

最优化方法

南京邮电大学理学院杨振华制作
yangzhenhua@njupt.edu.cn

目录

- 第一章 最优化问题概述
- 第二章 线性规划
- 第三章 无约束最优化方法
- 第四章 约束最优化方法

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

2

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

4

例 1.1.1 运输问题

设有 m 个水泥厂 A_1, A_2, \dots, A_m , 年产量各为 a_1, a_2, \dots, a_m 吨. 有 k 个城市 B_1, B_2, \dots, B_k , 用这些水泥厂生产的水泥, 年需求量 b_1, b_2, \dots, 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:

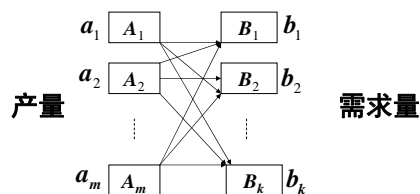
$$\sum_{i=1}^m a_i = \sum_{j=1}^k b_j$$

试设计一个调运方案, 在满足需要的同时使总运费最省.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

5

由题意可画出如下的运输费用图:



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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

6

数学模型:

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

$$s.t. \quad \sum_{j=1}^k x_{ij} = a_i (i=1, 2, \dots, m)$$

$$\sum_{i=1}^m x_{ij} = b_j (j=1, 2, \dots, k)$$

$$x_{ij} \geq 0 (i=1, 2, \dots, m, j=1, 2, \dots, k)$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

7

例1.1.2 生产计划问题

设某工厂有 m 种资源 B_1, B_2, \dots, B_m , 数量各为:
 b_1, b_2, \dots, b_m , 用这些资源产 n 种产品 A_1, A_2, \dots, 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 .

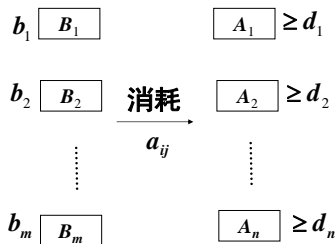
问如何安排生产计划, 才能既完成合同, 又使该厂总收入最多?

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

8

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:



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

9

数学模型

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

$$s.t. \quad \sum_{j=1}^n a_{ij} x_j \leq b_i (i=1, 2, \dots, m)$$

$$x_j \geq d_j (j=1, 2, \dots, n)$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

10

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

11

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:

$$\min S = \sum_{i=1}^4 \sum_{j=1}^4 c_{ij} x_{ij}$$

$$s.t. \quad \sum_{j=1}^4 x_{ij} = 1, i=1, 2, 3, 4$$

$$\sum_{i=1}^4 x_{ij} = 1, j=1, 2, 3, 4$$

$$x_{ij} \in \{0, 1\}, i, j=1, 2, 3, 4$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

12

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:

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

13

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

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

为基本函数.

一般选取 $1, x, x^2, \dots, x^n$ 为基本函数,即以

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

作为近似表达式.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

14

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

15

最优化问题

最优化问题的一般形式为:

$$\min f(x) \quad (1.1) \text{ (目标函数)}$$

$$s.t. \quad h_i(x) = 0, i = 1, 2, \dots, m \quad (1.2) \text{ (等式约束)}$$

$$g_j(x) \geq 0, j = 1, 2, \dots, p \quad (1.3) \text{ (不等式约束)}$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

16

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:

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

若 $h_i(x), g_j(x)$ 为连续函数,则 D 为闭集.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

17

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 $f(x^*) \leq f(x)$,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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

18

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 $f(x^*) \leq f(x)$, 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

19

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

20

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) $\|x\| \geq 0$ ($\|x\| = 0$ 当且仅当 $x = 0$) (正定性);
 - (2) $\|\alpha x\| = |\alpha| \cdot \|x\|$ (对于任意 $\alpha \in R$);
 - (3) $\|x + y\| \leq \|x\| + \|y\|$ (三角不等式);
- 则称 $\|x\|$ 为 R^n 上的一个向量范数.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

21

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) $\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}}$

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

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

22

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

23

最优化问题的分类

根据数学模型中是否有约束函数分为有约束的最优化问题 and 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

24

§ 1.2 最优化问题的一般算法

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

25

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:

$$x_1, x_2, \dots, x_k, \dots,$$

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

26

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 $f(x_{k+1}) \leq f(x_k)$.其中 α_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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

27

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 $g_k^T p_k \leq 0$ (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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

28

可行方向

定义1.2.2(可行方向) 已知区域 D , $x_k \in D$,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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

29

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)

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

30

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

31

收敛速度

定义1.2.3 设序列 $\{x_k\}$ 收敛于 x^* ,而且

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

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

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

则称 $\{x_k\}$ 为 p 阶收敛.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

32

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

33

§ 1.3 二维最优化问题的几何解释

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

34

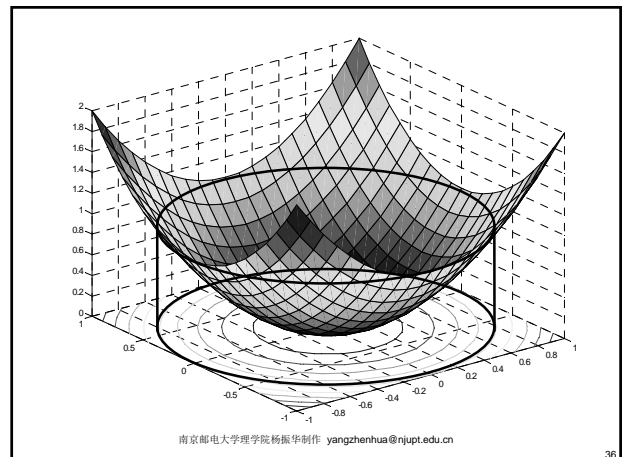
理论分析

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 $O-x_1x_2z$ 中,平面 $z=c$ 与曲面 $z=f(x_1, x_2)$ 的交线在 $O-x_1x_2$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

35



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

36

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

37

例1.3.1

$$\min (x_1 - 1)^2 + (x_2 - 2)^2$$

$$s.t. \quad x_1^2 + x_2 - \frac{7}{4} \leq 0$$

$$x_1, x_2 \geq 0$$

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

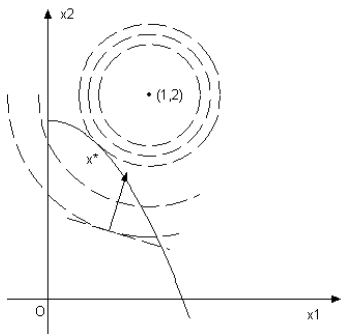
$$\nabla f(x_1, x_2) = (2(x_1 - 1), 2(x_2 - 2))^T$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

38

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 $1/2$.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

39

例1.3.2

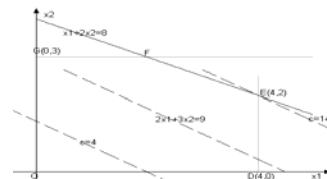
$$\max 2x_1 + 3x_2$$

$$s.t. \quad x_1 + 2x_2 \leq 8$$

$$x_1 \leq 4$$

$$x_2 \leq 3$$

$$x_1, x_2 \geq 0$$



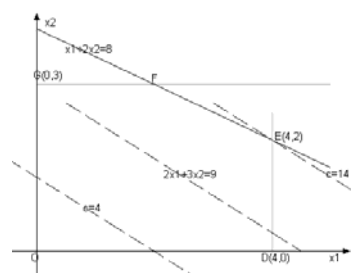
解 首先画出可行域 D 的图形. D 为凸多边形 $ODEFGO$. 再以 c 为参数画出目标函数的等值线 $2x_1 + 3x_2 = c$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

40

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 就是最优解, 最优值为 14 .



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

41

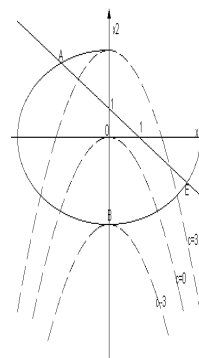
$$\min x_1^2 + x_2$$

$$s.t. \quad -x_1 - x_2 + 1 \geq 0$$

$$x_1^2 + x_2^2 - 9 = 0$$

例1.3.3

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 .



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

42

§ 1.4 一维搜索

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

43

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

$$\min_{a \leq x \leq b} f(x)$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

44

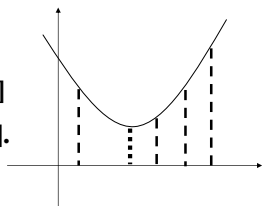
1.4.1 黄金分割法

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



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

45

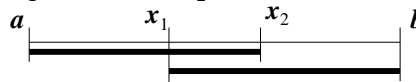
黄金分割法

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 $x_2 = a + (1 - p)(b - a)$.



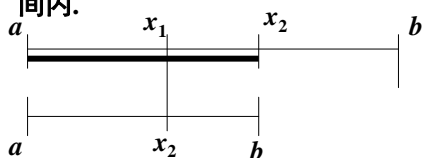
南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

46

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.



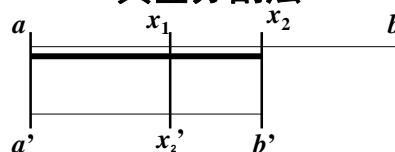
南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

47

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

48

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

49

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step 3.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

50

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 $x^*=(0.443+0.665)/2=0.554$						

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

51

0.618法和Fibonacci之间的关系

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

0.618法为Fibonacci法的极限形式.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

52

0.618法和Fibonacci之间的关系

迭代步骤数的比较

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

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

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

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

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

黄金分割法至多比Fibonacci法多一步

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

53

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

因为我们关心的是:

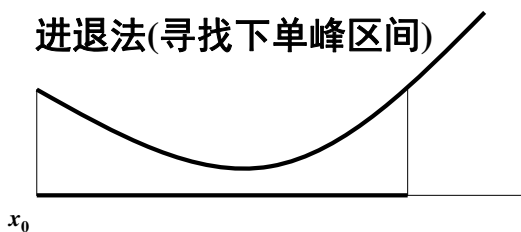
$$\min_{a > 0} f(x_k + \alpha p_k) = \min_{\alpha > 0} \phi(\alpha)$$

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

54

进退法(寻找下单峰区间)



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

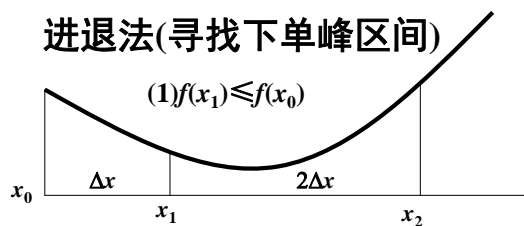
(1) $f(x_1)\leq f(x_0)$

x_1 对应着图上用红线标出的一部分

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

55

进退法(寻找下单峰区间)



(1) $f(x_1)\leq f(x_0)$

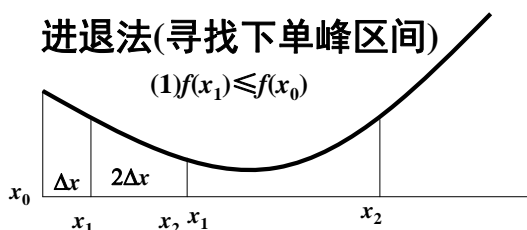
此时 x_1 取值小,我们加大步长向右搜索,取
 $\Delta x=2\Delta x$, $x_2=x_1+\Delta x$

若 $f(x_1)\leq f(x_2)$,则我们要找区间即为 $[x_0, x_2]$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

56

进退法(寻找下单峰区间)



(1) $f(x_1)\leq f(x_0)$

若 $f(x_1)>f(x_2)$,则我们所取的步长偏小.

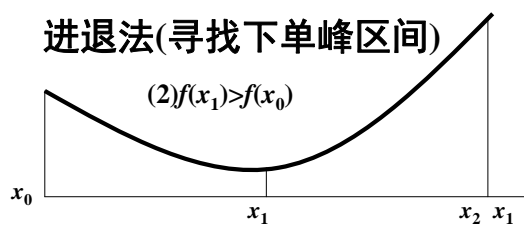
令 $x_1=x_2$, $\Delta x=2\Delta x$, $x_2=x_1+\Delta x$

继续往下判断,直到满足 $f(x_1)\leq f(x_2)$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

57

进退法(寻找下单峰区间)



(2) $f(x_1)>f(x_0)$

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 $f(x_1)\leq f(x_0)$,则我们要找区间即为 $[x_0, x_2]$

否则继续缩小区间,直到满足 $f(x_1)\leq f(x_0)$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

58

1.4.2 二分法

若 $f(x)$ 的导数存在且容易计算,则线性搜索的速度可以得到提高,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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

59

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 $y=\phi(x)$,即作一条抛物线,则可推导出:

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

60

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

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

61

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

62

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} 。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

63

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”。

下面令 $\phi(\alpha) = f(x_k + \alpha p_k)$,我们讨论 α_k 满足的条件。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

64

不精确一维搜索

对于一元函数 $\phi(\alpha)$,精确一维搜索的条件为 $\phi'(\alpha_k) = 0$ 。

不精确一维搜索的条件 $\phi'(\alpha_k) \approx 0$,或

$|\phi'(\alpha_k)| \leq \sigma$ 。

实际计算中上式不好控制,一般的方法是

$|\phi'(\alpha_k) / \phi'(0)| \leq \sigma$ 。

σ 的选取: 不宜太大——否则下降不够充分;

不宜太小——否则“太精确”。

适合的范围: 比1稍小一些。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

65

不精确一维搜索

例: 函数 $\phi(\alpha) = (\alpha - 1)^2$ 。

$\phi'(\alpha) = 2(\alpha - 1)$,

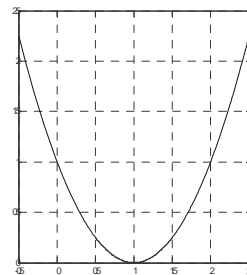
$\phi'(0) = -2$ 。

取 $\sigma = 0.5$,则控制条件为

$|\phi'(\alpha_k)| \leq \sigma |\phi'(0)| = 1$,

即 $|2(\alpha_k - 1)| \leq 1$,

$1/2 \leq \alpha_k \leq 3/2$ 。



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

66

不精确一维搜索

上述条件是不够的,甚至不能保证 $\phi(\alpha_k) < \phi(0)$.

例如: 对于一分段函数

$\alpha \leq 3/2$ 时, $\phi(\alpha) = (\alpha-1)^2$;

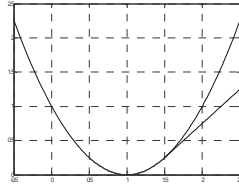
$\alpha > 3/2$ 时, $\phi(\alpha) = \alpha - 5/4$.

易见 $\phi(\alpha)$ 是一个连续可导函数,且 $\alpha > 3/2$ 时导数为1.

$|\phi'(\alpha_k)| \leq \sigma$ $|\phi'(0)| = 1$

确定的范围是

$1/2 \leq \alpha_k$, 不能保证函数值下降.

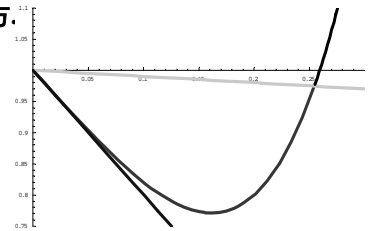


南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

67

不精确一维搜索

为此,我们在直线 $y=0$ 以及函数 $\phi(\alpha)$ 在 $\alpha=0$ 处的切线之间画一条直线,将搜索范围首先限制在这条直线下方.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

68

不精确一维搜索

上述直线的方程为

$y - \phi(0) = \mu \phi'(0) \alpha$.

因此搜索的条件为

$\phi(\alpha_k) - \phi(0) \leq \mu \phi'(0) \alpha_k$.

作业: 对 $\phi(\alpha) = f(x_k + \alpha p_k)$, 证明

$$\phi'(\alpha) = \nabla f(x_k + \alpha p_k)^T p_k$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

69

不精确一维搜索

$$\phi'(\alpha) = \nabla f(x_k + \alpha p_k)^T p_k$$

$$\implies \phi'(0) = \nabla f(x_k)^T p_k = g_k^T p_k$$

$$\phi(\alpha_k) - \phi(0) \leq \mu \phi'(0) \alpha \implies$$

$$f(x_k) - f(x_k + \alpha_k p_k) \geq -\mu \alpha_k g_k^T p_k$$

$$|\phi'(\alpha_k) / \phi'(0)| \leq \sigma \implies$$

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

70

不精确一维搜索的Wolfe原则

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

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

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

或用下面更强的条件代替(1.7)式:

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

71

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 $\mu \in (0, 1/2)$, $\sigma \in (\mu, 1)$, 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)).

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

72

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

step 2 $x_{k+1} = x_k + \alpha p_k$, 计算 f_{k+1}, g_{k+1} ;

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

73

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 $b = \alpha, \alpha = \frac{a+\alpha}{2}$ 转step 2

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

74

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

75

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 $\alpha_k = 0.125$

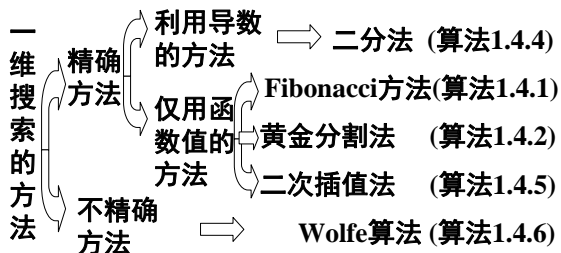
$x_{k+1} = x_k + \alpha_k p_k = (0.125, 1)^T$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

76

总结

求初始区间的方法 \Rightarrow 进退法 (算法1.4.3)



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

77

第二章 线性规划



§ 2.1 凸集与凸函数

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

79

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 $\alpha, 0 \leq \alpha \leq 1$,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 $H^+ = \{x \in R^n | a_1x_1 + a_2x_2 + \cdots + a_nx_n \geq b\}$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

80

凸集的例

例2.1.2 超球 $\|x\| \leq r$ 为凸集

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

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

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

即点 $\alpha x + (1-\alpha)y$ 属于超球,所以超球为凸集.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

81

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

82

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;

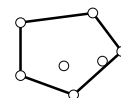
$D_1 + D_2 = R^2$ 是凸集

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

83

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 $\lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1$,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.

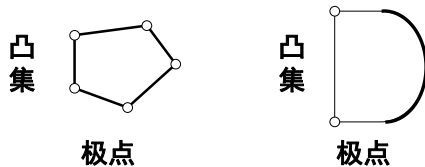
南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

84

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.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

85

极点

例 $D = \{x \in R^n \mid \|x\| \leq a\} (a > 0)$, 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

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

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

$$\leq a^2$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

86

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
 $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$
则称函数 $f(x)$ 为凸集 D 上的凸函数.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

87

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

88

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

$$\begin{aligned} & f(\alpha x + (1 - \alpha)y) - (\alpha f(x) + (1 - \alpha)f(y)) \\ &= (\alpha x + (1 - \alpha)y - 1)^2 - \alpha(x - 1)^2 - (1 - \alpha)(y - 1)^2 \\ &= -\alpha(1 - \alpha)(x - y)^2 < 0 \end{aligned}$$

因此 $f(x)$ 在 $(-\infty, +\infty)$ 上是严格凸函数.

例2.1.4 线性函数 $f(x) = c^T x = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$
既是 R^n 上凸函数也是 R^n 上凹函数.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

89

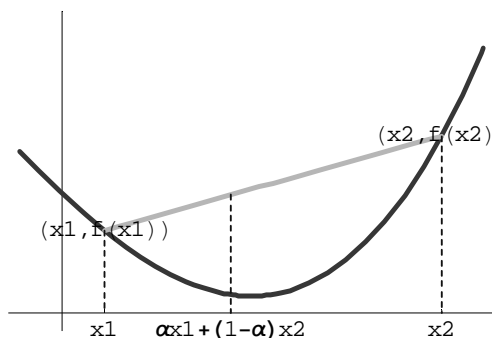
凸函数的几何性质

对一元函数 $f(x)$, 在几何上 $\alpha f(x_1) + (1 - \alpha)f(x_2)$ ($0 \leq \alpha \leq 1$) 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

90

凸函数的几何性质



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

91

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 $\text{epi}(f)$ 为凸集,其中 $\text{epi}(f) = \{(x, y) | x \in D, y \in R, y \geq f(x)\}$.

证明:作业

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

92

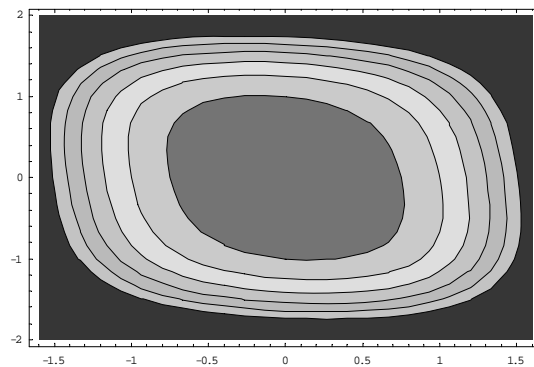
凸函数的性质

- (i) 设 $f(x)$ 是凸集 $D \subset R^n$ 上的凸函数,实数 $k \geq 0$,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 $\lambda, \mu \geq 0$,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 $S(f, \beta) = \{x | x \in D, f(x) \leq \beta\}$.
- 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

93

凸函数的性质



94

凸函数的判断

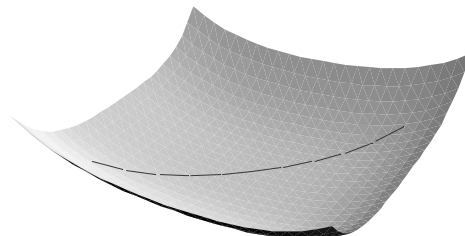
- 定理2.1.1 设 $f(x)$ 定义在凸集 $D \subset R^n$ 上, $x, y \in D$.令 $\phi(t) = f(tx + (1-t)y)$, $t \in [0, 1]$,则
- (i) $f(x)$ 是凸集 D 上的凸函数的充要条件是对任意的 $x \in D$,一元函数 $\phi(t)$ 为 $[0, 1]$ 上的凸函数.
 - (ii) $f(x)$ 是凸集 D 上的凸函数的充要条件是对任意的 $x, y \in D (x \neq y)$,一元函数 $\phi(t)$ 为 $[0, 1]$ 上的严格凸函数.

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

95

凸函数的判断



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

96

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

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

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

97

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

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

(ii) 若在 D 内 $G(x)$ 正定, 则 $f(x)$ 在 D 内是严格凸函数.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

98

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

99

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 $f(tx^* + (1-t)x_0) \leq tf(x^*) + (1-t)f(x_0)$ 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 $f(tx^* + (1-t)y^*) < tf(x^*) + (1-t)f(y^*) = f(x^*)$ 矛盾.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

100

§ 2.2 线性规划的标准型与基本概念

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

101

线性规划的一般形式

$$\min(\max) c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

$$\text{s.t. } a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \geq (\text{或} \leq, =) b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \geq (\text{或} \leq, =) b_2$$

... ..

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \geq (\text{或} \leq, =) b_m$$

$$x_1, x_2, \dots, x_n \geq 0$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

102

线性规划的标准型

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

103

矩阵-向量形式的标准型

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

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

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{array}{l} c: \text{价格向量} \\ A: \text{约束矩阵} \\ b: \text{右端向量} \end{array}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

104

矩阵-向量形式的标准型

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

$$\begin{aligned} \min & c^Tx \\ (LP) \text{ s.t. } & \sum_{j=1}^n x_j p_j = b \\ & x_j \geq 0, j=1,2,\dots,n \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

105

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 .

$D=\emptyset$ 时, 称线性规划无可行解;

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

106

可行域为凸集

定理2.2.1 线性规划问题对任意的 $\alpha \in [0,1]$, 设

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

的可行域 D 为凸集.

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

$$\begin{aligned} & \text{设 } z=\alpha x+(1-\alpha)y, \text{ 则 } z\geq 0, \text{ 且} \\ & Az=A(\alpha x+(1-\alpha)y) \\ & =\alpha Ax+(1-\alpha)Ay \\ & =\alpha b+(1-\alpha)b \\ & =b \\ & \text{因此 } z \in D \\ & D \text{ 为凸集.} \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

107

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 $\sum_{j=1}^n a_{ij}x_j \leq b_i$
- (iv) 变量 x_j 无非负约束 引入非负变量 $x_j' \geq 0, x_j'' \geq 0$, 令 $x_j = x_j' - x_j''$.

$$\begin{aligned} \text{令 } x_{n+i} &= b_i - \sum_{j=1}^n a_{ij}x_j \\ \text{则有 } x_{n+i} &\geq 0, \sum_{j=1}^n a_{ij}x_j + x_{n+i} = b_i \\ \text{类似可转化大于等于约束} \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

108

例2.2.1将线性规划

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

$$\text{s.t. } x_1+x_2+x_3 \leq 7$$

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

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

$$x_1, x_2 \geq 0, x_3 \text{ 是自由变量}$$

化为标准型

$$\text{解: 令 } x_4=7-(x_1+x_2+x_3),$$

$$x_5=(x_1-x_2+x_3)-2, \text{再令}$$

$$x_3=x_3'-x_3'', \text{得到标准型}$$

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

$$\text{s.t. } x_1+x_2+x_3'-x_3''+x_4=7$$

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

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

109

基本概念

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

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

110

基本概念

在约束方程组取定基矩阵

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

之后,令非基变量均为0,得到的方程组

$$p_1x_1+p_2x_2+\dots+p_mx_m=b$$

有唯一解,这样得到约束方程组的一个解向量

$$x=(x_1, x_2, \dots, x_m)^T$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

111

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

112

例 考虑线性规划

$$\min 2x_1-x_2$$

$$\text{s.t. } x_1+x_2+x_3=5$$

$$-x_1-x_2+x_4=0$$

$$2x_1+2x_2+x_5=22$$

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

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

$$(p_1, p_3, p_4), (11, 0, -6, 11, 0)^T \text{ 不可行}$$

$$(p_1, p_3, p_5), (0, 0, 5, 0, 22)^T \text{ 退化}$$

$$(p_1, p_4, p_5), (5, 0, 0, 5, 12)^T \text{ 非退化}$$

$$(p_2, p_3, p_4), (0, 11, -6, 11, 0)^T \text{ 不可行}$$

$$(p_2, p_3, p_5), (0, 0, 5, 0, 22)^T \text{ 退化}$$

$$(p_2, p_4, p_5), (0, 5, 0, 5, 12)^T \text{ 非退化}$$

$$(p_3, p_4, p_5), (0, 0, 5, 0, 22)^T \text{ 退化}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

113

§ 2.3 线性规划的基本定理

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

114

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

115

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 $x=(x_1, x_2, \dots, x_m, 0, \dots, 0)^T$ 是(LP)的基可行解,且 x_1, x_2, \dots, 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 $m+1 \leq j \leq n$ 时, $0 = x_j = \alpha u_j + (1 - \alpha)v_j$,因此 $u_j = v_j = 0$.

所以 $p_1 u_1 + p_2 u_2 + \dots + p_m u_m = p_1 v_1 + p_2 v_2 + \dots + p_m v_m = b$

从而 $p_1(u_1 - v_1) + p_2(u_2 - v_2) + \dots + p_m(u_m - v_m) = 0$

由于 x 是基可行解,所以 p_1, p_2, \dots, p_m 线性无关,

$u_j = v_j (j=1, 2, \dots, m)$.从而 $u=v$.这说明 x 为极点.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

116

充分性 设 $x=(x_1, x_2, \dots, x_k, 0, \dots, 0)^T$ 是可行域的极点,其中 $x_1, x_2, \dots, x_k > 0$.

假设 x 不是基可行解,于是 p_1, p_2, \dots, p_k 线性相关,即有一组不全为0的数 $\alpha_1, \alpha_2, \dots, \alpha_k$,使得

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

$$\text{又 } x \in D, \text{ 所以 } x_1 p_1 + x_2 p_2 + \dots + x_k p_k = b \quad (2.5)$$

用 $\varepsilon > 0$ 乘(2.4)再与(2.5)相加减得

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

117

$$\text{令 } u = (x_1 + \varepsilon \alpha_1, x_2 + \varepsilon \alpha_2, \dots, x_k + \varepsilon \alpha_k, 0, \dots, 0)^T$$

$$v = (x_1 - \varepsilon \alpha_1, x_2 - \varepsilon \alpha_2, \dots, x_k - \varepsilon \alpha_k, 0, \dots, 0)^T$$

则有 $Au=b, Av=b$,当 ε 充分小时,可使 $u \geq 0, v \geq 0$.

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 $D=\{x | Ax=b, x \geq 0\}$ 最多具有有限个极点

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

118

$(p_1, p_3, p_4), (11, 0, -6, 11, 0)^T$ 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)

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

119

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 $x=(x_1, x_2, \dots, x_n)^T$ 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, \dots, p_k 线性无关,则 x 为基可行解;

若 p_1, p_2, \dots, p_k 线性相关,即有一组不全为0的数 $\alpha_1, \alpha_2, \dots, \alpha_k$,使得 $\alpha_1 p_1 + \alpha_2 p_2 + \dots + \alpha_k p_k = 0$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

120

与定理2.3.2的证明类似,作

$$x_1 = x + \varepsilon \alpha, x_2 = x - \varepsilon \alpha, \text{其中 } \alpha = (\alpha_1, \dots, \alpha_k, 0, \dots, 0)^T$$

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 $x_j + \varepsilon \alpha_j, x_j - \varepsilon \alpha_j (j=1, \dots, k)$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

121

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 $c^T x + \varepsilon^T \alpha$ 与 $c^T x - \varepsilon^T \alpha$.

由于 x 是最优解, $c^T x + \varepsilon^T \alpha \geq c^T x$; $c^T x - \varepsilon^T \alpha \geq c^T x$.

因此 $\varepsilon^T \alpha = 0$.

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

122

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

123

单纯形方法的思路(几何)

线性规划

$$\min -72x_1 - 64x_2$$

$$\text{s.t. } x_1 + x_2 + x_3 = 50$$

$$12x_1 + 8x_2 + x_4 = 490$$

$$3x_1 + x_5 = 100$$

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

的等价形式为

$$\min -72x_1 - 64x_2$$

$$\text{s.t. } x_1 + x_2 \leq 50$$

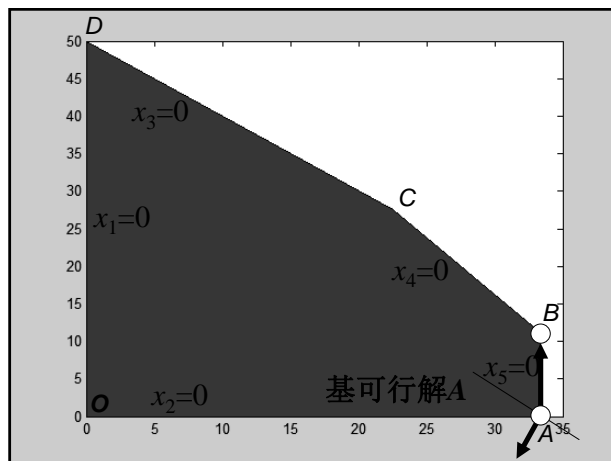
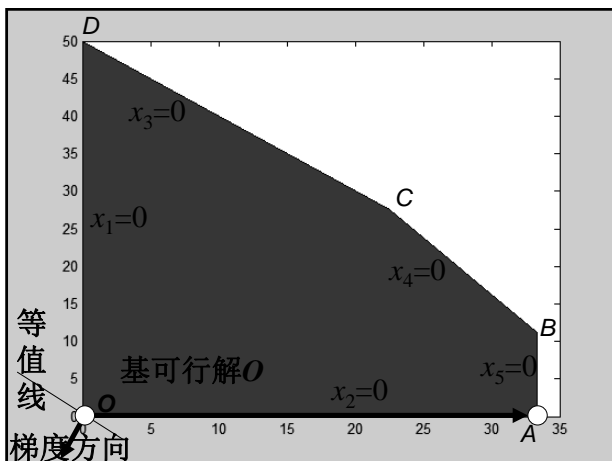
$$12x_1 + 8x_2 \leq 490$$

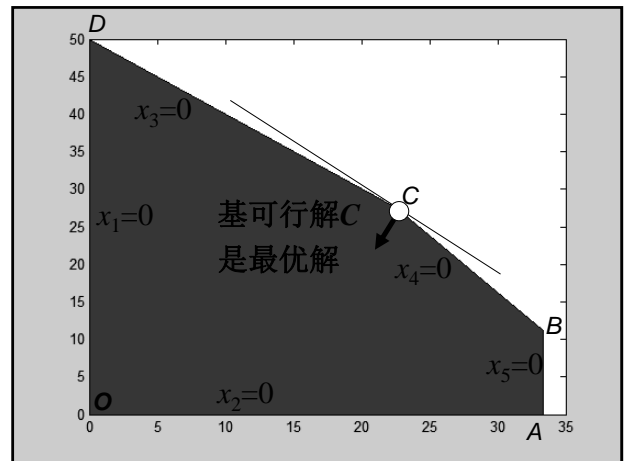
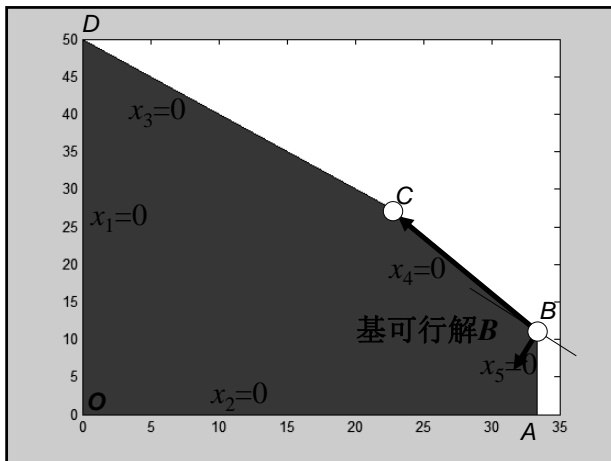
$$3x_1 \leq 100$$

$$x_1, x_2 \geq 0$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

124





单纯形方法的思路(代数)

例 考察线性规划

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

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 .

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

129

单纯形方法的思路(代数)

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

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”.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

130

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

$$\begin{aligned} \min & -64x_2 + 24x_5 - 2400 \\ \text{s.t.} & x_2 + x_3 - x_5/3 = 50/3 \\ & 8x_2 + x_4 - 4x_5 = 90 \\ & x_1 + x_5/3 = 100/3 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{aligned}$$

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

131

若 x_2 (其系数为负)的取值增加,可以使得目标函数值减少

$$\begin{aligned} \implies x_2 & \leq 50/3 \\ \implies x_2 & \leq 90/8 \end{aligned}$$

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

$$\begin{aligned} \min & 8x_4 - 8x_5 - 3120 \\ \text{s.t.} & x_3 - x_4/8 + x_5/6 = 65/12 \\ & x_2 + x_4/8 - x_5/2 = 45/4 \\ & x_1 + x_5/3 = 100/3 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{aligned}$$

解为 $(100/3, 45/4, 65/12, 0, 0)^T$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

132

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)

$$\min 48x_4 + 2x_5 - 3380$$

$$\text{s.t.} \quad 6x_3 - 3x_4/4 + x_5 = 65/2$$

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

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

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

133

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

134

§ 2.4 单纯形方法

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

135

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

$$\begin{array}{cccccccccccc} \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \end{array} = \begin{array}{c} \square \\ \square \\ \square \\ \square \end{array}$$

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

$$\begin{array}{cccccccccccc} \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \end{array} = \begin{array}{c} \square \\ \square \\ \square \\ \square \end{array}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

136

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

$$\begin{array}{cccccccccccc} \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \end{array} = \begin{array}{c} \square \\ \square \\ \square \\ \square \end{array}$$

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

$$\begin{array}{cccccccccccc} \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \end{array} = \begin{array}{c} \square \\ \square \\ \square \\ \square \end{array}$$

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

$$\begin{array}{cccccccccccc} \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square & \square \end{array} = \begin{array}{c} \square \\ \square \\ \square \\ \square \end{array}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

137

规范式

线性规划与基变量 x_1, \dots, x_m 对应的规范式

$$\min f(x) = f_0 + \sum_{j=m+1}^n \sigma_j x_j$$

$$\text{s.t.} \quad x_i = b'_i - \sum_{j=m+1}^n a'_{ij} x_j, i = 1, \dots, m$$

$$x_j \geq 0, j = 1, \dots, n$$

线性规划与基变量 x_{j_1}, \dots, x_{j_m} 对应的规范式

$$\begin{array}{ll} \min & f(x) = f_0 + \sum_{j \in T} \sigma_j x_j \\ \text{s.t.} & x_i = b'_i - \sum_{j \in T} a'_{ij} x_j, i \in S \\ & x_j \geq 0, j = 1, \dots, n \end{array}$$

$$\begin{array}{l} S = \{j_1, \dots, j_m\}, \\ T = \{1, 2, \dots, n\} \setminus S \end{array}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

138

2.4.1 基可行解是最优解的判断准则

在规范式 $\min f(x) = f_0 + \sum_{j \in T} \sigma_j x_j$
 $s.t. \quad x_i = b'_i - \sum_{j \in T} a'_{ij} x_j, i \in S$
 $x_j \geq 0, j = 1, \dots, n$

中,令非基变量 $x_j = 0, j \in T$, 得到一个基解
 $x_0 = (x_1, \dots, x_n)$ 其中 $\begin{cases} b'_j, & j \in S \\ 0 & j \in T \end{cases}$
 如果 $b'_j \geq 0 (j \in S)$, 则 x_0 是基可行解。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

139

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

140

最优性条件

在规范式 $\min f(x) = f_0 + \sum_{j \in T} \sigma_j x_j$ 中,
 $s.t. \quad x_i = b'_i - \sum_{j \in T} a'_{ij} x_j, i \in S$
 $x_j \geq 0, j = 1, \dots, n$

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 $B=(p_1, \dots, p_m)$ 的基可行解.与基变量 x_1, \dots, x_m 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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

141

判断无最优解

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

142

判断无最优解

证明:令 $d = (-a_{1k}, \dots, -a_{mk}, 0, \dots, 0, 1, 0, \dots, 0)^T$ (第 k 个分量为 1), $y =$

$$y_j = \begin{cases} b'_j - \lambda a'_{jk}, & j = 1, 2, \dots, m \\ \lambda, & j = k \\ 0, & j = m+1, m+2, \dots, n, j \neq k \end{cases}$$

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

143

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

144

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

$$x_i = b'_i - \sum_{j=m+1}^n a'_{ij} x_j, i = 1, 2, \dots, m$$

以及 $x_j=0, (j=m+1, \dots, n, j \neq k)$, 得到

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

我们选取 $x_k = \max\{x_k | x_i = b'_i -$

$$a'_{ik} x_k \geq 0, i = 1, 2, \dots, m\}$$

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

此时,我们选择 x_k 为离基变量.

145

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).

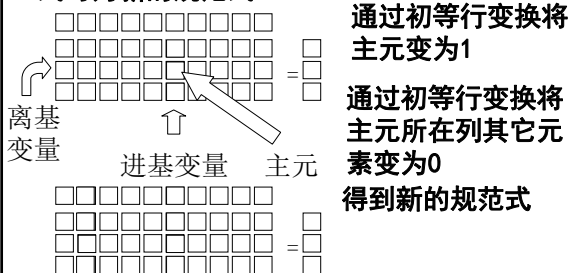
这种求解线性规划的方法称为单纯形方法.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

146

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.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

147

§ 2.5 单纯形表

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

148

例2.5.1 用单纯形法解 引入松弛变量将原问题解线性规划化为标准型

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

$$\text{s.t. } x_1 + x_2 \leq 6$$

$$x_1 + 2x_2 \leq 8$$

$$x_1 \leq 4$$

$$x_2 \leq 3$$

$$x_1, x_2 \geq 0$$

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

$$\text{s.t. } x_1 + x_2 + x_3 = 6$$

$$x_1 + 2x_2 + x_4 = 8$$

$$x_1 + x_5 = 4$$

$$x_2 + x_6 = 3$$

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

149

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

$$\text{s.t. } x_1 + x_2 + x_3 = 6$$

$$x_1 + 2x_2 + x_4 = 8$$

$$x_1 + x_5 = 4$$

$$x_2 + x_6 = 3$$

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

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	

基变量

150

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: $\sigma_2 = -3 = \min\{\sigma_j | \sigma_j < 0\}$, 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: $\theta_6 = \min \theta_i$, 所以 x_6 为离基变量 标出主元

c_j			写出基向量 (p_6 换成 p_2)			0	0	θ_i
c_B	B	b				p_5	p_6	
0	p_3	6	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			0	0	1	3		
p_2 所在行乘以-2加到 p_4 所在行			0	0	0			
p_5 所在行不变			0	0	-1			
p_2 所在行乘以3加到判别数所在行			1	0	-2			
-3	p_2	3	0	1	0	0	1	
σ_j			-2	0	0	0	3	

152

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	p_3	3	1	0	1	0	0	-1	
0	p_4	2	1	0	0	1	0	-2	
0	p_5	4	1	0	0	0	1	0	
-3	p_2	3	0	1	0	0	0	1	
σ_j			-2	0	0	0	0	3	

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: $\sigma_1 = -2 = \min\{\sigma_j | \sigma_j < 0\}$, 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: $\theta_4 = 2 = \min \theta_i$, x_4 为离基变量
 标出主元

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	p_3	1	0	0	1	-1	0	1	
-2	p_1	2	1	0	0	1	0	-2	
0	p_5	2	0	0	0	-1	1	2	
-3	p_2	3	0	1	0	0	0	1	
σ_j			0	0	0	2	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	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	p_3	1	0	0	1	-1	0	1	
-2	p_1	2	1	0	0	1	0	-2	
0	p_5	2	0	0	0	-1	1	2	
-3	p_2	3	0	1	0	0	0	1	
σ_j			0	0	0	2	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	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: $\sigma_6 = -1 = \min\{\sigma_j | \sigma_j < 0\}$, 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: $\theta_5 = 1 = \min \theta_i$, x_5 为离基变量(此处可选 x_3)
标出主元

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

157

c_j			-2	-3	0	0	0	0	写出基向量	
c_B	B	b	p_1	p_2	p_3	p_4	p_5	p_6	p (p_5 换成 p_6)	
0	p_3	1	0	0	1	-1	0	1	0	1
-2	p_1	2	1	0	0	1	0	-2	0	-2
0	p_5	2	0	0	0	-1	1	(2)	0	-2
-3	p_2	3	0	1	0	0	0	1	消去: 用初等行变换将主元所在列其它元素消为0	
σ_j			0	0	0	2	0	-1	0	0
			p_6 所在行乘以 -1 加到 p_3 所在行		0	0	1	3	0	-1
			p_6 所在行乘以 2 加到 p_1 所在行		-1/2	0				
			p_6 所在行乘以 -1 加到 p_2 所在行		0	1	0			
			p_6 所在行乘以 1 加到判别数所在行		1/2	1				
			-3	p_2	2	0	1	0	1/2	-1/2
σ_j			0	0	0	3/2	1/2	0		

158

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	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	

159

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

$f^* = (-2) \cdot 4 + (-3) \cdot 2 = -14$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

160

§ 2.6 初始基可行解的求法

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

161

对于线性规划问题 引入松弛变量化为标准型

$$\min c^T x$$

$$\text{s.t. } Ax \leq b (b \geq 0)$$

$$x \geq 0$$

$$\min c^T x$$

$$\text{s.t. } Ax + Ix_s = b$$

$$x, x_s \geq 0$$

其中 I 是单位矩阵, $x_s = (x_{n+1}, \dots, x_{n+m})^T$. 则可以将 x_s 作为基变量, 以 $(0, \dots, 0, b_1, \dots, b_m)^T$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

162

2.6.1 大M单纯形法

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

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

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

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

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

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

163

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

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

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

164

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 $(x_1, x_2, \dots, x_{n+m})$ 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 $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)^T$ 则 $x' = (\bar{x}_1, \dots, \bar{x}_n, 0, \dots, 0)^T$ 是辅助规划可行解,

于是 $Mx_r \leq \sum_{j=n+1}^{n+m} Mx_j \leq \bar{f}(x') - \sum_{j=1}^n c_j x_j = \sum_{j=1}^n c_j (\bar{x}_j - x_j)$ 这与 M 充分大矛盾.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

165

大M方法算例

例2.6.1 用大M单纯形法求解线性规划 引入松弛变量 x_5 , 人工变量 x_6, x_7 , 构造辅助线性规划

$$\min f(x) = -3x_1 + x_2 + x_3 \quad \min -3x_1 + x_2 + x_3 + Mx_6 + Mx_7$$

$$s.t. \begin{array}{lcl} x_1 - 2x_2 + x_3 & \leq & 11 \\ -4x_1 + x_2 + 2x_3 - x_4 & = & 3 \\ -2x_1 & + & x_3 = 1 \end{array} \quad s.t. \begin{array}{lcl} x_1 - 2x_2 + x_3 & + & x_5 = 11 \\ -4x_1 + x_2 + 2x_3 - x_4 & + & x_6 = 3 \\ -2x_1 & + & x_3 + x_7 = 1 \end{array}$$

$$x_1, x_2, x_3, x_4 \geq 0 \quad x_1, \dots, x_7 \geq 0$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

166

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: $\sigma_1 = -3 - 1 \cdot 0 - (-4)M - (-2)M = 6M - 3$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

167

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	
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		

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

168

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 $(4, 1, 9, 0, 0, 0)^T$,
原线性规划的最优解为 $(4, 1, 9, 0)^T$, 最优值为
 $-3 \cdot 4 + 1 \cdot 3 + 1 \cdot 9 = -2$.

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

170

2.6.2 两阶段单纯形法

对于线性规划问题 $\min f(x) = \sum_{j=1}^n c_j x_j$
 $s.t. \sum_{j=1}^n a_{ij} x_j = b_i (b_i \geq 0), i = 1, 2, \dots, m$
 $x_j \geq 0, j = 1, 2, \dots, n$

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

$\min w = \sum_{j=n+1}^{n+m} x_j$ 该规划有可行解 $(0, \dots, 0, b_1, \dots, b_n)^T$,
说明可行域非空.

$s.t. \sum_{j=1}^n a_{ij} x_j + x_{n+i} = b_i, i = 1, 2, \dots, m$ 目标函数显然
 $x_j \geq 0, j = 1, 2, \dots, n, n+1, \dots, n+m$ 有下界, 因此该
 规划有最优解.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

171

$$\min w = \sum_{j=n+1}^{n+m} x_j$$

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

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

172

两阶段法算例

例2.6.2 用两阶段
法求解线性规划

$$\min 4x_1 + x_2 + x_3$$

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

$$3x_1 + 3x_2 + x_3 = 3$$

$$x_1, x_2, x_3 \geq 0$$

解, 引入人工变量, 构造辅
助线性规划:

$$\min x_4 + x_5$$

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

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

173

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			

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

174

c_j		0	0	0	1	1	θ_i
c_B	B	b	p_1	p_2	p_3	p_4	

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.

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		

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

175

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 $f^*=11/5$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

176

§ 2.8 线性规划的对偶理论

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

177

假设某工厂有 m 种设备: B_1, B_2, \dots, B_m . 一年内各设备的生产能力(有效台时数)为 b_1, b_2, \dots, b_m . 利用这些设备可以加工 n 种产品: A_1, A_2, \dots, A_n . 单位产品的利润分别为 c_1, c_2, \dots, 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, 使全年总收入最大?

设 x_1, x_2, \dots, 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.

$$\begin{aligned} \max \quad & s = \sum_{j=1}^n c_j x_j \\ \text{s.t.} \quad & \sum_{j=1}^n a_{ij} x_j \leq b_i, i = 1, 2, \dots, m \\ & x_j \geq 0, j = 1, 2, \dots, n \end{aligned}$$

178

假设某工厂有 m 种设备: B_1, B_2, \dots, B_m . 一年内各设备的生产能力(有效台时数)为 b_1, b_2, \dots, b_m . 利用这些设备可以加工 n 种产品: A_1, A_2, \dots, A_n . 单位产品的利润分别为 c_1, c_2, \dots, 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, \dots, 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

179

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, 二是出租价格尽可能的低, 以利于市场竞争.

设第 i 种设备的出租价格为 y_i , 全年总收入为 w , 易建立该问题的数学模型.

$$\begin{aligned} \min \quad & w = \sum_{i=1}^m b_i y_i \\ \text{s.t.} \quad & \sum_{i=1}^m a_{ij} y_i \geq c_j, j = 1, 2, \dots, n \\ & y_i \geq 0, i = 1, 2, \dots, m \end{aligned}$$

180

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

181

原始线性规划

$$\begin{aligned} \max \quad & s = \sum_{j=1}^n c_j x_j \\ \text{s.t.} \quad & \sum_{j=1}^n a_{ij} x_j \leq b_i, i = 1, 2, \dots, m \\ & x_j \geq 0, j = 1, 2, \dots, n \end{aligned}$$

对偶线性规划

$$\begin{aligned} \min \quad & w = \sum_{i=1}^m b_i y_i \\ \text{s.t.} \quad & \sum_{i=1}^m a_{ij} y_i \geq c_j, j = 1, 2, \dots, n \\ & y_i \geq 0, i = 1, 2, \dots, m \end{aligned}$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

182

对偶规划(定义2.8.1)

原始线性规划

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

$$\begin{aligned} \min \quad & \boxed{c^T} \\ & \boxed{A} \geq \boxed{b} \\ & \boxed{x} \geq 0 \end{aligned}$$

对偶线性规划

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

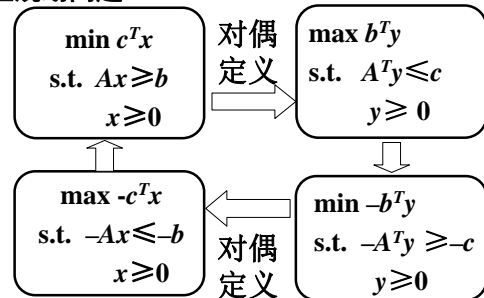
$$\begin{aligned} \max \quad & \boxed{b^T} \\ & \boxed{A^T} \leq \boxed{c} \\ & \boxed{y} \geq 0 \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

183

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



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

184

对偶规划的写法

min max, $A \quad A^T, c \quad b$, 约束 = 变量无正负

min + 约束 \geq → 变量非负 max + 约束 \geq → 变量非正

min + 约束 \leq → 变量非正 max + 约束 \leq → 变量非负

min + 变量非负 → 约束 \leq max + 变量非负 → 约束 \geq

min + 变量非正 → 约束 \geq max + 变量非正 → 约束 \leq

记忆方法: “max” = “非负” = “ \geq ” = “+” 类似定义 “min”

另外,左边的“约束” = “ \leq ”,然后将“+” “ \leq ”相结合

例如 “max + 约束 \geq ” 为两正一负,推出右边为负,即变量为非正

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

185

$$\max 16x_1 + 6x_2 - 3x_3$$

$$\text{s.t. } 2x_1 + x_2 - 4x_3 \geq -5$$

$$-3x_1 - x_2 = -2$$

$$2x_1 + 2x_2 + x_3 \leq 6$$

$$x_1 \geq 0, x_2 \leq 0,$$

$$x_3 \text{ 为自由变量}$$

先写出基本形式

max + 约束 \geq → 非正

max + 约束 \leq → 非负

$$\min -5y_1 - 2y_2 + 6y_3$$

$$\text{s.t. } 2y_1 - 3y_2 + 2y_3 \geq 16$$

$$y_1 - y_2 + 2y_3 \leq 6$$

$$-4y_1 + y_3 = -3$$

$$y_1 \leq 0, y_2 \text{ 为自由变量,}$$

$$y_3 \geq 0$$

max + 非负 → \geq

max + 非正 → \leq

完成!

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

186

非对称形式的对偶线性规划

标准型线性规划

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

对偶线性规划

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

187

对偶规划的性质

考虑如下的线性规划

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

其对偶规划为

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

188

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

189

弱对偶性定理

标准型线性规划

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

对偶线性规划

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

若 x_0, y_0 分别是(LP)与(DP)的可行解,则

$$Ax_0 = b, x_0 \geq 0, A^T y_0 \leq c$$

$$\text{于是 } y_0^T b = y_0^T (Ax_0) = (y_0^T A) x_0 = (A^T y_0)^T x_0 \leq c^T x_0$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

190

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 $c^T x_0 = b^T y_0$,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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

191

对偶规划的性质

考虑如下的线性规划

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

考虑基解 $(0,1,0,-2)^T$ (不可行).

对应的规范式为

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

其判别数向量

$(7/2, 0, 3/2, 0)$ 是非负的.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

192

对偶规划的性质

判别数向量用(LP)中的符号可以记为

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

$$\text{令 } y = (c_B^T B^{-1})^T,$$

则判别数向量为 $\sigma = c^T - y^T A$.

上述的判别数向量非负,因此有

$$c^T - y^T A \geq 0,$$

$$\text{即 } A^T y \leq c.$$

这说明 y 是对偶规划的可行解.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

193

对偶规划的性质

若 $B = (p_2, p_4), x = (0, 1, 0, -2)^T$,
 $\max 2y_1 + 3y_2$
 $(\text{DP}) \text{ s.t. } y_1 + 4y_2 \leq 5$
 $2y_1 + y_2 \leq 3$
 $y_1 \leq 0$
 $y_2 \leq 0$
 $\sigma = (7/2, 0, 3/2, 0)^T$, 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 \Rightarrow 松弛量非负 \Rightarrow 对偶解可行

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

194

对偶规划的性质

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

这一可行解对应的目标函数值为

$$y^T b = c_B^T B^{-1} b.$$

上式右端即为在原始线性规划中基解对应的
 目标函数值.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

195

对偶的线性规划问题的解

原始规划的最优性 $c^T - c_B^T B^{-1} A \geq 0$	\iff	对偶规划的可行性 $A^T y \leq c$
$y = (c_B^T B^{-1})^T$		

定理2.8.7 设 B 是问题

$$\min c^T x$$

$$\text{s.t. } Ax = b$$

$$x \geq 0$$

的一个基矩阵,对应的
 基解满足最优性条件

$$c^T - c_B^T B^{-1} A \geq 0,$$

则对偶线性规划问题

有可行解 $y = (c_B^T B^{-1})^T$,

$$\text{且 } c_B^T x_B = y^T b$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

196

对偶的线性规划问题的解

原始规划的最优性 $c^T - c_B^T B^{-1} A \geq 0$	\iff	对偶规划的可行性 $A^T y \leq c$
$y = (c_B^T B^{-1})^T$		

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

197

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

198

对偶性定理

证明:考虑到对合性,只需选 x_0 为最优解,判别数非负,有
 取两个规划中的任一个证明. $c^T - c_B^T B^{-1} A \geq 0$
 设 $\min c^T x$ $\quad c^T - c_B^T B^{-1} A \geq 0$
 s.t. $Ax = b$ \quad 令 $y_0 = (c_B^T B^{-1})^T$, 则
 $x \geq 0$ $\quad A^T y_0 \leq c, y_0$ 可行, 且
 有最优解 x_0 , 且其为基可行解. $y_0^T b = c_B^T B^{-1} b = c^T x_0$
 再设基矩阵为 $B, c^T x_0 = c_B^T B^{-1} b$ y_0 是对偶规划最优解

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

199

对偶的线性规划问题的解

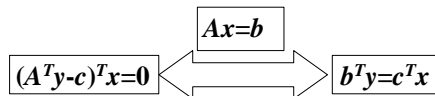
两个互为对偶的线性规划的解的情况

- (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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

200

互补松弛性



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 $(A^T y - c)^T x = 0$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

201

互补松弛性

对偶线性规划

原始线性规划
 $\max f = 6y_1 + 8y_2 + 4y_3 + 3y_4$
 $\min f = -2x_1 - 3x_2$
 s.t. $x_1 + x_2 + x_3 = 6$ $y_1 + 2y_2 + y_4 \leq -3$
 $x_1 + 2x_2 + x_4 = 8$ $y_1 \leq 0$
 $x_1 + x_5 = 4$ $y_2 \leq 0$
 $x_2 + x_6 = 3$ $y_3 \leq 0$
 $x_1, x_2, x_3, x_4, x_5, x_6 \geq 0$ $y_4 \leq 0$

原始规划最优解 $x^* = (4, 2, 0, 0, 0, 1)^T$, 互补松弛性:
 对偶规划最优解 $y^* = (0, -3/2, -1/2, 0)^T$ $q^T x = 0$
 对偶规划 y^* 的松弛量 $q = c - A^T y = (0, 0, 0, 3/2, 1/2, 0)^T$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

202

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 $(A^T y - c)^T x = 0$ 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.

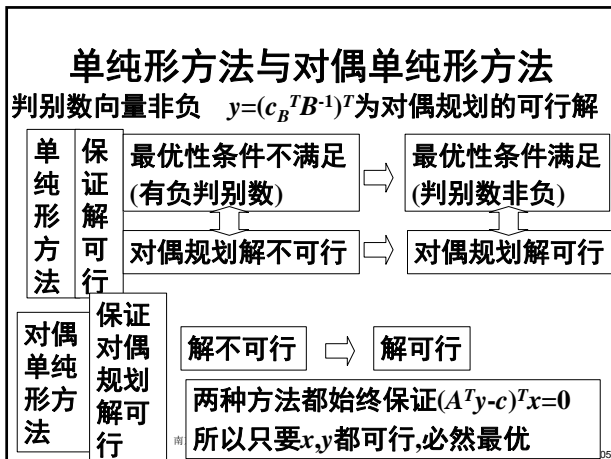
南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

203

§ 2.9 对偶单纯形法

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

204



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

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

则所得的基解为最优解

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn 206

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 $a_{lj} \geq 0 (j=1, 2, \dots, n)$, 则线性规划问题无可行解.

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn 207

3. 确定进基变量

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

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 ,

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

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

得:

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn 208

3. 确定进基变量

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

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

由(2.35)式, $a_{lk} < 0$

若 $a_{lj} \geq 0$, (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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn 209

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

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

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

引入松弛变量得到标准型线性规划

$$\begin{aligned} \min z &= 12x_1 + 8x_2 + 16x_3 + 12x_4 \\ \text{s.t. } &-2x_1 - x_2 - 4x_3 + x_5 = -2 \\ &-2x_1 - 2x_2 - 4x_4 + x_6 = -3 \\ &x_1, x_2, x_3, x_4, x_5, x_6 \geq 0 \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn 210

$\min z = 12x_1 + 8x_2 + 16x_3 + 12x_4$ 构造对偶单纯形表
 $\text{s.t. } -2x_1 - x_2 - 4x_3 + x_5 = -2$ 选取离基变量
 $-2x_1 - 2x_2 - 4x_4 + x_6 = -3$ $b_6 = -3 = \min\{b_j | b_j < 0\}$

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

211

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

212

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
$\sigma_j/a_{ij}(a_{ij} < 0)$			-3	-2	-4			
8	p_2	2	2	1	4	0	-1	0
12	p_4	-1/4	-1/2	0	(-2)	1	1/2	-1/4
σ_j			2	0	8	0	2	3
$\sigma_j/a_{ij}(a_{ij} < 0)$			-4		-4			-12
8	p_2	3/2	1	1	0	2	0	-1/2
16	p_3	1/8	1/4	0	1	-1/2	-1/4	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
 $z^* = 14$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

213

§ 2.10 灵敏度分析

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

214

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

215

灵敏度分析

设已经用单纯形法求得线性规划

$$\min c^T x$$

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

$$x \geq 0$$

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

216

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

217

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

218

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

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

219

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.

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

$$\text{s.t. } x_1 - 2x_2 + x_3 + x_5 = 11$$

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

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

$$x_1, \dots, x_7 \geq 0$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

220

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_i			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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

221

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

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

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

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

$$28/3 + 7/3 \Delta c_3 \geq 0.$$

$$\text{解得 } -1/4 \leq \Delta c_3 \leq 1/2$$

$$\text{即 } 3/4 \leq c_3 \leq 3/2$$

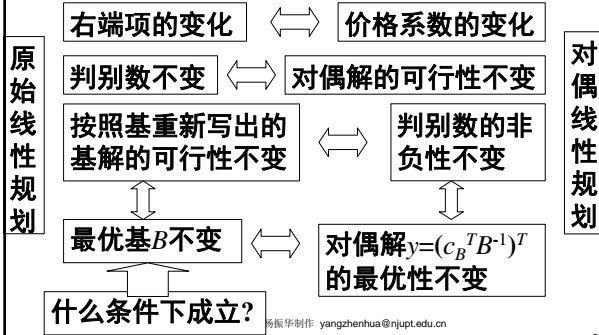
在实际应用中,如果价格系数的变化导致判别数变为负数,由于原来最优解的可行性不变,可以继续用单纯形方法求解.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

222

右端项的变化

右端项变化,原来的最优解不再可行.



B仍然是最优基?

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

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

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

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

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

即

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

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 $y = (c_B^T B^{-1})^T = (y_1, y_2, \dots, y_m)^T$, 是对偶规划最优解. $(c_B^T B^{-1})^T$ 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 σ_j 发生了变化.

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

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

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

若 $y_r < 0$, 则 $\Delta a_{rk} \geq \sigma_k / y_r$ 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?

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

$$\text{s.t. } x_1 - 2x_2 + x_3 + x_5 = 11$$

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

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

$$x_1, \dots, x_7 \geq 0$$

c_j			$-3 \quad 1$		$B^{-1} = \begin{pmatrix} 1/3 & 2/3 & -5/3 \\ 0 & 1 & 2 \\ 2/3 & 4/3 & -7/3 \end{pmatrix}$				
c_B	B	b	p_1	p_2					
-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 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 $x_B' = (4, 1, 9)^T + \Delta b_2 (2/3, 1, 4/3)^T \geq 0$,即
 $\Delta b_2 \geq -1, b \geq 2$ 时,最优基不变.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

229

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 $\Delta a_{24} \leq 1$ 时,最优基,最优解不变.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

230

§ 2.11 整数线性规划

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

231

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

232

2.11.1 割平面法

考虑整数规划

$\min -x_1 - x_2$

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

$4x_1 + 5x_2 \leq 20$

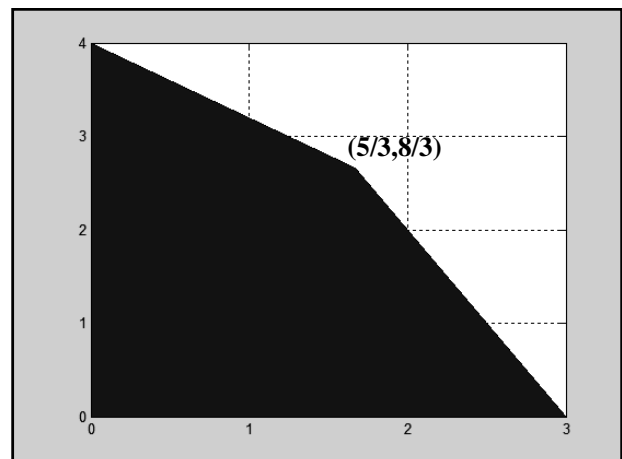
$x_1, x_2 \geq 0$ 且为整数

如果不考虑整数约束,其可行域见下图.

最优解为 $(5/3, 8/3)^T$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

233



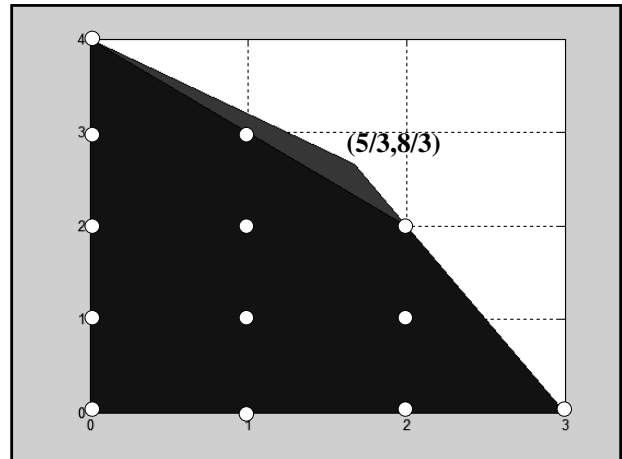
割平面法

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

235



解(LP)最终单纯形表为

考虑 x_2 对应的方程

		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	

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

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

化为等式约束

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

237

割平面法

考虑到

$$x_3 = 6 - (2x_1 + x_2),$$

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

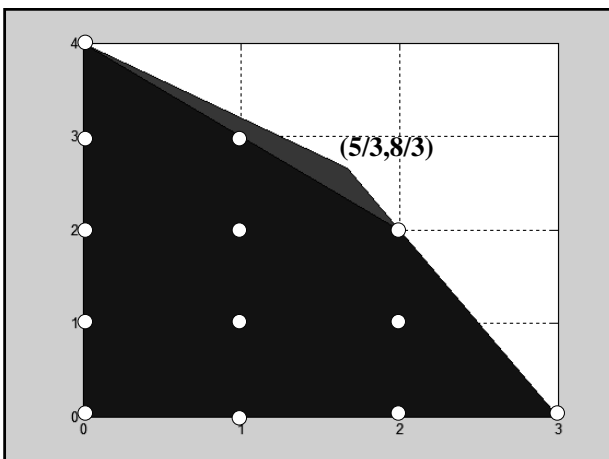
$$\text{根据 } 2/3 - x_3/3 - x_4/3 \leq 0$$

可以得到 $x_1 + x_2 \leq 4$

可以看出,(LP)中的可行域确实去掉了一块不含整点的部分.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

238



割平面法

设纯整数线性规划为

$$\min c^T x$$

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

$$x \geq 0$$

x_j 为整数, $j=1, 2, \dots, n$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

240

割平面方程

设对应的线性规划(LP)的最优解为 \bar{x} , 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 $x_r + \sum_{j \in T} \alpha_{rj} x_j = \bar{x}_r$

其中 T 表示非基变量的下标集.

把 \bar{x}_r 与 α_{rj} 分解为整数部分与非负真分数(纯小数), 即

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

代入到约束方程中得到

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

241

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.

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

242

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

$$\min c^T x$$

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

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

$$x \geq 0$$

$$x_j \text{ 为整数, } j=1, 2, \dots, n$$

该整数规划对应的线性规划可以用对偶单纯形方法求解.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

243

割平面法示例

例2.11.1 求解整数规划 解: 此规划的标准形式为

$$\min -x_1 - x_2$$

$$\min -x_1 - x_2$$

$$\text{s.t. } 2x_1 + x_2 \leq 6$$

$$\text{s.t. } 2x_1 + x_2 + x_3 = 6$$

$$4x_1 + 5x_2 \leq 20$$

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

$$x_1, x_2 \geq 0 \text{ 且为整数}$$

$$x_1, x_2, x_3, x_4 \geq 0 \text{ 且为整数}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

244

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

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

由于 $x_2 - x_3 - 2$ 为整数, 可得割平面方程

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

化为等式约束

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

245

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 .

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

246

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 $x=(0,4,2,0,0)^T$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

247

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

$x_r \leq N_r$ 或 $x_r \geq N_r+1$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

248

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.

$$\min z=c^T x$$

$$(P1) \text{ s.t. } Ax=b$$

$$x_r \leq N_r$$

$$x \geq 0$$

$$\min z=c^T x$$

$$(P2) \text{ s.t. } Ax=b$$

$$x_r \geq N_r+1$$

$$x \geq 0$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

249

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 $z_1 \leq z_2$ 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).

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

250

算例

例2.11.1用分枝定界法求解整数规划

$$\min z=5x_1+3x_2$$

$$(P0) \text{ s.t. } 3x_1+4x_2 \geq 12$$

$$5x_1+2x_2 \geq 10$$

$$x_1, x_2 \geq 0 \text{ 且为整数}$$

解:解与(P0)对应的线性规划,得到最优解与最优值分别为

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

$$z=85/7 \approx 12.14$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

251

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.

$$\min z=5x_1+3x_2$$

$$\text{s.t. } 3x_1+4x_2 \geq 12$$

$$(P1) \quad 5x_1+2x_2 \geq 10$$

$$x_1 \leq 1$$

$$x_1, x_2 \geq 0$$

$$\min z=5x_1+3x_2$$

$$\text{s.t. } 3x_1+4x_2 \geq 12$$

$$(P2) \quad 5x_1+2x_2 \geq 10$$

$$x_1 \geq 2$$

$$x_1, x_2 \geq 0$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

252

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.

$$\begin{array}{ll} \min z=5x_1+3x_2 & \min z=5x_1+3x_2 \\ \text{s.t. } 3x_1+4x_2 \geq 12 & \text{s.t. } 3x_1+4x_2 \geq 12 \\ \text{(P3)} \quad 5x_1+2x_2 \geq 10 & \text{(P4)} \quad 5x_1+2x_2 \geq 10 \\ x_1 \leq 1 & x_1 \leq 1 \\ \text{无可} & x_2 \leq 2 \\ \text{行解} & x_2 \geq 3 \\ x_1, x_2 \geq 0 & x_1, x_2 \geq 0 \end{array}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

253

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

254

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

256

§ 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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

257

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

258

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 $g^*=0$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

259

§ 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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

260

引理 对于正定二次函数 $H(x) = \frac{1}{2}x^T Gx + b^T x + c$

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

$$\alpha^2: \frac{1}{2} p_k^T G p_k \quad \alpha: p_k^T G x_k + b^T p_k = p_k^T g_k$$

因此

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

261

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

$$x_{k+1} = x_k - \frac{g_k^T g_k}{g_k^T G g_k} g_k$$

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 $x_k = 0.8^k [9, (-1)^k]^T (k=1, 2, \dots)$

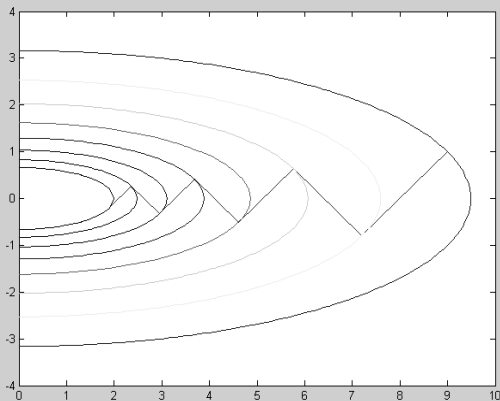
因此, $x_k \rightarrow x^* = (0, 0)^T$ 且 $\|x_{k+1} - x^*\| / \|x_k - x^*\| = 0.8$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

262



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 $L=\{x \in R^n | f(x) \leq f(x_0)\}$ 有界,令 $\{x_k\}$ 为最速下降法产生的点列, 则

或者(i)对某个 $k_0, g(x_{k_0})=0$;

或者(ii) $g_k \rightarrow 0 (k \rightarrow \infty)$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

264

2 用于二次函数时的收敛速度

定理 3.2.2 设 $f(x) = \frac{1}{2}x^T Gx$, 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

$$f(x_{k+1}) \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^2 f(x_k) \quad (k = 0, 1, 2, \dots)$$

$$\|x_k\| \leq \sqrt{\frac{\lambda_n}{\lambda_1} \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^k} \|x_0\| \quad (k = 0, 1, 2, \dots)$$

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

265

§ 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,

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

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 $x_{k+1} = x_k - G_k^{-1}g_k$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

266

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

267

Newton法算例

例3.3.2 用Newton法求解

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

设初始点为 $x_0 = (1, 1)^T$.

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

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

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

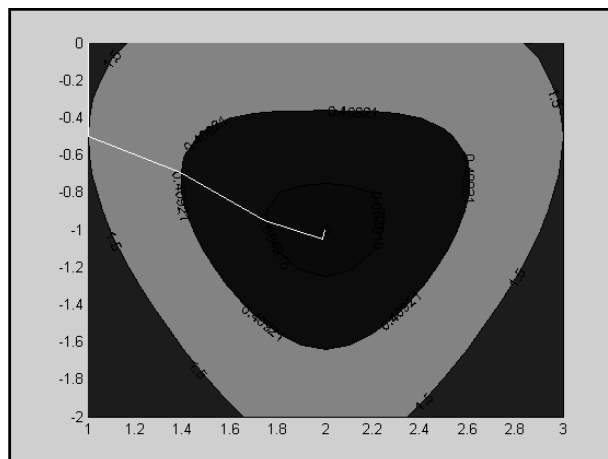
268

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

269



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 $h_k \rightarrow 0$ 且 $\|h_{k+1}\|=O(\|h_k\|^2)$ 其中 $h_k=x_k-x^*$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

271

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

272

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

273

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 $e_1=(1,0,\dots,0)^T$ 进行搜索,即求解下面问题

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

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

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

274

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

275

共扼方向法的思路

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

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

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

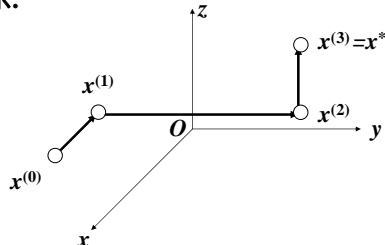
因此,上述的迭代过程每一步在一个分量上达到最优,且每一步求得了函数在一个集合中的极小点,这种集合在迭代过程中逐渐扩大,迭代 n 步之后得到原问题的最优解.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

276

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.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

277

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 $\min f(x) = \frac{1}{2}x^T Gx$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

278

共扼方向法的思路

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

在空间 R^n 上,重新定义内积与范数:

$$\langle x, y \rangle = x^T G y \quad \|x\|_G = \langle x, x \rangle^{1/2} = (x^T G x)^{1/2}$$

$$\begin{aligned} \text{则 } f(x) &= \frac{1}{2} \|x\|_G^2 + b^T x + c \\ &= \frac{1}{2} (x + G^{-1}b)^T G (x + G^{-1}b) + c - \frac{1}{2} b^T G^{-1}b \\ &= \frac{1}{2} \|x + G^{-1}b\|_G^2 + c - \frac{1}{2} b^T G^{-1}b \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

279

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, \dots, p_n ,

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

设问题的最优解 $x^* = -G^{-1}b$ 在这组基底下的

表示为 $x^* = u_1 p_1 + u_2 p_2 + \dots + u_n p_n$

任取初始点 $x^{(0)} = s_1 p_1 + s_2 p_2 + \dots + s_n p_n$, 在方向 p_1 上进行一维搜索,即求解问题

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

280

共扼方向法的思路

$$\begin{aligned} &\|(s_1 + \alpha_1 - u_1)p_1 + (s_2 - u_2)p_2 + \dots + (s_n - u_n)p_n\|_G^2 \\ &= \langle (s_1 + \alpha_1 - u_1)p_1 + (s_2 - u_2)p_2 + \dots + (s_n - u_n)p_n, \\ &\quad (s_1 + \alpha_1 - u_1)p_1 + (s_2 - u_2)p_2 + \dots + (s_n - u_n)p_n \rangle \\ &= (s_1 + \alpha_1 - u_1)^2 \|p_1\|_G^2 + \sum_{i=2}^n (s_i - u_i)^2 \|p_i\|_G^2 \end{aligned}$$

显然,当 $\alpha_1 = u_1 - s_1$ 时,上面的函数取最小值,

$$x^{(1)} = u_1 p_1 + s_2 p_2 + \dots + s_n p_n.$$

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 $\{x^{(0)} + \alpha_1 p_1, \alpha_1 \in R\}$ 中的极小点.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

281

共扼方向法的思路

类似的,再依次沿着 p_2, \dots, p_k 进行一维搜索,可以得到 $x^{(k)} = u_1 p_1 + \dots + u_k p_k + s_{k+1} p_{k+1} + \dots + u_n p_n$,

$x^{(k)}$ 是函数在 $\{x^{(0)} +$

$\alpha_1 p_1 + \alpha_2 p_2 + \dots + \alpha_k p_k, \alpha_1, \alpha_2, \dots, \alpha_k \in R\}$ 中的极小点.

最终 $x^{(n)} = u_1 p_1 + u_2 p_2 + \dots + u_n p_n = x^*$

即迭代过程同样在 n 步之后找到最优值.

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

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值.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

282

共扼方向法的思路

定义 3.4.2 设 n 维向量组 p_1, \dots, p_k 线性无关, $x^{(0)} \in R^n$, 称向量集合 $\{x^{(0)} + \sum_{i=1}^k \alpha_i p_i \mid \alpha_i \in R, i=1, 2, \dots, k\}$ 为由点 $x^{(0)}$ 与 p_1, p_2, \dots, 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

283

超平面上极小点的判断

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

$$H_k = \{x^{(0)} + \sum_{i=1}^k \alpha_i p_i \mid \alpha_i \in R, i=1, 2, \dots, k\} \quad \bar{x} = x^{(0)} + \sum_{i=1}^k \bar{\alpha}_i p_i \in H_k$$

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

$-p_1, -p_2, \dots, -p_k$ 也不是下降方向, 因此 $\nabla f(\bar{x})^T p_i \leq 0$ 于是有 $\nabla f(\bar{x})^T p_i = 0 (i=1, 2, \dots, k)$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

284

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

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

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

285

超平面上极小点的判断

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

286

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, \dots, 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, \dots, 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

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

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

287

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

288

共扼向量

对于 $f(x)=x^T Gx/2+b^T x+c$,有 $g(x)=Gx+b$,因此
 $g_{j+1}-g_j=G(x_{j+1}-x_j)=\alpha_j Gp_j$, 因此

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

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

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

p_1, p_2, \dots, 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

289

共扼方向法(用于二次函数)

注:在前面讨论思路时,根据方向的共扼性(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扼性(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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

290

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.

共扼方向法

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

291

共扼方向的性质

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

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

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

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 $\alpha_i = 0 (i=1, 2, \dots, k)$

于是 p_1, p_2, \dots, p_k 线性无关.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

292

共扼方向的性质

推论1 设 G 为 n 阶正定矩阵,非零向量组 p_1, p_2, \dots, p_n 关于 G 共扼,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, \dots, p_n 关于 G 共扼,若向量 v 与 p_1, p_2, \dots, p_n 关于 G 共扼,则 $v=0$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

293

共扼方向法(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, \dots, p_k 关于 G 共扼,对正定二次函数 $f(x)=x^T Gx/2+b^T x+c$ 由任意初始点 x_1 开始,依次进行 k 次一维搜索, $x_{i+1}=x_i+\alpha_i p_i (i=1, 2, \dots, k)$

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

(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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

294

共扼方向法(用于二次函数)

证明要点:只要证明 $g_{k+1}^T p_i = 0$.

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

精确一维搜索

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

295

共扼梯度法(共扼方向的形成)

我们首先讨论针对下面二次函数的共扼梯度法

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

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 共扼,于是 $0 = p_0^T G p_1 = -p_0^T G g_1 + \beta_0 p_0^T G p_0$

$$\text{因此 } \beta_0 = \frac{p_0^T G g_1}{p_0^T G p_0}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

296

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, \dots, 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, \dots, p_{k-1} 的线性组合.

$$\begin{aligned} p_k &= -g_k + \sum_{i=0}^{k-1} \beta_i p_i \\ \text{用 } p_j^T G (j=0, 1, \dots, k-1) \text{ 左乘上} \\ \text{式 } p_j^T G p_k &= -p_j^T G g_k + \sum_{i=0}^{k-1} \beta_i p_j^T G p_i = -p_j^T G g_k + \beta_j p_j^T G p_j \\ \text{因此 } \beta_j &= \frac{p_j^T G g_k}{p_j^T G p_j} \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

297

共扼梯度法(共扼方向的形成)

$$p_k = -g_k + \sum_{i=0}^{k-1} \beta_i p_i \quad 0 = -p_j^T G g_k + \beta_j p_j^T G p_j$$

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

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

$$\text{因此 } p_j^T G g_k = \frac{1}{\alpha_j} g_k^T (g_{j+1} - g_j)$$

由于 x_k 是由点 x_0 及向量 p_0, p_1, \dots, p_{k-1} 得到的 k 维超平面上的极小点,因此

$$g_k^T p_j = 0 \quad (j=0, 1, \dots, k-1) \quad \text{由 } p_j \text{ 的构造方式 } p_j = -g_j + \sum_{i=0}^{j-1} \beta_{ij} p_i$$

$$\text{因此 } g_k^T g_j = 0 \quad (j=0, 1, \dots, k-1).$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

298

共扼梯度法(共扼方向的形成)

$$p_j^T G g_k = \frac{1}{\alpha_j} g_k^T (g_{j+1} - g_j) \quad g_k^T g_j = 0 \quad (j=0, 1, \dots, k-1)$$

因此 $p_j^T G g_k = 0 \quad (j=0, 1, \dots, k-2)$

根据 $0 = -p_j^T G g_k + \beta_j p_j^T G p_j$ 得

$$\beta_j = 0 \quad (j=0, 1, \dots, k-2) \quad \beta_{k-1} = \frac{g_k^T G p_{k-1}}{p_{k-1}^T G p_{k-1}}$$

所以

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

299

共扼梯度法(用于二次函数)

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

定理3.4.4 对正定二次函数 $f(x) = \frac{1}{2} x^T G x + b^T x + c$

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 $m(\leq n)$ 次迭代后可函数的极小点,并且对所有 $i(1 \leq i \leq m)$ 有

$$p_i^T G p_j = 0, g_i^T g_j = 0, g_i^T p_j = 0, p_i^T g_i = -g_i^T g_i$$

其中 $j = 0, 1, \dots, i-1$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

300

FR算法

$\beta_{k-1} = \frac{g_k^T G p_{k-1}}{p_{k-1}^T G p_{k-1}}$ 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)Fletcher-Reeves公式

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

$$g_k^T G p_{k-1} = \frac{1}{\alpha_{k-1}} g_k^T (g_k - g_{k-1}) = \frac{1}{\alpha_{k-1}} g_k^T g_k$$

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

所以 $\beta_{k-1} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

301

PRP算法

FR算法中: $\beta_{k-1} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$

(2)Polak-Ribiere-Polyak公式

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

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 $p_k^T g_k = -g_k^T g_k < 0$, 因而都是下降算法.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

302

共扼梯度法算例

例3.4.1 用FR共扼梯度法求解 $(x_0=(0,0)^T)$

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

解 $g(x) = (3x_1 - x_2 - 2, x_2 - x_1)^T$ 注:此处不求 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 $\beta_0 = g_1^T g_1 / g_0^T g_0 = 1/9$

故 $p_1 = -g_1 + \beta_0 p_0 = (2/9, 2/3)^T$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

303

从 x_1 出发,沿 p_1 作一维搜索,求

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

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解

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

304

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

305

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 $\beta_0 = g_1^T g_1 / g_0^T g_0 = 6.763160$

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

306

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

307

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

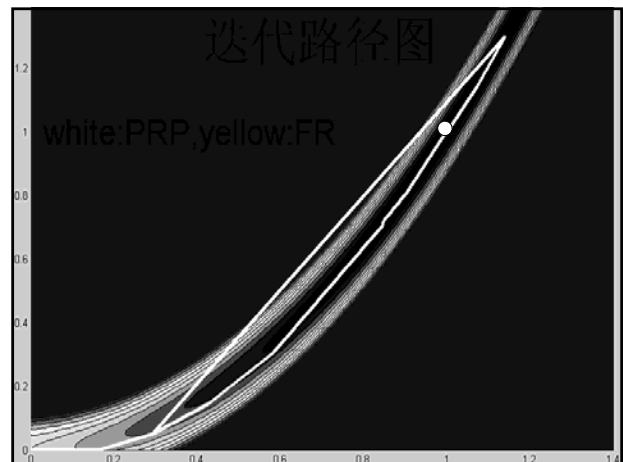
308

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	$(1+6.95e-10,1+9.93e-10)^T$	1.63e-17	1.79e-7
11	$(1+1.36e-9,2.69e-9)^T$	1.90e-18	1.27e-8
12	$(1+2.13e-10,1+4.66e-10)^T$	1.99e-19	1.71e-8
13	$(1-0.69e-11,1-1.35e-11)^T$	6.31e-23	1.85e-10
14	$(1-2.43e-13,1-5.50e-13)^T$	4.60e-27	2.79e-12
15	$(1-0.93e-14,1-1.82e-14)^T$	1.07e-28	2.15e-13

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

309



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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

311

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

312

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

313

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 $H_{k+1} y_k = s_k$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

314

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

315

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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

316

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

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

Step4 令 $x_{k+1} = x_k + \alpha_k p_k$

Step5 若 $\|g_{k+1}\| \leq \varepsilon$,则 $x^* = x_{k+1}$, Stop.

否则令 $s_k = x_{k+1} - x_k, y_k = g_{k+1} - g_k$

Step6 $H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}$

令 $k=k+1$,转Step2.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

317

算例(DFP算法)

例 3.5.1 用DFP算法求解($x_0 = (1, 1)^T, H_0 = I_2$)

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

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 $\phi(\alpha) = f(x_0 + \alpha p_0) = 40\alpha^2 - 20\alpha - 3$

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

$$H_1 = H_0 - \frac{H_0 y_0 y_0^T H_0}{y_0^T H_0 y_0} + \frac{s_0 s_0^T}{y_0^T s_0} = \frac{1}{100} \begin{pmatrix} 84 & 38 \\ 38 & 41 \end{pmatrix}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

318

下一搜索方向为 $p_1 = -H_1 g_1 = (1.6, 1.2)^T$.

(ii) 求迭代点 x_2 , 令 $\phi_1(\alpha) = f(x_1 + \alpha p_1) = 1.6\alpha^2 - 4\alpha - 5.5$, 其极小点为 $\alpha_1 = 1.25$, 于是

$$x_2 = x_1 + \alpha_1 p_1 = (4, 2)^T, g_2 = (0, 0)^T.$$

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 $H_2 = G^{-1} = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}^{-1}$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

319

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

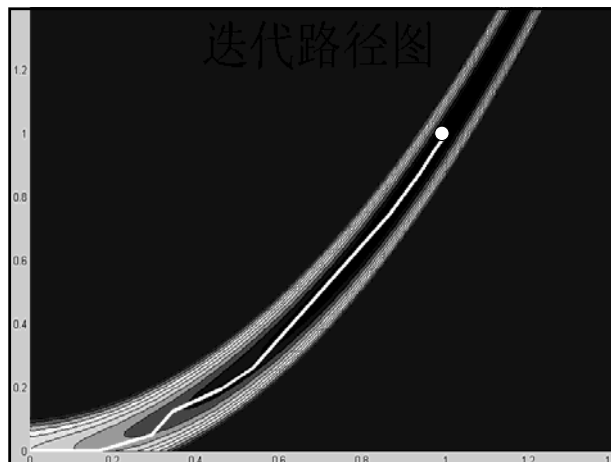
320

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

321



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

$$H_+ = H - \frac{Hy y^T H}{y^T H y} + \frac{ss^T}{y^T s} \quad (y \neq 0, s \neq 0)$$

如果 H_+ 也正定, 则

$$0 < y^T H_+ y = y^T H y - \frac{y^T H y y^T H y}{y^T H y} + \frac{y^T s s^T y}{y^T s} = y^T s \implies y^T s > 0$$

反之, 如果 $y^T s > 0$, 对任意的 $x (\neq 0) \in R^n$, 有

$$x^T H_+ x = \frac{x^T H x y^T H y - x^T H y y^T H x}{y^T H y} + \frac{(s^T x)^2}{y^T s}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

323

DFP修正公式的正定继承性

$$x^T H_+ x = \frac{x^T H x y^T H y - x^T H y y^T H x}{y^T H y} + \frac{(s^T x)^2}{y^T s}$$

由于 H 正定, 存在正定矩阵 D , 使得 $H = DD^T$, 令 $u = Dx, v = Dy$, 则有 $x^T H_+ x = \frac{u^T u v^T v - (u^T v)^2}{y^T H y} + \frac{(s^T x)^2}{y^T s}$

根据Cauchy-Schwartz不等式, 右端第一项非负, 根据 $y^T s > 0$, 右端第二项也非负, 因此 $x^T H_+ x$ 非负.

若第一项为0, 则 $u = \beta v (\beta \neq 0)$, 此时 $s^T x = \beta^T y \neq 0$, 因此第二项不为零, 所以总有 $x^T H_+ x > 0$.

所以, 在 $y^T s > 0$ 时, H_+ 正定.

于是得到引理3.5.1, H_+ 的充要条件为 $y^T s > 0$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

324

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 $y_k^T s_k$ 是否为正.

$$\begin{aligned} y_k^T s_k &= (g_{k+1} - g_k)^T (x_{k+1} - x_k) = \alpha_k (g_{k+1} - g_k)^T p_k \\ &= \alpha_k g_{k+1}^T p_k - \alpha_k g_k^T p_k \end{aligned}$$

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, $p_k = -H_k g_k$ 为下降方向,所以 $\alpha_k > 0$,从而 $y_k^T s_k > 0$.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).

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

325

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

应用DFP算法时,有

$$\begin{aligned} p_1^T G p_0 &= (G p_0)^T p_1 \\ s_0 &= x_1 - x_0 = \alpha_0 p_0, p_1 = -H_1 g_1 &= -\frac{1}{\alpha_0} (G s_0)^T H_1 g_1 \\ y_0 &= g_1 - g_0 = (G x_1 + b) - (G x_0 + b) = G s_0 &= -\frac{1}{\alpha_0} y_0^T H_1 g_1 \\ H_1 y_0 &= s_0 &= -\frac{1}{\alpha_0} g_1^T s_0 \\ & &= -g_1^T p_0 = 0 \end{aligned}$$

因此 p_1 与 p_0 共扼.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

326

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 $0 \leq i < j \leq l$ 时 $p_i^T G p_j = 0$.对 $0 \leq i \leq l$

$$\begin{aligned} p_i^T G p_{l+1} &= (G p_i)^T p_{l+1} & s_i &= x_{i+1} - x_i = \alpha_i p_i, \\ &= -\frac{1}{\alpha_i} (G s_i)^T H_{l+1} g_{l+1} & p_{l+1} &= -H_{l+1} g_{l+1} \\ &= -\frac{1}{\alpha_i} y_i^T H_{l+1} g_{l+1} & y_i &= g_{i+1} - g_i = G s_i \\ &= -\frac{1}{\alpha_i} g_{l+1}^T s_i & H_{l+1} y_i &= s_i? \\ &= -g_{l+1}^T p_i \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

327

DFP算法的二次终止性

$$\begin{aligned} g_{l+1}^T p_i &= (g_{l+1} - g_l + \cdots + g_{i+2} - g_{i+1} + g_{i+1})^T p_i & y_j &= G s_j \\ &= (\sum_{j=i+1}^l y_j + g_{i+1})^T p_i = g_{i+1}^T p_i + \sum_{j=i+1}^l s_j^T G p_i & s_j &= \alpha_j p_j, \\ & & p_j^T G p_i &= 0 \end{aligned}$$

$H_{l+1} y_i = s_i?$ $i=l$ 时显然成立, $0 \leq i \leq l-1$ 时

$$\begin{aligned} H_{l+1} y_i &= (H_l - \frac{H_l y_l y_l^T H_l}{y_l^T H_l y_l} + \frac{s_l s_l^T}{y_l^T s_l}) y_i & \text{归纳假设中} \\ &= H_l y_i - \frac{H_l y_l y_l^T H_l y_i}{y_l^T H_l y_l} + \frac{s_l s_l^T y_i}{y_l^T s_l} & \text{添上 } H_l y_i = s_i \\ &= s_i - \frac{H_l y_l y_l^T s_i}{y_l^T H_l y_l} + \frac{s_l s_l^T y_i}{y_l^T s_l} & y_l = G s_l, y_i = G s_i \\ &= s_i - \frac{H_l y_l (s_l^T G s_i)}{y_l^T H_l y_l} + \frac{s_l s_l^T G s_i}{y_l^T s_l} & s_i = \alpha_i p_i, s_l = \alpha_l p_l, \\ & & p_l^T G p_i = 0 \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

328

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) $p_i^T G p_j = 0, 0 \leq i < j \leq k$

(ii) $H_k y_i = s_i, 0 \leq i \leq k-1$.

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

329

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 $x_k \neq x^* (0 \leq k \leq n-1)$, 则 $H_n = G^{-1}$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

330

DFP算法的性质

- (1)对于正定二次函数
 (i)是共轭方向法(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
 $H_k y_i = s_i, 0 \leq i \leq k-1$.
 (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).

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

331

DFP算法的搜索方向

下面研究精确一维搜索时,DFP 根据 $g_{k+1}^T p_k = 0$ 得方法中第k+1步的下降方向.

$$\begin{aligned} H_{k+1}^{(DFP)} g_{k+1} &= H_k g_{k+1} - \frac{H_k y_k y_k^T H_k g_{k+1}}{y_k^T H_k y_k} + \frac{s_k s_k^T g_{k+1}}{y_k^T s_k} \\ &= H_k g_{k+1} - \frac{H_k y_k y_k^T H_k g_{k+1}}{y_k^T H_k y_k} \\ H_k g_{k+1} &= H_k (g_{k+1} - g_k) + H_k g_k = H_k y_k - p_k = H_k y_k - \alpha_k^{-1} s_k \\ H_{k+1}^{(DFP)} g_{k+1} &= H_k y_k - \alpha_k^{-1} s_k - \frac{H_k y_k y_k^T (H_k y_k - \alpha_k^{-1} s_k)}{y_k^T H_k y_k} \\ &= \frac{y_k^T s_k H_k y_k}{\alpha_k y_k^T H_k y_k} - \frac{s_k}{\alpha_k} \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

332

DFP算法的搜索方向

$$p_{k+1} = -H_{k+1} g_{k+1} = -\left(\frac{y_k^T s_k H_k y_k}{\alpha_k y_k^T H_k y_k} - \frac{s_k}{\alpha_k} \right) = \frac{y_k^T s_k}{\alpha_k} \left(\frac{s_k}{y_k^T s_k} - \frac{H_k y_k}{y_k^T H_k y_k} \right)$$

记 $w_k = (y_k^T H_k y_k)^{1/2} \left(\frac{s_k}{y_k^T s_k} - \frac{H_k y_k}{y_k^T H_k y_k} \right)$

显然 p_{k+1} 与 w_k 线性相关,

且 $w_k^T y_k = (y_k^T H_k y_k)^{1/2} \left(\frac{s_k^T y_k}{y_k^T s_k} - \frac{y_k^T H_k y_k}{y_k^T H_k y_k} \right) = 0$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

333

Broyden族拟Newton法

$$w_k = (y_k^T H_k y_k)^{1/2} \left(\frac{s_k}{y_k^T s_k} - \frac{H_k y_k}{y_k^T H_k y_k} \right) \quad w_k^T y_k = 0$$

在DFP公式中的 H_{k+1} 满足拟Newton方程 $H_{k+1} y_k = s_k$ 因此称3校正公式

即 $H_{k+1}(\varphi) = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k} + \varphi_k w_k w_k^T$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

334

Broyden族拟Newton法

$$H_{k+1}(\varphi) = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k} + \varphi_k w_k w_k^T$$

取 $\varphi_k = \frac{y_k^T s_k}{(s_k - H_k y_k)^T y_k}$ 就得到秩1校正公式

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

取 $\varphi=1$,得到的公式称为BFGS公式

$$\begin{aligned} H_{k+1} &= H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k} \\ &\quad + (y_k^T H_k y_k) \left(\frac{s_k}{y_k^T s_k} - \frac{H_k y_k}{y_k^T H_k y_k} \right)^T \left(\frac{s_k}{y_k^T s_k} - \frac{H_k y_k}{y_k^T H_k y_k} \right) \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

335

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

336

Broyden族拟Newton法

$$p_{k+1}^{BFGS} = -H_{k+1}^{BFGS} g_{k+1} = -(H_{k+1}^{DFP} + w_k w_k^T) g_k = p_{k+1}^{DFP} + (w_k^T g_k) w_k$$

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 $x (\neq 0) \in R^n$, $x^T H_{k+1}^{BFGS} x = x^T H_{k+1}^{DFP} x + x^T w_k w_k^T x > 0$

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).

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

337

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 ($\mu=0.1, \sigma=0.5$), 用BFGS方法21步收敛 ($\|g_k\| < 10^{-12}$), 而DFP方法30步收敛.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

338

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

340

§ 4.1 约束最优化问题的最优性条件

$$\begin{aligned} \text{问题 } & \min f(x), x \in R^n \\ \text{s.t. } & c_i(x) = 0, i \in E = \{1, 2, \dots, l\} \\ & c_i(x) \geq 0, i \in I = \{l+1, \dots, m\} \end{aligned}$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

341

4.1.1 等式约束问题的最优性条件

$$\begin{aligned} \text{问题 } & \min f(x), x \in R^n \\ \text{s.t. } & c_i(x) = 0, i \in E = \{1, 2, \dots, l\} \end{aligned}$$

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.

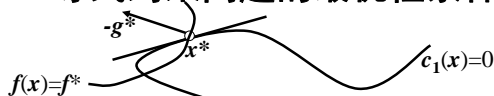
考虑 $f(x)$ 在最优点处的负梯度方向

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

342

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“最优性”矛盾,因此梯度方向与切线垂直.

或, $f(x)$ 在最优性点处的梯度方向就是 $c_1(x)=0$ 在该点处的法向.

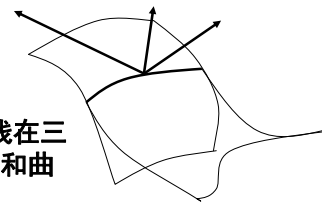
而 $c_1(x)=0$ 在该点处的法线方向为 $\nabla c_1(x^*)$

因此,存在数 λ_1 ,使得 $\nabla f(x^*) - \lambda_1 \nabla c_1(x^*) = 0$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

343

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 $\nabla f(x^*) - \lambda_1 \nabla c_1(x^*) - \lambda_2 \nabla c_2(x^*) = 0$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

344

等式约束问题的一阶必要条件

$$\begin{aligned} \min f(x), x \in R^n \\ \text{s.t. } c_i(x) = 0, i \in E = \{1, 2, \dots, l\} \end{aligned}$$

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 $c_i(x)(i=1, 2, \dots, l)$ 在 x^* 的某邻域内连续可微;

(iii) $\nabla c_i(x^*)(i=1, 2, \dots, l)$ 线性无关

则存在一组不全为零的数 $\lambda_1^*, \lambda_2^*, \dots, \lambda_l^*$ 使得

$$\nabla f(x^*) - \sum_{i=1}^l \lambda_i^* \nabla c_i(x^*) = 0.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

345

等式约束问题的一阶必要条件

$$\begin{aligned} \min f(x), x \in R^n \\ \text{s.t. } c_i(x) = 0, i \in E = \{1, 2, \dots, l\} \end{aligned}$$

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 $c(x)=(c_1(x), \dots, c_l(x))^T, \lambda=(\lambda_1, \dots, \lambda_l)^T$.

称 λ 为Lagrange乘子向量.

Lagrange函数的梯度为

$$\nabla L(x, \lambda) = \begin{pmatrix} \nabla_x L \\ \nabla_\lambda L \end{pmatrix} = \begin{pmatrix} \nabla f(x) - \sum_{i=1}^l \lambda_i \nabla c_i(x) \\ -(c_1(x), \dots, c_l(x))^T \end{pmatrix}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

346

等式约束问题的一阶必要条件

$$\nabla L(x, \lambda) = \begin{pmatrix} \nabla_x L \\ \nabla_\lambda L \end{pmatrix} = \begin{pmatrix} \nabla f(x) - \sum_{i=1}^l \lambda_i \nabla c_i(x) \\ -(c_1(x), \dots, c_l(x))^T \end{pmatrix}$$

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 $c_i(x^*), i=1, \dots, l$.

所以求含 $n+l$ 个未知数 $x_1, \dots, x_n, \lambda_1, \dots, \lambda_l$ 的非线性方程组的解 (x^*, λ^*) ,其中 $x^*=(x_1^*, \dots, x_n^*)^T$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

347

等式约束问题的二阶充分条件

$$\begin{aligned} \min f(x), x \in R^n \\ \text{s.t. } c_i(x) = 0, i \in E = \{1, 2, \dots, l\} \end{aligned}$$

定理4.1.2 在上面的等式约束问题中,若

(i) $f(x)$ 与 $c_i(x)(1 \leq i \leq l)$ 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

$$s^T \nabla c_i(x^*) = 0, (i=1, 2, \dots, l) \text{ 均有 } s^T \nabla_x^2 L(x^*, \lambda^*) s > 0$$

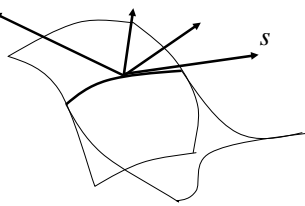
则 x^* 是上面问题的严格局部极小点.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

348

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.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

349

4.1.1 不等式约束问题的最优性条件

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) \geq 0, i \in \{1, 2, \dots, m\}$$

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 $c_j(x) \geq 0$ 中的等号成立,即 $c_j(\tilde{x}) = 0$,则该不等式约束 $c_j(x) \geq 0$ 称为关于 \tilde{x} 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 $c_k(x) \geq 0$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

350

Fritz-John一阶必要条件

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) \geq 0, i \in \{1, 2, \dots, m\}$$

定理4.1.6 设 x^* 为上述问题的局部最优解且 $f(x), c_i(x) (1 \leq i \leq m)$ 在 x^* 点可微,则存在非零向量 $\lambda^* = (\lambda_0^*, \lambda_1^*, \dots, \lambda_m^*)$ 使得

$$\lambda_0^* \nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla c_i(x^*) = 0,$$

$$\lambda_i^* c_i(x^*) = 0, i = 1, \dots, m$$

$$\lambda_i^* \geq 0, i = 0, 1, \dots, m$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

351

Fritz-John一阶必要条件

$$\lambda_0^* \nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla c_i(x^*) = 0, \quad \lambda_i^* c_i(x^*) = 0, i = 0, 1, \dots, m$$

$$\lambda_i^* \geq 0, i = 0, 1, \dots, m$$

证明概要 设 x^* 处的有效集为

$$I^* = I(x^*) = \{i \mid c_i(x^*) = 0, i = 1, 2, \dots, m\}.$$

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

$$\lambda_0^* \nabla f(x^*) - \sum_{i \in I^*} \lambda_i^* \nabla c_i(x^*) = 0, \quad \lambda_i^* \geq 0, i \in \{0\} \cup I^*.$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

352

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 $c_3(x) \geq 0$ 为无效约束, $c_1(x) \geq 0$, $c_2(x) \geq 0$ 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

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

353

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

$$I^* = \{i \mid c_i(x^*) = 0, i = 1, 2, \dots, m\};$$

(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 $G \cap S = \emptyset$

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

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

354

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

355

Gordan引理

引理4.1.4 设 a_1, \dots, a_r 是 n 维向量,则不存在向量 $d \in R^n$ 使得

$$a_i^T d < 0 (i=1, \dots, r)$$

成立的充要条件是,存在不全为零的非负实数组 $\lambda_1, \dots, \lambda_r$,使

$$\sum_{i=1}^r \lambda_i a_i = 0.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

356

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 $\lambda_0^* \geq 0, \lambda_i^* \geq 0 (i \in I^*)$,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。

$$\begin{aligned} \lambda_0^* \nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla c_i(x^*) &= 0, \\ \lambda_i^* c_i(x^*) &= 0, i = 0, 1, \dots, m \\ \lambda_i^* &\geq 0, i = 0, 1, \dots, m \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

357

例题 (Fritz-John条件)

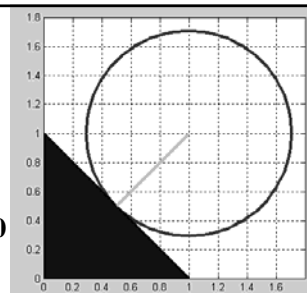
例4.1.1

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

$$\text{s.t. } c_1(x_1, x_2) = (1 - x_1 - x_2)^3 \geq 0$$

$$c_2(x) = x_1 \geq 0$$

$$c_3(x) = x_2 \geq 0$$



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 $x^* = (1/2, 1/2)^T$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

358

例题(Fritz-John条件)

$\min f(x) = (x_1 - 1)^2 + (x_2 - 1)^2$ 下面求满足 Fritz-John条件的点.

$$\text{s.t. } c_1(x_1, x_2) = (1 - x_1 - x_2)^3 \geq 0$$

$$c_2(x) = x_1 \geq 0$$

$$c_3(x) = x_2 \geq 0$$

即

$$\lambda_0^* \begin{pmatrix} 2x_1 - 2 \\ 2x_2 - 2 \end{pmatrix} - \lambda_1^* \begin{pmatrix} -3(1 - x_1 - x_2)^2 \\ -3(1 - x_1 - x_2)^2 \end{pmatrix} - \lambda_2^* \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \lambda_3^* \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0,$$

$$\lambda_1^* (1 - x_1 - x_2)^3 = 0, \quad \lambda_2^* x_1 = 0, \quad \lambda_3^* x_2 = 0. \quad \lambda_i^* \geq 0, i = 0, 1, 2, 3.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

359

例题

$$\lambda_0^* \begin{pmatrix} 2x_1 - 2 \\ 2x_2 - 2 \end{pmatrix} - \lambda_1^* \begin{pmatrix} -3(1 - x_1 - x_2)^2 \\ -3(1 - x_1 - x_2)^2 \end{pmatrix} - \lambda_2^* \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \lambda_3^* \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0,$$

$$\lambda_1^* (1 - x_1 - x_2)^3 = 0, \quad \lambda_2^* x_1 = 0, \quad \lambda_3^* x_2 = 0. \quad \lambda_i^* \geq 0, i = 0, 1, 2, 3.$$

若 $\lambda_1^* = 0, \lambda_2^* = 0, \lambda_3^* = 0$,则 $\lambda_0^* > 0, x = (1, 1)^T$.

该点虽然满足Fritz-John条件,但不是可行点。

若 $\lambda_0^* = 0, \lambda_2^* = 0, \lambda_3^* = 0$,则 $\lambda_1^* > 0, x = (u, 1 - u)^T$.

当 $0 \leq u \leq 1$ 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 $\lambda_0^* = 0, \lambda_1^* = 0$,则根据最上面的等式可以推出 $\lambda_2^* = 0, \lambda_3^* = 0$.不满足Fritz-John条件。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

360

Fritz-John条件的缺点

对于上面的例子,可行的Fritz-John点对应的 $\lambda_0^* = 0$.
 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

361

Kuhn-Tucker一阶必要条件

定理4.1.6 设 $\min f(x), x \in R^n$
 $s.t. \quad c_i(x) \geq 0, i \in \{1, 2, \dots, m\}$

(i) x^* 为上述问题的局部最优解, 有效集

$I^* = \{i | c_i(x^*) = 0, i = 1, 2, \dots, m\};$

(ii) $f(x), c_i(x) (1 \leq i \leq m)$ 在 x^* 点可微;

(iii) 对于 $i \in I^*$ 的 $\nabla c_i(x^*)$ 线性无关,
 则存在向量 $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)$ 使得

$$\nabla f(x^*) - \sum_{i \in I^*} \lambda_i^* \nabla c_i(x^*) = 0,$$

$$\lambda_i^* c_i(x^*) = 0, i = 1, \dots, m$$

$$\lambda_i^* \geq 0, i = 1, \dots, m$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

362

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 $\nabla c_i(x^*)^T d \geq 0 (i \in I^*),$

有

$$\nabla f(x^*)^T d \geq 0.$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

363

Kuhn-Tucker必要条件

$$\nabla c_i(x^*)^T d \geq 0 (i \in I^*) \implies \nabla f(x^*)^T d \geq 0.$$

借助于Farkas引理,可推出存在 $\lambda_i^* \geq 0 (i \in I^*),$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

364

Farkas引理

引理 4.1.3 设 a_1, a_2, \dots, a_r 和 b 均为 n 维向量, 则所有满足

$$a_i^T d \geq 0, i = 1, \dots, r$$

的向量 $d \in R^n$, 同时也满足不等式 $b^T d \geq 0$ 的充要条件是, 存在非负实数 $\lambda_1, \lambda_2, \dots, \lambda_r$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

365

Farkas引理的几何说明

考虑二维空间中两个向量 a_1, a_2 .

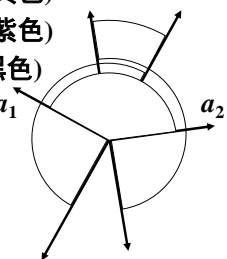
满足 $a_1^T d \geq 0$ 的方向 d 的范围(黄色)

满足 $a_2^T d \geq 0$ 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.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

366

Kuhn-Tucker必要条件

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) \geq 0, i \in \{1, 2, \dots, m\}$$

$$\nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla c_i(x^*) = 0,$$

$$\lambda_i^* c_i(x^*) = 0, i = 1, \dots, m$$

$$\lambda_i^* \geq 0, i = 1, \dots, m$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

367

线性规划

对于线性规划问题

$$\min f(y) = -b^T y$$

$$s.t. \quad -A^T y \geq -c$$

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 $x^* \geq 0$,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

$$\text{及 } (A^T y^* - c)^T x^* = 0$$

线性规划互补松弛条件

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

368

4.1.3 一般约束问题的最优性条件

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) = 0, i \in E = \{1, 2, \dots, l\}$$

$$c_i(x) \geq 0, i \in I = \{l+1, \dots, m\}$$

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) $f(x), c_i(x) (1 \leq i \leq m)$ 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 $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)$ 使得

$$\nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla c_i(x^*) = 0, \quad \lambda_i^* c_i(x^*) = 0, i \in I_1 \quad \lambda_i^* \geq 0, i \in I_2$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

369

一般约束问题的最优性条件

$$m+n \text{ 维函数 } L(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i c_i(x^*)$$

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.

$$\text{矩阵 } \nabla_x^2 L(x^*, \lambda^*) = \nabla^2 f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla^2 c_i(x^*)$$

称为Lagrange函数在 $\begin{pmatrix} x^* \\ \lambda^* \end{pmatrix}$ 处的Hesse矩阵,记为 $w^* = \nabla_x^2 L(x^*, \lambda^*)$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

370

二阶充分条件

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) = 0, i \in E = \{1, 2, \dots, l\}$$

$$c_i(x) \geq 0, i \in I = \{l+1, \dots, m\}$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

371

凸规划问题的充分条件

凸规划问题

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) = a_i^T x + b_i = 0, i \in E = \{1, \dots, l\}$$

$$c_i(x) \geq 0, i \in I = \{l+1, \dots, m\}$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

372

例题(KT点)

例4.1.3 已知约束问题

试验证最优点

$\min f(x) = -3x_1^2 - x_2^2 - 2x_3^2$ $x^* = (1, 1, 1)^T$ 为KT点.

$$\begin{aligned} \text{s.t. } c_1(x) &= x_1^2 + x_2^2 + x_3^2 - 3 = 0 \\ c_2(x) &= -x_1 + x_2 \geq 0 \\ c_3(x) &= x_1 \geq 0 \\ c_4(x) &= x_2 \geq 0 \\ c_5(x) &= x_3 \geq 0 \end{aligned}$$

解 $I^* = \{1, 2\}$

$$\nabla f(x^*) = (-6, -2, -4)^T,$$

$$\nabla c_1(x^*) = (2, 2, 2)^T,$$

$$\nabla c_2(x^*) = (-1, 1, 0)^T,$$

$$\text{令 } \begin{pmatrix} -6 \\ -2 \\ -4 \end{pmatrix} - \lambda_1 \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} - \lambda_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

得 $\lambda_1 = -2, \lambda_2 = 2$, 即

$$\nabla f(x^*) = -2\nabla c_1(x^*) + 2\nabla c_2(x^*) \quad \lambda_2 c_2(x^*) = 0, \lambda_2 > 0,$$

对于无效约束, 可取 $\lambda_i^* = 0$, 故 $x^* = (1, 1, 1)^T$ 为KT点.

373

§ 4.2 罚函数法与乘子法

对于一般的约束优化问题

$$\min f(x), x \in R^n$$

$$\text{s.t. } c_i(x) = 0, i \in E = \{1, \dots, l\}$$

$$c_i(x) \geq 0, i \in I = \{l+1, \dots, m\}$$

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.

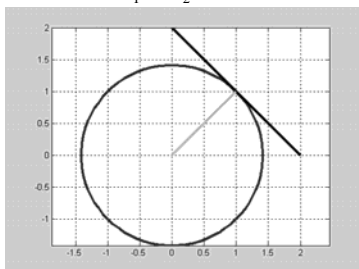
南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

374

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 $(1, 1)^T$.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

375

外罚函数法

设辅助函数为 $F(x_1, x_2) = \begin{cases} x_1^2 + x_2^2, & x_1 + x_2 = 2 \\ +\infty, & x_1 + x_2 \neq 2 \end{cases}$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

376

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 $\sigma \rightarrow +\infty$ 时, 有 $(x_1(\sigma), x_2(\sigma))^T \rightarrow (1, 1)^T = x^*$.

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 $\sigma \rightarrow +\infty$ 时也逼近于原问题的最优值 2.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

377

外罚函数法(等式约束问题)

$$\min f(x), x \in R^n$$

$$\text{s.t. } c_i(x) = 0, i \in E = \{1, 2, \dots, l\}$$

对于等式约束问题, 构造如下增广目标函数

$$P(x, \sigma) = f(x) + \sigma \sum_{i=1}^l |c_i(x)|^\beta, \beta \geq 1$$

其中 $\sigma > 0$ 为参数, 称为罚因子.

对于惩罚项 $\bar{P}(x) = \sum_{i=1}^l |c_i(x)|^\beta$,

当 x 为可行解时, $c_i(x) = 0, \bar{P}(x) = 0, P(x, \sigma) = f(x)$,

当 x 不是可行解时, $c_i(x) \neq 0, \bar{P}(x) > 0$,
 $P(x, \sigma) = f(x) + \sigma \bar{P}(x)$,

σ 越大, 惩罚越重.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

378

外罚函数法(等式约束问题)

$$P(x, \sigma) = f(x) + \sigma \bar{P}(x),$$

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 $\sigma \rightarrow \infty$ 时, $P(x, \sigma)$ 的极小点逼近于原问题的解.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

379

外罚函数法(不等式约束问题)

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) \geq 0, i \in I = \{1, 2, \dots, m\}$$

可以构造增广目标函数 $P(x, \sigma) = f(x) + \sigma \bar{P}(x)$,

其中
$$\bar{P}(x) = \begin{cases} 0 & c_i(x) \geq 0 \\ \sum_{i=1}^m |c_i(x)|^\alpha, \alpha \geq 1, c_i(x) < 0 \end{cases}$$

$$= \sum_{i=1}^m |\min(0, c_i(x))|^\alpha = \sum_{i=1}^m \left(\frac{|c_i(x)| - c_i(x)}{2} \right)^\alpha$$

“惩罚项”的作用与等式约束时的情形类似.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

380

例题(不等式约束问题)

例4.2.1求解约束问题

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

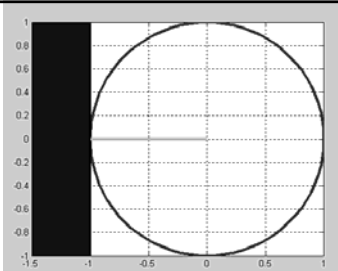
$$s.t. \quad x_1 + 1 \leq 0$$

问题是求点 $(0,0)^T$ 到半平面 $x \leq -1$ 的最短距离.

显然最优解为 $x^* = (-1, 0)^T$. 最优值为 $f(x^*) = 1$.

设增广目标函数为

$$P(x, \sigma) = x_1^2 + x_2^2 + \sigma [\min(0, -x_1 - 1)]^2$$



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

381

例题(不等式约束问题)

$$P(x, \sigma) = x_1^2 + x_2^2 + \sigma [\min(0, -x_1 - 1)]^2$$

$$= \begin{cases} x_1^2 + x_2^2, & x_1 + 1 \leq 0 \\ x_1^2 + x_2^2 + \sigma(x_1 + 1)^2, & x_1 + 1 > 0 \end{cases}$$

故
$$\frac{\partial P}{\partial x_1} = \begin{cases} 2x_1, & x_1 < -1 \\ 2x_1 + 2\sigma(x_1 + 1), & x_1 > -1 \end{cases} \quad \frac{\partial P}{\partial x_2} = 2x_2$$

令
$$\frac{\partial P}{\partial x_1} = \frac{\partial P}{\partial x_2} = 0 \quad \text{得} \quad x_1(\sigma) = -\frac{\sigma}{\sigma + 1}, x_2(\sigma) = 0.$$

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 $\sigma \rightarrow +\infty$ 时, $x_1(\sigma) \rightarrow -1, x_2(\sigma) \rightarrow 0$.

因此 $x(\sigma) \rightarrow x^*, P(x, \sigma) \rightarrow f(x^*) = 1$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

382

外罚函数法

在上面的两个例子中,当 $\sigma \rightarrow +\infty$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

383

外罚函数法(一般约束问题)

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) = 0, i \in E = \{1, 2, \dots, l\}$$

$$c_i(x) \geq 0, i \in I = \{l + 1, \dots, m\}$$

构造如下增广目标函数 $P(x, \sigma) = f(x) + \sigma \bar{P}(x)$,

其中
$$\bar{P}(x) = \sum_{i=1}^l |c_i(x)|^\beta + \sum_{j=l+1}^m |\min(0, c_j(x))|^\alpha \quad (\alpha \geq 1, \beta \geq 1)$$

$\bar{P}(x)$ 称为罚函数, $\sigma > 0$ 称为罚因子.

求解原问题转化为求解一系列的无约束问题

$$\min P(x, \sigma_k) (\sigma_k \rightarrow +\infty).$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

384

外罚函数法(算法步骤)

算法4.2.1 外罚函数法

取控制误差 $\varepsilon > 0$, 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

385

外罚函数法(收敛性)

对SUMT外点法产生的点列 $\{x_k\}$,

$$\begin{aligned} P(x_{k+1}, \sigma_{k+1}) &= f(x_{k+1}) + \sigma_{k+1} \tilde{P}(x_{k+1}) \\ &\geq f(x_{k+1}) + \sigma_k \tilde{P}(x_{k+1}) \quad \sigma_{k+1} \geq \sigma_k, \tilde{P}(x_{k+1}) \geq 0 \\ &= P(x_{k+1}, \sigma_k) \\ &\geq P(x_k, \sigma_k) \quad x_k \text{ 是 } P(x, \sigma_k) \text{ 的最优解.} \end{aligned}$$

因此有 $P(x_{k+1}, \sigma_{k+1}) \geq P(x_k, \sigma_k)$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

386

外罚函数法(收敛性)

$$f(x_{k+1}) + \sigma_k \tilde{P}(x_{k+1}) \geq f(x_k) + \sigma_k \tilde{P}(x_k) \quad x_k \text{ 是 } P(x, \sigma_k) \text{ 最优解.}$$

$$\Rightarrow f(x_{k+1}) - f(x_k) \geq \sigma_k [\tilde{P}(x_k) - \tilde{P}(x_{k+1})]$$

$$f(x_k) + \sigma_{k+1} \tilde{P}(x_k) \geq f(x_{k+1}) + \sigma_{k+1} \tilde{P}(x_{k+1}) \quad x_{k+1} \text{ 是 } P(x, \sigma_{k+1}) \text{ 最优解.}$$

$$\Rightarrow \sigma_{k+1} [\tilde{P}(x_k) - \tilde{P}(x_{k+1})] \geq f(x_{k+1}) - f(x_k)$$

$$\text{因此 } \sigma_{k+1} [\tilde{P}(x_k) - \tilde{P}(x_{k+1})] \geq \sigma_k [\tilde{P}(x_k) - \tilde{P}(x_{k+1})]$$

$$(\sigma_{k+1} - \sigma_k) [\tilde{P}(x_k) - \tilde{P}(x_{k+1})] \geq 0 \Rightarrow \tilde{P}(x_k) \geq \tilde{P}(x_{k+1}) \quad \sigma_{k+1} \geq \sigma_k$$

$$\text{再根据上面(绿色)标出的不等式, 有 } f(x_{k+1}) \geq f(x_k)$$

我们得到书中P157引理4.2.1

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

387

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

388

外罚函数法(收敛性)

$$\begin{aligned} \min f(x), x \in R^n \\ \text{s.t. } c_i(x) = 0, i \in E = \{1, 2, \dots, l\} \quad P(x, \sigma_k) = f(x) + \sigma_k \tilde{P}(x), \\ c_i(x) \geq 0, i \in I = \{l+1, \dots, m\} \end{aligned}$$

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 $\{\sigma_k\}$, $\sigma_{k+1} \geq \sigma_k$, 且 $\sigma_k \rightarrow +\infty$, 则由SUMT外点法产生的点列 $\{x_k\}$ 的任何聚点 \bar{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 $1, 0, -1, 1, 0, -1, 1, 0, -1, \dots$, 有三个聚点 $1, 0, -1$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

389

外罚函数法(收敛性)

证明 不妨设 $x_k \rightarrow \bar{x}$. 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

$$f(x^*) = f(x^*) + \sigma_k \tilde{P}(x^*)$$

$$\geq f(x_k) + \sigma_k \tilde{P}(x_k) = P(x_k, \sigma_k)$$

$$P(x_{k+1}, \sigma_{k+1}) \geq P(x_k, \sigma_k)$$

$$P(x_k, \sigma_k) \text{ 存在极限, 设为 } p^0.$$

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

$$f(x_k) \text{ 存在极限, 设为 } f^0.$$

$$\sigma_k \tilde{P}(x_k) = P(x_k, \sigma_k) - f(x_k) \rightarrow p^0 - f^0$$

$$\tilde{P}(x_k) \rightarrow 0$$

$$x_k \text{ 是 } P(x, \sigma_k) \text{ 最优解.}$$

$$\text{引理4.2.1}$$

$$\text{单调有界数列必收敛}$$

$$\text{引理4.2.1}$$

$$\text{单调有界数列必收敛}$$

$$\sigma_k \rightarrow +\infty$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

390

外罚函数法(收敛性)

由于 $x_k \rightarrow \underline{x}$,且 $\tilde{P}(x)$ 连续,因此 $\tilde{P}(\underline{x}) \rightarrow 0$.
即 \underline{x} 为可行解.而 x^* 为整体最优解,所以 $f(x^*) \leq f(\underline{x})$.
根据 $x_k \rightarrow \underline{x}$ 以及 $f(x_k) \leq P(x_k, \sigma_k) \leq f(x^*)$
得到 $f(\underline{x}) \leq f(x^*)$.
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 $\sigma_k \tilde{P}(x_k) \rightarrow 0$.
这是在算法中取 $\sigma_k \tilde{P}(x_k) < \varepsilon$ 作为终止准则的原因.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

391

算例(外罚函数法)

例4.2.3 $\min f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$
求解约束问题 s.t. $x_1^2 - x_2 = 0$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

392

算例(外罚函数法)

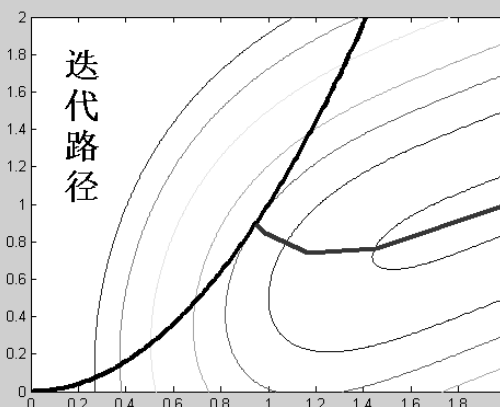
k	x_k	$f(x_k)$	$\sigma_k \tilde{P}(x_k)$
0	$(2, 1)^T$	0	0.9
1	$(1.45388, 0.76076)^T$	0.09353	1.83058
2	$(1.16872, 0.74067)^T$	0.57524	3.90930
3	$(0.99061, 0.84246)^T$	1.52013	1.92822
4	$(0.95076, 0.88749)^T$	1.89123	0.27170
5	$(0.94611, 0.89344)^T$	1.94052	0.02828
6	$(0.94563, 0.89405)^T$	1.94562	0.00284
7	$(0.94556, 0.89409)^T$	1.94619	6.8 e-6

精确解 $(0.94558299, 0.89412720)^T$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

393

迭代
路径



内罚函数法

$\min f(x), x \in R^n$
s.t. $c_i(x) \geq 0, i \in I = \{1, 2, \dots, m\}$

考虑上面的不等式约束问题.当 x 从可行域

$D = \{x \in R^n \mid c_i(x) \geq 0, i = 1, 2, \dots, m\}$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

395

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

396

内罚函数法(算法步骤)

算法4.2.2 内罚函数法

取控制误差 $\varepsilon > 0$, 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

397

内罚函数法(算例)

例4.2.4 用内点法求解 $\min f(x) = \frac{1}{3}(x_1 + 1)^3 + x_2$

解 增广目标函数为 $s.t. \begin{cases} 1 - x_1 \leq 0 \\ x_2 \geq 0 \end{cases}$

$$B(x_1, x_2, r) = \frac{1}{3}(x_1 + 1)^3 + x_2 + r\left(\frac{1}{x_1 - 1} + \frac{1}{x_2}\right).$$

$$\text{令 } \frac{\partial B}{\partial x_1} = (x_1 + 1)^2 - \frac{r}{(x_1 - 1)^2} = 0$$

$$\frac{\partial B}{\partial x_2} = 1 - \frac{r}{x_2^2} = 0$$

当 $r \rightarrow 0$ 时, 得 $x^* = (1, 0)^T$.

得 $x(r) = (\sqrt{1 + \sqrt{r}}, \sqrt{r})^T, f^* = 8/3$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

398

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

399

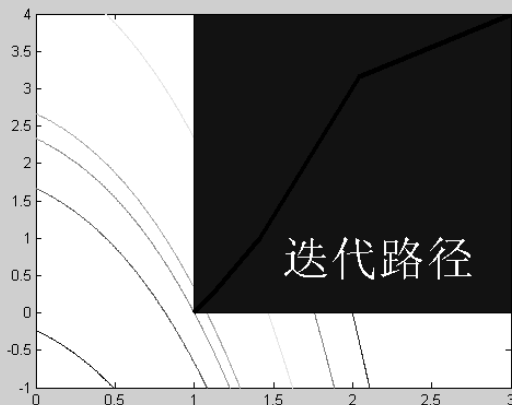
算例(内罚函数法)

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 $(1, 0)^T$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

400



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 $B(x_{k+1}, r_{k+1}) \leq B(x_k, r_k)$.

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 $r_k \rightarrow 0$, 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

402

4.2.4 乘子法

$$\begin{aligned} \min f(x), x \in R^n \\ \text{s.t. } c_i(x) = 0, i \in E = \{1, 2, \dots, l\} \end{aligned}$$

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 $L(x, \lambda) = f(x) - \sum_{j=1}^l \lambda_j c_j(x)$

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

403

乘子法

例4.2.5 求解约束问题 $\min f(x) = x_1^2 - 3x_2 - x_2^2$
 $\text{s.t. } x_2 = 0$

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 $\lambda, L(x, \lambda)$ 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 $\sigma \rightarrow \infty$,在一定条件下求得原问题的解。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

404

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

405

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 $\min f(x) = x_1^2 - 3x_2 - x_2^2$
 $\text{s.t. } x_2 = 0$

$$L(x, \lambda) = x_1^2 - (\lambda + 3)x_2 - x_2^2 \quad \text{取} \quad \tilde{P}(x) = x_2^2 / 2$$

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 $\lambda^* = -3, \sigma \geq \sigma^* = 2$ 时,原问题最优解 $(0,0)^T$ 是增广Lagrange函数的最优解。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

406

等式约束问题的乘子法

$$\begin{aligned} \min f(x) = x_1^2 - 3x_2 - x_2^2 \\ \text{s.t. } x_2 = 0 \end{aligned} \quad M(x, \lambda, \sigma) = x_1^2 - (\lambda + 3)x_2 + \frac{\sigma - 2}{2} x_2^2$$

反之,求解无约束问题

$$\begin{aligned} \min M(x, \lambda, \sigma) = x_1^2 + \frac{\sigma - 2}{2} x_2^2 - (\lambda + 3)x_2 \\ \text{令 } 0 = \frac{\partial M}{\partial x_1} = 2x_1, \quad \text{得 } x_0 = (0, \frac{\lambda + 3}{\sigma - 2})^T. \\ 0 = \frac{\partial M}{\partial x_2} = (\sigma - 2)x_2 - (\lambda + 3). \end{aligned}$$

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

407

等式约束问题的乘子法

考虑等式约束问题 $\min f(x), x \in R^n$
 $\text{s.t. } C(x) = 0$

其中 $C(x) = (c_1(x), \dots, c_l(x))^T$,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 $L(x^*, \lambda^*) = f(x^*) - \lambda^{*T} C(x^*) = f(x^*)$
 $\leq f(x) - \lambda^{*T} C(x) = L(x, \lambda^*)$

因此,原约束问题等价于下面的约束问题

$$\begin{aligned} \min L(x, \lambda^*) \\ \text{s.t. } C(x) = 0 \end{aligned}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

408

等式约束问题的乘子法

$\min L(x, \lambda^*)$ 对左边问题,构造增广Lagrange函数
 $s.t. \quad C(x) = 0 \quad M(x, \lambda, \sigma) = L(x, \lambda) + \frac{\sigma}{2} C(x)^T C(x)$

定理4.2.6 设上面等式约束问题中, $x^* \in R^n$ 和 $\lambda^* \in R^m$ 满足二阶充分条件(Th4.1.2), 则存在一个数 $\sigma^* > 0$, 对所有的 $\sigma \geq \sigma^*$, 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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

409

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 $x_k \rightarrow x^*, \lambda_k \rightarrow \lambda^*$ 。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

410

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 $x_k \rightarrow x^*, \lambda_k \rightarrow \lambda^*$, 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 $C(x_k) \rightarrow 0$ 。

若 $x_k \rightarrow x^*$, 则有 $C(x^*) = 0$, 即 x^* 为可行解。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

411

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

412

例题(等式约束问题的乘子法)

例4.2.7 求解约束问题 $\min f(x) = x_1^2 + x_2^2$
 $s.t. \quad x_1 + x_2 - 2 = 0$

解 增广Lagrange函数为

$$M(x_1, x_2, \lambda, \sigma) = x_1^2 + x_2^2 - \lambda(x_1 + x_2 - 2) + \frac{\sigma}{2}(x_1 + x_2 - 2)^2$$

$$\text{令 } \frac{\partial M}{\partial x_1} = 2x_1 - \lambda + \sigma(x_1 + x_2 - 2) = 0$$

$$\frac{\partial M}{\partial x_2} = 2x_2 - \lambda + \sigma(x_1 + x_2 - 2) = 0$$

$$\text{得 } x_1 = x_2 = \frac{2\sigma + \lambda}{2\sigma + 2}$$

将其中的 λ 视为 λ_k ,
 根据乘子迭代公式
 $\lambda_{k+1} = \lambda_k - \sigma(x_1 + x_2 - 2)$
 $= \frac{1}{\sigma+1} \lambda_k + \frac{2\sigma}{\sigma+1}$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

413

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 $\lambda_k \rightarrow \lambda^*$, 对上式取极限得

$$\lambda^* = \frac{1}{\sigma+1} \lambda^* + \frac{2\sigma}{\sigma+1}$$

因此 $\lambda^* = 2$ 。

在 $x_1 = x_2 = \frac{2\sigma + \lambda}{2\sigma + 2}$ 中, 令 $\lambda = 2$,

得原问题的最优解 $x^* = (1, 1)^T$ 。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

414

不等式约束问题的乘子法

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) \geq 0, i \in I = \{1, 2, \dots, m\}$$

对不等式约束问题, 引进辅助变量 $z_i (i=1, 2, \dots, m)$, 上面的问题转化为等价的等式约束问题

$$\min f(x), x \in R^n$$

$$s.t. \quad c_i(x) - z_i^2 = 0, i = 1, 2, \dots, m.$$

其增广Lagrange函数为

$$\tilde{M}(x, z, \lambda, \sigma) = f(x) - \sum_{i=1}^m \lambda_i (c_i(x) - z_i^2) + \frac{\sigma}{2} \sum_{i=1}^m (c_i(x) - z_i^2)^2.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

415

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 $\sigma c_i(x) - \lambda_i \geq 0$ 时, 要使函数取最小, $z_i^2 = c_i(x) - \frac{\lambda_i}{\sigma}$
 否则 $z_i = 0$. 因此 $z_i^2 = \frac{1}{\sigma} \max(0, \sigma c_i(x) - \lambda_i), i = 1, 2, \dots, m$
 得到增广的目标函数

$$M(x, \lambda, \sigma) = f(x) + \frac{1}{2\sigma} \sum_{i=1}^m \{[\max(0, \lambda_i - \sigma c_i(x))]^2 - \lambda_i^2\}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

416

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

$$(\lambda_{k+1})_i = \max[0, (\lambda_k)_i - \sigma c_i(x_k)], i = 1, 2, \dots, m$$

$$\text{终止准则} \quad \left(\sum_{i=1}^m [\max(c_i(x_k), \frac{(\lambda_k)_i}{\sigma})]^2 \right)^{1/2} < \varepsilon$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

417

例题(PHR算法)

例4.2.8 用PHR算法求解 $\min f(x) = x_1^2 + x_2^2$
 $s.t. \quad x_1 + x_2 \geq 2$

解 增广目标函数为

$$M(x_1, x_2, \lambda, \sigma) = x_1^2 + x_2^2 + \frac{1}{2\sigma} \{[\max(0, \lambda - \sigma(x_1 + x_2 - 2))]^2 - \lambda^2\}$$

$$= \begin{cases} x_1^2 + x_2^2 - \frac{\lambda^2}{2\sigma}, & x_1 + x_2 - 2 > \frac{\lambda}{\sigma} \\ x_1^2 + x_2^2 + \frac{1}{2\sigma} \{[\lambda - \sigma(x_1 + x_2 - 2)]^2 - \lambda^2\}, & x_1 + x_2 - 2 \leq \frac{\lambda}{\sigma} \end{cases}$$

当 $x_1 + x_2 - 2 > \frac{\lambda}{\sigma}$ 时, 令 $\frac{\partial M}{\partial x_1} = 2x_1 = 0, \frac{\partial M}{\partial x_2} = 2x_2 = 0$.

得 $\bar{x} = (0, 0)^T$. 当 σ 充分大时, 该点不满足 $x_1 + x_2 - 2 > \frac{\lambda}{\sigma}$.
 从而 $\bar{x} = (0, 0)^T$ 不是极小点.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

418

例题(PHR算法)

$$M = x_1^2 + x_2^2 + \frac{1}{2\sigma} \{[\lambda - \sigma(x_1 + x_2 - 2)]^2 - \lambda^2\}, x_1 + x_2 - 2 \leq \frac{\lambda}{\sigma}$$

当 $x_1 + x_2 - 2 \leq \frac{\lambda}{\sigma}$ 时, 令

$$\frac{\partial M}{\partial x_1} = 2x_1 - [\lambda - \sigma(x_1 + x_2 - 2)] = 0,$$

$$\frac{\partial M}{\partial x_2} = 2x_2 - [\lambda - \sigma(x_1 + x_2 - 2)] = 0,$$

$$\text{得} \quad \bar{x} = \left(\frac{2\sigma + \lambda}{2\sigma + 2}, \frac{2\sigma + \lambda}{2\sigma + 2} \right)^T.$$

当 σ 充分大时, 该点满足

$$x_1 + x_2 - 2 \leq \frac{\lambda}{\sigma}.$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

419

§ 4.3 投影梯度法

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

420

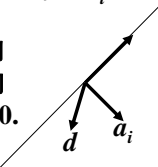
4.3.1 可行方向及其性质

$\min f(x), x \in R^n$ 本节研究左边的线性不等式约束问题。
 $s.t. A^T x \geq b$

对于线性约束 $a_i^T x \geq b_i$, 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 $a_i^T d \geq 0$ 。

若 $a_i^T d = 0$, 则沿该方向的点在边界上。



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

421

投影梯度法

例4.3.2 求解

$$\min f(x) = 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2$$

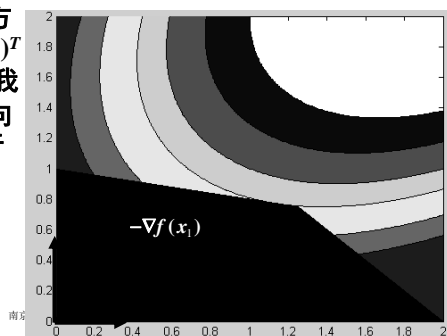
$$s.t. \begin{cases} 2 - x_1 - x_2 \geq 0 \\ 5 - x_1 - 5x_2 \geq 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

422

$\nabla f(x) = (4x_1 - 2x_2 - 4, 4x_2 - 2x_1 - 6)^T$ $-\nabla f(x_1) = (4, 6)^T$
 负梯度指向可行域内部 $-\nabla f(x_1) = 4(1, 0)^T + 6(0, 1)^T$
 仿照单纯形方法, 我们沿着边界使函数值下降。

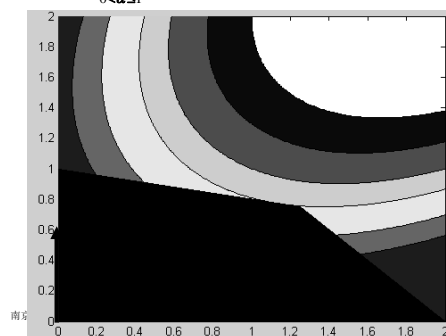
由于负梯度方向沿方向 $(0, 1)^T$ 的分量较大, 我们先沿此方向 $p_1 = (0, 1)^T$ 进行一维搜索。



$p_1 = (0, 1)^T$ 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 $\min_{0 \leq \alpha \leq 1} f(x_1 + \alpha p_1) = 72\alpha^2 - 36\alpha$

得 $\alpha_1 = 1$,
 $x_2 = x_1 + \alpha_1 p_1 = (0, 1)^T$ 。

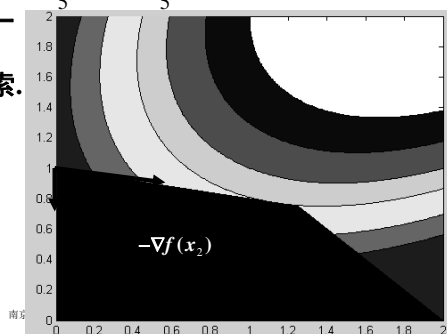


$\nabla f(x) = (4x_1 - 2x_2 - 4, 4x_2 - 2x_1 - 6)^T$ $x_2 = (0, 1)^T$ $-\nabla f(x_2) = (6, 2)^T$
 在 x_2 处沿边界指向可行域的方向为 $(0, -1)^T$ 和 $(5, -1)^T$ 。

$$(6, 2)^T = \frac{6}{5}(5, -1)^T - \frac{16}{5}(0, -1)^T$$

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 $5/4$ 。

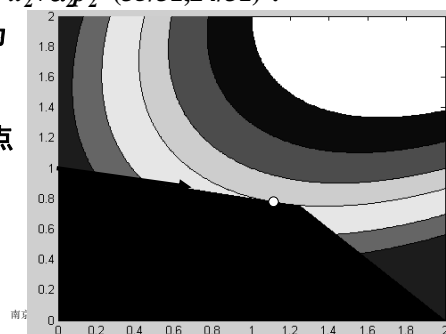


作线性搜索, $\min_{0 \leq \alpha \leq 5/4} f(x_2 + \alpha p_2) = 62\alpha^2 - 28\alpha - 4$ 即求解

得 $\alpha_2 = 7/31$, $x_3 = x_2 + \alpha_2 p_2 = (35/31, 24/31)^T$ 。

可以验证 x_3 为 KT 点。

又 $f(x)$ 为凸函数, x_3 为最优解。



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?

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

427

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

428

情形一

$\nabla f(x)$ 可以写为 $\{a_i (i \in I^*)\}$ 的线性组合.

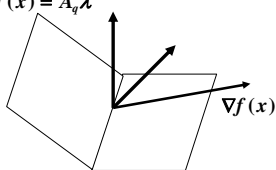
令 $\nabla f(x) = \sum_{i \in I^*} \lambda_i a_i$, 如果 $\lambda_i \geq 0$,则 x 为KT点.

如何求组合系数 λ_i ?

$$\nabla f(x) = \sum_{i \in I^*} \lambda_i a_i \text{ 可以写为 } \nabla f(x) = A_q \lambda$$

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 $\lambda = A_q^{-1} \nabla f(x)$.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

429

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 $A A^+ = I$ 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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

430

情形二

$\nabla f(x)$ 不能写为 $\{a_i (i \in I^*)\}$ 的线性组合.

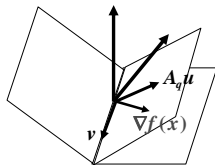
此时, $\nabla f(x)$ 可以分解成两个正交的向量,一个可以写为上述向量的线性组合,一个垂直于上述向量,即

$$\nabla f(x) = A_q u + v, A_q^T v = 0 (v \neq 0).$$

$A_q u$ 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.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

431

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

$$\begin{aligned} A_q A_q^+ \nabla f(x) &= A_q A_q^+ A_q u + A_q A_q^+ v \\ &= A_q A_q^+ A_q u + A_q (A_q^T A_q)^{-1} A_q^T v \\ &= A_q u. \end{aligned} \quad \boxed{A_q^+ A_q = I} \quad \boxed{A_q^T v = 0}$$

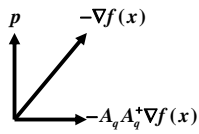
因此 $A_q u = A_q A_q^+ \nabla f(x), v = (I - A_q A_q^+) \nabla f(x)$.

$$\nabla f(x) = A_q A_q^+ \nabla f(x) + (I - A_q A_q^+) \nabla f(x).$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

432

情形二的可行下降方向



$$\nabla f(x) = A_q A_q^+ \nabla f(x) + (I - A_q A_q^+) \nabla f(x).$$

$$\text{令 } p = -(I - A_q A_q^+) \nabla f(x),$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

433

情形二的可行下降方向

$$p = -(I - A_q A_q^+) \nabla f(x),$$

考察矩阵 $P_q = I - A_q A_q^+$

$$\text{有 } P_q^2 = (I - A_q A_q^+)^2 = I - 2A_q A_q^+ + A_q A_q^+ A_q A_q^+ = I - A_q A_q^+ \quad A_q^+ A_q = I$$

$$\text{另外 } P_q^T = (I - A_q A_q^+)^T = (I - A_q (A_q^T A_q)^{-1} A_q^T)^T = P_q.$$

$$\text{因此 } p^T \nabla f(x) = -\nabla f(x)^T P_q^T \nabla f(x) \quad P_q^T = P_q = P_q^2 = P_q^T P_q \\ = -\nabla f(x)^T P_q^T P_q \nabla f(x) = -\|P_q \nabla f(x)\|^2 = -\|p\|^2 < 0.$$

方向 p 是下降方向.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

434

情形二的可行下降方向

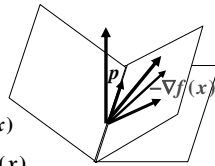
$$p = -P_q \nabla f(x) = -(I - A_q A_q^+) \nabla f(x)$$

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.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

435

情形二的可行下降方向

$$p = -P_q \nabla f(x) = -(I - A_q A_q^+) \nabla f(x) \quad A_q^T p = 0$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

436

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 $a_i^T x \geq b_i$, 将点 $x + \alpha p$ 代入此不等式有

$$a_i^T (x + \alpha p) \geq b_i \quad \text{即} \quad (a_i^T x - b_i) + \alpha a_i^T p \geq 0$$

由于 $a_i^T x - b_i \geq 0$, 在 $a_i^T p \geq 0$ 时, 上式对 $\alpha \geq 0$ 成立.

若 $a_i^T p < 0$, 则有 $0 \leq \alpha \leq \frac{a_i^T x - b_i}{-a_i^T p}$

因此 α 的取值上限为 $\alpha_{\max} = \min\{\frac{a_i^T x - b_i}{-a_i^T p} \mid a_i^T p < 0\}$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

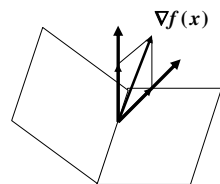
437

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 $\lambda \geq 0$, 显然 x 为 KT 点.



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

438

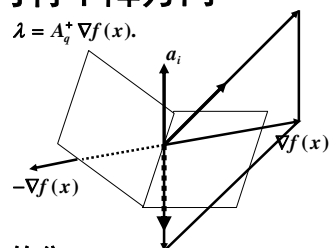
情形一的可行下降方向

$$\nabla f(x) = A_q \lambda, \text{ 其中 } \lambda = A_q^+ \nabla f(x).$$

若 λ 至少有一个分量小于0, 设为 $\lambda_i < 0$.

$$\begin{aligned} \nabla f(x) &= A_q \lambda = \sum_{j \in I^*} \lambda_j a_j \\ &= \lambda_i a_i + \sum_{j \in I^*, j \neq i} \lambda_j a_j \end{aligned}$$

此时负梯度投影到 a_i 上的分量是指向可行域之外的。



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

439

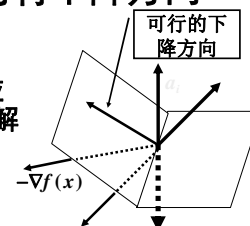
情形一的可行下降方向

$$\nabla f(x) = \lambda_i a_i + \sum_{j \in I^*, j \neq i} \lambda_j a_j$$

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 $a_i (j \in I^*, j \neq i)$ 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。



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

440

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

$$-\nabla f(x) = -A_{q-1} A_{q-1}^+ \nabla f(x) - (I - A_{q-1} A_{q-1}^+) \nabla f(x).$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

441

情形一的可行下降方向

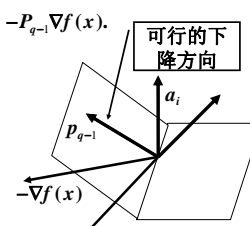
$$-\nabla f(x) = -A_{q-1} A_{q-1}^+ \nabla f(x) - (I - A_{q-1} A_{q-1}^+) \nabla f(x).$$

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) $p_{q-1}^T a_j = 0 (j \in I^*, j \neq i)$

$$p_{q-1}^T a_i > 0$$



南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

442

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 $p_{q-1} \neq 0$.

$$\begin{aligned} p_{q-1} &= -P_{q-1} \nabla f(x) = -(I - A_{q-1} A_{q-1}^+) (A_{q-1} \lambda_{q-1} + \lambda_i a_i) \\ &= \lambda_i (A_{q-1} A_{q-1}^+ a_i - a_i) \end{aligned}$$

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

443

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 $P_{q-1}^2 = P_{q-1}$, $P_{q-1}^T = P_{q-1}$.

$$\begin{aligned} \text{因此 } p_{q-1}^T \nabla f(x) &= -\nabla f(x)^T P_{q-1}^T \nabla f(x) = -\nabla f(x)^T P_{q-1} P_{q-1} \nabla f(x) \\ &= -\|P_{q-1} \nabla f(x)\|^2 = -\|p_{q-1}\|^2 < 0. \end{aligned}$$

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。

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

444

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.

$$\begin{aligned} A_{q-1}^T p_{q-1} &= -A_{q-1}^T (I - A_{q-1} (A_{q-1}^T A_{q-1})^{-1} A_{q-1}^T) \nabla f(x) \\ &= (-A_{q-1}^T + A_{q-1}^T A_{q-1} (A_{q-1}^T A_{q-1})^{-1} A_{q-1}^T) \nabla f(x) = 0. \end{aligned}$$

$$p_{q-1} = \lambda_i (A_{q-1} A_{q-1}^+ a_i - a_i)$$

$$a_i^T p_{q-1} = \lambda_i a_i^T A_{q-1} A_{q-1}^+ a_i - \lambda_i a_i^T a_i = -\lambda_i a_i^T a_i > 0$$

关于 x 点处的有效约束对应的方向 a_i 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 $P_q = I - A_q A_q^+$.

$P_q \nabla f(x) \neq 0 \implies$ 情形二

以 $-P_q \nabla f(x)$ 为可行下降方向进行迭代

$P_q \nabla f(x) = 0 \implies$ 情形一

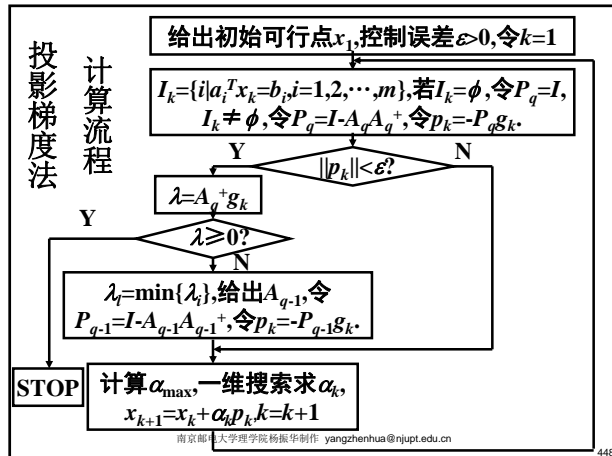
求乘子向量 λ , 若 $\lambda \geq 0$, 终止迭代

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

投影
梯度
法

计算
流程



算例(投影梯度法)

例4.3.2 求解

$$\begin{aligned} \min f(x) &= 2x_1^2 + 2x_2^2 - 2x_1x_2 - 4x_1 - 6x_2 \\ \text{s.t. } \begin{cases} 2 - x_1 - x_2 \geq 0 \\ 5 - x_1 - 5x_2 \geq 0 \\ x_1 \geq 0 \\ x_2 \geq 0 \end{cases} \quad \text{解 } A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix} \\ \nabla f(x) &= (4x_1 - 2x_2 - 4, 4x_2 - 2x_1 - 6)^T \end{aligned}$$

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 $P_2 = I - A_2 A_2^+ = I - A_2 A_2^{-1} = O$. $P_2 \nabla f(x_1) = 0$.

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 $\lambda_2 = -6$, 从 A_2 取中去掉第二列, 得到

$$A_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad A_1^+ = (A_1^T A_1)^{-1} A_1^T = A_1^T = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

$$P_1 = I - A_1 A_1^+ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$p_1 = -P_1 \nabla f(x_1) = -\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -4 \\ -6 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

算例(投影梯度法)

$$x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad p_1 = \begin{bmatrix} 0 \\ 6 \end{bmatrix} \quad A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix} \quad b = (-2, -5, 0, 0)^T$$

$$a_1^T p_1 = -6, a_2^T p_1 = -30,$$

$$a_1^T x_1 - b_1 = 2, a_2^T x_2 - b_2 = 5. \quad \alpha_{\max} = \min\left\{\frac{2}{6}, \frac{5}{30}\right\} = \frac{1}{6}$$

作线性搜索

$$\min f(x_1 + \alpha p_1) = 72\alpha^2 - 36\alpha, 0 < \alpha \leq \alpha_{\max} = 1/6$$

$$\text{得 } \alpha_1 = \frac{1}{6}, \quad x_2 = x_1 + \alpha_1 p_1 = (0, 1)^T.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

451

α_{\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 α_{\max} .

$$\begin{bmatrix} -1 & -1 \\ -1 & -5 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} -1 & -1 \\ -1 & -5 \end{bmatrix} \begin{bmatrix} 0 \\ 6 \end{bmatrix} \geq \begin{bmatrix} -2 \\ -5 \end{bmatrix}$$

$$\text{即 } \begin{bmatrix} -6\alpha \\ -30\alpha \end{bmatrix} \geq \begin{bmatrix} -2 \\ -5 \end{bmatrix}$$

$$\text{得 } \alpha \leq \frac{1}{6}. \quad \text{因此 } \alpha_{\max} = \frac{1}{6}.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

452

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 $\nabla f(x_2) = (-6, -2)^T, I_2 = \{2, 3\}$

$$A_2 = \begin{bmatrix} -1 & 1 \\ -5 & 0 \end{bmatrix} \quad \text{可逆. } P_2 = O, P_2 \nabla f(x_2) = 0.$$

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 $\lambda_2 = -28/5 < 0$, 从 A_2 取中去掉第二列, 得到

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

453

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 $\lambda_2 = -28/5 < 0$, 从 A_2 取中去掉第二列, 得到

$$A_1 = \begin{bmatrix} -1 \\ -5 \end{bmatrix} \quad A_1^+ = (A_1^T A_1)^{-1} A_1^T = \begin{bmatrix} -\frac{1}{26} & -\frac{5}{26} \end{bmatrix}$$

$$P_1 = I - A_1 A_1^+ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -1 \\ -5 \end{bmatrix} \begin{bmatrix} -\frac{1}{26} & -\frac{5}{26} \end{bmatrix} = \begin{bmatrix} 25/26 & -5/26 \\ -5/26 & 1/26 \end{bmatrix}$$

$$p_1 = -P_1 \nabla f(x_1) = -\begin{bmatrix} 25/26 & -5/26 \\ -5/26 & 1/26 \end{bmatrix} \begin{bmatrix} -6 \\ -2 \end{bmatrix} = \begin{bmatrix} 70/13 \\ -14/13 \end{bmatrix}$$

不妨取 $p_2 = (5, -1)^T$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

454

算例(投影梯度法)

$$x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad p_2 = \begin{bmatrix} 5 \\ -1 \end{bmatrix} \quad A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix} \quad b = (-2, -5, 0, 0)^T$$

$$a_1^T p_2 = -4, a_4^T p_2 = -1,$$

$$a_1^T x_2 - b_1 = 1, a_4^T x_2 - b_4 = 1. \quad \alpha_{\max} = \min\left\{\frac{1}{4}, \frac{1}{1}\right\} = \frac{1}{4}$$

作线性搜索

$$\min f(x_2 + \alpha p_2) = 62\alpha^2 - 28\alpha - 4, 0 < \alpha \leq \alpha_{\max} = 1/4$$

$$\text{得 } \alpha_2 = \frac{7}{31}, \quad x_3 = x_2 + \alpha_2 p_2 = (35/31, 24/31)^T.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

455

算例(投影梯度法)

$$x_3 = \begin{bmatrix} 35/31 \\ 24/31 \end{bmatrix} \quad A = \begin{bmatrix} -1 & -1 & 1 & 0 \\ -1 & -5 & 0 & 1 \end{bmatrix} \quad b = (-2, -5, 0, 0)^T$$

第三次迭代 $\nabla f(x_3) = (-32/31, -160/31)^T, I_3 = \{2\}$

$$A_1 = \begin{bmatrix} -1 \\ -5 \end{bmatrix} \quad A_1^+ = \begin{bmatrix} -\frac{1}{26} & -\frac{5}{26} \end{bmatrix} \quad P_1 = \begin{bmatrix} 25/26 & -5/26 \\ -5/26 & 1/26 \end{bmatrix}$$

$$P_1 \nabla f(x_3) = -\begin{bmatrix} 25/26 & -5/26 \\ -5/26 & 1/26 \end{bmatrix} \begin{bmatrix} -32/31 \\ -160/31 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\lambda = A_1^+ \nabla f(x_3) = \begin{bmatrix} -1/26 & -5/26 \end{bmatrix} \begin{bmatrix} -32/31 \\ -160/31 \end{bmatrix} = 32/31 > 0$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

456

§ 4.4 约束变尺度法

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

457

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

$$\min f(x) = \frac{1}{2} x^T G x + C^T x$$

$$(QP) \text{ s.t. } \begin{cases} c_i(x) = a_i^T x - b_i = 0, i \in E = \{1, \dots, l\} \\ c_i(x) = a_i^T x - b_i \geq 0, i \in I = \{l+1, \dots, m\} \end{cases}$$

其中 G 为 n 阶对称矩阵.当 G 正定时,(QP)为严格凸二次规划.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

458

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 $\lambda^*=(\lambda_1^*, \dots, \lambda_m^*)$,使得

$$Gx^* + C - \sum_{i \in E} \lambda_i^* a_i - \sum_{i \in I} \lambda_i^* a_i = 0$$

$$\begin{cases} a_i^T x^* - b_i = 0, i \in E \\ a_i^T x^* - b_i \geq 0, i \in I \\ \lambda_i^* \geq 0, i \in I \\ \lambda_i^* = 0, i \in I \setminus I^* \end{cases}$$

其中 I^* 是 x^* 处的有效集.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

459

等式约束二次规划

对于仅含等式约束的严格凸二次规划

$$\min f(x) = \frac{1}{2} x^T G x + C^T x$$

$$\text{s.t. } c_i(x) = a_i^T x - b_i = 0, i \in E = \{1, \dots, l\}$$

其中 $A=(a_1, \dots, a_l)$ 的秩为 l .

显然, x 是上述问题的解的充要条件是

$$\begin{cases} Gx + C - \sum_{i \in E} \lambda_i a_i = 0, \\ a_i^T x - b_i = 0, i \in E. \end{cases}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

460

等式约束二次规划

$$\begin{cases} Gx + C - \sum_{i \in E} \lambda_i a_i = 0, \\ a_i^T x - b_i = 0, i \in E. \end{cases}$$

令 $b=(b_1, \dots, b_l)^T, \lambda=(\lambda_1, \dots, \lambda_l)^T$,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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

461

等式约束二次规划(算例)

例4.4.1 求解严格凸二次规划

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

$$(QP) \text{ s.t. } \begin{cases} x_1 + 2x_2 - x_3 - 4 = 0 \\ x_1 - x_2 + x_3 + 2 = 0 \end{cases}$$

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

$$\left(\frac{2}{7}, \frac{10}{7}, -\frac{6}{7}, \frac{8}{7}, -\frac{4}{7}\right)^T.$$

因此最优解

$$x^* = (2/7, 10/7, -6/7)^T,$$

最优乘子向量

$$\lambda^* = (8/7, -4/7)^T.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

462

严格凸二次规划的有效集方法

$$\min f(x) = \frac{1}{2}x^T Gx + C^T x \quad (4.82)$$

$$(QP) \text{ s.t. } \begin{aligned} c_i(x) &= a_i^T x - b_i = 0, i \in E = \{1, \dots, l\} \\ c_i(x) &= a_i^T x - b_i \geq 0, i \in I = \{l+1, \dots, m\} \end{aligned}$$

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

$$\min f(x) = \frac{1}{2}x^T Gx + C^T x \quad (4.86)$$

$$\text{s.t. } c_i(x) = a_i^T x - b_i = 0, i \in I^*$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

463

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

$$\begin{aligned} Gx^* + C - \sum_{i \in E} \lambda_i^* a_i - \sum_{i \in I} \lambda_i^* a_i &= 0 \\ a_i^T x^* - b_i &\geq 0, i \in I \\ \lambda_i^* &= 0, i \in I \setminus I^* \end{aligned}$$

$$\text{显然有 } Gx^* + C - \sum_{i \in I^*} \lambda_i^* a_i = 0$$

这是问题(4.86)的最优解的充要条件.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

464

有效集方法

例:解二次规划

$$\min f(x) = (x_1 - 1)^2 + (x_1 - 2.5)^2$$

$$\text{s.t. } x_1 - 2x_2 + 2 \geq 0$$

$$-x_1 - 2x_2 + 6 \geq 0$$

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

$$x_1 \geq 0$$

$$x_2 \geq 0$$

解: 取初始可行点

$$x_0 = (2, 0)^T. \text{有效集为}$$

$$I_0 = \{3, 5\}.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

465

有效集方法

我们首先来求解等式约束二次规划

$$\min f(x) = (x_1 - 1)^2 + (x_1 - 2.5)^2$$

$$\text{s.t. } -x_1 + 2x_2 + 2 = 0$$

$$x_2 = 0$$

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.KT点.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

466

有效集方法

$$g_1 = (2, -5)^T, a_3 = (-1, 2)^T, a_5 = (0, 1)^T,$$

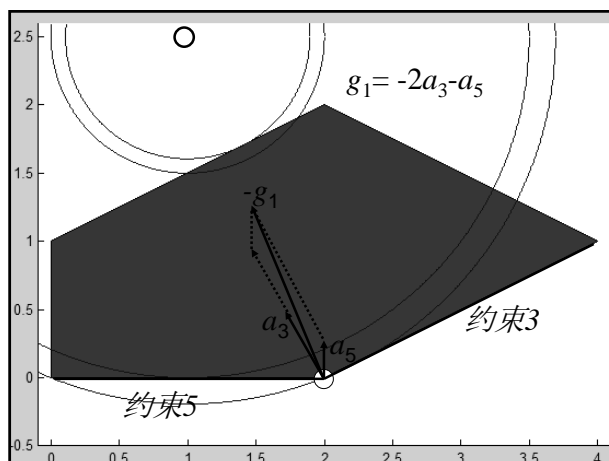
$$\text{令 } g_1 = \lambda_3 a_3 + \lambda_5 a_5, \text{ 则有 } \lambda_3 = -2, \lambda_5 = -1.$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

467



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

$$\min f(x)=(x_1-1)^2+(x_1-2.5)^2$$

$$\text{s.t. } x_2 = 0$$

$$\text{令 } x=x_k+d, g_k=Gx_k+C,$$

$$\text{则 } f(x)=x^T Gx/2+C^T x=(x_k+d)^T G(x_k+d)/2+C^T(x_k+d)$$

$$=d^T Gd/2+x_k^T Gd+x_k^T Gx_k/2+C^T x_k+C^T d$$

$$=d^T Gd/2+g_k^T d+x_k^T Gx_k/2+C^T d$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

469

有效集方法

约束 $a_j^T x \geq b_j$ 等价于 $a_j^T(x_k+d) \geq b_j$,即 $a_j^T d \geq 0$.

一般的,若已知可行解 x_k ,有效集 I_k ,我们求解二次规划

$$\min f(x) = \frac{1}{2} x^T Gx + C^T x$$

$$\text{s.t. } a_j^T x \geq b_j (j \in I_k)$$

它等价于求解

$$\min f(x) = \frac{1}{2} d^T Gd + g_k^T d$$

$$\text{s.t. } a_j^T d = 0 (j \in I_k)$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

470

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

$$\min d^T Gd/2+g_1^T d$$

$$\text{s.t. } x_2=0$$

其中

$$g_1 = \begin{bmatrix} 2 \\ -5 \end{bmatrix}, G = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

求解:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} d^{(1)} \\ d^{(2)} \\ \lambda \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \\ 0 \end{bmatrix}$$

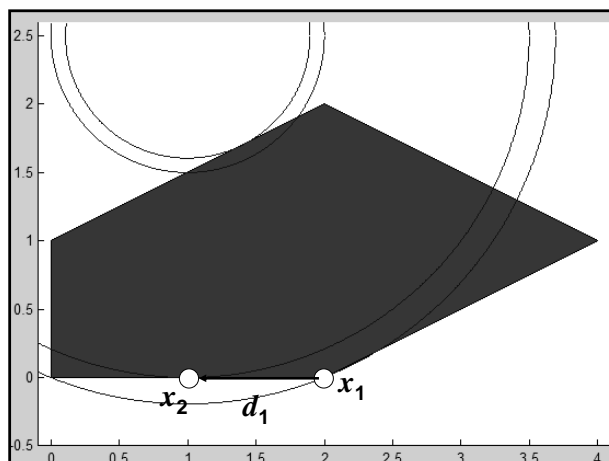
$$(d^{(1)}, d^{(2)}, \lambda)^T = (-1, 0, -5)^T$$

$$\text{因此 } d_1 = (-1, 0)^T$$

$$x_2 = x_1 + d_1 = (1, 0)^T.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

471



有效集方法

$$\text{因此 } d_1 = (-1, 0)^T$$

$$x_2 = x_1 + d_1 = (1, 0)^T. \text{有效集 } I_2 = I_1 = \{5\}.$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

473

有效集方法

定理若严格凸二次规划 $\min f(x) = \frac{1}{2} d^T Gd + g_k^T d$
 $\text{s.t. } a_j^T d = 0 (j \in I_k)$

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

$$\min f(x) = \frac{1}{2} d^T Gd + g_k^T d$$

$$\text{s.t. } a_j^T d = 0 (j \in I_k, j \neq s)$$

则有 $g_k^T p_k < 0, a_s^T p_k > 0$.

作业:证明此定理.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

474

有效集方法

点 x_2 对应的新二次规划为

$$\min d^T G d / 2 + g_2^T d$$

$$s.t. \ x_2 = 0$$

$$\text{其中 } g_2 = (0, -5)^T.$$

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 $I_3 = I_2 \setminus \{5\} = \emptyset$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

475

有效集方法

点 x_3 对应二次规划为(无约束) 因此有

$$\min d^T G d / 2 + g_3^T d$$

$$\text{求得 } d_3 = (0, 2.5)^T.$$

此时 $x_3 + d_3$ 不再可行.

我们求 α_3 使得 $x_3 + \alpha_3 d_3$ 可行, 且 α_3 尽可能大.

于是 $a_i^T(x_3 + \alpha_3 d_3) \geq b_i$.

$$1 - 5\alpha_3 + 2 \geq 0$$

$$-1 - 5\alpha_3 + 6 \geq 0$$

$$-1 + 5\alpha_3 + 2 \geq 0$$

$$1 \geq 0$$

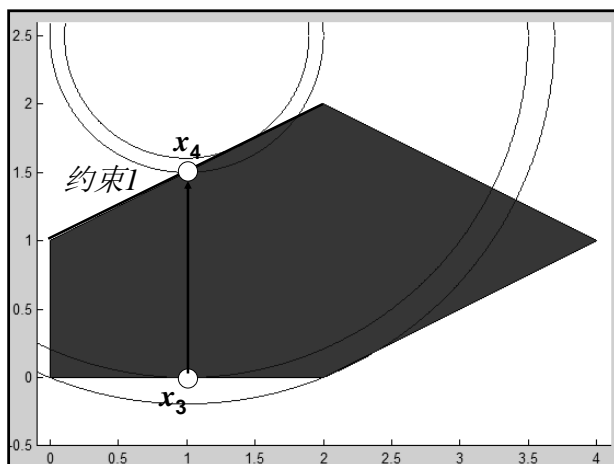
$$2.5\alpha_3 \geq 0$$

$$\text{解得 } \alpha_3 = 0.6.$$

(第一个约束有效)

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

476



有效集方法

$$x_4 = x_3 + \alpha_3 d_3 = (1, 1.5)^T, I_4 = \{1\}, g_4 = (0, -2)^T.$$

求解对应的二次规划

$$\min d^T G d / 2 + g_4^T d$$

$$s.t. \ x_1 - 2x_2 = 0$$

$$\text{解得 } d_4 = (0.4, 0.2)^T.$$

$$x_5 = x_4 + d_4 = (1.4, 1.7)^T (\text{可行}),$$

$$g_5 = (0.8, -1.6)^T.$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

478

有效集方法

x_5 对应的二次规划为

$$\min d^T G d / 2 + g_5^T d$$

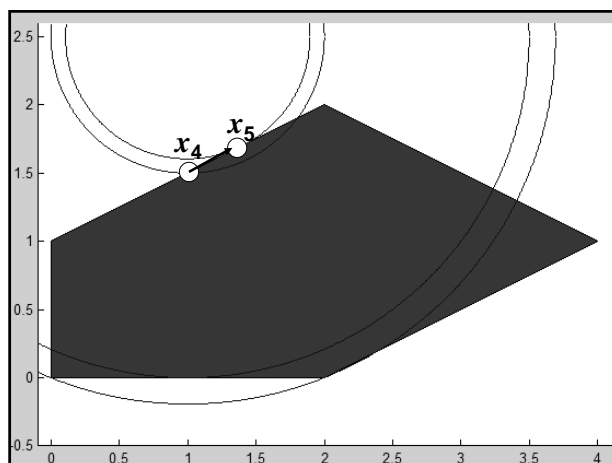
$$s.t. \ x_1 - 2x_2 = 0$$

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 $x^* = x_5 = (1.4, 1.7)^T$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

479



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, \dots, a_m) 中对应于 I_k 的子矩阵,求解

$$\min q(d) = \frac{1}{2}d^T G d + g_k^T d \quad (4.88)$$

$$s.t. \quad A_k^T d = 0$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

481

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 $a_i^T (x_k + \alpha_k d_k) \geq 0, i \notin I_k$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

482

有效集方法

$$a_i^T (x_k + \alpha_k d_k) \geq 0, i \notin I_k$$

即 $\alpha_k a_i^T d_k \geq b_i - a_i^T x_k, i \notin I_k$

由于 $b_i - a_i^T x_k \leq 0$,因此若 $a_i^T d_k \geq 0$,上面的式子恒成立.

我们只要考虑 $a_i^T d_k < 0$ 的情况,

$$\alpha_k = \bar{\alpha}_k = \min_{i \notin I_k} \left\{ \frac{b_i - a_i^T x_k}{a_i^T d_k} \mid a_i^T d_k < 0 \right\} = \frac{b_t - a_t^T x_k}{a_t^T d_k}$$

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

483

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

$$\alpha_k = \bar{\alpha}_k = \frac{b_t - a_t^T x_k}{a_t^T d_k}$$

此时有 $a_t^T x_{k+1} = a_t^T x_k + \alpha_k a_t^T d_k = b_t$,

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 $\alpha_k = 1$,则有效约束集不变 $I_{k+1} = I_k$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

484

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

$$\min q(d) = \frac{1}{2}d^T G d + g_k^T d$$

$$s.t. \quad a_j^T d = 0 (j \in I_k, j \neq s)$$

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.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

485

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

$$\min f(x) = \frac{1}{2}d^T G d + g_k^T d$$

$$s.t. \quad a_j^T d = 0 (j \in I_k)$$

(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 $x_{k+1} = x_k, I_{k+1} = I_k \setminus \{s\}$.

南京邮电大学理学院杨振华制作 yangzhenhua@njupt.edu.cn

486

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 $\alpha_k = 1$, 则有效集不变, $I_{k+1} = I_k$.

若 $\alpha_k < 1$, 则在 I_k 中添加指标得到新的有效集 I_{k+1} .