第十章 非线性规划方法

谭　忠

厦门大学数学科学学院

目录

1 源头问题与当今应用

在上一章，我们遇到了规划问题的约束条件和目标函数都是线性的情形，有时也会遇到目标函数或约束条件中包含有非线性函数的情形，这就是**非线性规划问题**.

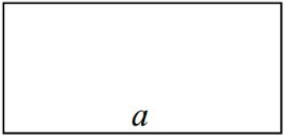
非线性规划是20世纪50年代才形成的一门学科. 1951年H.W. 库恩和A.W. 塔克发表的关于最优性条件即库恩—塔克条件的论文是非线性规划正式诞生的一个重要标志. 在50年代还得出了可分离规划和二次规划的多种解法，它们大都是以G.B. 丹齐克提出的解线性规划的单纯形法为基础的.

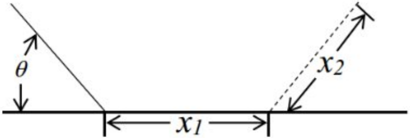
源头问题与当今应用

50年代末到60 年代末出现了许多解非线性规划问题的有效算法，70年代又得到进一步的发展．非线性规划是线性规划的进一步发展和继续．许多实际问题如设计问题、经济平衡问题都属于非线性规划的范畴．

例 10.1有一块薄的塑料板，宽为 a，对称地把两边折起，做成槽，底宽 x_1，槽边 x_2. 欲使槽的横截面积 S 最大，槽边与水平的夹角 θ 的最优值是多少?

源头问题与当今应用





分析： 该问题要找出最优参数底宽 x_1，槽边 x_2，槽边与水平的夹角θ，使槽的横截面积S 最大．所以，目标函数为

$$maxS = (x_1 + x_2 \cos\theta) \cdot x_2 \sin\theta$$

由于底边与两个斜边的总长度应等于塑料板宽度. 因此约束条件为

$$x_1 + 2x_2 = a$$

有许多最优化问题可以方便地将等式约束条件代入目标函数中，使原问题转换为无约束条件的最优化问题，便于求解。于是本例变为无约束条件的最优化问题时，目标函数为

$$maxS = (a - 2x_2 + x_2 \cos\theta) \cdot x_2 \sin\theta.$$

谢　谢!

2 非线性规划思想与建模方法

10.2.1基本概念

所谓**非线性规划**，是指目标函数或约束条件中包含有非线性函数的一类最优化问题。非线性规划问题简记为 NLP .

1、非线性规划的一般形式

$$\min z = f(x) \quad (10.2.1)$$

$s.t.$

$$\begin{cases} h_j(x) = 0, j = 1, 2, \cdots, q \\ g_i(x) \geq 0, i = 1, 2, \cdots, p \end{cases} \qquad (10.2.2)$$

式中

$$x = (x_1, x_2, \cdots, x_N)^T$$

为N维欧氏空间

中的向量点，

T表示转置；

$f(x)$为目标函数；

$h_j(x)$为 q 个等式约束条件；

$g_i(x)$ 为 p个不等式约束条件，

包括资源约束和变量约束.

定义 10.1:（可行域与可行解）

把满足上述非线性规划的

一般形式中的条件(10.2.2)的点

称为可行点(可行解),

所有可行点的集合称为可行域,

即

$$\boldsymbol{K} = \{x = (x_1, x_2, \cdots, x_N)^T |$$

$$g_i(x) \geq 0, i = 1, 2, \cdots, p;$$

$$h_j(x) = 0, j = 1, 2, \cdots, q\}$$

若某个可行解

使目标函数为最小,

即使(10.2.1)成立,

就称它为最优解.

2、非线性规划的标准式

NLP 模型的标准形式有两个要求：

其一，目标函数为极小形式，

若原 NLP 模型为 $\max f(x)$ 形式，

应变为 $\min[-f(x)] = minf'(x)$ ；

其二，约束条件为 $\geq$ 或 $=$ 或 $\leq$ 号.

这里有两点值得注意：

一是，原 NLP 模型的不等式约束

不必变为等式；

二是，$\leq$ 仅需用 “-1” 乘该约束的两端，

即可将这个约束变成 $\geq$ 的形式.

3、非线性规划

模型的特殊类型

(1) 按约束条件 $g_i(x)$ 或 $h_j(x)$ 有与无，

分为有约束 NLP 问题

和无约束 NLP 问题.

(2) 有约束 NLP 问题的目标函数
可以是非线性函数或线性函数
（此类NLP 问题的约束条件中
至少有一个约束条件是非线性函数）.

(3) 无约束的 NLP 问题的目标函数必须是非线性函数。其中，目标函数为单变量的非线性函数，称为 NLP 的一维无约束优化问题，又称一维搜索问题；若目标函数有多个变量，称为 NLP 的多维无约束优化问题，或称多维无约束极值问题.

(4) 二次规划：

有约束的 NLP 问题中，

目标函数为二次函数，

约束条件全为线性函数，称为二次规划.

(5) 几何规划：

NLP 问题中,

目标函数及约束条件的不等式

具有非线性的多元多项式形式,

如

$$\min f(x) = 6 + 3x_1 - 8x_2 + 7x_3 + 2x_1x_2$$

$$-3x_1x_3 - 9x_2^2 + x_3^2$$

$$s.t.$$

$$g(x) = 0.33x_2x_3 + 3x_1^{-0.5}x_2^{-0.75}x_3^{-1}$$

$$+4.5x_2^{0.5}x_3^{-1}x_4^{-0.5} - 1 \geq 0$$

$$x_1, x_2, x_3, x_4 > 0$$

谢　谢!

4、凸规划

(1) 凸集与凸函数的定义

设$\boldsymbol{K}$是$\boldsymbol{N}$维欧氏空间的一点集，

若任意两点$\boldsymbol{x}_1 \in \boldsymbol{K}$，$\boldsymbol{x}_2 \in \boldsymbol{K}$的

连线上的所有点

$$\alpha x_1 + (1-\alpha)x_2 \in K$$

其中$0 \leq \alpha \leq 1$

则称K为凸集。

实心圆，实心球体，实心立方体

等都是凸集。圆环不是凸集。

从直观上讲，

凸集没有凹入部分，

其内部没有空洞。

任何两个凸集

的交集是凸集。

定义 10.2 (凸函数)给定函数$f(x)(x \in D \subset R)$,

若 $\forall x_1, x_2 \in D \lambda \in [0,1]$，有

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2),$$

则称 $f(x)$ 为 D 上的凸函数;

特别地，若

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2),$$

则称 $f(x)$ 为 D 上的严格凸函数.

将上述两式中的不等号反向,

即可得到凹函数和严格凹函数的定义.

显然，若函数 $f(x)$ 是凸函数（严格凸函数），
则 $-f(x)$ 一定是凹函数（严格凹函数）.
凸函数和凹函数的几何意义十分明显，

若函数图形上任两点的连线
处处都不在这个函数图形的下方，
它当然是凸的.
线性函数既可看作凸函数，
也可看作凹函数.

(2)凸函数的性质

性质1：定义在凸集上的有限个凸函数的非负线性组合仍为凸函数.

性质2：设 $f(x)$ 为定义在凸集 K 上的凸函数，则对任一实数 α ，集合 $S_\alpha = \{x|x \in K, f(x) \leq \alpha\}$ 是凸集(S_α称为水平集).

证明：任取 $x_1 \in S_\alpha$ 和 $x_2 \in S_\alpha$，

则有 $f(x_1) \le \alpha, f(x_2) \le \alpha.$

由于 K 为凸集，故对任意实数 $t(0 < t < 1)$,

$tx_1 + (1-t)x_2 \in K,$

又因 $f(x)$ 为凸函数，故

$$f(tx_1+(1-t)x_2)\le tf(x_1)+(1-t)f(x_2)\le \alpha$$

这就表明点 $tx_1+(1-t)x_2\in S\alpha$,

于是，$S\alpha$ 为凸集.

(3) 函数凸性的判定

怎样判断一个函数是凸函数?

首先可以直接依据定义去判别.

对于可微凸函数，也可利用下述两个判别定理.

定理 10.1（一阶条件）

设 K 为 N 维欧氏空间 R^N 上的开凸集，

$f(x)$ 在 K 上具有一阶连续偏导数，

则 $f(x)$ 为 K 上的凸函数的充要条件是：

对任意两个不同点$x_1, x_2 \in K$，恒有

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T \cdot (x_2 - x_1) \quad (10.2.3)$$

证明: 必要性：设$f(x)$为K上的凸函数，

则对任何$t(0 << 1)$有

$$f(tx_2 + (1-t)x_1) \leq tf(x_2) + (1-t)f(x_1)$$

于是

$$\frac{f(x_1+t(x_2-x_1))-f(x_1)}{t} \leq f(x_2) - f(x_1)$$

令$t \to +0$，上式左端的极限为

$$\nabla f(x_1)^T(x_2 - x_1),$$

即

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T \cdot (x_2 - x_1)$$

充分性：任取 $x_1, x_2 \in K$，现令

$$x = tx_1 + (1-t)x_2,\ 0 < t < 1$$

分别以 x_1 和 x_2 为式(10.2.3)中的 x_2，以 x 为式(10.2.3)中的x_1，

则

$$f(x_1) \geq f(x) + \nabla f(x)^T \cdot (x_1 - x)$$

$$f(x_2) \geq f(x) + \nabla f(x)^T \cdot (x_2 - x)$$

用 t 乘上面的第一式，用 $(1-t)$ 乘上面的第二式，然后两端相加:

$$tf(x_1)+(1-t)f(x_2)$$

$$\geq f(x)+\nabla f(x)^T\cdot[tx_1-tx+(1-t)(x_2-x)]$$

$$=f(x)=f(tx_1+(1-t)x_2)$$

从而可知$f(x)$为K上的凸函数.

若式(10.2.3)为严格不等式，它就是严格凸函数的充要条件.

凸函数的定义，本质上是说凸函数上两点间的线性插值不低于这个函数的值；

而定理10.1则是说基于某点导数的线性近似不高于这个函数的值或曲线上各点的切线在曲线之下。

定义10.3(海赛矩阵)

多元函数 $f(x)(x = (x_1, x_2, \cdots, x_N)^T)$

的二阶偏导数组成的矩阵称为其海赛矩阵，记为

$$\nabla^2 f(x) = \begin{pmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_N} \\ \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_N} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial^2 f(x)}{\partial x_n \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f(x)}{\partial x_N^2} \end{pmatrix}$$

定理 10.2 （二阶条件）

设 K 为 N 维欧氏空间 R^N 上的开凸集，

$f(x)$ 在 K 上具有二阶连续偏导数，

则$f(x)$为K上的凸函数的充要条件是：

$f(x)$的海赛矩阵$\nabla^2 f(x^*)$在K上处处半正定.

证明：先证必要性

设 $f(x)$ 为 K 上的凸函数

任取 $x \in K, Z \in R^N$ ，现证

$$Z^T \nabla^2 f(x^*) Z \geq 0$$

因 K 为开集，故存在 $\overline{a} > 0$，

使当 $a \in [-\overline{a}, \overline{a}]$ 时，有 $x + aZ \in K$，

由定理 10.1 可得

$$f(x + aZ) \geq f(x) + a\nabla f(x)^T \cdot Z$$

再由泰勒公式

$$f(x+aZ)=f(x)+a\nabla f(x)^T\cdot Z$$

$$+\tfrac{1}{2}a^2Z^T\nabla^2 f(x^*)Z+o(a^2)$$

其中， $\lim\limits_{t\to 0}\frac{o(t^2)}{t^2}=0$,

由以上两式得

$\frac{1}{2}t^2 Z^T \nabla^2 f(x^*) Z + o(t^2) \geq 0$

从而$\frac{1}{2}Z^T \nabla^2 f(x^*) Z + \frac{o(t^2)}{t^2} \geq 0$。令 $t \to 0$，则得 $Z^T \nabla^2 f(x^*) Z \geq 0$，即 $\nabla^2 f(x^*)$ 为半正定矩阵。

下面证明充分性.

设对任意 $\boldsymbol{x} \in \boldsymbol{K}$， $\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}^*)$ 为半正定矩阵，

任取$\overline{\boldsymbol{x}} \in \boldsymbol{K}$，由泰勒公式，$\boldsymbol{\lambda} \in (\mathbf{0}, \mathbf{1})$，有

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\overline{\boldsymbol{x}}) + \boldsymbol{\nabla} \boldsymbol{f}(\overline{\boldsymbol{x}})^{T} \cdot (\boldsymbol{x} - \overline{\boldsymbol{x}})$$

$$+\frac{1}{2}(\boldsymbol{x} - \overline{\boldsymbol{x}})^{T} \boldsymbol{\nabla}^2 \boldsymbol{f}(\overline{\boldsymbol{x}} + \boldsymbol{\lambda}(\boldsymbol{x} - \overline{\boldsymbol{x}}))(\boldsymbol{x} - \overline{\boldsymbol{x}})$$

因 $\boldsymbol{K}$ 为凸集，$\overline{\boldsymbol{x}} + \boldsymbol{\lambda}(\boldsymbol{x} - \overline{\boldsymbol{x}}) \in \boldsymbol{K}$.

再由假设知 $\boldsymbol{\nabla}^2 \boldsymbol{f}(\overline{\boldsymbol{x}} + \boldsymbol{\lambda}(\boldsymbol{x} - \overline{\boldsymbol{x}}))$ 为半正定，从而

$$\boldsymbol{f}(\boldsymbol{x}) \geq \boldsymbol{f}(\overline{\boldsymbol{x}}) + \boldsymbol{\nabla} \boldsymbol{f}(\overline{\boldsymbol{x}})^{\boldsymbol{T}} \cdot (\boldsymbol{x} - \overline{\boldsymbol{x}})$$

由定理 10.1, $f(x)$ 为 K 上的凸函数.

若对一切 $x \in K$, $f(x)$ 的海赛矩阵都是正定的,

则 $f(x)$ 是 K 上的严格凸函数.

对于凹函数可以得到和上述类似的结果.

例10.2 试证明 $f(x) = -x_1^2 - x_2^2$ 为凹函数.

用三种方法证明：

证法一：由定义证明

$f_1(x_1) = -x_1^2$和$f_2(x_2) = -x_2^2$为凹函数.

根据性质 2 ,$f(X) = -x_1^2 - x_2^2$，为凹函数.

任意指定两点 a_1 和 a_2，看下式是否成立?

$$-[ta_1+(1-t)a_2]^2 \geq t(-a_1^2)+(1-t)(-a_2^2)$$

展开后即看下式是否成立

$$a_1^2(t-t^2)-2a_1a_2(t-t^2)+a_2^2(t-t^2) \geq 0$$

简化后即看下式是否成立

$(t - t^2)(a_1 - a_2)^2 \geq 0$

由于 $0 < t < 1$，

故 $t - t^2 > 0$.

显然，不管 a_1 和 a_2 取什么值，

总有 $(t-t^2)(a_1-a_2)^2 \geq 0$ 成立，

从而证明 $f_1(x_1) = -x_1^2$ 为凹函数。

证法二：再用定理 10.1 证明.

任意选取第一点 $\boldsymbol{X}^1=(\boldsymbol{a}_1,\boldsymbol{b}_1)^T$ ，

第二点 $\boldsymbol{X}^2=(\boldsymbol{a}_2,\boldsymbol{b}_2)^T$。

由此可得,

$$f(X^1) = -a_1^2 - b_1^2$$

$$f(X^2) = -a_2^2 - b_2^2$$

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

$$\nabla f(X^1) = (-2a_1, -2b_1)^T$$

现看下式是否成立?

$$-a_2^2-b_2^2 \leq -a_1^2-b_1^2+(-2a_1-2b_1)\begin{pmatrix} a_2-a_1 \\ b_2-b_1 \end{pmatrix}^T$$

或

$$-a_2^2-b_2^2 \leq -a_1^2-b_1^2-2a_1(a_2-a_1)-2b_1(b_2-b_1)$$

或

$$-(a_2^2 - 2a_1a_2 + a_1^2) - (b_2^2 - 2b_1b_2 + b_1^2) \leq 0$$

或

$$-(a_2 - a_1)^2 - (b_2 - b_1)^2 \leq 0$$

不管a_1、a_2、b_1和b_2取什么值，上式均成立，得证.

证法三：用定理 10.2 证明.

由于$\frac{\partial f(x)}{\partial x_1} = -2x_1$,

$$\frac{\partial f(x)}{\partial x_2} = -2x_2$$

$$\frac{\partial^2 f(x)}{\partial x_1^2} = -2 < 0,$$

$$\frac{\partial^2 f(x)}{\partial x_2^2} = -2 < 0$$

$$\frac{\partial^2 f(x)}{\partial x_1 \partial x_2} = \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} = 0$$

$$|\nabla^2 f(x)| = \begin{vmatrix} -2 & 0 \\ 0 & -2 \end{vmatrix} = 4 > 0$$

其海赛矩阵处处负定，故 $f(x)$ 为（严格）凹函数.

谢　谢!

定义 10.4（NLP 的凸规划）

对非线性规划问题 NLP，

若 $f(x)$ 为 K 上的凸函数,

$g_i(x)$ 为 R^N 上的凹函数

（$-g_i(x)$为R^N 上的凸函数）,

$h_j(x)$ 为 R^N 上的线性函数，

则称 NLP 为凸规划.

可以证明，上述凸规划的可行域为凸集，其局部最优解即全局最优解，

而且其最优解的集合形成一个凸集。

当凸规划的目标函数$f(X)$为严格凸函数时，其最优解必定唯一（假定最优解存在）.由此可见，凸规划是一类比较简单而又具有重要理论意义的非线性规划. 由于线性函数既可视为凸函数，又可视为凹函数，故线性规划也属于凸规划.

例题：试分析非线性规划

$$\begin{cases} minf(x) = x_1^2 + x_2^2 - 4x_1 + 4 \\ g_1(x) = x_1 - x_2 + 2 \geq 0 \\ g_2(x) = -x_1^2 + x_2 - 1 \geq 0 \\ x_1, x_2 \geq 0 \end{cases}$$

解：$f(x)$ 和$g_2(x)$ 的海赛矩阵的行列式分别是

$$|H| = \begin{vmatrix} \dfrac{\partial^2 f(x)}{\partial x_1^2} & \dfrac{\partial^2 f(x)}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f(x)}{\partial x_2^2} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0$$

$$|g_2| = \begin{vmatrix} \dfrac{\partial^2 g_2(x)}{\partial x_1^2} & \dfrac{\partial^2 g_2(x)}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 g_2(x)}{\partial x_2 \partial x_1} & \dfrac{\partial^2 g_2(x)}{\partial x_2^2} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 0 \end{vmatrix} = 0$$

知$f(x)$ 为严格凸函数，$g_2(x)$ 为凹函数。

由于其他约束条件均为线性函数，

所以这是一个凸规划问题.

C 点为其最优点： $X^* = (0.58, 1.34)^T$,

目标函数的最优值为 $f(X^*) = 3.8$.

谢　谢!

10.2.2无约束非线性规划的解法

1、最优解定义与结论

考虑无约束非线性规划问题

$$NLP: \quad \min z = f(x) \quad (10.2.4)$$

其中$x = (x_1, x_2, \cdots, x_N)^T$. 无约束非线性规划问题就是一个无条件极值问题.

定义10.5：局部最优解设$X^* \in D$，若存在$\delta > 0$，使得对一切$X \in D$，且$||X - X^*|| < \delta$，都有$f(X^*) \leq f(X)$，则称X^*是$f(X)$在D上的局部极小值点（局部最优解）．特别地，当$X \neq X^*$时，若$f(X^*) < f(X)$，则称X^*是$f(X)$在D上的严格局部极小值点（严格局部最优解）．

定义10.6：（全局最优解） 设$X^* \in D$，若对任意的$X \in D$，都有$f(X^*) \leq f(X)$，则称X^*是$f(X)$在D上的全局极小值点（全局最优解）. 特别地，当$X \neq X^*$时，若$f(X^*) < f(X)$，则称X^*是$f(X)$在D上的严格全局极小值点(严格全局最优解).

函数 $f(x) = f(x_1, x_2, \cdots, x_N)$

的一阶导数

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_N}\right)^T$$

称为其梯度.

记$f(x)$ 在点

$$x^0 = (x_1^0, x_2^0, \cdots, x_N^0)^T$$

处的梯度为:

$$\nabla f(x^0) = (\frac{\partial f(x^0)}{\partial x_1}, \frac{\partial f(x^0)}{\partial x_2}, \cdots, \frac{\partial f(x^0)}{\partial x_N})^T.$$

在微积分学中，一元函数取极值的必要条件：可导的一元函数 $f(x)$ 在 x^* 处取极值，则 $f'(x^*) = 0$（稳定点或驻点）.此结论可推广到多元函数的情形.

定理 10.3：（必要条件）

设函数$f(x)$ 可微，若 x^* 为无约束非线性规划问题

$$NLP: minz = f(x)$$

的局部最优解，

则 $\nabla f(x^*) = 0$（稳定点或驻点）.

定理10.4：（充分条件）

若梯度 $\nabla f(x^*) = 0$，且海赛矩阵$\nabla^2 f(x^*)$正定，

则 x^* 为无约束非线性规划问题

$NLP: minz = f(x)$

的局部最优解.

定理10.5：（必要条件）

若 $f(x)$ 为可微的凸函数，则梯度 $\nabla f(x^*) = 0$ 的充要条件是 x^* 为无约束非线性规划问题

$$NLP : minz = f(x)$$

的整体最优解.

例 10.4：求解无约束非线性规划问题NLP：

$$min\ \ f(x) = x_1^2 + 4x_2^2 + x_3^2 - 2x_1$$

解：$\nabla f(x) = (\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \frac{\partial f(x)}{\partial x_3})^T$

$= (2x_1 - 2, 8x_2, 2x_3)^T,$

$$\nabla f^2(x) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

令$\nabla f(x) = 0$，得$f(x)$ 的驻点为 $x^* = (1,0,0)^T$.

显然 $\nabla f^2(x)$ 正定，故 $f(x)$ 为凸函数，（NLP）为凸规划，因此 $x^* = (1, 0, 0)^T$ 为（NLP）的局部最优解，当然也是整体最优解.

定理 10.6若 $f(x)$ 为定义在凸集 K 上的凸函数，则它的任一极小点就是它在 K 上的最小点(全局极小点)，而且它的极小点形成一个凸集.

证明：设 x^* 是一个局部极小点，则对于充分小的邻域 $N\delta(x^*)$ 中的一切 x，均有$f(x) \geq f(x^*)$。

令 Y 是 K 中的任一点，对于充分小的$\boldsymbol{\lambda}$, $0 < \boldsymbol{\lambda} < 1$，就有

$$((1-\boldsymbol{\lambda})x^* + \boldsymbol{\lambda}Y) \in N\delta(x^*)$$

从而

$$f((1-\boldsymbol{\lambda})x^* + \boldsymbol{\lambda}Y)) \geq f(x^*)$$

由于$f(x)$ 为凸函数，故

$$(1-\boldsymbol{\lambda})f(x^*) + \boldsymbol{\lambda}f(Y) \geq f((1-\boldsymbol{\lambda})x^* + \boldsymbol{\lambda}Y)$$

将上述两个不等式相加，移项后除以 $\boldsymbol{\lambda}$，得到

$$\boldsymbol{f(Y) \geq f(x^*)}$$

这就是说，$\boldsymbol{x^*}$是全局极小点。

由性质2，所有极小点的集合形成一个凸集。

定理 10.7设 $f(x)$ 是定义在凸集 R 上的可微凸函数，若存在点$x^* \in \boldsymbol{K}$，使得对于所有的 $x \in \boldsymbol{K}$ 有

$$\nabla f(x^*)^T \cdot (X - x^*) \geq 0 \quad (10.2.5)$$

则 x^* 是 $f(x)$ 在 $\boldsymbol{K}$ 上的最小点(全局极小点)。

证明：由定理 10.1

$$f(x) \geq f(x^*) + \nabla f(x^*)^T \cdot (x - x^*)$$

如此，对所有 $x \in K$ 有

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

一种极为重要的情形是，当点 x^* 是$\boldsymbol{K}$ 的内点时，
这时 (10.2.3) 式对任意 $\boldsymbol{x}-\boldsymbol{x}^*$ 都成立，
这就意味着可将 (10.2.5) 式改为

$$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$$

以上两个定理说明，定义在凸集上的凸函数的稳定点，就是其全局极小点。

全局极小并不一定是唯一的，但若为严格凸函数，则其全局极小点就是唯一的了.

显然，上述求解无约束非线性规划问题的微分法是有局限性的，因为目标函数未必可微.

因此，需寻求其他方法.

谢　谢!

2、下降迭代算法

为了求某可微函数（假定无约束）的最优解，

根据前面的叙述，可如下进行:

令该函数的梯度等于零，

由此求得稳定点；

然后用充分条件进行判别，求出所要的解. 对某些较简单的函数，这样做有时是可行的；但对一般 N 元函数 $f(x)$ 来说，由条件 $\nabla f(x) = 0$ 得到的常常是一个非线性方程组，解它相当困难. 对于不可微函数，无法使用该方法.

为此，常直接使用迭代法.

迭代法的基本思想是：

为了求函数 $f(x)$ 的最优解，

首先给定一个初始估计 x^0，

然后按某种规则(即算法）找出比 x^0 更好的解 x^1，

对极小化问题， $f(x^1) < f(x^0)$;

对极大化问题， $f(x^1) > f(x^0)$,

再按此种规则找出比 x^1 更好的解 $x^2, \cdots ,$.

如此即可得到一个解的序列 $\{x^k\}$.

若这个解序列有极限 x^*，即

$$\lim_{k\to\infty} \|x^k - x^*\| = 0$$

则称它收敛于 x^* .

若这算法是有效的，那么它所产生的解的序列将收敛于该问题的最优解。

但计算机只能进行有限次迭代，一般说很难得到准确解，而只能得到近似解. 当满足所要求的精度时，即可停止迭代.

若由某算法所产生的解的序列 $\{x^k\}$ 使目标函数值 $f(x^k)$ 逐步减少，就称这算法为下降算法. 显然，求解极小化问题应采用下降算法.

现假定已迭代到点 x^k ，若从 x^k 出发沿任何方向移动都不能使目标函数值下降，则 x^k 是一局部极小点，迭代停止.

若从 x^k 出发至少存在一个方向可使目标函数值有所下降，则可选定能使目标函数值下降的某方向 P^k，沿这个方向迈进适当的一步，得到下一个迭代点 x^{k+1}，并使$f(x^{k+1}) < f(x^k)$

这相当于在射线：$\boldsymbol{x} = \boldsymbol{x}^k + \boldsymbol{\lambda} \boldsymbol{P}^k$上选定新点：

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{\lambda}_k \boldsymbol{P}^k$$

其中，$\boldsymbol{P}^k$ 称为搜索方向，$\boldsymbol{\lambda}_k$ 称为步长或步长因子。

下降迭代算法的步骤总结：

（1） 选定某一初始点 x^0，并令 $k = 0$；

（2） 确定搜索方向 p_k；

（3）从 x^k 出发，沿方向求步长 λ_k，以产生下一迭代点 x^{k+1}；

（4）检查得到的新点 x^{k+1} 是否为极小点或近似极小点(满足精度)，若是，则停止迭代.

否则令 $\boldsymbol{k=k+1}$，返回（2）继续迭代.

以上步骤中，确定搜索方向 p_k 是最关键的一步，也是各算法的区分之处；

确定步长 $\boldsymbol{\lambda}_k$ 可选用不同的方法.最简单的一种是令 $\boldsymbol{\lambda}_k$ 恒等于一个常数，计算简便但不保证目标函数值是下降的；

第二种称为可接受点算法：任意选取步长 $\boldsymbol{\lambda}_k$，只要它能使目标函数值下降；可任意选取步长 $\boldsymbol{\lambda}_k$.

第三种方法是基于沿搜索方向使目标函数值下降最多，即沿射线

$$x = x^k + \lambda P^k$$

求目标函数$f(x)$的极小：

$$\lambda_k : min f(x^k + \lambda P^k)$$

由于这项工作是求以 $\boldsymbol{\lambda}$ 为变量的一元函数

$f(x^k + \boldsymbol{\lambda} \boldsymbol{P}^k)$的极小点 $\boldsymbol{\lambda}_k$ ，故常称这一过程为（最优）一维搜索或线搜索，这样确定的步长为最佳步长.

一维搜索有个十分重要的性质：在搜索方向上所得最优点处的梯度和该搜索方向正交.

定理10.8设目标函数$f(x)$ 具有一阶连续偏导数，x^{k+1} 按下述重要的性质：规则产生 λ_k：

$$\begin{cases} \lambda_k : \min f(x^k + \lambda P^k) \\ x^{k+1} = x^k + \lambda_k P^k \end{cases}$$

则有$\nabla f(x^{k+1})^T \cdot P^k = 0$

证明：构造函数

$$\varphi(\boldsymbol{\lambda}) = f(x^k + \boldsymbol{\lambda} P^k),$$

则得

$$\begin{cases} \varphi(\boldsymbol{\lambda}_k) = \min\limits_{\lambda} \varphi(\boldsymbol{\lambda}) \\ x^{k+1} = x^k + \boldsymbol{\lambda}_k P^k \end{cases}$$

即 $\boldsymbol{\lambda}_k$ 为 $\varphi(\boldsymbol{\lambda})$ 的极小点. 此外

$$\varphi'(\boldsymbol{\lambda}) = \nabla f(x^k + \boldsymbol{\lambda} P^k)^T \cdot P^k$$

由 $\varphi'(\boldsymbol{\lambda})|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_k} = 0$, 可得

$$\nabla f(x^k + \boldsymbol{\lambda} P^k)^T \cdot P^k = \nabla f(x^{k+1})^T \cdot P^k = 0$$

定理得证.

对一个好的算法，不仅要求它产生的点列能收敛到问题的最优解，还要求具有较快的收敛速度.

设序列 $\{x^k\}$ 收敛于 x^* ，若存在与迭代次数 k 无关的数$0 < \beta < \infty$ 和 $\alpha \geq 1$ ，使 k 从某个 $k_0 > 0$ 开始都成立，$\|x^{k+1} - x^*\| \leq \beta \|x^k - x^*\|^\alpha$. 就称 $\{x^k\}$ 收敛的阶为 α ，或 $\{x^k\}\alpha$ 阶收敛.

当 $\alpha = 2$ 时，称为二阶收敛，也可说 $\{x^k\}$ 具有二阶敛速.

当 $1 < \alpha < 2$ 时，称超线性收敛.

当 $\alpha = 1$，且 $0 < \beta < 1$ 时，称线性收敛或一阶收敛.

谢　谢!

3、最速下降法

在求解无约束极值问题（10.2.4）时常使用迭代法，迭代法可大体分为两大类。

一类要用到函数的一阶导数和（或）二阶导数，用到了函数的解析性质，故称为解析法；

另一类在迭代过程中仅用到函数值，而不要求函数的解析性质，这类方法称为直接法.

一般说来，直接法的收敛速度较慢，只是在变量较少时才适用. 但是直接法的迭代步骤简单，特别是当目标函数的解析表达式十分复杂，甚至写不出具体表达式时，它们的导数很难求得，或者根本不存在，这时解析法就无能为力了.

本节只介绍在求解无约束极值问题的解析法中最为基本的一种数值方法，

梯度法. 假定无约束极值问题（10.2.4）式中的目标函数 $f(x)$ 有一阶连续偏导数，具有极小点 x^* .

以 x^k 表示极小点的第 k 次近似，为了求其第 $k+1$ 次近似点 x^{k+1}，我们在 x^k 点沿方向 P^k 做射线

$$x = x^k + \lambda P^k (\lambda \geq 0)$$

现将 $f(x)$ 在 x^k 点处展成泰勒级数

$$f(x) = f(x^k + \lambda P^k)$$

$$= f(x^k) + \lambda \nabla f(x^k)^T \cdot P^k + o(\lambda)$$

其中$\lim\limits_{\lambda \to 0} \frac{o(\lambda)}{\lambda} = 0$

对于充分小的 $\boldsymbol{\lambda}$，只要

$\nabla f(x^k)^T \cdot P^k < 0 \quad (10.2.6)$

即保证$f(x^k + \boldsymbol{\lambda} P^k) < f(x^k)$.

这时若取$x^{k+1} = x^k + \boldsymbol{\lambda} P^k$

就能使目标函数值得到改善.

现考查不同的方向 $\boldsymbol{P}^k$. 假定 $\boldsymbol{P}^k$ 的模一定（且不为零），并设 $\nabla f(\boldsymbol{x}^k) \neq \mathbf{0}$（否则，$\boldsymbol{x}^k$ 是稳定点），使式（10.2.6）成立的 $\boldsymbol{P}^k$ 有无限多个.

为了使目标函数值能得到尽量大的改善，必须寻求使 $\nabla f(x^k)^T \cdot P^k$ 取最小值的 P^k. 由线性代数学知道

$$\nabla f(x^k)^T \cdot P^k = \|\nabla f(x^k)\| \cdot \|P^k\| \cos\theta$$

式中 θ 为向量$\nabla f(x^k)$与P^k 的夹角.

当 $\boldsymbol{P}^k$与$\boldsymbol{f}(\boldsymbol{x}^k)$ 反向时，$\boldsymbol{\theta} = \mathbf{180}^\circ$，$\mathbf{cos}\,\boldsymbol{\theta} = -\mathbf{1}.$

这时（10.2.6）式成立，而且其左端取最小值.

我们称方向$\boldsymbol{P}^k = -\nabla \boldsymbol{f}(\boldsymbol{x}^k)$为负梯度方向，

它是使函数值下降最快的方向

（在 $\boldsymbol{x}^k$ 的某一小范围内）.

定理10.9函数$f(x)$在点x^0处的负梯度$-\nabla f(x^0)$是$f(x)$在点x^0 处下降最快的方向.

证明：由泰勒公式知

$$f(x^0+\lambda p)-f(x^0)=\nabla f(x^0)'\cdot\lambda p+o(||\lambda p||)$$

$$=\lambda\nabla f(x^o)'\cdot p+o(||\lambda p||).$$

因此，略去高阶无穷小量 $o(||\lambda p||)$ 不计，取 $p = -\nabla f(x^0)$ 时，函数 $f(x)$ 的值下降最快.

定理10.10若梯度$-\nabla f(x^0)=0$，则$f(x)$在点x^0处沿负梯度方向$-\nabla f(x^0)$不会再下降，故x^0为NLP的最优解.

注：为增强算法可操作性，判定“$\nabla f(x^0)=0$”可等价地转化为判定“$\nabla f(x^0)\approx 0$”.

为了得到下一个近似极小点，在选定了搜索方向之后，还要确定步长$\boldsymbol{\lambda}$. 当采用可接受点算法时，就是取某一$\boldsymbol{\lambda}$进行试算，看是否满足不等式

$$f(x^k - \boldsymbol{\lambda}\nabla f(x^k)) < f(x^k) \quad (10.2.7)$$

若上述不等式成立，就可以迭代下去。

否则，缩小 $\boldsymbol{\lambda}$ 使满足不等式（10.2.7）式. 由于采用负梯度方向，满足（10.2.7）式的$\boldsymbol{\lambda}$ 总是存在的. 另一种方法是通过在负梯度方向的一维搜索，来确定使 $f(x^k)$ 最小的 $\boldsymbol{\lambda}_k$ ，这种梯度法就是最速下降法.

总之，最速下降法的基本思想如下：给定初始点 x^0，若 $\nabla f(x^0) = 0$，则 x^0 即为 NLP 的最优解；否则，$f(x)$ 在点 x^0 处沿负梯度方向 $-\nabla f(x^0)$ 是 $f(x)$ 在点 x^0 处下降最快的方向.

于是，求解 $NLP \Longleftrightarrow$ 在点 x^0 处沿负梯度方向 $-\nabla f(x^0)$ 求函数$f(x)$ 的最小值$\Longleftrightarrow$ 求解极值问题P：

$$min\ f(x^0+\lambda\nabla f(x^0)),$$

其中λ 为步长. 这里 P 是一个以 λ 为自变量的

一元函数的极值问题，可利用微积分知识求解. 设求得 P 的极值点为 λ_0 (最优步长)，令

$$x^1 := x^0 + \lambda_0 \nabla(f(x^0)),$$

由定理易证,

$$f(x^1) < f(x^0).$$

重复上述步骤.

算法步骤:

步骤 1 取初始点 x^0 及允许误差 $\varepsilon > 0$，令 $k := 0$.

步骤 2计算 $P^k = -\nabla f(x^k)$.

步骤 3若 $||P^k|| \leq \varepsilon$，则 x^k 即为 NLP 的最优解，停；否则，转入步骤4.

步骤 4求解极值问题

$$\boldsymbol{min\{|f(x^k+\lambda P^k)|\}},$$

设得极值点 $\boldsymbol{\lambda_k}$.

步骤 5令 $\boldsymbol{x^{k+1} := x^k + \lambda_k P^k}$

$\boldsymbol{k := k+1}$，转步骤 2.

最速下降法的算法模式实质上是一种特定形式的迭代法，其中x^0为初始点，P^k为搜索方向，即下降方向，是区别不同算法的主要标志，

$\boldsymbol{\lambda}_k$为步长，$x^{k+1} := x^k + \boldsymbol{\lambda}_k P^k$为迭代格式.

例10.5：用最速下降法求解无约束非线性规划问题NLP：

$$minz = (x_1 - 1)^2 + 4(x_2 - 1)^2$$

取初始点$x^0 = (1, 0)'$，允许误差$\varepsilon = 0.01$.

解： (1) $f(x) = (x_1 - 1)^2 + 4(x_2 - 1)^2$,

$\nabla f(x) = (2(x_1 - 1), 8(x_2 - 1))'$

$p^0 = -\nabla f(x^0) = -(0, -8)' = (0, 8)'$

(2) $||p^0|| = \sqrt{0^2 + 8^2} = 8 > \varepsilon.$

$(3)\boldsymbol{x}^0+\boldsymbol{\lambda p}^0=(1,0)'+\boldsymbol{\lambda}(0,8)'=(1,8\boldsymbol{\lambda})'$

$f(\boldsymbol{x}^0+\boldsymbol{\lambda p}^0)=(1-1)^2+4(8\boldsymbol{\lambda}-1)^2=4(8\boldsymbol{\lambda}-1)^2$

求得极值点$\boldsymbol{\lambda}=\frac{1}{8}$，代入

$\boldsymbol{x}^0+\boldsymbol{\lambda p}^0=(1,0)'+\frac{1}{8}(0,8)'=(1,1)'=\boldsymbol{x}^1$

$$(4)\boldsymbol{p}^1 = -\boldsymbol{\nabla} \boldsymbol{f}(\boldsymbol{x}^1) = -(\mathbf{0}, \mathbf{0})' = (\mathbf{0}, \mathbf{0})'$$

$$|\boldsymbol{p}^1| = \sqrt{\mathbf{0}^2 + \mathbf{0}^2} = \mathbf{0} < \varepsilon$$

所以最优解为$\boldsymbol{x}^1 = (\mathbf{1}, \mathbf{1})$.

谢　谢!

10.2.3 约束非线性规划的解法

考虑约束非线性规划问题：

$$min \quad z = f(x) \tag{10.2.1}$$

$s.t.$

$$\begin{cases} h_j(x) = 0, i = 1, 2, \cdots, q \\ g_i(x) \geq 0, i = 1, 2, \cdots, p \end{cases} \tag{10.2.2}$$

.

其中$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)'$,

可行域

$$K=\{x=(x_1,x_2,\cdots,x_n)'|$$

$$g_i(x)\leq 0, i=1,2,\cdots,p;$$

$$h_j(x)=0, j=1,2,\cdots,q\}.$$

定义10.7：（$\boldsymbol{NLP}$的最优解）

对非线性规划问题（10.2.1）（10.2.2）及 $\boldsymbol{x}^* \in \boldsymbol{K}$，若 $\forall \boldsymbol{x} \in \boldsymbol{K}$，有 $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$，则称 $\boldsymbol{x}^*$ 为 $\boldsymbol{NLP}$ 的整体最优解，$f(\boldsymbol{x}^*)$为$\boldsymbol{NLP}$的(整体)最优值；

若在x^*的某领域内，有$f(x^*) \le f(x)$，则称x^*为NLP的局部最优解，$f(x^*)$为NLP的局部最优值. 现假定$f(x)$、$h_i(x)$ 和 $g_j(x)(i = 1, 2, \cdots, p;\ j = 1, 2, \cdots, q)$具有一阶连续偏导数.

设 x^0 是非线性规划的一个可行解，它当然满足所有约束.

现考虑某一不等式约条件$g_j(x) \geq 0$,x^0 满足它有两种可能: 其一为$g_j(x^0) > 0$ ，对 x^0 点的微小摄动，该约束仍然成立. 因而这一约束对 x^0 点的微小摄动不起限制作用，从而称这个约束条件是 x^0 点的无效约束（或不起作用约束）；

其二是$g_j(x^0)=0$：这时 x^0 点处于该约束条件形成的可行域边界上，只要 x^0 有摄动，该约束就发生变化. 因此，给约束对 x^0 的摄动起到了某种限制作用，故称这个约束是 x^0 点的有效约束（起作用约束）.

假定设 x^0 是非线性规划（10.2.1）（10.2.1）的一个可行点，现考虑此点的某一方向 P，若存在实数$\boldsymbol{\lambda}_0 > 0$，使对任意$\boldsymbol{\lambda} \in [0, \boldsymbol{\lambda}_0]$均有

$$\boldsymbol{x}^0 + \boldsymbol{\lambda P} \in \boldsymbol{K}$$

就称方向 P 是 x^0 点的一个可行方向。

根据梯度的性质，在可行点 x^0 处的曲面函数的梯度 $\nabla g_i(x^0)$ 方向一定指向 K 内，即可行方向与曲面 $g_i(x) = 0$ 本身以及曲面函数的梯度 $\nabla g_i(x^0)$ 都在可行点 x^0 处的切平面同侧.

因此，若 P 是可行点 x^0 处的任一可行方向，则对该点的所有有效约束$g_i(x) \geq 0$，均有

$$\nabla g_i(x^0)^T \cdot P \geq 0, i \in I$$

其中I为这个点所有有效约束下标的集合。

另一方面，由泰勒公式

$$g_i(x^0+\boldsymbol{\lambda} P)=g_i(x^0)+\boldsymbol{\lambda}\nabla g_i(x^0)^T\cdot P+o(\boldsymbol{\lambda})$$

对所有有效约束，当$\boldsymbol{\lambda}>0$ 足够小时，只要

$$\nabla g_i(x^0)^T P>0,\ i\in I \qquad (10.2.8)$$

就有$g_i(x^0+\boldsymbol{\lambda} P)\geq 0, i\in I.$

此外，对 x^0 点的无效约束，由约束函数的连续性，当$\boldsymbol{\lambda} > \mathbf{0}$足够小时亦有上式成立.

从而，只要方向P满足(10.2.8) 式，即可保证它是x^0点的可行方向。

考虑非线性规划的某一可行点 x^0，对该点的任一方向 P 来说，若存在实数 $\lambda_0' > 0$，使对任意 $\lambda \in [0, \lambda_0']$均有

$$f(x^0 + \lambda P) < f(x^0)$$

就称方向 P为x^0点的一个下降方向。

将目标函数 $f(x)$ 在点 x^0 处作一阶泰勒展开，可知满足条件$\nabla f(x^0)^T P < 0$的方向 P 必为x^0 点的下降方向. 如果方向 P 既是x^0 点的可行方向，又是这个点的下降方向，就称它是该点的**可行下降方向**.

假如 x^0 点不是极小点，继续寻优时的搜索方向，就应从该点的可行下降方向中去找. 显然，若某点存在可行下降方向，它就不会是极小点. 另一方面，若某点为极小点，则在该点不存在可行下降方向.

定理10.11设 x^* 是非线性规划(10.2.1) (10.2.2)的一个局部最优解，目标函数 $f(x)$ 在 x^* 处可微，而且 $g_i(x)$在 x^*处可微，当 $i \in I$ ，和 $g_i(x)$ 在 x^* 处连续，当$i \notin I$. 则在 x^* 点不存在可行下降方向，从而不存在向量 P 同时满足：

$$\begin{cases} \nabla f(x^*)^T P < 0 \\ \nabla g_i(x^*)^T P > 0, i \in I \end{cases} \qquad (10.2.9)$$

事实上，若存在满足 (10.2.9) 式的方向 $\boldsymbol{P}$，则沿该方向搜索找到更好的点，从而与 $\boldsymbol{x}^*$ 为局部最优解的假设矛盾. 满足 (10.2.9) 的方向 $\boldsymbol{P}$，与点 $\boldsymbol{x}^*$ 处目标函数负梯度方向的夹角为锐角，与点 $\boldsymbol{x}^*$ 处有效约束的梯度方向的夹角也为锐角.

谢　谢!

1、Kuhn-Tucher(简称 $K-T$库恩-塔克)条件

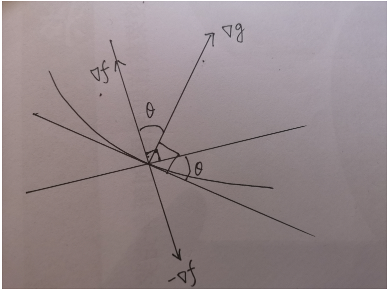
假定 x^* 是非线性规划(10.2.1) (10.2.2)的局部最优解，该点可能位于可行域的内部，也可能处于可行域的边界上。若为前者，这事实上是个无约束问题，x^*必满足条件

$$\nabla f(x^*) = 0 \;;$$

现在我们来讨论后一种情形. 不失一般性，设x^*位于第一个约束条件形成的可行域边界上，即第一个约束条件是x^*点的有效约束

$$g_1(x^*) = 0。$$

若x^*是局部最优解, 假定向量$\nabla g_1(x^*)$和$\nabla f(x^*)$皆不为零，则$\nabla g_1(x^*)$ 必与$-\nabla f(x^*)$在一条直线上且方向相反，否则必有方向 P 使(10.2.9)成立，即与有效约束的梯度方向的夹角成锐角，而与目标函数的梯度方向成钝角，



∇g
∇f
θ
θ
$-\nabla f$

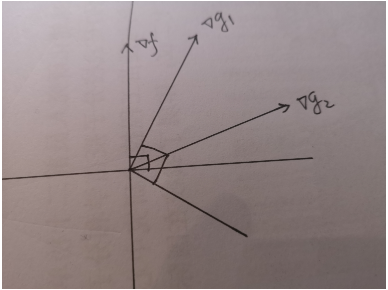
即在该点就一定存在可行下降方向，这与该点是局部最优解矛盾。

因此，在上述条件下，存在实数$\lambda_1 \geq 0$，使

$$\nabla f(x^*) - \lambda_1 \nabla g_1(x^*) = 0.$$

若 x^* 点有两个有效约束，例如说有$g_1(x^*) = 0$和$g_2(x^*$
0。在这种情况下，$\nabla f(x^*)$ 必处于$\nabla g_1(x^*)$和$\nabla g_2(x^*)$的夹角之内。

否则，必有方向 P 与 $\nabla g_1(x^*)$ 和 $\nabla g_2(x^*)$ 的夹角均成锐角，



而与目标函数的梯度方向成钝角，即在x^* 点就一定存在可行下降方向，这与该点是局部最优解矛盾。由此可见，如果x^*是局部最优解，而且x^*点的有效约束条件的梯度$\nabla g_1(x^*)$和$\nabla g_2(x^*)$线性无关。则可将$\nabla f(x^*)$表示成$\nabla g_1(x^*)$和$\nabla g_2(x^*)$的非负线性组合。

也就是说，在这种情况下存在实数$\boldsymbol{\lambda}_1 \geq 0$和$\boldsymbol{\lambda}_2 \geq 0$，使

$$\nabla f(x^*) - \boldsymbol{\lambda}_1 \nabla g_1(x^*) - \boldsymbol{\lambda}_2 \nabla g_2(x^*) = 0$$

如上类推，可以得到

$$\nabla f(x^*) - \sum_{i \in I} \boldsymbol{\lambda}_i \nabla g_i(x^*) = 0$$

为了把无效约束也包括进上式中，增加条件

$$\begin{cases} \lambda_i g_i(x^*) = 0 \\ \lambda_i \geq 0 \end{cases}$$

当 $g_i(x^*) = 0$ 时，λ_i 可不为零；当$g_i(x^*) \neq 0$时，必有$\lambda_i = 0$。如此即可得到著名的库恩- 塔克$Kuhn - Tucker$，简写为 $K - T$ 条件。

库恩- 塔克条件是确定某点为最优的必要条件。但一般说它并不是充分条件，因而满足这个条件的点不一定就是最优点（对于凸规划，它既是最优点存在的必要条件，同时也是充分条件）.

现可将库恩- 塔克条件叙述如下：设x^*是非线性规划(10.2.1) (10.2.2)式的局部最优解，且在x^*点的各有效约束的梯度线性无关，则存在向量

$$\boldsymbol{\Lambda}^* = (\boldsymbol{\lambda}_1^*, \boldsymbol{\lambda}_2^*, \cdots, \boldsymbol{\lambda}_p^*)^T,$$

使下述条件成立：

$$\begin{cases} \nabla f(x^*) - \sum\limits_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) = 0 \\ \lambda_i^* g_i(x^*) = 0, i = 1, 2, \cdots, I \\ \lambda_i^* \geq 0, i = 1, 2, \cdots, I \end{cases} \qquad (10.2.10).$$

条件（10.2.10）式常简称为$K-T$条件。

为了得出非线性规划(10.2.1)(10.2.2)式的库恩- 塔克条件，我们用

$$\begin{cases} h_j(x) \geq 0 \\ -h_j(x) \geq 0 \end{cases}$$

代替约束条件$h_j(x) = 0$，

这样即可使用条件(10.2.10)，得到这时的库恩- 塔克条件如下：

定理 10.13（必要条件） 设x^*是非线性规划问题(10.2.1)(10.2.2)的局部最优解，而且x^*点的所有起作用约束的梯度$h_j(x^*)(j = 1, 2, \cdots, q)$和$g_i(x^*)(i = 1, 2, \cdots, p)$线性无关，

则存在向量$\boldsymbol{\Lambda}^* = (\boldsymbol{\lambda}_1^*, \boldsymbol{\lambda}_2^*, \cdots, \boldsymbol{\lambda}_p^*)^T$

和$\boldsymbol{\Gamma}^* = (\gamma_1^*, \gamma^* 2, \cdots, \gamma_q^*)^T$使下述条件成立:

$$\begin{cases} \nabla f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^{q} \gamma_j^* \nabla h_j(x^*) \\ \lambda_i^* g_i(x^*) = 0, \qquad i = 1, 2, \cdots, p \\ \lambda_i^* \geq 0, \qquad i = 1, 2, \cdots, p \end{cases} \tag{10.2.11}$$

满足条件(10.2.11)式的点也称为库恩-塔克点。在条件(10.2.10)式和(10.2.11)式中，

$\boldsymbol{\lambda}_1^*, \boldsymbol{\lambda}_2^*, \cdots, \boldsymbol{\lambda}_p^*$

以及$\gamma_1^*, \gamma^*\mathbf{2}, \cdots, \gamma_{\boldsymbol{q}}^*$

称为广义拉格朗日（Lagrange）乘子.

定理10.14(充分条件)若 x 满足 $K-T$ 条件，则 x 必为凸规划的局部最优解，进而为整体最优解．定理10.13 和定理 10.14 给出了非线性规划问题的最优解应满足的充分条件和必要条件，二者合称为**非线性规划基本定理**．如果 x 为凸规划的（局部、整体）最优解$\Longleftrightarrow x$满足$K-T$条件．

例 10.6用库恩- 塔克条件解非线性规划

$$\begin{cases} \min f(x) = (x-3)^2 \\ 0 \le x \le 5 \end{cases}$$

解：先将该非线性规划问题写成以下形式

$$\begin{cases} \min f(x) = (x-3)^2 \\ g_1(x) = x \geq 0 \\ g_2(x) = 5 - x \geq 0 \end{cases}$$

写出其目标函数和约束函数的梯度：

$$f(x) = 2(x-3) \quad \nabla g_1(x) = 1 \quad \nabla g_2(x) = -1。$$

对第一个和第二个约束条件分别引入广义拉格朗日乘子 $\boldsymbol{\lambda}_1^*$ 和 $\boldsymbol{\lambda}_2^*$，

设 $K-T$ 点为x^*，则可写出该问题的 $K-T$ 条件如下:

$$\begin{cases} 2(x^*-3)-\lambda_1^*+\lambda_2^*=0 \\ \lambda_1^* x^*=0 \\ \lambda_2^*(5-x^*)=0 \\ \lambda_1^*, \lambda_2^* \geq 0 \end{cases}$$

为解上述方程组，考虑以下几种情形：

(1) 令$\boldsymbol{\lambda}_1^* \notin 0, \boldsymbol{\lambda}_2^* \notin 0$，无解。

(2) 令$\boldsymbol{\lambda}_1^* \notin 0, \boldsymbol{\lambda}_2^* = 0$，解之，得$\boldsymbol{x}^* = 0$，$\boldsymbol{\lambda}_1^* = -6$，不是$\boldsymbol{K}-\boldsymbol{T}$点。

(3) 令$\boldsymbol{\lambda}_1^* = \mathbf{0}$，$\boldsymbol{\lambda}_2^* \notin \mathbf{0}$，解之，得$x^* = \mathbf{5}, \boldsymbol{\lambda}_2^* = -\mathbf{4}$，不是$K-T$点。

(4) 令$\boldsymbol{\lambda}_1^* = \boldsymbol{\lambda}_2^* = \mathbf{0}$，解之，得$x^* = \mathbf{3}$，此为$K-T$点，其目标函数值$f(x^*) = \mathbf{0}$.

由于该非线性规划问题为凸规划，故$x^* = 3$就是其全局极小点。该点是可行域的内点，它也可直接由梯度等于零的条件求出。利用$K-T$条件(包括拉格朗日乘子法)求解约束非线性规划问题是有局限性的，故需寻找其他方法.

谢　谢!

2、惩罚函数法

求解约束非线性规划问题的一般思路是将其转化为等价的无约束非线性规划问题。一般做法是构造新函数

$$Q(x,M)=f(x)+M[\sum_{i=1}^{p}\varphi(g_i(x))+\sum_{j=1}^{q}\varphi(h_j^2(x))]$$
$$=f(x)+M[\sum_{i=1}^{p}(\min(0,g_i(x)))^2+\sum_{j=1}^{q}\varphi(h_j^2(x))],$$

其中，M是个充分大的正数．当$x \in K$时，有$\min(0,$
0，$h_j^2(x) = 0$，故

$$Q(x, M) = f(x) + M \cdot 0 = f(x);$$

当$x \notin K$时，有$\min(0, g_i(x)) = g_i(x)$，$h_j^2(x) =$
0不一定成立，故$Q(x, M) > f(x) + 0 = f(x)$.

综上，

$$Q(x,M)=\begin{cases} = f(x), & x\in K \\ > f(x), & x\notin K \end{cases}$$

构造无条件极值问题：$\min Q(x,M)$，并解得极值点x^*．若$x\notin K$时，则根据上述分析知，

$$\sum_{i=1}^{p}(\min(0,g_i(x^*)))^2+\sum_{j=1}^{q}h_j^2(x^*)$$

$$=\sum_{j=1}^{q}g_i^2(x^*)+\sum_{j=1}^{q}h_j^2(x^*)>0,$$

又M 是个充分大的正数,

故函数$Q(x,M)$的值是一个充分大的正数，Q不可能取得最小值（受到惩罚），因此称$Q(x,M)$为惩罚函数，

并称$M[\sum_{i=1}^{p}(\min(0,g_i(x^*)))^2]$为惩罚项，$M$为惩罚因子.

换言之，Q要取得最小值，须有$x \in \boldsymbol{K}$。此时，在一定的允许误差下，可将x^*作为NLP的（近似的）最优解；更进一步地，有如下结论：

定理 10.15 Q的极值点$x \in \boldsymbol{K}$，则x必为NLP的最优解.

惩罚函数的基本思想：取某正数$M = M_1$，求解无条件极值问题$\min Q(x, M_1)$，设得其极值点x'. 若$x \in K$，则由定理知，x' 即为（NLP）的最优解；否则，令$M = M_2 := 10M_1$，重复以上步骤.

算法步骤：

步骤1 令$M := M_1 > 0$。允许误差$\delta > 0$，$k := 1$.

步骤2 求解无条件极值问题$\min Q(x, M_1)$，设得其极值点 x_k.

步骤3 若

$g_i(x^k) \le \varepsilon$, $|h_j(x^k)| \le \varepsilon (i = 1, \cdots, p; j = 1, \cdots, q)$,

则x^k 即为NLP 的近似最优解，停；否则，转步骤4 .

步骤4 令$M = M_{k+1} := 10M_k$，转步骤2.

例 10.7利用惩罚函数法求解非线性规划问题NLP:

$$\begin{cases} min \quad z = x \\ s.t. \quad x \geq 0 \end{cases}$$

解：令

$$Q(x, M_k) = f(x) + M_k(\min(0, g_1(x)))^2$$

$$= x + M_k(\min(0, -x))^2$$

$$= \begin{cases} x, & x \geq 0 \\ x + M_k x^2, & x < 0 \end{cases}.$$

求解无条件极值问题$min\{Q(x, M_k)\}$：由

$$Q'(x, M_k) = \begin{cases} 1, & x \geq 0 \\ 1 + 2M_k x, & x < 0 \end{cases}$$

得极值点$x^k = -\frac{1}{2M_k}$. 因$\lim\limits_{k\to\infty} x^k = \lim\limits_{M_k\to\infty} -\frac{1}{2M_k} = 0$,

故NLP的最优解为$x = 0$，最优值为$z = 0$.

谢　谢！

3 案例分析

案例一、容器的设计问题

问题背景：某公司专门生产储蓄容器，订货合同要求该公司制造一种敞口的长方体容器，容器恰好为12立方米，该容器的底必须为正方形，容器总重量不超过68公斤。

已知用做容器四壁的材料为每平米 10 元，重 3 公斤；用做容器底的材料每平方米 20 元，重 2 公斤。试建立制造该容器费用最小的最优化模型.

【模型建立】

该容器底边长和高分别为x_1，x_2，以容器的费用为目标函数，

则问题的数学模型为：

$$\min f(x) = 40x_1x_2 + 20x_1^2$$

$s.t.$

$$\begin{cases} x_1^2x_2 = 12 \\ 12x_1x_2 + 2x_1^2 \leq 68 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

谢　谢!

案例二、地址问题

问题背景：某特大城市拟新建一个民用机场，在选址与可行性研究阶段，提出要在$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}, \boldsymbol{F}$ 6 个卫星城市都得到统筹兼顾的原则下，使 6 个城市客运费最低作为选址目标.

案例分析

卫星城	平面坐标位置(x,y)	货运量/万吨(W_i)
A	(40,200)	40
B	(160,210)	10
C	(250,160)	20
D	(220,80)	30
E	(100,40)	20
F	(30,100)	10

说明

(1)$x + y \geq 250$处于高山，不适合建机场；

(2)$(x-120)^2+(y-160)^2 \leq 400$，处在最大湖，不准建机场.

试构建机场选址的NLP模型.

【模型求解】

设(x,y)为机场选址的位置坐标。总运费极小化的目标函数$f(x,y)$。由 6 个卫星城市的货运量(吨/千米)之和决定，即

案例分析

$$min\ f(x,y)=W_AL_A+W_BL_B+W_CL_C+W_DL_D+W_EL_E+W_FL_F$$

$$=40\sqrt{(x-40)^2+(y-200)^2}+10\sqrt{(x-160)^2+(y-210)^2}$$

$$+20\sqrt{(x-250)^2+(y-160)^2}+30\sqrt{(x-220)^2+(y-80)^2}$$

$$+20\sqrt{(x-100)^2+(y-40)^2}+10\sqrt{(x-30)^2+(y-100)^2}$$

约束条件有4个，即高山处，大湖处不适宜建机场，位置参照系x与y为非负变量：

案例分析

$$x + y < 250$$

$$(x - 120)^2 + (y - 160)^2 > 400$$

$$x \geq 0, y \geq 0$$

可看出上式中目标函数和约束条件都有非线性函数.

案例三、森林救火最小费用问题

问题背景：在森林失火时，应派多少消防员去救火最合适？派的队员越多，灭火的速度越快，造成的损失越小，但救援的开支越大。那么该派出多少队员救火，才能使火灾损失费与救火费用之和（总费用）最小？

【模型分析】

记队员人数 x，失火时刻$t=0$，开始救火时刻t_1，灭火时刻t_2，时刻t森林烧毁面积$B(t)$。损失费$f_1(x)$是 x 的减函数，由烧毁面积$B(t_2)$决定。救援费$f_2(x)$是 x 的增函数，由队员人数 x 和救火时间决定.

案例分析

问题转换为，存在恰当的 x ，使得$f_1(x)$，$f_2(x)$之和最小。对分析$B(t)$比较困难，我们转而谈论森林烧毁速度 dB/dt

案例分析

【模型假设】

1） $0 \le t \le t_1$，dB/dt与 t 成正比，系数β(火势蔓延速度)

2） $t_1 \le t \le t_2$，β降为$\beta - \lambda x$（λ为队员的平均灭火速度）

3）$f_1(x)$与$B(t_2)$成正比，系数c_1(烧毁单位面积损失费)

4）每个队员的单位时间灭火费用c_2，一次性费用c_3.

【模型解析】

假设 1) 解释，火势以失火点为中心，均匀向四周呈圆形蔓延，半径 r 与 t 成正比。则面积 B 与t^2 成正比，dB/dt与 t 成正比。

即$r = vt$

$$B(t) = \pi(vt)^2$$

$$\beta = dB/dt = \pi v^2 t$$

假设2)令 b 为救火时刻火势蔓延的面积，$b = \beta t_1$，则灭火时间为

案例分析

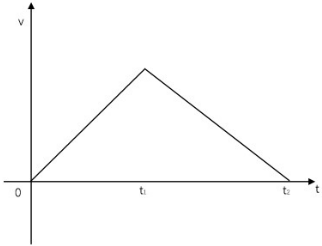
$$t_2 - t_1 = \frac{b}{\lambda x - \beta}$$

$$t_2 = t_1 + \frac{\beta t_1}{\lambda x - \beta}$$

则直至火势扑灭，森林烧毁面积为

$$B(t_2) = \int_0^{t_2} B'(t)dt = \frac{bt_2}{2} = \frac{\beta t_1^2}{2} + \frac{\beta^2 t_1^2}{2(\lambda x - \beta)}$$

案例分析



案例分析

假设3)、4)损失费$f_1(x)$为:

$$f_1(x) = c_1 B(t_2)$$

救援费$f_2(x)$为：

$$f_2(x) = c_2 x(t_2 - t_1) + c_3 x$$

案例分析

目标函数–总费用最低，即

$$C(x) = f_1(x) + f_2(x)$$

$$= \frac{c_1\beta t_1^2}{2} + \frac{c_1\beta^2 t_1^2}{2(\lambda x-\beta)} + \frac{c_2\beta t_1 x}{\lambda x-\beta} + c_3 x$$

其中$c_1, c_2, c_3, c_4, t_1, \beta, \lambda$视为已知参数.

案例分析

【模型建立】

森林救火模型：

$$Min\ \ C(x)$$

$$s.t \qquad x > 0$$

案例分析

其中$C(x) = f_1(x) + f_2(x) = \frac{c_1\beta t_1^2}{2} + \frac{c_1\beta^2 t_1^2}{2(\lambda x-\beta)} + \frac{c_2\beta t_1 x}{\lambda x-\beta} + c_3 x$

谢　谢!