3

Basic Concepts of Convex Optimization

Contents

In this chapter, we introduce some basic concepts of convex optimization and minimax theory, with a special focus on the question of existence of optimal solutions.

3.1 CONSTRAINED OPTIMIZATION

Let us consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \end{array}$$

where $f : \Re^n \mapsto (-\infty, \infty]$ is a function and X is a nonempty subset of $\Re^n$. Any vector $x \in X \cap \text{dom}(f)$ is said to be a *feasible solution* of the problem (we also use the terms *feasible vector* or *feasible point*). If there is at least one feasible solution, i.e., $X \cap \text{dom}(f) \neq \emptyset$, we say that the problem is *feasible*; otherwise we say that the problem is *infeasible*. Thus, when f is extended real-valued, we view only the points in $X \cap \text{dom}(f)$ as candidates for optimality, and we view $\text{dom}(f)$ as an implicit constraint set. Furthermore, feasibility of the problem is equivalent to $\inf_{x \in X} f(x) < \infty$.

We say that a vector x^* is a *minimum of f over X* if

$$x^* \in X \cap \text{dom}(f), \qquad \text{and} \qquad f(x^*) = \inf_{x \in X} f(x).$$

We also call x^* a *minimizing point* or *minimizer* or *global minimum of f over X*. Alternatively, we say that f *attains a minimum over X at x^**, and we indicate this by writing

$$x^* \in \arg\min_{x \in X} f(x).$$

If x^* is known to be the unique minimizer of f over X, with slight abuse of notation, we also write

$$x^* = \arg\min_{x \in X} f(x).$$

We use similar terminology for maxima, i.e., a vector $x^* \in X$ such that $f(x^*) = \sup_{x \in X} f(x)$ is said to be a *maximum of f over X* if x^* is a minimum of $(-f)$ over X, and we indicate this by writing

$$x^* \in \arg\max_{x \in X} f(x).$$

If $X = \Re^n$ or if the domain of f is the set X (instead of $\Re^n$), we also call x^* a (global) minimum or (global) maximum of f (without the qualifier "over X").

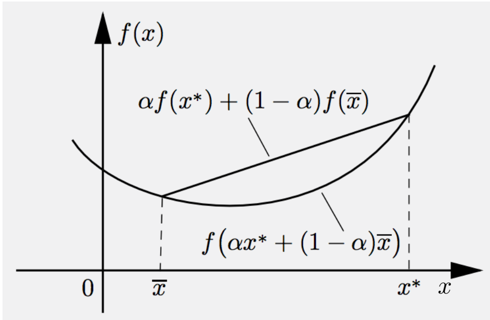

Figure 3.1.1. Illustration of why local minima of convex functions are also global (cf. Prop. 3.1.1). Given x^* and $\overline{x}$ with $f(\overline{x}) < f(x^*)$, every point of the form

$$x_\alpha = \alpha x^* + (1-\alpha)\overline{x}, \qquad \alpha \in (0,1),$$

satisfies $f(x_\alpha) < f(x^*)$. Thus x^* cannot be a local minimum that is not global.

Local Minima

Often in optimization problems we have to deal with a weaker form of minimum, one that is optimum only when compared with points that are "nearby." In particular, given a subset X of $\Re^n$ and a function $f : \Re^n \mapsto (-\infty, \infty]$, we say that a vector x^* is a *local minimum of f over X* if $x^* \in X \cap \text{dom}(f)$ and there exists some $\epsilon > 0$ such that

$$f(x^*) \leq f(x), \qquad \forall\, x \in X \text{ with } \|x - x^*\| < \epsilon.$$

If $X = \Re^n$ or the domain of f is the set X (instead of $\Re^n$), we also call x^* a local minimum of f (without the qualifier "over X"). A local minimum x^* is said to be *strict* if there is no other local minimum within some open sphere centered at x^*. Local maxima are defined similarly.

In practical applications we are typically interested in global minima, yet we have to contend with local minima because of the inability of many optimality conditions and algorithms to distinguish between the two types of minima. This can be a major practical difficulty, but an important implication of convexity of f and X is that all local minima are also global, as shown in the following proposition and in Fig. 3.1.1.

Proposition 3.1.1: If X is a convex subset of $\Re^n$ and $f : \Re^n \mapsto (-\infty, \infty]$ is a convex function, then a local minimum of f over X is also a global minimum. If in addition f is strictly convex, then there exists at most one global minimum of f over X.

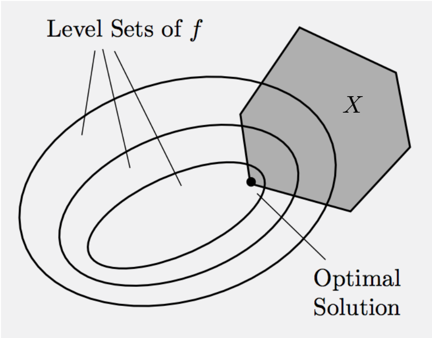

Figure 3.2.1. View of the set of optimal solutions of the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \end{array}$$

as the intersection of all the nonempty level sets of the form

$$\big\{x \in X \mid f(x) \leq \gamma\big\}, \quad \gamma \in \Re.$$

Proof: Let f be convex, and assume to arrive at a contradiction, that x^* is a local minimum of f over X that is not global (see Fig. 3.1.1). Then, there must exist an $\overline{x} \in X$ such that $f(\overline{x}) < f(x^*)$. By convexity, for all $\alpha \in (0,1)$,

$$f\big(\alpha x^* + (1-\alpha)\overline{x}\big) \leq \alpha f(x^*) + (1-\alpha) f(\overline{x}) < f(x^*).$$

Thus, f has strictly lower value than $f(x^*)$ at every point on the line segment connecting x^* with $\overline{x}$, except at x^*. Since X is convex, the line segment belongs to X, thereby contradicting the local minimality of x^*.

Let f be strictly convex, let x^* be a global minimum of f over X, and let x be a point in X with $x \neq x^*$. Then the midpoint $y = (x+x^*)/2$ belongs to X since X is convex, and by strict convexity, $f(y) < 1/2\big(f(x) + f(x^*)\big)$, while by the optimality of x^*, we have $f(x^*) \leq f(y)$. These two relations imply that $f(x^*) < f(x)$, so x^* is the unique global minimum. **Q.E.D.**

3.2 EXISTENCE OF OPTIMAL SOLUTIONS

A basic question in optimization problems is whether an optimal solution exists. It can be seen that the set of minima of a real-valued function f over a nonempty set X, call it X^*, is equal to the intersection of X and the level sets of f that have a common point with X:

$$X^* = \cap_{k=0}^{\infty} \big\{x \in X \mid f(x) \leq \gamma_k\big\},$$

where $\{\gamma_k\}$ is any scalar sequence with $\gamma_k \downarrow \inf_{x \in X} f(x)$ (see Fig. 3.2.1).

From this characterization of X^*, it follows that the set of minima is nonempty and compact if the sets

$$\big\{x \in X \mid f(x) \leq \gamma\big\},$$

are compact (since the intersection of a nested sequence of nonempty and compact sets is nonempty and compact). This is the essence of the classical theorem of Weierstrass (Prop. A.2.7), which states that a continuous function attains a minimum over a compact set. We will provide a more general version of this theorem, and to this end, we introduce some terminology.

We say that a function $f : \Re^n \mapsto (-\infty, \infty]$ is *coercive* if for every sequence $\{x_k\}$ such that $\|x_k\| \to \infty$, we have $\lim_{k\to\infty} f(x_k) = \infty$. Note that as a consequence of the definition, if $\text{dom}(f)$ is bounded, then f is coercive. Furthermore, all the nonempty level sets of a coercive function are bounded.

Proposition 3.2.1: (Weierstrass' Theorem) Consider a closed proper function $f : \Re^n \mapsto (-\infty, \infty]$, and assume that any one of the following three conditions holds:

(1) $\text{dom}(f)$ is bounded.

(2) There exists a scalar $\overline{\gamma}$ such that the level set

$$\big\{x \mid f(x) \leq \overline{\gamma}\big\}$$

is nonempty and bounded.

(3) f is coercive.

Then the set of minima of f over $\Re^n$ is nonempty and compact.

Proof: It is sufficient to show that each of the three conditions implies that the nonempty level sets $V_\gamma = \big\{x \mid f(x) \leq \gamma\big\}$ are compact for all $\gamma \leq \overline{\gamma}$, where $\overline{\gamma}$ is such that $V_{\overline{\gamma}}$ is nonempty and compact, and then use the fact that the set of minima of f is the intersection of its nonempty level sets. (Note that f is assumed proper, so it has some nonempty level sets.) Since f is closed, its level sets are closed (cf. Prop. 1.1.2). It is evident that under each of the three conditions the level sets are also bounded for γ less or equal to some $\overline{\gamma}$, so they are compact. **Q.E.D.**

The most common application of Weierstrass' Theorem is when we want to minimize a real-valued function $f : \Re^n \mapsto \Re$ over a nonempty set X. Then, by applying the proposition to the extended real-valued function

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

we see that the set of minima of f over X is nonempty and compact if X is closed, f is lower semicontinuous at each $x \in X$ (implying, by Prop. 1.1.3, that $\tilde{f}$ is closed), and one of the following conditions holds:

(1) X is bounded.

(2) Some set $\{x \in X \mid f(x) \leq \overline{\gamma}\}$ is nonempty and bounded.

(3) $\tilde{f}$ is coercive, or equivalently, for every sequence $\{x_k\} \subset X$ such that $\|x_k\| \to \infty$, we have $\lim_{k\to\infty} f(x_k) = \infty$.

The following is essentially Weierstrass' Theorem specialized to convex functions.

Proposition 3.2.2: Let X be a closed convex subset of $\Re^n$, and let $f : \Re^n \mapsto (-\infty, \infty]$ be a closed convex function with $X \cap \mathrm{dom}(f) \neq \emptyset$. The set of minima of f over X is nonempty and compact if and only if X and f have no common nonzero direction of recession.

Proof: Let $f^* = \inf_{x\in X} f(x)$ and note that $f^* < \infty$ since $X \cap \mathrm{dom}(f) \neq \emptyset$. Let $\{\gamma_k\}$ be a scalar sequence with $\gamma_k \downarrow f^*$, and consider the sets

$$V_k = \big\{x \mid f(x) \leq \gamma_k\big\}.$$

Then the set of minima of f over X is

$$X^* = \cap_{k=1}^{\infty} (X \cap V_k).$$

The sets $X \cap V_k$ are nonempty and have $R_X \cap R_f$ as their common recession cone, which is also the recession cone of X^*, when $X^* \neq \emptyset$ [cf. Props. 1.4.5, 1.4.2(c)]. It follows using Prop. 1.4.2(a) that X^* is nonempty and compact if and only if $R_X \cap R_f = \{0\}$. **Q.E.D.**

If X and f of the above proposition have a common direction of recession, then either the optimal solution set is empty [take for example, $X = \Re$ and $f(x) = e^x$] or else it is nonempty and unbounded [take for example, $X = \Re$ and $f(x) = \max\{0, x\}$]. Here is another result that addresses an important special case where the set of minima is compact.

Proposition 3.2.3: (Existence of Solution, Sum of Functions) Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, be closed proper convex functions such that the function $f = f_1 + \cdots + f_m$ is proper. Assume that the recession function of a single function f_i satisfies $r_{f_i}(d) = \infty$ for all $d \neq 0$. Then the set of minima of f is nonempty and compact.

Proof: By Prop. 3.2.2, the set of minima of f is nonempty and compact if and only if $R_f = \{0\}$, which by Prop. 1.4.6, is true if and only if $r_f(d) > 0$ for all $d \neq 0$. The result now follows from Prop. 1.4.8. **Q.E.D.**

As an example of application of the preceding proposition, if one of the functions f_i is a positive definite quadratic function, the set of minima of the sum f is nonempty and compact. In fact in this case f has a unique minimum because the positive definite quadratic is strictly convex, which makes f strictly convex.

The next proposition addresses cases where the optimal solution set is not compact.

Proposition 3.2.4: (Existence of Solution, Noncompact Level Sets) Let X be a closed convex subset of $\Re^n$, and let $f : \Re^n \mapsto (-\infty, \infty]$ be a closed convex function with $X \cap \text{dom}(f) \neq \emptyset$. The set of minima of f over X, denoted X^*, is nonempty under any one of the following two conditions:

(1) $R_X \cap R_f = L_X \cap L_f$.

(2) $R_X \cap R_f \subset L_f$ and X is a polyhedral set.

Furthermore, under condition (1),

$$X^* = \tilde{X} + (L_X \cap L_f),$$

where $\tilde{X}$ is some nonempty and compact set.

Proof: Let $f^* = \inf_{x \in X} f(x)$ and note that $f^* < \infty$ since $X \cap \text{dom}(f) \neq \emptyset$. Let $\{\gamma_k\}$ be a scalar sequence with $\gamma_k \downarrow f^*$, consider the level sets

$$V_k = \big\{x \mid f(x) \leq \gamma_k\big\},$$

and note that

$$X^* = \cap_{k=1}^{\infty}(X \cap V_k).$$

Let condition (1) hold. The sets $X \cap V_k$ are nonempty, closed, convex, and nested. Furthermore, they have the same recession cone, $R_X \cap R_f$, and the same lineality space $L_X \cap L_f$, while by assumption, $R_X \cap R_f = L_X \cap L_f$. By Prop. 1.4.11(a), it follows that X^* is nonempty and has the form

$$X^* = \tilde{X} + (L_X \cap L_f),$$

where $\tilde{X}$ is some nonempty compact set.

Let condition (2) hold. The sets V_k are nested and $X \cap V_k$ is nonempty for all k. Furthermore, all the sets V_k have the same recession cone, R_f, and the same lineality space, L_f, while by assumption, $R_X \cap R_f \subset L_f$, and X is polyhedral and hence retractive (cf. Prop. 1.4.9). By Prop. 1.4.11(b), it follows that X^* is nonempty. **Q.E.D.**

Note that in the special case $X = \Re^n$, conditions (1) and (2) of Prop. 3.2.4 coincide. Figure 3.2.3(b) provides a counterexample showing that if X is nonpolyhedral, the condition

$$R_X \cap R_f \subset L_f$$

is not sufficient to guarantee the existence of optimal solutions or even the finiteness of f^*. This counterexample also shows that the cost function may be bounded below and attain a minimum over any closed halfline contained in the constraint set, and yet it may not attain a minimum over the entire set. Recall, however, that in the special cases of linear and quadratic programming problems, boundedness from below of the cost function over the constraint set guarantees the existence of an optimal solution (cf. Prop. 1.4.12).

3.3 PARTIAL MINIMIZATION OF CONVEX FUNCTIONS

In our development of duality and minimax theory we will often encounter functions obtained by minimizing other functions partially, i.e., with respect to some of their variables. It is then useful to be able to deduce properties of the function obtained, such as convexity and closedness, from corresponding properties of the original.

There is an important geometric relation between the epigraph of a given function and the epigraph of its partially minimized version: except for some boundary points, *the latter is obtained by projection from the former* [see part (b) of the following proposition, and Fig. 3.3.1]. This is the key to understanding the properties of partially minimized functions.

Proposition 3.3.1: Consider a function $F : \Re^{n+m} \mapsto (-\infty, \infty]$ and the function $f : \Re^n \mapsto [-\infty, \infty]$ defined by

$$f(x) = \inf_{z \in \Re^m} F(x, z).$$

Then:

(a) If F is convex, then f is also convex.

(b) We have

$$P\big(\mathrm{epi}(F)\big) \subset \mathrm{epi}(f) \subset \mathrm{cl}\Big(P\big(\mathrm{epi}(F)\big)\Big), \tag{3.1}$$

where $P(\cdot)$ denotes projection on the space of (x, w), i.e., for any subset S of $\Re^{n+m+1}$, $P(S) = \big\{(x, w) \mid (x, z, w) \in S\big\}$.

Proof: (a) If $\mathrm{epi}(f) = \varnothing$, i.e., $f(x) = \infty$ for all $x \in \Re^n$, then $\mathrm{epi}(f)$ is convex, so f is convex. Assume that $\mathrm{epi}(f) \neq \varnothing$, and let $(\overline{x}, \overline{w})$ and $(\tilde{x}, \tilde{w})$

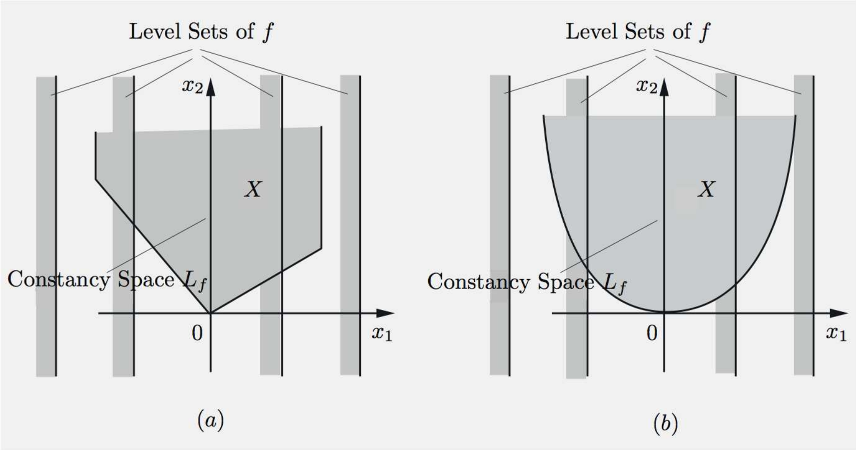

Figure 3.2.3. Illustration of the issues regarding existence of an optimal solution assuming $R_X \cap R_f \subset L_f$, i.e., that every common direction of recession of X and f is a direction in which f is constant [cf. Prop. 3.2.4 under condition (2)].

In both problems illustrated in (a) and (b) the cost function is

$$f(x_1, x_2) = e^{x_1}.$$

In the problem of (a), the constraint set X is the polyhedral set shown in the figure, while in the problem of (b), X is specified by a quadratic inequality:

$$X = \big\{(x_1, x_2) \mid x_1^2 \le x_2\big\},$$

as shown in the figure. In both cases we have

$$R_X = \big\{(d_1, d_2) \mid d_1 = 0,\ d_2 \ge 0\big\},$$

$$R_f = \big\{(d_1, d_2) \mid d_1 \le 0,\ d_2 \in \Re\big\}, \qquad L_f = \big\{(d_1, d_2) \mid d_1 = 0,\ d_2 \in \Re\big\},$$

so that $R_X \cap R_f \subset L_f$.

In the problem of (a) it can be seen that an optimal solution exists. In the problem of (b), however, we have $f(x_1, x_2) > 0$ for all (x_1, x_2), while for $x_1 = -\sqrt{x_2}$ where $x_2 \ge 0$, we have $(x_1, x_2) \in X$ with

$$\lim_{x_2 \to \infty} f\big(-\sqrt{x_2}, x_2\big) = \lim_{x_2 \to \infty} e^{-\sqrt{x_2}} = 0,$$

implying that $f^* = 0$. Thus f cannot attain the minimum value f^* over X. Note that f attains a minimum over the intersection of any line with X.

If in the problem of (b) the cost function were instead $f(x_1, x_2) = x_1$, we would still have $R_X \cap R_f \subset L_f$ and f would still attain a minimum over the intersection of any line with X, but it can be seen that $f^* = -\infty$. If the constraint set were instead $X = \big\{(x_1, x_2) \mid |x_1| \le x_2\big\}$, which is polyhedral, we would still have $f^* = -\infty$, but then the condition $R_X \cap R_f \subset L_f$ would be violated.

be two points in $\text{epi}(f)$. Then $f(\overline{x}) < \infty$, $f(\tilde{x}) < \infty$, and there exist sequences $\{\overline{z}_k\}$ and $\{\tilde{z}_k\}$ such that

$$F(\overline{x}, \overline{z}_k) \to f(\overline{x}), \qquad F(\tilde{x}, \tilde{z}_k) \to f(\tilde{x}).$$

Using the definition of f and the convexity of F, we have for all $\alpha \in [0, 1]$ and k,

$$\begin{aligned} f\big(\alpha\overline{x} + (1-\alpha)\tilde{x}\big) &\le F\big(\alpha\overline{x} + (1-\alpha)\tilde{x}, \alpha\overline{z}_k + (1-\alpha)\tilde{z}_k\big) \\ &\le \alpha F(\overline{x}, \overline{z}_k) + (1-\alpha)F(\tilde{x}, \tilde{z}_k). \end{aligned}$$

By taking the limit as $k \to \infty$, we obtain

$$f\big(\alpha\overline{x} + (1-\alpha)\tilde{x}\big) \le \alpha f(\overline{x}) + (1-\alpha)f(\tilde{x}) \le \alpha\overline{w} + (1-\alpha)\tilde{w}.$$

It follows that the point $\alpha(\overline{x}, \overline{w}) + (1-\alpha)(\tilde{x}, \tilde{w})$ belongs to $\text{epi}(f)$. Thus $\text{epi}(f)$ is convex, implying that f is convex.

(b) To show the left-hand side of Eq. (3.1), let $(x, w) \in P\big(\text{epi}(F)\big)$, so that there exists $\overline{z}$ such that $(x, \overline{z}, w) \in \text{epi}(F)$, or equivalently $F(x, \overline{z}) \le w$. Then

$$f(x) = \inf_{z \in \Re^m} F(x, z) \le w,$$

implying that $(x, w) \in \text{epi}(f)$.

To show the right-hand side, note that for any $(x, w) \in \text{epi}(f)$ and every k, there exists a z_k such that

$$(x, z_k, w + 1/k) \in \text{epi}(F),$$

so that $(x, w + 1/k) \in P\big(\text{epi}(F)\big)$ and $(x, w) \in \text{cl}\Big(P\big(\text{epi}(F)\big)\Big)$. **Q.E.D.**

Among other things, part (b) of the preceding proposition asserts that if F is closed, and if the projection operation preserves closedness of its epigraph, then partial minimization of F yields a closed function. Note also a connection between closedness of $P\big(\text{epi}(F)\big)$ and the attainment of the infimum of $F(x, z)$ over z. As illustrated in Fig. 3.3.1, for a fixed x, $F(x, z)$ attains a minimum over z if and only if $\big(x, f(x)\big)$ belongs to $P\big(\text{epi}(F)\big)$. Thus if $P\big(\text{epi}(F)\big)$ is closed, $F(x, z)$ attains a minimum over z for all x such that $f(x)$ is finite.

We now provide criteria guaranteeing that closedness is preserved under partial minimization, while simultaneously the partial minimum is attained.

Proposition 3.3.2: Let $F : \Re^{n+m} \mapsto (-\infty, \infty]$ be a closed proper convex function, and consider the function f given by

$$f(x) = \inf_{z \in \Re^m} F(x, z), \qquad x \in \Re^n.$$