6.7220 Nonlinear Optimization by ngorlo

Th. Triangle Inequality:
$$\|x+y\| \le \|x\| + \|y\|$$
$$\|b-a\| \le \|c-a\| + \|b-c\|$$
Cauchy-Schwarz Inequality:
$$|\langle x,y\rangle| \le \|x\| \cdot \|y\|$$
$$|\langle x,y\rangle|^2 \le \langle x,x\rangle \cdot \langle y,y\rangle$$
$$\left(\sum_{i=1}^n u_i v_i\right)^2 \le \left(\sum_{i=1}^n u_i^2\right)\left(\sum_{i=1}^n v_i^2\right)$$
Hölder's Inequality: Let $p,q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, then
$$\|f \cdot g\|_1 \le \|f\|_p \cdot \|g\|_q$$
Jensen's Inequality: Let f be convex, finite set of points $x_i \in \mathbb{R}^n$, weights $\alpha_i \ge 0$, $\sum_{i=1}^n \alpha_i = 1$, then
$$f\left(\sum_{i=1}^n \alpha_i x_i\right) \le \sum_{i=1}^n \alpha_i f(x_i)$$
AM-GM Inequality: Let $x_1, x_2, \ldots, x_n \in \mathbb{R}_{\ge 0}$, then
$$\frac{x_1 + x_2 + \ldots + x_n}{n} \ge \sqrt[n]{x_1 x_2 \ldots x_n}$$
Lipschitz Continuity: A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous with constant L if
$$\|f(x) - f(y)\| \le L\|x-y\|$$
Directional Derivative:: Let $f : \mathbb{R}^n \to \mathbb{R}$,
$$f'(x;d) = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} = \langle \nabla f(x), d\rangle$$
Product Rule:: Let $f : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable, then
$$\nabla(f^\top g) = f^\top \nabla g + g^\top \nabla f$$
Chain Rule:: Let $f : \mathbb{R} \to \mathbb{R}^m$, $g : \mathbb{R}^n \to \mathbb{R}$ be differentiable and let $h(t) = g(f(t))$, then
$$h'(t) = \nabla g^\top (f(t)) f'(t)$$
in particular, let $x, y \in \mathbb{R}^n$, and define $h(t) = g(x+ty)$, where $g : \mathbb{R}^n \to \mathbb{R}$.
$$h'(t) = \nabla g(x+ty)^T y = y^T \nabla g(x+ty)$$
Square decomposition:
$$\|x+y\|^2 = \|x\|^2 + 2\langle x,y\rangle + \|y\|^2$$
Cholesky Decomposition:
$$A = LL^T,$$
where L is lower triangular, A is symmetric positive definite.
Eigendecomposition:
$$A = Q\Lambda Q^\top,$$
where Q is orthogonal, Λ is diagonal. columns of Q are eigenvectors of A, A is symmetric.
Boundedness: Set is bounded iff $\exists r > 0$ s.t. $S \subseteq B_r = \{x : \|x\|_2 \le r\}$. **Openness**: for every $x \in S$, $\exists \epsilon > 0$ s.t. $B_\epsilon(x) \subseteq S$. **Closeness**: S contains all its limit points. Note: a set can be open, closed, neither or both.

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
$$\mathcal{N}_\Omega(x) := \left\{d \in \mathbb{R}^n : \langle d, y - x\rangle \le 0 \quad \forall y \in \Omega\right\}.$$
With this definition, the first-order necessary optimality condition for x, given in Th L2.1, can be equivalently written as
$$-\nabla f(x) \in \mathcal{N}_\Omega(x)$$
Normal cone at interior $N_\Omega(x) = \{0\}$ (consider $y = x + \delta d$ and realize $\langle d, y-x\rangle > 0$).

Th.L2.2 Normal cone to a hyperplane. Consider a hyperplane
$$\Omega := \{y \in \mathbb{R}^n : \langle a, y\rangle = b\},$$
where $a \in \mathbb{R}^n, a \ne 0$, and a point $x \in \Omega$. The normal cone at x is given by
$$\mathcal{N}_\Omega(x) = \operatorname{span}\{a\} = \{\lambda a : \lambda \in \mathbb{R}\}.$$
(consider $y \in \operatorname{span}\{a\}$ and $z = (y-x) = k \cdot a \notin \operatorname{span}\{a\}$) Similarly, for affine subspaces $\Omega = \{y \in \mathbb{R}^n : Ay = b\}$, the normal cone is $\mathcal{N}_\Omega(x) = \operatorname{colspan}\{A^\top\} = \{A^\top \lambda : \lambda \in \mathbb{R}^n\}$. (project $x+z$ onto Ω and $\operatorname{colspan}(A^\top)$)

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 $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$:
$$\mathcal{N}_\Omega(x) = \left\{\sum_{j=1}^m \lambda_j a_j : \sum_{j=1}^m \lambda_j \left(b_j - \langle a_j, x\rangle\right) = 0;\ \lambda_j \ge 0\right\}$$
$$= \left\{A^\top \lambda : \lambda^\top (b - Ax) = 0, \lambda \in \mathbb{R}^m_{\ge 0}\right\},$$
$$A = \begin{pmatrix} -a_1^\top - \\ \vdots \\ -a_m^\top - \end{pmatrix},$$
rewriting the condition $j \in I(x)$ (via **complementary slackness**).

Rm.L3.3 Primal and Dual Consider the linear program
$$\begin{aligned} \max_x \quad & c^\top x \\ \text{s.t.} \quad & Ax \le b \\ & x \in \mathbb{R}^n, \end{aligned} \tag{P}$$
$$\begin{aligned} \min_\lambda \quad & g(\lambda) := b^\top \lambda \\ \text{s.t.} \quad & A^\top \lambda = c \\ & \lambda \ge 0 \end{aligned} \tag{D}$$

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 $(\lambda^*)^\top (b - Ax^*) = 0$.

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

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. for all $x, y \in \Omega, t \in [0,1]$:
$$f((1-t) \cdot x + t \cdot y) \le (1-t) \cdot f(x) + t \cdot f(y)$$
$$f(x + t(y-x)) \le f(x) + t(f(y) - f(x))$$
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]

(1) $\Rightarrow$ (2) **Bounding by linearization**: Divide by t and take limit.
(2) $\Rightarrow$ (1) Sum linearization bounds (multiplied by t and $1-t$ accordingly) centered in point $z := t \cdot x + (1-t) \cdot y$ and look in directions $x - z$ and $y - z$.
(2) $\Rightarrow$ (3) Write condition (2) for (x,y) and (y,x) and take sum.
(3) $\Rightarrow$ (4) Consider $x_t := x + t \cdot (y-x)$ and plug into (3) for y. Divide by t^2, take limit.
(4) $\Rightarrow$ (3) Use $0 \le \langle y - x, \nabla^2 f(x + \tau(y-x)) \cdot (y-x)\rangle$ and integrate $\tau \in [0,1]$.
(3) $\Rightarrow$ (2) Define x_t as above, integrate $t \cdot \langle \nabla f(x_t) - \nabla f(x), y - x\rangle$ for $t \in [0,1]$.

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 $f \star_{\inf} g$ of two convex functions $f, g : \mathbb{R}^n \to \mathbb{R}$, defined as
$$\inf\left\{f(y) + g(x-y) : y \in \mathbb{R}^n\right\}$$

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 $t \in (0,1)$,
$$f((1-t) \cdot x + t \cdot y) < (1-t) \cdot f(x) + t \cdot f(y)$$
- For f twice differentiable and Ω open, $\nabla^2 f(x) \succ 0\ \forall x \in \Omega$ is sufficient for strict convexity.
- 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.
- For f twice differentiable and Ω open, strong convexity is equivalent to $\nabla^2 f(x) \succeq \mu I\ \forall x \in \Omega$.
- For f twice differentiable and Ω open, strong convexity is equivalent to $f(y) \ge f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y - x\|_2^2\ \forall x, y \in \Omega$.

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. (Set $f(x) = f(y) = f^\star$ and use strict convexity.)
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
$$\langle u, y\rangle < v, \quad \text{and} \quad \langle u, x\rangle \ge v \quad \forall x \in \Omega.$$
(Let $x^\star$ be projection of y on Ω, then $u = x^\star - y$ and $v = \langle u, x^\star\rangle$. **Note:** For strict inequality let $v' := \frac{1}{2}(\langle u, x'\rangle + \langle u, y\rangle)$)

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 $y \in S$ 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
$$\langle u, y\rangle < \langle u, x\rangle \quad \forall x \in \Omega$$

Note: separation oracle for convex polytope: If $Ay \le b$ return "$y \in \Omega$", otherwise return $(y \notin \Omega, -a)$, where a is a violated constraint.

Rm.L5.4 Ellipsoid method. at each iteration, start at a point c_t and a search space (that initially is $\Omega_1 = \Omega$).
- if c_t is not in the search space, then use a separating hyperplane to cut the search space in half
- if c_t is in the search space, intersect the search space with $H_t = \{x \in \mathbb{R}^n : \langle \nabla f(c_t), x - c_t\rangle \le 0\}$ (to make the search space contain only points x such that $f(x) \ge f(c_t)$ by Th.L4.3)
- set $\Omega_{t+1} = H_t \cap \Omega_t$

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

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 $b^\top y < 0$.

(Let $\Omega = \{Ax + s : x \in \mathbb{R}^n, s \in \mathbb{R}^m_{\ge 0}\}$. If $b \in \Omega$, $Ax \le b$ has a solution. Else, apply separation, first set $s = x = 0$: $v \le 0$, $\langle u, b\rangle < 0$, then $s = 0$: $\langle A^\top u, x\rangle \ge v$, finally set $x = 0$, $s = ke_i \ge 0$)

Rm.L7.1 Optimization Problem with differentiable functional constraints. We consider an optimization problem with constraint set defined as the intersection of differentiable functional constraints:
$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & h_i(x) = 0 \quad i \in \{1, \ldots, r\} \\ & g_j(x) \le 0 \quad j \in \{1, \ldots, s\}. \end{aligned} \tag{1, P}$$

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
$$\Omega := \left\{x \in \mathbb{R}^n : \begin{array}{ll} a_i^\top x = b_i & \forall i = 1, \ldots, r \\ c_j^\top x \le d_j & \forall j = 1, \ldots, s \end{array}\right\}$$
Given a point $x \in \Omega$, define the index set of the "active" inequality constraints
$$I(x) := \left\{j \in \{1, \ldots, s\} : c_j^\top x = d_j\right\}.$$
Then, the normal cone at any $x \in \Omega$ is given by $\mathcal{N}_\Omega(x) =$
$$= \left\{\sum_{i=1}^r \mu_i a_i + \sum_{j \in I(x)} \lambda_j c_j : \mu_i \in \mathbb{R}, \lambda_j \in \mathbb{R}_{\ge 0}\right\}$$
$$= \left\{\sum_{i=1}^r \mu_i a_i + \sum_{j=1}^s \lambda_j c_j : \mu_i \in \mathbb{R}, \lambda_j \in \mathbb{R}_{\ge 0},\ \lambda_j\left(d_j - c_j^\top x\right) = 0\ \forall j = 1, \ldots, s\right\},$$
where the second equality simply rewrites the condition $j \in I(x)$ via complementary slackness (see L3.1).

Def.L7.1 KKT conditions. Consider a nonlinear optimization problem with differentiable objective function and functional constraints, in the form given in (1), and let x be a point in the feasible set ("Primal Feasibility"). The KKT conditions at x are given by

- **Stationarity**: $-\nabla f(x) = \sum_{i=1}^r \mu_i \nabla h_i(x) + \sum_{j=1}^s \lambda_j \nabla g_j(x)$
- **Dual feasibility**: $\mu_i \in \mathbb{R}, \quad \lambda_j \geq 0 \quad \forall i = 1,\ldots,r, \quad j = 1,\ldots,s.$
- **Complementary slackness**: $\lambda_j \cdot g_j(x) = 0 \forall j = 1,\ldots,s.$

(Note: Example of failure of KKT: $f(x) = x, g(x) = x^2$)

Th.L7.2 Concave and linear constraints. Let $x \in \Omega \subseteq \mathbb{R}^n$ be a minimizer of (1). If

- the binding inequality constraints $\{g_j\}_{j \in I(x)}$ are concave differentiable functions in a convex neighborhood of x; and
- the equality constraints $\{h_i\}_{i=1}^r$ are affine functions on $\mathbb{R}^n$,

then the KKT conditions hold at x (Necessity of KKT conditions).

Th.L7.3 Linear independence of gradients. Let $x \in \Omega \subseteq \mathbb{R}^n$ be a minimizer of (1). If all functions h_i, g_j are continuously differentiable and the multiset of gradients at x of all active constraints

$$\{\nabla h_i(x) : i = 1,\ldots,r\} \cup \{\nabla g_j(x) : j \in I(x)\}$$

is linearly independent, then the KKT conditions hold at x.

Th.L7.4 Slater's condition. Let $x \in \Omega \subseteq \mathbb{R}^n$ be a minimizer of (1). If

- the binding inequality constraints $\{g_j\}_{j \in I(x)}$ are convex differentiable functions; and
- the equality constraints $\{h_i\}_{i=1}^r$ are affine functions; and
- there exists a feasible point x_0 that is strictly feasible for the binding inequality constraints, that is,
$$g_j(x_0) < 0 \quad \forall j \in I(x)$$
then the KKT conditions hold at x.

Th.L7.5 Necessity and sufficiency of KKT conditions. If f is convex and the constraints satisfy Slater's condition, then the KKT conditions are both necessary and sufficient for optimality. (Note: Proof by checking KKT conditions, then formulate Lagrangian, and conclude with $f(x) \geq L(x) = L(x^\star) = f(x^\star)$)

Rm.L8.1 Equivalence at Optimality The following statements are equivalent:

- The point $x^\star \in \Omega$ is optimal for (P).
- the point $x^\star$ admits $\mu^\star$, $\lambda^\star$ such that the KKT conditions hold.
- there exist $\mu^\star$, $\lambda^\star$ such that $x^\star$ is minimizer of Lagrangian.

Note: the Lagrangian is:

$$\mathcal{L}(x;\lambda,\mu) = f(x) + \sum_{i=1}^r \lambda_i h_i(x) + \sum_{j=1}^s \mu_j g_j(x)$$

Th.L8.1 Existence of finite penalization coefficients. If the constrained problem (P) has a minimizer $x^* \in \Omega$ and attains optimal value value (P), there exist concrete penalization coefficients $(\lambda^*, \mu^*) \in \mathbb{R}_{\geq 0}^r \times \mathbb{R}^s$ such that

1. x^* is a minimizer of $x \mapsto \mathcal{L}(x;\lambda^*,\mu^*)$ over $\mathbb{R}^n$; and
2. the value of $\min_{x \in \mathbb{R}^n} \mathcal{L}(x;\lambda^*,\mu^*)$ is exactly equal to value (P).

Conversely, if there is a triple $(x^*;\lambda^*,\mu^*) \in \mathbb{R}^n \times (\mathbb{R}_{\geq 0}^r \times \mathbb{R}^s)$ such that x^* is a minimizer of $\mathcal{L}(x;\lambda^*,\mu^*)$, and it satisfies $x^* \in \Omega, \lambda_j^* g_j(x^*) = 0$ for all $j = 1,\ldots,r$, then x^* is also an optimal solution of (P).

Th.L8.2 Weak duality. For any choice of penalization coefficients $\lambda \in \mathbb{R}_{\geq 0}^r, \mu \in \mathbb{R}^s$,

$$\inf_x \mathcal{L}(x;\lambda,\mu) \leq f(\tilde{x}) \quad \forall \tilde{x} \in \Omega$$

In fact, the inequality holds for any minimization problem with functional constraints – that is, even ignoring the requirement that g_j is convex, that h_i is affine, and any constraint qualification.
As a direct consequence, if (P) admits an optimal solution, then

$$\inf_x \mathcal{L}(x;\lambda,\mu) \leq \text{ value } (P)$$

Corollary (Strong Duality):
Corollary L8.1 (Strong duality). If (P) admits a minimizer, the optimization problem

$$(D) := \begin{cases} \max_{\mu,\lambda} & \inf_{x \in \mathbb{R}^n} \mathcal{L}(x;\lambda,\mu) \\ \text{s.t.} & \lambda \in \mathbb{R}_{\geq 0}^r \\ & \mu \in \mathbb{R}^s \end{cases}$$

admits an optimal solution λ^*, μ^*, and matches the value of the original problem (P).

Rm.L8.3 Strong Duality statement. If (P) has an optimal solution, then

$$\text{value}(D) = \max_{\lambda \in \mathbb{R}_{\geq 0}^r, \mu \in \mathbb{R}^s} \inf_{x \in \mathbb{R}^n} \mathcal{L}(x;\lambda,\mu) =$$
$$= \text{value}(P) = \min_{x \in \mathbb{R}^n} \sup_{\lambda \in \mathbb{R}_{\geq 0}^r, \mu \in \mathbb{R}^s} \mathcal{L}(x;\lambda,\mu).$$

Rm.L9.1 Conic Optimization Problem. Feasible set is intersection between affine subspace and nonempty closed convex cone $\mathcal{K}$:

$$\begin{array}{ll} \min_x & f(x) \\ \text{s.t.} & Ax = b \\ & x \in \mathcal{K} \end{array}$$

Def.L9.1 Lorentz cone. The ice-cream cone, or Lorentz cone, is defined as

$$\mathcal{L}^n := \{(x,z) \in \mathbb{R}^{n-1} \times \mathbb{R} : z \geq \|x\|_2\}$$

Def.L9.2 Semidefinite cone. The semidefinite cone $\mathcal{S}^n$ is the set of positive semidefinite $n \times n$ matrices:

$$\mathcal{S}^n := \{X \in \mathbb{S}^n : X \succeq 0\}$$
$$= \{X \in \mathbb{S}^n : a^\top X a \geq 0 \forall a \in \mathbb{R}^n\},$$

where $\mathbb{S}^n$ is the set of symmetric $n \times n$ real matrices.

Def.L9.3 Copositive cone. The copositive cone $\mathcal{C}^n$ is the set of symmetric $n \times n$ real matrices:

$$\mathcal{C}^n := \{X \in \mathbb{S}^n : a^\top X a \geq 0 \quad \forall a \in \mathbb{R}_{\geq 0}^n\}.$$

The difference with the positive semidefinite cone $\mathcal{S}^n$ lies in the fact that we need $a^\top X a \geq 0$ only for nonnegative vectors $a \in \mathbb{R}_{\geq 0}^n$, instead of all $a \in \mathbb{R}^n$.

Th.L9.1 Normal cone to the intersection of a hyperplane and a closed convex set. Let $H := \{x \in \mathbb{R}^n : Ax = b$, with $A \in \mathbb{R}^{m \times n}\}$ be an affine subspace, and S be a closed convex set (not necessarily a cone) such that $S^\circ \cap H$ is nonempty, where S° is the interior of S. For any $x \in H \cap S$,

$$\mathcal{N}_{H \cap S}(x) = \mathcal{N}_H(x) + \mathcal{N}_S(x).$$

So, using the fact (see Lecture 2) that $\mathcal{N}_H(x) = \operatorname{colspan}(A^\top)$, we have $\mathcal{N}_{H \cap S}(x) = \{A^\top \mu + z : \mu \in \mathbb{R}^m, z \in \mathcal{N}_S(x)\}$.

Rm.L9.1 Polar and dual cone. The polar cone of $\mathcal{K}$, often denoted $\mathcal{K}^\perp$, is exactly the normal cone at 0 :

$$\mathcal{K}^\perp := \mathcal{N}_\mathcal{K}(0) = \{d : \langle d, y\rangle \leq 0 \quad \forall y \in \mathcal{K}\}.$$

The dual cone of $\mathcal{K}$, often denoted $\mathcal{K}^\star$, is the opposite of the normal cone at 0 :

$$\mathcal{K}^\star := -\mathcal{K}^\perp = -\mathcal{N}_\mathcal{K}(0) = \{d : \langle d, y\rangle \geq 0 \quad \forall y \in \mathcal{K}\}$$

Th.L9.2 Self-duality of cones.

1. The nonnegative cone, the Lorentz cone, and the semidefinite cones are all self-dual cones, that is, $\mathcal{K}^\star = \mathcal{K}$.
2. The dual cone to the copositive cone $\mathcal{C}^n$ is called the totally positive cone, defined as
$$\mathcal{P}^n := \left\{\sum_{i=1}^k z_i z_i^\top : z_i \in \mathbb{R}_{\geq 0}^n, k \in \mathbb{N}\right\}$$
$$= \{BB^\top : B \in \mathbb{R}_{\geq 0}^{n \times m}, m \in \mathbb{N}\}$$

Th.L9.3 Normal cone to a closed convex cone. Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a nonempty, closed convex cone. The normal cone at any point $x \in \mathcal{K}$ is given by

$$\mathcal{N}_\mathcal{K}(x) = \{d \in \mathcal{K}^\perp : \langle d, x\rangle = 0\}.$$

Th.L9.4 Strictly feasible conic optimization. If f is convex and differentiable, and the conic optimization (2) is strictly feasible, that is, there exists $x \in \mathcal{K}^\circ$ such that $Ax = b$, then a point $x^\star \in \Omega := \mathcal{K} \cap \{x : Ax = b\}$ is a minimizer of f on Ω if and only if

$$-\nabla f(x^\star) \in \{A^\top \mu + z : \mu \in \mathbb{R}^m, z \in \mathcal{N}_\mathcal{K}(x^\star)\}$$

that is, expanding $\mathcal{N}_\mathcal{K}(x^\star)$ using Theorem L9.3, if and only if

$$-\nabla f(x^\star) \in \{A^\top \mu + z : \mu \in \mathbb{R}^m, z \in \mathcal{K}^\perp, \langle z, x^\star\rangle = 0\}$$
$$\Leftrightarrow$$
$$-\nabla f(x^\star) \in \{A^\top \mu - z : \mu \in \mathbb{R}^m, z \in \mathcal{K}^\star, \langle z, x^\star\rangle = 0\}$$

Th.L9.5 Conic duality for linear objectives. Theorem L9.5. If the primal problem

$$\min_x \langle c, x\rangle$$
$$\text{s.t. } Ax = b$$
$$x \in \mathcal{K}$$

is strictly feasible and admits an optimal solution $x^\star$, then the dual problem

$$\begin{array}{ll} \max_{\mu,z} & \langle b, \mu\rangle \\ \text{s.t.} & z = c - A^\top \mu \\ & \mu \in \mathbb{R}^m \\ & z \in \mathcal{K}^\star \end{array}$$

admits an optimal solution $(\mu^\star, z^\star)$ such that:

- the values of the two problems coincide: $\langle c, x^\star\rangle = \langle b, \mu^\star\rangle$; and
- the solution $z^\star$ satisfies the complementary slackness condition $\langle x^\star, z^\star\rangle = 0$.

Rm.L10.1 Polynomial optimization problems. We consider polynomial optimization problems of the form:

$$\begin{array}{ll} \min_x & f(x) \\ \text{s.t.} & g_j(x) \geq 0 \quad j = 1,\ldots,m \\ & x \in \mathbb{R}^n, \end{array}$$

Def.L10.1 Cone of nonnegative polynomials. The cone of nonnegative polynomials is the set of all polynomials $p \in \mathbb{R}[x]$ such that $p(x) \geq 0$ for all $x \in \mathbb{R}^n$. We will denote this set by $\mathbb{R}[x]_{\geq 0}$.

Th.L10.1 Deciding membership in the cone of nonnegative polynomials $\mathbb{R}[x]_{\geq 0}$ for a polynomial of degree ≥ 4 with $n \geq 2$ variables is computationally intractable.

Def.L10.2 Sum-of-squares polynomials, $\Sigma[x]$ An n-variate polynomial $p \in \mathbb{R}[x]$ is said to be a sum-of-squares polynomial (or SOS for short) if it can be written as

$$p(x) = \sum_k p_k(x)^2, \quad x \in \mathbb{R}^n$$

for appropriate polynomials $p_k \in \mathbb{R}[x]$.

Th.L10.2 Membership in the cone of sum-of-squares polynomials can be determined by checking the feasibility of a semidefinite program. In particular, let $p(x)$ be an arbitrary polynomial of degree $2d$ in n variables, and let v_d be the vector of all monomials of degree up to d that can be constructed using the variables x. Then, $p(x) \in \Sigma[x]$ if and only if there exists a positive semidefinite matrix Q such that

$$p(x) = (v_d(x))^\top Q v_d(x)$$

Corollary: The cone of n-variate sum-of-squares polynomials of degree $2d$ is a linear transformation of the cone of $s \times s$ positive semidefinite matrices, where $s = \binom{n+d}{d}$ is the dimension of the vector $v_d(x)$.

Th.L10.3 Hilbert's Theorem $\mathbb{R}[x]_{\geq 0} = \Sigma[x]$ if and only if:

- $n = 1$, no matter the degree d; or
- $d = 2$, no matter the number of variables n; or
- $n = 2$ and $d = 4$.

Th.L10.4 Putinar's Positivstellensatz Let

$$\Omega := \{x \in \mathbb{R}^n : g_j(x) \geq 0 \quad \forall j = 1,\ldots,m\}$$

where $\{x \in \mathbb{R}^n : g_j(x) \geq 0\}$ is compact for at least one $j \in \{1,\ldots,m\}$. Any polynomial $p \in \mathbb{R}[x]$ positive on Ω can be written in the form

$$p = \sigma_0 + \sum_{j=1}^m \sigma_j g_j$$

for appropriate SOS polynomials $\sigma_j \in \Sigma[x], j = 0,\ldots,m$.

Def.L11.X The depth function Let's define $d_y : \Omega \to \mathbb{R}$,

$$d_y(x) := \max\left\{\alpha \in \mathbb{R}_{\geq 0} : x + \alpha \frac{y - x_0}{\|y - x_0\|_2} \in \Omega\right\}$$

The depth function d_y measures how far we can move from x in the direction of $y - x_0$ while staying in Ω.

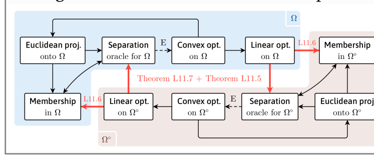
Th.L11.1-3 Properties of the depth function

- For any given $x \in \Omega$, we have $d_y(x) \in [0, 2R]$. Furthermore, we can compute $d_y(x)$ up to $\epsilon > 0$ error using $O(\log(R/\epsilon))$ calls to a membership oracle for Ω.
- The function d_y is concave (i.e., $-d_y$ is convex).
- Theorem L11.3. The function d_y restricted to the domain $\mathbb{B}_{r/3}(x_0)$ is $(3R/r)$-Lipschitz continuous.

Th.L11.4 Constructing supporting hyperplane with depth function If d_y is differentiable at x_0, the gradient $\nabla d_y(x_0)$ provides a separating direction, that is,

$$\langle \nabla d_y(x_0), y - x\rangle < 0 \quad \forall x \in \Omega$$

Rm.L11.3 Polarity relationships. Connections among oracles between set Ω and its polar Ω°:



Def.L11.1 Polar set Ω°. Let Ω be compact and convex, and such that $\mathbb{B}_r(0) \subseteq \Omega \subseteq \mathbb{B}_R(0)$ for some radii $0 < r \leq R$. The polar set Ω° to Ω is defined as

$$\Omega^\circ := \{y \in \mathbb{R}^n : \langle x, y\rangle \leq 1 \quad \forall x \in \Omega\}$$

Th.L11.5 Involution of the polar set. Theorem L11.5. Let the set $\Omega \subseteq \mathbb{R}^n$ be convex and compact, and such that $\mathbb{B}_r(0) \subseteq \Omega \subseteq \mathbb{B}_R(0)$ for some $0 < r \leq R$. Then,

1. the polar set Ω° is convex and compact, and satisfies $\mathbb{B}_{1/R}(0) \subseteq \Omega^\circ \subseteq \mathbb{B}_{1/r}(0)$; and
2. the bipolar $(\Omega^\circ)^\circ$ is equal to Ω.

Th.L11.6 Polar membership oracle. A membership oracle for Ω° can be constructed efficiently starting from a linear optimization oracle for Ω, even if the linear optimization oracle only returns the optimal objective value and not the minimizer. The construction only requires a single call to the linear optimization oracle.

Th.L11.7 Separation oracle for the polar set. A separation oracle for Ω° can be constructed efficiently, without use of Section L11.2, starting from a linear optimization oracle for Ω that returns the optimal point. The construction only requires a single call to the linear optimization oracle.

Rm.X Additional notes: with matrices, $Ax = b \Leftrightarrow \langle A_k, X\rangle = b_k \ \forall k = 1,\ldots,m$. Where $\langle A_k, X\rangle = \operatorname{tr}(AB^\top) = \sum_i \sum_j A_{ij} B_{ij}$.

Nonnegative Cone $\to$ Linear Programs
Lorentz Cone $\to$ Second Order Conic Programs

Failure modes of constraint qualification: Primal has optimal solution, dual matches value of primal, yet dual does not have maximizer. Primal has optimal solution, dual has optimal solution, but values of problems differ.