TMA947 / MMG621 — Nonlinear optimization

Lecture 3 — Introduction to optimality conditions
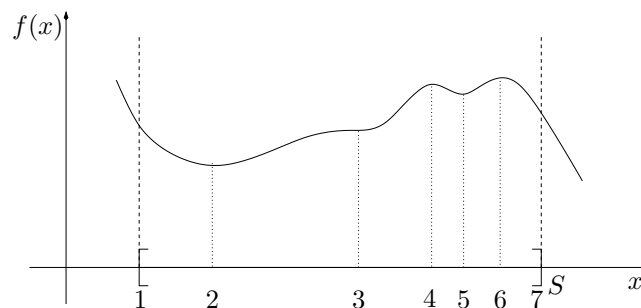
Emil Gustavsson, Zuzana Šabartová

November 7, 2014

Local and global optimality

We consider an optimization problem which is that to

$$\text{minimize } f(\boldsymbol{x}), \tag{1a}$$
$$\text{subject to } \boldsymbol{x} \in S, \tag{1b}$$

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

When $n = 1$, we know that an optimal solution will be found at

- boundary points of S,

- stationary points, that is, where $f'(x) = 0$,

- discontinuities of f or f'.

Definition (global minimum). $\boldsymbol{x}^* \in S$ is a *global minimum of f over S* if

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \qquad \boldsymbol{x} \in S$$

The next important definition is the one of a local minimum.

Definition (local minimum). *$\boldsymbol{x}^* \in S$ is a local minimum of f over S if*

$$\exists \varepsilon > 0 \text{ such that } f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}), \qquad \boldsymbol{x} \in S \cap B_\varepsilon(\boldsymbol{x}^*),$$

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

We denote $\boldsymbol{x}^* \in S$ as a *strict local minimum* of f over S if $f(\boldsymbol{x}^*) < f(\boldsymbol{x})$ holds above for $\boldsymbol{x} \neq \boldsymbol{x}^*$.

One of the most important theorems in optimization follows.

Theorem (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. See Theorem 4.3 in the book. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Existence of optimal solutions

First some basic notations.

- We say that a set $S \subseteq \mathbb{R}^n$ is *open* if for every $\boldsymbol{x} \in S$ there exists an $\varepsilon > 0$ such that $B_\varepsilon(\boldsymbol{x}) := \{\, \boldsymbol{y} \in \mathbb{R}^n \mid \|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon \,\} \subset S$.

- We say that a set $S \subseteq \mathbb{R}^n$ is *closed* if $\mathbb{R}^n \setminus S$ is open.

- A *limit point* of a set $S \subseteq \mathbb{R}^n$ is a point $\boldsymbol{x}$ such that there exists a sequence $\{\boldsymbol{x}_k\}_{k=1}^\infty \subset S$ fulfilling $\boldsymbol{x}_k \to \boldsymbol{x}$.

- We can then define a closed set as a set which contains all its limit points.

- If a set is both closed and bounded, we call it *compact*.

Two important definitions needed to formulate Weierstrass' Theorem are the following.

Definition (weakly coercive function). *A function f is said to be* weakly coercive *with respect to the set S if either S is bounded or*

$$\lim_{\substack{\|\boldsymbol{x}\| \to \infty \\ \boldsymbol{x} \in S}} f(\boldsymbol{x}) = \infty$$

Definition (lower semi-continuity). *A function f is said to be* lower semi-continuous *at $\boldsymbol{x}$ if the value $f(\boldsymbol{x})$ is less than or equal to every limit of f as $\boldsymbol{x}_k \to \boldsymbol{x}$*

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

$$\boldsymbol{x}_k \to \boldsymbol{x} \qquad \Longrightarrow \qquad f(\boldsymbol{x}) \leq \liminf_{k \to \infty} f(\boldsymbol{x}_k)$$
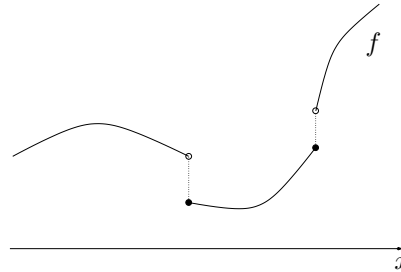
Figure 1: A lower semi-continuous function in one variable

Now we can formulate Weierstrass' Theorem which guarantees the existence of optimal solutions to an optimization problem as long as a few assumptions are satisfied.

Theorem (Weierstrass' Theorem). *Consider the problem* (1), *where S is a nonempty and closed set and f is lower semi-continuous on S. If f is weakly coercive with respect to S, then there exists a nonempty, closed and bounded (thus compact) set of optimal solutions to the problem* (1).

Proof. See Theorem 4.6 in the book.   □

One way to remember the assumptions in Weierstrass' Theorem is to imagine what can go wrong, i.e., when does a problem not have an optimal solution. One example of an optimization problem where the solution set is empty is when $f(x) = 1/x$ and $S = [1, \infty)$.

Optimality conditions when $S = \mathbb{R}^n$

When $S = \mathbb{R}^n$, i.e., the problem is an unconstrained optimization problem, then the following theorem holds.

Theorem (necessary condition for optimality, C^1). *If $f \in C^1$ on $\mathbb{R}^n$, then*

$$\boldsymbol{x}^* \text{ is a local minimum of } f \text{ on } \mathbb{R}^n \qquad \Longrightarrow \qquad \nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$$

Proof. See Theorem 4.13.   □

Note that $\nabla f(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_n} \right)^{\mathrm{T}}$. The opposite of the theorem is, however, not true. Take $f(x) = x^3$ and $x = 0$. We can strengthen the theorem by assuming that f is also in C^2.

Theorem (necessary condition for optimality, C^2). *If $f \in C^2$ on $\mathbb{R}^n$, then*

$$\boldsymbol{x}^* \text{ is a local minimum of } f \text{ on } \mathbb{R}^n \qquad \Longrightarrow \qquad \begin{cases} \nabla f(\boldsymbol{x}^*) = \boldsymbol{0} \\ \nabla^2 f(\boldsymbol{x}) \succeq 0 \end{cases}$$

Proof. See Theorem 4.16.   □

3

Remember that for a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, the notion $\boldsymbol{A} \succeq 0$ ($\boldsymbol{A}$ positive semidefinite) means that $\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} \geq 0$, for all $\boldsymbol{x} \in \mathbb{R}^n$. Now once again, the opposite direction in the theorem is not true. However, by assuming positive definiteness of the Hessian of f, we can obtain a sufficient condition.

Theorem (sufficient condition for optimality, C^2). *If $f \in C^2$ on $\mathbb{R}^n$, then*

$$\left\{ \begin{array}{l} \nabla f(\boldsymbol{x}^*) = \boldsymbol{0} \\ \nabla^2 f(\boldsymbol{x}) \succ 0 \end{array} \right. \qquad \Longrightarrow \qquad \boldsymbol{x}^* \text{ is a strict local minimum of } f \text{ on } \mathbb{R}^n$$

Proof. See Theorem 4.17. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

To get sufficient conditions for a point to be a local minimum, we need to assume convexity of the function f.

Theorem (necessary and sufficient condition for optimality, C^1). *If $f \in C^1$ is convex on $\mathbb{R}^n$, then*

$$\boldsymbol{x}^* \text{ is a global minimum of } f \text{ on } \mathbb{R}^n \qquad \Longleftrightarrow \qquad \nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$$

Proof. See Theorem 4.18. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Optimality conditions for $S \subseteq \mathbb{R}^n$

When $S = \mathbb{R}^n$ the directions we could move from a point $\boldsymbol{x}$ and still stay feasible were $\mathbb{R}^n$ itself. When we consider cases where $S \subset \mathbb{R}^n$ this might not hold.

Definition (feasible direction). *Let $\boldsymbol{x} \in S$. A vector $\boldsymbol{p} \in \mathbb{R}^n$ defines a* feasible direction *at $\boldsymbol{x}$ if*

$$\exists \delta > 0 : \quad \boldsymbol{x} + \alpha \boldsymbol{p} \in S, \quad \text{for all} \quad \alpha \in [0, \delta].$$

So the feasible directions at a point $\boldsymbol{x} \in S$ describes the directions in which we can "move" without becoming infeasible.

Definition (descent direction). *Let $\boldsymbol{x} \in \mathbb{R}^n$. A vector $\boldsymbol{p}$ defines a* descent direction *with respect to f at $\boldsymbol{x}$ if*

$$\exists \delta > 0 : \quad f(\boldsymbol{x} + \alpha \boldsymbol{p}) < f(\boldsymbol{x}), \quad \text{for all} \quad \alpha \in (0, \delta].$$

Suppose that $f \in C^1$ around a point $\boldsymbol{x} \in \mathbb{R}^n$, and that $\boldsymbol{p} \in \mathbb{R}^n$. If $\nabla f(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{p} < 0$ then the vector $\boldsymbol{p}$ defines a direction of descent with respect to f at $\boldsymbol{x}$. We can now state necessary optimality conditions for cases when $S \neq \mathbb{R}^n$.

Theorem (necessary optimality conditions). *Suppose that $S \subseteq \mathbb{R}^n$ and that $f \in C^1$.*

a) If $\boldsymbol{x}^* \in S$ is a local minimum of f over S, then

$$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}} \boldsymbol{p} \geq 0$$

holds for all feasible directions $\boldsymbol{p}$ at $\boldsymbol{x}^*$.

b) Suppose that S is convex. If $\boldsymbol{x}^*$ is a local minimum of f over S, then

$$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}} (\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \boldsymbol{x} \in S \tag{2}$$

Proof. See Proposition 4.22 in the book.   □

We refer to (2) as a *variational inequality* and we can now extend the notion of *stationary points* by denoting them as points fulfilling (2). This is first out of four definitions of a stationary point.

Now the necessary and sufficient conditions for optimality can be stated in the following theorem.

Theorem (necessary and sufficient optimality conditions). *Suppose $S \subseteq \mathbb{R}^n$ is a convex nonempty set and that $f \in C^1$ is a convex function on S. Then*

$$\boldsymbol{x}^* \text{ is a global minimum of } f \text{ over } S \quad \Longleftrightarrow \quad \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \boldsymbol{x} \in S.$$

Proof. See Theorem 4.23 in the book.   □

Note that when $S = \mathbb{R}^n$, the expression to the right just becomes $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$. Why?

We will now present three additional definitions of a stationary point which are all equivalent to (2). The first one we get by taking the minimum of the left-hand-side of (2) and then realizing that the optimal value must be zero, i.e.,

$$\min_{\boldsymbol{x} \in S} \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) = 0. \tag{3}$$

Convince yourself that (2) and (3) are equivalent! Now we claim that (2) and (3) are also equivalent with

$$\boldsymbol{x}^* = \mathrm{Proj}_S \left[\boldsymbol{x}^* - \nabla f(\boldsymbol{x}^*) \right]. \tag{4}$$

The equation (4) states that if you stand in a stationary point and take a step in the direction of the negative gradient and then project back to the feasible set, you should end up in the same point. The details for showing this equivalence can be found in the book (pp. 100–101). See also Figure 3.

For the last equivalent definition of a stationary point, we need to introduce the normal cone

Definition (normal cone). *Suppose the set S is closed and convex. Let $\boldsymbol{x} \in S$. Then the normal cone to S at $\boldsymbol{x}$ is the set*

$$N_S(\boldsymbol{x}) := \left\{ \boldsymbol{p} \in \mathbb{R}^n \mid \boldsymbol{p}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}) \leq 0, \quad \boldsymbol{y} \in S \right\}.$$

Think of the normal cone at a point $\boldsymbol{x}$ as all direction pointing "straight out" from the set. See Figure 3. Now the fourth equivalent definition of a stationary point is that

$$-\nabla f(\boldsymbol{x}^*) \in N_S(\boldsymbol{x}^*). \tag{5}$$

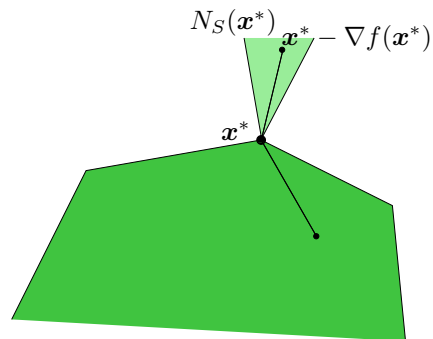That (5) is equivalent to (2) is trivial.

Figure 2: Illustration of the normal cone and the projection operator.

Summary of optimality condition for convex $S \subseteq \mathbb{R}^n$

Definition (stationary point). *Suppose that S is convex and that $f \in C^1$. A point $\boldsymbol{x} \in S$ fulfilling the four equivalent statements a)–b) are called a* stationary point.

a)
$$\nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \boldsymbol{x} \in S,$$

b)
$$\min_{\boldsymbol{x} \in S} \nabla f(\boldsymbol{x}^*)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{x}^*) = 0,$$

c)
$$\boldsymbol{x}^* = \mathrm{Proj}_S \left[\boldsymbol{x}^* - \nabla f(\boldsymbol{x}^*) \right],$$

d)
$$-\nabla f(\boldsymbol{x}^*) \in N_S(\boldsymbol{x}^*).$$

The two important theorems which will be utilized throughout the whole course are the following.

Theorem (necessary optimality conditions). *Suppose that S is convex and that $f \in C^1$. Then*

$$\boldsymbol{x}^* \text{ is a local minimum of } f \text{ over } S \quad \Longrightarrow \quad \boldsymbol{x} \text{ is stationary}$$

Theorem (necessary and sufficient optimality conditions). *Suppose that S is convex and that $f \in C^1$ is convex. Then*

$$\boldsymbol{x}^* \text{ is a global minimum of } f \text{ over } S \quad \Longleftrightarrow \quad \boldsymbol{x} \text{ is stationary}$$

As we will see later in the course, the last definition (the inclusion (5)) is the only one that can be extended to the case of non-convex sets S.

The separation theorem

Now we have the tools to prove the separation theorem.

Theorem (the separation theorem). *Suppose that $S \subseteq \mathbb{R}^n$ is closed and convex, and that the point $\boldsymbol{y}$ does not lie in S. Then there exists a vector $\boldsymbol{\pi} \neq \boldsymbol{0}$ and a scalar $\alpha \in \mathbb{R}$ such that $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{y} > \alpha$, and $\boldsymbol{\pi}^{\mathrm{T}}\boldsymbol{x} \leq \alpha$ for all $\boldsymbol{x} \in S$.*
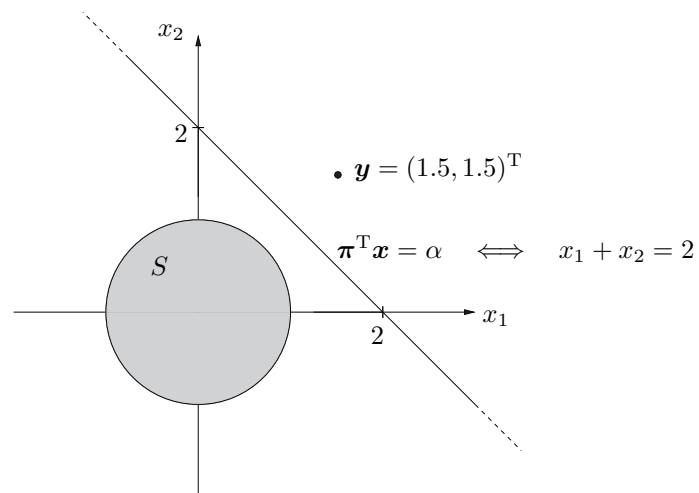
Proof. See Theorem 4.28.   □

Figure 3: Illustration of the separation theorem.