COMP6704 Lecture 1

Advanced Topics in Optimization

Introduction

Fall, 2022

Instructor: WU, Xiao-Ming

Class Information

- Lectures:
  - Tuesday <6:30 – 9:20 pm PQ303>
- Instructor: WU, Xiao-Ming @ Department of Computing
  - Office: PQ725
  - Phone: 27667261
  - Email: csxmwu@comp.polyu.edu.hk
- TA: FAN, Lu
  - Email: complu.fan@connect.polyu.hk

Assessment and Requirements

- Coursework

2 individual assignments (10% for each)	20%
Quiz (each class, in group)	25%
1 individual project	55%
Total	**100%**

- Assignments require programming in Python/matlab.
- Quiz – in-class group study of questions and reading materials.

Rules and Regulations: Assignment Submission

- All submissions (homework and project) will be done on Blackboard. Each assignment will be given a deadline by the Blackboard system. The normal cut-off time would be 11:59 pm on the specified date using the Blackboard clock.
- For late submissions, penalty will apply (33% penalty per day).
- You are encouraged to discuss with your teammates/classmates, but you should do the coding and writing independently.

Teaching Plan

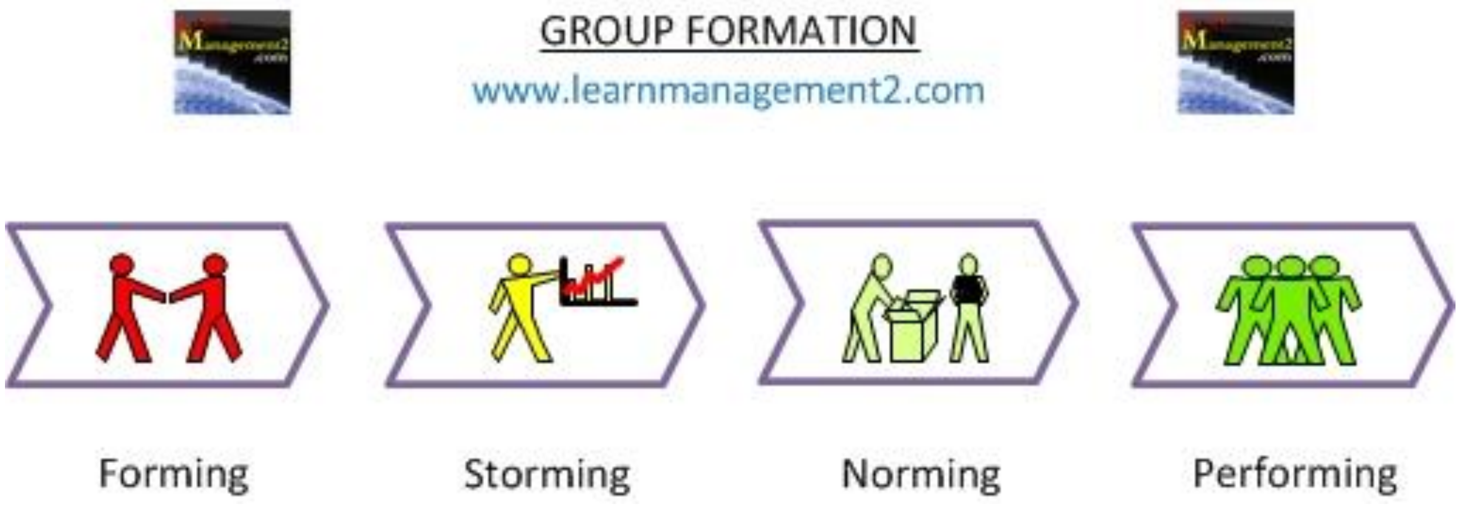
- **Basic knowledge (5 lectures)**
  - Least squares, linear programming.
  - Convex sets and functions.
  - Recognize and formulate convex optimization problems.
- **Optimization techniques & case studies in ML (8 lectures)**
  - Unconstrained optimization. Gradient decent, stochastic gradient decent, Newton's method, logistic regression, neural networks, back propagation, Bayesian optimization.
  - Constrained optimization. Lagrange multipliers, support vector machine, duality, KKT conditions, alternating direction method of multipliers (ADMM). Lasso, dimension reduction, PCA, Robust PCA, semi-definite programming.
- **Project presentations (1-2 lectures after the normal teaching period)**

Focus of the Course

- **Recognize** optimization problems
- **Formulate** optimization problems
- **Solve** optimization problems
  - Basic tools and techniques
- **Applications** of optimization techniques in Machine Learning

Group Study

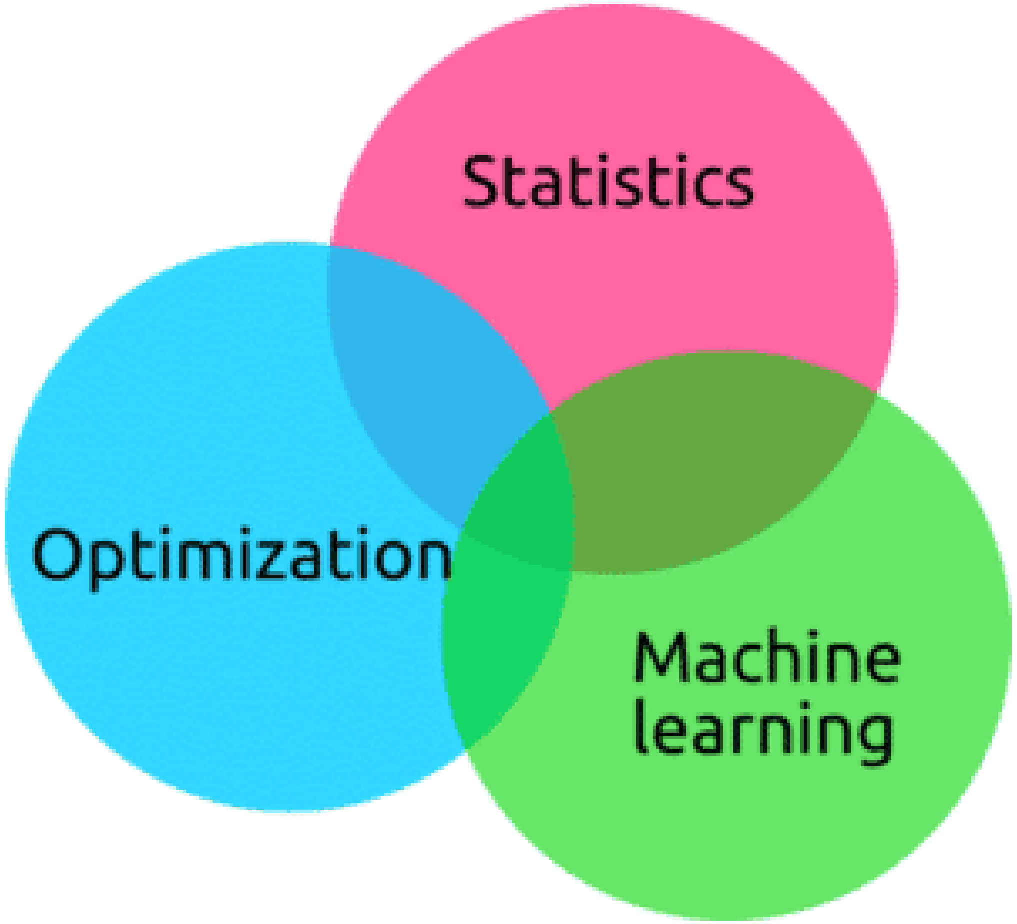
- Work with your group mates on a problem set
- Answer questions or present ideas



What Is Optimization?

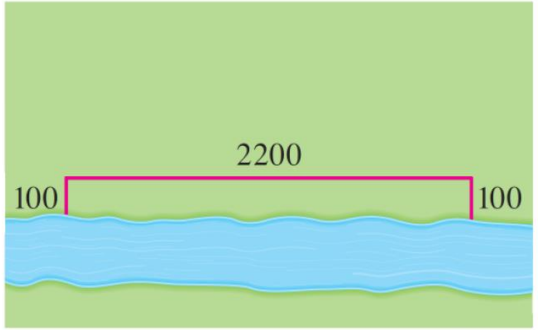
Optimization is Everywhere

Why Optimization Is Important for ML?

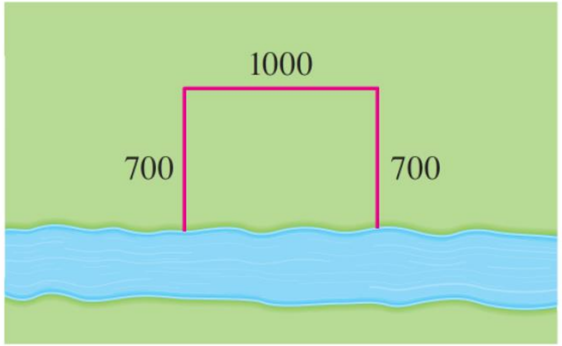


Optimization Problems

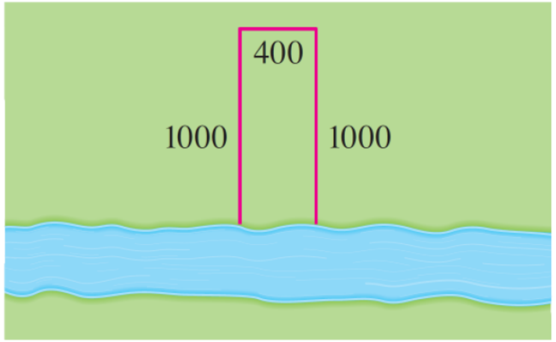
A farmer has 2400 ft of fencing and wants to fence off a rectangular field that borders a straight river. He needs no fence along the river. What are the dimensions of the field that has the largest area?



Area $= 100 \cdot 2200 = 220{,}000\ \text{ft}^2$

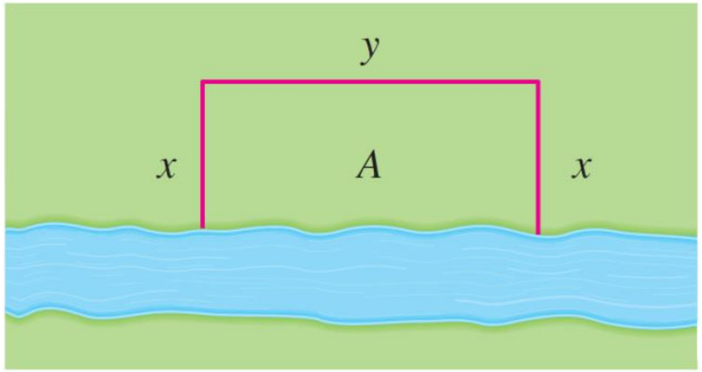


Area $= 700 \cdot 1000 = 700{,}000\ \text{ft}^2$



Area $= 1000 \cdot 400 = 400{,}000\ \text{ft}^2$

Optimization Problems



Maximize: $A = xy$

Constraint: $2x + y = 2400$

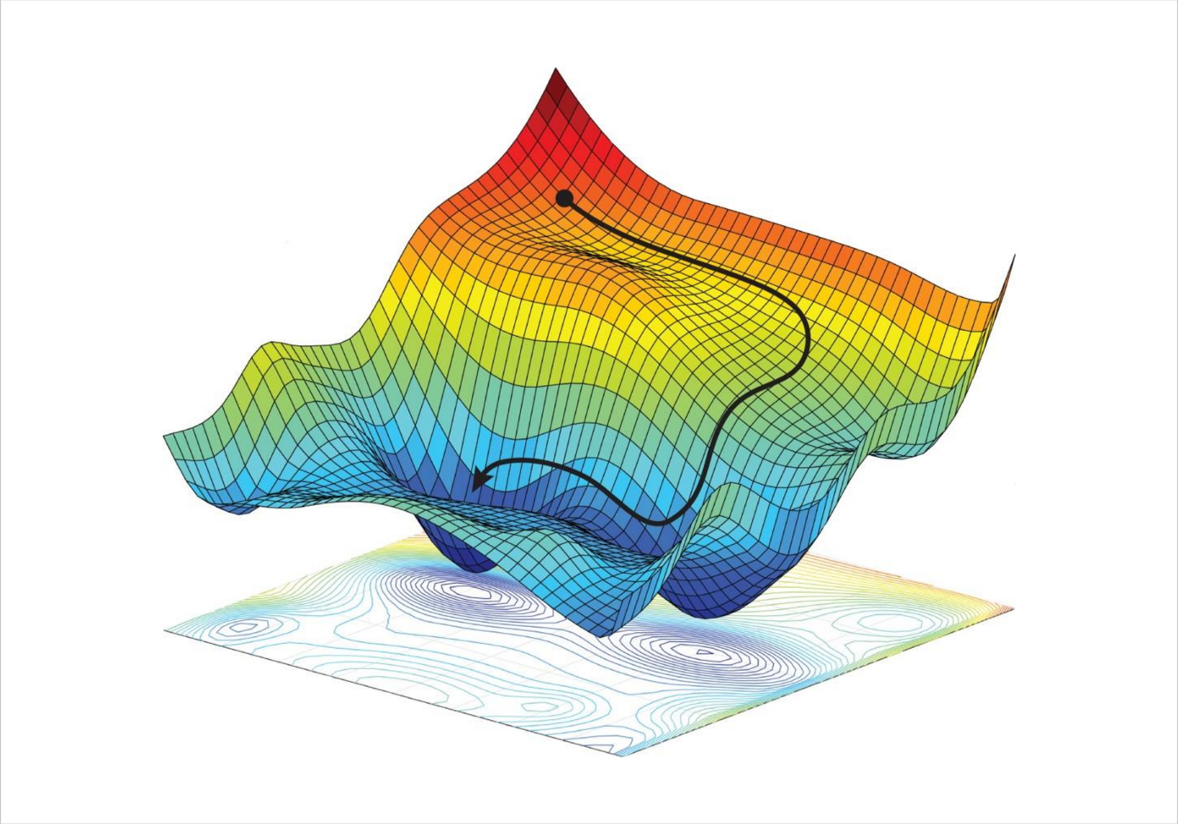
Optimization Problems

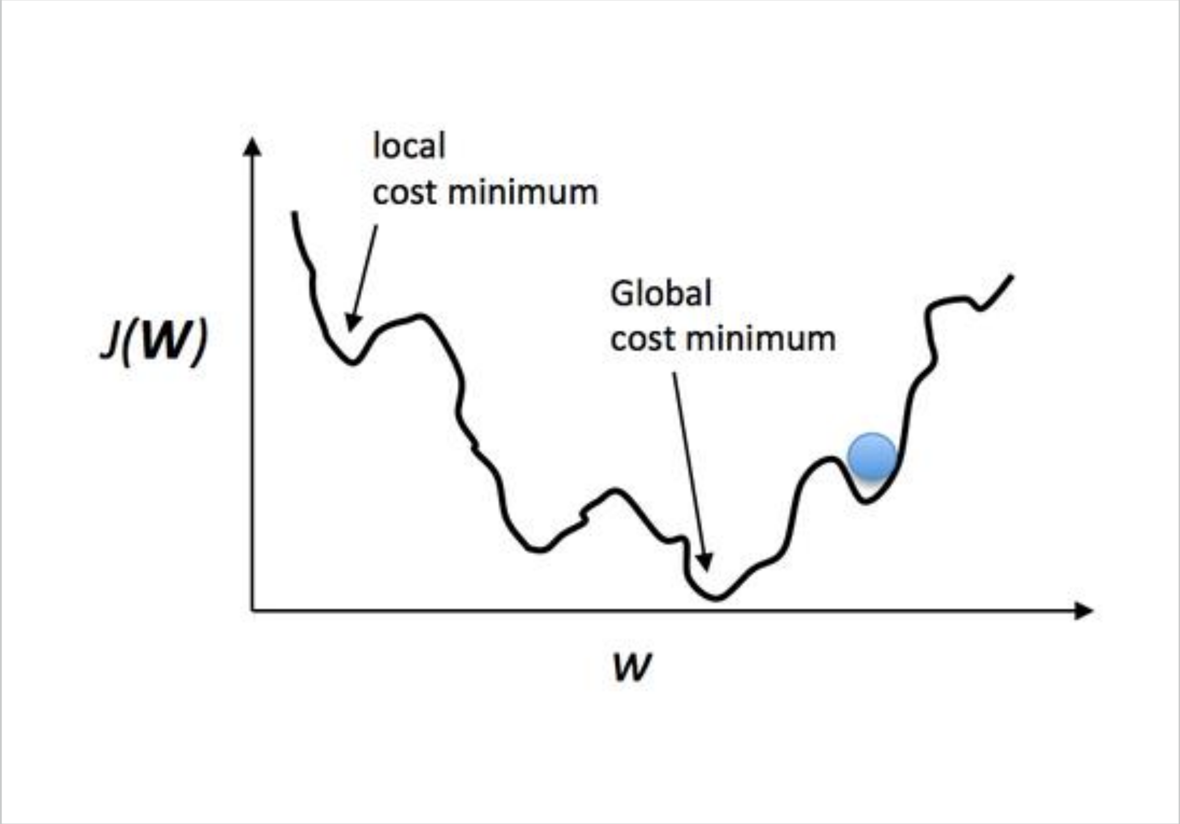
$$
\begin{array}{ll}
\text{minimize} & f_0(x_1, \ldots, x_n) \\
\text{subject to} & f_1(x_1, \ldots, x_n) \leq 0 \\
& \qquad \ldots \\
& f_m(x_1, \ldots, x_n) \leq 0
\end{array}
$$

- $x = (x_1, x_2, \ldots, x_n)$ are decision variables
- $f_0(x_1, x_2, \ldots, x_n)$ gives the cost of choosing x
- Inequalities give constraints that x must satisfy

Optimization is Difficult!

- **General optimization problems are intractable**
  - Local (non global) minima
  - All kinds of constraints

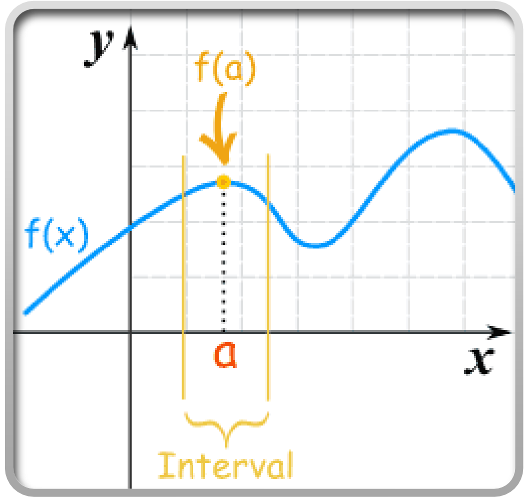




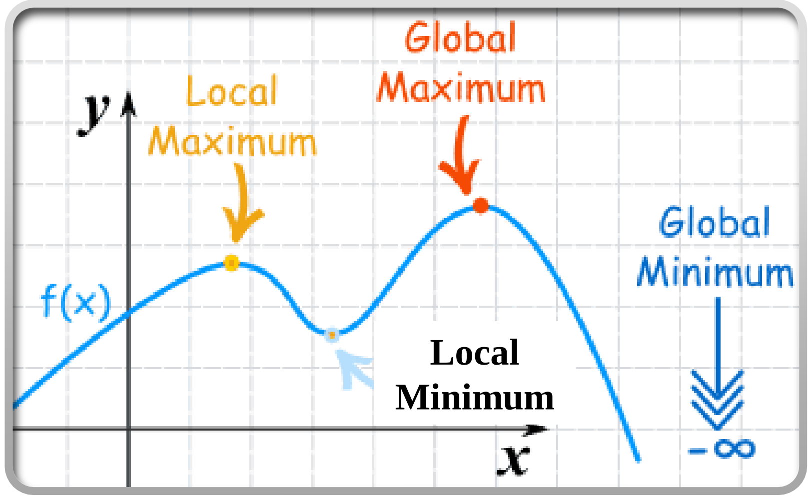
Local / Global Maximum / Minimum

- **Definition**. A point x^* is called a **local minimum** of f if there exists $\varepsilon > 0$ such that $f(x^*) < f(x)$ for all $||x - x^*|| \leq \varepsilon$.

- **Definition**. A point x^* is called a **global minimum** of f if $f(x^*) < f(x)$ for all feasible x.



Local maximum: f(a) ≥ f(x) for all x **in the interval**

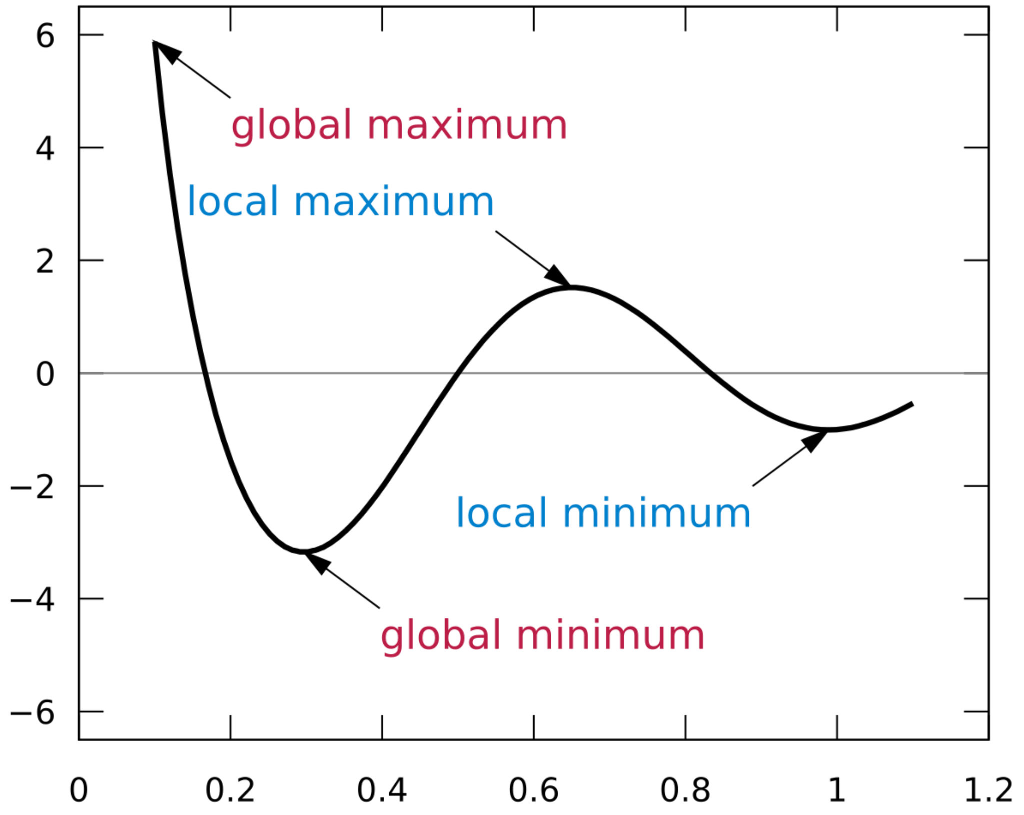


Global maximum: The maximum **over the entire function**.

Local / Global Maximum / Minimum

- **Definition**. A point x^* is called a local minimum of f if there exists $\varepsilon > 0$ such that $f(x^*) < f(x)$ for all $\left||x - x^*|\right| \leq \varepsilon$.

- **Definition**. A point x^* is called a global minimum of f if $f(x^*) < f(x)$ for all feasible x.



Convexity:
A Crucial Matter

- A convex function is one whose graph slopes everywhere toward its minimum value.

- A nonconvex function may have many basins, or local minima.

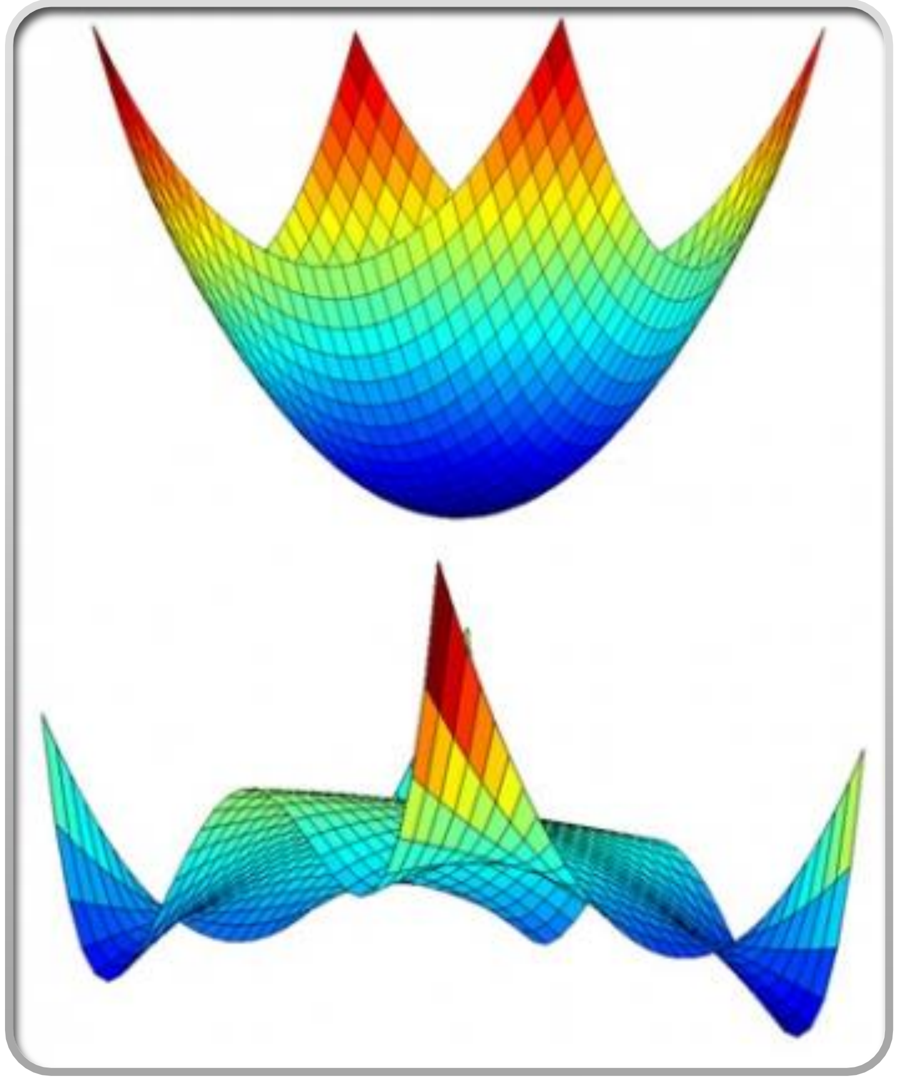


Image by Amir Ali Ahmadi

History

- 1940s: Linear programming
- 1950s: Quadratic programming
- 1960s: Geometric programming
- 1990s: Semidefinite programming, second-order cone programming, quadratically constrained quadratic programming
  - Around 1990 (Nesterov & Nemirovski): polynominal-time interior-point methods for nonlinear convex programming

From Lieven Vandenberghe

Two Special Subclasses

- Least Squares
- Linear Programming

Warm-up

- Problem: Suppose we measure a distance four times, and obtain the following results: 72, 69, 70 and 73 units.
- What is the **best estimate** of the correct measurement?
  - Use squared error

$$
\begin{aligned}
S &= (x-72)^2 + (x-69)^2 + (x-70)^2 + (x-73)^2 \\
&= 4(x-71)^2 + 10
\end{aligned}
$$

Statistics Review

Given, $x_1, x_2, \ldots, x_n$, what are the mean, variance, standard deviation of data?

Mean: $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$

Variance: $\sigma_x^2 = \frac{1}{N}\sum_{n=1}^{N}(x_i - \bar{x})^2$

Standard deviation: $\sigma_x = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_i - \bar{x})^2}$

Error Measurement

Why not consider other measurement such as …?

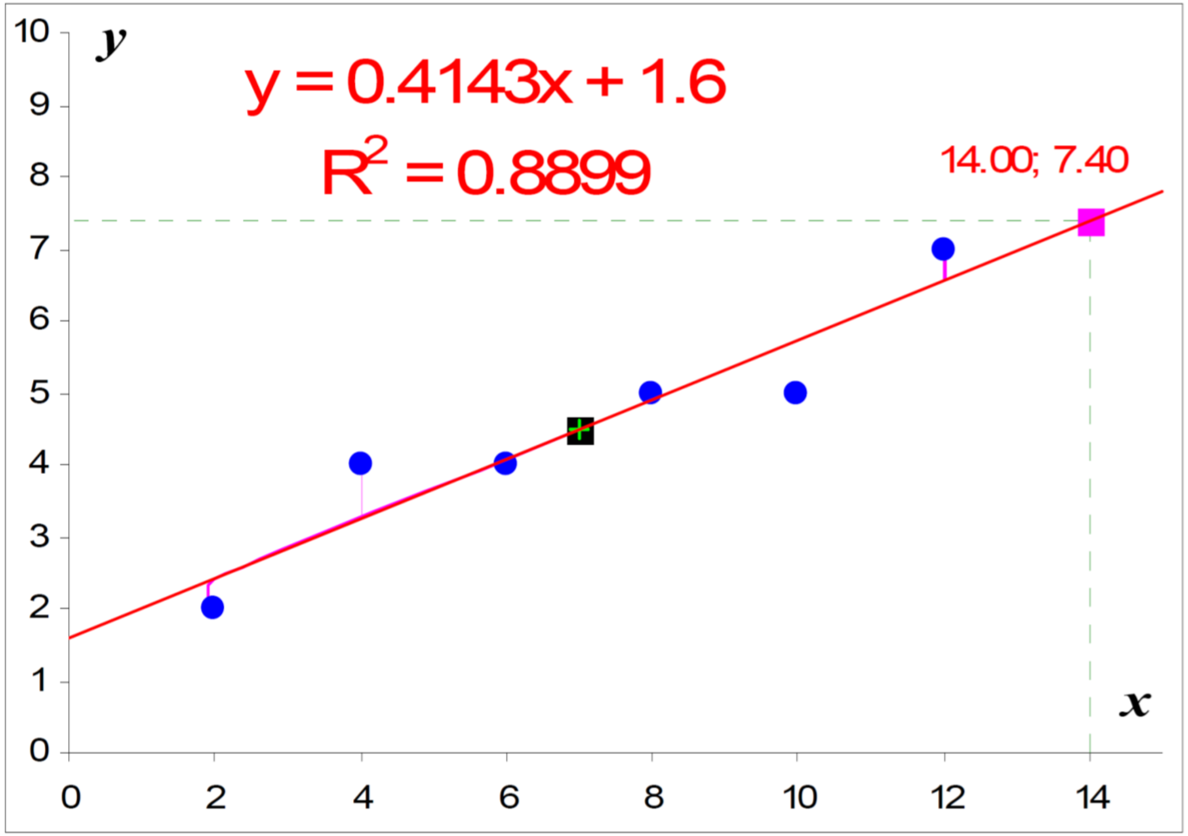
$$\frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})$$

$$\frac{1}{N}\sum_{n=1}^{N}|x_n - \bar{x}|$$

Least Square Regression

Carl Friedrich Gauss, 1794

DATA		MODEL
x	y	$y' = ax+b$
2	2	2.42857
4	4	3.25714
6	4	4.08571
8	5	4.91429
10	5	5.74286
12	7	6.57143
14	??	**7.4**



Find the line that best fits the data

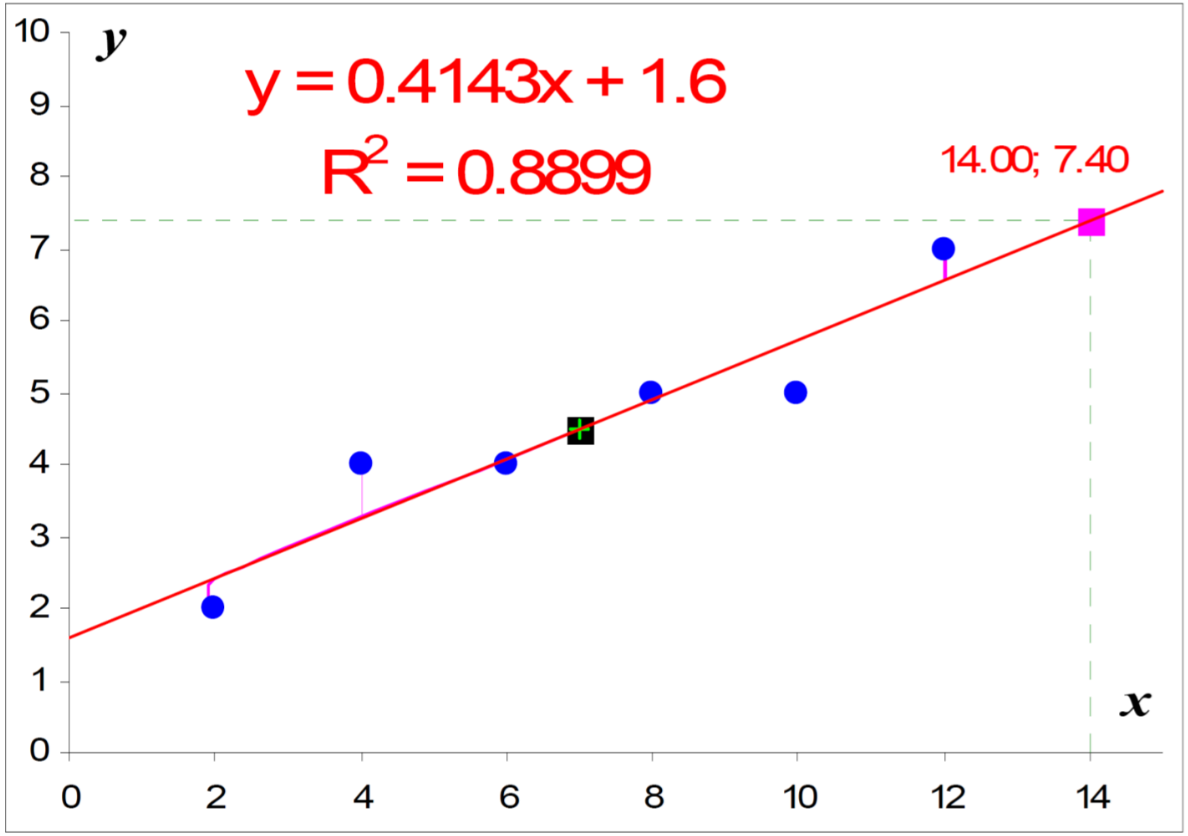
Meaning of the **Best Fit**

- What is the meaning of **best fit**?

$$\{(x_1, y_1), \ldots, (x_N, y_N)\}$$
$$y = ax + b$$

- Minimize the squared error

$$E(a, b) = \sum_{n=1}^{N} \left(y_n - (ax_n + b)\right)^2$$



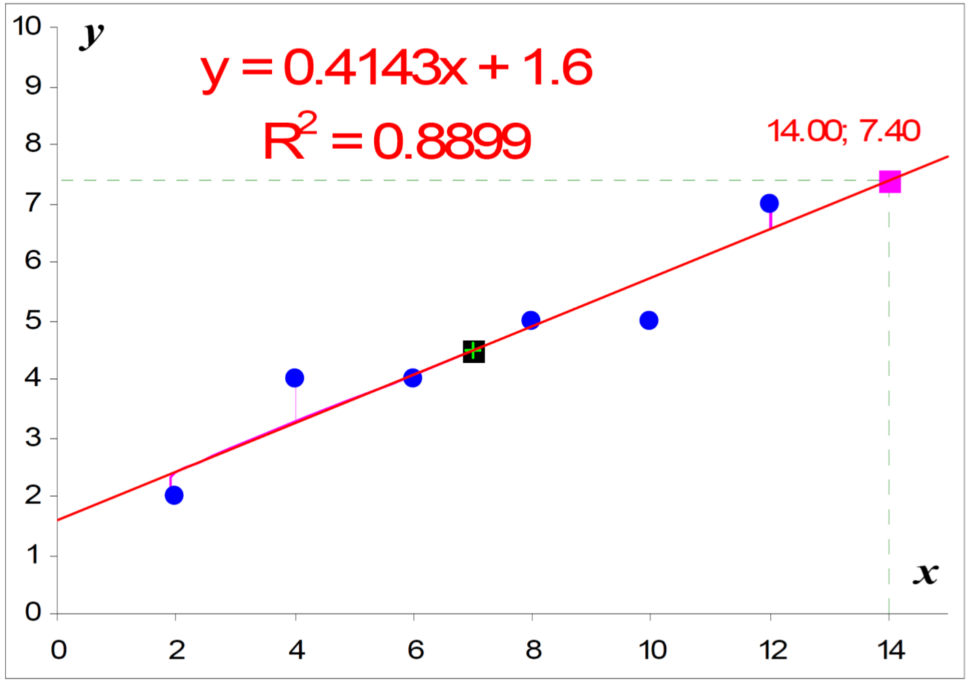
Solving Least Squares

$$E(a,b) = \sum_{n=1}^{N} \left(y_n - (ax_n + b)\right)^2$$

$$\frac{\partial E}{\partial a} = 0, \qquad \frac{\partial E}{\partial b} = 0$$

$$\frac{\partial E}{\partial a} = \sum_{n=1}^{N} 2\left(y_n - (ax_n + b)\right) * (-x_n)$$

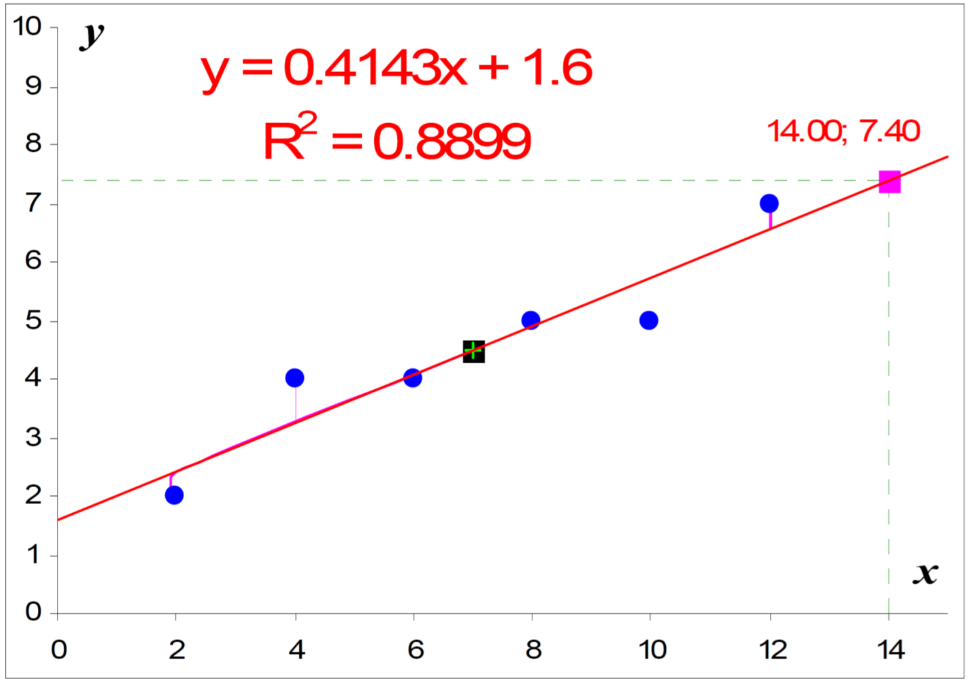
$$\frac{\partial E}{\partial \mathrm{b}} = \sum_{n=1}^{N} 2\left(y_n - (ax_n + b)\right) * 1$$



Solving Least Squares

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{n=1}^{N} x_n^2 & \sum_{n=1}^{N} x_n \\ \sum_{n=1}^{N} x_n & \sum_{n=1}^{N} 1 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{n=1}^{N} x_n y_n \\ \sum_{n=1}^{N} y_n \end{pmatrix}$$

$$\begin{aligned} \det \mathrm{M} &= N \sum_{n=1}^{N} x_n^2 - (N\bar{x})^2 \\ &= N^2 \left(\frac{1}{N} \sum_{n=1}^{N} x_n^2 - \bar{x}^2 \right) \\ &= N^2 * \frac{1}{\mathrm{N}} \sum_{n=1}^{N} (x_n - \bar{x})^2 \end{aligned}$$



Least Squares in Matrix Form

- Note that the notations are different from previous slides

$A \in \mathbb{R}^{k \times n} (with\ k \geq n), a_i^T$ are the rows of A, and the vector $x \in \mathbb{R}^n$

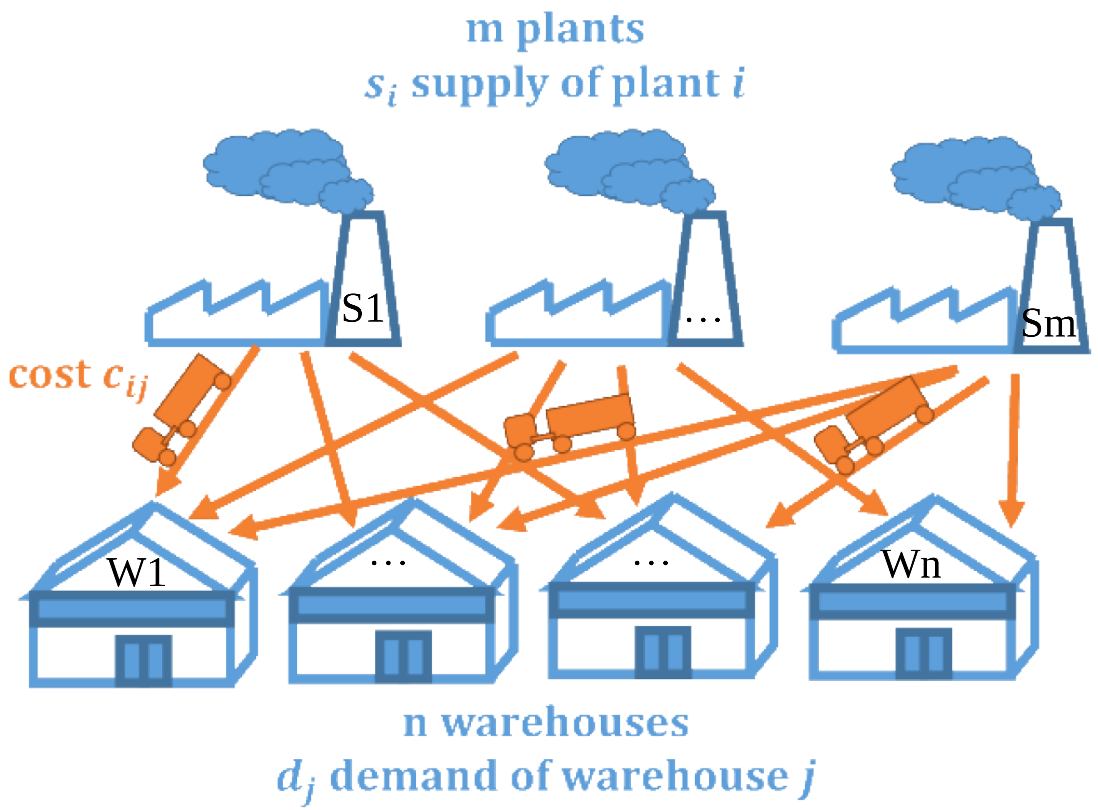
$$minimize\ f_0(x) = \|Ax - b\|_2^2 = \sum_{i=1}^{k} \left(a_i^T x - b_i\right)^2 .$$

The Transportation Problem

- Minimize the cost

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}\, x_{ij}$$

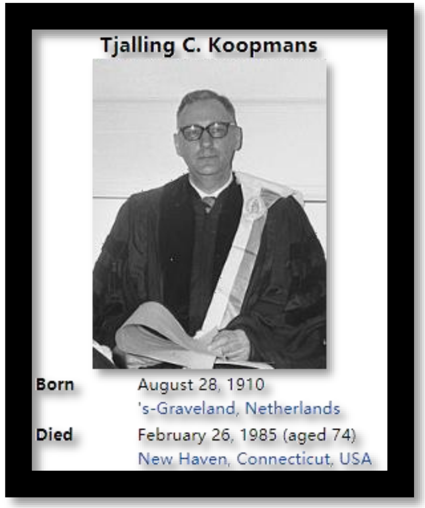
- Constraints
  - $\sum_{j=1}^{m} x_{ij} \leq s_i, \forall i = 1, \ldots, m$
  - $\sum_{i=1}^{n} x_{ij} \geq d_j, \forall j = 1, \ldots, n$
  - $x_{ij} \geq 0, \forall i, j$

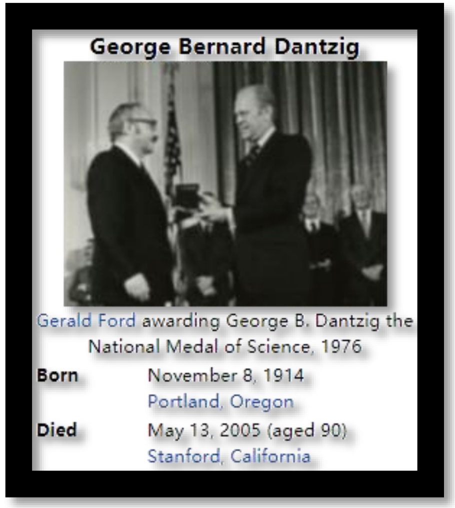


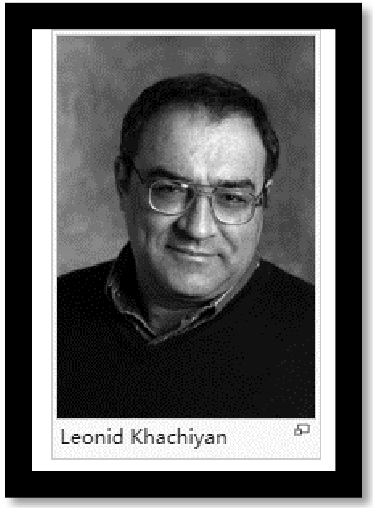
Linear Programming

$$\begin{array}{ll} minimize & c^T x \\ subject\ to & a_i^T x \leq b_i, \quad i = 1, \dots, m \end{array}$$

- Solving linear programs
  - No analytical formula for solution
  - Reliable and efficient algorithms and software
  - Computation time proportional to $n^2 m$ if $m \geq n$; less with structure
  - A mature technology







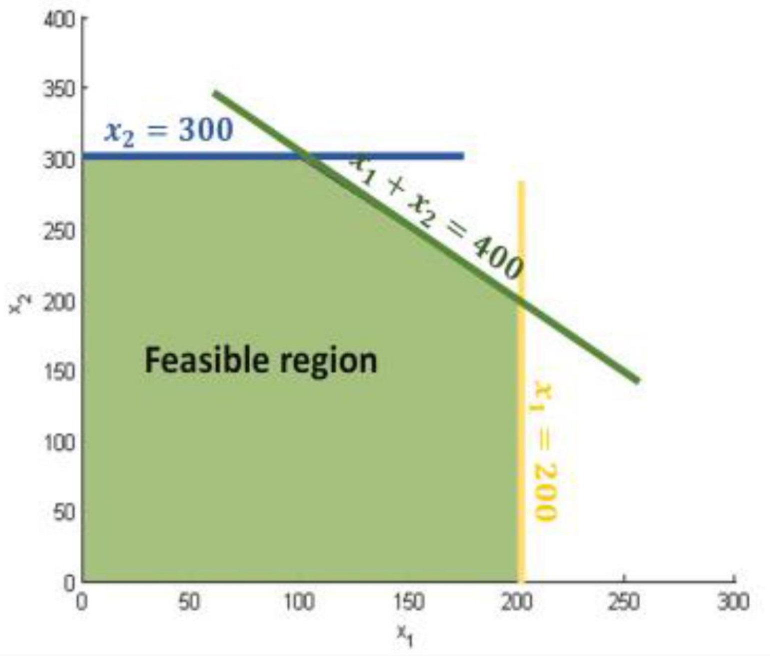
- 1700s, Fourier invented Fourier-Motzkin elimination method.
- 1930s, Kantorovich and Koopmans, optimal allocation of scarce resources.
- 1947, Dantzig invented the first practical algorithm for solving LPs: the simplex method.
- 1979, Khachiyan showed that LPs were solvable in polynomial time using the "ellipsoid method“, theoretical breakthrough.
- 1984, Karmarkar developed the "interior point method“, Along with the simplex method, this is the method of choice today for solving LPs.

History

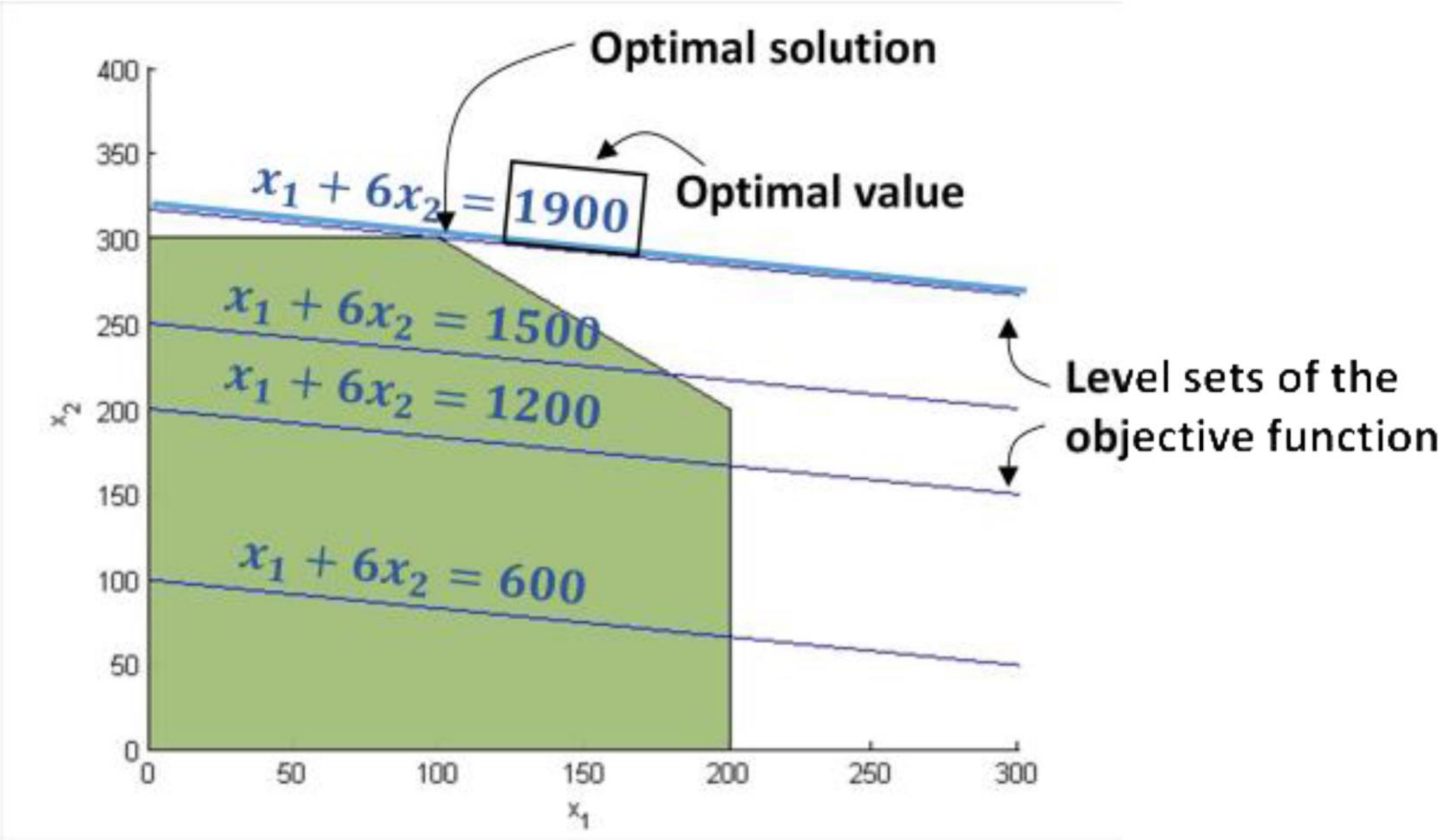
Solving systems of linear inequalities

Solving an LP

$$\begin{array}{ll} maximize & x_1 + 6x_2 \\ s.t. & x_1 \leq 200 \\ & x_2 \leq 300 \\ & x_1 + x_2 \leq 400 \\ & x_1, x_2 \geq 0 \end{array}$$



Level Sets

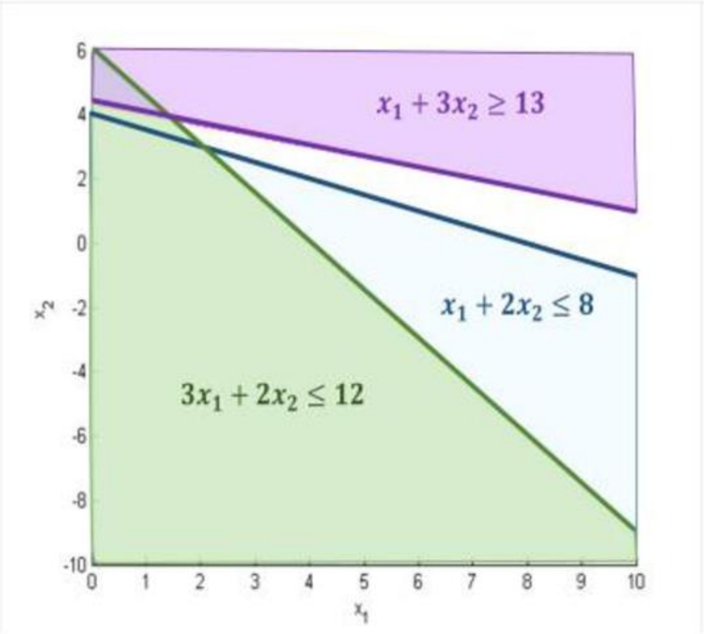


Optimal solution
$x_1 + 6x_2 = \boxed{1900}$
Optimal value
$x_1 + 6x_2 = 1500$
$x_1 + 6x_2 = 1200$
$x_1 + 6x_2 = 600$
Level sets of the
objective function
x_2
x_1
0
50
100
150
200
250
300
350
400

Case 1

$$\begin{aligned} &\textit{minimize} \quad x_1 + x_2 \\ &s.t. \quad x_1 + 2x_2 \le 8 \\ &\qquad 3x_1 + 2x_2 \le 12 \\ &\qquad x_1 + 3x_2 \ge 13 \\ &\qquad x_1 \ge 0 \end{aligned}$$

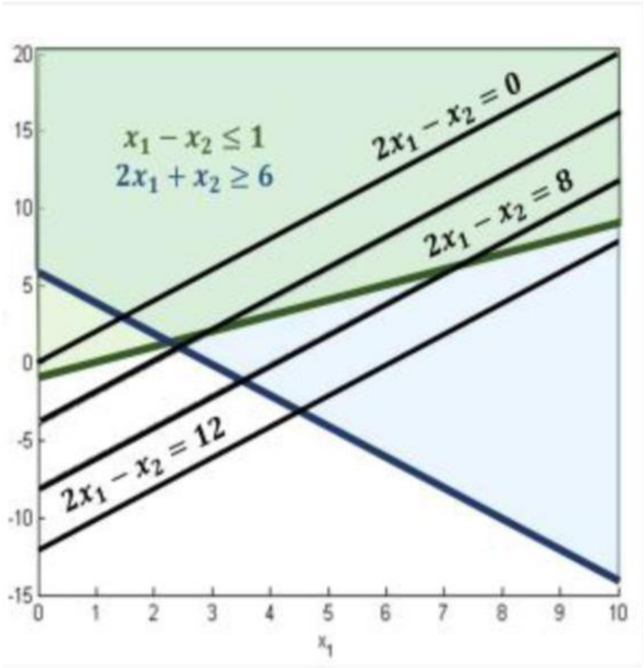
Infeasible



Case 2

$$\begin{array}{ll} minimize & 2x_1 - x_2 \\ s.t. & x_1 - x_2 \le 1 \\ & 2x_1 + x_2 \ge 6 \end{array}$$

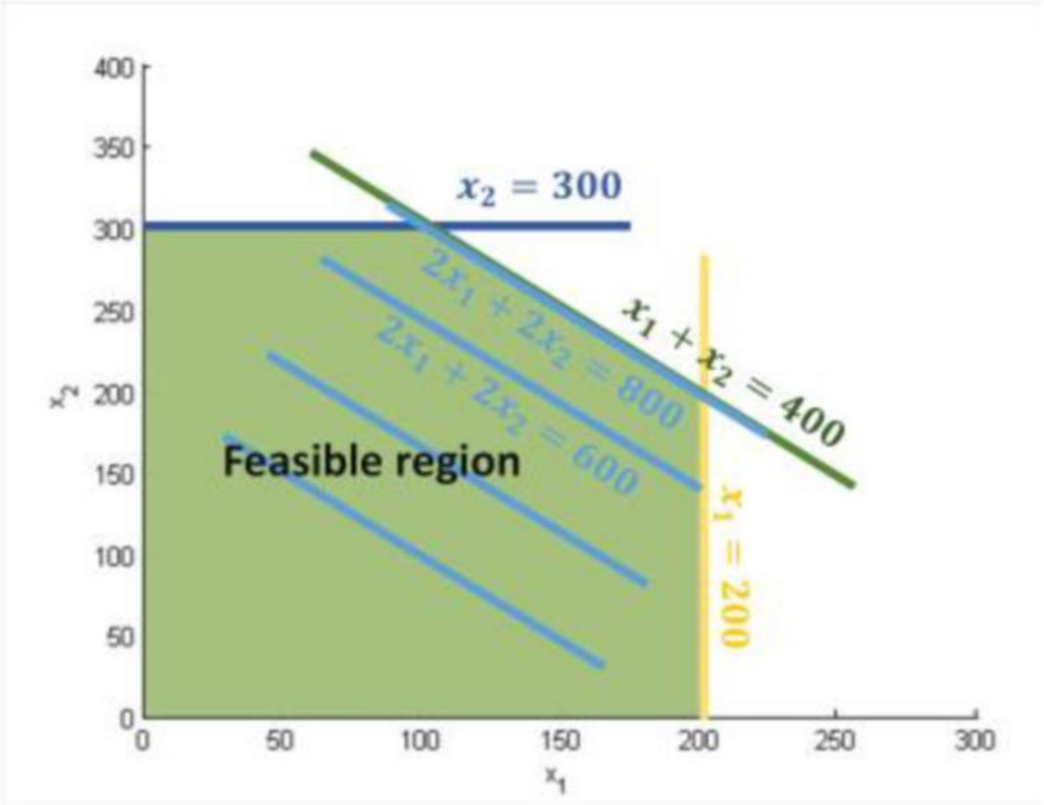
Unbounded



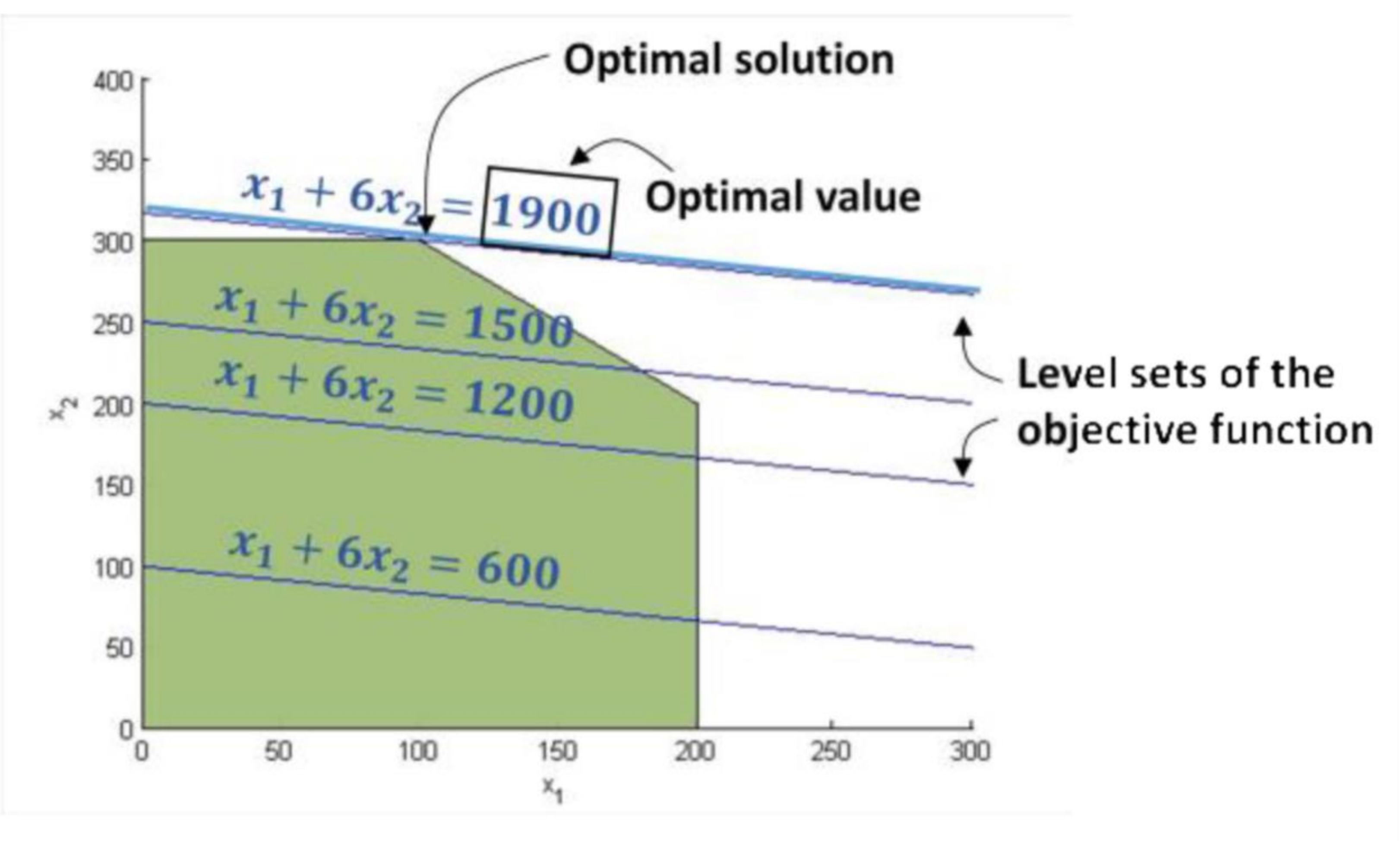
Case 3

Infinite number of optimal solutions

$maximize$	$2x_1 + 2x_2$
$s.t.$	$x_1 \leq 200$
	$x_2 \leq 300$
	$x_1 + x_2 \leq 400$
	$x_1,\ x_2 \geq 0$

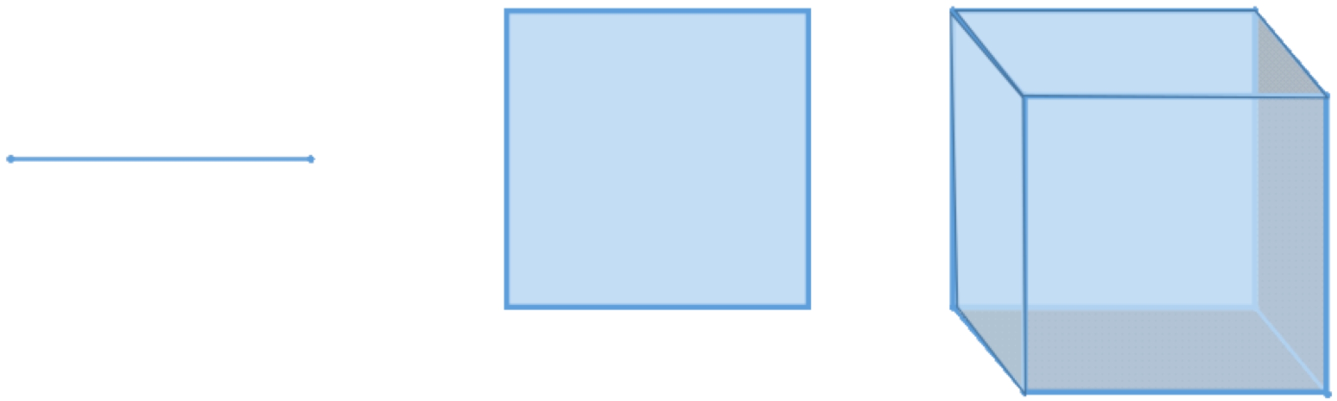


Looking at Extreme Points?



Too many Extreme Points

- consider the constraints $-1 \leq x_i \leq 1, i = 1, \ldots, n$
- Then we have in general $2n$ inequalities, but 2^n extreme points.



n=1
2 inequalities
2 extreme points

n=2
4 inequalities
4 extreme points

n=3
6 inequalities
8 extreme points