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.

2 Existence of optimal solutions

2.1 A classic result

Theorem 1 (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

$$\lim_{k\to\infty} f(x_k) = \inf_{x\in S} f(x). \tag{3}$$

(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,

$$f(\bar{x}) \leq \liminf_{k\in K} f(x_k) = \lim_{k\in K} f(x_k) = \inf_{x\in S} f(x). \tag{4}$$

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. □

Remark 1. *Case 1: Assuming S is bounded*

1. ***Construct a decreasing sequence:*** *Choose a sequence $\{x_k\} \subset S$ such that:*

$$\lim_{k\to\infty} f(x_k) = \inf_{x\in S} f(x). \tag{5}$$

   This means that we select a set of points $\{x_k\}$ where $f(x_k)$ approaches the minimum value of f on S.

2. ***Use boundedness:*** *Since S is bounded, the sequence $\{x_k\}$ must have a convergent subsequence (by the Bolzano-Weierstrass theorem).*

3. ***Use closedness:*** *Because S is closed, the limit point of the subsequence, denoted as $\bar{x}$, must lie within S.*

4. ***Use lower semi-continuity:*** *By the definition of lower semi-continuity:*

$$f(\bar{x}) \leq \liminf_{k\to\infty} f(x_k). \tag{6}$$

   Since $f(x_k) \to \inf_{x\in S} f(x)$, it follows that $\bar{x}$ attains the global minimum:

$$f(\bar{x}) = \inf_{x\in S} f(x). \tag{7}$$

5. ***Conclusion:*** *$\bar{x}$ is a global minimum of f on S.*

Case 2: Assuming S is unbounded

1. ***Apply weak coercivity:*** *By the definition of weak coercivity, if we construct the same sequence $\{x_k\} \subset S$, then:*
   - *Either $\{x_k\}$ is bounded (in which case it has a limit point, reducing to Case 1), or*
   - *$f(x_k) \to \infty$.*

2. ***Eliminate $f(x_k) \to \infty$:*** *If $f(x_k) \to \infty$, this contradicts the assumption that:*

$$\inf_{x\in S} f(x) < \infty, \tag{8}$$

   since $f(x_k)$ should approach a finite value. Thus, we can conclude that $\{x_k\}$ must be bounded.

3. ***Use boundedness result:*** *As in Case 1, the bounded sequence $\{x_k\}$ has a convergent subsequence with a limit point $\bar{x}$, which is the global minimum.*

Conclusion:

Regardless of whether S is bounded or unbounded, there exists a nonempty, closed, and bounded set of global optimal solutions.

Intuitive Explanation of the Theorem's Significance:

1. ***Why lower semi-continuity is necessary:*** *Lower semi-continuity ensures that the function value does not suddenly increase when taking a limit, allowing the limit point to still be a valid optimal solution.*

2. ***Why weak coercivity is necessary:*** *Weak coercivity ensures that even if S is unbounded, the function $f(x)$ does not grow uncontrollably in certain directions, restricting the range of optimal solutions.*

3. ***Objective:*** *By combining closedness, lower semi-continuity, and weak coercivity, the theorem successfully limits the range of global optimal solutions, guaranteeing their existence in practical optimization problems.*

Remark 2. *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 does not contain any constraints of the form

$$g_i(x) < 0, \quad i \in S_I, \tag{9}$$

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 2 (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:

$$S := \{x \in \mathbb{R}^n \mid Ax \leq b;\ Ex = d\}, \tag{10}$$

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:

$$\operatorname{rec} S := \{p \in \mathbb{R}^n \mid Ap \leq 0_m;\ Ep = 0_\ell\}. \tag{11}$$

We also suppose that

$$f(x) := \frac{1}{2}x^T Qx + q^T x, \quad x \in \mathbb{R}^n, \tag{12}$$

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 (12),

$$\operatorname{rec} f = \{p \in \mathbb{R}^n \mid Qp = 0_n;\ q^T p \leq 0\}. \tag{13}$$

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 (10), and f is the convex quadratic function given by (12), 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,

$$p \in \operatorname{rec} S \cap \mathcal{N}(Q) \implies q^T p \geq 0. \tag{14}$$

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 1:

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

$$\operatorname{rec} S \cap \operatorname{rec} f = \{0_n\}. \tag{15}$$

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 3. *Consider the special case of the problem 1 where*

$$f(x) := \frac{1}{x}, \quad S := [1, +\infty). \tag{16}$$

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,

$$p \in rec\, S \implies q^T p \geq 0. \tag{17}$$

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

Definition 1 (Convex Set)**.** *A set $S \subseteq \mathbb{R}^n$ is called* convex *if for any $x_1, x_2 \in S$ and $\lambda \in [0, 1]$, the convex combination $\lambda x_1 + (1-\lambda)x_2$ also belongs to S. Mathematically, this is expressed as:*

$$\lambda x_1 + (1-\lambda)x_2 \in S, \quad \forall x_1, x_2 \in S,\ \lambda \in [0,1]. \tag{18}$$

Definition 2 (Convex Function)**.** *A function $f : S \to \mathbb{R}$ defined on a convex set S is called* convex *if for any $x_1, x_2 \in S$ and $\lambda \in [0, 1]$, the following inequality holds:*

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda f(x_1) + (1-\lambda) f(x_2). \tag{19}$$

Definition 3 (Strictly Convex Function)**.** *A function $f : S \to \mathbb{R}$ defined on a convex set S is called* strictly convex *if for any $x_1, x_2 \in S$, $x_1 \neq x_2$, and $\lambda \in (0, 1)$, the following strict inequality holds:*

$$f\big(\lambda x_1 + (1-\lambda)x_2\big) < \lambda f(x_1) + (1-\lambda) f(x_2). \tag{20}$$

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

$$f\big(\lambda x^* + (1-\lambda)x^{**}\big) < \lambda f(x^*) + (1-\lambda) f(x^{**}) = f(x^*)\, [= f(x^{**})]. \tag{21}$$

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:

$$\text{maximize } f(x), \quad \text{subject to } x \in P, \tag{22}$$

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 3 (Optimal Extreme Point)**.** *An optimal solution to* (22) *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 (22) by Weierstrass' Theorem 1. The Representation Theorem implies that

$$\tilde{x} = \lambda_1 v_1 + \cdots + \lambda_k v_k, \tag{23}$$

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

$$f(\tilde{x}) = f\big(\lambda_1 v_1 + \cdots + \lambda_k v_k\big) \leq \lambda_1 f(v_1) + \cdots + \lambda_k f(v_k). \tag{24}$$

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

$$f(\tilde{x}) = f(v_i) \quad \text{for some } i = 1, \ldots, k. \tag{25}$$

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

Remark 4. *Every linear function is convex, so Theorem 3 implies, in particular, that every linear program over a nonempty and bounded polyhedron has an optimal extreme point.*

3 Optimality in unconstrained optimization

In Theorem[Fundamental Theorem of global optimality], we have established that locally optimal solutions are also global in the convex case. What are the necessary and sufficient conditions for a vector x^* to be a local optimum? This is an important question, because the algorithms that we will investigate for solving important classes of optimization problems are always devised based on those conditions that we would like to fulfill. This is a statement that seems to be true universally: efficient, locally or globally convergent iterative algorithms for an optimization problem are directly based on its necessary and/or sufficient local optimality conditions.

We begin by establishing these conditions for the case of unconstrained optimization, where the objective function is in C^1. Every proof is based on the Taylor expansion of the objective function up to order one or two. Our problem here is the following:

$$\text{minimize } f(x), \quad x \in \mathbb{R}^n, \tag{26}$$

where f is in C^1 on $\mathbb{R}^n$ (for short we say: in C^1 or $C^1(\mathbb{R}^n)$).

Theorem 4 (Necessary Optimality Conditions, C^1 Case)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is in C^1 on $\mathbb{R}^n$. Then,*

$$x^* \text{ is a local minimum of } f \text{ over } \mathbb{R}^n \implies \nabla f(x^*) = 0.$$

Note that:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix},$$

so the requirement is:

$$\frac{\partial f(x^*)}{\partial x_j} = 0, \quad j = 1, \ldots, n. \tag{27}$$

We refer to this condition as x^* being a stationary point of f.

Special Case: For $n = 1$, Theorem 4 reduces to:

$$x^* \in \mathbb{R} \text{ is a local minimum} \implies f'(x^*) = 0.$$

Proof. (By contradiction.) Suppose that x^* is a local minimum, but $\nabla f(x^*) \neq 0$. Let $p := -\nabla f(x^*)$, and consider the Taylor expansion of f around $x = x^*$ in the direction of p:

$$f(x^* + \alpha p) = f(x^*) + \alpha \nabla f(x^*)^\top p + o(\alpha), \tag{28}$$

where $o : \mathbb{R} \to \mathbb{R}$ satisfies $\frac{o(s)}{s} \to 0$ as $s \to 0$.

Substituting $p = -\nabla f(x^*)$, we get:

$$f(x^* + \alpha p) = f(x^*) - \alpha \|\nabla f(x^*)\|^2 + o(\alpha). \tag{29}$$

For sufficiently small $\alpha > 0$, this implies:

$$f(x^* + \alpha p) < f(x^*),$$

which contradicts the assumption that x^* is a local minimum. Hence, $\nabla f(x^*) = 0$. This completes the proof. $\square$

The opposite direction is false: take $f(x) = x^3$; then, $\bar{x} = 0$ is stationary, but it is neither a local minimum nor a local maximum.

The proof is instrumental in that it provides a sufficient condition for a vector p to define a descent direction, that is, a direction such that a small step along it yields a lower objective value. We first define this notion properly.

Definition 4 (Descent Direction)**.** *Let the function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be given. Let $x \in \mathbb{R}^n$ be a vector such that $f(x)$ is finite. Let $p \in \mathbb{R}^n$. We say that the vector $p \in \mathbb{R}^n$ is a* descent direction *with respect to f at x if there exists $\delta > 0$ such that:*

$$f(x + \alpha p) < f(x) \quad \textit{for every } \alpha \in (0, \delta]. \tag{30}$$

Proposition 2 (Sufficient Condition for Descent)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is in C^1 around a point x for which $f(x) < +\infty$, and that $p \in \mathbb{R}^n$. Then:*

$$\nabla f(x)^\top p < 0 \implies p \text{ is a descent direction with respect to } f \text{ at } x. \tag{31}$$

Proof. Since f is in C^1 around x, we can construct a Taylor expansion of f, as follows:

$$f(x + \alpha p) = f(x) + \alpha \nabla f(x)^\top p + o(\alpha), \tag{32}$$

where $o : \mathbb{R} \to \mathbb{R}$ satisfies $\frac{o(s)}{s} \to 0$ as $s \to 0$.

Since $\nabla f(x)^\top p < 0$, we have:

$$f(x + \alpha p) < f(x), \tag{33}$$

for all sufficiently small $\alpha > 0$. This completes the proof. □

Notice that at a point $x \in \mathbb{R}^n$, there may be other descent directions $p \in \mathbb{R}^n$ besides those satisfying $\nabla f(x)^\top p < 0$.

If f is additionally convex, then the opposite implication in the above proposition is true, thus making the descent property equivalent to the property that the directional derivative is negative. Since this result can also be stated for non-differentiable functions f (in which case we replace the expression $\nabla f(x)^\top p$ with the classic expression for the directional derivative):

$$f'(x; p) := \lim_{\alpha \to 0^+} \frac{1}{\alpha} \big[f(x + \alpha p) - f(x) \big], \tag{34}$$

If f has stronger differentiability properties, then we can make additional statements about the nature of a local optimum.

Theorem 5 (Necessary Optimality Conditions, C^2 Case)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is in C^2 on $\mathbb{R}^n$. Then,*

$$x^* \text{ is a local minimum of } f \implies \begin{cases} \nabla f(x^*) = 0, \\ \nabla^2 f(x^*) \text{ is positive semidefinite.} \end{cases}$$

Note: *For $n = 1$, Theorem 5 reduces to:*

$$x^* \in \mathbb{R} \text{ is a local minimum of } f \implies f'(x^*) = 0 \text{ and } f''(x^*) \geq 0.$$

Proof. Consider the Taylor expansion of f up to order two around x^* in the direction of a vector $p \in \mathbb{R}^n$:

$$f(x^* + \alpha p) = f(x^*) + \alpha \nabla f(x^*)^\top p + \frac{\alpha^2}{2} p^\top \nabla^2 f(x^*) p + o(\alpha^2). \tag{35}$$

Suppose that x^* satisfies $\nabla f(x^*) = 0$, but there exists a vector $p \neq 0$ such that $p^\top \nabla^2 f(x^*) p < 0$, meaning that $\nabla^2 f(x^*)$ is not positive semidefinite. Then the above expansion yields:

$$f(x^* + \alpha p) < f(x^*) \quad \text{for all sufficiently small } \alpha > 0,$$

which implies that x^* cannot be a local minimum. This completes the proof. □

The next result shows that under some circumstances, we can establish local optimality of a stationary point.

Theorem 6 (Sufficient Optimality Conditions, C^2 Case)**.** *Suppose that* $f : \mathbb{R}^n \to \mathbb{R}$ *is in* C^2 *on* $\mathbb{R}^n$*. Then:*

$$\nabla f(x^*) = 0 \quad \textit{and} \quad \nabla^2 f(x^*) \textit{ is positive definite} \implies x^* \textit{ is a strict local minimum of } f \textit{ over } \mathbb{R}^n.$$

Note: *For* $n = 1$*, Theorem 6 reduces to:*

$$f'(x^*) = 0 \quad \textit{and} \quad f''(x^*) > 0 \implies x^* \in \mathbb{R} \textit{ is a strict local minimum of } f \textit{ over } \mathbb{R}.$$

Proof. Suppose that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite. Take an arbitrary vector $p \in \mathbb{R}^n, p \neq 0$. Then, using the Taylor expansion:

$$f(x^* + \alpha p) = f(x^*) + \alpha \nabla f(x^*)^\top p + \frac{\alpha^2}{2} p^\top \nabla^2 f(x^*) p + o(\alpha^2). \tag{36}$$

Since $\nabla f(x^*)^\top p = 0$ and $p^\top \nabla^2 f(x^*) p > 0$ (due to positive definiteness of $\nabla^2 f(x^*)$), we have:

$$f(x^* + \alpha p) > f(x^*) \quad \text{for all small enough } \alpha > 0. \tag{37}$$

As p was arbitrary, this implies that x^* is a strict local minimum of f over $\mathbb{R}^n$. □

We naturally face the following question: When is a stationary point a global minimum? The answer is given in the next theorem.

Theorem 7 (Necessary and Sufficient Global Optimality Conditions)**.** *Suppose that* $f : \mathbb{R}^n \to \mathbb{R}$ *is convex and in* C^1 *on* $\mathbb{R}^n$*. Then:*

$$x^* \textit{ is a global minimum of } f \textit{ over } \mathbb{R}^n \iff \nabla f(x^*) = 0.$$

Proof. [$\Longrightarrow$] This has already been shown in Theorem 4, since a global minimum is also a local minimum.

[$\Longleftarrow$] The convexity of f yields that for every $y \in \mathbb{R}^n$:

$$f(y) \geq f(x^*) + \nabla f(x^*)^\top (y - x^*) = f(x^*), \tag{38}$$

where the equality stems from the property that $\nabla f(x^*) = 0$. This proves that x^* is a global minimum of f over $\mathbb{R}^n$. □

References

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