Nonlinear Optimization Cheatsheet

Preliminaries

Theorems

Th.L1.1 Weierstrass. Let $f : \Omega \to \mathbb{R}$ be a continuous function defined on a nonempty and compact (i.e., closed and bounded) set Ω. Then, there exists a minimizer $x^* \in \Omega$ of f on Ω, that is,

$$f(x^*) \le f(x) \quad \text{for all } x \in \Omega$$

Th.L1.2 Weierstrass for Compact Sublevel Sets. Let f be a continuous function defined on a set S. If f has a nonempty and compact sublevel set, that is, there exists $\alpha \in \mathbb{R}$ such that

$$\{x \in S : f(x) \le \alpha\}$$

is nonempty, bounded, and closed, then f has a minimizing point in S.

Def.L2.1 Star-convexity at x. A set $\Omega \subseteq \mathbb{R}^n$ is said to be star-convex at a point $x \in \Omega$ if, for all $y \in \Omega$, the entire segment from x to y is contained in Ω. In symbols, if

$$x + t \cdot (y - x) \in \Omega \quad \forall t \in [0,1]$$

(Note that the condition is equivalent to "$t \cdot y + (1-t) \cdot x \in \Omega \ \forall\, y \in \Omega, t \in [0,1]$", or also "$t \cdot x + (1-t) \cdot y \in \Omega \ \forall\, y \in \Omega, t \in [0,1]$".)

Def.L2.2 Convex set. A set Ω is convex if it is star-convex at all of its points $x \in \Omega$. In other words, Ω is convex if all segments formed between any two points $x, y \in \Omega$ are entirely contained in Ω. In symbols, if

$$t \cdot x + (1-t) \cdot y \in \Omega \quad \forall x, y \in \Omega \text{ and } t \in [0,1].$$

Th.L2.1 First-order necessary condition for a convex feasible set Let $\Omega \subseteq \mathbb{R}^n$ be convex and $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. For a point $x \in \Omega$ to be a minimizer of f over Ω it is necessary that

$$\langle \nabla f(x), y - x \rangle \ge 0 \quad \forall y \in \Omega$$

Def.L2.3 Normal cone. Let $\Omega \subseteq \mathbb{R}^n$ be convex, and let $x \in \Omega$. The normal cone to Ω at x, denoted $\mathcal{N}_\Omega(x)$, is defined as the set

$$\mathcal{N}_\Omega(x) := \{d \in \mathbb{R}^n : \langle d, y - x \rangle \le 0 \quad \forall y \in \Omega\}.$$

With this definition, the first-order necessary optimality condition for x, given in Th L2.1, can be equivalently written as

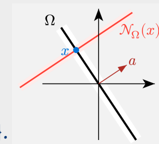
$$-\nabla f(x) \in \mathcal{N}_\Omega(x)$$

Th.L2.2 Normal cone to a hyperplane. Consider a hyperplane

$$\Omega := \{y \in \mathbb{R}^n : \langle a, y \rangle = 0\},$$

where $a \in \mathbb{R}^n, a \neq 0$, and a point $x \in \Omega$. The normal cone at x is given by

$$\mathcal{N}_\Omega(x) = \operatorname{span}\{a\} = \{\lambda a : \lambda \in \mathbb{R}\}.$$



Th.L3.1,L5.3 Normal cone to the intersection of m halfspaces. Let $\Omega \subseteq \mathbb{R}^n$ be given as the intersection of m linear inequalities $\langle a_j, x \rangle \le b_j$. Then, the normal cone at any point $x \in \Omega$ is obtained by taking nonnegative combinations of all those a_j's for which $\langle a_j, x \rangle = b_j$.

$$\mathcal{N}_\Omega(x) = \left\{ \sum_{j \in I(x)} \lambda_j \cdot a_j : \lambda_j \in \mathbb{R}_{\ge 0} \right\},$$

where $I(x) = \left\{ j \in \{1, \ldots, m\} : \langle a_j, x \rangle = b_j \right\}$.

The constraints j in $I(x)$ are often called the "active constraints" at $x \in \Omega$. Alternatively in vectorial form let $Ax \le b$, then with $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$:

$$\mathcal{N}_\Omega(x) = \left\{ A^\top \lambda : \lambda^\top (b - Ax) = 0, \lambda \in \mathbb{R}^m_{\ge 0} \right\},$$

$$A = \begin{pmatrix} -a_1^\top - \\ \vdots \\ -a_m^\top - \end{pmatrix},$$

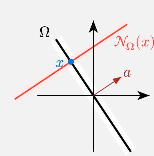
rewriting the condition $j \in I(x)$ (via **complementary slackness**).

Th.L3.2 Strong linear programming duality. If (P) admits an optimal solution x^*, then (D) admits an optimal solution λ^*, such that:

- the values of the two problems coincide: $c^\top x^* = b^\top \lambda^*$, and
- λ^* satisfies the complementary slackness condition $(\lambda^*)^\top (b - Ax^*) = 0$.

Def.L4.1 Convex function. Let $\Omega \subseteq \mathbb{R}^n$ be convex. A function $f : \Omega \to \mathbb{R}$ is convex if, for any two points $x, y \in \Omega$ and $t \in [0,1]$,

$$f((1-t) \cdot x + t \cdot y) \le (1-t) \cdot f(x) + t \cdot f(y)$$



Th.L4.1 Let $f : \Omega \to \mathbb{R}$ be a convex and differentiable function defined on a convex domain Ω. Then, at all $x \in \Omega$,

$$f(y) \ge \underbrace{f(x) + \langle \nabla f(x), y - x \rangle}_{\text{linearization of } f \text{ around } x} \quad \forall y \in \Omega.$$

Th.L4.2 Sufficiency of first-order optimality condition. Let $\Omega \subseteq \mathbb{R}^n$ be convex and $f : \Omega \to \mathbb{R}$ be a convex differentiable function. Then,

$$-\nabla f(x) \in \mathcal{N}_\Omega(x) \iff x \text{ is a minimizer of } f \text{ on } \Omega$$

Th.L4.3 Equivalent definitions of convexity. Let $\Omega \subseteq \mathbb{R}^n$ be a convex set, and $f : \Omega \to \mathbb{R}$ be a function. The following are equivalent definitions of convexity for f:

1. $f((1-t)x + ty) \le (1-t)f(x) + tf(y)$ for all $x, y \in \Omega, t \in [0,1]$.
2. $f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle$ for all $x, y \in \Omega$ [If f is differentiable]
3. $\langle \nabla f(y) - \nabla f(x), y - x \rangle \ge 0$ for all $x, y \in \Omega$ [If f is differentiable]
4. $\nabla^2 f(x) \succeq 0$ for all $x \in \Omega$ [If f is twice differentiable and Ω is open]

Th.L4.4 Operations preserving convexity.

- Multiplication of a convex function $f(x)$ by a nonnegative scalar $c \ge 0$;
- Addition of two convex functions $f(x), g(x)$;
- Pointwise supremum of a collection J of convex functions $\left\{ f_j(x) : j \in J \right\}$:
  $$f_{\max}(x) := \max_{j \in J} f_j(x);$$
- Pre-composition $f(Ax + b)$ of a convex function f with an affine function $Ax + b$.
- Post-composition $g(f(x))$ of a convex function with an increasing convex function g;
- Infimal convolution $f \pm g$ of two convex functions $f, g : \mathbb{R}^n \to \mathbb{R}$, defined as
  $$\inf\{f(y) + g(x - y) : y \in \mathbb{R}^n\}$$

Def.L4.2 Strict and strong convexity. Let $\Omega \subseteq \mathbb{R}^n$ be convex.

- A function $f : \Omega \to \mathbb{R}$ is strictly convex if, for any two distinct points $x, y \in \Omega$ and $t \in (0,1)$,
  $$f((1-t) \cdot x + t \cdot y) < (1-t) \cdot f(x) + t \cdot f(y)$$
- A function $f : \Omega \to \mathbb{R}$ is strongly convex with modulus $\mu > 0$ if the function
  $$f(x) - \frac{\mu}{2} \|x\|_2^2$$
  is convex. Note that strong convexity implies strict convexity, and strict convexity implies convexity. Neither of the reverse implications holds.

Th.L4.5 Strict convexity and uniqueness of minimizer. Let $\Omega \subseteq \mathbb{R}^n$ be convex, and $f : \Omega \to \mathbb{R}$ be a strictly convex function. Then, f has at most one minimizer.

Corollary (Projection onto convex set): Since the function $\|x - y\|_2^2$ is strongly convex, and hence strictly convex, it follows that any projection onto a convex set Ω, if it exists, is unique.

Th.L5.1 Separation Let $\Omega \subseteq \mathbb{R}^n$ be a nonempty, closed, and convex set, and let $y \in \mathbb{R}^n$ be a point. If $y \notin \Omega$, then there exist $u \in \mathbb{R}^n, v \in \mathbb{R}$ such that

$$\langle u, y \rangle < v, \quad \text{and} \quad \langle u, x \rangle \ge v \quad \forall x \in \Omega.$$

Note: Also works for strict inequalities if we choose v so that the havspace passes through midpoint of line connecting y and $x^\star$.

Def.L5.1 Cone A set S is a cone if, for any $x \in S$ and $\lambda \in \mathbb{R}_{\ge 0}$, the point $\lambda \cdot x \in S$.

Th.L5.2 Separation of a point from a cone. Let $S \subseteq \mathbb{R}^n$ be a nonempty closed convex cone, and $y \notin S$ be a point in $\mathbb{R}^n$. Then, there exists a hyperplane passing through the origin that separates y from S; formally, there exists $u \in \mathbb{R}^n$ such that

$$\langle u, y \rangle < 0 \quad \text{and} \quad \langle u, x \rangle \ge 0 \quad \forall x \in S$$

Def.L5.2 (Strong) separation oracle Let $\Omega \subseteq \mathbb{R}^n$ be convex and closed. A strong separation oracle for Ω is an algorithm that, given any point $y \in \mathbb{R}^n$, correctly outputs one of the following:

- "$y \in \Omega$", or
- $(y \notin \Omega, u)$", where the vector $u \in \mathbb{R}^n$ is such that $\quad \langle u, y \rangle < \langle u, x \rangle \quad \forall x \in \Omega$

Th.L5.4 Ellipsoid method for convex optimization. Theorem L5.4. Let R and r be as above, and let the range of the function f on Ω be bounded by $[-B, B]$. Then, the ellipsoid method described above run for $T \ge 2n^2 \log(R/r)$ steps either correctly reports that $\Omega = \emptyset$, or produces a point $x^\star$ such that

$$f\left(x^\star\right) \le f(x) + \frac{2BR}{r} \exp\left(-\frac{T}{2n(n+1)}\right) \quad \forall x \in \Omega.$$

Th.L6.1 Farkas lemma. Let $Ax \le b$ be a system of inequalities where $A \in \mathbb{R}^{m \times n}$. Then, exactly one of the following options is true:

- either $Ax \le b$ has a solution; or
- there exists a vector $y \ge 0$ such that $A^\top y = 0$ and $b^\top y < 0$.

Th.L7.1 Normal cone to the intersection of linear inequalities. (rewriting L3.1,L5.3)
Let $\Omega \subseteq \mathbb{R}^n$ be defined as the intersection of m linear inequalities

$$\Omega := \left\{ x \in \mathbb{R}^n : \begin{array}{ll} a_i^\top x = b_i & \forall i = 1, \dots, r \\ c_j^\top x \le d_j & \forall j = 1, \dots, s \end{array} \right\}$$

Given a point $x \in \Omega$, define the index set of the "active" inequality constraints

$$I(x) := \left\{ j \in \{1, \dots, s\} : c_j^\top x = d_j \right\}.$$

Then, the normal cone at any $x \in \Omega$ is given by $\mathcal{N}_\Omega(x) =$

$$
\begin{aligned}
&= \left\{ \sum_{i=1}^{r} \mu_i a_i + \sum_{j \in I(x)} \lambda_j c_j : \mu_i \in \mathbb{R}, \lambda_j \in \mathbb{R}_{\geq 0} \right\} \\
&= \Bigg\{ \sum_{i=1}^{r} \mu_i a_i + \sum_{j=1}^{s} \lambda_j c_j : \\
&\qquad \mu_i \in \mathbb{R}, \lambda_j \in \mathbb{R}_{\geq 0}, \\
&\qquad \lambda_j \left(d_j - c_j^\top x \right) = 0 \ \forall j = 1, \ldots, s \Bigg\},
\end{aligned}
$$

where the second equality simply rewrites the condition $j \in I(x)$ via complementary slackness (see Lecture 3).