最优化方法

（下 册）

张运模　杜文瑞　编

海军工程学院数学教研室

一九八四年十月

目　　录

第四章　非线性规划……（1）
§1　最优性条件……（1）
§2　二次规划……（16）
2.1　Wolfe方法……（16）
2.2　Hildreth—D'Espo 方法……（22）
§3　可行方向法……（25）
§4　梯度投影法……（33）
§5　既约梯度法……（39）
5.1　既约梯度法……（39）
5.2　广义既约梯度法（GRG方法）……（43）
§6　罚函数法（SUMT）……（48）
6.1　外点法……（48）
6.2　内点法……（52）
§7　乘子法……（55）
§8　网格法……（61）
习题四……（62）
第五章　几何规划……（65）
§1　几何不等式及其推广……（65）
§2　正项几何规划……（68）
§3　符号几何规划……（87）
§4　可变换为几何规划的问题举例……（91）
习题五……（98）
第六章　动态规划……（99）
§1　动态规划的基本概念和基本方程……（99）
§2　函数迭代法和策略迭代法……（105）
2.1　函数迭代法……（106）
2.2　策略迭代法……（108）
§3　资源分配问题和降维方法……（112）
§4　复合系统工作可靠性问题……（123）
§5　生产计划问题……（124）
§6　最优控制问题……（127）
习题六……（131）
习题答案……（132）
参考文献……（136）

第四章　非线性规划

上一章我们讨论了无约束的非线性规划，本章讨论有约束非线性规划，以下简称非线性规划。非线性规划问题是实际中最常见的最优化数学模型，正因为如此，研究的人多，提出的求解方法也很多。这里我们先对非线性规划的理论略作介绍，然后主要介绍一些常用的求解方法。

§1　最优性条件

本节我们分等式约束和不等式约束两种情况介绍约束极值的必要条件和充分条件。

1．等式约束下的极值

考虑非线性规划问题

$$(P)\begin{cases}\min f(x) & (1.1)\\ \quad h_i(x)=0,\ i=1,\cdots,m & (1.2)\end{cases}$$

假设 $f(x)$，$h_i(x)$（$i=1,\cdots,m$）有二阶连续偏导数。$x\in E_n$。

大家知道，求解等式约束极值问题的常用方法是拉格朗日乘子法，这种方法是基于下面的定理。

定理1　等式约束极值的必要条件　如果 x^* 是（P）的局部极小点，并且在 x^* 处，梯度向量 $\nabla h_i(x^*)$（$i=1,\cdots,m$）线性无关，则存在向量 $\lambda^*=(\lambda_1^*,\cdots,\lambda_m^*)^T$ 使得

$$\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla h_i(x^*)=0 \qquad (1.3)$$

证明从略。

引进拉格朗日函数

$$L(x,\lambda)=f(x)+\sum_{i=1}^{m}\lambda_i h_i(x)$$

于是（1.3）可简记为

$$\nabla_x L(x^*,\lambda^*)=0$$

由此可知，拉格朗日乘子法的基本思想是把等式约束极值问题转化为无约束极值问题求解。下面给出等式约束极值的充分条件。

定理2　等式约束极值的充分条件　（P）的可行解 x^* 为严格局部极小点的充分条件是存在 $\lambda^*=(\lambda_1^*,\cdots,\lambda_m^*)^T$ 使得

$$\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla h(x^*)=0 \qquad (1.3)$$

且对集合$M=\{y \mid \nabla h_i(x^*)^T y=0,\ i=1,\cdots,m\}$中的任何非零向量$y$均有

$$y^T[\nabla^2 f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla^2 h_i(x^*)]y>0 \tag{1.4}$$

证 用反证法。

若x^*不是(P)的严格局部极小点，则存在点列$\{x^k\}$：

$$x^k=x^*+\delta_k s^k$$

其中 $\delta_k\to 0$(当$k\to+\infty$时)，$\|s^k\|=1$，

满足 $h_i(x^k)=0,\ i=1,\cdots,m$

且 $f(x^k)\leqslant f(x^*)$

由于序列$\{s^k\}$有界，故必存在收敛的子序列，为了方便，不妨设$s^k\to s$。由台劳公式，得

$$\begin{aligned}0&=h_i(x^k)-h_i(x^*)\\&=\delta_k\nabla h_i(x^*)^T s^k+\frac{\delta_k^2}{2}s^{kT}\nabla^2 h_i(x^*)s^k+0(\delta_k^2)\end{aligned} \tag{1.5}$$

$$i=1,\cdots,m$$

$$\begin{aligned}0&\geqslant f(x^k)-f(x^*)\\&=\delta_k\nabla f(x^*)^T s^k+\frac{\delta_k^2}{2}s^{kT}\nabla^2 f(x^*)s^k+0(\delta_k^2)\end{aligned} \tag{1.6}$$

分别用λ_i^*($i=1,\cdots,m$)乘(1.5)中第i式，然后一起加到(1.6)上，即得

$$0\geqslant\delta_k\left[\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla h_i(x^*)\right]^T s^k$$

$$+\frac{\delta_k^2}{2}s^{kT}\left[\nabla^2 f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla^2 h_i(x^*)\right]s^k+0(\delta_k^2)$$

由条件(1.3)，从而即得

$$s^{kT}\left[\nabla^2 f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla^2 h_i(x^*)\right]s^k+0(1)\leqslant 0$$

当$k\to+\infty$时，$\delta_k\to 0$，$s^k\to s$，于是得

$$s^T\left[\nabla^2 f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla^2 h_i(x^*)\right]s\leqslant 0 \tag{1.7}$$

又由(1.5)得

$$\nabla h_i(x^*)^T s^k+\frac{\delta_k}{2}s^{kT}\nabla^2 h_i(x^*)s^k+0(\delta_k)=0,$$

$$i=1,\cdots,m$$

当$k\to+\infty$时，即得

$$\nabla h_i(x^*)^T s=0 \qquad i=1,\cdots,m$$

因此 $s \in M$，且 $s \neq 0$，这样（1.7）就与（1.4）矛盾，所以 x^* 为严格局部极小点。证毕。

为便于检验条件（1.4），可取子空间 M 中的一个基 $e_1, \cdots, e_{n-m}$，令

$$E=(e_1, \cdots, e_{n-m}),$$

E 为 $n\times(n-m)$ 阶矩阵，由于 $y\in M$ 等价于 $y=Ez$，$z\in E_{n-m}$，于是（1.4）等价于

$$E^T\left[\nabla^2 f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla^2 h_i(x^*)\right]E>0\text{（正定）} \qquad (1.8)$$

从定理的证明可知，如果（1.8）式反向为

$$E^T\left[\nabla^2 f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla^2 h_i(x^*)\right]E<0\text{（负定）}$$

其余条件不变，则 x^* 为严格局部极大点。

例1 求 $f(x_1, x_2, x_3)=x_1x_2+x_2x_3+x_3x_1$ 在约束条件 $x_1+x_2+x_3=3$ 下的极值点。

解 由约束条件及必要条件得

$$\begin{aligned}x_1+x_2+x_3&=3\\x_2+x_3+\lambda&=0\\x_1+x_3+\lambda&=0\\x_1+x_2+\lambda&=0\end{aligned}$$

解之，得 $\quad x_1^*=x_2^*=x_3^*=1$，$\lambda^*=-2$

这样 $x^*=(1, 1, 1)^T$ 可能是极值点，下面检验是否满足充分条件。因为

$$\nabla h(x^*)=(1, 1, 1)^T,$$

于是

$$M=\{y \mid \nabla h(x^*)^T y=0\}$$

即有 $\quad y_1+y_2+y_3=0$

因此 M 是三维空间中的一个平面，即是一个二维子空间，取基底为

$$e_1=(1, 0, -1)^T,$$
$$e_2=(1, -1, 0)^T$$

令

$$E=\begin{bmatrix}1 & 1\\0 & -1\\-1 & 0\end{bmatrix}$$

又

$$\nabla^2 f(x^*)=\begin{bmatrix}0 & 1 & 1\\1 & 0 & 1\\1 & 1 & 0\end{bmatrix}$$

$$\nabla^2 h(x^*)=0$$

于是

$$E^{T}[\nabla^2 f(x^*)+\lambda^*\nabla^2 h(x^*)]E$$

$$=\begin{bmatrix}1 & 0 & -1\\ 1 & -1 & 0\end{bmatrix}\begin{bmatrix}0 & 1 & 1\\ 1 & 0 & 1\\ 1 & 1 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 0 & -1\\ -1 & 0\end{bmatrix}$$

$$=\begin{bmatrix}-2 & -1\\ -1 & -2\end{bmatrix}<0\text{（负定）}$$

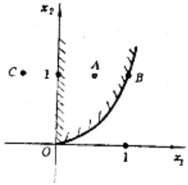

图4.1

所以$x^*=(1,1,1)^T$为严格极大点。

2. 有不等式约束的极值

考虑非线性规划问题

$$(P_1)\quad\begin{cases}\min f(x) & (1.9)\\ g_i(x)\geqslant 0,\ i=1,\cdots,p & (1.10)\end{cases}$$

假定$f(x)$，$g_i(x)$（$i=1,\cdots,p$）有二阶连续偏导数。

可行域　　$R=\{x\mid g_i(x)\geqslant 0,\ i=1,\cdots,p\}$

下面先介绍几个概念。

定义1　如果$x^0\in R$，且使$g_i(x^0)>0$，$i=1,\cdots,p$，则称x^0为R的内点。

如果$x^0\in R$，且至少有某个k（$1\leqslant k\leqslant p$）使$g_k(x^0)=0$，则称x^0为R的边界点。

如果$x^0\bar{\in} R$，即至少有某个k（$1\leqslant k\leqslant p$）使$g_k(x^0)<0$，则称x^0为R的外点。

例如　对可行域（见图4.1）

$$R=\{(x_1,x_2)\mid x_1\geqslant 0,\ x_2-x_1^2\geqslant 0\}$$

$A\left(\frac{1}{2},1\right)$为内点，$B(1,1)$为边界点，$C\left(-\frac{1}{2},1\right)$为外点。

定义2　设$x^0\in R$，若存在$\lambda_0>0$，使$\forall\lambda\in[0,\lambda_0]$均有

$$x^0+\lambda d\in R$$

即　　$$g_i(x^0+\lambda d)\geqslant 0,\quad i=1,\cdots,p$$

则称方向d为x^0处的一个可行方向。

对含有等式约束的问题，也可作出相应的可行方向的定义。

从定义可知，d为x^0处的可行方向即从x^0出发沿方向d有一条短线段整个包含在可行域内。显然，当x^0为R的内点时，任何方向都是可行方向。但当x^0为R的边界点时，情况就不同了。如图4.2，d^1是x^0处的可行方向，d^2就不是x^0处的可行方向。这里约束条件$g_1(x)\geqslant 0$对x^0处的可行方向起了限制作用，而$g_2(x)\geqslant 0$，$g_3(x)\geqslant 0$则不起这种作用，我们称$g_1(x)\geqslant 0$为点x^0的起作用约束。从图可以得知

$$g_1(x^0)=0,$$
$$g_2(x^0)>0,$$
$$g_3(x^0)>0$$

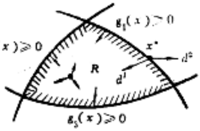

图4.2

于是一般我们定义如下：

如果可行点x^0使$g_i(x^0)=0$，则称约束条件$g_i(x)\geqslant 0$为点x^0的起作用约束。

在研究局部极值时，显然可以把注意力放在可行点的起作用约束上。一般我们用$I(x^0)$表示x^0的起作用约束的下标集（有时简记为I），即

$$I(x^0)=\{i \mid g_i(x^0)=0,\ 1\leqslant i\leqslant p\}$$

现在我们来看看，在怎样的条件下，d是一个可行方向。由台劳公式

$$g_i(x^0+\lambda d)=g_i(x^0)+\lambda\nabla g_i(x^0)^{\mathrm{T}}d+0(\lambda) \quad (1.11)$$

当$i\overline{\in} I(x^0)$时，$g_i(x^0)>0$，故当λ充分小时，都有$g_i(x^0+\lambda d)\geqslant 0$；当$i\in I(x^0)$时，$g_i(x^0)=0$，如果

$$\nabla g_i(x^0)^{\mathrm{T}}d>0$$

那么对充分小的λ就都有$g_i(x^0+\lambda d)\geqslant 0$。因此只要

$$\nabla g_i(x^0)^{\mathrm{T}}d>0,\quad i\in I(x^0) \quad (1.12)$$

d就是x^0处的可行方向。

下面先考虑线性不等式约束问题

$$(P_2)\quad\begin{cases}\min f(x)\\ a_i^{\mathrm{T}}x\geqslant b_i,\quad i=1,\cdots,p\end{cases} \quad (1.13)$$

其中 $a_i^{\mathrm{T}}=(a_{i1},\cdots,a_{in}),\ i=1,\cdots,p,\ x\in E_n$

对于线性约束函数$g_i(x)$，（1.11）中$0(\lambda)=0$，因此，d为x^0处的可行方向的充要条件是

$$\nabla g_i(x^0)^{\mathrm{T}}d\geqslant 0,\quad i\in I(x^0)$$

即
$$a_i^{\mathrm{T}}d\geqslant 0,\qquad i\in I(x^0) \quad (1.14)$$

如果d又满足

$$\nabla f(x^0)^{\mathrm{T}}d<0$$

那么，d又是$f(x)$在x^0处的下降方向，则称d为x^0处的下降可行方向。

下面将给出（P_2）的极小点的必要条件。先介绍一个著名的重要引理。

Farkas引理 已知矩阵A（$m\times n$），向量b（$n\times 1$）

$$(\mathrm{I})\begin{cases}A^{\mathrm{T}}x=b\\ x\geqslant 0\end{cases}$$

$$(\mathrm{II})\begin{cases}Ay\geqslant 0\\ b^{\mathrm{T}}y<0\end{cases}$$

其中$x\in E_m$，$y\in E_n$，则（Ⅰ）有解的充分必要条件是（Ⅱ）无解。

证　显然（Ⅱ）**无解**等价于

$$
(\text{Ⅲ})\begin{cases} Az \leqslant 0 \\ b^{\mathrm{T}}z > 0 \end{cases}
$$

无解。下面我们利用线性规划的对偶定理来证明本引理。

由（Ⅰ）引进人工变量，得到两段单纯形法中阶段Ⅰ的辅助规划（P'）及（P'）的对偶规划（D'）如下：

$$
(P')\quad \begin{cases} \min 1^{\mathrm{T}}t \\ \quad A^{\mathrm{T}}x + It = b \\ \quad x \geqslant 0,\quad t \geqslant 0 \end{cases}
$$

即

$$
\begin{cases} \min[O^{\mathrm{T}},\ 1^{\mathrm{T}}]\begin{bmatrix} x \\ t \end{bmatrix} \\ \quad [A^{\mathrm{T}}\quad I]\begin{bmatrix} x \\ t \end{bmatrix} = b \\ \quad \begin{bmatrix} x \\ t \end{bmatrix} \geqslant 0 \end{cases}
$$

$$
(D')\begin{cases} \max\ b^{\mathrm{T}}z \\ \quad \begin{bmatrix} A \\ I \end{bmatrix} z \leqslant \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{cases}
$$

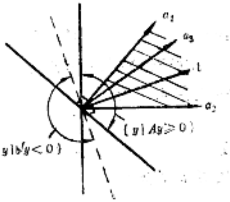

图4.3

即

$$
\begin{cases} \max\ b^{\mathrm{T}}z \\ \quad Az \leqslant 0 \\ \quad z \leqslant 1 \end{cases}
$$

其中 1 是分量都为 1 的 n 维向量，$t \in E_n$，$z \in E_n$。于是

（Ⅰ）有解 $\rightleftharpoons$（P'）的最优值为零

$\rightleftharpoons$（D'）的最优值为零

$\rightleftharpoons$（Ⅲ）无解

$\rightleftharpoons$（Ⅱ）无解

证毕。

Farkas 引理也可用图说明。大家知道，如果两个向量的内积为正，则向量之间成锐角；如果两个向量的内积为负，则向量之间成钝角。见图4.3，其中 a_1，a_2，a_3 表示 A^{T}的列向量。从图中可以看出，如果向量 b 能表为 A^{T}的列向量a_1，a_2，a_3的非负系数的线性组合，则满足$Ay \geqslant 0$ 的 y 必使 $b^{\mathrm{T}}y \geqslant 0$，这就是说，如果（Ⅰ）有解，则（Ⅱ）

无解。

定理3 设x^*为（P_2）的局部极小点，I为x^*的起作用约束的下标集，则存在μ_i^*（$i\in I$），使

$$\nabla f(x^*)-\sum_{i\in I}\mu_i^* a_i=0$$

$$\mu_i^*\geqslant 0,\quad i\in I$$

证 因x^*为局部极小点，故在x^*处不存在$f(x)$的下降可行方向，即

$$\begin{cases}a_i^T d\geqslant 0 & i\in I\\ \nabla f(x^*)^T d<0\end{cases}$$

无解，由Farkas引理

$$\begin{cases}\sum_{i\in I}a_i\mu_i^*=\nabla f(x^*)\\ \mu_i^*\geqslant 0,\quad i\in I\end{cases}$$

有解，于是定理得证。

现在我们考察非线性不等式约束问题，即（P_1）中的$g_i(x)$为非线性函数。自然我们希望定理3能推广到这种情形。也就是说在局部极小点x^*处有

$$\nabla f(x^*)-\sum_{i\in I}\mu_i^*\nabla g_i(x^*)=0 \tag{1.15}$$

$$\mu_i^*\geqslant 0,\quad i\in I \tag{1.16}$$

然而并不总是这样。那么在什么条件下才能保证在局部极小点x^*有上式成立呢？对此，人们进行了大量研究，提出各种各样的条件，通常把这类条件称为约束规范。其中一种较简单的约束规范是在x^*的起作用约束的梯度向量线性无关。

我们把非起作用约束的那些梯度向量也加到（1.15）中去，但对应系数μ_i^*取为零，这样就得到

$$\nabla f(x^*)-\sum_{i=1}^{p}\mu_i^*\nabla g_i(x^*)=0$$

其中当$i\overline{\in}I$时，$\mu_i^*=0$。另一方面，当$i\in I$时，有$g_i(x^*)=0$，所以总有

$$\mu_i^* g_i(x^*)=0$$

结合前面等式约束极值的必要条件（1.3）对一般非线性规划问题

$$(P)\begin{cases}\min f(x)\\ \quad h_i(x)=0,\quad i=1,\cdots,m\\ \quad g_j(x)\geqslant 0,\quad j=1,\cdots,p\end{cases}$$

就有下面的著名定理。

定理4 Kuhn—Tucker定理 设x^*为（P）的局部极小点，如果x^*的所有起作用

约束的梯度向量$\nabla g_i(x^*)$（$i\in I$）和$\nabla h_i(x^*)$（$i=1,\cdots,m$）线性无关，则存在向量$\lambda^*=(\lambda_1^*,\cdots,\lambda_m^*)^T$，$\mu^*=(\mu_1^*,\cdots,\mu_p^*)^T$使

$$\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla h_i(x^*)-\sum_{i=1}^{p}\mu_i^*\nabla g_i(x^*)=0 \tag{1.17}$$

$$\mu_j^* g_j(x^*)=0,\qquad j=1,\cdots,p \tag{1.18}$$

$$\mu_j^*\geqslant 0,\qquad j=1,\cdots,p \tag{1.19}$$

证明从略。

上述极值的必要条件（1.17）（1.18）（1.19)称为Kuhn—Tucker条件，简称K—T条件，满足K—T条件的点称为K—T点。上述定理是1951年由Kuhn和Tucker提出来的。它是带有不等式约束极值问题的重要理论基础，它是非线性最优化理论的重要发展。

引进广义拉格朗日函数

$$\begin{aligned}L(x,\lambda,\mu)&=f(x)+\sum_{i=1}^{m}\lambda_i h_i(x)-\sum_{i=1}^{p}\mu_i g_i(x)\\&=f(x)+\lambda^T h(x)-\mu^T g(x)\end{aligned}$$

K—T条件可简记为

$$\nabla_x L(x^*,\lambda^*,\mu^*)=0$$

$$\mu^{*T}g(x^*)=0$$

$$\mu^*\geqslant 0$$

注意，这里乘子μ_i^*的符号是确定的，它与前面等式约束极值问题的拉格朗日乘子（可正可负）不同。

对于无约束极值，当目标函数$f(x)$有一阶连续偏导数时，极值点一定是稳定点，而稳定点则不一定是极值点。但有一类函数例外，那就是凸函数，凸函数的稳定点不仅是局部极小点，而且是全局最小点。

类似地，对于有约束极值，一般说来，K—T点不一定是极值点。但有一种规划例外，这种规划不仅K—T点是它的局部极小点，而且还是它的全局最小点。这就是凸规划。

定义3 如果$f(x)$是凸函数，$g_i(x)$（$i=1,\cdots,p$）都是凹函数(即$-g_i(x)$为凸函数），则称规划

$$(P)\begin{cases}\min f(x)\\ \quad g_i(x)\geqslant 0,\ i=1,\cdots,p\end{cases}$$

为凸规划。

例如，线性规划就是凸规划。

定理5 凸规划的基本定理 如果x^*是凸规划（P）的K—T点，则x^*是（P）的最优解。

证 先证可行域

$$R=\{x\mid g_i(x)\geqslant 0,\ i=1,\cdots,p\}$$

为凸集。

因为$g_i(x)$为凹函数，故对$\forall x^1, x^2 \in R$，$x^1 \neq x^2$及$0<\alpha<1$，均有

$$g_i(\alpha x^1+(1-\alpha)x^2) \geqslant \alpha g_i(x^1)+(1-\alpha)g_i(x^2) \geqslant 0$$

因此 $\alpha x^1+(1-\alpha)x^2 \in R$

所以R为凸集。

下面再考虑函数

$$F(x)=f(x)-\sum_{i=1}^{p}\mu_i^* g_i(x)$$

其中 $\mu_i^*(i=1, \cdots, p)$为与x^*相对应的乘子。

由于$f(x)$，$-g_i(x)(i=1, \cdots, p)$都是凸函数，$\mu_i^* \geqslant 0(i=1, \cdots, p)$，所以当$x \in R$时，$F(x)$是凸函数。又因$x^*$是$(P)$的K—T点，所以$x^*$是凸函数$F(x)$的稳定点，从而是$F(x)$的最小点。于是，对$\forall x \in R$，都有

$$f(x^*)-\sum_{i=1}^{p}\mu_i^* g_i(x^*) \leqslant f(x)-\sum_{i=1}^{p}\mu_i^* g_i(x)$$

而

$$\mu_i^* g_i(x^*)=0, \qquad 1 \leqslant i \leqslant p$$
$$\mu_i^* g_i(x) \geqslant 0, \qquad 1 \leqslant i \leqslant p$$

所以

$$f(x^*) \leqslant f(x)$$

故x^*为(P)的最优解。证毕。

例2 求解非线性规划问题

$$\begin{cases} \min f(x)=x_1+x_2 \\ \quad g_1(x)=-x_1^2+x_2 \geqslant 0 \\ \quad g_2(x)=x_1 \geqslant 0 \end{cases}$$

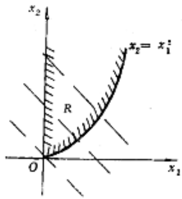

图4.4

解 从图解法不难得知，最优解为

$$x^*=(0, 0)^{\mathrm{T}}$$

下面用解析方法求解。

由于

$$\nabla^2 g_1(x)=\begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}<0$$

故知，$f(x)$，$-g_1(x)$，$-g_2(x)$都是凸函数，所以，这是一个凸规划，它的K—T点即为最优解。下面求K—T点。

$$\begin{cases} 1+2\mu_1 x-\mu_2=0 \\ 1-\mu_1=0 \\ \mu_1(-x_1^2+x_2)=0 \\ \mu_2 x_1=0 \end{cases}$$

解得

$$x_1^* = 0, \qquad x_2^* = 0$$
$$\mu_1^* = 1, \qquad \mu_2^* = 1$$

故最优解为

$$x^* = (0,\ 0)^T$$

对于一般有不等式约束的极值问题，K—T条件只是极小点的必要条件，而非充分条件。下面给出极小点的充分条件。考虑的问题与定理4相同。

定理6 有不等式约束极值的充分条件 （P）的可行解x^*为严格局部极小点的充分条件是存在向量λ^*，μ^*使

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) - \sum_{j=1}^{p} \mu_j^* \nabla g_j(x^*) = 0$$

$$\mu_j^* g_j(x^*) = 0, \quad j = 1, \cdots, p$$
$$\mu_j^* \geqslant 0, \quad j = 1, \cdots, p$$

且对集合$M = \{ y \mid \nabla h_i(x^*)^T y = 0,\ i = 1, \cdots, m;\ \nabla g_j(x^*)^T y = 0,\ j \in I \}$中的任何非零向量$y$均有

$$y^T \left[\nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) - \sum_{j=1}^{p} \mu_j^* \nabla^2 g_j(x^*) \right] y > 0$$

证明与定理2类似，从略。

例3 求解非线性规划问题

$$\begin{cases} \min -x_1 x_2 \\ \quad -x_1 - x_2^2 + 1 \geqslant 0 \\ \quad x_1 + x_2 \geqslant 0 \end{cases}$$

解 先求K—T点。

1）设$g_1(x_1,\ x_2) = -x_1 - x_2^2 + 1 = 0$

$g_2(x_1,\ x_2) = x_1 + x_2 = 0$（都是起作用约束）

由

$$\begin{cases} -x_2 + \mu_1 - \mu_2 = 0 \\ -x_1 + 2\mu_1 x_2 - \mu_2 = 0 \\ -x_1 - x_2^2 + 1 = 0 \\ x_1 + x_2 = 0 \end{cases}$$

解之，得

$$x_2 = \frac{1 \pm \sqrt{5}}{2},$$

$$x_1 = \frac{-1 \mp \sqrt{5}}{2},$$

$$\mu_1=\frac{1\pm\sqrt{5}}{\mp\sqrt{5}}<0。$$

所以当$-x_1-x_2^2+1=0$，$x_1+x_2=0$时，无约束极小。

2）设$g_1(x_1, x_2)=-x_1-x_2^2+1>0$，

$g_2(x_1, x_2)=x_1+x_2=0$（起作用约束）

$$\begin{cases}-x_2-\mu_2=0\\-x_1-\mu_2=0\\x_1+x_2=0\end{cases}$$

解之，得$x_1=x_2=\mu_2=0$，而可行点（ε，ε）（ε为小正数）上的函数值为$-\varepsilon^2<0$，故点$(0,\ 0)^T$并非极小点。

3）再考虑$g_1(x_1, x_2)=-x_1-x_2^2+1=0$（起作用约束），$g_2(x_1, x_2)=x_1+x_2>0$的情况

$$\begin{cases}-x_2+\mu_1=0\\-x_1+2\mu_1x_2=0\\-x_1-x_2^2+1=0\end{cases}$$

解之，得

$$x_1=\frac{2}{3},$$

$$x_2=\frac{1}{\sqrt{3}},$$

$$\mu_1=\frac{1}{\sqrt{3}}$$

所以$\left(\frac{2}{3},\ \frac{1}{\sqrt{3}}\right)^T$是K—T点。

下面再用二阶充分条件来检查此K—T点是否约束极小点，为此，先求

$$\nabla g_1\left(\frac{2}{3},\ \frac{1}{\sqrt{3}}\right)=\begin{bmatrix}-1\\ \frac{-2}{\sqrt{3}}\end{bmatrix}$$

再求

$$M=\left\{\ y\ \middle|\ \nabla g_1\left(\frac{2}{3},\ \frac{1}{\sqrt{3}}\right)^T y=0\ \right\}$$

即

$$\left[-1,\ \frac{-2}{\sqrt{3}}\right]\begin{bmatrix}y_1\\y_2\end{bmatrix}=0$$

即

$$y_1=\frac{-2}{\sqrt{3}}y_2$$

令

$$\nabla^2 f\left(\frac{2}{3},\ \frac{1}{\sqrt{3}}\right)=\begin{bmatrix}0 & -1\\ -1 & 0\end{bmatrix}$$

$$\nabla^2 g_1\left(\frac{2}{3},\ \frac{1}{\sqrt{3}}\right)=\begin{bmatrix}0 & 0\\ 0 & -2\end{bmatrix}$$

于是

$$y^{T}\left[\nabla^2 f\left(\frac{2}{3},\ \frac{1}{\sqrt{3}}\right)-\frac{1}{\sqrt{3}}\nabla^2 g_1\left(\frac{2}{3},\ \frac{1}{\sqrt{3}}\right)\right]y$$

$$=\left(\frac{-2}{\sqrt{3}},\ 1\right)\left(\begin{bmatrix}0 & -1\\ -1 & 0\end{bmatrix}-\frac{1}{\sqrt{3}}\begin{bmatrix}0 & 0\\ 0 & -2\end{bmatrix}\right)\begin{bmatrix}\frac{-2}{\sqrt{3}}\\ 1\end{bmatrix}$$

$$=\frac{6}{\sqrt{3}}>0$$

所以，点$\left(\frac{2}{3},\ \frac{1}{\sqrt{3}}\right)^{T}$是约束极小点。

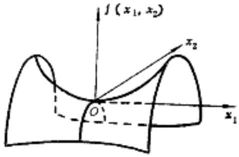

图4.5

前面介绍的最优性条件都要求函数可微，一般的函数不一定满足所要求的可微性，下面我们来介绍一般情况下约束极值的充分条件。这里涉及到函数的鞍点的概念。粗略地说，函数$f(x_1,\ x_2)$的鞍点就是这样的点：关于x_1它是$f(x_1,\ x_2)$的极小点，关于x_2它是$f(x_1,\ x_2)$的极大点。例如点$(0,\ 0)$，就是函数

$$f(x_1,\ x_2)=x_1^2-x_2^2$$

的鞍点(见图4.5)，它满足

$$f(0,\ x_2)\leqslant f(0,\ 0)\leqslant f(x_1,\ 0)$$

下面还是考虑问题

$$(P)\begin{cases}\min f(x)\\ \quad h_i(x)=0,\quad i=1,\ \cdots,\ m\\ \quad g_j(x)\geqslant 0,\quad j=1,\ \cdots,\ p\end{cases}$$

作拉格朗日函数

$$L(x,\ \lambda,\ \mu)=f(x)+\sum_{i=1}^{m}\lambda_i h_i(x)-\sum_{j=1}^{p}\mu_j g_j(x)$$

定义4 设有点$(\bar{x},\ \bar{\lambda},\ \bar{\mu})$，其中$\bar{x}\in E_n$，$\bar{\lambda}\in E_m$，$\bar{\mu}\in E_p$，且$\bar{\mu}\geqslant 0$，如果对任意$x\in E_n$，$\lambda\in E_m$，$\mu\in E_p$，且$\mu\geqslant 0$均有

$$L(\overline{x},\ \lambda,\ \mu)\leqslant L(\overline{x},\ \overline{\lambda},\ \overline{\mu})\leqslant L(x,\ \overline{\lambda},\ \overline{\mu}) \tag{1.20}$$

则称$(\overline{x},\ \overline{\lambda},\ \overline{\mu})$为$L(x,\ \lambda,\ \mu)$的鞍点。

定理7 点$(\overline{x},\ \overline{\lambda},\ \overline{\mu})$$(\overline{\mu}\geqslant 0)$是$(P)$的拉格朗日函数$L(x,\lambda,\mu)$的鞍点的充分必要条件是下列各式成立：

（1）$L(\overline{x},\ \overline{\lambda},\ \overline{\mu})=\min\limits_{x} L(x,\ \overline{\lambda},\ \overline{\mu})$

（2）$h_i(\overline{x})=0,\ i=1,\ \cdots,\ m$

（3）$g_j(\overline{x})\geqslant 0,\ j=1,\ \cdots,\ p$

（4）$\overline{\mu}_j g_j(\overline{x})=0,\ j=1,\ \cdots,\ p$

证 先证必要性。

由鞍点定义，（1）成立。

由鞍点定义，对任意$x\in E_n$，$\lambda\in E_m$，$\mu\in E_p$，$\mu\geqslant 0$均有

$$f(\overline{x})+\sum_{i=1}^{m}\lambda_i h_i(\overline{x})-\sum_{j=1}^{p}\mu_j g_j(\overline{x})$$

$$\leqslant f(\overline{x})+\sum_{i=1}^{m}\overline{\lambda}_i h_i(\overline{x})-\sum_{j=1}^{p}\overline{\mu}_j g_j(\overline{x}) \tag{1.21}$$

从而得

$$\sum_{i=1}^{m}(\lambda_i-\overline{\lambda}_i)h_i(\overline{x})+\sum_{j=1}^{p}(\overline{\mu}_j-\mu_j)g_j(\overline{x})\leqslant 0 \tag{1.22}$$

对于任意的$k$$(1\leqslant k\leqslant m)$，令$\lambda_i=\overline{\lambda}_i$$(i\neq k)$，$\lambda_k=\overline{\lambda}_k+1$，$\mu_j=\overline{\mu}_j$$(1\leqslant j\leqslant p)$，由（1.22）即得$h_k(\overline{x})\leqslant 0$；若在（1.22）中令$\lambda_i=\overline{\lambda}_i$$(i\neq k)$，$\lambda_k=\overline{\lambda}_k-1$，$\mu_j=\overline{\mu}_j$$(1\leqslant j\leqslant p)$，即得$h_k(\overline{x})\geqslant 0$，所以对所有$i$$(1\leqslant i\leqslant m)$，$h_i(\overline{x})=0$，即（2）成立。

由于$h_i(\overline{x})=0,\ i=1,\ \cdots,\ m$，从而由（1.22）得知

$$\sum_{j=1}^{p}(\overline{\mu}_j-\mu_j)g_j(\overline{x})\leqslant 0 \tag{1.23}$$

对任何$\mu=(\mu_1,\ \cdots,\ \mu_p)^T\geqslant 0$成立。下证$g_j(\overline{x})\geqslant 0$$(j=1,\ \cdots,\ p)$，如果对某$k$$(1\leqslant k\leqslant p)$有$g_k(\overline{x})<0$，在（1.23）中令$\mu_j=\overline{\mu}_j$$(j\neq k)$，$\mu_k=\overline{\mu}_k+1$，即得$g_k(\overline{x})\geqslant 0$，矛盾，所以

$$g_j(\overline{x})\geqslant 0,\ j=1,\ \cdots,\ p$$

即（3）成立。

在（1.23）中，令$\mu_j=0,\ j=1,\ \cdots,\ p$，则

$$\sum_{j=1}^{p}\overline{\mu}_j g_j(\overline{x})\leqslant 0$$

又由于$\overline{\mu}_j\geqslant 0$，$g_j(\overline{x})\geqslant 0$，$j=1,\cdots,p$
从而又有

$$\sum_{j=1}^{p}\overline{\mu}_j g_j(\overline{x})\geqslant 0$$

所以

$$\sum_{j=1}^{p}\overline{\mu}_j g_j(\overline{x})=0$$

且

$$\overline{\mu}_j g_j(\overline{x})=0,\quad j=1,\cdots,p$$

即（4）成立。

下证充分性。

由（1）得

$$L(\overline{x},\overline{\lambda},\overline{\mu})\leqslant L(x,\overline{\lambda},\overline{\mu}) \tag{1.24}$$

由$L(x,\lambda,\mu)$的定义及（2）

$$L(\overline{x},\lambda,\mu)=f(\overline{x})-\sum_{i=1}^{m}\lambda_i h_i(\overline{x})-\sum_{j=1}^{p}\mu_j g_j(\overline{x})$$

$$=f(\overline{x})-\sum_{j=1}^{p}\mu_j g_j(\overline{x})$$

由（3），对$\mu\geqslant 0$有

$$\mu_j g_j(\overline{x})\geqslant 0,\quad j=1,\cdots,p$$

从而

$$L(\overline{x},\lambda,\mu)\leqslant f(\overline{x})$$

由（2）及（4）

$$L(\overline{x},\overline{\lambda},\overline{\mu})=f(\overline{x})$$

因此

$$L(\overline{x},\lambda,\mu)\leqslant L(\overline{x},\overline{\lambda},\overline{\mu})$$

结合（1.24），所以（$\overline{x}$，$\overline{\lambda}$，$\overline{\mu}$）为$L(x,\lambda,\mu)$的鞍点。证毕。

定理8 如果($\overline{x}$，$\overline{\lambda}$，$\overline{\mu}$）是（P）的拉格朗日函数$L(x,\lambda,\mu)$的鞍点，则$\overline{x}$是（P）的最优解。

证 由鞍点定义

$$L(\overline{x},\overline{\lambda},\overline{\mu})\leqslant L(x,\overline{\lambda},\overline{\mu})$$

即

$$f(\bar{x})+\sum_{i=1}^{m}\bar{\lambda}_i h_i(\bar{x})-\sum_{j=1}^{p}\bar{\mu}_j g_j(\bar{x})$$

$$\leqslant f(x)+\sum_{i=1}^{m}\bar{\lambda}_i h_i(x)-\sum_{j=1}^{p}\bar{\mu}_j g_j(x)$$

由定理7，

$$h_i(\bar{x})=0,\quad i=1,\cdots,m$$

$$\bar{\mu}_j g_j(\bar{x})=0,\quad j=1,\cdots,p$$

于是得到

$$f(\bar{x})\leqslant f(x)+\sum_{i=1}^{m}\bar{\lambda}_i h_i(x)-\sum_{j=1}^{p}\bar{\mu}_j g_j(x)$$

对（P）的任意可行解x有

$$h_i(x)=0,\quad i=1,\cdots,m$$

$$g_j(x)\geqslant 0,\quad j=1,\cdots,p$$

又因 $\bar{\mu}_j\geqslant 0,\quad j=1,\cdots,p$

所以 $f(\bar{x})\leqslant f(x)$

故$\bar{x}$为（P）的最优解。

上述两个定理的主要特点是不要求函数具有可微性或凸性。有时可用它来解非线性规划问题。

例4 求解非线性规划问题

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

解 拉格朗日函数为

$$L(x,\mu_1,\mu_2)=(x-4)^2-\mu_1(x-1)-\mu_2(6-x)$$

由鞍点条件得

$$x-1\geqslant 0,\qquad 6-x\geqslant 0$$

$$\mu_1(x-1)=0,\ \mu_2(6-x)=0$$

$$\mu_1\geqslant 0,\qquad \mu_2\geqslant 0$$

1）若$\mu_1=\mu_2=0$，解

$$\min_x L(x,0,0)=(x-4)^2$$

得$x=4$，且$g_1(4)>0$，$g_2(4)>0$，故得鞍点（4，0，0）。

2）若$\mu_1=0$，$\mu_2\neq 0$，则$x=6$，对任意$\mu_2>0$，

$$L(6,0,\mu_2)\neq\min_x L(x,0,\mu_2),$$

这是因为 $L(6,0,\mu_2)=4$，