

似。在这个点附近取  $f(\mathbf{X})$  的二阶泰勒多项式逼近:

$$f(\mathbf{X}) \approx f(\mathbf{X}^{(k)}) + \nabla f(\mathbf{X}^{(k)})^T \Delta \mathbf{X} + \frac{1}{2} \Delta \mathbf{X}^T \nabla^2 f(\mathbf{X}^{(k)}) \Delta \mathbf{X} \quad (6.43)$$

其中,  $\Delta \mathbf{X} = \mathbf{X} - \mathbf{X}^{(k)}$ 。

这个近似函数的极小点应满足一阶必要条件,即

$$\nabla f(\mathbf{X}^{(k)}) + \nabla^2 f(\mathbf{X}^{(k)}) \Delta \mathbf{X} = \mathbf{0}$$

设  $\nabla^2 f(\mathbf{X}^{(k)})$  的逆阵存在,可得

$$\mathbf{X} = \mathbf{X}^{(k)} - [\nabla^2 f(\mathbf{X}^{(k)})]^{-1} \nabla f(\mathbf{X}^{(k)}) \quad (6.44)$$

由于式(6.43)仅是  $f(\mathbf{X})$  的近似表达式,由式(6.44)解得的该近似函数的极小点,也就仅是  $f(\mathbf{X})$  极小点的近似。为求得  $f(\mathbf{X})$  的极小点,可以  $-[\nabla^2 f(\mathbf{X}^{(k)})]^{-1} \nabla f(\mathbf{X}^{(k)})$  为搜索方向(牛顿方向),按下述公式进行迭代:

$$\begin{cases} \mathbf{P}^{(k)} = -[\nabla^2 f(\mathbf{X}^{(k)})]^{-1} \nabla f(\mathbf{X}^{(k)}) \\ \lambda_k: \min_{\lambda} f(\mathbf{X}^{(k)} + \lambda \mathbf{P}^{(k)}) \\ \mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \lambda_k \mathbf{P}^{(k)} \end{cases} \quad (6.45)$$

这就是所谓的阻尼牛顿法(广义牛顿法),可用于求解非正定二次函数的极小点。

牛顿法的优点是收敛速度快;缺点是有时进行不下去而需采取改进措施(可参看有关参考文献)。此外,当维数较高时,计算  $[\nabla^2 f(\mathbf{X}^{(k)})]^{-1}$  的工作量很大。

为克服梯度法收敛速度慢及牛顿法有时失效和在维数较高时计算工作量大的缺点,不少学者提出了一些更加实用的其他算法,如共轭梯度法、变尺度法等,限于篇幅此处从略,读者可参阅有关文献。

## 第四节 约束极值问题

绝大部分实际问题都受到某些条件的限制,这些限制条件(约束)常给寻优工作带来很大困难。下面首先说明约束极值问题解的最优性条件,然后研究几种基本的解法。

### 一、最优性条件

#### 1. 可行下降方向

##### (1) 起作用约束

假定  $\mathbf{X}^{(0)}$  是问题(6.3)的一个可行解,它满足所有约束条件。对某一个约束条件  $g_j(\mathbf{X}) \geq 0$  来说,  $\mathbf{X}^{(0)}$  满足它有两种情况:一种情况  $g_j(\mathbf{X}^{(0)}) > 0$ , 这时,  $\mathbf{X}^{(0)}$  不在由这个约束条件形成的可行域边界上,我们称这一约束为  $\mathbf{X}^{(0)}$  点的不起作用约束(或无效约束);另一种情况是  $g_j(\mathbf{X}^{(0)}) = 0$ , 这时,  $\mathbf{X}^{(0)}$  点处于由这个约束条件形成的可行域边界上,对  $\mathbf{X}^{(0)}$  点的进一步摄动来说,这一约束起到了某种限制作用,故称它为  $\mathbf{X}^{(0)}$  点的起作用约束(或有效约束)。显然,等式约束条件对所有可行点都是起作用约束。

## (2) 可行方向

设  $\mathbf{X}^{(0)}$  为任一可行点, 对某一方向  $\mathbf{P}$  来说, 若存在实数  $\lambda_0 > 0$ , 使对任意的  $\lambda \in [0, \lambda_0]$  均有下式成立:

$$\mathbf{X}^{(0)} + \lambda \mathbf{P} \in R$$

就称方向  $\mathbf{P}$  为  $\mathbf{X}^{(0)}$  点的一个可行方向。

以  $J$  记  $\mathbf{X}^{(0)}$  点所有起作用约束下标的集合, 即

$$J = \{j \mid g_j(\mathbf{X}^{(0)}) = 0, \quad 1 \leq j \leq l\}$$

显然, 如果  $\mathbf{P}$  为  $\mathbf{X}^{(0)}$  点的可行方向, 则存在  $\lambda_0 > 0$ , 使对任意  $\lambda \in [0, \lambda_0]$ , 有

$$g_j(\mathbf{X}^{(0)} + \lambda \mathbf{P}) \geq g_j(\mathbf{X}^{(0)}) = 0 \quad j \in J$$

从而

$$\left. \frac{dg_j(\mathbf{X}^{(0)} + \lambda \mathbf{P})}{d\lambda} \right|_{\lambda=0} = \nabla g_j(\mathbf{X}^{(0)})^T \mathbf{P} \geq 0 \quad j \in J$$

式中  $\nabla g_j(\mathbf{X}^{(0)})$  为约束函数  $g_j(\mathbf{X})$  在  $\mathbf{X}^{(0)}$  点的梯度。

另外, 由泰勒公式,

$$g_j(\mathbf{X}^{(0)} + \lambda \mathbf{P}) = g_j(\mathbf{X}^{(0)}) + \lambda \nabla g_j(\mathbf{X}^{(0)})^T \mathbf{P} + o(\lambda)$$

对  $\mathbf{X}^{(0)}$  点的所有起作用约束, 当  $\lambda > 0$  足够小时, 只要

$$\nabla g_j(\mathbf{X}^{(0)})^T \mathbf{P} > 0 \quad j \in J \quad (6.46)$$

就有

$$g_j(\mathbf{X}^{(0)} + \lambda \mathbf{P}) \geq 0 \quad j \in J$$

此外, 对  $\mathbf{X}^{(0)}$  点的所有不起作用约束,  $g_j(\mathbf{X}^{(0)}) > 0$ , 由  $g_j(\mathbf{X})$  的连续性, 当  $\lambda > 0$  足够小时, 亦有

$$g_j(\mathbf{X}^{(0)} + \lambda \mathbf{P}) \geq 0 \quad j \notin J$$

从而, 只要方向  $\mathbf{P}$  满足式(6.46), 即可保证它为  $\mathbf{X}^{(0)}$  点的可行方向。

## (3) 下降方向

设  $\mathbf{X}^{(0)} \in R$ , 对某一方向  $\mathbf{P}$  来说, 若存在实数  $\lambda'_0 > 0$ , 使对任意的  $\lambda \in [0, \lambda'_0]$  均有下式成立:

$$f(\mathbf{X}^{(0)} + \lambda \mathbf{P}) < f(\mathbf{X}^{(0)})$$

就称方向  $\mathbf{P}$  为  $\mathbf{X}^{(0)}$  点的一个下降方向。

由泰勒展开式

$$f(\mathbf{X}^{(0)} + \lambda \mathbf{P}) = f(\mathbf{X}^{(0)}) + \lambda \nabla f(\mathbf{X}^{(0)})^T \mathbf{P} + o(\lambda)$$

当  $\lambda$  足够小时, 只要

$$\nabla f(\mathbf{X}^{(0)})^T \mathbf{P} < 0 \quad (6.47)$$

就有  $f(\mathbf{X}^{(0)} + \lambda \mathbf{P}) < f(\mathbf{X}^{(0)})$ 。这说明, 只要方向  $\mathbf{P}$  满足式(6.47), 即可保证它为  $\mathbf{X}^{(0)}$  点的下降方向。

## (4) 可行下降方向

若  $\mathbf{X}^{(0)}$  点的某一方向  $\mathbf{P}$ , 既是该点的可行方向, 又是该点的下降方向, 就称它为这个点的可行下降方向。设  $\mathbf{X}^{(0)}$  不是极小点, 为求其极小点, 继续搜索时当然应沿该点的可行下降方向进行。显然, 对某一点  $\mathbf{X}^*$  来说, 若该点不存在可行下降方向, 它就可能是局部极小点; 若存在可行下降方向, 它当然就不是极小点。下面的定理 4 从另一角度说明了这一问题。

**定理 4** 设  $\mathbf{X}^*$  是问题(6.3)的一个局部极小点,  $f(\mathbf{X})$  在  $\mathbf{X}^*$  处可微, 而且

$g_j(\mathbf{X})$  在  $\mathbf{X}^*$  处可微, 当  $j \in J$  时

$g_j(\mathbf{X})$  在  $\mathbf{X}^*$  处连续, 当  $j \notin J$  时

则在  $\mathbf{X}^*$  点不存在可行下降方向, 从而不存在  $\mathbf{P}$  同时满足

$$\begin{cases} \nabla f(\mathbf{X}^*)^T \mathbf{P} < 0 \\ \nabla g_j(\mathbf{X}^*)^T \mathbf{P} > 0 \quad j \in J \end{cases} \quad (6.48)$$

其中, 指标集  $J = \{j | g_j(\mathbf{X}^*) = 0, 1 \leq j \leq l\}$ 。

这个定理是显然的。若存在满足式(6.48)的方向  $\mathbf{P}$ , 则沿  $\mathbf{P}$  搜索可找到比  $\mathbf{X}^*$  更好的可行点, 这与  $\mathbf{X}^*$  为局部极小点矛盾。

从几何上说, 满足式(6.48)的方向  $\mathbf{P}$ , 与该点目标函数负梯度方向的夹角成锐角, 而且, 与该点起作用约束梯度方向的夹角也成锐角。

## 2. 库恩-塔克(Kuhn-Tucker)条件

库恩-塔克条件是非线性规划领域中最重要理论成果之一, 具有很重要的理论价值。下面就来导出这个条件。

(1) **Gordan 引理** 设  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_l$  是  $l$  个  $n$  维向量, 不存在向量  $\mathbf{P}$  使

$$\mathbf{A}_j^T \mathbf{P} < 0 \quad (j = 1, 2, \dots, l)$$

成立的充要条件是, 存在不全为零的非负实数  $\mu_1, \mu_2, \dots, \mu_l$ , 使

$$\sum_{j=1}^l \mu_j \mathbf{A}_j = \mathbf{0}$$

本式的几何意义说明如下。设  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$  是三个二维向量, 若它们均位于某条直线  $H$  的同一侧, 则可找到某一向量  $\mathbf{P}$ , 使  $\mathbf{A}_j^T \mathbf{P} < 0 (j = 1, 2, 3)$  (参看图 6-11(a))。但是, 若  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$  不在任一条直线的同一侧 (见图 6-11(b)), 就无法找到使  $\mathbf{A}_j^T \mathbf{P} < 0 (j = 1, 2, 3)$  均满足的向量  $\mathbf{P}$ 。这时总可以适当缩小或放大各向量  $\mathbf{A}_j$  的长度, 使它们合成为零向量, 即可找到不全为零的非负实数  $\mu_j (j = 1, 2, 3)$ , 使

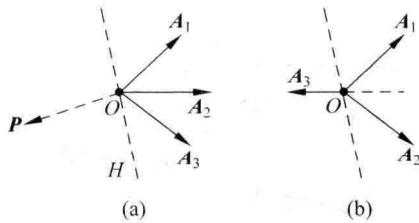


图 6-11

$$\sum_{j=1}^3 \mu_j \mathbf{A}_j = 0$$

## (2) Fritz John 定理

设  $\mathbf{X}^*$  是非线性规划(6.3)的局部最优点, 函数  $f(\mathbf{X})$  和  $g_j(\mathbf{X}) (j=1, 2, \dots, l)$  在  $\mathbf{X}^*$  点有连续一阶偏导数, 则必存在不全为零的数  $\mu_0, \mu_1, \mu_2, \dots, \mu_l$ , 使

$$\begin{cases} \mu_0 \nabla f(\mathbf{X}^*) - \sum_{j=1}^l \mu_j \nabla g_j(\mathbf{X}^*) = 0 \\ \mu_j g_j(\mathbf{X}^*) = 0 \quad (j = 1, 2, \dots, l) \\ \mu_j \geq 0 \quad (j = 1, 2, \dots, l) \end{cases} \quad (6.49)$$

证明从略, 读者可参看有关著作。

该定理给出了非线性规划(6.3)的(局部)最优点应满足的必要条件。式(6.49)称为 Fritz John 条件, 满足这个条件的点称为 Fritz John 点。

如果  $\mu_0 = 0$ ,  $\nabla f(\mathbf{X}^*)$  就从式(6.54)中消去, 说明在所讨论的点  $\mathbf{X}^*$  处, 起作用约束的梯度线性相关。这时 Fritz John 条件失效。人们自然想到, 对讨论点处起作用约束的梯度附加上线性无关的条件, 以保证  $\mu_0 > 0$ 。这样一来, 就引出了下面的库恩-塔克条件。

## (3) 库恩-塔克条件

设  $\mathbf{X}^*$  是非线性规划(6.3)的局部极小点,  $f(\mathbf{X})$  和  $g_j(\mathbf{X}) (j=1, 2, \dots, l)$  在点  $\mathbf{X}^*$  处有一阶连续偏导数, 而且  $\mathbf{X}^*$  处的所有起作用约束的梯度线性无关, 则存在数  $\mu_1^*, \mu_2^*, \dots, \mu_l^*$ , 使

$$\begin{cases} \nabla f(\mathbf{X}^*) - \sum_{j=1}^l \mu_j^* \nabla g_j(\mathbf{X}^*) = 0 \\ \mu_j^* g_j(\mathbf{X}^*) = 0 \quad (j = 1, 2, \dots, l) \\ \mu_j^* \geq 0 \quad (j = 1, 2, \dots, l) \end{cases} \quad (6.50)$$

因由 Fritz John 定理, 存在不全为零的非负数  $\mu_0, \mu_1, \dots, \mu_l$ , 使式(6.49)中的各式成立。又因  $\nabla g_j(\mathbf{X}^*) (j \in J)$  线性无关, 故  $\mu_0 \neq 0$ , 即  $\mu_0 > 0$ 。现用  $\mu_0$  除式(6.49)中各式的两边, 并令  $\mu_j^* = \mu_j / \mu_0, j=1, 2, \dots, l$ , 即可得条件(6.50)。

条件(6.50)称为库恩-塔克条件(简称为 K-T 条件), 满足这个条件的点称为库恩-塔克点或 K-T 点。

现在考虑非线性规划

$$\begin{cases} \min f(\mathbf{X}) \\ h_i(\mathbf{X}) = 0 \quad (i = 1, 2, \dots, m) \\ g_j(\mathbf{X}) \geq 0 \quad (j = 1, 2, \dots, l) \end{cases} \quad (6.51)$$

其中, 函数  $f(\mathbf{X}), h_i(\mathbf{X}) (i=1, 2, \dots, m)$  和  $g_j(\mathbf{X}) (j=1, 2, \dots, l)$  都具有有一阶连续偏导数。

对每一个  $i$ , 我们以

$$\begin{cases} h_i(\mathbf{X}) \geq 0 \\ -h_i(\mathbf{X}) \geq 0 \end{cases}$$

代替  $h_i(\mathbf{X})=0$ , 这样一来, 即可由条件(6.50)得到问题(6.51)的库恩-塔克条件如下:

若  $\mathbf{X}^*$  是非线性规划(6.51)的极小点, 且  $\mathbf{X}^*$  点的所有起作用约束的梯度  $\nabla h_i(\mathbf{X}^*)$  ( $i=1, 2, \dots, m$ ) 和  $\nabla g_j(\mathbf{X}^*)$  ( $j \in J$ ) 线性无关, 则存在向量  $\boldsymbol{\Gamma}^* = (\gamma_1^*, \gamma_2^*, \dots, \gamma_m^*)^\top$  和  $\mathbf{M}^* = (\mu_1^*, \mu_2^*, \dots, \mu_l^*)^\top$  使下述条件成立:

$$\begin{cases} \nabla f(\mathbf{X}^*) - \sum_{i=1}^m \gamma_i^* \nabla h_i(\mathbf{X}^*) - \sum_{j=1}^l \mu_j^* \nabla g_j(\mathbf{X}^*) = 0 \\ \mu_j^* g_j(\mathbf{X}^*) = 0 \quad (j=1, 2, \dots, l) \\ \mu_j^* \geq 0 \quad (j=1, 2, \dots, l) \end{cases} \quad (6.52)$$

其中  $\gamma_1^*, \gamma_2^*, \dots, \gamma_m^*$  和  $\mu_1^*, \mu_2^*, \dots, \mu_l^*$  称为广义拉格朗日(Lagrange)乘子。

库恩-塔克条件是确定某点为最优点的必要条件, 只要是最优点, 且此处起作用约束的梯度线性无关, 就必须满足这个条件。但一般说来它并不是充分条件, 因而, 满足这个条件的点不一定就是最优点。可是, 对于凸规划, 库恩-塔克条件不但是最优点存在的必要条件, 它同时也是充分条件。

为加深对库恩-塔克条件的直观理解, 考虑某非线性规划的可行解  $\mathbf{X}^{(k)}$ , 假定此处有两个起作用约束, 即  $g_1(\mathbf{X}^{(k)})=0, g_2(\mathbf{X}^{(k)})=0$ 。若  $\mathbf{X}^{(k)}$  是极小点, 则  $\nabla f(\mathbf{X}^{(k)})$  必处于  $\nabla g_1(\mathbf{X}^{(k)})$  和  $\nabla g_2(\mathbf{X}^{(k)})$  的夹角之内(见图 6-12)。如若不然,  $\mathbf{X}^{(k)}$  点处必存在可行下降方向, 它就不会是极小点。这说明, 若  $\mathbf{X}^{(k)}$  点是极小点, 而且此处起作用约束的梯度  $\nabla g_1(\mathbf{X}^{(k)})$  和  $\nabla g_2(\mathbf{X}^{(k)})$  线性无关, 则可将  $\nabla f(\mathbf{X}^{(k)})$  表示成  $\nabla g_1(\mathbf{X}^{(k)})$  和  $\nabla g_2(\mathbf{X}^{(k)})$  的非负线性组合, 即存在数  $\mu_1 \geq 0$  和  $\mu_2 \geq 0$ , 使下式成立

$$\nabla f(\mathbf{X}^{(k)}) - \mu_1 \nabla g_1(\mathbf{X}^{(k)}) - \mu_2 \nabla g_2(\mathbf{X}^{(k)}) = 0$$

如此推论, 并考虑到包含不起作用约束在内的所有约束条件, 也可从另一角度得出著名的库恩-塔克条件。

**例 9** 用库恩-塔克条件解非线性规划

$$\begin{cases} \max f(x) = (x-4)^2 \\ 1 \leq x \leq 6 \end{cases}$$

**解** 先将其变为问题(6.2)或问题(6.3)的形式:

$$\begin{cases} \min \bar{f}(x) = -(x-4)^2 \\ g_1(x) = x-1 \geq 0 \\ g_2(x) = 6-x \geq 0 \end{cases}$$

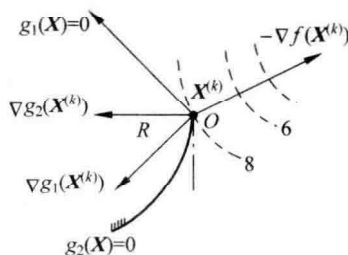


图 6-12

设库恩-塔克点为  $x^*$ , 各函数的梯度为

$$\nabla f(x) = -2(x-4), \quad \nabla g_1(x) = 1, \quad \nabla g_2(x) = -1$$

对第一个和第二个约束条件分别引入广义拉格朗日乘子  $\mu_1^*$  和  $\mu_2^*$ , 则得该问题的库恩-塔克条件如下:

$$\begin{cases} -2(x^* - 4) - \mu_1^* + \mu_2^* = 0 \\ \mu_1^*(x^* - 1) = 0 \\ \mu_2^*(6 - x^*) = 0 \\ \mu_1^* \geq 0, \mu_2^* \geq 0 \end{cases}$$

为解该方程组, 需分别考虑以下几种情况:

- (1)  $\mu_1^* > 0, \mu_2^* > 0$ : 无解;
- (2)  $\mu_1^* > 0, \mu_2^* = 0$ :  $x^* = 1, f(x^*) = 9$ ;
- (3)  $\mu_1^* = 0, \mu_2^* = 0$ :  $x^* = 4, f(x^*) = 0$ ;
- (4)  $\mu_1^* = 0, \mu_2^* > 0$ :  $x^* = 6, f(x^*) = 4$ 。

对应于上述(2)、(3)和(4)三种情形, 我们得到了三个库恩-塔克点, 其中  $x^* = 1$  和  $x^* = 6$  为极大点, 而  $x^* = 1$  为最大点, 最大值  $f(x^*) = 9$  (参看图 6-13);  $x^* = 4$  为可行域的内点, 它不是该问题的极大点, 而是极小点。

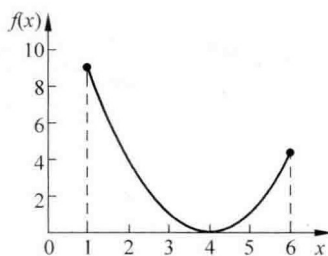


图 6-13

## 二、制约函数法

制约函数法是通过构造某种制约函数, 并将它加到非线性规划的目标函数上, 从而将原来的约束极值问题, 转化为无约束极值问题来求解。由于这里介绍的方法需要求解一系列无约束问题, 故称为序列无约束极小化技术 (sequential unconstrained minimization technique, SUMT)。下面简要介绍其中最基本的两种: 罚函数法 (也称外点法) 和障碍函数法 (也称内点法)。

### 1. 罚函数法 (外点法)

考虑非线性规划 (6.3)。

构造函数  $\psi(t)$ :

$$\psi(t) = \begin{cases} 0 & \text{当 } t \geq 0 \\ \infty & \text{当 } t < 0 \end{cases} \quad (6.53)$$

现把某一约束函数  $g_j(\mathbf{X})$  视为  $t$ , 显然, 当  $\mathbf{X}$  满足该约束时,  $g_j(\mathbf{X}) \geq 0$ , 从而  $\psi(g_j(\mathbf{X})) = 0$ ; 当  $\mathbf{X}$  不满足该约束条件时,  $\psi(g_j(\mathbf{X})) = \infty$ 。将各个约束条件的上述函数加到式 (6.3) 的目标函数上, 得一新函数如下:

$$\varphi(\mathbf{X}) = f(\mathbf{X}) + \sum_{j=1}^l \psi(g_j(\mathbf{X})) \quad (6.54)$$