最优化方法 (I)

张宏鑫，华炜

2010-06-10

浙江大学计算机学院

内容

- 线性规划
- 非线性优化
- 主要参考书：
  - 线性规划，张建中，许绍吉，科学出版社
  - 最优化理论与方法，袁亚湘，孙文瑜，科学出版社

一、线性规划

- 在人们的生产实践中，经常会遇到如何利用现有资源来安排生产，以取得最大经济效益的问题。

$利益驱动$

运筹学→

数学规划 →

线性规划(Linear Programming，LP)

- 1947年G. B. Dantzig 求解线性规划的单纯形方法
- 现有的计算机能处理成千上万个约束条件和决策变量的线性规划问题之后，线性规划的适用领域更为广泛了。

1. 基本理论

1.1 问题定义 (linear programming, LP)

- 某计算机公司提供大型计算机系统，有A, B, C三个技术团队
  - 高端系统
    - 每套售后的利润400万元
    - 需用A、B两个技术团队协作实施
    - A团队2个月，B团队1个月；
  - 中端系统
    - 每套售后的利润300万元
    - 需用A, B, C三个技术团队协作实施
    - 每个团队各需1个月
  - 2009年度，A队可工作10个月，B队8个月、C队7个月
- 问：怎样的销售策略能使总利润最大？

1.1 问题定义 (linear programming, LP)

- 设可销售x_1个中端系统， x_2个高端系统
- 目标函数

$$\max \quad z = 4x_1 + 3x_2$$

- 约束条件 $$\begin{cases} 2x_1 + x_2 \le 10 \\ x_1 + x_2 \le 8 \\ x_2 \le 7 \\ x_1, x_2 \ge 0 \end{cases}$$

附加条件：x_1和x_2是整数

讨论？！？

1.1 问题定义
(linear programming, LP)

- 在一组线性的等式或不等式约束下，求一个线性函数的最小值或最大值。
- 形式化的定义：

$$
\begin{aligned}
&\min\ c_1x_1 + \ldots + c_nx_n \\
&\text{s. t.} \\
&\qquad a_{11}x_1 + \ldots + a_{1n}x_n \le b_1 \\
&\qquad \ldots \\
&\qquad a_{m1}x_1 + \ldots + a_{mn}x_n \le b_m \\
&\qquad x_1, \ldots, x_n \ge 0
\end{aligned}
$$

1.1 问题定义 (linear programming, LP)

- 在一组线性的等式或不等式约束下，求一个线性函数的最小值或最大值。
- 形式化的定义：

目标函数
objective function

$$\begin{aligned} &\min \mathbf{c}^T\mathbf{x} \\ &\text{s.t. } A\mathbf{x} \leq \mathbf{b} \\ &\qquad \mathbf{x} \geq \mathbf{0} \end{aligned}$$

c: 价值向量，cost vector
A: 约束矩阵，constraint matrix
b: 右端向量，right-hand-side vector

1.1 问题定义
(linear programming, LP)

- **X**：满足约束条件，称为可行解或可行点
  - (feasible point)

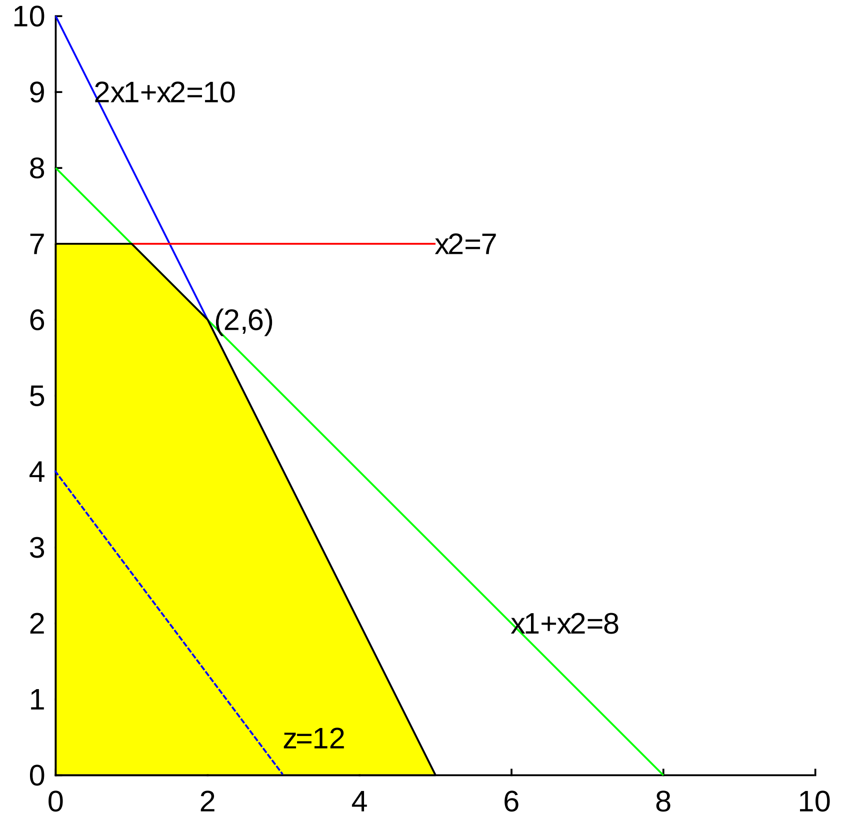
- D：所有可行点的集合,称为可行区域
  - (feasible region)

- LP问题：
  - D = ∅，无解或不可行
  - D ≠ ∅，但目标函数在D上无上界：无界
  - D ≠ ∅，且目标函数有限的最优解：有最优解

线性规划的图解法

$$\max \quad z = 4x_1 + 3x_2$$

$$\begin{cases} 2x_1 + x_2 \le 10 \\ x_1 + x_2 \le 8 \\ x_2 \le 7 \\ x_1, x_2 \ge 0 \end{cases}$$



线性规划在Matlab中的实现方法

- Matlab中规定线性规划的标准形式为:

$$\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x} \quad \text{such that} \quad \mathbf{Ax} \le \mathbf{b}$$

- 其中:
  - $\mathbf{c}$和$\mathbf{x}$为n维列向量,
  - $\mathbf{b}$为m维列向量,
  - $\mathbf{A}$为$m \times n$矩阵

线性规划在Matlab中的实现方法

- 基本函数形式为

  linprog(c, A, b)

  - 返回值是向量**x**的值
- 调用举例：

  [x,fval]=linprog(c, A, b, Aeq, beq, LB, UB, X0, OPTIONS)

  fval返回目标函数的值，

  Aeq、beq对应等式约束，Aeq X = beq

  LB、UB分别是变量X的下界和上界， X0是X的初始值

  OPTIONS是控制参数。

线性规划在Matlab中的实现方法

- 例：

$$\max z = 2x_1 + 3x_2 - 5x_3$$

$$\begin{cases} x_1 + x_2 + x_3 = 7 \\ 2x_1 - 5x_2 + x_3 \ge 10 \\ x_1, x_2, x_3 \ge 0 \end{cases}$$

- Matlab代码

```
c=[2,3,-5];
a=[-2,5,-1];
b=-10;
aeq=[1,1,1];
beq=7;
x=linprog(
  -c,a,b,aeq,beq,
  zeros(3,1))
value=c'*x
```

1.2 标准LP问题

- 从实际中总结出来的LP形式不完全一样：
  - 目标函数是最大值或最小值，
  - 约束条件是等式约束或不等式约束，
  - 变量有上界或下界或无界

$$
\boxed{\begin{aligned} \min\ z &= \mathbf{cx} \\ \text{s.t. } A\mathbf{x} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned}}
$$

标准形式的LP问题

$$\min z = \mathbf{cx}$$
$$\text{s.t. } A\mathbf{x} = \mathbf{b}$$
$$\mathbf{x} \geq \mathbf{0}$$

- 标准化
  - 目标函数的转换　max z → min (-z)
  - 约束条件的转换（引入松弛变量）

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i \Leftrightarrow \begin{cases} \sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i \\ x_{n+i} \geq 0 \end{cases}$$

  - 变量的非负约束

$$x_j \geq l_j \Leftrightarrow y_j \geq 0, y_j = x_j - l_j$$

$$x_j\text{自由变量} \Leftrightarrow u_j \geq 0, v_j \geq 0, x_j = u_j - v_j$$

1.3 可行区域D

- 下面讨论可行区域 $D=\{\mathbf{x} \mid A\mathbf{x}=\mathbf{b},\ \mathbf{x}\geq\mathbf{0}\}$ 的结构
  - 首先讨论集合 $K=\{\ \mathbf{x} \mid A\mathbf{x} = \mathbf{b}\ \}$

1.3.1 仿射集

- 仿射集：对于集合$S\subseteq E^n$，如果对任意$\mathbf{x}, \mathbf{y}\in S$, $\lambda\in E^1$, 都有$\lambda\mathbf{x}+(1-\lambda)\mathbf{y}\in S$,则$S$为仿射集

- 例：
  - E^3中的直线和平面都是仿射集，空集Ø也是仿射集
  - S是E^n中的子空间的充要条件是S为包含原点的仿射集
  - 每一非空的仿射集S都平行于唯一的子空间L
    - 集合的平移：$S+\mathbf{p}=\{\mathbf{x}+\mathbf{p} \mid \mathbf{x}\in S\}$

  - 非空仿射集S的维数定义为与它平行的子空间的维数，$\dim(S)$

1.3.1 仿射集

- 维数为0，1和2的仿射集也称为点、直线和平面
- E^n中的n-1维仿射集称为超平面（hyperplane）

- 给定$b \in E^1$及非零向量 $\mathbf{a} \in E^n$,则集合

$$H=\{\ \mathbf{x} \mid \mathbf{a}^T\mathbf{x}=b\ \}$$

是E^n中的一个超平面，并且E^n中的任一超平面都能表示成这种形式。

  - $\mathbf{a}$称为H的法向量
  - 当$\mathbf{a}$固定，b取不同值时，$\mathbf{a}^T\mathbf{x}=b$表示E^n中的一族平行的超平面

1.3.1 仿射集

- 一般仿射集的特征：
  - 给定 mxn实矩阵A和$\mathbf{b}\in E^m$，则$K=\{\mathbf{x}\in E^n|A\mathbf{x}=\mathbf{b}\}$为$E^n$中的仿射集，并且$E^n$中的任一仿射集均可表成此种形式。

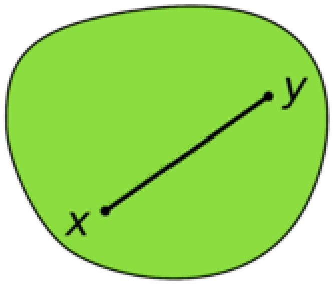
- 集合 $K=\{\ \mathbf{x}\ |\ A\mathbf{x} = \mathbf{b}\ \}$为 E^n 仿射集
  - 若 rank(A)=m，
  - 则 dim(K)=$n - m$

1.3.2 凸集（Convex Set）

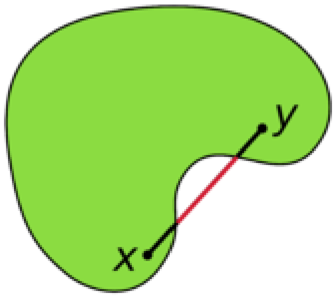
- 对任意的 $x, y \in C, \lambda \in (0,1)$，有

$$\lambda x + (1-\lambda) y \in C,$$

则C为凸集



凸集



非凸集

- E^n中有限个闭半空间的交是凸集，称为多面凸集，可表示成$\{\mathbf{x} \mid A\mathbf{x} \leq b\}$形式
- 可行区域 $D=\{\mathbf{x} \mid A\mathbf{x}=\mathbf{b}, \mathbf{x} \geq 0\}$ 为凸集。

1.3.2 凸集

- 凸集C的非空凸子集C'若满足:
  - 如果$\mathbf{u}\in C'$，且$\mathbf{u}$是C的一个线段的内点，即存在$\mathbf{x},\mathbf{y}\in C$，使$\mathrm{u}\in(\mathbf{x},\mathbf{y})$，
  - 则可推出$\mathbf{x},\mathbf{y}\in C'$，就称$C'$是$C$的一个界面。

  - 显然C本身也是一个界面。
  - 若C的界面$C'\neq C$，则C本身也是C的一个界面

1.3.3 界面、极点与极方向

- 定理：可行区域 D的子集D'为D的一个界面当且仅当存在指标集合$Q\subseteq\{1,\ldots,n\}$,使得

$$D'=\{\mathbf{x}|A\mathbf{x}=\mathrm{b},\mathbf{x}\geq 0,\text{且}x_j=0\text{当}j\in Q\}$$

  - 若dim(D)= n-m, D'为D的界面，且dim(D')=n-m-k，则|Q|≥k.

- 凸集C的1维界面称为边（edge）, 0维界面称为极点(extreme point)。若两个极点在同一条边上，称它们是相邻的。

- 极点的等价定义：$\mathbf{x}\in C$为极点,若$\mathbf{x}=\lambda\mathbf{x}_1+(1-\lambda)\mathbf{x}_2$,其中$\mathbf{x}_1,\mathbf{x}_2\in C$且$\lambda\in(0，1)$，则必有$\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}$.
- 直观的说，极点不是C中任一线段的内点。

极方向

- 设$C \neq \emptyset$为E^n中的一个凸集，$\mathbf{d} \in E^n$,d≠0。若对任一$\mathbf{x} \in C$ 及$\lambda>0$，均有$\mathbf{x}+\lambda\mathbf{d} \in C$，则称**d**为$C$的一个方向。

- 如果一个方向**d**不能表成两个方向的正线性组合，就成**d**是C的极方向（extreme direction）

- **d**是$D=\{\mathbf{x} \mid A\mathbf{x}=\mathbf{b},\ \mathbf{x}\geq\mathbf{0}\}\neq\emptyset$的方向等价于**d**满足条件$\mathbf{d}\geq 0,\ \mathbf{d}\neq 0,\ A\mathbf{d}=0$

1.4 基本可行解

假定rank(A)=m，则A中必有m个线性无关的列向量——构成满秩方阵B，A中其余各列组成子阵N,即A=(B, N)。相应的$\mathbf{x}=(\mathbf{x}_B, \mathbf{x}_N)$，则A$\mathbf{x}$=$b$可改写成

$$B\mathbf{x}_B+N\mathbf{x}_N=\mathbf{b}.$$

因为B为满秩，则B^{-1}存在，那么

$$\mathbf{x}_B=B^{-1}\mathbf{b}-B^{-1}N\mathbf{x}_N$$

任给一组值$\mathbf{x}_N$,则可得到相应的$\mathbf{x}_B$,

$$(\mathbf{x}_B, \mathbf{x}_N)$$

为A$\mathbf{x}$=b的一个解。

令$\mathbf{x}_N=0$, 则$\mathbf{x}_B=B^{-1}\mathbf{b}$, $\mathbf{x}=(B^{-1}\mathbf{b}, 0)$称为约束方程组的**基本解**。

1.4 基本可行解

- 设 rank(A)=m
  - A中m个线性无关的列向量构成满秩方阵B
  - A中其余各列组成子阵N
  - A=(B,N)
- $\boldsymbol{x}=(\boldsymbol{x}_B, \boldsymbol{x}_N)$
  - $A\boldsymbol{x}=b$ 可改写成 $B\boldsymbol{x}_B+N\boldsymbol{x}_N=b$

B称为**基**，B中的m个线性无关的列向量称为**基向量**，$\mathbf{x}_B$的m个分量称为**基变量**，其余的变量称为**非基变量**。

基本解不一定满足非负条件，故不一定是可行解。对于非负的基本解，称为**基本可行解**，此时B称为**可行基**。

1.4 基本可行解

- 定理：
  - 可行解**x**是基本可行解当且仅当它的正分量所对应的列向量线性无关
  - 可行解**x**是基本可行解当且仅当**x**为D的极点

证明：（充分性）

设$\mathbf{x}$为D的极点，设前k个分量取正值，欲证对应的列$\mathbf{a}_1,\ldots,\mathbf{a}_k$线性无关。

反证。若相关，则存在非零$\mathbf{y}=(y_1,\ldots,y_k,0,\ldots 0)^T$，使得

$$\sum_{j=1}^{k} y_j \mathbf{a}_j = 0.$$

则对任一正实数δ有：

$$\sum_{j=1}^{k} \left(x_j \pm \delta y_j\right)\mathbf{a}_j = \mathbf{b}$$

则取$x^1 = x+\delta y, x^2 = x-\delta y$，

容易验证当δ充分小时$x^1, x^2 \in D$。

因为$y \neq 0$，则$x^1 \neq x^2$，然而

$$x = \frac{1}{2}x^1 + \frac{1}{2}x^2，\text{与}x\text{为极点矛盾。}$$

故$a_1,\ldots,a_k$线性无关，从而x为基本可行解。

- 证明：（必要性）

设$\mathbf{x}$为基本可行解，设前k个分量取正值。假设存在$\mathbf{x}^1,\mathbf{x}^2 \in D$及$\lambda \in (0,1)$，使得

$$\mathbf{x} = \lambda \mathbf{x}^1 + (1-\lambda)\mathbf{x}^2$$

因为$j \geq k+1$时，$x_j = 0$，则有$x_j^1 = x_j^2 = 0$。由$A\mathbf{x}^1 = A\mathbf{x}^2 = \mathbf{b}$可得，

$$\sum_{j=1}^{k} \left(x_j^1 - x_j^2\right)\mathbf{a}_j = \mathbf{0}$$

因为$\mathbf{x}$为基本可行解，$\mathbf{a}_1,\ldots,\mathbf{a}_k$线性无关。故$x_j^1 = x_j^2 (j=1,\ldots,k)$，从而$\mathbf{x}^1 = \mathbf{x}^2 = \mathbf{x}$。

则$\mathbf{x}$为极点。

退化

- 基本可行解**x**，若它所有的基变量都取正值，则**x**为非退化的(non-degenerated); 反之，为退化的。

- 一个可行基对应一个基本可行解；反之，若一个基本可行解是非退化的，则它也对应着唯一的可行基。

- 若一个基本可行解是退化的，则它可以由不止一个可行基得到。

- 对应一个LP问题，如果它的所有基本可行解都是非退化，就说该问题是非退化的，否则为退化的。

可行区域的极方向

- 极方向的代数性质： $D=\{\mathbf{x} \mid A\mathbf{x}=\mathbf{b}, \mathbf{x}\geq\mathbf{0}\}$的方向**d**有$k$个非零分量，则**d**为$D$的极方向的充要条件是**d**的非零分量对应的列向量组秩为k-1

- 极方向的几何性质：**d**为D的极方向的充要条件是**d**为D的某个半直线界面方向

- 如果D有极方向，显然D为无界集；反之若D为无界集，则D有方向，且有极方向

1.5 线性规划基本定理

基本定理1（表示定理）

定理：设$D=\{\mathbf{x}\mid A\mathbf{x}=\mathbf{b},\mathbf{x}\geq 0\}$的所有极点为$\mathbf{x}^1,\ldots,\mathbf{x}^k$，所有极方向为$\mathbf{d}^1,\ldots,\mathbf{d}^l$，则$x\in D$的充要条件是存在一组$\lambda_i\,(i=1,\ldots,k)$和$\mu_j\,(j=1,\ldots,l)$使得

$$\mathbf{x}=\sum_{i=1}^{k}\lambda_i\mathbf{x}^i+\sum_{j=1}^{l}\mu_j\mathbf{d}^j,$$

$$\lambda_i\geq 0,\ i=1,\ldots,k;\ \mu_j\geq 0,\ j=1,\ldots,l$$

$$\sum_{i=1}^{k}\lambda_i=1$$

- 基本定理2

若LP问题

$$\min z = \mathbf{cx}$$

$$\text{s.t. } A\mathbf{x} = \mathbf{b}$$

$$\mathbf{x} \geq \mathbf{0}$$

存在有限最优值（即有最优解），则最优值必在可行区域D的某个极点上达到。并且，目标函数存在有限最优值的充要条件是对D的所有极方向$\mathbf{d}^j$,均有$\mathbf{cd}^j \geq 0$.

证明：设$D=\{\mathbf{x}\mid A\mathbf{x}=\mathbf{b},\mathbf{x}\ge\mathbf{0}\}$的所有极点为$\mathbf{x}^1,\dots,\mathbf{x}^k$，所有极方向为$\mathbf{d}^1,\dots,\mathbf{d}^l$，则由表示定理可知任意$\mathbf{x}\in D$可表示为

$$\mathbf{x}=\sum_{i=1}^{k}\lambda_i\mathbf{x}^i+\sum_{j=1}^{l}\mu_j\mathbf{d}^j,$$

$$\lambda_i\ge 0, i=1,\dots,k;\quad \mu_j\ge 0,\ j=1,\dots,l$$

$$\sum_{i=1}^{k}\lambda_i=1$$

则将关于$\mathbf{x}$的LP问题转换成关于λ和μ的LP问题。

$$\min\sum_{i=1}^{k}\left(\mathbf{c}\mathbf{x}^i\right)\lambda_i+\sum_{j=1}^{l}\left(\mathbf{c}\mathbf{d}^j\right)\mu_j$$

$$\text{s.t.}\ \sum_{i=1}^{k}\lambda_i=1;\ \lambda_i\ge 0, i=1,\dots,k;\quad \mu_j\ge 0,\ j=1,\dots,l$$

因为μ_j可以任意地大，那么若要目标函数存在下界，就需要所有的

$$\mathbf{c}\mathbf{d}^j\ge 0\quad\left(j=1,\dots,l\right)$$

或不存在极方向。

设$\mathbf{c}\mathbf{x}^p=\min\{\mathbf{c}\mathbf{x}^i\mid 1\le i\le k\}$，则可知目标函数不小于$\mathbf{c}\mathbf{x}^p$。于是令$\lambda_p=1$其余的$\lambda_i$=0，即使得目标函数达到最小值。即证明了在极点$\mathbf{x}_p$上达到最优值。

定理的注释与理解

- 定理说明：
  - 标准LP问题的目标函数的最优值必可在某个基本可行解处达到。
  - 求解标准LP问题，只需要在基本可行解的集合中进行搜索。

- 求出并比较所有基本可行解的方法通常不可行。且原因在于当n较大时，基本可行解个数将很大。

- 通常的方法根据一定的规则在基本可行解的一个子集中计算搜索，这就是单纯形法。

2. 单纯形方法（Simplex Method）

2.1 基本单纯形方法

- 主要思想：先找一个基本可行解，判断它是否为最优解。若不是，就找一个更好的基本可行解，在进行检验。如此迭代，直至最后找到最优解，或判定无界。

- 两个主要问题：
  - 寻找初始解
  - 如何判别和迭代（先考虑）

判别和迭代

- 假定rank(A)=m<n,且假定已找到一个非退化的基本可行解,即找到一个基B。则$A\mathbf{x}=\mathbf{b}$可写成

$$\mathbf{x}_B+B^{-1}N\mathbf{x}_N=B^{-1}\mathbf{b} \qquad (1)$$

记

$$B=(a_1,\ldots,a_m)$$

$$\bar{\mathbf{a}}_j=B^{-1}\mathbf{a}_j=(\bar{a}_{1j},\ldots,\bar{a}_{mj})^T,\ j=1,\ldots,n$$

$$\bar{\mathbf{b}}=B^{-1}\mathbf{b}=(\bar{b}_1,\ldots,\bar{b}_m)^T$$

$$\bar{N}=B^{-1}N$$

则上式(1)可写成$\mathbf{x}_B+\bar{N}\mathbf{x}_N=\bar{\mathbf{b}}$　　(2)

显然，不同的基对应的方程表达式也不同。

式(2)称为基B的典则方程组，简称典式。

判别和迭代

$$\mathbf{x}_B + \bar{N}\mathbf{x}_N = \bar{\mathbf{b}} \qquad (2-2)$$

显然，不同的基对应的方程表达式也不同。

式(2)称为基B的典则方程组，简称典式。

若$\bar{\mathbf{b}} \geq 0$,则式(2)就对应着基本可行解$\bar{\mathbf{x}} = \begin{pmatrix} \bar{\mathbf{b}} \\ 0 \end{pmatrix}$.

将目标函数也做相应的变换，

$$z = \mathbf{cx} = \mathbf{c}_B\bar{\mathbf{b}} - \left(\mathbf{c}_B\bar{N} - \mathbf{c}_N\right)\mathbf{x}_N = \mathbf{c}_B\bar{\mathbf{b}} - \sum_{j=m+1}^{n}\left(\mathbf{c}_B\bar{\mathbf{a}}_j - c_j\right)\mathbf{x}_j \qquad (3)$$

用z_0表示在基本可行解 $\bar{\mathbf{x}}$处的目标函数值，

则有$z_0 = \mathbf{c}\bar{\mathbf{x}} = \mathbf{c}_B\bar{\mathbf{b}}$（即上式的常数项）

判别和迭代

因为$\bar{\mathbf{a}}_1,\ldots\bar{\mathbf{a}}_m$是线性无关向量，

故当$j=1,\ldots,m$时

$\mathbf{c}_B\bar{\mathbf{a}}_j-c_j=0.$

引入记号

$\zeta_j=\mathbf{c}_B\bar{\mathbf{a}}_j-c_j,\ j=1,\ldots,n$

或用向量形式

$\zeta=\mathbf{c}_B B^{-1}A-\mathbf{c}=(\zeta_B,\zeta_N)=(\mathbf{0},\mathbf{c}_B B^{-1}N-\mathbf{c}_N)$

则式(3)改写成$z=\mathbf{c}_B\bar{\mathbf{b}}-\zeta\mathbf{x}.$

判别和迭代

$$z=\mathbf{c}_B\bar{\mathbf{b}}-\zeta\mathbf{x},\quad \zeta=\mathbf{c}_B B^{-1}A-\mathbf{c}=(\zeta_B,\zeta_N)=(\mathbf{0},\mathbf{c}_B B^{-1}N-\mathbf{c}_N)$$

经过变量变换后，LP问题可叙述成

$$\min z=z_0-\zeta\mathbf{x} \qquad (4)$$

$$s.t.\quad \mathbf{x}_B+B^{-1}N\mathbf{x}_N=\bar{\mathbf{b}}, \qquad (5)$$

$$\mathbf{x}\geq\mathbf{0}.$$

定理2.1:若（4）中$\zeta\leq 0$，则$\bar{\mathbf{x}}$为最优解。

定理2.2:若（4）中ζ存在某一分量$\zeta_k>0$，且$\bar{\mathbf{a}}_k\leq\mathbf{0}$，则原问题无解。

定理2.3:若（4）中有$\zeta_k>0$，且$\bar{\mathbf{a}}_k$至少存在一个正分量，则能找到基本可行解$\hat{\mathbf{x}}$，使得$\mathbf{c}\hat{\mathbf{x}}<\mathbf{c}\bar{\mathbf{x}}$。

构造（更好的基本可行解）$\hat{\mathbf{x}}$。令

$$\theta = \min\left\{ \frac{\bar{b}_i}{\bar{a}_{ik}} \middle| \bar{a}_{ik} > 0, i = 1, \ldots, m \right\} = \frac{\bar{b}_r}{\bar{a}_{rk}} > 0$$

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \theta\left(\begin{pmatrix} -\bar{\mathbf{a}}_k \\ 0 \end{pmatrix} + \mathbf{e}_k \right) = \begin{pmatrix} \bar{\mathbf{b}} - \theta\bar{\mathbf{a}}_k \\ 0 \end{pmatrix} + \theta\mathbf{e}_k .$$

下面将证明$\hat{\mathbf{x}}$是可行解。带入式(5),

$$\hat{\mathbf{x}}_B + \bar{N}\hat{\mathbf{x}}_N = \bar{\mathbf{b}} - \theta\bar{\mathbf{a}}_k + \sum_{j=m+1}^{n} \hat{x}_j \bar{\mathbf{a}}_j = \bar{\mathbf{b}} - \theta\bar{\mathbf{a}}_k + \theta\bar{\mathbf{a}}_k = \bar{\mathbf{b}}.(\text{注意到}\hat{x}_N = (0, \ldots, \theta, 0..0))$$

根据θ的构造，显然可知$\hat{\mathbf{x}} \geq 0$。故$\hat{\mathbf{x}}$是可行解。

下面再证明$\hat{\mathbf{x}}$是基本解。$\hat{\mathbf{x}}$的各分量为

$$\hat{x}_i = \bar{b}_i - \left(\bar{b}_r / \bar{a}_{rk}\right)\bar{a}_{ik}, i = 1, \ldots, m$$

$$\hat{x}_k = \bar{b}_k / \bar{a}_{rk}$$

$$\hat{x}_i = 0, j = m+1, \ldots, n, j \neq k.$$

由于这里$\hat{x}_r = 0$而$\hat{x}_k > 0$,

故只需要证明$a_1, \ldots, a_{r-1}, a_k, a_{r+1}, \ldots, a_m$线性无关。

证明$a_1,\ldots,a_{r-1},a_k,a_{r+1},\ldots,a_m$线性无关

反证法。若相关，则因为$\{\mathbf{a}_1,\ldots\mathbf{a}_m\}$是线性无关的，

从而$\mathbf{a}_k$必是其余的$m-1$个向量的线性组合。

即$\mathbf{a}_k = \sum_{i=1,i\neq r}^{m} y_i\mathbf{a}_i$

又$\bar{\mathbf{a}}_k = B^{-1}\mathbf{a}_k$，故$\mathbf{a}_k = B\bar{\mathbf{a}}_k = \sum_{i=1}^{m}\bar{a}_{ik}\mathbf{a}_i$

两式相减得

$$\bar{a}_{rk}\mathbf{a}_r + \sum_{i=1,i\neq r}^{m}(\bar{a}_{ik} - y_i)\mathbf{a}_i = \mathbf{0}$$

因为$\bar{a}_{rk} \neq 0$，故$\mathbf{a}_1,\ldots\mathbf{a}_m$线性无关，引出矛盾。

最后

$$z(\hat{\mathbf{x}}) = z_0 - \boldsymbol{\zeta}\hat{\mathbf{x}} = z_0 - \zeta_k\theta < z_0.$$

单纯形法的计算步骤

ζ称之为检验向量。
在迭代过程中，如果ζ有不止一个正分量，
则为了使目标函数下降得较快，
一般取最大的分量ζ_k所对应的列向量$\mathbf{a}_k$进入基。

1. 找初始的可行基;
2. 求出对应的典式;
3. 求$\zeta_k = \max\{\zeta_j \mid j = 1, \ldots, n\}$
4. 若$\zeta_k \le 0$，停止。已找到最优解$\mathbf{x} = \begin{pmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{pmatrix} = \begin{pmatrix} \overline{\mathbf{b}} \\ \mathbf{0} \end{pmatrix}$及最优值$z = \mathbf{c}_B \overline{\mathbf{b}}$
5. 若$\overline{\mathbf{a}}_k \le \mathbf{0}$，停止，原问题无界。
6. 求最小比$\dfrac{\bar{b}_r}{\bar{a}_{rk}} = \min\left\{ \dfrac{\bar{b}_i}{\bar{a}_{ik}} \middle| \bar{a}_{ik} > 0 \right\}$
7. 以$\mathbf{a}_k$代替$\mathbf{a}_{B_r}$得到新的基，回到2.

2.2 单纯形表

根据θ和$\hat{\mathbf{x}}$的构造方式可知，得到改进的基本可行解——

即使原来的非基变量x_k变成取正值的基本变量，

同时使原来的基本变量x_r值为零（变成非基变量）。

等价地，基$B=[\mathbf{a}_1,\dots,\mathbf{a}_m]$变成另一个基$\hat{B}=[\mathbf{a}_1,\dots,\mathbf{a}_r,\mathbf{a}_k,\mathbf{a}_{r+1},\dots\mathbf{a}_m]$.

* 换基变换

* x_k称为进基变量，x_r为离基变量。

在基B下，典式的系数增广矩阵为

$$
\begin{array}{cccccccccc}
x_1 & \dots & x_r & \dots & x_m & x_{m+1} & \dots & x_k & \dots & x_n & \\
1 & \dots & 0 & \dots & 0 & \bar{a}_{1,m+1} & \dots & \bar{a}_{1,k} & \dots & \bar{a}_{1,n} & \bar{b}_1 \\
\vdots & \ddots & \vdots & & \vdots & \vdots & & \vdots & & \vdots & \vdots \\
0 & \dots & 1 & \dots & 0 & \bar{a}_{r,m+1} & \dots & \bar{a}_{r,k} & \dots & \bar{a}_{r,n} & \bar{b}_r \\
\vdots & & \vdots & \ddots & \vdots & \vdots & & \vdots & & \vdots & \vdots \\
0 & \dots & 0 & \dots & 1 & \bar{a}_{m,m+1} & \dots & \bar{a}_{m,k} & \dots & \bar{a}_{m,n} & \bar{b}_m
\end{array}
$$

x_r与x_k的地位互换。

第k列通过初等变换变成单位向量$\mathbf{e}_r$.

1)第r行$\bar{\mathbf{a}}_r^T$除以$\bar{a}_{rk}$, $\hat{\mathbf{a}}_r^T=\bar{\mathbf{a}}_r^T/\bar{a}_{rk}$

2)对第i行（$i\neq r$）$\bar{\mathbf{a}}_i^T$，$\hat{\mathbf{a}}_r^T=\bar{\mathbf{a}}_i^T-\hat{\mathbf{a}}_r^T\bar{a}_{ik}$

即第i行减去新的第r行$\hat{\mathbf{a}}_r^T$乘以$\bar{a}_{ik}$

最后一列正是基$\hat{B}$所对应的基本可行解。

$\hat{b}_r=\bar{b}_r/\bar{a}_{rk}, \hat{b}_i=\bar{b}_i-(\bar{b}_r/\bar{a}_{rk})\bar{a}_{ik}, i\neq r$

换基时，目标函数$z=\mathbf{c}_B\bar{\mathbf{b}}-\left(\mathbf{c}_B\bar{N}-\mathbf{c}_N\right)\mathbf{x}_N$也要做相应的调整，即将$B$换成$\hat{B}$。
将上式的等价形式

$z+\left(\mathbf{c}_B B^{-1}N-\mathbf{c}_N\right)\mathbf{x}_N=\mathbf{c}_B\bar{\mathbf{b}}$，或$z+\zeta_N\mathbf{x}_N=z_0$

看成一个方程，放在典式中一同进行初等变换。

	z	x_1	$\cdots$	x_r	$\cdots$	x_m	x_{m+1}	$\cdots$	x_k	$\cdots$	x_n	RHS
z	1	0	$\cdots$	0	$\cdots$	0	ζ_{m+1}	$\cdots$	ζ_k	$\cdots$	ζ_n	z_0
x_1	0	1	$\cdots$	0	$\cdots$	0	$\bar{a}_{1,m+1}$	$\cdots$	$\bar{a}_{1,k}$	$\cdots$	$\bar{a}_{1,n}$	$\bar{b}_1$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_r	0	0	$\cdots$	1	$\cdots$	0	$\bar{a}_{r,m+1}$	$\cdots$	$\bar{a}_{r,k}$	$\cdots$	$\bar{a}_{r,n}$	$\bar{b}_r$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_m	0	0	$\cdots$	0	$\cdots$	1	$\bar{a}_{m,m+1}$	$\cdots$	$\bar{a}_{m,k}$	$\cdots$	$\bar{a}_{m,n}$	$\bar{b}_m$

单纯形表

转轴列

转轴行

转轴元

单纯形表可简记为

	z	$\mathbf{x}_B$	$\mathbf{x}_N$	RHS
z	1	$\mathbf{0}$	$\boldsymbol{\zeta}_N$	z_0
$\mathbf{x}_B$	$\mathbf{0}$	I	$\bar{N}$	$\bar{\mathbf{b}}$

上述变换又称之为旋转（pivot），
类似于解线性方程组的主元消去法.

2.3 初始解:两阶段法

设原问题为

$$\min \mathbf{cx}$$
$$s.t.\ \ A\mathbf{x} = \mathbf{b}\ (\mathbf{b} \geq \mathbf{0})$$
$$\mathbf{x} \geq \mathbf{0}$$

引入m个人工变量$\mathbf{x}_a = (x_{n+1}, \ldots, x_{n+m})^T$,考虑如下辅助问题

$$\min g = \mathbf{1x}_a$$
$$s.t.\ \ A\mathbf{x} + \mathbf{x}_a = \mathbf{b}$$
$$\mathbf{x}, \mathbf{x}_a \geq 0$$

其中$\mathbf{1} = (1, \ldots, 1)$。设原问题的可行域为$D$,辅助问题的可行域为$D'$,

显然$\mathbf{x} \in D \Leftrightarrow \begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix} \in D'$。而$\begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix} \in D' \Leftrightarrow \min g = 0.$

2.3 初始解:两阶段法

m个人工变量$\mathbf{x}_a=(x_{n+1},\ldots,x_{n+m})^T$,考虑如下辅助问题

$$\min\ g=\mathbf{1}\mathbf{x}_a$$
$$s.t.\quad A\mathbf{x}+\mathbf{x}_a=\mathbf{b}$$
$$\mathbf{x},\mathbf{x}_a\geq 0$$

$$\begin{pmatrix}\mathbf{x}\\0\end{pmatrix}\in D'\Leftrightarrow \min\ g=0.$$

对于辅助问题来说，一个基本可行解是$\begin{pmatrix}\mathbf{x}\\\mathbf{x}_a\end{pmatrix}=\begin{pmatrix}\mathbf{0}\\\mathbf{b}\end{pmatrix}$,

于是可进行单纯形迭代求解，其结果有两种可能：

1) $\min\ g>0$。说明$D=\varnothing$.

2) $\min\ g=0$。这是自然有$\mathbf{x}_a=0$,把它除去后即得到原问题的一个可行解。

2.3 初始解:两阶段法

m个人工变量$\mathbf{x}_a=\left(x_{n+1},\ldots,x_{n+m}\right)^T$,得到辅助问题

$$\begin{aligned}&\min\ g=\mathbf{1x}_a\\&s.t.\quad A\mathbf{x}+\mathbf{x}_a=\mathbf{b}\\&\qquad\quad \mathbf{x},\mathbf{x}_a\geq 0\end{aligned}$$

$$\begin{pmatrix}\mathbf{x}\\0\end{pmatrix}\in D'\Leftrightarrow \min\ g=0.$$

* 若此时所有人工变量都是非基变量，则解为基本可行解。
* 否则，进行旋转变换消去基变量中的人工变量。

  设基变量x_r为人工变量，

  ▹ 取第r行的前n个元素中不为零的元素$\bar{a}_{rs}$为转轴元进行变换，则非人工变量x_s进入基变量，同时从基本量消除了人工变量x_r。

  ▹ 若第r行的前n个元素都为零，则直接去掉该行以及对应的人工变量。

2.4 初始解:大M法

设原问题为

$$\min \mathbf{cx}$$
$$s.t.\ \ A\mathbf{x} = \mathbf{b}\ (\mathbf{b} \geq \mathbf{0})$$
$$\mathbf{x} \geq \mathbf{0}$$

引入$\mathbf{x}_a \in E^m$及足够大的正数M,得到新问题如下

$$\min \mathbf{cx} + M\,\mathbf{1x}_a$$
$$s.t.\ \ A\mathbf{x} + \mathbf{x}_a = \mathbf{b}$$
$$\mathbf{x}, \mathbf{x}_a \geq 0$$

其中$\mathbf{1} = (1,\ldots,1)$。只要$M$取得足够大，那么可以说

x是原问题的最优解 $\Leftrightarrow \begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix}$是新问题的最优解。

而新问题有初始解$\begin{pmatrix} \mathbf{0} \\ \mathbf{b} \end{pmatrix}$，因此总可以用单纯形法求解。

2.5 退化与防止循环

- 若单纯形表中右端向量存在分量为0，即出现退化解的情况下，可能出现循环。为了避免循环，需要再补充一些旋转规则。
  - 字典序方法
  - Bland规则

字典序方法

非零向量$\mathbf{x}$,并且第一个非零分量是非负的，称为按字典序非负，记为$\mathbf{x} \succ = \mathbf{0}$。

对向量$\mathbf{x}, \mathbf{y}$,若$\mathbf{x} - \mathbf{y} \succ = \mathbf{0}$,则称$\mathbf{x}$按字典序大于等于$\mathbf{y}$.

若一组$\mathbf{x}^i$,有$\mathbf{x}^t$,使得对所有$\mathbf{x}^i$,均有$\mathbf{x}^i \succ = \mathbf{x}^t$,则称$\mathbf{x}^t$为该组向量中按字典序最小的。

记为：$\mathbf{x}^t = lex \min \mathbf{x}^i$.

字典序法就是在选定了进基变量x_k后，令

$$\frac{\mathbf{p}_r}{\bar{a}_{rk}} = lex \min \left\{ \frac{\mathbf{p}_i}{\bar{a}_{ik}} \middle| \bar{a}_{ik} > 0, i = 1, \ldots, m \right\},$$其中$\mathbf{p}_i$为$\left(\bar{\mathbf{b}}, B^{-1}\right)$的第$i$行。

Bland规则

1) 选下标最小的正检验数ζ_k所对应的非基变量x_k作为进基变量。

2) 离基变量x_l的确定：如果同时有几个$\dfrac{b_r}{\bar{a}_{rk}}$达到最小，就选其中下标最小的那个基变量作为离基变量。即

$$l=\min\left\{r\left|\frac{b_r}{\bar{a}_{rk}}=\min\left\{\frac{b_i}{\bar{a}_{ik}}\middle|\bar{a}_{ik}>0\right\}\right.\right\}$$

- Bland规则的优点是简单易行，但缺点是只考虑最小下标，而不考虑目标函数值下降的快慢。故其效率比字典序或原来单纯形法低。
- 在实际中，退化是常见的，但退化不一定产生循环。事实上，产生循环是较罕见的。

2.6 修改单纯形法

- 当LP问题的规模很大时，如何减少存储量和计算时间，就是一个必须加以考虑的问题。在实际中，大多采用修改单纯形法。

逆矩阵法

一般单纯形表为

$$\begin{bmatrix} & z & \mathbf{x}_B & \mathbf{x}_N & RHS \\ z & 1 & \mathbf{0} & \mathbf{c}_B B^{-1}N-\mathbf{c}_N & \mathbf{c}_B B^{-1}\mathbf{b} \\ \mathbf{x}_B & \mathbf{0} & I & B^{-1}N & B^{-1}\mathbf{b} \end{bmatrix}$$

逆矩阵法对每一张单纯形表，仅存贮下列数据

$$\begin{bmatrix} \mathbf{w} & z_o \\ B^{-1} & \overline{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_B B^{-1} & \mathbf{c}_B B^{-1}\mathbf{b} \\ B^{-1} & B^{-1}\mathbf{b} \end{bmatrix},$$ 该表称为修改单纯形表。

根据这张表和原始数据进行迭代计算。由$\zeta_k = \max\{\mathbf{wa}_j - c_j \mid x_j\text{非基变量}\}$确定$x_k$为进基变量再求$\overline{\mathbf{a}}_k = B^{-1}\mathbf{a}_k$，用最小比确定离基变量$x_r$,得到新的基$\hat{B}$,最后构造对应于$\hat{B}$的修改单纯形表。如果每一次迭代都必须从构造$B^{-1}$,则计算量显然难以令人满意。

定理：对应基B的修改单纯形表的右边添加一列$\begin{pmatrix} \zeta_k \\ \overline{\mathbf{a}}_k \end{pmatrix}$,以$\overline{\mathbf{a}}_{rk}$为转轴元旋转，就得到了对应于基$\hat{B}$的修改单纯形表。

3. 最优性条件和对偶理论

对偶性

- 对于每一个线性规划P，总存在着另一个线性规划D，两者之间存在着密切的联系，人们可以通过求解对偶规划D来获得原规划P的最优解。

3.1 Kuhn Tucker条件

- 定理

设A为$m\times n$阶矩阵，$\mathbf{b}\in E^m,\mathbf{c}\in E^n,\mathbf{x}\in E^n$,则$\mathbf{x}^*$为$LP$问题

$$
\begin{aligned}
&\min \mathbf{cx}\\
&s.t.\quad A\mathbf{x}\ge\mathbf{b}\\
&\qquad\quad \mathbf{x}\ge\mathbf{0}
\end{aligned}
$$

的最优解的充要条件是存在$\mathbf{w}\in E^m,\mathbf{v}\in E^n$,使得下列$K-T$条件得到满足

$$
\begin{aligned}
&A\mathbf{x}^*\ge\mathbf{b},\mathbf{x}^*\ge\mathbf{0}\\
&\mathbf{c}-\mathbf{w}A-\mathbf{v}=\mathbf{0},\mathbf{w}\ge\mathbf{0},\mathbf{v}\ge\mathbf{0}\\
&\mathbf{w}(A\mathbf{x}^*-\mathbf{b})=0,\mathbf{vx}^*=0
\end{aligned}
$$

若将$\mathbf{v}$消去（利用$\mathbf{c}-\mathbf{w}A-\mathbf{v}=\mathbf{0}$），可得$K-T$条件的另一种形式

$$
\begin{aligned}
&A\mathbf{x}^*\ge\mathbf{b}，\ \mathbf{x}^*\ge\mathbf{0}\\
&\mathbf{w}A\le\mathbf{c}，\ \mathbf{w}\ge\mathbf{0}\\
&\mathbf{w}(A\mathbf{x}^*-\mathbf{b})=0，(\mathbf{w}A-\mathbf{c})\mathbf{x}^*=0
\end{aligned}
$$

3.1 Kuhn Tucker条件

对于标准形式的LP问题的$K-T$条件

只要将$A\mathbf{x}=\mathbf{b}$改写成$\begin{pmatrix} A \\ -A \end{pmatrix}\mathbf{x} \geq \begin{pmatrix} \mathbf{b} \\ -\mathbf{b} \end{pmatrix}$，可推出相应的$K-T$条件

$$A\mathbf{x}=\mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

$$\mathbf{c}-\mathbf{w}A-\mathbf{v}=\mathbf{0}, \mathbf{v} \geq \mathbf{0}$$

$$\mathbf{v}\mathbf{x}=0$$

或等价的

$$A\mathbf{x}=\mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

$$\mathbf{w}A \leq \mathbf{c}$$

$$(\mathbf{w}A-\mathbf{c})\mathbf{x}=0.$$

3.1 Kuhn Tucker条件

一个基本可行解$\begin{pmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{b}} \\ \mathbf{0} \end{pmatrix}$($\bar{\mathbf{b}} > \mathbf{0}$)为最优的充要条件是检验向量$\zeta \leq 0$，

和$K-T$是什么关系呢？

由$\mathbf{vx} = \mathbf{v}_B\mathbf{x}_B + \mathbf{v}_N\mathbf{x}_N = 0$，且 $\mathbf{x}_B > 0$，可知$\mathbf{v}_B = 0$.

条件$\mathbf{c} - \mathbf{w}A - \mathbf{v} = \mathbf{0}$可改写为$(\mathbf{c}_B, \mathbf{c}_N) - \mathbf{w}(B, N) - (\mathbf{v}_B, \mathbf{v}_N) = (\mathbf{0}, \mathbf{0})$，于是有

$$\begin{cases} \mathbf{c}_B - \mathbf{w}B = \mathbf{0} \\ \mathbf{c}_N - \mathbf{w}N - \mathbf{v}_N = \mathbf{0} \end{cases} \Rightarrow \begin{cases} \mathbf{w} = \mathbf{c}_B B^{-1} \\ \mathbf{v}_N = \mathbf{c}_N - \mathbf{w}N = \mathbf{c}_N - \mathbf{c}_B B^{-1} N \end{cases}$$

这就是说，若给一个基本可行解x,如果取

$\mathbf{v}_B = 0, \mathbf{w} = \mathbf{c}_B B^{-1}, \mathbf{v}_N = \mathbf{c}_N - \mathbf{w}N = \mathbf{c}_N - \mathbf{c}_B B^{-1} N$,

的$K-T$条件除了$\mathbf{v} \geq \mathbf{0}$外，其余条件都已满足。

而根据v的取法，易知$v \geq \mathbf{0} \Leftrightarrow \zeta \leq \mathbf{0}$.

3.2 对偶理论

- 对偶问题可被看作是原始问题的“行列转置”

$$\boxed{\min \quad \mathbf{c}^T\mathbf{x} \qquad \text{s.t.} \quad \mathbf{Ax}=\mathbf{b}, \mathbf{x}\geq 0}$$

$$\min \quad c^T x \qquad \text{s.t.} \quad \begin{bmatrix} A \\ -A \end{bmatrix} x \geq \begin{bmatrix} b \\ -b \end{bmatrix}, x \geq 0$$

- 对偶问题是：

$$\max \quad \begin{bmatrix} b^T & -b^T \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad \text{s.t.} \quad \begin{bmatrix} A^T & -A^T \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \leq c$$

- 其中y_1和y_2分别表示对应于约束Ax>=b和-Ax<=B的对偶变量组。令$y=y_1-y_2$

$$\boxed{\max \quad \mathbf{b}^T\mathbf{y} \qquad \text{s.t.} \qquad \mathbf{A}^T\mathbf{y} \leq \mathbf{c}}$$

3.2 对偶理论

对偶规划生成规则：

⑴ $\min \to \max$；

（2）$\mathbf{c} \to \mathbf{b}$；　（3）$A \to A^T$；

（4）按以下规则添上不等号

原问题		对偶问题	
变	≥ 0	行	$\leq$
量	≤ 0	约	$\geq$
	无限制	束	$=$
行	$\geq$		≥ 0
约	$\leq$	变	≤ 0
束	$=$	量	无限制

LP问题(P)

$$\min \mathbf{cx}$$
$$s.t.\quad A\mathbf{x} \geq \mathbf{b}$$
$$\mathbf{x} \geq \mathbf{0}$$

对偶问题(D)为

$$\max \mathbf{wb}$$
$$s.t.\quad \mathbf{w}A \leq \mathbf{c}$$
$$\mathbf{w} \geq \mathbf{0}$$

3.2 对偶理论

对偶规划生成规则：

(1) min $\to$ max；

(2) $\mathbf{c} \to \mathbf{b}$； (3) $A \to A^T$；

(4) 按以下规则添上不等号

原问题		对偶问题	
变	≥ 0	行	$\leq$
量	≤ 0	约	$\geq$
	无限制	束	$=$
行	$\geq$		≥ 0
约	$\leq$	变	≤ 0
束	$=$	量	无限制

标准LP问题

$$\begin{aligned} &\min \mathbf{cx} \\ &s.t. \quad A\mathbf{x} = \mathbf{b} \\ &\qquad\quad \mathbf{x} \geq \mathbf{0} \end{aligned}$$

对偶问题为

$$\begin{aligned} &\max \mathbf{wb} \\ &s.t. \quad \mathbf{w}A \leq \mathbf{c} \end{aligned}$$

3.2 对偶理论

对偶规划生成规则：

(1) min $\to$ max；

（2）$\mathbf{c} \to \mathbf{b}$；　（3）$A \to A^T$；

（4）按以下规则添上不等号

原问题		对偶问题	
变量	≥ 0	行约束	$\leq$
	≤ 0		$\geq$
	无限制		$=$
行约束	$\geq$	变量	≥ 0
	$\leq$		≤ 0
	$=$		无限制

*LP*问题

$$
\begin{aligned}
&\min \mathbf{c}_1\mathbf{x}_1 + \mathbf{c}_2\mathbf{x}_2 \\
&s.t. \quad A_{11}\mathbf{x}_1 + A_{12}\mathbf{x}_2 \geq \mathbf{b}_1 \\
&\qquad\ \ A_{21}\mathbf{x}_1 + A_{22}\mathbf{x}_2 = \mathbf{b}_2 \\
&\qquad\ \ A_{31}\mathbf{x}_1 + A_{32}\mathbf{x}_2 \leq \mathbf{b}_3 \\
&\qquad\ \ \mathbf{x}_1 \geq \mathbf{0}
\end{aligned}
$$

对偶问题为

$$
\begin{aligned}
&\max \mathbf{w}_1\mathbf{b}_1 + \mathbf{w}_2\mathbf{b}_2 + \mathbf{w}_3\mathbf{b}_3 \\
&s.t. \quad \mathbf{w}_1 A_{11} + \mathbf{w}_2 A_{21} + \mathbf{w}_3 A_{31} \leq \mathbf{c}_1 \\
&\qquad\ \ \mathbf{w}_1 A_{12} + \mathbf{w}_2 A_{22} + \mathbf{w}_3 A_{32} = \mathbf{c}_2 \\
&\qquad\ \ \mathbf{w}_1 \geq \mathbf{0}, \mathbf{w}_3 \leq \mathbf{0}
\end{aligned}
$$

对偶定理

- 设**x**和**w**分别为(P)和(D)的可行解，则有**cx≥wb**.
- 设**x***和**w***分别为(P)和(D)的可行解，则**x***和**w***分别为(P)和(D)的最优解的充要条件是**cx*=w*b**.
- 推论：若(P)有最优解**x***，则(D)有最优解**w***，且**cx*=w*b**；若(P)无界，则(D)无解。反之亦然。
- 互补松弛性：设**x***和**w***分别为(P)和(D)的最优解，则有**w***(A**x***-**b**)=**0**, (**c**-**w***A)**x***=**0**。

$K-T$条件:

$A\mathbf{x}^* \geq \mathbf{b}$，$\mathbf{x}^* \geq \mathbf{0}$ （原始可行性）

$\mathbf{w}A \leq \mathbf{c}, \mathbf{w} \geq \mathbf{0}$ （对偶可行性）

$\mathbf{w}(A\mathbf{x}^* - \mathbf{b}) = 0$，$(\mathbf{w}A - \mathbf{c})\mathbf{x}^* = 0$ （互补松弛性）

对偶定理

- 从K-T条件与单纯形方法中的最优准则的关系中可以看出：在单纯形方法中，除了K-T条件的**v**≥**0**外，其余都得到满足。即单纯形法就是保持了原始可行性、互补松弛性、**c**-**w***A*-**v**=**0**得到满足的条件下逐步改善基本可行解，使得对偶可行性中的**v**=**c**-**w***A*≥**0**得到满足。

- 启发：也可以在保证对偶可行性和互补松弛性满足的条件下，改善解使得原始可行性得到满足。

3.3 对偶单纯形法

$$(P)\ \min \mathbf{cx} \qquad\qquad (D)\ \max \mathbf{wb}$$
$$s.t. A\mathbf{x} = b \qquad\qquad s.t.\ \ \mathbf{w}A \le \mathbf{c}$$
$$\mathbf{x} \ge 0$$

D的基本可行解：设$A=(B,N)$，其中B为满秩方阵，则$\mathbf{w}B=\mathbf{c}_B$的解$\overline{\mathbf{w}}=\mathbf{c}_B B^{-1}$为$(D)$的基本解，若$\overline{\mathbf{w}}N \le \mathbf{c}_N$，则称$\overline{\mathbf{w}}$为$(D)$的基本可行解。

P的正则解：若原问题(P)的一个基本解$\mathbf{x}=\begin{pmatrix} B^{-1}\mathbf{b} \\ 0 \end{pmatrix}$对应的检验向量$\zeta=(\mathbf{0},\mathbf{c}_B B^{-1}N-\mathbf{c}_N)\le \mathbf{0}$,

则称$\mathbf{x}$为问题(P)的正则解。

此时的基B称为正则基。容易验证，D的基本可行解和P的正则解是一一对应的。

同单纯形法一样，求解对偶规划(D)从(D)的一个基本可行解迭代到另一个基本可行解，使得目标函数值增加。等价地，求解原规划(P)从一个正则解开始，迭代到另一个正则解，使目标函数$z=\mathbf{wb}=\mathbf{c}_B B^{-1}\mathbf{b}=\mathbf{cx}$增加，直到$B^{-1}\mathbf{b}\ge 0$，即正则解满足原始可行性时，也就找到最优解。这种方法称为对偶单纯性法。

3.3 对偶单纯形法

$$(P)\ \min \mathbf{cx} \qquad\qquad (D)\ \max \mathbf{wb}$$

$$s.t. A\mathbf{x} = b \qquad\qquad s.t.\ \ \mathbf{w}A \le \mathbf{c}$$

$$\mathbf{x} \ge 0$$

1。 找一个正则基B，建立单纯形表。

2。 若$\bar{\mathbf{b}} = B^{-1}\mathbf{b} \ge 0$，停止，已找到原问题最优解；

否则令$\bar{b}_r = \min\{\bar{b}_i \mid i = 1,\ldots,m\}$.

3。 若$\bar{\mathbf{a}}^r \ge 0$,停止，原问题无解；

否则令$\dfrac{\zeta_k}{\bar{a}_{rk}} = \min\{\dfrac{\zeta_j}{\bar{a}_{rj}} \mid a_{rj} < 0\}$

4。 以$\bar{a}_{rk}$为转转轴元旋转，返回2。

3.4 原始—对偶单纯形法

- 同时对对偶问题和第一阶段地辅助问题求解。从对偶问题的任意可行解开始，同时在迭代过程中保持对偶可行性和互补松弛性以及$x \geq 0$的情况下,通过剔除人工变量使$Ax=b$成立。

$$(P)\ \min \mathbf{cx} \qquad (D)\ \max \mathbf{wb}$$
$$s.t.\ \ A\mathbf{x} = b \qquad s.t.\ \ \mathbf{w}A \le \mathbf{c}$$
$$\mathbf{x} \ge 0$$

在(P)中引入人工变量后，得到辅助问题

$$\min g = \mathbf{1x}_a$$
$$s.t\ \ A\mathbf{x} + \mathbf{x}_a = \mathbf{b} \quad (\mathbf{b} \ge 0)$$
$$\mathbf{x}, \mathbf{x}_a \ge \mathbf{0}$$

若已知(D)的一个可行解$\overline{\mathbf{w}}$，为保持互补松弛性$(\mathbf{x}(\overline{\mathbf{w}}A - \mathbf{c}) = \mathbf{0})$,将令

$x_j = 0$（当$\overline{\mathbf{w}}\mathbf{a}_j \ne c_j$)，于是辅助问题为

$$(P')\ \min g = \mathbf{1x}_a$$
$$s.t\ \ A\mathbf{x} + \mathbf{x}_a = \mathbf{b}$$
$$x_j = 0,\ j \notin Q$$
$$x_j \ge 0,\ j \in Q$$
$$\mathbf{x}_a \ge \mathbf{0}$$

其中$Q = \{ j \mid \overline{\mathbf{w}}\mathbf{a}_j = c_j,\ j = 1,\ldots,n\}$。问题$(P')$称为对应于$\overline{\mathbf{w}}$的限定问题。

求解问题(P')，得到最优解$\begin{pmatrix} \mathbf{x}^* \\ \mathbf{x}_a^* \end{pmatrix}$，则它是$(P)$的辅助问题的基本可行解。

若$\mathbf{x}_a^* = \mathbf{0}$，则$\mathbf{x}^*$为问题$(P)$的最优解（因为$\mathbf{x}^*$和$\overline{\mathbf{w}}$分别是$(P)$和$(D)$的可行解，且满足互补松弛性）。

若$\mathbf{x}_a^* \ne \mathbf{0}$，找$(D)$的另一个基本可行解$\hat{\mathbf{w}}$，使得$(D)$目标函数值有所增加，同时对应于$\hat{\mathbf{w}}$的限定问题的最优值将较对应于$\overline{\mathbf{w}}$的限定问题的最优值有所减少（等价于剔除人工变量）。

考虑(P')的对偶问题(D')

$$(D')\ \max\ \mathbf{vb}$$
$$s.t.\ \ \mathbf{va}_j \le 0,\ j \in Q$$
$$\mathbf{v} \le \mathbf{1}$$

令$\mathbf{v}^*$为其最优解。如对于所有的$j=1,\ldots,n$，都要$\mathbf{v}^*\mathbf{a}_j \le \mathbf{0}$，则说明$\mathbf{v}^*$还是辅助问题的对偶问题的最优解，于是$\begin{pmatrix}\mathbf{x}^* \\ \mathbf{x}_a^*\end{pmatrix}$就是辅助问题的最优解。由于$\mathbf{x}_a^* \ne 0$，故原问题$(P)$无解。

构造$\hat{\mathbf{w}}=\overline{\mathbf{w}}+\theta\mathbf{v}^*$，其中

$$\theta=\min\left\{-\frac{\overline{\mathbf{w}}\mathbf{a}_j-c_j}{\mathbf{v}^*\mathbf{a}_j}\,\middle|\,\mathbf{v}^*\mathbf{a}_j>0,\ j=1,\ldots,n\right\}.$$

可以证明$\hat{\mathbf{w}}$满足要求。同时，$\begin{pmatrix}\mathbf{x}^* \\ \mathbf{x}_a^*\end{pmatrix}$还是对应于$\hat{\mathbf{w}}$的限定问题的一个基本可行解，因此可以使用它来作为初值来求解对应于$\hat{\mathbf{w}}$的限定问题。如此循环。。。

1。变求解的规划为

$$\min \mathbf{cx}$$
$$s.t.\ \ A\mathbf{x}=b$$
$$\mathbf{x}\geq 0$$

的形式。找其对偶规划的初始解$\mathbf{w}$，满足$\mathbf{w}A\leq \mathbf{c}$。（当$\mathbf{c}\geq 0$时，可取$\mathbf{w}$＝0作为初始解）

2。令$Q=\{j\mid \overline{\mathbf{w}}\mathbf{a}_j=c_j,\ j=1,\ldots,n\}$。求解$\mathbf{w}$的限定问题

$$\min g=\mathbf{1x}_a$$
$$s.t\ \ A\mathbf{x}+\mathbf{x}_a=\mathbf{b}$$
$$x_j=0,\ j\notin Q$$
$$x_j\geq 0,\ j\in Q$$
$$\mathbf{x}_a\geq \mathbf{0}$$

若它的最优值$g=0$，停止，以及找到原问题的最优解；否则，求其对偶问题并令其最优解为$\mathbf{v}$。

3。若$\mathbf{v}A\leq \mathbf{0}$，则停止，原问题无解；否则令

$$\theta=\min\left\{-\frac{\overline{\mathbf{w}}\mathbf{a}_j-c_j}{\mathbf{v}^*\mathbf{a}_j}\,\middle|\,\mathbf{v}^*\mathbf{a}_j>0,\ j=1,\ldots,n\right\},$$ 并且构造$\mathbf{w}=\mathbf{w}+\theta\mathbf{v}^*$，返回2。

3.5 对偶初始解

- 对偶单纯形法和原始－对偶单纯形法都需要一个对偶可行解。而且对偶单纯形法还要求是一个基本可行解。

对应任意一个LP问题，总可以使用高斯消去法找到一个基B,并化为典式。

此时若$\boldsymbol{\zeta}_N \le \mathbf{0}$,则说明已经找到对偶问题的一个基本可行解$w = c_B B^{-1}$.

否则，增加一个约束$\mathbf{1}\mathbf{x}_N \le M$, M是一个充分大的参数（从而这一约束不影响原问题）

增加约束后，称为扩充问题。（可证明：原问题有可行解 $\Leftrightarrow$ 对于充分大的M,扩充问题有可行解）

	z	$\mathbf{x}_B$	$\mathbf{x}_N$	x_{n+1}	RHS
z	1	0	$\boldsymbol{\zeta}_N$	0	z_0
$\mathbf{x}_B$	$\mathbf{0}$	I	$\overline{N}$	$\mathbf{0}$	$\overline{\mathbf{b}}$
x_{n+1}	0	$\mathbf{0}$	$\mathbf{1}$	1	M

设$\zeta_k = \max\{\zeta_i\}$,以$\overline{a}_{m+1,k}$为转轴旋转得到新的单纯形表。因为$\boldsymbol{\zeta} \le 0$，则新单纯形表对应一个正则解。于是可以开始用对偶单纯形法或原始－对偶单纯形法开始求解。

计算结果有两种可能：

1。扩充问题无解 $\Rightarrow$ 原问题无解。

2。扩充问题有最优解$\overline{\mathbf{x}} = \overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2 M$，最优值为$z_0 = z_1 + z_2 M$。这时原问题一定有可行解$\hat{\mathbf{x}}$,且$z_1 + z_2 M \le \mathbf{c}\hat{\mathbf{x}}$。由于$M$充分大，故一定有$z_2 \le 0$。（否则$\mathbf{c}\hat{\mathbf{x}}$会变的可以任意充分大了）

(i) $z_2 < 0$。当$M \to \infty, z_0 \to -\infty$。问题无界

(ii) z_2=0。$z_0 = z_1$为原问题最优值，$\overline{\mathbf{x}} \ge 0$为原问题的最优解。若$\overline{\mathbf{x}}_2 = 0$, $\overline{\mathbf{x}}$还是一个基本可行解。否则（$\overline{\mathbf{x}}_2 \ne 0$），令$M_o = \min\{M \mid \overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2 M \ge 0\}$,此时$\overline{\mathbf{x}}_0 = \overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2 M_0$也是基本可行解。

而$\{\mathbf{x} \mid \mathbf{x} = \overline{\mathbf{x}}_1 + \overline{\mathbf{x}}_2 M, M \ge M_0\}$则表示可行区域的一个半直线界面，其上任意一点均是最优解。

二、非线性最优化

引言

- 最优化的问题的一般形式为

$$\text{Min } f(x)$$
$$\text{s.t. } x \in X$$

f(x)为目标函数，$X \subset E^n$为可行域。
如$X = E^n$，则以上最优化问题为无约束最优化问题。
约束最优化问题通常写为

$$\text{Min } f(x)$$
$$\text{s.t. } c_i(x)=0,\ i \in E,$$
$$c_i(x) \ge 0,\ i \in I,$$

其中E,I分别为等式约束的指标集和不等式约束的指标集，$c_i(x)$是约束函数。

1. 无约束非线性最优化

1.1 无约束问题的最优条件

- min $f(x)$, $x \in R^n$

的最优性