Tutorial notes

2024-2025 Optimization theory

These notes are organized according to Section 4.1 and 4.2 of Andréasson et al. [2020].

1 Local and global optimality

Consider the problem to

$$\text{minimize } f(x), \tag{1}$$

$$\text{subject to } x \in S, \tag{2}$$

where $S \subseteq \mathbb{R}^n$ is a nonempty set and $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a given function.

Consider the function given in Figure 1.

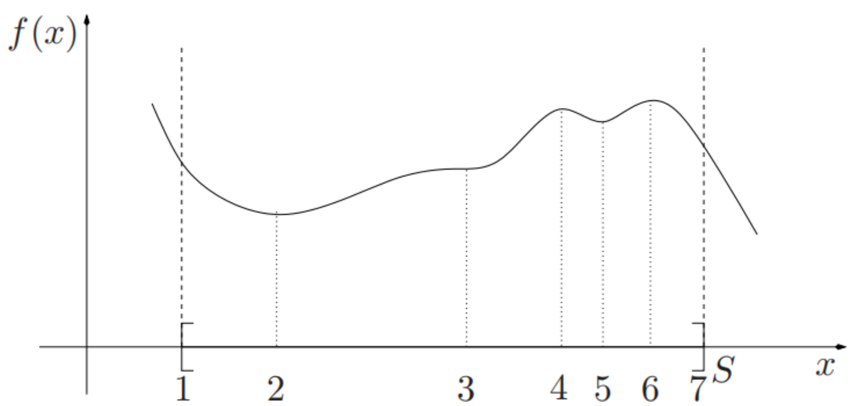

Figure 1: A one-dimensional function and its possible extremal points

For a minimization problem of f in one variable over a closed interval S, the interesting points are:

1. boundary points of S;
2. stationary points, that is, where $f'(x) = 0$;
3. discontinuities in f or f'.

In the case of the function in Figure 1, the points 1 and 7 are of category (i); 2, 3, 4, 5, and 6 of category (ii); and none of category (iii).

Definition 1 (Global Minimum)**.** *Consider the problem 1. Let $x^* \in S$. We say that x^* is a global minimum of f over S if f attains its lowest value over S at x^*.*

In other words, $x^* \in S$ is a global minimum of f over S if

$$f(x^*) \leq f(x), \quad \forall x \in S. \tag{3}$$

Let $B_\varepsilon(x^*) := \{y \in \mathbb{R}^n \mid \|y - x^*\| < \varepsilon\}$ be the open Euclidean ball with radius ε centered at x^*.

Definition 2 (Local Minimum)**.** *Consider the problem 1. Let $x^* \in S$.*

(a) We say that x^* is a local minimum of f over S if there exists a small enough ball intersected with S around x^* such that x^* is a globally optimal solution in that smaller set.

In other words, $x^* \in S$ is a local minimum of f over S if

$$\exists \varepsilon > 0 \text{ such that } f(x^*) \leq f(x), \quad \forall x \in S \cap B_\varepsilon(x^*). \tag{4}$$

(b) We say that $x^* \in S$ is a strict local minimum of f over S if, in (4), the inequality holds strictly for $x \neq x^*$.

Remark 1. *Note that a global minimum in particular is a local minimum. When is a local minimum a global one? This question is resolved in the case of convex problems, as the following fundamental theorem shows.*

Theorem 1 (Fundamental Theorem of Global Optimality)**.** *Consider the problem 1, where S is a convex set and f is convex on S. Then, every local minimum of f over S is also a global minimum.*

Proof. Suppose that x^* is a local minimum but not a global one, while $\bar{x}$ is a global minimum. Then, $f(\bar{x}) < f(x^*)$. Let $\lambda \in (0, 1)$.

By the convexity of S and f, it follows that

$$\lambda\bar{x} + (1 - \lambda)x^* \in S, \quad \text{and} \quad f\big(\lambda\bar{x} + (1 - \lambda)x^*\big) \leq \lambda f(\bar{x}) + (1 - \lambda)f(x^*). \tag{5}$$

Since $f(\bar{x}) < f(x^*)$, we have

$$\lambda f(\bar{x}) + (1 - \lambda)f(x^*) < f(x^*). \tag{6}$$

Choosing $\lambda > 0$ small enough then leads to

$$f\big(\lambda\bar{x} + (1 - \lambda)x^*\big) < f(x^*), \tag{7}$$

which contradicts the local optimality of x^*. Therefore, x^* must also be a global minimum. $\square$

Remark 2. *There is an intuitive image that can be seen from the proof design: If x^* is a local minimum, then f cannot "go down-hill" from x^* in any direction, but if $\bar{x}$ has a lower value, then f has to go down-hill sooner or later. No convex function can have this shape.*

2 Existence of optimal solutions

2.1 A classic result

Definition 3 (Weakly Coercive and Coercive Functions)**.** *Let $S \subseteq \mathbb{R}^n$ be a nonempty and closed set, and $f : S \to \mathbb{R}$ be a given function.*

(a) We say that f is weakly coercive with respect to the set S if S is bounded or for every $N > 0$ there exists an $M > 0$ such that

$$f(x) \geq N \quad \text{whenever} \quad \|x\| \geq M. \tag{8}$$

In other words, f is weakly coercive if either S is bounded or

$$\lim_{\|x\|\to\infty,\, x\in S} f(x) = \infty \tag{9}$$

holds.

(b) We say that f is coercive with respect to the set S if S is bounded or for every $N > 0$ there exists an $M > 0$ such that

$$\frac{f(x)}{\|x\|} \geq N \quad \text{whenever} \quad \|x\| \geq M. \tag{10}$$

In other words, f is coercive if either S is bounded or

$$\lim_{\|x\|\to\infty,\, x\in S} \frac{f(x)}{\|x\|} = \infty \tag{11}$$

holds.

The weak coercivity of $f : S \to \mathbb{R}$ (for nonempty closed sets S) is equivalent to the property that f has bounded level sets restricted to S.

A coercive function grows faster than any linear function. In fact, for convex functions f, f being coercive is equivalent to $x \mapsto f(x) - a^T x$ being weakly coercive for every vector $a \in \mathbb{R}^n$. This property is a very useful one for certain analyses in the context of Lagrangian duality.

We next introduce two extended notions of continuity.

Definition 4 (Semi-Continuity)**.** *Consider a function $f : S \to \mathbb{R}$, where $S \subseteq \mathbb{R}^n$ is nonempty.*

(a) The function f is said to be lower semi-continuous at $\bar{x} \in S$ if the value $f(\bar{x})$ is less than or equal to every limit of f as $x_k \to \bar{x}$. That is,

$$\lim_{k\to\infty} x_k = \bar{x} \implies f(\bar{x}) \leq \liminf_{k\to\infty} f(x_k).$$

In other words, f is lower semi-continuous at $\bar{x} \in S$ if

$$x_k \to \bar{x} \implies f(\bar{x}) \leq \liminf_{k\to\infty} f(x_k). \tag{12}$$

(b) *The function f is said to be upper semi-continuous at $\bar{x} \in S$ if the value $f(\bar{x})$ is greater than or equal to every limit of f as $x_k \to \bar{x}$. That is,*

$$\lim_{k\to\infty} x_k = \bar{x} \implies f(\bar{x}) \geq \limsup_{k\to\infty} f(x_k).$$

In other words, f is upper semi-continuous at $\bar{x} \in S$ if

$$x_k \to \bar{x} \implies f(\bar{x}) \geq \limsup_{k\to\infty} f(x_k). \tag{13}$$

We say that f is lower semi-continuous on S (respectively, upper semi-continuous on S) if it is lower semi-continuous (respectively, upper semi-continuous) at every $\bar{x} \in S$.

Lower semi-continuous functions in one variable have the appearance shown in Figure 2.

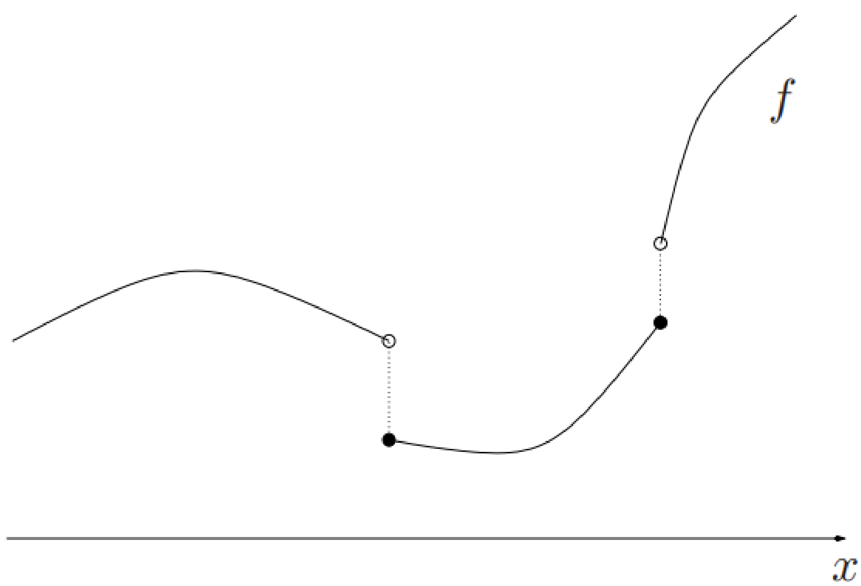

Figure 2: A lower semi-continuous function in one variable

Establish the following important relations:

(a) The function f mentioned in Definition 4 is *continuous* at $\bar{x} \in S$ if and only if it is both lower and upper semi-continuous at $\bar{x}$.

(b) The lower semi-continuity of f is equivalent to the closedness of all its level sets $\text{lev}_f(b)$, $b \in \mathbb{R}$, as well as the closedness of its epigraph.

Next follows the famous existence theorem credited to Karl Weierstrass.

Theorem 2 (Weierstrass' Theorem)**.** *Let $S \subseteq \mathbb{R}^n$ be a nonempty and closed set, and $f : S \to \mathbb{R}$ be a lower semi-continuous function on S. If f is weakly coercive with respect to S, then there exists a nonempty, closed, and bounded (thus compact) set of globally optimal solutions to the problem 1.*

Proof. We first assume that S is bounded, and proceed by choosing a sequence $\{x_k\}$ in S such that

$$\lim_{k\to\infty} f(x_k) = \inf_{x\in S} f(x). \tag{14}$$

(The infimum of f over S is the lowest limit of all sequences of the form $\{f(x_k)\}$ with $\{x_k\} \subset S$, so such a sequence of vectors $\{x_k\}$ is what we are choosing.)

Due to the boundedness of S, the sequence $\{x_k\}$ must have limit points, all of which lie in S because of the closedness of S. Let $\bar{x}$ be an arbitrary limit point of $\{x_k\}$, corresponding to the subsequence $K \subseteq \mathbb{Z}^+$. Then, by the lower semi-continuity of f,

$$f(\bar{x}) \leq \liminf_{k\in K} f(x_k) = \lim_{k\in K} f(x_k) = \inf_{x\in S} f(x). \tag{15}$$

Since $\bar{x}$ attains the infimum of f over S, $\bar{x}$ is a global minimum of f over S. This limit point of $\{x_k\}$ was arbitrarily chosen; any other choice (if more than one exists) has the same (optimal) objective value.

Suppose next that f is weakly coercive, and consider the same sequence $\{x_k\}$ in S. Then, by the weak coercivity assumption, either $\{x_k\}$ is bounded or the elements of the sequence $\{f(x_k)\}$ tend to infinity. The non-emptiness of S implies that $\inf_{x\in S} f(x) < \infty$, and hence we conclude that $\{x_k\}$ is bounded. We can then utilize the same arguments as in the previous paragraph and conclude that, also in this case, there exists a globally optimal solution. We are done. $\square$

Remark 3. *Before moving on, we take a closer look at the proof of this result, because it is instrumental in understanding the importance of some of the assumptions that we make about the optimization models that we pose.*

We notice that the closedness of S is really crucial; if S is not closed, then a sequence generated in S may converge to a point outside of S, which means that we would converge to an infeasible and, of course, also non-optimal solution.

This is the reason why the generic optimization model (1.1) stated in Chapter 1 does not contain any constraints of the form

$$g_i(x) < 0, \quad i \in S_I, \tag{16}$$

where S_I denotes strict inequality. The reason is that such constraints in general may describe non-closed sets.

2.2 Non-standard results

Theorem 3 (Existence of Optimal Solutions, Convex Polynomials)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a convex polynomial function. Suppose further that the set S can be described by inequality constraints of the form $g_i(x) \leq 0$, $i = 1, \ldots, m$, where each function g_i is convex and polynomial.*

The problem 1 then has a nonempty (as well as closed and convex) set of globally optimal solutions if and only if f is lower bounded on S.

In the following result, we let S be a nonempty polyhedron, and suppose that it is possible to describe it as the following finite set of linear constraints:

$$S := \{x \in \mathbb{R}^n \mid Ax \leq b;\, Ex = d\}, \tag{17}$$

where $A \in \mathbb{R}^{m\times n}$, $E \in \mathbb{R}^{\ell\times n}$, $b \in \mathbb{R}^m$, and $d \in \mathbb{R}^\ell$.

The recession cone to S then is the following set, defining the set of directions that are feasible at every point in S:

$$\operatorname{rec} S := \{p \in \mathbb{R}^n \mid Ap \leq 0_m;\, Ep = 0_\ell\}. \tag{18}$$

We also suppose that

$$f(x) := \frac{1}{2}x^T Qx + q^T x, \quad x \in \mathbb{R}^n, \tag{19}$$

where $Q \in \mathbb{R}^{n\times n}$ is a symmetric and positive semidefinite matrix, and $q \in \mathbb{R}^n$.

We define the recession cone to any convex function $f : \mathbb{R}^n \to \mathbb{R}$ as follows: the recession cone to f is the recession cone to the level set of f, defined for any value of b for which the corresponding level set of f is nonempty.

In the special case of the convex quadratic function given in (19),

$$\operatorname{rec} f = \{p \in \mathbb{R}^n \mid Qp = 0_n;\, q^T p \leq 0\}. \tag{20}$$

This is the set of directions that are nowhere ascent directions to f.

Corollary 1 (Frank–Wolfe Theorem). *Suppose that S is the polyhedron described by* (17)*, and f is the convex quadratic function given by* (19)*, so that the problem 1 is a convex quadratic programming problem. Then, the following three statements are equivalent:*

(a) The problem 1 has a nonempty (as well as a closed and convex) set of globally optimal solutions.

(b) f is lower bounded on S.

(c) For every vector p in the intersection of the recession cone $\operatorname{rec} S$ to S and the null space $\mathcal{N}(Q)$ of the matrix Q, it holds that $q^T p \geq 0$.

In other words,

$$p \in \operatorname{rec} S \cap \mathcal{N}(Q) \implies q^T p \geq 0. \tag{21}$$

The statement in (c) shows that the conditions for the existence of an optimal solution in the case of convex quadratic programs are milder than in the general convex case. In the latter case, we can state a slight improvement over the Weierstrass Theorem 2:

If, in the problem 1, f is convex on S, where the latter is nonempty, closed, and convex, then the problem has a nonempty, convex, and compact set of globally optimal solutions if and only if

$$\operatorname{rec} S \cap \operatorname{rec} f = \{0_n\}. \tag{22}$$

The improvements in the above results for polyhedral, in particular quadratic, programs stem from the fact that convex polynomial functions cannot be lower bounded and yet not have a global minimum.

Remark 4. *Consider the special case of the problem 1 where*

$$f(x) := \frac{1}{x}, \quad S := [1, +\infty). \tag{23}$$

It is clear that f is bounded from below on S, in fact by the value zero, which is the infimum of f over S, but it never attains the value zero on S. Therefore, this problem has no optimal solution. Of course, f is not a polynomial function.

Corollary 2 (A Fundamental Theorem in Linear Programming)**.** *Suppose, in the Frank–Wolfe Theorem, that f is linear, that is, $Q = 0_{n\times n}$. Then, the problem 1 is identical to a linear programming (LP) problem. The following three statements are equivalent:*

(a) The problem 1 has a nonempty (as well as a closed and convex polyhedral) set of globally optimal solutions.

(b) f is lower bounded on S.

(c) For every vector p in the recession cone $rec\, S$ to S, it holds that $q^T p \geq 0$.

In other words,

$$p \in rec\, S \implies q^T p \geq 0. \tag{24}$$

Corollary 2 will in fact be established later, by the use of polyhedral convexity, when we specialize our treatment of nonlinear optimization to that of linear optimization.

Since we have already established the Representation Theorem, proving Corollary 2 for the case of LP will be straightforward: since the objective function is linear, every feasible direction $p \in \operatorname{rec} S$ with $q^T p < 0$ leads to an unbounded solution from any vector $x \in S$.

2.3 Special optimal solution sets

Proposition 1 (Unique Optimal Solution under Strict Convexity)**.** *Suppose that in the problem 1, f is strictly convex on S and the set S is convex. Then, there can be at most one globally optimal solution.*

Proof. Suppose, by means of contradiction, that x^* and x^{**} are two different globally optimal solutions. Then, for every $\lambda \in (0, 1)$, we have that

$$f\big(\lambda x^* + (1-\lambda)x^{**}\big) < \lambda f(x^*) + (1-\lambda) f(x^{**}) = f(x^*) \, [= f(x^{**})]. \tag{25}$$

Since $\lambda x^* + (1-\lambda)x^{**} \in S$, we have found an entire interval of points that are strictly better than x^* or x^{**}. This is a contradiction. Hence, there can be at most one globally optimal solution. □

We now characterize a class of optimization problems over polytopes whose optimal solution set, if nonempty, includes an extreme point. Consider the maximization problem:

$$\text{maximize } f(x), \quad \text{subject to } x \in P, \tag{26}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function and $P \subset \mathbb{R}^n$ is a nonempty, bounded polyhedron (that is, a polytope). From the Representation Theorem, it follows that an optimal solution can be found among the extreme points of P.

Theorem 4 (Optimal Extreme Point)**.** *An optimal solution to* (26) *can be found among the extreme points of P.*

Proof. The function f is continuous; further, P is a nonempty and compact set. Hence, there exists an optimal solution $\tilde{x}$ to (26) by Weierstrass' Theorem 2. The Representation Theorem implies that

$$\tilde{x} = \lambda_1 v_1 + \cdots + \lambda_k v_k, \tag{27}$$

for some extreme points $v_1, \ldots, v_k$ of P and $\lambda_1, \ldots, \lambda_k \geq 0$ such that $\sum_{i=1}^k \lambda_i = 1$. By the convexity of f, we have

$$f(\tilde{x}) = f\big(\lambda_1 v_1 + \cdots + \lambda_k v_k\big) \leq \lambda_1 f(v_1) + \cdots + \lambda_k f(v_k). \tag{28}$$

Since $\tilde{x}$ is optimal, it follows that

$$f(\tilde{x}) = f(v_i) \quad \text{for some } i = 1, \ldots, k. \tag{29}$$

Thus, an optimal solution can be found among the extreme points of P. □

Remark 5. *Every linear function is convex, so Theorem 4 implies, in particular, that every linear program over a nonempty and bounded polyhedron has an optimal extreme point.*

References

Niclas Andréasson, Anton Evgrafov, and Michael Patriksson. *An introduction to continuous optimization: foundations and fundamental algorithms.* Courier Dover Publications, 2020.