Journal of Convex Analysis
Volume 15 (2008), No. 1, 179–190

Self-Dual Smoothing of Convex and Saddle Functions

Rafal Goebel
Research initiated at the University of Washington, Seattle, USA
Current address: 3518 NE 42 St., Seattle, WA 98105, USA
rafal.k.goebel@gmail.com

Received: August 6, 2006
Revised manuscript received: March 12, 2007

It is shown that any convex function can be approximated by a family of differentiable with Lipschitz continuous gradient and strongly convex approximates in a "self-dual" way: the conjugate of each approximate is the approximate of the conjugate of the original function. The approximation technique extends to saddle functions, and is self-dual with respect to saddle function conjugacy and also partial conjugacy that relates saddle functions to convex functions.

Keywords: Convex functions, approximation, Moreau envelopes, duality, saddle functions

1991 Mathematics Subject Classification: 52A41, 90C25, 90C59, 90C46, 26B25

1. Introduction

Numerous techniques exist for approximating a nondifferentiable convex function f with convex functions of different levels of regularity. The technique of Moreau [14], popular in convex analysis and optimization due to its relationship to proximal mappings and Yosida regularization of the subdifferential, yields a family of differentiable Moreau envelopes $e_\lambda f$, $\lambda > 0$, with the property (among other ones which we list in Section 2.1) that the convex conjugate of $e_\lambda f$ is the convex conjugate of f plus a quadratic function. In particular, the guaranteed differentiability of $e_\lambda f$ may be lost when passing to a conjugate. A similar phenomenon occurs for the "rolling a ball under the graph of f" smoothing technique of [22], where, thanks to the inf-convolution structure of the smoothing, the conjugate function can be found explicitly. For other smoothing techniques, say that of [25], that resembles an inf-convolution but where the quadratic functions used in Moreau envelopes are replaced by entropy-like distances, or [8], where integral convolutions are used, the conjugate function becomes hard to track. Smoothing techniques targetting particular classes of convex functions or optimization problems can be found in [23, 12] and [6, 7].

In this note, relying to an extent on the Moreau envelope, we propose a smoothing technique that, given any convex function f, yields differentiable approximates $s_\lambda f$, $\lambda \in (0,1)$, not only such that their conjugates are also differentiable, but in fact

$$(s_\lambda f)^* = s_\lambda f^*.$$

That is, the smoothing technique is "self-dual": the conjugate of the smoothed f is the smoothed conjugate of f. Furthermore, the smoothing technique extends to saddle functions, convex in some arguments, concave in the others. For such functions, it is self-dual with respect to saddle function conjugacy. Finally, in an appropriate sense, the

ISSN 0944-6532 / $ 2.50

smoothing is self-dual with respect to partial conjugacy that relates convex functions to saddle functions.

2. The case of convex functions

2.1. Preliminaries

Throughout the paper, X is a Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$, and $\overline{I\!R} = [-\infty, +\infty]$. Let $f : X \to \overline{I\!R}$ be a convex, lower semicontinuous (lsc), and proper (not identically $+\infty$ and never $-\infty$) function. Given f and $\lambda > 0$, the *Moreau envelope* $e_\lambda f$ of f is defined by

$$e_\lambda f(x) = \inf_{u\in X}\left\{f(u) + \frac{1}{2\lambda}\|x-u\|^2\right\}. \tag{1}$$

That is, $e_\lambda f$ is the *inf-convolution* of f and the function $x \mapsto \frac{1}{2\lambda}\|x\|^2$. For convenience, we will sometimes use the notation $j(x) = \frac{1}{2}\|x\|^2$ and write $\#$ for the operation of inf-convolution, so that $e_\lambda f = f \# \lambda^{-1} j$.

For any $\lambda > 0$, the envelope function $e_\lambda f$ is finite-valued, convex, continuous, and Frechet differentiable, with $\nabla e_\lambda f$ Lipschitz continuous with constant $1/\lambda$. (For second-order properties of $e_\lambda f$ see [13].) These, and other properties of $e_\lambda f$ will be reflected in the properties of the smoothing operation in Section 2.2. We add that the envelope functions $e_\lambda f$ are pointwise convergent and Mosco-epiconvergent to f as $\lambda \searrow 0$.

Given any function $g : X \to \overline{I\!R}$, its *convex conjugate* $g^* : X \to \overline{I\!R}$ is defined by

$$g^*(y) = \sup_{x\in X}\{\langle y, x\rangle - g(x)\}.$$

For f as above, f^* is a convex, lower semicontinuous, and proper function, and $(f^*)^* = f$. For the function j and $\lambda > 0$, we have $(\lambda^{-1}j)^* = \lambda j$. Finally, $(e_\lambda f)^* = (f \# \lambda^{-1}j)^* = f^* + \lambda j$, and symmetrically, $(f + \lambda j)^* = f^* \# \lambda^{-1} j$.

For more background and details, see [1], in particular Chapter 3, Proposition 3.3, Theorems 3.20 and 3.24, or [5], Chapter 2, Theorem 2.3 and Corollary 2.3. For the case of $X = I\!R^n$, consult [20], Theorem 2.26, Proposition 7.4, and Example 11.26.

2.2. Smoothing of convex functions

Definition 2.1. Given a convex, lower semicontinuous, and proper function $f : X \to \overline{I\!R}$ and any $\lambda \in (0, 1)$, the function $s_\lambda f : X \to I\!R$ is defined by

$$s_\lambda f(x) = \left(1-\lambda^2\right) e_\lambda f(x) + \frac{\lambda}{2}\|x\|^2. \tag{2}$$

The function $s_\lambda f$ inherits differentiability from the Moreau envelope $e_\lambda f$. We state this, and other properties of $s_\lambda f$ below, in Lemma 2.3. First, we note the most striking property of the operation defining $s_\lambda f$: it is symmetric with respect to convex conjugacy.

Theorem 2.2. *For any convex, lsc, and proper $f : X \to \overline{I\!R}$ and any $\lambda \in (0, 1)$,*

$$(s_\lambda f)^* = s_\lambda(f^*). \tag{3}$$

This was originally stated in [9] and shown via a direct calculation. It is also a special case of Theorem 3.2 (the proof of which simplifies to the direct proof of Theorem 2.2). An alternate proof, via Proposition 2.4, was suggested to the author by Stephen Simons [24].

To illustrate where the symmetry in (3) is coming from, we note that $\nabla s_\lambda f$ has the following property:

$$\operatorname{gph} \nabla s_\lambda f = \begin{bmatrix} I & \lambda I \\ \lambda I & I \end{bmatrix} \operatorname{gph} \partial f. \tag{4}$$

(Above, $\operatorname{gph} \partial f$ denotes the graph of the subdifferential mapping ∂f of f, etc., and I is the identity mapping from X to X.) Indeed, for any convex, lsc, and proper function g,

$$\operatorname{gph} \partial (g + \lambda j) = \begin{bmatrix} I & 0 \\ \lambda I & I \end{bmatrix} \operatorname{gph} \partial g, \qquad \operatorname{gph} \nabla e_\lambda g = \begin{bmatrix} I & \lambda I \\ 0 & I \end{bmatrix} \operatorname{gph} \partial g,$$

$$\operatorname{gph} \partial \left((1-\lambda^2)g\right) = \begin{bmatrix} I & 0 \\ 0 & (1-\lambda^2)I \end{bmatrix} \operatorname{gph} \partial g,$$

and so

$$\operatorname{gph} \partial s_\lambda f = \begin{bmatrix} I & 0 \\ \lambda I & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & (1-\lambda^2)I \end{bmatrix} \begin{bmatrix} I & \lambda I \\ 0 & I \end{bmatrix} \operatorname{gph} \partial f = \begin{bmatrix} I & \lambda I \\ \lambda I & I \end{bmatrix} \operatorname{gph} \partial f.$$

Now, since $\operatorname{gph} \partial g^* = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \operatorname{gph} \partial g$, we have, from (4), that

$$\begin{aligned} \operatorname{gph} \partial (s_\lambda f)^* &= \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \operatorname{gph} \partial s_\lambda f = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} I & \lambda I \\ \lambda I & I \end{bmatrix} \operatorname{gph} \partial f \\ &= \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} I & \lambda I \\ \lambda I & I \end{bmatrix} \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \operatorname{gph} \partial f^* = \begin{bmatrix} I & \lambda I \\ \lambda I & I \end{bmatrix} \operatorname{gph} \partial f^*. \end{aligned}$$

So, the graphs of $\partial s_\lambda f$ and of $\partial (s_\lambda f)^*$ are obtained from those of ∂f, ∂f^* via the same operation. (This effectively justifies (3) up to a constant of integration.)

Lemma 2.3. *For any convex, lsc, and proper $f : X \to \overline{\mathbb{R}}$ and any $\lambda \in (0,1)$,*

(*a*) *$s_\lambda f$ is strongly convex with constant λ;*

(*b*) *$s_\lambda f$ is Frechet differentiable, with the gradient Lipschitz continuous with constant $1/\lambda$ and given by*

$$\nabla s_\lambda f(x) = \frac{1-\lambda^2}{\lambda}\left(x - P_\lambda f(x)\right) + \lambda x,$$

where $P_\lambda f(x) = \arg\min_u \left\{ f(u) + \frac{1}{2\lambda}\|x-u\|^2 \right\}$ is the proximal mapping for f;

(*c*) *$\arg\min s_\lambda f$ is a singleton, and equals x_λ if and only if $x_\lambda = (1-\lambda^2)\, P_\lambda f(x_\lambda)$.*

Furthermore, as $\lambda \searrow 0$, $s_\lambda f$ converge to f pointwise and Mosco-epigraphically, and if $\arg\min f \neq \emptyset$, then $\lim x_\lambda = x_0$, where x_0 is the unique element of $\arg\min f$ of minimal norm.

Proof. The function $s_\lambda f - \lambda j = (1-\lambda^2)e_\lambda f$ is convex, so $s_\lambda f$ is strongly convex with constant λ. Differentiability, Lipschitz property of $\nabla s_\lambda f$ and the formula it follow directly from the properties of $e_\lambda f$.

The function $s_\lambda f$ is continuous and coercive, so $\arg\min s_\lambda f$ is nonempty. Strong, hence strict, convexity implies that this set must be a singleton, say x_λ. Since $s_\lambda f$ is convex and differentiable, $\nabla s_\lambda f(x_\lambda) = 0$, and (b) yields the desired condition. The Mosco-epigraphical and pointwise convergence of $s_\lambda f$ to f comes from such convergence of $e_\lambda f$ to f and continuity of j.

Now assume that $\arg\min f \neq \emptyset$. By convexity of f it is a convex set. Therefore, it contains a unique element of minimal norm, say x_0. For any $x_\lambda \in \arg\min s_\lambda f$ we have

$$\min s_\lambda f = s_\lambda f(x_\lambda) \leq s_\lambda f(x_0) = (1-\lambda^2)\min f + \frac{\lambda}{2}\|x_0\|^2$$

and also

$$\begin{aligned} \min s_\lambda f &= (1-\lambda^2)e_\lambda f(x_\lambda) + \frac{\lambda}{2}\|x_\lambda\|^2 \geq (1-\lambda^2)e_\lambda f(x_0) + \frac{\lambda}{2}\|x_\lambda\|^2 \\ &= (1-\lambda^2)\min f + \frac{\lambda}{2}\|x_\lambda\|^2. \end{aligned}$$

Thus $\|x_\lambda\| \leq \|x_0\|$, so any accumulation point of x_λ must have the norm less or equal to $\|x_0\|$. By the epiconvergence of $s_\lambda f$ to f, accumulation points of elements of $\arg\min s_\lambda f$), as $\lambda \searrow 0$, belong to $\arg\min f$ ([1, Proposition 2.9] or [20, Theorem 7.31] for the case of $X = I\!R^n$). Any such accumulation point must have the norm equal to at least $\|x_0\|$, and thus, it must actually equal x_0. □

The symmetry described by (3) turns out to be a special case of a more general fact about the structure of conjugates of functions that are a combination of Moreau envelopes and quadratic functions. That is, we have the following result:

Proposition 2.4. *Let $\gamma, \beta > 0$, $\alpha = \beta/(\beta+\gamma)$. Then for any convex, lsc, and proper $f : X \to \overline{I\!R}$,*

$$[(\alpha f \# \beta j) + \gamma j]^* = (\alpha f^* \# \rho j) + \sigma j, \tag{5}$$

where $\rho = \beta/\gamma(\beta+\gamma)$ and $\sigma = 1/(\beta+\gamma)$.

Proof. As

$$\begin{aligned} [(\alpha f \# \beta j) + \gamma j]^* &= (\alpha f \# \beta j)^* \# (\gamma j)^* = [(\alpha f)^* + (\beta j)^*] \# \gamma^{-1} j \\ &= [(\alpha f)^* + \beta^{-1} j] \# \gamma^{-1} j \end{aligned}$$

and $(\alpha f)^*(x) = \alpha f^*(x/\alpha)$, we have

$$\begin{aligned} [(\alpha f \# \beta j) + \gamma j]^*(x) &= \inf_{u\in X}\left\{\alpha f^*\left(\frac{u}{\alpha}\right) + \frac{\|u\|^2}{2\beta} + \frac{\|x-u\|^2}{2\gamma}\right\} \\ &= \inf_{u\in X}\left\{\alpha f^*(u) + \frac{\|\alpha u\|^2}{2\beta} + \frac{\|x-\alpha u\|^2}{2\gamma}\right\}. \end{aligned}$$

Now, some algebra shows that

$$\frac{\|\alpha u\|^2}{2\beta} + \frac{\|x-\alpha u\|^2}{2\gamma} = \frac{\rho}{2}\|x-u\|^2 + \frac{\sigma}{2}\|x\|^2,$$

and so $[(\alpha f \# \beta j) + \gamma j]^* = (\alpha f^* \# \rho j) + \sigma j$. □

Now note that having $\alpha, \beta, \gamma > 0$ with $\alpha = \beta/(\beta+\gamma)$, $\beta = \beta/\gamma(\beta+\gamma)$, $\gamma = 1/(\beta+\gamma)$ and such that

$$[(\alpha f \# \beta j) + \gamma j]^* = (\alpha f^* \# \beta j) + \gamma j,$$

amounts to having $\gamma \in (0,1)$, $\beta = \frac{1-\gamma^2}{\gamma}$, $\alpha = 1-\gamma^2$. Then, the equation above turns to

$$\left[\left(\left(1-\gamma^2\right) f \# \frac{1-\gamma^2}{\gamma} j\right) + \gamma j\right]^* = \left(\left(1-\gamma^2\right) f^* \# \frac{1-\gamma^2}{\gamma} j\right) + \gamma j,$$

which is exactly (3). This proves Theorem 2.2.

3. The case of saddle functions

3.1. Preliminaries

Let X and Y be Hilbert spaces. Inner products and norms in both will be denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$. By a saddle function we will understand $h : X \times Y \to \overline{I\!R}$ such that $h(x,y)$ is convex in x for each fixed y and concave in y for each fixed x. Properties of saddle functions that in a sense parallel properness and lower semicontinuity of convex functions are properness and closedness. A saddle function h is proper and closed if its convex parent $f : X \times Y \to \overline{I\!R}$ and its concave parent $g : X \times Y \to \overline{I\!R}$, obtained from h via partial conjugacy formulas

$$f(x,q) = \sup_{y\in Y}\left\{h(x,y) + \langle q,y\rangle\right\}, \qquad g(p,y) = \inf_{x\in X}\left\{h(x,y) - \langle p,x\rangle\right\}$$

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

$$-g(p,y) = \sup_{x\in X, q\in Y}\left\{\langle p,x\rangle + \langle y,q\rangle - f(x,q)\right\}.$$

The equivalence class of a proper and closed saddle function h consists of all proper and closed saddle functions that have the same parents as h, and has the least and the greatest elements given, respectively, by

$$\underline{h}(x,y) = \sup_{p\in X}\left\{g(p,y) + \langle x,p\rangle\right\}, \qquad \overline{h}(x,y) = \inf_{q\in Y}\left\{f(x,q) - \langle y,q\rangle\right\}.$$

Given a proper and closed saddle function h, the class conjugate to it (in the saddle sense) has the least and the greatest elements given by

$$\underline{h^*}(p,q) = \sup_x \inf_y \left\{\langle p,x\rangle + \langle q,y\rangle - h(x,y)\right\}, \quad \overline{h^*}(p,q) = \inf_y \sup_x \left\{\langle p,x\rangle + \langle q,y\rangle - h(x,y)\right\}.$$

In other words, the class conjugate to h comes from a convex parent $(p, y) \mapsto -g(p, -y)$ and a concave parent $(x, q) \mapsto -f(x, -q)$. If either function displayed above is finite-valued, then $\underline{h^*} = \overline{h^*}$ and the equivalence class conjugate to h consists of one element. For details, see [18], [3] and, for the finite-dimensional case, [15], [16].

An extension of the idea of Moreau envelope to saddle functions was proposed by [4]. Originally, the approximation used two parameters. Here, considering one will suffice. Given a proper and closed saddle function h and any $\lambda > 0$, the mixed Moreau envelope is defined by

$$E_\lambda h(x, y) = \inf_{u \in X} \sup_{v \in Y} \left\{ h(u, v) + \frac{1}{2\lambda}\|x - u\|^2 - \frac{1}{2\lambda}\|y - v\|^2 \right\}. \tag{6}$$

The order of taking the infimum and supremum in (6) is irrelevant, and the envelope $E_\lambda h$ depends only on the equivalence class of h. Basic properties of $E_\lambda h(x, y)$ are similar to those of the Moreau envelope for convex functions. That is, for each $\lambda > 0$, $E_\lambda h$ is a finite continuous and continuously differentiable saddle function, with the gradient Lipschitz continuous with constant $1/\lambda$ and given by $\nabla E_\lambda h(x, y) = (-(x - \bar{u})/\lambda, (y - \bar{v})/\lambda)$, where $(\bar{u}, \bar{v})$ is the unique saddle point in the minimax problem in (6). See [2].

3.2. Smoothing of saddle functions

Definition 3.1. Given a proper and closed saddle function h and any $\lambda \in (0, 1)$, the function $S_\lambda h : X \times Y \to I\!R$ is defined by

$$S_\lambda h(x, y) = \left(1 - \lambda^2\right) E_\lambda h(x, y) + \frac{\lambda}{2}\left(\|x\|^2 - \|y\|^2\right), \tag{7}$$

where $E_\lambda h$ is the mixed Moreau envelope of h given by (6).

The function $S_\lambda h$ inherits finiteness, continuity, differentiability, and Lipschitz continuity of $\nabla S_\lambda h$ from the corresponding properties of $E_\lambda h$, and is also strongly convex in x, strongly concave in y. We now turn to the symmetry properties of the smoothing operation in (7) and its relationship to the smoothing operation for convex functions (2). Below, given a convex function $\phi : X \times Y \to \overline{I\!R}$,

$$s_\lambda \phi(p, y) = \left(1 - \lambda^2\right) \inf_{u \in X, v \in Y} \left\{ \phi(u, v) + \frac{1}{2\lambda}\|p - u\|^2 + \frac{1}{2\lambda}\|y - v\|^2 \right\} + \frac{\lambda}{2}\|p\|^2 + \frac{\lambda}{2}\|y\|^2.$$

That is, $s_\lambda \phi$ is the smoothing of ϕ in the sense of (2).

Theorem 3.2. *Let $h : X \times Y \to \overline{I\!R}$ be a proper and closed saddle function and let $\phi : X \times Y \to \overline{I\!R}$ be a proper, lsc, and convex function given by*

$$\phi(p, y) = \sup_{x \in X} \left\{ \langle p, x \rangle - h(x, y) \right\}.$$

Then

$$s_\lambda \phi(p, y) = \left(S_\lambda h(\cdot, y)\right)^* (p) = \sup_{x \in X} \left\{ \langle p, x \rangle - S_\lambda h(x, y) \right\},$$

and, equivalently,

$$S_\lambda h(x, y) = \left(s_\lambda \phi(\cdot, y)\right)^* (x) = \sup_{p \in X} \{ \langle x, p \rangle - s_\lambda \phi(p, y) \}.$$

The proof is given at the end of this section. Here, we note that the result is written in terms of the function ϕ and not convex or concave parents of h, so that it reduces exactly to Theorem 2.2 for the special case of h being a convex function of x. However, as ϕ above equals $-g$, where g is the concave parent of h, the concave parent of $S_\lambda h$ is $-s_\lambda(-g)$. Similarly, the convex parent of $s_\lambda f$, equal to the convex conjugate of $-g = s_\lambda(-g)$, turns out to be $s_\lambda f$, where f is the convex parent of h.

Now, based on the just observed fact that smoothing of a saddle function according to (7) corresponds to smoothing of its parents according to (2), we can conclude that smoothing of saddle functions is self dual with respect to saddle function conjugacy. Indeed, the concave parent of $(S_\lambda h)^*$ is, in light of the said fact, given by $(x, q) \mapsto -s_\lambda f(x, -q)$. As for any saddle function h, $s_\lambda h$ depends only on the equivalence class of h and the equivalence class of $s_\lambda h$ has only one element, we obtain:

Corollary 3.3. *Let $h : X \times Y \to \overline{I\!R}$ be a proper and closed saddle function. Then*

$$(s_\lambda h)^* = s_\lambda h^*. \tag{8}$$

We now prove Theorem 3.2. Note that for the special case of $h : X \to \overline{I\!R}$ being a convex function, which is what one considers to deduce Theorem 2.2 from Theorem 3.2, the proof simplifies as no minimax theorems need to be invoked.

Proof. Let $\gamma = 1-\lambda^2$. Below, $\sup_p$ means $\sup_{p\in X}$, similarly for u and z, while $\inf_v$ means $\inf_{v\in Y}$.

$$(s_\lambda\phi(\cdot, y))^*(x) = \sup_p \{\langle x, p\rangle - s_\lambda\phi(p, y)\}$$

$$= \sup_p \left\{ \langle x, p\rangle - \gamma \inf_{u,v} \left\{ \phi(u, v) + \frac{1}{2\lambda}\|p-u\|^2 + \frac{1}{2\lambda}\|y-v\|^2 \right\} - \frac{\lambda}{2}\|p\|^2 - \frac{\lambda}{2}\|y\|^2 \right\}$$

$$= \sup_{p,u,v} \left\{ \langle x, p\rangle - \gamma \sup_z \{\langle u, z\rangle - h(z, v)\} - \frac{\gamma}{2\lambda}\|p-u\|^2 - \frac{\gamma}{2\lambda}\|y-v\|^2 - \frac{\lambda}{2}\|p\|^2 \right\} - \frac{\lambda}{2}\|y\|^2$$

$$= \sup_{p,u,v} \inf_z \left\{ \langle x, p\rangle - \gamma\langle u, z\rangle + \gamma h(z, v) - \frac{\gamma}{2\lambda}\|p-u\|^2 - \frac{\gamma}{2\lambda}\|y-v\|^2 - \frac{\lambda}{2}\|p\|^2 \right\} - \frac{\lambda}{2}\|y\|^2.$$

Now note that the function

$$(z, p, u, v) \mapsto \langle x, p\rangle - \gamma\langle u, z\rangle + \gamma h(z, v) - \frac{\gamma}{2\lambda}\|p-u\|^2 - \frac{\gamma}{2\lambda}\|y-v\|^2 - \frac{\lambda}{2}\|p\|^2$$

is convex in z for a fixed (p, u, v), concave in (p, u, v) for a fixed z, and as such a saddle function, it is proper and closed (since h is). In fact, it is strongly concave in (p, u, v): it is a sum of a concave in (p, u, v) function and of $-\frac{1}{2\lambda}\|p\|^2 - \frac{\gamma}{2\lambda}\|u\|^2 - \frac{\gamma}{2\lambda}\|v\|^2$. Consequently, the order of taking the supremum and infimum is irrelevant. (In finite dimensions, this follows from [16, Theorem 37.3]. In general, it can be deduced, for example, from [5, Chapter 2, Corollary 3.4], by using any point where h has a nonempty subdifferential.) Thus,

$$(s_\lambda\phi(\cdot, y))^*(x)$$

$$= \inf_z \sup_v \left\{ \gamma h(u, v) + \sup_{p,u} \left\{ \langle x, p\rangle - \gamma\langle u, z\rangle - \frac{\gamma}{2\lambda}\|p-u\|^2 - \frac{\lambda}{2}\|p\|^2 \right\} - \frac{\gamma}{2\lambda}\|y-v\|^2 \right\}$$

$$-\frac{\lambda}{2}\|y\|^2.$$

Now

$$\sup_{p,u} \left\{ \langle x, p\rangle - \gamma\langle u, z\rangle - \frac{\gamma}{2\lambda}\|p-u\|^2 - \frac{\lambda}{2}\|p\|^2 \right\} = \frac{\gamma}{2\lambda}\|x-z\|^2 + \frac{\lambda}{2}\|x\|^2,$$

and so

$$\begin{aligned} (s_\lambda \phi(\cdot, y))^*(x) &= \inf_z \sup_v \left\{ \gamma h(z,v) + \frac{\gamma}{2\lambda}\|x-z\|^2 - \frac{\gamma}{2\lambda}\|y-v\|^2 \right\} - \frac{\lambda}{2}\|x\|^2 + \frac{\lambda}{2}\|y\|^2 \\ &= \gamma e_\lambda h(x,y) + \frac{\lambda}{2}\left(\|x\|^2 - \|y\|^2\right). \end{aligned}$$

□

4. Applications

In what follows, $\|\cdot\|$ is the Euclidean norm and $x \cdot y$ is the dot product.

4.1. Dual problems in optimization

Given convex, lsc, and proper functions $g_i : I\!R^n \to \overline{I\!R}$, $i = 0, 1, \dots, m$, consider the (primal) optimization problem of minimizing $g_0(x)$ subject to $g_i(x) \le 0$ for $i = 1, 2, \dots, m$ over $x \in I\!R^n$. In other words, the problem is to minimize φ, where

$$\varphi(x) = g_0(x) + \sum_{i=1}^m \delta_{(-\infty,0]}(g_i(x)).$$

Above, $\delta_{(-\infty,0]}$ is the indicator of $(-\infty, 0]$, with $\delta_{(-\infty,0]}(z) = 0$ if $x \le 0$, $\delta_{(-\infty,0]}(z) = \infty$ if $z > 0$. Then $\phi(x) = f(x, 0)$ for a (convex, lsc, and proper) $f : I\!R^n \times I\!R^m \to \overline{I\!R}$ given by

$$f(x,u) = g_0(x) + \sum_{i=1}^m \delta_{(-\infty,0]}(g_i(x) - u_i).$$

The Lagrangian $l : I\!R^n \times I\!R^m \to \overline{I\!R}$ is a proper and closed saddle function given by

$$l(x,y) = \inf_u \{f(x,u) - y \cdot u\} = g_0(x) + \sum_{i=1}^m y_i g_i(x) - \sum_{i=1}^m \delta_{[0,\infty)}(y_i),$$

with the convention that $\infty - \infty = \infty$. The dual problem is that of maximizing ψ, where $\psi(y) = \inf_x l(x,y) = -f^*(0,y)$. For background, see [19] or [20, Chapter 11.H,I].

Applying the smoothing (7) to the Lagrangian l results in primal and dual problems having differentiable objective functions. Indeed, the primal problem corresponding to the Lagrangian $S_\lambda l$ is that of minimizing φ_λ where $\varphi_\lambda(x) = s_\lambda f(x,0)$ – the function $s_\lambda f(x,u)$ comes from considering $\sup_y \{S_\lambda l(x,y) + u \cdot y\}$, and this supremum is exactly $s_\lambda f(x,u)$, by Theorem 3.2. The dual is the problem of maximizing ψ_λ where $\psi_\lambda(y) = -(s_\lambda f)^*(0,y) = -s_\lambda f^*(0,y)$. We conclude by noting that φ_λ is exactly the function resulting from replacing, in φ, the constraints $g_i(x) \le 0$ by quadratic penalties, and then smoothing the result as in (2). Indeed,

$$\begin{aligned}
&\varphi_\lambda(x) = s_\lambda f(x,0)\\
=\ & \left(1-\lambda^2\right)\inf_{\alpha,\beta}\left\{f(\alpha,\beta)+\frac{1}{2\lambda}\|x-\alpha\|^2+\frac{1}{2\lambda}\|\beta\|^2\right\}+\frac{\lambda}{2}\|x\|^2\\
=\ & \left(1-\lambda^2\right)\inf_{\alpha,\beta}\left\{g_0(\alpha)+\sum_{i=1}^m \delta_{[g_i(\alpha),\infty)}(\beta_i)+\frac{1}{2\lambda}\|x-\alpha\|^2+\frac{1}{2\lambda}\|\beta\|^2\right\}+\frac{\lambda}{2}\|x\|^2\\
=\ & \left(1-\lambda^2\right)\inf_{\alpha}\left\{g_0(\alpha)+p_\lambda(\alpha)+\frac{1}{2\lambda}\|x-\alpha\|^2\right\}+\frac{\lambda}{2}\|x\|^2
\end{aligned}$$

where $p_\lambda(x) = \sum_{i=1}^m \begin{cases} 0 & g_i(x)\le 0,\\ \frac{1}{2\lambda}\|g_i(x)\|^2 & g_i(x)>0.\end{cases}$ Now, the last line displayed above is exactly the smoothing of $g_0 + p_\lambda$.

4.2. Dual problems of calculus of variations

Given convex, lsc, and proper functions $g : \mathbb{R}^n \to \overline{\mathbb{R}}$, $L : \mathbb{R}^{2n} \to \overline{\mathbb{R}}$, consider a pair of value functions defined by dual problems of Bolza type: for each $\tau \ge 0$, $\xi \in \mathbb{R}^n$, let

$$V(\tau,\xi) = \inf\left\{g(x(0)) + \int_0^\tau L\left(x(t),\dot{x}(t)\right)\,dt \;\middle|\; x(\tau)=\xi\right\},$$

$$W(\tau,\eta) = \inf\left\{g^*(p(0)) + \int_0^\tau L^*\left(\dot{p}(t),p(t)\right)\,dt \;\middle|\; p(\tau)=\eta\right\}.$$

Above, the minimization is over all absolutely continuous arcs x, respectively, p, on $[0,\tau]$ that meet the endpoint constraints. We pose the mild growth assumptions, required by the duality theory developed for such problems in [21]: there exist constants α,β and of a coercive, nondecreasing function θ on $[0,\infty)$ such that both $L(x,v)$ and $L^*(v,x)$ are bounded below by $\theta\left(\max\{0,\|v\|-\alpha\|x\|\}\right)-\beta\|x\|$. This assumption guarantees, for example, for each τ, $V(\tau,\cdot)$ and $W(\tau,\cdot)$ are convex, lsc, and proper functions conjugate to each other. Furthermore, numerous properties of the value functions can be studied, through the Hamilton-Jacobi partial differential equation and through the Hamiltonian differential inclusion, with the help of the Hamiltonian

$$H(x,p) = \sup_v\left\{p\cdot v - L(x,v)\right\},$$

which is a finite-valued (so proper and closed) saddle function (with the twist that $H(x,p)$ is concave-convex, in contrast to convex-concave functions we studied in this paper).

The value functions are of use in optimality verification techniques, and together with the Hamiltonian can be used to construct the optimal feedback mapping: essentially, optimal solutions to the problem defining $V(\tau,\xi)$ are the solutions to the differential inclusion $\dot{x} \in \partial_p H(x,\partial_\xi V(t,x))$. In general, the right-hand side of this inclusion can have bad regularity properties (it may even turn out to not have convex values; see [10]) and it may be desirable to approximate it with a more regular mapping.

Let us consider approximate problems, with L replaced by $s_\lambda L$, and accordingly, with L^* replaced by $s_\lambda L^*$, and denote the resulting value functions by V_λ and W_λ. The Hamiltonian corresponding to $s_\lambda L$ is $S_\lambda H$ (with appropriately changed definition of the smoothing for the concave-convex H). Note that we are not smoothing the initial costs g and g^*. Both V_λ and W_λ are finite-valued (since $s_\lambda L$, $s_\lambda L^*$ are). Functions $s_\lambda L$, as $\lambda \searrow 0$, meet the uniform growth conditions used to study convergence of problems of Bolza in [11]. We can then say that V_λ, W_λ are such that for each $\tau \geq 0$, $V_\lambda(\tau, \cdot)$, $W_\lambda(\tau, \cdot)$ converge epigraphically to V, W; this follows from [11, Theorem 4.6]. Also, $V_\lambda(\tau, \cdot)$, $W_\lambda(\tau, \cdot)$ are strictly convex – this can be easily shown from the strong, and thus strict, convexity of $s_\lambda L$, $s_\lambda L^*$ – and by conjugacy between them, they are also differentiable. (Alternatively, one can rely on the strict concavity, strict convexity of $S_\lambda H$, and [10, Theorem 4.3].) However, much more can be said about the regularity of V_λ, W_λ, if one accounts for both strong concavity, strong convexity of H and Lipschitz continuity of ∇H.

Proposition 4.1. *For each $\tau > 0$, the functions $V_\lambda(\tau, \cdot)$, $W_\lambda(\tau, \cdot)$ are differentiable and their gradients are Lipschitz continuous with constant $\kappa = 2\lambda^{-2}\left(1 - e^{-2\tau/\lambda}\right)^{-1}$. By duality, these functions are also strongly convex with constant κ^{-1}.*

Proof. We only show Lipschitz continuity of $\nabla V_\lambda(\tau, \cdot)$; all the other conclusions follow by symmetry of the assumptions and conjugacy between $V_\lambda(\tau, \cdot)$ and $W_\lambda(\tau, \cdot)$. We already know $V(\tau, \cdot)$ is finite-valued. Pick any $\xi_1 \neq \xi_2$ and any $\eta_1 \in \partial_\xi V(\tau, \xi_1)$, $\eta_2 \in \partial_\xi V(\tau, \xi_2)$. (Here, $\partial_\xi V(\tau, \cdot)$ denotes the subdifferential of the convex function $V(\tau, \cdot)$.) By [21, Theorem 2.4], for $i = 1, 2$, there exist solutions $(x_i, p_i) : [0, \tau] \to I\!R^n \times I\!R^n$ to the Hamiltonian dynamical system associated with $S_\lambda H$, such that $(x_i(\tau), p_i(\tau)) = (\xi_i, \eta_i)$ while $p_i(0) \in \partial g(x_i(0))$. Now, [17, Theorem 4] adapted for the case of a strongly concave, strongly convex Hamiltonian $S_\lambda H$, implies that

$$\frac{d}{dt}\left(x_1(t) - x_2(t)\right) \cdot \left(p_1(t) - p_2(t)\right) \geq \lambda c^2(t), \tag{9}$$

where $c(t) = \sqrt{\|x_1(t) - x_2(t)\|^2 + \|p_1(t) - p_2(t)\|^2}$. Since the Hamiltonian differential inclusion for $S_\lambda H$ reduces to an ordinary differential equation with Lipschitz continuous right-hand side, with constant $1/\lambda$, we have $c(t) \geq c(\tau)e^{(t-\tau)/\lambda}$ for all $t \in [0, \tau]$. Combining this bound, inequality (9), and the fact that

$$\left(x_1(0) - x_2(0)\right) \cdot \left(p_1(0) - p_2(0)\right) \geq 0$$

thanks to the monotonicity of the subdifferential mapping ∂g, we obtain

$$(\xi_1 - \xi_2) \cdot (\eta_1 - \eta_2) = \left(x_1(t) - x_2(t)\right) \cdot \left(p_1(t) - p_2(t)\right) \geq d,$$

where $d = \lambda^2 c^2(\tau)\left(1 - e^{-2\tau/\lambda}\right)/2$. Now $c(\tau) = \sqrt{\|\xi_1 - \xi_2\|^2 + \|\eta_1 - \eta_2\|^2}$, and so $c(\tau) \geq \|\eta_1 - \eta_2\|$. Then $\|\xi_1 - \xi_2\| c(\tau) \geq \|\xi_1 - \xi_2\| \|\eta_1 - \eta_2\| \geq d$ and the last two inequalities yield

$$\frac{\|\eta_1 - \eta_2\|}{\|\xi_1 - \xi_2\|} \leq \frac{c^2(\tau)}{d} = 2\lambda^{-2}\left(1 - e^{-2\tau/\lambda}\right)^{-1}.$$

This finishes the proof. □

In summary, by varying only L and L^* (in a symmetric way) and not g nor g^*, we can approximate both value functions V and W by families of value functions for which the gradient with respect to the state variable is Lipschitz continuous. As the approximate Hamiltonian also has a Lipschitz continuous gradient, the resulting optimal feedback mappings are Lipschitz continuous.

We conclude by illustrating that just strong concavity, strong convexity of the Hamiltonian is not sufficient for the conclusions of Proposition 4.1 to hold. Indeed, let $n = 1$ and consider $L(x, v) = \|x\| + x^2/2 + v^2/2$, so that $H(x, p) = -\|x\| - x^2/2 + p^2/2$. This H is strongly concave, strongly convex with constant 1, and so $V(\tau, \cdot)$ is differentiable, for each $\tau > 0$, independently of g. The Hamiltonian differential inclusion reduces to $\dot{x} = p$, $\dot{p} = x+1$ for all p and for $x > 0$, while for $x = 0$, we have $\dot{x} = p$, $\dot{p} \in [-1, 1]$. In particular, solutions to the inclusion on $[0, \tau]$ starting from $(0, 0)$ can remain at $(0, 0)$ for any amount of time, say up to time s, and then start evolving according to $\dot{x} = p$, $\dot{p} = x+1$, which, at time $\tau > s$, amounts to $x(\tau) = e^{\tau-s}/2+e^{-(\tau-s)}/2-1$, $p(\tau) = e^{\tau-s}/2-e^{-(\tau-s)}/2$. Pick any (convex, lsc) g with $g \geq 0$, $g(0) = 0$, so that in particular, $0 \in \partial g(0)$. By [21, Theorem 2.4], for each $\tau > 0$, $e^t/2 - e^{-t}/2 = \nabla V_\xi\left(\tau, e^t/2 + e^{-t}/2 - 1\right)$ for all $t \in [0, \tau]$. This is enough to check that $\nabla_\xi V(\tau, \cdot)$ is not Lipschitz continuous at 0.

References

[1] H. Attouch: Variational Convergence for Functions and Operators, Pitman, Boston (1984).

[2] H. Attouch, D. Aze, R. J.-B. Wets: On continuity properties of the partial Legendre-Fenchel transform: convergence of sequences of augmented Lagrangian functions, Moreau-Yosida approximates and subdifferential operators, in: Fermat Days 85: Mathematics for Optimization, J.-B. Hiriart-Urruty (ed.), North-Holland, Amsterdam (1986) 1–42.

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

[4] H. Attouch, R. J.-B. Wets: A convergence theory for saddle functions, Trans. Amer. Math. Soc. 280(1) (1983) 1–41.

[5] V. Barbu, Th. Precupanu: Convexity and Optimization in Banach Spaces, D. Reidel, Dordrecht (1986).

[6] A. Ben-Tal, M. Teboulle: A smoothing technique for nondifferentiable optimization problems, in: Optimization (Varetz, 1988), Lect. Notes Math. 1405, Springer, Berlin (1989) 1–11.

[7] A. Ben-Tal, M. Teboulle: A conjugate duality scheme generating a new class of differentiable duals, SIAM J. Optim. 6(3) (1996) 617–625.

[8] M. Ghomi: The problem of optimal smoothing for convex functions, Proc. Amer. Math. Soc. 130(8) (2002) 2255–2259.

[9] R. Goebel: Convexity, Convergence and Feedback in Optimal Control, PhD Thesis, University of Washington, Seattle (2000).

[10] R. Goebel: Regularity of the optimal feedback and the value function in convex problems of optimal control, Set-Valued Anal. 12(1-2) (2004) 127–145.

[11] R. Goebel: Convex optimal control problems with smooth Hamiltonians, SIAM J. Control Optim. 43(5) (2005) 1781–1811.

[12] S. Guillaume, A. Seeger: A higher-order smoothing technique for polyhedral convex functions: geometric and probabilistic considerations, J. Convex Analysis 8(1) (2001) 109–126.

[13] C. Lemaréchal, C. Sagastizábal: Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries, SIAM J. Optim. 7(2) (1997) 267–385.

[14] J.-J. Moreau: Proximité et dualité dans un espace hilbertien, Bull. Soc. Math. Fr. 93 (1965) 273–299.

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

[16] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton (1970).

[17] R. T. Rockafellar: Generalized Hamiltonian equations for convex problems of Lagrange, Pacific J. Math. 33(2) (1970) 411–427.

[18] R. T. Rockafellar: Saddle points of Hamiltonian systems in convex problems of Lagrange, J. Optimization Theory Appl. 12(4) (1973) 367–390.

[19] R. T. Rockafellar: Conjugate Duality and Optimization, CBMS-NSF Regional Conference Series in Applied Mathematics 16, SIAM, Philadelphia (1974).

[20] R. T. Rockafellar, R. J.-B. Wets: Variational Analysis, Springer, Berlin (1998).

[21] R. T. Rockafellar, P. R. Wolenski: Convexity in Hamilton-Jacobi theory. 1: Dynamics and duality, SIAM J. Control Optim. 39(5) (2000) 1323–1350.

[22] A. Seeger: Smoothing a nondifferentiable convex function: the technique of the rolling ball, Rev. Mat. Apl. 18(1) (1997) 45–60.

[23] A. Seeger: Smoothing a polyhedral convex function via cumulant transformation and homogenization, Ann. Polon. Math. 67 (1997) 259–268.

[24] S. Simons: Personal communication (2004).

[25] M. Teboulle: Entropic proximal mappings with applications to nonlinear programming, Math. Oper. Res. 17(3) (1992) 670–690.