CME 307 / MS&E 311 / OIT 676: Optimization

Introduction

Professor Udell

Management Science and Engineering
Stanford

September 20, 2024

Announcements

announcements:

- website: https://stanford-cme-307.github.io/web
- instructors: Madeleine Udell and Dan Iancu
- TAs: Zach Frangella and Pratik Rathore
- Ed for discussion and announcements
- fill out course survey (also linked on website)
- talk to instructors after class and/or at office hours (see website)
- class attendance is required. will post slides, generally no recordings

before class starts: find someone you haven't met and introduce yourselves.

- name, major, year
- something fun you did this summer
- why are you interested in optimization?
- what are you hoping to learn?

Agenda for today

- Understand course objectives and expectations
- Identify several types of optimization problem
- Meet someone you've not met before
- Discuss challenges in a real-world optimization problem
- Review basic linear algebra

Outline

(Integer) linear optimization problem

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & Ax = b \\
 & \ell \leq x \leq u \\
\text{variable} & x \in \mathbb{Z}^{n_1} \times \mathbf{R}^{n_2}
\end{array}
$$

- objective $c^T x$
- equality constraints $Ax = b$
- lower and upper bounds $\ell \leq x \leq u$
- integer variables if $n_1 > 0$

problem data:

- $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $n = n_1 + n_2$
- $c \in \mathbf{R}^n$
- $\ell \in \mathbf{R}^n$, $u \in \mathbf{R}^n$

LP example: diet problem

- x_j servings of food j, $j = 1, \ldots, n$
- c_j cost per serving
- a_{ij} amount of nutrient i in food j
- b_i required amount of nutrient i, $i = 1, \ldots, m$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

LP example: diet problem

- x_j servings of food j, $j = 1, \ldots, n$
- c_j cost per serving
- a_{ij} amount of nutrient i in food j
- b_i required amount of nutrient i, $i = 1, \ldots, m$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

extensions:

- foods come from recipes? $x = By$

LP example: diet problem

- x_j servings of food j, $j = 1, \ldots, n$
- c_j cost per serving
- a_{ij} amount of nutrient i in food j
- b_i required amount of nutrient i, $i = 1, \ldots, m$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

extensions:

- foods come from recipes?
- ensure diversity in diet?

LP example: diet problem

- x_j servings of food j, $j = 1, \ldots, n$
- c_j cost per serving
- a_{ij} amount of nutrient i in food j
- b_i required amount of nutrient i, $i = 1, \ldots, m$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

extensions:

- foods come from recipes?
- ensure diversity in diet? $y \leq u$

LP example: diet problem

- x_j servings of food j, $j = 1, \dots, n$
- c_j cost per serving
- a_{ij} amount of nutrient i in food j
- b_i required amount of nutrient i, $i = 1, \dots, m$

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

extensions:

- foods come from recipes?
- ensure diversity in diet? $y \leq u$
- ranges of nutrients?

LP example: diet problem

- x_j servings of food j, $j = 1, \ldots, n$
- c_j cost per serving
- a_{ij} amount of nutrient i in food j
- b_i required amount of nutrient i, $i = 1, \ldots, m$

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & Ax = b \\
& x \geq 0
\end{array}
$$

extensions:

- foods come from recipes?
- ensure diversity in diet? $y \leq u$
- ranges of nutrients? $Ax + s = b$, $l \leq s \leq u$

Nonlinear optimization problem

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m_1 \\
& h_i(x) = 0, \quad i = 1, \ldots, m_2 \\
\text{variable} & x \in \mathbf{R}^n
\end{array}
$$

- objective f_0
- inequality constraints f_i
- equality constraints h_i

problem data:

- (blackbox) code to evaluate f_i and h_i for any $x \in \mathbf{R}^n$
- (first order) and to compute gradients
- (second order) and to compute Hessians

Example: process control

You are the process engineer for a desalination plant that produces drinking water. The plant has a variety of knobs, collected in vector x, that you can turn to control the process. These control, *e.g.*, how much water is pumped into the plant, how much pressure is used to force the water through filters, and how much of each chemical is added to the water.

- $f_0(x)$: cost of water produced
- $f_i(x)$: level of each measured impurity in the water
- b_i: maximum allowable level of each impurity

Given a setting of the knobs, you can observe the cost of water produced and the levels of impurities.

What is the optimal setting of the knobs?

Optimization problems

important optimization problem classes:

- linear
- integer
- nonlinear (with linear or nonlinear constraints)
- quadratic
- unconstrained
- finite-sum
- conic
- convex
- black-box with (0, 1, or 2)-order oracle

Optimization problems

important optimization problem classes:

- linear
- integer
- nonlinear (with linear or nonlinear constraints)
- quadratic
- unconstrained
- finite-sum
- conic
- convex
- black-box with (0, 1, or 2)-order oracle

draw a picture relating these

Modularity in optimization

how to optimize:

1. model problem as a mathematical optimization problem
2. identify the properties of the problem
3. use an appropriate solver (or write a new one)

. . . and iterate:

- approximate the problem to make it easier
- solve a sequence of approximated problems that converge to solve the original problem
- or initialize ("warm-start") a solver for the original problem with a solution to the approximated problem

Outline

Course goals

look at goals, materials, and grading on course website:
https://stanford-cme-307.github.io/web/

- Which goals sound exciting?
- Which goals don't make sense?
- What else do you hope to accomplish?
- Do expectations make sense given course goals?

Outline

What is an optimization problem?

Course goals and expectations

Linear Algebra Review

Span and nullspace

matrix $A \in \mathbf{R}^{m \times n}$. define

- span of A
- nullspace of A
- rank of A

geometry? what is the relationship between these?

proof: on board

Span and nullspace

matrix $A \in \mathbf{R}^{m \times n}$. define

- span of A: $\mathrm{span}(A) = \{Ax \mid x \in \mathbf{R}^n\} \subseteq \mathbf{R}^m$
- nullspace of A: $\mathbf{nullspace}(A) = \{x \in \mathbf{R}^n \mid Ax = 0\} \subseteq \mathbf{R}^n$
- rank of A: $\mathrm{Rank}(A) = \dim(\mathrm{span}(A))$

geometry? what is the relationship between these?

proof: on board

Span and nullspace

matrix $A \in \mathbf{R}^{m \times n}$. define

- span of A: $\mathrm{span}(A) = \{Ax \mid x \in \mathbf{R}^n\} \subseteq \mathbf{R}^m$
- nullspace of A: $\mathbf{nullspace}(A) = \{x \in \mathbf{R}^n \mid Ax = 0\} \subseteq \mathbf{R}^n$
- rank of A: $\mathrm{Rank}(A) = \dim(\mathrm{span}(A))$

geometry? what is the relationship between these?

Rank Nullity Theorem:

$$\mathrm{Rank}(A) + \dim(\mathbf{nullspace}(A)) = n$$

proof: on board

Span and nullspace

matrix $A \in \mathbf{R}^{m \times n}$. define

- span of A: $\operatorname{span}(A) = \{Ax \mid x \in \mathbf{R}^n\} \subseteq \mathbf{R}^m$
- nullspace of A: $\mathbf{nullspace}(A) = \{x \in \mathbf{R}^n \mid Ax = 0\} \subseteq \mathbf{R}^n$
- rank of A: $\operatorname{Rank}(A) = \dim(\operatorname{span}(A))$

geometry? what is the relationship between these?

Rank Nullity Theorem:

$$\operatorname{Rank}(A) + \dim(\mathbf{nullspace}(A)) = n$$

proof: on board

Solution of linear system

matrix $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $m \leq n$ ("fat matrix"). define

- solution set of linear system $\{x : Ax = b\}$

Solution of linear system

matrix $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $m \leq n$ ("fat matrix"). define

- solution set of linear system $\{x : Ax = b\}$

geometry: solution set is a **linear subspace** of $\mathbf{R}^n$

Solution of linear system

matrix $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $m \leq n$ ("fat matrix"). define

- solution set of linear system $\{x : Ax = b\}$

geometry: solution set is a **linear subspace** of $\mathbf{R}^n$ **Q:** what is the dimension of the solution set? when is it unique?

Solution of linear system

matrix $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $m \leq n$ ("fat matrix"). define

- solution set of linear system $\{x : Ax = b\}$

geometry: solution set is a **linear subspace** of $\mathbf{R}^n$ **Q:** what is the dimension of the solution set? when is it unique?

- solution is unique if $m = n$ and A is full rank.

Solution of linear system

matrix $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $m \leq n$ ("fat matrix"). define

- solution set of linear system $\{x : Ax = b\}$

geometry: solution set is a **linear subspace** of $\mathbf{R}^n$ **Q:** what is the dimension of the solution set? when is it unique?

- solution is unique if $m = n$ and A is full rank. proof:
  - rank nullity theorem

Solution of linear system

matrix $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $m \leq n$ ("fat matrix"). define

- solution set of linear system $\{x : Ax = b\}$

geometry: solution set is a **linear subspace** of $\mathbf{R}^n$ **Q:** what is the dimension of the solution set? when is it unique?

- solution is unique if $m = n$ and A is full rank. proof:
  - rank nullity theorem
- if $m < n$ and A is full rank, solution set is a hyperplane of dimension $n - m$.

Solution of linear system

matrix $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $m \leq n$ ("fat matrix"). define

- solution set of linear system $\{x : Ax = b\}$

geometry: solution set is a **linear subspace** of $\mathbf{R}^n$ **Q:** what is the dimension of the solution set? when is it unique?

- solution is unique if $m = n$ and A is full rank. proof:
  - rank nullity theorem
- if $m < n$ and A is full rank, solution set is a hyperplane of dimension $n - m$. proof:
  - **nullspace**(A), is a hyperplane of dimension $n - m$ by rank-nullity theorem
  - solution set is $\{x : Ax = b\} = \{x_0 + v : v \in \textbf{nullspace}(A)\}$

Solution of linear system

matrix $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $m \leq n$ ("fat matrix"). define

- solution set of linear system $\{x : Ax = b\}$

geometry: solution set is a **linear subspace** of $\mathbf{R}^n$ **Q:** what is the dimension of the solution set? when is it unique?

- solution is unique if $m = n$ and A is full rank. proof:
  - rank nullity theorem
- if $m < n$ and A is full rank, solution set is a hyperplane of dimension $n - m$. proof:
  - $\textbf{nullspace}(A)$, is a hyperplane of dimension $n - m$ by rank-nullity theorem
  - solution set is $\{x : Ax = b\} = \{x_0 + v : v \in \textbf{nullspace}(A)\}$

if these are confusing: review linear algebra and prove them all!

What next?

- website: https://stanford-cme-307.github.io/web
- Ed for discussion and announcements
- fill out course survey (linked on website)
- talk to instructors after class and during office hours (see website)
- class attendance is required. will post some slides, generally no recordings

Outline

Quadratic optimization

Nonlinear optimization

Conic optimization

Integer programming

Convex optimization

Quadratic optimization

a **quadratic optimization** problem is written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax - b\|^2 := f_0(x) \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $A \in \mathbf{R}^{m \times n}$: matrix
- $b \in \mathbf{R}^m$: vector

how to solve?

Quadratic optimization

a **quadratic optimization** problem is written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax - b\|^2 := f_0(x) \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $A \in \mathbf{R}^{m \times n}$: matrix
- $b \in \mathbf{R}^m$: vector

how to solve? take gradient and set to 0:

$$\nabla f_0(x) = A^T(Ax - b) = 0$$

$\Longrightarrow$ linear system solvers also solve quadratic optimization problems matrix $A \in \mathbf{R}^{m \times n}$

- check matrix calculus results by checking dimensions

Linear algebra review

matrix $A \in \mathbf{R}^{m \times n}$

- check matrix calculus results by checking dimensions
- normal equations $A^T A x = A^T b$

Linear algebra review

matrix $A \in \mathbf{R}^{m \times n}$

- check matrix calculus results by checking dimensions
- normal equations $A^T A x = A^T b$
- solution to $Ax = b$ is unique if

Linear algebra review

matrix $A \in \mathbf{R}^{m \times n}$

- check matrix calculus results by checking dimensions
- normal equations $A^T A x = A^T b$
- solution to $Ax = b$ is unique if $m = n$ and A is full rank

Linear algebra review

matrix $A \in \mathbf{R}^{m \times n}$

- check matrix calculus results by checking dimensions
- normal equations $A^T A x = A^T b$
- solution to $Ax = b$ is unique if $m = n$ and A is full rank
- if $m < n$ and A is full rank
  - solution set is a hyperplane of dimension

Linear algebra review

matrix $A \in \mathbf{R}^{m \times n}$

- check matrix calculus results by checking dimensions
- normal equations $A^T A x = A^T b$
- solution to $Ax = b$ is unique if $m = n$ and A is full rank
- if $m < n$ and A is full rank
  - solution set is a hyperplane of dimension $n - m$
  - null space of A, **nullspace**(A), is a hyperplane of dimension

Linear algebra review

matrix $A \in \mathbf{R}^{m \times n}$

- check matrix calculus results by checking dimensions
- normal equations $A^T A x = A^T b$
- solution to $Ax = b$ is unique if $m = n$ and A is full rank
- if $m < n$ and A is full rank
  - solution set is a hyperplane of dimension $n - m$
  - null space of A, $\mathbf{nullspace}(A)$, is a hyperplane of dimension $n - m$
  - solution set is $\{x : Ax = b\} = \{x_0 + Vz\}$ where columns of $V \in \mathbf{R}^{n \times n-m}$ span $\mathbf{nullspace}(A)$

Linear algebra review

matrix $A \in \mathbf{R}^{m \times n}$

- check matrix calculus results by checking dimensions
- normal equations $A^T A x = A^T b$
- solution to $Ax = b$ is unique if $m = n$ and A is full rank
- if $m < n$ and A is full rank
  - solution set is a hyperplane of dimension $n - m$
  - null space of A, $\mathbf{nullspace}(A)$, is a hyperplane of dimension $n - m$
  - solution set is $\{x : Ax = b\} = \{x_0 + Vz\}$ where columns of $V \in \mathbf{R}^{n \times n-m}$ span $\mathbf{nullspace}(A)$
- $A^T A$ is symmetric positive semidefinite (proof on board)

Symmetric positive semidefinite matrices

Definition

a symmetric matrix $Q \in \mathbf{R}^{n\times n}$ is **positive semidefinite** (psd) if $x^T Qx \geq 0$ for all $x \in \mathbf{R}^n$.

these matrices are so important that there are many ways to write them! for $Q \in \mathbf{R}^{n\times n}$,

$$Q \in \mathbf{S}^n_+ \iff Q \succeq 0 \iff Q = Q^T,\ \lambda_{\min}(Q) \geq 0$$

Symmetric positive semidefinite matrices

Definition

a symmetric matrix $Q \in \mathbf{R}^{n \times n}$ is **positive semidefinite** (psd) if $x^T Q x \geq 0$ for all $x \in \mathbf{R}^n$.

these matrices are so important that there are many ways to write them! for $Q \in \mathbf{R}^{n \times n}$,

$$Q \in \mathbf{S}^n_+ \iff Q \succeq 0 \iff Q = Q^T, \ \lambda_{\min}(Q) \geq 0$$

$Q \in \mathbf{S}^n_+$ is **symmetric positive definite** (spd) ($Q \succ 0$) if $x^T Q x > 0$ for all $x \in \mathbf{R}^n$.

Symmetric positive semidefinite matrices

Definition

a symmetric matrix $Q \in \mathbf{R}^{n\times n}$ is **positive semidefinite** (psd) if $x^T Qx \geq 0$ for all $x \in \mathbf{R}^n$.

these matrices are so important that there are many ways to write them! for $Q \in \mathbf{R}^{n\times n}$,

$$Q \in \mathbf{S}^n_+ \iff Q \succeq 0 \iff Q = Q^T,\ \lambda_{\min}(Q) \geq 0$$

$Q \in \mathbf{S}^n_+$ is **symmetric positive definite** (spd) ($Q \succ 0$) if $x^T Qx > 0$ for all $x \in \mathbf{R}^n$.

why care about psd matrices Q?

- least-squares objective has a psd $Q = A^T A$
- level sets of $x^T Qx$ are (bounded) ellipsoids
- the quadratic form $x^T Qx$ is a metric iff $Q \succ 0$
- eigenvalue decomp and svd coincide for psd matrices

Quadratic program

a **quadratic program** is written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T Qx + c^T x \\ \text{subject to} & Ax = b \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $Q \in \mathbf{R}^{n \times n}$: symmetric positive semidefinite matrix
- $c \in \mathbf{R}^n$: vector

how to solve?

Quadratic program

a **quadratic program** is written as

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} & Ax = b \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $Q \in \mathbf{R}^{n \times n}$: symmetric positive semidefinite matrix
- $c \in \mathbf{R}^n$: vector

how to solve? reduce to quadratic optimization problem:

- (explicit) form solution set $\{x : Ax = b\} = \{x_0 + Vz \mid z \in \mathbf{R}^{n-m}\}$ by computing a solution $Ax_0 = b$ and a basis V for the null space of A
- (implicit) use duality to recast problem as larger linear (KKT) system

Quadratic program: application

Markowitz portfolio optimization problem:

$$
\begin{array}{ll}
\text{minimize} & \gamma x^T \Sigma x - \mu^T x \\
\text{subject to} & \sum_i x_i = 1 \\
& Ax = 0 \\
\text{variable} & x \in \mathbf{R}^n
\end{array}
$$

where

- $\Sigma \in \mathbf{R}^{n \times n}$: asset covariance matrix
- $\mu \in \mathbf{R}^n$: asset return vector
- $\gamma \in \mathbf{R}$: risk aversion parameter
- rows of $A \in \mathbf{R}^{m \times n}$ correspond to other portfolios
  - ensures new portfolio is independent, *e.g.*, of market returns

Outline

Quadratic optimization

Nonlinear optimization

Conic optimization

Integer programming

Convex optimization

Unconstrained smooth optimization

for $f : \mathbf{R}^n \to \mathbf{R}$ ctsly differentiable,

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

how to solve?

Unconstrained smooth optimization

for $f : \mathbf{R}^n \to \mathbf{R}$ ctsly differentiable,

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

how to solve? approximate as a quadratic problem

$$f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T H(x_0)(x - x_0)$$

and find solution x_{quad} to the quadratic problem.
then set $x_0 \leftarrow x_{\text{quad}}$ and repeat.

Finite sum

finite sum optimization problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} f_i(x) \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

key fact: can approximate gradient using gradient on **minibatch** $S \subseteq \{1, \ldots, m\}$:

$$\nabla f(x) \approx \frac{1}{|S|} \sum_{i \in S} \nabla f_i(x)$$

examples:

- statistical learning (logistic regression, SVM)
- deep learning

Background: classification

classification problem: m data points

- feature vector $a_i \in \mathbf{R}^n$, $i = 1, \dots, m$
- label $b_i \in \{-1, 1\}$, $i = 1, \dots, m$

choose decision boundary $a^T x = 0$ to separate data points into two classes

- $a^T x > 0 \implies$ predict class 1
- $a^T x < 0 \implies$ predict class -1

classification is correct if $b_i a^T x > 0$

Background: classification

classification problem: m data points

- feature vector $a_i \in \mathbf{R}^n$, $i = 1, \ldots, m$
- label $b_i \in \{-1, 1\}$, $i = 1, \ldots, m$

choose decision boundary $a^T x = 0$ to separate data points into two classes

- $a^T x > 0 \implies$ predict class 1
- $a^T x < 0 \implies$ predict class -1

classification is correct if $b_i a^T x > 0$

- projective transformation transforms affine boundary to linear boundary
- classification is invariant to scalar multiplication of x

Logistic regression

(regularized) **logistic regression** minimizes the **finite sum**

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} \log(1+\exp\left(-b_i a_i^T x\right)) + r(x) \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $b_i \in \{-1, 1\}$, $a_i \in \mathbf{R}^n$
- $r : \mathbf{R}^n \to \mathbf{R}$ is a **regularizer**, *e.g.*, $\|x\|^2$ or $\|x\|_1$

Support vector machine

support vector machine (SVM) minimizes the **finite sum**

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \max(0, 1 - b_i a_i^T x) + \gamma \|x\|^2 \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where $b_i \in \{-1, 1\}$ and $a_i \in \mathbf{R}^n$.

Support vector machine

support vector machine (SVM) minimizes the **finite sum**

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} \max(0, 1 - b_i a_i^T x) + \gamma \|x\|^2 \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where $b_i \in \{-1, 1\}$ and $a_i \in \mathbf{R}^n$. not differentiable!

Support vector machine

support vector machine (SVM) minimizes the **finite sum**

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} \max(0, 1 - b_i a_i^T x) + \gamma \|x\|^2 \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where $b_i \in \{-1, 1\}$ and $a_i \in \mathbf{R}^n$. not differentiable!

how to solve?

Support vector machine

support vector machine (SVM) minimizes the **finite sum**

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \max(0, 1 - b_i a_i^T x) + \gamma \|x\|^2 \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where $b_i \in \{-1, 1\}$ and $a_i \in \mathbf{R}^n$. not differentiable!

how to solve?

- use **subgradient** method
- transform to **conic form**
- solve **dual** problem instead
- **smooth** the objective

Nonlinear optimization

optimization problem: nonlinear form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le b_i, \quad i = 1, \ldots, m_1 \\ & h(x) = 0 \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

- $x = (x_1, \ldots, x_n)$: optimization variables
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: objective function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$: constraint functions

special case: **unconstrained optimization**

Example: process control

You are the process engineer for a desalination plant that produces drinking water. The plant has a variety of knobs, collected in vector x, that you can turn to control the process. These control, *e.g.*, how much water is pumped into the plant, how much pressure is used to force the water through filters, and how much of each chemical is added to the water.

- $f_0(x)$: cost of water produced
- $f_i(x)$: level of each measured impurity in the water
- b_i: maximum allowable level of each impurity

Given a setting of the knobs, you can observe the cost of water produced and the levels of impurities.

What is the optimal setting of the knobs?

Oracles

an optimization **oracle** is your interface for accessing the problem data:
e.g., an oracle for $f : \mathbf{R}^n \to \mathbf{R}$ can evaluate for any $x \in \mathbf{R}^n$:

- **zero-order:** $f_0(x)$
- **first-order:** $f_0(x)$ and $\nabla f_0(x)$
- **second-order:** $f_0(x)$, $\nabla f_0(x)$, and $\nabla^2 f_0(x)$

why oracles?

- can optimize real systems based on observed output (not just models)
- can use and extend old or complex but trusted code (*e.g.*, NASA, PDE simulations, . . .)
- can prove lower bounds on the oracle complexity of a problem class

source: Nesterov 2004 "Introductory Lectures on Convex Optimization"'

Nonlinear optimization: how to solve?

depends on the oracle:

- first- or second-order: approximate by a sequence of quadratic problems
- zero-order: harder, lots of methods
  - simulated annealing
  - Bayesian optimization
  - pseudo-higher-order methods, *e.g.*, compute approximate gradient

Solution of an optimization problem

$$\text{minimize} \quad f(x)$$

for $f : \mathcal{D} \to \mathbf{R}$. $x^\star$ is a

- **local minimizer** if there is a neighborhood $\mathcal{N}$ around $x^\star$ so that $f(x) \geq f(x^\star)$ for all $x \in \mathcal{N}$.
- **global minimizer** if $f(x) \geq f(x^\star)$ for all $x \in \mathcal{D}$.
- **strict local minimizer** if there is a neighborhood $\mathcal{N}$ around $x^\star$ so that $f(x) > f(x^\star)$ for all $x \in \mathcal{N}$.
- **isolated local minimizer** if the neighborhood $\mathcal{N}$ contains no other local minimizers.
- **unique minimizer** if it is the only global minimizer.

Solution of an optimization problem

$$\text{minimize} \quad f(x)$$

for $f : \mathcal{D} \to \mathbf{R}$. $x^\star$ is a

- **local minimizer** if there is a neighborhood $\mathcal{N}$ around $x^\star$ so that $f(x) \geq f(x^\star)$ for all $x \in \mathcal{N}$.
- **global minimizer** if $f(x) \geq f(x^\star)$ for all $x \in \mathcal{D}$.
- **strict local minimizer** if there is a neighborhood $\mathcal{N}$ around $x^\star$ so that $f(x) > f(x^\star)$ for all $x \in \mathcal{N}$.
- **isolated local minimizer** if the neighborhood $\mathcal{N}$ contains no other local minimizers.
- **unique minimizer** if it is the only global minimizer.

pictures!

First order optimality condition

Theorem

If $x^\star \in \mathbf{R}^n$ is a local minimizer of a differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, then $\nabla f(x^\star) = 0$.

First order optimality condition

Theorem

If $x^\star \in \mathbf{R}^n$ is a local minimizer of a differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, then $\nabla f(x^\star) = 0$.

proof: suppose by contradiction that $\nabla f(x^\star) \neq 0$. consider points of the form $x_\alpha = x^\star - \alpha \nabla f(x^\star)$ for $\alpha > 0$. by definition of the gradient,

$$\lim_{\alpha \to 0} \frac{f(x_\alpha) - f(x^\star)}{\alpha} = -\nabla f(x^\star)^\top \nabla f(x^\star) = -\|nablaf(x^\star)\|^2 < 0$$

so for any sufficiently small $\alpha > 0$, we have $f(x_\alpha) < f(x^\star)$, which contradicts the fact that $x^\star$ is a local minimizer.

Second order optimality condition

Theorem

If $x^\star \in \mathbf{R}^n$ is a local minimizer of a twice differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, then $\nabla^2 f(x^\star) \succeq 0$.

Second order optimality condition

Theorem

If $x^\star \in \mathbf{R}^n$ is a local minimizer of a twice differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, then $\nabla^2 f(x^\star) \succeq 0$.

proof: similar to the previous proof. use the fact that the second order approximation

$$f(x_\alpha) \approx f(x^\star) + \nabla f(x^\star)^\top (x_\alpha - x^\star) + \frac{1}{2}(x_\alpha - x^\star)^\top \nabla^2 f(x^\star)(x_\alpha - x^\star)$$

is accurate locally to show a contradiction unless $\nabla^2 f(x^\star) \succeq 0$: if not, there is a direction v such that $v^T \nabla^2 f(x^\star) v < 0$. then $f(x + \alpha v) < f(x^\star)$ for α arbitrarily small, which contradicts the fact that $x^\star$ is a local minimizer.

Outline

Linear program

a **linear program** is written as

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & b - Ax \geq 0 \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $A \in \mathbf{R}^{m \times n}$: matrix
- $b \in \mathbf{R}^m$: vector
- $c \in \mathbf{R}^n$: vector

how to solve?

Linear program

a **linear program** is written as

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & b - Ax \geq 0 \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

where

- $A \in \mathbf{R}^{m \times n}$: matrix
- $b \in \mathbf{R}^m$: vector
- $c \in \mathbf{R}^n$: vector

how to solve?

- use the simplex method
- use a conic solver

Conic form

conic form optimization problem generalizes LP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & b - Ax \in \mathcal{K}, \end{array}$$

where $\mathcal{K}$ is a **convex cone**:

$$x \in \mathcal{K} \iff rx \in \mathcal{K} \text{ for any } r > 0.$$

examples:

- zero cone $\mathcal{K}_0 = \{0\}$
- positive orthant $\mathcal{K}_+ = \{x : x_i >= 0, \; i = 1, \ldots, n\}$
- second order cone $\mathcal{K}_{\mathrm{SOC}} = \{(x, t) : \|x\|_2 \leq t\}$
- positive semidefinite (PSD) cone $\mathcal{K}_{\mathrm{SDP}} = \{X : X = X^T, \; v^T X v \geq 0, \; \forall v \in \mathbf{R}^n\}$
- cartesian products of cones

Conic form: how to solve?

Morally, conic problems are solved by reducing to a nonlinear optimization problem

- barrier methods (*e.g.*, interior point methods)
    - add a barrier term to the objective that goes to infinity when constraints are violated
- penalty methods (*e.g.*, augmented Lagrangian methods, ADMM, ...)
    - add a penalty term to the objective that depends on a dual variable
    - adjust the dual variable to enforce constraints

Conic form example: nonnegative least squares

$$\begin{array}{ll} \text{minimize} & \|Ax - b\| \\ \text{subject to} & x \geq 0 \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & x \in \mathcal{K}_+ \\ & (Ax - b, t) \in \mathcal{K}_{\text{SOC}} \end{array}$$

Conic form example: SVM

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \max(0, 1 - b_i a_i^T x) + \|x\|^2 \\ \text{variable} & x \in \mathbf{R}^n \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \text{minimize} & \sum_i s_i + t \\ \text{subject to} & s \geq \mathbf{diag}(b) A x - 1 \\ & s \geq 0 \\ & t \geq \|x\|^2 \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \text{minimize} & \sum_i s_i + t \\ \text{subject to} & s - \mathbf{diag}(b) A x + 1 \in \mathcal{K}_+ \\ & s \in \mathcal{K}_+ \\ & [t \ x; x^T \ I_n] \in \mathcal{K}_{\mathrm{SDP}} \end{array}$$

Schur complement

Consider the block matrix

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}.$$

- the **Schur complement** of A in X is $C - B^T A^{-1} B$.
- $X \succeq 0$ if and only if $A \succeq 0$ and $C - B^T A^{-1} B \succeq 0$.
  (proof by partial minimization of quadratic form $(u, v)^T X (u, v)$ over $u \in \mathbf{R}^m$ for fixed $v \in \mathbf{R}^n$)

Conic form example: semidefinite programming

$$
\begin{array}{ll}
\text{minimize} & \lambda_{\max}(X) + y^T X^{-1} y \\
\text{subject to} & X \succeq 0
\end{array}
$$

$$\Updownarrow$$

Conic form example: semidefinite programming

$$
\begin{array}{ll}
\text{minimize} & \lambda_{\max}(X) + y^T X^{-1} y \\
\text{subject to} & X \succeq 0
\end{array}
$$

$$\Updownarrow$$

$$
\begin{array}{ll}
\text{minimize} & t_1 + t_2 \\
\text{subject to} & t_1 I - X \in \mathcal{K}_{\mathsf{SDP}} \\
& \begin{bmatrix} t_2 & y^T \\ y & X \end{bmatrix} \in \mathcal{K}_{\mathsf{SDP}}
\end{array}
$$

Outline

Quadratic optimization

Nonlinear optimization

Conic optimization

Integer programming

Convex optimization

Integer programming

integer linear programming generalizes linear programming:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & b - Ax \geq 0 \\ \text{variable} & x \in \mathbf{Z}^n \end{array}$$

variants:

- **mixed integer linear programming** (MILP): $x \in \mathbf{Z}^{n-m} \cup \mathbf{R}^m$
- **mixed integer nonlinear programming** (MINLP): $x \in \mathbf{Z}^{n-m} \cup \mathbf{R}^m$ and nonlinear objective or constraints

how to solve?

Integer programming

integer linear programming generalizes linear programming:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & b - Ax \geq 0 \\ \text{variable} & x \in \mathbf{Z}^n \end{array}$$

variants:

- **mixed integer linear programming** (MILP): $x \in \mathbf{Z}^{n-m} \cup \mathbf{R}^m$
- **mixed integer nonlinear programming** (MINLP): $x \in \mathbf{Z}^{n-m} \cup \mathbf{R}^m$ and nonlinear objective or constraints

how to solve?

- use Gurobi, CPLEX, ...
- branch and bound and cut (*i.e.*, a sequence of LPs)
- use duality to decompose into a sequence of simpler LPs

Outline

Convex sets

Definition

A set $S \subseteq \mathbf{R}^n$ is convex if it contains every chord: for all $\theta \in [0,1]$, w, $v \in S$,

$$\theta w + (1-\theta) v \in S$$

Convex sets

Definition

A set $S \subseteq \mathbf{R}^n$ is convex if it contains every chord: for all $\theta \in [0, 1]$, w, $v \in S$,

$$\theta w + (1 - \theta) v \in S$$

Q: Which of these are convex?
ellipsoid, half moon

Convex functions

a function $f : \mathbf{R}^n \to \mathbf{R}$ is convex iff

Convex functions

a function $f : \mathbf{R}^n \to \mathbf{R}$ is convex iff

- **Chords.** it never lies above its chord: $\forall \theta \in [0, 1]$, $w, v \in \mathbf{R}^n$

$$f(\theta w + (1 - \theta)v) \leq \theta f(w) + (1 - \theta)f(v)$$

Convex functions

a function $f : \mathbf{R}^n \to \mathbf{R}$ is convex iff

- **Chords.** it never lies above its chord: $\forall \theta \in [0,1]$, $w, v \in \mathbf{R}^n$

$$f(\theta w + (1-\theta)v) \leq \theta f(w) + (1-\theta)f(v)$$

- **Epigraph.** $\mathbf{epi}(f) = \{(x,t) : \; t \geq f(x)\}$ is convex

Convex functions

a function $f : \mathbf{R}^n \to \mathbf{R}$ is convex iff

- **Chords.** it never lies above its chord: $\forall \theta \in [0, 1]$, $w, v \in \mathbf{R}^n$

$$f(\theta w + (1-\theta)v) \leq \theta f(w) + (1-\theta) f(v)$$

- **Epigraph.** $\textbf{epi}(f) = \{(x, t) : \; t \geq f(x)\}$ is convex
- **First order condition.** if f is differentiable,

$$f(v) - f(w) \geq \nabla f(w)^\top (v - w) \qquad \forall w, v \in \mathbf{R}^n$$

Convex functions

a function $f : \mathbf{R}^n \to \mathbf{R}$ is convex iff

- **Chords.** it never lies above its chord: $\forall \theta \in [0, 1]$, $w, v \in \mathbf{R}^n$

$$f(\theta w + (1-\theta)v) \leq \theta f(w) + (1-\theta)f(v)$$

- **Epigraph.** $\mathbf{epi}(f) = \{(x, t) : \ t \geq f(x)\}$ is convex
- **First order condition.** if f is differentiable,

$$f(v) - f(w) \geq \nabla f(w)^\top (v - w) \qquad \forall w, v \in \mathbf{R}^n$$

- **Second order condition.** If f is twice differentiable, its Hessian is always psd:

$$\lambda_{\min}(\nabla^2 f(x)) \geq 0 \qquad \text{for all} x \in \mathbf{R}^n$$

Convex functions

a function $f : \mathbf{R}^n \to \mathbf{R}$ is convex iff

- **Chords.** it never lies above its chord: $\forall \theta \in [0,1]$, $w, v \in \mathbf{R}^n$

$$f(\theta w + (1-\theta)v) \leq \theta f(w) + (1-\theta) f(v)$$

- **Epigraph.** $\textbf{epi}(f) = \{(x,t) :\ t \geq f(x)\}$ is convex
- **First order condition.** if f is differentiable,

$$f(v) - f(w) \geq \nabla f(w)^\top (v - w) \qquad \forall w, v \in \mathbf{R}^n$$

- **Second order condition.** If f is twice differentiable, its Hessian is always psd:

$$\lambda_{\min}(\nabla^2 f(x)) \geq 0 \qquad \text{for all} x \in \mathbf{R}^n$$

Q: Which of these are convex?
quadratic, l1, pwl, step, jump, logistic, logistic loss

Convex optimization

an optimization problem is convex if:

- **Geometrically:** the feasible set and the epigraph of the objective are convex
- **NLP:** the objective and inequality constraints are convex functions, and the equality constraints are affine
- **Conic:** all the cones are convex cones

Convex optimization

an optimization problem is convex if:

- **Geometrically:** the feasible set and the epigraph of the objective are convex
- **NLP:** the objective and inequality constraints are convex functions, and the equality constraints are affine
- **Conic:** all the cones are convex cones

why convex optimization?

- relatively complete theory
- efficient solvers
- conceptual tools that generalize

duality, stopping conditions, . . .

Convex optimization

an optimization problem is convex if:

- **Geometrically:** the feasible set and the epigraph of the objective are convex
- **NLP:** the objective and inequality constraints are convex functions, and the equality constraints are affine
- **Conic:** all the cones are convex cones

why convex optimization?

- relatively complete theory
- efficient solvers
- conceptual tools that generalize

duality, stopping conditions, . . .

- a function f is concave if $-f$ is convex
- concave maximization results in a **convex** optimization problem

Local minima are global for convex functions

Theorem

If $x^\star$ is a local minimizer of a convex function f, then $x^\star$ is a global minimizer.

Local minima are global for convex functions

Theorem

If $x^\star$ is a local minimizer of a convex function f, then $x^\star$ is a global minimizer.

proof: suppose by contradiction that another point x' is a global minimizer, with $f(x') < f(x^\star)$. draw the chord between x' and $x^\star$. since the chord lies above f, every convex combination $x = \theta x^\star + (1-\theta)x'$ of x' and $x^\star$ for $\theta \in (0,1)$ has a value $f(x) < f(x^\star)$. this is true even for $x \to x^\star$, contradicting our assumption that $x^\star$ is a local minimizer.

Corollary

Corollary

If f is convex and differentiable and $\nabla f(x^\star) = 0$, then $x^\star$ is a global minimizer.

Corollary

Corollary

If f is convex and differentiable and $\nabla f(x^\star) = 0$, then $x^\star$ is a global minimizer.

Q: Is a global minimizer of a convex function always unique?

Corollary

Corollary

If f is convex and differentiable and $\nabla f(x^\star) = 0$, then $x^\star$ is a global minimizer.

Q: Is a global minimizer of a convex function always unique?

A: No. Picture.

Modern solvers

- algebraic modeling languages, *e.g.*
  - JuMP facilitates nonlinear and mixed integer optimization
  - CVX* (CVX, CVXPY, Convex.jl, . . .) transform a problem into conic form
- and modern solvers

Optimization modeling

- Rocket control
- Power systems
- AML

Announcements

- website: https://stanford-cme-307.github.io/web
- Ed for discussion and announcements: https://edstem.org/us/courses/51411/
- fill out course survey (also linked on website):
  https://forms.gle/7hPniFeC576S12FAA
- talk to me after class and/or schedule office hours (see website)
- class attendance is required. will post some slides, generally no recordings