Modeling, Formulations and Applications

LI-PING ZHANG

Department of Mathematical Sciences

Tsinghua University, Beijing 100084

Office: New Science Building #A302, Tel: 62798531

E-mail: lipingzhang@tsinghua.edu.cn

Outline

- Formulation of Optimization Models
- Air Traffic Control; Transportation Problem 运输问题和空管问题
- Addwords Allocation; Supporting Vector Machine 把两类数据分开取优化函数
广告投放点击量问题
- Combinatorial Auction 赌球问题
- The Traveling Salesman Problem (TSP) 旅行商问题
- Minimize Max-TSP-Tour and Data Clustering Model 警察分区管制问题(数据聚类
- Data Fitting 数据拟合
- Fisher's Exchange Market
- Portfolio Management 投资组合管理

Formulation of Optimization Models: Four-Step Rule

- Sort out data and parameters from the verbal description
- Define the set of decision variables
- Formulate the objective function of data and decision variables
- Set up equality and/or inequality constraints

The Art of Modeling

The objective is to distill the real-world problem as accurately and succinctly as possible into a quantitative model.

- Don't want models to be too generalized. Otherwise, might not draw much real world value from your results. For example, analyzing traffic flows assuming every person has the same characteristics.

- Don't want models to be too specific. Otherwise, might lose the ability to solve problems or gain insights. For example, trying to analyze traffic flows by modeling every single individual using different assumptions.

Formulation 1: Air Traffic Control

Air plane $j, \quad j = 1, ..., n$ arrives at the airport within the time interval $[a_j, b_j]$ in the order of $1, 2, ..., n$. The airport wants to find the arrival time for each air plane such that the narrowest metering time (inter-arrival time between two consecutive airplanes) is the greatest.

Air Traffic Control continued

Let t_j be the arrival time of plane j. Then

maximize	$\min_{j=1,\ldots,n-1}(t_{j+1}-t_j)$	目标
subject to	$a_j \le t_j \le b_j, \quad j=1,2,\ldots,n.$	**限制**

Air Traffic Control continued

Equivalent smooth formulation:

$$
\begin{aligned}
\textbf{maximize} \quad & \Delta \\
\textbf{subject to} \quad & t_2 - t_1 - \Delta \geq 0, \\
& t_3 - t_2 - \Delta \geq 0, \\
& ..., \\
& t_n - t_{n-1} - \Delta \geq 0, \\
& a_j \leq t_j \leq b_j, \quad j = 1, 2, ..., n.
\end{aligned}
$$

Formulation 2: Transportation Problem

源

There are m origins that contain various amounts of a commodity that must be shipped to n destinations to meet demand requirements. Origin i contains an 汇 amount a_i, and destination j has a requirement of amount b_j. The numbers a_i and b_j, $i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$, are assumed to be nonnegative, and in many applications they are in fact nonnegative integers. There is a unit cost c_{ij} associated with the shipping of the commodity between origins and destinations that satisfied all the requirements and minimizes the shipping cost.

Transportation Problem continued

Let x_{ij} be the amounts of the commodity shipped from origin i to destination j.

Then in mathematical terms the above problem can be expressed as

$$
\begin{array}{lrl}
\min & \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} & \\
\textbf{s.t.} & \sum_{j=1}^{n} x_{ij} & \leq a_i, \ \forall i \\
& \sum_{i=1}^{m} x_{ij} & \geq b_j, \ \forall j \\
& x_{ij} & \geq 0, \ \forall i, j.
\end{array}
$$

张量互补问题或者是博弈问题

Formulation 3: Addwords Allocation

Given n different addwords that m companies are bidding to put their advertisements on the page when the addword is clicked.

An order, i, on a set of addwords includes price/per-click, b_{ij} on word j, and a budget b_i for a certain period.

Suppose the estimated total clicks for the period on word j is s_j.

Addwords Allocation continued

Let x_{ij} be the number of clicks awarded to order i on word j. Then one would like to

$$\begin{array}{rll} \max & \sum_{i,j} b_{ij}x_{ij} & \\ \textbf{s.t.} & \sum_j b_{ij}x_{ij} & \le b_i, \ \forall i \\ & \sum_i x_{ij} & \le s_j, \ \forall j \\ & x_{ij} & \ge 0, \ \forall i, j. \end{array}$$

Formulation 3: Addwords Allocation

Given n different addwords that m companies are bidding to put their advertisements on the page when the addword is clicked.

An order, i, on a set of addwords includes price/per-click, b_{ij} on word j, and a budget b_i for a certain period.

Suppose the estimated total clicks for the period on word j is s_j.

Addwords Allocation continued

Let x_{ij} be the number of clicks awarded to order i on word j. Then one would like to

$$\begin{array}{rll} \max & \sum_{i,j} b_{ij}x_{ij} & \\ \textbf{s.t.} & \sum_{j} b_{ij}x_{ij} & \leq b_i,\ \forall i \\ & \sum_{i} x_{ij} & \leq s_j,\ \forall j \\ & x_{ij} & \geq 0,\ \forall i,j. \end{array}$$

对偶问题为寻求单位点击量的定价
问题

Formulation 4: Supporting Vector Machine

Suppose we have two-class discrimination data. We assign the first class with 1 and the second with -1. A powerful discrimination method is the Supporting Vector Machine (SVM).

Let the data point i be given by $\mathbf{a}_i \in R^d$, $i = 1, ..., n$. With this data set, we have some $\bar{y}_i = 1$ (in the first class) and the rest $\bar{y}_i = -1$ (in the second class).

一般也可以考虑ai是随机变量的情况

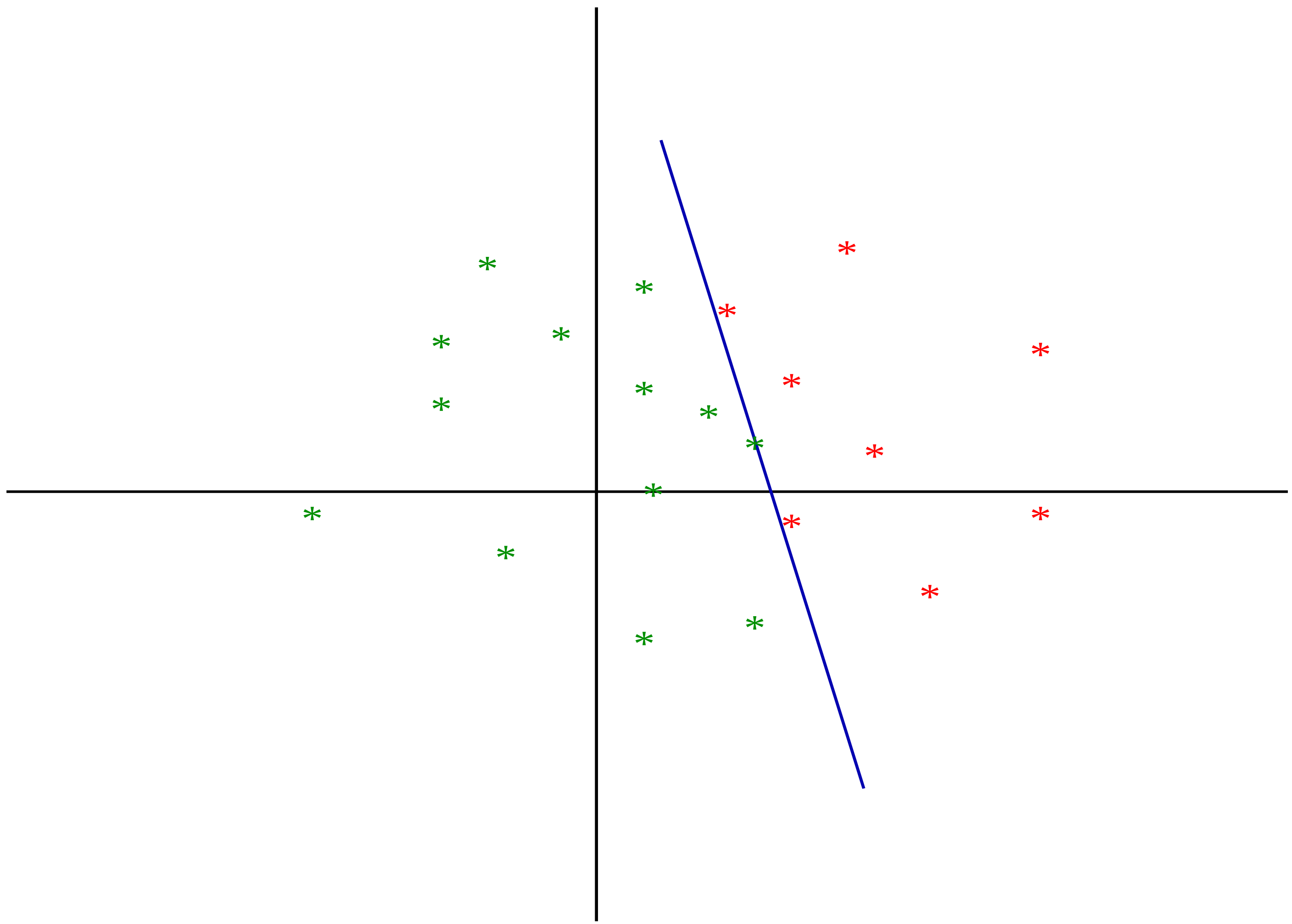

Figure 1: Linear Support Vector Machine

Supporting Vector Machine continued

We wish to find a hyper-plane in R^d to separate $\mathbf{a}_i$s with red from $\mathbf{a}_j$s with green. Mathematically, we wish to find $\omega \in R^d$ and $\beta \in R$ such that

$$\mathbf{a}_i^T \omega + \beta > 1 \quad \forall \{i : \bar{y}_i = 1\}$$

and

$$\mathbf{a}_i^T \omega + \beta < -1 \quad \forall \{i : \bar{y}_i = -1\},$$

that is,

$$\bar{y}_i(\mathbf{a}_i^T \omega + \beta) > 1 \quad \forall i.$$

The hyperplane would be

$$\{\mathbf{x} : \ \mathbf{x}^T \omega + \beta = 0\}.$$

Supporting Vector Machine continued

If a clean separation is possible, we can formulate the problem as a maximization problem:

$$\begin{array}{rl} \textbf{minimize} & \|\omega\|^2 \\ \textbf{subject to} & \bar{y}_i(\mathbf{a}_i^T\omega + \beta) > 1, \quad \forall i. \end{array}$$

目标二次约束一次：二次规划

Supporting Vector Machine continued

A clean separation may not be possible for noisy data. Another formulation of the problem is a minimization problem:

$$\begin{array}{rl} \textbf{minimize} & \|\omega\|^2 + \gamma \cdot \sum_{i=1}^{n} (\xi_i)^2 \quad \text{罚函数} \\ \textbf{subject to} & \bar{y}_i(\mathbf{a}_i^T \omega + \beta) > 1 + \xi_i, \ \forall i. \end{array}$$

综合看待间隔和错化程度的舍取

Formulation 5: Combinatorial Auction I

Given the m different states that are mutually exclusive and exactly one of them will be true at the maturity.

A contract on a state is a paper agreement so that on maturity it is worth a notional $\$w$ if it is on the winning state and worth $\$0$ if is not on the winning state. There are n orders betting on one or a combination of states, with a price limit and a quantity limit.

Combinatorial Auction II: an order

The jth order is given as $(\mathbf{a}_j \in R^m_+,\ \pi_j \in R_+,\ q_j \in R_+)$: $\mathbf{a}_j$ is the combination betting vector where each component is either 1 or 0

$$\mathbf{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \dots \\ a_{mj} \end{pmatrix},$$

where 1 is winning and 0 is non-winning; π_j is the price limit for one such a contract, and q_j is the maximum number of contracts the better like to buy.

World Cup Betting Market

- Market for World Cup Winner

  Assume 5 teams have a chance to win the World Cup: *Argentina, Brazil, Italy, Germany and France*. We'd like to have a standard payout of $1 per share if a participant has a winning order.

- Sample of Combinatorial Orders

Order	Price Limit π	Quantity Limit q	Argentina	Brazil	Italy	Germany	France
1	0.75	10	1	1	1		
2	0.35	5				1	
3	0.40	10	1		1		1
4	0.95	10	1	1	1	1	
5	0.75	5		1		1	

Combinatorial Auction III: qualified orders

Let x_j be the number of contracts awarded to the jth order. Then, the jth bidder will pay the amount $\pi_j \cdot x_j$ and the total amount paid is $\pi^T \mathbf{x}$.

If the ith state is the winning state, then the auction organizer need to pay back

$$w \cdot \left(\sum_{j=1}^{n} a_{ij} x_j \right) = w \cdot \mathbf{a}_{i\cdot} \mathbf{x}$$

where $\mathbf{a}_{i\cdot} \mathbf{x}$ is the number of contracts in the ith state.

Combinatorial Auction Pricing IV: Robust model

The question is, how to decide $\mathbf{x} \in R^n$, that is, to check if the auction is profitable or not, and how to fill the qualified orders if it is.

考虑最坏的情况

$$\begin{array}{lrl} \max & \pi^T \mathbf{x} - w \cdot \max_i \{\mathbf{a}_i . \mathbf{x}\} & \\ \textbf{s.t.} & \mathbf{x} & \leq \mathbf{q}, \\ & \mathbf{x} & \geq \mathbf{0}. \end{array}$$

$\pi^T \mathbf{x}$: the amount can be collected.

Combinatorial Auction Pricing IV: Robust model

$$
\begin{array}{lrl}
\max & \pi^T \mathbf{x} - w \cdot \max(A\mathbf{x}) & \\
\textbf{s.t.} & \mathbf{x} & \leq \mathbf{q}, \\
 & \mathbf{x} & \geq \mathbf{0}.
\end{array}
$$

Combinatorial Auction Pricing V: linear model

$$\begin{aligned} \max \quad & \pi^T \mathbf{x} - w \cdot s \\ \textbf{s.t.} \quad & A\mathbf{x} - \mathbf{e} \cdot s \le \mathbf{0}, \\ & \mathbf{x} \le \mathbf{q}, \\ & \mathbf{x} \ge \mathbf{0}. \end{aligned}$$

$\pi^T \mathbf{x}$: the potentially optimistic amount can be collected.

$w \cdot s$: the worst-case amount need to pay back (cost).

Numerical Result for World Cup Betting Market

Order	π	q	Argentina	Brazil	Italy	Germany	France	Qualified Order $\mathbf{x}$
1	0.75	10	1	1	1			5
2	0.35	5				1		5
3	0.40	10	1		1		1	5
4	0.95	10	1	1	1	1		
5	0.75	5		1		1		5

Formulation 6: The Traveling Salesman Problem

Given a finite set of points $V = \{1, 2, \ldots, n\}$ and a cost d_{ij} of travel between each pair $i, j \in V$. A tour is a circuit that passes exactly once through each point in V. The traveling salesman problem (TSP) is to find a tour of minimal cost.

The TSP can be modeled as a graph problem by considering a complete graph $G = (V, E)$, and assigning each edge $ij \in E$ the cost d_{ij}. A tour is then a circuit in G that meets every node. In this context, tours are sometimes called Hamiltonian circuits.

TSP continued

Let x_{ij} be the binary decision variable to choose the edge $ij \in E$. Then

$$\begin{array}{rl}
\textbf{minimize} & \sum_{i \neq j} d_{ij} x_{ij} \\
\textbf{subject to} & \sum_{j=1}^{n} x_{ij} = 1, \ i = 1, 2, \ldots, n, \\
& \sum_{i=1}^{n} x_{ij} = 1, \ j = 1, 2, \ldots, n, \\
& \sum_{i,j \in S} x_{ij} \leq |S| - 1, \ 2 \leq |S| \leq n - 2, \ S \subset V, \\
& x_{ij} \in \{0, 1\}, \ i, j = 1, 2, \ldots, n, \ i \neq j.
\end{array}$$

每个顶点经过且仅经过一次

不能是多个小圈

Enumeration method

枚举法

An array of n points is a feasible solution. We should enumerate $(n-1)!$ times. The computation time is listed in the following table.

$\|V\|$	24	25	26	27	28	29	30	31
Comp.T	1s	24s	10m	4.3h	4.9days	136.5days	10.8years	325years

Dynamic programming method

动态规划问题

Given a set $S \subseteq \{2, 3, \ldots, n\}$ and $k \in S$, compute

$$C(S, k) = \min_{m \in S \backslash \{k\}} \Big[C(S \backslash \{k\}, m) + d_{mk} \Big],$$

where $C(S, k)$ is the minimal cost on the path from 1 to k, passing exactly once through each point in S.

When $S = \{2, \ldots, n\}$, $\min_{k \in S}\{C(S, k) + d_{k1}\}$ is optimal.

The example of TSP

Let $V = \{1, 2, 3, 4\}$ and the distance matrix

$$D = \begin{pmatrix} 0 & 8 & 5 & 6 \\ 6 & 0 & 8 & 5 \\ 7 & 9 & 0 & 5 \\ 9 & 7 & 8 & 0 \end{pmatrix}.$$

When there is an element in S, we have

$$C(\{2\}, 2) = d_{12} = 8,\ C(\{3\}, 3) = d_{13} = 5,\ C(\{4\}, 4) = d_{14} = 6.$$

When there are two elements in S, we have

$$
\begin{aligned}
C(\{2,3\},2) &= C(\{3\},3) + d_{32} = 5 + 9 = 14, \\
C(\{2,3\},3) &= C(\{2\},2) + d_{23} = 8 + 8 = 16, \\
C(\{2,4\},2) &= C(\{4\},4) + d_{42} = 6 + 7 = 13, \\
C(\{2,4\},4) &= C(\{2\},2) + d_{24} = 8 + 5 = 13, \\
C(\{3,4\},3) &= C(\{4\},4) + d_{43} = 6 + 8 = 14, \\
C(\{3,4\},4) &= C(\{3\},3) + d_{34} = 5 + 5 = 10.
\end{aligned}
$$

When $S = \{2,3,4\}$, we have

$$\begin{aligned} C(\{2,3,4\},2) &= \min\Big\{C(\{3,4\},3) + d_{32}, C(\{3,4\},4) + d_{42}\Big\} \\ &= \min\Big\{14+9, 10+7\Big\} = 17, \\ C(\{2,3,4\},3) &= \min\Big\{C(\{2,4\},2) + d_{23}, C(\{2,4\},4) + d_{43}\Big\} \\ &= \min\Big\{13+8, 13+8\Big\} = 21, \\ C(\{2,3,4\},4) &= \min\Big\{C(\{2,3\},2) + d_{24}, C(\{2,3\},3) + d_{34}\Big\} \\ &= \min\Big\{14+5, 16+5\Big\} = 19. \end{aligned}$$

Thus,

$$\begin{aligned} \min_{k \in S}\Big\{C(S,k) + d_{k1}\Big\} &= \min\Big\{C(S,2) + d_{21}, C(S,3) + d_{31}, C(S,4) + d_{41}\Big\} \\ &= \min\Big\{17+6, 21+7, 19+9\Big\} = 23. \end{aligned}$$

The tour of minimal cost is $1 \to 3 \to 4 \to 2 \to 1$.

Formulation 7: Minimize max-TSP-tour

Given two-dimension sensor points $\mathbf{a}_j$, $j = 1, ..., n$, and the vehicle locations $\mathbf{b}_i$, $i = 1, ..., m$; find the best m clusters assigned to each vehicle such that the maximum of the TSP (Traveling Salesman Problem) tour length is minimized.

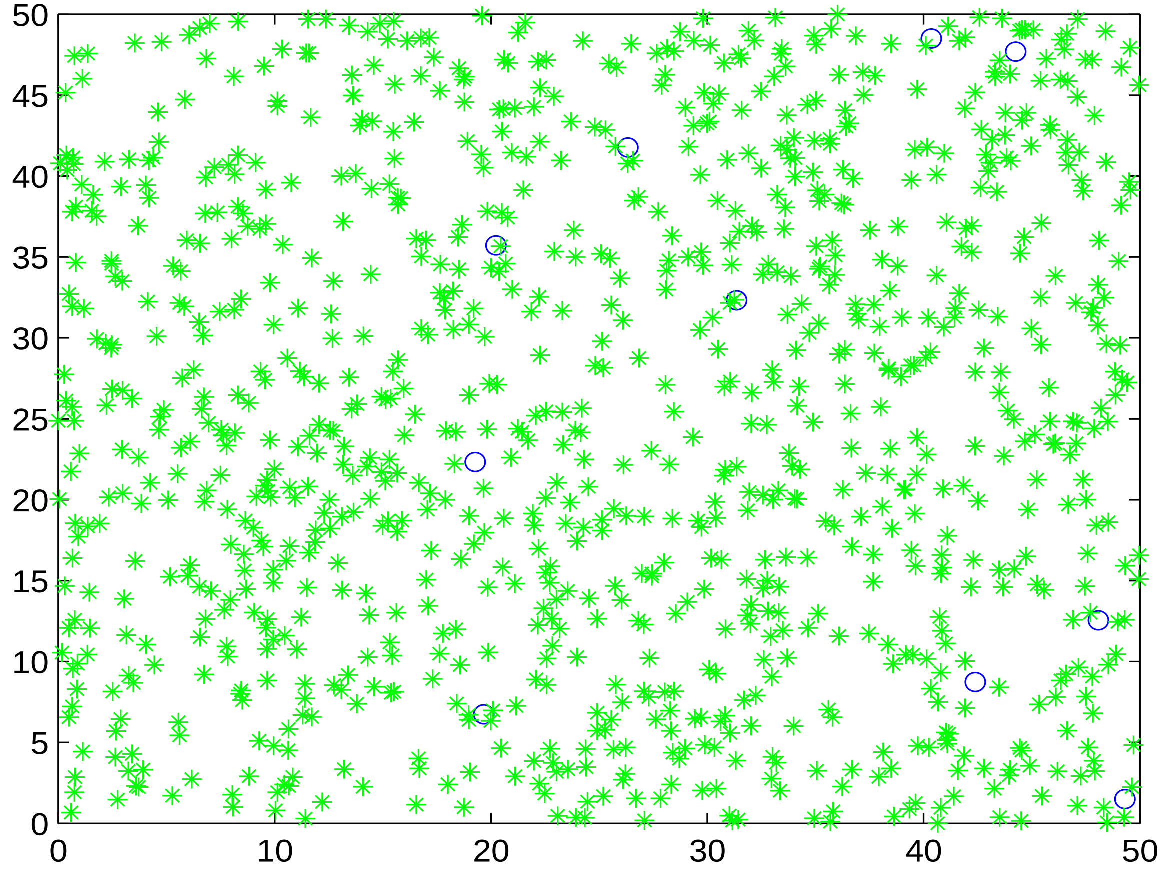

Figure 2: Base-Station Location

Optimization Model

$$\begin{aligned} \textbf{minimize} \quad & \lambda \\ \textbf{subject to} \quad & TSP(S_i) \le \lambda,\ \forall i \\ & \cup S_i = \mathcal{N}, \end{aligned}$$

where $\mathcal{N}$ is set of all customers, $S_i \subset \mathcal{N}$ is the subset of points assigned to vehicle i, and $TSP(S)$ is the minimal TSP tour-length to visit all points in set S by vehicle i.

Data Clustering Model

Let x_{ij} be the binary decision variable to assign sensor point j to vehicle i. Then

$$\begin{array}{rl} \textbf{minimize} & \sum_{i,j} \|\mathbf{a}_j - \mathbf{b}_i\| x_{ij} \\ \textbf{subject to} & \sum_j x_{ij} = \frac{n}{m},\ \forall i, \\ & \sum_i x_{ij} = 1,\ \forall j, \\ & x_{ij} \in \{0, 1\}. \end{array}$$

Data Clustering Model: LP relaxation

Let x_{ij} be the continuous variable to assign sensor point j to vehicle i. Then

$$
\begin{array}{rl}
\textbf{minimize} & \sum_{i,j} \|\mathbf{a}_j - \mathbf{b}_i\| x_{ij} \\
\textbf{subject to} & \sum_j x_{ij} = \frac{n}{m}, \ \forall i, \\
& \sum_i x_{ij} = 1, \ \forall j, \\
& x_{ij} \geq 0.
\end{array}
$$

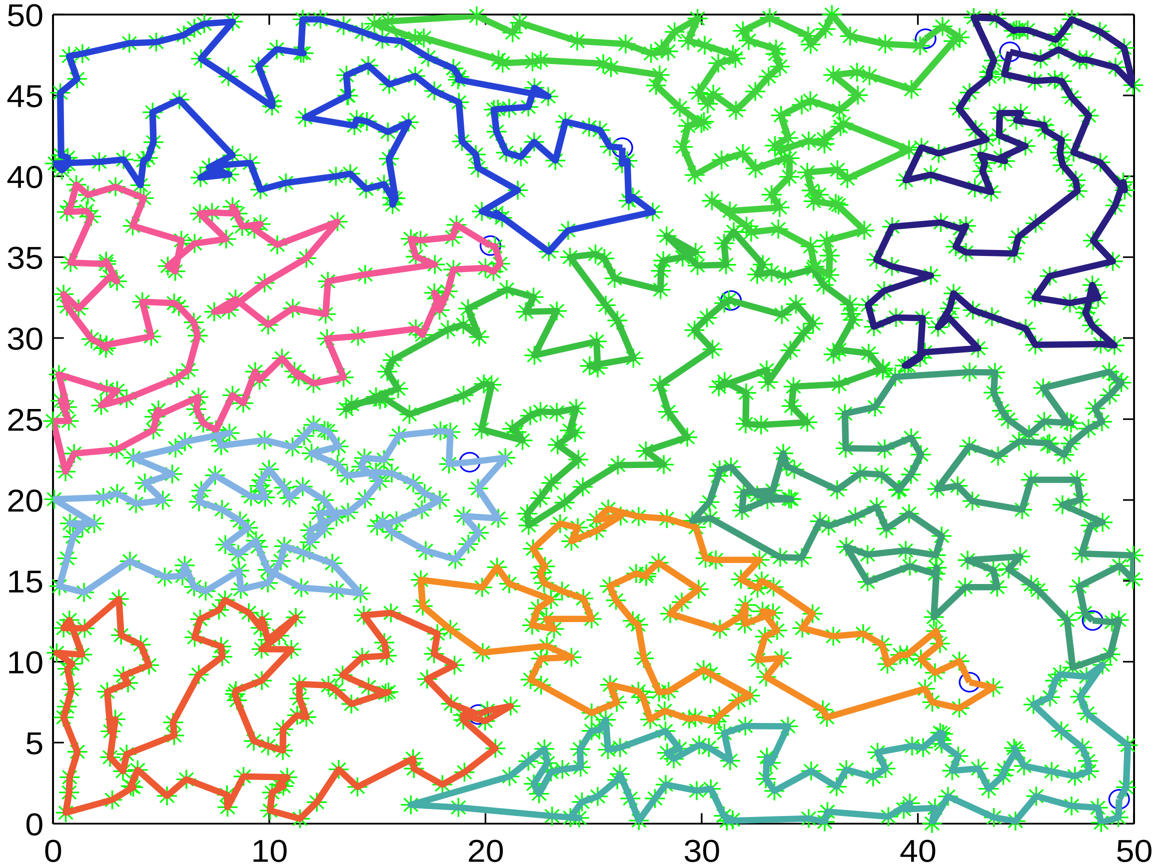

Figure 3: Base-Station Location

Formulation 8: Data Fitting

Given data points $\mathbf{a}_j$, $j = 1, ..., n$, and the observation value c_j at data point $\mathbf{a}_j$, the least squares problem is to find $\mathbf{y}$ such that

$$\sum_{j=1}^{n} (\mathbf{a}_j^T \mathbf{y} - c_j)^2$$

is minimized.

Sometime, it is desired to minimize the sum of p norm, $1 \leq p \leq \infty$

$$\sum_{j=1}^{n} \|\mathbf{a}_j^T \mathbf{y} - c_j\|_p$$

Data Fitting continued

Equivalent smooth formulation:

$$\begin{array}{ll} \textbf{minimize} & \sum_{j=1}^{n} \delta_j \\ \textbf{subject to} & \|\mathbf{a}_j^T \mathbf{y} - c_j\|_p \leq \delta_j, \quad j = 1, 2, ..., n. \end{array}$$

增加变量，转化为线性规划

Data Fitting continued

Constrained data fitting–Fingerprint Matching: c_j is the measured signal strength from base-station j at a location, and $\mathbf{a}_j$ contains base-station j's signal strengths for all known individual locations.

$$\begin{array}{rl} \textbf{minimize} & \sum_{j=1}^{n} |\mathbf{a}_j^T \mathbf{y} - c_j| \\ \textbf{subject to} & e^T \mathbf{y} = 1, \; y_i \in \{0, 1\}. \end{array}$$

LP relaxation:

$$\begin{array}{rl} \textbf{minimize} & \sum_{j=1}^{n} |\mathbf{a}_j^T \mathbf{y} - c_j| \\ \textbf{subject to} & e^T \mathbf{y} = 1, \; \mathbf{y} \geq \mathbf{0}. \end{array}$$

Formulation 9: Fisher's Exchange Market

Buyers have money (w_i) to buy goods and maximize their individual utility functions; Producers sell their goods for money. The equilibrium price is an assignment of prices to goods so as when every buyer buys an maximal bundle of goods then the market clears, meaning that all the money is spent and all goods are sold.

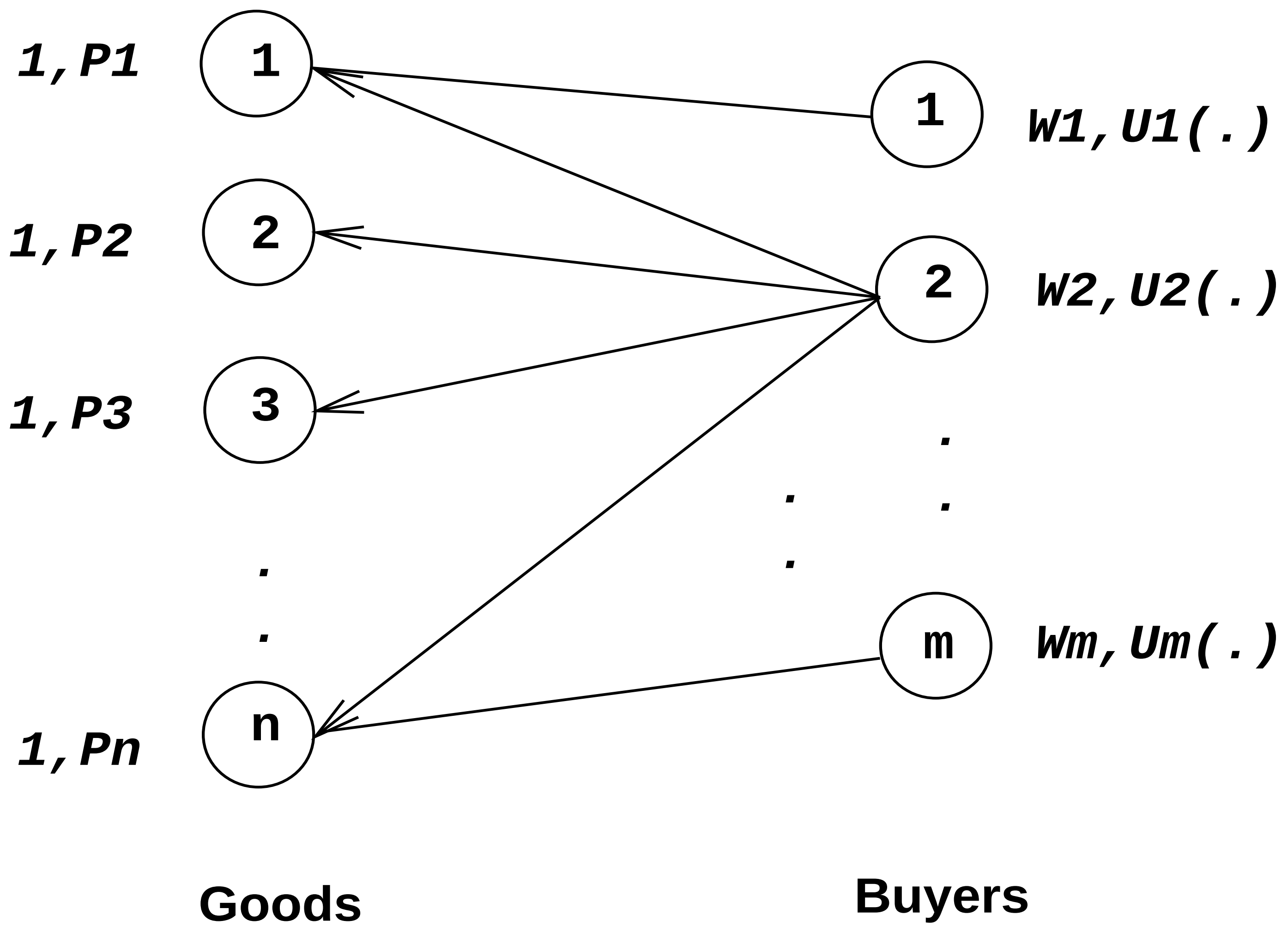

Figure 4: Fisher's Exchange Market Model

Each Buyer's Strategy and the Equilibrium Price

Player $i \in B$'s optimization problem for given prices p_j, $j \in G$.

$$\begin{array}{ll} \textbf{maximize} & \mathbf{u}_i^T \mathbf{x}_i := \sum_{j \in G} u_{ij} x_{ij} \\ \textbf{subject to} & \mathbf{p}^T \mathbf{x}_i := \sum_{j \in G} p_j x_{ij} \leq w_i, \\ & x_{ij} \geq 0, \quad \forall j, \end{array}$$

Without losing generality, assume that the amount of each good is 1. The equilibrium price vector is the one that for all $j \in G$

$$\sum_{i \in B} x(\mathbf{p})_{ij} = 1$$

where $x(\mathbf{p})_i$ is an optimal bundle solution for $i \in B$.

Example of Fisher's Model

Buyer $1, 2$'s optimization problems for given prices p_x, p_y.

$$
\begin{array}{rc}
\textbf{maximize} & 2x_1 + y_1 \\
\textbf{subject to} & p_x \cdot x_1 + p_y \cdot y_1 \le 5, \\
 & x_1, y_1 \ge 0; \\
\textbf{maximize} & 3x_2 + y_2 \\
\textbf{subject to} & p_x \cdot x_2 + p_y \cdot y_2 \le 8, \\
 & x_2, y_2 \ge 0.
\end{array}
$$

$$p_x = 26/3, \;\; p_y = 13/3, \;\; x_1 = 1/13, \;\; y_1 = 1, \;\; x_2 = 12/13, \;\; y_2 = 0.$$

Formulation 10: Portfolio Management

Let $\mathbf{r}$ denote the expected return vector and V denote the co-variance matrix of an investment portfolio, and let $\mathbf{x}$ be the investment proportion vector. Then, one management model is:

$$\begin{array}{rl} \textbf{minimize} & \mathbf{x}^T V \mathbf{x} \\ \textbf{subject to} & \mathbf{r}^T \mathbf{x} \geq \mu, \\ & \mathbf{e}^T \mathbf{x} = 1, \\ & \mathbf{x} \geq \mathbf{0}, \end{array}$$

where $\mathbf{e}$ is the vector of all ones. This is a quadratic program.

In real applications, $\mathbf{r}$ and V may be estimated under various scenarios, say $\mathbf{r}_i$ and V_i for $i = 1, ..., m$.

Robust Portfolio Management

$$
\begin{array}{rl}
\textbf{minimize} & \max_i \mathbf{x}^T V_i \mathbf{x} \\
\textbf{subject to} & \min_i \mathbf{r}_i^T \mathbf{x} \geq \mu, \\
& \mathbf{e}^T \mathbf{x} = 1, \\
& \mathbf{x} \ \geq \ \mathbf{0}.
\end{array}
$$

$$
\begin{array}{rl}
\textbf{minimize} & \alpha \\
\textbf{subject to} & \mathbf{r}_i^T \mathbf{x} \geq \mu, \ \forall i \\
& \mathbf{x}^T V_i \mathbf{x} \leq \alpha, \ \forall i \\
& \mathbf{e}^T \mathbf{x} = 1, \\
& \mathbf{x} \ \geq \ \mathbf{0}.
\end{array}
$$

This is called quadratically constrained optimization.