Math 408A: Non-Linear Optimization

Introduction

Professor James Burke

Math Dept, University of Washington

What is non-linear programming?

What is optimization?

A mathematical optimization problem is one in which a given function is either maximized or minimized relative to a given set of alternatives.

What is optimization?

A mathematical optimization problem is one in which a given function is either maximized or minimized relative to a given set of alternatives.

- The function to be minimized or maximized is called the *objective function*.

What is optimization?

A mathematical optimization problem is one in which a given function is either maximized or minimized relative to a given set of alternatives.

- The function to be minimized or maximized is called the *objective function*.
- The set of alternatives is called the constraint region (or feasible region).

What is optimization?

A mathematical optimization problem is one in which a given function is either maximized or minimized relative to a given set of alternatives.

- The function to be minimized or maximized is called the *objective function*.
- The set of alternatives is called the constraint region (or feasible region).
- In this course, the feasible region is always taken to be a subset of $\mathbb{R}^n$ (real n-dimensional space) and the objective function is a function from $\mathbb{R}^n$ to $\mathbb{R}$.

What is optimization?

$$\begin{array}{lll} \mathcal{P}: & \underset{x \in X}{\text{minimize}} & f_0(x) \\ & \text{subject to} & x \in \Omega. \end{array}$$

What is optimization?

$$\mathcal{P} : \quad \underset{x \in X}{\text{minimize}} \quad f_0(x) \\ \text{subject to} \quad x \in \Omega.$$

X is the variable space (or decision space),

What is optimization?

$$\mathcal{P}: \quad \begin{array}{ll} \underset{x \in X}{\text{minimize}} & f_0(x) \\ \text{subject to} & x \in \Omega. \end{array}$$

X is the variable space (or decision space),

$f_0 : X \to \mathbb{R} \cup \{\pm\infty\}$ is the objective function, and

What is optimization?

$$\begin{array}{lll} \mathcal{P}: & \underset{x \in X}{\text{minimize}} & f_0(x) \\ & \text{subject to} & x \in \Omega. \end{array}$$

X is the variable space (or decision space),

$f_0 : X \to \mathbb{R} \cup \{\pm\infty\}$ is the objective function, and

$\Omega \subset X$ is the constraint region.

What is optimization?

Problem Categories:

What is optimization?

Problem Categories:

- Variable Type:

What is optimization?

Problem Categories:

- Variable Type:
  - continuous variable: $X = \mathbb{R}^n$

What is optimization?

Problem Categories:

- Variable Type:
  - continuous variable: $X = \mathbb{R}^n$
  - discrete variable: $X = \mathbb{Z}^n$

What is optimization?

Problem Categories:

- Variable Type:
  - continuous variable: $X = \mathbb{R}^n$
  - discrete variable: $X = \mathbb{Z}^n$
  - mixed variable: $X = \mathbb{R}^s \times \mathbb{Z}^t$

What is optimization?

Problem Categories:

- Variable Type:
  - continuous variable: $X = \mathbb{R}^n$
  - discrete variable: $X = \mathbb{Z}^n$
  - mixed variable: $X = \mathbb{R}^s \times \mathbb{Z}^t$
- Constraint Type:

What is optimization?

Problem Categories:

- Variable Type:
  - continuous variable: $X = \mathbb{R}^n$
  - discrete variable: $X = \mathbb{Z}^n$
  - mixed variable: $X = \mathbb{R}^s \times \mathbb{Z}^t$
- Constraint Type:
  - unconstrained: $\Omega = X$

What is optimization?

Problem Categories:

- Variable Type:
  - continuous variable: $X = \mathbb{R}^n$
  - discrete variable: $X = \mathbb{Z}^n$
  - mixed variable: $X = \mathbb{R}^s \times \mathbb{Z}^t$
- Constraint Type:
  - unconstrained: $\Omega = X$
  - constrained: $\Omega \neq X$

Problem Types

Problem Types

Convex Programming:

Problem Types

Convex Programming: f_0 a convex function, Ω a convex set

Problem Types

Convex Programming: f_0 a convex function, Ω a convex set

Definition: $\Omega \subset \mathbb{R}^n$ is convex if for every $x, y \in \Omega$ one has $[x, y] \subset \Omega$ where $[x, y]$ is the line segment connecting x and y:

$$[x, y] = \{\lambda x + (1 - \lambda)y : 0 \leq \lambda \leq 1\}.$$

Problem Types

Convex Programming: f_0 a convex function, Ω a convex set

Definition: $\Omega \subset \mathbb{R}^n$ is convex if for every $x, y \in \Omega$ one has $[x, y] \subset \Omega$ where $[x, y]$ is the line segment connecting x and y:

$$[x, y] = \{\lambda x + (1 - \lambda)y : 0 \le \lambda \le 1\}.$$

Definition: $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ is convex if $epi(f) = \{(x, \mu) : f(x) \le \mu\}$ is a convex set in $\mathbb{R}^{n+1}$. In particular,

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y)$$

for all $0 \le \lambda \le 1$ and points x, y for which not both $f(x)$ and $f(y)$ are infinite.

Problem Types

Linear Programming:

The minimization or maximization of a linear functional subject to a finite number of linear inequality and/or equality constraints.

$f_0(x) := c^T x$ for some $c \in \mathbb{R}^n$ and

$$\Omega := \left\{ x : a_i^T x \begin{array}{ll} \le b_i & i = 1, \ldots, s \\ = b_i & i = s+1, \ldots, m \end{array} \right\}.$$

Problem Types

Linear Programming:
The minimization or maximization of a linear functional subject to a finite number of linear inequality and/or equality constraints.
$f_0(x) := c^T x$ for some $c \in \mathbb{R}^n$ and

$$\Omega := \left\{ x : a_i^T x \begin{array}{ll} \leq b_i & i = 1, \ldots, s \\ = b_i & i = s+1, \ldots, m \end{array} \right\}.$$

Linear programming is a special case of convex programming. In this case the constraint region Ω is called a polyhedral convex set. Polyhedra have a very special geometric structure.

Problem Types

Quadratic Programming:

Problem Types

Quadratic Programming:
The minimization or maximization of a quadratic objective functions over a convex polyhedron:

$$f_0(x) = \frac{1}{2}x^T Qx + b^T x + \alpha$$

Problem Types

Quadratic Programming:
The minimization or maximization of a quadratic objective functions over a convex polyhedron:

$$f_0(x) = \frac{1}{2}x^TQx + b^Tx + \alpha$$

Fact: f_0 is convex if and only if Q is positive semi-definite.

Problem Types

Mini-Max:

$$f_0(x) = \max\{f_i(x) : i = 1, \ldots, s\}.$$

Problem Types

Mini-Max:

$$f_0(x) = \max\{f_i(x) : \ i = 1, \ldots, s\}.$$

Nonlinear Programming:

$$\Omega := \{x \in X : f_i(x) \leq 0, i = 1, \ldots, s, f_i(x) = 0, i = j + 1, \ldots m\}$$

Problem Types

Mini-Max:

$$f_0(x) = \max\{f_i(x) : \; i = 1, \ldots, s\}.$$

Nonlinear Programming:

$$\Omega := \{x \in X : f_i(x) \leq 0, i = 1, \ldots, s, f_i(x) = 0, i = j + 1, \ldots m\}$$

- Differentiable: f_i is smooth $i = 0, \ldots, m$

Problem Types

Mini-Max:

$$f_0(x) = \max\{f_i(x) : i = 1, \ldots, s\}.$$

Nonlinear Programming:

$$\Omega := \{x \in X : f_i(x) \leq 0, i = 1, \ldots, s, f_i(x) = 0, i = j + 1, \ldots m\}$$

- Differentiable: f_i is smooth $i = 0, \ldots, m$
- Nonsmooth: at least one f_i is not smooth

Problem Types

Mini-Max:

$$f_0(x) = \max\{f_i(x) : i = 1, \dots, s\}.$$

Nonlinear Programming:

$$\Omega := \{x \in X : f_i(x) \leq 0, i = 1, \dots, s, f_i(x) = 0, i = j + 1, \dots m\}$$

- Differentiable: f_i is smooth $i = 0, \dots, m$
- Nonsmooth: at least one f_i is not smooth
- Semi-Infinite: $m = +\infty$.

Problem Types

Mini-Max:

$$f_0(x) = \max\{f_i(x) : i = 1, \ldots, s\}.$$

Nonlinear Programming:

$$\Omega := \{x \in X : f_i(x) \leq 0, i = 1, \ldots, s, f_i(x) = 0, i = j+1, \ldots m\}$$

- Differentiable: f_i is smooth $i = 0, \ldots, m$
- Nonsmooth: at least one f_i is not smooth
- Semi-Infinite: $m = +\infty$.

Box Constraints:

$$\begin{aligned} \Omega &:= \{x \in \mathbb{R}^n : l_i \leq x_i \leq u_i, i = 1, \ldots, n\} \\ l_i &\in \mathbb{R} \cup \{-\infty\},\ u_i \in \mathbb{R} \cup \{+\infty\},\ l_i \leq u_i \end{aligned}$$

Problem Types

Parameter Identification:

Given data points $\{(x_i, y_i)\}_{i=1}^m \subset \mathbb{R}^t \times \mathbb{R}^s$, find the function $f(x) := m(p, x)$ from a parametrized class of functions

$$\mathcal{M} := \{m(p, \cdot) \ |p \in \Gamma \subset \mathbb{R}^n\}$$

that "best" fits the data.

Parameter Identification

Polynomial Least-Squares:
The function class is the set of polynomials of degree n or less.

$$\mathcal{P}_n = \left\{ m(p,x) = p_0 + p_1 x + \cdots + p_n x^n \;\middle|\; p_i \in \mathbb{R},\ i = 0, 1, \ldots, n \right\}.$$

A "best" fit can be found by minimizing the sum of squares

$$f_0(p) = \sum_{i=1}^{n} (m(p,x_i) - y_i)^2$$

over all choices of $p \in \mathbb{R}^{n+1}$.

Polynomial Least-Squares

We would like to satisfy the following equations.

$$\begin{array}{rcl} p_0 + p_1 x_1 + p_2 x_1^2 + \cdots + p_n x_1^n & = & y_1 \\ p_0 + p_1 x_2 + p_2 x_2^2 + \cdots + p_n x_2^n & = & y_2 \\ \vdots & & \vdots \\ p_0 + p_1 x_m + p_2 x_m^2 + \cdots + p_n x_m^n & = & y_m \end{array}$$

Polynomial Least-Squares

We would like to satisfy the following equations.

$$\begin{array}{rcl} p_0 + p_1 x_1 + p_2 x_1^2 + \cdots + p_n x_1^n & = & y_1 \\ p_0 + p_1 x_2 + p_2 x_2^2 + \cdots + p_n x_2^n & = & y_2 \\ \vdots & & \vdots \\ p_0 + p_1 x_m + p_2 x_m^2 + \cdots + p_n x_m^n & = & y_m \end{array}$$

This is a matrix equation in the unknowns $p_0, \ldots, p_n$.

Polynomial Least-Squares

We would like to satisfy the following equations.

$$\begin{array}{rcl} p_0 + p_1 x_1 + p_2 x_1^2 + \cdots + p_n x_1^n & = & y_1 \\ p_0 + p_1 x_2 + p_2 x_2^2 + \cdots + p_n x_2^n & = & y_2 \\ \vdots & & \vdots \\ p_0 + p_1 x_m + p_2 x_m^2 + \cdots + p_n x_m^n & = & y_m \end{array}$$

This is a matrix equation in the unknowns $p_0, \ldots, p_n$.

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & & & \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Polynomial Least-Squares

Set

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & & & \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix}, \quad p = \begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_n \end{pmatrix}, \quad \text{and} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Then the matrix equation becomes $Xp = y$.

Polynomial Least-Squares

Set

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & & & \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix}, \quad p = \begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_n \end{pmatrix}, \quad \text{and} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Then the matrix equation becomes $Xp = y$.
Typically, $n << m$ so this equation has no solution.

Polynomial Least-Squares

Set

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & & & \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix}, \quad p = \begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_n \end{pmatrix}, \quad \text{and} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Then the matrix equation becomes $Xp = y$.
Typically, $n << m$ so this equation has no solution.
To get the "best" solution we solve

$$\min_{p \in \mathbb{R}^{n+1}} \frac{1}{2} \|Xp - y\|_2^2$$

Polynomial Least-Squares

Set

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & & & \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix}, \quad p = \begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_n \end{pmatrix}, \quad \text{and} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Then the matrix equation becomes $Xp = y$.
Typically, $n << m$ so this equation has no solution.
To get the "best" solution we solve

$$\min_{p \in \mathbb{R}^{n+1}} \frac{1}{2} \|Xp - y\|_2^2$$

This is an example of a *linear least squares* problem.

Linear Least-Squares

A linear least squares problem is any optimization problem of the form

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2,$$

for some $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

Linear Least-Squares

A linear least squares problem is any optimization problem of the form

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2,$$

for some $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

Linear least squares problems are convex quadratic programs:

$$\frac{1}{2}\|Ax-b\|_2^2 \;=\; \frac{1}{2}x^TA^TAx-(A^Tb)^Tx+\frac{1}{2}b^Tb.$$

Nonlinear Programming

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_j(x) \leq 0,\ j = 1, 2, \ldots, s \\ & f_j(x) = 0,\ j = s = 1, \ldots, m\ . \end{array}$$

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\overline{x} \in \Omega$ is said to be a global solution to the problem

$$\min\{f_0(x) : x \in \Omega\}$$

if $f_0(\overline{x}) \le f_0(x)$ for all $x \in \Omega$.

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\overline{x} \in \Omega$ is said to be a global solution to the problem
$$\min\{f_0(x) : x \in \Omega\}$$
if $f_0(\overline{x}) \leq f_0(x)$ for all $x \in \Omega$.
- If in fact $f_0(\overline{x}) < f_0(x)$ for all $x \in \Omega$, then $\overline{x}$ is said to be a strict global solution.

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\bar{x} \in \Omega$ is said to be a global solution to the problem
$$\min\{f_0(x) : x \in \Omega\}$$
if $f_0(\bar{x}) \le f_0(x)$ for all $x \in \Omega$.
- If in fact $f_0(\bar{x}) < f_0(x)$ for all $x \in \Omega$, then $\bar{x}$ is said to be a strict global solution.
- $\bar{x} \in \Omega$ is said to be a local solution to the problem $\min\{f_0(x) : x \in \Omega\}$ if there is an $\epsilon > 0$ such that
$$f_0(\bar{x}) \le f_0(x) \quad \text{for all } x \in \Omega \text{ satisfying} \quad \|\bar{x} - x\| \le \epsilon.$$

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\overline{x} \in \Omega$ is said to be a global solution to the problem
$$\min\{f_0(x) : x \in \Omega\}$$
if $f_0(\overline{x}) \le f_0(x)$ for all $x \in \Omega$.
- If in fact $f_0(\overline{x}) < f_0(x)$ for all $x \in \Omega$, then $\overline{x}$ is said to be a strict global solution.
- $\overline{x} \in \Omega$ is said to be a local solution to the problem $\min\{f_0(x) : x \in \Omega\}$ if there is an $\epsilon > 0$ such that
$$f_0(\overline{x}) \le f_0(x) \quad \text{for all } x \in \Omega \text{ satisfying} \quad \|\overline{x} - x\| \le \epsilon.$$
- If $f_0(\overline{x}) < f_0(x)$ for all $x \in \Omega$ with $\|x - \overline{x}\| \le \epsilon$, then $\overline{x}$ is called a strict local solution.

$f_0 : \mathbb{R}^n \to \mathbb{R}$ and $\Omega \subset \mathbb{R}^n$

- $\bar{x} \in \Omega$ is said to be a global solution to the problem
$$\min\{f_0(x) : x \in \Omega\}$$
if $f_0(\bar{x}) \le f_0(x)$ for all $x \in \Omega$.
- If in fact $f_0(\bar{x}) < f_0(x)$ for all $x \in \Omega$, then $\bar{x}$ is said to be a strict global solution.
- $\bar{x} \in \Omega$ is said to be a local solution to the problem $\min\{f_0(x) : x \in \Omega\}$ if there is an $\epsilon > 0$ such that
$$f_0(\bar{x}) \le f_0(x) \quad \text{for all } x \in \Omega \text{ satisfying } \quad \|\bar{x} - x\| \le \epsilon.$$
- If $f_0(\bar{x}) < f_0(x)$ for all $x \in \Omega$ with $\|x - \bar{x}\| \le \epsilon$, then $\bar{x}$ is called a strict local solution.
- The solution $\bar{x}$ is said to be isolated if $\bar{x}$ is the only local solution in the set $\{x \in \Omega : \|x - \bar{x}\| \le \epsilon\}$.

Optimality Conditions

These definitions, although sensible, are not practical since they require one to check infinitely many points to determine either local or global optimality.

Optimality Conditions

These definitions, although sensible, are not practical since they require one to check infinitely many points to determine either local or global optimality.

An optimality condition is a simple practical test for optimality.

Optimality Conditions

These definitions, although sensible, are not practical since they require one to check infinitely many points to determine either local or global optimality.

An optimality condition is a simple practical test for optimality.

Theorem: *If f_0 is differentiable at $\bar{x}$ and $\bar{x}$ is a local solution to the problem* $\min\{f_0(x) : x \in \mathbb{R}^n\}$, *then* $\nabla f_0(\bar{x}) = 0$.

Optimality Conditions

These definitions, although sensible, are not practical since they require one to check infinitely many points to determine either local or global optimality.

An optimality condition is a simple practical test for optimality.

Theorem: *If* f_0 *is differentiable at* $\overline{x}$ *and* $\overline{x}$ *is a local solution to the problem* $\min\{f_0(x) : x \in \mathbb{R}^n\}$, *then* $\nabla f_0(\overline{x}) = 0$.

The condition $\nabla f_0(\bar{x}) = 0$ is necessary, but clearly not sufficient for optimality.

Optimality Conditions

Optimality conditions play a key role in both the design of our algorithms and our tests for termination.

Optimality Conditions

Optimality conditions play a key role in both the design of our algorithms and our tests for termination.

We design our algorithms to locate points that satisfy some testable, or constructive, optimality conditions and then terminate when the procedure "nearly" solves these conditions.

Optimality Conditions

Optimality conditions play a key role in both the design of our algorithms and our tests for termination.

We design our algorithms to locate points that satisfy some testable, or constructive, optimality conditions and then terminate when the procedure “nearly” solves these conditions.

Our first order of business is to derive workable optimality conditions. In order to do this we must first review some facts from linear algebra and multi-variable calculus.