Introduction to convex optimization

Teemu Pennanen

January 8, 2019

Contents

Preface

Optimization problems arise whenever decisions are to be made. Many phenomena in natural sciences can also be described in terms of minima or maxima of certain functions. In the words of Leonhard Euler, *"...nothing whatsoever takes place in the universe in which some relation of maximum or minimum does not appear"*.

These lecture notes study general aspects of *convex optimization problems* where one seeks to minimize a convex function over a linear space. In a certain sense, convex optimization problems form the nicest class of optimization problems. Stronger analytical results e.g. on the existence and uniqueness of solutions and on optimality conditions are available as soon as the problem is known to be convex. Convexity is essential also in the duality theory of optimization. Under convexity, one can treat nonsmooth and infinite-dimensional problems with the same ease as the smooth and finite-dimensional ones. Convexity is indispensable in numerical optimization in higher dimensions. When analyzing a given optimization problem, convexity is the first thing to look for.

Many aspects of more specific problem classes such as stochastic optimization, robust optimization, calculus of variations, optimal control, semidefinite programming and multicriteria optimization can be treated under convex analysis. Although convexity rules out some important problems e.g. in combinatorial optimization, it arises quite naturally in many applications. Classical application fields include operations research, engineering, physics and economics. More recently, convex optimization has found important applications in mathematical finance and financial engineering. Even some combinatorial problems can be analyzed with techniques of convex analysis. Selected applications will be treated in the following sections.

These notes study convex optimization in general topological vector spaces. The generality is motivated by various important applications e.g. in physics and financial economics which go beyond finite-dimensional spaces. In stochastic optimization and mathematical finance, one often encounters topological vector spaces which are not even locally convex. The material is divided into three chapters according to mathematical structure. After outlining some applications, the first chapter studies convex optimization in general (real) vector spaces. Chapter 2 studies optimization problems in topological vector spaces. The last chapter is devoted to duality theory in locally convex topological vector spaces. The necessary topological and functional analytic concepts will be introduced as needed. Most of the material in these notes is collected from [11], [13], [16] and [7].

Chapter 1

Convexity

This chapter studies convex optimization problems in general vector spaces. The aim is to summarize some basic facts in convexity that do not require topology or duality. Most results are quite simple but already useful in practice. We start by studying convex functions on the real line and then proceed to study convexity preserving operations that result in convex functions on more general vector spaces. Knowledge of such operations is useful in identification of convexity in applications as well as in building tractable optimization models in decision sciences.

We will also study differential and asymptotic scaling properties of convex sets and functions. Such properties are involved in conditions for optimality and existence of solutions, respectively. We will see that convexity makes the scaling behaviour remarkably tractable and allows for useful geometric interpretations and calculus rules in practice. We end this chapter by presentting the fundamental separation theorems that yield more familiar separation theorems when applied in topological vector spaces.

Before going to the general theory, we outline some applications where convexity arises quite naturally. That these problems are in fact convex, will be verified as applications as we proceed with the general theory.

1.1 Examples

This section discusses some well-known optimization problems where convexity has an important role. The choice of applications is rather arbitrarily and many important applications e.g. in statistics and inverse problems have been omitted.

Mathematical programming

Mathematical programming is the discipline that studies numerical solution of optimization problems mostly in finite-dimensional spaces of real or integer variables. We will be concerned with problems with real variables. Optimization routines are designed for classes of optimization problems that can be written in the form of a generic model. The classical *linear programming* model can be written as

$$\begin{aligned}&\text{minimize} \quad x\cdot c \quad \text{over} \quad x\in\mathbb{R}^n\\ &\text{subject to} \quad Ax\le b,\end{aligned}\tag{LP}$$

where $c\in\mathbb{R}^n$, $A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$ and $c\cdot x=\sum_{i=1}^n c_ix_i$. Linear programming models were studied already by Kantorovich in the 1930's as a general formalism in operations research; see [6]. They gained popularity with George Dantzig's invention of the *simplex* algorithm [3] which is still widely used. For large problems, modern *interior point* methods often outperform simplex methods both in theory and practice.

An important special case of the LP-model is the *network optimization* problem

$$\begin{aligned}&\text{minimize} && \sum_{(i,j)\in\mathcal{A}} c_{i,j}x_{i,j} \quad \text{over} \quad x\in\mathbb{R}^{\mathcal{A}}\\ &\text{subject to} && \sum_{\{j\,|\,(i,j)\in\mathcal{A}\}} x_{i,j}-\sum_{\{j\,|\,(j,i)\in\mathcal{A}\}} x_{j,i}=s_i \quad \forall i\in\mathcal{N}\\ &&& l_{i,j}\le x_{i,j}\le u_{i,j} \quad \forall (i,j)\in\mathcal{A},\end{aligned}$$

where $\mathcal{N}$ is the set of *nodes* and $\mathcal{A}\subseteq\{(i,j)\,|\,i,j\in\mathcal{N}\}$ is the set of *arcs* of a network. A notable feature of such models is that if the parameters s_i, $l_{i,j}$ and $u_{i,j}$ have integer values, then there is an optimal solution x with all $x_{i,j}$ *integer*; see e.g. Rockafellar [15] or Bertsekas [2]. Important special cases of the general network optimization problem include *shortest path*, *assignment*, *transportation* and *max-flow* problems. There are also natural infinite-dimensional extensions where the finite network is replaced by a measurable space $(\Omega,\mathcal{F})$ and flows are described in terms of measures on the product space $(\Omega\times\Omega,\mathcal{F}\otimes\mathcal{F})$. The most famous instance is the Monge-Kantorovich *mass transportation problem*; see e.g. [9].

Replacing the linear functions in the LP-model by more general ones, one obtains the *nonlinear programming* model

$$\begin{aligned}&\text{minimize} \quad f_0(x) \quad \text{over} \quad x\in\mathbb{R}^n\\ &\text{subject to} \quad f_j(x)\le 0, \quad j=1,\dots,m.\end{aligned}\tag{NLP}$$

This is still the most common format for computational optimization routines. More flexible and often more natural format is the *composite optimization* model

$$\text{minimize} \quad h(F(x)) \quad \text{over} \quad x\in\mathbb{R}^n, \tag{CO}$$

where $F=(f_0,\dots,f_m)$ is a vector-valued function from $\mathbb{R}^n$ to $\mathbb{R}^{m+1}$ and h is a function on $\mathbb{R}^{m+1}$. In this model, all the functions f_j are treated symmetrically much as in *multicriteria* optimization. Allowing h to take on infinite-values, one obtains (NLP) as a special case of (CO); see Section 1.4 below.

Stochastic optimization

Many optimization problems involve uncertain factors that have an essential effect on the outcome. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $f : \mathbb{R}^n \times \Omega \to \mathbb{R}$ be such that $\omega \mapsto f(x, \omega)$ is measurable for every $x \in \mathbb{R}^n$. The basic static (as opposed to dynamic) *stochastic optimization* model can be written as

$$\text{minimize} \quad Ef(x, \cdot) \quad \text{over} \quad x \in \mathbb{R}^n. \tag{SP}$$

Instead of the expectation, one could also minimize other functions of the random variable $\omega \mapsto f(x, \omega)$. This leads to models of the form

$$\text{minimize} \quad \mathcal{V}(f(x, \cdot)) \quad \text{over} \quad x \in \mathbb{R}^n, \tag{CSP}$$

where $\mathcal{V}$ is a function on the space of random variables.

Problems where some of the decisions may react to realizations of random variables, often lead to *dynamic stochastic programming* models of the form

$$\text{minimize} \quad Ef(x(\cdot), \cdot) \quad \text{over} \quad x \in \mathcal{N}, \tag{DSP}$$

where f is a $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F}$-measurable function on $\mathbb{R}^n \times \Omega$ and

$$\mathcal{N} = \{(x_t)_{t=0}^T \,|\, x_t \in L^0(\Omega, \mathcal{F}_t, P; \mathbb{R}^{n_t})\}$$

for given integers n_t such that $n_0 + \ldots + n_T = n$ and an increasing sequence $(\mathcal{F}_t)_{t=0}^T$ of sub-sigma-algebras of $\mathcal{F}$. Here $L^0(\Omega, \mathcal{F}_t, P; \mathbb{R}^{n_t})$ denotes the space of (equivalence classes of) $\mathcal{F}_t$-measurable $\mathbb{R}^{n_t}$-valued functions. This kind of models are typical e.g. in mathematical finance; see the examples below or [8].

Price formation in a centralized market

Most modern securities markets are based on the so called *double auction* mechanism where market participants submit buying or selling offers characterized by limits on quantity and unit price. For example, a selling offer consists of an offer to sell up to x units of a security at the unit price of p (units of cash). All the selling offers can be combined into a function $x \mapsto c(x)$, which gives the *marginal price* when buying x units. The quantity available at the lowest submitted selling price is finite and when buying more one gets the second lowest price and so on. The marginal price $c(x)$ is thus a piecewise constant nondecreasing function of x. Equivalently, the *cost function*

$$C(x) = \int_0^x c(z)dz$$

is a piecewise linear convex function on $\mathbb{R}_+$. Buying offers are combined analogously into a piecewise linear concave function $R : \mathbb{R}_+ \to \mathbb{R}$ giving the maximum *revenue* obtainable by selling a given quantity to the willing buyers.

Given the cost and revenue functions C and R, the market is *cleared* by solving the optimization problem

$$\text{maximize} \quad R(x) - C(x) \quad \text{over} \quad x \in \mathbb{R}_+.$$

When multiple solutions exist, the greatest solution $\bar{x}$ is implemented by matching $\bar{x}$ units of the cheapest selling offers with with $\bar{x}$ units of the most generous buying offers. The prices of the remaining selling offers are then all strictly higher than the prices of the remaining buying offers and no more trades are possible before new offers arrive.

The offers remaining after market clearing are recorded in the *limit order book*. It gives the marginal prices for buying or selling a given quantity at the best available prices. Interpreting negative purchases as sales, the marginal prices can be incorporated into a single function $x \mapsto s(x)$ giving the marginal price for buying positive or negative quantities of the commodity. Since the highest buying price is lower than the lowest selling price, the marginal price curve s is a nondecreasing piecewise constant function on $\mathbb{R}$, or equivalently, the *cost function*

$$S(x) = \int_0^x s(z)dz$$

is a convex piecewise linear function on $\mathbb{R}$.

Price formation in a decentralized market

Centralized markets are used for trading goods that can be transported without a cost. When there are significant costs or constraints on transportation, the *location* of supply and demand plays a role. Examples include electricity markets where the traded good is transported through a network of transmission lines that have finite capacity and transmission losses cannot be ignored.

Assume that there are selling and buying offers for electricity at a finite set $\mathcal{N}$ of locations. Denote the cost and revenue functions at location $i \in \mathcal{N}$ by C_i and R_i, respectively. Let $\mathcal{A} \subseteq \{(i,j) \mid i,j \in \mathcal{N}\}$ be the set of *transmission lines* with *capacities* $u = \{u_{i,j}\}_{(i,j)\in\mathcal{A}}$ and *transmission losses* $a = \{a_{i,j}\}_{(i,j)\in\mathcal{A}} \subset [0,1]$. The capacity $u_{i,j}$ gives the maximum amount of energy that can be sent through line (i,j). If an amount $f_{i,j}$ of energy is sent from i to j, only the fraction $a_{i,j}f_{i,j}$ reaches j.

The market is cleared by solving the optimization problem

$$\begin{array}{lll}
\underset{s,d,f}{\text{minimize}} & \displaystyle\sum_{i\in\mathcal{N}}[C_i(s_i) - R_i(d_i)] & \\
\text{subject to} & \displaystyle\sum_{\{j\mid(i,j)\in\mathcal{A}\}} f_{i,j} - \sum_{\{j\mid(j,i)\in\mathcal{A}\}} a_{j,i}f_{j,i} = s_i - d_i & \forall i \in \mathcal{N}, \\
& s_i, d_i \geq 0 & \forall i \in \mathcal{N}, \\
& f_{i,j} \in [0, u_{i,j}] & \forall (i,j) \in \mathcal{A}.
\end{array}$$

Here s_i and d_i denote the supply and demand at node i and $f_{i,j}$ the amount of energy sent along the transmission line $(i,j) \in \mathcal{A}$. Note that if $u_{i,j} = 0$, this reduces to a situation where, at each node $i \in \mathcal{N}$, the market is cleared as in centralized markets. When transmission is possible, it may be beneficial for some market participants to trade across the network.

Mathematical finance

Consider a (centralized) securities market where trading occurs over several points $t = 0, \dots, T$ in time. At each time t, we can buy a *portfolio* $x_t \in \mathbb{R}^d$ of d different securities. Again, negative quantities are interpreted as sales. Buying a portfolio $x \in \mathbb{R}^d$ at time t costs $S_t(x)$ units of cash. The uncertainty about the future development of the markets will be modeled by allowing the cost functions S_t be random.

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t)_{t=0}^T$. The market is described by a sequence $(S_t)_{t=0}^T$ of real-valued functions on $\mathbb{R}^d \times \Omega$ such that

1. for every $\omega \in \Omega$, the function $x \mapsto S_t(x, \omega)$ is convex on $\mathbb{R}^d$ and vanishes at the origin,

2. S_t is $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{F}_t$-measurable.

The interpretation is that buying a portfolio $x_t \in \mathbb{R}^d$ at time t and state ω costs $S_t(x_t, \omega)$ units of cash. The measurability property implies that if the portfolio x_t is $\mathcal{F}_t$-measurable then the cost $\omega \mapsto S_t(x_t(\omega), \omega)$ is also $\mathcal{F}_t$-measurable. This just means that the cost is known at the time of purchase.

Many problems in financial risk management come down to *asset-liability management* problems of the form

$$\text{minimize} \quad \sum_{t=0}^{T} \mathcal{V}_t(S_t(\Delta x_t) + c_t) \quad \text{over} \quad x \in \mathcal{N}_D, \qquad \text{(ALM)}$$

where $c = (c_t)_{t=0}^T$ is a sequence of random cash-flows, $x_{-1} := 0$,

$$\mathcal{N}_D := \{x \in \mathcal{N} \mid x_t \in D_t,\ x_T = 0\},$$

the sets D_t describe portfolio constraints and the function $\mathcal{V}_t$ describe the "risk/disutility/regret" from the random expenditure $S_t(\Delta x_t) + c_t$ at time t. Various pricing principles can be derived by analyzing the optimum value $\varphi(c)$ of (ALM) as a function on the space $\mathcal{M} := \{(c_t)_{t=0}^T \mid c_t \in L^0(\Omega, \mathcal{F}_t, P)\}$ of adapted sequences of payments. The optimal value can be interpreted as the least risk an agent with liabilities c can achieve by optimally trading in financial markets. This is a fundamental problem in the basic operations of any financial institution.

For example, the least initial capital that would allow the agent to achieve an "acceptable" risk-return profile with liabilities c is given by

$$\pi^0(c) = \inf\{\alpha \in \mathbb{R} \mid \varphi(c - \alpha p_0) \le 0\},$$

where $p = (1, 0, \dots, 0)$. Here "acceptability" is defined as having nonpositive cumulative "risk". Similarly, the least multiple (*swap rate*) of a sequence $p \in \mathcal{M}$ an agent with initial liabilities $\bar{c} \in \mathcal{M}$ would accept in exchange for taking on an additional liability $c \in \mathcal{M}$ is given by

$$\pi(c) = \inf\{\alpha \in \mathbb{R} \mid \varphi(\bar{c} + c - \alpha p) \le \varphi(\bar{c})\},$$

Traditionally in mathematical finance, the focus has been on valuation of liabilities of the form $c = (0, \dots, 0, c_T)$, i.e. a random payment at a single date. In reality, where cash cannot be borrowed quite freely, the timing of payments is critical. Indeed, many financial contracts in practice involve several payments in time. Examples include bonds, interest rate swaps, insurance contracts and dividend paying stocks. The study of such products in realistic models of financial markets goes beyond the scope of traditional stochastic analysis but can be treated quite easily by combining stochastics with some convex analysis.

Optimal control and calculus of variations

Problems of *optimal control* can often be written in the form

$$\begin{aligned} &\text{minimize} \quad \int_{[0,T]} h_t(x_t, u_t)dt + h_0(x_0) + h_T(x_T) \quad \text{over} \quad x \in AC,\ u \in \mathcal{U} \\ &\text{subject to} \quad \dot{x}_t = g_t(x_t, u_t), \\ &\qquad\qquad\quad u_t \in U_t, \end{aligned}$$

where $AC = \{x \in L^1 \,|\, \dot{x} \in L^1\}$, $\mathcal{U}$ is the space of Lebesgue-measurable functions on $[0,T]$ and $(x,u,t) \mapsto h_t(x,u)$, $(x,u,t) \mapsto g_t(x,u)$, h_0 and h_T are sufficiently regular functions and $t \mapsto U_t$ is a measurable set-valued mapping. Defining

$$f_t(x,v) = \inf_{u \in U_t} \{h_t(x,u) \,|\, v = g_t(x,u)\},$$

we can write the optimal control problem as

$$\text{minimize} \quad \int_{[0,T]} f_t(x_t, \dot{x}_t)dt + h_0(x_0) + h_T(x_T) \quad \text{over} \quad x \in AC.$$

This is a special case of the generalized problem of *Bolza* which generalizes the classical problem in *calculus of variations* [12]. Such problems are central in variational principles e.g. in Lagrangian and Hamiltonian mechanics.

One can generalize the Bolza problem by replacing the domain of integration by an open set Ω in $\mathbb{R}^d$ and by allowing for higher order derivatives. This leads to problems of the form

$$\text{minimize} \quad \int_\Omega f(Dx(\omega), \omega)d\omega \quad \text{over} \quad u \in W^{m,p}(\Omega) \tag{CV}$$

where (we ignore the boundary term for simplicity)

$$W^{m,p}(\Omega) = \{x \in L^p(\Omega) \,|\, \partial^\alpha x \in L^p(\Omega)\ \forall \alpha:\ |\alpha| \le m\}$$

and $D : W^{m,p}(\Omega) \to L^p(\Omega)^n$ is the linear operator given by $Dx = \{\partial^\alpha x\}_{|\alpha| \le m}$. Here n denotes the cardinality of the index set $\{\alpha \in \mathbb{N}^d \,|\, |\alpha| \le m\}$ and ∂^α stands for the distributional derivative of order α. Elliptic partial differential equations often arise in this manner. In particular, when $D = \nabla$ and

$$f(u, \omega) = \frac{1}{2}|u|^2,$$

the above problem corresponds to the classical second order elliptic partial differential equation with Neumann boundary conditions. Indeed, we will see in Section 3.4 that the partial differential equation is obtained from the Fermat's rule for optimality which says that the "derivative" of a function vanishes at its minimum points.

1.2 Extended real-valued functions

This section introduces some basic concepts and notation that will be useful in the analysis of general optimization problems. Particularly useful will be the notion of "essential objective", that incorporates the constraints of a given minimization problem into the objective function. Although it hides much of the structure in a given problem, it is sufficient for many kinds of analysis, including the study of existence of solutions.

Let f_0 be a function on some space X and consider the problem of minimizing f_0 over a subset $C \subset X$. The problem is equivalent to minimizing the extended *real-valued* function $f : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$ defined by

$$f(x) = \begin{cases} f_0(x) & \text{if } x \in C, \\ +\infty & \text{otherwise,} \end{cases}$$

over all of X. The function f is called the *essential objective* of the original problem. Extended real-valued functions arise also via max and min operations. Consider, for example, the function

$$f(x) = \sup_{y \in Y} l(x, y),$$

where l is a function on a product space $X \times Y$. Even if l is real-valued at every point of $X \times Y$, the value of f may be infinite for some $x \in X$.

Identifying a set C by its *indicator function*

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

many properties of extended real-valued functions can be translated into properties of sets. Conversely, an extended real-valued function can be studied by examining its *epigraph*

$$\operatorname{epi} f := \{(x, \alpha) \in X \times \mathbb{R} \mid f(x) \le \alpha\},$$

which is a set in $X \times \mathbb{R}$. A function can also be characterized through its *level sets*

$$\operatorname{lev}_\alpha f := \{x \in X \mid f(x) \le \alpha\},$$

where $\alpha \in \mathbb{R}$ is fixed. In applications, the set of feasible points of an optimization problem is often described in terms of level sets of functions.

Example 1.1 (Nonlinear programming)**.** *The essential objective of the optimization model* (NLP) *can be expressed as*

$$f = f_0 + \sum_{j=1}^{m} \delta_{\mathrm{lev}_0 f_j}.$$

A function f is *proper* if $f(x) < \infty$ for at least one $x \in X$ and $f(x) > -\infty$ for all $x \in X$. Only proper functions make sensible objectives for optimization problems. Minimizing a proper function f over X is equivalent to minimizing a real-valued function over the set

$$\operatorname{dom} f := \{x \in X \mid f(x) < \infty\},$$

called the *effective domain* of f. The set of solutions is denoted by

$$\operatorname{argmin} f := \{x \in X \mid f(x) \leq \inf f\},$$

where $\inf f := \inf_{x \in X} f(x)$ is the *optimal value* of the problem. It may happen that $\operatorname{argmin} f = \emptyset$ but the set

$$\varepsilon\text{-}\operatorname{argmin} f := \{x \in X \mid f(x) \leq \inf f + \varepsilon\}$$

of *ε-optimal solutions* is nonempty whenever $\varepsilon > 0$ and $\inf f > -\infty$. Note that $\operatorname{argmin} f$ and ε-$\operatorname{argmin} f$ are level sets of f.

The above concepts are related to minimization problems but one could define corresponding concepts for maximization problems in an obvious way. That is unnecessary, however, since one can always revert maximization problems into minimization problems simply by minimizing the function $-f$ instead of maximizing f.

The following interchange rule is obvious but often useful.

Proposition 1.2. *Let f be an extended real-valued function on a product space $X \times U$ and define $\varphi(u) = \inf_x f(x,u)$ and $\psi(x) = \inf_u f(x,u)$. Then*

$$\inf f = \inf \varphi = \inf \psi$$

and

$$\begin{aligned} \operatorname{argmin} f &= \{(x,u) \mid u \in \operatorname{argmin} \varphi,\ x \in \operatorname{argmin} f(\cdot,u)\} \\ &= \{(x,u) \mid x \in \operatorname{argmin} \psi,\ u \in \operatorname{argmin} f(x,\cdot)\}. \end{aligned}$$

Example 1.3 (Decentralized markets)**.** *Consider the model of a decentralized market in Section 1.1 and define*

$$S_i(r) = \inf\{C_i(s) - R_i(d) \mid s, d \geq 0,\ s - d = r\}.$$

The market clearing problem can then be written in the reduced form

$$\begin{aligned} &\text{minimize} && \sum_{i \in \mathcal{N}} S_i \left(\sum_{\{j \mid (i,j) \in \mathcal{A}\}} f_{i,j} - \sum_{\{j \mid (j,i) \in \mathcal{A}\}} a_{j,i} f_{j,i} \right) \quad \text{over} \quad f \in \mathbb{R}^{\mathcal{A}} \\ &\text{subject to} && f_{i,j} \in [0, u_{i,j}] \quad \forall (i,j) \in \mathcal{A}. \end{aligned}$$

Example 1.4 (Mathematical finance). *The function π^0 in the financial model of Section 1.1 can be expressed as*

$$\pi^0(c) = \inf\{\alpha \,|\, c - \alpha p_0 \in \operatorname{lev}_0 \varphi\}.$$

Example 1.5 (Optimal control). *The reduction in the optimal control model sketched in Section 1.1 is obtained with*

$$\begin{aligned}\psi(x) &= \inf_{u \in \mathcal{U}} \left\{ \int_{[0.T]} h_t(x_t, u_t)dt + h_0(x_0) + h_T(x_T) \,\middle|\, \dot{x}_t = g_t(x_t, u_t),\ u_t \in U_t \right\} \\ &= \inf_{u \in \mathcal{U}} \left\{ \int_{[0.T]} h_t(x_t, u_t)dt \,\middle|\, \dot{x}_t = g_t(x_t, u_t),\ u_t \in U_t \right\} + h_0(x_0) + h_T(x_T)\end{aligned}$$

and using the interchange rule for minimization and integration; see [16, Theorem 14.60] for details.

While the order of minimization can be interchanged according to Proposition 1.2, the inequality in

$$\inf_x \sup_u f(x,u) \geq \sup_u \inf_x f(x,u)$$

is strict in general. If equality holds, the common value is known as the *saddle-value* of f. The existence of a saddle-value and of the points that attain it are central in the duality theory of convex optimization. That will be the topic of Chapter 3.

1.3 Convexity

In order to talk about convexity, one needs a vector space. A *real vector space* (or a real linear space) is a nonempty set X equipped with operations of *addition* and *scalar multiplication* and containing a *zero element* 0 (or the *origin*) such that for every $x, x_i \in X$ and $\alpha, \alpha_i \in \mathbb{R}$

$$\begin{aligned} x_1 + (x_2 + x_3) &= (x_1 + x_2) + x_3, \\ x_1 + x_2 &= x_2 + x_1, \\ \alpha(x_1 + x_2) &= \alpha x_1 + \alpha x_2, \\ (\alpha_1 + \alpha_2)x &= \alpha_1 x + \alpha_2 x, \\ (\alpha_1 \alpha_2)x &= \alpha_1(\alpha_2 x), \\ 1x &= x, \\ x + 0 &= x \end{aligned}$$

and such that for each $x \in X$ there is a unique element, denoted by $-x$ such that $x + (-x) = 0$. The *Euclidean space* $\mathbb{R}^n$ with componentwise addition and scalar multiplication is a vector space.

Example 1.6. *The function spaces $L^0(\Omega, F, P; \mathbb{R}^n)$, $\mathcal{N}$, $\mathcal{U}$, $W^{m,p}(\Omega)$ and AC in Section 1.1 are vector spaces if we define addition and scalar multiplication in the usual pointwise sense.*

A subset C of a vector space X is *convex* if $\alpha_1 x_1 + \alpha_2 x_2 \in C$ for every $x_i \in C$ and $\alpha_i > 0$ such that $\alpha_1 + \alpha_2 = 1$. An extended real-valued function f on a vector space X is *convex* if its epigraph is a convex subset of the vector space $X \times \mathbb{R}$. It is not hard to check that a function f is convex iff

$$f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$$

whenever $x_1, x_2 \in \operatorname{dom} f$ and $\alpha_1, \alpha_2 > 0$ are such that $\alpha_1 + \alpha_2 = 1$. A function f is *concave* if $-f$ is convex.

Exercise 1.7. *Let C be convex and $\alpha_i \geq 0$. Show that $(\alpha_1 + \alpha_2)C = \alpha_1 C + \alpha_2 C$.*

Convexity of a function f implies that $\operatorname{dom} f$ is convex. If a function is convex, then all its level sets are convex (the converse is not true in general). In particular, the set of solutions or ε-optimal solutions of the problem of minimizing a convex function is convex. A set is convex iff its indicator function is convex. A function is *strictly convex* if the above inequality holds as a strict inequality whenever $x_1 \neq x_2$. If f is strictly convex, then $\operatorname{argmin} f$ consists of at most one point.

A set C is a *cone* if $\alpha x \in C$ for every $x \in C$ and $\alpha > 0$. A function f is *positively homogeneous* if $f(\alpha x) = \alpha f(x)$ for every $x \in \operatorname{dom} f$ and $\alpha > 0$, or equivalently, if $\operatorname{epi} f$ is a cone. A function is *sublinear* if it is both convex and positively homogeneous, or equivalently, if its epigraph is a convex cone.

Proposition 1.8. *A set C is a convex cone if and only if $\alpha_1 x_1 + \alpha_2 x_2 \in C$ for every $x_i \in C$ and $\alpha_i > 0$. A function f is sublinear if and only if*

$$f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$$

for every $x_i \in \operatorname{dom} f$ and $\alpha_i > 0$. A positively homogeneous function is convex if and only if

$$f(x_1 + x_2) \leq f(x_1) + f(x_2)$$

for every $x_i \in \operatorname{dom} f$.

Proof. Exercise □

A *seminorm* is a real-valued sublinear function f which is *symmetric* in the sense that $f(-x) = f(x)$ for every x. A seminorm f is called a *norm* if $f(x) = 0$ only when $x = 0$.

Exercise 1.9. *Show that a seminorm is nonnegative and vanishes at the origin.*

A set C is *affine* if $\alpha_1 x_1 + \alpha_2 x_2 \in C$ for every $x_i \in C$ and $\alpha_i \in \mathbb{R}$ such that $\alpha_1 + \alpha_2 = 1$. A real-valued function f is *affine* if it is both convex and concave, or equivalently, if its graph is affine, or equivalently, if there is a linear function l and a scalar α such that $f(x) = l(x) + \alpha$ for every $x \in X$.

Even in infinite-dimensional applications, convexity can often be traced back to the convexity of functions on the real line.

Lemma 1.10. *Let $\phi : \mathbb{R} \to \overline{\mathbb{R}}$ be nondecreasing and $a \in \mathbb{R}$ such that $\phi(a)$ is finite. Then the function*

$$f(x) = \int_a^x \phi(t)dt$$

is convex. If ϕ is strictly increasing, f is strictly convex.

Proof. Let $x_i \subseteq \operatorname{dom} f$ such that $x_1 < x_2$ and $\alpha_i > 0$ such that $\alpha_1 + \alpha_2 = 1$. Denoting $\bar{x} = \alpha_1 x_2 + \alpha_2 x_2$, we have

$$f(\bar{x}) - f(x_1) = \int_{x_1}^{\bar{x}} \phi(t)dt \leq \phi(\bar{x})(\bar{x} - x_1)$$

and

$$f(x_2) - f(\bar{x}) = \int_{\bar{x}}^{x_2} \phi(t)dt \geq \phi(\bar{x})(x_2 - \bar{x}).$$

Thus,

$$\begin{aligned} f(\bar{x}) &\leq \alpha_1[f(x_1) + \phi(\bar{x})(\bar{x} - x_1)] + \alpha_1[f(x_2) - \phi(\bar{x})(x_2 - \bar{x})] \\ &= \alpha_1 f(x_1) + \alpha_1 f(x_2), \end{aligned}$$

which proves convexity. If ϕ is strictly increasing, then all the above inequalities are strict and we get strict convexity. □

Exercise 1.11. *Consider the problem of price formation in Section 1.1. Show that the cost and revenue functions C and R are convex and concave, respectively.*

Lemma 1.12. *A real-valued function f on an interval $I \subset \mathbb{R}$ is convex iff*

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(x)}{z - x} \leq \frac{f(z) - f(y)}{z - y}$$

whenever $x < y < z$ in I.

Proof. ?? □

Combined with the fundamental theorem of calculus, Lemma 1.10 gives a derivative test for convexity.

Theorem 1.13. *A proper function f on $\mathbb{R}$ is convex if and only if $\operatorname{dom} f$ is an interval and f is absolutely continuous with nondecreasing f' on compact subintervals of $\operatorname{dom} f$.*

Proof. By the fundamental theorem of calculus, an absolutely continuous function f on an interval $[a, b]$ is differentiable almost everywhere (with respect to the Lebesgue measure) and

$$f(x) - f(a) = \int_a^x f'(t)dt$$

for every $x \in [a, b]$. The first claim thus follows from Lemma 1.10.

Conversely, if f is convex, Lemma 1.12 implies that f is Lipschitz continuous, and thus absolutely continuous, on every compact subinterval of $\operatorname{dom} f$. The monotonicity of the difference quotient in Lemma 1.12 also implies that f' is nondecreasing. $\square$

Example 1.14. *The convexity/concavity of the following elementary functions on the real line is easily verified with the derivative test.*

1. $f(x) = x$ *is both convex and concave on* $\mathbb{R}$*,*
2. $f(x) = e^x$ *is convex on* $\mathbb{R}$*,*
3. $f(x) = \ln x$ *is concave on* $\mathbb{R}_{++}$*,*
4. $f(x) = x \ln x$ *is convex on* $\mathbb{R}_{++}$*,*
5. $f(x) = x^p$ *is concave on* $\mathbb{R}_{++}$ *if* $p \in (0, 1]$*. It is convex if* $p \notin [0, 1)$*. Note that* f *is increasing when* $p > 0$ *and decreasing when* $p < 0$*,*

More examples are obtained by convexity-preserving algebraic operations from the above; see Section 1.4 below.

The derivative criterion in Theorem 1.13 yields derivative criteria also in higher dimensions.

Corollary 1.15 (Hessians)**.** *A twice continuously differentiable function* f *on an open subset* $C \subset \mathbb{R}^n$ *is convex if and only if its Hessian* $\nabla^2 f(x)$ *is positive semidefinite for all* $x \in C$*.*

Proof. The function f is convex if and only if for every $x, z \in \mathbb{R}^n$, the function $g(\alpha) := f(x + \alpha z)$ is convex on $\{\alpha \,|\, x + \alpha z \in C\}$. The first derivative of g is given by $g'(\alpha) = \nabla f(x + \alpha z) \cdot z$. This is increasing if and only if the second derivative

$$g''(\alpha) = z \cdot \nabla^2 f(x + \alpha z) z$$

is positive. This holds for all $x, z \in \mathbb{R}^n$ if and only if the Hessian is positive semidefinite on all of C. $\square$

Example 1.16 (Quadratics and ellipsoids)**.** *A quadratic function* $f(x) = x \cdot Hx$ *is convex if and only if* H *is positive semidefinite. In particular, if* H *is a positive semidefinite matrix, the corresponding* ellipsoid

$$\{x \in \mathbb{R}^n \,|\, x \cdot Hx \leq 1\}$$

is a convex set.

The following provides an important source of convex functions on function spaces and, in particular, Euclidean spaces. Throughout these notes, the *positive* and *negative parts* of an $\alpha \in \overline{\mathbb{R}}$ are denoted by $\alpha^+ := \max\{\alpha, 0\}$ and $\alpha^- := [-\alpha]^+$, respectively.

Example 1.17 (Convex integral functionals)**.** *Let f be an extended real-valued $\mathcal{B}(\mathbb{R}^n)\otimes\mathcal{F}$-measurable function on $\mathbb{R}^n\times\Omega$ such that $f(\cdot,\omega)$ is proper and convex for every $\omega\in\Omega$. Then the function $I_f : L^0(\Omega,\mathcal{F},\mu;\mathbb{R}^n)\to\overline{\mathbb{R}}$ defined by*

$$I_f(x) = \begin{cases} \int_\Omega f(x(\omega),\omega)d\mu(\omega) & \textit{if } f(x(\cdot),\cdot)^+\in L^1,\\ +\infty & \textit{otherwise}\end{cases}$$

is convex. The function I_f is the integral functional *associated with f; see [10, 14] or [16, Chapter 14].*

If f is of the form $f(x,\omega)=\delta_{D(\omega)}(x)$ for a convex valued mapping $\omega\mapsto D(\omega)\subset\mathbb{R}^n$, then I_f is the indicator function of the set

$$\mathcal{D} := \{x\in L^0(\Omega,\mathcal{F},\mu)\,|\,x\in D \ \mu\text{-a.e.}\}$$

of measurable selectors *of D. If D is convex-valued, then $f(\cdot,\omega)$ is convex so that I_f is convex function and thus, $\mathcal{D}$ is a convex set.*

Proof. Let $x_i\in\operatorname{dom} I_f$ so that $f(x_i(\cdot),\cdot)^+\in L^1$. Then, for every $\alpha_i>0$ such that $\alpha_1+\alpha_2=1$, the pointwise convexity of f gives

$$f(\alpha_1x_1(\omega)+\alpha_2x_2(\omega),\omega)\le\alpha_1f(x_1(\omega),\omega)+\alpha_1f(x_1(\omega),\omega)$$

so $\alpha_1x_1+\alpha_2x_2\in\operatorname{dom} I_f$ and

$$\begin{aligned} I_f(\alpha_1x_1+\alpha_2x_2) &\le \int_\Omega[\alpha_1f(x_1(\omega),\omega)+\alpha_1f(x_1(\omega),\omega)]d\mu(\omega)\\ &= \int_\Omega \alpha_1f(x_1(\omega),\omega)d\mu(\omega)+\int_\Omega\alpha_1f(x_1(\omega),\omega)d\mu(\omega)\\ &= \alpha_1\int_\Omega f(x_1(\omega),\omega)d\mu(\omega)+\alpha_1\int_\Omega f(x_1(\omega),\omega)d\mu(\omega)\\ &= \alpha_1I_f(x_1)+\alpha_2I_f(x_2),\end{aligned}$$

where the first equality holds since the positive parts of the two integrands are in L^1. □

The *convex hull* of a set C is the intersection of all convex sets that contain C. It is denoted by $\operatorname{co} C$. The *convex hull* of a function f is the pointwise supremum of all convex functions dominated by f. It is denoted by $\operatorname{co} f$.

Lemma 1.18. *We have* $\inf\operatorname{co} f=\inf f$, $\operatorname{argmin}\operatorname{co} f\supseteq\operatorname{co}\operatorname{argmin} f$ *and* $(\operatorname{co} f)(x)=\inf\{\alpha\,|\,(x,\alpha)\in\operatorname{co}\operatorname{epi} f\}$*.*

Proof. Clearly, $\operatorname{co} f\le f$. On the other hand, the constant function $x\mapsto\inf f$ is convex and dominated by f, so that $(\operatorname{co} f)(x)\ge\inf f$ for every x. This proves the first clam and also shows that $\operatorname{argmin} f\subseteq\operatorname{argmin}\operatorname{co} f$. Since the latter set is convex, we have $\operatorname{co}\operatorname{argmin} f\subseteq\operatorname{argmin}\operatorname{co} f$. If $g\le f$ is convex, then $\operatorname{co}\operatorname{epi} f\subseteq\operatorname{epi} g$, so that

$$g(x)=\inf\{\alpha\,|\,(x,\alpha)\in\operatorname{epi} g\}\le\inf\{\alpha\,|\,(x,\alpha)\in\operatorname{co}\operatorname{epi} f\},$$

where the last expression is a convex function of x by Lemma **??**. □

Exercise 1.19. *Show that*

$$\operatorname{co} C = \left\{ \sum_{i=1}^n \alpha_i x_i \,\middle|\, n \in \mathbb{N},\ \alpha_i \geq 0,\ \sum_{i=1}^n \alpha_i = 1,\ x_i \in C \right\}$$

and

$$(\operatorname{co} f)(x) = \inf \left\{ \sum_{i=1}^n \alpha_i f(x_i) \,\middle|\, n \in \mathbb{N},\ \alpha_i \geq 0,\ x_i \in \operatorname{dom} f,\ \sum_{i=1}^n \alpha_i = 1,\ x = \sum_{i=1}^n \alpha_i x_i \right\}.$$

1.4 Convexity in algebraic operations

One of the reasons why convexity appears so frequently in applications is that it is well preserved under various algebraic operations. One often encounters functions that have been constructed through various operations from some elementary functions on the real line. Convexity of such a function can often be verified by first inspecting the derivatives of the elementary functions (see Theorem 1.13) and then checking that the involved operations preserve convexity. This section is devoted to convexity preserving algebraic operations.

Theorem 1.20. *Let X, X_i and U be vector spaces.*

1. *If $C_1, C_2 \subset X$ are convex, then $C_1 + C_2 := \{x_1 + x_2 \,|\, x_i \in C_i\}$ is convex.*
2. *Given any collection $\{C_j\}_{j \in J}$ of convex sets in X, their intersection $\cap_{j \in J} C_j$ is convex.*
3. *If $C_i \subset X_i$ are convex, then $\prod_i C_i$ is convex.*
4. *If $A : X \to U$ is linear and $C \subset X$ is convex, then $AC := \{Ax \,|\, x \in C\}$ is convex.*
5. *If $A : X \to U$ is linear and $C \subset U$ is convex, then $A^{-1}C = \{x \in X \,|\, Ax \in C\}$ is convex.*

All the above hold with "convex" replaced by "cone" or "affine".

Proof. Exercise □

Unions of convex sets are not convex in general. However, if $C \subset X$ is convex, then its *positive hull*

$$\operatorname{pos} C = \bigcup_{\alpha > 0} \alpha C$$

is convex. It is the smallest cone containing C.

Exercise 1.21. *Let A be a totally ordered set and $(C_\alpha)_{\alpha \in A}$ a nondecreasing family of convex sets, then $\bigcup_{\alpha \in A} C_\alpha$ is convex.*

We now turn to operations on convex functions. Intersecting the epigraphs of convex functions gives the following.

Theorem 1.22 (Pointwise supremum)**.** *The pointwise supremum of an arbitrary collection of convex functions is convex.*

Example 1.23. *Here are some useful constructions in Euclidean spaces:*

1. *The function $x \mapsto |x|$ is convex on $\mathbb{R}$.*
2. *The function*
$$f(x) = \max_{i=1,\dots,n} x_i$$
is convex on $\mathbb{R}^n$.
3. *Given a set $C \subset \mathbb{R}^n$, the function*
$$\sigma_C(x) = \sup_{v \in C} x \cdot v$$
is convex on $\mathbb{R}^n$.
4. *For any function g on $\mathbb{R}^n$, the function*
$$f(x) = \sup\{x \cdot v - g(v)\}$$
is convex on $\mathbb{R}^n$.

Note that 3 is a special case of 4 with $g = \delta_C$ while 2 is a special case of 3 with $C = \{v \in \mathbb{R}^n_+ \mid \sum_{i=1}^n x_i = 1\}$. In 1, one simply takes the supremum of the functions, $x \mapsto x$ and $x \mapsto -x$. We will see later that 2-4 can be extended to general locally convex vector spaces.

Let X and U be vector spaces and let $K \subseteq U$ be a convex cone containing the origin. A function F from a subset $\operatorname{dom} F$ of X to U is K*-convex* if

$$\operatorname{epi}_K F = \{(x, u) \mid x \in \operatorname{dom} F,\ F(x) - u \in K\}$$

is a convex set in $X \times U$. The domain $\operatorname{dom} F$ of a K-convex function is convex, by Theorem 1.20(c), since it is the projection of the convex set $\operatorname{epi}_K F$ to X.

Lemma 1.24. *A function F from X to U is K-convex if and only if* $\operatorname{dom} F$ *is convex and*
$$F(\alpha_1 x_1 + \alpha_2 x_2) - \alpha_1 F(x_1) - \alpha_2 F(x_2) \in K$$
for every $x_i \in \operatorname{dom} F$ and $\alpha_i > 0$ such that $\alpha_1 + \alpha_2 = 1$.

Proof. Since K is a convex cone containing the origin, the condition in the lemma is equivalent, by Proposition 1.8, to the seemingly stronger condition that
$$F(\alpha_1 x_1 + \alpha_2 x_2) - \alpha_1 (F(x_1) - u_1) - \alpha_2 (F(x_2) - u_2) \in K$$

for every $x_i \in \operatorname{dom} F$, $u_i \in K$ and $\alpha_i > 0$ such that $\alpha_1 + \alpha_2 = 1$. Since $x_i \in \operatorname{dom} F$ and $u_i \in K$ is equivalent to $(x_i, F(x_i) - u_i) \in \operatorname{epi}_K F$, the condition can be written as

$$F(\alpha_1 x_1 + \alpha_2 x_2) - \alpha_1 v_1 - \alpha_2 v_2 \in K$$

for every $(x_i, v_i) \in \operatorname{epi}_K F$ and $\alpha_i > 0$ such that $\alpha_1 + \alpha_2 = 1$. This means that $\operatorname{epi}_K F$ is convex. □

If h is an extended real-valued function on X, the *composition* $h \circ F$ of F with h is defined by

$$\operatorname{dom}(h \circ F) = \{x \in \operatorname{dom} F \mid F(x) \in \operatorname{dom} h\},$$
$$(h \circ F)(x) = h(F(x)) \quad \forall x \in \operatorname{dom}(h \circ F).$$

The *range* of F will be denoted by $\operatorname{rge} F$.

Theorem 1.25 (Composition)**.** *If F is a K-convex function from X to U and h is a convex function on U such that*

$$u_1 \in \operatorname{rge} F,\ u_1 - u_2 \in K \implies h(u_1) \leq h(u_2)$$

then $h \circ F$ is convex.

Proof. Let $x_i \in \operatorname{dom}(h \circ F)$ and $\alpha_i > 0$ such that $\alpha_1 + \alpha_2 = 1$. Then,

$$h(F(\alpha_1 x_1 + \alpha_2 x_2)) \leq h(\alpha_1 F(x_1) + \alpha_2 F(x_2)) \leq \alpha_1 h(F(x_1)) + \alpha_2 h(F(x_2)),$$

where the first inequality comes from the K-convexity of F, Lemma 1.24 and the growth property of h. The second comes from the convexity of h. □

A function h certainly satisfies the growth condition in Theorem 1.25 if $h(u_1) \leq h(u_2)$ whenever $u_1 - u_2 \in K$. Imposing the growth condition only on the range of F is often essential.

Example 1.26. *The following functions are convex:*

1. *$f(x) = h(x)^p$, where $p \geq 1$ and h is convex and nonnegative,*
2. *$f(x) = e^{h(x)}$, where h is convex,*
3. *$f(x) = |x|^p$, where $p \geq 1$,*
4. *$f(x) = x^x$ on $\mathbb{R}$ (this follows from 2 and Example 1.14.4 by writing $f(x) = e^{x \ln x}$).*

Example 1.27 (Composite model)**.** *Theorem 1.25 can be directly applied to the composite model* (CO) *of Section 1.1. We do not need to assume that the functions have full domains so we can incorporate various constraints in the definition of the functions F and h. If all the $m+1$ components of the function $F = (f_0, \dots, f_m)$ are proper and convex, then F is an $\mathbb{R}^{m+1}_-$-convex function* *with* $\operatorname{dom} F = \bigcap_{j=0}^m \operatorname{dom} f_j$*. Thus, for the composite model to be convex, it suffices that* h *be convex and* $h(u_1) \le h(u_2)$ *whenever* $u_1 \in \operatorname{rge} F$ *and* $u_1 \le u_2$ *componentwise. This is certainly satisfied by*

$$h(u) = \begin{cases} u_0 & \text{if } u_j \le 0 \text{ for } j = 1, \dots, m, \\ +\infty & \text{otherwise,} \end{cases}$$

which corresponds to the classical nonlinear programming model (NLP) *with*

$$(h \circ F)(x) = \begin{cases} f_0(x) & \text{if } f_j(x) \le 0 \text{ for } j = 1, \dots, m, \\ +\infty & \text{otherwise.} \end{cases}$$

Example 1.28 (Stochastic programming)**.** *Consider the stochastic optimization model* (CSP) *in the generalized formulation where* f *is allowed to be extended real-valued. Let* $\operatorname{dom} F = \{x \in \mathbb{R}^n \mid f(x, \cdot) \in L^0(\Omega, \mathcal{F}, P)\}$*,* $F(x) = f(x, \cdot)$ *for* $x \in \operatorname{dom} F$ *and*

$$K = \{u \in L^0(\Omega, \mathcal{F}, P) \mid u \le 0 \ P\text{-a.s.}\}.$$

If $f(\cdot, \omega)$ *is a proper convex function for every* ω*, then* F *is a* K*-convex function from* $\mathbb{R}^n$ *to* $L^0(\Omega, \mathcal{F}, P)$*. Indeed, if* $x_i \in \operatorname{dom} F$ *and* $\alpha > 0$ *are such that* $\alpha_1 + \alpha_2 = 1$*, then* $f(\alpha_1 x_1 + \alpha_2 x_2, \omega) \le \alpha_1 f(x_1, \omega) + \alpha_2 f(x_2, \omega)$ *so that* $\alpha_1 x_1 + \alpha_2 x_2 \in \operatorname{dom} F$ *and* $F(\alpha_1 x_1 + \alpha_2 x_2) - \alpha_1 F(x_1) - \alpha_2 F(x_2) \in K$*.*

If $\mathcal{V}$ *is a convex function on* $L^0(\Omega, \mathcal{F}, P)$ *such that* $\mathcal{V}(u_1) \le \mathcal{V}(u_2)$ *whenever* $u_1 \le u_2$*, then by Theorem 1.25, the composition* $\mathcal{V} \circ f$ *is a convex function on* $\mathbb{R}^n$*. In the special case where*

$$f(x, \omega) = \begin{cases} f_0(x, \omega) & \text{if } f_j(x, \omega) \le 0 \text{ for } j = 1, \dots, m, \\ +\infty & \text{otherwise,} \end{cases}$$

with $f_j(x, \cdot) \in L^0(\Omega, \mathcal{F}, P)$ *for every* $j = 0, \dots, m$ *and* $x \in \mathbb{R}^n$*, we have* $\operatorname{dom} F = \{x \in \mathbb{R}^n \mid f_j(x, \cdot) \le 0 \ P\text{-a.s. } j = 1, \dots, m\}$ *so that* (CSP) *can be written with pointwise constraints as*

$$\begin{array}{llll} \text{minimize} & \mathcal{V}(f_0(x, \cdot)) & \text{over} & x \in \mathbb{R}^n \\ \text{subject to} & f_j(x, \cdot) \le 0, & P\text{-a.s. } j = 1, \dots, m & \end{array}$$

much like in the traditional nonlinear programming model.

Example 1.29 (Convex integral functionals)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and define* F *like in the previous example. The function* $I : L^0(\Omega, \mathcal{F}, P) \to \overline{\mathbb{R}}$ *defined by*

$$I(u) = \begin{cases} \int_\Omega u(\omega) d\mu(\omega) & \text{if } u^+ \in L^1, \\ +\infty & \text{otherwise.} \end{cases}$$

satisfies the growth condition in Theorem 1.25 so the function $I \circ F$ *is convex on* $L^0(\Omega, \mathcal{F}, \mu; \mathbb{R}^n)$*. Since* $I \circ F = I_f$*, we obtain an alternative proof of the convexity of an integral functional defined in Example 1.17.*

Corollary 1.30. *Let X and U be vector spaces.*

(a) If A is a linear mapping from a linear subspace $\operatorname{dom} A$ *of X to U and h is a convex function on U, then $h \circ A$ is convex on X.*

(b) If f_j are proper and convex then $f_1 + \cdots + f_m$ is convex.

Proof. In part (a), let $F = A$ and $K = \{0\}$ and in (b), $U = \mathbb{R}^m$, $F(x) = (f_1(x), \ldots, f_m(x))$, $K = \mathbb{R}^m_-$ and $h(u) = u_1 + \cdots + u_m$. □

The following gives another important class of convexity preserving operations.

Theorem 1.31 (Infimal projection)**.** *If f is a convex function on the product $X \times U$ of two vector spaces, then $\varphi(u) := \inf_{x \in X} f(x, u)$ is a convex function on U. If f is sublinear, then φ is sublinear as well.*

Proof. We have $\operatorname{dom} \varphi = \{u \in U \mid \exists x \in X : \; f(x, u) < \infty\}$. Given $u_1, u_2 \in \operatorname{dom} \varphi$ and $\alpha_1, \alpha_2 > 0$ such that $\alpha_1 + \alpha_2 = 1$, the convexity of f gives

$$\begin{aligned}
\alpha_1 \varphi(u_1) + \alpha_2 \varphi(u_2) &= \alpha_1 \inf_{x_1 \in X} f(x_1, u_1) + \alpha_2 \inf_{x_2 \in X} f(x_2, u_2) \\
&= \inf_{x_1, x_2 \in X} \{\alpha_1 f(x_1, u_1) + \alpha_2 f(x_2, u_2)\} \\
&\geq \inf_{x_1, x_2 \in X} f(\alpha_1 x_1 + \alpha_2 x_2, \alpha_1 u_1 + \alpha_2 u_2) \\
&= \varphi(\alpha_1 u_1 + \alpha_2 u_2).
\end{aligned}$$

Sublinearity is proved similarly. □

The function φ in Theorem 1.31 is called the *inf-projection* of f. It may be viewed as the *value function* of a parametric optimization problem. The value function plays a key role in the duality theory of convex optimization to be studied in Chapter 3.

Corollary 1.32. *Let X and U be vector spaces.*

(a) If F is a K-convex function from X to U and f is a convex function on X, then the function

$$(Ff)(u) = \inf\{f(x) \mid x \in \operatorname{dom} F, \; F(x) - u \in K\}$$

is convex on U.

(b) If $A : X \to U$ is linear and k is a convex function on X, then the function

$$(Af)(u) = \inf_x \{f(x) \mid Ax = u\}$$

is convex on U.

(c) If f_1 and f_2 are convex functions on U, then the function

$$(f_1 \,\square\, f_2)(u) = \inf_{u_1,u_2} \{f_1(u_1) + f_2(u_2) \,|\, u_1 + u_2 = u\}$$

is convex.

Proof. In (a), apply Theorem 1.31 to the function $(x, u) \mapsto f(x) + \delta_{\mathrm{epi}_K F}(x, u)$. In (b), we can take $K = \{0\}$. In part (c), let $X = U \times U$, $A(u_1, u_1) = u_1 + u_2$, $f(u_1, u_2) = f_1(u_1) + f_2(u_2)$ and apply part (b). □

The following simple construction will be useful.

Corollary 1.33. *If C is a convex subset of $X \times \mathbb{R}$, then the function*

$$g(u) = \inf\{\alpha \,|\, (u, \alpha) \in C\}$$

is convex. If C is a convex cone, g is sublinear.

Proof. This applies Theorem 1.31 to the function $f(x, \alpha) = \alpha + \delta_C(x, \alpha)$. □

Example 1.34 (Distance function)**.** *Given a convex set C in a normed space X, the associated* distance function

$$d_C(x) := \inf\{\|x - z\| \,|\, z \in C\}$$

is convex.

Proof. This applies Theorem 1.31 to the function $f(x, z) = \|x - z\| + \delta_C(z)$. □

Example 1.35 (Risk measures)**.** *Let $\mathcal{V}$ be a convex function on $L^0(\Omega, \mathcal{F}, P)$ such that $\mathcal{V}(0) = 0$ and $\mathcal{V}(c_1) \le \mathcal{V}(c_2)$ whenever $c_1 \le c_2$ almost surely (i.e. $\mathcal{V}$ in* nondecreasing*). The function $\mathcal{R}(c) := \inf\{w \,|\, \mathcal{V}(c - w) \le 0\}$ is a* convex risk measure *in the sense that it is convex, nondecreasing and $\mathcal{R}(c + \alpha) = \mathcal{R}(c) + \alpha$ for every $\alpha \in \mathbb{R}$.*

Example 1.36. *The function*

$$f(x) = \ln E e^x$$

is a convex on $L^0(\Omega, \mathcal{F}, P)$. In particular, the function

$$f(x) = \ln \sum_{i=1}^{n} e^{x_i}$$

is convex on $\mathbb{R}^n$.

Proof. We have

$$\begin{aligned} f(x) &= \inf\{\alpha \,|\, \ln E e^x \le \alpha\} \\ &= \inf\{\alpha \,|\, E e^x \le e^\alpha\} \\ &= \inf\{\alpha \,|\, E e^{x-\alpha} \le 1\} \\ &= \inf\{\alpha \,|\, (x, \alpha) \in \mathrm{lev}_1\, g\}, \end{aligned}$$

where $g(x, \alpha) := Ee^{x-\alpha}$ is convex by Example 1.17 and Corollary 1.32. Convexity of f thus follows from Corollary 1.33. When $\Omega := \{1, \dots, n\}$, $\mathcal{F}$ is the power set of Ω and P is the uniform measure on Ω, we have $L^0(\Omega, \mathcal{F}, P) = \mathbb{R}^n$ and we can write the expectation as the weighted sum. □

Exercise 1.37 (Exponential family of probability distributions)**.** *In statistics, an* exponential family *is a parameteric set of probability distributions which is often convenient in numerical computations. Given a measure space* $(\Omega, \mathcal{F}, \mu)$ *and a parameter* $\eta \in \mathbb{R}^n$*, the probability measure* P *is defined by*

$$\frac{dP}{d\mu} = e^{\eta \cdot T - A(\eta)},$$

where T *is a* $\mathbb{R}^n$*-valued random variable (known as "sufficient statistic") and* A *is an extended real-valued function on* $\mathbb{R}^n$*. Several familiar parametric distributions such as normal, log-normal, gamma, binomial, Poisson, etc. can be parameterized as exponential families.*

Show that A *is convex and thus that the log-likelihood function given a sample of* N *observations of* T *is concave with respect to* θ *(thus making max-likelihood estimation of* θ *a convex optimization problem). Hint: use Exercise 1.36 and the fact that* $P(\Omega) = 1$*.*

It can be shown that members of the exponential family associated with given $(\Omega, \mathcal{F}, \mu)$ *and* T *are the measures* P *whose* entropy *with respect to* μ *is maximal among all measures under which the expectation of* T *equals a given* $t \in \mathbb{R}^n$*. The proof is a simple application of the duality theory to be developed in Section 3.2 below.*

Exercise 1.38 (Decentralized markets)**.** *Show that the functions* S_i *in Example 1.3 are convex.*

The function $(h_1 \square h_2)$ in the above corollary is known as the *inf-convolution* of h_1 and h_2. It turns out to be dual to the operation of addition under the so called Legendre-Fenchel transform; see Section 3.1.

Exercise 1.39. *Show that*

1. *the infimal projection* φ *of a function* f *can be expressed as*

$$\varphi(u) = \inf\{\alpha \mid \exists x \in X : (x, u, \alpha) \in \operatorname{epi} f\}.$$

Because of this, the inf-projection is sometimes called the epi-projection.

2. *the inf-convolution can be expressed as*

$$(h_1 \square h_2)(u) = \inf\{\alpha \mid (u, \alpha) \in \operatorname{epi} h_1 + \operatorname{epi} h_2\}.$$

Because of this, the inf-convolution is sometimes called the epi-sum*.*

3. *for any* $\alpha > 0$,

$$\lambda f(x/\lambda) = \inf\{\alpha \mid (x, \alpha) \in \lambda \operatorname{epi} f\}.$$

This function is sometimes called the epi-multiplication of f*.*

Since sums, projections and scalar multiples of convex sets are convex (see Theorem 1.20), the above expressions combined with Corollary 1.33, yield quick proofs of convexity of the functional operations.

The following shows that the epi-multiplication is jointly convex in (x, α).

Proposition 1.40. *If f is convex, then*

$$g(x,\alpha) = \begin{cases} \alpha f(x/\alpha) & \text{if } \alpha > 0, \\ +\infty & \text{otherwise} \end{cases}$$

is a sublinear function on $X \times \mathbb{R}$.

Proof. By convexity of f,

$$f\left(\frac{x_1 + x_2}{\alpha_1 + \alpha_2}\right) \le \frac{\alpha_1}{\alpha_1 + \alpha_2} f(x_1/\alpha_1) + \frac{\alpha_2}{\alpha_1 + \alpha_2} f(x_2/\alpha_2)$$

for any $x_i \in X$ and $\alpha_i > 0$. The claim then follows by multiplying by $(\alpha_1 + \alpha_2)$ and using Proposition 1.8. □

Example 1.41 (Gauge). *Given a convex set $C \subset X$, the associated* gauge

$$g_C(x) := \inf\{\alpha > 0 \,|\, z/\alpha \in C\}$$

is convex.

Proof. By Proposition 1.40, the function $f(x, \alpha) = \alpha + \delta_C(z/\alpha)$ is convex. The convexity of g_C thus follows from Theorem 1.31. □

Example 1.42 (Geometric mean). *The function*

$$u(x) = \exp(E \ln x)$$

is concave and positively homogeneous on $L^0(\Omega, \mathcal{F}, P)_+$ (compare with Example 1.36). In particular, given $\alpha_i > 0$ such that $\sum_{i=1}^n \alpha_i = 1$, the geometric mean (aka Cobb–Douglas utility function)

$$u(x) = \prod_{i=1}^n x_i^{\alpha_i}$$

is concave and positively homogeneous on on $\mathbb{R}^n_+$. If $\sum_{i=1}^n \alpha_i < 1$, this function is still concave but not sublinear (use the composition rule).

Proof. We can express u as

$$\begin{aligned} u(x) &= \sup\{\lambda \,|\, \lambda > 0,\ E \ln x \ge \ln \lambda\} \\ &= \sup\{\lambda \,|\, \lambda > 0,\ E \ln(x/\lambda) \ge 0\} \\ &= \sup\{\lambda \,|\, \lambda > 0,\ \lambda E \ln(x/\lambda) \ge 0\} \\ &= \sup\{\lambda \,|\, \lambda > 0,\ \lambda g(x/\lambda) \le 0\}, \end{aligned}$$

where $g(x) = -E \ln x$. Since g is convex the concavity of u follows from Proposition 1.40 and Corollary 1.33. □

Example 1.43 (Generalized mean)**.** *If $p \leq 1$, the function*

$$f(x) = (Ex^p)^{1/p}$$

is concave and positively homogeneous on L^0_+. In particular, the harmonic mean

$$f(x) = \left(\sum_{i=1}^n x_i^{-1}\right)^{-1}$$

is a concave function on $\mathbb{R}^n_+$.

Proof. Assume first that $p < 0$ so that $x \mapsto x^p$ is convex and decreasing. We have

$$\begin{aligned} f(x) &= \sup\{\lambda \mid (Ex^p)^{1/p} \geq \lambda\} \\ &= \sup\{\lambda \mid Ex^p \leq \lambda^p\} \\ &= \sup\{\lambda \mid E(x/\lambda)^p \leq 1\} \\ &= \sup\{\lambda \mid \lambda E(x/\lambda)^p \leq \lambda\} \\ &= \sup\{\lambda \mid g(x,\lambda) \leq 0\}, \end{aligned}$$

where the function

$$g(x,\lambda) = \lambda E(x/\lambda)^p - \lambda$$

is convex by Proposition 1.40. Thus, the concavity of f follows from Corollary 1.33. When $p \in (0,1]$, the function $x \mapsto x^p$ is concave and increasing and the above proof goes through with obvious modifications. □

No that the functions in Examples 1.36, 1.42 and 1.43 are all of the form

$$f(x) = g^{-1}[Eg(x)]$$

for a function g that is strictly increasing or strictly decreasing as well as convex or concave. Functions of this form are known as Kolmogorov means or quasi-arithmetic means.

1.5 Convex sets and functions under scaling

Convex sets and functions have regular behavior under scaling by positive constants. The main purpose of this section is to study the limiting behavior as the scaling parameter approaches zero or infinity. This gives rise to directional derivatives and recession functions, which are involved e.g. in optimality conditions and criteria for existence of solutions.

Lemma 1.44. *If C is a convex set with $0 \in C$ and if $0 < \alpha_1 < \alpha_2$, then $\alpha_1 C \subseteq \alpha_2 C$. If f is a convex function finite at $\bar{x}$, then for every $x \in X$*

$$\frac{f(\bar{x} + \lambda x) - f(\bar{x})}{\lambda}$$

is increasing in $\lambda > 0$.

Proof. The first claim follows from $\alpha_1/\alpha_2 C + (1-\alpha_1/\alpha_2)\{0\} \subseteq C$. Applying the first part to the epigraph of the function $h(x) = f(\bar{x}+x) - f(\bar{x})$ we get that the set $\alpha \operatorname{epi} h$ grows with α. Since $\alpha \operatorname{epi} h$ is the epigraph of the function $x \mapsto \alpha h(x/\alpha)$, this means that

$$\alpha[f(\bar{x}+x/\alpha) - f(\bar{x})]$$

decreases with $\alpha > 0$. It now suffices to make the substitution $\alpha = 1/\lambda$. □

The union

$$\operatorname{pos}(C-x) = \bigcup_{\alpha>0} \alpha(C-x)$$

may be viewed as a *local approximation* of the set C at a point $x \in C$. It is a convex cone containing the origin. Given a convex function f and a point $\bar{x}$ where f is finite, the *directional derivative*

$$f'(\bar{x};x) = \lim_{\lambda \searrow 0} \frac{f(\bar{x}+\lambda x) - f(\bar{x})}{\lambda}$$

gives a local approximation of f at $\bar{x}$. By Lemma 1.44, the directional derivative is well-defined and

$$f'(\bar{x};x) = \inf_{\lambda>0} \frac{f(\bar{x}+\lambda x) - f(\bar{x})}{\lambda}.$$

Exercise 1.45. *Let f be a convex function on $\mathbb{R}$. Show that*

$$f'(\bar{x};x) = \begin{cases} f'_+(\bar{x})x & \text{if } x \ge 0, \\ f'_-(\bar{x})x & \text{if } x \le 0, \end{cases}$$

where f'_+ and f_- denote the right- and left-derivatives of f, respectively.

The following will be useful.

Lemma 1.46. *If f is a convex function finite at $\bar{x}$, then*

$$f'(\bar{x};x) = \inf\{\alpha \mid (x,\alpha) \in \operatorname{pos}(\operatorname{epi} f - (\bar{x}, f(\bar{x})))\}.$$

Proof. The condition $(x,\alpha) \in \operatorname{pos}(\operatorname{epi} f - (\bar{x}, f(\bar{x})))$ can be written as

$$\begin{aligned} &\exists \gamma > 0 : \ (x,\alpha) \in \gamma(\operatorname{epi} f - (\bar{x}, f(\bar{x}))) \\ \iff &\exists \gamma > 0 : \ (\bar{x} + x/\gamma, f(\bar{x}) + \alpha/\gamma) \in \operatorname{epi} f \\ \iff &\exists \gamma > 0 : \ \alpha \ge \gamma[f(\bar{x}+x/\gamma) - f(\bar{x})], \end{aligned}$$

so

$$\begin{aligned} &\inf\{\alpha \mid (x,\alpha) \in \operatorname{pos}(\operatorname{epi} f - (\bar{x}, f(\bar{x})))\} \\ =&\inf\{\alpha \mid \exists \gamma > 0 : \ \alpha \ge \gamma[f(\bar{x}+x/\gamma) - f(\bar{x})]\} \\ =&\inf_{\gamma>0} \gamma[f(\bar{x}+x/\gamma) - f(\bar{x})] \end{aligned}$$

which equals the above expression for $f(\bar{x};x)$. □

Theorem 1.47. *If f is a convex function finite at a point $\bar{x}$, then the directional derivative $f'(\bar{x};\cdot)$ is a well-defined sublinear function. Moreover, f attains its minimum at $\bar{x}$ if and only if $f'(\bar{x};\cdot)$ is a nonnegative function.*

Proof. The first claim follows from Lemma 1.46 and Corollary 1.33. The rest follows from Lemma 1.44 once we observe that an $\bar{x}$ minimizes f if and only if the difference quotient in Lemma 1.44 is nonnegative for every $x \in X$ and $\lambda > 0$. □

The optimality condition in the above theorem can be seen as an abstract version of the classical Fermat's rule which says that the derivative of a differentiable function vanishes at a point where the function attains its minimum.

While $\operatorname{pos}(C - x)$ gives a local approximation of a set C at a point $x \in C$, the set

$$C^\infty = \bigcap_{x\in C}\bigcap_{\alpha>0} \alpha(C - x)$$

describes the shape of C infinitely far from the origin. The set C^∞ is called the *recession cone* of C. It is a convex cone containing the origin. It can also be expressed as

$$C^\infty = \{y \,|\, x + \alpha y \in C \;\forall x \in C \;\forall \alpha > 0\}.$$

The recession cone thus gives the directions in which a set is "unbounded".

We will say that a set C is *algebraically open* (or linearly open) if for any $x, y \in X$ the preimage of C under the mapping $\alpha \mapsto x + \alpha y$ is open in $\mathbb{R}$. A set is *algebraically closed* (or linearly closed) if its complement is algebraically open, or equivalently, if the above preimages are closed in $\mathbb{R}$. We will see in the next chapter that *topologically closed* sets in a topological vector space are always algebraically closed and that, for a convex set in a Euclidean space, the two notions coincide. Clearly, if C is convex, then the preimages of C under the mapping $\alpha \mapsto x + \alpha y$ are intervals.

Theorem 1.48. *Let C be convex and algebraically closed. Then C^∞ is algebraically closed and*

$$C^\infty = \bigcap_{\alpha>0} \alpha(C - x)$$

for every $x \in C$. In other words, $y \in C^\infty$ if there exists even one $x \in C$ such that $x + \alpha y \in C$ for every $\alpha > 0$.

Proof. Let $x \in C$ and $y \neq 0$ be such that $x + \alpha y \in C$ for every $\alpha > 0$ and let $x' \in C$ and $\alpha' > 0$ be arbitrary. It suffices to show that $x' + \alpha' y \in C$. Since $x + \alpha y \in C$ for every $\alpha \geq \alpha'$, we have, by convexity of C,

$$x' + \alpha' y + \frac{\alpha'}{\alpha}(x - x') = (1 - \frac{\alpha'}{\alpha})x' + \frac{\alpha'}{\alpha}(x + \alpha y) \in C \quad \forall \alpha \geq \alpha'.$$

If C is algebraically closed, we must have $x' + \alpha' y \in C$. □

If C is an algebraically closed convex set containing the origin, Theorem 1.48 gives $C^\infty = \bigcap_{\alpha>0} \alpha C$. In particular, if C is an algebraically closed convex cone, then necessarily $0 \in C$ and thus, $C^\infty = C$. In general, $C \subseteq C^\infty$ for every convex cone C. The equality fails e.g. for $C = \{(x_1, x_2) \in \mathbb{R}^2 \,|\, x_i > 0\}$.

Given a convex function f, the function

$$f^\infty(x) = \inf\{\alpha \,|\, (x, \alpha) \in (\operatorname{epi} f)^\infty\}$$

is called the *recession function* of f. By Corollary 1.33, f^∞ is sublinear. We will say that a function is *algebraically closed* if its epigraph is algebraically closed.

Theorem 1.49. *Let f be a proper convex function. Then*

$$f^\infty(x) = \sup_{\bar{x}\in\operatorname{dom} f} \sup_{\lambda>0} \frac{f(\bar{x} + \lambda x) - f(\bar{x})}{\lambda}.$$

If f is algebraically closed, then f^∞ is algebraically closed and

$$f^\infty(x) = \sup_{\lambda>0} \frac{f(\bar{x} + \lambda x) - f(\bar{x})}{\lambda} = \lim_{\lambda\nearrow\infty} \frac{f(\bar{x} + \lambda x) - f(\bar{x})}{\lambda}$$

for every $\bar{x} \in \operatorname{dom} f$.

Proof. The condition $(x, \alpha) \in (\operatorname{epi} f)^\infty$ means that

$$\begin{aligned}
&(\bar{x} + \lambda x, \bar{\alpha} + \lambda\alpha) \in \operatorname{epi} f \quad \forall(\bar{x}, \bar{\alpha}) \in \operatorname{epi} f,\ \lambda > 0\\
\iff &f(\bar{x} + \lambda x) \le \bar{\alpha} + \lambda\alpha \quad \forall(\bar{x}, \bar{\alpha}) \in \operatorname{epi} f,\ \lambda > 0\\
\iff &f(\bar{x} + \lambda x) \le f(\bar{x}) + \lambda\alpha \quad \forall\bar{x} \in \operatorname{dom} f,\ \lambda > 0.
\end{aligned}$$

When f is proper, the last condition can be written as

$$\sup_{\bar{x}\in\operatorname{dom} f} \sup_{\lambda>0} \frac{f(\bar{x} + \lambda x) - f(\bar{x})}{\lambda} \le \alpha,$$

which gives the first expression. The rest follows from Theorem 1.48 and Lemma 1.44. □

Corollary 1.50. *Given a proper convex function f, one has*

$$f(\bar{x} + \lambda x) \le f(\bar{x}) \quad \forall\bar{x} \in \operatorname{dom} f,\ \lambda > 0$$

if and only if $f^\infty(x) \le 0$. If f is algebraically closed, this holds if there is even one $\bar{x} \in \operatorname{dom} f$ such that $f(\bar{x} + \lambda x)$ is nonincreasing in λ.

By the above, a convex function is constant with respect to the linear space $\{x \in X \,|\, f^\infty(x) \le 0,\ f^\infty(-x) \le 0\}$ known as the *constancy space* of f.

Exercise 1.51. *Let f be a convex function with $\operatorname{dom} f = \mathbb{R}$. Show that*

$$f^\infty(x) = \begin{cases} v^+x & \text{if } x \ge 0,\\ v^-x & \text{if } x \le 0,\end{cases}]$$

where $v^+ = \operatorname{ess\,sup} f'$ and $v^- = \operatorname{essinf} f'$; see Theorem 1.13. How does the expression change when $\operatorname{dom} f$ is not all of $\mathbb{R}$?

Exercise 1.52. *Define* $f(x) = \ln Ee^x$ *on* L^0 *as in Example 1.36 and show that* $f^\infty(x) = \text{ess sup}\, x$.

Exercise 1.53. *Let* f *be sublinear. Show that* $f^\infty \le f$ *and that equality holds when* f *is algebraically closed.*

Theorem 1.54. *If* f *is proper, convex and algebraically closed, then*

$$(\text{lev}_\alpha f)^\infty = \text{lev}_0 f^\infty$$

for every $\alpha \in \mathbb{R}$ *with* $\text{lev}_\alpha f \neq \emptyset$.

Proof. The condition $y \in (\text{lev}_\alpha f)^\infty$ means that $f(x+\lambda y) \le \alpha$ for all $x \in \text{lev}_\alpha f$ and $\lambda > 0$, or equivalently, that $(x, \alpha) + \lambda(y, 0) \in \text{epi} f$ for all $x \in \text{lev}_\alpha f$ and $\lambda > 0$. When f is algebraically closed, Theorem 1.48 says that the latter property is equivalent to $(y, 0) \in (\text{epi} f)^\infty$. Since by Theorem 1.49, $(\text{epi} f)^\infty$ is algebraically closed, this means that $f^\infty(y) \le 0$. □

Exercise 1.55. *Show that the level sets of an algebraically closed function are algebraically closed.*

1.6 Separation of convex sets

A *hyperplane* is an affine set of the form $\{x \in X \,|\, l(x) = \alpha\}$, where α is a scalar and l is a linear functional (i.e. a real-valued function) not identically zero. A hyperplane is said to *separate* two sets C_1 and C_2 if the sets belong to opposite sides of the hyperplane in the sense that

$$l(x_1) \ge \alpha \quad \forall x_1 \in C_1 \quad \text{and} \quad l(x_2) \le \alpha \quad \forall x_2 \in C_2$$

The separation is said to be *proper* unless both sets are contained in the hyperplane. In other words, proper separation means that

$$\inf\{l(x_1 - x_2) \,|\, x_i \in C_i\} \ge 0 \quad \text{and} \quad \sup\{l(x_1 - x_2) \,|\, x_i \in C_i\} > 0.$$

A proper linear subspace of X is said to be *maximal* if it cannot be properly enlarged to another proper subspace of X.

Lemma 1.56. *A proper linear subspace is a maximal if and only if it is a hyperplane containing the origin.*

Proof. If L is a maximal proper subspace and if $y \notin L$, then X is the linear span of $L \cup \{y\}$. For every $x \in X$, there are unique $\alpha \in \mathbb{R}$ and a $z \in L$ such that $x = \alpha y + z$. Indeed, if $(\alpha', z') \in \mathbb{R} \times L$ is another such pair, we have $(\alpha' - \alpha)y = z - z' \in L$, so $\alpha' = \alpha$ and $z' = z$. As is easily checked, the map $x \mapsto \alpha$ is linear and has L as its null space. Conversely, let $l : X \to \mathbb{R}$ be linear and $L := \{x \,|\, l(x) = 0\}$. If l is nonzero, L is a proper subspace of X. If $y \notin L$, then for every $x \in X$ there is an $\alpha \in \mathbb{R}$ such that $l(x - \alpha y) = l(x) - \alpha l(y) = 0$. Thus, X is the linear span of $L \cup \{y\}$ so L is maximal. □

Theorem 1.57. *If C is a nonempty algebraically open convex set with $0 \notin C$, then there exists a linear functional l such that*

$$\inf\{l(x) \,|\, x \in C\} \geq 0 \quad \textit{and} \quad \sup\{l(x) \,|\, x \in C\} > 0.$$

Proof. By Zorn's lemma, there exists a maximal element L among all subspaces disjoint from C. By Lemma 1.56, it suffices to show that L is a maximal proper subspace. Since L is disjoint from C, it is also disjoint from $\operatorname{pos} C$ and thus, from $D = L + \operatorname{pos} C$. In particular, $0 \notin D$ so that $D \cap (-D) = \emptyset$. We also have $D \cup (-D) \cup L = X$ since if there was an $x \notin D \cup (-D) \cup L$ the linear span of $L \cup \{x\}$ would be disjoint from D, and thus, also from C, contradicting the maximality of L. Indeed, if $\alpha x + \alpha' x' \in D$ for some $\alpha, \alpha' \in \mathbb{R}$ and $x' \in L$, then since $D + L = D$, we would have $\alpha x \in D$ which is in contradiction with $x \notin D \cup (-D)$ since D is a cone not containing the origin.

Suppose that L is not a maximal proper subspace. Then there exist $x_1 \in D$ and $x_2 \in -D$ such that x_2 is not in the linear span of $L \cup \{x_1\}$. Since C is algebraically open, D is also algebraically open (exercise) so the preimage of $D \cup (-D)$ under the mapping $\alpha \mapsto \alpha x_1 + (1 - \alpha_2) x_2$ is the union of two disjoint open intervals in $\mathbb{R}$. One of the intervals contains 0 and the other 1, so there exists an $\alpha \in (0, 1)$ such that $\alpha x_1 + (1 - \alpha_2) x_2 \in L$. But this implies that x_2 is in the linear span of $L \cup \{x_1\}$ which is a contradiction. □

The conclusion of Theorem 1.57 implies that $l(x) > 0$ for all $x \in C$. Indeed, since $l(\bar{x}) > 0$ for some $\bar{x} \in C$, one cannot have $l(x') = 0$ for an $x' \in C$ since the algebraic openness of C would give the existence of an $\alpha > 1$ such that $\alpha x' + (1 - \alpha)\bar{x} \in C$ and the linearity of l would give $l(\alpha x' + (1 - \alpha)\bar{x}) < 0$.

Corollary 1.58. *If C_1 and C_2 are nonempty disjoint convex sets such that $C_1 - C_2$ is algebraically open, then C_1 and C_2 can be properly separated.*

Proof. Apply Theorem 1.57 to the set $C_1 - C_2$. □

In a typical situation, we have two convex sets one of which is algebraically open.

Exercise 1.59. *Show that the sum of two sets is algebraically open if one of the sets is algebraically open.*

Corollary 1.60 (Hahn–Banach)**.** *Let p be a seminorm on X and f a linear function on a linear subspace L such that $f \leq p$ on L. There exists a linear function $\bar{f}$ on X such that $\bar{f} = f$ on L and $\bar{f} \leq p$ on X.*

Proof. Apply Corollary 1.58 to the sets $C_1 = \{(x, \alpha) \,|\, p(x) < \alpha\}$ and $C_2 = \{(x, \alpha) \,|\, x \in L,\ f(x) = \alpha\}$. To see that C_1 is algebraically open, let $(x, \alpha) \in C_1$, $(y, \beta) \in X \times \mathbb{R}$ and $f(\lambda) = p(x + \lambda y) - \alpha - \lambda\beta$. It suffices to show that there is a $\lambda > 0$ such that $f(\lambda) < 0$. Since f is convex, $f(\lambda) \leq \lambda f(1) + (1 - \lambda) f(0)$, where $f(0) < 0$ and $f(1)$ is finite. □

Chapter 2

Topological properties

Topology has an important role in optimization. For example, existence of solutions is often established through compactness arguments as e.g. in the "direct method in the calculus of variations". Continuity and differentiability properties of optimal value functions turn out to be intimately related to existence criteria and to optimality conditions.

This chapter studies some basic properties of optimization problems in topological spaces. Section 2.2 studies lower semicontinuous functions and the existence of optimal solutions in general topological spaces. Sections 2.4 and 2.5 study the topological properties of convex sets and functions in topological vector spaces. Neither Hausdorff property nor local convexity will be assumed in this chapter.

2.1 Topological spaces

A *topology* on a set X is a collection τ of subsets of X such that

1. $X, \emptyset \in \tau$
2. $\bigcup_{i\in I} U_i \in \tau$ for any collection $\{U_i\}_{i\in I} \subset \tau$
3. $\bigcap_{i\in I} U_i \in \tau$ for any finite collection $\{U_i\}_{i\in I} \subset \tau$

Members of τ are called *open* sets. Their complements are said to be *closed*.

Example 2.1 (Metric spaces)**.** *A* metric *on a set X is a function $d : X \times X \to \mathbb{R}_+$ such that $d(x,y) = d(y,x)$, $d(x,z) \le d(x,y) + d(y,z)$ for all $x,y,z \in X$ and $d(x,y) = 0$ if and only if $x = y$. An* open ball *is a set of the form $\mathbb{B}^o(x,\delta) = \{y \in X \mid d(x,y) < \delta\}$. If we define τ as the collection of sets $U \subseteq X$ such that for every $x \in U$ there is a $\delta > 0$ such that $\mathbb{B}^o(x,\delta) \subset U$, then τ is a topology on X known as the* metric topology*. A set is closed in the metric topology if and only if it is* sequentially closed*, i.e. if every converging sequence in the set has its limit in the set; see e.g. [5, Proposition 4.6].*

Example 2.2 (Euclidean topology)**.** *The* Euclidean topology *on* $\mathbb{R}^n$ *is the metric topology associated with*

$$d(x,y) = |x-y| := \left(\sum_{i=1}^n |x_i - y_i|^2\right)^{\frac{1}{2}}.$$

Example 2.3 (Lebesgue spaces)**.** *For* $p \in [1,\infty)$*, the function*

$$d(x,y) = \|x-y\|_{L^p} := \left[\int_\Omega |x-y|^p d\mu\right]^{\frac{1}{p}}$$

is a metric on $L^p(\Omega, \mathcal{F}, \mu)$*. The function*

$$d(x,y) = \|x-y\|_{L^\infty} := \inf\{\alpha \,|\, \mu(\{\omega \,|\, |x(\omega) - y(\omega)| \geq \alpha\}) = 0\}$$

is a metric on $L^\infty(\Omega, \mathcal{F}, \mu)$*. If* $\mu(\Omega) < \infty$*, the function*

$$d(x,y) = \int_\Omega \frac{|x-y|}{|x-y|+1} d\mu$$

is a metric on $L^0(\Omega, \mathcal{F}, \mu)$*.*

Not all topologies correspond to metrics. We will see later that in convex analysis, some of the most interesting topologies (e.g. weak topologies generated by collections of linear functionals) go beyond metric topologies.

The *interior* of a set C is the union of all open sets contained in C. The *closure* of C is the intersection of all closed sets that contain C. The interior and the closure of a set C will be denoted by $\operatorname{int} C$ and $\operatorname{cl} C$, respectively.

Exercise 2.4. *Show that* $\operatorname{int} C = (\operatorname{cl}(C^c))^c$ *and* $\operatorname{cl} C = (\operatorname{int}(C^c))^c$*.*

A set $U \subseteq X$ is a *neighborhood* of a point x if $x \in \operatorname{int} U$. A collection $\mathcal{B}(x)$ of neighborhoods of a point $x \in X$ is said to be a *neighborhood base* at x if for any neighborhood U of x there is a $V \in \mathcal{B}(x)$ such that $V \subseteq U$. In Example 2.1, the collection of open balls $\mathbb{B}^o(x,\delta)$ is a neighborhood base of x, as is the collection of *closed balls* $\mathbb{B}(x,\delta) = \{y \in X \,|\, d(y,x) \leq \delta\}$ for $\delta > 0$.

It is easily checked that the intersection of a collection of topologies on X is again a topology on X. If $\mathcal{E}$ is a collection of subsets of X, then the intersection of all topologies containing $\mathcal{E}$ is called the *topology generated by* $\mathcal{E}$.

Exercise 2.5. *Let* τ *be the topology generated by a collection* $\mathcal{E}$ *of sets. Show that*

1. τ *consists of* X *and the unions of finite intersections of the members of* $\mathcal{E}$*.*
2. *each* $x \in X$ *has a neighborhood base* $\mathcal{B}(x)$ *consisting of finite intersections of members of* $\mathcal{E}$*.*
3. *if* $\mathcal{E}$ *is the set of open balls in a metric space, then* τ *is the metric topology.*

If $(X_i, \tau_i)_{i\in I}$ is a collection of topological spaces, then the *product topology* on the product space $\prod_{i\in I} X_i$ is the topology generated by sets of the form

$$\{x \in \prod_{i\in I} X_i \mid x_j \in V_j\},$$

where $j \in I$ and $V_j \in \tau_j$. Every point $x \in \prod_{i\in I} X_i$ has a neighborhood base consisting of sets of the form $\prod_{i\in I} V_i$ where $V_i \in \tau_i$ differs from X_i only for finitely many $i \in I$.

A function f from a topological space (X, τ) to another (Y, σ) is *continuous at a point* x if the preimage of every neighborhood of $f(x)$ is a neighborhood of x. The function is *continuous* if it is continuous at every point.

Exercise 2.6. *Show that a function is continuous if and only if the preimage of every open set is open. If the topology σ is generated by a collection $\mathcal{E}$ of sets and if $f^{-1}(V) \in \tau$ for every $V \in \mathcal{E}$, then f is continuous.*

Lemma 2.7. *If f is a continuous function on $X_1 \times X_2$, then for every $x_2 \in X_2$, the function $x_1 \mapsto f(x_1, x_2)$ is continuous on X_1.*

Proof. The preimage of any neighborhood V of $f(x_1, x_2)$ contains a neighborhood of (x_1, x_2) of the form $U_1 \times U_2$. In particular, $U_1 \times \{x_2\} \subset U_1 \times U_2 \subset F^{-1}(V)$, which means that U_1 is in the preimage of V under $x_1 \mapsto f(x_1, x_2)$. □

If $\{f_i\}_{i\in I}$ is a collection of functions f_i from X to topological spaces (Y_i, τ_i), then the *weak topology* on X generated by the collection is the topology generated by sets of the form $f_i^{-1}(U_i)$ where $U_i \in \tau_i$ and $i \in I$. In other words, the weak topology is the intersection of all topologies under which every f_i is continuous.

A set C is *compact* if any collection of open sets whose union contains C has a finite subcollection whose union also contains C. A collection of sets has the *finite intersection property* if every finite subcollection has a nonempty intersection.

Lemma 2.8. *A set is compact if and only if every collection of its closed subsets having the finite intersection property has a nonempty intersection.*

Proof. Exercise on de Morgan's laws. □

Lemma 2.9. *Let f be a contiuous function from (X, τ) to (Y, σ). If $C \subset X$ is compact, then $f(C)$ is compact.*

Proof. Let (V_j) be a collection of open sets whose union contains $f(C)$. The union of the preimages $f^{-1}(V_j)$ then contains C. Since C is compact, it is contained also in a finite subcollection. The corresponding finite collection of V_j's then contains $f(C)$. □

Theorem 2.10 (Tychonoff product theorem)**.** *The product of compact sets is compact in the product topology.*

Proof. See e.g. [5, Section 4.6]. □

A sequence $(x_n)_{n=1}^\infty$ in a metric space (X, d) is said to be a *Cauchy sequence* if for every $\varepsilon > 0$ there is an N such that $d(x_n, x_m) \le \varepsilon$ for every $n, m \ge N$. A metric space is *complete* if every Cauchy sequence converges in the sense that there is an $\bar{x} \in X$ such that for every $\varepsilon > 0$ there is an N such that $d(x_n, \bar{x}) \le \varepsilon$ for $n \ge N$. Euclidean, Lebesgue and the L^0-spaces are examples of complete metric spaces.

Theorem 2.11 (Baire category theorem)**.** *Let (X, d) be a complete metric space. If $(C_n)_{n=1}^\infty$ is a countable collection of closed sets whose union equals X, then $\operatorname{int} C_n \neq \emptyset$ for some n.*

Proof. See e.g. [5, Theorem 5.9]. □

2.2 Existence of optimal solutions

According to the classical theorem of Weierstrass, a continuous function achieves it minimum and maximum over a compact set. If we are only interested in one extreme, continuity can be replaced by "semicontinuity". This principle is sometimes referred to as the "direct method" in calculus of variations; see e.g. [1, **?**].

An extended real-valued function on a topological space X is *lower semicontinuous* (lsc) if its epigraph is a closed set in $X \times \mathbb{R}$ with respect to the product topology. A set C is closed iff its indicator function δ_C is lsc.

Theorem 2.12. *The following are equivalent*

(a) f *is lsc,*

(b) $\operatorname{lev}_\alpha f$ *is closed for every* $\alpha \in \mathbb{R}$,

(c) For every $x \in X$,

$$\sup_{U \in N(x)} \inf_{x' \in U} f(x') \ge f(x),$$

where $N(x)$ is the collection of neighborhoods of x.

Proof. Recall that each point in $X \times \mathbb{R}$ has a neighborhood base consisting of sets of the form $U \times (\alpha_1, \alpha_2)$ where $U \subset X$ is open. Assume f is lsc. To show that $\operatorname{lev}_\alpha f$ is closed it suffices to show that for every $x \notin \operatorname{lev}_\alpha f$ there is an open set $U \ni x$ which is disjoint from $\operatorname{lev}_\alpha f$. If $x \notin \operatorname{lev}_\alpha f$ then $(x, \alpha) \notin \operatorname{epi} f$ so there exists and open set $U \ni x$ and an interval $(\alpha_1, \alpha_2) \ni \alpha$ such that $[U \times (\alpha_1, \alpha_2)] \cap \operatorname{epi} f = \emptyset$. In particular, $f(x') > \alpha$ for every $x' \in U$, so U is disjoint of $\operatorname{lev}_\alpha f$.

Assume (b) holds, or in other words, that the sets $U_\alpha = \{x' \mid f(x') > \alpha\}$ are open. Given an x, we have $U_\alpha \in N(x)$ for every $\alpha < f(x)$ and thus,

$$\sup_{U \in N(x)} \inf_{x' \in U} f(x') \ge \sup_{\alpha < f(x)} \inf_{x' \in U_\alpha} f(x') \ge \sup_{\alpha < f(x)} \alpha = f(x),$$

so (c) holds.

Assume (c) and let $(x, \alpha) \notin \operatorname{epi} f$. Since $\alpha < f(x)$, there is a $U \in N(x)$ such that

$$\inf_{x' \in U} f(x') > \alpha.$$

Choosing $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $\alpha_1 < \alpha < \alpha_2 < \inf_{x' \in U} f(x)$, we have that $[U \times (\alpha_1, \alpha_2)]$ is a neighborhood of (x, α) which is disjoint from $\operatorname{epi} f$. Since $(x, \alpha) \notin \operatorname{epi} f$ was arbitrary, $\operatorname{epi} f$ must be closed. □

Example 2.13 (Integral functionals)**.** *Define I_f as in Example 1.17. Assume that f is nonnegative and that $f(\cdot, \omega)$ is lower semicontinuous on $\mathbb{R}^n$ for each ω. Then I_f is lower semicontinuous on $L^0(\Omega, \mathcal{F}, \mu; \mathbb{R}^n)$.*

Proof. Since L^0 is a metric space, it suffices to verify that $\operatorname{lev}_\alpha I_f$ is sequentially closed for every $\alpha \in \mathbb{R}$; see e.g. [5, Section 4.1]. Let $x^\nu \in \operatorname{lev}_\alpha I_f$ such that $x^\nu \to x$. The convergence in the metric given in Example 2.3 implies convergence in measure. Passing to a subsequence if necessary, we may assume that $x^\nu \to x$ almost everywhere; see e.g. [5, Theorem 2.30]. Then, by Fatou's lemma,

$$\liminf_\nu I_f(x^\nu) \geq \int_\Omega \liminf_\nu f(x^\nu(\omega), \omega) d\mu(\omega) \geq \int_\Omega f(x(\omega), \omega) d\mu(\omega) = I_f(x),$$

so that $x \in \operatorname{lev}_\alpha I_f$ □

Theorem 2.14. *Let f be a lower semicontinuous function such that $\operatorname{lev}_\alpha f$ is nonempty and compact for some $\alpha \in \mathbb{R}$. Then $\operatorname{argmin} f$ is nonempty and compact.*

Proof. Let $\bar{\alpha} \in \mathbb{R}$ be such that $\operatorname{lev}_{\bar{\alpha}} f$ is nonempty and compact. If $\bar{\alpha} = \inf f$ the claim clearly holds. Otherwise, the collection of closed sets $\{\operatorname{lev}_\alpha f \mid \alpha \in (\inf f, \bar{\alpha})\}$ has the finite intersection property. Since the sets are contained in the compact set $\operatorname{lev}_{\bar{\alpha}} f$, their intersection, which equals $\operatorname{argmin} f$, is nonempty; see Lemma 2.8. Being a closed subset of a compact set, $\operatorname{argmin} f$ is compact. □

The above result is a topological version of the "direct method" in calculus of variations; see e.g. [1, **?**]. The more common "sequential" version will be studied in the exercises.

Theorem 2.12 implies that the pointwise supremum of a collection of lsc functions is again lsc. Indeed, the epigraph of the pointwise supremum is the intersection of closed epigraphs. The *lower semicontinuous hull* $\operatorname{lsc} f$ of a function f is the pointwise supremum of all the lower semicontinuous functions dominated by f. The optimum value of a minimization problem is not affected if the essential objective is replaced by its lower semicontinuous hull.

Proposition 2.15. *If U is open, then $\inf_U \operatorname{lsc} f = \inf_U f$. In particular, $\inf f = \inf \operatorname{lsc} f$. Moreover,*

$$(\operatorname{lsc} f)(x) = \sup_{U \in N(x)} \inf_{x' \in U} f(x')$$

and $\operatorname{epi} \operatorname{lsc} f = \operatorname{cl} \operatorname{epi} f$.

Proof. The piecewise constant function

$$g(x) = \begin{cases} \inf_U f & x \in U, \\ -\infty & x \notin U \end{cases}$$

is lower semicontinuous (see Theorem 2.12) and dominated by f. Thus $\operatorname{lsc} f \geq g$ and $\inf_U \operatorname{lsc} f \geq \inf_U g = \inf_U f$. This must hold as an equality since $\operatorname{lsc} f \leq f$. By Theorem 2.12 and the first part of the lemma,

$$(\operatorname{lsc} f)(x) = \sup_{U \in N(x)} \inf_{x' \in U} (\operatorname{lsc} f)(x') = \sup_{U \in N(x)} \inf_{x' \in U} f(x').$$

The function

$$\bar{f}(x) = \inf\{\alpha \in \mathbb{R} \mid (x, \alpha) \in \operatorname{cl} \operatorname{epi} f\},$$

has $\operatorname{epi} \bar{f} = \operatorname{cl} \operatorname{epi} f$ so it is lower semicontinuous. If $g \leq f$ is lsc, then $\operatorname{cl} \operatorname{epi} f \subseteq \operatorname{epi} g$ so that $g \leq \bar{f}$. Thus $\bar{f} = \operatorname{lsc} f$. □

Much like convexity, lower semicontinuity is preserved under many algebraic operations.

Theorem 2.16. *If F is a continuous function from a topological space X to another U and h is a lsc function on U, then $h \circ F$ is lsc.*

Proof. Exercise. □

Corollary 2.17. *If f_i are proper and lsc then $f_1 + \cdots + f_m$ is lsc.*

In the convex case, Theorem 2.16 can be considerably strengthened; see Section 2.43.

Exercise 2.18 (Nonlinear programming)**.** *Show that the essential objective of the optimization model* (NLP) *is lower semicontinuous if the functions f_j are lower semicontinuous.*

Example 2.19 (Calculus of variations)**.** *Consider the problem of calculus of variations in Section 1.1. If f is nonnegative, then by Example 2.13, the integral functional I_f is lower semicontinuous on $L^p(\Omega)^n$. Since the differential operator $D : W^{m,p}(\Omega) \to L^p(\Omega)^n$ is continuous from the strong topology of $W^{m,p}(\Omega)$ to that of $L^p(\Omega)^n$, the objective $I_f \circ D$ is a lower semicontinuous, by Theorem 2.16. In order to guarantee the existence of minimizers, it suffices by Theorem 2.14, to show that $I_f \circ D$ has a compact level set. One possibility is to apply Poincaré inequality.*

The following gives a criterion for the lower semicontinuity of the optimum value function of a parametric optimization problem.

Theorem 2.20. *Let f be a lsc function on $X \times U$ such that for every $\bar{u} \in U$ and $\alpha \in \mathbb{R}$ there is a neighborhood $U_{\bar{u}}$ of $\bar{u}$ and a compact set $K_{\bar{u}} \subseteq X$ such that*

$$\{x \in X \mid \exists u \in U_{\bar{u}} : \ f(x, u) \leq \alpha\} \subseteq K_{\bar{u}}.$$

Then, $\varphi(u) = \inf_x f(x, u)$ is lsc and the infimum is attained for every $u \in U$.

Proof. The attainment of the infimum follows from Theorem 2.14 since the lower semicontinuity of f implies the lower semicontinuity of $x \mapsto f(x,u)$ for each fixed u. This can be verified by following the proof of Lemma 2.7 in showing that the inverse image of every set of the form (α,∞) is open.

Let $\alpha \in \mathbb{R}$. If $\bar{u} \notin \operatorname{lev}_\alpha \varphi$, then $X \times \{\bar{u}\} \cap \operatorname{lev}_\alpha f = \emptyset$. Since $\operatorname{lev}_\alpha f$ is closed, there exist for every $x \in X$ an $X_x \in \mathcal{B}(x)$ and a $U_x \in \mathcal{B}(\bar{u})$ such that $(X_x \times U_x) \cap \operatorname{lev}_\alpha f = \emptyset$. Since $K_{\bar{u}}$ is compact, there is a finite set of points $x_1,\dots,x_n$ such that $K_{\bar{u}} \subset \bigcup_{i=1}^n X_{x_i}$. Then $V = \bigcap_{i=1}^n (U_{x_i} \cap U_{\bar{u}})$ is a neighborhood of $\bar{u}$ and $V \cap \operatorname{lev}_\alpha \varphi = \emptyset$. Indeed, if $u \in V \cap \operatorname{lev}_\alpha \varphi$, there is an $x \in X$ such that $(x,u) \in \operatorname{lev}_\alpha f$. But this is impossible since $u \in V \subset U_{\bar{u}}$ and $(x,u) \in \operatorname{lev}_\alpha f$ imply $x \in K_{\bar{u}}$ while

$$K_{\bar{u}} \times V \subset \bigcup_{i=1}^n (X_{x_i} \times U_{x_i}),$$

which is disjoint from $\operatorname{lev}_\alpha f$. □

The condition in Theorem 2.20 is certainly satisfied if f is such that $\operatorname{dom} f(\cdot,u)$ is contained in a fixed compact set for every $u \in U$. When f is the indicator of a set, Theorem 2.20 can be written as follows.

Corollary 2.21. *Let $C \subset X \times U$ be a closed set such that for every $\bar{u} \in U$ there is a neighborhood $U_{\bar{u}}$ of $\bar{u}$ and a compact set $K_{\bar{u}} \subseteq X$ such that*

$$\{x \in X \mid \exists u \in U_{\bar{u}} : \ (x,u) \in C\} \subseteq K_{\bar{u}}.$$

Then, $\{u \in U \mid \exists x \in X : \ (x,u) \in C\}$ is closed.

It is clear that $f_1 + \dots + f_m \geq \operatorname{lsc} f_1 + \dots + \operatorname{lsc} f_m$ for any functions f_i. Thus, by Corollary 2.17, $\operatorname{lsc}(f_1 + \dots + f_m) \geq \operatorname{lsc} f_1 + \dots + \operatorname{lsc} f_m$. In general, however, the inequality may be strict.

Lemma 2.22. *If f_1 is continuous at x_0, then $\operatorname{lsc}(f_1 + f_2)(x_0) = f_1(x_0) + \operatorname{lsc} f_2(x_0)$. In particular, if f_1 is continuous, then $\operatorname{lsc}(f_1 + f_2) = f_1 + \operatorname{lsc} f_2$.*

Proof. For any $\varepsilon > 0$ there is an open set $U' \ni x_0$ such that $f_1(x) \leq f_1(x_0) + \varepsilon$ for every $x \in U'$ and thus,

$$\begin{aligned}
\operatorname{lsc}(f_1 + f_2)(x_0) &= \sup_{U \in N(x_0)} \inf_{x \in U} [f_1(x) + f_2(x)] \\
&\leq \sup_{U \in N(x_0)} \inf_{x \in U \cap U'} [f_1(x) + f_2(x)] \\
&\leq f_1(x_0) + \varepsilon + \sup_{U \in N(x_0)} \inf_{x \in U \cap U'} f_2(x) \\
&= f_1(x_0) + \varepsilon + (\operatorname{lsc} f_2)(x_0).
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we get $\operatorname{lsc}(f_1 + f_2)(x_0) \leq f_1(x_0) + (\operatorname{lsc} f_2)(x_0)$. □

In the convex case, the above result can be significantly strengthened; see Section 2.5.

2.3 Compactness in Euclidean spaces

The Heine–Borel theorem states that a subset in a Euclidean space is compact if and only if it is closed and bounded. This fact works particularly nicely with the scaling properties of convex sets; see Section 1.5.

Theorem 2.23. *A nonempty closed convex set $C \subset \mathbb{R}^d$ is bounded if and only if $C^\infty = \{0\}$.*

Proof. If C^∞ contains a nonzero vector then C contains a half-line so it is unbounded. On the other hand, if C is unbounded and $x \in C$, then the sets $D_\alpha = \alpha(C - x) \cap \{x \mid |x| = 1\}$ are nonempty, closed and increasing in α so they have the finite intersection property. Since they are contained in the compact set $\mathbb{B}_1$, their intersection is nonempty, by Lemma 2.8. By Theorem 1.48, the intersection equals $C^\infty \cap \{x \in \mathbb{R}^n \mid |x| = 1\}$. □

Corollary 2.24. *If f is a lower semicontinuous proper convex function on $\mathbb{R}^n$ such that*

$$\{x \in \mathbb{R}^n \mid f^\infty(x) \leq 0\} = \{0\},$$

then argmin f *is nonempty and compact.*

Proof. By Theorem 2.14, it suffices to prove that $\operatorname{lev}_\alpha f$ is nonempty and compact for some $\alpha \in \mathbb{R}$. By Theorem 1.54, the recession condition implies that $(\operatorname{lev}_\alpha f)^\infty = \{0\}$ for all $\alpha \geq \inf f$, so the compactness follows from Theorem 2.23. □

Theorem 2.25. *Let $A : \mathbb{R}^m \to \mathbb{R}^n$ be linear and $C \subseteq \mathbb{R}^m$ be closed and convex. If* $\ker A \cap C^\infty$ *is a linear space, then AC is closed and $(AC)^\infty = AC^\infty$.*

Proof. ?? □

Theorem 2.26. *Let f be a lower semicontinuous proper convex function on $\mathbb{R}^m \times \mathbb{R}^n$. If $\{x \in \mathbb{R}^n \mid f^\infty(x, 0) \leq 0\}$ is a linear space, then*

$$\varphi(u) = \inf_{x \in \mathbb{R}^n} f(x, u)$$

is a lower semicontinuous proper convex function and the infimum is attained for every u.

Proof. By Corollary 1.50, the linearity condition implies that $f(\cdot, u)$ is constant in the directions of $L := \{x \in \mathbb{R}^n \mid f^\infty(x, 0) \leq 0\}$. Thus,

$$\varphi(u) = \inf_{x \in L^\perp} f(x, u).$$

Each $\bar{u} \in \mathbb{R}^m$ has the unit ball $\mathbb{B}(\bar{u})$ as its neighborhood. To prove the lower semicontinuity and attainment of the infimum, it suffices, by Theorem 2.20, to show that the set

$$K_{\bar{u}} := \{x \mid \exists u \in \mathbb{B}(\bar{u}) : \ f(x, u) \leq \alpha, \ x \in L^\perp\}$$

is compact. By ??, it's recession cone can be expressed as

$$K_{\bar{u}}^{\infty} := \{x \in \mathbb{R}^n \mid f(x,0) \leq 0,\ x \in L^{\perp}\} = \{0\}.$$

The compactness thus follows from Theorem 2.23. □

2.4 Interiors of convex sets

For convex sets and functions, stronger topological properties are available as soon as the topology is compatible with the vector space operations. A *topological vector space* is a vector space X equipped with a topology τ such that

1. the function $(x_1, x_2) \mapsto x_1 + x_2$ is continuous from the product topology of $X \times X$ to X,
2. the function $(x, \alpha) \mapsto \alpha x$ is continuous from the product topology of $X \times \mathbb{R}$ to X,

Example 2.27. *Let $\|\cdot\|$ be a* norm *on a vector space X. Then the topology generated by the metric $d(x,y) = \|x - y\|$ makes X a topological vector space. In particular, Euclidean, Lebesgue and Sobolev spaces for $p \in [1, \infty]$ are topological vector spaces.*

Example 2.28. *Given a finite measure μ, the translation invariant metric*

$$d(x,y) = \int_{\Omega} \frac{|x-y|}{|x-y|+1} d\mu$$

is not a norm but, nevertheless, makes $L^0(\Omega, \mathcal{F}, \mu)$ a topological vector space.

Exercise 2.29. *If a set is open/closed, then it is algebraically open/closed.*

The following lists some fundamental consequences of the continuity of the vector space operations.

Lemma 2.30. *Let V be a neighborhood of the origin.*

(a) If $\alpha \neq 0$, then αV is a neighborhood of the origin

(b) There exists a neighborhood W of the origin such that $W + W \subset V$.

(c) If $x \in X$, then $V + x$ is a neighborhood of x.

Proof. Use Lemma 2.7 and the fact that products of open sets form a base for the product topology. □

If $\mathcal{B}(0)$ is a neighborhood base at the origin, then part (c) of Lemma 2.30 implies that $\{U + x \mid U \in \mathcal{B}(0)\}$ is a neighborhood base at x. In a topological vector space, the topology is thus completely determined by a neighborhood base $\mathcal{B}(0)$ of the origin. Indeed, a set U is open if and only if for each $x \in U$ there is a $V \in \mathcal{B}(0)$ such that $x + V \subset U$.

Lemma 2.31. *If C is convex then* $\operatorname{int} C$ *and* $\operatorname{cl} C$ *are convex.*

Proof. Let $\alpha \in (0,1)$. Since $\operatorname{int} C \subset C$ and C is convex, we have $\alpha \operatorname{int} C + (1-\alpha)\operatorname{int} C \subset C$. Since $\alpha \operatorname{int} C + (1-\alpha)\operatorname{int} C$ is open, it is contained in $\operatorname{int} C$. Since $\alpha \in (0,1)$ was arbitrary, this means that $\operatorname{int} C$ is convex.

Let $x_i \in \operatorname{cl} C$, $\alpha \in (0,1)$ and let U be a neighborhood of the origin. It suffices to show that $C \cap (\alpha x_1 + (1-\alpha)x_2 + U) \neq \emptyset$. Lemma 2.30 gives the existence of neighborhoods U_i of the origin such that $\alpha U_1 + (1-\alpha)U_2 \subset U$. Since $x_i \in \operatorname{cl} C$, there exist $z_i \in C \cap (x_i + U_i) \neq \emptyset$. We get

$$\alpha z_1 + (1-\alpha)z_2 \in \alpha(x_1 + U_1) + (1-\alpha)(x_2 + U_2) \subset \alpha x_1 + (1-\alpha)x_2 + U,$$

where the point on the left belongs to C by convexity. □

Theorem 2.32. *Let C be a convex set, $x \in \operatorname{int} C$ and $\bar{x} \in \operatorname{cl} C$. Then*

$$\alpha x + (1-\alpha)\bar{x} \in \operatorname{int} C \quad \forall \alpha \in (0,1].$$

If $\operatorname{int} C \neq \emptyset$, *then* $\operatorname{cl}\operatorname{int} C = \operatorname{cl} C$ *and* $\operatorname{int}\operatorname{cl} C = \operatorname{int} C$.

Proof. Let $U \subseteq C$ be an open neighborhood of x. The set $V = -\alpha/(1-\alpha)(U-x)$ is a neighborhood of the origin so there is a $z \in C \cap (\bar{x} + V)$. We get

$$\alpha x + (1-\alpha)\bar{x} \in \alpha x + (1-\alpha)(z - V) = \alpha U + (1-\alpha)z,$$

where the open set on the right is contained in C by convexity. This proves the first claim.

To prove that $\operatorname{cl}\operatorname{int} C = \operatorname{cl} C$, it suffices to show that for any $\bar{x} \in \operatorname{cl} C$ and any neighborhood U of $\bar{x}$, there is an $x' \in \operatorname{int} C \cap U$. Let $x \in \operatorname{int} C$. Then, by the continuity of $\alpha \mapsto \alpha x + (1-\alpha)\bar{x}$, there is an $\alpha \in (0,1)$ such that $\alpha x + (1-\alpha)\bar{x} \in U$. By the first claim, $\alpha x + (1-\alpha)\bar{x} \in \operatorname{int} C$.

To prove the last claim, let $z \in \operatorname{int}\operatorname{cl} C$ and $x \in \operatorname{int} C$. By continuity of the vector space operations, there is an $\varepsilon > 0$ such that $\bar{x} := z + \varepsilon(z - x) \in \operatorname{cl} C$. Let $\alpha = \varepsilon/(1+\varepsilon)$. By the first claim, $\alpha x + (1-\alpha)\bar{x} \in \operatorname{int} C$. The last claim now follows since $\alpha x + (1-\alpha)\bar{x} = z$ and $z \in \operatorname{int}\operatorname{cl} C$ was arbitrary. □

The *core* (or the algebraic interior) of a set C is the set of points $x \in C$ such that, for every $y \in X$, the set $\{\alpha \,|\, x + \alpha y \in C\}$ is a neighborhood of the origin in $\mathbb{R}$. The algebraic interior will be denoted by $\operatorname{core} C$. Note that $x \in \operatorname{core} C$ if and only if $\operatorname{pos}(C - x) = X$.

Exercise 2.33. *Show that* $\operatorname{core} C \supseteq \operatorname{int} C$. *Give an example of a set $C \subset \mathbb{R}^2$ for which* $\operatorname{core} C \neq \operatorname{int} C$.

A topological vector space X is said to be *barrelled* if every closed convex set C with $\operatorname{pos} C = X$ is a neighborhood of the origin.

Proof. If C is convex, we have $\operatorname{pos} C = \bigcup_{n=1}^{\infty} nC$. If the union equals X and C is closed then by Theorem 2.11, one of the sets nC, and thus, C has a nonempty interior. □

Theorem 2.34. *Let C be convex. Then* $\operatorname{core} C = \operatorname{int} C$ *holds in the following situations*

(a) *X is Euclidean,*

(b) $\operatorname{int} C \neq \emptyset$,

(c) *X is barrelled and C is closed.*

(c) *X is a complete metric space (e.g. a Banach space) and C is closed.*

The space X is barrelled, in particular, if the topology of X is induced by a metric under which X is complete.

Proof. If $x \in \operatorname{core} C$ and X is Euclidean, then there is a finite set of points $\{x_i\}_{i\in I} \subset C$ such that $x \in \operatorname{int}\operatorname{co}\{x_i\}_{i\in I}$ (exercise). By convexity, $\operatorname{co}\{x_i\}_{i\in I} \subset C$, so that $x \in \operatorname{int} C$.

Let $x' \in \operatorname{int} C$. If $x \in \operatorname{core} C$, there is an $\alpha \in (0,1)$ and a $z \in C$ such that $x = \alpha z + (1-\alpha)x'$. By convexity,

$$x + (1-\alpha)[\operatorname{int} C - x'] = \alpha z + (1-\alpha)\operatorname{int} C \subseteq C$$

where the set on the left is a neighborhood x.

When C is closed and $\operatorname{pos}(C-x) = X$, barrelledness of X implies that $C-x$ is a neighborhood of the origin so $x \in \operatorname{int} C$, by Lemma2.30(c).

If C is convex, $\operatorname{pos} C = \bigcup_{n=1}^{\infty} nC$. Thus, if X is a complete metric space and C is closed with $\operatorname{pos} C = X$, Theorem 2.11 implies that one of the sets nC, and thus, Lemma2.30(a), C has a nonempty interior. Part (b) then implies that $0 \in \operatorname{int} C$. □

Note that the last part of Theorem 2.34 covers e.g. Banach spaces or, more generally, Frechet spaces which are the completely metrizable locally convex vector spaces.

The *relative interior* of a set is its interior relatively its closed affine hull. That is, the relative interior $\operatorname{rint} C$ of a set C is the set of points $x \in C$ that have a neighborhood U such that $U \cap \operatorname{cl}\operatorname{aff} C \subset C$. In a Euclidean space, every convex set has a nonempty relative interior; see [11, Theorem 6.2]. Many results involving interiors of convex sets have natural counterparts involving relative interiors. In particular, it is easily checked that the above results hold with "rint" in the place of "int" throughout.

We end this section with some topological refinements of the algebraic separation theorems from Section 1.6. These will be based on the following simple but significant observation.

Exercise 2.35. *If a linear function is bounded from above on an open set, then it is continuous.*

Theorem 2.36. *Let C_1 and C_2 be disjoint convex sets of a topological vector space. If one of the sets is open, then there is a continuous linear functional that separates them properly.*

Proof. By Exercise 2.29, a topologically open set is algebraically open. Corollary 1.58 thus gives the existence of a linear function l that properly separates C_1 and C_2. If one of the sets is open, l must be continuous, by Exercise 2.35. □

The following is a version of the Hahn–Banach extension theorem.

Corollary 2.37. *Let L be a linear subspace of a topological vector space X. If $l : L \to \mathbb{R}$ is linear and continuous in the relative topology of L, then there exists a linear continuous $\bar{l} : X \to \mathbb{R}$ such that $\bar{l} = l$ on L. The extension $\bar{l}$ is unique if L is dense in X.*

Proof. The continuity of $l(x)$ on L means that there is an $\alpha > 0$ and an open convex set $U \ni 0$ such that $l(x) \leq \alpha$ for $x \in U \cap L$. Applying Theorem 2.36 to the sets $\{(x, l(x)) \,|\, x \in L\}$ and $U \times (\alpha, \infty)$, we get the existence of a continuous linear functional $k : X \times \mathbb{R} \to \mathbb{R}$ such that

$$k(x_1, l(x_1)) \leq k(x_2, \beta) \quad \forall x_1 \in L,\ (x_2, \beta) \in U \times (\alpha, \infty)$$

where the inequality does not reduce to equality. In particular, $x \mapsto k(x, l(x))$ is bounded on L so, by linearity, it has to be identically zero on L. By linearity, k is necessarily of the form $k(x, r) = \tilde{l}(x) + \gamma r$, where $\gamma \neq 0$ since otherwise we could not have $0 \leq k(x, \beta)$ for all $x \in U$ and $\beta > \alpha$ without this reducing to an equality. It is easily checked that the linear function $\bar{l} = -1/\gamma\tilde{l}$ is a continuous extension of l.

If l has two continuous extensions $\bar{l}_1$ and $\bar{l}_2$, then L is contained in the closed set $\ker(l_1 - l_2)$. If L is dense, then its closure and, thus, $\ker(\bar{l}_1 - \bar{l}_2)$ equals X so $\bar{l}_1 = \bar{l}_2$. □

2.5 Continuity of convex functions

If a function f is a continuous at a point $x \in \operatorname{dom} f$, then it is bounded on a neighborhood of that point. Indeed, for any $M > f(x)$, the preimage of $(-\infty, M)$ under f is a neighborhood of x. For convex functions, we have the following converse.

Theorem 2.38. *If a convex function is bounded from above on an open set, then the function is continuous throughout the core of its domain.*

Proof. Assume that $f(x) \leq M$ on a neighborhood U of a point $\bar{x}$. By translation, we may assume that $\bar{x} = 0$ and $f(0) = 0$. By convexity,

$$f(\alpha x) = f(\alpha x + (1 - \alpha)0) \leq \alpha f(x) + (1 - \alpha) f(0) = \alpha M$$

for any $x \in U$ and $\alpha \in (0, 1)$. Also,

$$\frac{1}{2} f(x) + \frac{1}{2} f(-x) \geq f(0) = 0,$$

so that $-f(\alpha x) \leq f(-\alpha x)$ for any $x \in X$. We thus have that $|f(x)| \leq \alpha M$ on the set $\alpha[U \cap (-U)]$ which is a neighborhood of the origin. Since $\alpha \in (0,1)$ was arbitrary, this implies that f is continuous at 0.

To finish the proof, it suffices to show that if $f(x) \leq M$ for x in an open set U, then f is bounded from above on a neighborhood of every point of $\operatorname{core}\operatorname{dom} f$. Let $x' \in U$. If $x \in \operatorname{core}\operatorname{dom} f$, there is an $\alpha \in (0,1)$ and a $z \in \operatorname{dom} f$ such that $x = \alpha z + (1-\alpha)x'$. By convexity,

$$f(\alpha z + (1-\alpha)w) \leq \alpha f(z) + (1-\alpha)f(w) \leq \alpha f(z) + (1-\alpha)M \quad \forall w \in U.$$

Thus, f is bounded from above on $W = \alpha z + (1-\alpha)U$, which is an open set containing x. □

Note that if f is linear, then Theorem 2.38 reduces to Exercise 2.35.

Corollary 2.39. *A convex function f is continuous on* $\operatorname{core}\operatorname{dom} f$ *in the following situations*

(a) X is Euclidean,

(b) X is a complete metric space and f is lower semicontinuous.

Proof. If $x \in \operatorname{core}\operatorname{dom} f$ and X is Euclidean, there is a finite set of points $\{x_i\}_{i\in I} \subset \operatorname{dom} f$ such that $x \in \operatorname{int}\operatorname{co}\{x_i\}_{i\in I}$ (exercise). The number $\alpha = \max_{i\in I} f(x_i)$ is then finite and, by convexity, $\operatorname{co}\{x_i\}_{i\in I} \subset \operatorname{lev}_\alpha f$, so the continuity follows from Theorem 2.38. In part (b), let $x \in \operatorname{core}\operatorname{dom} f$ and $\alpha > f(x)$. The lower semicontinuity of f implies, by Theorem 2.12, that the convex set $\operatorname{lev}_\alpha f$ is closed. Let $z \in X$. By part (a), $x \in \operatorname{core}\operatorname{dom} f$ implies that the function $\lambda \mapsto f(x+\lambda z)$ is continuous at the origin, so the interval $\{\lambda \,|\, f(x+\lambda z) < \alpha\}$ is a neighborhood of the origin. Since $z \in X$ was arbitrary, we have $x \in \operatorname{core}\operatorname{lev}_\alpha f$. By Theorem 2.34(c), $\operatorname{int}\operatorname{lev}_\alpha f \neq \emptyset$, so the continuity follows from Theorem 2.38. □

Part (b) of the above result gives the famous uniform boundedness principle.

Corollary 2.40 (Banach–Steinhaus)**.** *Given a family $\{A_i\}_{i\in I}$ of bounded linear operators form a Banach space X to a normed space, let*

$$f(x) = \sup_{i\in I} \|A_i(x)\|.$$

If $\operatorname{dom} f = X$, *then* $\sup_{i\in I} \|A_i\| < \infty$.

Proof. Exercise □

The lower semicontinuous hull $\operatorname{lsc} f$ of a convex function f is convex. Indeed, by Proposition 2.15 $\operatorname{epi}\operatorname{lsc} f = \operatorname{cl}\operatorname{epi} f$, where the latter set is convex by Lemma 2.31. The second part of the following result shows how the values of $\operatorname{lsc} f$ can be recovered from the values of f along a single line.

Theorem 2.41. *Let f be a convex function finite and continuous at a point. Then*

$$\operatorname{int}\operatorname{epi} f = \{(x,\alpha) \mid x \in \operatorname{int}\operatorname{dom} f,\ \alpha > f(x)\}$$

and for any $x_0 \in \operatorname{core}\operatorname{dom} f$ and any x

$$(\operatorname{lsc} f)(x) = \lim_{\lambda \nearrow 1} f(\lambda x + (1-\lambda)x_0).$$

Proof. If $(x,\alpha) \in \operatorname{int}\operatorname{epi} f$, then (x,α) has a neighborhood contained in $\operatorname{epi} f$. The neighborhood in $X \times \mathbb{R}$ contains a neighborhood of the product form $U \times (\alpha-\varepsilon, \alpha+\varepsilon)$ so $x \in U \subset \operatorname{int}\operatorname{dom} f$ and $f(x) < \alpha - \varepsilon$. On the other hand, if $x \in \operatorname{int}\operatorname{dom} f$ and $\alpha > f(x)$, then by Theorem 2.38, f is continuous at x, so for any $\beta \in (f(x), \alpha)$, there is a neighborhood U of x such that $f(z) < \beta$ for $z \in U$. Thus, $(x,\alpha) \in U \times (\beta, \infty) \subset \operatorname{epi} f$, so $(x,\alpha) \in \operatorname{int}\operatorname{epi} f$.

Since $\lambda \mapsto \lambda x + (1-\lambda)x_0$ is continuous, property (c) in Theorem 2.12 gives

$$(\operatorname{lsc} f)(x) \le \liminf_{\lambda \nearrow 1} f(\lambda x + (1-\lambda)x_0).$$

Let $(x,\alpha) \in \operatorname{cl}\operatorname{epi} f$. By the first claim, we have $(x^0, \alpha^0) \in \operatorname{int}\operatorname{epi} f$ for any $\alpha^0 > f(x^0)$. Applying the line segment principle of Theorem 2.41 to $\operatorname{epi} f$, we get $\lambda(x,\alpha) + (1-\lambda)(x^0,\alpha^0) \in \operatorname{int}\operatorname{epi} f$ for every $\lambda \in (0,1)$. Thus,

$$f(\lambda x + (1-\lambda)x^0) \le \lambda\alpha + (1-\lambda)\alpha^0$$

so $\limsup_{\lambda \nearrow 1} f(\lambda x + (1-\lambda)x_0) \le \alpha$. Since $(x,\alpha) \in \operatorname{cl}\operatorname{epi} f$ was arbitrary, the last part of Proposition 2.15 gives

$$(\operatorname{lsc} f)(x) \ge \limsup_{\lambda \nearrow 1} f(\lambda x + (1-\lambda)x_0).$$

which completes the proof. □

Corollary 2.42. *A lower semicontinuous convex function on the real line is continuous with respect to the closure of its domain.*

Proof. If the domain has a nonempty interior, the claim follows from Theorem 2.41. Otherwise it is trivial. □

In the convex case, Lemma 2.22 can be improved as follows.

Proposition 2.43. *Let f_1 and f_2 be convex functions such that f_1 is continuous at a point where f_2 is finite. Then*

$$\operatorname{lsc}(f_1 + f_2) = \operatorname{lsc} f_1 + \operatorname{lsc} f_2.$$

Proof. Let $x_0 \in \operatorname{dom} f_2$ be a point where f_1 is continuous and let $x \in X$. Since $x_0 \in \operatorname{dom}\operatorname{lsc}(f_1 + f_2)$, Corollary 2.42 gives

$$\operatorname{lsc}(f_1 + f_2)(x) = \lim_{\lambda \nearrow 1} \operatorname{lsc}(f_1 + f_2)(\lambda x + (1-\lambda)x_0)$$

(here both sides equal $+\infty$ unless $x \in \operatorname{dom} \operatorname{lsc}(f_1 + f_2)$). By Theorems 2.32 and 2.38, f_1 is continuous at $\lambda x + (1-\lambda)x_0$ for every $\lambda \in (0,1)$, so by Lemma 2.22,

$$\begin{aligned}\operatorname{lsc}(f_1 + f_2)(x) &= \lim_{\lambda \nearrow 1} [f_1(\lambda x + (1-\lambda)x_0) + (\operatorname{lsc} f_2)(\lambda x + (1-\lambda)x_0)] \\ &= (\operatorname{lsc} f_1)(x) + (\operatorname{lsc} f_2)(x),\end{aligned}$$

where the second equality comes from Theorem 2.41 and Corollary 2.42. □

Exercise 2.44. *Show that the conclusion of Proposition 2.43 may fail to hold if either convexity or the continuity condition is relaxed.*

Exercise 2.45. *Show that if A is a continuous linear mapping from X to U and h is a convex function on U continuous at a point of* $\operatorname{rge} A$*, then* $\operatorname{lsc}(h \circ A) = (\operatorname{lsc} h) \circ A$.

Chapter 3

Duality

Duality theory of convex optimization has important applications e.g. in mechanics, statistics, economics and finance. Duality theory yields optimality conditions and it is the basis for many algorithms for numerical optimization. This chapter studies the conjugate duality framework of Rockafellar [13] which unifies various other duality frameworks for convex optimization.

We start by studying conjugate convex functions on dual pairs of topological vector spaces. Conjugate convex functions are involved in many fundamental results in functional analysis. They also form the basis for dualization of optimization problems which will be studied in the remainder of this chapter.

3.1 Conjugate convex functions

Two vector spaces X and V are in *separating duality* if there is a bilinear form $(x, v) \mapsto \langle x, v\rangle$ on $X \times V$ such that $x = 0$ whenever $\langle x, v\rangle = 0$ for every $v \in V$ and $v = 0$ whenever $\langle x, v\rangle = 0$ for every $x \in X$. It is easily checked that $X = \mathbb{R}^n$ and $V = \mathbb{R}^n$ are in separating duality under the bilinear form $\langle x, v\rangle := x \cdot v$. Moreover, every continuous linear functional on $\mathbb{R}^n$ can be expressed as $x \mapsto x \cdot v$ for some $v \in \mathbb{R}^n$. To have a similar situation in the general case, we need appropriate topologies on X and V.

We will denote by $\sigma(X, V)$ the *weakest* topology on X under which the linear functions $x \mapsto \langle x, v\rangle$ are continuous for every $v \in V$. In other words, $\sigma(X, V)$ is the topology generated by sets of the form $\{x \in X \,|\, |\langle x, v\rangle| < 1\}$ where $v \in V$. It turns out (see e.g. [**?**]) that every $\sigma(X, V)$-continuous real-valued linear function can be expressed in the form $x \mapsto \langle x, v\rangle$. In general, there are many topologies with this property but, by the Mackey–Arens theorem, there is a *strongest* one among them; see e.g. [**?**]. This topology, know as the Mackey topology, is generated by sets of the form

$$\{x \in X \mid \sup_{v \in D} \langle x, v\rangle < 1\},$$

where $D \subset V$ is convex symmetric and $\sigma(V, X)$-compact. The Mackey topology is denoted by $\tau(X, V)$. Corresponding topologies $\sigma(V, X)$ and $\tau(V, X)$ are defined similarly on V. All the above topologies are *locally convex*, i.e. there is a neighborhood base $\mathcal{B}(0)$ consisting of convex sets.

Even though a dual pair of vector spaces is a purely algebraic concept, the most important pairs are constructed from vector spaces that already have a locally convex topology.

Example 3.1. *Let X be a locally convex Hausdorff topological vector space and let V be the set of continuous linear functionals on X. Pointwise addition and scalar multiplication make V a vector space and the bilinear form $\langle x, v\rangle := v(x)$ puts X and V in separating duality. If X is a normed space, then $\tau(X, V)$ is the norm topology and $\sigma(V, X)$ is known as the weak* topology.*

Proof. See e.g. [**?**]. □

The space V in Example 3.1 is called the *topological dual* of X. It can be given an explicit description in many important cases. Several examples can be found e.g. in [4]. Below are two.

Example 3.2. *For $p, q \in [1, \infty]$ such that $1/p + 1/q = 1$, the spaces $X = L^p(\Omega, \mathcal{F}, P)$ and $V = L^q(\Omega, \mathcal{F}, P)$ (assuming that μ is σ-finite if $p = 1$) are in separating duality under the bilinear form*

$$\langle x, v\rangle = \int_\Omega x(\omega)v(\omega)d\mu(\omega).$$

Indeed, for $p \in [1, \infty)$, L^q is the dual of the normed space L^p (see e.g. [5, Theorem 6.15]).

Example 3.3. *Let Ω be a compact Hausdorff space, X the space of continuous functions on Ω and V the space of Radon measures on Ω. The bilinear form*

$$\langle x, v\rangle = \int_\Omega x(\omega)dv(\omega).$$

puts X and V in separating duality. Indeed, V is the topological dual of $(X, \|\cdot\|)$, where $\|\cdot\|$ is the supremum norm; see [5, Theorem 7.17].

Exercise 3.4. *Show that a continuous linear functional on a linear subspace of a locally convex topological vector space has a unique extension if and only if the linear subspace is dense.*

The topologies $\sigma(X, V)$ and $\tau(X, V)$ are perfectly suited for convex sets.

Theorem 3.5. *The $\tau(X, V)$-closed convex hull of a set is the intersection of a collection of half-spaces of the form $\{x \mid \langle x, v\rangle \le \alpha\}$. In particular, a $\tau(X, V)$-closed convex set is $\sigma(X, V)$-closed.*

Proof. Let C be the $\tau(X,V)$-closed convex hull in question and $x \notin C$. Since C^c is open there exists, by local convexity, an open convex set $D \ni x$ which is disjoint from C. By Theorem 2.36, there is a continuous linear functional l separating C and D. The continuous linear functional can be expressed as $l(x) = \langle x, v\rangle$ for some $v \in V$. For the constant α, one can take $\alpha := \sup_C \langle x, v\rangle$. Since these half-spaces are $\sigma(X,V)$-closed, the set C must be $\sigma(X,V)$-closed as well. □

The *polar* of a set $C \subset X$ is defined by

$$C^\circ = \{v \in V \mid \langle x, v\rangle \leq 1,\ \forall x \in C\}.$$

Being the intersection of closed half-spaces containing the origin, C° is always $\sigma(V,X)$-closed and contains the origin.

Corollary 3.6 (Bipolar theorem)**.** *If $C \subset X$ contains the origin, then $C^{\circ\circ} = \operatorname{cl}\operatorname{co} C$.*

Proof. By Theorem 3.5, $\operatorname{cl}\operatorname{co} C$ is the intersection of closed half-spaces of the form $\{x \mid \langle x, v\rangle \leq \alpha\}$. When $0 \in C$, the half-spaces must have $\alpha \geq 0$. Half-spaces of the form $\{x \mid \langle x, v\rangle \leq 0\}$ can be expressed as intersections of half-spaces $\{x \mid \langle x, v\rangle \leq \alpha\}$ where $\alpha > 0$. Replacing each v by v/α, we see that $\operatorname{cl}\operatorname{co} C$ is the intersection of half-spaces of the form $\{x \mid \langle x, v\rangle \leq 1\}$. This proves the claim since, by definition, $C^{\circ\circ}$ is the intersection of the closed half-spaces $\{x \mid \langle x, v\rangle \leq 1\}$ which contain C. □

If C is a cone, C° equals the *polar cone* of C defined by

$$C^* := \{v \in V \mid \langle x, v\rangle \leq 0\ \forall x \in C\}.$$

This is a $\sigma(V,X)$-closed convex cone. For a cone, Corollary 3.6 can thus be written as $C^{**} = \operatorname{cl}\operatorname{co} C$.

Given an extended real-valued function f on X, its *conjugate* is the extended real-valued function f^* on V defined by

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

The conjugate of a function on V is defined analogously. The conjugate f^{**} of the conjugate f^* of f is known as the *biconjugate* of f. Being the pointwise supremum of continuous linear functions, f^* is convex and $\sigma(V,X)$-lsc on V. Similarly, the biconjugate f^{**} is convex and $\sigma(X,V)$-lsc on X. The following is the most useful form of the separation theorem.

Theorem 3.7. *If $f : X \to \overline{\mathbb{R}}$ is such that $(\operatorname{lsc}\operatorname{co} f)(x) > -\infty$ for every $x \in X$, then $f^{**} = \operatorname{lsc}\operatorname{co} f$. Otherwise, $f^{**} \equiv -\infty$. In particular, if f is a lsc proper convex function, then $f^{**} = f$.*

Proof. We have $(x, \alpha) \in \operatorname{epi} f^{**}$ if and only if

$$\alpha \geq \langle x, v\rangle - \beta \quad \forall (v, \beta) \in \operatorname{epi} f^*.$$

Here $(v, \beta) \in \operatorname{epi} f^*$ if and only if $f(x) \geq \langle x, v\rangle - \beta$ for every x. Thus, $\operatorname{epi} f^{**}$ is the intersection of the epigraphs of all continuous affine functionals dominated by f.

On the other hand, by Theorem 3.5, $\operatorname{epi} \operatorname{lsc} \operatorname{co} f$ is the intersection of all closed half-spaces

$$H_{v,\beta,\gamma} = \{(x, \alpha) \,|\, \langle x, v\rangle + \alpha\beta \leq \gamma\}$$

containing $\operatorname{epi} f$. We have $(\operatorname{lsc} \operatorname{co} f)(x) > -\infty$ for every $x \in X$ if and only if one of the half-spaces has $\beta \neq 0$, or in other words, there is an affine function h_0 dominated by f.

It thus suffices to show that if there is a half-space $H_{v,\beta,\gamma}$ containing $\operatorname{epi} f$ but not a point $(\bar{x}, \bar{\alpha})$, then there is an affine function h such that $f \geq h$ but $h(\bar{x}) > \bar{\alpha}$. If $\operatorname{epi} f \subseteq H_{v,\beta,\gamma}$, then necessarily $\beta \leq 0$. If $\beta < 0$, then the function $h(x) = \langle x, v/(-\beta)\rangle + \gamma/\beta$ will do. If $\beta = 0$, then $\operatorname{dom} f$ is contained in $\{x \,|\, \langle x, v\rangle \leq \gamma\}$ while $\bar{x}$ is not. It follows that f dominates the affine function $h(x) = h_0(x) + \lambda(\langle x, v\rangle - \gamma)$ for any $\lambda \geq 0$. Since $\langle \bar{x}, v\rangle > \gamma$, we have $h(\bar{x}) > \bar{\alpha}$ for λ large enough. □

The mapping $f \mapsto f^*$ is known as the *Legendre-Fenchel* transform. By Theorem 3.7, it provides a one-to-one correspondence between proper lower semicontinuous convex functions on X and V, respectively.

Exercise 3.8. *Let $X = V = \mathbb{R}$. Calculate the conjugate of*

(a) $f(x) = 1/p|x|^p$, *where* $p \in [1, \infty)$.

(b) $f(x) = e^x$.

(c) $f(x) = \begin{cases} -\ln x, & x > 0, \\ +\infty, & x \leq 0. \end{cases}$

(d) $f(x) = \max\{x, 0\}$.

Applying Theorem 3.7 to the indicator function of a closed convex set $C \subset X$ gives

$$\delta_C(x) = \sup_v \{\langle x, v\rangle - \sigma_C(v)\},$$

where $\sigma_C(v) = \sup\{\langle x, v\rangle \,|\, x \in C\}$ is known as the *support function* of C. The above says that $x \in C$ if and only if $\langle x, v\rangle \leq \sigma_C(v)$ for every $v \in V$. We thus obtain Theorem 3.5 as a special case of Theorem 3.7. When X is a normed space and C is the unit ball, σ_C is the dual norm on V. When C is a cone, we have

$$\sigma_C = \delta_{C^*},$$

where C^* is the polar cone of C.

Example 3.9 (Integral functionals)**.** *Let (Ω, F, μ) be a complete measure space, $X = L^p(\omega, \mathcal{F}, \mu; \mathbb{R}^n)$ and $V = L^q(\omega, \mathcal{F}, \mu; \mathbb{R}^n)$. Let f be an extended real-valued $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F}$-measurable function on $\mathbb{R}^n \times \Omega$ such that $f(\cdot, \omega)$ is lsc proper and convex for every ω. If $I_f(x) < \infty$ for some $x \in X$, then the conjugate (with respect to X and V) of the integral functional $I_f : X \to \overline{\mathbb{R}}$ can be expressed as $I_f^* = I_{f^*}$, where*

$$f^*(v, \omega) = \sup_x \{x \cdot v - f(x, \omega)\}$$

is a $\mathcal{B}(\mathbb{R}^n) \otimes \mathcal{F}$-measurable function on $\mathbb{R}^n \times \Omega$.

Proof. For any $v \in V$,

$$\begin{aligned}
I_f^*(v) &= \sup_{x \in X} \{\int x(\omega) \cdot v(\omega) d\mu(\omega) - \int f(x(\omega), \omega) d\mu(\omega)\} \\
&= \sup_{x \in X} \int [x(\omega) \cdot v(\omega) - f(x(\omega), \omega)] d\mu(\omega) \\
&\leq \int \sup_{x \in \mathbb{R}^n} [x \cdot v(\omega) - f(x, \omega)] d\mu(\omega) \\
&= \int f^*(v(\omega), \omega) d\mu(\omega) = I_{f^*}(v).
\end{aligned}$$

On the other hand, let $\varepsilon > 0$ and $x^1 \in X$ such that $I_f(x^1) < \infty$. By the measurable selection theorem (see e.g. [16, Theorem 14.6]), there is an $x^0 \in L^0(\Omega, F, \mu)$ such that

$$x^0(\omega) \cdot v(\omega) - f(x^0(\omega), \omega) \geq \min\{f^*(v(\omega), \omega), 1/\varepsilon\} - \varepsilon.$$

Defining $A_\nu = \{\omega \,|\, |x^0(\omega)| \leq \nu\}$ and $x^\nu = x^0 \mathbb{1}_{A_\nu} + x^1 \mathbb{1}_{A_\nu^c}$, we have $x^\nu \in X$ and

$$x^\nu(\omega) \cdot v(\omega) - f(x^\nu(\omega), \omega) \to x^0(\omega) \cdot v(\omega) - f(x^0(\omega), \omega)$$

almost surely. Thus, by Fatou's lemma,

$$\begin{aligned}
\liminf_{\nu \to \infty} \{\langle x^\nu, v \rangle - I_f(x^\nu)\} &\geq \int [x^0(\omega) \cdot v(\omega) - f(x^0(\omega), \omega)] d\mu(\omega) \\
&\geq \int [\min\{f^*(v(\omega), \omega), 1/\varepsilon\} - \varepsilon] d\mu(\omega).
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary and $f^*(v(\omega), \omega) \geq x^1(\omega) \cdot v(\omega) - f(x^1(\omega), \omega)$, the monotone convergence theorem gives

$$\liminf_{\nu \to \infty} \{\langle x^\nu, v \rangle - I_f(x^\nu)\} \geq I_{f^*}(v)$$

which completes the proof. □

The following gives some basic rules for calculating conjugates of functions obtained by elementary algebraic operations from another functions.

Exercise 3.10. *Show that if*

(b) $g(x) = f(x) + c$ *for* $c \in \mathbb{R}$*, then* $g^*(v) = f^*(v) - c$*,*

(a) $g(x) = f(x) - \langle x, \bar{v} \rangle$ *for* $\bar{v} \in V$*, then* $g^*(v) = f^*(v + \bar{v})$*,*

(b) $g(x) = f(x + \bar{x})$ *for* $\bar{x} \in X$*, then* $g^*(v) = f^*(v) - \langle \bar{x}, v \rangle$*,*

(b) $g(x) = \lambda f(x)$ *for* $\lambda > 0$*, then* $g^*(v) = \lambda f^*(v/\lambda)$*,*

(b) $g(x) = \lambda f(x/\lambda)$ *for* $\lambda > 0$*, then* $g(v) = \lambda f^*(v)$*.*

More general calculus rules for the Legendre–Fenchel transform will be derived in Section 3.3.

The following is an important source of compactness in locally convex spaces.

Theorem 3.11. *Let f be a lsc proper convex function on X such that $f(0) < \infty$. The following are equivalent:*

(a) $\operatorname{lev}_\alpha f$ *is a* $\tau(X, V)$*-neighborhood of the origin for some* $\alpha > f(0)$*,*

(b) $\operatorname{lev}_\alpha f$ *is a* $\tau(X, V)$*-neighborhood of the origin for every* $\alpha > f(0)$*,*

(c) $\operatorname{lev}_\beta f^*$ *is* $\sigma(V, X)$*-compact for some* $\beta > \inf f^*$*,*

(d) $\operatorname{lev}_\beta f^*$ *is* $\sigma(V, X)$*-compact for every* $\beta > \inf f^*$*.*

Even if f is not lsc, (a) implies (d).

Proof. Clearly, (b) implies (a). By Theorem 2.38, (a) implies the continuity of f at the origin which, in turn implies (b).

Part (a) means that there is a $C \in \tau(X, V)$ such that $0 \in C$ and $f \leq \delta_C + \alpha$. Since the topology $\tau(X, V)$ is generated by polars of convex symmetric and $\sigma(V, X)$-compact sets $D \subset V$, we may assume that $C = D^\circ$ for such a D. We have

$$f^*(v) \geq \sup\{\langle x, v \rangle - \delta_C(x) - \alpha\} = \sigma_C(v) - \alpha.$$

In particular, $\inf f^* \geq -\alpha$, so if $\beta > \inf f^*$, we have $\alpha + \beta > 0$ and then, by the bipolar theorem,

$$\begin{aligned} \operatorname{lev}_\beta f^* &\subseteq \{v \,|\, \sigma_C(v) \leq \alpha + \beta\} \\ &= (\alpha + \beta)\{v \,|\, \sigma_C(v) \leq 1\} \\ &= (\alpha + \beta)C^\circ = (\alpha + \beta)D^{\circ\circ} = (\alpha + \beta)D. \end{aligned}$$

Since $\operatorname{lev}_\beta f^*$ is $\sigma(V, X)$-closed, by the $\sigma(V, X)$-lower semicontinuity of f^*, it is $\sigma(V, X)$-compact since D is.

To prove that (c) implies (a), we assume temporarily that $f^*(0) = \inf f^* = 0$. Denote $D := \operatorname{lev}_\beta f^*$. If $v \notin D$, we have

$$\begin{aligned} \gamma_D(v) :&= \inf\{\eta > 0 \,|\, v \in \eta D\} \\ &= \inf\{\eta > 1 \,|\, v \in \eta D\} \\ &= \inf\{\eta > 1 \,|\, f^*(v/\eta) \leq \beta\} \\ &\leq \inf\{\eta > 1 \,|\, f^*(v)/\eta \leq \beta\} \\ &= f^*(v)/\beta, \end{aligned}$$

where the inequality holds since $f^*(v/\eta+(1-1/\eta)0) \leq f^*(v)/\eta+(1-1/\eta)f^*(0)$, by convexity. Letting $\delta := \inf f^* - \beta$, we thus have $f^*(v) \geq \beta\gamma_D(v) + \delta$ for all $v \in V$ and consequently, $f^* \geq \beta\gamma_{\bar{D}} + \delta$, where $\bar{D} := \mathrm{co}[D \cup (-D)]$. Exercise 3.10 gives

$$f^{**}(x) \leq \beta\gamma^*_{\bar{D}}(x/\beta) - \delta,$$

where $\gamma^*_{\bar{D}} = \delta_{\bar{D}^\circ}$ as is easily checked. The set $\bar{D}$ is convex, symmetric and $\sigma(V,X)$-compact[1] so $\bar{D}^\circ$ is a $\tau(X,V)$-neghborhood of the origin. This proves the claim in the case $f^*(0) = \inf f^* = 0$. The general case follows by considering the function $\tilde{g}(v) = f^*(\bar{v}+v)-f^*(\bar{v})$, where $\bar{v} \in \operatorname{argmin} f^*$. Indeed, by Exercise 3.10, g^* is $\tau(X,V)$-continuous at the origin if and only if f^{**} is. □

The following is probably the best known special case of Theorem 3.11.

Corollary 3.12 (Banach–Alaoglu). *The unit ball of the dual of a normed space is weak*-compact.*

Proof. The indicator δ_B of the unit ball $B \subset X$ satisfies conditions (a) and (b) of Theorem 3.11. Thus, the level sets of δ^*_B are weakly compact. But δ^*_B is the dual norm on V. □

Differential properties of convex functions can be conveniently described in terms of duality. A $v \in V$ is said to be a *subgradient* of f at $\bar{x}$ if

$$f(x) \geq f(\bar{x}) + \langle x - \bar{x}, v\rangle \quad \forall x \in X.$$

The set of all subgradients of f at $\bar{x}$ is denoted by $\partial f(\bar{x})$ and it is called the *subdifferential* of f at $\bar{x}$. Since $\partial f(\bar{x})$ is the intersection of closed half-spaces, it is a closed convex set. Writing the subdifferential condition as

$$\langle \bar{x}, v\rangle - f(\bar{x}) \geq \langle x, v\rangle - f(x) \quad \forall x \in X$$

we see that

$$v \in \partial f(\bar{x}) \quad \iff \quad f(\bar{x}) + f^*(v) \leq \langle \bar{x}, v\rangle.$$

The reverse inequality $f(\bar{x}) + f^*(v) \geq \langle \bar{x}, v\rangle$ is always valid, by definition of the conjugate.

Theorem 3.13. *Let f be a convex function and $\bar{x}$ a point where f is finite. The conjugate of the sublinear function $f'(\bar{x};\cdot)$ is the indicator function of $\partial f(\bar{x})$. If f is continuous at $\bar{x}$, then $\partial f(\bar{x})$ is a nonempty $\sigma(V,X)$-compact set and $f'(\bar{x};\cdot)$ is the support function of $\partial f(\bar{x})$.*

[1] By Exercise 1.19, the set $\bar{D}$ can be expressed as $F(\Delta, D, -D)$ where $\Delta = \{(\alpha_1,\alpha_2) \in \mathbb{R}^2_+ \mid \alpha_i \geq 0,\ \alpha_1 + \alpha_2 = 1\}$ and $F(\alpha, x_1, x_2) = \alpha_1 x_1 + \alpha_2 x_2$. Thus, $\bar{D}$ is compact since it is the continuous image of a compact set.

Proof. By Lemma 1.44,

$$\begin{aligned}
f'(\bar{x};\cdot)^*(v) &= \sup_{x\in X}\left\{\langle x,v\rangle - \inf_{\lambda>0}\frac{f(\bar{x}+\lambda x)-f(\bar{x})}{\lambda}\right\}\\
&= \sup_{x\in X,\lambda>0}\{\langle x,v\rangle - [f(\bar{x}+\lambda x)-f(\bar{x})]/\lambda\}\\
&= \sup_{x\in X,\lambda>0}\frac{1}{\lambda}\{\langle \lambda x,v\rangle - f(\bar{x}+\lambda x)+f(\bar{x})\}\\
&= \sup_{\lambda>0}\frac{1}{\lambda}\sup_{x\in X}\{\langle x,v\rangle - f(\bar{x}+x)+f(\bar{x})\}\\
&= \sup_{\lambda>0}\frac{1}{\lambda}\{f^*(v)+f(\bar{x})-\langle \bar{x},v\rangle\},
\end{aligned}$$

which equals the indicator function of $\partial f(\bar{x})$. If f is continuous at $\bar{x}$ then it is bounded from above on a neighborhood of $\bar{x}$. Since $f'(\bar{x};x)\le[f(\bar{x}+\lambda x)-f(\bar{x})]/\lambda$, we then get that $f'(\bar{x};\cdot)$ is bounded from above on a neighborhood of the origin. Since $f'(\bar{x};\cdot)$ is sublinear, it is then continuous throughout X, by Theorem 2.38. By Theorem 3.7, $f'(\bar{x};\cdot)$ equals its biconjugate, which by the first part is the support function of $\partial f(\bar{x})$. The compactness now comes from Theorem 3.11. □

If $f'(\bar{x};\cdot)$ is linear and continuous, f is said to be *Gateaux differentiable* at $\bar{x}$. By Theorem 3.13, $\partial f(\bar{x})$ consists then of a single point. That point is denoted by $\nabla f(\bar{x})$ and it is called the *gradient* of f at $\bar{x}$. Subgradients and subdifferentials provide useful substitutes for the gradient in many situations where the latter fails to exist. In particular, a point $\bar{x}\in X$ minimizes a function f if and only if $0\in\partial f(\bar{x})$. This generalizes the classical *Fermat's rule* which says that the gradient of a differentiable function vanishes at a minimizing point. The subdifferential makes good sense for nonsmooth and even extended real-valued functions. It will be seen in Section 3.3, that subdifferentials follow calculus rules reminiscent of those in classical analysis.

The subdifferential $\partial\delta_C(x)$ of the indicator function of a convex set $C\subseteq X$ is known as the *normal cone* of C at x. It is denoted by $N_C(x)$. If $x\notin C$ then $N_C(x)=\emptyset$ while for $x\in C$

$$N_C(x)=\{v\in V\,|\,\langle z-x,v\rangle\le 0\quad\forall z\in C\}.$$

Example 3.14 (Integral functionals)**.** *Let X, V and f be as in Example 3.9. The subdifferential of the integral functional $I_f: X\to\overline{\mathbb{R}}$ can be expressed as*

$$\partial I_f(x)=\{v\in V\,|\,v(\omega)\in\partial f(x(\omega),\omega)\ \mu\text{-a.e.}\}.$$

Proof. By Example 3.9, the subgradient condition $I_f(x)+I_f^*(v)\le\langle x,v\rangle$ can be written as

$$\int_\Omega f(x(\omega),\omega)d\mu(\omega)+\int_\Omega f^*(v(\omega),\omega)d\mu(\omega)\le\int_\Omega x(\omega)\cdot v(\omega)d\mu(\omega).$$

Since $f(x,\omega)+f^*(v,\omega) \geq x\cdot v$ for every $x,v\in\mathbb{R}^n$, the subgradient condition is equivalent to

$$f(x(\omega),\omega)+f^*(v(\omega),\omega)\leq x(\omega)\cdot v(\omega) \quad \mu\text{-a.e.}$$

which, in turn, means that $v(\omega)\in\partial f(x(\omega),\omega)$ almost everywhere. □

We have already seen that the subdifferential and the Legendre–Fenchel transform are closely related. This is further elaborated in the following.

Theorem 3.15. *If $\partial f(x)\neq\emptyset$, then $f^{**}(x)=f(x)$. On the other hand, if $f^{**}(x)=f(x)$, then $\partial f(x)=\partial f^{**}(x)$ and*

$$\partial f(x)=\{v\,|\,x\in\partial f^*(v)\}.$$

In particular, if $f=f^{**}$, then $\partial f^*=(\partial f)^{-1}$ and $\operatorname{argmin} f=\partial f^*(0)$.

Proof. The inequality $f^{**}\leq f$ is always valid, by Theorem 3.7. We have $v\in\partial f(x)$ if and only if $\langle x,v\rangle - f^*(v)\geq f(x)$ which implies

$$f^{**}(x)\geq f(x),$$

proving the first claim. The second follows from the expressions $\partial f^{**}(x)=\{v\,|\,f^{**}(x)+f^*(v)\leq\langle x,v\rangle\}$ and $\partial f^*(v)=\{x\,|\,f^{**}(x)+f^*(v)\leq\langle x,v\rangle\}$, where the former is valid since $f^{***}=f^*$, by Theorem 3.7. □

Recession functions can also be described in terms of conjugates.

Theorem 3.16. *Let f be a proper convex function. Then the recession function of f^* is the support function of* $\operatorname{dom} f$*. Conversely, if f is lower semicontinuous, then its recession function is the support function of* $\operatorname{dom} f^*$*.*

Proof. By ??, the recession function of a pointwise supremum of proper lsc convex functions equals the pointwise supremum of the recession functions of the individual functions. If $f(x)$ is finite, the recession function of $x\mapsto\langle x,v\rangle-f(x)$ is $x\mapsto\langle x,v\rangle$. The first claim then follows by writing the conjugate as

$$f^*(v)=\sup_{x\in\operatorname{dom} f}\{\langle x,v\rangle-f(x)\}.$$

The second claim follows by applying the first to f^*. □

Exercise 3.17. *Let h be a finite convex function on $\mathbb{R}^n$ and z an $\mathbb{R}^n$-valued random vector with full support. Use Exercise 1.52 to show that*

$$\sup_{\alpha>0}\frac{\ln E\exp[\alpha(z\cdot v-h(z))]}{\alpha}=h^*(v)$$

for every $v\in\mathbb{R}^n$.

3.2 Parametric optimization and saddle-point problems

For the remainder of this chapter, we will study Rockafellar's conjugate duality framework which addresses parametric optimization problems, minimax problems and the associated dual pairs of optimization problems. In minimax theory, central concepts are saddle-values and subdifferential characterizations of saddle-points which translate to the absence of a duality gap and generalized Karush–Kuhn–Tucker conditions for dual pairs of convex optimization problems. This theory is fundamental e.g. in Hamiltonian mechanics, partial differential equations, mathematical finance as well as many numerical optimization algorithms.

Let U and Y be vector spaces in separating duality. Let X be another vector space and let f be a proper convex function on $X \times U$ such that $f(x, \cdot)$ is lsc for every $x \in X$. Consider the parametric optimization problem

$$\text{minimize} \quad f(x,u) \quad \text{over} \quad x \in X \tag{3.1}$$

and denote the optimal value by

$$\varphi(u) = \inf_{x \in X} f(x,u).$$

By Theorem 1.31, φ is a convex function on U. If φ is proper and lower semicontinuous, Theorem 3.7 gives the dual representation

$$\varphi(u) = \sup_{y \in Y}\{\langle u, y\rangle - \varphi^*(y)\}.$$

This simple formula is behind many fundamental duality results e.g. in mathematical finance.

Even when studying a fixed optimization problem without obvious parameters, an appropriate parameterization often leads to duality relations which yield valuable information about the original problem. The optimization problems

$$\begin{aligned} &\text{minimize} \quad f(x,0) \quad &\text{over} \quad x \in X,\\ &\text{maximize} \quad -\varphi^*(y) \quad &\text{over} \quad y \in Y \end{aligned}$$

are called the *primal* and the *dual problem*, respectively, associated with f. The optimum values can be expressed as $\varphi(0)$ and $\varphi^{**}(0)$. The conjugacy correspondences of Section 3.1 translate to relations between the primal and the dual problems. In particular, by Theorem 3.7, $\varphi(0) \geq \varphi^{**}(0)$. If the inequality is strict, a *duality gap* is said to exist.

Theorem 3.18. *The optimal value and the optimal solutions of the dual problem are given by $\varphi^{**}(0)$ and $\partial\varphi^{**}(0)$, respectively. We have $y \in \partial\varphi(0)$ if and only if there is no duality gap and y solves the dual.*

Proof. The optimal value of the dual can be written as

$$\sup_y\{-\varphi^*(y)\} = \sup_y\{\langle 0, y\rangle - \varphi^*(y)\} = \varphi^{**}(0).$$

The rest follows from Theorem 3.15. □

The relations between the primal and the dual problems are thus largely determined by the local behavior of the value function φ at the origin. Indeed, by Theorem 3.7, the biconjugate φ^{**} in Theorem 3.18 equals $\operatorname{lsc}\varphi$ as soon as $\operatorname{lsc}\varphi$ is proper. Thus, if $\operatorname{lsc}\varphi$ is proper and $\varphi(0) = (\operatorname{lsc}\varphi)(0)$, then the primal and dual optimal values are equal. Combining Theorem 3.18 with Theorem 3.13 gives the following.

Corollary 3.19. *Assume that φ is finite and continuous at the origin. Then the optimal values of the primal and the dual are equal and the set of dual solutions is nonempty and $\sigma(Y, U)$-compact.*

In general, the exact calculation of $\varphi(u)$ even for one $u \in U$ is impossible so it may seem hopeless to get anything useful out of φ^* or φ^{**}. This is not necessarily so. The conjugate of the value function can be expressed as

$$\begin{aligned}\varphi^*(y) &= \sup_u\{\langle u, y\rangle - \varphi(u)\}\\ &= \sup_u \sup_x\{\langle u, y\rangle - f(x, u)\}\\ &= -\inf_x l(x, y),\end{aligned}$$

where

$$l(x, y) = \inf_{u\in U}\{f(x, u) - \langle u, y\rangle\}.$$

is the *Lagrangian* associated with f. The Lagrangian is an extended real-valued function on $X\times Y$, convex in x and concave in y. Minimization of the Lagrangian with respect to x may be considerably easier than the minimization of $f(\cdot, u)$.

Example 3.20 (Composite model)**.** *Let*

$$f(x, u) = \begin{cases} +\infty & \text{if } x \notin \operatorname{dom} F,\\ h(F(x) + u) & \text{if } x \in \operatorname{dom} F,\end{cases}$$

where F is a K-convex function from X to U and h is a proper convex function on U such that $u_1 \in \operatorname{rge} F,\ u_1 - u_2 \in K \implies h(u_1) \le h(u_2)$. The Lagrangian becomes

$$\begin{aligned} l(x, y) &= \begin{cases} +\infty & x \notin \operatorname{dom} F,\\ \inf_u\{h(F(x) + u) - \langle u, y\rangle\} & x \in \operatorname{dom} F\end{cases}\\ &= \begin{cases} +\infty & x \notin \operatorname{dom} F,\\ \langle F(x), y\rangle - h^*(y) & x \in \operatorname{dom} F.\end{cases}\end{aligned}$$

In many applications, the Lagrangian has stronger continuity properties (over $\operatorname{dom} F \times \operatorname{dom} h^*$*) than* f*. This is used by certain optimization algorithms.*

When h *is nonlinear (see Example 1.27), minimization of* $f(\cdot, u)$ *may be much more difficult than minimization of* $l(\cdot, y)$ *for a fixed* y*. In particular, when* $X = \prod_{i \in I} X_i$ *and* $F(x) = \sum_{i \in I} F_i(x_i)$ *for a finite index set* I*. The Lagrangian can now be written as*

$$l(x, y) = \begin{cases} \sum_{i \in I} \langle F_i(x_i), y \rangle - h^*(y) & \textit{if } x_i \in \operatorname{dom} F_i \ \forall i \in I, \\ +\infty & \textit{otherwise.} \end{cases}$$

In this case, the problem of minimizing the Lagrangian decomposes into minimizing $\langle F_i(x_i), y \rangle$ *over* $\operatorname{dom} F_i$ *separately for each* $i \in I$*. This decomposition technique can be seen as an instance of a more general "Lagrangian relaxation" technique, where certain complicating features of an optimization problem are replaced by penalty terms in the objective.*

Optimality conditions for the primal and the dual problems can be written conveniently in terms of the Lagrangian. Indeed, when f is proper and $f(x, \cdot)$ is lsc for every $x \in X$, Theorem 3.7 gives

$$f(x, u) = \sup_{y \in Y} \{l(x, y) + \langle u, y \rangle\}.$$

In particular, the primal and the dual objectives can then be expressed symmetrically by

$$f(x, 0) = \sup_{y \in Y} l(x, y),$$

and

$$-\varphi^*(y) = \inf_{x \in X} l(x, y).$$

The primal and the dual problems can be seen as dual-halfs of the Lagrangian saddle-point problem. A pair $(x, y) \in X \times Y$ is said to be a *saddle-point* of l if

$$l(x, y') \leq l(x, y) \leq l(x', y) \quad \forall x' \in X,\ y' \in Y.$$

It is clear that

$$\inf_x \sup_y l(x.y) \geq \sup_y l \inf_x l(x.y).$$

If this holds as an equality, a *saddle value* is said to exist. In fact, the above inequality is nothing but a restatement of the Fenchel inequality $\varphi(0) \geq -\varphi^{**}(0)$.

Theorem 3.18 gives sufficient conditions for the existence of a saddle-value for convex-concave saddle functions l which are closed in the first argument. Indeed, such functions are in one-to-one correspondence with convex functions f closed in the first argument. Writing the saddle-point condition as

$$f(x, 0) \leq l(x, y) \leq -\varphi^*(y)$$

we get the following.

Theorem 3.21. *Assume that f is lsc in u. Then a saddle-value of the Lagrangian exists if and only if φ is closed at the origin. A pair (x,y) is a* saddle-point *of the Lagrangian if and only if x solves the primal, y solves the dual and the optimal values are equal.*

The following is the source of more concrete optimality conditions to be derived later.

Corollary 3.22. *Assume that f is lsc in u, that there is no duality gap and that the dual optimum is attained. Then an x solves the primal if and only if there is a y such that (x,y) is a saddle-point of the Lagrangian.*

Although, far from being necessary, the following condition is often easy to check when dealing with the composite model.

Example 3.23 (Slater condition)**.** *Consider the composite model of Example 3.20 and assume that there is an $x \in \operatorname{dom} F$ such that h is bounded from above on a neighborhood of $F(x)$. Then φ is continuous at the origin.*

We will assume from now on that X is in separating duality with a vector space V. This allows us to write the saddle-point conditions of the previous section in differential form. The following result, originally due to Rockafellar, gives a far reaching generalization of the classical Karush–Kuhn–Tucker condition for the classical nonlinear programming model.

Theorem 3.24 (KKT-conditions)**.** *Assume that f is lsc in u, that there is no duality gap and that the dual optimum is attained (as happens e.g. when φ is bounded from above on a neighborhood of the origin). Then an x solves the primal if and only if there is a y such that*

$$0 \in \partial_x l(x,y) \quad \text{and} \quad 0 \in \partial_y[-l](x,y).$$

Note that $X \times V$ is in separating duality with $V \times Y$ and that the conjugate of f can be expressed as

$$\begin{aligned} f^*(v,y) &= \sup_{x\in X, u\in U}\{\langle x,v\rangle + \langle u,y\rangle - f(x,u)\} \\ &= \sup_{x\in X}\{\langle x,v\rangle + \sup_{u\in U}\{\langle u,y\rangle - f(x,u)\}\} \\ &= \sup_{x\in X}\{\langle x,v\rangle - l(x,y)\}. \end{aligned}$$

Theorem 3.25. *Assume that f is closed in u. The following are equivalent*

(a) $(v,y) \in \partial f(x,u)$,

(b) $v \in \partial_x l(x,y)$ *and* $u \in \partial_y[-l](x,y)$.

If f is closed, then the above are equivalent also to

(c) $(x,u) \in \partial f^*(v,y)$.

Proof. Part (a) means that

$$f(x',u') \geq f(x,u) + \langle x'-x, v\rangle + \langle u'-u, y\rangle \quad \forall x' \in X,\ u' \in U,$$

or equivalently,

$$l(x',y) \geq f(x,u) - \langle u,y\rangle + \langle x'-x, v\rangle \quad \forall x' \in X. \tag{3.2}$$

This implies $v \in \partial_x l(x,y)$. It is clear that (a) also implies $y \in \partial_u f(x,u)$. Since f is closed in u, this is equivalent to $u \in \partial_y[-l](x,y)$, by Theorem 3.15. Conversely, $u \in \partial_y[-l](x,y)$ means that $l(x,y) = f(x,u) - \langle u,y\rangle$, and then $v \in \partial_x l(x,y)$ is equivalent to (3.2). The last claim is proved by a symmetric argument. □

3.3 Calculus of subgradients and conjugates

In order to apply the KKT-conditions in practice, one needs rules for calculating the subdifferential of a given function. The following abstract result turns out to be quite useful in this respect. Its proof is a direct application of Theorem 3.24.

Corollary 3.26. *Assume that f is lsc in u and that for every $v \in V$ the function*

$$\varphi_v(u) = \inf_{x \in X}\{f(x,u) - \langle x,v\rangle\}$$

is bounded from above on a neighborhood of the origin. Then

$$\partial_x f(x,0) = \bigcup\{\partial_x l(x,y) \mid \exists y \in Y :\ 0 \in \partial_y[-l](x,y)\}$$

and

$$\sup_x\{\langle x,v\rangle - f(x,0)\} = \min_y f^*(v,y)$$

for every $v \in V$.

Proof. Applying Theorem 3.24 to the function $f_v(x,u) := f(x,u) - \langle x,v\rangle$, we see that an x minimizes $f_v(\cdot,0)$ if and only if there is a y such that

$$0 \in \partial_x l_v(x,y) \quad \text{and} \quad 0 \in \partial_y[-l_v](x,y),$$

where

$$l_v(x,y) = \inf_u\{f_v(x,u) - \langle u,y\rangle\} = l(x,y) - \langle x,v\rangle.$$

It now suffices to note that an x minimizes $f_v(\cdot,0)$ if and only if $v \in \partial_x f(x,0)$ while $\partial_x l_v(x,y) = \partial_x l(x,y) - v$ and $\partial_y[-l_v](x,y) = \partial_y[-l](x,y)$. □

Given a K-convex function F from X to U and a $y \in Y$, the composition of F and the linear function $u \mapsto \langle u,y\rangle$ will be denoted by $\langle y,F\rangle$. i.e.

$$\langle y,F\rangle(x) = \begin{cases} \langle F(x),y\rangle & \text{if } x \in \operatorname{dom} F,\\ +\infty & \text{otherwise.}\end{cases}$$

Corollary 3.27 (Subdifferential chain rule)**.** *Let F be a K-convex function from X to U, let h be lsc proper convex function on U such that*

$$u_1 \in \operatorname{rge} F,\ u_1 - u_2 \in K \implies h(u_1) \le h(u_2).$$

If there is an $x \in \operatorname{dom} F$ such that h bounded from above on a neighborhood of $F(x)$, then

$$\partial(h \circ F)(x) = \bigcup\{\partial\langle y, F\rangle(x) \mid y \in \partial h(F(x))\}$$

for every $x \in \operatorname{dom}(h \circ F)$ and

$$(h \circ F)^*(v) = \min_{y \in Y}\{\langle y, F\rangle^*(v) + h^*(y)\}.$$

Proof. We apply Corollary 3.26 to the function $f(x,u) = h(F(x) + u)$. The boundedness assumption on h implies that on φ_v. We have

$$\begin{aligned} l(x,y) &= \inf\{f(x,u) - \langle u, y\rangle\} \\ &= \begin{cases} +\infty & x \notin \operatorname{dom} F, \\ \langle F(x), y\rangle - h^*(y) & x \in \operatorname{dom} F. \end{cases} \end{aligned}$$

so that

$$\begin{aligned} \partial_x l(x,y) &= \partial\langle y, F\rangle(x), \\ \partial_y[-l](x,y) &= \partial h^*(y) - F(x). \end{aligned}$$

Since h is lsc and proper, the condition $0 \in \partial_y[-l](x,y)$ can be written as $y \in \partial h(F(x))$, by Theorem 3.15. □

The above chain rule can be written in a more convenient form when F is linear. Let A be a linear mapping from a linear subspace $\operatorname{dom} A$ of X to U. The *adjoint* of A is the (a priori set-valued) mapping from Y to V defined through its graph by

$$\operatorname{gph} A^* = \{(y,v) \mid \langle Ax, y\rangle = \langle x, v\rangle \quad \forall x \in \operatorname{dom} A\}.$$

In other words, we have $v \in A^*y$ if and only if the linear functional $x \mapsto \langle Ax, y\rangle$ on $\operatorname{dom} A$ can be represented by the continuous linear functional $x \mapsto \langle x, v\rangle$. The set $\operatorname{dom} A^* = \{y \in Y \mid A^*y \ne \emptyset\}$ is called the *domain* of the adjoint. If A^* is single-valued on $\operatorname{dom} A^*$, then

$$\langle Ax, y\rangle = \langle x, A^*y\rangle \quad \forall x \in \operatorname{dom} A,\ y \in \operatorname{dom} A^*$$

in accordance with the definition of the adjoint of a continuous linear mapping. Indeed, if A is continuous with $\operatorname{dom} A = X$, then the linear functional $x \mapsto \langle Ax, y\rangle$ is always continuous, so there is (since X and V are in separating duality) a unique $v \in V$ such that $\langle Ax, y\rangle = \langle x, v\rangle$. Thus, in the continuous case, $\operatorname{dom} A^* = Y$ and $\langle Ax, y\rangle = \langle x, A^*y\rangle$ holds for every $x \in X$ and $y \in Y$. For general linear mappings, we have the following.

Lemma 3.28. *We have $y \in \operatorname{dom} A^*$ if and only if the functional $x \mapsto \langle Ax, y\rangle$ is continuous on $\operatorname{dom} A$. The adjoint A^* is single-valued on $\operatorname{dom} A^*$ if and only if $\operatorname{dom} A$ is dense in X. If A has closed graph, then $A^{**} = A$ and $\operatorname{dom} A^*$ is dense in Y.*

Proof. If $x \mapsto \langle Ax, y\rangle$ is continuous on $\operatorname{dom} A$ then, by Corollary 2.37, it can be extended to a continuous linear functional on X. This yields the first claim while the second claim follows from Exercise 3.4. When $\operatorname{gph} A$ is closed, the bipolar theorem gives $A^{**} = A$ once we notice that $\operatorname{gph} A^* = \{(y, v) \mid (y, -v) \in (\operatorname{gph} A)^*\}$. The denseness of $\operatorname{dom} A^*$ then follows by applying the second claim to A^{**}. □

Example 3.29 (Differential operators)**.** *The differential operator $D : W_0^{m,p}(\Omega) \to L^p(\Omega)^n$ referred to at the end of Section 1.1 is linear and continuous. The space $L^p(\Omega)^n$ is in separating duality with $L^q(\Omega)^n$ under the usual bilinear form. The dual space of $W_0^{m,p}(\Omega)$ is denoted by $W^{-m,q}(\Omega)$. The adjoint of D can be expressed in terms of distributional derivatives as*

$$D^*y = \sum_{|\alpha|\le m} (-1)^{|\alpha|}\partial^\alpha y_\alpha.$$

Defining $\operatorname{dom} D = W_0^{m,p}(\Omega)$, we can view D also as a densely defined linear operator from $L^p(\Omega)$ to $L^p(\Omega)^n$. The adjoint D^* can still be expressed as above but now

$$\operatorname{dom} D^* = \{y \in L^q(\Omega)^n \mid \sum_{|\alpha|\le m} (-1)^{|\alpha|}\partial^\alpha y_\alpha \in L^q(\Omega)\},$$

which is only dense in $L^q(\Omega)^m$. Since the norm of a point $(x, u) \in \operatorname{gph} D$ is simply the $W^{m,p}(\Omega)$-norm of x, the graph of D is closed, by completeness of $W^{m,p}(\Omega)$.

For linear F, Theorem 3.27 can be written as follows.

Corollary 3.30. *Let A be a densely defined linear mapping from X to U and let h be a lsc proper convex function. If h is continuous at a point of $\operatorname{rge} A$, then*

$$\partial(h \circ A)(x) = A^*\partial h(Ax)$$

for $x \in \operatorname{dom}(h \circ A)$.

Proof. The function $F = A$ is K-convex for $K = \{0\}$ so the assumptions of Theorem 3.27 are satisfied. Thus, for $x \in \operatorname{dom}(h \circ A)$,

$$\partial(h \circ A)(x) = \bigcup\{\partial\langle y, A\rangle(x) \mid y \in \partial h(A(x))\}.$$

By Lemma 3.28, $x \mapsto \langle Ax, y\rangle$ is continuous on $\operatorname{dom} A$ if and only if $y \in \operatorname{dom} A^*$. Thus,

$$\partial\langle y, A\rangle(x) = \begin{cases} \emptyset & \text{for } y \notin \operatorname{dom} A^*, \\ A^*y & \text{for } y \in \operatorname{dom} A^*, \end{cases}$$

which yields the desired expression. □

Example 3.31 (Generalized Laplacians)**.** *The objective in the problem of calculus of variations* (CV) *in Section 1.1 fits the format of Corollary 3.30. If I_f is lsc proper and continuous somewhere on* $\operatorname{rge} D$*, then*

$$\partial(I_f \circ D) = D^* \partial I_f D.$$

Thus, by Example 3.14, $v \in \partial(I_f \circ D)(x)$ if and only if there is a $y \in L^q(\Omega)^n$ such that $y(\omega) \in \partial f(Dx(\omega), \omega)$ and $v = D^* y$. In particular, if $m = 1$ and $f(u, \omega) = \frac{1}{p} \sum_{|\alpha|=1} |u_\alpha|^p$, we get the so called p-Laplacian given by $\partial(I_f \circ D)(x) = -\operatorname{div}(|\nabla x|^{p-2} \nabla x)$.

When $m = 1$ we can write the differential operator as $Dx = (x, \nabla x)$. If the continuity condition holds, then an x minimizes $I_f(x, \nabla x)$ if and only if there exist a $y \in L^q(\Omega)^d$ such that

$$((\operatorname{div} y)(\omega), y(\omega)) \in \partial f(x(\omega), \nabla x(\omega), \omega).$$

In the one-dimensional case, this becomes the classical Euler–Lagrange condition in calculus of variations.

Corollary 3.32 (Subdifferential sum-rule)**.** *Let f_1 and f_2 be proper convex functions such that f_2 is continuous at a point of* $\operatorname{dom} f_1$*. If f_2 is lsc, then*

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x).$$

Proof. This is obtained by applying Corollary 3.26 to the function $f(x, u) = f_1(x) + f_2(x + u)$. The details are left as an exercise. □

For the infimal-projection, we have the following simple result whose proof does not require the use of Corollary 3.26 like the above results.

Theorem 3.33. *If the infimum $\varphi(u) = \inf_x f(x, u)$ is attained at x, then*

$$\partial\varphi(u) = \{y \mid (0, y) \in \partial f(x, u)\}.$$

Proof. If $f(x, u) = \varphi(u)$, then

$$\varphi(u') \geq \varphi(u) + \langle y, u' - u\rangle \quad \forall u' \in U$$

can be written as

$$f(x', u') \geq f(x, u) + \langle (0, y), (x', u') - (x, u)\rangle \quad \forall (x', u') \in X \times U,$$

which means that $(0, y) \in \partial f(x, u)$. □

Corollary 3.34. *Let f_i be convex functions such that* $\operatorname{int}\operatorname{dom} f_i \neq \emptyset$*. If the infimum in the definition of $f_1 \,\square \cdots \square\, f_n$ is attained at $(x_i)_{i=1}^n$, then*

$$\partial(f_1 \,\square \cdots \square\, f_n)(x) = \bigcap_{i=1}^n \partial f_i(x_i).$$

3.4 Parametric optimization and saddle-point problems: Part II

Equipped with the above rules for calculating subdifferentials, we now return to the KKT-conditions. The first example treats an instance of the composite model in Example 3.20. The special form is a generalization of the classical nonlinear programming model.

Example 3.35 (Composite model)**.** *Let F and h be as in Example 3.20 and let k be a convex function on X which is bounded from above on a neighborhood of an $x \in \operatorname{dom} F$. The Lagrangian corresponding to*

$$f(x,u) = \begin{cases} k(x) + h(F(x)+u) & \text{if } x \in \operatorname{dom} F, \\ +\infty & \text{otherwise} \end{cases}$$

can be written as

$$l(x,y) = \begin{cases} +\infty & \text{if } x \notin \operatorname{dom} k \cap \operatorname{dom} F, \\ k(x) + \langle F(x), y\rangle - h^*(y) & \text{otherwise.} \end{cases}$$

When $U = \mathbb{R}^m$ and $h = \delta_{\mathbb{R}^m_-}$, we recover the classical nonlinear programming model (NLP) *in Section 1.1. By Corollary 3.32, the KKT conditions can be written as*

$$\begin{aligned} \partial k(x) + \partial \langle y, F\rangle(x) &\ni 0, \\ \partial h^*(y) - F(x) &\ni 0. \end{aligned}$$

When $h = \delta_K$ for a closed convex cone, then by ??,

$$\partial h^*(y) - F(x) \ni 0 \iff F(x) \in K,\ y \in K^*,\ \langle F(x), y\rangle = 0.$$

If there is an $x \in \operatorname{dom} k \cap \operatorname{dom} F$ such that h is bounded from above on a neighborhood of $F(x)$ (the Slater condition holds), then an x solves the primal if and only if there is a y such that the KKT conditions are satisfied.

When F is linear, the dual problem can be written more explicitly.

Example 3.36 (Fenchel–Rockafellar model)**.** *Consider Example 3.20 in the case $F(x) = Ax$, where A is a linear mapping from a dense subset* $\operatorname{dom} A$ *of X to U and $K = \{0\}$ so that*

$$f(x,u) = \begin{cases} +\infty & \text{if } x \notin \operatorname{dom} A, \\ k(x) + h(Ax+u) & \text{if } x \in \operatorname{dom} A. \end{cases}$$

The Lagrangian becomes

$$l(x,y) = \begin{cases} +\infty & x \notin \operatorname{dom} A \cap \operatorname{dom} k, \\ k(x) + \langle Ax, y\rangle - h^*(y) & x \in \operatorname{dom} A \cap \operatorname{dom} k, \end{cases}$$

so the KKT-condition can be written as (see the proof of Corollary 3.30)

$$\partial k(x) + A^* y \ni 0,$$
$$\partial h^*(y) - Ax \ni 0.$$

Since k is bounded from above on an open set, the dual objective

$$-\varphi^*(y) = \inf_x \{k(x) + \langle Ax, y\rangle - h^*(y) \,|\, x \in \operatorname{dom} A \cap \operatorname{dom} k\}$$

equals $-\infty$ unless $x \mapsto \langle Ax, y\rangle$ is continuous on $\operatorname{dom} A$, or in other words, unless $y \in \operatorname{dom} A^*$ and, thus

$$-\varphi^*(y) = \begin{cases} \inf_x\{k(x) + \langle x, A^*y\rangle - h^*(y) \,|\, x \in \operatorname{dom} k\} & y \in \operatorname{dom} A^*, \\ -\infty & y \notin \operatorname{dom} A^*, \end{cases}$$
$$= \begin{cases} k^*(-A^*y) + h^*(y) & y \in \operatorname{dom} A^*, \\ -\infty & y \notin \operatorname{dom} A^*. \end{cases}$$

If there is an $x \in \operatorname{dom} k \cap \operatorname{dom} A$ such that h is continuous at Ax, then Theorem 3.19 implies that the optimal values of the two problems

$$\text{minimize} \quad k(x) + h(Ax) \quad \text{over} \quad x \in \operatorname{dom} A$$

and

$$\text{maximize} \quad -h^*(y) - k^*(-A^*y) \quad \text{over} \quad y \in \operatorname{dom} A^*$$

are equal and the set of optimal solutions to the latter is nonempty and $\sigma(Y, U)$-compact. In this situation, the KKT-conditions are necessary and sufficient for an x to solve the former problem.

Example 3.37 (Linear Programming)**.** *Consider the Fenchel–Rockafellar duality framework in the case where $h(u) = \delta(u \,|\, u + b \in K\}$ and $k(x) = \langle x, c\rangle$ for a given convex cone $K \subseteq U$, $b \in U$ and $c \in V$. In this case, the primal and dual problems can be written as*

$$\begin{aligned} &\text{minimize} && \langle x, c\rangle \quad \text{over} \quad x \in \operatorname{dom} A \\ &\text{subject to} && Ax + b \in K, \end{aligned}$$

and

$$\begin{aligned} &\text{maximize} && \langle b, y\rangle \quad \text{over} \quad y \in \operatorname{dom} A^* \\ &\text{subject to} && A^*y = c, \\ &&& y \in K^*, \end{aligned}$$

where K^* is the polar of K. The KKT-conditions can be written as

$$c + A^*y = 0,$$
$$N_{K^*}(y) - b - Ax \ni 0.$$

In the finite-dimensional case, where K is the negative orthant, the polar cone K^* is the positive orthant and we recover the classical LP-duality framework.

The Fenchel–Rockafellar model can be generalized as follows.

Example 3.38. *Let*

$$f(x,u) = \begin{cases} +\infty & \text{if } x \notin \operatorname{dom} A, \\ \tilde{f}(x, Ax+u) & \text{if } x \in \operatorname{dom} A. \end{cases}$$

The Lagrangian becomes

$$\begin{aligned} l(x,y) &= \inf_u \{\tilde{f}(x, Ax+u) - \langle u, y\rangle\} \\ &= \begin{cases} +\infty & \text{if } x \notin \operatorname{dom} A, \\ \tilde{l}(x,y) + \langle Ax, y\rangle & \text{if } x \in \operatorname{dom} A, \end{cases} \end{aligned}$$

where $\tilde{l}(x,y) = \inf_u\{\tilde{f}(x,u) - \langle u,y\rangle\}$*. If for each* y*, the function* $\tilde{l}(\cdot,y)$ *is bounded from above on an open set, then*

$$-\varphi^*(y) = \begin{cases} -\tilde{f}^*(-A^*y, y) & \text{if } y \in \operatorname{dom} A^*, \\ -\infty & \text{if } y \notin \operatorname{dom} A^* \end{cases}$$

and the KKT conditions can be written as

$$\begin{aligned} \partial_x \tilde{l}(x,y) + A^*y &\ni 0, \\ \partial_y[-\tilde{l}](x,y) + Ax &\ni 0. \end{aligned}$$

Example 3.39 (Calculus of variations)**.** *Consider the problem of calculus of variations* (CV) *in Section 1.1. When* $m = 1$*,* $d = 1$ *and* $\Omega = [0,T]$*, we can write the problem as*

$$\text{minimize} \quad I_f(x,\dot{x}) \quad \text{over} \quad x \in W^{1,p}([0,T])$$

This fits the format of Example 3.38 with $X = U = L^p([0,T])$*,* $A = \nabla$*,* $\operatorname{dom} A = W^{1,p}([0,T])$ *and* $\tilde{f}(x,u) = I_f(x,u)$*. We have* $\operatorname{dom}\nabla^* = W^{1,q}([0,T])$ *and* $\nabla^* y = -\dot{y}$*, so if the function* $\tilde{f}(\cdot,y)$ *defined by*

$$\begin{aligned} \tilde{l}(x,y) &= \inf_{u \in L^p([0,T])} \{I_f(x,u) - \langle u,y\rangle\} \\ &= \inf_{u \in L^p([0,T])} \int_{[0,T]} \{f(x(t), u(t), t) - u(t)\cdot y(t)\} d\mu(t) \end{aligned}$$

is bounded from above on an open set, then

$$-\varphi^*(y) = \begin{cases} -I_{f^*}(\dot{y}, y) & \text{if } y \in W^{1,p}([0,T]), \\ -\infty & \text{if } y \notin W^{1,p}([0,T]), \end{cases}$$

where we have used that fact that $I_f^* = I_{f^*}$*, by Example 3.9. The dual problem can then be written as*

$$\text{maximize} \quad -I_{f^*}(\dot{y}, y) \quad \text{over} \quad y \in W^{1,q}([0,T]),$$

which is of the same for as the primal problem.

Bibliography

[1] H. Attouch, G. Buttazzo, and G. Michaille. *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*, volume 6 of *MPS/SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.

[2] Dimitri P. Bertsekas. *Linear network optimization: algorithms and codes.* MIT Press, Cambridge, MA, 1991.

[3] George B. Dantzig. Maximization of a linear function of variables subject to linear inequalities. In *Activity Analysis of Production and Allocation*, Cowles Commission Monograph No. 13, pages 339–347. John Wiley & Sons Inc., New York, N. Y., 1951.

[4] N. Dunford and J. T. Schwartz. *Linear operators. Part I.* Wiley Classics Library. John Wiley & Sons Inc., New York, 1988. General theory, With the assistance of William G. Bade and R. G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

[5] G. B. Folland. *Real analysis.* Pure and Applied Mathematics. John Wiley & Sons Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[6] L. V. Kantorovich. Mathematical methods of organizing and planning production. *Management Sci.*, 6:366–422, 1959/1960.

[7] J.-J. Moreau. *Fonctionelles convexes.* Séminaire sur les Equations aux Dérivées Partielles. Collège de France, Paris, 1967.

[8] T. Pennanen. Convex duality in stochastic optimization and mathematical finance. *Mathematics of Operations Research*, 36(2):340–362, 2011.

[9] Svetlozar T. Rachev and Ludger Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[10] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. Math.*, 24:525–539, 1968.

[11] R. T. Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[12] R. T. Rockafellar. Existence and duality theorems for convex problems of Bolza. *Trans. Amer. Math. Soc.*, 159:1–40, 1971.

[13] R. T. Rockafellar. *Conjugate duality and optimization.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1974.

[14] R. T. Rockafellar. Integral functionals, normal integrands and measurable selections. In *Nonlinear operators and the calculus of variations (Summer School, Univ. Libre Bruxelles, Brussels, 1975)*, pages 157–207. Lecture Notes in Math., Vol. 543. Springer, Berlin, 1976.

[15] R. T. Rockafellar. *Network flows and monotropic optimization.* Pure and Applied Mathematics (New York). John Wiley & Sons Inc., New York, 1984. A Wiley-Interscience Publication.

[16] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1998.