Math 164: Introduction to Optimization

Instructor: Wotao Yin
Department of Mathematics, UCLA

Winter 2014

online discussions on piazza.com

Cabinet manufacture

Table 1.1. *Cabinet data.*

Cabinet	Wood	Labor	Revenue
Bookshelf	10	2	100
With Doors	12	4	150
With Drawers	25	8	200
Custom	20	12	400

Available resources:

- Wood 5000
- Label 1500

Objective: to maximize the total revenue

$$
\begin{array}{ll}
\text{maximize} & z = 100x_1 + 150x_2 + 200x_3 + 400x_4 \\
\text{subject to} & 10x_1 + 12x_2 + 25x_3 + 20x_4 \leq 5000 \\
& \;\;2x_1 + \;\;4x_2 + \;\;8x_3 + 12x_4 \leq 1500 \\
& x_1, x_2, x_3, x_4 \geq 0.
\end{array}
$$

Resource-constrained revenue optimization

- m resources; resource i has b_i units available
- n products; product j uses a_{ij} units of resource i, generates revenue c_j
- find the maximal revenue subject to available resources

$$
\begin{aligned}
\text{maximize} \quad & z = \sum_{j=1}^{n} c_j x_j \\
\text{subject to} \quad & \sum_{j=1}^{n} a_{ij} \leq b_i, \quad i = 1, \ldots, m \\
& x_j \geq 0, \qquad\quad j = 1, \ldots, n.
\end{aligned}
$$

Compact form

$$\begin{array}{ll} \text{maximize} & z = c^T x \\ \text{subject to} & Ax \leq b \\ & x \geq 0. \end{array}$$

Job assignment

- An insurance office handles three types of work: Information, Policy, Claims.
- There are five workers, and each of them can handle all three types of work
- The office wants to minimize the process time by designing an call assignment strategy.

Table 1.2. *Work times (in minutes).*

Worker	Information	Policy	Claim
1	10	28	31
2	15	22	42
3	13	18	35
4	19	25	29
5	17	23	33

Variables:

- Information calls: fraction p_i assigned to worker i, $\sum_i p_i = 1$
- Policy calls: fraction q_i assigned to worker i, $\sum_i q_i = 1$
- Claim calls: fraction r_i assigned to worker i, $\sum_i r_i = 1$
- Process time t

Objective: to minimize t

Optimization model:

$$
\begin{aligned}
\text{minimize} \quad & z = t \\
\text{subject to} \quad & p_1 + p_2 + p_3 + p_4 + p_5 = 1 \\
& q_1 + q_2 + q_3 + q_4 + q_5 = 1 \\
& r_1 + r_2 + r_3 + r_4 + r_5 = 1 \\
& 10p_1 + 28q_1 + 31r_1 \le t \\
& 15p_2 + 22q_2 + 42r_2 \le t \\
& 13p_3 + 18q_3 + 35r_3 \le t \\
& 19p_4 + 25q_4 + 29r_4 \le t \\
& 17p_5 + 23q_5 + 33r_5 \le t \\
& p_i, q_i, r_i \ge 0, \ i = 1, \ldots, 5.
\end{aligned}
$$

Electrical wire design

- Four building to be connected by electrical wires at one point
- Two circle buildings and two square buildings

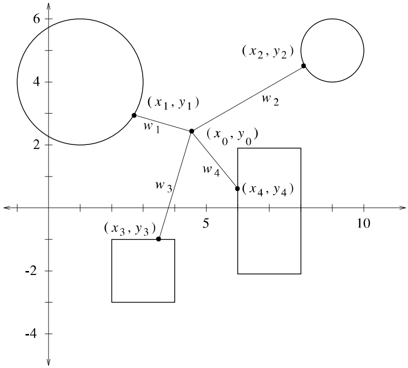


Figure 1.6. *Electrical connections.*

Variables:

- (x_0, y_0): central point
- (x_i, y_i): connection point on building i, $i = 1, 2, 3, 4$

Objective: to minimize the total wire length

Wire lengthen between the central point and building i is

$$w_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}$$

Constraints on (x_i, y_i):

- circle building: $(x_i - x_{\text{center}})^2 + (y_i - y_{\text{center}})^2 \leq \text{radius}^2$
- square building: $x_{\text{left}} \leq x_i \leq x_{\text{right}}$ and $y_{\text{bottom}} \leq y_i \leq y_{\text{top}}$

Model:

$$
\begin{array}{ll}
\text{minimize} & z = w_1 + w_2 + w_3 + w_4 \\
\text{subject to} & w_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2},\ i = 1, 2, 3, 4, \\
& (x_1 - 1)^2 + (y_1 - 4)^2 \le 4 \\
& (x_2 - 9)^2 + (y_2 - 5)^2 \le 1 \\
& 2 \le x_3 \le 4 \\
& -3 \le y_3 \le -1 \\
& 6 \le x_4 \le 8 \\
& -2 \le y_4 \le 2.
\end{array}
$$

Traffic optimization

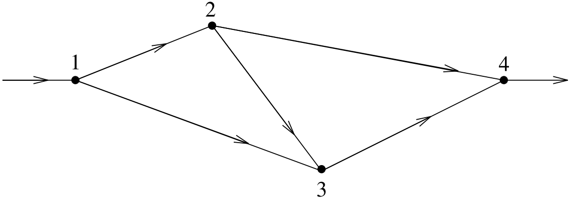


Figure 1.8. *Traffic network.*

Travel time model: between nodes i and j

- if traffic is light, travel time is $t_{i,j}$
- otherwise, define $x_{i,j}$ – $\#$.cars entering per hour, $c_{i,j}$ – maximal $\#$.cars entering per hour, $\alpha_{i,j}$ a slow-down factor

$$T_{i,j}(x_{i,j}) = t_{i,j} + \alpha_{i,j} \frac{x_{i,j}}{1 - x_{i,j}/c_{i,j}}.$$

If there is no traffic, the travel time is $t_{i,j}$. If $x_{i,j} \to c_{i,j}$, then $T_{i,j} \to \infty$.

Objective: to minimize the total travel time through the network from node 1 to node 4 for a volume of X cars per hour.

Optimization model:

$$\text{minimize } f(x) = \sum x_{i,j} T_{i,j}(x_{i,j})$$

subject to the constraints

$$\begin{aligned} x_{1,2} + x_{1,3} &= X \\ x_{2,3} + x_{2,4} - x_{1,2} &= 0 \\ x_{3,4} - x_{1,3} - x_{2,3} &= 0 \\ x_{2,4} + x_{3,4} &= X \\ 0 \le x_{i,j} &\le c_{ij}. \end{aligned}$$

Potential snag: any $x_{i,j} = c_{i,j}$ causing undefined objective ($+\infty$).

Fix: use $0 \le x_{i,j} \le c_{i,j} - \epsilon$ instead.

Projection to hyperplane

Find a minimum distance from a point r to the set $S = \{x : a^T x = b\}$.

- In 2D, S is a line
- In 3D, S is a plane
- in 4 or high dimensions, S is called a hyperplane.

The least distance problem can be formulated as

$$\begin{array}{ll} \text{minimize} & f(x) = \frac{1}{2}(x - r)^T(x - r) \\ \text{subject to} & a^T x = b. \end{array}$$

Unlike most NLP, this problem has a *closed-form* solution, which is unique,

$$x^* = r + \frac{b - a^T r}{a^T a} a.$$

Quadratic programming (QP)

General QP:

$$\begin{array}{ll} \text{minimize} & f(x) = \frac{1}{2} x^T Q x \\ \text{subject to} & Ax \geq b. \end{array}$$

- Easy if Q is positive semi-definite (PSD or $Q \succeq 0$)
- Can be difficult otherwise.

Support vector machine (SVM)

Task: to classify a set of data points into two sets.

Examples:

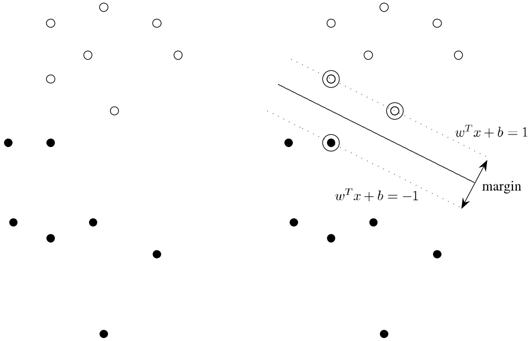
- emails: legitimate or spam
- medical data (age, sex, weight, blood pressure, cholesteral levels, etc.) of a person: high risk or low risk for a heart attack
- handwritten digits: digit 0 or not.

Refined statement: to determine a rule based on the existing data (*training data*) that characterizes the set of points that have the property, distinguishing them from those not having the property.

Training data are *labeled*.

The rule, once determined, can be used to tell whether a new point (*test data*) has the property or not. The rule is called *classifier*.

Simplest form: a set of m training data $x_i \in \mathbb{R}^n$ and labels $y_i = \pm 1$.



Assume it is possible to find a hyperplane $\{x : w^T x + b = 0\}$ that separates the two sets of data points. We would like to find the one with margins. Thus we require (two hyperplanes)

$$\begin{aligned} w^T x_i + b &\geq +1 \quad \text{for } y_i = +1, \\ w^T x_i + b &\leq -1 \quad \text{for } y_i = -1. \end{aligned}$$

w and b can be scaled so that the separation will be ± 1. Signs can be flipped.

There are **infinite many hyperplanes** that separate the two sets of points.

We would like to have the one separate the positive from the negative *as far apart as possible*.

The separation margin is given by $2/\|w\|$**.** To find w and b with maximal margin, it is equivalent to minimize $w^T w$.

$$\begin{array}{ll} \text{minimize} & f(w,b) = \frac{1}{2} w^T w \\ \text{subject to} & y_i(w^T x_i + b) \geq 1, \quad i = 1, \ldots, m. \end{array}$$

Coefficient $1/2$ is included for convenience of analysis and computation.

Among the training points, the few that lie on the boundary of either hyperplane are called *support vectors*. They determine w^* and b^*.

Removing the remaining training points does not change the solution.

Finally, given a new point x, we can use $f(x) = w^T x + b$ (*learning machine*) to tell the label of x, which is the sign of $f(x)$.

So far we have assumed that the training set is perfectly separable.

For the non-separable case, we must allow some points to violate the separating hyperplane. But, we impose a penalty ξ_i for the violation.

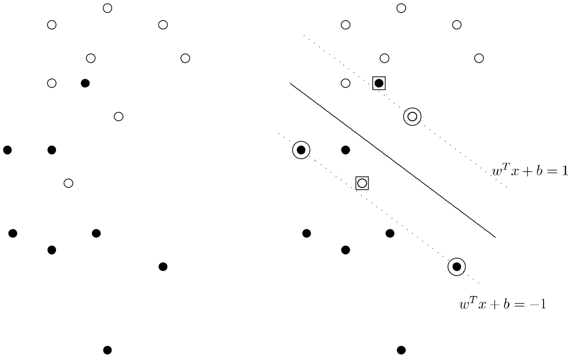
$$\begin{aligned} w^T x_i + b &\geq +1 - \xi_i \quad \text{for } y_i = +1 \\ w^T x_i + b &\leq -1 + \xi_i \quad \text{for } y_i = -1. \end{aligned}$$

We only allow $\xi_i \geq 0$; otherwise, reward will be given to non-violating points.

Our new model is

$$\begin{array}{ll} \text{minimize} & f(w, b, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i \\ \text{subject to} & y_i (w^T x_i + b) \geq 1 - \xi_i, \quad i = 1, \ldots, m, \\ & \xi_i \geq 0. \end{array}$$

where C is a penalty parameter balancing between the margin maximization and violation penalty.



Points inside squares are *misclassified points*. They lie on the incorrect side of the hyperplane $w^T x + b = 0$. But there are only two of them.

Portfolio optimization

Goal: selects a set of assets to achieve a good return with a controlled amount of risk.

The work was pioneered by Nobel Prize laureate Harry Markowitz. He demonstrated how to reduce the risk of investment by selecting a portfolio of stocks rather than picking individual attractive stocks.

Definitions:

- Investment proportion: $x = [x_1, \ldots, x_n]^T$, x_j is the proportion invested in asset j.
- Expected return: $\mu = [\mu_1, \ldots, \mu_n]^T$, μ_j for asset j. Portfolio expected rate of return $= \mu^T x$.
- Risk: Σ is the matrix of variances and covariances of the assets' returns.

Understanding risks:

- A high variance Σ_{jj} indicates high volatility or risk of asset j alone.
- A positive covariance Σ_{ij} indicates that the values of assets i and j tend to move in the same direction. For example, stocks in the same industry.
- A negative covariance Σ_{ij} indicates that the values of assets i and j tend to move in the opposite direction, a desirable feature in a diversified portfolio.
- The portfolio risk is defined as the expected variance $x^T \Sigma x$.

μ and Σ can be computed from the assets' history.

We have two conflicting objectives:

- to maximize the expected return $\mu^T x$
- to minimize the risk $x^T \Sigma x$.

Therefore, we introduce a parameter $\alpha > 0$ to reflect the trade-off between them.

$$\text{maximize}\, \mu^T x - \alpha x^T \Sigma x \quad \textsf{subject to} \ \sum_i x_i = 1,\ x \geq 0.$$

Parameter α:

- larger $\Rightarrow$ more risk aversion
- smaller $\Rightarrow$ higher tolerance for risk, more emphasis on the expected return
- it can be difficult to choose a sensible value, so we commonly solve this model for a range of α and obtain a solution path.

Example

Table 1.3. *Past rates of return of stocks.*

Period	Stock 1	Stock 2	Stock 3	Stock 4
1	0.08	0.05	0.01	0.08
2	0.06	0.17	0.09	0.12
3	0.07	0.05	0.10	0.07
4	0.04	−0.07	0.04	−0.01
5	0.08	0.12	0.08	0.09
6	0.07	0.22	0.11	0.09

Sample mean and variance:

$$r = \begin{pmatrix} 0.0667 & 0.0900 & 0.0717 & 0.0733 \end{pmatrix},$$

$$V = \begin{pmatrix} 0.00019 & 0.00065 & 0.00004 & 0.00038 \\ 0.00065 & 0.00883 & 0.00218 & 0.00327 \\ 0.00004 & 0.00218 & 0.00125 & 0.00063 \\ 0.00308 & 0.00327 & 0.00063 & 0.00162 \end{pmatrix}.$$

Table 1.4. *Optimal portfolio for selected values of* α.

α	Stock 1	Stock 2	Stock 3	Stock 4	Mean	Variance
1	0	1	0	0	0.090	8.8×10^{-3}
2	0.12	0.65	0.23	0	0.083	4.5×10^{-3}
5	0.57	0.19	0.24	0	0.072	8.0×10^{-4}
10	0.71	0.04	0.25	0	0.069	2.6×10^{-4}
100	0.87	0	0.13	0	0.067	1.7×10^{-4}

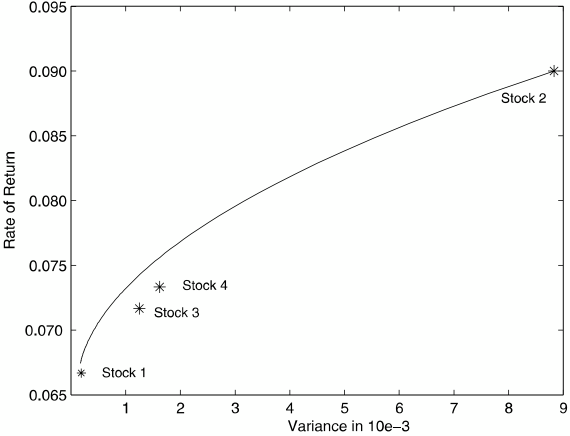


Figure 1.12. *Efficient frontier.*