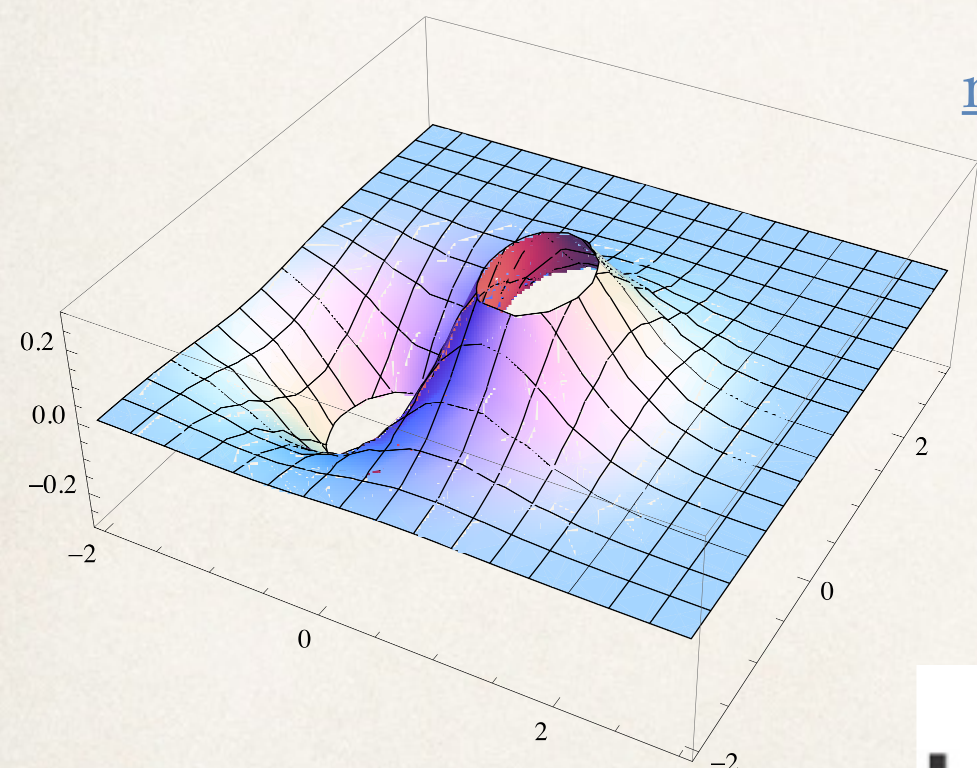
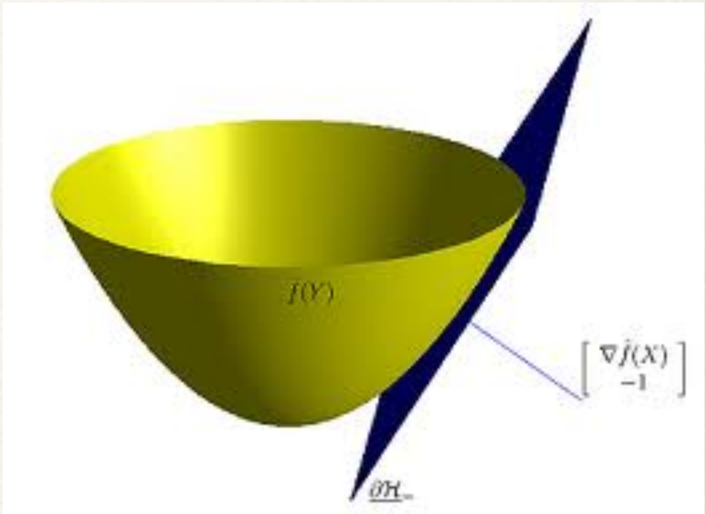
690-OP: Optimization for Computer Science

Sridhar Mahadevan

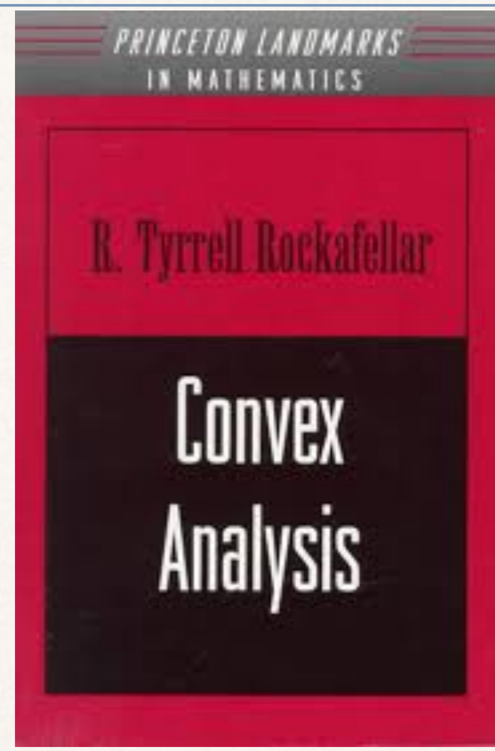
mahadeva@cs.umass.edu

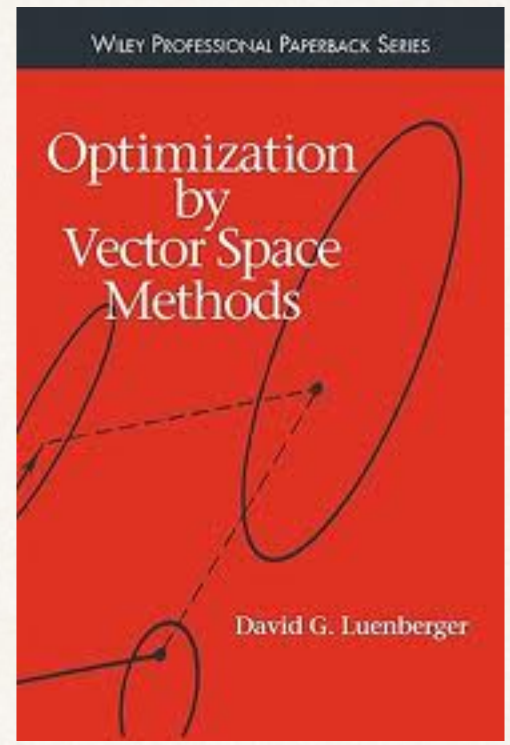


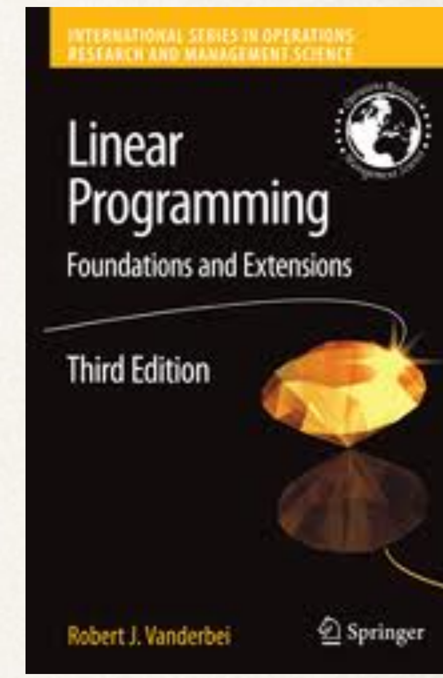


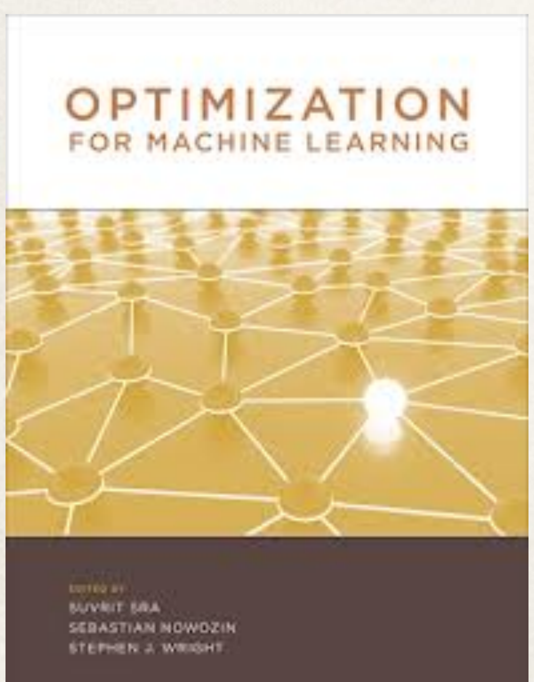
UMASS CS
SCHOOL OF COMPUTER SCIENCE

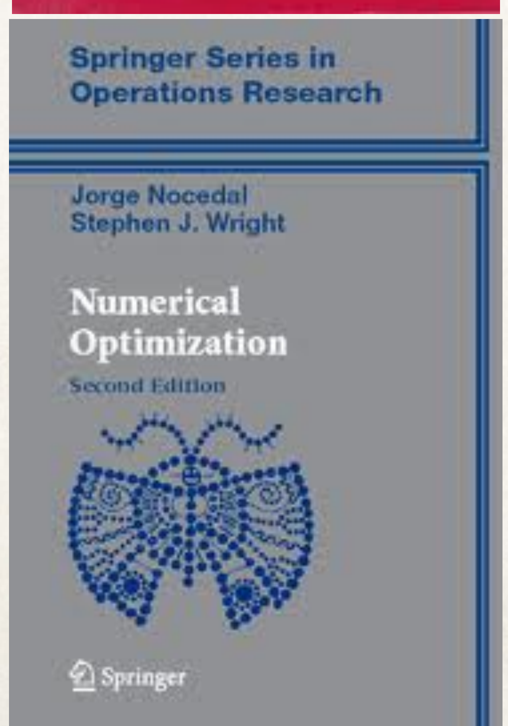
Some References

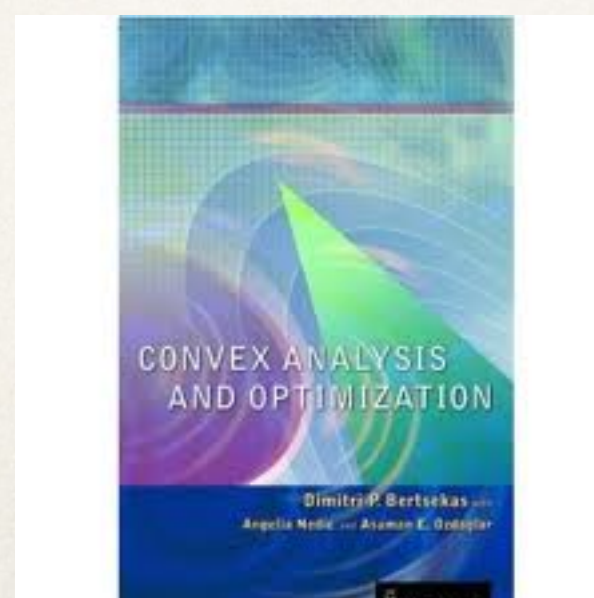


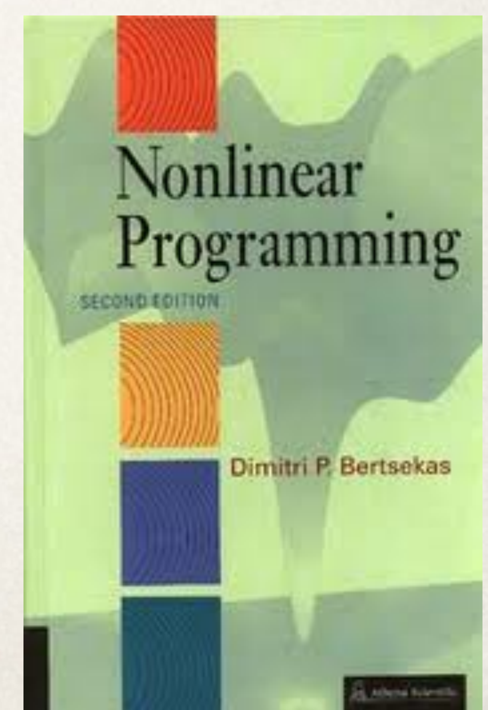












Overview

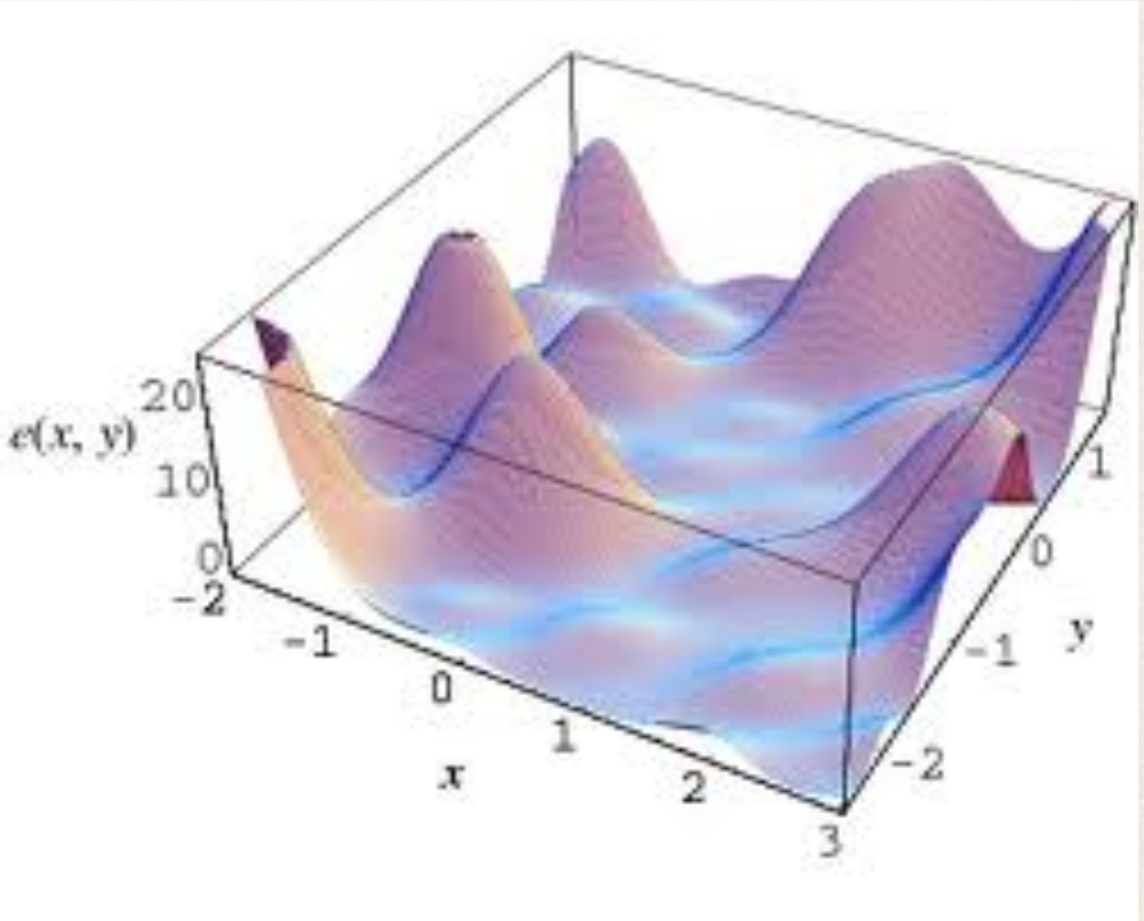
- In this lecture, we provide an overview of the course
  - What is optimization, and why is it useful?
  - What will we broadly cover?
- Reading material and background
- Course organization and evaluation
  - Grader: Stefan Dernbach (dernbach@cs.umass.edu)

Definition

- **Definition of OPTIMIZATION**

: an act, process, or methodology of making something (as a design, system, or decision) as fully perfect, functional, or effective as possible;

specifically **:** the mathematical procedures (as finding the maximum of a function) involved in this



Major Applications in CS, Engineering, and Science

- Resource allocation problems
- Control, planning, and decision-making
- Approximation and Estimation, machine learning, data mining
- Game theory, economics, network optimization

Major Topics

- Convex optimization over convex sets
- Optimization over manifolds
- Non-convex optimization in high-dimensional spaces
- Generalizations of optimization: variational inequalities

Optimization Problems

Game Theory

Linear Programming

Engineering

$$\min_x c^T x$$

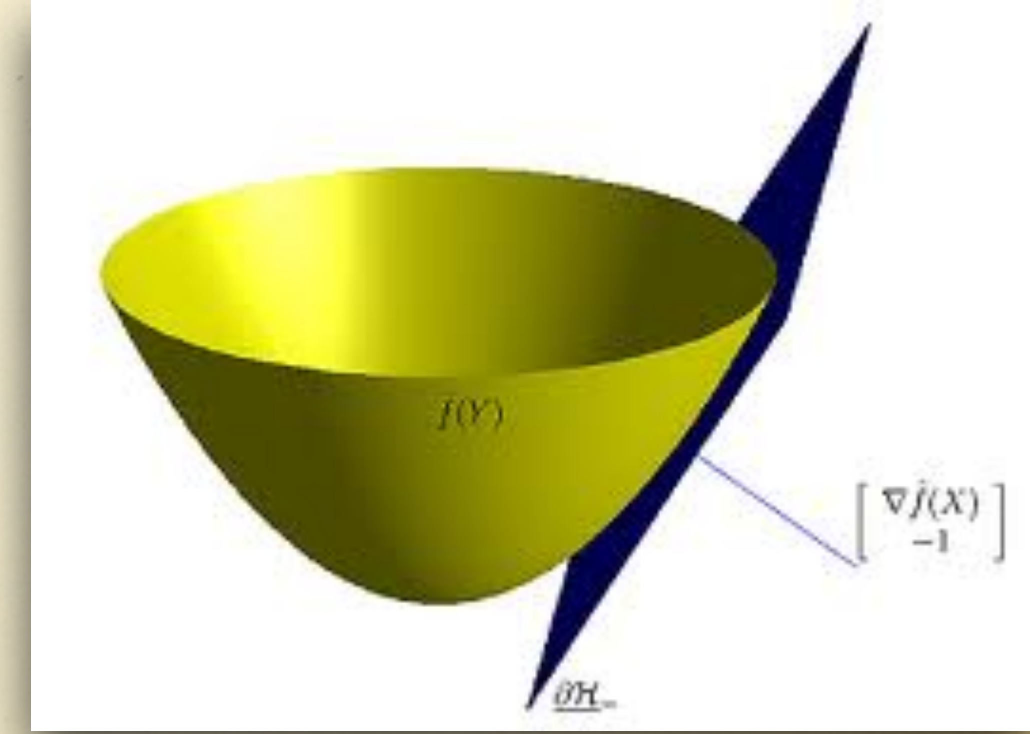
$$\text{s.t. } Ax = b, x \geq 0$$

ML

Business

Convex Optimization

$x^* = \text{argmin}_x f(x)$ such that $x \in \mathcal{K}$



- If f is a convex function, x* is its unique minimum whenever

$$f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle, \ \forall x \in K$$

Optimization: Machine Learning

"Sparse" Supervised learning

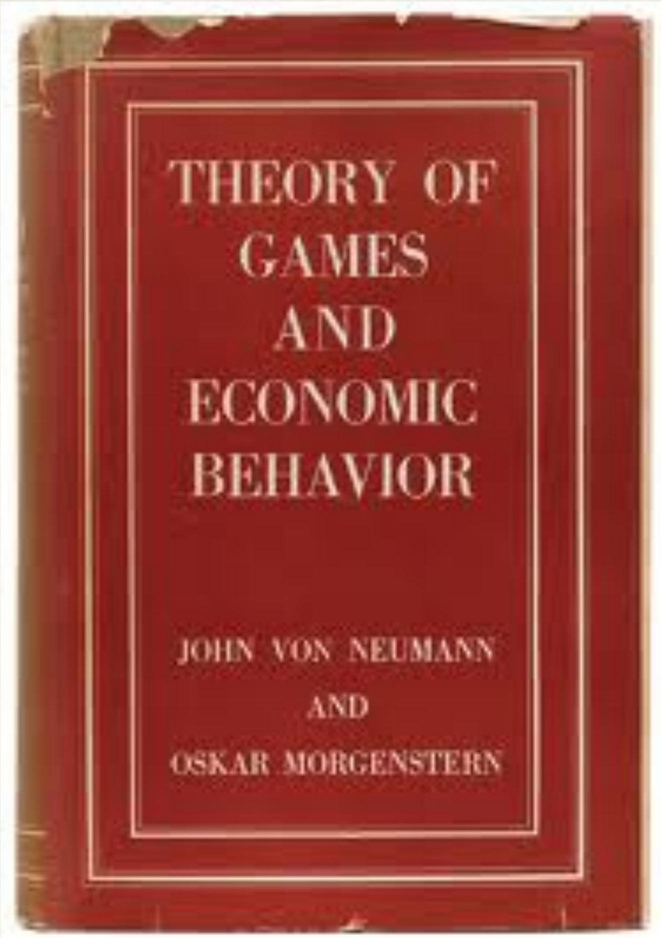
Lasso: $$\min_{x \in X} f(x) + g(x) : \min_{\beta \in \mathbb{R}^k} \|X\beta - y\|_2^2 + \lambda \|\beta\|_1$$

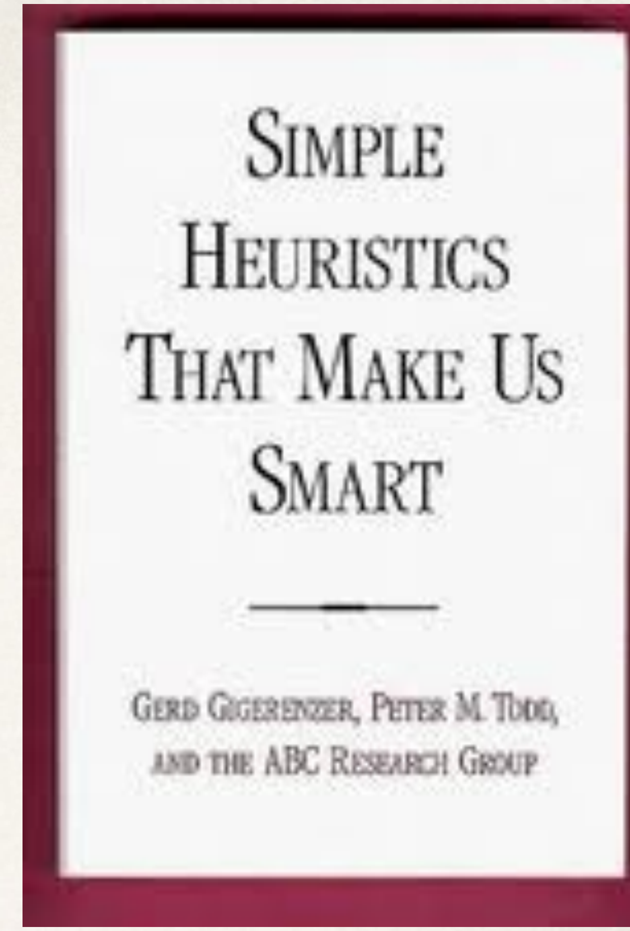
RO-TD: "Saddle Point" Reinforcement Learning

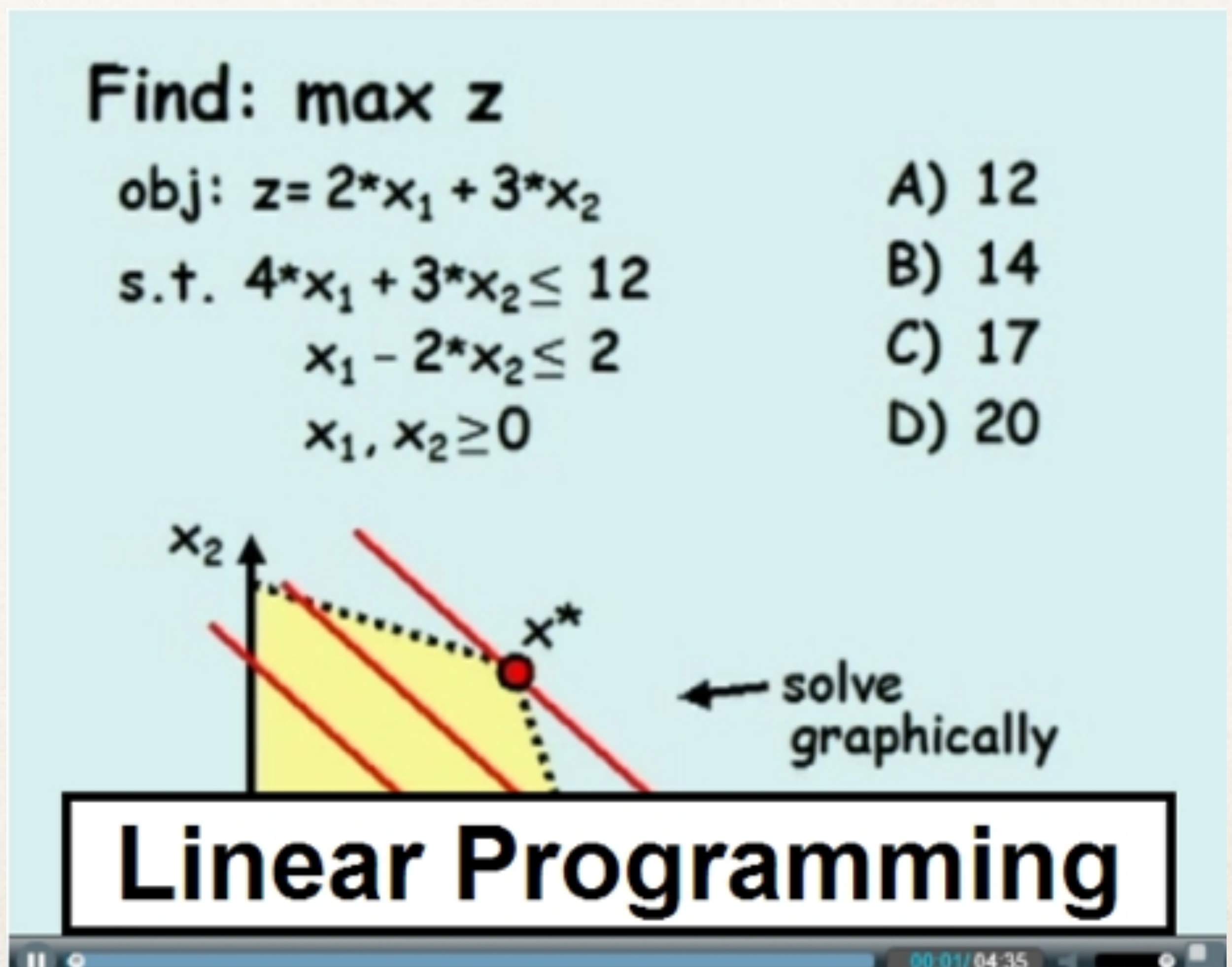
$$\min_x \|Ax - b\|_m + h(x) = \min_x \max_{\|y\|_n \leq 1} y^T (Ax - b) + h(x)$$

Unsupervised learning

Low-rank embedding: $$\min_R \frac{1}{2} ||X - XR||_F^2 + \lambda ||R||_*$$


THEORY OF GAMES AND ECONOMIC BEHAVIOR
JOHN VON NEUMANN
AND
OSKAR MORGENSTERN


SIMPLE HEURISTICS THAT MAKE US SMART
GERD GIGERENZER, PETER M. TODD,
AND THE ABC RESEARCH GROUP


Find: max z
obj: z = 2*x1 + 3*x2
s.t. 4*x1 + 3*x2 ≤ 12
x1 - 2*x2 ≤ 2
x1, x2 ≥ 0
A) 12
B) 14
C) 17
D) 20
x2
x*
solve graphically
Linear Programming
00:01/04:35

Example ~ Vacation in Europe!

Italy

England

France

How to plan a vacation to Europe given a fixed budget?

Optimization Problem: Europe

- Given a fixed budget M, design a travel vacation in Europe that visits S cities, using route R, in an "optimal" way
  - Which cities to place in S?
  - How to find the cheapest route R?
  - Does the problem even have a feasible solution?

Feasible Solution: European Vacation

- Suppose budget M = $10
  - No feasible solution (or stay at home and watch a movie about Europe on Netflix!)
- Budget M = $100
  - Still tough!
  - Buy lots of Powerball lottery tickets?
- Budget M = $1000
  - Looks feasible
  - Perhaps $500 for airfare
  - Remainder for hotels / trains
- Budget M = $5000
  - Looks better...

One Week in Paris...ooh la la!

Louvre Museum

Seine river

Cafe au lait!

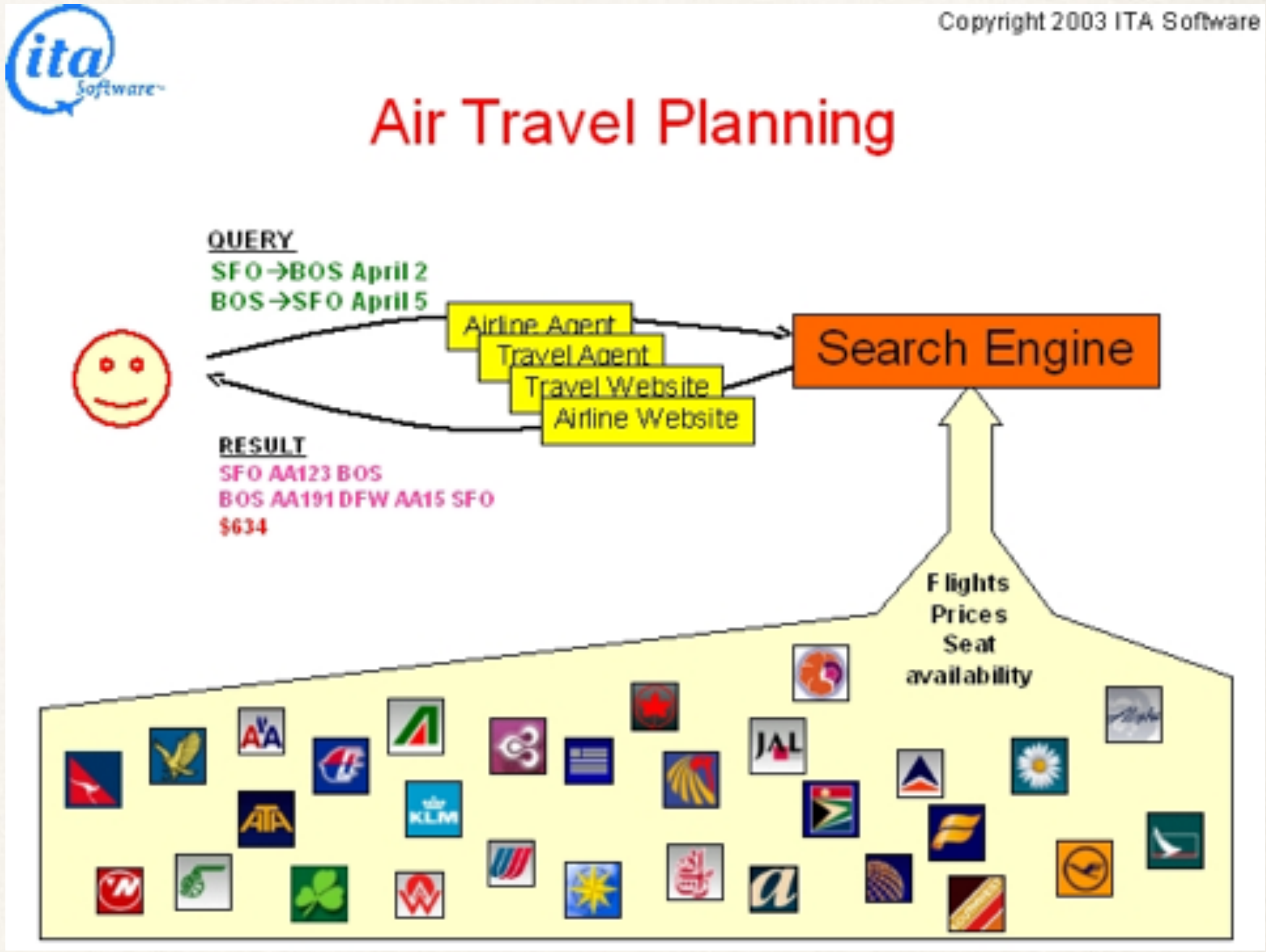
Mona Lisa

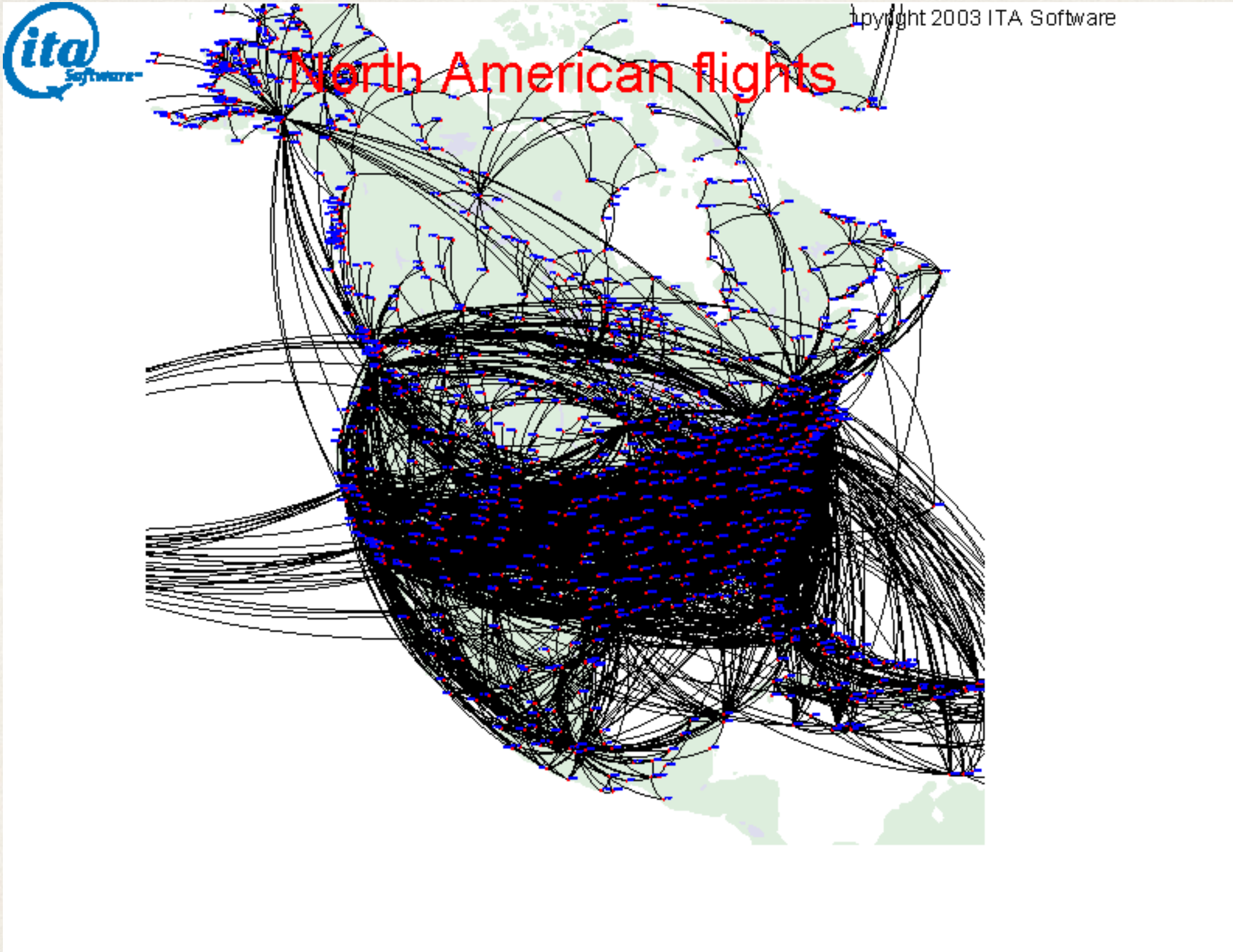
European Vacation: Optimization

- How to select cities S to visit?
  - Let S = cities in England?
  - Law of diminishing returns (submodularity)
- Find the cheapest airfare
  - Use Orbitz, Expedia, Cheaptickets,....
  - Optimization problem
- Best route to travel
  - Traveling salesman problem
- What hotels to stay in?
  - Youth hostels, bed / breakfast, Univ. dorms

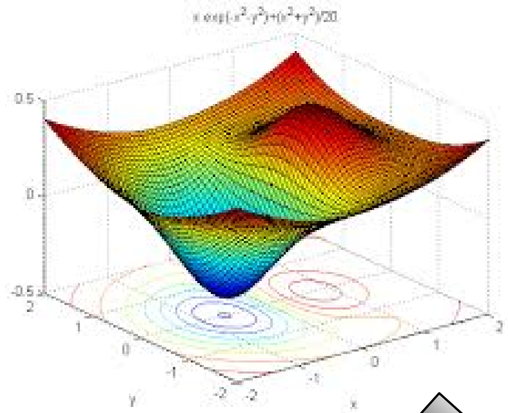
Air Travel Optimization

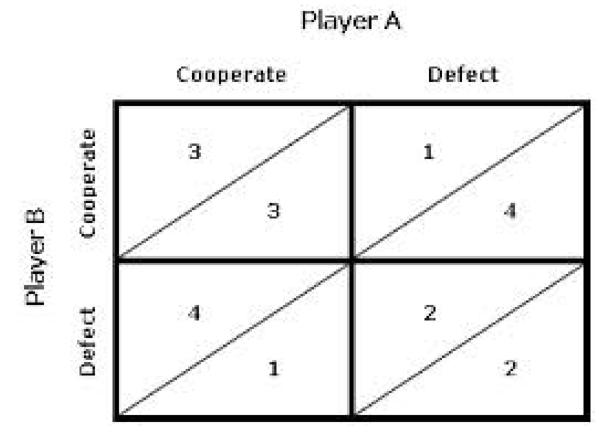
- Carl de Marcken, ITA software
  - Used by many companies (Orbitz, Expedia)
- Hard optimization problem
- http://www.demarcken.org/carl/




ita Software
North American flights
opyright 2003 ITA Software

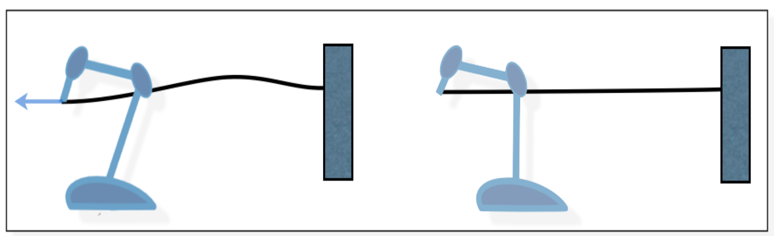
Optimization

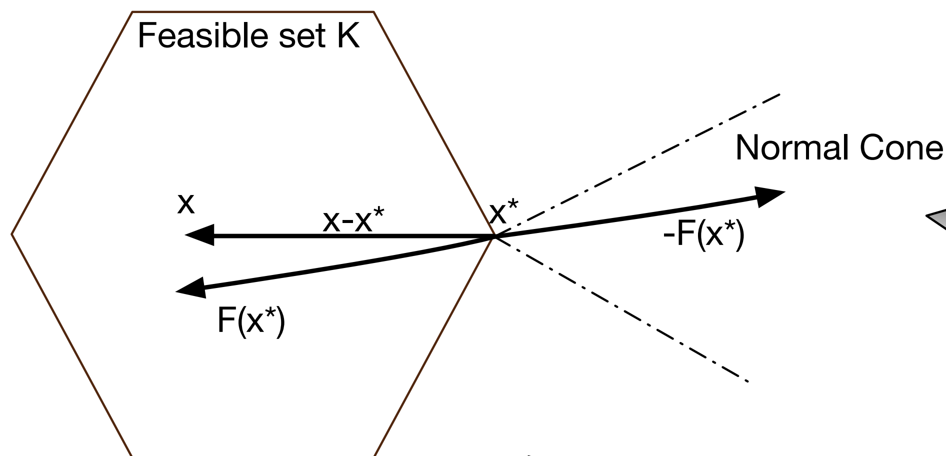




Game theory

Complementarity problems





Variational Inequalities

$$\frac{\partial u}{\partial x_1} + \frac{\partial u}{\partial x_2} = 0 \quad \text{is linear .}$$

$$\frac{\partial u}{\partial x_1} + \left(\frac{\partial u}{\partial x_2}\right)^2 = 0 \quad \text{is nonlinear .}$$

$$\frac{\partial u}{\partial x_1} + \frac{\partial u}{\partial x_2} + u^2 = 0 \quad \text{is nonlinear .}$$

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = x_1 \quad \text{is linear .}$$

$$\frac{\partial^2 u}{\partial x_1^2} + u\frac{\partial^2 u}{\partial x_2^2} = 0 \quad \text{is quasilinear .}$$

Nonlinear equation solving

Traffic equilibrium problem

Next Generation Internet Model [Nagurney et al., 2014]

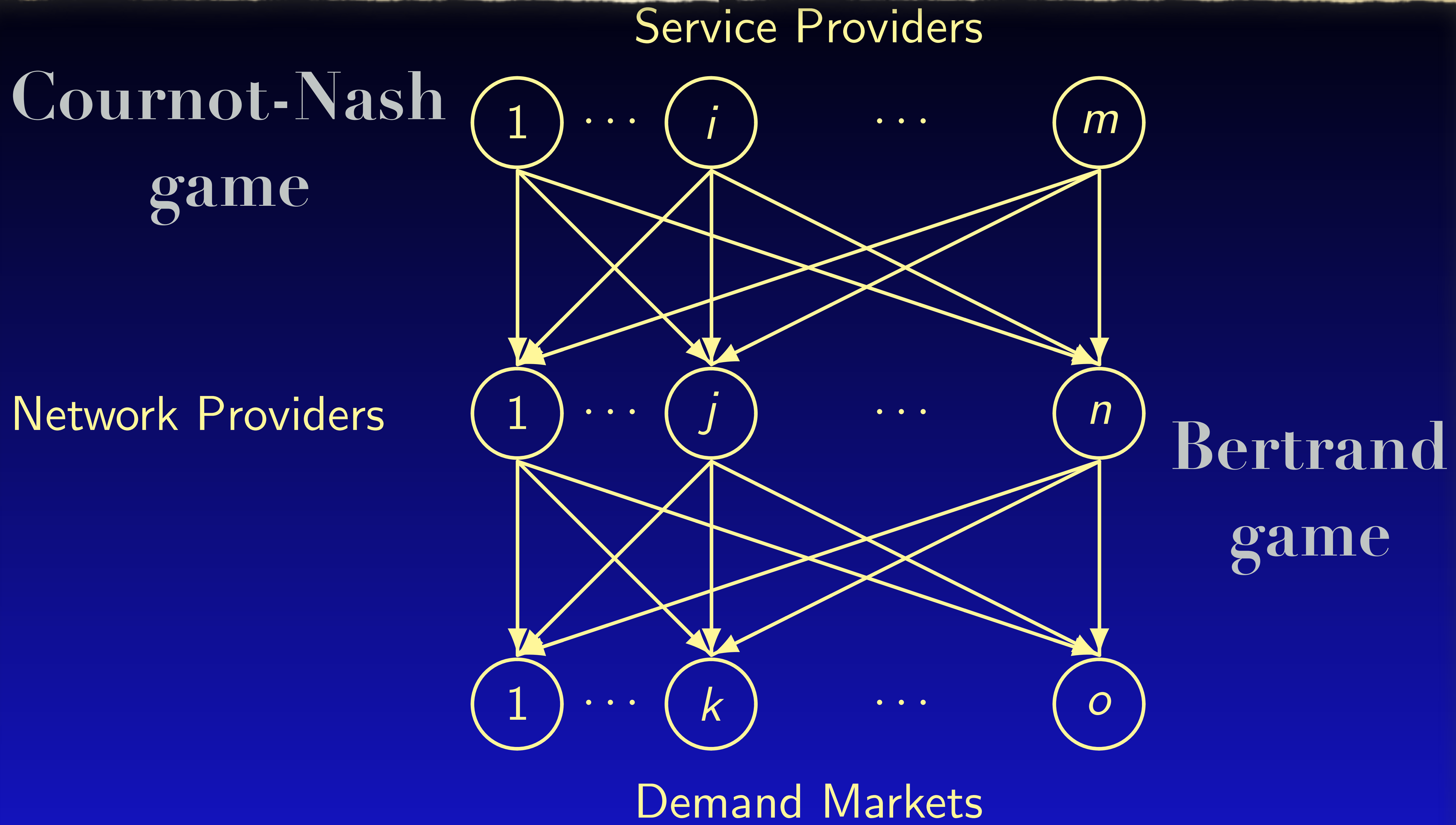
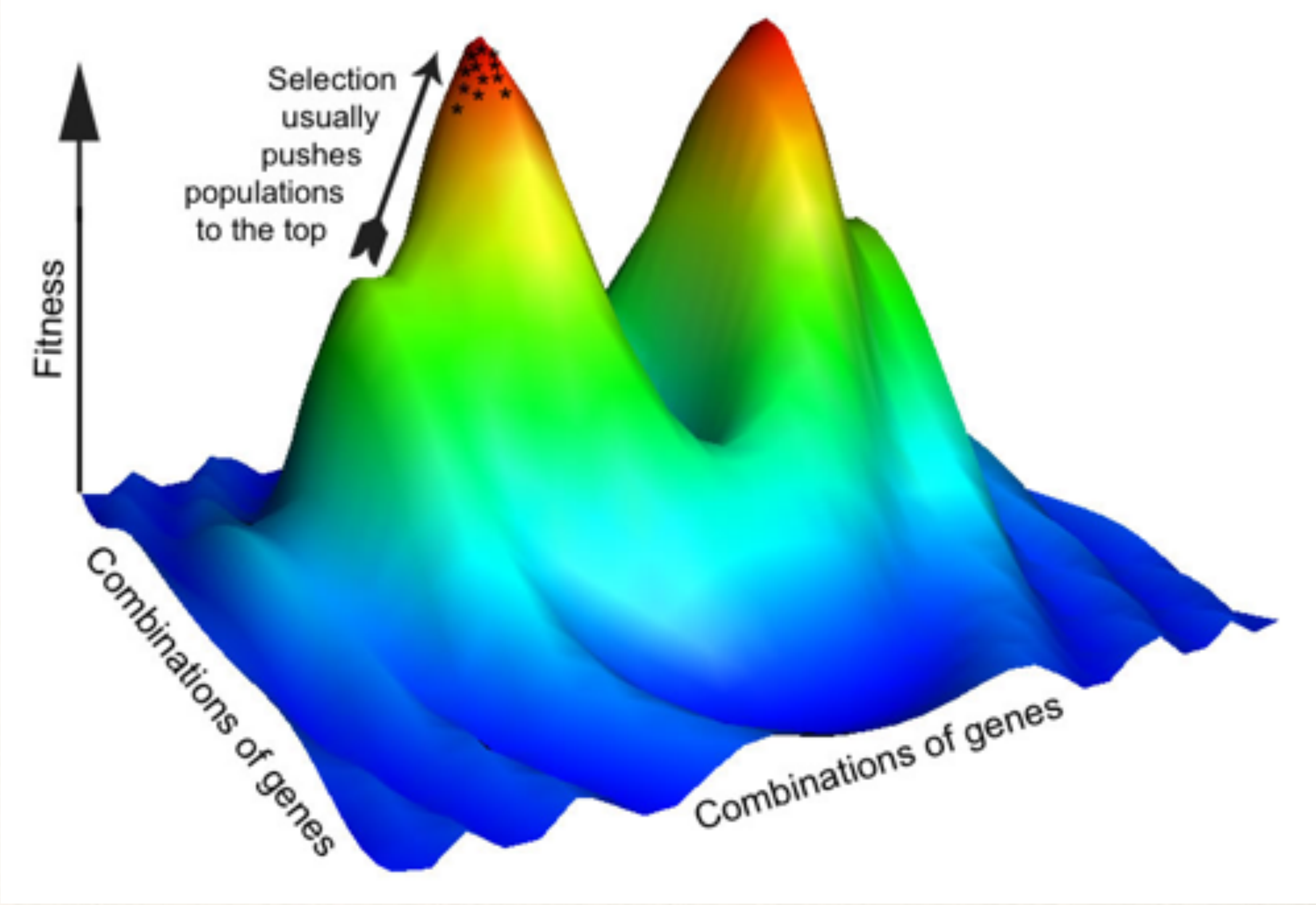


Figure 1: The Network Structure of the Cournot-Nash-Bertrand Model for a Service-Oriented Internet

Optimization in Science

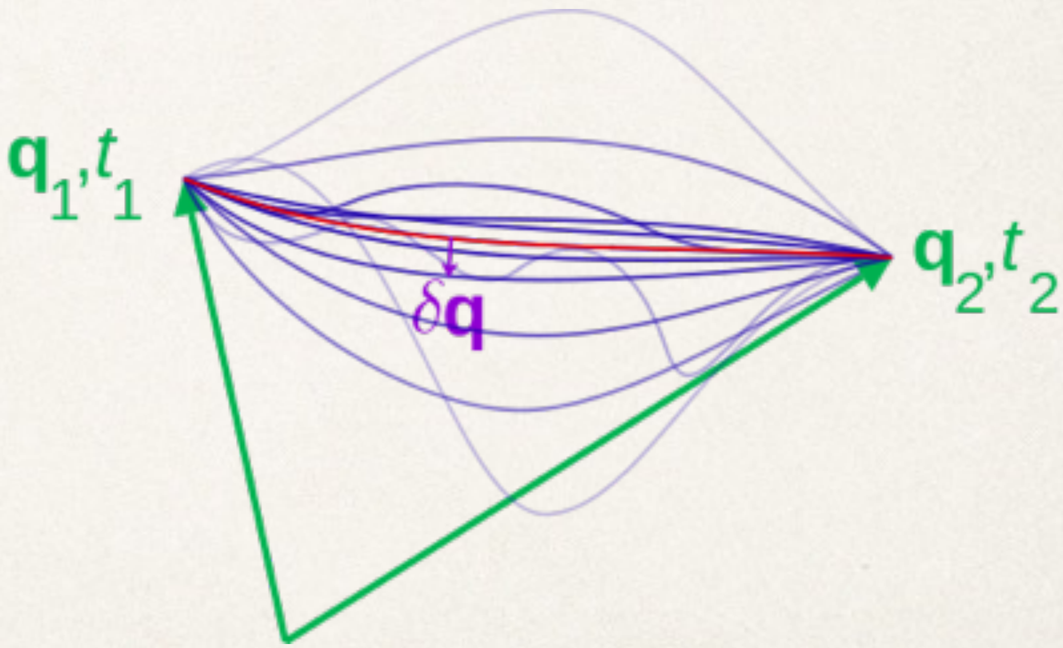
- Biology
  - Evolution can be modeled as optimization in the space of genomes (Dawkins)
  - Organisms must optimize in face of limited resources
  - Fitness landscapes in evolutionary dynamics



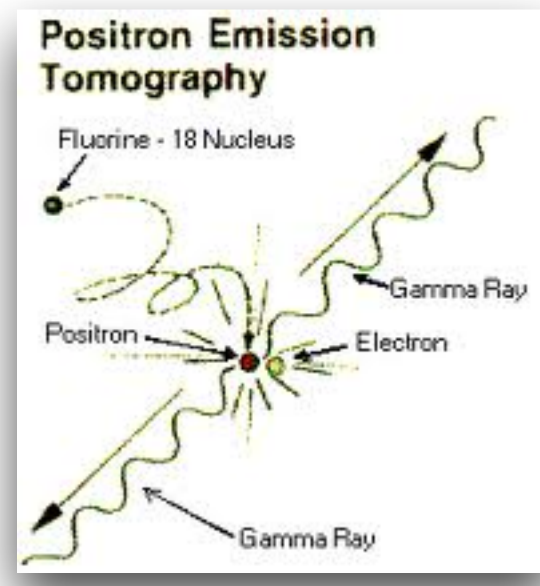
Optimization in Science

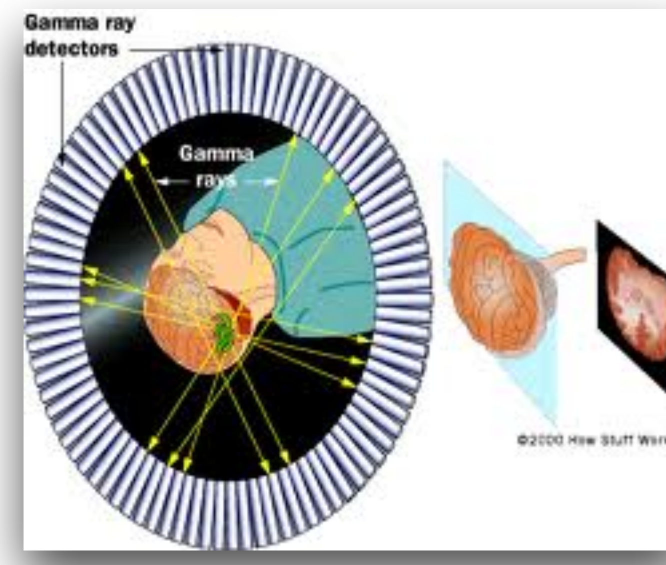
- Physics
  - Many physical laws can be formulated using optimization principles
  - Fermat: principle of least time
  - Least action principle
  - Lagrangian dynamics

$$\delta \mathcal{S} = \delta \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}}, t)\, dt = 0\,,$$

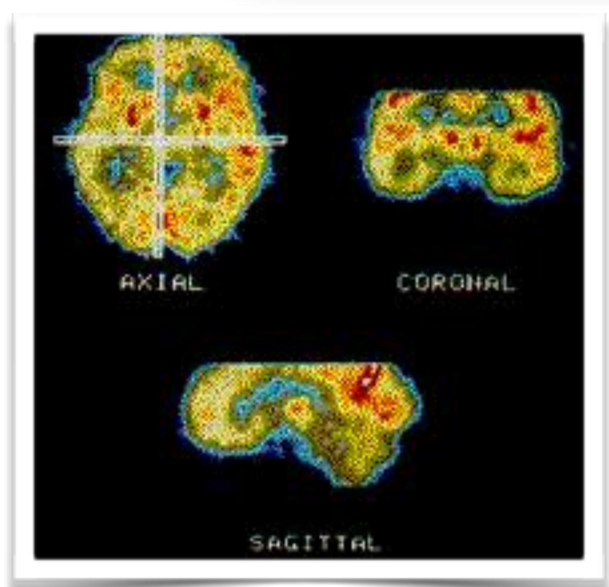


Optimization in Medicine: Positron Emission Tomography





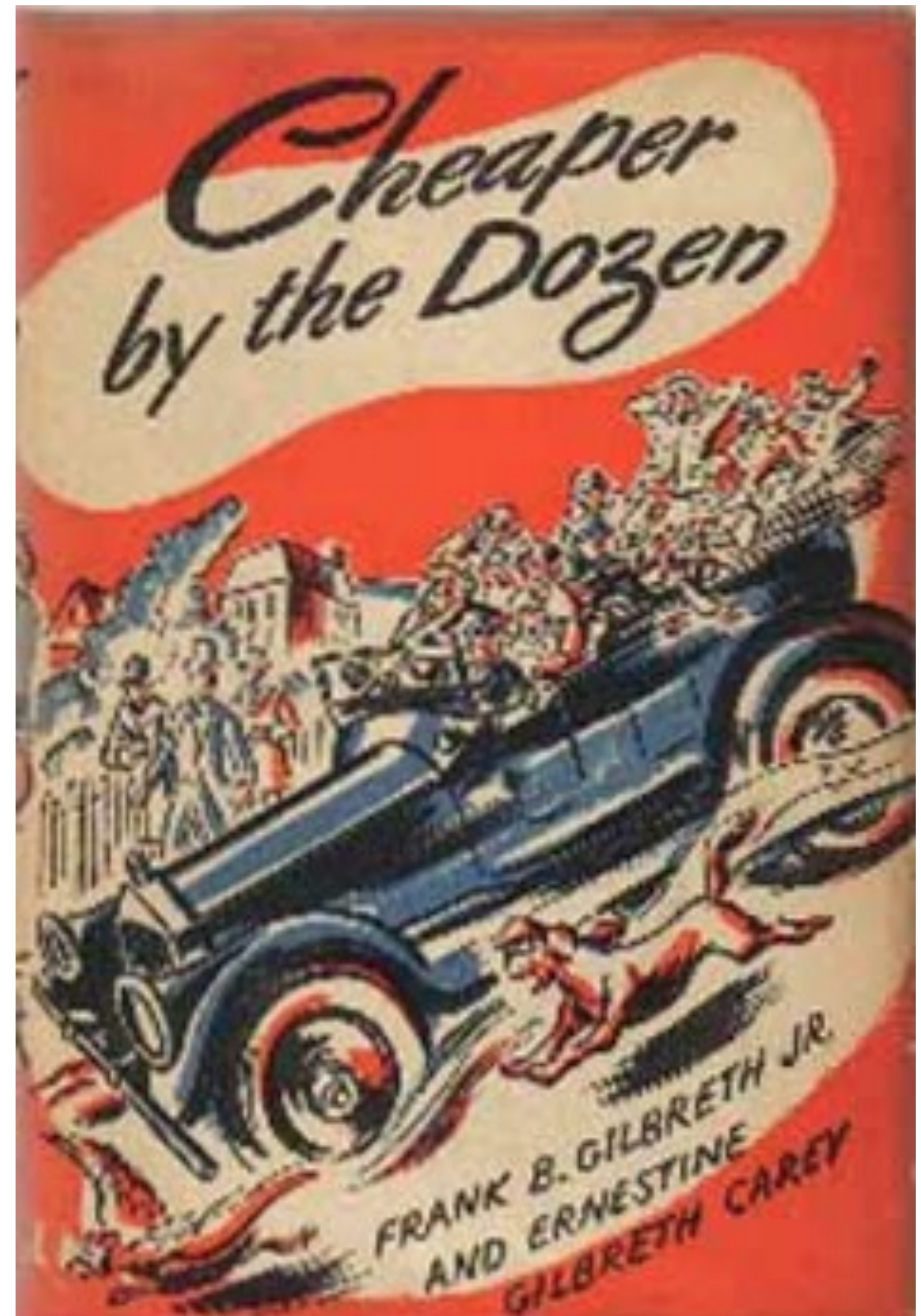
Reconstruction involves optimization problem



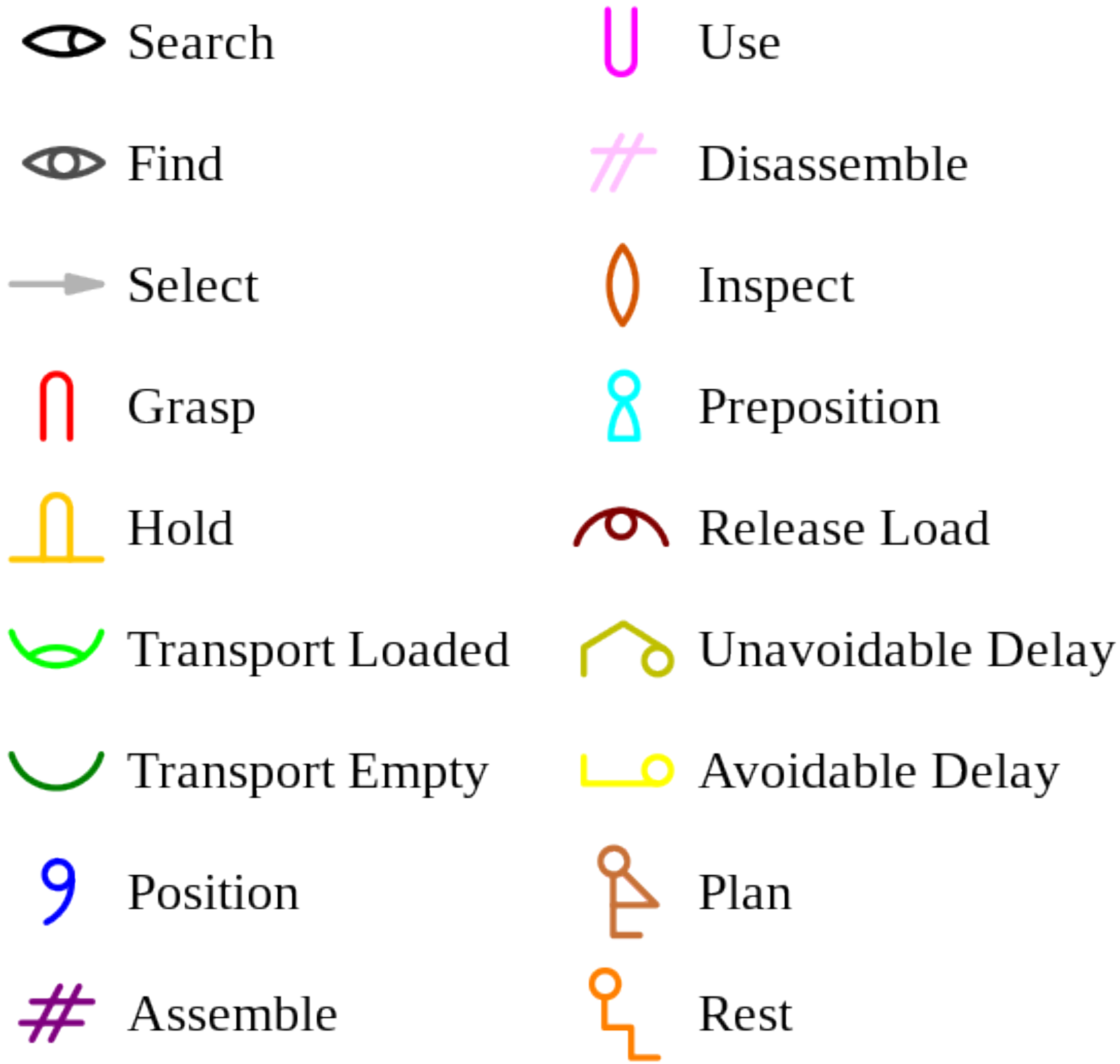
Time-Motion Optimization

http://en.wikipedia.org/wiki/Frank_Bunker_Gilbreth

Frank Bunker Gilbreth

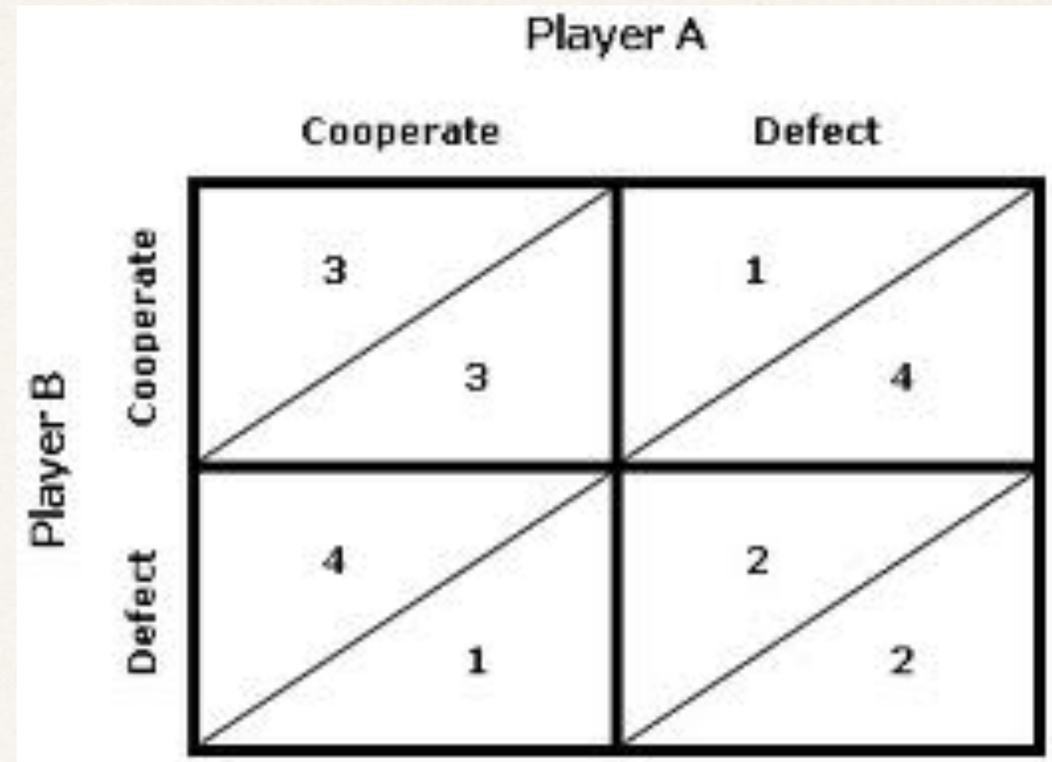


Therblig: Time/Motion Study

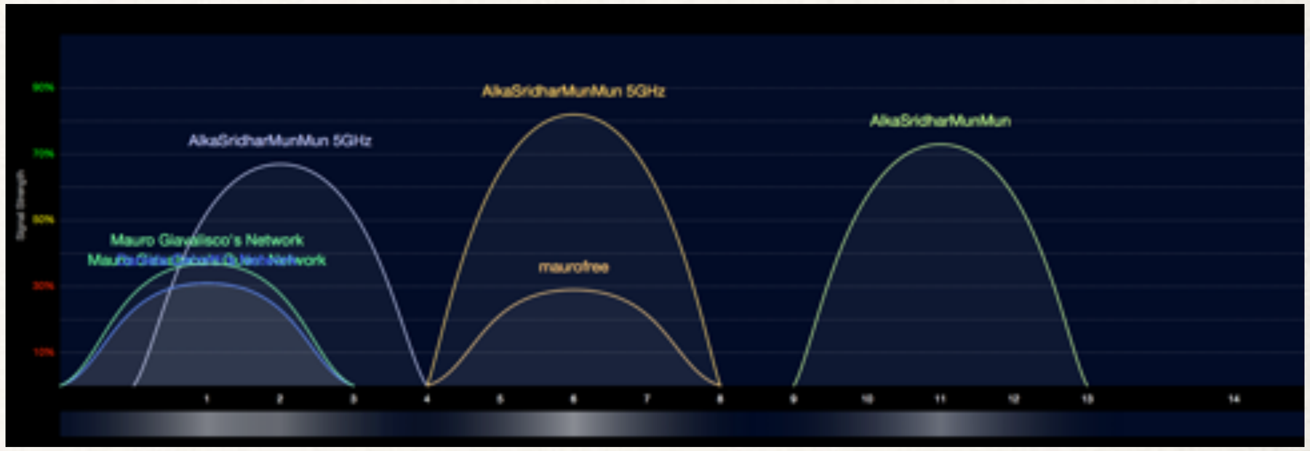


Optimization in Social Sciences

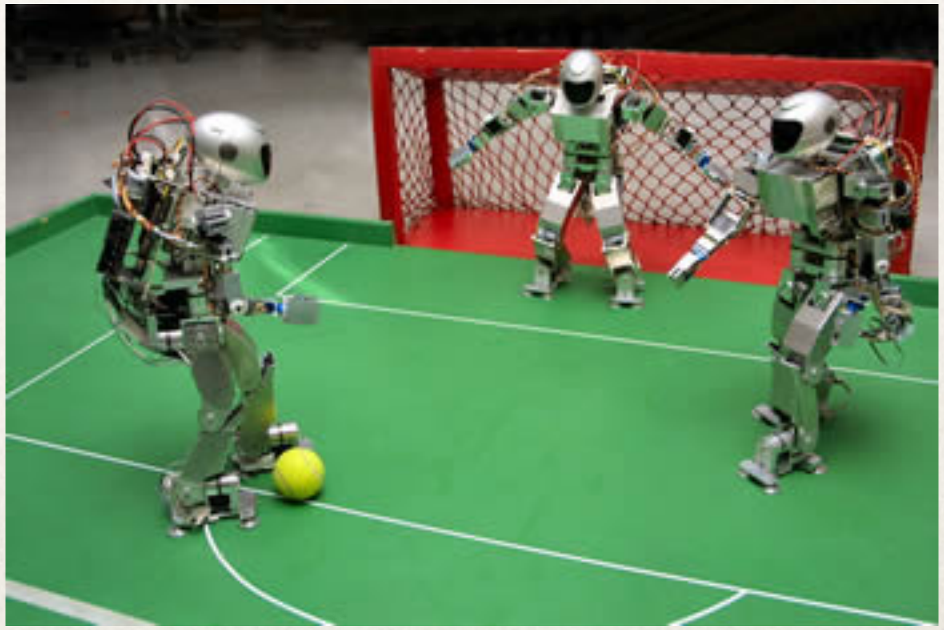
- Economics
  - Study of rational behavior
  - Maximize expected utility
- Game theory
  - Self-interested agents
- Von Neumann & Morgenstern, Nash



Optimization in Computer Sci.



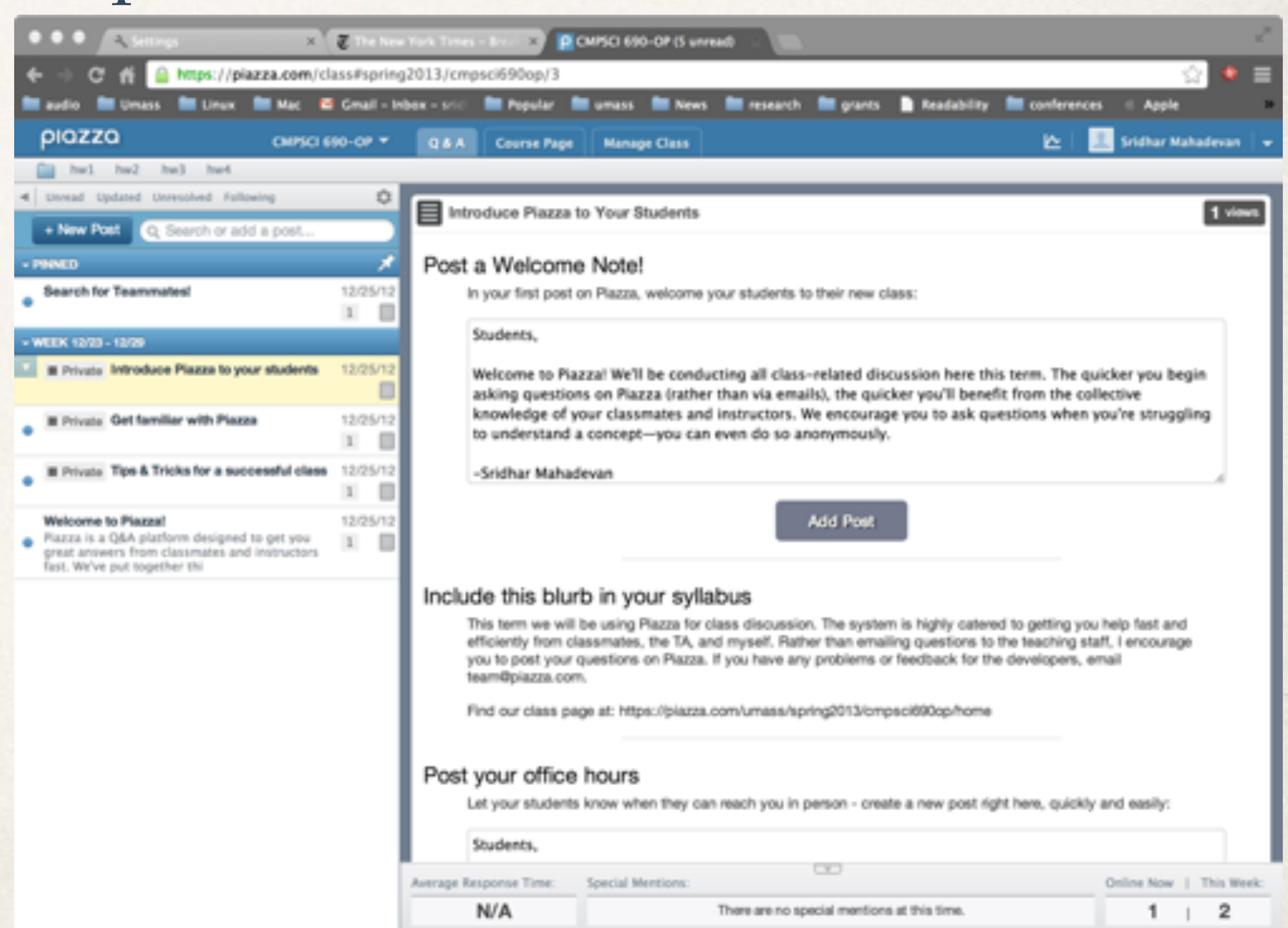
- Resource allocation
  - Wi-fi bandwidth allocation
  - Sensor network placement
- Game theory
  - Multiagent E-commerce
- Approximation & estimation: ML



Structure of the Course

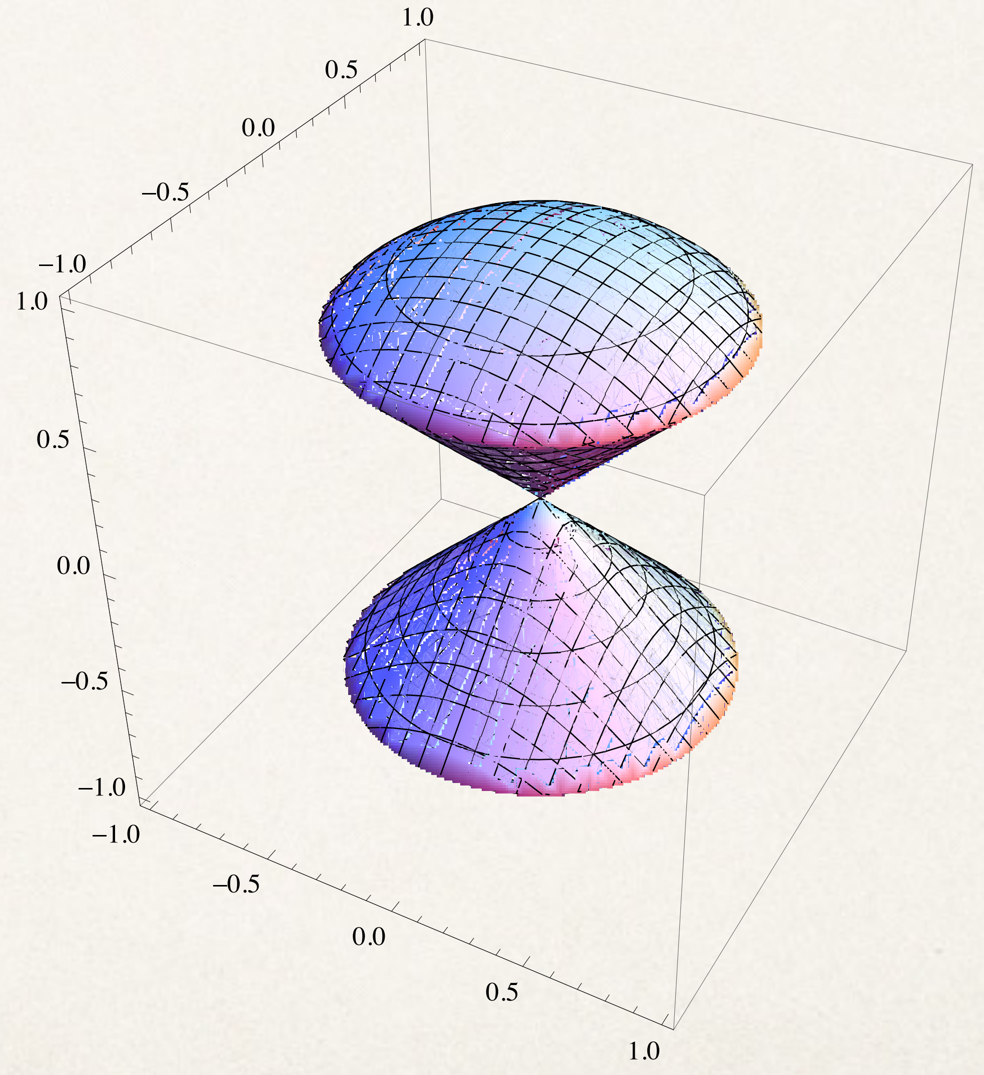
- Weekly lectures
    - Occasional guest lectures
- Moodle reading assignments
    - All students should have a Moodle account
- Moodle quizzes
- Piazza discussion forum

piazza.com discussion forum



Prerequisites

- Basic calculus
    - Know how to compute gradients, partial derivates
- Basic linear algebra
    - Matrix representations
- Mathematical maturity



Old Chinese Saying…

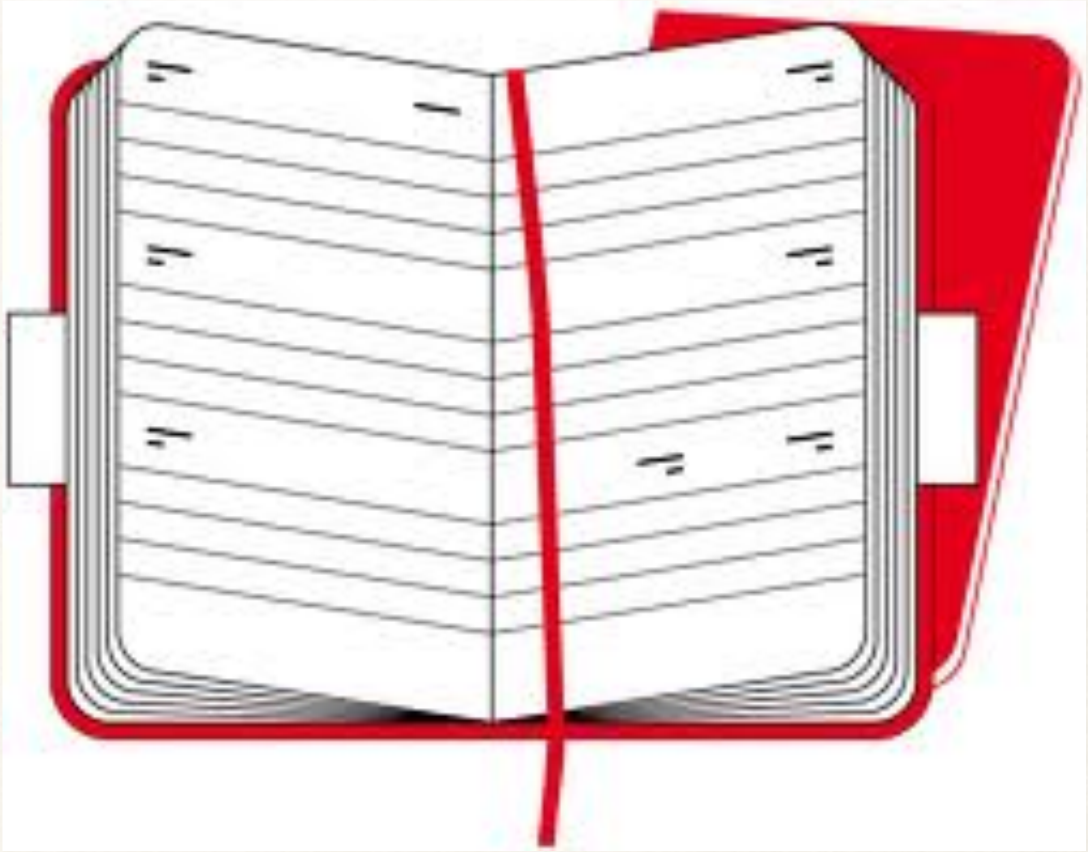
- I hear, I forget
- I see, I remember
- **I do, I understand**

Moodle quizzes

- Designed to test grasp of lecture material
- Examples
  - If A and B are two convex sets, is their intersection convex?
  - What is the sub differential of the indicator function of a convex set?
- There will be 5 quizzes in all.
  - In grading, we will record your performance on the best 4 scores.

Journal of Activity

- Registered students must maintain weekly journal
  - Problems attempted and solved
  - Algorithms implemented
  - Reading assignments
  - Papers read



Midterm Project

- An individual midterm project is required
- Students will be assigned a real-world optimization problem
- This problem will require developing some optimization code and testing it

Final Project

- A final (group) project is also required
- Students are expected to present their final projects at the end

Course Grading

- Quizzes
  - 40%
- Class participation, Self-assessment (journal)
  - 20%
- Midterm group project
  - 20%
- Final group project
  - 20%

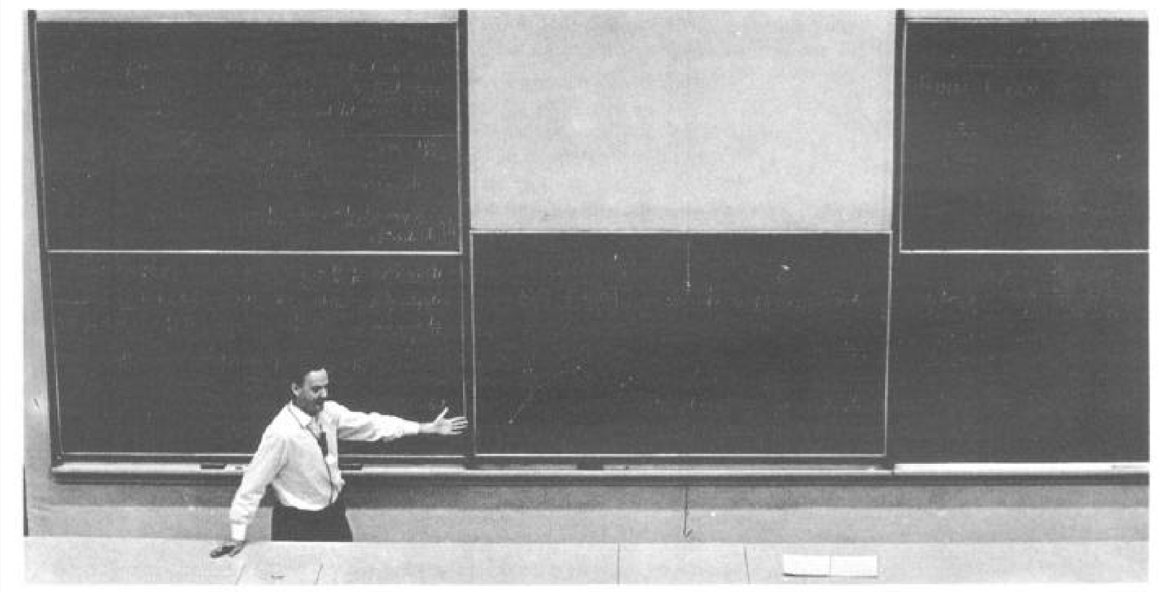
Grading will be ``seminar" style

Reading Assignments

- Log in to Moodle to download lectures
- You will also find suggestions for weekly reading
  - Principle of least action in physics
  - Mathematics and economics

19

The Principle of Least Action



A special lecture—almost verbatim*

Mathematical Methods of Economics

Joel Franklin
California Institute of Technology, Pasadena, California 91125

The American Mathematical Monthly, April 1983, Volume 90, Number 4, pp. 229–244.

When Dr. Golomb and Dr. Bergquist asked me to give a talk on *economics*, my first impulse was to try to get out of it.

"Sol," I said, "I'm not an economist. You know that."

"I know," said Golomb.

"If you want an economist, I can get you one," I said. "I know some excellent economists."

"No," he said, "we want a mathematician to talk about the subject to other mathematicians from their own point of view."

That made sense, and I hit on this idea: I won't try to tell you what mathematics has done for economics. Instead, I'll do the reverse: I'll tell you some things economics has done for mathematics. I'll describe some mathematical discoveries that were motivated by problems in economics, and I'll suggest to you that some of the new mathematical methods of economics might come into your own teaching and research.

Optimization Problem

$$\min_{x \in \Omega} f(x)$$

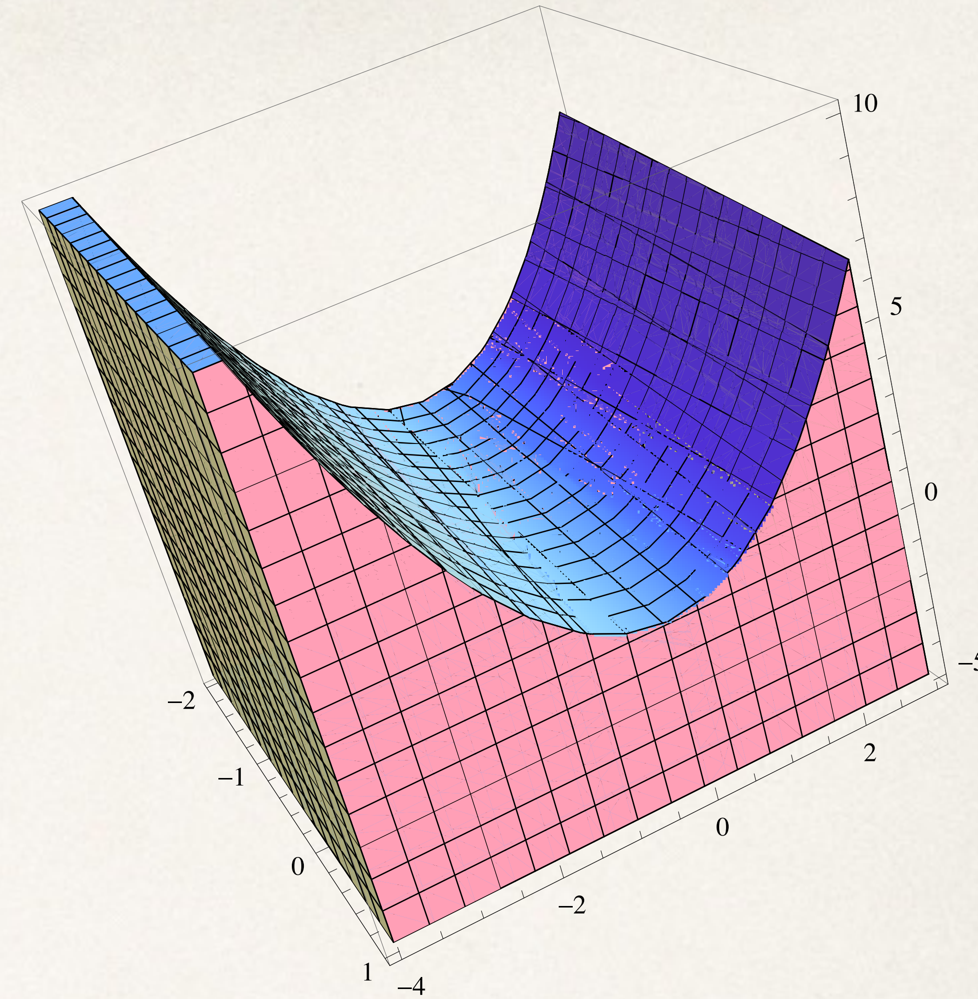
Optimization problems vary depending on the function and the constraint set

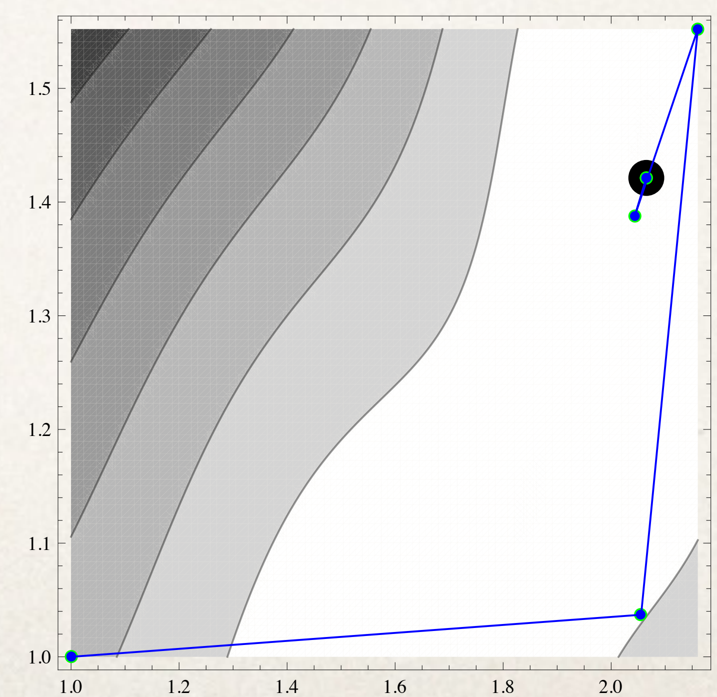
Four Unifying Principles

- Projection Theorem
- Hahn-Banach Theorem
- Duality
- Differentials

Categories

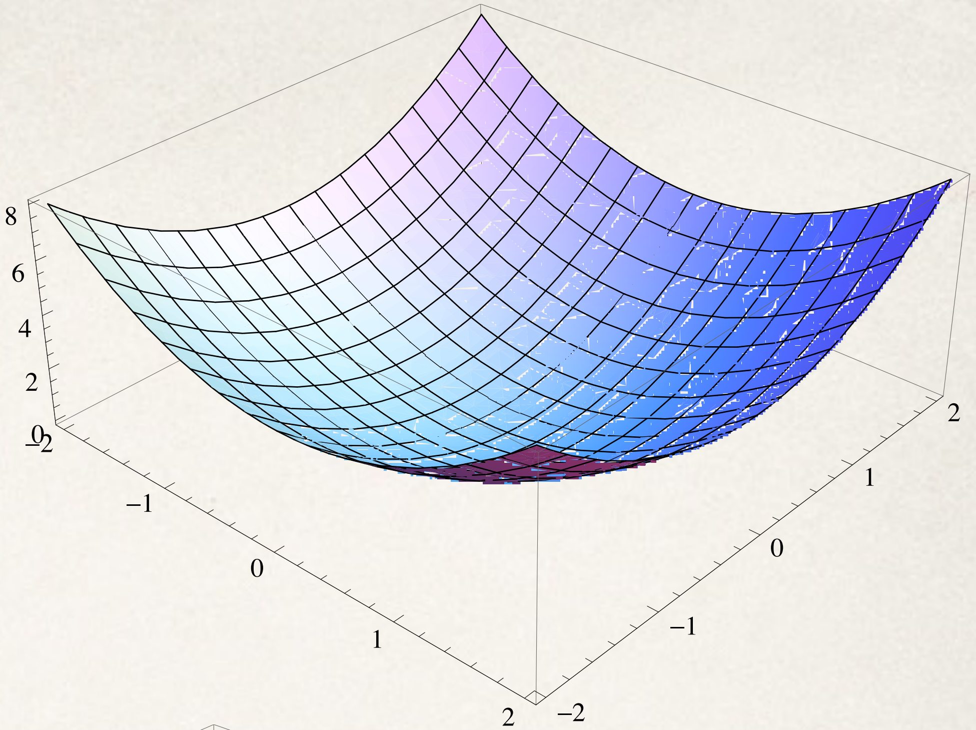
- Constrained optimization
  - Linear programming
  - Quadratic programming
  - Linear complementarity
- Unconstrained optimization
  - Global optimization
  - Local optimization

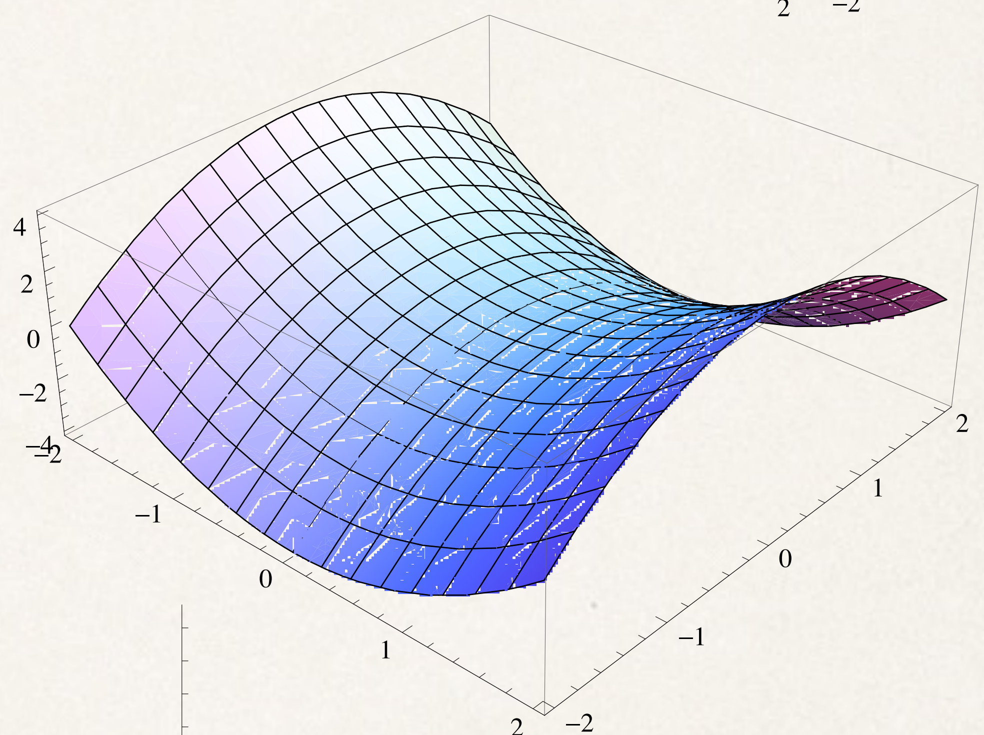


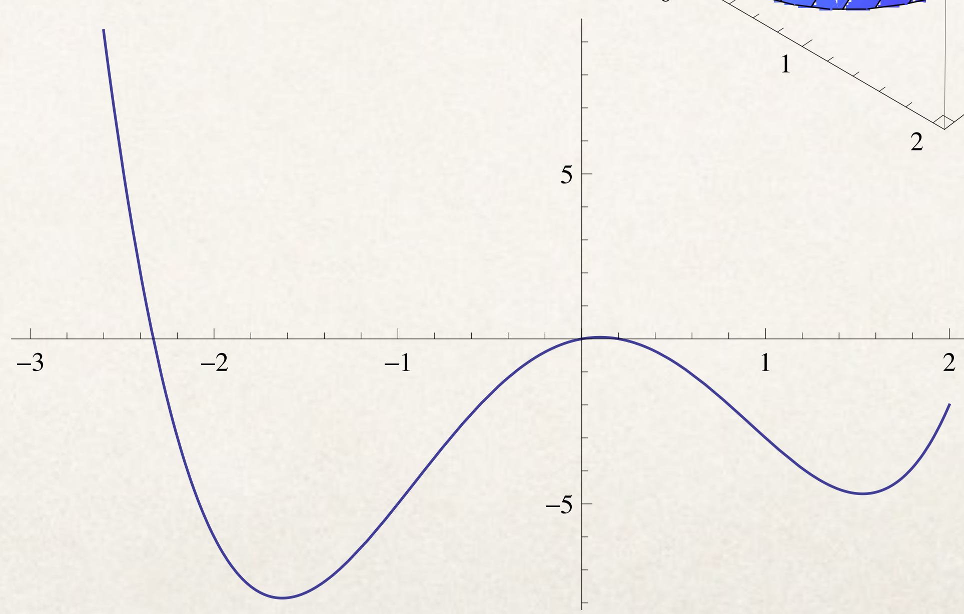


Types of Optimization

- Discrete
  - Submodularity
- Continuous
  - Smooth, Non-smooth
  - Finite-dimensional
  - Infinite-dimensional
  - Convexity

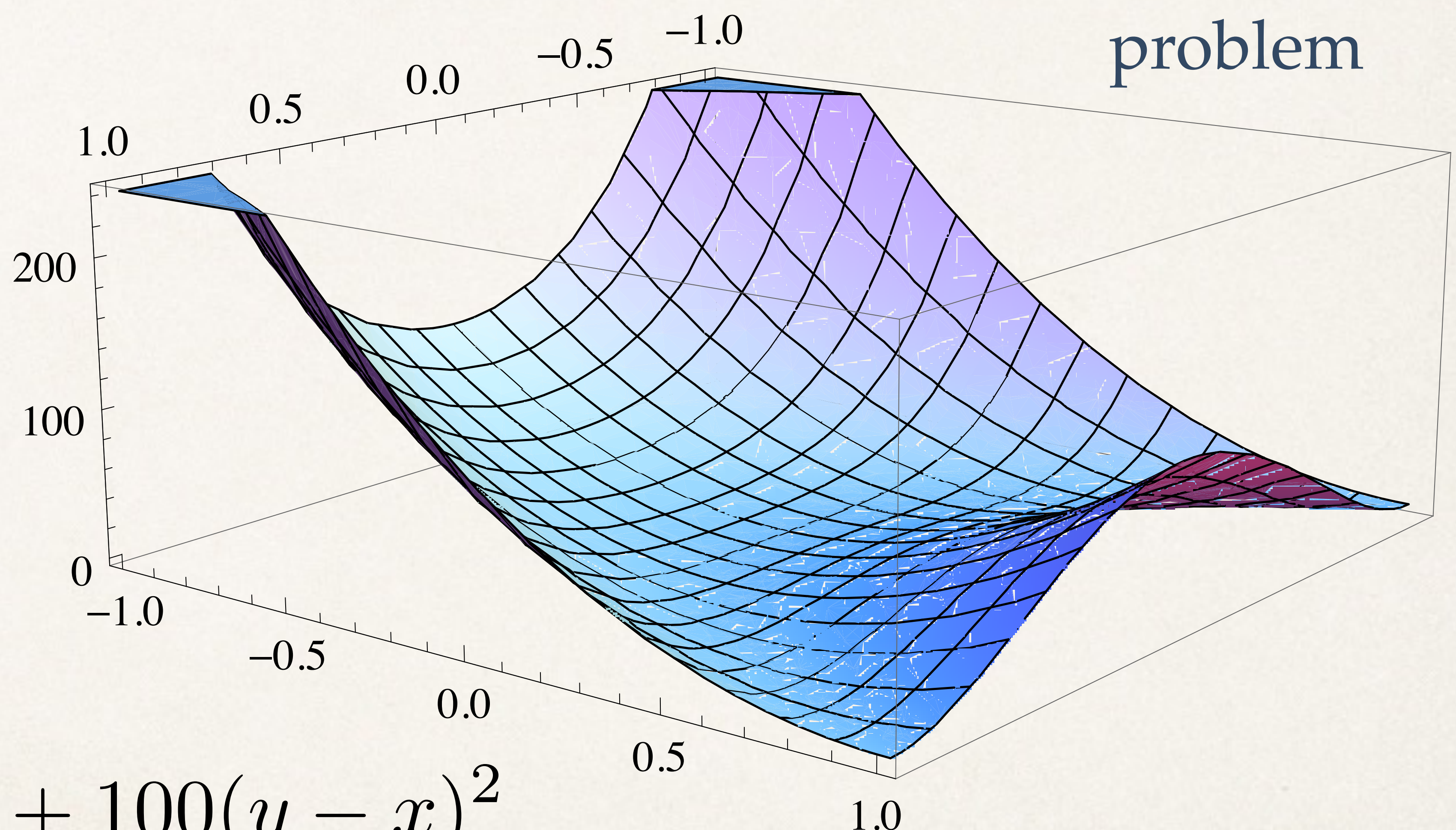






Unconstrained Optimization

Rosenbrock problem



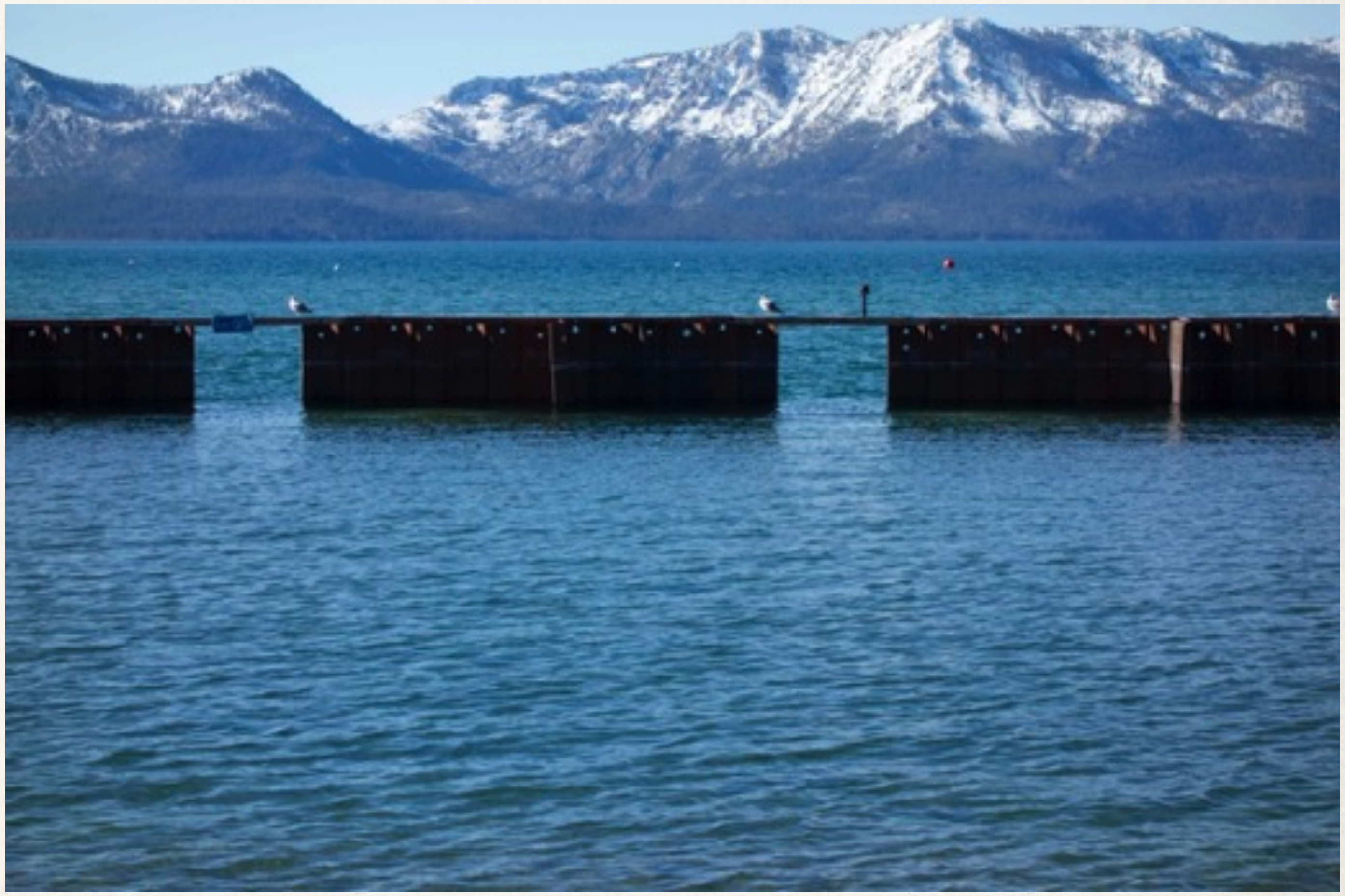
$$f(x, y) = (1 - x)^2 + 100(y - x)^2$$

Discrete Optimization

- Discrete optimization covers a large variety of CS problems
  - Integer programming
  - Bin packing
  - Combinatorial optimization
  - Max flow problem

Sensor Placement

- Suppose we want to monitor the water quality of Lake Taohe (nice skiing area)
- We place wi-fi equipped sensors that beam back measurements
- Where do we place sensors?
- Submodularity principle shows greedy method achieve near optimality



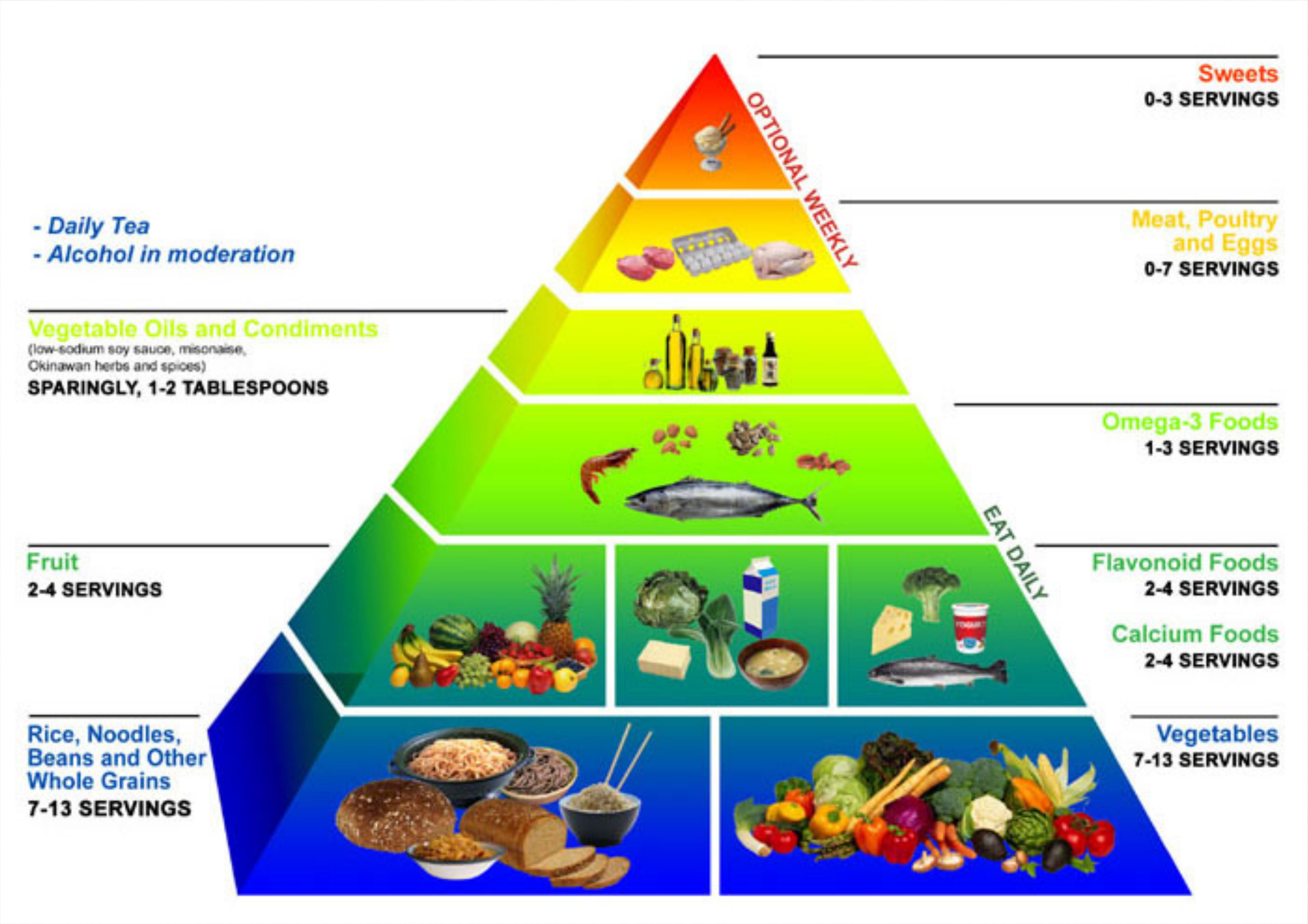
Submodularity

- Often, in discrete optimization, we want to select an optimal subset S from a set X
  - Best faculty candidates from applicant pool
  - Best sensor network locations
- Problem: maximize set-valued function f over

Submodularity property has deep ties to convexity

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$$

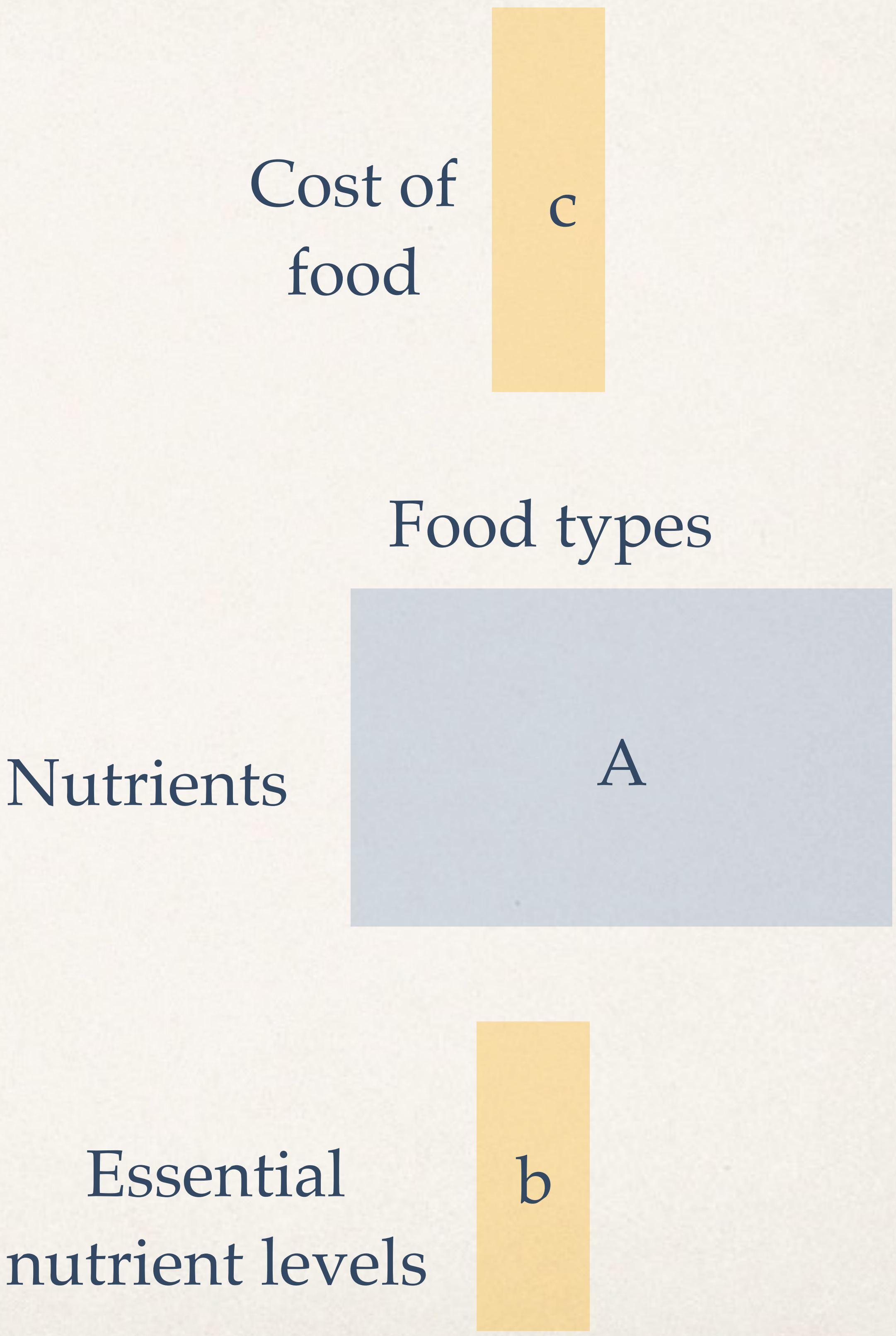
``Principle of diminishing returns”



Okinawa Diet Food Pyramid

Diet Problem

- Suppose there are M basic nutrients that are essential
- Also, there are N > M types of food that contain nutrients
- The goal is to choose quantities of food sufficient to obtain good nutrition as cheaply as possible



Optimal Assignment Problem

- Suppose there are M basic tasks that need to be done
- Also, there are N > M agents that can carry out these tasks
- The goal is to assign agents to tasks so that all tasks are carried out as cheaply as possible

Linear Programming

- The modern study of optimization began with linear programming
- The simplex algorithm was a significant breakthrough
- It is still a widely used optimization method
- Interior point methods developed in the 1980s introduced a new paradigm

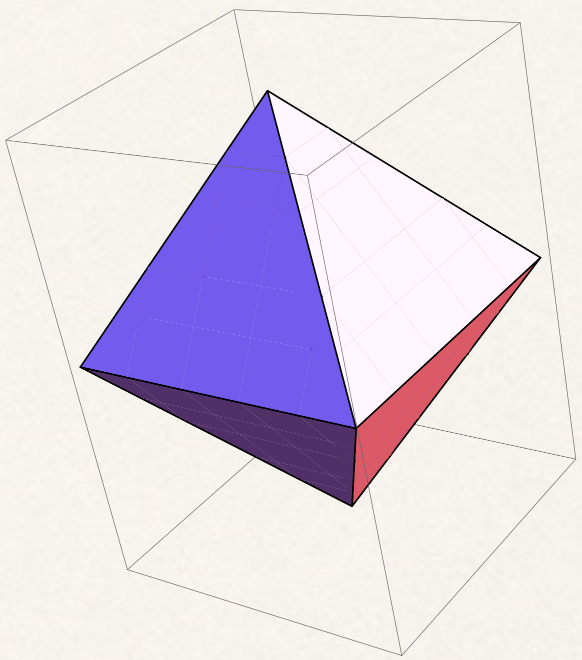
$$\min_x c^T x$$

$$Ax = b$$

$$x \geq 0$$

Geometry of LP

- The constraints in LP can be viewed as an intersecting set of hyperplanes
- Each (row,column) multiple defines a hyperplane
- The region defined by Ax=b for all non-negative x is called a polyhedron



$$Ax = b$$

$$x \geq 0$$

Media Blitz about Interior Point Methods: Karmakar's breakthrough

Breakthrough in Problem Solving

By JAMES GLEICK

A 28-year-old mathematician at A.T.&T. Bell Laboratories has made a startling theoretical breakthrough in the solving of systems of equations that often grow too vast and complex for the most powerful computers.

The discovery, which is to be formally published next month, is already circulating rapidly through the mathematical world. It has also set off a deluge of inquiries from brokerage houses, oil companies and airlines, industries with millions of dollars at stake in problems known as linear programming.

Faster Solutions Seen

These problems are fiendishly complicated systems, often with thousands of variables. They arise in a variety of commercial and government applications, ranging from allocating time on a communications satellite to routing millions of telephone calls over long distances, or whenever a limited, expensive resource must be spread most efficiently among competing users. And investment companies use them in creating portfolios with the best mix of stocks and bonds.

The Bell Labs mathematician, Dr. Narendra Karmarkar, has devised a radically new procedure that may speed the routine handling of such problems by businesses and Government agencies and also make it possible to tackle problems that are now far out of reach.

"This is a path-breaking result," said Dr. Ronald L. Graham, director of mathematical sciences for Bell Labs in Murray Hill, N.J.

"Science has its moments of great progress, and this may well be one of them."

Because problems in linear programming can have billions or more possible answers, even high-speed computers cannot check every one. So computers must use a special procedure, an algorithm, to examine as few answers as possible before finding the best one — typically the one that minimizes cost or maximizes efficiency.

A procedure devised in 1947, the simplex method, is now used for such problems,

Continued on Page A19, Column 1

Karmarkar at Bell Labs: an equation to find a new way through the maze

Folding the Perfect Corner

A young Bell scientist makes a major math breakthrough

Every day 1,200 American Airlines jets crisscross the U.S., Mexico, Canada and the Caribbean, stopping in 110 cities and bearing over 80,000 passengers. More than 4,000 pilots, copilots, flight personnel, maintenance workers and baggage carriers are shuffled among the flights; a total of 3.6 million gal. of high-octane fuel is burned. Nuts, bolts, altimeters, landing gears and the like must be checked at each destination. And while performing these scheduling gymnastics, the company must keep a close eye on costs, projected revenue and profits.

Like American Airlines, thousands of companies must routinely untangle the myriad variables that complicate the efficient distribution of their resources. Solving such monstrous problems requires the use of an abstruse branch of mathematics known as linear programming. It is the kind of math that has frustrated theoreticians for years, and even the fastest and most powerful computers have had great difficulty juggling the bits and pieces of data. Now Narendra Karmarkar, a 28-year-old Indian-born mathematician at Bell Laboratories in Murray Hill, N.J., after only a years' work has cracked the puzzle of linear programming by devising a new algorithm, a step-by-step mathematical formula. He has translated the procedure into a program that should allow computers to track a greater combination of tasks than ever before and in a fraction of the time.

Unlike most advances in theoretical mathematics, Karmarkar's work will have an immediate and major impact on the real world. "Breakthrough is one of the most abused words in science," says Ronald Graham, director of mathematical sciences at Bell Labs. "But this is one situation where it is truly appropriate."

Before the Karmarkar method, linear equations could be solved only in a cumbersome fashion, ironically known as the simplex method, devised by Mathematician George Dantzig in 1947. Problems are conceived of as giant geodesic domes with thousands of sides. Each corner of a facet on the dome

TIME MAGAZINE, December 3, 1984

AT&T Markets Problem Solver, Based On Math Whiz's Find, for $8.9 Million

By ROGER LOWENSTEIN
Staff Reporter of THE WALL STREET JOURNAL

NEW YORK—American Telephone & Telegraph Co. has called its math whiz, Narendra Karmarkar, a latter-day Isaac Newton. Now, it will see if he can make the firm some money.

Four years after AT&T announced an "astonishing" discovery by the Indian-born Mr. Karmarkar, it is marketing an $8.9 million problem solver based on his invention.

Dubbed Korbx, the computer-based system is designed to solve major operational problems of both business and government. AT&T predicts "substantial" sales for the product, but outsiders say the price is high and point out that its commercial viability is unproven.

"At $9 million a system, you're going to have a small number of users," says Thomas Magnanti, an operations-research specialist at Massachusetts Institute of Technology. "But for very large-scale problems, it might make the difference."

Korbx uses a unique algorithm, or step-by-step procedure, invented by Mr. Karmarkar, a 32-year old, an AT&T Bell Laboratories mathematician.

"It's designed to solve extremely difficult or previously unsolvable resource-allocation problems—which can involve hundreds of thousands of variables—such as personnel planning, vendor selection, and equipment scheduling," says Aristides Fronistas, president of an AT&T division created to market Korbx.

Potential customers might include an airline trying to determine how to route many planes between numerous cities and an oil company figuring how to feed different grades of crude oil into various refineries and have the best blend of refined products emerge.

AT&T says that fewer than 10 companies, which it won't name, are already using Korbx. It adds that, because of the price, it is targeting only very large companies—mostly in the Fortune 100.

Korbx "won't have a significant bottom-line impact initially" for AT&T, though it might in the long term, says Charles Nichols, an analyst with Bear, Stearns & Co. "They will have to expose it to users and demonstrate" it uses.

AMR Corp.'s American Airlines says it's considering buying AT&T's system. Like other airlines, the Fort Worth, Texas, carrier has the complex task of scheduling pilots, crews and flight attendants on thousands of flights every month.

Thomas M.Cook, head of operations research at American, says, "Every airline has programs that do this. The question is: Can AT&T do it better and faster? The jury is still out."

The U.S. Air Force says it is considering using the system at the Scott Air Force Base in Illinois.

One reason for the uncertainty is that AT&T has, for reasons of commercial secrecy, deliberately kept the specifics of Mr. Karmarkar's algorithm under wraps.

"I don't know the details of their system," says Eugene Bryan, president of Decision Dynamics Inc., a Portland, Ore., consulting firm that specializes in linear programming, a mathematical technique that employs a series of equations using many variables to find the most efficient way of allocating resources.

Mr. Bryan says, though, that if the Karmarkar system works, it would be extremely useful. "For every dollar you spend on optimization," he says, "you usually get them back many-fold."

AT&T has used the system in-house to help design equipment and routes on its Pacific Basin system, which involves 22 countries. It's also being used to plan AT&T's evolving domestic network, a problem involving some 800,000 variables.

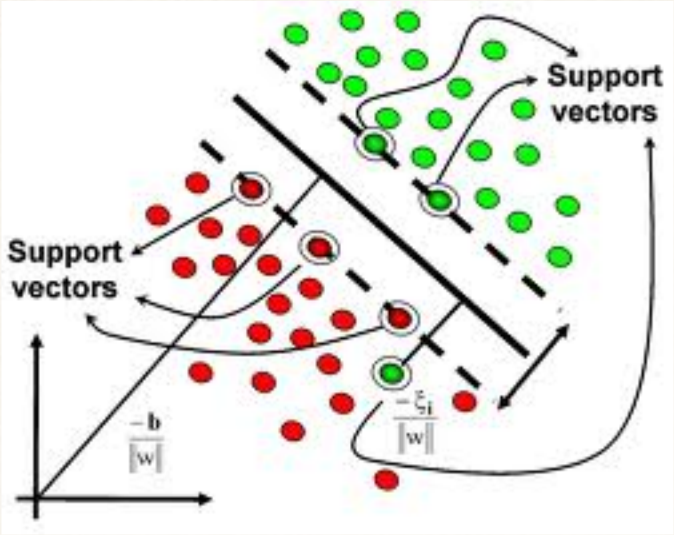
THE WALL STREET JOURNAL, August 15, 1988

Integer Programming

- A large variety of interesting problems can be formulated as IP problems
  - Traveling salesman
  - Scheduling
  - Boolean satisfiability

$$\begin{aligned} \min_x \, c^T x \\ Ax = b \\ x \geq 0 \\ x \quad \text{integer} \end{aligned}$$

Quadratic Programming



- When the function to be minimized is quadratic, the optimization problem becomes a QP
- Simple example of a QP is least-squares optimization
- Generally, QPs are harder to solve than LPs

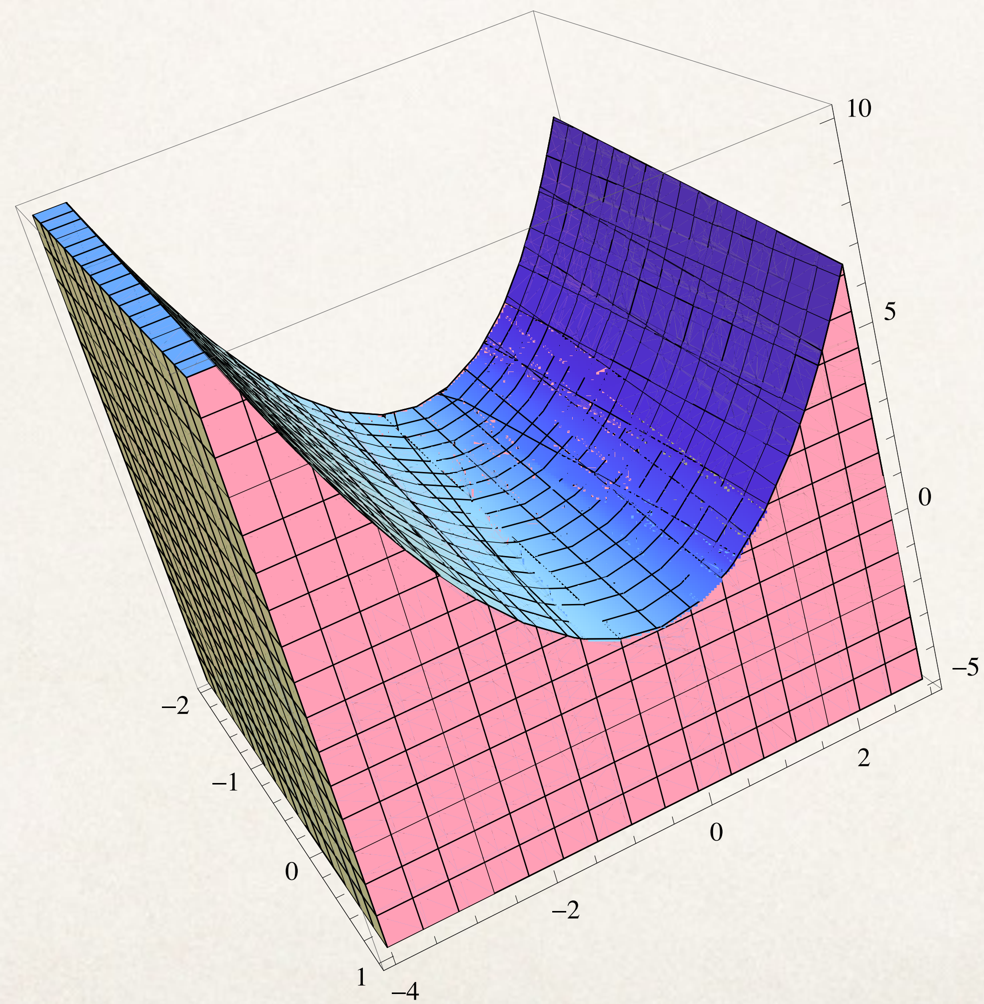
$$\min_x \frac{1}{2} x^T A x + p^T x + r$$

$$Ax = b$$

$$x \geq 0$$

Convexity

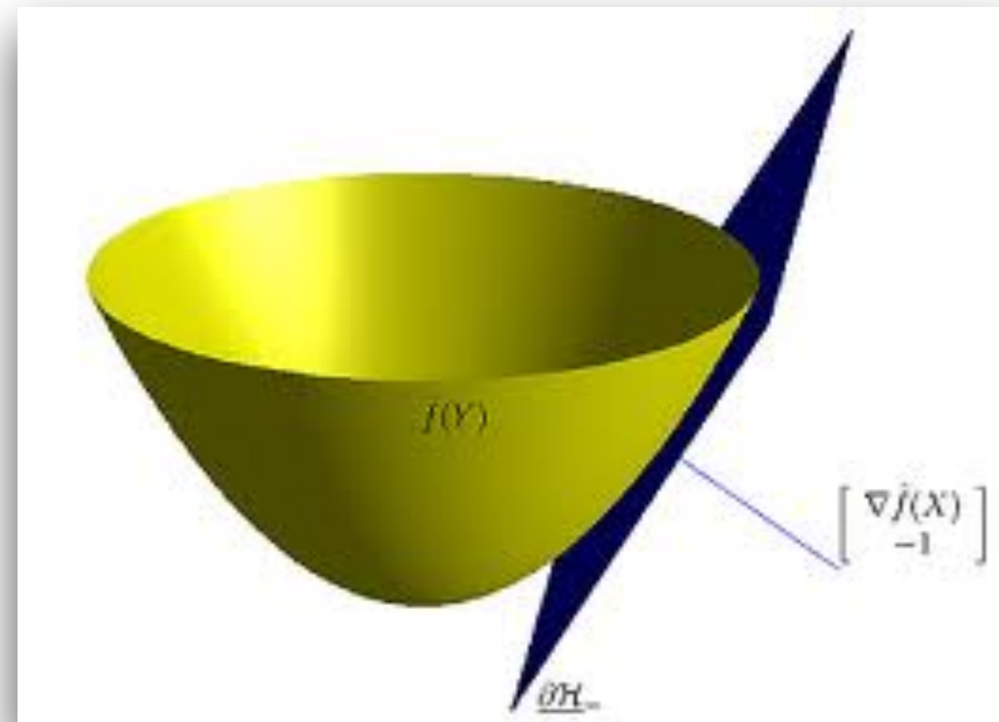
$$f(\lambda x + (1-\lambda)y) \leq (1-\lambda)f(x) + \lambda f(y)$$

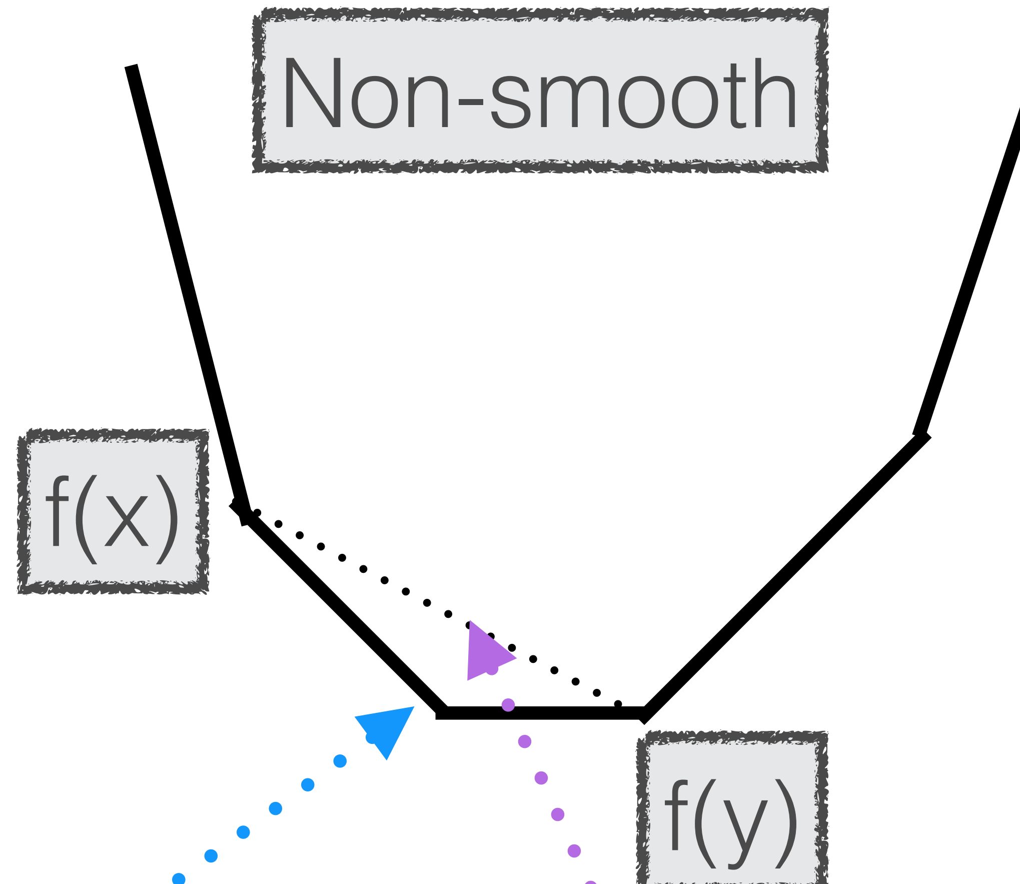


The division between tractable and intractable optimization is not linear vs. nonlinear, but convex vs. non-convex

Convex functions

Smooth



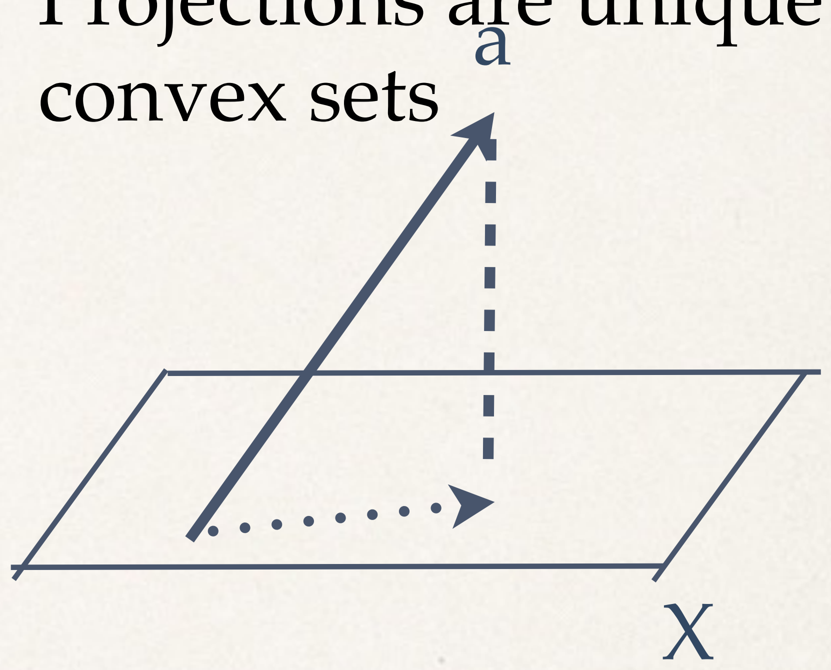


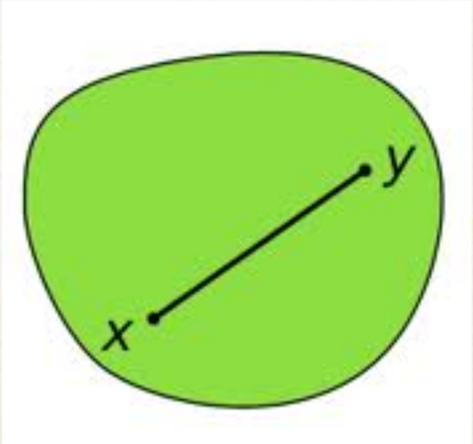
$$f((1-\lambda)x+\lambda y) \leq (1-\lambda)f(x)+\lambda f(y)$$

Geometric Principles

- Projection theorem in vector spaces
  - Find the closest element in set X to a vector **a** (not in C)
  - Holds in any Hilbert space
- Foundational principle of machine learning and statistics

- Projections are unique for convex sets





Least Squares Estimation

- Invented by Gauss, the most widely used estimation method

$$y = X\beta + \epsilon$$

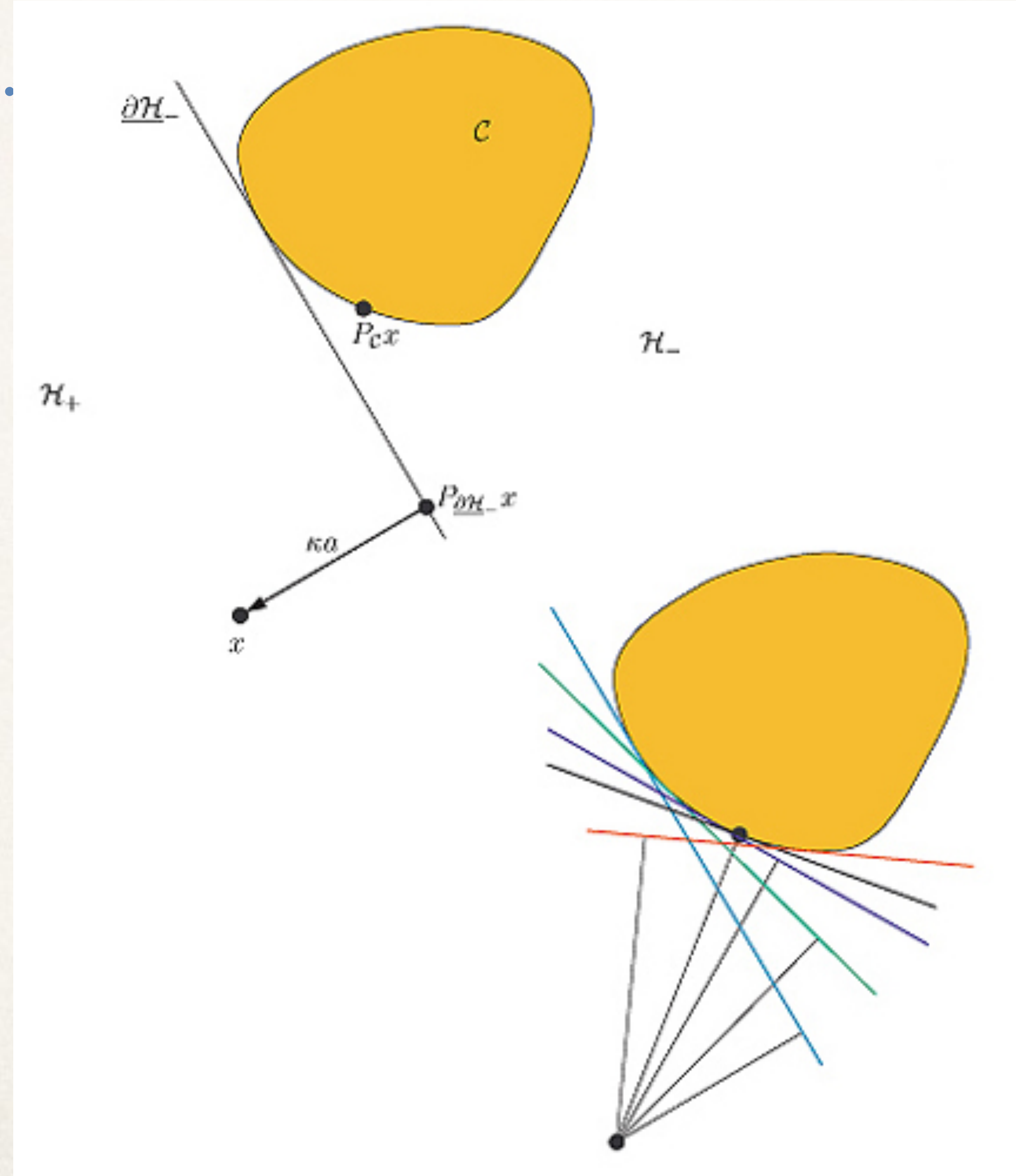
- Using least squares, the solution is given as

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

- Fourier analysis: generalized least squares in infinite

Hahn-Banach Extension Theorem

- Considered the most important theorem in optimization!
- Geometric form: given a point outside a convex set, there is a separating hyperplane
- Foundational principle underlying duality theory



Duality Theory

- Hahn-Banach theorem shows how to convert minimization problems into maximization problems
- Basis for Lagrange Multiplier theory
- Instead of finding closest element in a set, find the furthest separating hyperplane!
- Does not require orthogonality!

$$\min_x f(x) \text{ such that}$$

$$g_i(x) \geq 0$$

$$h_i(x) = 0$$

$$max_{\alpha,\beta} \mathcal{L}(\alpha, \beta)$$

Dual Spaces

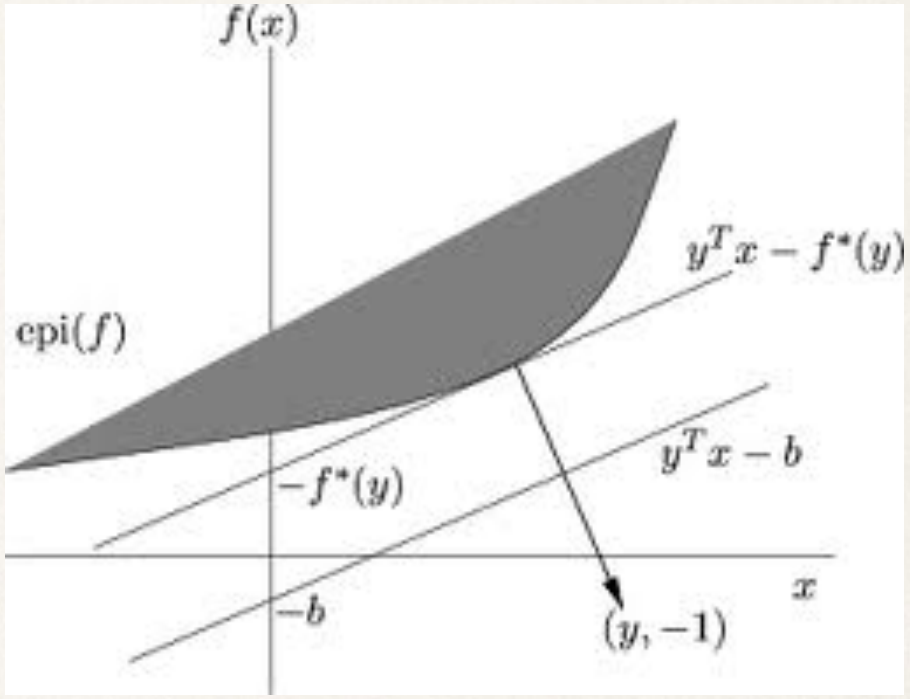
- We will explore the concept of duality in many forms

- Dual norms

$$\|x\|^* = \max_z \{z^T x \,|\, \|z\| = 1\}$$

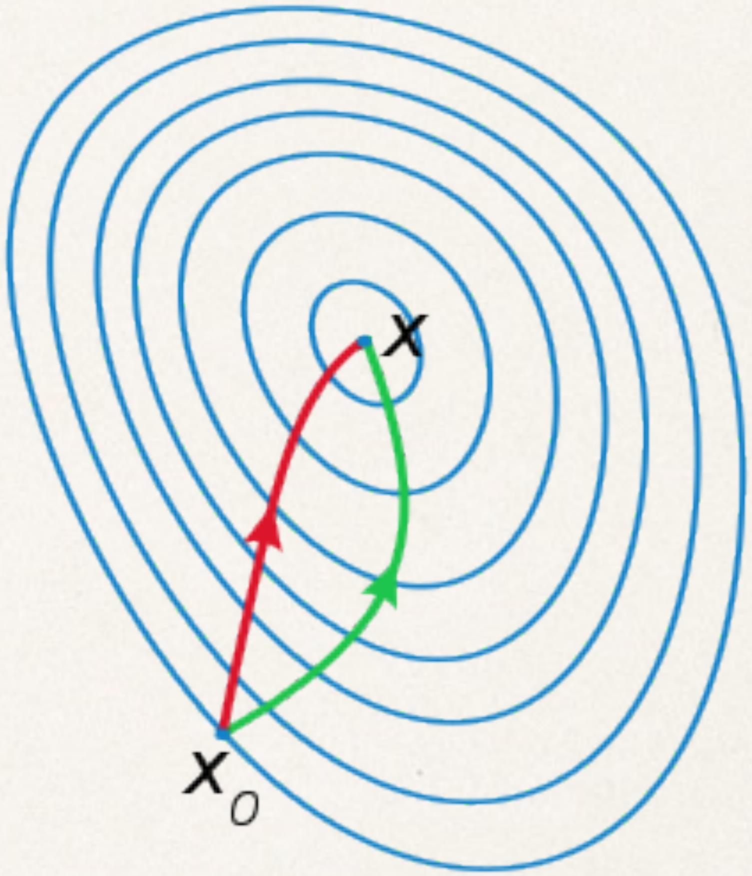
- Conjugate function

$$f^*(y) = \sup_x (y^T x - f(x))$$

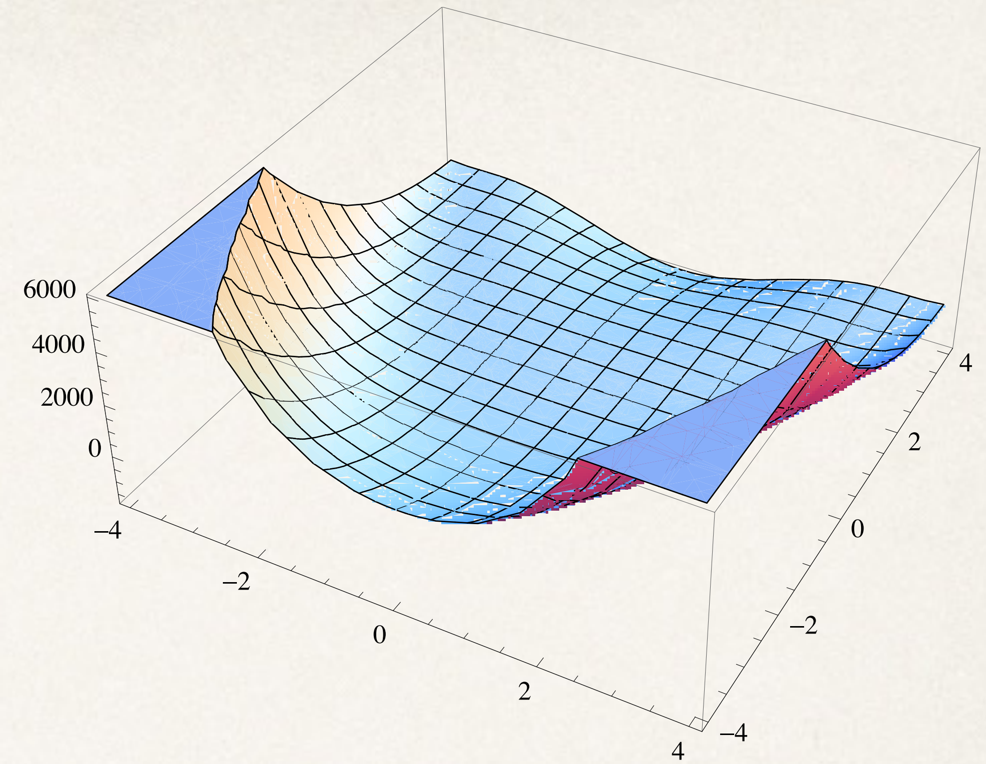


Algorithms

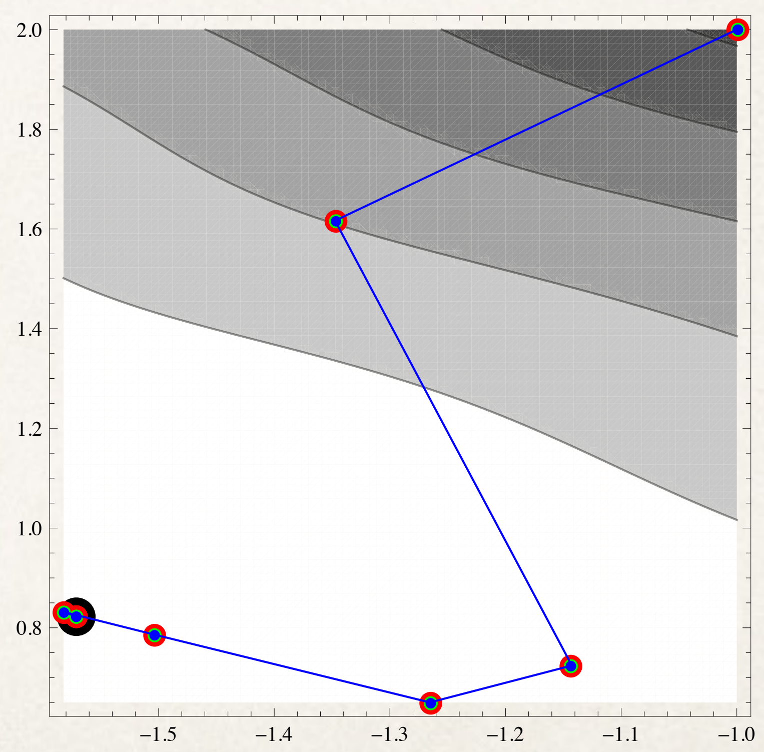
- First-order methods
- Second-order methods
- Incremental methods
- Batch methods
- Deterministic methods
- Stochastic methods



Unconstrained Optimization

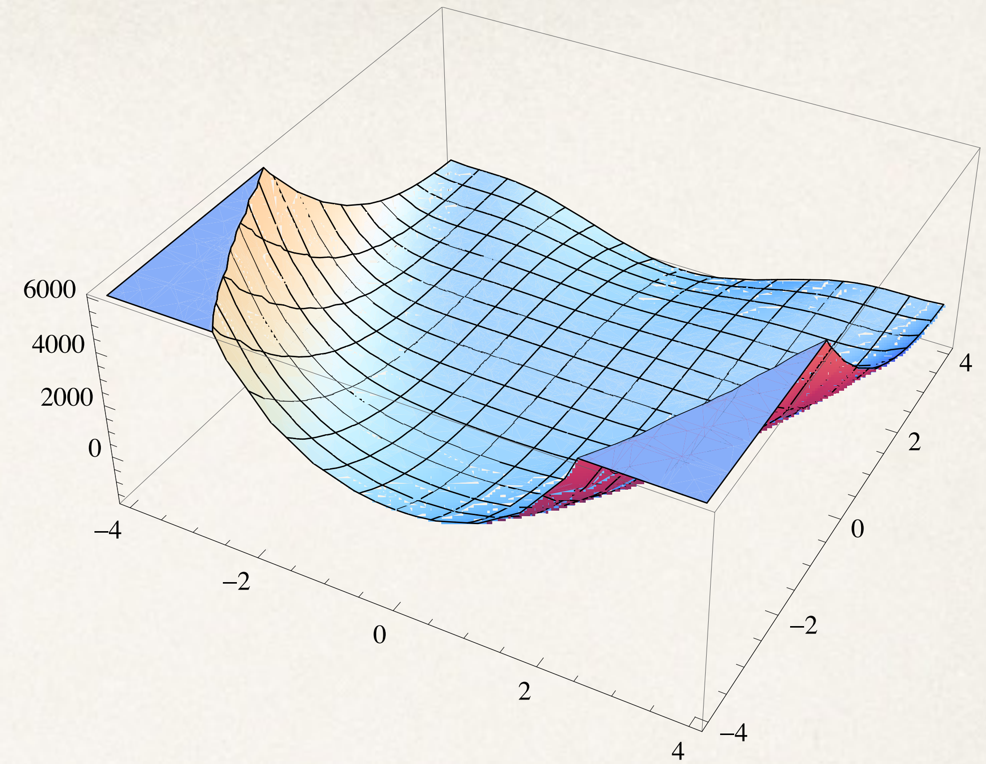


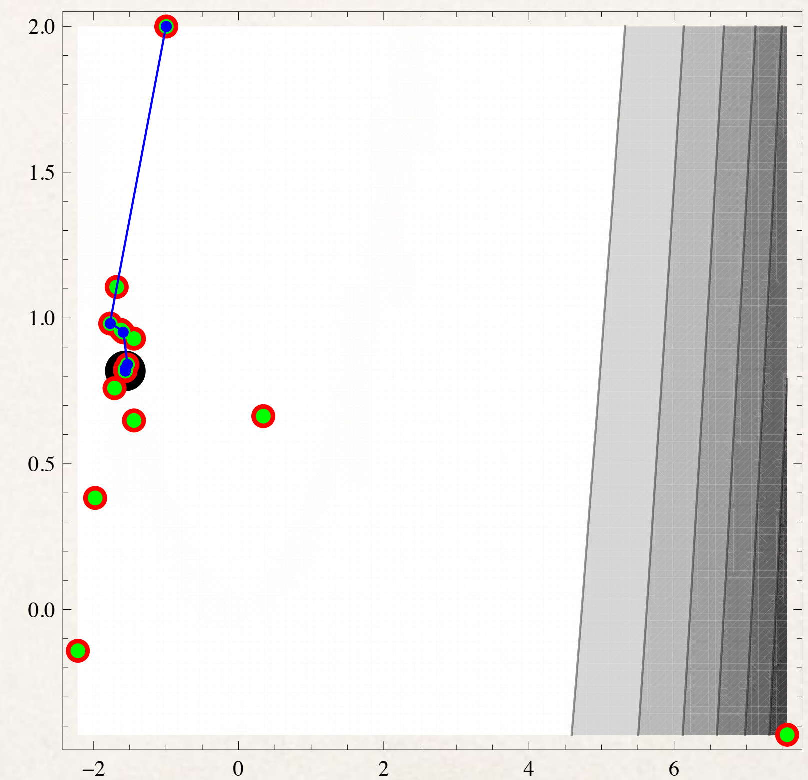
- Simplest example of unconstrained optimization is to find the global minimum of a function
- Newton's method is widely used
- Second-order algorithm



First-order methods

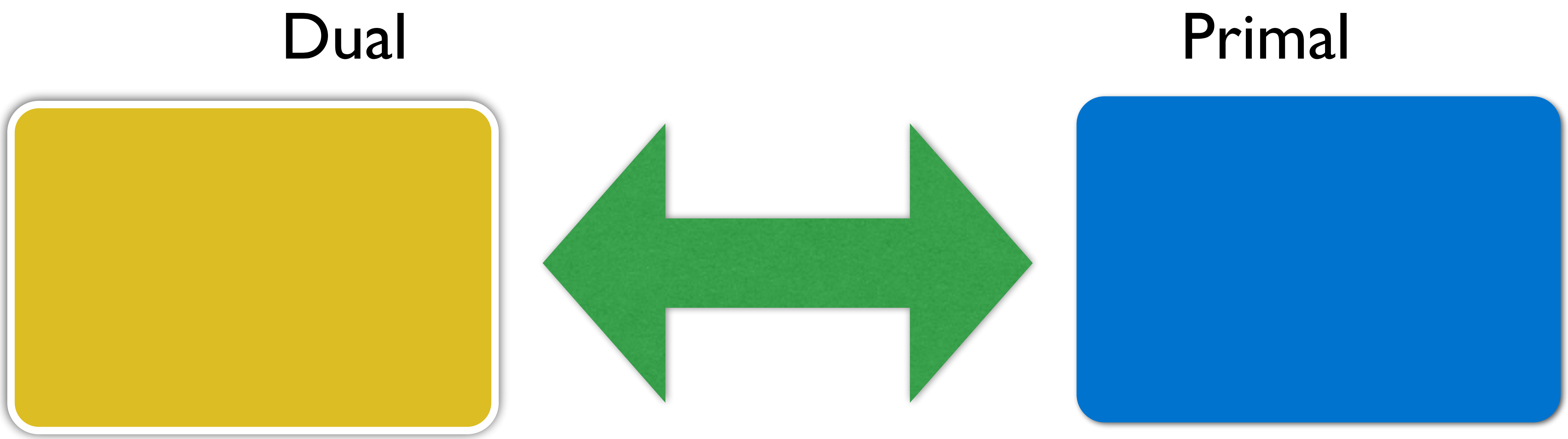
- Among the simplest optimization methods are first-order gradient methods
- These fell out of favor till recently when they have become popular again due to increased interest in large problems
- Conjugate gradient methods
- Proximal gradient methods



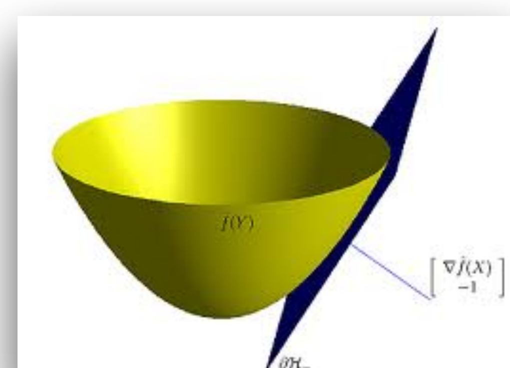


Mirror Descent

(Nemirovsky and Yudin)



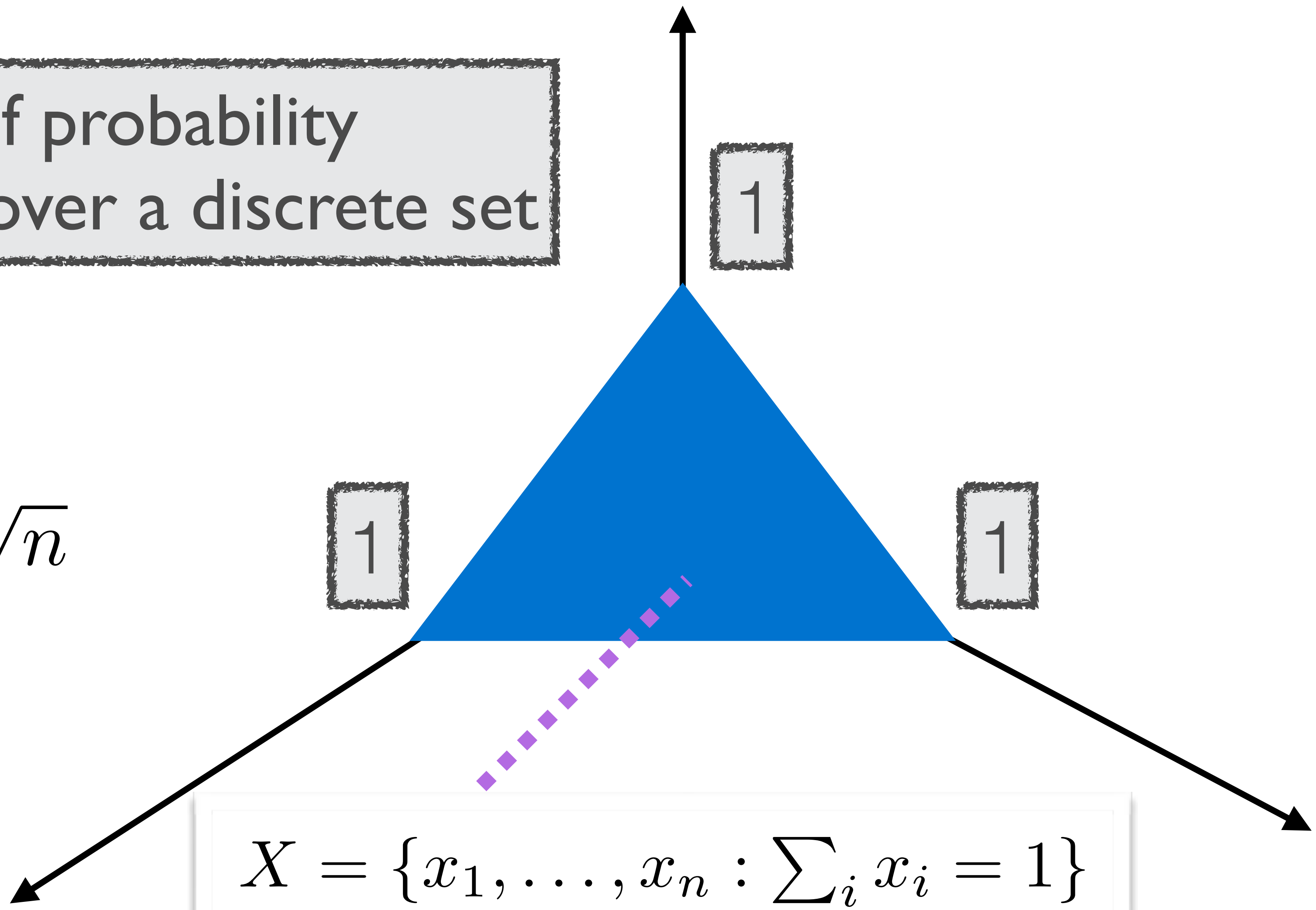
$$x_{k+1} = \nabla\psi^* \left(\nabla\psi(x_k) - t_k \partial f(x_k)\right)$$



Unit Simplex

Space of probability distributions over a discrete set

$$d = \sqrt{n}$$



$$X = \{x_1, \ldots, x_n : \textstyle\sum_i x_i = 1\}$$

Example: Unit Simplex

- Consider minimizing a convex function over the unit simplex

- Subgradient descent converges at the rate

$$\boxed{f(x_k) - \min_{x \in X} f(x) = \frac{O(1) L_f \sqrt{n}}{\sqrt{k}}}$$

- Mirror descent converges much faster

$$\boxed{f(x_k) - \min_{x \in X} f(x) = \frac{O(1)\sqrt{\ln n}\, \max_s \|\partial f(x_s)\|_\infty}{\sqrt{k}}}$$