned}$$

And from the first proof of Proposition 6.2.1 we also see that

$$g_1^*(x^*) = \frac{1}{8n^2} \sum_{i=1}^n \sum_{j=1}^n \|x_i^* - x_j^*\|^2 = \frac{1}{2n} \sum_{i=1}^n \frac{1}{2} \|x_i^*\|^2.$$

So that the following three formulations for the conjugate of $g_1$ are equivalent:

(1) $g_1^*(x^*) = \frac{1}{8n^2} \sum_{i=1}^n \sum_{j=1}^n \|x_i^* - x_j^*\|^2 = \sum_{i=1}^n \sum_{j=1}^n \frac{1}{2} \|\frac{x_i^* - x_j^*}{2n}\|^2$

(2) $g_1^*(x^*) = \frac{1}{8n^2} \sum_{i=1}^n \sum_{j=1}^n \|x_i^* + x_j^*\|^2 = \sum_{i=1}^n \sum_{j=1}^n \frac{1}{2} \|\frac{x_i^* + x_j^*}{2n}\|^2$

(3) $g_1^*(x^*) = \frac{1}{2n} \sum_{i=1}^{n} \frac{1}{2}\|x_i\|^2$

when $x_1^* + \cdots + x_n^* = 0$ and $g_1^*(x^*) = +\infty$ otherwise.

The next step we wish to consider is the case where $g = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{p}\|x_i - x_j\|^p$ for both general $p > 1$ and general $n > 1$. The conjugate is known for general $p > 1$ and $n = 2$, so in the next section we begin looking at general $p$ with $n = 3$.

**Example 6.2.2 (General $p$, $n = 2$)** *Let $f(x_1, x_2) = \frac{1}{p}\|x_1 - x_2\|^p$ with $p > 1$. Then,*

$$\begin{aligned} f^*(y_1, y_2) &= \sup_{x_1,x_2} \left( \langle x_1, y_1 \rangle + \langle x_2, y_2 \rangle - \frac{1}{p}\|x_1 - x_2\|^p \right) \\ &= \sup_{x_1,x_2} \left( \langle x_1 - x_2, y_1 \rangle + \langle x_2, y_1 + y_2 \rangle - \frac{1}{p}\|x_1 - x_2\|^p \right). \end{aligned}$$

*And using Example 3.2.30, with $\frac{1}{p} + \frac{1}{q} = 1$, we get*

$$f^*(y_1, y_2) = \begin{cases} \frac{1}{q}\|y_1\|^q, & \text{if } y_1 + y_2 = 0; \\ +\infty, & \text{otherwise.} \end{cases}$$

## 6.3 P-norm Kernel Conjugate for General $p$ Case when $n = 3$

We now wish to consider the case where $g_2 = \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{1}{p}\|x_i - x_j\|^p$, that is

$$g_2(x_1, x_2, x_3) = \frac{1}{p}\|x_1 - x_2\|^p + \frac{1}{p}\|x_2 - x_3\|^p + \frac{1}{p}\|x_3 - x_1\|^p.$$

Then $g_2^*(y_1, y_2, y_3)$ is equal to

$$\sup_{x_1,x_2,x_3} \{x_1y_1 + x_2y_2 + x_3y_3 - \frac{1}{p}\|x_1 - x_2\|^p - \frac{1}{p}\|x_2 - x_3\|^p - \frac{1}{p}\|x_3 - x_1\|^p\}$$
$$= \sup_{x_1,x_2} \{x_1y_1 + x_2y_2 - \frac{1}{p}\|x_1 - x_2\|^p + \sup_{x_3}(x_3y_3 - \frac{1}{p}\|x_2 - x_3\|^p - \frac{1}{p}\|x_3 - x_1\|^p)\} \tag{6.2}$$

Recognizing that $\sup_{x_3}\{x_3y_3 - \frac{1}{p}\|x_2 - x_3\|^p - \frac{1}{p}\|x_3 - x_1\|^p\} = (\frac{1}{p}\|x_1 - \cdot\|^p + \frac{1}{p}\|x_2 - \cdot\|^p)^*(y_3)$, then applying Fact 3.2.42(i)

$$\sup_{x_3}\{x_3y_3 - \frac{1}{p}\|x_2 - x_3\|^p - \frac{1}{p}\|x_3 - x_1\|^p\} = ((\frac{1}{p}\| \cdot - x_1\|^p)^* \Box (\frac{1}{p}\| \cdot - x_2\|^p)^*)(y_3) \tag{6.3}$$

Using Proposition 3.2.28 and Example 3.2.30 we get

$$(\frac{1}{p}\|x - z\|^p)^*(y) = \langle y, z\rangle + \frac{1}{q}\|y\|^q. \tag{6.4}$$

Combining (6.4) and (6.3)

$$((\frac{1}{p}\| \cdot - x_1\|^p)^* \Box (\frac{1}{p}\| \cdot - x_2\|^p)^*)(y_3) = (\frac{1}{q}\| \cdot \|^q + \langle x_1, \cdot\rangle) \Box (\frac{1}{q}\| \cdot \|^q + \langle x_2, \cdot\rangle)$$
$$= \inf_{u+v=y_3} \{\frac{1}{q}\|u\|^q + \frac{1}{q}\|v\|^q + \langle x_1, u\rangle + \langle x_2, v\rangle\}. \tag{6.5}$$

Substituting (6.5) back into (6.2) and setting $v = y_3 - u$ yields

$$g_2^*(y_1, y_2, y_3) = \sup_{x_1,x_2} \inf_u [\langle x_1, y_1\rangle + \langle x_2, y_2\rangle - \frac{1}{p}\|x_1 - x_2\|^p + \frac{1}{q}\|u\|^q + \frac{1}{q}\|u - y_3\|^q$$
$$+ \langle x_1, u\rangle + \langle x_2, y_3 - u\rangle].$$

Looking at the above equation we can see that

$$F((x_1,x_2),u) = \langle x_1,y_1\rangle+\langle x_2,y_2\rangle-\frac{1}{p}\|x_1-x_2\|^p+\frac{1}{q}\|u\|^q+\frac{1}{q}\|u-y_3\|^q+\langle x_1,u\rangle+\langle x_2,y_3-u\rangle$$

is concave-convex since $F((x_1,x_2),\cdot)$ is a sum of linear and convex functions and $F(\cdot,u)$ is a sum of linear and concave functions.

Fix $x = (x_1,x_2) \in \mathcal{H}\times\mathcal{H}$, and let $F((x_1,x_2),u) = F_x(u)$. Using Fact 3.4.3 we show that a saddle value exists by showing that $0^+(\operatorname{epi} F_x) = \{(0,\lambda) : \lambda \geq 0\}$.

By definition, $0^+(\operatorname{epi} F_x) = \operatorname{epi} F_x^\infty$, so

$$\begin{aligned} (u,\lambda) \in 0^+(\operatorname{epi} F_x) &\Leftrightarrow (u,\lambda) \in \operatorname{epi} F_x^\infty \\ &\Leftrightarrow \lambda \geq F_x^\infty(u). \end{aligned}$$

Using Proposition 3.2.19 to compute $F_x^\infty(u)$ yields

$$\begin{aligned} F_x^\infty(u) &= \lim_{\lambda\to\infty}\left(\frac{F_x(\lambda u)-F_x(0)}{\lambda}\right) \\ &= \lim_{\lambda\to\infty}\left(\frac{\frac{1}{q}\|\lambda u\|^q+\frac{1}{q}\|\lambda u-y_3\|^q+\langle x_1,\lambda u\rangle+\langle x_2,y_3-\lambda u\rangle-\frac{1}{q}\|y_3\|^q-\langle x_2,y_3\rangle}{\lambda}\right) \\ &= \lim_{\lambda\to\infty}\left(\frac{\lambda^q\frac{1}{q}\|u\|^q+\lambda^q\frac{1}{q}\|u-\frac{y_3}{\lambda}\|^q+\lambda\langle x_1,u\rangle+\lambda\langle x_2,\frac{y_3}{\lambda}-u\rangle}{\lambda}\right) \\ &= \begin{cases} +\infty, & \text{if } u \neq 0; \\ 0, & \text{if } u = 0. \end{cases} \end{aligned}$$

Therefore $(u,\lambda) \in \operatorname{epi} F_x^\infty \Leftrightarrow u = 0$. Since there is no common nonzero direction of recession for $F_x$, we can use Fact 3.4.3 to swap the positions of the infimum and supremum, and combining the appropriate inner product terms produces

$$g_2^*(y_1, y_2, y_3) = \inf_u \sup_{x_1,x_2} \left( \langle x_1, y_1+u\rangle + \langle x_2, y_2+y_3-u\rangle - \frac{1}{p}\|x_1-x_2\|^p + \frac{1}{q}\|u\|^q + \frac{1}{q}\|u-y_3\|^q \right). \tag{6.6}$$

Considering the inner supremum first, we will fix $x_1$ and let $x_2 - x_1 = z$. Then (6.6) becomes

$$\inf_u \left( \frac{1}{q}\|u\|^q + \frac{1}{q}\|u - y_3\|^q + \sup_{x_1}\langle x_1, y_1 + u\rangle + \sup_z \langle x_1 + z, y_2 + y_3 - u\rangle - \frac{1}{p}\|z\|^p \right)$$
$$= \inf_u \left( \frac{1}{q}\|u\|^q + \frac{1}{q}\|u - y_3\|^q + \sup_{x_1}\langle x_1, y_1 + y_2 + y_3\rangle + \sup_z \langle z, y_2 + y_3 - u\rangle - \frac{1}{p}\|z\|^p \right).$$

Here we can see that the supremum on the right is the definition of $(\frac{1}{p}\|\cdot\|^p)^*$ evaluated at $y_2+y_3-u$. Using Example 3.2.30 and the fact that $y_1+y_2+y_3 = 0$ we then get

$$g_2^*(y_1, y_2, y_3) = \inf_u \left( \frac{1}{q}\|u\|^q + \frac{1}{q}\|u - y_3\|^q + \frac{1}{q}\|y_2 + y_3 - u\|^q \right)$$

where $y_1 + y_2 + y_3 = 0$ and $\frac{1}{p} + \frac{1}{q} = 1$.

Since $g_2$ is symmetric under permutation of its variables, we interchange $y_1$ and $y_3$ so that the previous description of $g_2^*$ turns into the more symmetric form

$$g_2^*(y_1, y_2, y_3) = \inf_x \left( \frac{1}{q}\|x - y_1\|^q + \frac{1}{q}\|x - (y_1 + y_2)\|^q + \frac{1}{q}\|x - (y_1 + y_2 + y_3)\|^q \right)$$

where $y_1 + y_2 + y_3 = 0$ and $\frac{1}{p} + \frac{1}{q} = 1$.

In $\mathbb{R}$, the problem becomes

$$\min_{x \in \mathbb{R}} \left( \frac{1}{q}|x - y_1|^q + \frac{1}{q}|x - (y_1 + y_2)|^q + \frac{1}{q}|x|^q \right) \tag{6.7}$$

with $\frac{1}{p} + \frac{1}{q} = 1$ and $y_1 + y_2 + y_3 = 0$. We will define

$$h(x) := \frac{1}{q}|x - y_1|^q + \frac{1}{q}|x - (y_1 + y_2)|^q + \frac{1}{q}|x|^q, \tag{6.8}$$

so that we are solving $\min\limits_{x \in \mathbb{R}} h(x)$. Since the case with $p = 2$ and $q = 2$ was already solved, we consider this problem with the next simplest case, $q = 3$ which corresponds to $p = \frac{3}{2}$.

## 6.4 P-norm Kernel Conjugate when $p = \frac{3}{2}$, $q = 3$, and $n = 3$

Considering (6.8) with $q = 3$ gives the problem

$$\min_{x \in \mathbb{R}} h(x) \tag{6.9}$$

where

$$h(x) = \frac{1}{3}|x - y_1|^3 + \frac{1}{3}|x - (y_1 + y_2)|^3 + \frac{1}{3}|x|^3, \tag{6.10}$$

and $y_1, y_2$ are constants.

In order to find the optimal value of (6.9) we need to consider eight cases, which cover all possible values of the three absolute values in (6.10). The eight cases are as follows:

(1) $x \geq y_1$, $x \geq y_1 + y_2$, and $x \geq 0$

(2) $x \leq y_1$, $x \geq y_1 + y_2$, and $x \geq 0$

(3) $x \geq y_1$, $x \leq y_1 + y_2$, and $x \geq 0$

(4) $x \geq y_1$, $x \geq y_1 + y_2$, and $x \leq 0$

(5) $x \leq y_1$, $x \leq y_1 + y_2$, and $x \geq 0$

(6) $x \leq y_1$, $x \geq y_1 + y_2$, and $x \leq 0$

(7) $x \geq y_1$, $x \leq y_1 + y_2$, and $x \leq 0$

(8) $x \leq y_1$, $x \leq y_1 + y_2$, and $x \leq 0$

We now consider each case in depth, and set $y = (y_1, y_2)$.

### 6.4.1 Case 1: $x \geq y_1$, $x \geq y_1 + y_2$, $x \geq 0$

Using the constraints of this case, we define the function to be minimized as

$$h_{1,y}(x) = \frac{1}{3}(x - y_1)^3 + \frac{1}{3}(x - (y_1 + y_2))^3 + \frac{1}{3}x^3,$$

and we minimize $h_{1,y}$ over the region where $\max\{y_1, y_1 + y_2, 0\} \leq x$. Because $h_{1,y}(x)$ is convex with respect to $x$, see Example 3.2.31, then by Fact 3.2.23 any critical point will be the minimizer. Differentiating to solve for critical points yields

$$\frac{\partial h_{1,y}}{\partial x} = (x - y_1)^2 + (x - (y_1 + y_2))^2 + x^2.$$

Setting $\frac{\partial h_{1,y}}{\partial x} = 0$ and solving for $x$, we get the critical points

$$\begin{aligned}
x_1 &= \frac{1}{3}y_2 + \frac{2}{3}y_1 + \frac{1}{3}\sqrt{-2y_2^2 - 2y_1y_2 - 2y_1^2} \\
x_2 &= \frac{1}{3}y_2 + \frac{2}{3}y_1 - \frac{1}{3}\sqrt{-2y_2^2 - 2y_1y_2 - 2y_1^2}.
\end{aligned}$$

To see if the critical points are real, the value of $-2y_2^2 - 2y_1y_2 - 2y_1^2$ must be examined and we see that

$$\begin{aligned}
-2y_2^2 - 2y_1y_2 - 2y_1^2 &= -(y_1^2 + y_2^2) - (y_1^2 + 2y_1y_2 + y_2^2) \\
&= -(y_1^2 + y_2^2) - (y_1 + y_2)^2 \leq 0.
\end{aligned}$$

Since $-2y_2^2 - 2y_1y_2 - 2y_1^2 \leq 0$ for all values of $y_1$ and $y_2$ then there are no real critical points of $h_{1,y}$, so we next check the boundary points, $x \geq \max\{0, y_1, y_1 + y_2\}$.

If $\max\{0, y_1, y_1 + y_2\} = y_1 + y_2$, i.e. when $y_1 \geq 0$ and $y_2 \geq 0$, or when $y_2 \geq 0$ and $-y_2 \leq y_1 \leq 0$, then the minimum value of $h^1_{y_1,y_2}(x)$ is

$$h_{1,y}(y_1 + y_2) = \frac{1}{3}y_2^3 + \frac{1}{3}(y_1 + y_2)^3.$$

If $\max\{0, y_1, y_1 + y_2\} = y_1$, or rather when $y_1 \geq 0$ and $y_2 \leq 0$, then we get a minimum value at

$$h_{1,y}(y_1) = \frac{1}{3}|y_2|^3 + \frac{1}{3}y_1^3.$$

If $\max\{0, y_1, y_1 + y_2\} = 0$, i.e. if $y_1 \leq 0$ and $y_2 \leq 0$, or $y_2 \geq 0$ and $y_2 \leq -y_1$, then the minimum value is

$$h_{1,y}(0) = \frac{1}{3}|y_1|^3 + \frac{1}{3}|y_1 + y_2|^3.$$

This case is summarized graphically in Figure 6.1.

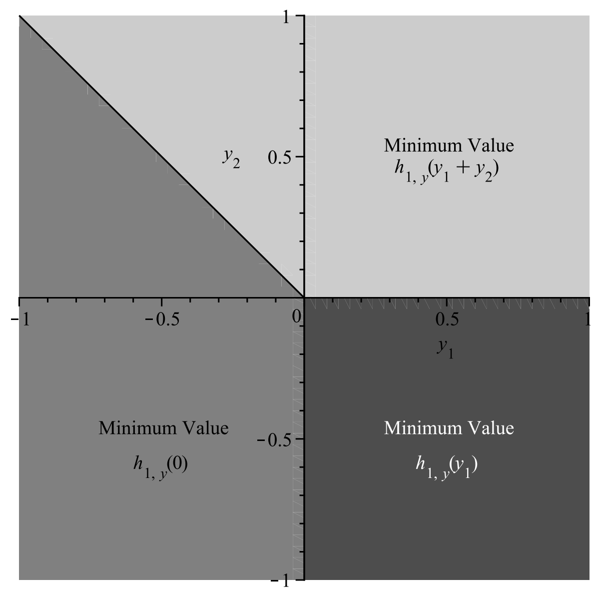



Figure 6.1: Case 1 summary

### 6.4.2 Case 2: $x \leq y_1$, $x \geq y_1 + y_2$, $x \geq 0$

The conditions for this case give the following function,

$$h_{2,y}(x) = \frac{1}{3}(y_1 - x)^3 + \frac{1}{3}(x - (y_1 + y_2))^3 + \frac{1}{3}x^3,$$

for minimization on the region where $\max\{0, y_1 + y_2\} \leq x \leq y_1$. Again, differentiating to find the critical points gives us

$$\frac{\partial h_{2,y}}{\partial x} = -(y_1 - x)^2 + (x - (y_1 + y_2))^2 + x^2.$$

And solving $\frac{\partial h_{2,y}}{\partial x} = 0$ yields the two critical points

$$x_1 = y_2 + \sqrt{-2y_1y_2}$$
$$x_2 = y_2 - \sqrt{-2y_1y_2}.$$

Considering the constraints, we see that $y_1 \geq x \geq 0$ and $y_1 + y_2 \leq x \leq y_1$, so that $y_1 \geq 0$ and $y_2 \leq 0$ is the region of interest. This makes $-2y_1y_2 \geq 0$ so that the critical points are real. We need $x \geq 0$, but $x_2 \leq 0$ for all $y_1, y_2$ in this region, and if $x_2 = 0$ then $y_2 = y_1 = 0$ and $x_1 = x_2 = 0$. Hence, it suffices to consider only $x_1$.

Next, we check that $x_1$ satisfies the three conditions of this case: $x \geq 0$, $x \leq y_1$, and $x \geq y_1 + y_2$. For $x_1 = y_2 + \sqrt{-2y_1y_2} \geq 0$, it is required that

$$-y_2 = |y_2| \leq \sqrt{-2y_1y_2} \Leftrightarrow y_2^2 \leq -2y_1y_2 \Leftrightarrow |y_2| \leq 2y_1 \Leftrightarrow \frac{1}{2}|y_2| \leq y_1.$$

For $x_1 \leq y_1$, this is equivalent to

$$y_2 + \sqrt{-2y_1y_2} \leq y_1 \Leftrightarrow -2y_1y_2 \leq (y_1 - y_2)^2 \Leftrightarrow 0 \leq y_1^2 + y_2^2.$$

Therefore this condition is always true. For $x_1 \geq y_1 + y_2$, we get

$$y_2 + \sqrt{-2y_1y_2} \geq y_1 + y_2 \Leftrightarrow \sqrt{-2y_1y_2} \geq y_1 \Leftrightarrow 2|y_2| \geq y_1.$$

Since $h_{2,y}$ is convex with respect to $x$, then if these three conditions hold then $x_1$ is the minimizer. So if $\frac{1}{2}|y_2| \leq y_1 \leq 2|y_2|$ then the critical point is the minimizer and we get the minimum value of $h_{2,y}(x)$ is

$$h_{2,y}(x_1) = -\frac{1}{3}y_1y_2(3y_1 - 3y_2 - 4\sqrt{-2y_1y_2})$$

If the conditions for the critical point are not satisfied then we must check the boundary conditions. In this case the boundary points are $\max\{y_1 + y_2, 0\} \leq x \leq y_1$, and we consider two subcases:

**Subcase 1:** $y_1 + y_2 \geq 0$

To determine which is the minimum, we evaluate the difference between the upper boundary value and the lower boundary value. When $y_1 + y_2 \geq 0$, the $\max\{y_1 + y_2, 0\} = y_1 + y_2$ so we look at the difference

$$\begin{aligned} h_{2,y}(y_1) - h_{2,y}(y_1 + y_2) &= \frac{1}{3}(-y_2)^3 + \frac{1}{3}y_1^3 - \frac{1}{3}(-y_2)^3 - \frac{1}{3}(y_1 + y_2)^3 \\ &= \frac{1}{3}y_1^3 - \frac{1}{3}(y_1 + y_2)^3 = -\frac{1}{3}y_2^3 - y_1y_2(y_1 + y_2). \end{aligned}$$

Since $y_1 \geq 0$, $y_2 \leq 0$ and $y_1 + y_2 \geq 0$ the above difference is always positive. This means that the minimum value is

$$h_{2,y}(y_1 + y_2) = \frac{1}{3}(y_1 + y_2)^3 - \frac{1}{3}y_2^3.$$

**Subcase 2:** $y_1 + y_2 \leq 0$

With $y_1 + y_2 \leq 0$ the $\max\{y_1 + y_2, 0\} = 0$ so we calculate the difference

$$\begin{aligned} h_{2,y}(y_1) - h_{2,y}(0) &= \frac{1}{3}y_1^3 + \frac{1}{3}(-y_2)^3 - \frac{1}{3}y_1^3 - \frac{1}{3}(-y_1 - y_2)^3 \\ &= \frac{1}{3}(-y_2)^3 - \frac{1}{3}(-y_1 - y_2)^3 = \frac{1}{3}y_1^3 + y_1 y_2 (y_1 + y_2). \end{aligned}$$

Again, because of the signs of $y_1$, $y_2$, and $y_1 + y_2$ the difference is positive so the minimum is

$$h_{2,y}(0) = \frac{1}{3}y_1^3 - \frac{1}{3}(y_1 + y_2)^3.$$

Case 2 is summarized graphically in Figure 6.2.

### 6.4.3 Case 3: $x \geq y_1$, $x \leq y_1 + y_2$, $x \geq 0$

For this case, the function we are looking to minimize is

$$h_{3,y} = \frac{1}{3}(x - y_1)^3 + \frac{1}{3}(y_1 + y_2 - x)^3 + \frac{1}{3}x^3,$$

over the region where $\max\{y_1, 0\} \leq x \leq y_1 + y_2$. Differentiating $h_{3,y}$ yields

$$\frac{\partial h_{3,y}}{\partial x} = (x - y_1)^2 - (y_1 + y_2 - x)^2 + x^2.$$

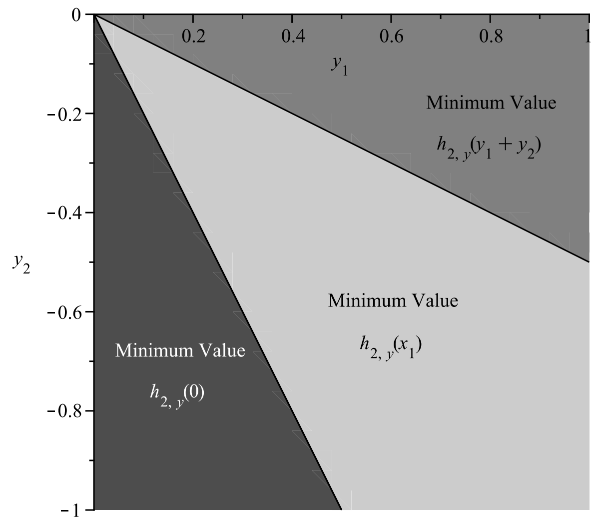



Figure 6.2: Case 2 summary

And setting $\frac{\partial h_{3,y}}{\partial x} = 0$ and solving for $x$ gives us the critical points

$$
\begin{aligned}
x_1 &= -y_2 + \sqrt{2y_2^2 + 2y_1 y_2} \\
x_2 &= -y_2 - \sqrt{2y_2^2 + 2y_1 y_2}.
\end{aligned}
$$

Looking at the constraints for this case we can see that $y_1 + y_2 \geq x \geq y_1$, so we must have $y_2 \geq 0$. And $y_1 + y_2 \geq x \geq 0$ gives us that $y_1 + y_2 \geq 0$. Knowing that $y_2 \geq 0$, it is obvious that $x_2 \leq 0$ and when $x_2 = 0$ then $y_2 = y_1 = 0$ and $x_1 = x_2 = 0$. So this case can be covered by considering only $x_1$.

For $x_1$ to be a real critical point, we need $2y_2^2 + 2y_1 y_2 \geq 0 \Leftrightarrow 2y_2(y_1 + y_2) \geq 0$. Since both $y_1 \geq 0$ and $y_1 + y_2 \geq 0$ always holds for this case then this is always true. We also require that the critical point $x_1$ satisfies the conditions for this case. For $x_1 \geq 0$,

$$-y_2 + \sqrt{2y_2^2 + 2y_1y_2} \geq 0 \Leftrightarrow 2y_2^2 + 2y_1y_2 \geq y_2^2 \Leftrightarrow 2y_1y_2 \geq -y_2^2 \Leftrightarrow 2y_1 \geq -y_2$$

For $x_1 \geq y_1$

$$-y_2+\sqrt{2y_2^2 + 2y_1y_2} \geq y_1 \Leftrightarrow \sqrt{2y_2^2 + 2y_1y_2} \geq y_1+y_2 \Leftrightarrow 2y_2^2+2y_1y_2 \geq y_1^2+2y_1y_2+y_2^2$$
$$\Leftrightarrow y_2^2 \geq y_1^2 \Leftrightarrow y_2 \geq |y_1|$$

And for $x_1 \leq y_1 + y_2$

$$-y_2 + \sqrt{2y_2^2 + 2y_1y_2} \leq y_1 + y_2 \Leftrightarrow 2y_2^2 + 2y_1y_2 \leq y_1^2 + 4y_1y_2 + 4y_2^2$$
$$\Leftrightarrow 0 \leq y_1^2 + 2y_1y_2 + 2y_2^2 \Leftrightarrow 0 \leq (y_1 + y_2)^2 + y_2^2$$

So $x_1 \leq y_1 + y_2$ is always true, and $x_1 \geq 0$ and $x_1 \geq y_1$ hold when $2y_1 \geq -y_2$ and $y_2 \geq |y_1|$, respectively. This means that when both $2y_1 \geq -y_2$ and $y_2 \geq |y_1|$ are true, the critical point $x_1$ will produce the minimum,

$$h_{3,y}(x_1) = \frac{1}{3}y_2(y_1 + y_2)(3y_1 + 6y_2 - 4\sqrt{2y_2(y_1 + y_2)}).$$

Next, we will need to consider two subcases of $y_1 \geq 0$ and $y_1 \leq 0$ separately in order to find the minimum value for each region where the critical point does not satisfy the conditions of this case.

**Subcase 1:** $y_2 \geq 0$**,** $y_1 \leq 0$

If $y_1 \leq 0$ and $2y_1 \geq -y_2$ then the critical point $x_1$ is the minimizer and the minimum value is as stated above. If $2y_1 < -y_2$ then there are no critical points and we look at the boundary points, which in this subcase are $\max\{0, y_1\} = 0 \leq x \leq y_1 + y_2$. Taking the difference of the two potential minimums yields

$$\begin{aligned} h_{3,y}(y_1 + y_2) - h_{3,y}(0) &= \frac{1}{3}y_2^3 + \frac{1}{3}(y_1 + y_2)^3 - \frac{1}{3}|y_1|^3 - \frac{1}{3}(y_1 + y_2)^3 \\ &= \frac{1}{3}y_2^3 - \frac{1}{3}|y_1|^3. \end{aligned}$$

Since $y_1 + y_2 \geq 0 \Leftrightarrow y_2 \geq -y_1 \Leftrightarrow y_2 \geq |y_1|$, then $h_{3,y}(0) \leq h_{3,y}(y_1 + y_2)$ and the lower bound is the minimizer and the minimum value is

$$h_{3,y}(0) = \frac{1}{3}(y_1 + y_2)^3 - \frac{1}{3}y_1^3.$$

**Subcase 2:** $y_2 \geq 0$**,** $y_1 \geq 0$

Since both $y_1$ and $y_2$ are always positive then $2y_1 \geq -y_2$ always holds. Therefore the critical point is good everywhere within the region where $y_2 \geq y_1$. When $y_1 \geq y_2$ then the boundary points are $\max\{0, y_1\} = y_1 \leq x \leq y_1 + y_2$. Taking the difference produces

$$\begin{aligned} h_{3,y}(y_1 + y_2) - h_{3,y}(y_1) &= \frac{1}{3}y_2^3 + \frac{1}{3}(y_1 + y_2)^3 - \frac{1}{3}y_2^3 - \frac{1}{3}y_1^3 \\ &= \frac{1}{3}(y_1 + y_2)^3 - \frac{1}{3}y_1^3 \geq 0. \end{aligned}$$

So the minimum value is

$$h_{3,y}(y_1) = \frac{1}{3}y_2^3 + \frac{1}{3}y_1^3.$$

Case 3 is summarized graphically in Figure 6.3.

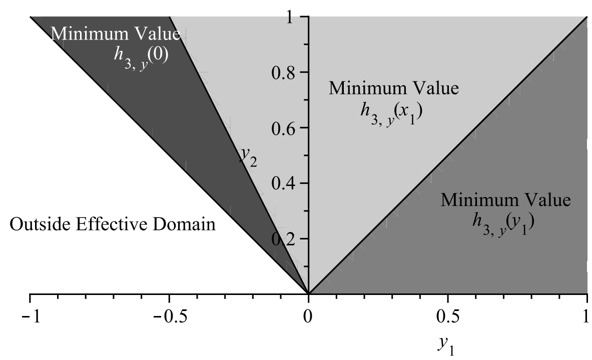



Figure 6.3: Case 3 summary

### 6.4.4 Case 4: $x \geq y_1$, $x \geq y_1 + y_2$, $x \leq 0$

For this case we minimize the function

$$h_{4,y}(x) = \frac{1}{3}(x - y_1)^3 + \frac{1}{3}(x - y_1 - y_2)^3 - \frac{1}{3}x^3$$

over the region where $\max\{y_1, y_1 + y_2\} \leq x \leq 0$. Solving $\frac{\partial h_4}{\partial x} = (x - y_1)^2 + (x - y_1 - y_2)^2 - x^2 = 0$ yields the critical points

$$x_1 = 2y_1 + y_2 + \sqrt{2y_1y_2 + 2y_1^2}$$
$$x_2 = 2y_1 + y_2 - \sqrt{2y_1y_2 + 2y_1^2}.$$

The constraints for this case lead us to the inequalities $y_1 \leq 0$ and $y_1 + y_2 \leq 0$, or $y_2 \leq -y_1 = |y_1|$. In order for the critical points to be real, we require $2y_1y_2 + 2y_1^2 \geq 0 \Leftrightarrow |y_1| \geq y_2$ which is always true in this case. Therefore the critical point will give the minimum value. To satisfy $x \leq 0$ and determine which critical point is the one we want, we need $x_1, x_2 \leq 0$. Looking first at $x_1$:

$$y_2 + 2y_1 + \sqrt{2y_1y_2 + 2y_1^2} \leq 0 \Leftrightarrow (-y_2 - 2y_1)^2 \geq 2y_1y_2 + 2y_1^2$$
$$\Leftrightarrow y_2^2 + 4y_1y_1 + 4y_1^2 \geq 2y_1y_2 + 2y_1^2 \Leftrightarrow y_2^2 + 2y_1y_1 + 2y_1^2 \geq 0$$
$$\Leftrightarrow (y_1 + y_2)^2 + y_1^2 \geq 0.$$

This always holds, so we check the next condition, $x_1 \geq y_1 + y_2$:

$$y_2 + 2y_1 + \sqrt{2y_1y_2 + 2y_1^2} \geq y_1 + y_2 \Leftrightarrow y_1 + \sqrt{2y_1y_2 + 2y_1^2} \geq 0$$
$$\Leftrightarrow 2y_1y_2 + 2y_1^2 \geq y_1^2 \Leftrightarrow 2y_1y_2 + y_1^2 \geq 0 \Leftrightarrow y_1(2y_2 + y_1) \geq 0$$
$$\Leftrightarrow 2y_2 + y_1 \leq 0 \Leftrightarrow y_2 \leq \frac{1}{2}|y_1|.$$

And the final condition, $x_1 \geq y_1$, yields

$$y_2 + 2y_1 + \sqrt{2y_1y_2 + 2y_1^2} \geq y_1 \Leftrightarrow y_1 + y_2 + \sqrt{2y_1y_2 + 2y_1^2} \geq 0$$
$$\Leftrightarrow 2y_1y_2 + 2y_1^2 \geq (-y_1 - y_2)^2 \Leftrightarrow 2y_1y_2 + 2y_1^2 \geq y_1^2 + 2y_1y_2 + y_2^2$$
$$\Leftrightarrow y_1^2 - y_2^2 \geq 0 \Leftrightarrow |y_1| \geq |y_2|.$$

So the critical point $x_1$ is in the region of interest when $y_2 \leq \frac{1}{2}|y_1|$ and $|y_1| \geq |y_2|$ both hold.

Next looking at $x_2$, if we look at the condition $x_2 \geq y_1$ we see that

$$2y_1 + y_2 - \sqrt{2y_1y_2 + 2y_1^2} \geq y_1 \Leftrightarrow (y_1 + y_2) - \sqrt{2y_1y_2 + 2y_1^2} \geq 0.$$

This only holds if $y_1 = y_2 = 0$, in which case $x_1 = x_2$ so this point is not in the interior of the region and we do not need to consider it. So when $y_2 \leq \frac{1}{2}|y_1|$ and $|y_1| \geq |y_2|$ both hold then the minimum value is

$$h_{4,y}(x_1) = -\frac{1}{3}y_1(y_1 + y_2)(6y_1 + 3y_2 + 4\sqrt{2y_1(y_1 + y_2)}).$$

Next, we examine the boundary conditions $\max\{y_1, y_1 + y_2\} \leq x \leq 0$ for the two subcases $y_2 \geq 0$ and $y_2 \leq 0$, to determine the minimum when the constraints needed for the critical point do not hold.

**Subcase 1:** $y_2 \geq 0$

When $y_2 \geq 0$ then the $\max\{y_1, y_1+y_2\} = y_1+y_2$, so we look at the difference

$$\begin{aligned} h_{4,y}(0) - h_{4,y}(y_1+y_2) &= -\frac{1}{3}y_1^3 - \frac{1}{3}(y_1+y_2)^3 - \frac{1}{3}y_2^3 + \frac{1}{3}(y_1+y_2)^3 \\ &= \frac{1}{3}|y_1|^3 - \frac{1}{3}y_2^3. \end{aligned}$$

Since $|y_1| \geq y_2$ then the above difference is positive, so the minimum is

$$h_{4,y}(y_1+y_2) = \frac{1}{3}y_2^3 - \frac{1}{3}(y_1+y_2)^3.$$

**Subcase 2:** $y_2 \leq 0$

When $y_2 \leq 0$ then the $\max\{y_1, y_1+y_2\} = y_1$, so we look at the difference

$$\begin{aligned} h_{4,y}(0) - h_{4,y}(y_1) &= \frac{1}{3}|y_1|^3 + \frac{1}{3}|y_1+y_2|^3 - \frac{1}{3}|y_2|^3 - \frac{1}{3}|y_1|^3 \\ &= \frac{1}{3}|y_1+y_2|^3 - \frac{1}{3}|y_2|^3. \end{aligned}$$

Since $|y_1+y_2| \geq |y_2|$ then the difference is positive and the minimum is

$$h_{4,y}(y_1) = \frac{1}{3}|y_1|^3 + \frac{1}{3}|y_2|^3.$$

Case 4 is summarized graphically in Figure 6.4.

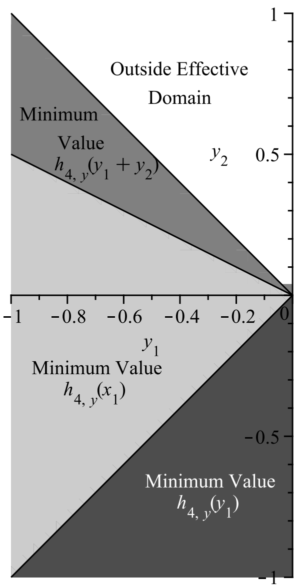



Figure 6.4: Case 4 summary

### 6.4.5 Case 5: $x \leq y_1$, $x \leq y_1 + y_2$, $x \geq 0$

For this case we look at minimizing

$$h_{5,y}(x) = \frac{1}{3}(y_1 - x)^3 + \frac{1}{3}(y_1 + y_2 - x)^3 + \frac{1}{3}x^3,$$

over the region where $0 \leq x \leq \min\{y_1, y_1 + y_2\}$. Differentiating $h_{5,y}$ with respect to $x$ yields

$$\frac{\partial h_{5,y}}{\partial x} = -(y_1 - x)^2 - (y_1 + y_2 - x)^2 + x^2.$$

Then solving $\frac{\partial h_{5,y}}{\partial x} = 0$ gives us the critical points

$$x_1 = 2y_1 + y_2 + \sqrt{2y_1y_2 + 2y_1^2}$$
$$x_2 = 2y_1 + y_2 - \sqrt{2y_1y_2 + 2y_1^2}.$$

These critical points are the same as in the previous case, but must be re-examined using the new constraints of this case. We see that $y_1 \geq x \geq 0$, so $y_1 \geq 0$ and similarly $y_1 + y_2 \geq 0$. In order for the critical points to be real, we need $2y_1y_2 + 2y_1^2 \geq 0 \Leftrightarrow y_1 + y_2 \geq 0$ which always holds for this case. Next, we check where the critical points are valid.

When we check $x_1 \leq y_1$, we see that $2y_1 + y_2 + \sqrt{2y_1y_2 + 2y_1^2} \leq y_1 \Leftrightarrow y_1 + y_2 + \sqrt{2y_1y_2 + 2y_1^2} \leq 0$, but since $y_1 + y_2 \geq 0$ and $\sqrt{2y_1y_2 + 2y_1^2} \geq 0$ then this does not hold except when $y_1 = y_2 = 0$, and it that situation $x_1 = x_2$. Therefore, for this case we need only consider $x_2$.

Looking at $x_2$, we see for $x_2 \geq 0$,

$$2y_1 + y_2 - \sqrt{2y_1y_2 + 2y_1^2} \geq 0 \Leftrightarrow (2y_1 + y_2)^2 \geq 2y_1y_2 + 2y_1^2$$
$$\Leftrightarrow 4y_1^2 + 4y_1y_2 + y_2^2 \geq 2y_1y_2 + 2y_1^2 \Leftrightarrow 2y_1^2 + 2y_1y_2 + y_2^2 \geq 0$$
$$\Leftrightarrow y_1^2 + (y_1 + y_2)^2 \geq 0.$$

This always holds, so next we check $x_2 \leq y_1$:

$$2y_1 + y_2 - \sqrt{2y_1y_2 + 2y_1^2} \leq y_1 \Leftrightarrow (y_1 + y_2)^2 \leq 2y_1y_2 + 2y_1^2$$
$$\Leftrightarrow y_1^2 + 2y_1y_2 + y_2^2 \leq 2y_1y_2 + 2y_1^2 \Leftrightarrow y_2^2 - y_1^2 \leq 0 \Leftrightarrow |y_1| \geq |y_2|.$$

And checking the last condition we see that for $x_2 \leq y_1 + y_2$,

$$\begin{aligned} 2y_1 + y_2 - \sqrt{2y_1y_2 + 2y_1^2} \leq y_1 + y_2 \Leftrightarrow y_1 - \sqrt{2y_1y_2 + 2y_1^2} \leq 0 \\ \Leftrightarrow 2y_1y_2 + 2y_1^2 \geq y_1^2 \Leftrightarrow 2y_1y_2 + y_1^2 \geq 0 \\ \Leftrightarrow y_1(2y_2 + y_1) \geq 0 \Leftrightarrow 2y_2 + y_1 \geq 0 \end{aligned}$$

So the critical point $x_2$ is in the region of interest when both $|y_1| \geq |y_2|$ and $2y_2 + y_1 \geq 0$ hold, and the minimum value is

$$h_{5,y}(x_2) = -\frac{1}{3}y_1(y_1 + y_2)(-6y_1 - 3y_2 + 4\sqrt{2y_1(y_1 + y_2)}).$$

Now we look at the boundary conditions $0 \leq x \leq \min\{y_1, y_1 + y_2\}$ in the two subcases $y_2 \geq 0$ and $y_2 \leq 0$ to determine the minimum when the critical point is not in the region of interest. That is, when $|y_1| < |y_2|$ and $2y_2 + y_1 < 0$.

**Subcase 1:** $y_2 \geq 0$

With $y_2 \geq 0$ the $\min\{y_1, y_1 + y_2\} = y_1$ and the difference that we need to consider is

$$\begin{aligned} h_{5,y}(y_1) - h_{5,y}(0) &= \frac{1}{3}y_2^3 + \frac{1}{3}y_1^3 - \frac{1}{3}y_1^3 - \frac{1}{3}(y_1 + y_2)^3 \\ &= \frac{1}{3}y_2^3 - \frac{1}{3}(y_1 + y_2)^3 = -\frac{1}{3}y_1^3 - y_1y_2(y_1 + y_2) \leq 0 \end{aligned}$$

So the upper bound is the minimizer and the minimum value is

$$h_{5,y}(y_1) = \frac{1}{3}y_1^3 + \frac{1}{3}y_2^3.$$

**Subcase 2:** $y_2 \leq 0$

When $y_2 \leq 0$ then the $\min\{y_1, y_1+y_2\} = y_1+y_2$ and we look at the difference

$$\begin{aligned} h_{5,y}(y_1+y_2) - h_{5,y}(0) &= \frac{1}{3}(-y_2)^3 + \frac{1}{3}(y_1+y_2)^3 - \frac{1}{3}y_1^3 - \frac{1}{3}(y_1+y_2)^3 \\ &= \frac{1}{3}(-y_2)^3 - \frac{1}{3}y_1^3 \end{aligned}$$

Since $y_2 \leq 0$, $y_1 \geq 0$, and $y_1 + y_2 \geq 0$ then $y_1 \geq |y_2|$ and the above equation is always less than or equal to zero, which makes the upper bound, $y_1 + y_2$ the minimizer with a minimum value of

$$h_{5,y}(y_1+y_2) = \frac{1}{3}(y_1+y_2)^3 - \frac{1}{3}y_2^3.$$

Case 5 is summarized graphically in Figure 6.5.

### 6.4.6 Case 6: $x \leq y_1$, $x \geq y_1 + y_2$, $x \leq 0$

For this case we have the function

$$h_{6,y}(x) = \frac{1}{3}(y_1 - x)^3 + \frac{1}{3}(x - y_1 - y_2)^3 + \frac{1}{3}(-x)^3$$

to minimize. Differentiating with respect to $x$ to get our critical points yields

$$\frac{\partial h_{6,y}}{\partial x} = -(y_1 - x)^2 + (x - y_1 - y_2)^2 - x^2$$

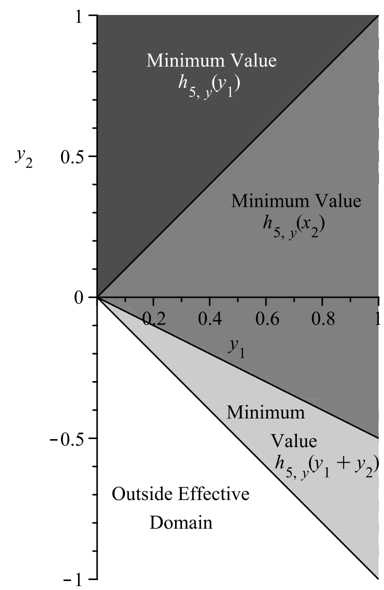



Figure 6.5: Case 5 summary

Setting this equal to zero and solving for $x$ then gives us the critical points

$$x_1 = -y_2 + \sqrt{2y_2^2 + 2y_1y_2}$$
$$x_2 = -y_2 - \sqrt{2y_2^2 + 2y_1y_2}.$$

Looking at the constraints of the case, notice that $y_1 + y_2 \leq x \leq y_1$ and $y_1 + y_2 \leq x \leq 0$ imply that both $y_1 + y_2 \leq 0$ and $y_2 \leq 0$. For the critical points to be real we need $2y_2^2 + 2y_1y_2 \geq 0 \Leftrightarrow 2y_2(y_1 + y_2) \geq 0$, which is always true since $y_2$ and $y_1 + y_2$ are both always negative. It is easy to see that critical point $x_1$ is always positive, so it does not satisfy the conditions of this case, except when $y_1 = y_2 = 0$ which makes $x_1 = x_2$. So it suffices to check only $x_2$ for this case.

Checking the condition $x_2 \geq y_1 + y_2$,

$$\begin{aligned} y_1 + y_2 \leq -y_2 - \sqrt{2y_2^2 + 2y_1y_2} &\Leftrightarrow \sqrt{2y_2^2 + 2y_1y_2} \leq -y_1 - 2y_2 \\ &\Leftrightarrow 2y_2^2 + 2y_1y_2 \leq y_1^2 + 4y_1y_2 + 4y_2^2 \\ &\Leftrightarrow 0 \leq y_1^2 + 2y_1y_2 + 2y_2^2 \Leftrightarrow 0 \leq (y_1 + y_2)^2 + y_2^2. \end{aligned}$$

This condition always holds. Next, we check $x_2 \leq y_1$,

$$\begin{aligned} -y_2 - \sqrt{2y_2^2 + 2y_1y_2} \leq y_1 &\Leftrightarrow -y_1 - y_2 \leq \sqrt{2y_2^2 + 2y_1y_2} \\ &\Leftrightarrow y_1^2 + 2y_1y_2 + y_2^2 \leq 2y_2^2 + 2y_1y_2 \\ &\Leftrightarrow y_1^2 \leq y_2^2 \Leftrightarrow |y_1| \leq |y_2|. \end{aligned}$$

And the last condition is $x_2 \leq 0$,

$$\begin{aligned} -y_2 - \sqrt{2y_2^2 + 2y_1y_2} \leq 0 &\Leftrightarrow -y_2 \leq \sqrt{2y_2^2 + 2y_1y_2} \\ &\Leftrightarrow y_2^2 \leq 2y_2^2 + 2y_1y_2 \Leftrightarrow 0 \leq y_2^2 + 2y_1y_2 \\ &\Leftrightarrow 0 \leq y_2(y_2 + 2y_1) \Leftrightarrow y_2 + 2y_1 \leq 0 \Leftrightarrow 2y_1 \leq |y_2|. \end{aligned}$$

So $x_2$ is the minimizer when both $|y_1| \leq |y_2|$ and $2y_1 \leq |y_2|$ hold and the minimum value is

$$h_{6,y}(x_2) = -\frac{1}{3}y_2(y_1 + y_2)(3y_1 + 6y_2 + 4\sqrt{2y_2(y_1 + y_2)}).$$

Outside this region we check the boundary conditions $y_1 + y_2 \leq x \leq \min\{0, y_1\}$ to determine the minimizer.

**Subcase 1:** $y_1 \geq 0$

With $y_1 \geq 0$ the $\min\{y_1, 0\} = 0$ so we look at the difference

$$\begin{aligned} h_{6,y}(0) - h_{6,y}(y_1 + y_2) &= \frac{1}{3}y_1^3 + \frac{1}{3}(-y_1 - y_2)^3 - \frac{1}{3}(-y_2)^3 - \frac{1}{3}(-y_1 - y_2)^3 \\ &= \frac{1}{3}y_1^3 - \frac{1}{3}(-y_2)^3 = \frac{1}{3}y_1^3 - \frac{1}{3}|y_2|^3. \end{aligned}$$

Since $y_1 + y_2 \leq 0$ we know that $|y_2| \geq y_1$ and the above equation is less than or equal to zero. This makes the upper boundary the minimum and the minimum value is

$$h_{6,y}(0) = \frac{1}{3}y_1^3 - \frac{1}{3}(y_1 + y_2)^3.$$

**Subcase 2:** $y_1 \leq 0$

When $y_1 \leq 0$ the $\min\{y_1, 0\} = y_1$ and we consider the difference

$$\begin{aligned} h_{6,y}(y_1) - h_{6,y}(y_1 + y_2) &= \frac{1}{3}(-y_2)^3 + \frac{1}{3}(-y_1)^3 - \frac{1}{3}(-y_2)^3 - \frac{1}{3}(-y_1 - y_2)^3 \\ &= \frac{1}{3}|y_1|^3 - \frac{1}{3}|y_1 + y_2|^3. \end{aligned}$$

Because $y_1$ and $y_2$ are both negative then $|y_1 + y_2| \geq |y_1|$ and the equation above is less than or equal to zero, giving us a minimum value of

$$h_{6,y}(y_1) = -\frac{1}{3}y_1^3 - \frac{1}{3}y_2^3.$$

Case 6 is summarized graphically in Figure 6.6.

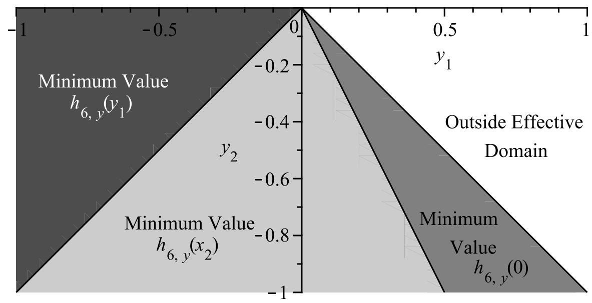



Figure 6.6: Case 6 summary

### 6.4.7 Case 7: $x \geq y_1$, $x \leq y_1 + y_2$, $x \leq 0$

In this case the equation to be minimized is

$$h_{7,y}(x) = \frac{1}{3}(x - y_1)^3 + \frac{1}{3}(y_1 + y_2 - x)^3 + \frac{1}{3}(-x)^3,$$

over the region where $y_1 \leq x \leq \min\{0, y_1 + y_2\}$. Differentiating $h_{7,y}$ with respect to $x$ yields

$$\frac{\partial h_{7,y}}{\partial x} = (x - y_1)^2 - (y_1 + y_2 - x)^2 - x^2.$$

Setting $\frac{\partial h_{7,y}}{\partial x} = 0$ and solving for $x$ gives us the critical points

$$x_1 = y_2 + \sqrt{-2y_1y_2}$$
$$x_2 = y_2 - \sqrt{-2y_1y_2}.$$

Looking at the constraints we see that $y_1 \leq x \leq 0$ and $y_1 \leq x \leq y_1 + y_2$ so we have $y_1 \leq 0$ and $y_2 \geq 0$ for this case. For the critical points to be real, $-2y_1y_2 \geq 0$ must hold, and because of the signs of $y_1$ and $y_2$ this is always true. The critical point $x_1 \leq 0$ so it does not lie in the interior of the region of interest. And when $x_1 = 0$ then $y_1 = y_2 = 0$ and so $x_1 = x_2$. Therefore we need only consider $x_2$, and we check $x_2$ against the constraints. First, $x_2 \geq y_1$,

$$y_2 - \sqrt{-2y_1y_2} \geq y_1 \Leftrightarrow (y_2 - y_1)^2 \geq -2y_1y_2 \Leftrightarrow y_2^2 + y_1^2 \geq 0.$$

This will always hold, so next we look at $x_2 \leq 0$

$$y_2 - \sqrt{-2y_1y_2} \leq 0 \Leftrightarrow y_2^2 \leq -2y_1y_2 \Leftrightarrow y_2^2 + 2y_1y_2 \leq 0 \Leftrightarrow y_2(y_2 + 2y_1) \leq 0$$
$$\Leftrightarrow y_2 + 2y_1 \leq 0 \Leftrightarrow 2|y_1| \geq y_2.$$

And finally at $x_2 \leq y_1 + y_2$,

$$y_2 - \sqrt{-2y_1y_2} \leq y_1 + y_2 \Leftrightarrow y_1^2 \leq -2y_1y_2 \Leftrightarrow |y_1| \leq 2y_2 \Leftrightarrow \frac{1}{2}|y_1| \leq y_2.$$

Then the critical point $x_2$ is the minimizer if $\frac{1}{2}|y_1| \leq y_2 \leq 2|y_1|$, and the minimum value is

$$h_{7,y} = \frac{1}{3}y_1y_2(3y_1 - 3y_2 + 4\sqrt{-2y_1y_2}).$$

Outside this region, we look at the boundary conditions $y_1 \leq x \leq \min\{0, y_1 + y_2\}$ to determine the minimizer.

**Subcase 1:** $y_1 + y_2 \geq 0$

When $y_1 + y_2 \geq 0$ the $\min\{0, y_1 + y_2\} = 0$ so we look at the difference

$$\begin{aligned} h_{7,y}(0) - h_{7,y}(y_1) &= \frac{1}{3}(-y_1)^3 + \frac{1}{3}(y_1 + y_2)^3 - \frac{1}{3}y_2^3 - \frac{1}{3}(-y_1)^3 \\ &= \frac{1}{3}(y_1 + y_2)^3 - \frac{1}{3}y_2^3 = \frac{1}{3}y_1^3 + y_1y_2(y_1 + y_2). \end{aligned}$$

Because $y_1 \leq 0$, $y_2 \geq 0$, then $y_1 + y_2 \leq y_2$ and $(y_1 + y_2)^3 \leq y_2^3$ so that the above equation is always negative and therefore the minimum value is

$$h_{7,y}(0) = \frac{1}{3}(y_1 + y_2)^3 - \frac{1}{3}y_1^3.$$

**Subcase 2:** $y_1 + y_2 \leq 0$

Here the $\min\{0, y_1 + y_2\} = y_1 + y_2$ so the difference to consider is

$$\begin{aligned} h_{7,y}(y_1 + y_2) - h_{7,y}(y_1) &= \frac{1}{3}y_2^3 + \frac{1}{3}(-y_1 - y_2)^3 - \frac{1}{3}y_2^3 - \frac{1}{3}(-y_1)^3 \\ &= \frac{1}{3}(-y_1 - y_2)^3 - \frac{1}{3}(-y_1)^3 = -\frac{1}{3}y_2^3 - y_1y_2(y_1 + y_2) \end{aligned}$$

Again, because of the signs of $y_1$ and $y_2$, we get that $(-y_1 - y_2) \leq -y_1$ and $(-y_1 - y_2)^3 \leq (-y_1)^3$ so that this equation is always negative so the minimum value is

$$h_{7,y}(y_1 + y_2) = \frac{1}{3}y_2^3 - \frac{1}{3}(y_1 + y_2)^3.$$

Case 7 is summarized graphically in Figure 6.7.

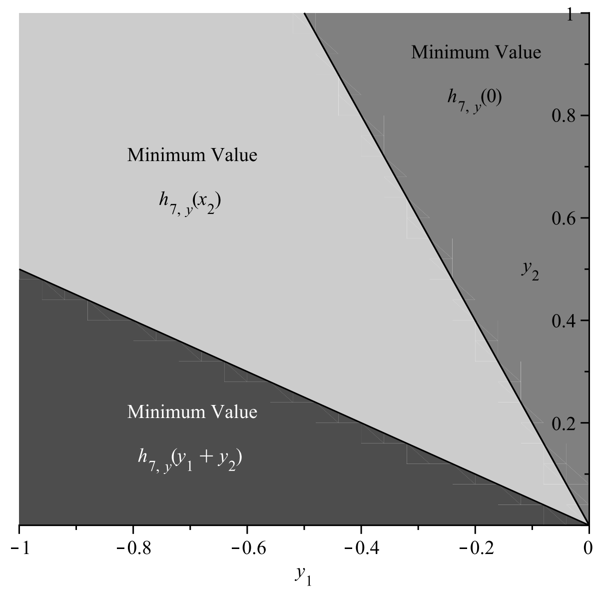



Figure 6.7: Case 7 summary

### 6.4.8 Case 8: $x \leq y_1$, $x \leq y_1 + y_2$, $x \leq 0$

For the final case the function we minimize is

$$h_{8,y}(x) = \frac{1}{3}(y_1 - x)^3 + \frac{1}{3}(y_1 + y_2 - x)^3 + \frac{1}{3}(-x)^3,$$

over the region where $x \leq \min\{0, y_1, y_1 + y_2\}$. Differentiating $h_{8,y}$ with respect to $x$ yields

$$\frac{\partial h_{8,y}}{\partial x} = -(y_1 - x)^2 - (y_1 + y_2 - x)^2 - x^2.$$

Setting this equal to zero and solving for $x$ gives the critical points

$$\begin{aligned} x_1 &= \frac{2}{3}y_1 + \frac{1}{3}y_2 + \frac{1}{3}\sqrt{-2y_1^2 - 2y_1y_2 - 2y_2^2} \\ x_2 &= \frac{2}{3}y_1 + \frac{1}{3}y_2 - \frac{1}{3}\sqrt{-2y_1^2 - 2y_1y_2 - 2y_2^2}. \end{aligned}$$

As we saw in case 1,

$$-2y_1^2 - 2y_1y_2 - 2y_2^2 = -(y_1^2 + y_2^2) - (y_1 + y_2)^2 \leq 0$$

so the critical points are only real when $y_1 = y_2 = 0$. Thus we are only left with the boundary conditions $x \leq \min\{0, y_1, y_1 + y_2\}$. If $\min\{0, y_1, y_1 + y_2\} = 0$ then the minimum value is

$$h_{8,y}(0) = \frac{1}{3}y_1^3 + \frac{1}{3}(y_1 + y_2)^3.$$

When the $\min\{0, y_1, y_1 + y_2\} = y_1$ then the minimum value is

$$h_{8,y}(y_1) = \frac{1}{3}y_2^3 + \frac{1}{3}(-y_1)^3.$$

And finally, if $\min\{0, y_1, y_1 + y_2\} = y_1 + y_2$, the minimum value is

$$h_{8,y}y_1 + y_2) = \frac{1}{3}(-y_2)^3 + \frac{1}{3}(-y_1 - y_2)^3.$$

Case 8 is summarized graphically in Figure 6.8.

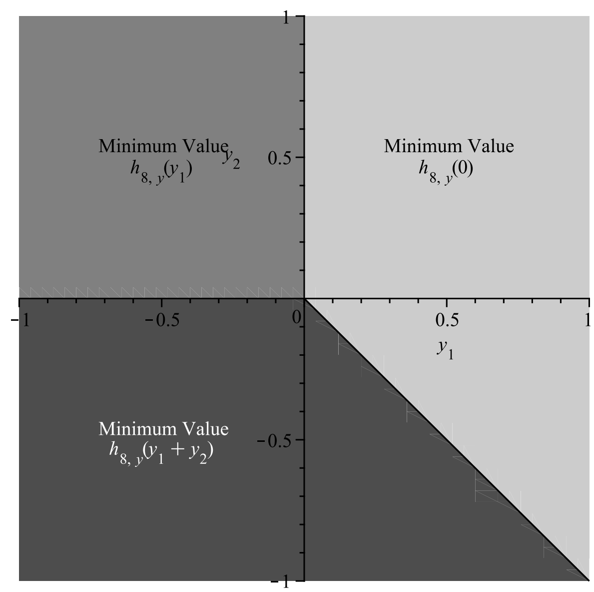



Figure 6.8: Case 8 summary

### 6.4.9 Combining the Eight Cases

Now that each possibility has been examined, each case must be compared against the others to determine the global minimum in each region. Cases 2,3, and 4 and Cases 5,6, and 7 can be plotted without any overlap, as seen in Figure 6.9 on page 97. This leaves 12 regions each with four possible minimizers.

Since we can see from equation (6.10) that $h(x)$ is convex with respect to $x$, the critical point will be the minimum in regions where there is a valid critical point. Combining all of the eight cases, we can see that there is a valid critical point for every possible region. We therefore have only six regions, as seen in Figure 6.10 on page 98. The regions are divided by the lines $y_1 - y_2 = 0$, $2y_1 + y_2 = 0$, and $y_1 + 2y_2 = 0$, and each region has a critical point minimizer.

The six regions are defined as follows. Let $y = (y_1, y_2)$, then

$$\begin{aligned} y \in A &\text{ if } -\frac{1}{2}y_1 \leq y_2 \leq y_1, \\ y \in B &\text{ if } -\frac{1}{2}y_2 \leq y_1 \leq y_2, \\ y \in C &\text{ if } -2y_2 \leq y_1 \leq -\frac{1}{2}y_2, \\ y \in D &\text{ if } y_1 \leq y_2 \leq -\frac{1}{2}y_1, \\ y \in E &\text{ if } y_2 \leq y_1 \leq -\frac{1}{2}y_2, \text{ and} \\ y \in F &\text{ if } -2y_1 \leq y_2 \leq -\frac{1}{2}y_1. \end{aligned} \tag{6.11}$$

The minimum for each region is outlined below

(A) When $y \in A$, then the minimizer is the critical point $x_2$ from Case 5, and the minimum value is

$$\begin{aligned} h_{5,y}(x_2) &= h_{5,y}(2y_1 + y_2 - \sqrt{2y_1y_2 + 2y_1^2}) \\ &= -\frac{1}{3}y_1(y_1 + y_2)(-6y_1 - 3y_2 + 4\sqrt{2y_1(y_1 + y_2)}). \end{aligned}$$

(B) When $y \in B$, the minimizer is the critical point $x_1$ from Case 3, and the minimum value is

$$\begin{aligned} h_{3,y}(x_1) &= h_{3,y}(-y_2 + \sqrt{2y_2^2 + 2y_1y_2}) \\ &= -\frac{1}{3}y_2(y_1 + y_2)(-3y_1 - 6y_2 + 4\sqrt{2y_2(y_1 + y_2)}). \end{aligned}$$

(C) When $y \in C$, the minimizer is the critical point $x_2$ from Case 7, and the minimum value is

$$\begin{aligned} h_{7,y}(x_2) &= h_{7,y}(y_2 - \sqrt{-2y_1y_2}) \\ &= \frac{1}{3}y_1y_2(3y_1 - 3y_2 + 4\sqrt{-2y_1y_2}). \end{aligned}$$

(D) When $y \in D$, the minimizer is the critical point $x_1$ from Case 4, and the minimum value is

$$\begin{aligned} h_{4,y}(x_1) &= h_{4,y}(2y_1 + y_2 + \sqrt{2y_1y_2 + 2y_1^2}) \\ &= -\frac{1}{3}y_1(y_1 + y_2)(6y_1 + 3y_2 + 4\sqrt{2y_1(y_1 + y_2)}). \end{aligned}$$

(E) When $y \in E$, the minimizer is the critical point $x_2$ from Case 6, and the minimum value is

$$\begin{aligned} h_{6,y}(x_2) &= h_{6,y}(-y_2 - \sqrt{2y_2^2 + 2y_1y_2}) \\ &= -\frac{1}{3}y_2(y_1 + y_2)(3y_1 + 6y_2 + 4\sqrt{2y_2(y_1 + y_2)}). \end{aligned}$$

(F) And when $y \in F$, the minimizer is the critical point $x_1$ from Case 2, and the minimum value is

$$\begin{aligned} h_{2,y}(x_1) &= h_{2,y}(y_2 + \sqrt{-2y_1y_2}) \\ &= \frac{1}{3}y_1y_2(-3y_1 + 3y_2 + 4\sqrt{-2y_1y_2}). \end{aligned}$$

### 6.4.10 Bringing It All Together

If you recall, the goal was to solve

$$\min_{x\in\mathbb{R}} h(x) = \min_{x\in\mathbb{R}} \left( \frac{1}{3}|x - y_1|^3 + \frac{1}{3}|x - (y_1 + y_2)|^3 + \frac{1}{3}|x|^3, \right).$$

in order to get $g_2^*(y_1, y_2, y_3)$, where $y_1 + y_2 + y_3 = 0$

$$g_2(x_1, x_2, x_3) = \frac{2}{3}\|x_1 - x_2\|^{\frac{3}{2}} + \frac{2}{3}\|x_2 - x_3\|^{\frac{3}{2}} + \frac{2}{3}\|x_3 - x_1\|^{\frac{3}{2}}$$

The minimizer of $h(x)$ has been found for each of the six regions, and so we have found that $g_2(y_1, y_2, -y_1 - y_2) =$

$$\begin{cases} -\frac{1}{3}y_1(y_1+y_2)(-6y_1 - 3y_2 + 4\sqrt{2y_1(y_1+y_2)}), & \text{if } y \in A; \\ -\frac{1}{3}y_2(y_1+y_2)(-3y_1 - 6y_2 + 4\sqrt{2y_2(y_1+y_2)}), & \text{if } y \in B; \\ \frac{1}{3}y_1y_2(3y_1 - 3y_2 + 4\sqrt{-2y_1y_2}), & \text{if } y \in C; \\ -\frac{1}{3}y_1(y_1+y_2)(6y_1 + 3y_2 + 4\sqrt{2y_1(y_1+y_2)}), & \text{if } y \in D; \\ -\frac{1}{3}y_2(y_1+y_2)(3y_1 + 6y_2 + 4\sqrt{2y_2(y_1+y_2)}), & \text{if } y \in E; \\ \frac{1}{3}y_1y_2(-3y_1 + 3y_2 + 4\sqrt{-2y_1y_2}), & \text{if } y \in F, \end{cases} \tag{6.12}$$

where the regions $A, \cdots, F$ are as defined in (6.11) on page 90.

Examining a plot of the above function and its contour plot, in Figure 6.11 on page 98, we see that $g_2^*$ is convex which is what we would expect from a conjugate function, even though $g_2^*$ is not obviously convex upon inspection.

Recall from (6.7) on page 63 that we had three conjugate variables, $y_1, y_2$, and $y_3$ such that $y_1 + y_2 + y_3 = 0$.

Making the substitution $y_3 = -(y_1 + y_2)$ allows us to write (6.12) in the more symmetric form

$$g_2^*(y_1, y_2, y_3) = \begin{cases} \frac{1}{3}y_1y_3(3y_3 - 3y_1 + 4\sqrt{-2y_1y_3}), & \text{if } (y_1, y_2) \in A; \\ \frac{1}{3}y_2y_3(3y_3 - 3y_2 + 4\sqrt{-2y_2y_3}), & \text{if } (y_1, y_2) \in B; \\ \frac{1}{3}y_1y_2(3y_1 - 3y_2 + 4\sqrt{-2y_1y_2}), & \text{if } (y_1, y_2) \in C; \\ \frac{1}{3}y_1y_3(3y_1 - 3y_3 + 4\sqrt{-2y_1y_3}), & \text{if } (y_1, y_2) \in D; \\ \frac{1}{3}y_2y_3(3y_2 - 3y_3 + 4\sqrt{-2y_2y_3}), & \text{if } (y_1, y_2) \in E; \\ \frac{1}{3}y_1y_2(3y_2 - 3y_1 + 4\sqrt{-2y_1y_2}), & \text{if } (y_1, y_2) \in F, \end{cases} \tag{6.13}$$

where the regions $A, \cdots, F$ are as defined in (6.11) on page 90.

With the $y_3$ variable added back into the equation, we can then recognize that the three boundaries $y_1 - y_2 = 0$, $2y_1 + y_2 = 0$, and $y_1 + 2y_2 = 0$ are equivalent to $y_1 = y_2$, $y_1 = y_3$, and $y_2 = y_3$. If we consider region $A$ defined by $y_1 \geq y_2 \geq -\frac{1}{2}y_1$, and look at the difference,

$$\begin{aligned} y_2 - y_3 &= y_2 - (-y_1 - y_2) = y_1 + 2y_2 \\ &\geq y_1 + 2(-\frac{1}{2}y_1) = 0, \end{aligned}$$

so $\min\{y_1, y_2, y_3\} = y_3$ and $\max\{y_1, y_2, y_3\} = y_1$. Thus, we can rewrite using $y_{max} = \max\{y_1, y_2, y_3\}$ and $y_{min} = \min\{y_1, y_2, y_3\}$

$$\begin{aligned} g_2^*(y_1, y_2, y_3) &= \frac{1}{3}y_1y_3(3y_3 - 3y_1 + 4\sqrt{-2y_1y_3}) \\ &= y_{max}y_{min}^2 - y_{max}^2y_{min} + \frac{4}{3}y_{max}y_{min}\sqrt{-2y_{max}y_{min}}, \end{aligned}$$

if $(y_1, y_2) \in A$.

Similarly, all of the other regions can be rewritten in the same manner so that (6.13) can be rewritten using $y_{max} = \max\{y_1, y_2, y_3\}$ and $y_{min} = \min\{y_1, y_2, y_3\}$ as

$$g_2^*(y_1, y_2, y_3) = y_{max}y_{min}^2 - y_{max}^2 y_{min} + \frac{4}{3}y_{max}y_{min}\sqrt{-2y_{max}y_{min}}, \quad (6.14)$$

without the need to specify the region.

**Remark 6.4.1** *Although $g_2^*$ does not look convex, it is because it arose as a conjugate function. Convexity can also be shown with calculus if you proceed as follows.*

*We have three variables such that $y_{max} \geq y_0 \geq y_{min}$ and $y_{max} + y_0 + y_{min} = 0$, so $y_0 = -y_{max} - y_{min}$ and hence*

$$y_{max} \geq -y_{max} - y_{min} \geq y_{min}.$$

*Equivalently, $2y_{max} + y_{min} \geq 0$ and $-2y_{min} - y_{max} \geq 0$. Now define $x := y_{max}$ and $y := -y_{min}$, then both $x \geq 0$ and $y \geq 0$ and we care about the region where*

$$2x - y \geq 0 \text{ and } 2y - x \geq 0. \quad (6.15)$$

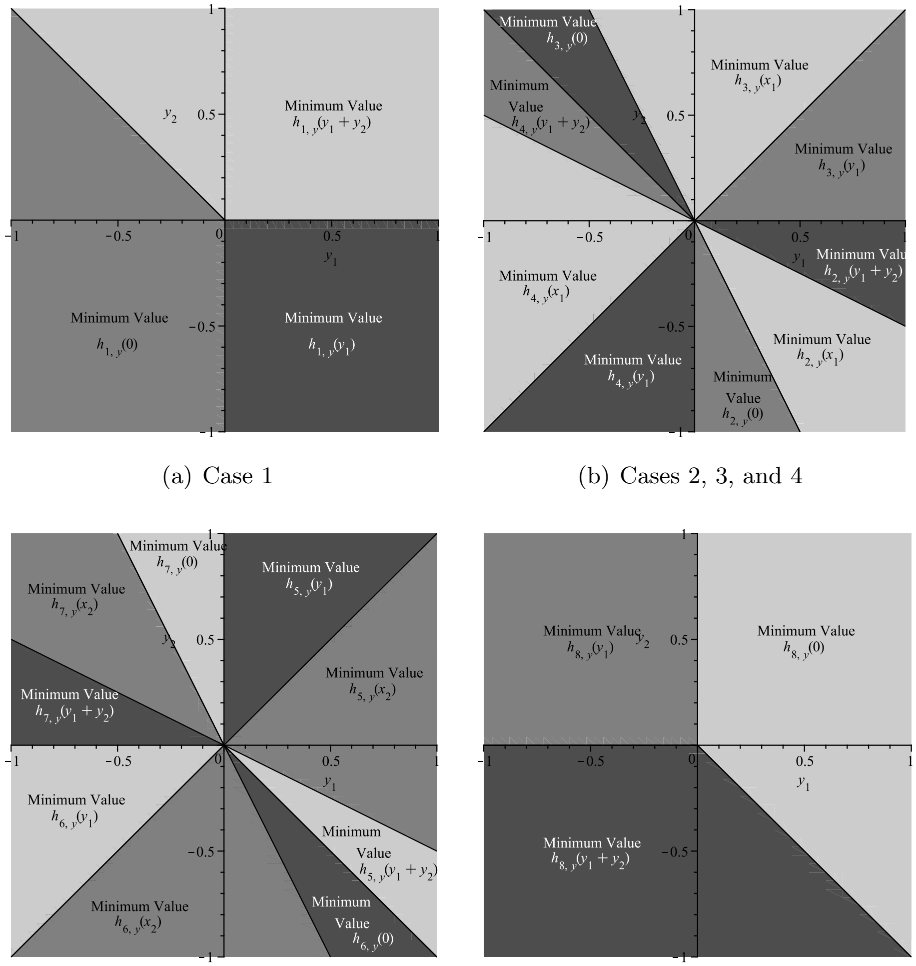
*Then we get the function*

$$f(x, y) = xy^2 + yx^2 - \frac{4}{3}xy\sqrt{2xy},$$

*with*

$$\nabla^2 f(x, y) = \begin{pmatrix} \frac{(2\sqrt{xy}-\sqrt{2}y)y}{\sqrt{xy}} & \frac{2y\sqrt{xy}+2x\sqrt{xy}-3\sqrt{2}xy}{\sqrt{xy}} \\ \frac{2y\sqrt{xy}+2x\sqrt{xy}-3\sqrt{2}xy}{\sqrt{xy}} & \frac{(2\sqrt{xy}-\sqrt{2}x)x}{\sqrt{xy}} \end{pmatrix}$$

*It can be shown that the* $(1, 1)$ *and* $(2, 2)$ *elements of* $\nabla^2 f(x, y)$ *are positive using the constraints in* (6.15). *The determinant of* $\nabla^2 f(x, y)$ *can be shown to be positive by assuming* $x = a^2$ *and* $y = b^2$*, factoring the resulting equation, and considering the signs of each of the factors. Since the Hessian is positive semidefinite then* $f$ *is convex.*




(a) Case 1

(b) Cases 2, 3, and 4

(c) Cases 5, 6, and 7

(d) Case 8

Figure 6.9: Overview of the eight cases

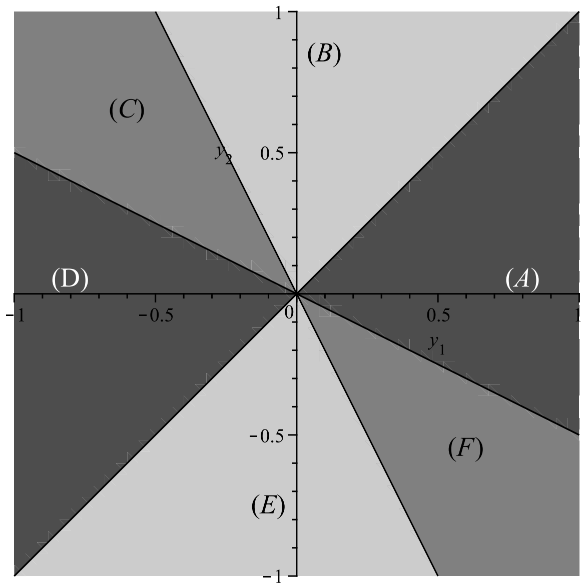
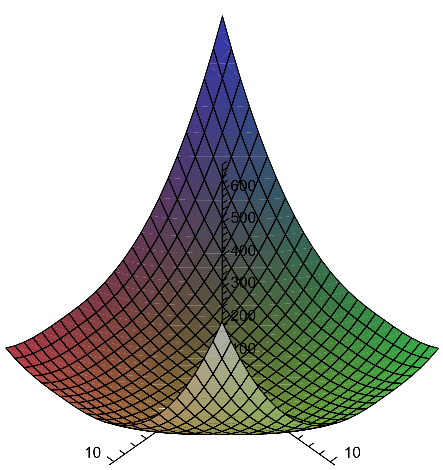
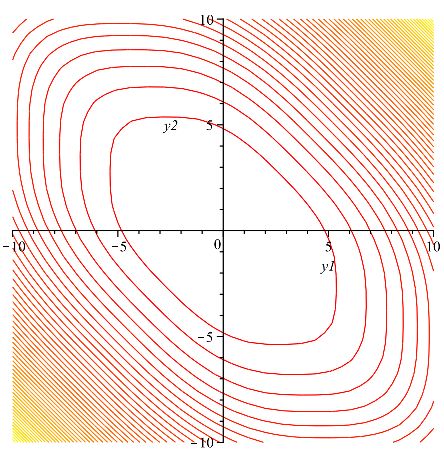



Figure 6.10: The six regions



(a) The function $g_2^*(y_1, y_2, -y_1 - y_2)$




(b) The contour plot of $g_2^*(y_1, y_2, -y_1 - y_2)$

Figure 6.11: Plots of $g_2^*(y_1, y_2, -y_1 - y_2)$

# Chapter 7

# Conclusion

The kernel average was previously only defined for two functions. We have used the identity

$$\sum_{i=1}^{n} \lambda_i \mathfrak{q}(y_i) - \mathfrak{q}(\sum_{i=1}^{n} \lambda_i y_i) = \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \|y_i - y_j\|^2,$$

and the definition of the proximal average to define the kernel average for $n$ functions,

$$Q_g(\mathbf{f}, \boldsymbol{\lambda})(x) := \inf_{\sum \lambda_i y_i = x} \big( \sum \lambda_i f_i(y_i) + \sum_{i<j} \lambda_i \lambda_j g(y_i - y_j) \big).$$

We then examined a specific case of the kernel average, namely the proximal average, and its conjugate and calculated that for

$$g_1(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2} \|x_i - x_j\|^2 = 2 \sum_{i<j} \frac{1}{2} \|x_i - x_j\|^2,$$

the conjugate function is

$$g_1^*(x^*) = \begin{cases} \frac{1}{8n^2} \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} \|x_i^* - x_j^*\|^2, & \text{if } x_1^* + \cdots + x_n^* = 0; \\ \\ +\infty, & \text{otherwise.} \end{cases}$$

This can also be written using any of the following equivalent formulations when $x_1^* + \cdots + x_n^* = 0$:

(i) $g_1^*(x^*) = \frac{1}{8n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|x_i^* - x_j^*\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2} \|\frac{x_i^* - x_j^*}{2n}\|^2$

(ii) $g_1^*(x^*) = \frac{1}{8n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|x_i^* + x_j^*\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2} \|\frac{x_i^* + x_j^*}{2n}\|^2$

(iii) $g_1^*(x^*) = \frac{1}{2n} \sum_{i=1}^{n} \frac{1}{2} \|x_i\|^2$.

Next, we computed the conjugate when

$$g_2(x_1, x_2, x_3) = \frac{2}{3}\|x_1 - x_2\|^{\frac{3}{2}} + \frac{2}{3}\|x_2 - x_3\|^{\frac{3}{2}} + \frac{2}{3}\|x_3 - x_1\|^{\frac{3}{2}},$$

and found that

$$g_2^*(y_1, y_2, y_3) = \begin{cases} \frac{1}{3}y_1y_3(3y_3 - 3y_1 + 4\sqrt{-2y_1y_3}), & \text{if } (y_1, y_2) \in A; \\ \frac{1}{3}y_2y_3(3y_3 - 3y_2 + 4\sqrt{-2y_2y_3}), & \text{if } (y_1, y_2) \in B; \\ \frac{1}{3}y_1y_2(3y_1 - 3y_2 + 4\sqrt{-2y_1y_2}), & \text{if } (y_1, y_2) \in C; \\ \frac{1}{3}y_1y_3(3y_1 - 3y_3 + 4\sqrt{-2y_1y_3}), & \text{if } (y_1, y_2) \in D; \\ \frac{1}{3}y_2y_3(3y_2 - 3y_3 + 4\sqrt{-2y_2y_3}), & \text{if } (y_1, y_2) \in E; \\ \frac{1}{3}y_1y_2(3y_2 - 3y_1 + 4\sqrt{-2y_1y_2}), & \text{if } (y_1, y_2) \in F, \end{cases}$$

where the regions $A, B, C, D, E$, and $F$ are as defined in (6.11), or equivalently using $y_{max} = \max\{y_1, y_2, y_3\}$ and $y_{min} = \min\{y_1, y_2, y_3\}$,

$$g_2^*(y_1, y_2, y_3) = y_{max}y_{min}^2 - y_{max}^2 y_{min} + \frac{4}{3}y_{max}y_{min}\sqrt{-2y_{max}y_{min}}. \quad (7.1)$$

It was expected that we might find a similar form for $g_2^*$ as was found for $g_1^*$, which would help formulate a closed form solution for $\tilde{g}^*$ in the general case where

$$\tilde{g}(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j g(x_i - x_j),$$

for any function $g$. However, due to the fact that there does not appear to be any correlation between the solutions for $g_1^*$ and $g_2^*$, it seems unlikely that a general solution will be found.

# Bibliography

[1] H. H. Bauschke, R. Goebel, Y. Lucet and X. Wang, "The proximal average: basic theory," *SIAM Journal on Optimization,* vol 19(2), pp. 766-785, 2008.

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

[3] H. H. Bauschke and X. Wang, "The kernel average for two convex functions and its application to the extension and representation of monotone operators," *Transactions of the American Mathematical Society* 361, pp. 5947-5965, 2009

[4] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization,* Springer, 2006.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization,* Cambridge University Press, 2004.

[6] P. L. Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inverse Problems,* vol. 24(6), 2008.

[7] K. R. Davidson and A. P. Donsig, *Real Analysis with Real Applications,* Prentice Hall, 2002.

[8] R. Ellaia and J.-B. Hiriart-Urruty, “The conjugate of the difference of functions,” *Journal of Optimization Theory and Applications,* vol 49(3), pp.493-498, 1986.

[9] R. Goebel, “Self-dual smoothing of convex and saddle functions,” *Journal of Convex Analysis,* vol 15(1), pp.179-190, 2008.

[10] E. Kreyszig, *Introductory Functional Analysis with Applications,* John Wiley and Sons, 1978.

[11] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra,* SIAM, 2000.

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

[13] R. T. Rockafellar and R. J-B Wets, *Variational Analysis,* Springer-Verlag, 2004.

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

# Appendix A

# Maple Code

The following is the Maple code used to verify the calculations for chapter 6.

```
>restart: with(plots):
Case 1:
> h1 := x-> (1/3)*(x-y[1])^3+(1/3)*(x-y[1]-y[2])^3+(1/3)*x^3;
```

$$x- > \frac{1}{3}(x-y_1)^3 + \frac{1}{3}(x-y_1-y_2)^3 + \frac{1}{3}x^3$$

```
> dh1 := diff(h1(x), x); criticalpoints1 := solve(dh1 = 0, x);
```

$$\frac{1}{3}y_2 + \frac{2}{3}y_1 + \frac{1}{3}\sqrt{-2y_2^2 - 2y_1y_2 - 2y_1^2}, \frac{1}{3}y_2 + \frac{2}{3}y_1 - \frac{1}{3}\sqrt{-2y_2^2 - 2y_1y_2 - 2y_1^2}$$

```
Case 2:
> h2 := x-> (1/3)*(y[1]-x)^3+(1/3)*(x-y[1]-y[2])^3+(1/3)*x^3;
```

$$x- > \frac{1}{3}(y_1-x)^3 + \frac{1}{3}(x-y_1-y_2)^3 + \frac{1}{3}x^3$$

```
> dh2 := diff(h2(x), x): criticalpoints2 := solve(dh2 = 0, x);
```

$$y_2 + \sqrt{-2y_1y_2}, y_2 - \sqrt{-2y_1y_2}$$

```
> h2soln1 := factor(simplify(h2(criticalpoints2[1])));
```

$$-\frac{1}{3}y_1y_2(3y_1 - 3y_2 - 4\sqrt{2}\sqrt{-y_1y_2})$$

```
> h2soln2 := factor(simplify(h2(criticalpoints2[2])));
```

$$-\frac{1}{3}y_1y_2(3y_1 - 3y_2 + 4\sqrt{2}\sqrt{-y_1y_2})$$

Case 3:

```
> h3 := x-> (1/3)*(x-y[1])^3+(1/3)*(y[1]+y[2]-x)^3+(1/3)*x^3;
```

$$x- > \frac{1}{3}(x - y_1)^3 + \frac{1}{3}(y_1 + y_2 - x)^3 + \frac{1}{3}x^3$$

```
>dh3 := diff(h3(x), x): criticalpoints3 := solve(dh3 = 0, x);
```

$$-y_2 + \sqrt{2y_2^2 + 2y_1y_2}, -y_2 - \sqrt{2y_2^2 + 2y_1y_2}$$

```
>h3soln1 := factor(simplify(h3(criticalpoints3[1])));
```

$$-\frac{1}{3}y_2(y_2 + y_1)(-3y_1 - 6y_2 + 4\sqrt{2}\sqrt{y_2(y_2 + y_1)})$$

```
>h3soln2 := factor(simplify(h3(criticalpoints3[2])));
```

$$\frac{1}{3}y_2(y_2+y_1)(3y_1+6y_2+4\sqrt{2}\sqrt{y_2(y_2+y_1)})$$

```
Case 4:
> h4 := x-> (1/3)*(x-y[1])^3+(1/3)*(x-y[1]-y[2])^3-(1/3)*x^3;
```

$$x->\frac{1}{3}(x-y_1)^3+\frac{1}{3}(x-y_1-y_2)^3+\frac{1}{3}x^3$$

```
>dh4 := diff(h4(x), x): criticalpoints4 := solve(dh4 = 0, x);
```

$$y_2+2y_1+\sqrt{2y_1y_2+2y_1^2}, y_2+2y_1-\sqrt{2y_1y_2+2y_1^2}$$

```
>h4soln1 := factor(simplify(h4(criticalpoints4[1])));
```

$$-\frac{1}{3}y_1(y_2+y_1)(3y_2+6y_1+4\sqrt{2y_1(y_2+y_1)})$$

```
>h4soln2 := factor(simplify(h4(criticalpoints4[2])));
```

$$\frac{1}{3}y_1(y_2+y_1)(-3y_2-6y_1+4\sqrt{2y_1(y_2+y_1)})$$

```
Case 5:
> h5 := x-> (1/3)*(y[1]-x)^3+(1/3)*(y[1]+y[2]-x)^3+(1/3)*x^3;
```

$$x->\frac{1}{3}(y_1-x)^3+\frac{1}{3}(y_1+y_2-x)^3+\frac{1}{3}x^3$$

```
>dh5 := diff(h5(x), x): criticalpoints5 := solve(dh5 = 0, x);
```

$$y_2 + 2y_1 + \sqrt{2y_1y_2 + 2y_1^2}, y_2 + 2y_1 - \sqrt{2y_1y_2 + 2y_1^2}$$

```
>h5soln1 := factor(simplify(h5(criticalpoints5[1])));
```

$$\frac{1}{3}y_1(y_2 + y_1)(3y_2 + 6y_1 + 4\sqrt{2y_1(y_2 + y_1)})$$

```
>h5soln2 := factor(simplify(h5(criticalpoints5[2])));
```

$$-\frac{1}{3}y_1(y_2 + y_1)(-3y_2 - 6y_1 + 4\sqrt{2y_1(y_2 + y_1)})$$

```
Case 6:
> h6 := x-> (1/3)*(y[1]-x)^3+(1/3)*(x-y[1]-y[2])^3-(1/3)*x^3;
```

$$x- > \frac{1}{3}(y_1 - x)^3 + \frac{1}{3}(x - y_1 - y_2)^3 - \frac{1}{3}x^3$$

```
dh6 := diff(h6(x), x): criticalpoints6 := solve(dh6 = 0, x);
```

$$-y_2 + \sqrt{2y_2^2 + 2y_1y_2}, -y_2 - \sqrt{2y_2^2 + 2y_1y_2}$$

```
>h6soln1 := factor(simplify(h6(criticalpoints6[1])));
```

$$\frac{1}{3}y_2(y_2 + y_1)(-3y_1 - 6y_2 + 4\sqrt{2y_2(y_2 + y_1)})$$

```
>h6soln2 := factor(simplify(h6(criticalpoints6[2])));
```

$$-\frac{1}{3}y_2(y_2+y_1)(3y_1+6y_2+4\sqrt{2y_2(y_2+y_1)})$$

Case 7:

```
> h7 := x-> (1/3)*(x-y[1])^3+(1/3)*(y[1]+y[2]-x)^3-(1/3)*x^3;
```

$$x->\frac{1}{3}(x-y_1)^3+\frac{1}{3}(y_1+y_2-x)^3-\frac{1}{3}x^3$$

```
>dh7 := diff(h7(x), x): criticalpoints7 := solve(dh7 = 0, x);
```

$$y_2+\sqrt{-2y_1y_2}, y_2-\sqrt{-2y_1y_2}$$

```
>h7soln1 := factor(simplify(h7(criticalpoints7[1])));
```

$$-\frac{1}{3}y_1y_2(3y_2-3y_1+4\sqrt{-2y_1y_2})$$

```
>h7soln2 := factor(simplify(h7(criticalpoints7[2])));
```

$$\frac{1}{3}y_1y_2(3y_1-3y_2+4\sqrt{-2y_1y_2})$$

Case 8:

```
> h8 := x-> (1/3)*(y[1]-x)^3+(1/3)*(y[1]+y[2]-x)^3-(1/3)*x^3;
```

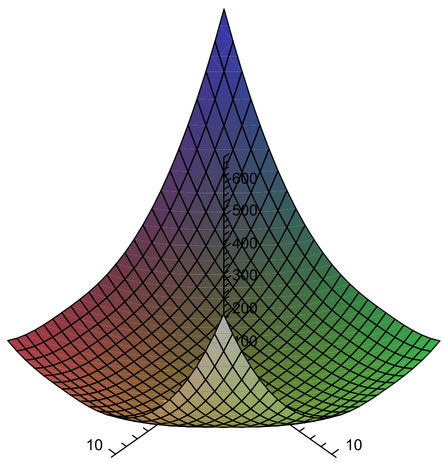
$$x->\frac{1}{3}(y_1-x)^3+\frac{1}{3}(y_1+y_2-x)^3-\frac{1}{3}x^3$$

```
>dh8 := diff(h8(x), x): criticalpoints8 := solve(dh8 = 0, x);
```

$$\frac{1}{3}y_2+\frac{2}{3}y_1+\frac{1}{3}\sqrt{-2y_2^2-2y_1y_2-2y_1^2}, \frac{1}{3}y_2+\frac{2}{3}y_1-\frac{1}{3}\sqrt{-2y_2^2-2y_1y_2-2y_1^2}$$

```
Plotting:
>restart: with(plots):
> f1 := (y1,y2)-> -(1/3)*y1*(y2+y1)*(-6*y1-3*y2+4*sqrt(2)*
sqrt(y1*(y2+y1))):
> f2 := (y1,y2)-> -(1/3)*y2*(y2+y1)*(-3*y1-6*y2+4*sqrt(2)*
sqrt(y2*(y2+y1))):
> f3 := (y1,y2)-> (1/3)*y1*y2*(3*y1-3*y2+4*sqrt(2)*sqrt(-y1
*y2)):
> f4 := (y1,y2)-> -(1/3)*y1*(y2+y1)*(6*y1+3*y2+4*sqrt(2)*
sqrt(y1*(y2+y1))):
> f5 := (y1,y2)-> -(1/3)*y2*(y2+y1)*(3*y1+6*y2+4*sqrt(2)*
sqrt(y2*(y2+y1))):
> f6 := (y1,y2)-> (1/3)*y1*y2*(-3*y1+3*y2+4*sqrt(2)*sqrt(-y1
*y2)):
>fpiece := (y1,y2)-> piecewise(0 <= y1 and -(1/2)*y1 <= y2
and y2 <= y1, f1(y1,y2), 0 <= y2 and -(1/2)*y2 <= y1 and
y1 <= y2, f2(y1,y2), y1 <= 0 and -(1/2)*y1 <= y2 and y2 <=
-2*y1, f3(y1,y2), y1 <= 0 and y1 <= y2 and y2 <= -(1/2)*y1,
```

```
f4(y1,y2), y2 <= 0 and y2 <= y1 and y1 <= -(1/2)*y2,
f5(y1,y2), 0 <= y1 and -2*y1 <= y2 and y2 <= -(1/2)*y1,
f6(y1,y2)):
>plot3d(fpiece(y1,y2), y1=-10..10, y2=-10..10, axes=normal);
```




```
>plots[contourplot](fpiece(y1,y2), y1=-10..10, y2=-10..10,
 contours=100);
```