Elements of Optimality Analysis

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

Real n-Space; Euclidean Space

- $\mathcal{R}, \quad \mathcal{R}_+, \quad \text{int}\,\mathcal{R}_+$
- $\mathcal{R}^n, \quad \mathcal{R}^n_+, \quad \text{int}\,\mathcal{R}^n_+$
- $\mathbf{x} \geq \mathbf{y}$ means $x_j \geq y_j$ for $j = 1, 2, ..., n$
- $\mathbf{0}$: all zero vector; and $\mathbf{e}$: all one vector
- Column vector:

$$\mathbf{x} = (x_1; x_2; \dots; x_n)$$

and row vector:

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

- Inner-Product:

$$\mathbf{x} \bullet \mathbf{y} := \mathbf{x}^T \mathbf{y} = \sum_{j=1}^{n} x_j y_j$$

- Vector norm: $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}}$, $\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, ..., |x_n|\}$, in general, for $p \geq 1$

$$\|\mathbf{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}$$

- A set of vectors $\mathbf{a}_1, ..., \mathbf{a}_m$ is said to be linearly dependent if there are multipliers $\lambda_1, ..., \lambda_m$, not all zero, the linear combination

$$\sum_{i=1}^{m} \lambda_i \mathbf{a}_i = \mathbf{0}$$

- A linearly independent set of vectors that span R^n is a basis.

Matrices

- $A \in \mathcal{R}^{m\times n}$; $\mathbf{a}_{i.}$, the ith row vector; $\mathbf{a}_{.j}$, the jth column vector; a_{ij}, the i,jth entry
- $\mathbf{0}$: all zero matrix, and I: the identity matrix
- The null space $\mathcal{N}(A)$, the row space $\mathcal{R}(A^T)$, and they are orthogonal.
- $\det(A)$, tr(A): the sum of the diagonal entries of A
- Inner Product:

$$A \bullet B = \mathrm{tr} A^T B = \sum_{i,j} a_{ij} b_{ij}$$

- The operator norm of matrix A:

$$\|A\|^2 := \max_{\mathbf{0}\neq\mathbf{x}\in\mathcal{R}^n} \frac{\|A\mathbf{x}\|^2}{\|\mathbf{x}\|^2}$$

The Frobenius norm of matrix A:

$$\|A\|_f^2 := A \bullet A = \sum_{i,j} a_{ij}^2$$

- Sometimes we use $X = \text{diag}(\mathbf{x})$
- Eigenvalues and eigenvectors

$$A\mathbf{v} = \lambda \cdot \mathbf{v}$$

Symmetric Matrices

- $\mathcal{S}^n$
- The Frobenius norm:

$$\|X\|_f = \sqrt{\text{tr} X^T X} = \sqrt{X \bullet X}$$

- Positive Definite (PD): $Q \succ \mathbf{0}$ iff $\mathbf{x}^T Q \mathbf{x} > 0, \quad \text{for all} \quad \mathbf{x} \neq \mathbf{0}$. The sum of PD matrices is PD.
- Positive Semidefinite (PSD): $Q \succeq \mathbf{0}$ iff $\mathbf{x}^T Q \mathbf{x} \geq 0, \quad \text{for all} \quad \mathbf{x}$. The sum of PSD matrices is PSD.
- PSD matrices: $\mathcal{S}^n_+$, $\quad \text{int}\, \mathcal{S}^n_+$ is the set of all positive definite matrices.

Affine Set

$S \subset R^n$ is affine if

$$[\,\mathbf{x}, \mathbf{y} \in S \text{ and } \alpha \in R\,] \Longrightarrow \alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in S.$$

When x and y are two distinct points in R^n and α runs over R ,

$$\{\mathbf{z} : \mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y}\}$$

is the affine combination of $\mathbf{x}$ and $\mathbf{y}$.

When $0 \le \alpha \le 1$, it is called the convex combination of $\mathbf{x}$ and $\mathbf{y}$.

Convex Set

- Ω is said to be a convex set if for every $\mathbf{x}^1, \mathbf{x}^2 \in \Omega$ and every real number $\alpha \in [0, 1]$, the point $\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2 \in \Omega$.
- Ball and Ellipsoid: for given $\mathbf{y} \in \mathcal{R}^n$ and positive definite matrix Q,

$$E(\mathbf{y}, Q) = \{\mathbf{x} : \ (\mathbf{x}-\mathbf{y})^T Q(\mathbf{x}-\mathbf{y}) \leq 1\}$$

- The Intersection of convex sets is convex
- The convex hull of a set Ω is the intersection of all convex sets containing Ω
- An extreme point in a convex set is a point that cannot be expressed as a convex combination of other two distinct points of the set.
- A set is polyhedral if it has finitely many extreme points; $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0}\}$ and $\{\mathbf{x} : \ A\mathbf{x} \leq \mathbf{b}\}$ are convex polyhedral.

Cone and Convex Cone

- A set C is a cone if $\mathbf{x} \in C$ implies $\alpha\mathbf{x} \in C$ for all $\alpha > 0$
- The intersection of cones is a cone
- A convex cone is a cone and also a convex set
- A pointed cone is a cone that does not contain a line
- Dual cone:

$$C^* := \{\mathbf{y} : \ \mathbf{x} \bullet \mathbf{y} \geq 0 \quad \text{for all} \quad \mathbf{x} \in C\}.$$

The dual cone is a closed convex cone.

Cone Examples

- Example 1: The n-dimensional non-negative orthant, $\mathcal{R}^n_+ = \{\mathbf{x} \in \mathcal{R}^n : \ \mathbf{x} \geq \mathbf{0}\}$, is a convex cone. Its dual is itself.

- Example 2: The set of all PSD matrices in $\mathcal{S}^n$, $\mathcal{S}^n_+$, is a convex cone, called the PSD matrix cone. Its dual is itself.

- Example 3: The set $\{(t; \mathbf{x}) \in \mathcal{R}^{n+1} : \ t \geq \|\mathbf{x}\|_p\}$ for a $p \geq 1$ is a convex cone in $\mathcal{R}^{n+1}$, called the p-order cone. Its dual is the q-order cone with $\frac{1}{p} + \frac{1}{q} = 1$.

- The dual of the second-order cone ($p = 2$) is itself.

Polyhedral Convex Cones

- A cone C is (convex) polyhedral if C can be represented by

$$C = \{\mathbf{x} : A\mathbf{x} \leq \mathbf{0}\}$$

or

$$C = \{A\mathbf{x} : \mathbf{x} \geq \mathbf{0}\}$$

for some matrix A.

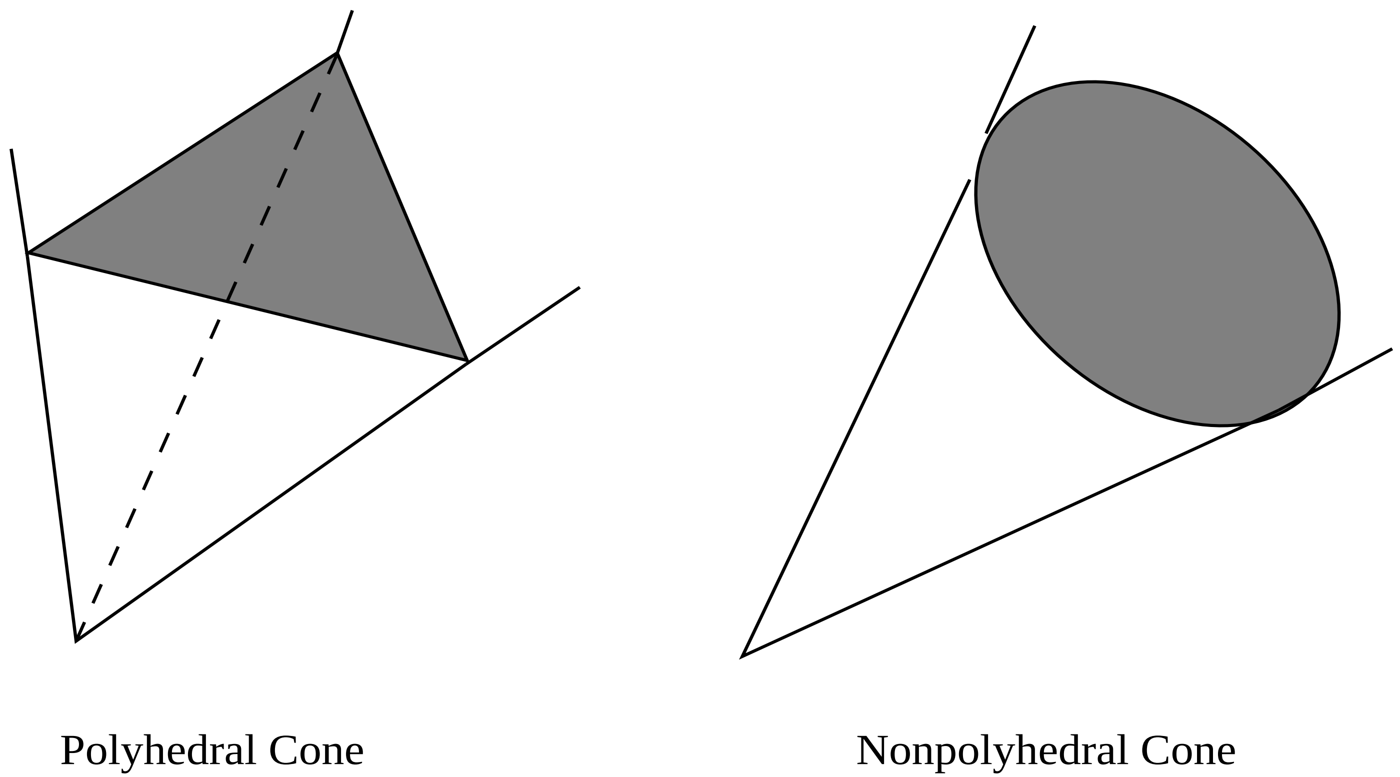

Figure 1: Polyhedral and nonpolyhedral cones.

- The non-negative orthant is a polyhedral cone, and neither the PSD matrix cone nor the second-order cone is polyhedral.

Real Functions

- Continuous functions
- Weierstrass theorem: a continuous function f defined on a compact set (bounded and closed) $\Omega \subset \mathcal{R}^n$ has a minimizer in Ω.
- A function $f(\mathbf{x})$ is called homogeneous of degree k if $f(\alpha\mathbf{x}) = \alpha^k f(\mathbf{x})$ for all $\alpha \geq 0$.
- Example: Let $\mathbf{c} \in \mathcal{R}^n$ be given and $\mathbf{x} \in \operatorname{int} \mathcal{R}^n_+$. Then $\mathbf{c}^T\mathbf{x}$ is homogeneous of degree 1 and

$$\mathcal{P}(\mathbf{x}) = n\log(\mathbf{c}^T\mathbf{x}) - \sum_{j=1}^{n} \log x_j$$

  is homogeneous of degree 0.
- Example: Let $C \in \mathcal{S}^n$ be given and $X \in \operatorname{int} \mathcal{S}^n_+$. Then $\mathbf{x}^T C\mathbf{x}$ is homogeneous of degree 2, $C \bullet X$ and $\det(X)$ are homogeneous of degree

1 and n, respectively; and

$$\mathcal{P}(X) = n\log(C \bullet X) - \log\det(X)$$

is homogeneous of degree 0.

- The gradient vector:

$$\nabla f(\mathbf{x}) = \{\partial f/\partial x_i\}, \quad \text{for} \quad i = 1, ..., n.$$

- The Hessian matrix:

$$\nabla^2 f(\mathbf{x}) = \left\{\frac{\partial^2 f}{\partial x_i \partial x_j}\right\} \quad \text{for} \quad i = 1, ..., n;\ j = 1, ..., n.$$

- Vector function: $\mathbf{f} = (f_1; f_2; ...; f_m)$

- The Jacobian matrix of $\mathbf{f}$ is

$$\nabla \mathbf{f}(x) = \begin{pmatrix} \nabla f_1(\mathbf{x}) \\ \dots \\ \nabla f_m(\mathbf{x}) \end{pmatrix}.$$

- The least upper bound or supremum of f over Ω

$$\sup\{f(\mathbf{x}) : \ x \in \Omega\}$$

and the greatest lower bound or infimum of f over Ω

$$\inf\{f(\mathbf{x}) : \ x \in \Omega\}$$

Convex Functions

- f is a convex function iff for $0 \leq \alpha \leq 1$,

$$f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}).$$

- The sum of convex functions is a convex function; the max of convex functions is a convex function;

- The level set of f is convex:

$$L(z) = \{\mathbf{x} : \quad f(\mathbf{x}) \leq z\}.$$

- Convex set $\{(z; \mathbf{x}) : \ f(\mathbf{x}) \leq z\}$ is called the epigraph of f.

- $tf(\mathbf{x}/t)$ is a convex function of $(t; \mathbf{x})$ for $t > 0$; it's homogeneous with degree 1.

Convex Function Examples

- $\|\mathbf{x}\|_p$ for $p \geq 1$.

$$\|\alpha\mathbf{x}+(1-\alpha)\mathbf{y}\|_p \leq \|\alpha\mathbf{x}\|_p + \|(1-\alpha)\mathbf{y}\|_p \leq \alpha\|\mathbf{x}\|_p + (1-\alpha)\|\mathbf{y}\|_p,$$

  from the triangle inequality.

- $e^{x_1} + e^{x_2} + e^{x_3}$.
- $\log(e^{x_1} + e^{x_2} + e^{x_3})$: we will prove it later.

Example: Proof of convex function

Consider the minimal-objective function of $\mathbf{b}$ for fixed A and $\mathbf{c}$:

$$\begin{array}{rl} z(\mathbf{b}) := \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}, \end{array}$$

where $f(\mathbf{x})$ is a convex function.

Show that $z(\mathbf{b})$ is a convex function in $\mathbf{b}$.

Theorems on functions

Taylor's theorem or the mean-value theorem:

Theorem 1 *Let* $f \in C^1$ *be in a region containing the line segment* $[\mathbf{x}, \mathbf{y}]$*. Then there is a* α*,* $0 \le \alpha \le 1$*, such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})(\mathbf{y} - \mathbf{x}).$$

Furthermore, if $f \in C^2$ *then there is a* α*,* $0 \le \alpha \le 1$*, such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y}-\mathbf{x}) + (1/2)(\mathbf{y}-\mathbf{x})^T \nabla^2 f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})(\mathbf{y}-\mathbf{x}).$$

Theorem 2 *Let* $f \in C^1$*. Then* f *is convex over a convex set* Ω *if and only if*

$$f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

for all $\mathbf{x},\ \mathbf{y} \in \Omega$*.*

Theorem 3 *Let $f \in C^2$. Then f is convex over a convex set Ω if and only if the Hessian matrix of f is positive semi-definite throughout Ω.*

Theorem 4 *Suppose we have a set of m equations in n variables*

$$h_i(\mathbf{x}) = 0, \; i = 1, ..., m$$

where $h_i \in C^p$ for some $p \geq 1$. Then, a set of m variables can be expressed as implicit functions of the other $n - m$ variables in the neighborhood of a feasible point when the Jacobian matrix of the m variables is nonsingular.

Known Inequalities

- Cauchy-Schwarz: given $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$, $|\mathbf{x}^T\mathbf{y}| \le \|\mathbf{x}\|_p \|\mathbf{y}\|_q$, where $\frac{1}{p} + \frac{1}{q} = 1$ and $p \ge 1$.
- Triangle: given $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$, $\|\mathbf{x} + \mathbf{y}\|_p \le \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$ for $p \ge 1$.
- Arithmetic-geometric mean: given $\mathbf{x} \in \mathcal{R}^n_+$,

$$\frac{\sum x_j}{n} \ge \left(\prod x_j\right)^{1/n}.$$

System of linear equations

Given $A \in \mathcal{R}^{m \times n}$ and $\mathbf{b} \in \mathcal{R}^m$, the problem is to determine n unknowns from m linear equations:

$$A\mathbf{x} = \mathbf{b}$$

Theorem 5 *Let $A \in \mathcal{R}^{m \times n}$ and $\mathbf{b} \in \mathcal{R}^m$. The system $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}\}$ has a solution if and only if that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{b}^T\mathbf{y} \neq 0$ has no solution.*

A vector $\mathbf{y}$, with $A^T\mathbf{y} = 0$ and $\mathbf{b}^T\mathbf{y} \neq 0$, is called an infeasibility certificate for the system.

Alternative system pairs: $\{\mathbf{x} : \ A\mathbf{x} = \mathbf{b}\}$ and $\{\mathbf{y} : \ A^T\mathbf{y} = \mathbf{0}, \ \mathbf{b}^T\mathbf{y} \neq 0\}$.

Gaussian elimination method

$$\begin{pmatrix} a_{11} & A_{1.} \\ 0 & A' \end{pmatrix} \begin{pmatrix} x_1 \\ x' \end{pmatrix} = \begin{pmatrix} b_1 \\ b' \end{pmatrix}.$$

$$A = L \begin{pmatrix} U & C \\ 0 & 0 \end{pmatrix}$$

The method runs in $O(n^3)$ time for n equations with n unknowns.