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://faculty.uestc.edu.cn/zhangxiaowei

Version: 20210905223500

前 言

最优化理论与方法讲义[1–5] . . . .

目　录

算 法

勘误

(1)、 例1.2中，设由A_i到B_j的运输量为x_{ij}吨。

(2)、 定理1.1中，$\boldsymbol{e}^i=(0,\cdots,0,1_{[i]},0,\cdots,0)^T$。

(3)、 定理1.2中，$\phi(t)=f(\boldsymbol{X}^0+t\boldsymbol{P})$。

(4)、 定理1.8中，$\lim_{\alpha\to 0^+}\cdots=\nabla f(\boldsymbol{X}^1)^T(\boldsymbol{X}^2-\boldsymbol{X}^1)$。

(5)、 定理1.10中，$f\big(\alpha\boldsymbol{X}+(1-\alpha)\boldsymbol{Y}\big)<\alpha f(\boldsymbol{X})+(1-\alpha)f(\boldsymbol{X})$。

(6)、 定理1.11中，$f\big(\alpha\boldsymbol{X}^1+(1-\alpha)\boldsymbol{X}^2\big)<\alpha f(\boldsymbol{X}^1)+(1-\alpha)f(\boldsymbol{X}^2)$。

(7)、 定理1.12中，$f\big(\boldsymbol{X}^\star+\alpha(\boldsymbol{X}^0-\boldsymbol{X}^\star)\big)-f(\boldsymbol{X}^\star)=\cdots<0$。

(8)、 修改了换基运算的推导逻辑。

(9)、 换基运算中，若$\sigma_l>0$，$a_{kl}>0$，则利用$\cdots$第l列变成$\cdots$。

(10)、 修改了§ 2.4中部分例题。

(11)、 定理2.8中，$f(\boldsymbol{X}^\theta)=\cdots=\sum_{i=1}^m c_i(b_i-\theta a_{il})+c_l\theta$。

(12)、 在大M法中，最优解改为$\begin{pmatrix}X^\star\\ X_a^\star\end{pmatrix}$。

(13)、 $3-1-7$，$3-1-8$中，min 改为max 。

(14)、 修改了例子、定义、定理等环境的计数规则：(a)例子、练习共同计数；(b)定理、引理、公理、推论、性质共同计数；(c)定义单独使用计数器。

(15)、 修改了已知明显角标错误。

第一部分 算法篇

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，用数学语言来说，就是求一个一元函数或多元函数的极值。

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)，$\cdots$

(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)，$\cdots$

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。

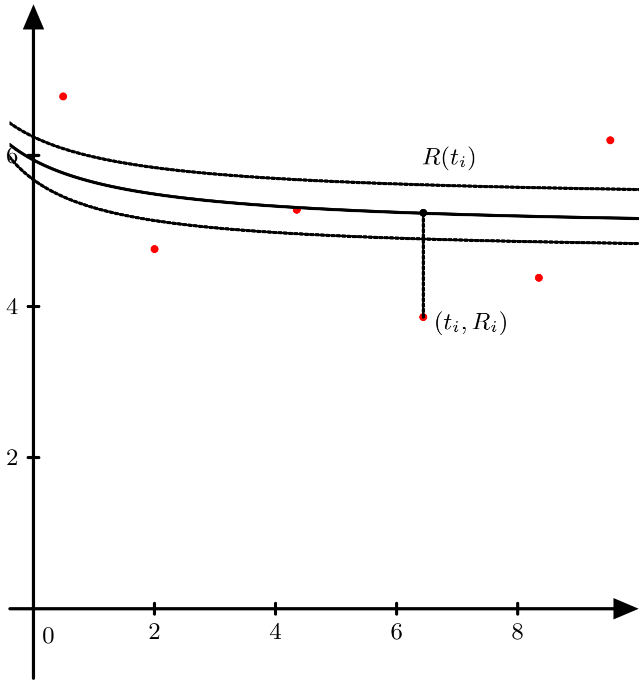
(2) 20世纪50年代中后期，钱学森、华罗庚、许国志等引入。1965年以后，华罗庚的“优选法”，“统筹法”。

下面通过具体例子来看看什么是最优化问题。

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)
6
4
2
0
2
4
6
8

解 函数关系R在几何上对应一条平面曲线。用所有测量点沿垂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A_i到B_j的运输量为x_{ij}吨，则有数学模型

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

$s.t.$

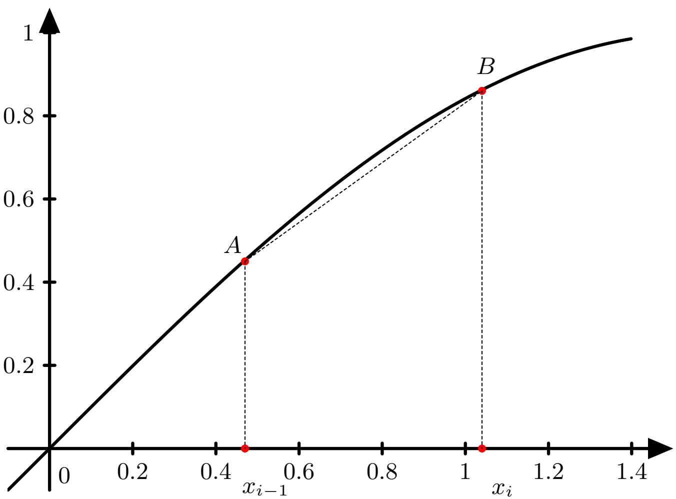
$$\sum_{j=1}^{n} x_{ij} = a_i, \quad i = 1, 2, \cdots, m$$

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

$$x_{ij} \geq 0$$

∎

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：

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

$s.t.$

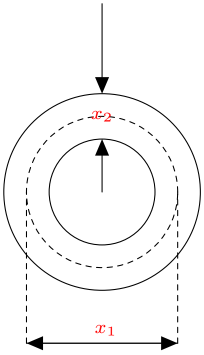
$$0=x_0<x_1<\cdots<x_7=\frac{\pi}{2}\quad 或$$

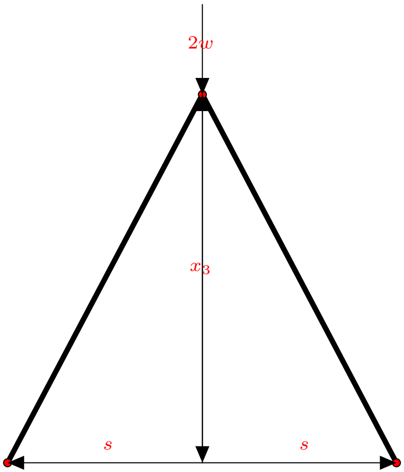
$$g_i(\boldsymbol{X})=x_i-x_{i-1}>0, x_i\in\left(0,\frac{\pi}{2}\right), i=1,\cdots,6$$

∎

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

$$x_3 \le h.$$

(2)钢管的压应力不能超过临界应力（弯曲应力）σ，即

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

综上所述，数学模型为：

$$\min \ x_1x_2\sqrt{s^2+x_3^2}$$

$$s.t.$$

$$x_3 \le h,$$

$$w\sqrt{s^2+x_3^2} \le \sigma\pi x_1x_2x_3,$$

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

∎

1.1.3 数学模型

(1)一般形式

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

定义 1.1 称$x_1, x_2, \cdots, x_n$为自变量或决策变量；$f(x_1, x_2, \cdots, x_n)$为目标函数；

$$\boldsymbol{S} = \left\{ \boldsymbol{X} \in \boldsymbol{R}^n \;\middle|\; \begin{array}{l} g_i(x_1,x_2,\cdots,x_n) \geq 0,\ i=1,2,\cdots,m; \\ h_j(x_1,x_2,\cdots,x_n)=0,\ j=1,2,\cdots,k. \end{array} \right\} \tag{1-1-2}$$

为可行集或可行域。

(2)向量形式

$$
\begin{aligned}
& \min \ f(\boldsymbol{X}) \\
s.t. & \\
& \boldsymbol{G}(\boldsymbol{X}) \geq \mathbf{0} \\
& \boldsymbol{H}(\boldsymbol{X}) = \mathbf{0}
\end{aligned}
\tag{1-1-3}
$$

这里，

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

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

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

$\mathbf{0}$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

(4)整数/实数

§ 1.2 数学基础

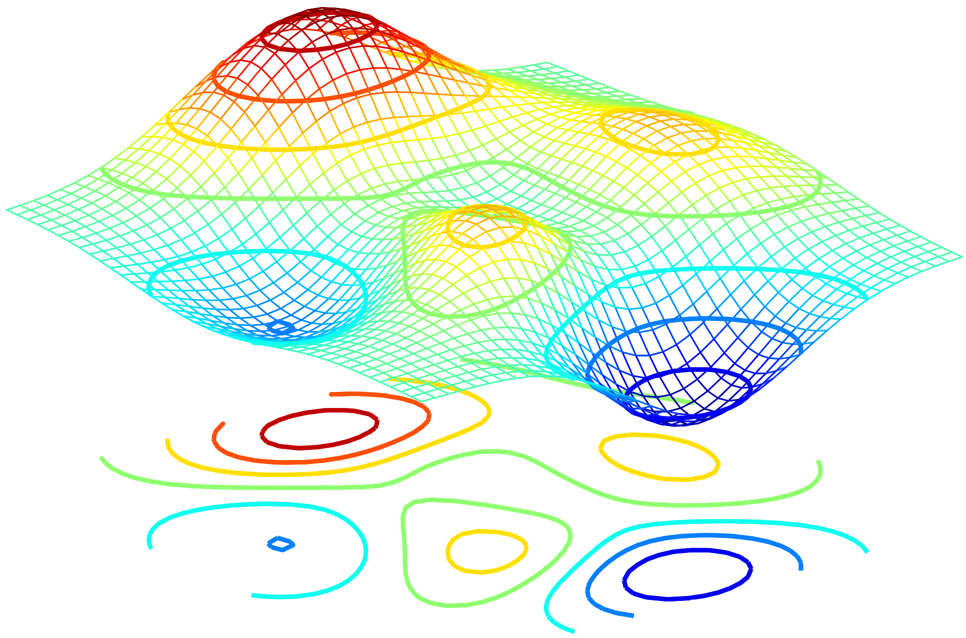
1.2.1 等值线

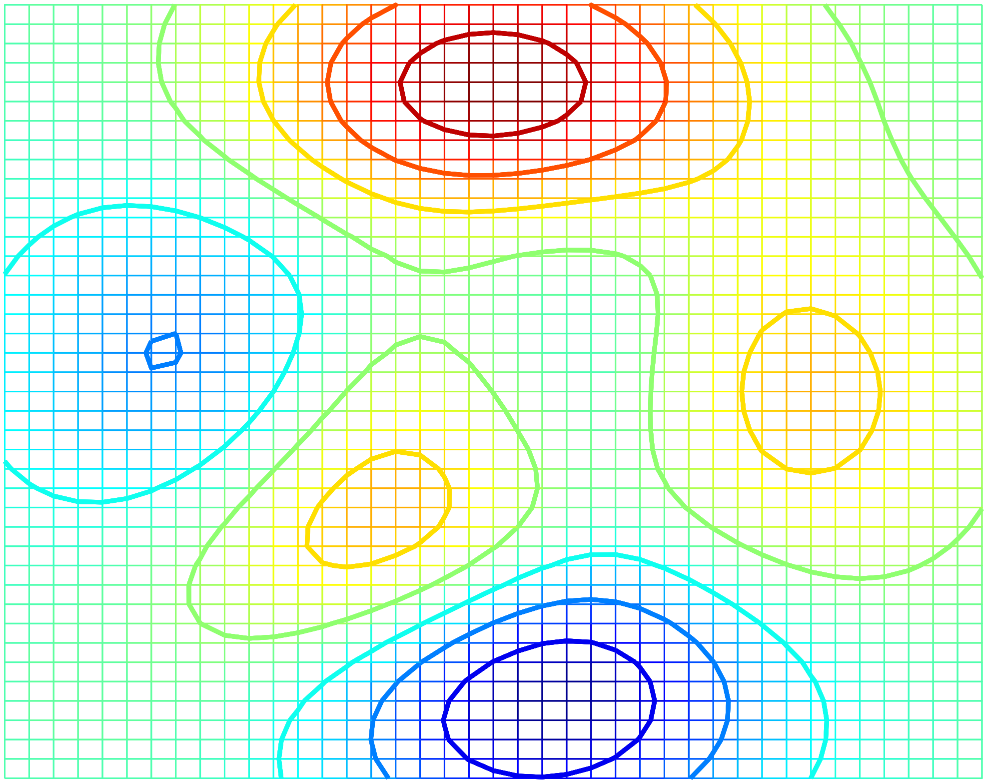
定义 1.2 在高维空间($n \geq 3$)中，使目标函数值取同一常数的点集$\{\boldsymbol{X}|f(\boldsymbol{X})=c, c$为一常数$\}$称为$f(\boldsymbol{X})$的等值线（或等高线、等值面）。

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 可微与梯度

定义 1.3 设$f:\boldsymbol{D}\subset\boldsymbol{R}^n\to\boldsymbol{R}$，且$\boldsymbol{X}^0\in\boldsymbol{D}$，若存在$n$维向量$\boldsymbol{L}$，对任意$n$维向量$\boldsymbol{P}$，都有

$$\lim_{\|\boldsymbol{P}\|\to 0}\frac{f(\boldsymbol{X}^0+\boldsymbol{P})-f(\boldsymbol{X}^0)-\boldsymbol{L}^T\boldsymbol{P}}{\|\boldsymbol{P}\|}=0. \tag{1-2-4}$$

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

$$f(\boldsymbol{X}^0 + \boldsymbol{P}) - f(\boldsymbol{X}^0) = \boldsymbol{L}^T\boldsymbol{P} + \alpha\|\boldsymbol{P}\| = \boldsymbol{L}^T\boldsymbol{P} + o(\|\boldsymbol{P}\|). \quad \text{(1-2-5)}$$

下面定理给出了n维向量$\boldsymbol{L}$的具体表达式。

定理 1.1 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

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

证明 令$\boldsymbol{e}^i=(0,\cdots,0,1_{[i]},0,\cdots,0)^T$，依次取$\boldsymbol{P}=p_i\boldsymbol{e}^i$，$i=1,\cdots,n$，则由式(1-2-5)得证！ ▌

定义 1.4 (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(1-2-5)可记为：

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

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。

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

1.2.3 方向导数

定义 1.5 (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$\boldsymbol{P}$的方向导数为：

$$
\begin{aligned}
\frac{df(\boldsymbol{X})}{d\boldsymbol{P}} &= \lim_{\alpha \to 0^+} \frac{f(\boldsymbol{X} + \alpha \boldsymbol{P}) - f(\boldsymbol{X})}{\|\alpha \boldsymbol{P}\|} \\
&= \lim_{\alpha \to 0^+} \frac{\nabla f(\boldsymbol{X})^T (\alpha \boldsymbol{P}) + o(\|\alpha \boldsymbol{P}\|)}{\|\alpha \boldsymbol{P}\|} \\
&= \nabla f(\boldsymbol{X})^T \boldsymbol{P} \\
&= \|\nabla f(\boldsymbol{X})\| \cos(\nabla f(\boldsymbol{X}), \boldsymbol{P})
\end{aligned}
\tag{1-2-6}
$$

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 1.5 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$\boldsymbol{X}^1$的函数值。

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

1.2.4 *Hesse*矩阵

定义 1.6 (向量值函数) $\boldsymbol{g}(\boldsymbol{X})$是一个向量值函数，若$\boldsymbol{g}:\boldsymbol{D}\subset\boldsymbol{R}^n\to\boldsymbol{R}^m$，即

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

定义 1.7 (可微) 设$\boldsymbol{g}(\boldsymbol{X}):\boldsymbol{D}\subset\boldsymbol{R}^n\to\boldsymbol{R}^m$，$\boldsymbol{X}^0\in\boldsymbol{D}$，

若$\boldsymbol{g}(\boldsymbol{X})$的所有分量$g_1(\boldsymbol{X})$，$g_2(\boldsymbol{X})$，$\cdots$，$g_m(\boldsymbol{X})$在都可微，则称$\boldsymbol{g}(\boldsymbol{X})$在$\boldsymbol{X}^0$可微。

这时称

$$\nabla \boldsymbol{g}(\boldsymbol{X}^0) = \begin{pmatrix} \frac{\partial g_1(\boldsymbol{X}^0)}{\partial x_1} & \frac{\partial g_1(\boldsymbol{X}^0)}{\partial x_2} & \cdots & \frac{\partial g_1(\boldsymbol{X}^0)}{\partial x_n} \\ \frac{\partial g_2(\boldsymbol{X}^0)}{\partial x_1} & \frac{\partial g_2(\boldsymbol{X}^0)}{\partial x_2} & \cdots & \frac{\partial g_2(\boldsymbol{X}^0)}{\partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_m(\boldsymbol{X}^0)}{\partial x_1} & \frac{\partial g_m(\boldsymbol{X}^0)}{\partial x_2} & \cdots & \frac{\partial g_m(\boldsymbol{X}^0)}{\partial x_n} \end{pmatrix} \tag{1-2-7}$$

为向量值函数$\boldsymbol{g}(\boldsymbol{X})$在$\boldsymbol{X}^0$处的导数或$Jacobi$矩阵。

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

设$m=n$，则

$$\nabla^2 f(\boldsymbol{X}) = \nabla \boldsymbol{g}(\boldsymbol{X}) = \begin{pmatrix} \frac{\partial^2 f(\boldsymbol{X})}{\partial^2 x_1} & \frac{\partial^2 f(\boldsymbol{X})}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(\boldsymbol{X})}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(\boldsymbol{X})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\boldsymbol{X})}{\partial^2 x_2} & \cdots & \frac{\partial^2 f(\boldsymbol{X})}{\partial x_n \partial x_2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f(\boldsymbol{X})}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\boldsymbol{X})}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(\boldsymbol{X})}{\partial^2 x_n} \end{pmatrix}. \tag{1-2-8}$$

显然其对称。

定义 1.8 $\nabla^2 f(\boldsymbol{X})$称为$f(\boldsymbol{X})$关于$\boldsymbol{X}$的二阶导数，矩阵(1-2-8)称为$f(\boldsymbol{X})$的$\boldsymbol{Hesse}$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

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

证明 因为

$$\phi(t) = f(x_1^0 + tp_1, \cdots, x_i^0 + tp_i, \cdots, x_n^0 + tp_n) \tag{1-2-9}$$

所以

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

进而

$$\begin{aligned}\phi''(t) &= \sum_{i=1}^{n}\frac{d}{dt}\Big(\frac{\partial f(\boldsymbol{X}^0+t\boldsymbol{P})}{\partial x_i}\Big)p_i\\ &= \sum_{i=1}^{n}\Big(\sum_{j=1}^{n}\frac{\partial^2 f(\boldsymbol{X}^0+t\boldsymbol{P})}{\partial x_j\partial x_i}p_j\Big)p_i \\ &= \boldsymbol{P}^T\nabla^2 f(\boldsymbol{X}^0+t\boldsymbol{P})\boldsymbol{P}\end{aligned} \tag{1-2-11}$$

∎

1.2.5 多元函数的*Taylor*展式

定理 1.2 设$f : \boldsymbol{D} \subset \boldsymbol{R}^n \to \boldsymbol{R}$，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

$$
\begin{aligned}
f(\boldsymbol{X}) = f(\boldsymbol{X}^0) \\
&+ \nabla f(\boldsymbol{X}^0)^T(\boldsymbol{X}-\boldsymbol{X}^0) \\
&+ \frac{1}{2}(\boldsymbol{X}-\boldsymbol{X}^0)^T\nabla^2 f(\boldsymbol{X}^0)(\boldsymbol{X}-\boldsymbol{X}^0) \\
&+ o(\|\boldsymbol{X}-\boldsymbol{X}^0\|^2)
\end{aligned}
\tag{1-2-12}
$$

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

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！ ▌

1.2.6 开集与闭集

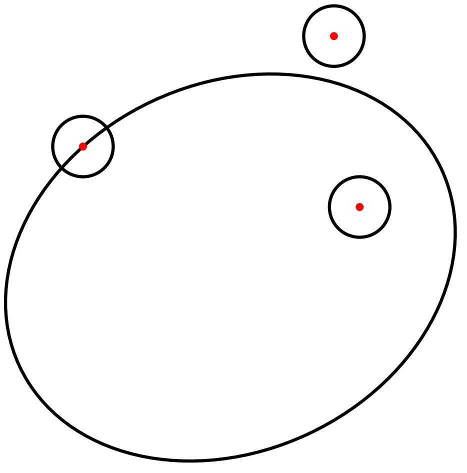
定义 1.9 (内点、边界点与极限点) 设$\boldsymbol{D} \in \boldsymbol{R}^n$，$\boldsymbol{X}^0 \in \boldsymbol{R}^n$。

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

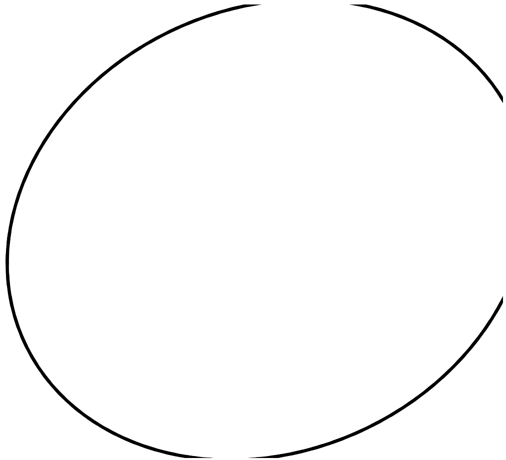
如果在$\boldsymbol{X}^0$的任意领域中，既有$\boldsymbol{X}^1 \in \boldsymbol{D}$，又有$\boldsymbol{X}^2 \notin \boldsymbol{D}$，则称$\boldsymbol{X}^0$为$\boldsymbol{D}$的边界点；

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$\boldsymbol{D}$的极限点（聚点）。

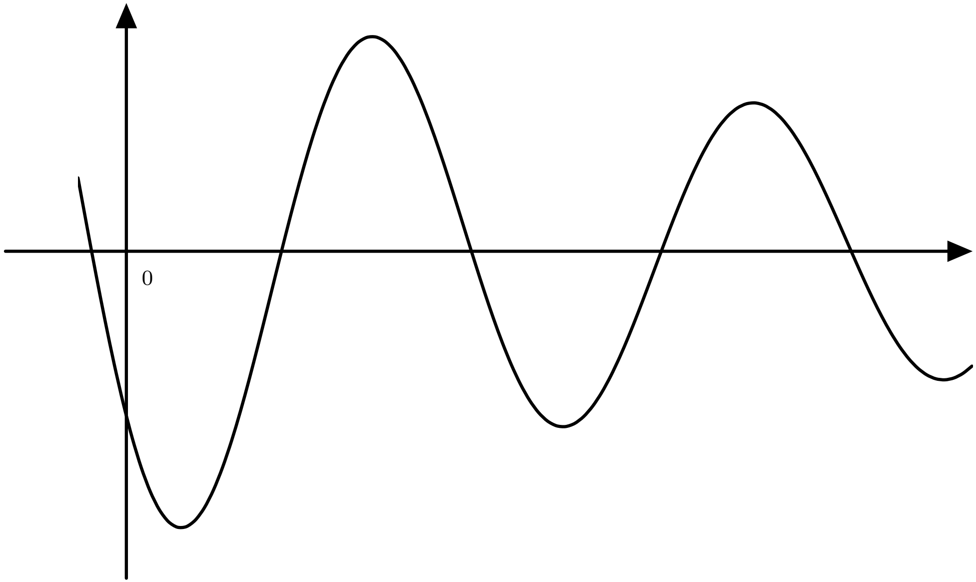


定义 1.10 (开集与闭集) 设$\boldsymbol{D} \subset \boldsymbol{R}^n$，如果$\boldsymbol{D}$的每一个点都是$\boldsymbol{D}$的内点，则称$\boldsymbol{D}$为开集；如果$\boldsymbol{D}$的每一个极限点都属于$\boldsymbol{D}$，则称$\boldsymbol{D}$为闭集。



1.2.7 局部极小点

定义 1.11 (极小点与最优解) 设$f: \boldsymbol{D} \subset \boldsymbol{R}^n \to \boldsymbol{R}$，若存在$\boldsymbol{X}^\star \in \boldsymbol{D}$及实数$\delta > 0$，使得$\forall \boldsymbol{X} \in N^\circ(\boldsymbol{X}^\star, \delta) \bigcap \boldsymbol{D}$都有$f(\boldsymbol{X}^\star) \leq f(\boldsymbol{X})$，则称$\boldsymbol{X}^\star$为$f(\boldsymbol{X})$的局部极小点；若$f(\boldsymbol{X}^\star) < f(\boldsymbol{X})$，则称$\boldsymbol{X}^\star$为$f(\boldsymbol{X})$的严格局部极小点；若对$\forall \boldsymbol{X} \in \boldsymbol{D}$，都有$f(\boldsymbol{X}^\star) \leq f(\boldsymbol{X})$，则称称$\boldsymbol{X}^\star$为$f(\boldsymbol{X})$的全局极小点；若$f(\boldsymbol{X}^\star) < f(\boldsymbol{X})$，则称$\boldsymbol{X}^\star$为$f(\boldsymbol{X})$的全局严格极小点。



1.2.8 最优性条件

定义 1.12 (驻点) 设 $f : \boldsymbol{D} \subset \boldsymbol{R}^n \to \boldsymbol{R}$, $\boldsymbol{X}^{\star}$ 是 $\boldsymbol{D}$ 的内点,

若$\nabla f(\boldsymbol{X}^\star)=0$，则称$\boldsymbol{X}^\star$为$f(\boldsymbol{X})$的驻点。

定理 1.3 (一阶必要条件) 设$f:\boldsymbol{D}\subset\boldsymbol{R}^n\to\boldsymbol{R}$，$f$具有连续的一阶偏导数，若$\boldsymbol{X}^\star$是$f(\boldsymbol{X})$的局部极小点且是$\boldsymbol{D}$的内点，则$\nabla f(\boldsymbol{X}^\star)=0$。

定理 1.4 (二阶必要条件) 设$f:\boldsymbol{D}\subset\boldsymbol{R}^n\to\boldsymbol{R}$，$f$具有连续的二阶偏导数，若$\boldsymbol{X}^\star$是$f(\boldsymbol{X})$的局部极小点且是$\boldsymbol{D}$的内点，则$\nabla^2 f(\boldsymbol{X}^\star)$半正定。

证明 ▮

定理 1.5 (二阶充分条件) 设$f:\boldsymbol{D}\subset\boldsymbol{R}^n\to\boldsymbol{R}$，具有连续

的二阶偏导数，$\boldsymbol{X}^\star$为$f(\boldsymbol{X})$的驻点，且$\nabla^2 f(\boldsymbol{X}^\star)$是正定矩阵，则$\boldsymbol{X}^\star$是$f(\boldsymbol{X})$的严格局部极小点。

证明

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

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

∎

1.2.9 凸集

几何直观上，若集合$\boldsymbol{D}$中的任意两点的连线仍在$\boldsymbol{D}$中，则

称$\boldsymbol{D}$为凸集。

定义 1.13 (凸集) 设$\boldsymbol{D} \subset \boldsymbol{R}^n$，若对所有的$\boldsymbol{X}^1$，$\boldsymbol{X}^2 \in \boldsymbol{D}$，以及$\alpha \in [0,1]$，都有$\alpha \boldsymbol{X}^1 + (1-\alpha)\boldsymbol{X}^2 \in \boldsymbol{D}$，则称$\boldsymbol{D}$为凸集。

例 1.6 平面$\boldsymbol{D} = \{(x_1, x_2, x_3) | x_1 + 2x_2 - x_3 = 4\}$是凸集。

例 1.7 $\boldsymbol{D} = \{\boldsymbol{X} | \boldsymbol{A}\boldsymbol{X} \leq \boldsymbol{b}, \boldsymbol{X} \in \boldsymbol{R}^n, \boldsymbol{A} \in \boldsymbol{R}^{m \times n}\}$是凸集。

例 1.8 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 1.14 (凸组合) 设$\boldsymbol{X}^1$，$\boldsymbol{X}^2$，$\cdots$，$\boldsymbol{X}^m \in \boldsymbol{R}^n$，$\alpha_1$，

α_2，$\cdots$，α_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 1.6 $\boldsymbol{D} \subset \boldsymbol{R}^n$为凸集的充要条件是$\boldsymbol{D}$中任意有限个点的凸组合仍在$\boldsymbol{D}$中。

证明 数学归纳法。 ∎

1.2.10 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 1.15 (凸函数) 设$f : \boldsymbol{D} \subset \boldsymbol{R}^n \to \boldsymbol{R}$，$\boldsymbol{D}$是凸集，若对所有的$\boldsymbol{X}^1$，$\boldsymbol{X}^2 \in \boldsymbol{D}$，以及$\alpha \in (0,1)$，都有

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

则称$f(\boldsymbol{X})$为$\boldsymbol{D}$上的凸函数。

例 1.9 $f(x) = 3x + 4$，$g(x) = |x|$

例 1.10 $f(\boldsymbol{X}) = 2x_1^2 + x_2^2 - 2x_1 x_2$

例 1.11 $f(\boldsymbol{X}) = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$

例 1.12 $f(\boldsymbol{X})=\boldsymbol{X}^T(\boldsymbol{A}\boldsymbol{X})$，$\boldsymbol{A}=\boldsymbol{A}^T$，$\boldsymbol{X}\in\boldsymbol{R}^n$

(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 1.7 设$f:\boldsymbol{D}\subset\boldsymbol{R}^n\to\boldsymbol{R}$，$\boldsymbol{D}$是凸集，则$f(\boldsymbol{X})$是$\boldsymbol{D}$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$\boldsymbol{X}^1$，$\boldsymbol{X}^2$，$\cdots$，$\boldsymbol{X}^m\in\boldsymbol{D}$，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 1.8 设$f:\boldsymbol{D}\subset\boldsymbol{R}^n\to\boldsymbol{R}$，$\boldsymbol{D}$是非空凸集，$f(\boldsymbol{X})$在$\boldsymbol{D}$上可

微，则$f(\boldsymbol{X})$在$\boldsymbol{D}$上是凸函数的充要条件是对任意的$\boldsymbol{X}^1$，$\boldsymbol{X}^2 \in \boldsymbol{D}$，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

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

所以

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

又由凸函数的定义可得

$$\frac{f\big(\boldsymbol{X}^1+\alpha(\boldsymbol{X}^2-\boldsymbol{X}^1)\big)-f(\boldsymbol{X}^1)}{\alpha}\le f(\boldsymbol{X}^2)-f(\boldsymbol{X}^1).$$

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

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

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

分别乘以α，$1-\alpha$后相加即可。 ∎

定理 1.9 设$f:\boldsymbol{D}\subset\boldsymbol{R}^n\to\boldsymbol{R}$，$\boldsymbol{D}$是非空凸集，$f(\boldsymbol{X})$在$\boldsymbol{D}$上具有连续的二阶偏导数，则$f(\boldsymbol{X})$在$\boldsymbol{D}$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：将$Taylor$公式展到二阶，由定理1.8可得。

必要性：对$\forall\boldsymbol{X}\in\boldsymbol{D}$，$\boldsymbol{Y}\in\boldsymbol{D}$，$\exists\lambda$使得$\boldsymbol{X}+\lambda\boldsymbol{Y}\in\boldsymbol{D}$，因此由定理1.8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 1.13 $f(\boldsymbol{X})=3x_1^2-2x_1+2x_2^2-x_2+10$是凸函数。

解 $\nabla^2 f(\boldsymbol{X})=\begin{pmatrix}6&0\\0&4\end{pmatrix}$. ∎

1.2.11 凸规划

定义 1.16 (凸规划) 对于

$$\begin{aligned}&\min\ f(\boldsymbol{X})\\ s.t.&\\ &g_i(\boldsymbol{X})\geq 0\ \ (i=1,2,\cdots,m)\end{aligned}\tag{1-2-13}$$

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

例 1.14 线性规划

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

$$s.t.$$

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

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

为凸规划。

定理 1.10 若设规划(1-2-13)为凸规划，则有

(1)可行集$\boldsymbol{S}$为凸集；

(2)最优解集$\boldsymbol{S}^\star$为凸集；

(3)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$\boldsymbol{X}^1$，$\boldsymbol{X}^2 \in \boldsymbol{S}^\star$，由凸性知，

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

所以，$\alpha \boldsymbol{X}^1 + (1-\alpha)\boldsymbol{X}^2 \in \boldsymbol{S}^\star$，即$\boldsymbol{S}^\star$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$\boldsymbol{Y} \in \boldsymbol{S}$，使得

$$f\big(\alpha \boldsymbol{X} + (1-\alpha)\boldsymbol{Y}\big) < \alpha f(\boldsymbol{X}) + (1-\alpha) f(\boldsymbol{X}) = f(\boldsymbol{X}),$$

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

两式矛盾。 ▌

定理 1.11 设规划(1-2-13)为凸规划，且$f(\boldsymbol{X})$为严格凸函数，则当$\boldsymbol{S}^\star \neq \emptyset$时，其最优解是唯一的。

证明 反证。若规划(1-2-13)的最优解不唯一，则$\exists \boldsymbol{X}^1 \neq \boldsymbol{X}^2 (\in \boldsymbol{S}^\star)$，使得

$$\begin{aligned} f\big(\alpha \boldsymbol{X}^1 + (1-\alpha)\boldsymbol{X}^2\big) < \alpha f(\boldsymbol{X}^1) + (1-\alpha) f(\boldsymbol{X}^2) = f(\boldsymbol{X}^1) \\ = \min_{\boldsymbol{X} \in \boldsymbol{S}} f(\boldsymbol{X}). \end{aligned}$$

因此与$\boldsymbol{X}^1 \in \boldsymbol{S}^\star$矛盾。 ▌

定理 1.12 设凸规划(1-2-13)的目标函数$f(\boldsymbol{X})$可微，则$\boldsymbol{X}^\star$为凸规划(1-2-13)的最优解的充要条件是，$\forall \boldsymbol{X} \in \boldsymbol{S}$，有

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

证明 充分性：

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

必要性：

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

$$f\big(\boldsymbol{X}^\star + \alpha(\boldsymbol{X}^0 - \boldsymbol{X}^\star)\big) - f(\boldsymbol{X}^\star)$$

$$
\begin{aligned}
&= \alpha \nabla f(\boldsymbol{X}^{\star})^T (\boldsymbol{X}^0 - \boldsymbol{X}^{\star}) + o(\alpha \|\boldsymbol{X}^0 - \boldsymbol{X}^{\star}\|) \\
&= \alpha \left((\boldsymbol{X}^0 - \boldsymbol{X}^{\star})^T \nabla f(\boldsymbol{X}^{\star}) + \frac{o(\alpha \|\boldsymbol{X}^0 - \boldsymbol{X}^{\star}\|)}{\alpha} \right) \\
&< 0.
\end{aligned}
$$

这与$\boldsymbol{X}^{\star}$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

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

$s.t.$

$$\begin{array}{lllllllll}
2x_1 & +x_2 & +x_3 & +x_4 & & & & & = 100 \\
 & 2x_2 & +x_3 & & +3x_5 & +2x_6 & +x_7 & & \geq 150 \\
x_1 & & +x_3 & +3x_4 & & +2x_6 & +3x_7 & +5x_8 & \geq 120 \\
\multicolumn{9}{l}{x_i \geq 0, i = 1, 2, \cdots, 8}
\end{array}$$

∎

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

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

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

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

$s.t.$

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

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**：

$$
\begin{aligned}
&\min \ f(\boldsymbol{X})=\sum_{j=1}^{n} c_j x_j \\
&s.t. \\
&\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i=1,2\cdots,m \\
&x_j \geq 0, \quad j=1,2,\cdots,n
\end{aligned}
\tag{2-1-2}
$$

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

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

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

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

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

$$s.t.$$

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

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

$$x_1 + x_2 + x_3 = 5$$

$$x_1, x_2 \geq 0, x_3 \in \boldsymbol{R}$$

§ 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)可行域$\boldsymbol{S} = \varnothing$，无最优解；

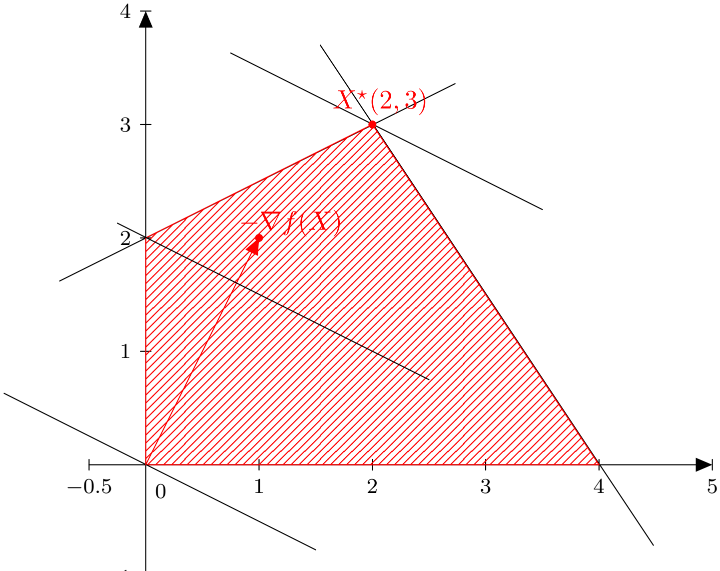
(2)$\boldsymbol{S}$为有界闭集，有唯一的最优解或无穷多个最优解；

(3)$\boldsymbol{S}$为无界集，可能有最优解，也可能无最优解。

§ 2.3 基本概念与解的性质

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。

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



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$n \geq m$。

定义 2.1 (可行集,可行域)

$$\boldsymbol{S} \triangleq \{\boldsymbol{X}|\boldsymbol{A}\boldsymbol{X} = \boldsymbol{b},\ \boldsymbol{X} \geq 0\}.$$

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**

$\boldsymbol{X}\in\boldsymbol{S}$

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

(6)**顶点，极点**

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

或者

设$\boldsymbol{X}^0 \in \boldsymbol{D} \subset \boldsymbol{R}^n$，$\boldsymbol{D}$是凸集，如果$\forall \boldsymbol{X}^1$，$\boldsymbol{X}^2 \in \boldsymbol{D}$，且$\boldsymbol{X}^1 \neq \boldsymbol{X}^2$，则$\nexists\ \alpha^0 \in (0,1)$，$s.t.$，$\boldsymbol{X}^0 = \alpha \boldsymbol{X}^1 + (1+\alpha)\boldsymbol{X}^2$，则称$\boldsymbol{X}^0$为$\boldsymbol{D}$的顶点。

2.3.2 一个例子

例 2.6 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：

	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)

若每生产一件产品A和B分别可获利1万和3万，问如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

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

$$s.t.$$

$$x_1 \leq 4$$

$$x_2 \leq 3$$

$$x_1 + 2x_2 \leq 8$$

引入松弛变量x_3，x_4，x_5，化为标准型

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

$s.t.$

$$x_1 + x_3 = 4$$

$$x_2 + x_4 = 3$$

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

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

■

2.3.3 解的性质

定理 2.1 (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 2.2 线性规划(2-1-3)的可行解$\boldsymbol{X}$是可行集$\boldsymbol{S}$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$\boldsymbol{S}$的顶点，则存

在$\boldsymbol{U}, \boldsymbol{V} \in \boldsymbol{S}$，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$\boldsymbol{S}$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$\boldsymbol{X} \in \boldsymbol{S}$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$\boldsymbol{U}, \boldsymbol{V} \in \boldsymbol{S}$，$\boldsymbol{X} = \frac{1}{2}\boldsymbol{U} + \frac{1}{2}\boldsymbol{V}$，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 2.3 若线性规划(2-1-3)有最优解，则必在其可行集$\boldsymbol{S}$的顶

点处取得。

证明 设$\boldsymbol{X}^0$是最优解，若$\boldsymbol{X}^0=0$，则必为$\boldsymbol{S}$的顶点。下设$\boldsymbol{X}^0=(x_1,x_2,\cdots,x_l,0,\cdots,0)^T\neq 0$，且不是$\boldsymbol{S}$顶点，其中$x_i>0,(i=1,2,\cdots,l)$。因$\boldsymbol{X}^0$不是$\boldsymbol{S}$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$\varepsilon=\min_{\substack{j=1,\cdots,l\\ y_j\neq 0}}\left\{\frac{x_j}{|y_j|}\right\}$，$\boldsymbol{Y}=(y_1,\cdots,y_l,0,\cdots,0)^T$，则由定理2.2可知

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

因为$f(\boldsymbol{X}) = \boldsymbol{C}^T\boldsymbol{X}$，所以

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

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

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。

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：何时迭代终止，找到最优解。

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

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

于是线性规划(2-1-3)可记为

$$
\begin{aligned}
&\min \ \ f(\boldsymbol{X}) = \boldsymbol{C}_{\boldsymbol{I}}^T \boldsymbol{X}_{\boldsymbol{I}} + \boldsymbol{C}_{\boldsymbol{N}}^T \boldsymbol{X}_{\boldsymbol{N}} \\
s.t. \quad & \\
&\boldsymbol{I}\boldsymbol{X}_{\boldsymbol{I}} + \boldsymbol{N}\boldsymbol{X}_{\boldsymbol{N}} = \boldsymbol{b} \\
&\boldsymbol{X}_{\boldsymbol{I}} \geq 0, \quad \boldsymbol{X}_{\boldsymbol{N}} \geq 0
\end{aligned}
\tag{2-4-5}
$$

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

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

设$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_{\boldsymbol{I}} \\ \boldsymbol{X}_{\boldsymbol{N}} \end{pmatrix}$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 2.4 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$\boldsymbol{X}^0 = \begin{pmatrix} \boldsymbol{b} \\ 0 \end{pmatrix}$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 2.5 对于线性规划(2-4-5)，当$\forall \sigma_j \leq 0$时，则$\boldsymbol{X}^0 = \begin{pmatrix} \boldsymbol{b} \\ 0 \end{pmatrix}$就是这个线性规划的最优解。

例 2.7 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 2.6 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 2.8 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

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

$$s.t.$$

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

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

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

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

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

$$\begin{aligned} f(\boldsymbol{X}) &= \boldsymbol{C}_{\boldsymbol{B}}^T(\boldsymbol{B}^{-1}\boldsymbol{b} - \boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{X}_{\boldsymbol{N}}) + \boldsymbol{C}_{\boldsymbol{N}}^T\boldsymbol{X}_{\boldsymbol{N}} \\ &= \boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{b} + (\boldsymbol{C}_{\boldsymbol{N}}^T - \boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{N})\boldsymbol{X}_{\boldsymbol{N}} \end{aligned} \tag{2-4-9}$$

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

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

是线性规划(2-4-8)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(2-4-9)可知，对于任一可行解$\widetilde{\boldsymbol{X}}$，必有

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

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

又因为

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{A}-\boldsymbol{C}^T&=\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}(\boldsymbol{B},\boldsymbol{N})-(\boldsymbol{C}_{\boldsymbol{B}}^T,\boldsymbol{C}_{\boldsymbol{N}}^T)\\&=(\boldsymbol{C}_{\boldsymbol{B}}^T,\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{N})-(\boldsymbol{C}_{\boldsymbol{B}}^T,\boldsymbol{C}_{\boldsymbol{N}}^T)\\&=(0,\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{N}-\boldsymbol{C}_{\boldsymbol{N}}^T)\end{aligned}\tag{2-4-10}$$

所以

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

若令$\sigma_j=\boldsymbol{C}_{\boldsymbol{B}}^T\boldsymbol{B}^{-1}\boldsymbol{P}_j-c_j,j=1,\cdots,n$，则有下面类似定理成立：

定理 2.7 对于线性规划(2-4-8)，当$\forall \sigma_j \leq 0$，则$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{B}^{-1}\boldsymbol{b} \\ 0 \end{pmatrix}$为最优解。

(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

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

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

若$\sigma_l > 0$，$a_{kl} > 0$，则利用矩阵的初等行变换（不换行）将

第l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

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

其中

$$
\begin{aligned}
a'_{kj} &= \frac{a_{kj}}{a_{kl}}, \quad j = 1, \cdots, n \\
a'_{ij} &= a_{ij} - \frac{a_{kj}}{a_{kl}} a_{il}, \quad i = 1, \cdots, m; \quad i \neq k; \quad j = 1, \cdots, n \\
b'_k &= \frac{b_k}{a_{kl}} \\
b'_i &= b_i - \frac{b_k}{a_{kl}} a_{il}, \quad i = 1, \cdots, m; \quad i \neq k
\end{aligned}
\tag{2-4-11}
$$

于是得到新基

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

及其基本解

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

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？

因为$b_i' = b_i - \frac{b_k}{a_{kl}} a_{il}$，所以若$a_{il} \leq 0$时，则$b_i' \geq 0$；若$a_{il} > 0$时，则取符合下面条件

$$\frac{b_k}{a_{kl}} = \min_{i=1,\cdots,m} \left\{ \frac{b_i}{a_{il}} \middle| a_{il} > 0 \right\} \tag{2-4-12}$$

成立的a_{kl}作为主元，则$b_i' \geq 0$。因此$\boldsymbol{X}'$为基可行解。

例 2.9 考虑线性规划

$-\frac{19}{4}$

$$\min \ -x_3$$

$$s.t.$$

$$x_1 + 3x_4 - x_5 + 3x_6 = 2$$

$$x_2 + 2x_4 - 2x_5 + x_6 = 1$$

$$x_3 - 2x_4 - x_5 + 2x_6 = 3$$

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

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

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

若$\boldsymbol{P}_4$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

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

新的基可行解为

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

注：$f(\boldsymbol{X}^1) < f(\boldsymbol{X}^0)$。

(3)进基列的选择：如何选择进基列（换基运算）可以使目标函数有较大的下降。

定理 2.8 若线性规划(2-4-8)，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$\boldsymbol{X}^0 = \begin{pmatrix} \boldsymbol{B}^{-1}b \\ 0 \end{pmatrix}$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$b_i - \theta a_{il} (\triangleq x_i') \geq$

$b_i - \frac{b_i}{a_{il}} a_{il} = 0$。

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

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

因为$\theta > 0$以及$\sigma_l > 0$，所以

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

∎

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 2.10

1

$$\min \ -3x_1 - 5x_2 \quad -36$$

$$s.t.$$

$$x_1 + x_3 = 4$$

$$x_2 + x_4 = 6$$

$$3x_1 + 2x_2 + x_5 = 18$$

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

2.4.2 单纯形方法

(1)初始单纯形表

对于线性规划(2-4-8)，称矩阵

$$\begin{pmatrix} \boldsymbol{B}^{-1}\boldsymbol{A} & \boldsymbol{B}^{-1}\boldsymbol{b} \\ \boldsymbol{C}_{\boldsymbol{B}}^{T}\boldsymbol{B}^{-1}\boldsymbol{A} - \boldsymbol{C}^{T} & \boldsymbol{C}_{\boldsymbol{B}}^{T}\boldsymbol{B}^{-1}\boldsymbol{b} \end{pmatrix} \tag{2-4-13}$$

为初始单纯形表。

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

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{N} & \boldsymbol{b} \\ \boldsymbol{0} & \boldsymbol{\sigma}_{\boldsymbol{N}}^{T} & \boldsymbol{C}_{\boldsymbol{I}}^{T}\boldsymbol{b} \end{pmatrix} \tag{2-4-14}$$

初始单纯形表为

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

初始单纯形表中记录以下信息：

(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2.8确定进基列$\boldsymbol{P}_l$，按式(2-4-12)确定主元a_{kl}，然后按照式(2-4-11)作换基运算。

于是上述单纯形表变成下面的新单纯形表。

容易证明下面的定理成立:

定理 2.9 (I)$\sigma_j' = \sigma_j - \frac{a_{kj}}{a_{kl}}\sigma_l$就是新表$\boldsymbol{P}_j$中的判别数；

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'

(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)

$$
\begin{aligned}
\sigma_j' &= \sigma_j - \frac{a_{kj}}{a_{kl}}\sigma_l \\
&= \Big(\sum_{i=1}^{m} c_i a_{ij} - c_j\Big) - \frac{a_{kj}}{a_{kl}}\Big(\sum_{i=1}^{m} c_i a_{il} - c_l\Big) \\
&= \sum_{i=1}^{m} c_i\Big(a_{ij} - \frac{a_{kj}}{a_{kl}} a_{il}\Big) + c_l\frac{a_{kj}}{a_{kl}} - c_j \\
&= \sum_{\substack{i=1 \\ i\neq k}}^{m} c_i a_{ij}' + c_l a_{kj}' - c_j \\
&= \boldsymbol{C}_{\boldsymbol{I}}^T \boldsymbol{P}_j - c_j
\end{aligned} \tag{2-4-15}
$$

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)

$$
\begin{aligned}
f' &= f_0 - \frac{b_k}{a_{kl}}\sigma_l \\
&= \sum_{i=1}^{m} c_i b_i - \frac{b_k}{a_{kl}}\Big(\sum_{i=1}^{m} c_i a_{il} - c_l\Big) \\
&= \sum_{\substack{i=1 \\ i\neq k}}^{m} c_i\Big(b_i - \frac{b_k}{a_{kl}}a_{il}\Big) + c_l\frac{b_k}{a_{kl}} \\
&= \sum_{\substack{i=1 \\ i\neq k}}^{m} c_i b_i' + c_l b_k' \\
&= f(\boldsymbol{X}^1)
\end{aligned}
\tag{2-4-16}
$$

∎

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：

$$\begin{aligned} \sigma_j' &= \sigma_j - \frac{a_{kj}}{a_{kl}}\sigma_l \\ f' &= f_0 - \frac{b_k}{a_{kl}}\sigma_l \end{aligned} \tag{2-4-17}$$

(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 单纯形算法

步骤 1 构造初始单纯形表

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

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$\sigma_l = \max_{1 \leq j \leq n} \{\sigma_j\}$。

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

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

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

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 2.11

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

$s.t.$

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

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

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

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

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，按下表2.6作换基运算：

到此，所有判别数小于等于0，故当前基可行解$\boldsymbol{X}^\star =$

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

$(0,4,5,0,0,11)^T$为最优解，最优值为$f(\boldsymbol{X}^\star)=-11$。 ∎

例 2.12

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

$s.t.$

$$x_1 \le 4$$

$$x_2 \le 3$$

$$x_1 + 2x_2 \le 8$$

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

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
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
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

§ 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

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

$s.t.$

$$\begin{aligned} & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ & a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ & \qquad \vdots \\ & a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \\ & x_j \geq 0, \quad j = 1, 2, \cdots, n \end{aligned} \tag{2-5-18}$$

2.5.1 两阶段方法

引入人工变量$y_1, y_2, \cdots, y_m$，构造辅助线性规划：

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

$s.t.$

$$\begin{aligned}
&y_1 + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1\\
&y_2 + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2\\
&\qquad\vdots\\
&y_m + a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m\\
&y_i \geq 0,\quad x_j \geq 0,\quad i = 1 \sim m,\quad j = 1 \sim n
\end{aligned} \tag{2-5-19}$$

称人工变量所对应的列$\boldsymbol{d}_i = (0,\cdots,0,1_{[i]},0,\cdots,0)^T$为人工向量，$i=1,2,\cdots,m$。

显然，$\boldsymbol{I} = (\boldsymbol{d}_1,\boldsymbol{d}_2,\cdots,\boldsymbol{d}_m)$为单位矩阵，$\boldsymbol{Y}^0 = (b_1,b_2,\cdots,b_m,0,\cdots,0)^T$为初始基可行解。

从$\boldsymbol{Y}^0$出发，对线性规划(2-5-19)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(2-5-19)的最优解，设最后一张单纯

形表为

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

其中，$g^\star = \sum_{i=1}^m y_i^\star$。

分两种情况讨论。

（1）$g^\star > 0$

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(2-5-19)的可行解，且$g(\boldsymbol{Y}^0) = 0 < g^\star$。

（2）$g^\star = 0$

则对$\forall i = 1 \sim m, y_i^\star = 0$，

（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$y_i^\star = 0, \forall i = 1 \sim m$可知，此时上式即为$0 = 0$，即在规划(2-5-19)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 2.13 (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 构造辅助线性规划

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

$s.t.$

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

$$y_2 + x_2 + x_4 = 2$$

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

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

按照单纯形法，作两次换基运算后，得到

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

表 2.8 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 2.9 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

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

例 2.14

$$\min \ x_1 - x_2$$

$$s.t.$$

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

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

$$x_1 - x_3 = 0$$

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

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 2.10 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，$X^\star = (0,1,0)^T$，$f(X^\star) = -1$。 ∎

2.5.2 大M法

初始基未知的情况下，也可采用大M方法。

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

$$\begin{aligned} &\min \ \ \boldsymbol{C}^T\boldsymbol{X} + M\boldsymbol{e}^T\boldsymbol{X}_a \\ &s.t. \\ &\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}_a = \boldsymbol{b} \\ &\boldsymbol{X} \geq 0, \boldsymbol{X}_a \geq 0. \end{aligned} \tag{2-5-20}$$

用单纯形法求解线性规划(2-5-20)，其结果为:

(1)线性规划(2-5-20)有最优解$\begin{pmatrix} \boldsymbol{X}^\star \\ \boldsymbol{X}_a^\star \end{pmatrix}$，且$\boldsymbol{X}_a^\star = 0$，则$\boldsymbol{X}^\star$为原线性规划的最优解;

(2)线性规划(2-5-20)有最优解$\begin{pmatrix} \boldsymbol{X}^\star \\ \boldsymbol{X}_a^\star \end{pmatrix}$，且$\boldsymbol{X}_a^\star \neq 0$，即$\boldsymbol{e}^T\boldsymbol{X}_a^\star > 0$，则原线性规划无可行解。

证明 若原线性规划有可行解$\boldsymbol{X}$，则$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{0} \end{pmatrix}$是线性规划(2-5-20)的可行解，则$f\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{0} \end{pmatrix} = \boldsymbol{C}^T\boldsymbol{X} + M\boldsymbol{e}^T\boldsymbol{0} = \boldsymbol{C}^T\boldsymbol{X}$，然而，$f\begin{pmatrix} \boldsymbol{X}^\star \\ \boldsymbol{X}_a^\star \end{pmatrix} = \boldsymbol{C}^T\boldsymbol{X}^\star + M\boldsymbol{e}^T\boldsymbol{X}_a^\star > \boldsymbol{C}^T\boldsymbol{X} = f\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{0} \end{pmatrix}$. ∎

(3)线性规划(2-5-20)无最优解，则原规划无解。(证明略)

例 2.15

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

$s.t.$

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

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

$$x_1 - 2x_3 = 1$$

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

解 化为标准型为

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

$$s.t.$$

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

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

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

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

用单纯形法求解如下：见表2.11。

所以$\boldsymbol{X}^\star = (9, 1, 4)^T$，$f(\boldsymbol{X}^\star) = -2$。 ∎

表 2.11 单纯形表

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

§ 2.6 单纯形法的改进

2.6.1 避免循环

在定理2.8中，曾假设基可行解$\begin{pmatrix} \boldsymbol{b} \\ 0 \end{pmatrix}$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:

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

$s.t.$

$$\frac{1}{4}x_1 - 60x_2 - \frac{1}{25}x_3 + 9x_4 \le 0$$

$$\frac{1}{2}x_1 - 90x_2 - \frac{1}{50}x_3 + 3x_4 \le 0$$

$$x_3 \le 1$$

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

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(2-4-8)，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$\boldsymbol{B}^{-1}\boldsymbol{N} = (\boldsymbol{B}^{-1}\boldsymbol{P}_{m+1}, \cdots, \boldsymbol{B}^{-1}\boldsymbol{P}_n)$。即在求解当前基可行解$\begin{pmatrix} \boldsymbol{B}^{-1}\boldsymbol{b} \\ 0 \end{pmatrix}$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)$\boldsymbol{X}_B = \boldsymbol{B}^{-1}\boldsymbol{b}$；

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

定理 2.10 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，

$$
\boldsymbol{E}_{kl} = \begin{array}{c}
\begin{pmatrix}
1 & & & -\dfrac{a_{1l}}{a_{kl}} & & & \\
 & \ddots & & \vdots & & & \\
 & & 1 & -\dfrac{a_{k-1,l}}{a_{kl}} & & & \\
 & & & \dfrac{1}{a_{kl}} & & & \\
 & & & -\dfrac{a_{k+1,l}}{a_{kl}} & 1 & & \\
 & & & \vdots & & \ddots & \\
 & & & -\dfrac{a_{ml}}{a_{kl}} & & & 1
\end{pmatrix} \\
k\text{列}
\end{array}
\tag{2-6-21}
$$

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 修正单纯形算法

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 2.16

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

$$s.t.$$

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

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

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

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

第一步：

取$\boldsymbol{B}^0 = (\boldsymbol{P}_4, \boldsymbol{P}_5) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$,

$(\boldsymbol{B}^0)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$,

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

$\sigma_1 = \pi\boldsymbol{P}_1 - c_1 = 4$，$\sigma_2 = \pi\boldsymbol{P}_2 - c_2 = 3$，$\sigma_3 = \pi\boldsymbol{P}_3 - c_3 = 6$,

$(\boldsymbol{B}^0)^{-1}\boldsymbol{P}_1 = (3,2)^T$,

$\min\left\{\frac{30}{3}, \frac{40}{2}\right\} = \frac{30}{3}$,

所以主元为a_{11}，进基列为$\boldsymbol{P}_1$，出基列为$\boldsymbol{P}_4$.

第二步：

$\boldsymbol{E}_{11} = \begin{pmatrix} \frac{1}{3} & 0 \\ -\frac{2}{3} & 1 \end{pmatrix}$,

$(\boldsymbol{B}^1)^{-1} = (\boldsymbol{P}_1, \boldsymbol{P}_5)^{-1} = \boldsymbol{E}_{11}(B^0)^{-1} = \boldsymbol{E}_{11}$,

$(\boldsymbol{B}^1)^{-1}\boldsymbol{b} = \boldsymbol{E}_{11}\boldsymbol{b} = (10,20)^T$，$\boldsymbol{C}_{\boldsymbol{B}^1}^T = (-4,0)$，$\pi = \boldsymbol{C}_{B^1}^T(\boldsymbol{B}^1)^{-1} = (-4/3, 0)$,

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

$(\boldsymbol{B}^1)^{-1}\boldsymbol{P}_2 = (\frac{2}{3}, \frac{2}{3})^T$,

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

所以主元为a_{12}，进基列为$\boldsymbol{P}_2$，出基列为$\boldsymbol{P}_1$.

第三步:

$\boldsymbol{E}_{12} = \begin{pmatrix} \frac{3}{2} & 0 \\ -1 & 1 \end{pmatrix}$,

$(\boldsymbol{B}^2)^{-1} = (\boldsymbol{P}_2, \boldsymbol{P}_5)^{-1} = \boldsymbol{E}_{12}(\boldsymbol{B}^1)^{-1} = \boldsymbol{E}_{12}\boldsymbol{E}_{11} = \begin{pmatrix} \frac{1}{2} & 0 \\ -1 & 1 \end{pmatrix}$,

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

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

$(\boldsymbol{B}^2)^{-1}\boldsymbol{P}_3 = (\frac{3}{2}, 0)^T$,

所以主元为a_{13}，进基列为$\boldsymbol{P}_3$，出基列为$\boldsymbol{P}_2$.

第四步:

$\boldsymbol{E}_{13}=\begin{pmatrix}\frac{2}{3} & 0\\ 0 & 1\end{pmatrix}$,

$(\boldsymbol{B}^3)^{-1}=(\boldsymbol{P}_3,\boldsymbol{P}_5)^{-1}=\boldsymbol{E}_{13}(\boldsymbol{B}^2)^{-1}=\boldsymbol{E}_{13}\boldsymbol{E}_{12}\boldsymbol{E}_{11}=\begin{pmatrix}\frac{1}{3} & 0\\ -1 & 1\end{pmatrix}$,

$(\boldsymbol{B}^3)^{-1}\boldsymbol{b}=(10,10)^T$，$\boldsymbol{C}_{\boldsymbol{B}^3}^T=(-6,0)$，$\pi=\boldsymbol{C}_{\boldsymbol{B}^3}^T(\boldsymbol{B}^3)^{-1}=(-2,0)$,

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

所以最优解为$\boldsymbol{X}^\star=(0,0,10)^T$，最优值$f(\boldsymbol{X}^\star)=-60$. ∎

解 (对比单纯形方法) ∎

表 2.12 单纯形表

P_1	P_2	P_3	P_4	P_5	b
3	2	3	1	0	30
$\boxed{2}$	2	3	0	1	40
4	3	6	0	0	0
1	$\boxed{2/3}$	1	1/3	0	10
0	2/3	1	−2/3	1	20
0	1/3	2	−4/3	0	−40
3/2	1	$\boxed{3/2}$	1/2	0	15
−1	0	0	−1	1	10
−1/2	0	3/2	−3/2	0	−45
1	2/3	1	1/3	0	10
−1	0	0	−1	1	10
−2	−1	0	−2	0	−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w_1，w_2分别表示在一周内产品I、II的产量，则有数学模

型：

$$\max \ g(\boldsymbol{W}) = 2w_1 + 3w_2$$

$$s.t.$$

$$\begin{aligned} & 2w_1 + 2w_2 \leq 12 \\ & w_1 + 2w_2 \leq 8 \\ & 4w_1 \leq 16 \\ & 4w_2 \leq 12 \\ & w_1, w_2 \geq 0 \end{aligned} \tag{3-1-1}$$

▮

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x_1，x_2，x_3，x_4千

元，那么这个问题的数学模型为

$$\min\ f(\boldsymbol{X}) = 12x_1 + 8x_2 + 16x_3 + 12x_4$$

$s.t.$

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

$$2x_1 + 2x_2 + 4x_4 \geq 3$$

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

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$\boldsymbol{W}^\star = (4,2)^T$、$\boldsymbol{X}^\star = (\frac{1}{2}, 1, 0, 0)^T$。这时$f(\boldsymbol{X}^\star) = g(\boldsymbol{W}^\star) = 14$。

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

3.1.2 对称形式

定义 3.1 对于线性规划

$$(P)\qquad \begin{array}{l} \min\ \boldsymbol{C}^T\boldsymbol{X} \\ s.t. \\ \boldsymbol{A}\boldsymbol{X} \geq \boldsymbol{b} \\ \boldsymbol{X} \geq 0 \end{array} \qquad (3\text{-}1\text{-}3)$$

线性规划

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

称为(D)的对偶线性规划，而(P)称为原规划。

定理 3.1 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 3.2 写出线性规划

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

$s.t.$

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

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

$$x_3 + x_4 \geq 2$$

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

的对偶线性规划。

解

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

$s.t.$

$$w_1 + 3w_2 \leq 8$$

$$2w_1 + w_2 \leq 6$$

$$w_2 + w_3 \leq 3$$

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

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

3.1.3 非对称形式

对于线性规划

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

因为$\boldsymbol{AX} = \boldsymbol{b}$等价于

$$
\boldsymbol{AX} \geq \boldsymbol{b},
$$

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

故(P')可变为

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

$$s.t.$$

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

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

其对偶线性规划为

$$\max \begin{pmatrix} \boldsymbol{b} \\ -\boldsymbol{b} \end{pmatrix}^T \begin{pmatrix} \boldsymbol{W}^1 \\ \boldsymbol{W}^2 \end{pmatrix}$$

$$s.t.$$

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

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

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

$$
(D') \qquad \begin{aligned} & \max\ \boldsymbol{b}^T\boldsymbol{W} \\ & s.t. \\ & \boldsymbol{A}^T\boldsymbol{W}\le \boldsymbol{C} \end{aligned} \tag{3-1-8}
$$

例 3.3 写出线性规划

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

$s.t.$

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

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

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

的对偶线性规划。

解

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

$$s.t.$$

$$w_1 + 3w_2 \leq 8$$

$$2w_1 + w_2 \leq 6$$

$$w_2 \leq 3$$

$$w_1 + w_2 \leq 6$$

$$w_j \in \boldsymbol{R},\ j = 1, 2.$$

■

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。

例 3.4 写出线性规划

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

表 3.2 原(对偶)线性规划变换规则表

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 $=$

$s.t.$

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

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

$$x_3 + x_4 = 2$$

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

的对偶线性规划。

解

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

$s.t.$

$$w_1 + 3w_2 \leq 8$$

$$2w_1 + w_2 \leq 6$$

$$w_2 + w_3 \geq 3$$

$$w_1 + w_2 + w_3 = 6$$

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

∎

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

定理 3.2 (弱对偶性) $\forall \boldsymbol{X} \in \boldsymbol{S}_P$，$\boldsymbol{W} \in \boldsymbol{S}_D$，必有

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

证明

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

∎

推论 3.3 (无界性) 如果$(P)(D)$具有无界解，则其对偶$(D)(P)$无可行解。

定理 3.4 (P)与对偶(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$\forall \boldsymbol{X} \in \boldsymbol{S}_P, \boldsymbol{W} \in \boldsymbol{S}_D$,

有$\boldsymbol{C}^T\boldsymbol{X} \geq \boldsymbol{b}^T\boldsymbol{W}$。因此$(P)$的目标函数在$\boldsymbol{S}_P$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 3.5 (最优性) 如果$\boldsymbol{X} \in \boldsymbol{S}_P$，$\boldsymbol{W} \in \boldsymbol{S}_D$，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 3.6 (强对偶性) 若(P)与其对偶(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(P)化为标准形

式

$$
\begin{aligned}
&\min\ \begin{pmatrix} \boldsymbol{C}^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \\
&s.t. \\
&\begin{pmatrix} \boldsymbol{A} & -\boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} = \boldsymbol{b} \\
&\boldsymbol{X}, \boldsymbol{Y} \geq 0
\end{aligned}
\tag{3-2-9}
$$

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

$$\boldsymbol{C}_B^T \boldsymbol{B}^{-1} \begin{pmatrix} \boldsymbol{A} & -\boldsymbol{I} \end{pmatrix} \leq \begin{pmatrix} \boldsymbol{C}^T & \boldsymbol{0}^T \end{pmatrix}.$$

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

$$(\boldsymbol{W}^0)^T\begin{pmatrix}\boldsymbol{A} & -\boldsymbol{I}\end{pmatrix}\leq\begin{pmatrix}\boldsymbol{C}^T & \boldsymbol{0}^T\end{pmatrix}.$$

即，$\boldsymbol{W}^0$是(D)的可行解。

又

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

故由推论知命题成立。 ∎

注: 从上面定理的证明过程看，若B是线性规划(P)的最优基，那么$(\boldsymbol{C}_B^T\boldsymbol{B}^{-1})^T$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 3.5 对于线性规划

$$\min\ 2x_1+3x_2+x_3$$

$$s.t.$$

$$x_1+4x_2+2x_3\geq 8$$

$$3x_1 + 2x_2 \geq 6$$

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

引入剩余变量x_4，x_5后，用单纯形法求得最终单纯形表如下(表3.3)：

表 3.3 单纯形表

P_1	P_2	P_3	P_4	P_5	b
0	1	$\frac{3}{5}$	$-\frac{3}{10}$	$\frac{1}{10}$	$\frac{9}{5}$
1	0	$-\frac{2}{5}$	$\frac{1}{5}$	$-\frac{2}{5}$	$\frac{4}{5}$
0	0	0	$-\frac{1}{2}$	$-\frac{1}{2}$	7

则对偶线性规划

$$\max\ 8w_1 + 6w_2$$

$$s.t.$$

$$w_1 + 3w_2 \leq 2$$

$$4w_1 + 2w_2 \leq 3$$

$$2w_1 \leq 1$$

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

的最优解$W^\star = (1/2, 1/2)^T$。 ∎

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 3.7 (对称形式的松弛定理,或互补性) 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$\boldsymbol{X}$，$\boldsymbol{W}$是最优解，所以

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

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

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

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3.7等价于

定理 3.8 (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 3.2 (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 3.3 (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 3.9 设(P)与(D)都有可行解，则松约束的对偶约束是紧约束。

定理 3.10 (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 3.6 已知线性规划

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

$$s.t.$$

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

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

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

的最优解是$\boldsymbol{X}^\star = (6, 2, 0)^T$，求其对偶规划的最优解$\boldsymbol{Y}^\star$。

解 对偶规划为

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

$$s.t.$$

$$y_1 + 2y_2 \geq 3$$

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

$$y_1 + y_2 \geq 1$$

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

设对偶规划的最优解为$\boldsymbol{Y}^\star = (y_1^\star, y_2^\star)^T$，因为$x_1^\star = 6 > 0$，$x_2^\star = 2 > 0$，由松弛定理得

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

所以对偶规划的最优解为$\boldsymbol{Y}^\star = (1, 1)^T$，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。

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

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

定理 3.11 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

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

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 3.4 (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$\boldsymbol{X}^\star$。这就是对偶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；

$$\begin{pmatrix} \boldsymbol{A} \mid \boldsymbol{b} \\ \sigma \quad f \end{pmatrix} = \begin{pmatrix} 1 & & & & & a_{1,m+1} & \cdots & a_{1l} & \cdots & a_{1n} & b_1 \\ & \ddots & & & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & 1 & & & a_{k,m+1} & \cdots & a_{kl} & \cdots & a_{kn} & b_k \\ & & & \ddots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & 1 & a_{m,m+1} & \cdots & a_{ml} & \cdots & a_{mn} & b_m \\ 0 & \cdots & 0 & \cdots & 0 & \sigma_{m+1} & \cdots & \sigma_l & \cdots & \sigma_n & f_0 \end{pmatrix}$$

(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$\boldsymbol{W}$ 的目标函数值$\boldsymbol{b}^T\boldsymbol{W} = \boldsymbol{C}_B^T\boldsymbol{B}^{-1}\boldsymbol{b}$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

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

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

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

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

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

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

算法 3.1 对偶单纯形算法

步骤 1 构造对偶初始单纯形表

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

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。

例 3.7 用对偶单纯形法求解下面的线性规划

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

$s.t.$

$$x_1 + 2x_2 \geq 4$$

$$x_1 \leq 5$$

$$3x_1 + x_2 \geq 6$$

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

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：

这时$\boldsymbol{b} > 0$，所以当前正则解$\boldsymbol{X}^\star = (4, 0, 0, 1, 6)^T$是最优解，因

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

此原规划的最优解为$\boldsymbol{X}^\star = (4,0)^T$。 ∎

§ 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

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

其中$\boldsymbol{C} \geq 0$。

引入剩余变量$\boldsymbol{Y}$，原规划变为

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

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

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

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

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

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

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

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

$$s.t.$$

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

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

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

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

这时

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

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

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

于是

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

3.4.2 影子价格

继续考虑问题(例3.1):

$g(\boldsymbol{W}^{\star}) = 14$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$f(\boldsymbol{X}^{\star}) = 14$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。

(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

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

所以$y_i^\star$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

$$\begin{aligned}\sigma_j &= \boldsymbol{C}_{\boldsymbol{B}^\star}^T(\boldsymbol{B}^\star)^{-1}\boldsymbol{P}_j - c_j \\ &= (\boldsymbol{W}^\star)^T\boldsymbol{P}_j - c_j \\ &= \sum_{i=1}^{m} a_{ij}y_i^\star - c_j, \quad j = 1 \sim n.\end{aligned} \tag{3-4-27}$$

这里，c_j表示第j种产品的利润(产值)，$\sum_{i=1}^m a_{ij}y_i^\star$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)。所以产品利润小于隐含成本时，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：

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

(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$f(\boldsymbol{X}^{k+1}) < f(\boldsymbol{X}^k)$，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$\boldsymbol{X}^{\star}$，即

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

则称**算法**收敛到$\boldsymbol{X}^{\star}$。

(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$\boldsymbol{t}_{k}$**。

下降迭代算法的一般格式：

算法 4.1 下降迭代算法

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 4.1 (收敛速度) 设序列$\{\boldsymbol{X}^k\}$收敛于解$\boldsymbol{X}^\star$，若

$$\lim_{k\to\infty}\frac{\|\boldsymbol{X}^{k+1}-\boldsymbol{X}^\star\|}{\|\boldsymbol{X}^k-\boldsymbol{X}^\star\|} = \beta. \tag{4-1-5}$$

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

$$\lim_{k\to\infty}\frac{\|\boldsymbol{X}^{k+1}-\boldsymbol{X}^\star\|}{\|\boldsymbol{X}^k-\boldsymbol{X}^\star\|^p}=\beta<+\infty. \tag{4-1-6}$$

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 4.1 (1)$\boldsymbol{X}^k=a^{2^k}$，这里$0<\|a\|<1$，为二阶收敛；

(2) $\boldsymbol{X}^k = k^{-2}$为1阶收敛，但不是线性收敛。

(3) $\boldsymbol{X}^k = k^{-k}$为1阶收敛，且为超线性收敛。

定义 4.2 (二次收敛性,或二次终止性) 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

$$\|\boldsymbol{X}^{k+1}-\boldsymbol{X}^{k}\|<\varepsilon,\quad or\quad \frac{\|\boldsymbol{X}^{k+1}-\boldsymbol{X}^{k}\|}{\|\boldsymbol{X}^{k}\|}<\varepsilon. \tag{4-1-7}$$

$$\|\boldsymbol{f}^{k+1}-\boldsymbol{f}^{k}\|<\varepsilon,\quad or\quad \frac{\|\boldsymbol{f}^{k+1}-\boldsymbol{f}^{k}\|}{\|\boldsymbol{f}^{k}\|}<\varepsilon. \tag{4-1-8}$$

$$\|\nabla f(\boldsymbol{X}^{k})\|<\varepsilon. \tag{4-1-9}$$

§ 4.2 精确一维搜索

4.2.1 黄金分割法

黄金分割法(**0.618**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 4.3 设$\varphi : [a, b] \subset \boldsymbol{R} \to \boldsymbol{R}$，$t^\star$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

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

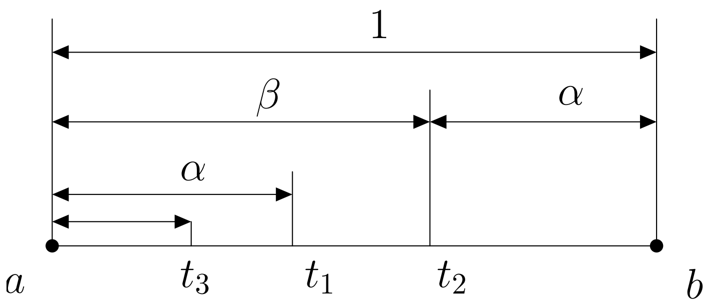
那么称$\varphi(t)$是区间$[a,b]$上的单峰函数。若$t_1 < t^\star < t_2$，称$[t_1, t_2]$为搜索区间。

性质 4.3 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

$$\begin{aligned}\beta&=\frac{\sqrt{5}-1}{2}\approx 0.618\\ \alpha&=\frac{3-\sqrt{5}}{2}\approx 0.382\end{aligned}\tag{4-2-10}$$

于是有

$$
\begin{aligned}
t_2 &= a + \overline{at_2} \\
&= a + \overline{ab}\frac{\overline{at_1}}{\overline{at_2}} \\
&= a + (b-a)\frac{\alpha}{\beta} \\
&= a + \beta(b-a),
\end{aligned}
\tag{4-2-11}
$$

$$\begin{aligned}
t_1 &= a + \overline{at_1} \\
&= a + (\overline{t_2 b}) \\
&= a + [b - a - \beta(b - a)] \\
&= a + (1 - \beta)(b - a) \\
&= a + \alpha(b - a).
\end{aligned} \tag{4-2-12}$$

4.2.1.3 算法分析

$$[a^{n+1}, b^{n+1}] \subset [a^n, b^n] \tag{4-2-13}$$

$$b^n - a^n = \beta^n \cdot (b-a) \to 0, \quad n \to \infty. \tag{4-2-14}$$

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$a < t_1 < t_2 < b$，则

$$L_n \leq L_{n-2} + L_{n-1}. \tag{4-2-15}$$

显然，$L_0 = L_1 = 1$。

因此

$$L_n = L_{n-2} + L_{n-1}, \tag{4-2-16}$$

$$L_0 = L_1 = 1. \tag{4-2-17}$$

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

$$L_n = Ar_1^n + Br_2^n. \tag{4-2-18}$$

所以

$$L_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right). \tag{4-2-19}$$

因为

$$\frac{L_{n-2}}{L_n} + \frac{L_{n-1}}{L_n} = 1, \tag{4-2-20}$$

所以如下选择n个试验点，

$$t_1 = a + \frac{L_{n-2}}{L_n}(b-a), \tag{4-2-21}$$

$$t_2 = a + \frac{L_{n-1}}{L_n}(b-a). \tag{4-2-22}$$

后续过程同黄金分割法。

4.2.2.3 算法分析

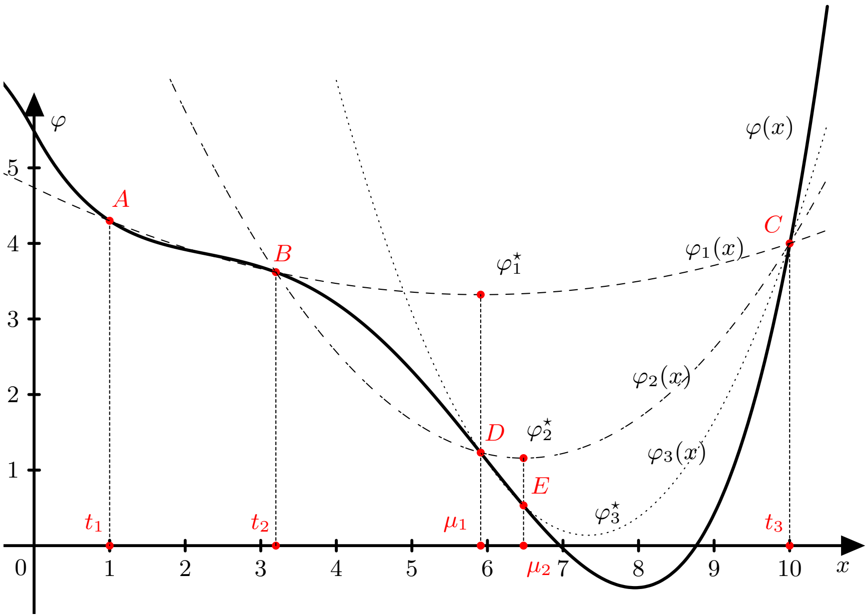
由式子(4-2-19)易得，

$$\lim_{n\to\infty} \frac{L_{n-1}}{L_n} = \frac{\sqrt{5}-1}{2} \approx 0.618. \tag{4-2-23}$$

4.2.3 二次插值法

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$t^{\star}$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。二次插值法也称为抛物线插值法。


φ
$\varphi(x)$
A
B
C
D
E
$\varphi_1(x)$
$\varphi_2(x)$
$\varphi_3(x)$
$\varphi_1^\star$
$\varphi_2^\star$
$\varphi_3^\star$
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

$$\begin{aligned} P(t) = {} & \varphi_1 \frac{(t-t_2)(t-t_3)}{(t_1-t_2)(t_1-t_3)} + \varphi_2 \frac{(t-t_1)(t-t_3)}{(t_2-t_1)(t_2-t_3)} \\ & + \varphi_3 \frac{(t-t_1)(t-t_2)}{(t_3-t_1)(t_3-t_2)}. \end{aligned} \tag{4-2-24}$$

令$P'(t) = 0$，得

$$\mu = \frac{\varphi_1(t_2^2 - t_3^2) + \varphi_2(t_3^2 - t_1^2) + \varphi_3(t_1^2 - t_2^2)}{2\big((t_2 - t_3)\varphi_1 + (t_3 - t_1)\varphi_2 + (t_1 - t_2)\varphi_3\big)}. \tag{4-2-25}$$

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

$$P(t)=\frac{\varphi_2-\varphi_1-\varphi_1'(t_2-t_1)}{(t_2-t_1)^2}(t-t_1)^2+\varphi_1'(t-t_1)+\varphi_1 \quad (4\text{-}2\text{-}26)$$

令$P'(t)=0$，得

$$\mu=t_1-\frac{\varphi_1'(t_2-t_1)^2}{2[\varphi_2-\varphi_1-\varphi_1'(t_2-t_1)]} \quad (4\text{-}2\text{-}27)$$

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

$$\mu = t_1 + (t_2 - t_1)\left(1 - \frac{\varphi_2' + \omega + \kappa}{\varphi_2' - \varphi_1' + 2\omega}\right). \tag{4-2-28}$$

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$|\varphi_1'| \leq \varepsilon$，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$|\varphi_2'| \leq \varepsilon$，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$t^\star = \mu$；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

$$\begin{cases} (i) \quad f(\boldsymbol{X}^k + t_k\boldsymbol{P}^k) \leq f(\boldsymbol{X}^k) + \rho t_k \nabla f(\boldsymbol{X}^k)^T\boldsymbol{P}^k; \\ (ii) \quad f(\boldsymbol{X}^k + t_k\boldsymbol{P}^k) \geq f(\boldsymbol{X}^k) + (1-\rho) t_k \nabla f(\boldsymbol{X}^k)^T\boldsymbol{P}^k. \end{cases} \tag{4-3-29}$$

这里，$0 < \rho < \frac{1}{2}$。

$Goldstein$准则的直观意义是：避免t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$\varphi(t)=f(\boldsymbol{X}^k+t\boldsymbol{P}^k)$，那么$Goldstein$准则可改写为：

$$\begin{cases}\varphi(t_k)\leq\varphi(0)+\rho t_k\varphi'(0);\\ \varphi(t_k)\geq\varphi(0)+(1-\rho)t_k\varphi'(0).\end{cases}\tag{4-3-30}$$

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：

$$\begin{cases} (i) \quad f(\boldsymbol{X}^k + t_k \boldsymbol{P}^k) \leq f(\boldsymbol{X}^k) + \rho t_k \nabla f(\boldsymbol{X}^k)^T \boldsymbol{P}^k; \\ (ii) \quad \nabla f(\boldsymbol{X}^k + t_k \boldsymbol{P}^k)^T \boldsymbol{P}^k \geq \sigma \nabla f(\boldsymbol{X}^k)^T \boldsymbol{P}^k, \\ \qquad \text{亦即} \varphi'(t_k) \geq \sigma \varphi'(0). \end{cases} \tag{4-3-31}$$

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，

$$\nabla f(\boldsymbol{X}^k + t_k\boldsymbol{P}^k)^T\boldsymbol{P}^k \geq \sigma\nabla f(\boldsymbol{X}^k)^T\boldsymbol{P}^k \tag{4-3-32}$$

是精确一维搜索满足的正交条件$\nabla f(\boldsymbol{X}^{k+1})^T\boldsymbol{P}^k = 0$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

$$\left|\nabla f(\boldsymbol{X}^k + t_k\boldsymbol{P}^k)^T\boldsymbol{P}^k\right| \leq -\sigma\nabla f(\boldsymbol{X}^k)^T\boldsymbol{P}^k, \tag{4-3-33}$$

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*准则

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

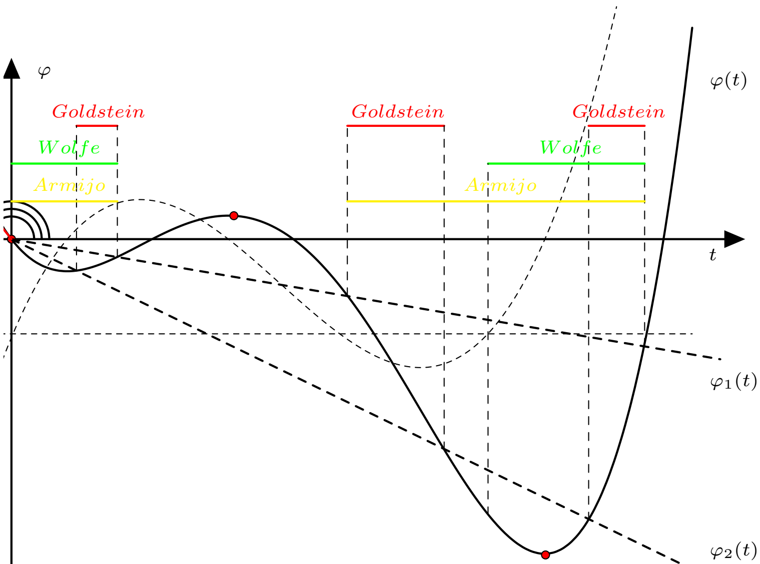
$$f(\boldsymbol{X}^k + \beta^{m_k}\tau\boldsymbol{P}^k) \le f(\boldsymbol{X}^k) + \rho\beta^{m_k}\tau\nabla f(\boldsymbol{X}^k)^T\boldsymbol{P}^k \tag{4-3-34}$$

成立的最小非负整数。

若令$t_k = \beta^{m_k}\tau$，其就是$Goldstein$准则中的第一个准则

$$f(\boldsymbol{X}^k + t_k\boldsymbol{P}^k) \leq f(\boldsymbol{X}^k) + \rho t_k \nabla f(\boldsymbol{X}^k)^T \boldsymbol{P}^k, \tag{4-3-35}$$

即$\varphi(t_k) \leq \varphi(0) + \rho t_k \varphi'(0)$。


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$\boldsymbol{P}^k$与其梯度方向$-g_k = -\nabla f(\boldsymbol{X}^k)$成锐角。并且要求其夹角$\theta_k$满足：$\theta_k \leq \frac{\pi}{2} - \mu$，$\forall k \ (\mu > 0)$。

采用非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$\boldsymbol{X}^0 \in \boldsymbol{R}^n$，允许误差$0 \leq \varepsilon < 1$，置$k = 0$;

步骤 2 若$\|\nabla f(\boldsymbol{X}^k)\| \leq \varepsilon$，算法停止，$\boldsymbol{X}^k \approx \boldsymbol{X}^\star$；否则求出满足$\nabla f(\boldsymbol{X}^k)^T \boldsymbol{P}^k < 0$ 的下降方向$\boldsymbol{P}^k$;

步骤 3 求出步长因子t_k，使其满足$Goldstein$准则或$Wolfe$准则；

步骤 4 令$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k + t_k \boldsymbol{P}^k$，$k := k+1$，转步骤(2)。

下面给出基于这些非精确一维搜索的一般下降算法的总体收敛性定理。

定理 4.4 若上述算法中，步长因子t_k满足$Goldstein$准则或$Wolfe$准则，且对$\forall k$，有

$$\theta_k \le \frac{\pi}{2} - \mu, \tag{4-3-36}$$

若$\nabla f(\boldsymbol{X})$在水平集$L = \left\{\boldsymbol{X} \mid f(\boldsymbol{X}) < f(\boldsymbol{X}^0)\right\}$上一致连续。那么，或者对某个$k$，有$\nabla f(\boldsymbol{X}^k) = 0$，或者$f(\boldsymbol{X}^k) \to -\infty$，或者$\nabla f(\boldsymbol{X}^k) \to 0$。

定理 4.5 设函数$f(\boldsymbol{X})$连续可微，$\nabla f(\boldsymbol{X})$满足$Lipschitz$条件：

$$\|\nabla f(\boldsymbol{X}) - \nabla f(\boldsymbol{Y})\| \leq M \|\boldsymbol{X} - \boldsymbol{Y}\|. \tag{4-3-37}$$

又设算法采用$Wolfe$准则，且$\boldsymbol{P}^k$与$-\nabla f(\boldsymbol{X}^k)$的夹角$\theta_k$满足

$$\theta_k \leq \frac{\pi}{2} - \mu, \quad (0 < \mu < \frac{\pi}{2}) \tag{4-3-38}$$

那么由算法产生的点列$\{\boldsymbol{X}^k\}$，或者对某个k，有$\nabla f(\boldsymbol{X}^k) = 0$，或者$f(\boldsymbol{X}^k) \to -\infty$，或者$\nabla f(\boldsymbol{X}^k) \to 0$。

下面定理给出非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 4.6 设t_k满足非精确一维搜索准则中第一条，若函数还满足：

$$m\|\boldsymbol{Y}\|^2 \le \boldsymbol{Y}^T\nabla^2 f(\boldsymbol{X})\boldsymbol{Y} \le M\|\boldsymbol{Y}\|^2, \tag{4-3-39}$$

$$(\boldsymbol{Y}-\boldsymbol{Z})^T\big(\nabla f(\boldsymbol{Y})-\nabla f(\boldsymbol{Z})\big) \ge \eta\|\boldsymbol{Y}-\boldsymbol{Z}\|^2. \tag{4-3-40}$$

则必有

$$f(\boldsymbol{X}^k)-f(\boldsymbol{X}^k+t_k\boldsymbol{P}^k) \ge \frac{\rho\eta}{1+\sqrt{M/m}}\|t_k\boldsymbol{P}^k\|^2. \tag{4-3-41}$$

§ 4.4 最速下降法

4.4.1 基本思想

求解

$$\min_{\boldsymbol{X}\in\boldsymbol{R}^n} f(\boldsymbol{X}) \tag{4-4-42}$$

假设上式中目标函数具有一阶连续偏导数。且具有极小点$\boldsymbol{X}^\star$，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

$$f\big(\boldsymbol{X}^k + t_k \nabla f(\boldsymbol{X}^k)\big) = \min_t f\big(\boldsymbol{X}^k + t \nabla f(\boldsymbol{X}^k)\big). \tag{4-4-43}$$

令

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - t_k \nabla f(\boldsymbol{X}^k). \tag{4-4-44}$$

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$\boldsymbol{X}^\star$。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，$\boldsymbol{g}^0 \triangleq \nabla f(\boldsymbol{X}^0)$，置$k=0$。

步骤 2 一维搜索$t_k = arg\min f(\boldsymbol{X}^k - t\nabla f(\boldsymbol{X}^k))$，$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - t_k\nabla f(\boldsymbol{X}^k)$，计算$f^{k+1}$，$\boldsymbol{g}^{k+1} \triangleq \nabla f(\boldsymbol{X}^{k+1})$。

步骤 3 若$\|\boldsymbol{g}^{k+1}\| \le \varepsilon$，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 4.7 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$\boldsymbol{g}^\star\neq 0$，设子列

$$\boldsymbol{X}^{k_i}\rightarrow\boldsymbol{X}^\star,\quad \boldsymbol{X}^{k_i+1}\rightarrow\boldsymbol{X}^{\star\star},\quad i\rightarrow\infty.$$

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

$$f(\boldsymbol{X}^{k_{i+1}})\leq f(\boldsymbol{X}^{k_i+1})\leq f(\boldsymbol{X}^{k_i}). \tag{4-4-45}$$

因而

$$f(\boldsymbol{X}^{\star}) = f(\boldsymbol{X}^{\star\star}). \tag{4-4-46}$$

但$\boldsymbol{g}^{\star} \neq 0$，故存在$t^{\star} > 0$（充分小），有

$$\begin{aligned} f(\boldsymbol{X}^{\star} - t\boldsymbol{g}^{\star}) &= f(\boldsymbol{X}^{\star}) - t\nabla f(\boldsymbol{X}^{\star})^T \boldsymbol{g}^{\star} + o(t\|\boldsymbol{g}^{\star}\|) \\ &= f(\boldsymbol{X}^{\star}) - t(\boldsymbol{g}^{\star})^T \boldsymbol{g}^{\star} + o(t\|\boldsymbol{g}^{\star}\|) \\ &= f(\boldsymbol{X}^{\star}) - t\|\boldsymbol{g}^{\star}\|^2 + o(t\|\boldsymbol{g}^{\star}\|) \end{aligned} \tag{4-4-47}$$

$$\Longrightarrow$$

$$f(\boldsymbol{X}^{\star} - t^{\star}\boldsymbol{g}^{\star}) - f(\boldsymbol{X}^{\star}) < 0$$

又因为

$$f(\boldsymbol{X}^{k_i+1}) = \min f(\boldsymbol{X}^{k_i} - t_{k_i}\boldsymbol{g}^{k_i}) \leq f(\boldsymbol{X}^{k_i} - t^{\star}\boldsymbol{g}^{k_i}). \tag{4-4-48}$$

所以

$$f(\boldsymbol{X}^{\star\star}) \leq f(\boldsymbol{X}^{\star} - t^{\star}\boldsymbol{g}^{\star}) < f(\boldsymbol{X}^{\star}). \tag{4-4-49}$$

矛盾！ ∎

4.4.4 最优步长

若$f(\boldsymbol{X})$是（正定）二次函数，由$Taylor$公式得

$$\begin{aligned} f\left(\boldsymbol{X}^k - t\nabla f(\boldsymbol{X}^k)\right) = & f(\boldsymbol{X}^k) - t\nabla f(\boldsymbol{X}^k)^T \nabla f(\boldsymbol{X}^k) \\ & + \frac{1}{2}t^2 \nabla f(\boldsymbol{X}^k)^T \boldsymbol{H}(\boldsymbol{X}^k) \nabla f(\boldsymbol{X}^k). \end{aligned} \tag{4-4-50}$$

其中$\boldsymbol{H}(\boldsymbol{X}^k)$是$f(\boldsymbol{X})$在点$\boldsymbol{X}^k$的$\boldsymbol{Hesse}$矩阵。

令

$$\begin{aligned} \frac{df\left(\boldsymbol{X}^k - t\nabla f(\boldsymbol{X}^k)\right)}{dt} = & -\nabla f(\boldsymbol{X}^k)^T \nabla f(\boldsymbol{X}^k) \\ & + t\nabla f(\boldsymbol{X}^k)^T \boldsymbol{H}(\boldsymbol{X}^k) \nabla f(\boldsymbol{X}^k) = 0. \end{aligned} \tag{4-4-51}$$

可得最优步长

$$t_k = \frac{\nabla f(\boldsymbol{X}^k)^T \nabla f(\boldsymbol{X}^k)}{\nabla f(\boldsymbol{X}^k)^T \boldsymbol{H}(\boldsymbol{X}^k) \nabla f(\boldsymbol{X}^k)} = \frac{(\boldsymbol{g}^k)^T \boldsymbol{g}^k}{(\boldsymbol{g}^k)^T \boldsymbol{H}^k \boldsymbol{g}^k} \tag{4-4-52}$$

由此得到第$k+1$步的迭代点为：

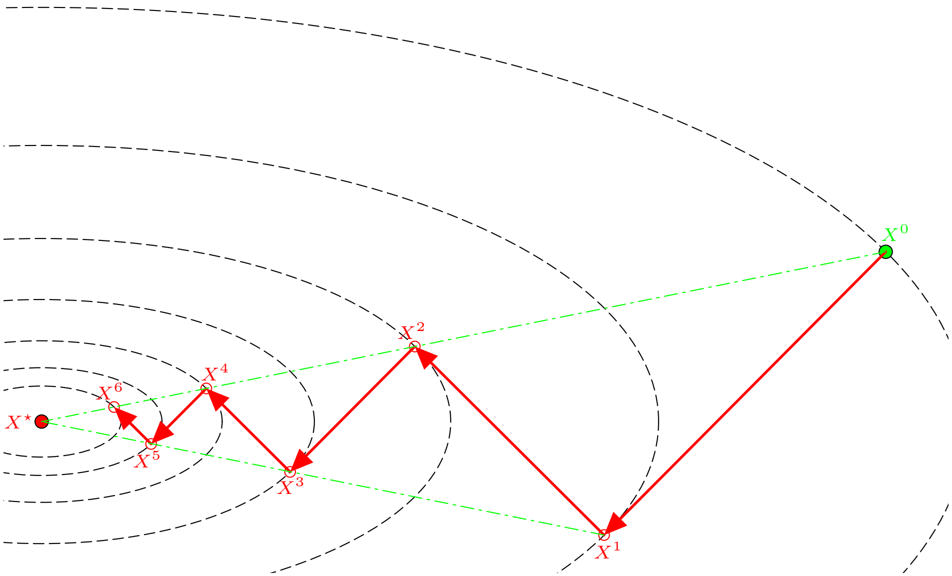
$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - \frac{(\boldsymbol{g}^k)^T \boldsymbol{g}^k}{(\boldsymbol{g}^k)^T \boldsymbol{H}^k \boldsymbol{g}^k} \boldsymbol{g}^k. \tag{4-4-53}$$

因为$\varphi'(t) = \nabla f(\boldsymbol{X} + t\boldsymbol{P})^T \boldsymbol{P}$，所以令其为零，得

$$(\boldsymbol{g}^{k+1})^T \boldsymbol{g}^k = 0. \tag{4-4-54}$$

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 4.2 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 $\boldsymbol{g}^k = \nabla f(\boldsymbol{X}) = \big(2(x_1 - 1), 2(x_2 - 1)\big)^T$

$$\boldsymbol{H}^k = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\boldsymbol{X}^0 = (0, 0)^T$$

$$t_0 = \frac{(\boldsymbol{g}^0)^T \boldsymbol{g}^0}{(\boldsymbol{g}^0)^T \boldsymbol{H}^0 \boldsymbol{g}^0} = \frac{1}{2}$$

$$\boldsymbol{X}^1 = \boldsymbol{X}^0 - t_0 \boldsymbol{g}^0 = (0, 0)^T - \frac{1}{2}(-2, -2)^T = (1, 1)^T$$

$\nabla f(\boldsymbol{X}^1) = (0, 0)^T$，$\|\nabla f(\boldsymbol{X}^1)\| < \varepsilon$

$$\boldsymbol{X}^\star = \boldsymbol{X}^1. \qquad \blacksquare$$

§ 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$\boldsymbol{X}^{\star}$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

$$\begin{aligned} f(\boldsymbol{X}) \approx q(\boldsymbol{X}) \triangleq & f(\boldsymbol{X}^k) + \boldsymbol{g}(\boldsymbol{X}^k)^T(\boldsymbol{X} - \boldsymbol{X}^k) \\ & + \frac{1}{2}(\boldsymbol{X} - \boldsymbol{X}^k)^T \boldsymbol{H}(\boldsymbol{X}^k)(\boldsymbol{X} - \boldsymbol{X}^k). \end{aligned} \tag{4-5-55}$$

其中$q_k(\boldsymbol{X})$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$\boldsymbol{Hesse}$矩阵。

令$\nabla q_k(\boldsymbol{X}) = \boldsymbol{H}(\boldsymbol{X}^k)(\boldsymbol{X} - \boldsymbol{X}^k) + \boldsymbol{g}(\boldsymbol{X}^k) = 0$，$\boldsymbol{X}_k^\star$为$q_k(\boldsymbol{X})$的最优解，则有

$$\boldsymbol{H}(\boldsymbol{X}^k)(\boldsymbol{X}_k^\star - \boldsymbol{X}^k) = -\boldsymbol{g}(\boldsymbol{X}^k). \tag{4-5-56}$$

若$\boldsymbol{Hesse}$矩阵正定，则

$$\boldsymbol{X}_k^\star - \boldsymbol{X}^k = -\boldsymbol{H}^{-1}(\boldsymbol{X}^k)\boldsymbol{g}(\boldsymbol{X}^k). \tag{4-5-57}$$

因此，

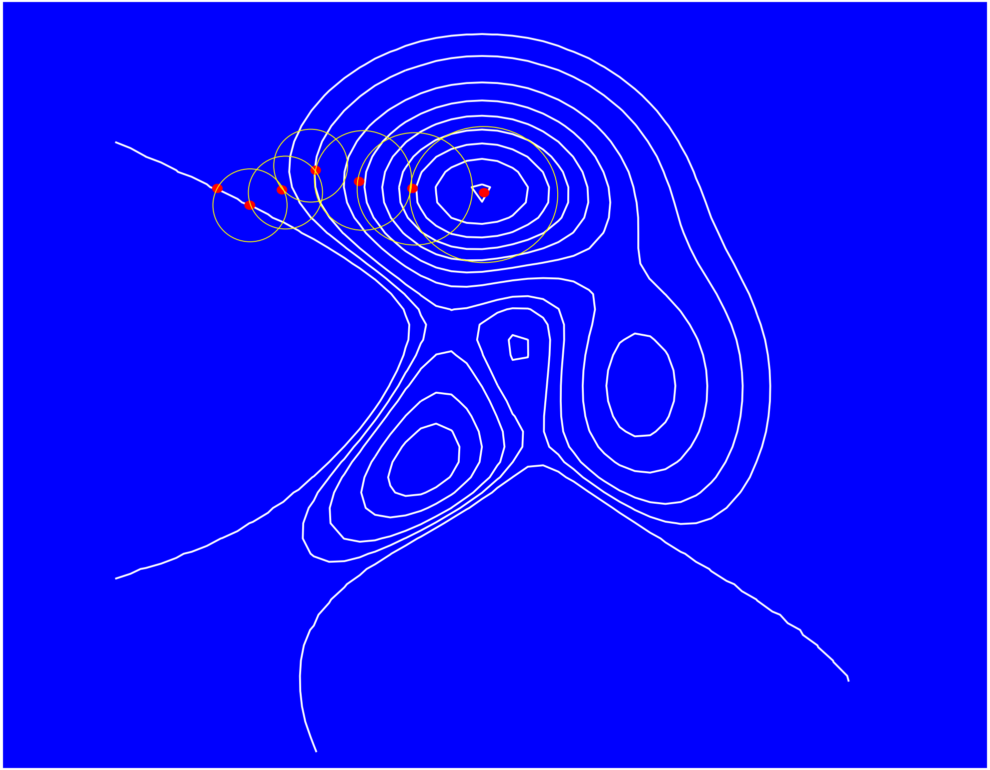
$$\boldsymbol{X}_k^\star = \boldsymbol{X}^k - \boldsymbol{H}^{-1}(\boldsymbol{X}^k)\boldsymbol{g}(\boldsymbol{X}^k). \tag{4-5-58}$$

令$\boldsymbol{X}_k^\star \triangleq \boldsymbol{X}^{k+1}$，即利用$\boldsymbol{X}^{k+1}$作为$f(\boldsymbol{X})$的极小点$\boldsymbol{X}^\star$的新的近似。式(4-5-59)就是**牛顿迭代公式**。

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k - \boldsymbol{H}^{-1}(\boldsymbol{X}^k)\boldsymbol{g}(\boldsymbol{X}^k). \tag{4-5-59}$$

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$\boldsymbol{X}^{\star}$的新的近似。

得到$f(\boldsymbol{X})$的极小点$\boldsymbol{X}^{\star}$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$\boldsymbol{X}^{\star}$ 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 4.3 求$f(\boldsymbol{X}) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$的极小点。

算法 4.6 牛顿算法

步骤 1 给定目标函数$f(\boldsymbol{X})$以及梯度$\boldsymbol{g}(\boldsymbol{X})$，$\boldsymbol{Hesse}$矩阵$\boldsymbol{H}(\boldsymbol{X})$，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	0	1	2	3	4	5	6
$\boldsymbol{X}^k$	$\binom{0.00}{3.00}$	$\binom{0.67}{0.33}$	$\binom{1.11}{0.56}$	$\binom{1.41}{0.70}$	$\binom{1.61}{0.80}$	$\binom{1.74}{0.87}$	$\binom{1.83}{0.91}$
$f(\boldsymbol{X}^k)$	52	3.13	0.63	0.12	0.02	0.005	0.0009

解 $\boldsymbol{X}^0=(0,3)^T$，$\boldsymbol{g}^0=(-44,24)^T$，$\boldsymbol{H}^0=\begin{pmatrix}50 & -4\\ -4 & 8\end{pmatrix}$,

$$\boldsymbol{P}^0=-(\boldsymbol{H}^0)^{-1}\boldsymbol{g}^0=(0.67,-2.67)^T,$$

$$\boldsymbol{X}^1=\boldsymbol{X}^0+\boldsymbol{P}^0=(0.67,0.33)^T,\ f(\boldsymbol{X}^1)=3.13。$$

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$\boldsymbol{Hesse}$矩阵奇异，则用最速下降方向代替牛顿方向。

(2)$\boldsymbol{Hesse}$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：

(I)当$|(\boldsymbol{g}^k)^T\boldsymbol{P}^k| \leq \varepsilon\|\boldsymbol{g}^k\|\|\boldsymbol{P}^k\|$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$(\boldsymbol{g}^k)^T\boldsymbol{P}^k \leq -\varepsilon\|\boldsymbol{g}^k\|\|\boldsymbol{P}^k\|$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。

(III)当$(\boldsymbol{g}^k)^T\boldsymbol{P}^k > \varepsilon\|\boldsymbol{g}^k\|\|\boldsymbol{P}^k\|$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 4.8 (牛顿法收敛性定理) 设

(1)$f(\boldsymbol{X})\in C^2$;

(2)$\boldsymbol{H}(\boldsymbol{X})>0$;

(3)水平集$\boldsymbol{C}_0=\{\boldsymbol{X}|f(\boldsymbol{X})\leq f(\boldsymbol{X}^0)\}$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$\boldsymbol{X}^\star$,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

$$\begin{aligned} f(\boldsymbol{X}^k + \lambda \boldsymbol{P}^k) &= f[\boldsymbol{X}^k - \lambda(\boldsymbol{H}^k)^{-1}\boldsymbol{g}^k] \\ &= f(\boldsymbol{X}^k) + (g^k)^T[-\lambda(\boldsymbol{H}^k)^{-1}\boldsymbol{g}^k] + o(\lambda) \\ &= f(\boldsymbol{X}^k) + \lambda\Big(- (\boldsymbol{g}^k)^T(\boldsymbol{H}^k)^{-1}\boldsymbol{g}^k + \frac{o(\lambda)}{\lambda}\Big) \end{aligned} \tag{4-5-60}$$

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$\boldsymbol{X}^k \in \boldsymbol{C}_0 = \{\boldsymbol{X} | f(\boldsymbol{X}) \leq f(\boldsymbol{X}^0)\}$，故$\{\boldsymbol{X}^k\}$为有界点列，则其必有极限点。

设$\boldsymbol{X}^\star$为任一极限点，且设$\lim_{i\to+\infty} \boldsymbol{X}^{k_i} = \boldsymbol{X}^\star$，再由$f(\boldsymbol{X})$的连续性有$\lim_{i\to+\infty} f(\boldsymbol{X}^{k_i}) = f(\boldsymbol{X}^\star)$，又因$\{\boldsymbol{X}^k\}$为为单调下降且有下界的数列，因而

$$\lim_{i\to+\infty} f(\boldsymbol{X}^{k_i}) = \lim_{i\to+\infty} f(\boldsymbol{X}^{k_i+1}) = f(\boldsymbol{X}^\star).$$

假设$\boldsymbol{g}^\star \neq 0$，则

$$(\boldsymbol{g}^\star)^T(\boldsymbol{H}^\star)^{-1}\boldsymbol{g}^\star > 0. \tag{4-5-61}$$

其中$(\boldsymbol{H}^\star)^{-1} > 0$，所以$\exists\delta > 0$，当$0 < \lambda < \delta$时有

$$\begin{aligned} &= f(\boldsymbol{X}^\star) - \lambda(\boldsymbol{g}^\star)^T(\boldsymbol{H}^\star)^{-1}\boldsymbol{g}^\star + o(\lambda) \\ &= f(\boldsymbol{X}^\star) + \lambda\Big(- (\boldsymbol{g}^\star)^T(\boldsymbol{H}^\star)^{-1}\boldsymbol{g}^\star + \frac{o(\lambda)}{\lambda}\Big) \end{aligned} \tag{4-5-62}$$

因此

$$f\big(\boldsymbol{X}^\star + \lambda(\boldsymbol{H}^\star)^{-1}\boldsymbol{g}^\star\big) < f(\boldsymbol{X}^\star), \quad 0 < \lambda < \delta. \tag{4-5-63}$$

又

$$f(\boldsymbol{X}^{k_i+1}) = \min_{\lambda>0} f(\boldsymbol{X}^{k_i} + \lambda \boldsymbol{P}^{k_i}) \leq f\big(\boldsymbol{X}^{k_i} + \lambda(\boldsymbol{H}^{k_i})^{-1}\boldsymbol{g}^{k_i}\big). \quad (4\text{-}5\text{-}64)$$

令$i \to +\infty$，则有

$$f(\boldsymbol{X}^{\star}) \leq f\big(\boldsymbol{X}^{\star} + \lambda(\boldsymbol{H}^{\star})^{-1}\boldsymbol{g}^{\star}\big) < f(\boldsymbol{X}^{\star}). \quad (4\text{-}5\text{-}65)$$

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。

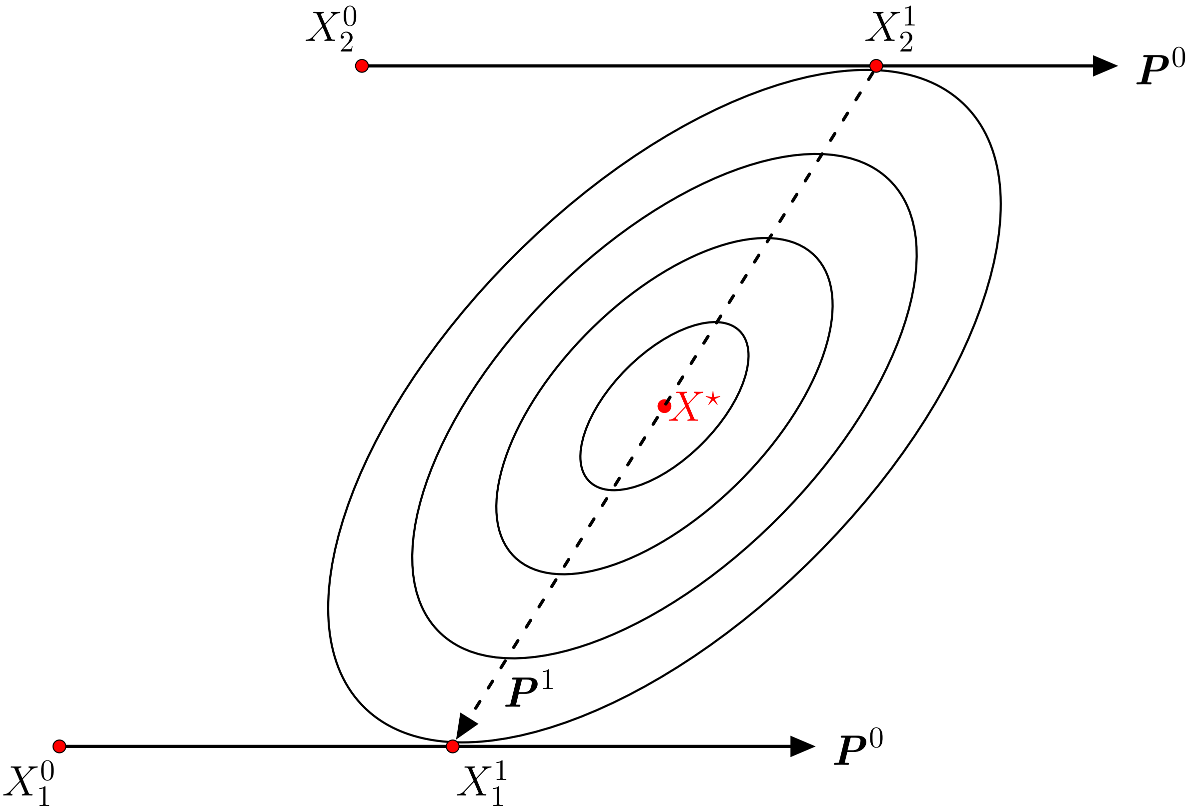
而严格凸函数的最小点必唯一，所以命题得证！ ∎

§ 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$\boldsymbol{Hesse}$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$\boldsymbol{Hesse}$矩阵，且具有二次终止性。

$$f(\boldsymbol{X})=\frac{1}{2}\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}+\boldsymbol{b}^T\boldsymbol{X}+\boldsymbol{c}. \tag{4-6-66}$$

在这里，$\boldsymbol{Q}>0$。


X_2^0
X_2^1
$\boldsymbol{P}^0$
$X^\star$
$\boldsymbol{P}^1$
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$\boldsymbol{X}^\star$。

$$\boldsymbol{X}_2^0 \xrightarrow{\quad \boldsymbol{P}^0 \quad} \boldsymbol{X}_2^1 \xrightarrow{\quad \boldsymbol{P}^1=\boldsymbol{X}_1^1-\boldsymbol{X}_2^1 \quad} \boldsymbol{X}^\star$$

考察$\boldsymbol{P}^1$，

$$\nabla f(\boldsymbol{X}^\star)=\boldsymbol{Q}\boldsymbol{X}^\star+b=0. \tag{4-6-67}$$

即，

$$\nabla f(\boldsymbol{X}_2^1)+t_1\boldsymbol{Q}\boldsymbol{P}^1=\boldsymbol{Q}(X^\star-t_1\boldsymbol{P}^1)+b+t_1\boldsymbol{Q}\boldsymbol{P}^1=0. \tag{4-6-68}$$

所以，

$$(\boldsymbol{P}^0)^T\nabla f(\boldsymbol{X}_2^1)+t_1(P^0)^T\boldsymbol{Q}\boldsymbol{P}^1=0\Rightarrow(P^0)^T\boldsymbol{Q}\boldsymbol{P}^1=0. \tag{4-6-69}$$

定义 4.4 (共轭) 设$\boldsymbol{Q}$是$n\times n$对称正定矩阵，如果$\boldsymbol{X}$和$\boldsymbol{Q}\boldsymbol{Y}$正交，即

$$\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{Y}=0. \tag{4-6-70}$$

则称$\boldsymbol{X}$和$\boldsymbol{Y}$关于$\boldsymbol{Q}$共轭。

定义 4.5 (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$\boldsymbol{Q}$共轭，即

$$(\boldsymbol{P}^i)^T\boldsymbol{Q}(\boldsymbol{P}^j) = 0; \qquad i, j = 0 \sim n, i \neq j. \tag{4-6-71}$$

则称这组向量关于$\boldsymbol{Q}$共轭。

注意：共轭和正交的联系。

定理 4.9 如果非零向量组$\boldsymbol{P}^0$、$\boldsymbol{P}^1$、$\cdots$、$\boldsymbol{P}^{m-1}$（是）关于$\boldsymbol{Q}$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个的向量组必线性相关，所以，

推论 4.10 在$\boldsymbol{R}^n$空间中，互相共轭的非零向量的向量个数不超过n个。

定理 4.11 (扩张子空间定理) 设

$$f(\boldsymbol{X}) = \frac{1}{2}\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{b}^T\boldsymbol{X} + c \tag{4-6-72}$$

为正定二次函数，向量组$\boldsymbol{P}^i, (i = 0 \sim n-1)$为$\boldsymbol{Q}$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，那么至多经过$n$次迭代，必然收敛于问题(4-6-72) 的最优解$X^\star$。

证明 因为，

$$\nabla f(\boldsymbol{X}^{k+1}) = \nabla f(\boldsymbol{X}^k) + t_k \boldsymbol{Q}\boldsymbol{P}^k, \quad k = 0 \sim n-1. \tag{4-6-73}$$

所以若$\nabla f(\boldsymbol{X}^k) \neq 0$，$k = 0 \sim n-1$，则，

$$\nabla f(\boldsymbol{X}^n) = \nabla f(\boldsymbol{X}^{k+1}) + \sum_{i=k+1}^{n-1} t_i \boldsymbol{Q}\boldsymbol{P}^i. \tag{4-6-74}$$

因此对$k = 0 \sim n-1$有，

$$(\boldsymbol{P}^k)^T \nabla f(\boldsymbol{X}^n) = (\boldsymbol{P}^k)^T \nabla f(\boldsymbol{X}^{k+1}) + \sum_{i=k+1}^{n-1} t_i (\boldsymbol{P}^k)^T \boldsymbol{Q}\boldsymbol{P}^i = 0. \tag{4-6-75}$$

即，$\nabla f(\boldsymbol{X}^n)$与$\boldsymbol{P}^0$、$\cdots$、$\boldsymbol{P}^{n-1}$正交。

则下面线性系统，

$$(\boldsymbol{P}^0, \boldsymbol{P}^1, \cdots, \boldsymbol{P}^{n-1})^T \cdot \nabla f(\boldsymbol{X}^n) = 0. \tag{4-6-76}$$

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$\boldsymbol{Q}$共轭的向量组？

(1)设$\boldsymbol{v}^i$，$i = 0 \sim n-1$是一组线性无关的向量组，首先取

$$\boldsymbol{P}^0 = \boldsymbol{v}^0. \tag{4-6-77}$$

设已求得$\boldsymbol{P}^i$，$i = 0 \sim k$，现构造$\boldsymbol{P}^{k+1}$。令

$$\boldsymbol{P}^{k+1} = \boldsymbol{v}^{k+1} + \sum_{r=0}^{k} \beta_{k+1,r} \boldsymbol{P}^r.$$

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

$$(\boldsymbol{P}^j)^T \boldsymbol{Q} \cdot \boldsymbol{P}^{k+1} = (\boldsymbol{P}^j)^T \boldsymbol{Q} \cdot \boldsymbol{v}^{k+1} + \sum_{r=0}^{k} \beta_{k+1,r} (\boldsymbol{P}^j)^T \boldsymbol{Q} \cdot \boldsymbol{P}^r = 0.$$

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，

$$\boldsymbol{P}^{k+1}=\boldsymbol{v}^{k+1}-\sum_{j=0}^{k}\frac{(\boldsymbol{P}^j)^T\boldsymbol{Q}\boldsymbol{v}^{k+1}}{(\boldsymbol{P}^j)^T\boldsymbol{Q}\boldsymbol{P}^j}\boldsymbol{P}^j. \tag{4-6-78}$$

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组，相应的共轭方向法称为共轭梯度法。

4.6.1 共轭梯度法

首先构造

$$\boldsymbol{P}^0=-\boldsymbol{g}^0. \tag{4-6-79}$$

对$k=1\sim n-1$，令

$$\boldsymbol{P}^{k+1}=-\boldsymbol{g}^{k+1}+\alpha_k\boldsymbol{P}^k.$$

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

$$(\boldsymbol{P}^{k+1})^T\cdot\boldsymbol{Q}\boldsymbol{P}^k=-(\boldsymbol{g}^{k+1})^T\cdot\boldsymbol{Q}\boldsymbol{P}^k+\alpha_k(\boldsymbol{P}^k)^T\cdot\boldsymbol{Q}\boldsymbol{P}^k=0.$$

从而

$$\alpha_k=\frac{(\boldsymbol{g}^{k+1})^T\boldsymbol{Q}\boldsymbol{P}^k}{(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^k}.$$

即，

$$\boldsymbol{P}^{k+1}=-\boldsymbol{g}^{k+1}+\frac{(\boldsymbol{g}^{k+1})^T\boldsymbol{Q}\boldsymbol{P}^k}{(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^k}\boldsymbol{P}^k. \tag{4-6-80}$$

注：一些著名的的公式

推论 4.12 ($Daniel$,1967)

$$\alpha_k^D = \frac{(\boldsymbol{g}^{k+1})^T \boldsymbol{Q} \boldsymbol{P}^k}{(\boldsymbol{P}^k)^T \boldsymbol{Q} \boldsymbol{P}^k}.$$

推论 4.13 ($Sorenson$ &. $Wolfe$,1972)

$$\alpha_k^{SW} = \frac{(\boldsymbol{g}^{k+1})^T (\boldsymbol{g}^{k+1} - \boldsymbol{g}^k)}{(\boldsymbol{P}^k)^T (\boldsymbol{g}^{k+1} - \boldsymbol{g}^k)}.$$

推论 4.14 ($Myers$,1972)

$$\alpha_k^M = -\frac{(\boldsymbol{g}^{k+1})^T \boldsymbol{g}^{k+1}}{(\boldsymbol{P}^k)^T \boldsymbol{g}^k}.$$

推论 4.15 ($Flecher$ &. $Reeves$,**1964**)

$$\alpha_k^{FR} = \frac{\|\boldsymbol{g}^{k+1}\|^2}{\|\boldsymbol{g}^k\|^2}$$

推论 4.16 ($Polyak, Polak$ &. $Ribiere$,**1969**)

$$\alpha_k^{PRP} = \frac{(\boldsymbol{g}^{k+1})^T(\boldsymbol{g}^{k+1} - \boldsymbol{g}^k)}{(\boldsymbol{g}^k)^T\boldsymbol{g}^k}.$$

推论 4.17 (Dai &. $Yuan$,**1999**)

$$\alpha_k^{DY} = \frac{(\boldsymbol{g}^{k+1})^T\boldsymbol{g}^{k+1}}{(\boldsymbol{P}^k)^T(\boldsymbol{g}^{k+1} - \boldsymbol{g}^k)}.$$

4.6.1.1 *FR*共轭梯度法

算法 4.7 ***FR*共轭梯度算法**

步骤 1 给定目标函数$f(\boldsymbol{X})$，初始点$\boldsymbol{X}^0$，梯度$\boldsymbol{g}^0$，精度ε，置$k=0$。

步骤 2 若满足停机条件，输出$\boldsymbol{X}^k$。

步骤 3 计算$\boldsymbol{P}^k = \begin{cases} -\boldsymbol{g}^k, & k=0; \\ -\boldsymbol{g}^k + \frac{\|\boldsymbol{g}^k\|^2}{\|\boldsymbol{g}^{k-1}\|^2}\boldsymbol{P}^{k-1}, & k>0. \end{cases}$

步骤 4 计算$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k + t_k\boldsymbol{P}^k$，$\boldsymbol{g}^{k+1}$，$k=k+1$，转步骤2。

例 4.4 (*FR*共轭梯度法) 用***FR***方法求解$f(\boldsymbol{X}) = x_1^2 + 2x_2^2$，取初始点$\boldsymbol{X}^0 = (5,5)^T$。

解 $\boldsymbol{g} \triangleq \nabla f(\boldsymbol{X}) = (2x_1, 4x_2)^T, \boldsymbol{H} \triangleq \nabla^2 f(\boldsymbol{X}) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}.$

$\|\boldsymbol{g}^0\|^2 > \varepsilon$,

$\boldsymbol{P}^0 = -\boldsymbol{g}^0 = (-10, -20)^T$,

$$\begin{aligned}
\boldsymbol{X}^1 &= \boldsymbol{X}^0 + t_0 \boldsymbol{P}^0 = \boldsymbol{X}^0 - t_0 \boldsymbol{g}_0 \\
&= (5,5)^T + \frac{-(\boldsymbol{g}^0)^T \boldsymbol{P}^0}{(\boldsymbol{P}^0)^T \boldsymbol{H} \boldsymbol{P}^0}(-10, -20)^T \\
&= (5,5)^T + \frac{5}{18}(-10, -20)^T \\
&= \left(\frac{20}{9}, -\frac{5}{9}\right)^T.
\end{aligned}$$

$\|\boldsymbol{g}^1\| > \varepsilon$

$$
\begin{aligned}
\boldsymbol{P}^1 &= -\boldsymbol{g}^1 + \alpha_0 \boldsymbol{P}^0 \\
&= \left(-\frac{40}{9}, \frac{20}{9}\right)^T + \frac{\|\boldsymbol{g}^1\|^2}{\|\boldsymbol{g}^0\|^2}(-10, -20)^T \\
&= \left(-\frac{40}{9}, \frac{20}{9}\right)^T + \frac{4}{81}(-10, -20)^T \\
&= \left(-\frac{400}{81}, \frac{100}{81}\right)^T.
\end{aligned}
$$

$$\boldsymbol{X}^2 = \boldsymbol{X}^1 + t_1 \boldsymbol{P}^1$$

$$
\begin{aligned}
&= \left(\frac{20}{9}, -\frac{5}{9}\right)^T + \frac{-(\boldsymbol{g}^1)^T \boldsymbol{P}^1}{(\boldsymbol{P}^1)^T \boldsymbol{H} \boldsymbol{P}^1} \left(-\frac{400}{81}, \frac{100}{81}\right)^T \\
&= \left(\frac{20}{9}, -\frac{5}{9}\right)^T + \frac{9}{20} \left(-\frac{400}{81}, \frac{100}{81}\right)^T \\
&= (0,0)^T.
\end{aligned}
$$

$\|\boldsymbol{g}^2\|^2 = 0 < \varepsilon$，$\boldsymbol{X}^\star = \boldsymbol{X}^2 = (0,0)^T$。 ∎

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 4.18　对于二次正定函数$f(\boldsymbol{X}) = \frac{1}{2}\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{b}^T\boldsymbol{X} + c$，$\boldsymbol{X} \in \boldsymbol{R}^n$，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：

$$(\boldsymbol{g}^k)^T\boldsymbol{P}^j=0,\qquad 0\le j\le k-1;\tag{4-6-81}$$

$$(\boldsymbol{g}^k)^T\boldsymbol{g}^j=0,\qquad 0\le j\le k-1;\tag{4-6-82}$$

$$(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^j=0,\qquad 0\le j\le k-1;\tag{4-6-83}$$

$$(\boldsymbol{P}^k)^T\boldsymbol{g}^k=-\|\boldsymbol{g}^k\|^2\tag{4-6-84}$$

证明 $\boldsymbol{P}^0=-\boldsymbol{g}^0,\boldsymbol{P}^1=-\boldsymbol{g}^1+\alpha_0\boldsymbol{P}^0,$

$$(\boldsymbol{g}^1)^T\boldsymbol{P}^0=0,$$

$$(\boldsymbol{g}^1)^T\boldsymbol{g}^0=-(\boldsymbol{g}^1)^T\boldsymbol{P}^0=0,$$

$$(\boldsymbol{P}^1)^T\boldsymbol{Q}\boldsymbol{P}^0=0,$$

$$(\boldsymbol{P}^1)^T\boldsymbol{g}^1 = (-\boldsymbol{g}^1 + \alpha_0\boldsymbol{P}^0)^T\boldsymbol{g}^1 = -\|\boldsymbol{g}^1\|^2.$$

假设k成立。

$$\boldsymbol{g}^{k+1} - \boldsymbol{g}^k = \boldsymbol{Q}(\boldsymbol{X}^{k+1} - \boldsymbol{X}^k) = t_k\boldsymbol{Q}\boldsymbol{P}^k$$

$$\begin{aligned}
(\boldsymbol{g}^{k+1})^T\boldsymbol{P}^j &= (\boldsymbol{g}^k + t_k\boldsymbol{Q}\boldsymbol{P}^k)^T\boldsymbol{P}^j \\
&= \begin{cases} (\boldsymbol{g}^k)^T\boldsymbol{P}^j + t_k(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^j, & 0 \le j \le k-1; \\ (\boldsymbol{g}^{k+1})^T\boldsymbol{P}^k, & j = k. \end{cases} \\
&= 0, \quad 0 \le j \le k.
\end{aligned} \tag{4-6-85}$$

$$
\begin{aligned}
(\boldsymbol{g}^{k+1})^T\boldsymbol{g}^j &= (\boldsymbol{g}^k + t_k\boldsymbol{Q}\boldsymbol{P}^k)^T\boldsymbol{g}^j \\
&= (\boldsymbol{g}^k)^T\boldsymbol{g}^j + t_k(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{g}^j \\
&= (\boldsymbol{g}^k)^T\boldsymbol{g}^j - t_k(\boldsymbol{P}^k)^T\boldsymbol{Q}(\boldsymbol{P}^j - \alpha_{j-1}\boldsymbol{P}^{j-1}) \\
&= \begin{cases} 0, & 0 \le j \le k-1; \\ (\boldsymbol{g}^k)^T\boldsymbol{g}^k - t_k(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^k, & j = k. \end{cases} \\
&= \begin{cases} 0, & 0 \le j \le k-1; \\ (\boldsymbol{g}^k)^T\boldsymbol{g}^k + \frac{(\boldsymbol{g}^k)^T\boldsymbol{P}^k}{(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^k}(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^k, & j = k. \end{cases} \\
&= 0, \quad 0 \le j \le k.
\end{aligned}
\tag{4-6-86}
$$

$$
\begin{aligned}
&(\boldsymbol{P}^{k+1})^T\boldsymbol{Q}\boldsymbol{P}^j = (-\boldsymbol{g}^{k+1}+\alpha_k\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^j\\
&=\frac{1}{t_j}(\boldsymbol{g}^{k+1})^T(\boldsymbol{g}^j-\boldsymbol{g}^{j+1})+\alpha_k(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^j\\
&=\begin{cases} 0, & 0\le j\le k-1;\\ -\frac{1}{t_k}(\boldsymbol{g}^{k+1})^T\boldsymbol{g}^{k+1}+\alpha_k(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^k, & j=k.\end{cases}\\
&=\begin{cases} 0, & 0\le j\le k-1;\\ \frac{(\boldsymbol{g}^{k+1})^T\boldsymbol{g}^{k+1}}{(\boldsymbol{g}^k)^T\boldsymbol{P}^k}(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^k+\frac{(\boldsymbol{g}^{k+1})^T\boldsymbol{g}^{k+1}}{(\boldsymbol{g}^k)^T\boldsymbol{g}^k}(\boldsymbol{P}^k)^T\boldsymbol{Q}\boldsymbol{P}^k, & j=k.\end{cases}\\
&=0,\quad 0\le j\le k.
\end{aligned}
\tag{4-6-87}
$$

$$
\begin{aligned}
(\boldsymbol{P}^{k+1})^T\boldsymbol{g}^{k+1} &= (-\boldsymbol{g}^{k+1}+\alpha_k\boldsymbol{P}^k)^T\boldsymbol{g}^{k+1} \\
&= -(\boldsymbol{g}^{k+1})^T\boldsymbol{g}^{k+1} \qquad (4\text{-}6\text{-}88) \\
&= -\|\boldsymbol{g}^{k+1}\|^2.
\end{aligned}
$$

∎

4.6.2 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$\boldsymbol{Hesse}$矩阵$\boldsymbol{Q}$以及其逆$\boldsymbol{Q}^{-1}$，当问题的维数n较大时，计算量迅速增加。

4.6.2.2 对称秩一公式

拟牛顿法的基本思想：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$\boldsymbol{Hesse}$矩阵近似。

拟牛顿法也称为**变尺度法**。

对于二次函数

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

这里$\boldsymbol{Q}>0.$

牛顿迭代公式为

$$\boldsymbol{X}^{k+1}=\boldsymbol{X}^k-(\boldsymbol{Q}^k)^{-1}\boldsymbol{g}^k. \tag{4-6-89}$$

取$(\boldsymbol{Q}^k)^{-1}$某种近似$\boldsymbol{H}^k$，将其变为如下迭代

$$\boldsymbol{X}^{k+1}=\boldsymbol{X}^k-\boldsymbol{H}^k\boldsymbol{g}^k. \tag{4-6-90}$$

更一般地

$$\boldsymbol{X}^{k+1}=\boldsymbol{X}^{k}-t_k\boldsymbol{H}^{k}\boldsymbol{g}^{k}. \tag{4-6-91}$$

问题：怎么构造$\boldsymbol{H}^k$？

(1)搜索方向$\boldsymbol{P}^k=-\boldsymbol{H}^k\boldsymbol{g}^k$应该为下降方向，即

$$(\boldsymbol{g}^k)^T\boldsymbol{P}^k=-(\boldsymbol{g}^k)^T\boldsymbol{H}^k\boldsymbol{g}^k<0.$$

不妨令

$$\boldsymbol{H}^k>0. \tag{4-6-92}$$

(2)

$$
\begin{aligned}
f(\boldsymbol{X}^k) =& f(\boldsymbol{X}^{k+1}) + \nabla f(\boldsymbol{X}^{k+1})^T(\boldsymbol{X}^k - \boldsymbol{X}^{k+1}) \\
&+ \frac{1}{2}(\boldsymbol{X}^k - \boldsymbol{X}^{k+1})^T \nabla^2 f(\boldsymbol{X}^{k+1})(\boldsymbol{X}^k - \boldsymbol{X}^{k+1}).
\end{aligned} \tag{4-6-93}
$$

关于$\boldsymbol{X}^k$求导，得

$$
\nabla f(\boldsymbol{X}^k) = \nabla f(\boldsymbol{X}^{k+1}) + \nabla^2 f(\boldsymbol{X}^{k+1})(\boldsymbol{X}^k - \boldsymbol{X}^{k+1}). \tag{4-6-94}
$$

即

$$
\boldsymbol{g}^k = \boldsymbol{g}^{k+1} + \boldsymbol{Q}^{k+1}(\boldsymbol{X}^k - \boldsymbol{X}^{k+1}). \tag{4-6-95}
$$

因此，若$\boldsymbol{Q}^{k+1}$正定时，有

$$(\boldsymbol{Q}^{k+1})^{-1}(\boldsymbol{g}^{k+1}-\boldsymbol{g}^{k})=\boldsymbol{X}^{k+1}-\boldsymbol{X}^{k}. \tag{4-6-96}$$

令$\boldsymbol{Y}^{k}\triangleq\boldsymbol{g}^{k+1}-\boldsymbol{g}^{k}$，$\boldsymbol{S}^{k}\triangleq\boldsymbol{X}^{k+1}-\boldsymbol{X}^{k}$，有

$$(\boldsymbol{Q}^{k+1})^{-1}\boldsymbol{Y}^{k}=\boldsymbol{S}^{k}. \tag{4-6-97}$$

于是，不妨令$\boldsymbol{H}^{k+1}$满足上式，得

$$\boldsymbol{H}^{k+1}\boldsymbol{Y}^{k}=\boldsymbol{S}^{k}. \tag{4-6-98}$$

称式(4-6-98)为拟牛顿方程(条件).

(3)为了便于计算，令

$$\boldsymbol{H}^{k+1}=\boldsymbol{H}^{k}+\Delta\boldsymbol{H}^{k}. \tag{4-6-99}$$

并考虑修正项$\Delta\boldsymbol{H}^{k}$为秩1对称矩阵，即

$$\boldsymbol{H}^{k+1}-\boldsymbol{H}^{k}=\Delta\boldsymbol{H}^{k}=\boldsymbol{P}\begin{pmatrix}\lambda & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0\end{pmatrix}\boldsymbol{P}^{T}. \tag{4-6-100}$$

令$\boldsymbol{u}$表示$\boldsymbol{P}$的第一列，即

$$\boldsymbol{H}^{k+1}=\boldsymbol{H}^{k}+\lambda\boldsymbol{u}\boldsymbol{u}^{T}. \tag{4-6-101}$$

将此式带入拟牛顿方程，得

$$\lambda \boldsymbol{u}\boldsymbol{u}^T \boldsymbol{Y}^k = \boldsymbol{S}^k - \boldsymbol{H}^k \boldsymbol{Y}^k \triangleq \boldsymbol{Z}^k. \tag{4-6-102}$$

不妨令

$$\boldsymbol{u} = \alpha \boldsymbol{Z}^k. \tag{4-6-103}$$

将此式带入上式

$$\lambda \alpha^2 \boldsymbol{Z}^k (\boldsymbol{Z}^k)^T \boldsymbol{Y}^k = \boldsymbol{Z}^k. \tag{4-6-104}$$

上式左乘以$(\boldsymbol{Z}^k)^T$，得

$$\lambda\alpha^2 = \frac{1}{(\boldsymbol{Z}^k)^T\boldsymbol{Y}^k} = \frac{1}{(\boldsymbol{Y}^k)^T\boldsymbol{Z}^k}. \tag{4-6-105}$$

于是

$$\boldsymbol{H}^{k+1} = \boldsymbol{H}^k + \frac{\boldsymbol{Z}^k(\boldsymbol{Z}^k)^T}{(\boldsymbol{Y}^k)^T\boldsymbol{Z}^k} \tag{4-6-106}$$

称式(4-6-106)为对称秩一($\boldsymbol{SR1}$)校正公式。

4.6.2.3 对称秩一算法

性质 4.19 *在$\boldsymbol{SR1}$算法中，由式(4-6-106)确定的$\boldsymbol{H}^{k+1}$满足拟牛*

算法 4.8 $\boldsymbol{SR}1$算法

步骤 1 给定目标函数$f(\boldsymbol{X})$，初始点$\boldsymbol{X}^0$，变尺度矩阵$\boldsymbol{H}^0$，梯度$\boldsymbol{g}^0$，搜索方向$\boldsymbol{P}^0$，精度ε，置$k=0$。

步骤 2 计算$\boldsymbol{X}^{k+1}=\boldsymbol{X}^k+t_k\boldsymbol{P}^k$，$\boldsymbol{g}^{k+1}$。

步骤 3 判断终止条件是否满足？若是，则停机并输出$\boldsymbol{X}^{k+1}$。

步骤 4 计算$\boldsymbol{S}^k$，$\boldsymbol{Y}^k$，$\boldsymbol{Z}^k$。

步骤 5 计算

$$\boldsymbol{H}^{k+1}=\boldsymbol{H}^k+\frac{\boldsymbol{Z}^k(\boldsymbol{Z}^k)^T}{(\boldsymbol{Y}^k)^T\boldsymbol{Z}^k}.$$

步骤 6 计算$\boldsymbol{P}^{k+1}=-\boldsymbol{H}^{k+1}\boldsymbol{g}^{k+1}$，$k=k+1$，转步骤2。

顿条件，即

$$\boldsymbol{H}^{k+1}\boldsymbol{Y}^k = \boldsymbol{S}^k.$$

性质 4.20 $\boldsymbol{SR}\mathbf{1}$算法用于$f(\boldsymbol{X}) = \frac{1}{2}\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{b}^T\boldsymbol{X} + c$，$\boldsymbol{Q} > 0$，如果

(1)$\boldsymbol{H}^0$为对称矩阵；

(2)迭代点互异；

(3)$(\boldsymbol{Z}^k)^T\boldsymbol{Y}^k \neq 0$.

那么，

$$(\boldsymbol{P}^i)^T\boldsymbol{Q}\boldsymbol{P}^j = 0; i, j = 0 \sim k+1, i \neq j \tag{4-6-107}$$

$$\boldsymbol{H}^{k+1}\boldsymbol{Q}\boldsymbol{P}^{j}=\boldsymbol{P}^{j}; j=0\sim k; \tag{4-6-108}$$

注：*SR*1算法是一种共轭方向法，当然具有二次收敛性。

证明 因为

$$\boldsymbol{Q}\boldsymbol{S}^{j}=\boldsymbol{Y}^{j}, \qquad j=0\sim k+1.$$

$$\boldsymbol{H}^{j+1}\boldsymbol{Y}^{j}=\boldsymbol{S}^{j}, \qquad j=0\sim k+1.$$

$$\boldsymbol{S}^{j}=-t_j\boldsymbol{H}^{j}\boldsymbol{g}^{j}=t_j\boldsymbol{P}^{j}, \qquad j=0\sim k+1.$$

所以

$$\boldsymbol{H}^{j+1}\boldsymbol{Q}\boldsymbol{P}^{j}=\boldsymbol{P}^{j},\quad j=0\sim k. \tag{4-6-109}$$

显然$k=0$时，有

$$\begin{aligned}\boldsymbol{H}^{1}\boldsymbol{Q}\boldsymbol{P}^{0}&=\boldsymbol{P}^{0},\\(\boldsymbol{P}^{1})^{T}\boldsymbol{Q}\boldsymbol{P}^{0}&=-(\boldsymbol{g}^{1})^{T}\boldsymbol{H}^{1}\boldsymbol{Q}\boldsymbol{P}^{0}\\&=-(\boldsymbol{g}^{1})^{T}\boldsymbol{P}^{0}\\&=0.\end{aligned}$$

假设$k=m-1$成立；

现在证明$k=m$也成立：

因为

$$
\begin{aligned}
(\boldsymbol{S}^m)^T \boldsymbol{Q} \boldsymbol{P}^j &= t_m (\boldsymbol{P}^m)^T \boldsymbol{Q} \boldsymbol{P}^j \\
&= 0, \qquad j = 0 \sim m-1;
\end{aligned} \tag{4-6-110}
$$

$$
\begin{aligned}
(\boldsymbol{Y}^m)^T \boldsymbol{H}^m \boldsymbol{Q} \boldsymbol{P}^j &= (\boldsymbol{Y}^m)^T \boldsymbol{P}^j \\
&= (\boldsymbol{S}^m)^T \boldsymbol{Q} \boldsymbol{P}^j \\
&= 0, \qquad j = 0 \sim m-1;
\end{aligned} \tag{4-6-111}
$$

所以

$$
\boldsymbol{H}^{m+1} \boldsymbol{Q} \boldsymbol{P}^j = \boldsymbol{H}^m \boldsymbol{Q} \boldsymbol{P}^j = \boldsymbol{P}^j, \quad j = 0 \sim m-1; \tag{4-6-112}
$$

又

$$\boldsymbol{H}^{m+1}\boldsymbol{Q}\boldsymbol{P}^{m}=\boldsymbol{P}^{m}. \tag{4-6-113}$$

最后

$$\begin{aligned}(\boldsymbol{P}^{m+1})^{T}\boldsymbol{Q}\boldsymbol{P}^{j}&=-(\boldsymbol{g}^{m+1})^{T}\boldsymbol{H}^{m+1}\boldsymbol{Q}\boldsymbol{P}^{j}\\&=-(\boldsymbol{g}^{m+1})^{T}\boldsymbol{P}^{j}=0.\quad j=0\sim m.\end{aligned} \tag{4-6-114}$$

∎

定理 4.21 对于正定二次函数，$\boldsymbol{SR1}$方法至多n步终止，即$\boldsymbol{H}^{n}=\boldsymbol{Q}^{-1}$。

证明 由性质4.20，不妨设$\boldsymbol{B} \triangleq (\boldsymbol{P}^0, \cdots, \boldsymbol{P}^{n-1})$，则有

$$\boldsymbol{H}^n \boldsymbol{Q} \boldsymbol{B} = \boldsymbol{H}^n \boldsymbol{Q}(\boldsymbol{P}^0, \cdots, \boldsymbol{P}^{n-1}) = (\boldsymbol{P}^0, \cdots, \boldsymbol{P}^{n-1}) = \boldsymbol{B}. \tag{4-6-115}$$

因为$\boldsymbol{P}^j$线性无关，$j = 0, \cdots, n-1$，所以$\boldsymbol{H}^n \boldsymbol{Q} = \boldsymbol{I}$，

$$\boldsymbol{H}^n = \boldsymbol{Q}^{-1}. \tag{4-6-116}$$

∎

注： 对称秩一校正公式具有:

(1)简单、对称迭代形式。

$$\boldsymbol{H}^{k+1}=\boldsymbol{H}^k+\frac{(\boldsymbol{S}^k-\boldsymbol{H}^k\boldsymbol{Y}^k)(\boldsymbol{S}^k-\boldsymbol{H}^k\boldsymbol{Y}^k)^T}{(\boldsymbol{Y}^k)^T(\boldsymbol{S}^k-\boldsymbol{H}^k\boldsymbol{Y}^k)}, \tag{4-6-117}$$

这里,

$$\boldsymbol{S}^k=\boldsymbol{X}^{k+1}-\boldsymbol{X}^k,\quad \boldsymbol{Y}^k=\boldsymbol{g}^{k+1}-\boldsymbol{g}^k. \tag{4-6-118}$$

(2)满足拟牛顿条件.

$$\boldsymbol{H}^{k+1}\boldsymbol{Y}^k=\boldsymbol{S}^k. \tag{4-6-119}$$

然而$\boldsymbol{SR1}$有两个明显的缺点:

(3)$\boldsymbol{SR}\mathbf{1}$公式可能出现

$$\boldsymbol{S}^k-\boldsymbol{H}^k\boldsymbol{Y}^k=0 \tag{4-6-120}$$

$$(\boldsymbol{Y}^k)^T(\boldsymbol{S}^k-\boldsymbol{H}^k\boldsymbol{Y}^k)=0 \tag{4-6-121}$$

(4)不能保证$\boldsymbol{P}^k=-\boldsymbol{H}^k\boldsymbol{g}^k$是下降方向，即$\boldsymbol{H}^k$不正定。

4.6.2.4 对称秩二公式

$\boldsymbol{SR}\mathbf{1}$公式的推导是假设变尺度矩阵修正项$\Delta\boldsymbol{H}^k$的为对称矩阵，且$rank(\Delta\boldsymbol{H}^k)=1$。

因此$\Delta \boldsymbol{H}^k$具有下面的简单形式：

$$\boldsymbol{Q}\begin{pmatrix} \lambda & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}\boldsymbol{Q}^T = \lambda \boldsymbol{u}\boldsymbol{u}^T, \tag{4-6-122}$$

这里，$\lambda(\neq 0) \in \boldsymbol{R}, \boldsymbol{u} \in \boldsymbol{R}^n$.

若让$\Delta \boldsymbol{H}^k$的形式稍复杂些，令$rank(\Delta \boldsymbol{H}^k) = 2$，那么会得到怎样的迭代公式呢？得到的公式是否不会具有$\boldsymbol{SR1}$的缺点。

这就是下面要讲的具有对称秩二($\boldsymbol{SR2}$)校正公式的$\boldsymbol{DFP}$算法。

DFP算法首先是由$\boldsymbol{D}avidon$(1959年)提出来的，后来，$\boldsymbol{F}letcher$和$\boldsymbol{P}owell$(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。

根据前面的分析，若$rank(\Delta \boldsymbol{H}^k) = 2$，则

$$\Delta \boldsymbol{H}^k = \boldsymbol{P} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \boldsymbol{P}^T = \lambda_1 \boldsymbol{u}_1 \boldsymbol{u}_1^T + \lambda_2 \boldsymbol{u}_2 \boldsymbol{u}_2^T. \quad (4\text{-}6\text{-}123)$$

这里，$\lambda_i(\neq 0) \in \boldsymbol{R}$，$\boldsymbol{u}_i \in \boldsymbol{R}^n$，$i = 1, 2$。

由拟牛顿条件，

$$\Delta \boldsymbol{H}^k \boldsymbol{Y}^k = \boldsymbol{S}^k - \boldsymbol{H}^k \boldsymbol{Y}^k.$$

即

$$\lambda_1 \boldsymbol{u}_1 \boldsymbol{u}_1^T \boldsymbol{Y}^k + \lambda_2 \boldsymbol{u}_2 \boldsymbol{u}_2^T \boldsymbol{Y}^k = \boldsymbol{S}^k - \boldsymbol{H}^k \boldsymbol{Y}^k. \tag{4-6-124}$$

与$\boldsymbol{SR1}$不同，上式(4-6-124)不能唯一确定$\boldsymbol{u}_1$，$\boldsymbol{u}_2$，因此，特殊有：

$$\lambda_1 \boldsymbol{u}_1 \boldsymbol{u}_1^T \boldsymbol{Y}^k = \boldsymbol{S}^k, \tag{4-6-125}$$

$$\lambda_2 \boldsymbol{u}_2 \boldsymbol{u}_2^T \boldsymbol{Y}^k = -\boldsymbol{H}^k \boldsymbol{Y}^k. \tag{4-6-126}$$

所以$\exists \alpha, \beta \in \boldsymbol{R}$，使得

$$\boldsymbol{u}_1 = \alpha \boldsymbol{S}^k, \tag{4-6-127}$$

$$\boldsymbol{u}_2 = -\beta \boldsymbol{H}^k \boldsymbol{Y}^k. \tag{4-6-128}$$

所以

$$\lambda_1 \alpha^2 \boldsymbol{S}^k (\boldsymbol{S}^k)^T \boldsymbol{Y}^k = \boldsymbol{S}^k, \tag{4-6-129}$$

$$\lambda_2 \beta^2 \boldsymbol{H}^k \boldsymbol{Y}^k (\boldsymbol{H}^k \boldsymbol{Y}^k)^T \boldsymbol{Y}^k = -\boldsymbol{H}^k \boldsymbol{Y}^k. \tag{4-6-130}$$

因而

$$\lambda_1\alpha^2 = \frac{1}{(\boldsymbol{S}^k)^T\boldsymbol{Y}^k} \tag{4-6-131}$$

$$\lambda_2\beta^2 = -\frac{1}{(\boldsymbol{H}^k\boldsymbol{Y}^k)^T\boldsymbol{Y}^k} \tag{4-6-132}$$

即

$$\lambda_1\boldsymbol{u}_1\boldsymbol{u}_1^T = \lambda_1\alpha^2\boldsymbol{S}^k(\boldsymbol{S}^k)^T = \frac{\boldsymbol{S}^k(\boldsymbol{S}^k)^T}{(\boldsymbol{S}^k)^T\boldsymbol{Y}^k} \tag{4-6-133}$$

$$\begin{aligned}\lambda_2\boldsymbol{u}_2\boldsymbol{u}_2^T &= \lambda_2\beta^2\boldsymbol{H}^k\boldsymbol{Y}^k(\boldsymbol{H}^k\boldsymbol{Y}^k)^T \\ &= -\frac{\boldsymbol{H}^k\boldsymbol{Y}^k(\boldsymbol{H}^k\boldsymbol{Y}^k)^T}{(\boldsymbol{H}^k\boldsymbol{Y}^k)^T\boldsymbol{Y}^k}\end{aligned} \tag{4-6-134}$$

从而得到$\boldsymbol{DFP}$算法的对称秩2($\boldsymbol{SR2}$)校正公式：

$$\boldsymbol{H}^{k+1}=\boldsymbol{H}^{k}+\frac{\boldsymbol{S}^{k}(\boldsymbol{S}^{k})^{T}}{(\boldsymbol{S}^{k})^{T}\boldsymbol{Y}^{k}}-\frac{\boldsymbol{H}^{k}\boldsymbol{Y}^{k}(\boldsymbol{H}^{k}\boldsymbol{Y}^{k})^{T}}{(\boldsymbol{H}^{k}\boldsymbol{Y}^{k})^{T}\boldsymbol{Y}^{k}} \tag{4-6-135}$$

定理 4.22 (正定遗传性) 若$\boldsymbol{H}^{0}$是对称正定矩阵，则$\boldsymbol{SR2}$校正公式(4-6-135)所产生的$\boldsymbol{H}^{k}$是对称正定矩阵。

证明 对称性显然，下只证正定性。

$k=1$，成立；假设$\boldsymbol{H}^{k}$成立；下证$\boldsymbol{H}^{k+1}$成立：

$\forall\ \boldsymbol{X}(\neq 0) \in \boldsymbol{R}^n$，有

$$\begin{aligned}
\boldsymbol{X}^T\boldsymbol{H}^{k+1}\boldsymbol{X} &= \boldsymbol{X}^T\boldsymbol{H}^k\boldsymbol{X} + \frac{\boldsymbol{X}^T\boldsymbol{S}^k(\boldsymbol{S}^k)^T\boldsymbol{X}}{(\boldsymbol{S}^k)^T\boldsymbol{Y}^k} - \frac{\boldsymbol{X}^T\boldsymbol{H}^k\boldsymbol{Y}^k(\boldsymbol{H}^k\boldsymbol{Y}^k)^T\boldsymbol{X}}{(\boldsymbol{H}^k\boldsymbol{Y}^k)^T\boldsymbol{Y}^k} \\
&= \frac{\boldsymbol{X}^T\boldsymbol{H}^k\boldsymbol{X}(\boldsymbol{Y}^k)^T\boldsymbol{H}^k\boldsymbol{Y}^k - \boldsymbol{X}^T\boldsymbol{H}^k\boldsymbol{Y}^k(\boldsymbol{Y}^k)^T\boldsymbol{H}^k\boldsymbol{X}}{(\boldsymbol{Y}^k)^T\boldsymbol{H}^k\boldsymbol{Y}^k} \\
&\qquad + \frac{\boldsymbol{X}^T\boldsymbol{S}^k(\boldsymbol{S}^k)^T\boldsymbol{X}}{(\boldsymbol{S}^k)^T\boldsymbol{Y}^k} \\
&= \frac{\boldsymbol{X}^T\boldsymbol{H}^k\boldsymbol{X}(\boldsymbol{Y}^k)^T\boldsymbol{H}^k\boldsymbol{Y}^k - (\boldsymbol{X}^T\boldsymbol{H}^k\boldsymbol{Y}^k)^2}{(\boldsymbol{Y}^k)^T\boldsymbol{H}^k\boldsymbol{Y}^k} \\
&\qquad + \frac{(\boldsymbol{X}^T\boldsymbol{S}^k)^2}{(\boldsymbol{S}^k)^T\boldsymbol{Y}^k}
\end{aligned} \tag{4-6-136}$$

又因为

$$\begin{aligned}(\boldsymbol{S}^k)^T\boldsymbol{Y}^k &= t_k(\boldsymbol{P}^k)^T(\boldsymbol{g}^{k+1}-\boldsymbol{g}^k)\\ &= -t_k(\boldsymbol{P}^k)^T\boldsymbol{g}^k = t_k(\boldsymbol{g}^k)^T\boldsymbol{H}^k\boldsymbol{g}^k > 0, (\boldsymbol{H}^k > 0)\end{aligned} \tag{4-6-137}$$

$$\boldsymbol{X}^T\boldsymbol{H}^k\boldsymbol{X}(\boldsymbol{Y}^k)^T\boldsymbol{H}^k\boldsymbol{Y}^k-(\boldsymbol{X}^T\boldsymbol{H}^k\boldsymbol{Y}^k)^2 \geq 0, (Cauchy-Schwarz) \tag{4-6-138}$$

$$\begin{aligned}(\boldsymbol{X}^T\boldsymbol{S}^k)^T &= (\boldsymbol{S}^k)^T\boldsymbol{X} = \alpha(\boldsymbol{S}^k)^T\boldsymbol{Y}^k\\ &= \alpha t_k(\boldsymbol{g}^k)^T\boldsymbol{H}^k\boldsymbol{g}^k \neq 0, (\boldsymbol{X} = \alpha\boldsymbol{Y}^k, \alpha \neq 0)\end{aligned} \tag{4-6-139}$$

证毕! ∎

4.6.2.5 *DFP*算法

算法 4.9 ***DFP*算法**

替换***SR*1**算法4.8中***SR*1**校正公式(4-6-106)为***SR*2**校正公式(4-6-135)。

注：*DFP*算法具有类似于*SR*1算法的性质，本质上也是一种共轭方向法，具有二次收敛性。特别地，当$\boldsymbol{H}^0 = \boldsymbol{I}$时，*DFP*算法本质上是共轭梯度法。

例 4.5 用***DFP***算法求$\min\ 2x_1^2 + x_2^2 - 4x_1 + 2$，取$\boldsymbol{X}^0 = (2,1)^T$。

解 $\boldsymbol{g} = \begin{pmatrix} 4(x_1-1) \\ 2x_2 \end{pmatrix}$，$\boldsymbol{Q} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$，$\boldsymbol{H}^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$，

$\boldsymbol{g}^0 = (4,2)^T$，$\|\boldsymbol{g}^0\| \neq 0$；

第一次迭代:

$\boldsymbol{P}^0 = -\boldsymbol{H}^0\boldsymbol{g}^0 = -\boldsymbol{g}^0$,

$\boldsymbol{X}^1 = \boldsymbol{X}^0 + t_0\boldsymbol{P}^0 = \boldsymbol{X}^0 + \frac{-\boldsymbol{g}^0\boldsymbol{P}^0}{(\boldsymbol{P}^0)^T\boldsymbol{Q}\boldsymbol{P}^0}\boldsymbol{P}^0 = \left(\frac{8}{9}, \frac{4}{9}\right)^T$,

$\boldsymbol{g}^1 = \left(-\frac{4}{9}, \frac{8}{9}\right)^T$, $\|\boldsymbol{g}^1\| \neq 0$;

第二次迭代:

$\boldsymbol{S}^0 = \boldsymbol{X}^1 - \boldsymbol{X}^0 = \left(-\frac{10}{9}, -\frac{5}{9}\right)^T$,

$\boldsymbol{Y}^0 = \boldsymbol{g}^1 - \boldsymbol{g}^0 = \left(-\frac{40}{9}, -\frac{10}{9}\right)^T$,

所以

$$\boldsymbol{H}^1 = \boldsymbol{H}^0 + \frac{\boldsymbol{S}^0(\boldsymbol{S}^0)^T}{(\boldsymbol{S}^0)^T\boldsymbol{Y}^0} - \frac{\boldsymbol{H}^0\boldsymbol{Y}^0(\boldsymbol{H}^0\boldsymbol{Y}^0)^T}{(\boldsymbol{H}^0\boldsymbol{Y}^0)^T\boldsymbol{Y}^0}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{18}\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} - \frac{1}{17}\begin{pmatrix} 16 & 4 \\ 4 & 1 \end{pmatrix}$$
$$= \frac{1}{306}\begin{pmatrix} 86 & -38 \\ -38 & 305 \end{pmatrix}$$

搜索方向为$\boldsymbol{P}^1 = -\boldsymbol{H}^1\boldsymbol{g}^1 = \frac{4}{17}(1,-4)^T$,

从$\boldsymbol{X}^1$出发沿$\boldsymbol{P}^1$进行一维搜索，即

$$\boldsymbol{X}^2 = \boldsymbol{X}^1 + t_1\boldsymbol{P}^1 = (1,0)^T,$$

$\|\boldsymbol{g}^2\| = 0$，所以$\boldsymbol{X}^2$是极小点。 ∎

思考：

$$(1)\boldsymbol{H}^1=\frac{1}{306}\begin{pmatrix}86 & -38\\ -38 & 305\end{pmatrix}\neq \boldsymbol{Q}^{-1}=\begin{pmatrix}1/4 & 0\\ 0 & 1/2\end{pmatrix}?$$

(2)第一次迭代执行最速下降法，那么第二次(最后一次)迭代是不是执行牛顿法？

练习 4.6 计算$\boldsymbol{P}^1_{\boldsymbol{Newton}}=-\boldsymbol{Q}^{-1}\boldsymbol{g}^1$，$\boldsymbol{H}^2$。

解

$$\begin{aligned}P^1_{\boldsymbol{Newton}}&=-\boldsymbol{Q}^{-1}\boldsymbol{g}^1\\ &=-\begin{pmatrix}1/4 & 0\\ 0 & 1/2\end{pmatrix}\begin{pmatrix}-4/9\\ 8/9\end{pmatrix}\end{aligned}$$

$$= \frac{4}{9}\begin{pmatrix} 1 \\ -4 \end{pmatrix}.$$

$$\begin{aligned} \boldsymbol{H}^2 &= \boldsymbol{H}^1 + \frac{\boldsymbol{S}^1(\boldsymbol{S}^1)^T}{(\boldsymbol{S}^1)^T\boldsymbol{Y}^1} - \frac{\boldsymbol{H}^1\boldsymbol{Y}^1(\boldsymbol{H}^1\boldsymbol{Y}^1)^T}{(\boldsymbol{H}^1\boldsymbol{Y}^1)^T\boldsymbol{Y}^1} \\ &= \frac{1}{306}\begin{pmatrix} 86 & -38 \\ -38 & 305 \end{pmatrix} + \frac{1}{9}\begin{pmatrix} 1/4 & -1 \\ -1 & 4 \end{pmatrix} - \frac{1}{17}\begin{pmatrix} 1 & -4 \\ -4 & 16 \end{pmatrix} \\ &= \begin{pmatrix} 1/4 & 0 \\ 0 & 1/2 \end{pmatrix} \end{aligned}$$

∎

练习 4.7 证明:

$$\boldsymbol{H}^n\boldsymbol{g}^{n-1} = t_{n-1}\boldsymbol{H}^{n-1}\boldsymbol{g}^{n-1}. \tag{4-6-140}$$

拟牛顿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(4-6-124)的$\boldsymbol{u}_1, \boldsymbol{u}_2$有无穷多个，因此上述拟牛顿法就构成一族算法。常见的有$\boldsymbol{Broyden}$族和$\boldsymbol{Huang}$族两大类。

4.6.2.6 *Broyden*和*Huang*类校正公式

(1)1967年$\boldsymbol{Broyden}$给出：

***Broyden*类校正公式：**

$$\begin{aligned}\boldsymbol{H}^{k+1} = \boldsymbol{H}^{k} + \frac{\boldsymbol{S}^{k}(\boldsymbol{S}^{k})^{T}}{(\boldsymbol{Y}^{k})^{T}\boldsymbol{S}^{k}} - \frac{\boldsymbol{H}^{k}\boldsymbol{Y}^{k}(\boldsymbol{Y}^{k})^{T}\boldsymbol{H}^{k}}{(\boldsymbol{Y}^{k})^{T}\boldsymbol{H}^{k}\boldsymbol{Y}^{k}} \\ + \beta(\boldsymbol{Y}^{k})^{T}\boldsymbol{S}^{k}(\boldsymbol{Y}^{k})^{T}\boldsymbol{H}^{k}\boldsymbol{Y}^{k}\omega^{k}(\omega^{k})^{T}.\end{aligned} \tag{4-6-141}$$

这里，

$$\omega^{k} = \frac{\boldsymbol{S}^{k}}{(\boldsymbol{Y}^{k})^{T}\boldsymbol{S}^{k}} - \frac{\boldsymbol{H}^{k}\boldsymbol{Y}^{k}}{(\boldsymbol{Y}^{k})^{T}\boldsymbol{H}^{k}\boldsymbol{Y}^{k}}. \tag{4-6-142}$$

- 当$\beta = \frac{1}{(\boldsymbol{Y}^{k})^{T}(\boldsymbol{S}^{k}-\boldsymbol{H}^{k}\boldsymbol{Y}^{k})}$时，$\Rightarrow$ ***SR*1**。
- 当$\beta = 0$时，$\Rightarrow$ ***SR*2**。

(2)1970年***Huang*给出：**

***Huang*类校正公式：**

$$\boldsymbol{H}^{k+1} = \boldsymbol{H}^k + \boldsymbol{S}^k(\boldsymbol{u}^k)^T + \boldsymbol{H}^k\boldsymbol{Y}^k(\boldsymbol{v}^k)^T. \tag{4-6-143}$$

在这里，

$$\boldsymbol{u}^k = a_k\boldsymbol{S}^k + b_k(\boldsymbol{H}^k)^T\boldsymbol{Y}^k, \tag{4-6-144}$$

$$\boldsymbol{v}^k = c_k\boldsymbol{S}^k + d_k(\boldsymbol{H}^k)^T\boldsymbol{Y}^k. \tag{4-6-145}$$

且满足

$$(\boldsymbol{u}^k)^T\boldsymbol{Y}^k = \omega, \tag{4-6-146}$$

$$(\boldsymbol{v}^k)^T\boldsymbol{Y}^k = -1 \tag{4-6-147}$$

- 当$\omega \neq 1$，$\boldsymbol{H}uang$族校正公式不满足拟牛顿方程，因此不是拟牛顿算法。

$$\boldsymbol{H}^{k+1} = \boldsymbol{H}^k + \beta_k \frac{\boldsymbol{S}^k(\boldsymbol{S}^k)^T - \boldsymbol{H}^k\boldsymbol{Y}^k(\boldsymbol{S}^k)^T - \boldsymbol{S}^k(\boldsymbol{Y}^k)^T\boldsymbol{H}^k}{(\boldsymbol{S}^k)^T\boldsymbol{Y}^k},$$
$$\beta_k = 1 + \frac{(\boldsymbol{Y}^k)^T\boldsymbol{H}^k\boldsymbol{Y}^k}{(\boldsymbol{S}^k)^T\boldsymbol{Y}^k}. \tag{4-6-148}$$

- 当$\omega = 1$，$b_k = c_k$时，$\Rightarrow$ ***Broyden***。
- 当$\omega = 1$，$b_k = c_k = 0$时，$\Rightarrow$ ***DFP***。

- 当$\omega=1$，$b_k=c_k$，$d_k=0$时，$\Rightarrow$ $\boldsymbol{BFGS}$，即

$$\boldsymbol{H}^{k+1}=\boldsymbol{H}^{k}+\beta_k\frac{\boldsymbol{S}^{k}(\boldsymbol{S}^{k})^{T}-\boldsymbol{H}^{k}\boldsymbol{Y}^{k}(\boldsymbol{S}^{k})^{T}-\boldsymbol{S}^{k}(\boldsymbol{Y}^{k})^{T}\boldsymbol{H}^{k}}{(\boldsymbol{S}^{k})^{T}\boldsymbol{Y}^{k}},$$

$$\beta_k=1+\frac{(\boldsymbol{Y}^{k})^{T}\boldsymbol{H}^{k}\boldsymbol{Y}^{k}}{(\boldsymbol{S}^{k})^{T}\boldsymbol{Y}^{k}}. \tag{4-6-149}$$

$\boldsymbol{BFGS}$由$\boldsymbol{B}roy den$，$\boldsymbol{F}letcher$，$\boldsymbol{G}oldfarb$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。**$\boldsymbol{BFGS}$比$\boldsymbol{DFP}$具有更好的数值稳定性。** 为保证$\boldsymbol{H}^{k}$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$\boldsymbol{X}^0=\boldsymbol{X}^n$，$\boldsymbol{H}^0=\boldsymbol{I}$。

(3) $\boldsymbol{Broyden}$族和$\boldsymbol{Huang}$族关系为：

$$\boldsymbol{Huang}\succ\left|\begin{array}{c}\boldsymbol{DFP}\\ \boldsymbol{Broyden}\succ\left|\begin{array}{c}\boldsymbol{SR1}\\ \boldsymbol{SR2}(\boldsymbol{DFP})\end{array}\right.\\ \boldsymbol{BFGS}\end{array}\right.$$

§ 4.7 信赖域方法

牛顿法的基本思想是在迭代点$\boldsymbol{X}^k$附近用二次函数

$$q^k(\boldsymbol{S})=f(\boldsymbol{X}^k)+(\boldsymbol{g}^k)^T\boldsymbol{S}+\frac{1}{2}\boldsymbol{S}^T\boldsymbol{G}^k\boldsymbol{S} \tag{4-7-150}$$

逼近$f(\boldsymbol{X})$，并以$q^k(\boldsymbol{S})$的极小点修正$\boldsymbol{X}^k$，得到

$$\boldsymbol{X}^{k+1} = \boldsymbol{X}^k + \boldsymbol{S}^k. \tag{4-7-151}$$

这里，$\boldsymbol{g}^k \triangleq \nabla f(\boldsymbol{X}^k)$，$\boldsymbol{G}^k \triangleq \nabla^2 f(\boldsymbol{X}^k)$。

但是，这种方法只能保证算法的局部收敛性，即只有当$\boldsymbol{S}$充分小时，$q^k(\boldsymbol{S})$才能逼近$f(\boldsymbol{X})$。

4.7.1 基本思想

信赖域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，并由此定义$\boldsymbol{X}^k$的

邻域Ω^k，

$$\Omega^k = \{\boldsymbol{X} \mid \|\boldsymbol{X} - \boldsymbol{X}^k\| \leq h^k\}. \tag{4-7-152}$$

假定在这个邻域中$q^k(\boldsymbol{S})$与目标函数$f(\boldsymbol{X}^k+\boldsymbol{S})$一致，即二次模型是目标函数$f(\boldsymbol{X})$的一个合适的模拟，然后用$n$维二次模型$q^k(\boldsymbol{S})$确定搜索方向$\boldsymbol{S}^k$，并取$\boldsymbol{X}^{k+1}=\boldsymbol{X}^k+\boldsymbol{S}^k$。

信赖域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。

不仅可以用来代替一维搜索，而且也可以解决$\boldsymbol{Hesse}$矩阵$\boldsymbol{G}^k$不正定和$\boldsymbol{X}^k$为鞍点等困难。

信赖域方法的模型问题是

$$\min \ q^k(\boldsymbol{S}) = f(\boldsymbol{X}^k) + (\boldsymbol{g}^k)^T\boldsymbol{S} + \frac{1}{2}\boldsymbol{S}^T\boldsymbol{G}^k\boldsymbol{S}$$

$$s.t. \qquad (4\text{-}7\text{-}153)$$

$$\|\boldsymbol{S}\| \le h^k.$$

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$f(\boldsymbol{X})$在第k步的实际下降量:

$$\Delta f^k = f(\boldsymbol{X}^k) - f(\boldsymbol{X}^k + \boldsymbol{S}^k), \tag{4-7-154}$$

Δq^k为对应的预测下降量:

$$\Delta q^k = f(\boldsymbol{X}^k) - q^k(\boldsymbol{S}^k). \tag{4-7-155}$$

定义比值:

$$r^k = \Delta f^k / \Delta q^k, \tag{4-7-156}$$

它衡量二次模型$q^k(\boldsymbol{S}^k)$近似目标函数$f(\boldsymbol{X}^k + \boldsymbol{S}^k)$的程度。$r^k$越接近1，表明近似程度越好。

信赖域方法自适应地改变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。

注意：

(1)算法4.10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)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$(0.1\|\boldsymbol{S}^k\|, 0.5\|\boldsymbol{S}^k\|)$中由多项式插值选取$h^{k+1}$。

(3)一般地，h^k变大，算法4.10性能趋向于牛顿方法；h^k变小，性能趋向于最速下降方法。

例 4.8 求解$\min\ f(\boldsymbol{X}) = x_1^4 + x_1^2 + x_2^2 - 4x_2$，令$\boldsymbol{X}^0 =$

算法 4.10 信赖域算法

步骤 1　给出$\boldsymbol{X}^0$，h^0，令$k=0$。

步骤 2　给出$\boldsymbol{X}^k$和h^k，计算$\boldsymbol{g}^k$和$\boldsymbol{G}^k$。

步骤 3　解信赖域模型(4-7-153)，求出$\boldsymbol{S}^k$。

步骤 4　求$f(\boldsymbol{X}^k+\boldsymbol{S}^k)$和$r^k$的值。

步骤 5

if $r^k<0.25$ **then**

　$h^{k+1}=\|\boldsymbol{S}^k\|/4;$

else if $r^k>0.75$ 和$\|\boldsymbol{S}^k\|=h^k$ **then**

　$h^{k+1}=2\cdot h^k;$

else

　$h^{k+1}=h^k$。

end if

步骤 6　若$r^k\le 0$，则$\boldsymbol{X}^{k+1}=\boldsymbol{X}^k$；否则，$\boldsymbol{X}^{k+1}=\boldsymbol{X}^k+\boldsymbol{S}^k$。

$(0,1)^T$，$h^0 = 1/2$。

解 $g(\boldsymbol{X}) = \nabla f(\boldsymbol{X}) = (4x_1^3 + 2x_1, 2x_2 - 4)^T$,

$$G(\boldsymbol{X}) = \nabla^2 f(\boldsymbol{X}) = \begin{pmatrix} 12x_1 + 2, & 0 \\ 0, & 2 \end{pmatrix}$$

$$\boldsymbol{g}^0 = (0, -2)^T,\ \|\boldsymbol{g}^0\| \neq 0,\ \boldsymbol{G}^0 = \begin{pmatrix} 2, & 0 \\ 0, & 2 \end{pmatrix},$$

建立信赖域模型：

$$\min\ q^0(\boldsymbol{S}) = f(\boldsymbol{X}^0) + (\boldsymbol{g}^0)^T \boldsymbol{S} + \frac{1}{2}\boldsymbol{S}^T \boldsymbol{G}^0 \boldsymbol{S}$$

$s.t.$

$$\|\boldsymbol{S}\|_2 \leq h_0.$$

即

$$\min \ q^0(\boldsymbol{S}) = -3 - 2s_2 + s_1^2 + s_2^2$$

$s.t.$

$$s_1^2 + s_2^2 \leq \frac{1}{4}.$$

解之, $\boldsymbol{S}^0 = (0, \frac{1}{2})$,

$\boldsymbol{X}^1 = \boldsymbol{X}^0 + \boldsymbol{S}^0 = (0,1)^T + (0, \frac{1}{2})^T = (0, \frac{3}{2})^T$,

$\boldsymbol{g}^1 = (0, -1)^T, \|\boldsymbol{g}^1\| \neq 0$,

$r^0 = \frac{\Delta f^0}{\Delta q^0} = \frac{-3+15/4}{-3+15/4} = 1 > 0.75$,

$h_1 = 2h_0 = 1$,

建立信赖域模型：

$$\min\ q^1(\boldsymbol{S}) = -\frac{15}{4} - s_2 + s_1^2 + s_2^2$$

$s.t.$

$$s_1^2 + s_2^2 \leq 1.$$

解之, $\boldsymbol{S}^1 = (0, \frac{1}{2})$,

$\boldsymbol{X}^2 = \boldsymbol{X}^1 + \boldsymbol{S}^1 = (0, \frac{3}{2})^T + (0, \frac{1}{2})^T = (0, 2)^T$,

$\boldsymbol{g}^2 = (0, 0)^T$,$\|\boldsymbol{g}^2\| = 0$,

所以$\boldsymbol{X}^\star = \boldsymbol{X}^2 = (0, 2)^T$. ∎

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。

定理 4.23 (总体收敛性定理) 设$\boldsymbol{D} \in \boldsymbol{R}^n$是有界集，对$\forall k$，有$\boldsymbol{X}^k \in \boldsymbol{D}$，$\|\boldsymbol{G}^k\|_2 \leq M$，$M > 0$，$f \in C^2$，则信赖域算法产生一个满足一阶和二阶必要条件的聚点$\boldsymbol{X}^\star$.

在较强的假设下，可以得出算法具有二阶收敛速度.

第五章　约束最优化方法

约束最优化问题为：

$$\min \ f(\boldsymbol{X}) \tag{NLP}$$

$s.t.$

$$g_i(\boldsymbol{X}) \geq 0, \quad i = 1 \sim m; \tag{5-0-1}$$

$$h_j(\boldsymbol{X}) = 0, \quad j = 1 \sim l. \tag{5-0-2}$$

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

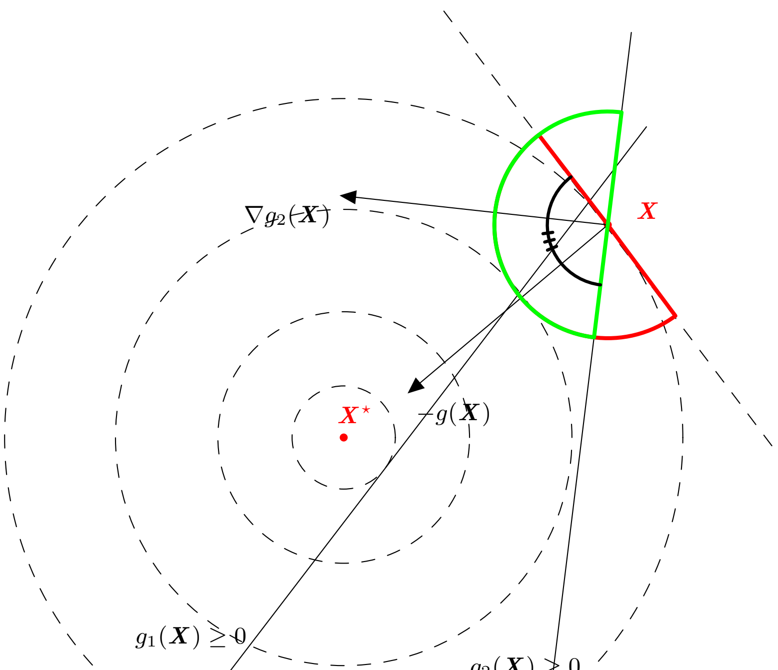
5.1.1 可行方向

定义 5.1 (可行域) $\boldsymbol{S} \triangleq \{\boldsymbol{X} | g_i(\boldsymbol{X}) \geq 0, h_j(\boldsymbol{X}) = 0, i = 1 \sim m, j = 1 \sim l\}$。

定义 5.2 (可行方向) 设$\boldsymbol{X} \in \boldsymbol{S}$，若对某一个非零向量$\boldsymbol{P} \in \boldsymbol{R}^n$，存在一个$\delta > 0$，使得对所有$\lambda \in (0, \delta)$时，有$\boldsymbol{X} + \lambda \boldsymbol{P} \in \boldsymbol{S}$，则称$\boldsymbol{P}$为点$\boldsymbol{X}$处的一个可行方向。

定义 5.3 (下降方向) 设$\boldsymbol{X} \in \boldsymbol{S}$，存在$\delta > 0$，当$\lambda \in (0, \delta)$时，若存在非零向量$\boldsymbol{P}$使得$f(\boldsymbol{X} + \lambda\boldsymbol{P}) < f(\boldsymbol{X})$成立，则称$\boldsymbol{P}$为$f(\boldsymbol{X})$在点$\boldsymbol{X}$处的一个下降方向。

定义 5.4 (可行下降方向) 设$\boldsymbol{X} \in \boldsymbol{S}$，非零向量$\boldsymbol{P}$既是点$\boldsymbol{X}$处的可行方向，又是$f(\boldsymbol{X})$在点$\boldsymbol{X}$处的一个下降方向。


$\nabla g_2(\boldsymbol{X})$
$\boldsymbol{X}$
$\boldsymbol{X}^\star$
$-g(\boldsymbol{X})$
$g_1(\boldsymbol{X}) \geq 0$
$g_2(\boldsymbol{X}) \geq 0$

定义 5.5 设$\overline{\boldsymbol{X}} \in \boldsymbol{S}$，若$\exists k \in \{1, \cdots, m\}$，使得$g_k(\overline{\boldsymbol{X}}) = 0$，则称不等式$g_k(\boldsymbol{X}) \geq 0$为关于$\overline{\boldsymbol{X}}$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$\mathcal{I}(\overline{\boldsymbol{X}}) \triangleq \{i \mid g_i(\overline{\boldsymbol{X}}) = 0, i = 1 \sim m\}$，$\mathcal{J} \triangleq \{1, 2, \cdots, l\}$，则称$\mathcal{A} = \mathcal{I}(\overline{\boldsymbol{X}}) \bigcup \mathcal{J}$为关于$\overline{\boldsymbol{X}}$的起作用(积极、有效)约束集。

定理 5.1 设$\overline{\boldsymbol{X}} \in \boldsymbol{S}$，若$\boldsymbol{P}$是$\overline{\boldsymbol{X}}$的一个可行下降方向，则有

$$\nabla f(\overline{\boldsymbol{X}})^T \boldsymbol{P} < 0; \tag{5-1-3}$$

$$\nabla g_k(\overline{\boldsymbol{X}})^T \boldsymbol{P} > 0; \qquad k \in \mathcal{A}. \tag{5-1-4}$$

5.1.2 一阶必要条件

定义 5.6 (正则点,正则解) 设$\overline{\boldsymbol{X}} \in \boldsymbol{S}$，若$\nabla g_i(\overline{\boldsymbol{X}})$，$\nabla h_j(\overline{\boldsymbol{X}})$，$i \in \mathcal{I}(\overline{\boldsymbol{X}})$，$j \in \mathcal{J}$线性无关，则称$\overline{\boldsymbol{X}}$为$\boldsymbol{S}$的一个正则点。

定理 5.2 (一阶必要条件) 设$f(\boldsymbol{X})$，$g_i(\boldsymbol{X})$，$h_j(\boldsymbol{X})$可 微 ，若$\boldsymbol{X}^\star$是(NLP)的 一 个 局 部 极 小 点 且 为 正 则 点 ， 则 存 在$\lambda_i^\star(i \in \mathcal{I}(\boldsymbol{X}^\star))$以及$\mu_j^\star(j \in \mathcal{J})$，使得

$$\nabla f(\boldsymbol{X}^\star) = \sum_{i \in \mathcal{I}(\boldsymbol{X}^\star)} \lambda_i^\star \nabla g_i(\boldsymbol{X}^\star) + \sum_{j=1}^{l} \mu_j^\star \nabla h_j(\boldsymbol{X}^\star), \tag{5-1-5}$$

$$\lambda_i^\star \geq 0, \qquad i = 1, 2, \cdots, m; \tag{5-1-6}$$

$$\lambda_i^\star g_i(\boldsymbol{X}^\star) = 0, \qquad i = 1, 2, \cdots, m; \tag{5-1-7}$$

$$g_i(\boldsymbol{X}^\star) \geq 0, \qquad i = 1, 2, \cdots, m; \tag{5-1-8}$$

$$h_j(\boldsymbol{X}^\star) = 0, \qquad j = 1, 2, \cdots, l. \tag{5-1-9}$$

成立。

定义 5.7 (*KKT*点) 称(5-1-5)-(5-1-9)为***KKT***条件；称满足***KKT***条件的点为***KKT***点，或***K-T***点。

例 5.1

$$\min\ f(\boldsymbol{X}) = (x_1 - 2)^2 + (x_2 - 1)^2$$

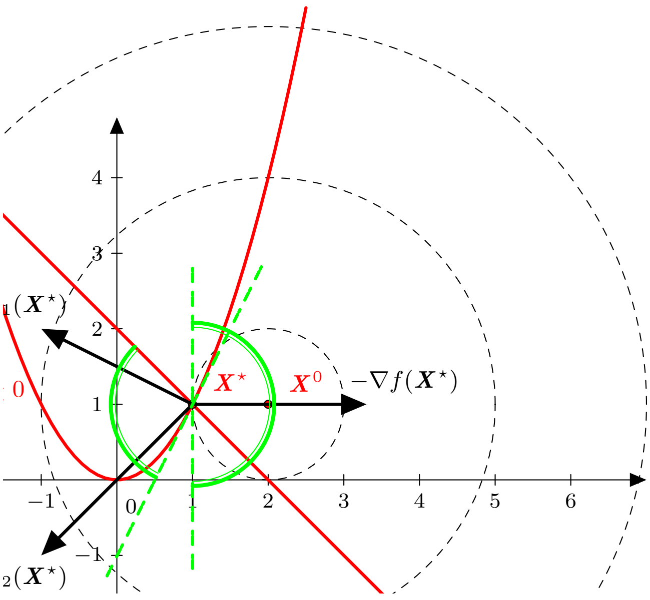
$$s.t.$$

$$g_1(\boldsymbol{X}) = x_2 - x_1^2 \geq 0$$

$$g_2(\boldsymbol{X}) = 2 - x_1 - x_2 \geq 0$$

解

$\blacksquare$


$_1(\boldsymbol{X}^\star)$
$\boldsymbol{X}^\star$
$\boldsymbol{X}^0$
$-\nabla f(\boldsymbol{X}^\star)$
$_2(\boldsymbol{X}^\star)$
0

例 5.2 试验证$\boldsymbol{X}^1 = (1,1,1)^T$,$\boldsymbol{X}^2 = (-\sqrt{3},0,0)^T$是下面问题的$\boldsymbol{KKT}$点。

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

$$s.t.$$

$$-x_1 + x_2 \geq 0$$

$$x_1^2 + x_2^2 + x_3^2 = 3.$$

解 存在$\lambda^\star = 2$，$\mu^\star = -2$，使得

$$\nabla f(\boldsymbol{X}^\star) = \lambda^\star \nabla g_1(\boldsymbol{X}^\star) + \mu^\star \nabla h_1(\boldsymbol{X}^\star),$$

$$\lambda^\star \geq 0,$$

$$\lambda^\star g_1(\boldsymbol{X}^\star) = 0.$$

∎

5.1.3 二阶充分条件

定理 5.3 (二阶充分条件) 设$f(\boldsymbol{X})$，$g_i(\boldsymbol{X})$，$h_j(\boldsymbol{X})$二次连续可微，若$\boldsymbol{X}^\star \in \boldsymbol{S}$，存在$\lambda_i^\star$，$\mu_j^\star$，$i = 1 \sim m$，$j = 1 \sim l$，使

得$\boldsymbol{KKT}$条件成立，且对$\forall\ \boldsymbol{Z}(\neq 0)\in\mathcal{F}$有

$$\boldsymbol{Z}^T\left(\nabla^2 f(\boldsymbol{X}^\star)-\sum_{i=1}^{m}\lambda_i^\star\nabla^2 g_i(\boldsymbol{X}^\star)-\sum_{j=1}^{l}\mu_j^\star\nabla^2 h_j(\boldsymbol{X}^\star)\right)\boldsymbol{Z}>0 \tag{5-1-10}$$

成立，则$\boldsymbol{X}^\star$为(NLP)的严格局部最小点。

这里，

$$\mathcal{F}\triangleq\left\{\boldsymbol{Z}\in R^n\;\middle|\;\begin{array}{ll}\boldsymbol{Z}^T\nabla g_i(\boldsymbol{X}^\star)=0, & i\in\mathcal{I}(\boldsymbol{X}^\star),\lambda_i^\star>0;\\ \boldsymbol{Z}^T\nabla g_i(\boldsymbol{X}^\star)\geq 0, & i\in\mathcal{I}(\boldsymbol{X}^\star),\lambda_i^\star=0;\\ \boldsymbol{Z}^T\nabla h_j(\boldsymbol{X}^\star)=0, & j=1,2,\cdots,l.\end{array}\right\} \tag{5-1-11}$$

例 5.3

$$\min\ \ f(\boldsymbol{X})=x_1^2-8x_1+x_2+16$$

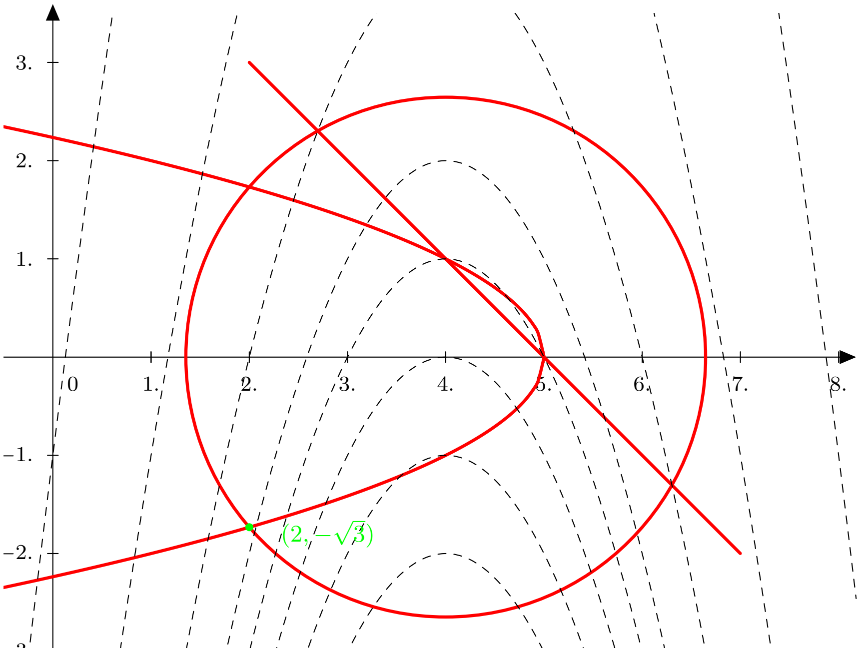
$$s.t.$$

$$g_1(\boldsymbol{X}) = -(x_1 + x_2^2) + 5 \geq 0$$

$$g_2(\boldsymbol{X}) = -(x_1 + x_2) + 5 \geq 0$$

$$h_1(\boldsymbol{X}) = (x_1 - 4)^2 + x_2^2 - 7 = 0$$

解 $\boldsymbol{X}^\star = (2, -\sqrt{3})^T$。


3.
2.
1.
−1.
−2.
0
1.
2.
3.
4.
5.
6.
7.
8.
$(2, -\sqrt{3})$

易得，$\lambda^\star = \frac{36+2\sqrt{3}}{15} > 0$，$\mu^\star = \frac{24-\sqrt{3}}{30}$。

所以，

$$\begin{aligned}
& det\left(\nabla^2 f(\boldsymbol{X}^\star) - \lambda^\star \nabla^2 g_1(\boldsymbol{X}^\star) - \mu^\star \nabla^2 h_1(\boldsymbol{X}^\star)\right) \\
=& det\left(\begin{pmatrix} 2, & 0 \\ 0, & 0 \end{pmatrix} - \lambda^\star \cdot \begin{pmatrix} 0, & 0 \\ 0, & -2 \end{pmatrix} - \mu^\star \cdot \begin{pmatrix} 2, & 0 \\ 0, & 2 \end{pmatrix}\right) \\
=& det\begin{pmatrix} 2-2\mu^\star, & 0 \\ 0, & 2\lambda^\star - 2\mu^\star \end{pmatrix} \\
=& 4(\lambda^\star - \mu^\star)(1-\mu^\star) > 0.
\end{aligned}$$

■

例 5.4 求下面问题的最优解

$$\min\ x_1^2 - x_2 - 3x_3$$

$$s.t.$$

$$-x_1 - x_2 - x_3 \geq 0$$

$$x_1^2 + 2x_2 - x_3 = 0$$

解 设$\boldsymbol{X}^\star = (x_1^\star, x_2^\star, x_3^\star)^T$为$\boldsymbol{KT}$点，则

$$\nabla f(\boldsymbol{X}^\star) = \lambda_1^\star \nabla g_1(\boldsymbol{X}^\star) + \mu_1^\star \nabla h_1(\boldsymbol{X}^\star)$$

$$\lambda_1^\star \geq 0$$

$$\lambda_1^\star(-x_1^\star - x_2^\star - x_3^\star) = 0$$

$$-x_1^\star - x_2^\star - x_3^\star \geq 0$$

$$(x_1^\star)^2 + 2x_2^\star - x_3^\star = 0$$

即

$$2x_1^\star + \lambda_1^\star - 2\mu_1^\star x_1^\star = 0$$

$$-1 + \lambda_1^\star - 2\mu_1^\star = 0$$

$$-3 + \lambda_1^\star + \mu_1^\star = 0$$

$$\lambda_1^\star \geq 0$$

$$\lambda_1^\star(-x_1^\star - x_2^\star - x_3^\star) = 0$$

$$-x_1^\star - x_2^\star - x_3^\star \geq 0$$

$$(x_1^\star)^2 + 2x_2^\star - x_3^\star = 0$$

解之，

$$\begin{cases} \boldsymbol{X}^\star = \left(-\frac{7}{2}, -\frac{35}{12}, \frac{77}{12}\right)^T \\ \lambda_1^\star = \frac{7}{3} \\ \mu_1^\star = \frac{2}{3} \end{cases}$$

因为

$$\nabla^2 f(\boldsymbol{X}^\star) - \lambda_1^\star \nabla^2 g_1(\boldsymbol{X}^\star) - \mu_1^\star \nabla^2 h_1(\boldsymbol{X}^\star) = \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{cases} \boldsymbol{Z}^T \nabla g_1(\boldsymbol{X}^\star) = 0 \\ \boldsymbol{Z}^T \nabla h_1(\boldsymbol{X}^\star) = 0 \end{cases} \Rightarrow \boldsymbol{Z} = (z_1, 2z_1, -3z_1)^T.$$

所以

$$\boldsymbol{Z}^T \begin{pmatrix} \frac{2}{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \boldsymbol{Z} = \frac{2}{3} z_1^2 > 0.$$

即$\boldsymbol{X}^{\star}$为(局部)最优解。 ▮

§ 5.2 惩罚函数法

5.2.1 基本思想

通过罚因子将约束优化问题变为一系列求无约束优化（罚函数）的极小值问题。

考虑

$$\min \ f(\boldsymbol{X}) \tag{5-2-12}$$

$$s.t.$$

$$h_j(\boldsymbol{X}) = 0, j = 1 \sim l. \tag{5-2-13}$$

罚函数法将其转化为

$$\min \ P(\boldsymbol{X}, m_k) = \min \ f(\boldsymbol{X}) + m_k \sum_{j=1}^{l} \Big(h_j(\boldsymbol{X})\Big)^2 \tag{5-2-14}$$

这里，称m_k为罚因子，其为一个很大的实数；称$P(\boldsymbol{X}, m_k)$为罚函数。通常令$\alpha(\boldsymbol{X}) \triangleq \sum_{j=1}^{l} (h_j(\boldsymbol{X}))^2$ 为罚项。

5.2.2 罚因子与拉格朗日乘子之间的关系

对于问题(5-2-12)：

令拉格朗日函数为

$$L(\boldsymbol{X},\lambda) = f(\boldsymbol{X}) + \sum_{j=1}^{l} \lambda_j h_j(\boldsymbol{X}) \tag{5-2-15}$$

罚函数为

$$P(\boldsymbol{X},m) = f(\boldsymbol{X}) + \sum_{j=1}^{l} m_j \Big(h_j(\boldsymbol{X})\Big)^2 \tag{5-2-16}$$

则有

$$\frac{\lambda_i}{2m_j} = h_j(\boldsymbol{X}^\star). \tag{5-2-17}$$

在最优解处，上式表明了三者之间的关系。

§ 5.3 外点罚函数法

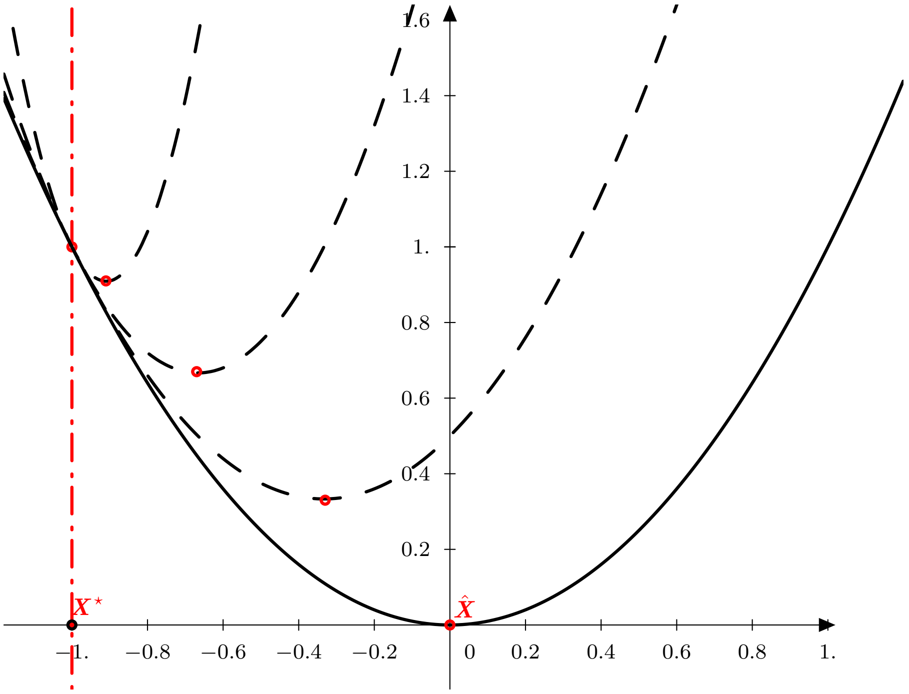
5.3.1 基本思想

例 5.5

$$\min \ f(x) = x^2$$

$s.t.$

$$g(x) = x + 1 \leq 0$$


1.6
1.4
1.2
1.
0.8
0.6
0.4
0.2
$X^\star$
$\hat{X}$
−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

$$\min\ f(\boldsymbol{X}) \tag{5-3-18}$$

$$s.t.$$

$$g_i(\boldsymbol{X}) \geq 0, i = 1 \sim m.$$

构造

$$P(\boldsymbol{X}, m_k) = f(\boldsymbol{X}) + m_k \sum_{i=1}^{m} \Big(\min\{g_i(\boldsymbol{X}), 0\} \Big)^2 \tag{5-3-19}$$

仿此，对于

$$\min \ f(\boldsymbol{X}) \tag{5-3-20}$$

$$s.t.$$

$$h_j(\boldsymbol{X}) = 0, j = 1 \sim l.$$

构造

$$P(\boldsymbol{X}, m_k) = f(\boldsymbol{X}) + m_k \sum_{j=1}^{l} \big(h_j(\boldsymbol{X})\big)^2. \tag{5-3-21}$$

因此，对于一般问题(NLP)，可如下构造罚函数

$$P(\boldsymbol{X}, m_k) = f(\boldsymbol{X}) + m_k\Big(\sum_{i=1}^{m}\big(\min\{g_i(\boldsymbol{X}), 0\}\big)^2 + \sum_{j=1}^{l}\big(h_j(\boldsymbol{X})\big)^2\Big). \tag{5-3-22}$$

若令

$$\alpha(\boldsymbol{X}) = \sum_{i=1}^{m}\big(\min\{g_i(\boldsymbol{X}), 0\}\big)^2 + \sum_{j=1}^{l}\big(h_j(\boldsymbol{X})\big)^2. \tag{5-3-23}$$

则约束优化问题(NLP)可转化为求解下面的无约束优化问题：

$$\min\ P(\boldsymbol{X}, m_k) = \min\ f(\boldsymbol{X}) + m_k\alpha(\boldsymbol{X}). \tag{5-3-24}$$

如果$\boldsymbol{X}^k \in \boldsymbol{S}$，则对任意$\boldsymbol{X} \in \boldsymbol{S}$，有

$$f(\boldsymbol{X}^k) = P(\boldsymbol{X}^k, m_k) \leq P(\boldsymbol{X}, m_k) = f(\boldsymbol{X}). \tag{5-3-25}$$

即，$\boldsymbol{X}^k$是约束优化问题(NLP)的最优解。

5.3.3 算法

见算法5.1。

例 5.6

$$\min \ x_1 + x_2$$

算法 5.1 外点(罚函数)法

步骤 1 选定$\boldsymbol{X}^0$，$m_1 > 0(m_1 = 1)$，$C > 1(C = 10)$，$k = 1$。

步骤 2 以$\boldsymbol{X}^{k-1}$为初始点，求解惩罚函数(5-3-22)的最优解$\boldsymbol{X}^k$。

步骤 3 若$m_k\alpha(\boldsymbol{X}^k) < \varepsilon$，则$\boldsymbol{X}^k$为约束优化问题($NLP$)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)。

$$s.t.$$

$$g_1(\boldsymbol{X}) = -x_1^2 + x_2 \geq 0.$$

$$g_2(\boldsymbol{X}) = x_1 \geq 0.$$

解

$$\boldsymbol{X}^m = \left(-\frac{1}{2(1+m)}, \left(\frac{1}{4(1+m)^2} - \frac{1}{2m}\right) \right)^T.$$

∎

例 5.7

$$\min\ (x-1)^2$$

$$s.t.$$

$$x - 2 \geq 0.$$

解 $x^\star = 2$。 ∎

5.3.4 收敛性定理

引理 5.4 设点列$\{\boldsymbol{X}^k\}$是由算法5.1所产生的序列，则有

$$P(\boldsymbol{X}^{k+1}, m_{k+1}) \geq P(\boldsymbol{X}^k, m_k) \tag{5-3-26}$$

$$\alpha(\boldsymbol{X}^{k+1}) \leq \alpha(\boldsymbol{X}^k) \tag{5-3-27}$$

$$f(\boldsymbol{X}^{k+1}) \geq f(\boldsymbol{X}^k) \tag{5-3-28}$$

这里，$k \geq 1$。

证明 首先

$$\begin{aligned}P(\boldsymbol{X}^{k+1}, m_{k+1}) &= f(\boldsymbol{X}^{k+1}) + m_{k+1}\alpha(\boldsymbol{X}^{k+1}) \\ &\geq f(\boldsymbol{X}^{k+1}) + m_k\alpha(\boldsymbol{X}^{k+1}) \\ &= P(\boldsymbol{X}^{k+1}, m_k) \\ &\geq P(\boldsymbol{X}^k, m_k).\end{aligned} \tag{5-3-29}$$

接着

$$f(\boldsymbol{X}^{k+1}) + m_k\alpha(\boldsymbol{X}^{k+1}) \geq f(\boldsymbol{X}^k) + m_k\alpha(\boldsymbol{X}^k), \tag{5-3-30}$$

$$f(\boldsymbol{X}^k) + m_{k+1}\alpha(\boldsymbol{X}^k) \geq f(\boldsymbol{X}^{k+1}) + m_{k+1}\alpha(\boldsymbol{X}^{k+1}). \tag{5-3-31}$$

即

$$
\begin{aligned}
m_k\Big(\alpha(\boldsymbol{X}^k)-\alpha(\boldsymbol{X}^{k+1})\Big) &\le f(\boldsymbol{X}^{k+1})-f(\boldsymbol{X}^k) \\
&\le m_{k+1}\Big(\alpha(\boldsymbol{X}^k)-\alpha(\boldsymbol{X}^{k+1})\Big).
\end{aligned}
\tag{5-3-32}
$$

则有

$$
(m_{k+1}-m_k)\Big(\alpha(\boldsymbol{X}^k)-\alpha(\boldsymbol{X}^{k+1})\Big) \ge 0. \tag{5-3-33}
$$

因此

$$
\alpha(\boldsymbol{X}^{k+1}) \le \alpha(\boldsymbol{X}^k). \tag{5-3-34}
$$

最后，显然有$f(\boldsymbol{X}^{k+1}) \ge f(\boldsymbol{X}^k)$成立。 ∎

定理 5.5 设f，g_i，h_j都是$\boldsymbol{R}^n$上的连续函数，则由算法5.1产生的任何收敛序列$\{\boldsymbol{X}^k\}$的极限点必是约束优化问题(NLP)的极小点。

证明 设$\boldsymbol{X}^k \to \overline{\boldsymbol{X}}$，$\boldsymbol{X}^\star$为问题($NLP$)的极小点。因为

$$f(\boldsymbol{X}^\star) = P(\boldsymbol{X}^\star, m_k) \geq P(\boldsymbol{X}^k, m_k) \geq f(\boldsymbol{X}^k), \tag{5-3-35}$$

所以

$$\lim_{k\to\infty} P(\boldsymbol{X}^k, m_k) = \overline{P} \leq f(\boldsymbol{X}^\star), \tag{5-3-36}$$

$$\lim_{k\to\infty} f(\boldsymbol{X}^k) = \overline{f} \leq f(\boldsymbol{X}^\star). \tag{5-3-37}$$

所以

$$\lim_{k\to\infty}\alpha(\boldsymbol{X}^k)=\lim_{k\to\infty}\frac{P(\boldsymbol{X}^k,m_k)-f(\boldsymbol{X}^k)}{m_k}=0. \tag{5-3-38}$$

因此

$$\alpha(\overline{\boldsymbol{X}})=\lim_{k\to\infty}\alpha(\boldsymbol{X}^k)=0. \tag{5-3-39}$$

所以

$$\overline{\boldsymbol{X}}\in\boldsymbol{S}. \tag{5-3-40}$$

从而

$$f(\boldsymbol{X}^{\star}) \leq f(\overline{\boldsymbol{X}}). \tag{5-3-41}$$

又因为

$$f(\boldsymbol{X}^{\star}) \geq f(\overline{\boldsymbol{X}}). \tag{5-3-42}$$

所以

$$f(\boldsymbol{X}^{\star}) = f(\overline{\boldsymbol{X}}). \tag{5-3-43}$$

∎

§ 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。

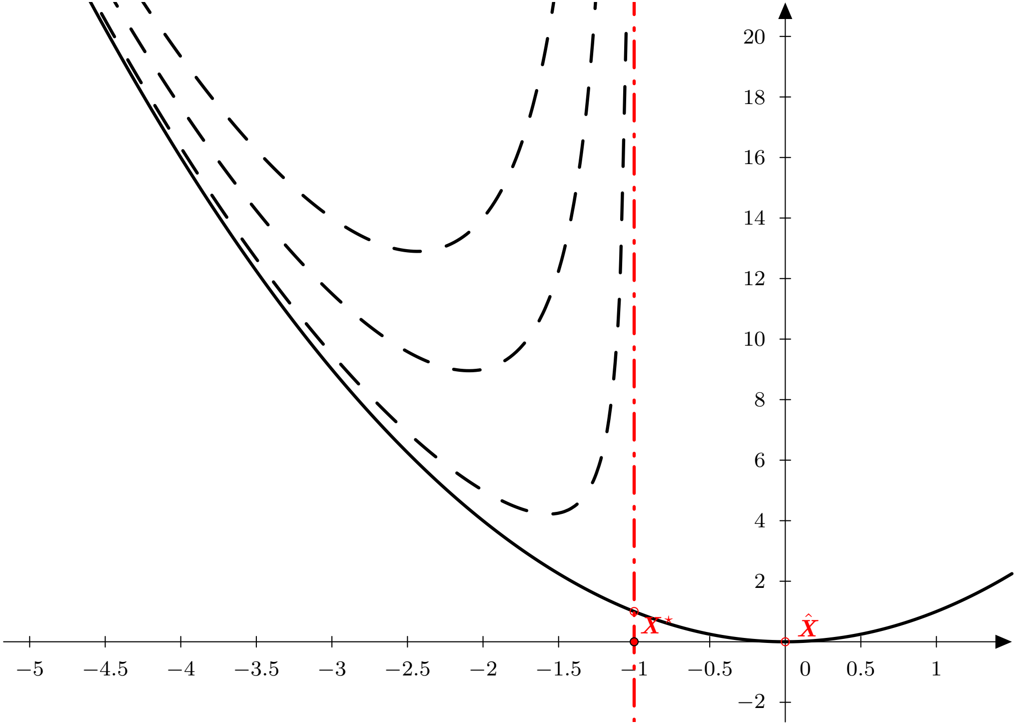
5.4.1 基本思想

例 5.8

$$\min \quad f(x) = x^2$$

$$s.t.$$

$$g(x) = x + 1 \leq 0$$


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
$X^\star$
$\hat{X}$

考虑

$$\min \ f(\boldsymbol{X}) \tag{5-4-44}$$

$$s.t.$$

$$g_i(\boldsymbol{X}) \geq 0, i = 1 \sim m.$$

这里$\boldsymbol{S} \triangleq \{\boldsymbol{X} \mid g_i(\boldsymbol{X}) \geq 0, i = 1 \sim m\}$，并要求$\boldsymbol{S}^o \neq \emptyset$。

定义罚函数(障碍函数)为

$$P(\boldsymbol{X}, r_k) = f(\boldsymbol{X}) + r_k \sum_{i=1}^{m} \frac{1}{g_i(\boldsymbol{X})}. \tag{5-4-45}$$

这里称

$$\beta(\boldsymbol{X}) \triangleq \sum_{i=1}^{m} \frac{1}{g_i(\boldsymbol{X})}. \tag{5-4-46}$$

为罚项。

或定义罚函数为

$$P(\boldsymbol{X}, r_k) = f(\boldsymbol{X}) - r_k \sum_{i=1}^{m} \ln g_i(\boldsymbol{X}), \tag{5-4-47}$$

$$P(\boldsymbol{X}, r_k) = f(\boldsymbol{X}) + r_k \sum_{i=1}^{m} \frac{1}{\big(g_i(\boldsymbol{X})\big)^2}. \tag{5-4-48}$$

例 5.9

$$\min\ f(x) = x$$

$s.t.$

$$g(x) = x \geq 0.$$

5.4.2 算法

例 5.10

$$\min\ f(\boldsymbol{X}) = \frac{1}{3}(x_1+1)^3 + x_2$$

$s.t.$

算法 5.2 内点(罚函数)法

步骤 1 选定$\boldsymbol{X}^0 \in \boldsymbol{S}^o$，$r_1 > 0(r_1 = 10)$，$C < 1(C = 0.1)$，$k = 1$。

步骤 2 以$\boldsymbol{X}^{k-1}$为初始点，求解障碍函数(5-4-45)的最优解$\boldsymbol{X}^k$。

步骤 3 若$r_k\beta(\boldsymbol{X}^k) < \varepsilon$，则$\boldsymbol{X}^k$为约束优化问题(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)。

$$g_1(\boldsymbol{X}) = x_1 - 1 \geq 0,$$

$$g_2(\boldsymbol{X}) = x_2 \geq 0.$$

解 $\boldsymbol{X}^\star = (1, 0)^T$。 ∎

例 5.11 用对数罚函数法(5-4-47)求解

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

$$s.t.$$

$$g_1(\boldsymbol{X}) = -x_1^2 + x_2 \geq 0,$$

$$g_2(\boldsymbol{X}) = x_1 \geq 0.$$

解 $\boldsymbol{X}^{\star} = (0,0)^T$。 ∎

5.4.3 收敛性定理

引理 5.6 由内点法(5.2)产生的点列$\boldsymbol{X}^k$，有

$$P(\boldsymbol{X}^{k+1}, r_{k+1}) \leq P(\boldsymbol{X}^k, r_k). \tag{5-4-49}$$

证明

$$\begin{aligned} P(\boldsymbol{X}^k, r_k) &= f(\boldsymbol{X}^k) + r_k\beta(\boldsymbol{X}^k) \\ &\geq f(\boldsymbol{X}^k) + r_{k+1}\beta(\boldsymbol{X}^k) \\ &= P(\boldsymbol{X}^k, r_{k+1}) \\ &\geq P(\boldsymbol{X}^{k+1}, r_{k+1}). \end{aligned} \tag{5-4-50}$$

∎

定理 5.7 设f，g_i都是$\boldsymbol{R}^n$上的连续函数，则由内点算法5.2产生的任何收敛序列$\{\boldsymbol{X}^k\}$的极限点必是约束优化问题(5-4-44)的极小点。

证明 设$\boldsymbol{X}^k \to \overline{\boldsymbol{X}}$，并令$\boldsymbol{X}^\star$是优化问题(5-4-44) 的极小点。因为

$$P(\boldsymbol{X}^k, r_k) \geq f(\boldsymbol{X}^k) \geq f(\boldsymbol{X}^\star), \tag{5-4-51}$$

所以

$$\lim_{k\to\infty} P(\boldsymbol{X}^k, r_k) \triangleq \overline{P} \geq f(\boldsymbol{X}^\star). \tag{5-4-52}$$

因为f连续，所以对$\forall \varepsilon > 0$，都存在$\widetilde{\boldsymbol{X}} \in \boldsymbol{S}^o$，使得

$$f(\widetilde{\boldsymbol{X}}) - f(\boldsymbol{X}^\star) < \varepsilon, \tag{5-4-53}$$

又因为

$$P(\boldsymbol{X}^k, r_k) - f(\boldsymbol{X}^\star) \le f(\widetilde{\boldsymbol{X}}) - f(\boldsymbol{X}^\star) + r_k\beta(\widetilde{\boldsymbol{X}}), \tag{5-4-54}$$

所以当$k \to \infty$，就有

$$P(\boldsymbol{X}^k, r_k) - f(\boldsymbol{X}^\star) < \varepsilon. \tag{5-4-55}$$

因此

$$\lim_{k\to\infty} P(\boldsymbol{X}^k, r_k) = f(\boldsymbol{X}^\star) = \lim_{k\to\infty} f(\boldsymbol{X}^k). \tag{5-4-56}$$

∎

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

考虑等式约束问题：

$$\begin{aligned} &\min \ f(\boldsymbol{X}) \\ &s.t. \end{aligned} \tag{5-5-57}$$

$$h_j(\boldsymbol{X}) = 0,\ j = 1, \cdots, l.$$

这里，$f(\boldsymbol{X})$，$h_j(\boldsymbol{X})$是二次连续可微函数。

定义如下**乘子罚函数**(**增广*Lagrange*函数**)：

$$\phi(\boldsymbol{X}, \boldsymbol{V}, \sigma) = f(\boldsymbol{X}) - \sum_{j=1}^{l} v_j h_j(\boldsymbol{X}) + \frac{\sigma}{2} \sum_{j=1}^{l} h_j^2(\boldsymbol{X}). \tag{5-5-58}$$

这里，$\boldsymbol{V} = (v_1, \cdots, v_l)^T$，$\sigma > 0$。

定理 5.8 设$\widetilde{\boldsymbol{X}}$，$\widetilde{\boldsymbol{V}}$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$\forall \sigma > \sigma'$，$\widetilde{\boldsymbol{X}}$是$\phi(\boldsymbol{X}, \widetilde{\boldsymbol{V}}, \sigma)$的严格局部最小点。

反之，若存在$\widetilde{\boldsymbol{X}}$是问题(5-5-57)的可行解，且对于某个$\widetilde{\boldsymbol{V}}$，$\widetilde{\boldsymbol{X}}$是$\phi(\boldsymbol{X}, \widetilde{\boldsymbol{V}}, \sigma)$的极小点，又满足极小点的二阶充分条件，则$\widetilde{\boldsymbol{X}}$是问题(5-5-57)的最优解。

因此，若知道最优乘子$\widetilde{\boldsymbol{V}} \triangleq \boldsymbol{V}^{\star}$，只要取充分大的罚因子$\sigma < +\infty$，就能通过$\phi(\boldsymbol{X}, \boldsymbol{V}^{\star}, \sigma)$ 得到问题(5-5-57)的最优解。

但是，最优乘子$\boldsymbol{V}^{\star}$一般事前是未知的。所以，一般先给充分大的σ和初始估计$\boldsymbol{V}^0$，然后在迭代过程中，使得$\boldsymbol{V}^k \rightarrow \boldsymbol{V}^{\star}$。

设在第k次迭代，$Lagrange$乘子的估计为$\boldsymbol{V}^k$，罚因子

为σ，$\boldsymbol{X}^k$为(5-5-58)的最优解，则

$$
\begin{aligned}
\nabla\phi(\boldsymbol{X}^k,\boldsymbol{V}^k,\sigma) &= \nabla f(\boldsymbol{X}^k) - \sum_{j=1}^{l} v_j^k \nabla h_j(\boldsymbol{X}^k) \\
&+ \sigma\sum_{j=1}^{l} h_j(\boldsymbol{X}^k)\nabla h_j(\boldsymbol{X}^k). \\
&= \nabla f(\boldsymbol{X}^k) - \sum_{j=1}^{l}\big(v_j^k - \sigma h_j(\boldsymbol{X}^k)\big)\nabla h_j(\boldsymbol{X}^k) \\
&= 0.
\end{aligned}
\tag{5-5-59}
$$

又因问题(5-5-57)的最优解$\boldsymbol{X}^\star$满足

$$\nabla f(\boldsymbol{X}^\star) - \sum_{j=1}^{l} v_j^\star \nabla h_j(\boldsymbol{X}^\star) = 0. \tag{5-5-60}$$

所以令

$$v_j^{k+1} = v_j^k - \sigma h_j(\boldsymbol{X}^k),\ j = 1, \cdots, l. \tag{5-5-61}$$

例 5.12

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

$$s.t.$$

$$h_1(\boldsymbol{X}) = x_1 + x_2 - 1 = 0.$$

令$\sigma = 2$，$v^1 = 1$。

解 考虑

$$\begin{aligned}\phi(\boldsymbol{X}, v^k, \sigma) = 2x_1^2 + x_2^2 - 2x_1x_2 - v^k(x_1 + x_2 - 1)\\ + \frac{\sigma}{2}(x_1 + x_2 - 1)^2.\end{aligned}$$

易得，在第k次迭代时，$\phi(\boldsymbol{X}, v^k, \sigma)$的最优解为

$$\boldsymbol{X}^k = \left(\frac{v^k + 2}{6}, \ \frac{v^k + 2}{4}\right)^T.$$

因此

$$
\begin{aligned}
v^{k+1} &= v^k - \sigma h_1(\boldsymbol{X}^k) \\
&= v^k - 2\left(\frac{v^k+2}{6} + \frac{v^k+2}{4} - 1\right) \\
&= \frac{v^k}{6} + \frac{1}{3}.
\end{aligned}
$$

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，$v^\star = \frac{2}{5}$，$\boldsymbol{X}^\star = (\frac{2}{5}, \frac{3}{5})^T$。 ∎

考虑不等式约束问题：

$$\min \ f(\boldsymbol{X})$$

$$s.t. \tag{5-5-62}$$

$$g_i(\boldsymbol{X}) \geq 0, \ i = 1, \cdots, m.$$

这里，$f(\boldsymbol{X})$，$g_i(\boldsymbol{X})$是二次连续可微函数。

引入变量y_i，$i = 1, \cdots, m$，问题可变为：

$$\min \ f(\boldsymbol{X})$$

$$s.t. \tag{5-5-63}$$

$$g_i(\boldsymbol{X}) - y_i^2 = 0,\ i = 1, \cdots, m.$$

定义如下**增广*Lagrange*函数**：

$$\phi(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, \sigma) = f(\boldsymbol{X}) - \sum_{i=1}^{m} w_i \left(g_i(\boldsymbol{X}) - y_i^2\right) + \frac{\sigma}{2}\sum_{i=1}^{m}\left(g_i(\boldsymbol{X}) - y_i^2\right)^2. \tag{5-5-64}$$

这里， $\boldsymbol{W} = (w_1, \cdots, w_m)^T$， $\boldsymbol{Y} = (y_1, \cdots, y_m)^T$。

因为

$$\phi(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{W}, \sigma)$$

$$
\begin{aligned}
&= f(\boldsymbol{X}) + \sum_{i=1}^{m} \left(-w_i \left(g_i(\boldsymbol{X}) - y_i^2 \right) + \frac{\sigma}{2} \left(g_i(\boldsymbol{X}) - y_i^2 \right)^2 \right) \\
&= f(\boldsymbol{X}) + \sum_{i=1}^{m} \left(\frac{\sigma}{2} \left(y_i^2 - \frac{1}{\sigma} (\sigma g_i(\boldsymbol{X}) - w_i) \right)^2 - \frac{w_i^2}{2\sigma} \right).
\end{aligned}
\tag{5-5-65}
$$

令

$$
y_i^2 = \frac{1}{\sigma} \max\left\{0, \sigma g_i(\boldsymbol{X}) - w_i\right\}, \ i = 1, \cdots, m. \tag{5-5-66}
$$

则问题(5-5-62)转化成求解下面问题

$$\phi(\boldsymbol{X}, \boldsymbol{W}, \sigma) = f(\boldsymbol{X}) + \frac{1}{2\sigma}\sum_{i=1}^{m}\Big(\big(\max\{0, w_i - \sigma g_i(\boldsymbol{X})\}\big)^2 - w_i^2\Big). \tag{5-5-67}$$

设在第k次迭代，$Lagrange$乘子的估计为$\boldsymbol{W}^k$，罚因子为σ，$\boldsymbol{X}^k$为(5-5-67)的最优解，则

$$\nabla\phi(\boldsymbol{X}^k, \boldsymbol{W}^k, \sigma) = \nabla f(\boldsymbol{X}^k) - \sum_{i=1}^{m}\left(\max\{0, w_i^k - \sigma g_i(\boldsymbol{X}^k)\}\right)\nabla g_i(\boldsymbol{X}^k)$$

$$= 0. \tag{5-5-68}$$

又因问题(5-5-62)的最优解$\boldsymbol{X}^\star$满足

$$\nabla f(\boldsymbol{X}^\star) - \sum_{i=1}^{m} w_i^\star \nabla g_i(\boldsymbol{X}^\star) = 0. \tag{5-5-69}$$

所以令

$$w_i^{k+1} = \max\{0, w_i^k - \sigma g_i(\boldsymbol{X}^k)\},\ \ i = 1, \cdots, m. \tag{5-5-70}$$

对于一般情形：

$$\min\ \ f(\boldsymbol{X})$$

$$s.t. \tag{5-5-71}$$

$$g_i(\boldsymbol{X}) \geq 0,\ i = 1, \cdots, m.$$

$$h_j(\boldsymbol{X}) = 0,\ j = 1, \cdots, l.$$

定义乘子罚函数如下：

$$\begin{aligned}
&\phi(\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{V}, \sigma) \\
&\quad = f(\boldsymbol{X}) + \frac{1}{2\sigma}\sum_{i=1}^{m}\Big(\big(\max\{0, w_i - \sigma g_i(\boldsymbol{X})\}\big)^2 - w_i^2\Big) \\
&\quad - \sum_{j=1}^{l} v_j h_j(\boldsymbol{X}) + \frac{\sigma}{2}\sum_{j=1}^{l} h_j^2(\boldsymbol{X}). \qquad (5\text{-}5\text{-}72)
\end{aligned}$$

这里,

$$\begin{aligned} w_i^{k+1} &= \max\left\{0, w_i^k - \sigma g_i(\boldsymbol{X}^k)\right\}; \\ v_j^{k+1} &= v_j^k - \sigma h_j(\boldsymbol{X}^k). \end{aligned} \tag{5-5-73}$$

乘子算法如下，见算法5.3。

例 5.13

$$\min\ x_1^2 + 2x_2^2$$

$$s.t.$$

$$x_1 + x_2 \geq 1.$$

算法 5.3 乘子算法

步骤 1 给定初始点$\boldsymbol{X}^0$，乘子向量估计$\boldsymbol{W}^1$，$\boldsymbol{V}^1$，参数σ，误差$\varepsilon > 0$，常数$C > 1$，$0 < \beta < 1$，令$k = 1$。

步骤 2 求解$\min\ \phi(\boldsymbol{X}, \boldsymbol{W}^k, \boldsymbol{V}^k, \sigma)$，得到$\boldsymbol{X}^k$。

步骤 3

if $\|\alpha(\boldsymbol{X}^k)\| < \varepsilon$ **then**

停机, **return** $\boldsymbol{X}^k$;

else if $\frac{\|\alpha(\boldsymbol{X}^k)\|}{\|\alpha(\boldsymbol{X}^{k-1})\|} \geq \beta$ **then**

$\sigma = C \cdot \sigma$;

end if

步骤 4 通过公式(5-5-73)修正$\boldsymbol{W}^k$和$\boldsymbol{V}^k$，令$k = k + 1$，转步骤2。

这里，$\sigma = 2$，$w^1 = 1$。

解 易得$\phi(\boldsymbol{X}, w^k, \sigma)$的最优解为

$$\boldsymbol{X}^k = \begin{cases} \frac{2+w^k}{5} \\ \frac{2+w^k}{10} \end{cases}$$

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，$w^\star = \frac{4}{3}$，$\boldsymbol{X}^\star = (\frac{2}{3}, \frac{1}{3})^T$。 ∎

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 5.14

$$\min \ x_1 + \frac{1}{3}(x_2+1)^2$$

$$s.t.$$

$$x_1 \geq 0,$$

$$x_2 \geq 1.$$

解 $\boldsymbol{X}^{\star} = (0,1)^T$。 ∎

§ 5.6 *Rosen*梯度投影法

考虑

$$\min \ f(\boldsymbol{X}) \tag{5-6-74}$$

$s.t.$

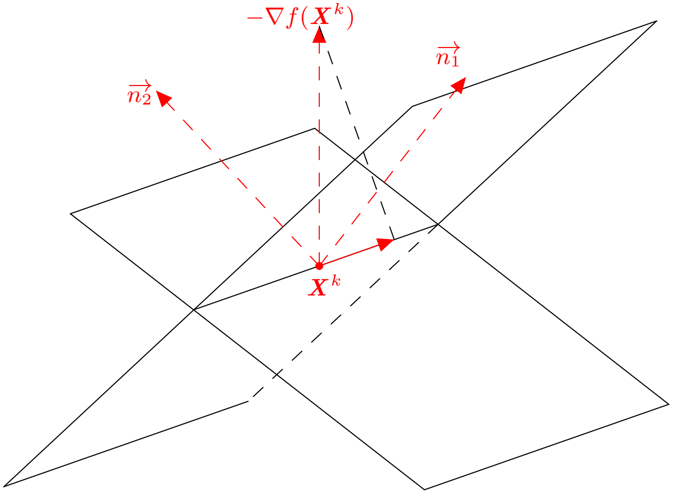
$$\boldsymbol{A}\boldsymbol{X} \geq \boldsymbol{b}, \tag{5-6-75}$$

$$\boldsymbol{C}\boldsymbol{X} = \boldsymbol{d}. \tag{5-6-76}$$

这里$\boldsymbol{A}$是$m \times n$矩阵，$\boldsymbol{C}$是$l \times n$矩阵，$\boldsymbol{b}$是m维向量，$\boldsymbol{d}$是l维向量，$f(\boldsymbol{X})$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$\boldsymbol{X}^k$的负梯度方向$-\nabla f(\boldsymbol{X}^k)$不是可行方向，则将其投影到$\boldsymbol{X}^k$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。


$-\nabla f(\boldsymbol{X}^k)$
$\overrightarrow{n_1}$
$\overrightarrow{n_2}$
$\boldsymbol{X}^k$

5.6.2 下降可行方向的确定

定义 5.8 设$\mathbb{P}$为$n\times n$矩阵，若$\mathbb{P}=\mathbb{P}^T$且$\mathbb{P}^2=\mathbb{P}$，则称$\mathbb{P}$为投影矩阵。

引理 5.9 设$\mathbb{P}$为投影矩阵，则

(1)$\mathbb{P}$为半正定；

(2)充要条件为$\mathbb{Q}\triangleq \boldsymbol{I}-\mathbb{P}$是投影矩阵；

(3)对$\forall \boldsymbol{X}\in \boldsymbol{R}^n$，可唯一的表示成$\boldsymbol{X}=p+q$，$p\in \boldsymbol{L}$，$q\in \boldsymbol{L}^{\perp}$，这里

$$\boldsymbol{L}\triangleq\{\mathbb{P}\boldsymbol{X}\mid \boldsymbol{X}\in \boldsymbol{R}^n\}, \tag{5-6-77}$$

$$\boldsymbol{L}^{\perp} \triangleq \{\mathbb{Q}\boldsymbol{X} \mid \boldsymbol{X} \in \boldsymbol{R}^n\}. \tag{5-6-78}$$

定理 5.10 对$\forall \boldsymbol{X} \in \boldsymbol{R}^n$，$\boldsymbol{X} \overset{1}{=} \mathbb{P}\boldsymbol{X} + \mathbb{Q}\boldsymbol{X}$，并称$\mathbb{P}\boldsymbol{X}$为$\boldsymbol{X}$在子空间$\boldsymbol{L}$上的投影，$\mathbb{P}$为$\boldsymbol{R}^n$到$\boldsymbol{L}$的投影矩阵；对于$\mathbb{Q}$类似。

在可行点$\boldsymbol{X}^k$处，将$\boldsymbol{A}$，$\boldsymbol{b}$分解为

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_k' \\ \boldsymbol{A}_k'' \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_k' \\ \boldsymbol{b}_k'' \end{pmatrix}, \tag{5-6-79}$$

并要求

$$\boldsymbol{A}_k' \boldsymbol{X}^k = \boldsymbol{b}_k' \tag{5-6-80}$$

$$\boldsymbol{A}_k^{''}\boldsymbol{X}^k > \boldsymbol{b}_k^{''}. \tag{5-6-81}$$

令

$$\boldsymbol{N}_k = \begin{pmatrix} \boldsymbol{A}_k^{'} \\ \boldsymbol{C}_k \end{pmatrix}. \tag{5-6-82}$$

并假设$rank(\boldsymbol{N}_k) = r$。

定义

$$\boldsymbol{N}_k \triangleq \begin{pmatrix} \boldsymbol{n}_1^T \\ \boldsymbol{n}_2^T \\ \vdots \\ \boldsymbol{n}_r^T \end{pmatrix}. \tag{5-6-83}$$

这里，$\boldsymbol{n}_i^T$表示$\boldsymbol{N}_k$的第i行，$i=1,\cdots,r$。

考虑

$$\mathbb{P}_k=\boldsymbol{N}_k^T(\boldsymbol{N}_k\boldsymbol{N}_k^T)^{-1}\boldsymbol{N}_k, \tag{5-6-84}$$

$$\mathbb{Q}_k=\boldsymbol{I}-\mathbb{P}_k \tag{5-6-85}$$

定理 5.11 对于优化问题(5-6-74)，搜索方向$\boldsymbol{P}^k$为点$\boldsymbol{X}^k$处的可行方向的充分必要条件是

$$\boldsymbol{A}_k'\boldsymbol{P}^k\geq 0; \tag{5-6-86}$$

$$\boldsymbol{C}_k\boldsymbol{P}^k=0. \tag{5-6-87}$$

定理 5.12 若$\boldsymbol{P}^k = -\mathbb{Q}_k \nabla f(\boldsymbol{X}^k) \neq 0$，则$\boldsymbol{P}^k$为点$\boldsymbol{X}^k$处的可行下降方向，这里$\mathbb{Q}_k = \boldsymbol{I} - \boldsymbol{N}_k^T(\boldsymbol{N}_k \boldsymbol{N}_k^T)^{-1}\boldsymbol{N}_k$。

证明

$$\nabla f(\boldsymbol{X}^k)^T \boldsymbol{P}^k = -\Big(\mathbb{Q}_k \nabla f(\boldsymbol{X}^k)\Big)^T \Big(\mathbb{Q}_k \nabla f(\boldsymbol{X}^k)\Big) < 0; \quad \text{(5-6-88)}$$

$$\boldsymbol{N}_k \boldsymbol{P}^k = -\boldsymbol{N}_k \Big(\boldsymbol{I} - \boldsymbol{N}_k^T(\boldsymbol{N}_k \boldsymbol{N}_k^T)^{-1}\boldsymbol{N}_k\Big) \nabla f(\boldsymbol{X}^k) = 0. \quad \text{(5-6-89)}$$

$\blacksquare$

5.6.3 直线搜索及终止准则

一维搜索：

$$\min_{0\leq t\leq \overline{t_k}} f(\boldsymbol{X}^k+t\boldsymbol{P}^k). \tag{5-6-90}$$

因为

$$\boldsymbol{A}_k'\boldsymbol{P}^k\geq 0, \boldsymbol{C}_k\boldsymbol{P}^k=0. \tag{5-6-91}$$

所以只需

$$\boldsymbol{A}_k''(\boldsymbol{X}^k+t\boldsymbol{P}^k)\geq \boldsymbol{b}_k''. \tag{5-6-92}$$

即

$$\boldsymbol{A}_k''\boldsymbol{X}^k - \boldsymbol{b}_k'' + t\boldsymbol{A}_k''\boldsymbol{P}^k \geq 0. \tag{5-6-93}$$

令

$$\boldsymbol{A}_k''\boldsymbol{X}^k - \boldsymbol{b}_k'' \triangleq \boldsymbol{u} = (u_1, u_2, \cdots, u_\theta)^T. \tag{5-6-94}$$

$$\boldsymbol{A}_k''\boldsymbol{P}^k \triangleq \boldsymbol{v} = (v_1, v_2, \cdots, v_\theta)^T. \tag{5-6-95}$$

这里，$\theta \leq m$。

按照下式确定$\overline{t_k}$：

$$\overline{t_k} = \begin{cases} +\infty, & \boldsymbol{v} \geq 0; \\ \min_{1\leq i\leq\theta}\left\{-\frac{u_i}{v_i} \mid v_i < 0\right\}, & \boldsymbol{v} \ngeq 0. \end{cases} \tag{5-6-96}$$

5.6.4 算法

当$\boldsymbol{P}^k = -\mathbb{Q}_k\nabla f(\boldsymbol{X}^k) \neq 0$时，其就是可行下降方向；

若当$\boldsymbol{P}^k = 0$时，有

$$
\begin{aligned}
\mathbb{Q}_k \nabla f(\boldsymbol{X}^k) &= \left(\boldsymbol{I} - \boldsymbol{N}_k^T(\boldsymbol{N}_k\boldsymbol{N}_k^T)^{-1}\boldsymbol{N}_k\right)\nabla f(\boldsymbol{X}^k) \\
&= \nabla f(\boldsymbol{X}^k) - \boldsymbol{N}_k^T(\boldsymbol{N}_k\boldsymbol{N}_k^T)^{-1}\boldsymbol{N}_k\nabla f(\boldsymbol{X}^k) \\
&= \nabla f(\boldsymbol{X}^k) - \boldsymbol{N}_k^T\boldsymbol{q}_k \\
&= 0.
\end{aligned}
\tag{5-6-97}
$$

这里，$\boldsymbol{q}_k \triangleq (\boldsymbol{N}_k\boldsymbol{N}_k^T)^{-1}\boldsymbol{N}_k\nabla f(\boldsymbol{X}^k)$。

此时，

$$\begin{aligned}\nabla f(\boldsymbol{X}^k) &= \boldsymbol{N}_k^T \boldsymbol{q}_k \\ &= \begin{pmatrix} \boldsymbol{A}_k' \\ \boldsymbol{C}_k \end{pmatrix}^T \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix} \\ &= (\boldsymbol{A}_k')^T \boldsymbol{y} + \boldsymbol{C}_k^T \boldsymbol{z}.\end{aligned} \tag{5-6-98}$$

当$\boldsymbol{y} \geq 0$时，$\boldsymbol{X}^k$是$\boldsymbol{KT}$点；

若$\boldsymbol{y} \ngeq 0$时，不妨假设$y_i < 0$，从$\boldsymbol{N}_k$中删去第i行向量得到$\overline{\boldsymbol{N}}_k$，并令

$$\overline{\mathbb{Q}}_k = \boldsymbol{I} - \overline{\boldsymbol{N}}_k^T (\overline{\boldsymbol{N}}_k \overline{\boldsymbol{N}}_k^T)^{-1} \overline{\boldsymbol{N}}_k. \tag{5-6-99}$$

容易证明$\overline{\mathbb{Q}}_k$为投影矩阵，这时，

$$\overline{\boldsymbol{P}}^k = -\overline{\mathbb{Q}}_k \nabla f(\boldsymbol{X}^k) \tag{5-6-100}$$

就是点$\boldsymbol{X}^k$处的一个下降方向。

又因为

$$\begin{aligned}\overline{\boldsymbol{P}}^k &= -\overline{\mathbb{Q}}_k \nabla f(\boldsymbol{X}^k)\\ &= -\overline{\mathbb{Q}}_k(\boldsymbol{A}'_k)^T\boldsymbol{y} - \overline{\mathbb{Q}}_k\boldsymbol{C}_k^T\boldsymbol{z}\\ &= -\overline{\mathbb{Q}}_k(\boldsymbol{A}'_k)^T\boldsymbol{y} \qquad (5\text{-}6\text{-}101)\\ &= -\sum_{\forall j}\overline{\mathbb{Q}}_k\boldsymbol{n}_j y_j\end{aligned}$$

$$= -y_i \overline{\mathbb{Q}}_k \boldsymbol{n}_i.$$

这里，$\boldsymbol{n}_j^T$表示$\boldsymbol{N}_k$的第j行。

所以

$$\boldsymbol{n}_j^T \overline{\boldsymbol{P}}^k = -y_i \boldsymbol{n}_j^T \overline{\mathbb{Q}}_k \boldsymbol{n}_i$$
$$= \begin{cases} 0, & j \neq i; \\ -y_i \|\overline{\mathbb{Q}}_k n_i\|^2 \geq 0, & j = i. \end{cases} \tag{5-6-102}$$

从而$\overline{\boldsymbol{P}}^k$为点$\boldsymbol{X}^k$处的一个可行下降方向。

算法 5.4 *Rosen*梯度投影算法

步骤 1 选定$\boldsymbol{X}^0$，$k=0$。

步骤 2 令$\boldsymbol{N}_k=((\boldsymbol{A}'_k)^T,\boldsymbol{C}_k^T)^T$。

步骤 3 若$r>0$，计算$\mathbb{Q}_k=\boldsymbol{I}-\boldsymbol{N}_k^T(\boldsymbol{N}_k\boldsymbol{N}_k^T)^{-1}\boldsymbol{N}_k$。转步骤(5)。

步骤 4 若$r=0$，计算$\mathbb{Q}_k=\boldsymbol{I}$。

步骤 5 计算$\boldsymbol{P}^k=-\mathbb{Q}_k\nabla f(\boldsymbol{X}^k)$。

步骤 6 若$\|\boldsymbol{P}^k\|<\varepsilon$，转步骤(7)；否则转步骤(9)。

步骤 7 若$r=0$，则$\boldsymbol{X}^k$就是最优解；若$r\neq 0$，计算$\boldsymbol{q}_k=(\boldsymbol{y}^T,\boldsymbol{z}^T)^T$。

步骤 8 若$\boldsymbol{y}\geq 0$，则$\boldsymbol{X}^k$是最优解($\boldsymbol{KT}$点)；若$\boldsymbol{y}\ngeq 0$，去掉$\boldsymbol{N}_k$中相应的行向量。转步骤(2)。

步骤 9 求解$\min_{0\leq t\leq \overline{t_k}} f(\boldsymbol{X}^k+t\boldsymbol{P}^k)$，$k=k+1$，转步骤(2)。

例 5.15

$$
\begin{aligned}
&\min \ f(\boldsymbol{X}) = x_1^2 + x_2^2 + 2x_2 + 5 \\
&s.t. \\
&\qquad x_1 - 2x_2 \geq 0 \\
&\qquad x_j \geq 0, \quad j = 1, 2.
\end{aligned}
\tag{5-6-103}
$$

初始点$\boldsymbol{X}^0 = (2, 0)^T$。

解

$$\boldsymbol{g} = \nabla f(\boldsymbol{X}) = (2x_1, 2x_2 + 2)^T$$

第一次迭代：

(1)求$\boldsymbol{P}^0$

$$\boldsymbol{g}^0 = (4,2)^T,$$

$$\boldsymbol{N}_0 = (0,1),$$

$$\boldsymbol{M}_0 = (\boldsymbol{N}_0\boldsymbol{N}_0^T)^{-1} = \big((0,1)(0,1)^T\big)^{-1} = 1,$$

$$\mathbb{Q}_0 = \boldsymbol{I} - \boldsymbol{N}_0^T\boldsymbol{M}_0\boldsymbol{N}_0 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}0\\ 1\end{pmatrix}\cdot 1\cdot(0,1) = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},$$

$$\boldsymbol{P}^0 = -\mathbb{Q}_0\boldsymbol{g}^0 = -\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}(4,2)^T = (-4,0)^T,$$

$\|\boldsymbol{P}^0\| \neq 0;$

(2)求$\boldsymbol{X}^1$

$$\boldsymbol{A}_0^{''} = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix},$$

$$\boldsymbol{b}_0^{''} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\boldsymbol{u} = \boldsymbol{A}_0^{''}\boldsymbol{X}^0 - \boldsymbol{b}_0^{''} = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

$$\boldsymbol{v} = \boldsymbol{A}_0^{''}\boldsymbol{P}^0 = \begin{pmatrix} 1 & -2 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} -4 \\ 0 \end{pmatrix} = \begin{pmatrix} -4 \\ -4 \end{pmatrix},$$

$$\overline{t_0} = \min\left\{-\frac{u_j}{v_j}\middle| v_j < 0\right\} = \min\left\{-\frac{2}{-4}, -\frac{2}{-4}\right\} = \frac{1}{2},$$

求

$$
\begin{aligned}
&\min_{0\le t\le \overline{t_0}} f(\boldsymbol{X}^0 + t\boldsymbol{P}^0)\\
=&\min_{0\le t\le \frac{1}{2}} f\left(\left(2+t(-4), 0+t(0)\right)^T\right)\\
=&\min_{0\le t\le \frac{1}{2}} (2-4t)^2+5
\end{aligned}
$$

得

$$
\begin{aligned}
t_0 &= \frac{1}{2}\\
\boldsymbol{X}^1 &= \boldsymbol{X}^0 + t_0\boldsymbol{P}^0 = (0,0)^T\\
\boldsymbol{g}^1 &= (0,2)^T
\end{aligned}
$$

第二次迭代：

(1)求$\boldsymbol{P}^1$

$$\boldsymbol{N}_1=\begin{pmatrix}1&-2\\1&0\\0&1\end{pmatrix},\boldsymbol{N}_1=\begin{pmatrix}1&-2\\0&1\end{pmatrix},x_1\geq 2x_2,x_2\geq 0,$$

$$\boldsymbol{M}_1=(\boldsymbol{N}_1\boldsymbol{N}_1^T)^{-1}=\begin{pmatrix}5&-2\\-2&1\end{pmatrix}^{-1}=\begin{pmatrix}1&2\\2&5\end{pmatrix},$$

$$\begin{aligned}\mathbb{Q}_1&=\boldsymbol{I}-\boldsymbol{N}_1^T\boldsymbol{M}_1\boldsymbol{N}_1\\&=\begin{pmatrix}1&0\\0&1\end{pmatrix}-\begin{pmatrix}1&0\\-2&1\end{pmatrix}\begin{pmatrix}1&2\\2&5\end{pmatrix}\begin{pmatrix}1&-2\\0&1\end{pmatrix}\end{aligned}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\boldsymbol{P}^1 = -\mathbb{Q}_1 \boldsymbol{g}^1 = (0,0)^T,$$

$$\|\boldsymbol{P}^1\| = 0,$$

$$\boldsymbol{N}_1 \neq \varnothing;$$

(2)修正$\boldsymbol{P}^1$

$$\boldsymbol{q}_1 = \boldsymbol{M}_1 \boldsymbol{N}_1 \boldsymbol{g}^1 = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

因为$\boldsymbol{q}_1 > 0$，所以$\boldsymbol{X}^1$为$\boldsymbol{KT}$点；又因为问题为凸规划，因此最优解$\boldsymbol{X}^\star = (0,0)^T$。 ▮

练习 5.16

$$
\begin{aligned}
&\min \ f(\boldsymbol{X}) = x_1^2 + x_2^2 + 6x_2 \\
s.t.& \\
&2x_1 + x_2 \geq 4 \\
&x_j \geq 0, \quad j = 1, 2.
\end{aligned}
\tag{5-6-104}
$$

初始点$\boldsymbol{X}^0 = (2,0)^T$。

第六章 直接搜索方法

若最优化问题具有目标函数不可导，或没有显式解析表达式等特征，那么经典优化方法难以求解。针对此类问题，人们提出了一种直接搜索法($Direct\ Search$)，该方法仅仅利用问题的目标函数值信息，因此该方法又称免导数方法($Derivative-free$)。

§ 6.1 步长加速法

针对无约束优化问题

$$\min_{\boldsymbol{X}\in R^n} f(\boldsymbol{X}), \tag{6-1-1}$$

1961年*Hooke*和*Jeeves*提出步长加速法，又称之为模式搜索法(*Pattern Search*)。

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$\boldsymbol{X}^0 \in \boldsymbol{R}^n$，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

$$\boldsymbol{e}^k \triangleq (0, \cdots, 0, 1_k, 0, \cdots, 0)^T, k = 1 \sim n. \tag{6-1-2}$$

显然，$\boldsymbol{e}^k, k = 1 \sim n$，是$n$个互相正交的方向。

用$\boldsymbol{Y}$表示参考点，$\boldsymbol{Z}$表示基点。

令$\boldsymbol{Y}^1 = \boldsymbol{X}^0$，$\boldsymbol{Z}^1 = \boldsymbol{Z}^0 = \boldsymbol{X}^0$，

以$\boldsymbol{Y}^1$为参考点，令$\boldsymbol{L}^k \triangleq \boldsymbol{Y}^k - \delta \boldsymbol{e}^k$，$\boldsymbol{U}^k \triangleq \boldsymbol{Y}^k + \delta \boldsymbol{e}^k$，

则沿n个方向进行如下探测

$$\boldsymbol{Y}^{k+1}=\begin{cases}\boldsymbol{Y}^k+\delta\boldsymbol{e}^k, & f(\boldsymbol{U}^k)<f(\boldsymbol{Y}^k);\\ \boldsymbol{Y}^k-\delta\boldsymbol{e}^k, & f(\boldsymbol{L}^k)<f(\boldsymbol{Y}^k)\le f(\boldsymbol{U}^k);\\ \boldsymbol{Y}^k, & f(\boldsymbol{L}^k)\ge f(\boldsymbol{Y}^k)\le f(\boldsymbol{U}^k);\end{cases}\tag{6-1-3}$$

若$f(\boldsymbol{Y}^1)\ge f(\boldsymbol{Z}^1)$，$\delta=\alpha\delta$，从$\boldsymbol{Y}^1$开始，重新寻找基点。

若$f(\boldsymbol{Y}^1)<f(\boldsymbol{Z}^1)$，则令新的基点为$\boldsymbol{Y}^{n+1}$，这时可考虑沿方向$\boldsymbol{Z}^1-\boldsymbol{Y}^1$进行搜索，以期望能够得到更好的下降，即

$$\boldsymbol{Y}^1=\boldsymbol{Z}^1+\beta(\boldsymbol{Z}^1-\boldsymbol{Y}^1)=2\boldsymbol{Z}^1-\boldsymbol{Y}^1(\beta=1).\tag{6-1-4}$$

6.1.3 *Hooke-Jeeves*步长加速法

见算法6.1。

算法 6.1 *Hooke-Jeeves*步长加速算法

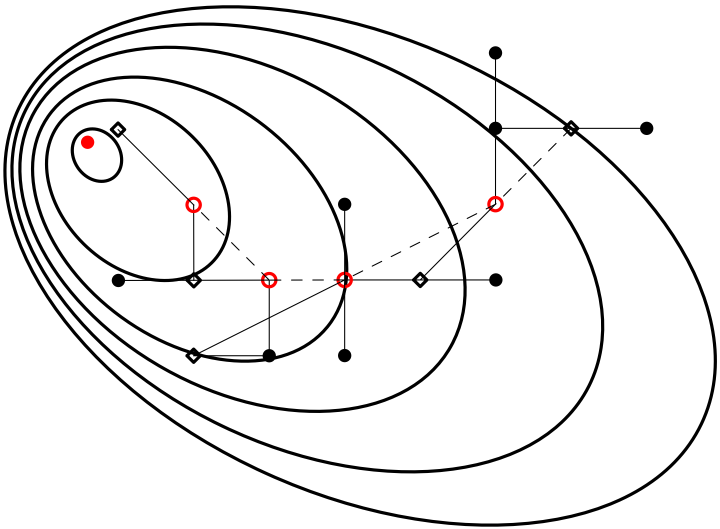
步骤 1 选定$\boldsymbol{Y}^1 = \boldsymbol{X}^0$，$\boldsymbol{Z}^1 = \boldsymbol{Z}^0 = \boldsymbol{X}^0$，$\boldsymbol{L}^k \triangleq \boldsymbol{Y}^k - \delta \boldsymbol{e}^k$，$\boldsymbol{U}^k \triangleq \boldsymbol{Y}^k + \delta \boldsymbol{e}^k$，$k = 1$。

步骤 2 若$f(\boldsymbol{U}^k) < f(\boldsymbol{Y}^k)$，则令$\boldsymbol{Y}^{k+1} = \boldsymbol{Y}^k + \delta \boldsymbol{e}^k$，转步骤(3)；否则若$f(\boldsymbol{L}^k) < f(\boldsymbol{Y}^k)$，则令$\boldsymbol{Y}^{k+1} = \boldsymbol{Y}^k - \delta \boldsymbol{e}^k$，转步骤(3)；若$f(\boldsymbol{L}^k) \geq f(\boldsymbol{Y}^k)$，则令$\boldsymbol{Y}^{k+1} = \boldsymbol{Y}^k$，转步骤(3)。

步骤 3 若$k < n$，则令$k = k+1$，转步骤(2)；否则$k = n$，若$f(\boldsymbol{Y}^{n+1}) < f(\boldsymbol{Z}^1)$，则令$\boldsymbol{Z}^1 = \boldsymbol{Y}^{n+1}$，转步骤(5)；否则若$f(\boldsymbol{Y}^{n+1}) \geq f(\boldsymbol{Z}^1)$，$\boldsymbol{Z}^0 = \boldsymbol{Z}^1$，再若$\delta \leq \varepsilon$，则停止，$\boldsymbol{X}^\star = \boldsymbol{Z}^1$；否则令$\delta = \alpha\delta$，$\boldsymbol{Y}^1 = \boldsymbol{Z}^1$，$k = 1$，转步骤(2)。

步骤 4 若$f(\boldsymbol{Y}^{n+1}) \geq f(\boldsymbol{Z}^1)$，$\boldsymbol{Z}^0 \neq \boldsymbol{Z}^1$，则令$\boldsymbol{Y}^1 = \boldsymbol{Z}^1$，$\boldsymbol{Z}^0 = \boldsymbol{Z}^1$，$k = 1$，转步骤(2)。

步骤 5 令$\boldsymbol{Y}^1 = 2\boldsymbol{Z}^1 - \boldsymbol{Y}^1$，$k = 1$，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

$$f(\boldsymbol{X})=\frac{1}{2}\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X}+\boldsymbol{b}^T\boldsymbol{X}+c, \tag{6-2-5}$$

其基本思想是在不使用导数的情况下，在迭代过程中逐步构造至多n个关于$\boldsymbol{Q}$共轭的方向组。

6.2.1 基本算法

见算法6.2。

算法 6.2 ***Powell***方向加速算法

步骤 1 选定$\boldsymbol{X}^0$，置$\boldsymbol{P}^i=\boldsymbol{e}_{i+1}$，$i=0\sim n-1$。

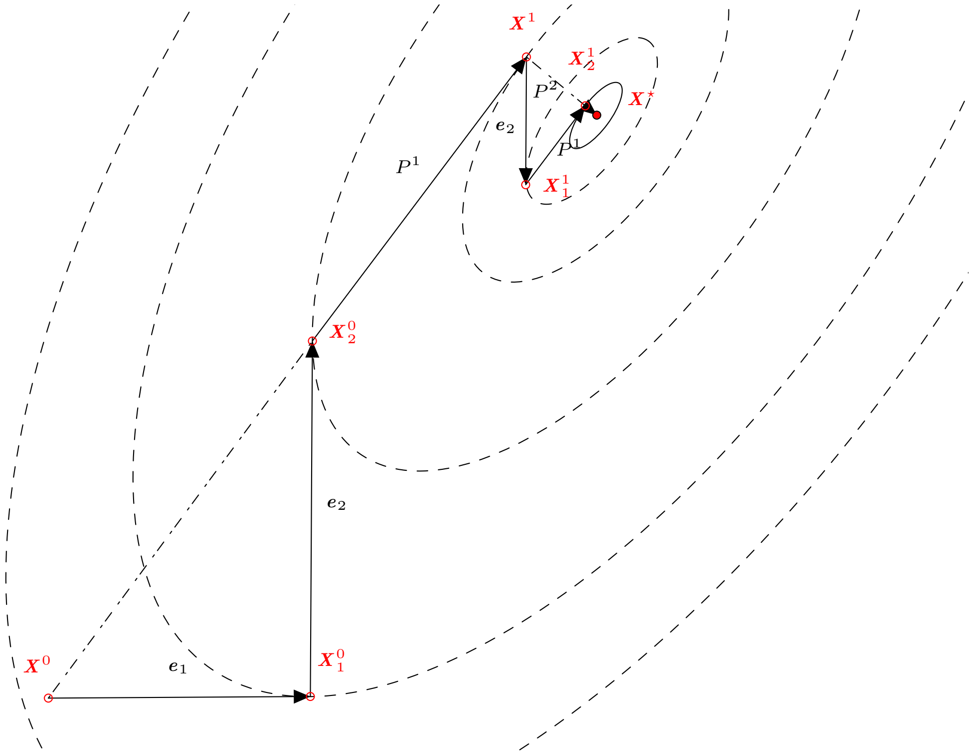
步骤 2 对$\forall i=0\sim n-1$，依次对目标函数作直线搜索$\min f(\boldsymbol{X}^i+t\boldsymbol{P}^i)$。

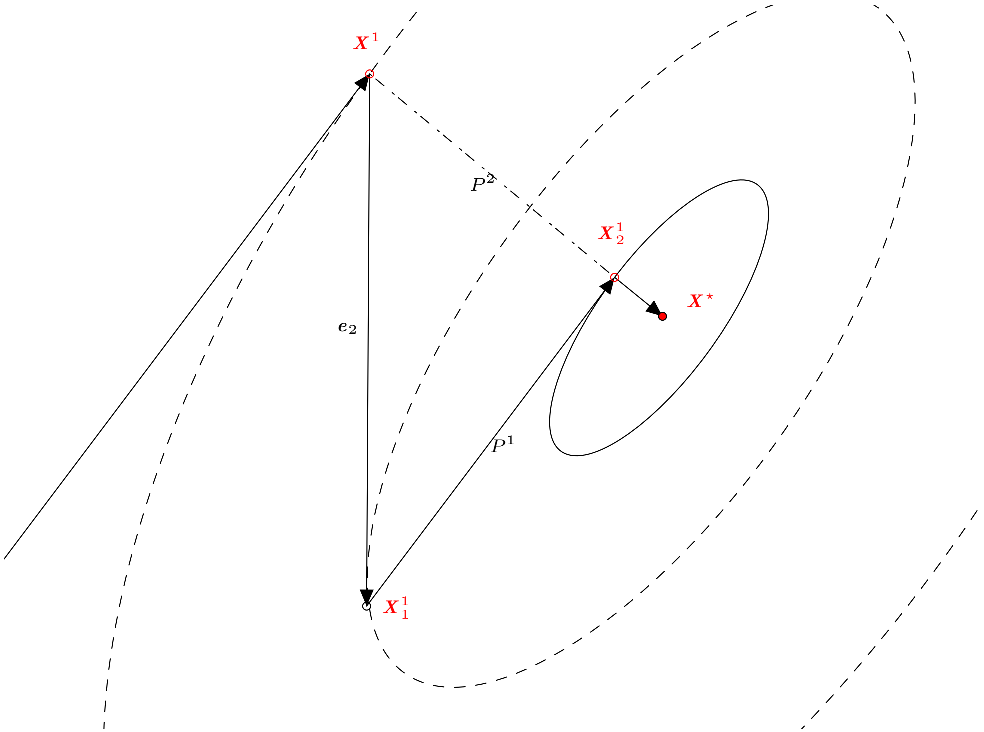
步骤 3 对$\forall i=0\sim n-2$，置$\boldsymbol{P}^i=\boldsymbol{P}^{i+1}$，$\boldsymbol{P}^{n-1}=\boldsymbol{X}^n-\boldsymbol{X}^0$。

步骤 4 一维搜索$\min\ f(\boldsymbol{X}^n+t\boldsymbol{P}^{n-1})$。

步骤 5 若$\|\boldsymbol{X}^{n+1}-\boldsymbol{X}^0\|<\varepsilon$，则$\boldsymbol{X}^{n+1}$就是最优解，停机。否则转步骤(6)。

步骤 6 置$\boldsymbol{X}^0=\boldsymbol{X}^{n+1}$。转步骤(2)。


X^1
X_2^1
P^2
$X^\star$
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
$X^\star$
e_2
P^1
X_1^1

定理 6.1 对于正定二次函数$f(\boldsymbol{X}) = \frac{1}{2}\boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{b}^T\boldsymbol{X} + c$，设$\boldsymbol{P}^0, \boldsymbol{P}^1, \cdots, \boldsymbol{P}^{m-1}$关于$\boldsymbol{Q}$共轭，$m < n$，分别以$\boldsymbol{X}^1$，$\boldsymbol{X}^2$为初始点依次沿$\boldsymbol{P}^i$，$i = 0, \cdots, m-1$，进行一维搜索得到$\boldsymbol{X}_1^\star$，$\boldsymbol{X}_2^\star$，令$\boldsymbol{P}^m = \boldsymbol{X}_1^\star - \boldsymbol{X}_2^\star$，则

$$(\boldsymbol{P}^m)^T\boldsymbol{Q}\boldsymbol{P}^j = 0, \quad j = 0, \cdots, m-1. \tag{6-2-6}$$

证明 由定理4.18易得。 ∎

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。

例 6.1

$$\min \ f(\boldsymbol{X}) = x_1^2 + x_2^2 + x_3^2,$$

这里，$\boldsymbol{X}^0 = (1,2,3)^T$。

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$\boldsymbol{V}$是向量空间，$\boldsymbol{L} \subseteq \boldsymbol{V}$，$\boldsymbol{L} \neq \emptyset$,

若$\exists v \in \boldsymbol{V}$，使得

$$\boldsymbol{L} + v = \{l + v \mid l \in \boldsymbol{L}\} \tag{6-2-7}$$

是$\boldsymbol{V}$的子空间，那么$\boldsymbol{L}$是一个线性流形，并且$dim(\boldsymbol{L}) = dim(\boldsymbol{L} + v)$。若$dim(\boldsymbol{L}) = dim(\boldsymbol{V}) - 1$，则$\boldsymbol{L}$为超平面。

例 6.2

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

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

解 $\boldsymbol{e}^1 = (1, 0, 0)^T$

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

$\boldsymbol{e}^2 = (0, 1, 0)^T$

$\boldsymbol{X}^2 = (\frac{1}{2}, \frac{1}{3}, \frac{1}{2})^T$

$\boldsymbol{e}^3 = (0, 0, 1)^T$

$\boldsymbol{X}^3 = (\frac{1}{2}, \frac{1}{3}, \frac{5}{18})^T$

$\boldsymbol{P}^4 = \boldsymbol{X}^3 - \boldsymbol{X}^0 = (0, -\frac{2}{3}, -\frac{2}{9})^T$

因此在随后的迭代中，永远也不会收敛到问题的最优解$\boldsymbol{X}^\star = (0,0,0)^T$，往往得到维数小于3的线性流形上的极小点。 ∎

失败原因：$\boldsymbol{e}^2$，$\boldsymbol{e}^3$，$\boldsymbol{P}^4$线性相关。**处理技巧：**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

$$\Delta = |\boldsymbol{P}^1 \times \boldsymbol{P}^2| = |det(\boldsymbol{P}^1, \boldsymbol{P}^2)|, \tag{6-2-8}$$

$$\Delta = |(\boldsymbol{P}^1 \times \boldsymbol{P}^2) \cdot \boldsymbol{P}^3| = |det(\boldsymbol{P}^1, \boldsymbol{P}^2, \boldsymbol{P}^3)|. \tag{6-2-9}$$

定义 6.2 设$\boldsymbol{P}^1$，$\cdots$，$\boldsymbol{P}^n$是$\boldsymbol{R}^n$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

$$\begin{aligned}\Delta &= \left|det\left(\frac{\boldsymbol{P}^1}{\|\boldsymbol{P}^1\|},\cdots,\frac{\boldsymbol{P}^n}{\|\boldsymbol{P}^n\|}\right)\right| \\ &= \frac{|det(\boldsymbol{P}^1,\cdots,\boldsymbol{P}^n)|}{\prod_{i=1}^n \|\boldsymbol{P}^i\|}\end{aligned} \tag{6-2-10}$$

这样衡量$\boldsymbol{P}^1$，$\cdots$，$\boldsymbol{P}^n$的正交程度是合理的，因为

定理 6.2 设$\boldsymbol{P}^1$，$\cdots$，$\boldsymbol{P}^n$是$\boldsymbol{R}^n$空间的n个向量，则其正交程度

$$\Delta(\boldsymbol{P}^1,\cdots,\boldsymbol{P}^n) \leq 1. \tag{6-2-11}$$

证明 令$\boldsymbol{B}=(\boldsymbol{P}^1,\cdots,\boldsymbol{P}^n)$，显然当$\boldsymbol{P}^i$线性相关时，$det(\boldsymbol{B})=0$，所以$\Delta=0$。

当$\boldsymbol{P}^i$线性无关时，不妨假设

$$\|\boldsymbol{P}^i\| = 1, i = 1 \sim n. \tag{6-2-12}$$

因为

$$\boldsymbol{A} = \boldsymbol{B}^T\boldsymbol{B} = \left((\boldsymbol{P}^i)^T\boldsymbol{P}^j\right)_{n\times n} = \begin{pmatrix} (\boldsymbol{P}^1)^T\boldsymbol{P}^1 & \cdots & (\boldsymbol{P}^1)^T\boldsymbol{P}^n \\ \vdots & \vdots & \vdots \\ (\boldsymbol{P}^n)^T\boldsymbol{P}^1 & \cdots & (\boldsymbol{P}^n)^T\boldsymbol{P}^n \end{pmatrix}, \tag{6-2-13}$$

所以

$$det^2(\boldsymbol{B}) = det(\boldsymbol{A}). \tag{6-2-14}$$

设$\boldsymbol{A}$的特征根为λ_1，$\cdots$，λ_n，因此

$$det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i, \tag{6-2-15}$$

$$trace(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i = n. \tag{6-2-16}$$

所以

$$\sqrt[n]{\prod_{i=1}^{n} \lambda_i} \leq \frac{1}{n} \sum_{i=1}^{n} \lambda_i. \tag{6-2-17}$$

当且仅当$\lambda_i = \lambda_j$时上式等号成立。因而

$$det^2(\boldsymbol{B}) = det(\boldsymbol{A}) \leq 1. \tag{6-2-18}$$

即

$$\Delta = \Delta(\boldsymbol{B}) \leq 1. \tag{6-2-19}$$

当且仅当$\boldsymbol{P}^i$为正交时上式等号成立。 ∎

作如下变换

$$\boldsymbol{P}^i = \sqrt{\boldsymbol{Q}}\boldsymbol{d}^i, i = 1\cdots, n. \tag{6-2-20}$$

则$\boldsymbol{P}^1, \cdots, \boldsymbol{P}^n$的正交程度就反映了$\boldsymbol{d}^1, \cdots, \boldsymbol{d}^n$的关于$\boldsymbol{Q}$共轭程度。

定义 6.3 设$\boldsymbol{Q}$是$n\times n$的正定矩阵，$\boldsymbol{d}^1$，$\cdots$，$\boldsymbol{d}^n$是$\boldsymbol{R}^n$空间的n个

向量，若其中有零向量，则定义其关于$\boldsymbol{Q}$共轭程度Δ为0；否则，定义为

$$\Delta = \frac{|det(\sqrt{\boldsymbol{Q}})| \cdot |det(\boldsymbol{P}^1, \cdots, \boldsymbol{P}^n)|}{\sqrt{\prod_{i=1}^{n} (\boldsymbol{P}^i)^T \boldsymbol{Q} \boldsymbol{P}^i}}. \tag{6-2-21}$$

6.2.3 *Powell*改进算法

6.2.4 基本算法

见算法6.3。

算法 6.3 改进的$\boldsymbol{Powell}$方向加速算法

步骤 1 选定$\boldsymbol{X}^0$，线性无关向量组$\boldsymbol{P}^i$，$i = 0 \sim n-1$，$\varepsilon > 0$。

步骤 2 对$\forall i = 0 \sim n-1$，依次一维搜索，得到$\boldsymbol{X}^n$。

步骤 3 找出下降最多的方向$\boldsymbol{P}^m$，并计算下降量μ。

步骤 4 以$\boldsymbol{X}^n$为起点，得到$\boldsymbol{X}^{n+1} = \boldsymbol{X}^n + t^\star \boldsymbol{u}\left(\boldsymbol{u} \triangleq \frac{\boldsymbol{X}^n - \boldsymbol{X}^0}{\|\boldsymbol{X}^n - \boldsymbol{X}^0\|}\right)$。

步骤 5

if $\|\boldsymbol{X}^{n+1} - \boldsymbol{X}^0\| \leq \varepsilon$ **then** 输出$\boldsymbol{X}^{n+1}$，停机。**end if**

步骤 6

if $\left|1 + \frac{t^\star}{\|\boldsymbol{X}^n - \boldsymbol{X}^0\|}\right| > \sqrt{\frac{f(\boldsymbol{X}^0) - f(\boldsymbol{X}^{n+1})}{\mu}}$ **then**

$$\boldsymbol{P}^i = \begin{cases} \boldsymbol{P}^{i+1} & i = m \sim n-2; \\ \boldsymbol{u} & i = n-1. \end{cases}$$

end if

步骤 7 置$\boldsymbol{X}^0 = \boldsymbol{X}^{n+1}$。转步骤(2)。

第二部分 应用篇

索引

这里列出了重要定义、定理、模型以及算法的索引。

习 题

1、 填空题

(1)、 函数$f(\boldsymbol{X}) = e^{x_1x_2}\sin x_2$在$\boldsymbol{X} = (0,\pi/2)^T$处的最速下降方向为__________，$\boldsymbol{X}$在该方向上的方向导数为__________.

(2)、 若$f(\boldsymbol{X}) = \boldsymbol{X}^T\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{b}^T\boldsymbol{X} + c$，且$\boldsymbol{Q} = \boldsymbol{Q}^T$，$\boldsymbol{Q} > 0$，$\boldsymbol{b} \in \boldsymbol{R}^n$，$c \in \boldsymbol{R}$，则$\nabla f(\boldsymbol{X}) =$ __________；若在$\boldsymbol{X}^k$处沿下降方向$\boldsymbol{P}^k$进行精确一维搜索，那么$\nabla f(\boldsymbol{X}^{k+1})^T\boldsymbol{P}^k =$ __________，最优步长$t_k =$ __________；又若某种算法能够在有限步之内找到此$f(\boldsymbol{X})$的最优解$\boldsymbol{X}^\star$，则称此算法具有__________性.

(3)、 设$\boldsymbol{A}, \boldsymbol{B}$为凸集，则在$\boldsymbol{A}\bigcup\boldsymbol{B}, \boldsymbol{A}\bigcap\boldsymbol{B}, \boldsymbol{A}+\boldsymbol{B}$中，不是凸集的是__________.

(4)、 线性规划$\min\ f(\boldsymbol{X});\ s.t.\ 2x_1 - x_2 = 1; x_1 + x_3 = 1; x_i \geq 0, i = 1 \sim 3.$的可行区域的一个顶点为__________.

(5)、 在两阶段（单纯形）法中，若辅助线性规划的最优值不为零，那么原线性规划的最优解情况为__________.

(6)、 若用三点二次插值法求解$\min_{x\geq 0}\ f(x)=x^3-3x$，初始三个点分别取$t_1=0$、$t_2=1$、$t_3=2$，则第一次迭代后$t_1,t_2,t_3$分别为__________.

(7)、 设$f(x)=(x-3/2)^2$，若用黄金分割法求解此问题，设初始搜索区间为$[0,2]$，则第一次迭代后得到的搜索区间为__________.

(8)、 设$\{\boldsymbol{X}^k\}=k^{-k}$，$\{\boldsymbol{Y}^k\}=a^{2^k}$，$0<a<1$，分别为$\boldsymbol{A},\boldsymbol{B}$两种算法得到的迭代点列，则收敛速度最快的为算法__________.

2、 已知用单纯形方法求解某一线性规划的初始单纯形表和最终单纯形表，试求$a\sim l$.

表 6.1 初始单纯形表

P_1	P_2	P_3	P_4	P_5	b
b	c	d	1	0	6
-1	3	e	0	1	1
a	-1	-3	0	0	

最终单纯形表

P_1	P_2	P_3	P_4	P_5	b
g	2	-1	1/2	0	f
h	i	1	1/2	1	4
0	-7	j	k	l	

3、 已知线性规划

$$\max \ c_1x_1 + c_2x_2 + c_3x_3$$

$s.t.$

$$x_1 + 2x_2 + x_3 \le b_1$$

$$2x_1 + x_2 + 3x_3 \le 2b_2$$

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

用单纯形法求得最终单纯形表如下，试求所有参数.

表 6.2 单纯形表

P_1	P_2	P_3	P_4	P_5	b
$1/5$	1	0	$3/5$	$-1/5$	1
$3/5$	0	1	$-1/5$	$2/5$	3
$-7/10$	0	0	$-3/5$	$-4/5$	

4、 已知线性规划

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

$$s.t.$$

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

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

$$x_1 \geq 0, \ j = 1 \sim 3.$$

的最优解为$\boldsymbol{X}^\star = (6, 2, 0)^T$，试用两种方法求其对偶规划的最优解.

5、 已知下面线性规划的最优解为$\boldsymbol{X}^\star = (2, 2, 4, 0)^T$ ，试求其对偶规划的最优解。

$$\max \ 2x_1 + 4x_2 + x_3 + x_4$$

$$s.t.$$

$$x_1 + 3x_2 + x_4 \leq 8$$

$$2x_1 + x_2 \leq 6$$

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

$$x_1 + x_2 + x_3 \leq 9$$

$$x_i \geq 0, i = 1 \sim 4.$$

6、 试叙述惩罚函数法的基本思想及其优缺点；并用外部惩罚函数法求解下面的优化问题：

$$\min \ {x_1}^2 + 2{x_2}^2$$

$$s.t.$$

$$x_1 + x_2 \geq 1.$$

7、 试用$Rosen$梯度投影法求解下面的优化问题，以$\boldsymbol{X}^0 = (1,2)^T$为初始迭代点，

$$\min\ f(\boldsymbol{X}) = (4 - x_2)(x_1 - 3)^2$$

$s.t.$

$$x_1 + x_2 \leq 3$$

$$x_1 \leq 2$$

$$x_2 \leq 2$$

$$x_i \geq 0, \quad i = 1, 2.$$

答 案

1、 填空题

(1)、 最速下降方向为$\left(-\frac{\pi}{2},0\right)^T$，方向导数为$-\frac{\pi}{2}$.

(2)、 $\nabla f(X)=2QX+b$；$\nabla f(X^{k+1})^TP^k=0$；$t_k=\frac{-\nabla f(X^k)^TP^k}{2(P^k)^TQP^k}$；具有二次终止性.

(3)、 $A\cup B$.

(4)、 $(1,1,0)^T$或$\left(\frac{1}{2},0,\frac{1}{2}\right)^T$.

(5)、 无可行解或无最优解.

(6)、 $t_1=\frac{5}{6},t_2=1,t_3=2$.

(7)、 $[3-\sqrt{5},2]$.

(8)、 B.

2、 **解** $a=-3,b=2,c=4,d=-2,e=2,f=3,g=1,h=0,i=5,j=-5,k=$

$-3/2, l = 0.$ ∎

3、 **解** $g = 1, h = 0.$ ∎

参 考 文 献

[1] 傅英定. 最优化理论与方法. 国防工业出版社, 2008.

[2] 陈宝林. 最优化理论与方法. 北京：清华大学出版社, 1999.

[3] 袁亚湘, 孙文瑜. 最优化理论与方法. 北京：科学出版社, 2002.

[4] Jorge Nocedal, Stephen J., Wright. Numerical optimization. *Springer*, 2006.

[5] Horst, Pardalos, Thoai. Introduction to global optimization. *Kluwer*, 2000.