1. Course Overview
2. Administrivia
3. Linear Programming
Outline

1. Course Overview
2. Administrivia
3. Linear Programming
Mathematical Optimization

The task of selecting the “best” configuration of a set of variables from a “feasible” set of configurations.

minimize (or maximize) \( f(x) \)
subject to \( x \in \mathcal{X} \)

- Terminology: decision variable(s), objective function, feasible set, optimal solution, optimal value
- Two main classes: continuous and combinatorial
Continuous Optimization Problems

Optimization problems where feasible set $\mathcal{X}$ is a connected subset of Euclidean space, and $f$ is a continuous function.

- Instances typically formulated as follows.

  \[
  \begin{align*}
  \text{minimize} & \quad f(x) \\
  \text{subject to} & \quad g_i(x) \leq b_i, \quad \text{for } i \in C.
  \end{align*}
  \]

- Objective function $f : \mathbb{R}^n \to \mathbb{R}$.
- Constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$. The inequality $g_i(x) \leq b_i$ is the $i$’th constraint.
- In general, intractable to solve efficiently (NP hard)
Convex Optimization Problem

A continuous optimization problem where $f$ is a convex function on $\mathcal{X}$, and $\mathcal{X}$ is a convex set.

- **Convex function:** $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$ for all $x, y \in \mathcal{X}$ and $\alpha \in [0, 1]$
- **Convex set:** $\alpha x + (1 - \alpha)y \in \mathcal{X}$, for all $x, y \in \mathcal{X}$ and $\alpha \in [0, 1]$
- Convexity of $\mathcal{X}$ implied by convexity of $g_i$’s
- For maximization problems, $f$ should be **concave**
- Typically solvable efficiently (i.e. in polynomial time)
- Encodes optimization problems from a variety of application areas
Convex Optimization Example: Least Squares Regression

Given a set of measurements \((a_1, b_1), \ldots, (a_m, b_m)\), where \(a_i \in \mathbb{R}^n\) is the \(i\)'th input and \(b_i \in \mathbb{R}\) is the \(i\)'th output, find the linear function \(f : \mathbb{R}^n \rightarrow \mathbb{R}\) best explaining the relationship between inputs and outputs.

- \(f(a) = x^\top a\) for some \(x \in \mathbb{R}^n\)
- Least squares: minimize mean-square error.

\[
\text{minimize} \quad \|Ax - b\|_2^2
\]
Given a directed network $G = (V, E)$ with cost $c_e \in \mathbb{R}_+$ per unit of traffic on edge $e$, and capacity $d_e$, find the minimum cost routing of $r$ divisible units of traffic from $s$ to $t$. 

\[
\begin{align*}
\text{minimize} & \quad \sum_{e \in E} c_e x_e \\
\text{subject to} & \quad \sum_{e \leftarrow v} x_e = \sum_{e \rightarrow v} x_e, \quad \text{for } v \in V \setminus \{s, t\}.
\end{align*}
\]

\[
\sum_{e \leftarrow s} x_e \leq d_e, \quad \text{for } e \in E.
\]

\[
x_e \geq 0, \quad \text{for } e \in E.
\]
Given a directed network $G = (V, E)$ with cost $c_e \in \mathbb{R}_+$ per unit of traffic on edge $e$, and capacity $d_e$, find the minimum cost routing of $r$ divisible units of traffic from $s$ to $t$.

minimize $\sum_{e \in E} c_e x_e$

subject to $\sum_{e \leftarrow v} x_e = \sum_{e \rightarrow v} x_e$, for $v \in V \setminus \{s, t\}$.

$\sum_{e \leftarrow s} x_e = r$

$x_e \leq d_e$, for $e \in E$.

$x_e \geq 0$, for $e \in E$. 

Generalizes to traffic-dependent costs. For example $c_e(x_e) = a_e x_e^2 + b_e x_e + c_e$. 

Course Overview 5/29
Given a directed network $G = (V, E)$ with cost $c_e \in \mathbb{R}_+$ per unit of traffic on edge $e$, and capacity $d_e$, find the minimum cost routing of $r$ divisible units of traffic from $s$ to $t$.

\[
\begin{align*}
\text{minimize} & & \sum_{e \in E} c_e x_e \\
\text{subject to} & & \sum_{e \leftarrow v} x_e = \sum_{e \rightarrow v} x_e, \quad \text{for } v \in V \setminus \{s, t\}.
\sum_{e \leftarrow s} x_e = r \\
x_e & \leq d_e, \quad \text{for } e \in E.
\end{align*}
\]

Generalizes to traffic-dependent costs. For example
\[
c_e(x_e) = a_e x_e^2 + b_e x_e + c_e.
\]
Combinatorial Optimization

Combinatorial Optimization Problem

An optimization problem where the feasible set $\mathcal{X}$ is finite.

- e.g. $\mathcal{X}$ is the set of paths in a network, assignments of tasks to workers, etc...
- Again, NP-hard in general, but many are efficiently solvable (either exactly or approximately)
Given a directed network $G = (V, E)$ with cost $c_e \in \mathbb{R}_+$ on edge $e$, find the minimum cost path from $s$ to $t$. 
Combinatorial Optimization Example: Traveling Salesman Problem

Given a set of cities $V$, with $d(u, v)$ denoting the distance between cities $u$ and $v$, find the minimum length tour that visits all cities.
Some optimization problems are best formulated as one or the other.

Many problems, particularly in computer science and operations research, can be formulated as both.

This dual perspective can lead to structural insights and better algorithms.
The shortest path problem can be encoded as a minimum cost flow problem, using distances as the edge costs, unit capacities, and desired flow rate 1.

The optimum solution of the (linear) convex program above will assign flow only on a single path — namely the shortest path.
Course Goals

- Recognize and model convex optimization problems, and develop a general understanding of the relevant algorithms.
- Formulate combinatorial optimization problems as convex programs
- Use both the discrete and continuous perspectives to design algorithms and gain structural insights for optimization problems
Anyone planning to do research in the design and analysis of algorithms
  - Convex and combinatorial optimization have become an indispensable part of every algorithmist’s toolkit

Students interested in theoretical machine learning and AI
  - Convex optimization underlies much of machine learning
  - Submodularity has recently emerged as an important abstraction for feature selection, active learning, planning, and other applications

Anyone else who solves or reasons about optimization problems: electrical engineers, control theorists, operations researchers, economists . . .
  - If there are applications in your field you would like to hear more about, let me know.
Course Outline

- Weeks 1-4: Convex optimization basics and duality theory
- Week 5: Algorithms for convex optimization
- Weeks 6-8: Viewing discrete problems as convex programs; structural and algorithmic implications.
- Weeks 9-14: Matroid theory, submodular optimization, and other applications of convex optimization to combinatorial problems
- Week 15: Project presentations (or additional topics)
Outline

1. Course Overview
2. Administrivia
3. Linear Programming
Basic Information

- Lecture time: Tuesdays and Thursdays 2 pm - 3:20 pm
- Lecture place: KAP 147
- Instructor: Shaddin Dughmi
  - Email: shaddin@usc.edu
  - Office: SAL 234
  - Office Hours: TBD
- Course Homepage: www.cs.usc.edu/people/shaddin/cs599fa13
- References: Convex Optimization by Boyd and Vandenberghe, and Combinatorial Optimization by Korte and Vygen. (Will place on reserve)
Prerequisites

- Mathematical maturity: Be good at proofs
- Substantial exposure to algorithms or optimization
  - CS570 or equivalent, or
  - CS303 and you did really well
Requirements and Grading

- This is an advanced elective class, so grade is not the point.
  - I assume you want to learn this stuff.
- 3-4 homeworks, 75% of grade.
  - Proof based.
  - Challenging.
  - Discussion allowed, even encouraged, but must write up solutions independently.
- Research project or final, 25% of grade. Project suggestions will be posted on website.
- One late homework allowed, 2 days.
Survey

- Name
- Email
- Department
- Degree
- Relevant coursework/background
- Research project idea
Outline

1. Course Overview
2. Administrivia
3. Linear Programming
A Brief History

- The forefather of convex optimization problems, and the most ubiquitous.
- Developed by Kantorovich during World War II (1939) for planning the Soviet army’s expenditures and returns. Kept secret.
- Discovered a few years later by George Dantzig, who in 1947 developed the simplex method for solving linear programs
- John von Neumann developed LP duality in 1947, and applied it to game theory
- Polynomial-time algorithms: Ellipsoid method (Khachiyan 1979), interior point methods (Karmarkar 1984).
minimize (or maximize) $c^T x$
subject to
\[ a_i^T x \leq b_i, \quad \text{for } i \in C^1. \]
\[ a_i^T x \geq b_i, \quad \text{for } i \in C^2. \]
\[ a_i^T x = b_i, \quad \text{for } i \in C^3. \]

- **Decision variables:** $x \in \mathbb{R}^n$
- **Parameters:**
  - $c \in \mathbb{R}^n$ defines the linear objective function
  - $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ define the $i$'th constraint.
maximize \: \mathbf{c}^\top \mathbf{x} \\
subject to \: \mathbf{a}_i^\top \mathbf{x} \leq b_i, \quad \text{for } i = 1, \ldots, m. \\
\quad \mathbf{x}_j \geq 0, \quad \text{for } j = 1, \ldots, n.

Every LP can be transformed to this form

- minimizing $\mathbf{c}^\top \mathbf{x}$ is equivalent to maximizing $-\mathbf{c}^\top \mathbf{x}$
- $\geq$ constraints can be flipped by multiplying by $-1$
- Each equality constraint can be replaced by two inequalities
- Unconstrained variable $\mathbf{x}_j$ can be replaced by $\mathbf{x}_j^+ - \mathbf{x}_j^-$, where both $\mathbf{x}_j^+$ and $\mathbf{x}_j^-$ are constrained to be nonnegative.
Geometric View
A 2-D example

maximize \( x_1 + x_2 \)
subject to
\[
\begin{align*}
x_1 + 2x_2 &\leq 2 \\
2x_1 + x_2 &\leq 2 \\
x_1, x_2 &\geq 0
\end{align*}
\]
Application: Optimal Production

- $n$ products, $m$ raw materials
- Product $j$ uses $a_{ij}$ units of raw material $i$
- There are $b_i$ units of material $i$ available
- Product $j$ yields profit $c_j$ per unit
- Facility wants to maximize profit subject to available raw materials

\[
\begin{align*}
\text{maximize} & \quad c^T x \\
\text{subject to} & \quad a_i^T x \leq b_i, \quad \text{for } i = 1, \ldots, m. \\
& \quad x_j \geq 0, \quad \text{for } j = 1, \ldots, n.
\end{align*}
\]
Terminology

- **Hyperplane**: The region defined by a linear equality
- **Halfspace**: The region defined by a linear inequality $a_i^Tx \leq b_i$.
- **Polytope**: The intersection of a set of linear inequalities in Euclidean space
  - Feasible region of an LP is a polytope
  - Equivalently: convex hull of a finite set of points
- **Vertex**: A point $x$ is a vertex of polytope $P$ if $\exists y \neq 0$ with $x + y \in P$ and $x - y \in P$
- **Face of $P$**: The intersection with $P$ of a hyperplane $H$ disjoint from the interior of $P$
Basic Facts about LPs and Polytopes

Fact

Feasible regions of LPs (i.e. polytopes) are convex
Basic Facts about LPs and Polytopes

**Fact**
Feasible regions of LPs (i.e. polytopes) are convex

**Fact**
Set of optimal solutions of an LP is convex
- In fact, a face of the polytope
- intersection of $P$ with hyperplane $c^T x = OPT$
Basic Facts about LPs and Polytopes

Fact
Feasible regions of LPs (i.e. polytopes) are convex

Fact
Set of optimal solutions of an LP is convex
- In fact, a face of the polytope
- Intersection of $P$ with hyperplane $c^T x = OPT$

Fact
At a vertex, $n$ linearly independent constraints are satisfied with equality (a.k.a. **tight**)
Basic Facts about LPs and Polytopes

**Fact**

An LP either has an optimal solution, or is *unbounded* or *infeasible*.
Fundamental Theorem of LP

If an LP in standard form has an optimal solution, then it has a vertex optimal solution.

Proof
Assume not, and take a non-vertex optimal solution $x$ with the maximum number of tight constraints.

There is $y \neq 0$ s.t. $x \pm y$ are feasible.

$y$ is perpendicular to the objective function and the tight constraints at $x$.

i.e., $c^\top y = 0$, and $a_i^\top y = 0$ whenever the $i$'th constraint is tight for $x$.

Can choose $y$ s.t. $y_j < 0$ for some $j$.

Let $\alpha$ be the largest constant such that $x + \alpha y$ is feasible.

Such an $\alpha$ exists.

An additional constraint becomes tight at $x + \alpha y$, a contradiction.
**Fundamental Theorem of LP**

If an LP in standard form has an optimal solution, then it has a vertex optimal solution.

**Proof**

- Assume not, and take a non-vertex optimal solution \( x \) with the maximum number of tight constraints.
- There is \( y \neq 0 \) s.t. \( x \pm y \) are feasible.
- \( y \) is perpendicular to the objective function and the tight constraints at \( x \).
  - i.e. \( c^T y = 0 \), and \( a_i^T y = 0 \) whenever the \( i \)'th constraint is tight for \( x \).
- Can choose \( y \) s.t. \( y_j < 0 \) for some \( j \).
- Let \( \alpha \) be the largest constant such that \( x + \alpha y \) is feasible.
  - Such an \( \alpha \) exists.
- An additional constraint becomes tight at \( x + \alpha y \), a contradiction.
Corollary

If an LP in standard form has an optimal solution, then there is an optimal solution with at most \( m \) non-zero variables.

\[
\begin{align*}
\text{maximize} & \quad c^T x \\
\text{subject to} & \quad a_i^T x \leq b_i, \quad \text{for } i = 1, \ldots, m. \\
& \quad x_j \geq 0, \quad \text{for } j = 1, \ldots, n.
\end{align*}
\]

- e.g. for optimal production with \( n \) products and \( m \) raw materials, there is an optimal plan with at most \( m \) products.
Next Lecture

- LP Duality and its interpretations
- Examples of duality relationships
- Implications of Duality