Convex Sets (II)

Lecture 2, Convex Optimization

National Taiwan University

March 3, 2023

Table of contents

Orthogonal Complements

Orthogonal Complement

The **orthogonal complement** of a nonempty subset $\mathcal{S}$ of $\mathbf{R}^n$, denoted by $\mathcal{S}^\perp$ (read "$\mathcal{S}$ perp"), is the set of all vectors in $\mathbf{R}^n$ that are **orthogonal** to every vector in $\mathcal{S}$. That is,

$$\mathcal{S}^\perp = \{\mathbf{v} \in \mathbf{R}^n \mid \mathbf{v} \cdot \mathbf{u} = 0, \forall \mathbf{u} \in \mathcal{S}\}.$$

Properties of Orthogonal Complements

1. The orthogonal complement of any nonempty subset of $\mathbf{R}^n$ is a subspace of $\mathbf{R}^n$.
2. For any nonempty subset $\mathcal{S}$ of $\mathbf{R}^n$, we have $\mathcal{S}^\perp = (\text{Span } \mathcal{S})^\perp$. In particular, the orthogonal complement of a basis for a subspace is the same as the orthogonal complement of the subspace.
3. For any matrix A, with its ith row being denoted by a_i^T, the orthogonal complement of the row space of A is the null space of A; that is

$$(\text{Span }\{a_i\})^\perp = (\mathcal{R}(A^T))^\perp = \mathcal{N}(A).$$

Orthogonal Projections

Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbf{R}^n$. Then, for any vector $\mathbf{u}$ in $\mathbf{R}^n$, there exist unique vectors $\mathbf{w}$ in W and $\mathbf{z}$ in $W^\perp$ such that $\mathbf{u} = \mathbf{w} + \mathbf{z}$. In addition, if $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ is an orthonormal basis for W, then

$$\mathbf{w} = (\mathbf{u} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{u} \cdot \mathbf{v}_2)\mathbf{v}_2 + \cdots + (\mathbf{u} \cdot \mathbf{v}_k)\mathbf{v}_k.$$

Orthogonal Projection

Let W be a subspace of $\mathbf{R}^n$ and $\mathbf{u} \in \mathbf{R}^n$. The **orthogonal projection of $\mathbf{u}$ on W** is the unique vector $\mathbf{w}$ such that $\mathbf{u} - \mathbf{w} \in W^\perp$.

Closest Vector Property

Let W be a subspace of $\mathbf{R}^n$ and $\mathbf{u}$ be a vector in $\mathbf{R}^n$. Among all vectors in W, the vector **closest** to $\mathbf{u}$ is the **orthogonal projection** of $\mathbf{u}$ on W.

Eigenvalues and Eigenvectors

Eigenvalues and Eigenvectors

- Let T be a linear operator on $\mathbf{R}^n$. A nonzero vector v in $\mathbf{R}^n$ is called an **eigenvector** of T if $T(v)$ is a multiple of v, that is,

$$T(v) = \lambda v$$

for some λ. The scalar λ is called the **eigenvalue** of T that corresponds to v.

- Let A be an $n \times n$ matrix. A nonzero vector v in $\mathbf{R}^n$ is called an **eigenvector** of A if

$$Av = \lambda v$$

for some scalar λ. The scalar λ is called the **eigenvalue** that corresponds to v.

Diagonalizability

Diagonalizability

An $n \times n$ matrix A is called **diagonalizable** if $A = PDP^{-1}$ for some diagonal $n \times n$ matrix D and some invertible $n \times n$ matrix P.

Diagonalizability and Eigen-decomposition

a) An $n \times n$ matrix A is diagonalizable if and only if there is a basis for $\mathbf{R}^n$ consisting of eigenvectors of A.

b) If $P \in \mathbf{R}^{n \times n}$ is invertible and $D \in \mathbf{R}^{n \times n}$ is diagonal , then

$$A = PDP^{-1}$$

if and only if the columns of P are a basis for $\mathbf{R}^n$ consisting of eigenvectors of A and the diagonal entries of D are the eigenvalues corresponding to the respective columns of P.

Orthogonal Diagonalizability

Orthogonal Matrix

A matrix P is called an **orthogonal matrix** if $P^T P = I_n$.

Orthogonal Diagonalizability

An $n \times n$ matrix A is called **orthogonally diagonalizable** if $A = PDP^{-1}$ for some diagonal $n \times n$ matrix D and some $n \times n$ orthogonal matrix P.

- If A is orthogonally diagonalizable, then $A = PDP^{-1} = PDP^T$.
- Then, we have $A^T = (PDP^T)^T = PD^T P^T = PDP^T = A$, i.e, A is symmetric.
- Conversely, if $A \in \mathbf{S}^n$, then A is orthogonally diagonalizable.
- There exists an orthonormal basis for $\mathbf{R}^n$ in which every vector is an eigenvector of A.

Properties of Symmetric Matrices (1/2)

- Let $A \in \mathbf{S}^n$. Then, there exist a diagonal matrix D and an orthogonal matrix P such that
$$A = PDP^T.$$
- For any $n \geq 1$, we have
$$A^n = \underbrace{(PDP^T)(PDP^T)\cdots(PDP^T)}_{n \text{ times}} = \cdots = PD^nP^T.$$
- If $A \in \mathbf{S}^n_+$, then all diagonal entries of D are nonnegative, i.e., $D = \mathbf{diag}\left([\lambda_1, \cdots, \lambda_n]^T\right)$ where $\lambda_i \geq 0$ for all i.

Properties of Symmetric Matrices (2/2)

Positive definite matrices and Positive semidefinite matrices

A symmetric matrix $A \in \mathbf{S}^n$ is said to be **positive definite** if for all $v \in \mathbf{R}^n \backslash \{0\}$, $v^T A v > 0$.

A symmetric matrix $A \in \mathbf{S}^n$ is said to be **positive semidefinite** if for all $v \in \mathbf{R}^n$, $v^T A v \geq 0$.

Eigenvalues of positive-definite and positive semidefin matrices

A symmetric matrix $A \in \mathbf{S}^n$ is positive definite if and only if every eigenvalue of A is positive.

A symmetric matrix $A \in \mathbf{S}^n$ is positive semidefinite if and only if every eigenvalue of A is nonnegative.

Square root of a (positive semidefinite) symmetric matrix

We define $D^{1/2} = \mathbf{diag}\left([\lambda_1^{1/2}, \cdots, \lambda_n^{1/2}]^T\right)$ and

$$A^{1/2} = PD^{1/2}P^T.$$

Properties of Matrices

Properties of Symmetric Matrices

Let $X \in \mathbf{S}^n$ with eigenvalues $\lambda_1, ..., \lambda_n \in \mathbf{R}$, then

- $\mathbf{tr}\ X = \sum_{k=1}^{n} \lambda_k$.
- $\mathbf{det}\ X = \prod_{k=1}^{n} \lambda_k$.

Some more properties of traces and determinants

- $\mathbf{tr}\ (AB) = \mathbf{tr}\ (BA)$ for any $A \in \mathbf{R}^{m \times n}$ and $B \in \mathbf{R}^{n \times m}$.
- $\mathbf{det}\ (AB) = \mathbf{det}\ (BA)$ for any $A, B \in \mathbf{R}^{n \times n}$

Singular Value Decomposition (1/4)

Singular Value Decomposition

Suppose $A \in \mathbf{R}^{m \times n}$ with rank $A = r$. Then A can be factored as

$$A = U \Sigma V^T,$$

where $U \in \mathbf{R}^{m \times r}$ satisfies $U^T U = I$, $V \in \mathbf{R}^{n \times r}$ satisfies $V^T V = I$, and $\Sigma = \mathbf{diag}\ (\sigma_1, \cdots, \sigma_r)$, with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.
Such a factorization is called the **singular value decomposition (SVD)** of A.
The columns of U are called **left singular vectors** of A, the columns of V are right singular vectors, and the numbers σ_i are the **singular values**.

The singular value decomposition can be written

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T,$$

where $u_i \in \mathbf{R}^m$ are the left singular vectors, and $v_i \in \mathbf{R}^n$ are the right singular vectors.

Singular Value Decomposition (2/4)

- The singular value decomposition of a matrix A is closely related to the eigenvalue decomposition of the (symmetric, nonnegative definite) matrix $A^T A \in \mathbf{R}^{n \times n}$.
- Using $A = U\Sigma V^T$ we can write

$$A^T A = V\Sigma^2 V^T = \left[\begin{array}{cc} V & \tilde{V} \end{array}\right] \left[\begin{array}{cc} \Sigma^2 & 0 \\ 0 & 0 \end{array}\right] \left[\begin{array}{cc} V & \tilde{V} \end{array}\right]^T$$

  where $\tilde{V}$ is any matrix for which $\left[\begin{array}{cc} V & \tilde{V} \end{array}\right]$ is orthogonal.
- It is observed that the nonzero eigenvalues of $A^T A$ are the singular values of A squared, and the associated eigenvectors of $A^T A$ are the right singular vectors of A.
- Similarly, nonzero eigenvalues of AA^T are the singular values of A squared, and the associated eigenvectors of AA^T are the left singular vectors of A.

Singular Value Decomposition (3/4)

- The first or largest singular value is also written as $\sigma_{\max}(A)$, and can be expressed as

$$\sigma_{\max}(A) = \sup_{x,y \neq 0} \frac{x^T A y}{\|x\|_2 \|y\|_2} = \sup_{y \neq 0} \frac{\|Ay\|_2}{\|y\|_2},$$

[1] showing that the maximum singular value is the ℓ_2 operator norm[2] of A.

- The minimum singular value of $A \in \mathbf{R}^{m \times n}$ is given by

$$\sigma_{\min}(A) = \begin{cases} \sigma_r(A) & r = \min\{m, n\} \\ 0 & r < \min\{m, n\}, \end{cases}$$

which is positive if and only if A is full rank.

[1] The notation sup denotes supremum, which will be introduced in a subsequent lecture. It can be thought of as $\max$ for now.

[2] The operator norm is to be defined shortly.

Singular Value Decomposition (4/4)

- The singular values of a symmetric matrix are the absolute values of its nonzero eigenvalues, sorted into descending order.
- The singular values of a symmetric positive semidefinite matrix are the same as its nonzero eigenvalues.
- The **condition number** of a nonsingular $A \in \mathbf{R}^{n\times n}$, denoted $\textbf{cond}(A)$ or $\kappa(A)$, is defined as

$$\textbf{cond}(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}.$$

Norms

Norms

A function $f : \mathbf{R}^n \to \mathbf{R}$ (with $\mathbf{dom}\ f = \mathbf{R}^n$) is called a **norm** if for any $x, y \in \mathbf{R}^n, t \in \mathbf{R}$, we have

- $f(x) \geq 0$ (f is nonnegative).
- $f(x) = 0$ only if $x = 0$ (f is definite).
- $f(tx) = |t| f(x)$ (f is homogeneous).
- $f(x + y) \leq f(x) + f(y)$ (f satisfies the triangle inequality).

ℓ_p-norm

Let $p \geq 1$. Then the **ℓ_p-norm** is defined as

$$||x||_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

Question: When $p < 1$, is $||x||_p$ still a norm?

Examples of ℓ_p-norm

- When $p = 2$, the ℓ_2-norm is actually the Euclidean norm:

$$||x||_2 = (x_1^2 + \cdots + x_n^2)^{1/2}.$$

- When $p = 1$, the ℓ_1-norm is the sum-absolute-value:

$$||x||_1 = |x_1| + \cdots + |x_n|.$$

- When $p \to \infty$, the ℓ_∞-**norm** is defined as:

$$||x||_\infty \triangleq \lim_{p\to\infty} ||x||_p = \lim_{p\to\infty} (|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

It can be shown that $||x||_\infty = \max\{|x_1|, \cdots, |x_n|\}$.

Unit Ball

Unit ball

Given a norm $||\cdot||$, the **unit ball** with respect to the norm is defined as

$$\{x \mid ||x|| \leq 1\}.$$

- What are the unit balls with respect to the ℓ_p-norm with $p = 1, 2, \infty$?

P-Quadratic Norms

- For $P \in \mathbf{S}^n_{++}$, the P-**quadratic norm** is defined as

$$||x||_P = (x^T P x)^{1/2} = ||P^{1/2} x||_2.$$

  - If rank $P < n$, is $||x||_P$ still a norm?
- The **unit ball** of a quadratic norm,

$$\{x \in \mathbf{R}^n \mid ||x||_P \leq 1\},$$

is an ellipsoid.

Matrix Norms – Norms Defined On $\mathbf{R}^{m\times n}$ (1/2)

Matrix Norms

A function $f : \mathbf{R}^{m\times n} \to \mathbf{R}$ (with $\mathbf{dom}\ f = \mathbf{R}^{m\times n}$) is called a **norm** if for any $X, Y \in \mathbf{R}^{m\times n}, t \in \mathbf{R}$, we have

- $f(X) \geq 0$ (f is nonnegative).
- $f(X) = 0$ only if $X = 0$ (f is definite).
- $f(tX) = |t|f(X)$ (f is homogeneous).
- $f(X + Y) \leq f(X) + f(Y)$ (f satisfies the triangle inequality).

- The **Frobenius norm**, defined on $\mathbf{R}^{m\times n}$, is

$$||X||_F = (\mathbf{tr}\ (X^T X))^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2\right)^{1/2}.$$

- The **sum-absolute-value norm**: $||X||_{\text{sav}} = \sum_{i=1}^{m}\sum_{j=1}^{n} |X_{ij}|$.
- The **maximum-absolute-value norm**:
  $||X||_{\text{mav}} = \max\{|X_{ij}| \mid i = 1, \cdots, m,\ j = 1, \cdots, n\}$.

Matrix Norms – Norms Defined On $\mathbf{R}^{m\times n}$ (2/2)

- Suppose $||\cdot||_a$ and $||\cdot||_b$ are norms on $\mathbf{R}^m$ and $\mathbf{R}^n$, respectively. The **operator norm** of $X \in \mathbf{R}^{m\times n}$, induced by the norms $||\cdot||_a$ and $||\cdot||_b$, is defined as

$$||X||_{a,b} = \sup\{||Xu||_a \mid ||u||_b \leq 1\}.$$

- If $a = b = 2$, we obtain the **spectral norm** of X, which equals to the maximum singular value of X:

$$||X||_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^T X))^{1/2}.$$

It is also called **ℓ_2-norm** of X.

Schur complement

Schur complement

Consider a matrix $X \in \mathbf{S}^n$ partitioned as

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

where $A \in \mathbf{S}^k$. If $\mathbf{det}\ A \neq 0$, the matrix

$$S = C - B^T A^{-1} B$$

is called the **Schur complement** of A in X.

- It can be shown that
  1. $\mathbf{det}\ X = \mathbf{det}\ A \mathbf{det}\ S$.
  2. $X \succ 0$ if and only if $A \succ 0$ and $S \succ 0$.
  3. If $A \succ 0$, then $X \succeq 0$ if and only if $S \succeq 0$.

Norm Balls and Norm Cones

Norm Balls and Norm Cones

- Suppose $||\cdot||$ is a norm on $\mathbf{R}^n$.
- It can be shown that a **norm ball** of radius r and center x_c, given by $\{x \mid ||x - x_c|| \leq r\}$, is convex.
- The **norm cone** associated with the norm $||\cdot||$ is the set

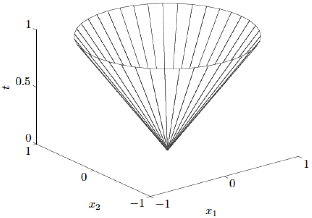
$$C = \{(x, t) \mid ||x|| \leq t\} \subseteq \mathbf{R}^{n+1}.$$

Second-Order Cone

The **second-order cone** is the norm cone for the Euclidean norm, i.e.,

$$\begin{aligned} C &= \left\{ (x,t) \in \mathbf{R}^{n+1} \,\middle|\, \|x\|_2 \le t \right\} \\ &= \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \,\middle|\, \begin{bmatrix} x \\ t \end{bmatrix}^T \begin{bmatrix} I & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \le 0, \quad t \ge 0 \right\}. \end{aligned}$$

It is also known as the quadratic cone, the Lorentz cone, or ice-cream cone.



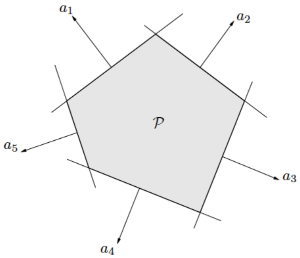
Polyhedra

Polyhedra

A **polyhedron** is defined as the solution set of a finite number of linear equations and linear inequalities:

$$\mathcal{P} = \left\{ x \mid a_i^T x \le b_i,\ i = 1, \cdots, m,\ c_j^T x = d_j, j = 1, \cdots, p \right\}.$$

- A polyhedron is the intersection of a finite number of halfspaces and hyperplanes.



Polyhedra

- Polyhedra are convex sets.
- Affine sets (including subspaces, hyperplanes, and lines) are polyhedra.
- Rays, line segments, and hyperplanes are polyhedra.
- A bounded[3] polyhedron is called a **polytope**.

[3] A subset C of $\mathbf{R}^n$ is called bounded if there exists $B > 0$ such that any $x \in C$ satisfies $|x_i| \leq B$ for any $i \in \{1, 2, ...n\}$.

Polyhedra

The polyhedron

$$\mathcal{P} = \{x \mid a_i^T x \leq b_i,\ i = 1, \cdots, m,\ c_j^T x = d_j, j = 1, \cdots, p\}$$

can be rewritten as

$$\mathcal{P} = \{x \mid Ax \preceq b, Cx = d\}$$

where

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \text{ and } C = \begin{bmatrix} c_1^T \\ \vdots \\ c_p^T \end{bmatrix},$$

and the symbol $\preceq$ denotes **vector inequality** or **componentwise inequality** in $\mathbf{R}^m$: $u \preceq v$ means $u_i \leq v_i$ for $i = 1, \cdots, m$.

Polyhedra – An example

The set of nonnegative numbers

Let $\mathbf{R}_+$ denote the set of nonnegative numbers. Let $\mathbf{R}_{++}$ denote the set of positive numbers.

Nonnegative orthant

The **nonnegative orthant** in $\mathbf{R}^n$ is

$$\mathbf{R}_+^n = \{x \in \mathbf{R}^n \mid x_i \geq 0,\ i = 1, \cdots, n\} = \{x \in \mathbf{R}^n \mid x \succeq 0\}.$$

- The nonnegative orthant is a polyhedron and a cone (called a **polyhedral cone**).

Simplexes – Another example of polyhedra

Affinely Independent

The $k+1$ points $v_0, v_1, \cdots, v_k \in \mathbf{R}^n$ are called **affinely independent** if $\{v_1 - v_0, \cdots, v_k - v_0\}$ is linearly independent.

Simplex

Suppose the $k+1$ points $v_0, v_1, \cdots, v_k \in \mathbf{R}^n$ are affinely independent. The **simplex** determined by these $k+1$ points is

$$C = \mathbf{conv}\ \{v_0, \cdots, v_k\} = \left\{\theta_0 v_0 + \cdots + \theta_k v_k \mid \theta \succeq 0, \mathbf{1}^T\theta = 1\right\}$$

where $\mathbf{1}$ is the vector with all entries one.

- The above defined simplex is sometimes called a **k-dimensional simplex in $\mathbf{R}^n$**, since its affine dimension is k.

Examples of Simplexes

- A 1-dimensional simplex is a line segment.
- A 2-dimensional simplex is a triangle (including its interior).
- A 3-dimensional simplex is a tetrahedron.

Unit Simplex

The **unit simplex** in $\mathbf{R}^n$ is the n-dimensional simplex determined by the zero vector and the unit vectors: $\{0, e_1, \cdots, e_n\}$.

The unit simplex can be expressed as

$$\left\{x \mid x \succeq 0, \quad \mathbf{1}^T x \leq 1\right\}.$$

Example of Simplexes – Probability Simplex

- The **probability simplex** in $\mathbf{R}^n$ is the $(n-1)$-dimensional simplex determined by the unit vectors $\{e_1, \cdots, e_n\}$.
- It can be expressed as

$$\left\{x \mid x \succeq 0, \ \ \mathbf{1}^T x = 1\right\}.$$

- Vectors in the probability simplex correspond to probability distributions on a set with n elements.

Expressing A Simplex as A Polyhedron

- Consider the simplex

$$C = \textbf{conv}\ \{v_0, \cdots, v_k\} = \left\{\theta_0 v_0 + \cdots + \theta_k v_k \mid \theta \succeq 0, \mathbf{1}^T\theta = 1\right\}$$

- Let

$$B = \begin{bmatrix} v_1 - v_0 & \cdots & v_k - v_0 \end{bmatrix} \in \mathbf{R}^{n \times k}$$

and $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \in \mathbf{R}^{n\times n}$ be a nonsingular matrix such that

$$AB = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} B = \begin{bmatrix} I_k \\ 0_{(n-k)\times k} \end{bmatrix}.$$

- Then, we have $x \in C$ if and only if

$$A_2 x = A_2 v_0, \quad A_1 x \succeq A_1 v_0, \quad \mathbf{1}^T A_1 x \le 1 + \mathbf{1}^T A_1 v_0.$$

(a form of a polyhedron)

Convex Hull Description of Polyhedra (1/2)

- Consider the convex hull of the finite set $\{v_1, \cdots, v_k\}$,

$$\begin{aligned} & \textbf{conv}\,\{v_1, \cdots, v_k\} \\ = \ & \{\theta_1 v_1 + \cdots + \theta_k v_k \mid \theta_i \geq 0,\ i = 1, \cdots, k,\ \theta_1 + \cdots + \theta_k = 1\} \\ = \ & \{\theta_1 v_1 + \cdots + \theta_k v_k \mid \theta \succeq 0,\ \mathbf{1}^T\theta = 1\} \end{aligned}$$

- It is a polyhedron and is bounded. (why?)
- How can we express **conv** $\{v_1, \cdots, v_k\}$ in the form

$$\mathcal{P} = \{x \mid a_i^T x \leq b_i,\ i = 1, \cdots, m,\ c_j^T x = d_j, j = 1, \cdots, p\}?$$

Convex Hull Description of Polyhedra (2/2)

- Conversely, how do we express a polyhedron

$$\mathcal{P} = \left\{ x \,\middle|\, a_i^T x \le b_i,\ i = 1, \cdots, m,\ c_j^T x = d_j,\ j = 1, \cdots, p \right\}$$

in the form of convex hull description **conv** $\{v_1, \cdots, v_k\}$?

- Example: consider

$$C = \{x \mid |x_i| \le 1,\ i = 1, \cdots, n\}$$

(with $2n$ linear inequalities). Then we have

$$C = \textbf{conv}\,\{v_1, \cdots, v_{2^n}\},$$

where $v_1, \cdots, v_{2^n}$ are the 2^n vectors whose components are all 1 or -1.

Notations for Sets of Symmetric Matrices

- The notation $\mathbf{S}^n$ denotes the set of symmetric $n \times n$ matrices:

$$\mathbf{S}^n = \{X \in \mathbf{R}^{n \times n} \mid X = X^T\}.$$

- $\mathbf{S}^n$ is a vector space with dimension $n(n+1)/2$.
- The notation $\mathbf{S}^n_+$ denotes the set of symmetric positive semidefinite matrices:

$$\mathbf{S}^n_+ = \{X \in \mathbf{S}^n \mid X \succeq 0\}.$$

- The notation $\mathbf{S}^n_{++}$ denotes the set of symmetric positive definite matrices:

$$\mathbf{S}^n_{++} = \{X \in \mathbf{S}^n \mid X \succ 0\}.$$

Isomorphism of Symmetric Matrix Subspace (1/2)

Symmetric Vectorization

We define the **symmetric vectorization** function $\mathrm{svec} : \mathbf{S}^n \to \mathbf{R}^{n(n+1)/2}$ with $\mathbf{dom}\,\mathrm{svec} = \mathbf{S}^n$ and

$$\mathrm{svec}(Y) \triangleq \begin{bmatrix} Y_{11} \\ \sqrt{2}Y_{12} \\ Y_{22} \\ \sqrt{2}Y_{13} \\ \sqrt{2}Y_{23} \\ Y_{33} \\ \vdots \\ Y_{nn} \end{bmatrix} \in \mathbf{R}^{n(n+1)/2}.$$

Isomorphism of Symmetric Matrix Subspace (2/2)

The symmetric vectorization is an isometric isomorphism

It can be shown that svec is an isometric isomorphism:

- $\|\mathrm{svec}\ X - \mathrm{svec}\ Y\|_2 = \|X - Y\|_F$.
- $(\mathrm{svec}\ X)^T \mathrm{svec}\ Y = \mathbf{tr}\ (XY)$.

Positive Semidefinite Cone

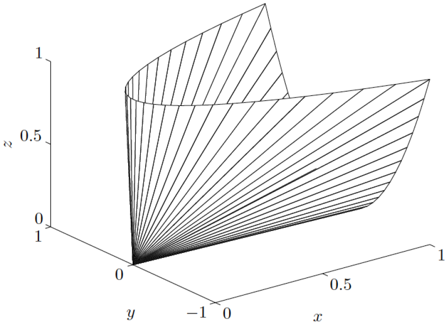
Convexity of Positive Semidefinite Cones

The set $\mathbf{S}_+^n$ is a convex cone:
if $\theta_1, \theta_2 \geq 0$ and $A, B \in \mathbf{S}_+^n$, then $\theta_1 A + \theta_2 B \in \mathbf{S}_+^n$.

Proof:

Positive Semidefinite Cone in $\mathbf{S}^2$

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in \mathbf{S}_+^2 \iff x \geq 0, \quad z \geq 0, \quad xz \geq y^2.$$



Intersection Preserves Convexity

Intersection Preserves Convexity

If S_1 and S_2 are convex, then $S_1 \cap S_2$ is convex.

Intersection of an infinite number of sets

If S_α is convex for every $\alpha \in \mathcal{A}$, then

$$\bigcap_{\alpha \in \mathcal{A}} S_\alpha$$

is convex. Here, $\mathcal{A}$ is the set of indices and can be finite or infinite.

- Example: A polyhedron is the intersection of halfspaces and hyperplanes (which are convex), and therefore is convex.

Positive Semidefinite Cone

Positive Semidefinite Cone

The positive semidefinite cone $\mathbf{S}^n_+$ can be expressed as

$$\bigcap_{z \neq 0} \{X \in \mathbf{S}^n \mid z^T X z \geq 0\}$$

and is convex.

- For each $z \neq 0$, $z^T X z$ is a linear function of X, so the set

$$\{X \in \mathbf{S}^n \mid z^T X z \geq 0\}$$

is a halfspace in $\mathbf{S}^n$.

An Example

- Consider the set

$$S = \{x \in \mathbf{R}^m \mid |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}$$

where $p(t) = \sum_{k=1}^{m} x_k \cos kt$.

- The set S can be expressed as the intersection of an infinite number of slabs:

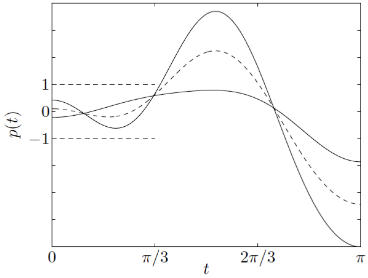
$$S = \bigcap_{|t| \leq \pi/3} S_t$$

where

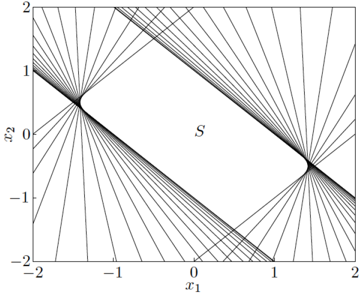
$$S_t = \left\{x \mid -1 \leq (\cos t, \cdots, \cos mt)^T x \leq 1\right\}.$$

- So, S is convex.

An Example



$$p(t) = \sum_{k=1}^{m} x_k \cos kt$$



$$S = \bigcap_{|t| \le \pi/3} S_t,$$

where $S_t =$

$$\{x \mid -1 \le [\cos t, \cdots, \cos mt]x \le 1\}$$

Convex Sets as Intersection of Halfspaces

- We have seen that the intersection of (possibly infinite) halfspaces is convex.
- It will be shown that a converse is true: every closed convex set S is the intersection of (usually infinite) halfspaces.
- A closed convex set $S \subseteq \mathbf{R}^n$ is the intersection of all halfspaces that contain it:

$$S = \bigcap_{\substack{S \subseteq \mathcal{H} \subseteq \mathbf{R}^n \\ \mathcal{H} \text{ is a halfspace}}} \mathcal{H}.$$

Affine functions

Affine function

A function $f : \mathbf{R}^n \to \mathbf{R}^m$ is **affine** if it is a sum of a linear function and a constant. That is, it has the form

$$f(x) = Ax + b,$$

where $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$.

Affine functions preserve convexity

Suppose $S \subseteq \mathbf{R}^n$ is convex and $f : \mathbf{R}^n \to \mathbf{R}^m$ is an affine function. Then the **image** of S under f,

$$f(S) = \{f(x) \mid x \in S\},$$

is convex.

Affine functions

Affine functions preserve convexity

Suppose $S \subseteq \mathbf{R}^n$ is convex and $f : \mathbf{R}^n \to \mathbf{R}^m$ is an affine function. Then the **image** of S under f,

$$f(S) = \{f(x) \mid x \in S\},$$

is convex.

Inverse Image under Affine functions

Suppose $S \subseteq \mathbf{R}^n$ is convex and $f : \mathbf{R}^k \to \mathbf{R}^n$ is an affine function. Then the **inverse image** of S under f,

$$f^{-1}(S) = \{x \mid f(x) \in S\},$$

is convex.

Examples – Scaling, Translation, and Projection

- **Scaling**: If $S \subseteq \mathbf{R}^n$ is convex, then for any $\alpha \in \mathbf{R}$, the set

$$\alpha S = \{\alpha x \mid x \in S\}$$

  is convex.

- **Translation**: If $S \subseteq \mathbf{R}^n$ is convex, then for any $a \in \mathbf{R}^n$, the set

$$S + a = \{x + a \mid x \in S\}$$

  is convex.

- **Projection** onto some coordinates: If $S \subseteq \mathbf{R}^m \times \mathbf{R}^n$ is convex, then

$$T = \{x_1 \in \mathbf{R}^m \mid (x_1, x_2) \in S \text{ for some } x_2 \in \mathbf{R}^n\}$$

  is convex.

Cartesian Products of Sets and Sums of Sets

Cartesian Product of two sets

Suppose $S_1 \subseteq \mathbf{R}^m, S_2 \subseteq \mathbf{R}^n$, then the **Cartesian product** of S_1 and S_2 is defined as

$$S_1 \times S_2 = \{(x_1, x_2) \mid x_1 \in S_1, x_2 \in S_2\}.$$

- If S_1 and S_2 are convex, then $S_1 \times S_2$ is convex.

Sum of two sets

The **sum** of two sets, $S_1, S_2 \subseteq \mathbf{R}^n$, is defined as

$$S_1 + S_2 = \{x + y \mid x \in S_1, y \in S_2\}.$$

- If S_1 and S_2 are convex, then $S_1 + S_2$ is convex.

Partial Sums of Sets

Partial sum of two sets

The partial sum of $S_1, S_2 \subseteq \mathbf{R}^n \times \mathbf{R}^m$ is defined as

$$S = \{(x, y_1 + y_2) \mid (x, y_1) \in S_1, (x, y_2) \in S_2, x \in \mathbf{R}^n, y_i \in \mathbf{R}^m, i = 1, 2\}.$$

- Partial sums of convex sets are convex.
- Partial sums are general cases for set intersection $(m = 0)$ and set addition $(n = 0)$.

Examples – Polyhedra

- The polyhedron $\{x \mid Ax \preceq b\}$ can be expressed as the inverse image of the nonnegative orthant under the affine function $f(x) = b - Ax$:

$$\{x \mid Ax \preceq b\} = \left\{x \mid f(x) \in \mathbf{R}_+^m\right\}.$$

- More generally, the polyhedron $\{x \mid Ax \preceq b,\ Cx = d\}$ can be expressed as the inverse image of the Cartesian product of the nonnegative orthant and the origin under the affine function $f(x) = (b - Ax, d - Cx)$:

$$\{x \mid Ax \preceq b,\ Cx = d\} = \left\{x \mid f(x) \in \mathbf{R}_+^m \times \{0\}\right\}.$$

Examples – Ellipsoid

- The ellipsoid

$$\mathcal{E} = \left\{ x \,\middle|\, (x - x_c)^T P^{-1} (x - x_c) \leq 1 \right\},$$

where $P \in \mathbf{S}^n_{++}$ is the image of the unit Euclidean ball $\{u \mid ||u||_2 \leq 1\}$ under the affine mapping $f(u) = P^{1/2}u + x_c$.

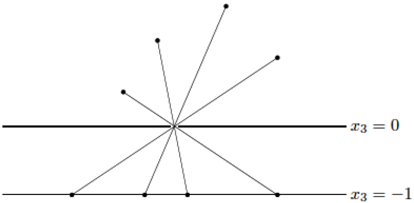
- It is also the inverse image of the unit Euclidean ball under the affine mapping $g(x) = P^{-1/2}(x - x_c)$.

Perspective Functions

Perspective function

The **perspective function** $P : \mathbf{R}^{n+1} \to \mathbf{R}^n$, with domain **dom** $P = \mathbf{R}^n \times \mathbf{R}_{++}$, is defined as $P(z, t) = z/t$.

The perspective function can be interpreted as the action of a pin-hole camera.



Perspective Functions Preserve Convexity

- Let $C \subseteq \textbf{dom}\ P = \mathbf{R}^n \times \mathbf{R}_{++}$ be convex, then its image under the perspective function $P : \mathbf{R}^{n+1} \to \mathbf{R}^n$, defined as $P(z,t) = z/t$, i.e.,

$$P(C) = \{P(x) \mid x \in C\}$$

is also convex.

Proof idea: A line segment in C is mapped to a line segment in $P(C)$ under $P(\cdot)$.

Perspective Functions Preserve Convexity

- The inverse image of a convex set under the perspective function is also convex:
- If $C \subseteq \mathbf{R}^n$ is convex, then

$$P^{-1}(C) = \{(x,t) \in \mathbf{R}^{n+1} \mid x/t \in C,\ t > 0\}$$

is convex.

Linear-fractional functions

- A **linear-fractional function** is formed by composing the perspective function with an affine function.

Linear-fractional functions

Let $g : \mathbf{R}^n \to \mathbf{R}^{m+1}$ be affine:

$$g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix},$$

where $A \in \mathbf{R}^{m\times n}, b \in \mathbf{R}^m, c \in \mathbf{R}^n$, and $d \in \mathbf{R}$. The function $f : \mathbf{R}^n \to \mathbf{R}^m$ given by $f = P \circ g$, i.e.,

$$f(x) = \frac{Ax+b}{c^T x + d}, \quad \mathbf{dom}\ f = \left\{ x \mid c^T x + d > 0 \right\},$$

is called a **linear-fractional** (or **projective**) function.

- Affine functions and linear functions are special cases of linear-fractional functions.

Projective Interpretation (1/2)

- A linear-fractional function can be represented as a matrix

$$Q = \begin{bmatrix} A & b \\ c^T & d \end{bmatrix} \in \mathbf{R}^{(m+1)\times(n+1)}.$$

- The matrix Q maps the point $\begin{bmatrix} x \\ 1 \end{bmatrix}$ to $\begin{bmatrix} Ax+b \\ c^Tx+d \end{bmatrix}$, a scalar multiple of $\begin{bmatrix} f(x) \\ 1 \end{bmatrix}$.

Projective Interpretation (2/2)

- Let us associate $\mathbf{R}^n$ with a set of rays in $\mathbf{R}^{n+1}$ as follows.
- For any $z \in \mathbf{R}^n$ we associate the ray

$$\mathcal{P}(z) = \left\{ t \begin{bmatrix} z \\ 1 \end{bmatrix} \,\middle|\, t \geq 0 \right\}$$

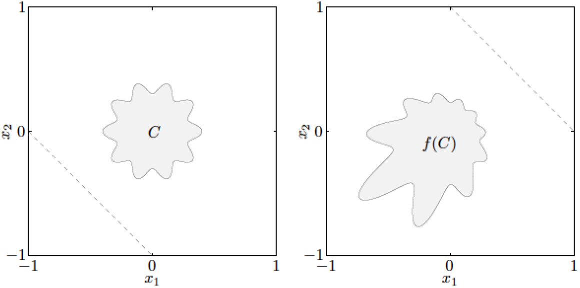
  in $\mathbf{R}^{n+1}$.
- Conversely, any ray in $\mathbf{R}^{n+1}$, with base at the origin and last component which takes on positive value, can be written as $\mathcal{P}(v) = \left\{ t \begin{bmatrix} v \\ 1 \end{bmatrix} \,\middle|\, t \geq 0 \right\}$ for some $v \in \mathcal{R}^n$.
- The correspondence $\mathcal{P}$ is therefore one-to-one and onto.
- The linear-fractional function f can be expressed as

$$f(x) = \mathcal{P}^{-1}(Q\mathcal{P}(x)).$$

Linear-fractional Functions Preserve Convexity

- Linear-fractional functions preserve convexity.
- If C is convex and $C \subseteq \textbf{dom } f = \{x \mid c^T x + d > 0\}$, then its image $f(C)$ is convex.
  - Proof idea: $f = P \circ g$ where P is the perspective function and g is an affine function.
- Similarly, if $C \subseteq \mathbf{R}^n$ is convex, then the inverse image $f^{-1}(C)$ is convex.

Linear-fractional Functions – An Example



- A set $C \subseteq \mathbf{R}^2$ and its image under the linear-fractional function

$$f(x) = \frac{x}{x_1 + x_2 + 1}, \quad \textbf{dom } f = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \,\middle|\, x_1 + x_2 + 1 > 0 \right\}.$$

Linear-fractional Functions – An Example

- Suppose u and v are random variables that take on values in $\{1, ..., n\}$ and $\{1, ..., m\}$, respectively.
- Let $p_{ij} = \textbf{prob}(u = i, v = j)$. Then the conditional probability $f_{ij} = \textbf{prob}(u = i | v = j)$ is

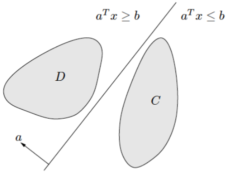
$$f_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n} p_{kj}}.$$

- Then, f is obtained by a linear-fractional mapping from p. (what is the mapping?)

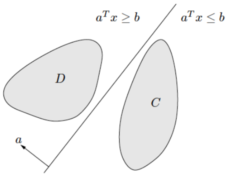
Separating Hyperplane Theorem (1/2)

Separating Hyperplane

The hyperplane $\{x \mid a^T x = b\}$ is called a **separating hyperplane** for the sets C and D, or is said to **separate** the sets C and D if $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$.



Separating Hyperplane Theorem (2/2)



Separating Hyperplane Theorem

Suppose C and D are two convex sets that do not intersect, i.e., $C \cap D = \varnothing$. Then, there exist $a \neq 0$ and b such that the hyperplane $\{x \mid a^T x = b\}$ separates C and D.

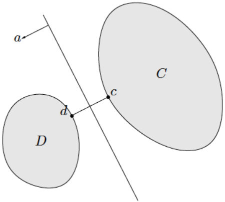
Separating Hyperplane Theorem – Proof of a Special Case (1/2)

- Consider that C and D are both convex, closed, and bounded.
- Assume that the **Euclidean distance** between C and D, defined as

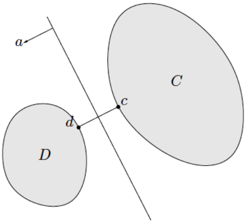
$$\textbf{dist}(C, D) = \inf\{||u - v||_2 \mid u \in C, v \in D\},$$

  is positive.
- Since C and D are both closed and bounded, there exist $c \in C$ and $d \in D$ such that

$$||c - d||_2 = \textbf{dist}(C, D).$$



Separating Hyperplane Theorem – Proof of a Special Case (2/2)



- Let

$$a = d - c, \quad b = \frac{||d||_2^2 - ||c||_2^2}{2}.$$

- Then, it can be shown that the affine function

$$f(x) = a^T x - b = (d - c)^T \left(x - \frac{d + c}{2} \right)$$

is nonpositive on C and nonnegative on D.

Example – A Convex Set and An Affine Set

- Suppose $C \subseteq \mathbf{R}^n$ is convex and $D \subseteq \mathbf{R}^n$ is affine, i.e., $D = \{Fu + g | u \in \mathbf{R}^m\}$, where $F \in \mathbf{R}^{n\times m}$, $g \in \mathbf{R}^n$.
- Suppose C and D are disjoint, so by the separating hyperplane theorem there are $a \neq 0$ and b such that $a^T x \leq b$ for all $x \in C$ and $a^T x \geq b$ for all $x \in D$.
- $\because a^T x \geq b$ for all $x \in D$, $\therefore a^T F u \geq b - a^T g$ for all $u \in \mathbf{R}^m$.
- But a linear function is bounded below on $\mathbf{R}^m$ only when it is zero, so we conclude $a^T F = 0$ (and hence, $b \leq a^T g$).
- Thus we conclude that there exists $a \neq 0$ such that $F^T a = 0$ and $a^T x \leq a^T g$ for all $x \in C$.

Convex Sets as Intersection of Halfspaces (Revisit)

- We have seen that the intersection of (possibly infinite) halfspaces is convex.
- It will be shown that a converse is true: every closed convex set S is the intersection of (usually infinite) halfspaces.
- A closed convex set S is the intersection of all halfspaces that contain it:

$$S = \bigcap_{\substack{S \subseteq \mathcal{H} \subseteq \mathbf{R}^n \\ \mathcal{H} \text{ is a halfspace}}} \mathcal{H}.^4$$

[4] In the text book, it was written as $S = \bigcap \{\mathcal{H} \mid \mathcal{H} \text{ halfspace}, S \subseteq \mathcal{H}\}$.

Strict Separation of Convex Sets

Strict separation

For two sets $C, D \subseteq \mathbf{R}^n$, if there exists $a \in \mathbf{R}^n, b \in \mathbf{R}$ such that

$$a^T x < b \ \forall x \in C \text{ and } a^T x > b \ \forall x \in D,$$

then C and D are said to be **strictly separable**, and the hyperplane $\{x \mid a^T x = b\}$ is called **strict separation** of C and D.

- Remark: The separating hyperplane theorem only dictates that two disjoint convex sets are separated by a hyperplane. A strict separation is not guaranteed (even when the sets are closed).

Example – A Point and A Closed Convex Set

- Let C be a closed convex set and $x_0 \notin C$. Then there exists a hyperplane that strictly separates $\{x_0\}$ from C.
- Proof idea:
  - The two sets C and $B(x_0, \epsilon)$ do not intersect for some $\epsilon > 0$.
  - Apply the separating hyperplane theorem on C and $B(x_0, \epsilon)$ (getting a^T and b).
  - The affine function

$$f(x) = a^T x - b - \epsilon||a||_2/2$$

    strictly separates C and $\{x_0\}$.
- Corollary: A closed convex set is the intersection of all halfspaces that contain it.

Converse of Separating Hyperplane Theorems

- Question: If there exists a hyperplane that separates convex sets C and D, does this imply C and D are disjoint?
  - (No. Consider $C = D = \{0\} \subseteq \mathbf{R}$.)
- Suppose C and D are convex sets, with C open, and there exists an affine function f that is nonpositive on C and nonnegative on D. Then C and D are disjoint.
  - *Hint: f is negative on C.*

Theorem

Any two convex sets, at least one of which is open, are disjoint if and only if there exists a separating hyperplane.

Supporting Hyperplanes (1/2)

Supporting hyperplanes

Suppose $C \subseteq \mathbf{R}^n$, and x_0 is a point in its boundary $\mathbf{bd}\ C$, i.e.,

$$x_0 \in \mathbf{bd}\ C = \mathbf{cl}\ C \backslash \mathbf{int}\ C.^5$$

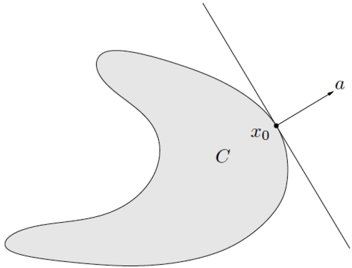
If $a \neq 0$ satisfies $a^T x \leq a^T x_0$ for all $x \in C$, then the hyperplane $\{x \mid a^T x = a^T x_0\}$ is called a **supporting hyperplane** to C at the point x_0.

- This is equivalent to the statement that $\{x_0\}$ and C are separated by the hyperplane $\{x \mid a^T x = a^T x_0\}$.
- The hyperplane is tangent to C at x_0, and the halfspace $\{x \mid a^T x \leq a^T x_0\}$ contains C.

[5] The notation $\mathbf{cl}$ means the closure of a set, a concept that will be formally introduced in the next lecture.

Supporting Hyperplanes (2/2)

- This is equivalent to the statement that $\{x_0\}$ and C are separated by the hyperplane $\{x|a^Tx = a^Tx_0\}$.
- The hyperplane is tangent to C at x_0, and the halfspace $\{x \mid a^Tx \leq a^Tx_0\}$ contains C.



Supporting Hyperplane Theorem

Supporting Hyperplane Theorem

For any nonempty convex set C, and any $x_0 \in \mathbf{bd}\ C$, there exists a supporting hyperplane to C at x_0.

Proof: Use the separating hyperplane theorem.

- If $\mathbf{int}\ C \neq \varnothing$: then by applying the separating hyperplane theorem on $\{x_0\}$ and $\mathbf{int}\ C$, the statement is proved.
- If $\mathbf{int}\ C = \varnothing$: then C lies in an affine set of dimension less than n. Then any hyperplane that contains this affine set contains both C and x_0 and therefore is a supporting hyperplane.

Converse of the Supporting Hyperplane Theorem

Converse of the Supporting Hyperplane Theorem

If a set C is closed, has nonempty interior, and has a supporting hyperplane at any $x_0 \in \mathbf{bd}\ C$, then C is convex.