Notes for Stanford EE364a – Convex Optimization I

Pranay Pasula

Intro

Notes are from both the course lectures and the course textbook *Convex Optimization* by Stephen Boyd and Lieven Vandenberghe.

In addition, I added my own notes wherever I saw fit (e.g., p-norms in section Norm Ball and Norm Cone).

$\forall x$ means "for every x."

Sorry for any typos.

Lecture 2 (Chapter 2 – Convex Sets)

Affine Set

A line through two points x_1 and x_2 can be represented by

$$x = \theta x_1 + (1-\theta)x_2 \qquad \forall \theta \in \mathbb{R}.$$

An **affine set** contains all points on the line connecting any two distinct points in the set.

The solution set $\{x|Ax = b\}$ of a linear system of equations is an affine set.

Conversely, any affine set is a solution set of some linear system of equations.

Convex Set

A line segment connecting two points x_1 and x_2 can be represented by

$$x = \theta x_1 + (1-\theta)x_2, \qquad 0 \le \theta \le 1.$$

A **convex set** contains the line segment connecting any two points in the set:

$$x_1, x_2 \in C \implies \theta x_1 + (1-\theta)x_2 \in C \qquad \forall \theta \in [0,1].$$

So, an affine set is a convex set.

Also, the null set is a convex set, as it is not non-convex.

Convex Combination

Convex combination of set $S = \{x_1, ..., x_k\}$:

$$x = \theta_1 x_1 + ... + \theta_k x_k, \quad \sum_{i=1}^{k} \theta_i = 1, \;\; \theta_i \geq 0 \quad \forall i = 1, ..., k).$$

Convex hull of S, *conv* $S = \{x | x$ is convex combination of S $\}$.

In $\mathbb{R}^n$, *conv* $S =$ set of points within or on boundary of border line segments.

Convex hull of a convex set is the convex set itself.

The convex hull of an open set is itself open, but the convex hull of a closed set is not necessarily closed.

Convex combinations differ from affine combinations only by the constraints above.

Convex Cone

A set C is a **cone** if $\forall x \in C$ and $\forall \theta \geq 0$, we have $\theta x \in C$.

A set C is a convex cone if C is convex and a cone, which means $\forall x_1, x_2 \in C$ and $\forall \theta \geq 0$, we have

$$\theta_1 x_1 + (1 - \theta) x_2 \in C.$$

Conic (non-neg) combination of $C = \{x_1, ..., x_k\}$:

$$x = \theta_1 x_1 + ... + \theta_k x_k, \quad \theta_i \geq 0 \quad \forall i = 1, ..., k.$$

Conic hull: (1) set of all conic combinations of points in C, (2) smallest convex cone that contains C.

- Here, C is any set, not just a cone or convex cone.

Convex cone is convex because the definition of a convex set is a subset of the definition of a convex cone (similar to why an affine set is a convex set).

Hyperplane

A **hyperplane** is a set of points of the form

$$\{x | a^T x = b\}$$

where $a \in \mathbb{R}^n$, $a \neq 0$, and $b \in \mathbb{R}$.

This is simply the solution set to a non-trivial linear equation.

$a^T x = b$ is equivalent to $a_1 x_1 + ... + a_n x_n = b$, so a is the normal vector to the hyperplane.

a points in the positive direction.

A hyperplane need not pass through the origin.
$\implies$ A hyperplane need not be a vector space.

A hyperplane in $\mathbb{R}^n$ is an **affine subspace** with dimension $n-1$ and **codimension** 1.
• An affine subspace is a vector subspace that has been shifted by a fixed vector.
• No vector is denoted as the origin.
• The codimension of an affine subspace W of a vector space V is

$$codim(W) = dim(V) - dim(W).$$

$\implies$ A hyperplane is affine and convex.

Halfspace

A **halfspace** is a set of points of the form

$$\{x | a^T x \square b\}$$

where $a \in \mathbb{R}^n$, $a \neq 0$, and $b \in \mathbb{R}$ and where the square can be $<, >$ (open halfspace), $\leq$, or $\geq$ (closed halfspace).

A halfspace is not a vector space.

A halfspace is convex but not affine.

A hyperplane splits the surrounding space into two halfspaces.

Euclidean Ball and Ellipsoid

A **euclidean ball** $B(x_c, r)$ is a set of points of the form

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\}.$$

Alternatively, this can be written as

$$B(x_c, r) = \{x_c + ru \mid \|u\|_2 \leq 1\}.$$

An **ellipsoid** is a set of the form

$$\mathcal{E} = \{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$$

with $P \in \boldsymbol{S}^n_{++}$ (i.e., P is symmetric positive definite).

Alternatively, this can be written as

$$\mathcal{E} = \{x_c + Au \mid \|u\|_2 < 1\}$$

where $A = P^{1/2} \implies A \in \boldsymbol{S}^n_{++}$.

The lengths of the semi-major axes of the ellipsoid are equal to square roots of the eigenvalues of P.

A ball is an ellipsoid with $P = r^2 I$.

When A is positive semidefinite but singular, the ellipsoid is called **degenerate**, and the affine dimension equals the rank of A.
$\implies$ Degenerate ellipsoids are convex.

Norm Ball and Norm Cone

A **norm** is a function $\|\cdot\| : \mathbb{R}^n \to [0, \infty)$ that satisfies

- (Non-negativity) $\|x\| \geq 0$ and 0 iff $x = 0$
- (Absolute homogeneity) $\|tx\| = |t|\|x\|$
- (Triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$

$\forall x, y \in \mathbb{R}^n$ and $t \in R$.

We treat $\|\cdot\|$ as a general (unspecified) norm. Only $\|\cdot\|_{\text{symb}}$ is a specific norm.

The **absolute value** function is an L1 norm over $\mathbb{R}$ (or $\mathbb{C}$).

For $p \geq 1$, the ***p*-norm** of a vector $x \in \mathbb{R}^n$ is

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

A **norm ball** of radius r and center x_c is the set of points

$$\{x \mid \|x - x_c\| \leq r\}.$$

The **norm cone** C associated with $\|\cdot\|$ is the set of points

$$C = \{(x, t) \mid \|x\| \leq t\}.$$

Unit norm ball in R^n is cross section (**level set** at $t = 1$) of corresponding norm cone.

All norm balls and norm cones are convex.

Polyhedra

A **polyhedron** $\mathcal{P}$ is defined as the solution set of a finite number of linear inequalities and equalities

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

$\implies$ a polyhedron is an intersection of halfspaces and hyperplanes.

$\preceq$ can be another component-wise inequality.

Affine sets, rays, line segments, and halfspaces are all polyhedra.

Polyhedra are convex.

A bounded polyhedron is sometimes called a **polytope**.

The **nonnegative orthant** $\mathbb{R}^n_+$ is

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x \succeq 0\}.$$

$\mathbb{R}^n_+$ is a polyhedron and a cone, sometimes called a **polyhedral cone**.

Positive Semidefinite Cone

$\boldsymbol{S}^n$, the set of all symmetric n x n matrices is convex, affine, and linear.

$\boldsymbol{S}^n_+ = \{X \in S \mid X \succeq 0\}$, the set of all positive semidefinite n x n matrices is a convex cone.
• Note, here and for matrix inequalities in general, $\succeq$ denotes definiteness.

$\boldsymbol{S}^n_{++}$ is not a cone.

Can use quadratic forms of $X, Y \in$ either $\boldsymbol{S}^n_+$ or $\boldsymbol{S}^n_{++}$ to show that each set is convex.

Operations That Preserve Convexity

Some ways to determine convexity of a set:

1. Use the definition of convexity (often difficult to do).
2. Show that the set is obtained from convexity-preserving operations on simple convex sets:
    - Intersection
    - Affine functions
    - Perspective functions
    - Linear-fractional functions
3. "Programming approach": For random x_1, x_2 in the set, test if $\theta_1 x_1 + (1-\theta)x_2$ is in the set.
   – This is just to check for non-convexity.

The **intersection** of any number of convex sets is convex.

If a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ is an **affine function** ($f(x) = Ax + b$, where $A \in \mathbb{R}^{mxn}, b \in \mathbb{R}^m$) then

- The image of a convex set under f is convex:

$$S \subseteq \mathbb{R}^n \text{ is convex } \implies f(S) = \{f(x) \mid x \in S\} \text{ is convex.}$$

- The inverse image $f^{-1}(C)$ of a convex set under f is convex:

$$C \subseteq \mathbb{R}^m \text{ is convex } \implies f^{-1}(C) = \{x \in \mathbb{R}^n \mid f(x) \in C\} \text{ is convex.}$$

- The converses of these are not necessarily true.

Examples of sets that can be shown to be convex through affine functions:

- Scaling, translation, rotation, projection
- Solution set of a **linear matrix inequality**: $\{x \mid x_1 A_1 + ... + x_m A_m \preceq B\}$, where $A_i, B \in \boldsymbol{S}^p$.
  – Here, $\preceq$ means $\lambda_{min}(LHS) \leq \lambda_{min}(B)$.
- Hyperbolic cone

A **perspective function** $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$ has the form

$$P(x,t) = \frac{x}{t}, \quad \text{where } \mathbf{dom}\ P = \{(x,t) \mid t > 0\}.$$

P divides elements $x_i, ..., x_n$ by x_{n+1} and removes x_{n+1} from the vector.

A generalization of the perspective function is the **linear-fractional function** $f : \mathbb{R}^n \to \mathbb{R}^m$:

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

If the image of a line segment under a function remains a line segment, then the function preserves convexity.

Generalized Inequalities

A set $K \subseteq \mathbb{R}^n$ is a **proper cone** iff

- K is closed (roughly, the entire boundary exists)
- K has nonempty interior (roughly, the interior is the set of points not on the boundary)
- K is pointed (contains no line)

Examples of proper cones: nonneg orthant, positive semidefinite cone (i.e., the set of positive semidefinite matrices).

A **generalized inequality** parametrized by proper cone K:

$$x \preceq_K y \iff y - x \in K \quad \text{and} \quad x \prec y \iff y - x \in \mathbf{int}\ K.$$

Examples: component-wise inequality ($K = \mathbb{R}^n_+$), matrix-wise inequality ($K = \boldsymbol{S}^n_+$).

Many properties of $\preceq_K$ are similar to $\leq$ on $\mathbb{R}$. For example,

$$u \preceq_K v,\ x \preceq_K y \implies u + x \preceq_K v + y.$$

Some are not: in general, $\preceq_K$ is not a **linear ordering** (possible for $x \not\preceq_K y$ and $y \not\preceq x$.)

Minimum and Minimal Elements

$x \in S$ is **the minimum element** of S wrt $\preceq_K$ if

$$y \in S \implies x \preceq_k y.$$

$x \in S$ is **a minimal element** of S wrt $\preceq_K$ if

$$y \in S,\ y \preceq_k x \implies y = x.$$

Unambiguous ordering is defined only for $x, y \in \{K \cup -K\}$, so for regions outside this set, the ordering is ambiguous.

Roughly, minimum if all other points are more, minimal if no other points are less.

Separating Hyperplane Theorem

If C, D are disjoint convex sets, then $\exists a \neq 0,\ b$ s.t.

$$a^T x \leq b \quad \forall x \in C \qquad \text{and} \qquad a^T x \geq b \quad \forall x \in D.$$

Hyperplane separates space into two halfspaces, each containing either C or D.

Strict separation requires closed C and singleton D.

Supporting Hyperplane Theorem

Suppose we have a point x_0 on the boundary of a set C. If $a \neq 0$ and $a^T x \leq a^T x_0\ \forall x \in C$, then

$$\{x \mid a^T x = a^T x_0\}$$

is the **supporting hyperplane** to set C at boundary point x_0.

I.e., hyperplane separates x_0 and C.

Hyperplane is tangent to C at x_0.

Supporting hyperplane theorem: If C is convex and nonempty, then $\exists$ a supporting hyperplane $\forall$ boundary point x_0.

Dual Cones and Generalized Inequalities

Dual Cones

The **dual cone** K^* of a cone K is the set

$$\{y \mid x^T y \geq 0 \quad \forall x \in K\}.$$

Equivalently, it is the set of y s.t. y is a normal vector to a supporting hyperplane of K at the origin 0.
-I.e., the set of all vectors within 90 degrees of all vectors in K.

K^* is always a convex cone, even if K is not convex.

K^* of a subspace $V \subseteq \mathbb{R}^n$ (which is a cone) is the orthogonal complement of V.

$$K^* = \{y \mid v^T y = 0 \quad \forall v \in V\}.$$

Examples

- The nonneg orthant is **self-dual**.
- The PSD cone is self-dual. The **standard inner product of two matrices** X, Y is

$$\mathbf{Tr}(XY) = \sum_{i=1}^{n} \sum_{j=1}^{n} X_{ij} Y_{ij}.$$

- The 2-norm cone is self-dual.
- For the 1-norm cone, the dual cone is the infinity-norm cone.

Properties

- $K*$ is convex and closed.
- If $K1 \subseteq K2$, then $K2^* \subseteq K1^*$.
- If K has a nonempty interior, then K^* is pointed.
- If the closure of K is pointed, then K^* has a nonempty interior.
- K^{**} is the closure of the convex hull of K (So, if K is convex, then $K^{**} = K$).

This properties imply that if K is a proper cone, then K^* is a proper cone.

Dual Generalized Inequalities

If K is a proper cone, it induces a generalized inequality $\preceq_K$, and $K*$ is a proper cone.

So, K^* induces a generalized inequality $\preceq_{K^*}$, which we refer to as the **dual of** $\preceq_K$.

Some properties relating a generalized inequality and its dual are

- $x \preceq_K y \quad \iff \quad \lambda^T x \preceq_K \lambda^T y \quad \forall \lambda \succeq_{K^*} 0.$
- Similar for strict generalized inequalities
- Similar for flipped K and K^* generalized inequalities (because $K = K^{**}$ when K is a proper cone).

Minimum and Minimal Elements via Dual Inequalities

Dual Characterization of Minimum Element

x is **the minimum element** of S wrt $\preceq_K$ if and only if for all $\lambda \succ_{K^*} 0$, x is the unique minimizer for $\lambda^T z$ over all $z \in S$.

Equivalently, the hyperplane $\{z \mid \lambda^T(z - x) = 0\}$ is a **strictly supporting hyperplane** at x.
-Strictly supporting at x means that it intersects S at only x.

S does not have to be convex.

Dual Characterization of Minimal Elements

x is **a minimal element** of S wrt $\preceq_K$ if x minimizes $\lambda^T z$ over all $z \in S$ for some $\lambda \succ_{K^*} 0$.
-I.e., for some $\lambda \succ_{K^*} 0$, the (biased) hyperplane orthogonal to λ is tangent to S.

S does not have to be convex.

If S is convex, then for any minimal element x_i of S (note: not the dual characterization), there exists a nonzero $\lambda_i \succ_{K^*} 0$ s.t. each x_i minimizes $\lambda_i^T z$ over $z \in S$.
-Generally not true if S is not convex.

Optimal Production Frontier

Efficient (**Pareto optimal**) solutions are minimal wrt $\boldsymbol{R}_+^n$.

Lecture 3 (Chapter 3 – Convex Functions)

Definition

A function $f : \mathbb{R}^n : \mathbb{R}$ is **convex** if its domain is a convex set and

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$

for every $x, y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$.

Or, any chord of the graph lies above the graph (except at the end points).

f is **strictly convex** if the inequality is strict and it holds for every $x, y \in \mathbf{dom}\, f,\ x \ne y,\ 0 < \theta < 1$.

f is **concave** if $-f$ is convex.

Examples on R

Convex:

- Affine: $ax + b$ on $\mathbb{R}$ for any $a, b \in \mathbb{R}$
  –Equality holds

- Exponential: e^{ax} for any $a \in \mathbb{R}$
- Powers: x^p on $\boldsymbol{R}_{++}$, for any $p \leq 0$ or $p \geq 1$
- Absolute powers: $|x|^p$ on $\mathbb{R}$, for any $p \in \mathbb{R}$
- Negative log entropy: $x\log x$ on $\boldsymbol{R}_{++}$

Concave:

- Affine
- Powers, $p \in [0, 1]$
- Logarithm on $\boldsymbol{R}_{++}$

Examples on $\mathbb{R}^n$ and $\mathbb{R}^{m\mathbf{x}n}$

$\mathbb{R}^n$:

- Affine functions: $f(x) = a^T x + b$
- Norms ($p \geq 1$ for p-norms)
  - $p < 1$ of interest for sparsity

$\mathbb{R}^{mxn}$:

- Affine functions: $f(X) = \mathbf{Tr}(A^T X) + b$
  - $\mathbf{Tr}(A^T X) =$ **standard inner product** of A and X.
- **Spectral norm**: $\|X\|_\sigma = \sigma_{max} X = (\lambda_{max}(X^T X))^{1/2}$

Restriction of a Convex Function to a Line

Theorem: $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if $g : \mathbb{R} \to \mathbb{R}$

$$g(t) = f(x + tv), \qquad \mathbf{dom}\, g = \{t | x + tv \in \mathbf{dom}\, f\}$$

is convex for any $x \in \mathbf{dom}\, f,\ v \in \mathbb{R}^n$, and $t \in \mathbf{dom}\ g$.

$\implies$ can check convexity of function on $\mathbb{R}^n$ by checking convexity of functions of $\mathbb{R}$.

Super interesting example: $f : \mathbb{R}^n \to \mathbb{R}$ with $f(X) = \log \det(X)$

So, domain of f restricted to $\boldsymbol{S}^n_{++}$.

To show that f is convex (or concave), we must show that for any $X \in \boldsymbol{S}^n_{++}$ and $V \in \boldsymbol{S}^n$

$$g(t) = \log \det(X + tV)$$

is convex (or concave) in t.

$$\begin{aligned}
g(t) &= \log \det(X + tV) \\
&= \log \det(X^{1/2}(I + tX^{-1/2}VX^{-1/2})X^{1/2}) \\
&= \log \det(X(I + tX^{-1/2}VX^{-1/2})) \\
&= \log \det(X) + \log \det(I + tX^{-1/2}VX^{-1/2}) \\
&= \log \det(X) + \log \det(Q^TQ + tQ^T\Lambda Q) \quad \text{(by orthogonal diagonalization)} \\
&= \log \det(X) + \log \det(Q^T(I + t\Lambda)Q) \\
&= \log \det(X) + \log \det(Q^TQ(I + t\Lambda)) \\
&= \log \det(X) + \log \det(I + t\Lambda) \\
&= \log \det(X) + \sum_{i=1}^{n} \log(1 + t\lambda_i)
\end{aligned}$$

where $\lambda_i \geq 0$ because they are the eigenvalues of $X^{-1/2}VX^{-1/2}$, which is symmetric PSD (any matrix of the form B^TAB is PSD).

$\log(1+t\lambda_i)$ is concave, and the sum of concave functions is itself concave. $\log \det(X) \in \boldsymbol{R}_{++}$, so g is concave in t, which implies that f is concave.