15.093 Optimization Methods

Lecture 17: Applications of Nonlinear Optimization
1   Lecture Outline

- History of Nonlinear Optimization
- Where do NLPs Arise?
- Portfolio Optimization
- Traffic Assignment
- The general problem
- The role of convexity
- Convex optimization
- Examples of convex optimization problems

2   History of Optimization

Fermat, 1638; Newton, 1670

$$\min f(x) \quad x: \text{scalar}$$

$$\frac{df(x)}{dx} = 0$$

Euler, 1755

$$\min f(x_1, \ldots, x_n)$$

$$\nabla f(x) = 0$$

Lagrange, 1797

$$\min f(x_1, \ldots, x_n)$$

s.t. $$g_k(x_1, \ldots, x_n) = 0 \quad k = 1, \ldots, m$$

Euler, Lagrange Problems in infinite dimensions, calculus of variations.

Kuhn and Tucker, 1950s Optimality conditions.

1950s Applications.

1960s Large Scale Optimization.

Karmakar, 1984 Interior point algorithms.
3 Where do NLPs Arise?

3.1 Wide Applicability

- Transportation
  - Traffic management, Traffic equilibrium...
  - Revenue management and Pricing
- Finance - Portfolio Management
- Equilibrium Problems
- Engineering
  - Data Networks and Routing
  - Pattern Classification
- Manufacturing
  - Resource Allocation
  - Production Planning

4 A Simple Portfolio Selection Problem

4.1 Data

- $x_i$: decision variable on amount to invest in stock $i = 1, 2$
- $r_i$: reward from stock $i = 1, 2$ (random variable)

Data:

- $\mu_i = E(r_i)$: expected reward from stock $i = 1, 2$
- $Var(r_i)$: variance in reward from stock $i = 1, 2$
- $\sigma_{ij} = E[(r_j - \mu_j)(r_i - \mu_i)] = Cov(r_i, r_j)$

- Budget $B$, target $\beta$ on expected portfolio reward
5 A Simple Portfolio Selection Problem

5.1 The Problem

Objective: Minimize total portfolio variance so that:
- Expected reward of total portfolio is above target $\beta$
- Total amount invested stay within our budget
- No short sales

\[
\begin{align*}
\min f(x) &= x_1^2 \text{Var}(r_1) + x_2^2 \text{Var}(r_2) + 2x_1 x_2 \sigma_{12} \\
\text{subject to} & \sum_i x_i \leq B \\
E[\sum_i r_i x_i] &= \sum_i \mu_i x_i \geq \beta, \quad (\text{exp reward of portf.}) \\
x_i \geq 0, & \quad i = 1, 2
\end{align*}
\]

(Linearly constrained NLP)

6 A Real Portfolio Optimization Problem

6.1 Data

- We currently own $z_i$ shares from stock $i, i \in S$
- $P_i$: current price of stock $i$
- We consider buying and selling stocks in $S$, and consider buying new stocks from a set $B$ ($B \cap S = \emptyset$)
- Set of stocks $B \cup S = \{1, \ldots, n\}$

- Data: Forecasted prices next period (say next month) and their correlations:

\[
\begin{align*}
E[\hat{P}_i] &= \mu_i \\
\text{Cov}(\hat{P}_i, \hat{P}_j) &= E[(\hat{P}_i - \mu_i)(\hat{P}_j - \mu_j)] = \sigma_{ij} \\
\mu &= (\mu_1, \ldots, \mu_n)', \quad \Sigma = \sigma_{ij}
\end{align*}
\]
6.2 Issues and Objectives
- Mutual funds regulations: we cannot sell a stock if we do not own it
- Transaction costs
- Turnover
- Liquidity
- Volatility
- **Objective**: Maximize expected wealth next period minus transaction costs

6.3 Decision variables
\[ x_i = \begin{cases} 
\# \text{ shares bought or sold} & \text{if } i \in S \\
\# \text{ shares bought} & \text{if } i \in B 
\end{cases} \]

By convention:
\[ x_i \geq 0 \quad \text{buy} \]
\[ x_i < 0 \quad \text{sell} \]

6.4 Transaction costs
- Small investors only pay commision cost: \( a_i \) $/share traded
- Transaction cost: \( a_i |x_i| \)
- Large investors (like portfolio managers of large funds) may affect price: price becomes \( P_i + b_i x_i \)
- Price impact cost: \( (P_i + b_i x_i) x_i - P_i x_i = b_i x_i^2 \)
- Total cost model:
\[ c_i(x_i) = a_i |x_i| + b_i x_i^2 \]

6.5 Liquidity
- Suppose you own 50% of all outstanding stock of a company
- How difficult is to sell it?
- Reasonable to bound the percentage of ownership on a particular stock
- Thus, for liquidity reasons \( \frac{z_i + x_i}{z_i^{out}} \leq \gamma_i \)
- \( z_i^{out} \) = # outstanding shares of stock \( i \)
- \( \gamma_i \) maximum allowable percentage of ownership
6.6 Turnover

- Because of transaction costs: $|x_i|$ should be small
  $$|x_i| \leq \delta_i \Rightarrow -\delta_i \leq x_i \leq \delta_i$$
- Alternatively, we might want to bound turnover:
  $$\sum_{i=1}^{n} p_i |x_i| \leq t$$

6.7 Balanced portfolios

- Need the value of stocks we buy and sell to balance out:
  $$\left| \sum_{i=1}^{n} p_i x_i \right| \leq L \Rightarrow -L \leq \sum_{i=1}^{n} p_i x_i \leq L$$
- No short sales:
  $$z_i + x_i \geq 0, \quad i \in B \cup S$$

6.8 Expected value and Volatility

- Expected value of portfolio:
  $$E \left[ \sum_{i=1}^{n} \hat{p}_i (z_i + x_i) \right] = \sum_{i=1}^{n} \mu_i (z_i + x_i)$$
- Variance of the value of the portfolio:
  $$Var \left[ \sum_{i=1}^{n} \hat{p}_i (z_i + x_i) \right] = (z + x)' \Sigma (z + x)$$

6.9 Overall formulation

$$\begin{align*}
\max & \quad \sum_{i=1}^{n} \mu_i(z_i + x_i) - \sum_{i=1}^{n} (a_i |x_i| + b_i x_i^2) \\
\text{s.t.} & \quad (z + x)' \Sigma (z + x) \leq \sigma^2 \\
& \quad z_i + x_i \leq \gamma_i z_i^{\text{total}} \\
& \quad -\delta_i \leq x_i \leq \delta_i \\
& \quad -L \leq \sum_{i=1}^{n} p_i x_i \leq L \\
& \quad \sum_{i=1}^{n} p_i |x_i| \leq t \\
& \quad z_i + x_i \geq 0
\end{align*}$$
7 The general problem

\[ f(x): \mathbb{R}^n \rightarrow \mathbb{R} \]

\[ g_i(x): \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \ldots, m \]

\[
\begin{align*}
NLP: & \quad \min \quad f(x) \\
\text{s.t.} & \quad g_1(x) \leq 0 \\
& \quad \vdots \\
& \quad g_m(x) \leq 0
\end{align*}
\]

7.1 Is Portfolio Optimization an NLP?

\[
\begin{align*}
\max & \quad \sum_{i=1}^{n} \mu_i(z_i + x_i) - \sum_{i=1}^{n} (a_i|x_i| + b_ix_i^2) \\
\text{s.t.} & \quad (z + x)^T \Sigma (z + x) \leq \sigma^2 \\
& \quad z_i + x_i \leq \gamma_i x_i^{\text{total}} \\
& \quad -\delta_i \leq x_i \leq \delta_i \\
& \quad -L \leq \sum_{i=1}^{n} P_i x_i \leq L \\
& \quad \sum_{i=1}^{n} P_i|x_i| \leq t \\
& \quad z_i + x_i \geq 0
\end{align*}
\]

8 Geometry Problems

8.1 Fermat-Weber Problem

Given \( m \) points \( c_1, \ldots, c_m \in \mathbb{R}^n \) (locations of retail outlets) and weights \( w_1, \ldots, w_m \in \mathbb{R} \). Choose the location of a distribution center.

That is, the point \( x \in \mathbb{R}^n \) to minimize the sum of the weighted distances from \( x \) to each of the points \( c_1, \ldots, c_m \in \mathbb{R}^n \) (\textbf{minimize total daily distance traveled}).

\[
\min \sum_{i=1}^{m} w_i \|x - c_i\| \\
\text{s.t.} \quad x \in \mathbb{R}^n
\]

or
\[
\min \sum_{i=1}^{m} w_i \|x - c_i\| \\
\text{s.t.} \quad x \geq 0 \\
Ax \leq b, \text{ feasible sites}
\]

(Linearly constrained NLP)

8.2 The Ball Circumscription Problem

Given \( m \) points \( c_1, \ldots, c_m \in \mathbb{R}^n \), locate a distribution center at point \( x \in \mathbb{R}^n \) to minimize the maximum distance from \( x \) to any of the points \( c_1, \ldots, c_m \in \mathbb{R}^n \).

\[
\min \delta \\
\text{s.t.} \quad \|x - c_i\| \leq \delta, \quad i = 1, \ldots, m
\]

9 Transportation

9.1 Traffic Assignment

- OD \( w \), paths \( p \in P_w \), demand \( d_w \), \( x_p \): flow of \( p \)
- \( c_{ij}(\sum_{p} \text{crossing} (i,j) x_p) \): travel cost of link \((i,j)\).
- \( c_p(x) \) is the travel cost of path \( p \) and

\[
c_p(x) = \sum_{(i,j) \text{ on } p} c_{ij}(x_{ij}), \quad \forall p \in P_w, \quad \forall w \in W.
\]

System optimization principle: Assign flow on each path to satisfy total demand and so that the total network cost is minimized.

\[
\text{Min } C(x) = \sum_{p} c_p(x) x_p \\
\text{s.t. } x_p \geq 0, \quad \sum_{p \in P_w} x_p = d_w, \quad \forall w
\]

9.2 Example

Consider a three path network, \( d_w = 10 \).

With travel costs \( c_{p_1}(x) = 2x_{p_1} + x_{p_2} + 15, c_{p_2}(x) = 3x_{p_2} + x_{p_1} + 11, c_{p_3}(x) = x_{p_3} + 48 \)

\[
C(x) = c_{p_1}(x) x_{p_1} + c_{p_2}(x) x_{p_2} + c_{p_3}(x) x_{p_3} = 2x_{p_1}^2 + 3x_{p_2}^2 + x_{p_3}^2 + 2x_{p_1} x_{p_2} + 15x_{p_1} + 11x_{p_2} + 38x_{p_3} \\
x_{p_1}^* = 6, \quad x_{p_2}^* = 4, \quad x_{p_3}^* = 0
\]
• **User – optimization principle**: Each user of the network chooses, among all paths, a path requiring minimum travel cost, i.e., for all \( w \in W \) and \( p \in P_w \),

\[
x^*_p > 0 : \quad \Rightarrow \quad c_p(x^*) \leq c_{p'}(x^*) \quad \forall p' \in P_w, \quad \forall w \in W
\]

where \( c_p(x) \) is the travel time of path \( p \) and

\[
c_p(x) = \sum_{(i,j) \in p} c_{ij}(x_{ij}), \quad \forall p \in P_w, \quad \forall w \in W
\]

10 **Optimal Routing**

• Given a data net and a set \( W \) of OD pairs \( w = (i, j) \) each OD pair \( w \) has input traffic \( d_w \)

• Optimal routing problem:

\[
\begin{align*}
\text{Min} & \quad C(x) = \sum_{i,j} C_{i,j} \left( \sum_{y: (i,j) \in p} x_p \right) \\
\text{s.t.} & \quad \sum_{p \in P_w} x_p = d_w, \quad \forall w \in W \\
                & \quad x_p \geq 0, \quad \forall p \in P_w, \quad w \in W
\end{align*}
\]

11 **The general problem again**

\[
f(x) : \mathbb{R}^n \rightarrow \mathbb{R}
\]

is a continuous (usually differentiable) function of \( n \) variables

\[
g_i(x) : \mathbb{R}^n \rightarrow \mathbb{R}, \quad i = 1, \ldots, m,
\]

\[
h_j(x) : \mathbb{R}^n \rightarrow \mathbb{R}, \quad j = 1, \ldots, l
\]

\[
\begin{align*}
\text{NLP:} & \quad \min & f(x) \\
\text{s.t.} & & g_1(x) \leq 0 \\
& & \vdots \\
& & g_m(x) \leq 0 \\
& & h_1(x) = 0 \\
& & \vdots \\
& & h_l(x) = 0
\end{align*}
\]
11.1 Definitions

- The feasible region of $NLOP$ is the set:

$$\mathcal{F} = \{ x | g_1(x) \leq 0, \ldots, g_m(x) \leq 0 \}$$

$$h_1(x) = 0, \ldots, h_n(x) = 0$$

11.2 Where do optimal solutions lie?

Example:

$$\min f(x, y) = (x - a)^2 + (y - b)^2$$

Subject to

$$(x - 8)^2 + (y - 9)^2 \leq 49$$

$$2 \leq x \leq 13$$

$$x + y \leq 24$$

Optimal solution(s) do not necessarily lie at an extreme point!

Depends on $(a, b)$.

- $(a, b) = (16, 14)$ then solution lies at a corner
- $(a, b) = (11, 10)$ then solution lies in interior
- $(a, b) = (14, 14)$ then solution lies on the boundary
  (not necessarily corner)

11.3 Local vs Global Minima

- The ball centered at $\bar{x}$ with radius $\epsilon$ is the set:

$$B(\bar{x}, \epsilon) := \{ x ||x - \bar{x}|| \leq \epsilon \}$$

- $x \in \mathcal{F}$ is a local minimum of $NLOP$ if there exists $\epsilon > 0$ such that $f(x) \leq f(y)$ for all $y \in B(x, \epsilon) \cap \mathcal{F}$

- $x \in \mathcal{F}$ is a global minimum of $NLOP$ if $f(x) \leq f(y)$ for all $y \in \mathcal{F}$

12 Convex Sets

- A subset $S \subseteq \mathbb{R}^n$ is a convex set if

$$x, y \in S \Rightarrow \lambda x + (1 - \lambda) y \in S \quad \forall \lambda \in [0, 1]$$

- If $S, T$ are convex sets, then $S \cap T$ is a convex set

- Implication: The intersection of any collection of convex sets is a convex set
13 Convex Functions

- A function \( f(x) \) is a convex function if
  \[
  f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y)
  \]
  \( \forall x, y \quad \forall \lambda \in [0, 1] \)

- A function \( f(x) \) is a concave function if
  \[
  f(\lambda x + (1 - \lambda) y) \geq \lambda f(x) + (1 - \lambda) f(y)
  \]
  \( \forall x, y \quad \forall \lambda \in [0, 1] \)

13.1 Examples in one dimension

- \( f(x) = ax + b \)
- \( f(x) = x^2 + bx + c \)
- \( f(x) = |x| \)
- \( f(x) = -\ln(x) \) for \( x > 0 \)
- \( f(x) = \frac{1}{x} \) for \( x > 0 \)
- \( f(x) = e^x \)

13.2 Properties

- If \( f_1(x) \) and \( f_2(x) \) are convex functions, and \( a, b \geq 0 \), then \( f(x) := af_1(x) + bf_2(x) \) is a convex function
- If \( f(x) \) is a convex function and \( x = Ay + b \), then \( g(y) := f(Ay + b) \) is a convex function

13.3 Recognition of a Convex Function

A function \( f(x) \) is twice differentiable at \( \tilde{x} \) if there exists a vector \( \nabla f(\tilde{x}) \) (called the gradient of \( f(\cdot) \)) and a symmetric matrix \( H(\tilde{x}) \) (called the Hessian of \( f(\cdot) \)) for which:

\[
    f(x) = f(\tilde{x}) + \nabla f(\tilde{x})'(x - \tilde{x})
    + \frac{1}{2}(x - \tilde{x})'H(\tilde{x})(x - \tilde{x}) + R(x)\|x - \tilde{x}\|^2
\]

where \( R(x) \to 0 \) as \( x \to \tilde{x} \)

The gradient vector is the vector of partial derivatives:

\[
    \nabla f(\tilde{x}) = \left(\frac{\partial f(\tilde{x})}{\partial x_1}, \ldots, \frac{\partial f(\tilde{x})}{\partial x_n}\right)
\]
The Hessian matrix is the matrix of second partial derivatives:

\[ H(\mathbf{x})_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \]

### 13.4 Examples

- For LP, \( f(x) = c'x, \nabla f(\mathbf{x}) = c \)
- For NLP,
  \[ f(x) = 8x_1^2 - x_1x_2 + x_2^2 + 8x_1, \text{ at } \mathbf{x} = (1, 0), \]
  \[ f(\mathbf{x}) = 16 \text{ and } \nabla f(\mathbf{x})' = (16x_1 - x_2 + 8, -x_1 + 2x_2) = (24, -1). \]

Property: \( f(x) \) is a convex function if and only if \( H(x) \) is positive semi-definite (PSD) for all \( x \)

Recall that \( A \) is PSD if \( u'Au \geq 0, \forall u \)

Property: If \( H(x) \) is PD for all \( x \), then \( f(x) \) is a strictly convex function

### 13.5 Examples in n Dimensions

- \( f(x) = a'x + b \)
- \( f(x) = \frac{1}{2}x'Mx - c'x \) where \( M \) is PSD
- \( f(x) = \|x\| \text{ for any norm } \|\cdot\| \)
- \( f(x) = \sum_{i=1}^{m} -\ln(b_i - a_i'x) \) for \( x \) satisfying \( Ax < b \)

### 14 Convex Optimization

#### 14.1 Convexity and Minima

\[ \min_{x \in \mathcal{F}} f(x) \]

**Theorem:** Suppose that \( \mathcal{F} \) is a convex set, \( f : \mathcal{F} \to \mathbb{R} \) is a convex function, and \( x^* \) is a local minimum of \( P \). Then \( x^* \) is a global minimum of \( f \) over \( \mathcal{F} \).

**Proof:** Assume that \( x^* \) is not the global minimum. Let \( y \) be the global minimum. Let \( \lambda \) be a convex combination of \( x^* \) and \( y \),

\[ f(y(\lambda)) = f(\lambda x^* + (1 - \lambda)y) \leq \lambda f(x^*) + (1 - \lambda) f(y) \]

\[ < \lambda f(x^*) + (1 - \lambda) f(y) = f(x^*) \]

11
for all \( \lambda \in (0,1) \).

Therefore, \( f(y(\lambda)) < f(x^*) \) for all \( \lambda \in (0,1) \), and so \( x^* \) is not a local minimum, resulting in a contradiction.

### 14.2 COP

\[
COP: \quad \min \ f(x) \\
\text{s.t.} \quad g_1(x) \leq 0 \\
\quad \vdots \\
\quad g_m(x) \leq 0 \\
\quad Ax = b
\]

\( COP \) is called a \textit{convex optimization problem} if \( f(x), g_1(x), \ldots, g_m(x) \) are convex functions.

Note that this implies that the feasible region \( \mathcal{F} \) is a convex set.

In \( COP \) we are minimizing a convex function over a convex set.

Implication: If \( COP \) is a convex optimization problem, then any local minimum will be a global minimum.

### 15 Examples of COPs

The Fermat-Weber Problem - \( COP \)

\[
\min \sum_{i=1}^{m} w_i ||x - c_i|| \\
\text{s.t.} \quad x \in P
\]

The Ball Circumscription Problem - \( COP \)

\[
\min \ \delta \\
\text{s.t.} \quad ||x - c_i|| \leq \delta, \quad i = 1, \ldots, m
\]
15.1 Is Portfolio Optimization a COP?

\[
\begin{align*}
\max & \quad \sum_{i=1}^{n} \mu_i(z_i + x_i) - \sum_{i=1}^{n} (a_i|x_i| + b_i x_i^2) \\
\text{s.t.} & \quad (z + x)' \Sigma (z + x) \leq \sigma^2 \\
& \quad z_i + x_i \leq \gamma_i x_i^{(\text{total})} \\
& \quad -\delta_i \leq x_i \leq \delta_i \\
& \quad -L \leq \sum_{i=1}^{n} P_i x_i \leq L \\
& \quad \sum_{i=1}^{n} P_i |x_i| \leq \ell \\
& \quad z_i + x_i \geq 0
\end{align*}
\]

15.2 Quadratically Constrained Problems

\[
\begin{align*}
\min & \quad (A_0 x + b_0)'(A_0 x + b_0) - c_0' x - d_0 \\
\text{s.t.} & \quad (A_i x + b_i)'(A_i x + b_i) - c_i' x - d_i \leq 0 \\
& \quad i = 1, \ldots, m
\end{align*}
\]

This is a COP

16 Classification of NLPs

- **Linear**: \( f(x) = c' x, \quad g_i(x) = A_i' x - b_i, \quad i = 1, \ldots, m \)
- **Unconstrained**: \( f(x), \quad \mathbb{R}^n \)
- **Quadratic**: \( f(x) = c' x + x' Q x, \quad g_i(x) = A_i' x - b_i \)
- **Linearly Constrained**: \( g_i(x) = A_i' x - b_i \)
- **Quadratically Constrained**: \( g_i(x) = (A_i x + b_i)'(A_i x + b_i) - c_i' x - d_i \leq 0, \quad i = 1, \ldots, m \)
- **Separable**: \( f(x) = \sum_j f_j(x_j), \quad g_i(x) = \sum_j g_{ij}(x_j) \)

17 Two Main Issues

- Characterization of minima
  - Necessary — Sufficient Conditions
  - Lagrange Multiplier and KKT Theory
• Computation of minima via iterative algorithms
  Iterative descent Methods
  Interior Point Methods

18 Summary

• Convex optimization is a powerful modeling framework
• Main message: convex optimization can be solved efficiently