Semidefinite Optimization and Relaxation

Heng Yang

2024-01-22

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

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 generation 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 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

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

where $B(u, \epsilon)$ denotes a ball centered at u with radius ϵ (using the usual 2-norm). Each point in $\operatorname{int}(C)$ is called an *interior point* of C. If $\operatorname{int}(C) = C$, 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 not in the interior of C and we denote it as ∂C.

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

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

For a nonempty convex set, the relative interior always exists. If $\operatorname{ri}(C) = C$, 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 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

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

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 nonnegative scalings. 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

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

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.

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

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

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

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

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

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

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

This is easy to verify because

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

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

Conversely, if (1.7) holds, then take

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

we have

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

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

Not very 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 convex a minimax problem into a single-level minimization problem.

1.3.2 Lagrangian Duality

Consider a nonlinear optimization problem

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

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

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

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 $h_j(z) = 0, j = 1, \ldots, p$ and $g_i(z) < 0, i = 1, \ldots, m$.

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

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

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

Chapter 2

Semidefinite Optimization

- Positive Semidefinite Matrices
- Spectrahedra

Chapter 3

Problem Sets

Exercise 3.1 (Test)**.** Test

Exercise 3.2 (Test)**.** Test

Bibliography

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

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