Introduction

Lecture 1, Nonlinear Programming, (Part a)

National Taiwan University

September 13, 2016

Table of contents

Mathematical Optimization Problems

Mathematical Optimization Problems

A **mathematical optimization problem**, or just **optimization problem**, has the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

- $x = [x_1, \ldots, x_n]^T \in \mathbf{R}^n$: **optimization variable** of the problem
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: **objective function**
- $f_i : \mathbf{R}^n \to \mathbf{R}$: **constraint functions**

Mathematical Optimization Problems

Mathematical Optimization Problems

A **mathematical optimization problem**, or just **optimization problem**, has the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

- $x = [x_1, \ldots, x_n]^T \in \mathbf{R}^n$: **optimization variable** of the problem
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: **objective function**
- $f_i : \mathbf{R}^n \to \mathbf{R}$: **constraint functions**

Optimal Solutions

A vector x^* is called **optimal**, or a **solution** of the problem, if it has the smallest objective value among all vectors that satisfy the constraints: for any z with $f_1(z) \leq b_1, \ldots, f_m(z) \leq b_m$, we have

$$f_0(z) \geq f_0(x^*).$$

Linear Programming

Linear Programming

An optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \dots, m, \end{array}$$

is called a **linear program** if the objective and constraint functions $f_0, f_1, \dots, f_m$ are linear:

$$f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y)$$

for all $x, y \in \mathbf{R}^n$ and all $\alpha, \beta \in \mathbf{R}$.

Linear Programming & Nonlinear Programming

Linear Programming

An optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

is called a **linear program** if the objective and constraint functions $f_0, f_1, \ldots, f_m$ are linear:

$$f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y)$$

for all $x, y \in \mathbf{R}^n$ and all $\alpha, \beta \in \mathbf{R}$.

Nonlinear programming

An optimization problem that is not linear is called a **nonlinear program**.

Convex Optimization Problems

Convex Optimization Problems

An optimization problem is called a **convex optimization problem** if the objective and constraint functions $f_0, f_1, \ldots, f_m$ are convex:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$

for all $x, y \in \mathbf{R}^n$ and all $\alpha, \beta \in \mathbf{R}$, with $\alpha, \beta \geq 0, \alpha + \beta = 1$.

- Convex optimization problems are more general than linear programs.
- There are very effective algorithms that can reliably and efficiently solve even large convex problems.

Applications

- A great variety of practical problems can be cast in the form of a mathematical optimization problem.
- It is widely used in engineering, in electronic design automation, automatic control systems, and optimal design problems arising in civil, chemical, mechanical, and aerospace engineering.
- Optimization is used for problems arising in network design and operation, finance, supply chain management, scheduling, and many other areas.
- The list of applications is still steadily expanding.

Linear Programming and Least-Squares Problems

- We briefly describe two commonly used problems, namely, least-squares problems and linear programming.
- Least-squares: quadratic objective function; no constraints (nonlinear).
- Linear programming: linear objective functions; linear constraint functions.
- Both special cases of convex optimization problems.

Least-Squares Problems

Least-Squares Problems

A least-squares problem is an optimization problem with no constraints (i.e., $m = 0$) and an objective which is a sum of squares of terms of the form $a_i^T x - b_i$:

$$\text{minimize } f_0(x) = ||Ax - b||_2^2 = \sum_{i=1}^{k} (a_i^T x - b_i)^2,$$

where $A \in \mathbf{R}^{k \times n}$ (with $k \geq n$), a_i^T are the rows of A, $b_i \in \mathbf{R}$, and the vector $x \in \mathbf{R}^n$ is the optimization variable.

Solving Least-Squares Problems

- The solution of a least-squares problem

$$\text{minimize } f_0(x) = ||Ax - b||_2^2 = \sum_{i=1}^{k} (a_i^T x - b_i)^2$$

can be reduced to solving a set of linear equations,

$$(A^T A)x = A^T b.$$

so we have the analytical solution $x = (A^T A)^{-1} A^T b$.

- The least-squares problem can be solved in a time approximately proportional to $n^2 k$, with a known constant.

Using Least-Squares

- Least-squares has many statistical interpretations, *e.g.*, as maximum likelihood estimation of a vector x, given linear measurements corrupted by Gaussian measurement errors (e.g., $b = Ax + n$).
- To recognize a problem as a least-squares problem, we need to verify that
    - the objective is a quadratic function;
    - whether the associated quadratic form is positive semidefinite.
- Examples:
    - **weighted least-squares**: minimize $\sum_{i=1}^{k} w_i(a_i^T x - b_i)^2$, where $w_1, \ldots, w_k$ are positive.
    - **regularization**: minimize $\sum_{i=1}^{k}(a_i^T x - b_i)^2 + \rho \sum_{i=1}^{n} x_i^2$ where $\rho > 0$.

Linear Programming

Linear Programming

A linear programming has the following form:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

where the vectors $c, a_1, \cdots, a_m \in \mathbf{R}^n$ and scalars $b_1, \cdots, b_m \in \mathbf{R}$.

Solving Linear Programs

- No simple analytical formula for the solution of a linear problem.
- Very effective methods exist for solving them
    - Simplex method (Dantzig 1947).
    - Interior-point method.
- Complexity
    - Simplex method: usually efficient (average case polynomial time, but worst-case exponential time)
    - Interior-point method: in the order of $n^2 m$.
- Considered a mature technology,
    - although it's still challenging to solve extremely large problems, or with real-time computing requirements.

Using Linear Programming

Example – Chebyshev approximation problem

Consider the **Chebyshev approximation problem**:

$$\text{minimize} \qquad \max_{i=1,\ldots,k} |a_i^T x - b_i|,$$

where $x \in \mathbf{R}^n$ is the variable, and $a_1, \ldots, a_k \in \mathbf{R}^n$, $b_1, \ldots, b_k \in \mathbf{R}$.

The objective can be rewritten as

$$\max_{i=1,\ldots,k} |a_i^T x - b_i| = ||Ax - b||_\infty,$$

where $A \in \mathbf{R}^{k \times n}$ whose ith row is a_i^T and $b = \begin{bmatrix} b_1, \ldots, b_k \end{bmatrix}^T$.

Using Linear Programming

Example – Chebyshev approximation problem

Consider the **Chebyshev approximation problem**:

$$\text{minimize} \quad \max_{i=1,\dots,k} |a_i^T x - b_i|,$$

where $x \in \mathbf{R}^n$ is the variable, and $a_1, \dots, a_k \in \mathbf{R}^n$, $b_1, \dots, b_k \in \mathbf{R}$.

The Chebyshev approximation problem can be solved by solving the linear program

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x - t \leq b_i, i = 1, \dots, k \\ & -a_i^T x - t \leq -b_i, i = 1, \dots, k \end{array}$$

with variables $x \in \mathbf{R}^n$ and $t \in \mathbf{R}$.

Convex Optimization Problems

Convex Optimization Problems

A **convex optimization problem** has the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

where the objective and constraint functions $f_0, f_1, \ldots, f_m$ are convex:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$

for all $x, y \in \mathbf{R}^n$ and all $\alpha, \beta \in \mathbf{R}$, with $\alpha, \beta \geq 0, \alpha + \beta = 1$.

- Least-squares and linear programming are both special cases of convex optimization problems.

Solving Convex Optimization Problems

- No analytical formula for the solution of convex optimization problems in general.
- There are very effective methods for solving convex optimization problems.
    - E.g., interior-point methods.

Nonlinear Programming

- Nonlinear optimization (or nonlinear programming) is the term used to describe an optimization problem when the objective or constraint functions are not linear, but not known to be convex.
- No effective methods for solving the general nonlinear programming problem yet.
- Methods for the general nonlinear programming problem therefore take several different approaches, each of which involves some compromise.

Outline of the course

- Theory
  - Convex sets.
  - Convex functions.
  - Convex optimization problems.
  - Duality.
- Algorithms
  - Unconstrained minimization.
  - Equality constrained minimization.
  - Interior-point methods.
- Applications (when time allows).
  - Approximation and fitting.
  - Statistical estimation.
  - Geometric problems.

Notations

- $\mathbf{R}$: the set of real numbers.
- $\mathbf{R}_+ = \{x \in \mathbf{R} \mid x \geq 0\}$: the set of nonnegative real numbers.
- $\mathbf{R}_{++} = \{x \in \mathbf{R} \mid x > 0\}$: the set of positive numbers.
- $\mathbf{R}^n$: the set of real n-vectors.
- $\mathbf{R}^{m \times n}$: the set of real $m \times n$ matrices.
- $\mathbf{1}$: a vector whose components are all one.
- If $a, b, c \in \mathbf{R}$, then the following notations denote the same vector in $\mathbf{R}^3$.

$$(a, b, c) = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a & b & c \end{bmatrix}^T$$

Notations

- $\mathbf{S}^k$: the set of symmetric $k \times k$ matrices ($\mathbf{S}^k \subseteq \mathbf{R}^{k \times k}$).
- $\mathbf{S}^k_+$: the set of symmetric nonnegative definite (i.e., positive semidefinite) $k \times k$ matrices.
- $\mathbf{S}^k_{++}$: the set of symmetric positive definite $k \times k$ matrices.
- $f : \mathbf{R}^p \to \mathbf{R}^q$
  - denotes an $\mathbf{R}^q$-valued function on some subset of $\mathbf{R}^p$.
  - $\mathbf{dom}\ f$: denotes the domain of f (i.e., $\mathbf{dom}\ f \subseteq \mathbf{R}^p$).
- Example: $\log : \mathbf{R} \to \mathbf{R}$ with $\mathbf{dom}\ \log = \mathbf{R}_{++}$.
  - Note: The base of the logarithm used in this series of slides, if not explicitly indicated, is e instead of 10.