Convex Optimization Theory and Applications

Lecture 1
Introduction

Prof. Chun-Hung Liu

Dept. of Electrical and Computer Engineering
National Chiao Tung University
Fall 2016
Course Information

• **Instructor:** Prof. Chun-Hung Liu (劉俊宏)
  • Email: chungliu@nctu.edu.tw
  • Phone: 03-571-2121 ext. 54512
  • Office: ED 703

• **Class Website:** e3 campus [http://web.it.nctu.edu.tw/~chungliu/](http://web.it.nctu.edu.tw/~chungliu/)
  • Textbooks: *Convex Optimization, S. Boyd and and L. Vandenberghe, Cambridge University, 2004*. (This book is downloadable)
  • All course materials are on the above website.
  • Prerequisites: Undergraduate linear algebra and/or probability

• Office Hours: 9:00am~12pm on Wednesday (But, you are always welcome to talk to me when you see me in my office)

• **Teaching Assistant:** 胡恆鳴、陳柏嘉 (ED 717)

(* Part of the lecture slides is based on the slides made by Profs. S. Boyd and L. Vandenberghe)
Course Grading and Coverage

- **Grading**
  - Homework (20%)
    - Including written and computer simulation problems
  - Midterm Exams (45%)
    - Two short exams in class
  - Final Exam/Project (35~40%)
    - Which one would you like?

- **Categories of Optimization Theory**
  - Linear programming
  - Nonlinear programming
  - Convex/Non-Convex optimization
    - Linear and Nonlinear
  - More advanced optimization
    - Dynamic programming
    - Stochastic dynamic programming, Markov decision processes
Introduction

• Mathematical optimization

• Least-squares and linear programming

• Convex optimization

• Example

• Linear programming

• Nonlinear optimization

• Brief history of convex optimization
Mathematical Optimization

(mathematical) optimization problem

minimize \( f_0(x) \)
subject to \( f_i(x) \leq b_i, \quad i = 1, \ldots, m \)

• \( x = (x_1, \ldots, x_n) \): optimization variables

• \( f_0 : \mathbb{R}^n \to \mathbb{R} \): objective function

• \( f_i : \mathbb{R}^n \to \mathbb{R}, \quad i = 1, \ldots, m \): constraint functions

optimal solution \( x^* \) has smallest value of \( f_0 \) among all vectors that satisfy the constraints
Examples

• **Portfolio optimization**
  • variables: amounts invested in different assets
  • constraints: budget, max./min. investment per asset, minimum Return
  • objective: overall risk or return variance

• **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
Solving Optimization Problems

• **General optimization problem**
  • very difficult to solve *(Does this truth frustrate you?)*
  • methods involve some compromise, e.g., very long computation time, or not always finding the solution

• **Exceptions**: Certain problem classes can be solved efficiently and reliably
  • least-squares problems
  • linear programming problems
  • convex optimization problems
Least-Squares

minimize $\|Ax - b\|_2^2$

solving least-squares problems

- analytical solution: $x^* = (A^TA)^{-1}A^Tb$
- reliable and efficient algorithms and software
- computation time proportional to $n^2k$ ($A \in \mathbb{R}^{k \times n}$); less if structured
- a mature technology

using least-squares

- least-squares problems are easy to recognize
- a few standard techniques increase flexibility (e.g., including weights, adding regularization terms)
Linear Programming

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad a_i^T x \leq b_i, \quad i = 1, \ldots, m
\end{align*}
\]

solving linear programs

- no analytical formula for solution
- reliable and efficient algorithms and software
- computation time proportional to \( n^2m \) if \( m \geq n \); less with structure
- a mature technology

using linear programming

- not as easy to recognize as least-squares problems
- a few standard tricks used to convert problems into linear programs (\( e.g., \) problems involving \( l_1 \)- or \( l_\infty \)-norms, piecewise-linear functions)
Convex Optimization Problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq b_i, \quad i = 1, \ldots, m
\end{align*}
\]

- objective and constraint functions are convex:

\[
f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)
\]

if $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$

- includes least-squares problems and linear programs as special cases
Solving Optimization Problems

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq b_i, \quad i = 1, \ldots, m
\end{align*}
\]

- objective and constraint functions are convex:

\[
f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)
\]

if \( \alpha + \beta = 1, \alpha \geq 0, \beta \geq 0 \)

- includes least-squares problems and linear programs as special cases
Solving Optimization Problems

solving convex optimization problems

- no analytical solution
- reliable and efficient algorithms
- computation time (roughly) proportional to $\max\{n^3, n^2m, F\}$, where $F$ is cost of evaluating $f_i$'s and their first and second derivatives
- almost a technology

using convex optimization

- often difficult to recognize
- many tricks for transforming problems into convex form
- surprisingly many problems can be solved via convex optimization
**Example**

$m$ lamps illuminating $n$ (small, flat) patches

\[ I_k = \sum_{j=1}^{m} a_{kj} p_j, \quad a_{kj} = r_{kj}^{-2} \max\{\cos \theta_{kj}, 0\} \]

**Problem**: achieve desired illumination $I_{\text{des}}$ with bounded lamp powers

\[
\begin{align*}
\text{minimize} & \quad \max_{k=1,\ldots,n} | \log I_k - \log I_{\text{des}} | \\
\text{subject to} & \quad 0 \leq p_j \leq p_{\text{max}}, \quad j = 1, \ldots, m
\end{align*}
\]
How to Solve?

1. use uniform power: \( p_j = p \), vary \( p \)
2. use least-squares:
   \[
   \text{minimize} \quad \sum_{k=1}^{n} (I_k - I_{\text{des}})^2
   \]
   round \( p_j \) if \( p_j > p_{\text{max}} \) or \( p_j < 0 \)
3. use weighted least-squares:
   \[
   \text{minimize} \quad \sum_{k=1}^{n} (I_k - I_{\text{des}})^2 + \sum_{j=1}^{m} w_j (p_j - p_{\text{max}}/2)^2
   \]
   iteratively adjust weights \( w_j \) until \( 0 \leq p_j \leq p_{\text{max}} \)
4. use linear programming:
   \[
   \text{minimize} \quad \max_{k=1,\ldots,n} |I_k - I_{\text{des}}|
   \]
   subject to \( 0 \leq p_j \leq p_{\text{max}}, \quad j = 1, \ldots, m \)
   which can be solved via linear programming
   of course these are approximate (suboptimal) ‘solutions’
How to Solve?

5. use convex optimization: problem is equivalent to

\[
\begin{align*}
\text{minimize} & \quad f_0(p) = \max_{k=1, \ldots, n} h(I_k/I_{\text{des}}) \\
\text{subject to} & \quad 0 \leq p_j \leq p_{\text{max}}, \quad j = 1, \ldots, m
\end{align*}
\]

with \( h(u) = \max\{u, 1/u\} \)

\( f_0 \) is convex because maximum of convex functions is convex.

**exact** solution obtained with effort \( \approx \) modest factor \( \times \) least-squares effort.
Nonlinear Optimization

• Traditional techniques for general nonconvex problems involve compromises
  • local optimization methods (nonlinear programming)
  • find a point that minimizes $f_0$ among feasible points near it
  • fast, can handle large problems
  • require initial guess

• provide no information about distance to (global) optimum

• global optimization methods
  • find the (global) solution
  • worst-case complexity grows exponentially with problem size

• these algorithms are often based on solving convex subproblems
Brief History of Convex Optimization

• **Theory (convex analysis): ca1900–1970**

• **Algorithms**
  • 1947: simplex algorithm for linear programming (Dantzig)
  • 1960s: early interior-point methods (Fiacco & McCormick, Dikin, . . . )
  • 1970s: ellipsoid method and other subgradient methods
  • 1980s: polynomial-time interior-point methods for linear programming (Karmarkar 1984)
  • late 1980s–now: polynomial-time interior-point methods for nonlinear
  • convex optimization (Nesterov & Nemirovski 1994)

• **Applications**
  • before 1990: mostly in operations research; few in engineering
  • since 1990: many new applications in engineering (control, signal processing, communications, circuit design, machine learning, . . . ); new problem classes
  • (semidefinite and second-order cone programming, robust optimization)
Some Useful Reviews

(column) vector $x \in \mathbb{R}^n$:

$$x = \begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_n
\end{bmatrix}$$

- $x_i \in \mathbb{R}$: $i$th component or element of $x$
- also written as $x = (x_1, x_2, \ldots, x_n)$

some special vectors:

- $x = 0$ (zero vector): $x_i = 0$, $i = 1, \ldots, n$
- $x = 1$: $x_i = 1$, $i = 1, \ldots, n$
- $x = e_i$ (ith basis vector or ith unit vector): $x_i = 1$, $x_k = 0$ for $k \neq i$

($n$ follows from context)
Vector Operations

 multiplying a vector $x \in \mathbb{R}^n$ with a scalar $\alpha \in \mathbb{R}$:

$$\alpha x = \begin{bmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{bmatrix}$$

 adding and subtracting two vectors $x, y \in \mathbb{R}^n$:

$$x + y = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}, \quad x - y = \begin{bmatrix} x_1 - y_1 \\ \vdots \\ x_n - y_n \end{bmatrix}$$
Inner Products

\[ x, y \in \mathbb{R}^n \]
\[ \langle x, y \rangle := x_1y_1 + x_2y_2 + \cdots + x_ny_n = x^T y \]

important properties

- \[ \langle \alpha x, y \rangle = \alpha \langle x, y \rangle \]
- \[ \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \]
- \[ \langle x, y \rangle = \langle y, x \rangle \]
- \[ \langle x, x \rangle \geq 0 \]
- \[ \langle x, x \rangle = 0 \iff x = 0 \]

linear function: \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is linear, i.e.

\[ f(\alpha x + \beta y) = \alpha f(x) + \beta f(y), \]

if and only if \( f(x) = \langle a, x \rangle \) for some \( a \)
Euclidean Norm

for $x \in \mathbb{R}^n$ we define the (Euclidean) norm as

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x}$$

$\|x\|$ measures length of vector (from origin)

important properties:

- $\|\alpha x\| = |\alpha| \|x\|$ (homogeneity)
- $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)
- $\|x\| \geq 0$ (nonnegativity)
- $\|x\| = 0 \iff x = 0$ (definiteness)

$distance$ between vectors: $\text{dist}(x, y) = \|x - y\|$
Inner Products and Angles

angle between vectors in $\mathbb{R}^n$:

$$\theta = \angle(x, y) = \cos^{-1} \frac{x^T y}{\|x\|\|y\|}$$

i.e., $x^T y = \|x\|\|y\| \cos \theta$

- $x$ and $y$ aligned: $\theta = 0$; $x^T y = \|x\|\|y\|$  
- $x$ and $y$ opposed: $\theta = \pi$; $x^T y = -\|x\|\|y\|$  
- $x$ and $y$ orthogonal: $\theta = \pi/2$ or $-\pi/2$; $x^T y = 0$ (denoted $x \perp y$)  
- $x^T y > 0$ means $\angle(x, y)$ is acute; $x^T y < 0$ means $\angle(x, y)$ is obtuse
Cauchy-Schwarz Inequality

Cauchy-Schwarz inequality:

\[ |x^T y| \leq \|x\| \|y\| \]

**Projection of** $x$ **on** $y$

The projection of $x$ on $y$ is given by

\[
\left( \frac{x^T y}{\|y\|^2} \right) y
\]
Hyperplanes

A hyperplane in $\mathbb{R}^n$ can be defined as:

$$\{x \mid a^T x = b\} \quad (a \neq 0)$$

- Solution set of one linear equation $a_1 x_1 + \cdots + a_n x_n = b$ with at least one $a_i \neq 0$
- Set of vectors that make a constant inner product with vector $a = (a_1, \ldots, a_n)$ (the normal vector)

In $\mathbb{R}^2$: a line, in $\mathbb{R}^3$: a plane, \ldots
Halfspaces

(closed) halfspace in $\mathbb{R}^n$: 

$$\{x \mid a^T x \leq b\} \quad (a \neq 0)$$

- solution set of one linear inequality $a_1x_1 + \cdots + a_nx_n \leq b$ with at least one $a_i \neq 0$
- $a = (a_1, \ldots, a_n)$ is the (outward) normal

- $\{x \mid a^T x < b\}$ is called an open halfspace
Affine Sets

solution set of a set of linear equations

\[
\begin{align*}
    a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
    a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_1 \\
    &\vdots \\
    a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{align*}
\]

intersection of \( m \) hyperplanes with normal vectors \( a_i = (a_{i1}, a_{i2}, \ldots, a_{in}) \) (w.l.o.g., all \( a_i \neq 0 \))

in matrix notation:

\[
Ax = b
\]

with

\[
A = \begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}, \quad b = \begin{bmatrix}
    b_1 \\
    b_2 \\
    \vdots \\
    b_m
\end{bmatrix}
\]
Polyhedra

solution set of system of linear inequalities

\[
\begin{align*}
    a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & \leq b_1 \\
    \vdots & \\
    a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & \leq b_m
\end{align*}
\]

intersection of \( m \) halfspaces, with normal vectors \( a_i = (a_{i1}, a_{i2}, \ldots, a_{in}) \) (w.l.o.g., all \( a_i \neq 0 \))
**Polyhedra**

Matrix notation

\[ Ax \leq b \]

with

\[
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}, \quad b = \begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_m
\end{bmatrix}
\]

\( Ax \leq b \) stands for *componentwise* inequality, i.e., for \( y, z \in \mathbb{R}^n \),

\[ y \leq z \iff y_1 \leq z_1, \ldots, y_n \leq z_n \]
Examples of Polyhedra

- a hyperplane \( \{ x \mid a^T x = b \} \):
  \[
  a^T x \leq b, \quad a^T x \geq b
  \]

- solution set of system of linear equations/inequalities
  \[
  a_i^T x \leq b_i, \quad i = 1, \ldots, m, \quad c_i^T x = d_i, \quad i = 1, \ldots, p
  \]

- a slab \( \{ x \mid b_1 \leq a^T x \leq b_2 \} \)

- the probability simplex \( \{ x \in \mathbb{R}^n \mid 1^T x = 1, \ x_i \geq 0, \ i = 1, \ldots, n \} \)

- (hyper)rectangle \( \{ x \in \mathbb{R}^n \mid l \leq x \leq u \} \) where \( l < u \)
Sets of Linear Equations

\[ Ax = y \]

given \( A \in \mathbb{R}^{m \times n}, y \in \mathbb{R}^m \)

- solvable if and only if \( y \in \mathcal{R}(A) \)
- unique solution if \( y \in \mathcal{R}(A) \) and \( \text{rank}(A) = n \)
- general solution set:
  \[ \{ x_0 + v \mid v \in \mathcal{N}(A) \} \]
  where \( Ax_0 = y \)

A square and invertible: unique solution for every \( y \):

\[ x = A^{-1}y \]
Polyhedron (inequality form)

\[ A = [a_1 \cdots a_m]^T \in \mathbb{R}^{m \times n}, \ b \in \mathbb{R}^m \]

\[ \mathcal{P} = \{ x \mid Ax \leq b \} = \{ x \mid a_i^T x \leq b_i, \ i = 1, \ldots, m \} \]

\( \mathcal{P} \) is convex:

\[ x, y \in \mathcal{P}, \ 0 \leq \lambda \leq 1 \implies \lambda x + (1 - \lambda)y \in \mathcal{P} \]

i.e., the line segment between any two points in \( \mathcal{P} \) lies in \( \mathcal{P} \)
**Extreme Points and Vertices**

$x \in \mathcal{P}$ is an **extreme point** if it cannot be written as

$$x = \lambda y + (1 - \lambda)z$$

with $0 \leq \lambda \leq 1$, $y, z \in \mathcal{P}$, $y \neq x$, $z \neq x$

$x \in \mathcal{P}$ is a **vertex** if there is a $c$ such that $c^T x < c^T y$ for all $y \in \mathcal{P}$, $y \neq x$

**Fact:** $x$ is an extreme point $\iff$ $x$ is a vertex (proof later)
Basic Feasible Solutions

define $I$ as the set of indices of the active or binding constraints (at $x^*$):

$$a_i^T x^* = b_i, \quad i \in I, \quad a_i^T x^* < b_i, \quad i \notin I$$

define $\bar{A}$ as

$$\bar{A} = \begin{bmatrix} a_{i_1}^T \\ a_{i_2}^T \\ \vdots \\ a_{i_k}^T \end{bmatrix}, \quad I = \{i_1, \ldots, i_k\}$$

$x^*$ is called a basic feasible solution if

$$\text{rank } \bar{A} = n$$

**fact:** $x^*$ is a vertex (extreme point) $\iff x^*$ is a basic feasible solution
Example

\[
\begin{bmatrix}
-1 & 0 \\
2 & 1 \\
0 & -1 \\
1 & 2 \\
\end{bmatrix}
\begin{bmatrix}
x \\
\end{bmatrix}
\leq
\begin{bmatrix}
0 \\
3 \\
0 \\
3 \\
\end{bmatrix}
\]

• (1,1) is an extreme point

• (1,1) is a vertex: unique minimum of \( c^T x \) with \( c = (-1, -1) \)

• (1,1) is a basic feasible solution: \( I = \{2, 4\} \) and \( \text{rank } \overline{A} = 2 \), where

\[
\overline{A} = \begin{bmatrix}
2 & 1 \\
1 & 2 \\
\end{bmatrix}
\]
Convex Functions

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if for $0 \leq \lambda \leq 1$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$
Convex Sets

A set $C$ is a convex set if for any two elements $x, y \in C$ and any $0 \leq \theta \leq 1$, $\theta x + (1 - \theta)y \in C$. 
Closed and Open Sets

• We say that \( x \) is a **closure point** of a subset \( X \) of \( \mathbb{R}^n \) if there exists a sequence \( \{x_k\} \subset X \) that converges to \( x \). The closure of \( X \), denoted \( \text{cl}(X) \), is the set of all closure points of \( X \).

• **Definition of Closed Set**: A subset \( X \) of \( \mathbb{R}^n \) is called **closed** if it is equal to its closure. It is called **open** if its complement, \( \{x|x \notin X\} \), is closed. It is called **bounded** if there exists a scalar \( c \) such that \( \|x\| \leq c \) for all \( x \in X \). It is called **compact** if it is closed and bounded.

• For any \( \epsilon > 0 \) and \( x^* \in \mathbb{R}^n \), consider the sets
  \[
  \{x|\|x - x^*\| < \epsilon\}, \quad \{x|\|x - x^*\| \leq \epsilon\}.
  \]
  The first set is open and is called an open ball centered at \( x^* \), while the second set is closed and is called a closed ball centered at \( x^* \).

• A subset \( X \) of \( \mathbb{R}^n \) is open if and only if for every \( x \in X \) there is an open ball that is centered at \( x \) and is contained in \( X \). A **neighborhood** of a vector \( x \) is an open set containing \( x \).
Interior Point and Boundary Point

• We say that $x$ is an interior point of a subset $X$ of $\mathbb{R}^n$ if there exists a neighborhood of $x$ that is contained in $X$. The set of all interior points of $X$ is called the interior of $X$, and is denoted by $\text{int}(X)$.

• A vector $x \in \text{cl}(X)$ which is not an interior point of $X$ is said to be a boundary point of $X$. The set of all boundary points of $X$ is called the boundary of $X$.

• A set is open if and only if all of its elements are interior points.
Supremum and Infimum of Sets

• The **supremum** of a nonempty set $X$ of scalars, denoted by $\sup X$, is defined as the smallest scalar $y$ such that $y \geq x$ for all $x \in X$. If no such scalar exists, we say that the supremum of $X$ is $\infty$.
  • If $\sup X$ is equal to a scalar $\bar{x}$ that belongs to the set, we say that $\bar{x}$ is the maximum point of $X$ and we write $\bar{x} = \max X$.

• Similarly, the **infimum** of $X$, denoted by $\inf X$, is defined as the largest scalar $y$ such that $y \leq x$ for all $x \in X$, and is equal to $-\infty$ if no such scalar exists.
  • Similarly, if $\inf X$ is equal to a scalar $\bar{x}$ that belongs to the set $X$, we say that $\bar{x}$ is the minimum point of $X$ and we write $\bar{x} = \min X$.

• Thus, when we write $\max X$ (or $\min X$) in place of $\sup X$ (or $\inf X$, respectively), we do so just for emphasis: we indicate that it is either evident, or it is known through earlier analysis, or it is about to be shown that the maximum (or minimum, respectively) of the set $X$ is attained at one of its points.
**Continuity**

- Let $X$ be a subset of $\mathbb{R}^n$

- A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called **continuous** at a vector $x \in X$ if
  $$\lim_{z \to x} f(z) = f(x)$$

- A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is called **right-continuous** (respectively, **left-continuous**) at a vector $x \in X$ if
  $$\lim_{z \downarrow x} f(z) = f(x) \quad (\text{respectively, } \lim_{z \uparrow x} f(z) = f(x))$$

- A real-valued function $f : X \to \mathbb{R}$ is called **upper** semicontinuous (respectively, **lower** semicontinuous) at a vector $x \in X$ if
  $$f(x) \geq \limsup_{k \to \infty} f(x_k) \quad (\text{respectively, } f(x) \leq \liminf_{k \to \infty} f(x_k) \text{ for every sequence } \{x_k\} \subset X \text{ that converges to } x.)$$