Convex Optimization

Sasank

April 27, 2015

1 Affine Sets, Convex Sets, Cones etc.

Affine Set contains the *line* through any two points in the set. Equivalently S is affine if $S = \{x \mid Ax = b\}$ for some matrix A. This can be proven by showing that $S - x_p$ is a vector space for some $x_p \in S$

Convex Set contains *line segment* between any two points in the set. Convex hull of a set is set of all convex combinations of points of the set.

Cone S is cone if $x \in S$ then $\theta x \in S$ for $\theta \geq 0$.

Convex Cone S is convex cone if $x_1, x_2 \in S$ then $\theta_1 x_1 + \theta_2 x_2 \in S$ for $\theta_1, \theta_2 \geq 0$.

Hyperplane Sets of the form $\{x \mid a^T x = b\}$. a is the normal vector.

Halfspace Sets of the form $\{x \mid a^T x \leq b\}$

Euclidean ball and Ellipsoid Ball is set of the form $B(x_c, r) = \{x \mid \|x - x_c\| \leq r\}$. An ellipsoid is set of the form $\{x \mid (x - x_c)^T P^{-1}(x - x_c) \leq 1\}$ with $P \in S^n_{++}$

Note that Affine sets, convex cones, hyperplanes, halfspaces, Euclidean balls and Ellipsoids are convex and hyperplanes are also affine.

Following are the definitions of some combinations of points x_1 and x_2. Here $\theta_i \in R$.

Linear combination $\theta_1 x_1 + \theta_2 x_2$

Affine combination $\theta_1 x_1 + \theta_2 x_2$, where $\theta_1 + \theta_2 = 1$

Convex combination $\theta_1 x_1 + \theta_2 x_2$, where $\theta_1 + \theta_2 = 1$, $\theta_1, \theta_2 \geq 0$

Conic combination $\theta_1 x_1 + \theta_2 x_2$, where $\theta_1, \theta_2 \geq 0$

2 Some spaces

2.1 General Topological Spaces

General topological spaces have some sense of neighbourhood. Following are the definitions for such spaces.

Closure of a set $\mathrm{cl}(S) = S \cup \{\lim x_i \text{ for convergent } \{x_i\} \subseteq S\}$

Closed set if $\mathrm{cl}(S) = S$

Open set if S^c is closed

Interior(S) $= \bigcup_{S' \subseteq S \text{ open}} S'$

Boundary(S) $= \mathrm{cl}(S) - \mathrm{int}(S)$

2.2 Normed Vector Space

A function $\|.\|$ is called norm if it satisfies

- $\|x\| \geq 0$; $\|x\| = 0$ iff $x = 0$
- $\|tx\| = |t|\|x\|$ for $t \in R$
- $\|x + y\| \geq \|x\| + \|y\|$

A vector space on which a norm is defined is called normed vector space. These are equivalent characterizations for sets defined before for normed vector spaces.

Closure(S) $= \{x \mid \forall \epsilon > 0, S \cap B(x, \epsilon) \neq \emptyset\}$

Closed set if $\mathrm{cl}(S) = S$

Open set $\forall x \in S, \exists \epsilon > 0$ s.t. $B(x, \epsilon) \subseteq S$

Interior(S) $= \{x \mid x \in S, B(x, \epsilon) \subseteq S$ for some $\epsilon > 0\}$

Boundary(S) $= \mathrm{cl}(S) - \mathrm{int}(S)$

Norm ball is of the form $\{x \mid \|x - x_c\| \leq r\}$ and norm cone is defined as $\{(x, t) \mid \|x\| \leq t\}$.

2.3 Inner Product Space

Inner product space is a vector space with a inner product defined on it. Inner product needs to satisfy following properties:

Conjugate symmetry $\langle x, y\rangle = \overline{\langle y, x\rangle}$

Linearity $\langle ax, y\rangle = a\langle x, y\rangle$, $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$

Positive definiteness $\langle x, x\rangle \geq 0$ with equality iff $x = 0$

An inner product induces a norm as follows $\|x\| = \sqrt{\langle x, x\rangle}$. Thus every inner product space is a normed vector space. Note that not all norms are inner products. For example consider p-norm $\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$ for $x \in \mathbb{R}^n$. This is an inner product if and only if $p = 2$

2.4 Banach and Hilbert Spaces

A metric space is complete if every Cauchy sequence is convergent.

Banch space is a complete normed vector space

Hilbert space is a complete inner product space

2.5 Some Examples of Vector Spaces

Sequences Infinite sequences, sequences with finite non zero elements, l^p = sequences with bounded p-norm $\|x\|_p = (\sum_{i=1}^{\infty} |x_i|^p)^{1/p}$. l^p is a Banach space for $p \geq 1$ and a Hilbert space for $p = 2$.

Polynomials Polynomials, polynomials with degree $\leq n$, polynomials in multiple variables.

Functions $f : X \rightarrow V$ has dimension $|X| \dim(V)$. $L_p = \{f \mid$ measurable f with $\|f\|_p = \left(\int_{i=1}^{\infty} |f(x)|^p dx\right)^{1/p} < \infty\}$ is a Banach space for $p \geq 1$ and a Hilbert space for $p = 2$.

Matrices $m \times n$ matrices. Let $N(x)$ be a vector norm. Then a matrix norm can be defined as

$$\|A\| = \sup_{x \neq 0} \frac{N(Ax)}{N(x)}$$

Frobenius norm of matrices is defined as

$$\|A\| = \sqrt{\sum_{i,j} a_{ij}} = \text{trace}(A^T A)$$

and Frobenius inner product is defined as $\langle A, B \rangle = \text{trace}(A^T B)$

3 Operators & Duals

3.1 Linear Operators

A linear operator T between vector spaces X and Y is $T : X \to Y$ such that $\forall x, x', \lambda, \mu$

$$T(\lambda x + \mu x') = \lambda T(x) + \mu T(x')$$

These are some properties and definitions of linear operators:

- If X & Y are normed vector spaces, we can define operator norm as

$$\|T\| = \sup_{x \neq 0} \frac{\|T(x)\|}{\|x\|}$$

- T is called bounded if $\|T\| \leq N$ for some $N \in R$. Equivalently if $\exists M$ such that $\|Tx\| \leq M\|x\| \; \forall x \in X$
- T is continuous if and only if T is bounded
- Can generalize eigenfunctions and eigenvalues to operators.
- X and Y are called linearly isomorphic if T is bijection.
- Kernel of T = $\{x \in X \mid Tx = 0\}$. Range of T = $\{y \in Y \mid y = Tx \text{ for some } x \in X\}$
- If X is a normed v.s. and Y is Banach, then $T : X \to Y$ is Banach with respect to operator norm.
- $T^* : Y \to X$ is called adjoint linear map of $T : X \to Y$ if $\langle Tx, y \rangle_Y = \langle x, T^* y \rangle_X \; \forall x \in X, y \in Y$

3.2 Dual Spaces & Cones

Linear operator $T : X \to R$ is called a linear functional. We define the following

Algebraic dual $X^\# = \{T \mid T : X \to R\}$. i.e, set of all linear functionals. If X is fdvs, $X \cong X^\#$.

Topological dual $X^* = \{T \mid \text{continuous } T : X \to R\}$. i.e, set of all continuous/bounded linear functionals.

Algebraic dual cone $C^\# = \{T \in X^\# \mid T(x) \geq 0 \; \forall x \in C\}$. Note that C can be any subset of X.

Topological dual cone $C^* = \{T \in X^* \mid T(x) \geq 0 \; \forall x \in C\}$. It can be easily seen that C^* is always a convex cone irrespective of C.

If X is finite dimensional vector space with a norm, $X^\# = X^* \cong X$ and $C^\# = C^*$.

3.2.1 Riesz Representation Theorem

If X is a Hilbert space and T is a bounded linear functional, then there exists a *unique* $y \in X$ such that

$$T(x) = \langle y, x \rangle \; \forall x \in X$$

In fact $X^* = \{T_y(x) = \langle y, x \rangle \mid x \in X\}$. Furthermore $X \cong X^*$.

Therefore we can simplify our notion of (topological) dual cone as

$$C^* = \{y \in X \mid \langle y, x \rangle \geq 0 \; \forall x \in C\}$$

.

Properties of dual cones

- C^* is closed because it is intersection of (closed) halfspaces. Thus C^* is closed convex cone for all $C \subseteq X$.
- $C_1 \subseteq C_2 \implies C_2^* \subseteq C_1^*$
- $\text{int}(C^*) = \{y \in X \mid \langle y, x \rangle > 0 \; \forall x \in X\}$
- If C is a cone and $\text{int}(C) \neq \emptyset$, then C^* is pointed. i.e if $x \in C^*$ and $-x \in C^*$ then $x = 0$.
- If C is a closed convex cone, $C^{**} = (C)$. Prove this using strict separating hyperplane theorem.
- If C is a convex cone, $C^{**} = \text{closure}(C)$
- If $\text{closure}(C)$ is pointed then $\text{int}(C^*) \neq \emptyset$

3.3 Positive Semidefinite Cone

S^n_+, set of all symmetric positive semidefinite matrices is a convex cone. On the other hand, S^n_{++}, set of all symmetric positive definite matrices is a convex set but not cone ($0 \notin S^n_{++}$). Some notable properties of positive semidefinite cone:

- $(S^n_+)^* = S^n_+$
- $\text{int}(S^n_+) = S^n_{++}$

4 Linear & Conic Programs and Weak Duality

4.1 Linear Program and Its Dual

Following optimization problem is called linear program:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & -Ax + b \leq 0 \end{aligned} \tag{LP}$$

We'll derive its dual. For all $\lambda \in \mathbb{R}^n_+$, we have $\lambda^T(-Ax+b) \leq 0$. Therefore for all $\lambda \in \mathbb{R}^n_+$ we have,

$$\begin{aligned} c^T x &\geq c^T x + \lambda^T(-Ax+b) \\ &= \lambda^T b + (c - A^T\lambda)^T x \\ &\geq \min_x \lambda^T b + (c - A^T\lambda)^T x \\ &= \begin{cases} \lambda^T b & A^T\lambda = c \\ -\infty & A^T\lambda \neq c \end{cases} \end{aligned}$$

Thus we can write

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & c^T x \\ \text{subject to} & -Ax+b \leq 0 \end{array} \leq \begin{array}{ll} \underset{\lambda\geq 0}{\text{maximize}} & b^T\lambda \\ \text{subject to} & A^T\lambda = c \end{array}$$

We call the problem

$$\begin{array}{ll} \underset{\lambda\geq 0}{\text{maximize}} & b^T\lambda \\ \text{subject to} & A^T\lambda = c \end{array} \tag{LD}$$

as dual linear program. The property LP $\leq$ LD is called weak duality.

4.2 Conic Program and Its Dual

4.2.1 Generalized Inequality

Let K be a cone. We define $a \geq_K b$ if $a - b \in K$. This is a valid inequality iff K is a convex pointed cone. i.e, following properties need to be satisfied

Reflexivity $a \geq a$

Anti-symmetry If both $a \geq b$ and $b \geq a$, then $a = b$

Transitivity If both $a \geq b$ and $b \geq c$, then $a \geq c$

Compatible with linear operators :

- **Homogenity** If $a \geq b$ and $\mu \in R_+$, then $\mu a \geq \mu b$
- **Additivity** If $a \geq b$ and $c \geq d$, then $a + c \geq b + d$

A convex cone is called proper if it is closed, solid and pointed.

4.2.2 Conic Program & Weak Duality

Following optimization problem is called conic program:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{maximize}} & c^T x \\ \text{subject to} & -Ax+b \leq_k 0 \end{array} \tag{CP}$$

To derive a weak dual of this, we need a set D such that $\forall\lambda \in D, \lambda^T(Ax-b) \geq 0$. Dual cone K^* satisfies this criterion. Therefore the following is the conic dual program:

$$\begin{array}{ll} \underset{\lambda\in K^*}{\text{maximize}} & b^T\lambda \\ \text{subject to} & A^T\lambda = c \end{array} \tag{CD}$$

We have CP $\leq$ CD. Note that LP is a special case of CP where $K = \mathbb{R}^n$.

Following is the alternate standard notation and generalization for conic program and its dual.

$$\begin{aligned} \text{maximize} \quad & \langle c, x \rangle_V \\ \text{subject to} \quad & Ax = b, x \in K \subseteq V \end{aligned} \tag{CP}$$

where $A : V \to \mathbb{R}^n$ is a linear map. (If $K = S_n^+ \subset V = S^n$, then CP is called semi-definite program.) Denote feasible set of CP by F_p. Then conic dual is

$$\begin{aligned} \text{maximize} \quad & \langle b, \lambda \rangle_{\mathbb{R}^n} \\ \text{subject to} \quad & c - A^*\lambda \in K^*, \lambda \in \mathbb{R}^n \end{aligned} \tag{CD}$$

where $A^* : \mathbb{R}^n \to V$ is adjoint of A. Denote feasible set of CD by F_d. In this notation we have CP $\geq$ CD. Some definitions and properties:

- CP $-$ CD is called duality gap and is always ≥ 0
- If CP or CD is feasible but unbounded, then the other is infeasible or has no feasible solution
- If a pair of feasible solutions can be found to the both primal and dual problems with equal objective, then they are both optimal

5 Strong Duality

Theorem:

1. Let CP or CD be infeasible and let other be feasible and have an interior. Then the other is unbounded.
2. Let CP and CD be both feasible and let one of them have an interior. Then there is 0 duality gap.
3. Let CP and CD be both feasible and have an interior. Then both have optimal solutions with 0 duality gap.

To prove this we need to prove theorem of alternatives/Farkas' lemma.

5.1 Theorem of alternatives

1. Consider $\{x \mid Ax = b, x \in K\}$ for a proper cone $K \subseteq V$ and $A : V \to \mathbb{R}^n$. Suppose $\exists\lambda$ such that $-A^*\lambda \in \text{int}(K^*)$.
   Then $\{x \mid Ax = b, x \in K\}$ has a feasible solution x iff $\{\lambda \mid -A^*\lambda \in K^*, \langle b, \lambda \rangle > 0\}$ has no feasible solution.
2. Consider $\{(\lambda, s) \mid c - A^*\lambda = s \in K\}$ for a proper cone $K \subseteq V$ and $A : V \to \mathbb{R}^n$. Suppose $\exists x$ such that $Ax = 0, x \in \text{int}(K^*)$.
   Then, $\{(\lambda, s) \mid c - A^*\lambda = s \in K\}$ has a feasible solution (λ, s) iff $\{x \mid Ax = 0, x \in K^*, \langle c, x \rangle < 0\}$ has no feasible solution.

Proof of 1: Let $\bar{\lambda}$ be such that $-A^*\bar{\lambda} \in K^*$ and let $\{x \mid Ax = b, x \in K\}$ has a feasible solution $\bar{x}$. Then

$$-\langle \bar{\lambda}, b \rangle = -\langle \bar{\lambda}, A\bar{x} \rangle = \langle -A^*\bar{\lambda}, \bar{x} \rangle \geq 0$$

Thus $\{\lambda \mid -A^*\lambda \in K^*, \langle b, \lambda \rangle > 0\}$ has no solution, i.e. empty.

$C = \{y \mid y = Ax, x \in K\}$ is a closed convex set. Let $\{x \mid Ax = b, x \in K\}$ be empty. i.e $b \notin C$. By strict separating hyperplane theorem, $\exists \lambda \in \mathbb{R}^n$ such that

$$\langle \lambda, b \rangle > \langle \lambda, y \rangle \ \forall y \in C$$

Using the definition of C, we have

$$\langle \lambda, b \rangle > \langle \lambda, Ax \rangle = \langle A^*\lambda, x \rangle \ \forall x \in K$$

Thus $\langle A^*\lambda, x \rangle$ is bounded above $\forall x \in K$. Additionally $\langle A^*\lambda, x \rangle \leq 0$ as otherwise $\langle A^*\lambda, \alpha x \rangle$ can made as large as needed contradicting previous statement. Since $0 \in K$, $\langle \lambda, b \rangle > 0$. Thus the set $\{\lambda \mid -A^*\lambda \in K^*, \langle b, \lambda \rangle > 0\}$ is non empty. $\square$

Proof of 2: Let $\{(\lambda, s) \mid c - A^*\lambda = s \in K\}$ has a feasible solution $(\bar{\lambda}, \bar{s})$ and $\bar{x} \in K^*$ be such that $A\bar{x} = 0$. Then

$$\langle \bar{s}, \bar{x} \rangle = \langle c - A^*\bar{\lambda}, \bar{x} \rangle = \langle c, \bar{x} \rangle - \langle A^*\bar{\lambda}, \bar{x} \rangle = \langle c, \bar{x} \rangle - \langle \bar{\lambda}, A\bar{x} \rangle = \langle c, \bar{x} \rangle \geq 0$$

Thus $\{x \mid Ax = 0, x \in K^*, \langle c, x \rangle < 0\}$ has no feasible solution.

$C' = \{t \mid t = A^*\lambda + s, \lambda \in \mathbb{R}^n, s \in K\}$ is a closed convex set. Let $\{(\lambda, s) \mid c - A^*\lambda = s \in K\}$ has no feasible solution. i.e $c \notin C'$. By strict separating hyperplane theorem, $\exists x \in V$ such that

$$\langle x, c \rangle < \langle x, t \rangle \ \forall t \in C'$$

Using the definition of C', we have

$$\langle x, c \rangle < \langle x, s + A^*\lambda \rangle = \langle x, s \rangle + \langle x, A^*\lambda \rangle = \langle x, s \rangle + \langle Ax, \lambda \rangle \ \forall \lambda \in \mathbb{R}^n, s \in K$$

Thus $\langle x, s + A^*\lambda \rangle$ is bounded below $\forall \lambda \in \mathbb{R}^n, s \in K$. Additionally $\langle x, s + A^*\lambda \rangle \geq 0$ as otherwise $\langle x, \alpha s + \alpha A^*\lambda \rangle$ can made as small as needed contradicting previous statement. Since $0 \in K$, $\langle x, c \rangle < 0$.

Also, $\langle x, A^*\lambda \rangle = 0$ as otherwise $\langle x, s + A^*\lambda \rangle$ can be made negative. Therefore $\langle x, A^*\lambda \rangle = \langle Ax, \lambda \rangle = 0$. Thus $Ax = 0$ and $\langle x, s \rangle \geq 0 \ \forall s \in K$ or equivalently $x \in K^*$. This proves that $\{x \mid Ax = 0, x \in K^*, \langle c, x \rangle < 0\}$ is non empty. $\square$

5.2 Proof of Strong Duality

We will now apply Farkas' lemma to prove strong duality theorem. Nope.

A Appendix

A.1 Characterizations of positive definite matrices

If $A \in S^n$, the following are equivalent

1. $x^T Ax > 0$ for all $x \in \mathbb{R}^n \setminus 0$ i.e $A \in S^n_+$
2. All n eigenvalues $\lambda > 0$
3. $A = LL^T$ where L is lower triangular with positive diagonal elements
4. $A = Q\Lambda Q^T$ where Q is orthonormal matrix and Λ is a positive diagonal matrix
5. $x^T Ay$ is a inner product