Chap 8. Optimization

8.1. Linear Programming

$$\begin{cases} \text{Minimize } \vec{c}^T\vec{x} & \text{(cost function or objective function)} \\ \text{Subject to } A\vec{x}=\vec{b} \text{ and } \vec{x}\geq 0 & (\vec{x}: \text{feasible}) \end{cases}$$

where A: $m\times n$ matrix ($m<n$)

If $\vec{x}$ is the vector that gives the minimum,
$n-m$ of its components are zero.
$\Rightarrow$ We need to discover where they belong.

Simplex Method

Stage 1: Find a feasible vector $\vec{x}$ with $n-m$ zero components ($A\vec{x}=\vec{b}$ and $\vec{x}\geq\vec{0}$)

Stage 2: Allow one of the zero components to become positive and force one of the positive components to become zero, while reducing the cost $\vec{c}^T\vec{x}$ and keeping $A\vec{x}=\vec{b}$ and $\vec{x}\geq\vec{0}$.

There are $\binom{n}{n-m}$ possible positions for the $(n-m)$ zeros.
Thus the algorithm stops at a finite number of steps.

Three Approaches

(1) Geometric: the feasible set (Section 8.1)
(2) Computational: the simplex method (Section 8.2)
(3) Algebraic: the theory of duality (Section 8.3)

Extreme Points of Feasible Sets

Example:

Minimize $5x_1 + 4x_2 + 8x_3$

Subject to $x_1 + x_2 + x_3 = 1$, $x_1, x_2, x_3 \geq 0$

$$\left(\begin{array}{l} \vec{c}^T = [5\ 4\ 8] \\ m=1,\ n=3,\ A=[1\ 1\ 1],\ \vec{b}=[1] \end{array} \right)$$

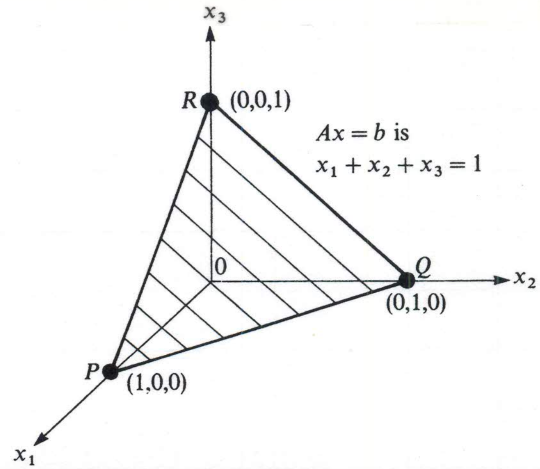

Fig. 8.1. A feasible set: the triangle PQR.

- The minimum occurs at one of the corners P, Q, R, and not in the interior.
  - $\Rightarrow$ minimum at $Q = (0,1,0)$ with $n-m=2$ zero components
- Geometrically, all vectors $\vec{x}$ with zero cost would lie on the plane $5x_1 + 4x_2 + 8x_3 = 0$ passing through the origin. None of these vectors are feasible.
- As the cost increases above zero, the plane moves toward the triangle. When the cost reaches 4, the plane $5x_1 + 4x_2 + 8x_3 = 4$ first makes contact with the feasible set at the point Q.

2D Example

Minimize $x_1 + x_2$

Subject to $x_1 + 2x_2 \geq 4$, $x_1, x_2 \geq 0$

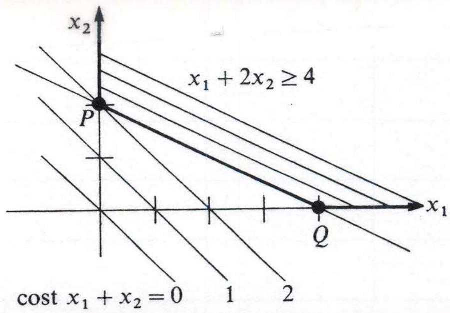

Fig. 8.2. The feasible set touched by $cx = x_1 + x_2 = 2$.

The line $x_1 + x_2 = 0$ of zero cost misses the feasible set. So does $x_1 + x_2 = 1$. The line $x_1 + x_2 = 2$ touches the set at $P = (0, 2)$. The optimal solution is $P = (0, 2)$ or

$x_1 = 0, x_2 = 2$

Other Possibilities in Linear Programming

(1) Maximize $\vec{c}^T\vec{x}$ instead of minimizing it.

(2) The cost $\vec{c}^T\vec{x}$ could be negative

(E.g., $\vec{c}^T = (5, -3, 8)$, $\min \vec{c}^T\vec{x} = -3$ at $Q = (0, 1, 0)$)

(3) The feasible set could be unbounded even with equality constraints

[E.g., If the constraint $A\vec{x} = \vec{b}$ is $x_1 + x_2 - x_3 = 1$, the component x_3 can be arbitrarily large, say $x_3 = 999$, $x_1 = x_2 = 500$. Therefore, $\vec{c}^T\vec{x}$ can be very large]

(4) The feasible set can be empty.

[E.g., If $x_1 + x_2 + x_3 = -1$, there is no solution that meets $x_1, x_2, x_3 \geq 0$.]

Inequality Constraints

Slack variables turn inequalities to equalities

$x_1+x_2+x_3 \le 1 \Rightarrow x_1+x_2+x_3+\underset{\text{slack variable}}{x_4}=1.$

$$\begin{cases} \text{Minimize } 5x_1+4x_2+8x_3 \\ \text{Subject to } x_1+x_2+x_3 \le 1, \quad x_1, x_2, x_3 \ge 0 \end{cases}$$

$\Rightarrow$

$$\begin{cases} \text{Minimize } 5x_1+4x_2+8x_3+0\cdot x_4 \\ \text{Subject to } x_1+x_2+x_3+x_4=1, \quad x_1, x_2, x_3, x_4 \ge 0 \end{cases}$$

- The feasible set is changed, from the tetrahedron PQRO in 3-dim space to some set in 4-dim space with corners $P=(1,0,0,0)$, $Q=(0,1,0,0)$, $R=(0,0,1,0)$ and $S=(0,0,0,1)$

- Maximize $\vec{c}^T\vec{x}$ subject to $A\vec{x} \le \vec{b}$, $\vec{x} \ge 0$
$\Rightarrow$ Maximize $\vec{c}^T\vec{x}$ subject to $[A \ I]\begin{bmatrix}\vec{x} \\ \vec{w}\end{bmatrix} = A\vec{x}+\vec{w}=\vec{b}$
($\vec{w}$: a vector of m slack variables) $\vec{x} \ge \vec{0}$, $\vec{w} \ge \vec{0}$.
$\Rightarrow$ Maximize $\vec{c}^T\vec{x}$ subject to $A\vec{x}=\vec{b}$, $\vec{x} \ge \vec{0}$.
($\vec{c}^T = [\vec{c}^T \ 0]$, $\vec{x}^T = [\vec{x}^T \ \vec{w}^T]$)

- Minimize $\vec{c}^T\vec{x}$ Subject to $A\vec{x} \ge \vec{b}$, $\vec{x} \ge \vec{0}$
$\Rightarrow$ Minimize $\vec{c}^T\vec{x}$ subject to $[A \ -I]\begin{bmatrix}\vec{x} \\ \vec{w}\end{bmatrix} = A\vec{x}-\vec{w}=\vec{b}$,
$\vec{x} \ge \vec{0}$, $\vec{w} \ge \vec{0}$.
$\Rightarrow$ Minimize $\vec{c}^T\vec{x}$ subject to $A\vec{x}=\vec{b}$, $\vec{x} \ge \vec{0}$
($\vec{c}^T = [\vec{c}^T \ 0]$, $A = [A \ -I]$, $\vec{x}^T = [\vec{x}^T \ \vec{w}^T]$.)

8.2 The Simplex Method and Karmarkar's Method

The Simplex Method

Example:

Minimize $\vec{c}^T\vec{x} = 3x_1 + x_2 + 9x_3 + x_4$

Subject to
$$\begin{cases} x_1 \quad\; + 2x_3 + x_4 = 4 \\ \quad x_2 + x_3 - x_4 = 2 \end{cases} \Big) \; A\vec{x} = \vec{b}$$
$$x_1, x_2, x_3, x_4 \geq 0 \quad \big) \; \vec{x} \geq \vec{0}$$

Starting from a corner $\vec{x} = (4, 2, 0, 0)$
with $m=2$ positive entries and $n-m=2$ zeros.

① If x_3 is increased (keeping $x_4=0$),

$x_1 = 4 - 2x_3$

$x_2 = 2 - x_3$

the reduced cost

$\vec{c}^T\vec{x} = 3(4-2x_3) + (2-x_3) + 9x_3 = 14 + 2x_3$

The cost increases with x_3.

② On the other hand, if x_4 is increased (keeping $x_3=0$),

$x_1 = 4 - x_4$

$x_2 = 2 + x_4$

$\vec{c}^T\vec{x} = 3(4-x_4) + (2+x_4) + x_4 = 14 - x_4$

This edge decreases the total cost.

③ When x_4 enters, either x_1 or x_2 must leave.

From $\begin{cases} x_1 = 4 - x_4 \\ x_2 = 2 + x_4 \end{cases}$

as x_4 increases, it is only x_1 that moves toward zero.

It reaches $x_1 = 0$ when $x_4 = 4$.

The new vertex: $\vec{x} = (0, 6, 0, 4)$

Its cost: $\vec{c}^T\vec{x} = 10$

④ The fourth and second columns are now basic.

$B = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \Rightarrow$ It can be made into an identity by adding the first row of A to the second:

$$\begin{cases} x_1 \quad + 2x_3 + x_4 = 4 \\ x_1 + x_2 + 3x_3 \quad = 6 \end{cases} \quad \cdots\cdots (*)$$

To verify the current corner $\vec{x} = (0,6,0,4)$ is optimal, the simplex method computes the reduced costs.

(4-1) If x_1 were introduced (keeping $x_3 = 0$),

$$x_2 = 6 - x_1$$

$$x_4 = 4 - x_1$$

positive reduced cost

$\uparrow (r_1 = 1)$

$$\vec{c}^T\vec{x} = 3x_1 + (6 - x_1) + (4 - x_1) = 10 + x_1$$

(4-2) If x_3 were introduced (keeping $x_1 = 0$)

$$x_2 = 6 - 3x_3$$

$$x_4 = 4 - 2x_3$$

positive reduced cost.

$\uparrow (r_3 = 4)$

$$\vec{c}^T\vec{x} = (6 - 3x_3) + 9x_3 + (4 - 2x_3) = 10 + 4x_3$$

$\therefore$ The current corner $\vec{x} = (0,6,0,4)$ is optimal

Remark:

Matrix algebra gives an explicit formula for $r_1 = 1$, $r_3 = 4$

$$[r_1 \; r_3] = [c_1 \; c_3] - [c_4 \; c_2] \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$$

$$= [3 \; 9] - [1 \; 1] \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = [1 \; 4]$$

The matrix came from columns 1 and 3 of the constraints in equation (*).

Tableau Method

$$T_0 = \left[\begin{array}{c|c} A & \vec{b} \\ \hline \vec{c}^T & 0 \end{array}\right] = \left[\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 4 \\ 0 & 1 & 1 & -1 & 2 \\ \hline 3 & 1 & 9 & 1 & 0 \end{array}\right]$$

① Subtracting row 2 and also 3 times row 1 from the last row completes the elimination:

$$T_1 = \left[\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 4 \\ 0 & 1 & 1 & -1 & 2 \\ \hline 0 & 0 & 2 & -1 & -14 \end{array}\right]$$

The last row contains the reduced costs $r_3 = 2$ and $r_4 = -1$.
It also shows the cost $\vec{c}^T\vec{x} = 14$.
We bring in x_4 because r_4 is negative.
We can also decide whether x_1 or x_2 should leave.

$$\frac{4}{1} = 4, \quad \frac{2}{-1} = -2$$

Those ratios give the limit on x_4, when the end of the edge is reached. We can go as far as $x_4 = 4$, when x_1 touches zero and must leave.
The negative ratio can be ignored.

② A new elimination prepares for the next step.
Adding the first row of T_1 to the second and third gives

$$T_2 = \left[\begin{array}{cccc|c} 1 & 0 & 2 & 1 & 4 \\ 1 & 1 & 3 & 0 & 6 \\ \hline 1 & 0 & 4 & 0 & -10 \end{array}\right]$$

The last column shows $x_4 = 4$ and $x_2 = 6$, and $\vec{c}^T\vec{x} = 10$.
The reduced costs are $r_1 = 1$, $r_3 = 4$ in the last row.
Neither one is negative. The simplex method stops.

The Algebra of a Simplex Step

$$A\vec{x} = [B\ N]\begin{bmatrix}\vec{x}_B\\ \vec{0}\end{bmatrix} = \vec{b} \quad \text{or} \quad \vec{x}_B = B^{-1}\vec{b}$$

$$\vec{c}^T\vec{x} = [\vec{c}_B\ \vec{c}_N]\begin{bmatrix}\vec{x}_B\\ \vec{0}\end{bmatrix} = \vec{c}_B \cdot B^{-1}\vec{b}$$

Where to go next?

$$[I\ B^{-1}N]\begin{bmatrix}\vec{x}_B\\ \vec{0}\end{bmatrix} = B^{-1}\vec{b}$$

When the zero components of $\vec{x}$ increase to $\vec{x}_N$, the nonzero components $\vec{x}_B$ must drop by $B^{-1}N\vec{x}_N$.

$$\Rightarrow \quad \vec{c}^T\vec{x} = \vec{c}_B^{\,T}(\vec{x}_B - B^{-1}N\vec{x}_N) + \vec{c}_N^{\,T}\vec{x}_N$$
$$= (\vec{c}_N^{\,T} - \vec{c}_B^{\,T}B^{-1}N)\vec{x}_N + \vec{c}_B^{\,T}\vec{x}_B$$

Let $\vec{r}^{\,T} = \vec{c}_N^{\,T} - \vec{c}_B^{\,T}B^{-1}N$: the vector of reduced costs

① If $\vec{r} \geq \vec{0}$, the current corner is optimal.

② Otherwise, choose the most negative component i of $\vec{r}$.
It allows that component x_i to increase from zero.
Which component x_j should leave?
It will be the first to reach zero as x_i increases.
Let $\vec{v}$ be the column of $B^{-1}N$ that multiplies x_i. Then,

$$\vec{x}_B + x_i\vec{v} = B^{-1}\vec{b}$$

The kth component of $\vec{x}_B$ will drop to zero when

$$x_i = \frac{k\text{th component of } B^{-1}\vec{b}}{k\text{th component of } \vec{v}}$$

The smallest of these ratios determines how large x_i can be.
If the jth ratio is the smallest, the leaving variable is x_j.
At the new corner, x_i has become positive and x_j has become zero.

Karmarkar's Method

Karmarkar proposes to move within the feasible set, subject to the constraints $A\vec{x}=\vec{b}$ and $\vec{x}\geq\vec{0}$ toward the minimum. At the same time, he has introduced a change of variables that transforms the problem. In its simplest form, it is a rescaling problem.

A simplified first step of Karmarkar's method

(1) Choose the direction $\Delta\vec{x}$ that makes $\cos\theta$ $\left(\vec{c}^T\vec{x}=|\vec{c}||\vec{x}|\cdot\cos\theta\right)$ as large as possible.

It is the projection of $-\vec{c}$ onto the nullspace of A.

(2) Stop close to, but a little before, the boundary before which a constraint is violated.

(A typical constant is $\alpha=0.98$)

Since $A\vec{x}^0=\vec{b}$ and $A\vec{x}^1=\vec{b}$, $A\Delta\vec{x}=\vec{0}$.

Each step must lie in the nullspace of A, which is parallel to the feasible set. Projecting $-\vec{c}$ onto the nullspace gives the one of steepest descent.

The formula for projection onto the nullspace of A:

$$P=I-A^T(AA^T)^{-1}A$$

Pf) ① If $A\vec{z}=\vec{0}$, then $P\vec{z}=\vec{z}$

$\therefore$ a vector in the nullspace is not moved.

② P takes every vector into the null space since $AP=A-(AA^T)(AA^T)^{-1}A=0$ and $AP\vec{c}=\vec{0}$,

$\therefore P\vec{c}$ is the projection of $\vec{c}$ onto the nullspace. ⌟

- In computations, $(AA^T)^{+}$ is never formed.
  Instead we compute $(AA^T)\vec{y} = A\vec{c}$ and $P\vec{c} = \vec{c} - A^T\vec{y}$.
- The move $\Delta\vec{x}$ is a multiple of $-P\vec{c}$.
  The cost is reduced because $\vec{c}^T\Delta\vec{x}$ is the same multiple of $-\vec{c}^TP\vec{c}$, which cannot be positive.
  ($-P$ is negative semidefinite)

Example

Minimize $\vec{c}^T\vec{x} = 5x_1 + 4x_2 + 8x_3$

Subject to $x_1 + x_2 + x_3 = 1$

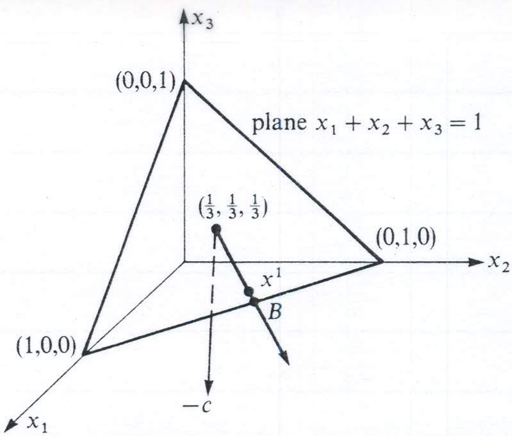

Fig. 8.4. A step of the rescaling algorithm.

The optimal corner is $Q = (0,1,0)$

The initial guess: $\vec{x}^0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ at the center of the feasible set.

The first step projects $\vec{c}$ onto the plane of triangle.

$A = [1 \ 1 \ 1]$

$AA^T\vec{y} = A\vec{c}$ becomes $3y = 17$

$$P\vec{c} = \vec{c} - A^T\vec{y} = \begin{bmatrix} 5 \\ 4 \\ 8 \end{bmatrix} - \frac{17}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -2 \\ -5 \\ 7 \end{bmatrix}$$

Moving from $\vec{x}^0$ in the opposite direction $-P\vec{c}$ gives

$$\vec{x}^0 - sP\vec{c} = \frac{1}{3}\begin{bmatrix} 1+2s \\ 1+5s \\ 1-7s \end{bmatrix}$$

a boundary point.

The largest possible step is $s=\frac{1}{7}$ (that reaches $(\frac{3}{7},\frac{4}{7},0)$)

For safety, the step length s is reduced by $\alpha=0.98$:

$$s = 0.98\left(\frac{1}{7}\right) = 0.14 \text{ and } \vec{x}^1 = \vec{x}^0 - 0.14P\vec{c} = \frac{1}{3}\begin{bmatrix} 1.28 \\ 1.70 \\ 0.02 \end{bmatrix}$$

For the second step, Karmarkar suggests to transform $\vec{x}^1$ back to the central position $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ by rescaling

$$D^{-1}\vec{x}^1 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \text{ where } D = \begin{bmatrix} 1.28 & & \\ & 1.70 & \\ & & 0.02 \end{bmatrix}$$

The rescaling from $\vec{x}$ to $\vec{X} = D^{-1}\vec{x}$ has three effects:

① $A\vec{x}=\vec{b}$ (or $ADD^{-1}\vec{x}=\vec{b}$) becomes $AD\vec{X}=\vec{b}$

② $\vec{c}^T\vec{x}$ (or $\vec{c}^TDD^{-1}\vec{x}$) becomes $(D^T\vec{c})^T\vec{X}$.

③ $\vec{x}\geq\vec{0}$ becomes $\vec{X}\geq\vec{0}$.

$$A = [1 \ 1 \ 1], \quad \vec{c}^T = (5, 4, 8)$$

$$\Rightarrow AD = [1.28, 1.70, 0.02] \text{ and } D\vec{c} = (6.4, 8.0, 0.16)^T$$

$$\Delta\vec{X} = -s\left[I - D^TA^T(ADD^TA^T)^{-1}AD\right]D^T\vec{c} \qquad (15)$$

Rescaling Algorithm

(1) Construct a diagonal matrix D from the components of x^k, so that $D^{-1}x^k = (1,1,\ldots,1) = e$
(2) Apply the projection in equation (15)
(3) Determine s so that $e+\Delta X$ has a zero component
(4) Multiply that correction ΔX by a factor α $(0.9 \leq \alpha < 1)$
(5) The new vector is $x^{k+1} = D(e+\alpha\Delta X) = x^k + D\alpha\Delta X$.

o. It is a weighted projection with $W = D^T$ and $C = W^T W = DD^T$.

The weighted normal equations are solved for $\vec{y}$:

$$(ADD^T A^T)\vec{y} = ADD^T \vec{c} \qquad (16)$$

This is the heart of the problem — it is weighted least squares.

o. The projection formula is familar.

At each step, we find the vector $\vec{z}$ that is closest to $\vec{c}$ but still in the nullspace of A. It is a weighted least squares, with the constraint $A\vec{z} = \vec{0}$.

Equilibrium equations:

$$\begin{bmatrix} (DD^T)^{-1} & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \vec{z} \\ \vec{y} \end{bmatrix} = \begin{bmatrix} \vec{c} \\ \vec{0} \end{bmatrix}$$

o. Starting values:

By adding an extra column $\vec{b} - A\vec{e}$ to the matrix A, where $\vec{e} = (1, \cdots, 1)^T$, the vector $\vec{e}$ is made feasible. The larger matrix is renamed A, so that $A\vec{e} = \vec{b}$ and the algorithm starts with $\vec{x}^0 = \vec{e}$.

When the new variable is given a high cost, it is driven to zero and the solution of the enlarged problem is also optimal for the original problem. Phases I & II go simultaneously. Only one artificial variable is used whereas the simplex method uses m artificial variables.

o. Karmarkar replaces the rescaling by

$$\hat{\vec{x}} = \frac{n D^{-1}\vec{x}}{\vec{e}^T D^{-1}\vec{x}} \quad \text{with } \vec{e}^T = [1\ 1 \cdots 1]$$

This is a projective transformation. $\vec{e}^T\hat{\vec{x}} = n$ or $\hat{x}_1 + \hat{x}_2 + \cdots + \hat{x}_n = n$.

We skip the details of further transformations.