最优化理论与方法

OPTIMIZATION THEORY AND METHODS

张晓伟

数学科学学院

Zhangxiaowei@uestc.edu.cn

http://staff.uestc.edu.cn/zhangxiaowei

VERSION: 20150901003400

序　言

资料：

[1] 傅英定，等，最优化理论与方法，国防工业出版社，2008.

[2] 陈宝林，最优化计算方法，清华大学出版社，1999.

[3] 袁亚湘等，最优化理论与方法，科学出版社，2000.

[4] *Jorge Nocedal*，*Stephen J. Wright*，*Numerical Opti–mization*，*Springer*，2006.

目　录

算　法

第一章　最优化问题与数学基础

§ 1.1 最优化问题

所谓最优化问题，用数学语言来说，就是求一个一元函数或多元函数的极值。下面通过具体例子来看看什么是最优化问题。

1.1.1 发展史

(1) 萌芽期：Lanchester战斗方程（1914）、排队论（1917，Erlang公式）、LP模型（1939，康托罗洛维

奇，1960，Nobel Prize）、单纯形法（1947，Dantzig）、对策论（1944，Von Neumann，Morgenstern），⋯

(2) 成长期：20世纪30年代末（二战），运筹学（Operational Research）或者最优化（Optimization）作为一个名词出现（诞生），主要应用于军事作战、防御等方面。

(3) 发展期：20世纪50年代至今，相继应用到工业、农业、经济、社会问题等领域，并形成许多分支和社团：IFORS（1959）、EUOR（1975）、APORS（1985），⋯

中国

(1)《史记·高祖本纪》：“运筹策帷幄之中，决胜于千里之外”

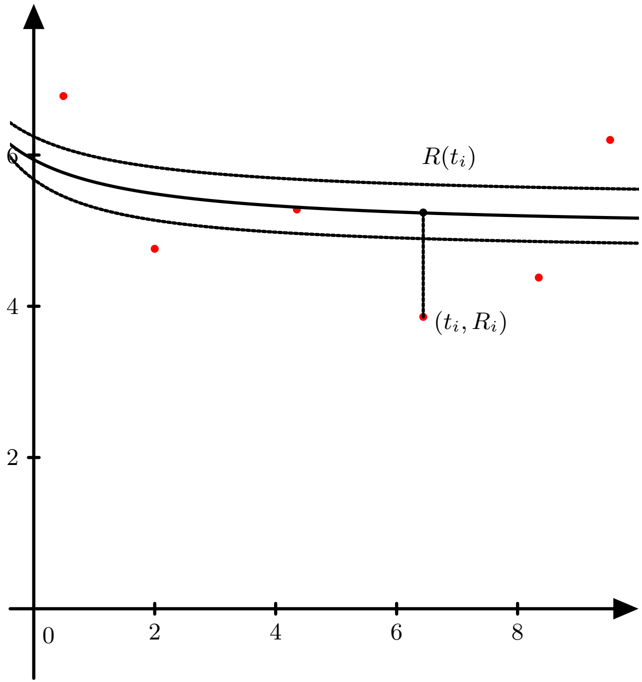
(2) 战国田忌赛马、宋朝丁渭修皇宫。

(2) 20世纪50年代中后期，钱学森、华罗庚、许国志等引入。1965年以后，华罗庚的“优选法”，“统筹法”。

1.1.2 一些例子

例 1.1 已知热敏电阻的阻值R与温度t的函数关系为$R(t) = x_1 \cdot e^{\frac{x_2}{t+x_3}}$，这里$x_1$，$x_2$，$x_3$ 为待定参数。通过实验测得在温度为t_i时，阻值为R_i，从而得到一组数据(t_1, R_1)，(t_2, R_2)，$\cdots$，(t_m, R_m)，问怎样根据这一组测量

数据来确定参数x_1，x_2，x_3?


$R(t_i)$
(t_i, R_i)
0
2
4
6
8
2
4
6

解 函数关系R在几何上对应一条平面曲线。用所有测量点沿铅直方向到曲线距离的平方和来描述这种偏差，则此问题的数学模型为

$$\min \sum_{i=1}^{m} \left[R_i - x_1 \exp\left(\frac{x_2}{t_i + x_3} \right) \right]^2$$

∎

例 1.2 (运输问题) 已知某煤炭集团公司有m个产地A_1，A_2，$\cdots$，A_m，其产量分别为a_1，a_2，$\cdots$，a_m（吨）。有n个销地B_1，B_2，$\cdots$，B_n，其

销售量分别为b_1，b_2，$\cdots$，b_n（吨）。假设产销平衡，即

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j.$$

由A_i到B_j的运费为c_{ij}（元/吨），（$i = 1,2,\cdots,m; j = 1,2,\cdots,n$）。问在保障供给的条件下，由每个产地到每个销地的运输量为多少吨时，总运费最少？

解 设由到的运输量为吨，则有数学模型

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$s.t.$$

$$\sum_{j=1}^{n} x_{ij} = a_i, \quad i = 1, 2, \cdots, m$$

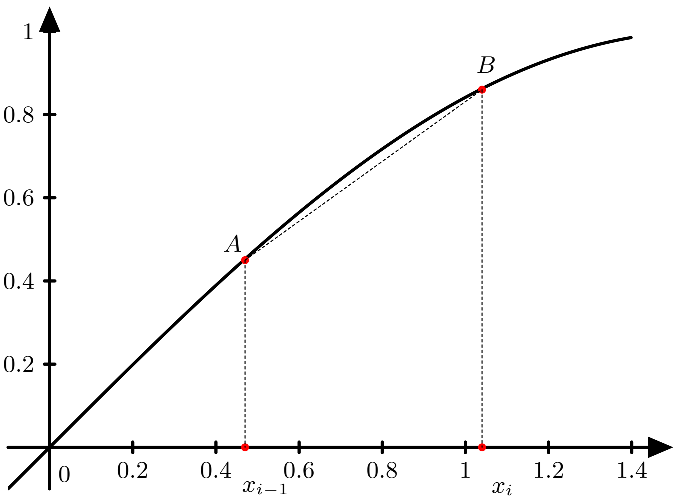
$$\sum_{i=1}^{m} x_{ij} = b_j, \quad j = 1, 2, \cdots, n$$

$$x_{ij} \geq 0$$

∎

例 1.3 (信号发生仪中正弦波形逼近的优化设计) 一个实际

的电路设计问题，要求用折线近似的代替正弦曲线，并要求在$\left(0, \frac{\pi}{2}\right)$内确定6个点，使得将$(0,0)$，$(x_1, \sin x_1)$，$\cdots$，$(x_6, \sin x_6)$，$(\pi/2, 1)$等点连接所得折线代替$\left[0, \frac{\pi}{2}\right]$上的正弦曲线时失真度最小。


1
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
A
B
x_{i-1}
x_i

解 数学上就是，使该折线与正弦曲线之间所围成的平面图形面积最小。

正弦曲线、$x = \frac{\pi}{2}$、x轴所围成的面积为

$$\int_0^{\frac{\pi}{2}} \sin x dx = 1.$$

折线、$x = \frac{\pi}{2}$、x轴所围成的面积（梯形）为

$$\frac{1}{2} \sum_{i=1}^{7} (x_i - x_{i-1})(\sin x_i + \sin x_{i-1}).$$

其中，$x_0 = 0$，$x_7 = \frac{\pi}{2}$。问题数学模型为：

$$\min f(\boldsymbol{X}) = 1 - \frac{1}{2}\sum_{i=1}^{7}(x_i - x_{i-1})(\sin x_i + \sin x_{i-1})$$

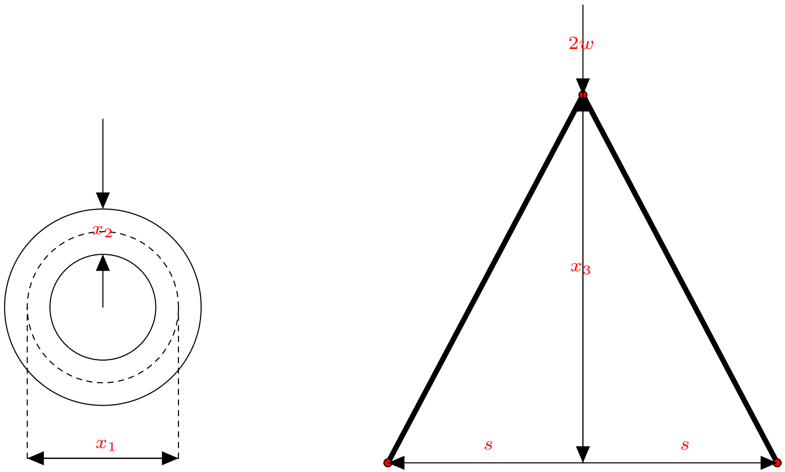
$s.t.$

$$0 = x_0 < x_1 < \cdots < x_7 = \frac{\pi}{2} \quad \text{或}$$

$$g_i(\boldsymbol{X}) = x_i - x_{i-1} > 0, x_i \in \Big(0, \frac{\pi}{2}\Big), i = 1, \cdots, 6$$

∎

例 1.4 已知由两根钢管组成的对称桁架的跨度为$2s$，钢管的平均直径为x_1，厚度为x_2，桁架的高度为x_3，现在要求桁架能够承受$2w$ 的负荷，问如何设计桁架，使得其重量最小？



解 因为钢管的截面积为$\pi x_1 x_2$，长度为$\sqrt{s^2+x_3^2}$，设钢管的

密度为ρ，则钢管的重量为

$$\rho\pi x_1 x_2 \sqrt{s^2 + x_3^2}.$$

问题为求上式的最小值，但必须满足下面几个条件：

(1)空间有限，桁架的高度不能超过h，即

$$x_3 \leq h.$$

(2)钢管的压应力不能超过临界应力（弯曲应力）σ，即

$$w\sqrt{s^2 + x_3^2} \leq \sigma\pi x_1 x_2 x_3.$$

综上所述，数学模型为：

$$\min x_1 x_2 \sqrt{s^2 + x_3^2}$$

$$s.t.$$

$$x_3 \leq h,$$

$$w\sqrt{s^2 + x_3^2} \leq \sigma \pi x_1 x_2 x_3,$$

$$x_1, x_2, x_3 \geq 0.$$

∎

1.1.3 数学模型

（1）一般形式

$$
\begin{aligned}
&\min f(x_1, x_2, \cdots, x_n) \\
s.t.& \\
&g_i(x_1, x_2, \cdots, x_n) \geq 0, \quad i = 1, 2, \cdots, m \\
&h_j(x_1, x_2, \cdots, x_n) = 0, \quad j = 1, 2, \cdots, k
\end{aligned}
\tag{1-1-1}
$$

（2）向量形式

$$\begin{aligned} &\min f(\boldsymbol{X}) \\ s.t.\quad & \\ &\boldsymbol{G}(\boldsymbol{X}) \geq \mathbf{0} \\ &\boldsymbol{H}(\boldsymbol{X}) = \mathbf{0} \end{aligned} \tag{1-1-2}$$

这里，

$\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^T$

$\boldsymbol{G}(\boldsymbol{X}) = (g_1(\boldsymbol{X}), g_2(\boldsymbol{X}), \cdots, g_m(\boldsymbol{X}))^T$

$\boldsymbol{H}(\boldsymbol{X}) = (h_1(\boldsymbol{X}), h_2(\boldsymbol{X}), \cdots, h_k(\boldsymbol{X}))^T$

$\mathbf{0}$表示零向量，以后在不混淆的前提下，为了方便，用0表

示零向量。

1.1.4 问题分类

（1）静态/动态

（2）约束/无约束

（3）线性/非线性

§ 1.2 梯度与*Hesse*矩阵

1.2.1 等值线

定义 1.5 在高维空间$(n \geq 3)$中，使目标函数值取同一常数

的点集$\{\boldsymbol{X}|f(\boldsymbol{X})=c, c$为一常数$\}$称为$f(\boldsymbol{X})$的等值线（或等值面）。

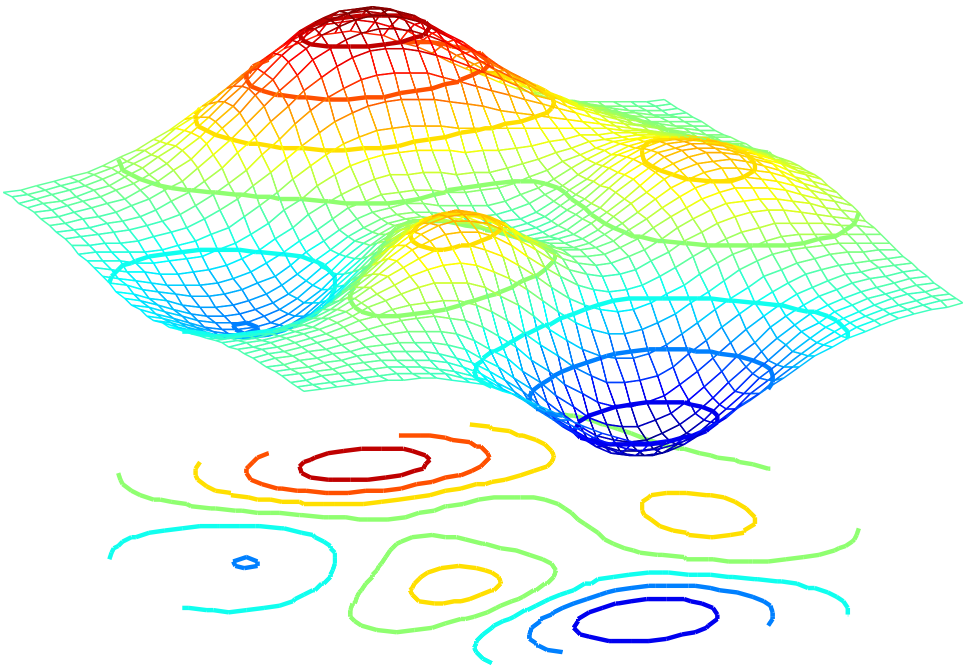
在通常情况下，目标函数是连续的单值函数，则其等值线具有以下性质：

（1）不同的等值线不相交；

（2）除极点所在的等值线外，等值线不会中断；

（3）等值线稠密的地方，目标函数值变化较快，稀疏的地方，变化较慢；

（4）在极值点附近，等值线近似地为同心椭圆族。



1.2.2 n元函数的可微性与梯度

（1）可微与梯度的定义

定义 1.6 设$f : D \subset \boldsymbol{R}^n \to \boldsymbol{R}$，且$\boldsymbol{X}^0 \in D$，若存在$n$维向量$\boldsymbol{L}$，对任意$n$维向量$\boldsymbol{P}$，都有

$$\lim_{\|\boldsymbol{P}\|\to 0} \frac{f(\boldsymbol{X}^0 + \boldsymbol{P}) - f(\boldsymbol{X}^0) - \boldsymbol{L}^T\boldsymbol{P}}{\|\boldsymbol{P}\|} = 0. \tag{1-2-3}$$

则称$f(\boldsymbol{X})$在$\boldsymbol{X}^0$可微。

其中，$\|\boldsymbol{P}\| = \sqrt{p_1^2 + p_2^2 + \cdots + p_n^2}$是向量$\boldsymbol{P}$的模。

若令

$$\frac{f(\boldsymbol{X}^0+\boldsymbol{P})-f(\boldsymbol{X}^0)-\boldsymbol{L}^T\boldsymbol{P}}{\|\boldsymbol{P}\|}=\alpha,$$

则$\lim_{\|\boldsymbol{P}\|\to 0}\alpha=0$。于是上式与下式等价

$$f(\boldsymbol{X}^0+\boldsymbol{P})-f(\boldsymbol{X}^0)=\boldsymbol{L}^T\boldsymbol{P}+\alpha\|\boldsymbol{P}\|=\boldsymbol{L}^T\boldsymbol{P}+o(\|\boldsymbol{P}\|). \quad (1\text{-}2\text{-}4)$$

下面定理给出的表达式。

定理 1.7 若$f(\boldsymbol{X})$在$\boldsymbol{X}^0$处可微，则$f(\boldsymbol{X})$在$\boldsymbol{X}^0$关于各变量的

一阶偏导数存在，且

$$\boldsymbol{L}=\Big(\frac{\partial f(\boldsymbol{X}^0)}{\partial x_1},\frac{\partial f(\boldsymbol{X}^0)}{\partial x_2},\cdots,\frac{\partial f(\boldsymbol{X}^0)}{\partial x_n}\Big)^T.$$

证明 令$e^i=(0,\cdots,0,1_{[i]},0,\cdots,0)$，依次取$\boldsymbol{P}=p_ie^i$，$i=1,\cdots,n$，则由式（1-2-4）得证！ ∎

定义 1.8 (梯度) 以$f(\boldsymbol{X})$的n个偏导数为分量的向量$\nabla f(\boldsymbol{X})$称为$f(\boldsymbol{X})$的梯度。显然$\nabla f(\boldsymbol{X})=\boldsymbol{L}$。

若$\boldsymbol{P}=\boldsymbol{X}-\boldsymbol{X}^0$，则式（1-2-4）可记为：

$$f(\boldsymbol{X})-f(\boldsymbol{X}^0)=\nabla f(\boldsymbol{X}^0)^T(\boldsymbol{X}-\boldsymbol{X}^0)+o(\|\boldsymbol{X}-\boldsymbol{X}^0\|).$$

（2）梯度的性质

(I)若$\nabla f(\boldsymbol{X}^0) \neq 0$，则$\nabla f(\boldsymbol{X}^0)$与过$\boldsymbol{X}^0$点的等值线垂直。

(II)沿梯度方向函数具有最大的变化率。

（3）方向导数

定义 1.9 (方向导数) 设$\boldsymbol{P} \in \boldsymbol{R}^n$，$\|\boldsymbol{P}\| = 1$，可微函

数$f(\boldsymbol{X})$在$\boldsymbol{X}$点沿方向$\boldsymbol{P}$的方向导数定义为：

$$
\begin{aligned}
\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{P}} &= \lim_{\alpha \to 0^+} \frac{f(\boldsymbol{X}+\alpha \boldsymbol{P}) - f(\boldsymbol{X})}{\alpha} \\
&= \lim_{\alpha \to 0^+} \frac{\nabla f(\boldsymbol{X})^T(\alpha \boldsymbol{P}) + o(\|\alpha \boldsymbol{P}\|)}{\alpha} \\
&= \lim_{\alpha \to 0^+} \nabla f(\boldsymbol{X})^T(\boldsymbol{P}) + \frac{o(\|\alpha \boldsymbol{P}\|)}{\|\alpha \boldsymbol{P}\|}\|\boldsymbol{P}\| \\
&= \nabla f(\boldsymbol{X})^T \boldsymbol{P} \\
&= \|\nabla f(\boldsymbol{X})\| \cos(\nabla f(\boldsymbol{X}), \boldsymbol{P})
\end{aligned} \tag{1-2-5}
$$

其中，$(\nabla f(\boldsymbol{X}), \boldsymbol{P})$表示向量$\nabla f(\boldsymbol{X})$与$\boldsymbol{P}$的夹角。

（I） 若$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{P}} = \nabla f(\boldsymbol{X})^T \boldsymbol{P} > 0$，则$\boldsymbol{P}$的方向是$f(\boldsymbol{X})$在$\boldsymbol{X}$处的上升方向。

（II） 若$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{P}} = \nabla f(\boldsymbol{X})^T\boldsymbol{P} < 0$，则$\boldsymbol{P}$的方向是$f(\boldsymbol{X})$在$\boldsymbol{X}$处的下降方向。

（III）若$\nabla f(\boldsymbol{X}) = 0$，则对任何方向$\boldsymbol{P}$，$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{P}} = 0$。

（IV） 若$\nabla f(\boldsymbol{X}) \neq 0$，则当$\boldsymbol{P} = \frac{1}{\|\nabla f(\boldsymbol{X})\|}\nabla f(\boldsymbol{X})$时，$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{P}}$取得最大值，当$\boldsymbol{P} = -\frac{1}{\|\nabla f(\boldsymbol{X})\|}\nabla f(\boldsymbol{X})$时，$\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{P}}$取得最小值。

例 1.10 求函数$f(\boldsymbol{X}) = x_1^2 + x_2^2 + 1$在$\boldsymbol{X}^0 = (0,3)^T$处的最速下降方向，并求沿此方向移动一个单位长度后，所得新点的函数值。

解 $-\nabla f(\boldsymbol{X}^0) = (0,-6)^T$，$f(\boldsymbol{X}^1) = 5$。 ∎

（4）几种特殊函数的梯度

（I）对任意常数c，$\nabla c = 0$。

（II）$\nabla(\boldsymbol{b}^T\boldsymbol{X}) = \boldsymbol{b}$，$\boldsymbol{b} = (b_1, b_2, \cdots, b_n)^T \in \boldsymbol{R}^n$。

（III）$\nabla(\boldsymbol{X}^T\boldsymbol{X}) = 2\boldsymbol{X}$。

（IV）$\nabla(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}) = 2\boldsymbol{A}\boldsymbol{X}$，这里$\boldsymbol{A} = \boldsymbol{A}^T$。

（V）$\nabla(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{A}^T\boldsymbol{X}$。

（5）$Hesse$矩阵

定义 1.11 (向量值函数) $g(\boldsymbol{X})$是一个向量值函数，若$g : D \subset \boldsymbol{R}^n \to \boldsymbol{R}^m$，即

$$g(\boldsymbol{X}) = (g_1(\boldsymbol{X}), g_2(\boldsymbol{X}), \cdots, g_m(\boldsymbol{X}))^T.$$

定义 1.12 (可微) 设$g(\boldsymbol{X}) : D \subset \boldsymbol{R}^n \to \boldsymbol{R}^m$，$\boldsymbol{X}^0 \in D$，若$g(\boldsymbol{X})$的所有分量$(g_1(\boldsymbol{X})$，$g_2(\boldsymbol{X})$，$\cdots$，$g_m(\boldsymbol{X})$在都可微，则称$g(\boldsymbol{X})$在$\boldsymbol{X}^0$可微。

这时称

$$\nabla g(\boldsymbol{X}^0) = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{X}^0)}{\partial x_1} & \frac{\partial g_1(\boldsymbol{X}^0)}{\partial x_2} & \cdots & \frac{\partial g_1(\boldsymbol{X}^0)}{\partial x_n} \\ \frac{\partial g_2(\boldsymbol{X}^0)}{\partial x_1} & \frac{\partial g_2(\boldsymbol{X}^0)}{\partial x_2} & \cdots & \frac{\partial g_2(\boldsymbol{X}^0)}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_m(\boldsymbol{X}^0)}{\partial x_1} & \frac{\partial g_m(\boldsymbol{X}^0)}{\partial x_2} & \cdots & \frac{\partial g_m(\boldsymbol{X}^0)}{\partial x_n} \end{bmatrix} \tag{1-2-6}$$

为向量值函数在处的导数或$Jacobi$矩阵。

设$f : D \subset \boldsymbol{R}^n \to \boldsymbol{R}$，且$f(\boldsymbol{X})$具有二阶连续偏导数，又设$m = n$，则

$$\nabla^2 f(\boldsymbol{X}) = \nabla g(\boldsymbol{X}) = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{X})}{\partial^2 x_1} & \frac{\partial^2 f(\boldsymbol{X})}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{X})}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(\boldsymbol{X})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\boldsymbol{X})}{\partial^2 x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{X})}{\partial x_n \partial x_2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f(\boldsymbol{X})}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\boldsymbol{X})}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(\boldsymbol{X})}{\partial^2 x_n} \end{bmatrix}. \quad (1\text{-}2\text{-}7)$$

显然其对称。

定义 1.13 $\nabla^2 f(\boldsymbol{X})$称为$f(\boldsymbol{X})$关于$\boldsymbol{X}$的二阶导数，矩阵（1-2-7）称为$f(\boldsymbol{X})$的$Hesse$矩阵。

对于向量值函数，有以下几个常用公式。

（I）$\nabla \boldsymbol{C} = 0$。$\boldsymbol{C} = (c_1, c_2, \cdots, c_n)^T$，$c_i$为常数。

（II）$\nabla \boldsymbol{X} = \boldsymbol{I}$。$\boldsymbol{I}$是$n$阶单位方阵。

（III）$\nabla(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}$。$A$为$n$阶方阵。

（IV）设$\phi(t) = f(\boldsymbol{X}^0 + t\boldsymbol{P})$，则

$$\phi'(t) = \nabla f(\boldsymbol{X}^0 + t\boldsymbol{P})^T \boldsymbol{P}, \quad \phi''(t) = \boldsymbol{P}^T \nabla^2 f(\boldsymbol{X}^0 + t\boldsymbol{P})\boldsymbol{P}$$

证明 因为

$$\phi(t) = f(x_1^0 + tp_1, \cdots, x_i^0 + tp_i, \cdots, x_n^0 + tp_n) \tag{1-2-8}$$

所以

$$\begin{aligned}\phi'(t) &= \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{X}^0 + t\boldsymbol{P})}{\partial(x_i^0 + tp_i)} p_i \\ &= \sum_{i=1}^{n} \frac{\partial f(\boldsymbol{X})}{\partial x_i} p_i \\ &= \nabla f(\boldsymbol{X}^0 + t\boldsymbol{P})^T \boldsymbol{P}\end{aligned} \tag{1-2-9}$$

进而

$$
\begin{aligned}
\phi''(t) &= \sum_{i=1}^{n} \frac{d}{dt}\Big(\frac{\partial f(\boldsymbol{X}^0 + t\boldsymbol{P})}{\partial x_i}\Big)p_i \\
&= \sum_{i=1}^{n}\Big(\sum_{j=1}^{n}\frac{\partial^2 f(\boldsymbol{X}^0 + t\boldsymbol{P})}{\partial x_j \partial x_i}p_j\Big)p_i \\
&= \boldsymbol{P}^T\nabla^2 f(\boldsymbol{X}^0 + t\boldsymbol{P})\boldsymbol{P}
\end{aligned}
\tag{1-2-10}
$$

∎

§ 1.3 多元函数的*Taylor*展式

定理 1.14 设$f : D \subset \boldsymbol{R}^n \to \boldsymbol{R}$，若$f(\boldsymbol{X})$在$\boldsymbol{X}^0$的某个领域$N(\boldsymbol{X}^0, \delta)$内二阶连续可微，则对任意的$\boldsymbol{X} \in N(\boldsymbol{X}^0, \delta)$在$\boldsymbol{X}^0$

处有$Taylor$展式

$$
\begin{aligned}
f(\boldsymbol{X}) = & f(\boldsymbol{X}^0) \\
& + \nabla f(\boldsymbol{X}^0)^T(\boldsymbol{X} - \boldsymbol{X}^0) \\
& + \frac{1}{2}(\boldsymbol{X} - \boldsymbol{X}^0)^T \nabla^2 f(\boldsymbol{X})(\boldsymbol{X} - \boldsymbol{X}^0) \\
& + o(\|\boldsymbol{X} - \boldsymbol{X}^0\|^2)
\end{aligned} \tag{1-3-11}
$$

证明 设$\phi(t) = f(\boldsymbol{X} + t\boldsymbol{P})$，

则

$$
\begin{aligned}
f(\boldsymbol{X}^0+\boldsymbol{P}) &= \phi(1) \\
&= \phi(0)+\phi'(0)+\frac{1}{2}\phi''(\theta) \\
&= f(\boldsymbol{X}^0)+\nabla f(\boldsymbol{X}^0)^T\boldsymbol{P}+\frac{1}{2}\boldsymbol{P}^T\nabla^2 f(\boldsymbol{X}^0+\theta\boldsymbol{P})\boldsymbol{P},
\end{aligned}
$$

又$\nabla^2 f(\boldsymbol{X})$连续，

所以

$$
\frac{\partial^2 f(\boldsymbol{X}^0+\theta\boldsymbol{P})}{\partial x_i\partial x_j}=\frac{\partial^2 f(\boldsymbol{X}^0)}{\partial x_i\partial x_j}+\delta_{ij}. \quad (\lim_{\|\boldsymbol{P}\|\to 0}\delta_{ij}=0)
$$

于是

$$\begin{aligned}\boldsymbol{P}^T\nabla^2 f(\boldsymbol{X}^0+\theta\boldsymbol{P})\boldsymbol{P} &= \boldsymbol{P}^T\nabla^2 f(\boldsymbol{X}^0)\boldsymbol{P}+\boldsymbol{P}^T[\delta_{ij}]_{n\times n}\boldsymbol{P}\\ &= \boldsymbol{P}^T\nabla^2 f(\boldsymbol{X}^0)\boldsymbol{P}+\sum_{i=1}^n\sum_{j=1}^n \delta_{ij}p_ip_j.\end{aligned}$$

因此

$$f(\boldsymbol{X}^0+\boldsymbol{P}) = f(\boldsymbol{X}^0)+\nabla f(\boldsymbol{X}^0)^T\boldsymbol{P}+\frac{1}{2}\boldsymbol{P}^T\nabla^2 f(\boldsymbol{X}^0)\boldsymbol{P}+o(\|\boldsymbol{P}\|^2).$$

令$\boldsymbol{P}=\boldsymbol{X}-\boldsymbol{X}^0$。证毕！ ▮

§ 1.4 极小点及其判定条件

定义 1.15 (内点、边界点与极限点) 设$D \in \boldsymbol{R}^n$，$\boldsymbol{X}^0 \in$

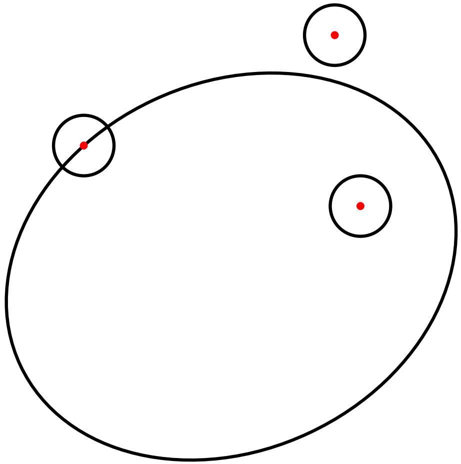
$\boldsymbol{R}^n$。

若存在$\boldsymbol{X}^0$的领域$N(\boldsymbol{X}^0,\delta)\subset D$，则称$\boldsymbol{X}^0$为$D$的内点；

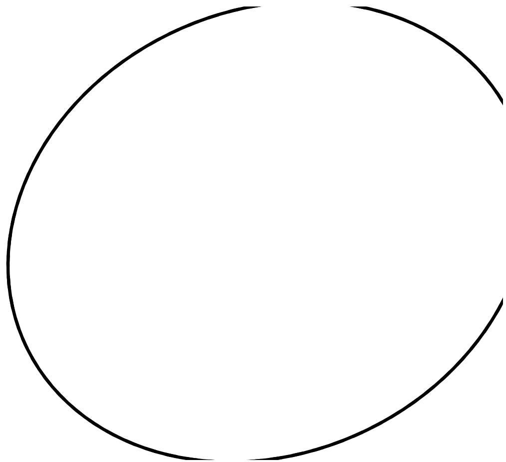
如果在$\boldsymbol{X}^0$的任意领域中，既有$\boldsymbol{X}^1\in D$，又有$\boldsymbol{X}^2\notin D$，则称$\boldsymbol{X}^0$为D的边界点；

既不是内点，又不是边界点的点，称为外点；

如果存在点列$\{\boldsymbol{X}^k\}$，（$\boldsymbol{X}^k$互异），且$\lim_{k\to+\infty}\|\boldsymbol{X}^k-\boldsymbol{X}^0\|=0$，则称$\boldsymbol{X}^0$为$D$的极限点（聚点）。

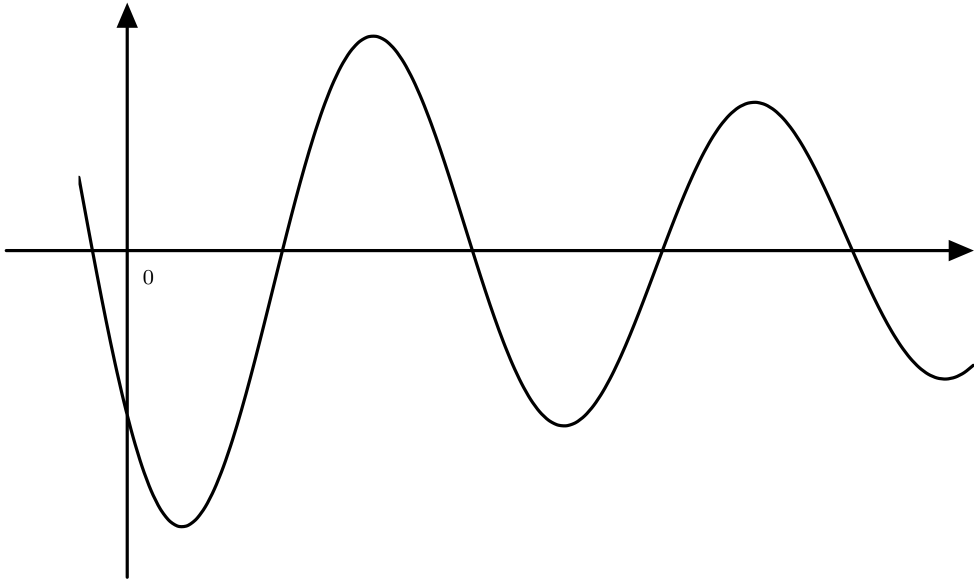


定义 1.16 (开集与闭集) 设$D \subset \boldsymbol{R}^n$，如果$D$的每一个点都是$D$的内点，则称$D$为开集；如果$D$的每一个极限点都属于$D$，则称$D$为闭集。



定义 1.17 (极小点与最优解) 设$f : D \subset \boldsymbol{R}^n \to \boldsymbol{R}$，若

存在$\boldsymbol{X}^* \in D$及实数$\delta > 0$，使得$\forall \boldsymbol{X} \in N^\circ(\boldsymbol{X}^*, \delta) \bigcap D$都有$f(\boldsymbol{X}^*) \leq f(\boldsymbol{X})$，则称$\boldsymbol{X}^*$为$f(\boldsymbol{X})$的局部极小点；若$f(\boldsymbol{X}^*) < f(\boldsymbol{X})$，则称$\boldsymbol{X}^*$为$f(\boldsymbol{X})$的严格局部极小点；若对$\forall \boldsymbol{X} \in D$，都有$f(\boldsymbol{X}^*) \leq f(\boldsymbol{X})$，则称称$\boldsymbol{X}^*$为$f(\boldsymbol{X})$的全局极小点；若$f(\boldsymbol{X}^*) < f(\boldsymbol{X})$，则称$\boldsymbol{X}^*$为$f(\boldsymbol{X})$的全局严格极小点。



1.4.1 局部极小点的判定条件

定理 1.18 设$f: D \subset \boldsymbol{R}^n \to \boldsymbol{R}$，$f$具有连续的一阶偏导数，

若$\boldsymbol{X}^*$是$f(\boldsymbol{X})$的局部极小点且是D的内点，则$\nabla f(\boldsymbol{X}^*)=0$。

定义 1.19 (驻点) 设$f: D\subset \boldsymbol{R}^n\rightarrow\boldsymbol{R}$，$\boldsymbol{X}^*$是$D$的内点，若$\nabla f(\boldsymbol{X}^*)=0$，则称$\boldsymbol{X}^*$为$f(\boldsymbol{X})$的驻点。

定理 1.20 设$f: D\subset \boldsymbol{R}^n\rightarrow\boldsymbol{R}$，具有连续的二阶偏导数，$\boldsymbol{X}^*$为$f(\boldsymbol{X})$的驻点，且$\nabla^2 f(\boldsymbol{X}^*)$是正定矩阵，则$\boldsymbol{X}^*$是$f(\boldsymbol{X})$的严格局部极小点。

证明

$$f(\boldsymbol{X})-f(\boldsymbol{X}^*)=\frac{1}{2}(\boldsymbol{X}-\boldsymbol{X}^*)^T\nabla^2 f(\boldsymbol{X}^*)(\boldsymbol{X}-\boldsymbol{X}^*)+o(\|(\boldsymbol{X}-\boldsymbol{X}^*)\|^2)$$

$$\lambda_1\|\boldsymbol{X}-\boldsymbol{X}^*\|^2\leq(\boldsymbol{X}-\boldsymbol{X}^*)^T\nabla^2 f(\boldsymbol{X}^*)(\boldsymbol{X}-\boldsymbol{X}^*)\leq\lambda_n\|\boldsymbol{X}-\boldsymbol{X}^*\|^2$$

§ 1.5 凸集、凸函数与凸规划

1.5.1 凸集

几何直观上，若集合D中的任意两点的连线仍在D中，则称D为凸集。

定义 1.21 (凸集) 设$D \subset \boldsymbol{R}^n$，若对所有的$\boldsymbol{X}^1$，$\boldsymbol{X}^2 \in D$，以及$\alpha \in [0,1]$，都有$\alpha \boldsymbol{X}^1 + (1-\alpha)\boldsymbol{X}^2 \in D$，则称$D$为凸集。

例 1.22 平面$S = \{(x_1, x_2, x_3) | x_1 + 2x_2 - x_3 = 4\}$是凸集。

例 1.23 $S = \{\boldsymbol{X} | \boldsymbol{A}\boldsymbol{X} \leq \boldsymbol{b}, \boldsymbol{X} \in \boldsymbol{R}^n, \boldsymbol{A} \in \boldsymbol{R}^{m\times n}\}$是凸集。

例 1.24 若A，B是$\boldsymbol{R}^n$中的凸集，则$A\bigcap B$，$A+B$，$A-B$也是凸集，但$A\bigcup B$一般不是。

定义 1.25 (凸组合) 设$\boldsymbol{X}^1$，$\boldsymbol{X}^2$，$\cdots$，$\boldsymbol{X}^m \in \boldsymbol{R}^n$，$\alpha_1$，$\alpha_2$，$\cdots$，$\alpha_m$，是一组非负实数，且$\sum_{i=1}^m \alpha_i = 1$，则

$$\alpha_1\boldsymbol{X}^1+\alpha_2\boldsymbol{X}^2+\cdots+\alpha_m\boldsymbol{X}^m=\sum_{i=1}^m \alpha_i\boldsymbol{X}^i$$

称为的一个凸组合。

定理 1.26 $D\subset \boldsymbol{R}^n$为凸集的充要条件是$D$中任意有限个点的凸组合仍在$D$中。

证明 数学归纳法。 ▮

1.5.2 凸函数

从几何上看，曲线上任意两点的连线在相应弧段的上方，即弦在弧之上。

定义 1.27 (凸函数) 设$f: D \subset \boldsymbol{R}^n \to \boldsymbol{R}$，$D$是凸集，若对所有的$\boldsymbol{X}^1$，$\boldsymbol{X}^2 \in D$，以及$\alpha \in (0,1)$，都有

$$f[\alpha \boldsymbol{X}^1 + (1-\alpha)\boldsymbol{X}^2] \leq \alpha f(\boldsymbol{X}^1) + (1-\alpha) f(\boldsymbol{X}^2),$$

则称$f(\boldsymbol{X})$为D上的凸函数。

例 1.28 $f(x) = 3x + 4$，$g(x) = |x|$

例 1.29　$f(\boldsymbol{X}) = 2x_1^2 + x_2^2 - 2x_1x_2$

例 1.30　$f(\boldsymbol{X}) = c_1x_1 + c_2x_2 + \cdots + c_nx_n$

例 1.31　$f(\boldsymbol{X}) = \boldsymbol{X}^T\boldsymbol{(}AX)$，$\boldsymbol{A} = \boldsymbol{A}^T$，$\boldsymbol{X} \in \boldsymbol{R}^n$

（I）若$\boldsymbol{A}$为半正定矩阵，则$f(\boldsymbol{X})$为凸函数。

（II）若$\boldsymbol{A}$为正定矩阵，则$f(\boldsymbol{X})$为严格凸函数。

定理 1.32　设$f : D \subset \boldsymbol{R}^n \to \boldsymbol{R}$，$D$是凸集，则$f(\boldsymbol{X})$是$D$上凸函数的充要条件是对任意正整数$m(m \geq 2)$及任意不全相同的$\boldsymbol{X}^1$，$\boldsymbol{X}^2$，$\cdots$，$\boldsymbol{X}^m \in D$，如果$\alpha_i \geq 0(i =$

$1, 2, \cdots, m$)，$\sum_{i=1}^{m} \alpha_i = 1$，则有

$$f(\sum_{i=1}^{m} \alpha_i \boldsymbol{X}^i) \leq \sum_{i=1}^{m} \alpha_i f(\boldsymbol{X}^i).$$

定理 1.33 设$f : D \subset \boldsymbol{R}^n \to \boldsymbol{R}$，$D$是非空凸集，$f(\boldsymbol{X})$在$D$上可微，则$f(\boldsymbol{X})$在$D$上是凸函数的充要条件是对任意的$\boldsymbol{X}^1$，$\boldsymbol{X}^2 \in D$，都有

$$f(\boldsymbol{X}^2) \geq f(\boldsymbol{X}^1) + \nabla f(\boldsymbol{X}^1)^T(\boldsymbol{X}^2 - \boldsymbol{X}^1).$$

证明 必要性：由可微得

$$f[\boldsymbol{X}^1 + \alpha(\boldsymbol{X}^2 - \boldsymbol{X}^1)] =$$

$$f(\boldsymbol{X}^1)+\nabla f(\boldsymbol{X}^1)^T\alpha(\boldsymbol{X}^2-\boldsymbol{X}^1)+o(\alpha\|\boldsymbol{X}^2-\boldsymbol{X}^1\|).$$

所以

$$\lim_{\alpha\to 0^+}\frac{f[\boldsymbol{X}^1+\alpha(\boldsymbol{X}^2-\boldsymbol{X}^1)]-f(\boldsymbol{X}^1)}{\alpha}=\nabla f(\boldsymbol{X}^1)^T\alpha(\boldsymbol{X}^2-\boldsymbol{X}^1).$$

又由凸函数的定义可得

$$\frac{f[\boldsymbol{X}^1+\alpha(\boldsymbol{X}^2-\boldsymbol{X}^1)]-f(\boldsymbol{X}^1)}{\alpha}\leq f(\boldsymbol{X}^2)-f(\boldsymbol{X}^1).$$

充分性：令 $\boldsymbol{Y}=\alpha\boldsymbol{X}^1+(1-\alpha)\boldsymbol{X}^2\in D$(凸集)，所以

$$f(\boldsymbol{X}^1)\geq f(\boldsymbol{Y})+\nabla f(\boldsymbol{Y})^T(\boldsymbol{X}^1-\boldsymbol{Y}),$$

$$f(\boldsymbol{X}^2) \geq f(\boldsymbol{Y}) + \nabla f(\boldsymbol{Y})^T(\boldsymbol{X}^2 - \boldsymbol{Y}),$$

分别乘以α，$1-\alpha$后相加即可。 ∎

定理 1.34 设$f : D \subset \boldsymbol{R}^n \to \boldsymbol{R}$，$D$是非空凸集，$f(\boldsymbol{X})$在$D$上具有连续的二阶偏导数，则$f(\boldsymbol{X})$在$D$上是凸函数的充要条件是$\nabla^2 f(\boldsymbol{X})$是半正定矩阵。

证明 充分性：$Taylor$公式展到二阶。由定理1.33可得。

必要性：对$\forall \boldsymbol{X} \in D$，$\boldsymbol{Y} \in D$，$\exists\lambda$使得$\boldsymbol{X} + \lambda\boldsymbol{Y} \in D$，因此由定理1.33和$Taylor$公式得：

$$\frac{1}{2}\boldsymbol{Y}^T\nabla^2 f(\boldsymbol{X})\boldsymbol{Y} + \frac{o(\lambda^2\|\boldsymbol{Y}\|^2)}{\lambda^2\|\boldsymbol{Y}\|^2}\|\boldsymbol{Y}\|^2 \geq 0.$$

■

例 1.35 $f(\boldsymbol{X}) = 3x_1^2 - 2x_1 + 2x_2^2 - x_2 + 10$是凸函数。$\left(\nabla^2 f(\boldsymbol{X}) = \begin{bmatrix} 6 & 0 \\ 0 & 4 \end{bmatrix}\right)$

1.5.3 凸规划

定义 1.36 (凸规划) 对于

$$\begin{aligned} &\min f(\boldsymbol{X}) \\ s.t. & \\ &g_i(\boldsymbol{X}) \geq 0 \quad (i = 1, 2, \cdots, m) \end{aligned} \tag{1-5-12}$$

若$f(\boldsymbol{X})$与$-g_i(\boldsymbol{X})$都是凸函数，则其为凸规划。

例 1.37 线性规划

$$\min \boldsymbol{C}^T\boldsymbol{X}$$

$$s.t.$$

$$\boldsymbol{A}\boldsymbol{X} \geq b$$

$$\boldsymbol{X} \geq 0$$

为凸规划。

定理 1.38 若设规划(1-5-12)为凸规划，则有

（1）规划(1-5-12)的可行集R为凸集；

（2）规划(1-5-12)的最优解集R^*为凸集；

（3）规划(1-5-12)的任何局部最优解为全局最优解。

证明 （1）由$-g_i(\boldsymbol{X})$的凸性易证。

（2）设$\boldsymbol{X}^1$，$\boldsymbol{X}^2 \in R^*$，由凸性知，

$$f(\boldsymbol{X}^1) \leq [\alpha \boldsymbol{X}^1 + (1-\alpha)\boldsymbol{X}^2] \leq f(\boldsymbol{X}^1).$$

所以，$\alpha \boldsymbol{X}^1 + (1-\alpha)\boldsymbol{X}^2 \in R^*$，$R^*$为凸集。

（3）若$\boldsymbol{X}$是局部最优解，而不是全局最优解，那么存在$\boldsymbol{Y} \in R$，使得

$$f[\alpha \boldsymbol{X} + (1-\alpha)\boldsymbol{Y}] \leq \alpha f(\boldsymbol{X}) + (1-\alpha) f(\boldsymbol{X}) = f(\boldsymbol{X}),$$

$$\alpha \boldsymbol{X} + (1-\alpha)\boldsymbol{Y} \in N(\boldsymbol{X}, \delta) \xrightarrow[\alpha \to 1^-]{0<\alpha<1} f(\boldsymbol{X}) \leq f(\alpha \boldsymbol{X} + (1-\alpha)\boldsymbol{Y}).$$

两式矛盾。 ∎

定理 1.39 设规划(1-5-12)为凸规划，且$f(\boldsymbol{X})$为严格凸函数，则当规划(1-5-12)的最优解集$R^* \neq \emptyset$时，规划(1-5-12)的最优解是唯一的。

证明 反证。规划(1-5-12)的最优解不唯一，则$\exists \boldsymbol{X}^1 \neq \boldsymbol{X}^2 (\in R^*)$，使得

$$\begin{aligned} f[\alpha \boldsymbol{X} + (1-\alpha)\boldsymbol{X}^2] \leq \alpha f(\boldsymbol{X}^1) + (1-\alpha) f(\boldsymbol{X}^2) &= f(\boldsymbol{X}^1) \\ &= \min_{\boldsymbol{X} \in R} f(\boldsymbol{X}). \end{aligned}$$

因此与$\boldsymbol{X}^1 \in R^*$矛盾。 ∎

定理 1.40 设凸规划(1-5-12)的目标函数$f(\boldsymbol{X})$可微，则$\boldsymbol{X}^*$为规划1-5-12的最优解的充要条件是，$\forall \boldsymbol{X} \in R$，有

$$(\boldsymbol{X} - \boldsymbol{X}^*)^T \nabla f(\boldsymbol{X}^*) \geq 0.$$

证明 充分性：

$$f(\boldsymbol{X}) \geq f(\boldsymbol{X}^*) + \nabla f(\boldsymbol{X}^*)^T (\boldsymbol{X} - \boldsymbol{X}^*) \geq f(\boldsymbol{X}^*).$$

必要性：

若存在$\boldsymbol{X}^0 \in R$，使得$(\boldsymbol{X}^0 - \boldsymbol{X}^*)^T f(\boldsymbol{X}^*) < 0$，则

$$
\begin{aligned}
& f[\boldsymbol{X}^* + \alpha(\boldsymbol{X}^0 - \boldsymbol{X}^*)] - f(\boldsymbol{X}^*) \\
& \quad = \alpha \nabla f(\boldsymbol{X}^*)^T (\boldsymbol{X} - \boldsymbol{X}^*) + o(\alpha \| \boldsymbol{X}^0 - \boldsymbol{X}^* \|) \\
& \quad = \alpha \left((\boldsymbol{X}^0 - \boldsymbol{X}^*)^T f(\boldsymbol{X}^*) + \frac{o(\alpha \| \boldsymbol{X}^0 - \boldsymbol{X}^* \|)}{\alpha} \right) \\
& \quad \leq 0.
\end{aligned}
$$

这与$\boldsymbol{X}^*$为最优解矛盾。 ▮

第二章 线性规划和单纯形方法

线性规划是最优化理论与方法中理论完整、方法成熟、应用广泛的一个分支。

§ 2.1 例子与标准形式

例 2.1 (下料问题) 某车间有长度为180cm的钢管，今要将其截成三种不同长度不同的管料，长度分别为70cm、52cm、35cm。生产任务规定，70cm的管料只需100根，而52cm、35cm的管料分别不得少于150根、120根，问应采取怎样的截法，才能完成任务，同时

使剩下的余料最少？

解 所有可能的截法见下表：

表 2.1 截法

截法	一	二	三	四	五	六	七	八	需要量
70	2	1	1	1	0	0	0	0	100
52	0	2	1	0	3	2	1	0	150
35	1	0	1	3	0	2	3	5	120
余料	5	6	23	5	24	6	23	5	

设第i种截法被采用x_i次，则数组$(x_i, x_2, \cdots, x_8)$可描述为

一个截料方案。则数学模型为

$$\min f(\boldsymbol{X}) = 5x_1 + 6x_2 + 23x_3 + 5x_4 + 24x_5 + 6x_6 + 23x_7 + 5x_8$$

$s.t.$

$$\begin{array}{llllllllll} 2x_1 & +x_2 & +x_3 & +x_4 & & & & & = 100 \\ & 2x_2 & +x_3 & & +3x_5 & +2x_6 & +x_7 & & \geq 150 \\ x_1 & & +x_3 & +3x_4 & & +2x_6 & +3x_7 & +5x_8 & \geq 120 \\ x_i \geq 0, i = 1, 2, \cdots, 8 \end{array}$$

∎

例 2.2 (资源利用问题) 设某企业有m种不同的资源（如原料、能源、资金等）用来生产n种产品。用a_{ij}表示生产一个单位第j种产品所消耗的第i种资源的数量，用c_j表示第j种产

品的单价。而这个企业现有的第i种资源的数量是b_i，现在要作一个能够充分利用现有资源的生产计划，使每种产品在不超过现有资源的条件下，总产值最大。

解 用x_j表示生产第j种产品的数量，由于所消耗的资源不能超过现有的资源数量，所以

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \le b_i, \quad (i = 1, 2, \cdots, m).$$

该问题的数学模型为：

$$\min(\max) f(\boldsymbol{X}) = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

$s.t.$

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m$$

$$x_j \geq 0, \quad j = 1, 2, \cdots, n$$

∎

一般地，线性规划的的数学模型为

$$\min f(\boldsymbol{X}) = \sum_{j=1}^{n} c_j x_j$$

$s.t.$

$$\begin{aligned}
&\sum_{j=1}^{n} a_{pj} x_j \leq b_p, \quad p = 1, 2, \cdots, u \\
&\sum_{j=1}^{n} a_{qj} x_j \geq b_q, \quad q = u+1, u+2, \cdots, u+v \\
&\qquad \vdots \\
&\sum_{j=1}^{n} a_{rj} x_j = b_r, \quad p = u+v+1, \cdots, m \\
&x_j \geq 0, \quad j = 1, 2, \cdots, n
\end{aligned} \tag{2-1-1}$$

通过

（1）$\max f(\boldsymbol{X}) = -\min f(\boldsymbol{X})$（极大问题极小化）

（2）$\sum_{j=1}^{n} a_{pj}x_j + x_{n+p} = b_p$（$x_{n+p} \geq 0$：松弛变量）

（3）$\sum_{j=1}^{n} a_{qj}x_j + x_{n+q} = b_q$（$x_{n+q} \geq 0$：剩余变量）

（4）$\exists x_i \in \boldsymbol{R}$，则

（I）$x_i = x_i^+ - x_i^-$；（x_i^+，$x_i^- \geq 0$：自由变量）

（II）通过解含有x_i的等式约束将变量消去。

可转化为等价的**标准形式**：

$$\min f(\boldsymbol{X}) = \sum_{j=1}^{n} c_j x_j$$

$$s.t. \tag{2-1-2}$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, 2 \cdots, m$$

$$x_j \geq 0, \quad j = 1, 2, \cdots, n$$

注 在标准型中，还要求$b_i \geq 0$，$(i = 1, 2, \cdots, m)$。

线性规划的**矩阵形式：**

$$
\begin{aligned}
&\min f(\boldsymbol{X}) = \boldsymbol{C}^T \boldsymbol{X} \\
s.t.\quad & \\
&\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b} \\
&\boldsymbol{X} \geq 0
\end{aligned}
\tag{2-1-3}
$$

这里

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \boldsymbol{X}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\boldsymbol{C}=\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}, \boldsymbol{b}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

若令$\boldsymbol{P}_j = (a_{1j}, a_{2j}, \cdots, a_{mj})^T$，则线性规划(2-1-3)可化成下面的**向量形式**

$$\min f(\boldsymbol{X}) = \boldsymbol{C}^T\boldsymbol{X}$$

$$s.t.$$

$$\sum_{j=1}^{n} x_j\boldsymbol{P}_j = \boldsymbol{b} \quad or \quad \sum_{j=1}^{n} \boldsymbol{P}_j x_j = \boldsymbol{b} \tag{2-1-4}$$

$$\boldsymbol{X} \geq 0$$

例 2.3 将下面的线性规划化为标准型

$$\max f(\boldsymbol{X}) = x_1 - 2x_2 + 3x_3$$

$$s.t.$$

$$2x_1 - 7x_3 \leq 0$$

$$3x_1 + 2x_2 \geq 0$$

$$x_1 + x_2 + x_3 = 5$$

$$x_1, x_2 \geq 0, x_3 \in R$$

§ 2.2 二维线性规划的图解法

其对于提出和理解一般的线性规划的理论与求解方法有很

大的帮助。

例 2.4

$$\min f(\boldsymbol{X}) = -x_1 - 2x_2$$

$$s.t.$$

$$-x_1 + 2x_2 \leq 4$$

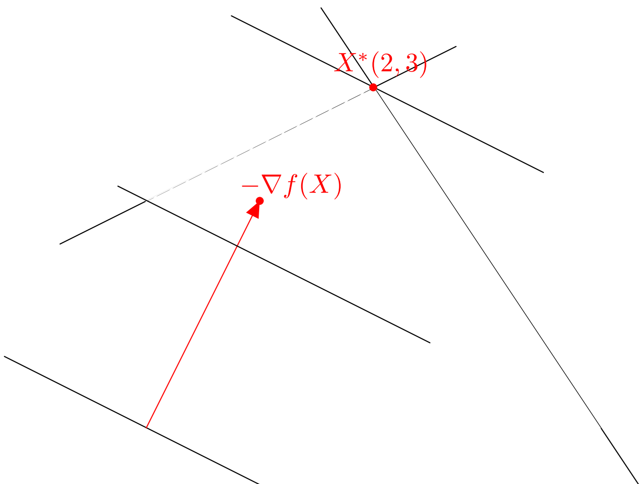
$$3x_1 + 2x_2 \leq 12$$

$$x_1, x_2 \geq 0.$$

由前面的例子可以看到

（1）$R = \emptyset$，无最优解；

$$
\begin{aligned}
&\quad \min f(\boldsymbol{X}) = -x_1 - 2x_2 \\
&s.t. \\
&\quad -x_1 + 2x_2 \leq 4 \\
&\quad 3x_1 + 2x_2 \leq 12 \\
&\quad x_1, x_2 \geq 0.
\end{aligned}
$$



（2）R为有界闭集，有唯一的最优解或无穷多个最优解；

（3）R为无界集，可能有最优解，也可能无最优解。

§ 2.3 基本概念与解得性质

此节为下面介绍线性规划的单纯形方法奠定基础，展开思路。

对于线性规划(2-1-3)，即

$$\min f(\boldsymbol{X}) = \boldsymbol{C}^T\boldsymbol{X}$$

$$s.t.$$

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b}$$

$$\boldsymbol{X} \geq 0$$

其中$\boldsymbol{A} = (a_{ij})_{m\times n}$，设$rank(\boldsymbol{A}) = m$，则$n \geq m$。

2.3.1 基本概念

（1）基

记$j1, j2, \cdots, jm$为$1, 2, \cdots, n$中m个数的一个组合，若$\boldsymbol{B} =$

$(\boldsymbol{P}_{j1}, \boldsymbol{P}_{j2}, \cdots, \boldsymbol{P}_{jm})$可逆，则$\boldsymbol{B}$称为线性规划的基。$\boldsymbol{P}_{ji}$，$j = 1, 2, \cdots, m$称为基向量。$x_{ji}$称为基变量，其余的变量称为非基变量。

（2）基本解

设$\boldsymbol{B} = (\boldsymbol{P}_{j1}, \boldsymbol{P}_{j2}, \cdots, \boldsymbol{P}_{jm})$是线性规划(2-1-3)的基，其相应的基变量$\boldsymbol{X}_B = (x_{j1}, x_{j2}, \cdots, x_{jm})^T$，则方程组$\boldsymbol{B}\boldsymbol{X}_B = \boldsymbol{b}$有唯一解

$$\boldsymbol{X}_B = \boldsymbol{B}^{-1}\boldsymbol{b} = (x_{j1}^0, x_{j2}^0, \cdots, x_{jm}^0)^T,$$

令非基变量全部为零，则得到$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b}$的一组解

$$x_{j1} = x_{j1}^0, \quad \forall j = 1, 2, \cdots, m. \text{ 其余} x_j = 0.$$

。这个解称为基本解。

（3）**可行解**

$\{\boldsymbol{X}|\boldsymbol{AX}=\boldsymbol{b},\ \boldsymbol{X}\geq 0\}$

（4）**基可行解**

既是基本集，又是可行解。即所有分量非负的基本解。

注 不同的基最多有$\binom{n}{m}$个，而一个基最多对应一个基可行解。

（5）**最优基可行解、最优基**

若$\boldsymbol{X}^0$是一个基可行解，且对任意的基可行解$\boldsymbol{X}$，都有$f(\boldsymbol{X}^0)\leq f(\boldsymbol{X})$，则称$\boldsymbol{X}^0$为最优基可行解。而$\boldsymbol{X}^0$所对应

的基为最优基。

例 2.5 求约束

$$x_1 + x_2 + x_3 = 2$$

$$x_1 + 2x_2 + 4x_3 = 6$$

$$x_1, x_2, x_3 \geq 0$$

的所有基可行解。

解 $\boldsymbol{X}^1 = (0, 1, 1)^T$，$\boldsymbol{X}^2 = \left(\frac{2}{3}, 0, \frac{4}{3}\right)^T$。 ∎

例 2.6 求约束

$$x_1 + x_2 + x_3 = 2$$

$$x_1 + 2x_2 + 4x_3 = 4$$

$$x_1, x_2, x_3 \geq 0$$

所确定的全部基可行解。

（6）**顶点**

设$\boldsymbol{X}^0 \in D \subset \boldsymbol{R}^n$，$D$是凸集。如果$\boldsymbol{X}^0$不能表示为$D$中其它任意两个不同点的凸组合，则称$\boldsymbol{X}^0$为$D$的顶点。

2.3.2 一个例子

例 2.7 某工厂在生产计划期内要生产A，B两种产品，已知生产单位产品所需的I，II两种原材料以及设备台时，如下表所示：若每生产一件产品A和B分别可获利1万和3万，问如

	A	B	限额
原材料I	1	0	4(吨)
原材料II	0	1	3(吨)
设备	1	2	8(小时)

何安排生产计划可使工厂获利最大？

解 设生产产品A，B分别为x_1，x_2件，则有

$$\max f(X) = x_1 + 3x_2$$

$$s.t.$$

$$x_1 \leq 4$$

$$x_2 \leq 3$$

$$x_1 + 2x_2 \leq 8$$

引入松弛变量x_3，x_4，x_5，化为标准型

$$\max f(X) = x_1 + 3x_2$$

$$s.t.$$

$$x_1 + x_3 = 4$$

$$x_2 + x_4 = 3$$

$$x_1 + 2x_2 + x_5 = 8$$

$$x_i \geq 0,\ i = 1 \sim 5.$$

$\blacksquare$

2.3.3 解的性质

定理 2.8 (引理) 设$\boldsymbol{X}$是线性规划(2-1-3)的可行解，则$\boldsymbol{X}$是基可行解的充要条件是$\boldsymbol{X}$的非零分量在$\boldsymbol{A}$中所对应的列向量组线性无关。

证明 不妨设$\boldsymbol{X} = (x_1, x_2, \cdots, x_l, 0, \cdots, 0)^T$，其中$x_i > 0 (i = 1, 2, \cdots, l)$。

必要性：显然。

充分性：由题设$\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_l$线性无关，故$l \leq m$。若$l = m$，则$\boldsymbol{X}$就是基可行解（非退化）。若$l < m$，则可在$\boldsymbol{A}$中选择$(m-l)$个列向量与$\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_l$一起构成$\boldsymbol{A}$的$m$个线性无关列向量组，这时$\boldsymbol{X}$就是基可行解（退化）。 ∎

定理 2.9 线性规划(2-1-3)的可行解$\boldsymbol{X}$是可行集S的顶点的充要条件是$\boldsymbol{X}$是基可行解。

证明 充分性：设$\boldsymbol{X} = (x_1, x_2, \cdots, x_m, 0, \cdots, 0)^T$是一个基可行解。则

$$\boldsymbol{A}\boldsymbol{X} = \sum_{i=1}^{m} x_i \boldsymbol{P}_i,$$

其中，$\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_m$线性无关，$\boldsymbol{A} = (\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_m, \boldsymbol{P}_{m+1}, \cdots, \boldsymbol{P}_n)$。若$\boldsymbol{X}$不是$S$的顶点，则存在$\boldsymbol{U}, \boldsymbol{V} \in R$，且$\boldsymbol{U} \neq \boldsymbol{V}$，使得

$$\boldsymbol{X} = a_1 \boldsymbol{U} + a_2 \boldsymbol{V}, \quad a_1, a_2 > 0, \quad a_1 + a_2 = 1.$$

由$\boldsymbol{U}, \boldsymbol{V} \geq 0$可设

$$\boldsymbol{U} = (u_1, u_2, \cdots, u_m, 0, \cdots, 0)^T,$$

$$\boldsymbol{V} = (v_1, v_2, \cdots, v_m, 0, \cdots, 0)^T,$$

又$\boldsymbol{U}, \boldsymbol{V}$是可行解，所以

$$\sum_{i=1}^{m} u_i \boldsymbol{P}_i = \boldsymbol{b},$$

$$\sum_{i=1}^{m} v_i \boldsymbol{P}_i = \boldsymbol{b}.$$

由$\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_m$线性无关，得

$$u_1 - v_1 = u_2 - v_2 = \cdots = u_m - v_m = 0.$$

即$\boldsymbol{U} = \boldsymbol{V}$。与$\boldsymbol{X}$为顶点矛盾。

必要性：设$\boldsymbol{X}$是S的顶点，且不妨设$\boldsymbol{X}$的前l个分量大于

零，而其余分量全部为零，即

$$\boldsymbol{X}=(x_1,x_2,\cdots,x_l,0,\cdots,0)^T.$$

又$\boldsymbol{X}\in S$知

$$\boldsymbol{A}\boldsymbol{X}=\boldsymbol{b}.$$

若$\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_l$线性相关，则存在一组不全为零的数$y_1,\cdots,y_l$，使

$$\sum_{i=1}^{l}y_i\boldsymbol{P}_i=0.$$

令$\boldsymbol{Y}=(y_1,\cdots,y_l,0,\cdots,0)^T$

则任意的ε，令

$$\boldsymbol{U}=\boldsymbol{X}+\varepsilon\boldsymbol{Y},\quad \boldsymbol{V}=\boldsymbol{X}-\varepsilon\boldsymbol{Y}.$$

则

$$\boldsymbol{AU}=\boldsymbol{AX}+\varepsilon\boldsymbol{AY}=\boldsymbol{b},\quad \boldsymbol{AV}=\boldsymbol{AX}-\varepsilon\boldsymbol{AY}=\boldsymbol{b}$$

取$\varepsilon = \min_{\substack{j=1,\cdots,l\\ y_j\neq 0}}\left\{\frac{x_j}{|y_j|}\right\}$，则$\varepsilon > 0$，且$\varepsilon|y_j| \leq x_j$，$j = 1,\cdots,l$.

所以

$$\boldsymbol{U}=\boldsymbol{X}+\varepsilon\boldsymbol{Y}\geq 0,\quad \boldsymbol{V}=\boldsymbol{X}-\varepsilon\boldsymbol{Y}\geq 0.$$

于是$\boldsymbol{U},\boldsymbol{V}\in S$，$\boldsymbol{X}=\frac{1}{2}\boldsymbol{U}+\frac{1}{2}\boldsymbol{V}$，这与$\boldsymbol{X}$为顶点矛盾。所以$\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_l$线性无关，从而由$rank(\boldsymbol{A})=m$可知$l\leq m$，所以$\boldsymbol{X}$就是基可行解。 ▮

定理 2.10 若线性规划(2-1-3)有最优解，则必在其可行集S的顶点处取得。

证明 设$\boldsymbol{X}^0$是最优解，若$\boldsymbol{X}^0=0$，则必为S的顶点。下设$\boldsymbol{X}^0=(x_1,x_2,\cdots,x_l,0,\cdots,0)^T\neq 0$，且不是$S$顶点，其中$x_i>0,(i=1,2,\cdots,l)$。因$\boldsymbol{X}^0$不是$S$的顶点，从而不是基可行解，因此存在一组不全为零的数$y_1,\cdots,y_l$，使

$$y_1\boldsymbol{P}_1+y_2\boldsymbol{P}_2+\cdots+y_l\boldsymbol{P}_l=0.$$

设$\varepsilon=\min_{\substack{j=1,\cdots,l\\ y_j\neq 0}}\left\{\frac{x_j}{|y_j|}\right\}$，$\boldsymbol{Y}=(y_1,\cdots,y_l,0,\cdots,0)^T$，则由上一个定理的证明可知

$$\boldsymbol{X}^1=\boldsymbol{X}^0+\varepsilon\boldsymbol{Y}\in S,\quad \boldsymbol{X}^2=\boldsymbol{X}^0-\varepsilon\boldsymbol{Y}\in S.$$

由$f(\boldsymbol{X})$线性性知

$$f(\boldsymbol{X}^1)=f(\boldsymbol{X}^0+\varepsilon\boldsymbol{Y})=f(\boldsymbol{X}^0)+f(\varepsilon\boldsymbol{Y}),$$

$$f(\boldsymbol{X}^2)=f(\boldsymbol{X}^0-\varepsilon\boldsymbol{Y})=f(\boldsymbol{X}^0)-f(\varepsilon\boldsymbol{Y}).$$

又由$\boldsymbol{X}^0$是最优解，得

$$f(\boldsymbol{X}^1)-f(\boldsymbol{X}^0)=f(\varepsilon\boldsymbol{Y})\geq 0,$$

$$f(\boldsymbol{X}^2) - f(\boldsymbol{X}^0) = -f(\varepsilon \boldsymbol{Y}) \geq 0.$$

所以$f(\varepsilon \boldsymbol{Y}) = 0$，$f(\boldsymbol{X}^1) = f(\boldsymbol{X}^2) = f(\boldsymbol{X}^0)$。

即$\boldsymbol{X}^1$，$\boldsymbol{X}^2$都是最优解。

由ε得取法知，$\boldsymbol{X}^1$，$\boldsymbol{X}^2$中必有一个的非零向量比$\boldsymbol{X}^0$的至少少一个(由ε的定义知)。不妨设$\boldsymbol{X}^2$的非零分量比$\boldsymbol{X}^0$的少。如果$\boldsymbol{X}^2$还不是基可行解，则可仿照此法，从$\boldsymbol{X}^2$ 出发，构造出新的最优解$\boldsymbol{X}^3$，而$\boldsymbol{X}^3$的非零分量的个数比$\boldsymbol{X}^2$少。继续下去，在有限步骤后，必得到最优解$\boldsymbol{X}^k$。若$\boldsymbol{X}^k = 0$，则为顶点；若$\boldsymbol{X}^k \neq 0$，且其非零分量对应的列向量组线性无关，则其仍然是基可行解，从而也是顶点。 ∎

§ 2.4 单纯形法

找出线性规划(2-1-3)所有的基可行解很困难，尤其当$n >> m$比较大时。单纯形法的基本思想是，从线性规划的某一个顶点出发，沿着使目标函数值下降的方向寻找下一个顶点。

2.4.1 准备工作

（1）最优解判别准则：迭代何时终止，最优解找到。

设

$$\boldsymbol{A} = (\boldsymbol{I}, \boldsymbol{N}),$$

这里$\boldsymbol{I} = (\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_m)$为单位矩阵，$\boldsymbol{N} =$

$(\boldsymbol{P}_{m+1}, \cdots, \boldsymbol{P}_n)$。

$$\boldsymbol{X_I} = (x_1, x_2, \cdots, x_m)^T,$$

$$\boldsymbol{X_N} = (x_{m+1}, x_{m+2}, \cdots, x_n)^T,$$

$$\boldsymbol{C_I} = (c_1, c_2, \cdots, c_m)^T,$$

$$\boldsymbol{C_N} = (c_{m+1}, c_{m+2}, \cdots, c_n)^T,$$

则有

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X_I} \\ \boldsymbol{X_N} \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{C_I} \\ \boldsymbol{C_N} \end{bmatrix}.$$

于是线性规划(2-1-3)可记为

$$\min f(\boldsymbol{X}) = \boldsymbol{C}_{\boldsymbol{I}}^T \boldsymbol{X}_{\boldsymbol{I}} + \boldsymbol{C}_{\boldsymbol{N}}^T \boldsymbol{X}_{\boldsymbol{N}}$$

$$s.t. \qquad \begin{aligned} &\boldsymbol{I}\boldsymbol{X}_{\boldsymbol{I}} + \boldsymbol{N}\boldsymbol{X}_{\boldsymbol{N}} = \boldsymbol{b} \\ &\boldsymbol{X}_{\boldsymbol{I}} \geq 0, \quad \boldsymbol{X}_{\boldsymbol{N}} \geq 0 \end{aligned} \tag{2-4-5}$$

显然$\boldsymbol{X}^0=\begin{bmatrix}\boldsymbol{b}\\ \boldsymbol{0}\end{bmatrix}$，所对应的目标函数值为

$$f(\boldsymbol{X}^0)=\begin{pmatrix}\boldsymbol{C}_{\boldsymbol{I}}^T & \boldsymbol{C}_{\boldsymbol{N}}^T\end{pmatrix}\begin{bmatrix}\boldsymbol{b}\\ \boldsymbol{0}\end{bmatrix}=\boldsymbol{C}_{\boldsymbol{I}}^T\boldsymbol{b}.$$

设$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}_{\boldsymbol{I}}\\ \boldsymbol{X}_{\boldsymbol{N}}\end{bmatrix}$是线性规划的任一可行解，相应的函数值为

$$f(\boldsymbol{X})=\boldsymbol{C}_{\boldsymbol{I}}^T(\boldsymbol{b}-\boldsymbol{N}\boldsymbol{X}_N)+\boldsymbol{C}_{\boldsymbol{N}}^T\boldsymbol{X}_{\boldsymbol{N}}=f(\boldsymbol{X}^0)-(\boldsymbol{C}_{\boldsymbol{I}}^T\boldsymbol{N}-\boldsymbol{C}_{\boldsymbol{N}}^T)\boldsymbol{X}_{\boldsymbol{N}}.$$

由$\boldsymbol{X}_{\boldsymbol{N}} \geq 0$可知，若$\boldsymbol{C}_{\boldsymbol{I}}^T\boldsymbol{N} - \boldsymbol{C}_{\boldsymbol{N}}^T \leq 0$，则有$f(\boldsymbol{X}) \geq f(\boldsymbol{X}^0)$。即此时$f(\boldsymbol{X}^0)$是线性规划的最优解。

因此得到下面的定理

定理 2.11 对于线性规划(2-4-5)，当$\boldsymbol{C}_{\boldsymbol{I}}^T\boldsymbol{N} - \boldsymbol{C}_{\boldsymbol{N}}^T \leq 0$时，则$\boldsymbol{X}^0 = (\boldsymbol{b} \quad \boldsymbol{0})^T$就是这个线性规划的最优解。

用分量的形式表示

$$\sigma_j = \boldsymbol{C}_{\boldsymbol{I}}^T\boldsymbol{P}_j - c_j, \quad j = m+1, \cdots, n.$$

$$\sigma_j = \boldsymbol{C}_{\boldsymbol{I}}^T\boldsymbol{P}_j - c_j = 0, \quad j = 1, \cdots, m.$$

这里$\boldsymbol{P}_j = (0, \cdots, 1_{[j]}, \cdots, 0)^T$)。

因此上述定理可叙述为：

定理 2.12 若线性规划(2-4-5)，当$\sigma_j \leq 0, (j = 1, \cdots, n)$时，则$\boldsymbol{X}^0 = (\boldsymbol{b} \quad \boldsymbol{0})^T$就是这个线性规划的最优解。

例 2.13 判断$\boldsymbol{X}^0 = (4, 0, 5, 0)^T$是否为下面线性规划的最优

解。

$$\min f(\boldsymbol{X}) = x_1 + 2x_2 + 3x_3 - 2x_4$$

$s.t.$

$$x_1 - 2x_2 + 4x_4 = 4 \tag{2-4-6}$$

$$x_2 + x_3 - 2x_4 = 5$$

$$x_j \geq 0, \quad j = 1, \cdots, 4$$

定理 2.14 若线性规划(2-4-5)的某个判别数$\sigma_j > 0$，而相应的列向量$\boldsymbol{P}_j \leq 0$，则线性规划(2-4-5)无最优解。

证明 设$(\boldsymbol{P}_1, \cdots, \boldsymbol{P}_m) = \boldsymbol{I}$，

则

$$b_1\boldsymbol{P}_1+\cdots+b_m\boldsymbol{P}_m=\boldsymbol{b},$$

$$a_{1j}\boldsymbol{P}_1+\cdots+a_{mj}\boldsymbol{P}_m=\boldsymbol{P}_j.$$

所以

$$(b_1-\theta a_{1j})\boldsymbol{P}_1+\cdots+(b_m-\theta a_{mj})\boldsymbol{P}_m+\theta\boldsymbol{P}_j=\boldsymbol{b}.$$

故当$\theta>0$时，

$$b_i-\theta a_{ij}\geq 0,\quad i=1,\cdots,m.$$

设$\boldsymbol{X}^\theta=(b_1-\theta a_{1j},\cdots,b_m-\theta a_{mj},\cdots,\theta_{[j]},\cdots)^T$，则$\boldsymbol{X}^\theta$是

线性规划(2-4-5)的可行解。

因为

$$
\begin{aligned}
f(\boldsymbol{X}^{\theta}) &= \sum_{i=1}^{m} c_i(b_i - \theta a_{ij}) + c_j\theta \\
&= \sum_{i=1}^{m} c_i b_i - \theta\Big(\sum_{i=1}^{m} c_i a_{ij} - c_j\Big) \\
&= \sum_{i=1}^{m} c_i b_i - \theta\sigma_j
\end{aligned}
$$

所以当$\theta \to +\infty$时

$$
f(\boldsymbol{X}^{\theta}) \to -\infty.
$$

即$f(\boldsymbol{X})$在可行集中无下界，因此线性规划无最优解。 ▮

例 2.15 线性规划

$$\min f(\boldsymbol{X}) = x_1 - x_2$$

$$s.t.$$

$$3x_1 - 2x_2 + x_3 = 1 \tag{2-4-7}$$

$$2x_1 - x_2 + x_4 = 5$$

$$x_j \geq 0, \quad j = 1, \cdots, 4$$

无最优解。

解 因$\sigma_2 > 0$，$\boldsymbol{P}_2 < 0$。 ▮

对于一般问题

线性规划(2-1-3)，即

$$
\begin{aligned}
&\min f(\boldsymbol{X}) = \boldsymbol{C}^T\boldsymbol{X} \\
s.t.\quad & \\
&\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b} \\
&\boldsymbol{X} \geq 0
\end{aligned}
\tag{2-4-8}
$$

设$rank(\boldsymbol{A}) = m$，$\boldsymbol{A} = (\boldsymbol{B}, \boldsymbol{N})$，$\boldsymbol{B}$可逆，取$\boldsymbol{B}$为基。

所以

$$\min f(\boldsymbol{X}) = \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{X}_{\boldsymbol{B}} + \boldsymbol{C}_{\boldsymbol{N}}^T \boldsymbol{X}_{\boldsymbol{N}}$$

$$s.t. \qquad (2\text{-}4\text{-}9)$$

$$\boldsymbol{B}\boldsymbol{X}_{\boldsymbol{B}} + \boldsymbol{N}\boldsymbol{X}_{\boldsymbol{N}} = \boldsymbol{b}$$

$$\boldsymbol{X}_{\boldsymbol{B}} \geq 0, \quad \boldsymbol{X}_{\boldsymbol{N}} \geq 0$$

由$\boldsymbol{B}$可逆，得

$$\boldsymbol{X}_{\boldsymbol{B}} = \boldsymbol{B}^{-1}\boldsymbol{b} - \boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{X}_{\boldsymbol{N}}.$$

于是

$$\begin{aligned} f(\boldsymbol{X}) &= \boldsymbol{C}_{\boldsymbol{B}}^{T}(\boldsymbol{B}^{-1}\boldsymbol{b} - \boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{X}_{\boldsymbol{N}}) + \boldsymbol{C}_{\boldsymbol{N}}^{T}\boldsymbol{X}_{\boldsymbol{N}} \\ &= \boldsymbol{C}_{\boldsymbol{B}}^{T}\boldsymbol{B}^{-1}\boldsymbol{b} + (\boldsymbol{C}_{\boldsymbol{N}}^{T} - \boldsymbol{C}_{\boldsymbol{B}}^{T}\boldsymbol{B}^{-1}\boldsymbol{N})\boldsymbol{X}_{\boldsymbol{N}} \end{aligned} \tag{2-4-10}$$

若$\boldsymbol{B}^{-1}\boldsymbol{b} \geq 0$，则

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_B \\ \boldsymbol{X}_{\boldsymbol{N}} \end{bmatrix} = \begin{bmatrix} B^{-1}\boldsymbol{b} \\ 0 \end{bmatrix}.$$

是线性规划2-4-9的一个基可行解。

若$\boldsymbol{C}_{\boldsymbol{B}}^{T}\boldsymbol{B}^{-1}\boldsymbol{N} - \boldsymbol{C}_{\boldsymbol{N}}^{T} \leq 0$

则由式（2-4-10）可知，对于任一可行解$\boldsymbol{X}$，必有

$$f(\boldsymbol{X}) \geq \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{b}.$$

故$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{B}^{-1}\boldsymbol{b} \\ 0 \end{bmatrix}$为线性规划的最优解。

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{A} - \boldsymbol{C}^T &= \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1}(\boldsymbol{B}, \boldsymbol{N}) - (\boldsymbol{C}_{\boldsymbol{B}}, \boldsymbol{C}_{\boldsymbol{N}})^T \\ &= (\boldsymbol{C}_{\boldsymbol{B}}^T, \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{N}) - (\boldsymbol{C}_{\boldsymbol{B}}, \boldsymbol{C}_{\boldsymbol{N}})^T \quad (2\text{-}4\text{-}11) \\ &= (0, \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{N} - \boldsymbol{C}_{\boldsymbol{N}}^T) \end{aligned}$$

所以$\boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{N} - \boldsymbol{C}_{\boldsymbol{N}}^T \leq 0 \Leftrightarrow \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{A} - \boldsymbol{C}^T \leq 0$

事实上，已经证明了对于上式定义的判别数$\sigma_j =$

$\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{P}_j - c_j, j = 1, \cdots, n$，前面的两个定理成立。

（2）换基运算：从一个基可行解（顶点）迭代出（转到）另一个基可行解（顶点）。

考虑下面特殊的约束

$$\begin{cases} x_1 + a_{1,m+1}x_{m+1} + \cdots + a_{1,n}x_n = b_1 \\ x_2 + a_{2,m+1}x_{m+1} + \cdots + a_{2,n}x_n = b_2 \\ \\ x_m + a_{m,m+1}x_{m+1} + \cdots + a_{m,n}x_n = b_m \\ x_j \geq 0, \quad j = 1, \cdots, n \end{cases}$$

现在从基可行解$\boldsymbol{X}$ 出发寻找新的基可行解。

令

$$(\boldsymbol{A}\boldsymbol{b}) = \begin{bmatrix} 1 & & & & & a_{1,m+1} & \cdots & a_{1l} & \cdots & a_{1n} & b_1 \\ & \ddots & & & & & & \cdots & & & \\ & & 1 & & & a_{k,m+1} & \cdots & a_{kl} & \cdots & a_{kn} & b_k \\ & & & \ddots & & & & \cdots & & & \\ & & & & 1 & a_{m,m+1} & \cdots & a_{ml} & \cdots & a_{mn} & b_m \end{bmatrix}$$

其中，$\boldsymbol{I} = (\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_m)$。

若$a_{kl} \neq 0$，则可用矩阵的初等行变换（不换行）将第i列变

成初始单位向量

$$(0,\cdots,0,1_{[l]},0,\cdots,0)^T.$$

这时$\boldsymbol{P}_k$变为非初始单位向量，同时$(\boldsymbol{A}\boldsymbol{b})$变为

$$(\boldsymbol{A}'\boldsymbol{b}')=\begin{bmatrix} 1 & & a'_{1k} & & a'_{1,m+1} & \cdots & 0 & \cdots & a'_{1n} & b'_1 \\ & \ddots & \vdots & & & & \cdots & & & \\ & & a'_{kk} & & a_{k,m+1} & \cdots & 1 & \cdots & a'_{kn} & b'_k \\ & & \vdots & \ddots & & & \cdots & & & \\ & & a'_{mk} & 1 & a'_{m,m+1} & \cdots & 0 & \cdots & a'_{mn} & b'_m \end{bmatrix}$$

其中

$$
\begin{aligned}
a'_{kj} &= \frac{a_{kj}}{a_{kl}}, \quad j = 1, \cdots, n \\
a'_{ij} &= a_{ij} - \frac{a_{kj}}{a_{kl}} a_{il}, \quad i = 1, \cdots, m; \quad i \neq k; \quad j = 1, \cdots, n \\
b'_k &= \frac{b_k}{a_{kl}} \\
b'_i &= b_i - \frac{b_k}{a_{kl}} a_{il}, \quad i = 1, \cdots, m; \quad i \neq k
\end{aligned}
\tag{2-4-12}
$$

于是得新基

$$\boldsymbol{I} = (\boldsymbol{P}_1, \cdots, \boldsymbol{P}_{k-1}, \boldsymbol{P}_l, \boldsymbol{P}_{k+1}, \cdots, \boldsymbol{P}_m).$$

这里

a_{kl}称为主元，

$\boldsymbol{P}_l$为进基列，

$\boldsymbol{P}_k$出基列，

x_l为进基变量，

x_k出基变量。

在换基运算中，如何选择主元？

令

$$\boldsymbol{X}' = (b_1', \cdots, b_{k-1}', 0, b_{k+1}', \cdots, b_m, 0, \cdots, b_{k[l]}', 0, \cdots, 0)^T,$$

由$b'_i + a_{il}b'_k = b_i - \frac{b_k}{a_{kl}}a_{il} + \frac{b_k}{a_{kl}}a_{il} = b_i$ 得

$$\boldsymbol{A}\boldsymbol{X}' = \begin{bmatrix} b'_1 + a_{1l}b'_k \\ b'_2 + a_{2l}b'_k \\ \vdots \\ b'_m + a_{ml}b'_k \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \boldsymbol{b}.$$

若$\boldsymbol{X}' \geq 0$，则$\boldsymbol{X}'$就是一个基可行解，而$\boldsymbol{X}' \geq 0$就是$b'_i \geq 0, i = 1, 2, \cdots, m$。

所以取符合条件

$$\frac{b_k}{a_{kl}} = \min_{i=1,\cdots,m} \left\{ \frac{b_i}{a_{il}} \middle| a_{il} > 0 \right\} \tag{2-4-13}$$

的a_{kl}为主元，对进行换基运算。

从而得到新基

$$\boldsymbol{I} = (\boldsymbol{P}_1, \cdots, \boldsymbol{P}_{k-1}, \boldsymbol{P}_l, \boldsymbol{P}_{k+1}, \cdots, \boldsymbol{P}_m)$$

和新基可行解

$$\boldsymbol{X}' = (b'_1, \cdots, b'_{k-1}, 0, b'_{k+1}, \cdots, b_m, 0, \cdots, b'_{k[l]}, 0, \cdots, 0)^T. \tag{2-4-14}$$

因而换基运算可按照下面步骤进行：

（I）在进基列$\boldsymbol{P}_l$中按照式(2-4-13)选择主元；

（II）按照式(2-4-12)计算b'，得到如式(2-4-14)所示的新基

可行解。

例 2.16 给定约束

$$
\begin{cases}
x_1 + 3x_4 - x_5 + 3x_6 = 2 \\
x_2 + 2x_4 - 2x_5 + x_6 = 1 \\
x_3 - 2x_4 - x_5 + 2x_6 = 3 \\
x_j \geq 0, \quad j = 1, \cdots, 6
\end{cases}
$$

显然，$\boldsymbol{X}^0=(2,1,3,0,0,0)^T$是一个初始基可行解。

$$(\boldsymbol{Ab})=\begin{bmatrix} 1 & 0 & 0 & 3 & -1 & 3 & 2 \\ 0 & 1 & 0 & \boxed{2} & -2 & 1 & 1 \\ 0 & 0 & 1 & -2 & -1 & 2 & 3 \end{bmatrix}$$

若将$\boldsymbol{P}_4$为进基向量，则主元a_{k4}满足

$$\frac{b_k}{a_{k4}}=\min_i\left\{\frac{b_i}{a_{i4}}\middle|a_{i4}>0\right\}=\min_i\left\{\frac{2}{3},\frac{1}{2}\right\}=\frac{1}{2}=\frac{b_2}{a_{24}}.$$

进行换基运算后得

$$(\boldsymbol{A}\boldsymbol{b}) = \begin{bmatrix} 1 & -\frac{3}{2} & 0 & 0 & 2 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & 1 & -1 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 1 & 0 & -3 & 3 & 4 \end{bmatrix}$$

新的基可行解为

$$\boldsymbol{X}^1 = (\frac{1}{2}, 0, 4, \frac{1}{2}, 0, 0)^T.$$

（3）进基列的选择：选择怎样的进基列（换基运算）可以使目标函数有较大的下降。

定理 2.17 若线性规划(2-4-9)，即

$$\min f(\boldsymbol{X}) = \boldsymbol{C}_{\boldsymbol{B}}^{T}\boldsymbol{X}_{\boldsymbol{B}} + \boldsymbol{C}_{\boldsymbol{N}}^{T}\boldsymbol{X}_{\boldsymbol{N}}$$

$s.t.$

$$\boldsymbol{B}\boldsymbol{X}_{\boldsymbol{B}} + \boldsymbol{N}\boldsymbol{X}_{\boldsymbol{N}} = \boldsymbol{b}$$

$$\boldsymbol{X}_{\boldsymbol{B}} \geq 0, \quad \boldsymbol{X}_{\boldsymbol{N}} \geq 0$$

满足以下条件

（I）基可行解$\boldsymbol{X}^0 = \begin{bmatrix} \boldsymbol{B}^{-1}b \\ 0 \end{bmatrix}$非退化；

（II）$\boldsymbol{P}_l$的判别数$\sigma_l > 0$；

（III）$\boldsymbol{P}_l$的分量中至少有一个为正。

则用$\boldsymbol{P}_l$作为进基列将得到使目标函数下降的基可行解。

证明 这里只讨论$\boldsymbol{B}=\boldsymbol{I}$的情形

因为$\boldsymbol{I}=(\boldsymbol{P}_1,\cdots,\boldsymbol{P}_m)$，所以$\sum_{i=1}^{m}(b_i-\theta a_{il})\boldsymbol{P}_i+\theta\boldsymbol{P}_l=\boldsymbol{b}$。

令

$$\theta=\frac{b_k}{a_{kl}}=\min_{1\le i\le m}\left\{\frac{b_i}{a_{il}}\middle|a_{il}>0\right\}.$$

则$\theta>0$且上式中$b_k-\theta a_{kl}=0$，其余各项系数$b_i-\theta a_{il}(\triangleq x_i')\ge b_i-\frac{b_i}{a_{il}}a_{il}=0$。

所以$\boldsymbol{X}^{\theta}=(x_1',\cdots,x_{k-1}',0,x_{k+1}',\cdots,x_m',\cdots,0,\theta_{[l]},0,\cdots,0)^T$

是基可行解，且相应的目标函数值为

$$\begin{aligned}
f(\boldsymbol{X}^{\theta}) &= \sum_{\substack{i=1\\ i\neq k}}^{m} c_i x_i' + c_l\theta \\
&= \sum_{i=1}^{m} c_i(b_i - \theta a_{il}) + c_i\theta \\
&= \sum_{i=1}^{m} c_i b_i - \theta\Big(\sum_{i=1}^{m} c_i a_{il} - c_l\Big) \\
&= f(\boldsymbol{X}^0) - \theta\sigma_l
\end{aligned}$$

因为$\theta > 0$以及$\sigma_l > 0$，所以

$$f(\boldsymbol{X}^{\theta}) < f(\boldsymbol{X}^0).$$

$\blacksquare$

通常情况下，满足进基条件的列很多，则选择判别数最大的那一列作为进基列（$\Delta f(\boldsymbol{X}) = f(\boldsymbol{X}^{\theta}) - f(\boldsymbol{X}^{0}) = -\theta\sigma_l$），这时目标函数值**可能将获得最大的下降**。

例 2.18

$$\min \ x_2 - 3x_3 + 2x_5$$

$$s.t.$$

$$x_1 + 3x_2 - x_3 + 2x_5 = 7$$

$$-2x_2 + 4x_3 + x_4 = 12$$

$$-4x_2 + 3x_3 + 8x_5 + x_6 = 10$$

$$x_i \geq 0, i = 1 \sim 6.$$

2.4.2 单纯形算法

（1）初始单纯形表

对于线性规划(2-4-9)，称矩阵

$$\begin{bmatrix} \boldsymbol{B}^{-1}\boldsymbol{A} & \boldsymbol{B}^{-1}\boldsymbol{b} \\ \boldsymbol{C}_{\boldsymbol{B}}^{T}\boldsymbol{B}^{-1}\boldsymbol{A}-\boldsymbol{C} & \boldsymbol{C}_{\boldsymbol{B}}^{T}\boldsymbol{B}^{-1}\boldsymbol{b} \end{bmatrix} \tag{2-4-15}$$

为初始单纯形表。

特别地，当时$\boldsymbol{B}=\boldsymbol{I}$，有

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{N} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{\sigma}_{\boldsymbol{N}}^{T} & f_0=\boldsymbol{C}_{\boldsymbol{I}}^{T}\boldsymbol{b} \end{bmatrix} \tag{2-4-16}$$

初始单纯形表为

初始单纯形表中记录以下信息：

表 2.2 初始单纯形表

$\boldsymbol{P}_1$	$\cdots$	P_k	$\cdots$	P_m	P_{m+1}	$\cdots$	P_l	$\cdots$	P_n	b
1					$a_{1,m+1}$	$\cdots$	a_{1l}	$\cdots$	a_{1n}	b_1
	$\ddots$				$\vdots$		$\vdots$		$\vdots$	$\vdots$
		1			$a_{k,m+1}$	$\cdots$	a_{kl}	$\cdots$	a_{kn}	b_k
			$\ddots$		$\vdots$		$\vdots$		$\vdots$	$\vdots$
				1	$a_{m,m+1}$	$\cdots$	a_{ml}	$\cdots$	a_{mn}	b_m
0	$\cdots$	0	$\cdots$	0	σ_{m+1}	$\cdots$	σ_l	$\cdots$	σ_n	f_0

（I）等式约束的有关数据；

（II）各列向量的判别数；

（III）初始基可行解；

（IV）对应初始基可行解的目标函数值。

（2）换基运算

在单纯形表上作换基运算，由定理2.17确定进基列$\boldsymbol{P}_l$，按式(2-4-13)确定主元a_{kl}，然后按照式(2-4-12)作换基运算。

于是上述单纯形表变成下面的新单纯形表。

表 2.3 新单纯形表

P_1	$\cdots$	P_k	$\cdots$	P_m	P_{m+1}	$\cdots$	P_l	$\cdots$	P_n	b
1		a'_{1k}			$a'_{1,m+1}$	$\cdots$	0	$\cdots$	a'_{1n}	b'_1
	$\ddots$	$\vdots$			$\vdots$		$\vdots$		$\vdots$	$\vdots$
		a'_{kk}			$a'_{k,m+1}$	$\cdots$	1	$\cdots$	a'_{kn}	b'_k
		$\vdots$	$\ddots$		$\vdots$		$\vdots$		$\vdots$	$\vdots$
		a'_{mk}		1	$a'_{m,m+1}$	$\cdots$	0	$\cdots$	a'_{mn}	b'_m
0	$\cdots$	σ'_k	$\cdots$	0	σ'_{m+1}	$\cdots$	0	$\cdots$	σ'_n	f'

容易证明下面的定理成立:

定理 2.19 （I）$\sigma'_j = \sigma_j - \frac{a_{kj}}{a_{kl}}\sigma_l$就是新表$\boldsymbol{P}_j$中的判别数；

（II）$f' = f_0 - \frac{b_k}{a_{kl}}\sigma_l$就是新基可行解

$$\boldsymbol{X}^1 = (b'_1, \cdots, b'_{k-1}, 0, b'_{k+1}, \cdots, b'_m, 0, \cdots, 0, b'_k, 0, \cdots, 0)^T$$

所对应的目标函数值。

证明 （I）

$$
\begin{aligned}
\sigma_j' &= \sigma_j - \frac{a_{kj}}{a_{kl}}\sigma_l \\
&= \Big(\sum_{i=1}^{m} c_i a_{ij} - c_j\Big) - \frac{a_{kj}}{a_{kl}}\Big(\sum_{i=1}^{m} c_i a_{il} - c_l\Big) \\
&= \sum_{i=1}^{m} c_i\Big(a_{ij} - \frac{a_{kj}}{a_{kl}}a_{il}\Big) + c_l\frac{a_{kj}}{a_{kl}} - c_j \qquad (2\text{-}4\text{-}17) \\
&= \sum_{\substack{i=1\\ i\neq k}}^{m} c_i a_{ij}' + c_l a_{kj}' - c_j \\
&= \boldsymbol{C}_{\boldsymbol{I}}^T \boldsymbol{P}_j - c_j
\end{aligned}
$$

这里$\boldsymbol{I} = (\boldsymbol{P}_1, \cdots, \boldsymbol{P}_{k-1}, \boldsymbol{P}_l, \boldsymbol{P}_{k+l}, \cdots, \boldsymbol{P}_m)$是新基。

（II）

$$
\begin{aligned}
f' &= f_0 - \frac{b_k}{a_{kl}}\sigma_l \\
&= \sum_{i=1}^{m} c_i b_i - \frac{b_k}{a_{kl}}\Big(\sum_{i=1}^{m} c_i a_{il} - c_l\Big) \\
&= \sum_{\substack{i=1 \\ i\neq k}}^{m} c_i\Big(b_i - \frac{b_k}{a_{kl}}a_{il}\Big) + c_l\frac{b_k}{a_{kl}} \\
&= \sum_{\substack{i=1 \\ i\neq k}}^{m} c_i b_i' + c_l b_k' \\
&= f(\boldsymbol{X}^1)
\end{aligned}
\tag{2-4-18}
$$

∎

由上述定理可知，在进行换基运算时，可以一并对单纯形表的最后一行做如下变化：

$$\begin{aligned} \sigma_j' &= \sigma_j - \frac{a_{kj}}{a_{kl}}\sigma_l \\ f' &= f_0 - \frac{b_k}{a_{kl}}\sigma_l \end{aligned} \tag{2-4-19}$$

（3）单纯形算法

设$\boldsymbol{A}$中有m个列向量$\boldsymbol{P}_{j1}, \boldsymbol{P}_{j2}, \cdots, \boldsymbol{P}_{jm}$构成单位矩阵$\boldsymbol{I}$，则

$$\boldsymbol{A} = (\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_n)$$

$$\boldsymbol{I} = (\boldsymbol{P}_{j1}, \boldsymbol{P}_{j2}, \cdots, \boldsymbol{P}_{jm})$$

$$\boldsymbol{b} = (b_1, b_2, \cdots, b_m)$$

$$\boldsymbol{C}^T = (c_1, c_2, \cdots, c_n)$$

$$\boldsymbol{C}_{\boldsymbol{I}}^T = (c_{j1}, c_{j2}, \cdots, c_{jm})$$

算法 2.1 单纯形方法

步骤 1 构造初始单纯形表

$$\begin{bmatrix} \boldsymbol{P}_1 & \boldsymbol{P}_2 & \cdots & \boldsymbol{P}_m & \boldsymbol{b} \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n & f_0 \end{bmatrix}$$

其中，

$$\sigma_j = \boldsymbol{C}_{\boldsymbol{I}}^T \boldsymbol{P}_j - c_j, \quad j = 1, \cdots, n,$$

$$f_0 = \boldsymbol{C}_{\boldsymbol{I}}^T \boldsymbol{b}.$$

步骤 2 求$\sigma_l = \max_{1 \le j \le n}\{\sigma_j\}$。

步骤 3 若$\sigma_l \le 0$，则$\boldsymbol{X}$是最优解，否则转步骤4。

步骤 4 若$\boldsymbol{P} \le 0$，则无最优解，否则转步骤5。

步骤 5 求$\frac{b_k}{a_{kl}} = \min_{1 \le i \le n}\left\{\frac{b_i}{a_{il}} \middle| a_{il} > 0\right\}$。

步骤 6 以a_{kl}为主元对单纯形表作换基运算得到新单纯形表，转步骤2。

例 2.20

$$\min f(\boldsymbol{X}) = x_2 - 3x_3 + 2x_5$$

$$s.t.$$

$$x_1 + 3x_2 - x_3 + 2x_5 = 7$$

$$-2x_2 + 4x_3 + x_4 = 12$$

$$-4x_2 + 3x_3 + 8x_5 + x_6 = 10$$

$$x_j \geq 0, \quad j = 1, \cdots, 6$$

解 初始基$\boldsymbol{I} = (\boldsymbol{P}_1, \boldsymbol{P}_4, \boldsymbol{P}_6)$，初始基可行解$\boldsymbol{X}^0 = (7, 0, 0, 12, 0, 10)^T$，$C_{\boldsymbol{I}}^T = (0, 0, 0)^T$。

由$\sigma_j = \boldsymbol{C}_{\boldsymbol{I}}^T \boldsymbol{P}_j - c_j$得：$\sigma_2 = -1$，$\sigma_3 = 3$，$\sigma_5 = -2$。显然基变量所对应的判别数$\sigma_1 = \sigma_4 = \sigma_6 = 0$。这时$f_0 = \boldsymbol{C}_{\boldsymbol{I}}^T \boldsymbol{b} = 0$。

初始单纯形表如下

表 2.4 单纯形表

P_1	P_2	P_3	P_4	P_5	P_6	b
1	3	-1	0	2	0	7
0	-2	$\boxed{4}$	1	0	0	12
0	-4	3	0	8	1	10
0	-1	3	0	-2	0	0

在非零判别数中，只有$\sigma_3 > 0$，且$\boldsymbol{P}_3$有正分量，故将其引入基底。

由$\frac{b_k}{a_{k3}} = \min\left\{\frac{b_i}{a_{i3}} \middle| a_{i3} > 0\right\} = \min\left\{\frac{12}{4}, \frac{10}{3}\right\} = \frac{12}{4}$，可知，

选a_{23}为主元，然后作换基运算。新单纯形表如下：

表 2.5 单纯形表

P_1	P_2	P_3	P_4	P_5	P_6	b
1	$\boxed{\frac{5}{2}}$	0	$\frac{1}{4}$	2	0	10
0	$-\frac{1}{2}$	1	$\frac{1}{4}$	0	0	3
0	$-\frac{5}{2}$	0	$-\frac{3}{4}$	8	1	1
0	$\frac{1}{2}$	0	$-\frac{3}{4}$	-2	0	-9

在此表中，同法，$\boldsymbol{P}_2$进基，a_{12}为主元。作换基运算。

表 2.6 单纯形表

P_1	P_2	P_3	P_4	P_5	P_6	b
$\boxed{\frac{2}{5}}$	1	0	$\frac{1}{10}$	$\frac{4}{5}$	0	4
$\frac{1}{5}$	0	1	$\frac{3}{10}$	$\frac{2}{5}$	0	5
1	0	0	$-\frac{1}{2}$	10	1	11
$-\frac{1}{5}$	0	0	$-\frac{4}{5}$	$-\frac{12}{5}$	0	-11

到此，所有判别数小于等于0，故当前基可行解$\boldsymbol{X}^* = (0,4,5,0,0,11)^T$为最优解，最优值为$f(\boldsymbol{X}^*) = -11$。 ▮

例 2.21

$$\min f(\boldsymbol{X}) = -x_1 - 2x_2$$

$$s.t.$$

$$x_1 \leq 4$$

$$x_2 \leq 3$$

$$x_1 + 2x_2 \leq 8$$

$$x_j \geq 0, \quad j = 1,2$$

解 转化为标准型

$$\min f(\boldsymbol{X}) = -x_1 - 2x_2$$

$s.t.$

$$x_1 + x_3 = 4$$

$$x_2 + x_4 = 3$$

$$x_1 + 2x_2 + x_5 = 8$$

$$x_j \geq 0, \quad j = 1, \cdots, 5$$

∎

表 2.7 单纯形表

P_1	P_2	P_3	P_4	P_5	b
$\boxed{1}$	0	1	0	0	4
0	1	0	1	0	3
1	2	0	0	1	8
1	2	0	0	0	0

表 2.8 单纯形表

P_1	P_2	P_3	P_4	P_5	b
1	0	1	0	0	4
0	1	0	1	0	3
0	$\boxed{2}$	-1	0	1	4
0	2	-1	0	0	-4

§ 2.5 初始基可行解的确定法

对一般线性规划，$\boldsymbol{A}$未必刚好有一个m阶单位矩阵，因此

表 2.9 单纯形表

P_1	P_2	P_3	P_4	P_5	b
1	0	1	0	0	4
0	0	$\frac{1}{2}$	1	$-\frac{1}{2}$	1
0	1	$-\frac{1}{2}$	0	$\frac{1}{2}$	2
0	0	0	0	-1	-8

没有现成的初始基可行解，但这可以通过引入人工变量的方法解决。

对于线性规划

$$\min f(\boldsymbol{X}) = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

$s.t.$

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ &a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ &\qquad \vdots \\ &a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \\ &x_j \geq 0, \quad j = 1, 2, \cdots, n \end{aligned} \tag{2-5-20}$$

2.5.1 两阶段法

引入人工变量$y_1, y_2, \cdots, y_m$，构成线性规划

$$\min g(\boldsymbol{Y}) = y_1 + y_2 + \cdots + y_m$$

$s.t.$

$$\begin{aligned}
&y_1 + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
&y_2 + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\
&\qquad\vdots \\
&y_m + a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \\
&y_i \geq 0, \quad x_j \geq 0, \quad i = 1, 2, \cdots, m, \quad j = 1, 2, \cdots, n
\end{aligned} \tag{2-5-21}$$

人工变量所对应的列$\boldsymbol{d}_i = (0,\cdots,0,1_{[i]},0,\cdots,0)^T$，$i = 1,2,\cdots,m$称为人工向量。$\boldsymbol{I} = (\boldsymbol{d}_1,\boldsymbol{d}_2,\cdots,\boldsymbol{d}_m)$为单位矩阵。$\boldsymbol{Y}^0 = (b_1,b_2,\cdots,b_m)^T$为初始基可行解。

从Y^0出发，对线性规划(2-5-21)做换基运算，当基变量全部换为x_j时，相应的基可行解就是原规划的初始基可行解。分析如下。

用单纯形法求解线性规划(2-5-21)的最优解，设最后一张单

纯形表为

$$\begin{bmatrix} d'_{11} & \cdots & d'_{1m} & a'_{11} & \cdots & a'_{1n} & b'_1 \\ d'_{21} & \cdots & d'_{2m} & a'_{21} & \cdots & a'_{2n} & b'_2 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ d'_{m1} & \cdots & d'_{mm} & a'_{m1} & \cdots & a'_{mn} & b'_m \\ \sigma''_1 & \cdots & \sigma''_m & \sigma'_1 & \cdots & \sigma'_n & g^* \end{bmatrix}$$

其中，$g^* = \sum_{i=1}^{m} y_i^*$。

分两种情况讨论。

（1）$g^* > 0$

原规划无可行解。这是因为，若原规划有可行解$\boldsymbol{X}^0 = (x_1, \cdots, x_n)^T$，令$\boldsymbol{Y}^0 = (0, \cdots, 0, x_1, \cdots, x_n)^T$就是线性规划(2-5-21)的可行解，且$g(\boldsymbol{Y}^0) = 0 < g^*$。

（2）$g^* = 0$

则对$\forall i = 1 \sim m, y_i = 0$，

（I）基变量全在x_i，$i = 1 \sim n$中，则基可行解就是原规划的初始基可行解。

（II）基变量不全在x_i，$i = 1 \sim n$中，比如y_k仍然是基变

量，则有$0 = y_k = b'_k$，此时，第k个约束条件为

$$\sum_{i=1}^{m} d'_{ki} y_i + \sum_{j=1}^{n} a'_{kj} x_j = 0.$$

- $a'_{kj} = 0, \forall j = 1 \sim n$，则由$y_i = 0, \forall i = 1 \sim m$可知，此时上式即为$0 = 0$，即在规划2-5-21中可去掉上式，也就是在基变量中消掉了y_k。
- a'_{kj}不全为零，设$a'_{kl} \neq 0$，则以a'_{kl}为主元对最后一张单纯形表作换基运算，$\boldsymbol{P}_k$出基，y_k由基变量变为非基变量。这时，无论a'_{kl}正负，均以其为主元。

考虑：这样做，在新的单纯形表中可能会出现负的$b'_i, i \in$

$\{1, \cdots, m\}$，因此基解不一定为基可行解，那么后续过程还有意义吗？

例 2.22 (两阶段单纯形法)

$$\min f(\boldsymbol{X}) = -x_1 - x_2$$

$s.t.$

$$x_1 + x_2 + x_3 + x_4 = 5$$

$$x_2 + x_4 = 2$$

$$x_1 + 2x_2 + x_3 + 2x_4 = 7$$

$$x_j \geq 0, \quad j = 1, 2, 3, 4$$

解 作辅助线性规划

$$\min f(\boldsymbol{X}) = y_1 + y_2 + y_3$$

$s.t.$

$$y_1 + x_1 + x_2 + x_3 + x_4 = 5$$

$$y_2 + x_2 + x_4 = 2$$

$$y_3 + x_1 + 2x_2 + x_3 + 2x_4 = 7$$

$$y_i \geq 0, \quad i = 1, 2, 3; \quad x_j \geq 0, \quad j = 1, 2, 3, 4$$

按照单纯形法，作两次换基运算后，得到

其中基为$(\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{d}_3)$，$\boldsymbol{Y}^* = (0, 0, 0, 3, 2, 0, 0)^T$为最优解。

表 2.10 单纯形表

d_1	d_2	d_3	P_1	P_2	P_3	P_4	b
1	0	0	1	1	1	1	5
0	1	0	0	1	0	1	2
0	0	1	1	2	1	2	7
0	0	0	2	4	2	4	14

表 2.11 单纯形表

d_1	d_2	d_3	P_1	P_2	P_3	P_4	b
1	-1	0	1	0	1	0	3
0	1	0	0	1	0	1	2
-1	-1	1	0	0	0	0	0
-2	-2	0	0	0	0	0	0

因此$\boldsymbol{X}^0=(3,2,0,0)^T$为原规划的初始基可行解。 ∎

例 2.23

$$\min \ x_1 - x_2$$

$$s.t.$$

$$-x_1 + 2x_2 + x_3 \leq 2$$

$$-4x_1 + 4x_2 - x_3 = 4$$

$$x_1 - x_3 = 0$$

$$x_i \geq 0, i = 1 \sim 3.$$

解 引入人工变量x_5，x_6，构造辅助线性规划

$$\min\ x_5 + x_6$$

$$s.t.$$

$$-x_1 + 2x_2 + x_3 + x_4 = 2$$

$$-4x_1 + 4x_2 - x_3 + x_5 = 4$$

$$x_1 - x_3 + x_6 = 0$$

$$x_i \geq 0, i = 1 \sim 6.$$

第一阶段

用单纯形法求解得

表 2.12 单纯形表

P_1	P_2	P_3	P_4	d_5	d_6	b
-1	$\boxed{2}$	1	1	0	0	2
-4	4	-1	0	1	0	4
1	0	-1	0	0	1	0
-3	4	-2	0	0	0	4
-1/2	1	1/2	1/2	0	0	1
-2	0	-3	-2	1	0	0
$\boxed{1}$	0	-1	0	0	1	0
-1	0	-4	-2	0	0	0

因为人工变量x_5，x_6是基变量，所以应该替换出。

P_1	P_2	P_3	P_4	d_5	d_6	b
0	1	0	1/2	0	1/2	1
0	0	$\boxed{-5}$	-2	1	2	0
1	0	-1	0	0	1	0
0	1	0	1/2	0	1/2	1
0	0	1	2/5	-1/5	-2/5	0
1	0	0	2/5	-1/5	3/5	0

第二阶段

P_1	P_2	P_3	P_4	b
0	1	0	1/2	1
0	0	1	2/5	0
1	0	0	2/5	0
0	0	0	-1/10	-1

所以，$X^* = (0,1,0)^T$，$f(X^*) = -1$。 ∎

2.5.2 大M法

初始基未知的情况下，也可采用大M法。

基本思想：在约束中增加人工变量$X_a = (y_1, \cdots, y_m)^T$，同时在目标函数上加上罚项$Me^T X_a$，这样，在最优化目标函数的过程中，会迫使人工变量$X_a$ 离基。这里$e^T = (1, \cdots, 1)$。

$$
\begin{aligned}
& \min\ C^T X + Me^T X_a \\
& s.t. \\
& AX + X_a = b \\
& X \geq 0, X_a \geq 0.
\end{aligned}
\tag{2-5-22}
$$

用单纯形法求解线性规划(2-5-22)，其结果为:

(1)线性规划(2-5-22)有最优解$(X^*, X_a)^T$，且$X_a = 0$，则X^*为原线性规划的最优解；

(2)线性规划(2-5-22)有最优解$(X^*, X_a)^T$，且$X_a \neq 0$，即$e^T X_a > 0$，则原线性规划无可行解。

证明 若原线性规划有可行解X，则$(X, 0)^T$是线性规划(2-5-22)的可行解，则，

$$f\big((X, 0)^T\big) = C^T X + Me^T 0 = C^T X,$$

然而，

$$f\Big((X^*, X_a)^T\Big) = C^T X^* + M e^T X_a.$$

$\blacksquare$

(3)线性规划(2-5-22)无最优解，则原规划无解。（证明略）

例 2.24

$$\min \; x_1 + x_2 - 3x_3$$

$$s.t.$$

$$x_1 - 2x_2 + x_3 \leq 11$$

$$2x_1 + x_2 - 4x_3 \geq 3$$

$$x_1 - 2x_3 = 1$$

$$x_i \geq 0, i = 1 \sim 3.$$

解 化为标准型为

$$\min \ x_1 + x_2 - 3x_3 + M(x_6 + x_7)$$

$s.t.$

$$x_1 - 2x_2 + x_3 + x_4 = 11$$

$$2x_1 + x_2 - 4x_3 - x_5 + x_6 = 3$$

$$x_1 - 2x_3 + x_7 = 1$$

$$x_i \geq 0, i = 1 \sim 7.$$

P_1	P_2	P_3	P_4	P_5	P_6	P_7	b
1	-2	1	1	0	0	0	11
2	1	-4	0	-1	1	0	3
$\boxed{1}$	0	-2	0	0	0	1	1
$3M$-1	M-1	-6M+3	0	-M	0	0	4M
0	-2	3	1	0	0	-1	10
0	$\boxed{1}$	0	0	-1	1	-2	1
1	0	-2	0	0	0	1	1
0	M-1	1	0	-M	0	1-3M	M+1
0	0	$\boxed{3}$	1	-2	2	-5	12
0	1	0	0	-1	1	-2	1
1	0	-2	0	0	0	1	1
0	0	1	0	-1	1-M	-1-M	2
0	0	1	1/3	-2/3	2/3	-5/3	4
0	1	0	0	-1	1	-2	1
1	0	0	2/3	-4/3	4/3	-7/3	9
0	0	0	-1/3	-1/3	1/3-M	-2/3-M	-2

所以$X^* = (9,1,4)^T$，$f(X^*) = -2$。 ∎

§ 2.6 单纯形法的改进

2.6.1 避免循环

在定理2.17中，曾假设基可行解$\begin{bmatrix} \boldsymbol{b} \\ 0 \end{bmatrix}$非退化，即$\boldsymbol{b} > 0$。在实际计算中，在有退化解的情况下，单纯形法一般仍然有效。然而1955年$Beale$给出了单纯形法不能求解的例子。

$$\min f(\boldsymbol{X}) = -\frac{3}{4}x_1 + 150x_2 - \frac{1}{50}x_3 + 6x_4$$

$s.t.$

$$\frac{1}{4}x_1 - 60x_2 - \frac{1}{25}x_3 + 9x_4 \leq 0$$
$$\frac{1}{2}x_1 - 90x_2 - \frac{1}{50}x_3 + 3x_4 \leq 0$$

$$x_3 \leq 1$$

$$x_j \geq 0, \quad j = 1, 2, 3, 4$$

若每次迭代都把左起第一个非基向量引入基底，那么经过6次换基运算后，单纯形表又回到初始单纯形表。为了避免

循环出现，1976年$Bland$提出了一种简单易行的方法：

（1）在所有判别数为正的那些列中，以最左边的那一列为进基列。即

$$l = \min\{j|\sigma_j > 0\}.$$

（2）在进基列中有多个分量符合主元条件，则选择基变量下标最小的那一个a_{kl}作为主元。即

$$\theta = \min_i \left\{ \frac{b_i}{a_{il}} \middle| a_{il} > 0 \right\},$$

$$k = \min_i \left\{ i \middle| \theta = \frac{b_i}{a_{il}}, a_{il} > 0 \right\}.$$

2.6.2 修正单纯形法

对于线性规划(2-4-9)，若$\boldsymbol{B}$可逆，则约束条件可写作为

$$\boldsymbol{I}\boldsymbol{X}_{\boldsymbol{B}}+\boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{X}_{\boldsymbol{N}}=\boldsymbol{B}^{-1}\boldsymbol{b}.$$

设$\boldsymbol{N}=(\boldsymbol{P}_{m+1},\cdots,\boldsymbol{P}_n)$，则$\boldsymbol{B}^{-1}\boldsymbol{N}=(\boldsymbol{B}\boldsymbol{P}_{m+1},\cdots,\boldsymbol{B}^{-1}\boldsymbol{P}_n)$。即在求解当前基可行解$\begin{bmatrix}\boldsymbol{B}^{-1}\boldsymbol{b} & 0\end{bmatrix}^T$的过程中，原$\boldsymbol{N}$中的列向量$\boldsymbol{P}_l$变成了新的列向量$\boldsymbol{B}^{-1}\boldsymbol{P}_l$。若$\boldsymbol{P}_l$在下一次是进基列，则其在进基前已变成

$$\boldsymbol{B}^{-1}\boldsymbol{P}_l=(a'_{1l},\cdots,a'_{ml})^T.$$

于是，只需关心下面的数据：

（1）$\sigma_j = \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{P}_j - c_j$;

（2）$\boldsymbol{X}_{\boldsymbol{B}} = \boldsymbol{B}^{-1} \boldsymbol{b}$;

（3）$\boldsymbol{B}^{-1} \boldsymbol{P}_l$。

定理 2.25 设在单纯形法的某次迭代中可行基为$\boldsymbol{B}$，则以a_{kl}为主元作为换基运算后，所得新基的逆为

$$\boldsymbol{B}'^{-1} = \boldsymbol{E}_{kl} \boldsymbol{B}^{-1}.$$

其中，

$$
\boldsymbol{E}_{kl} = \begin{array}{c}
\begin{bmatrix}
1 & & & -\frac{a_{1l}}{a_{kl}} & & & \\
 & \ddots & & \vdots & & & \\
 & & 1 & -\frac{a_{k-1,l}}{a_{kl}} & & & \\
 & & & \frac{1}{a_{kl}} & & & \\
 & & & -\frac{a_{k+1,l}}{a_{kl}} & 1 & & \\
 & & & \vdots & & \ddots & \\
 & & & -\frac{a_{ml}}{a_{kl}} & & & 1
\end{bmatrix} \\
k\text{列}
\end{array}
\tag{2-6-23}
$$

修正单纯形方法优点：

（1）存储少，每次迭代只存储一个基矩阵的逆；

（2）运算量减少，去掉了不参与迭代的列向量的运算。

算法 2.2 修正单纯形方法

步骤 1 计算$\boldsymbol{B}^{-1}$，$\pi = \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1}$。

步骤 2 计算$\sigma_j = \pi \boldsymbol{P}_j - c_j$，若所有$\sigma_j$非正，则当前基可行解为最优解。否则转步骤3。

步骤 3 $l = \min\{j|\sigma_j > 0\}$，计算$\boldsymbol{B}^{-1}\boldsymbol{P}_l = (a_{1l}, \cdots, a_{ml})^T$，若所有的$a_{il}$非正，则原规划无最优解。否则转步骤4。

步骤 4 求$\theta = \min_i \left\{ \frac{b_i}{a_{il}} \middle| a_{il} > 0 \right\}$，$k = \min_i \left\{ i \middle| \theta = \frac{b_i}{a_{il}}, a_{il} > 0 \right\}$。

步骤 5 形成矩阵$\boldsymbol{E}_{kl}$。

步骤 6 计算$\boldsymbol{B}'^{-1} = \boldsymbol{E}_{kl}\boldsymbol{B}^{-1}$，$\boldsymbol{X}_{\boldsymbol{B}'} = \boldsymbol{B}'^{-1}\boldsymbol{b}$，$\boldsymbol{B}^{-1} := \boldsymbol{B}'^{-1}$，转步骤1。

例 2.26

$$\min\ -4x_1-3x_2-6x_3$$

$$s.t.$$

$$3x_1+2x_2+3x_3\le 30$$

$$2x_1+2x_2+3x_3\le 40$$

$$x_j\ge 0, i=1,\cdots,3.$$

解 引入松弛变量x_4、x_5，将其化为标准形式：

$$\min\ -4x_1-3x_2-6x_3$$

$$s.t.$$

$$3x_1 + 2x_2 + 3x_3 + x_4 = 30$$

$$2x_1 + 2x_2 + 3x_3 + x_5 = 40$$

$$x_j \geq 0, i = 1, \cdots, 5.$$

这里，

$$A = \begin{bmatrix} 3 & 2 & 3 & 1 & 0 \\ 2 & 2 & 3 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 30 \\ 40 \end{bmatrix}.$$

第一步：

取$B^0 = (P_4, P_5) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (B^0)^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$

$(B^0)^{-1}b = b$，$C_{B^0}^T = (0,0)$，$\pi = C_{B^0}^T(B^0)^{-1} = (0,0)$，

$\sigma_1 = \pi P_1 - c_1 = 4$，$\sigma_2 = \pi P_2 - c_2 = 3$，$\sigma_3 = \pi P_3 - c_3 = 6$，

$(B^0)^{-1}P_1 = (3,2)^T$，$\min\left\{\frac{30}{3}, \frac{40}{2}\right\} = \frac{30}{3}$，

所以主元为a_{11}，进基列为P_1，出基列为P_4.

第二步：

$$E_{11} = \begin{bmatrix} 1/3 & 0 \\ -2/3 & 1 \end{bmatrix},$$

$(B^1)^{-1} = (P_1, P_5)^{-1} = E_{11}(B^0)^{-1} = E_{11}$，

$(B^1)^{-1}b = E_{11}b = (10,20)^T$，$C_{B^1}^T = (-4,0)$，$\pi = C_{B^1}^T(B^1)^{-1} = (-4/3\ ,0)$，

$\sigma_2=\frac{1}{3}$，$\sigma_3=2$，$\sigma_4=-\frac{4}{3}$，

$(B^1)^{-1}P_2=(\frac{2}{3},\frac{2}{3})^T$，

$\min\left\{10/(2/3\ ),20/(2/3\ )\ \right\}=15$，

所以主元为a_{12}，进基列为P_2，出基列为P_1.

第三步：

$$E_{12}=\begin{bmatrix}3/2 & 0\\ -1 & 1\end{bmatrix},$$

$$(B^2)^{-1}=(P_2,P_5)^{-1}=E_{12}(B^1)^{-1}=E_{12}E_{11}=\begin{bmatrix}1/2 & 0\\ -1 & 1\end{bmatrix},$$

$(B^2)^{-1}b=(15,10)^T$，$C_{B^2}^T=(-3,0)$，$\pi=C_{B^2}^T(B^2)^{-1}=$

$(-3/2\ ,0)$,

$\sigma_1 = -\frac{1}{2}$，$\sigma_3 = \frac{3}{2}$，$\sigma_4 = -\frac{3}{2}$,

$(B^2)^{-1}P_3 = (\frac{3}{2}, 0)^T$,

所以主元为a_{13}，进基列为P_3，出基列为P_2.

第四步：

$E_{13} = \begin{bmatrix} 2/3 & 0 \\ 0 & 1 \end{bmatrix}$,

$(B^3)^{-1} = (P_3, P_5)^{-1} = E_{13}(B^2)^{-1} = E_{13}E_{12}E_{11} = \begin{bmatrix} 1/3 & 0 \\ -1 & 1 \end{bmatrix}$,

$(B^3)^{-1}b = (10, 10)^T$，$C_{B^3}^T = (-6, 0)$，$\pi = C_{B^3}^T(B^3)^{-1} =$

$(-2, 0)$，

$\sigma_1 = -2$，$\sigma_2 = -1$，$\sigma_4 = -2$，

所以最优解为$X^* = (0, 0, 10)^T$，最优值$f(X^*) = -60$. ∎

第三章　对偶线性规划

对每一个给定的线性规划，都存在着与之对应的对偶线性规划。这两种线性规划的最优解之间存在着密切的联系。

§ 3.1 对偶问题的提出

3.1.1 经济问题

例 3.1　某工厂在一周中要安排生产I、II两种产品，这两种产品分别要在A，B，C，D四种不同的设备上加工。它们在各设备上所需要的加工时数列于下表中。已知各设备在一周

内可提供的最大加工时数分别为12h，8h，16h，12h。该厂每生产一件产品I可获利2千元，每生产一件产品II可获利3千元。问应该如何安排生产计划，才能获得最大利润？

表 3.1 加工时数

产品\ 设备	A	B	C	D
I	2	1	4	0
II	2	2	0	4

解 设x_1，x_2分别表示在一周内产品I、II的产量，则有数学模

型：

$$\max f(\boldsymbol{X}) = 2x_1 + 3x_2$$

$$s.t.$$

$$\begin{aligned} &2x_1 + 2x_2 \le 12 \\ &x_1 + 2x_2 \le 8 \\ &4x_1 \le 16 \\ &4x_2 \le 12 \\ &x_1, x_2 \ge 0 \end{aligned} \tag{3-1-1}$$

▮

另一方面，假设工厂不生产这两种产品，而是将这四种生产设备用来接收对外加工，通过收加工费来获得最大利润。那么哪一种方案可获得最大利润？

用于对外加工，工厂的决策者据需考虑如何对生产设备的工时进行定价。因此，定价应尽可能的低，但不能低于生产产品I、II所获得的利润。

设这四种设备对外加工1h所获得利润分别

为w_1，w_2，w_3，w_4千元，那么这个问题的数学模型为

$$\min g(\boldsymbol{W}) = 12w_1 + 8w_2 + 16w_3 + 12w_4$$

$s.t.$

$$2w_1 + w_2 + 4w_3 \geq 2 \tag{3-1-2}$$

$$2w_1 + 2w_2 + 4w_4 \geq 3$$

$$w_j \geq 0, \quad j = 1, \cdots, 4$$

当$\min g(\boldsymbol{W}) > \max f(\boldsymbol{X})$，应采取对外加工方案；当$\min g(\boldsymbol{W}) = \max f(\boldsymbol{X})$，两种方案都为最优；不可能出现$\min g(\boldsymbol{W}) < \max f(\boldsymbol{X})$的情形。

上面的两个规划互为（对称的）对偶线性规划。

这两种线性规划的最优解分别为$\boldsymbol{X}^* = (4,2)^T$、$\boldsymbol{W}^* = (\frac{1}{2},1,0,0)^T$。这时$f(\boldsymbol{X}^*) = g(\boldsymbol{W}^*) = 14$。

$f(\boldsymbol{X}^*) = 14$千元是在不超过设备能力的条件下，工厂安排生产I、II两种产品所能获得的最大利润。而$g(\boldsymbol{W}^*) = 14$千元则是工厂将设备用来对外加工时所获得的最低利润。于是，当对外加工的利润可以做到C，D 不亏本，而A的利润大于每台时0.5千元，B的利润大于每台时1千元时，就可以接受对外加工，否则宁可生产产品I、II。这些w_i值就相当于对第i种资源在实现最大利润时的一种价格估计，这种估计是针对企业产品而存在的一种特殊价值，称为**影子价格**。

影子价格可以分析增加那种资源能为企业带来更大利润；反映资源的短缺程度等。

3.1.2 对称形式

定义 3.2 对于线性规划

$$(P)\quad \begin{aligned} &\min \boldsymbol{C}^T\boldsymbol{X} \\ s.t.\quad & \\ &\boldsymbol{A}\boldsymbol{X} \geq \boldsymbol{b} \\ &\boldsymbol{X} \geq 0 \end{aligned} \tag{3-1-3}$$

线性规划

$$(D)\quad \begin{aligned} &\max \boldsymbol{b}^T \boldsymbol{W} \\ s.t.& \\ &\boldsymbol{A}^T \boldsymbol{W} \leq \boldsymbol{C} \\ &\boldsymbol{W} \geq 0 \end{aligned} \tag{3-1-4}$$

称为(D)的对偶线性规划，而(P)称为原规划。

定理 3.3 如果将线性规划(D)看作为原规划，则线性规划(P)就是(D)的对偶线性规划。

证明 显然。 ∎

例 3.4 写出线性规划

$$\min f(X) = 8x_1 + 6x_2 + 3x_3 + 6x_4$$

$s.t.$

$$x_1 + 2x_2 + x_4 \geq 3$$

$$3x_1 + x_2 + x_3 + x_4 \geq 6$$

$$x_3 + x_4 \geq 2$$

$$x_j \geq 0, j = 1, \cdots, 4.$$

的对偶线性规划。

解

$$\max\ \ g(W) = 3w_1 + 6w_2 + 2w_3$$

$s.t.$

$$w_1 + 3w_2 \leq 8$$

$$2w_1 + w_2 \leq 6$$

$$w_2 + w_3 \leq 3$$

$$w_1 + w_2 + w_3 \leq 6$$

$$w_j \geq 0, j = 1, \cdots, 3.$$

3.1.3 非对称形式

对于线性规划

$$
(P') \qquad \begin{aligned} & \min \boldsymbol{C}^T\boldsymbol{X} \\ s.t. \quad & \\ & \boldsymbol{A}\boldsymbol{X} = \boldsymbol{b} \\ & \boldsymbol{X} \geq 0 \end{aligned} \qquad (3\text{-}1\text{-}5)
$$

因为$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b}$等价于

$$\boldsymbol{A}\boldsymbol{X} \geq \boldsymbol{b},$$

$$\boldsymbol{A}\boldsymbol{X} \leq \boldsymbol{b}.$$

故(P')可变为

$$\min \boldsymbol{C}^T\boldsymbol{X}$$

$$s.t.$$

$$\begin{bmatrix} \boldsymbol{A} \\ -\boldsymbol{A} \end{bmatrix} \boldsymbol{X} \geq \begin{bmatrix} \boldsymbol{b} \\ -\boldsymbol{b} \end{bmatrix} \tag{3-1-6}$$

$$\boldsymbol{X} \geq 0$$

其对偶线性规划为

$$\min \begin{bmatrix} \boldsymbol{b} \\ -\boldsymbol{b} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{W}^1 \\ \boldsymbol{W}^2 \end{bmatrix}$$

$$s.t.$$

$$\begin{bmatrix} \boldsymbol{A} \\ -\boldsymbol{A} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{W}^1 \\ \boldsymbol{W}^2 \end{bmatrix} \leq \boldsymbol{C} \tag{3-1-7}$$

$$\begin{bmatrix} \boldsymbol{W}^1 \\ \boldsymbol{W}^2 \end{bmatrix} \geq 0$$

令 $\boldsymbol{W} = \boldsymbol{W}^1 - \boldsymbol{W}^2$，则$(P')$的对偶线性规划又可写成

$$(D') \qquad \begin{array}{c} \min \boldsymbol{b}^T \boldsymbol{W} \\ s.t. \\ \boldsymbol{A}^T \boldsymbol{W} \leq \boldsymbol{C} \end{array} \qquad (3\text{-}1\text{-}8)$$

例 3.5 写出线性规划

$$\min f(X) = 8x_1 + 6x_2 + 3x_3 + 6x_4$$

$s.t.$

$$x_1 + 2x_2 + x_4 = 3$$

$$3x_1 + x_2 + x_3 + x_4 = 6$$

$$x_j \geq 0, j = 1, \cdots, 4.$$

的对偶线性规划。

解

$$\max g(W) = 3w_1 + 6w_2$$

$$s.t.$$

$$w_1 + 3w_2 \leq 8$$

$$2w_1 + w_2 \leq 6$$

$$w_2 \leq 3$$

$$w_1 + w_2 \leq 6$$

$$w_j \in R,\ \ j = 1, 2.$$

∎

3.1.4 混合形式

对于混合形式的线性规划，可按表3.2所列的规则进行变换。

表 3.2 变换规则

原规划（对偶规划）	对偶规划（原规划）
min	max
目标中系数	约束条件右端项
约束条件右端项	目标中系数
约束条件 $\geq$	变量 $\geq$
约束条件 $\leq$	变量 $\leq$
约束条件 $=$	变量无约束
变量 $\geq$	约束条件 $\leq$
变量 $\leq$	约束条件 $\geq$
变量无约束	约束条件 $=$

例 3.6 写出线性规划

$$\min f(X) = 8x_1 + 6x_2 + 3x_3 + 6x_4$$

$s.t.$

$$x_1 + 2x_2 + x_4 \geq 3$$

$$3x_1 + x_2 + x_3 + x_4 \leq 6$$

$$x_3 + x_4 = 2$$

$$x_j \geq 0, j = 1, 2; x_3 \leq 0, x_4 \in R.$$

的对偶线性规划。

解

$$\max g(W) = 3w_1 + 6w_2 + 2w_3$$

$s.t.$

$$w_1 + 3w_2 \leq 8$$

$$2w_1 + w_2 \leq 6$$

$$w_2 + w_3 \geq 3$$

$$w_1 + w_2 + w_3 = 6$$

$$w_1 \geq 0, w_2 \leq 0, w_3 \in R.$$

∎

§ 3.2 对偶定理

本节考虑对称形式的对偶定理。相应的结论可以容易的推广到非对称形式的对偶规划问题上。

设$R_P = \{\boldsymbol{X} | \boldsymbol{A}\boldsymbol{X} \geq \boldsymbol{b}, \boldsymbol{X} \geq 0\}$是$(P)$的可行集；$R_D = \{\boldsymbol{W} | \boldsymbol{A}^T\boldsymbol{W} \leq \boldsymbol{C}, \boldsymbol{W} \geq 0\}$是$(D)$的可行集。

定理 3.7 $\forall \boldsymbol{X} \in R_P$，$\boldsymbol{W} \in R_D$，必有

$$\boldsymbol{C}^T\boldsymbol{X} \geq \boldsymbol{b}^T\boldsymbol{W}.$$

证明

$$\boldsymbol{C}^T\boldsymbol{X} \geq (\boldsymbol{A}^T\boldsymbol{W})^T\boldsymbol{X} = \boldsymbol{W}^T(\boldsymbol{A}\boldsymbol{X}) \geq \boldsymbol{W}^T\boldsymbol{b} = \boldsymbol{b}^T\boldsymbol{W}.$$

∎

定理 3.8 线性规划(P)与其对偶线性规划(D)都有最优解的充要条件是它们都有可行解。

证明 必要性：显然。

充分性：由上一个定理知，对$\forall \boldsymbol{X} \in R_P, \boldsymbol{W} \in R_D$，有$\boldsymbol{C}^T\boldsymbol{X} \geq \boldsymbol{b}^T\boldsymbol{W}$。因此的目标函数$(P)$在$R_P$上有下界，而$(P)$是求极小，故必有最优解。同理$(D)$也有最优解。

∎

推论 3.9 如果$\boldsymbol{X} \in R_P$，$\boldsymbol{W} \in R_D$，且$\boldsymbol{C}^T\boldsymbol{X} = \boldsymbol{b}^T\boldsymbol{W}$，则$\boldsymbol{X}$，$\boldsymbol{W}$分别是$(P)$与$(D)$的最优解。

定理 3.10 若线性规划(P)与其对偶线性规划(D)中有一个有最优解，则另一个必有最优解，且最优值相等。

证明 不妨设(P)有最优解。引入剩余变量$\boldsymbol{Y}$，将化为标准形

式

$$\min \begin{bmatrix} \boldsymbol{C}^T & \boldsymbol{0}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix}$$

$$s.t. \tag{3-2-9}$$

$$\begin{bmatrix} \boldsymbol{A} & -\boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} = \boldsymbol{b}$$

$$\boldsymbol{X}, \boldsymbol{Y} \geq 0$$

设$\boldsymbol{X}^0$是上面标准线性规划的一个最优基可行解，$\boldsymbol{B}$为对应

的最优基，则由最优解判别准则知

$$\boldsymbol{C}_B^T \boldsymbol{B}^{-1}[\boldsymbol{A} \quad -\boldsymbol{I}] \leq [\boldsymbol{C}^T \quad \boldsymbol{0}^T].$$

令$\boldsymbol{W}^0 = (\boldsymbol{C}_B^T \boldsymbol{B}^{-1})^T$，则

$$(\boldsymbol{W}^0)^T[\boldsymbol{A} \quad -\boldsymbol{I}] \leq [\boldsymbol{C}^T \quad \boldsymbol{0}^T].$$

即，$\boldsymbol{W}^0$是(D)的可行解。

又

$$\boldsymbol{b}^T \boldsymbol{W}^0 = \boldsymbol{b}^T(\boldsymbol{C}_B^T \boldsymbol{B}^{-1})^T = \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{b} = \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{X}_{\boldsymbol{B}}^0 = \boldsymbol{C}^T \boldsymbol{X}^0.$$

故由推论知命题成立。 ∎

从上面定理的证明过程看，若$\boldsymbol{B}$是线性规划(P)的最优基，那么$(\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1})^T$就是对偶线性规划$(D)$的最优解。

对于原规划(P)引入剩余变量$x_{n+l}, l=1,\cdots,m$，使得变成标准的线性规划(P')，然后对其进行单纯形求解，当全部判别数非正时，得到最终单纯形表。设最优基为$\boldsymbol{B}$，则引入的剩余变量的判别数是

$$\sigma_{n+l}=\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{P}_{n+l}-c_{n+l},\quad l=1,2,\cdots,m.$$

因x_{n+l}是引入变量，故$c_{n+l}=0$。而$\boldsymbol{P}_{n+l}=$

$[0,\cdots,0,-1_{[l]},0,\cdots,0]^T$。所以

$$\sigma_{n+l}=\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{P}_{n+l}=-(\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1})_l=-(\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1})_l^T.$$

又由$\boldsymbol{W}^0=(y_1^0,\cdots,y_l^0,\cdots,y_m^0)^T$知

$$y_l^0=-(\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1})_l^T=-\sigma_{n+l}.$$

即对偶规划(D)的最优解的第l个分量就是原规划(P)的最终单纯形表中剩余变量$x_{n+l}(l=1\sim m)$的判别数σ_{n+l}的相反数。

例 3.11 求线性规划

$$\max 2x_1 + 3x_2$$

$$s.t.$$

$$2x_1 + 2x_2 \leq 12$$

$$x_1 + 2x_2 \leq 8$$

$$4x_1 \leq 16$$

$$4x_2 \leq 12$$

$$x_j \geq 0, j = 1 \sim 2.$$

的对偶线性规划的最优解。

解 对偶线性规划为

$$\min 12w_1 + 8w_2 + 16w_3 + 12w_4$$

$s.t.$

$$2w_1 + w_2 + 4w_3 \geq 2$$

$$2w_1 + 2w_2 + 4w_4 \geq 3$$

$$w_j \geq 0, j = 1 \sim 4.$$

将原线性规划化为标准型:

$$\min -2x_1 - 3x_2$$

$s.t.$

$$2x_1 + 2x_2 + x_3 \leq 12$$

$$x_1 + 2x_2 + +x_4 \leq 8$$

$$4x_1 + x_5 \leq 16$$

$$4x_2 + x_6 \leq 12$$

$$x_j \geq 0, j = 1 \sim 6.$$

用单纯形方法求解，见最后一个单纯形表格如下(表3.3)。因此原线性规划的最优解为$X^* = (4,2)^T$，又因为$\sigma_3 = -1/2$，$\sigma_4 = -1$，$\sigma_5 = 0$，$\sigma_6 = 0$，所以对偶线性规划的

表 3.3 单纯形表

P_1	P_2	P_3	P_4	P_5	P_6	b
0	0	2	-4	0	1	4
1	0	1	-1	0	0	4
0	0	-4	4	1	0	0
0	1	-1/2	1	0	0	2
0	0	-1/2	-1	0	0	-14

最优解为$W^* = (1/2, 1, 0, 0)^T$。 ∎

注：对于相互对偶的线性规划，它们的解之间有三种关系：

（1）两个规划都有最优解（可行解）；

（2）两个规划都无最优解（无可行解）；

（3）一个无可行解，另一个有可行解，但目标函数在可

行集上无界。

定理 3.12 (对称形式的松弛定理) 若$\boldsymbol{X}$与$\boldsymbol{W}$分别是(P)与(D)的可行解，则$\boldsymbol{X}$与$\boldsymbol{W}$分别是(P)与(D)的最优解的充要条件是

$$\begin{aligned}(\boldsymbol{A}^T\boldsymbol{W}-\boldsymbol{C})^T\boldsymbol{X}&=0\\ \boldsymbol{W}^T(\boldsymbol{A}\boldsymbol{X}-\boldsymbol{b})&=0\end{aligned} \tag{3-2-10}$$

证明 充分性：由条件3-2-10易得$\boldsymbol{C}^T\boldsymbol{X}=\boldsymbol{b}^T\boldsymbol{W}$。

必要性：因为$\boldsymbol{X}$，$\boldsymbol{W}$是可行解，所以

$$\boldsymbol{W}^T\boldsymbol{A}\boldsymbol{X}\geq\boldsymbol{W}^T\boldsymbol{b}=\boldsymbol{b}^T\boldsymbol{W},$$

$$\boldsymbol{W}^T\boldsymbol{A}\boldsymbol{X} = \boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{W} \leq \boldsymbol{X}^T\boldsymbol{C} = \boldsymbol{C}^T\boldsymbol{X}.$$

所以

$$\boldsymbol{C}^T\boldsymbol{X} = \boldsymbol{b}^T\boldsymbol{W}.$$

因此

$$\boldsymbol{W}^T\boldsymbol{A}\boldsymbol{X} = \boldsymbol{C}^T\boldsymbol{X},$$

$$\boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{W} = \boldsymbol{b}^T\boldsymbol{W}.$$

$\blacksquare$

若$\boldsymbol{A}_i$表示矩阵$\boldsymbol{A}$的行，而$\boldsymbol{P}_j$表示列。由

$$(\boldsymbol{A}^T\boldsymbol{W} - \boldsymbol{C})^T\boldsymbol{X} = (\boldsymbol{W}^T\boldsymbol{A} - \boldsymbol{C}^T)\boldsymbol{X} = 0$$

可得

$$(\boldsymbol{W}^T\boldsymbol{P}_j - c_j)x_j = 0, \quad j = 1, \cdots, n.$$

因此

$$\begin{cases} x_j = 0, & if \quad \boldsymbol{W}^T\boldsymbol{P}_j - c_j < 0 \\ \boldsymbol{W}^T\boldsymbol{P}_j - c_j = 0, & if \quad x_j > 0 \end{cases} \tag{3-2-11}$$

同理有

$$\begin{cases} y_i = 0, & if \quad \boldsymbol{A}_i\boldsymbol{X} - b_i > 0 \\ \boldsymbol{A}_i\boldsymbol{X} - b_i = 0, & if \quad y_i > 0 \end{cases} \tag{3-2-12}$$

因此定理3.12等价于

定理 3.13 (对称形式的松弛定理)　若$\boldsymbol{X}$与$\boldsymbol{W}$分别是(P)与(D)的可行解，则$\boldsymbol{X}$与$\boldsymbol{W}$分别是(P)与(D)的最优解的充要条件是

$$\begin{cases} y_i = 0, \quad i = 1 \sim m, \;\; if \quad \boldsymbol{A}_i\boldsymbol{X} - b_i > 0 \\ x_j = 0, \quad j = 1 \sim n, \;\; if \quad \boldsymbol{W}^T\boldsymbol{P}_j - c_j < 0 \end{cases} \tag{3-2-13}$$

若将符号约束$x_j \geq 0$，$y_i \geq 0$视作为约束条件，那么对偶线性规划的$2(m+n)$约束之间存在如下对偶关系。

定义 3.14 (对偶约束)　$\boldsymbol{A}_i\boldsymbol{X} \geq b_i$与$y_i \geq 0$称为一对对偶约束；$\boldsymbol{W}^T\boldsymbol{P}_j \leq c_j$与$x_j \geq 0$也称为一对对偶约束。

定义 3.15 (紧、松约束) 如果线性规划的每一个最优解都使得某个约束取等号，就称这个约束是紧的；否则称为松的。

定理 3.16 设(P)与(D)都有可行解，则松约束的对偶约束是紧约束。

定理 3.17 (非对称形式的松弛定理) 若$\boldsymbol{X}$与$\boldsymbol{W}$分别是(P')与(D')的可行解，则$\boldsymbol{X}$与$\boldsymbol{W}$分别是(P')与(D')的最优解的充要条件是

$$(\boldsymbol{A}^T\boldsymbol{W}-\boldsymbol{C})^T\boldsymbol{X}=0. \tag{3-2-14}$$

例 3.18 已知线性规划

$$\max f(X) = 3x_1 + 4x_2 + x_3$$

$$s.t.$$

$$x_1 + 2x_2 + x_3 \leq 10$$

$$2x_1 + 2x_2 + x_3 \leq 16$$

$$x_j \geq 0, j = 1, 2, 3.$$

的最优解是$X^* = (6, 2, 0)^T$，求其对偶规划的最优解Y^*。

解 对偶规划为

$$\min g(Y) = 10y_1 + 16y_2$$

$$s.t.$$

$$y_1 + 2y_2 \geq 3$$

$$2y_1 + 2y_2 \geq 4$$

$$y_1 + y_2 \geq 1$$

$$y_j \geq 0, j = 1, 2.$$

设对偶规划的最优解为$Y^* = (y_1^*, y_2^*)^T$，因为$x_1^* = 6 >$

0，$x_2^* = 2 > 0$，由松弛定理得

$$\begin{cases} y_1^* + 2y_2^* = 3 \\ 2y_1^* + 2y_2^* = 4 \end{cases} \Rightarrow \ y_1^* = y_2^* = 1.$$

所以对偶规划的最优解为$Y^* = (1,1)^T$，最优值为26. ▮

§ 3.3 对偶单纯形方法

3.3.1 基本思想

对于线性规划与有以下定理。

$$
(P') \quad \begin{aligned} & \min \boldsymbol{C}^T \boldsymbol{X} \\ s.t. \quad & \\ & \boldsymbol{A}\boldsymbol{X} = \boldsymbol{b} \\ & \boldsymbol{X} \geq 0 \end{aligned}
$$

$$(D')\qquad \begin{aligned} &\max \boldsymbol{b}^T\boldsymbol{W} \\ &s.t. \\ &\boldsymbol{A}^T\boldsymbol{W} \leq \boldsymbol{C} \end{aligned}$$

定理 3.19 设$\boldsymbol{X}$是(P')的任一基本解，$\boldsymbol{X}$对应于基$\boldsymbol{B}$，$\boldsymbol{W} = (\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1})^T$，若$\boldsymbol{X}$与$\boldsymbol{W}$分别是$(P')$与$(D')$的可行解，则$\boldsymbol{X}$与$\boldsymbol{W}$分别是$(P')$与$(D')$的最优解。

证明 $\boldsymbol{W} = (\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1})^T$是的可行解，所以有

$$\boldsymbol{A}^T(\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1})^T \leq \boldsymbol{C} \Rightarrow \boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{A} - \boldsymbol{C}^T \leq 0.$$

即基可行解$\boldsymbol{X}$的所有判别数$\sigma_j = \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{P}_j - c_j$非正，因此为$(P')$的最优解。

又

$$\boldsymbol{b}^T \boldsymbol{W} = \boldsymbol{b}^T (\boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1})^T = \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1} \boldsymbol{b} = \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{X}_{\boldsymbol{B}} = \boldsymbol{C}^T \boldsymbol{X}.$$

证毕！ ∎

注 $\boldsymbol{X}$的判别数σ_j全部非正与$\boldsymbol{W} = (\boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{B}^{-1})^T$为$(D')$的可行解等价。

定义 3.20 (正则解、正则基) 设$\boldsymbol{X}$是原规划(P')的一个基本解，且$\boldsymbol{X}$所有的判别数非正，则称$\boldsymbol{X}$为(P')的一个正则

解。$\boldsymbol{X}$所对应的基称作为正则基。

设$\boldsymbol{X}^0$是一个正则解，$\boldsymbol{B}$是$\boldsymbol{X}^0$对应的正则基，则$\sigma_j = \boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{P}_j - c_j \leq 0$，因此$\boldsymbol{W} = (\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1})^T$是$(D')$的可行解。从正则解$\boldsymbol{X}^0$出发，寻找下一个正则解$\boldsymbol{X}^1$，在迭代过程中，使得$\boldsymbol{X}^i$中的负分量逐次减少，同时保持$\boldsymbol{W} = (\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1})^T$对于$(D')$的可行性（$\sigma_j \leq 0$）。当某个正则解$\boldsymbol{X}^k$的分量全部非负时，$\boldsymbol{X}^k$就成为$(P')$的基可行解。因为$\sigma_j \leq 0$，所以$\boldsymbol{X}^k$就是的最优解$\boldsymbol{X}^*$。这就是单纯形法的基本思想。

对偶单纯形方法也需要解决下面三个问题：

（1）确定初始正则基。

（2）换基运算：寻找下一个正则解。

（3）终止准则。

换基运算过程如下：

设(P')有初始正则解$\boldsymbol{X}^0 = (b_1, b_2, \cdots, b_n)^T$与初始正则基$\boldsymbol{B} = [\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_m] = \boldsymbol{I}$。

（1）建立初始对偶单纯形表；

$$(\boldsymbol{A}|\boldsymbol{b}) = \begin{bmatrix} 1 & & & & a_{1,m+1} & \cdots & a_{1l} & \cdots & a_{1n} & b_1 \\ & \ddots & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & 1 & & a_{k,m+1} & \cdots & a_{kl} & \cdots & a_{kn} & b_k \\ & & & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & 1 & a_{m,m+1} & \cdots & a_{ml} & \cdots & a_{mn} & b_m \\ 0 & \cdots & 0 & \cdots \; 0 & \sigma_{m+1} & \cdots & \sigma_l & \cdots & \sigma_n & f_0 \end{bmatrix}$$

（2）确定主元a_{kl}；

设$b_k < 0$，以a_{kl}为主元做换基运算，则有

$$\begin{aligned} b'_k &= \frac{b_k}{a_{kl}} \\ \sigma'_j &= \sigma_j - \frac{a_{kj}}{a_{kl}}\sigma_l, \quad j = m+1, \cdots, n \end{aligned} \tag{3-3-15}$$

为了使$b_k \geq 0$，且$\sigma_j \leq 0$，a_{kl}应满足

$$\begin{aligned} &a_{kl} < 0 \\ &\frac{\sigma_l}{a_{kl}} = \min\left\{\frac{\sigma_j}{a_{kj}}\middle| a_{kj} < 0\right\} \end{aligned} \tag{3-3-16}$$

（3）换基运算：

$$
\begin{aligned}
b'_k &= \frac{b_k}{a_{kl}} \\
b'_i &= b_i - \frac{b_k}{a_{kl}} a_{il}, \quad i = 1, \cdots, m; \quad i \neq k \\
a'_{kj} &= \frac{a_{kj}}{a_{kl}}, \quad j = 1, \cdots, n \\
a'_{ij} &= a_{ij} - \frac{a_{kj}}{a_{kl}} a_{il}, \quad i = 1, \cdots, m; \quad i \neq k; \quad j = 1, \cdots, n
\end{aligned}
\tag{3-3-17}
$$

注(1) 单纯形法中由$\sigma_l > 0$确定进基列$\boldsymbol{P}_l$，再在$\boldsymbol{P}_l$中选择主元a_{kl}；而在对偶单纯形法中，是先由$b_k < 0$确定所在的行标k，再在这个行中选择主元a_{kl}。因此，单纯形法是按照列

选择主元，而对偶单纯形法是按行的。

注(2) 因为

$$f_1 = f_0 - \frac{\sigma_l}{a_{kl}} b_k \tag{3-3-18}$$

所以$\sigma_l < 0$时，

$$f_1 - f_0 > 0 \tag{3-3-19}$$

即线性规划(P')正则解的目标函数值越来越大，也就是其对偶线性规划(D')在可行解W 的目标函数值$b^T W = C_B^T B^{-1} b$越来越大。又因为基的个数有限，所以对偶单纯形法会在有限步内找到问题的最优解(若最优解存在)。

注(3) 若$b_k < 0$，但是$a_{kj} \geq 0$，即无法选择主元，这时线

性规划(P')无可行解。

注(4) 在最优对偶单纯表中，若$b_i = 0$，则线性规划(P')有无穷多解。

3.3.2 对偶单纯形法

已知

$$\boldsymbol{A} = [\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_n],$$

$$\boldsymbol{B} = [\boldsymbol{P}_{j1}, \boldsymbol{P}_{j2}, \cdots, \boldsymbol{P}_{jm}] = \boldsymbol{I},$$

$$\boldsymbol{b} = (b_1, b_2, \cdots, b_m)^T,$$

$$\boldsymbol{C}^T = (c_1, c_2, \cdots, c_n),$$

$$\boldsymbol{C}_{\boldsymbol{B}}^T = (c_{j1}, c_{j2}, \cdots, c_{jm}),$$

$$\sigma_j = \boldsymbol{C}_{\boldsymbol{B}}^T \boldsymbol{P}_j - c_j \leq 0.$$

例 3.21 用对偶单纯形法求解下面的线性规划

$$\min f(\boldsymbol{X}) = x_1 + 2x_2$$

$s.t.$

$$x_1 + 2x_2 \geq 4$$

$$x_1 \leq 5$$

$$3x_1 + x_2 \geq 6$$

$$x_j \geq 0, \quad j = 1 \sim 2.$$

算法 3.1 对偶单纯形方法

步骤 1 构造对偶初始单纯形表

$$\begin{bmatrix} \boldsymbol{P}_1 & \boldsymbol{P}_2 & \cdots & \boldsymbol{P}_m & \boldsymbol{b} \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n & f_0 \end{bmatrix}$$

步骤 2 若$\boldsymbol{b} \geq 0$，则当前正则解就为最优解，否则，取

$$k = \min\{i|b_i < 0\}.$$

步骤 3 若$a_{kj} \geq 0, j = 1 \sim n$，则原规划$(P')$无可行解，否则转步骤5。

步骤 4 求$\frac{\sigma_l}{a_{kl}} = \min_{1\leq j\leq n}\left\{\frac{\sigma_j}{a_{kj}}\middle|a_{kj} < 0\right\}$。

步骤 5 以a_{kl}为主元按照式（3-3-17）对对偶单纯形表作换基运算得到新对偶单纯形表，转步骤2。

解 转化为下面的标准型

$$\min f(\boldsymbol{X}) = x_1 + 2x_2$$

$$s.t.$$

$$-x_1 - 2x_2 + x_3 = -4$$

$$x_1 + x_4 = 5$$

$$-3x_1 - x_2 + x_5 = -6$$

$$x_j \geq 0, \quad j = 1 \sim 5.$$

显然$\boldsymbol{X}^0 = (0, 0, -4, 5, -6)^T$是一个正则解，从$\boldsymbol{X}^0$出发寻找下一个正则解。

（1）初始对偶单纯形表：

表 3.4 对偶单纯形表

P_1	P_2	P_3	P_4	P_5	b
$\boxed{-1}$	-2	1	0	0	-4
1	0	0	1	0	5
-3	-1	0	0	1	-6
-1	-2	0	0	0	0

（2）$x_{j1} = x_3 = -4$，$x_{j2} = x_4 = 5$，$x_{j3} = x_5 = 6$，$j = \min\{j_1, j_3\} = 3$，$k = 1$。

（3）$\theta = \min\left\{\frac{-1}{-1}, \frac{-2}{-2}\right\} = 1$，$l = \min\{1, 2\} = 1$

（4）以a_{11}为主元对上表做换基运算，得下一个对偶单纯形表：

表 3.5 对偶单纯形表

P_1	P_2	P_3	P_4	P_5	b
1	2	-1	0	0	4
0	-2	1	1	0	1
0	5	-3	0	1	6
0	0	-1	0	0	4

这时$\boldsymbol{b} > 0$，所以当前正则解$\boldsymbol{X}^* = (4,0,0,1,6)^T$是最优解，因此原规划的最优解为$\boldsymbol{X}^1 = (4,0)^T$。 ∎

§ 3.4 对偶线性规划的应用

3.4.1 对偶单纯形法的应用

寻找正则解不是那么容易，因此在实际计算中常用到单纯

形法，但是，对于下面的几种情况，对偶单纯形法求解比较方便。

（1）如果线性规划具有下面的形式

$$\begin{aligned} &\min \boldsymbol{C}^T\boldsymbol{X} \\ s.t.& \\ &\boldsymbol{A}\boldsymbol{X} \geq \boldsymbol{b} \\ &\boldsymbol{X} \geq 0 \end{aligned} \tag{3-4-20}$$

其中$\boldsymbol{C} \geq 0$。

引入剩余变量$\boldsymbol{Y}$，原规划变为

$$\begin{gathered}\min \boldsymbol{C}^T\boldsymbol{X} \\ s.t. \\ -\boldsymbol{A}\boldsymbol{X}+\boldsymbol{Y}=-\boldsymbol{b} \\ \boldsymbol{X}\geq 0,\quad \boldsymbol{Y}\geq 0\end{gathered} \tag{3-4-21}$$

取$\boldsymbol{Y}$所在的列为基$\boldsymbol{B}=(\boldsymbol{I})$，由于$\boldsymbol{C}\geq 0$，$\boldsymbol{C}_{\boldsymbol{B}}^T=0$，所以$\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{A}-\boldsymbol{C}^T=-\boldsymbol{C}^T\leq 0$。于是

$$\boldsymbol{Y}=-\boldsymbol{b} \tag{3-4-22}$$

就是一个正则解。

（2）如果需要计算下面两个线性规划

$$(P^1)\quad \begin{aligned} &\min \boldsymbol{C}^T\boldsymbol{X} \\ s.t.\ & \\ &\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b}^1 \\ &\boldsymbol{X} \geq 0 \end{aligned}$$

$$(P^2)\quad \begin{aligned} &\min \boldsymbol{C}^T\boldsymbol{X} \\ s.t.\ & \\ &\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b}^2 \\ &\boldsymbol{X} \geq 0 \end{aligned}$$

这时采用单纯形法求得(P^1)的最优基$\boldsymbol{B}$，于是$\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{A}-\boldsymbol{C}^T \leq 0$，因此$\boldsymbol{B}$是$(P^2)$的正则基，其相应的基本解$\boldsymbol{X}^0$是$(P^2)$的正则解。

此法对多个具有相同形式的线性规划同样适用。

（3）如果已求得线性规划

$$\begin{aligned} &\min \boldsymbol{C}^T\boldsymbol{X} \\ s.t.& \\ &\boldsymbol{A}\boldsymbol{X}=\boldsymbol{b} \\ &\boldsymbol{X}\geq 0 \end{aligned} \tag{3-4-23}$$

的最优基$\boldsymbol{B} = (\boldsymbol{P}_{j1}, \boldsymbol{P}_{j2}, \cdots, \boldsymbol{P}_{jm})$，接下来需要求线性规划

$$\min \boldsymbol{C}^T \boldsymbol{X}$$

$$s.t.$$

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b} \tag{3-4-24}$$

$$\boldsymbol{a}^T \boldsymbol{X} \leq d \ \ (\text{一个新加约束})$$

$$\boldsymbol{X} \geq 0$$

其中$\boldsymbol{a}^T = (a_1, a_2, \cdots, a_n)$，$d$是常数。

这时

$$\boldsymbol{B}' = \begin{bmatrix} \boldsymbol{P}_{j1} & \boldsymbol{P}_{j2} & \cdots & \boldsymbol{P}_{jm} & 0 \\ a_{j1} & a_{j2} & \cdots & a_{jm} & 1 \end{bmatrix} \tag{3-4-25}$$

就是其正则基，其相应的基本解就是正则解。

这是因为下面的缘故。

设$\boldsymbol{B}=[b_{ij}]_{m\times m}$，则

$$|\boldsymbol{B}'|=\begin{vmatrix} b_{11} & \cdots & b_{1m} & 0 \\ & \cdots & & \\ b_{m1} & \cdots & b_{mm} & 0 \\ a_{j1} & \cdots & a_{jm} & 1 \end{vmatrix}=|\boldsymbol{B}|\neq 0.$$

于是$\boldsymbol{B}'$可作为线性规划(3-4-25)的一个基（需添加一个松弛变量）。

又记线性规划(3-4-25)的约束矩阵为

$$\begin{bmatrix} \boldsymbol{A} & 0 \\ \boldsymbol{a}^T & 1 \end{bmatrix}.$$

于是

$$
\begin{aligned}
\sigma' &= (\boldsymbol{C}_{\boldsymbol{B}}^T, 0)\boldsymbol{B}'^{-1}\begin{bmatrix} \boldsymbol{A} & 0 \\ \boldsymbol{a}^T & 1 \end{bmatrix} - (\boldsymbol{C}^T, 0) \\
&= (\boldsymbol{C}_{\boldsymbol{B}}^T, 0)\begin{bmatrix} \boldsymbol{B}^{-1} & 0 \\ -\boldsymbol{a}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1} & 1 \end{bmatrix}\begin{bmatrix} \boldsymbol{A} & 0 \\ \boldsymbol{a}^T & 1 \end{bmatrix} - (\boldsymbol{C}^T, 0) \\
&= (\boldsymbol{C}_{\boldsymbol{B}}^T, 0)\begin{bmatrix} \boldsymbol{B}^{-1}\boldsymbol{A} & 0 \\ -\boldsymbol{a}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{A} + \boldsymbol{a}^T & 1 \end{bmatrix} - (\boldsymbol{C}^T, 0) \\
&= (\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{A}, 0) - (\boldsymbol{C}^T, 0) \\
&= (\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{A} - \boldsymbol{C}^T, 0) \\
&\leq 0.
\end{aligned}
$$

3.4.2 影子价格

继续考虑问题(例3.1):

(1)影子价格是对资源在实现利润最大化时的一种价格估计，这种估计是针对具体企业产品而存在的一种特殊价格;

(2)影子价格会随着企业生产任务、产品结构等因素的变化而改变;

(3)若某种资源的市场价格高于影子价格时，企业应该卖出这种资源，否则应该买进。所以影子价格是一种机会成本;

(4)因为

$$\frac{\partial(b^T W^*)}{\partial b_i} = y_i^*, \quad i = 1 \sim m, \tag{3-4-26}$$

所以y_i^*可以看作b_i每增加一个单位时目标函数的增量。因此影子价格是一种边际价格(在资源得到最优利用的情况下，增加每单位资源所能获得产品的价值)；

(5)由松弛定理得，若某种资源未得到充分利用，则该种资源的影子价格为零，若某种资源的影子价格为零，则该种资源在生产中已经消耗完毕；

(6)因为

$$\begin{aligned}\sigma_j &= C_{B^*}^T (B^*)^{-1} P_j - c_j \\ &= (W^*)^T P_j - c_j \\ &= \sum_{i=1}^{m} a_{ij} y_i^* - c_j, \quad j = 1 \sim n.\end{aligned} \tag{3-4-27}$$

这里，c_j表示第j种产品的利润(产值)，$\sum_{i=1}^m a_{ij}y_i^*$表示生产一个单位的第j种产品所消耗的各种资源的影子价格总和(隐含成本)。所以产品利润大于隐含成本时，表明生产该种产品有利，否则生产其他产品。

第四章　无约束最优化计算方法

在实际问题的数学模型中，很多是无约束极值问题，而其求解往往可归结为反复地求解一系列无约束条件下单变量函数的最优问题。因此，无约束条件下单变量函数最优化问题是解非线性优化问题的基础。

§ 4.1 下降迭代算法

4.1.1 基本思想

考虑问题：

$$\min_{\boldsymbol{X}\in\boldsymbol{R}} f(\boldsymbol{X}). \tag{4-1-1}$$

(1)首先确定目标函数$f(\boldsymbol{X})$的极小点的一个初始估计点$\boldsymbol{X}^k$，$k=0$。

(2)然后按照一定规则产生一个下降方向$\boldsymbol{P}^k$。

(3)再沿方向$\boldsymbol{P}^k$求得下一个迭代点$\boldsymbol{X}^{k+1}$，即在射线

$$\boldsymbol{X}^k + t\boldsymbol{P}^k, t > 0.$$

上确定一个新点$\boldsymbol{X}^{k+1}=\boldsymbol{X}^{k}+t_k\boldsymbol{P}^{k}$，使得$f(\boldsymbol{X}^{k+1})<f(X^{k})$，其中$t_k$称为步长因子。

(4)若满足停机条件则输出$\boldsymbol{X}^{k}$，否则$k=k+1$，转(1)。

注：

(i)按照上面的规则，一般会产生一个迭代序列，若此序列的极限是式(4-1-1)的极小点$\boldsymbol{X}^{*}$，即

$$\lim_{k\to\infty}\boldsymbol{X}^{k}=\boldsymbol{X}^{*},\ or\ \lim_{k\to\infty}\|\boldsymbol{X}^{k}-\boldsymbol{X}^{*}\|=0. \tag{4-1-2}$$

则称算法收敛到$\boldsymbol{X}^{*}$。

(ii)如果每迭代一步都使得目标函数值下降，则称为下降

算法。

(iii)在迭代中，一是要选择**下降方向$\boldsymbol{P}^k$**，二是要计算**最优步长因子$\boldsymbol{t}_k$**。

下降算法的一般格式：

算法 4.1 下降算法

步骤 1 选择初始点$\boldsymbol{X}^0$，置$k=0$。

步骤 2 按某种规则确定$\boldsymbol{P}^k$，使得$\nabla f(\boldsymbol{X}^k)^T\boldsymbol{P}^k<0$。

步骤 3 按某种规则确定t_k，使得$f(\boldsymbol{X}^k+t_k\boldsymbol{P}_k)<f(\boldsymbol{X}^k)$。

步骤 4 计算$\boldsymbol{X}^{k+1}=\boldsymbol{X}^k+t_k\boldsymbol{P}^k$。

步骤 5 判定$\boldsymbol{X}^{k+1}$是否满足终止准则，若满足则输出$\boldsymbol{X}^{k+1}$，$f(\boldsymbol{X}^{k+1})$，停机；否则置$k=k+1$，转步骤2。

4.1.2 一维搜索

如何求解步长因子t_k？即选取使得

$$f(\boldsymbol{X}+t_k\boldsymbol{P}^k)=\min_t f(\boldsymbol{X}^k+t\boldsymbol{P}^k) \tag{4-1-3}$$

这时，t_k称为最优步长。

求一元函数

$$\varphi(t)=f(\boldsymbol{X}^k+t\boldsymbol{P}^k) \tag{4-1-4}$$

极小点的迭代法称为**一维搜索**或**直线搜索**。

一维搜索的优点：

（1）在下降方向下降最多；

（2）一元函数极值的求法相对容易。

缺点：计算量大。

定理 4.1 设$f(\boldsymbol{X})$具有连续的偏导数，给定搜索方向$\boldsymbol{P}^k$，则$\varphi'(t) = \nabla f(\boldsymbol{X}^k + t\boldsymbol{P}^k)^T \boldsymbol{P}^k$。

证明 显然。 ∎

定理 4.2 设$f(\boldsymbol{X})$具有连续的偏导数，设$f(\boldsymbol{X}^{k+1}) = \min_t f(\boldsymbol{X}^k + t\boldsymbol{P}^k)$，则$\nabla f(\boldsymbol{X}^{k+1})^T \boldsymbol{P}^k = 0$。

证明 显然。 ∎

4.1.3 收敛速度

定义 4.3 (收敛阶) 设序列$\{\boldsymbol{X}^k\}$收敛于解$\boldsymbol{X}^*$，若存在常数$p \geq 0$，L，k_0，使得当k从某个k_0开始下式成立：

$$\|\boldsymbol{X}^{k+1} - \boldsymbol{X}^*\| \leq L\|\boldsymbol{X}^k - \boldsymbol{X}^*\|^p \tag{4-1-5}$$

则称$\{\boldsymbol{X}^k\}$为p阶收敛。

定义 4.4 (线性收敛) 设序列$\{\boldsymbol{X}^k\}$收敛于解$\boldsymbol{X}^*$，若存在常数k_0，L，$\theta \in (0,1)$，使得当k从某个k_0开始下式成立：

$$\|\boldsymbol{X}^{k+1} - \boldsymbol{X}^*\| \leq L\theta^k \tag{4-1-6}$$

则称$\{\boldsymbol{X}^k\}$为线性收敛。

定义 4.5 (超线性收敛) 设序列$\{\boldsymbol{X}^k\}$收敛于解$\boldsymbol{X}^*$，若任给$\beta > 0$，都存在$k_0 > 0$，使得当k从某个k_0开始下式成立：

$$\|\boldsymbol{X}^{k+1} - \boldsymbol{X}^*\| \leq \beta\|\boldsymbol{X}^k - \boldsymbol{X}^*\| \tag{4-1-7}$$

则称$\{\boldsymbol{X}^k\}$为超线性收敛。

定义 4.6 (收敛速度) 设序列$\{\boldsymbol{X}^k\}$收敛于解$\boldsymbol{X}^*$，若

$$\lim_{k\to\infty} \frac{\|\boldsymbol{X}^{k+1} - \boldsymbol{X}^*\|}{\|\boldsymbol{X}^k - \boldsymbol{X}^*\|} = \beta. \tag{4-1-8}$$

则$0 < \beta < 1$时，称$\{\boldsymbol{X}^k\}$为β线性收敛；$\beta = 0$时，称$\{\boldsymbol{X}^k\}$为

超线性收敛；$\beta = 1$时，称$\{\boldsymbol{X}^k\}$为次线性收敛；

又若

$$\lim_{k\to\infty} \frac{\|\boldsymbol{X}^{k+1} - \boldsymbol{X}^*\|}{\|\boldsymbol{X}^k - \boldsymbol{X}^*\|^p} = \beta < +\infty. \tag{4-1-9}$$

则称$\{\boldsymbol{X}^k\}$为p阶收敛，这里$p \geq 1$。

注(1) p阶收敛，则对$\forall q < p$，必有q阶收敛；

注(2) 线性与次线性收敛都是一阶收敛，反之不然；

注(3) $p > 1$，p阶收敛必为超线性收敛；

注(4) 超线性收敛不一定$p > 1$阶收敛。

例 4.7 (1)$X^k = a^{2^k}$，这里$0 < a < 1$，为二阶收敛；

(2)$X^k = k^{-2}$为1阶收敛，但不是线性收敛。

定义 4.8 (二次收敛) 一个算法用于求解具有正定矩阵的二次函数

$$f(\boldsymbol{X}) = \frac{1}{2}\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}^T\boldsymbol{X} + c$$

时，在有限步内可以达到它的极小点。

4.1.4 终止准则

对精度ε，有

$$\|\boldsymbol{X}^{k+1} - \boldsymbol{X}^k\| < \varepsilon, \quad or \quad \frac{\|\boldsymbol{X}^{k+1} - \boldsymbol{X}^k\|}{\|\boldsymbol{X}^k\|} < \varepsilon. \tag{4-1-10}$$

$$\|\boldsymbol{f}^{k+1} - \boldsymbol{f}^k\| < \varepsilon, \quad or \quad \frac{\|\boldsymbol{f}^{k+1} - \boldsymbol{f}^k\|}{\|\boldsymbol{f}^k\|} < \varepsilon. \tag{4-1-11}$$

$$\|\nabla f(\boldsymbol{X}^k)\| < \varepsilon. \tag{4-1-12}$$

§ 4.2 精确一维搜索

4.2.1 黄金分割法（*0.618*法）

黄金分割法适用于任何单峰值函数$\varphi(t)$求极小点的问题，甚至对函数可以不要求连续。

4.2.1.1 单峰函数

定义 4.9 设$\varphi : [a, b] \subset \boldsymbol{R} \to \boldsymbol{R}$，$t^*$是在$[a, b]$上的全局极小

点，如果对于$[a,b]$上的任意两点t_1, t_2，且$t_1 < t_2$都有

$$\begin{cases} \varphi(t_1) > \varphi(t_2), \ \ if \ \ t_2 \le t^* \\ \varphi(t_1) < \varphi(t_2), \ \ if \ \ t_1 \ge t^* \end{cases}$$

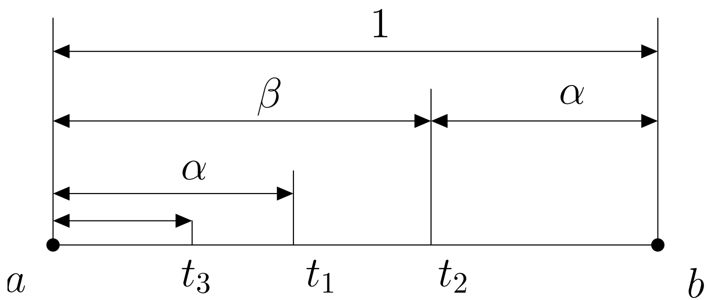
那么称$\varphi(t)$是区间$[a,b]$上的单峰函数。若$t_1 < t^* < t_2$，称$[t_1, t_2]$为搜索区间。

性质 4.10 设$[a,b]$是单峰函数$\varphi(t)$极小点的一个搜索区间，在$[a,b]$上任取两点t_1, t_2且$t_1 < t_2$。若$\varphi(t_1) < \varphi(t_2)$，则$[a, t_2]$是$\varphi(t)$极小点的一个搜索区间；若$\varphi(t_1) > \varphi(t_2)$，则$[t_1, b]$是$\varphi(t)$极小点的一个搜索区间。

4.2.1.2 基本思想

在搜索区间$[a,b]$内适当插入两点t_1,t_2，将$[a,b]$分成三段，通过比较这两点的函数值，然后由单峰函数的性质，就可以删去最左端或者最右端的一段，这算一次迭代。然后在留下来的区间上再插入一点，就可以重复上述过程，如此下去，可将区间无限缩小。

问题是：每次迭代中如何插入t_1,t_2的位置，才能使得在函数值计算同样多的条件下，区间缩短得最快？下面仅做简单直观分析。


1
β
α
α
a
t3
t1
t2
b

设区间的长为1，如下图，在与a相距分别为α和β的点插入t_1, t_2，为确定α和β，规定：

（1）要求插入的两点在搜索区间中是对称的。因此无论删除哪一端，留下的总是长为β的区间。于是$\alpha + \beta = 1$。

（2）保证了每次迭代都以同一个的比率缩短区间。不妨设去掉$[t_2, b]$，在留下的区间$[a, t_2]$里在插入新的一个点t_3，使得t_3, t_1在$[a, t_2]$中的位置与t_1, t_1 在$[a, b]$的位置有相同的比例。这样就有

$$\frac{\overline{at_1}}{\overline{at_2}} = \frac{\overline{at_2}}{\overline{ab}},$$

即$\beta^2=\alpha$。因此就有

$$\begin{aligned} \beta &= \frac{\sqrt{5}-1}{2} \approx 0.618 \\ \alpha &= \frac{3-\sqrt{5}}{2} \approx 0.382 \end{aligned} \tag{4-2-13}$$

于是有

$$\begin{aligned} t_2 &= a + \overline{at_2} \\ &= a + \overline{ab}\frac{\overline{at_1}}{\overline{at_2}} \\ &= a + (b-a)\frac{\alpha}{\beta} \\ &= a + \beta(b-a), \end{aligned} \tag{4-2-14}$$

$$
\begin{aligned}
t_1 &= a + \overline{at_1} \\
&= a + (\overline{t_2b}) \\
&= a + [b - a - \beta(b - a)] \\
&= a + (1 - \beta)(b - a) \\
&= a + \alpha(b - a).
\end{aligned} \tag{4-2-15}
$$

4.2.1.3 算法分析

$$
[a^{n+1}, b^{n+1}] \subset [a^n, b^n] \tag{4-2-16}
$$

$$b^n - a^n = \beta^n \cdot (b-a) \to 0, \quad n \to \infty. \tag{4-2-17}$$

4.2.2 *Fibonacci*法

4.2.2.1 基本思想

若不要求每次迭代使得区间的收缩比不变，而希望在试验点个数相同的前提下，找出一种选取试验点的最佳策略，使得最终的区间长度最小。

如果规定试验点的个数为n，最终区间长度为1，那么问题就是如何选取这n个点，使得原始的区间长度最大？

算法 4.2 黄金分割算法

步骤 1 确定$\varphi(t)$的初始搜索区间$[a, b]$。

步骤 2 计算$t_2 = a + \beta(b - a)$，$\varphi_2 = \varphi(t_2)$。

步骤 3 计算$t_1 = a + \alpha(b - a)$，$\varphi_1 = \varphi(t_1)$，。

步骤 4 若$t_1 - t_2 \leq \varepsilon$，停机；否则，转步骤5。

步骤 5 若$\varphi(t_1) \leq \varphi(t_2)$，则

$$b = t_2, \quad t_2 = t_1, \quad \varphi(t_2) = \varphi(t_1), \quad t_1 = a + \alpha(b - a), \quad \varphi_1 = \varphi(t_1);$$

否则，

$$a = t_1, \quad t_1 = t_2, \quad \varphi(t_1) = \varphi(t_2), \quad t_2 = a + \beta(b - a), \quad \varphi_2 = \varphi(t_2).$$

然后转步骤4。

设L_n表示试验点个数为n，最终区间长度为1时原始区间$[a,b]$的最大长度。

4.2.2.2 算法过程

令$a < x_1 < x_2 < b$，则

$$L_n \leq L_{n-2} + L_{n-1}. \tag{4-2-18}$$

显然，$L_0 = L_1 = 1$。

因此

$$L_n = L_{n-2} + L_{n-1}, \tag{4-2-19}$$

$$L_0 = L_1 = 1. \tag{4-2-20}$$

因为差分方程$L_n = L_{n-2} + L_{n-1}$的通解有如下形式：

$$L_n = Ar_1^n + Br_2^n. \tag{4-2-21}$$

所以

$$L_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right). \tag{4-2-22}$$

因为

$$\frac{L_{n-2}}{L_n} + \frac{L_{n-1}}{L_n} = 1, \tag{4-2-23}$$

所以如下选择n个试验点，

$$x_1 = a + \frac{L_{n-2}}{L_n}(b-a), \tag{4-2-24}$$

$$x_2 = a + \frac{L_{n-1}}{L_n}(b-a). \tag{4-2-25}$$

后续过程同黄金分割法。

4.2.2.3 算法分析

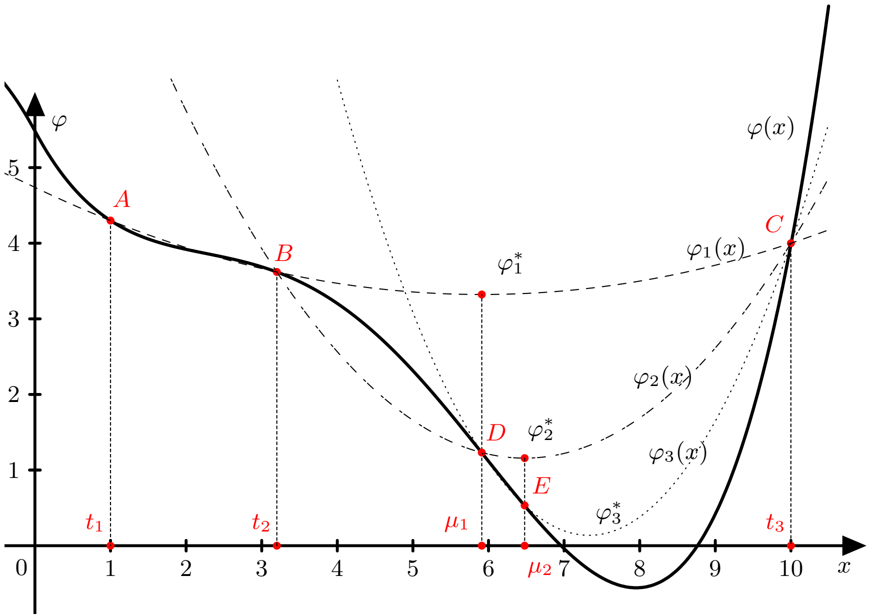
由式子(4-2-22)易得，

$$\lim_{n\to\infty} \frac{L_{n-1}}{L_n} = \frac{\sqrt{5}-1}{2} \approx 0.618. \tag{4-2-26}$$

4.2.3 二次插值法（抛物线插值法）

4.2.3.1 基本思想

利用$\varphi(t)$在某些点的信息去构造一个插值多项式$P(t)$，用$P(t)$去拟合$\varphi(t)$，然后求出$P(t)$的极小点μ，以μ作为t^*的估计值。通常取$P(t)$为二次或三次多项式，即二次或三次插值法。


φ
$\varphi(x)$
$\varphi_1(x)$
$\varphi_2(x)$
$\varphi_3(x)$
φ_1^*
φ_2^*
φ_3^*
A
B
C
D
E
t_1
t_2
t_3
μ_1
μ_2
x
0
1
2
3
4
5
6
7
8
9
10

4.2.3.2 三点二次插值法

设函数在三点t_1, t_2, t_3的函数值为$\varphi_1, \varphi_2, \varphi_3$（$t_1 < t_2 < t_3$），为了保证在$(t_1, t_3)$内存在函数$\varphi(t)$的一个极小点，在选取初始点时，要求（两边高中间低）

$$\varphi(t_1) > \varphi(t_2), \quad \varphi(t_3) > \varphi(t_2).$$

通过点(t_1, φ_1)，(t_2, φ_2)，(t_3, φ_3)作一条二次插值多项式（抛物线）$P(t)$，并且认为这条抛物线在区间(t_1, t_3)上近似于曲线$\varphi(t)$。

设过三点的抛物线为

$$P(t_i) = a_0 + a_1 t_i + a_2 t_i^2 = \varphi_i, \quad a_2 \neq 0, \quad i = 1, 2, 3.$$

从而得到

$$\begin{aligned} P(t) = {} & \varphi_1 \frac{(t-t_2)(t-t_3)}{(t_1-t_2)(t_1-t_3)} + \varphi_2 \frac{(t-t_1)(t-t_3)}{(t_2-t_1)(t_2-t_3)} \\ & + \varphi_3 \frac{(t-t_1)(t-t_2)}{(t_3-t_1)(t_3-t_2)}. \end{aligned} \tag{4-2-27}$$

令$P'(t) = 0$，得

$$\mu = \frac{\varphi_1(t_2^2 - t_3^2) + \varphi_2(t_3^2 - t_1^2) + \varphi_3(t_1^2 - t_2^2)}{2\big((t_2 - t_3)\varphi_1 + (t_3 - t_1)\varphi_2 + (t_1 - t_2)\varphi_3\big)}. \tag{4-2-28}$$

点μ就是$\varphi(t)$的极小点的一次近似。然后再在四个点中找出

相邻且满足两边高中间低的三点。然后对着三点作二次抛物线，如此反复。

其它形式 如果知道一点t_1的函数值$\varphi(t_1)$和导数值$\varphi'(t_1) < 0$以及另一点t_2的函数值$\varphi(t_2)$，做二次多项式$P(t)$，使得下面条件成立

$$P(t_1) = \varphi(t_1), \quad P'(t_1) = \varphi'(t_1), \quad P(t_2) = \varphi(t_2).$$

为了保证二次插值$p(t)$有极小点，要求

$$\varphi(t_2) > \varphi_1 + \varphi'(t_1)(t_2 - t_1).$$

得

$$P(t)=\frac{\varphi_2-\varphi_1-\varphi_1'(t_2-t_1)}{(t_2-t_1)^2}(t-t_1)^2+\varphi_1'(t-t_1)+\varphi_1 \quad (4\text{-}2\text{-}29)$$

令$P'(t)=0$，得

$$\mu=t_1-\frac{\varphi_1'(t_2-t_1)^2}{2[\varphi_2-\varphi_1-\varphi_1'(t_2-t_1)]} \quad (4\text{-}2\text{-}30)$$

4.2.4 两点三次插值法

4.2.4.1 基本思想

取$P(t)$为三次多项式来拟合$\varphi(t)$，然后用其极小点作为$\varphi(t)$极小点的近似值。

4.2.4.2 三次多项式

设两点t_1, t_2，为了保证$\varphi(t)$在搜索区间(t_1, t_2)内有极小点，假定$\varphi'_1 < 0 < \varphi'_2$。

设

$$P(t) = A(t-t_1)^3 + B(t-t_1)^2 + C(t-t_1) + D.$$

由$P(t_i) = \varphi(t_i)$，$P'(t_i) = \varphi'(t_i)$，$i = 1, 2$，得

$$\mu = t_1 + (t_2 - t_1)\left(1 - \frac{\varphi'_2 + \omega + \kappa}{\varphi'_2 - \varphi'_1 + 2\omega}\right). \tag{4-2-31}$$

其中

$$\kappa = \frac{2(\varphi_2 - \varphi_1)}{t_2 - t_1} - \varphi_1' - \varphi_2',$$

$$\omega = sign(t_2 - t_1)\sqrt{\kappa^2 - \varphi_1'\varphi_2'}.$$

§ 4.3 非精确一维搜索

精确一维搜索往往需要花费大量计算量，导致整个算法不是十分有效。另外，在很多算法如牛顿算法和拟牛顿算法，其收敛速度也并不依赖于精确一维搜索。因此，另一种变通的方法是在每次一维搜索过程中，保证目标函数都有满意的下降就够了，这就是所谓的不精确一维搜索。

算法 4.3 两点三次插值算法

步骤 1 置初始步长α及精度ε。

步骤 2 $t_1 = 0$，若$|\varphi_1'| \le \varepsilon$，则停止计算。

步骤 3 若$\varphi_1' > 0$，则$\alpha = -|\alpha|$；否则$\alpha = |\alpha|$。

步骤 4 $t_2 = t_1 + \alpha$，若$|\varphi_2'| \le \varepsilon$，则停止计算。

步骤 5 若$\varphi_1'\varphi_2' > 0$，则$\alpha = 2\alpha$，$t_1 = t_2$，$\varphi_1' = \varphi_2'$，然后转步骤4。

步骤 6 计算κ，ω，μ。

步骤 7 若$\varphi' < \varepsilon$，则停止计算，打印$t^* = \mu$；否则$\alpha = \frac{\alpha}{10}$，$t_1 = \mu$，$\varphi_1 = \varphi$，$\varphi_1' = \varphi'$，然后转步骤3 。

4.3.1 *Goldstein*准则

$$\begin{cases} (i) \quad f(X_k+\alpha_k P_k) \le f(X_k) + \rho\alpha_k \nabla f(X_k)^T P_k; \\ (ii) \quad f(X_k+\alpha_k P_k) \ge f(X_k) + (1-\rho)\alpha_k \nabla f(X_k)^T P_k. \end{cases} \tag{4-3-32}$$

这里，$0 < \rho < \frac{1}{2}$。

$Goldstein$准则的直观意义是：避免α取在区间的两个端点附近，因为取在两端点附近都会导致目标函数的改进量不大。若记$\varphi(\alpha) = f(X_k + \alpha P_k)$，那么$Goldstein$准则可改写

为：

$$\begin{cases} \varphi(\alpha_k) \leq \varphi(0) + \rho\alpha_k\varphi'(0); \\ \varphi(\alpha_k) \geq \varphi(0) + (1-\rho)\alpha_k\varphi'(0). \end{cases} \tag{4-3-33}$$

注意到$\varphi'(0) < 0$，因此当$\rho > \frac{1}{2}$时，上面两个不等式互相矛盾。由此可见，$\rho < \frac{1}{2}$ 的要求是自然的。ρ一旦给定，图中两条直线就同时给定，可接受区间也随之确定，而且ρ越接近于0，可接受区间越大，而ρ越趋近于$\frac{1}{2}$，则可接受区间越小。

4.3.2 *Wolfe*准则

$Goldstein$准则有时候会将最佳步长因子排斥在可接受区间之外。为此，$Wolfe$给出了一个简单的替代准则：

$$\begin{cases} (i) \ \ f(X_k+\alpha_k P_k) \le f(X_k) + \rho\alpha_k \nabla f(X_k)^T P_k; \\ (ii) \ \ \nabla f(X_k+\alpha_k P_k)^T P_k \ge \sigma \nabla f(X_k)^T P_k, \\ \qquad \text{亦即} \varphi'(\alpha_k) \ge \sigma\varphi'(0). \end{cases} \tag{4-3-34}$$

注：(1) 准则(ii)的几何解释：在可接受点处的切线斜率大于或等于初始斜率的σ倍。由于$\sigma \in (\rho, 1)$，因而接收点处的切线更平坦些；

(2) 可以证明：当$\rho < \sigma < 1$时，满足$Wolfe$准则的可接受步长因子是存在的；

(3) 从另一个观点看，

$$\nabla f(X_k + \alpha_k P_k)^T P_k \geq \sigma \nabla f(X_k)^T P_k \tag{4-3-35}$$

是精确一维搜索满足的正交条件$\nabla f(X_{k+1})^T P_k = 0$的某种近似。

但这种近似，即使$\sigma \to 0$，也不能导致精确一维搜索。若将$Wolfe$准则中第(ii)式换为

$$\left| \nabla f(X_k + \alpha_k P_k)^T P_k \right| \leq -\sigma \nabla f(X_k)^T P_k, \tag{4-3-36}$$

则当$\sigma \to 0$时，接近精确搜索准则。而且σ取得越小，越接近精确搜索，工作量也越大。应该指出的是不精确搜索，不要求过小的σ，应用上通常取$\rho = 0.1$，$\sigma \in [0.6, 0.8]$。

4.3.3 *Armijo*准侧

给定$\beta \in (0,1)$，$\rho \in (0, \frac{1}{2})$，$\tau > 0$，设$m_k$是使得下述不等式

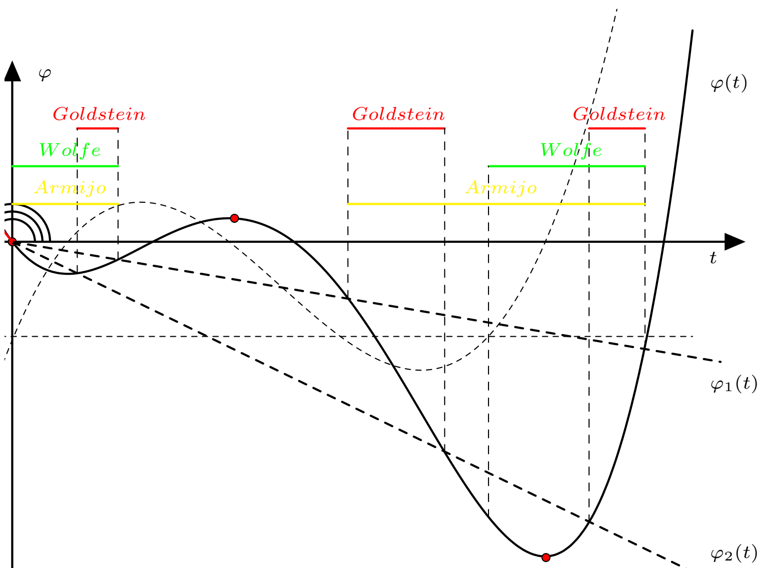
$$f(X_k + \beta^{m_k}\tau P_k) \le f(X_k) + \rho\beta^{m_k}\tau {g_k}^T P_k \tag{4-3-37}$$

成立的最小非负整数。

若令$\alpha_k = \beta^{m_k}\tau$，其就是$Goldstein$准则中的第一个准则

$$f(X_k + \alpha P_k) \leq f(X_k) + \rho\alpha {g_k}^T P_k, \tag{4-3-38}$$

即$\varphi(\alpha_k) \leq \varphi(0) + \rho\alpha_k\varphi'(0)$。


φ
Goldstein
Goldstein
Goldstein
Wolfe
Wolfe
Armijo
Armijo
$\varphi(t)$
t
$\varphi_1(t)$
$\varphi_2(t)$

4.3.4 收敛性定理

为了保证算法的下降性，要求每次搜索方向P_k与其梯度方向$-g_k = -\nabla f(X_k)$成锐角。并且要求其夹角θ_k满足：$\theta_k \leq \frac{\pi}{2} - \mu$，$\forall k\ (\mu > 0)$。

采用不精确一维搜索的一般下降算法（模式算法）

算法 4.4 模式算法

步骤 1 给出初始点$X_0 \in R^n$，允许误差$0 \leq \varepsilon < 1$，置$k := 0$;

步骤 2 若$\|\nabla f(X_k)\| \leq \varepsilon$，算法停止，$X_k \approx X^*$；否则求出满足$\nabla f(X_k)^T P_k < 0$的下降方向$P_k$;

步骤 3 求出步长因子α_k，使其满足$Goldstein$准则或$Wolfe$准则;

步骤 4 令$X_{k+1} = X_k + \alpha_k P_k$，$k := k + 1$，转步骤(2)。

下面给出基于这些不精确一维搜索的一般下降算法的总体收敛性定理。

定理 4.11 若上述算法中，步长因子α_k满足$Goldstein$准则或$Wolfe$准则，且对$\forall k$，有

$$\theta_k \leq \frac{\pi}{2} - \mu, \tag{4-3-39}$$

若$\nabla f(X)$在水平集$L = \{X | f(X) < f(X_0)\}$上一致连续。那么，或者对某个$k$，有$\nabla f(X_k) = 0$，或者$f(X_k) \to -\infty$，或者$\nabla f(X_k) \to 0$。

定理 4.12 设函数$f(X)$连续可微，$\nabla f(X)$满足$Lipschitz$条

件：

$$\|\nabla f(X) - \nabla f(Y)\| \leq M \|X - Y\|. \tag{4-3-40}$$

又设算法采用$Wolfe$准则，且P_k与$(-\nabla f(X_k))$的夹角θ_k满足

$$\theta_k \leq \frac{\pi}{2} - \mu, \quad (0 < \mu < \frac{\pi}{2}) \tag{4-3-41}$$

那么由算法产生的点列$\{X_k\}$，或者对某个k，有$\nabla f(X_k) = 0$，或者$f(X_k) \to -\infty$，或者$\nabla f(X_k) \to 0$。

下面定理给出不精确一维搜索条件下，每步迭代目标函数下降量的估计式。

定理 4.13 设α_k满足不精确一维搜索准则中第一条，若函数

还满足:

$$m\|Y\|^2 \leq Y^T \nabla^2 f(X) Y \leq M\|Y\|^2, \tag{4-3-42}$$

$$(Y-Z)^T (\nabla f(Y) - \nabla f(Z)) \geq \eta \|Y-Z\|^2. \tag{4-3-43}$$

则必有

$$f(X_k) - f(X_k + \alpha_k P_k) \geq \frac{\rho\eta}{1+\sqrt{M/m}} \|\alpha_k P_k\|^2. \tag{4-3-44}$$

§ 4.4 最速下降法

4.4.1 基本思想

求解

$$\min_{\boldsymbol{X}\in\boldsymbol{R}^n} f(\boldsymbol{X}) \tag{4-4-45}$$

假设上式中目标函数具有一阶连续偏导数。且具有极小点$\boldsymbol{X}^*$，从第k个迭代点$\boldsymbol{X}^k$出发，沿最速下降方向$\boldsymbol{P}^k = -\nabla f(\boldsymbol{X}^k)$搜索，即在射线上$\boldsymbol{X} = \boldsymbol{X}^k + t\boldsymbol{P}^k$作直线搜索，确

定最优步长t_k，使得

$$f(\boldsymbol{X}^k + t_k \nabla f(\boldsymbol{X}^k)) = \min_t f(\boldsymbol{X}^k + t\nabla f(\boldsymbol{X}^k)). \tag{4-4-46}$$

令

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - t_k \nabla f(\boldsymbol{X}^k). \tag{4-4-47}$$

因此得到序列$\{\boldsymbol{X}^k\}$，当满足一定条件时，该序列收敛于$f(\boldsymbol{X})$的极小点$\boldsymbol{X}^*$。以上式为迭代公式的算法为**最速下降算法**。

算法 4.5 最速下降算法

步骤 1 选择初始点$\boldsymbol{X}^0$，计算f^0，$\boldsymbol{g}^0$，置$k=0$。

步骤 2 直线搜索$t_k = arg\min f(\boldsymbol{X}^k - tg(\boldsymbol{X}^k))$，$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - t_k\nabla f(\boldsymbol{X}^k)$，计算$f^{k+1}$，$\boldsymbol{g}^{k+1}$。

步骤 3 若$\|\boldsymbol{g}^{k+1}\| \leq \varepsilon$，则停机，并输出$\boldsymbol{X}^{k+1}$，$f^{k+1}$；否则$k = k+1$，转步骤2。

4.4.2 最速下降法

4.4.3 收敛性

定理 4.14 设

（1）$f(\boldsymbol{X}) \in C^1$；

（2）水平集$\boldsymbol{C}_0=\{\boldsymbol{X}|f(\boldsymbol{X})\leq f(\boldsymbol{X}^0)\}$有界，

则最速下降法或者在有限步迭代后停止（由终止条件得为驻点）；或者得到点列$\{\boldsymbol{X}^k\}$，它的任何极限点都是$f(\boldsymbol{X})$的驻点。

证明 假设$\boldsymbol{g}^*\neq 0$，设子列

$$\boldsymbol{X}^{k_i}\to\boldsymbol{X}^*,\quad \boldsymbol{X}^{k_i+1}\to\boldsymbol{X}^{**},\quad i\to\infty.$$

由算法知，$\{f(\boldsymbol{X}^k)\}$单调减少，$\boldsymbol{X}^k\in\boldsymbol{C}_0(k=0,1,2,\cdots)$。

所以

$$f(\boldsymbol{X}^{k_{i+1}}) \leq f(\boldsymbol{X}^{k_i+1}) \leq f(\boldsymbol{X}^{k_i}). \tag{4-4-48}$$

因而

$$f(\boldsymbol{X}^{*}) = f(\boldsymbol{X}^{**}). \tag{4-4-49}$$

但$\boldsymbol{g}^* \neq 0$，故存在$t^* > 0$（充分小），有

$$\begin{aligned} f(\boldsymbol{X}^* - t\boldsymbol{g}^*) &= f(\boldsymbol{X}^*) - t\nabla f(\boldsymbol{X}^*)^T \boldsymbol{g}^* + o(t\|\boldsymbol{g}^*\|) \\ &= f(\boldsymbol{X}^*) - t(\boldsymbol{g}^*)^T \boldsymbol{g}^* + o(t\|\boldsymbol{g}^*\|) \\ &= f(\boldsymbol{X}^*) - t\|\boldsymbol{g}^*\|^2 + o(t\|\boldsymbol{g}^*\|) \end{aligned} \tag{4-4-50}$$

$$\Longrightarrow$$

$$f(\boldsymbol{X}^* - t^*\boldsymbol{g}^*) - f(\boldsymbol{X}^*) < 0$$

又因为

$$f(\boldsymbol{X}^{k_i+1}) = \min f(\boldsymbol{X}^{k_i} - t_{k_i}\boldsymbol{g}^{k_i}) \leq f(\boldsymbol{X}^{k_i} - t^*\boldsymbol{g}^{k_i}). \tag{4-4-51}$$

所以

$$f(\boldsymbol{X}^{**}) \leq f(\boldsymbol{X}^{*} - t^{*}\boldsymbol{g}^{*}) < f(\boldsymbol{X}^{*}). \tag{4-4-52}$$

矛盾！ ∎

4.4.4 最优步长

若$f(\boldsymbol{X})$具有二阶连续偏导数，由$Taylor$公式得

$$\begin{aligned} f[\boldsymbol{X}^{k} - t\nabla f(\boldsymbol{X}^{k})] \approx f(\boldsymbol{X}^{k}) &- t\nabla f(\boldsymbol{X}^{k})^{T}\nabla f(\boldsymbol{X}^{k}) \\ &+ \frac{1}{2}t^{2}\nabla f(\boldsymbol{X}^{k})^{T}\boldsymbol{H}(\boldsymbol{X}^{k})\nabla f(\boldsymbol{X}^{k}). \end{aligned} \tag{4-4-53}$$

其中$\boldsymbol{H}(\boldsymbol{X}^{k})$是$f(\boldsymbol{X})$在点$\boldsymbol{X}^{k}$的$Hesse$矩阵。

令

$$\frac{df[\boldsymbol{X}^k - t\nabla f(\boldsymbol{X}^k)]}{dt} = -\nabla f(\boldsymbol{X}^k)^T \nabla f(\boldsymbol{X}^k) + t\nabla f(\boldsymbol{X}^k)^T \boldsymbol{H}(\boldsymbol{X}^k)\nabla f(\boldsymbol{X}^k) = 0. \tag{4-4-54}$$

可得最优步长

$$t_k = \frac{\nabla f(\boldsymbol{X}^k)^T \nabla f(\boldsymbol{X}^k)}{\nabla f(\boldsymbol{X}^k)^T \boldsymbol{H}(\boldsymbol{X}^k)\nabla f(\boldsymbol{X}^k)} = \frac{(\boldsymbol{g}^k)^T \boldsymbol{g}^k}{(\boldsymbol{g}^k)^T \boldsymbol{H}^k \boldsymbol{g}^k} \tag{4-4-55}$$

由此得到第$k+1$步的迭代点为：

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - \frac{(\boldsymbol{g}^k)^T \boldsymbol{g}^k}{(\boldsymbol{g}^k)^T \boldsymbol{H}^k \boldsymbol{g}^k}\boldsymbol{g}^k. \tag{4-4-56}$$

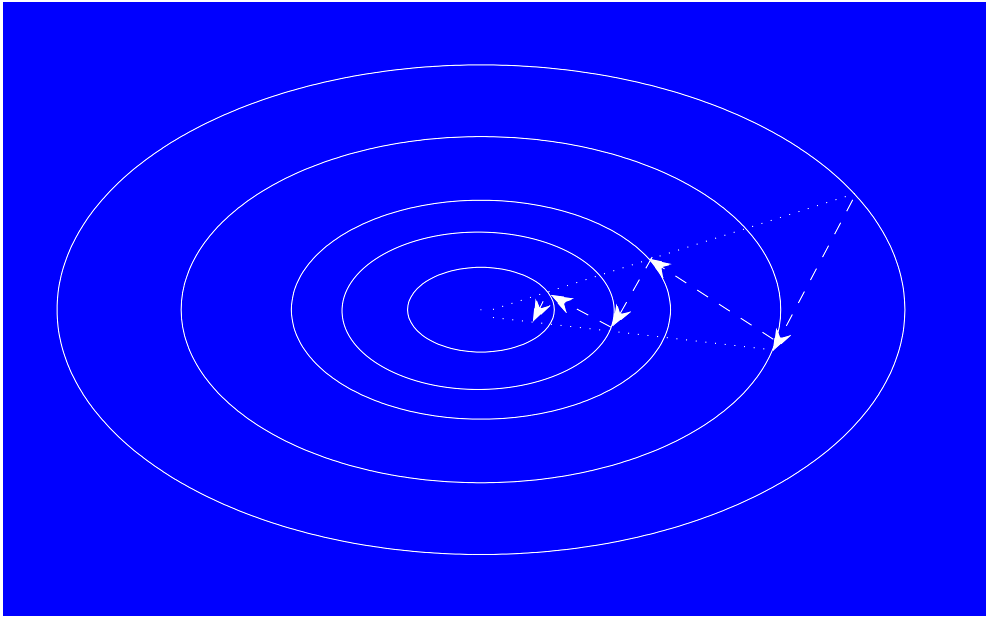
因为$\varphi'(t) = \nabla f(\boldsymbol{X} + t\boldsymbol{P})^T\boldsymbol{P}$，所以令其为零，得

$$(\boldsymbol{g}^{k+1})^T\boldsymbol{g}^k = 0. \tag{4-4-57}$$

所以最速下降法相邻两次迭代的方向互相垂直，这就影响了它的收敛速度，搜索呈锯齿状前进。因此开始搜索时，目标函数下降快，但接近极小点时，呈锯齿状搜索，目标函数变化较慢。

例 4.15 用最速下降法求解$f(\boldsymbol{X}) = (x_1 - 1)^2 + (x_2 - 1)^2$的极小点。已知$\varepsilon$。

解 $\boldsymbol{g}^k = \nabla f(\boldsymbol{X}) = [2(x_1 - 1), 2(x_2 - 1)]^T$



$$\boldsymbol{H}^k = \begin{bmatrix} 2 & 0 \\ \\ 0 & 2 \end{bmatrix}$$

$$\boldsymbol{X}^0 = (0,0)^T$$

$$t_0 = \frac{(\boldsymbol{g}^0)^T \boldsymbol{g}^0}{(\boldsymbol{g}^0)^T \boldsymbol{H}^0 \boldsymbol{g}^0} = \frac{1}{2}$$

$$\boldsymbol{X}^1 = \boldsymbol{X}^0 - t_0 \boldsymbol{g}^0 = (0,0)^T - \frac{1}{2}(-2,-2)^T = (1,1)^T$$

$$\nabla f(\boldsymbol{X}^1) = (0,0)^T\text{，}\ \|\nabla f(\boldsymbol{X}^1)\| < \varepsilon$$

$$\boldsymbol{X}^* = \boldsymbol{X}^1.$$

▮

§ 4.5 牛顿法

4.5.1 基本思想

牛顿法是用一个二次函数去近似一个目标函数，然后精确地求出这个二次函数的极小点，以它作为目标函数极小点$\boldsymbol{X}^*$的近似值。设已经迭代到$\boldsymbol{X}^k$，在点$\boldsymbol{X}^k$处对目标函数按$Taylor$公式展开，即

$$\begin{aligned} f(\boldsymbol{X}) \approx Q(\boldsymbol{X}) = f(\boldsymbol{X}^k) + \boldsymbol{g}(\boldsymbol{X}^k)^T(\boldsymbol{X} - \boldsymbol{X}^k) \\ + \frac{1}{2}(\boldsymbol{X} - \boldsymbol{X}^k)^T \boldsymbol{H}(\boldsymbol{X}^k)(\boldsymbol{X} - \boldsymbol{X}^k). \end{aligned} \tag{4-5-58}$$

其 中$Q(\boldsymbol{X})$是$\boldsymbol{X}$的 二 次 函 数 ，$\boldsymbol{g}(\boldsymbol{X}^k)$和$\boldsymbol{H}(\boldsymbol{X}^k)$分 别

为$f(\boldsymbol{X})$在$\boldsymbol{x}^k$点的梯度和$Hesse$矩阵。

令$\nabla Q(\boldsymbol{X}) = \boldsymbol{H}(\boldsymbol{X}^k)(\boldsymbol{X} - \boldsymbol{X}^k) + \boldsymbol{g}(\boldsymbol{X}^k) = 0$，得

$$\boldsymbol{H}(\boldsymbol{X}^k)(\boldsymbol{X} - \boldsymbol{X}^k) = -\boldsymbol{g}(\boldsymbol{X}^k). \tag{4-5-59}$$

若$Hesse$矩阵正定，则

$$\boldsymbol{H}^{-1}(\boldsymbol{X}^k)(\boldsymbol{X} - \boldsymbol{X}^k) = -\boldsymbol{g}(\boldsymbol{X}^k). \tag{4-5-60}$$

则由上式解出的$\boldsymbol{X} = \boldsymbol{X}^{k+1}$，就是二次函数$Q(\boldsymbol{X})$的极小点。

即

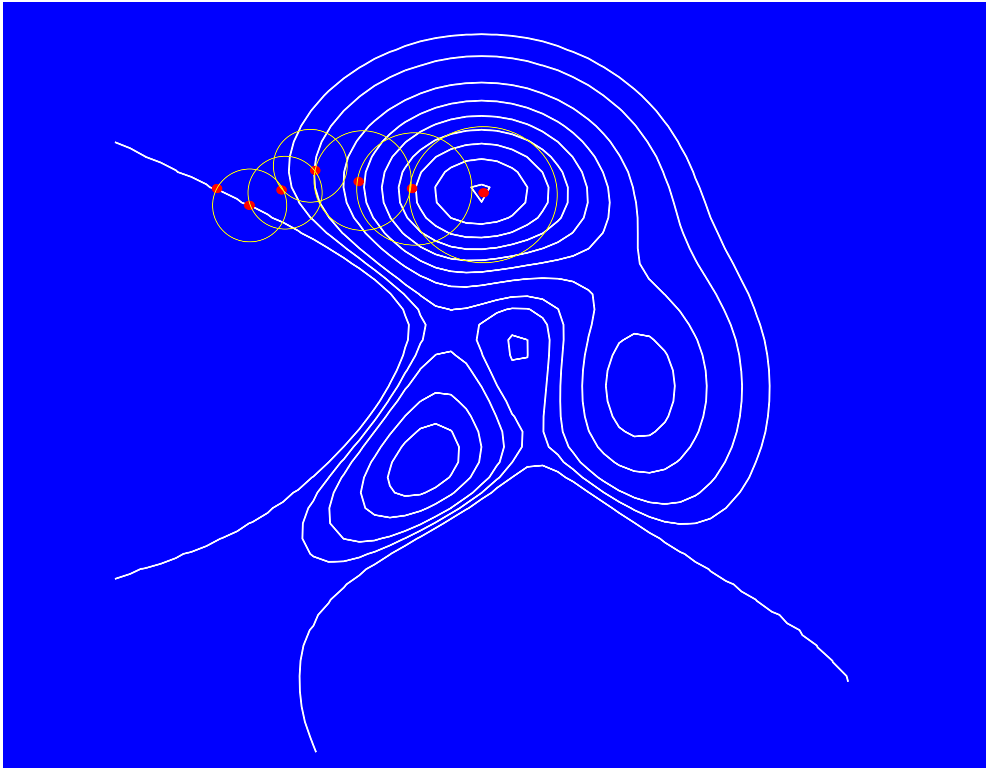
$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - \boldsymbol{H}^{-1}(\boldsymbol{X}^k)\boldsymbol{g}(\boldsymbol{X}^k). \tag{4-5-61}$$

用$\boldsymbol{X}^{k+1}$作为$f(\boldsymbol{X})$的极小点$\boldsymbol{X}^*$新的新的近似。式（4-5-61）就是**牛顿迭代公式**。

4.5.2 几何解释

当$\boldsymbol{H}(\boldsymbol{X}^k)$正定时，$Q(\boldsymbol{X}) = f(\boldsymbol{X}^k)$是一个超椭球面，在二维时$Q(\boldsymbol{X})$右边所表示的就是等高线为$f(\boldsymbol{X}^k)$的一族椭圆，$Q(\boldsymbol{X})$的极小点$\boldsymbol{X}^{k+1}$就是这个椭圆的中心，以此中心作为$f(\boldsymbol{X})$极小点$\boldsymbol{X}^*$的新的近似。

得到$f(\boldsymbol{X})$的极小点$\boldsymbol{X}^*$的近似$\boldsymbol{X}^{k+1}$，然后又在$\boldsymbol{X}^{k+1}$点对$f(\boldsymbol{X})$作二次近似，用上面同样的方法又可以得到$f(\boldsymbol{X})$的



极小点$\boldsymbol{X}^*$ 的新的近似$\boldsymbol{X}^{k+2}$。如此迭代下去，就能得到满足精度要求的$f(\boldsymbol{X})$的极小点。

4.5.3 牛顿法

例 4.16 求$f(\boldsymbol{X})=(x_1-2)^4+(x_1-2x_2)^2$的极小点。

解 $\boldsymbol{X}^0=(0,3)^T$

$\boldsymbol{g}^0=(-44,24)^T$

$$\boldsymbol{H}^0=\begin{pmatrix}50 & -4\\ -4 & 8\end{pmatrix}$$

求解$\boldsymbol{H}^0\boldsymbol{P}^0=-\boldsymbol{g}^0$，得$\boldsymbol{P}^0=(0.67,-2.67)^T$

算法 4.6 牛顿算法

步骤 1 给定目标函数$f(\boldsymbol{X})$以及梯度$\boldsymbol{g}(\boldsymbol{X})$，$Hesse$矩阵，精度$\varepsilon$。

步骤 2 选定初始点$\boldsymbol{X}^0$，计算$f^0 = f(\boldsymbol{X}^0)$，$\boldsymbol{g}^0 = \boldsymbol{g}(\boldsymbol{X}^0)$，置$k = 0$。

步骤 3 计算$\boldsymbol{H}^k = \boldsymbol{H}(\boldsymbol{X}^k)$。

步骤 4 由方程$\boldsymbol{H}^k\boldsymbol{P}^k = -\boldsymbol{g}^k$解出$\boldsymbol{P}^k$（没有求逆，考虑到计算量）。

步骤 5 计算$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k + \boldsymbol{P}^k$，$f^{k+1} = f(\boldsymbol{X}^{k+1})$，$\boldsymbol{g}^{k+1} = \boldsymbol{g}(\boldsymbol{X}^{k+1})$。

步骤 6 判断终止条件是否满足？若是，则停机并输出$\boldsymbol{X}^{k+1}$及f^{k+1}；否则$k = k + 1$，转步骤3。

表 4.1 计算结果

k	1	2	3	4	5	6	7
$\boldsymbol{X}^k$	$\begin{pmatrix}0.00\\3.00\end{pmatrix}$	$\begin{pmatrix}0.67\\0.33\end{pmatrix}$	$\begin{pmatrix}1.11\\0.56\end{pmatrix}$	$\begin{pmatrix}1.41\\0.70\end{pmatrix}$	$\begin{pmatrix}1.61\\0.80\end{pmatrix}$	$\begin{pmatrix}1.74\\0.87\end{pmatrix}$	$\begin{pmatrix}1.83\\0.91\end{pmatrix}$
$f(\boldsymbol{X}^k)$	52	3.13	0.63	0.12	0.02	0.005	0.0009

于是$\boldsymbol{X}^1 = \boldsymbol{X}^0 + \boldsymbol{P}^0 = (0.67, 0.33)^T$，$f(\boldsymbol{X}^1) = 3.13$。

按照同样的过程，下表给出了前六次的计算结果。这个问题在$\boldsymbol{X} = (2,1)^T$处达到其极小值0，所得的序列$\{\boldsymbol{X}^k\}$能够无限靠近或取得极小点。 ∎

4.5.4 优缺点及其改进

优点：当初始点离极小点较近时，牛顿迭代法公式所产生

的序列不仅能收敛到，而且收敛速度相当快，尤其是目标函数是正定二次函数时，牛顿法能够一次达到极小点，其具有二次终止性。

主要确定和改进方法：

（1）若在某点处$Hesse$矩阵奇异，这是可用最速下降方向代替牛顿方向，再做一维搜索。

（2）$Hesse$矩阵非奇异，则按照牛顿法计算法。有时候，$\boldsymbol{P}^k$方向不能保证为下降方向，即使为下降方向，也未必能保证函数值的减少。因此需要分以下几种情形来处理：

（1）若$f(\boldsymbol{X}^{k+1}) < f(\boldsymbol{X}^k)$，则按照牛顿法进行。

（2）否则，按照下面两种方式处理。

（I）当$|(\boldsymbol{g}^k)^T\boldsymbol{P}^k| \le \varepsilon\|\boldsymbol{g}^k\|\|\boldsymbol{P}^k\|$时，表明$\boldsymbol{P}^k$几乎与$\boldsymbol{g}^k$垂直，这时按照最速下降方法。

（II）当$(\boldsymbol{g}^k)^T\boldsymbol{P}^k \le -\varepsilon\|\boldsymbol{g}^k\|\|\boldsymbol{P}^k\|$时，说明牛顿方向是下降方向，此时可沿这个方向进行直线搜索。否则$(\boldsymbol{g}^k)^T\boldsymbol{P}^k > \varepsilon\|\boldsymbol{g}^k\|\|\boldsymbol{P}^k\|$时，说明牛顿方向不是下降方向，此时取反方向为搜索方向，进行一维搜索。

4.5.5 收敛性

定理 4.17 (牛顿法收敛性定理) 设

（1）$f(\boldsymbol{X}) \in C^2$;

（2）$\boldsymbol{H}(\boldsymbol{X}) > 0$（正定）；

（3）水平集$C_0 = \{\boldsymbol{X} | f(\boldsymbol{X}) \leq f(\boldsymbol{X}^0)\}$有界，

则牛顿法或者在有限步迭代终止；或者得到无穷点列$\{\boldsymbol{X}^k\}$，且有如下性质：

（1）$\{f(\boldsymbol{X}^k)\}$为严格单调下降数列；

（2）$\{\boldsymbol{X}^k\}$有唯一极限点$\boldsymbol{X}^*$，它是$f(\boldsymbol{X})$的最小点。

证明 设$\{\boldsymbol{X}^k\}$为无穷点列，$\boldsymbol{g}^k \neq 0$，因$\boldsymbol{H}^k > 0$，所以$(\boldsymbol{g}^k)^T \boldsymbol{H}^k \boldsymbol{g}^k > 0$。

（1）当$\lambda > 0(\lambda < \delta)$充分小时，有

$$\begin{aligned} f(\boldsymbol{X}^k + \lambda \boldsymbol{P}^k) &= f[\boldsymbol{X}^k - \lambda (\boldsymbol{H}^k)^{-1} \boldsymbol{g}^k] \\ &= f(\boldsymbol{X}^k) + (g^k)^T [-\lambda (\boldsymbol{H}^k)^{-1} \boldsymbol{g}^k] + o(\lambda) \\ &= f(\boldsymbol{X}^k) + \lambda \Big(-(\boldsymbol{g}^k)^T (\boldsymbol{H}^k)^{-1} \boldsymbol{g}^k + \frac{o(\lambda)}{\lambda} \Big) \end{aligned} \tag{4-5-62}$$

因为$-(\boldsymbol{g}^k)^T(\boldsymbol{H}^k)^{-1}\boldsymbol{g}^k < 0$，所以

$$f(\boldsymbol{X}^k + \lambda \boldsymbol{P}^k) < f(\boldsymbol{X}^k), \quad 0 < \lambda < \delta, \quad k = 0, 1, 2, \cdots.$$

从而存在$\lambda > 0$，使得

$$f(\boldsymbol{X}^k + \lambda_k \boldsymbol{P}^k) = \min_{\lambda>0} f(\boldsymbol{X}^k + \lambda \boldsymbol{P}^k) \leq f(\boldsymbol{X}^k + \lambda \boldsymbol{P}^k) < f(\boldsymbol{X}^k).$$

（2）因$\boldsymbol{X}^k \in \{\boldsymbol{X} | f(\boldsymbol{X}) \leq f(\boldsymbol{X}^0)\}$，故$\{\boldsymbol{X}^k\}$为有界点列，则其必有极限点。

设$\boldsymbol{x}^*$为任一极限点，且设$\lim_{i\to+\infty} \boldsymbol{X}^{k_i} = \boldsymbol{X}^*$，再由$f(\boldsymbol{X})$的连续性有$\lim_{i\to+\infty} f(\boldsymbol{X}^{k_i}) = f(\boldsymbol{X}^*)$，又因$\{\boldsymbol{X}^k\}$为为单调下降且有下界的数列，因而

$$\lim_{i\to+\infty} f(\boldsymbol{X}^{k_i}) = \lim_{i\to+\infty} f(\boldsymbol{X}^{k_i+1}) = f(\boldsymbol{X}^*).$$

假设$\boldsymbol{g}^* \neq 0$，则

$$(\boldsymbol{g}^*)^T(\boldsymbol{H}^*)^{-1}\boldsymbol{g}^* > 0. \tag{4-5-63}$$

其中$(\boldsymbol{H}^*)^{-1} > 0$，所以$\exists\delta > 0$，当$0 < \lambda < \delta$时有

$$\begin{aligned} &= f(\boldsymbol{X}^*) - \lambda(\boldsymbol{g}^*)^T(\boldsymbol{H}^*)^{-1}\boldsymbol{g}^* + o(\lambda) \\ &= f(\boldsymbol{X}^*) + \lambda\Big(- (\boldsymbol{g}^*)^T(\boldsymbol{H}^*)^{-1}\boldsymbol{g}^* + \frac{o(\lambda)}{\lambda}\Big) \end{aligned} \tag{4-5-64}$$

因此

$$f\big(\boldsymbol{X}^* + \lambda(\boldsymbol{H}^*)^{-1}\boldsymbol{g}^*\big) < f(\boldsymbol{X}^*), \quad 0 < \lambda < \delta. \tag{4-5-65}$$

又

$$f(\boldsymbol{X}^{k_i+1}) = \min_{\lambda>0} f(\boldsymbol{X}^{k_i} + \lambda \boldsymbol{P}^{k_i}) \leq f\Big(\boldsymbol{X}^{k_i} + \lambda(\boldsymbol{H}^{k_i})^{-1}\boldsymbol{g}^{k_i}\Big). \tag{4-5-66}$$

令$i \to +\infty$，则有

$$f(\boldsymbol{X}^*) \leq f\Big(\boldsymbol{X}^* + \lambda(\boldsymbol{H}^*)^{-1}\boldsymbol{g}^*\Big) < f(\boldsymbol{X}^*). \tag{4-5-67}$$

矛盾！

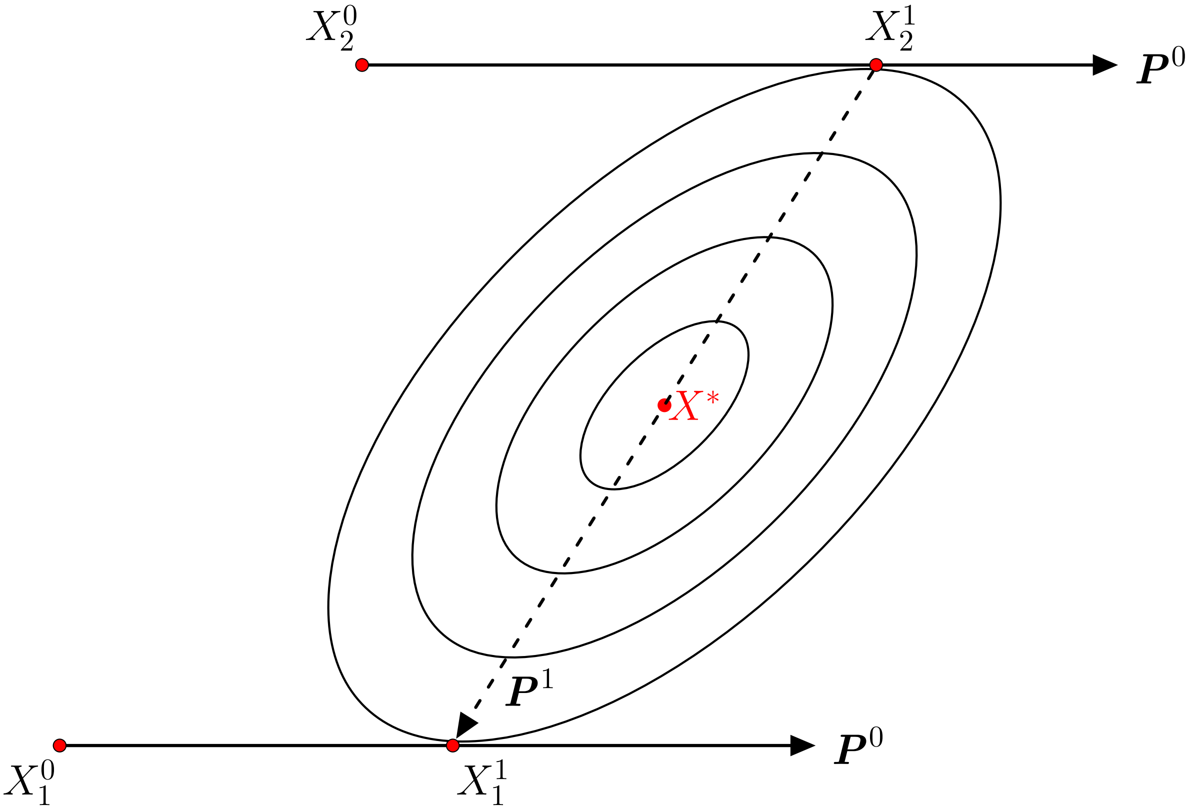
因为$\boldsymbol{H} > 0$，所以$f(\boldsymbol{X})$为严格凸函数，故驻点必为最小点。而严格凸函数的最小点必唯一，所以命题得证！ ∎

§ 4.6 共轭方向法

最速下降法有锯齿现象，收敛速度慢；而牛顿法要计算$Hesse$矩阵，计算量较大。而共轭方向法的收敛速度介于前两者之间，无需计算$Hesse$矩阵，且具有二次终止性。

$$f(\boldsymbol{X}) = \frac{1}{2}\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{b}^T\boldsymbol{X} + \boldsymbol{c}. \tag{4-6-68}$$

在这里，$\boldsymbol{Q} > 0$。


X_2^0
X_2^1
$\boldsymbol{P}^0$
X^*
$\boldsymbol{P}^1$
X_1^0
X_1^1
$\boldsymbol{P}^0$

即，从$\boldsymbol{X}_2^0$出发，以$\boldsymbol{P}^0$为方向进行精确一维搜索得到$\boldsymbol{X}_2^1$，在以此按照方向$\boldsymbol{P}^1 = \boldsymbol{X}_1^1 - \boldsymbol{X}_2^1$进行精确一维搜索便得最优解$\boldsymbol{X}^*$。

$$\boldsymbol{X}_2^0 \xrightarrow{\boldsymbol{P}^0} \boldsymbol{X}_2^1 \xrightarrow{\boldsymbol{P}^1 = \boldsymbol{X}_1^1 - \boldsymbol{X}_2^1} \boldsymbol{X}^*$$

考察$\boldsymbol{P}^1$，

$$\nabla f(\boldsymbol{X}^*) = \boldsymbol{Q}\boldsymbol{X}^* + b = 0. \tag{4-6-69}$$

即，

$$\nabla f(\boldsymbol{X}_2^1) + t_1\boldsymbol{Q}\boldsymbol{P}^1 = \boldsymbol{Q}(X^* - t_1\boldsymbol{P}^1) + b + t_1\boldsymbol{Q}\boldsymbol{P}^1 = 0. \tag{4-6-70}$$

所以，

$$(\boldsymbol{P}^0)^T\nabla f(\boldsymbol{X}_2^1)+t_1(P^0)^T\boldsymbol{Q}\boldsymbol{P}^1=0\Rightarrow(P^0)^T\boldsymbol{Q}\boldsymbol{P}^1=0. \tag{4-6-71}$$

定义 4.18 (共轭) 设$\boldsymbol{Q}$是$n\times n$对称正定矩阵，如果$\boldsymbol{X}$和$\boldsymbol{Q}Y$正交，即

$$\boldsymbol{X}^TQY=0. \tag{4-6-72}$$

则称$\boldsymbol{X}$和$\boldsymbol{Y}$关于$\boldsymbol{Q}$共轭。

推广：

定义 4.19 (共轭) 设$\boldsymbol{Q}$是$n\times n$对称正定矩阵，如果n维空间

中的非零向量组$\boldsymbol{P}^0$、$\boldsymbol{P}^1$、$\cdots$、$\boldsymbol{P}^{m-1}$两两关于Q 共轭，即

$$(\boldsymbol{P}^i)^T Q(\boldsymbol{P}^j) = 0; \quad i, j = 0 \sim n, i \neq j. \tag{4-6-73}$$

则称这组向量关于$\boldsymbol{Q}$共轭。

注意：共轭和正交的联系。

定理 4.20 如果非零向量组$\boldsymbol{P}^0$、$\boldsymbol{P}^1$、$\cdots$、$\boldsymbol{P}^{m-1}$（是）关于Q共轭的，则这m个向量必然线性无关。

证明 易得。 ∎

因为在n维空间中，个数多于n个的向量组比线性相关，所以，

推论 4.21 在R^n空间中，互相共轭的非零向量的向量个数不超过n个。

定理 4.22 (扩张子空间定理) 设

$$f(\boldsymbol{X}) = \frac{1}{2}\boldsymbol{X}^T Q X + \boldsymbol{b}^T X + c \tag{4-6-74}$$

为正定二次函数，向量组$\boldsymbol{P}^i$,$(i = 0 \sim n-1)$为Q共轭，则从任意一点$\boldsymbol{X}^0$出发，依次按照$\boldsymbol{P}^0$、$\boldsymbol{P}^1$、$\sim$、$\boldsymbol{P}^{n-1}$为搜索方向进行精确一维搜索，那么经过n后迭代点必然收敛于问题(4-6-74)的最优解X^*。

证明 因为，

$$\nabla f(\boldsymbol{X}^{k+1}) = \nabla f(\boldsymbol{X}^k) + t_k \boldsymbol{Q} P^k, \quad k = 0 \sim n-1. \tag{4-6-75}$$

所以若$\nabla f(\boldsymbol{X}^k) \neq 0$，$k = 0 \sim n-1$，则，

$$\nabla f(\boldsymbol{X}^n) = \nabla f(\boldsymbol{X}^{k+1}) + \sum_{i=k+1}^{n-1} t_i \boldsymbol{Q} P^i. \tag{4-6-76}$$

因此对$k = 0 \sim n-1$有，

$$(\boldsymbol{P}^k)^T \nabla f(\boldsymbol{X}^n) = (\boldsymbol{P}^k)^T \nabla f(\boldsymbol{X}^{k+1}) + \sum_{i=k+1}^{n-1} t_i (\boldsymbol{P}^k)^T \boldsymbol{Q} P^i = 0. \tag{4-6-77}$$

即，$\nabla f(X^n)$与$\boldsymbol{P}^0$、$\cdots$、$\boldsymbol{P}^{n-1}$正交。

则下面线性系统，

$$(\boldsymbol{P}^0, \boldsymbol{P}^1, \cdots, \boldsymbol{P}^{n-1})^T \cdot \nabla f(\boldsymbol{X}^n) = 0. \tag{4-6-78}$$

只有零解，即$\nabla f(\boldsymbol{X}^n) = 0$。又因为$f(\boldsymbol{X})$为凸函数，从而定理得证！ ∎

问题：怎样构造一组关于Q共轭的向量组？

(1)设v^i，$i = 0 \sim n-1$是一组线性无关的向量组，首先取

$$\boldsymbol{P}^0 = v^0. \tag{4-6-79}$$

设已求得$\boldsymbol{P}^i$，$i=0\sim k$，现构造$\boldsymbol{P}^{k+1}$。令

$$\boldsymbol{P}^{k+1}=v^{k+1}+\sum_{r=0}^{k}\beta_{k+1,r}\boldsymbol{P}^r.$$

这里$\beta_{k+1,r}$为待定系数，为了使$\boldsymbol{P}^{k+1}$与$\boldsymbol{P}^r$，$r=0\sim k$关于$\boldsymbol{Q}$共轭，应有

$$(\boldsymbol{P}^j)^T\boldsymbol{Q}\cdot\boldsymbol{P}^{k+1}=(\boldsymbol{P}^j)^T\boldsymbol{Q}\cdot v^{k+1}+\sum_{r=0}^{k}\beta_{k+1,r}(\boldsymbol{P}^j)^T\boldsymbol{Q}\cdot P^r=0.$$

从而得

$$\beta_{k+1,r}=-\frac{(\boldsymbol{P}^r)^T\boldsymbol{Q}\boldsymbol{v}^{k+1}}{(\boldsymbol{P}^r)^T\boldsymbol{Q}\boldsymbol{P}^r}.$$

于是，

$$\boldsymbol{P}^{k+1} = v^{k+1} - \sum_{j=0}^{k} \frac{(\boldsymbol{P}^j)^T \boldsymbol{Q} \boldsymbol{v}^{k+1}}{(\boldsymbol{P}^j)^T \boldsymbol{Q} \boldsymbol{P}^j} \boldsymbol{P}^j. \tag{4-6-80}$$

此过程称作为***Gram-Schmidt*共轭化方法**。

(2)通过负梯度来构造共轭向量，即共轭梯度法。

4.6.1 共轭梯度法

首先构造

$$\boldsymbol{P}^0 = -g^0. \tag{4-6-81}$$

对$k = 1 \sim n-1$，令

$$\boldsymbol{P}^{k+1} = -g^{k+1} + \alpha_k \boldsymbol{P}^k.$$

其中α_k为待定参数，为$\boldsymbol{P}^{k+1}$、$\boldsymbol{P}^k$关于$\boldsymbol{Q}$共轭，所以

$$(\boldsymbol{P}^{k+1})^T \cdot \boldsymbol{Q} P^k = -(g^{k+1})^T \cdot \boldsymbol{Q} P^k + \alpha_k (\boldsymbol{P}^k)^T \cdot \boldsymbol{Q} P^k = 0.$$

从而

$$\alpha_k = \frac{(g^{k+1})^T \boldsymbol{Q} P^k}{(\boldsymbol{P}^k)^T \boldsymbol{Q} P^k}.$$

即，

$$\boldsymbol{P}^{k+1} = -g^{k+1} + \frac{(g^{k+1})^T \boldsymbol{Q} P^k}{(\boldsymbol{P}^k)^T \boldsymbol{Q} P^k} \boldsymbol{P}^k. \tag{4-6-82}$$

注：一些著名的的公式

推论 4.23 (*Daniel*,1967)

$$\alpha_k = \frac{(g^{k+1})^T Q P^k}{(P^k)^T Q P^k}.$$

推论 4.24 (*Sorenson &. Wolfe*,1972)

$$\alpha_k = \frac{(g^{k+1})^T (g^{k+1} - g^k)}{(P^k)^T (g^{k+1} - g^k)}.$$

推论 4.25 (*Myers*,1972)

$$\alpha_k = \frac{(g^{k+1})^T g^{k+1}}{(P^k)^T g^{k+1}}.$$

推论 4.26 **(**$Flecher$ &. $Reeves$**,1964)**

$$\alpha_k = \frac{\|g^{k+1}\|^2}{\|g^k\|^2}$$

推论 4.27 **(**$Polyak, Polak$ &. $Ribiere$**,1969)**

$$\alpha_k = \frac{(g^{k+1})^T(g^{k+1} - g^k)}{(g^k)^T g^k}.$$

推论 4.28 **(**Dai &. $Yuan$**,1999)**

$$\alpha_k = \frac{(g^{k+1})^T P^{k+1}}{(P^k)^T g^k}.$$

例 4.29 **(**FR**共轭梯度法)** 用FR方法求解$f(X) = x_1^2 + 2x_2^2$，取初始点$X^0 = (5,5)^T$。

解 $g \triangleq \nabla f(X) = \begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix}, H \triangleq \nabla^2 f(X) = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$。

$\|g^0\|^2 > \varepsilon$,

$P^0 = -g^0 = (-10, -20)^T$,

$$\begin{aligned} X^1 &= X^0 + t_0 P_0 = X^0 - t_0 g_0 \\ &= \begin{bmatrix} 5 \\ 5 \end{bmatrix} + \frac{-(g^0)^T P^0}{(P^0)^T H P^0} \begin{bmatrix} -10 \\ -20 \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ 5 \end{bmatrix} + \frac{5}{18} \begin{bmatrix} -10 \\ -20 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \frac{20}{9} \\ \\ -\frac{5}{9} \end{bmatrix}$$

$\|g^1\| > \varepsilon$

$$\begin{aligned} P^1 &= -g^1 + \alpha_0 P_0 \\ &= \begin{bmatrix} -\frac{40}{9} \\ \\ \frac{20}{9} \end{bmatrix} + \frac{\|g^1\|^2}{\|g^0\|^2} \begin{bmatrix} -10 \\ \\ -20 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{40}{9} \\ \\ \frac{20}{9} \end{bmatrix} + \frac{4}{81} \begin{bmatrix} -10 \\ \\ -20 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} -\frac{400}{81} \\ \\ \frac{100}{81} \end{bmatrix}$$

$$\begin{aligned} X^2 &= X^1 + t_1 P_1 \\ &= \begin{bmatrix} \frac{20}{9} \\ \\ -\frac{5}{9} \end{bmatrix} + \frac{-(g^1)^T P^1}{(P^1)^T H P^1} \begin{bmatrix} -\frac{400}{81} \\ \\ \frac{100}{81} \end{bmatrix} \\ &= \begin{bmatrix} \frac{20}{9} \\ \\ -\frac{5}{9} \end{bmatrix} + \frac{9}{20} \begin{bmatrix} -\frac{400}{81} \\ \\ \frac{100}{81} \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\|g^2\|^2 = 0 < \varepsilon$，$X^* = X^2 = (0,0)^T$。 ∎

注意初始方向应为负梯度方向，否则所产生一般不是为共轭方向。

定理 4.30 对于二次正定函数$f(X) = \frac{1}{2}X^TGX + b^TX + c$，$X \in R^n$，若采用精确一维搜索，则共轭梯度法对任意一迭代步$k$，有下面的式子成立：

$$(g^k)^T P^j = 0, \qquad 0 \le j \le k-1; \tag{4-6-83}$$

$$(P^k)^T G P^j = 0, \qquad 0 \le j \le k-1; \tag{4-6-84}$$

$$(g^k)^T g^j = 0, \qquad 0 \le j \le k-1; \tag{4-6-85}$$

$$(P^k)^T g^k = -\|g^k\|^2 \tag{4-6-86}$$

证明 $P^0 = -g^0$，$(g^1)^T P^0 = 0$，$(g^1)^T g^0 = 0$。

$P^1 = -g^1 + \alpha_0 P^0$，$(P^1)^T G P^0 = 0$。

现假设k成立。

$$g^{k+1} - g^k = G(X^{k+1} - X^k) = t_k G P^k$$

$$\begin{aligned}(g^{k+1})^T g^j &= (g^k)^T g^j + t_k (P^k)^T G g^j \\ &= (g^k)^T g^j - t_k (P^k)^T G (P^j - \alpha_{j-1} P^{j-1}).\end{aligned} \tag{4-6-87}$$

$$
\begin{aligned}
(P^{k+1})^T G P^j &= (-g^{k+1} + \alpha_k P^k)^T G P^j \\
&= \frac{1}{t_j}(g^{k+1})^T (g^j - g^{j+1}) + \alpha_k (P^k)^T G P^j.
\end{aligned}
\tag{4-6-88}
$$

当$j < k$时，

$$
(P^{k+1})^T G P^j = 0, \qquad j \le k-1; \tag{4-6-89}
$$

$$
(g^{k+1})^T g^j = 0, \qquad 1 \le j \le k-1. \tag{4-6-90}
$$

即当$j < k$时对$k+1$成立。

当$j=k$时，

$$\begin{aligned}(g^{k+1})^T g^k &= (g^k)^T g^k + t_k (P^k)^T G P^k \\ &= (g^k)^T g^k - \frac{(g^k)^T g^k}{(P^k)^T G P^k} (P^k)^T G P^k \\ &= 0.\end{aligned} \tag{4-6-91}$$

$$\begin{aligned}(P^{k+1})^T G P^k &= -\frac{1}{t_k}(g^{k+1})^T g^{k+1} + \alpha_k (P^k)^T G P^k \\ &= -\frac{(g^{k+1})^T g^{k+1}}{(g^k)^T g^k}(P^k)^T G P^k + \frac{(g^{k+1})^T g^{k+1}}{(g^k)^T g^k}(P^k)^T G P^k \\ &= 0.\end{aligned} \tag{4-6-92}$$

即当$j=k$时对$k+1$成立。

最后

$$\begin{aligned}(P^{k+1})^T g^{k+1} &= (-g^{k+1}+\alpha_k P^k)^T g^{k+1} \\ &= -(g^{k+1})^T g^{k+1} \\ &= -\|g^{k+1}\|^2.\end{aligned} \tag{4-6-93}$$

■

4.6.2 变尺度法（拟牛顿法）

4.6.2.1 一般格式

我们知道，牛顿法的收敛速度很快，但其一个致命弱点就

是每次迭代都要计算目标函数的$Hesse$矩阵Q以及其逆Q^{-1}，当问题的维数n较大时，计算量迅速增加。

4.6.2.2 对称秩1公式（*SR1*法）

变尺度法的基本思想：利用在迭代过程中某些已知信息，去构造一个新的矩阵（变尺度矩阵），使得这个矩阵与$Hesse$ 矩阵近似。

对于二次函数

$$f(X) = \frac{1}{2}X^TQX + b^TX + c$$

这里$Q > 0$.

牛顿迭代公式为

$$X^{k+1} = X^k - (Q^k)^{-1} g^k. \tag{4-6-94}$$

取$(Q^k)^{-1}$某种近似H^k，将其变为如下迭代

$$X^{k+1} = X^k - H^k g^k. \tag{4-6-95}$$

更一般地

$$X^{k+1} = X^k - t_k H^k g^k. \tag{4-6-96}$$

问题：怎么构造H^k？

(1)搜索方向$P^k = -H^k g^k$应该为下降方向，即

$$(g^k)^T P^k = -(g^k)^T H^k g^k < 0.$$

不妨令

$$H^k > 0. \tag{4-6-97}$$

(2)

$$\begin{aligned} f(X^k) =& f(X^{k+1}) + \nabla f(X^{k+1})^T (X^k - X^{k+1}) \\ &+ \frac{1}{2}(X^k - X^{k+1})^T \nabla^2 f(X^{k+1})(X^k - X^{k+1}). \end{aligned} \tag{4-6-98}$$

关于X^k求导，得

$$\nabla f(X^k) = \nabla f(X^{k+1}) + \nabla^2 f(X^{k+1})(X^k - X^{k+1}). \tag{4-6-99}$$

即

$$g^k = g^{k+1} + Q^{k+1}(X^k - X^{k+1}). \tag{4-6-100}$$

因此，若Q^{k+1}正定时，有

$$(Q^{k+1})^{-1}(g^{k+1} - g^k) = X^{k+1} - X^k. \tag{4-6-101}$$

记$Y^k = g^{k+1} - g^k$，$S^k = X^{k+1} - X^k$，有

$$(Q^{k+1})^{-1}Y^k = S^k. \tag{4-6-102}$$

于是，不妨令H^{k+1}满足上式，得

$$H^{k+1}Y^k = S^k. \tag{4-6-103}$$

称式(4-6-103)为拟牛顿方程(条件).

(3)为了便于计算，令

$$H^{k+1} = H^k + \Delta H^k. \tag{4-6-104}$$

并考虑修正项ΔH^k为秩1对称矩阵，即

$$H^{k+1} - H^k = \Delta H^k = P \begin{pmatrix} \lambda & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} P^T. \tag{4-6-105}$$

令u表示P的第一列，即

$$H^{k+1} = H^k + \lambda u u^T. \tag{4-6-106}$$

将此式带入拟牛顿方程，得

$$\lambda u u^T Y^k = S^k - H^k Y^k \triangleq Z^k. \tag{4-6-107}$$

不妨令

$$u = \alpha Z^k. \tag{4-6-108}$$

将此式带入上式

$$\lambda \alpha^2 Z^k (Z^k)^T Y^k = Z^k. \tag{4-6-109}$$

上式左乘以$(Z^k)^T$，得

$$\lambda \alpha^2 = \frac{1}{(Z^k)^T Y^k} = \frac{1}{(Y^k)^T Z^k}. \tag{4-6-110}$$

于是

$$H^{k+1} = H^k + \frac{Z^k (Z^k)^T}{(Y^k)^T Z^k} \tag{4-6-111}$$

性质 4.31 在$SR1$算法中，由式(4-6-111)确定的H^i，$i = 1, 2, \cdots, k+1$。都满足拟牛顿条件，即

$$H^{k+1} Y^k = S^k.$$

性质 4.32 $SR1$算法用于$f(X) = \frac{1}{2} X^T Q X + b^T X + c$，$Q > 0$，如果

(1)$H^0 > 0$;

(2)迭代点互异;

(3)$(Z^k)^T Y^k \neq 0.$

那么，

$$(P^i)^T QP^j = 0; i, j = 0 \sim k+1, i \neq j \tag{4-6-112}$$

$$H^{k+1}QP^j = P^j; j = 0 \sim k; \tag{4-6-113}$$

注：$SR1$算法是一种共轭方向法，当然具有二次收敛性。

证明 因为

$$QS^j = Y^j, \qquad j = 0 \sim k+1.$$

$$H^{j+1}Y^j = S^j, \qquad j = 0 \sim k+1.$$

$$S^j = -t_j H^k g^k = t_j P^k, \qquad j = 0 \sim k+1.$$

所以

$$H^{j+1}QP^j = P^j, \quad j = 0 \sim k. \tag{4-6-114}$$

显然$k=0$时，有

$$\begin{aligned} H^1QP^0 &= P^0, \\ (P^1)^TQP^0 &= -(g^1)^TH^1QP^0 \\ &= -(g^1)^TP^0 \\ &= 0. \end{aligned}$$

假设$k=m-1$成立；

现在证明$k=m$也成立：

因为

$$\begin{aligned}(S^m)^T QP^j &= t_m (P^m)^T QP^j \\ &= 0, \quad j=0\sim m-1;\end{aligned} \tag{4-6-115}$$

$$\begin{aligned}(Y^m)^T H^m QP^j &= (Y^m)^T P^j \\ &= (S^m)^T QP^j \\ &= 0, \quad j=0\sim m-1;\end{aligned} \tag{4-6-116}$$

所以

$$H^{m+1}QP^j = H^mQP^j = P^j, \quad j = 0 \sim m-1; \tag{4-6-117}$$

又

$$H^{m+1}QP^m = P^m. \tag{4-6-118}$$

最后

$$\begin{aligned}(P^{m+1})^TQP^j &= -(g^{m+1})^TH^{m+1}QP^j \\ &= -(g^{m+1})^TP^j = 0. \quad j = 0 \sim m.\end{aligned} \tag{4-6-119}$$

∎

定理 4.33 (对称秩一校正性质定理) 设$S^0, S^1, \cdots, S^{n-1}$线

性无关，那么对于二次函数，$SR1$方法至多$n+1$步终止，即$H^n = Q^{-1}$。

证明 设二次函数$f(X)$的$Hesse$矩阵Q是正定的,在所有二次终止性证明中我们都利用

$$Y^k = QS^k$$

首先，我们用归纳法证明

$$H^iY^j = S^j, \quad j = 0, 1, \cdots, i-1.$$

对于$i = 1$，直接由$SR1$公式可知上式成立。今假定上式对

于$i \geq 1$成立, 我们证明它对于$i+1$也成立。我们有

$$H^{i+1}Y^j = H^iY^j + \frac{(S^i - H^iY^i)(S^i - H^iY^i)^TY^j}{(S^i - H^iY^i)^TY^i},$$

当$j < i$时，由归纳法假设有

$$\begin{aligned}
(S^i - H^iY^i)^TY^j &= (S^i)^TY^j - (Y^i)^TH^iY^j \\
&= (S^i)^TY_j - (Y^i)^TS^j \\
&= (S^i)^TGS^j - (S^i)^TGS^j \\
&= 0,
\end{aligned}$$

故

$$H^{i+1}Y^j = H^iY^j = S^j, \quad j < i.$$

当$j = i$时，有

$$H^{i+1}Y^i = S^i.$$

从而，

$$S^j = H^nY^j = H^nQS^j, \quad j = 0, 1, \cdots, n-1.$$

由于P^j线性无关，故$H^nQ = I$，即$H^n = Q^{-1}$。 ∎

注： 对称秩一校正公式具有:

(1)简单、对称迭代形式。

$$H^{k+1} = H^k + \frac{(S^k - H^kY^k)(S^k - H^kY^k)^T}{(Y^k)^T(S^k - H^kY^k)}, \tag{4-6-120}$$

这里，

$$S^k = X^{k+1} - X^k, \quad Y^k = g^{k+1} - g^k. \tag{4-6-121}$$

(2)满足拟牛顿条件.

$$H^{k+1}Y^k = S^k. \tag{4-6-122}$$

然而$SR1$有两个明显的缺点：

(3)$SR1$公式可能出现

$$S^k - H^kY^k = 0 \tag{4-6-123}$$

$$(Y^k)^T(S^k - H^kY^k) = 0 \tag{4-6-124}$$

(4)不能保证$P^k = -H^kg^k$是下降方向，即H^k不正定。

4.6.2.3 对称秩2公式（*DFP*算法）

$SR1$公式的推导是假设变尺度矩阵修正项ΔH^k的为对称矩阵，且$rank(\Delta H^k) = 1$。

因此ΔH^k具有下面的简单形式：

$$Q\begin{pmatrix} \lambda & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}Q^T = \lambda uu^T, \lambda(\neq 0) \in R, u \in R^n. \tag{4-6-125}$$

若让ΔH^k的形式稍复杂些，令$rank(\Delta H^k) = 2$，那么会得到怎样的迭代公式呢？得到的公式是否不会具有$SR1$的缺点。

这就是下面要讲的具有对称秩二校正的DFP算法。

DFP算法首先是由$Davidon$（1959年）提出来的，后来，$Fletcher$和$Powell$（1963年）对$Davidon$的方法作了改进。这种算法是无约束最优化方法最有效的方法之一。其主要的特点就是具有正定遗传性。

(1)对称秩二校正公式推导

根据前面的分析，若$rank(\Delta H^k) = 2$，则

$$\Delta H^k = P \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} P^T = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T. \quad (4\text{-}6\text{-}126)$$

这里，$\lambda_i(\neq 0) \in R$，$u_i \in R^n$，$i = 1, 2$。

由拟牛顿条件，

$$\Delta H^k Y^k = S^k - H^k Y^k.$$

即

$$\lambda_1 u_1 u_1^T Y^k + \lambda_2 u_2 u_2^T Y^k = S^k - H^k Y^k. \tag{4-6-127}$$

与$SR1$不同，上式(4-6-127)不能唯一确定u_1，u_2，因此，特殊有：

$$\lambda_1 u_1 u_1^T Y^k = S^k, \tag{4-6-128}$$

$$\lambda_2 u_2 u_2^T Y^k = -H^k Y^k. \tag{4-6-129}$$

所以$\exists \alpha, \beta \in R$，使得

$$u_1 = \alpha S^k, \tag{4-6-130}$$

$$u_2 = -\beta H^k Y^k. \tag{4-6-131}$$

所以

$$\lambda_1 \alpha^2 S^k (S^k)^T Y^k = S^k, \tag{4-6-132}$$

$$\lambda_2 \beta^2 H^k Y^k (H^k Y^k)^T Y^k = -H^k Y^k. \tag{4-6-133}$$

因而

$$\lambda_1\alpha^2 = \frac{1}{(S^k)^T Y^k} \tag{4-6-134}$$

$$\lambda_2\beta^2 = -\frac{1}{(H^k Y^k)^T Y^k} \tag{4-6-135}$$

即

$$\lambda_1 u_1 u_1^T = \lambda_1\alpha^2 S^k (S^k)^T = \frac{S^k (S^k)^T}{(S^k)^T Y^k} \tag{4-6-136}$$

$$\begin{aligned} \lambda_2 u_2 u_2^T &= \lambda_2\beta^2 H^k Y^k (H^k Y^k)^T \\ &= -\frac{H^k Y^k (H^k Y^k)^T}{(H^k Y^k)^T Y^k} \end{aligned} \tag{4-6-137}$$

从而得到DFP算法的对称秩二校正公式：

$$H^{k+1} = H^k + \frac{S^k(S^k)^T}{(S^k)^T Y^k} - \frac{H^k Y^k (H^k Y^k)^T}{(H^k Y^k)^T Y^k} \qquad (Eq:SR2)$$

(2)正定遗传性

定理 4.34 若取$H^1 = I$，则$(Eq:SR2)$所产生的H^k是对称正定的。

证明 对称性显然，下只证正定性。

$k=1$，成立；假设H^k成立；下证H^{k+1}成立：

$\forall\ X(\neq 0) \in R^n$，有

$$
\begin{aligned}
X^T H^{k+1} X &= X^T H^k X + \frac{X^T S^k (S^k)^T X}{(S^k)^T Y^k} - \frac{X^T H^k Y^k (H^k Y^k)^T X}{(H^k Y^k)^T Y^k} \\
&= \frac{X^T H^k X (Y^k)^T H^k Y^k - X^T H^k Y^k (Y^k)^T H^k X}{(Y^k)^T H^k Y^k} \\
&\qquad + \frac{X^T S^k (S^k)^T X}{(S^k)^T Y^k} \\
&= \frac{X^T H^k X (Y^k)^T H^k Y^k - (X^T H^k Y^k)^2}{(Y^k)^T H^k Y^k} \\
&\qquad + \frac{(X^T S^k)^2}{(S^k)^T Y^k}
\end{aligned}
\tag{4-6-138}
$$

又因为

$$\begin{aligned}(S^k)^T Y^k &= t_k(P^k)^T(g^{k+1} - g^k) \\ &= -t_k(P^k)^T g^k = t_k(g^k)^T H^k g^k > 0, (H^k > 0)\end{aligned} \tag{4-6-139}$$

$$X^T H^k X (Y^k)^T H^k Y^k - (X^T H^k Y^k)^2 \geq 0, (Cauchy - Schwarz) \tag{4-6-140}$$

$$\begin{aligned}(X^T S^k)^T &= (S^k)^T X = \alpha (S^k)^T Y^k \\ &= \alpha t_k (g^k)^T H^k g^k \neq 0, (X = \alpha Y^k, \alpha \neq 0)\end{aligned} \tag{4-6-141}$$

证毕! ∎

(3)算法描述

算法 4.7 DFP算法

在具有对称秩一校正公式的变尺度法中，只要把尺度矩阵迭代公式改为式($Eq:SR2$)，该算法就为DFP算法。

例 4.35 用DFP算法求$\min 2x_1^2+x_2^2-4x_1+2$，取$X_1=(2,1)^T$。

解 $g=\begin{pmatrix}4(x_1-1)\\2x_2\end{pmatrix}$，$Q=\begin{pmatrix}4&0\\0&2\end{pmatrix}$，$H^1=\begin{pmatrix}1&0\\0&1\end{pmatrix}$，$g^1=\begin{pmatrix}4\\2\end{pmatrix}$，$\|g^1\|\neq 0$；

第一次迭代：

$P^1=-H^1g^1=-g^1$，$X^2=X^1+t_1P^1=X^1+\frac{-g^1P^1}{(P^1)^TQP^1}P^1=\begin{pmatrix}\frac{8}{9}\\ \\ \frac{4}{9}\end{pmatrix}$，$g^2=\begin{pmatrix}-\frac{4}{9}\\ \\ \frac{8}{9}\end{pmatrix}$，$\|g^2\|\neq 0$；

第二次迭代：

$S^1=X^2-X^1=\begin{pmatrix}-\frac{10}{9}\\ \\ -\frac{5}{9}\end{pmatrix}$，$Y^1=g^2-g^1=\begin{pmatrix}-\frac{40}{9}\\ \\ -\frac{10}{9}\end{pmatrix}$，

所以

$$H^2=H^1+\frac{S^1(S^1)^T}{(S^1)^TY^1}-\frac{H^1Y^1(H^1Y^1)^T}{(H^1Y^1)^TY^1}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{18}\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} - \frac{1}{17}\begin{pmatrix} 16 & 4 \\ 4 & 1 \end{pmatrix}$$

$$= \frac{1}{306}\begin{pmatrix} 86 & -38 \\ -38 & 305 \end{pmatrix}$$

搜索方向为$P^2 = -H^2 g^2 = \frac{12}{51}\begin{pmatrix} 1 \\ -4 \end{pmatrix}$

从X^2出发沿P^2进行一维搜索，即

$$X^3 = X^2 + t_2 P^2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

$\|g^3\| = 0$，所以X^3是极小点。 ▮

变尺度法是求解无约束优化问题的最有效方法之一，其具有二次收敛性。满足式(4-6-127)的u_1, u_2有无穷多个，因此上述变尺度法就构成一族算法。常见的有$Broyden$族和$Huang$族两大类。

4.6.2.4 修正公式

(1)1967年$Broyden$给出：

$Broyden$类公式：

$$
\begin{aligned}
H^{k+1} = H^k &+ \frac{S^k (S^k)^T}{(Y^k)^T S^k} - \frac{H^k Y^k (Y^k)^T H^k}{(Y^k)^T H^k Y^k} \\
&+ \beta (Y^k)^T S^k (Y^k)^T H^k Y^k \omega^k (\omega^k)^T.
\end{aligned} \tag{4-6-142}
$$

在这里，

$$
\omega^k = \frac{S^k}{(Y^k)^T S^k} - \frac{H^k Y^k}{(Y^k)^T H^k Y^k}. \tag{4-6-143}
$$

- 当$\beta = \frac{1}{(Y^k)^T (S^k - H^k Y^k)}$时，$\Rightarrow SR1$。
- 当$\beta = 0$时，$\Rightarrow SR2$。

(2)1970年$Huang$给出：

$Huang$类公式：

$$H^{k+1} = H^k + S^k (u^k)^T + H^k Y^k (v^k)^T. \tag{4-6-144}$$

在这里，

$$u^k = a_k S^k + b_k (H^k)^T Y^k, \tag{4-6-145}$$

$$v^k = c_k S^k + d_k (H^k)^T Y^k. \tag{4-6-146}$$

且满足

$$(u^k)^T Y^k = \omega, \tag{4-6-147}$$

$$(v^k)^T Y^k = -1 \tag{4-6-148}$$

- 当$\omega=1$，$a_k=b_k=0$时，$\Rightarrow DFP$。
- 当$\omega=1$，$b_k=c_k$时，$\Rightarrow Broyden$。
- 当$\omega=1$，$b_k=c_k$，$d_k=0$时，$\Rightarrow BFGS$，即

$$H^{k+1}=H^k+\beta_k\frac{S^k(S^k)^T-H^kY^k(S^k)^T-S^k(Y^k)^TH^k}{(S^k)^TY^k},$$
$$\beta_k=1+\frac{(Y^k)^TH^kY^k}{(S^k)^TY^k}. \tag{4-6-149}$$

$\boldsymbol{BFGS}$由$\boldsymbol{B}royden$，$\boldsymbol{F}letcher$，$\boldsymbol{G}oldfarb$和$\boldsymbol{S}hanno$于1970年提出。由于实际计算中，在舍去误差的影响，特别是直线搜索不精确的影响，一般会破坏尺度矩阵的正定性，从而导

致算法失效。$BFGS$比DFP具有更好的数值稳定性。为保证H^k的正定性，常采取以下重置措施：迭代$n+1$次后，重置初始点和迭代矩阵，即$X^1 = X^n$，$H^1 = I$。

(3)$Broyden$族和$Huang$族关系为：

$$Huang \succ \left| \begin{array}{l} DFP \\ \\ Broyden \succ \left| \begin{array}{l} SR1 \\ \\ SR2 \end{array} \right. \\ \\ BFGS \end{array} \right.$$

§ 4.7 信赖域方法

牛顿法的基本思想是在迭代点x_k附近用二次函数

$$q^{(k)}(s) = f(x_k) + g_k^T s + \frac{1}{2} s^T G_k s \tag{4-7-150}$$

逼近$f(x)$，并以$q^{(k)}(s)$的极小点修正x_k，得到

$$x_{k+1} = x_k + s_k. \tag{4-7-151}$$

但是，这种方法只能保证算法的局部收敛性，即只有当s充分小时，$q^{(k)}(s)$才能逼近$f(x)$。

为了建立算法的总体收敛性，我们前面介绍了一维搜索技

术，在采用一维搜索的方法中，我们保持同样的搜索方向，而选择一个缩短了的步长。虽然这种策略是成功的，但它有一个缺点，即没有进一步利用n维二次模型。

4.7.1 基本思想

信赖域方法不仅可以用来代替一维搜索，而且也可以解决$Hesse$矩阵G_k不正定和x_k为鞍点等困难。

这种方法首先选择一个缩短了的步长，然后利用n维二次模型选择搜索方向。即先确定一个步长上界h_k，并由此定

又x_k的邻域Ω_k，

$$\Omega_k = \{x \mid \|x - x_k\| \leq h_k\}. \tag{4-7-152}$$

假定在这个邻域中$q^{(k)}(s)$与目标函数$f(x_k+s)$一致，即二次模型是目标函数$f(x)$的一个合适的模拟，然后用n维二次模型$q^{(k)}(s)$确定搜索方向s_k，并取$x_{k+1}=x_k+s_k$。

这种方法既具有牛顿法的快速局部收敛性，又具有理想的总体收敛性。由于步长受到使$Taylor$展式有效的信赖域的限制，故方法又称为限步长方法。

信赖域方法的模型问题是

$$\min q^{(k)}(s) = f(x_k) + g_k^T s + \frac{1}{2} s^T G_k s$$

$$s.t. \qquad \text{(4-7-153)}$$

$$\|s\| \le h_k.$$

注意，这里的范数没有指明，可以利用2-范数$\|\cdot\|_2$，∞-范数$\|\cdot\|_\infty$，也可以利用G-范数$\|\cdot\|_G$或其它范数，多数方法采用$\|\cdot\|_2$。

如何选择h_k？设Δf_k为f在第k步的实际下降量:

$$\Delta f_k = f_k - f(x_k + s_k), \qquad \text{(4-7-154)}$$

$\Delta q^{(k)}$为对应的预测下降量:

$$\Delta q^{(k)} = f_k - q^{(k)}(s). \tag{4-7-155}$$

定义比值:

$$r_k = \Delta f_k / \Delta q^{(k)}, \tag{4-7-156}$$

它衡量二次模型$q^{(k)}(s_k)$近似目标函数$f(x_k + s_k)$的程度。r_k越接近1，表明近似程度越好。

下面给出一个简单的模式算法，它自适应地改变h_k，并且在使h_k尽可能大的同时，尽量保持二次模型与目标函数的一致程度。

算法 4.8 信赖域方法

步骤 1 给出x_0，令$h_0 = \|g_0\|$，$k = 0$。

步骤 2 给出x_k和h_k，计算g_k和G_k。

步骤 3 解信赖域模型(4-7-153)，求出s_k。

步骤 4 求$f(x_k + s_k)$和r_k的值。

步骤 5

if $r_k < 0.25$ **then**

$h_{k+1} = \|s_k\|/4;$

else if $r_k > 0.75$和$\|s_k\| = h_k$ **then**

$h_{k+1} = 2h_k$；

else

$h_{k+1} = h_k$。

end if

步骤 6 若$r_k \leq 0$，则$x_{k+1} = x_k$；否则，$x_{k+1} = x_k + s_k$。

应该指出，(1)算法中的常数0.25，0.75等是根据经验选取的，算法对这些常数的变化不太敏感。(2)也可根据多项式插值选取h_k，例如当$r_k < 0.25$时，可以在区间$(0.1\|s_k\|, 0.5\|s_k\|)$中由多项式插值选取h_{k+1}。

例 4.36 求解$\min f(X) = x_1^4+x_1^2+x_2^2-4x_2+5$，这里，$X^0 = (0,0)^T$。

解 ∎

4.7.2 信赖域方法的收敛性

信赖域方法的一个突出优点是它具有总体收敛性.下面我们

仅就l_2范数给出方法的总体收敛性证明.类似地,利用范数等价性定理,可以知道其总体收敛性对其它范数也成立.

定理 4.37 (总体收敛性定理) 设$B \in R^n$是有界集,$x_k \in B$, $\forall k$,若$f \in C^2$,在有界集B上$\|G_k\|_2 \leq M$, $M > 0$,则信赖域算法产生一个满足一阶和二阶必要条件的聚点x^∞.

定理 4.38 如果聚点x^∞还满足f的Hesse矩阵G^∞是正定的条件,那么,对于主序列,有$r_k \to 1$, $x_k \to x^\infty$, $\mathrm{glb}\,(x_k) > 0$,以及对于充分大的k,约束$\|s\|_2 < h_k$.此外,收敛速度是二阶的.

第五章　约束最优化方法

问题为：

$$\min f(X) \qquad (NLP)$$

$$s.t.$$

$$g_i(X) \geq 0, \quad i = 1 \sim m; \qquad (5\text{-}0\text{-}1)$$

$$h_j(X) = 0, \quad j = 1 \sim l. \qquad (5\text{-}0\text{-}2)$$

方法大致分为以下几类：

- 可行方向法（$Feasible\ Direction$）

- 线性化（*Linearization Technique*）
- 惩罚函数法（*Penalty Function*）

§ 5.1 最优性条件

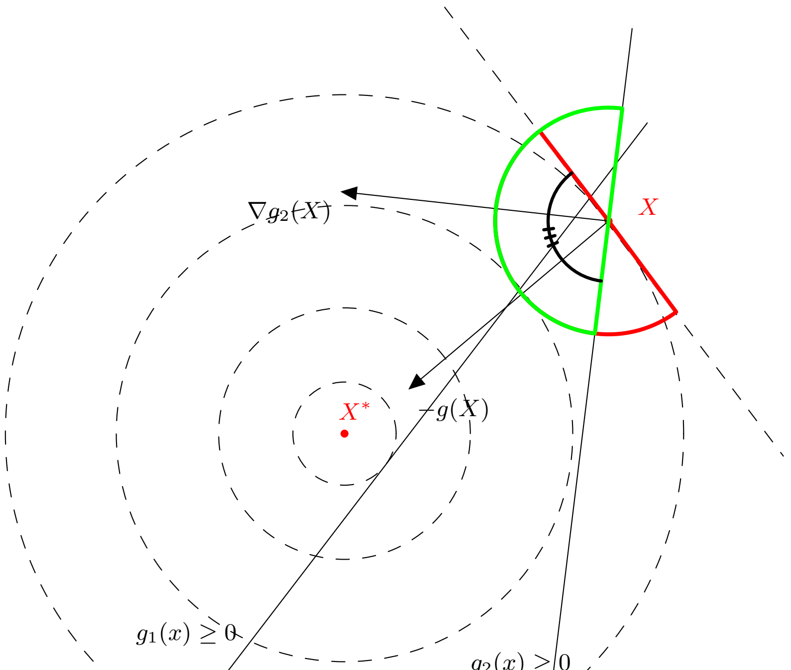
5.1.1 可行方向

定义 5.1 (可行域) $R = \{X|g_i(X) \geq 0, h_j(X) = 0, i = 1 \sim m, j = 1 \sim l\}$。

定义 5.2 (可行方向) 设$X \in R$，若对某一个非零向量$P \in R^n$，存在一个$\delta > 0$，使得对所有$\lambda \in (0, \delta)$时，有$X + \lambda P \in R$，则称P 为点X处的一个可行方向。

定义 5.3 (下降方向)　设X是(NLP)的一个可行点，存在$\delta>0$，当$\lambda\in(0,\delta)$时，若存在非零向量P使得$f(X+\lambda P)<f(X)$成立，则称P为$f(X)$在点X处的一个下降方向。

定义 5.4 (可行下降方向)　设X是(NLP)的一个可行点，非零向量P既是点X处的可行方向，又是$f(X)$在点X处的一个下降方向。


$\nabla g_2(X)$
X
X^*
$-g(X)$
$g_1(x) \geq 0$
$g_2(x) \geq 0$

定义 5.5 设$\overline{X}$是(NLP)的一个可行点，若$\exists k \in \{1,\cdots,m\}$，使得$g_k(\overline{X})=0$，则称不等式$g_k(X)\geq 0$为关于$\overline{X}$的起作用约束(积极约束、有效约束)；否则称为不起作用约束(非积极约束、非有效约束)。若令$\mathcal{I}(\overline{X})\triangleq\{i \mid g_i(\overline{X})=0, i=1\sim m\}$，$\mathcal{J}\triangleq\{1,2,\cdots,l\}$，则称$\mathcal{A}=\mathcal{I}(\overline{X})\bigcup\mathcal{J}$为关于$\overline{X}$的起作用(积极、有效)约束集。

定理 5.6 设P是(NLP)的一个可行下降方向，则有

$$\nabla f(\overline{X})^T P<0; \tag{5-1-3}$$

$$\nabla g_k(\overline{X})^T P>0; \qquad k\in\mathcal{A}. \tag{5-1-4}$$

5.1.2 一阶必要条件

定义 5.7 (正则点) 设$\overline{X} \in R$，若$\nabla g_i(\overline{X})$，$\nabla h_j(\overline{X})$，$i \in \mathcal{I}(\overline{X})$，$j \in \mathcal{J}$线性无关，则称$\overline{X}$为$R$的一个正则点。

定理 5.8 (一阶必要条件) 设$f(X)$，$g_i(X)$，$h_j(X)$可微，若X^*是(NLP)的一个局部极小点且为正则点，则存在$\lambda_i^*(i \in \mathcal{I}(\overline{X^*}))$以及$\mu_j^*(j \in \mathcal{J})$，使得

$$\nabla f(X^*) = \sum_{i \in \mathcal{I}(X^*)} \lambda_i^* \nabla g_i(X^*) + \sum_{j=1}^{l} \mu_j^* \nabla h_j(X^*), \tag{5-1-5}$$

$$\lambda_i^* \geq 0, \quad \lambda_i^* g_i(X^*) = 0, \qquad i = 1, 2, \cdots, m. \tag{5-1-6}$$

成立。

定义 5.9 (KKT点) 称(5-1-5)和(5-1-6)为KKT条件；称满足KKT条件的点为KKT点，或$K-T$点。

例 5.10

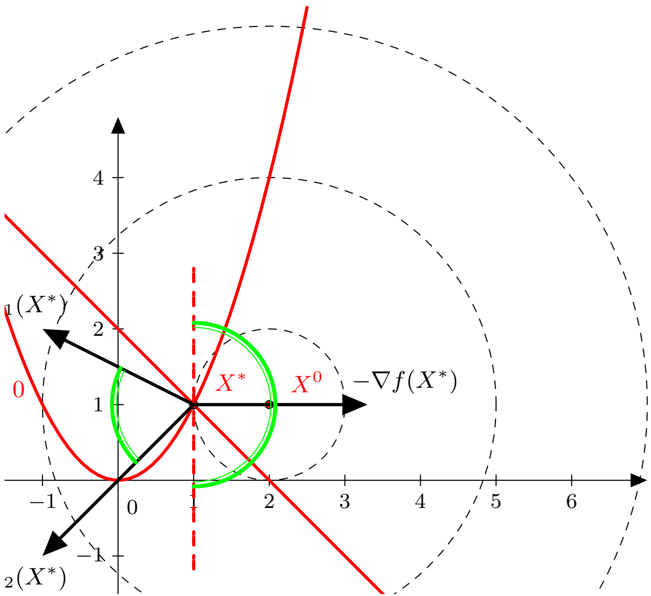
$$\min f(X) = (x_1 - 2)^2 + (x_2 - 1)^2$$

$$s.t.$$

$$g_1(X) = x_2 - x_1^2 \geq 0$$

$$g_2(X) = 2 - x_1 - x_2 \geq 0$$

解 ▮


$_1(X^*)$
$_2(X^*)$
$-\nabla f(X^*)$
X^*
X^0
0
4
3
2
1
−1
0
1
2
3
4
5
6
−1

例 5.11 试验证$X^* = (1,1,1)^T$时下面问题的KKT点。

$$\min f(X) = -3x_1^2 - x_2^2 - 2x_3^2$$

$s.t.$

$$-x_1 + x_2 \geq 0$$

$$x_1^2 + x_2^2 + x_3^2 = 3$$

$$x_i \geq 0, i = 1,2,3.$$

解 存在$\lambda^* = 2$，$\mu^* = -2$，使得

$$\nabla f(X^*) = \lambda^* \nabla g_1(X^*) + \mu^* \nabla h_1(X^*),$$

$$\lambda^* \geq 0,$$

$$\lambda^* g_1(X^*) = 0.$$

∎

5.1.3 二阶充分条件

定理 5.12 (二阶充分条件) 设$f(X)$，$g_i(X)$，$h_j(X)$二次连续可微，若$X^* \in R$，存在λ_i^*，μ_j^*，$i = 1 \sim m$，$j = 1 \sim l$，使得KKT条件成立且对满足条件

$$Z^T \nabla g_i(X^*) = 0, \qquad i \in \mathcal{I}(X^*), \lambda_i^* > 0; \tag{5-1-7}$$

$$Z^T \nabla g_i(X^*) \geq 0, \qquad i \notin \mathcal{I}(X^*), \lambda_i^* = 0; \tag{5-1-8}$$

$$Z^T \nabla h_j(X^*) = 0. \tag{5-1-9}$$

的非零向量Z有

$$Z^T \Big(\nabla^2 f(X^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(X^*) - \sum_{j=1}^{l} \mu_j^* \nabla^2 h_j(X^*) \Big) Z > 0 \tag{5-1-10}$$

成立，则X^*为(NLP)的严格局部最小点。

例 5.13

$$\min f(X) = x_1^2 - 8x_1 + x_2 + 16$$

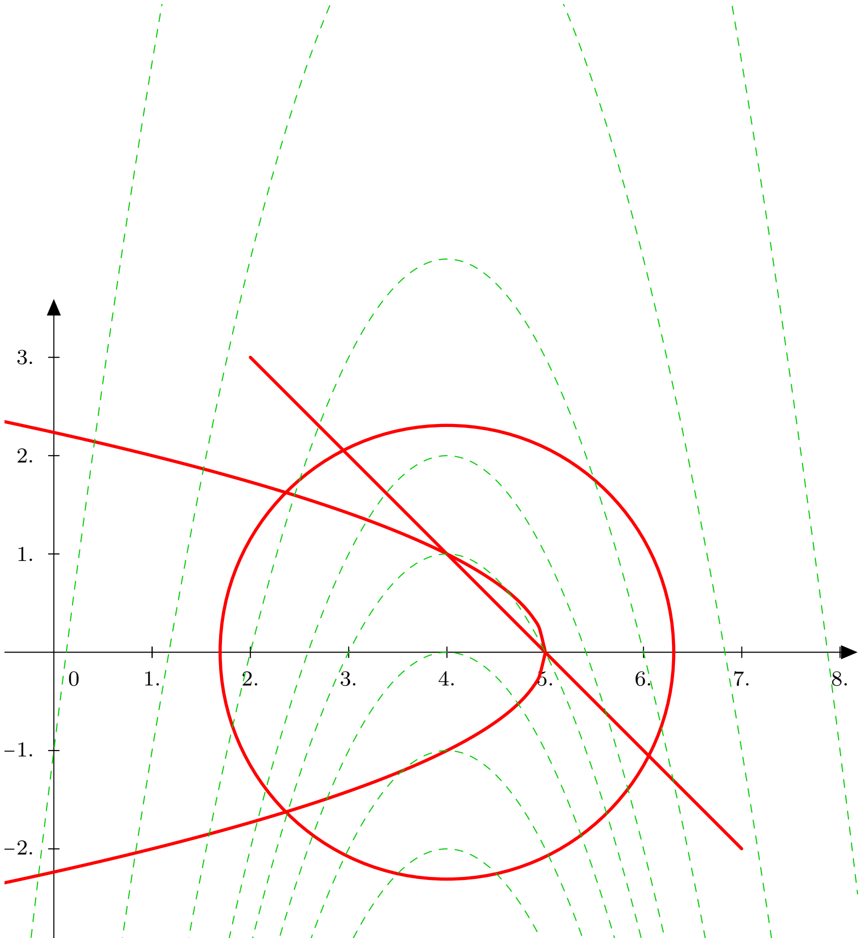
$s.t.$

$$g_1(X) = -(x_1 + x_2^2) + 5 \geq 0$$

$$g_2(X) = -(x_1 + x_2) + 5 \geq 0$$

$$h_1(X) = (x_1 - 4)^2 + x_2^2 - 9 = 0$$

解


3.
2.
1.
−1.
−2.
0
1.
2.
3.
4.
5.
6.
7.
8.

§ 5.2 惩罚函数法

5.2.1 基本思想

基本思想是有约束的最优化问题通过罚因子的选择变为一系列的求罚函数的极小值，即将原问题转化为无约束优化问题。考虑

$$\min f(X) \tag{5-2-11}$$

$s.t.$

$$g_i(X) = 0, i = 1 \sim m. \tag{5-2-12}$$

罚函数法将其转化为

$$\min P(X, m_k) = \min f(X) + m_k \sum_{i=1}^{m} \Big(g_i(X)\Big)^2 \tag{5-2-13}$$

这里，称m_k为罚因子，其为一个很大的实数；称$P(X, m_k)$为罚函数。通常令$\alpha(X) \triangleq m_k \sum_{i=1}^{m} \Big(g_i(X)\Big)^2$为罚项。

5.2.2 经济解释

若把目标函数看成是价格，约束条件看成是规定，采购人员可在规定的范围内购置最便宜的东西。对违反规定制定了惩罚政策。

若符合规定，罚款为零；若违反政策，要收罚款，接受惩

罚。结果是采购人员付出的总代价是价格和罚款的总和，迫使总代价最小。

5.2.3 罚因子与拉格朗日乘子之间的关系

对于问题(5-2-11)：

令拉格朗日函数为

$$L(X,\lambda)=f(X)+\sum_{i=1}^{m}\lambda_i g_i(X) \tag{5-2-14}$$

罚函数为

$$P(X,m)=f(X)+\sum_{i=1}^{m}m_i\Big(g_i(X)\Big)^2 \tag{5-2-15}$$

则有

$$\frac{\lambda_i}{2m_i} = g_i(X). \tag{5-2-16}$$

§ 5.3 外点罚函数法

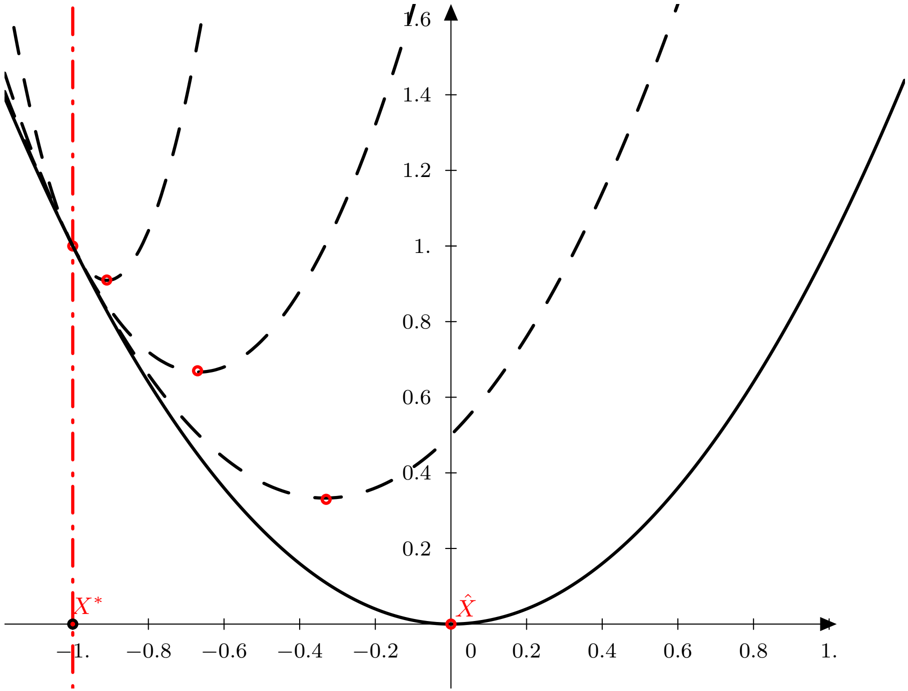
5.3.1 基本思想

例 5.14

$$\min f(X) = x^2$$

$s.t.$

$$g(X) = x + 1 \leq 0$$


1.6
1.4
1.2
1.
0.8
0.6
0.4
0.2
X^*
$\hat{X}$
−1.
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1.

5.3.2 一般约束最优化处理

首先针对

$$minf(x) \tag{5-3-17}$$

$$s.t.$$

$$h_j(X) = 0, j = 1 \sim l.$$

构造

$$P(X, m_k) = f(X) + m_k \sum_{j=1}^{l} \Big(h_j(X)\Big)^2. \tag{5-3-18}$$

仿此，对于

$$minf(X) \tag{5-3-19}$$

$$s.t.$$

$$g_i(X) \geq 0, i = 1 \sim m.$$

构造

$$P(X, m_k) = f(X) + m_k \sum_{i=1}^{m} \Big(g_i(X)\Big)^2 u_i(g_i). \tag{5-3-20}$$

这里，$m_{k+1} > m_k$，$u_i(g_i)$为**单位阶跃函数**，即

$$u_i(g_i) = \begin{cases} 0, & if\ g_i \geq 0 \\ 1, & otherwise \end{cases} \tag{5-3-21}$$

或者，如下构造

$$P(X, m_k) = f(X) + m_k \sum_{i=1}^{m} \left(\min \Big(g_i(X), 0 \Big) \right)^2 \tag{5-3-22}$$

因此，对于一般问题(NLP)，可如下构造罚函数

$$p(X, m_k) = f(X) + m_k \left(\sum_{j=1}^{l} \Big(h_j(X) \Big)^2 + \sum_{i=1}^{m} \Big(g_i(X) \Big)^2 u_i(g_i) \right). \tag{5-3-23}$$

若令

$$\alpha(X) = \sum_{j=1}^{l} \Big(h_j(X)\Big)^2 + \sum_{i=1}^{m} \Big(g_i(X)\Big)^2 u_i(g_i). \tag{5-3-24}$$

则约束优化问题(NLP)可转化为求解下面的无约束优化问题：

$$\min P(X, m_k) = \min \Big(f(X) + m_k\alpha(X)\Big). \tag{5-3-25}$$

如果$X_k^* \in R$，则对任意$X \in R$，有

$$f(X_k^*) = P(X_k^*, m_k) \le P(X, m_k) = f(X). \tag{5-3-26}$$

即X_k^*是问题(5-3-23)的最优解。显然若X_k^*是问题(NLP)的最

优解，自然$X_k^* \in R$。

5.3.3 算法

见算法5.1。

算法 5.1 外点法

步骤 1 选定X^0，$m_1 > 0(m_1 = 1)$，$C > 1(C = 10)$，$k = 1$。

步骤 2 以X^{k-1}为初始点，求解无约束优化问题(5-3-23)的最优解X^k。

步骤 3 若$m_k\alpha(X^k) < \varepsilon$，则$X^k$为优化问题($NLP$)的最优解，否则转步骤(4)。

步骤 4 令$m_{k+1} = Cm_k$，$k = k+1$，转步骤(2)。

例 5.15

$$\min f(X)$$

$$s.t.$$

$$g_1(X) = -x_1^2 + x_2 \geq 0.$$

$$g_2(X) = x_1 \geq 0.$$

解

$$X(m) = \left(-\frac{1}{2(1+m)}, \left(\frac{1}{4(1+m)^2} - \frac{1}{2m} \right) \right)^T.$$

∎

例 5.16

$$\min(x-1)^2$$

$$s.t.$$

$$x-2 \geq 0.$$

解 $x^* = 2$。 ∎

5.3.4 收敛性定理

引理 5.17 设点列$\{X^k\}$是由算法5.1所产生的序列，则有

$$P(X^{k+1}, m_{k+1}) \geq P(X^m, m_k) \tag{5-3-27}$$

$$\alpha(X^{k+1}) \leq \alpha(X^k) \tag{5-3-28}$$

$$f(X^{k+1}) \geq f(X^k) \tag{5-3-29}$$

这里，$k \geq 1$。

证明 首先

$$\begin{aligned} P(X^{k+1}, m_{k+1}) &= f(X^{k+1}) + m_{k+1}\alpha(X^{k+1}) \\ &\geq f(X^{k+1}) + m_k\alpha(X^{k+1}) \\ &= P(X^{k+1}, m_k) \\ &\geq P(X^k, m_k). \end{aligned} \tag{5-3-30}$$

接着

$$f(X^{k+1}) + m_k\alpha(X^{k+1}) \geq f(X^k) + m_k\alpha(X^k), \tag{5-3-31}$$

$$f(X^k) + m_{k+1}\alpha(X^k) \geq f(X^{k+1}) + m_{k+1}\alpha(X^{k+1}). \tag{5-3-32}$$

即

$$\begin{aligned} m_k\Big(\alpha(X^k) - \alpha(X^{k+1})\Big) &\leq f(X^{k+1}) - f(X^k) \\ &\leq m_{k+1}\Big(\alpha(X^k) - \alpha(X^{k+1})\Big). \end{aligned} \tag{5-3-33}$$

则有

$$(m_{k+1}-m_k)\Big(\alpha(X^k)-\alpha(X^{k+1})\Big)\geq 0. \tag{5-3-34}$$

因此

$$\alpha(X^{k+1})\leq\alpha(X^k). \tag{5-3-35}$$

最后，显然有$f(X^{k+1})\geq f(X^k)$成立。 ∎

定理 5.18 设f，g_i，h_j都是R^n上的连续函数，则由算法5.1产生的任何收敛序列$\{X^k\}$的极限点必是约束优化问题(NLP)的极小点。

证明 设$X^k \to \overline{X}$，X^*为问题NLP的极小点。因为

$$f(X^*) = P(X^*, m_k) \geq P(X^k, m_k) \geq f(X^k), \tag{5-3-36}$$

所以

$$f(X^*) \geq f(X^k). \tag{5-3-37}$$

下证明

$$f(X^*) \leq f(X^k). \tag{5-3-38}$$

因为

$$\lim_{k\to\infty} P(X^k, m_k) = \overline{P} \leq f(X^*), \tag{5-3-39}$$

$$\lim_{k\to\infty} f(X^k) = \overline{f} \leq f(X^*). \tag{5-3-40}$$

所以

$$\lim_{k\to\infty} \alpha(X^k) = \lim_{k\to\infty} \frac{P(X^k, m_k) - f(X^k)}{m_k} = 0. \tag{5-3-41}$$

因此

$$\alpha(\overline{X}) = \lim_{k\to\infty} \alpha(X^k) = 0. \tag{5-3-42}$$

所以

$$\overline{X} \in R. \tag{5-3-43}$$

从而式子(5-3-38)成立。

由式子(5-3-37)和式子(5-3-38)得证！ ∎

§ 5.4 内点罚函数法

外点法特点：

(1)等式约束，非凸规划；

(2)初始点可以任意选择；

但是，

(3)可行区域外函数无定义或性质复杂，外点法失效。

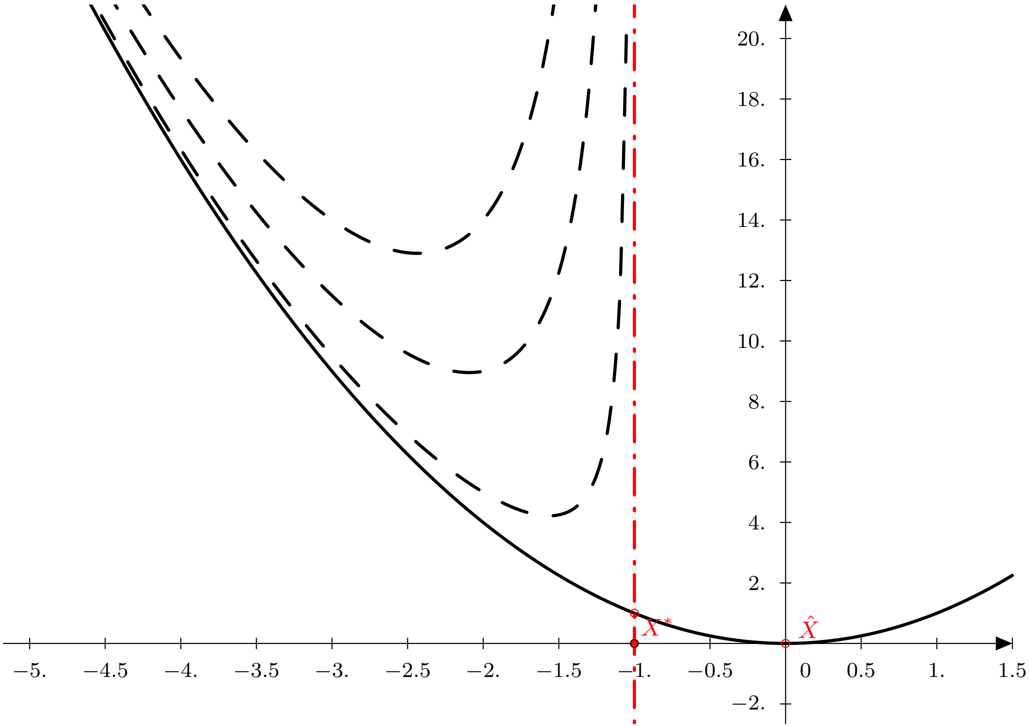
5.4.1 基本思想

例 5.19

$$\min f(X) = x^2$$

$s.t.$

$$g(X) = x + 1 \le 0$$


20.
18.
16.
14.
12.
10.
8.
6.
4.
2.
−2.
−5.
−4.5
−4.
−3.5
−3.
−2.5
−2.
−1.5
−1.
−0.5
0
0.5
1.
1.5
X^*
$\hat{X}$

考虑

$$\min f(X) \tag{5-4-44}$$

$$s.t.$$

$$g_i(X) \geq 0, i = 1 \sim m.$$

这里$S \triangleq \{X \mid g_i(X) \geq 0, i = 1 \sim m\}$，并要求$S^o \neq \emptyset$。

定义罚函数为

$$P(X, r_k) = f(X) + r_k \sum_{i=1}^{m} \frac{1}{g_i(X)}. \tag{5-4-45}$$

这里称

$$\beta(X)=\sum_{i=1}^{m}\frac{1}{g_i(X)}. \tag{5-4-46}$$

为罚项。

或定义罚函数为

$$P(X,r_k)=f(X)+r_k\sum_{i=1}^{m}\ln g_i(X), \tag{5-4-47}$$

$$P(X,r_k)=f(X)+r_k\sum_{i=1}^{m}\frac{1}{\Big(g_i(X)\Big)^2}. \tag{5-4-48}$$

例 5.20

$$\min f(X) = x$$

$s.t.$

$$g(x) = x \geq 0.$$

5.4.2 算法

例 5.21

$$\min f(X) = \frac{1}{3}(x_1 + 1)^3 + x_2$$

$s.t.$

算法 5.2 内点法

步骤 1 选定$X^0 \in S^o$，$r_1 > 0(r_1 = 10)$，$C < 1(C = 0.1)$，$k = 1$。

步骤 2 以X^{k-1}为初始点，求解无约束优化问题(5-4-45)的最优解X^k。

步骤 3 若$r_k\beta(X^k) < \varepsilon$，则$X^k$为优化问题(5-4-44)的最优解，否则转步骤(4)。

步骤 4 令$r_{k+1} = Cr_k$，$k = k + 1$，转步骤(2)。

$$g_1(X) = x_1 - 1 \geq 0,$$

$$g_2(X) = x_2 \geq 0.$$

解 $X^* = (1, 0)^T$。 ∎

例 5.22 用对数罚函数法(5-4-47)求解

$$\min f(X) = x_1 + 2x_2$$

$s.t.$

$$g_1(X) = -x_1^2 + x_2 \geq 0,$$

$$g_2(X) = x_1 \geq 0.$$

解 $X^* = (0, 0)^T$。 ∎

5.4.3 收敛性定理

引理 5.23 由内点法(5.2)产生的点列X^k，有

$$P(X^{k+1}, r_{k+1}) \le P(X^k, r_k). \tag{5-4-49}$$

证明

$$\begin{aligned} P(X^k, r_k) &= f(X^k) + r_k\beta(X^k) \\ &\ge f(X^k) + r_{k+1}\beta(X^k) \\ &= P(X^k, r_{k+1}) \\ &\ge P(X^{k+1}, r_{k+1}). \end{aligned} \tag{5-4-50}$$

∎

定理 5.24 设f，g_i都是R^n上的连续函数，则由内点算法5.2产生的任何收敛序列$\{X^k\}$的极限点必是约束优化问题(5-4-44)的极小点。

证明 设$X^k \to \overline{X}$，并令X^*是优化问题(5-4-44) 的极小点。因为

$$P(X^k, r_k) \geq f(X^k) \geq f(X^*), \tag{5-4-51}$$

所以

$$\lim_{k\to\infty} P(X^k, r_k) \triangleq \overline{P} \geq f(X^*). \tag{5-4-52}$$

因为f连续，所以对$\forall\varepsilon>0$，都存在$\widetilde{X}\in S^o$，使得

$$f(\widetilde{X})-f(X^*)<\varepsilon, \tag{5-4-53}$$

又因为

$$P(X^k,r_k)-f(X^*)\leq f(\widetilde{X})-f(X^*)+r_k\beta(\widetilde{X}), \tag{5-4-54}$$

所以当$k\to\infty$，就有

$$P(X^k,r_k)-f(X^*)<\varepsilon. \tag{5-4-55}$$

因此

$$\lim_{k\to\infty} P(X^k, r_k) = f(X^*) = \lim_{k\to\infty} f(X^k). \tag{5-4-56}$$

∎

5.4.4 小结

惩罚函数法基本思想：若当前迭代点不满足可行性或有不满足可行性的趋势，则对其函数值添加一个比较大的数字（惩罚项），迫使迭代点在极小化的过程中向可行区域靠近或满足可行性。

优缺点：算法简单，可以用求解无约束优化问题的方法

求解约束优化问题，然而罚因子选择困难，容易出现数值不稳定情形；外点法通常得到的解不满足可行性，而内点法满足；外点法可以处理所有约束，而内点法只能处理不等式约束。

§ 5.5 乘子法

考虑

$$\min\ f(X)$$

$$s.t. \tag{5-5-57}$$

$$h_j(X) = 0,\ j = 1, \cdots, l.$$

这里，$f(X)$，$h_j(X)$是二次连续可微函数。

定义如下**乘子罚函数**(**增广*Lagrange*函数**)：

$$\phi(X,V,\sigma)=f(X)-\sum_{j=1}^{l}v_jh_j(X)+\frac{\sigma}{2}\sum_{j=1}^{l}h_j^2(X). \qquad (5\text{-}5\text{-}58)$$

这里，$V=(v_1,\cdots,v_l)^T$，$\sigma>0$。

定理 5.25 设$\widetilde{X}$，$\widetilde{V}$满足问题(5-5-57)的局部最优解的二阶充分条件，则存在$\sigma'\geq 0$，使得对$\forall\sigma>\sigma'$，$\widetilde{X}$是$\phi(X,\widetilde{V},\sigma)$的严格局部最小点。

反之，若存在X^0是问题(5-5-57)的可行解，且对于某个V^0，X^0是$\phi(X,V^0,\sigma)$的极小点，又满足极小点的二阶充

分条件，则X^0是问题(5-5-57)的最优解。

因此，若知道最优乘子V^*，只要取充分大的罚因子$\sigma < +\infty$，就能通过$\phi(X, V^*, \sigma)$得到问题(5-5-57)的最优解。

但是，最优乘子V^*一般事前是未知的。所以，一般先给充分大的σ和初始估计V^1，然后在迭代过程中，使得$V^k \to V^*$。

设在第k次迭代，$Lagrange$乘子的估计为V^k，罚因子

为σ，X^k为(5-5-58)的最优解，则

$$\begin{aligned}\nabla\phi(X^k,V^k,\sigma) &= \nabla f(X^k)-\sum_{j=1}^{l}v_j^k\nabla h_j(X^k)\\ &\quad+\sigma\sum_{j=1}^{l}h_j(X^k)\nabla h_j(X^k).\\ &= \nabla f(X^k)-\sum_{j=1}^{l}\left(v_j^k-\sigma h_j(X^k)\right)\nabla h_j(X^k)\\ &= 0.\end{aligned} \tag{5-5-59}$$

又因问题(5-5-57)的最优解X^*满足

$$\nabla f(X^*) - \sum_{j=1}^{l} v_j^* \nabla h_j(X^*) = 0. \tag{5-5-60}$$

所以令

$$v^{k+1} = v^k - \sigma h_j(X^k),\ j = 1, \cdots, l. \tag{5-5-61}$$

例 5.26

$$\min\ f(X) = 2x_1^2 + x_2^2 - 2x_1x_2$$

$s.t.$

$$h_1(X) = x_1 + x_2 - 1 = 0.$$

令$\sigma=2$，$v^1=1$。

解 考虑

$$\phi(X,v^k,\sigma)=2x_1^2+x_2^2-2x_1x_2-v^k(x_1+x_2-1)$$
$$+\frac{1}{2}(x_1+x_2-1)^2.$$

易得，在第k次迭代时，$\phi(X,v^k,\sigma)$的最优解为

$$X^k=\left(\frac{v^k+2}{6},\ \frac{v^k+2}{4}\right)^T.$$

因此

$$v^{k+1}=v^k-\sigma h_1(X^k)$$

$$= v^k - 2\sigma\left(\frac{v^k+2}{6} + \frac{v^k+2}{4} - 1\right)$$
$$= \frac{v^k}{6} + \frac{1}{3}.$$

显然，当$v^k > \frac{2}{5}$时，$v^{k+1} - v^k < 0$，所以当$k \to \infty$时，v^k收敛。

即，$v^* = \frac{2}{5}$，$X^* = (\frac{2}{5}, \frac{3}{5})^T$。 ▮

考虑不等式约束问题：

$$\min \ f(X)$$
$$s.t. \tag{5-5-62}$$

$$g_j(X) \geq 0,\ j = 1, \cdots, m.$$

这里，$f(X)$，$g_j(X)$是二次连续可微函数。

引入变量y_j，$j = 1, \cdots, m$，问题可变为：

$$\begin{aligned} &\min\ f(X) \\ s.t.& \\ &g_j(X) - y_j^2 = 0,\ j = 1, \cdots, m. \end{aligned} \tag{5-5-63}$$

定义如下**增广*Lagrange*函数**：

$$\phi(X, Y, W, \sigma) = f(X) - \sum_{j=1}^{m} w_j \left(g_j(X) - y_j^2\right)$$

$$+\frac{\sigma}{2}\sum_{j=1}^{m}\left(g_j(X)-y_j^2\right)^2. \tag{5-5-64}$$

这里，$W=(w_1,\cdots,w_m)^T$，$Y=(y_1,\cdots,y_m)^T$。

因为

$$\begin{aligned}
&\phi(X,Y,W,\sigma)\\
&\quad=f(X)+\sum_{j=1}^{m}\left(-w_j\left(g_j(X)-y_j^2\right)+\frac{\sigma}{2}\left(g_j(X)-y_j^2\right)^2\right)\\
&\quad=f(X)+\sum_{j=1}^{m}\left(\frac{\sigma}{2}\left(y_j^2-\frac{1}{\sigma}(\sigma g_j(X)-w_j)\right)^2-\frac{w_j^2}{2\sigma}\right).
\end{aligned} \tag{5-5-65}$$

令

$$y_j^2 = \frac{1}{\sigma}\max\{0, \sigma g_j(X) - w_j\}, \ j = 1, \cdots, m. \tag{5-5-66}$$

则问题(5-5-62)转化成求解下面问题

$$\phi(X, W, \sigma) = f(X) + \frac{1}{2\sigma}\sum_{j=1}^{m}\Big(\big(\max\{0, w_j - \sigma g_j(X)\}\big)^2 - w_j^2\Big). \tag{5-5-67}$$

对于一般情形：

$$\min \ f(X)$$

$$s.t. \tag{5-5-68}$$

$$g_j(X) \geq 0,\ \ j = 1, \cdots, m.$$

$$h_j(X) = 0,\ \ j = 1, \cdots, l.$$

定义**乘子罚函数**如下：

$$\begin{aligned}
&\phi(X, W, V, \sigma)\\
&\quad = f(X) + \frac{1}{2\sigma}\sum_{j=1}^{m}\Big(\big(\max\{0, w_j - \sigma g_j(X)\}\big)^2 - w_j^2\Big)\\
&\quad - \sum_{j=1}^{l} v_j h_j(X) + \frac{\sigma}{2}\sum_{j=1}^{l} h_j^2(X). \qquad (5\text{-}5\text{-}69)
\end{aligned}$$

这里，

$$\begin{aligned} w_j^{k+1} &= \max\left\{0, w_j - \sigma g_j(X^k)\right\}; \\ v_j^{k+1} &= v_j^k - \sigma h_j(X^k). \end{aligned} \tag{5-5-70}$$

乘子算法如下：

例 5.27

$$\min \ x_1^2 + 2x_2^2$$

$$s.t.$$

$$x_1 + x_2 \geq 1.$$

这里，$\sigma = 2$，$w^1 = 1$。

算法 5.3 乘子法

步骤 1 给定初始点X^0，乘子向量估计W^1，V^1，参数σ，误差$\varepsilon > 0$，常数$\alpha > 1$，$0 < \beta < 1$，令$k = 1$。

步骤 2 求解$\min\ \phi(X, W^k, V^k, \sigma)$，得到$X^k$。

步骤 3

if $\|h(X^k)\| < \varepsilon$ **then**

  停机, **return** X^k;

else if $\frac{\|h(X^k)\|}{\|h(X^{k-1})\|} \geq \beta$ **then**

  $\sigma = \alpha \cdot \sigma$;

end if

步骤 4 通过公式(5-5-70)修正W^k和V^k，令$k = k + 1$，转步骤2。

解 易得$\phi(X, w^k, \sigma)$的最优解为

$$X^k = \begin{cases} \frac{2+w^k}{5} \\ \frac{2+w^k}{10} \end{cases}$$

而且，

$$w^{k+1} = \frac{2w^k + 4}{5}.$$

显然，当$w^k < \frac{4}{3}$时，$w^{k+1} - w^k > 0$，所以w^k收敛。

因此，$w^* = \frac{4}{3}$，$X^* = (\frac{2}{3}, \frac{1}{3})^T$。 ∎

注意：σ不能太大，也不能太小。一般地，乘子法优于罚函数法。

练习 5.28

$$\min \ x_1 + \frac{1}{3}(x_2+1)^2$$

$s.t.$

$$x_1 \geq 0,$$

$$x_2 \geq 1.$$

解 $X^* = (0,1)^T$。 ∎

§ 5.6 *Rosen*梯度投影法

考虑

$$\min f(X) \tag{5-6-71}$$

$$s.t.$$

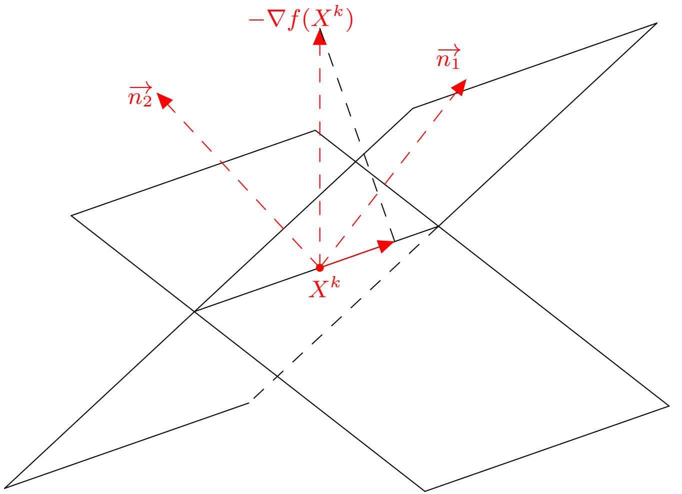
$$AX \geq b, \tag{5-6-72}$$

$$CX = d. \tag{5-6-73}$$

这里A是$m \times n$矩阵，C是$l \times n$矩阵，b是m维向量，d是l维向量，$f(X)$连续可微。

5.6.1 基本思想

1960年由$Rosen$提出，用于求解线性约束的非线性规划。若当前迭代点X^k的负梯度方向$-\nabla f(X^k)$不是可行方向，则将其投影到X^k的积极约束的法向量张成空间的正交补空间中，即积极约束的交线上，从而使其成为下降可行方向。


$-\nabla f(X^k)$
$\overrightarrow{n_1}$
$\overrightarrow{n_2}$
X^k

5.6.2 下降可行方向的确定

定义 5.29 设P为$n\times n$矩阵，若$P=P^T$且$P^2=P$，则称P为投影矩阵。

引理 5.30 设P为投影矩阵，则

(1)P为半正定；

(2)充要条件为$I-P$是投影矩阵；

(3)对$\forall X\in R^n$，可唯一的表示成$X=p+q$，$p\in L$，$q\in L^{\perp}$，这里

$$L\triangleq\{PX\mid X\in R^n\},\tag{5-6-74}$$

$$L^{\perp} \triangleq \{QX \mid X \in R^n\}. \tag{5-6-75}$$

定理 5.31 对$\forall X \in R^n$，$X \overset{1}{=} PX + QX$，并称PX为X 在子空间L上的投影，P为R^n到L的投影矩阵；对于Q 类似。

在可行点X^k处，将A，b分解为

$$A = \begin{pmatrix} A' \\ A'' \end{pmatrix}, b = \begin{pmatrix} b' \\ b'' \end{pmatrix}, \tag{5-6-76}$$

并要求

$$A'X^k = b' \tag{5-6-77}$$

$$A^{''}X^k > b^{''}. \tag{5-6-78}$$

令

$$N = \begin{pmatrix} A' \\ C \end{pmatrix}. \tag{5-6-79}$$

并假设$rank(N) = r$。

定义

$$N \triangleq \begin{pmatrix} n_1^T & n_2^T & \cdots & n_r^T \end{pmatrix}^T. \tag{5-6-80}$$

考虑

$$P = N^T(NN^T)^{-1}N, \tag{5-6-81}$$

$$Q = I - P \tag{5-6-82}$$

定理 5.32　对于优化问题(5-6-71)，搜索方向P^k为点X^k处的可行方向的充分必要条件是

$$A'P^k \geq 0; \tag{5-6-83}$$

$$CP^k = 0. \tag{5-6-84}$$

定理 5.33　若$P^k = -Q\nabla f(X^k) \neq 0$，则$P^k$为点$X^k$处的可行

下降方向，这里$Q = I - N^T(NN^T)^{-1}N$。

证明

$$\nabla f(X^k)^T P^k = -\big(Q\nabla f(X^k)\big)^T \big(Q\nabla f(X^k)\big) < 0; \tag{5-6-85}$$

$$NP^k = -N\big(I - N^T(NN^T)^{-1}N\big)\nabla f(X^k) = 0. \tag{5-6-86}$$

∎

5.6.3 直线搜索及终止准则

一维搜索：

$$\min_{0\le t\le \bar{t}} f(X^k + tP^k). \tag{5-6-87}$$

因为

$$A'P^k \geq 0, CP^k = 0. \tag{5-6-88}$$

所以只需

$$A''(X^k + tP^k) \geq b''. \tag{5-6-89}$$

即

$$A''X^k - b'' + tA''P^k \geq 0. \tag{5-6-90}$$

令

$$A^{''}X^k - b^{''} = \boldsymbol{u} \triangleq (u_1, u_2, \cdots, u_\theta)^T. \tag{5-6-91}$$

$$A^{''}P^k = \boldsymbol{v} \triangleq (v_1, v_2, \cdots, v_\theta)^T. \tag{5-6-92}$$

这里，$\theta \leq m$。

按照下式确定$\bar{t}$：

$$\bar{t} = \begin{cases} +\infty, & \boldsymbol{v} \geq 0; \\ \min_{1 \leq i \leq \theta} \left\{ -\frac{u_i}{v_i} \mid v_i < 0 \right\}, & \boldsymbol{v} \ngeq 0. \end{cases} \tag{5-6-93}$$

5.6.4 算法

当$P^k = -Q\nabla f(X^k) \neq 0$时，其就是可行下降方向；

若当$P^k = 0$时，有

$$\begin{aligned} Q\nabla f(X^k) &= \Big(I - N^T(NN^T)^{-1}N\Big)\nabla f(X^k) \\ &= \nabla f(X^k) - N^T(NN^T)^{-1}N\nabla f(X^k) \\ &= \nabla f(X^k) - N^T q \\ &= 0. \end{aligned} \tag{5-6-94}$$

这里，$q \triangleq (NN^T)^{-1}N\nabla f(X^k)$。

此时，

$$\begin{aligned}\nabla f(X^k) &= N^T q \\ &= \begin{pmatrix} A' \\ \\ C \end{pmatrix}^T \begin{pmatrix} y \\ \\ z \end{pmatrix} \\ &= (A')^T y + C^T z.\end{aligned} \tag{5-6-95}$$

当$y \geq 0$时，X^k是KT点；

若$y \not\geq 0$时，不妨假设$y_i < 0$，从N中删去第i行向量得到$\overline{N}$，并令

$$\overline{Q} = I - \overline{N}^T(\overline{N}\,\overline{N}^T)^{-1}\overline{N}. \tag{5-6-96}$$

容易证明$\overline{Q}$为投影矩阵，这时，

$$\overline{P}^k = -\overline{Q}\nabla f(X^k) \tag{5-6-97}$$

就是点X^k处的一个下降方向。

又因为

$$\begin{aligned}\overline{P}^k &= -\overline{Q}\nabla f(X^k) \\ &= -\overline{Q}(A')^T y - \overline{Q}C^T z \\ &= -\overline{Q}(A')^T y \qquad (5\text{-}6\text{-}98) \\ &= -\sum_{\forall k} \overline{Q} n_k^T y_k\end{aligned}$$

$$= -y_i\overline{Q}n_i^T.$$

这里，n_k表示N的第k行。

所以

$$n_k\overline{P}^k = -y_i n_k\overline{Q}n_i^T = \begin{cases} 0, & k \neq i; \\ -y_i\|\overline{Q}n_i^T\|^2 \geq 0, & k = i. \end{cases} \tag{5-6-99}$$

从而$\overline{P}^k$为点X^k处的一个可行下降方向。

算法 5.4 *Rosen*梯度投影法

步骤 1 选定X^0，$k=0$。

步骤 2 令$N_k=((A')^T,C^T)^T$。

步骤 3 若$r>0$，计算$Q_k=I-N_k^T(N_kN_k^T)^{-1}N_k$。转步骤(5)。

步骤 4 若$r=0$，计算$Q_k=I$。

步骤 5 计算$P^k=-Q_k\nabla f(X^k)$。

步骤 6 若$\|P^k\|<\varepsilon$，转步骤(7)；否则转步骤(9)。

步骤 7 若$r=0$，则X^k就是最优解；若$r\neq 0$，计算$q=(y^T,z^T)^T$。

步骤 8 若$y\geq 0$，则X^k是最优解(KT 点)；若$y\ngeq 0$，去掉N_k中相应的行向量。转步骤(2)。

步骤 9 求解$\min_{0\leq t\leq\bar{t}} f(X^k+tP^k)$，$k=k+1$，转步骤(2)。

例 5.34

$$
\begin{aligned}
&\min f(X) = x_1^2 + x_2^2 + 2x_2 + 5 \\
s.t.\quad & \\
&x_1 - 2x_2 \geq 0 \\
&x_j \geq 0, \quad j = 1, 2.
\end{aligned}
\tag{5-6-100}
$$

初始点$X_1 = (2, 0)^T$。

解

$$g = \nabla f(X) = (2x_1, 2x_2 + 2)^T$$

第一次迭代：

(1)求P_1

$$g_1 = (4,2)^T,$$

$$N_1 = (0,1),$$

$$M_1 = (N_1 N_1^T)^{-1} = \Big((0,1)(0,1)^T\Big)^{-1} = 1,$$

$$Q_1 = I - N_1^T M_1 N_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot 1 \cdot (0,1) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$P_1 = -Q_1 g_1 = -\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} (4,2)^T = (-4,0)^T,$$

$$\|P_1\| \neq 0;$$

(2)求X_2

$$A_1'' = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix},$$

$$b_1'' = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$u = A_1''X_1 - b_1'' = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

$$v = A_1''P_1 = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -4 \\ 0 \end{pmatrix} = \begin{pmatrix} -4 \\ -4 \end{pmatrix},$$

$$\overline{t_1} = \min\{-\frac{u_j}{v_j} | v_j < 0\} = \min\{-\frac{2}{-4}, -\frac{2}{-4}\} = \frac{1}{2},$$

求

$$\begin{aligned}
&\min_{0 \le t_1 \le \overline{t_1}} f(X_1 + t_1 P_1) \\
&= \min_{0 \le t_1 \le \frac{1}{2}} f\left(\left(2 + t_1(-4), 0 + t_1(0)\right)^T\right) \\
&= \min_{0 \le t_1 \le \frac{1}{2}} (2 - 4t_1)^2 + 5
\end{aligned}$$

得

$$t_1^* = \frac{1}{2}$$

$$X_2 = (0, 0)^T$$

$$g_2 = (0, 2)^T$$

第二次迭代：

(1)求P_2

$$N_2 = \begin{pmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, N_2 = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}, x_1 \geq 2x_2, x_2 \geq 0,$$

$$M_2 = (N_2 N_2^T)^{-1} = \begin{pmatrix} 5 & -2 \\ -2 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix},$$

$$\begin{aligned} Q_2 &= I - N_2^T M_2 N_2 \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

$$P_2 = -Q_2 g_2 = (0, 0)^T,$$

$$\|P_2\| = 0,$$

$$N_2 \neq \emptyset;$$

(2)修正P_2

$$q = M_2 N_2 g_2 = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

因为$q > 0$，所以X_2为KT点；又因为问题为**凸规划**，因此最优解$X^* = (0, 0)^T$。 ∎

第六章　直接搜索方法

目标函数不一定可导，或没有显式解析表达式，针对此类问题，人们提出了直接搜索法($Direct\ Search$)，该方法仅仅利用到问题的目标函数值，因此此方法有称免导数方法（$Derivative-free$）。

§ 6.1 步长加速法

1961年 由*Hooke*和*Jeeves*提出， 又称之为模式搜索法（*Pattern Search*）。针对无约束优化问题

$$\min_{X \in R^n} f(X) \tag{6-1-1}$$

6.1.1 基本思想

步长加速法由**探测搜索**和**模式移动**组成。探测搜索是在出发点（**参考点**）的周围寻找比它更好的点，从而确定一个有利的前进方向（**基点**）。模式移动则是从基点出发沿着有利

的方向进行加速，得到新的参考点。该方法就是反复探测再移动，从而使得迭代点向极小点移动。

6.1.2 探索性移动

取初始点$X^0 \in R^n$，初始步长$\delta > 0$，收缩因子$\alpha(0.1 \sim 0.5)$，加速因子$\beta(1 \sim 2)$，令

$$e^k \triangleq (0, \cdots, 0, 1_k, 0, \cdots, 0)^T, k = 1 \sim n. \tag{6-1-2}$$

显然$e^k, k = 1 \sim n$是n个互相正交的方向。

现在用Y表示参考点，Z表示基点。

令$Y^1 = X^0$，$Z^1 = Z^0 = X^0$，以Y^1为参考点，令$L^k \triangleq Y^k - \delta e^k$，$U^k \triangleq Y^k + \delta e^k$，则沿$n$个方向进行如下探测

$$Y^{k+1} = \begin{cases} Y^k + \delta e^k, & f(U^k) < f(Y^k); \\ Y^k - \delta e^k, & f(L^k) < f(Y^k) \le f(U^k); \\ Y^k, & f(L^k) \ge f(Y^k) \le f(U^k); \end{cases} \tag{6-1-3}$$

若$f(Y^1) \ge f(Z^1)$，$\delta = \alpha\delta$，从Y^1开始，重新寻找基点。

若$f(Y^1) < f(Z^1)$，则令新的基点为Y^{n+1}，这时可考虑沿方向$Z^1 - Y^1$进行搜索，以期望能够得到更好的下降，即

$$Y^1 = Z^1 + \beta(Z^1 - Y^1) = 2Z^1 - Y^1 (\beta = 1). \tag{6-1-4}$$

6.1.3 *Hooke-Jeeves*步长加速法

见算法6.1。

算法 6.1 $Hooke-Jeeves$步长加速法

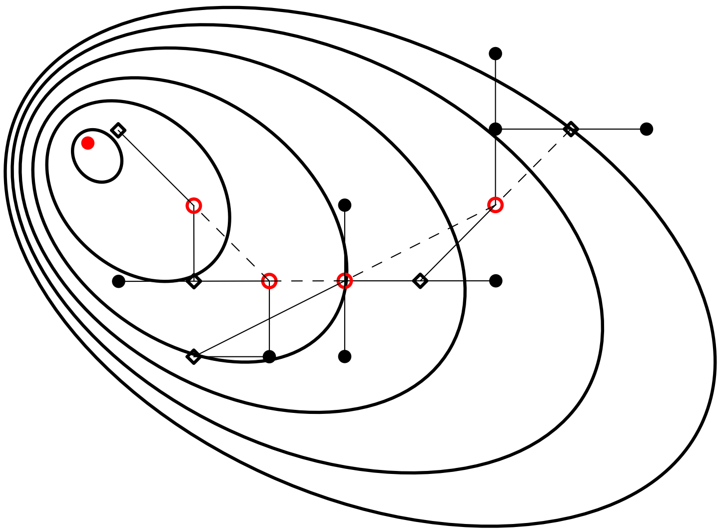
步骤 1 选定$Y^1 = X^0$，$Z^1 = Z^0 = X^0$，$L^k \triangleq Y^k - \delta e^k$，$U^k \triangleq Y^k + \delta e^k$，$k = 1$。

步骤 2 若$f(U^k) < f(Y^k)$，则令$Y^{k+1} = Y^k + \delta e^k$，转步骤(3)；否则若$f(L^k) < f(Y^k)$，则令$Y^{k+1} = Y^k - \delta e^k$，转步骤(3)；若$f(L^k) \geq f(Y^k)$，则令$Y^{k+1} = Y^k$，转步骤(3)。

步骤 3 若$k < n$，则令$k = k + 1$，转步骤(2)；否则$k = n$，若$f(Y^{n+1}) < f(Z^1)$，则令$Z^1 = Y^{n+1}$，转步骤(5)；否则若$f(Y^{n+1}) \geq f(Z^1)$，$Z^0 = Z^1$，再若$\delta \leq \varepsilon$，则停止，$X^* = Z^1$；否则令$\delta = \alpha\delta$，$Y^1 = Z^1$，$k = 1$，转步骤(2)。

步骤 4 若$f(Y^{n+1}) \geq f(Z^1)$，$Z^0 \neq Z^1$，则令$Y^1 = Z^1$，$Z^0 = Z^1$，$k = 1$，转步骤(2)。

步骤 5 令$Y^1 = 2Z^1 - Y^1$，$k = 1$，转步骤(2)。



§ 6.2 *Powell*方向加速法

1964年由$Powell$提出，研究具有对称正定的二次函数

$$f(X)=\frac{1}{2}X^TQX+b^TX+c, \tag{6-2-5}$$

其基本思想是在不用导数的情况下，在迭代中采用精确一维搜索的条件下逐步构造Q共轭方向。

6.2.1 基本算法

见算法6.2。

算法 6.2 *Powell*方法

步骤 1 选定X^0，置$P^i=e_{i+1}$，$i=0\sim n-1$。

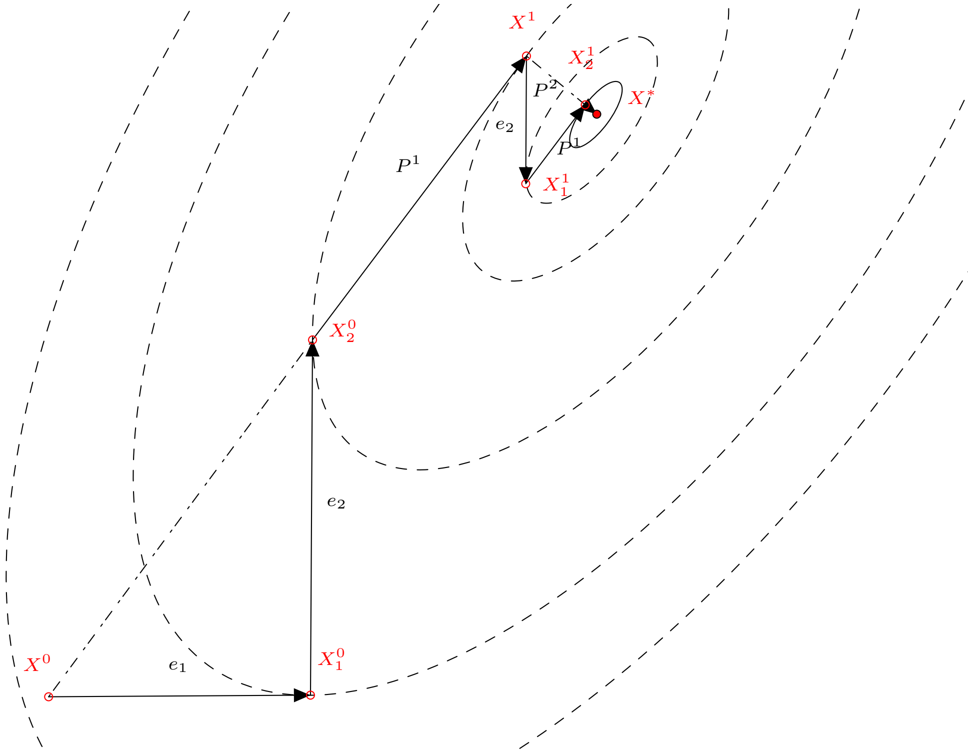
步骤 2 对$\forall i=0\sim n-1$，依次对目标函数作直线搜索$\min f(X^i+\lambda P^i)$。

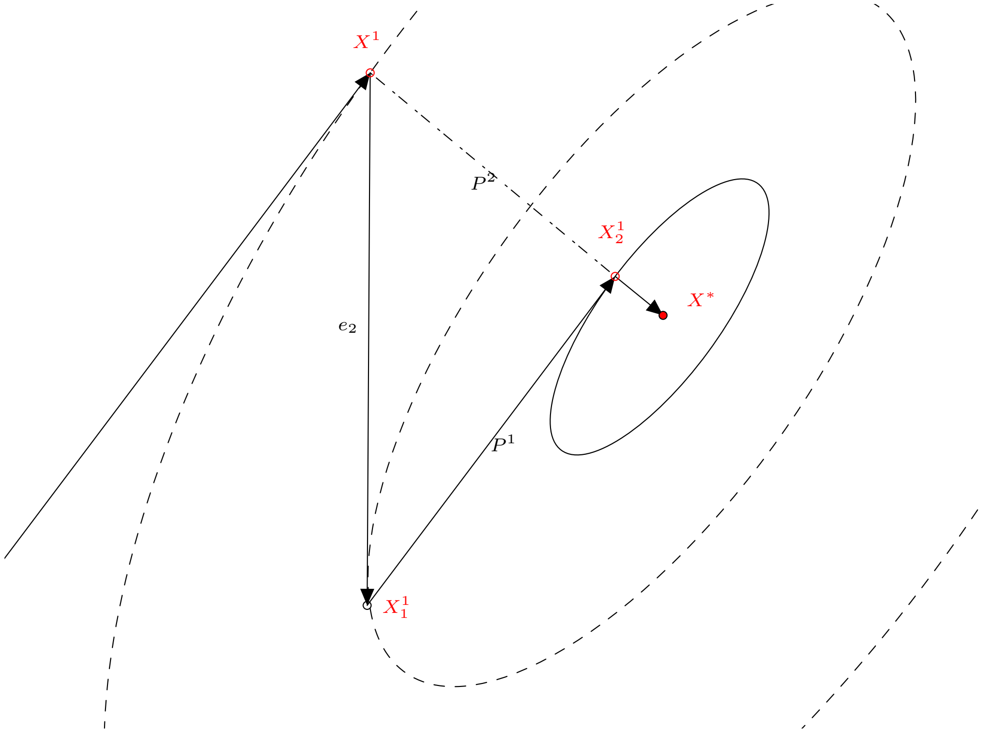
步骤 3 对$\forall i=0\sim n-2$，置$P^i=P^{i+1}$，$P^{n-1}=X^n-X^0$。

步骤 4 作直线搜索$\min f(X^n+\lambda P^{n-1})$。

步骤 5 若$\|X^{n+1}-X^0\|<\varepsilon$，则$X^{n+1}$就是最优解，停机。否则转步骤(6)。

步骤 6 置$X^0=X^{n+1}$。转步骤(2)。


X^1
X_2^1
P^2
X^*
e_2
P^1
P^1
X_1^1
X_2^0
e_2
X^0
e_1
X_1^0


X^1
P^2
X_2^1
X^*
e_2
P^1
X_1^1

因此理论上对于正定二次函数$Powell$法至多次n迭代就会求得最优解，但是一味的简单更替可能导致所得到的n个方向是线性相关的，因而不能构成为线性空间，所得到的极小点只能是某一维数小于n的线性流形上的极小点，从而导致算法失效。

定义 6.1 (线性流形) 设V是向量空间，$L \subseteq V$，$L \neq \emptyset$，若$\exists v \in V$，使得

$$L + v = \{l + v \mid l \in L\} \tag{6-2-6}$$

是V的子空间，那么L是一个线性流形，并且$dim(L) = dim(L+v)$。若$dim(L) = dim(V) - 1$，则L为超平面。

例 6.2

$$\min f(X) = (x_1 - x_2 + x_3)^2 + (-x_1 + x_2 + x_3)^2 + (x_1 + x_2 - x_3)^2$$

这里，$X^0 = (\frac{1}{2}, 1, \frac{1}{2})^T$。

解 $e^1 = (1, 0, 0)^T$

$X^1 = (\frac{1}{2}, 1, \frac{1}{2})^T$

$e^2 = (0, 1, 0)^T$

$X^2 = (\frac{1}{2}, \frac{1}{3}, \frac{1}{2})^T$

$e^3 = (0, 0, 1)^T$

$X^3 = (\frac{1}{2}, \frac{1}{3}, \frac{5}{18})^T$

$P^4 = X^3 - X^0 = (0, -\frac{2}{3}, -\frac{2}{9})^T$

因此在随后的迭代中，永远也不会收敛到问题的最优解$X^* = (0, 0, 0)^T$，往往得到维数小于3的线性流形上的极小点。 ∎

原因： e^2，e^3，P^4线性相关。

处理： 从$n + 1$个方向中选出最好的n个方向。

6.2.2 共轭程度的判别

考虑

$$\Delta = |P^1 \times P^2| = |det(P^1, P^2)|, \tag{6-2-7}$$

$$\Delta = |(P^1 \times P^2) \cdot P^3| = |det(P^1, P^2, P^3)|. \tag{6-2-8}$$

定义 6.3 设P^1，$\cdots$，P^n是R^n空间的n个向量，若其中有零向量，则定义其正交程度Δ为0；否则，定义为

$$\begin{aligned}\Delta &= \left| det(\frac{P^1}{\|P^1\|}, \cdots, \frac{P^n}{\|P^n\|}) \right| \\ &= \frac{|det(P^1, \cdots, P^n)|}{\prod_{i=1}^n \|P^i\|}\end{aligned} \tag{6-2-9}$$

这样衡量P^1，$\cdots$，P^n的正交成都是合理的，因为

定理 6.4 设P^1，$\cdots$，P^n是R^n空间的n个向量，则其正交程度

$$\Delta(P^1, \cdots, P^n) \leq 1. \tag{6-2-10}$$

证明 令$B = (P^1, \cdots, P^n)$，显然当P^i线性相关时，$det(B) = 0$，所以$\Delta = 0$。

当P^i线性无关时，不妨假设

$$\|P^i\| = 1, i = 1 \sim n. \tag{6-2-11}$$

因为

$$A = B^T B = ((P^i)^T P^j)_{n\times n} = \begin{pmatrix} (P^1)^T P^1 & \cdots & (P^1)^T P^n \\ \vdots & \vdots & \vdots \\ (P^n)^T P^1 & \cdots & (P^n)^T P^n \end{pmatrix}, \tag{6-2-12}$$

所以

$$det^2(B) = det(A). \tag{6-2-13}$$

设A的特征根为λ_1，$\cdots$，λ_n，因此

$$det(A) = \prod_{i=1}^{n} \lambda_i, \tag{6-2-14}$$

$$trace(A) = \sum_{i=1}^{n} \lambda_i = n. \tag{6-2-15}$$

所以

$$\left(\prod_{i=1}^{n} \lambda_i\right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^{n} \lambda_i. \tag{6-2-16}$$

当且仅当$\lambda_i = \lambda_j$时上式等号成立。因而

$$det^2(B) = det(A) \leq 1. \tag{6-2-17}$$

即

$$\Delta = \Delta(B) \leq 1. \tag{6-2-18}$$

当且仅当P^i为正交时上式等号成立。 ▮

作如下变换

$$P^i = \sqrt{Q}d^i, i = 1 \cdots, n. \tag{6-2-19}$$

则$P^1, \cdots, P^n$的正交程度就反映了$d^1, \cdots, d^n$的关于Q共轭程度。

定义 6.5 设Q是$n \times n$的正定矩阵，d^1，$\cdots$，d^n是R^n空间的n个向量，若其中有零向量，则定义其关于Q共轭程

度Δ为0；否则，定义为

$$\Delta = \frac{|det(\sqrt{Q})| \cdot |det(P^1, \cdots, P^n)|}{\prod_{i=1}^{n} \sqrt{(P^i)^T Q P^i}}. \tag{6-2-20}$$

6.2.3 *Powell*改进算法

第七章　进化计算

进化计算是一种模拟自然界生物进化过程与机制进行问题求解的自组织、自适应的随机搜索技术。它以达尔文（Darwin）进化论的“物竞天择、适者生存”作为算法的进化规则，并结合孟德尔（Mendel）的遗传变异理论，将生物进化过程中的繁殖（Reproduction）、变异（Mutation）、竞争（Competition）、选择（Selection）引入到了算法中。

§ 7.1 生物学基础

自然界生物进化过程是进化计算的生物学基础，它主要包

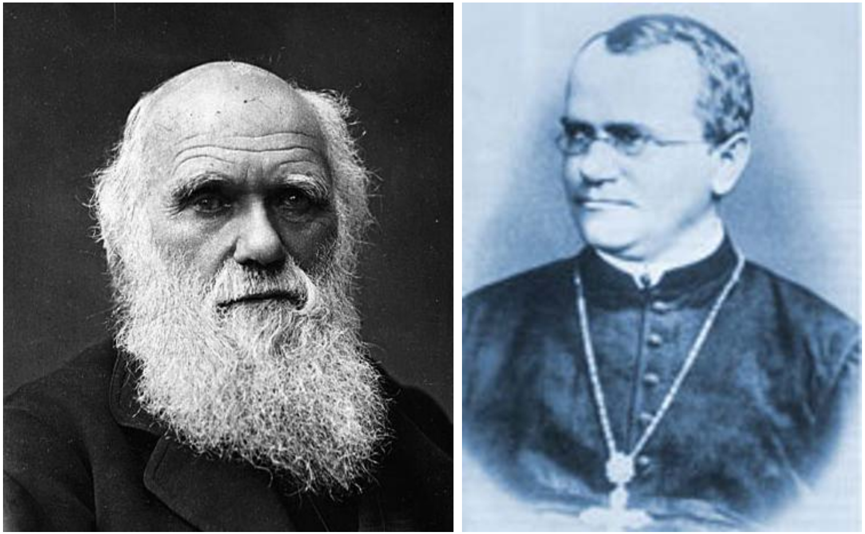
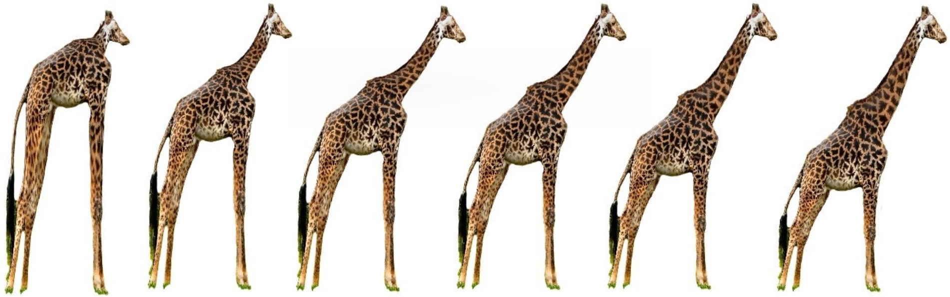


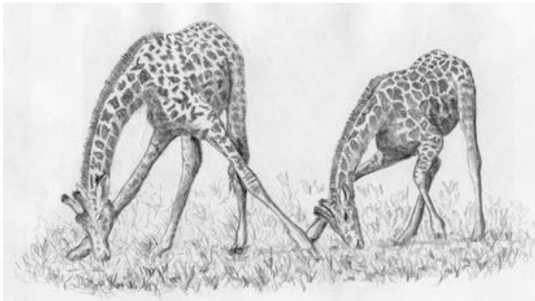
图 7.1 达尔文(1809-1882)，英国人，进化论之父。孟德尔(1822—1884)，奥地利人，遗传学的奠基人。

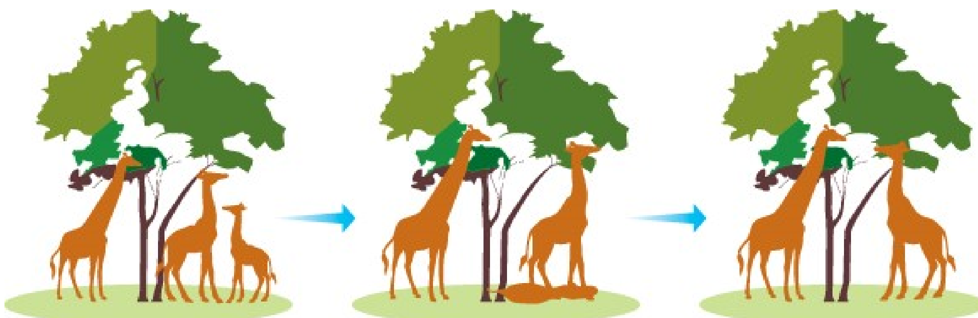


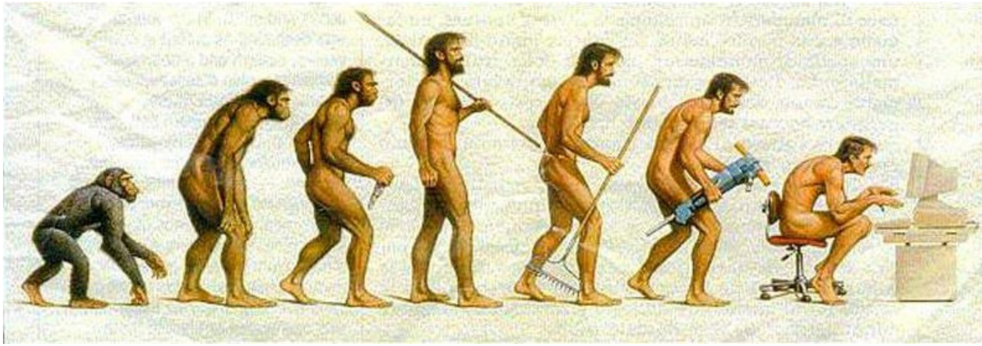
括**遗传**(Heredity)、**变异**(Mutation)、**进化**(Evolution)**理论**。

进化理论：

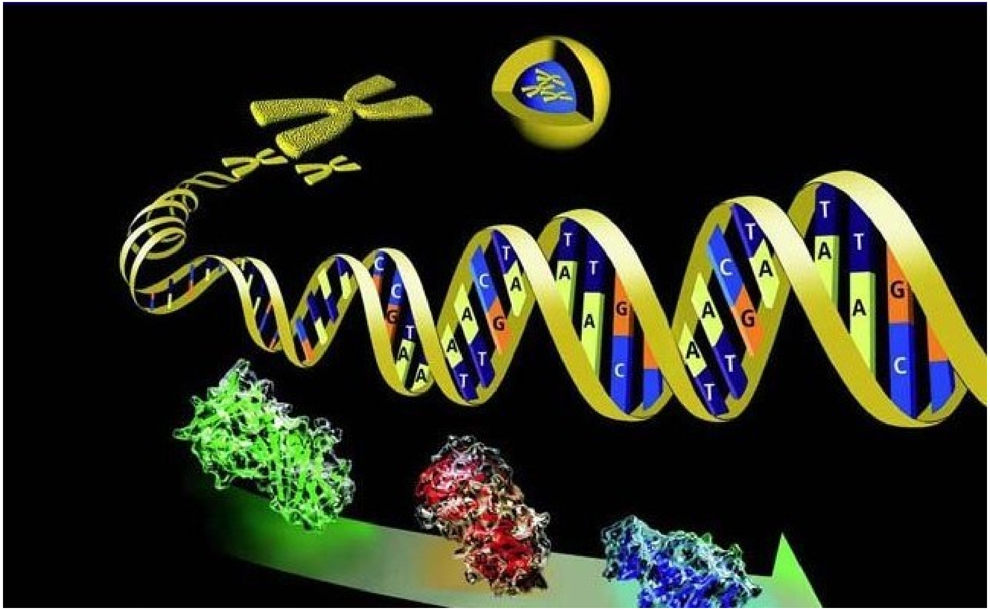
进化是指在生物延续生存过程中，逐渐适应其生存环境，使得其品质不断得到改良的这种生命现象。遗传和变异是生物进化的两种基本现象，优胜劣汰、适者生存是生物进化的基本规律。







达尔文的自然选择学说认为：在生物进化中，一种基因有可能发生变异而产生出另一种新的生物基因。这种新基因将依据其与生存环境的适应性而决定其增殖能力。一般情况下，适应性强的基因会不断增多，而适应性差的基因则会逐渐减少。通过这种自然选择，物种将逐渐向适应于生存环境的方向进化，甚至会演变成为另一个新的物种，而那些不适



应于环境的物种将会逐渐被淘汰。

在自然界，构成生物基本结构与功能的单位是细胞

（Cell）。细胞中含有一种包含着所有遗传信息的复杂而又微小的丝状化合物，人们称其为染色体（Chromosome）。在染色体中，遗传信息由基因（Gene）所组成，基因决定着生物的性状，是遗传的基本单位。染色体的形状是一种双螺旋结构，构成染色体的主要物质叫做脱氧核糖核酸(DNA)，每个基因都在DNA长链中占有一定的位置。一个细胞中的所有染色体所携带的遗传信息的全体称为一个基因组(Genome)。细胞在分裂过程中，其遗传物质DNA通过复制转移到新生细胞中，从而实现了生物的遗传功能。

7.1.1 遗传理论

遗传是指父代（或亲代）利用遗传基因将自身的基因信息传递给下一代（或子代)，使子代能够继承其父代的特征或性状的这种生命现象。正是由于遗传的作用，自然界才能有稳定的物种。

7.1.2 变异理论

变异是指子代和父代之间，以及子代的各个不同个体之间产生差异的现象。变异是生物进化过程中发生的一种随机现象，是一种不可逆过程，在生物多样性方面具有不可替代的

图 7.2 “种瓜得瓜，种豆得豆”，“龙生龙，凤生凤，老鼠生儿打地洞”

作用。引起变异的主要原因有以下两种：

杂交: 指有性生殖生物在繁殖下一代时两个同源染色体之间的交配重组，即两个染色体在某一相同处的DNA被切断后再进行交配重组，形成两个新的染色体。

复制差错: 指在细胞复制过程中因DNA上某些基因结构的随机改变而产生出新的染色体。

§ 7.2 基本概念

20世纪50年代后期，一些生物学家在研究如何用计算机模拟生物遗传系统中，产生了遗传算法的基本思想，并于1962年由美国密执安（Michigan）大学霍兰德（Holland）

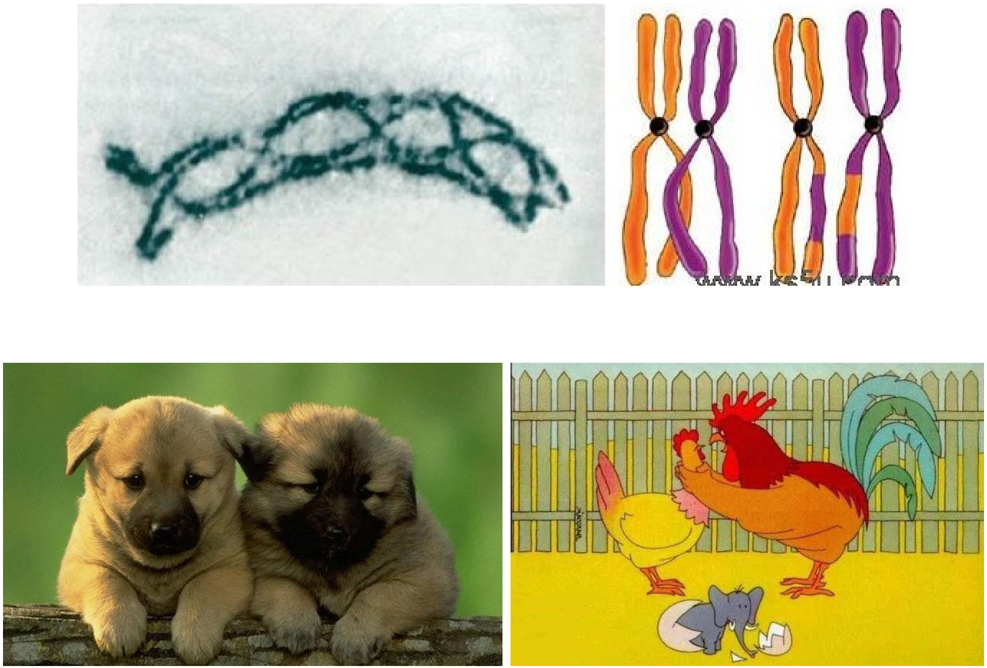


图 7.3 “龙生九子，九子九样”

图 7.4 "Prof. Holland，遗传算法提出者。"

提出。

遗传算法的基本思想是从初始种群出发，采用优胜劣汰、适者生存的自然法则选择个体，并通过杂交、变异来产生新一代种群，如此逐代进化，直到满足目标为止。

遗传算法所涉及到的基本概念主要有以下几个：

种群（Population）：种群是指用遗传算法求解问题时，初始给定的多个解的集合。遗传算法的求解过程是从这个子集开始的。

个体（Individual）：个体是指种群中的单个元素，它通常由一个用于描述其基本遗传结构的数据结构来表示。例如，可以用0、1组成的长度为l的串来表示个体。

染色体（Chromos）：染色体是指对个体进行编码后所得到的编码串。染色体中的每1位称为基因，染色体上由若干个基因构成的一个有效信息段称为基因组。

适应度（Fitness）函数：适应度函数是一种用来对种群中各个个体的环境适应性进行度量的函数。其函数值是遗传算法实现优胜劣汰的主要依据。

遗传操作（Genetic Operator）:遗传操作是指作用于种群而产生新的种群的操作。标准的遗传操作包括以下三种基本形式：选择（Selection）、交叉（Crosssover）、变异（Mutation）。

§ 7.3 遗传算法

遗传算法主要由染色体编码、初始种群设定、适应度函数设定、遗传操作设计等几大部分所组成，其算法主要内

容和基本步骤可描述如下：若用$X(t)$表示第t代种群，则标准GA的迭代过程见算法7.1.

算法 7.1 遗传算法

步骤 1 (初始化)确定选择概率P_s，种群规模$Popsize$，交叉概率P_c，变异概率P_m和终止原则，令$t=0$，随机产生初始种群$X(t)$.

步骤 2 (个体评价)计算或估价$X(t)$中个体的适应度.

步骤 3 (选择)从当前种群$X(t)$中选取母体.

步骤 4 (交叉)独立地对所选母体实施杂交生成个中间个体.

步骤 5 (变异)独立地对中间个体进行变异得到个候选个体.

步骤 6 (选择)从候选个体中依适应度高低选出个个体组成下一代新种群$X(t+1)$.

步骤 7 (终止检验)若满足终止原则，则停止；否则并返回步骤2.

7.3.1 遗传编码

常用的遗传编码算法有二进制码、格雷码（Gray Code）、实数编码和字符编码等。

二进制编码（Binary encoding）

二进制编码是将原问题的结构变换为染色体的位串结构。在二进制编码中，首先要确定二进制字符串的长度l，该长度与变量的定义域和所求问题的计算精度有关。

例 7.1 假设变量x的定义域为[5，10]，要求的计算精度为10E-5，则需要将[5，10]至少分为600000个等长小区间，每个小区间用一个二进制串表示。于是，串长至少等于20，

原因是：524288=219¡600000¡220=1048576 这样，对应于区间[5，10]内满足精度要求的每个值x，都可用一个20位编码的二进制串¡b19,b18,···,b0¿来表示。二进制编码存在的主要缺点是汉明（Hamming）悬崖。例如，7和8的二进制数分别为0111和1000，当算法从7改进到8时，就必须改变所有的位。

7.3.2 适应度函数

适应度函数是一个用于对个体的适应性进行度量的函数。通常，一个个体的适应度值越大，它被遗传到下一代种群中的概率也就越大。

标准适应度函数在遗传算法中，一般要求适应度函数非负，并其适应度值越大越好。这就往往需要对原始适应函数进行某种变换，将其转换为标准的度量方式，以满足进化操作的要求，这样所得到的适应度函数被称为标准适应度函数fNormal(x)。

7.3.3 轮盘赌选择

轮盘赌选择法又被称为转盘赌选择法或轮盘选择法。在这种方法中，个体被选中的概率取决于该个体的相对适应度。

7.3.4 交叉操作

交叉（Crossover）操作是指按照某种方式对选择的父代个体的染色体的部分基因进行交配重组，从而形成新的个体。

单点交叉

单点交叉也称简单交叉，它是先在两个父代个体的编码串中随机设定一个交叉点，然后对这两个父代个体交叉点前面或后面部分的基因进行交换，并生成子代中的两个新的个体。

例 7.2 设有两个父代的个体串$A = 0\ 0\ 1\ 1\ 0\ 1$ 和$B = 1\ 1\ 0\ 0\ 1\ 0$，若随机交叉点为4，则交叉后生成的两个新的个

体是：$A. = 0\ 0\ 1\ 1\ 1\ 0\ B. = 1\ 1\ 0\ 0\ 0\ 1$

7.3.5 变异操作

变异（Mutation）是指对选中个体的染色体中的某些基因进行变动，以形成新的个体。

二进制变异

当个体的染色体采用二进制编码表示时，其变异操作应采用二进制变异方法。该变异方法是先随机地产生一个变异位，然后将该变异位置上的基因值由“0”变为“1”，或由“1”变为“0”，产生一个新的个体。

例 7.3 设变异前的个体为$A = 0\ 0\ 1\ 1\ 0\ 1$，若随机产生的变异位置是2，则该个体的第2位由“0”变为“1”。变异后的新的个体是$A. = 0\ 1\ 1\ 1\ 0\ 1$。

§ 7.4 遗传算法应用简例

例 7.4 用遗传算法求函数$f(x) = x^2$的最大值，其中x为$[0，31]$间的整数。

解 (1) 编码

由于的定义域是区间$[0，31]$上的整数，由5位二进制数即可全部表示。因此，可采用二进制编码方法，其编码

串的长度为5。例如，用二进制串00000来表示=0,11111来表示=31等。其中的0和1为基因值。

(2) 生成初始种群

若假设给定的种群规模=4，则可用4个随机生成的长度为5的二进制串作为初始种群。再假设随机生成的初始种群（即第0代种群）为：S01=0 1 1 0 1 S02=1 1 0 0 1 S03=0 1 0 0 0 S04=1 0 0 1 0

(3) 计算适应度

要计算个体的适应度，首先应该定义适应度函数。由于本例是求的最大值，因此可直接用来作为适应度函数。即：

其中的二进制串S对应着变量的值。根据此函数，初始种群中各个个体的适应值及其所占比例如表5-5所示。表1 初始种群情况表编号个体串（染色体）适应值百分比S01 0 1 1 0 1 13 169 14.44 14.44 1 S02 1 1 0 0 1 25 625 52.88 67.18 2 S03 0 1 0 0 0 8 64 5.41 72.59 0 S04 1 0 0 1 0 18 324 27.41 100 1 可以看出，在4个个体中S03的适应值最大，是当前最佳个体。

(4) 选择操作

假设采用轮盘赌方式选择个体，且依次生成的4个随机数（相当于轮盘上指针所指的数）为0.85、0.32、0.12和0.46，经选择后得到的新的种群为：S01=1 0 0 1 0 S02=1 1 0 0 1

S03=0 1 1 0 1 S04=1 1 0 0 1 其中，染色体1 1 0 0 1在种群中出现了2次，而原染色体0 1 0 0 0则因适应值太小而被淘汰。

(5) 交叉

假设交叉概率Pi为50%，则种群中只有1/2的染色体参与交叉。若规定种群中的染色体按顺序两两配对交叉，且有S01与S02交叉，S03与S04不交叉，则交叉情况如表5-6所示。表2 初始种群的交叉情况表编号个体串（染色体）交叉对象交叉位子代适应值S01 1 0 0 1 0 S02 3 1 0 0 0 1 289 S02 1 1 0 0 1 S01 3 1 1 0 1 0 676 S03 0 1 1 0 1 S04 N 0 1 1 0 1 169 S04 1 1 0 0 1 S03 N 1 1 0 0 1 625

可见，经交叉后得到的新的种群为：S01=1 0 0 0 1 S02=1 1 0 1 0 S03=0 1 1 0 1 S04=1 1 0 0 1

(6) 变异

变异概率Pm一般都很小，假设本次循环中没有发生变异，则变异前的种群即为进化后所得到的第1代种群。即：S11=1 0 0 0 1 S12=1 1 0 1 0 S13=0 1 1 0 1 S14=1 1 0 0 1 然后，对第1代种群重复上述(4)-(6)的操作。

对第1代种群，同样重复上述(4)-(6)的操作。其选择情况如表5-7所示。表3 第1代种群的选择情况表编号个体串（染色体）适应值百分比S11 1 0 0 0 1 27 289 16.43 16.437 1 S12 1

1 0 1 0 26 676 38.43 54.86 2 S13 0 1 1 0 1 13 169 9.61 64.47 0 S14 1 1 0 0 1 25 625 35.53 100 1 其中若假设按轮盘赌选择时依次生成的4个随机数为0.14、0.51、0.24和0.82，经选择后得到的新的种群为：S11=1 0 0 0 1 S12=1 1 0 1 0 S13=1 1 0 1 0 S14=1 1 0 0 1 可以看出，染色体1 1 0 1 0被选择了2次，而原染色体0 1 1 0 1则因适应值太小而被淘汰。

对第1代种群，其交叉情况如表5-8所示。表5-8 第1代种群的交叉情况表编号个体串（染色体）交叉对象交叉位子代适应值S11 1 0 0 0 1 S12 3 1 0 0 1 0 324 S12 1 1 0 1 0 S11 3 1 1 0 0 1 625 S13 1 1 0 1 0 S14 2 1 1 0 0 1 625 S14 1 1 0 0 1 S13

2 1 1 0 1 0 675 可见，经杂交后得到的新的种群为：S11=1 0 0 1 0 S12=1 1 0 0 1 S13=1 1 0 0 1 S14=1 1 0 1 0 可以看出，第3位基因均为0，已经不可能通过交配达到最优解。这种过早陷入局部最优解的现象称为早熟。为解决这一问题，需要采用变异操作。

对第1代种群，其变异情况如表5-9所示。表5-9 第1代种群的变异情况表

编号个体串（染色体）是否变异变异位子代适应值S11 1 0 0 1 0 N 1 0 0 1 0 324 S12 1 1 0 0 1 N 1 1 0 0 1 625 S13 1 1 0 0 1 N 1 1 0 0 1 625 S14 1 1 0 1 0 Y 3 1 1 1 1 0 900 它是通

过对S14的第3位的变异来实现的。变异后所得到的第2代种群为：S21=1 0 0 1 0 S22=1 1 0 0 1 S23=1 1 0 0 1 S24=1 1 1 1 0

对第2代种群，同样重复上述(4)-(6)的操作。其选择情况如表5-10所示。表5-10 第2代种群的选择情况表编号个体串（染色体）适应值百分比S21 1 0 0 1 0 18 324 23.92 23.92 1 S22 1 1 0 0 1 25 625 22.12 46.04 1 S23 1 1 0 0 1 25 625 22.12 68.16 1 S24 1 1 1 1 0 30 900 31.84 100 1 其中若假设按轮盘赌选择时依次生成的4个随机数为0.42、0.15、0.59和0.91，经选择后得到的新的种群为：S21=1 1 0 0 1 S22=1 0 0 1 0 S23=1 1 0 0 1 S24=1 1 1 1 0

对第2代种群，其交叉情况如表5-11所示。编号个体串（染色体）交叉对象交叉位子代适应值S21 1 1 0 0 1 S22 3 1 1 0 1 0 676 S22 1 0 0 1 0 S21 3 1 0 0 0 1 289 S23 1 1 0 0 1 S24 4 1 1 0 0 0 576 S24 1 1 1 1 0 S23 4 1 1 1 1 1 961 这时，函数的最大值已经出现，其对应的染色体为1 1 1 1 1，经解码后可知问题的最优解是在点=31处。求解过程结束。

§ 7.5 模拟退火

1982年，Kirkpatrick等将金属热加工中的退火工艺的思想应用于组合优化中，提出了一种新的搜索技术：模拟退火(Simulated Annealing，SA)算法．模拟退火采用Metropolis

接受准则，并使用一组称为冷却进度表的参数控制算法进程，使算法在多项式时间内给出一个近似最优解.

该方法包括两种运算：(1) 当前状态的变化；(2) 变化的接受和舍弃. 模拟退火算法的思想是允许以一定的概率从现有解移动到一个较差的解，以期能够跳出局部最小. 这种移动的允许概率随着搜索的进行逐渐减小.

算法7.2中，$gen(T_t, t)$和$gen(L_t, t)$分别表示第t次迭代温度控制、生成次数的函数发生器. 算法从一个初始解开始，该初始解可以是随机产生的，也可以是按照某种启发式方法产生的. 同时初始化温度参数. 然后，在每一次迭代中，从邻

算法 7.2 模拟退火算法

步骤 1 选定一个初始解x_1，初始温度T_1，L_1，$t=1$.

步骤 2

for $k=1$ to L_t **do**

从领域$N(x_i)$中随机选取x_j

if $f(x_j)<f(x_i)$ **then**

$x_i=x_j$

else if $\exp\left(\frac{f(x_i)-f(x_j)}{T}\right)>rand(0,1)$ **then**

$x_i=x_j$

end if

end for

步骤 3 $T_{t+1}=gen(T_t,t)$，$L_{t+1}=gen(T_t,t)$，$t=t+1$；若满足终止条件，则停止计算；否则，转步骤2.

域中随机地抽取一个解x_j，然后根据$f(x_i)$，$f(x_j)$，以及T来决定足否接受这个解．如果$f(x_j) < f(x_i)$，则用x_j替代x_i，否则以波尔兹曼(Boltzmann)概率分布接受x_j．

§ 7.6 蚁群算法

蚁群算法(Ant Colony Optimization，ACO)由意大利学者Dorigo等人提出．其充分利用了蚁群搜索食物的过程与著名的旅行商问题(Traveling Salesman Problem，TSP)之间的相似性，通过人工模拟蚂蚁搜索食物的过程，即通过个体之间的信息交流与相互协作最终找到从蚁穴到食物源的最短路径来求解TSP．蚂蚁这种社会性动物，虽然个体行为及其

简单，但是由这些简单个体所组成的群体却表现出及其复杂的行为特征．这是因为蚂蚁在寻找食物时，能在其经过的路径上释放一种叫做信息素的物质，使得一定范围内的其他蚂蚁能够感觉到这种物质，且倾向于朝着该物质强度高的方向移动．蚁群的集体行为表现为一种正反馈现象．

ACO算法通过一群蚂蚁在TSP拓扑图中的行走来共同寻求问题的解．在每一次迭代过程中，每个蚂蚁根据路径上的信息素分布情况，随机地选择下一个未曾走过的城市进行访问，直至完成对于所有城市的访问，形成问题的一个解．具体来说，在第t次迭代中，第k个蚂蚁在第i个城市选择下一个

城市j的转移规则为：

$$j=\begin{cases} arg\max_{l\in N_i^k}\{\tau_{i,l}(t)\cdot\eta_{i,l}^{\beta}(t)\}, & \text{如果} rand(0,1)<P; \\ J, & \text{否则}. \end{cases} \tag{7-6-1}$$

其中，$\tau_{i,j}$是分布在路径$i\to j$上的信息素，$\eta_{i,j}$是该路径上的局部启发式值，一般取值为该路径长度的倒数，即$\eta=\frac{1}{d_{i,j}}$，P为$0\leq P\leq 1$的参数，N_i^k表示蚂蚁k还未曾走过的城市集合．$J\in N_i^k$是按照下面的概率公式所选择的一个随机城市：

$$p_{i,j}^k(t)=\frac{\tau_{i,j}(t)\cdot\eta_{i,j}^{\beta}(t)}{\sum_{l\in N_i^k}\tau_{i,l}(t)\cdot\eta_{i,l}^{\beta}(t)}. \tag{7-6-2}$$

在第t次迭代中．每个蚂蚁在从第i个城市转移到下一个城市j，在所经过的路径上进行局部信息素更新，其更新规则如下：

$$\tau_{i,j}(t+1) = (1-\rho)\tau_{i,j}(t) + \rho(\frac{1}{L_{Gbest}}), \forall(i,j) \in T_{Gbest}. \quad (7\text{-}6\text{-}3)$$

上式中，T_{Gbest}表示当前最优路径，L_{Gbest}表示T_{Gbest}的总路程长度．在所有的蚂蚁完成一次迭代过程后，ACO算法计算所有本次迭代所得解的质量，并将本次迭代最优解与全局历史最优解X_{Gbest}进行比较，取两者之间的较优解作为新

的X_{Gbest}．随后进行信息素的全局更新，更新规则为：

$$\tau_{i,j}(t+1) = (1-\rho)\tau_{i,j}(t) + \rho\tau_0. \tag{7-6-4}$$

通过这种方法，迭代过程将更有针对性的在现有最优解的基础上探索更好的解．算法流程见算法7.3.

§ 7.7 粒子群优化

粒子群优化(particle Swarm Optimization，PSO)算法最初是由Kennedy和Eberhart于1995年受人工生命研究结果启发，在模拟鸟群觅食过程中的迁徙和群集行为时提出的一种基于群体智能的演化计算技术．其思想来源于对鸟群飞行的研究

算法 7.3 蚁群算法

步骤 1 初始化.

步骤 2

while 终止条件未满足 **do.**

随机选取每一个蚂蚁的初始城市.

for $i = 1$ to n **do**

对于每一个蚂蚁，按用状态转移规则，随机选取下一个城市.

局部信息素更新.

end for

局部搜索法.

全局信息素更新.

end while

步骤 3 输出最优解.

发现：鸟仅仅是追踪它有限数量的邻居，但最终的整体结果是整个鸟群好像在一个中心的控制之下，即复杂的全局行为是由简单规则的相互作用引起的．不同于达尔文“适者生存，优胜劣汰”进化思想，粒子群优化算法是通过个体之间的协作来寻找最优解的．生物学家Wilson关于生物体曾经说过这样一段话：“至少在理论上，一个生物群体中的一员可以从这个群体中所有其它成员以往在找寻食物过程中积累的经验和发现中获得好处．只要食物源不可预知地分布于不同地方，这种协作带来的优势可能变成决定性的，超过群体中个体之间对食物竞争带来的劣势”．这段话的意思是说生物群体中信

息共享会产生进化优势，这也正是粒子群优化算法的基本思想.

为了改善基本PSO算法易发散的缺点，Eberhart和Shi提出了一种惯性粒子群优化算法. 该算法通过动态调整惯性系数的方式有效地提高算法的收敛性能，因而常被称为标准的PSO算法，惯性粒子群优化算法见算法7.4.

算法 7.4 粒子群优化算法

步骤 1 初始化种群P，速度V，认知系数c_1，社会系数c_2，惯性权重w. 令$t=0$.

步骤 2 计算每一个粒子$X_i \in P(t)$的适应值，经历最优位置$P_i(t)$，种群全局最优位置$G(t)$.

步骤 3 更新每一个粒子$X_i \in P(t)$的速度$V_i(t)$和位置$X_i(t+1)$:

for $X_i(t) \in P(t)$ **do**

  for $j = 1$ to n **do**

$$
\begin{aligned}
V_{i,j}(t+1) = & \, w \cdot X_{i,j}(t) \\
& + c_1 \cdot rand(0,1) \cdot (P_{i,j}(t) - X_{i,j}(t)) \\
& + c_2 \cdot rand(0,1) \cdot (G_j(t) - X_{i,j}(t)) \\
X_i(t+1) = & \, X_i(t) + V_i(t+1).
\end{aligned}
$$

  end for

end for

步骤 4 终止条件是否满足？若是，停机；否则，$t=t+1$，转步骤2.

第八章　多目标最优化

第九章　$MATLAB$优化工具包

参考文献