CS599: Convex and Combinatorial Optimization
Fall 2013
Lecture 1: Introduction to Optimization

Instructor: Shaddin Dughmi

Outline

Outline

Mathematical Optimization

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

$$\begin{array}{ll} \text{minimize (or maximize)} & f(x) \\ \text{subject to} & x \in \mathcal{X} \end{array}$$

- 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.

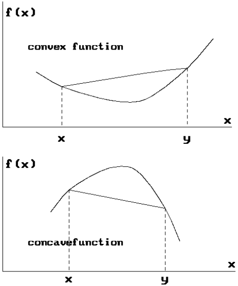
$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq b_i, \quad \text{for } i \in \mathcal{C}. \end{array}$$

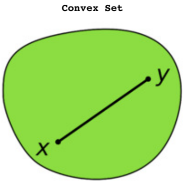
- Objective function $f : \mathbb{R}^n \to \mathbb{R}$.
- Constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$. The inequality $g_i(x) \leq b_i$ is the i'th constraint.
- In general, intractable to solve efficiently (NP hard)

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



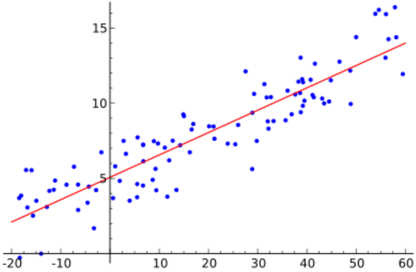


Convex Optimization Example: Least Squares Regression

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 $f : \mathbb{R}^n \to \mathbb{R}$ best explaining the relationship between inputs and outputs.

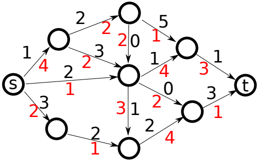
- $f(a) = x^\intercal a$ for some $x \in \mathbb{R}^n$
- Least squares: minimize mean-square error.

$$\text{minimize} \quad ||Ax - b||_2^2$$



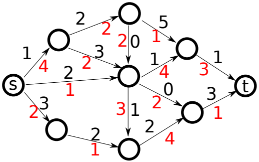
Convex Optimization Example: Minimum Cost Flow

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.



Convex Optimization Example: Minimum Cost Flow

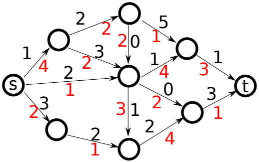
Given a directed network $G = (V, E)$ with cost $c_e \in \mathbb{R}_+$ per unit of traffic on edge e, and capacity d_e, find the minimum cost routing of r divisible units of traffic from s to t.



$$
\begin{array}{lll}
\text{minimize} & \sum_{e \in E} c_e x_e & \\
\text{subject to} & \sum_{e \leftarrow v} x_e = \sum_{e \rightarrow v} x_e, & \text{for } v \in V \setminus \{s, t\}. \\
& \sum_{e \leftarrow s} x_e = r & \\
& x_e \leq d_e, & \text{for } e \in E. \\
& x_e \geq 0, & \text{for } e \in E.
\end{array}
$$

Convex Optimization Example: Minimum Cost Flow

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.



$$
\begin{array}{lll}
\text{minimize} & \sum_{e \in E} c_e x_e & \\
\text{subject to} & \sum_{e \leftarrow v} x_e = \sum_{e \rightarrow v} x_e, & \text{for } v \in V \setminus \{s, t\}\,. \\
& \sum_{e \leftarrow s} x_e = r & \\
& x_e \leq d_e, & \text{for } e \in E. \\
& x_e \geq 0, & \text{for } e \in E.
\end{array}
$$

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

Combinatorial Optimization

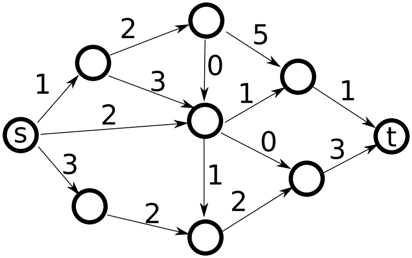
Combinatorial Optimization Problem

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)

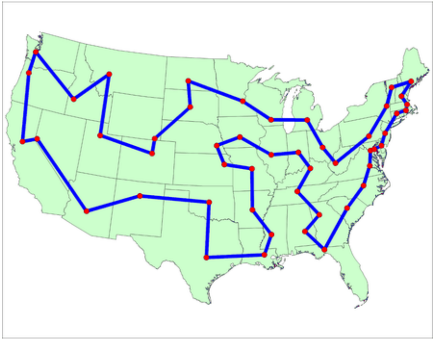
Combinatorial Optimization Example: Shortest Path

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.

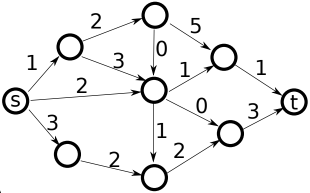


Continuous vs Combinatorial Optimization

- 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

Example: Shortest Path

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



$$
\begin{array}{lll}
\text{minimize} & \sum_{e \in E} c_e x_e & \\
\text{subject to} & \sum_{e \leftarrow v} x_e = \sum_{e \rightarrow v} x_e, & \text{for } v \in V \setminus \{s,t\}. \\
& \sum_{e \leftarrow s} x_e = 1 & \\
& x_e \leq 1, & \text{for } e \in E. \\
& x_e \geq 0, & \text{for } e \in E.
\end{array}
$$

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

Who Should Take this Class

- Anyone planning to do research in the design and analysis of algorithms
  - Convex and combinatorial optimization have become an indispensible 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

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).

LP General Form

$$
\begin{array}{lll}
\text{minimize (or maximize)} & c^\intercal x & \\
\text{subject to} & a_i^\intercal x \leq b_i, & \text{for } i \in \mathcal{C}^1. \\
& a_i^\intercal x \geq b_i, & \text{for } i \in \mathcal{C}^2. \\
& a_i^\intercal x = b_i, & \text{for } i \in \mathcal{C}^3.
\end{array}
$$

- 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.

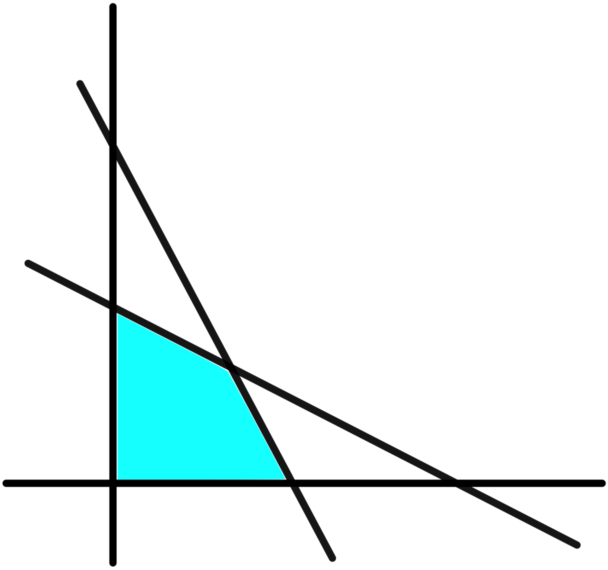
Standard Form

$$\begin{array}{lll} \text{maximize} & c^\intercal x & \\ \text{subject to} & a_i^\intercal x \le b_i, & \text{for } i = 1, \ldots, m. \\ & x_j \ge 0, & \text{for } j = 1, \ldots, n. \end{array}$$

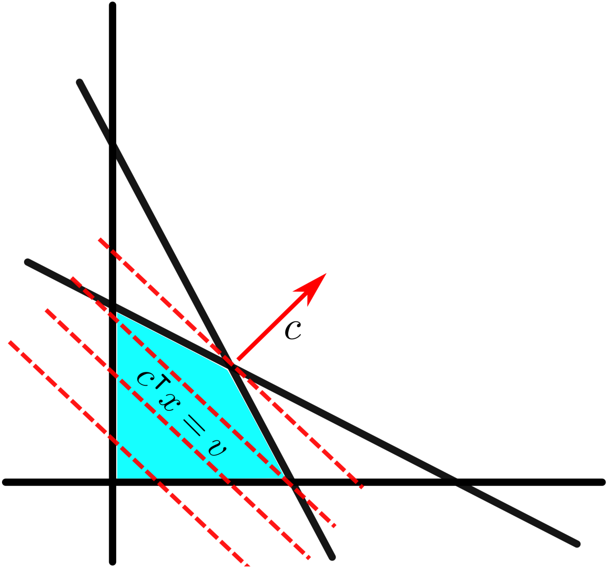
Every LP can be transformed to this form

- minimizing $c^\intercal x$ is equivalent to maximizing $-c^\intercal x$
- $\geq$ constraints can be flipped by multiplying by -1
- Each equality constraint can be replaced by two inequalities
- Uconstrained variable x_j can be replaced by $x_j^+ - x_j^-$, where both x_j^+ and x_j^- are constrained to be nonnegative.

Geometric View

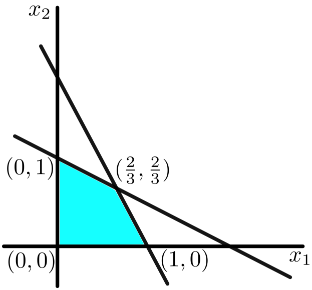


Geometric View



A 2-D example

$$
\begin{array}{ll}
\text{maximize} & x_1 + x_2 \\
\text{subject to} & x_1 + 2x_2 \leq 2 \\
& 2x_1 + x_2 \leq 2 \\
& x_1, x_2 \geq 0
\end{array}
$$



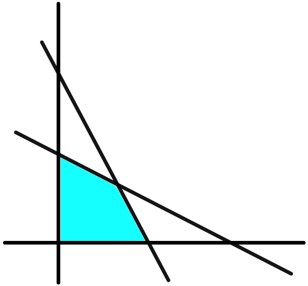
Application: Optimal Production

- 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

$$\begin{array}{lll} \text{maximize} & c^\intercal x & \\ \text{subject to} & a_i^\intercal x \leq b_i, & \text{for } i = 1, \ldots, m. \\ & x_j \geq 0, & \text{for } j = 1, \ldots, n. \end{array}$$

Terminology

- Hyperplane: The region defined by a linear equality
- Halfspace: The region defined by a linear inequality $a_i^\intercal x \leq b_i$.
- 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 $\nexists y \neq 0$ 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 $c^\intercal x = OPT$

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

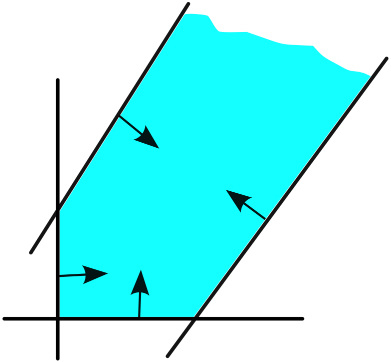
- In fact, a face of the polytope
- intersection of P with hyperplane $c^\intercal x = OPT$

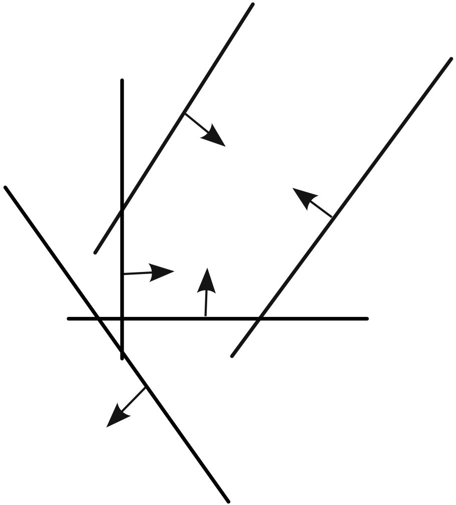
Fact

At a vertex, n linearly independent constraints are satisfied with equality (a.k.a. tight)

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.

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. $c^\intercal y = 0$, and $a_i^\intercal y = 0$ whenever the i'th constraint is tight for x.
- Can choose y s.t. $y_j < 0$ for some j
- Let α be the largest constant such that $x + \alpha y$ is feasible
  - Such an α exists
- An additional constraint becomes tight at $x + \alpha y$, a contradiction.

Counting non-zero Variables

Corollary

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

$$\begin{array}{lll} \text{maximize} & c^\intercal x & \\ \text{subject to} & a_i^\intercal x \le b_i, & \text{for } i = 1, \ldots, m. \\ & x_j \ge 0, & \text{for } j = 1, \ldots, n. \end{array}$$

- 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