MGMT 690—Convex Optimization

Alex L. Wang

Last updated February 28, 2024

Contents

1

Linear algebra review

1.1 Euclidean space

Definition 1. A *Euclidean space*[1] is a set of elements V called *vectors* or *points* endowed with

1. addition: for any $u, v \in V$, $u + v \in V$
2. real scalar multiplication: for any $u \in V$ and $\alpha \in \mathbb{R}$, $\alpha u \in V$
3. a finite basis: there exists finitely many $u_1, \ldots, u_k$ so that for any $v \in V$, we can express $v = \sum_{i=1}^{k} \alpha_i u_i$ for a unique choice of $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$
4. an inner product: there exists a nonnegative symmetric bilinear function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfying $\langle v, v \rangle = 0$ if and only if $v = 0$. □

[1] Also known as a finite-dimensional real inner product space

Example 1.

- $\mathbb{R}^n$ with the standard inner product

$$\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i$$

  is a Euclidean space.

- $\mathbb{R}^{n \times m}$ with the trace inner product

$$\langle X, Y \rangle := \operatorname{tr}(X^\intercal Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} X_{i,j} Y_{i,j}$$

  is a Euclidean space.

- Let $Q \in \mathbb{S}^n_{++}$ be a positive definite matrix. Then, $\mathbb{R}^n$ with the Q-weighted inner product

$$\langle x, y \rangle := x^\intercal Q y$$

  is a Euclidean space.

□

Definition 2. A *norm* on a Euclidean space V is a function $\|\cdot\| : V \to \mathbb{R}$ so that

- Positivity: $\|v\| \geq 0$ for all $v \in V$ and $\|v = 0\|$ if and only if $v = 0$
- Homogeneity: $\|\lambda v\| = |\lambda| \, \|v\|$ for all $\lambda \in \mathbb{R}$ and $v \in V$
- Triangle inequality: $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in V$ □

Example 2.

- In any Euclidean space V, the *induced norm*

$$\|v\| := \sqrt{\langle v, v \rangle}$$

  is a norm.[2]

- Let $p \in [1, \infty)$, the ℓ_p norm[3] on $\mathbb{R}^n$ is defined as

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

  The ℓ_∞ norm is defined as $\lim_{p \to \infty} \|x\|_p$. It is equivalently,

$$\|x\|_\infty = \max_i |x_i| \, .$$

  The ℓ_2 norm is equal to the norm induced by the standard inner product.

- Let $p \in [1, \infty]$. The Schatten-p norm is a norm defined on $\mathbb{R}^{n \times m}$. Given $X \in \mathbb{R}^{n \times m}$, let

$$\mathrm{svals}(X) := (\sigma_1, \ldots, \sigma_{\min(n,m)})$$

  denote the list of singular values of X. The Schatten-p norm is

$$\|X\|_{\text{Sch-}p} := \|\mathrm{svals}(X)\|_p \, .$$

  For example, the Schatten-1 norm is the sum of the singular values and the Schatten-∞ norm is the maximum singular value.

  The Schatten-2 norm is also known as the Frobenius norm, the Schatten-1 norm is also known as the trace-class norm or the nuclear norm, and the Schatten-∞ norm is also known as the operator norm. □

[2] **Exercise:** Verify that this is indeed a norm.
[3] **Exercise:** Verify that the same construction is *not* a norm for $p \in (0, 1)$

1.2 PSD matrices and the Singular Value Decomposition

Definition 3. A matrix $X \in \mathbb{R}^{n \times n}$ is *orthogonal* if

$$X^\intercal X = I.$$

The set of orthogonal matrices is denoted $O(n)$. □

Theorem 1 (Spectral theorem for symmetric matrices)**.** *Given $A \in \mathbb{S}^n$, there exists a $U \in O(n)$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ so that*

$$A = U \operatorname{Diag}(\lambda_1, \ldots, \lambda_n) U^\intercal.$$

The values of $\lambda_1, \ldots, \lambda_n$ are unique up to reordering and are referred to as the eigenvalues of A. The matrix U is referred to as an eigenbasis of A; it is not unique in general. The columns of U are eigenvectors of A.

Definition 4. A matrix $A \in \mathbb{S}^n$ is positive semidefinite, denoted $A \in \mathbb{S}^n_+$, if any of the equivalent definitions hold:

- There exists a spectral decomposition of A with $\lambda_1, \ldots, \lambda_n \geq 0$
- $x^\intercal A x \geq 0$ for all $x \in \mathbb{R}^n$

A matrix $A \in \mathbb{S}^n$ is positive definite, denoted $A \in \mathbb{S}^n_{++}$, if any of the equivalent definitions hold:

- There exists a spectral decomposition of A with $\lambda_1, \ldots, \lambda_n > 0$
- $x^\intercal A x > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$ □

The definitions above are equivalent by the spectral theorem: Write $A = U \operatorname{Diag}(\lambda_1, \ldots, \lambda_n) U^\intercal$. The set of values of $x^\intercal A x$ as x range over $\mathbb{R}^n$ is equal to the set of values of $y^\intercal \operatorname{Diag}(\lambda_1, \ldots, \lambda_n) y$ as $y = (U^\intercal x)$ ranges over $\mathbb{R}^n$. The latter expression is

$$y^\intercal \operatorname{Diag}(\lambda_1, \ldots, \lambda_n) y = \sum_{i=1}^n \lambda_i y_i^2.$$

This is nonnegative for all choices of y if and only if $\lambda_i \geq 0$ for all i.

This calculation also shows that the following variational characterization of the minimum eigenvalue holds:

Theorem 2 (Courant-Fischer Theorem)**.** *Let $A \in \mathbb{S}^n$ and let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ denote the eigenvalues of A in nondecreasing order. Then,*

$$\lambda_1 = \min_{x \in \mathbb{R}^n \setminus \{0\}} \frac{x^\intercal A x}{x^\intercal x}.$$

More generally, the kth smallest eigenvalue λ_k is given by

$$\lambda_k = \min_{W \text{ a subspace of dimension } k} \max_{x \in W \setminus \{0\}} \frac{x^\intercal A x}{x^\intercal x}$$

Lemma 1. *Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of A. It holds that* $\operatorname{tr}(A) = \sum_{i=1}^n \lambda_i$ *and* $\det(A) = \prod_{i=1}^n \lambda_i$.

Proof. Let $A = UDU^\intercal$ denote an eigendecomposition of A. Then, the cyclic property of the trace proves that

$$\operatorname{tr}(A) = \operatorname{tr}(UDU^\intercal) = \operatorname{tr}(DUU^\intercal) = \operatorname{tr}(D).$$

The commutative property of the determinant gives

$$\det(A) = \det(UDU^\intercal) = \det(DUU^\intercal) = \det(D).$$ ■

Problems

1. Given $A \in \mathbb{S}^n$ and $B \in \mathbb{S}^m$, the Kronecker product $A \otimes B$ is the $\mathbb{S}^{mn}$ matrix given in block form as

$$A \otimes B = \begin{pmatrix} A_{1,1}B & \dots & A_{1,n}B \\ \vdots & \ddots & \vdots \\ A_{n,1}B & \dots & A_{n,n}B \end{pmatrix}$$

   Suppose $A \in \mathbb{S}^n_+$ and $B \in \mathbb{S}^m_+$. Show that $A \otimes B \succeq 0$.

2. Given $A \in \mathbb{S}^n$ and $B \in \mathbb{S}^n$, the Schur product is the S^n matrix given by

$$(A \odot B)_{i,j} = A_{i,j}B_{i,j}.$$

   Suppose $A \in \mathbb{S}^n_+$ and $B \in \mathbb{S}^n_+$. Show that $A \odot B \succeq 0$.

3. Given a symmetric matrix $A \in \mathbb{S}^n$, let $\text{Inertia}(A) := (n_-, n_0, n_+)$ denote the number of negative eigenvalues, number of zero eigenvalues, and number of positive eigenvalues of A. Prove that for any invertible $P \in \mathbb{R}^{n \times n}$, that

$$\text{Inertia}(A) = \text{Inertia}(P^\mathsf{T} A P).$$

4. Let $A \in \mathbb{S}^n_{++}$, $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{S}^m$. Prove that

$$\begin{pmatrix} A & B \\ B^\mathsf{T} & C \end{pmatrix} \succeq 0 \qquad \iff \qquad C - B^\mathsf{T} A^{-1} B \succeq 0.$$

2

Elementary convex analysis I

2.1 Convex sets

Definition 5. A set $S \subseteq \mathbb{R}^n$ is

- *affine* if for all $x, y \in S$ and $\theta \in \mathbb{R}$, we have $\theta x + (1-\theta)y \in S$
- *convex conic* if for all $x, y \in S$ and $\lambda, \mu \geq 0$, we have $\lambda x + \mu y \in S$.
- *convex* if for all $x, y \in S$ and $\theta \in [0, 1]$, we have $\theta x + (1-\theta)y \in S$. $\square$

Definition 6. Fix $x_1, \ldots, x_k \in \mathbb{R}^n$. Let $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$. We say that $\sum_{i=1}^k \alpha_i x_i$

- is an *affine* combination of $x_1, \ldots, x_k$ if $\sum_{i=1}^k \alpha_i = 1$
- is a *conic* combination of $x_1, \ldots, x_k$ if $\alpha_i \geq 0$
- is a *convex* combination of $x_1, \ldots, x_k$ if $\sum_{i=1}^k \alpha_i = 1$ and $\alpha_i \geq 0$ $\square$

Example 3.

- $\mathbb{R}^n$ and $\{0\}$ are affine sets
- An affine hyperplane $\{x \in \mathbb{R}^n : \langle a, x \rangle = b\}$ is an affine set
- A closed halfspace $\{x \in \mathbb{R}^n : \langle a, x \rangle \leq b\}$ is a convex set
- $\{x \in \mathbb{R}^n : \langle a, x \rangle \leq 0\}$ is a convex cone $\square$

Lemma 2. *Affine $\implies$ convex. Similarly, convex conic $\implies$ convex.*

Lemma 3. *An arbitrary intersection of affine sets is an affine set. A finite product of affine sets is an affine set.*

Both statements also hold if we replace "affine set" throughout with "convex cone" or "convex set."

Proof. We prove the affine set statements. The other claims are similar.

Suppose $S_\alpha \subseteq \mathbb{R}^d$ is an affine set for every $\alpha \in A$. Let $x, y \in \bigcap_{\alpha \in A} S_\alpha$. Let $\theta \in \mathbb{R}$ and $\alpha \in A$. As S_α is an affine set, we have that $\theta x + (1-\theta)y \in S_\alpha$. Thus, $\theta x + (1-\theta)y \in \bigcap_{\alpha \in A} S_\alpha$.

Suppose $S_i \subseteq \mathbb{R}^{d_i}$ is an affine set for every $i \in [k]$. Let

$$\prod_{i=1}^k S_i \coloneqq S_1 \times \cdots \times S_k \coloneqq \left\{(x_1, \ldots, x_k) \in \prod_{i=1}^k \mathbb{R}^{d_i} : x_i \in S_i\right\}$$

denote the product of $S_1, \ldots, S_k$. Suppose $(x_1, \ldots, x_k)$ and $(y_1, \ldots, y_k) \in \prod_{i=1}^k S_i$ and let $\theta \in \mathbb{R}$. By definition, we have that $x_i, y_i \in S_i$. As $\theta \in \mathbb{R}$ and S_i is affine, we have $\theta x_i + (1-\theta)y_i \in S_i$. Thus,

$$\theta(x_1, \ldots, x_k) + (1-\theta)(y_1, \ldots, y_k) \in \prod_{i=1}^k S_i. \qquad \blacksquare$$

Example 4.

- Any affine subspace $\left\{x \in \mathbb{R}^d : Ax = b\right\}$ is an affine set
- Any polyhedral set $\left\{x \in \mathbb{R}^d : Ax \leq b\right\}$ is a convex set
- Any polyhedral cone $\left\{x \in \mathbb{R}^d : Ax \leq 0\right\}$ is a convex cone $\square$

Example 5. Let $\mathbb{R}[x]_{\leq d}$ denote the polynomials in x with degree at most d. We can identify $\mathbb{R}[x]_{\leq d}$ with $\mathbb{R}^{d+1}$ as

$$\sum_{i=0}^d c_i x^i \equiv (c_0, c_1, \ldots, c_d).$$

- The set of nonnegative polynomials,
$$\left\{p \in \mathbb{R}[x]_{\leq d} : p(x) \geq 0, \forall x \in \mathbb{R}\right\},$$
is a convex cone.
- The set of polynomials with some prespecified evaluations $\{(x_i, \alpha_i)\}_{i=1}^k$,
$$\left\{p \in \mathbb{R}[x]_{\leq d} : p(x_i) = \alpha_i, \forall i \in [k]\right\},$$
is an affine space. $\square$

Example 6. Some important cones

- The nonnegative orthant
$$\mathbb{R}^n_+ \coloneqq \{x \in \mathbb{R}^n : x \geq 0\}$$
- The second order cone
$$\mathcal{L}^{1+n} \coloneqq \left\{\begin{pmatrix} t \\ x \end{pmatrix} \in \mathbb{R}^{1+n} : \|x\|_2 \leq t\right\}$$

- The semidefinite cone

$$\mathbb{S}^n_+ \coloneqq \left\{X \in \mathbb{S}^n_+ : X \succeq 0\right\}. \qquad \square$$

Lemma 4. *The affine image of a convex set is convex.*

This proof is left as an **Exercise**.

2.2 The convex hull

Definition 7. Let $S \subseteq \mathbb{R}^n$. The convex hull of S is the smallest convex set containing S and is well-defined by Lemma 3. $\square$

Theorem 3. *Let $S \subseteq \mathbb{R}^n$ and let C denote the set of convex combinations of points in S:*

$$C \coloneqq \bigcup_{k=1}^{\infty} \left\{ \sum_{i=1}^{k} \lambda_i s_i : \begin{array}{l} \lambda_1 + \cdots + \lambda_k = 1 \\ \lambda_i \geq 0, \forall i \\ s_i \in S, \forall i \end{array} \right\}.$$

Then, $C = \operatorname{conv}(S)$.

Proof. We begin by showing that C is convex. Supppose $x, y \in C$ and $\theta \in [0,1]$. As $x \in C$, we can write $x = \sum_{i=1}^k \lambda_i x^i$ where $\lambda_1, \dots, \lambda_k$ is a set of convex combination weights and $x^i \in S$. Similarly, we can write $y = \sum_{i=1}^m \mu_i y^i$ where $\mu_1, \dots, \mu_m$ is a set of convex combination weights and $y^i \in S$. Then,

$$\theta x + (1-\theta) y = \sum_{i=1}^{k} (\theta\lambda_i) x^i + \sum_{i=1}^{m} ((1-\theta)\mu_i) y^i \in C.$$

We deduce that $\operatorname{conv}(S) \subseteq C$.

The direction $C \subseteq \operatorname{conv}(S)$ is direct. ∎

Theorem 4 (Carathéodory's theorem)**.** *Let $S \subseteq \mathbb{R}^n$. For any $x \in \operatorname{conv}(S)$, there exists $\lambda_1, \dots, \lambda_{n+1}$ and $s_1, \dots, s_{n+1} \in S$ so that*

$$x = \sum_{i=1}^{n+1} \lambda_i s_i.$$

Proof. By the inner representation of the convex hull, there exists some $k \geq 1$ and $\lambda_1, \dots, \lambda_k$ and $s_1, \dots, s_k \in S$ so that

$$x = \sum_{i=1}^{k} \lambda_i s_i.$$

If $k \leq n+1$ then we are done. Otherwise, $k \geq n+2$. Consider the set of vectors $\{x_i - x_k\}_{i=1}^{k-1}$. As this set contains $k-1 > n$ elements, it is linearly dependent and there exists nonzero $\theta_1, \dots, \theta_{k-1}$ so that

$$\sum_{i=1}^{k-1} \theta_i (x_i - x_k) = 0.$$

Now consider the modified convex combination weights:

$$\lambda_i = \lambda_i + \delta\theta_i,\ \forall i \in [k-1]$$
$$\lambda_k = \lambda_k - \delta\sum_{i=1}^{k-1}\theta_i.$$

This is a valid set of convex combination weights as long as all multipliers are nonnegative. Take δ either large enough or small enough to zero out at least one of these weights while maintaining that all weights are nonnegative. Repeat until $k \leq n+1$. ■

2.3 Sets related to a convex set

Definition 8. Let $S \subseteq \mathbb{R}^n$. The affine hull of S, denoted $\text{aff}(S)$ is the smallest affine set containing S. The conic hull of S, denoted $\text{cone}(S)$ is the smallest cone containing S.

These sets are well-defined by Lemma 3. □

Let $\mathbb{B}(x,\epsilon) := \{y \in \mathbb{R}^n : \|x-y\| \leq \epsilon\}$.

Definition 9. Let $C \subseteq \mathbb{R}^n$.

- The *interior* of C is the set

$$\text{int}(C) := \{x \in C : \exists\epsilon > 0,\ \mathbb{B}(x,\epsilon) \subseteq C\}.$$

- The *boundary* of C is the set $\text{bd}(C) := \text{cl}(C) \setminus \text{int}(C)$.
- The *relative interior* of C is the set

$$\text{rint}(C) := \{x \in C : \exists\epsilon > 0,\ \mathbb{B}(x,\epsilon) \cap \text{aff}(C) \subseteq C\}.$$

- The *relative boundary* of C is the set $\text{rbd}(C) := \text{cl}(C) \setminus \text{rint}(C)$.
- The *recessive cone* of C is the set

$$\text{rec}(C) := \{x \in \mathbb{R}^n : \forall y \in C, \forall t \geq 0,\ y + tx \in C\}.$$ □

Lemma 5. *Suppose $C \subseteq \mathbb{R}^n$ is a convex set. Then, $\text{int}(C)$ and $\text{rint}(C)$ are convex sets and $\text{rec}(C)$ is a cone.*

Lemma 6. *Suppose $C \subseteq \mathbb{R}^n$ is a convex set, $x \in \text{rint}(C)$ and $y \in \text{cl}(C)$. Then for all $\theta \in [0,1)$, $(1-\theta)x + \theta y \in \text{rint}(C)$.*

This proof is left as an **Exercise**.

Corollary 1. *Let $C \subseteq \mathbb{R}^n$ be a convex set. Then,*

- $\text{rint}(C)$ *is dense in* $\text{cl}(C)$*, i.e., for any $c \in \text{cl}(C)$, there exists a sequence $c_i \in \text{rint}(C)$ so that $c_i \to c$.*
- $\text{rint}(C) = \text{rint}(\text{cl}(C))$.

3

Elementary convex analysis II

3.1 Convex functions

Definition 10. A function $f : \mathbb{R}^n \to \mathbb{R}$ is

- *affine* if $f(\theta x + (1 - \theta y)) = \theta f(x) + (1 - \theta) f(y)$ for all $x, y \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$. Equivalently, $f(x)$ is affine if it can be written as $f(x) = \langle a, x \rangle + b$.
- *convex* if $f(\theta x + (1 - \theta y)) \leq \theta f(x) + (1 - \theta) f(y)$ for all $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$. □

We can generalize this definition to a convex function over a convex set $\Omega \subseteq \mathbb{R}^n$ by restricting $x, y \in \Omega$ in the definition above.

Example 7.

- Any norm is a convex function.
- Suppose $Q \succeq 0$. Then $x \mapsto x^\intercal Q x$ is convex
- Let $\mathcal{A} : \mathbb{R}^m \to \mathbb{S}^n$ be a linear map. Then, $y \mapsto \lambda_{\max}(\mathcal{A}(y))$ is a convex function. □

Lemma 7. *Suppose $f, g : \mathbb{R}^n \to \mathbb{R}$ are convex functions and $\alpha \geq 0$*

- *αf is convex*
- *$f + g$ is convex*
- *$\max \{f(x), g(x)\}$ is convex*
- *$y \mapsto f(Ay + b)$ is convex if $A \in \mathbb{R}^{n \times m}$ and $b \in \mathbb{R}^n$*

If $f(x, y)$ is jointly convex in (x, y), then $x \mapsto \inf_y f(x, y)$ is convex.

Lemma 8. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex and $\alpha \in \mathbb{R}$. Then the α sublevel set of f,*

$$\{x \in \mathbb{R}^n : f(x) \leq \alpha\},$$

is a convex set.

Example 8.

- The unit ball of any norm is convex as it is the sublevel set of a convex function.
- The distance function to a convex set S in a given norm $\|\cdot\|$ is defined as

$$\operatorname{dist}_{S,\|\cdot\|}(x) \coloneqq \inf_{y\in S} \|x-y\|$$

  is a convex function. Its ϵ-sublevel set is called the ϵ-neighborhood of S and is a convex set. □

3.2 *Separation of convex sets*

All norms $\|\cdot\|$ in this section are the ℓ_2 norm (the induced norm).

Definition 11. Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Given a point $x \in \mathbb{R}^n$, we define its projection onto C, i.e., $\Pi_C(x) : \mathbb{R}^n \to C$, as

$$\Pi_C(x) = \arg\min_{y\in C} \|x-y\|^2 .$$ □

Remark 1. $\Pi_C(x)$ is well-defined (i.e., it exists and is unique): Fix an arbitrary $c \in C$. Note that the optimum value of $\inf_{y\in C} \|x-y\|^2$ is achieved if and only if the optimum value of

$$\inf_{y\in C} \left\{ \|x-y\|^2 : \|x-y\|^2 \le \|x-c\|^2 \right\}$$

is achieved. The feasible domain of this problem is compact whence the continuous function $\|x-y\|^2$ achieves its minimum value.

To see that the minimizer is unique, suppose $y_1 \neq y_2$ both achieve the minimum. Then by convexity, we have $y \coloneqq (y_1+y_2)/2 \in C$. However $\|x-y\|^2 < \|x-y_1\|^2$, a contradiction. □

Π_C admits a variational characterization.

Theorem 5. *$y_x = \Pi_C(x)$ if and only if $y_x \in C$ and*

$$\langle x-y_x, y-y_x\rangle \le 0,\ \forall y \in C.$$

Proof. $(\Rightarrow)$. By definition of $\Pi_C(x)$, we have that $y_x \in C$. Let $y \in C$ and $\alpha \in (0,1]$. As C is convex, $(1-\alpha)y_x + \alpha y \in C$. Then

$$\begin{aligned}
\|x-y_x\|^2 &\le \|(1-\alpha)y_x + \alpha y - x\|^2 \\
&= \|\alpha(y-y_x) - (x-y_x)\|^2 \\
&= \alpha^2 \|y-y_x\|^2 - 2\alpha \langle x-y_x, y-y_x\rangle + \|x-y_x\|^2 .
\end{aligned}$$

Rearranging and dividing by α, we get

$$\langle x-y_x, y-y_x\rangle \le \frac{\alpha}{2} \|y-y_x\|^2 .$$

This holds for all $\alpha \in (0,1]$. Taking $\alpha \to 0$ gives the desired inequality.

($\Leftarrow$). Suppose $\bar{y} \in C$ is such that for all $y \in C$, we have $\langle x - \bar{y}, y - \bar{y}\rangle \leq 0$. Then for all $y \in C$,

$$\begin{aligned}\|x-\bar{y}\|\left(\|x-\bar{y}\| - \|x-y\|\right) &= \|x-\bar{y}\|^2 - \|x-\bar{y}\|\,\|y-x\| \\ &\leq \|x-\bar{y}\|^2 + \langle x-\bar{y}, y-x\rangle \\ &= \langle x-\bar{y}, y-x+x-\bar{y}\rangle \\ &\leq \langle x-\bar{y}, y-\bar{y}\rangle \\ &\leq 0.\end{aligned}$$

Thus, $\|x-\bar{y}\| \leq \|x-y\|$ for all $y \in C$ implying $\bar{y} = \Pi_C(x)$. ■

Theorem 6. *Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed, convex set. Let $x \notin C$. Then there exists $s \in \mathbb{R}^n$, such that*

$$\langle s, x\rangle > \sup_{y\in C} \langle s, y\rangle .$$

Proof. Let $s = x - \Pi_C(x)$. Note because $x \notin C$, we have that s is nonzero. Applying our variational characterization, we have that for all $y \in C$,

$$\begin{aligned}0 &\geq \langle x - \Pi_C(x), y - \Pi_C(x)\rangle \\ &= \langle s, y - x + s\rangle \\ &= \|s\|^2 + \langle s, y - x\rangle .\end{aligned}$$

Thus, for all $y \in C$, we have that $\langle s, y\rangle \leq \langle s, x\rangle - \|s\|^2$ is uniformly bounded away from $\langle s, x\rangle$. Taking the supremum over $y \in C$ concludes the proof. ■

Definition 12. Given a nonzero vector $a \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, define

- Hyperplane: $H_{a,\alpha} = \{x \in \mathbb{R}^n : \langle a, x\rangle = \alpha\}$
- (Closed) halfspace: $H^{\geq}_{a,\alpha} = \{x \in \mathbb{R}^n : \langle a, x\rangle \geq \alpha\}$
- Open halfspace: $H^{>}_{a,\alpha} = \{x \in \mathbb{R}^n : \langle a, x\rangle > \alpha\}$

Similarly define $H^{\leq}_{a,\alpha}$ and $H^{<}_{a,\alpha}$. □

Definition 13. Suppose C and D are nonempty subsets of $\mathbb{R}^n$. Let $a \in \mathbb{R}^n$ be nonzero and $\alpha, \beta \in \mathbb{R}$.

- We say $H_{a,\alpha}$ *separates* C and D if $C \subseteq H^{\leq}_{a,\alpha}$ and $D \subseteq H^{\geq}_{a,\alpha}$.
- We say $H_{a,\alpha}$ *strictly separates* C and D if $C \subseteq H^{<}_{a,\alpha}$ and $D \subseteq H^{>}_{a,\alpha}$.
- We say C and D can be *strongly separated* if there exists nonzero $a \in \mathbb{R}^n$ and $\alpha < \beta$ such that $C \subseteq H^{\leq}_{a,\alpha}$ and $D \subseteq H^{\geq}_{a,\beta}$. □

In this language, our previous theorem stated:

Theorem 7 (Strong separation of convex set and point)**.** *Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and let $x \notin C$. Then $\{x\}$ and C can be strongly separated.*

Corollary 2 (Strong separation of convex sets)**.** *Let $C, D \subseteq \mathbb{R}^n$ be nonempty closed convex sets with an empty intersection. Suppose further that C is bounded. Then C and D can be strongly separated.*

Proof. Consider the function $c \mapsto \operatorname{dist}(c, D)$. This function is continuous. Thus as C is compact, the minimum value

$$\min_{c \in C} \operatorname{dist}(c, D)$$

is achieved. Recalling that $\Pi_D(c)$ is well-defined, we have a pair $(\bar{c}, \bar{d})$ minimizing $\min_{c \in C,\, d \in D} \|c - d\|$. Note that $\Pi_D(\bar{c}) = \bar{d}$ and $\Pi_C(\bar{d}) = \bar{c}$.

Let $s = \bar{d} - \bar{c}$ and note that s is nonzero. Applying the variational characterization to $\{\bar{d}\}$ and C, we have that for all $c \in C$:

$$\langle s, c \rangle \leq \langle s, \bar{d} \rangle - \|s\|^2 =: \alpha$$

so that $C \subseteq H^{\leq}_{s,\alpha}$. Applying the variational characterization to $\{\bar{c}\}$ and D, we have that for all $d \in D$:

$$\langle s, d \rangle \geq \langle s, \bar{c} \rangle + \|s\|^2 =: \beta$$

so that $D \subseteq H^{\geq}_{s,\beta}$. Note that $\alpha < \beta$. ■

Theorem 8. *Suppose $C, D \subseteq \mathbb{R}^n$ are disjoint nonempty convex sets. Then, C and D can be separated.*

Proof. As C, D are disjoint, it holds that $0 \notin C - D$. For convenience, write $K \coloneqq C - D$. We have that K is a convex set not containing 0. Note that C and D can be separated if and only if 0 and K can be separated. We will without loss of generality assume that K is full-dimensional.

If $0 \notin \operatorname{cl}(K)$, then we can apply the previous theorem to separate 0 and K.

Else, suppose $0 \in \operatorname{bd}(K)$. By Corollary 1 (i.e., that the relative interior of a convex set is dense in its closure), there exists $x_i \in \operatorname{int}(K)$ so that $x_i \to 0$. As $0 \notin K$, we have that $-x_i \notin \operatorname{cl}(K)$. By the previous theorem, there exists a hyperplane v_i strongly separating $\operatorname{cl}(K)$ with $-x_i$. Without loss of generality $\|v_i\| = 1$. We have

$$-\langle v_i, x_i \rangle \geq \inf_{x \in K} \langle v_i, x \rangle .$$

Taking subsequential limits on both sides gives $\inf_{x \in K} \langle w, x \rangle \geq 0$. ■

3.3 Dual cones

Definition 14. Let K be a convex cone, the *dual cone* K_* is

$$K_* = \{y : \langle x, y\rangle \geq 0, \forall x \in K\} . \qquad \square$$

Example 9. Examples of cones and their duals:

- The SDP cone $\mathbb{S}^n_+$, the Lorentz cone $\mathcal{L}^n$, and the nonnegative orthant $\mathbb{R}^n_+$ are all self-dual.
- Let $p \in [1, \infty]$ and let $K_p = \Big\{(x, t) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_p \leq t\Big\}$. Then $(K_p)_* = K_q$ where q is the Hölder dual of p. $\square$

Lemma 9. *For any closed convex cone K, we have $(K_*)_* = K$.*

Proof. $K \subseteq (K_*)_*$ by definition: Indeed, suppose $y \in K_*$ and $x \in K$, then $\langle y, x\rangle \geq 0$.

Next, suppose $\bar{x} \notin K$. As K is a closed convex set and $\{\bar{x}\}$ is compact convex, Corollary 2 implies there exists v so that

$$\langle v, \bar{x}\rangle < \inf_{x\in K} \langle v, x\rangle .$$

As K is a cone, the RHS must equal zero so that $v \in K_*$. We deduce that the LHS is negative so that $\bar{x} \notin (K_*)_*$. ■

Definition 15. A cone $K \subseteq \mathbb{R}^n$ is *pointed* if $K \cap -K = \{0\}$. Alternatively, a cone K is pointed if and only if does not contain any lines. $\square$

3.4 Basic definitions about general convex programs

Definition 16. A *convex optimization problem/convex program* is a problem of the form

$$\inf_{x\in\Omega} f(x)$$

where $\Omega \subseteq \mathbb{R}^n$ is a convex set and $f : \Omega \to \mathbb{R}$ is convex. The objects x, Ω, f are referred to as the *decision variable*, the *domain*, and *objective function* respectively.

- An *optimal solution* $x^\star$ is a point $x^\star \in \Omega$ so that $f(x^\star) \leq f(x)$ for all $x \in \Omega$. An optimal solution does not have to exist or be unique. When an optimal solution exists we say that the problem is *solvable.*
- The *optimal value* is $\inf_{x\in\Omega} f(x)$. We define the value to be ∞ if Ω is empty (in which case we say the problem is *infeasible*). If the value is $-\infty$, we say the problem is *unbounded below.* Else, it is *bounded below.*

- Often, the domain will be defined by *constraints*, for example,

$$\Omega = \{x \in \mathbb{R}^n : \text{some constraints}\}. \qquad \square$$

Definition 17. Given a feasible point $x^\star \in \Omega$, the descent cone at $x^\star$ is

$$\operatorname{cone}\left(\left\{\delta \in \mathbb{R}^n : \begin{array}{l} f(x^\star + \delta) \le f(x^\star) \\ x^\star + \delta \in \Omega \end{array}\right\}\right).$$

It is the set of infinitesimal directions so that moving in that direction produces a feasible point with nonincreasing objective value. $\square$

Exercise

1. Give an example of a pair of disjoint nonempty closed convex sets that cannot be strictly separated.

Problems

1. Prove that the nonnegative orthant, second-order cone, and semidefinite cones are self-dual.

2. In sparse recovery, the goal is to recover a sparse vector $x^\star \in \mathbb{R}^n$ given linear measurements $(A, b) \in \mathbb{R}^{m \times n} \times \mathbb{R}^m$ where $b = Ax^\star$. A convex-optimization approach to this problem is to output the optimizer of

$$\min_{x \in \mathbb{R}^n} \left\{\|x\|_1 : Ax = b\right\}.$$

   This problem gives a necessary and sufficient condition for when this convex-optimization approach correctly recovers any $\le k$-sparse vector $x^\star$.

   Given a subset $S \subseteq [n]$ and a vector $x \in \mathbb{R}^n$, let x_S denote the restriction of x onto the set S. Let S^c denote the complement of S.

   - Compute the *descent cone* of this convex-optimization problem at the solution $x^\star$.

     The descent cone at $x^\star$ is defined as

$$\left\{\delta \in \mathbb{R}^n : \begin{array}{l} \forall \epsilon > 0 \text{ small enough}: \\ x^\star + \epsilon\delta \text{ is feasible} \\ \text{obj. value at } x^\star + \epsilon\delta \le \text{obj. value at } x^\star \end{array}\right\}$$

   - The matrix A is said to satisfy the *nullspace property at order k* if for all sets $S \subseteq [n]$ with $|S| \le k$ and for all $\delta \in \ker(A) \setminus \{0\}$, we have

$$\|\delta_S\|_1 < \|\delta_{S^c}\|_1.$$

Show that the descent cone at $x^\star$ is trivial if A satisfies the nullspace property at order k.

- Show that if A does not satisfy the nullspace property, then there exists a k-sparse $x^\star$ for which the convex-optimization approach may fail to recover $x^\star$.

3. Given a permutation σ of $[n]$, we can associate σ with the $n \times n$ permutation matrix

$$(X^\sigma)_{i,j} = \begin{cases} 1 & \text{if } \sigma(i) = j \\ 0 & \text{else} \end{cases}.$$

Prove that the convex hull of the $n!$ permutation matrices is given by the set of doubly stochastic matrices:

$$\mathrm{DS}(n) := \left\{ X \in \mathbb{R}^{n\times n} : \begin{array}{c} X \geq 0 \\ X^\mathsf{T} 1_n = 1_n \\ X 1_n = 1_n \end{array} \right\}.$$

Hint: Use Hall's marriage theorem to prove that the support of any doubly stochastic matrix contains a permutation matrix.

4
Conic programming I

4.1 What is a conic program

Recall the standard linear program with inequality constraints and equality constraints:

$$\min_{x\in\mathbb{R}^n}\left\{c^\intercal x:\begin{array}{c}Ax\geq a\\ Bx=b\end{array}\right\}.$$

The constraint $Ax\geq b$ can be rewritten $Ax-b\geq 0$ or $Ax-b\in\mathbb{R}^m_+$. Central to the definition of a linear program is the cone $\mathbb{R}^m_+$ that gives us an ordering on vectors, i.e., for vectors x and $y\in\mathbb{R}^m$, the cone $\mathbb{R}^m_+$ imposes a partial ordering where $x\geq y$ if and only if $x-y\in\mathbb{R}^m_+$. A conic program generalizes a linear program by consider other interesting orderings on vectors.

Definition 18. A *conic program* in standard form is an optimization problem of the form

$$\inf_{x\in\mathbb{R}^n}\left\{c^\intercal x:\begin{array}{c}Ax-a\in K\\ Bx-b=0\end{array}\right\},$$

where c,A,a,B,b are matrices/vectors of compatible dimensions and K is a convex cone.[1] $\square$

[1] We will usually impose additional constraints on the convex cone to get "well-behaved" conic programs.

Example 10. The optimization problem,

$$\inf_{x\in\mathbb{R}^n}\left\{c^\intercal x:\ \|x-\mu_i\|_2\leq r_i,\ \forall i\in[m]\right\},\tag{4.1}$$

is a conic program. Here, $c\in\mathbb{R}^n$, $\mu_i\in\mathbb{R}^n$, and $r_i\in\mathbb{R}$. To put this program into the standard form we can write

$$(4.1)=\inf_{x\in\mathbb{R}^n}\left\{c^\intercal x:\begin{array}{c}\begin{pmatrix}I_n\\0^\intercal\\ \vdots\\ I_n\\ 0^\intercal\end{pmatrix}x-\begin{pmatrix}\mu_1\\-r_1\\ \vdots\\ \mu_m\\-r_m\end{pmatrix}\in(\mathcal{L}^{n_1+1})^m\\ 0x=0\end{array}\right\}.\qquad\square$$

4.2 Weak Conic Duality

Consider a standard conic program

$$\text{(Primal)} \qquad \inf_{x\in\mathbb{R}^n} \left\{ c^\intercal x : \begin{array}{l} Ax - a \in K \\ Bx - b = 0 \end{array} \right\}.$$

For concreteness, suppose $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^k$.

Duality theory begins with the question: "how do we prove lower bounds on the optimal value of (Primal)?"

Recall the definition of the dual cone

$$K_* := \{y \in \mathbb{R}^m : \langle y, u\rangle \geq 0, \forall u \in K\}.$$

Then, for any $y \in K_*$ and any $z \in \mathbb{R}^k$ and any feasible x in (Primal), we can derive the valid inequality

$$0 \leq \langle Ax - a, y\rangle + \langle Bx - b, z\rangle = \langle A^\intercal y + B^\intercal z, x\rangle - \langle a, y\rangle - \langle b, z\rangle.$$

Rearranging, we have that $\langle A^\intercal y + B^\intercal z, x\rangle \geq \langle a, y\rangle + \langle b, z\rangle$. Thus, if $y \in K_*$, $z \in \mathbb{R}^k$ satisfies $A^\intercal y + B^\intercal z = c$ then $\langle a, y\rangle + \langle b, z\rangle$ is a valid lower bound on the optimal value of (Primal). The dual conic program optimizes this lower bound:

$$\begin{aligned} \text{(Dual)} \qquad & \sup_{y\in\mathbb{R}^m, z\in\mathbb{R}^k} \left\{ \langle a, y\rangle + \langle b, z\rangle : \begin{array}{l} A^\intercal y + B^\intercal z = c \\ y \in K_* \end{array} \right\} \\ &= \sup_{y\in\mathbb{R}^m, z\in\mathbb{R}^k} \left\{ \left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} y \\ z \end{pmatrix} \right\rangle : \begin{array}{l} \begin{pmatrix} I_m & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} \in K_* \\ \begin{pmatrix} A^\intercal & B^\intercal \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} - c = 0 \end{array} \right\}. \end{aligned}$$

Thus, the dual of a conic program is again a conic program.

Theorem 9 (Weak conic duality)**.** Opt(Primal) $\geq$ Opt(Dual).

Proof. Suppose $x \in \mathbb{R}^n$ is feasible in the primal and suppose $(y, z) \in \mathbb{R}^m \times \mathbb{R}^k$ is feasible in the dual. Then

$$\begin{aligned} \langle c, x\rangle &= \langle A^\intercal y + B^\intercal z, x\rangle = \langle y, Ax\rangle + \langle z, Bx\rangle \\ &= \langle a, y\rangle + \langle b, z\rangle + \langle Ax - a, y\rangle + \langle Bx - b, z\rangle \\ &\geq \langle a, y\rangle + \langle b, z\rangle. \end{aligned}$$

■

This is known as *weak* conic duality because of the inequality in the theorem and is not a fully satisfactory duality theory. Specifically, compare the case of Linear Programming where equality always holds. In many situations, we can prove a stronger version of this result called *strong conic duality* where the inequality is replaced with an equality.

Remark 2. The definition of the dual of a conic program assumes that the conic program comes in standard form. In practice, this is usually not the case and we may see programs that look like

$$\inf_{x \in \mathbb{R}^n} \left\{ \langle c, x \rangle : \begin{array}{l} A_1 x - a_1 \in K_1 \\ \vdots \\ A_r x - a_r \in K_r \\ Bx - b = 0 \end{array} \right\}.$$

Recall that the dual of the product of cones is the product of the duals. In particular, the dual of this conic program is

$$\sup_{y_1, \ldots, y_r, z} \left\{ \sum_{i=1}^{r} \langle a_i, y_i \rangle + \langle b, z \rangle : \begin{array}{l} \sum_{i=1}^{r} A_i^{\intercal} y_i + B^{\intercal} z = c \\ y_i \in (K_i)_*, \, \forall i = 1, \ldots, r \end{array} \right\}. \qquad \square$$

4.3 Cones and inequalities

In order to prove strong duality, we will need to impose further assumptions on the cone K.

We note some properties of $\geq$ in the LP setting. First, the $\geq$ inequality gives a partial ordering over $\mathbb{R}^m$:

Definition 19. A binary relation $\succeq$ is a *partial ordering* on $\mathbb{R}^m$ if for all $x, y, z \in \mathbb{R}^m$

- (Reflexive) $x \succeq x$
- (Antisymmetric) If $x \succeq y$ and $y \succeq x$, then $x = y$
- (Transitive) If $x \succeq y$ and $y \succeq z$, then $x \succeq z$. $\square$

Additionally, the $\geq$ inequality is compatible with linear operations:

1. (Additive) if $a \geq b$ and $c \geq d$, then $a + c \geq b + d$
2. (Positive homogeneous) if $a \geq b$ and $\lambda \in \mathbb{R}_+$, then $\lambda a \geq \lambda b$

Definition 20. Let $\succeq$ be a binary relation on vectors. We call this relation a *good inequality* if it induces a partial ordering on $\mathbb{R}^m$ that is additive and positively homogeneous. $\square$

A good inequality $\succeq$ is completely characterized by the set $K_{\succeq} = \{y \in \mathbb{R}^m : y \succeq 0\}$. This is immediate as, for a good inequality $\succeq$, we have $a \succeq b$ if and only if $a - b \succeq 0$.

Lemma 10. *If $\succeq$ is a good inequality, then $K_{\succeq}$ is a nonempty pointed cone. Conversely, if K is a nonempty pointed cone, then the binary relation $\succeq_K$ where $a \succeq_K b$ if and only if $a - b \in K$ is a good inequality.*

In addition, we will assume two more properties so that we can pass to limits and talk about strict inequalities. In particular, from now on we will only work with *closed* pointed convex cones with *nonempty interior*.

Definition 21. A *proper cone*[2] is a closed pointed convex cone with nonempty interior. □

[2] Also known as a regular cone

When K is a proper cone, we can pass to limits on both sides of the associated conic inequality: Suppose K is a proper cone, $\lim_{i\to\infty} a_i = a$ and $\lim_{i\to\infty} b_i = b$ then

$$a_i \succeq_K b_i, \forall i \implies a \succeq_K b.$$

We can also *define* a strict inequality, i.e.

$$a \succ_K b \Leftrightarrow a - b \in \operatorname{int}(K).$$

This strict inequality has the property that for any $a \in \mathbb{R}^m$ and any $b \succ_K 0$ and all λ large enough, that $a + \lambda b \succ_K 0$.

Example 11. The cone of positive semidefinite matrices $\mathbb{S}^m_+$ is a proper cone. Let $\mathcal{A} : \mathbb{R}^n \to \mathbb{S}^m$ be a linear operator and let $A \in \mathbb{S}^m$. Let $B \in \mathbb{R}^{k\times n}$ and $b \in \mathbb{R}^k$. The following conic program

$$\inf_{x\in\mathbb{R}^n} \left\{ c^\mathsf{T} x : \begin{array}{l} \mathcal{A}(x) - A \in \mathbb{S}^m_+ \\ Bx - b = 0 \end{array} \right\}$$

is known as a *semidefinite program*. □

5

Conic programming II

5.1 Strong Conic Duality

Definition 22. We say that (Primal) is strictly feasible if there exists $\bar{x} \in \mathbb{R}^n$ so that

$$A\bar{x} - a \in \text{int}(K) \quad \text{and} \quad B\bar{x} - b = 0.$$

(Dual) is strictly feasible if there exists $(\bar{y}, \bar{z})$ so that

$$y \in \text{int}(K_*) \quad \text{and} \quad A^\intercal \bar{y} + B^\intercal \bar{z} = c. \qquad \square$$

We are now ready to state the strong conic duality theorem.

Theorem 10 (Strong conic duality)**.** *Consider primal* (Primal) *and its dual* (Dual)*. Suppose K is a regular cone and suppose the linear systems in both* (Primal) *and* (Dual) *are feasible, i.e.,*

$$\exists \bar{x} : \; B\bar{x} - b = 0$$
$$\exists (\bar{y}, \bar{z}) : \; A^\intercal \bar{y} - B^\intercal \bar{z} - c = 0.$$

Then

- Symmetry*: the dual problem to* (Dual) *is (*Primal*).*
- Weak duality*: for primal feasible $\bar{x}$ and dual feasible* $(\bar{y}, \bar{z})$*,*

$$\langle c, \bar{x} \rangle \geq \langle a, \bar{y} \rangle + \langle b, \bar{z} \rangle .$$

- Strong duality under strict feasibility*: if* (Primal) *is strictly feasible with bounded objective, then* (Dual) *is solvable and* Opt(Primal) = Opt(Dual)*.*

  The same statement holds with the roles of (Primal) *and* (Dual) *interchanged. In particular, if both are strictly feasible, then both are solvable.*

Proof. **Proof of symmetry:** We can write (Dual) as

$$\sup_{y\in\mathbb{R}^m, z\in\mathbb{R}^k} \left\{ \left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} y \\ z \end{pmatrix} \right\rangle : \begin{array}{l} \begin{pmatrix} I_m & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} \in K_* \\ \begin{pmatrix} A^\intercal & B^\intercal \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} - c = 0 \end{array} \right\}$$

Thus the dual to (Dual) is

$$\begin{aligned} &\inf_{\xi\in\mathbb{R}^m, \zeta\in\mathbb{R}^n} \left\{ \left\langle \begin{pmatrix} 0 \\ c \end{pmatrix}, \begin{pmatrix} \xi \\ \zeta \end{pmatrix} \right\rangle : \begin{array}{l} \begin{pmatrix} I_m & 0 \end{pmatrix} \begin{pmatrix} \xi \\ \zeta \end{pmatrix} \in K \\ \begin{pmatrix} I_m & A \\ 0 & B \end{pmatrix} \begin{pmatrix} \xi \\ \zeta \end{pmatrix} - \begin{pmatrix} a \\ b \end{pmatrix} = 0 \end{array} \right\} \\ &= \inf_{\xi\in\mathbb{R}^m, \zeta\in\mathbb{R}^n} \left\{ \langle c, \zeta \rangle : \begin{array}{c} \xi \in K \\ \xi + A\zeta = a \\ B\zeta = b \end{array} \right\} \\ &= \inf_{\zeta\in\mathbb{R}^n} \left\{ \langle c, \zeta \rangle : \begin{array}{c} A\zeta - a \in K \\ B\zeta - b = 0 \end{array} \right\}. \end{aligned}$$

We recognize (Primal).

Proof of weak duality: this was already done.

Proof of strong duality under strict feasibility: Assume that (Primal) is strictly feasible with bounded objective. As weak duality holds, it suffices to construct a dual feasible solution with value $\geq$ Opt(Primal).

Define

$$\mathcal{S}_1 := \left\{ \begin{pmatrix} \langle c, x \rangle \\ Ax - a \end{pmatrix} : Bx - b = 0 \right\}$$
$$\mathcal{S}_2 := \left\{ \begin{pmatrix} \lambda \\ \zeta \end{pmatrix} : \begin{array}{l} \lambda < \mathrm{Opt(Primal)} \\ \zeta \in K \end{array} \right\}$$

We apply[1] the hyerplane separation theorem to $\mathcal{S}_1$ and $\mathcal{S}_2$ to get a nonzero vector $(t, \bar{y}) \in \mathbb{R}^{1+m}$ so that

$$\inf_{(\lambda,\zeta)\in\mathcal{S}_1} t\lambda - \langle \bar{y}, \zeta \rangle \geq \sup_{(\lambda,\zeta)\in\mathcal{S}_2} t\lambda - \langle \bar{y}, \zeta \rangle$$

Notice that $t \geq 0$ and $\bar{y} \in K_*$. Indeed, if $t < 0$, then we may approach $(-\infty, 0) \in \mathcal{S}_2$ to set the RHS arbitrarily positive, a contradiction. Similarly, if $d \notin K_*$, then there exists $\zeta \in K$ so that $\langle \bar{y}, \zeta \rangle < 0$. Again, we can approach $(\mathrm{Opt(Primal)} - 1, \infty\zeta) \in \mathcal{S}_2$ to set the RHS arbitrarily positive, a contradiction.

We must also have that $t \neq 0$. Indeed, suppose $t = 0$ and recall that by strict feasibility,[2] there exists $\bar{x}$ so that $B\bar{x} - b = 0$ and

[1] **Exercise:** Verify that the assumptions of the hyperplane separation theorem hold.

[2] This is the only place we used strict feasibility

$A\bar{x} - a \in \operatorname{int}(K)$. As $(t, \bar{y})$ is nonzero, we must have that $\bar{y} \in K_*$ is nonzero. Thus, there exists $(\lambda, \zeta) \in \mathcal{S}_1$ achieving

$$t\lambda - \langle \bar{y}, \zeta \rangle = -\langle \bar{y}, A\bar{x} - a \rangle < 0.$$

On the other hand, $(\operatorname{Opt}(\text{Primal}) - 1, 0) \in \mathcal{S}_2$ achieves

$$t\lambda - \langle \bar{y}, \zeta \rangle = 0,$$

a contradiction.

As $t > 0$, we can normalize $t = 1$ in the separation statement. We now rewrite the separation statement. The RHS is equal to $\operatorname{Opt}(\text{Primal})$. Thus, for all $x \in \mathbb{R}^n$ satisfying $Bx - b = 0$, we have that

$$\langle c - A^\intercal \bar{y}, x \rangle + \langle a, \bar{y} \rangle = \langle c, x \rangle - \langle \bar{y}, Ax - a \rangle \geq \operatorname{Opt}(\text{Primal}).$$

We conclude that $c - A^\intercal \bar{y} \in \ker(B)^\perp = \operatorname{range}(B^\intercal)$ so that there exists $\bar{z} \in \mathbb{R}^k$ satisfying

$$c - A^\intercal \bar{y} = B^\intercal \bar{z}.$$

Let $\bar{x}$ satisfy $Bx - b = 0$. We deduce that $(\bar{y}, \bar{z})$ satisfies

$$\begin{aligned}
\langle a, \bar{y} \rangle + \langle b, \bar{z} \rangle &= \langle a, \bar{y} \rangle + \langle B\bar{x}, \bar{z} \rangle \\
&= \langle a, \bar{y} \rangle + \langle \bar{x}, B^\intercal \bar{z} \rangle = \langle a, \bar{y} \rangle + \langle C - A^\intercal \bar{y}, \bar{x} \rangle \\
&\geq \operatorname{Opt}(\text{Primal}).
\end{aligned}$$
■

Corollary 3. *Suppose* (Primal) *and* (Dual) *are strictly feasible. Let $\bar{x}$ and $(\bar{y}, \bar{z})$ be primal and dual feasible solutions. Then the following are equivalent*

- *$\bar{x}$ and $(\bar{y}, \bar{z})$ are both optimal*
- *Zero duality gap: $\langle c, \bar{x} \rangle = \langle a, \bar{y} \rangle + \langle b, \bar{z} \rangle$*
- *Complementary slackness: $\langle \bar{y}, A\bar{x} - a \rangle = 0$*

Example 12. Consider the following problem: Given $\mu_1, \ldots, \mu_k \in \mathbb{R}^n$, find $x \in \mathbb{R}^n$ minimizing $\sum_{i=1}^k \|x - \mu_i\|_2$.

Recall that

$$\begin{pmatrix} t \\ x - \mu_i \end{pmatrix} \in \mathcal{L}^{1+n} \quad \iff \quad \|x - \mu_i\|_2 \leq t.$$

Thus, we can write the above problem as

$$\inf_{x \in \mathbb{R}^n, t_1, \ldots, t_k \in \mathbb{R}} \left\{ \sum_{i=1}^k t_i : \begin{pmatrix} t_i \\ x \end{pmatrix} - \begin{pmatrix} 0 \\ \mu_i \end{pmatrix} \in \mathcal{L}^{1+n}, \, \forall i = 1, \ldots, k \right\}$$

The dual problem has k variables of the form $(\xi_i, \zeta_i) \in \mathcal{L}_*^{1+n} = \mathcal{L}^{1+n}$. It is given by

$$\sup_{(\xi_1,\zeta_1),\ldots,(\xi_k,\zeta_k)\in\mathbb{R}^{1+n}} \left\{ \sum_{i=1}^{k} \langle \mu_i, \zeta_i \rangle : \begin{array}{l} \xi_1, \ldots, \xi_k = 1 \\ \sum_{i=1}^{k} \zeta_i = 0 \\ \begin{pmatrix} \xi_i \\ \zeta_i \end{pmatrix} \in \mathcal{L}^{1+n}, \, \forall i = 1, \ldots, k \end{array} \right\}$$

$$= \sup_{\zeta_1,\ldots,\zeta_k\in\mathbb{R}^n} \left\{ \sum_{i=1}^{k} \langle \mu_i, \zeta_i \rangle : \begin{array}{l} \sum_{i=1}^{k} \zeta_i = 0 \\ \|\zeta_i\|_2 \leq 1, \, \forall i = 1, \ldots, k \end{array} \right\}.$$

The primal and dual are both strictly feasible so that both programs achieve their optimal solutions and the optimal values are equal. ☐

Remark 3. Up to now, the conic programs we have considered write the affine constraints $Bx - b$ separately from the conic constraint $Ax - a \in K$. In the future, we will combine the two and simply write

$$\inf_{x\in\mathbb{R}^n} \left\{ c^{\mathsf{T}} x : \, Ax - a \in K \right\}.$$

In this form, we can still apply the results in the conic programming lectures. Obviously, we could treat this as a conic program in the previous form without the $Bx - b = 0$ term and apply the previous results verbatim. Alternatively, we can get a more powerful duality result by first "pulling out" the affine constraints implied by $Ax - a \in K$ before applying the duality results. The effect of this is that the strict feasibility conditions will become weaker conditions. ☐

Problems

1. Consider an optimization problem of the form

$$\inf_{x\in\mathbb{R}^n} \{f(x) : g_i(x) \le 0,\ \forall i \in [m]\}.$$

We make no assumptions on whether f or $g_1, \dots, g_m$ is convex.
Define

$$\mathcal{I} := \left\{ \begin{pmatrix} f(x) \\ g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix} : x \in \mathbb{R}^n \right\} + \mathbb{R}^{1+m}_+.$$

- Show that if f and $g_1, \dots, g_m$ are convex functions, then $\mathcal{I}$ is a convex set.
- Adapt the proof of strong conic duality to show that if $\mathcal{I}$ is convex and there exists $\bar{x}$ so that $g_i(\bar{x}) < 0$ for all $i \in [m]$, then

$$\begin{aligned} &\inf_{x\in\mathbb{R}^n} \{f(x) : g_i(x) \le 0,\ \forall i \in [m]\} \\ &= \sup_{u\in\mathbb{R},\,\lambda\in\mathbb{R}^m} \left\{ u : \begin{array}{l} \lambda \ge 0 \\ f(x) + \sum_{i=1}^m \lambda_i g_i(x) \ge u,\ \forall x \in \mathbb{R}^n \end{array} \right\}. \end{aligned}$$

This statement is known as hidden convexity and allows us to extend convex optimization theory to some very special nonconvex optimization problems where $\mathcal{I}$ is convex despite f, g_i being possibly nonconvex.

2. Suppose K is a proper cone and consider the primal and dual conic problems:

$$\inf_{x\in\mathbb{R}^n} \left\{ c^\intercal x : \begin{array}{l} Ax - a \in K \\ Bx - b = 0 \end{array} \right\} \ge \sup_{y\in\mathbb{R}^m,\, z\in\mathbb{R}^k} \left\{ \langle a, y\rangle + \langle b, z\rangle : \begin{array}{l} A^\intercal y + B^\intercal z = c \\ y \in K_* \end{array} \right\}.$$

Assume that the primal problem is feasible. Prove that the primal problem has bounded sublevel sets, i.e.,

$$\forall t \in \mathbb{R},\ \text{ the set } \left\{ x \in \mathbb{R}^n : \begin{array}{c} c^\intercal x \le t \\ Ax - a \in K \\ Bx - b = 0 \end{array} \right\} \text{ is bounded}$$

if and only if the dual problem is strictly feasible.

3. We show that strong duality may fail in general for conic programs without further assumptions. Consider the following SDP.

$$\inf_{X\in\mathbb{S}^2} \left\{ 2X_{1,2} : \begin{array}{l} X_{1,1} = 0 \\ X \succeq 0 \end{array} \right\}$$

Write its dual and compute the optimal value for both the primal and dual.

6
SOCP representability

The following two lectures introduce two classes of conic optimization problems: second-order cone programs (SOCPs) and semidefinite programs (SDPs).

Any LP is an SOCP and any SOCP is an SDP. Thus, SDPs give the most modeling power of these three classes of conic programs. On the other hand, algorithms for solving LPs generally run faster than algorithms for solving SOCPs, than algorithms for solving SDPs.

This motivates the need to understand what can be modeled in the class of SOCPs and what can be modeled in the class of SDPs.

6.1 Second-order cone programming/conic quadratic program

Definition 23. A second-order cone program (SOCP), also known as a *Conic quadratic program* (CQP), is a conic program where the cone K is a direct product of finitely many second-order cones:

$$\inf_{x\in\mathbb{R}^n} \left\{ c^\intercal x : \begin{array}{l} Ax - a \in K \\ Bx - b = 0 \end{array} \right\}, \qquad K = \mathcal{L}^{1+n_1} \times \cdots \times \mathcal{L}^{1+n_k}. \qquad \square$$

Example 13 (Any LP is an SOCP)**.** Consider a linear constraint in x:

$$a^\intercal x \geq \alpha \iff \begin{pmatrix} a^\intercal x - \alpha \\ 0 \end{pmatrix} \in \mathcal{L}^2. \qquad \square$$

Definition 24. We say that $X \subseteq \mathbb{R}^n$ is a second-order cone representable (SOCR) set if there exists a set

$$S = \left\{ (x,u) \in \mathbb{R}^n \times \mathbb{R}^{n'} : A_i(x,u) - b_i \in \mathcal{L}^{1+n_j}, \forall i \in [k] \right\}$$

such that $X = \Pi_x S$ where $\Pi_x(x,u) \coloneqq x$.

We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is SOCR if

$$\mathrm{epi}(f) \coloneqq \left\{ (x,t) \in \mathbb{R}^{n+1} : f(x) \leq t \right\}$$

is a SOCR set. $\square$

We care about SOCR sets and functions because they can be used as building blocks for SOCPs. Suppose $f_0, \dots, f_k : \mathbb{R}^n \to \mathbb{R}$ are SOCR functions and $\mathcal{X} \subseteq \mathbb{R}^n$ is a SOCR set. Then,

$$\inf_{x \in \mathbb{R}^n} \left\{ f_0(x) : \begin{array}{l} f_i(x) \leq 0, \ \forall i \in [k] \\ x \in \mathcal{X} \end{array} \right\}$$

can be converted into an SOCP.[1]

[1] **Exercise**: Verify this.

Example 14. • $f(x) = \|x\|$ is SOCR:

$$\|x\| \leq t \iff \begin{pmatrix} t \\ x \end{pmatrix} \in \mathcal{L}^{1+n}$$

• $f(x) = \|x\|^2$ is SOCR:

$$\begin{aligned} \|x\|^2 \leq t &\iff \|x\|^2 + \left(\frac{t-1}{4}\right)^2 \leq \left(\frac{t+1}{4}\right)^2 \\ &\iff t+1 \geq 0 \quad \text{and} \quad \begin{pmatrix} (t+1)/4 \\ (t-1)/4 \\ x \end{pmatrix} \in \mathcal{L}^{1+(1+n)}. \end{aligned}$$

□

Lemma 11. *Suppose $X_1, \dots X_k$ are SOCR sets where $X_i \subseteq \mathbb{R}^{n_i}$. Then,*

- *(Direct product) $\Pi_i X_i$ is SOCR*
- *(Affine image) $\{Ax + b : x \in X_1\}$ is SOCR.*
- *(Inverse affine image) $\{y : Ay + b \in X_1\}$ is SOCR.*

If additionally, $n_1 = \cdots = n_k$ then,

- *(Intersection) $\bigcap_i X_i$ is SOCR.*
- *(Minkowski sum) $\sum_i X_i$ is SOCR.*

Proof. Left as an exercise.[2] ■

[2] **Exercise**: Verify.

Example 15. • Consider a quadratic function

$$f(x) = x^\intercal A x + b^\intercal x + c.$$

We will assume that f is convex, i.e., that $A \succeq 0$. Let D so that $D^\intercal D = A$, for example, we could take $D = A^{1/2}$ which exists because A is PSD. Then,

$$\begin{aligned} f(x) \leq t &\iff x^\intercal D^\intercal D x \leq t - c - b^\intercal x \\ &\iff \|Dx\|^2 \leq t - c - b^\intercal x \end{aligned}$$

Using Lemma 11 (Inverse affine image) and the fact that $\left\{(x,t) : \|x\|^2 \leq t\right\}$ is SOCR, we see that $f(x)$ is SOCR.

- Hypograph of geometric mean of two variables

$$\left\{ \begin{pmatrix} x \\ y \\ t \end{pmatrix} \in \mathbb{R}^3 : \begin{array}{l} x, y \geq 0 \\ t \leq \sqrt{xy} \end{array} \right\}$$
$$= \left\{ \begin{pmatrix} x \\ y \\ t \end{pmatrix} \in \mathbb{R}^3 : \begin{array}{l} \exists u \in \mathbb{R} : \\ x, y \geq 0,\ 0, t \leq u \\ u \leq \sqrt{xy} \end{array} \right\}$$
$$= \left\{ \begin{pmatrix} x \\ y \\ t \end{pmatrix} \in \mathbb{R}^3 : \begin{array}{l} \exists u \in \mathbb{R} : \\ x, y \geq 0,\ 0, t \leq u \\ \begin{pmatrix} (x+y)/4 \\ (x-y)/4 \\ u \end{pmatrix} \in \mathcal{L}^{1+2} \end{array} \right\}$$

- Hypograph of geometric mean of 2^ℓ variables: We show this inductively. The case $\ell = 1$ is the previous example. Now let $\ell > 1$ and suppose the claim holds inductively for all smaller ℓ. Then,

$$\left\{ (x_1, \ldots, x_{2^\ell}, t) : \begin{array}{l} x_i \geq 0,\ \forall i \\ t \leq \left(\prod_i x_i\right)^{1/2^\ell} \end{array} \right\}$$
$$= \left\{ (x_1, \ldots, x_{2^\ell}, t) : \begin{array}{l} \exists u_{\text{left}}, u_{\text{right}} : x_i \geq 0,\ \forall i \\ 0 \leq u_{\text{left}} \leq \left(\prod_{i=1}^{2^{\ell-1}} x_i\right)^{1/2^{\ell-1}} \\ 0 \leq u_{\text{right}} \leq \left(\prod_{i=2^{\ell-1}+1}^{2^\ell} x_i\right)^{1/2^{\ell-1}} \\ t \leq \sqrt{u_{\text{left}} u_{\text{right}}} \end{array} \right\}$$

  We introduce $O(2^\ell)$ new variables and new constraints in order to represent the hypograph of the geometric mean of 2^ℓ numbers in a SOCP.

- Convex powers with rational exponents: Let $p, q \in \mathbb{N}$ so that $p/q \geq 1$. Note that $x^{p/q}$ is a convex function on $x \geq 0$. The following set is SOCR:

$$\left\{ (x,t) \in \mathbb{R}^2 : \begin{array}{l} x \geq 0 \\ x^{p/q} \leq t \end{array} \right\} = \left\{ (x,t) \in \mathbb{R}^2 : \begin{array}{l} x \geq 0 \\ x \leq \left(x^{2^\ell - p} t^q 1^{p-q} \right)^{1/2^\ell} \end{array} \right\}$$

  Here, $\ell \geq 0$ is any number so that $2^\ell \geq p, q$.

- $\ell_{p/q}$ norm: Let $p, q \in \mathbb{N}$ so that $p/q \geq 1$. We claim that

$$\left\{ (x,t) : \|x\|_{p/q} \leq t \right\} = \left\{ (x,t) : \begin{array}{l} \exists u_1, \ldots, u_n, v_1, \ldots, v_n \geq 0 : \\ u_i \geq x_i,\ u_i \geq -x_i \forall i \\ u_i \leq \left(u_i^{2^\ell - p} v_i^q t^{p-q} \right)^{1/2^\ell} \\ \sum_i v_i \leq t \end{array} \right\}.$$

The $\subseteq$ relation follows by setting $u_i = |x_i|$ and $v_i = u_i^{p/q} t^{1-p/q}$.

The $\supseteq$ relation follows as

$$\begin{aligned}
\|u\|_{p/q} &= \left(\sum_i u_i^{p/q}\right)^{q/p} \\
&\leq \left(\sum_i v_i t^{p/q-1}\right)^{q/p} \\
&\leq t. \qquad \square
\end{aligned}$$

Exercises

- Show that the following branch of the hyperbola is a SOCR set.

$$\left\{ (x, y) \in \mathbb{R}^2 : \begin{array}{l} xy \geq 1 \\ x, y \geq 0 \end{array} \right\}$$