A regularization technique in dynamic optimization

Alvaro Guevara
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://repository.lsu.edu/gradschool_dissertations

Part of the Applied Mathematics Commons

Recommended Citation
https://repository.lsu.edu/gradschool_dissertations/3915
A regularization technique in dynamic optimization

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Mathematics

by

Alvaro José Guevara
B.S., Universidad de Costa Rica, 2002
M.Sc., Louisiana State University, 2006
August 2009
Acknowledgments

I thank God for taking good care of me and my family during all these years.

I want to especially thank my advisor Peter Wolenski for his support and mentoring during my graduate studies. It has been a great pleasure and honor to work with him. During our conversations, he conveyed a passion about mathematics and science in general that inspired me to learn more and more.

I also want to thank the Department of Mathematics for providing me with resources to excel as a graduate student. Moreover I want to thank all the members of my final exam committee. Particularly, I want to thank Prof Frank Neubrander and Prof Perlis, for their support in order for me to participate in workshops and give talks; to Prof Shipman, with whom I shared several conversations about math and life in general; Prof Oxley, for his real interest in helping me become a better teacher; and Prof Estrada, who guided me during my undergraduate studies and remains a very good friend.

I worked with excellent researchers outside the Math department, to whom I express my gratitude: Brad Manor, Prof Li Li, Prof Alex Cohen and Prof Lee Hong. It has been great to work with people with such a tremendous appreciation of mathematics.

Thank you to my friends in Baton Rouge: Patricio, Jasson, Raquel, Silvia, Jens, Rick, Armin, Santiago, Maria, Carolyn, Jeremy, Kevin, Qingxia, Lee, Jake, Luis, Ana; and also those who left Baton Rouge: Bacim, Stanislav, Nils, Tania, Martin, Vinicio.

I left many friends in Costa Rica as well who, regardless of the physical distance, have been very close to me during this time: Guis, Hazel, Bobby (Heiner), Nubia, Sharold, Estela, Jose Manuel, and Bea. Also, my friends from Costa Rica who are working/studying in the US: Leo, Nacho, and David.

I want to dedicate this to my family: Ma, Pa, Andres, Diego, Esteban, Jose Pablo, and Daniela. Thank you Ma for your love and support. And Tefi, for making me smile every time I see you.

Finally, Dally, my dear, this is for you. Thank you very much for your true love and deep affection. Enormous blessings to all your family, especially to the “little” girls Keren and Danielita.
# Table of Contents

Acknowledgments ................................................................. ii

Abstract ................................................................. iv

Introduction .............................................................. 1

Chapter 1: Convex Analysis and Saddle Function Theory .............. 7
  1.1 Convex Sets and Functions ............................................. 7
      1.1.1 Subgradients .................................................. 11
      1.1.2 Conjugation and Inf-convolution ............................ 11
      1.1.3 Epi-convergence .............................................. 12
      1.1.4 Fenchel Duality Theorem .................................... 13
  1.2 Saddle Functions ................................................... 14
      1.2.1 Convex and Concave Parents ................................ 17
      1.2.2 Closure and Properness ..................................... 19
      1.2.3 Saddle Points and Dual Saddle Functions ................. 24
      1.2.4 Subgradients ................................................. 27
      1.2.5 Hypo/epi-convergence ....................................... 28

Chapter 2: Regular Approximation ...................................... 33
  2.1 Moreau Envelope and Prox-bounded Functions .................... 33
  2.2 Regular Approximation ............................................. 37
      2.2.1 Prox-regular Functions .................................... 41
  2.3 Moreau Envelope Under Convexity Assumptions .................. 42
  2.4 Self-dual Regular Approximation ................................. 44
  2.5 Self-dual Regular Approximation For Saddle Functions ........ 46
  2.6 Applications to Optimization .................................... 48
      2.6.1 Basic Duality Theory For Convex Programs ............... 48
      2.6.2 First Regularization Scheme .............................. 52
      2.6.3 Second Regularization Scheme ............................ 53

Chapter 3: Calculus of Variations ....................................... 59
  3.1 Calculus of Variations Under Full Convexity Assumptions ....... 59
      3.1.1 Value Function Regularization ............................. 62

References .............................................................. 65

Appendix: Symbolic Computation of the Regular Approximation .... 68

Vita ................................................................. 70
Abstract

In this dissertation we discuss certain aspects of a parametric regularization technique which is based on recent work by R. Goebel. For proper, lower semicontinuous, and convex functions, this regularization is self-dual with respect to convex conjugation, and a simple extension of this smoothing exhibits the same feature when applied to proper, closed, and saddle functions.

In Chapter 1 we give an introduction to convex and saddle function theory, which includes new results on the convergence of saddle function values that were not previously available in the form presented. In Chapter 2, we define the regularization and extend some of the properties previously shown in the convex case to the saddle one. Furthermore, we investigate the properties of this regularization without convexity assumptions. In particular, we show that for a prox-bounded function the family of infimal values of the regularization converges to the infimal values of the given function, even when the given function might not have a minimizer. Also we show that for a general type of prox-regular functions the regularization is locally convex, even though their Moreau envelope might fail to have this property. Moreover, we apply the regularization technique to Lagrangians of convex optimization problems in two different settings, and describe the convergence of the associated saddle values and the value functions.

We also employ the regularization in fully convex problems in calculus of variations, in Chapter 3, in the setting studied by R. Rockafellar and P. Wolenski. In this case, we extend a result by Rockafellar on the Lipschitz continuity of the proximal mapping of the value function jointly in the time and state variables, which in turn implies the same regularity for the gradient of the self-dual regularization.

Finally, we attach a software code to use with SCAT (Symbolic Convex Analysis Toolbox) in order to symbolically compute the regularization for functions of one variable.
Introduction

Optimization is a central branch of applied mathematics whose goal is to find optimal values with respect to certain criteria restricted to specific constraints. The basic problem in mathematical optimization can be formulated as follows: for a vector space $X$ and a function $f : X \to (-\infty, \infty]$, determine

$$\min f(x) \text{ subject to } x \in C \subset X.$$  \hspace{1cm} (1)

The function $f$ is called the objective function and the set $C$ the feasible set. In this general formulation, one can see how optimization serves as a modeling tool to different areas of science. In fact, the objective function describes some criteria to be minimized (total energy of the system, cost to complete a task, error produced by an approximation) while the feasible set describes either the amount of resources available (fuel, speed of a chemical reaction) or a desirable feature of the solution (nonnegativity if one deals with prices of items).

There are many variations to the basic model (1), which are grouped according to the features of the objective function and the feasible set. For instance, problems are called finite or infinite-dimensional, depending on the dimension of the space $X$.

The typical questions in optimization usually fall into one of the following categories

- existence theory: state hypothesis which ensure a solution exists;
- necessary conditions: find properties that feasible elements must satisfy if they are to be optimal;
- sufficient conditions: find a test that can be applied to feasible solutions to verify their optimality;
- duality theory: introduce dual problems and exploit duality-preserving properties;
- regularity: find conditions to be imposed on the objective function and constraint set so that the problem behaves “nicely” with respect to parameter perturbation;
- approximation: construct approximation schemes with stronger regularity properties than the given problem to estimate optimal values;
- algorithm design: construct and implement algorithms to solve optimization problems.

One important question arises: what is the distinctive hypothesis to impose on the objective function and constraint set such that the optimization problem
(a) is not very restrictive as to serve as a model for several real-life applications;
(b) has a rich structure which helps to exploit tools in duality theory;
(c) has necessary conditions that become sufficient and vice versa under minor
or no extra assumptions;
(d) can be used as a local approximation for other more general problems.

The immediate answer that comes to mind is to use a linear model, in order to
mimic the success of linear systems in the theory of dynamical system, where those
systems have a rich structure and nonlinear problems are, with a few exceptions,
very complicated to solve. The idea of linearity as the distinctive factor in optimization
was reinforced with the publication of the simplex method to solve these prob-
lems. The term linear programming was coined by the method's author, G. Dantzig.
The term “program”, not to be confused with its use in computer science, meant
that no “closed form” solution was sought, but rather an algorithm or program was
provided instead. The word became a synonym for optimization. Nevertheless, the
research of prominent figures such as Fenchel, Moreau, and Rockafellar revealed
that nonlinear problems which were convex had a powerful structure similar to
those with linear assumptions; their results had extensions to infinite dimensions.
As Rockafellar pointed out in [Rock93], “the great watershed in optimization is not
between linearity and nonlinearity, but convexity and nonconvexity”.

Convex optimization had also a great impulse with the success of algorithms
such as the proximal point and the interior-point methods, which made possible
to solve many convex problems almost as easily as linear problems. This fact,
together with the strong theoretical development already established made (and
still makes) convex programs a very attractive modeling tool in applications. As
matter of fact, convex optimization has been applied in areas such as automatic
control systems, estimation and signal processing, communications and networks,
electronic circuit design, data analysis and modeling, statistics, and finance (see
[Boyd04] for details).

One of the features behind the success of convexity is the interplay among the
geometrical concepts, the analytical tools, the duality structure, and their appli-
cations in optimization problems. To exemplify this situation, we first recall
the concept of convex conjugate. For an extended-value function \( f \) on \( \mathbb{R}^n \), its convex
conjugate (also called the Legendre-Fenchel conjugate), denoted by \( f^* \), is given by

\[
 f^*(y) = \sup_x \{ y \cdot x - f(x) \}. 
\]

So, let us look for instance at the concept of minimizer of a finite convex function.
The following statements turn out to be equivalent

- a point \( x \) is a (global!) minimizer of a finite convex function \( f \) (optimization
problem),
- the zero vector is a subgradient of \( f \) at \( x \) (analytical concept),
• there exists a horizontal supporting hyperplane to the epigraph of \( f \) at \( x \) (geometrical concept)

• \( x \) belongs to the subgradient of the convex conjugate function \( f^* \) at zero (duality).

However, even in the advantageous convex setting, problems in mathematical programming arise for which either the objective function or the constraint set lack some regularity, which brings up mathematical and computational challenges to understand the behavior of optimal values and minimizers. It is desirable then to regularize the problem. A strategy for this purpose is to transform the optimization problem into a family of smooth problems for which the limiting behavior of their optimal solutions can shed light about the solutions of the original problem.

More concretely, a regularization method for an optimization problem consists of a parametrized family of optimization problems, each of which has objective function with enhanced regularity properties compared to the original objective function, whose solutions approximate the solutions of the given problem in a suitable sense, after a particular choice of the parameters. Regularization techniques have been used extensively in inverse problems (see for instance [Engl00]) to solve ill-posed problems in a least square sense. The Tikhonov regularization, which consists of adding an extra term of the form \( \lambda \| \cdot \|_2^2 \), for a parameter \( \lambda \), is a prominent tool in this field. Applications are also known in image processing [Mara96], calculus of variations [Goeb05], fluid dynamics [Asto08], among others fields. Regularizations are not only important from a theoretical standpoint, but also in numerical schemes, and the research in this area is oriented to find computationally cheap and efficient algorithms to guarantee the stability and fast convergence of the solutions of the regularized problems.

Among the regularization techniques, the envelope methods play a significant role. There, the regularizing family is monotonically increasing and has the objective function as an upper bound, thus providing an approximation “from below”. The Moreau envelope (also known in the literature as the Moreau-Yosida approximate) is a regularization of this type widely used in convex analysis, variational principles, and optimization (see for instance [Lema97], [Rock98], [Baus05], [Meng05], [Luce06]). For a function \( f : \mathbb{R}^n \to \mathbb{R} \) and a positive value \( \lambda \), the Moreau envelope \( e_\lambda f \) is given by

\[
e_\lambda f(x) = \inf_u \left\{ f(u) + \frac{1}{2\lambda} \| x - u \|^2 \right\}.
\]

In general, the inf-convolution \( f \# g \) of two extended-valued functions \( f \) and \( g \) in \( \mathbb{R}^n \) is given by

\[
(f \# g)(x) = \inf_u \{ f(u) + g(x - u) \};
\]

its name of course comes as a result of the resemblance with the usual convolution of functions in integration theory, provided the inf sign is interpreted as an integral
sign and the sum operation as a product. Defining \( j_\lambda(x) = \frac{1}{2\lambda} \|x\|^2 \) (here \( \|\cdot\| \) indicates the Euclidean norm), then it holds that

\[
e_\lambda f = f \# j_\lambda.
\]

Since \( j_\lambda(x) \) converges pointwise to the indicator function of zero, \( \delta_0 \), as \( \lambda \downarrow 0 \) - the latter meaning \( \lambda \to 0 \) with \( \lambda > 0 \) -, and \( f \# \delta_0 = f \), for every \( f \), \( e_\lambda f \) is the convolution of \( f \) with an approximating identity family. This type of regularization technique is predominant in partial differential equations and functional analysis.

Under mild growth conditions on \( f \), \( e_\lambda f \) is finite and continuous, and converges pointwise to \( f \) as \( \lambda \) converges to zero. An overview of the main properties of the Moreau envelope is given in Chapter 2. However, at this point, we want to emphasize how this envelope behaves under convex conjugation. For a sufficiently regular convex function \( f \),

\[
(e_\lambda f)^* = f^* + \frac{\lambda}{2} \|\cdot\|^2,
\]

and

\[
\left(f + \frac{\lambda}{2} \|\cdot\|^2\right)^* = e_\lambda f^*.
\]

In other words, the conjugate of the Moreau envelope function is the Tikhonov regularization of the conjugate function and the conjugate of the Tikhonov regularization of a function is the Moreau envelope of its conjugate. This situation was established by Schade in [Scha94] in a far more general setting, which can be stated as follows: a hullfunction (such as the Moreau envelope) of the primal problem \((P)\) corresponds to a regularization of the dual problem \((D)\), and vice versa.

As can be seen from the two previous equations, the regularity conditions that could be achieved by using the Moreau envelope are not necessarily preserved under conjugation. A challenge then is to construct regularizations which preserve the duality structure. This has importance in convex optimization, because every convex program can be associated with an auxiliary or dual problem, which is defined in terms of the convex conjugate. The optimal value of the dual problem provides a lower bound on the optimal value of original (primal) problem, and these two values match in many situations. Specific details on the relationship between the primal and its dual problem will be given in Chapter 2, and more extensive material can be found in [Rock74], [Boyd04], [Bert03], and [Borw06].

In this direction, R. Goebel introduced in [Goeb08] a smoothing for convex functions which is self-dual with respect to the convex conjugation. For a function \( f : \mathbb{R}^n \to (-\infty, \infty] \) and \( 0 < \lambda < 1 \), it is given by

\[
s_\lambda f(x) = (1 - \lambda^2) e_\lambda f(x) + \frac{\lambda}{2} \|x\|^2,
\]

and its striking feature is that, for a sufficiently regular convex function \( f \),

\[
(s_\lambda f)^* = s_\lambda f^*.
\]
The reason behind the success of the regular approximation $s_\lambda$ lies in the fact that, according to the terminology of Schade mentioned before, it combines the features of a hullfunction and a regularization simultaneously. Furthermore, an extension of $s_\lambda$ (denoted by $S_\lambda$) can be defined for functions $K : \mathbb{R}^m \times \mathbb{R}^n \to [-\infty, \infty]$ which are concave in the first argument and convex in the second one, called saddle functions, and it turns out to be self-dual as well under the saddle conjugation operation.

The following is the structure of this dissertation. Chapter 1 contains background material in convex analysis and provides a constructive and concise approach to the theory of saddle functions, including two new results on hypo/epi-convergent sequences in the modulated sense, which is a refinement of a result previously known only for everywhere-finite saddle functions. The modulated convergence for saddle functions was introduced by Rockafellar in [Rock90], and it is based on a previous definition made by Attouch and Wets in [Atto83b].

In Chapter 2 we introduce the Moreau envelope $e_\lambda$ and the regular approximation $s_\lambda$. Here we address the following questions:

- Assuming no previous knowledge on the existence of minimizers for the objective function $f$, do the infimal values of the regularized problems $s_\lambda f$ converge to the infimal value of $f$?
- Is there a more general type of function $f$ (not necessarily convex) for which $s_\lambda f$ is still convex?
- Under what conditions is $s_\lambda f$ coercive?
- For a saddle function $K$, do the saddle value (or points) of $S_\lambda K$ converge to the saddle value (or points) of $K$?

Also, we give a short overview of the duality theory in convex optimization and apply the regularizations $s_\lambda$ and $S_\lambda$ to convex programs, using two different approaches: one which follows Goebel in [Goeb08] and the other one inspired by a theorem of Attouch and Wets in [Atto83a]. The specific questions addressed are the following:

- Do the Lagrangians associated with the regularized problems hypo-epi converge to the Lagrangian associated with the original problem?
- Do the value functions associated with the regularized problem converge (epi or pointwise) to the value function of the original problem?
- If the constraint qualification known as Slater’s condition is satisfied for the original problem, do the regularized problems also satisfy it?

In Chapter 3, the regular approximation is applied to fully convex problems in calculus of variations, in the setting studied by Rockafellar and Wolenski in [Rock01a] and [Rock01b]. In this case, we extend a result by Rockafellar in [Rock05] on the Lipschitz continuity (in both the time and state variables) of the proximal
mapping of the regularized value function, which in turn implies the same regularity for the gradient of the self-dual regularization.

Finally, in the appendix we attach a software code to compute the self-dual smoothing $s_\lambda$ for a convex function in one variable. It was implemented in SCAT (Symbolic Convex Analysis Toolbox), a software code for use in MAPLE developed by Borwein and Hamilton [Borw09].
Chapter 1
Convex Analysis and Saddle Function Theory

In this chapter we will give some definitions and present important results in convex analysis and saddle function theory that will be used throughout the thesis. For the sake of completeness, most of the proofs for the results herein are either provided or referenced. Most of the material is classical; however, a few results are new in the form presented here, and they have been adapted from the literature either to fit a particular definition (for instance, for concave-convex functions instead of convex-concave) or a specific setting (for example, convergence of saddle points for modulated saddle sequences).

1.1 Convex Sets and Functions

Definition 1.1.1 A set $C \subset \mathbb{R}^n$ is said to be convex if for any elements $x, y \in C$, the line segment joining $x$ and $y$, namely, the set $\{\alpha x + (1 - \alpha) y | \alpha \in (0, 1)\}$ is contained in $C$.

From the definition (1.1.1), it is clear that any arbitrary intersection of convex sets is itself a convex set. This result allows us to form “convex closures”, analogous to the situation in topological spaces with closed sets.

Definition 1.1.2 For any set $P \subset \mathbb{R}^n$, its convex hull $\text{conv}(P)$ is defined as the intersection of all the convex sets containing $P$.

So, $\text{conv}(P)$ is the smallest convex set in $\mathbb{R}^n$ containing $P$, and clearly, $P$ is convex if and only if $P = \text{conv}(P)$. In general, $\text{conv}(P)$ is the set of all convex combinations of elements in $P$, that is,

$$\text{conv}(P) = \left\{ p \in \mathbb{R}^n \mid p = \sum_{i=1}^{l} \alpha_i p_i, p_i \in P, \alpha_i \geq 0, \sum_{i} \alpha_i = 1 \right\}$$

If in the previous equation the coefficients $\alpha_i$ are allowed to take any real value, then one obtains the affine hull of $P$, written $\text{aff} P$, which consists of all possible affine combinations of elements of $P$. Clearly, $\text{conv}(P) \subset \text{aff}(P)$. The interior and closure of $P \subset \mathbb{R}^n$ are indicated by $\text{int} P$ and $\text{cl} P$ respectively. We write $B(x, \rho)$ for the open ball centered at a point $x$ of radius $\rho$. The relative interior of $P$, denoted by $\text{ri} P$ is the set of all $x$ such that $B(x, \rho) \cap \text{aff} C \subset P$ for some $\rho > 0$. By definition, $\text{ri} P \subset \text{int} P$.

Convexity is invariant under a variety of set operations, as described in the theorem below.

Proposition 1.1.3. Let $C$ and $D$ be convex sets in $\mathbb{R}^m$. Then the following sets are convex:

1. $C \cup D$
2. $C \cap D$
3. $C + D$
4. $tC$ for any $t \geq 0$
5. $\text{int}(C)$
6. $\text{cl}(C)$
7. $B(x, \rho)$
8. $\text{aff}(P)$
9. $\text{conv}(P)$
• the vector sum $C + D = \{x = c + d \mid c \in C, d \in D\}$,

• the product set $C \times D = \{x = (c, d) \mid c \in C, d \in D\}$,

• the set $\mu C = \{\mu c \mid c \in C\}$, for every scalar $\mu$.

**Proposition 1.1.4.** For every linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ and convex set $C$ in $\mathbb{R}^n$, the image set $L(C)$ is a convex set in $\mathbb{R}^m$. Furthermore, for each convex subset $E$ in $\mathbb{R}^m$, the set $L^{-1}(E)$ is a convex set in $\mathbb{R}^n$.

In order to study variational principles, it is necessary to define a sense of convergence for sets in $\mathbb{R}^n$.

**Definition 1.1.5** For a sequence of sets $\{C_n\}$, let

$$\lim \sup_{n \to \infty} C_n = \bigcap_{n} \bigcup_{m>n} \text{cl} C_m = \{x \mid \exists \text{ a subsequence } x_{n_k} \in C_{n_k} \text{ with } x_{n_k} \to x \text{ as } k \to \infty\}$$

$$\lim \inf_{n \to \infty} C_n = \bigcup_{n} \bigcap_{m>n} \text{cl} C_m = \{x \mid \text{ for all but finite } n, x_n \in C_n \text{ with } x_n \to x \text{ as } k \to \infty\}$$

(1.1) 

(1.2)

The sequence $\{C_n\}$ is said to converge to a set $C$ in the Painlevé-Kuratowski sense provided $C = \lim_{n \to \infty} \sup C_n = \lim_{n \to \infty} \inf C_n$, and it is written $C = \lim_{n \to \infty} C_n$.

If $C_n = C$ for every $n$, then $\lim_{n \to \infty} C_n = \text{cl} C$.

We write $\overline{\mathbb{R}}$ to denote the extended-real line $[-\infty, \infty]$ with the usual ordering and the following extended arithmetic rules

• $r + \infty = \infty + r = \infty$, $r - \infty = -\infty + r = -\infty$, for $r \in \mathbb{R}$,

• $r \cdot \pm \infty = \pm \infty$, for $r > 0$,

• $r \cdot \pm \infty = \mp \infty$, for $r < 0$,

• $\infty + \infty = \infty$, $-\infty - \infty = -\infty$,

• $\infty - \infty = -\infty + \infty = \infty$,

• $0 \cdot \pm \infty = 0$.

**Definition 1.1.6** Let $f$ be a extended-real valued function defined on a set $S \subset \mathbb{R}^n$. The function $f$ is said to be convex if its epigraph

$$\text{epi } f = \{(x, \mu) \mid x \in S, \mu \geq f(x)\}$$

is a convex set in $\mathbb{R}^{n+1}$.
As a consequence, the effective domain

$$\text{dom } f = \{x \mid f(x) < \infty\}$$

of a convex function $f$ is itself convex, because it is the image set of a convex set ($\text{epi } f$) under a linear transformation (the projection map on $\mathbb{R}^n$). A convex function $f$ defined on a subset $S$ of $\mathbb{R}^n$ can be extended as a function $\tilde{f}$ on $\mathbb{R}^n$ by setting

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in S \\ \infty & \text{otherwise.} \end{cases}$$

The function $\tilde{f}$ thus defined is convex. Conversely, every convex function $\tilde{f}$ on $\mathbb{R}^n$ can be identified with a convex function defined on a subset $S$ of $\mathbb{R}^n$, namely, the function $f$ defined as the restriction of $\tilde{f}$ to its effective domain. In virtue of this correspondence, it will be assumed that any given convex function $f$ is actually a convex function defined on the whole space, unless otherwise noted. This notational convention has advantages, as it makes unnecessary to explicitly state the effective domains of the convex functions being studied, and it allows to introduce infinite penalties very naturally in optimization problems.

Moreover, if one defines the indicator function $\delta_S$ of a set $S$ as

$$\delta_S(x) = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{otherwise,} \end{cases}$$

then $S$ is a convex set if and only if $\delta_S$ is a convex function. Moreover,

$$\inf_{x \in S} f(x) = \inf_{x \in \mathbb{R}^n} \tilde{f}(x) = \inf_{x \in \mathbb{R}^n} \{f(x) + \delta_S(x)\} \quad (1.3)$$

and

$$f(\bar{x}) \leq f(x) \text{ for every } x \in S \text{ if and only if } \tilde{f}(\bar{x}) \leq \tilde{f}(x) \text{ for every } x \in \mathbb{R}^n \quad (1.4)$$

This fact has importance in optimization, as it allows to introduce infinite penalties convexly.

Under the extended arithmetic rules explained before, we have the discrete version of Jensen’s inequality.

**Theorem 1.1.7.** Let $f$ be an extended-valued function on $\mathbb{R}^n$. Then $f$ is convex if and only if

$$f \left( \sum_{i=1}^l \alpha_i x_i \right) \leq \sum_{i=1}^l \alpha_i f(x_i)$$

whenever $x_i \in \mathbb{R}^n$, $\alpha_i \geq 0$, for $i = \{1, 2, \ldots, l\}$ and $\sum_{i=1}^l \alpha_i = 1$.

It is possible to “convexify” a function $f$, by defining $\text{conv } f$ as the supremum of all the convex functions minorizing $f$. The function $\text{conv } f$ is convex since $\text{epi conv } f = \text{conv}(\text{epi } f)$, and is called the convex hull of the function $f$. For a convex function $f$, $\text{conv } f = f$. 

9
Definition 1.1.8 A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be lower semicontinuous (lsc) at a point $\bar{x}$ if
\[ f(x) \leq \liminf_{x \to \bar{x}} f(x) = \lim_{\delta \to 0} \left[ \inf_{x \in B(\bar{x}, \delta)} f(x) \right]. \]

Theorem 1.1.9. For a function $f : \mathbb{R}^n \to \mathbb{R}$, the following are equivalent

(a) $f$ is lsc,

(b) the epigraph of $f$ is a closed set in $\mathbb{R}^{n+1}$,

(c) for each $\alpha$, the level set $\{ x \mid f(x) \leq \alpha \}$ is closed.

Definition 1.1.10 A function $f$ is proper if $f$ is not the constant function $\infty$ and $f(x) > -\infty$ for every $x$, that is, its epigraph is nonempty and contains no vertical lines.

A stronger, global version of lower semicontinuity is needed in order to handle the conjugation transforms to be presented later.

Definition 1.1.11 For a function $f : \mathbb{R}^n \to \mathbb{R}$, the closure of $f$ is given by
\[ \text{cl} f(x) = \begin{cases} -\infty \text{ if for some } z, \liminf_{y \to z} f(y) = -\infty \\ \liminf_{y \to x} f(y) \text{ otherwise} \end{cases} \]

By definition, $\text{cl} f \leq f$ and if $f_1 \leq f_2$, then $\text{cl} f_1 \leq \text{cl} f_2$. Moreover, for a proper function $f$, $\text{cl} f$ is the supremum of all the lsc functions minorizing $f$, and in this case $\text{epi} \text{cl} f = \text{cl}(\text{epi} f)$.

A function $f$ is said to be closed if $\text{cl} f = f$. Moreover, a proper function is lsc if and only if it is closed; therefore, under properness assumptions, these two terms can be used interchangeably.

A function $f$ is upper semicontinuous (usc) if $-f$ is lsc. A function is continuous if and only if it is both lsc and usc.

Definition 1.1.12 An extended-valued function $g$ on $\mathbb{R}^m$ is concave if $-g$ is convex, that is, if its hypograph
\[ \text{hypo} g = \{(x, \mu) \mid x \in \mathbb{R}^n, \mu \leq g(x)\} \]
is convex in $\mathbb{R}^{n+1}$.

It is possible to carry out a similar set of definitions for concave case to the ones already presented in the convex case, by making the pertinent sign and inequality changes. For simplicity, however, we use the same notation as in the convex case. For instance, for a concave function $g$, we have $\text{dom} g = \{ x \mid g(x) > -\infty \}$ and $\text{cl} g = -\text{cl}(-g)$. In most cases it is clear from the context whether the convex or the concave definition is used, otherwise one specifies that a definition is given “in the concave sense”.

10
1.1.1 Subgradients

**Definition 1.1.13** Let $f$ be an extended-valued proper, closed and convex function on $\mathbb{R}^n$, and let $x_0 \in \mathbb{R}^n$. A vector $v \in \mathbb{R}^n$ is called a subgradient of $f$ at $x_0$ if

$$f(x) \geq f(x_0) + \langle x - x_0, v \rangle, \quad \text{for all } x$$  \hspace{1cm} (1.5)

For $f, x_0$ and $v$ as before, let $\pi(x_0,v,f)$ be the hyperplane that passes through $(x_0, f(x_0))$ and has $v$ as a normal vector. Then the epigraph of $f$ is fully contained in the upper hyperspace determined by $\pi(x_0,v,f)$. In this case, $\pi(x_0,v,f)$ is said to be a supporting hyperplane of $f$ at $x_0$. Conversely, if the last statement is satisfied, then inequality (1.5) holds and so $v$ is a subgradient of $f$ at $x_0$. Summarizing, an extended-valued proper, closed and convex function $f$ has a subgradient at $x_0$ if and only if there exists a supporting hyperplane to $f$ at $x_0$.

The set of all subgradients of $f$ at $x_0$ is called the subdifferential of $f$ at $x_0$, and is denoted by $\partial f(x_0)$. It can be readily checked that $\partial f(x_0) = \emptyset$ for $x_0 \notin \text{dom } f$. One says that $f$ is subdifferentiable at $x_0$ provided $\partial f(x_0) \neq \emptyset$.

The following theorem collects important characteristics of the subdifferential set.

**Theorem 1.1.14.** For $f$ as before, the following hold:

(a) the map $\partial f : x \to \partial f(x)$ is a closed and convex valued mapping,

(b) for each $x \in \text{ri dom } f$, $\partial f(x_0)$ is nonempty,

(c) $\partial f(x_0)$ is a bounded set if and only if $x_0 \in \text{int } \text{dom } f$,

(d) if $f$ is differentiable at $x_0$, then $\partial f(x_0) = \{\nabla f(x_0)\}$,

(e) $f$ is differentiable almost everywhere on $\text{dom } f$ if $\text{int dom } f \neq \emptyset$.

1.1.2 Conjugation and Inf-convolution

**Definition 1.1.15** Let $f$ be an extended-valued function on $\mathbb{R}^n$. The convex conjugate or the Legendre transform of $f$ is the function $f^*$ given by

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{ \langle y, x \rangle - f(x) \}.$$  

where $\langle y, x \rangle$ denotes the usual inner product in $\mathbb{R}^n$.

The function $f^*$ is always convex, regardless of the properties of $f$. One can then reapply the conjugation to $f^*$ to obtain the biconjugate $f^{**} = (f^*)^*$. Moreover,

**Theorem 1.1.16** (Theorem 11.1, [Rock98]). If the function $\text{conv } f$ is proper, then $f^{**} = \text{cl conv } f$.

Thus $f^{**} = f$ if and only if $f$ is proper, lsc, and convex.
Then, for a convex function $f$,

$$\text{cl } f(x) = \sup_v \inf_z \{ \langle x - z, v \rangle + f(z) \}$$

Theorem 1.1.16 is the motivation behind our previous definitions: dual statements between $f$ and $f^*$ are best carried out in the family of proper, lsc, and convex functions.

The following is a list of well-known conjugate dualizing statements for a proper, lsc, and convex function $f$. Here coercivity means faster growth at infinity than a linear function, and argmax (respectively argmin $f$) denotes the set of maximizers (respect. minimizers) of the function $f$.

(a) $f^*$ is everywhere finite if and only if $f$ is coercive,

(b) $f^*$ is differentiable if and only if $f$ is strictly convex,

(c) $y \in \partial f(x)$ if and only if $x \in \partial f^*(y)$ if and only if $f(x) + f^*(y) = \langle x, y \rangle$,

(d) $\partial f(x) = \text{argmax}_y \{ \langle x, y \rangle - f^*(y) \}$ and $\partial f^*(y) = \text{argmax}_x \{ \langle y, x \rangle - f(x) \}$

Notice that, in particular, $\inf f = f^*(0)$, and $\text{argmin } f = \partial f^*(0)$, thus $f$ is bounded below if and only $0 \in \text{dom } f^*$.

**Definition 1.1.17** Let $f$ and $g$ be extended-valued functions on $\mathbb{R}^n$. The inf-convolution (or epi-addition) $f \# g$ is defined as

$$(f \# g)(x) = \inf_u \{ f(u) + g(x - u) \} \quad (1.6)$$

For a positive scalar $\mu$, the epi-multiplication $\mu \ast f$ is given by

$$(\mu \ast f)(x) = \mu f(\mu^{-1}x) \quad (1.7)$$

It can be shown that, for proper functions $f$ and $g$, $\text{epi}(f \# g) = \text{epi } f + \text{epi } g$ and also $\text{epi } \mu \ast f = \mu \ast \text{epi } f$. Moreover, for $f$ and $g$ proper, then $(f \# g)^* = f^* + g^*$, and if $f$ and $g$ are also lsc and convex, with $g$ finite, then $(f + g)^* = f^* \# g^*$. Also, for arbitrary $f$, $(\mu \ast f)^* = \mu f^*$ and $(\mu f)^* = \mu \ast f^*$.

### 1.1.3 Epi-convergence

**Definition 1.1.18** For a sequence of functions $\{f_n\}$ on $\mathbb{R}^n$, we define the lower epi-limit $\text{e-liminf}_n f_n$ and the upper epi-limit $\text{e-limsup}_n f_n$ by

$$\text{e-liminf}_n f_n(x) = \min \left\{ \alpha \in \mathbb{R} \mid \exists x_n \to x \text{ with } \liminf_n f_n(x_n) = \alpha \right\}$$

$$\text{e-limsup}_n f_n(x) = \min \left\{ \alpha \in \mathbb{R} \mid \exists x_n \to x \text{ with } \limsup_n f_n(x_n) = \alpha \right\}$$

The sequence $\{f_n\}$ is said to epi-converge to $f$ provided

$$\text{e-liminf}_n f_n = \text{e-limsup}_n f_n = f$$
The epi-convergence of a sequence \( \{f_n\} \) to \( f \) is equivalent to the convergence of \( \{\text{epi } f_n\} \) to \( \text{epi } f \) in the Painlevé-Kuratowski sense defined in (1.1). It is also equivalent to having

\[
\begin{align*}
\liminf_n f_n(x_n) &\geq f(x) \text{ for every sequence } x_n \to x \\
\limsup_n f_n(x_n) &\leq f(x) \text{ for some sequence } x_n \to x
\end{align*}
\]

If the latter holds for every sequence \( x_n \to x \), then the sequence is said to converge continuously.

Epi-convergence is not implied nor implies pointwise convergence. It is a variational convergence, as shown by the next result

**Proposition 1.1.19** (Epigraphical nesting, Theorem 7.30 in [Rock98]). If the sequence \( f_n \) epi-converges to \( f \), then

(a) \( \limsup_n (\inf f_n) \leq \inf f \),

(b) If \( \text{argmin } f \neq \emptyset \), then given a sequence \( x_n \in \text{argmin}_n f_n \), every accumulation point of \( \{x_n\} \) converges to some point \( x \in \text{argmin } f \).

Moreover, the Fenchel transform is epi-continuous, in the following sense:

**Theorem 1.1.20** (Theorem 11.34, [Rock98]). If the functions \( f_n \) and \( f \) on \( \mathbb{R}^n \) are proper, lsc, and convex, one has

\[
f_n \text{ epi-converges to } f \text{ if and only if } f^*_n \text{ epi-converges to } f^*.
\]

These results show that epi-convergence is a much more powerful and robust sense of convergence than pointwise convergence.

### 1.1.4 Fenchel Duality Theorem

This theorem transforms, under a constraint qualification, a minimization problem into a maximization problem involving the convex conjugation.

**Theorem 1.1.21.** Let \( f \) be a proper convex function on \( \mathbb{R}^n \) and let \( g \) be a proper concave function on \( \mathbb{R}^n \). One has

\[
\inf_x \{f(x) - g(x)\} = \sup_y \{g^*(y) - f^*(y)\}
\]

provided one of the following conditions holds:

(a) \( \text{ri}(\text{dom } f) \cap \text{ri}(\text{dom } g) \neq \emptyset \),

(b) \( f \) and \( g \) are lsc and \( \text{ri}(\text{dom } g^*) \cap \text{ri}(\text{dom } f^*) \neq \emptyset \).

**Remark 1.1.22** The constraint qualification in (a) has a nice geometrical interpretation, namely, the convex sets \( \text{dom } f \) and \( \text{dom } g \) cannot be properly separated, which means that it is not possible to find a hyperplane \( \pi \) such that each set is strictly contained on each hyperspace determined by \( \pi \).
The next result provides a criterion to find the minimizers and maximizers in equation (1.8).

**Theorem 1.1.23.** Let $f$ be a proper, lsc, and convex function on $\mathbb{R}^n$ and let $g$ be a proper, usc, and concave function on $\mathbb{R}^n$. In order that $x$ and $u$ to be vectors such that

$$f(x) - g(x) = \inf (f - g) = \sup (g^* - f^*) = g^*(u) - f^*(u)$$

(1.9)

it is necessary and sufficient that $x$ and $u$ satisfy

$$u \in \partial f(x), \quad x \in \partial g^*(u)$$

(1.10)

**Proof.** For each $z$ and $w$ in $\mathbb{R}^n$, Fenchel’s inequality for convex and concave conjugates ensures that

$$\langle z, w \rangle \leq f(z) + f^*(w)$$

and

$$\langle w, z \rangle \geq g^*(w) + g(z).$$

So, $f(z) - g(z) \geq g^*(w) - f^*(w)$ for every $z$ and $w$, thus

$$\inf (f - g) \geq \sup (g^* - f^*).$$

(1.11)

The conditions (1.10) are equivalent to

$$f(x) + f^*(u) = \langle x, u \rangle = g^*(u) + g(x).$$

In virtue of (1.11), this is equivalent to

$$\inf (f - g) = \sup (g^* - f^*),$$

as required. \qed

## 1.2 Saddle Functions

The goal of this section is to give a concise presentation of the main properties of saddle functions and describe their connection with convex functions in optimization and duality.

**Definition 1.2.1** A function $K : C \times D \to \mathbb{R}$, with $C \times D \subset \mathbb{R}^m \times \mathbb{R}^n$, is a saddle function (or concave-convex) provided

(i) for each $y \in D$, $K(\cdot, y)$ is concave and

(ii) for each $x \in C$, $K(x, \cdot)$ is convex.

**Remark 1.2.2** A similar definition applies to convex-concave functions. In other references, both concave-convex and convex-concave functions are called saddle functions. However, in this thesis, we will use the term saddle function only to denote concave-convex function.
A typical example of saddle function is given by $K(x, y) = -f(x) + g(y)$, where $f$ and $g$ are real-valued convex functions.

Saddle functions appear naturally in the duality theory for convex optimization problem, as it will be detailed later.

The theory of saddle functions is more subtle than what might appear at first, due to the possibility of extended-real valuedness, and a careful treatment is required. Consider for example the problem of how to extend a saddle function $K$ on $C \times D$ to a saddle function on $\mathbb{R}^m \times \mathbb{R}^n$. The convex functions $K(x, \cdot)$ can be extended to values of $y$ not in $D$ by assigning them $+\infty$ and similarly, redefining the concave functions $K(\cdot, y)$ as $-\infty$ at those $x \notin C$. One possibility is then to define

$$K_{\text{low}}(x, y) = \begin{cases} 
K(x, y) & \text{if } x \in C, y \in D \\
+\infty & \text{if } x \in C, y \notin D \\
-\infty & \text{if } x \notin C 
\end{cases}$$

The function $K_{\text{low}}$ is called the lower extension of $K$. Moreover, equally possible is to extend $K$ by means of an upper extension

$$K_{\text{up}}(x, y) = \begin{cases} 
K(x, y) & \text{if } x \in C, y \in D \\
-\infty & \text{if } x \notin C, y \in D \\
+\infty & \text{if } y \notin D 
\end{cases}$$

The discrepancy between extensions occurs at the points $x \notin C, y \notin D$, and it is not clear which one (if there is one) should be preferred. The non-uniqueness of the saddle extensions (except of course in the case where the saddle function is finite everywhere) gives rise to consider saddle functions as equivalence classes instead of single objects. In order for this theory to attain cohesiveness, this equivalence relation should be compatible in some sense with the duality operations and the optimization principles in which saddle functions intervene naturally. Rockafellar gave such a definition in [Rock64] and later refined it in [Rock70]. A particular example of an equivalent pair is given by the lower and upper extensions of a saddle function, and this fact is what makes any of the two extensions appropriate.

From this point on, unless explicitly stated otherwise, we assume that every saddle function $K$ is actually defined on $\mathbb{R}^m \times \mathbb{R}^n$, by substituting $K$ with either its lower or its upper extension if needed.

**Definition 1.2.3** For a saddle function $K$, the function $\text{cl}_1 K$ is the function obtained by taking, for each $y$, the concave closure of $K(\cdot, y)$. Therefore,

$$\text{cl}_1 K(\cdot, y) = \text{cl} K(\cdot, y) \text{ in the concave sense}$$

Similarly, the function $\text{cl}_2 K$ is the function obtained by taking, for each $x$, the convex closure of $K(x, \cdot)$. Thus,

$$\text{cl}_2 K(x, \cdot) = \text{cl} K(x, \cdot) \text{ in the convex sense}$$
For a saddle function $K$, the functions $c_1 K$ and $c_2 K$ are also saddle functions that satisfy $c_2 K(x, y) \leq K(x, y) \leq c_1 K(x, y)$ for each $x$ and $y$. Furthermore, the crossed closures $\overline{K} = c_2 c_1 K$ and $\underline{K} = c_1 c_2 K$, known as the lower and upper closures of $K$ respectively, are saddle functions as well. The following example shows that the two crossed closures do not necessarily match.

**Example 1.2.4** On $\mathbb{R} \times \mathbb{R}$, let

$$K(x, y) = \begin{cases} 0 & \text{if } xy = 0 \\ +\infty & \text{if } xy > 0 \\ -\infty & \text{otherwise} \end{cases}$$

$K$ is in fact a saddle function: for $x > 0$ (respect. $x < 0$) the epigraph of $K(x, \cdot)$ is the union of the left-hand (respect. right-hand) side half-plane and the ray $\{(0, y) \mid y \geq 0\}$, thus is a convex set in $\mathbb{R}^2$. Moreover, $K(0, y) = 0$ for each $y \in \mathbb{R}$ and consequently $K(x, \cdot)$ is convex for all values of $x$. A similar verification can be performed for the functions $K(\cdot, y)$ to prove their concavity.

Moreover,

$$c_1 K(x, y) = \begin{cases} 0 & \text{if } y = 0 \\ +\infty & \text{otherwise} \end{cases}$$

$$c_2 K(x, y) = \begin{cases} 0 & \text{if } x = 0 \\ -\infty & \text{otherwise} \end{cases}$$

and therefore,

$$c_2 c_1 K = c_1 K \neq c_2 K = c_1 c_2 K$$

**Example 1.2.5** [Rock70] Let $K$ be any of the extensions of the finite saddle function $x^y$, $0 < x < 1, 0 < y < 1$. In this case, we have

$$\overline{K}(x, y) = c_1 c_2 K = \begin{cases} x^y & \text{if } x \in [0, 1], y \in [0, 1], (x, y) \neq (0, 0) \\ 1 & \text{if } x = 0, y = 0 \\ +\infty & \text{if } x \in [0, 1], y \notin [0, 1] \\ -\infty & \text{if } x \notin [0, 1], y \in [0, 1] \\ +\infty & \text{if } x \notin [0, 1], y \notin [0, 1] \end{cases}$$

On the other hand,

$$\underline{K}(x, y) = c_2 c_1 K = \begin{cases} x^y & \text{if } x \in [0, 1], y \in [0, 1], (x, y) \neq (0, 0) \\ 0 & \text{if } x = 0, y = 0 \\ +\infty & \text{if } x \in [0, 1], y \notin [0, 1] \\ -\infty & \text{if } x \notin [0, 1], y \in [0, 1] \\ -\infty & \text{if } x \notin [0, 1], y \notin [0, 1] \end{cases}$$

Notice that the function $\overline{K}$ is lsc in $y$, while $\underline{K}$ is usc in $x$. 

16
Next, we present the equivalence relation for saddle functions, which is based on their partial closures.

**Definition 1.2.6** Two saddle functions $H$ and $K$ are equivalent provided

$$\text{cl}_1 H = \text{cl}_1 K$$

and

$$\text{cl}_2 H = \text{cl}_2 K$$

We denote the equivalence class of $K$ by $[K]$.

Notice that, by definition, equivalent saddle functions have the same lower closure and the same upper closure, even though the two might not agree.

**Example 1.2.7** For a saddle function $K : C \times D \to \mathbb{R}$, where $C$ and $D$ are proper subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, the lower extension $K_{\text{low}}^\up$ and the upper extension $K_{\text{up}}^\down$ are equivalent. In fact,

$$\text{cl}_2 K_{\text{low}}(x, \cdot) = \text{cl}_2 K_{\text{up}}(x, \cdot) = \begin{cases} 
\text{cl} \{ K(x, y), \ y \in D \} & x \in C \\
+\infty, & y \notin D \\
-\infty, & x \notin C
\end{cases}$$

and a similar formula can be found for $\text{cl}_1 K_{\text{low}}$ and $\text{cl}_1 K_{\text{up}}$.

### 1.2.1 Convex and Concave Parents

This section explores how saddle functions can be associated with jointly convex or concave functions. When the functions involved are closed, this association is actually a one-to-one correspondence, so in this case the unique convex and concave functions associated with the saddle function are called its convex and concave parents.

Given a (nonnecessarily closed) jointly convex function $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, its convex partial conjugate is given by

$$K(x, y) = (f(x, \cdot))^*(y) = \sup_v \{ \langle v, y \rangle - f(x, v) \}.$$  

(1.12)

Since the function $(f(x, \cdot))^*$ is convex and closed, then $K(x, \cdot)$ is convex and closed as well. Also, $K(\cdot, y)$ is concave on $\mathbb{R}^n$, for $y$ fixed, because

$$-K(x, y) = \inf_v h(x, v)$$

for the jointly convex function $h(x, v) = f(x, v) - \langle v, y \rangle$. This shows that $K$ is a saddle function. Moreover, we have that for each $x$, the biconjugate $(f(x, \cdot))^\text{**}$ equals the (convex) closure of $f(x, \cdot)$, $\text{cl} f(x, \cdot)$, and the following formula holds

$$\text{cl} f(x, \cdot)(v) = K(x, \cdot)^*(v) = \sup_y \{ \langle v, y \rangle - K(x, y) \}.$$
A similar reasoning can be applied, in reverse, by starting with a concave-convex function $K$, its convex partial conjugate gives defines a function $f$ by

$$f(x, v) = K(x, \cdot)^*(v) = \sup_y \{\langle v, y \rangle - K(x, y)\} \quad (1.13)$$

that is convex and satisfies both that $f(x, \cdot)$ is closed and the equation

$$\text{cl}_2 K(x, y) = \sup_v \{\langle v, y \rangle - f(x, v)\}. \quad (1.14)$$

Analogously, all the results in the convex case can be carried for the concave case. In fact, given a (nonnecessarily closed) jointly concave function $g$, its concave partial conjugate is a concave-convex function $L$ is given by

$$L(x, y) = (g(\cdot, y))^*(x) = \inf_w \{\langle w, x \rangle - g(w, y)\} \quad (1.15)$$

and this function satisfies both $\text{cl}_1 L = L$ and the formula

$$\text{cl} g(\cdot, y)(w) = L(\cdot, y)^*(w) = \inf_x \{\langle w, x \rangle - L(x, y)\}.$$

Also, starting with $L$ concave-convex, the concave partial conjugate gives defines a function $g$ by

$$g(w, y) = L(\cdot, y)^*(w) = \inf_x \{\langle w, x \rangle - L(x, y)\} \quad (1.16)$$

that is concave and satisfies both that $g(\cdot, y)$ is closed and the equation

$$\text{cl}_1 L(x, y) = \inf_w \{\langle w, x \rangle - g(w, y)\}. \quad (1.17)$$

Are both the convex and the concave approaches related? That is, given $f$ convex, is it possible to choose $g$ in such a way that $K$ defined as in (1.12) and $L$ defined as in (1.15) are, say, equivalent to each other? The answer is positive and a formula for such $g$ is given by

$$g(w, y) = -f^*(-w, y). \quad (1.18)$$

Let us verify this statement. We have that

$$g(w, y) = -f^*(-w, y)
= -\sup_{x, v} \{\langle v, y \rangle + \langle x, -w \rangle - f(x, v)\}
= -\sup_x \left\{\langle x, -w \rangle + \sup_v \{\langle v, y \rangle - f(x, v)\} \right\}
= -\sup_x \{\langle x, -w \rangle + K(x, y)\} \quad (\text{using (1.12)})
= \inf_x \{\langle x, w \rangle - K(x, y)\}$$
Therefore, $K$ satisfies (1.16) with $L$ changed by $K$, so according to (1.17),
\[
\text{cl}_1 K(x, y) = \inf \{ \langle w, x \rangle - g(w, y) \} = L(x, y) = \text{cl}_1 L(x, y).
\] (1.19)

Symmetrically, given a concave function $g$ and a saddle function $L$ as in (1.15), for $f$ defined as
\[
f(x, v) = -g^*(x, -v)
\] (1.20)
and $K$ given by (1.12), we have that
\[
\text{cl}_2 L(x, y) = \sup \{ \langle v, y \rangle - f(x, v) \} = K(x, y) = \text{cl}_2 K(x, y).
\] (1.21)

We conclude that in order for a convex function $f$ and a concave function $g$ to generate equivalent saddle functions it is necessary that equations (1.20) and (1.18) hold. This condition can be shown to be sufficient as well, and the proof is just essentially reversing the steps.

Now, assume that $f$ is convex, $g$ is concave, and they satisfy (1.20) and (1.18). We want to fully characterize such $f$ and $g$. To do this, let $h(x, v) = -g(x, -v) = f^*(-x, -v)$. Then
\[
f(x, v) = -g^*(x, -v)
= h^*(-x, -v) \quad (\text{by definition of concave conjugate})
= (f^*(x, v))^*
= (\text{cl } f)(x, v)
\]
Therefore, $f$ is closed and analogously, $g$ is also closed. Conversely, given a closed convex function $f$ and defining $g$ as in (1.18), then equation (1.20) holds and $f$ and $g$ generate equivalent saddle functions $K$ and $L$. In this case, $f$ and $g$ are called respectively the convex and the concave parents of $[K]$, and the equivalence class is completely determined by $f$. The concept of parent is well-defined, because every element in $[K]$ has $f$ and $g$ as parents, as a consequence of equations (1.19) and (1.21), and their uniqueness follows from the fact that $f$ and $g$ are closed.

### 1.2.2 Closure and Properness

Let us now expand the concept of closedness given previously for convex functions to saddle functions.

**Definition 1.2.8** A saddle function $K$ is closed provided $\text{cl}_1 K$ and $\text{cl}_2 K$ are both equivalent to $K$.

A characterization of $K$ being closed is that
\[
\text{cl}_1 \text{cl}_2 K = \text{cl}_1 K
\]
\[
\text{cl}_2 \text{cl}_1 K = \text{cl}_2 K
\]
Clearly, if a saddle function is closed, then every member of its equivalence class is closed as well.

The following example provides a connection between closed convex functions and closed saddle functions. This will be expanded in Theorem 1.2.11

**Example 1.2.9** Let $f$ be a closed convex function and consider the equivalence class that it generates. Then each saddle function in this equivalence class is closed. To show this, take a saddle function $M$ in the equivalence class, and let $K$ be the saddle function defined by (1.12) and $L$ defined by (1.15). Then, as a consequence of the discussion following equation (1.17), $K$ and $L$ belong to the equivalence class generated by $f$, so they both are equivalent to $M$. Furthermore,

$$
\text{cl}_1 M = L \\
\text{cl}_2 M = K
$$

and so $\text{cl}_1 M$ and $\text{cl}_2 M$ must be equivalent to $K$, thus in turn equivalent to $M$. This, by definition, means that $M$ is closed.

The concept of properness of a convex function is important in that it keeps certain degenerate examples from appearing, thus ensuring stronger duality results. We now turn to the definition of properness for saddle functions. First it is necessary to find a suitable definition for the domain of a saddle function. Looking back into the convex (respect. concave) case, for a nontrivial convex (respect. concave) function $f$, $\text{dom } f$ satisfies the following two conditions:

(i) it is a convex set,

(ii) it contains all the points at which $f$ is finite.

In optimization problems, condition (i) is definitely important, due to the properties of convex sets explained in Section 1.1 On the other hand, condition (ii) is convenient because in particular $\text{dom } f$ will then contain the set of minimizers of $f$, namely, $\text{argmin } f = \{ x \mid f(x) = \min f \}$. Taking this into account, an approach would be to ensure that, for a saddle function $K$, $\text{dom } K$ satisfies

(i) $\text{dom } K$ is the product of convex sets,

(ii) $\text{dom } K$ contains all the points at which $K$ is finite.

A consequent definition would be

$$
\text{dom } K = \{ (x, y) \mid -\infty < K(x, y) < \infty \}
$$

This definition clearly satisfies the finiteness requirement; however, in general, it fails to satisfy the convexity in condition (i). For instance, for the saddle function $K$ in Example 1.2.4, according to this definition, $\text{dom } K = \{ (x, 0) \mid x \in \mathbb{R} \} \cup \{ (0, y) \mid y \in \mathbb{R} \}$, which cannot be written as a product of two convex sets in $\mathbb{R}$.

Rockafellar approaches the issue to require a more substantial property than mere finiteness.
(i) \( \text{dom} \, K \) is a product of convex sets,

(ii) \( \text{dom} \, K \) might not contain all the points at which \( K \) is finite, but at least those which are fundamental in optimization (in this case the saddle points, a concept we will address in detail later)

and his definition, which fulfills these two last requirements, is given by

\[
\text{dom} \, K = \text{dom}_1 \, K \times \text{dom}_2 \, K
\]

where

\[
\text{dom}_1 \, K = \{ x \mid K(x, y) > -\infty \text{ for every } y \} = \bigcap_y \text{dom} \, K(\cdot, y)
\]

\[
\text{dom}_2 \, K = \{ y \mid K(x, y) < \infty \text{ for every } x \} = \bigcap_x \text{dom} \, K(x, \cdot)
\]

Having an extension of the concept of the domain for a saddle function, one can now define properness for a saddle function.

**Definition 1.2.10** A saddle function \( K \) is proper provided \( \text{dom} \, K \neq \emptyset \).

The saddle function in Example 1.2.4 is proper: in fact, its domain is the singleton \((0, 0)\). An example of a saddle function that is not proper is

\[
K(x, y) = \begin{cases} 
0 & \text{if } x > y \\
+\infty & \text{otherwise},
\end{cases}
\]

since \( \text{dom}_2 \, K = \bigcap_x \text{dom} \, K(x, \cdot) = \bigcap_x (-\infty, x) = \emptyset \).

The following questions are important to address

(i) What is the connection between the lower and upper extensions of a saddle function and its equivalence class?

(ii) For which values of \( x \) and \( y \) can we guarantee that \( K(x, y) = \text{cl}_1 \, K(x, y) = \text{cl}_2 \, K(x, y) \)?

(iii) How is the domain of a saddle function related to that of its convex and concave parent?

The following theorem and lemma address these questions, and groups together the results previously shown. The theorem exhibits a one-to-one correspondence between closed jointly convex functions and closed saddle functions, and characterizes the equivalence class \([K]\) as an “interval” having as lower bound the lower extension \( \overline{K} \) and as upper bound the upper extension \( \bar{K} \), as detailed below.
Theorem 1.2.11 (Theorem 34.2, [Rock70]). Given a closed and jointly convex function \( f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R} \), let

\[
\overline{K}(x, y) = \sup_v \{ \langle v, y \rangle - f(x, v) \} \tag{1.22}
\]
\[
\underline{K}(x, y) = \inf_w \{ \langle w, x \rangle + f^*(-w, y) \} \tag{1.23}
\]

and consider the collection

\[
\Lambda(f) = \{ K : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}, K \text{ is a saddle function and satisfies } \underline{K} \leq K \leq \overline{K} \}.
\]

Then \( \Lambda(f) \) is an equivalence class of saddle functions (the one containing both \( \underline{K} \) and \( \overline{K} \)) and each member of this equivalence class is closed.

Conversely, given an equivalence class of saddle functions \([K]\), there exists a unique closed jointly convex function \( f \) (namely, the convex parent) such that \([K] = \Lambda(f)\). For any \( K \in \Lambda(f)\),

\[
\text{cl}_1 K = \overline{K}, \quad \text{cl}_2 K = \underline{K},
\]
\[
f(x, v) = \sup_y \{ \langle v, y \rangle - K(x, y) \} \tag{1.24}
\]
\[
-f^*(-w, y) = \inf_x \{ \langle x, w \rangle - K(x, y) \} \tag{1.25}
\]

Moreover,

\[
K(x, y) = \sup_v \{ \langle v, y \rangle - f(x, v) \} = \inf_w \{ \langle w, x \rangle + f^*(-w, y) \}
\]

if \( x \in \text{ri}(\text{dom } f(\cdot, \hat{y})) \) for some \( \hat{y} \) or if \( y \in \text{ri}(\text{dom } (-f^*(-\hat{x}, \cdot))) \) for some \( \hat{x} \).

Corollary 1.2.12. The formulas

\[
K(x, y) = \sup_v \{ \langle v, y \rangle - f(x, v) \} \tag{1.26}
\]
\[
f(x, v) = \sup_y \{ \langle v, y \rangle - K(x, y) \} \tag{1.27}
\]

define a one-to-one correspondence between the closed convex functions \( f \) on \( \mathbb{R}^m \times \mathbb{R}^n \) and the closed saddle functions \( K \) on \( \mathbb{R}^m \times \mathbb{R}^n \) satisfying

\[
\text{cl}_2 \text{cl}_1 K = K. \tag{1.28}
\]

Each equivalence class of closed saddle functions on \( \mathbb{R}^m \times \mathbb{R}^n \) contains exactly one \( K \) satisfying (1.28).

Proof. For a given \( f \) that satisfies the hypothesis, define \( K \) by formula (1.26). Then, by the previous theorem, \( K = \overline{K} \) is the lower extension of its equivalence class, \( f \) can be recovered from \( K \) using formula (1.24) and \( K \) satisfies \( \text{cl}_2 \text{cl}_1 K = \text{cl}_2 \overline{K} = \underline{K} \), as desired. The converse follows similarly. \( \square \)
Lemma 1.2.13. Let $K$ be a closed saddle function and let $f$ and $g$ be their parents. Then

$$\text{dom}_1 K = \bigcup_{y \in \mathbb{R}^n} \text{dom} f(\cdot, y)$$

(1.29)

$$\text{dom}_2 K = \bigcup_{x \in \mathbb{R}^m} \text{dom} g(x, \cdot)$$

(1.30)

Proof. Suppose that $u \notin \text{dom}_1 K$. By definition there exists $\tilde{y} \in \mathbb{R}^n$ such that $u$ is not in the domain of the concave function $K(\cdot, \tilde{y})$. Thus, $K(u, \tilde{y}) = -\infty$. Then, according to (1.27), $f(u, y) = \sup_w \{\langle y, w \rangle - K(u, w)\} \geq \langle v, \tilde{y} \rangle - K(x, \tilde{y}) = +\infty$, for every $y \in \mathbb{R}^n$. Hence, $u \notin \bigcup_{y \in \mathbb{R}^n} \text{dom} f(\cdot, y)$. For the reverse inclusion, let $u \notin \bigcup_{y \in \mathbb{R}^n} \text{dom} f(\cdot, y)$. Then, for every $y$, $u \notin \text{dom} f(\cdot, y)$, that is, $f(u, y) = +\infty$, for every $y$. Then, according to (1.26), $K(u, w) = \sup_y \{\langle y, w \rangle - f(u, y)\} = -\infty$, for every $w \in \mathbb{R}^n$. Hence, $u \notin \text{dom}_1 K$. Equation (1.30) follows by symmetry.

Corollary 1.2.14. A closed saddle function $K$ is proper if and only if its convex parent $f$ is proper.

Proof. Suppose that convex parent $f$ is not proper. Then $f \equiv +\infty$ or $f \equiv -\infty$, and is such case, the conjugation formulas 1.22 and 1.23 imply that $[K] \equiv \{-\infty\}$ or $[K] \equiv \{+\infty\}$, which are not proper saddle functions. On the other hand, if the convex parent $f$ of $K$ is closed and proper, then by definition $\text{dom} f \neq \emptyset$, so there exists $u$ and $v$ satisfying $\text{dom} f(\cdot, v) \neq \emptyset$ and $\text{dom} f(u, \cdot) \neq \emptyset$ which in turn means that the set unions in (1.29) and (1.29) are nonempty. Thus $\text{dom} K$ is the product of two nonempty sets and this implies that $K$ is proper.

Thus, the only closed saddle functions which are not proper are $K(x, y) = +\infty$ and $K(x, y) = -\infty$. This is the reason why when we restrict our attention to only proper closed saddle function. As a matter of fact, for the latter type of saddle functions, there is another simple characterization of equivalence that we include next.

Definition 1.2.15 The kernel of a saddle function $K$ is the restriction of $K$ to $\text{ri}(\text{dom} K) = \text{ri}(\text{dom}_1 K) \times \text{ri}(\text{dom}_2 K)$.

As a consequence of Theorem 1.2.11, the kernel of a saddle function is finite and it is useful as a condition for saddle function equivalence, as can be seen below.

Theorem 1.2.16 (Theorem 34.4, [Rock70]). Two proper and closed saddle functions $K$ and $L$ are equivalent to each other if and only if they have the same kernel.

Although there are a few technical details to verify in order to prove this theorem, it is based on the last statement in Theorem 1.2.11 and the fact that a convex (and also concave) function is fully determined by its values on the relative interior of its domain.

Remark 1.2.17 Let $K$ be a saddle function such that for every $y$, $K(\cdot, y)$ is usc and for every $x$, $K(x, \cdot)$ is lsc. Then $K$ is closed.
1.2.3 Saddle Points and Dual Saddle Functions

So far, the approach used to understand equivalence classes of saddle function has relied solely in terms of closure operations. Now, the idea is to highlight that equivalent closed saddle functions determine the same regularized minimax problems. Actually, when Rockafellar first presented the equivalence class concept in [Rock64], equivalent saddle functions are those that are minimax equivalent extensions of each other.

The properties of convex functions make them an essential in minimization problems, whereas concave functions appear in maximization problems. Since saddle functions are a combination between these two type of functions, it is natural to study maxima values in the concave variables and minima in the convex ones. More precisely, let $K$ be a closed proper saddle function. Consider its lower saddle value

$$\sup_x \inf_y K(x,y)$$

and its upper saddle value

$$\inf_y \sup_x K(x,y)$$

Clearly, the lower saddle value is always less than or equal to the upper saddle value. When they both match, that common quantity is called the saddle value of $K$. In general, this value (when it exists) is not necessarily attained. Also, a pair $(\bar{x}, \bar{y})$ is called a saddle point for $K$ provided

$$K(x, \bar{y}) \leq K(x, y) \leq K(\bar{x}, y).$$

If a saddle point $(\bar{x}, \bar{y})$ exists, then $K(\bar{x}, \bar{y})$ is the saddle value for $K$.

**Theorem 1.2.18.** Let $K$ and $L$ be equivalent closed saddle functions. Then for each $w \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, the saddle functions $L_{w,v} = L(x, y) - \langle x, w \rangle - \langle y, v \rangle$ and $K_{w,v}(x, y) = K(x, y) - \langle x, w \rangle - \langle y, v \rangle$ have the same lower saddle and upper saddle values, and the same saddle points (if they exist).

**Proof.** First, let us verify that $K_{w,v}$ and $L_{w,v}$ have the upper saddle value. Let $f$ and $g$ be the convex and concave parents of the equivalence class containing $K$ and $L$. We will show that for each $w$ and $v$, the upper saddle value of the functions $K_{w,v}$ and $L_{w,v}$ equals the common value $(f(\cdot, v))^*(-w)$. We have that

$$\sup_x \inf_y \{K(x, y) - \langle x, w \rangle - \langle y, v \rangle\} = -\inf_x \sup_y \{\langle x, w \rangle + \langle y, v \rangle - K(x, y)\}$$

$$= -\inf_x \{\langle x, w \rangle + f(x, v)\}$$

$$= -(f(\cdot, v))^*(-w)$$

Since $f$ is also the convex parent of $L$, all the steps just shown are justified if $K$ is changed by $L$. Thus, we conclude that $K$ and $L$ have the upper saddle value. An analogous reasoning applies to show the result for lower saddle values.
Now, suppose that \((\bar{x}, \bar{y})\) is a saddle point of \(K_{w,v}\). Then the lower and upper saddle values of \(K_{w,v}\) agree and equal \(K_{w,v}(\bar{x}, \bar{y})\). Since \(g\) is the concave parent of both \(K\) and \(L\), it is true that
\[
g(w, \bar{y}) = \inf_x \{\langle x, w \rangle - L(x, \bar{y})\} = \inf_x \{\langle x, w \rangle - K(x, \bar{y})\}
\]
So,
\[
\sup_x \{L(x, \bar{y}) - \langle x, w \rangle\} = \sup_x \{K(x, \bar{y}) - \langle x, w \rangle\}
\]
Thus,
\[
L_{w,v}(\bar{x}, \bar{y}) \leq \sup_x \{L(x, \bar{y}) - \langle x, w \rangle - \langle v, \bar{y} \rangle\} = \sup_x \{K(x, \bar{y}) - \langle x, w \rangle - \langle v, \bar{y} \rangle\} \leq K_{w,v}(\bar{x}, \bar{y})
\]
where the last inequality holds by definition of saddle point. A symmetric argument shows the other inequality, and therefore \((\bar{x}, \bar{y})\) is a saddle point of \(L\). By switching the roles of \(K\) and \(L\) in the previous discussion, we conclude that \(K_{w,v}\) and \(L_{w,v}\) have the same saddle points, as desired. 

**Corollary 1.2.19.** Equivalent closed saddle functions have the same saddle values and saddle points, if they exist.

**Proof.** This follows applying the previous lemma for the values \(w = v = 0\). 

The fact that the saddle values for saddle functions are preserved under linear perturbations is the key for the duality theory that is presented next.

**Definition 1.2.20** For each saddle function \(K\) on \(\mathbb{R}^m \times \mathbb{R}^n\), the lower conjugate \(K^\ast\) and the upper conjugate \(\overline{K}^\ast\) are defined as follows
\[
K^\ast(w, v) = \sup_y \inf_x \{\langle x, w \rangle + \langle y, v \rangle - K(x, y)\} \quad (1.31)
\]
\[
\overline{K}^\ast(w, v) = \inf_x \sup_y \{\langle x, w \rangle + \langle y, v \rangle - K(x, y)\} \quad (1.32)
\]
Both conjugates are saddle functions, as can be verified directly from the definition. Moreover, the function \(-\overline{K}^\ast(w, v)\) equals the upper saddle value of the linearly perturbed saddle function \(K_{w,y}\). The proof of Theorem 1.2.18 implies that the values of \(\overline{K}^\ast\) and \(K^\ast\) do not depend on the representative chosen for equivalence class \([K]\). The notation suggests that both \(\overline{K}^\ast\) and \(K^\ast\) are the lower and upper closures for a equivalence class of saddle functions. Let us show that this actually
holds. By definition of closure and (1.32),
\[
\text{cl}_2 K^* (w, v) = \sup_{u} \inf_{z} \left\{ \langle v - z, u \rangle + K^* (w, z) \right\} \\
= \sup_{u} \inf_{z} \left\{ \langle v - z, u \rangle + \inf_{y} \sup_{x} \{ \langle x, w \rangle + \langle y, z \rangle - K(x, y) \} \right\} \\
= \sup_{u} \inf_{y} \sup_{x} \{ \langle x, w \rangle + \langle v, u \rangle + \langle y - u, z \rangle - K(x, y) \} \\
= \sup_{u} \inf_{y} \inf_{x} \sup_{z} \{ \langle y - u, z \rangle - \inf_{w} \sup_{y} \{ \langle x, w \rangle - K(w, z) \} \} \\
= \sup_{u} \inf_{y} \inf_{x} \sup_{z} \{ \langle v, u \rangle - \text{cl}_2 K(x, u) \} \\
= \sup_{u} \inf_{x} \{ \langle x, w \rangle + \langle v, u \rangle - K(x, u) \} \\
= K^* (w, v)
\]
where the \( \text{cl}_2 \) can be dropped because by assumption \( K \) is closed and so \( \text{cl}_2 K \) and \( K \) are equivalent and thus determine the same conjugate saddle. Similarly, \( \text{cl}_1 K^* = K^* \), so \( \text{cl}_2 \text{cl}_1 K^* = \text{cl}_2 K^* = K^* \). In virtue of Corollary 1.2.12, \( K^* \) determines an equivalence class of closed saddle functions, called the conjugate class of \([K]\) for which \( K^* \) and \( \overline{K}^* \) are the lower and upper closures, respectively. The same corollary helps us determine its convex parent \( \tilde{f} \). If we let \( g \) be the concave parent of \( K \), then \( \tilde{f} \) is given by formula 1.27
\[
\tilde{f} (x, v) = \sup_{y} \{ \langle v, y \rangle - K^* (x, y) \} \\
= \sup_{y} \inf_{z} \{ \langle v - z, y \rangle - \inf_{w} \{ \langle x, w \rangle - K(w, z) \} \} \\
= \sup_{y} \inf_{z} \{ \langle v - z, y \rangle - g(x, z) \} \\
= \text{cl} (-g(x, \cdot)) (v) \\
= - \text{cl} (g(x, \cdot)) (v) \\
= -g(x, v)
\]
Having determined the convex parent of the equivalence class \([K^*]\), we can invoke the structural theorem 1.2.11 and summarize our findings in the next theorem.

**Theorem 1.2.21.** For a given closed saddle function \( K \), the lower conjugate \( K^* \) and the upper conjugate \( \overline{K}^* \) determine an equivalence class of closed saddle functions, whose lower closure is \( \overline{K}^* \) and upper closure is \( K^* \), and has as convex and concave parents \(-g \) and \(-f \), where \( g \) and \( f \) are the convex and concave parents of \( K \), respectively. Equivalent saddle functions determine the same conjugate equivalence class. A saddle equivalence class is proper and closed if and only if its conjugate class is proper and closed.
1.2.4 Subgradients

Definition 1.2.22 Let $K$ be a saddle function. The set of subgradients of $K$ at a point $(x, y)$, written $\partial K(x, y)$, is

$$\partial K(x, y) = \partial_1 K(x, y) \times \partial_2 K(x, y)$$

where $\partial_1 K(x, y)$ is the set of all subgradients of the concave function $K(\cdot, y)$ at $x$ and $\partial_2 K(x, y)$ is the set of all subgradients of the convex function $K(x, \cdot)$ at $y$.

Accordingly, $(\bar{u}, \bar{v}) \in \partial K(\bar{x}, \bar{y})$ if and only if

$$K(x, \bar{y}) \leq K(x, y) + \langle x - \bar{x}, \bar{u} \rangle \text{ for all } x$$

(1.33)

$$K(\bar{x}, y) \geq K(\bar{x}, \bar{y}) + \langle y - \bar{y}, \bar{v} \rangle \text{ for all } y$$

(1.34)

A direct verification shows that if $(\bar{x}, \bar{y}) \in \text{ri dom } K = \text{ri dom}_1 K \times \text{ri dom}_2 K$, then $K$ has at least a subgradient at $(\bar{x}, \bar{y})$. The following theorem connects saddle conjugation, saddle points and subgradients in a very elegant fashion.

Theorem 1.2.23 (Theorem 6, [Rock64]). Let $K$ and $L$ be proper and closed saddle functions which are conjugate to each other. Then the following are equivalent

(a) $(\bar{u}, \bar{v}) \in \partial K(\bar{x}, \bar{y})$

(b) $(\bar{x}, \bar{y}) \in \partial L(\bar{u}, \bar{v})$

(c) $(\bar{x}, \bar{y})$ is a saddle point of $K_{\bar{u}, \bar{v}}(x, y) = K(x, y) - \langle x, \bar{u} \rangle - \langle y, \bar{v} \rangle$

(d) $(\bar{u}, \bar{v})$ is a saddle point of $L_{\bar{x}, \bar{y}}(u, v) = L(u, v) - \langle \bar{x}, u \rangle - \langle \bar{y}, v \rangle$

Proof. Assume (a). Then formulas (1.33) and (1.34) hold and they can be combined to obtain

$$K_{\bar{u}, \bar{v}}(x, \bar{y}) \leq K_{\bar{u}, \bar{v}}(\bar{x}, y) \leq K_{\bar{u}, \bar{v}}(\bar{x}, \bar{y}),$$

which is the statement in (c). All the steps can be reversed, so we conclude that (a) and (c) are equivalent. By symmetry, (b) and (d) are also equivalent. So assume (a) again. Consider the upper closure $\overline{L}$ of the equivalence class containing $L$. Then $L$ is bounded from above by $\overline{L}$ and by the equation (1.32) defining $\overline{L}$, it is verified that for all $u$,

$$L(u, \bar{v}) \leq \overline{L}(u, \bar{v}) \leq \sup_y \{ \langle x, u \rangle + \langle y, \bar{v} \rangle - K(\bar{x}, y) \}$$

$$\leq \langle \bar{x}, u \rangle + \langle \bar{y}, \bar{v} \rangle - K(\bar{x}, \bar{y})$$

where the last step is a consequence of inequality (1.34). Similarly, $L(\bar{u}, v) \geq \langle x, \bar{u} \rangle + \langle \bar{y}, v \rangle - K(\bar{x}, \bar{y})$ for all $v$. So the last two inequalities imply that for each $u$ and $v$,

$$L_{\bar{x}, \bar{y}}(u, \bar{v}) \leq L_{\bar{x}, \bar{y}}(\bar{u}, v)$$

which is the statement in (d). Finally, (c) and (d) are equivalent due to \qed
1.2.5 Hypo/epi-convergence

Definition 1.2.24 A sequence of closed proper saddle functions \( \{K_n\} \) for \( n \in \mathbb{N} \) is modulated as \( n \to \infty \) if for some \( \rho \geq 0 \) and \( N \) sufficiently large, one has that, for \( n > N \),

\[
\inf_{\|y\| \leq \rho} K_n(x, y) \leq \rho (1 + \|x\|), \quad \text{for all } x, \tag{1.35}
\]

\[
\sup_{\|x\| \leq \rho} K_n(x, y) \geq -\rho (1 + \|y\|), \quad \text{for all } y. \tag{1.36}
\]

A sequence of closed proper saddle functions \( \{K_n\} \) is said to hypo/epi-converge to a closed proper saddle function \( K \) provided

\[
\tilde{K}(x, y) = \sup_{x_n \to x} \inf_{y_n \to y} \limsup_n K_n(x_n, y_n) \leq K(x, y) \tag{1.37}
\]

\[
\bar{K}(x, y) = \inf_{y_n \to y} \sup_{x_n \to x} \liminf_n K_n(x_n, y_n) \geq K(x, y) \tag{1.38}
\]

If the sequence is modulated and hypo/epi-convergent, it is said to hypo/epi-converge in the modulated sense.

This definition of modulated family first appeared in [Rock90], and it was later used in [Goeb05]. A hypo/epi-convergent sequence does not necessarily have a unique limit; in fact, if \( K \) is a limit of such sequence, then every element in the equivalence class \([K]\) is a limit as well. The following result links the concept of hypo/epi-convergence for closed saddle functions with that of epi- and hypo-convergence of their corresponding parents.

Theorem 1.2.25. Let \( \{K_n\}_{n \in \mathbb{N}} \) be a sequence of closed proper saddle functions and let \( f_n \) and \( g_n \) be the convex and concave parents of \( K_n \), respectively.

(i) \( K_n \) is modulated as \( n \to \infty \) if and only if for some \( \rho \geq 0 \) and \( n \) sufficiently large

\[
\min_{\|x\| \leq \rho, \|v\| \leq \rho} f_n(x, v) \leq \rho, \tag{1.39}
\]

and

\[
\max_{\|w\| \leq \rho, \|y\| \leq \rho} g_n(w, y) \geq -\rho. \tag{1.40}
\]

(ii) Let \( K_n, f_n \) and \( g_n \) defined as before and let \( K \) be a closed proper saddle functions with convex and concave parents \( f \) and \( g \) respectively. The following are equivalent

(a) \( K_n \) hypo/epi-converges to \( K \) in the modulated sense as \( n \to \infty \).

(b) \( f_n \) epi-converges to \( f \) as \( n \to \infty \).
\( g_n \) hypo-converges to \( g \) as \( n \to \infty \).

**Proof.** To show (i), notice that property (1.36) is equivalent to
\[
\inf_y \sup_{\|x\| \leq \rho} \{ K_n(x, y) + \rho \|y\| \} \geq -\rho.
\]
Let
\[
K^\rho_n(x, y) = \begin{cases} 
K_n(x, y) + \rho \|y\| & \text{if } \|x\| \leq \rho, \\
-\infty & \text{if } \|x\| > \rho.
\end{cases}
\]
Then
\[
\inf_y \sup_{\|x\| \leq \rho} \{ K_n(x, y) + \rho \|y\| \} = \inf_y \sup_x K^\rho_n(x, y).
\]
For \( \rho \) sufficiently large, \( K^\rho_n \) is a closed proper saddle function with \( \text{dom}_1 K^\rho_n \) a bounded set. According to Corollary 37.3.2 in [Rock70], \( K^\rho_n \) has a saddle value and the order in which \( \inf \) and \( \sup \) are written in the previous equation does not matter, so
\[
\inf_y \sup_x K^\rho_n(x, y) = \sup_x \inf_y \{ K_n(x, y) + \rho \|y\| \} \quad (1.41)
\]
For the inner infimum, the Fenchel duality theorem (Theorem 1.1.21) applies for the convex function \( p(y) = K_n(x, y) \) and the concave function \( q(y) = -\rho \|y\| \), since the constraint qualification on this theorem is satisfied because \( \text{dom} q = \mathbb{R}^m \). Therefore, (1.41) is equivalent to
\[
\inf_y \sup_x K^\rho_n(x, y) = \sup_x \inf_y \{ K_n(x, y) + \rho \|y\| \} = \sup_x \inf_y \left\{ -\left( K_n(x, \cdot) \right)^* (u) + (\rho \|\cdot\|)^* (u) \right\} \quad (1.42)
\]
by the formula of the scalar multiplication conjugate and the partial conjugate formula (1.24). So, if (1.36) holds, then
\[
\sup_x \inf_y \left\{ -f_n(u, v) \right\} \geq -\rho,
\]
thus (1.39) holds. On the other hand, under the assumption that formula (1.39) holds, the reverse formula (1.22) holds since \( K_n \) is closed for every \( n \), therefore (1.36) must hold due to the equivalences previously shown. We conclude that (1.36) and (1.39) are equivalent. The equivalence between (1.35) and (1.40) can be deduced similarly. This completes (i).

For (ii), (b) is equivalent to (c) due to the continuity of the Fenchel conjugate and the relationship between convex and concave parents. Now, assume that (b) holds. Then by Lemma 3.3 in [Atto88], \( \tilde{K}(x, y) \leq K(x, y) \). Since the conjugates \( f_n^* \) also epi-converge to \( f^* \), again Lemma 3.3 in [Atto88] ensures that \( K(x, y) \geq \)
This proves (a). For the reverse implication, the same lemma shows that $f \leq e^{-\liminf_n f_n}$ and $f^* \leq e^{-\liminf_n f_n^*}$. Under the upper modulated hypothesis, this means that $f_n$ epi-converges to $f$.

**Remark 1.2.26** The bounds in (1.39) that characterize the modulated hypothesis are satisfied if and only if the sequence $f_n$ does not escape epigraphically to the horizon, that is, if $f_n$ does not epi-converges to the constant function $+\infty$ (see Exercise 7.5, [Rock98]). Therefore, if a sequence of proper, closed, and convex function epi-converges to a proper function, then the modulated hypothesis is automatically satisfied.

The following result was proved in the extended hypo/epi-convergence setting in [Atto88]. We provide a proof under the modulated sense just studied.

**Theorem 1.2.27.** Let $\{K_n\}_{n \in \mathbb{N}}$ be a sequence of closed proper saddle functions that hypo/epi-converges to $K$. Suppose that there exists a subsequence $\{K_{n_j}\}_{j \in \mathbb{N}}$ such that for each $j$, $(\bar{x}_j, \bar{y}_j)$ is a saddle point of $K_{n_j}$, with

$$\lim_{j \to \infty} \bar{x}_j = \bar{x} \text{ and } \lim_{j \to \infty} \bar{y}_j = \bar{y}.$$

Then $(\bar{x}, \bar{y})$ is a saddle point of $K$ and

$$\lim_{j \to \infty} K_{n_j}(\bar{x}_j, \bar{y}_j) = K(\bar{x}, \bar{y}).$$

**Proof.** For each $j$, $K_{n_j}$, $K_{n_j}^*$ and $\overline{K}_{n_j}$ are equivalent closed saddle functions, so by Corollary 1.2.19 we have that

$$K_{n_j}(\bar{x}_j, \bar{y}_j) = K_{n_j}^*(\bar{x}_j, \bar{y}_j) = \overline{K}_{n_j}(\bar{x}_j, \bar{y}_j) \quad (1.44)$$

By definition of saddle point, for each $j$, $\overline{K}_{n_j}(x, \bar{y}_j) \leq K_{n_j}(\bar{x}_j, \bar{y}_j) \leq \overline{K}_{n_j}(\bar{x}_j, y)$, for every $x$ and $y$.

Let $\{\bar{\xi}_n\}$ be a sequence such that $\bar{\xi}_{n_j} = \bar{x}_j$ and $\lim_{n \to \infty} \bar{\xi}_n = \bar{x}$. Pick $y \in \mathbb{R}^m$ and let $\{y_n\}$ be a sequence that converges to $y$ as $n \to \infty$. Then,

$$\limsup_n \overline{K}_n(\bar{\xi}_n, y_n) \geq \limsup_j \overline{K}_{n_j}(\bar{x}_j, y_j) \geq \limsup_j \overline{K}_{n_j}(\bar{x}_j, \bar{y}_j)$$

The last inequality holds for every sequence $\{y_n\}$ that converges to $y$, so

$$\inf_{y_n \to y} \limsup_n \overline{K}_n(\bar{\xi}_n, y_n) \geq \limsup_j \overline{K}_{n_j}(\bar{x}_j, \bar{y}_j)$$

Since $\bar{\xi}_n$ converges to $\bar{x}$, the following holds

$$\bar{K}(\bar{x}, y) \geq \limsup_j \overline{K}_{n_j}(\bar{x}_j, \bar{y}_j)$$

30
An analogous argument allows to conclude also that

$$K(x, \bar{y}) \leq \lim \inf_j K_{n_j}(\bar{x}_j, \bar{y}_j).$$

Combining the last two inequalities with the two inequalities in the definition of hypo/epi-convergence, we obtain, for every $x$ and $y$,

$$K(x, \bar{y}) \leq \tilde{K}(x, \bar{y}) \leq \lim \inf_j K_{n_j}(\bar{x}_j, \bar{y}_j),$$

and

$$\lim \sup_j K_{n_j}(\bar{x}_j, \bar{y}_j) \leq \widetilde{K}(\bar{x}, \bar{y}) \leq K(\bar{x}, \bar{y}).$$

Since (1.45) holds for all $y$, it is also true that

$$K(x, \bar{y}) = \text{cl} \ K(x, \bar{y}) \leq \lim \inf_j K_{n_j}(\bar{x}_j, \bar{y}_j).$$

Thus combining (1.46) and (1.47),

$$K(x, \bar{y}) \leq \lim \inf_j K_{n_j}(\bar{x}_j, \bar{y}_j) \leq \lim \sup_j K_{n_j}(\bar{x}_j, \bar{y}_j) \leq \tilde{K}(\bar{x}, \bar{y}) \leq K(\bar{x}, \bar{y}).$$

This proves that $(\bar{x}, \bar{y})$ is a saddle point of $K$, thus in turn a saddle point of $K$ due to the equivalence between $K$ and $\overline{K}$. The limit formula follows from the last observation and equation (1.44).

The following result was only known for finite saddle function. Its proof requires a slight modification of the proof for the finite case.

**Theorem 1.2.28.** Let $\{K_n\}_{n \in \mathbb{N}}$ be a sequence of closed proper saddle function that hypo/epi-converges to a closed proper saddle function $K$. Let $(\bar{x}, \bar{y}) \in \text{int dom } K \neq \emptyset$. Then

$$K(\bar{x}, \bar{y}) = \lim_{n \to \infty} K_n(\bar{x}, \bar{y}).$$

**Proof.** The given hypothesis imply, according to Theorem 4.3 in [Rock90], that the sequence $\{\partial K_n\}$ converges graphically to $\partial K$. The assumptions on $\bar{x}$ and $\bar{y}$ assure that both sets $\partial_1 K(\bar{x}, \bar{y})$ and $\partial_2 K(\bar{x}, \bar{y})$ are nonempty, convex, closed, and bounded. Thus, the same can be said about the subgradient $\partial K(\bar{x}, \bar{y})$. Exercise 5.34 in [Rock98] applied to $\partial K = \partial \overline{K} = \partial K$ ensures that for each sequence $\{x_n\}$ converging to $\bar{x}$, there exist $M > 0$ and $N \in \mathbb{N}$ such that

$$\text{for } n > N, \|s_n\| \leq M, \text{ for every } s_n \in \partial_2 \overline{K}_n(x_n, \bar{y}),$$

and a similar statement holds for $\partial_1 \overline{K}$. Let $\{y_n\}$ be a sequence converging to $\bar{y}$. By definition of subgradient, for $n$ sufficiently large,

$$\overline{K}_n(x_n, y_n) \geq \overline{K}_n(x_n, \bar{y}) + \langle y_n - \bar{y}, s_n \rangle, \text{ with } \|s_n\| \leq M.$$
Notice that
\[
\limsup_{n \to \infty} |\langle \bar{y} - y_n, \varsigma_n \rangle| \leq \limsup_{n \to \infty} \| \bar{y} - y_n \| \| \varsigma_n \| \\
\leq M \limsup_{n \to \infty} \| \bar{y} - y_n \| = 0.
\]

Inequality (1.37) in the definition of hypo/epi-convergence, combined with the last observation, implies that
\[
\overline{\mathcal{K}}(\bar{x}, \bar{y}) \geq \sup_{x_n \to x} \inf_{y_n \to y} \limsup_{n} \mathcal{K}_n(x_n, y_n)
\]
\[
= \sup_{x_n \to x} \inf_{y_n \to y} \left\{ \limsup_{n} \mathcal{K}_n(x_n, y_n) + \limsup_{n \to \infty} \langle \bar{y} - y_n, \varsigma_n \rangle \right\}
\]
\[
\geq \sup_{x_n \to x} \inf_{y_n \to y} \left\{ \limsup_{n} \left( \overline{\mathcal{K}}_n(x_n, y_n) + \langle \bar{y} - y_n, \varsigma_n \rangle \right) \right\}
\]
\[
\geq \sup_{x_n \to x} \inf_{y_n \to y} \limsup_{n} \overline{\mathcal{K}}_n(x_n, \bar{y})
\]
\[
\geq \sup_{x_n \to x} \limsup_{n} \overline{\mathcal{K}}_n(x_n, \bar{y})
\]
\[
\geq \limsup_{n} \overline{\mathcal{K}}_n(\bar{x}, \bar{y}),
\]
therefore, \(K(\bar{x}, \bar{y}) \geq \limsup_{n} K_n(\bar{x}, \bar{y})\), because \(K\) and \(\overline{\mathcal{K}}\) agree on int dom \(K\). A similar argument shows that \(K(\bar{x}, \bar{y}) \leq \liminf_{n} K_n(\bar{x}, \bar{y})\), and this proves the result. \(\square\)
Chapter 2
Regular Approximation

In this chapter we introduce a regularization that can be applied to an ample spectrum of functions. Its definition relies on that of the celebrated Moreau envelope. A remarkable property of the regularization is its self-duality with respect to the Fenchel conjugation, when restricted to the class of proper, lsc and convex functions. Furthermore, an extension of this regularization to saddle functions preserves the same feature when applied to the class of proper, closed saddle functions. These interesting facts were first developed in [Goeb08]. Here, we highlight the main features of this smoothing and the Moreau envelope, extend some of the results presented by Goebel, and study its behavior on a broader class of nonconvex functions. Finally, we present two applications to convex optimization problems.

2.1 Moreau Envelope and Prox-bounded Functions

Definition 2.1.1 For a proper, lsc and extended-valued function $f$ and $\lambda > 0$, the Moreau envelope of $f$, denoted by $e_\lambda f$, is defined as

$$e_\lambda f(x) = \inf_u \left\{ f(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\}. \quad (2.1)$$

The proximal mapping $P_\lambda f(x)$ is a set-valued mapping given by

$$P_\lambda f(x) = \arg\min_u \left\{ f(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\} \quad (2.2)$$

Therefore, $e_\lambda f$ is the inf-convolution of $f$ with the function $j_\lambda(x) = \|x\|^2 / (2\lambda)$. In our notation, $e_\lambda f = f \# j_\lambda(x)$.

It is important first to distinguish a class of functions for which the Moreau envelope is not trivial.

Definition 2.1.2 A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is prox-bounded if there exists $\lambda > 0$ such that $e_\lambda f(x) > -\infty$ for some $x \in \mathbb{R}^n$. For a prox-bounded function $f$, the supremum of all such $\lambda$ is the threshold $t_f$ of prox-boundedness for $f$.

For a proper, lsc function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $p > 0$, the following formula will be used

$$\liminf \frac{f(x)}{\|x\|^{-\infty}} = \sup \left\{ \gamma \in \mathbb{R} \mid \exists \beta \in \mathbb{R} \text{ with } f(x) \geq \gamma \|x\|^p + \beta \text{ for all } x \right\}. \quad (2.3)$$

Here is a characterization of prox-bounded functions.
Lemma 2.1.3 (Exercise 1.24, [Rock98]). For a proper and lsc function \( f : \mathbb{R}^n \to \mathbb{R} \), the following are equivalent:

(a) \( f \) is prox-bounded,

(b) \( f \) majorizes a quadratic function (i.e., \( f \geq q \) for a polynomial \( q \) of degree two or less),

(c) for some \( r \in \mathbb{R} \), \( f + \frac{1}{2} r \| \cdot \| \) is bounded from below on \( \mathbb{R}^n \),

(d) \( \lim \inf_{\|x\|\to\infty} \frac{f(x)}{\|x\|} > -\infty \).

Moreover, if \( r_f \) is the infimum of all \( r \) for which (c) holds, the limit in (d) is \( -\frac{1}{2} r_f \) and the proximal threshold for \( f \) is \( t_f = 1/\max \{0, r_f\} \), interpreting \( 1/0 = \infty \).

Proof. ((a) \Rightarrow (b)) Let \( \lambda \) and \( x \) be such that \( e^\lambda f(x) > -\infty \). By definition, \( f(u) + \frac{1}{2} \lambda \| x - u \| \geq \alpha \), (2.4) for every \( u \in \mathbb{R}^n \) and for some real number \( \alpha \). Thus, \( f \) majorizes the quadratic function \( q(u) = -\frac{1}{2} \lambda \| x - u \| + \alpha \).

((a) \Rightarrow (c)) It follows from equation (2.4) after taking \( r = \frac{1}{\lambda} \).

((c) \Rightarrow (a)) There exist real numbers \( r \) and \( \alpha \) such that \( f(u) + \frac{1}{2} r \| u \|^2 \geq \alpha \). Then, for every \( \lambda > 0 \),

\[
f(u) + \frac{1}{2\lambda} \| x - u \|^2 \geq \left( \frac{1}{2} r + \frac{1}{2\lambda} \right) \| u \|^2 + \alpha.
\]

If \( r \geq 0 \), the right-hand side of the inequality is bounded below by \( \alpha \), and so taking the infimum over \( u \), \( e^\lambda f(0) \geq \alpha > -\infty \). The same conclusion can be obtained for \( r < 0 \), for the fixed value \( \lambda = -\frac{1}{r} \). Hence, in both cases, (a) holds.

((c) \Rightarrow (d)) The function \( f \) satisfies \( f(x) \geq -\frac{1}{2} r \| u \|^2 + \alpha \). Then, by equation (2.3), \( \lim \inf_{\|x\|\to\infty} \frac{f(x)}{\|x\|^2} \geq -\frac{1}{2} r > -\infty \).

((d) \Rightarrow (c)) If the inequality is satisfied, then there exist real numbers \( \gamma \) and \( \beta \) such that \( f(x) \geq \gamma \| x \|^2 + \beta \). Thus, the function \( f + \frac{1}{2}(-2\gamma) \| \cdot \|^2 \) is bounded below.

((b) \Rightarrow (d)) By assumption, the function \( f \) majorizes a quadratic function \( q \).

Therefore, \( \lim \inf_{\|x\|\to\infty} \frac{f(x)}{\|x\|^2} \geq \lim \inf_{\|x\|\to\infty} \frac{q(x)}{\|x\|^2} > -\infty \).

This completes the required equivalences. Now, from the proof of ((c) \Rightarrow (d)), \( \lim \inf_{\|x\|\to\infty} \frac{f(x)}{\|x\|^2} \geq -\frac{1}{2} r_f \), after taking infimum over \( r \). Moreover, from the proof of ((d)
⇒ (c), \( r_f \leq -2\gamma \), after taking the infimum over \( r \). Maximizing over all such \( \gamma \),

\[
-\frac{1}{2} r_f \leq \liminf_{\|x\| \to -\infty} \frac{f(x)}{\|x\|^2}.
\]

Hence, \(-\frac{1}{2} r_f = \liminf_{\|x\| \to -\infty} \frac{f(x)}{\|x\|^2} \).

Finally, we verify the formula for \( \lambda_f \). If \( r_f \leq 0 \), then \( f + \frac{1}{2\lambda} \|\cdot\|^2 \) is bounded below for every \( \lambda \), so \( e_\lambda f(0) > -\infty \) and \( \lambda_f = \infty \). On the other hand, if \( r_f > 0 \), then \( f + \frac{1}{2\lambda} \|\cdot\|^2 \) is bounded below for every \( \lambda \leq \frac{1}{r_f} \), and unbounded below for \( \lambda > \frac{1}{r_f} \). Hence, \( \lambda_f = \frac{1}{r_f} \), and the formula for \( \lambda_f \) follows.

\[ \text{Example 2.1.4} \] Proper convex functions and also functions which are bounded from below are examples of prox-bounded functions with infinite threshold.

\[ \text{Proposition 2.1.5} \] Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a proper, lsc, and prox-bounded function with threshold \( \lambda_f \). Then, for \( 0 < \lambda \leq \mu < \lambda_f \),

(a) \( e_\mu f \leq e_\lambda f \leq f \). Thus, as \( \lambda \downarrow 0 \), the envelopes \( e_\lambda f \) form an increasing family of functions bounded above by \( f \), and

\[
\sup_{\lambda > 0} e_\lambda f(x) = f(x) \text{ for all } x,
\]

(b) \( e_\lambda(e_\mu f) = e_{\lambda+\mu} f \).

(c) \( \inf e_\lambda f = \inf f \) and \( \argmin e_\lambda f = \argmin f \).

\[ \text{Proof.} \] If \( \lambda \leq \mu > 0 \), then \( f(u) + \frac{1}{2\lambda} \|x-u\|^2 \geq f(u) + \frac{1}{2\mu} \|x-u\|^2 \), for all \( u \), so taking the infimum over \( u \), \( e_\mu f \leq e_\lambda f \). The other inequality follows after taking \( u = x \) in the definition of the Moreau envelope.

For (b),

\[
e_\lambda(e_\mu f) = e_\mu f \#j_\lambda(x)
= f \#(j_\lambda(x) \#j_\lambda(x))
= f \#(j_{\lambda+\mu}(x))
= e_{\lambda+\mu} f.
\]

To show (c), it is clear from (a) that \( \inf e_\lambda f \leq \inf f \). On the other hand, \( f(u) \leq f(u) + \frac{1}{2\lambda} \|x-u\|^2 \) for each \( x \) and \( u \), thus taking infimum over \( w \) on both sides, \( \inf f \leq e_\lambda f(x) \), thus \( \inf f \leq \inf e_\lambda f \), and this shows the first part of (c). Its second part can be verified similarly.

The following two theorems summarize the basic properties of the Moreau envelope and its convergence behavior as \( \lambda \downarrow 0 \).
Theorem 2.1.6 (Theorem 1.25, [Rock98]). Let \( f : \mathbb{R}^n \to \mathbb{R} \) be proper, lsc, and prox-bounded function with positive threshold \( t_f \). Then for every \( \lambda \in (0, t_f) \) the set \( P_\lambda f(x) \) is nonempty and compact, while the value \( e_\lambda f(x) \) is finite and depends continuously on \((\lambda, x)\), with

\[
\lim_{\lambda \searrow 0} e_\lambda f(x) = \sup_{\lambda > 0} e_\lambda f(x) = f(x) \quad \text{for all } x.
\]

Furthermore, provided that \( x_m \to x \) and \( \lambda_m \searrow 0 \) in \((0, t_f)\) in such a way that the sequence \( \{\|x_m - x\|/\lambda_m\}_{m \in \mathbb{N}} \) is bounded, then \( e_\lambda f \) converges continuously to \( f \), namely,

\[
e_{\lambda_m} f(x_m) \to f(x). \tag{2.5}
\]

Moreover, if \( w_m \in P_{\lambda_m} f(x_m), x_m \to x \) and \( \lambda_m \to \lambda \in (0, t_f) \), then the sequence \( \{w_m\}_{m \in \mathbb{N}} \) is bounded and all its cluster points lie in \( P_\lambda f(x) \).

Corollary 2.1.7. For a proper, lsc, and prox-bounded function \( f \), the map \( \lambda \to e_\lambda f \) is continuous for \( 0 < \lambda < \lambda_f \).

Proof. This is a direct consequence of statement (b) in Proposition 2.1.5 and Theorem 2.1.6.

Proposition 2.1.8. Let \( f : \mathbb{R}^n \to \mathbb{R} \) be proper, lsc, and prox-bounded function. Then \( e_\lambda f \) epi-converges to \( f \) as \( \lambda \searrow 0 \).

Proof. Since \( e_\lambda f \) is a non-decreasing family, then by Proposition 7.4 in [Rock98], \( e_\lambda f \) epi-converges to \( \sup_\lambda (\text{cl } e_\lambda f) \), which equals \( f \) in virtue of Theorem 2.1.6.

In summary, given a proper, lsc and prox-bounded function \( f \) and \( \lambda \in (0, t_f) \), the functions \( e_\lambda f \) form a finite, continuous, infima-preserving, and nondecreasing family that converges to \( f \) pointwise and epigraphically as \( \lambda \searrow 0 \).

Remark 2.1.9 Let \( h \) be an extended-valued proper and prox-bounded function on \( \mathbb{R}^n \). Then both \( h \) and \( \text{cl } h \) have the same Moreau envelope. In fact,

\[
e_\lambda h(x) = \inf_u \left\{ h(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\}
\]

\[
= \inf_u \left\{ \text{cl } \left\{ h(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\} \right\}
\]

\[
= \inf_u \left\{ \text{cl } h(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\}
\]

\[
= e_\lambda \text{cl } h(x),
\]

therefore, in this situation, by Theorem 2.1.6 we conclude

\[
\lim_{\lambda \searrow 0} e_\lambda h(x) = \sup_{\lambda > 0} e_\lambda \text{cl } h(x) = \text{cl } h(x) \quad \text{for all } x.
\]

Moreover, \( e_\lambda h \) epi-converges to \( \text{cl } h \) as \( \lambda \searrow 0 \), by Theorem 2.1.8.
2.2 Regular Approximation

Definition 2.2.1 Let \( f : \mathbb{R}^n \to \mathbb{R} \) be proper, lsc, and prox-bounded with threshold \( \lambda_f \). For each \( \lambda \in (0, \lambda_f) \), the regular approximation \( s_\lambda f \) is defined as

\[
s_\lambda f(x) = (1 - \lambda^2) e_\lambda f(x) + \frac{\lambda}{2} \|x\|^2.
\]

The motivation behind this particular definition will be apparent in Section 2.4. It is clear from its definition that the regularization \( s_\lambda \) inherits many properties from those of \( e_\lambda \).

Theorem 2.2.2. Let \( f : \mathbb{R}^n \to \mathbb{R} \) be proper, lsc, and prox-bounded function with positive threshold \( t_f \). Then for every \( \lambda \in (0, t_f) \) the value \( s_\lambda f(x) \) is finite and depends continuously on \((\lambda, x)\), with

\[
\lim_{\lambda \downarrow 0} s_\lambda f(x) = f(x) \text{ for all } x.
\]

Moreover, \( s_\lambda f \) epi-converges to \( f \) as \( \lambda \searrow 0 \), and converges continuously as well provided the conditions in Theorem 2.1.6 are satisfied.

Proof. This result follows due to the continuity of the function \( j_\lambda \) and the fact that pointwise and epi-convergence are preserved under continuous perturbations.

It is important to point out, however, that \( s_\lambda \) does not preserve other features of the Moreau envelope. As a matter of fact, none of the statements listed in Proposition 2.1.5 holds in general if \( e_\lambda f \) is substituted by \( s_\lambda f \).

Example 2.2.3 For the function \( f : \mathbb{R} \to \mathbb{R}, f(x) = x \), then \( e_\lambda f(x) = x - \frac{1}{2} \lambda \), and \( s_\lambda f(x) = \frac{1}{2} \lambda x^2 + (1 - \lambda^2) x - \frac{1}{2} \lambda (1 - \lambda^2) \). In this case, a direct verification shows that neither (a) nor (b) in Proposition 2.1.5 holds. Also, \( \text{argmin} \ s_\lambda f = \left\{ \frac{\lambda^2 - 1}{\lambda} \right\} \).

In general, a sequence that epi-converges to a proper function that has at least a minimizer satisfies that the property that the infimum values of the sequence converge to the infimum value of the limit function. The next result establishes this result for the regular approximation \( s_\lambda \) with the improvement that no assumptions are necessary on the existence of minimizers of the limit function.

Proposition 2.2.4. Let \( f : \mathbb{R}^n \to \mathbb{R} \) be proper, lsc and prox-bounded with positive threshold \( t_f \). Then, for every sequence \( \lambda_m \searrow 0 \) with \( \lambda_m \in (0, t_f) \),

\[
\inf s_{\lambda_m} f \to \inf f
\]

Proof. By Proposition 2.1.8, \( f_m = s_{\lambda_m} f \) epi-converges to \( f \), so by Proposition (1.1.19),

\[
\lim sup_m (\inf s_{\lambda_m} f) \leq \inf f
\]
If $\inf f = -\infty$, then
\[
\limsup_m (\inf s_{\lambda m} f) \leq \inf f \leq \liminf_m (\inf s_{\lambda m} f),
\] (2.9)

thus the result follows in this case. Suppose now that $\inf f$ is finite. By definition of $s_{\lambda m} f$,
\[
\inf s_{\lambda m} f \geq (1 - \lambda_m^2) \inf e_{\lambda m} f + \frac{\lambda_m^2}{2} \inf \| \cdot \|^2 = (1 - \lambda_m^2) \inf f,
\]
so, after taking $\liminf$ on both sides,
\[
\liminf_m (\inf s_{\lambda m} f) \geq \inf f.
\] (2.10)

The result follows after combining the inequalities (2.10) and (2.8).

**Remark 2.2.5** If $s_\lambda$ is substituted by $e_\lambda$ in the previous statement, the result is straightforward, since in this case, $\inf e_\lambda f$ equals $\inf f$ by statement (c) in Proposition 2.1.5.

Another important property to establish is the coercivity of $s_\lambda f$, because a continuous and coercive function has at least one minimizer, i.e., its argmin is nonempty.

**Definition 2.2.6** A function $f : \mathbb{R}^n \to \mathbb{R}$ is

(a) coercive if it is bounded below on bounded sets and
\[
\liminf_{\| x \| \to \infty} \frac{f(x)}{\| x \|} = \infty;
\] (2.11)

(b) counter-coercive if it is bounded below on bounded sets and
\[
\liminf_{\| x \| \to \infty} \frac{f(x)}{\| x \|} = -\infty;
\] (2.12)

There exists a connection between these growth conditions and the value of the corresponding recession functions.

**Theorem 2.2.7** (Theorem 3.26, [Rock98]). Let $f$ be proper and lsc on $\mathbb{R}^n$. Then

(i) $f$ is coercive if and only if $f^\infty(x) = \infty$ for all $x \neq 0$.

(ii) $f$ is counter-coercive if and only if $f^\infty(x) = -\infty$ for some $x \neq 0$, or equivalently, $f^\infty(0) = -\infty$.

Since $s_\lambda f$ is the sum of two functions, one of them which is always coercive, namely, $\frac{\lambda}{2} \| \cdot \|^2$, it is necessary to figure out how coercivity behaves under addition. The following basic rules address that.
Proposition 2.2.8. Let \( f_1 \) and \( f_2 \) be proper, lsc on \( \mathbb{R}^n \), and suppose that neither is countercoercive. Then
\[
(f_1 + f_2)^\infty \geq f_1^\infty + f_2^\infty
\]
where the inequality becomes an equation when both functions are convex and \( \text{dom } f_1 \cap \text{dom } f_2 \neq \emptyset \).

Proof. We have that
\[
\inf_{\eta \in (0,\delta)} \eta(f_1 + f_2)(\eta^{-1}x) = \inf_{\eta \in (0,\delta)} \left( \eta f_1(\eta^{-1}x) + \eta f_2(\eta^{-1}x) \right)
\geq \inf_{\eta \in (0,\delta)} \eta f_1(\eta^{-1}x) + \inf_{\eta \in (0,\delta)} \eta f_2(\eta^{-1}x).
\]
where all the steps are justified because none of the previous sums is of the form \( \infty - \infty \) since neither \( f_1 \) nor \( f_2 \) is countercoercive. Taking \( \lim \) as \( \delta \searrow 0 \) on both sides of the last expression we obtain the required inequality. Let us show that the equality holds in the convex case. Let \( \bar{x} \in \text{dom } f_1 \cap \text{dom } f_2 \), which is nonempty by hypothesis. Then, for each \( w \in \mathbb{R}^n \),
\[
(f_1 + f_2)^\infty (w) = \lim_{\eta \to \infty} \frac{(f_1 + f_2)(\bar{x} + \eta w) - (f_1 + f_2)(x)}{\eta}
= \lim_{\eta \to \infty} \frac{(f_1(\bar{x} + \eta w) - f_1(x)) + (f_2(\bar{x} + \eta w) - f_2(x))}{\eta}
= f_1^\infty(w) + f_2^\infty(w)
\]
as desired. \( \square \)

Lemma 2.2.9 (Corollary 3.33, [Rock98]). Suppose that \( f = f_1 \# f_2 \) for proper, lsc functions \( f_1 \) and \( f_2 \) on \( \mathbb{R}^n \) such that
\[
f_1^\infty(-w) + f_2^\infty(w) > 0 \quad \text{for all } w \neq 0. \tag{2.13}
\]
Then \( f \) is a proper, lsc function and the infimum in its definition is attained when finite. Moreover, \( f^\infty \geq f_1^\infty \# f_2^\infty \). When \( f_1 \) and \( f_2 \) are both convex, this holds as an equation.

As an application of the two previous results, we have the following

Proposition 2.2.10. Let \( f_1 \) and \( f_2 \) be proper, lsc on \( \mathbb{R}^n \), and assume that \( f_1 \) is not counter-coercive and \( f_2 \) is coercive. Then
\begin{enumerate}[(i)]
\item the function \( f = f_1 + f_2 \) is proper, lsc, and coercive.
\item the function \( g = f_1 \# f_2 \) is proper, lsc, and not counter-coercive.
\end{enumerate}
Proof. Let us begin by showing (i). Since $f_2$ is coercive, then in particular is not counter-coercive, and so by Proposition (2.2.8), $f$ is proper, lsc and satisfies $f^\infty \geq f_1^\infty + f_2^\infty$. By Theorem 2.2.7, this implies that $f^\infty(x) = \infty$ for all $x \neq 0$, and the coercivity of $f$ follows. For (ii), the hypothesis on $f_1$ and $f_2$ imply that (2.13) holds and so invoking Lemma 2.2.9, $g$ is proper, lsc and satisfies $g^\infty \geq f_1^\infty \# f_2^\infty = f_1^\infty \# \delta_0 = f_1^\infty$. Therefore, since $f_1$ is not counter-coercive, the same property holds for $g$. \hfill \Box

**Theorem 2.2.11.** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc, which is not counter-coercive. Then $s_\lambda f$ is coercive.

**Proof.** By definition, $e_\lambda f = f \# j_\lambda$. For the choices $f_1 = f$ and $f_2 = j_\lambda$, the Moreau envelope is the epi-sum of a function which is not counter-coercive and a function which is coercive. Thus, by Proposition (2.2.10), part (ii), $e_\lambda f$ is not counter-coercive. This in turn means that $s_\lambda f$ is the sum of two proper, lsc functions, one of which is coercive, namely, $\lambda \frac{1}{2} \| \cdot \|^2$, and one which is not counter-coercive, namely, $(1 - \lambda^2) e_\lambda f$. Therefore, by Proposition (2.2.10), part (i), we conclude that $s_\lambda f$ is coercive. \hfill \Box

The functions which are not counter-coercive belong to a particular class of prox-bounded functions, as it is shown below.

**Lemma 2.2.12.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be proper, lsc, which is not counter-coercive. Then $f$ is prox-bounded with threshold $t_f = \infty$.

**Proof.** If $f$ is not counter-coercive, then there exist constants $\alpha$ and $\beta$ such that

$$f(x) \geq \alpha \| x \| + \beta,$$

for all $x$. Dividing by $\| x \|^2$ both sides of the previous inequality, we obtain

$$\frac{f(x)}{\| x \|^2} \geq \frac{\alpha}{\| x \|} + \frac{\beta}{\| x \|^2},$$

for all $x \neq 0$. Thus, the following estimate holds

$$r_f = -2 \left( \liminf_{\| x \| \to \infty} \frac{f(x)}{\| x \|^2} \right) \leq 0.$$

Therefore, by Lemma 2.1.3, $f$ is prox-bounded with threshold

$$t_f = 1 / \max \{ 0, r_f \} = \infty.$$

\hfill \Box

The converse of the previous lemma does not hold in general: the functions $g(x) = -\| x \|^p$, with $p \in (1, 2)$ are prox-bounded with threshold $t_g = \infty$ but they are counter-coercive as well.
2.2.1 Prox-regular Functions

Theorem 2.1.6 states that a proper, lsc, and prox-bounded function has a finite and continuous Moreau envelope. However, higher degrees of regularity, for instance smoothness, do not necessarily hold. A subclass of prox-bounded functions, called prox-regular, has been studied in connection with these features.

Definition 2.2.13 An extended-valued function \( f \) is locally lsc at \( \bar{x} \) if \( f \) is lsc relative to the set \( \{ x \mid \| x - \bar{x} \| < \varepsilon, f(x) < \alpha \} \) for some \( \varepsilon > 0 \) and \( \alpha > f(\bar{x}) \).

The next definition involves the set of proximal subgradients of \( f \), \( \partial_P f(x) \). This is a generalization of the concept of convex subgradient introduced in Chapter 1. For more details on this approach, please refer to [Clar98].

Definition 2.2.14 A function \( f : \mathbb{R}^n \to \mathbb{R} \) that is finite at \( \bar{x} \) is prox-regular at \( \bar{x} \) for \( \bar{v} \), where \( \bar{v} \in \partial_P f(\bar{x}) \), if \( f \) is locally lsc at \( \bar{x} \) and there exist \( \varepsilon > 0 \) and \( r > 0 \) such that

\[
\begin{align*}
\n(f(x') - f(x)) + \langle v, x' - x \rangle &< r \| x' - x \|^2 \quad (2.14)
\end{align*}
\]

whenever \( \| x' - \bar{x} \| < \varepsilon \) and \( \| x - \bar{x} \| < \varepsilon \) with \( x' \neq x \) and \( \| f(x') - f(\bar{x}) \| < \varepsilon \), while \( \| v - \bar{v} \| < \varepsilon \) with \( v \in \partial_P f(x) \).

The following two results provide a connection between prox-regular functions and the self-dual smoothing \( s_\lambda \).

Theorem 2.2.15 (Theorem 5.2, [Rock96]). Suppose that \( f \) is prox-regular at \( \bar{x} = 0 \) for \( \bar{v} = 0 \) with respect to \( \varepsilon \) and \( r \), and let \( \mu \in (0, 1/r) \). Then on some neighborhood of \( 0 \) the function

\[
e_{\mu}f + \frac{r}{2(1 - \mu r)} \| \cdot \|^2
\]

is convex.

Theorem 2.2.16. Suppose that \( f \) is prox-regular at \( \bar{x} = 0 \) for \( \bar{v} = 0 \) with respect to \( \varepsilon \) and \( r \), for \( r < 1 \). Then, on some neighborhood of \( 0 \), \( s_\lambda f \) is

\[
\begin{align*}
\text{strongly convex with modulus } \frac{\lambda - r}{2(1 - \lambda r)} & \quad \text{for } r < \lambda < 1 \\
\text{convex} & \quad \text{for } r = \lambda \\
\text{paraconvex with modulus } \frac{\lambda - r}{2(1 - \lambda r)} & \quad \text{for } 0 < \lambda < r
\end{align*}
\]

Proof. We have that

\[
\begin{align*}
\frac{s_\lambda f}{1 - \lambda^2} &= e_{\lambda}f + \frac{\lambda}{2(1 - \lambda^2)} \| \cdot \|^2 \\
&= e_{\lambda}f + \frac{r}{2(1 - \lambda r)} \| \cdot \|^2 + \frac{\lambda - r}{2(1 - \lambda^2)(1 - \lambda r)} \| \cdot \|^2.
\end{align*}
\]
We can take $\lambda = \mu$ in Theorem 2.2.15 (since $\lambda < 1 < 1/r$) to conclude that

$e_\lambda f + \frac{r}{2(1 - \lambda r)} \| \cdot \|^2$

is a convex function. The result follows after considering the sign of the term $\lambda - r$.

\[ \square \]

### 2.3 Moreau Envelope Under Convexity Assumptions

In this section we present properties of the Moreau envelope and proximal mapping of a proper, lsc, and convex function. For this type of function, the Moreau envelope has a stronger regularizing effect, as can be seen in the next theorem.

**Theorem 2.3.1** (Theorem 2.26 and Example 11.26, [Rock98]). Let $f : \mathbb{R}^n \to \mathbb{R}$ be lsc, proper, and convex. Then the following properties hold for every $\lambda > 0$.

(i) The proximal mapping $P_\lambda f$ is single-valued and continuous. In fact, one has that $P_\lambda f(x) \to P_\lambda f(x)$ whenever $(\lambda, x) \to (\bar{\lambda}, \bar{x})$.

(ii) The envelope function $e_\lambda f$ is convex and continuously differentiable, the gradient being

$\nabla e_\lambda f = \frac{1}{\lambda} [x - P_\lambda f(x)]$.

(iii) The functions $e_\lambda f$ and $f^* + \frac{1}{2} \lambda \| \cdot \|^2$ are conjugate to each other.

**Remark 2.3.2** The differentiability of $e_\lambda f$ in the previous statement is not carried over its conjugate, since $f^*$ might not be smooth. Therefore, it is important to find a regularization for which regularity is preserved through conjugation. As will be seen later in this Chapter, $s_\lambda$ satisfies this requirement.

The following is a decomposition theorem which will be useful in Chapter 3.

**Theorem 2.3.3.** Let $f$ be a proper, lsc, and convex function on $\mathbb{R}^n$ and let $\lambda > 0$. Then every $z \in \mathbb{R}^n$ can be decomposed uniquely as

$z = x + y$

with $x = P_\lambda f(z)$ and $(x, \frac{1}{\lambda} y) \in \text{gph} \, \partial f(x)$

**Proof.** For each $\lambda > 0$, the functions $e_\lambda f$ and $f^* + \lambda j$ are conjugate to each other, with $j(x) = \frac{1}{2} \| x \|^2$. Explicitly, this means that, for every $z \in \mathbb{R}^n$,

$$\inf_x \left\{ f(x) + \frac{1}{2\lambda} \| z - x \|^2 \right\} = \sup_{x^*} \left\{ \langle x^*, z \rangle - \frac{\lambda}{2} \| x^* \|^2 - f^*(x^*) \right\}. \quad (2.15)$$

42
For \( g(x) = -\frac{1}{\lambda}j(z - x) \) we have that
\[
g^*(v) = -\left(\lambda^{-1}j(z - x)\right)^* = -\left(\lambda^* j(z - x)\right)^* = -\lambda j(x) + \langle z, x \rangle.
\]

Then equation (2.15) is of the form (1.9) with the choices of \( f \) and \( g \) just mentioned. The convexity of \( f \) and the strict convexity of \(-g\) ensures that both infima are finite and attained at unique points \( x \) and \( x^* \), which by Theorem 1.1.23 satisfy
\[
x^* \in \partial f(x), \quad x = \nabla g(x^*) = \frac{1}{\lambda}(x^* - z).
\]

Notice that by definition, \( x = \arg\min \left\{ f(x) + \frac{1}{2\lambda} \|z - x\|^2 \right\} = P_\lambda f(z) \). Moreover, letting \( y = \lambda x^* \), we see that \( z = x + y \) and \( \frac{1}{\lambda}y \in \partial f(x) \). This completes the proof.

**Corollary 2.3.4.** Let \( I \) be the identity map in \( \mathbb{R}^n \) and define \( Q_\lambda f = I - P_\lambda f \). Then, for every \( z \in \mathbb{R}^n \),
\[
(a) \quad (P_\lambda f(z), \frac{1}{\lambda}Q_\lambda f(z)) \in \text{gph} \partial f(P_\lambda f(z)),
(b) \quad \frac{1}{\lambda}y \in \partial f(x) \text{ if and only if } (x, y) = (P_\lambda f(z), Q_\lambda f(z)) \text{ for some } z \text{ (in fact, } z = x + y),
(c) \quad P_\lambda f = (I + \lambda \partial f)^{-1},
(d) \quad \nabla e_\lambda f = \frac{1}{\lambda}Q_\lambda f.
\]

**Proof.** Statements (a) and (b) follow directly from the previous theorem. The equation in (c) is merely a restatement of (b). Equation in (d) holds since
\[
\nabla e_\lambda f(z) = \frac{1}{\lambda}(z - P_\lambda f(z)) = \frac{1}{\lambda}Q_\lambda f(z),
\]
for all \( z \).

**Remark 2.3.5** The last result is a generalization of Moreau’s Theorem, which specializes to the case \( \lambda = 1 \). In this case,
\[
\bullet \quad Q_1 f = P_1 f^*
\bullet \quad \text{For } f = \delta_C, \text{ for a closed convex set } C, \text{ then } P_1 f(z) \text{ is the projection of } z \text{ on } C, \text{ which justifies the name of proximal mapping for } P.
\bullet \quad \text{For } f = \delta_L, \text{ for } L \text{ a subspace of } \mathbb{R}^n, \text{ then the corollary reduces to the well-known decomposition of a vector with respect to the subspaces } L \text{ and its orthogonal complement } L^\perp.
\]

43
Lemma 2.3.6. Let \( f \) be a proper, lsc, and convex function on \( \mathbb{R}^n \). Then for each \( \lambda > 0 \), the proximal mapping \( P_{\lambda}f \) is nonexpansive, that is,

\[
\| P_{\lambda}f(z_2) - P_{\lambda}f(z_1) \| \leq \| z_2 - z_1 \|
\]

for every \( z_1, z_2 \) in \( \mathbb{R}^n \).

Proof. Let \( x_i = P_{\lambda}f(z) \) and \( y_i = Q_{\lambda}f(z_i) \), for \( i = 1, 2 \). Then \( z_i = x_i + y_i, i = 1, 2 \) and so,

\[
\| z_2 - z_1 \|^2 = \| x_2 - x_1 \|^2 + 2 \langle x_2 - x_1, y_2 - y_1 \rangle + \| y_2 - y_1 \|^2 \quad (2.16)
\]

By Corollary 2.3.4, \( \frac{y_i}{\lambda} \in \partial f(x_i) \). The convexity of \( f \) implies that \( \partial f \) is a monotone mapping, and this means that

\[
\langle x_2 - x_1, \frac{y_2}{\lambda} - \frac{y_1}{\lambda} \rangle \geq 0
\]

which is of course equivalent to

\[
\langle x_2 - x_1, y_2 - y_1 \rangle \geq 0.
\]

Substituting the last expression into (2.16), we obtain

\[
\| z_2 - z_1 \|^2 \geq \| x_2 - x_1 \|^2,
\]

and the result follows after extracting the square root on both sides. \( \square \)

2.4 Self-dual Regular Approximation

We include the main properties of \( s_{\lambda}f \) presented in [Goeb08] in the next two theorems. The same reference contains the proof of these results.

Theorem 2.4.1. For any convex, lsc, and proper \( f : \mathbb{R}^n \to \mathbb{R} \) and any \( \lambda \in (0, 1) \),

(a) \( s_{\lambda}f \) is strongly convex with constant \( \lambda \);

(b) \( s_{\lambda}f \) is differentiable, with gradient Lipschitz continuous with constant \( 1/\lambda \) and given by

\[
\nabla s_{\lambda}f = \frac{1 - \lambda^2}{\lambda} (x - P_{\lambda}f(x)) + \lambda x; \quad (2.17)
\]

(c) \( \text{argmin } s_{\lambda}f \) is a singleton and equals \( x_{\lambda} \) if and only if \( x_{\lambda} = (1 - \lambda^2) P_{\lambda}f(x_{\lambda}) \).

Furthermore, as \( \lambda \downarrow 0, s_{\lambda}f \) converges to \( f \) pointwise and epigraphically, and if \( \text{argmin } f \neq \emptyset \), then \( \lim x_{\lambda} = x_0 \), where \( x_0 \) is the unique element of \( \text{argmin } f \) of minimal norm.
Recall that for a function $g$, we denote by $g^*$ its convex conjugate, i.e.,

$$g^*(y) = \sup_x \{ y \cdot x - g(x) \}.$$ 

The most striking feature of this smoothing is its self-duality with respect to the convex conjugate operation, and indeed this was largely the motivation behind its definition.

**Theorem 2.4.2** (Theorem 2.2, [Goeb08]). For any convex, lsc, and proper $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and any $\lambda \in (0, 1)$,

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

(2.18)

Another remarkable feature of the regularization $s_\lambda f$ is that it satisfies the cancellation rule presented below.

**Proposition 2.4.3.** For each $\lambda \in (0, 1)$, the map $s_\lambda (\cdot)$ is a one-to-one map from the space of proper, lsc, and convex functions into the space of differentiable functions with Lipschitzian gradient with Lipschitz constant $1/\lambda$.

**Proof.** The fact that $s_\lambda (\cdot)$ maps a proper, lsc, and convex function into the described space is a consequence of Theorem 2.4.1. Now, let $f$ and $g$ be proper, lsc and convex functions such that $s_\lambda f = s_\lambda g$. Then, by definition, for each $x \in \mathbb{R}^n$,

$$(1 - \lambda^2) e_\lambda f(x) + \frac{\lambda}{2} \|x\|^2 = (1 - \lambda^2) e_\lambda g(x) + \frac{\lambda}{2} \|x\|^2,$$

which amounts to

$$e_\lambda f(x) = e_\lambda g(x), \text{ for every } x.$$ 

Taking convex conjugate on both sides and simplifying

$$f^*(x) + \frac{\lambda}{2} \|x\|^2 = g^*(x) + \frac{\lambda}{2} \|x\|^2 \text{ for every } x,$$

thus $f^* = g^*$ and this in turn implies that $f = g$. \qed

**Definition 2.4.4** A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be piecewise linear-quadratic if $\text{dom } f$ can be represented as the union of finitely many polyhedral sets, relative to each of which $f(x)$ is given by an expression of the form $\frac{1}{2} \langle x, Ax \rangle + \langle a, x \rangle + \alpha$ for some scalar $\alpha \in \mathbb{R}$, vector $a \in \mathbb{R}^n$, and a symmetric matrix $A \in \mathbb{R}^{n \times n}$.

A feature which makes this class of non-differentiable functions popular in the literature is its nice behavior under the convex conjugate operation. In fact, according to Theorem 11.14 in [Rock98], a proper, lsc, and convex function $f$ is piecewise linear-quadratic if and only if its conjugate $f^*$ is piecewise linear-quadratic. Using this result, we prove the following

**Proposition 2.4.5.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a proper, convex, and piecewise linear-quadratic function. Then, for each $\lambda \in (0, 1)$, both $s_\lambda f$ and $s_\lambda f^*$ are piecewise linear-quadratic functions.
Proof. Let us show first that, under the given hypothesis, the Moreau envelope $e_\lambda f$ is piecewise linear-quadratic. This holds provided

$$(e_\lambda f)^* = f^* + \frac{\lambda}{2} \| \cdot \|^2$$

is piecewise linear-quadratic, as justified previously. Since $f$ itself is piecewise linear-quadratic, we have that $f^*$ is also piecewise quadratic and therefore the right-hand side of (2.19) is piecewise linear-quadratic. This shows our initial claim. Finally, from the equation (2.6) defining $s_\lambda f$, we conclude that $s_\lambda f$ is piecewise linear-quadratic. The last statement follows applying the conjugate operation and the self-duality of the smoothing.

\[\square\]

2.5 Self-dual Regular Approximation For Saddle Functions

In this section an extension of the self-dual smoothing for saddle functions is presented. This extension was introduced in [Goeb08] and it is based on the concept of mixed Moreau envelope for saddle functions.

Definition 2.5.1 For a proper and closed saddle function $K$, the one-parameter mixed Moreau envelope $E_\lambda K$ for $\lambda > 0$ is given by

$$E_\lambda K(x, y) = \inf_{q \in \mathbb{R}^n} \sup_{p \in \mathbb{R}^m} \left\{ K(p, q) - \frac{1}{2\lambda} \| x - p \|^2 + \frac{1}{2\lambda} \| y - q \|^2 \right\}.$$  \hspace{1cm} (2.20)

and the smoothing $S_\lambda K$ for $0 < \lambda < 1$ is given by

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

Theorem 2.5.2 (Theorem 5.1, [Atto86]). For a proper and closed saddle function $K$,

(a) The order in which $\inf$ and $\sup$ appear in 2.20 is irrelevant, hence,

$$E_\lambda K(x, y) = \sup_{p \in \mathbb{R}^m} \inf_{q \in \mathbb{R}^n} \left\{ K(p, q) - \frac{1}{2\lambda} \| x - p \|^2 + \frac{1}{2\lambda} \| y - q \|^2 \right\}.$$  \hspace{1cm} (2.21)

(b) $K$ and $E_\lambda K$ have the same saddle points and values.

The self-dual smoothing for saddle functions satisfies analogous properties to the ones included in Theorem 2.4.1 for the convex case.

Theorem 2.5.3. For any proper and closed saddle function $K : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ and any $\lambda \in (0, 1)$,

(a) $S_\lambda K$ is strongly concave, strongly convex with constant $\lambda$,
(b) $S_\lambda K$ is differentiable, with gradient Lipschitz continuous with constant $1/\lambda$ and given by

$$\nabla S_\lambda K = \left( \frac{1 - \lambda^2}{\lambda} (x - \hat{x}) - \lambda x, \frac{1 - \lambda^2}{\lambda} (y - \hat{y}) + \lambda y \right),$$  \hspace{1cm} (2.21)

where $(\hat{x}, \hat{y})$ is the (unique) saddle point of the function $K'(p, q) = K(p, q) - \frac{1}{2\lambda} \|x - p\|^2 + \frac{1}{2\lambda} \|y - q\|^2$,

(c) $S_\lambda K$ has a unique saddle point $(x_\lambda, y_\lambda)$ with

$$(x_\lambda, y_\lambda) = \left( (1 - \lambda^2) \hat{x}, (1 - \lambda^2) \hat{y} \right).$$

Furthermore, as $\lambda \searrow 0$, $S_\lambda K$ hypo/epi-converges to $K$ (in the modulated sense), and if $K$ has a saddle point, then $\lim (x_\lambda, y_\lambda) = (x_0, y_0)$, where $(x_0, y_0)$ is the (unique) saddle point of $K$ of minimal norm.

Proof. Statements (a), (b), and the first part of (c) are straightforward adaptations of the corresponding statements for the fully convex case in Theorem 2.4.1. Let $f$ be the convex parent of $K$. Then, the convex parent of $S_\lambda K$ is $s_\lambda f$ by Theorem 3.2 in [Goeb08]. As $\lambda \searrow 0$, $s_\lambda f$ epi-converges to $f$ by Theorem 2.4.1. Therefore, the result follows by the equivalence shown in Theorem 1.2.25 and the remark after it. Suppose that $K$ has a saddle point. The set of all saddle points of $K$ is a convex set in $\mathbb{R}^m \times \mathbb{R}^n$, thus it has a unique element $(x_0, y_0)$ of minimal norm. Let $(x_\lambda, y_\lambda)$ be the saddle point of $S_\lambda K$. Then, by definition of saddle point,

$$S_\lambda K(x_\lambda, y_\lambda) \leq S_\lambda K(x_0, y_\lambda) \leq S_\lambda K(x_\lambda, y_0),$$

so

$$(1 - \lambda^2) E_\lambda K(x_\lambda, y_\lambda) + \frac{\lambda}{2} (\|y_\lambda\|^2 - \|x_\lambda\|^2) \leq (1 - \lambda^2) E_\lambda K(x_\lambda, y_0) + \frac{\lambda}{2} (\|y_0\|^2 - \|x_\lambda\|^2).$$ \hspace{1cm} (2.22)

By Theorem 2.5.2, the point $(x_0, y_0)$ is also a saddle point of $E_\lambda K$. Consequently,

$$E_\lambda K(x_\lambda, y_0) \leq E_\lambda K(x_0, y_0) \leq E_\lambda K(x_0, y_\lambda),$$

thus

$$(1 - \lambda^2) E_\lambda K(x_\lambda, y_0) + \frac{\lambda}{2} (\|y_\lambda\|^2 - \|x_\lambda\|^2) \leq (1 - \lambda^2) E_\lambda K(x_\lambda, y_\lambda) + \frac{\lambda}{2} (\|y_\lambda\|^2 - \|x_\lambda\|^2),$$ \hspace{1cm} (2.23)

Combining inequalities (2.22) and (2.23) and simplifying, we obtain

$$\|y_\lambda\|^2 - \|x_\lambda\|^2 \leq \|y_0\|^2 - \|x_\lambda\|^2,$$

or

$$\|y_\lambda\|^2 + \|x_\lambda\|^2 \leq \|x_0\|^2 + \|y_0\|^2.$$
Thus, every accumulation point of \((x_\lambda, y_\lambda)\) has norm bounded above by the norm of \((x_0, y_0)\). Moreover, since \(S_\lambda K\) hypo/epi-converges to \(K\) as \(\lambda \searrow 0\), Theorem 1.2.27 ensures that the accumulation points of \((x_\lambda, y_\lambda)\) are saddle points of \(K\). The minimality of \((x_0, y_0)\) means that every such accumulation point has norm bounded below by the norm of \((x_0, y_0)\). The uniqueness of the latter point proves the result.

2.6 Applications to Optimization

In this section we provide an overview of the duality theory for convex programs and apply the regular approximation to these programs under two different approaches.

2.6.1 Basic Duality Theory For Convex Programs

A (primal) convex program is an optimization problem of the form

\[
(P) \quad \begin{cases} 
\inf h_0(x) \\
\text{subject to } h_i(x) \leq 0, \text{ for } i = 1, 2, \ldots, m \\
x \in \mathbb{R}^n 
\end{cases}
\]

where the \(h_i : \mathbb{R}^n \to \mathbb{R}\) are proper, lsc, and convex functions for \(i = 0, 1, 2, \ldots, m\). Notice that this definition includes the case where the values of the state \(x\) are constrained to some closed convex set \(C \subset \mathbb{R}^m\) by adding the constraint \(h(x) = \delta_C(x) \leq 0\). This setting also handles situations where affine equality constraints are present, in fact, a constraint of the type \(l(x) = 0\) where \(l\) is affine can be incorporated by means of a double inequality constraint \(l(x) \leq 0\) and \(-l(x) \leq 0\).

The problem \((P)\) is feasible if there exists a point \(x_0 \in \text{dom } h_0\) which satisfies the constraints. In this case, the point \(x_0\) is called a feasible solution.

Let us develop the duality theory in this setting. The idea is to introduce associated perturbed problems by adding parameters convexly. One of the most common settings is to consider the family of convex problems

\[
V(u) = \begin{cases} 
\inf h_0(x) \\
\text{subject to } h_i(x) \leq u_i, \text{ for } i = 1, 2, \ldots, m \\
x \in \mathbb{R}^n, u \in \mathbb{R}^m 
\end{cases}
\]

If we let

\[
f(u, x) = \begin{cases} 
h_0(x) \quad \text{if } h_i(x) \leq u_i, \text{ for } i = 1, 2, \ldots, m \\
\infty \quad \text{otherwise} 
\end{cases} = h_0(x) + \sum_{i=1}^{m} \delta_{(-\infty, 0]}(h_i(x) - u_i),
\]

48
then
\[ V(u) = \inf_x f(u, x) \] (2.24)

and \((P)\) is recovered by means of \(V(0) = \inf_x \phi(x)\), where \(\phi(x) = f(0, x)\). The function \(V : \mathbb{R}^m \to \mathbb{R}\) is called the value function associated to \((P)\), \(V(0)\) is the optimal value for problem \((P)\) and \(f\) is called either a representation [Rock74] or dualizing parametrization [Rock98] for \(\phi\).

The convexity assumptions imposed on the functions \(h_i\) make \(f\) not just convex in \(x\) for each \(u\), but also convex jointly in these variables. Consequently, \(V\) is convex as well. Moreover, since each \(h_i\) is proper and lsc, then the same can be said about \(f\), thus making it a closed function.

In order to develop the duality theory for this optimization problem, the customary path followed in convex optimization (see for instance [Rock74], [Boyd04], [Borw06]) is by means of the so-called Lagrangian function \(L\) defined as
\[
L(y, x) = h_0(x) + \sum_{i=1}^m y_i h_i(x) - \sum_{i=1}^m \delta_{(-\infty, 0)}(y_i) \tag{2.25}
\]
\[= \inf \{ f(u, x) + \langle u, y \rangle \mid u \in \mathbb{R}^m \} \tag{2.26}
\]

with the convention that \(\infty - \infty = \infty\). Notice that \(L\) is concave in \(y\) regardless of the features of the functions \(h_i\). Moreover,
\[
\phi(x) = \sup_y L(y, x) \\
= \sup_{y \geq 0} L(y, x) \\
= \begin{cases} h_0(x) & \text{if } x \text{ is feasible} \\ +\infty & \text{otherwise} \end{cases}
\]

where \(y \geq 0\) means that \(y_i \geq 0\), for each \(i\). Thus, the primal problem \((P)\) can be recovered from \(L\) as a \(\inf\)-\(\sup\) problem
\[
V(0) = \inf_x \sup_y L(y, x)
\]

The dual problem arises from the question on whether the minimization and maximization in the previous formula can be reordered.

**Definition 2.6.1** The dual problem \((D)\) for the primal problem \((P)\) is
\[
\sup_y \gamma(y)
\]
with
\[
\gamma(y) = \inf_x L(y, x)
\]
The concavity of $L(y, \cdot)$ entails that of $\gamma$, and so the dual problem is always a concave program even in the absence of convexity assumptions on ($P$).

Equation (2.26) allows to link the optimization problem with the saddle function theory developed in Chapter 1. In fact, according to formula (2.26), the Lagrangian $L$ is the partial concave conjugate of the jointly concave function $-f$. Therefore, invoking Theorem 1.2.11, $L$ is a closed saddle function and the upper closure of the equivalence class generated by $-f$. In other words, $L$ is a closed saddle function which has convex and concave parents are given by $(u, x) \rightarrow f^*(-u, x)$ and $(u, x) \rightarrow -f(u, x)$ respectively. In this case, $f$ satisfies

$$f(u, x) = \sup_y \{L(y, x) - \langle y, u \rangle\} \quad \text{(2.27)}$$

Under the previous assumptions, the objective function $\gamma$ in the dual problem equals satisfies

$$\gamma(y) = -f^*(-y, 0).$$

Motivated by the previous expression for the dual objective function, a natural candidate for a representation for $\gamma$ is

$$g(w, y) = \inf_x \{L(y, x) - \langle w, x \rangle\} = -f^*(-y, w),$$

and this makes possible to define the dual value function

$$\tilde{V}(w) = \sup_y g(w, y).$$

Furthermore, the primal value function $V$ is related to the objective function $\gamma$ of the dual problem as follows

$$\gamma(y) = -f^*(-y, 0)$$

$$= \inf_u \inf_x \{f(u, x) + \langle u, y \rangle\}$$

$$= \inf_u \inf_x \{f(u, x) + \langle u, y \rangle\}$$

$$= \inf_u \{V(u) + \langle u, y \rangle\}$$

$$= (-V)^*(y),$$

consequently

$$\gamma^* = ((-V)^*)^* = -V^{**} = -\text{cl} V.$$

Thus,

$$\tilde{V}(0) = \sup_y \gamma(y) = -\inf_y \{\langle 0, y \rangle - \gamma(y)\}$$

$$= -\gamma^*(0)$$

$$= \text{cl} V(0),$$

50
while on the other hand, by definition,

\[ V(0) = \inf_x \phi(x). \]

Therefore, the dual optimal value \( \tilde{V}(0) \) is a lower bound for the primal optimal value \( V(0) \) and they match provided \( V(0) \) is lsc at \( y = 0 \). The difference \( V(0) - \tilde{V}(0) \) (assuming \( \infty - \infty = \infty \)) is called the duality gap. Thus, the lower semicontinuity of the value function at 0 means that there is zero duality gap. In general, convex programs have a positive duality gap, but there is a significant number for which the duality gap is zero, including linear programs which are feasible. A central part of the research in duality theory for convex optimization is to find constraint qualifications (CQ) that ensure that \( V \) is lsc at 0. In Section 2.6.3 we introduce a well-known CQ of this type.

The convexity of \( V \) implies that \( V \) is continuous with bounded subdifferential on \( \text{int dom } V \). Consequently, in the case that \( 0 \in \text{int dom } V \), then \( V \) is not just lsc at 0, but continuous there, and the duality gap vanishes.

The following theorem summarizes our previous discussions.

**Theorem 2.6.2** (Theorem 11.39, [Rock98]). Let \( f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R} \) be proper, lsc, and convex and consider the primal problem

\[
\inf_x \phi(x), \text{ where } \phi(x) = f(0, x),
\]

along with the dual problem

\[
\sup_y \gamma(y), \text{ where } \gamma(y) = -f^*(-y, 0).
\]

The function \( \phi \) is convex and lsc, while \( \gamma \) is concave and usc. Define the primal and dual value functions

\[
V(u) = \inf_x f(u, x), \quad \tilde{V}(w) = \sup_y (-f^*(-y, w)); \quad (2.28)
\]

the former being a convex function, while the latter is concave. Then

(a) \( \sup_y \gamma(y) \leq \inf_x \phi(x) \) and \( \inf_x \phi(x) < \infty \) if and only if \( 0 \in \text{dom } V \), whereas \( \sup_y \gamma(y) > -\infty \) if and only if \( 0 \in \text{dom } \tilde{V} \). Moreover,

\[
\sup_y \gamma(y) = \inf_x \phi(x) \text{ if either } 0 \in \text{int dom } V \text{ or } 0 \in \text{int dom } \tilde{V}. \quad (2.29)
\]

(b) The set \( \text{argmax}_y \gamma(y) \) is nonempty and bounded if and only if \( 0 \in \text{int dom } V \) and the value \( V(0) = \inf_x \phi(x) \) is finite, in which case \( \text{argmax}_y \gamma(y) = \partial V(0) \).

(c) The set \( \text{argmin}_x \phi(x) \) is nonempty and bounded if and only if \( 0 \in \text{int dom } \tilde{V} \) and the value \( \tilde{V}(0) = \sup_y \gamma(y) \) is finite, in which case \( \text{argmin}_x \phi(x) = \partial \tilde{V}(0) \).
Optimal solutions are characterized jointly through primal and dual forms of Fermat’s rule:

\[-y, 0) \in \partial f(0, \bar{x}) \iff (0, \bar{x}) \in \partial f^*(-y, 0) \iff \begin{cases} \bar{x} \in \text{argmin}_x \phi(x) \\ \bar{y} \in \text{argmin}_y \gamma(y) \\ \inf_x \phi(x) = \sup_y \gamma(y) \end{cases}\]

The following is a useful result on the epi-convergence of value functions.

**Proposition 2.6.3** (Exercise 7.57, [Rock98]). Consider proper and lsc functions \(f, f_n : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}\) and let

\[V(u) = \inf_x f(x, u), \quad V_n(u) = \inf_x f_n(x, u)\]

If \(f_n\) epi-converges to \(f\) and at least one of the following conditions holds

- the functions \(f_n\) are (jointly) convex,
- the functions \(f_n\) are positively homogeneous,
- the sequence \(\{f_n\}\) is nondecreasing,

then \(V_n\) epi-converges to \(V\)

### 2.6.2 First Regularization Scheme

This setting was first studied in [Goeb08]. For the Lagrangian \(L\) in problem \((P)\), let \(L_\lambda = S_\lambda L\), and consider the regularized problem \((P_\lambda)\) generated by \(L_\lambda\). According to (2.27), the dualizing parametrization \(f_\lambda\) is given by

\[f_\lambda(u, x) = \sup_y \{S_\lambda L(y, x) - \langle y, u \rangle\} = s_\lambda f(u, x)\]

by Theorem 3.2, [Goeb08]. Therefore, the primal problem \((P_\lambda)\) is given by

\[(P_\lambda) \begin{cases} \inf_x \phi_\lambda(x) = s_\lambda f(0, x) = s_\lambda \phi(x) \\ x \in \mathbb{R}^n \end{cases}\]

with dual problem \((D_\lambda)\) given by

\[(D_\lambda) \begin{cases} \sup_x \gamma_\lambda(y) = -(s_\lambda f^*(-y, 0) = -s_\lambda f^*(-y, 0) = -s_\lambda(-\gamma(y)) \\ x \in \mathbb{R}^n \end{cases}\]

Now we show that \(f_\lambda(u, x)\) can be written as the self-dual smoothing of a function in terms of quadratic penalties of the constraints functions in the problem \((P)\). In fact, let \(h(u, x, \lambda) = \frac{\lambda}{2} (\|u\|^2 + \|x\|^2)\). Then, for fixed \(u = (u_1, \ldots, u_m)\),
\[ f_\lambda(u, x) - h(u, x, \lambda) = (1 - \lambda^2) e_\lambda f(u, x) \]

\[
= (1 - \lambda^2) \inf_{\alpha, \beta} \left\{ h_0(\alpha) + \sum_{i=1}^{m} \delta_{[h_i(\alpha), +\infty)}(\beta_i) + \frac{1}{2\lambda} \|x - \alpha\|^2 + \frac{1}{2\lambda} \|u - \beta\|^2 \right\}
\]

\[
= (1 - \lambda^2) \inf_{\alpha} \left\{ h_0(\alpha) + \frac{1}{2\lambda} \|x - \alpha\|^2 + \inf_{\beta} \left\{ \sum_{i=1}^{m} \delta_{[h_i(\alpha), +\infty)}(\beta_i) + \frac{1}{2\lambda} \|u - \beta\|^2 \right\} \right\}
\]

\[
= (1 - \lambda^2) \inf_{\alpha} \left\{ h_0(\alpha) + \frac{1}{2\lambda} \|x - \alpha\|^2 + \inf_{\beta} \left\{ \frac{1}{2\lambda} \|u - \beta\|^2 \text{ if } h_i(\alpha) \leq \beta_i \right\} \right\}
\]

Let us compute the internal infimum. For a given \( \alpha \), if \( h_i(\alpha) \leq u_i \), then we can take \( \beta_i = u_i \) to make the ith summand of the norm equal to zero. On the other hand, \( h_i(\alpha) > u_i \), then the best possible selection is \( \beta_i = h_i(\alpha) \), and this makes the ith summand equal to \( \frac{1}{2\lambda}(u_i - h_i(\alpha))^2 \). Thus,

\[
f_\lambda(u, x) - h(u, x, \lambda)
\]

\[
= (1 - \lambda^2) \inf_{\alpha} \left\{ h_0(\alpha) + \sum_{i=1}^{m} \begin{cases} 0 & \text{if } h_i(\alpha) \leq u_i \\ \frac{1}{2\lambda}(u - h_i(\alpha))^2 & \text{if } h_i(\alpha) > u_i \end{cases} \right\} + \frac{1}{2\lambda} \|x - \alpha\|^2
\]

\[
= (1 - \lambda^2) e_\lambda (h_0(x) + p_\lambda(u, x)),
\]

where \( p_\lambda(u, x) = \sum_{i=1}^{m} \begin{cases} 0 & \text{if } h_i(x) \leq u_i \\ \frac{1}{2\lambda}(u - h_i(x))^2 & \text{if } h_i(x) > u_i \end{cases} \).

Therefore we conclude that

\[
f_\lambda(u, x) = s_\lambda (h_0(x) + p_\lambda(u, x)). \tag{2.30}
\]

as desired. In particular, notice that for \( u = 0 \), \( f_\lambda(0, x) = \phi_\lambda(x) \) is exactly the function resulting from replacing, in the definition of \( \phi \), the constraints \( h_i(x) \leq 0 \) by quadratic penalties, and then smoothing the result. This fact was already shown in [Goeb08].

The problem \( (P_\lambda) \) is an unconstrained, differentiable, and convex optimization problem. Therefore, the perturbation technique performed in the previous section does not apply for \( (P_\lambda) \).

### 2.6.3 Second Regularization Scheme

In this section, another approximation scheme to a convex program is presented. The same assumptions are kept for the primal problem \( (P) \), but we make the state constraints explicit. For a closed and convex set \( C \) in \( \mathbb{R}^n \),

\[
(P) \quad \begin{cases} 
\inf_{x} & h_0(x) \\
\text{subject to} & h_i(x) \leq 0, \text{ for } i = 1, 2, \ldots, m \\
x \in C
\end{cases}
\]
with associated Lagrangian

\[ L(y, x) = \begin{cases} 
  h_0(x) + \sum_{i=1}^{m} y_i h_i(x) & \text{if } x \in C \text{ and } y \geq 0 \\
  \infty & \text{if } x \notin C \text{ and arbitrary } y \\
  -\infty & \text{otherwise}.
\end{cases} \]

and representation

\[ f(u, x) = \sup_y \{ L(y, x) - \langle y, u \rangle \} = \begin{cases} 
  h_0(x) & \text{if } x \in C \text{ and } h_i(x) \leq u_i, \text{ for } i = 1, 2, \ldots, m \\
  \infty & \text{otherwise}.
\end{cases} \]

and value function \( V \) given by (2.24).

We regularize (P) by means of the family \((P^\lambda)\) as follows

\[ (P^\lambda) \begin{cases} 
  \inf h^\lambda_0(x) \\
  \text{subject to } h^\lambda_i(x) \leq 0, \text{ for } i = 1, 2, \ldots, m \\
  x \in C.
\end{cases} \]

where \( h^\lambda_i \) is a proper, lsc and convex function for \( i = 0, \ldots, m \), and consider the associated Lagrangian \( L^\lambda \) to this problem. The following theorem by Attouch and Wets finds conditions that guarantee that \( \{ L^\lambda \} \) will hypo/epi-converge to \( L \).

**Theorem 2.6.4** (Proposition 1.17, [Atto83a]). Suppose that \( \{ h^\lambda_0 \} \) epi-converges to \( h_0 \) as \( \lambda \searrow 0 \) and for all \( i = 1, 2, \ldots, m \), \( \{ h^\lambda_i \} \) converges continuously to \( h_i \) as \( \lambda \searrow 0 \). Then, for each \( \lambda \), the associated Lagrangian \( L^\lambda \) is a proper, closed saddle function, \( T^\lambda = L^\lambda \), and the family \( \{ L^\lambda \} \) hypo/epi-converge to \( L \).

**Proof.** The stated properties of \( L^\lambda \) hold as a consequence of the discussion right before equation (2.27). For the hypo/epi-convergence, given \( x \) and \( y \), we need to verify, according to (1.37) and (1.38), that

(a) for every sequence \( \lambda_n \searrow 0 \) and \( x_n \to x \), there exists a sequence \( y_n \to y \) such that \( \liminf_n L^\lambda_n(y_n, x_n) \geq L^\lambda(y, x) \),

(b) for every sequence \( \lambda_n \searrow 0 \) and \( y_n \to y \), there exists a sequence \( x_n \to x \) such that \( \limsup_n L^\lambda_n(y_n, x_n) \leq L^\lambda(y, x) \).

For (a), first we have that

\[ L^\lambda_n(y, x) = \begin{cases} 
  h^\lambda_n(x) + \sum_{i=1}^{m} y_i h^\lambda_i(x) & \text{if } x \in C \text{ and } y \geq 0 \\
  \infty & \text{if } x \notin C \text{ and } y \geq 0 \\
  -\infty & \text{otherwise}.
\end{cases} \]
So, let \( \lambda_n \searrow 0 \), \( x_n \to x \) and define \( y_n = y \) for every \( n \). If \( y \not\preceq 0 \), then clearly \( L^{\lambda_n}(y_n, x_n) = L^{\lambda_n}(y, x) = -\infty \), so (a) holds. Suppose that \( y \succeq 0 \). Then, in this case, \( L^{\lambda_n}(y_n, x_n) = L^{\lambda_n}(y, x_n) \). If \( x \notin C \), since \( C \) is closed, \( x_n \notin C \) for \( n \) sufficiently large, thus
\[
\infty = \lim \inf_n L^{\lambda_n}(y_n, x_n) \geq L^\lambda(y, x) = \infty.
\]
On the other hand, if \( x \in C \), we have that
\[
\lim \inf_n L^{\lambda_n}(y, x_n) = \lim \inf_n \left\{ h_0^{\lambda_n}(x_n) + \sum_{i=1}^{m} y_i h_i^{\lambda_n}(x_n) \right\}
\geq \lim \inf_n \left\{ h_0^{\lambda_n}(x_n) \right\} + \sum_{i=1}^{m} y_i \lim \inf_n \{ h_i^{\lambda_n}(x_n) \}
\geq h_0(x) + \sum_{i=1}^{m} y_i h_i(x) \quad \text{(by convergence hypothesis)}
= L(y, x)
\]
This proves (a). For (b), let \( \lambda_n \searrow 0 \) and \( y_n \to y \). We need to find \( x_n \to x \) such that the inequality in (b) holds. If \( x \notin C \), then taking \( x_n = x \), we obtain
\[
\infty = L^{\lambda_n}(y_n, x_n) \leq L^\lambda(y, x) = \infty.
\]
So assume that \( x \in C \). If \( y \not\preceq 0 \), then \( y_n \not\preceq 0 \) eventually, and so again taking \( x_n = x \),
\[
-\infty = L^{\lambda_n}(y_n, x_n) \leq L^\lambda(y, x) = -\infty.
\]
On the other hand, if \( y \succeq 0 \), since \( h_0^{\lambda_n} \) epi-converges to \( h_0 \), there exists \( x_n \to x \) such that
\[
\lim \sup_n \{ h_0^{\lambda_n}(x_n) \} \leq h_0(x).
\] (2.31)
Since each \( h_i^{\lambda_n} \) converges continuously to \( h_i \), then for each \( i \),
\[
\lim \sup_n \{ h_i^{\lambda_n}(x_n) \} \leq h_i(x).
\] (2.32)
The inequality in (b) holds in this case after combining inequalities (2.31) and (2.32). This finishes the proof. \( \square \)

The following is the main result of the section.

**Theorem 2.6.5.** For the problem \((P^\lambda)\) as before, let \( h_i^{\lambda} = e_\lambda h_i \), for \( i = 0, \ldots, m \). Assume either one of the following conditions

(a) the functions \( h_i \) are continuous on \( \mathbb{R}^n \),

(b) whenever \( x_m \to x \) and \( \lambda_m \searrow 0 \), the sequence \( \{ \|x_m - x\| / \lambda_m \} \) is bounded.
Then the family \( \{L^\lambda\} \) hypo/epi-converge in the modulated sense to \( L \). Accordingly, if a subsequence \( \{(y_{\lambda_k}, x_{\lambda_k})\} \) converges to some value \((\bar{y}, \bar{x})\), where \( x_{\lambda_k} \) solves \((P^\lambda)\) and \( y_{\lambda_k} \) is an associated Lagrange multiplier, then \( \bar{x} \) solves \((P)\) with associated multiplier \( \bar{y} \).

**Proof.** By Proposition 2.1.8, \( e_{\lambda} h_0 \) epi-converges to \( h_0 \) as \( \lambda \downarrow 0 \). Either one of the conditions ensures that \( e_{\lambda} h_i \) converges continuously to \( h_i \) for \( i = 1, \ldots, m \), see Theorem 2.1.6. Therefore, the Lagrangians \( \{L^\lambda\} \) hypo/epi-converges to \( L \) by Theorem 2.6.5.

By definition, \( \{L^\lambda\} \) is modulated if there exists \( \rho > 0 \) such that

\[
\inf_{\|x\| \leq \rho} L^\lambda(y, x) \leq \rho \left(1 + \|y\|\right), \quad \text{for all } y, \tag{2.33}
\]

\[
\sup_{\|y\| \leq \rho} L^\lambda(y, x) \geq -\rho \left(1 + \|x\|\right), \quad \text{for all } x. \tag{2.34}
\]

for all positive \( \lambda \) sufficiently small. To show (2.33), let \( \rho_1 > 0 \) be such that the convex set \( C \) contains at least one element of norm less than \( \rho_1 \). If \( y \not\preceq 0 \), \( L^\lambda(y, x) = -\infty \) for \( x \in C \) and the inequality (2.33) is satisfied. Suppose then that \( y \succeq 0 \). Define

\[
M_{\rho_1} = \max_{i=1,\ldots,m, \|x\| \leq \rho_1} \{|e_{\lambda} h_i(x)|\}
\]

and define

\[
\rho = \max\{\rho_1, (\sqrt{m} + 1)M_{\rho_1}\}
\]

Then, for each \( \lambda > 0 \),

\[
\inf_{\|x\| \leq \rho} L^\lambda(y, x) \leq \inf_{\|x\| \leq \rho_1} L^\lambda(y, x)
\]

\[
= \inf_{\|x\| \leq \rho_1} e_{\lambda} h_0(x) + \sum_{i=1}^{m} y_i e_{\lambda} h_i(x) + \delta_C(x)
\]

\[
\leq M_{\rho_1} + \sqrt{m}M_{\rho_1} \|y\| \quad (\text{by the Cauchy-Schwartz inequality})
\]

\[
\leq \rho (1 + \|y\|)
\]

and thus (2.33) holds. Now, to verify (2.34), notice that, for all positive \( \rho \) and \( \lambda \),

\[
\sup_{\|y\| \leq \rho} L^\lambda(y, x) \geq L^\lambda(0, x)
\]

\[
= e_{\lambda} h_0(x) + \delta_C(x).
\]

If \( x \notin C \), (2.34) holds trivially; otherwise, the last expression equals \( e_{\lambda} h_0(x) \), which epi-converges to \( h_0 \), thus this guarantees the existence of some \( \tilde{\rho} \) such that

\[
e_{\lambda} h_0(x) \geq -\tilde{\rho} \left(1 + \|x\|\right) \quad \text{for all } x \in C
\]

across to Example 7.34, [Rock98]. Consequently, the same inequality holds for

\[\sup_{\|y\| \leq \tilde{\rho}} L^\lambda(y, x),\]

and this means that (2.34) is satisfied, as desired. Therefore, \( \{L^\lambda\} \) is modulated as \( \lambda \downarrow 0 \).
Finally, the statement on the convergence of optimal points and multipliers is a consequence of the convergence of saddle points of hypo/epi-convergent sequences shown in Theorem 1.2.27.

**Corollary 2.6.6.** Under the hypothesis of the previous theorem, let

\[ f^\lambda(u, x) = e^\lambda h_0(x) + \delta_C(x) + \sum_{i=1}^m \delta_{(-\infty, 0]}(e^\lambda h_i(x) - u_i). \]

Then \( f^\lambda(\cdot, \cdot) \) is a representation for \( (P^\lambda) \) and \( -(f^\lambda)^*(-\cdot, \cdot) \) is a representation for the dual problem, identified by \( (D^\lambda) \). Moreover, as \( \lambda \downarrow 0 \),

(i) \( f^\lambda(\cdot, \cdot) \) epi-converges to \( f(\cdot, \cdot) \),

(ii) \( -(f^\lambda)^*(-\cdot, \cdot) \) hypo-converges to \( -f^*(-\cdot, \cdot) \).

Furthermore, let \( V^\lambda(u) = \inf_x f^\lambda(u, x) \) and \( \tilde{V}^\lambda(w) = \sup_y -(f^\lambda)^*(-y, w) \) be the primal and dual value functions respectively. Then, as \( \lambda \downarrow 0 \),

(iii) \( V^\lambda(\cdot) \) epi-converges to \( V(\cdot) \),

(iv) \( \tilde{V}^\lambda(w) \) hypo-converges to \( \tilde{V}(w) \).

**Remark 2.6.7** In the previous results in this section, \( e^\lambda \) can be substituted by \( s^\lambda \) and the results hold as well.

In the problem \( (P^\lambda) \), the Lagrangian \( L^\lambda \) is not everywhere differentiable, as it is the case in \( (P_\lambda) \); however, on the interior of its domain, it is differentiable and has Lipschitz gradient. We can then rely on necessary conditions such as the Karush-Kuhn-Tucker (KKT) conditions to find explicit equations that must be satisfied by the optimal values and Lagrange multipliers.

In fact, a pair \( (x^\lambda, y^\lambda) \in \text{int dom } L^\lambda \) is an optimal pair in the sense that \( x^\lambda \) is optimal for the problem \( (P^\lambda) \) and \( y^\lambda = (y^\lambda_1, \ldots, y^\lambda_m) \) is an associated multiplier if and only if the following KKT conditions hold:

\[
\begin{align*}
s^\lambda h_i(x^\lambda) &\leq 0 \\
\nabla s^\lambda h_0(x^\lambda) + \sum_{i=1}^m y^\lambda_i \nabla s^\lambda h_i(x^\lambda) &\leq 0 \\
\nabla_i h^\lambda &\geq 0 \\
y^\lambda_i s^\lambda h_i(x^\lambda) &\geq 0
\end{align*}
\]

and in this case \( (x^\lambda, y^\lambda) \) is a saddle point of \( L^\lambda \).

**Definition 2.6.8** For the convex program \( (P) \), let

\[ J = \{ i \in \{1, \ldots, m\} \mid h_i \text{ is affine and } h_i = -h_j \text{ for some } j \}. \]

It is said that \( (P) \) satisfies Slater’s condition if there exists \( x \in C \) such that \( h_j(x) \leq 0 \) for \( j \in J \) and \( h_j(x) < 0 \) for \( j \notin J \). The point \( x \) is called a Slater point for \( (P) \).
Slater’s condition is a constraint qualification that ensures
(a) attainment in the dual problem (provided it is feasible) and
(b) zero duality gap
in convex programs. So, we study the circumstances under which Slater’s condition
is satisfied in \((P^\lambda)\) by using either regularization \(e_\lambda\) or \(s_\lambda\).

**Proposition 2.6.9.** Assume that Slater’s condition is satisfied for the convex problem \((P)\). Then

(a) if the problem \((P^\lambda)\) is defined by means of \(e_\lambda\), then Slater’s condition for
\((P^\lambda)\) is satisfied for every \(\lambda > 0\),

(b) if the problem \((P^\lambda)\) is defined by means of \(s_\lambda\), then there exists \(\lambda_0 > 0\) such
that Slater’s condition for \((P^\lambda)\) is satisfied for every \(0 < \lambda < \lambda_0\).

**Proof.** For simplicity, we assume no affine constraints are present in \((P)\). The other
case can be handled after a minor change in the proof. So, assume Slater’s condition
holds for \((P)\). Then there exists a Slater point \(z\), that is, there exists \(z \in C\) with
\(h_i(z) < 0\), \(i = 1, \ldots, m\). Consequently, \(e_\lambda h_i(z) \leq h_i(z) < 0\), and \(z\) is a Slater point
for the problem \((P^\lambda)\) is defined by means of \(e_\lambda\). This shows (a). For (b), choose \(\lambda_0\)
such that
\[
\max_{i=1,\ldots,m} \left\{ \frac{\|x\|^2}{-2h_i(z)} \right\} < \frac{1}{\lambda_0} - \lambda_0.
\]
Since for each \(\lambda > 0\) and \(i = 1, \ldots, m\), \(e_\lambda h_i(z) \leq h_i(z) < 0\), it is clear that
\[
\frac{\|z\|^2}{-2e_\lambda h_i(z)} \leq \frac{\|z\|^2}{-2h_i(z)}.
\]
Thus, for \(0 < \lambda < \lambda_0\) and \(i = 1, \ldots, m\), the inequality
\[
\frac{\|z\|^2}{-2e_\lambda h_i(z)} < \frac{1}{\lambda_0} - \lambda_0 < \frac{1}{\lambda} - \lambda.
\]
holds. Rearranging terms, the last expression is equivalent to
\[
(1 - \lambda^2) e_\lambda h_i(z) + \frac{\lambda}{2} \|z\|^2 < 0,
\]
that is, \(s_\lambda h_i(z) < 0\). Therefore, \(z\) is a Slater point for \((P^\lambda)\). This completes the proof.
Chapter 3
Calculus of Variations

In this chapter we identify properties of fully convex problems in calculus of variations under the Moreau envelope and the regular approximation introduced in Chapter 2.

3.1 Calculus of Variations Under Full Convexity Assumptions

The interest in (time-independent) calculus of variation centers about finding the minimum value of the integral functional

$$J_\tau(x(\cdot)) = g(x(0)) + \int_0^\tau L(x(t), \dot{x}(t)) dt$$  \hspace{1cm} (3.1)

where \(\tau \in [0, \infty)\), \(g : \mathbb{R}^n \to \mathbb{R} = [-\infty, \infty]\) is the initial cost function, and \(L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\) called the Lagrangian, which dictates the way that the initial cost function propagates forward in time, and the arc \(x\) belongs to some specific function space.

Therefore, special attention is placed on the value function \(V : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}\) defined as

$$V(\tau, \xi) = \inf \{ J_\tau(x(\cdot)) \mid x(\cdot) \in A^1_n[0, \tau], x(\tau) = \xi \} ,$$

$$V(0, \xi) = g(\xi)$$  \hspace{1cm} (3.2)

where \(A^p_n[\tau_1, \tau_2]\) denotes the space of absolutely continuous maps \(x(\cdot) : [\tau_1, \tau_2] \to \mathbb{R}^n\) with derivative \(\dot{x}(t) \in L^p_n[\tau_1, \tau_2]\).

We adopt the setting introduced by Rockafellar and Wolenski in [Rock01a]. The basic assumptions introduced there are denoted by (A).

(A0) The initial cost function \(g\) is convex, proper, and lsc on \(\mathbb{R}^n\).

(A1) The Lagrangian function \(L\) is convex, proper, and lsc on \(\mathbb{R}^n \times \mathbb{R}^n\).

(A2) The set \(F(x) := \text{dom}L(x, \cdot)\) is nonempty for all \(x\), and there is a constant \(\rho\) such that \(\text{dist}(0, F(x)) \leq \rho(1 + |x|)\) for all \(x\).

(A3) There are constants \(\alpha\) and \(\beta\) and a coercive, proper, nondecreasing function \(\theta\) on \([0, \infty]\) such that

$$L(x, v) \geq \theta(\max \{0, |v| - \alpha |x|\}) - \beta |x|$$

for all \(x\) and \(v\).
Let us discuss the assumptions (A). The joint convexity of \( L \) in (A1) together with the convexity of \( g \) in (A0) guarantee that the functional \( J_\tau \) in (3.1) is well-defined and convex in \( A^1_n[0, \tau] \). The way \( \infty \) is admitted in the definition of \( L \) and \( g \) allows to introduce certain abstract constraints on the problem. For instance, according to (A2), \( J_\tau(x(\cdot)) = \infty \) unless the arc \( x(\cdot) \) satisfies the constraints

\[
\dot{x}(t) \in F(x(t)) \quad \text{a.e.} \ t, \text{with } x(0) \in D := \text{dom} \ g.
\]

Thus the Lagrangian \( L \) implicitly defines a differential inclusion in terms of the mapping \( F \). Also, the nonemptiness of \( F(x) \) implies that the function \( x(\cdot) \) has no restriction on which values to attain, that is, there are no state constraints implicitly imposed by \( L \). The growth condition in (A2) means that the differential inclusion has no “forced escape time”. This means that from any point it provides at least one trajectory over \([0, \infty)\). Moreover, the function \( L(x, \cdot) \) is convex by (A1), proper by (A2) and coercive by (A3).

Next, we introduce the convex duality theory in this problem. Define the dual Lagrangian \( \tilde{L} \) by

\[
\tilde{L}(y, w) = \sup_{x,v} \left\{ \langle x, w \rangle + \langle v, y \rangle - L(x,v) \right\}
\]

and the dual value function \( \tilde{V} \) formed by replacing \( g \) and \( L \) by \( g^* \) and \( \tilde{L} \) respectively in (3.2); that is,

\[
\tilde{V}(\tau, \xi) := \inf \left\{ g^*(x(0)) + \int_0^\tau \tilde{L}(x(t), \dot{x}(t))dt \mid x(\cdot) \in A^1_n[0, \tau], x(\tau) = \xi \right\}, \quad (3.3)
\]

\[
\tilde{V}(0, \xi) = g^*(\xi)
\]

One of the advantages of the assumptions (A) is that they fully dualizable. In fact, by Proposition 3.5, [Rock01a],

- \( g \) satisfies (A0) if and only if \( g^* \) does,
- \( L \) satisfies (A1) if and only if \( \tilde{L} \).
- \( L \) satisfies (A2) if and only if \( \tilde{L} \) satisfies (A3).
- \( L \) satisfies (A3) if and only if \( \tilde{L} \) satisfies (A2).

Therefore, results that can be shown for \( V \) hold in a parallel form for \( \tilde{V} \).

The Hamiltonian is a key function in order to understand the relationship between optimal solutions of the problem defining \( V \) and optimal solutions of the problem defining \( \tilde{V} \). It is defined as the partial conjugate of \( L \) on its second variable, namely,

\[
H(x, y) = \sup_v \left\{ v \cdot y - L(x, v) \right\}. \quad (3.4)
\]
The properness, lower semicontinuity and joint convexity of $L$ imply that $H$ is a proper and closed saddle function, which is the lower closure of its equivalence class, according to equation (1.22) in Theorem 1.2.11. Moreover, the coercivity assumption on $L$ implies that actually $H$ is everywhere finite, thus the only member of its equivalence class.

The following theorem illustrates how assumptions in (A) transform into properties for the Hamiltonian.

**Theorem 3.1.1.** A function $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is the Hamiltonian of a Lagrangian satisfying (A1), (A2) and (A3) if and only if $H(x, y)$ is everywhere finite, concave in $x$, convex in $y$, and the following growth conditions hold, where (a) corresponds to (A3) and (b) corresponds to (A2)

(a) There are constants $\alpha$ and $\beta$ and a finite, convex function $\varphi$ such that

$$H(x, y) \leq \varphi(y) + (\alpha |y| + \beta) |x|$$

for all $x, y$.

(b) There are constants $\gamma$ and $\delta$ and a finite, concave function $\psi$ such that

$$H(x, y) \geq \psi(y) - (\gamma |x| + \delta) |y|$$

for all $x, y$.

**Corollary 3.1.2.** Under assumptions (A), the Hamiltonian $H$ given by (3.4) is locally Lipschitz continuous and has a nonempty closed, convex, and bounded subdifferential at each $(x, y)$.

Moreover, an arc $(x(\cdot), y(\cdot)) \in A^1_{2n}[\tau_0, \tau_1]$ is a Hamiltonian trajectory if, for a.e. $t$,

$$\dot{x}(t) = \partial_y H(x(t), y(t)), \quad \dot{y}(t) = -\partial_x H(x(t), y(t))$$

where $\partial_y H(x(t), y(t))$ denotes the subdifferential of the saddle function $H$, as defined in Chapter 1.

The following result provides a characterization of the minimizers of problems (3.2) and (3.3) as Hamiltonian trajectories over $[0, \tau]$. More precisely,

**Theorem 3.1.3** (Theorem 2.4, [Rock01a]). A pair of arcs $x(\cdot)$ and $y(\cdot)$ gives a Hamiltonian trajectory over $[0, \tau]$ that starts in graph $\partial g$ and ends at a point $(\xi, \eta) \in \text{graph } \partial V$ if and only if

(a) $x(\cdot)$ is optimal in the minimization problem in (3.2) that defines $V(\tau, \xi)$

(b) $y(\cdot)$ is optimal in the minimization problem in (3.3) that defines $\tilde{V}(\tau, \eta)$

Associated with the trajectories, we have the Hamiltonian flow, which is the one-parameter family of set-valued mappings $S_\tau$ defined by

$$S_\tau(\xi_0, \eta_0) := \{ (\xi, \eta) | \exists \text{ a Hamiltonian trajectory over } [0, \tau]\text{ from } (\xi_0, \eta_0) \text{ to } (\xi, \eta) \}$$
The following theorem shows that the dynamics of
\[ \text{graph} \, \partial V_{\tau} := \{(\xi, \eta)|\eta \in \partial V_{\tau}\} \]
are determined by the Hamiltonian flow.

**Theorem 3.1.4.** Under (A), the following statements are equivalent

(I) \( \eta \in \partial V_{\tau} \).

(II) for some \( \eta_0 \in \partial g(\xi_0) \), there is a Hamiltonian trajectory \((x(\cdot), y(\cdot))\) over \([0, \tau]\)
with \((x(0), y(0)) = (\xi_0, \eta_0)\) and \((x(\tau), y(\tau)) = (\xi, \eta)\).

### 3.1.1 Value Function Regularization

The idea of this section is to regularize the proper, lsc and convex value function \( V_{\tau} \) by applying to it either \( e_\lambda \) or \( s_\lambda \) and study the regularity properties of the gradient of such regularization. For convenience, we will write \( \bar{P}_\lambda(\tau, \xi) = P_\lambda V_{\tau}(\xi) \), \( \bar{Q}_\lambda(\tau, \xi) = Q_\lambda V_{\tau}(\xi) \), \( \bar{e}_\lambda(\tau, \xi) = e_\lambda V_{\tau}(\xi) \), and \( \bar{s}_\lambda(\tau, \xi) = s_\lambda V_{\tau}(\xi) \).

The properties of the Moreau envelope and the regular approximation in Theorem 2.4.1 imply that, for fixed \( \tau \) and \( \lambda \in (0, 1) \), the gradient mappings \( \nabla \bar{e}_\lambda(\tau, \cdot) \) and \( \nabla \bar{s}_\lambda(\tau, \cdot) \) are locally Lipschitz continuous. The goal now is to show that the latter is true jointly with respect to \((\tau, \zeta)\). This was shown by Rockafellar in [Rock05] for the particular case \( \lambda = 1 \).

**Theorem 3.1.5** (Theorem 4 and Corollary, [Rock04]). For \( \lambda > 0 \), the Moreau envelope function
\[ \bar{e}_\lambda(\tau, \xi) = \min_{\zeta \in \mathbb{R}^n} \left\{ V_{\tau}(\zeta) + \frac{\|\xi - \zeta\|^2}{2\lambda} \right\} \]
is continuously differentiable with respect to \((\lambda, \xi, \tau)\). Moreover,
\[ \frac{\partial (\bar{e}_\lambda(\tau, \xi))}{\partial \tau} = -H(x, y) \]
where \( x = \bar{P}_\lambda(\tau, \xi) \) and \( y = \bar{Q}_\lambda(\tau, \xi) \).

The following is the most important theorem of this section.

**Theorem 3.1.6.** Under assumptions (A), for each \( \lambda > 0 \), the maps \( \bar{P}_\lambda(\tau, \zeta) \), \( \bar{Q}_\lambda(\tau, \zeta) \), \( \nabla \bar{e}_\lambda(\tau, \zeta) \) and \( \nabla \bar{s}_\lambda(\tau, \zeta) \) are locally Lipschitz continuous jointly with respect to \((\tau, \zeta)\).

**Proof.** Fix \( \lambda > 0 \). First we show the claim for \( \bar{P}_\lambda \). By Lemma 2.3.6,
\[ \|\bar{P}_\lambda(\tau, \zeta) - \bar{P}_\lambda(\tau, \zeta')\| \leq \|\zeta - \zeta'\| \] for all \( \zeta, \zeta' \in \mathbb{R}^n \), and \( \tau \geq 0 \).
so the result will follow if it is shown that \( \bar{P}_\lambda \) is locally Lipschitz in \( \tau \) with a Lipschitz constant that is locally uniform in \( \zeta \). Fix \( \tau^* > 0 \) and \( \zeta^* \in \mathbb{R}^n \), and define \( M : [0, \infty) \times \mathbb{R}^n \rightarrow \mathbb{R}^{2n} \) by
\[ M(\tau, \zeta) = (\bar{P}_\lambda(\tau, \zeta), \bar{Q}_\lambda(\tau, \zeta)). \]
The continuity of $M$ is a consequence of the continuity of both $P$ and $Q$ (Lemma 2.3.6). Thus, given a compact neighborhood $T_0 \times Z_0$ of $(\tau^*, \xi^*)$, the image set $M(T_0 \times Z_0)$ is a compact set in $\mathbb{R}^{2n}$. Since the Hamiltonian map $H$ is locally Lipschitz continuous on $\mathbb{R}^{2n}$, it is possible to find compact sets $U_0$ and $U_1$ satisfying

- $M(T_0, Z_0) \subset U_1 \subset \text{int} U_0$
- for $(\xi, \eta) \in U_0$, if $u \in \partial_x H(\xi, \eta)$ and $v \in \partial_y H(\xi, \eta)$, then there exists a constant $\kappa > 0$ such that $\|u\| \leq \kappa$, $\|v\| \leq \kappa$.

Therefore, Hamiltonian trajectories $(x, y)$ over time intervals in which they are contained in $U_0$ must necessarily be Lipschitz continuous with constant $\kappa$. One can choose an interval $T_1$ small enough such that $\tau^* \in T_1$, $T_1 \subset T_0$, and each Hamiltonian trajectory over $T_1$ that touches $U_1$ remains inside of $U_0$. This choice of $T_1$ guarantees that Hamiltonian trajectories over that interval are locally Lipschitz of constant $\kappa$. Finally, it is possible to choose a neighborhood $T \times Z$ of $(\tau^*, \xi^*)$ that is contained in $T_0 \times Z_0$ such that $M(T \times Z) \subset U_1$.

Now, pick $\zeta \in Z$ and two values $\tau < \tau'$ such that the interval $[\tau, \tau'] \subset T$. Let $\xi = P_\lambda(\tau, \zeta)$, $\eta = Q_\lambda(\tau, \zeta)$. Then $(\xi, \eta) \in U_1$ and

$$\lambda^{-1} \eta \in \partial f(\xi) \quad \text{and} \quad \zeta = \xi + \eta$$

in virtue of Corollary 2.3.4. According to Theorem 2.4 in [Rock01a], there exists a Hamiltonian trajectory $(x, y)$ on $[0, \tau]$ satisfying $(x(\tau), y(\tau)) = (\xi, \lambda^{-1} \eta)$. Moreover, the trajectories can be extended to the interval $[\tau, \tau']$, and over that interval they are Lipschitz continuous with constant $\kappa$, by the previous considerations.

Let $t \in [\tau, \tau']$ and on that interval define $z(t) = x(t) + y(t)$. Then $z(\tau) = \zeta$ and $z$ is Lipschitz continuous of constant $2\kappa$. Moreover, the inclusion $y(t) \in \partial f(x(t))$ holds in one substitutes $\tau$ by $t$. Consequently, using again Corollary 2.3.4,

$$x(t) = P_\lambda(t, z(t)), \quad \lambda y(t) = Q_\lambda(t, z(t)) \quad \text{for } t \in [\tau, \tau'].$$ 

Hence,

$$\|P_\lambda(\tau', \zeta) - P_\lambda(\tau, \zeta)\|$$

$$= \|P_\lambda(\tau', \zeta) - P_\lambda(\tau', z(\tau')) + P_\lambda(\tau', z(\tau')) - P_\lambda(\tau, \zeta)\|$$

$$\leq \|P_\lambda(\tau', z(\tau)) - P_\lambda(\tau', z(\tau'))\| + \|P_\lambda(\tau', z(\tau')) - P_\lambda(\tau, z(\tau))\|$$

$$\leq \|z(\tau) - z(\tau')\| + \|x(\tau') - x(\tau)\|$$

$$\leq 3\kappa |\tau' - \tau|.$$
Since $\zeta \in Z$ and $[\tau, \tau'] \subset T$ were arbitrarily chosen, this proves that $\bar{P}_\lambda$ is locally Lipschitz in $\tau$ with a Lipschitz constant that is locally uniform in $\zeta$, and this completes the claim on $\bar{P}_\lambda$. The result holds for $\bar{Q}_\lambda$ since $\bar{Q}_\lambda = I - \bar{P}_\lambda$.

The only statement left to prove is the local Lipschitz property on $\nabla \bar{e}_\lambda$. From Corollary 2.3.4, $\nabla \bar{e}_\lambda(\tau, \cdot) = \frac{1}{\lambda} Q_\lambda(\tau, \cdot)$, thus $\nabla_\zeta \bar{e}_\lambda(\tau, \zeta) = \frac{1}{\lambda} Q_\lambda(\tau, \zeta)$. From Theorem 3.1.5,

$$\frac{\partial (\bar{e}_\lambda(\tau, \xi))}{\partial \tau} = -H(x, y)$$

where $x = \bar{P}_\lambda(\tau, \xi)$ and $y = \bar{Q}_\lambda(\tau, \xi)$. Consequently,

$$\nabla (\bar{e}_\lambda(\tau, \xi)) = \left(-H(\bar{P}_\lambda(\tau, \xi), \bar{Q}_\lambda(\tau, \xi)), \frac{1}{\lambda} Q_\lambda(\tau, \zeta)\right).$$

The map $H$ is (jointly) locally Lipschitz and it was just shown that the maps $\bar{P}_\lambda$ and $\bar{Q}_\lambda$ are also (jointly) locally Lipschitz. Therefore, the same must hold for $\nabla \bar{e}_\lambda$ by the last formula, and in turn for $\nabla \bar{s}_\lambda$, by the formula of $s_\lambda$. This completes the proof. \qed
References


Appendix: Symbolic Computation of the Regular Approximation

Here we include a code that computes the regular approximation $s_\lambda$ for a function of one variable. The function can be added to the SCAT (Symbolic Convex Analysis Toolbox) software, created by Jonathan M. Borwein and Chris H. Hamilton [Borw09]. The code, user guide and several examples are freely available at http://ddrive.cs.dal.ca/projects/scat/

# Performs regularizing envelope on a one-dimensional PWF.
RegApprox := proc(_pwf::PWF,val,constant)
options remember:
local pwf, v, val1, val2, quadc, quad, constantc1, constantc2,
summand1, summand2, c:

# Get the PWF
pwf := _pwf:

# Get its variable
val1 := op(2,pwf):

# Get the variable for the smoothing
v := val:

# Get the constant lambda in the regularization
c :=constant:

# Ensure non-zero dimension
if nops(v) = 0 then
error "zero dimension PWF":
end if:

# Computes quadratic function with parameter c
quadc := convert(1/(2*c)*v*v, PWF, v, {c >0,c<1}):
val2 := op(2,quadc):

# Compare the variable of the input and the quadratic function
if val2 <> val1 then
error "variable mismatch":
end if:

# Defines quadratic function
quad := convert( v*v, PWF, v, {c >0,c<1}):

# Defines constant 1-c*c
constantc1 := convert(1-c*c,PWF,v,{c >0,c<1}):

# Defines constant c/2
constantc2 := convert(c/2,PWF,v,{c >0,c<1}):

# Computes first summand in the definition of the smoothing
summand1 := simplify('*'(constantc1,InfConv(pwf,quadc))):

# Computes second summand in the definition of the smoothing
summand2 := simplify('*'(constantc2,quad)):

# Computes regular approximate by adding both summands
pwf := simplify('+'(summand1,summand2)):

return pwf:
end proc:

Example 3.1.7 The following code computes the self-dual smoothing of the function

\[ f(x) = \begin{cases} 
-x & \text{if } -2 \leq x \leq 0 \\
2x & \text{otherwise}
\end{cases} \]

and verifies its self-duality.

f := piecewise( x<=0,-x,2*x);
f1 := convert( f, PWF, [x=-2..1]);
regf1 := RegApprox(f1,x,lambda);
f1star := Conj(f1,y);
regf1star := RegApprox(f1star,y,lambda);
starregf1 := simplify(Conj(regf1,y));
Equal(regf1star,starregf1);
Vita

Alvaro José Guevara was born on June, 1980, in San José, Costa Rica. He finished his undergraduate studies at the Universidad de Costa Rica in August of 2002. In August of 2003 he came to Louisiana State University and earned a master of science in mathematics in 2006. Currently, he is a candidate for the degree of Doctor of Philosophy in mathematics, which will be awarded in August 2009.