Convexity of the Proximal Average

by

Jennifer Johnstone

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

BACHELOR OF SCIENCE HONOURS

in

The I. K. Barber School of Arts & Sciences

(Computer Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Kelowna, Canada)

December, 2008

Abstract

The proximal average operator is recognized for its ability to transform two convex functions into another convex function. However, we prove with examples that the proximal average operator does have limitations, with respect to convexity. We also look at the importance of $\lambda \in [0, 1]$ and describe an idea of how to plot the proximal average of two convex functions more efficiently.

Table of Contents

Acknowledgements

To start, the author would like to express her gratitude to her supervisor Dr. Yves Lucet for his constant support and guidance throughout this process. His ideas and topic suggestions made this thesis possible.

The author would also like to thank all those who encouraged her to take on this project and supported her throughout.

Dedication

To my Mother for her continuous encouragement.

Chapter 1

Introduction

The purpose of this thesis is to investigate the convexity of the proximal average and elaborate on some of the properties of the proximal average. In this chapter, we review some of the standard facts on Convex Analysis followed by a review of some of the basics of the proximal average in Chapter 2, in which we also provide the main results of our investigation into the convexity of the proximal average.

1.1 Convex Functions

In this section we recall some of the basics of Convex Analysis to help the reader better understand the results in Chapter 2. We assume that we are working in a real *Hilbert space*, H, which is defined as a complete, real, inner product space. The Hilbert space H is *complete* if all Cauchy sequences in H converge, with respect to the defined norm, and H is an *inner product space* if an inner product exists such that $\|x\| = \sqrt{< x, x >}$, for all $x \in H$. So we define $<\cdot, \cdot>: H \times H \to \mathbb{R}$ such that for all $x, y, z \in H$ and $\lambda, \mu \in \mathbb{R}$ we have

$$\begin{array}{rcl} < x, y > & = & < y, x > \\ < \lambda x + \mu y, z > & = & \lambda < x, z > + \mu < y, z > \\ < x, x > \geq 0 & \text{and} & < x, x >= 0 \text{ iff } x = 0. \end{array}$$

Then working in H we can extend any function $g : \Omega \subset H \to \mathbb{R}$ to $\tilde{g} : H \to \mathbb{R} \cup \{+\infty\}$ with

$$\tilde{g}(x) = \begin{cases} g(x) & \text{if } x \in \Omega, \\ +\infty & \text{if not.} \end{cases}$$

We also define dom $\tilde{g}$ as

$$\text{dom } \tilde{g} := \{x \in H | \tilde{g}(x) < +\infty\}$$

Definition 1.1 (Convex Sets)**.** Let $\Omega \subset H$ then Ω is a *convex set* if for all $x_1 \in \Omega$ and $x_2 \in \Omega$ it contains all points

$$\alpha x_1 + (1-\alpha)x_2, \qquad 0 < \alpha < 1.$$

Definition 1.2 (Convex Functions)**.** Let C be a non-empty convex set in H. A function $g : C \to \mathbb{R}$ is said to be *convex* on C when, for all pairs $(x_1, x_2) \in C \times C$ and all $0 < \alpha < 1$, we have

$$g(\alpha x_1 + (1-\alpha)x_2) \leq \alpha g(x_1) + (1-\alpha)g(x_2).$$

For a convex function, g, we see that the line segment

$$\{(\alpha x_1 + (1-\alpha)x_2, \alpha g(x_1) + (1-\alpha)g(x_2) : \alpha \in [0,1]\}$$

is always above the graph of g, as illustrated in Figure 1.1. On the other hand, in Figure 1.2 we see that this is not the case for all pairs $(x_1, x_2) \in C \times C$ and all $0 < \alpha < 1$.

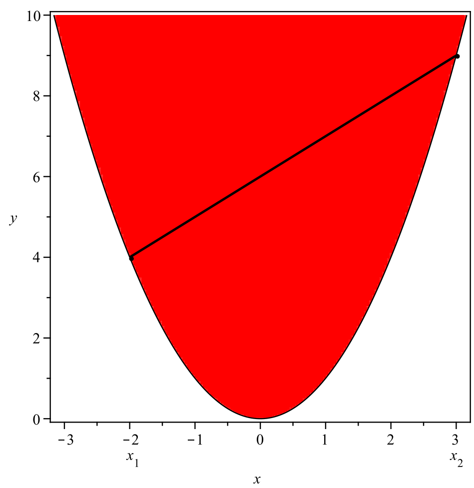

Figure 1.1: For $g = x^2$ we see that the line segment $\{(\alpha x_1 + (1-\alpha)x_2, \alpha g(x_1) + (1-\alpha)g(x_2) : \alpha \in [0,1]\}$ is always above the graph g thus g is a convex function.

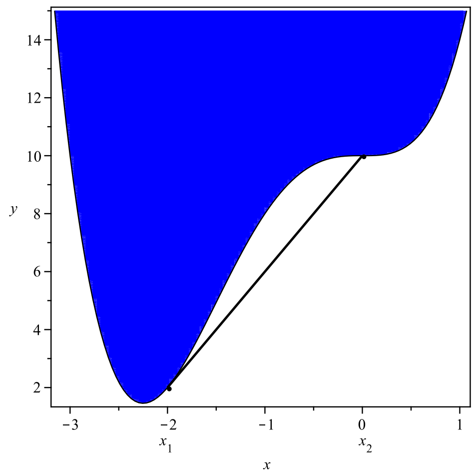

Figure 1.2: For $g = x^4 + 3x^3 + 10$ we see that the line segment $\{(\alpha x_1 + (1-\alpha)x_2, \alpha g(x_1) + (1-\alpha)g(x_2) : \alpha \in [0,1]\}]$ is not above the graph of g at $x = -1$ thus g is a non-convex function.

We will only consider proper functions, since we are not interested in degenerate functions such as $g = +\infty$ everywhere.

Definition 1.3 (Proper)**.** A function g is called *proper* if $g(x) > -\infty$ for all x and $g(x) < +\infty$ for at least one x.

An example of a proper function would be the *indicator function* which is defined on a nonempty set C by

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

Definition 1.4 (Lower Semi-continuous)**.** A function f is said to be *lower semi-continuous* (lsc) at a point $\bar{x} \in \operatorname{dom} f$ if

$$f(\bar{x}) \leq \liminf_{x \to \bar{x}} f(x)$$

While continuous functions are lsc, Figure 1.3 gives an example of a lsc function which is not continuous everywhere.

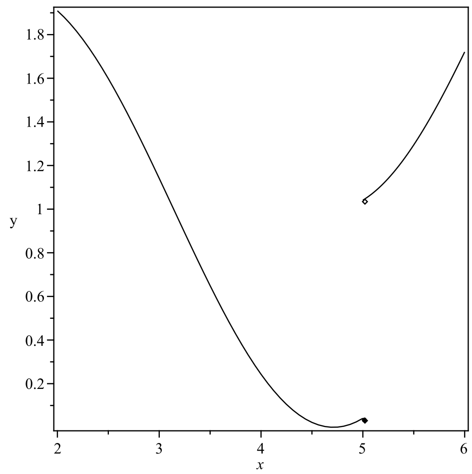

Figure 1.3: An example of a lower semi-continuous function (lsc): it is lsc at x=5 and continuous everywhere else.

1.2 Subdifferential

Since convex functions are not always differentiable we need to introduce the concept of a subgradient.

Definition 1.5 (Subgradients)**.** Let $g : H \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function and let $x \in \operatorname{dom} g$. A vector s in H satisfying

$$g(y) \geq g(x) + < s, y - x > \qquad \forall y \in H$$

is called a *subgradient* of g at x.

Moreover, the set of all subgradients of g at x is called the *subdifferential* of g at x, denoted by $\partial g(x)$, see Figure 1.4 for a graph of a function with some subtangent lines. We note that if g is convex and differentiable the subgradient of g at x is $\bigtriangledown g(x)$ such that $\partial g(x) = \{\bigtriangledown g(x)\}$. An example of when $\partial g(x) = \{\bigtriangledown g(x)\}$ is represented in the tangent of Figure 1.4 when x = 3, as this function is differentiable there.

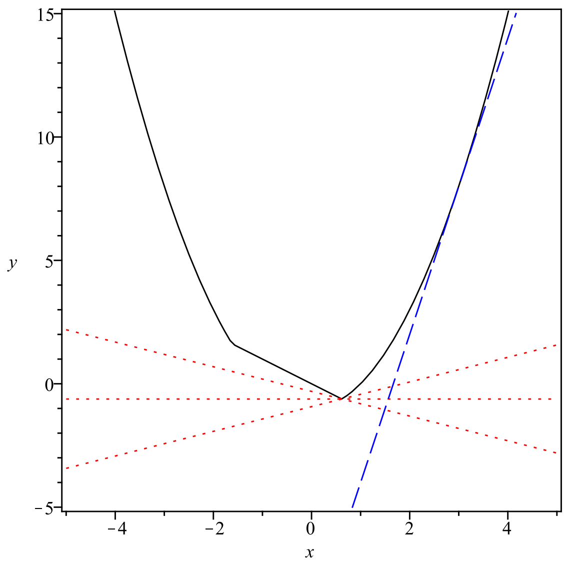

Figure 1.4: An example of a convex function with both subtangents and tangents.

When g is twice continuously differentiable we denote $\bigtriangledown^2 g(x)$ its *Hessian* at x. The Hessian of g is defined as the following symmetric matrix:

$$\bigtriangledown^2 g(x) = \begin{bmatrix} \frac{\partial^2 g(x)}{\partial x_1^2} & \frac{\partial^2 g(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 g(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 g(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 g(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 g(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 g(x)}{\partial x_n \partial x_1} & \frac{\partial^2 g(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 g(x)}{\partial x_n^2} \end{bmatrix}.$$

It is also worth mentioning that if the Hessian of a function exists then we have the following convexity test.

Fact 1.1 (Convexity Test)**.** [7, Theorem 2.69(i)] If we assume that $g : H \to \mathbb{R}$ is a twice continuously differentiable function then g is convex if and only if its Hessian $\bigtriangledown^2 g(x)$ is positive semi-definite for all $x \in H$.

Recall that a Hessian matrix is *positive semi-definite* if and only if all its eigenvalues are greater or equal to zero [6].

1.3 Legendre-Fenchel Conjugate

Now that we have defined a convex, lsc and proper function let

$$X = \{f : H \to \mathbb{R} \cup \{+\infty\} \,|f \text{ is convex, lsc and proper}\}$$

be the set of functions that we are working with for the remainder of this thesis. In this section we define the Fenchel Conjugate that is important in the field of convex analysis.

Definition 1.6 (Legendre-Fenchel Conjugate)**.** The *Legendre-Fenchel Conjugate* (aka convex conjugate) of $f \in X$ is the function $f^* \in X$ defined by

$$f^*(s) := \sup_x \{< s, x > -f(x)\} \qquad \forall s \in H.$$

Furthermore, we note that the *Fenchel Biconjugate Theorem*, as seen in [3, 4], states that

$$f \in X \Longleftrightarrow f^{**} = f.$$

We also define the relative interior of a convex set.

Definition 1.7. The *relative interior* of a convex set $C \subset H$ is the interior of C for the topology relative to the affine hull of C.

We now present the Fenchel Duality Theorem which allows us to solve convex optimization problems.

Theorem 1.1 (Fenchel Duality)**.** *[3] Given two functions f and g in X such that*

$$y = \inf_{x \in H} \{f(x) + g(x)\}$$

is a finite number and assume the relative interiors of dom f and dom g intersect. Then

$$-y = \min_{x^* \in H} [f^*(x^*) + g^*(-x^*)]$$

is actually attained.

Chapter 2

The Proximal Average

From now on we will assume that f_0 and f_1 are in X; λ_0 and λ_1 are two real numbers strictly greater than zero, such that $\lambda_0 + \lambda_1 = 1$; and μ is strictly greater than zero.

2.1 Main Results

We first start by defining the proximal average.

Definition 2.1 (Proximal Average)**.** The λ weighted *proximal average* of f_0 and f_1 with parameter μ is

$$p_\mu(f_0, f_1; \lambda_0, \lambda_1) = \frac{1}{\mu}\left[-\frac{1}{2}\|x\|^2 + \inf_{x_1+x_2=x}\left[\lambda_0\left(\mu f_0\left(\frac{x_0}{\lambda_0}\right) + \frac{1}{2}\left\|\frac{x_0}{\lambda_0}\right\|^2\right)\right.\right.$$
$$\left.\left.+\lambda_1\left(\mu f_1\left(\frac{x_1}{\lambda_1}\right) + \frac{1}{2}\left\|\frac{x_1}{\lambda_1}\right\|^2\right)\right]\right] \quad (2.1)$$

Remark 2.1 (Simplification of p)**.** Since $\lambda_0 + \lambda_1 = 1$ we can re-write equation (2.1) as follows:

$$p_\mu(f_0, f_1; \lambda_0, \lambda_1) = p_\mu(f_0, f_1; (1-\lambda_1), \lambda_1).$$

Then for $\lambda = \lambda_1$ we have

$$p_\mu(f_0, f_1; \lambda) = p_\mu(f_0, f_1; (1-\lambda), \lambda)).$$

Fact 2.1 (Reformulations)**.** [2, Proposition 4.3] By changing variables we see that Equation (2.1) is equivalent to the following:

$$p_\mu(f_0, f_1; \lambda)(x) = \inf_{(1-\lambda)x_0+\lambda x_1=x}\left[(1-\lambda)f_0(x_0) + \lambda f_1(x_1) + \right.$$
$$\left.\frac{1}{\mu}\left((1-\lambda)q(x_0) + \lambda q(x_1) - q(x)\right)\right].$$

where $q(x) = \frac{\|x\|^2}{2}$.

One of the immediate consequences of Definition 2.1, as seen in [2], is that $p_\mu(f_0, f_1; \lambda) = \mu^{-1} p_1(\mu f_0, \mu f_1; \lambda)$. This consequence, in conjunction with the definition of the proximal average, provides us with the following two properties:

$$p_\mu(f_0, f_1; 0) = f_0 \text{ and } p_\mu(f_0, f_1; 1) = f_1.$$

The above properties are similarly seen in [2] for $\mu = 1$. As a visualization we see in Figure 2.1 that $p_\mu(f_0, f_1; \lambda)$ is the conversion of f_0 into f_1 over $\lambda \in [0, 1]$, with constant μ.

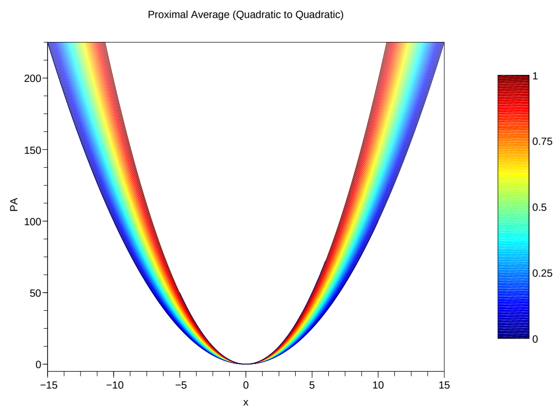

Figure 2.1: Plot of $p_1(x^2, 2x^2; \lambda)$, for $\lambda \in [0, 1]$.

We now state some useful and interesting properties that have be previously discovered.

Fact 2.2 (Domain)**.** [2, Theorem 4.6] We have the following domain property

$$\operatorname{dom} p_\mu(f_0, f_1; \lambda) = (1 - \lambda) \operatorname{dom} f_0 + \lambda \operatorname{dom} f_1.$$

Fact 2.3. [3, Proposition 2.8] Let $f \in X$. Then

$$p_1(f, f^*; \frac{1}{2}) = \frac{1}{2} \|\cdot\|^2 .$$

Fact 2.4 (Fenchel Conjugate of the Proximal Average)**.** [2, Theorem 5.1][3, Fact 2.3]

$$(p_\mu(f_0, f_1; \lambda))^* = p_{\mu^{-1}}(f_0^*, f_1^*; \lambda)$$

2.2 Convexity Results

2.2.1 Convexity of the Proximal Average with respect to x

In this section we will show that the function $\phi(x) := p_\mu(f_0, f_1; \lambda)(x)$, with f_0, f_1, λ and μ fixed, is convex lsc, and proper.

Proposition 2.1. *[2, Corollary 5.3] Assume that $\mu > 0, \lambda \in [0,1]$ and $f_0, f_1 \in X$. Then the function*

$$\phi := x \mapsto p_\mu(f_0, f_1; \lambda)(x)$$

is convex for all $x \in$ dom p.

Proof. Applying Fact 2.4 twice, we see that

$$\begin{aligned} (p_\mu(f_0, f_1; \lambda))^{**} &= (p_{\mu^{-1}}(f_0^*, f_1^*; \lambda))^* \\ &= p_{(\mu^{-1})^{-1}}(f_0^{**}, f_1^{**}; \lambda) \\ &= p_\mu(f_0, f_1; \lambda). \end{aligned}$$

Hence, by the Fenchel Biconjugate Theorem we have that $\phi(x)$ is convex. □

2.2.2 Convexity of the Proximal Average with respect to λ

In this section we will show that the function $\phi(\lambda) := p_\mu(f_0, f_1; \lambda)(x)$ is convex, with f_0, f_1, μ and x fixed. In order to show that $\phi(\lambda)$ is convex we need to first recall some properties of marginal and perspective functions. We start with a property pertaining to marginal functions.

Fact 2.5. [8, Theorem 2.1.3 (v)] Let $\Omega, O \subset H$. If $g := \Omega \times O \mapsto \mathbb{R} \cup \{+\infty\}$ is convex then the marginal function γ associated to g is convex where

$$\gamma : O \mapsto \mathbb{R},\ \gamma(y) := \inf_{x \in X} g(x, y)$$

Before we state a useful property of perspective functions we first need to define them.

Definition 2.2 (Perspective function)**.** The *perspective function* of $f : \mathbb{R}^n \to \mathbb{R} \cup +\infty$ is the function from $\mathbb{R} \times \mathbb{R}^n$ to $\mathbb{R} \cup \{+\infty\}$ given by

$$\operatorname{Persp}(f)(u, x) = \begin{cases} uf\left(\frac{x}{u}\right) & \text{if } u > 0 \\ +\infty & \text{if not.} \end{cases}$$

Fact 2.6. [5, Proposition IV.2.2.1] [8, Section 1.2] If $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper convex, then its perspective function Persp (f) is proper and convex on $\mathbb{R} \times \mathbb{R}^n$.

Now we prove our main result for this section.

Proposition 2.2. *[1, Proposition 6.1] Assume that the functions f_0, f_1 are in X, $x \in$ dom $p_\mu(f_0, f_1; \lambda)$, and $\mu > 0$. Then the function*

$$\phi : \lambda \mapsto p_\mu(f_0, f_1; \lambda)(x)$$

is convex for $\lambda \in [0, 1]$.

Proof. Using Definition 2.1 and Remark 2.1, with $x_1 = x - x_0$, we have

$$\phi : \lambda \mapsto p_\mu(f_0, f_1; \lambda)(x) = \inf_{x_0} [g(\lambda, x_0)] - \frac{q(x)}{\mu}$$

where $q(x) = \frac{\|x\|^2}{2}$ and

$$g(\lambda, x_0) = (1 - \lambda)(\mu f_0 + q)\left(\frac{x_0}{1 - \lambda}\right) + \lambda(\mu f_1 + q)\left(\frac{x - x_0}{\lambda}\right).$$

Now in order to apply Fact 2.5 we need to first show that g is convex as a function of both x_0 and λ. We note that

$$\phi(\lambda) = \frac{1}{\mu} \min_{x_0} [\mathrm{Persp}(\mu f_0 + q)(1 - \lambda, x_0) + \mathrm{Persp}(\mu f_1 + q)(\lambda, x - x_0)] - \frac{q(x)}{\mu}$$

based on Definition 2.2. So using Fact 2.6 and [5, Proposition IV.2.1.5] (the composition of a convex function with an affine mapping is convex), the function g is convex. Now, using the convexity of the functions f_0 and f_1 we have that the function g is bounded below for each value of λ. Hence, Fact 2.5 applies thereby showing that ϕ is convex.

□

2.2.3 Convexity of the Proximal Average with respect to μ

In this section we will show that the function $\phi(\mu) := p_\mu(f, \lambda)(x)$ is convex, with $f_0, f_1 \in X$, $\lambda_0, \lambda_1 \in [0, 1]$ such that $\lambda_0 + \lambda_1 = 1$ and x fixed.

Proposition 2.3. *[1, Proposition 5.7] Assume $\mu > 0, \lambda_0 + \lambda_1 = 1, \lambda_i \geq 0, f_i \in X$ for $i = 1, 2$ and take $x \in$ dom $p_\mu(f_0, f_1; \lambda_0, \lambda_1) = \lambda_0$ dom $f_0 + \lambda_1$ dom f_1. Then the function*

$$\phi : \mu \mapsto p_\mu(f_0, f_1; \lambda_0, \lambda_1)(x)$$

is convex on $]0, +\infty[$.

Proof. We begin by substituting $x_1 = (x - \lambda_0 x_0)/\mu$ into Equation (2.1) and use $\lambda_0 + \lambda_1 = 1$ to obtain

$$\phi(\mu) = \inf_{x_0} g(\mu, x_0)$$

where

$$g(\mu, x_0) = \lambda_0 f_0(x_0) + \lambda_1 f_1\left(\frac{x - \lambda_0 x_0}{\lambda_1}\right) + \frac{\lambda_0}{2\mu}\left\|x_0 - x\right\|^2.$$

Now we see that the function g is lower bounded on any set $\{\mu\} \times H$ for any $\mu > 0$, and is convex (as a composition of convex functions). So, Fact 2.5 applies thereby proving that ϕ is convex.

□

2.2.4 Further Investigation of the Proximal Average

In this section we present our results on the join convexity of the proximal average. We begin with a computation of the proximal average that will be used to build our examples.

Lemma 2.1 (Proximal Average of energy functions)**.** *[2, Example 4.5] Let $f_0 = \alpha_0 q$ and $f_1 = \alpha_1 q$ where α_0 and α_1 are strictly positive numbers and $q(x) = \left\|x\right\|^2/2$. Then*

$$p_\mu(f_0, f_1; \lambda) = \left(\left(\frac{(1-\lambda)}{\alpha_0 - \mu^{-1}} + \frac{\lambda}{\alpha_1 - \mu^{-1}}\right)^{-1} - \mu^{-1}\right) q.$$

Proof. For $q(x) = \left\|x\right\|^2/2$ in Remark 2.1 we see that using Fact 2.4 and some basic properties of conjugacy, namely $(\alpha q)^* = q^*(\alpha\cdot)$, we have

$$\begin{aligned}
p_{\mu^{-1}}(f_0, f_1; \lambda) &= ((1-\lambda)(\alpha_0 q + \mu q)^* + \lambda(\alpha_1 q + \mu q)^*)^* - \mu q \\
&= \left(\frac{1-\lambda}{\alpha_0 + \mu} q + \frac{\lambda}{\alpha_1 + \mu} q\right)^* - \mu q \\
&= \left(\frac{1-\lambda}{\alpha_0 + \mu} q + \frac{\lambda}{\alpha_1 + \mu}\right)^{-1} q - \mu q.
\end{aligned}$$

Thus,

$$p_\mu(f_0, f_1; \lambda) = \left(\left(\frac{1-\lambda}{\alpha_0 - \mu^{-1}} + \frac{\lambda}{\alpha_1 - \mu^{-1}}\right)^{-1} - \mu^{-1}\right) q.$$

□

We now present the main results for this section.

Proposition 2.4. *The following functions are not always convex:*

$$\begin{aligned}
\phi_{x\lambda} : (x, \lambda) \quad &\mapsto p_\mu(f_0, f_1; \lambda)(x) \\
\phi_{x\mu} : (x, \mu) \quad &\mapsto p_\mu(f_0, f_1; \lambda)(x) \\
\phi_{\lambda\mu} : (\lambda, \mu) \quad &\mapsto p_\mu(f_0, f_1; \lambda)(x)
\end{aligned}$$

when f_0, f_1 are two convex lsc proper functions, $\lambda \in [0, 1]$ and $\mu > 0$.

Proof. We need to show that $\phi_{x\lambda}, \phi_{x\mu}$ and $\phi_{\lambda\mu}$ are not always convex. So consider the following quadratic example:

Let f_0 and f_1 be in X such that

$$\begin{aligned}
f_0 \quad &= \alpha_0 q \\
f_1 \quad &= \alpha_1 q
\end{aligned}$$

where α_0 and α_1 are strictly positive numbers and q is the quadratic energy function: $q(x) = \frac{1}{2}x^2$ (when $x \in \mathbb{R}$). Then using Lemma 2.1 we have

$$p_\mu(f_0, f_1; \lambda) = \left(\left(\frac{(1-\lambda)}{\alpha_0 - \mu^{-1}} + \frac{\lambda}{\alpha_1 - \mu^{-1}} \right)^{-1} - \mu^{-1} \right) q.$$

To show that $\phi_{x\lambda}$ (resp. $\phi_{x\mu}$ and $\phi_{\lambda\mu}$) is not convex we use Fact 1.1 and show that the Hessian of $p_\mu(f_0, f_1; \lambda)$ is not positive semi-definite, for μ (resp. λ and x) constant.

To start, consider $p_\mu(f_0, f_1; \lambda)$ with $\mu = 1$, so as to show that $\phi_{x,\lambda}(x, \lambda)$ is not convex. Now,

$$p_1(f_0, f_1; \lambda)(x) = \frac{1}{2} \left(\left(\frac{(1-\lambda)}{\alpha_0 - 1} + \frac{\lambda}{\alpha_1 - 1} \right)^{-1} - 1 \right) x^2.$$

Then with $\alpha_0 = 2$ and $\alpha_1 = 4$ we have

$$\phi_{x\lambda}(x, \lambda) = p_1(f_0, f_1; \lambda)(x) = \frac{1}{2} \left(\left((1-\lambda) + \frac{\lambda}{3} \right)^{-1} - 1 \right) x^2$$

and the Hessian of $\phi_{x\lambda}$, at $x = 2$ and $\lambda = \frac{1}{2}$, is

$$H_{x\lambda} = \begin{bmatrix} \frac{1}{2} & 3 \\ 3 & 6 \end{bmatrix}.$$

The determinant of $H_{x\lambda}$ is -6 which is enough to show that $H_{x\lambda}$ is not positive semi-definite, since -6 implies that one of the eigenvalues of the determinant must be negative. Hence $\phi_{x\lambda}(x,\lambda)$ is not convex.

Similarly, we can show that $\phi_{x,\mu}$ is not always convex by holding λ constant. So, let $\lambda = \frac{1}{2}$ then

$$p_\mu(f_0, f_1; \frac{1}{2})(x) = \frac{1}{2}\left(\left(\frac{\frac{1}{2}}{\alpha_0 - \mu^{-1}} + \frac{\frac{1}{2}}{\alpha_1 - \mu^{-1}}\right)^{-1} - \mu^{-1}\right) x^2$$

and when $\alpha_0 = 2$ and $\alpha_1 = 4$ we have

$$\phi_{x\mu}(x,\mu) = p_\mu(f_0, f_1; \frac{1}{2})(x) = \frac{1}{2}\left(\left(\frac{\frac{1}{2}}{2 - \mu^{-1}} + \frac{\frac{1}{2}}{4 - \mu^{-1}}\right)^{-1} - \mu^{-1}\right) x^2.$$

Now, the Hessian of $\phi_{x\mu}$, at $x = 2$ and $\mu = 1$, is

$$H_{x\mu} = \begin{bmatrix} \frac{1}{2} & \frac{9}{2} \\ \frac{9}{2} & -\frac{19}{2} \end{bmatrix}.$$

The determinant of $H_{x\mu}$ is -25 so $\phi_{x\mu}(x,\mu)$ is not convex.

Finally, to show that $\phi_{\lambda\mu}(\lambda,\mu)$ is not always convex we consider $p_\mu(f,\lambda)$ with $f_0(x) = 2q(x)$ and $f_1(x) = 4q(x)$, evaluated at $x = 2$. Then

$$\phi_{\lambda\mu}(\lambda,\mu) = p_\mu(f_0, f_1; \lambda)(2) = 2\left(\left(\frac{(1-\lambda)}{2 - \mu^{-1}} + \frac{\lambda}{4 - \mu^{-1}}\right)^{-1} - \mu^{-1}\right)$$

and the Hessian of $\phi_{\lambda\mu}$, at $\mu = 1$ and $\lambda = \frac{1}{2}$, is

$$H_{\lambda\mu} = \begin{bmatrix} 6 & 1 \\ 1 & -\frac{19}{2} \end{bmatrix}.$$

The determinant of $H_{\lambda\mu}$ is -58 so $\phi_{\lambda\mu}(\lambda,\mu)$ is not convex.

Altogether, we proved that each of $\phi_{x\lambda}, \phi_{x\mu}$ and $\phi_{\lambda\mu}$ are not always convex.

□

Corollary 2.1. *The following function is not always convex:*

$$\phi : (x, \lambda, \mu) \mapsto p_\mu(f,\lambda)(x)$$

where $f = (f_0, f_1) \in X \times X$, $\lambda \in [0,1]$ *and* $\mu > 0$.

Proof. Corollary 2.1 is a direct result of Proposition 2.1. □

Proposition 2.5. *The following function is not always convex:*

$$\phi : (f_0, f_1) \mapsto p_\mu(f_0, f_1; \lambda)$$

where $f_0, f_1 \in X$, $\lambda \in [0, 1]$ *and* $\mu > 0$.

Proof. In order to find a function $\phi(f_0, f_1)$ which is not always convex we need to show, by Definition 1.2, that there exists an $x \in \mathbb{R}$ such that

$$p_\mu(\tau(f_0, f_1) + (1-\tau)(g_0, g_1), \lambda)(x) > \tau p_\mu((f_0, f_1), \lambda) + (1-\tau)p_\mu((g_0, g_1), \lambda)(x)$$

where f_0, f_1, g_0 and g_1 are all in X and $\lambda, \tau \in [0, 1]$. So, let

$$\begin{aligned} f_0 &= \alpha_0 q \\ f_1 &= \alpha_1 q \\ g_0 &= \beta_0 q \\ g_1 &= \beta_1 q \end{aligned}$$

where q is the quadratic energy function $q(x) = \frac{1}{2}x^2$, when $x \in \mathbb{R}$. Then for $\alpha_0 = 1, \alpha_1 = 2, \beta_0 = 3$ and $\beta_1 = 4$ the left-hand side is

$$p_\mu(\tau(f_0, f_1) + (1-\tau)(g_0, g_1), \lambda) = p_\mu((\tau q + (1-\tau)3q, \tau 2q + (1-\tau)4q), \lambda).$$

Then with $\mu = 5, \tau = \frac{1}{2}$ and $\lambda = \frac{1}{2}$, we have

$$p_\mu((\tau q + (1-\tau)3q, \tau 2q + (1-\tau)4q), \lambda) \quad = p_5\left(2q, 3q, \tfrac{1}{2}\right) = \tfrac{65}{27}q.$$

On the other hand, the right-hand side is

$$\begin{aligned} \tau p((f_0, f_1), \lambda) + \tau p((g_0, g_1), \lambda) &= \tau p((q, 2q), \lambda) + (1-\tau)p((3q, 4q), \lambda) \\ &= \frac{1}{2}p((q, 2q), \frac{1}{2}) + \frac{1}{2}p((3q, 4q), \frac{1}{2}) \\ &= \frac{1}{2}\left(\frac{23}{17}q\right) + \frac{1}{2}\left(\frac{127}{37}q\right) \\ &= \frac{1505}{629}q. \end{aligned}$$

Therefore,

$$p_\mu(\tau(f_0, f_1) + (1-\tau)(g_0, g_1), \lambda)(x) > \tau p_\mu((f_0, f_1), \lambda) + (1-\tau)p_\mu((g_0, g_1), \lambda)(x)$$

since

$$\frac{65}{27} \approx 2.41 > 2.39 \approx \frac{1505}{629}.$$

Hence, $\phi(f_0 f_1)$ is not always convex.

$\square$

To conclude this section, we show that extending the proximal average to $\lambda \in \mathbb{R}$ does not provide a useful tool as the following example shows the infimum may no longer be attained.

Example 2.1. Using Definition 2.1, with Remark 2.1, $\mu = 1$ and $x_1 = x - x_0$, we have

$$\begin{aligned} p_1(f_0, f_1; \lambda) = -\frac{\|x\|^2}{2} + \inf_{x_0} \Bigg[\lambda \left(f_0 \left(\frac{x_0}{(1-\lambda)} \right) + \frac{1}{2} \left\| \frac{x_0}{(1-\lambda)} \right\|^2 \right) \\ + \lambda \left(f_1 \left(\frac{x - x_0}{\lambda} \right) + \frac{1}{2} \left\| \frac{x - x_0}{\lambda} \right\|^2 \right) \Bigg] . \qquad (2.2) \end{aligned}$$

Then in Equation (2.2) for $f_0(x) = \alpha_0 x + \beta$ and $f_1(x) = \alpha_1 x + \beta_1$, with $x \in \mathbb{R}$, we have

$$-\frac{x^2}{2} + \inf_{x_0} \left[(1-\lambda) \left(\alpha_0 \frac{x_0}{(1-\lambda)} + \beta_0 + \frac{x_0^2}{2(1-\lambda)^2} \right) + \lambda \left(\alpha_1 \frac{x - x_0}{\lambda} + \beta_1 + \frac{(x-x_0)^2}{2\lambda^2} \right) \right]$$

which can be further reduced to

$$-\frac{x^2}{2} + \inf_{x_0} \left[\alpha_0 x_0 + \beta_0 - \lambda\beta_0 + \frac{x_0^2}{2(1-\lambda)} + \alpha_1 (x - x_0) + \lambda\beta_1 + \frac{x^2}{2\lambda} - \frac{x_0}{\lambda} + \frac{x_0^2}{2\lambda} \right] \quad (2.3)$$

Then the dominating term of Equation (2.3) is

$$\begin{aligned} \frac{x_0^2}{2(1-\lambda)} + \frac{x_0^2}{2\lambda)} &= \frac{x_0^2 \lambda + x_0^2 (1-\lambda)}{2\lambda(1-\lambda)} \\ &= \frac{x_0^2}{2\lambda(1-\lambda)} \end{aligned}$$

which requires $2\lambda(1-\lambda) > 0$ for the infimum to be finite. Thus we require $\lambda \in]0, 1[$.

Chapter 3

Plotting the Proximal Average

Given two functions f_0 and f_1 we want to plot the proximal average of these functions for $\lambda \in [0, 1]$.

3.1 New Plotting function

Currently the Computation Convex Analysis Numerical Library has Scilab functions that can be used to plot the proximal average of two piecewise linear-quadratic (PLQ) functions, namely plq_plotpa. A PLQ function is defined on a set of disjoint domains where for each domain the function is either a linear or quadratic polynomial. The plq_plotpa function is defined to plot the proximal average curve based on λ, as seen in Figure 3.1. However, plq_plotpa is not always very efficient as it may plot the same proximal average curve for a given λ value when the λ values are close together.

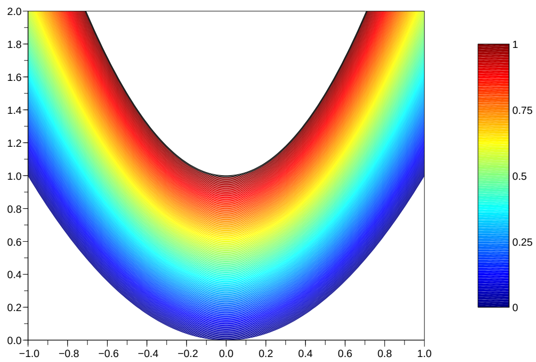

Figure 3.1: Plot of $p_1(x^2, 2x^2 + 1; \lambda)$, for $\lambda \in [0, 1]$, done in Scilab using the original command plq_plotpa.

So, we created a new plotting function, plot_proxavg, that plots each pixel of the proximal average given two PLQ functions, f_0 and f_1. That is to say for each (x, y) we assign the appropriate λ value where

$$y = p_1(f_0, f_1; \lambda)(x), \tag{3.1}$$

as seen in Figure 3.2.

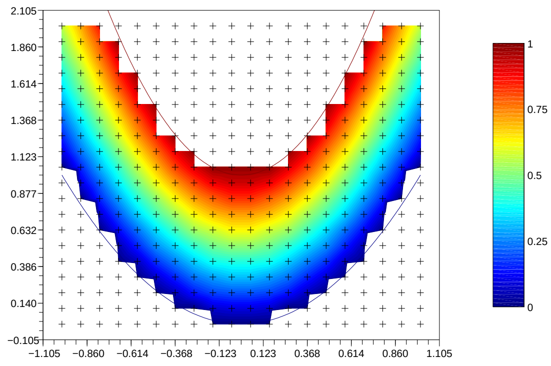

Figure 3.2: Plot of $p_1(x^2, 2x^2 + 1; \lambda)$, for $\lambda \in [0, 1]$, done in Scilab using the new command plot_proxavg .

The plot_proxavg command requires solving Equation (3.1) for λ. This however is not a simple task given the equation of the proximal average, which is why we have taken a different approach. The approach that we have taken assumes f_0, f_1 are convex and $f_0 \leq f_1$ so that λ is always increasing. Then for every x value we can determine λ starting at $y_0 = f_0(x)$ for $\lambda = 0$ and going to $y_1 = f_1(x)$ for $\lambda = 1$, since

$$p_1(f_0, f_1; 0)(x) = f_0(x) \text{ and } p_1(f_0, f_1; 1)(x) = f_1(x).$$

The λ values in between y_0 and y_1, for each (x, y) where $y_0 < y < y_1$, can be determined by incrementally increasing λ until it generates a *newy* value that is as close to y as possible.

Chapter 4

Conclusion

We have answered many of the remaining questions pertaining to the proximal average including showing that

$$(x, \lambda, \mu, f) \mapsto p_\mu(f, \lambda)(x)$$

is always convex in x, λ and μ but not always convex in $(x, \lambda), (x, \mu)$, (λ, μ) and (f_0, f_1). Our plotting algorithm needs further refinement. It currently uses a linear search to compute λ. Implementing a binary search would reduce the computation time. Furthermore, the bounds used in the binary search could be improved by using the fact that the function $\lambda \mapsto p(f_0, f_1; \lambda)(x)$ is convex.

Future work in this area may focus on defining a partial differential equation equivalent to the proximal average. It is believed that one exists as the proximal average can be described as the curve evolution from f_0 to f_1 over λ [1].

Bibliography

[1] Heinz H. Bauschke, Rafal Goebel, Yves Lucet, and X. Wang. The proximal average: applications, extensions and computation. 2008.

[2] Heinz H. Bauschke, Rafal Goebel, Yves Lucet, and X. Wang. The proximal average: Basic theory. *SIAM J. Optim.*, 2008.

[3] Heinz H. Bauschke, Yves Lucet, and Michael Trienis. How to transform one convex function continuously into another. *SIAM Rev.*, 50(1):115–132, July 2008.

[4] Jonathan M. Borwein and Adrian S. Lewis. *Convex Analysis and Nonlinear Optimization.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 3. Springer-Verlag, New York, 2000. Theory and examples.

[5] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms*, volume 305–306 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1993. Vol I: Fundamentals, Vol II: Advanced theory and bundle methods.

[6] D. C. Lay. *Linear Algebra and Its Applications.* Addison-Wesley Longman Inc., Reading, Massachusetts, second edition, 2000.

[7] A. Ruszczyński. *Nonlinear Optimization.* Princeton University Press, Princeton, New York, 2006.

[8] C. Zălinescu. *Convex Analysis in General Vector Spaces.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. World Scientific, New Jersey, 2002.