Linear Programming for Machine Learning

Yinyu Ye
Stanford University and CUHKSZ (Sabbatical Leave)
Currently Visiting CUHK and HK PolyU
https://web.stanford.edu/class/msande211x/handout.shtml

Read Chapter 1.1, 1.2, 2.1, 2.2, Appendices A, B&D
of the Text-Book

1^{st} Day Questions

- Assistant Instructor: Zaikun Zhang

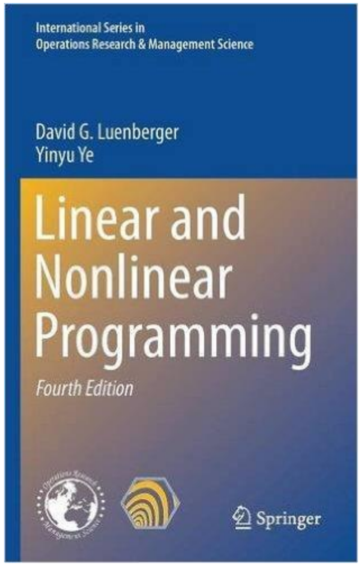
Websites:

https://web.stanford.edu/class/msande211x/handout.shtml

And

https://web.stanford.edu/class/msande211x/assignment.shtml

- Textbook: Linear and Nonlinear Programming (LY 5^{th} edition)



- Prerequisite: calculus and linear algebra classes
- The software use will help: Solvers in Matlab, R, Python or other public free software. It is mostly a "PAPER AND PENCIL" class!
- Form a "diversified" study/project group

Mathematical Optimization Model

- Often consider the common quantitative model of data/decision/management science & engineering:

  - Maximize or Minimize $f(\boldsymbol{x})$
    for all $\boldsymbol{x} \in$ some set X

- Decision variables represented by a vector $\boldsymbol{x}$, Objective function $f(\boldsymbol{x})$, Constraint set X

- Applications in:
  - Applied Science, Engineering, Economics, Finance, Medicine, Statistics, Business
  - General Decision and Policy Making

- The famous Eighteenth Century Swiss mathematician and physicist Leonhard Euler (1707-1783) proclaimed that "…nothing at all takes place in the Universe in which some rule of maximum or minimum does not appear."

The Prototypical Optimization Problem

Max (or Max): $f(\mathbf{x})$

s.t. :

$$h_1(\mathbf{x}) = 0$$

$$\ldots$$

$$h_m(\mathbf{x}) = 0$$

$$g_1(\mathbf{x}) \leq 0$$

$$\ldots$$

$$g_r(\mathbf{x}) \leq 0$$

The Function could be:

x_1+2x_2, $x^2+2xy+2y^2$, $xln(x)+e^y$, $|x|+max\{x,y\}$, etc

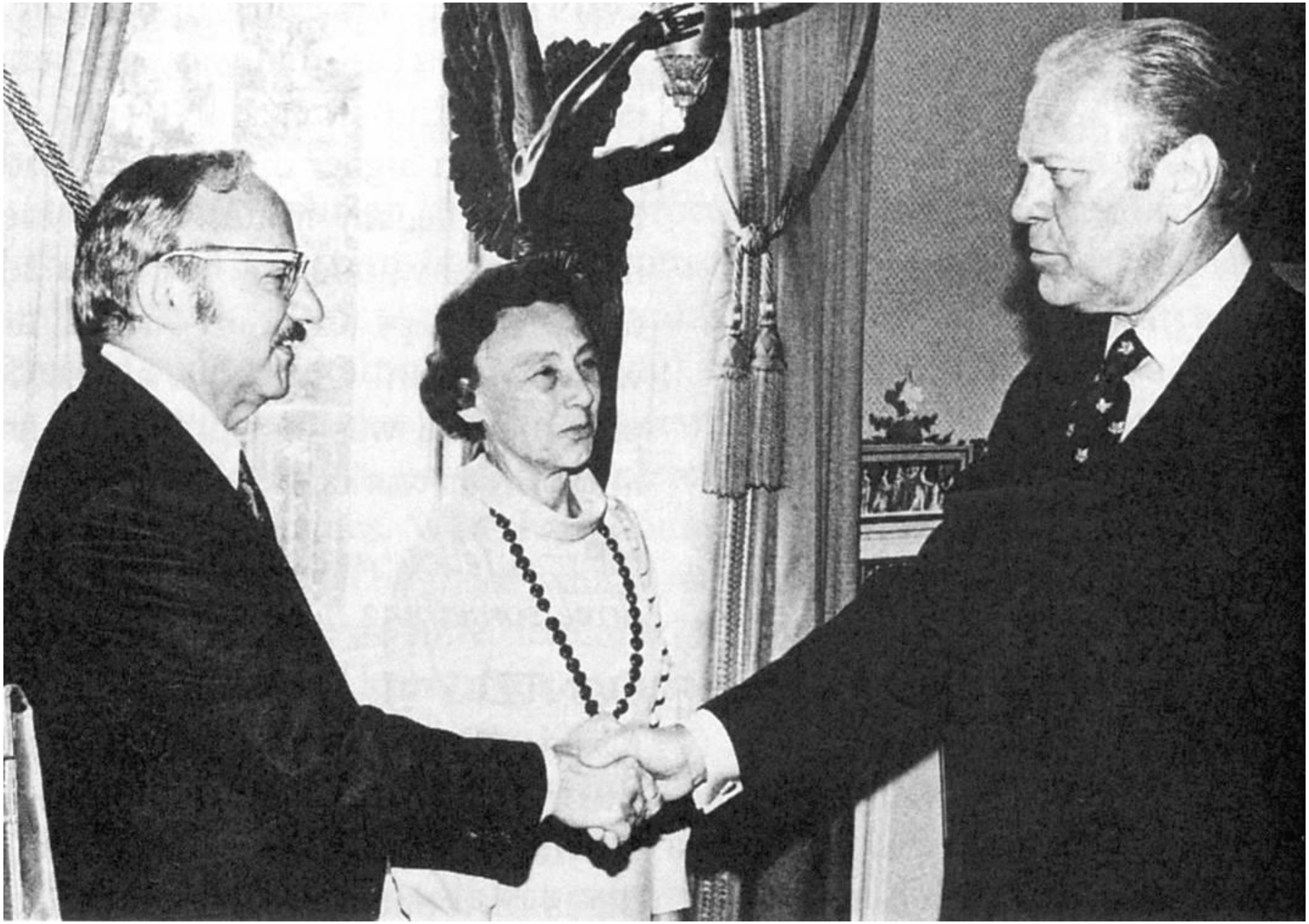
Linear Programming/Optimization: **all functions are linear/affine**

Linear Programming

- Why do we study LP's
  - Not just because solving non-linear problems are difficult
  - But also real-world Machine-Learning problems are often formulated as linear equations and inequalities
    - Either because they indeed are linear
    - Or because it is unclear how to represent them and linear is an intuitive compromise
  - A stepping stone for solving more complicated nonlinear optimization problems, which you would see later.

LP Giants won Nobel Prize…

… and National Medal of Science



Contents Covered in this Course

- **Linear Optimization** (Programming)
  - Model
  - Math Preliminaries
- **LP in Machine Learning**
  - Support Vector Machine
  - Information Markets
  - Wasserstein Barry Center
  - Reinforcement Learning

What do you learn?

- Models – the Art: intuition and common sense
  - How formulate real problems using quantitative models
- Little Theory – the Science: theorems, geometries and universal rules
  - Necessary and Sufficient Conditions that must be true for the optimality of different classes of problems.
- Applications – AI, Machine Learning and Data Science
  - SVM, the Wasserstain barycenter, Reinforced learning/MDP, Information market,…

Art of Modeling, Formulation & Vocabulary

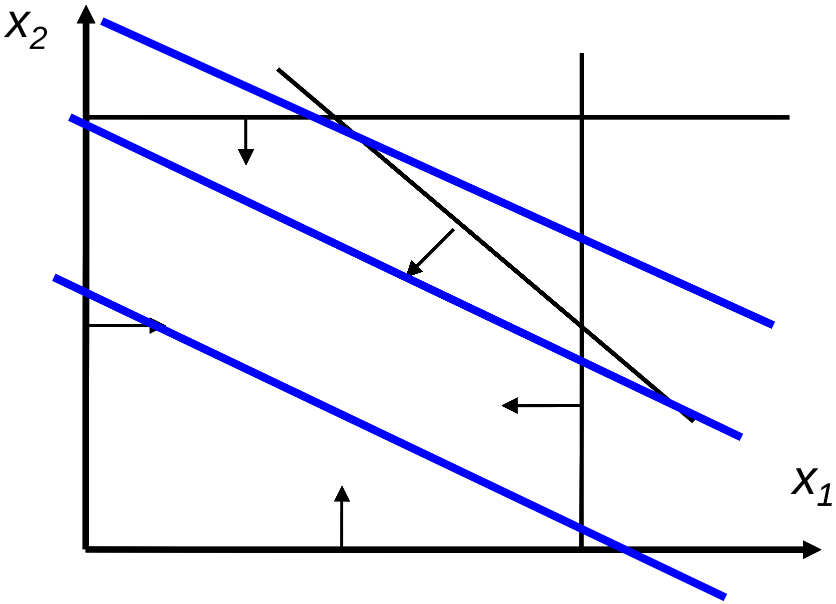
- Decision Variables $\boldsymbol{x} \in R^n$, yet to be decide
- Data/Coefficients, $\boldsymbol{c} \in R^n$, that are given and fixed
- Objective inner product $f = \boldsymbol{c}^T\boldsymbol{x}$: $R^n \rightarrow R$
- Constraint Set $X \subset R^n$
- Feasible solution $\boldsymbol{x} \in X$
- Optimal solution $\boldsymbol{x}^* \in X^*$
- Optimal value $z^* = f(\boldsymbol{x}^*)$

LP Example 1: Resource Allocation/Production Management

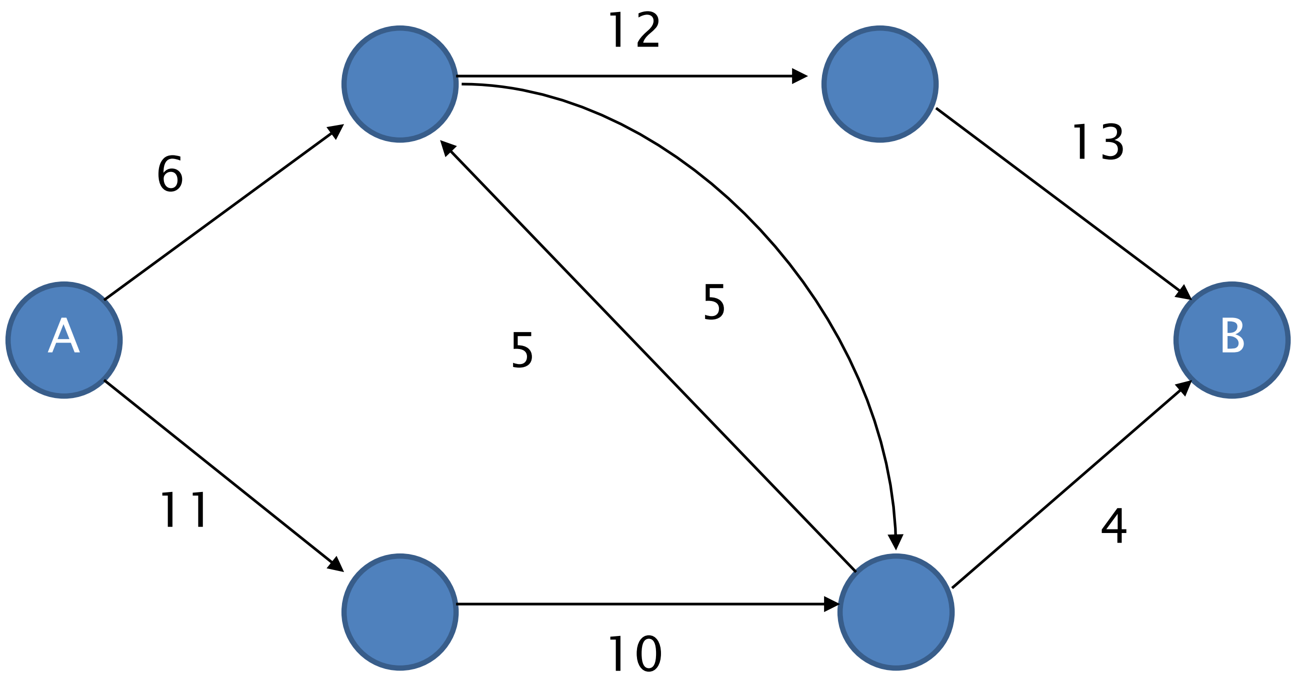
The Wyndor Glass Co. is a producer of high-quality glass products. It has three plants. Aluminum frames and hardware are made in Plant 1, wood frames are made in Plant 2, and Plant 3 is used to produce glass and assemble the products. Wyndor produces two products which require the resources of the three plants as follows:

Plant	Aluminum	Wood	Resources
1	1	0	100
2	0	2	200
3	1	1	150
Unit Profit	$1000	$2000	

$$
\begin{array}{ll}
\max & x_1 + 2x_2 \\
\text{s.t.} & x_1 \le 1, \\
 & 2x_2 \le 2, \\
 & x_1 + x_2 \le 1.5, \\
 & x_1, x_2 \ge 0
\end{array}
$$

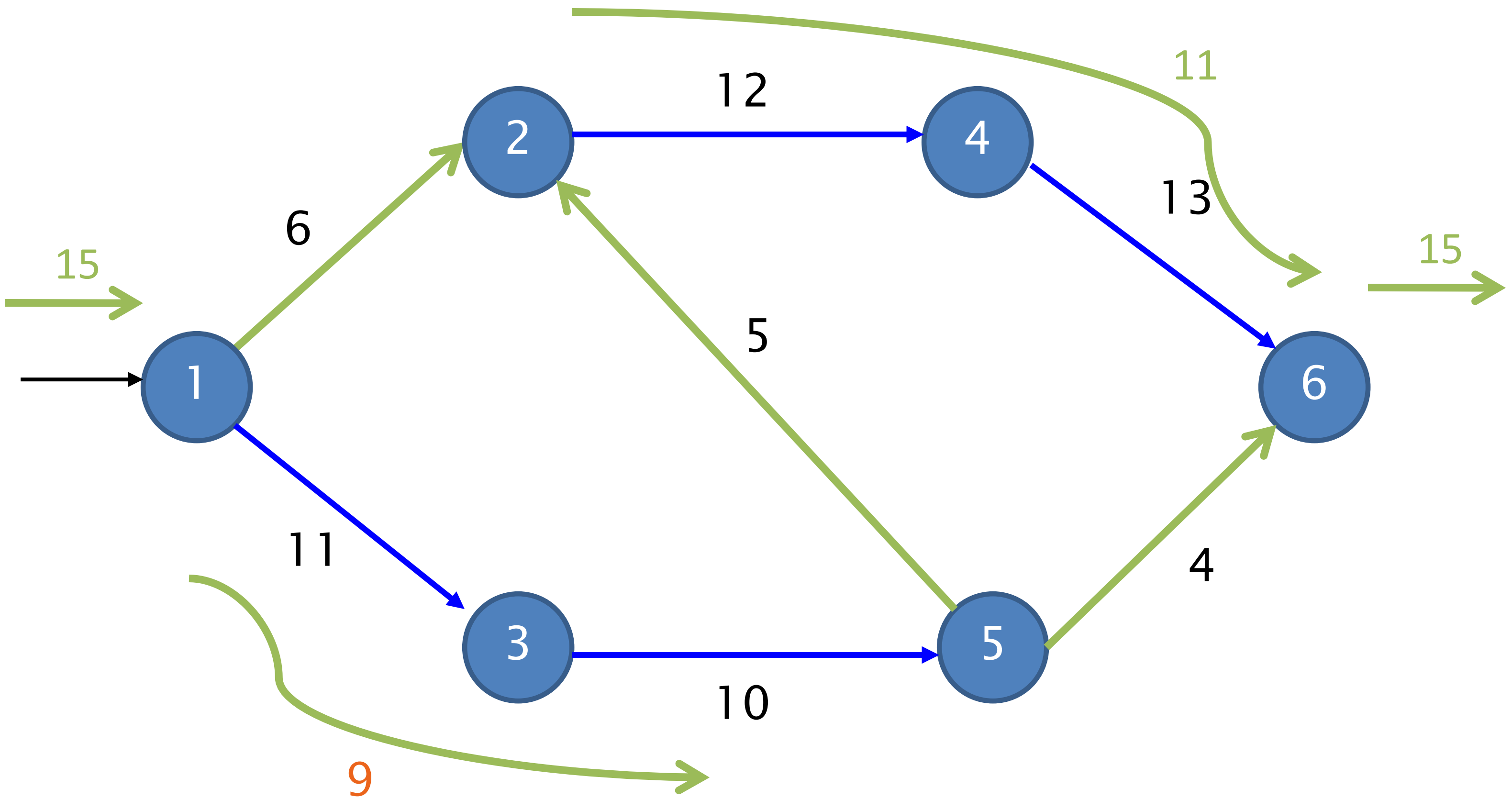


LP Example 2: Maximum Flow



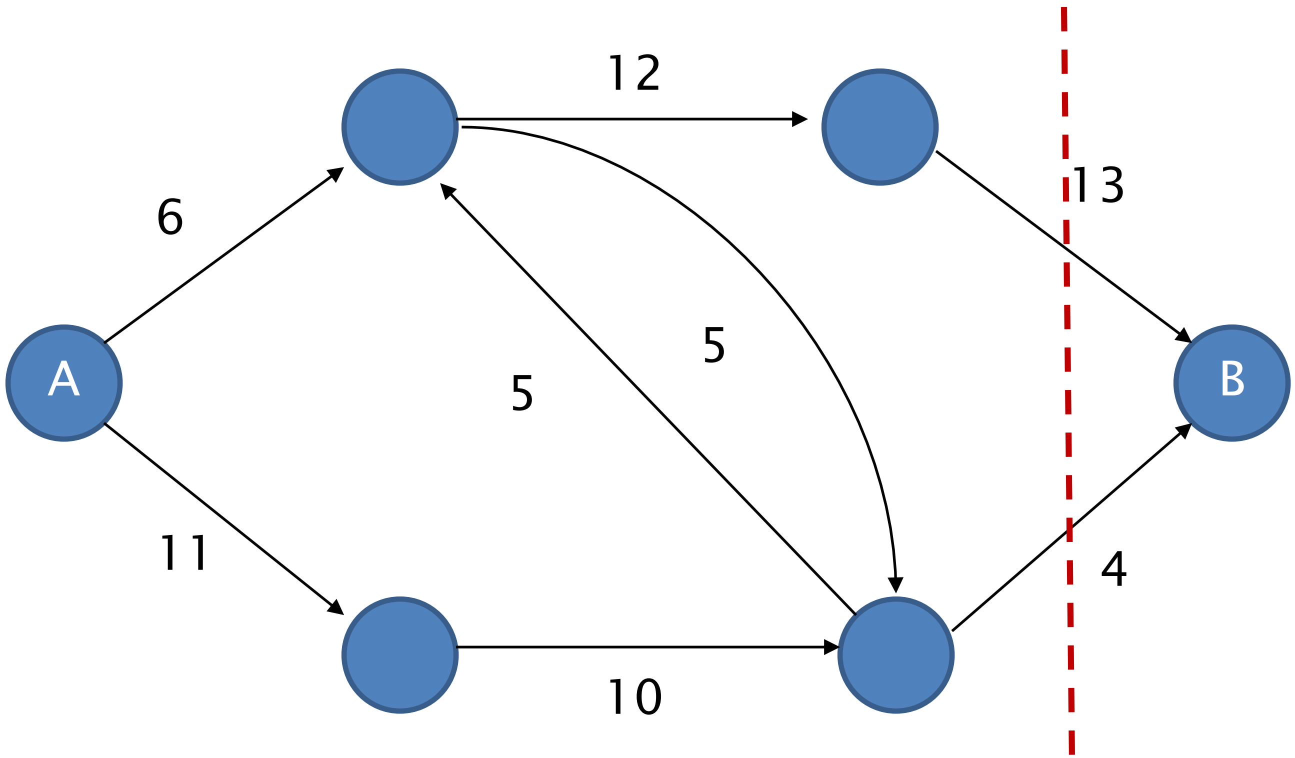
How much flow per unit time can travel from A to B, given that each of the directed connecting routes have flow limits/capacities?

Maximum Flow by Inspection



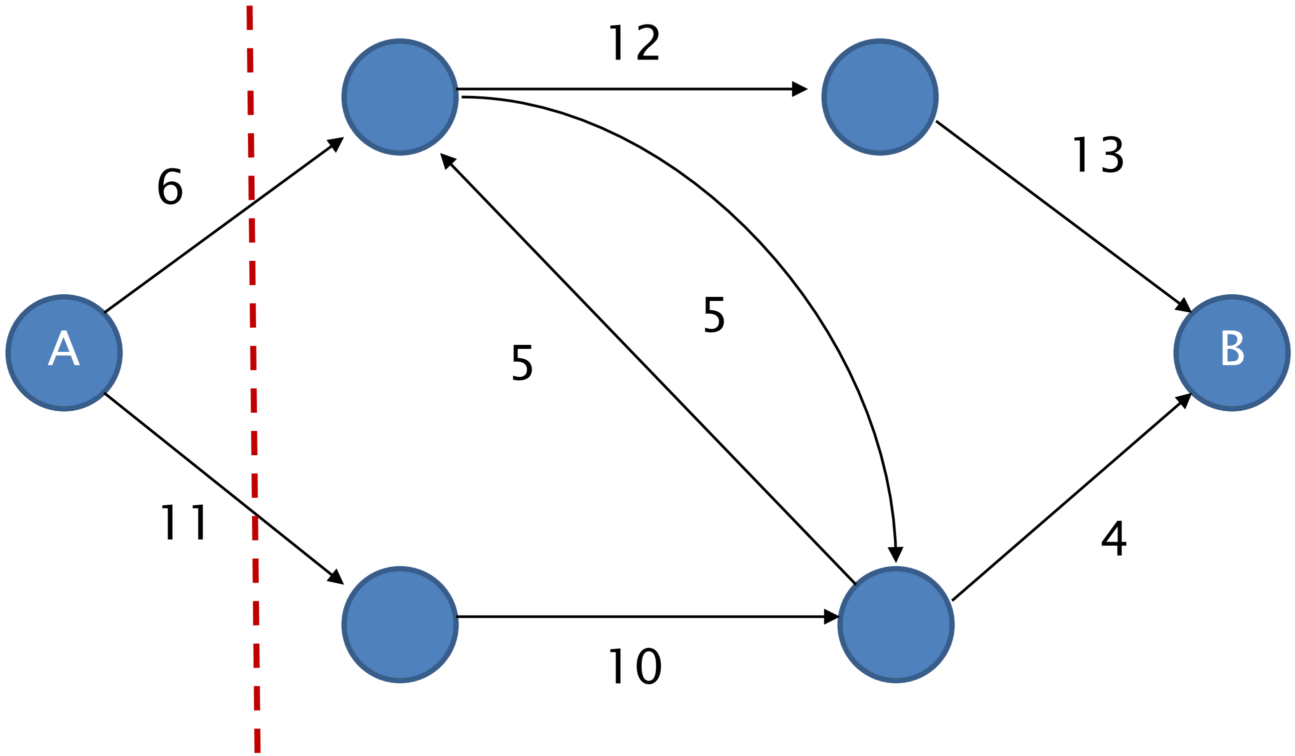
How to certify that 15 is maximal?

Cut in Maximum Flow I



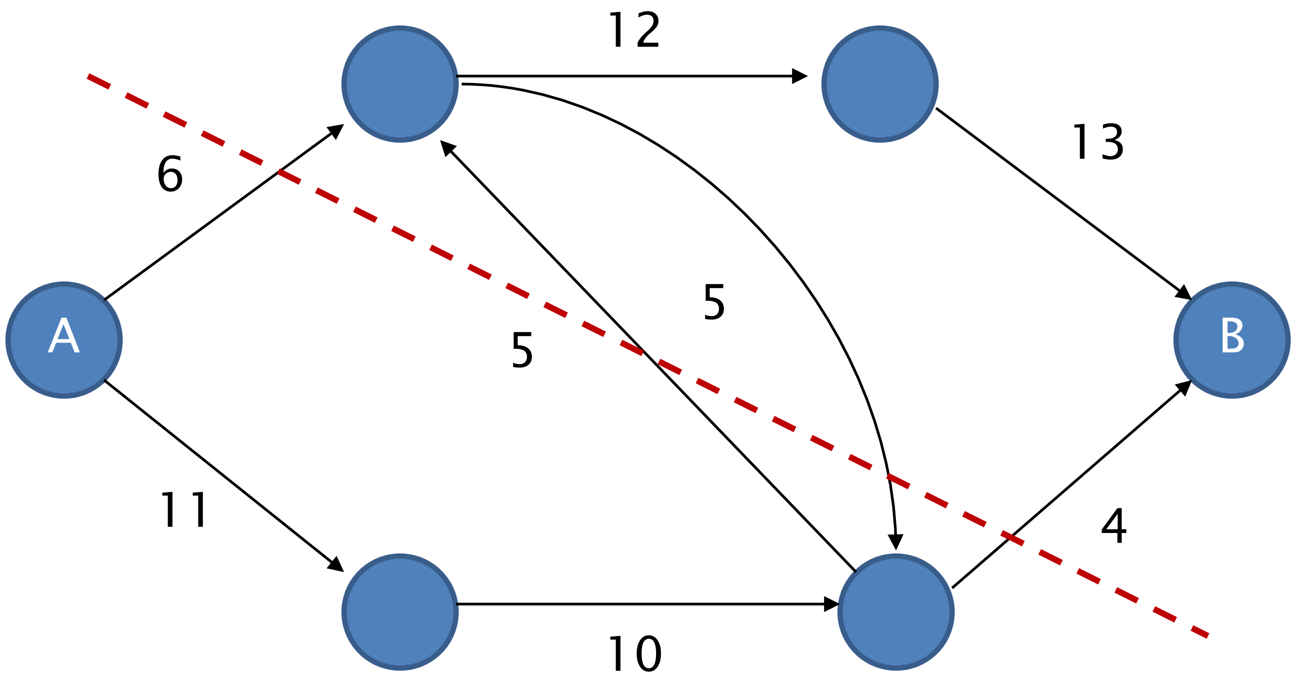
Cut value from Source site to Sink site=17

Cut in Maximum Flow II



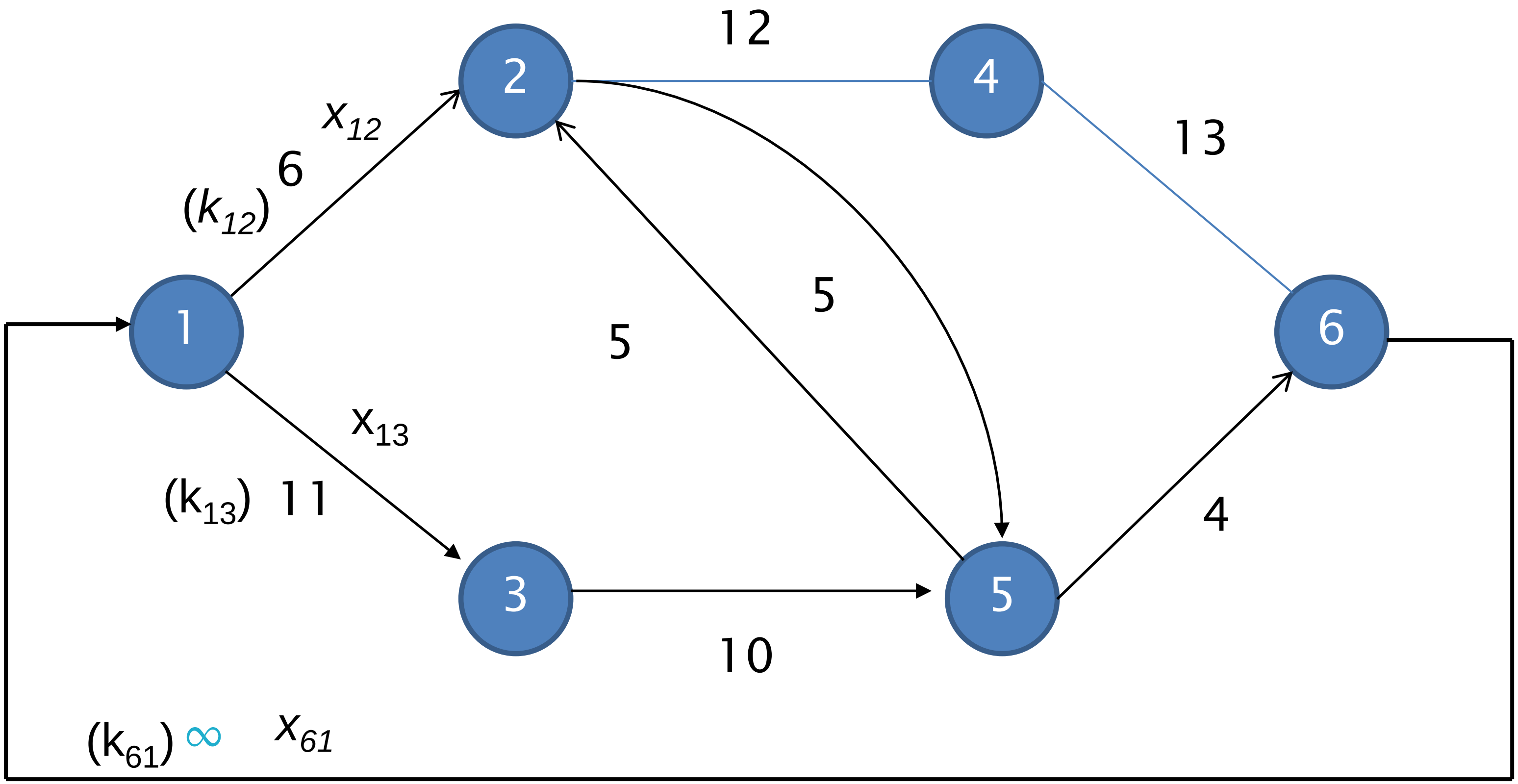
Cut value from Source site to Sink site=17

Cut in Maximum Flow III



Cut value from Source site to Sink site=15

Data points classification application in Machine Learning and Data Science



$$\max \quad x_{61}$$

$$\text{s.t.} \quad \sum_k x_{ki} = \sum_j x_{ij} \quad \forall\, i = 1,2,3,4,5,6$$

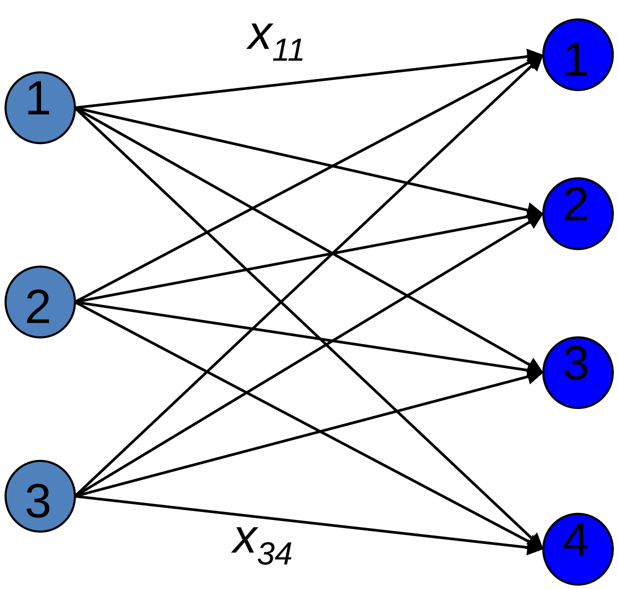
$$0 \le x_{ij} \le k_{ij} \quad \forall\, i, j = 1,2,3,4,5,6$$

inflow

outflow

LP Example 3: Transportation and Assignment

	Retailer 1	Retailer 2	Retailer 3	Retailer 4	SUPPLY
Warehouse 1	**12 (c_{11})**	**13**	**4**	**6**	**500 (s_1)**
Warehouse 2	**6**	**4**	**10**	**11**	**700 (s_2)**
Warehouse 3	**10**	**9**	**12**	**14 (c_{34})**	**800 (s_3)**
DEMAND	**400 (d_1)**	**900 (d_2)**	**200 (d_3)**	**500 (d_4)**	**2000 (s_4)**



$$\min \quad \sum_{i=1}^{3}\sum_{j=1}^{4} c_{ij}x_{ij}$$

Abstract Model

$$\text{s.t.} \quad \sum_{j=1}^{4} x_{ij} = s_i, \quad \forall\, i = 1,2,3$$

$$\sum_{i=1}^{3} x_{ij} = d_j, \quad \forall\, j = 1,2,3,4$$

$$x_{ij} \geq 0, \quad \forall\, i, j$$

Inventory Planning: ***s*** is part of the decision vars.

Machine Learning: The Wassestein Barycenter Problem I

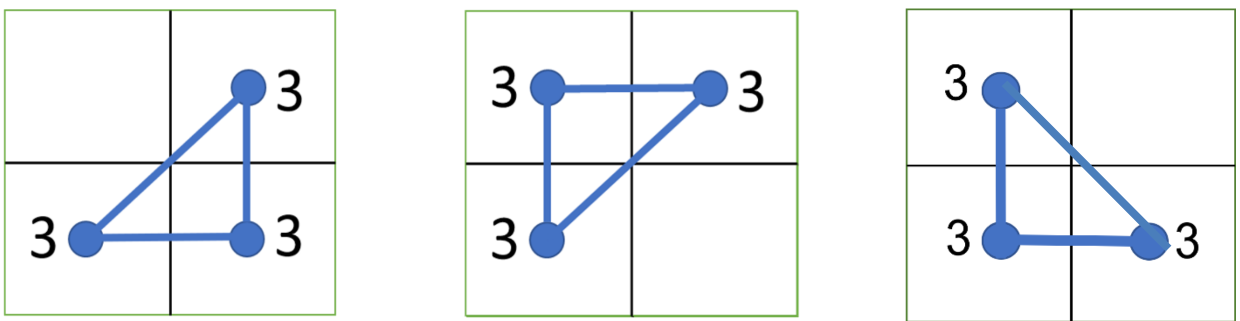
The minimal transportation cost in Data Science is called the Wasserstein distance between a supply distribution and a demand distribution.

The **Wasserstein Barycenter Problem** is to find a distribution/points such that the sum of its Wasserstein distances to each of a set of distributions/points would be minimized

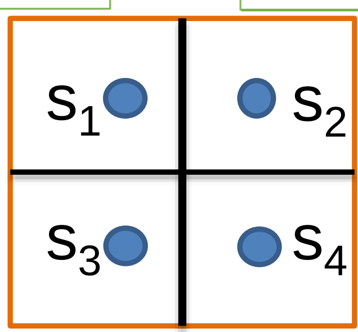
$$\min_s \sum_k \mathbf{WD(s, d^k)} \text{ s.t. total mass constraint}$$

WD(s, d^k)=

$$\begin{aligned} \min \quad & \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}x_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{N} x_{ij} = s_i, \quad \forall\ i = 1,\dots,N \\ & \sum_{i=1}^{N} x_{ij} = d_j, \quad \forall\ j = 1,\dots,N \\ & x_{ij} \geq 0, \qquad \forall\ i,j \end{aligned}$$



Three possible demand distribution scenario of 4 cities



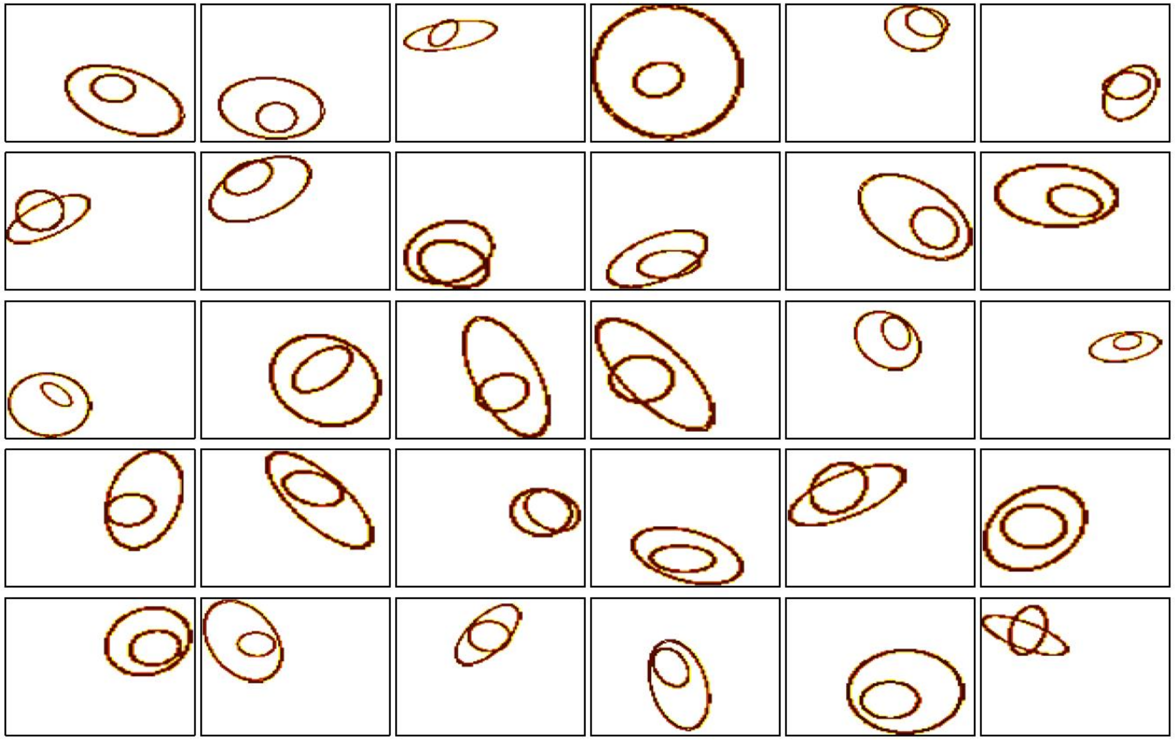
Constraints:

$s_1+s_2+s_3+s_4=9$

$(s_1,s_2,s_3,s_4)>=0$

$$C=\begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}$$

Machine Learning: The Wassestein Barycenter Problem II



What is the best ``mean or consensus‘’ image from a set of images (pixel distributions)?

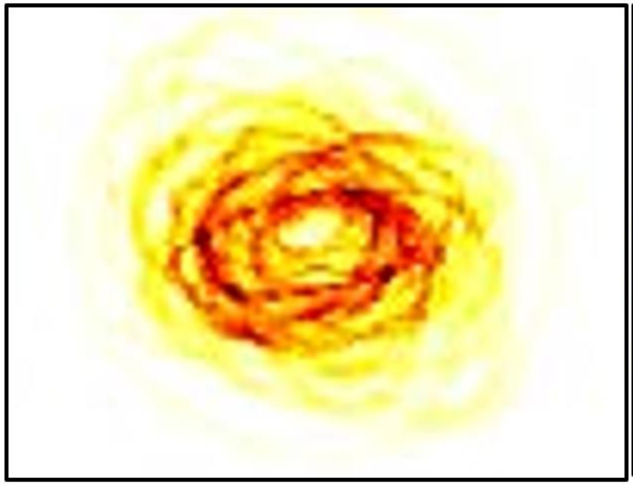
- Simple average
- Simple average after re-centering
- The Wasserstein Barycenter of the set of images (self re-center and rotation)

Machine Learning: The Wassestein Barycenter Problem III

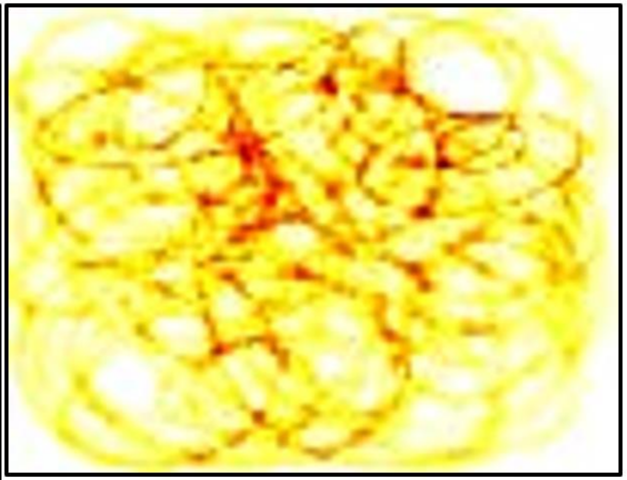
The simple avarage of n points is

$$s=(\sum_k d^k)/n \quad \text{or} \quad \min_s \sum_k (\| s - d^k \|_2)^2$$

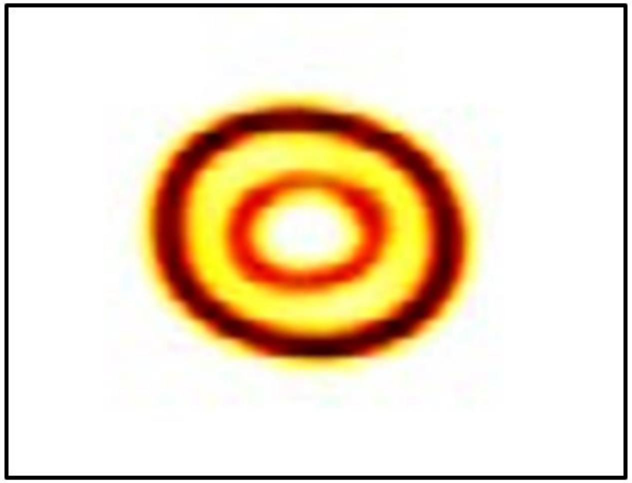
The **Wasserstein Barycenter Problem** is to find a distribution/points such that the sum of its Wasserstein distances to each of a set of distributions/points would be minimized (self re-center and rotation).



Simple average after re-centering



Simple average



the Barycenter image

LP Example 4: Electric Vehicle Charging Schedule and Inventory Control

	Period 1	Period 2	Period 3	Period 4	Period 5
Price ($)	1.25 (c_1)	1.35 (c_2)	1.25 (c_3)	1.10 (c_4)	1.05 (c_5)
Demand (kw)	60 (d_1)	110 (d_2)	100 (d_3)	40 (d_4)	0 (d_5)
Charging (kw)	x_1	x_2	x_3	x_4	x_5
Inventory (I_0)	I_1	I_2	I_3	I_4	I_5

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{5} c_i x_i \\
\text{s.t.} \quad & I_{i-1} + x_i - d_i = I_i, \quad \forall\, i = 1,2,3,4,5 \\
& I_{i-1} + x_i \le K, \quad \forall\, i = 1,2,3,4,5 \\
& x_i \ge 0, I_i \ge 0, \quad \forall\, i.
\end{aligned}
$$

LP Example 4: When Discharge is Allowed

	Period 1	Period 2	Period 3	Period 4	Period 5
Price ($)	1.25 (c_1)	1.35 (c_2)	1.25 (c_3)	1.10 (c_4)	1.05 (c_5)
Demand (kw)	60 (d_1)	110 (d_2)	100 (d_3)	40 (d_4)	0 (d_5)
Charging (kw)	x_1	x_2	x_3	x_4	x_5
Inventory (I_0)	I_1	I_2	I_3	I_4	I_5

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{5} c_i x_i \\
\text{s.t.} \quad & I_{i-1} + x_i - d_i = I_i, \quad \forall\, i = 1,2,3,4,5 \\
& I_{i-1} + x_i \le K, \quad \forall\, i = 1,2,3,4,5 \\
& I_i \ge 0, \quad \forall\, i.
\end{aligned}
$$

Linear Programming Abstract Form

$$\begin{array}{ll} \max & x_1 + 2x_2 \\ \text{s.t.} & x_1 \le 1, \\ & 2x_2 \le 2, \\ & x_1 + x_2 \le 1.5, \\ & x_1 \ge 0, x_2 \ge 0 \end{array}$$

$$\Rightarrow$$

$$\begin{array}{ll} \max & 1\cdot x_1 + 2\cdot x_2 \\ \text{s.t.} & 1\cdot x_1 + 0\cdot x_2 \le 1, \\ & 0\cdot x_1 + 2\cdot x_2 \le 2, \\ & 1\cdot x_1 + 1\cdot x_2 \le 1.5, \\ & x_1 \ge 0, x_2 \ge 0. \end{array}$$

$$\begin{array}{ll} \max & c_1 x_1 + c_2 x_2 \\ \text{s.t.} & a_{11} x_1 + a_{12} x_2 \le b_1 \\ & a_{21} x_1 + a_{22} x_2 \le b_2 \\ & a_{31} x_1 + a_{32} x_2 \le b_3 \\ & x_1 \ge 0, \; x_2 \ge 0.. \end{array}$$

Abstract Linear Programming Model

$$\max\ (\min)\quad c_1 x_1 + c_2 x_2 + \dots + c_n x_n$$

$$\text{s.t.}\quad \begin{array}{l} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n \ \{\le, =, \ge\}\ b_1 \\ a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n \ \{\le, =, \ge\}\ b_2 \\ \qquad \dots \qquad \dots \\ a_{m1} x_1 + a_{m2} x_2 + \dots + a_{mn} x_n \{\le, =, \ge\}\ b_m \\ x_1 \ge 0,\ x_2 \text{ free},\ \dots,\ x_n \le 0. \end{array}$$

Input: $c_1, \dots, c_2$, objective coef.; $b_1, \dots, b_m$, constraint right-hand-side coef.

$a_{ij}, i = 1, \dots, m; j = 1, \dots, n$, constraint left-hand-side table or matrix coef.

Output: $x_1, \dots, x_n$, decision v ariables

LP in Compact Matrix Form

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{pmatrix}, c = \begin{pmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

Coefficient matrix

RHS vector

Obj. vector

decision vector

$$\begin{array}{ll} \text{max(min)} & c^T x \\ \text{s.t.} & A\,x\ \{\le, =, \ge\}\ b\,, \\ & \quad x\ \{\ge, \le\}\ 0 \text{ or free.} \end{array}$$

Some Facts of Linear Programming

- Add a constant to the objective function does not change the optimality
- Scale the objective coefficients does not change the optimality
- Scale the right-hand-side coefficients does not change the optimality but the solution scaled accordingly
- Reorder the decision variables (together with their corresponding objective and constraint coefficients) does not change the optimality
- Reorder the constraints (together with their right-hand-side coefficients) does not change the optimality
- Multiply both sides of an equality constraint by a constant does not change the optimality
- Pre-multiply both sides of all equality constraints by a non-singular matrix does not change the optimality