Semidefinite Optimization and Relaxation

Heng Yang

2024-01-27

Contents

Preface

This is the textbook for Harvard ENG-SCI 257: Semidefinite Optimization and Relaxation.

Feedback

I would like to invite you to provide comments to the textbook via the following two ways:

- Inline comments with Hypothesis:
    - Go to Hypothesis and create an account
    - Install the Chrome extension of Hypothesis
    - Provide public comments to textbook contents and I will try to address them

- Blog-style comments with Disqus:
    - At the end of each Chapter, there is a Disqus module where you can leave feedback

I would recommend using Disqus for high-level and general feedback regarding the entire Chapter, but using Hypothesis for feedback and questions about the technical details.

Offerings

Information about the offerings of the class is listed below.

2024 Spring

Time: Mon/Wed 2:15 - 3:30pm

Location: Science and Engineering Complex, 1.413

Instructor: Heng Yang

Teaching Fellow: Safwan Hossain

Syllabus

Notation

We will use the following standard notation throughout this book.

Basics

$\mathbb{R}$	real numbers
$\mathbb{R}_+$	nonnegative real
$\mathbb{R}_{++}$	positive real
$\mathbb{Z}$	integers
$\mathbb{N}$	nonnegative integers
$\mathbb{N}_+$	positive integers
$\mathbb{R}^n$	n-D column vector
$\mathbb{R}^n_+$	nonnegative orthant
$\mathbb{R}^n_{++}$	positive orthant
e_i	standard basic vector
$\Delta_n := \{x \in \mathbb{R}^n_+ \mid \sum x_i = 1\}$	standard simplex

Matrices

$\mathbb{R}^{m \times n}$	$m \times n$ real matrices
$\mathbb{S}^n$	$n \times n$ symmetric matrices
$\mathbb{S}^n_+$	$n \times n$ positive semidefinite matrices
$\mathbb{S}^n_{++}$	$n \times n$ positive definite matrices
$\langle A, B \rangle$ or $\bullet$	inner product in $\mathbb{R}^{m \times n}$
$\mathrm{tr}(A)$	trace of $A \in \mathbb{R}^{n \times n}$
$A^\top$	matrix transpose
$\det(A)$	matrix determinant
$\mathrm{rank}(A)$	rank of a matrix
$\mathrm{diag}(A)$	diagonal of a matrix A as a vector

$\mathrm{Diag}(a)$	turning a vector into a diagonal matrix
$\mathrm{BlkDiag}(A, B, \dots)$	block diagonal matrix with blocks $A, B, \dots$
$\succeq 0$ and $\preceq 0$	positive / negative semidefinite
$\succ 0$ and $\prec 0$	positive / negative definite
$\lambda_{\max}$ and $\lambda_{\min}$	maximum / minimum eigenvalue
$\sigma_{\max}$ and $\sigma_{\min}$	maximum / minimum singular value
$\mathrm{vec}(A)$	vectorization of $A \in \mathbb{R}^{m \times n}$
$\mathrm{svec}(A)$	symmetric vectorization of $A \in \mathbb{S}^n$
$\|A\|_{\mathrm{F}}$	Frobenius norm
$\mathrm{Range}(A)$	span of the column vectors
$\ker(A)$	right null space

Geometry

$\|a\|_p$	p-norm
$\|a\|$	2-norm
$B(o, r)$	ball with center o and radius r
$\mathrm{aff}(S)$	affine hull of set S
$\mathrm{conv}(S)$	convex hull of set S
$\mathrm{cone}(S)$	conical hull of set S
$\mathrm{int}(S)$	interior of set S
$\mathrm{ri}(S)$	relative interior of set S
∂S	boundary of set S
P°	polar of convex body
P^*	dual of set P
$\mathrm{SO}(d)$	special orthogonal group of dimension d
$\mathcal{S}^{d-1}$	unit sphere in $\mathbb{R}^d$

Optimization

KKT	Karush–Kuhn–Tucker
LP	linear program

QP	quadratic program
SOCP	second-order cone program
SDP	semidefinite program

Algebra

$\mathbb{R}[x]$	polynomial ring in x with real coefficients
deg	degree of a monomial / polynomial
$\mathbb{R}[x]_d$	polynomials in x of degree up to d
$[x]_d$	vector of monomials of degree up to d
$[\![x]\!]_d$	vector of monomials of degree d

Chapter 1

Mathematical Background

1.1 Convexity

A very important notion in modern optimization is that of *convexity*. To a large extent, an optimization problem is "easy" if it is convex, and "difficult" when convexity is lost, i.e., *nonconvex*. We give a basic review of convexity here and refer the reader to (Rockafellar, 1970), (Boyd and Vandenberghe, 2004), and (Bertsekas et al., 2003) for comprehensive treatments.

We will work on a finite-dimensional real vector space, which we will identify with $\mathbb{R}^n$.

Definition 1.1 (Convex Set)**.** A set S is convex if $x_1, x_2 \in S$ implies $\lambda x_1 + (1 - \lambda)x_2 \in S$ for any $\lambda \in [0, 1]$. In other words, if $x_1, x_2 \in S$, then the line segment connecting x_1 and x_2 lies inside S.

Conversely, a set S is nonconvex if Definition 1.1 does not hold.

Given $x_1, x_2 \in S$, $\lambda x_1 + (1 - \lambda)x_2$ is called a *convex combination* when $\lambda \in [0, 1]$. For convenience, we will use the following notation

$$
\begin{aligned}
(x_1, x_2) &= \{\lambda x_1 + (1 - \lambda)x_2 \mid \lambda \in (0, 1)\}, \\
[x_1, x_2] &= \{\lambda x_1 + (1 - \lambda)x_2 \mid \lambda \in [0, 1]\}.
\end{aligned}
\tag{1.1}
$$

A **hyperplane** is a common convex set defined as

$$
H = \{x \in \mathbb{R}^n \mid \langle c, x \rangle = d\}
\tag{1.2}
$$

for some $c \in \mathbb{R}^n$ and scalar d. A **halfspace** is a convex set defined as

$$
H^+ = \{x \in \mathbb{R}^n \mid \langle c, x \rangle \geq d\}.
\tag{1.3}
$$

Given two nonempty convex sets C_1 and C_2, the **distance** between C_1 and C_2 is defined as

$$\mathrm{dist}(C_1, C_2) = \inf\{\|c_1 - c_2\| \mid c_1 \in C_1, c_2 \in C_2\}. \tag{1.4}$$

For a convex set C, the hyperplane H in (1.2) is called a **supporting hyperplane** for C if C is contained in the half space H^+ and the distance between H and C is zero. For example, the hyperplane $x_1 = 0$ is supporting for the hyperboloid $\{(x_1, x_2) \mid x_1 x_2 \geq 1, x_1 \geq 0, x_2 \geq 0\}$ in $\mathbb{R}^2$.

An important property of a convex set is that we can *certify* when a point is not in the set. This is usually done via a separation theorem.

Theorem 1.1 (Separation Theorem)**.** *Let S_1, S_2 be two convex sets in $\mathbb{R}^n$ and $S_1 \cap S_2 = \emptyset$, then there exists a hyperplane that separates S_1 and S_2, i.e., there exists c and d such that*

$$\begin{aligned} \langle c, x \rangle &\geq d, \forall x \in S_1, \\ \langle c, x \rangle &\leq d, \forall x \in S_2. \end{aligned} \tag{1.5}$$

Further, if S_1 is compact (i.e., closed and bounded) and S_2 is closed, then the separation is strict, i.e., the inequalities in (1.5) are strict.

The strict separation theorem is used typically when S_1 is a single point (hence compact).

We will see a generalization of the separation theorem for nonconvex sets later after we introduce the idea of sums of squares.

Exercise 1.1. Provide examples of two disjoint convex sets such that the separation in (1.5) is not strict in one way and both ways.

Exercise 1.2. Provide a constructive proof that the separation hyperplane exists in Theorem 1.1 when (1) both S_1 and S_2 are closed, and (2) at least one of them is bounded.

The intersection of convex sets is always convex (try to prove this).

1.2 Convex Geometry

1.2.1 Basic Facts

Given a set S, its **affine hull** is the set

$$\mathrm{aff}(S) = \left\{ \sum_{i=1}^{k} \lambda_i u_i \mid \lambda_1 + \cdots + \lambda_k = 1, u_i \in S, k \in \mathbb{N}_+ \right\},$$

where $\sum_{i=1}^{k} \lambda_i u_i$ is called an *affine combination* of $u_1, \ldots, u_k$ when $\sum_i \lambda_i = 1$. The affine hull of a set is the smallest affine subspace that contains S, and the **dimension** of S is the dimension of its affine hull. The affine hull of the emptyset is the emptyset, of a singleton is the singleton itself. The affine hull of a set of two different points is the line going through them. The affine hull of a set of three points not on one line is the plane going through them. The affine hull of a set of four points not in a plane in $\mathbb{R}^3$ is the entire space $\mathbb{R}^3$.

For a convex set $C \subseteq \mathbb{R}^n$, the **interior** of C is defined as

$$\text{int}(C) := \{u \in C \mid \exists \epsilon > 0, B(u, \epsilon) \subseteq C\},$$

where $B(u, \epsilon)$ denotes a ball centered at u with radius ϵ (using the usual 2-norm). Each point in $\text{int}(C)$ is called an *interior point* of C. If $\text{int}(C) = C$, then C is said to be an **open set**. A convex set with nonempty interior is called a **convex domain**, while a compact (i.e., closed and bounded) convex domain is called a **convex body**.

The **boundary of** C is the subset of points that are in the **closure**[1] of C but are not in the interior of C, and we denote it as ∂C. For example, the closed line segment $C = [0, 1]$ has two points on the boundary: 0 and 1; the open line segment $C = (0, 1)$ has the same two points as its boundary.

It is possible that a convex set has empty interior. For example, a hyperplane has no interior, and neither does a singleton. In such cases, the **relative interior** can be defined as

$$\text{ri}(C) := \{u \in C \mid \exists \epsilon > 0, B(u, \epsilon) \cap \text{aff}(C) \subseteq C\}.$$

For a nonempty convex set, the relative interior always exists. If $\text{ri}(C) = C$, then C is said to be **relatively open**. For example, the relative interior of a singleton is the singleton itself, and hence a singleton is relatively open.

For a convex set C, a point $u \in C$ is called an **extreme point** if

$$u \in (x, y), x \in C, y \in C \quad \Rightarrow u = x = y.$$

For example, consider $C = \{(x, y) \mid x^2 + y^2 \leq 1\}$, then all the points on the boundary $\partial C = \{(x, y) \mid x^2 + y^2 = 1\}$ are extreme points.

A subset $F \subseteq C$ is called a **face** if F itself is convex and

$$u \in (x, y), u \in F, x, y \in C \quad \Rightarrow x, y \in F.$$

Clearly, the empty set $\emptyset$ and the entire set C are faces of C, which are called *trivial faces*. The face F is said to be *proper* if $F \neq C$. The set of any single

---

[1]The closure of a subset C of points, denoted $\text{cl}(C)$, consists of all points in C together with all limit points of C. The closure of C may equivalently be defined as the intersection of all closed sets containing C. Intuitively, the closure can be thought of as all the points that are either in C or "very near" C. For example, the closure of the open line segment $C = (0, 1)$ is the closed line segment $C = [0, 1]$.

extreme point is also a face. A face F of C is called **exposed** if there exists a supporting hyperplane H for C such that

$$F = H \cap C.$$

1.2.2 Cones, Duality, Polarity

Definition 1.2 (Polar)**.** For a nonempty set $T \subseteq \mathbb{R}^n$, its polar is the set

$$T^\circ := \{y \in \mathbb{R}^n \mid \langle x, y \rangle \leq 1, \forall x \in T\}. \tag{1.6}$$

The polar T° is a closed convex set and contains the origin. Note that T is always contained in the polar of T°, i.e., $T \subseteq (T^\circ)^\circ$. Indeed, they are equal under some assumptions.

Theorem 1.2 (Bipolar)**.** *If $T \subseteq \mathbb{R}^n$ is a closed convex set containing the origin, then $(T^\circ)^\circ = T$.*

An important class of convex sets are those that are invariant under positive scalings.[2] A set $K \subseteq \mathbb{R}^n$ is a **cone** if $tx \in K$ for all $x \in K$ and for all $t > 0$. For example, the positive real line $\{x \in \mathbb{R} \mid x > 0\}$ is a cone. The cone K is **pointed** if $K \cap -K = \{0\}$. It is said to be **solid** if its interior $\mathrm{int}(K) \neq \emptyset$. Any nonzero point of a cone cannot be extreme. If a cone is pointed, the only extreme point is the origin.

The analogue of extreme point for convex cones is the **extreme ray**. For a convex cone K and $0 \neq u \in K$, the line segment

$$u \cdot [0, \infty) := \{tu \mid t \geq 0\}$$

is called an extreme ray of K if

$$u \in (x, y), x, y \in K \quad \Rightarrow \quad u, x, y \text{ are parallel to each other.}$$

If $u \cdot [0, \infty)$ is an extreme ray, then we say u generates the extreme ray.

Definition 1.3 (Proper Cone)**.** A cone K is proper if it is closed, convex, pointed, and solid.

A proper cone K induces a **partial order** on the vector space, via $x \succeq y$ if $x - y \in K$. We also use $x \succ y$ if $x - y$ is in $\mathrm{int}(K)$. Important examples of proper cones are the nonnegative orthant, the second-order cone, the set of symmetric positive semidefinite matrices, and the set of nonnegative polynomials, which we will describe later in the book.

---

[2]Some authors define a cone using nonnegative scalings.

Definition 1.4 (Dual)**.** The dual of a nonempty set S is

$$S^* := \{y \in \mathbb{R}^n \mid \langle y, x \rangle \geq 0, \forall x \in S\}.$$

Given any set S, its dual S^* is always a closed convex cone. Duality reverses inclusion, that is,

$$S_1 \subseteq S_2 \quad \Rightarrow \quad S_1^* \supseteq S_2^*.$$

If S is a closed convex cone, then $S^{**} = S$. Otherwise, S^{**} is the closure of the smallest convex cone that contains S.

For a cone $K \subseteq \mathbb{R}^n$, one can show that

$$K^\circ = \{y \in \mathbb{R}^n \mid \langle x, y \rangle \leq 0, \forall x \in K\}.$$

The set K° is called the **polar cone** of K. The negative of K° is just the **dual cone**

$$K^* = \{y \in \mathbb{R}^n \mid \langle x, y \rangle \geq 0, \forall x \in K\}.$$

Definition 1.5 (Self-dual)**.** A cone K is self-dual if $K^* = K$.

As an easy example, the nonnegative orthant $\mathbb{R}_+^n$ is self-dual.

Example 1.1 (Second-order Cone)**.** The second-order cone, or the Lorentz cone, or the ice cream cone

$$\mathcal{Q}_n := \{(x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sqrt{x_1^2 + \cdots + x_n^2} \leq x_0\}$$

is a proper cone of $\mathbb{R}^{n+1}$. We will show that it is also self-dual.

Proof. Consider $(y_0, y_1, \ldots, y_n) \in \mathcal{Q}_n$, we want to show that

$$x_0 y_0 + x_1 y_1 + \cdots + x_n y_n \geq 0, \forall (x_0, x_1, \ldots, x_n) \in \mathcal{Q}_n. \tag{1.7}$$

This is easy to verify because

$$x_1 y_1 + \cdots + x_n y_n \geq -\sqrt{x_1^2 + \cdots + x_n^2}\sqrt{y_1^2 + \cdots + y_n^2} \geq -x_0 y_0.$$

Hence we have $\mathcal{Q}_n \subseteq \mathcal{Q}_n^*$.

Conversely, if (1.7) holds, then take

$$x_1 = -y_1, \ldots, x_n = -y_n, \quad x_0 = \sqrt{x_1^2 + \cdots + x_n^2},$$

we have

$$y_0 \geq \sqrt{y_1^2 + \cdots + y_n^2},$$

hence $\mathcal{Q}_n^* \subseteq \mathcal{Q}_n$. ∎

Not every proper cone is self-dual.

Exercise 1.3. Consider the following proper cone in $\mathbb{R}^2$

$$K = \{(x_1, x_2) \mid 2x_1 - x_2 \geq 0, 2x_2 - x_1 \geq 0\}.$$

Show that it is not self-dual.

1.3   Convex Optimization

Definition 1.6 (Convex Function). A function $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \forall \lambda \in [0, 1], \forall x, y \in \mathbb{R}^n.$$

A function f is convex if and only if its **epigraph** $\{(x, t) \in \mathbb{R}^{n+1} \mid f(x) \leq t\}$ is a convex set.

When a function f is differentiable, then there are several equivalent characterizations of convexity, in terms of the gradient $\nabla f(x)$ or the Hessian $\nabla^2 f(x)$.

Theorem 1.3 (Equivalent Characterizations of Convexity). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function. The following propositions are equivalent.*

  i. *f is convex, i.e.,*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \forall \lambda \in [0, 1], x, y \in \mathbb{R}^n.$$

  ii. *The first-order convexity condition holds:*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \forall x, y \in \mathbb{R}^n,$$

  i.e., the hyperplane going through $(x, f(x))$ with slope $\nabla f(x)$ supports the epigraph of f.

  iii. *The second-order convexity condition holds:*

$$\nabla^2 f(x) \succeq 0, \forall x \in \mathbb{R}^n,$$

  i.e., the Hessian is positive semidefinite everywhere.

Let's work on a little exercise.

Exercise 1.4. Which one of the following functions $f : \mathbb{R}^n \to \mathbb{R}$ is not convex?

  a. $\exp(-c^\top x)$, with c constant

  b. $\exp(c^\top x)$, with c constant

  c. $\exp(x^\top x)$

  d. $\exp(-x^\top x)$

1.3.1 Minimax Theorem

Given a function $f : X \times Y \to \mathbb{R}$, the following inequality always holds

$$\max_{y \in Y} \min_{x \in X} f(x, y) \leq \min_{x \in X} \max_{y \in Y} f(x, y). \tag{1.8}$$

If the maximum or minimum is not attained, then (1.8) holds with max / min replaced by sup and inf, respectively.

Exercise 1.5. Provide examples of f such that the inequality in (1.8) is strict.

It is of interest to understand when equality holds in (1.8).

Theorem 1.4 (Minimax Theorem)**.** *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^n$ be compact convex sets, and $f : X \times Y \to \mathbb{R}$ be a continuous function that is convex in its first argument and concave in the second. Then*

$$\max_{y \in Y} \min_{x \in X} f(x, y) = \min_{x \in X} \max_{y \in Y} f(x, y).$$

A special case of this theorem, used in game theory to prove the existence of equilibria for zero-sum games, is when X and Y are standard unit simplicies and the function $f(x, y)$ is bilinear. In a research from our group (Tang et al., 2023), we used the minimax theorem to convert a minimax problem into a single-level minimization problem.

1.3.2 Lagrangian Duality

Consider a nonlinear optimization problem

$$\begin{aligned}
u^{\star} = \min_{x \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \leq 0, i = 1, \dots, m, \\
& h_j(x) = 0, j = 1, \dots, p.
\end{aligned} \tag{1.9}$$

Define the **Lagrangian** associated with the optimization problem (1.9) as

$$\begin{aligned}
L : \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p \quad &\to \quad \mathbb{R}, \\
(x, \lambda, \mu) \quad &\mapsto \quad f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x).
\end{aligned} \tag{1.10}$$

The **Lagrangian dual function** is defined as

$$\phi(\lambda, \mu) := \min_{x \in \mathbb{R}^n} L(x, \lambda, \mu). \tag{1.11}$$

Maximizing this function over the dual variables (λ, μ) yields

$$v^{\star} := \max_{\lambda \geq 0, \mu \in \mathbb{R}^p} \phi(\lambda, \mu) \tag{1.12}$$

Applying the minimax Theorem 1.4, we can see that

$$v^\star = \max_{(\lambda,\mu)} \min_x L(x, \lambda, \mu) \leq \min_x \max_{(\lambda,\mu)} L(x, \lambda, \mu) = u^\star.$$

That is to say solving the dual problem (1.12) always provides a lower bound to the primal problem (1.9).

If the functions f, g_i are convex and h_i are affine, the Lagrangian is convex in x and convex in (λ, μ). To ensure strong duality (i.e., $u^\star = v^\star$), compactness or other **constraint qualifications** are needed. An often used condition is the Slater constraint qualification.

Definition 1.7 (Slater Constraint Qualification). There exists a strictly feasible point for (1.9), i.e., a point $z \in \mathbb{R}^n$ such that $h_j(z) = 0, j = 1, \dots, p$ and $g_i(z) < 0, i = 1, \dots, m$.

Under these conditions, we have strong duality.

Theorem 1.5 (Strong Duality). *Consider the optimization* (1.9) *and assume f, g_i are convex and h_j are affine. If Slater's constraint qualification holds, then the optimal value of the primal problem* (1.9) *is the same as the optimal value of the dual problem* (1.12).

1.3.3   KKT Optimality Conditions

Consider the nonlinear optimization problem (1.9). A pair of primal and dual variables $(x^\star, \lambda^\star, \mu^\star)$ is said to satisfy the Karush-Kuhn-Tucker (KKT) optimality conditions if

$$
\begin{aligned}
\text{primal feasibility}: \ & g_i(x^\star) \leq 0, \forall i = 1, \dots, m; h_j(x^\star) = 0, \forall j = 1, \dots, p \\
\text{dual feasibility}: \ & \lambda_i^\star \geq 0, \forall i = 1, \dots, m \\
\text{stationarity}: \ & \nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0 \\
\text{complementarity}: \ & \lambda_i^\star \cdot g_i(x^\star) = 0, \forall i = 1, \dots, m.
\end{aligned}
\tag{1.13}
$$

Under certain constraint qualifications, the KKT conditions are necessary for local optimality.

Theorem 1.6 (Necessary Optimality Conditions). *Assume any of the following constraint qualifications hold:*

- *The gradients of the constraints $\{\nabla g_i(x^\star)\}_{i=1}^m$, $\{\nabla h_j(x^\star)\}_{j=1}^p$ are linearly independent.*

- *Slater's constraint qualification (cf. Definition 1.7).*

- *All constraints $g_i(x)$ and $h_j(x)$ are affine functions.*

Then, at every local minimum $x^\star$ of (1.9), the KKT conditions (1.13) hold.

On the other hand, for convex optimization problems, the KKT conditions are sufficient for global optimality.

Theorem 1.7 (Sufficient Optimality Conditions). *Assume optimization (1.9) is convex, i.e., f, g_i are convex and h_j are affine. Every point $x^\star$ that satisfies the KKT conditions (1.13) is a global minimizer.*

1.4 Linear Optimization

1.4.1 Polyhedra

In $\mathbb{R}^n$, a **polyhedron** is a set defined by finitely many linear inequalities, i.e.,

$$P = \{x \in \mathbb{R}^n \mid Ax \geq b\}, \tag{1.14}$$

for some matrix $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. In (1.14), the inequality should be interpreted as $Ax - b \in \mathbb{R}_+^m$, i.e., every entry of Ax is no smaller than the corresponding entry of b.

The convex hull of finitely many points in $\mathbb{R}^n$ is called a **polytope**, where the convex hull of a set S is defined as

$$\mathrm{conv}(S) = \left\{ \sum_{i=1}^{k} \lambda_i u_i \mid k \in \mathbb{N}_+, \sum_{i=1}^{k} \lambda_i = 1, \lambda_i \geq 0, i = 1, \dots, k, u_i \in S, \forall i = 1, \dots, k \right\},$$
$$\tag{1.15}$$

i.e., all possible convex combinations of points in S. Clearly, a polytope is bounded.

The conic hull of finitely many points in $\mathbb{R}^n$ is called a **polyhedral cone**, where the conic hull of a set S is defined as

$$\mathrm{cone}(S) = \left\{ \sum_{i=1}^{k} \lambda_i u_i \mid k \in \mathbb{N}_+, \lambda_i \geq 0, i = 1, \dots, k, u_i \in S, \forall i = 1, \dots, k \right\}. \tag{1.16}$$

The only difference between (1.16) and (1.15) is the removal of $\sum_i \lambda_i = 1$. Clearly, the origin belongs to the conic hull of any nonempty set, and the conic hull of any nonempty set is unbounded.

The next theorem characterizes a polyhedron.

Theorem 1.8 (Polyhedron Decomposition)**.** *Every polyhedron P is finitely generated, i.e., it can be written as the Minkowski sum of a polytope and a polyhedral cone:*

$$P = \operatorname{conv}(u_1, \dots, u_r) + \operatorname{cone}(v_1, \dots, v_s),$$

where the Minkowski sum of two sets is defined as $X + Y := \{x + y \mid x \in X, y \in Y\}$.

Further, a bounded polyhedron is a polytope.

An extreme point of a polytope is called a **vertex**. A 1-dimensional face of a polytope is called an **edge**. A $d-1$-dimensional face of a d-dimensional polytope is called a **facet**.

1.4.2 Linear Program

We will now give a brief review of important results in linear programming (LP). The standard reference for linear programming is (Bertsimas and Tsitsiklis, 1997). In some sense, the theory of semidefinite programming (SDP) has been developed in order to generalize those of LP to the setup where the decision variable becomes a symmetric matrix and the inequality is interpreted as being positive semidefinite.

A standard form linear program (LP) reads

$$\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & \langle c, x \rangle \\
\text{s.t.} \quad & Ax = b \\
& x \geq 0
\end{aligned} \tag{1.17}$$

for given $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. Often the tuple (A, b, c) is called the *problem data* because the LP (1.17) is fully defined once the tuple is given (indeed many LP numerical solvers take the tuple (A, b, c) as input). Clearly, the feasible set of the LP (1.17) is a polyhedron. The LP (1.17) is often referred to as the **primal** LP. Associated with (1.17) is the following **dual** LP

$$\begin{aligned}
\max_{y \in \mathbb{R}^m} \quad & \langle b, y \rangle \\
\text{s.t.} \quad & c - A^\top y \geq 0
\end{aligned} \tag{1.18}$$

It is worth noting that the dimension of the dual variable y is exactly the number of constraints in the primal LP.

Lagrangian duality. Let us use the idea of Lagrangian duality introduced in Section 1.3.2 to verify that (1.18) is indeed the Lagrangian dual problem of (1.17). The Lagrangian associated with (1.17) is

$$\begin{aligned}
L(x, \lambda, \mu) &= \langle c, x \rangle + \langle \mu, Ax - b \rangle + \langle \lambda, -x \rangle, \quad \mu \in \mathbb{R}^m, \lambda \in \mathbb{R}^n_+ \\
&= \langle c + A^\top \mu - \lambda, x \rangle - \langle \mu, b \rangle, \quad \mu \in \mathbb{R}^m, \lambda \in \mathbb{R}^n_+.
\end{aligned} \tag{1.19}$$

The Lagrangian dual function is therefore

$$\phi(\lambda, \mu) = \min_x L(x, \lambda, \mu) = \begin{cases} -\langle \mu, b \rangle & \text{if } c + A^\top \mu - \lambda = 0 \\ -\infty & \text{Otherwise} \end{cases}, \mu \in \mathbb{R}^m, \lambda \in \mathbb{R}^n_+.$$

The Lagrangian dual problem seeks to maximize the dual function $\phi(\lambda, \mu)$, and hence it must set $c + A^\top \mu - \lambda = 0$ (otherwise it leads to $-\infty$). As a result, the dual problem is

$$\begin{aligned} \max_{\mu \in \mathbb{R}^m} \quad & \langle b, -\mu \rangle \\ \text{s.t.} \quad & c + A^\top \mu = \lambda \geq 0 \end{aligned} \tag{1.20}$$

With a change of variable $y := -\mu$, we observe that problem (1.20) is precisely problem (1.18).

Weak duality. For the pair of primal-dual LPs, it is easy to verify that, for any x that is feasible for the primal (1.17) and y that is feasible for the dual (1.18), we have

$$\langle c, x \rangle - \langle b, y \rangle = \langle c, x \rangle - \langle Ax, y \rangle = \langle c, x \rangle - \langle A^\top y, x \rangle = \langle c - A^\top y, x \rangle \geq 0. \tag{1.21}$$

Therefore, denoting $p^\star$ as the optimum of (1.17) and $d^\star$ as the optimum of (1.18), we have the weak duality

$$p^\star \geq d^\star.$$

Note that such weak duality can also be directly obtained since (1.20) is the Lagrangian dual of (1.17).

If $p^\star = d^\star$, then we say **strong duality** holds. The LP (1.17) is said to be **feasible** if its feasible set is nonempty. It is said to be **unbounded below** if there exists a sequence $\{u_i\}_{i=1}^\infty \subseteq \mathbb{R}^n_+$ such that $\langle c, u_i \rangle \to -\infty$ and $Au_i = b$. If the primal (1.17) is infeasible (resp. unbounded below), we set $p^\star = +\infty$ (resp. $p^\star = -\infty$). Similar characteristics are defined for the dual LP (1.18). In particular, if the dual (1.18) is unbounded, then we set $d^\star = +\infty$. If the dual is infeasible, then we set $d^\star = -\infty$.

Strong duality is well understood in linear programming.

Theorem 1.9 (LP Strong Duality)**.** *For the LP primal-dual pair* (1.17) *and* (1.18)*, we have*

- *If one of* (1.17) *and* (1.18) *is feasible, then* $p^\star = d^\star$ *(i.e., finite, $+\infty$, or $-\infty$).*

- *If one of* $p^\star$ *or* $d^\star$ *is finite, then* $p^\star = d^\star$ *is finite, and both* (1.17) *and* (1.18) *achieve the same optimal value (i.e., they botb have optimizers).*

- *A primal feasible point $x^\star$ of (1.17) is a minimizer if and only if there exists a dual feasible point $y^\star$ such that $\langle c, x^\star \rangle = \langle b, y^\star \rangle$.*

For example, consider the following primal-dual LP pair

$$
\begin{cases}
\min_{x \in \mathbb{R}^3_+} & x_1 + x_2 + 2x_3 \\
\text{s.t.} & \begin{bmatrix} -1 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\end{cases}
,
\begin{cases}
\max_{y \in \mathbb{R}^2} & y_2 \\
\text{s.t.} & \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \geq 0
\end{cases}
.
$$

$$(1.22)$$

$x^\star = [1/2, 1/2, 0]^\top$ is feasible for the primal and attains $p^\star = 1$. $y^\star = [0, 1]^\top$ is feasible for the dual and attains $d^\star = 1$. Therefore, both $x^\star$ and $y^\star$ are optimizers for the primal and dual, respectively.

Complementary slackness. Strong duality, when combined with (1.21), implies that

$$x_i^\star (c - A^\top y^\star)_i = 0, \forall i = 1, \dots, n,$$

where $(\cdot)_i$ denotes the i-th entry of a vector. This is known as complementary slackness, which states that whenever a primal optimal solution has a nonzero entry, the corresponding dual inequality must be tight.

An important property of LP is that if the primal problem is feasible and bounded below, then it must have an optimizer that is a **basic feasible point**, i.e., a feasible point has at most m nonzero entries. The simplex method (Bertsimas and Tsitsiklis, 1997) for solving LPs searches for optimizers among the basic feasible points.

We also introduce how to detect infeasibility and unboundedness of LPs.

Theorem 1.10 (LP Infeasibility and Unboundedness)**.** *Infeasibility and Unboundedness of LP can be certified by existence of an improving/decreasing ray for the primal and dual:*

- *When the primal (1.17) is feasible, it is unbounded below if and only if it has a decreasing ray, i.e., there exists $u \in \mathbb{R}^n$ such that*

$$Au = 0, \quad u \geq 0, \quad \langle c, u \rangle < 0.$$

- *When the dual (1.18) is feasible, it is unbounded above if and only if it has an improving ray, i.e., there exists $u \in \mathbb{R}^m$ such that*

$$A^\top u \leq 0, \quad \langle b, u \rangle > 0.$$

- *The primal problem (1.17) is infeasible if and only if the dual problem (1.18) has an improving ray, i.e., there exists $u \in \mathbb{R}^m$ such that*

$$A^\top u \leq 0, \quad \langle b, u \rangle > 0.$$

- *The dual problem* (1.18) *is infeasible if and only if the primal problem* (1.17) *has a decreasing ray, i.e., there exists $u \in \mathbb{R}^n$ such that*

$$Au = 0, \quad u \geq 0, \quad \langle c, u \rangle < 0.$$

It is important to note that both the primal and dual can be infeasible, as in the following example.

$$\begin{cases} \min_{x \in \mathbb{R}^2_+} & -x_1 - x_2 \\ \text{s.t.} & \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} x = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \end{cases}, \begin{cases} \max_{y \in \mathbb{R}^2} & 2y_1 + 3y_2 \\ \text{s.t.} & \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} y \geq 0 \end{cases}. \tag{1.23}$$

1.4.3 Farkas Lemma

A foundational result in linear programming is the Farkas Lemma.

Theorem 1.11 (Farkas Lemma). *For a given $A \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}^n$, if $\langle c, x \rangle \geq 0$ for all x satisfying $Ax \geq 0$, then there exists $\lambda \in \mathbb{R}^m$ such that*

$$c = A^\top \lambda, \quad \lambda \geq 0.$$

As a simple example, take $A = I_n$ as the identity matrix, then Farkas Lemma says if $\langle c, x \rangle \geq 0$ for all $x \geq 0$, then c must be that $c \geq 0$ – this is exactly the fact that the nonnegative orthant $\mathbb{R}^n_+$ is self-dual.

In general, the Farkas Lemma states if the linear function $\langle c, x \rangle$ is nonnegative on the space $\{Ax \geq 0\}$, then there exists $\lambda \in \mathbb{R}^m_+$ such that

$$\langle c, x \rangle = \langle \lambda, Ax \rangle = \sum_{i=1}^m \lambda_i(a_i^\top x), \tag{1.24}$$

where $a_i^\top$ is the i-th row of A. Note that (1.24) is a polynomial identity. As we will see later in the course, the idea of sums of squares (SOS), to some extent, is to generalize Farkas Lemma to the case where the function is a polynomial and the set is a basic semialgebraic set (i.e., defined by polynomial equalities and inequalities).

A generalization of Farkas Lemma to inhomogeneous affine functions is stated below.

Theorem 1.12 (Inhomogeneous Farkas Lemma). *Suppose the set $P = \{x \in \mathbb{R}^n \mid Ax \geq b\}$ with $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ is nonempty. If a linear function $\langle c, x \rangle - d$ is nonnegative on P, then there exists $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}$ such that*

$$\langle c, x \rangle - d = \nu + \langle \lambda, Ax - b \rangle, \quad \lambda \geq 0, \nu \geq 0.$$

A more general result is called the Theorem of Alternatives, which states that
a polyhedral set is empty if and only if another polyhedral set is nonempty.

Theorem 1.13 (Theorem of Alternatives). *Given $A_1 \in \mathbb{R}^{m_1 \times n}, A_2 \in \mathbb{R}^{m_2 \times n}$,
$b_1 \in \mathbb{R}^{m_1}$, and $b_2 \in \mathbb{R}^{m_2}$, the set*

$$\{x \in \mathbb{R}^n \mid A_1 x > b_1, A_2 x \geq b_2\}$$

is empty if and only if the following set

$$\left\{ (\lambda_1, \lambda_2) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \;\middle|\; \begin{array}{r} \lambda_1 \geq 0, \lambda_2 \geq 0, \\ b_1^\top \lambda_1 + b_2^\top \lambda_2 \geq 0, \\ A_1^\top \lambda_1 + A_2^\top \lambda_2 = 0, \\ (e + b_1)^\top \lambda_1 + b_2^\top \lambda_2 = 1 \end{array} \right\}$$

is nonempty, with e being the vector of all ones.

Chapter 2

Semidefinite Optimization

2.1 Positive Semidefinite Matrices

A real matrix $A = (A_{ij}) \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^\top$, i.e., $A_{ij} = A_{ji}$ for all i, j. Let $\mathbb{S}^n$ be the space of all real symmetric matrices.

Any symmetric matrix A defines a **quadratic form** $x^\top A x$. A matrix A is said to be **positive semidefinite** (PSD) if and only if its associated quadratic form is nonnegative, i.e.,

$$x^\top A x \geq 0, \quad \forall x \in \mathbb{R}^n.$$

We use $\mathbb{S}^n_+$ to denote the set of $n \times n$ PSD matrices. We also write $A \succeq 0$ to denote positive semidefiniteness when the dimension is clear.

There are several equivalent characterizations of positive semidefiniteness.

Lemma 2.1 (Positive Semidefinite Matrices). *Let $A \in \mathbb{S}^n$ be a symmetric matrix, the following statements are equivalent:*

1. *A is positive semidefinite.*

2. *$x^\top A x \geq 0, \forall x \in \mathbb{R}^n$.*

3. *All eigenvalues of A are nonnegative.*

4. *All $2^n - 1$ principal minors of A are nonnegative.*

5. *The coefficients of $p_A(\lambda)$ weakly alternate in sign, i.e., $(-1)^{n-k} p_k \geq 0$ for $k = 0, \ldots, n-1$, where $p_A(\lambda) = \det(A - \lambda \mathrm{I}_n)$ is the characteristics polynomial of A.*

6. *There exists a factorization $A = BB^\top$, where $B \in \mathbb{R}^{n \times r}$ with r the rank of A.*

Similarly, we say a matrix $A \in \mathbb{S}^n$ is **positive definite** (PD) is its associated quadratic form is always positive, i.e.,

$$x^\top A x > 0, \quad \forall x \in \mathbb{R}^n.$$

We use $\mathbb{S}^n_{++}$ to denote the set of $n \times n$ PD matrices, and also write $A \succ 0$ when the dimension is clear.

Below is set of equivalent characterizations of positive definite matrices.

Lemma 2.2 (Positive Definite Matrices). *Let $A \in \mathbb{S}^n$ be a symmetric matrix, the following statements are equivalent:*

1. *A is positive definite.*

2. *$x^\top A x > 0, \forall x \in \mathbb{R}^n$.*

3. *All eigenvalues of A are strictly positive.*

4. *All n leading principal minors of A are strictly positive.*

5. *The coefficients of $p_A(\lambda)$ strictly alternate in sign, i.e., $(-1)^{n-k} p_k > 0$ for $k = 0, \ldots, n-1$, where $p_A(\lambda) = \det(A - \lambda I_n)$ is the characteristics polynomial of A.*

6. *There exists a factorization $A = BB^\top$ with B square and nonsingular (full-rank).*

Schur Complements. A useful technique to check whether a matrix is positive (semi-)definite is to use the Schur Complements. Consider a block-partitioned matrix

$$M = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}, \tag{2.1}$$

where A and C are symmetric matrices. If A is invertible, then the Schur complement of A is

$$M/A = C - B^\top A^{-1} B.$$

Similarly, if C is invertible, then the Schur complement of C is

$$M/C = A - BC^{-1}B^\top.$$

We have the following result relating the Schur Complements to positive (semi-)definiteness.

Proposition 2.1 (Schur Complements and PSD). *Consider the block-partitioned matrix M in* (2.1)*,*

- *M is positive definite if and only if both A and M/A are positive definite:*

$$M \succ 0 \Leftrightarrow A \succ 0, M/A = C - B^\top A^{-1} B \succ 0.$$

- *M is positive definite if and only if both C and M/C are positive definite:*

$$M \succ 0 \Leftrightarrow C \succ 0, M/C = A - BC^{-1}B^\top \succ 0.$$

- *If A is positive definite, then M is positive semidefinite if and only if M/A is positive semidefinite:*

$$\text{If } A \succ 0, \text{ then } M \succeq 0 \Leftrightarrow M/A \succeq 0.$$

- *If C is positive definite, then M is positive semidefinite if and only if M/C is positive semidefinite:*

$$\text{If } C \succ 0, \text{ then } M \succeq 0 \Leftrightarrow M/C \succeq 0.$$

2.1.1 Geometric Properties

The set $\mathbb{S}_+^n$ is a proper cone (cf. Definition 1.3). Its interior is $\mathbb{S}_{++}^n$. Under the inner product

$$\langle A, B \rangle = \text{tr}(AB^\top), \quad A, B \in \mathbb{R}^{n \times n},$$

the PSD cone $\mathbb{S}_+^n$ is self-dual.

Next we want to characterize the face of the PSD cone. We first present the following lemma which will turn out to be useful afterwards.

Lemma 2.3 (Range of PSD Matrices). *Let $A, B \in \mathbb{S}_+^n$, then we have*

$$\text{Range}(A) \subseteq \text{Range}(A + B), \tag{2.2}$$

where $\text{Range}(A)$ *denotes the span of the column vectors of A.*

Proof. For any symmetric matrix S, we know

$$\text{Range}(S) = \ker(S)^\perp.$$

Therefore, to prove (2.2), it is equivalent to prove

$$\ker(A) \supseteq \ker(A + B).$$

Pick any $u \in \ker(A + B)$, we have

$$(A + B)u = 0 \Rightarrow u^\top(A + B)u = 0 \Rightarrow u^\top Au + u^\top Bu = 0 \Rightarrow u^\top Au = u^\top Bu = 0,$$

where the last derivation is due to $A, B \succeq 0$. Now that we have $u^\top Au = 0$, we claim that $Au = 0$ must hold, i.e., $u \in \ker(A)$. To see this, write

$$u = \sum_{i=1}^{n} a_i v_i,$$

where $a_i = \langle u, v_i \rangle$ and $v_i, i = 1, \dots, n$ are the eigenvectors of A corresponding to eigenvalues $\lambda_i, i = 1, \dots, n$. Then we have

$$Au = \sum_{i=1}^{n} a_i A v_i = \sum_{i=1}^{n} a_i \lambda_i v_i,$$

and

$$u^\top A u = \sum_{i=1}^{n} \lambda_i a_i^2 = 0.$$

Since $\lambda_i \geq 0, a_i^2 \geq 0$, we have

$$\lambda_i a_i^2 = 0, \forall i = 1, \dots, n.$$

This indicates that if $\lambda_i > 0$, then $a_i = 0$. Therefore, a_i can only be nonzero for $\lambda_i = 0$, which leads to

$$Au = \sum_{i=1}^{n} a_i \lambda_i v_i = 0.$$

Therefore, $u \in \ker(A)$, proving the result. $\qquad\qquad\qquad\qquad\qquad\square$

Lemma 2.3 indicates that if $A \succeq B$, then $\mathrm{Range}(B) \subseteq \mathrm{Range}(A)$.

For a subset $T \subseteq \mathbb{S}_+^n$, we use $\mathrm{face}(T, \mathbb{S}_+^n)$ to denote the smallest face of $\mathbb{S}_+^n$ that contains T. We first characterize the smallest face that contains a given PSD matrix, i.e., $\mathrm{face}(A, \mathbb{S}_+^n)$ for $A \succeq 0$. Clearly, if A is PD, then $\mathrm{face}(A, \mathbb{S}^n) = \mathbb{S}_+^n$ is the entire cone.

2.2   Semidefinite Programming

2.2.1   Geometric Properties

2.3   Interior Point Algorithm

Chapter 3

Problem Sets

Consider the SDP

$$\min_{X \in \mathbb{S}^n} \{ \langle C, X \rangle \mid \mathcal{A}(X) = b, X \succeq 0 \}$$

The ALM of the SDP reads

$$L_\sigma(X, \mu) = \langle C, X \rangle - \langle \mu, \mathcal{A}(X) - b \rangle + \frac{\sigma}{2} \| \mathcal{A}(X) - b \|^2.$$

Let $x := \text{svec}(X)$, $c := \text{svec}(C)$, and

$$A = \begin{bmatrix} \text{svec}(A_1) & \dots & \text{svec}(A_m) \end{bmatrix},$$

The ALM is equivalent to

$$L_\sigma(x, \mu) = c^\top x - \mu^\top (A^\top x - b) + \frac{\sigma}{2} \| A^\top x - b \|^2.$$

If instead, we use B-M factorization, then we have $X = YY^\top$ with $Y \in \mathbb{R}^{n \times r}$. Let $y = \text{vec}(Y)$, then it is clear we have

$$x = \text{svec}(X) = \text{svec}(YY^\top) = \text{svec}(\text{mat}(y)\text{mat}(y)^\top) := g_r(y),$$

with g a quadratic function depending on r. Let

$$J(y) = \frac{\partial x}{\partial y} = \frac{\partial g_r(y)}{\partial y}.$$

To get the gradient of $L_\sigma(x, \mu)$ w.r.t. y, we have

$$\nabla_y L_\sigma = \nabla_x L_\sigma \frac{\partial x}{\partial y} = J(y)^\top (c - A\mu + \sigma A(A^\top x - b)).$$

Exercise 3.1 (Test). Test

Bibliography

Bertsekas, D., Nedic, A., and Ozdaglar, A. (2003). *Convex analysis and optimization*, volume 1. Athena Scientific. 11

Bertsimas, D. and Tsitsiklis, J. N. (1997). *Introduction to linear optimization*, volume 6. Athena scientific Belmont, MA. 20, 22

Boyd, S. P. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press. 11

Rockafellar, R. T. (1970). Convex analysis. 11

Tang, Y., Lasserre, J.-B., and Yang, H. (2023). Uncertainty quantification of set-membership estimation in control and perception: Revisiting the minimum enclosing ellipsoid. *arXiv preprint arXiv:2311.15962*. 17