MGMT 690—Convex Optimization

Alex L. Wang

Last updated April 26, 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 symmetric bilinear function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfying $\langle v, v \rangle \geq 0$ for all v and $\langle v, v \rangle = 0$ if and only if $v = 0$. $\square$

[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

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

  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. $\square$

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. It may be useful to first verify that the Cauchy-Schwarz inequality holds for the induced norm on any Euclidean space.

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

1.2 PSD matrices and the Singular Value Decomposition

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

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

That is, if its rows (or columns) form a set of orthonormal vectors. The set of orthogonal matrices is denoted $O(n)$. □

Definition 4. Given $A \in \mathbb{S}^n$, we say that $\lambda \in \mathbb{R}$ is an eigenvalue of A if

$$\det(A - \lambda I) = 0.$$

Equivalently, if there exists a nonzero vector $v \in \mathbb{R}^n$ so that $Av = \lambda v$. We call such a v, an eigenvector of A (corresponding to eigenvalue λ). $\square$

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 the eigenvalues of A. The ith column of U is an eigenvector of A corresponding to eigenvalue λ_i; it is not unique in general.

Definition 5. 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 $x \in \mathbb{R}^n \setminus \{0\}$ $\square$

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

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

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 $\text{Inertia}(A) \coloneqq (n_-, n_0, n_+)$ 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

$$\text{Inertia}(A) = \text{Inertia}(P^\intercal AP).$$

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

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

2

Elementary convex analysis I

2.1 Convex sets

Definition 6. A set $S \subseteq \mathbb{R}^n$ is

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

[1] Some authors call this *convex conic*.

Definition 7. 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/nonnegative* 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 $\alpha_i \geq 0$ □

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 cone □

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

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

denote the product of $S_1, \dots, S_k$. Suppose $(x_1, \dots, x_k)$ and $(y_1, \dots, y_k) \in \prod_{i=1}^k S_i$ 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,

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

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 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

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

- 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

  $$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x \geq 0\}$$

- The second order cone

  $$\mathcal{L}^{1+n} := \left\{\begin{pmatrix} t \\ x \end{pmatrix} \in \mathbb{R}^{1+n} : \|x\|_2 \leq t\right\}$$

- The semidefinite cone

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

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

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

2.2 The convex hull

Definition 8. 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:*

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

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

Compare these two definitions: The original definition of a convex hull is an “outer description”. It defines the convex hull as the intersection of all possible convex sets containing the original set S. The equivalent definition given by the theorem is an “inner description”. It defines the convex hull as the union of all the points that can be produced via convex combinations.

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 $\lambda_1, \ldots, \lambda_k$ 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 $\mu_1, \ldots, \mu_m$ 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 $\lambda_1, \ldots, \lambda_{n+1}$ and $s_1, \ldots, s_{n+1} \in S$ 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 $\lambda_1, \ldots, \lambda_k$ and $s_1, \ldots, s_k \in S$ 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 $\theta_1, \ldots, \theta_{k-1}$ 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 9. Let $S \subseteq \mathbb{R}^n$. The affine hull of S, denoted $\operatorname{aff}(S)$ is the smallest affine set containing S. The conic hull of S, denoted $\operatorname{cone}(S)$ 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 10. Let $C \subseteq \mathbb{R}^n$.

- The *interior* of C is the set

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

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

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

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

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

What is the point of the definition for relative interior and relative boundary? At times (often) we will care more about a convex set thought of as a full-dimensional set in its affine hull instead of as a "degenerate" object in a larger ambient space. For example, consider the set $S = [0,1] \times \{0\}$. The affine hull of S is $\mathbb{R} \times \{0\}$. Then, $\operatorname{rint}(S) = (0,1) \times \{0\}$ and $\operatorname{rbd}(S) = \{(0,0),(1,0)\}$. On the other hand, $\operatorname{int}(S) = \varnothing$ and $\operatorname{bd}(S) = S$.

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

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

Proof sketch. Assume that $y \in C$. The case $y \in \operatorname{cl}(C)$ is similar and requires just one extra limiting argument. [2]

[2] **Exercise**: complete the proof.

As $x \in \operatorname{rint}(C)$, there exists an $\epsilon > 0$ so that $\mathbb{B}(x,\epsilon) \cap \operatorname{aff}(C) \subseteq C$. That is for all $\delta \in \mathbb{B}(0,\epsilon) \operatorname{aff}(C)$, $x + \delta \in C$. As C is convex, we have that

$$(1-\theta)(x+\delta) + \theta y \in C.$$

Thus, $\mathbb{B}((1-\theta)x + \theta y, (1-\theta)\epsilon) \operatorname{aff}(C) \subseteq C$ and $(1-\theta)x + \theta y \in \operatorname{rint}(C)$. ∎

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

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

3
Elementary convex analysis II

3.1 Convex functions

Definition 11. 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.

  Proof. Given $x, y \in V$ and $\theta \in [0, 1]$, $\|\theta x + (1-\theta)y\| \leq \|\theta x\| + \|(1-\theta)y\| = \theta \|x\| + (1-\theta) \|y\|$.

- Any squared-norm is a convex function.

  Proof. Let $x, y \in V$ and $\theta \in [0, 1]$. Then, $\|(1-\theta)x + \theta y\|^2 \leq ((1-\theta) \|x\| + \theta \|y\|)^2 \leq (1-\theta) \|x\|^2 + \theta \|y\|^2$. Here, the first inequality follows from convexity of a norm and the fact that $(\cdot)^2$ is an increasing function on $\mathbb{R}_+$. The second inequality follows from convexity of $(\cdot)^2$. □

Lemma 7. *The following functions are convex:*

- *If f is convex and $\alpha \geq 0$, then αf is convex*
- *If f, g are convex, then $f + g$ is convex*
- *If f is convex and $Ay + b$ is affine, then $y \mapsto f(Ay + b)$ is convex.*
- *Suppose f_α is convex for all $\alpha \in A$ and $\sup_{\alpha \in A} f_\alpha(x) < \infty$ for all x. Then, $g(x) := \sup_{\alpha \in A} f_\alpha(x)$ is convex.*[1]

[1] It is possible to get rid of the finiteness assumption here, but we would need to go back and define convex *extended-valued* functions. This is easy to do if you need it.

- *Suppose $f(x, y)$ is jointly convex in (x, y) and assume that for $\inf_y f(x, y) > -\infty$ for all x. Then, $g(x) := \inf_y f(x, y)$ is convex.*[2]

[2] If $\inf_y f(x, y) = -\infty$ for some x, then one can show that $g(x) = -\infty$ for all x so that it is vacuously "convex."

Proof. Most are straightforward. We prove the last two:

Let $x, y \in V$ and $\theta \in [0, 1]$. Let $\epsilon > 0$ and let $\alpha \in A$ so that

$$g((1-\theta)x + \theta y) - \epsilon \leq f_\alpha((1-\theta)x + \theta y).$$

Then, by convexity and definition of g, we have that

$$\begin{aligned} g((1-\theta)x + \theta y) - \epsilon &\leq (1-\theta)f_\alpha(x) + \theta f_\alpha(y) \\ &\leq (1-\theta)g(x) + \theta g(y). \end{aligned}$$

Letting $\epsilon \to 0$ completes the proof.

Suppose $x, z \in V$ and $\theta \in [0, 1]$. Let $\epsilon > 0$ and let y_x and y_z so that

$$\begin{aligned} g(x) &\geq f(x, y_x) - \epsilon \\ g(z) &\geq f(z, y_z) - \epsilon. \end{aligned}$$

Then,

$$\begin{aligned} g(\theta x + (1-\theta)z) &\leq f(\theta x + (1-\theta)z, \theta y_x + (1-\theta)y_z) \\ &\leq \theta f(x, y_x) + (1-\theta)f(z, y_z) \\ &\leq \theta g(x) + (1-\theta)g(z) + \epsilon. \end{aligned}$$

Letting $\epsilon \to 0$ completes the proof. ■

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. □

3.2 Separation of convex sets

All norms $\|\cdot\|$ in this section are the ℓ_2 norm (or the induced norm in a general Euclidean space).

Definition 12. 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

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

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*

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

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

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

The derivative of this expression at $\alpha = 0$ must be nonnegative by definition of the projection.

($\Leftarrow$). 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 $y \in C$,

$$\begin{aligned}
\|x-y\|^2 &= \|x-\bar{y}\|^2 + 2\langle x-\bar{y}, \bar{y}-y\rangle + \|\bar{y}-y\|^2 \\
&\geq \|x-\bar{y}\|^2 + \|\bar{y}-y\|^2
\end{aligned}$$

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

Definition 13. 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 14. 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 $D \subseteq H^{\geq}_{a,\beta}$. □

Theorem 6 (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 $c \in C$:

$$\begin{aligned} \langle s, c \rangle &= \left\langle \bar{d} - \bar{c}, c - \bar{c} + \bar{c} \right\rangle \\ &\leq \left\langle \bar{d} - \bar{c}, \bar{c} \right\rangle =: \alpha \end{aligned}$$

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

$$\begin{aligned} \langle s, d \rangle &= \left\langle \bar{d} - \bar{c}, d - \bar{d} + \bar{d} \right\rangle \\ &\geq \left\langle \bar{d} - \bar{c}, \bar{d} \right\rangle =: \beta \end{aligned}$$

so that $D \subseteq H^{\geq}_{s,\beta}$. Note that $\beta - \alpha = \left\|\bar{d} - \bar{c}\right\|^2 > 0$ so that $\beta > \alpha$. ■

Theorem 7. *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 $K := C - D$. 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.

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

Else, suppose $0 \in \operatorname{cl}(K)$. By Corollary 1 (i.e., that the relative interior of a convex set is dense in its closure), there exists $x_i \in \operatorname{rint}(K)$ 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

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

Now, as the unit sphere is compact, we may assume that v_i converges to some nonzero w (else pass to a subsequence). We claim that

$$\inf_{x \in K} \langle w, x \rangle \geq 0.$$

To see this, suppose otherwise and let $\bar{x} \in K$ so that $\langle w, \bar{x} \rangle = -\epsilon < 0$. Now, for all i large enough,

$$- \|x_i\| \leq \inf_{x \in K} \langle v_i, x \rangle \leq \langle v_i, \bar{x} \rangle \leq -\epsilon/2.$$

This contradicts the assumption that $x_i \to 0$. ■

3.3 Basic definitions about general convex programs

Definition 15. 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,

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

Definition 16. Given a feasible point $x^\star \in \Omega$, the descent cone at $x^\star$ is

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

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

Exercise

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

Problems

1. 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 $x^\star$.

We say that a vector is k-sparse if it has at most k nonzero entries. 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. For a vector $x \in \mathbb{R}^n$, let $\operatorname{sign}(x)$ denote the $\{-1, 0, 1\}$-valued vector giving the individual signs of the coordinates of x.

(a) The *descent cone* of a convex-optimization problem at a feasible solution $\bar{x}$ is defined as

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

Show that for this problem, the descent cone at the optimal solution $x^\star$ is

$$\left\{ \delta \in \mathbb{R}^n : \begin{array}{l} \delta \in \ker(A) \\ \langle \operatorname{sign}(x^\star)_{S^\star}, \delta_{S^\star}\rangle + \left\|\delta_{(S^\star)^C}\right\|_1 \leq 0 \end{array} \right\}$$

where $S^\star$ is the support of $x^\star$.

(b) The matrix A is said to satisfy the *nullspace property at order k* if for all sets $S \subseteq [n]$ with $|S| \leq k$ 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, i.e., equal to $\{0\}$, if A satisfies the nullspace property at order k and $x^\star$ is k-sparse.

(c) Show that if A does not satisfy the nullspace property at order k, then there exists a k-sparse $x^\star$ for which the convex-optimization approach may fail to recover $x^\star$. That is, for which the descent cone at $x^\star$ is nontrivial.

2. 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 Dual cones

Definition 17. Let K be a cone, the *dual cone* K_* is

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

In other words, it is the set of linear functions that are nonnegative on K. □

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 $K_p = \left\{(x, t) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_p \leq t\right\}$. Then $(K_p)_* = K_q$ where q is the Hölder dual of p. □

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, Theorem 6 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 18. 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. □

4.2 What is a conic program?

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

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

The constraint $Ax \geq a$ can be rewritten $Ax - a \geq 0$ or $Ax - a \in \mathbb{R}^m_+$. Central to the definition of a linear program is the cone $\mathbb{R}^m_+$ that gives us a partial 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 partial orderings on vectors.

Definition 19. A *conic program* in standard form is an optimization problem of the form

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

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

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

Example 10. We will not always work with conic programs *in standard form.* For example, the following problem

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

is a conic program. Here, $c \in \mathbb{R}^n$, $\mu_i \in \mathbb{R}^n$, and $r_i \in \mathbb{R}$. The theory that we develop for conic programs in standard form apply to this program if we write it in standard form as:

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

Putting a conic program into standard form can be tedious so it will be useful to pay attention to how the theory we develop applies to conic programs *not* in standard form. □

4.3 Weak Conic Duality

Consider a standard conic program

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

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:

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

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

Theorem 8 (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

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

4.4 Cones and inequalities

In order to prove strong duality, we will need to impose further assumptions on the cone K. As we will see, this is equivalent to imposing additional assumptions on the partial ordering.

Definition 20. Given a set $K \subseteq \mathbb{R}^n$, define the binary relation $\succeq_K$ where

$$a \succeq_K b \iff a - b \in K. \qquad \square$$

Definition 21. Given a binary relation $\succeq$ on $\mathbb{R}^n$, define the set

$$K_\succeq := \{a - b : a \succeq b\}. \qquad \square$$

Definition 22. A "proper" binary relation on $\mathbb{R}^n$ satisfies:

- (Reflexive) For any $x \in \mathbb{R}^n$, we have $x \succeq x$
- (Antisymmetric) If $x \succeq y$ and $y \succeq x$, then $x = y$
- (Transitive) If $x \succeq y$ and $y \succeq z$, then $x \succeq z$.
- (Additive) If $a \succeq b$ and $c \succeq d$, then $a + c \succeq b + d$
- (Positively homogeneous) If $a \succeq b$ and $\lambda \in \mathbb{R}_+$, then $\lambda a \succeq \lambda b$
- (Stable w.r.t. limits) If $a_i \to a$ and $a_i \succeq 0$ for all i, then $a \succeq 0$
- ("Existence of a strict relation") There exists a so that for all b there exists $\lambda \in \mathbb{R}_+$ so that $\lambda a \succeq b$. $\square$

Definition 23. A cone $K \subseteq \mathbb{R}^n$ is proper if it is pointed, closed, and has nonempty interior. $\square$

Lemma 10. *If $\succeq$ is a proper binary relation, then $K_\succeq$ is a proper cone. Conversely, if K is a proper cone, then $\succeq_K$ is a proper binary relation.*

Proof. First, suppose $\succeq$ is a proper binary relation. We will check that $K_\succeq$ is a proper cone. The proof that $K_\succeq$ is a pointed cone is straightforward.[2] We check that $K_\succeq$ is closed: Suppose $a_i \in K_\succeq$ converge to a. By definition, $a_i \succeq 0$ for each i. By stability w.r.t. limits, $a \succeq 0$ and $a \in K_\succeq$. Next, we check that $K_\succeq$ has a nonempty interior. By "existence of a strict relation", there exists an a such that for all b, there exists $\lambda \in \mathbb{R}_+$ so that $\lambda a \succeq b$. We claim that $a \in$

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

$\text{int}(K_{\succeq})$. First, taking $b = -a$, and using additivity and homogeneity, we have that $a \succeq 0$. Thus, for all b, there exists $\lambda \in \mathbb{R}_{++}$ so that $\lambda a \succeq b$. Once more by additivity and homogeneity, taking $b = e_i$, there exists $\delta_i > 0$ so that $a + \delta_i e_i \succeq 0$. We deduce that $\{a + \delta_i e_i\}_i \subseteq K_{\succeq}$. As $K_{\succeq}$ is convex, we deduce that $a \in \text{int}(K_{\succeq})$.

In the other direction, suppose K is a proper cone. The reflexivity, antisymmetry, transitivity, additiviy, positive homogeneity, and stability w.r.t. limits are easy to check.[3] We check the "existence of a strict relation". Let $a \in \text{int}(K)$ and let b be arbitrary. Then, for all $\delta > 0$ small enough, $a + \delta b \in K$. Thus, $\frac{1}{\delta} a + b \in K$ so that $\frac{1}{\delta} a \succeq b$. ■

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

Definition 24. Given a cone K, we will denote by $a \succ_K 0$ the fact that $a \in \text{int}(K)$. This is equivalent to saying that for all b, there exists $\lambda \geq 0$ so that

$$\lambda a \succeq_K b.$$ □

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*. □

Problems

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

2. This problem derives a dual description of the *Wasserstein distance* for discrete probability distributions.

   Fix a discrete metric space $\mathcal{X} = \{x_1, \ldots, x_n\}$.[4] Let $D \in \mathbb{R}^{n \times n}$ denote the matrix where $D_{i,j}$ is the distance between x_i and x_j.

   Let P be a probability distribution on $\mathcal{X}$ defined by $P = (p_1, \ldots, p_n)$. Similarly, let $Q = (q_1, \ldots, q_n)$ be a probability distribution on $\mathcal{X}$.

   The *Wasserstein distance* between P and Q is defined as follows: We can think of P as placing some amount of "earth/dirt" at each of the n points in $\mathcal{X}$. We want to move this earth as efficiently as possible to transform P into Q. That is, we require a transportation schedule, called a *coupling*, that tells us how much earth to move from x_i to x_j. Formally, the matrix $\Gamma \in \mathbb{R}^{n \times n}$ is a coupling if

$$\begin{aligned} \sum_{j=1}^{n} \Gamma_{i,j} &= p_i, & \forall i \in [n] \\ \sum_{i=1}^{n} \Gamma_{i,j} &= q_j, & \forall j \in [n] \\ \Gamma_{i,j} &\geq 0, & \forall i, j. \end{aligned}$$

   The cost of a coupling is given by $\langle \Gamma, D \rangle$, i.e., it is the linear cost function where moving one unit of mass from x_i to x_j costs $D_{i,j}$.

   - Write the Wasserstein distance as the optimum value of a minimization LP. We will refer to this as the primal LP.
   - Derive the dual of this LP.
   - Explain what complementary slackness means for this primal-dual pair.

[4] For concreteness, you could think of this as n points in $\mathbb{R}^d$ for example.

5
Conic programming II

5.1 Strong Conic Duality

Definition 25. We say that (Primal) is strictly feasible if there exists $\bar{x} \in \mathbb{R}^n$ so that

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

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

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

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

Theorem 9 (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

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

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 $\geq \mathrm{Opt}(\text{Primal})$.

- Define

$$\mathcal{I} := \left\{ \begin{pmatrix} \langle c, x\rangle \\ Ax - a \end{pmatrix} : Bx - b = 0 \right\}$$

$$\mathcal{S} := \left\{ \begin{pmatrix} \lambda \\ \zeta \end{pmatrix} : \begin{array}{l} \lambda < \mathrm{Opt}(\text{Primal}) \\ \zeta \in K \end{array} \right\}$$

  Note that $\mathcal{I}$ is an affine subspace and $\mathcal{S}$ is a convex set. Furthermore, $\mathcal{I}$ and $\mathcal{S}$ are disjoint: Otherwise, there exists some $x \in \mathbb{R}^n$ for which $\langle c, x\rangle < \mathrm{Opt}(\text{Primal})$ and $Ax - a \in K$.

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

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

  Notice that $\bar{t} \geq 0$ and $\bar{y} \in K_*$: Indeed, if $\bar{t} < 0$, then we may approach $(-\infty, 0) \in \mathcal{S}$ 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 $(\mathrm{Opt}(\mathrm{Primal}) - 1, \infty\zeta) \in \mathcal{S}$ to set the RHS arbitrarily positive, a contradiction.

We must also have that $\bar{t} \neq 0$. Indeed, suppose $\bar{t} = 0$ and recall that by strict feasibility,[1] there exists $\bar{x}$ so that $B\bar{x} - b = 0$ and $A\bar{x} - a \in \mathrm{int}(K)$. As $(\bar{t}, \bar{y})$ is nonzero, we must have that $\bar{y} \in K_*$ is nonzero. Thus, there exists $(\lambda, \zeta) \in \mathcal{I}$ achieving

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

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

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

a contradiction.

As $t > 0$, we can normalize $\bar{t} = 1$ in the separation statement (i.e., replace $\bar{y} \leftarrow \bar{y}/\bar{t}$).

- We now rewrite the separation statement:

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

The RHS is equal to Opt(Primal). Thus, for all $x \in \mathbb{R}^n$ satisfying $Bx - b = 0$, we have that

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

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

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

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

$$\begin{aligned}\langle a, \bar{y} \rangle + \langle b, \bar{z} \rangle &= \langle a, \bar{y} \rangle + \langle B\bar{x}, \bar{z} \rangle \\ &= \langle a, \bar{y} \rangle + \langle \bar{x}, B^\intercal \bar{z} \rangle = \langle a, \bar{y} \rangle + \langle c - A^\intercal \bar{y}, \bar{x} \rangle \\ &\geq \mathrm{Opt}(\mathrm{Primal}). \qquad \blacksquare\end{aligned}$$

Remark 3. The second step in the proof of strong duality constructed a separating hyperplane between $\mathcal{I}$ and $\mathcal{S}$ of the form

$$\inf_{(\lambda,\zeta)\in\mathcal{I}} \left\langle \begin{pmatrix} 1 \\ -\bar{y} \end{pmatrix}, \begin{pmatrix} \lambda \\ \zeta \end{pmatrix} \right\rangle \geq \sup_{(\lambda,\zeta)\in\mathcal{S}} \left\langle \begin{pmatrix} 1 \\ -\bar{y} \end{pmatrix}, \begin{pmatrix} \lambda \\ \zeta \end{pmatrix} \right\rangle.$$

This is the "core difficulty" in proving conic strong duality. You can drop almost all assumptions[2] in the theorem statement if you can construct this separating hyperplane using other methods (for example, in the LP setting).

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

[2] The only remaining assumptions will be feasibility of the primal and dual linear systems and that $(K_*)_* = K$

It is natural that this is the "core difficulty" as it is in fact *equivalent* to solving the dual with value Opt(Primal). Specifically, let us define

$$\mathcal{S}_t := \left\{ \begin{pmatrix} \lambda \\ \zeta \end{pmatrix} : \begin{array}{l} \lambda < t \\ \zeta \in K \end{array} \right\}.$$

With this notation, our previous $\mathcal{S}$ can be written as $\mathcal{S}_{\text{Opt(Primal)}}$.

Now suppose $(\bar{y}, \bar{z})$ is feasible in the dual with value t. Then,

$$\begin{aligned} \inf_{(\lambda,\zeta)\in\mathcal{I}} \lambda - \langle \bar{y}, \zeta \rangle &= \inf_{x\in\mathbb{R}^n} \left\{ \langle c, x \rangle - \langle \bar{y}, A\bar{x} - a \rangle : \quad Bx - b = 0 \right\} \\ &= \inf_{x\in\mathbb{R}^n} \left\{ \langle a, \bar{y} \rangle + \langle B^\mathsf{T} \bar{z}, x \rangle : \quad Bx - b = 0 \right\} \\ &= t. \end{aligned}$$

On the other hand,

$$\sup_{(\lambda,\zeta)\in\mathcal{S}_t} \lambda - \langle \bar{y}, \zeta \rangle = t.$$

We conclude that if $(\bar{y}, \bar{z})$ is feasible with value t, then

$$\inf_{(\lambda,\zeta)\in\mathcal{I}} \lambda - \langle \bar{y}, \zeta \rangle \geq \sup_{(\lambda,\zeta)\in\mathcal{S}_t} \lambda - \langle \bar{y}, \zeta \rangle .$$

Now, suppose $\bar{y}$ and $\bar{t}$ satisfy:

$$\inf_{(\lambda,\zeta)\in\mathcal{I}} \lambda - \langle \bar{y}, \zeta \rangle \geq \sup_{(\lambda,\zeta)\in\mathcal{S}_t} \lambda - \langle \bar{y}, \zeta \rangle .$$

Then, we must have that $\bar{y} \in K_*$ and the value of the RHS is equal to t. Thus, we deduce that

$$\inf_{x\in\mathbb{R}^n} \{ c^\mathsf{T} x - \langle \bar{y}, Ax - a \rangle : Bx - b = 0 \} \geq t.$$

This is a *bounded* affine function on an affine space. Thus, it must be constant on the affine space $Bx = b$. In other words, there exists $\bar{z}$ so that

$$c - A^\mathsf{T} \bar{y} = B^\mathsf{T} \bar{z} \qquad \text{and} \qquad \langle a, \bar{y} \rangle + \langle b, \bar{z} \rangle \geq t.$$

From all of this, we conclude that

$$\begin{aligned} &\sup_{y,z} \left\{ \langle a, \bar{y} \rangle + \langle b, \bar{z} \rangle : \begin{array}{l} y \in K_* \\ A^\mathsf{T} y + B^\mathsf{T} z - c = 0 \end{array} \right\} \\ &= \sup_{t,y} \left\{ t : \inf_{(\lambda,\zeta)\in\mathcal{I}} \lambda - \langle y, \zeta \rangle \geq \sup_{(\lambda,\zeta)\in\mathcal{S}_t} \lambda - \langle \bar{y}, \zeta \rangle \right\} \end{aligned}$$

and that solvability of either implies solvability of the other. In words, constructing a dual feasible solution with a given value t is equivalent to constructing a separating hyperplane between $\mathcal{I}$ and $\mathcal{S}_t$ of a particular form. □

Remark 4. What can go wrong without primal strict feasibility?

Without primal strict feasibility, we can be in one of the following situations:

- Strong duality and dual solvability still holds (e.g, in the case of LPs).
- Strong duality still holds, but the dual is not solvable: For example, consider the following primal and dual conic programs

$$\begin{aligned}
&\inf_{(\lambda,t)\in\mathbb{R}^2}\left\{t : \begin{array}{c}\lambda = 1\\ \begin{pmatrix}\lambda\\ \lambda\\ t\end{pmatrix}\in\mathcal{L}^{1+2}\end{array}\right\}\\
&\geq \sup_{\alpha\in\mathbb{R},\,(\beta,\gamma)\in\mathbb{R}^{1+2}}\left\{\alpha : \begin{array}{l}\alpha+\beta+\gamma_1 = 0\\ \gamma_2 = 1\\ (\beta,\gamma)\in\mathcal{L}^{1+2}\end{array}\right\}\\
&= \sup_{\alpha,\,\beta\in\mathbb{R}}\left\{\alpha : \begin{pmatrix}\beta\\ -\alpha-\beta\\ 1\end{pmatrix}\in\mathcal{L}^{1+2}\right\}\\
&= \sup_{\beta\in\mathbb{R}}\left\{2(\sqrt{\beta^2-1}-\beta) : \beta\geq 1\right\}.
\end{aligned}$$

  We see that both the primal and the dual have the optimum value 0 and the dual is *not* solvable.

- Strong duality fails, either the primal or dual is feasible with bounded objective value and the other is infeasible. See **Problem 1**.
- Strong duality fails, both primal and dual have bounded objective value and are feasible, but there is a positive duality gap: For example, consider the following primal and dual conic programs

$$\begin{aligned}
&\inf_{x\in\mathbb{R}^2}\left\{x_2 : \begin{pmatrix}1-x_1 & 0 & x_2\\ 0 & 1+x_2 & 0\\ x_2 & 0 & 0\end{pmatrix}\succeq 0\right\}\\
&\geq \sup_{Y\in\mathbb{S}^3}\left\{-Y_{1,1}-Y_{2,2} : \begin{array}{l}Y_{1,1}=0\\ Y_{1,3}+Y_{2,2}+Y_{3,1}=1\\ Y\succeq 0\end{array}\right\}.
\end{aligned}$$

  The primal has value 0 and the dual has value -1.

In all of these cases, the proof of strong duality that we did breaks. Specifically, the separating hyperplane that we would find via the hyperplane separation theorem would fall entirely in the subspace corresponding to the "conic directions." □

Corollary 2. *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:* $\langle \bar{y}, A\bar{x} - a\rangle = 0$

Proof. Follows from

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

■

We will not prove this theorem, but it is useful to know and compare.

Theorem 10. *In a Linear Program, if both primal and dual are feasible, then strong duality holds and both are solvable.*

Example 12. Consider the following problem: Given $\mu_1, \ldots, \mu_k \in \mathbb{R}^n$, solve

$$\inf_{x\in\mathbb{R}^n} \sum_{i=1}^{k} \|x - \mu_i\|_2 .$$

Our goal is to recognize this as a conic program, construct its dual, deduce that strong duality holds, that both programs are solvable, and to understand what complementary slackness says about the structure of primal and dual optima.

Recall that

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

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 $(\xi_i, \zeta_i) \in \mathcal{L}_*^{1+n} = \mathcal{L}^{1+n}$. We now derive the dual problem. First, we collect all the inequalities and then rearrange to derive a lower bound on some linear form evaluated at our primal variables x, t. Let $(\xi_i, \zeta_i) \in \mathcal{L}^{1+n}$. Then,

$$\begin{aligned} &\sum_{i=1}^{k} (\xi_i t_i + \langle \zeta_i, x - \mu_i\rangle) \geq 0 \\ \Longleftrightarrow \qquad &\sum_i \xi_i t_i + \left\langle \sum_i \zeta_i, x \right\rangle - \sum_i \langle \zeta_i, \mu_i\rangle \geq 0 \\ \Longleftrightarrow \qquad &\sum_i \xi_i t_i + \left\langle \sum_i \zeta_i, x \right\rangle \geq \sum_i \langle \zeta_i, \mu_i\rangle . \end{aligned}$$

Thus, the dual is given by

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

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

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

What does complementary slackness mean? It means that the conic inequality is tight at the optimal solution, i.e., let $(\bar{t},\bar{x})$, $(\bar{\xi}_i,\bar{\zeta}_i)$ be primal and dual optimal solutions. Then, by complementary slackness, for all $i = 1,\dots,k$,

$$\bar{\xi}_i \bar{t}_i + \left\langle \bar{\zeta}_i, \bar{x} - \mu_i \right\rangle = 0.$$

Note that $\bar{\xi}_i = 1$ in any dual feasible solution and $\bar{t}_i = \|\bar{x} - \mu_i\|$ in any primal optimal solution. Thus, complementary slackness tells us that

$$\|x - \mu_i\| = \left\langle \bar{\zeta}_i, \bar{x} - \mu_i \right\rangle.$$

In other words (as long as $\bar{x} \neq \mu_i$), we have that $\bar{\zeta}_i$ must be the unit vector in the direction $\bar{x} - \mu_i$. $\square$

The following lemma explains what complementary slackness means for $\mathbb{R}^n_+$, $\mathcal{L}^{1+n}$, and $\mathbb{S}^n_+$.

Lemma 11. *Let $n \geq 1$.*

- *If $x, y \in \mathbb{R}^n_+$ are such that $\langle x, y \rangle = 0$, then the support of x and y are disjoint.*
- *If $(s,x),(t,y) \in \mathcal{L}^{1+n}$ are such that*

$$\left\langle \begin{pmatrix} s \\ x \end{pmatrix}, \begin{pmatrix} t \\ y \end{pmatrix} \right\rangle = 0,$$

  then $\langle x, y \rangle = -st$. More cogently, if both (s,x) and (t,y) are nonzero, then $s = \|x\|$, $t = \|y\|$, and x and y are collinear pointed in opposite directions.
- *If $X, Y \in \mathbb{S}^n_+$ are such that $\langle X, Y \rangle = 0$, then* $\operatorname{range}(X) \subseteq \ker(Y)$ *and* $\operatorname{range}(Y) \subseteq \ker(X)$.

Proof. First, suppose $x, y \in \mathbb{R}^n_+$. Then, $0 = \langle x, y \rangle = \sum_i x_i y_i$. Thus, for each $i \in [n]$, at least one of x_i or y_i must be zero.

Next, suppose $(s,x),(t,y) \in \mathcal{L}^{1+n}$. Then,

$$0 = st + \langle x, y \rangle \geq 0.$$

We deduce that $\langle x, y\rangle = -st$. If (s, x) and (t, y) are both nonzero, then st is nonzero and $\langle x, y\rangle \geq -\|x\|\,\|y\| \geq -st$ where equality holds throughout the chain only if $\|x\| = s$, $\|y\| = t$, and x, y are collinear in opposite directions.

Next, suppose $X, Y \in \mathbb{S}^n_+$ are such that $\langle X, Y\rangle = 0$. We will show that $\operatorname{range}(X) \subseteq \ker(Y)$. The second statement follows by symmetry. By the spectral theorem, we can write $X = \sum_{i=1}^n \lambda_i v_i v_i^\intercal$. Suppose $\lambda_i > 0$ for $i \in [k]$ and $\lambda_i = 0$ for $i \in [k+1, n]$. Then, $\operatorname{range}(X)$ is given by

$$\begin{aligned}
\operatorname{range}(X) &= \{Xu :\ u \in \mathbb{R}^n\} \\
&= \left\{\sum_{i=1}^n \lambda_i v_i v_i^\intercal u :\ u \in \mathbb{R}^n\right\} \\
&= \left\{\sum_{i=1}^k \lambda_i v_i v_i^\intercal u :\ u \in \mathbb{R}^n\right\} \\
&= \operatorname{span}\{v_1, \dots, v_k\}.
\end{aligned}$$

On the other hand,

$$0 = \langle X, Y\rangle = \sum_{i=1}^n \lambda_i v_i^\intercal Y v_i = \sum_{i=1}^k \lambda_i v_i^\intercal Y v_i.$$

Note that for all $i = 1, \dots, k$, we have $\lambda_i > 0$ and $v_i^\intercal Y v_i \geq 0$. Thus, $v_i^\intercal Y v_i = 0$ so that $v_i \in \ker(Y)$. ■

Remark 5. 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

$$\inf_{x\in\mathbb{R}^n} \{c^\intercal x :\ Ax - a \in K\}.$$

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. 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.

2. Consider an optimization problem of the form

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

   We make no assumptions on whether f or $g_1, \ldots, g_m$ 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 $g_1, \ldots, g_m$ are convex functions, then $\mathcal{I}$ is a convex set.
   - We now will only assume that $\mathcal{I}$ is a convex set (while $f, g_1, \ldots, g_m$ may not necessarily be convex).

     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

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

     where the dual problem is solvable (i.e., the supremum is achieved).

   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.

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

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

Furthermore, assume that the primal problem is feasible and that:

$$\ker\left(\begin{pmatrix} c^\intercal \\ A \\ B \end{pmatrix}\right) = \{0\}\,.$$

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.

Hint: in the only if direction, consider the set

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

You must justify why this set needs to be $\{0\}$. Now, take the dual cone of either side of this equation. You may use the fact that the relative interior of an affine image of a convex set is the affine image of the relative interior of the convex set.

6
SOCP representability

The following two lectures will investigate in detail 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 26. 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:

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

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

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

Definition 27. We say that $X \subseteq \mathbb{R}^n$ is a second-order cone representable (SOCR) set if there exists a set

$$S = \left\{ (x,u) \in \mathbb{R}^n \times \mathbb{R}^{n'} : A(x,u) - b \in \mathcal{K} \right\}$$

such that $X = \Pi_x S$ where $\Pi_x(x,u) := x$ and $\mathcal{K}$ is a product of second-order cones.

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

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

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 $\mathcal{X}_1, \dots, \mathcal{X}_m \subseteq \mathbb{R}^n$ are SOCR set. Then,

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

can be converted into an SOCP.[1] We will slightly abuse terminology and even refer to this problem as an SOCP (albeit one that is not in standard form).

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

Example 14. • $f(x) = \|x\|$ is SOCR:

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

- $f(x) = \|x\|^2$ is SOCR:

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

The "trick" here[2] is that we can get a linear form as a difference of quadratic functions $(t+1)^2 - (t-1)^2 = 4t$. □

[2] **Hint**: this may be useful in the exercises.

Lemma 12. *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. The first four are left as an exercise.[3]

[3] **Exercise**: Verify.

We prove only the last statement: Suppose $n_1 = \cdots = n_k$ so that $\sum_i X_i$ is defined. By assumption, each X_i is SOCR so that

$$X_i = \Pi_{x_i} \left\{ (x_i, u_i) : A_{i,j}(x_i, u_i) - b_{i,j} \in \mathcal{L}^{1+n_{i,j}},\ \forall i \in [m_i] \right\}.$$

Then,

$$\sum_i X_i = \Pi_\xi \left\{ (\xi, x_i, u_i) : \begin{array}{l} \xi = \sum_i x_i \\ A_{i,j}(x_i, u_i) - b_{i,j} \in \mathcal{L}^{1+n_{i,j}},\ \forall i, j \end{array} \right\}. \quad \blacksquare$$

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

6.2 *Rational convex powers and ℓ_p norms are SOCR*

In this section, we will prove that the following two commonly occurring functions are SOCR:

Lemma 13. *The hypograph of the geometric mean of two nonnegative variables:*

$$\left\{ \begin{pmatrix} x \\ y \\ t \end{pmatrix} \in \mathbb{R}^3 : \begin{array}{l} x, y \geq 0 \\ t \leq \sqrt{xy} \end{array} \right\}$$

is SOCR.

Proof. We will give a SOC representation of this set. Suppose $x, y \geq 0$. We would like to square both sides of the inequality $t \leq \sqrt{xy}$ to remove the square-root. However, we may not be allowed to do this if t is negative. Thus, we will introduce a variable $u \geq 0$ and say

$$\begin{aligned} t \leq \sqrt{xy} &\iff \exists u \geq 0,\, t \leq u \leq \sqrt{xy} \\ &\iff \exists u,\, t \leq u,\, 0 \leq u,\, u^2 \leq xy \\ &\iff \exists u,\, t \leq u,\, 0 \leq u,\, (2u)^2 + (x-y)^2 \leq (x+y)^2 \\ &\iff \exists u,\, t \leq u,\, 0 \leq u \\ &\qquad \begin{pmatrix} x+y \\ x-y \\ 2u \end{pmatrix} \in \mathcal{L}^{1+2}. \end{aligned}$$

Thus, we may write

$$\left\{\begin{pmatrix} x \\ y \\ t \end{pmatrix} \in \mathbb{R}^3 : \begin{array}{l} x, y \geq 0 \\ t \leq \sqrt{xy} \end{array}\right\}$$

$$= \Pi_{x,y,t}\left\{(x, y, t, u) : \begin{array}{l} x, y, u \geq 0 \\ t \leq u \\ \begin{pmatrix} x+y \\ x-y \\ 2u \end{pmatrix} \in \mathcal{L}^{1+2} \end{array}\right\}. \qquad \blacksquare$$

Lemma 14. *The hypograph of the geometric mean of 2^ℓ nonnegative variables,*

$$\left\{(x, t) \in \mathbb{R}^{2^\ell} \times \mathbb{R} : \begin{array}{l} x \geq 0 \\ t \leq (\prod_i x_i)^{1/2^\ell} \end{array}\right\},$$

is SOCR.

Proof. We show this inductively. The case $\ell = 1$ is the previous example. Now let $\ell \geq 1$ and suppose the claim holds inductively up to ℓ. Now, suppose $x \in \mathbb{R}^{2^\ell+1}$ and $x \geq 0$. Then,

$$t \leq \left(\prod_i x_i\right)^{1/2^{\ell+1}} \iff t \leq \sqrt{\left(\prod_{i=1}^{2^\ell} x_i\right)^{1/2^\ell} \left(\prod_{i=2^\ell+1}^{2^{\ell+1}} x_i\right)^{1/2^\ell}}$$

$$\iff \begin{cases} \exists\, u_{\text{left}},\, u_{\text{right}} \geq 0 : \\ u_{\text{left}} \leq \left(\prod_{i=1}^{2^\ell} x_i\right)^{1/2^\ell} \\ u_{\text{right}} \leq \left(\prod_{i=2^\ell+1}^{2^{\ell+1}} x_i\right)^{1/2^\ell} \\ t \leq \sqrt{u_{\text{left}} u_{\text{right}}} \end{cases}.$$

We deduce that

$$\left\{(x, t) \in \mathbb{R}^{2^{\ell+1}\times\mathbb{R}} : \begin{array}{l} x \geq 0,\ \forall i \\ t \leq (\prod_i x_i)^{1/2^{\ell+1}} \end{array}\right\}$$

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

Proposition 1. *The epigraph of a convex power of a nonnegative variable, i.e.,*

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

where $p, q \in \mathbb{N}$ and $p/q \geq 1$, is SOCR.

Proof. Suppose $x \geq 0$ and $t \geq 0$ and let $\ell \in \mathbb{N}$ so that $2^\ell \geq p, q$. Then we will write the condition that $x^{p/q} \leq t$ as

$$\begin{aligned} x^{p/q} \leq t &\iff x^p \leq t^q \\ &\iff x^{2^\ell} \leq x^{2^\ell - p} t^q \\ &\iff x \leq (x^{2^\ell - p} t^q)^{1/2^\ell}. \end{aligned}$$

This happens if and only if the vector

$$\left(\underbrace{x, \ldots, x}_{2^\ell - p \text{ times}}, \underbrace{t, \ldots, t}_{q \text{ times}}, \underbrace{1, \ldots, 1}_{p-q \text{ times}}, x \right)$$

is in the hypograph of the geometric mean of 2^ℓ nonnegative variables. Thus, by the preceding lemma and the fact that the affine preimage of an SOCR set is SOCR, we deduce that the epigraph of a convex power of a nonnegative variable is SOCR. ∎

Proposition 2. *Let $p, q \in \mathbb{N}$ so that $p/q \geq 1$. Then the set*

$$\left\{ (x, t) \in \mathbb{R}^{n+1} : \|x\|_{p/q} \leq t \right\}$$

is SOCR.

Proof. Our first step in constructing the SOC representation is to "get the absolute values" of each x_i. It is clear that if $t \geq 0$, then:

$$\|x\|_{p/q} \leq t \iff \begin{cases} \exists u_1, \ldots, u_n \geq 0 : \\ x_i \leq u_i, \, -x_i \leq u_i, \, \forall i \in [n] \\ \sum_i u_i^{p/q} \leq t^{p/q} \end{cases}$$

Our next step is to "linearize" both sides of the nonlinear equation. We will do this by multiplying both sides by $t^{1-p/q}$ and introducing new variables $v_1, \ldots, v_n$ for the resulting expressions in the summation

$$\cdots \iff \begin{cases} \exists u_1, \ldots, u_n, \, v_1, \ldots, v_n \geq 0 : \\ x_i \leq u_i, \, -x_i \leq u_i, \, \forall i \in [n] \\ u_i^{p/q} \leq v_i t^{p/q-1}, \, \forall i \in [n] \\ \sum_i v_i \leq t \end{cases}$$

We should be careful here to write $u_i^{p/q} \leq v_i t^{p/q-1}$ instead of $u_i^{p/q} t^{1-p/q} \leq v_i$ to handle the case $t = 0$ correctly. Finally, let ℓ so that $2^\ell \geq p$ and rewrite the remaining nonlinearities as geometric mean constraints:

$$\cdots \iff \begin{cases} \exists u_1, \ldots, u_n, \, v_1, \ldots, v_n \geq 0 : \\ x_i \leq u_i, \, -x_i \leq u_i, \, \forall i \in [n] \\ u_i \leq \left(u_i^{2^\ell - p} v_i^q t^{p-q} \right)^{1/2^\ell}, \, \forall i \in [n] \\ \sum_i v_i \leq t \end{cases}.$$ ∎

Exercises

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

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

7

SDP representability

Definition 28. A semidefinite program (SDP) is a conic program where the cone K is the PSD cone. $\square$

The primal and dual SDPs in standard form look like

$$\inf_x \left\{ \langle c, x\rangle : \begin{array}{l} \mathcal{A}(x) - A \succeq 0 \\ Bx - b = 0 \end{array} \right\}$$
$$\geq \sup_{Y,z} \left\{ \langle A, Y\rangle + \langle b, z\rangle : \begin{array}{l} \mathcal{A}^*(Y) + B^\intercal z - c = 0 \\ Y \succeq 0 \end{array} \right\}.$$

Here $\mathcal{A} : \mathbb{R}^n \to \mathbb{S}^m$ is a linear map. Explicitly, one can write $\mathcal{A}(x) = \sum_{i=1}^n x_i A^{(i)}$ for some $A^{(i)} \in \mathbb{S}^m$. The adjoint $\mathcal{A}^* : \mathbb{S}^m \to \mathbb{R}^n$ is also a linear map. Explicitly, it is given by $\mathcal{A}^*(Y) = \left(\left\langle A^{(i)}, Y\right\rangle\right)_i$.

Lemma 15. *Any SOCP can be written as an SDP.*

Proof. Consider an SOCP:

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

Define $\mathcal{M} : \mathbb{R}^{(1+n_1)+\cdots+(1+n_k)} \to \mathbb{S}^{(1+n_1)+\cdots+(1+n_k)}$ to be the linear map

$$\begin{pmatrix} t_1 \\ x_1 \\ \vdots \\ t_k \\ x_k \end{pmatrix} \mapsto \begin{pmatrix} t_1 & x_1^\intercal & & & \\ x_1 & t_1 I_{n_1} & & & \\ & & \ddots & & \\ & & & t_k & x_k^\intercal \\ & & & x_k & t_k I_{n_k} \end{pmatrix}.$$

Note that $\mathcal{M}(t_1, x_1, \dots) \succeq 0$ if and only if for all $i \in [k]$

$$\begin{aligned} & \begin{pmatrix} t_i & x_i^\intercal \\ x_i & t_i I_{n_i} \end{pmatrix} \succeq 0 \\ \iff \quad & (t_i = 0 \text{ and } x_i = 0) \quad \text{or} \quad (t_i > 0 \text{ and } t_i \geq \|x_i\|^2 / t_i) \\ \iff \quad & (t_i, x_i) \in \mathcal{L}^{1+n_i}. \end{aligned}$$

Thus, the SOCP can be rewritten as

$$\inf_{x \in \mathbb{R}^n} \left\{ c^\intercal x : \begin{array}{l} (\mathcal{M} \circ A)(x) - \mathcal{M}(a) \succeq 0 \\ Bx - b = 0 \end{array} \right\}. \qquad \blacksquare$$

Definition 29. A set $\mathcal{X} \subseteq \mathbb{R}^n$ is *semidefinite representable* (SDr) if there exists a representation

$$\mathcal{X} = \Pi_x \left\{ (x,u) \in \mathbb{R}^n \times \mathbb{R}^{n'} : \mathcal{A}(x,u) - A \succeq 0 \right\}.$$

for some $\mathcal{A}$ and A.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *semidefinite representable* (SDr) if

$$\operatorname{epi}(f) := \left\{ (x,t) \in \mathbb{R}^{n+1} : f(x) \le t \right\}$$

is SDr. $\square$

Remark 6. The set $\mathbb{R}^n$ that shows up in the definition of SDr is not inherently important. We can replace it with any other Euclidean space. For example, we can replace $\mathbb{R}^n$ with $\mathbb{S}^n$ by identifying $\mathbb{S}^n \simeq \mathbb{R}^{\binom{n+1}{2}}$. Alternatively, we can replace $\mathbb{R}^n$ with $\mathbb{R}^{n_1 \times n_2}$ by identifying $\mathbb{R}^{n_1 \times n_2} \simeq \mathbb{R}^{n_1 n_2}$. Thus, we can also define a SDr sets and SDr functions on these spaces. $\square$

The operations in Lemma 12 that preserve SOCr also preserve SDr.

Lemma 16. *Suppose $\mathcal{X}_1, \ldots \mathcal{X}_k$ are SDR sets where $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$. Then,*

- *(Direct product) $\Pi_i \mathcal{X}_i$ is SDr*
- *(Affine image) $\{Ax + b : x \in \mathcal{X}_1\}$ is SDr.*
- *(Inverse affine image) $\{y : Ay + b \in \mathcal{X}_1\}$ is SDr.*

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

- *(Intersection) $\bigcap_i \mathcal{X}_i$ is SDr.*
- *(Minkowski sum) $\sum_i \mathcal{X}_i$ is SDr.*

Example 16. Consider the following functions on $\mathbb{S}^n$:

- The maximum eigenvalue function, $f(X) := \lambda_{\max}(X)$ is SDr

$$\begin{aligned} \operatorname{epi}(f) &= \{(X,t) : \lambda_{\max}(X) \le t\} \\ &= \{(X,t) : tI - X \succeq 0\}. \end{aligned}$$

- The Schatten ∞-norm (the operator norm) is SDr:

$$\operatorname{epi}(\|\cdot\|_{\mathrm{op}}) = \{(X,t) : -tI \preceq X \preceq tI\}.$$

- The sum of the k-largest eigenvalues is SDr, i.e., $S_k(X) := \sum_{i=1}^k \lambda_i(X)$ where the eigenvalues are arranged in nonincreasing order. One can verify that

$$\begin{aligned} \operatorname{epi}(S_k) &= \left\{(X,t) : \quad \sum_{i=1}^k \lambda_i(X) \le t \right\} \\ &= \Pi_{(X,t)} \left\{ (X,t,Z,s) : \begin{array}{l} Z \succeq 0 \\ Z + sI \succeq X \\ \operatorname{tr}(Z) + sk \le t \end{array} \right\}. \end{aligned}$$

First, the $\subseteq$ direction: WLOG, we may assume that $X = \operatorname{Diag}(\lambda_1, \ldots, \lambda_n)$ by the spectral theorem. Let $s = \lambda_k$ and set $Z = \operatorname{Diag}(\lambda_1 - s, \ldots, \lambda_k - s, 0, \ldots, 0)$. Then, $Z \succeq 0$, $Z + sI \succeq X$, and

$$\operatorname{tr}(Z) + sk = \sum_{i=1}^k \lambda_i - sk + sk \le t.$$

Next, the $\supseteq$ direction: WLOG we may assume that $X = \operatorname{Diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal, and in turn that $Z = \operatorname{Diag}(z_1, \ldots, z_n)$ is diagonal. Now,

$$\sum_{i=1}^k \lambda_i \le \sum_{i=1}^k (z_i + s) \le \operatorname{tr}(Z) + sk \le t.$$

This completes the proof. We also see from this that $S_k(X)$ is a convex function in X.[1] □

[1] **Exercise**: Prove that
$$S_k(X) = \max_{Y \in \mathbb{S}^n} \left\{ \langle X, Y\rangle : \begin{array}{c} 0 \preceq Y \preceq I \\ \operatorname{tr}(Y) = k \end{array} \right\}$$
and use this fact to give an alternate proof of convexity of S_k. What do you get when you take the dual to this SDP?

Example 17. The set of positive *definite* matrices is SDr:

$$\{X \in \mathbb{S}^n : X \succ 0\} = \left\{ X \in \mathbb{S}^n : \begin{array}{c} \exists t \in \mathbb{R} \\ \begin{pmatrix} X & I_n \\ I_n & tI_n \end{pmatrix} \succeq 0 \end{array} \right\}.$$

To see this, note that by the Schur-Complement Lemma, the lifted matrix is PSD if and only if $t > 0$ and $X \succeq \frac{1}{t} I_n$. □

Example 18. The maximum singular value $f(X) = \|X\|_{\mathrm{op}} = \sigma_{\max}(X)$ defined for $X \in \mathbb{R}^{n_1 \times n_2}$ is SDr

$$\begin{aligned} \operatorname{epi}(f) &= \{(X,t) : \sigma_{\max}(X) \le t\} \\ &= \left\{ (X,t) : \begin{pmatrix} tI_{n_1} & X \\ X^\intercal & tI_{n_2} \end{pmatrix} \succeq 0 \right\}. \end{aligned}$$ □

7.1 Schatten-norms

The remainder of this lecture will prove that the Schatten-p norms are also SDr. This is very useful result.

Recall that the Schatten-p norm is defined as

$$\|X\|_p = \|(\lambda_1, \ldots, \lambda_n)\|_p .$$

We will use the Birkhoff-von Neumann Theorem (which was proved in Homework 1 Problem 5). Recall $P \in \mathbb{R}^{n\times n}$ is a permutation matrix if it is a $\{0,1\}$ matrix where every row and column has exactly one 1, and $D \in \mathbb{R}^{n\times n}$ is a doubly stochastic matrix if it is a nonnegative matrix where every row and column sums to 1.

Theorem 11 (Birkhoff-von Neumann)**.** *The convex hull of the permutation matrices in $\mathbb{R}^{n\times n}$ is the set of doubly stochastic matrices.*

Corollary 3. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex and permutation invariant, i.e., $f(Px) = f(x)$ for any permutation matrix P and any $x \in \mathbb{R}^n$. Then, for any doubly stochastic matrix D and any $x \in \mathbb{R}^n$, we have $f(Dx) \le f(x)$.*

Proof. By the lemma, we can write $D = \sum_{i=1}^k \lambda_i P_i$ where λ_i are convex combination weights. Then,

$$\begin{aligned} f(Dx) &= f(\sum_{i=1}^k \lambda_i (P_i x)) \\ &\le \sum_{i=1}^K \lambda_i f(P_i x) \\ &= f(x). \end{aligned}$$ ■

We will also need the following characterization of doubly stochastic matrices:

Lemma 17. *Let $y, x \in \mathbb{R}^n$. There exists a doubly stochastic matrix P so that $y = Px$ if and only if x and y satisfy the majorization inequalities:*

$$\begin{cases} \sum_{i=1}^n x_i = y_i \\ S_k(x) \ge S_k(y), \quad \forall k = 1, \dots, n-1 \end{cases}$$

Proof. The forward direction follows from the previous corollary with the convex permutation-invariant function S_k.

Now, suppose the majorization inequalities hold. Our goal is to show that $y = Px$ for some doubly stochastic matrix P. We will induct on the dimension n.

If $n = 1$, then $x = y$ and there is nothing to prove.

Now, suppose $n \ge 1$. Without loss of generality, we will assume that x and y are in nonincreasing order.

If there is any $i \in [n]$ for which $x_i = y_i$, then we can form $\hat{x}$ and $\hat{y}$ indexed by $[n] \setminus i$ by deleting the ith coordinate. The resulting vectors $\hat{x}$ and $\hat{y}$ still satisfy the majorization inequalities so that by induction, we can write $\hat{x} = D\hat{y}$. Break up D into blocks

$$\begin{pmatrix} y_{1:i-1} \\ y_{i+1:n} \end{pmatrix} = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \begin{pmatrix} x_{1:i-1} \\ x_{i+1:n} \end{pmatrix}$$

of the appropriate dimensions. Then,

$$\begin{pmatrix} y_{1:i-1} \\ y_i \\ y_{i+1:n} \end{pmatrix} = \begin{pmatrix} D_{11} & & D_{12} \\ & 1 & \\ D_{21} & & D_{22} \end{pmatrix} \begin{pmatrix} x_{1:i-1} \\ x_i \\ x_{i+1:n} \end{pmatrix}.$$

Now, we may assume that $x_i \neq y_i$ for any $i \in [n]$. Then, $x_1 > y_1$. Let k be the first index so that $y_k > x_k$. This must exist as the sums are equal. Thus,

$$x_1 > y_1 \geq y_k > x_k.$$

Now let D be the doubly stochastic matrix that acts as

$$Dx = \begin{pmatrix} (1-\mu)x_1 + \mu x_k \\ x_{2:k-1} \\ \mu x_1 + (1-\mu)x_k \\ x_{k+1:n} \end{pmatrix}.$$

We will increase μ from $0 \to 1$ until either $(1-\mu)x_1 + \mu x_k = y_1$ or $\mu x_1 + (1-\mu)x_k = y_k$, whichever occurs first.

The rest of the proof has two cases depending on which stopping condition was hit. The proofs are analogous so we will assume that

$$\begin{aligned} (Dx)_1 &= (1-\mu)x_1 + \mu x_k = y_1 \\ (Dx)_k &= \mu x_1 + (1-\mu)x_k \leq y_k. \end{aligned}$$

Now, let $\hat{x}$ and $\hat{y}$ denote the vectors achieved by dropping the first coordinates of Dx and y respectively. We will index $\hat{x}$ and $\hat{y}$ by $[2, n]$.

We verify that the majorization inequalities hold between $\hat{x}$ and $\hat{y}$. Note that $\hat{y}$ is still in sorted order, whereas $\hat{x}$ may no longer be in sorted order. For $t \in [2, k-1]$, we have that

$$S_{t-1}(\hat{x}) \geq \sum_{i=2}^{t} \hat{x}_i = \sum_{i=2}^{t} x_i \geq \sum_{i=2}^{t} \hat{y}_i.$$

The inequality here holds because by assumption $x_i > y_i$ for all $i \in [2, k-1]$. For $t \geq k$, we have

$$S_{t-1}(\hat{x}) \geq \sum_{i=2}^{t} \hat{x}_i = \sum_{i=1}^{t} x_i - (Dx)_1 \geq \sum_{i=1}^{t} y_i - y_1 = \sum_{i=2}^{k} \hat{y}_i.$$

Finally,

$$\sum_{i=2}^{n} \hat{x}_i = \sum_{i=1}^{n} x_i - (Dx)_1 = \sum_{i=2}^{n} y_i.$$

Thus, by induction there exists a doubly stochastic $\hat{D}$ acting on $\mathbb{R}^{[2,n]}$ so that $\hat{D}\hat{x} = \hat{y}$. Then,

$$\begin{pmatrix} 1 & \\ & D' \end{pmatrix} Dx = \begin{pmatrix} 1 & \\ & D' \end{pmatrix} \begin{pmatrix} y_1 \\ \hat{x} \end{pmatrix} = \begin{pmatrix} y_1 \\ \hat{y} \end{pmatrix} = y.$$

It remains to note that the product of doubly stochastic matrices is doubly stochastic. ■

Theorem 12. *Suppose f is any convex permutation-invariant function that is SDr. Then, $F(X) := f(\lambda(X))$ is SDr.*

Proof. Assume that

$$\text{epi}(f) = \Pi_{x,t}\left\{(x,t,u) : \mathcal{A}(x,t,u) - A \succeq 0\right\}.$$

We claim that

$$\text{epi}(F) = \Pi_{X,t}\left\{(X,t,x,u) : \begin{array}{l} x_1 \geq x_2 \geq \cdots \geq x_n \\ S_k(X) \leq \sum_{i=1}^k x_i,\ \forall k = 1,\ldots,n-1 \\ \text{tr}(X) = \sum_i x_i \\ \mathcal{A}(x,t,u) - A \succeq 0 \end{array}\right\}.$$

First, for the $\subseteq$ direction. Suppose X, t are such that $F(X) \leq t$. Then

$$f(\lambda(X)) \leq t.$$

Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of X arranged in nonincreasing order. By assumption, there exists u so that $\mathcal{A}(\lambda, t, u) - A \succeq 0$. Set $x_i = \lambda_i$. Then, it is clear that (X, t, x, u) satisfies the inequalities of the right hand set.

Now, suppose (X, t, x, u) satisfy the inequalities in the right hand set. By assumption λ is majorized by x. Thus, there exists a doubly stochastic matrix P so that

$$\lambda = Px.$$

Thus,

$$F(X) = f(\lambda) \leq f(x) \leq t.$$ ■

Corollary 4. *If $p \in [1, \infty]$ and p is rational, then the Schatten-p norm is SDr.*

7.2 *Some comments on lifting*

The ability to *project* in the definition of SOCR and SDR is very natural from the point of optimization: Additional lifting variables simply mean additional decision variables in our optimization problem. However, there are also important benefits to allowing lifting in the definition of representability.

First, lifting and projection can dramatically reduce the problem "complexity". Specifically, consider the following LP-representable set[2]

[2] LP representability is the same as SOC representability and SDP representability where the cone $\mathcal{K}$ is the nonnegative orthant.

$$\mathcal{X} := \{x \in \mathbb{R}^n : \|x\|_1 \le 1\}.$$

A naive LP representation of this set uses 2^n constraints:

$$\mathcal{X} := \left\{ x \in \mathbb{R}^n : \begin{pmatrix} \sigma_1^\intercal \\ \sigma_2^\intercal \\ \vdots \\ \sigma_{2^n}^\intercal \end{pmatrix} x + \mathbf{1}_{2^n} \in \mathbb{R}_+^{2^n} \right\}$$

where $\sigma_1, \ldots, \sigma_{2^n}$ are the 2^n sign vectors in $\{\pm 1\}^n$. On the other hand, we can also write it as

$$\mathcal{X} := \Pi_x \left\{ (x,u) \in \mathbb{R}^n \times \mathbb{R}^n : \begin{pmatrix} u - x \\ u + x \\ 1 - \mathbf{1}^\intercal u \end{pmatrix} \in \mathbb{R}_+^{2n+1} \right\}.$$

Thus, if we were to use these descriptions within an LP, the number of decision variables would go from n to $2n$, but the number of constraints would decrease from 2^n to $2n+1$.

Next, sets with a lifted description may not have a "non-lifted" descriptions. A well-known example is the following set:

$$\Pi_{(x,z)} \left\{ (x,y,z) \in \mathbb{R}^3 : \begin{array}{l} 0 \le z \le 1 \\ \sqrt{(x-z)^2 + y^2} \le z/2 \end{array} \right\}.$$

This set is SOCR, however there does not exist even an SDP representation of this set that does not use lifting variables.

8
SDP applications

8.1 Stability analysis and synthesis

For $A \in \mathbb{R}^{n\times n}$ let $\rho(A)$ be the *spectral radius* of A:

$$\rho(A) \coloneqq \max_{i\in[n]} |\lambda_i(A)| \,.$$

Note that for a general matrix $A \in \mathbb{R}^{n\times n}$, the eigenvalues of A need not be real so $|\lambda_i(A)|$ is the modulus of the possibly complex eigenvalue $\lambda_i(A)$.

8.1.1 Analysis

Consider the following discrete-time dynamical system

$$x_{t+1} = Ax_t.$$

Stability analysis asks whether this system is stable, i.e., whether $\lim_{t\to\infty} x_t = 0$ for all starting conditions $x_0 \in \mathbb{R}^n$.

Lemma 18. *The following are equivalent*

(i) $\lim_{t\to\infty} x_t = 0$ *for any initial* $x_0 \in \mathbb{R}^n$

(ii) $\lim_{t\to\infty} x_t = 0$ *for any initial* $x_0 \in \mathbb{C}^n$

(iii) $\rho(A) < 1$

(iv) There exists $P \in \mathbb{S}^n$ *with* $P \succ 0$ *so that* $A^\intercal PA - P \prec 0$.

Proof. $(i) \implies (ii)$ Let $x_0 \in \mathbb{C}^n$ and write $x_0 = a_0 + ib_0$ where $a_0, b_0 \in \mathbb{R}^n$. Then,

$$x_t = A^t x_0 = (A^t a_0) + i(A^t b_0).$$

By assumption, this converges to zero.

$(ii) \implies (iii)$ Consider an arbitrary eigenvalue λ of A and let $v \in \mathbb{C}^n$ be a corresponding eigenvector with $\|v\| = 1$. Set $x_0 = v$. By assumption

$$0 = \lim_{t\to\infty} x_t = \lim_{t\to\infty} A^t v = \lim_{t\to\infty} \lambda^t v.$$

We deduce that $|\lambda| < 1$.

$(iii) \implies (iv)$ Define $P = \sum_{t=0}^{\infty} (A^t)^\intercal (A^t)$. This is well-defined as $\rho(A) < 1$.[1] Now, $P \succeq (A^0)^\intercal (A^0) = I \succ 0$. Furthermore,

$$A^\intercal P A - P = -I \prec 0.$$

$(iv) \implies (i)$ Finally, let $x_0 \in \mathbb{R}^n$. We track the evolution of $x_t^\intercal P x_t$. Let $\epsilon > 0$ so that $A^\intercal P A \preceq (1-\epsilon)P$.

$$x_t^\intercal P x_t = x_{t-1}^\intercal A^\intercal P A x_{t-1} \leq (1-\epsilon) x_{t-1}^\intercal P x_{t-1}.$$

We deduce that $x_t^\intercal P x_t \to 0$. As $P \succ 0$, we must have that $\lim_{t\to\infty} x_t = 0$. ■

[1] **Careful**: $\rho(A) < 1$ does *not* imply that $\|A\|_{\mathrm{op}} < 1$. Even so, $\rho(A) < 1$ ensures that this "geometric series" converges. If you haven't seen this before, it can be proved by checking the case of a single Jordan canonical block. This is done in the Notes section at the end of this chapter.

The function $f(x) = x^\intercal P x$ is called a *quadratic Lyapunov function*. You can think of it as some generalization or formalization of the notion of "energy." The function $f(x)$ assigns some nonnegative "energy" to every state x, and $f(x)$ is shown to be decreasing on every trajectory of our system $x \mapsto Ax$.

Thus, checking whether a system $x \mapsto Ax$ is stable is equivalent to checking whether

$$\inf_{P\in\mathbb{S}^n} \left\{ \lambda_{\max}(A^\intercal P A - P) : \begin{array}{l} P \succ 0 \\ \mathrm{tr}(P) \leq 1 \end{array} \right\}$$

is negative. This problem is SDP representable for any fixed A.

8.1.2 Analysis

Now, suppose the system is given by

$$x_{t+1} = Ax_t + Bu_t,$$

where $B \in \mathbb{R}^{n\times m}$ and $u \in \mathbb{R}^m$ is a control that we get to design to attempt to stabilize our system. We will consider the case of a linear control $u_t = Kx_t$, i.e. $K \in \mathbb{R}^{m\times n}$ is our *controller*. Thus, our goal is to find K so that

$$\rho(A + BK) < 1.$$

We cannot simply plug $A + BK$ into the previous stability analysis SDP as then $(A + BK)^\intercal P(A + BK)$ would be nonlinear in our variables

P and K. We make the nonlinear change of variables $Q = P^{-1}$ and $Y = KQ$ to rewrite the stability condition as follows:

$$\begin{aligned} & P \succ 0,\, P - (A+BK)^\intercal P(A+BK) \succ 0 \\ \iff \quad & Q \succ 0,\, Q - (AQ+BY)^\intercal Q^{-1}(AQ+BY) \\ \iff \quad & \begin{pmatrix} Q & (AQ+BY)^\intercal \\ AQ+BY & Q \end{pmatrix} \succ 0. \end{aligned}$$

We can now solve the synthesis problem by solving the SDP

$$\inf_{Q \in \mathbb{S}^n,\, Y \in \mathbb{R}^{m\times n}} \left\{ \|Y\|_{\mathrm{op}} : \begin{pmatrix} Q & (AQ+BY)^\intercal \\ AQ+BY & Q \end{pmatrix} \succ 0 \right\}$$

8.2 SDP Relaxation of Max-Cut

Let $G = ([n], E)$ be a graph on the vertex set $[n]$. Suppose each edge (i,j) has weight $w_{i,j}$. In the Max-Cut problem, we are asked to find a partition of $[n]$ into $S \subseteq [n]$ and S^c, in order to maximize

$$\sum_{i \in S} \sum_{j \in S^c} w_{i,j}.$$

This is an NP-hard problem but we will see an approximation algorithm for this problem based on semidefinite programming. This is called the Goemans–Williamson MaxCut SDP relaxation.

First, we rewrite the problem as minimizing a quadratic form over $\{\pm 1\}^n$: Define the following matrix[2]

$$L := \frac{1}{4} \sum_{(i,j)\in E} w_{i,j}(e_i - e_j)(e_i - e_j)^\intercal.$$

[2] This matrix is called the *Laplacian* matrix and acts as a discrete second-order derivative.

Now, suppose $x \in \{\pm 1\}^n$. It holds that

$$x^\intercal L x = \sum_{(i,j)\in E} w_{i,j} \mathbf{1}[\![x_i \neq x_j]\!],$$

i.e., if we identify x with the set of coordinates where it is equal to one, then $x^\intercal L x$ is the weight of the edges cut by the partition S, S^c.

Thus, the MaxCut problem can be relaxed as

$$\begin{aligned} & \max_{x \in \mathbb{R}^n} \left\{ x^\intercal L x : x_i^2 = 1,\, \forall i \in [n] \right\} \\ & = \max_{X \in \mathbb{S}^n} \left\{ \langle L, X \rangle : \begin{array}{l} X_{i,i} = 1,\, \forall i \in [n] \\ \operatorname{rank}(X) = 1 \\ X \succeq 0 \end{array} \right\} \\ & \leq \max_{X \in \mathbb{S}^n} \left\{ \langle L, X \rangle : \begin{array}{l} X_{i,i} = 1,\, \forall i \in [n] \\ X \succeq 0 \end{array} \right\}. \end{aligned}$$

Note, the problem on the first two lines is nonconvex and the problem on the last line is an SDP. Note also that the feasible domain of the SDP is compact so that the SDP optimizer exists.

Now consider the following procedure[3] for taking the SDP optimizer G and generating a vector $x \in \{\pm 1\}^n$.

[3] **Exercise**: Verify that the following procedure is equivalent (i.e., generates the same distribution on x): Sample $y \sim N(0, G)$ and output $x = \mathrm{sign}(y)$.

- Let U such that $G = U^\intercal U$ (for example, $U = G^{1/2}$ is one such option). Let u_i denote the columns of U.
- Sample $z \sim N(0, I)$
- Let $x = \mathrm{sign}(\langle z, u_i \rangle)$

Theorem 13 (Goemans–Williamson)**.** *It holds that*

$$\mathrm{Opt}(\text{Max-Cut}) \geq \mathop{\mathbb{E}}_x \left[x^\intercal L x\right] \geq (0.868\dots)\,\mathrm{Opt}(\mathrm{SDP}).$$

Proof. Define β to be the largest value so that

$$\arccos(x) \geq \beta(1 - x)$$

for all $x \in [-1, 1]$.

Consider any edge (i, j). This edge is cut if and only if $\langle u_i, z\rangle \leq 0 \leq \left\langle u_j, z\right\rangle$ or $\langle u_i, z\rangle \geq 0 \geq -\left\langle u_j, z\right\rangle$ (up to a probability zero event). This happens with probability $\frac{\theta_{i,j}}{\pi}$ where $\theta_{i,j} = \arccos(G_{i,j})$ is the angle between u_i and u_j.

Now, the expected value of $x^\intercal L x$ is

$$\begin{aligned}
\mathop{\mathbb{E}}_x \left[x^\intercal L x\right] &= \sum_{(i,j)\in E} \frac{w_{i,j} \arccos(G_{i,j})}{\pi} \\
&\geq \frac{\beta}{\pi} \sum_{(i,j)\in E} w_{i,j}(1 - G_{i,j}) \\
&= \frac{2\beta}{\pi} \left\langle L, G \right\rangle .
\end{aligned}$$

■

8.3 SDP relaxations of polynomial optimization problems

Let $\mathbb{R}[x]_d$ denote the polynomials in x with real coefficients and with degree at most d.

Let $f \in \mathbb{R}[x]_{2d}$ and consider the problem of minimizing

$$\inf_{x\in\mathbb{R}} f(x),$$

We will introduce one additional variable t and then think of the problem above as:

$$\sup_t \left\{ t :\ f(x) - t \geq 0,\ \forall x \right\}.$$

The optimum values are the same.

The Sum-of-Squares (SOS) hierarchy is a sequence of increasingly large and increasingly accurate SDP relaxations of this problem. It is parameterized by a degree $d \in 2\mathbb{N}$.

We will say that a polynomial $p \in \mathbb{R}[x]$ is a sum-of-squares if it can be written in the form

$$p = \sum_{i=1}^{k} q_i(x)^2$$

where each $q_i(x) \in \mathbb{R}[x]$ is itself a polynomial in x.

The following lemma states that in the *univaraite case*, the cone of nonnegative polynomials of degree $2d$ and the sum-of-squares polynomials of degree $2d$ are equal.

Lemma 19. *Let $p \in \mathbb{R}[x]_{2d}$. The following are equivalent*

1. *p is nonnegative*
2. *There exists $q_1, \ldots, q_k \in \mathbb{R}[x]_d$ so that*

$$p = \sum_{i=1}^{k} (q_i)^2.$$

Proof. The backwards direction is trivial.

For the forward direction, without loss of generality, we may assume that the coefficient on x^{2d} is 1. By the fundamental theorem of algebra, we can write

$$p(x) = \prod_{i=1}^{2d} (x - \lambda_i)$$

where $\lambda_i \in \mathbb{C}$ are the (possibly complex) roots of p (without repetition). As p is nonnegative and real, every real root must have an even multiplicity, and every complex root must also come with its conjugate (with multiplicity). Thus, we can write

$$p(x) = \prod_{i=1}^{d} (x - \lambda_i)(x - \bar{\lambda}_i) = |q(x)|^2$$

where $q(x) \coloneqq \prod_{i=1}^{d} (x - \lambda_i)$. Let $q_1(x)$ and $q_2(x)$ be the polynomials attained by taking the real parts of the coefficients of q and the imaginary parts of the coefficient of q respectively. Then,

$$q_1(x)^2 + q_2(x)^2 = |q(x)|^2 = p(x). \qquad \blacksquare$$

Let $\hat{x}$ denote the following symbolic vector $(1, x, \ldots, x^d)$. We can map symmetric matrices of size $\mathbb{S}^{1+d}$ to polynomials in $\mathbb{R}[x]_{2d}$ in the following way

$$A \mapsto p_A := \hat{x}^\intercal A \hat{x} = \left\langle A, \begin{pmatrix} 1 & x & x^2 & \dots & x^d \\ x & x^2 & \iddots & \iddots & x^{d+1} \\ x^2 & \iddots & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & \iddots & x^{2d-1} \\ x^d & x^{d+1} & \dots & x^{2d-1} & x^{2d} \end{pmatrix} \right\rangle$$

We will index the columns and rows of A by $[0, n]$ to correspond to the degrees of the vector $(1, x, \dots, x^n)$.

Lemma 20. *$p \in \mathbb{R}[x]_{2d}$ is a sum-of-squares if and only if we can write $p = p_A$ with a positive semidefinite matrix A.*

Proof. First, suppose $p \in \mathbb{R}[x]_{2d}$ is a sum-of-squares. Then, there exist $q_1, \dots, q_k \in \mathbb{R}[x]_d$ such that

$$p(x) = \sum_{i=1}^k (q_i(x))^2.$$

Each $q_i(x)$ is of the form

$$q_i(x) = \left\langle \alpha^{(i)}, \hat{x} \right\rangle.$$

Thus,

$$p(x) = \hat{x}^\intercal \left(\sum_{i=1}^k (\alpha_i \alpha_i^\intercal) \right) \hat{x}.$$

On the other hand, suppose

$$p(x) = \hat{x}^\intercal A \hat{x}$$

for some PSD matrix A. By the spectral decomposition, we can write $A = \sum_{i=1}^k (\alpha_i)(\alpha_i)^\intercal$. Then,

$$p(x) = \sum_{i=1}^k (\alpha_i^\intercal \hat{x})^2.$$

Each $\alpha_i^\intercal \hat{x}$ is a real polynomial in x of degree at most d. ■

We deduce that

$$\begin{aligned} \inf_{x \in \mathbb{R}} f(x) &= \sup_{t \in \mathbb{R}} \{ t : p - t \text{ is a sum-of-squares} \} \\ &= \sup_{t \in \mathbb{R},\, A \in \mathbb{S}^{1+d}} \left\{ t : \begin{array}{l} A \succeq 0 \\ p_A = p - t \end{array} \right\}. \end{aligned}$$

The linear constraint here imposes linear constraints on the matrix A. Specifically, it specifies the sum on each antidiagonal of A.

Notes

Lemma 21. *Let J be an $n \times n$ Jordan canonical block corresponding to the eigenvalue $\lambda \in \mathbb{C}$ with $|\lambda| < 1$. Then,*

$$\left\| J^k \right\|_{\mathrm{op}}$$

is exponentially small in k.

Proof. For all $k \geq n-1$ we have

$$J^k = \begin{pmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \ddots & \binom{k}{n-1}\lambda^{k-n+1} \\ 0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & \ddots & \ddots \\ \vdots & \vdots & \ddots & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} \\ 0 & 0 & \dots & \lambda^k & \binom{k}{1}\lambda^{k-1} \\ 0 & 0 & \dots & 0 & \lambda^k \end{pmatrix}.$$

The binomial coefficients are at most polynomially large in k whereas $|\lambda^{k-n+1}|$ is exponentially small in k. ∎

Lemma 22. *Let J be an $n \times n$ Jordan canonical block corresponding to the eigenvalue $\lambda \in \mathbb{C}$ with $|\lambda| < 1$. Then,*

$$\sum_{k=0}^{\infty} (J^k)^* J^k$$

is well-defined, i.e., the partial sums converge.

Proof. It suffices to show that

$$\sum_{k=0}^{\infty} \left\| (J^k)^* J^k \right\|_{\mathrm{op}}$$

is bounded. This is bounded as $\left\| (J^k)^*(J^k) \right\|_{\mathrm{op}} = \left\| J^k \right\|_{\mathrm{op}}$ is exponentially small in k. ∎

9

Subgradient descent for nonsmooth convex optimization

This chapter will begin our study of *first-order methods*. These are iterative algorithms that rely only on first-order information, i.e., function value, gradient, or subgradient information (notably omitting Hessian information).

Remark 7. In this chapter, all norms are the ℓ_2 norm. □

We consider the problem of solving

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

where $\Omega \subseteq \mathbb{R}^n$ is closed and convex and $f : \mathbb{R}^n \to \mathbb{R}$ is convex and L-Lipschitz continuous:

Definition 30. $f : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz continuous if

$$|f(x) - f(y)| \leq L \|x - y\| \qquad \forall x, y \in \mathbb{R}^n.$$ □

9.1 Subgradients of convex functions

Definition 31. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then $g \in \mathbb{R}^n$ is a subgradient of f at x if

$$f(x) + \langle g, y - x\rangle \leq f(y) \qquad \forall y \in \mathbb{R}^n.$$

The set of subgradients of f at $x \in \mathbb{R}^n$ is denoted $\partial f(x)$. □

Proposition 3. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex. Then, $\partial f(x)$ is nonempty for all $x \in \mathbb{R}^n$.*

Proof. Consider the strict epigraph

$$\mathcal{S} = \left\{(y, t) \in \mathbb{R}^{n+1} : \quad f(y) < t \quad\right\}$$

and the point $(x, f(x))$. By construction, this is a pair of disjoint nonempty convex sets. Thus, the hyperplane separation theorem gives $(g, \alpha) \in \mathbb{R}^{n+1}$ nonzero so that

$$\langle -g, x\rangle + \alpha f(x) \leq \inf_{(y,t)\in\mathcal{S}} \langle -g, y\rangle + \alpha t.$$

By taking $t \to \infty$, we see that $\alpha \geq 0$. We claim that $\alpha \neq 0$. Indeed, suppose $\alpha = 0$ and consider $y = x + \epsilon g$. Then,

$$\langle -g, x\rangle \leq \langle -g, x\rangle - \epsilon \|g\|^2 < \langle -g, x\rangle,$$

a contradiction. We deduce that $a > 0$.

We may thus WLOG assume that $a = 1$. Then, for all $y \in \mathbb{R}^n$,

$$f(x) + \langle g, y - x\rangle \leq f(y).$$

We deduce that $g \in \partial f(x)$. ∎

Lemma 23. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then, f is L-Lipschitz if and only if $\|g\| \leq L$ for all $x \in \mathbb{R}^n$ and all $g \in \partial f(x)$.*

Proof. First, suppose $x, y \in \mathbb{R}^n$. Let $g \in \partial f(x)$. Then,

$$f(x) - f(y) \leq \langle g, x - y\rangle \leq L \|x - y\|.$$

Reversing the roles with $g \in \partial f(y)$, we also have that

$$f(y) - f(x) \leq L \|x - y\|.$$

We deduce that $|f(x) - f(y)| \leq L \|x - y\|$ and f is L-Lipschitz.

Now, suppose f is L-Lipschitz and assume for the sake of contradiction that there exists x and $g \in \partial f(x)$ with $\|g\| > L$. Let $y = x + g$. Then,

$$f(y) - f(x) \geq \langle g, y - x\rangle = \|g\|^2 > L \|y - x\|,$$

a contradiction. ∎

The following lemma relates subgradients to gradients.[1]

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

Lemma 24. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable. Then, $\partial f(x) = \{\nabla f(x)\}$. Thus, for all $x, y \in \mathbb{R}^n$,*

$$f(x) + \langle \nabla f(x), y - x\rangle \leq f(y).$$

9.2 The projected subgradient algorithm

Recall the problem that we are trying to solve is

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and L-Lipschitz and $\Omega \subseteq \mathbb{R}^n$ is closed and convex.

The following algorithm is the *projected subgradient method*.

Algorithm 1 Projected subgradient method

Given: Initial iterate $x_0 \in \Omega$, step lengths $\eta_0, \dots, \eta_T > 0$, time horizon T

- For $t = 0, \dots, T-1$, set

$$\begin{aligned} y_{t+1} &= x_t - \eta_t g_t, \qquad \text{for some } g_t \in \partial f(x_t) \\ x_{t+1} &= \Pi_\Omega(y_t). \end{aligned}$$

- Let $\mu = \sum_{t=0}^T \eta_t$ and define $\bar{x} \coloneqq \sum_{t=0}^T \frac{\eta_t}{\mu} x_t$

Theorem 14. *Suppose* $\inf_{x\in\Omega} f(x)$ *has a minimizer* $x^\star$ *with optimal value* $f^\star$ *and* $\|x_0 - x^\star\| \leq R$. *The projected subgradient method guarantees*

$$\begin{aligned} f(\bar{x}) - f^\star &\leq \frac{R^2}{2\mu} + \frac{\sum_{t=0}^T \eta_t^2 \|g_t\|^2}{2\mu} \\ &\leq \frac{R^2}{2\mu} + \frac{L^2 \sum_{t=0}^T \eta_t^2}{2\mu}. \end{aligned}$$

Proof. For the sake of the proof, we will imagine simulating one additional step of the method so that x_{T+1} and y_{T+1} are also defined.

Let $t \in [0, T]$. We compute

$$\begin{aligned} f(x_t) - f^\star &\leq \langle g_t, x_t - x^\star \rangle && \text{(definition of subgradient)} \\ &= \frac{1}{\eta_t} \langle x_t - y_{t+1}, x_t - x^\star \rangle && \text{(definition of } y_{t+1}\text{)} \\ &= \frac{1}{2\eta_t} \left(\|x_t - x^\star\|^2 + \|x_t - y_{t+1}\|^2 - \|y_{t+1} - x^\star\|^2 \right) && \text{(Parallelogram law)} \\ &= \frac{1}{2\eta_t} \left(\|x_t - x^\star\|^2 - \|y_{t+1} - x_*\|^2 \right) + \frac{\eta_t}{2} \|g_t\|^2 . \end{aligned}$$

Next, we will use the fact that $\|y_{t+1} - x^\star\| \geq \|x_{t+1} - x^\star\|$. Thus,

$$f(x_t) - f^\star \leq \frac{1}{2\eta_t} \left(\|x_t - x^\star\|^2 - \|x_{t+1} - x^\star\|^2 \right) + \frac{\eta_t \|g_t\|^2}{2}.$$

Let $\mu = \sum_{t=0}^T \eta_t$. We will take an (η_t/μ)-weighted sum of these inequalities to get

$$\begin{aligned}
\sum_{t=0}^{T} \frac{\eta_t}{\mu}(f(x_t) - f^\star) &\le \frac{\|x_0 - x^\star\|^2 - \|x_{T+1} - x^\star\|^2}{2\mu} + \frac{\sum_{t=0}^{T} \eta_t^2 \|g_t\|^2}{2\mu} \\
&\le \frac{R^2}{2\mu} + \frac{\sum_{t=0}^{T} \eta_t^2 \|g_t\|^2}{2\mu} \\
&\le \frac{R^2}{2\mu} + \frac{L^2 \sum_{t=0}^{T} \eta_t^2}{2\mu}.
\end{aligned}$$

The fact that $f(\bar{x}) - f^\star$ is at most the LHS follows from convexity. ■

Corollary 5. *Suppose $\eta_t > 0$ satisfies $\sum_{t=0}^{\infty} \eta_t = \infty$ and $\sum_{t=0}^{\infty} \eta_t < \infty$. Then, $f(\bar{x}_T) - f^\star \to 0$.*

Corollary 6. *Taking $\eta_t = \frac{R}{\|g_t\|\sqrt{t+1}}$ gives*

$$f(\bar{x}) - f^\star \le \frac{LR(2 + \ln(T+1))}{2(\sqrt{T+2} - 1)}$$

Proof. For any T we have that

$$\begin{aligned}
f(\bar{x}) - f^* &\le \frac{R^2}{2\mu} + \frac{R^2 \sum_{t=0}^{T} \frac{1}{t+1}}{2\mu} \\
&\le \frac{R^2(2 + \ln(T+1))}{2\mu}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mu &= \sum_{t=0}^{T} \frac{R}{\|g_t\| \sqrt{t+1}} \\
&\ge \frac{R}{L} \sum_{t=0}^{T} \frac{1}{\sqrt{t+1}} \\
&\ge \frac{2R(\sqrt{T+2} - 1)}{L}.
\end{aligned}$$

■

Suppose we fix the time horizon T and want to pick the set of step sizes $(\eta_0, \ldots, \eta_T) \in \mathbb{R}^{T+1}$ to minimize the upper bound

$$\frac{R^2}{2\mu} + \frac{L^2 \sum_{t=0}^{T} \eta_t^2}{2\mu}.$$

First, note that for any fixed total μ, the optimal η to pick is the one that is constant with step sizes $\eta_t = \frac{\mu}{T+1}$ (think: minimize ℓ_2 norm subject to constant ℓ_1 norm). Then, restricting ourselves to constant stepsizes $\eta_t = \eta$, the upper bound simplifies to

$$\frac{R^2}{2(T+1)\eta} + \frac{L^2\eta}{2}.$$

This is the arithmetic mean of $\frac{R^2}{(T+1)\eta}$ and $L^2\eta$. Note that the geometric mean is unchanged upon varying η. Thus, the upper bound is always at least $LR/\sqrt{T+1}$. On the other hand, we can set the two to be equal by setting $\eta = \frac{R}{L\sqrt{T+1}}$ (so that the AM-GM inequality is tight).

Corollary 7. *Suppose* $\inf_{x\in\Omega} f(x)$ *has a minimizer* $x^\star$ *with optimal value* $f^\star$ *and* $\|x_0 - x^\star\| \le R$*. The projected subgradient method with* $\eta = \frac{R}{L\sqrt{T+1}}$ *guarantees*

$$f(\bar{x}) - f^\star \le \frac{LR}{\sqrt{T+1}}.$$

In particular, it achieves an ϵ *suboptimal solution in* $O\left(\left(\frac{LR}{\epsilon}\right)^2\right)$ *iterations.*

Exercises

- Consider $f : \mathbb{R}^2 \to \mathbb{R}$ by
$$f(x) = |x_1| + 2|x_2|.$$
Show that $\partial f(1,0) = \{(1,y) : |y| \le 2\}$. Thus, $(1,2) \in \partial f(1,0)$. Next, show that $f((1,0) - t(1,2)) > f(1,0)$ for all $t > 0$. Thus, $-(1,2)$ is not a descent direction.

- Let $f_i : \mathbb{R}^d \to \mathbb{R}$ be convex and differentiable for $i = 1, \dots, n$. Let $F(x) \coloneqq \max_i f_i(x)$. Show that
$$\partial F(x) = \operatorname{conv}(\{\nabla f_i(x) : f_i(x) = F(x)\}).$$

Problems

1. Let $\gamma > 1$ and consider the following function $f : \mathbb{R}^2 \to \mathbb{R}$
$$f(x) = \begin{cases} \sqrt{x_1^2 + \gamma x_2^2} & \text{if } |x_2| \le x_1 \\ \frac{x_1 + \gamma |x_2|}{\sqrt{1+\gamma}} & \text{else} \end{cases}$$
This function is convex and $\sqrt{\gamma}$-Lipschitz (you do not need to prove this).

   Consider the subgradient method with *exact* line-search initialized at $x^{(0)} = (\gamma, 1)$, i.e., for $t \ge 1$, let $g \in \partial f(x^{(t-1)})$ and set
$$x^{(t)} = \operatorname*{arg\,min}_{x \in x^{(t-1)} - \mathbb{R}_+ g} f(x)$$

   (a) Prove that for a general convex function $f : \mathbb{R}^n \to \mathbb{R}$, if f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$. Recall, if f is differentiable at x, then $\nabla f(x)$ is defined to be the unique vector in $\mathbb{R}^n$ so that for all $u \in \mathbb{R}^n$,
$$\frac{d}{dt} f(x + tu) = \langle \nabla f(x), u \rangle.$$

   (b) Prove by induction that $x^{(t)} = \left(\gamma \left(\frac{\gamma-1}{\gamma+1}\right)^t, \left(\frac{1-\gamma}{\gamma+1}\right)^t\right)$ for all $t \ge 0$.

   This shows that the subgradient method with exact line-search converges to the origin where $f(0) = 0$. On the other hand, f can be made arbitrarily negative by sending $x_1 \to -\infty$.

2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a L-Lipschitz convex function with minimizer $x^\star$ and minimum value f^*. Suppose that f satisfies the following growth condition parameterized by $\delta > 0, \alpha > 0$:
$$f(x) - f^\star \le \delta \quad \implies \quad f(x) - f^* \ge \alpha \|x - x^\star\|^2.$$

Suppose we are given $x_0 \in \mathbb{R}^n$ with $\|x_0 - x^\star\| \leq R$.

Fill in the missing details (i.e., replace the ?s) in the following *restarted* subgradient method. Consider the following algorithm:

Algorithm 2 Restarted subgradient method

Given: $L, R, \alpha, \delta, x_0$

- For each $k = 0, \dots$
  - Run the subgradient method with constant stepsizes (see Corollary 11) with initial iterate x_k for
$$T_k = ?$$
    iterations. Let x_{k+1} to be the output of the subgradient method.

By setting $T_0 = ?$, we can ensure the following property:

Lemma 25. *It holds that* $f(x_1) - f^\star \leq \delta$.

Proof. ? ■

For $k \geq 1$, define $\delta_k = \frac{2}{2^k}\delta \leq \delta$. By setting $T_k = ?$ for $k \geq 1$, we can ensure the following property:

Lemma 26. *It holds that* $f(x_k) - f^\star \leq \delta_k$.

Proof. ? ■

We conclude that:

Proposition 4. *The restarted subgradient method with constant stepsizes and horizons* $T_0 = ?$ *and* $T_k = ?$ *for all* $k \geq 1$ *achieves a gap* $f(x) - f^\star \leq \epsilon$ *after at most*

$$O\left(\frac{L^2R^2}{\delta^2} + \frac{L^2}{\alpha\epsilon}\right)$$

total (inner) iterations. Thus for $\epsilon \ll \frac{\delta^2}{\alpha R^2}$*, this convergence rate is* $O\left(\frac{L^2}{\alpha\epsilon}\right)$.

Compare this rate with Corollary 11.

10
Gradient descent for smooth and strongly convex optimization

Remark 8. All norms in this lecture are Euclidean norms. □

10.1 Smoothness and strong convexity

Definition 32. Let $L \geq 0$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth if f is differentiable and for all $x, y \in \mathbb{R}^n$

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| . \quad \square$$

Definition 33. Let $\mu \geq 0$. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex if for all $x, y \in \mathbb{R}^n$ and all $t \in [0, 1]$,

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) - \frac{\mu}{2}(1-t)(t) \|x - y\|^2 . \quad \square$$

10.1.1 Properties of smooth and strongly convex functions

Lemma 27. *Let $L \geq 0$ and suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex. If f is L-smooth, then for all $x, y \in \mathbb{R}^n$*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2$$

Proof. Define

$$g(t) \coloneqq f(x + t(y - x)) - [f(x) + t \langle \nabla f(x), y - x \rangle] .$$

Then, $g(t)$ is differentiable and

$$\begin{aligned} f(y) - f(x) - \langle \nabla f(x), y - x \rangle &= g(1) \\ &= g(0) + \int_0^1 g'(t)\, dt \\ &= \int_0^1 \langle \nabla f(x + t(y-x)) - \nabla f(x), y - x \rangle \; dt \\ &\leq \int_0^1 Lt \left\| y - x \right\|^2 \; dt \\ &= \frac{L \left\| y - x \right\|^2}{2}. \end{aligned}$$

■

Lemma 28. *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ satisfies for all $x, y \in \mathbb{R}^n$*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \left\| y - x \right\|^2$$

Then, $y = x - \frac{1}{L} \nabla f(x)$ satisfies

$$f(y) \leq f(x) - \frac{1}{2L} \left\| \nabla f(x) \right\|^2.$$

Lemma 29. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Furthermore, suppose that for all $x, y \in \mathbb{R}^n$*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \left\| x - y \right\|^2.$$

Then, f is L-smooth.

Proof. Our goal is to show that $\left\| \nabla f(x) - \nabla f(y) \right\| \leq L \left\| x - y \right\|^2$ for all $x, y \in \mathbb{R}^n$. It suffices to prove this statement in the case where $x = 0$, $f(x) = 0$, and $\nabla f(x) = 0$ as otherwise we can consider the function

$$g(\delta) := f(\delta + x_0) - f(x_0) - \langle \nabla f(x_0), \delta \rangle$$

instead.

Now, suppose $x = 0$, $f(x) = 0$, and $\nabla f(x) = 0$. Let y be arbitrary and set $z = y - \frac{1}{L} \nabla f(y)$. Then,

$$0 \leq f(z) \leq f(y) - \frac{1}{2L} \left\| \nabla f(y) \right\|^2 \leq \frac{L}{2} \left\| x - y \right\|^2 - \frac{1}{2L} \left\| \nabla f(y) \right\|^2.$$

Rearranging completes the proof. ■

Lemma 30. *Suppose $\mu \geq 0$ and $f : \mathbb{R}^n \to \mathbb{R}$ is convex. Then, f is μ-strongly convex if and only if for all $x, y \in \mathbb{R}^n$ and all $g \in \partial f(x)$,*

$$f(y) \geq f(x) + \langle g, y - x \rangle + \frac{\mu}{2} \left\| y - x \right\|^2.$$

Proof. Throughout this proof let $x_t := (1-t)x + ty$

First, suppose f is μ-strongly convex. By definition, for all $t \in (0,1]$,

$$\begin{aligned} f(y) &\geq \frac{f(x_t) - (1-t)f(x)}{t} + \frac{\mu}{2}(1-t)\|x-y\|^2 \\ &= f(x) + \frac{\mu}{2}(1-t)\|x-y\|^2 + \frac{f(x_t)-f(x)}{t} \\ &\geq f(x) + \frac{\mu}{2}(1-t)\|x-y\|^2 + \langle g, y-x\rangle . \end{aligned}$$

Taking the limit as $t \to 0$ shows that

$$f(y) \geq f(x) + \frac{\mu}{2}\|x-y\|^2 + \langle g, y-x\rangle .$$

In the other direction, fix $t \in [0,1]$ and set $g \in \partial f(x_t)$. Invoke the supplied inequality twice to get

$$\begin{aligned} f(y) &\geq f(x_t) + \langle g, y - x_t\rangle + \frac{\mu}{2}\|x_t - y\|^2 \\ f(x) &\geq f(x_t) + \langle g, x - x_t\rangle + \frac{\mu}{2}\|x_t - x\|^2 . \end{aligned}$$

Note that $y - x_t = (1-t)(y-x)$ and $x - x_t = t(x-y)$. Thus, this is equivalent to

$$\begin{aligned} f(y) &\geq f(x_t) + (1-t)\langle g, y-x\rangle + \frac{\mu}{2}(1-t)^2\|y-x\|^2 \\ f(x) &\geq f(x_t) - t\langle g, y-x\rangle + \frac{\mu}{2}t^2\|y-x\|^2 . \end{aligned}$$

Taking the $t, (1-t)$ weighted average of these inequalities proved that f is μ-strongly convex. ∎

Lemma 31. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$. Then, f is μ-strongly convex if and only if $f(x) - \frac{\mu}{2}\|x\|^2$ is convex.*

Proof. Let $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$.

Note that $g(x)$ is convex if and only if for all $x, y \in \mathbb{R}^n$ and $t \in [0,1]$,

$$g((1-t)x + ty) \leq (1-t)g(x) + tg(y).$$

This is if and only if

$$\begin{aligned} &f((1-t)x+ty) - \frac{\mu}{2}\|(1-t)x+ty\|^2 \\ &\qquad \leq (1-t)\left[f(x) - \frac{\mu}{2}\|x\|^2\right] + t\left[f(y) - \frac{\mu}{2}\|y\|^2\right] . \end{aligned}$$

Rearranging, this is

$$\begin{aligned} &f((1-t)x+ty) \\ &\qquad \leq (1-t)f(x) + tf(y) + \frac{\mu}{2}t(1-t)\|x-y\|^2 . \end{aligned}$$ ∎

Lemma 32. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex and suppose $x^\star$ is a minimizer of f. Then,*

$$f(y) \geq f(x^*) + \frac{\mu}{2} \|y - x^\star\|^2 .$$

Lemma 33. *Suppose $f, g : \mathbb{R}^n \to \mathbb{R}$ are α-strongly convex and β-strongly convex respectively. Then, $f + g$ is $\alpha + \beta$ strongly convex. Suppose $\lambda \geq 0$, then λf is $\lambda\alpha$-strongly convex.*

10.2 The Prox Point Method

Consider the following algorithm, known as the Prox Point Method:

Algorithm 3 Prox Point Method

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an *arbitrary* convex function. Let $x_0 \in \mathbb{R}^n$ and let $\eta_0, \eta_1, \dots > 0$

- For $t = 1, \dots,$ set

$$x_t \in \operatorname*{arg\,min}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2\eta_{t-1}} \|x - x_{t-1}\|^2 \right\}$$

This algorithm is not practically implementable (usually). However, it will serve as a template for understanding other algorithms.

Theorem 15. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function with a minimizer $x^\star$ with value $f^\star$. Suppose $\eta_0, \dots, \eta_{T-1} > 0$. Then,*

$$f(x_T) - f^\star \leq \left(\sum_{t=0}^{T-1} \eta_{t-1} \right)^{-1} \frac{\|x_0 - x_\star\|^2}{2}.$$

Proof. Suppose $t \geq 1$. Then,

$$\phi(x) := \eta_{t-1} f(x) + \frac{1}{2} \|x - x_{t-1}\|^2$$

is a 1-strongly convex function. Thus,

$$\phi(x_t) + \frac{1}{2} \|x_t - x_\star\|^2 \leq \phi(x_\star).$$

Expanding this and dropping the term $\|x_t - x_{t-1}\|^2$ gives

$$\eta_{t-1}(f(x_t) - f^\star) + \frac{1}{2} \left(\|x_t - x_\star\|^2 - \|x_{t-1} - x_\star\|^2 \right) \leq 0.$$

Summing up these inequalities over t gives

$$\sum_{t=1}^{T} \eta_{t-1}(f(x_t) - f^\star) + \frac{1}{2} \left(\|x_T - x_\star\|^2 - \|x_0 - x_\star\|^2 \right) \leq 0.$$

Note also that $f(x_0) \geq f(x_1) \geq \dots$. This follows as

$$\eta_{t-1} f(x_t) + \frac{1}{2} \|x_t - x_{t-1}\|^2 \leq \eta_{t-1} f(x_{t-1}).$$

We conclude that

$$f(x_T) - f^\star \leq \left(\sum_{t=1}^{T} \eta_{t-1} \right)^{-1} \frac{\|x_0 - x_\star\|^2}{2}. \qquad \blacksquare$$

10.3 Gradient descent for smooth convex functions

We will attempt to approximate the prox-point method by the update rule

$$x_t \in \arg\min_{x \in \mathbb{R}^n} \left\{ f(x_{t-1}) + \langle \nabla f(x_{t-1}), x - x_{t-1} \rangle + \frac{1}{2\eta_{t-1}} \|x - x_{t-1}\|^2 \right\}.$$

That is, we replace $f(x)$ by its first-order approximation at x_{t-1}.

Another way to write this update rule is as

$$x_t = x_{t-1} - \eta_{t-1} \nabla f(x_{t-1}).$$

This is the *gradient descent* update rule.

Theorem 16. *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is an L-smooth convex function with minimizer $x^\star$ and value $f^\star$. Let $x_0 \in \mathbb{R}^d$ and iteratively set $x_t = x_{t-1} - \eta \nabla f(x_{t-1})$, where $\eta = \frac{1}{L}$. Then,*

$$f_T - f^\star \leq \frac{L \|x_0 - x^\star\|^2}{2T}.$$

Proof. We will attempt to use the same proof strategy as for the prox-point method, but will have to keep track of potential errors.

Suppose $t \geq 1$. Then,

$$\phi(x) \coloneqq \eta \left(f_{t-1} + \langle g_{t-1}, x - x_{t-1} \rangle \right) + \frac{1}{2} \|x - x_{t-1}\|^2$$

is 1-strongly convex. Thus,

$$\phi(x_t) + \frac{1}{2} \|x_t - x^\star\|^2 \leq \phi(x^\star).$$

We will use the bounds:

$$f_t - \frac{L}{2} \|x_t - x_{t-1}\|^2 \leq f_{t-1} + \langle g_{t-1}, x_t - x_{t-1} \rangle$$
$$f_{t-1} + \langle g_{t-1}, x^\star - x_{t-1} \rangle \leq f(x^\star)$$

Thus,

$$\eta (f_t - f^\star) + \frac{1}{2} \left(\|x_t - x^\star\|^2 - \|x_{t-1} - x^\star\|^2 \right) \leq 0.$$

Adding up these inequalities gives

$$\sum_{t=1}^{T} \eta(f_t - f^\star) \leq \frac{\|x_0 - x^\star\|^2}{2}.$$

Next, we have that $f_0 \geq f_1 \geq \dots$. This holds because

$$f_t \leq f_{t-1} - \frac{\|g_{t-1}\|^2}{2L}.$$

We conclude that

$$f_T - f^\star \leq \frac{L\,\|x_0 - x^\star\|^2}{2T}. \qquad \blacksquare$$

10.4 Accelerated gradient descent for smooth minimization

It turns out that gradient descent does not achieve the optimal convergence rate. We can do much better if we decouple the location where we query first order information from our sequence x_t. Consider the following scheme

---

Algorithm 4 Accelerated gradient descent for smooth convex minimization

---

Given $x_0 \in \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ convex and L-smooth

- Set $y_0 = x_0$
- For $t = 0, \dots$

$$x_{t+1} = y_t - \frac{1}{L}\nabla f(y_t)$$
$$y_{t+1} = x_{t+1} + \gamma_t(x_{t+1} - x_t)$$

We will attempt to prove a convergence rate for this method that has the usual telescoping structure. There will be a natural choice of γ_t that will appear in the proof that will allow for telescoping:

Proof/Derivation of γ_t. Let $\delta_t = f(x_t) - f^\star$, $g_t = \nabla f(y_t)$, and $\Delta_t = y_t - x_t$.

Now, as f is L-smooth, we have that for all $t \geq 0$.

$$f(x_{t+1}) \leq f(y_t) - \frac{1}{2L}\|\nabla f(y_t)\|^2.$$

Combining this with $f(y_t) \leq f(x_t) + \langle \nabla f(y_t), y_t - x_t \rangle$ gives

$$\delta_{t+1} - \delta_t \leq \langle g_t, \Delta_t \rangle - \frac{1}{2L}\|g_t\|^2.$$

Combining this with $f(y_t) \le f^\star + \langle \nabla f(y_t), y_t - x^\star \rangle$ gives

$$\delta_{t+1} \le \langle g_t, \Delta_t + x_t - x^\star \rangle - \frac{1}{2L} \|g_t\|^2$$

Now, let us take the first inequality weighted by $(\lambda_t - 1)$ for some $\lambda_t \ge 1$ and add it to the second inequality to get

$$\lambda_t \delta_{t+1} - (\lambda_t - 1)\delta_t \le \langle g_t, \lambda_t \Delta_t + (x_t - x^\star) \rangle - \frac{\lambda_t}{2L} \|g_t\|^2 .$$

We will complete the square on the right hand side to write it as

$$\begin{aligned}
&\langle g_t, \lambda_t \Delta_t + (\lambda_t - 1)(x_t - x^\star) \rangle - \frac{\lambda_t}{2L} \|g_t\|^2 \\
&\quad = \frac{L}{2\lambda_t} \left(2 \left\langle \frac{\lambda_t g_t}{L}, \lambda_t \Delta_t + (x_t - x^\star) \right\rangle - \left\| \frac{\lambda_t g_t}{L} \right\|^2 \right) \\
&\quad = \frac{L}{2\lambda_t} \left(\|\lambda_t \Delta_t + (x_t - x^\star)\|^2 - \left\| \lambda_t \Delta_t + (x_t - x^\star) - \frac{\lambda_t g_t}{L} \right\|^2 \right) .
\end{aligned}$$

We will choose λ_t and γ_t so that

$$\lambda_t \Delta_t + (x_t - x^\star) - \frac{\lambda_t g_t}{L} = \lambda_{t+1} \Delta_{t+1} + (x_{t+1} - x^\star).$$

This can be achieved by setting $\lambda_t = 1 + \lambda_{t+1}\gamma_{t+1}$. Finally, set $\lambda_{t-1}^2 = \lambda_t^2 - \lambda_t$ where $\lambda_{-1} := 1$. This gives us for all $t \ge 0$,

$$\lambda_t^2 \delta_{t+1} - \lambda_{t-1}^2 \delta_t \le \frac{L}{2} \left(\|\lambda_t \Delta_t + (x_t - x^\star)\|^2 - \|\lambda_{t+1} \Delta_{t+1} + (x_{t+1} - x^\star)\|^2 \right) .$$

Now, telescoping this inequality gives us

$$\lambda_T^2 \delta_{T+1} \le \frac{L}{2} \|x_0 - x^\star\|^2 + \delta_0. \qquad \blacksquare$$

We summarize the derivation above:

Theorem 17. *Consider the accelerated gradient descent method where $\lambda_{-1} = 1$ and we inductively define for $t \ge 0$*

$$\begin{cases} \lambda_t = \frac{1 + \sqrt{1 + 4\lambda_{t-1}^2}}{2} \\ \gamma_t = \frac{\lambda_{t-1} - 1}{\lambda_t} \end{cases} .$$

Then,

$$f(x_T) - f^\star \le \frac{4L}{(T+2)^2} \|x_0 - x^\star\|^2 ,$$

Proof. It suffices to check that

$$\lambda_{-1} = 1 \qquad \lambda_t \ge \frac{1}{2} + \lambda_{t-1}$$

so that $\lambda_{T-1} \ge \frac{T+2}{2}$. $\blacksquare$

	L-smooth	L-smooth, μ-SC
GD	$\frac{L\|x_0-x^\star\|^2}{T}$	$\left(1-\kappa^{-1}\right)^T\left(L\left\|x_0-x^\star\right\|^2\right)$
Accel. GD	$\frac{L\|x_0-x^\star\|^2}{T^2}$	$\left(1-\kappa^{-1/2}\right)^T\left(L\left\|x_0-x^\star\right\|^2\right)$

Table 10.1: Bounds on $f(x_T)-f^\star$ for gradient descent and accelerated gradient descent for L-smooth and L-smooth and μ-strongly convex minimization up to $O(\cdot)$.

10.5 (Accelerated) gradient descent for smooth strongly convex minimization

A similar story holds for smooth and strongly convex minimization.

Algorithm 5 Accelerated gradient descent for smooth and strongly convex minimization

Given $x_0 \in \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ that is L-smooth and μ-strongly convex

- Set $x_1 = x_0 - \frac{1}{L}\nabla f(x_0)$
- For $t = 1, \ldots$

$$y_t = x_t + \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)(x_t - x_{t-1})$$
$$x_{t+1} = y_t - \frac{1}{L}\nabla f(y_t)$$

Theorem 18. *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth and μ-strongly convex (set $\kappa = L/\mu$). Suppose f has a minimizer $x^\star$ with optimal value $f^\star$. Then, for any $x_0 \in \mathbb{R}^d$, we have*

$$f(x_T) - f^\star \le \left(1 - \frac{1}{\sqrt{\kappa}}\right)^k \left(L\left\|x_0 - x^\star\right\|^2\right).$$

10.6 Minimizing a quadratic function

Consider a quadratic function of the form

$$f(x) = \frac{x^\intercal A x}{2} + b^\intercal x + c.$$

We will assume that $A \succ 0$ so that the unique minimizer of this problem is $x^\star = A^{-1}b$. We can alternatively write

$$f(x) = \frac{1}{2}(x-x^\star)^\intercal A(x-x^\star) + c + (x^\star)^\intercal A x^\star =: \frac{1}{2}(x-x^\star)^\intercal A(x-x^\star) + c'.$$

Note that $\nabla f(x) = A(x - x^\star)$.

Lemma 34. *$f(x)$ is L-smooth μ-strongly convex function if and only if $\mu I \preceq A \preceq L$.*

Now, suppose we employ a first-order method to minimize this function beginning at some $x_0 \in \mathbb{R}^n$.

Having learned $g_0 = A(x_0 - x^\star)$, we will form $x_1 \in x_0 + \operatorname{span}(g_0)$. Suppose $x_1 = x_0 + \alpha g_0$ for some $\alpha \geq 0$. Then,

$$\begin{aligned} g_1 &= A(x_1 - x^\star) \\ &= A(x_0 - x^\star + \alpha g_0) \\ &= (A + \alpha A^2)(x_0 - x^\star). \end{aligned}$$

Thus, after querying $\nabla f(x_1)$ we will have learned $A^2(x_0 - x^\star)$.

Repeating this logic, one can check that after T queries to the first-order oracle, we can learn

$$A(x_0 - x^\star), A^2(x_0 - x^\star), \ldots, A^{T-1}(x_0 - x^\star).$$

Now, we ask what $\bar{x}$ should we output to minimize $\frac{\|\bar{x} - x^\star\|}{\|x_0 - x^\star\|}$ in the worst-case? Equivalently, we ask how should we set $c_1, c_2, \ldots, c_{T-1}$ to minimize

$$\begin{aligned} &\max_{x_0 \in \mathbb{R}^n} \frac{\left\| x_0 - x^\star + \sum_{i=0}^{T-1} c_i A^i (x_0 - x^\star) \right\|}{\|x_0 - x^\star\|} \\ &= \max_{x_0 \in \mathbb{R}^n} \frac{\left\| \left(I + \sum_{i=0}^{T-1} c_i A^i \right) (x_0 - x^\star) \right\|}{\|x_0 - x^\star\|} \\ &= \max_{x_0 \in \mathbb{R}^n} \frac{\|p(A)(x_0 - x^\star)\|}{\|x_0 - x^\star\|} \\ &= \|p(A)\|_2 \end{aligned}$$

where $p(x) = 1 + c_1 x + c_2 x^2 + \cdots + c_{T-1} x^{T-1}$. In other words, we get to design a polynomial $p(x)$ whose constant term is 1 in order to minimize $\|p(A)\|_2$ in the worst-case over A.

It is not too hard to check that if $A = U \operatorname{Diag}(\lambda_i) U^\intercal$ is an eigenvalue decomposition of A, then

$$p(A) = U \begin{pmatrix} p(\lambda_1) & & & \\ & p(\lambda_2) & & \\ & & \ddots & \\ & & & p(\lambda_n) \end{pmatrix} U^\intercal.$$

Thus,

$$\|p(A)\|_2 \leq \max_{\lambda \in [\mu, L]} p(\lambda).$$

Our goal is now to pick $c_1, \ldots, c_{T-1}$ in order to minimize

$$\max_{\lambda \in [\mu, L]} p(\lambda).$$

Thankfully, this is a well-studied problem. The degree $T-1$-polynomial that minimizes this quantity is a shifted and scaled version of the

$(T-1)$th Chebyshev polynomial, $p_{T-1}(\lambda)$. These polynomials can be defined via the following recurrence:

$$\begin{aligned} p_0(\lambda) &:= 1 \\ \delta_1 &:= \frac{L-\mu}{L+\mu} \\ p_1(\lambda) &:= 1 - \frac{2}{L+\mu}\lambda \\ \delta_k &:= \frac{1}{2\frac{L+\mu}{L-\mu} - \delta_{k-1}} \quad \forall k \geq 2 \\ p_k(\lambda) &:= \frac{2\delta_k}{L-\mu}(L+\mu-2\lambda)p_{k-1}(\lambda) + \left(1 - \frac{2\delta_k(L+\mu)}{L-\mu}\right)p_{k-2}(\lambda) \quad \forall k \geq 2. \end{aligned}$$

It is not too important to know what this recurrence is, just that it satisfies the above recursive formula. This tell us that we can iteratively maintain $\bar{x}_t := p_t(A)(x_0 - x^\star) + x^\star$ as follows:

$$\begin{aligned} \bar{x}_0 &= x_0 \\ \bar{x}_1 &= x_0 - \frac{2}{L+\mu}\nabla f(x_0) \\ \bar{x}_k &= \frac{2\delta_k(L+\mu)}{L-\mu}\left[\bar{x}_{k-1} - \frac{2}{L+\mu}\nabla f(\bar{x}_{k-1})\right] + \left(1 - \frac{2\delta_k(L+\mu)}{L-\mu}\right)\bar{x}_{k-2} \quad \forall k \geq 2 \end{aligned}$$

This is gradient descent with a step size of $\frac{2}{L+\mu}$ plus a momentum term weighted by $\left(\frac{2\delta_k(L+\mu)}{L-\mu} - 1\right)$.

Theorem 19. *The iterates $\bar{x}_k$ satisfy*

$$\frac{\|\bar{x}_k - x^\star\|}{\|\bar{x}_0 - x^\star\|} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k.$$

Problems

1. This problem extends the accelerated gradient descent method for L-smooth convex functions and its analysis to other "smoothly-proxable" convex problems.

   Formally, consider a minimization problem of the form

   $$\min_{x \in \Omega} F(x)$$

   where $F : \mathbb{R}^n \to \mathbb{R}$ is an arbitrary function and $\Omega \subseteq \mathbb{R}^n$ is an arbitrary set. We say that

   $$\texttt{prox} : \mathbb{R}^n \to \Omega$$

   is a *smooth prox-oracle* for this problem if `prox` satisfies the following property: Given $y \in \mathbb{R}^n$, define $g(y) := L(y - \texttt{prox}(y))$. Then, for all $z \in \Omega$, it holds that

   $$F(\texttt{prox}(y)) \leq f(z) + \langle g(y), y - z \rangle - \frac{\|g(y)\|^2}{2L}. \tag{10.1}$$

   We will replace the gradient step in accelerated gradient descent with the prox oracle:

Algorithm 6 Accelerated gradient descent for smoothly proxable problems

Given $x_0 \in \mathbb{R}^d$, $F : \mathbb{R}^n \to \mathbb{R}$ and $\texttt{prox} : \mathbb{R}^n \to \Omega$

- Set $y_0 = x_0$ and $\lambda_{-1} = 1$
- For $t = 0, \ldots$

$$\begin{aligned}
\lambda_t &= \frac{1 + \sqrt{1 + 4\lambda_{t-1}^2}}{2} \\
\gamma_t &= \frac{\lambda_{t-1} - 1}{\lambda_t} \\
x_{t+1} &= \texttt{prox}(y_t) = y_t - \frac{1}{L} g(y_t) \\
y_{t+1} &= x_{t+1} + \gamma_t (x_{t+1} - x_t)
\end{aligned}$$

   (a) Modify the analysis of Theorem 17 to show that:

   Theorem 20. *Suppose $F : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subseteq \mathbb{R}^n$ and suppose $\texttt{prox} : \mathbb{R}^n \to \Omega$ is a smooth prox-oracle for $\min_{x \in \Omega} F(x)$. Furthermore, suppose F has minimizer $x^\star$ with minimum value $F^\star$. Then, it holds that*

   $$F(x_T) - F^\star = O\left(\frac{L \|x_0 - x^\star\|^2}{T^2} \right).$$

(b) Suppose $F : \mathbb{R}^n \to \mathbb{R}$ is an L-smooth convex function and $\Omega \subseteq \mathbb{R}^n$ is nonempty, closed, and convex. Define

$$\texttt{prox}(y) \coloneqq \arg\min_{x \in \Omega} \left\{ F(y) + \langle \nabla F(y), x - y \rangle + \frac{L}{2} \|x - y\|^2 \right\}.$$

Prove that this map is well-defined, is equal to

$$\texttt{prox}(y) = \Pi_\Omega \left(y - \frac{1}{L} \nabla F(y) \right),$$

and is a smooth prox-oracle for $\min_{x \in \Omega} F(x)$.

(c) Suppose $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ are L-smooth convex functions and define

$$F(x) \coloneqq \max_{i \in [k]} f_i(x).$$

Define

$$\texttt{prox}(y) \coloneqq \arg\min_{x \in \mathbb{R}^n} \max_{i \in [k]} \left\{ f_i(y) + \langle \nabla f_i(y), x - y \rangle + \frac{L}{2} \|x - y\|^2 \right\}.$$

Prove that this map is well-defined and is a smooth prox-oracle for $\min_{x \in \Omega} F(x)$.

11
Oracle lower bounds

In this lecture, we will prove that the convergence rates attained by subgradient descent for nonsmooth minimization and accelerated gradient descent for both smooth and smooth and strongly convex minimization are optimal up to constants.

11.1 Oracle complexity of nonsmooth convex minimization

Let $L, D > 0$ and define

$$\mathcal{P}_{L,D} \coloneqq \left\{ (f, x_0) : \begin{array}{l} f : \mathbb{R}^n \to \mathbb{R} \text{ is convex and } L\text{-Lipschitz with minimizer } x^\star \\ \|x_0 - x^\star\| \leq D \end{array} \right\}.$$

This is the family of problem instances that one may encounter in Lipschitz convex minimization.

We would like to argue that subgradient descent is the best possible algorithm within a given class of candidate algorithms. Our class of candidate algorithms will be any deterministic algorithm that interacts with the objective function f only through at most T first-order oracle calls

- For $t \geq 0, \dots, T-1$
  - Invoke the first-order oracle to receive: $f(x_t)$ and $g_t \in \partial f(x_t)$
  - Use a deterministic procedure applied to $x_0, \dots, x_t, f_0, \dots, f_t, g_0, \dots, g_t$ to construct x_{t+1}
- Output x_T

We call such methods *first-order methods*. Note that T is not directly related to computational complexity. For example, in our definition of a first-order method, the determinstic procedure is given unlimited computational power. The parameter T only controls the number of calls to the first-order oracle.

Remark 9. Note that the deterministic procedure for constructing x_{t+1} has access to *all* first-order information (not just first-order information at time t) so that first-order mechanisms like momentum or accelerated gradient descent can be written in this form. □

Our goal is to construct a worst-case function for 1-Lipschitz convex minimization where $x_0 = 0$ and $D = 1$. The general case follows by rescaling. For notational simplicity, we will also assume $T + 1 = 2^k$ for some k. The general case then follows by taking T' be the first power of 2 larger than or equal to T.

Let Σ denote the kth Hadamard matrix (scaled):

$$\Sigma := \frac{1}{\sqrt{2^k}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{\otimes k}.$$

This is an orthonormal matrix.

We construct f in response to the algorithm. Let I_0 denote the columns of Σ as a set and $f_0(x) := -\frac{\epsilon}{2}$ where ϵ will be fixed later. Note that I_0 has size $2^k = T + 1$. Pick arbitrary numbers $\epsilon/2 \geq \delta_0 > \delta_1 > \cdots > \delta_T > 0$.

For $t = 0, \ldots, T$, let

$$\sigma_t \in \operatorname*{arg\,max}_{\sigma \in I_t} |\langle \sigma, x_t \rangle|$$

and define $\omega_t = \operatorname{sign}(\langle \sigma_t, x_t \rangle)$. Update $I_{t+1} := I_t \setminus \{\sigma_t\}$ and define $f_{t+1}(x) := \max(f_t(x), \langle \omega_t \sigma_t, x \rangle + \delta_t)$. Then, return $f_{t+1}(x_t)$ and any subgradient $g_t \in \partial f_{t+1}(x_t)$.

We claim that the algorithm performs poorly on $f := f_{T+1}$:

First, note that $f(x) = f_{t+1}(x)$ on a neighborhood of x_t for all $t = 0, \ldots, T$. To see this, note that

$$f(x) = \max\left(f_{t+1}(x), \max_{j \in [t+1, T-1]} \left\{\left\langle \omega_j \sigma_j, x \right\rangle + \delta_j\right\}\right).$$

Next, for any $j \geq t$,

$$\begin{aligned} f_{t+1}(x_t) &\geq \langle \omega_t \sigma_t, x_t \rangle + \delta_t \\ &> \left\langle \omega_j \sigma_j, x_t \right\rangle + \delta_j. \end{aligned}$$

The second inequality follows as $\sigma_j \in I_t$ and $\delta_t > \delta_j$. Thus, $g_t \in \partial f(x_t)$.

From this, we also deduce that

$$f(x_T) \geq \langle w_T \sigma_T, x_T \rangle + \delta_T \geq 0.$$

On the other hand, as Σ is orthonormal, so too is

$$\begin{pmatrix} \omega_0 \sigma_0 & \ldots \omega_T \sigma_T \end{pmatrix}.$$

Thus, there exists some x with $\|x\| = 1$ so that $\langle \omega_i \sigma_i, x\rangle = -\frac{1}{\sqrt{2^k}}$ for all i. Then,

$$f(x) = \max\left\{-\frac{\epsilon}{2}, \frac{-1}{\sqrt{2^k}} + \delta_0\right\} \leq \max\left\{-\frac{\epsilon}{2}, \frac{-1}{\sqrt{2^k}} + \frac{\epsilon}{2}\right\}.$$

Setting $\epsilon = \frac{1}{\sqrt{2^k}}$, we conclude that

$$f(x_T) - f^\star \geq \frac{1}{2\sqrt{2^k}} = \frac{1}{2\sqrt{T+1}}.$$

Theorem 21. *Consider any deterministic method that makes at most T calls to a first-order oracle for f before outputting x_T. Then, there exists an L-Lipschitz convex function f with optimizer $x^\star$ and $\|x^\star\| \leq D$ so that*

$$f(x_T) - f^\star \geq \frac{LD}{2\sqrt{2T+1}}.$$

11.2 Oracle complexity for smooth convex minimization

A similar story holds for smooth and smooth and strongly convex minimization.

Theorem 22. *Consider any deterministic method that makes at most T calls to a first-order oracle for f before outputting x_T. Then, there exists an L-smooth convex function f with optimizer $x^\star$ and $\|x^\star\| \leq D$ so that*

$$f(x_T) - f^\star = \Omega\left(\frac{LD^2}{T^2}\right).$$

Theorem 23. *Consider any deterministic method that makes at most T calls to a first-order oracle for f before outputting x_T. Then, there exists an L-smooth and μ-strongly convex function f with optimizer $x^\star$ and $\|x^\star\| \leq D$ so that*

$$f(x_T) - f^\star = \Omega\left(\mu\left(1 - c\kappa^{-1/2}\right)^T D^2\right),$$

where c is an absolute constant.

We will prove just the smooth (non-strongly convex) statement. We will slightly cheat and make the assumption the following span-respecting assumption:

$$x_{t+1} \in x_0 + \operatorname{span}\{g_0, g_1, \ldots, g_t\}.$$

This is not a big deal and the same proof strategy can be made to work without this assumption using a “doubling trick.”

Proof of Theorem 22. Define the following sequence of matrices

$$A_k = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}$$

Let

$$f_k(x) = \frac{L}{4}\left(x^\intercal A_k x - x_1\right).$$

We will imagine running some first-order method on $f_N(x)$ where $N \gg T$.

Note that

$$\nabla f_N(x) = \frac{L}{4}\left(A_N x - e_1\right).$$

Thus, if x is supported on the first k coordinates, then $\nabla f_N(x)$ is supported on the first $k+1$ coordinates. By the span respecting assumption, if $x_0 = 0$, then x_T is supported on the first T coordinates.

Note that f_N on the first k coordinates is equal to f_T. Thus,

$$f_N(x_T) \geq \min_x f_T(x).$$

Our goal now is to understand the minimum value and minimizer of a general f_k. One can show that[1]

$$\min_x f_k(x) = \frac{L}{8}\left(-1 + \frac{1}{k+1}\right)$$
$$\left\|\arg\min_x f_k(x)\right\|^2 = O(k).$$

Thus, by setting $N = 2T$, we have that

$$f_N(x_T) - f_N^\star = \Omega\left(\frac{L}{T}\right)$$

despite $\|x^\star\|^2 \leq O(T)$. This matches the claimed lower bound:

$$f_N(x_T) - f_N^\star = \Omega\left(\frac{L\left\|x_0 - x^\star\right\|^2}{T^2}\right)$$

One can normalize[2] the constructed function f appropriately, to get a family of lower bounds with arbitrary $\|x_0 - x^\star\|$. ∎

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

[2] **Exercise**: Explain how this normalization works.

12

Performance Estimation Programming

This chapter introduces performance estimation programming (PEP). We begin by reviewing the convex conjugate of a function. This will be used in developing the PEP SDP.

12.1 The convex conjugate

Definition 34. Let $f : \mathbb{R}^n \to [-\infty \to \infty]$. The convex conjugate f^* of f is the extended-valued function $f^* : \mathbb{R}^n \to [-\infty, +\infty]$ given by

$$f^*(y) \coloneqq \sup_{x \in \mathbb{R}^n} \left\{ \langle y, x \rangle - f(x) \right\}. \qquad \square$$

Remark 10. How should one think about the convex conjugate? Up to some convex analysis technicalities (that we will formalize soon), we can think of any convex function as a supremum over affine functions (possibly infinitely many). We can parameterize an affine function $x \mapsto \langle y, x \rangle - c$ by some $y \in \mathbb{R}^n$ and some $c \in \mathbb{R}$. Thus, there exists some function $c : \mathbb{R}^n \to \mathbb{R}$ so that

$$f(x) = \sup_{y \in \mathbb{R}^n} \langle y, x \rangle - c(y).$$

This function c is "the definition" of f^*. Furthermore, you may have noticed there is a nice symmetry that goes from f to f^* and back. This intuition is basically all true except for some technicalities that we now make formal. $\square$

Definition 35. An extended-valued function $f : \mathbb{R}^n \to [-\infty, \infty]$ is closed if the epigraph

$$\{(x, t) : t \geq f(x)\}$$

is closed. We say f is convex if the epigraph is convex. We say f is proper if the epigraph is nonempty. $\square$

Lemma 35. *Let $f : \mathbb{R}^n \to [-\infty, \infty]$. Then f^* is a* closed and convex *function.*

Proof. Note that the epigraph is given by

$$\bigcap_{x \in \mathbb{R}^n} \{(y,t) : t \geq \langle y, x \rangle - f(x)\},$$

where the set in the intersection is closed and convex for each $x \in \mathbb{R}^n$. Recalling that an arbitrary intersection of closed convex sets is closed and convex proves the lemma. ■

Lemma 36 (Fenchel-Young)**.** *Suppose $f : \mathbb{R}^n \to [-\infty, \infty]$ and $x, y \in \mathbb{R}^n$. Then,*

$$f(x) + f^*(y) \geq \langle x, y \rangle .$$

Proof. By definition,

$$f^*(y) = \sup_{z \in \mathbb{R}^n} \langle z, y \rangle - f(z) \geq \langle x, y \rangle - f(x).$$ ■

Lemma 37. *Suppose $f : \mathbb{R}^n \to [-\infty, \infty]$ is closed and convex. Then,*

$$\ell \in \partial f(x) \iff x \in \partial f^*(\ell) \iff x \in \arg\max_{\tilde{x}} \langle \ell, \tilde{x} \rangle - f(\tilde{x})$$
$$\iff \ell \in \arg\max_{\tilde{\ell}} \left\langle \tilde{\ell}, x \right\rangle - f^*(x) \iff f(x) + f^*(\ell) = \langle x, \ell \rangle .$$

Proof. Note $\bar{\ell} \in \partial f(\bar{x})$ if and only if 0 is in the subgradient of

$$f(x) - \left\langle \bar{\ell}, x \right\rangle$$

at $\bar{x}$ if and only if

$$f^*(\bar{\ell}) = \sup_x \left\langle \bar{\ell}, x \right\rangle - f(x) = \left\langle \bar{\ell}, \bar{x} \right\rangle - f(\bar{x}).$$

if and only if

$$f(\bar{x}) + f^*(\bar{\ell}) = \left\langle \bar{\ell}, \bar{x} \right\rangle .$$

Reversing the roles completes the proof. ■

Lemma 38. *Suppose $f : \mathbb{R}^n \to [-\infty, \infty]$. Then, $f^{**}(x) \leq f(x)$ for all $x \in \mathbb{R}^n$.*

Proof. Let $x, y \in \mathbb{R}^n$. Then,

$$f^*(y) \geq \langle y, x \rangle - f(x)$$

Thus,

$$f(x) \geq \langle y, x \rangle - f^*(y)$$

Taking the supremum of the RHS in y gives

$$f(x) \geq \sup_{y \in \mathbb{R}^n} \{\langle y, x \rangle - f^*(y)\} = f^{**}(y).$$ ■

Lemma 39. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex. Then, $f^{**} = f$.*[1]

[1] This can be extended to the setting of an extended-real-valued function f as long as it is proper, closed, and convex.

Proof. We have shown that $f^{**} \leq f$ pointwise. Now, for the sake of contradiction, suppose there exists $\bar{x}$ so that $f^{**}(\bar{x}) < f(\bar{x})$. As

$$\mathcal{S} = \{(x, t) : t \geq f(x)\}$$

is a closed convex set, and $(\bar{x}, f^{**}(\bar{x}))$ is not in this set, by the *strict* hyperplane separation theorem, there exists $(a, b) \in \mathbb{R}^{1+n}$ nonzero so that

$$\langle a, \bar{x} \rangle - b f^{**}(\bar{x}) > \sup_{(x,t) \in \mathcal{S}} \langle a, x \rangle - bt.$$

We have that $b \geq 0$ (else send $t \to \infty$ for a contradiction) and that $b \neq 0$ (else send $x \to \infty a$).

Thus, we assume WLOG that $b = 1$ and get

$$\langle a, \bar{x} \rangle - f^{**}(\bar{x}) > \sup_{x \in \mathbb{R}^n} \langle a, x \rangle - f(x) = f^*(a).$$

This contradicts Fenchel's inequality:

$$f^*(a) + f^{**}(\bar{x}) \geq \langle a, \bar{x} \rangle .$$ ■

Lemma 40. *Suppose $f, g : \mathbb{R}^n \to [-\infty, +\infty]$ are such that $f \geq g$ pointwise. Then, $f^* \leq g^*$ pointwise.*

Proof.

$$\begin{aligned} f^*(y) &:= \sup_{x \in \mathbb{R}^n} \langle x, y \rangle - f(x) \\ &\leq \sup_{x \in \mathbb{R}^n} \langle x, y \rangle - g(x) \\ &= g^*(y). \end{aligned}$$ ■

Corollary 8. *Suppose $f, g : \mathbb{R}^n \to \mathbb{R}$ and g is convex. If $f \geq g$ pointwise, then $f^{**} \geq g$ pointwise.*

Proof. Assume $f \geq g$ pointwise. By previous lemma, $f^* \leq g^*$ pointwise. Applying the lemma once more gives, $f^{**} \geq g^{**}$ pointwise. Finally, note that as g is a real-valued convex function, $g^{**} = g$. ■

In other words, given $f : \mathbb{R}^n \to \mathbb{R}$, the function $f^{**} : \mathbb{R}^n \to \mathbb{R}$ is the pointwise *largest* convex function laying below f. To make this more precise,

$$f^{**}(x) = \max_g \left\{ g(x) : \begin{array}{l} g : \mathbb{R}^n \to \mathbb{R} \text{ is convex} \\ f \geq g \text{ pointwise} \end{array} \right\}.$$

Lemma 41. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly convex, then $f^* : \mathbb{R}^n \to \mathbb{R}$ is a $\frac{1}{\mu}$-smooth convex function.*

Proof. First, note that

$$\sup_{x \in \mathbb{R}^n} \{\langle \ell, x\rangle - f(x)\}$$

has finite value as f is strongly convex.

In fact, this maximization problem has a unique maximizer. By previous lemma, the subdifferential of f^* at any point is unique so that f^* is differentiable.

By μ-strong convexity, we have that for any $x, x' \in \mathbb{R}^n$ and $\ell \in \partial f(x)$ and $\ell' \in \partial f(x')$, that

$$\left\langle \ell - \ell', x - x' \right\rangle \geq \mu \left\| x - x' \right\|^2.$$

Recognizing that $x = \nabla f^*(\ell)$ and $x' = \nabla f^*(\ell')$ gives us

$$\left\langle \ell - \ell', \nabla f^*(\ell) - \nabla f^*(\ell') \right\rangle \geq \mu \left\| \nabla f^*(\ell) - \nabla f^*(\ell') \right\|^2.$$

By Cauchy-Schwarz,

$$\left\| \nabla f^*(\ell) - \nabla f^*(\ell') \right\| \leq \frac{1}{\mu} \left\| \ell - \ell' \right\|. \qquad \blacksquare$$

12.2 *PEP and interpolation*

For concreteness, consider the following first-order method for minimizing a 1-smooth convex function $f : \mathbb{R}^n \to \mathbb{R}$:

$$\begin{aligned} x_0 &\text{ is given satisfying } \|x_0 - x_\star\|^2 \leq 1 \\ x_1 &= x_0 - h_{1,0}\nabla f(x_0) \\ x_2 &= x_1 - h_{2,0}\nabla f(x_0) - h_{2,1}\nabla f(x_1) \\ x_k &= x_{k-1} - \sum_{i=0}^{k-1} h_{k,i}\nabla f(x_i) \qquad \forall k = 1, \ldots, T \end{aligned}$$

This first-order method is defined by a lower triangular matrix $h \in \mathbb{R}^{T\times T}$ where the columns and rows are both indexed by $0, \ldots, T-1$.

Now, we will attempt to find the worst-case function

$$(\text{PEP}) = \max_{f, x_0, x_T, x_\star} \left\{ f(x_k) - f(x_\star) : \begin{array}{l} f : \mathbb{R}^n \to \mathbb{R} \text{ is convex and 1-smooth} \\ \nabla f(x_\star) = 0 \\ \|x_0 - x_\star\|^2 \leq 1 \\ x_T \text{ is produced by FOM starting at } x_0 \end{array} \right\}.$$

On the surface, this is an infinite-dimensional nonconvex optimization problem. We will rewrite this problem in several ways to end up with a finite-dimensional convex optimization problem (an SDP).

The first step is to reduce optimizing over f to only optimizing over the first-order data $\mathcal{D} = \{(f_\star, g_\star, x_\star), (f_0, g_0, x_0), \ldots, (f_T, g_T, x_T)\}$. For notational convenience, let $\mathcal{I} = \{\star, 0, 1, \ldots, T\}$. Then, the above is equal to

$$(\text{PEP}) = \max_{\substack{f_\star, f_0, \ldots, f_T \in \mathbb{R} \\ g_\star, g_0, \ldots, g_T \in \mathbb{R}^n \\ x_\star, x_0, \ldots, x_T \in \mathbb{R}^n}} \left\{ f_T - f_\star : \begin{array}{l} \exists f : \mathbb{R}^n \to \mathbb{R} \text{ convex and 1-smooth interpolating } \{(f_i, g_i, x_i)\}_{i \in \mathcal{I}} \\ g_\star = 0 \\ \|x_0 - x_\star\|^2 \leq 1 \\ x_T \text{ is produced by FOM given } \mathcal{D} \text{ starting at } x_0 \end{array} \right\}.$$

Here, we say that:

Definition 36. A function $f : \mathbb{R}^n \to \mathbb{R}$ interpolates $\{(f_i, g_i, x_i)\}_{i \in \mathcal{I}}$ if

$$f(x_i) = f_i \qquad \text{and} \qquad \nabla f(x_i) = g_i \qquad \forall i \in \mathcal{I}. \qquad \square$$

Theorem 24. *Fix a set $\mathcal{D} \coloneqq \{(f_i, g_i, x_i)\}_{i \in \mathcal{I}}$. There exists a convex 1-smooth function f interpolating $\mathcal{D}$ if and only if*

$$f_i \geq f_j + \langle g_j, x_i - x_j \rangle + \frac{1}{2} \|g_i - g_j\|^2 \qquad \forall i, j \in \mathcal{I}.$$

Proof. In the forward direction, suppose there exists a convex 1-smooth function for which

$$f(x_i) = f_i \qquad \text{and} \qquad \nabla f(x_i) = g_i \qquad \forall i \in \mathcal{I}.$$

Fix an arbitrary $i, j \in \mathcal{I}$. Set

$$h(x) = f(x) - \langle \nabla f(x_j), x - x_j \rangle .$$

Note that h is a convex 1-smooth function for which $\nabla h(x_j) = 0$. Thus,

$$\begin{aligned} h(x_j) &= \min_{x \in \mathbb{R}^n} h(x) \\ &\leq \min_{x \in \mathbb{R}^n} h(x_i) + \langle \nabla h(x_i), x - x_i \rangle + \frac{1}{2} \|x - x_i\|^2 \\ &= h(x_i) - \frac{1}{2} \|\nabla h(x_i)\|^2 . \end{aligned}$$

Expanding the definition of h gives us

$$f(x_j) \leq f(x_i) - \langle \nabla f(x_j), x_i - x_j \rangle - \frac{1}{2} \|\nabla f(x_i) - \nabla f(x_j)\|^2 .$$

In the reverse direction, our goal is: Given $\mathcal{D} \coloneqq \{(f_i, g_i, x_i)\}_{i \in \mathcal{I}}$ satisfying

$$f_i \geq f_j + \langle g_j, x_i - x_j \rangle + \frac{1}{2} \|g_i - g_j\|^2 \qquad \forall i, j \in \mathcal{I},$$

construct a convex 1-smooth function f interpolating $\mathcal{D}$.

We will do this as follows. Define

$$h(x) = \min_{i \in \mathcal{I}} f_i + \langle g_i, x - x_i \rangle + \frac{1}{2} \|x - x_i\|^2$$

and $f := h^{**}$.

There are two things to check. First, we must check that f is 1-smooth. It suffices to check that h^* is 1-strongly convex. Second, we must check that $f(x_i) = f_i$ and $\nabla f(x_i) = g_i$.

For the first assertion, we compute

$$\begin{aligned}
h^*(y) &= \sup_{x\in\mathbb{R}^n} \langle y, x\rangle - \left[\min_{i\in\mathcal{I}} f_i + \langle g_i, x - x_i\rangle + \frac{1}{2}\|x - x_i\|^2\right] \\
&= \max_{i\in\mathcal{I}} \sup_{x\in\mathbb{R}^n} \left(\langle y, x\rangle - f_i - \langle g_i, x - x_i\rangle - \frac{1}{2}\|x - x_i\|^2\right) \\
&= \max_{i\in\mathcal{I}} \sup_{x\in\mathbb{R}^n} \left(\langle y, x_i\rangle - f_i + \langle y - g_i, x - x_i\rangle - \frac{1}{2}\|x - x_i\|^2\right) \\
&= \max_{i\in\mathcal{I}} \left\{\langle y, x_i\rangle - f_i + \frac{\|y - g_i\|^2}{2}\right\} \\
&=: \max_{i\in\mathcal{I}} s_i(y).
\end{aligned}$$

Thus, $h^*(y)$ is the pointwise maximum of 1-strongly convex functions so is also 1-strongly convex. We conclude that $f(x)$ is 1-smooth.

Now, suppose $i \in \mathcal{I}$. Our goal is to check that $f(x_i) = f_i$ and $\nabla f(x_i) = g_i$. We claim it suffices to check that

$$i \in \arg\max_{j\in\mathcal{I}} s_j(g_i).$$

Indeed, supposing this is true, then

$$x_i = \nabla s_i(g_i) \in \partial h^*(g_i) \iff g_i = \nabla f(x_i).$$

Furthermore, g_i maximizes $\langle y, x_i\rangle - h^*(y)$. Thus, by definition, $f(x_i) = \langle g_i, x_i\rangle - s_i(g_i) = f_i$.

This condition is equivalent to saying that for all $i, j \in \mathcal{I}$,

$$\begin{aligned}
& s_i(g_i) \geq s_j(g_i) \\
\iff\quad & \langle g_i, x_i\rangle - f_i + \frac{\|g_i - g_i\|^2}{2} \geq \langle g_i, x_j\rangle - f_j + \frac{\|g_i - g_j\|^2}{2} \\
\iff\quad & f_j \geq f_i + \langle g_i, x_j - x_i\rangle + \frac{\|g_i - g_j\|^2}{2}.
\end{aligned}$$

■

With this interpolation theorem in hand, we may now rewrite the PEP as

$$(\text{PEP}) = \max_{\substack{f_\star, f_0,\dots,f_T\in\mathbb{R}\\ g_\star, g_0,\dots,g_T\in\mathbb{R}^n\\ x_\star, x_0,\dots,x_T\in\mathbb{R}^n}} \left\{ f_T - f_\star : \begin{array}{l} f_j \geq f_i + \langle g_i, x_j - x_i\rangle + \frac{\|g_i - g_j\|^2}{2} \quad \forall i, j \in \mathcal{I} \\ g_\star = 0 \\ \|x_0 - x_\star\|^2 \leq 1 \\ x_t = x_{t-1} + \sum_{i=0}^{t-1} H_{t,i} g_i \quad \forall t = 1, \dots, T \end{array} \right\}.$$

This is now a finite-dimensional nonconvex problem. We will fix $x^\star = 0$ without loss of generality and get rid of the optimization on

x_is as they are completely determined by the g_i. Instead, we will treat $x_1, \ldots, x_T$ as *linear functions* in x_0 and $g_0, \ldots, g_T$. We can arrange

$$G = \begin{pmatrix} x_0 & g_0 & \ldots & g_{T-1} & g_T \end{pmatrix}.$$

Thus,

$$= \max_{\substack{f_\star, f_0, \ldots, f_T \in \mathbb{R} \\ G \in \mathbb{R}^{n \times (T+2)}}} \left\{ f_T - f_\star : \begin{array}{ll} f_j \geq f_i + \langle g_i, x_j(G) - x_i(G) \rangle + \frac{\|g_i - g_j\|^2}{2} & \forall i, j \in \mathcal{I} \\ \|x_0\|^2 \leq 1 & \end{array} \right\}.$$

Now, we observe that the dependence on the columns of G in this problem are all quadratic, i.e., the constraints are linear in $(f_\star, \ldots, f_T)$ and $Q \coloneqq G^\intercal G \in \mathbb{S}^{T+2}$.

Thus, there exist $M_{i,j} \in \mathbb{S}^{T+2}$ depending on H so that

$$\langle M_{i,j}, G^\intercal G \rangle = \langle g_i, x_j(G) - x_i(G) \rangle + \frac{\|g_i - g_j\|^2}{2}.$$

Then,

$$(\text{PEP}) \leq \max_{\substack{f_\star, f_0, \ldots, f_T \in \mathbb{R} \\ Q \in \mathbb{S}^{T+2}}} \left\{ f_T - f_\star : \begin{array}{ll} f_j \geq f_i + \langle Q, M_{i,j} \rangle & \forall i, j \in \mathcal{I} \\ Q_{1,1} \leq 1 & \\ Q \succeq 0 & \end{array} \right\}.$$

This relaxation is exact if $d \geq T + 2$ as given any $Q \succeq 0$, we can take any matrix satisfying $G^\intercal G = Q$ (which exists as Q is positive semidefinite) and set $x_0, g_0, \ldots, g_T$ to be the columns of G.

13
Mirror descent

In this lecture, we will discuss mirror descent. This is an extension of the projected subgradient method to nonsmooth non-Euclidean settings.[1]

[1] There are also extensions of gradient descent and accelerated gradient descent to smooth non-Euclidean settings.

13.1 Mirror descent setup and algorithm

In Mirror Descent, we will assume we have the following setup:

- A **norm** $\|\cdot\|$
- Problem **domain** $\mathcal{X} \subseteq \mathbb{R}^n$ nonempty, closed, convex, with nonempty interior[2]
- **Objective function** $f : \mathcal{X} \to \mathbb{R}$ is closed and convex, i.e.,

$$\left\{ (x,t) : \begin{array}{l} x \in \mathcal{X} \\ f(x) \le t \end{array} \right\}$$

  is closed and convex and subdifferentiable on $\mathcal{X}$, i.e., $\partial f(x)$ is nonempty for all $x \in \mathcal{X}$. Further, assume that for all $x \in \mathcal{X}$, we can algorithmically find $g \in \partial f(x)$ with

$$\|g\|_* \le L.$$

  For example, if f is L-Lipschitz and defined on an open neighborhood of $\mathcal{X}$, then any subgradient suffices. In general, if $f : \mathcal{X} \to \mathbb{R}$ is only defined on $\mathcal{X}$ and is L-Lipschitz, then any subgradient suffices on $\mathrm{int}(\mathcal{X})$, but some care will need to be taken at $\mathrm{bd}(\mathcal{X})$.
- A **distance generating function** $\omega : \mathcal{X} \to \mathbb{R}$ that is closed and convex. We assume that ω is differentiable over $\mathrm{dom}(\partial(\omega))$ and is 1-strongly convex on $\mathcal{X}$, i.e., for all $x, y \in \mathcal{X}$ and $\alpha \in [0,1]$

$$\omega((1-\alpha)x + \alpha y) \le (1-\alpha)\omega(x) + \alpha\omega(y) - \frac{1}{2}\alpha(1-\alpha)\|x-y\|^2 .$$

[2] Nonempty interior is not really required but makes the exposition easier

Remark 11. Recall that the subgradient of a convex function is always nonempty within the interior of the domain (here, $\mathcal{X}$). Thus, the assumption that ω is differentiable on $\mathrm{dom}(\partial(\omega))$ implies that ω is differentiable on $\mathrm{int}(\mathcal{X})$. In some cases $\partial(\omega) = \mathcal{X}$, but this is not always the case. For example, consider $\mathcal{X} = [0,1]$ and $\omega(x) = -\sqrt{x}$. Then, ω is a closed convex function. The subgradient $\partial\omega$ is defined for all $x \in (0,1]$ but is not defined at $0 \in \mathcal{X}$. In this case, ω is differentiable on $\mathrm{dom}(\partial\omega) = (0,1]$. □

Recall that the basic step in the projected subgradient method is

$$x_{k+1} = \operatorname*{arg\,min}_{x\in\Omega} \left\{ f(x_k) + \langle g_k, x - x_k\rangle + \frac{1}{2\eta_k} \|x - x_k\|_2^2 \right\},$$

where $g_k \in \partial f(x_k)$. In the non-Euclidean setting, we will want to replace $\frac{1}{2}\|x - x_k\|_2^2$ with something more specific to the norm $\|\cdot\|$. We will do so with what is called a Bregman divergence (to be defined below) $D(y||x)$. Then, the basic step in mirror descent will be of the form

$$x_{k+1} = \operatorname*{arg\,min}_{x\in\Omega} \left\{ f(x_k) + \langle g_k, x - x_k\rangle + \frac{1}{\eta_k} D(x||x_k) \right\}.$$

Definition 37. For $y \in \Omega$ and $x \in \mathrm{dom}(\partial\omega)$, the *Bregman divergence* is

$$D(y||x) \coloneqq \omega(y) - (\omega(x) + \langle \nabla\omega(x), y - x\rangle).$$ □

Note that for all $y \in \Omega$ and $x \in \mathrm{dom}(\partial\omega)$, $D(y||x) \geq \frac{1}{2}\|x - y\|^2$.

Example 19. Example mirror setups and their Bregman divergences:

- Take $\|\cdot\|$ to be the Euclidean norm and define $\omega(x) = \frac{1}{2}\|x - x_0\|^2$. Then,

  $$\begin{aligned} D(y||x) &= \frac{1}{2}\|y - x_0\|^2 - \left(\frac{1}{2}\|x - x_0\|^2 + \langle x - x_0, y - x\rangle\right) \\ &= \frac{1}{2}\|x - y\|^2. \end{aligned}$$

  Thus, the mirror descent step with this mirror setup is

  $$x_{k+1} = \operatorname*{arg\,min}_{x\in\Omega} \left\{ f(x_k) + \langle g_k, x - x_k\rangle + \frac{1}{2\eta_k} \|x - x_k\|^2 \right\}$$

  and recovers the projected subgradient step.

- Take $\|\cdot\|$ to be the ℓ_1-norm, $\mathcal{X} = \mathbb{R}^n_+$ and

  $$\omega(x) \coloneqq \sum_{i=1}^n x_i \log(x_i)$$

  where we take the convention $0\log 0 \coloneqq 0$. We will see in the homework that this is a 1-strongly convex function w.r.t. the ℓ_1 norm, compute the Bregman divergence, and give a closed-form solution to the mirror descent step.

- A suitable mirror-descent setups for the ℓ_p norms $p \in (1,2]$ is $\omega(x) := \frac{\alpha_p}{2}\|x - x_0\|_p^2$ for any base point x_0. The normalizing constant α_p is set to make ω 1-strongly convex. $\square$

Algorithm 7 Mirror Descent

Given mirror setup, initial $x_0 \in \mathrm{dom}(\partial(\omega))$, step lengths $\eta_0, \dots, \eta_T > 0$, time horizon T

- For $t = 0, \dots T-1$
  - Let $g_t \in \partial(f(x_t))$
  - Set $x_{t+1} = \arg\min_{x\in\Omega} \left\{ f(x_t) + \langle g_t, x - x_t\rangle + \frac{1}{\eta_t} D(x\|x_t) \right\}$
- Return $\bar{x} := \frac{\sum_{t=0}^{T} \eta_t x_t}{\sum_{t=0}^{T} \eta_t}$.

Remark 12. Why do we care about the mirror descent algorithm? The guarantees for mirror descent will look quite similar to the guarantees for subgradient descent:

$$f(\bar{x}_T) - f^\star \le O\left(\frac{L \cdot (\text{some distance measure})}{\sqrt{T}}\right).$$

The main advantage is that, if the objective function has a geometry which is "non-Euclidean", then we may be able to drastically reduce the Lipschitz constant L by working in a more appropriate norm. For example consider the function

$$f(x) = \|x - x_0\|_1 .$$

This function is $\sqrt{n}$-Lipschitz in the Euclidean norm so that the guarantees of the projected subgradient method would include $\sqrt{n}$. On the other hand, this function is 1-Lipschitz in the ℓ_1 norm. $\square$

It will be useful to simplify the objective function in the definition of x_{t+1}. An equivalent definition of x_{t+1} is

$$x_{t+1} = \arg\min_{x\in\mathcal{X}} \left\{ \langle \eta_t g_t - \nabla\omega(x_t), x\rangle + \omega(x) \right\}.$$

Lemma 42. *Mirror descent is well-defined, i.e., for all $\eta > 0$, $\bar{x} \in \mathrm{dom}(\partial(\omega))$, and $\bar{g} \in \partial f(\bar{x})$, then*

$$\arg\min_{x\in\mathcal{X}} \left\{ \langle \eta\bar{g} - \nabla\omega(\bar{x}), x\rangle + \omega(x) \right\} = \{\tilde{x}\},$$

for some $\tilde{x} \in \mathrm{dom}(\partial(\omega))$.

Proof. This objective function is 1-strongly convex so that the minimizer exists and is unique. Call this minimizer $\bar{x} \in \Omega$. Now, by first-order optimality, we deduce that

$$0 \in \eta \bar{g} - \nabla\omega(\bar{x}) + \partial\omega(\tilde{x}).$$

In other words, $\nabla\omega(\bar{x}) - \eta\bar{g} \in \partial\omega(\tilde{x})$ so that $\tilde{x} \in \mathrm{dom}(\partial\omega)$.[3] ■

[3] Note: this does *not* imply that $\nabla\omega(\tilde{x}) = \nabla\omega(\bar{x}) - \eta\bar{g}$. Specifically, if $\tilde{x}$ is on the boundary of $\mathcal{X}$, then it may be the case that $\{\nabla\omega(\tilde{x})\} \subsetneq \partial\omega(\tilde{x})$.

We now know that $x_t \in \mathrm{dom}(\partial(\omega))$ for all $t = 0, \ldots, T$.

13.2 Convergence analysis

The following lemma follows by simply expanding definitions and is omitted.

Lemma 43 (Three point identity)**.** *Suppose $x, y \in \mathrm{dom}(\partial\omega)$ and $z \in \Omega$. Then,*

$$D(z||x) - D(z||y) - D(y||x) = \langle \nabla\omega(y) - \nabla\omega(x), z - y\rangle .$$

We apply the first order optimality condition to the mirror descent step: let $t \geq 0$, then

$$\langle \nabla\omega(x_{t+1}) + \eta_t g_t - \nabla\omega(x_t), y - x_{t+1}\rangle \geq 0, \qquad \forall y \in \mathcal{X},$$

rearranging,

$$\langle \eta_t g_t, x_{t+1} - y\rangle \leq \langle \nabla\omega(x_{t+1}) - \nabla\omega(x_t), y - x_{t+1}\rangle , \qquad \forall y \in X.$$

Applying the three point identity,

$$\langle \eta_t g_t, x_{t+1} - y\rangle \leq D(y||x_t) - D(y||x_{t+1}) - D(x_{t+1}||x_t), \qquad \forall y \in X.$$

This is going to give us an opportunity to create a telescoping sum (take $y = x^\star$). Additionally, the final term is negative!

Theorem 25. *Suppose* $\inf_{x\in\mathcal{X}} f(x)$ *has a minimizer $x^\star$ with optimal value $f^\star$ and $D(x^\star||x_0) \leq \frac{R^2}{2}$. The Mirror Descent method guarantees*

$$\begin{aligned} f(\bar{x}) - f^\star &\leq \frac{R^2}{2H} + \frac{\sum_{t=0}^T \eta_t^2 \|g_t\|_*^2}{2H} \\ &\leq \frac{R^2}{2H} + \frac{L^2 \sum_{t=0}^T \eta_t^2}{2H}, \end{aligned}$$

where $H = \sum_{t=0}^T \eta_t$.

Proof. For the sake of the proof, we will imagine simulating one additional step of the method so that x_{T+1} and y_{T+1} are also defined.

Let $t \in [0, T]$. By the previous inequality, we know that

$$\langle \eta_t g_t, x_{t+1} - x^\star\rangle \leq D(x^\star||x_t) - D(x^\star||x_{t+1}) - D(x_{t+1}||x_t).$$

Thus,

$$\begin{aligned}
\langle \eta_t g_t, x_t - x^\star \rangle &\le \langle \eta_t g_t, x_t - x_{t+1} \rangle + D(x^\star || x_t) - D(x^\star || x_{t+1}) - D(x_{t+1} || x_t) \\
&\le \eta_t \|g_t\|_* \|x_t - x_{t+1}\| + D(x^\star || x_t) - D(x^\star || x_{t+1}) - \frac{1}{2}\|x_{t+1} - x_t\|^2 \\
&\le D(x^\star || x_t) - D(x^\star || x_{t+1}) + \max_\alpha \left(\eta_t \|g_t\|_* \alpha - \frac{1}{2}\alpha^2 \right) \\
&= D(x^\star || x_t) - D(x^\star || x_{t+1}) + \frac{\eta_t^2 \|g_t\|_*^2}{2}.
\end{aligned}$$

Now, we also have by the definition of the subgradient that

$$\begin{aligned}
\eta_t (f(x_t) - f^*) &\le \langle \eta_t g_t, x_t - x^\star \rangle \\
&\le D(x^\star || x_t) - D(x^\star || x_{t+1}) + \frac{\eta_t^2 \|g_t\|_*^2}{2}.
\end{aligned}$$

Let $H = \sum_{t=0}^T \eta_t$. We will take a $\frac{1}{H}$-weighted combination of the above inequalities for $t = 0, \dots, T$ to get

$$\begin{aligned}
\sum_{t=0}^T \frac{\eta_t}{H}(f(x_t) - f^\star) &\le \frac{D(x^\star || x_0) - D(x^* || x_{T+1})}{H} + \frac{\sum_{t=0}^T \eta_t^2 \|g_t\|_*^2}{2H} \\
&\le \frac{R^2}{H} + \frac{\sum_{t=0}^T \eta_t^2 \|g_t\|_*^2}{2H} \\
&\le \frac{R^2}{H} + \frac{L^2 \sum_{t=0}^T \eta_t^2}{2H}.
\end{aligned}$$

The fact that $f(\bar{x}) - f^\star$ is at most the LHS follows from convexity. ■

The following corollaries from this base guarantee are proved in exactly the same way as were proved for projected subgradient descent:

Corollary 9. *Suppose $\eta_t > 0$ satisfies $\sum_{t=0}^\infty \eta_t = \infty$ and $\sum_{t=0}^\infty \eta_t < \infty$. Then, $f(\bar{x}_T) - f^\star \to 0$.*

Corollary 10. *Taking $\eta_t = \frac{R}{\|g_t\|_* \sqrt{t+1}}$ gives*

$$f(\bar{x}_T) - f^\star \le \frac{LR(2 + \ln(T+1))}{2(\sqrt{T+2} - 1)}$$

Corollary 11. *Taking $\eta = \frac{R}{L\sqrt{T+1}}$ guarantees*

$$f(\bar{x}) - f^\star \le \frac{LR}{\sqrt{T+1}}.$$

In particular, it achieves an ϵ suboptimal solution in $O\left(\left(\frac{LR}{\epsilon}\right)^2\right)$ iterations.

14
Frank–Wolfe / Conditional Gradient Descent

This lecture studies the Frank–Wolfe algorithm (also known as Conditional Gradient Descent) for smooth convex minimization[1] over a compact convex set $\mathcal{X} \subseteq \mathbb{R}^n$ in an arbitrary norm $\|\cdot\|$:

$$\min_{x \in \mathcal{X}} f(x).$$

[1] Homework 4 will contain an extension to smooth strongly convex objective functions and strongly convex sets.

One algorithm we have already seen for problems of this form (for the Euclidean norm) is the accelerated projected gradient descent method (Homework 3 Problem 4.b). That algorithm achieves a $O\left(\frac{LD^2}{T^2}\right)$ convergence rate, which we have also shown is optimal among first-order methods. In each iteration, the accelerated projected gradient method requires a projection:

$$x_{t+1} = \Pi_{\mathcal{X}}\left(y_t - \frac{1}{L}\nabla f(y_t)\right).$$

We assumed that this projection could be done cheaply and deferred its computation to a projection oracle. In some applications, however, this projection is expensive to compute. For example, if $\mathcal{X} = \left\{X \in \mathbb{S}^n_+ : \operatorname{tr}(X) \leq 1\right\}$ is the set of positive semidefinite matrices with bounded trace, then this projection requires performing an SVD (practically $O(n^3)$ time).

The Frank–Wolfe method, which we will study in this lecture, has a worse convergence rate $O\left(\frac{LD^2}{T}\right)$, however will not require a projection in each iteration. Instead, Frank–Wolfe will only need to "access" Ω through a linear minimization oracle: Given $\ell \in \mathbb{R}^n$, find a minimizer of

$$\min_{x \in \Omega} \langle \ell, x \rangle .$$

For example, if $\mathcal{X} = \left\{X \in \mathbb{S}^n_+ : \operatorname{tr}(X) \leq 1\right\}$ is the set of positive semidefinite matrices with bounded trace, then linear minimization requires computing just a single leading eigenvector (practically $O(n^2)$ time or even smaller).

Algorithm 8 Frank–Wolfe

Given $x_0 \in \mathcal{X}$ and smooth convex function $f : \mathbb{R}^n \to \mathbb{R}$ and step-sizes $\eta_0, \dots, \eta_{T-1} \in [0, 1]$.

- For $t = 0, \dots, T-1$
  - $y_t \in \arg\min_{y \in \mathcal{X}} \langle \nabla f(x_t), y \rangle$
  - $x_{t+1} = (1 - \eta_t)x_t + \eta_t y_t$

Let $x^\star$ be a minimizer of $\min_{x \in \mathcal{X}} f(x)$. We will bound the primal gap $f(x_t) - f(x^\star)$ by what is known as the Wolfe–gap (the last expression below):

$$f(x_t) - f^\star \le \langle \nabla f(x_t), x_t - x^\star \rangle \le \max_{y \in \mathcal{X}} \langle \nabla f(x_t), x_t - y \rangle .$$

Note that by definition, y_t is the maximizer of the Wolfe–gap. So

$$f(x_t) - f^\star \le \langle \nabla f(x_t), x_t - y_t \rangle .$$

The following lemma gives the per-step improvement in the smooth setting:

Lemma 44. *Suppose f is convex and L-smooth w.r.t. $\|\cdot\|$ and the diameter of Ω is bounded by D w.r.t. $\|\cdot\|$. Then,*

$$f(x_{t+1}) \le f(x_t) - \eta_t \langle \nabla f(x_t), x_t - y_t \rangle + \frac{L\eta_t^2}{2} \|x_t - y_t\|^2 .$$

Proof. Recall that $x_{t+1} = (1 - \eta_t)x_t + \eta_t y_t$ where

$$y_t \in \arg\min_{y \in \mathcal{X}} \langle \nabla f(x_t), y \rangle .$$

By smoothness,

$$f(x_{t+1}) \le f(x_t) + \langle \nabla f(x_t), x_{t+1} - x_t \rangle + \frac{L}{2} \|x_t - x_{t+1}\|^2 .$$

Plugging in the definition of x_{t+1}, we get

$$f(x_{t+1}) \le f(x_t) - \eta_t \langle \nabla f(x_t), x_t - y_t \rangle + \frac{L\eta_t^2}{2} \|x_t - y_t\|^2 . \qquad \blacksquare$$

Theorem 26. *Suppose we pick $\eta_t \in [0, 1]$ to minimize the upper bound at each iteration. Let $\delta_t := f(x_t) - f^\star$. Then,*

$$\delta_1, \delta_2, \dots$$

is a nonincreasing sequence and $\delta_T \le \epsilon$ for all

$$T \ge O\left(\frac{LD^2}{\epsilon}\right).$$

Remark 13. Some notes:

- The same rate can be achieved by explicitly setting $\eta_t = \frac{2}{t+2}$.
- The proof below is not as "elegant" as the standard proof of this result. The standard proof gives a better bound and is also shorter but relies on guessing a nice inductive hypothesis. The proof I present below is not as "elegant" but is easier to come up with. □

Proof. First, note that

By the previous lemma, for all $t \geq 0$,

$$\delta_{t+1} \leq \min_{\eta_t \in [0,1]} \left(\delta_t - \eta_t \langle \nabla f(x_t), x_t - y_t \rangle + \frac{L\eta_t^2}{2} \|x_t - y_t\|^2 \right).$$

Thus, by taking $\eta_0 = 1$ and noting that $\langle \nabla f(x_t), x_t - y_t \rangle \geq \delta_t$, we deduce that

$$\delta_1 \leq \frac{LD^2}{2}.$$

Next, note that δ_t is a nonincreasing sequence as we may take $\eta = 0$ at each step. Explicitly, the upper bound on δ_{t+1} is given by

$$\delta_{t+1} \leq \begin{cases} \delta_t - \frac{\langle \nabla f(x_t), x_t - y_t \rangle^2}{L\|x_t - y_t\|^2} & \text{if } \langle \nabla f(x_t), x_t - y_t \rangle \leq L \|x_t - y_t\|^2 \\ \delta_t - \langle \nabla f(x_t), x_t - y_t \rangle + \frac{L}{2} \|x_t - y_t\|^2 & \text{else.} \end{cases}$$

In the first case, we may bound $\langle \nabla f(x_t), x_t - y_t \rangle \geq \delta_t$ and $\|x_t - y_t\| \leq D$. In the second case, we can bound $\langle \nabla f(x_t), x_t - y_t \rangle \geq \frac{L}{2} \|x_t - y_t\|^2 + \frac{1}{2}\delta_t$. Thus,

$$\delta_{t+1} \leq \max \left(\delta_t - \frac{\delta_t^2}{LD^2}, \delta_t/2 \right).$$

Now, fix $\epsilon > 0$. For each index $t = 1, 2, \ldots$ we will place t in the box $\mathcal{B}_k$ where

$$\frac{LD^2}{2^{k+1}} < \delta_t \leq \frac{LD^2}{2^k}.$$

Note that every index $t = 1, 2, \ldots$ falls in some box $\mathcal{B}_k$ for $k \geq 1$.

We will now bound the size of $\mathcal{B}_k$ for $k \geq 1$. There is at most one index $t \in \mathcal{B}_k$ satisfying

$$\delta_{t+1} \leq \delta_t/2.$$

Every other index in $\mathcal{B}_k$ satisfies

$$\delta_{t+1} \leq \delta_t - \frac{\delta_t^2}{LD^2} \leq \delta_t - \frac{LD^2}{(2^{k+1})^2}.$$

Thus,

$$|\mathcal{B}_k| = O\left(2^k\right).$$

We note that $\delta_T \leq \epsilon$ if

$$T > |\mathcal{B}_1| + |\mathcal{B}_2| + \cdots + \left|\mathcal{B}_{\lceil \log_2(LD^2/2\epsilon) \rceil}\right| = O\left(\frac{LD^2}{\epsilon}\right). \qquad \blacksquare$$

14.1 Lower bounds

We now show that this $O(\frac{LD^2}{T})$ convergence rate is in fact optimal (up to constants) if one assumes to only have first-order access to f and linear minimization oracle (LMO) access to Ω.

Theorem 27. *Consider an algorithm that makes T calls to a LMO (receiving response $x_1, \ldots, x_T$) and an arbitrary number of calls to a first-order oracle, and that outputs $\bar{x} \in \operatorname{conv}(x_1, \ldots, x_T)$. Then, there exists an L-smooth convex function in the Euclidean norm and a closed convex set $\mathcal{X} \subseteq \mathbb{R}^n$ with diameter $\leq D$ in the Euclidean norm s.t.*

$$f(\bar{x}) - f^\star \geq \frac{LD^2}{8T}$$

Proof. Let $n = 2T$ and define $\Delta \coloneqq \frac{D}{\sqrt{2}} \operatorname{conv}(\{e_1, \ldots, e_n\})$. Consider

$$\min_{x \in \Delta} \frac{L}{2} \|x\|^2 .$$

This objective function is L-smooth in the Euclidean norm. The diameter of Δ in the Euclidean norm is D.

By symmetry, the optimal value is achieved by $\frac{D}{n\sqrt{2}}\mathbf{1}$ and is

$$f^\star = \frac{LD^2}{4n} = \frac{LD^2}{8T}.$$

On the other hand, for any algorithm satisfying the assumptions, the responses $x_1, \ldots, x_T$ will each have support 1 so that $\operatorname{conv}(\{x_1, \ldots, x_T\})$ contains only vectors with support at most T. Thus,

$$f(\bar{x}) \geq \frac{LD^2}{4T}$$

and $f(\bar{x}) - f^\star \geq \frac{LD^2}{8T}$. ∎