Lecture Note 1: Introduction and convex sets

Xiaoqun Zhang
Shanghai Jiao Tong University

Last updated: March 2, 2017
1.1 Introduction

1. Classification of optimization problems

<table>
<thead>
<tr>
<th>Continuous</th>
<th>vs</th>
<th>Discrete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constrained</td>
<td>vs</td>
<td>Unconstrained</td>
</tr>
<tr>
<td>Global</td>
<td>vs</td>
<td>Local</td>
</tr>
<tr>
<td>Stochastic</td>
<td>vs</td>
<td>Deterministic</td>
</tr>
<tr>
<td>Convex</td>
<td>vs</td>
<td>Nonconvex</td>
</tr>
</tbody>
</table>

2. Mathematical optimization

\[
\min_{x} f_0(x) \\
\text{subject to (s.t.) } f_i(x) \leq 0, i = 1, \cdots, m
\]

where \(x = (x_1, \cdots, x_n)\) optimization variables; \(f_0 : \mathbb{R}^n \to \mathbb{R}\) objective function and \(f_i : \mathbb{R}^n \to \mathbb{R}\) constraint functions. **optimal solution** \(x^*\) has smallest value of \(f_0\) among all vectors that satisfy the constraints.

3. Examples:

- Portfolio optimization. variables: amounts invested in different assets; constraints: budget, max./min. investment per asset, minimum return; objective: overall risk or return variance.
Lecture note 1

Convex optimization

- device sizing in electronic circuits, variables: device widths and lengths; constraints: manufacturing limits, timing requirements, maximum area objective: power consumption
- data fitting, variables: model parameters; constraints: prior information, parameter limits; objective: measure of misfit or prediction error

4. Nonlinear programming: local methods: find a point that minimize \( f_0 \) among feasible points near it, require initial guess and usually does not provide information about distance to (global) optimum. Global optimization: find the global solution and worst case complexity can grow exponentially with problem size.

5. Convex programming:
\[
\min f(x) \text{ s.t. } X \in C
\]
both \( f \) and \( C \) are convex. Special cases:
- linear programming
\[
\min c^T x \text{ s.t. } Ax \leq b; Bx = d;
\]
- least square problems:
\[
\min \frac{1}{2} \|Ax - b\|^2
\]
- Quadratic programming
\[
\min x^T Qx + c^T x \text{ s.t. } Ax \leq b; Bx = d;
\]
for \( Q \succeq 0 \) being (symmetric) positive semi-definite (SPSD).

6. Optimization algorithms requirement
- Robustness: perform well on a wide variety of problem
- Efficiency: reasonable computer time or storage
- Accuracy: precision, not being sensitive to norms
- Solving convex optimization problems: reliable and efficient algorithms, often only involves first and second order derivatives, and almost a technology
- Using convex optimization problem: often difficult to recognize, many tricks for transforming problems into convex form, surprisingly many problems can be solved via convex optimization.

7. Brief history of convex optimization
- Theory: convex analysis 1900-1970 (Minkowski, Fenchel, Moreau, Rockafellar etc)
Lecture note 1

Convex optimization

- Algorithms: simplex algorithm 1947 (Dantzig); interior-point methods (1960s); 1970s: ellipsoid method and other subgradient methods; polynomial time interior point methods for nonlinear convex optimization.

- Applications mostly in operational research; few in engineering before 1990; many new applications in engineering (control and signal, image processing, machine learning, data science)

- New problem classes: second order cone programming (SOCP), semi-definite programming (SDP), robust optimization etc.

8. Application:

- Solving linear equation: $Ax = b$, for $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Recover $x$. Compressive sensing/Sparse learning (Lasso): find the sparsest solution.

  $$\min \|x\|_0 \quad \min \|x\|_2 \quad \min \|x\|_1$$

  $$s.t. Ax = b \quad s.t. Ax = b \quad s.t. Ax = b$$

  $\ell_1$ norm is sparsity promoting! Donoho et al 98, Candes&Tao 05,06.

  Image processing:

  $$\min \|\Phi x\|_1$$

  $$s.t. Ax = b$$

  $$\min \|\Phi x\|_1 + \frac{\mu}{2} \|Ax - b\|^2$$

  Subsampled (Accelerated) MRI imaging.

- Low rank matrix completion problem: Netflix database contains about a million users, 25,000 movies; People rate movies. Sparsely sampled entries. Million dollar award fill out the "Netflix matrix".

  $$\min \text{rank}(X) \quad \min \text{rank}(X) \quad \min \|X\|_*$$

  $$s.t. X_{ij} = Y_{ij}, (i, j) \in \Omega \quad s.t. A(X) = Y \quad s.t. A(X) = Y$$

- Portfolio optimization: $r_i$ the rate of return for stock $i$ (random variable with expectation $\mu_i$, $i = 1, \cdots, n$ and variance matrix $\Sigma$. Let $x_i$ be the relative amount invested in stock $i$.

  $$\min \frac{1}{2} x^T \Sigma x$$

  $$s.t. \sum \mu_i x_i \geq r_0$$

  $$\sum x_i = 1$$

  $$x_i \geq 0$$

  or other risk measure (value at risk (VaR), Conditional VaR etc).
• Correlation matrices: $X = X^T, X_{ii} = 1, i = 1, \cdots, n, X \succeq 0.$

$$
\min \frac{1}{2} \|X - C\|_F
\quad s.t. \ X = X^T, X_{ii} = 1, i = 1, \cdots, n, X \succeq 0
$$

One can add rank constraints $\text{rank}(X) \leq r.$

• And many more.

1.2 Notations and elementary results

• Let $E$ be an ordered set in $R$: greatest lower bound, $\inf E$ and $\sup E$: least upper bound, when they exists. When they belongs to $E$, then $\inf$ becomes $\min$ and $\sup$ becomes $\max$.

• Computing the number:

$$
\mathcal{J} := \inf\{f(x) : x \in X\} \tag{1.1}
$$

finding a minimizing sequence: i.e. $(x_k) \subset X$ such that $f(x_k) \to \mathcal{J}$ when $k \to \infty$. Other notations $\inf_{x \in X} f$ and $\inf_{X} f$. $f$ is often called objective function, or infimand.

$$
\inf\{f(x) : x \in X_1 \cup X_2\} = \min\{\mathcal{J}_1, \mathcal{J}_2\}
$$

$$
X_1 \subset X_2 \Rightarrow \mathcal{J}_1 \geq \mathcal{J}_2
$$

$$
\inf\{f(x) : x \in X_1 \cap X_2\} \geq \max\{\mathcal{J}_1, \mathcal{J}_2\}
$$

$$
\inf\{f(x_1) + g(x_2) : x_1 \in X_1; x_2 \in X_2\} = \mathcal{J}_1 + \mathcal{J}_2
$$

$$
\inf\{f(x) + g(x) : x \in X\} \geq \mathcal{J} + \mathcal{J}
$$

$$
\inf\{tf(x) : x \in X\} = t\mathcal{J}_1
$$

$$
\inf\{-f(x) : x \in X\} = -\sup\{f(x) : x \in X\}
$$

$$
\inf\{g(x, y) : x \in X \text{ and } y \in Y\} = \inf_{x \in X} \inf_{y \in Y} g(x, y) = \inf_{y \in Y} \inf_{x \in X} g(x, y)
$$

• An optimal solution of [1] is an $\bar{x} \in X$ such that $f(\bar{x}) = \mathcal{J} \leq f(x)$ for all $x \in X$. Such an $\bar{x}$ is often called a minimizer, a minimum point, global minimum or more simply a minimum of $f$ on $X$ and the minimal value. It can be sometimes denoted as $\text{arg min}\{f(x) : x \in X\}$.

• If $(\bar{x}, \bar{y})$ minimizes $g$ over $X \times Y$, then $\bar{y}$ minimizes $g(\bar{x}, \cdot)$ over $Y$ and $\bar{x}$ minimizes over $X$ the function

$$
\phi(x) := \inf\{g(x, y) : y \in Y\}
$$

Conversely, if $\bar{x}$ minimizes $\phi$ over $X$ and $\bar{y}$ minimizes $g(\bar{x}, \cdot)$ over $Y$, then $(\bar{x}, \bar{y})$ minimize $g$ over $X \times Y$. 

6
Continuity: let \( x^* \in X \). If \( f(x^*) \leq \lim \inf_{x \to x^*} f(x) \geq \lim \inf_{x \to x^*} f(x) \), then \( f \) is said to be lower (upper) semi-continuous (l.s.c) at \( x^* \). If \( X \) is compact and \( f \) is continuous, the lower bound \( \bar{f} \) exists. The l.s.c (of \( f \) on the whole compact \( X \)) suffices: if \( (x_k) \) is a minimizing sequence, with some cluster point \( x^* \in X \), we have

\[
f(x^*) \leq \lim \inf f(x_k) = \lim f(x_k) = \bar{f}
\]

- A given minimizing sequence \( (x_k) \) converges to an optimal solution when \( k \to \infty \), \( x_k \) is not necessarily converging to the an optimal solution. With \( X = \mathbb{R}, f(0) = 0, f(x) : 1/|x| \) for \( x \neq 0 \). The sequence \( x_k = k \) is minimizing but does not converge to the minimum 0 when \( k \to \infty \).

Differentiation in a Euclidean space: \( f : \Omega \to \mathbb{R} \) where \( \Omega \) is an open set \( \Omega \subset X \) is said to be differentiable if there exists a linear form \( l \) on \( X \) such that

\[
f(x + h) = f(x) + l(h) + o(\|h\|)
\]

The linear form \( l(h) \) is given by \( l(h) = \langle \nabla f, h \rangle \). \( \nabla f \) is called the gradient of \( f \) at \( x \):

- Gradient:
  \[
  \nabla f(x) = \left[ \frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right]
  \]
  (column vector)

Hessian of \( f \) (symmetric matrix)

\[
H(x) = J(\nabla f(x)) = D^2 f(x) = \begin{pmatrix}
\frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\
\cdots & \cdots & \cdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x)
\end{pmatrix}
\]

- Gradient and Hessian are local properties that help us recognize local solutions and determine a direction to move at toward the next point.

Taylor expansion: consider \( x, d \in \mathbb{R}^n \) and \( f \) is second-order differentiable. Define \( \phi(\alpha) = f(x + \alpha d) \) Then

\[
\phi'(\alpha) = \nabla f(x + \alpha d)^T d
\]

\[
\phi''(\alpha) = d^T \nabla^2 f(x + \alpha d) d
\]

Hence:

\[
f(x + \alpha d) = f(x) + \alpha \nabla f(x)^T d + o(\alpha) \quad \text{1st order approximation}
\]

\[
= f(x) + \alpha \nabla f(x)^T d + \frac{d^T \nabla^2 f(x) d}{2} \alpha^2 + o(\alpha^2) \quad \text{2nd order approximation}
\]
1.3 Convex sets

1.3.1 Definitions

**Definition 1** A set $C \subset \mathbb{R}^n$ is called convex if for every pair of $x, y \in C$, the entire line segment: $[x, y] := \{z : z = \lambda x + (1 - \lambda) y : 0 \leq \lambda \leq 1\} \subset C$.

- By convention: empty set $\emptyset$ is convex.
- A consequence of the definition is that $C$ is also path-connected, i.e., two arbitrary points in $C$ can be linked by a continuous path.
- **Affine set**: contain the line $x = \theta x_1 + (1 - \theta)x_2$, $\theta \in \mathbb{R}$ for any two distinct points in the set.
- **Affine combination**: any point $x$ of the form $x = \sum^n \theta_i x_i$ with $\sum \theta_i = 1$ is called an affine combination of the points $x_1, \ldots, x_n \in \mathbb{R}^n$.
- **Affine hull**: the set of all affine combinations of points in $S$, denoted as $\text{Aff}(S)$.
- **Convex Combination**: any point $x$ of the form $x = \sum^n \theta_i x_i$ with $\sum \theta_i = 1, \theta_i \geq 0$ is called a convex combination of the points $x_1, \ldots, x_n \in \mathbb{R}^n$.
- **Convex hull**: set of all convex combinations of points in $S$. Denoted as $\text{Conv}(S)$.
  - Among all convex sets containing $S$, $\text{Conv}(S)$ is the smallest convex set.
- **Cone**: a cone is a set that is closed under multiplication by positive scalars, i.e. if $x \in C$, then $tx \in C$ for any $t > 0$. In general cone is not supposed to contain 0 for notational reasons. (a subspace is a cone but has no apex)

**Examples of convex sets in $\mathbb{R}^n$**

- **(Affine) Hyperplane**: set of form $\{x|a^\top x = b\}, a \neq 0$, denoted as $H_{a,b}$, $a$ is the normal vector. Special case for $b = 0, 1$. Generalization to affine subspace $V$ is a translation of some vector space $V_0$. Affine subspace: $M = a + L = \{y = a + x | x \in L\}$ where $L$ is a linear subspace in $\mathbb{R}^n$ and $a$ is a vector from $\mathbb{R}^n$.
- **(Closed) Halfspace**: set of the form $\{x|a^\top x \leq b\}$ $(a \neq 0)$, and $a$ is the normal vector.
- Hyperplanes are affine and convex, halfspaces are convex.
- **Polyhedra**: solution sets of many half spaces $\{x|Ax \leq b; Cx = d\}$.
- **Simplices**: the (standard) unit simplex in $\mathbb{R}^k$ is defined as $\Delta_k = \{\alpha \in \mathbb{R}^k : \sum_{i=1}^k \alpha_i = 1, \alpha_i \geq 0 \text{ for } i = 1, \ldots, k\}$.
- **Norm ball**: $\{x||x| \leq 1\}$ and $\{x||x - x_0|| \leq r\} = \{x_0 + ru ||u|| \leq r\}$. Special case: Euclidean balls, and $||x||_p \leq 1$, $(p \geq 1)$.  

8
Lecture note 1

Convex optimization

- Ellipsoid: set of the form $E = \{x | (x - x_0)^TP^{-1}(x - x_0) \leq 1\}$ with $P \in S^n_+$ being symmetric positive definite. The lengths of the semi-axis of $E$ are given by $\sqrt{\lambda_i}$, where $\lambda_i$ are the eigenvalues of $P$. Other representation: $\{x | x_0 + Au, \|u\|_2 \leq 1\}$ with $A = P^{1/2}$ being square and nonsingular.

- Neighborhood of a convex set. Let $M$ be convex set in $\mathbb{R}^n$. Define $M_\epsilon = \{y \in \mathbb{R}^n | \text{dist}_{\|\cdot\|}(y, M) \leq \epsilon\}$ is convex.

- Convex hull and simplex.

- Convex cone: convex and cone. A cone $K$ is convex if and only if $x + y \in K$ for any $x, y \in K$. Example: subspaces, non-negative orthant of $\mathbb{R}^n$: $\mathbb{R}_+^n := \{x = (x_1, \cdots, x_n); x_i \geq 0, \text{for } i = 1, \cdots, n\} = \{x \in \mathbb{R}^n : \langle e_i, x \rangle \geq 0, \text{for } i = 1, \cdots, n\}$.

- Norm cone $\{(x, t) | \|x\| \leq t\}$. Euclidean norm cone is called second order cone.

- Positive semi-definite cone: $S^n$ is the set of symmetric $n \times n$ matrices, and $S^n_+$: positive semi-definite $n \times n$ matrices is a convex cone. $S^n_+$ comprise the cone interior; all singular positive semidefinite matrices reside on the cone boundary.

1.3.2 Calculus of convex sets

Practical methods to check the convexity of a set $C$

- Apply definition

- Show that $C$ is obtained from simple convex sets (hyperplanes, halfspaces, norm balls) by operations that preserve convexity.

Operations that preserve convexity of sets

- Taking intersections. If $M_i, i \in I$ ($I$ be arbitrary set) are convex, so is the set $\bigcap M_i$. Example: $\{x \in \mathbb{R}^n : \langle s_j, x \rangle \leq r_j \text{ for } j = 1, \cdots, m\}$. Example: $S = \{x \in \mathbb{R}^n | |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}$ where $p(t) = x_1 \cos t + x_2 \cos(2t) + \cdots + x_m \cos(mt)$.

- Taking direct product. $C_1 \subset \mathbb{R}^{n_1}, C_2 \subset \mathbb{R}^{n_2}, \cdots, C_k \subset \mathbb{R}^{n_k}$ are convex. So is the set $C_1 \times C_2 \times \cdots \times C_k$.

- Taking the image under an affine mapping. $C$ is convex and $x \to Ax + b$, then $A(C) = \{y : Ax = Ax + b, x \in C\}$ is convex. Taking the inverse image of $C$ is also convex, $A^{-1}C = \{y \in \mathbb{R}^n, A(y) \in C\}$. examples:
  - the opposite $-C$ of a convex set is convex.
  - scaling, translation, projection
Lecture note 1 Convex optimization

- Arithmetic summation and multiplication by reals: if $C_1, \ldots, C_k$ are convex sets in $\mathbb{R}^n$ and $\lambda_1, \ldots, \lambda_k$ are arbitrary real, the set $\sum \lambda_i C_i$ is still convex.
- Solution set of linear matrix inequality $\{x | x_1 A_1 + \cdots + x_m A_m \preceq B\}$ with $A_i, B \in \mathbb{S}^p$.
- Hyperbolic cone $\{x | x^T P x \leq (c^T x)^2, c^T x \geq 0\}$ with $P \in \mathbb{S}^n_{++}$.
- Perspective function $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$:
  $$f(x, t) = \frac{x}{t}, \quad \text{dom}(f) = \{(x, t) | t > 0\}$$
images and inverse images of convex sets under perspective are convex.
- Linear fractional function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$:
  $$f(x) = \frac{Ax + b}{c^T x + d}, \quad \text{dom}(f) = \{x | c^T x + d > 0\}$$
images and inverse images of convex sets under linear-fractional functions are convex. Example: $f(x) = \frac{1}{x_1 + x_2 + 1} x$, where $x = (x_1, x_2)$.

Topological properties of convex sets
- The closure of a set $C \subset \mathbb{R}^n$ is the set comprised of the limits of all converging sequences of elements of $C$, denoted as $\text{cl}(C)$. Example: $\text{cl}(\{x : \|x - a\| < r\}, r > 0) = \{x : \|x - a\| \leq r\}$; $\text{cl}(\{x : s^T x < b\}) = \{x : s^T x \leq b\}$.
- The interior of a set $C \subset \mathbb{R}^n$ is the set comprised of all interior points of $C$ ($x \in C$ is an interior point if some neighborhood of $x$ is still contained in $M$), denoted as $\text{int}(C)$. It is easy to see that $\text{int}(C) \subset C \subset \text{cl}(C)$.
- If $C$ is convex, so are its interior $\text{int}(C)$ and its closure $\text{cl}(C)$.
- Boundary of $C$: $\partial C = \text{cl}(C) \setminus \text{int}(C)$.
- A point $x \in C$ is relative interior of $C$, if $C$ contains the intersection: $\exists r > 0$, $B(x, r) \cap \text{Aff}(C) \subset C$, where $B(x, r)$ denotes the $r$-neighborhood of $x$ (closed ball). The set of all interior points of $C$ is called its relative interior, denoted as $\text{ri}(C)$ (or relint$(C)$).
- The relative boundary $\partial \text{ri}(C) = \text{cl}(C) \setminus \text{ri}(C)$.

<table>
<thead>
<tr>
<th>$C$</th>
<th>Aff$(C)$</th>
<th>dim$(C)$</th>
<th>ri$(C)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>${x}$</td>
<td>${x}$</td>
<td>0</td>
<td>${x}$</td>
</tr>
<tr>
<td>$[x, y]$</td>
<td>line through $x$ and $y$</td>
<td>1</td>
<td>$(x, y)$</td>
</tr>
<tr>
<td>$\Delta_n$</td>
<td>affine subspace $e^T \alpha = 1$</td>
<td>$n - 1$</td>
<td>${\alpha \in \Delta_n : \alpha_i &gt; 0}$</td>
</tr>
<tr>
<td>$B(x_0, r)$</td>
<td>$\mathbb{R}^n$</td>
<td>$n$</td>
<td>$\text{int}(B(x_0, r))$.</td>
</tr>
</tbody>
</table>

- If $C \neq \emptyset$, then $\text{ri}(C) \neq \emptyset$.
- The three convex sets $\text{ri}(C)$, $C$ and $\text{cl}(C)$ have the same affine hull (and hence the same dimension), the same relative interior and the same closure.
1.3.3 Projection onto convex sets

In what follows, $C$ is a nonempty closed convex set of $\mathbb{R}^n$.

**Definition 2** Let $P_C(x)$ be the minimizer of $\inf \{ \frac{1}{2} \| x - y \|^2 : y \in C \}$. Projection map: $x \rightarrow P_C(x)$.

**Remarks:**

- *Existence:* consider $f_x(y) := \frac{1}{2} \| x - y \|^2$ and $S = \{ y : f_x(y) \leq f_x(c) \}$ for a fixed $c \in C$. Then the set $C \cap S$ is compact and $\inf_C f_x(y)$ is equivalent to $\inf_{C \cap S} f_x(y)$ and $f_x$ is continuous, thus there exists a closest point in $C$ to $x$. inf becomes min.

- *Uniqueness:* relies on convexity. Let $y_1$ and $y_2$ be two distinct solutions. We obtain $f_x(\frac{1}{2}(y_1+y_2)) = \frac{1}{2}(f_x(y_1)+f_x(y_2))-\frac{1}{8}\|y_2-y_1\|^2$, which is contradiction.

**Theorem 1** (Variational inequality) A point $y_x = P_C(x)$ iff

\[ \langle x - y_x, y - x \rangle \leq 0, \quad \text{for all } y \in C \]  \hspace{1cm} (1.2)

**Proof:** Let $y_x = P_C(x)$. Take $y \in C$ arbitrary and $y_x + \alpha(y - y_x) \in C$ for any $\alpha \in (0, 1)$. Then

\[ f_x(y_x) \leq f_x(y_x + \alpha(y - y_x)) = \frac{1}{2} \| y_x - x + \alpha(y - y_x) \|^2 \]

leads to

\[ 0 \leq \alpha \langle x - y_x, y - y_x \rangle + \frac{1}{2} \alpha^2 \| y - y_x \|^2 \]

Divide by $\alpha > 0$ and let $\alpha \searrow 0$ to obtain the variational inequality (1.2).

Conversely, suppose $y_x \in C$ satisfies (1.2), if $y_x = x$, then $y_x$ is the $p_C(x)$. If not, write for arbitrary $y \in C$,

\[ 0 \geq \langle x - y_x, y - y_x \rangle = \langle x - y_x, y - x + x - y_x \rangle = \| x - y_x \|^2 + \langle x - y_x, y - x \rangle \]

By Cauchy-Schwartz inequality, we have $\| x - y_x \|^2 + \langle x - y_x, y - x \rangle \geq \| x - y_x \|^2 - \| x - y_x \| \| y - x \|$, which leads to $\| x - y_x \| \leq \| y - x \|$ for any $y$. Thus $y_x = p_C(x)$.

**Some Remarks:**

- Suppose that $C$ is actually an affine set (manifold), then for any $y$, $\langle x - y_x, y - y_x \rangle \leq 0$. Meanwhile, for $y' = y_x - (y - y_x)$, it is still in $C$. This leads to $y' - y_x = y_x - y$ and $\langle x - y_x, y' - y_x \rangle \leq 0$. Thus the variational inequality becomes $\langle x - y_x, y - y_x \rangle = 0$ for all $y \in C$. This is the classical characterization of the projection onto a subspace, namely $x - y_x \in C^\perp$ (the subspace orthogonal to $C$).

- The set $\{ x \in \mathbb{R}^n : p_C(x) = x \}$ of fixed point of $p_C(x)$ is $C$ itself, from which results that $p_C \circ p_C = p_C$.

- $p_C(x)$ is linear operator iff $C$ is a subspace.
Proposition 1 For all \((x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n\), there holds
\[
\|p_C(x_1) - p_C(x_2)\|^2 \leq \langle p_C(x_1) - p_C(x_2), x_1 - x_2 \rangle.
\]
This leads to two immediate results: monotone increasing and nonexpansive
\[
0 \leq \langle p_C(x_1) - p_C(x_2), x_1 - x_2 \rangle
\]
\[
0 \leq \|p_C(x_1) - p_C(x_2)\| \leq \|x_1 - x_2\|
\]

1.3.4 Separation between convex sets

Theorem 2 (convex and singleton) Let \(C \subset \mathbb{R}^n\) be nonempty closed convex, and let \(x \notin C\), then there exists \(s \in \mathbb{R}^n\) such that
\[
s^T x > \sup\{s^T y : y \in C\}
\]

Proof. Let \(s = x - P_c(x) \neq 0\). we have
\[
0 \geq s^T (y - x + s) = s^T y - s^T x + \|s\|^2
\]
Thus \(s^T x - \|s\|^2 \geq s^T y\) for all \(y \in C\). That means \(s^T x > \sup\{s^T y : y \in C\}\).

Theorem 3 (Separating hyperplane theorem) If \(C\) and \(D\) are disjoint convex sets (\(C \cap D = \emptyset\)), then there exists \(a \neq 0, b\), such that \(a^T x \leq b\) for \(x \in C\) and \(a^T x \geq b\) for \(x \in D\). The hyperplane \(\{x | a^T x = b\}\) separates \(C\) and \(D\).

Definition 3 (Supporting hyperplane) A hyperplane \(H_{a,b}\) \((a \neq 0)\) is said to support the set \(C\) at boundary point \(x_0\) when \(C\) is entirely contained in one of the two closed half-spaces delimited by \(H_{a,b}\): say
\[
a^T y \leq b \quad \text{for all } y \in C
\]
and \(a^T x_0 = b\).

Theorem 4 (Supporting hyperplane theorem) Let \(C\) be a convex set and \(\text{cl}(C) \neq \mathbb{R}^n\), \(C \neq \emptyset\), then there exists a supporting hyperplane at every boundary point \(x \in \text{bd}(C)\).

Proof. Let \(\{x_k\}\) be a sequence such that \(x_k \notin \text{cl}(C)\) and \(\lim_k x_k = x\). For each \(k\) we have that there exists some \(s_k\) of norm 1 such that \(\langle s_k, x_k - y \rangle > 0\) for all \(y \in C \subset \text{cl}(C)\). Extract a subsequence if necessary so that \(s_k \to s\) (note \(s \neq 0\)) and pass to the limit \(\langle s, x - y \rangle \geq 0\) for all \(y \in C\). This is the required result \(\langle s, x \rangle = r \geq \langle s, y \rangle\) for all \(y \in C\).
1.4 Generalized inequalities

A convex cone $K \subset \mathbb{R}^n$ is a proper cone if

- $K$ is closed (contains its boundary)
- $K$ is solid (has nonempty interior)
- $K$ is pointed (contains no line)

Examples

- Nonnegative orthant $K = \mathbb{R}^n = \{ x \in \mathbb{R}^n | x_i \geq 0, i = 1, \cdots, n \}$
- $K = S^n_+$.

Appendix

1.4.1 Separation between convex sets

**Theorem 5 (Strict separation of Convex sets)** Let $C_1, C_2$ be two nonempty closed convex sets with $C_1 \cap C_2 = \emptyset$. If $C_2$ is bounded, there exists $s \in \mathbb{R}^n$ such that

$$\sup_{y \in C_1} s^T y < \min_{y \in C_2} s^T y$$

*Proof.*: Consider $C_1 - C_2$, it is convex and closed ($C_2$ is compact). $C_1$ and $C_2$ is disjoint, then $0 \notin C_1 - C_2$. Thus by the above theorem and $s \in \mathbb{R}^n$ separating $\{0\}$ and $C_1 - C_2$:

$$\sup \{ s^T y : y \in C_1 - C_2 \} < s^T 0 = 0$$

this yields

$$0 > \sup_{y_1 \in C_1} s^T y_1 + \sup_{y_2 \in C_2} s^T (-y_2)$$

$$= \sup_{y_1 \in C_1} s^T y_1 - \inf_{y_2 \in C_2} s^T y_2$$

Since $C_2$ is bounded, and the last infinitum is finite and can be attained.

Note: when $C_1$ and $C_2$ are both unbounded, the strict separation may fail. Properly separated:

$$\sup_{y_1 \in C_1} s^T y_1 \leq \inf_{y_2 \in C_2} s^T y_2, \text{ and } \inf_{y_1 \in C_1} s^T y_1 < \sup_{y_2 \in C_2} s^T y_2$$

**Theorem 6 (Proper separation of convex sets)** If the two nonempty convex sets $C_1$ and $C_2$ satisfy $ri(C_1) \cap ri(C_2) = \emptyset$, they can be properly separated.

Denote $H_{s,r}^- := \{ y \in \mathbb{R}^n : (s, y) \leq r \}$, $\Sigma_S := \{ (s, r) \in \mathbb{R}^n \times \mathbb{R} : S \subset H_{s,r}^- \}$ is the closed half space that contain a given set $S$. 

13
Theorem 7 (Outer description of closed convex sets) The closed convex hull of a set $S \neq \emptyset$ is either the whole of $\mathbb{R}^n$ or the set $C^* = \{z \in \mathbb{R}^n : \langle s, z \rangle \leq r \text{ whenever } \langle s, y \rangle \leq r \text{ for all } y \in S\} = \bigcap_{(s,r) \in \Sigma_S} H_{s,r} = \{z \in \mathbb{R}^n : \langle s, z \rangle \leq \sup_{y \in S} \langle s, y \rangle\}.

Corollary 1 The data $(s_j, r_j) \in \mathbb{R}^n \times \mathbb{R}$ for $j$ in an arbitrary index set $J$ is equivalent to the data of a closed convex set $C$ via the relation

$$
C = \{x \in \mathbb{R}^n : \langle s_j, x \rangle \leq r_j \text{ for } j \in J\}
$$

The $C = \mathbb{R}^n$ with $J = \emptyset$ is an extreme case.

Definition 4 (Polyhedral sets) A closed convex polyhedron is an intersection of finitely many half-spaces. Take $(s_1, r_1), \cdots, (s_m, r_m)$ in $\mathbb{R}^n \times \mathbb{R}$, with $s_i \neq 0$ for $i = 1, \cdots, m$ then define

$$
P = \{x \in \mathbb{R}^n : \langle s_j, x \rangle \leq r_j \text{ for } j = 1, \cdots, m\},
$$

or in matrix form $P = \{x \in \mathbb{R}^n : Ax \leq b\}$.

Definition 5 (Extreme points) Let $C$ be nonempty convex. We say that $x \in C$ is an extreme point of $C$ if there are no two different points $x_1$ and $x_2$ in $C$ such that $x = 1/2(x_1 + x_2)$.

Examples:

- Let $C$ be the unit ball $B(0,1)$. Then every $x$ of norm 1 is an extreme point of $B(0,1)$.
- If $C$ is a convex cone, a nonzero $x \in C$ has no chance of being an extreme point.
- An affine manifold, a half space have no extreme point.
- If $C$ is compact, then the extreme point set is nonempty.

Theorem 8 (H. Minkowski, inner description of convex sets) Let $C$ be compact, convex in $\mathbb{R}^n$. Then $C$ is the convex hull of its extreme points.

1.4.2 Tangent cones and normal cones of convex sets**

Tangent space for smooth surface $\leftrightarrow$ Tangent cones for convex sets.

Orthogonality of subspaces $\leftrightarrow$ Polarity of cones

Consider first an arbitrary set $S$. 

14
Definition 6 (Tangent direction/cone) Let $S \subset \mathbb{R}^n$ be nonempty. We say that $d \in \mathbb{R}^n$ is a direction tangent to $S$ at $x \in S$ when there exists a sequence $x_k \subset S$ and a sequence $t_k$ such that when $k \to \infty$:

$$x_k \to x, t_k \downarrow 0, \frac{x_k - x}{t_k} \to d$$

The set of all such direction is called the tangent cone to $S$ at $x \in S$, denoted as $T_S(x)$.

Obviously 0 is always a tangent direction and if $d$ is tangent so is $\alpha d$ for any $\alpha > 0$. Thus tangent direction is a cone. If $x \in \text{int}(S)$, $T_S(x)$ is clearly the whole space, so that the only interesting point are those on $\text{bd}(S)$.

Proposition 2 (Equivalent definition) A direction $d$ is tangent to $S$ at $x \in S$ if and only if $\exists d_k \to d, t_k \to 0$ such that $x + t_k d_k \in S$ for all $k$.

The tangent cone is closed.

Definition 7 (Normal cone of a closed convex set) The direction $s \in \mathbb{R}^n$ is said to be normal to $C$ at $x \in C$ when

$$\langle s, y - x \rangle \leq 0$$

for all $y \in C$. the set of all such directions is called normal cone to $C$ at $x$, denoted by $N_C(x)$.

Now we assume that $C$ is a closed convex set.

Proposition 3 The tangent cone to a closed convex set $C$ at $x \in C$ is the closure of the cone generated by $C - \{x\}$:

$$T_C(x) = \text{cl}[\mathbb{R}^+(C - x)] = \text{cl}\{d \in \mathbb{R}^n : d = \alpha(y - x), y \in C, \alpha \geq 0\}$$

It is therefore convex.

Definition 8 (Polar) Let $K$ be a convex cone. Its polar (also called negative polar cone) is

$$K^0 := \{s \in \mathbb{R}^n : \langle s, x \rangle \leq 0 \text{ for all } x \in K\}$$

Theorem 9 The normal cone is the polar of the tangent cone. And the tangent cone is the polar of the normal cone: $T_C(x) = \{d \in \mathbb{R}^n : s^T d \leq 0, \text{ for all } s \in N_C(x)\}$.

Remarks:

- The $N_C(x)$ is closed convex cone.
- By the supporting hyperplane theorem, for each $x \in \text{bd}(C)$, there is a $s \neq 0$ which is a normal vector. $(v - p_C(v) \in N_C(p_C(v))$ for all $v \in \mathbb{R}^N$.
- By contrast, $N_C(x) = \{0\}$ for $x \in \text{int}(C)$.
- For a closed half-space $C = H_{s,r} = \{y : \langle s, y \rangle \leq r\}$, the normals at any point of $C$ are the nonnegative multiples of $s$. 

15