Convex Optimization: Introduction and Basic Terminologies

A supplementary note to Chapter 2 of *Convex Optimization* by S. Boyd and L. Vandenberghe

Optimization Lab.

IE department
Seoul National University

22nd July 2009

Optimization

$$
\begin{array}{lll l}
\min & f_0(x) & & \text{“Objective”} \\
\text{sub. to} & f_i(x) \leq b_i & i = 1, \ldots, m, & \text{“Constraints”}
\end{array} \tag{1}
$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 0, 1, \ldots, m$.

Definition

Vector x is called feasible if it satisfies all constraints. A feasible solution x^* is called optimal if its objective value is minimum: $f_0(x^*) \leq f_0(x)$ for all feasible x.

- Tractability of (1), namely possibility of an *efficient* solution method for (1), depends on characteristics of f_i's.
- In general, easy to devise a problem whose feasibility problem is believed to have no efficient method.

Convex optimization

Definition

Optimization (1) is convex if f_i's are all convex: $\forall\ x, y\ \forall$ $0 \leq \lambda \leq 1$, *we have*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \tag{2}$$

"Function value of a convex combination of any two points is no greater than the same convex combination of the two function values."

Convex optimization. Why care?

- Easy!

  "In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity." - Rockafellar

  We can find global optima in polynomial time of input sizes of problem size and numerical accuracy (modulo some technical conditions).

- Prevalent!

  We are discovering new applications that can be formulated as a convex optimization problem. Especially, it contains conic programs such as linear programs, second-order cone programs, and semidefinite programs.

The goals

1. To develop the skills and background needed to recognize, formulate, and solve convex optimization problems.
2. To perform in-depth review on how conic programs offer tight relaxations of NP-hard combinatorial optimization problems to yield better approximation algorithms.
3. To survey convex optimization problems discovered recently in such areas as control, signal processing, circuit design, data modeling, and finance, and, more ambitiously, to discover new such practical problems.

Lines and affine sets

- A line through x_1 and x_2 is defined to a set of points y such that $y = (1-\lambda)x_1 + \lambda x_2 = x_1 + \lambda(x_2 - x_1)$ where $\lambda \in \mathbb{R}$. Thus, y is the sum of the base point x_2 and the direction $x_1 - x_2$ scaled by λ.
- An affine set is defined to be a set that contains the line through any two distinct points in the set: for any x_1 , $x_2 \in C$, with $x_1 \neq x_2$, and $\lambda \in \mathbb{R}$, we have $\lambda x_1 + (1-\lambda)x_2 \in C$. (Extendible to an equivalent definition in terms of a finite number of points.)
- If C is affine, then for *any* $x_0 \in C$, $V = C - x_0 = \{x - x_0 | x \in C\}$ is a subspace as closed for scalar multiplication and addition: $\forall\ x, y \in C$ and $\forall \lambda \in \mathbb{R}$, $\lambda(x - x_0) = \lambda x + (1-\lambda)x_0 - x_0 \in C - x_0$, and $x - x_0 + y - x_0 = 2(\frac{1}{2}x + \frac{1}{2}y - x_0) \in C - x_0$. Thus, an affine set C is a translation of a subspace V, $C = V + x_0$. The *dimension* of C, dim C is defined as the dimension of V.

Lines and affine sets

- An affine combination of the points $x_1, \dots, x_k$ is a point of the form: $\lambda_1 x_1 + \cdots + \lambda_k x_k$ with $\sum_{i=1}^k \lambda_i = 1$.
- The affine hull of a set C, denoted by $\mathrm{aff}\ C$, is defined to be the set of all affine combinations of points in C. Thus, $\mathrm{aff}\ C$ is the smallest affine set that contains C.
- The *affine dimension* of a set C is defined as $\dim(\mathrm{aff}\ C)$.
  e.g. $C = \{x \in \mathbb{R}^2 | x_1^2 + x_2^2 = 1\} \Rightarrow \dim(\mathrm{aff}\ C) = 2$, while the "dimension" of C is < 2 in usual senses.
- Relative interior:

$$x \in \mathrm{relint}\ C \ \Leftrightarrow\ \exists \delta > 0\ s.t.\ B(x, \delta) \cap \mathrm{aff}\ C \subseteq C.$$

Affine sets: Some exercises

- $x_2 - x_1$, $x_3 - x_1$, ..., $x_n - x_1$ are linearly independent $\Leftrightarrow$ $x_1 - x_2$, $x_3 - x_2$, ..., $x_n - x_2$ are linearly independent $\Leftrightarrow \cdots \Leftrightarrow$ $x_1 - x_n$, $x_2 - x_n$, ..., $x_{n-1} - x_n$ are linearly independent.
- In this case, we say x_1, x_2, ..., x_n are *affinely independent*.
- When an affine set C contains the origin the maximum number of affinely independent points in C is one plus the maximum number of linearly independent points in C. Otherwise they are the same.
- The dimension of affine set C is one less than maximum number of affinely independent points in C.

Line segments and convex sets

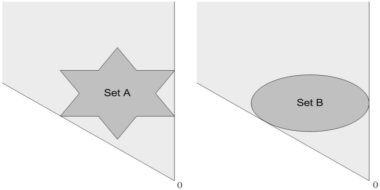
- The line segment between x_1 and x_2 is the set of points $y = (1-\lambda)x_1 + \lambda x_2 = x_1 + \lambda(x_2 - x_1)$ where $0 \leq \lambda \leq 1$.
- A set is called *convex* if it contains the line segment between any two points in the set: for any $x_1, x_2 \in C$, and for any $0 \leq \lambda \leq 1$, we have $(1-\lambda)x_1 + \lambda x_2 \in C$.
  1. Finitely many points: If C is convex, then $x_1, \ldots, x_k \in C, \sum_i \lambda_i = 1, \lambda_i \geq 0 \ \Rightarrow \ \sum_i \lambda_i x_i \in C$.
  2. Countably many points: If C is convex, then $\{x_i\} \subseteq C$, $\sum_{i=1}^{\infty} \lambda_i = 1$, $\lambda_i \geq 0$, $\sum_{i=1}^{\infty} \lambda_i x_i$ convergent $\Rightarrow$ $\sum_{i=1}^{\infty} \lambda_i x_i \in C$.

Line segments and convex sets

- A *convex combination* of $x_1, \ldots, x_k$ is a point of the form: $\lambda_1 x_1 + \cdots + \lambda_k x_k$ with $\sum_{i=1}^{k} \lambda_i = 1, \forall \lambda_i \geq 0$.
- The *convex hull* of a set C, denoted by $\mathrm{conv}\, C$, is defined to be the set of convex combinations of points from C. Thus, $\mathrm{conv}\, C$ is the smallest convex set that contains C.
- The *dimension* of a convex set is defined to be its affine dimension.

Cones

- A *cone* is a set closed under scalar multiplication: $\forall x \in C$ and $\forall \lambda \geq 0$, $\lambda x \in C$.
- A *convex cone* is a set which is convex as well as a cone.
- A *conic combination* of points $x_1, \ldots, x_k$ is a point of the form: $\lambda_1 x_1 + \cdots + \lambda_k x_k$ with $\lambda_i \geq 0 \; \forall i$.
- The *conic hull* of a set C, $\mathrm{cone}C$ is defined to be the set of conic combinations of points in C. Thus, $\mathrm{cone}C$ is the smallest convex cone containing C.
- We say $\mathrm{cone}C$ is *finitely generated* if $|C| < \infty$.



Examples

- The empty set $\emptyset$, any single point $\{x_0\}$, and $\mathbb{R}^n$ are affine (hence convex).
- Any line is affine. If it passes through zero, then it is a subspace (hence a convex cone).
- A ray, $\{x_0 + \lambda v | \lambda \geq 0\}$, where $v \neq 0$, is convex, but not affine.
- Any subspace is affine, and a convex cone.

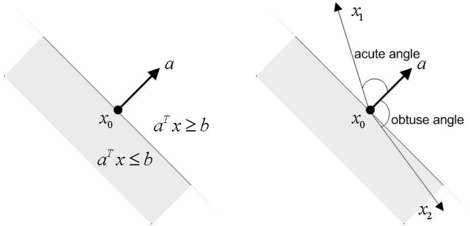
Hyperplanes and halfspaces

Definition

A hyperplane is a set $H = \{x : a^T x = b\}$*, where* $a \in \mathbb{R}^n$, $a \neq 0$, *and* $b \in \mathbb{R}$.

For any $x_0 \in H$, $\{x | a^T(x - x_0) = 0\} = x_0 + a^{\perp}$, where $a^{\perp} = \{v | a^T v = 0\}$

A (closed) halfspace is a set of the form $\{x : a^T x \leq b\}$, where $a \neq 0$.



Euclidean balls and ellipsoids

Definition

A (Euclidean) ball in $\mathbb{R}^n$ is

$$B(x_c, r) = \{x : \|x - x_c\|_2 \leq r\} = \{x_c + ru : \|u\|_2 \leq 1\}$$

where $r > 0$, and $\|\cdot\|_2$ denotes the Euclidean norm. The vector x_c is the center of the ball and the scalar r is its radius.

Definition

An ellipsoid is

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

where $P = P^T \succ 0$. (Notice it is a ball with radius r when $P = r^2 I$.)

The length of axes are $\sqrt{\lambda_i}$ where λ_i are eigenvalues of P. The triangle property of the norm implies the convexity of an ellipsoid.

Euclidean balls and ellipsoids

Theorem

An ellipsoid can be represented as

$$\{Au + b : \|u\|_2 \leq 1\}$$

where A is nonsingular.

Proof. For $u \in B(0,1)$, let $x = Au + b$ or $u = A^{-1}(x-b)$. Then,

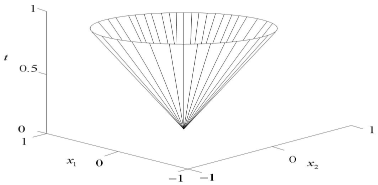
$$u^T I u \leq 1 \Leftrightarrow (x-b)^T (A^{-1})^T A^{-1}(x-b) = (x-b)^T (A^{-1})^T A^{-1}(x-b) \leq 1$$

By denoting $(A^{-1})^T A^{-1}$, symmetric and positive-definite, by P^{-1}, we get

$$(x-b)^T P^{-1}(x-b) \leq 1. \quad \square$$

Norm balls and norm cones

- A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a *norm* if
  1. nonnegative: $f(x) \geq 0, \forall x$,
  2. definite: $f(x) = 0$ only if $x = 0$,
  3. homogeneous: $f(tx) = |t| f(x), \forall x, \forall t \in \mathbb{R}$, and
  4. satisfies triangle inequality: $f(x+y) \leq f(x) + f(y)$, $\forall x, y$.
- A *norm ball* of radius r and center x_c is $B_{\|\cdot\|}(x_c, r) = \{x : \|x - x_c\| \leq r\}$.
- The *norm cone* is $C = \{(x,t) : \|x\| \leq t\} \subseteq \mathbb{R}^{n+1}$.



Polyhedra

Definition

A polyhedron P is the intersection of a finite number of halfspaces:

$$P = \{x | a_i^T x \leq b_i,\ i = 1, ..., m\} = \{x | Ax \leq b\}.$$

Definition

A simplex C is the convex hull of a set of affinely indep vectors:

$$C = conv\{v_0, ..., v_k\} = \{\lambda_0 v_0 + ... + \lambda_k v_k | \lambda \geq 0, \mathbf{1}^T \lambda = 1\}$$

where $v_0, ..., v_k \in \mathbb{R}^n$ *are affinely independent.*

Fundamental theorem of linear inequalities

Theorem

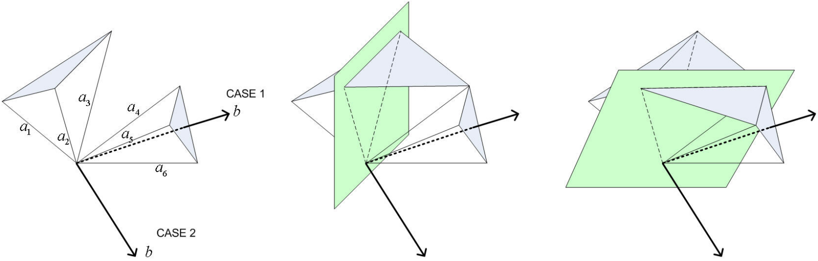
Let $a_1, a_2, \ldots, a_m$ $b \in \mathbb{R}^n$. Then, either $b \in cone\{a_1, a_2, \ldots, a_m\}$ or there is a hyperplane $c^T x = 0$ containing $t-1$ independent vectors from $a_1, a_2, \ldots, a_m$ such that $c^T b < 0$ and $c^T a_i \geq 0$ for $i = 1, \ldots, m$, where $t = rank\{a_1, a_2, \ldots, a_m, b\}$. But never both.

Proof May assume $t = n$. Consider any basis $B = \{a_{i_1}, a_{i_2}, \ldots, a_{i_n}\}$ of n independent vectors from $a_1, a_2, \ldots, a_m$.

1. If $b = B\lambda$ with $\lambda \geq 0$. Then we are in the first case. Done.
2. Choose smallest h among $i_1, i_2, \ldots, i_n$ such that $\lambda_h < 0$. Let $c^T x = 0$ be hyperplane generated by $D - \{a_h\}$. Normalize c so that $c^T a_h = 1$. (Then $c^T b = \lambda_h < 0$.)
3. If $c^T a_i \geq 0$ for $i = 1, \ldots, m$, then we are in the second case. Done.
4. Otherwise, choose smallest s with $c^T a_s < 0$ and $D \leftarrow (D - \{a_h\}) \cup \{a_l\}$ and repeat.

Fundamental theorem of linear inequalities (*cont'd*)

Feasibility simplex method



Fundamental theorem of linear inequalities (*cont'd*)

Proof(*cont'd*) Suffices to show this process terminates. Denote, by B_k, the basis B in kth iteration. If process does not terminate, then there is $k < l$ with $B_k = B_l$. Let r be largest index for which a_r has been removed from D at the end of one of iterations k, $k+1$, ..., $l-1$. Then a_r must have been added back to B. Thus notice that

$$B_p \cap \{a_{r+1}, \ldots, a_m\} = B_q \cap \{a_{r+1}, \ldots, a_m\}. \tag{3}$$

Let $B = \{a_{i_1}, a_{i_2}, \ldots, a_{i_n}\}$ and $b = \lambda_{i_1} a_{i_1} + \cdots + \lambda_{i_n} a_{i_n}$. Let c' be the vector c in Step 2 of iteration q. Then we have the following contradiction:

$$0 > (c')^T b = (c')^T(\lambda_{i_1} a_{i_1} + \cdots + \lambda_{i_n} a_{i_n}) = \lambda_{i_1}(c')^T a_{i_1} + \cdots + \lambda_{i_n}(c')^T a_{i_n} > 0.$$

The last inequality follows from that

- if $i_j < r$ then $\lambda_{i_j} \geq 0$ and $(c')^T a_{i_j} \geq 0$;
- if $i_j = r$ then $\lambda_{i_j} < 0$ and $(c')^T a_{i_j} < 0$, and
- if $i_j > r$ then $(c')^T a_{i_j} = 0$ (from (3) and def of c'). $\square$

Polyhedra(*cont'd*)

Farkas-Minkowski-Weyl theorem

Corollary

A cone C is polyhedral, i.e. $C = \{x | Ax \leq 0\}$ for some $A \in \mathbb{R}^{m \times n}$ if and only if C is finitely generated.

Proof $(\Leftarrow)$ Let $C = \mathrm{cone}\{x_1, x_2, \ldots, x_m\}$ with $x_i \in \mathbb{R}^n$. May assume $\mathrm{span}\{x_1, x_2, \ldots, x_m\} = \mathbb{R}^n$ as we can extend a halfspace H of $\mathrm{span}\{x_1, x_2, \ldots, x_m\}$ to a halfspace H' of $\mathbb{R}^n$ so that $H = H' \cap \mathrm{span}\{x_1, x_2, \ldots, x_m\}$.
From Fundamental theorem, for any $y \notin \mathrm{cone}\{x_1, x_2, \ldots, x_m\}$ there is a separating hyperplane $c^T x = 0$ containing $n - 1$ independent vectors from x_1, x_2, ..., x_m. Since there are only finite such combinations, C is the intersection of finite number of corresponding halfspaces, namely polyhedral.

Polyhedra(*cont'd*)

Farkas-Minkowski-Weyl theorem

Proof (*cont'd*) ($\Rightarrow$) Let $C = \{x | a_i^T x \leq 0,\ i = 1, \ldots, m\}$. Consider $\mathrm{cone}\{a_1, a_2, \ldots, a_m\}$ which is, by the above, polyhedral: there are b_1, b_2, ..., b_t such that $\mathrm{cone}\{a_1, a_2, \ldots, a_m\} = \{x | b_i^T x \leq 0,\ i = 1, \ldots, t\}$. Our claim is $C = C' := \mathrm{cone}\{b_1, b_2, \ldots, b_t\}$. First notice that $C' \subseteq C$ as $b_i \in C$ (since $b_i^T a_j \leq 0$).
To establish $C' \supseteq C$, suppose $y \notin C'$. Since $C' := \mathrm{cone}\{b_1, b_2, \ldots, b_t\}$ is polyhedral, Fundamental theorem implies $\exists$ w such that $w^T b_i \leq 0\ \forall\ i$ and $w^T y > 0$. Hence $w \in \mathrm{cone}\{a_1, a_2, \ldots, a_m\}$ and $w^T x \leq 0$ for all $x \in C$. Since $w^T y > 0$ we have $y \notin C$. $\square$

Polyhedra(*cont'd*)

We call a finitely generated convex hull *polytope*.

Corollary

A set $P \subseteq \mathbb{R}^n$ is polyhedral if and only if $P = Q + C$ for some polytope Q and finitely generated cone C.

Proof ($\Rightarrow$) Say $P = \{x | Ax \leq b\}$. Then consider homogenized cone

$$\left\{ \begin{pmatrix} x \\ \lambda \end{pmatrix} | x \in \mathbb{R}^n,\ \lambda \in \mathbb{R},\ Ax - \lambda b \leq 0 \right\}.$$

Polyhedra

Example

Octahedron in $\mathbb{R}^3$

$$\left.\begin{array}{rcl} x_1+x_2+x_3 & \leq & 1 \\ -x_1+x_2+x_3 & \leq & 1 \\ \vdots & \vdots & \vdots \\ -x_1-x_2-x_3 & \leq & 1 \end{array}\right\} 8 \textit{ equations vs } 6 \textit{ points} = conv\{\pm e_1, \pm e_2, \pm e_3\}$$

Need 2^n equations for polyhedral description but $2n$ points in conv hull description

Example

Cube in $\mathbb{R}^3$

$$\left.\begin{array}{rcl} x_1 & \leq & 1 \\ -x_1 & \leq & 1 \\ \vdots & \vdots & \vdots \\ -x_3 & \leq & 1 \end{array}\right\} 6 \textit{ equations vs } 8 \textit{ points} = conv\{(1,1,1),(-1,1,1),\dots,(-1,-1,-1)\}$$

Need $2n$ equations in polyhedral edescription, but 2^n points for convex hull description.

Polyhedral representation of a simplex

Suppose that C is a simplex defined by the affinely indep points $v_0, \ldots, v_k$.

$$\begin{aligned} x \in C \quad & \Leftrightarrow x = \lambda_0 v_0 + \lambda_1 v_1 + \cdots + \lambda_k v_k \ \text{ for some } \lambda \text{ s.t } \forall \lambda_i \geq 0, \sum \lambda_i = 1 \\ & \Leftrightarrow x = v_0 + \lambda_1(v_1 - v_0) + \cdots + \lambda_k(v_k - v_0) \\ & \Leftrightarrow x = v_0 + A\lambda' \end{aligned}$$

where $A = \begin{bmatrix} | & | & & | \\ v_1 - v_0 & v_2 - v_0 & \cdots & v_k - v_0 \\ | & | & & | \end{bmatrix}$, $\lambda' = [\lambda_1, \ldots, \lambda_k], 1^T\lambda' \leq 1$.

Since $v_0, \ldots, v_k$ are affinely independent,

$$\begin{aligned} \operatorname{rank}(A) = k \quad & \Leftrightarrow \text{By row operations, we can reduce } A \text{ to } \begin{bmatrix} I_k \\ 0 \end{bmatrix} \\ & \Leftrightarrow \exists B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \in \mathbb{R}^{n \times n} \text{ s.t.} BA = \begin{bmatrix} I_k \\ 0 \end{bmatrix} \end{aligned}$$

Polyhedral representation of a simplex

$$
\begin{aligned}
& x = v_0 + A\lambda', \lambda' \geq 0, 1^T\lambda' \leq 1 \\
& \Leftrightarrow Bx = Bv_0 + BA\lambda', \lambda' \geq 0, 1^T\lambda' \leq 1 \\
& \Leftrightarrow Bx = Bv_0 + \begin{bmatrix} I_k \\ 0 \end{bmatrix} \lambda', \lambda' \geq 0, 1^T\lambda' \leq 1
\end{aligned}
$$

Thus,

$$
\begin{aligned}
x \in C \quad & \Leftrightarrow B_1 x = B1 v_0 + \lambda', B_2 x = B_2 v_0, \lambda' \geq 0, 1^T\lambda' \leq 1 \\
& \Leftrightarrow B_2 x = B_2 v_0, B_1 x \geq B_1 v_0, 1^T B_1 x \leq 1 + 1^T B_1 v_0
\end{aligned}
$$

The positive semidefinite cone

- Some sets of matrices:

$$\mathbb{R}^{n\times n} \xrightarrow{symmetry} \mathbb{S}^n \xrightarrow{PSDness} \mathbb{S}^n_+ \xrightarrow{PDness} \mathbb{S}^n_{++}.$$

- $\mathbb{S}^n_+$ is a convex cone: for nonnegative $\alpha, \beta \in \mathbb{R}$ and $M, N \in \mathbb{S}^n_+$,

$$x^T(\alpha M + \beta N)x = \alpha x^T Mx + \beta x^T Nx \geq 0.$$

- Positive semidefinite cone in $\mathbb{S}^2$

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \Longleftrightarrow x \geq 0, z \geq 0, xz \geq y^2.$$

Intersection

Convexity is preserved under intersection:

S_α is convex for $\alpha \in A \Rightarrow \cap_{\alpha \in A} S_\alpha$ is convex.

Example:

- $\mathbb{S}^n_+ = \cap_{z \neq 0} \{X \in \mathbb{S}^n : z^T X z \geq 0\}$ where $\{X \in \mathbb{S}^n | z^T X z \geq 0\}$ is a linear function of X.
- Let $p(t) = \sum_{k=1}^m x_k \cos kt$. Then,

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

can be expressed as $S = \bigcap_{|t| \leq \pi/3} S_t$ where

$$S_t = \{x | -1 \leq (\cos t, ..., \cos mt)^T x \leq 1\}.$$

Affine functions

Definition

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is affine if it is a sum of linear function and a constant, i.e., if it has the form $f(x) = Ax + b$, where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

The image and inverse image of a convex set under an affine function f is convex:

- $S \subseteq \mathbb{R}^n$ *convex* $\Rightarrow f(S) = \{f(x)|x \in S\}$ is convex.
- $C \subseteq \mathbb{R}^m$ *convex* $\Rightarrow f^{-1}(C) = \{x \in \mathbb{R}^n|f(x) \in C\}$ is convex.

Affine functions(*cont'd*)

Example

1. *Scaling and translation preserve convexity.*
2. *So does a projection:* $[\mathrm{I}_m \vdots O_n]\begin{bmatrix} x \\ y \end{bmatrix} = x$, *where* $x \in \mathbb{R}^m, y \in \mathbb{R}^n$
3. S_1, S_2 *convex* $\Rightarrow$ *So are their sum* $S_1 + S_2$ *and product* $S_1 \times S_2 := \{(x_1, x_2)|x_1 \in S_1, x_2 \in S_2\}$,
4. *and partial sum,* $S := \{(x, y_1 + y_2)|(x, y_1) \in S_1, (x, y_2) \in S_2\}$.

Affine functions(*cont'd*)

Example

$$\begin{aligned} \textit{Polyhedron} &= \{x | Ax \leq b, Cx = d\} \\ &= \{x | f(x) \in \mathbb{R}^m_+ \times \{0\}\} \end{aligned}$$

where $f(x) = (b - Ax, d - Cx)$

Example

$$\begin{aligned} \textit{Ellipsoid} &= \{x | (x - x_c)^T P^{-1} (x - x_c) \leq 1\}, P \in \mathbb{S}^n_{++} \\ &= \textit{the image of } \{u | \|u\|_2 \leq 1\} \\ &\quad \textit{under the affine mapping } f(u) = P^{1/2} u + x_c \\ &= \textit{the inverse image of unit ball} \\ &\quad \textit{under the affine mapping } g(x) = P^{-1/2}(x - x_c) \end{aligned}$$

Linear-fractional and perspective functions

Definition

A perspective function $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, *with* dom $P = \mathbb{R}^n \times \mathbb{R}_{++}$ *is defined as*

$$P(z, t) = z/t.$$

- If $C \subseteq \mathrm{dom} P$ is convex, $P(C)$ is convex.
- If $C \subseteq \mathrm{R}^n$ is convex. $P^{-1}(C)$ is convex.

Linear-fractional and perspective functions

Definition

Let $g : \mathbb{R}^n \to \mathbb{R}^{m+1}$ *be affine, namely,* $g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}$. *The function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *given by* $f = P \circ g$, *i.e.,*

$$f(x) = (Ax + b)/(c^T x + d), \quad \mathrm{dom} f = \{x : c^T x + d > 0\},$$

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

- Both image and inverse image of a convex set under linear-fractional are convex.

$$x \times \{1\} \xrightarrow[\text{affine mapping}]{} \begin{bmatrix} A & b \\ C^T & d \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} \xrightarrow[\text{scaling}]{} \begin{bmatrix} \frac{Ax+b}{c^T x+d} \\ 1 \end{bmatrix} \xrightarrow[\text{projection}]{} \frac{Ax+b}{c^T x+d}$$

Proper cones and generalized inequalities

Definition

A cone $K \subseteq \mathbb{R}^n$ is called a proper cone if

1. *K is convex and closed,*
2. *K is solid, or K has nonempty interior,*
3. *K is pointed, or K contains no line.*

Example :

a. Nonnegative orthant, $K = \mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0, i = 1, \ldots, n\}$.

b. Positive semidefinite cone, $K = \mathbb{S}^n_+$.

c. Nonnegative polynomials in $[0, 1]$

$$K = \{x \in \mathbb{R}^n : x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1} \geq 0 \text{ for } t \in [0, 1]\}.$$

K is proper and its interior is the set of coeffi. of polynomials positive on $[0, 1]$ (from uniform continuity).

Proper cones and generalized inequalities

A proper cone K can be used to define a generalized inequality, which is a partial ordering on $\mathbb{R}^n$:

$$x \preceq_K y \leftrightarrow y - x \in K, x \prec_K y \leftrightarrow y - x \in \mathrm{int}K.$$

Example:

- Let K be the nonnegative orthant, that is $K = \mathbb{R}^n_+$. Then,

$$x \preceq_K y \Leftrightarrow x_i \leq y_i, i = 1, \ldots, n$$

- Let K be the PSD cone, that is $K = \mathbb{S}^n_+$. Then,

$$X \preceq_K Y \Leftrightarrow Y - X \in \mathbb{S}^n_+$$

Proper cones and generalized inequalities

Generalized inequality $\preceq_K$ satisfies

1. $x \preceq_K y, u \preceq_K v \Rightarrow x + u \preceq_K y + v$,
2. $x \preceq_K y, y \preceq_K z \Rightarrow x \preceq_K z$,
3. $x \preceq_K y, \alpha \geq 0 \Rightarrow \alpha x \preceq_K \alpha y$,
4. $x \preceq_K x$,
5. $x \preceq_K y, y \preceq_K x \Rightarrow x = y$, and
6. $x_i \preceq_K y_i$ for $i = 1, \ldots,$ and $x_i \to x, y_i \to y$ as $i \to \infty \Rightarrow x \preceq_K y$.

Strict generalized inequality $\prec$ satisfies

1. $x \prec_K y \Rightarrow x \preceq_K y$,
2. $x \prec_K y, u \preceq_K v \Rightarrow x + u \prec_K y + v$,
3. $x \prec_K y, \alpha > 0 \Rightarrow \alpha x \preceq_K \alpha y$,
4. $x \nprec_K x$, and
5. $x \prec_K y \Rightarrow x + u \prec_K y + v$ for small enough u and v.

Minimum and minimal elements

In general, $\preceq_K$ is not a linear ordering: there can be x and y such that $x \npreceq_K y$ and $y \npreceq_K x$.

Definition

A point $x \in S$ is the minimum element of S w.r.t. $\preceq_K$ if $x \preceq_K y, \forall y \in S$. Equivalently, $x \in S$ is the minimum element iff $S \subseteq x + K$. A point $x \in S$ is a minimal element of S w.r.t. $\preceq_K$ if $y \in S, y \preceq_K x \Rightarrow y = x$. Equivalently, $x \in S$ is a minimal element iff $(x - K) \cap S = \{x\}$.

Separating hyperplane theorem

Theorem

Suppose C and D are disjoint convex sets. Then, $\exists a \neq 0,\ b$ s.t.

$$a^T x \leq b, \forall x \in C, a^T x \geq b, \forall x \in D.$$

Then $\{x | a^T x = b\}$ is called a separating hyperplane for C and D.
In some cases, a *strict* separation can be established: Let C be a closed convex set and $x_0 \notin C$. Then there exists a hyperplane $\{x | a^T x = b\}$ that strictly separates x_0 from C, namely, $a^T x \leq b$ for every $x \in C$ and $a^T x_0 > b$.

Separating hyperplane theorem(*cont'd*)

Case 1 : C is compact, and D is closed.
Define $\mathrm{dist}(C, D) = \inf\{\|u - v\|_2 | u \in C, v \in D\}$. Then, $\exists c \in C$ and $d \in D$ s.t. $\|c - d\|_2 = \mathrm{dist}(C, D)$ and $\|c - d\|_2 > 0$. (Argue that we may also assume D is also bounded as far as such d is concerned.)
Consider hyperplane $f(x) = (d - c)^T(x - \frac{c+d}{2}) = 0$ so that $a := d - c$ and $b := \frac{\|d\|_2^2 - \|c\|_2^2}{2}$. Need to show f is nonpositive on C and nonnegative on D.
Suppose not: $\exists u \in D$ s.t. $f(u) = (d - c)^T(u - \frac{d+c}{2}) < 0$.
Consider the distance between c and the points on line segment from d to u, $\|d + \lambda(u - d) - c\|_2$ and its derivative w.r.t. λ, $\frac{1}{2}\frac{d}{d\lambda}\|d + \lambda(u - d) - c\|_2^2\big|_{\lambda=0}$
$= (d - c)^T(u - d) = (d - c)^T(u - \frac{d+c}{2}) - \frac{1}{2}(d - c)^T(d - c) < 0$.
It means that the point $d + \lambda(u - d)$ is closer to c than d for sufficiently small $\lambda > 0$. A contradiction.

Separating separating hyperplane(*cont'd*)

Case 2 : C is convex and $D = \{d\}$, with $d \notin \operatorname{int} C$

Subcase 1 $d \notin \mathrm{cl}C$. As the point d itself is compact, from the Case 1, and we can separate d from $\mathrm{cl}C$ and hence also from C.

Subcase 2 $d \in \mathrm{bd}C$. Then $\exists\{d^n\}$ with $d^n \notin \mathrm{cl}C$, which converges to d. Thus, from Subcase 1, $\exists \begin{bmatrix} a^1 \\ b^1 \end{bmatrix}, \ldots, \begin{bmatrix} a^n \\ b^n \end{bmatrix}, \ldots,$ which satisfy $(a^n)^T x \leq b^n$ and $(a^n)^T d^n \leq b^n$ for all n. Taking $\left\| \begin{bmatrix} a^n \\ b^n \end{bmatrix} \right\|_2 = 1$ guarantees a convergent subsequence: $\begin{bmatrix} a^{n_k} \\ b^{n_k} \end{bmatrix} \rightarrow \begin{bmatrix} a \\ b \end{bmatrix}$. Then $\begin{bmatrix} a \\ b \end{bmatrix}$ defines a hyperplane separating d from C as easily checked.

Case 3 : C, D are convex. (Most general case)

D is convex. $\Rightarrow -D$ is convex. $\Rightarrow C + (-D) = C - D$ is convex. Since $0 \notin C - D$, from Case 2, there is a hyperplane separating 0 from $C - D$. Thus there is a such that $a^T x \leq a^T y$ for any pair $x \in C$ and $y \in D$. Can complete the proof by considering $b := \sup_{x \in C} a^T x$. $\square$

Converse of separating separating hyperplane

Theorem

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

Only if: Say C is open and an affine function is nonpositive on C and nonnegative on D. Any common element of C and D should have the function value 0. But, C's being open implies the affine function is strictly negative on it and can not have a zero on C. $\square$

Application

Theorem of alternatives for strict linear inequalities

$$\begin{array}{ll} & Ax \prec b \text{ are infeasible.} \\ \Longleftrightarrow & C = \{b - Ax \mid x \in \mathbb{R}^n\}, D = \mathbb{R}^m_{++} \text{ do not intersect.} \\ \Longleftrightarrow & \exists \text{ a separating hyperplane: } \exists \text{ a nonzero } \lambda \in \mathbb{R}^m, \mu \in \mathbb{R} \\ & \text{s.t. } \lambda^T(b - Ax) \leq \mu \ \forall x \in \mathbb{R}^n \text{ and } \lambda^T y \geq \mu \ \forall y \succ 0. \\ \Longleftrightarrow & \lambda \neq 0, \lambda \succeq 0, A^T\lambda = 0, \lambda^T b \leq 0 \text{ as we may assume } \mu = 0. \text{ (Why?)} \end{array}$$

Supporting hyperplanes

Definition

Suppose $x_0 \in C \subseteq \mathbb{R}^n$. *If* $a \neq 0$ *satisfies* $a^T x \leq a^T x_0, \forall x \in C$, *then the hyperplane* $\{x | a^T x = a^T x_0\}$ *is called a supporting hyperplane to* C *at* $x = x_0$.

Separating hyperplane theorem immediately implies the following.

Theorem

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

Dual cones

Definition

The dual cone of a cone K is

$$K^* = \{y | x^T y \geq 0, \forall x \in K\}$$

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

Definition

When a cone is its own dual, we call it self-dual.
e.g. $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+, (\mathbb{S}^n_+)^* = \mathbb{S}^n_+$

1. K^* is closed and convex.
2. $K_1 \subseteq K_2 \Longrightarrow K_2^* \subseteq K_1^*$
3. If K has nonempty interior, then K^* is pointed.
4. If *cl* K is pointed, then K^* has nonempty interior.
5. K^{**} is the closure of the convex hull of K.

Dual cones

Example

Self-duality of positive semidefinite cones: $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$

1. $\subseteq$: *Suppose* $P \in \mathbb{S}^n_+$ *satisfies* $P \cdot Q \geq 0, \forall Q \in \mathbb{S}^n_+$. *We want to show that* $P \in \mathbb{S}^n_+$.
   Suppose, on the contrary, that $P \notin \mathbb{S}^n_+$. *Then,* $\exists y \neq 0$ *s.t.* $y^T Py = tr(y^T Py) = tr(P(yy^T)) = P \cdot (yy^T) < 0$. *As* $yy^T \in \mathbb{S}^n_+$, *contradiction.*
2. $\supseteq$: *Pick any* $Q \in \mathbb{S}^n_+$. *We want to show that* $P \cdot Q \geq 0 \; \forall P \in \mathbb{S}^n_+$.
   Consider Spectral decomposition $P = \lambda_1 p_1 p_1^T + \cdots + \lambda_k p_k p_k^T$. $\Rightarrow P \cdot Q = tr(\sum_{i=1}^k \lambda_i (p_i p_i^T) \cdot Q) = tr(\sum_{i=1}^k \lambda_i p_i^T Q p_i) \geq 0$.

Example

Dual of a norm cone $K = \{(x,t) \in \mathbb{R}^{n+1} : \|x\,| \leq t\}$ *is* $K^* = \{(y,s) \in \mathbb{R}^{n+1} : \|y\|_* \leq s\}$, *where* $\|y\|_* = \sup\{y^T u | \; \|u\| \leq 1\}$, *a norm called dual norm. For instance,* $(\|\;\|_p)_* = \|\;\|_q$ *when* $p^{-1} + q^{-1} = 1$.

Dual generalized inequalities

Theorem

K is a proper cone. $\Rightarrow K^*$ is a proper cone.

It induces a generalized inequality.

$$y \succeq_{K^*} 0 \iff y^T x \geq 0, \forall x \succeq_K 0$$

- Dual characterization of minimum element
  x is the minimum element of S w.r.t. $\preceq_K$
  $\iff \forall\lambda \succ_{K^*} 0$, x is the unique minimizer of $\lambda^T z$ over $z \in S$.
  $\iff \{z|\lambda^T(z-x)=0\}$ is a strict supporting hyperplane to S at x_0.
- Dual characterization of minimal element
  x is minimal if $\lambda \succ_{K^*} 0$ and x minimizes $\lambda^T z$ over $z \in S$.
  If S is convex, then for any minimal element x,
  there exists a nonzero $\lambda \succeq_{K^*} 0$ s.t. x minimizes $\lambda^T z$ over $z \in S$.

Homework

2.1, 2.4, 2.9, 2.10, 2.12(e-g), 2.13, 2.18, 2.20, 2.24, 2.28, 2.31, 2.33, 2.35, 2.37, 2.38