Mathematical Preliminaries

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

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

LY, Appendices A, B, and Chapter 1.
Real $n$-Space; Euclidean Space

- $\mathbb{R}$, $\mathbb{R}_+$, $\text{int} \mathbb{R}_+$
- $\mathbb{R}^n$, $\mathbb{R}_+^n$, $\text{int} \mathbb{R}_+^n$
- $\mathbf{x} \geq \mathbf{y}$ means $x_j \geq y_j$ for $j = 1, 2, \ldots, n$
- $\mathbf{0}$ denotes the zero vector and $\mathbf{e}$ denotes the vector of ones
- Inner-Product:
  \[ \mathbf{x} \cdot \mathbf{y} := \mathbf{x}^T \mathbf{y} = \sum_{j=1}^{n} x_j y_j \]
- Norm: $\|\mathbf{x}\|_2 := \sqrt{\mathbf{x}^T \mathbf{x}}$, $\|\mathbf{x}\|_{\infty} := \max\{|x_1|, |x_2|, \ldots, |x_n|\}$, and $\|\mathbf{x}\|_p := \left( \sum_{j=1}^{n} |x_j|^p \right)^{1/p}$
- The dual of the $p$ norm, denoted by $\|\cdot\|^*$, is the $q$ norm where $\frac{1}{p} + \frac{1}{q} = 1$ and $1 \leq p, q < \infty$. 
• Column vector:

\[ \mathbf{x} = (x_1; x_2; \ldots; x_n) \]

and row vector:

\[ \mathbf{x} = (x_1, x_2, \ldots, x_n) \]

• A set of vectors \( \mathbf{a}_1, \ldots, \mathbf{a}_m \) is said to be linearly dependent if there exists some scalars \( \lambda_1, \ldots, \lambda_m \), not all zero, such that the linear combination

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

• A linearly independent set of vectors that spans \( \mathbb{R}^n \) is a basis.
Matrices

- $\mathcal{R}^{m \times n}$, $a_i$, $a_j$, $a_{ij}$

- $A_I$ denotes the submatrix of $A$ whose rows belong to $I$, $A_J$ denotes the submatrix whose columns belong to $J$, and $A_{IJ}$ denotes the submatrix whose rows belong to $I$ and whose columns belong to $J$.

- $0$ denotes the zero matrix and $I$ denotes the identity matrix

- $\mathcal{N}(A)$, $\mathcal{R}(A)$:

  **Theorem 1** Each linear subspace of $\mathcal{R}^n$ can be generated by finitely many vectors and is also an intersection of finitely many hyperplanes; that is, for each linear subspace of $L$ of $\mathcal{R}^n$ there are matrices $A$ and $C$ such that $L = \mathcal{N}(A) = \mathcal{R}(C)$.

- $\text{det}(A)$, $\text{tr}(A)$
• Inner Product:

$$A \bullet B := \text{tr}(A^T B) = \sum_{i,j} a_{ij} b_{ij}$$

• The operator norm of $A$:

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

• Sometimes we use $X = \text{diag}(x)$

• Eigenvalues and eigenvectors

$$Av = \lambda v$$
Symmetric Matrices

- $S^n$

- The Frobenius norm:

$$\|X\|_f := \sqrt{\text{tr}(X^T X)} = \sqrt{X \cdot X}$$

- Positive Definite (PD): $Q \succ 0$ iff $x^T Q x > 0$, for all $x \neq 0$

- Positive SemiDefinite (PSD): $Q \succeq 0$ iff $x^T Q x \geq 0$, for all $x$

- The set of PSD matrices: $S^n_+, \ \text{int} \ S^n_+$
Cauchy-Schwarz Inequality: given $x, y \in \mathbb{R}^n$, we have $x^T y \leq \|x\| \|y\|$. 

Triangle Inequality: given $x, y \in \mathbb{R}^n$, we have $\|x + y\| \leq \|x\| + \|y\|$. 

Arithmetic Mean-Geometric Mean Inequality: given $x \in \mathbb{R}_+^n$, we have

$$\frac{\sum x_j}{n} \geq \left(\prod x_j\right)^{1/n}.$$
Hyperplane and Half-spaces

\[ H = \{ x : ax = \sum_{j=1}^{n} a_j x_j = b \} \]

\[ H^+ = \{ x : ax = \sum_{j=1}^{n} a_j x_j \leq b \} \]

\[ H^- = \{ x : ax = \sum_{j=1}^{n} a_j x_j \geq b \} \]
Figure 1: Plane and Half-Spaces
System of Linear Equations

Solve for $x \in \mathbb{R}^n$ from:

\[
\begin{align*}
    a_1 x &= b_1 \\
    a_2 x &= b_2 \\
    \vdots & \quad \vdots \\
    a_m x &= b_m
\end{align*}
\]

⇒ $Ax = b$
Figure 2: System of Linear Equations
**Fundamental Theorem of Linear Equations**

**Theorem 2**  Given $A \in \mathcal{R}^{m \times n}$ and $b \in \mathcal{R}^m$, the system $\{x : Ax = b\}$ has a solution if and only if that $A^T y = 0$ and $b^T y \neq 0$ has no solution.

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

**Example** Let $A = (1; -1)$ and $b = (1; 1)$. Then, $y = (1/2; 1/2)$ is an infeasibility certificate.

**Alternative systems:** $\{x : Ax = b\}$ and $\{y : A^T y = 0, \ b^T y \neq 0\}$. 
Figure 3: $\mathbf{b}$ is not in the set $\{A\mathbf{x} : \mathbf{x}\}$, and $\mathbf{y}$ is the distance vector from $\mathbf{b}$ to the set.
Affine, Convex, Linear, and Conic Combinations

When \( x \) and \( y \) are two distinct points in \( \mathbb{R}^n \) and \( \alpha \) runs over \( \mathbb{R} \),

\[
\{ z : z = \alpha x + (1 - \alpha)y \}
\]

is the line connecting \( x \) and \( y \). When \( 0 \leq \alpha \leq 1 \), it is called the convex combination of \( x \) and \( y \) and it is the line segment between \( x \) and \( y \). Also, the set

\[
\{ z : z = \alpha x + \beta y \},
\]

for multipliers \( \alpha \) and \( \beta \) is the linear combination of \( x \) and \( y \), and it is the hyperplane containing the origin and \( x \) and \( y \). When \( \alpha \geq 0 \) and \( \beta \geq 0 \), such \( z \) is called a conic combination.
Convex Sets

- Set notations: $x \in \Omega$, $y \notin \Omega$, $S \cup T$, and $S \cap T$

- $\Omega$ is said to be a convex set if for every $x^1, x^2 \in \Omega$ and every real number $\alpha \in [0, 1]$, the linear combination satisfies $\alpha x^1 + (1 - \alpha) x^2 \in \Omega$.

- The convex hull of a set $\Omega$ is the intersection of all convex sets containing $\Omega$.

- Any Intersection of convex sets is convex.

- A point in a convex set is an extreme point if and only if it cannot be represented as a convex combination of two distinct points in the set.

- A set is polyhedral if and only if it has finite number of extreme points.
Proof of convex set

• All solutions to the system of linear equations \( \{x : Ax = b\} \) form a convex set.

• All solutions to the system of linear inequalities \( \{x : Ax \leq b\} \) form a convex set.

• All solutions to the system of linear equations and inequalities \( \{x : Ax = b, x \geq 0\} \) form a convex set.

• Ball is a convex set. The ball with a center \( y \in \mathbb{R}^n \) and a radius \( r > 0 \) is denoted by \( B(y, r) := \{x : \|x - y\| \leq r\} \).

• Ellipsoid is a convex set. The ellipsoid with a center \( y \in \mathbb{R}^n \) and a positive definite matrix \( Q \) is denoted by \( E(y, Q) = \{x : (x - y)^T Q (x - y) \leq 1\} \).
More Proofs on Convexity

Given a matrix $A$, let’s consider the set $B$ of all $b$ such that the set

$$\{x : Ax = b, \ x \geq 0\}$$

is feasible. Show that $B$ is a convex set.

Example:

$$B = \{b : \ (x_1, x_2) : \ x_1 + x_2 = b, \ (x_1, x_2) \geq 0\}$$

is feasible.
Convex Cones

- A set $C$ is a cone if $x \in C$ implies $\alpha x \in C$ for all $\alpha > 0$.

- A **convex cone** is a cone which is also convex.

- Dual cone:
  \[ C^* := \{ y : y \cdot x \geq 0 \quad \text{for all} \quad x \in C \} \]
Cone Examples

- Example 2.1: The $n$-dimensional non-negative orthant
  \[ \mathcal{R}^n_+ = \{ x \in \mathbb{R}^n : x \geq 0 \} \]
  is a convex cone.

- Example 2.2: The set of all positive semi-definite matrices in $\mathcal{S}^n$, $\mathcal{S}^n_+$, is a convex cone, called the positive semi-definite matrix cone.

- Example 2.3: The set $\mathcal{N}_2^n := \{ x \in \mathcal{R}^n : x_1 \geq \| x_{-1} \| \}$ is a convex cone in $\mathcal{R}^n$ called the second-order cone.

- Example 2.4: The set $\mathcal{N}_p^n := \{ x \in \mathcal{R}^n : x_1 \geq \| x_{-1} \|_p \}$ is a convex cone in $\mathcal{R}^n$ called the $p$-order cone with $p \geq 1$. 
A cone $C$ is a (convex) polyhedral if $C$ can be represented as

$$C = \{ x : Ax \leq 0 \} \quad \text{or} \quad \{ x : x = Ay, \ y \geq 0 \}$$

for some matrix $A$. In the latter case, $C$ is generated by the columns of $A$. 

**Polyhedral Convex Cones**
The nonnegative orthant is a polyhedral cone but the second-order cone is not polyhedral.

Figure 4: Polyhedral and non-polyhedral cones.
The following theorem states that a polyhedral cone can be generated by a set of basic directional vectors.

**Theorem 3** Given matrix $A \in \mathbb{R}^{m \times n}$ where $n > m$, take a convex polyhedral cone $C = \{Ax : x \geq 0\}$. Then for any $b \in C$,

$$b = \sum_{i=1}^{d} a_{j_i}x_{j_i}, \ x_{j_i} \geq 0, \forall i$$

for some linearly independent vectors $a_{j_1}, \ldots, a_{j_d}$ chosen from $a_1, \ldots, a_n$. 
Real Functions

- **Continuous functions** $C$

- **Weierstrass theorem**: a continuous function $f(x)$ defined on a compact set (bounded and closed) $\Omega \subset \mathbb{R}^n$ has a minimizer in $\Omega$.

- The least upper bound or supremum of $f$ over $\Omega$

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

and the greatest lower bound or infimum of $f$ over $\Omega$

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

- A function $f(x)$ is said to be **homogeneous** of degree $k$ if

$$f(\alpha x) = \alpha^k f(x)$$

for all $\alpha \geq 0$.

Let $c \in \mathbb{R}^n$ be given and $x \in \text{int} \, \mathbb{R}_+^n$. Then $c^T x$ is homogeneous of
degree 1 and

\[ \phi(x) = n \log(c^T x) - \sum_{j=1}^{n} \log x_j \]

is homogeneous of degree 0.

Let \( C \in S^n \) be given and \( X \in \text{int} S^n_+ \). Then \( x^T C x \) is homogeneous of degree 2, \( C \cdot X \) and \( \det(X) \) are homogeneous of degree 1 and \( n \), respectively, and

\[ \Phi(X) = n \log(C \cdot X) - \log \det(X) \]

is homogeneous of degree 0.

- The gradient vector \( C^1 \):

\[ \nabla f(x) = \{ \partial f / \partial x_i \}, \quad \text{for} \quad i = 1, \ldots, n. \]
• The Hessian matrix $C^2$:

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

• Vector function: $f = (f_1; f_2; \ldots; f_m)$

• The Jacobian matrix of $f$:

\[
\nabla f(x) := \begin{pmatrix}
\nabla f_1(x) \\
\vdots \\
\nabla f_m(x)
\end{pmatrix}
\]
Convex Functions

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

\[ f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y). \]

- The level set of $f$ is convex:

\[ L(z) = \{ x : f(x) \leq z \}. \]

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

- $tf(x/t)$ is a convex function of $(t; x)$ for $t > 0$ and it’s homogeneous of degree 1.
Proof of convex function

Consider the minimal-objective value function of $b$ for fixed $A$ and $c$:

$$z(b) := \text{minimize} \quad c^T x$$

subject to $A x = b$, $x \geq 0$.

Show that $z(b)$ is a convex function in $b$ for all feasible $b$. 
Taylor’s theorem or the mean-value theorem:

**Theorem 4** Let \( f \in C^1 \) be in a region containing the line segment \([x, y]\). Then there is \( \alpha \) with \( 0 \leq \alpha \leq 1 \) such that

\[
f(y) = f(x) + \nabla f(\alpha x + (1 - \alpha)y)(y - x).
\]

Furthermore, if \( f \in C^2 \) then there is \( \alpha \) with \( 0 \leq \alpha \leq 1 \) such that

\[
f(y) = f(x) + \nabla f(x)(y - x) + (1/2)(y - x)^T \nabla^2 f(\alpha x + (1 - \alpha)y)(y - x).
\]

**Theorem 5** Let \( f \in C^1 \). Then \( f \) is convex over a convex set \( \Omega \) if and only if

\[
f(y) \geq f(x) + \nabla f(x)(y - x)
\]

for all \( x, y \in \Omega \).
Theorem 6 Let \( f \in C^2 \). Then \( f \) is convex over a convex set \( \Omega \) if and only if the Hessian matrix of \( f \) is positive semi-definite throughout \( \Omega \).
Linear Least Squares Problems

Given $A \in \mathcal{R}^{m \times n}$ and $c \in \mathcal{R}^n$, 

\[(LS) \quad \text{minimize} \quad \|c - A^T y\|^2 \]
\[\text{subject to} \quad y \in \mathcal{R}^m.\]

A close form solution:

$$AA^T y = Ac \quad \text{or} \quad y = (AA^T)^{-1} Ac.$$ 

$$c - A^T y = c - A^T (AA^T)^{-1} Ac = c - Pc$$

Projection matrix: $P = A^T (AA^T)^{-1} A$ or $P = I - A^T (AA^T)^{-1} A.$
Figure 5: Projection of $c$ onto a subspace
Choleski decomposition method

\[ AA^T = LL^T \]

\[ LL^Ty^* = Ac \]
System of nonlinear equations

Given $f(x) : \mathbb{R}^n \to \mathbb{R}^n$, the problem is to solve $n$ equations for $n$ unknowns:

$$f(x) = 0.$$ 

Given a point $x^k$, Newton’s Method sets

$$f(x) \simeq f(x^k) + \nabla f(x^k)(x - x^k) = 0.$$ 

or solve for direction vector $d_x$:

$$x^{k+1} = x^k - (\nabla f(x^k))^{-1}f(x^k)$$

and

$$\nabla f(x^k)d_x = -f(x^k) \quad \text{and} \quad x^{k+1} = x^k + d_x.$$
Figure 6: Newton’s method for root finding
The quasi Newton method

\[ x^{k+1} = x^k - \alpha (\nabla f(x^k))^{-1} f(x^k) \]

where scalar \( \alpha \geq 0 \) is called the step-size. More generally, we may use

\[ x^{k+1} = x^k - \alpha M^k f(x^k) \]

where \( M^k \) is an \( n \times n \) symmetric matrix. In particular, if \( M^k = I \), then the method is called the gradient method, where \( f \) is viewed as the gradient vector of a real function.
Convergence and Big O

- \( \{x^k\}_{0}^{\infty} \) denotes a sequence \( x^0, x^1, x^2, \ldots, x^k, \ldots \).
- We denote \( x^k \to \bar{x} \) when \( \|x^k - \bar{x}\| \to 0 \).
- \( g(x) \geq 0 \) is a real valued function of a real nonnegative variable, the notation \( g(x) = O(x) \) means that \( g(x) \leq \bar{c}x \) for some constant \( \bar{c} \).
- \( g(x) = \Omega(x) \) means that \( g(x) \geq cx \) for some constant \( c \).
- \( g(x) = \theta(x) \) means that \( cx \leq g(x) \leq \bar{c}x \).
- \( g(x) = o(x) \) means that \( g(x) \) goes to zero faster than \( x \) does:
  \[
  \lim_{x \to 0} \frac{g(x)}{x} = 0
  \]