最优化方法

南京邮电大学数理学院杨振华制作

njuptshumo2006@126.com

目录

第一章 最优化问题概述

§1.1 最优化问题的数学模型与基本概念

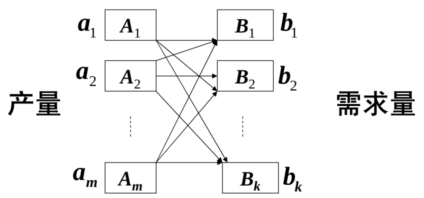
例 1.1.1 运输问题

设有m个水泥厂$A_1,A_2,\ldots,A_m$,年产量各为$a_1,a_2,\ldots,a_m$吨.有k个城市$B_1,B_2,\ldots,B_k$用这些水泥厂生产的水泥,年需求量$b_1,b_2,\ldots,b_k$吨.再设由A_i到B_j每吨水泥的运价为c_{ij}元.假设产销是平衡的,即:

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

试设计一个调运方案,在满足需要的同时使总运费最省.

由题意可画出如下的运输费用图:



设$A_i \to B_j$的水泥量为x_{ij},已知$A_i \to B_j$单价为c_{ij},单位为元,则总运费为:

$$S = \sum_{i=1}^{m} \sum_{j=1}^{k} c_{ij} x_{ij}$$

数学模型: $\min \sum_{i=1}^{m}\sum_{j=1}^{k} c_{ij}x_{ij}$

$$
\begin{aligned}
s.t.\quad & \sum_{j=1}^{k} x_{ij} = a_i (i=1,2,\cdots,m)\\
& \sum_{i=1}^{m} x_{ij} = b_j (j=1,2,\cdots,k)\\
& x_{ij} \ge 0 (i=1,2,\cdots,m, j=1,2,\cdots,k)
\end{aligned}
$$

注:平衡条件 $\sum_{i=1}^{m} a_i = \sum_{j=1}^{k} b_j$ 作为已知条件并不出现在约束条件中.

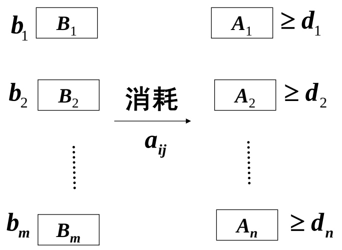
例1.1.2 生产计划问题

设某工厂有m种资源$B_1,B_2, \ldots,B_m$,数量各为: $b_1,b_2, \ldots, b_m$,用这些资源产n种产品$A_1,A_2, \ldots, A_n$.每生产一个单位的A_j产品需要消耗资源B_i的量为a_{ij},根据合同规定,产品A_j的量不少于d_j.再设A_j的单价为c_j.

问如何安排生产计划,才能既完成合同,又使该厂总收入最多?

假设产品A_j的计划产量为x_j.

由题意可画出如下的生产与消耗的关系图:



数学模型

$$
\begin{aligned}
& \max \sum_{j=1}^{n} c_j x_j\\
s.t.\quad & \sum_{j=1}^{n} a_{ij}x_j \le b_i (i=1,2,\cdots,m)\\
& x_j \ge d_j (j=1,2,\cdots,n)
\end{aligned}
$$

例 1.1.3 指派问题

设有四项任务B_1,B_2,B_3,B_4派四个人A_1,A_2, A_3,A_4去完成.每个人都可以承担四项任务中的任何一项,但所消耗的资金不同.设A_i完成B_j所需资金为c_{ij}.

如何分配任务,使总支出最少?

设变量

$$
x_{ij} = \begin{cases} 1 & \text{指派}A_i\text{完成}b_j \\ 0 & \text{不指派}A_i\text{完成}b_j \end{cases}
$$

则总支出可表示为: $S = \sum_{i=1}^{4}\sum_{j=1}^{4} c_{ij}x_{ij}$

数学模型:

$$
\begin{aligned}
\min\quad & S = \sum_{i=1}^{4}\sum_{j=1}^{4} c_{ij}x_{ij}\\
s.t.\quad & \sum_{j=1}^{4} x_{ij} = 1,\ i=1,2,3,4\\
& \sum_{i=1}^{4} x_{ij} = 1,\ j=1,2,3,4\\
& x_{ij} \in \{0,1\}, i,j = 1,2,3,4
\end{aligned}
$$

例 1.1.4 数据拟合问题

在实验数据处理或统计资料分析中常遇到如下问题.设两个变量x和y,已知存在函数关系,但其解析表达式或者是未知的或者虽然为已知的但过于复杂.

设已取得一组数据:

$$(x_i,y_i)\quad i=1,2,\ldots,m.$$

根据这一组数据导出函数$y=f(x)$的一个简单而近似的解析表达式.

最小二乘法

解这种问题常用的方法是最小二乘法,以一个简单的函数序列

$$\varphi_1(x),\ \varphi_2(x),\cdots,\ \varphi_n(x)$$

为基本函数.

一般选取$1,x,x^2,\cdots,x^n$为基本函数,即以

$$f(x)=\sum_{j=0}^{n}a_jx^j$$

作为近似表达式.

最小二乘法

系数的选取要使得下面得平方和最小:

$$Q=\sum_{i=1}^{m}(y_i-\sum_{j=0}^{n}a_j\varphi_j(x_i))^2$$

因此,数据拟合问题得数学模型为

$$\min\sum_{i=1}^{m}(y_i-\sum_{j=0}^{n}a_j\varphi_j(x_i))^2$$

其中$x_i,y_i(i=1,2,\ldots,m)$及$\varphi_j(x)(j=0,1,\ldots,n)$为已知.

最优化问题

最优化问题的一般形式为:

$$\min\quad f(x)\qquad (1.1)(目标函数)$$

$$s.t.\quad h_i(x)=0,i=1,2,\cdots,m\qquad (1.2)(等式约束)$$

$$g_j(x)\ge 0,j=1,2,\cdots,p\qquad (1.3)(不等式约束)$$

其中x是n维向量.

在实际应用中,可以将求最大值的目标函数取相反数后统一成公式中求最小值的形式.

相关定义

定义1.1.1 (可行解) 满足约束条件(1.2)和(1.3)的x称为可行解,也称为可行点或容许点.

定义1.1.2 (可行域)全体可行解构成的集合称为可行域,也称为容许集,记为D,即:

$D=\{x|h_i(x)=0,i=1,\cdots,m,g_j(x)\geqslant 0,j=1,\cdots,p,x\in R^n\}$.

若$h_i(x)$, $g_j(x)$为连续函数,则D为闭集.

相关定义

定义1.1.3 (整体最优解) 若$x^*\in D$,对于一切$x\in D$恒有$f(x^*)\leqslant f(x)$,则称x^*为最优化问题(P)的整体最优解.

若$x^*\in D,x\neq x^*$, 恒有$f(x^*)<f(x)$,则称x^*为最优化问题(P)的严格整体最优解.

相关定义

定义1.1.4 (局部最优解) 若$x^*\in D$,存在x^*的某邻域$N_\varepsilon(x^*)$,使得对于一切$x\in D\cap N_\varepsilon(x^*)$,恒有$f(x^*)\leqslant f(x)$,则称为最优化问题(P)的局部最优解,其中$N_\varepsilon(x^*)=\{x \mid \|x-x^*\|<\varepsilon,\varepsilon>0\}$.

当$x\neq x^*$时,若上面的不等式为严格不等式则称x^*为问题(P)的严格局部最优解.

显然,整体最优解一定是局部最优解,而局部最优解不一定是整体最优解.

相关定义

求解最优化问题(P),就是求目标函数$f(x)$在约束条件(1.2),(1.3)下的极小点,实际上是求可行域D上的整体最优解.但是,在一般情况下,整体最优解是很难求出的,往往只能求出局部最优解.

向量范数

定义1.1.5 如果向量$x\in R^n$的某个实值函数$\|x\|$,满足条件

(1)$\|x\|\geqslant 0$($\|x\|=0$当且仅当$x=0$)(正定性);

(2)$\|\alpha x\|=|\alpha|\cdot\|x\|$(对于任意$\alpha\in R$);

(3) $\|x+y\|\leqslant\|x\|+\|y\|$(三角不等式);

则称$\|x\|$为R^n上的一个向量范数.

常用的向量范数

1-范数 $\|x\|_1=\sum_{i=1}^{n}|x_i|$

2-范数(欧式范数) $\|x\|_2=(\sum_{i=1}^{n}|x_i|^2)^{\frac{1}{2}}$

∞-范数 $\|x\|_\infty=\max_{1\le i\le n}|x_i|$

p-范数 $\|x\|_p=(\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}}$

∞-范数是p-范数的极限 $\|x\|_\infty=\lim_{p\to\infty}\|x\|_p$

常用的向量范数

对向量$x=(1,-2,3)^T$,有

$$\|x\|_1=6,$$

$$\|x\|_2=\sqrt{14}\approx 3.74166,$$

$$\|x\|_3=\sqrt[3]{36}\approx 3.30193,$$

$$\|x\|_\infty=3.$$

$\|x\|_p$是p的单调递减函数.

最优化问题的分类

根据数学模型中有无约束函数分为有约束的最优化问题和无约束的最优化问题.

根据目标函数和约束函数的函数类型分类:线性最优化问题,非线性最优化问题,二次规划,多目标规划,动态规划,整数规划,0-1规划.

§ 1.2 最优化问题的一般算法

迭代算法

迭代算法 选取一个初始可行点$x_0 \in D$,由这个初始可行点出发,依次产生一个可行点列:

$x_1, x_2, \cdots, x_k, \cdots$,

记为$\{x_k\}$,使得某个x_k恰好是问题的一个最优解,或者该点列收敛到问题的一个最优解x^*.

下降算法 在迭代算法中一般要求

$f(x_{k+1}) \leqslant f(x_k)$.

可行点列的产生

在x_k处求得一个方向p_k(下降方向),在射线$x_k+\alpha p_k(\alpha>0)$上求一点:

$$x_{k+1}=x_k+\alpha_k p_k$$

使得$f(x_{k+1}) \leqslant f(x_k)$.其中$\alpha_k$称为步长.

定义1.2.1(下降方向) 在点x_k处,对于方向$p_k \neq 0$,若存在实数$\beta>0$,使得任意的$\alpha \in (0,\beta)$,有:

$f(x_k+\alpha p_k)<f(x_k)$,

则称p_k为函数$f(x)$在点x_k处的一个下降方向.

下降方向

若$f(x)$具有连续的一阶偏导数,令

由Taylor公式:

当$g_k^T p_k<0$时, $f(x_k+\alpha p_k)<f(x_k)$,所以p_k是$f(x)$在x_k处的一个下降方向.

反之,当p_k是$f(x)$在x_k处的一个下降方向时,有$g_k^T p_k \leqslant 0$ (书中有错) .

通常称满足$g_k^T p_k<0$的方向p_k是为$f(x)$在x_k处的一个下降方向.

可行方向

定义1.2.2(可行方向) 已知区域 ,$x_k \in D$,对于向量$p_k \neq 0$,若存在实数$\beta>0$, 使得对任意的$\alpha \in (0,\beta)$,有:$x_k+\alpha p_k \in D$,

则称p_k为点x_k处关于区域D的可行方向.

对于D的内点(存在邻域包含于D),任意方向可行,对于边界点(任意邻域既有D的点也有不在D中的点),则有些方向可行,有些方向不可行.

若下降方向关于域D可行,则称为可行下降方向.

最优化问题的算法的一般迭代格式

给定初始点x_0,令$k=0$.

(1) 确定x_k处的可行下降方向p_k;

(2) 确定步长α_k,使得$f(x_k+\alpha_k p_k)<f(x_k)$,

(3) 令$x_{k+1}=x_k+\alpha_k p_k$;

(4) 若x_{k+1}满足某种终止准则,则停止迭代,

以x_{k+1}为近似最优解;或者已经达到最大迭代步数,也可终止迭代.

否则令$k:=k+1$, 转(1)

收敛性

如果一个算法只有当初始点x_0充分接近x^*时，产生的点列才收敛于x^*,则称该算法为具有局部收敛的算法.

如果对任意的$x_0 \in D$,由算法产生的点列都收敛x^*,则称该算法为具有全局收敛的算法.

由于一般情况下最优解x^*是未知的,所以只有具有全局收敛性的算法才有实用意义.但算法的局部收敛性分析,在理论上是重要的,因为它是全局收敛性分析的基础。

收敛速度

定义1.2.3 设序列$\{x_k\}$收敛于x^*,而且

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}=\beta,$$

若$0<\beta<1$,则称$\{x_k\}$为线性收敛的,称β为收敛比;

若$\beta=1$,则称$\{x_k\}$为超线性收敛的.

定义1.2.4 设序列$\{x_k\}$收敛于x^*,而且

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|^p}=\beta,$$

则称$\{x_k\}$为p阶收敛.

终止准则

对于一种算法,应该有某种终止准则,当某次迭代满足终止准则时,就停止迭代.常用的终止准则有:

(1) $\|x_{k+1}-x_k\|<\varepsilon$ 或 $\dfrac{\|x_{k+1}-x_k\|}{\|x_k\|}<\varepsilon$;

(2) $|f(x_{k+1})-f(x_k)|<\varepsilon$ 或 $\dfrac{|f(x_{k+1})-f(x_k)|}{|f(x_k)|}<\varepsilon$;

(3) $\|\nabla f(x_k)\|=\|g_k\|<\varepsilon$;

(4) 上面三种准则的组合.

注:其中$\varepsilon>0$是预先给定的.

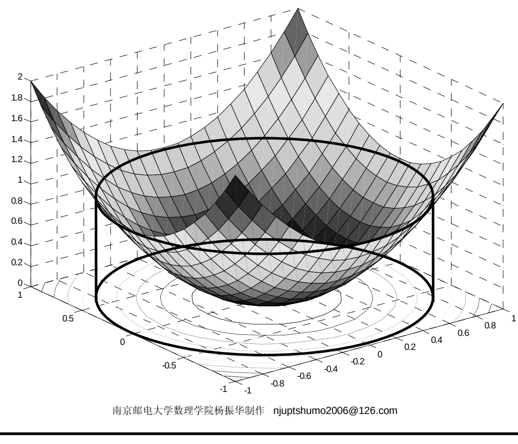
§1.3 二维最优化问题的几何解释

理论分析

二维最优化问题的目标函数$z=f(x_1,x_2)$表示三维空间R^3中的曲面.在空间直角坐标系$O\text{-}x_1x_2z$中,平面$z=c$与曲面$z=f(x_1,x_2)$ 的交线在$0\text{-}x_1x_2$平面上的投影曲线为:

$$\begin{cases} z=0 \\ f(x_1,x_2)=c \end{cases}$$

取不同的c值得到不同的投影曲线,每一条投影曲线对应一个c值,称投影曲线为目标函数的等值线或等高线.



理论分析

求目标函数$z=f(x_1,x_2)$ 在可行域D上的极小点,是在与可行域D有交集的等值线中找出具有最小值的等值线.也就是在可行域D上沿着$f(x_1,x_2)$的负梯度方向或某种下降方向上找取得最小值c的点.

例1.3.1

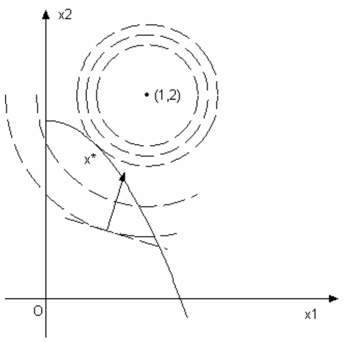
$$\min (x_1-1)^2+(x_2-2)^2$$
$$s.t.\quad x_1^2+x_2-\frac{7}{4}\le 0$$
$$x_1,x_2\ge 0$$

解 首先画出可行域D,目标函数的等值线是以点(1,2)为圆心的一族圆. $f(x_1,x_2)$的梯度为

$$\nabla f(x_1,x_2)=(2(x_1-1),2(x_2-2))^T$$

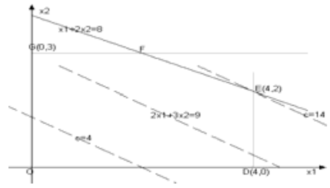
例1.3.1

负梯度方向指向等值线圆心,所以等值线与可行域D的边界相切的点$x^*=(1/2,3/2)^T$是此问题的最优解,目标函数的最优值为1/2.



例1.3.2

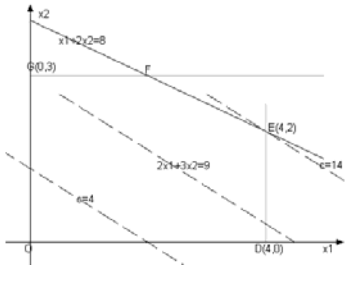
$$\max 2x_1+3x_2$$
$$s.t.\quad x_1+2x_2\le 8$$
$$x_1\le 4$$
$$x_2\le 3$$
$$x_1,x_2\ge 0$$



解 首先画出可行域D的图形.D为凸多边形ODEFGO.再以c为参数画出目标函数的等值线$2x_1+3x_2=c$.

例1.3.2

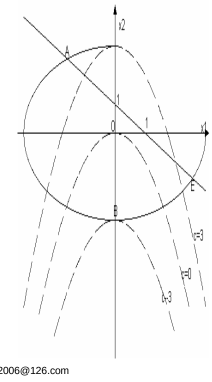
目标函数c的值由小到大逐渐增加,等值线沿着目标函数的梯度方向平行移动.当移动到点E时,再移动就与可行域D不相交了,所以顶点E就是最优点,最优值为14.



例1.3.3

$$\min x_1^2+x_2$$
$$s.t.\quad -x_1-x_2+1\ge 0$$
$$x_1^2+x_2^2-9=0$$

解 如图所示,可行域只能是圆弧ABE,其中点A和点E是等值线$-x_1-x_2+1=0$和圆$x_1^2+x_2^2-9=0$的交点.注意到等值线是平行的抛物线,图中画出了几条目标函数的等值线.容易看出B点是最优点,所以最优解是$(0,-3)^T$,最优值为-3.



§1.4 一维搜索

问题描述

已知x_k,并且求出了x_k处的可行下降方向p_k ,从x_k出发,沿方向p_k求目标函数的最优解,即求解问题:
$$\min_{\alpha>0} f(x_k+\alpha p_k)=\min_{\alpha>0}\phi(\alpha)$$
设其最优解为α_k,于是得到一个新点
$$x_{k+1}=x_k+\alpha_k p_k$$
所以一维搜索是求解一元函数$\phi(\alpha)$ 的最优化问题(也叫一维最优化问题).我们来求解
$$\min_{a\le x\le b} f(x)$$

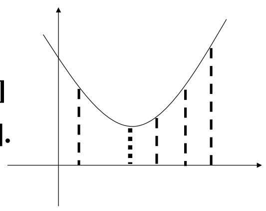
1.4.1 Fibonacci法与黄金分割法

设$f(x)$在$[a,b]$上为下单峰函数,即有唯一的极小点x^*,在x^*左边$f(x)$ 严格下降,在x^*右边 $f(x)$严格上升.

在$[a,b]$内任取$x_1<x_2$,

若$f(x_1) < f(x_2)$,则$x^*\in[a,x_2]$

若$f(x_1)\geqslant f(x_2)$,则$x^*\in[x_1,b]$.



Fibonacci方法

如果只有一个试点x,我们无法将区间缩小.

如果知道两个试点$x_1< x_2$,根据$f(x_1)$,$f(x_2)$的大小关系,可以得到缩小的区间$[a,x_2]$或$[x_1,b]$.

我们的目的是使$\max(x_2-a,b-x_1)$尽量小.

由于$x_2-a+b-x_1= b-a+x_2-x_1>b-a$,

因此$\max(x_2-a,b-x_1)>(b-a)/2$.

我们尽量靠近$[a,b]$中点来取点.一般取

$x_1=(b-a)/2$, $x_2=x_1+0.1(b-a)$,

或者$x_2=(b-a)/2$,$x_1=x_2-0.1(b-a)$.

Fibonacci法

下面我们考虑任给一个$f(x)$的单峰区间$[a,b]$,如果给定试点的个数n,如何使最后确定的包含最优值的区间尽量的小.

另一种思维方式为,按什么方式取点,求n次函数值之后,可最多将多长的原始区间长度缩小为1.设L_n为试点个数为n,最终区间长度为1时,原始区间$[a,b]$的最大可能长度。

Fibonacci法

设最初两个试点为x_1和x_2,若极小点在$[a,x_1]$内,至多还有$n-2$个试点,则$x_1-a\leqslant L_{n-2}$.

若极小点在$[x_1,b]$内,包括x_2在内可以有$n-1$个试点,则$b-x_1\leqslant L_{n-1}$.

因此, $L_n\leqslant L_{n-1}+L_{n-2}$.

如果我们采取合适的技巧,可以使得$L_n=L_{n-1}+L_{n-2}$.

另外,显然有$L_0=L_1=1$.

Fibonacci数列

从而L_n满足差分方程: $\begin{cases} L_n = L_{n-1} + L_{n-2} & (n \ge 2) \\ L_0 = L_1 = 1 \end{cases}$

此为Fibonacci数列,一般写为:

$$\begin{cases} F_n = F_{n-1} + F_{n-2} & (n \ge 2) \\ F_0 = F_1 = 1 \end{cases}$$

Fibonacci数列的前几个数据为:

Fibonacci法

若原始区间为[a,b],要求最终的区间长度≤ε,则F_n≥$(b-a)/\varepsilon$,由此可确定n,区间缩短之后与之前的比依次为: $\dfrac{F_{n-1}}{F_n}, \dfrac{F_{n-2}}{F_{n-1}}, \cdots, \dfrac{3}{5}, \dfrac{2}{3}, \dfrac{1}{2}$

n确定之后,最初两个试点(关于[a,b]对称)分别为:

$$x_1 = a + \frac{F_{n-2}}{F_n}(b-a), \quad x_2 = a + \frac{F_{n-1}}{F_n}(b-a)$$

Fibonacci法

上述过程完成了一次迭代,新区间仍记为[a,b].

若已经进行了$i-1$次迭代,第i次迭代时,还有$n-i+1$个试点(包括已计算过的一个函数值).

$$x_1 = a + \frac{F_{n-i-1}}{F_{n-i+1}}(b-a), \quad x_2 = a + \frac{F_{n-i}}{F_{n-i+1}}(b-a)$$

注:(1)最后的两个试点取的方式:

$x_1=(b-a)/2, x_2=x_1+0.1(b-a)$.

(2)若$f(x_1)$与 $f(x_2)$近似相等,则认为x^*在[x_1,x_2]内.

例1.4.1(Fibonacci方法)

用Fibonacci法求函数$f(x)=x^2-x+2$在区间[-1,3]上的极小点.要求最终区间长度不大于原始区间长度的0.08倍。

解　函数$f(x)$在区间[-1,3]上为下单峰函数,

由F_n≥1/0.08=12.5,可知n应取6(F_6=13).

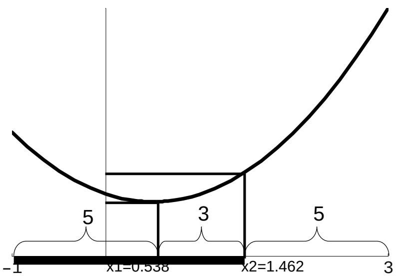
第一次迭代

(a= –1,b=3)

$x_1 = a + \dfrac{F_4}{F_6}(b-a)$

$= -1 + \dfrac{5}{13} \cdot 4$

$= 0.538$



$x_2 = a + \dfrac{F_5}{F_6}(b-a) = -1 + \dfrac{8}{13} \bullet 4 = 1.462$

$f_1 = 1.751, \ f_2 = 2.675$

$f_1 < f_2$,缩短后区间为[-1,1.462]

第二次迭代

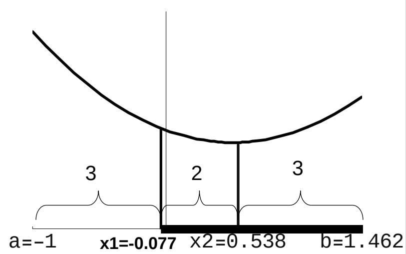
(a= –1,b=1.462)

$x_2 = 0.538,$

$f_2 = 1.751$

$x_1 = -1 + \dfrac{F_3}{F_5}(1.462+1)$

$= -0.077$



$f_1 = 2.083 \qquad f_1 > f_2$　新区间为[-0.077,1.462]

第三次迭代

(a=-0.077,b=1.462)

$x_1 = 0.538,\ f_1 = 1.751$

$$x_2 = -0.077 + \frac{F_3}{F_4}(1.462 + 0.077) = 0.846,\ f_2 = 1.870$$

$f_1 < f_2$ **新区间为** $[-0.077, 0.846]$

第四次迭代

(a=-0.077,b=0.846)

$x_2 = 0.538,\ f_2 = 1.751$

$$x_1 = -0.077 + \frac{F_1}{F_3}(0.846 + 0.077) = 0.231,\ f_2 = 1.822$$

$f_1 > f_2$ **新区间为** $[0.231, 0.846]$

第五次迭代

(a=0.231,b=0.846)

$x_2 = 0.538,\ f_2 = 1.751$

$$x_1 = x_2 - 0.1 \times (0.846 - 0.231) = 0.477,\ f_1 = 1.751$$

$f_1 = f_2$ **取最优解** $x^* = \dfrac{x_1 + x_2}{2} = 0.508$

Fibonacci方法评价

Fibonacci方法的优点

(1)如果缩小的区间包含原来的试点,则该试点在下一步被利用

(2)效率最高,有限个试点的情况下,将最优点所在的区间缩小到最小.

Fibonacci方法的缺点

(1)搜索前先要计算搜索的步数

(2)每次搜索试点计算的公式不一致

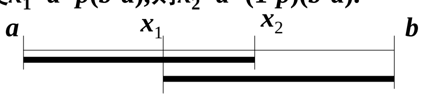
黄金分割法

我们希望保留Fibonacci方法的优点(效率最高是不可能保留的),改进其缺点.

若第一次选取的试点为$x_1<x_2$,则下一步保留的区间为$[a,x_2]$或$[x_1,b]$,两者的机会是均等的.

因此我们选取试点时希望$x_2-a=b-x_1$.

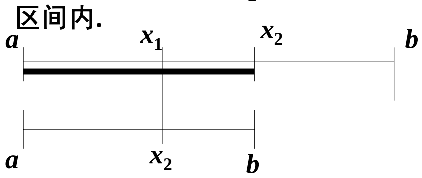
设$x_1=a+p(b-a)$,则$x_2=a+(1-p)(b-a)$.



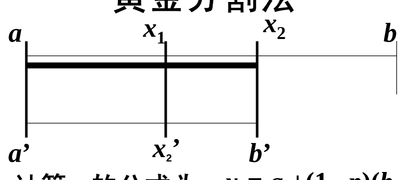
黄金分割法

另外,我们希望如果缩小的区间包含原来的试点,则该试点在下一步被利用.

若保留的区间为$[a,x_2]$,前一次的试点x_1在这个区间内.



黄金分割法



我们希望原来的x_1,在缩小的区间内成为新的"x_2".

我们根据这一条件来计算p.

计算x_2的公式为 $x_2 = a + (1-p)(b-a)$.

因此我们希望 $x'_2 = a' + (1-p)(b' - a')$.

即 $a+p(b-a)=a+(1-p)(a+(1-p)(b-a)-a)$

化简得 $p^2-3p+1=0$ $p = \dfrac{3-\sqrt{5}}{2} \approx 0.382,\ 1-p \approx 0.618$

黄金分割法

若保留区间为$[x_1,b]$,我们得到的结果是一致的.
该方法称为黄金分割法,实际计算取近似值:
$x_1=a+0.382(b-a)$, $x_2=a+0.618(b-a)$,
所以黄金分割法又称为0.618法.
黄金分割法每次缩小区间的比例是一致的,每次将区间长度缩小到原来的0.618倍.

算法1.4.2 黄金分割法

给定$a,b(a<b)$以及$\varepsilon>0$,
step 1 令$x_2=a+0.618(b-a)$, $f_2=f(x_2)$;
step 2 令$x_1=a+0.382(b-a)$, $f_1=f(x_1)$;
step 3 若$|b-a|<\varepsilon$, 则 $x^*=(a+b)/2$,Stop.
step 4 若$f_1<f_2$, 则$b=x_2,x_2=x_1,f_2=f_1$,转step 2;
若$f_1=f_2$, 则$a=x_1,b=x_2$,转step 1;
若$f_1>f_2$, 则$a=x_2,x_1=x_2,f_1=f_2$,转step 5;
step 5 令$x_2=a+0.618(b-a)$，$f_2=f(x_2)$,转step3.

用0.618法求解例1.4.1的数据表

迭代次数	$[a,b]$	x_1	x_2	f_1	f_2	$\|b-a\|<\varepsilon$
0	[-1,3]	0.528	1.472	1.751	2.695	否
1	[-1,1.472]	-0.056	0.528	2.059	1.751	否
2	[-0.056,1.472]	0.528	0.888	1.751	1.901	否
3	[-0.056,0.888]	0.305	0.528	1.788	1.751	否
4	[0.305,0.888]	0.528	0.665	1.751	1.777	否
5	[0.305,0.665]	0.443	0.528	1.753	1.751	否
6	[0.443,0.665]	0.528	0.580	1.751	1.757	是
最优解$x^*=(0.443+0.665)/2=0.554$						

0.618法和Fibonacci之间的关系

$$\lim_{n\to\infty}\frac{F_{n-1}}{F_n}=\frac{\sqrt{5}-1}{2}\approx 0.618$$

0.618法为Fibonacci法的极限形式.

0.618法和Fibonacci之间的关系

迭代步数的比较

0.618法: $(\frac{\sqrt{5}-1}{2})^{n-1}\le\frac{\varepsilon}{b-a}$ $\quad n\ge\frac{\ln(\frac{\varepsilon}{b-a})}{\ln\frac{\sqrt{5}-1}{2}}+1=\frac{\ln(\frac{b-a}{\varepsilon})}{\ln\frac{\sqrt{5}+1}{2}}+1$

Fibonacci方法: $F_n\ge\frac{b-a}{\varepsilon}$

$$\frac{1}{\sqrt{5}}((\frac{1+\sqrt{5}}{2})^{n+1}-(\frac{1-\sqrt{5}}{2})^{n+1})\ge\frac{b-a}{\varepsilon}$$

忽略 $(\frac{1-\sqrt{5}}{2})^{n+1}$

得到 $\frac{1}{\sqrt{5}}(\frac{1+\sqrt{5}}{2})^{n+1}\ge\frac{b-a}{\varepsilon}$

$$n\ge\frac{\ln(\frac{b-a}{\varepsilon})}{\ln\frac{\sqrt{5}+1}{2}}+\frac{\ln\sqrt{5}}{\ln\frac{\sqrt{5}+1}{2}}-1\approx\frac{\ln(\frac{b-a}{\varepsilon})}{\ln\frac{\sqrt{5}+1}{2}}+0.672$$

黄金分割法至多比Fibonacci法多一步

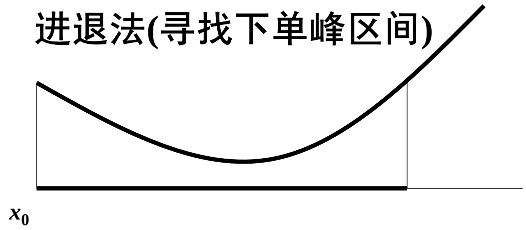
进退法(寻找下单峰区间)

在一维搜索之前,必须先知道一个$f(x)$的下单峰区间.书中的算法1.4.3给出了求函数的一般的下单峰区间的算法.此处我们对算法1.4.3加以改进,求出$f(x)$的一个形如$[0,b]$形式的下单峰区间
因为我们关心的问题是:

$$\min_{\alpha>0} f(x_k+\alpha p_k)=\min_{\alpha>0}\phi(\alpha)$$

我们的目的是找出两个点$x_1<x_2$,使得
$f(x_1)\leqslant f(x_2)$,$f(x_1)\leqslant f(0)$.

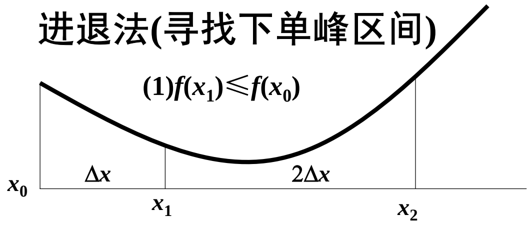
进退法(寻找下单峰区间)



给定初始点x_0=0,初始步长Δx>0,$x_1=x_0+\Delta x$.
下面分两种情况讨论.

(1)$f(x_1) \leqslant f(x_0)$
x_1对应着图上用红线标出的一部分

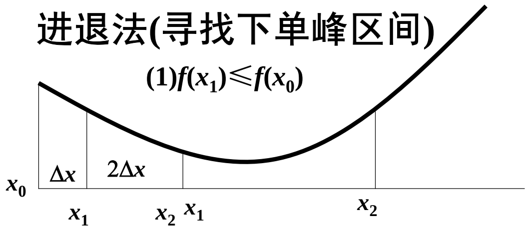
进退法(寻找下单峰区间)



此时x_1取值小,我们加大步长向右搜索,取
$\Delta x=2\Delta x, x_2=x_1+\Delta x$

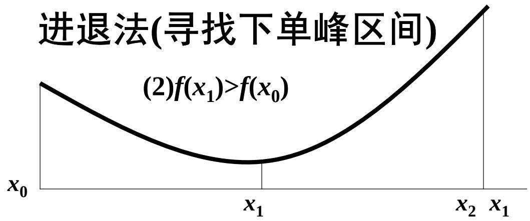
若$f(x_1) \leqslant f(x_2)$,则我们要找的区间即为$[x_0,x_2]$

进退法(寻找下单峰区间)



若$f(x_1)>f(x_2)$,则我们所取的步长偏小.
令$x_1=x_2$, $\Delta x=2\Delta x, x_2=x_1+\Delta x$
继续往下判断,直到满足$f(x_1) \leqslant f(x_2)$.

进退法(寻找下单峰区间)



此时x_1取值大,我们缩小步长向x_1左边搜索,
取$\Delta x=\Delta x/2$, $x_2=x_1$, $x_1=x_2 - \Delta x$

若$f(x_1) \leqslant f(x_0)$,则我们要找的区间即为$[x_0,x_2]$
否则继续缩小区间,直到满足$f(x_1) \leqslant f(x_0)$.

1.4.2 二分法

若$f(x)$的导数存在且容易计算,则线性搜索的速多可以得到提高,下面的二分法每次将区间缩小至原来的二分之一.

设$f(x)$为下单峰函数,若$f(x)$在[a,b]具有连续的一阶导数,且$f'(a)<0, f'(b)>0$

取 $c=(a+b)/2$,若$f'(c)=0$,则c为极小点;

若$f'(c)>0$,则以$[a,c]$代替$[a,b]$作为新区间;

若$f'(c)<0$,则以$[b,c]$代替$[a,b]$作为新区间.

1.4.3 抛物线法

在求一元函数的极小点问题上,我们可以利用若干点处的函数值来构造一个多项式,用这个多项式的极小点作为原来函数极小点的近似值.

抛物线法就是一个用二次函数来逼近$f(x)$的方法,这也是我们常说的二次插值法.

设在已知的三点$x_1<x_0<x_2$处对应的函数值$f(x_i)=f_i$,且满足：$f_1>f_0, f_0<f_2$

过三点$(x_1,f_1),(x_0,f_0),(x_2,f_2)$作二次函数$y=\varphi(x)$,即作一条抛物线,则可推导出：

$$\varphi(x)=\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f_1+\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f_0+\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}f_2$$

为求$\varphi(x)$的极小点,令$\varphi'(x)=0$,得:

$$\bar{x}=\frac{1}{2}\bullet\frac{(x_2^2-x_0^2)f_1+(x_1^2-x_2^2)f_0+(x_0^2-x_1^2)f_2}{(x_2-x_0)f_1+(x_1-x_2)f_0+(x_0-x_1)f_2}$$

若 充分接近x_0,即对于预先给定的精度$\varepsilon>0$,有$|x_0-\bar{x}|<\varepsilon$,则把$\bar{x}$作为近似极小点.

否则计算$f(\bar{x})=\bar{f}$,找出f_0和$\bar{f}$之间的大者,去掉x_1或x_2,使新的三点仍具有两端点的函数值大于中间点的函数值的性质.利用新的点再构造二次函数,继续进行迭代.

1.4.4 不精确的一维搜索

前面介绍的得几种一维搜索方法,都是为了获得一元函数$f(x)$的最优解,所以习惯上称为精确一维搜索.

在解非线性规划问题中,一维搜索一般很难得到真正的精确值.

因此,不精确的一维搜索开始为人们所重视.

即在x_k点确定了下降方向p_k后,只计算少量的几个函数就可得到一个满足$f(x_{k+1})<f(x_k)$的近似点x_{k+1}.

不精确的一维搜索

对于不精确的一维搜索,要求产生的点列具有某种收敛性.所以除了对下降方向p_k有要求之外,对步长α_k也有要求,即要求目标函数要"充分的下降"

不精确一维搜索的Wolfe原则

设$f(x)$可微,取$\mu\in(0,1/2)$, $\sigma\in(0,\mu)$,选取$\alpha_k>0$, 使

$$f(x_k)-f(x_k+\alpha_k p_k)\ge-\mu\alpha_k g_k^T p_k \quad (1.6)$$

$$\nabla f(x_k+\alpha_k p_k)^T p_k\ge\sigma g_k^T p_k \quad (1.7)$$

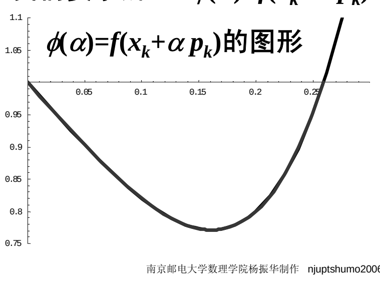
或用下面更强的条件代替(1.7)式:

$$|\nabla f(x_k+\alpha_k p_k)^T p_k|\le-\sigma g_k^T p_k \quad (1.8)$$

Wolfe原则的解释

给定函数$f(x)$,迭代点x_k,下降方向p_k.

我们要求解$\min\phi(\alpha)=f(x_k+\alpha p_k)$的近似极小值.



我们首先要求$f(x_k+\alpha_k p_k)<f(x_k)$,即$\phi(\alpha_k)<\phi(0)$

α_k的取值对应与曲线上红色的部分.

Wolfe原则的解释

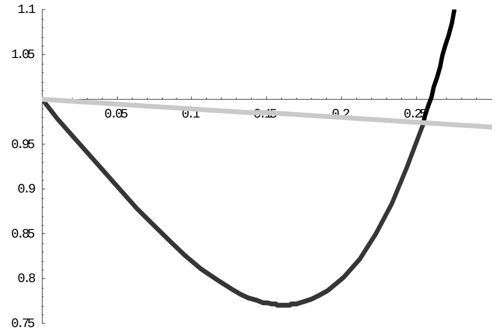
在实际计算中,我们不仅要求函数值下降,而且对下降的量有一定的要求.

如果$\phi'(0)$ "负得比较多",我们希望下降得多,另外,如果α_k取得比较大,我们也希望下降量大.

通常的要求是原则一

$\phi(\alpha_k) \leqslant f(x_k) + \mu\, \alpha_k\, \phi'(0)\ (\mu > 0)$

从图形上看, α_k 的合理取值对应曲线上位于一条直线下方的部分.

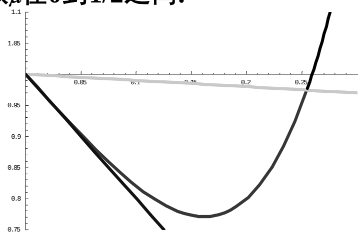


$\mu=1$时,对应的直线是曲线在$\alpha=0$处的切线.

为了保证直线在曲线的上方, μ的取值应小于1.

另外,如果直线偏下(μ的取值接近于1),可能导致α的取值范围过小,不利于一维搜索.

通常取μ在0到1/2之间.



Wolfe原则的解释

从图形可以看出,在我们上面所要求的α的范围内,函数值的下降还有可能比较小.

我们还希望达到比较高的下降效率.

如果$\phi(\alpha_k)$的导数"负得比较大",说明函数值还可以继续下降.

因此我们通常还要求$\phi(\alpha)$有一个下界,此下界通常比$\phi'(0)$小一些,我们要求

$\phi'(\alpha_k) \geqslant \sigma\, \phi'(0)(0 < \sigma < 1)$.

Wolfe原则的解释

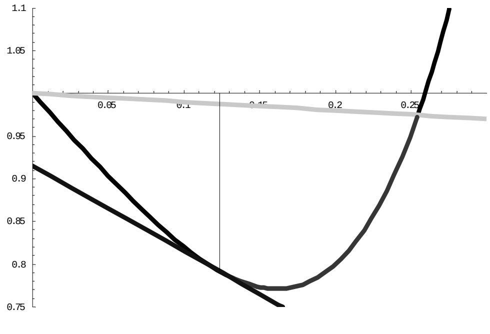
因此我们又有如下的原则二

$\phi'(\alpha_k) \geqslant \sigma\, \phi'(0)(0 < \sigma < 1)$

另外,如果σ的取值的取值过小,则α_k的取值范围过小,同样不利于一维搜索(已接近于精确搜索).

通常取σ在μ与1之间.

由于原则二的限制,α_k的取值范围缩小.



Wolfe原则的解释

从上面的图形还可以看出,在α_k的容许取值范围内,还有一段曲线上升的部分.可能导致函数值下降不大.

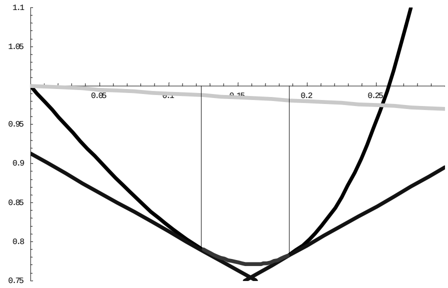
因此,我们有时还要求$\phi'(\alpha_k)$ 的值不能"正"得太大.

于是原则二改为

$|\phi'(\alpha_k)| \leqslant -\sigma\phi'(0)$ $(\mu < \sigma < 1)$

利用改进的原则二,函数值下降的效果可以让人接受.

注意:原则一是不可缺少的,它严格地划定了原则二搜索的范围.



Wolfe原则的解释

原则一 $\phi(\alpha_k) \leqslant f(x_k) + \mu\,\alpha_k\,\phi'(0)$ $(0<\mu<1/2)$

原则二 $\phi'(\alpha_k) \geqslant \sigma\,\phi'(0)$ $(\mu<\sigma<1)$

改进的原则二$|\phi'(\alpha_k)| \leqslant -\sigma\phi'(0)$ $(\mu<\sigma<1)$

由$\phi(\alpha)=f(x_k+\alpha\,p_k)$可知$\phi'(\alpha_k) = \nabla f(x_k+\alpha_k p_k)^T p_k$

于是得到

原则一 $f(x_k+\alpha_k x_k) \leqslant f(x_k) + \mu\,\alpha_k \nabla f_k^T p_k$

原则二$\nabla f(x_k+\alpha_k p_k)^T p_k \geqslant \sigma\nabla f_k^T p_k$

改进的原则二$|\nabla f(x_k+a_k p_k)^T p_k| \leqslant -\sigma\nabla f_k^T p_k$

Wolfe原则

关于满足Wolfe原则的步长α_k的存在性:

定理1.4.2 设$f(x)$有下界且$g_k^T p_k<0$.令$\mu\in(0,1/2)$, $\sigma\in(0,\mu)$, 则存在区间$[c_1,c_2]$,使得任意的$\alpha\in[c_1,c_2]$均满足式(1.6)和(1.7)(也满足(1.8)).

不精确一维搜索Wolfe算法

问题:设已知x_k和下降方向p_k,求问题

$$\min_{\alpha>0} f(x_k+\alpha\,p_k)$$

的近似值α_k,使α_k 满足(1.6)和(1.7).

算法1.4.6 不精确一维搜索Wolfe算法

step 1 给定$\mu\in(0,1)$, $\sigma\in(\mu,1)$,

令$a=0$, $b=\infty$, $\alpha=1$, $j=0$;

step2 $x_{k+1}=x_k+\alpha\,p_k$,计算f_{k+1}, g_{k+1};

若α满足(1.6)和(1.7)式,则令$\alpha_k=\alpha$;

若α不满足(1.6)式,则令$k:=k+1$,转step 3;

若α不满足(1.7)式,则令$k:=k+1$,转step 4;

step 3 令 $b=\alpha, \alpha=\dfrac{a+\alpha}{2}$ 转step 2

step 4 令 $a=\alpha, \alpha=\min\{2\alpha,\dfrac{\alpha+b}{2}\}$ 转step 2.

Wolfe准则(例1.4.2)

用不精确一维搜索求Rosenbrock函数

$f(x)=100(x_2-x_1^2)^2+(1-x_1)^2$

在点$x_k=(0,0)^T$处沿方向$p_k=(1,0)^T$的近似步长α_k.

解
$$\nabla f(x)=\begin{pmatrix}-400(x_2-x_1^2)x_1-2(1-x_1)\\ 200(x_2-x_1^2)\end{pmatrix}$$

$$f_k=f(0,0)=1,\ g_k=(-2,0)^T,\ g_k^T p_k=-2$$

step 1 给定$\mu=0.1,\sigma=0.5$,

令$a=0,\ b=\infty,\ \alpha=1,\ j=0$;

step 2 $x_{k+1}=x_k+\alpha p_k=(1,0)^T,\ f_{k+1}=f(1,0)=100$

因为$f_k-f_{k+1}=1-100=-99<-\mu\alpha g_k^T p_k=0.2$

所以(1.6)式成立。转step 3

step 3 令$b=1,\alpha=(a+\alpha)/2=(1+0)/2=0.5$,转step 2.

重新计算x_{k+1}.

迭代四次得到满足Wolfe条件的步长$\alpha_k=0.125$

$x_{k+1}=x_k+\alpha_k p_k=(0.125,1)^T$.

总结

求初始区间的方法 ⇒ 进退法 (算法1.4.3)

一维搜索的方法
- 利用导数的方法 ⇒ 二分法 (算法1.4.4)
- 仅用函数值的方法
  - 精确方法
    - Fibonacci方法(算法1.4.1)
    - 黄金分割法 (算法1.4.2)
    - 二次插值法 (算法1.4.5)
  - 不精确方法 ⇒ Wolfe算法 (算法1.4.6)

第二章

线性规划

§2.1 凸集与凸函数

凸 集

定义2.1.1 设集合$D\subset R^n$,若对于任意点$x,y\in D$,

及实数$\alpha,0\leqslant\alpha\leqslant1$,都有

$$\alpha x+(1-\alpha)y\in D,$$

则称集合D为凸集.

常见的凸集:空集(补充定义),整个欧式空间R^n,

超平面 $H=\{x\in R^n|a_1x_1+a_2x_2+\cdots a_nx_n=b\}$

半空间 $H^+=\{x\in R^n|a_1x_1+a_2x_2+\cdots a_nx_n\geqslant b\}$

凸集的例

例2.1.2 超球$\|x\|\leqslant r$为凸集

证明 设x,y为超球中任意两点, $0\leqslant\alpha\leqslant 1$,则有

$\|\alpha x+(1-\alpha)y\|\leqslant\alpha\|x\|+(1-\alpha)\|y\|$

$\leqslant\alpha r+(1-\alpha)r = r$,

即点$\alpha x+(1-\alpha)y$属于超球,所以超球为凸集.

凸集的性质

(i)有限个(可以改成无限)凸集的交集为凸集.

即:若$D_j(j\in J)$是凸集,则它们的交集

$D=\{x|x\in D_j, j\in J\}$

是凸集.

(ii)设D是凸集,β是一实数,则下面集合是凸集

$$\beta D=\{y\mid y=\beta x, x\in D\}.$$

凸集的性质

(iii)设D_1,D_2是凸集,则D_1与D_2的和集

$D_1+D_2=\{y|y=x+z, x\in D_1, z\in D_2\}$是凸集.

注:和集与并集有很大的区别,凸集的并集未必是凸集, 而凸集的和集是凸集.

例:$D_1=\{(x,0)^T|x\in R\}$表示 x 轴上的点,

$D_2=\{(0,y)^T|y\in R\}$,表示 y 轴上的点.

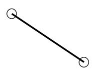
则$D_1\cup D_2$表示两个轴的所有点,它不是凸集;

$D_1+D_2=R^2$是凸集

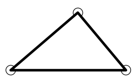
推论 凸集的线性组合是凸集.

定义2.1.2 设$x_i\in R^n, i=1,\cdots,k$,实数$\lambda_i\geqslant 0$, $\sum_{i=1}^{k}\lambda_i=1$,

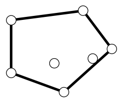
则 $x=\sum_{i=1}^{k}\lambda_i x_i$ 称为$x_1,x_2,\cdots,x_k$的凸组合.



两点的凸组合



三点的凸组合



多点的凸组合

容易证明:凸集中任意有限个点的凸组合仍然在该凸集中.

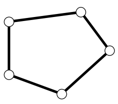
极 点

定义2.1.3 设D为凸集, $x\in D$.若D中不存在两个相异的点y,z及某一实数$\alpha\in(0,1)$使得

$x=\alpha y+(1-\alpha)z$

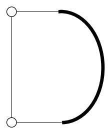
则称x为D的极点.

凸集



极点

凸集



极点

极 点

例 $D=\{x\in R^n|\ \|x\|\leqslant a\}(a>0)$,则$\|x\|=a$上的点均为极点

证明:设$\|x\|=a$,若存在$y,z\in D$及$\alpha\in(0,1)$,使得$x=\alpha y+(1-\alpha)z$.则

$a^2=\|x\|^2=\langle\alpha y+(1-\alpha)z,\alpha y+(1-\alpha)z\rangle$

$\leqslant\alpha^2\|y\|^2+(1-\alpha)^2\|z\|^2+2\alpha(1-\alpha)\|y\|\|z\|$

$\leqslant a^2$

不等式要取等号,必须$\|y\|=\|z\|=a$,且$\langle y,z\rangle=\|y\|\|z\|$,容易证明$y=z=x$,根据定义可知,$x$为极点.

凸 函 数

定义2.1.4 设函数$f(x)$定义在凸集$D\subset R^n$上,若对任意的$x,y\in D$,及任意的$\alpha\in[0,1]$都有
$f(\alpha x+(1-\alpha)y)\leqslant\alpha f(x)+(1-\alpha)f(y)$
则称函数$f(x)$为凸集D上的凸函数.

凸 函 数

定义2.1.5 设函数$f(x)$定义在凸集$D\subset R^n$上,若对任意的$x,y\in D,x\neq y$,及任意的$\alpha\in(0,1)$都有
$f(\alpha x+(1-\alpha)y)<\alpha f(x)+(1-\alpha)f(y)$
则称函数$f(x)$为凸集D上的严格凸函数.

将上述定义中的不等式反向,可以得到凹函数和严格凹函数的定义.

凸函数的例

例2.1.3 设$f(x)=(x-1)^2$,试证明$f(x)$在$(-\infty,+\infty)$上是严格凸函数.

证明:设$x,y\in R$,且$x\neq y$, $\alpha\in(0,1)$都有

$f(\alpha x+(1-\alpha)y)-(\alpha f(x)+(1-\alpha)f(y))$
$=(\alpha x+(1-\alpha)y-1)^2-\alpha(x-1)^2-(1-\alpha)(y-1)^2$
$=-\alpha(1-\alpha)(x-y)^2<0$

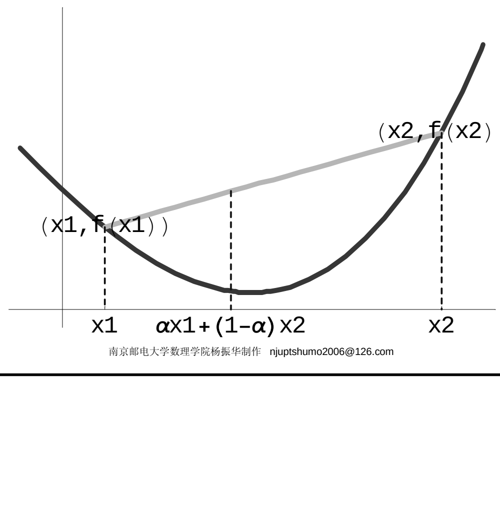
因此$f(x)$在$(-\infty,+\infty)$上是严格凸函数.

例2.1.4 线性函数$f(x)=c^Tx=c_1x_1+c_2x_2+\cdots+c_nx_n$既是$R^n$上凸函数也是$R^n$上凹函数.

凸函数的几何性质

对一元函数$f(x)$,在几何上$\alpha f(x_1)+(1-\alpha)f(x_2)$ $(0\leqslant\alpha\leqslant1)$表示连接$(x_1,f(x_1)),(x_2,f(x_2))$的线段, $f(\alpha x_1+(1-\alpha)x_2)$表示在点$\alpha x_1+(1-\alpha)x_2$处的函数值,所以一元凸函数表示连接函数图形上任意两点的线段总是位于曲线弧的上方.

凸函数的几何性质



上 图

对于一元凸函数$f(x)$,可以发现,位于函数曲线上方的图形是凸集.事实上这一结论对于多元函数也是成立的,而且是充要条件,即有下面的定理.

定理:设$f(x)$是定义在凸集$D\subset R^n$上的函数,则$f(x)$是凸函数的充要条件是其上图epi(f)为凸集,其中epi(f)=$\{(x,y)|x\in D,y\in R,y\geqslant f(x)\}$.

证明:作业

凸函数的性质

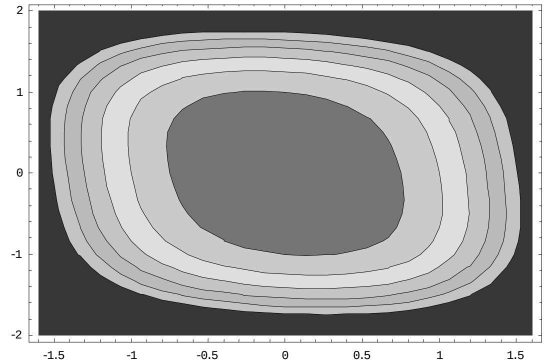
(i)设$f(x)$是凸集$D\subset R^n$上的凸函数,实数$k\geqslant 0$,则$kf(x)$也是D上的凸函数.

(ii)设$f_1(x)$, $f_2(x)$是凸集$D\subset R^n$上的凸函数,实数$\lambda,\mu\geqslant 0$,则$\lambda f_1(x)+\mu f_2(x)$也是$D$上的凸函数.

(iii)设$f(x)$是凸集$D\subset R^n$上的凸函数,β为实数,则水平集$S(f,\beta)=\{x|x\in D,f(x)\leqslant\beta\}$.

下面的图形给出了凸函数$f(x,y)=x^4+3x^2+y^4+y^2+xy$的等值线($f(x,y)$=2,4,6,8,10,12)的图形.可以看出水平集为凸集.

凸函数的性质



凸函数的判断

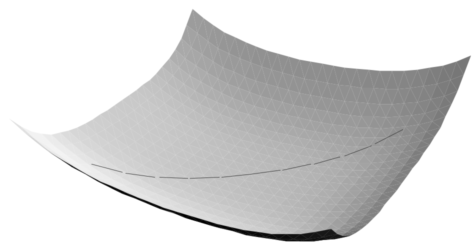
定理2.1.1 设$f(x)$定义在凸集$D\subset R^n$上,$x,y\in D$.
令$\Phi(t)=f(tx+(1-t)y)$, $t\in[0,1]$,则

(i) $f(x)$是凸集D上的凸函数的充要条件是对任意的$x\in D$,一元函数$\Phi(t)$为[0,1]上的凸函数.

(ii) $f(x)$是凸集D上的凸函数的充要条件是对任意的$x,y\in D(x\neq y)$,一元函数$\Phi(t)$为[0,1]上的严格凸函数.

该定理的几何意义是:凸函数上任意两点之间的部分是一段向下凸的弧线.

凸函数的判断



一阶条件

定理2.1.2 (一阶条件)

设在凸集$D\subset R^n$上$f(x)$可微,则$f(x)$在D上为凸函数的充要条件是对任意的$x,y\in D$,都有

$f(y)\geqslant f(x)+\nabla f(x)^T(y-x)$

定理2.1.3 (一阶条件)

设在凸集$D\subset R^n$上$f(x)$可微,则$f(x)$在D上为凸函数的充要条件是对任意的$x,y\in D$, $x\neq y$,都有

$f(y)>f(x)+\nabla f(x)^T(y-x)$

二阶条件

设在开凸集$D\subset R^n$上$f(x)$可微,则

(i) $f(x)$是D内的凸函数的充要条件为,在D内任一点x处, $f(x)$的Hesse矩阵$G(x)$半正定,其中

$$G(x)=\nabla^2 f(x)=\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\partial x_1} & \frac{\partial^2 f}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{vmatrix}$$

(ii) 若在D内$G(x)$正定,则$f(x)$在D内是严格凸函数.

凸规划

定义2.1.6 设$D\subset R^n$为凸集,则$f(x)$为D上的凸函数,则称规划问题

min $f(x)$

s.t. $x \in D$

为凸规划问题.

定理2.1.5 (i)凸规划的任一局部极小点x是整体极小点,全体极小点组成凸集.

(ii)若$f(x)$是$D\subset R^n$上的严格凸函数,且凸规划问题

min $f(x)$

s.t. $x \in D$

的整体极小点存在,则整体极小点唯一.

定理2.1.5证明(思路)

(i)x^*为局部极小点,若存在x_0使得$f(x_0)<f(x^*)$,则$f(t x^*+(1-t) x_0)\leqslant t f(x^*)+(1-t) f(x_0)$

令 t 取一个足够小的正数,可导出矛盾.

(ii)若存在x^*,y^*都是整体极小点($f(x^*)=f(y^*)$),则$f(t x^*+(1-t)y^*)<t f(x^*)+(1-t) f(y^*)=f(x^*)$矛盾.

§2.2 线性规划的标准型与基本概念

线性规划的一般形式

$$\begin{aligned}
&\min(\max)\ c_1x_1+c_2x_2+\cdots+c_nx_n\\
&\text{s.t.}\ \ a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n\geqslant(\text{或}\leqslant,=)b_1\\
&\qquad a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n(\text{或}\leqslant,=)b_2\\
&\qquad \cdots\ \cdots\\
&\qquad a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n(\text{或}\leqslant,=)b_m\\
&\qquad x_1,x_2,\cdots,x_n\geqslant 0
\end{aligned}$$

线性规划的标准型

$$\begin{aligned}
&\min\ c_1x_1+c_2x_2+\cdots+c_nx_n\\
&\text{s.t.}\ \ a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1\\
&\qquad a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2\\
&\qquad \cdots\ \cdots\\
&\qquad a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m\\
&\qquad x_1,x_2,\cdots,x_n\geqslant 0
\end{aligned}$$

矩阵-向量形式的标准型

$$\begin{aligned}
&\min\ c^Tx\\
&(LP)\quad \text{s.t.}\ \ Ax=b\\
&\qquad\qquad\quad x\geqslant 0
\end{aligned}$$

其中$c=(c_1,c_2,\cdots,c_n)^T,x=(x_1,x_2,\cdots,x_n)^T,b=(b_1,b_2,\cdots,b_m)^T$

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{pmatrix}$$

c:价格向量

A:约束矩阵

b:右端向量

矩阵-向量形式的标准型

记$A=(p_1,p_2,\cdots,p_n)$,其中$p_j=(a_{1j},a_{2j},\cdots,a_{mj})^T$,线性规划($LP$)又可以表示为

$$(LP)\quad \begin{aligned} &\min c^T x \\ &s.t. \sum_{j=1}^{n} x_j p_j = b \\ &\quad x_j \ge 0, j=1,2,\cdots,n \end{aligned}$$

线性规划解的情况

满足约束条件的向量x是可行解,全体可行解构成可行域D.

$D=\Phi$时,称线性规划无可行解;

$D\neq\Phi$时但目标函数无下界时,称线性规划(LP)无界或无最优解;

$D\neq\Phi$时若目标函数有下界,可以证明线性规划(LP)必有最优解.

可行域为凸集

定理2.2.1线性规划问题

$$(LP)\quad \begin{aligned} &\min c^T x \\ &s.t.\ Ax=b \\ &\quad x\geqslant 0 \end{aligned}$$

的可行域D为凸集.

证明 任取$x,y\in D$,则有$Ax=b,x\geqslant 0$, $Ay=b,y\geqslant 0$

对任意的$\alpha\in[0,1]$,设$z=\alpha x+(1-\alpha)y$,则$z\geq 0$,且

$Az=A(\alpha x+(1-\alpha)y)$

$=\alpha Ax+(1-\alpha)Ay$

$=\alpha b+(1-\alpha)b$

$=b$

因此$z\in D$

D为凸集.

一般形式转化为标准型

(i)极大$\rightarrow$极小

$\max f(x)\rightarrow\min -f(x)$

(ii)小于等于约束

$$\sum_{j=1}^{n} a_{ij}x_j \le b_i$$

令$x_{n+i}=b_i-\sum_{j=1}^{n} a_{ij}x_j$

则有 $x_{n+i}\ge 0, \sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i$

类似可转化大于等于约束

(iv) 变量x_j无非负约束

引入非负变量

$x_j'\geqslant 0, x_j''\geqslant 0$,

令$x_j=x_j'-x_j''$.

例2.2.1将线性规划

$\min y=2x_1-x_2-3x_3$

s.t. $x_1+x_2+x_3\leqslant 7$

$x_1-x_2+x_3\geqslant 2$

$-3x_1-x_2+2x_3=5$

$x_1,x_2\geq 0,x_3$是自由变量

化为标准型

解:令$x_4=7-(x_1+x_2+x_3)$,

$x_5=(x_1-x_2+x_3)-2$,再令

$x_3=x_3'-x_3''$,得到标准型

$\min y=2x_1-x_2-3x_3'+x_3''$

s.t. $x_1+x_2+x_3'-x_3''+x_4=7$

$x_1-x_2+x_3'-x_3''-x_5=2$

$-3x_1-x_2+2x_3'-2x_3''=5$

$x_1,x_2,x_3',x_3'',x_4,x_5\geqslant 0$

基本概念

设约束矩阵A的秩为,有$m\leqslant n$,则A中必存在m阶非奇异子阵B,不妨设

$B=(p_1,p_2,\cdots,p_m)$

称B为线性规划问题(LP)的一个基矩阵,或称为基,基矩阵中的列向量称为基向量,

对应的变量称为基变量,

其余变量称为非基变量.

基本概念

在约束方程组取定基矩阵
$B=(p_1,p_2,\cdots,p_m)$
之后,令非基变量均为0,得到的方程组
$p_1x_1+p_2x_2+\cdots+p_mx_m=b$
有唯一解,这样得到约束方程组的一个解向量
$x=(x_1,x_2,\cdots x_m)^T$
通过这种方法得到的满足约束方程组的解称为基矩阵B对应的基解.

基本概念

线性规划(LP)的基解个数不会超过 C_n^m

如果基解又满足非负条件,则称之为基可行解.

此时的基B称为可行基.

基可行解中非零分量的个数不会超过m,若基可行解中非零分量的个数恰为m,称此基可行解为非退化的基可行解,否则称为退化的基可行解.

若一个线性规划的所有基可行解都是非退化的,称此线性规划是非退化的.

例 考虑线性规划

$$\min\ 2x_1-x_2$$
$$\text{s.t.}\ \begin{cases} x_1+x_2+x_3=5 \\ -x_1-x_2+x_4=0 \\ 2x_1+2x_2+x_5=22 \\ x_1,x_2,x_3,x_4,x_5\geqslant 0\end{cases}$$

该线性规划有5个变量,3个约束,最多10个基解.

事实上,该线性规划只有7个基解(p_1,p_2线性相关)

下面列出7个基解及对应的基

(p_1,p_3,p_4),$(11,0,-6,11,0)^T$不可行
(p_1,p_3,p_5),$(0,0,5,0,22)^T$退化
(p_1,p_4,p_5),$(5,0,0,5,12)^T$非退化
(p_2,p_3,p_4),$(0,11,-6,11,0)^T$不可行
(p_2,p_3,p_5),$(0,0,5,0,22)^T$退化
(p_2,p_4,p_5),$(0,5,0,5,12)^T$非退化
(p_3,p_4,p_5),$(0,0,5,0,22)^T$退化

§2.3 线性规划的基本定理

本节的基本定理要说明要找线性规划的最优解只需在基可行解中选择就可以了,这样将选择的范围控制在有限个.

定理2.3.1 设x是标准型线性规划(LP)的可行解,x为(LP)的基可行解的充要条件是,x的正分量对应的系数列向量线性无关.

定理2.3.2 设x是标准型线性规划(LP)的可行解,x为(LP)的基可行解的充要条件是,x为可行域D的极点.

证明: 必要性 不妨设$x=(x_1,x_2,\cdots,x_m,0,\cdots,0)^T$是($LP$)的基可行解,且$x_1,x_2,\cdots,x_m$是基变量,假设有$u,v\in D$, $0<\alpha<1$,使得 $x=\alpha u+(1-\alpha)v$

当$m+1\leqslant j\leqslant n$时,$0=x_j=\alpha u_j+(1-\alpha)v_j$,因此$u_j=v_j=0$.
所以$p_1u_1+p_2u_2+\cdots+p_mu_m=p_1v_1+p_2v_2+\cdots+p_mv_m=b$
从而$p_1(u_1-v_1)+p_2(u_2-v_2)+\cdots+p_m(u_m-v_m)=0$
由于x是基可行解,所以$p_1,p_2,\cdots,p_m$线性无关,
$u_j=v_j\ (i=1,2,\cdots,m)$.从而$u=v$.这说明$x$为极点.

充分性 设$x=(x_1,x_2,\cdots,x_k,0,\cdots,0)^T$是可行域的极点,其中$x_1,x_2,\cdots,x_k>0$.
假设x不是基可行解,于是$p_1,p_2,\cdots,p_k$线性相关,即有一组不全为0的数$\alpha_1,\alpha_2,\cdots,\alpha_k$,使得

$$\alpha_1 p_1+\alpha_2 p_2+\cdots+\alpha_k p_k=0 \quad (2.4)$$

又$x\in D$,所以 $x_1p_1+x_2p_2+\cdots+x_kp_k=b$ (2.5)
用$\varepsilon>0$乘(2.4)再与(2.5)相加减得

$$(x_1+\varepsilon\alpha_1)p_1+(x_2+\varepsilon\alpha_2)p_2+\cdots+(x_k+\varepsilon\alpha_k)p_k=b$$
$$(x_1-\varepsilon\alpha_1)p_1+(x_2-\varepsilon\alpha_2)p_2+\cdots+(x_k-\varepsilon\alpha_k)p_k=b$$

令$u=(x_1+\varepsilon\alpha_1,x_2+\varepsilon\alpha_2,\cdots,x_k+\varepsilon\alpha_k,0,\cdots,0)^T$
$v=(x_1-\varepsilon\alpha_1,x_2-\varepsilon\alpha_2,\cdots,x_k-\varepsilon\alpha_k,0,\cdots,0)^T$
则有$Au=b,Av=b$,当ε充分小时,可使$u\geqslant 0,v\geqslant 0$.
因此,当ε充分小时,u,v都是(LP)的可行解,且$u\neq v, x=1/2\ u+1/2\ v$,
这与x是D的极点相矛盾.
因此x是基可行解.

推论:线性规划(LP)的可行域$D=\{x|Ax=b,x\geqslant 0\}$最多具有有限个极点

$(p_1,p_3,p_4),(11,0,-6,11,0)^T$不可行
$(p_1,p_3,p_5),(0,0,5,0,22)^T$退化
$(p_1,p_4,p_5),(5,0,0,5,12)^T$非退化
$(p_2,p_3,p_4),(0,11,-6,11,0)^T$不可行
$(p_2,p_3,p_5),(0,0,5,0,22)^T$退化
$(p_2,p_4,p_5),(0,5,0,5,12)^T$非退化
$(p_3,p_4,p_5),(0,0,5,0,22)^T$退化

前例中三个退化的基可行解对应着同一个极点(基可行解与极点不是一一对应)

有可行解→有基可行解

定理2.3.3 若线性规划(LP)存在可行解,则它一定存在基可行解.
证明 设$x=(x_1,x_2,\cdots,x_n)^T$是(LP)的可行解.不失一般性,设其前k个分量为正,其余分量为零.则有

$$\sum_{j=1}^{k} x_j p_j = b$$

若$p_1,p_2,\cdots,p_k$线性无关,则x为基可行解;
若$p_1,p_2,\cdots,p_k$线性相关,即有一组不全为0的数$\alpha_1,\alpha_2,\cdots,\alpha_k$,使得$\alpha_1p_1+\alpha_2p_2+\cdots+\alpha_kp_k=0$

与定理2.3.2的证明类似,作
$x_1=x+\varepsilon\alpha, x_2=x-\varepsilon\alpha$,其中$\alpha=(\alpha_1,\cdots,\alpha_k,0,\cdots,0)^T$
当ε充分小时,x_1,x_2是线性规划(LP)的可行解.
选择适当的ε,使得$x_j+\varepsilon\alpha_j, x_j-\varepsilon\alpha_j (j=1,\cdots,k)$中至少有一个为零,而其余的值大于零.
这样得到一个新的可行解,其中非零分量的个数比x至少减少一个.
如果新的可行解正分量对应的列向量线性无关,则问题得证.否则重复上面的过程直到正分量对应的列向量线性无关为止.

有最优解→有最优的基可行解

定理2.3.4 若线性规划(LP)存在最优解,则必存在基可行解是最优解.
证明: 设x是最优解.若x不是基可行解,作出两个新的可行解:$x+\varepsilon\alpha$, $x-\varepsilon\alpha$,对应的目标函数值为$c^Tx+\varepsilon c^T\alpha$与$c^Tx-\varepsilon c^T\alpha$.
由于x是最优解, $c^Tx+\varepsilon c^T\alpha \geqslant c^Tx$; $c^Tx-\varepsilon c^T\alpha \geqslant c^Tx$.
因此$c^T\alpha=0$.
于是,当$\varepsilon>0$充分小时, $x+\varepsilon\alpha$, $x-\varepsilon\alpha$也是可行最优解.
仿照定理2.3.3的证明,可以得到最优的基可行解.

单纯形方法的思路

找出一基可行解(极点)

若其不是最优,找到一个相邻极点

新的目标函数值不大于原目标函数值

经过有限次迭代给出最优解或判断无最优解

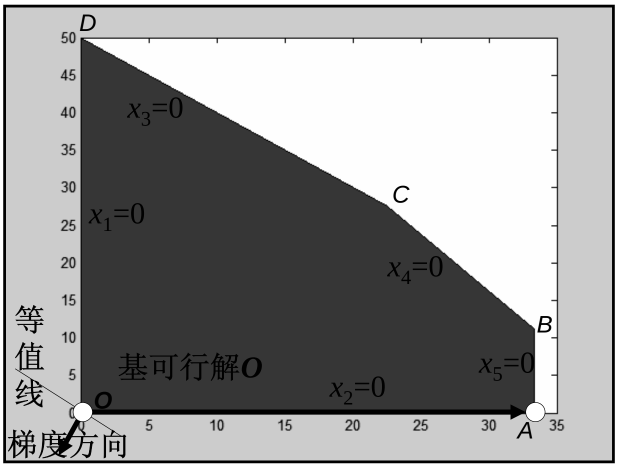
单纯形方法的思路(几何)

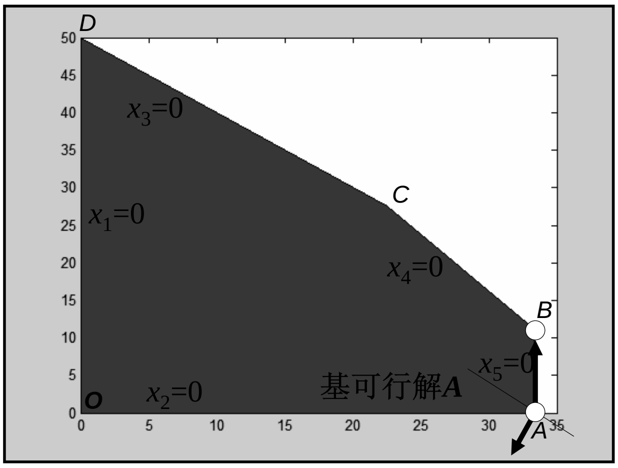
线性规划

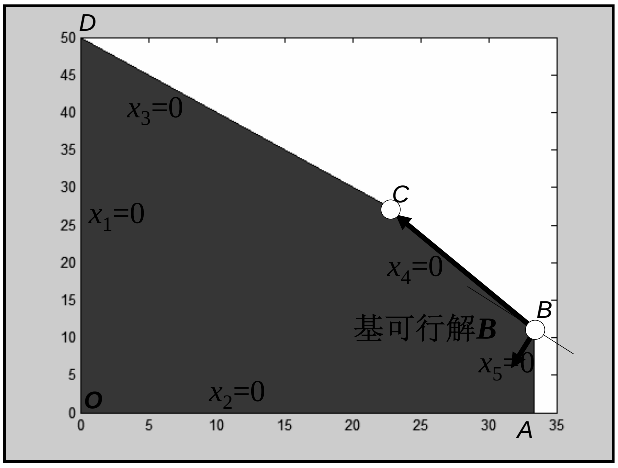
$$\begin{aligned}\min\ & -72x_1-64x_2\\ \text{s.t.}\ & x_1+x_2+x_3=50\\ & 12x_1+8x_2+x_4=490\\ & 3x_1+x_5=100\\ & x_1,x_2,x_3,x_4,x_5\geqslant 0\end{aligned}$$

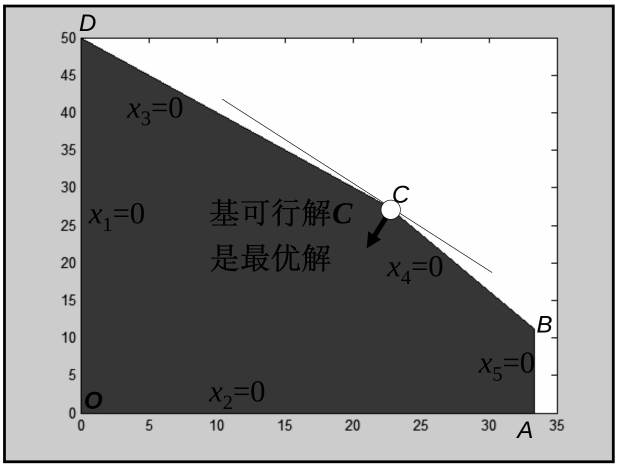
的等价形式为

$$\begin{aligned}\min\ & -72x_1-64x_2\\ \text{s.t.}\ & x_1+x_2\leqslant 50\\ & 12x_1+8x_2\leqslant 490\\ & 3x_1\leqslant 100\\ & x_1,x_2\geqslant 0\end{aligned}$$









单纯形方法的思路(代数)

例 考察线性规划

min $-72x_1-64x_2$

s.t. $x_1+x_2+x_3=50$

$12x_1+8x_2+x_4=490$

$3x_1+x_5=100$

$x_1,x_2,x_3,x_4,x_5\geqslant 0$

以x_3,x_4,x_5为基变量,容易得到基可行解$(0,0,50,490,100)^T$.

由于x_1的价格系数为负数,增加x_1的取值可以使得目标函数值减少.

类似的,我们也可以增加x_2的取值,使得目标函数值减少.

由于-72负得多一些,我们先增加x_1.

单纯形方法的思路(代数)

min $-72x_1-64x_2$

s.t. $x_1+x_2+x_3=50 \Rightarrow x_1\leqslant 50$

$12x_1+8x_2+x_4=490 \Rightarrow x_1\leqslant 490/12$

$3x_1+x_5=100 \Rightarrow x_1\leqslant 100/3$

$x_1,x_2,x_3,x_4,x_5\geqslant 0$

x_1可以增加多少?

因此x_1的最大取值为min(50,490/12,100/3)=100/3

此时x_5的取值为0, x_5“出基”.

单纯形方法的思路(代数)

根据$3x_1+x_5=100$,我们将原来的线性规划改写如下

min $-64x_2+24x_5-2400$

s.t. $x_2+x_3-x_5/3=50/3 \Rightarrow x_2\leqslant 50/3$

$8x_2+x_4-4x_5=90 \Rightarrow x_2\leqslant 90/8$

$x_1+x_5/3=100/3$

$x_1,x_2,x_3,x_4,x_5\geqslant 0$

此时,基变量为x_1,x_3,x_4,基可行解为$(100/3,0,50/3,90,0)^T$.

若x_2(其系数为负)的取值增加,可以使得目标函数值减少

因此x_2的最大取值为min(50/3,90/8)=90/8

x_4“出基”.

单纯形方法的思路(代数)

此时,x_4,x_5是非基变量,将原规划化为

min $8x_4-8x_5-3120$

s.t. $x_3-x_4/8+x_5/6=65/12$

$x_2+x_4/8-x_5/2=45/4$

$x_1+x_5/3=100/3$

$x_1,x_2,x_3,x_4,x_5\geqslant 0$

解为$(100/3,45/4,65/12,0,0)^T$.

x_5最大可以取为65/2.

对应的,线性规划可以转化为下页的形式

单纯形方法的思路(代数)

min $48x_4+2x_5-3380$

s.t. $6x_3-3x_4/4+x_5=65/2$

$x_2+3x_3-x_4/4=55/2$

$x_1-2x_3+x_4/4=45/2$

$x_1,x_2,x_3,x_4,x_5\geqslant 0$

对应的解为$(45/2,55/2,0,0,65/2)^T$.

此时,目标函数中非基变量的系数为正,因此目标函数的取值不能再减少.最优值为-3380.

单纯形方法的思路(代数)

单纯形方法求解线性规划,首先找出一个基可行解.将目标函数写成非基变量的线性组合(再加上一个常数)的形式.

如果组合的系数全部非负,则已经找到最优解.

如果组合的系数中有负数,从中选取一个变量(“进基”)来增加取值,可以使得函数值减少.

根据约束条件,可以控制增加的范围.

在进基变量取最大值时,有一个变量出基,从而得到另一个基可行解.

重复上面的过程,可以求得最优解.

§ 2.4 单纯形方法

$$\begin{aligned} \min \quad & z = c^T x \\ s.t. \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & z = (C_B^T, C_N^T)\begin{pmatrix} x_B \\ x_N \end{pmatrix} \\ s.t. \quad & (B, N)\begin{pmatrix} x_B \\ x_N \end{pmatrix} = b \\ & x_B, x_N \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & z = c_B^T c_B + c_N^T x_N \\ s.t. \quad & Bx_B + Nx_N = b \\ & x_B, x_N \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & z = c_B^T x_B + c_N^T x_N \\ s.t. \quad & x_B + B^{-1}Nx_N = B^{-1}b \\ & x_B, x_N \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & z = c_B^T B^{-1} b + (c_N^T - c_B^T B^{-1} N) x_N \\ s.t. \quad & x_B + B^{-1}Nx_N = B^{-1}b \\ & x_B, x_N \geq 0 \end{aligned}$$

规范式

线性规划与基变量$x_1,\cdots,x_m$对应的规范式

$$\begin{aligned} \min \quad & f(x) = f_0 + \sum_{j=m+1}^{n} \sigma_j x_j \\ s.t. \quad & x_i = b_i' - \sum_{j=m+1}^{n} a_{ij}' x_j, \ i = 1, \cdots, m \\ & x_j \geq 0, \ j = 1, \cdots, n \end{aligned}$$

线性规划与基变量$x_{j1},\cdots,x_{jm}$对应的规范式

$$\begin{aligned} \min \quad & f(x) = f_0 + \sum_{j \in T} \sigma_j x_j \\ s.t. \quad & x_i = b_i' - \sum_{j \in T} a_{ij}' x_j, \ i \in S \\ & x_j \geq 0, \ j = 1, \cdots, n \end{aligned}$$

$S=\{j_1,\cdots,j_m\}$, $T=\{1,2,\cdots,n\}\backslash S$

2.4.1 基可行解是最优解的判断准则

在规范式

$$\begin{aligned} \min \quad & f(x) = f_0 + \sum_{j \in T} \sigma_j x_j \\ s.t. \quad & x_i = b_i' - \sum_{j \in T} a_{ij}' x_j, \ i \in S \\ & x_j \geq 0, \ j = 1, \cdots, n \end{aligned}$$

中,令非基变量$x_j=0, j\in T$,得到一个基解$x_0=(x_1,\cdots,x_n)^T$,
其中

$$x_j = \begin{cases} b_j' & j \in S \\ 0 & j \in T \end{cases}$$

如果$b'_j \geqslant 0 (j\in S)$,则$x_0$是基可行解.

判别数

定义2.4.1 令 $z_j = c_B^T p_j' = \sum_{i \in S} c_i a_{ij}'$,则称

$$\sigma_j = c_j - z_j = c_j - c_B^T p_j' = c_j - c_B^T B^{-1} p_j = c_j - \sum_{i \in S} c_i a_{ij}'$$

为变量x_j的判别数.

其中p_j为A的列向量, p_j'为$B^{-1}A$的列向量

写成向量形式为

$$\begin{aligned} \sigma_B &= c_B^T - c_B^T B^{-1} B = 0 \\ \sigma_N &= c_N^T - c_B^T B^{-1} N \\ \sigma &= c^T - c_B^T B^{-1} A \end{aligned}$$

最优性条件

在规范式
$$\begin{aligned} &\min\ f(x) = f_0 + \sum_{j\in T} \sigma_j x_j \\ &s.t.\quad x_i = b_i' - \sum_{j\in T} a_{ij}' x_j,\ i \in S \\ &\qquad x_j \ge 0,\ j = 1,\cdots,n \end{aligned}$$
中,

目标函数中各个变量的系数就是判别数.

定理2.4.1 设x_0是线性规划(LP)对应于基$B=(p_{j1},\cdots,p_{jm})$的基可行解.与基变量$x_{j1},\cdots,x_{jm}$对应的规范式中,若x_0的全体判别数非负,则x_0是(LP)的最优解.

判断无最优解

定理2.4.2设x_0是线性规划(LP)对应于基$B=(p_1,\cdots,p_m)$的基可行解.与基变量$x_1,\cdots,x_m$对应的规范式中,若存在$\sigma_k<0$,且对所有的$i=1,2,\cdots,m$有$a_{ik}'\leqslant 0$,则线性规划(LP)无最优解.

判断无最优解

证明:令$d=(-a_{1k}',\cdots,-a_{mk}',0,\cdots,0,1,0,\cdots,0)^T$(第$k$个分量为1),$y=x_0+\lambda d$,则

$$y_j = \begin{cases} b_j' - \lambda a_{jk}', & j = 1,2,\cdots,m \\ \lambda, & j = k \\ 0, & j = m+1, m+2,\cdots,n,\ j \ne k \end{cases}$$

由于$a_{ik}'\leqslant 0, b_i'\geqslant 0, i=1,\cdots,m$,所以对任意$\lambda>0$,$y$是可行解.$y$对应的目标函数值为$f=f_0+\sigma_k\lambda\to-\infty$ ($\lambda\to+\infty$ 时).目标函数在可行域无解,(LP)无最优解.

基可行解的转换

从上面定理的证明中可以看出,如果某个判别数非负时,可以构造新的可行解,使得目标函数值减少.

1. 确定进基变量

负的判别数对应的变量都可以作为进基变量.

一般的,若$\sigma_k=\min\{\sigma_j|\sigma_j<0\}$

则以x_k为进基变量.

基可行解的转换

2. 确定离基变量

设已确定x_k为进基变量,根据
$$x_i = b_i' - \sum_{j=m+1}^{n} a_{ij}' x_j,\ i = 1,2,\cdots,m$$
以及$x_j=0,(j=m+1,\cdots,n,j\neq k)$,得到

$x_i=b'_i-a'_{ik}x_k, i=1,2,\cdots,m$

我们选取$x_k=\max\{x_k|\ x_i=b'_i-a'_{ik}x_k\geqslant 0, i=1,2,\cdots,m\}$

显然$x_k=\min\{\theta_i=b'_i/a'_{ik}|a'_{ik}>0\}=b'_l/a'_{lk}=\theta_l$,

此时,我们选择x_l为离基变量.

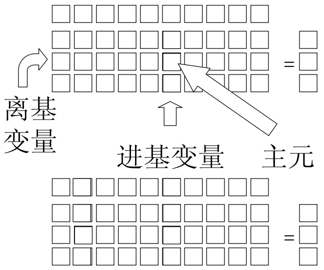
单纯形方法

如果线性规划是非退化的,则按照上面的方法(进基,离基)迭代一次后,目标函数值有所下降.经过有限次迭代之后,一定可以得到一个基可行解,使得其所有判别数非负(得到最优解),或者其有一个判别数是负的,但对应列向量的所有分量非正(线性规划无最优解).

这种求解线性规划的方法称为单纯形方法.

单纯形法的迭代步骤

单纯形法的迭代的主要工作是将原来的规范式写为新的规范式.



通过初等行变换将主元变为1

通过初等行变换将主元所在列其它元素变为0

得到新的规范式

§2.5 单纯形表

例2.5.1 用单纯形法解线性规划

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

s.t. $x_1+x_2\leqslant 6$

$x_1+2x_2\leqslant 8$

$x_1\leqslant 4$

$x_2\leqslant 3$

$x_1,x_2\geqslant 0$

解 引入松弛变量将原问题化为标准型

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

s.t. $x_1+x_2+x_3=6$

$x_1+2x_2+x_4=8$

$x_1+x_5=4$

$x_2+x_6=3$

$x_1,x_2,x_3,x_4,x_5,x_6\geqslant 0$

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

$s.t.\ x_1+x_2+x_3=6$

$x_1+2x_2+x_4=8$

$x_1+x_5=4$

$x_2+x_6=3$

$x_1,x_2,x_3,x_4,x_5,x_6\geqslant 0$

目标函数中的系数

基向量

基变量的价格系数

右端系数

约束等式的系数

c_j			-2	-3	0	0	0	0	
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	
0	p_3	6	1	1	1	1	0	0	
0	p_4	8	1	2	0	1	1	0	
0	p_5	4	1	0	0	0	1	0	
0	p_6	3	0	1	0	0	0	1	

c_j			-2	-3	0	0	0	0	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	
0	p_3	6	1	1	1	0	0	0	**6**
0	p_4	8	1	2	0	1	0	0	**4**
0	p_5	4	1	0	0	0	1	0	**/**
0	p_6	3	0	(1)	0	0	0	1	**3**
σ_j			**-2**	**-3**	**0**	**0**	**0**	**0**	

第一步的判别数:由于在此例中基变量的价格系数均为0,所以判别数就是价格系数.

进基变量:σ_2**=-3=min{**$\sigma_j|\sigma_j$**<0},所以**x_2**为进基变量**

离基变量的判别标准:$\theta_i=b_i/a_{i2}$ $(a_{i2}>0)$

离基变量: θ_6**=min**θ_i**,所以**x_6**为离基变量 标出主元**

c_j							0	0	θ_i
c_B	B	b					p_5	p_6	
0	p_3	6							
						0	0	1	3
						0	0	0	
						0	**0**	**-1**	
		2	**1**	**0**	**0**	**1**	**0**	**-2**	
							1	**0**	
-3	p_2	**3**	**0**	**1**	**0**	**0**	**0**	**1**	
σ_j			**-2**	**0**	**0**	**0**	**0**	**3**	

写出基向量(p_6换成p_2)

归一化:若主元不等于1,则进行行变换,将主元变为1(此处不变)

写出价格系数

消去:用初等行变换将主元所在列其它元素消为0

p_2所在行乘以-1加到p_3所在行

p_2所在行乘以-2加到p_4所在行

p_5所在行不变

p_2所在行乘以3加到判别数所在行

c_j			-2	-3	0	0	0	0	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	
0	p_3	6	1	1	1	0	0	0	6
0	p_4	8	1	2	0	1	0	0	4
0	p_5	4	1	0	0	0	1	0	/
0	p_6	3	0	(1)	0	0	0	1	3
σ_j			-2	-3	0	0	0	0	
0	$\boldsymbol{p_3}$	**3**	**1**	**0**	**1**	**0**	**0**	**-1**	
0	$\boldsymbol{p_4}$	**2**	**1**	**0**	**0**	**1**	**0**	**-2**	
0	$\boldsymbol{p_5}$	**4**	**1**	**0**	**0**	**0**	**1**	**0**	
-3	$\boldsymbol{p_2}$	**3**	**0**	**1**	**0**	**0**	**0**	**1**	
σ_j			**-2**	**0**	**0**	**0**	**0**	**3**	

169

c_j			-2	-3	0	0	0	0	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	
0	p_3	3	1	0	1	0	0	-1	**3**
0	p_4	2	(1)	0	0	1	0	-2	**2**
0	p_5	4	1	0	0	0	1	0	**4**
-3	p_2	3	0	1	0	0	0	1	**/**
σ_j			**-2**	**0**	**0**	**0**	**0**	**3**	

进基变量:σ_1=-2=min{σ_j|σ_j<0},所以x_1为进基变量

离基变量的判别标准:$\theta_i = b_i / a_{i1}$ (a_{i1}>0)

离基变量: θ_4=2=minθ_i, x_4为离基变量

标出主元

归一化:若主元不等于1,则进行行变换,将主元变为1(此处不变)

写出基向量($\boldsymbol{p_4}$换成$\boldsymbol{p_1}$)

写出价格系数

消去:用初等行变换将主元所在列其它元素消为0

$\boldsymbol{p_1}$所在行乘以-1加到$\boldsymbol{p_3}$所在行

$\boldsymbol{p_1}$所在行乘以-1加到$\boldsymbol{p_5}$所在行

$\boldsymbol{p_2}$所在行不变

$\boldsymbol{p_1}$所在行乘以2加到判别数所在行

-3	$\boldsymbol{p_2}$	**3**	**0**	**1**	**0**	**0**	**0**	**1**	
σ_j			**0**	**0**	**0**	**2**	**0**	**-1**	

171

c_j			-2	-3	0	0	0	0	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	
0	p_3	3	1	0	1	0	0	-1	3
0	p_4	2	(1)	0	0	1	0	-2	2
0	p_5	4	1	0	0	0	1	0	4
-3	p_2	3	0	1	0	0	0	1	/
σ_j			-2	0	0	0	0	3	
0	$\boldsymbol{p_3}$	**1**	**0**	**0**	**1**	**-1**	**0**	**1**	
-2	$\boldsymbol{p_1}$	**2**	**1**	**0**	**0**	**1**	**0**	**-2**	
0	$\boldsymbol{p_5}$	**2**	**0**	**0**	**0**	**-1**	**1**	**2**	
-3	$\boldsymbol{p_2}$	**3**	**0**	**1**	**0**	**0**	**0**	**1**	
σ_j			**0**	**0**	**0**	**2**	**0**	**-1**	

172

c_j			-2	-3	0	0	0	0	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	
0	p_3	1	0	0	1	-1	0	1	**1**
-2	p_1	2	1	0	0	1	0	-2	**/**
0	p_5	2	0	0	0	-1	1	(2)	**1**
-3	p_2	3	0	1	0	0	0	1	**3**
σ_j			**0**	**0**	**0**	**2**	**0**	**-1**	

进基变量:σ_6=-1=min{σ_j|σ_j<0},所以x_6为进基变量

离基变量的判别标准:$\theta_i = b_i / a_{i6}$ (a_{i6}>0)

离基变量: θ_5=1=minθ_i, x_5为离基变量(此处可选x_3)

标出主元

归一化:若主元不等于1,则进行行变换,将主元变为1(此处除以2)

写出基向量($\boldsymbol{p_5}$换成$\boldsymbol{p_6}$)

写出价格系数

消去:用初等行变换将主元所在列其它元素消为0

$\boldsymbol{p_6}$所在行乘以-1加到$\boldsymbol{p_3}$所在行

$\boldsymbol{p_6}$所在行乘以2加到$\boldsymbol{p_1}$所在行

$\boldsymbol{p_6}$所在行乘以-1加到$\boldsymbol{p_2}$所在行

$\boldsymbol{p_6}$所在行乘以1加到判别数所在行

-3	$\boldsymbol{p_2}$	**2**	**0**	**1**	**0**	**½**	**-½**	**0**	
σ_j			**0**	**0**	**0**	**3/2**	**½**	**0**	

174

		c_j	-2	-3	0	0	0	0	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	
0	p_3	1	0	0	1	-1	0	1	1
-2	p_1	2	1	0	0	1	0	-2	/
0	p_5	2	0	0	0	-1	1	(2)	1
-3	p_2	3	0	1	0	0	0	1	3
	σ_j		0	0	0	2	0	-1	
0	$\boldsymbol{p_3}$	**0**	**0**	**0**	**1**	**-½**	**-½**	**0**	
-2	$\boldsymbol{p_1}$	**4**	**1**	**0**	**0**	**0**	**1**	**0**	
0	$\boldsymbol{p_6}$	**1**	**0**	**0**	**0**	**-½**	**½**	**1**	
-3	$\boldsymbol{p_2}$	**2**	**0**	**1**	**0**	**½**	**-½**	**0**	
	σ_j		**0**	**0**	**0**	**3/2**	**½**	**0**	

		c_j	-2	-3	0	0	0	0	
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	
0	p_3	0	0	0	1	-1/2	-1/2	0	
-2	p_1	4	1	0	0	0	1	0	
0	p_6	1	0	0	0	-1/2	1/2	1	
-3	p_2	2	0	1	0	1/2	-1/2	0	
	σ_j		**0**	**0**	**0**	**3/2**	**1/2**	**0**	

此时所有的判别数都非负,迭代终止.

最优解为x^*=(4,2,0,0,0,1)^T,

原问题的最优解为x^*=(4,2)^T,最优值为

f^*=(-2)·4+(-3)·2=-14.

§2.6 初始基可行解的求法

对于线性规划问题 引入松弛变量化为标准型

$$\min c^Tx \qquad\qquad \min c^Tx$$

$$\text{s.t. } Ax \leqslant b\ (b \geqslant 0) \qquad \text{s.t. } Ax+Ix_S=b$$

$$x \geqslant 0 \qquad\qquad x, x_S \geqslant 0$$

其中I是单位矩阵,$x_S=(x_{n+1},\cdots,x_{n+m})^T$.则可以将$x_S$作为基变量,以$(0,\cdots,0,b_1,\cdots,b_m)^T$为初始基可行解进行单纯形迭代.

对于一般的线性规划问题,无法简单给出初始基可行解.

2.6.1 大M单纯形法

对于线性规划问题 $\min f(x)=\sum_{j=1}^{n} c_j x_j$

$$s.t. \sum_{j=1}^{n} a_{ij}x_j = b_i (b_i \geq 0), i=1,2,\cdots,m$$

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

引入人工变量$x_{n+1},\cdots,x_{n+m}$,构造辅助线性规划问题

$$\min \bar{f}(x)=\sum_{j=1}^{n} c_j x_j + M\sum_{j=n+1}^{n+m} x_j$$

$$s.t. \sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i, i=1,2,\cdots,m$$

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

$$\min \bar{f}(x)=\sum_{j=1}^{n} c_j x_j + M\sum_{j=n+1}^{n+m} x_j$$

$$s.t. \sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i, i=1,2,\cdots,m$$

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

M是相当大的正数(可以理解为正无穷),对人工变量起到惩罚的作用,逼迫辅助线性规划的最优解中人工变量均为0.

考虑辅助规划中的目标函数

$$\sum_{j=1}^{n} c_j x_j + M \sum_{j=n+1}^{n+m} x_j$$

注:教材中"不全为非基变量"应改为"不全为0"

事实上,若辅助规划的最优解$(x_1,x_2,\cdots,x_{n+m})$中人工变量不全为0(设$x_r\neq 0$),则原规划无可行解.

如若不然,设原规划有可行解 $\bar{x}=(\bar{x}_1,\bar{x}_2,\cdots,\bar{x}_n)^T$
则 $x'=(\bar{x}_1,\cdots,\bar{x}_n,0,\cdots,0)^T$ 是辅助规划可行解,

于是 $Mx_r \le \sum_{j=n+1}^{n+m} Mx_j \le \bar{f}(x') - \sum_{j=1}^{n} c_j x_j = \sum_{j=1}^{n} c_j(\bar{x}_j - x_j)$

这与M充分大矛盾.

大M方法算例

例2.6.1 用大M单纯形法求解线性规划

$\min f(x)=-3x_1+x_2+x_3$
s.t. $x_1-2x_2+x_3 \leqslant 11$
$-4x_1+x_2+2x_3-x_4=3$
$-2x_1 \quad +x_3 \quad =1$
$x_1,x_2,x_3,x_4\geqslant 0$

引入松弛变量x_5,人工变量x_6,x_7,构造辅助线性规划

$\min -3x_1+x_2+x_3+Mx_6+Mx_7$
s.t. $x_1-2x_2+x_3 \quad +x_5 \quad =11$
$-4x_1+x_2+2x_3-x_4 \quad +x_6 \quad =3$
$-2x_1 \quad +x_3 \quad +x_7=1$
$x_1,\cdots,x_7\geqslant 0$

注:根据线性规划问题本身的形式,可以少引进一些人工变量.

	c_j		-3	1	1	0	0	M	M	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	p_7	
0	p_5	11	1	-2	1	0	1	0	0	11
M	p_6	3	-4	1	2	-1	0	1	0	3/2
M	p_7	1	-2	0	(1)	0	0	0	1	1
	σ_j		$6M$-3	$-M$+1	$-3M$+1	M	0	0	0	

第一步的判别数:σ_1=-3-1·0-(-4)M-(-2)M=6M-3
类似地可以给出其它各个判别数.
x_3进基, x_7离基,标出主元.
注:人工变量一旦离基,则在迭代时不再参与计算.

	c_j		-3	1	1	0	0	M	M	
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	p_7	
0	p_5	11	1	-2	1	0	1	0	0	
M	p_6	3	-4	1	2	-1	0	1	0	
M	p_7	1	-2	0	(1)	0	0	0	1	
	σ_j		$6M$-3	$-M$+1	$-3M$+1	M	0	0	0	
0	p_5	10	3	-2	0	0	1	0		
M	p_6	1	0	1	0	-1	0	1		
1	p_3	1	-2	0	1	0	0	0		
	σ_j		-1	$-M$+1	0	M	0	0		

0	p_5	10	3	-2	0	0	1	0	/
M	p_6	1	0	(1)	0	-1	0	1	1
1	p_3	1	-2	0	1	0	0	0	/
	σ_j		-1	$-M$+1	0	M	0	0	
0	p_5	12	(3)	0	0	-2	1		4
1	p_2	1	0	1	0	-1	0		/
1	p_3	1	-2	0	1	0	0		/
	σ_j		-1	0	0	1	0		
-3	p_1	4	1	0	0	-2/3	1/3		
1	p_2	1	0	1	0	-1	0		
1	p_3	9	0	0	1	-4/3	2/3		
	σ_j		0	0	0	1/3	1/3		

求得辅助规划问题的最优解为 $(4,1,9,0,0,0,0)^T$,原线性规划的最优解为$(4,1,9,0)^T$,最优值为-3·4+1·3+1·9=-2.

注:根据各个判别数,可以发现M应满足
$-3M+1<-M+1, -3M+1<6M-3, -M+1<-1$
即$M>2$.

在实际问题中可以取M为适当大的一个数,比如比问题中的系数大一个数量级.

2.6.2 两阶段单纯形法

对于线性规划问题 $\min\ f(x)=\sum_{j=1}^{n}c_jx_j$

$$s.t.\ \sum_{j=1}^{n}a_{ij}x_j=b_i(b_i\ge 0),i=1,2,\cdots,m$$
$$x_j\ge 0,j=1,2,\cdots,n$$

引入人工变量$x_{n+1},\cdots,x_{n+m}$,构造辅助线性规划问题

$$\min\ w=\sum_{j=n+1}^{n+m}x_j$$
$$s.t.\ \sum_{j=1}^{n}a_{ij}x_j+x_{n+i}=b_i,i=1,2,\cdots,m$$
$$x_j\ge 0,j=1,2,\cdots,n,n+1,\cdots,n+m$$

该规划有可行解$(0,\cdots,0,b_1,\cdots,b_n)^T$,说明可行域非空.

目标函数显然有下界,因此该规划有最优解.

$$\min\ w=\sum_{j=n+1}^{n+m}x_j$$
$$s.t.\ \sum_{j=1}^{n}a_{ij}x_j+x_{n+i}=b_i,i=1,2,\cdots,m$$
$$x_j\ge 0,j=1,2,\cdots,n,n+1,\cdots,n+m$$

若最优值$w^*\neq 0$,显然原规划问题无可行解;

若最优值$w^*=0$,从最优解中删去人工变量就得到原规划的可行解.此时如果辅助线性规划是非退化的,那么此可行解必是原规划的基可行解.

两阶段法算例

例2.6.2 用两阶段法求解线性规划

$\min\ 4x_1+x_2+x_3$

s.t. $2x_1+x_2+2x_3=4$

$3x_1+3x_2+x_3=3$

$x_1,x_2,x_3\geqslant 0$

解,引入人工变量,构造辅助线性规划:

$\min\ x_4+x_5$

s.t. $2x_1+x_2+2x_3+x_4\quad=4$

$3x_1+3x_2+x_3\quad+x_5=3$

$x_1,x_2,x_3,x_4,x_5\geqslant 0$

c_j			0	0	0	1	1	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	p_5	
1	p_4	4	2	1	2	1	0	2
1	p_5	3	(3)	3	1	0	1	1
σ_j			-5	-4	-3	0	0	
1	p_4	2	0	-1	(4/3)	1		3/2
0	p_1	1	1	1	1/3	0		3
σ_j			0	1	-4/3	0		
0	p_3	3/2	0	-3/4	1			
0	p_1	1/2	1	5/4	0			
σ_j			0	0	0			

c_j			0	0	0	1	1	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	p_5	
σ_j			-5	-4	-3	0	0	
σ_j			0	1	-4/3	0		
0	p_3	3/2	0	-3/4	1			
0	p_1	1/2	1	5/4	0			
σ_j			0	0	0			

最优解$(1/2,0,3/2\ 0,0)^T$,最优值$w^*=0$,由此得到原线性规划的一个基可行解$x_0=(1/2,0,3/2)^T$.

将上表的最后部分转到新的单纯形表中(求解原线性规划),不过判别数要重新计算.

求解原线性规划的单纯形表

c_j			4	1	1	θ_i
c_B	B	b	p_1	p_2	p_3	
1	p_3	3/2	0	-3/4	1	/
4	p_1	1/2	1	(5/4)	0	2/5
σ_j			0	-13/4	0	
1	p_3	9/5	3/5	0	1	
1	p_2	2/5	4/5	1	0	
σ_j			13/5	0	0	

得到原线性规划的最优解$x^*=(0,2/5,9/5)^T$,最优值$f^*=11/5$.

§ 2.8 线性规划的对偶理论

假设某工厂有m种设备:$B_1,B_2,\cdots,B_m$.一年内各设备的生产能力(有效台时数)为$b_1,b_2,\cdots,b_m$.利用这些设备可以加工n种产品:$A_1,A_2,\cdots,A_n$,单位产品的利润分别为$c_1,c_2,\cdots,c_n$.第j种产品需要在第i种设备上加工的台时数为a_{ij}.问在设备能力允许 全年总收入最

$$\max \quad s=\sum_{j=1}^{n} c_j x_j$$

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

$$x_j \ge 0, j=1,2,\cdots,n$$

设x_1,x_2 为全年总收入

假设某工厂有m种设备:$B_1,B_2,\cdots,B_m$.一年内各设备的生产能力(有效台时数)为$b_1,b_2,\cdots,b_m$.利用这些设备可以加工n种产品:$A_1,A_2,\cdots,A_n$,单位产品的利润分别为$c_1,c_2,\cdots,c_n$.第j种产品需要在第i种设备上加工的台时数为a_{ij}.问在设备能力允许的条件下怎样安排生产计划,使全年总收入最多?

设$x_1,x_2,\cdots,x_n$为各产品的计划年产量,s为全年总收入,易建立该问题的数学模型.

对偶问题

假设工厂将所有的设备用于出租,需要给各种设备制定出租价格.定价原则有两条:一是出租后得到的单位利润不得少于直接生产时的收入,二是出租价格尽量的低,以利于市场竞争.

$$\min \quad w=\sum_{i=1}^{m} b_i y_i$$

$$s.t.\quad \sum_{i=1}^{m} a_{ij} y_i \ge c_j, j=1,2,\cdots,n$$

$$y_i \ge 0, i=1,2,\cdots,m$$

设第i种设 为y_i,全年总收入

对偶问题

假设工厂将所有的设备用于出租,需要给各种设备制定出租价格.定价原则有两条:一是出租后得到的单位利润不得少于直接生产时的收入,二是出租价格尽量的低,以利于市场竞争.

设第i种设备B_i的单位台时的出租价格为y_i,全年总收入为w,则该问题的数学模型为

原始线性规划

$$\max \quad s=\sum_{j=1}^{n} c_j x_j$$

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

$$x_j \ge 0, j=1,2,\cdots,n$$

对偶线性规划

$$\min \quad w=\sum_{i=1}^{m} b_i y_i$$

$$s.t.\quad \sum_{i=1}^{m} a_{ij} y_i \ge c_j, j=1,2,\cdots,n$$

$$y_i \ge 0, i=1,2,\cdots,m$$

可以看出,原始规划与对偶规划是同一组数据参数,只是位置有所不同,所描述的问题实际上是同一个问题从另一种角度去描述.

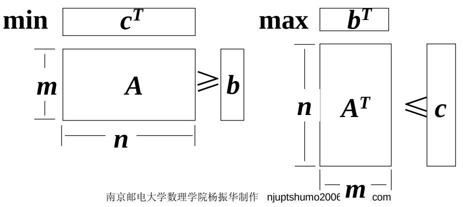
对偶规划(定义2.8.1)

原始线性规划

$$\text{(LP)}\quad \begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax \geqslant b \\ & x \geqslant 0 \end{aligned}$$

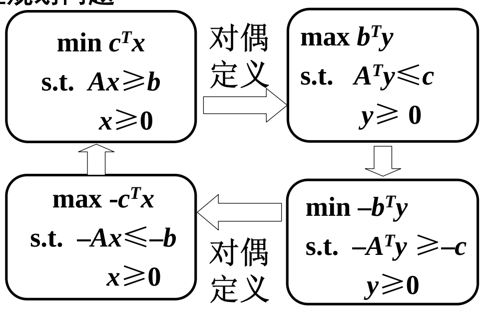
对偶线性规划

$$\text{(DP)}\quad \begin{aligned} \max\ & b^T y \\ \text{s.t.}\ & A^T y \leqslant c \\ & y \geqslant 0 \end{aligned}$$



对合性

定理2.8.1 对偶线性规划的对偶问题是原始线性规划问题



对偶规划的写法

min　max, A　A^T, c　b, 约束=　变量无正负

min + 约束≥ → 变量非负　max + 约束≥ → 变量非正

min + 约束≤ → 变量非正　max + 约束≤ → 变量非负

min + 变量非负 →约束≤　max +变量非负 → 约束≥

min + 变量非正 →约束≥　max +变量非正 → 约束≤

记忆方法: “max” = “非负” = “≥” = +　类似定义“–”

另外,左边的“约束”= “–”,然后将“+” “–”相结合

例如 “max + 约束≥” 为两正一负, 推出右边为负,即变量为非正

$$\begin{aligned} \max\ & 16x_1+6x_2-3x_3 \\ \text{s.t.}\ & 2x_1+x_2-4x_3 \geqslant -5 \\ & -3x_1-x_2 = -2 \\ & 2x_1+2x_2+x_3 \leqslant 6 \\ & x_1 \geqslant 0, x_2 \leqslant 0, \end{aligned}$$

x_3为自由变量

先写出基本形式

max+约束+≥→ 非正

max+约束+≤→ 非负

$$\begin{aligned} \min\ & -5y_1-2y_2+6y_3 \\ \text{s.t.}\ & 2y_1-3y_2+2y_3 \geqslant 16 \\ & y_1-y_2+2y_3 \leqslant 6 \\ & -4y_1+y_3 = -3 \end{aligned}$$

$y_1 \leqslant 0$, y_2为自由变量, $y_3 \geqslant 0$

max+非负→ ≥

max+非正→ ≤

完成!

非对称形式的对偶线性规划

标准型线性规划

$$\begin{aligned} \min\ & c^T x \\ \text{s.t.}\ & Ax = b \\ & x \geqslant 0 \end{aligned}$$

对偶线性规划

$$\begin{aligned} \max\ & b^T y \\ \text{s.t.}\ & A^T y \leqslant c \end{aligned}$$

对偶规划的性质

考虑如下的线性规划

$$\text{(LP)}\quad \begin{aligned} \min\ & 5x_1+3x_2 \\ \text{s.t.}\ & x_1+2x_2-x_3 = 2 \\ & 4x_1+x_2-x_4 = 3 \\ & x_1,x_2,x_3,x_4 \geqslant 0 \end{aligned}$$

其对偶规划为

$$\text{(DP)}\quad \begin{aligned} \max\ & 2y_1+3y_2 \\ \text{s.t.}\ & y_1+4y_2 \leqslant 5 \\ & 2y_1+y_2 \leqslant 3 \\ & -y_1 \leqslant 0 \\ & -y_2 \leqslant 0 \end{aligned}$$

对偶规划的性质

(LP)的三个基可行解(只写出x_1,x_2)及对应的目标函数值为

(2,0)	10
(0,3)	9
(4/7,5/7)	5

(DP)的四个基可行解(只写出y_1,y_2)及对应的目标函数值为

(0,0)	0
(3/2,0)	3
(0,5/4)	15/4
(1,1)	5

可见:(LP)在可行解上的取值不小于(DP)在可行解上的取值;

若两者取值相等,则对应的解都是最优解.

弱对偶性定理

标准型线性规划

$$\text{(LP)}\quad \begin{array}{l}\min\ c^Tx\\ \text{s.t. } Ax=b\\ \quad x\geqslant 0\end{array}$$

对偶线性规划

$$\text{(DP)}\quad \begin{array}{l}\max\ b^Ty\\ \text{s.t. } A^Ty\leqslant c\end{array}$$

若x_0,y_0分别是(LP)与(DP)的可行解,则

$Ax_0=b,\ x_0\geqslant 0,\ A^Ty_0\leqslant c$

于是$y_0^Tb=y_0^T(Ax_0)=(y_0^TA)x_0=(A^Ty_0)^Tx_0\leqslant c^Tx_0$

定理2.8.4 极小化线性规划问题的目标函数值不小于其对偶规划的目标函数值.

推论 若x_0是原始线性规划的可行解,y_0是对偶线性规划的可行解,且$c^Tx_0=b^Ty_0$,则x_0与y_0分别是原始线性规划问题与对偶线性规划问题的最优解.

推论2(补充) 对偶的线性规划都有最优解的充要条件是两者都有可行解.

定理2.8.6 若原始线性规划问题与对偶线性规划问题之一具有无界的目标函数值,则另一个无可行解.

对偶规划的性质

考虑如下的线性规划

$$\text{(LP)}\quad \begin{array}{l}\min\ 5x_1+3x_2\\ \text{s.t. } x_1+2x_2-x_3=2\\ \quad 4x_1+x_2-x_4=3\\ \quad x_1,x_2,x_3,x_4\geqslant 0\end{array}$$

考虑基解$(0,1,0,-2)^T$(不可行).

对应的规范式为

$$\begin{array}{l}\min\ 3+7x_1/2+3x_3/2\\ \text{s.t. } x_1/2+x_2-x_3/2=1\\ \quad 7x_1/2-x_3/2+x_4=-2\\ \quad x_1,x_2,x_3,x_4\geqslant 0\end{array}$$

其判别数向量$(7/2,0,3/2,0)$是非负的.

对偶规划的性质

判别数向量用(LP)中的符号可以记为

$\sigma=c^T-c_B^TB^{-1}A$

令$y=(c_B^TB^{-1})^T$,

则判别数向量为$\sigma=c^T-y^TA$.

上述的判别数向量非负,因此有

$c^T-y^TA\geqslant 0$,

即$A^Ty\leqslant c$.

这说明y是对偶规划的可行解.

对偶规划的性质

$$\text{(DP)}\quad \begin{array}{l}\max\ 2y_1+3y_2\\ \text{s.t. } y_1+4y_2\leqslant 5\\ \quad 2y_1+y_2\leqslant 3\\ \quad y_1\leqslant 0\\ \quad y_2\leqslant 0\end{array}$$

若$B=(p_2,p_4)$,$x=(0,1,0,-2)^T$,

$\sigma=(7/2,0,3/2,0)^T$,则$y=(3/2,0)^T$.

将其代入到(DP)的约束条件中,可以发现,约束的松弛量为$(7/2,0,3/2,0)^T$,恰好为(LP)中基解的判别数向量.

判别数=松弛量

判别数非负 ⇒ 松弛量非负 ⇒ 对偶解可行

对偶规划的性质

$y=(c_B^TB^{-1})^T$

这一可行解对应的目标函数值为

$y^Tb=c_B^TB^{-1}b.$

上式右端即为在原始线性规划中基解对应的目标函数值.

对偶的线性规划问题的解

原始规划的最优性	⇔	对偶规划的可行性
$c^T-c_B^TB^{-1}A\geqslant 0$	$y=(c_B^TB^{-1})^T$	$A^Ty\leqslant c$

定理2.8.7 设B是问题

$\min c^Tx$

s.t. $Ax=b$

$x\geqslant 0$

的一个基矩阵,对应的基解满足最优性条件

$c^T-c_B^TB^{-1}A\geqslant 0$,

则对偶线性规划问题有可行解$y=(c_B^TB^{-1})^T$,且$c_B^Tx_B=y^Tb$

对偶的线性规划问题的解

原始规划的最优性	⇔	对偶规划的可行性
$c^T-c_B^TB^{-1}A\geqslant 0$	$y=(c_B^TB^{-1})^T$	$A^Ty\leqslant c$

在上述条件下,(LP)的基解的目标函数值等于(DP)的解的目标函数值.

如果x_B是可行解,则显然是(LP)的最优解.

此时对应的y是(DP)的最优解.

对偶性定理

定理2.8.5 若原始线性规划问题与对偶线性规划问题之一有最优解,则另一个也有最优解,并且它们目标函数的最优值相等.

对偶性定理

证明:考虑到对合性,只需选取两个规划中的任一个证明.

设 $\min c^Tx$

s.t. $Ax=b$

$x\geqslant 0$

有最优解x_0,且其为基可行解.

再设基矩阵为B,$c^Tx_0=c_B^TB^{-1}b$

x_0为最优解,判别数非负,有

$c^T-c_B^TB^{-1}A\geqslant 0$

令$y_0=(c_B^TB^{-1})^T$,则

$A^Ty_0\leqslant c$, y_0可行,且

$y_0^Tb=c_B^TB^{-1}b=c^Tx_0$

y_0是对偶规划最优解

对偶的线性规划问题的解

两个互为对偶的线性规划的解的情况

(1) 两个都有可行解

两个都有最优解,最优值相等

一个有最优解

(2)两个都无可行解 (书中有错)

(3)一个有可行解,无最优解(目标函数无界),则另一个无可行解

互补松弛性

$Ax=b$

$(A^Ty-c)^Tx=0 \Leftrightarrow b^Ty=c^Tx$

定理2.8.8 x,y分别是原始线性规划问题与对偶线性规划的可行解,则x,y分别是最优解的充要条件为$(A^Ty-c)^Tx=0$.

互补松弛性

原始线性规划

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

$$\text{s.t.}\ \begin{aligned} x_1+x_2+x_3 &=6\\ x_1+2x_2+x_4 &=8\\ x_1+x_5 &=4\\ x_2+x_6 &=3\\ x_1,x_2,x_3,x_4,x_5,x_6 &\geqslant 0\end{aligned}$$

对偶线性规划

$$\max f = 6y_1+8y_2+4y_3+3y_4$$

$$\text{s.t.}\ \begin{aligned} y_1+y_2+y_3 &\leqslant -2\\ y_1+2y_2+y_4 &\leqslant -3\\ y_1 &\leqslant 0\\ y_2 &\leqslant 0\\ y_3 &\leqslant 0\\ y_4 &\leqslant 0\end{aligned}$$

原始规划最优解$x^*=(4,2,0,0,0,1)^T$,

对偶规划最优解$y^*=(0,-3/2,-1/2,0)^T$.

互补松弛性: $q^Tx=0$

对偶规划y^*的松弛量$q=c-A^Ty=(0,0,0,3/2,1/2,0)^T$

单纯形方法→对偶单纯形法

用单纯形方法求解线性规划时,互补松弛条件
$(A^Ty-c)^Tx=0$
始终满足(非基变量为零,基变量判别数为零).

求解时,x的可行性是满足的,但是最优性条件不满足即y不可行.最终最优性条件满足,x,y都可行,从而求得最优解.

我们构造另一种方法,最优性条件始终满足,即y可行,而x不可行.最终x可行,从而是最优解.

§2.9 对偶单纯形法

单纯形方法与对偶单纯形方法

判别数向量非负　$y=(c_B^TB^{-1})^T$为对偶规划的可行解

单纯形方法 | 保证解可行：最优性条件不满足(有负判别数) ⇒ 最优性条件满足(判别数非负)；对偶规划解不可行 ⇒ 对偶规划解可行

对偶单纯形方法 | 保证对偶规划解可行：解不可行 ⇒ 解可行

两种方法都始终保证$(A^Ty-c)^Tx=0$
所以只要x,y都可行,必然最优

1. 最优解的判别

已知线性规划问题的基矩阵B及它对应的基解,并且此基解的所有判别数非负.

若

$$x_B=B^{-1}b\geqslant 0$$

则所得的基解为最优解

2. 确定离基变量

令min{$(B^{-1}b)_i$|$(B^{-1}b)_i<0$}=$(B^{-1}b)_l$,

则以x_l为离基变量

若x_l所在行的所有系数$a_{lj}\geqslant 0(j=1,2,\cdots,n)$,则线性规划问题无可行解.

(此时若有可行解$0 \leqslant b_l=a_{l1}x_1+a_{l2}x_2+\cdots+a_{ln}x_n$)

3.确定进基变量

设目标函数的形式为 $f = f_0+\sum_{j\in T}\sigma_j x_j$

已确定离基变量为x_l,设进基变量为x_k.在目标函数中,用x_k替换x_l,

$$b_l=\sum_{\substack{j\in T\\ j\neq k}} a_{lj}x_j + a_{lk}x_k + x_l$$

$$x_l=\frac{1}{a_{lk}}(b_l-\sum_{\substack{j\in T\\ j\neq k}} a_{lj}x_j - x_l)$$

得:

$$f=(f_0+\sigma_k\frac{b_l}{a_{lk}})+\sum_{\substack{j\in T\\ j\neq k}}(\sigma_j-\sigma_k\frac{a_{lj}}{a_{lk}})x_j-\frac{\sigma_k}{a_{lk}}x_l$$

3.确定进基变量

$$f=(f_0+\sigma_k\frac{b_l}{a_{lk}})+\sum_{\substack{j\in T\\ j\neq k}}(\sigma_j-\sigma_k\frac{a_{lj}}{a_{lk}})x_j-\frac{\sigma_k}{a_{lk}}x_l$$

要求
$$\begin{cases}-\dfrac{\sigma_k}{a_{lk}}\geq 0 & (2.35)\\ \sigma_j-\sigma_k\dfrac{a_{lj}}{a_{lk}}\geq 0 & (2.36)\end{cases}$$

由(2.35)式,$a_{lk}<0$

若$a_{lj}\geqslant 0$,(2.36)恒成立

若$a_{lj}<0$,要求

令 $\frac{\sigma_k}{a_{lk}}=\max\{\frac{\sigma_j}{a_{lj}}\,|\,a_{lj}<0\}$ 则x_k为进基变量

算例

例2.5.1 用对偶单纯形法解线性规划

min $z=12x_1+8x_2+16x_3+12x_4$

s.t. $2x_1+x_2+4x_3\geqslant 2$

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

$x_1,x_2,x_3,x_4\geqslant 0$

引入松弛变量得到标准型线性规划

min $z=12x_1+8x_2+16x_3+12x_4$

s.t. $-2x_1-x_2-4x_3+x_5=-2$

$-2x_1-2x_2-4x_4+x_6=-3$

$x_1,x_2,x_3,x_4,x_5,x_6\geqslant 0$

min $z=12x_1+8x_2+16x_3+12x_4$

s.t. $-2x_1-x_2-4x_3+x_5=-2$

$-2x_1-2x_2-4x_4+x_6=-3$

构造对偶单纯形表

选取离基变量

$b_6=-3=\min\{b_j|b_j<0\}$

c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6
c_j			12	8	16	12	0	0
0	p_5	-2	-2	-1	4	0	1	0
0	p_6	-3	-2	-2	0	(-4)	0	1
σ_j			12	8	16	12	0	0
$\sigma_j/a_{lj}(a_{lj}<0)$			-6	-4		-3		

选取进基变量 $\sigma_4/a_{64}=\max\{\sigma_j/a_{6j}|a_{6j}<0\}$ 主元

c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6
c_j			12	8	16	12	0	0
0	p_5	-2	-2	-1	4	0	1	0
0	p_6	-3	-2	-2	0	-4	0	1
σ_j			12	8	16	12	0	0
$\sigma_j/a_{lj}(a_{lj}<0)$			-6	-4		-3		
0	p_5	-2	-2	(-1)	-4	0	1	0
12	p_4	3/4	1/2	1/2	0	1	0	-1/4
σ_j			6	2	16	0	0	3
$\sigma_j/a_{lj}(a_{lj}<0)$			-3	-2	-4			

0	p_5	-2	-2	(-1)	-4	0	1	0
12	p_4	3/4	1/2	1/2	0	1	0	-1/4
σ_j			6	2	16	0	0	3
$\sigma_j/a_{lj}(a_{lj}<0)$			-3	-2	-4			
8	p_2	2	2	1	4	0	-1	0
12	p_4	-¼	- ½	0	(-2)	1	½	-¼
σ_j			2	0	8	0	2	3
$\sigma_j/a_{lj}(a_{lj}<0)$			-4		-4			-12
8	p_2	3/2	1	1	0	2	0	-1/2
16	p_3	1/8	¼	0	1	-½	-¼	1/8
σ_j			0	0	0	4	4	2

基解已可行,最优解为

$x^*=(0,3/2,1/8,0)^T$

最优值为

$z^*=14$

§2.10 灵敏度分析

灵敏度分析

在线性规划中,已经求得了最优解.

若系数c_j,b_i,a_{ij}等发生变化的时候,最优解是否发生变化?怎样变化?

这一类问题称为灵敏度分析.

通常,我们进行灵敏度分析时,都假设只有一个系数发生变化.

灵敏度分析

设已经用单纯形法求得线性规划

$$\text{(LP)}\quad \begin{aligned} &\min c^T x \\ &\text{s.t. } Ax=b \\ &\quad x\geqslant 0 \end{aligned}$$

的最优解$x_B=B^{-1}b$,基矩阵$B=(p_1,p_2,\cdots,p_m)$(称为最优基),最优值$z^*=c_B^T x_B=c_B^T B^{-1}b$.

非基变量价格系数的变化

若x_r是非基变量,其价格系数c_r的变化不影响解的可行性.

由于$\sigma_k=c_k-c_B^T B^{-1}p_k$,$c_r$的变化只影响$x_r$的判别数,不影响其它变量的判别数.

设c_r的变化量为Δc_r,

$\sigma_r'=c_r+\Delta c_r-c_B^T B^{-1}p_r=\sigma_r+\Delta c_r$

因此,若$\sigma_r+\Delta c_r\geqslant 0$,即$\Delta c_r\geqslant -\sigma_r$
最优解,最优值都不变.

基变量价格系数的变化

若x_r是基变量,其价格系数c_r的变化不影响解的可行性.

由于$\sigma_k=c_k-c_B^T B^{-1}p_k$,$c_r$变化时,$c_B$发生变化,所有的非基变量的判别数都发生变化.

设c_r的变化量为Δc_r,$c_B'=c_B+\Delta c_B$

其中$\Delta c_B=(0,\cdots,0,\Delta c_r,0,\cdots,0)^T$

$\sigma_j'=c_j-c_B'^T B^{-1}p_j=c_j-c_B^T B^{-1}p_j-\Delta c_B^T B^{-1}p_j=\sigma_j-\Delta c_r a_{rj}$

基变量价格系数的变化

$\sigma_j' = \sigma_j - \Delta c_r a_{rj}$

在所有的判别数仍然非负时,最优解不变,根据上面的式子,可以得到

$\max\{\sigma_j/a_{rj} \mid a_{rj}<0\} \leqslant \Delta c_r \leqslant \min\{\sigma_j/a_{rj} \mid a_{rj}>0\}$

在满足上述条件的情况下,最优值由原来的z^*变为 $z^*+\Delta c_r x_r$

实际讨论时,只要写出新的判别数,然后解不等式即可求得价格系数的变化范围.

例2.10.1

已知用单纯形方法解线性规划的最后的单纯形表,问价格系数c_3在什么范围内变化时,最优解不变.

$\min y=-3x_1+x_2+x_3+10x_6+10x_7$

s.t. $x_1-2x_2+x_3+x_5=11$

$-4x_1+x_2+2x_3-x_4+x_6=3$

$-2x_1+x_3+x_7=1$

$x_1,\cdots,x_7\geqslant 0$

c_j			-3	1	1	0	0	10	10
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	p_7
-3	p_1	4	1	0	0	-2/3	1/3	2/3	-5/3
1	p_2	1	0	1	0	-1	0	1	-2
1	p_3	9	0	0	1	-4/3	2/3	4/3	-7/3
σ_j			0	0	0	1/3	1/3	29/3	28/3

若c_3的系数增加了变化量Δc_3,则目标函数变化了

$\Delta c_3\, x_3=\Delta c_3\,(4x_4/3-2x_5/3-4x_6/3+7x_7/3)$

因此,x_4到x_7的判别数分别变为

$1/3+4/3\,\Delta c_3, 1/3-2/3\,\Delta c_3, 29/3-4/3\,\Delta c_3, 28/3+7/3\,\Delta c_3$

因此,x_4到x_7的判别数分别变为

$1/3+4/3\,\Delta c_3, 1/3-2/3\,\Delta c_3, 29/3-4/3\,\Delta c_3, 28/3+7/3\,\Delta c_3$

要使得最优解不变,只需判别数均非负,即

$1/3+4/3\,\Delta c_3\geqslant 0$,

$1/3-2/3\,\Delta c_3\geqslant 0$,

$29/3-4/3\,\Delta c_3\geqslant 0$,

$28/3+7/3\,\Delta c_3\geqslant 0$.

解得$-1/4\leqslant\Delta c_3\leqslant 1/2$

即$3/4\leqslant c_3\leqslant 3/2$

在实际应用中,如果价格系数的变化导致判别数变为负数,由于原来最优解的可行性不变,可以继续用单纯形方法求解.

右端项的变化

右端项变化,原来的最优解不再可行.

原始线性规划

右端项的变化 ⟺ 价格系数的变化

判别数不变 ⟺ 对偶解的可行性不变

按照基重新写出的基解的可行性不变 ⟺ 判别数的非负性不变

最优基B不变 ⟺ 对偶解$y=(c_B^T B^{-1})^T$的最优性不变

什么条件下成立?

对偶线性规划

B仍然是最优基?

$x_B'=B^{-1}(b+\Delta b)=B^{-1}b+B^{-1}\Delta b=x_B+B^{-1}\Delta b$,

$B^{-1}=(b_{ij})_{m\times m}, \Delta b=(0,\cdots,0,\Delta b_r,0,\cdots,0)^T$

$x_B'=x_B+\Delta b_r(b_{1r},b_{2r},\cdots,b_{mr})^T$,

$(x_B')_i=(x_B)_i+\Delta b_r\, b_{ir}, (i=1,2,\cdots,m)$.

当$x_B'\geqslant 0$,即$(x_B)_i+\Delta b_r\, b_{ir}\geqslant 0$时,最优基$B$不变.

即

$\max\{-(x_B)_i/b_{ir} \mid b_{ir}>0\}\leqslant\Delta b_r\leqslant\min\{-(x_B)_i/b_{ir} \mid b_{ir}<0\}$

影子价格

最优基不变时,容易写出新的最优解.但是直接求最优值的变化比较困难.

从对偶线性规划的角度考虑,容易知道,最优值的变化为 $y_r\Delta b_r$,

其中$y=(c_B^T B^{-1})^T=(y_1,y_2,\cdots,y_m)^T$,是对偶规划最优解.

$y=(c_B^T B^{-1})^T$称为线性规划的影子价格向量,它的第r个分量为线性规划关于第r种资源的影子价格.

如果由于Δb_r的变化导致了x_B'不再可行,可以用对偶单纯形方法继续求解.

约束矩阵中系数的变化

设系数a_{rk}发生变化,此时若x_k为基变量,则原来的最优解的可行性,判别数都发生变化,讨论相当复杂.

如果x_k为非基变量,由于x_k的取值为0,不影响可行性($Ax=b$),不过判别数可能发生变化.

约束矩阵中系数的变化

根据$\sigma_j=c_j-c_B^T B^{-1}p_j$可知只有判别数$\sigma_j$发生了变化.

$\sigma_k'=c_j-c_B^T B^{-1}(a_{1k},\cdots,a_{rk}+\Delta a_{rk},\cdots,a_{mk})^T$

$=c_k-c_B^T B^{-1}p_k-c_B^T B^{-1}(0,\cdots,\Delta a_{rk},\cdots,0)^T=\sigma_k-y_r\Delta a_{rk}$

若$y_r>0$,则$\Delta a_{rk}\leqslant\sigma_k/y_r$时,最优基,最优解,最优值不变,

若$y_r<0$,则$\Delta a_{rk}\geqslant\sigma_k/y_r$时,最优基,最优解,最优值不变.

例2.10.2

已知用单纯形方法解线性规划的最后的单纯形表,问b_2在什么范围内变化时,最优基不变?

系数a_{24}在什么范围内变化时,最优基,最优解不变?

$$\min y=-3x_1+x_2+x_3+10x_6+10x_7$$
$$\text{s.t.}\quad \begin{aligned} x_1-2x_2+x_3+x_5&=11\\ -4x_1+x_2+2x_3-x_4+x_6&=3\\ -2x_1+x_3+x_7&=1\\ x_1,\cdots,x_7&\geqslant 0\end{aligned}$$

c_j			-3	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
c_B	B	b	p_1	p_2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
-3	p_1	4	1	0	0	-2/3	1/3	2/3	-5/3
1	p_2	1	0	1	0	-1	0	1	-2
1	p_3	9	0	0	1	-4/3	2/3	4/3	-7/3
σ_j			0	0	0	1/3	1/3	29/3	28/3

$$B^{-1}=\begin{pmatrix}1/3 & 2/3 & -5/3\\ 0 & 1 & 2\\ 2/3 & 4/3 & -7/3\end{pmatrix}$$

若b_2有改变量Δb_2,新的基解 $x_B=(4,1,9)^T$

$x_B'=x_B+B^{-1}\Delta b=(4,1,9)^T+\Delta b_2(2/3,1,4/3)^T$

当$x_B'=(4,1,9)^T+\Delta b_2(2/3,1,4/3)^T\geqslant 0$,即

$\Delta b_2\geqslant -1$,$b\geqslant 2$时,最优基不变.

由于 $B^{-1}=\begin{pmatrix}1/3 & 2/3 & -5/3\\ 0 & 1 & 2\\ 2/3 & 4/3 & -7/3\end{pmatrix}$

可以算出影子价格向量$y=(c_B^T B^{-1})^T=(-1/3,1/3,2/3)^T$

当a_{24}有改变量Δa_{24}时,新的判别数$\sigma_4'=\sigma_4-\Delta a_{24}/3$,而$\sigma_4=1/3$,所以$\Delta a_{24}\leqslant 1$时,最优基,最优解不变.

§2.11 整数线性规划

整数线性规划

如果线性规划的全部或部分变量要求取为整数,就称为整数线性规划,有时简称为整数规划.

所有的变量都要求是整数时,称为纯整数线性规划;部分变量要求取整数时,称为混合型整数线性规划.

求解的简单思路:先不考虑整数要求,求解一般的线性规划.若求得的最优解不满足整数要求时,用"舍入取整"的方法处理.

该方法有时会带来很大的误差,甚至得到的解不可行.

2.11.1 割平面法

考虑整数规划

$\min -x_1-x_2$

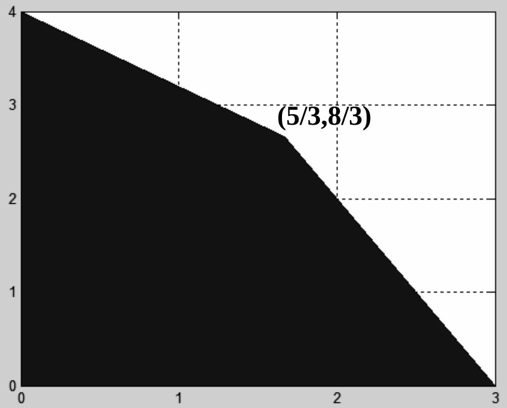
s.t. $2x_1+x_2\leqslant 6$

$4x_1+5x_2\leqslant 20$

$x_1,x_2\geqslant 0$且为整数

如果不考虑整数约束,其可行域见下图.

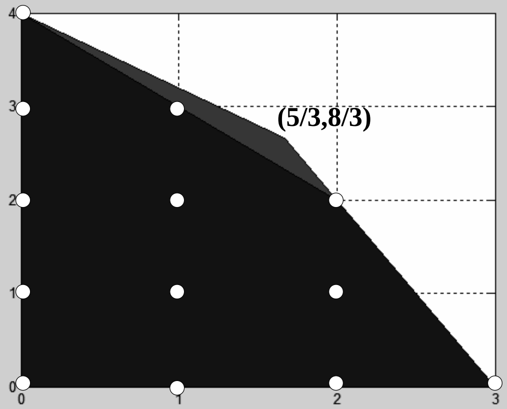
最优解为$(5/3,8/3)^T$.



割平面法

如果考虑整数约束,那么可行域是(LP)可行域内的13个整点.

我们考虑的方法之一是首先将(LP)的最优点附近的一块区域"挖掉"再求解,挖掉的区域内不含任何整点.



解(LP)最终单纯形表为

	c_j		-1	-1	0	0
c_B	B	b	p_1	p_2	p_3	p_4
-1	p_1	5/3	1	0	5/6	-1/6
-1	p_2	8/3	0	1	-2/3	1/3
	σ_j		0	0	1/6	1/6

考虑x_2对应的方程

由单纯形表得到

$$x_2 - \frac{2}{3}x_3 + \frac{1}{3}x_4 = \frac{8}{3}$$

$$x_2 - x_3 - 2 = \frac{2}{3} - \frac{1}{3}x_3 - \frac{1}{3}x_4$$

由于x_2–x_3–2为整数,可得

$$\frac{2}{3} - \frac{1}{3}x_3 - \frac{1}{3}x_4 \le 0$$

化为等式约束

$-x_3-x_4+x_5= -2$

割平面法

考虑到

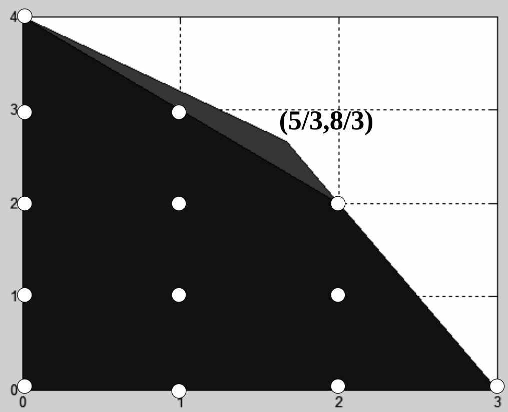
x_3=6-(2x_1+x_2),

x_4=20-(4x_1+5x_2),

根据2/3-x_3/3-x_4/3 ≤0

可以得到x_1+x_2 ≤4

可以看出,(LP)中的可行域确实去掉了一块不含整点的部分.



割平面法

设纯整数线性规划为

min c^Tx

(ILP) s.t. $Ax=b$

$x \geqslant 0$

x_j为整数,j=1,2,···,n

其中A,c,b中的元素为整数.

去掉其中x_j为整数的要求,得到的线性规划(LP)称为与(ILP)相对应的线性规划.

割平面方程

设对应的线性规划(LP)的最优解为 ,再设其中的x_r为非整数的基变量.由最终的单纯形表可写出约束方程

$$x_r + \sum_{i \in T} \alpha_{rj} x_j = \bar{x}_r$$

其中T表示非基变量的下标集.

把 与α_{rj}分解为整数部分与非负真分数(纯小数),即

$$\bar{x}_r = N_r + f_r, 0 < f_r < 1,$$
$$\alpha_{rj} = N_{rj} + f_{rj}, 0 \le f_{rj} < 1, j \in T$$

代入到约束方程中得到

割平面方程

$$x_r + \sum_{j \in T} N_{rj} x_j - N_r = f_r - \sum_{j \in T} f_{rj} x_j$$

所有的变量都取非负整数时,上式的左端是整数,因此右端也是整数.

而右端是小于1的,从而小于等于0.

$$f_r - \sum_{j \in T} f_{rj} x_j \le 0$$

这是所有变量取非负整数时必须满足的条件.称之为割平面方程.

增加割平面方程的规划

将割平面方程添加到纯整数规划的约束条件中,得到新的规划问题

$$\min\ c^T x$$

$$\text{(ILP) s.t. } Ax=b$$

$$f_r - \sum_{j\in T} f_{rj}x_j \le 0$$

$$x \geqslant 0$$

$$x_j\text{为整数}, j=1,2,\cdots,n$$

该整数规划对应的线性规划可以用对偶单纯形方法求解.

割平面法示例

例2.11.1 求解整数规划

min $-x_1-x_2$

s.t. $2x_1+x_2 \leqslant 6$

$4x_1+5x_2 \leqslant 20$

$x_1,x_2 \geqslant 0$且为整数

解:此规划的标准形式为

min $-x_1-x_2$

s.t. $2x_1+x_2+x_3 \quad =6$

$4x_1+5x_2 \quad +x_4=20$

$x_1,x_2,x_3,x_4 \geqslant 0$且为整数

解与之对应的线性规划,得到最优解$(5/3,8/3,0,0)^T$.

相应的最终单纯形表为

x_1与x_2都不是整数.我们不妨先考虑x_2.

由单纯形表得到

$$x_2 - \frac{2}{3}x_3 + \frac{1}{3}x_4 = \frac{8}{3}$$

$$x_2 - x_3 - 2 = \frac{2}{3} - \frac{1}{3}x_3 - \frac{1}{3}x_4$$

由于x_2-x_3-2为整数,可得割平面方程

$$\frac{2}{3} - \frac{1}{3}x_3 - \frac{1}{3}x_4 \le 0$$

化为等式约束

$-x_3-x_4+x_5=-2$

解与之对应的线性规划,得到最优解$(5/3,8/3,0,0)^T$.

相应的最终单纯形表为

	c_j		-1	-1	0	0
c_B	B	b	p_1	p_2	p_3	p_4
-1	p_1	5/3	1	0	5/6	-1/6
-1	p_2	8/3	0	1	-2/3	1/3
	σ_j		0	0	1/6	1/6

x_1与x_2都不是整数.我们不妨先考虑x_2.

将约束$-x_3-x_4+x_5=-2$添加到原问题中,再求解对应的线性规划.将新增加的约束置于单纯形表的最后一行.

	c_j		-1	-1	0	0	0
c_B	B	b	p_1	p_2	p_3	p_4	p_5
-1	p_1	5/3	1	0	5/6	-1/6	0
-1	p_2	8/3	0	1	-2/3	1/3	0
0	p_5	-2	0	0	-1	-1	1
	σ_j		0	0	1/6	1/6	0

该解已经是整数解.从而原规划的最优解为$x^*=(0,4)^T$,最优值为-4.

用对偶单纯形方法求解此问题,得$x=(0,4,2,0,0)^T$

分枝定界法

对于整数规划(ILP),求解与之对应的线性规划(LP),如果(LP)的最优解满足整数要求,那么就是(ILP)的最优解.

否则,设(LP)的最优解的分量x_r不是整数,设

$x_r=N_r+f_r, 0<f_r<1, N_r$为整数,

(ILP)的整数可行解必满足下面两个条件之一

$x_r \leqslant N_r$或$x_r \geqslant N_r+1$

分枝定界法

根据上面的两个条件将原规划划分为两个子问题.对这两个子问题,依然不考虑整数要求,求解对应的线性规划的最优解,即求解下面两个线性规划问题.

$$(P1)\quad \begin{aligned} \min\ & z=c^Tx \\ \text{s.t. } & Ax=b \\ & x_r \leqslant N_r \\ & x \geqslant 0 \end{aligned} \qquad (P2)\quad \begin{aligned} \min\ & z=c^Tx \\ \text{s.t. } & Ax=b \\ & x_r \geqslant N_r+1 \\ & x \geqslant 0 \end{aligned}$$

分枝定界法

设(P1),(P2)的最优解分别为x_1,x_2最优值为z_1,z_2.考虑$z_1 \leqslant z_2$的情况.

若x_1为整数解,则x_1为原来的问题(ILP)的解;

若x_1不是整数解,再将(P1)划分为两个线性规划子问题.在求出这两个子问题的解之后再考虑是否需要划分(P2).

算例

例2.11.1用分枝定界方法求解整数规划

$$(P0)\quad \begin{aligned} \min\ & z=5x_1+3x_2 \\ \text{s.t. } & 3x_1+4x_2 \geqslant 12 \\ & 5x_1+2x_2 \geqslant 10 \\ & x_1,x_2 \geqslant 0 \text{且为整数} \end{aligned}$$

解:解与(P0)对应的线性规划,得到最优解与最优值分别为

$x=(8/7,15/7)^T$

$z=85/7 \approx 12.14$

算例

x_1,x_2都不是整数,不妨先根据x_1进行分枝.

x_1=1+1/7,构造以下两个线性规划子问题.

$$(P1)\quad \begin{aligned} \min\ & z=5x_1+3x_2 \\ \text{s.t. } & 3x_1+4x_2 \geqslant 12 \\ & 5x_1+2x_2 \geqslant 10 \\ & x_1 \leqslant 1 \\ & x_1,x_2 \geqslant 0 \end{aligned} \qquad (P2)\quad \begin{aligned} \min\ & z=5x_1+3x_2 \\ \text{s.t. } & 3x_1+4x_2 \geqslant 12 \\ & 5x_1+2x_2 \geqslant 10 \\ & x_1 \geqslant 2 \\ & x_1,x_2 \geqslant 0 \end{aligned}$$

算例

分别解上面两个线性规划,(P1)的最优解$(1,2.5)^T$,最优值为12.5,

(P2)的最优解$(2,1.5)^T$,最优值为14.5.

(P1)的最优值较小,在将(P1)分划为两个子问题.

$$(P3)\quad \begin{aligned} \min\ & z=5x_1+3x_2 \\ \text{s.t. } & 3x_1+4x_2 \geqslant 12 \\ & 5x_1+2x_2 \geqslant 10 \\ & x_1 \leqslant 1 \\ & x_2 \leqslant 2 \\ & x_1,x_2 \geqslant 0 \end{aligned} \qquad (P4)\quad \begin{aligned} \min\ & z=5x_1+3x_2 \\ \text{s.t. } & 3x_1+4x_2 \geqslant 12 \\ & 5x_1+2x_2 \geqslant 10 \\ & x_1 \leqslant 1 \\ & x_2 \geqslant 3 \\ & x_1,x_2 \geqslant 0 \end{aligned}$$

无可行解

算例

(P4)的最优解为$(1,3)^T$(已是整数解), 最优值为14.

由于(P2)的最优值比(P4)的最优值大.所以(P2)不再分划.

原问题的最优解为$(1,3)^T$, 最优值为14.

第三章
无约束最优化方法

作业

(1)对于任意给定的一个函数$f(x)$,用最速下降法求解其最小值,若一维搜索是精确的,是否两个相邻的搜索方向一定正交?

若是,请证明;若否,请举反例.

(2)对P137中3.9题的两小题,分别用最速下降法,Newton算法,FR算法,PRP算法,DFP算法计算两步.

(3,4,5)P137 3.10,3.11, P138 3.17

§3.1 最优性条件

设$\varphi(\alpha)$为一元函数,若α^*为$\varphi(\alpha)$的局部极小点,则$\varphi'(\alpha)=0$.

定理 3.1.1 (一阶必要条件)

若x^*为$f(x)$的局部极小点,且在x^*的某邻域内具有一阶连续偏导数,则

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

二阶充分条件

设$\varphi(\alpha)$为一元函数,若$\varphi'(\alpha^*)=0$, $\varphi''(\alpha^*)>0$ 则α^*为$\varphi(\alpha)$的严格局部极小点,

定理 3.1.2(二阶充分条件)

若在x^*的某邻域内$f(x)$有二阶连续偏导数且$g^*=0$,$G(x^*)$正定,则x^*为问题(3.1)的严格局部极小点.

定理 3.1.3(二阶必要条件)

若x^*为$f(x)$的局部极小点,且在x^*的某邻域内$f(x)$有二阶连续偏导数,则$g^*=0$,$G(x^*)$半正定.

定理 3.1.4 设$f(x)$ 在R^n上是凸函数且有一阶连续偏导数,则x^*为$f(x)$的整体极小点的充分必要条件是$g^*=0$.

§3.2 最速下降法

在某一点x,负梯度方向$p=-g(x)$是使目标函数$f(x)$下降最快的方向,称为最速下降方向.

例 3.2.1用最速下降法求解

$$\min\ f(x)=\frac{1}{2}x_1^2+\frac{9}{2}x_2^2$$ 设初始点为$(9,1)^T$.

引理 对于正定二次函数 $H(x)=\frac{1}{2}x^TGx+b^Tx+c$
精确一维搜索问题 $\min_{\alpha>0} H(x_k+\alpha p_k)$ 的解为

$$\alpha_k=-\frac{g_k^T p_k}{p_k^T G_k p_k}.$$

证明思路：考虑函数$H(x_k+\alpha p_k)$,其二次项与一次项系数分别为

$\alpha^2:\frac{1}{2}p_k^TGp_k$ $\qquad$ $\alpha: p_k^TGx_k+b^Tp_k=p_k^Tg_k$

因此 $\qquad \alpha_k=-\frac{g_k^T p_k}{p_k^TGp_k}$

例 3.2.1的解 $g(x)=\begin{pmatrix}x_1\\9x_2\end{pmatrix},G(x)=\begin{pmatrix}1&0\\0&9\end{pmatrix}$

对于正定二次函数,最速下降法的迭代公式为

$$x_{k+1}=x_k-\frac{g_k^Tg_k}{g_k^TGg_k}g_k$$

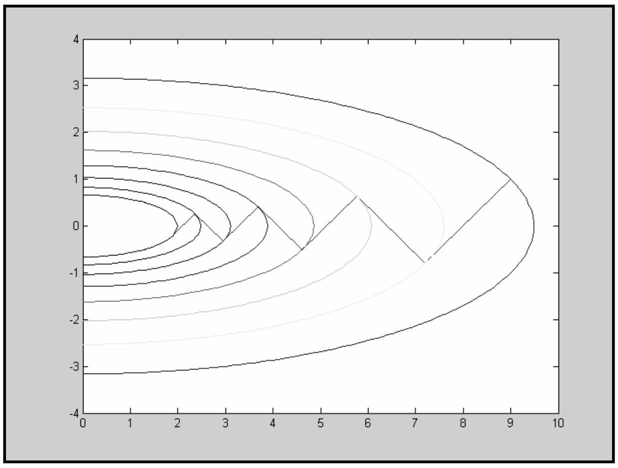
由于$x_0=(9,1)^T$,$g_0=(9,9)^T$,可得 $x_1=(7.2,-0.8)^T$

由归纳法可以证明 $x_k=0.8^k\bullet(9,(-1)^k)^T\ (k=1,2,\cdots)$

因此， $x_k\to x^*=(0,0)^T$ 且 $\|x_{k+1}-x^*\|/\|x_k-x^*\|=0.8$

算法是线性收敛的.

根据图形发现,两个相邻的方向是正交的.



3.2.2 收敛性

1 整体收敛性

定理 3.2.1 设$f(x)$具有一阶连续偏导数,给定$x_0\in R^n$,假定水平集$L=\{x\in R^n|f(x)\leqslant f(x_0)\}$有界,令$\{x_k\}$为最速下降法产生的点列，则

或者(i)对某个k_0,$g(x_{k0})=0$;

或者(ii) $g_k\to 0(k\to\infty)$.

2 用于二次函数时的收敛速度

定理 3.2.2 设 $f(x)=\frac{1}{2}x^TGx$,其中G为正定矩阵，设λ_1,λ_n表示G的最小与最大特征值，则最速下降法产生的点列$\{x_k\}$满足

$$f(x_{k+1})\le(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1})^2 f(x_k)(k=0,1,2,\cdots)$$

$$\|x_k\|\le\sqrt{\frac{\lambda_n}{\lambda_1}}(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1})^k\|x_0\|(k=0,1,2,\cdots)$$

最速下降法是线性收敛的,可以用 $\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}$ 来衡量其收敛速度。

§ 3.3 Newton法

设x_k为$f(x)$的一个近似极小点,将$f(x)$在x_k附近作Taylor展开,

$$f(x)\approx q_k(x)=f_k+g_k^T(x-x_k)+\frac{1}{2}(x-x_k)^TG_k(x-x_k)$$

其中$f_k=f(x_k)$,$g_k=g(x_k)$,$G_k=G(x_k)$.

若G_k正定,则$q_k(x)$有唯一极小点,该极小点即为Newton法取的x_{k+1}.

显然 $0=\nabla q_k(x_{k+1})=G_k(x_{k+1}-x_k)+g_k$

Newton迭代公式为 $x_{k+1}=x_k-G_k^{-1}g_k$

Newton法算例

例3.3.1 用Newton法求解 $\min f(x)=\frac{1}{2}x_1^2+\frac{9}{2}x_2^2$

设初始点为 $x_0=(9,1)^T$.

解 $g(x)=\begin{pmatrix}x_1\\9x_2\end{pmatrix},G(x)=\begin{pmatrix}1&0\\0&9\end{pmatrix}$

因此 $x_1=x_0-G_0^{-1}g_0=\begin{pmatrix}9\\1\end{pmatrix}-\begin{pmatrix}1&0\\0&9\end{pmatrix}^{-1}\begin{pmatrix}9\\9\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}=x^*$

对于正定二次函数,Newton法总是一步迭代得到极小点.

Newton法算例

例3.3.2 用Newton法求解

$$\min f(x)=(x_1-2)^4+(x_1-2)^2x_2^2+(x_2+1)^2$$

设初始点为 $x_0=(1,1)^T$.

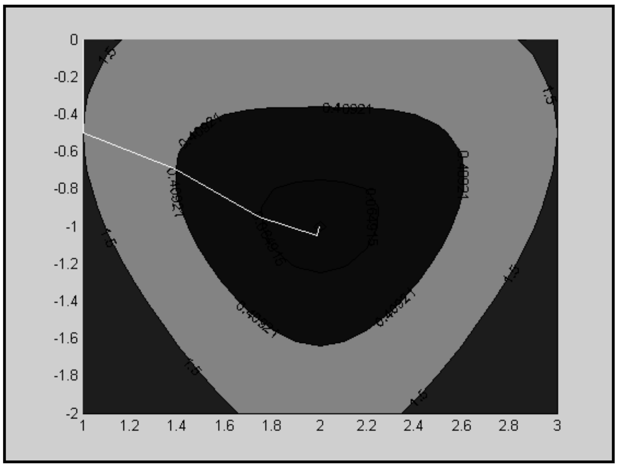
解 $g(x)=\begin{pmatrix}4(x_1-2)^3+2(x_1-2)x_2^2\\2(x_1-2)^2x_2+2(x_2+1)\end{pmatrix}$

$$G(x)=\begin{pmatrix}12(x_1-2)^2+2x_2^2 & 4(x_1-2)x_2\\4(x_1-2)x_2 & 4(x_1-2)^2+2\end{pmatrix}$$

根据Newton迭代公式$x_{k+1}=x_k-G_k^{-1}g_k$可以得到下面的结果.

结果:

k	x_k	$f(x_k)$
0	$(1,1)^T$	6.0
1	$(1,-0.5)^T$	1.5
2	$(1.3913043,-0.69565217)^T$	4.09e-1
3	$(1.7459441,-0.94879809)^T$	6.49e-2
4	$(1.9862783,-1.0482081)^T$	2.53e-3
5	$(1.9987342,-1.0001700)^T$	1.63e-6
6	$(1.9999996,-1.0000016)^T$	2.75e-12



Newton法的收敛性

定理3.3.1设$f(x)$是某一开集内的三阶连续可微函数,且它在该开集内有极小点x^*,设$G^*=G(x^*)$正定,则当x_0与x^*充分接近时,对一切k,Newton法有定义,且当$\{x_k\}$为无穷点列时,$\{x_k\}$二阶收敛于x^*, 即$h_k\to 0$且$\|h_{k+1}\|=O(\|h_k\|^2)$

其中$h_k=x_k-x^*$.

Newton法的优缺点

优点

(1)如果G^*正定且初始点合适,算法二阶收敛;

(2)对正定二次函数,迭代一次就可以得到极小点;

缺点

(1)对多数问题算法不是整体收敛的;

(2)在每次迭代中需要计算G_k;

(3)每次迭代时需要求解线性方程组;该方程组可能是奇异或病态的;求出的方向可能不下降;

(4)收敛于鞍点或极大点的可能性与收敛于极小点的可能性一致.

Newton法的改进

(1)阻尼Newton法

以 $p_k = -G_k^{-1} g_k$ 为下降方向进行精确一维搜索.

(2)在 G_k 正定时,取 $p_k = -G_k^{-1} g_k$;否则取 $p_k = -M_k^{-1} g_k$

(3)拟Newton方法

共扼方向法的思路

对于简单的二次函数

$$\frac{1}{2}x^T x + b^T x + c = \frac{1}{2}(x+b)^T(x+b) + c - b^T b$$

任给一个初始向量 $x^{(0)}$,沿着方向 $e_1=(1,0,\cdots,0)^T$ 进行搜索,即求解下面问题

$$\min_{\alpha_1} f_1(\alpha_1) = (x^{(0)} + \alpha_1 e_1 + b)^T (x^{(0)} + \alpha_1 e_1 + b)$$

由于 $f_1(\alpha_1) = (x_1^{(0)} + \alpha_1 + b_1)^2 + \sum_{i=2}^{n}(x_i^{(0)} + b_i)^2$

因此 $\alpha_1^* = -x_1^{(0)} - b_1,\ x^{(1)} = x^{(0)} + \alpha_1^* e_1 = (-b_1, x_2^{(0)}, \cdots, x_n^{(0)})^T.$

注:此处的一维搜索中 α_1 的范围是整个实数集,即 $x^{(1)}$ 是函数在集合 $\{x^{(0)}+\alpha_1 e_1, \alpha_1 \in R\}$ 中的极小点.

共扼方向法的思路

$x^{(1)}$ 与 $x^{(0)}$ 唯一不同的是它们的第一个分量.其中 $x^{(1)}$ 的第一个分量与原问题最优解 $-b$ 的第一个分量一致,其余的分量未发生变化.

下面再沿着方向 $e_2=(0,1,0,\cdots,0)^T$ 进行搜索,得到的 $x^{(2)}$ 的前两个分量与最优解 $-b$ 的的前两个分量一致,其余分量不变. $x^{(2)} = (-b_1, -b_2, x_3^{(1)}, \cdots, x_n^{(1)})^T.$

显然, $x^{(2)}$ 是函数在集合 $\{x^{(0)} + \alpha_1 e_1 + \alpha_2 e_2, \alpha_1, \alpha_2 \in R\}$ 中的极小点.

共扼方向法的思路

将此过程进行下去有 $x^{(k)} = (-b_1, \cdots, -b_k, x_{k+1}^{(1)}, \cdots, x_n^{(1)})^T.$

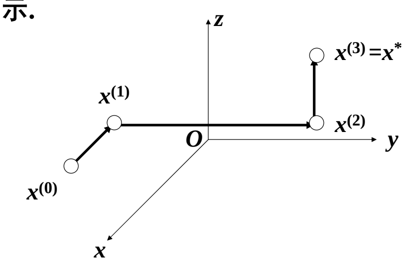
$x^{(k)}$ 是函数在 $\{x^{(0)} + \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_k e_k, \alpha_1, \alpha_2 \cdots, \alpha_k \in R\}$ 中的极小点.

进行 n 步后有 $x^{(n)} = (-b_1, -b_2, \cdots, -b_n)^T = -b.$

因此,上述的迭代过程每一步在一个分量上达到最优,且每一步求得了函数在一个集合中的极小点,这种集合在迭代过程中逐渐扩大,迭代 n 步之后得到原问题的最优解.

共扼方向法的思路

在三维情况下,上面的迭代过程可以用图形来表示.



共扼方向法的思路

上面的方法对一般的二次函数是否适用呢?

考虑问题 $\min f(x) = \frac{1}{2}x^T G x$ 其中 $G = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$

易见 G 是正定的, $f(x)$ 的极小点为 $(0,0)^T$.

以 $x^{(0)}=(-1,-1)^T$ 为初始点,在方向 $e_1=(1,0)^T$ 上进行一维搜索.即求解问题

$$\min_{\alpha} f((-1,-1)^T + \alpha_1(1,0)^T) = (\alpha_1 - 1)^2 - 4(\alpha_1 - 1) + 5$$

易求得 $\alpha_1^*=3$, $x^{(1)}=x^{(0)}+\alpha_1^* e_1=(2,-1)^T$.

第一个分量没有变为0,进一步沿 $e_2=(0,1)^T$ 搜索显然不能达到 $f(x)$ 的极小点.

共扼方向法的思路

给定最优化问题 $\min f(x)=\frac{1}{2}x^TGx+b^Tx+c$
(其中G对称正定)

在空间R^n上,重新定义内积与范数:

$\langle x,y\rangle = x^TGy \qquad \|x\|_G=\langle x,x\rangle^{1/2}=(x^TGx)^{1/2}$

则 $f(x)=\frac{1}{2}\|x\|_G^2+b^Tx+c$

$$=\frac{1}{2}(x+G^{-1}b)^TG(x+G^{-1}b)+c-\frac{1}{2}b^TG^{-1}b$$

$$=\frac{1}{2}\|x+G^{-1}b\|_G^2+c-\frac{1}{2}b^TG^{-1}b$$

共扼方向法的思路

因此原问题等价于 $\min\|x+G^{-1}b\|_G^2$

在R^n上,按照上面定义的内积给出一组正交基$p_1,p_2,\cdots,p_n$,

即$p_1,p_2,\cdots,p_n$线性无关,且$\langle p_i,p_j\rangle=0(i\neq j)$

设问题的最优解$x^*=-G^{-1}b$在这组基底下的表示为$x^*=u_1p_1+u_2p_2+\cdots+u_np_n$

任取初始点$x^{(0)}=s_1p_1+s_2p_2+\cdots+s_np_n$, 在方向$p_1$上进行一维搜索,即求解问题

$$\min\|(s_1+\alpha_1-u_1)p_1+(s_2-u_2)p_2+\cdots+(s_n-u_n)p_n\|_G^2$$

共扼方向法的思路

$$\|(s_1+\alpha_1-u_1)p_1+(s_2-u_2)p_2+\cdots+(s_n-u_n)p_n\|_G^2$$

$$=\langle(s_1+\alpha_1-u_1)p_1+(s_2-u_2)p_2+\cdots+(s_n-u_n)p_n,\ (s_1+\alpha_1-u_1)p_1+(s_2-u_2)p_2+\cdots+(s_n-u_n)p_n\rangle$$

$$=(s_1+\alpha_1-u_1)^2\|p_1\|_G^2+\sum_{i=2}^{n}(s_i-u_i)^2\|p_i\|_G^2$$

显然,当$\alpha_1=u_1-s_1$时,上面的函数取最小值,

$x^{(1)}=u_1p_1+s_2p_2+\cdots+s_np_n$.

即$x^{(1)}$与最优点在基底中第一个向量p_1前的系数达到一致.

$x^{(1)}$是函数在集合$\{x^{(0)}+\alpha_1p_1,\alpha_1\in R\}$中的极小点.

共扼方向法的思路

类似的,再依次沿着$p_2,\cdots,p_k$进行一维搜索,可以得到$x^{(k)}=u_1p_1+\cdots+u_kp_k+s_{k+1}p_{k+1}+\cdots+u_np_n$,

$x^{(k)}$是函数在$\{x^{(0)}+\alpha_1p_1+\alpha_2p_2+\cdots+\alpha_kp_k,\alpha_1,\alpha_2\cdots,\alpha_k\in R\}$中的极小点.

最终$x^{(n)}=u_1p_1+u_2p_2+\cdots+u_np_n=x^*$

即迭代过程同样在n步之后找到最优点.

因此,对二次函数 $f(x)=\frac{1}{2}x^TGx+b^Tx+c$

我们可以找到n个方向(向量),对其依次进行一维搜索,最多n步可以找到函数的最优点.

共扼方向法的思路

定义 3.4.2设n维向量组$p_1,\cdots,p_k$线性无关, $x^{(0)}\in R^n$,

称向量集合$\{x^{(0)}+\sum_{i=1}^{k}\alpha_ip_i\mid\alpha_i\in R,i=1,2,\cdots,k\}$为由点$x^{(0)}$与$p_1,p_2,\cdots,p_k$ 生成的k维超平面.

若$k=1$,上述集合表示以p_1为方向向量,且过点$x^{(0)}$的一条直线.

我们希望$x^{(k)}$是k维超平面的极小点,于是$x^{(n)}$是n维超平面(即整个R^n空间)的极小点.

超平面上极小点的判断

若函数$f(x)$连续可微, $p_1,p_2,\cdots,p_k$为一组线性无关的n维向量, $x^{(0)}\in R^n$,

$H_k=\{x^{(0)}+\sum_{i=1}^{k}\alpha_ip_i\mid\alpha_i\in R,i=1,2,\cdots,k\} \qquad \bar{x}=x^{(0)}+\sum_{i=1}^{k}\bar{\alpha}_ip_i\in H_k$

若 是$f(x)$在H_k上的极小点,则$p_1,p_2,\cdots,p_k$都不是下降方向,因此

$-p_1,-p_2,\cdots,-p_k$也不是下降方向,因此 $\nabla f(\bar{x})^Tp_i\le 0$

于是有 $\nabla f(\bar{x})^Tp_i=0(i=1,2,\cdots,k)$

超平面上极小点的判断

严格证明:H_k上的点可以表示为$x=x^{(0)}+\sum_{i=1}^{k}\alpha_i p_i$

定义 $h(\alpha_1,\cdots,\alpha_k)=f(x^{(0)}+\sum_{i=1}^{k}\alpha_i p_i)$

若 $\bar{x}=x^{(0)}+\sum_{i=1}^{k}\bar{\alpha}_i p_i\in H_k$ 是$f(x)$在H_k上的极小点.则

$0=\nabla h(\bar{\alpha}_1,\cdots,\bar{\alpha}_k)=(\bar{g}^T p_1,\cdots,\bar{g}^T p_k)^T$

其中$\bar{g}=\nabla f(\bar{x})$. 因此 $\bar{g}^T p_i=0(i=1,2,\cdots,k)$

超平面上极小点的判断

反之,若 $\nabla f(\bar{x})^T p_i=0(i=1,2,\cdots,k)$

如果$f(x)$是严格凸函数,对于 $x=x^{(0)}+\sum_{i=1}^{k}\alpha_i p_i\in H_k$

则 $f(x)>f(\bar{x})+\nabla f(\bar{x})^T(x-\bar{x})$

$$=f(\bar{x})+\sum_{i=1}^{k}(\alpha_i-\bar{\alpha}_i)\nabla f(\bar{x})^T p_i=f(\bar{x})$$

因此 $\bar{x}$ 是$f(x)$在H_k上的唯一极小点.

超平面上极小点的判断

引理 3.4.2 设$f(x)$为连续可微的严格凸函数,又$p_1,p_2,\cdots,p_k$为一组线性无关的n维向量, $x_1\in R^n$,则 $x_{k+1}=x_1+\sum_{i=1}^{k}\bar{\alpha}_i p_i$ 是$f(x)$在x_1与$p_1,p_2,\cdots,p_k$所生成的k维超平面H_k上唯一极小点的充分必要条件是

$$g_{k+1}^T p_i=0(i=1,2,\cdots,k).$$

注:若$k=n$,易推出在x_{k+1}的梯度为零向量.因此,这一引理是一常用定理(极小点梯度为0)的推广.

共扼方向法(用于二次函数)

给定一个初始点x_1,给出一个下降方向p_1,令$x_2=x_1+\alpha_1 p_1$,进行精确一维搜索,确定x_2,再确定p_2(方法待定).

已知k个点与k个方向之后,令$x_{k+1}=x_k+\alpha_k p_k$,进行精确一维搜索,确定x_{k+1},再确定p_{k+1}.

对于正定二次函数,确定p_{k-1}的准则是希望 x_k是目标函数在k维超平面上的极小点. x_n就是目标函数在整个空间的极小点.

共扼向量

对于$f(x)=x^TGx/2+b^Tx+c$,有$g(x)=Gx+b$,因此

$g_{j+1}-g_j=G(x_{j+1}-x_j)=\alpha_j Gp_j$, 因此

$$g_{j+1}^T p_i-g_j^T p_i=\alpha_j p_i^T Gp_j$$

根据引理3.4.2,左边应为零,于是搜索方向满足性质$p_i^TGp_j=0(i\neq j)$.

共扼向量:设G为n阶正定矩阵,$p_1,p_2,\cdots,p_k$为n维向量组,如果$p_i^TGp_j=0(i\neq j)$则称向量组 $p_1,p_2,\cdots,p_k$关于G共扼.

实际上是在新的内积意义下,这是一组正交向量.

共扼方向法(用于二次函数)

注:在前面讨论思路时,根据方向的共轭性(正交性)得到x_{k+1}是目标函数在k维超平面上的极小点(后面的定理3.4.3给出严格证明).

根据上一页的推导,根据极小点可以推出共轭性(正交性),即若一种迭代方法每次求出的是二次函数在k维超平面上的极小点,则对应的方向是共扼的.

基本概念

二次终止性

如果一个算法经过有限次迭代就得到正定二次函数的极小点,称该算法具有二次终止性.

共轭方向法

是一种迭代方法,每次所取方向与前面的方向关于G共轭,然后进行精确一维搜索确定步长及下一步的迭代点.

共轭方向的性质

定理3.4.1设G为n阶正定矩阵,非零向量组 $p_1,p_2,\cdots,p_k$关于G共轭,则此向量组线性无关.

证明:设存在常数$\alpha_1,\alpha_2,\cdots,\alpha_k$使得

$\alpha_1 p_1+\alpha_2 p_2+\cdots+\alpha_k p_k=0$,

以$p_i^T G$左乘上式,显然有$\alpha_i p_i^T G p_i=0$.

又,G是正定矩阵,$p_i\neq 0$,因此$\alpha_i=0(i=1,2,\cdots,k)$

于是$p_1,p_2,\cdots,p_k$线性无关.

共轭方向的性质

推论1设G为n阶正定矩阵,非零向量组$p_1,p_2,\cdots,p_n$关于G共轭,则此向量组构成 R^n的一组基.

推论2设G为n阶正定矩阵,非零向量组$p_1,p_2,\cdots,p_n$关于G共轭,若向量v与$p_1,p_2,\cdots,p_n$ 关于G共轭，则$v=0$.

共轭方向法(用于二次函数)

定理 3.4.3 设G是n阶正定阵,向量组$p_1,p_2,\cdots,p_k$关于G共轭,对正定二次函数$f(x)=x^TGx/2+b^Tx+c$

由任意初始点x_1开始,依次进行k次一维搜索,$x_{i+1}=x_i+\alpha_i p_i(i=1,2,\cdots,k)$

则(i)$g^T_{k+1}p_i=0$ $(i=1,2,\cdots,k)$.

(ii)x_{k+1}是二次函数在k维超平面H_k上的极小点.

推论 当$k=n$时,x_{n+1}为二次函数在R^n上的极小点.

共轭方向法(用于二次函数)

证明要点:只要证明$g^T_{k+1}p_i=0$.

$$
\begin{aligned}
g_{k+1}^T p_i &= (g_{k+1}^T - g_k^T)p_i + \cdots + (g_{i+2}^T - g_{i+1}^T)p_i + g_{i+1}^T p_i \\
&= (Gx_{k+1} - Gx_k)^T p_i + \cdots + (Gx_{i+2} - Gx_{i+1})^T p_i + g_{i+1}^T p_i \\
&= \alpha_k p_k^T G p_i + \cdots + \alpha_{i+1} p_{i+1}^T G p_i + g_{i+1}^T p_i \\
&= g_{i+1}^T p_i \\
&= 0
\end{aligned}
$$

$x_{k+1}=x_k+\alpha_k p_k$

$p_j^T G p_i=0$

精确一维搜索

共轭梯度法(共轭方向的形成)

我们首先讨论针对下面二次函数的共轭梯度法

$$f(x)=\frac{1}{2}x^TGx+b^Tx+c$$

给定初始点x_0,初始下降方向取为$p_0= -g_0$

从x_0出发,沿方向p_0进行一维搜索得x_1.

设p_1是$-g_1$与p_0的线性组合$p_1= -g_1+\beta_0 p_0$,

p_1与p_0共轭,于是 $0=p_0^TGp_1=-p_0^TGg_1+\beta_0 p_0^TGp_0$

因此 $\beta_0=\dfrac{p_0^TGg_1}{p_0^TGp_0}$

共扼梯度法(共扼方向的形成)

假设已经沿k个共扼方向$p_0, p_1,\cdots, p_{k-1}$逐次进行一维搜索得x_k.

若$g_k=g(x_k)=0$,则x_k已是极小点,否则构造下一个方向p_k.令p_k为$-g_k$以及$p_0,p_1,\cdots,p_k$的线性组合.

$$p_k=-g_k+\sum_{i=0}^{k-1}\beta_i p_i$$

用$p_j^TG(j=0,1,\cdots,k-1)$左乘上式

$$0=p_j^TGp_k=-p_j^TGg_k+\sum_{i=0}^{k-1}\beta_i p_j^TGp_i \quad =-p_j^TGg_k+\beta_j p_j^TGp_j$$

因此 $\beta_j=\dfrac{p_j^TGg_k}{p_j^TGp_j}$

共扼梯度法(共扼方向的形成)

$$p_k=-g_k+\sum_{i=0}^{k-1}\beta_i p_i \qquad 0=-p_j^TGg_k+\beta_j p_j^TGp_j$$

由于 $x_{j+1}=x_j+\alpha_j p_j$ 有 $Gp_j=\dfrac{1}{\alpha_j}G(x_{j+1}-x_j)$

再根据二次函数的性质,有 $G(x_{j+1}-x_j)=g_{j+1}-g_j$

因此 $p_j^TGg_k=\dfrac{1}{\alpha_j}g_k^T(g_{j+1}-g_j)$

由于x_k是由点x_0及向量$p_0,p_1,\cdots,p_{k-1}$得到的k维超平面上的极小点,因此$g_k^Tp_j=0(j=0,1,\cdots,k-1)$.

由p_j的构造方式 $p_j=-g_j+\sum_{i=0}^{j-1}\beta_{ij}p_i$

因此$g_k^Tg_j=0(j=0,1,\cdots,k-1)$.

共扼梯度法(共扼方向的形成)

$p_j^TGg_k=\dfrac{1}{\alpha_j}g_k^T(g_{j+1}-g_j)$ $\quad g_k^Tg_j=0(j=0,1,\cdots,k-1)$

因此 $p_j^TGg_k=0(j=0,1,\cdots,k-2)$

根据 $0=-p_j^TGg_k+\beta_j p_j^TGp_j$ 得

$\beta_j=0(j=0,1,\cdots,k-2)$ $\qquad \beta_{k-1}=\dfrac{g_k^TGp_{k-1}}{p_{k-1}^TGp_{k-1}}$

所以

$$p_k=-g_k+\sum_{i=0}^{k-1}\beta_i p_i=-g_k+\beta_{k-1}p_{k-1}=-g_k+\frac{g_{k-1}^TGp_{k-1}}{p_{k-1}^TGp_{k-1}}p_{k-1}$$

共扼梯度法(用于二次函数)

$$p_0=-g_0,\ p_k=-g_k+\beta_{k-1}p_{k-1},\ \beta_{k-1}=-\frac{g_k^TGp_{k-1}}{p_{k-1}^TGp_{k-1}}$$

定理3.4.4 对正定二次函数 $f(x)=\frac{1}{2}x^TGx+b^Tx+c$ 由上面三式所确定共扼方向并采用精确一维搜索得到的共扼梯度法,在$m(\leqslant n)$次迭代后可函数的极小点,并且对所有$i(1\leqslant i\leqslant m)$有

$$p_i^TGp_j=0,\ g_i^Tg_j=0,\ g_i^Tp_j=0,\ p_i^Tg_i=-g_i^Tg_i$$

其中 $j=0,1,\cdots,i-1.$

FR算法

$\beta_{k-1}=\dfrac{g_k^TGp_{k-1}}{p_{k-1}^TGp_{k-1}}$ 为了能将上述方法用于其它函数,我们必须消去系数中的G.

(1)Flecher-Reeves公式

$$Gp_{k-1}=G\frac{1}{\alpha_{k-1}}(x_k-x_{k-1})=\frac{1}{\alpha_{k-1}}(g_k-g_{k-1})$$

$$g_k^TGp_{k-1}=\frac{1}{\alpha_{k-1}}g_k^T(g_k-g_{k-1}) \quad =\frac{1}{\alpha_{k-1}}g_k^Tg_k$$

$$p_{k-1}^TGp_{k-1}=\frac{1}{\alpha_{k-1}}(-g_{k-1}+\beta_{k-2}p_{k-2})^T(g_k-g_{k-1})=\frac{1}{\alpha_{k-1}}g_{k-1}^Tg_{k-1}$$

所以 $\beta_{k-1}=\dfrac{g_k^Tg_k}{g_{k-1}^Tg_{k-1}}$

PRP算法

FR算法中: $\beta_{k-1}=\dfrac{g_k^Tg_k}{g_{k-1}^Tg_{k-1}}$

(2)Polak-Ribiere-Polyak公式

由于$g_k^Tg_{k-1}=0$,所以有 $\beta_{k-1}=\dfrac{g_k^T(g_k-g_{k-1})}{g_{k-1}^Tg_{k-1}}$

对于二次函数,这两个函数是等价的,但对于一般的函数,根据这两个公式的出的算法的计算效果有差异.

注:对于这两个算法,可以证明$p_k^Tg_k=-g_k^Tg_k<0$,因而都是下降算法.

共扼梯度法算例

例3.4.1 用FR共扼梯度法求解($x_0=(0,0)^T$)

$$\min f(x)=\frac{3}{2}x_1^2+\frac{1}{2}x_2^2-x_1x_2-2x_1$$

解 $g(x)=(3x_1-x_2-2,x_2-x_1)^T$　　注:此处不需求G.

由$g_0=(-2,0)^T\neq 0$,故取$p_0=(2,0)^T$,从x_0出发,沿p_0作一维搜索,

即求$\min f(x_0+\alpha p_0)=6\alpha^2-4\alpha$的极小点,

得$\alpha_0=1/3$,于是$x_1=x_0+\alpha_0 p_0=(2/3,0)^T$,$g_1=(0,-2/3)^T$,

由FR公式得$\beta_0=g_1^Tg_1/g_0^Tg_0=1/9$

故$p_1=-g_1+\beta_0 p_0=(2/9,2/3)^T$.

从x_1出发,沿p_1作一维搜索,求

$$\min f(x_1+\alpha p_1)=\frac{4}{27}\alpha^2-\frac{4}{9}\alpha+\frac{2}{3}$$ 的极小点

解得$\alpha_1=3/2$,于是 $x_2=x_1+\alpha_1 p_1=(1,1)^T$.

此时 $g_2=(0,0)^T$,

故 $x^*=x_2=(1,1)^T$, $f^*=-1$

此算例中,$f(x)$为二元的正定二次函数,因此FR算法迭代两次得到最优点

共扼梯度法算例

例3.3.2 用FR方法与PRP方法求解

$$\min f(x)=100(x_2-x_1^2)^2+(1-x_1)^2$$

设初始点为$x_0=(0,0)^T$.

解: $g(x)=(-400(x_2-x_1^2)x_1-2(1-x_1),200(x_2-x_1^2))^T$

由$g_0=(-2,0)^T\neq 0$,故取$p_0=(2,0)^T$,从x_0出发,沿p_0作一维搜索,即求

$\min f(x_0+\alpha p_0)=1600\alpha^4+4\alpha^2-4\alpha+1$的极小点,

得$\alpha_0=0.080632$,(精确一维搜索方法求得,$\varepsilon=10^{-5}$,)

于是$x_1=x_0+\alpha_0 p_0=(0.161264,0)^T$,

$g_1=(0.000065,-5.201215)^T$.

共扼梯度法算例

$$\min f(x)=100(x_2-x_1^2)^2+(1-x_1)^2$$

$$g(x)=(-400(x_2-x_1^2)x_1-2(1-x_1),200(x_2-x_1^2))^T$$

$p_0=(2,0)^T$,$x_1=(0.161264,0)^T$,$g_1=(0.000065,-5.201215)^T$,

由FR公式得$\beta_0=g_1^Tg_1/g_0^Tg_0=6.763160$

故$p_1=-g_1+\beta_0 p_0=(13.526254,5.201215)^T$.

进一步可以以下的迭代,所得的结果(终止准则为$\|g_k\|<10^{-12}$,55步收敛)见下表.

最终得到$x^*\approx(1,1)^T$.

k	x_k	$f(x_k)$	$\|g(x_k)\|$
0	$(0,0)^T$	1	2
1	$(0.161264,0)^T$	0.771110	5.201215
2	$(0.292861,0.050603)^T$	0.623703	7.535261
10	$(1.006492,1.015405)^T$	6.07e-4	1.057204
20	$(1.000035,1.000074)^T$	3.02e-9	0.001843
30	$(1+1.31e-7,1+2.69e-7)^T$	2.21e-14	2.89e-6
40	$(1+0.51e-9,1+1.03e-9)^T$	2.79e-19	5.40e-9
50	$(1+2.10e-12,1+4.26e-12)^T$	4.74e-24	2.16e-11
54	$(1-1.14e-13,1-2.51e-13)^T$	6.14e-26	9.63e-12
55	$(1-1.42e-13,1-2.86e-13)^T$	2.06e-26	5.55e-13

从最后两组数据可以看出,虽然函数值下降,但是迭代点离最优点的距离却有所增加.

对于PRP算法,计算过程类似.

计算15步收敛,

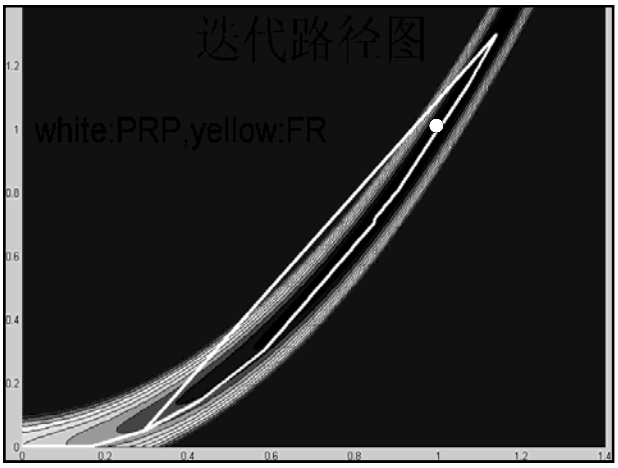
$x^*\approx(1,1)^T$

对于此例,PRP方法比FR方法收敛快.

计算结果见下表.

PRP方法计算结果

k	x_k	$f(x_k)$	$\|g(x_k)\|$
0	$(0,0)^T$	1	2
1	$(0.161264,0)^T$	0.771110	5.201215
2	$(0.292861,0.050603)^T$	0.623704	7.535350
3	$(1.139761,1.300789)^T$	0.019834	0.617648
10	$(1+6.95\text{e-}10,1+9.93\text{e-}10)^T$	1.63e-17	1.79e-7
11	$(1+1.36\text{e-}9,2.69\text{e-}9)^T$	1.90e-18	1.27e-8
12	$(1+2.13\text{e-}10,1+4.66\text{e-}10)^T$	1.99e-19	1.71e-8
13	$(1-0.69\text{e-}11,1-1.35\text{e-}11)^T$	6.31e-23	1.85e-10
14	$(1-2.43\text{e-}13,1-5.50\text{e-}13)^T$	4.60e-27	2.79e-12
15	$(1-0.93\text{e-}14,1-1.82\text{e-}14)^T$	1.07e-28	2.15e-13



重新开始的共扼梯度法

对于FR算法和PRP算法,如果初始方向不取负梯度方向,即使对于二次函数,也不能产生n个共扼方向.

因此,在用这两个方法时,如果迭代到距离最优点比较近,函数接近与一个二次函数时,我们重新取搜索方向为负梯度方向.

一般在实际应用中迭代n步或n+1步时重新设定搜索方向为负梯度方向.

重新开始的共扼梯度法

对于前面的例子,采用重新开始的共扼梯度法得到的收敛步数为:

FR算法: 迭代n步重新开始,49步收敛
迭代n+1步重新开始,31步收敛

PRP算法: 迭代n步重新开始,27步收敛
迭代n+1步重新开始,13步收敛

由于此处n比较小,改善并不明显.

拟Newton法的基本思想

最速下降法和阻尼Newton法可以统一成

$$x_{k+1}=x_k-\alpha_k H_k g_k$$

希望选取合适的H_k既能逐步逼近G_k^{-1},又不需要计算二阶导数.

H_k满足的条件:

C1. H_k对称正定,保证下降

C2. $H_{k+1}=H_k+E_k$,其中E_k的计算较为简单;

C3.拟Newton方程

拟Newton方程

为了希望H_0的逐步修正能逼近G_k^{-1}.

令$s_k=\alpha_k p_k=x_{k+1}-x_k, y_k=g_{k+1}-g_k$,

由Taylor公式

$$g_k\approx g_{k+1}=G_{k+1}(x_k-x_{k+1}).$$

当G_{k+1}非奇异时,有$G_{k+1}^{-1}y_k\approx s_k$.

对于二次函数上式精确成立.

为了达到逼近的效果,用H_{k+1}代替上式中的G_{k+1}^{-1},就得到拟Newton方程 $H_{k+1}y_k=s_k$.

秩1校正(P129(3.49))

考虑$H_{k+1}=H_k+E_k$,为了使E_k计算简单,我们考虑最简单的形式:秩1校正

$H_{k+1}=H_k+\alpha_k u_k u_k^T$

再根据拟Newton方程$H_{k+1}y_k=s_k$有

$$(H_k+\alpha_k u_k u_k^T)\, y_k=s_k$$

从而$\alpha_k (u_k^T y_k)\, u_k=s_k-H_k y_k$,

直接令$u_k=s_k-H_k y_k$,则

$$\alpha_k=\frac{1}{(s_k-H_ky_k)^T y_k},\quad H_{k+1}=H_k+\frac{(s_k-H_ky_k)(s_k-H_ky_k)^T}{(s_k-H_ky_k)^T y_k},$$

DFP算法

秩1校正公式不能保证H_k的正定性,从而不能保证得到的方向是下降的.

下面考虑秩2校正$H_{k+1}=H_k+\alpha_k u_k u_k^T+\beta_k v_k v_k^T$

根据拟Newton方程$H_{k+1}y_k=s_k$有

$$(H_k+\alpha_k u_k u_k^T+\beta_k v_k v_k^T)\, y_k=s_k$$

即$\alpha_k (u_k^T y_k)\, u_k+\beta_k(v_k^T y_k)\, v_k=s_k-H_k y_k$,

可取 $u_k=s_k$,$v_k=H_k y_k$,可得DFP修正公式

$$H_{k+1}=H_k-\frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}+\frac{s_k s_k^T}{y_k^T s_k}$$

DFP算法

Step1 给定控制误差ε,初始点x_0,初始矩阵$H_0=I$,计算g_0,令$k=0$.

Step2 令$p_k=-H_k g_k$.

Step3 由精确一维搜索确定步长

α_k $\quad f(x_k+\alpha_k p_k)=\min\limits_{\alpha>0} f(x_k+\alpha_k p_k)$

Step4 令$x_{k+1}=x_k+\alpha_k p_k$

Step5 若$\|g_{k+1}\|\leqslant\varepsilon$,则$x^*=x_{k+1}$,Stop.

否则令$s_k=x_{k+1}-x_k$,$y_k=g_{k+1}-g_k$

Step6 $\quad H_{k+1}=H_k-\frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}+\frac{s_k s_k^T}{y_k^T s_k}$

令$k=k+1$,转Step2.

算例(DFP算法)

例 3.5.1 用DFP算法求解($x_0=(1,1)^T$,$H_0=I_2$)

$$\min f(x)=x_1^2+2x_2^2-2x_1x_2-4x_1$$

解 $g(x)=(2x_1-2x_2-4,-2x_1+4x_2)^T$, $g_0=(-4,2)^T$, $p_0=(4,-2)^T$.

(i)求迭代点x_1,令$\varphi_0(\alpha)=f(x_0+\alpha p_0)=40\alpha^2-20\alpha-3$

得其极小点为$\alpha_0=1/4$,所以

$$x_1=x_0+\alpha_0 p_0=(2,0.5)^T,\ g_1=(-1,-2)^T,$$

$$s_0=x_1-x_0=(1,-0.5)^T,\ y_0=g_1-g_0=(3,-4)^T$$

由DFP修正公式有

$$H_1=H_0-\frac{H_0y_0y_0^TH_0}{y_0^T H y_0}+\frac{s_0s_0^T}{y_0^T s_0}=\frac{1}{100}\begin{pmatrix}84&38\\38&41\end{pmatrix}$$

下一搜索方向为$p_1=-H_1g_1=(1.6,1.2)^T$.

(ii)求迭代点x_2,令$\varphi_1(\alpha)=f(x_1+\alpha p_1)=1.6\alpha^2-4\alpha-5.5$,

其极小点为$\alpha_1=1.25$,于是

$$x_2=x_1+\alpha_1 p_1=(4,2),\ g_2=(0,0)^T.$$

所以$x^*=x_2=(4,2)^T$,$f^*=-8$.

由于$G(x)=\begin{pmatrix}2&-2\\-2&4\end{pmatrix}$为正定矩阵,$f(x)$为严格凸函数,所以$x^*$为整体极小点.

另外,可验证 $\quad H_2=G^{-1}=\begin{pmatrix}2&-2\\-2&4\end{pmatrix}^{-1}$

可以看出,对所给的二次正定函数,DFP算法迭代两次得到极小点,即DFP算法具有二次终止性.

算例(DFP算法)

例3.3.2 用DFP方法求解

$$\min f(x)=100(x_2-x_1^2)^2+(1-x_1)^2$$

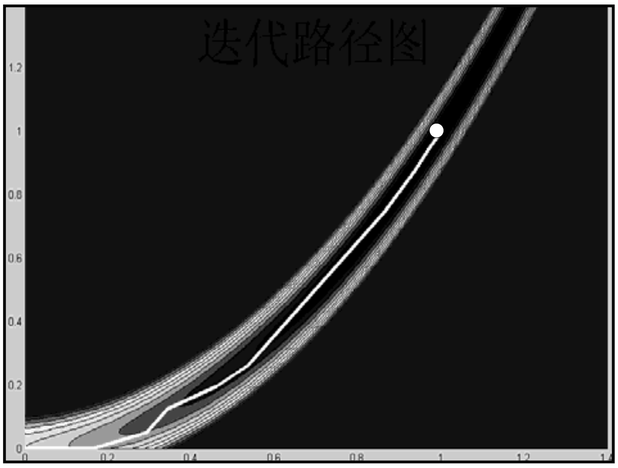
设初始点为$x_0=(0,0)^T$.

用DFP算法解决此问题,15步收敛(终止条件依然取为$\|g_k\|<10^{-12}$.)

具体计算结果见下表.

DFP方法计算结果

k	x_k	$f(x_k)$	$\|g(x_k)\|$
0	$(0,0)^T$	1	2
1	$(0.161264,0)^T$	0.771101	5.201215
2	$(0.292837,0.050595)^T$	0.623691	7.533679
3	$(0.344614,0.127077)^T$	0.436450	2.967550
10	$(0.990468,0.980310)^T$	0.000142	0.300874
11	$(1.000795,1.001389)^T$	4.67e-6	0.091333
12	$(0.999929,0.999854)^T$	6.10e-9	0.001279
13	$(1-4.13e-7,1-7.37e-7)^T$	9.82e-13	0.000041
14	$(1+3.46e-9,1+7.08e-9)^T$	1.44e-19	6.30e-9
15	$(1-1.67e-15,1-2.66e-15)^T$	4.71e-29	3.01e-13



DFP修正公式的正定继承性

假设H正定,那么由DFP修正公式得到的矩阵为

$$H_+ = H - \frac{Hyy^T H}{y^T Hy} + \frac{ss^T}{y^T s}(y \neq 0, s \neq 0)$$

如果H_+也正定,则

$$0 < y^T H_+ y = y^T Hy - \frac{y^T Hyy^T Hy}{y^T Hy} + \frac{y^T ss^T y}{y^T s} = y^T s \Longrightarrow y^T s > 0$$

反之,如果$y^Ts>0$,对任意的$x(\neq 0)\in R^n$,有

$$x^T H_+ x = \frac{x^T Hxy^T Hy - x^T Hyy^T Hx}{y^T Hy} + \frac{(s^T x)^2}{y^T s}$$

DFP修正公式的正定继承性

$$x^T H_+ x = \frac{x^T Hxy^T Hy - x^T Hyy^T Hx}{y^T Hy} + \frac{(s^T x)^2}{y^T s}$$

由于H正定,存在正定矩阵D,使得$H=DD^T$,令$u=Dx,v=Dy$,则有 $x^T H_+ x = \frac{u^T uv^T v - (u^T v)^2}{y^T Hy} + \frac{(s^T x)^2}{y^T s}$

根据Cauchy-Schwartz不等式,右端第一项非负,根据$y^Ts>0$,右端第二项也非负,因此x^TH_+x非负.

若第一项为0,则$u=\beta v(\beta\neq 0)$,此时$s^Tx=\beta s^Ty\neq 0$,因此第二项不为零,所以总有$x^TH_+x>0$.

所以,在$y^Ts>0$时,H_+正定.

于是得到引理3.5.1,H_+的充要条件为$y^Ts>0$.

DFP修正公式的正定继承性

对于DFP算法,要研究矩阵H_{k+1}是否正定,只需考察$y_k^Ts_k$是否为正.

$$y_k^T s_k = (g_{k+1} - g_k)^T (x_{k+1} - x_k) = \alpha_k (g_{k+1} - g_k)^T p_k$$
$$= \alpha_k g_{k+1}^T p_k - \alpha_k g_k^T p_k$$

在进行精确的一维搜索时,有 $g_{k+1}^T p_k = 0$.

因此 $y_k^T s_k = -\alpha_k g_k^T p_k = \alpha_k g_k^T H_k g_k$

在H_k正定时,$p_k=-H_kg_k$为下降方向,所以$\alpha_k>0$,从而$y_k^Ts_k>0$.于是由H_k正定可推出H_{k+1}正定.

因此,在DFP算法(精确的一维搜索)中,若H_0正定,则矩阵列$\{H_k\}$正定(定理3.5.2).

DFP算法的二次终止性

在前面的算例中,已经验证了DFP算法具有二次终止性.下面在理论上加以研究.

对正定二次函数 $f(x) = \frac{1}{2}x^T Gx + b^T x + c$

应用DFP算法,有

$s_0 = x_1 - x_0 = \alpha_0 p_0, p_1 = -H_1 g_1$

$y_0 = g_1 - g_0 = (Gx_1 + b) - (Gx_0 + b) = Gs_0$

$H_1 y_0 = s_0$

$$p_1^T G p_0 = (Gp_0)^T p_1 = -\frac{1}{\alpha_0}(Gs_0)^T H_1 g_1 = -\frac{1}{\alpha_0} y_0^T H_1 g_1 = -\frac{1}{\alpha_0} g_1^T s_0 = -g_1^T p_0 = 0$$

因此,p_1与p_0共扼.

DFP算法的二次终止性

为证明二次终止性,我们只需证明得到的方向是共扼方向.

若$0\leqslant i<j\leqslant l$时,$p_i^T G p_j=0$.对$0\leqslant i\leqslant l$

$$
\begin{aligned}
p_i^T G p_{l+1} &= (Gp_i)^T p_{l+1} \\
&= -\frac{1}{\alpha_i}(Gs_i)^T H_{l+1} g_{l+1} \\
&= -\frac{1}{\alpha_i} y_i^T H_{l+1} g_{l+1} \\
&= -\frac{1}{\alpha_i} g_{l+1}^T s_i \\
&= -g_{l+1}^T p_i
\end{aligned}
$$

$s_i = x_{i+1} - x_i = \alpha_i p_i,$
$p_{l+1} = -H_{l+1} g_{l+1}$

$y_i = g_{i+1} - g_i = Gs_i$

$H_{l+1} y_i = s_i\ ?$

南京邮电大学数理学院杨振华制作 njuptshumo2006@126.com

343

DFP算法的二次终止性

$$
\begin{aligned}
g_{l+1}^T p_i &= (g_{l+1} - g_l + \cdots + g_{i+2} - g_{i+1} + g_{i+1})^T p_i \\
&= (\sum_{j=i+1}^{l} y_j + g_{i+1})^T p_i = g_{i+1}^T p_i + \sum_{j=i+1}^{l} s_j^T G p_i
\end{aligned}
$$

$y_j = Gs_j$

$s_j = \alpha_j p_j,$
$p_j^T G p_i = 0$

$H_{l+1} y_i = s_i\ ?$ $i=l$时显然成立.$0\leqslant i\leqslant l-1$时

$$
\begin{aligned}
H_{l+1} y_i &= (H_l - \frac{H_l y_l y_l^T H_l}{y_l^T H_l y_l} + \frac{s_l s_l^T}{y_l^T s_l}) y_i \\
&= H_l y_i - \frac{H_l y_l y_l^T H_l y_i}{y_l^T H_l y_l} + \frac{s_l s_l^T y_i}{y_l^T s_l} \\
&= s_i - \frac{H_l y_l y_l^T s_i}{y_l^T H_l y_l} + \frac{s_l s_l^T y_i}{y_l^T s_l} \\
&= s_i - \frac{H_l y_l (s_l^T G s_i)}{y_l^T H_l y_l} + \frac{s_l s_l^T G s_i}{y_l^T s_l}
\end{aligned}
$$

归纳假设中添上$H_l y_i = s_i$

$y_l = Gs_l, y_i = Gs_i$

$s_l = \alpha_l p_l, s_i = \alpha_i p_i,$
$p_l^T G p_i = 0$

南京邮电大学数理学院杨振华制作 njuptshumo2006@126.com

344

DFP算法的二次终止性

根据上面的推导,我们得到

定理3.5.3 将DFP算法用于正定二次目标函数,设初始矩阵H_0正定,产生的迭代点互异,并设产生的搜索方向为$p_0,p_1,\cdots,p_k,\cdots$,则

(i)$p_i^T G p_j=0, 0\leqslant i<j\leqslant k$

(ii)$H_k y_i=s_i$, $0\leqslant i\leqslant k-1$.

DFP方法用于二次函数时产生的方向是共扼的,是一个共扼方向法,最多n步得到极小点.

南京邮电大学数理学院杨振华制作 njuptshumo2006@126.com

345

DFP算法的二次终止性

若DFP方法n步收敛,则n个共扼方向$p_0,p_1,\cdots,p_{n-1}$, 线性无关.故$s_0,s_1,\cdots,s_{n-1}$线性无关.

由于$H_n y_i=H_n G s_i=s_i (i=0,1,\cdots,n-1)$.

$H_n G[s_0,s_1,\cdots,s_{n-1}]=[s_0,s_1,\cdots,s_{n-1}]$.

矩阵$[s_0,s_1,\cdots,s_{n-1}]$非奇异,因此$H_n=G^{-1}$.

推论 在定理3.5.3的条件下,有

(i)DFP方法至多迭代n次就可得到极小点.

(ii)若$x_k\neq x^*(0\leqslant k\leqslant n-1)$,则$H_n=G^{-1}$.

南京邮电大学数理学院杨振华制作 njuptshumo2006@126.com

346

DFP算法的性质

(1)对于正定二次函数

(i)是共扼方向法(ii)迭代至多n次收敛,$H_n=G^{-1}$

(iii)满足比拟Newton方程更强的条件

$H_k y_i=s_i$, $0\leqslant i\leqslant k-1$.

(2)对于一般的函数

(i)矩阵H_k保持正定,是下降算法(ii)超线性收敛

(iii)对于凸函数整体收敛

(iv)每次迭代需要$3n^2+O(n)$次乘除运算(不包括一维搜索).

南京邮电大学数理学院杨振华制作 njuptshumo2006@126.com

347

DFP算法的搜索方向

下面研究精确一维搜索时,DFP方法中第k+1步的下降方向.

根据 $g_{k+1}^T p_k = 0$ 得

$$g_{k+1}^T s_k = g_{k+1}^T (x_{k+1} - x_k) = 0$$

$$
\begin{aligned}
H_{k+1}^{(DFP)} g_{k+1} &= H_k g_{k+1} - \frac{H_k y_k y_k^T H_k g_{k+1}}{y_k^T H_k y_k} + \frac{s_k s_k^T g_{k+1}}{y_k^T s_k} \\
&= H_k g_{k+1} - \frac{H_k y_k y_k^T H_k g_{k+1}}{y_k^T H_k y_k}
\end{aligned}
$$

$x_{k+1} = x_k + \alpha_k p_k$

$$H_k g_{k+1} = H_k (g_{k+1} - g_k) + H_k g_k = H_k y_k - p_k = H_k y_k - \alpha_k^{-1} s_k$$

$$
\begin{aligned}
H_{k+1}^{DFP} g_{k+1} &= H_k y_k - \alpha_k^{-1} s_k - \frac{H_k y_k y_k^T (H_k y_k - \alpha_k^{-1} s_k)}{y_k^T H_k y_k} \\
&= \frac{y_k^T s_k H_k y_k}{\alpha_k y_k^T H_k y_k} - \frac{s_k}{\alpha_k}
\end{aligned}
$$

南京邮电大学数理学院杨振华制作 njuptshumo2006@126.com

348

DFP算法的搜索方向

$$p_{k+1}=-H_{k+1}g_{k+1}=-\left(\frac{y_k^Ts_kH_ky_k}{\alpha_ky_k^TH_ky_k}-\frac{s_k}{\alpha_k}\right)=\frac{y_k^Ts_k}{\alpha_k}\left(\frac{s_k}{y_k^Ts_k}-\frac{H_ky_k}{y_k^TH_ky_k}\right)$$

记 $w_k=(y_k^TH_ky_k)^{1/2}\left(\frac{s_k}{y_k^Ts_k}-\frac{H_ky_k}{y_k^TH_ky_k}\right)$

显然p_{k+1}与w_k线性相关,

且 $w_k^Ty_k=(y_k^TH_ky_k)^{1/2}\left(\frac{s_k^Ty_k}{y_k^Ts_k}-\frac{y_k^TH_ky_k}{y_k^TH_ky_k}\right)=0$

Broyden族拟Newton法

$w_k=(y_k^TH_ky_k)^{1/2}\left(\frac{s_k}{y_k^Ts_k}-\frac{H_ky_k}{y_k^TH_ky_k}\right)$ $w_k^Ty_k=0$

在DFP公式中的H_{k+1}满足拟Newton方程$H_{k+1}y_k=s_k$

因此秩3校正公式

即 $H_{k+1}(\varphi)=H_k-\frac{H_ky_ky_k^TH_k}{y_k^TH_ky_k}+\frac{s_ks_k^T}{y_k^Ts_k}+\varphi_kw_kw_k^T$

也满足拟Newton方程,称为Broyden族修正公式.

其中φ_k可以与k有关,也可以是一个常数,从而可以给出无穷多的一族公式.

Broyden族拟Newton法

$$H_{k+1}(\varphi)=H_k-\frac{H_ky_ky_k^TH_k}{y_k^TH_ky_k}+\frac{s_ks_k^T}{y_k^Ts_k}+\varphi_kw_kw_k^T$$

取 $\varphi_k=\frac{y_k^Ts_k}{(s_k-H_ky_k)^Ty_k}$ 就得到秩1校正公式

$$H_{k+1}=H_k+\frac{(s_k-H_ky_k)(s_k-H_ky_k)^T}{(s_k-H_ky_k)^Ty_k},$$

取φ=1,得到的公式称为BFGS公式

$$H_{k+1}=H_k-\frac{H_ky_ky_k^TH_k}{y_k^THy_k}+\frac{s_ks_k^T}{y_k^Ts_k}+(y_k^TH_ky_k)\left(\frac{s_k}{y_k^Ts_k}-\frac{H_ky_k}{y_k^TH_ky_k}\right)^T\left(\frac{s_k}{y_k^Ts_k}-\frac{H_ky_k}{y_k^TH_ky_k}\right)$$

Broyden族拟Newton法

Broyden族公式可以看作DFP公式与BFGS公式的加权平均.

$$H_{k+1}(\varphi)=(1-\varphi)H_{k+1}^{DFP}+\varphi H_{k+1}^{BFGS}$$

对于相同的函数和相同的H_k,以及迭代点x_{k+1},下面研究DFP公式和BFGS公式所得搜索方向p_{k+1}有什么关系.

Broyden族拟Newton法

$$p_{k+1}^{BFGS}=-H_{k+1}^{BFGS}g_{k+1}=-(H_{k+1}^{DFP}+w_kw_k^T)g_k=p_{k+1}^{DFP}+(w_k^Tg_k)w_k$$

p_{k+1}^{DFP}与w_k线性相关,因此 p_{k+1}^{DFP} 与 p_{k+1}^{BFGS} 线性相关.

对任意的$x(\neq 0)\in R^n$,$x^TH_{k+1}^{BFGS}x=x^TH_{k+1}^{DFP}x+x^Tw_kw_k^Tx>0$

从而H_{k+1}^{BFGS}是正定矩阵,p_{k+1}^{BFGS}是下降方向.

因此,只要初始条件相同DFP方法和BFGS方法在精确一维搜索情况下得到的迭代点列是一致的.进一步,所有的Broyden族算法得到的点列是一致的(定理3.5.4).

Broyden族拟Newton法

在精确一维搜索的条件下,BFGS算法与DFP算法产生的点列是一致的.

在实际应用中,经常采用不精确的一维搜索.此时,两种算法产生的点列不再一致,BFGS算法要优于DFP算法.

对于Rosenbrock函数,采用Wolfe原则(μ=0.1,σ=0.5),用BFGS方法21步收敛($\|g_k\|<10^{-12}$.),而DFP方法30步收敛.

§ 3.6 Powell方向加速法

前面所介绍的各种方法都要用到目标函数的一阶或二阶导数.

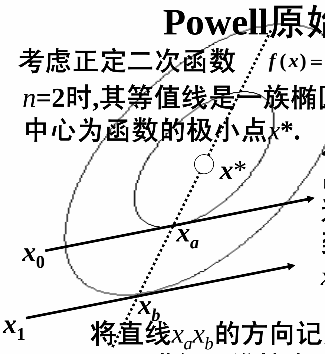
在实际应用中出现的例子有时导数难于计算,甚至不可导.

因此,有必要研究只利用函数值的方法,这一类方法称为直接法.

Powell原始算法

考虑正定二次函数 $f(x)=\frac{1}{2}x^TGx+b^Tx+c$

n=2时,其等值线是一族椭圆,这族椭圆的共同中心为函数的极小点x*.



从不同的初始点x_0,x_1出发,沿相同的方向p_0进行一维搜索,分别得到极小点x_a,x_b.连接x_a,x_b的直线必通过x*.

将直线x_ax_b的方向记为p_1,则从x_0出发,沿p_0, p_1进行一维搜索,可以得到极小点.

Powell原始算法

对于上面的二维的二次函数,从一点出发,沿两个方向进行一维搜索即可得到极小点,可以推测这两个方向是共扼的.

定理3.6.1 对于n维正定二次函数 $f(x)=\frac{1}{2}x^TGx+b^Tx+c$

设$p_0,p_1,\cdots,p_{k-1}$(k<n)关于G共扼,x_0与x_1为不同的任意两点,分别从x_0和x_1出发,依次沿$p_0,p_1,\cdots,p_{k-1}$作精确一维搜索,并设最后得到的极小点为x_a和x_b.如果$x_a \neq x_b$,则$x_b - x_a$与关于$p_0,p_1,\cdots,p_{k-1}$关于G共扼.即$(x_b - x_a)^TGp_i=0(i=0,1,\cdots,k-1)$.

Powell原始算法

证明 由于$p_0,p_1,\cdots,p_{k-1}$是关于G共扼的方向. x_a和x_b.是根据上述方向进行一维搜索得到的极小点,因此有

$g(x_a)^Tp_i=0(i=0,1,\cdots,k-1), g(x_b)^Tp_i=0(i=0,1,\cdots,k-1)$.

两式相减得

$(g(x_a)-g(x_b))^Tp_i=0$.

而$g(x_a)-g(x_b)=G(x_a-x_b)$,所以

$(x_a-x_b)^TGp_i=0(i=0,1,\cdots,k-1)$.

Powell原始算法

算法描述:给定初始点$x_0^{(0)}$,给定n个线性无关的向量$p_1,p_2,\cdots,p_n$(一般取为$e_1,e_2,\cdots,e_n$).

对$x_0^{(0)}$依次沿上面的n个方向进行一维搜索,最终得到的点为$x_1^{(0)}$,

以$x_1^{(0)}-x_0^{(0)}$为方向p_{n+1},去掉方向p_1.

$x_1^{(0)}$沿方向p_n进行搜索得到新的初始点$x_0^{(1)}$.

再由初始点$x_0^{(1)}$沿方向$p_2,p_3,\cdots,p_{n+1}$进行搜索得到点$x_1^{(1)}$,…

算例(Powell原始算法)

用Powell算法求解min $f(x)=(x_1+x_2)^2+(x_2-1)^2$,初始点$x_0=(2,1)^T$,初始搜索方向为$e_1,e_2$.

解 以$x_0^{(0)}$为初始点,$p_1=e_1$为搜索方向(搜索范围是整个实数集),得到点为$(0,1)^T$.再沿方向$p_2=e_2$进行搜索得$x_1^{(0)}=(0,0)^T$.

$p_3=x_1-x_0=(-2,-1)^T$.从$(0,0)^T$出发,沿p_3进行一维搜索得到$x_0^{(1)}=(4/13,2/13)^T$.

第一轮完成.

算例(Powell原始算法)

$x_0^{(1)}=(4/13,2/13)^T,p_2=(0,1)^T,p_3=(-2,-1)^T$.

从$x_0^{(1)}$出发,沿p_2进行一维搜索,得到点为$(4/13,-4/13)^T$.

再沿方向p_3进行搜索得$x_1^{(1)}=(88/169,-34/169)^T$.

$p_4=x_1^{(1)}-x_0^{(1)}=(36/169,-60/169)^T$.

从$x_1^{(1)}$出发,沿p_4进行一维搜索得到$x_0^{(2)}=(1,-1)^T=x^*$.

第二轮完成,得到目标函数的极小点.

Powell原始算法

在上面的算例中,可以发现$x_1^{(0)}$和$x_1^{(1)}$是不同的点沿同一个方向p_3搜索得到,

根据定理3.6.1, $p_4=x_1^{(1)}-x_0^{(1)}$和p_3关于G共扼.

$x_0^{(2)}$是由某个点沿p_3,p_4搜索而得,从而是最优点.

对于一般的n元正定二次函数,在一定条件下Powell算法至多经过n轮即可找到最优点.

Powell改进算法

Powell原始算法中,每次得到的n个向量可能线性无关或接近与线性相关,得不到极小点.

Powell对此作了改进,得到Powell改进算法.在此略去.

第四章

约束最优化方法

作业

P212 4.4 (ii),(iii)

P213 4.7 (ii)

P214 4.9 (ii) 4.11

§4.1 约束最优化问题的最优性条件

问题
$$\begin{aligned} &\min\ f(x), x\in R^n\\ &s.t.\quad c_i(x)=0, i\in E=\{1,2,\cdots,l\}\\ &\qquad\ \ c_i(x)\ge 0, i\in I=\{l+1,\cdots,m\}\end{aligned}$$

在求解问题之前,我们先讨论其最优解的必要条件,充分条件和充要条件.

这些条件是最优化理论的重要组成部分,对讨论算法起着关键的作用.

有的算法甚至可以直接用来求解问题.

4.1.1等式约束问题的最优性条件

问题　$\min\ f(x), x \in R^n$

$s.t.\quad c_i(x) = 0, i \in E = \{1, 2, \cdots, l\}$

考虑n=2,l=1的情况.$c_1(x)$=0表示二维平面的一条曲线.最优点满足约束,必落在这一曲线上.

在最优点处作曲线的切线.

考虑$f(x)$在最优点处的负梯度方向

$$-g^* = -\nabla f(x^*)$$

等式约束问题的最优性条件

若$-g^*$与上述切线不垂直,则可以在曲线上移动充分小的距离,使 f 的函数值下降.

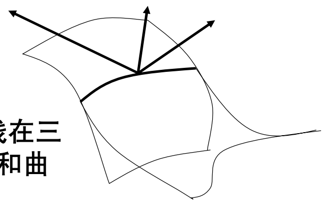
这与”最优点”矛盾.因此梯度方向与切线垂直.

或,$f(x)$在最优点处的梯度方向就是$c_1(x)$=0在该点处的法向.

而$c_1(x)$=0在该点处的法线方向为 $\nabla c_1(x^*)$

因此,存在数λ_1,使得 $\nabla f(x^*) - \lambda_1 \nabla c_1(x^*) = 0$

等式约束问题的最优性条件



如果n=3,l=2,约束曲线在三维空间中曲面$c_1(x)$=0和曲面$c_2(x)$=0的交线.

同样可以说明(-)g^*与曲线的切线垂直.

因此,曲面在x^*处的法向量 $\nabla c_1(x^*), \nabla c_2(x^*)$ 与梯度向量g^*共面.

存在数λ_1, λ_2,使得 $\nabla f(x^*) - \lambda_1 \nabla c_1(x^*) - \lambda_2 \nabla c_2(x^*) = 0$

等式约束问题的一阶必要条件

$\min\ f(x), x \in R^n$

$s.t.\quad c_i(x) = 0, i \in E = \{1, 2, \cdots, l\}$

定理4.1.1(一阶必要条件)

若(i)x^*是上述问题的局部最优解;

(ii)$f(x)$与$c_i(x)$(i=1,2,···,l)在x^*的某邻域内连续可微;

(iii) $\nabla c_i(x^*)(i = 1, 2 \cdots, l)$ 线性无关

则存在一组不全为零的数 $\lambda_1^*, \lambda_2^*, \cdots, \lambda_l^*$ 使得

$$\nabla f(x^*) - \sum_{i=1}^{l} \lambda_i^* \nabla c_i(x^*) = 0.$$

等式约束问题的一阶必要条件

$\min\ f(x), x \in R^n$

$s.t.\quad c_i(x) = 0, i \in E = \{1, 2, \cdots, l\}$

对于上述问题,引入$n+l$元的Lagrange函数

$$L(x, \lambda) = f(x) - \lambda^T c(x) = f(x) - \sum_{i=1}^{l} \lambda_i c_i(x)$$

其中$c(x)=(c_1(x),\cdots,c_l(x))^T$,$\lambda=(\lambda_1,\cdots,\lambda_l)^T$.

称λ为Lagrange乘子向量.

Lagrange函数的梯度为

$$\nabla L(x, \lambda) = \begin{pmatrix} \nabla_x L \\ \nabla_\lambda L \end{pmatrix} = \begin{pmatrix} \nabla f(x) - \sum_{i=1}^{l} \lambda_i \nabla c_i(x) \\ -(c_1(x), \cdots, c_l(x))^T \end{pmatrix}$$

等式约束问题的一阶必要条件

$$\nabla L(x, \lambda) = \begin{pmatrix} \nabla_x L \\ \nabla_\lambda L \end{pmatrix} = \begin{pmatrix} \nabla f(x) - \sum_{i=1}^{l} \lambda_i \nabla c_i(x) \\ -(c_1(x), \cdots, c_l(x))^T \end{pmatrix}$$

因此无约束问题min $L(x,\lambda)$的最优性条件

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

恰好是原来问题的一阶必要条件及$c_i(x^*)$,i=1,···,l.

所以求含$n+l$个未知数$x_1,\cdots,x_n,\lambda_1,\cdots,\lambda_l$的非线性方程组的解$(x^*,\lambda^*)$,其中$x^*=(x_1^*,\cdots,x_n^*)^T$在一定条件下就是原来约束问题的最优解.

点(x^*,λ^*)称为Lagrange函数$L(x,\lambda)$的驻点.

等式约束问题的二阶充分条件

$$\min\ f(x), x\in R^n$$
$$s.t.\quad c_i(x)=0, i\in E=\{1,2,\cdots,l\}$$

定理4.1.2 在上面的等式约束问题中,若

(i)$f(x)$与$c_i(x)(1\leqslant i\leqslant l)$是二阶连续可微函数

(ii)存在$x^*\in R^n$与$\lambda^*\in R^l$使得Lagrange函数的梯度为零,即 $\nabla L(x^*,\lambda^*)=0$

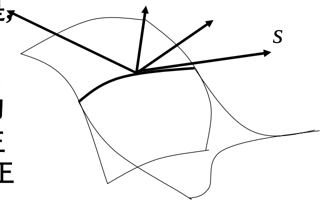
(iii)对于任意非零向量$s\in R^n$,且

$s^T\nabla c_i(x^*)=0,(i=1,2,\cdots,l)$ 均有 $s^T\nabla_x^2 L(x^*,\lambda^*)s>0$

则x^*是上面问题的严格局部极小点.

等式约束问题的二阶充分条件

定理4.1.2的几何意义是,在Lagrange函数$L(x,\lambda)$的驻点处,若$L(x,\lambda)$函数关于x的Hesse矩阵在约束超曲面的切平面上正定(不要求在整个空间正定),则x^*就是严格局部极小点.



4.1.1不等式约束问题的最优性条件

$$\min\ f(x), x\in R^n$$
$$s.t.\quad c_i(x)\ge 0, i\in\{1,2,\cdots,m\}$$

定义4.1.1 若上述问题的一个可行点$\tilde{x}$使得某个不等式约束$c_j(x)\geqslant 0$中的等号成立,即$c_j(\tilde{x})=0$,则该不等式约束$c_j(x)\geqslant 0$称为关于 的有效约束.否则,若对某个k,使得$c_k(\tilde{x})>0$,则该不等式约束$c_k(x)\geqslant 0$称为关于$\tilde{x}$的非有效约束.

称所有在$\tilde{x}$处的有效约束的指标组成的集合.

$\tilde{I}=I(\tilde{x})=\{i\mid c_i(\tilde{x})=0\}$为$\tilde{x}$处的有效(约束)集

注:有时我们也将等式约束也视为有效约束.

在教材中有说法不一致的地方.

Fritz-John一阶必要条件

$$\min\ f(x), x\in R^n$$
$$s.t.\quad c_i(x)\ge 0, i\in\{1,2,\cdots,m\}$$

定理4.1.6 设x^*为上述问题的局部最优解且$f(x),c_i(x)(1\leqslant i\leqslant m)$在$x^*$点可微,则存在非零向量$\lambda^*=(\lambda_0^*,\lambda_1^*,\cdots,\lambda_m^*)$使得

$$\lambda_0^*\nabla f(x^*)-\sum_{i=1}^m\lambda_i^*\nabla c_i(x^*)=0,$$
$$\lambda_i^*c_i(x^*)=0, i=1,\cdots,m$$
$$\lambda_i^*\ge 0, i=0,1,\cdots,m$$

满足上面的条件的点称为Fritz-John点.

上面的条件仅仅是必要条件.

Fritz-John一阶必要条件

$$\lambda_0^*\nabla f(x^*)-\sum_{i=1}^m\lambda_i^*\nabla c_i(x^*)=0,\quad \begin{array}{l}\lambda_i^*c_i(x^*)=0, i=0,1,\cdots,m\\ \lambda_i^*\ge 0, i=0,1,\cdots,m\end{array}$$

证明概要 设x^*处的有效集为$I^*=I(x^*)=\{i|c_i(x^*)=0,i=1,2,\cdots,m\}$.

对于无效约束,由于$c_i(x)>0$,若定理的结论成立,显然有$\lambda_i^*=0$.

定理结论可以描述为存在λ_0及$\lambda_i(i\in I^*)$,使得

$$\lambda_0^*\nabla f(x^*)-\sum_{i\in I^*}\lambda_i^*\nabla c_i(x^*)=0,\ \lambda_i^*\ge 0, i=\{0\}\cup I^*.$$

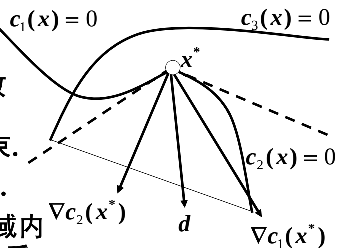
因为x^*是局部最优解,在"指向有效约束的内部的方向中"不含$f(x)$的下降方向.

因为x^*是局部最优解,在"指向有效约束的内部的方向中"不含$f(x)$的下降方向.

如图显示的是三个约束的例子

其中$c_3(x)\geqslant 0$为无效约束, $c_1(x)\geqslant 0$, $c_2(x)\geqslant 0$为有效约束.

黑色部分为可行域.



由最优点指向可行域内部的方向d都具有性质

$d^T\nabla c_i(x^*)>0, i\in I^*$

这种方向都不是下降方向,因此

$\nabla f(x^*)^T d\ge 0$

即由 $d^T\nabla c_i(x^*)>0, i\in I^*$ 可以推出 $\nabla f(x^*)^T d\geq 0$
因此有下面的引理

引理4.1.5 在不等式约束问题中,假设

(i)x^*为问题的局部最优解,且
$I^*=\{i|c_i(x^*)=0, i=1,2,\cdots,m\}$;

(ii)$f(x)$和$c_i(x)(i\in I^*)$在x^*可微;

(iii)$c_i(x)(i\in I\setminus I^*)$在$x^*$连续;则$G\cap S=\phi$.

其中 $S=\{d\in R^n \mid \nabla f(x^*)^T d<0\}$ 表示下降方向

$G=\{d\in R^n \mid \nabla c_i(x^*)^T d>0, i\in I^*\}$

表示指向可行域内部的方向

Fritz-John一阶必要条件

证明概要(续)根据上述引理,不存在$d\in R^n$,使得

$$\nabla f(x^*)^T d<0,\quad \nabla c_i(x^*)^T d>0, i\in I^*$$

即 $$\nabla f(x^*)^T d<0,\quad -\nabla c_i(x^*)^T d<0, i\in I^*$$

$\nabla f(x^*), -\nabla c_i(x^*)(i\in I^*)$是这样一组向量,它们不在过原点的任何超平面的同一侧.

于是我们总可以适当放大或缩小各向量的长度,使得变化后的各向量的合成向量为零向量.

注:这一结论的依据是下面的Gordan引理.

Gordan引理

引理4.1.4 设$a_1,\cdots,a_r$是n维向量,则不存在向量$d\in R^n$使得

$$a_i^T d<0(i=1,\cdots,r)$$

成立的充要条件是,存在不全为零的非负实数组$\lambda_1,\cdots,\lambda_r$,使

$$\sum_{i=1}^r \lambda_i a_i=0.$$

Fritz-John一阶必要条件

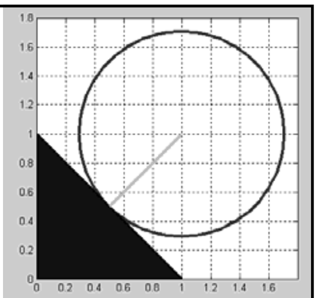
证明概要(续)根据Gordan引理,存在不全为零的数$\lambda_0^*\geqslant 0$, $\lambda_i^*\geqslant 0(i\in I^*)$,使得

$$\lambda_0^*\nabla f(x^*)-\sum_{i\in I^*}\lambda_i^*\nabla c_i(x^*)=0,$$

对于$i\in I\setminus I^*$,只要令$\lambda_i^*=0$,即可得到Fritz-John条件.

$$\lambda_0^*\nabla f(x^*)-\sum_{i=1}^m\lambda_i^*\nabla c_i(x^*)=0,$$
$$\lambda_i^* c_i(x^*)=0, i=0,1,\cdots,m$$
$$\lambda_i^*\geq 0, i=0,1,\cdots,m$$

例题 (Fritz-John条件)



例4.1.1

min $f(x)=(x_1-1)^2+(x_2-1)^2$

s.t. $c_1(x_1,x_2)=(1-x_1-x_2)^3\geqslant 0$

$c_2(x)=x_1\geqslant 0$

$c_3(x)=x_2\geqslant 0$

解:本问题是求点$(1,1)^T$到如图三角形区域的最短距离.显然唯一最优解为$x^*=(1/2,1/2)^T$.

例题(Fritz-John条件)

min $f(x)=(x_1-1)^2+(x_2-1)^2$

s.t. $c_1(x_1,x_2)=(1-x_1-x_2)^3\geqslant 0$

$c_2(x)=x_1\geqslant 0$

$c_3(x)=x_2\geqslant 0$

下面求满足Fritz-John条件的点.

$$\lambda_0^*\nabla f(x^*)-\sum_{i=1}^m\lambda_i^*\nabla c_i(x^*)=0,$$
$$\lambda_i^* c_i(x^*)=0, i=1,\cdots,m$$
$$\lambda_i^*\geq 0, i=0,1,\cdots,m$$

即

$$\lambda_0^*\begin{pmatrix}2x_1-2\\2x_2-2\end{pmatrix}-\lambda_1^*\begin{pmatrix}-3(1-x_1-x_2)^2\\-3(1-x_1-x_2)^2\end{pmatrix}-\lambda_2^*\begin{pmatrix}1\\0\end{pmatrix}-\lambda_3^*\begin{pmatrix}0\\1\end{pmatrix}=0,$$

$$\lambda_1^*(1-x_1-x_2)^3=0,\quad \lambda_2^* x_1=0,\quad \lambda_3^* x_2=0.\quad \lambda_i^*\geq 0, i=0,1,2,3.$$

例题

$$\lambda_0^*\begin{pmatrix}2x_1-2\\2x_2-2\end{pmatrix}-\lambda_1^*\begin{pmatrix}-3(1-x_1-x_2)^2\\-3(1-x_1-x_2)^2\end{pmatrix}-\lambda_2^*\begin{pmatrix}1\\0\end{pmatrix}-\lambda_3^*\begin{pmatrix}0\\1\end{pmatrix}=0,$$

$$\lambda_1^*(1-x_1-x_2)^3=0,\ \ \lambda_2^*x_1=0,\ \ \lambda_3^*x_2=0.\ \ \lambda_i^*\ge 0,i=0,1,2,3.$$

若λ_1*= 0, λ_2*= 0, λ_3*= 0,则λ_0*>0,$x=(1,1)^T$.
该点虽然满足Fritz-John条件,但不是可行点.

若λ_0*= 0, λ_2*= 0, λ_3*= 0,则λ_1*>0,$x=(u,1-u)^T$.
当0≤u≤1时,x为可行点.
三角形斜边上的点都是可行的Fritz-John点,
但只有$(1/2,1/2)^T$是最优点.

若λ_0*= 0, λ_1*= 0,则根据最上面的等式可以推出λ_2*= 0, λ_3*= 0.不满足Fritz-John条件.

Fritz-John条件的缺点

对于上面的例子,可行的Fritz-John点对应的λ_0*= 0.
这表明在这些点处的有效约束函数的梯度是线性相关的.
但是目标函数的信息并未出现在其中,这样对判断是否为最优解就不起作用了.
因为只要约束函数不变,改变目标函数,Fritz-John条件始终是成立的.
Kuhn-Tucker条件针对这一缺点作了改进.

Kuhn-Tucker一阶必要条件

$$\min\ f(x), x\in R^n$$
$$s.t.\quad c_i(x)\ge 0, i\in\{1,2,\cdots,m\}$$

定理4.1.6 设
(i)x*为上述问题的局部最优解, 有效集
$I^*=\{i|c_i(x^*)=0,i=1,2,\cdots,m\}$;
(ii)$f(x),c_i(x)(1\leqslant i\leqslant m)$在$x$*点可微;
(iii)对于$i\in I^*$的 线性无关,
则存在向量$\lambda^*=(\lambda_1^*,\cdots,\lambda_m^*)$使得

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla c_i(x^*)=0,$$
$$\lambda_i^*c_i(x^*)=0, i=1,\cdots,m$$
$$\lambda_i^*\ge 0, i=1,\cdots,m$$

满足左边的条件的点称为KT点.

$m+n$维向量$\begin{pmatrix}x^*\\\lambda^*\end{pmatrix}$称为KT对.

向量λ*称为Lagrange乘子向量.

Kuhn-Tucker必要条件

显然,由Fritz-John条件立刻可以推出Kuhn-Tucker条件.
另一证明思路:
x*处不存在可行下降方向,即若$d(\neq 0)\in R^n$,且

$$\nabla c_i(x^*)^T d\ge 0(i\in I^*),$$

有

$$\nabla f(x^*)^T d\ge 0.$$

注:此处可行方向的条件比Fritz-John条件中的证明中的条件多了等号,在此不详细讨论其中的区别.

Kuhn-Tucker必要条件

$$\nabla c_i(x^*)^T d\ge 0(i\in I^*)\Longrightarrow\nabla f(x^*)^T d\ge 0.$$

借助于Farkas引理,可推出存在$\lambda_i^*\geqslant 0(i\in I^*)$,使得

$$\nabla f(x^*)-\sum_{i\in I^*}\lambda_i^*\nabla c_i(x^*)=0.$$

类似与Fritz-John条件的证明,可以证明Kuhn-Tucker条件.

有效约束函数的梯度线性无关称为Kuhn-Tucker约束规范.

如果该约束规范不满足,最优点不一定是KT点.

Farkas引理

引理 4.1.3 设$a_1,a_2,\cdots,a_r$和b均为n维向量,则所有满足

$$a_i^Td\geqslant 0, i=1,\cdots,r$$

的向量$d\in R^n$,同时也满足不等式$b^Td\geqslant 0$的充要条件是,存在非负实数$\lambda_1,\lambda_2,\cdots,\lambda_r$使得

$$b=\sum_{i=1}^{r}\lambda_i a_i$$

注: 此引理的充分性证明是显然的.
必要性的证明较为复杂, 此处略去.

Farkas引理的几何说明

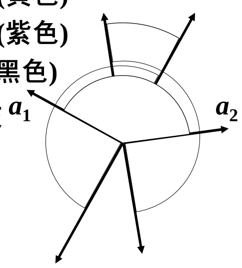
考虑二维空间中两个向量a_1,a_2.

满足$a_1^T d\geqslant 0$的方向d的范围(黄色)

满足$a_2^T d\geqslant 0$的方向d的范围(紫色)

同时满足上面两个条件的d(黑色)

要与上述黑色区域的方向交角为锐角的方向应在a_1与a_2两个方向之间(红色).



从而可以表示为这两个向量的非负的线性组合.

Kuhn-Tucker必要条件

$$\min\ f(x), x\in R^n$$
$$s.t.\quad c_i(x)\ge 0, i\in\{1,2,\cdots,m\}$$

$$\nabla f(x^*)-\sum_{i=1}^m \lambda_i^*\nabla c_i(x^*)=0,$$
$$\lambda_i^* c_i(x^*)=0, i=1,\cdots,m$$
$$\lambda_i^*\ge 0, i=1,\cdots,m$$

当所有的有效约束的乘子都不为零时,称互补松弛条件为严格互补松弛条件.

KT条件中$\lambda_i^* c_i(x^*)=0$称为互补松弛条件.

它表明λ_i^*与$c_i(x^*)$不能同时不为0.

线性规划

对于线性规划问题

$\min f(y)=-b^T y$

s.t. $-A^T y\geqslant -c$

其中$y\in R^m$,$A\in R^{m\times n}$,$b\in R^m$,$c\in R^n$

问题有n个约束条件.

各个约束条件关于y的梯度为$-A^T$的行向量$(-p_i)$.

若A^T的各个行向量线性无关.根据Kuhn-Tucker条件,在该线性规划的最优点y^*处存在乘子向量$x^*\geqslant 0$,使得

$$-b-\sum_{i=1}^n x_i^*(-p_i)=0$$
$$x_i^*(-p_i^T y_i^*+c_i)=0$$

即$Ax^*=b$

对偶规划约束条件

及$(A^T y^*-c)^T x^*=0$

线性规划互补松弛条件

4.1.3 一般约束问题的最优性条件

$$\min\ f(x), x\in R^n$$
$$s.t.\quad c_i(x)=0, i\in E=\{1,2,\cdots,l\}$$
$$c_i(x)\ge 0, i\in I=\{l+1,\cdots,m\}$$

定理4.1.8 在上述问题中,若

(i)x^*为局部最优解, 有效集$I^*=\{i|c_i(x^*)=0,i\in I\}$;

(ii)$f(x)$,$c_i(x)(1\leqslant i\leqslant m)$在$x^*$点可微;

(iii)对于$i\in E\cup I^*$,$\nabla c_i(x^*)$线性无关,

则存在向量$\lambda^*=(\lambda_1^*,\cdots,\lambda_m^*)$使得

$$\nabla f(x^*)-\sum_{i=1}^m \lambda_i^*\nabla c_i(x^*)=0,\quad \lambda_i^* c_i(x^*)=0, i\in I,\quad \lambda_i^*\ge 0, i\in I.$$

一般约束问题的最优性条件

$m+n$维函数 $L(x,\lambda)=f(x)-\sum_{i=1}^m \lambda_i c_i(x^*)$

称为前述问题的Lagrange函数.

KT条件的第一式 $\nabla f(x^*)-\sum_{i=1}^m \lambda_i^*\nabla c_i(x^*)=0,$

可以写为 $\nabla_x L(x^*,\lambda^*)=0.$

其中λ^*称为Lagrange乘子向量.

矩阵 $\nabla_x^2 L(x^*,\lambda^*)=\nabla^2 f(x^*)-\sum_{i=1}^m \lambda_i^*\nabla^2 c_i(x^*)$

称为Lagrange函数在$\begin{pmatrix}x^*\\ \lambda^*\end{pmatrix}$处的Hesse矩阵,记为

$$w^*=\nabla_x^2 L(x^*,\lambda^*).$$

二阶充分条件

$$\min\ f(x), x\in R^n$$
$$s.t.\quad c_i(x)=0, i\in E=\{1,2,\cdots,l\}$$
$$c_i(x)\ge 0, i\in I=\{l+1,\cdots,m\}$$

定理4.1.9 设$f(x)$和$c_i(x)(i\in E\cup I)$是二阶连续可微函数,若存在$x^*\in R^n$满足

(i)$\begin{pmatrix}x^*\\ \lambda^*\end{pmatrix}$为KT对,且严格互补松弛条件成立;

(ii)对子空间 $M=\{d\in R^n\mid d^T\nabla c_i(x^*)=0, i\in I^*\}$ 中的任意$d\neq 0$,有$d^T w^* d>0$,

则x^*为上述问题的严格局部最优解.

凸规划问题的充分条件

凸规划问题

$$\min\ f(x), x\in R^n$$
$$s.t.\quad c_i(x)=a_i^T x+b_i=0, i\in E=\{1,\cdots,l\}$$
$$c_i(x)\ge 0, i\in I=\{l+1,\cdots,m\}$$

其中$f(x)$为凸函数,$c_i(x)(i\in I)$为凹函数.

定理4.1.10 设上面的凸规划问题中$f(x)$和$c_i(x)(i\in I)$为可微函数,若x^*为该问题的KT点,则x^*为其整体最优解.

例题(KT点)

例4.1.3 已知约束问题

$$\min\ f(x)=-3x_1^2-x_2^2-2x_3^2$$
$$s.t.\quad c_1(x)=x_1^2+x_2^2+x_3^2-3=0$$
$$c_2(x)=-x_1+x_2\ \ge 0$$
$$c_3(x)=x_1\ \ge 0$$
$$c_4(x)=x_2\ \ge 0$$
$$c_5(x)=x_3\ \ge 0$$

试验证最优点 $x^*=(1,1,1)^T$为KT点.

解 $I^*=\{1,2\}$

$\nabla f(x^*)=(-6,-2,-4)^T$,

$\nabla c_1(x^*)=(2,2,2)^T$,

$\nabla c_2(x^*)=(-1,1,0)^T$,

令 $\begin{pmatrix}-6\\-2\\-4\end{pmatrix}-\lambda_1\begin{pmatrix}2\\2\\2\end{pmatrix}-\lambda_2\begin{pmatrix}-1\\1\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}$

得λ_1= -2,λ_2=2,即

$\nabla f(x^*)=-2\nabla c_1(x^*)+2\nabla c_2(x^*)$ $\lambda_2 c_2(x^*)=0, \lambda_2>0$,

对于无效约束,可取λ_i^*=0, 故$x^*=(1,1,1)^T$为KT点.

§4.2 罚函数法与乘子法

对于一般的约束优化问题

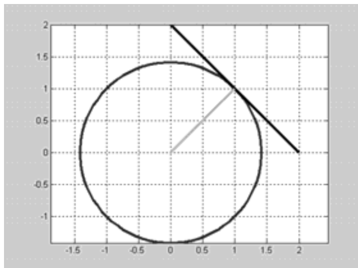
$$\min\ f(x), x\in R^n$$
$$s.t.\quad c_i(x)=0, i\in E=\{1,\cdots,l\}$$
$$c_i(x)\ge 0, i\in I=\{l+1,\cdots,m\}$$

一种常用的方法是将其转化为无约束问题.
其中的约束转化为目标函数的一部分.
满足约束时,对应的函数值很小;
不满足约束时,对应的函数值很大——"惩罚".
在求解过程中,为使得总的目标函数值最小,得到的解一般会满足约束.
这类方法一般称为罚函数法.

外罚函数法

例4.2.1 求解约束问题 $\min\ f(x_1,x_2)=x_1^2+x_2^2$, $s.t.\ x_1+x_2-2=0$

解: 本问题就是求原点到直线x_1+x_2-2=0上的点的最短距离.显然最优解为$(1,1)^T$.



外罚函数法

设辅助函数为 $F(x_1,x_2)=\begin{cases}x_1^2+x_2^2, & x_1+x_2=2\\ +\infty, & x_1+x_2\ne 2\end{cases}$

则以$F(x_1,x_2)$为目标函数的无约束问题的极小点必在直线x_1+x_2-2=0上,且目标函数的取值与原来问题的目标函数的取值一致.

只是该函数的性质比较糟糕,无法用一般的算法求解.

外罚函数法

考虑下面的增广目标函数

$$P(x_1,x_2,\sigma)=x_1^2+x_2^2+\sigma(x_1+x_2-2)^2$$

其中σ是很大的正数.
以其为目标函数,解无约束问题,得到最优解为

$$x_1(\sigma)=x_2(\sigma)=\frac{2\sigma}{2\sigma+1}$$

当$\sigma\to+\infty$时,有$(x_1(\sigma),x_2(\sigma))^T\to(1,1)^T=x^*$.
即无约束问题最优解的极限为原问题最优解.

注:辅助函数的目标函数值为$\frac{4\sigma}{2\sigma+1}$

在$\sigma\to+\infty$时也逼近于原问题的最优值 2.

外罚函数法(等式约束问题)

$$\min\ f(x), x \in R^n$$
$$s.t.\quad c_i(x) = 0, i \in E = \{1,2,\cdots,l\}$$

对于等式约束问题,构造如下增广目标函数

$$P(x,\sigma) = f(x) + \sigma\sum_{i=1}^{l}|c_i(x)|^{\beta}, \beta \ge 1$$

其中σ>0为参数,称为罚因子.

对于惩罚项 $\tilde{P}(x) = \sum_{i=1}^{l}|c_i(x)|^{\beta}$,

当x为可行解时, $c_i(x) = 0, \tilde{P}(x) = 0, P(x,\sigma) = f(x)$,

当x不是可行解时, $c_i(x) \ne 0, \tilde{P}(x) > 0,$
$P(x,\sigma) = f(x) + \sigma\tilde{P}(x)$,

σ越大,惩罚越重.

外罚函数法(等式约束问题)

$$P(x,\sigma) = f(x) + \sigma\tilde{P}(x),$$

当σ充分大时,要使$P(x,\sigma)$取极小值, 应充分小,即$P(x,\sigma)$的极小点充分逼近可行域.
在一定条件下,当$\sigma\to\infty$时, $P(x,\sigma)$的极小点逼近于原问题的解.

外罚函数法(不等式约束问题)

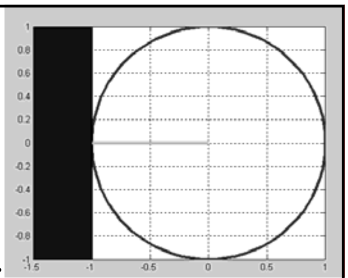
$$\min\ f(x), x \in R^n$$
$$s.t.\quad c_i(x) \ge 0, i \in I = \{1,2,\cdots,m\}$$

可以构造增广目标函数 $P(x,\sigma) = f(x) + \sigma\tilde{P}(x)$,

其中
$$\tilde{P}(x) = \begin{cases} 0 & c_i(x) \ge 0 \\ \sum_{i=1}^{m}|c_i(x)|^{\alpha}, \alpha \ge 1, & c_i(x) < 0 \end{cases}$$
$$= \sum_{i=1}^{m}|\min(0, c_i(x))|^{\alpha} \quad = \sum_{i=1}^{m}\left(\frac{|c_i(x)| - c_i(x)}{2}\right)^{\alpha}$$

“惩罚项”的作用与等式约束时的情形类似.

例题(不等式约束问题)



例4.2.1求解约束问题

$$\min\ f(x) = x_1^2 + x_2^2$$
$$s.t.\ x_1 + 1 \le 0$$

问题是求点$(0,0)^T$到半平面x≤-1的最短距离.

显然最优点为$x^*=(-1,0)^T$.最优值为$f(x^*)=1$.

设增广目标函数为

$$P(x,\sigma) = x_1^2 + x_2^2 + \sigma[\min(0, -x_1 - 1)]^2$$

例题(不等式约束问题)

$$P(x,\sigma) = x_1^2 + x_2^2 + \sigma[\min(0, -x_1 - 1)]^2$$
$$= \begin{cases} x_1^2 + x_2^2, & x_1 + 1 \le 0 \\ x_1^2 + x_2^2 + \sigma(x_1+1)^2, & x_1 + 1 > 0 \end{cases}$$

故 $\dfrac{\partial P}{\partial x_1} = \begin{cases} 2x_1, & x_1 < -1 \\ 2x_1 + 2\sigma(x_1+1), & x_1 > -1 \end{cases}$ $\quad \dfrac{\partial P}{\partial x_2} = 2x_2$

令 $\dfrac{\partial P}{\partial x_1} = \dfrac{\partial P}{\partial x_2} = 0$ 得 $x_1(\sigma) = -\dfrac{\sigma}{\sigma+1}, x_2(\sigma) = 0.$

它是min $P(x,\sigma)$的最优解,最优值为

$$P(x,\sigma) = \left(-\frac{\sigma}{\sigma+1}\right)^2 + \sigma\left(\frac{1}{\sigma+1}\right)^2 = \frac{\sigma}{\sigma+1}.$$

当$\sigma\to+\infty$时, $x_1(\sigma)\to-1$, $x_2(\sigma)\to0$.
因此$x(\sigma)\to x^*$,$P(x,\sigma)\to f(x^*)=1$.

外罚函数法

在上面的两个例子中,当$\sigma\to+\infty$时,$P(x,\sigma)$的最优解$x(\sigma)$趋向于极限x^*.而x^*即为原约束问题的最优解.

不过$x(\sigma)$往往不满足约束条件,在迭代过程中, $x(\sigma)$从可行域的外部逐步趋于原问题的最优解.

因此上面的方法称为外罚函数法.

通过一系列无约束最优化问题来求解约束最优化问题称为序列无约束极小化方法SUMT (Sequential Unconstrained Minimization Technique),因此外罚函数法又称为SUMT外点法.

外罚函数法(一般约束问题)

$$\min\ f(x),\ x\in R^n$$
$$s.t.\quad c_i(x)=0,\ i\in E=\{1,2,\cdots,l\}$$
$$c_i(x)\ge 0,\ i\in I=\{l+1,\cdots,m\}$$

构造如下增广目标函数 $P(x,\sigma)=f(x)+\sigma\tilde{P}(x)$,

其中 $\tilde{P}(x)=\sum_{i=1}^{l}|c_i(x)|^{\beta}+\sum_{j=l+1}^{m}|\min(0,c_j(x))|^{\alpha}\quad(\alpha\ge 1,\beta\ge 1)$

$\tilde{P}(x)$ 称为罚函数，σ>0称为罚因子.

求解原问题转化为求解一系列的无约束问题

$\min P(x,\sigma_k)(\sigma_k\to+\infty)$.

外罚函数法(算法步骤)

算法4.2.1 外罚函数法

取控制误差ε>0,罚因子放大系数c >1,初始点x_0,

初始罚因子σ_1,令k=1.

Step1 以x_{k-1}为初始点解无约束问题

$$\min P(x,\sigma_k)=f(x)+\sigma_k\tilde{P}(x),$$

得最优解$x_k=x(\sigma_k)$.

Step2 若 $\sigma_k\tilde{P}(x_k)<\varepsilon$, 则以$x_k$为问题的近似最优解.

Stop.

否则,令$\sigma_{k+1}=c\sigma_k$,$k=k$+1,转Step1.

外罚函数法(收敛性)

对SUMT外点法产生的点列$\{x_k\}$,

$$\begin{aligned}P(x_{k+1},\sigma_{k+1})&=f(x_{k+1})+\sigma_{k+1}\tilde{P}(x_{k+1})\\&\ge f(x_{k+1})+\sigma_k\tilde{P}(x_{k+1})\qquad \sigma_{k+1}\ge\sigma_k,\ \tilde{P}(x_{k+1})\ge 0\\&=P(x_{k+1},\sigma_k)\\&\ge P(x_k,\sigma_k)\qquad \boxed{x_k\text{是}P(x,\sigma_k)\text{的最优解.}}\end{aligned}$$

因此有 $\boxed{P(x_{k+1},\sigma_{k+1})\geqslant P(x_k,\sigma_k)}$

外罚函数法(收敛性)

$f(x_{k+1})+\sigma_k\tilde{P}(x_{k+1})\ge f(x_k)+\sigma_k\tilde{P}(x_k)$ $\boxed{x_k\text{是}P(x,\sigma_k)\text{ 最优解.}}$

$\Rightarrow$ $\boxed{f(x_{k+1})-f(x_k)\ge\sigma_k[\tilde{P}(x_k)-\tilde{P}(x_{k+1})]}$

$f(x_k)+\sigma_{k+1}\tilde{P}(x_k)\ge f(x_{k+1})+\sigma_{k+1}\tilde{P}(x_{k+1})$ $\boxed{x_{k+1}\text{是}P(x,\sigma_{k+1})\text{ 最优解.}}$

$\Rightarrow$ $\sigma_{k+1}[\tilde{P}(x_k)-\tilde{P}(x_{k+1})]\ge f(x_{k+1})-f(x_k)$

因此 $\sigma_{k+1}[\tilde{P}(x_k)-\tilde{P}(x_{k+1})]\ge\sigma_k[\tilde{P}(x_k)-\tilde{P}(x_{k+1})]$

$(\sigma_{k+1}-\sigma_k)[\tilde{P}(x_k)-\tilde{P}(x_{k+1})]\ge 0\Rightarrow$ $\tilde{P}(x_k)\ge\tilde{P}(x_{k+1})$ $\sigma_{k+1}\ge\sigma_k$

再根据上面(绿色)标出的不等式,有 $f(x_{k+1})\ge f(x_k)$

我们得到书中P157引理4.2.1

外罚函数法(收敛性)

如果原来约束问题有多个最优解,那么SUMT外点法产生的点列不能保证收敛.

例如, $\min f(x)=x_1+x_2$ 的最优解有无穷多个

$s.t.\ x_1+x_2=1$

问题的增广目标函数为 $P(x,\sigma)=x_1+x_2+\sigma(x_1+x_2-1)^2$

该函数的极小点(无穷多)位于直线 $x_1+x_2=1-\frac{1}{2\sigma}$

在求解过程中,虽然迭代点向直线x_1+x_2=1靠近.

但是不能保证迭代点列收敛.

外罚函数法(收敛性)

$$\min\ f(x),\ x\in R^n$$
$$s.t.\quad c_i(x)=0,\ i\in E=\{1,2,\cdots,l\}\qquad P(x,\sigma_k)=f(x)+\sigma_k\tilde{P}(x),$$
$$c_i(x)\ge 0,\ i\in I=\{l+1,\cdots,m\}$$

定理4.2.2 设上面优化问题与增广目标函数的整体最优解分别是x^*与x_k,对于正数序列$\{\sigma_k\}$, $\sigma_{k+1}\geqslant\sigma_k$,且$\sigma_k\to+\infty$,则由SUMT外点法产生的点列$\{x_k\}$的任何聚点$\underline{x}$必是约束优化问题的整体最优解.

注:所谓聚点,是指子序列的极限.

例:点列1,0,-1,1,0,-1,1,0,-1,···,有三个聚点1,0,-1.

外罚函数法(收敛性)

证明 不妨设$x_k \to \underline{x}$.由于x^*和x_k分别是原问题与增广目标函数的整体最优解,$\tilde{P}(x^*)=0$,从而

$$f(x^*)=f(x^*)+\sigma_k\tilde{P}(x^*)$$

$$\geq f(x_k)+\sigma_k\tilde{P}(x_k)=P(x_k,\sigma_k)$$

x_k是$P(x,\sigma_k)$ 最优解.

$P(x_{k+1},\sigma_{k+1})\geq P(x_k,\sigma_k)$ 引理4.2.1

$P(x_k,\sigma_k)$存在极限,设为p^0. 单调有界数列必收敛

$f(x_{k+1})\geq f(x_k)$ 引理4.2.1

$f(x_k)$存在极限,设为f^0. 单调有界数列必收敛

$$\sigma_k\tilde{P}(x_k)=P(x_k,\sigma_k)-f(x_k)\to p^0-f^0$$

$$\tilde{P}(x_k)\to 0$$ $\sigma_k\to+\infty$

外罚函数法(收敛性)

由于$x_k \to \underline{x}$,且 $\tilde{P}(x)$连续,因此 $\tilde{P}(\underline{x})\to 0.$

即$\underline{x}$为可行解.而x^*为整体最优解,所以$f(x^*)\leqslant f(\underline{x})$.

根据$x_k \to \underline{x}$以及 $f(x_k)\leq P(x_k,\sigma_k)\leq f(x^*)$

得到$f(\underline{x})\leqslant f(x^*)$.

因此$f(\underline{x})=f(x^*)$.

所以$\underline{x}$也是整体最优解.

注:对 $f(x_k)\leq P(x_k,\sigma_k)\leq f(x^*)$两边取极限可得$f^0=p^0$.

因此 $\sigma_k\tilde{P}(x_k)\to 0.$

这是在算法中取 $\sigma_k\tilde{P}(x_k)<\varepsilon$作为终止准则的原因.

算例(外罚函数法)

例4.2.3 求解约束问题

$$\min\ f(x)=(x_1-2)^4+(x_1-2x_2)^2$$
$$s.t.\quad x_1^2-x_2=0$$

解 取增广目标函数为$(x_1-2)^4+(x_1-2x_2)^2+\sigma(x_1^2-x_2)^2$, $x_0=(2,1)^T,\sigma_1=0.1,c=10,\varepsilon=10^{-3}$.对于无约束问题,采用重新开始的PRP算法($\varepsilon=10^{-3}$)求解.得到的结果见下表.

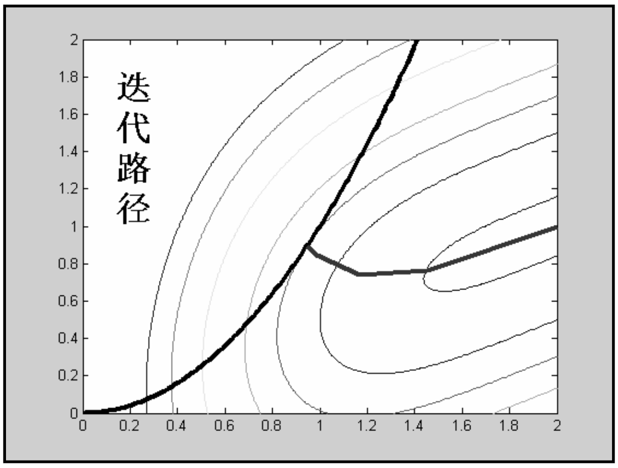
注: (1)函数中应为$(x_1-2x_2)^2$

(2)表4-1中有许多打印错误.

算例(外罚函数法)

k	x_k	$f(x_k)$	$\sigma_k\tilde{P}(x_k)$
0	$(2,1)^T$	0	0.9
1	$(1.45388,0.76076)^T$	0.09353	1.83058
2	$(1.16872,0.74067)^T$	0.57524	3.90930
3	$(0.99061,0.84246)^T$	1.52013	1.92822
4	$(0.95076,0.88749)^T$	1.89123	0.27170
5	$(0.94611,0.89344)^T$	1.94052	0.02828
6	$(0.94563,0.89405)^T$	1.94562	0.00284
7	$(0.94556,0.89409)^T$	1.94619	6.8 e-6

精确解$(0.94558299,0.89412720)^T$.



内罚函数法

$$\min\ f(x),x\in R^n$$
$$s.t.\quad c_i(x)\geq 0,i\in I=\{1,2,\cdots,m\}$$

考虑上面的不等式约束问题.当x从可行域

$$D=\{x\in R^n\mid c_i(x)\geq 0,i=1,2,\cdots,m\}$$

的内部趋于边界时,至少有一个$c_i(x)$趋于零,因此我们构造下面的增广目标函数

$$B(x,r)=f(x)+r\tilde{B}(x)$$

其中 $\tilde{B}(x)=\sum_{i=1}^{m}\frac{1}{c_i(x)}$ 或 $\tilde{B}(x)=-\sum_{i=1}^{m}\ln c_i(x)$

称为内罚函数或障碍函数(Barrier function),参数r称为罚因子.

内罚函数法

$B(x,r)= f(x)+r\tilde{B}(x)$ $\tilde{B}(x)=\sum_{i=1}^{m}\frac{1}{c_i(x)}$ $\tilde{B}(x)=-\sum_{i=1}^{m}\ln c_i(x)$

当x为可行域D的内点时,$\tilde{B}(x)$的值是有限的正数,而r>0很小,几乎不受惩罚.

当x接近D的边界时,$\tilde{B}(x)$的值趋于无穷大,受惩罚很大,迫使极小点落在D的内部,最终逼近$f(x)$的约束极小点.

令正数列$\{r_k\}$逐步减小,并趋于零,则原约束问题转化为系列无约束问题.

内罚函数法(算法步骤)

算法4.2.2 内罚函数法

取控制误差ε>0,罚因子缩小系数0<c<1,初始点$x_0\in D_0$(可行域内部),初始罚因子r_1,令k=1.

Step1 以x_{k-1}为初始点解无约束问题

$$\min B(x,r_k)= f(x)+r_k\tilde{B}(x),$$

得最优解$x_k=x(r_k)$.

Step2 若 $r_k\tilde{B}(x_k)<\varepsilon$, 则以$x_k$为问题的近似最优解. Stop.

否则,令$r_{k+1}=cr_k$,$k=k+1$,转Step1.

内罚函数法(算例)

例4.2.4 用内点法求解 $\min f(x)=\frac{1}{3}(x_1+1)^3+x_2$

$$s.t.\quad 1-x_1\le 0,\quad x_2\ge 0$$

解 增广目标函数为

$$B(x_1,x_2,r)=\frac{1}{3}(x_1+1)^3+x_2+r\left(\frac{1}{x_1-1}+\frac{1}{x_2}\right).$$

令 $\frac{\partial B}{\partial x_1}=(x_1+1)^2-\frac{r}{(x_1-1)^2}=0$

$\frac{\partial B}{\partial x_2}=1-\frac{r}{x_2^2}=0$ 　　当$r\to 0$时,得$x^*=(1,0)^T$.

得 $x(r)=(\sqrt{1+\sqrt{r}},\sqrt{r})^T$, 　　$f^*=8/3$.

内罚函数法(算例)

一般而言,$B(x,r)$的最优点要用无约束问题的算法来求解.

$x_0=(3,4)^T$,r_1=10,c=10,$\varepsilon=10^{-3}$.对于无约束问题,采用重新开始的PRP算法($\varepsilon=10^{-3}$)求解.得到的结果见下表.

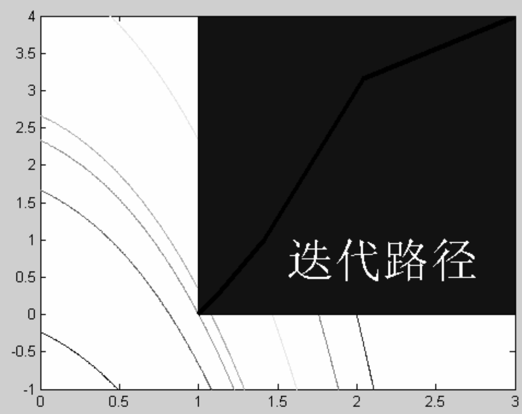
注:在求解无约束问题时,要注意限制一维搜索的初始区间,即保证迭代点始终在可行域之内.

在本问题中,如果对一维搜索的初始区间不加限制,函数值会趋于负无穷.

算例(内罚函数法)

k	x_k	$f(x_k)$	$r_k\tilde{B}(x_k)$
0	$(3,4)^T$	25.3333	250
1	$(2.04017,3.16226)^T$	12.5286	1.27761
2	$(1.41421,1.00007)^T$	5.69042	0.34141
3	$(1.14727,0.31630)^T$	3.61648	0.09952
4	$(1.04881,0.10004)^T$	2.96674	0.03048
5	$(1.01569,0.03162)^T$	2.76154	0.00954
6	$(1.00499,0.01000)^T$	2.69667	0.00300
7	$(1.00158,0.00316)^T$	2.67615	0.00095

精确解$(1,0)^T$.



内罚函数法(收敛性)

关于内罚函数法,有类似于外罚函数法的收敛性结论.

引理4.2.3 对于由SUMT内点法产生的点列$\{x_k\}$,总有$B(x_{k+1},r_{k+1})\leqslant B(x_k,r_k)$.

定理4.2.4 设可行域内点集$D_0=\{x\in R^n|c_i(x)>0,i\in I\}$非空,$f(x)$在$D$上存在整体极小点$x^*$,对于严格单调递减正数序列$\{r_k\}$, $r_{k+1}<r_k$,且$r_k\to 0$,则由SUMT内点法产生的点列$\{x_k\}$的任何聚点必是不等式约束优化问题的整体最优解.

4.2.4 乘子法

$$\begin{aligned}&\min\ f(x),x\in R^n\\ &s.t.\quad c_i(x)=0,i\in E=\{1,2,\cdots,l\}\end{aligned}$$

考虑上面的等式约束问题.

设x^*为该问题的最优解,在一定条件下,存在λ^*使(x^*,λ^*)为Lagrange函数 $L(x,\lambda)=f(x)-\sum_{j=1}^{l}\lambda_i c_j(x)$ 的驻点,即 $\nabla L(x^*,\lambda^*)=\begin{pmatrix}\nabla_x L\\ \nabla_\lambda L\end{pmatrix}=0.$

一个自然的问题是,能否找到λ^*使得(x^*,λ^*)是Lagrange函数的极小点.那样的话,约束问题就转化为无约束问题.

乘子法

例4.2.5 求解约束问题 $\begin{aligned}&\min\ f(x)=x_1^2-3x_2-x_2^2\\ &s.t.\quad x_2=0\end{aligned}$

此问题的最优解为$(0,0)^T$. Lagrange函数为

$$L(x,\lambda)=x_1^2-3x_2-x_2^2-\lambda x_2=x_1^2-(\lambda+3)x_2-x_2^2$$

对于任何λ,$L(x,\lambda)$关于x的极小点不存在.

对于等式约束问题,我们构造了辅助函数

$$P(x,\sigma)=f(x)+\sigma\sum_{i=1}^{l}c_i(x)^2$$

然后令$\sigma\to\infty$,在一定条件下求得原问题的解.

乘子法

$$P(x,\sigma)=f(x)+\frac{\sigma}{2}\sum_{i=1}^{l}c_i(x)^2$$

另一个问题是,能否找到σ^*,使得$P(x,\sigma^*)$的无约束极小点是原约束问题的极小点.

如果x^*是$P(x,\sigma^*)$的极小点,则有

$$0=\nabla_x P(x^*,\sigma^*)=\nabla f(x^*)+\sigma^*\sum_{i=1}^{l}c_i(x^*)\nabla c_i(x^*)$$

由于x^*是可行点,$c_i(x^*)=0$,因此$\nabla f(x^*)=0.$

这在一般情况下是不成立的.

等式约束问题的乘子法

我们将上述两种思路结合起来,即考虑问题,能否找到λ^*,σ^*,使得x^*是下面的增广Lagrange函数的极小点. $L(x,\lambda)+\sigma\tilde{P}(x)$

考虑例4.2.5中的问题 $\begin{aligned}&\min\ f(x)=x_1^2-3x_2-x_2^2\\ &s.t.\quad x_2=0\end{aligned}$

$L(x,\lambda)=x_1^2-(\lambda+3)x_2-x_2^2$ 取 $\tilde{P}(x)=x_2^2/2$

增广Lagrange函数为 $M(x,\lambda,\sigma)=x_1^2-(\lambda+3)x_2+\frac{\sigma-2}{2}x_2^2$

当$\lambda^*=-3$,$\sigma\geqslant\sigma^*=2$时,原问题最优解$(0,0)^T$是增广Lagrange函数的最优解.

等式约束问题的乘子法

$\begin{aligned}&\min\ f(x)=x_1^2-3x_2-x_2^2\\ &s.t.\quad x_2=0\end{aligned}$ $M(x,\lambda,\sigma)=x_1^2-(\lambda+3)x_2+\frac{\sigma-2}{2}x_2^2$

反之,求解无约束问题

$$\min M(x,\lambda,\sigma)=x_1^2+\frac{\sigma-2}{2}x_2^2-(\lambda+3)x_2$$

令 $0=\frac{\partial M}{\partial x_1}=2x_1,$ $0=\frac{\partial M}{\partial x_2}=(\sigma-2)x_2-(\lambda+3).$ 得 $x_0=(0,\frac{\lambda+3}{\sigma-2})^T.$

要求x_0满足约束条件$x_2=0$,必须取$\lambda=-3$,从而$x_0=(0,0)^T=x^*$,得到原约束问题的最优解.

等式约束问题的乘子法

考虑等式约束问题 $\begin{array}{l}\min f(x), x\in R^n\\ s.t.\quad C(x)=0\end{array}$

其中$C(x)=(c_1(x),\cdots,c_l(x))^T$,目标函数和约束函数二次连续可微.

设$\lambda\in R^l$为Lagrange乘子向量,则上面问题的Lagrange 函数为 $L(x,\lambda)=f(x)-\lambda^T C(x)$

对任意的$x^*\in D$,有 $L(x^*,\lambda^*)=f(x^*)-\lambda^{*T}C(x^*)=f(x^*)$
$\le f(x)-\lambda^{*T}C(x)=L(x,\lambda^*)$

因此,原约束问题等价于下面的约束问题

$$\begin{array}{l}\min L(x,\lambda^*)\\ s.t.\quad C(x)=0\end{array}$$

等式约束问题的乘子法

$\begin{array}{l}\min L(x,\lambda^*)\\ s.t.\quad C(x)=0\end{array}$ 对左边问题,构造增广Lagrange函数

$$M(x,\lambda,\sigma)=L(x,\lambda)+\frac{\sigma}{2}C(x)^T C(x)$$

定理4.2.6 设在上面等式约束问题中,$x^*\in R^n$和$\lambda^*\in R^n$满足二阶充分条件(Th4.1.2),则存在一个数$\sigma^*>0$,对所有的$\sigma\geqslant\sigma^*$,x^*是增广目标函数的严格局部极小点;

反之,若$C(x_0)=0$且x_0对某个λ_0是增广目标函数的局部极小点,则x_0是等式约束问题的局部极小点.

等式约束问题的乘子法

乘子法并不要求σ趋于无穷大.只要σ大于某个正数σ^*,就能保证无约束问题min $M(x,\lambda^*,\sigma)$的最优解为原问题的最优解.

要解决的问题是,如何确定λ^*?

我们采用迭代的方法求出λ^*.

求解无约束问题min $M(x,\lambda_k,\sigma)$,其解为x_k,

然后修正λ_k为λ_{k+1},再求解min $M(x,\lambda_{k+1},\sigma)$.

得到两个点列$\{x_k\}$,$\{\lambda_k\}$,希望$x_k\to x^*$,$\lambda_k\to\lambda^*$.

等式约束问题的乘子法

如何对λ_k进行修正?

设已有λ_k和x_k,由$M(x,\lambda,\sigma)$的定义

$$\nabla_x M(x_k,\lambda_k,\sigma)=\nabla f(x_k)-\nabla C(x_k)(\lambda_k-\sigma C(x_k))=0$$

希望$x_k\to x^*$,$\lambda_k\to\lambda^*$, 又 $\nabla f(x^*)-\nabla C(x^*)\lambda^*=0$

采取公式$\lambda_{k+1}=\lambda_k-\sigma C(x_k)$是比较合理的.

若$\{\lambda_k\}$收敛,则有$C(x_k)\to 0$.

若$x_k\to x^*$,则有$C(x^*)=0$,即x^*为可行解.

等式约束问题的乘子法——PH算法

Step1选定初始点x_0,初始乘子向量λ_1,初始罚因子σ_1,放大系数$c>1$,控制误差ε,常数$\theta\in(0,1)$,令$k=1$;

Step2以x_1为初始点求解无约束问题

$$\min M(x,\lambda_k,\sigma)=f(x_k)-\lambda_k^T C(x)+\frac{\sigma_k}{2}C(x)^T C(x)$$

得到的最优解即为x_k;

Step3当$\|C(x_k)\|<\varepsilon$时,x_k为所求的最优解,Stop;

Step4当$\|C(x_k)\|/\|C(x_{k-1})\|>\theta$时,令$\sigma_{k+1}=c\sigma_k$;

Step5令$\lambda_{k+1}=\lambda_k-\sigma_k C(x_k)$,$k=k+1$,转Step2.

例题(等式约束问题的乘子法)

例4.2.7求解约束问题 $\begin{array}{l}\min f(x)=x_1^2+x_2^2\\ s.t.\ x_1+x_2-2=0\end{array}$

解 增广Lagrange函数为

$$M(x_1,x_2,\lambda,\sigma)=x_1^2+x_2^2-\lambda(x_1+x_2-2)+\frac{\sigma}{2}(x_1+x_2-2)^2$$

令 $\frac{\partial M}{\partial x_1}=2x_1-\lambda+\sigma(x_1+x_2-2)=0$

$\frac{\partial M}{\partial x_2}=2x_2-\lambda+\sigma(x_1+x_2-2)=0$

得 $x_1=x_2=\dfrac{2\sigma+\lambda}{2\sigma+2}$.

将其中的λ视为λ_k,根据乘子迭代公式

$$\lambda_{k+1}=\lambda_k-\sigma(x_1+x_2-2)=\frac{1}{\sigma+1}\lambda_k+\frac{2\sigma}{\sigma+1}$$

例题(等式约束问题的乘子法)

$$\lambda_{k+1}=\frac{1}{\sigma+1}\lambda_k+\frac{2\sigma}{\sigma+1}$$

当$\sigma>0$时,$\{\lambda_k\}$收敛.设$\lambda_k\to\lambda^*$,对上式取极限得

$$\lambda^*=\frac{1}{\sigma+1}\lambda^*+\frac{2\sigma}{\sigma+1}$$

因此$\lambda^*=2$.

在 $x_1=x_2=\dfrac{2\sigma+\lambda}{2\sigma+2}$ 中,令$\lambda=2$,

得原问题的最优解$x^*=(1,1)^T$.

不等式约束问题的乘子法

$$\min f(x), x\in R^n$$
$$s.t.\quad c_i(x)\ge 0, i\in I=\{1,2,\cdots,m\}$$

对不等式约束问题,引进辅助变量$z_i(i=1,2,\cdots,m)$,上面的问题转化为等价的等式约束问题.

$$\min f(x), x\in R^n$$
$$s.t.\quad c_i(x)-z_i^2=0, i=1,2,\cdots,m.$$

其增广Lagrange函数为

$$\tilde{M}(x,z,\lambda,\sigma)=f(x)-\sum_{i=1}^m\lambda_i(c_i(x)-z_i^2)+\frac{\sigma}{2}\sum_{i=1}^m(c_i(x)-z_i^2)^2.$$

不等式约束问题的乘子法

$$\tilde{M}(x,z,\lambda,\sigma)=f(x)-\sum_{i=1}^m\lambda_i(c_i(x)-z_i^2)+\frac{\sigma}{2}\sum_{i=1}^m(c_i(x)-z_i^2)^2.$$

先考虑这一函数关于z的极小化函数,关于变量z_i,它是z_i^2的二次函数

$$\tilde{M}(x,z,\lambda,\sigma)=(z_i^2-\frac{1}{\sigma}(\sigma c_i(x)-\lambda_i))^2+u_i$$

当 $\sigma c_i(x)-\lambda_i\ge 0$ 时,要使函数取最小, $z_i^2=c_i(x)-\dfrac{\lambda_i}{\sigma}$

否则$z_i=0$. 因此 $z_i^2=\dfrac{1}{\sigma}\max(0,\sigma c_i(x)-\lambda_i), i=1,2,\cdots,m$

得到增广的目标函数

$$M(x,\lambda,\sigma)=f(x)+\frac{1}{2\sigma}\sum_{i=1}^m\{[\max(0,\lambda_i-\sigma c_i(x))]^2-\lambda_i^2\}$$

不等式约束问题的乘子法

$$M(x,\lambda,\sigma)=f(x)+\frac{1}{2\sigma}\sum_{i=1}^m\{[\max(0,\lambda_i-\sigma c_i(x))]^2-\lambda_i^2\}$$

乘子迭代公式

$$(\lambda_{k+1})_i=\max[0,(\lambda_k)_i-\sigma c_i(x_k)], i=1,2,\cdots,m$$

终止准则 $\left(\sum_{i=1}^m[\max(c_i(x_k),\frac{(\lambda_k)_i}{\sigma})]^2\right)^{1/2}<\varepsilon$

对于一般约束问题,只要综合等式约束和不等式约束的情况写出增广目标函数来求解.

其算法称为PHR算法.

例题(PHR算法)

例4.2.8 用PHR算法求解 $\min f(x)=x_1^2+x_2^2$, $s.t.\quad x_1+x_2\ge 2$

解 增广目标函数为

$$M(x_1,x_2,\lambda,\sigma)=x_1^2+x_2^2+\frac{1}{2\sigma}\{[\max(0,\lambda-\sigma(x_1+x_2-2))]^2-\lambda^2\}$$

$$=\begin{cases}x_1^2+x_2^2-\dfrac{\lambda^2}{2\sigma}, & x_1+x_2-2>\dfrac{\lambda}{\sigma}\\ x_1^2+x_2^2+\dfrac{1}{2\sigma}\{[\lambda-\sigma(x_1+x_2-2)]^2-\lambda^2\}, & x_1+x_2-2\le\dfrac{\lambda}{\sigma}\end{cases}$$

当 $x_1+x_2-2>\dfrac{\lambda}{\sigma}$ 时,令 $\dfrac{\partial M}{\partial x_1}=2x_1=0,\ \dfrac{\partial M}{\partial x_2}=2x_2=0.$

得 $\tilde{x}=(0,0)^T$.当σ充分大时,该点不满足 $x_1+x_2-2>\dfrac{\lambda}{\sigma}$.

从而 $\tilde{x}=(0,0)^T$. 不是极小点.

例题(PHR算法)

$$M=x_1^2+x_2^2+\frac{1}{2\sigma}\{[\lambda-\sigma(x_1+x_2-2)]^2-\lambda^2\}, x_1+x_2-2\le\frac{\lambda}{\sigma}$$

当 $x_1+x_2-2\le\dfrac{\lambda}{\sigma}$ 时, 令

$$\frac{\partial M}{\partial x_1}=2x_1-[\lambda-\sigma(x_1+x_2-2)]=0,$$
$$\frac{\partial M}{\partial x_2}=2x_2-[\lambda-\sigma(x_1+x_2-2)]=0,$$

得 $\tilde{x}=(\dfrac{2\sigma+\lambda}{2\sigma+2},\dfrac{2\sigma+\lambda}{2\sigma+2})^T.$

当σ充分大时,该点满足

$$x_1+x_2-2\le\frac{\lambda}{\sigma}.$$

将其中的λ视为λ_k,采用下面的公式得出λ_{k+1}.

$$\lambda_{k+1}=\max(0,\lambda_k-\sigma(x_1+x_2-2))$$
$$=\max(0,\frac{2\sigma+\lambda_k}{\sigma+1})$$

若给定$\lambda_1>0$,且$\sigma>0$,则

$$\lambda_{k+1}=\frac{1}{\sigma+1}+\frac{2\sigma}{\sigma+1}>0.$$

以下过程同例4.2.7.

§ 4.3 投影梯度法

4.3.1 可行方向及其性质

$\min\ f(x), x \in R^n$
$s.t.\quad A^T x \ge b$

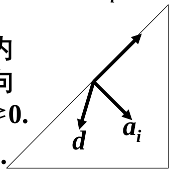
本节研究左边的线性不等式约束问题.

对于线性约束$a_i^T x \geqslant b_i$,在几何上表示半空间.

其边界是一个超平面$a_i^T x = b_i$,其法向量指向a_i半空间的内部.

对于超平面上的一点,指向半空间内部的方向是可行方向,易见可行方向d与a_i的夹角为锐角或直角,即$a_i^T d \geqslant 0$.

若$a_i^T d = 0$,则沿该方向的点在边界上.



投影梯度法

例4.3.2 求解

$$\min\ f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2$$

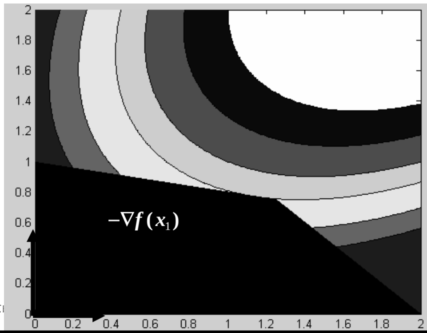
$$s.t.\ \begin{cases} 2 - x_1 - x_2 \ge 0 \\ 5 - x_1 - 5x_2 \ge 0 \\ x_1 \ge 0 \\ x_2 \ge 0 \end{cases}$$

$\nabla f(x) = (4x_1 - 2x_2 - 4, 4x_2 - 2x_1 - 6)^T \quad -\nabla f(x_1) = (4,6)^T$

负梯度指向可行域内部 $-\nabla f(x_1) = 4\bullet(1,0)^T + 6\bullet(0,1)^T$

仿照单纯形方法,我们沿着边界使函数值下降.

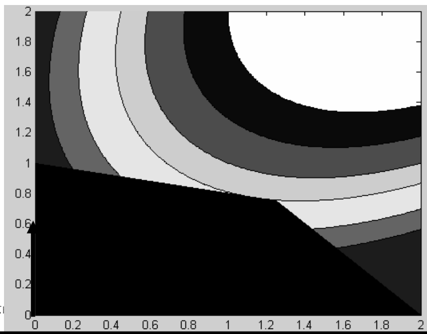
由于负梯度方向沿方向$(0,1)^T$的分量较大,我们先沿此方向$p_1=(0,1)^T$进行一维搜索.



$p_1=(0,1)^T$ 为了保证迭代点在可行域内部,显然步长α_k的最大值为1.

作线性搜索,即求解 $\min\limits_{0<\alpha\le 1} f(x_1 + \alpha p_1) = 72\alpha^2 - 36\alpha$

得$\alpha_1=1$,
$x_2=x_1+\alpha_1 p_1 =(0,1)^T$.



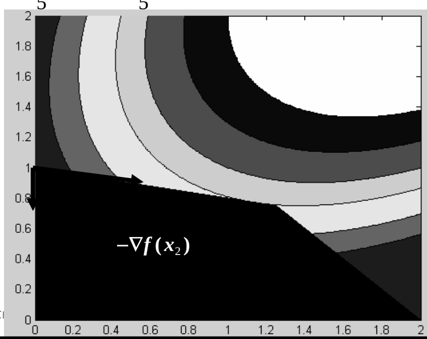
$\nabla f(x) = (4x_1 - 2x_2 - 4, 4x_2 - 2x_1 - 6)^T \quad x_2 = (0,1)^T \quad -\nabla f(x_2) = (6,2)^T$

在x_2处沿边界指向可行域的方向为$(0,-1)^T$和$(5,-1)^T$.

$$(6,2)^T = \frac{6}{5}(5,-1)^T - \frac{16}{5}(0,-1)^T$$

因此我们下一步沿着方向$p_2=(5,-1)^T$搜索.

可以求得步长α_k的最大值为5/4.

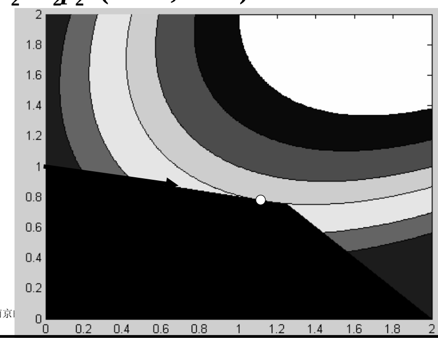


作线性搜索, $\min\limits_{0<\alpha\le 5/4} f(x_2+\alpha p_2)=62\alpha^2-28\alpha-4$ 即求解

得α_2=7/31,x_3=x_2+$\alpha_2 p_2$=(35/31,24/31)^T.

可以验证x_3为KT点.

又$f(x)$为凸函数,x_3为最优点



投影梯度法

上述思路的总结

方法:在任一点,找出有效约束,沿着与梯度成锐角的边界方向向可行域内搜索.

问题:

不是极点的点如何搜索?(最优点可能在内部)

沿边界的方向通过什么方法确定?

如何确定已达到最优?

投影梯度法

如何判断是否达到最优?通常用KT条件来判断.

假设当前的迭代点为x,设I^*为有效约束集,即有 $a_i^T x=b_i (i\in I)$,

下面假设$\{a_i (i\in I^*)\}$线性无关,

我们根据 $\nabla f(x)$ 是否可以写为$\{a_i (i\in I^*)\}$的线性组合来分情况讨论迭代方法以及乘子向量的计算方法.

情形一

$\nabla f(x)$可以写为$\{a_i (i\in I^*)\}$的线性组合.

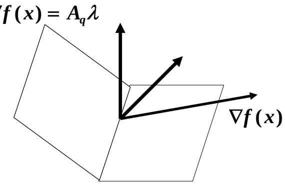
令 $\nabla f(x)=\sum\limits_{i\in I^*}\lambda_i a_i$, 如果$\lambda_i\geqslant 0$,则$x$为KT点.

如何求组合系数λ_i?

$\nabla f(x)=\sum\limits_{i\in I^*}\lambda_i a_i$ 可以写为 $\nabla f(x)=A_q\lambda$

其中A_q是以$\{a_i (i\in I^*)\}$为列向量构成的矩阵.

若A_q为方阵.则A_q非奇异,且 $\lambda=A_q^{-1}\nabla f(x)$.



情形一

$\nabla f(x)=A_q\lambda$

在A_q为列满秩时,$A_q^T A_q$为可逆矩阵.

记$A_q^+=(A_q^T A_q)^{-1}A_q^T$,则$A_q^+A_q=I$.

在$\nabla f(x)=A_q\lambda$的两边同时左乘A_q^+,得 $\lambda=A_q^+\nabla f(x)$.

注: A_q^+为A_q的广义逆矩阵, $A^+A=I$.

但是$AA^+=I$不一定成立.

当A为非奇异方阵时,$A^+=A^{-1}$.

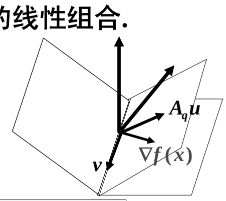
此时如何求下降方向进行迭代在后面讨论.

情形二

$\nabla f(x)$不能写为$\{a_i (i\in I^*)\}$的线性组合.

此时, $\nabla f(x)$可以分解成两个正交的向量,一个可以写为上述向量的线性组合,一个垂直于上述向量,即

$\nabla f(x)=A_q u+v,\ A_q^T v=0 (v\neq 0)$.



$A_q u$表示$\{a_i (i\in I^*)\}$的一个线性组合

$A_q^T v$=0表示与$\{a_i (i\in I^*)\}$中每个向量正交

由图形可以看出-v是一个沿着x点有效约束边界的可行下降方向.

情形二

$$\nabla f(x) = A_q u + v, A_q^T v = 0.$$

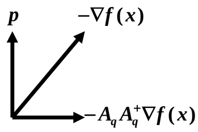
下面求出梯度的分解中的两个向量$A_q u$与v.
根据上面的等式有

$$A_q A_q^+ \nabla f(x) = A_q A_q^+ A_q u + A_q A_q^+ v$$
$$= A_q A_q^+ A_q u + A_q (A_q^T A_q)^{-1} A_q^T v$$
$$= A_q u. \quad \boxed{A_q^+ A_q = I} \quad \boxed{A_q^T v = 0}$$

因此 $A_q u = A_q A_q^+ \nabla f(x), v = (I - A_q A_q^+)\nabla f(x)$.

$$\nabla f(x) = A_q A_q^+ \nabla f(x) + (I - A_q A_q^+)\nabla f(x).$$

情形二的可行下降方向



$$\nabla f(x) = A_q A_q^+ \nabla f(x) + (I - A_q A_q^+)\nabla f(x).$$

令 $p = -(I - A_q A_q^+)\nabla f(x)$,

显然p与$-\nabla f(x)$的夹角是锐角,
因此是下降方向.

从图形上可以看出 $p^T(-\nabla f(x)) = \| p \|^2$.

下面从理论上给出推导.

情形二的可行下降方向

$$p = -(I - A_q A_q^+)\nabla f(x),$$

考察矩阵 $P_q = I - A_q A_q^+$

有 $P_q^2 = (I - A_q A_q^+)^2 = I - 2A_q A_q^+ + A_q A_q^+ A_q A_q^+$ $\boxed{A_q^+ A_q = I}$
$= I - A_q A_q^+$

另外 $P_q^T = (I - A_q A_q^+)^T = (I - A_q (A_q^T A_q)^{-1} A_q^T)^T = P_q$.

因此 $p^T \nabla f(x) = -\nabla f(x)^T P_q^T \nabla f(x)$ $\boxed{P_q^T = P_q = P_q^2 = P_q^T P_q}$

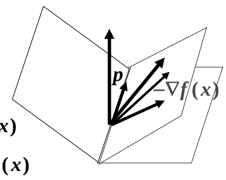
$$= -\nabla f(x)^T P_q^T P_q \nabla f(x) = -\| P_q \nabla f(x) \|^2 = -\| p \|^2 < 0.$$

方向p是下降方向.

情形二的可行下降方向

$$p = -P_q \nabla f(x) = -(I - A_q A_q^+)\nabla f(x)$$

那么p是否是可行方向呢?
从图形上可以看出p是位
于有效约束的边界上.



$$A_q^T p = -A_q^T (I - A_q (A_q^T A_q)^{-1} A_q^T)\nabla f(x)$$
$$= (-A_q^T + A_q^T A_q (A_q^T A_q)^{-1} A_q^T)\nabla f(x)$$

即有效约束对应的超平面的法向量与方向p
均是正交的.

情形二的可行下降方向

$$p = -P_q \nabla f(x) = -(I - A_q A_q^+)\nabla f(x) \qquad A_q^T p = 0$$

对于点$x+\alpha p$,有

$$A_q^T (x + \alpha p) = A_q^T x + \alpha A_q^T p = A_q^T x = b_q$$

因此点$x+\alpha p$依然满足有效约束的相关不等式.

对于无效约束,只要α取值足够小,不等式就
能够成立.

α的取值上限

为了保证$x+\alpha p$在可行域内, α取值必须足够小,
使得在x点处的无效约束依然成立.

对约束$a_i^T x \geqslant b_i$,将点$x+\alpha p$代入此不等式有

$a_i^T (x + \alpha p) \ge b_i$ 即 $(a_i^T x - b_i) + \alpha a_i^T p \ge 0$

由于 $a_i^T x - b_i \ge 0$, 在$a_i^T p \geqslant 0$时,上式对$\alpha \geqslant 0$成立.

若$a_i^T p < 0$,则有 $0 \le \alpha \le \dfrac{a_i^T x - b_i}{-a_i^T p}$

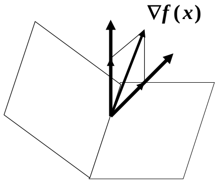
因此α的取值上限为 $\alpha_{\max} = \min\{\dfrac{a_i^T x - b_i}{-a_i^T p} \mid a_i^T p < 0\}$

情形一的可行下降方向

$$\nabla f(x) = A_q\lambda.$$

其中 $\lambda = A_q^+\nabla f(x).$

若$\lambda\geqslant 0$,显然x为KT点.

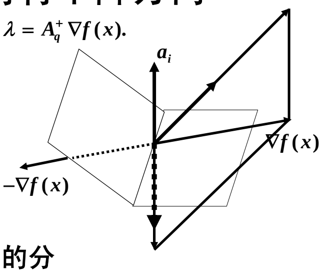


情形一的可行下降方向

$\nabla f(x) = A_q\lambda.$ 其中 $\lambda = A_q^+\nabla f(x).$

若λ至少有一个分量小于0,设为λ_i<0.

$$\nabla f(x) = A_q\lambda = \sum_{j\in I^*}\lambda_j a_j = \lambda_i a_i + \sum_{j\in I^*, j\neq i}\lambda_j a_j$$



此时负梯度投影到a_i上的分量是指向可行域之外的.

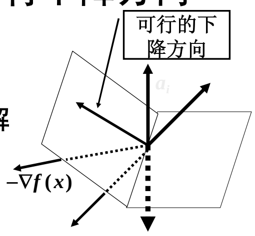
情形一的可行下降方向

$$\nabla f(x) = \lambda_i a_i + \sum_{j\in I^*, j\neq i}\lambda_j a_j$$

因此此时不考虑λ_i对应的约束,而将负梯度分解为两个垂直的向量,

一个是$a_j(j\in I^*, j\neq i)$的线性组合,

一个与这些向量垂直.



由图形上可以看出,我们得到一个可行的下降方向,而且是沿着有效约束(去掉a_i对应的约束)的边界的.

情形一的可行下降方向

$$\nabla f(x) = \lambda_i a_i + \sum_{j\in I^*, j\neq i}\lambda_j a_j$$

设$\{a_j, j\in I^*, j\neq i\}$构成的矩阵为$A_{q-1}$, A_{q-1}为A_q去掉列向量a_i得到的矩阵.再设λ去掉分量λ_i得到的向量为λ_{q-1}. 显然有 $\nabla f(x) = A_{q-1}\lambda_{q-1} + \lambda_i a_i$

现在将负梯度写为如下的形式

$$-\nabla f(x) = A_{q-1}u + v, A_{q-1}^T v = 0.$$

类似与情形二的推导,有

$$-\nabla f(x) = -A_{q-1}A_{q-1}^+\nabla f(x) - (I - A_{q-1}A_{q-1}^+)\nabla f(x).$$

情形一的可行下降方向

$$-\nabla f(x) = -A_{q-1}A_{q-1}^+\nabla f(x) - (I - A_{q-1}A_{q-1}^+)\nabla f(x).$$

设 $P_{q-1} = I - A_{q-1}A_{q-1}^+,\ p_{q-1} = -P_{q-1}\nabla f(x).$

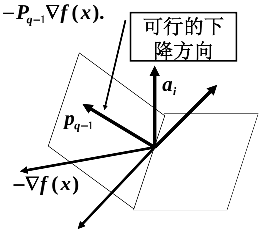
由图形可以看出

(1) $p_{q-1} \neq 0$

(2) $p_{q-1}^T(-\nabla f(x)) = \| p_{q-1}\|^2$

(3) $p_{q-1}^T a_j = 0 (j\in I^*, j\neq i)$

$p_{q-1}^T a_i > 0$



情形一的可行下降方向

$$P_{q-1} = I - A_{q-1}A_{q-1}^+,\quad p_{q-1} = -P_{q-1}\nabla f(x).$$

下面证明p_{q-1}为可行的下降方向.首先证$p_{q-1}\neq 0$.

$$p_{q-1} = -P_{q-1}\nabla f(x) = -(I - A_{q-1}A_{q-1}^+)(A_{q-1}\lambda_{q-1} + \lambda_i a_i) = \lambda_i(A_{q-1}A_{q-1}^+ a_i - a_i)$$

若p_{q-1}=0,则 $A_{q-1}A_{q-1}^+ a_i - a_i = 0.$

是一个向量

$A_{q-1}A_{q-1}^+ a_i - a_i = 0$ 说明a_i可以写为$a_j(j\in I^*, j\neq i)$的线性组合,这与A_q列满秩矛盾.

情形一的可行下降方向

$$p_{q-1} = -P_{q-1}\nabla f(x), \quad P_{q-1} = I - A_{q-1}A_{q-1}^{+},$$

其次证p_{q-1}为下降方向.

类似与情形二中的讨论,有 $P_{q-1}^2 = P_{q-1}, P_{q-1}^T = P_{q-1}$.

因此 $p_{q-1}^T \nabla f(x) = -\nabla f(x)^T P_{q-1}^T \nabla f(x) = -\nabla f(x)^T P_{q-1}^T P_{q-1} \nabla f(x)$

$$= -\| P_{q-1}\nabla f(x) \|^2 = -\| p_{q-1} \|^2 < 0.$$

注:若有效约束只有一个,对应矩阵取为单位阵I,此时下降方向即为负梯度方向.

情形一的可行下降方向

$$p_{q-1} = -P_{q-1}\nabla f(x), \quad P_{q-1} = I - A_{q-1}A_{q-1}^{+},$$

最后证p_{q-1}为可行方向.

$$A_{q-1}^T p_{q-1} = -A_{q-1}^T (I - A_{q-1}(A_{q-1}^T A_{q-1})^{-1} A_{q-1}^T)\nabla f(x)$$

$$= (-A_{q-1}^T + A_{q-1}^T A_{q-1}(A_{q-1}^T A_{q-1})^{-1} A_{q-1}^T)\nabla f(x) = 0.$$

$$p_{q-1} = \lambda_i (A_{q-1}A_{q-1}^{+} a_i - a_i)$$

$$a_i^T p_{q-1} = \lambda_i a_i^T A_{q-1}A_{q-1}^{+} a_i - \lambda_i a_i^T a_i = -\lambda_i a_i^T a_i > 0$$

关于x点处的有效约束对应的方向a_j与p_{q-1}的夹角为锐角或直角,因此p_{q-1}为可行方向.

情形一或情形二的判断?

前面在两种情形下讨论了如何确定迭代的可行下降方向.

关于属于何种情形是通过描述性的语言”梯度可以写为有效约束对应的向量的线性组合”来定义的.

那么如何通过数据特征来区分两种情形?

根据前面的讨论,我们可以通过梯度的投影 $P_q\nabla f(x)$ 是否为零向量来判断.

投影梯度法

已知迭代点x,根据有效约束集$I^*=\{i|a_i^T x=b_i(i\in I)\}$写出对应的矩阵$A_q$及投影矩阵$P_q=I-A_qA_q^{+}$.

$P_q\nabla f(x) \neq 0 \Longrightarrow$ 情形二

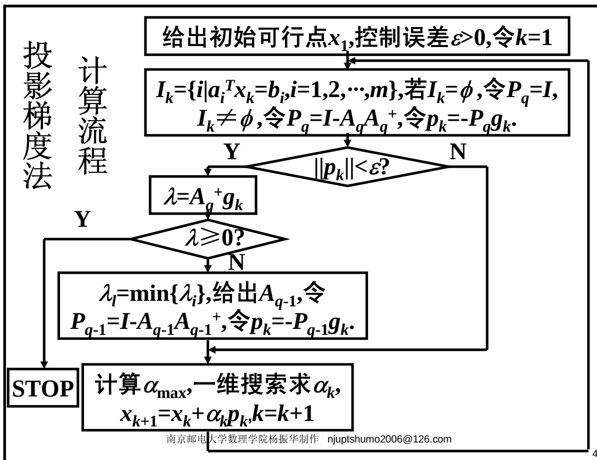
以 $-P_q\nabla f(x)$ 为可行下降方向进行迭代

$P_q\nabla f(x) = 0 \Longrightarrow$ 情形一

求乘子向量λ, 若$\lambda\geqslant 0$,终止迭代

否则,选取一个负分量,给出A_{q-1},P_{q-1}

以$-P_{q-1}\nabla f(x)$为可行下降方向进行迭代



算例(投影梯度法)

例4.3.2 求解

$$\min\ f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2$$

$$s.t. \begin{cases} 2 - x_1 - x_2 \geq 0 \\ 5 - x_1 - 5x_2 \geq 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

解 $A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix}$

$\nabla f(x) = (4x_1 - 2x_2 - 4, 4x_2 - 2x_1 - 6)^T$

取初始点$x_1=(0,0)^T$. 第一次迭代

$-\nabla f(x_1) = (4,6)^T$ 有效集为$\{3,4\}$. $A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$可逆.

因此 $P_2 = I - A_2A_2^{+} = I - A_2A_2^{-1} = O$. $P_2\nabla f(x_1) = 0$.

算例(投影梯度法)

乘子向量 $\lambda = A_2^+\nabla f(x_1) = \nabla f(x_1) = (-4,-6)^T$.

取λ_2=-6，从A_2取中去掉第二列,得到

$$A_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad A_1^+ = (A_1^T A_1)^{-1} A_1^T = A_1^T = [1 \quad 0]$$

$$P_1 = I - A_1 A_1^+ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix}[1 \quad 0] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$p_1 = -P_1\nabla f(x_1) = -\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -4 \\ -6 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

算例(投影梯度法)

$$x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad p_1 = \begin{bmatrix} 0 \\ 6 \end{bmatrix} \quad A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix} \quad b = (-2,-5,0,0)^T$$

$a_1^T p_1 = -6, a_2^T p_2 = -30,$

$a_1^T x_1 - b_1 = 2, a_2^T x_2 - b_2 = 5.$ $\quad \alpha_{\max} = \min\{\frac{2}{6}, \frac{5}{30}\} = \frac{1}{6}$

作线性搜索

$$\min f(x_1 + \alpha p_1) = 72\alpha^2 - 36\alpha, 0 < \alpha \le \alpha_{\max} = 1/6$$

得 $\alpha_1 = \frac{1}{6}$, $\quad x_2 = x_1 + \alpha_1 p_1 = (0,1)^T$.

$\alpha_{\max}$的手工计算

$$x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad p_1 = \begin{bmatrix} 0 \\ 6 \end{bmatrix} \quad A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix} \quad b = (-2,-5,0,0)^T$$

在手算时,可以解下面的不等式来确定$\alpha_{\max}$.

$$\begin{bmatrix} -1 & -1 \\ -1 & -5 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \alpha\begin{bmatrix} -1 & -1 \\ -1 & -5 \end{bmatrix}\begin{bmatrix} 0 \\ 6 \end{bmatrix} \ge \begin{bmatrix} -2 \\ -5 \end{bmatrix}$$

即 $\begin{bmatrix} -6\alpha \\ -30\alpha \end{bmatrix} \ge \begin{bmatrix} -2 \\ -5 \end{bmatrix}$

得 $\alpha \le \frac{1}{6}$. $\quad$ 因此 $\alpha_{\max} = \frac{1}{6}$.

算例(投影梯度法)

$$x_2 = (0,1)^T \quad A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix}$$

第二次迭代 $\nabla f(x_2) = (-6,-2)^T, I_2 = \{2,3\}$

$A_2 = \begin{bmatrix} -1 & 1 \\ -5 & 0 \end{bmatrix}$ 可逆. $P_2 = O, P_2\nabla f(x_2) = 0.$

乘子向量 $\lambda = A_2^+\nabla f(x_2) = A_2^{-1}\nabla f(x_2)$

$$= \begin{bmatrix} 0 & -1/5 \\ 1 & -1/5 \end{bmatrix}\begin{bmatrix} -6 \\ -2 \end{bmatrix} = \begin{bmatrix} 2/5 \\ -28/5 \end{bmatrix}$$

取λ_2=-28/5<0, 从A_2取中去掉第二列,得到

算例(投影梯度法)

$$x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix} \quad A_2 = \begin{bmatrix} -1 & 1 \\ -5 & 0 \end{bmatrix} \quad \lambda = \begin{bmatrix} 2/5 \\ -28/5 \end{bmatrix}$$

取λ_2= -28/5<0,从A_2取中去掉第二列,得到

$$A_1 = \begin{bmatrix} -1 \\ -5 \end{bmatrix} \quad A_1^+ = (A_1^T A_1)^{-1} A_1^T \quad = [-\frac{1}{26} \quad -\frac{5}{26}]$$

$$P_1 = I - A_1 A_1^+ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -1 \\ -5 \end{bmatrix}\begin{bmatrix} -\frac{1}{26} & -\frac{5}{26} \end{bmatrix} = \begin{bmatrix} 25/26 & -5/26 \\ -5/26 & 1/26 \end{bmatrix}$$

$$p_1 = -P_1\nabla f(x_1) = -\begin{bmatrix} 25/26 & -5/26 \\ -5/26 & 1/26 \end{bmatrix}\begin{bmatrix} -6 \\ -2 \end{bmatrix} = \begin{bmatrix} 70/13 \\ -14/13 \end{bmatrix}$$

不妨取p_2=(5,-1)^T.

算例(投影梯度法)

$$x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad p_2 = \begin{bmatrix} 5 \\ -1 \end{bmatrix} \quad A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix} \quad b = (-2,-5,0,0)^T$$

$a_1^T p_2 = -4, a_4^T p_2 = -1,$

$a_1^T x_2 - b_1 = 1, a_4^T x_2 - b_4 = 1.$ $\quad \alpha_{\max} = \min\{\frac{1}{4}, \frac{1}{1}\} = \frac{1}{4}$

作线性搜索

$$\min f(x_2 + \alpha p_2) = 62\alpha^2 - 28\alpha - 4, 0 < \alpha \le \alpha_{\max} = 1/4$$

得 $\alpha_2 = \frac{7}{31}$, $\quad x_3 = x_2 + \alpha_2 p_2 = (35/31, 24/31)^T$.

算例(投影梯度法)

$x_3=\begin{bmatrix}35/31\\24/31\end{bmatrix}$ $A=\begin{bmatrix}-1&-1&1&0\\-1&-5&0&1\end{bmatrix}$ $b=(-2,-5,0,0)^T$

第三次迭代 $\nabla f(x_2)=(-32/31,-160/31)^T, I_2=\{2\}$

$A_1=\begin{bmatrix}-1\\-5\end{bmatrix}$ $A_1^+=[-\frac{1}{26}\quad -\frac{5}{26}]$ $P_1=\begin{bmatrix}25/26&-5/26\\-5/26&1/26\end{bmatrix}$

$$P_1\nabla f(x_3)=-\begin{bmatrix}25/26&-5/26\\-5/26&1/26\end{bmatrix}\begin{bmatrix}-32/31\\-160/31\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

$$\lambda=A_1^+\nabla f(x_3)=\begin{bmatrix}-1/26&-5/26\end{bmatrix}\begin{bmatrix}-32/31\\-160/31\end{bmatrix}=32/31>0$$

因此x_3=(35/31,24/31)^{T}为KT点.

本问题是凸规划问题,x_3是最优解.

§4.4 约束变尺度法

4.4.1 二次规划

目标函数是二次函数,约束函数为线性函数的规划问题称为二次规划,其一般形式为

$$\begin{aligned}&\min f(x)=\frac{1}{2}x^TGx+C^Tx\\(QP)\ s.t.\quad &c_i(x)=a_i^Tx-b_i=0, i\in E=\{1,\cdots,l\}\\&c_i(x)=a_i^Tx-b_i\ge 0, i\in I=\{l+1,\cdots,m\}\end{aligned}$$

其中G为n阶对称矩阵.当G正定时,(QP)为严格凸二次规划.

二次规划KT条件

定理4.4.1 点x^*是严格凸二次规划(QP)的严格整体最优解的充要条件是x^*满足KT条件,即存在乘子向量$\lambda^*=(\lambda_1^*,\cdots,\lambda_m^*)$,使得

$$\begin{aligned}&Gx^*+C-\sum_{i\in E}\lambda_i^*a_i-\sum_{i\in I}\lambda_i^*a_i=0\\&a_i^Tx^*-b_i=0, i\in E\\&a_i^Tx^*-b_i\ge 0, i\in I\\&\lambda_i^*\ge 0,\qquad i\in I\\&\lambda_i^*=0,\qquad i\in I\setminus I^*\end{aligned}$$

其中I^*是x^*处的有效集.

等式约束二次规划

对于仅含等式约束的严格凸二次规划

$$\begin{aligned}&\min f(x)=\frac{1}{2}x^TGx+C^Tx\\s.t.\quad &c_i(x)=a_i^Tx-b_i=0, i\in E=\{1,\cdots,l\}\end{aligned}$$

其中$A=(a_1,\cdots,a_l)$的秩为l.

显然,x是上述问题的解的充要条件是

$$\begin{aligned}&Gx+C-\sum_{i\in E}\lambda_ia_i=0,\\&a_i^Tx-b_i=0, i\in E.\end{aligned}$$

等式约束二次规划

$$\begin{aligned}&Gx+C-\sum_{i\in E}\lambda_ia_i=0,\\&a_i^Tx-b_i=0, i\in E.\end{aligned}$$

令$b=(b_1,\cdots,b_l)^T$, $\lambda=(\lambda_1,\cdots,\lambda_l)^T$,上式可以记为

$$\begin{pmatrix}G&-A\\A^T&0\end{pmatrix}\begin{pmatrix}x\\\lambda\end{pmatrix}=\begin{pmatrix}-C\\b\end{pmatrix}$$

由于G正定,A列满秩,该线性方程组有唯一解.

等式约束二次规划(算例)

例4.4.1 求解严格凸二次规划

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

(QP)s.t. $x_1+2x_2-x_3-4=0$

$x_1-x_2+x_3+2=0$

解 (QP)的最优解及乘子满足

$$\begin{bmatrix} 2 & 0 & 0 & -1 & -1 \\ 0 & 2 & 0 & -2 & 1 \\ 0 & 0 & 2 & 1 & -1 \\ 1 & 2 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda_1 \\ \lambda_2 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \\ 4 \\ -2 \end{bmatrix}$$

该方程组有唯一解

$(\frac{2}{7},\frac{10}{7},-\frac{6}{7},\frac{8}{7},-\frac{4}{7})^T$.

因此最优解

$x^*=(2/7,10/7,-6/7)^T$,

最优乘子向量

$\lambda^*=(8/7,-4/7)^T$.

严格凸二次规划的有效集方法

$$\min f(x)=\frac{1}{2}x^TGx+C^Tx$$

$$(QP)\ s.t.\ \ c_i(x)=a_i^Tx-b_i=0, i\in E=\{1,\cdots,l\}$$

$$c_i(x)=a_i^Tx-b_i\ge 0, i\in E=\{l+1,\cdots,m\} \qquad (4.82)$$

定理4.4.2 设x^*是上面问题的最优解,且在x^*处的有效集为I^*,则x^*也是下列问题的最优解

$$\min f(x)=\frac{1}{2}x^TGx+C^Tx$$

$$s.t.\ \ c_i(x)=a_i^Tx-b_i=0, i\in I^* \qquad (4.86)$$

有效集方法

证明:由于x^*是(4.82)问题的最优解,存在乘子向量λ^*满足

$$Gx^*+C-\sum_{i\in E}\lambda_i^*a_i-\sum_{i\in I}\lambda_i^*a_i=0$$

$$a_i^Tx^*-b_i\ge 0, i\in I$$

$$\lambda_i^*=0, \quad i\in I\setminus I^*$$

显然有 $Gx^*+C-\sum_{i\in I^*}\lambda_i^*a_i=0$

这是问题(4.86)的最优解的充要条件.

有效集方法

例:解二次规划

$\min f(x)=(x_1-1)^2+(x_1-2.5)^2$

s.t. $x_1-2x_2+2\geqslant 0$

$-x_1-2x_2+6\geqslant 0$

$-x_1+2x_2+2\geqslant 0$

$x_1\geqslant 0$

$x_2\geqslant 0$

解: 取初始可行点$x_0=(2,0)^T$.有效集为$I_0=\{3,5\}$.

有效集方法

我们首先来求解等式约束二次规划

$\min f(x)=(x_1-1)^2+(x_1-2.5)^2$

s.t. $-x_1+2x_2+2=0$

$x_2=0$

显然求得的最优解$x_1=x_0$.

我们来判断x_1是否是原来的二次规划的KT点.

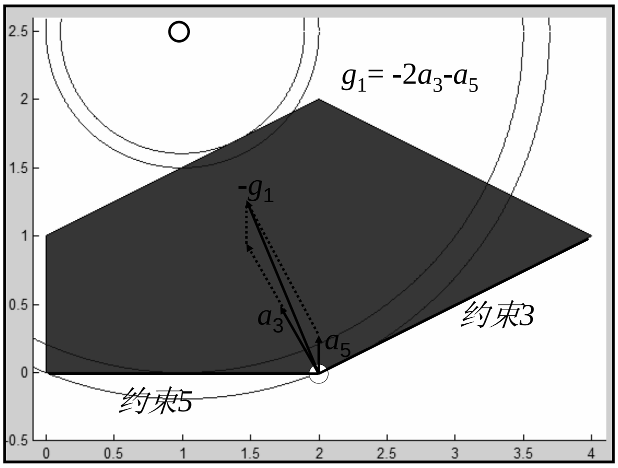
有效集方法

$g_1=(2,-5)^T, a_3=(-1,2)^T, a_5=(0,1)^T$,

令$g_1=\lambda_3a_3+\lambda_5a_5$,则有$\lambda_3=-2$, $\lambda_5=-1$.

因此$x_1=(2,0)^T$不是KT点.

事实上, $a_3=(-1,2)^T, a_5=(0,1)^T$都是可行的下降方向.



有效集方法

去掉约束i_1=3,得有效集I_1={5},求解下面的二次规划

min $f(x)=(x_1-1)^2+(x_1-2.5)^2$

s.t. x_2 =0

令$x=x_k+d, g_k=Gx_k+C$,

则$f(x)=x^TGx/2+C^Tx=(x_k+d)^TG(x_k+d)/2+C^T(x_k+d)$

$=d^TGd/2+x_k^TGd+x_k^TGx_k/2+C^Tx_k+C^Td$

$=d^TGd/2+g_k^Td+x_k^TGx_k/2+C^Td$

有效集方法

约束$a_j^Tx \geqslant b_j$等价于$a_j^T(x_k+d) \geqslant b_j$,即$a_j^Td=0$.

一般的,若已知可行解x_k,有效集I_k,我们求解二次规划

$$\min f(x)=\frac{1}{2}x^TGx+C^Tx$$
$$s.t.\quad a_j^Tx \ge b_j (j \in I_k)$$

它等价于求解

$$\min f(x)=\frac{1}{2}d^TGd+g_k^Td$$
$$s.t.\quad a_j^Td=0 (j \in I_k)$$

有效集方法

为了求得最优得x,只要求得对应的最优的d,因此,我们只要求解下面的二次规划

min $d^TGd/2+g_1^Td$

$s.t.$ x_2=0

其中

$$g_1=\begin{bmatrix}2\\-5\end{bmatrix}, G=\begin{bmatrix}2&0\\0&2\end{bmatrix}$$

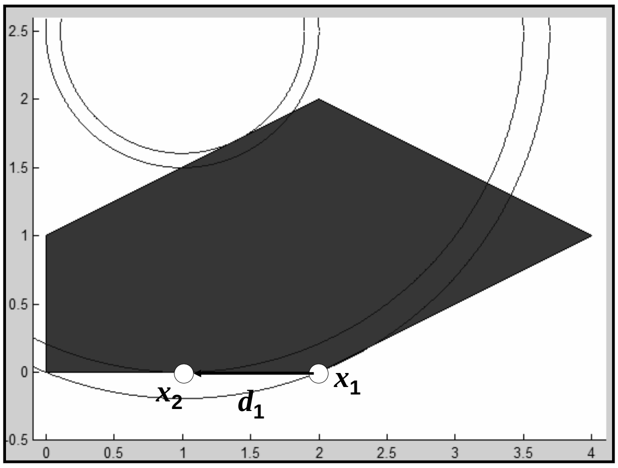
求解:

$$\begin{bmatrix}2&0&0\\0&2&-1\\0&1&0\end{bmatrix}\begin{bmatrix}d^{(1)}\\d^{(2)}\\\lambda\end{bmatrix}=\begin{bmatrix}-2\\5\\0\end{bmatrix}$$

$$(d^{(1)},d^{(2)},\lambda)^T=(-1,0,-5)^T$$

因此d_1=(-1,0)^T

$x_2=x_1+d_1=(1,0)^T$.



有效集方法

因此d_1=(-1,0)^T

$x_2=x_1+d_1=(1,0)^T$.有效集$I_2=I_1$={5}.

在上面求得的结果中x_1+d_1对于原来的二次规划是可行解.

在有些情况下,可能出现不可行解.

不过可以确定d_1一定是可行的下降方向.

有效集方法

定理若严格凸二次规划 $\min f(x)=\frac{1}{2}d^TGd+g_k^Td$

$s.t.\quad a_j^Td=0(j\in I_k)$

最优解为$d_k=0$,且Lagrange乘子有负分量$(\lambda_k)_s<0$.

设p_k为下面二次规划的最优解

$$\min f(x)=\frac{1}{2}d^TGd+g_k^Td$$
$$s.t.\quad a_j^Td=0(j\in I_k, j\neq s)$$

则有$g_k^Tp_k<0$, $a_sp_k>0$.

作业:证明此定理.

有效集方法

点x_2对应的新二次规划为

$\min d^TGd/2+g_2^Td$

$s.t.\quad x_2=0$

其中$g_2=(0,-5)^T$.

显然其最优解为$d_2=(0,0)^T$,Lagrange乘子向量为(-5).

取$x_3=x_2+d_2=x_2=(1,0)^T$,有效集$I_3=I_2\backslash\{5\}=\phi$.

有效集方法

点x_3对应二次规划为(无约束)

$\min d^TGd/2+g_3^Td$

求得$d_3=(0,2.5)^T$.

此时x_3+d_3不再可行.

我们求α_3使得$x_3+\alpha_3d_3$可行,且α_3尽可能大.

于是$a_i^T(x_3+\alpha_3d_3)\geqslant b_i$.

因此有

$1-5\alpha_3+2\geqslant 0$

$-1-5\alpha_3+6\geqslant 0$

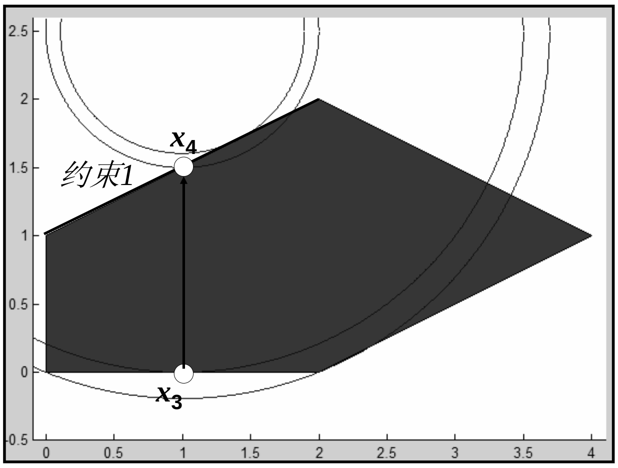
$-1+5\alpha_3+2\geqslant 0$

$1\geqslant 0$

$2.5\alpha_3\geqslant 0$

解得$\alpha_3=0.6$.

(第一个约束有效)



有效集方法

$x_4=x_3+\alpha_3d_3=(1,1.5)^T$,$I_4=\{1\}$,$g_4=(0,-2)^T$.

求解对应的二次规划

$\min d^TGd/2+g_4^Td$

$s.t.\quad x_1-2x_2=0$

解得$d_4=(0.4,0.2)^T$.

$x_5=x_4+d_4=(1.4,1.7)^T$(可行),

$g_5=(0.8,-1.6)^T$.

有效集方法

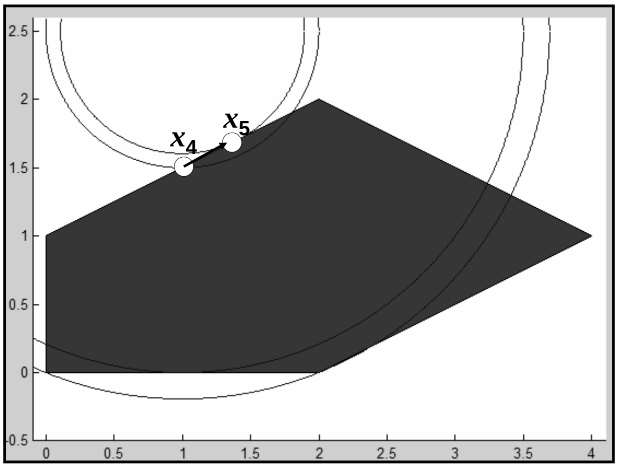
x_5对应的二次规划为

$\min d^TGd/2+g_5^Td$

$s.t.\quad x_1-2x_2=0$

其最优解为$d_5=(0,0)^T$.

Lagrange乘子向量为(0.8)>0,算法终止,最优解为$x^*=x_5=(1.4,1.7)^T$.



有效集方法

设x_k是问题(4.82)的一个可行解,相应的有效集为I_k,A_k为$(a_1,\cdots,a_m)$中对应于I_k的子矩阵,求解

$$\min q(d)=\frac{1}{2}d^TGd+g_k^Td \qquad (4.88)$$
$$s.t.\quad A_k^Td=0$$

若(4.88)的最优解为d_k=0,则x_k为(4.87)的最优解.若此时Lagrange乘子向量非负,则x_k是问题(4.82)的KT点,从而是最优解.

有效集方法

若(4.88)的最优解$d_k\neq 0$,并且x_k+d_k为(4.82)的可行解,则令$x_{k+1}=x_k+d_k$;

否则,以d_k为方向进行线性搜索,求得(4.82)最好的可行点.

由于$f(x)$为凸二次函数,这一点必在边界上达到.设步长为α_k, $x_{k+1}=x_k+\alpha_k d_k$.

根据约束条件有 $a_i^T(x_k+\alpha_k d_k)\geq 0, i\notin I_k$

有效集方法

$$a_i^T(x_k+\alpha_k d_k)\geq 0, i\notin I_k$$

即 $\alpha_k a_i^T d_k\geq b_i-a_i^Tx_k, i\notin I_k$

由于$b_i-a_i^Tx_k\leqslant 0$,因此若$a_i^Td_k\geqslant 0$,上面的式子恒成立.

我们只要考虑$a_i^Td_k<0$的情况,

$$\alpha_k=\bar{\alpha}_k=\min_{i\notin I_k}\left\{\frac{b_i-a_i^Tx_k}{a_i^Td_k}\middle| a_i^Td_k<0\right\}=\frac{b_t-a_t^Tx_k}{a_t^Td_k}$$

有效集方法

综合上面的两种情况,取 $\alpha_k=\min\{1,\bar{\alpha}_k\}$

如果$\alpha_k<1$,即 $\alpha_k=\bar{\alpha}_k$ 则有某个$t\notin I_k$ 使得

$$\alpha_k=\bar{\alpha}_k=\frac{b_t-a_t^Tx_k}{a_t^Td_k}$$

此时有$a_t^Tx_{k+1}=a_t^Tx_k+\alpha_k a_t^Td_k=b_t$,

因此在x_{k+1}处增加了一个有效约束$I_{k+1}=I_k\cup\{t\}$

如果α_k=1,则有效约束集不变$I_{k+1}=I_k$.

有效集方法

如果(4.88)的最优解d_k=0,且Lagrange乘子向量有负分量,设$(\lambda_k)_s<0$,有定理4.1.1可知x_k不是(4.82)的最优解.

此时我们去掉第s个约束求解下面的二次规划

$$\min q(d)=\frac{1}{2}d^TGd+g_k^Td$$
$$s.t.\quad a_j^Td=0(j\in I_k, j\neq s)$$

其最优解是原二次规划的可行下降方向.类似于前面的情况定出步长及下一迭代点.

有效集方法

总结:给定了二次规划可行解x_k,其有效集为I_k,
d_k为下面的二次规划的解

$$\min f(x)=\frac{1}{2}d^TGd+g_k^Td$$
$$s.t.\quad a_j^Td=0(j\in I_k)$$

(1)d_k=0,且Lagrange乘子非负,则x_k为最优解;

(2)d_k=0,且Lagrange乘子有负分量,设为$(\lambda_k)_s<0$,
则令$x_{k+1}=x_k$,$I_{k+1}=I_k\backslash\{s\}$.

有效集方法

(3)$d_k\neq 0$,取合适的α_k使得$x_k+\alpha_k d_k$为原规划可行解,令$x_{k+1}=x_k+\alpha_k d_k$,并根据的取值情况判断是否改变有效集.

若α_k=1,则有效集不变,$I_{k+1}=I_k$.

若α_k<1,则在I_k.中添加指标得到新的有效集I_{k+1}.