薛毅 编著

最优化原理与方法

ZuiYouHua YuanLi Yu FangFa

北京工业大学出版社

最优化原理与方法

薛 毅 编著

北京工业大学出版社

图书在版编目（CIP）数据

最优化原理与方法／薛毅编著．—北京：北京工业大学出版社，2001．2

ISBN 7-5639-0961-3

Ⅰ．最…　Ⅱ．薛…　Ⅲ．①最佳化—数学理论②最佳化—数学方法　Ⅳ．O224

中国版本图书馆 CIP 数据核字（2001）第 03796 号

最优化原理与方法

薛 毅　编著

*

北京工业大学出版社出版发行

邮编：100022　电话：（010）67392308

各地新华书店经销

徐水宏远印刷厂印刷

*

2001 年 2 月第 1 版　　2001 年 2 月第 1 次印刷

850mm×1168mm　32 开本　11.75 印张　280 千字

印数：1～2000 册

ISBN7-5639-0961-3/T·172

定价：20.00 元

序　言

最优化，是指从问题的许多可能的解答中，选择依某种指标最好的解答，它是数学的一个重要分支．在50年代后，随着计算机的发展和生产的需要，逐渐形成了最优化理论，以及相应的求解方法——最优化方法．目前这些方法已在生产管理、工程设计、系统控制、经济规划等领域内得到了广泛的应用．

本书是为应用数学系本科生、工科硕士研究生所写的有关最优化知识的一本教材．作为教材，本书的基本观点是：采用简单、基本直观的方法，向学生介绍最优化的有关理论、基本原理和相应的算法，并试图让学生了解算法的来龙去脉，以便使他们在解决实际问题的过程中，更好地运用这些方法．

本书的基础是"数学分析"和"线性代数"，即学生只需具备"数学分析"和"线性代数"知识就可读懂全部内容．对于工科学生，只需具备"高等数学"和"线性代数"知识就可读懂大部分内容．

本书共有十一章．第一章最优化方法，作为引言，向读者介绍最优化的基本概念和本书欲解决的问题．为了便于读者更好地学习本书的知识，特增加了两节数学预备知识．第二章线性规划与第三章线性规划的对偶问题主要涉及线性规划的基本内容．第四章无约束最优化问题的一般结构，第五章一维搜索，第六章使用导数的最优化方法和第七章直接方法，主要讨论无约束最优化问题的求解方法．第八章约束问题的最优性条件，第九章二次规划，第十章可行方向法和第十一章乘子法，主要讨论约束最优化问题的求解方法．作为教材，在每章的后面均列有习题，便于学

生复习和巩固该章所学的知识.

本书在编排上采用分块式，即本书分为三大块，线性规划、无约束最优化问题和约束最优化问题. 各块相对独立，教师可根据情况作一定的删减，可作为不同类型的教材. 如选用无约束最优化问题和约束最优化问题部分可作为"非线性规划"教材，选用线性规划、无约束最优化问题和约束最优化问题（或一部分）可作为"最优化方法"教材. 去掉较难的定理证明和较繁的算法可作为工科学生"最优化方法"方面的教材.

具体授课方法可作如下安排：

1. 作为应用数学系本科生"最优化计算方法"课程的教材，讲授全部内容，大约 80 学时.

2. 作为应用数学系本科生"非线性规划"课程的教材，可去掉本书的第二章、第三章，大约教授 60 学时.

3. 作为工科硕士生"最优化计算方法"课程的教材，可去掉较难的理论证明和第九章二次规划的内容，大约讲授 60 学时.

4. 作为工科本科生"最优化计算方法"课程的教材，可去掉较难的理论证明以及第八章约束问题的最优性条件和第九章二次规划及第四章无约束最优化问题的一般结构的部分内容，大约讲授 40 学时.

当然，任课教师可根据授课内容和对象，对教材进行适当的删减或增加.

本书除作为教材外，也可供从事优化方面的科技工作者和工程技术人员学习和参考.

在编写本书之前，编者曾为我校应用数学系本科生的"非线性规则"课程和工科研究生的"最优化方法"课程编写过讲义，本书是在此基础上，通过多次的教学实践修改而成的. 在编写过程中，得到了邓乃扬教授、诸梅芳教授的很大帮助，陈志教授、杨中华副教授在使用过程中曾提出过许多宝贵修改意见，在此向

他们表示衷心的感谢.

由于受编者水平限制，可能在内容的取材、结构的编排以及课程的讲法上存在着不妥之处，我们希望使用本书的老师、同学以及同行专家和其他读者提出宝贵的批评和建议.

在本书出版之际，谨向为本书出版予以大力支持和帮助的北工大教材建设委员会和北工大出版社表示衷心的感谢.

编　者

目 录

第一章 绪 论

§1.1 引 言

最优化（Optimization），就是在复杂环境中遇到的许多可能的决策中，挑选“最好”的决策的科学．最优化方法——即求最优决策的方法．在本世纪 30 年代末，由于军事和工业生产发展的需要，提出了一些不能用古典微分法和变分法解决的问题．在许多学者和广大科技工作者的共同努力下，逐渐产生、发展和形成了一些新的方法．

前面提到，最优化就是选择“最好”决策．那么什么是最好呢？在不同的意义下，有不同的标准，因此它包括的范围很广，我们无法在短时间内全面而又系统地进行介绍．这里我们首先介绍一个名词——**数学规划**（Mathematical Programming）．数学规划是最优化理论的一个重要分支．数学规划，是指对 n 个变量对单目标（或多目标）函数求极小（或极大）．而这些变量也可能受到某些条件（等式方程或不等式方程）的限制．其数学表达式为：

$$\begin{aligned}
\min\quad & f(\boldsymbol{x}),\quad \boldsymbol{x}\in R^n,\\
\text{s.t.}\quad & c_i(\boldsymbol{x})=0,\ i\in E=\{1,2,\cdots,l\},\\
& c_i(\boldsymbol{x})\leqslant 0,\ i\in I=\{l+1,l+2,\cdots,l+m\},
\end{aligned}\tag{1.1.1}$$

这里 min 表示求极小，s. t. = subject to 意思是受限制于…，$\boldsymbol{x}$ 是 n 维向量，其分量为 x_1，x_2，…，x_n．在问题（1.1.1）中称 $f(\boldsymbol{x})$ 为**目标函数**（objective function），称 $c_i(\boldsymbol{x})=0,(i\in E)$ $c_i(\boldsymbol{x})\leqslant 0(i\in I)$ 为**约束条件**（constraint condition）．若求极大，可

以将目标函数写成 min $(-f(x))$，若不等式约束 $c_i(x) \geqslant 0$ 可以写成 $-c_i(x) \leqslant 0$.

在问题（1.1.1）中，若 $f(x)$、$c_i(x)(i \in E \cup I)$ 均是线性函数，则得到的规划称为**线性规划**（Linear Programming），其一般表达式为

$$
\begin{aligned}
\min \quad & f(x)=c^T x, x \in R^n, \\
\text{s.t.} \quad & c_i(x)=a_i^T x-b_i=0, i \in E=\{1,2,\cdots,l\} \\
& c_i(x)=a_i^T x-b_i \leqslant 0, i \in I=\{l+1, l+2, \cdots, l+m\}.
\end{aligned} \tag{1.1.2}
$$

在问题（1.1.1）中，$f(x)$、$c_i(x)(i \in E \cup I)$ 中之一是非线性函数，则称为**非线性规划**（Nonlinear Programming）. 若去掉问题（1.1.1）的约束条件，则得到

$$
\min \quad f(x),\ x \in R^n, \tag{1.1.3}
$$

称为**无约束最优化问题**（unconstraint optimmization problem）. 因此，也称问题（1.1.1）为**约束最优化问题**（constraint optimization problm）.

§1.2 最优化问题

为进一步说明最优化问题，我们先举例说明.

1.2.1 无约束最优化问题

【例 1.2.1】 曲线拟合问题.

设有两个物理量 ξ 和 η，根据某一物理定律得知它们满足如下关系：

$$
\eta=a+b\xi^c,
$$

其中 a、b、c 三个常数在不同情况下取不同的值. 现由实验得到一给数据

$$
(\xi_1, \eta_1), (\xi_2, \eta_2), \cdots, (\xi_m, \eta_m),
$$

试选择 a、b、c 的值，使曲线 $\eta = a + b\xi^c$ 尽可能靠近所有的实验点（ξ_i，η_i），$i=1, 2, \cdots, m$.

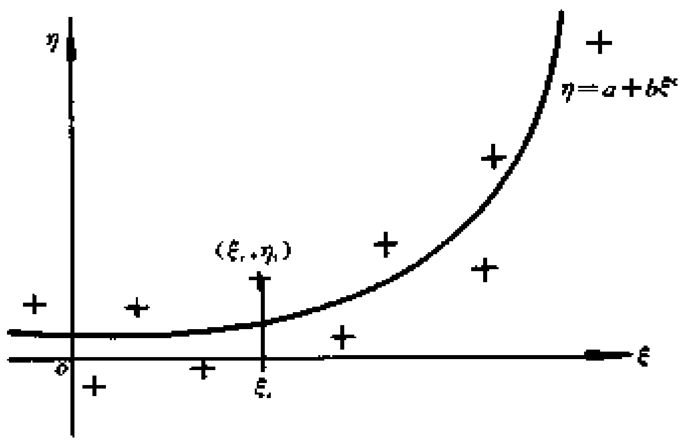


图 1.2.1　曲线拟合问题

这个问题可用最小二乘原理求解，即选择 a、b、c 的一组值，使得偏差的平方和

$$\delta(a,b,c) = \sum_{i=1}^{m}(a + b\xi_i^c - \eta_i)^2 \tag{1.2.1}$$

达到最小. 换句话说，就是求三个变量的函数 δ（a，b，c）的极小点作为问题的解.

为了便于今后的讨论，我们把它写成统一的形式，把 a、b、c 换成 x_1、x_2、x_3，记为 $\boldsymbol{x}$ = （x_1，x_2，x_3）^{T}，把 δ 换成 f，这样（1.2.1）式改写为

$$f(\boldsymbol{x}) = f(x_1,x_2,x_3) = \sum_{i=1}^{m}(x_1 + x_2\xi_i^{x_3} - \eta_i)^2, \tag{1.2.2}$$

问题归纳为

$$\min \quad f(\boldsymbol{x}), \; \boldsymbol{x}\in R^3,$$

即无约束最优化问题.

无约束最优化问题的一般形式为

$$\min \quad f(\boldsymbol{x}), \quad \boldsymbol{x} \in R^n. \tag{1.2.3}$$

那么什么是它的最优解呢？我们给出严格的定义.

定义 1.2.1 若存在 $\boldsymbol{x}^* \in R^n$，使得对任意的 $\boldsymbol{x} \in R^n$，均有

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*), \tag{1.2.4}$$

则称 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 的**全局极小点**（global minimum point）或称 $\boldsymbol{x}^*$ 是无约束问题**全局解**（global solution）. 若对任意的 $\boldsymbol{x} \in R^n$，$\boldsymbol{x} \neq \boldsymbol{x}^*$，均有

$$f(\boldsymbol{x}) > f(\boldsymbol{x}^*), \tag{1.2.5}$$

则称 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 的严格全局极小点，或称 $\boldsymbol{x}^*$ 是无约束问题严格全局解.

由此可知，所谓无约束最优化问题的**最优解**（optimal solution），就是满足（1.2.4）式或（1.2.5）式的 $\boldsymbol{x}^*$. ●

1.2.2 约束最优化问题

【例 1.2.2】 生产安排问题.

某工厂生产 A、B、C 三种产品，每件产品所消耗的材料、工时以及盈利见表 1.2.1.

表 1.2.1

产　　品	A	B	C
材料（千克/件）	4	4	5
工时（小时/件）	4	2	3
盈利（元/件）	7	3	6

已知该工厂每天的材料消耗不超过 600 千克，工时不得超过 1400 小时，问每天生产 A、B、C 三种产品各多少可使得盈利最大？

解 设每天生产 A、B、C 三种产品分别是 x_1、x_2、x_3 件，因此共盈利

$$7x_1 + 3x_2 + 6x_3,$$

其相应的材料限制为

$$4x_1+4x_2+5x_3\leqslant 600,$$

工时限制为

$$4x_1+2x_2+3x_3\leqslant 1400,$$

再考虑自然限制

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

因此生产安排问题就是在上述限制条件下，使得其盈利达到最大，其数学表达式为

$$\begin{aligned}
\min\quad & f(\boldsymbol{x})=-7x_1-3x_2-6x_3,\\
\text{s.t.}\quad & c_1(\boldsymbol{x})=4x_1+4x_2+5x_3-600\leqslant 0,\\
& c_2(\boldsymbol{x})=4x_1+2x_2+3x_3-1400\leqslant 0,\\
& c_3(\boldsymbol{x})=-x_1\leqslant 0,\\
& c_4(\boldsymbol{x})=-x_2\leqslant 0,\\
& c_4(\boldsymbol{x})=-x_3\leqslant 0.
\end{aligned}$$

【例 1.2.3】 人字架最优设计问题.

考虑如图 1.2.2 所示的钢管构造的人字架，设钢管壁厚 $t=\bar{t}$ 和半跨度 $s=\bar{s}$ 已给定，试求能承受负荷 $2P$ 的最轻设计.

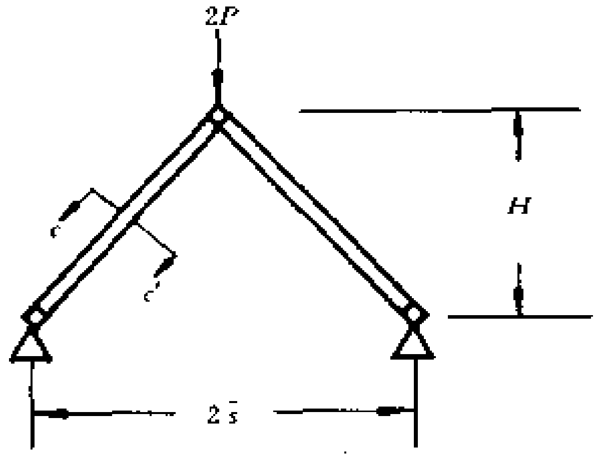
首先我们定性地分析此问题. 壁厚和跨度一定，欲求最轻设计，需要杆短. 这样做必使张角增大，则负荷 $2P$ 就会在钢管上有很大的张力. 为了能承受这样的应力，钢管需变粗，其结果是杆变重.

下面进行定量的分析. 给定一组 d 和 H 值后，可以计算出钢管的截面积 A 和钢管的长度 L，即

$$A=\frac{1}{4}\pi(D_2^2-D_1^2)=\frac{\pi}{4}(D_2+D_1)(D_2-D_1)=\pi d\,\bar{t},$$

$$L=(\bar{s}^2+H^2)^{\frac{1}{2}},$$

因此钢管的重量为



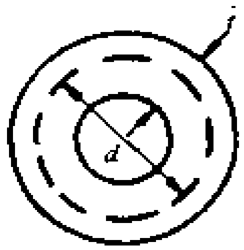


图 1.2.2 人字架最优设计问题

$$W(d,H)=2\rho\pi d\,\bar{t}(\bar{s}^2+H^2)^{\frac{1}{2}},$$

其中 ρ 是比重.

下面考虑 d、H 受到的限制，当负荷为 $2P$ 时，杆件受到的压力为

$$\sigma(d,H)=\frac{P}{\pi\bar{t}}\cdot\frac{(\bar{s}^2+H^2)^{\frac{1}{2}}}{Hd}.$$

根据结构力学原理，对于选定的钢管，不出现屈服的条件是

$$\sigma(d,H)\leqslant\sigma_y,$$

其中 σ_y 是钢管最大许可的抗压强度. 而不出现弹性变曲（屈曲条件）为

$$\sigma(d,H)\leqslant\frac{\pi^2E(d^2+\bar{t}^2)}{8(\bar{s}^2+H^2)},$$

其中 E 为钢管材料的扬氏模量. 因此人字架最优设计问题是在上述两个条件下使 W（d，H）达到最小.

为了方便起见，写成统一的数学表达式，分别用 x_1、x_2 代替 d、H，记 $\boldsymbol{x}=(x_1,x_2)^T$，用 f 代替 W，因此得到

$$f(x) = 2\pi\rho \bar{t} x_1 (\bar{s}^2 + x_2^2)^{\frac{1}{2}},$$

$$c_1(x) = \frac{P}{\pi \bar{t}} \cdot \frac{(\bar{s}^2 + x_2^2)^{\frac{1}{2}}}{x_1 x_2} - \sigma_y,$$

$$c_2(x) = \frac{P}{\pi \bar{t}} \cdot \frac{(\bar{s}^2 + x_2^2)^{\frac{1}{2}}}{x_1 x_2} - \frac{\pi^2 E}{8} \cdot \frac{\bar{t}^2 + x_1^2}{\bar{s}^2 + x_2^2}.$$

人字架最优设计问题的数学表达式为

$$\begin{aligned} &\min \quad f(x), \quad x \in R^n, \\ &\text{s.t.} \quad c_i(x) \leqslant 0, \; i = 1,2. \end{aligned}$$

综上所述，约束最优化问题的一般形式为

$$\begin{aligned} &\min \quad f(x), \qquad x \in R^n, \\ &\text{s.t.} \quad c_i(x) = 0, i \in E = \{1,2,\cdots,l\}, \\ &\qquad\; c_i(x) \leqslant 0, i \in I = \{l+1, l+2, \cdots, l+m\}, \end{aligned} \tag{1.2.6}$$

称 $f(x)$ 为目标函数，称 $c_i(x) = 0$ $(i \in E)$ 或 $c_i(x) \leqslant 0$ $(i \in I)$ 为约束条件.

在例 1.2.2 中，所有函数均是线性的，因此是线性规划问题，而例 1.2.3 则属于非线性规划问题.

定义 1.2.2 称满足约束条件的点为**可行点** (feasible point)，称可行点全体组成的集合为**可行域** (feasible region)，记作 D，即

$$D = \{x \mid c_i(x) \leqslant 0, i \in E, c_i(x) \leqslant 0, i \in I, x \in R^n\}. \tag{1.2.7}$$

这样我们可以定义约束问题的最优解.

定义 1.2.3 若存在 $x^* \in D$，使得对任意的 $x \in D$，均有

$$f(x) \geqslant f(x^*), \tag{1.2.8}$$

则称 x^* 是 $f(x)$ 在可行域上的全局极小点，或称 x^* 是约束问题全局解. 若对任意的 $x \in D$，$x \neq x^*$，均有

$$f(\boldsymbol{x}) > f(\boldsymbol{x}^*), \tag{1.2.9}$$

则称 $\boldsymbol{x}^*$ 是 f（$\boldsymbol{x}$）在可行域上的严格全局极小点，或称 $\boldsymbol{x}^*$ 是约束问题严格全局解.

与无约束问题一样，满足（1.2.8）式或（1.2.9）式的 $\boldsymbol{x}^*$，称为约束问题的最优解.

1.2.3 最优化问题的图解法

【例 1.2.4】 根据图像求解无约束问题

$$\min f(\boldsymbol{x}) = (x_1 - 2)^2 + (x_2 - 2)^2 + 4.$$

解 先画出曲面 $y = f(x)$ 的图像，该曲面的最低点在 $x_1 0 x_2$ 上的投影为问题的解. 为寻找该投影点，考虑 f 的等高线，$f(\boldsymbol{x}) = k$，即

$$(x_1 - 2)^2 + (x_2 - 2)^2 = k - 4,$$

（见图 1.2.3）当 $k>4$ 时，等高线为同心圆，其同心圆的中心是 $(2,\ 2)^T$，即是问题的最优解 $\boldsymbol{x}^*$，其目标函数值为 f（$\boldsymbol{x}^*$）$=4$.

【例 1.2.5】 用图解法求解约束问题

$$\begin{aligned}
\min \quad & f(\boldsymbol{x}) = -3x_1 - 4x_2, \\
\text{s.t} \quad & c_1(\boldsymbol{x}) = x_1 + 2x_2 - 6 \leqslant 0, \\
& c_2(\boldsymbol{x}) = 3x_1 + 2x_2 - 12 \leqslant 0, \\
& c_3(\boldsymbol{x}) = x_2 - 2 \leqslant 0, \\
& c_4(\boldsymbol{x}) = -x_1 \leqslant 0, \\
& c_5(\boldsymbol{x}) = -x_2 \leqslant 0.
\end{aligned}$$

解 首先画出问题的可行域 D，它是由五条直线组成的区域（见图 1.2.4），然后再画出等高线

$$-3x_1 - 4x_2 = k,$$

这是若干条直线. 由图形可知，在两条直线 $x_1 + 2x_2 = 6$、$3x_1 +$

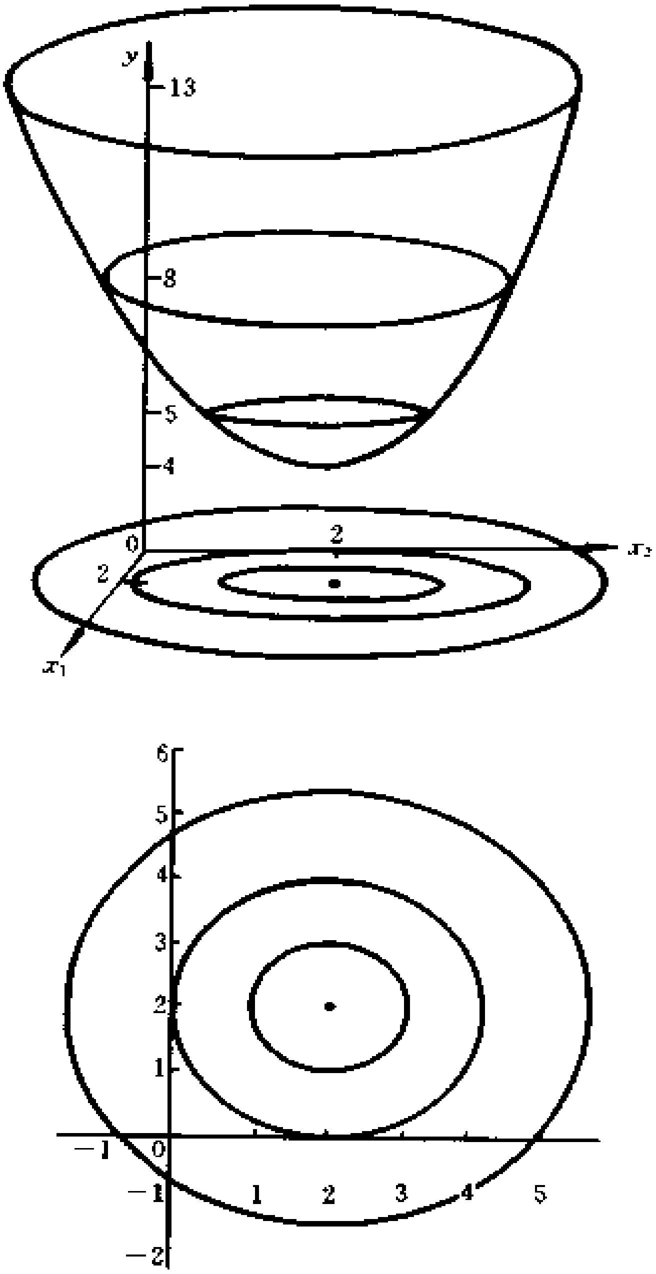


图 1.2.3

$2x_2=12$ 的交点处，等高线达到最小，因此该交点就是我们要求的最优解．解联立方程得到 $\boldsymbol{x}^*=\left(3,\ \dfrac{3}{2}\right)^T$.

【例 1.2.6】　用图解法求解约束问题

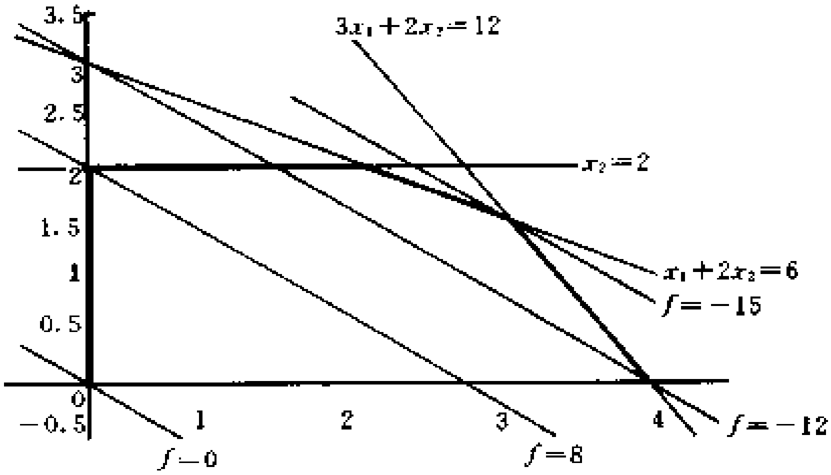


图 1.2.4

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = (x_1-2)^2+(x_2-2)^2, \\ \text{s.t.} \quad & c_1(\boldsymbol{x}) = x_1+2x_2-4 \leqslant 0, \\ & c_2(\boldsymbol{x}) = -x_1 \leqslant 0, \\ & c_3(\boldsymbol{x}) = -x_2 \leqslant 0. \end{aligned}$$

解 首先画出问题的可行域 D，即

$$D = \{\boldsymbol{x} \mid x_1+2x_2 \leqslant 4,\ x_1 \geqslant 0,\ x_2 \geqslant 0\},$$

再考虑 f 的等高线

$$(x_1-2)^2+(x_2-2)^2=k,$$

这是一族同心圆．由图形可知，某一等高线（$f(\boldsymbol{x})=k$）与约束 $c_1(\boldsymbol{x})=0$ 相切，即其偏导数成比例．由于

$$\frac{\partial f}{\partial x_1}=2x_1-4,\ \frac{\partial f}{\partial x_2}=2x_2-4$$

和

$$\frac{\partial c_1}{\partial x_1}=1,\ \frac{\partial c_1}{\partial x_2}=2,$$

则最优点满足方程组

$$\begin{cases}2x_1 - 4 = \lambda, \\ 2x_2 - 4 = 2\lambda, \\ x_1 + 2x_2 = 4.\end{cases}$$

解方程组得到 $\boldsymbol{x}^* = \left(\frac{8}{5}, \frac{6}{5}\right)^T$，其最优目标函数值为 $f(\boldsymbol{x}^*) = \frac{4}{5}$.

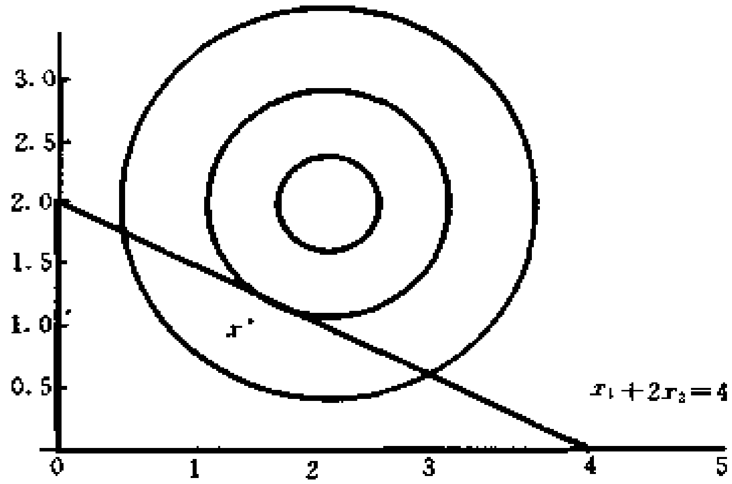


图 1.2.5

§1.3 数学预备知识

为了便于后面的讨论，我们在介绍正文内容之前，先复习一下与最优化有关的知识.

1.3.1 线性代数知识

1. 线性空间

设 E 是一个集合，在此集合上定义加法和数乘运算，并且运算是封闭的，这样形式的 E 称为线性空间，其中 E 的元素 $\boldsymbol{x}$ 称为向量.

对于 n 维线性空间中，记 n 维列向量为

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^T, \tag{1.3.1}$$

其中 T 表示转置.

若 $\boldsymbol{x}$、$\boldsymbol{y}$ 是线性空间 E 中的两个向量，通常的加法定义为

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n)^T, \tag{1.3.2}$$

因此有

$$\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}, \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}. \tag{1.3.3}$$

若 $x_i = 0$，$i = 1, 2, \cdots, n$，则记为 $\boldsymbol{x} = 0$，称为零向量.

若 λ 是一个纯量，通常的数乘定义为

$$\lambda \boldsymbol{x} = (\lambda x_1, \lambda x_2, \cdots, \lambda x_n)^T. \tag{1.3.4}$$

数乘有如下性质：

$$\begin{aligned} \lambda(\mu x) &= (\lambda\mu)\boldsymbol{x}, \\ (\lambda + \mu)\boldsymbol{x} &= \lambda\boldsymbol{x} + \mu\boldsymbol{x}, \\ \lambda(\boldsymbol{x} + \boldsymbol{y}) &= \lambda\boldsymbol{x} + \lambda\boldsymbol{y}. \end{aligned} \tag{1.3.5}$$

设 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，$\cdots$，$\boldsymbol{a}_m$ 为 E 中 m 个向量，称

$$\boldsymbol{x} = \lambda_1 \boldsymbol{a}_1 + \lambda_2 \boldsymbol{a}_2 + \cdots + \lambda_m \boldsymbol{a}_m = \sum_{i=1}^{m} \lambda_i \boldsymbol{a}_i \tag{1.3.6}$$

为 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，$\cdots$，$\boldsymbol{a}_m$ 的线性组合.

如果存在 $\lambda_i \neq 0$（即 λ_i 不全为零），使得

$$\lambda_1 \boldsymbol{a}_1 + \lambda_2 \boldsymbol{a}_2 + \cdots + \lambda_m \boldsymbol{a}_m = 0, \tag{1.3.7}$$

则称 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，$\cdots$，$\boldsymbol{a}_m$ 线性相关，反之则称 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，$\cdots$，$\boldsymbol{a}_m$ 线性无关，也就是说，若（1.3.7）式成立，则 $\lambda_i = 0$，$i = 1, 2, \cdots, m$ 是唯一解.

2. Euclid 空间（欧氏空间）

所谓欧氏空间，就是在线性空间上定义一个度量. 对于 n 维欧氏空间（记为 R^n），向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的内积定义为

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x}^T \boldsymbol{y}. \tag{1.3.8}$$

由（1.3.8）式，容易得到内积具有如下性质：

(1) $\langle x, x\rangle \geqslant 0$，且 $\langle x, x\rangle = 0$ 的充分必要条件是 $x=0$.

(2) $\langle x, y\rangle = \langle y, x\rangle$.

(3) $\langle \lambda x+\mu y, z\rangle = \lambda\langle x, z\rangle + \mu\langle y, z\rangle$.

n 维欧氏空间 R^n 上的范数定义为

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} = (\langle x,x\rangle)^{\frac{1}{2}}, \qquad (1.3.9)$$

即通常意义下的距离，或称为 l_2——范数.

范数具有如下性质：

(1) $\|x\| \geqslant 0$，且 $\|x\| = 0$ 的充分必要条件是 $x=0$.

(2) $\|\lambda x\| = |\lambda|\,\|x\|$.

(3) $\|x+y\| \leqslant \|x\| + \|y\|$（三角不等式）. 以及（Cauchy-Schwarz）不等式

$$|\langle x,y\rangle| \leqslant \|x\| \cdot \|y\|, \qquad (1.3.10)$$

且等号成立的充分必要条件是：x 与 y 共线，即存在 λ，使得

$$x = \lambda y. \qquad (1.3.11)$$

将（1.3.10）写成分量形式为

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leqslant \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{n} y_i^2\right)^{\frac{1}{2}}, \qquad (1.3.12)$$

若 $\|x\| = 1$，则称 x 为单位向量. 对于作一个非零向量 x，可构造出同方向上的单位向量 $x/\|x\|$.

若 x、y 满足

$$\langle x,y\rangle = 0, \qquad (1.3.13)$$

则称 x 与 y 正交. 由（1.3.13）式可知，零向量与任何向量均正交. 反过来，若 x 与任何向量正交，则 x 必为零向量. 事实上，x 与任何向量正交，那么也与它本身正交，即

$$\langle x, x\rangle = 0,$$

由内积的性质，得到 $x=0$.

3. 矩阵

称

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \tag{1.3.14}$$

为 $m \times n$ 阶矩阵，记为 $\boldsymbol{A}=\boldsymbol{A}_{m\times n}$.

对于矩阵 $\boldsymbol{A}$，可以进行分块，即

$$\boldsymbol{A}=\left[\begin{array}{c:c} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \hdashline \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{array}\right], \tag{1.3.15}$$

其中 $\boldsymbol{A}_{11}$为 $m_1 \times n_1$，$\boldsymbol{A}_{12}$为 $m_1 \times n_2$，$\boldsymbol{A}_{21}$为 $m_2 \times n_1$，$\boldsymbol{A}_{22}$为 $m_2 \times n_2$ 阶矩阵，且 $m_1+m_2=m$，$n_1+n_2=n$. 这里特别要提到两种特殊的分块形式，按列分块

$$\boldsymbol{A}=[\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_n],\ \boldsymbol{P}_j=\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \tag{1.3.16}$$

和按行分块

$$\boldsymbol{A}=\begin{bmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \\ \vdots \\ \boldsymbol{a}_m^T \end{bmatrix},\ \boldsymbol{a}_i^T=(a_{i1},a_{i2},\cdots,a_{in}). \tag{1.3.17}$$

称矩阵 $\boldsymbol{A}$ 的行（或列）的极大无关组的个数为矩阵 $\boldsymbol{A}$ 的秩，记为 rank（$\boldsymbol{A}$）. 若

$$\operatorname{rank}(\boldsymbol{A})=\min\{m,n\} \tag{1.3.18}$$

时，则称矩阵 $\boldsymbol{A}$ 是满秩的．若 $m<n$，则称为行满秩，若 $m>n$，则称为列满秩．当若 $m=n$ 时，则矩阵为 n 阶非奇异（可逆）方阵．

当 $\boldsymbol{A}$ 为方阵时，可计算矩阵 $\boldsymbol{A}$ 的行列式

$$\det(\boldsymbol{A})=\sum_{i_1i_2\cdots i_n}(-1)^{\tau(i_1i_2\cdots i_n)}a_{1i_1}a_{2i_2}\cdots a_{ni_n}, \tag{1.3.19}$$

其中 τ $(i_1i_2\cdots i_n)$ 是 i_1，i_2，$\cdots$，i_n 的逆序数，$\sum\limits_{i_1i_2\cdots i_n}$ 表示对所有的 n 阶排列求和．

矩阵非奇异的充分必要条件是：$\det$ $(\boldsymbol{A})$ $\neq 0$．

若 $\boldsymbol{A}$ 为 $n\times n$ 阶矩阵，若 $\boldsymbol{A}$ 满足

$$\boldsymbol{A}^T=\boldsymbol{A}, \tag{1.3.20}$$

则称 $\boldsymbol{A}$ 为对称矩阵．若对于一切 $\boldsymbol{x}\neq 0$，均有

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}>0, \tag{1.3.21}$$

则称 $\boldsymbol{A}$ 为正定对称矩阵．若对一切 $\boldsymbol{x}$，均有

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}\geqslant 0, \tag{1.3.22}$$

则称 $\boldsymbol{A}$ 为半正定矩阵．

1.3.2　多变量函数的方向导数

首先介绍梯度和 Hesse 矩阵的概念．

定义 1.3.1　称 n 维向量$\left(\frac{\partial}{\partial x_1}f(\bar{\boldsymbol{x}}),\frac{\partial}{\partial x_2}f(\bar{\boldsymbol{x}}),\cdots,\frac{\partial}{\partial x_n}f(\bar{\boldsymbol{x}})\right)^T$为函数$f(\boldsymbol{x})$在 $\bar{\boldsymbol{x}}$ 处的**梯度**（gradient），记为∇f $(\bar{\boldsymbol{x}})$，即

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

称$\nabla f(\boldsymbol{x})$为梯度函数，简称梯度．有时将$\nabla f(\boldsymbol{x})$记为$\boldsymbol{g}(\boldsymbol{x})$．

定义 1.3.2　称 $n\times n$ 阶矩阵

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

为函数 $f(\boldsymbol{x})$ 在 $\bar{\boldsymbol{x}}$ 处的 Hesse 矩阵（Hessian matrtix），记为 $\nabla^2 f(\bar{\boldsymbol{x}})$，即

$$\nabla^2 f(\bar{\boldsymbol{x}}) = \left(\frac{\partial^2}{\partial x_i \partial x_j}f(\bar{\boldsymbol{x}})\right)_{n\times n}.$$

类似地，称$\nabla^2 f(\boldsymbol{x})$为 Hesse 矩阵函数，有时记为 $G(\boldsymbol{x})$.

【例 1.3.1】 求 $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x}$、$f(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x}$ 的梯度和 Hesse 矩阵. 这里 $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^T$，$\boldsymbol{A} = (a_{ij})_{n\times n}$.

解 (1) $f(\boldsymbol{x}) = a^T\boldsymbol{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n$，

所以

$$\frac{\partial f}{\partial x_k} = a_k,\ k = 1,2,\cdots,n, \tag{1.3.23}$$

因此，

$$\nabla f(\boldsymbol{x}) = (a_1, a_2, \cdots, a_n)^T = \boldsymbol{a}. \tag{1.3.24}$$

对 (1.3.23) 式再求导数，得到

$$\frac{\partial^2 f}{\partial x_k \partial x_l} = 0,$$

因此得到$\nabla^2 f(\boldsymbol{x}) = 0$(0 矩阵).

(2) $f(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$，

因此

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{n} a_{ik}x_i + \sum_{j=1}^{n} a_{kj}x_j, k = 1,2,\cdots,n \tag{1.3.25}$$

和

$$\frac{\partial^2 f}{\partial x_k \partial x_l} = a_{lk} + a_{kl}, \quad k, l = 1, 2, \cdots, n, \tag{1.3.26}$$

所以梯度为

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} +$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{1.3.27}$$

$$= A^T x + A x$$

其 Hesse 矩阵为

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

$$= \boldsymbol{A}^T + \boldsymbol{A}. \tag{1.3.28}$$

若 $\boldsymbol{A}$ 为对称矩阵，即 $\boldsymbol{A}^T = \boldsymbol{A}$，则（1.3.27）式和（1.3.28）式可以写成

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}, \tag{1.3.29}$$

$$\nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}. \tag{1.3.30}$$

与导数有关的另一个概念是**方向导数**.

定义 1.3.3 对于任意给定的 $\boldsymbol{d} \neq 0$，若极限

$$\lim_{\alpha \to 0^+} \frac{f(\bar{\boldsymbol{x}} + \alpha \boldsymbol{d}) - f(\bar{\boldsymbol{x}})}{\alpha \| \boldsymbol{d} \|}$$

存在，则称该极限值为函数 $f(x)$ 在 $\bar{x}$ 处沿方向 $\boldsymbol{d}$ 的一阶方向导数，简称为方向导数，记为 $\frac{\partial}{\partial \boldsymbol{d}}f(\bar{\boldsymbol{x}})$，即

$$\frac{\partial}{\partial \boldsymbol{d}}f(\bar{\boldsymbol{x}}) = \lim_{\alpha \to 0^+} \frac{f(\bar{\boldsymbol{x}} + \alpha \boldsymbol{d}) - f(\bar{\boldsymbol{x}})}{\alpha \| \boldsymbol{d} \|}. \tag{1.3.31}$$

由定义 1.3.3 求方向导数相当繁琐，这里给出它的另一种表达式.

定理 1.3.1 若函数 $f(x)$ 具有连续的一阶偏导数，则它在 $\bar{x}$ 处沿方向 $\boldsymbol{d}$ 的一阶方向导数为

$$\frac{\partial}{\partial \boldsymbol{d}}f(\bar{\boldsymbol{x}}) = \langle \nabla f(\bar{\boldsymbol{x}}), \frac{\boldsymbol{d}}{\| \boldsymbol{d} \|} \rangle$$

$$= \frac{1}{\| \boldsymbol{d} \|} \boldsymbol{d}^T f(\bar{\boldsymbol{x}}). \tag{1.3.32}$$

证明 记 $\bar{\boldsymbol{x}} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n)^T$，$\boldsymbol{d} = (d_1, d_2, \cdots, d_n)^T$，考虑单变量函数

$$\phi(\alpha) = f(\bar{\boldsymbol{x}} + \alpha \bar{\boldsymbol{d}})$$

$$= f(\bar{\boldsymbol{x}}_1 + \alpha d_1, \bar{\boldsymbol{x}}_2 + \alpha d_2, \cdots, \bar{\boldsymbol{x}}_n + \alpha d_n), \tag{1.3.33}$$

由定理条件知 $\phi(\alpha)$ 可微，因此

$$\phi'(\alpha) = \frac{\partial f}{\partial x_1} d_1 + \frac{\partial f}{\partial x_2} d_2 + \cdots \frac{+ \partial f}{\partial x_n} d_n$$

$$= \langle \nabla f(\bar{\boldsymbol{x}} + \alpha \boldsymbol{d}), \boldsymbol{d} \rangle, \tag{1.3.34}$$

当 $\alpha = 0$ 时，有

$$\phi'(0) = \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{d} \rangle. \tag{1.3.19}$$

另一方面，由 (1.3.31) 式及 (1.3.33) 式得到

$$\frac{\partial}{\partial \boldsymbol{d}}f(\bar{\boldsymbol{x}}) = \lim_{\alpha \to 0^+} \frac{f(\bar{\boldsymbol{x}} + \alpha \boldsymbol{d}) - f(\bar{\boldsymbol{x}})}{\alpha \| \boldsymbol{d} \|}$$

$$= \frac{1}{\| \boldsymbol{d} \|} \lim_{\alpha \to 0^+} \frac{\phi(\alpha) - \phi(0)}{\alpha}$$

$$= \frac{1}{\| \boldsymbol{d} \|} \phi'(0). \tag{1.3.36}$$

由（1.3.36）式和（1.3.35）式得到

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}}) &= \frac{1}{\| \boldsymbol{d} \|} \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{d} \rangle \\ &= \langle \nabla f(\bar{\boldsymbol{x}}), \frac{\boldsymbol{d}}{\| \boldsymbol{d} \|} \rangle \\ &= \frac{1}{\| \boldsymbol{d} \|} \boldsymbol{d}^T \nabla f(\bar{\boldsymbol{x}}). \end{aligned}$$

设 $\boldsymbol{d}$ 是单位向量，即 $\| \boldsymbol{d} \| = 1$，那么沿 $\boldsymbol{d}$ 的方向导数可表示为

$$\frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}}) = \langle \nabla f(\bar{\boldsymbol{x}}), \boldsymbol{d} \rangle. \tag{1.3.37}$$

方向导数的几何意义是：函数 $f(x)$ 在 $\bar{\boldsymbol{x}}$ 处沿 $\boldsymbol{d}$ 方向的变化率．若 $\frac{\partial f}{\partial \boldsymbol{d}} > 0$，则函数沿 $\boldsymbol{d}$ 方向增加时，函数值上升，因此称 $\boldsymbol{d}$ 为上升方向．若 $\frac{\partial f}{\partial \boldsymbol{d}} < 0$，称 $\boldsymbol{d}$ 是下降方向．

由（1.3.32）式和 Cauchy-Schwarz 不等式得到

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}}) &= \langle \nabla f(\bar{\boldsymbol{x}}), \frac{\boldsymbol{d}}{\| \boldsymbol{d} \|} \rangle \\ &\leqslant \| \nabla f(\bar{\boldsymbol{x}}) \| \left\| \frac{\boldsymbol{d}}{\| \boldsymbol{d} \|} \right\| \\ &= \| \nabla f(\bar{\boldsymbol{x}}) \|. \end{aligned} \tag{1.3.38}$$

特别当 $\boldsymbol{d} = \nabla f(\bar{\boldsymbol{x}})$ 时，有

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}}) &= \langle \nabla f(\bar{\boldsymbol{x}}), \frac{\boldsymbol{d}}{\| \boldsymbol{d} \|} \rangle \\ &= \langle \nabla f(\bar{\boldsymbol{x}}), \frac{\nabla f(\bar{\boldsymbol{x}})}{\| \nabla f(\bar{\boldsymbol{x}}) \|} \rangle = \| \nabla f(\bar{\boldsymbol{x}}) \|, \end{aligned} \tag{1.3.39}$$

结合（1.3.38）式和（1.3.39）式，$\boldsymbol{d} = \nabla f(\bar{\boldsymbol{x}})$ 是在 $\bar{\boldsymbol{x}}$ 处使方

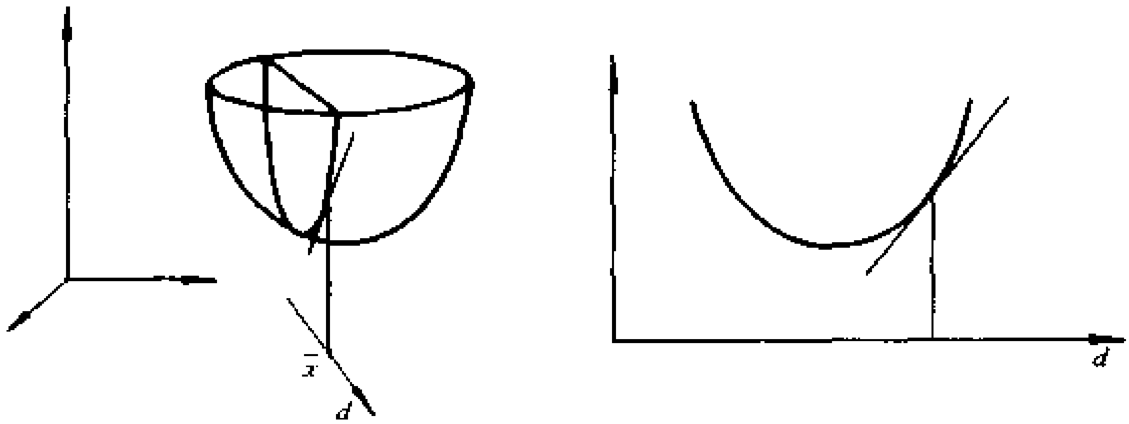


图 1.3.1　方向导数的几何意义

向导数达到最大的方向，称为最速上升方向.

同理得到$\frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}}) \geqslant -\|\nabla f(\bar{\boldsymbol{x}})\|$，当 $\boldsymbol{d}=-\nabla f(\bar{\boldsymbol{x}})$时，有$\frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}})=-\|\nabla f(\bar{\boldsymbol{x}})\|$. 因此称 $\boldsymbol{d}=-\nabla f(\bar{\boldsymbol{x}})$为 $\bar{\boldsymbol{x}}$ 处的**最速下降方向** (direction of steepest descent).

下面介绍二阶方向导数.

定义 1.3.4　对任意的 $\boldsymbol{d} \neq 0$，若极限

$$\lim_{\alpha \rightarrow 0^{+}} \frac{\frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}}+\alpha \boldsymbol{d})-\frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}})}{\alpha\|\boldsymbol{d}\|}$$

存在，则称该极限值为函数$f(\boldsymbol{x})$在 $\bar{\boldsymbol{x}}$ 处沿方向$\boldsymbol{d}$ 的二阶方向导数，记为$\frac{\partial^{2}}{\partial \boldsymbol{d}^{2}} f(\bar{\boldsymbol{x}})$，即

$$\frac{\partial^{2}}{\partial \boldsymbol{d}^{2}} f(\bar{\boldsymbol{x}})=\lim_{\alpha \rightarrow 0^{+}} \frac{\frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}}+\alpha \boldsymbol{d})-\frac{\partial}{\partial \boldsymbol{d}} f(\bar{\boldsymbol{x}})}{\alpha\|\boldsymbol{d}\|}. \tag{1.3.40}$$

并有相应的定理.

定理 1.3.2　若函数$f(\boldsymbol{x})$具有连续的二阶偏导数，则它在 $\bar{\boldsymbol{x}}$

处沿方向 $\boldsymbol{d}$ 的二阶方向导数为

$$\frac{\partial^2}{\partial \boldsymbol{d}^2} f(\overline{\boldsymbol{x}}) = \frac{1}{\|\boldsymbol{d}\|^2} \boldsymbol{d}^T \nabla^2 f(\overline{\boldsymbol{x}}) \boldsymbol{d}. \tag{1.3.41}$$

证明 设 $\phi(\alpha) = f(\overline{\boldsymbol{x}} + \alpha \boldsymbol{d})$，由(1.3.34)式得到

$$\phi'(\alpha) = \langle \nabla f(\overline{\boldsymbol{x}} + \alpha \boldsymbol{d}), \boldsymbol{d} \rangle = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(\overline{\boldsymbol{x}} + \alpha \boldsymbol{d}) d_i,$$

所以

$$\begin{aligned} \phi''(\alpha) &= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} f(\overline{\boldsymbol{x}} + \alpha \boldsymbol{d}) d_i d_j \\ &= \boldsymbol{d}^T \nabla^2 f(\overline{\boldsymbol{x}} + \alpha \boldsymbol{d}) \boldsymbol{d}. \end{aligned} \tag{1.3.42}$$

当 $\alpha = 0$ 时，有

$$\phi''(0) = \boldsymbol{d}^T \nabla^2 f(\overline{\boldsymbol{x}}) \boldsymbol{d}. \tag{1.3.43}$$

另一方面，由定理 1.3.1 的证明过程，得到

$$\frac{\partial}{\partial \boldsymbol{d}} f(\overline{\boldsymbol{x}}) = \frac{1}{\|\boldsymbol{d}\|} \phi'(0)$$

和

$$\frac{\partial}{\partial \boldsymbol{d}} f(\overline{\boldsymbol{x}} + \alpha \boldsymbol{d}) = \frac{1}{\|\boldsymbol{d}\|} \phi'(\alpha),$$

因此有

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{d}^2} f(\overline{\boldsymbol{x}}) &= \lim_{\alpha \to 0^+} \frac{\frac{\partial}{\partial \boldsymbol{d}} f(\overline{\boldsymbol{x}} + \alpha \boldsymbol{d}) - \frac{\partial}{\partial \boldsymbol{d}} f(\overline{\boldsymbol{x}})}{\alpha \|\boldsymbol{d}\|} \\ &= \frac{1}{\|\boldsymbol{d}\|^2} \lim_{\alpha \to 0^+} \frac{\phi'(\alpha) - \phi'(0)}{\alpha} \\ &= \frac{1}{\|\boldsymbol{d}\|^2} \phi''(0). \end{aligned} \tag{1.3.44}$$

二阶方向导数的几何意义：描述函数 $f(\boldsymbol{x})$ 在 $\overline{\boldsymbol{x}}$ 处沿方向 $\boldsymbol{d}$ 的凹凸性和弯曲的程度.

在这一小节的最后，我们介绍一下多元函数的 Taylor 展开式.

定理 1.3.3 若函数 $f(x)$ 具有连续的一阶偏导数，则在 $\bar{x}$ 处具有下列 Taylor 展开式

$$f(x)=f(\bar{x})+\langle\nabla f(\bar{x}+\theta(x-\bar{x})),x-\bar{x}\rangle,0<\theta<1,\tag{1.3.45}$$

$$f(x)=f(\bar{x})+\langle\nabla f(\bar{x}),x-\bar{x}\rangle+o(\|x-\bar{x}\|).\tag{1.3.46}$$

证明 先证明 (1.3.45) 式成立. 当 $x=\bar{x}$ 时, (1.3.45) 式显然成立，因此我们考虑 $x\neq\bar{x}$ 的情况. 令 $d=x-\bar{x}$，构造辅助函数

$$\phi(\alpha)=f(\bar{x}+\alpha d),$$

所以 $\phi(1)=f(\bar{x}+d)=f(x)$, $\phi(0)=f(\bar{x})$. 由一元函数的 Taylor 展开式和 (1.3.34) 式，得到

$$\begin{aligned}\phi(1)&=\phi(0)+\phi'(0+\theta(1-0))=\phi(0)+\phi'(\theta)\\&=f(\bar{x})+\langle\nabla f(\bar{x}+\theta d),d\rangle\\&=f(\bar{x})+\langle\nabla f(\bar{x}+\theta(x-\bar{x})),d\rangle,\end{aligned}$$

其中 $0<\theta<1$，即 (1.3.45) 式成立.

下面证明 (1.3.46) 式. 因为 $f(x)$ 具有连续的一阶偏导数，所以 $\nabla f(x)$ 为连续函数，由连续函数的性质，有

$$\nabla f(\bar{x}+\theta(x-\bar{x}))=\nabla f(\bar{x})+r(x),\tag{1.3.47}$$

其中

$$\lim_{x\to\bar{x}}r(x)=0.\tag{1.3.48}$$

将 (1.3.47) 式代入 (1.3.45) 式，得到

$$\begin{aligned}f(x)=&f(\bar{x})+\langle\nabla f(\bar{x}),x-\bar{x}\rangle\\&+\langle r(x),x-\bar{x}\rangle.\end{aligned}\tag{1.3.49}$$

下面只需证明 (1.3.49) 式的最后一项是高阶无穷小量.

由 Cauchy-Schwarz 不等式得到

$$\frac{\langle r(x),x-\bar{x}\rangle}{\|x-\bar{x}\|}\leqslant\frac{\|r(x)\|\cdot\|x-\bar{x}\|}{\|x-\bar{x}\|}=\|r(x)\|,\tag{1.3.50}$$

由（1.3.48）式，当 $x \to \bar{x}$ 时，$r(x) \to 0$，即

$$\langle r(x), x-\bar{x}\rangle = o(\|x-\bar{x}\|). \qquad (1.3.51)$$

用同样的方法可以得到下面的定理.

定理 1.3.4 若函数 $f(x)$ 具有连续的二阶偏导数，则在 $\bar{x}$ 处具有下列 Taylor 展开式

$$f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x-\bar{x}\rangle + \frac{1}{2}(x-\bar{x})^T \nabla^2 f(\bar{x}+\theta(x-\bar{x}))(x-\bar{x}), \qquad (1.3.52)$$

$$0<\theta<1,$$

$$f(x) = f(\bar{x}) + \langle \nabla f(\bar{x}), x-\bar{x}\rangle + \frac{1}{2}(x-\bar{x})^T \nabla^2 f(\bar{x})(x-\bar{x}) + o(\|x-\bar{x}\|^2). \qquad (1.3.53)$$

此外，我们还可以得到关于梯度的 Taylor 展开式

$$\nabla f(x) = \nabla f(\bar{x}) + \nabla^2 f(\bar{x})(x-\bar{x}) + o(\|x-\bar{x}\|). \qquad (1.3.54)$$

注意，（1.3.54）式的 $o(\|x-\bar{x}\|)$ 是向量.

§1.4 凸集和凸函数

凸集和凸函数在最优化的理论中是较为重要的一部分内容. 本节扼要地介绍凸集和凸函数的基本概念和基本结果.

1.4.1 凸集

定义 1.4.1 设集合 $D \subset R^n$，如果 $\forall x^{(1)}, x^{(2)} \in D$，$\forall \alpha \in [0, 1]$，均有

$$\alpha x^{(1)} + (1-\alpha)x^{(2)} \in D, \qquad (1.4.1)$$

则称集合 D 为**凸集**（comvex set）.

凸集的几何意义是：若两个点属于此集合，则这两点连线上的任意一点均属于此集合（见图 1.4.1）.

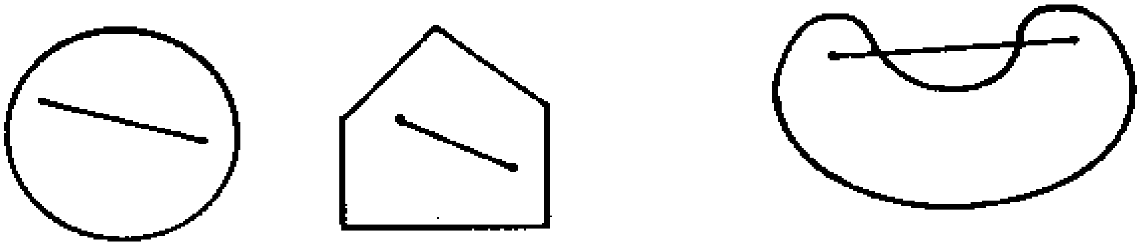


图 1.4.1 凸集与非凸集

定义 1.4.2 设 $x^{(1)}$，$x^{(2)}$，…，$x^{(m)}\in R^n$，若 $\alpha_i\geqslant 0$，$i=1$，2，…，m，$\sum_{i=1}^{m}\alpha_i=1$，则称线性组合

$$\alpha_1x^{(1)}+\alpha_2x^{(2)}+\cdots+\alpha_mx^{(m)}$$

为 $x^{(1)}$，$x^{(2)}$，…，$x^{(m)}$的**凸组合**（convex combination）.

显然，两个点的凸组合表示一条线段，三个点的凸组合表示的是一个三角形，m 个点的凸组合构成一个**凸多面体**（convex hull）（见图 1.4.2）.

由凸集的定义可知，超平面 $\{x\mid c^Tx=\alpha\}$ 是凸集，半空间 $\{x\mid c^Tx\geqslant\alpha\}$，$\{x\mid c^Tx\leqslant\alpha\}$ 是凸集，凸集的交集仍是凸集.

定理 1.4.1 D 是凸集的充分必要条件是：对任意的 $m\geqslant 2$，任给 $x^{(1)}$，$x^{(2)}$，…，$x^{(m)}\in D$ 和实数 α_1，α_2，…，α_m，且 $\alpha_i\geqslant 0$，$i=1$，2，…，m，$\sum_{i=1}^{m}\alpha_i=1$，均有

$$\alpha_1x^{(1)}+\alpha_2x^{(2)}+\cdots+\alpha_mx^{(m)}\in D. \qquad (1.4.2)$$

证明 “充分性”. 当 $m=2$ 时，由凸集的定义可知 D 是凸集.

“必要性”. 当 $m=2$ 时，由定义 1.4.1，命题成立.

假设当 $m=k$ 时命题成立，即当 $x^{(i)}\in D$，$i=1$，2，…，

图 1.4.2　凸组合的几何意义

k，$\alpha_i \geqslant 0$，$i=1, 2, \cdots, k$，$\sum_{i=1}^{k} \alpha_i = 1$ 时，有 $\sum_{i=1}^{k} \alpha_i x^{(i)} \in D$.

当 $m=k+1$ 时，$x^{(i)} \in D$，$i=1, 2, \cdots, k, k+1$，$\alpha_i \geqslant 0$，$i=1, 2, \cdots, k+1$，$\sum_{i=1}^{k+1} \alpha_i = 1$，有

$$\begin{aligned}\sum_{i=1}^{k+1} \alpha_i x^{(i)} &= \sum_{i=1}^{k} \alpha_i x^{(i)} + \alpha_{k+1} x^{(k+1)} \\ &= \left(\sum_{j=1}^{k} \alpha_j\right)\left[\sum_{i=1}^{k} \frac{\alpha_i}{\sum_{j=1}^{k} \alpha_j} x^{(i)}\right] + \alpha_{k+1} x^{(k+1)}.\end{aligned} \tag{1.4.3}$$

由于 $\sum_{i=1}^{k} \frac{\alpha_i}{\sum_{j=1}^{k} \alpha_j} = 1$，且 $\frac{\alpha_i}{\sum_{j=1}^{k} \alpha_j} \geqslant 0$，由归纳法假设有

$$\sum_{i=1}^{k} \frac{\alpha_i}{\sum_{j=1}^{k} \alpha_j} x^{(i)} \in D. \tag{1.4.4}$$

注意到 $\sum_{i=1}^{k} \alpha_i + \alpha_{k+1} = 1$，由(1.4.3)式和(1.4.4)式及凸集的定义，得到 $\sum_{i=1}^{k+1} \alpha_i x^{(i)} \in D$.

定义 1.4.3　设 D 是非空凸集，$x \in D$. 如果 x 不能表示成 D 中另外两个点的严格凸组合，则称 x 为凸集 D 的极点，即若

$$x=\alpha x^{(1)}+(1-\alpha)x^{(2)},\ \alpha\in(0,1),\ x^{(1)},\ x^{(2)}\in D,$$

则有

$$x=x^{(1)}=x^{(2)}.$$

在图 1.4.1 的凸集中，圆周上的点和多边形的顶点都是极点.

定义 1.4.4 设 D 是闭凸集，$0\neq \boldsymbol{d}\in R^n$. 如果对每个 $\boldsymbol{x}\in D$，有

$$\boldsymbol{x}+\alpha \boldsymbol{d}\in D,\forall \alpha\geqslant 0, \tag{1.4.5}$$

则称 $\boldsymbol{d}$ 是凸集 D 的方向.

考虑集合 $D=\{x\mid A\boldsymbol{x}=\boldsymbol{b},\ \boldsymbol{x}\geqslant 0\}$，由定义可知，$\boldsymbol{d}$ 是 D 的一个方向的充分必要条件是：

$$A(\boldsymbol{x}+\alpha \boldsymbol{d})=\boldsymbol{b}, \tag{1.4.6}$$

$$\boldsymbol{x}+\alpha \boldsymbol{d}\geqslant 0, \tag{1.4.7}$$

$\forall \alpha\geqslant 0$，$\boldsymbol{x}\in D$ 成立.（1.4.6）式和（1.4.7）式等价于

$$A\boldsymbol{d}=0,\boldsymbol{d}\geqslant 0,\boldsymbol{d}\neq 0. \tag{1.4.8}$$

定义 1.4.5 设 $\boldsymbol{d}$ 是 D 的一个方向. 如果 $\boldsymbol{d}$ 不能表示成两个不同方向的正线性组合，则称 $\boldsymbol{d}$ 是极方向，即如果

$$\boldsymbol{d}=\alpha_1\boldsymbol{d}^{(1)}+\alpha_2\boldsymbol{D}^{(2)},\ \alpha_1>0,\ \alpha_2>0,$$

则有

$$\boldsymbol{d}^{(1)}=\alpha \boldsymbol{d}^{(2)},\ \alpha>0.$$

下面证明凸集中较为重要的两个定理.

定理 1.4.2 设 D 是非空闭凸集，若 $\boldsymbol{y}\notin D$，则存在唯一的点 $\bar{\boldsymbol{x}}\in D$，使得它与 $\boldsymbol{y}$ 的距离最短，且满足

$$\langle \boldsymbol{x}-\bar{\boldsymbol{x}},\boldsymbol{y}-\bar{\boldsymbol{x}}\rangle\leqslant 0,\forall \boldsymbol{x}\in D. \tag{1.4.9}$$

证明 令 $\inf\{\|\boldsymbol{y}-\boldsymbol{x}\|\mid \boldsymbol{x}\in D\}=\gamma>0$. 由下确界的定义，存在序列 $\{\boldsymbol{x}^{(k)}\}\subset D$，使得 $\|\boldsymbol{y}-\boldsymbol{x}^{(k)}\|\to\gamma$.

下面证明 $\{\boldsymbol{x}^{(k)}\}$ 是 Cauchy 序列.

由平行四边形法则，有

$$\begin{aligned}\| x^{(k)} - x^{(m)} \|^2 &= 2\| y - x^{(k)} \|^2 + 2\| y - x^{(m)} \|^2 - \\ &\quad \| 2y - x^{(k)} - x^{(m)} \|^2 \\ &= 2\| y - x^{(k)} \|^2 + 2\| y - x^{(m)} \|^2 - \\ &\quad 4\| y - \frac{x^{(k)} + x^{(m)}}{2} \|^2,\end{aligned} \tag{1.4.10}$$

由于 D 是凸集，$\frac{x^{(k)} + x^{(m)}}{2} \in D$，由 γ 的定义有

$$\| y - \frac{x^{(k)} + x^{(m)}}{2} \|^2 \geqslant \gamma^2, \tag{1.4.11}$$

因而

$$\| x^{(k)} - x^{(m)} \|^2 \leqslant 2\| y - x^{(k)} \|^2 + 2\| y - x^{(m)} \|^2 - 4\gamma^2,$$

当 k 和 m 充分大时，有

$$\| x^{(k)} - x^{(m)} \|^2 \to 0,$$

因此 $\{x^{(k)}\}$ 是 Cauchy 序列. 这样存在点 $\bar{x}$，使得

$$x^{(k)} \to \bar{x}.$$

由于 D 是闭集，所以 $\bar{x} \in D$.

再证明唯一性. 假设存在着 $\bar{x}$，$\bar{x}' \in D$，满足

$$\| y - \bar{x} \| = \| y - \bar{x}' \| = \gamma, \tag{1.4.12}$$

由 D 的凸性，$\frac{\bar{x} + \bar{x}'}{2} \in D$，因此

$$\| y - \frac{\bar{x} + \bar{x}'}{2} \| \leqslant \frac{1}{2}\| y - \bar{x} \| + \frac{1}{2}\| y - \bar{x}' \| = \gamma. \tag{1.4.13}$$

如果 (1.4.13) 式中不等式严格成立，则与 γ 的定义矛盾，因此只有等号成立. 从而存在着某个常数 α，使得

$$y - \bar{x} = \alpha\ (y - \bar{x}'),$$

由 (1.4.12) 式得到 $|\alpha| = 1$. 如果 $\alpha = -1$，则有 $y = \frac{\bar{x} + \bar{x}'}{2} \in$

D，这与 $y \notin D$ 矛盾．因此 $\alpha=1$，即 $\bar{x}=\bar{x}'$，唯一性得证．

最后证明 $\bar{x}$ 满足

$$\langle x-\bar{x},\ y-\bar{x}\rangle \leqslant 0,\ \forall x \in D.$$

对于 $x \in D$，$\alpha \in (0,1)$，有 $\bar{x}+\alpha(x-\bar{x})=\alpha x+(1-\alpha)\bar{x} \in D$．因为 $\bar{x}$ 是距离 y 的最小点，因此有

$$\| y-\bar{x} \|^2 \leqslant \| y-\bar{x}-\alpha(x-\bar{x}) \|^2,$$

即有

$$0 \leqslant -2\langle y-\bar{x}, x-\bar{x}\rangle+\alpha \| x-\bar{x} \|^2, \qquad (1.4.14)$$

在 (1.4.14) 式中令 $\alpha \to 0^+$，得证．

定理 1.4.2 的几何意义见图 1.4.3．

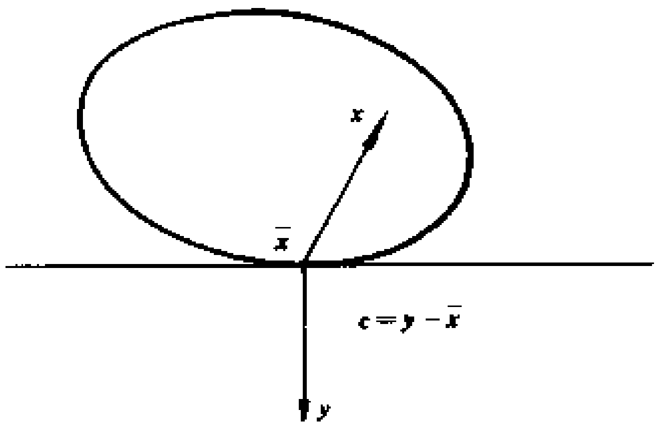


图 1.4.3

定理 1.4.3 设 D 是非空闭凸集，若 $y \notin D$，则存在 $c \in R^n$，$c \neq 0$ 和 $\alpha \in R^1$，使得

$$c^T x \leqslant \alpha, \forall x \in D, \qquad (1.4.15)$$

$$c^T y > \alpha. \qquad (1.4.16)$$

证明 由定理 1.4.2，存在 $\bar{x} \in D$，使得

$$\langle x-\bar{x},\ y-\bar{x}\rangle \leqslant 0,\ \forall x \in D.$$

令 $c=y-\bar{x}$，则

$$\langle x-\bar{x},\ c\rangle \leqslant 0,$$

即

$$c^T x \leqslant c^T \bar{x}, \quad \forall x \in D.$$

令 $\alpha = c^T \bar{x}$，因此（1.4.15）式成立.

由于 $c = y - \bar{x} \neq 0$，因此，

$$0 < \langle y - \bar{x}, y - \bar{x} \rangle = \langle y - \bar{x}, c \rangle,$$

得到

$$c^T y > c^T \bar{x} = \alpha.$$

即（1.4.16）式成立.

1.4.2 凸函数

定义 1.4.6 设 D 是非空凸集，f 是定义在 D 上的函数，如果对任意的 $x^{(1)}$，$x^{(2)} \in D$，$\alpha \in (0, 1)$，均有

$$f(\alpha x^{(1)} + (1-\alpha) x^{(2)}) \leqslant \alpha f(x^{(1)}) + (1-\alpha) f(x^{(2)}),$$

则称 f 为 D 上的**凸函数**（convex function）.

若对任意的 $x^{(1)}$，$x^{(2)} \in D$，$x^{(1)} \neq x^{(2)}$，$\alpha \in (0, 1)$ 均有

$$f(\alpha x^{(1)} + (1-\alpha) x^{(2)}) < \alpha f(x^{(1)}) + (1-\alpha) f(x^{(2)}),$$

则称 f 为 D 上的严格凸函数.

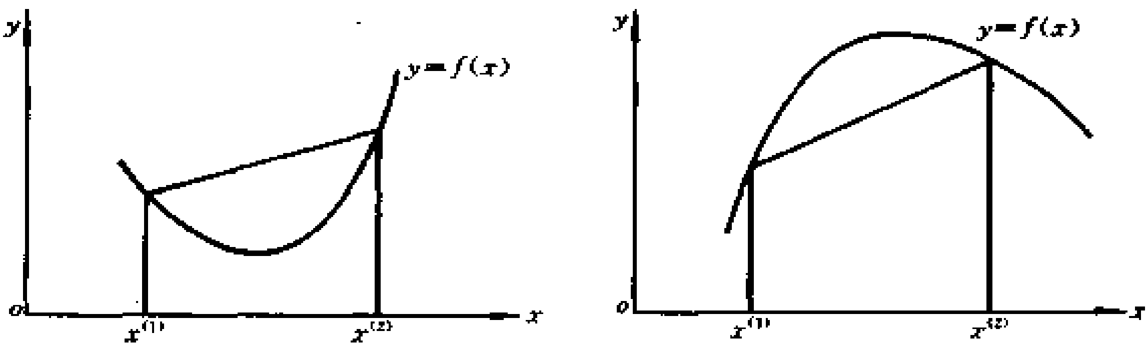


图 1.4.4 凸函数与凹函数的几何意义

若 $-f$ 为凸函数，则称 f 为**凹函数**（concave function），若 $-f$ 为严格凸函数，则称 f 为严格凹函数.

凸函数的几何意义：当 x 为单变量时，凸函数的任意两点间的曲线段总在弦的下方. 凹函数在弦的上方（见图 1.4.4）.

下列函数是 R^n 上的凸函数：

(1) $f(\boldsymbol{x})=\boldsymbol{c}^T\boldsymbol{x}$；

(2) $f(\boldsymbol{x})=\|\boldsymbol{x}\|$；

(3) $f(\boldsymbol{x})=\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$，其中 $\boldsymbol{A}$ 是正定对称矩阵.

定义 1.4.7 称集合

$$D_\alpha=\{\boldsymbol{x}\mid f(\boldsymbol{x})\leqslant\alpha,\ \boldsymbol{x}\in D\}$$

为函数 f 的**水平集** (level set).

定理 1.4.4 若 D 是非空凸集，f 是定义在 D 上的凸函数，则对任意的 $\alpha\in R$，水平集 D_α 为凸集.

证明 $\forall\boldsymbol{x}^{(1)},\ \boldsymbol{x}^{(2)}\in D_\alpha$，即 $f(\boldsymbol{x}^{(1)})\leqslant\alpha$，$f(\boldsymbol{x}^{(2)})\leqslant\alpha$，$\forall\lambda\in(0,1)$，有

$$\begin{aligned}f(\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)})&\leqslant\lambda f(\boldsymbol{x}^{(1)})+(1-\lambda)f(\boldsymbol{x}^{(2)})\\&\leqslant\lambda\alpha+(1-\lambda)\alpha=\alpha,\end{aligned}$$

所以

$$\lambda\boldsymbol{x}^{(1)}+(1-\lambda)\boldsymbol{x}^{(2)}\in D_\alpha.$$

下面讨论可微凸函数的性质.

定理 1.4.5 设 $D\subset R^n$ 为非空开凸集，$f(\boldsymbol{x})$ 在 D 上可微，则 $f(\boldsymbol{x})$ 为 D 上的凸函数的充分必要条件是：$\forall\boldsymbol{x},\ \boldsymbol{y}\in D$，恒有

$$f(\boldsymbol{y})\geqslant f(\boldsymbol{x})+(\boldsymbol{y}-\boldsymbol{x})^T\nabla f(\boldsymbol{x}).\qquad(1.4.17)$$

$f(\boldsymbol{x})$ 为 D 上的严格凸函数的充分必要条件是：$\forall\boldsymbol{x},\ \boldsymbol{y}\in D$，$\boldsymbol{x}\neq\boldsymbol{y}$，恒有

$$f(\boldsymbol{y})>f(\boldsymbol{x})+(\boldsymbol{y}-\boldsymbol{x})^T\nabla f(\boldsymbol{x}).\qquad(1.4.18)$$

证明 “必要性”.

(1) 设 $f(\boldsymbol{x})$ 为 D 上的凸函数，则 $\forall\boldsymbol{x},\ \boldsymbol{y}\in D$，$0<\alpha<1$，恒有

$$f(\alpha\boldsymbol{y}+(1-\alpha)\boldsymbol{x})\leqslant\alpha f(\boldsymbol{y})+(1-\alpha)f(\boldsymbol{x}),$$

即

$$f(\boldsymbol{x}+\alpha(\boldsymbol{y}-\boldsymbol{x}))-f(\boldsymbol{x})\leqslant\alpha[f(\boldsymbol{y})-f(\boldsymbol{x})].\tag{1.4.19}$$

在（1.4.19）两端同除 α，得到

$$\frac{f(\boldsymbol{x}+\alpha(\boldsymbol{y}-\boldsymbol{x}))-f(\boldsymbol{x})}{\alpha}\leqslant f(\boldsymbol{y})-f(\boldsymbol{x}),\tag{1.4.20}$$

在（1.4.20）式中，令 $\alpha\to0^+$，得到

$$(\boldsymbol{y}-\boldsymbol{x})^T\nabla f(\boldsymbol{x})\leqslant f(\boldsymbol{y})-f(\boldsymbol{x}),$$

故（1.4.17）式成立.

（2）设 $f(\boldsymbol{x})$ 为 D 上的严格凸函数，则有

$$f\left(\frac{\boldsymbol{x}+\boldsymbol{y}}{2}\right)<\frac{1}{2}f(\boldsymbol{x})+\frac{1}{2}f(\boldsymbol{y}).\tag{1.4.21}$$

另一方面，$f(\boldsymbol{x})$ 也是凸函数，故(1.4.17)式成立，即

$$\begin{aligned}f\left(\frac{\boldsymbol{x}+\boldsymbol{y}}{2}\right)&\geqslant f(\boldsymbol{x})+\left(\frac{\boldsymbol{x}+\boldsymbol{y}}{2}-\boldsymbol{x}\right)^T\nabla f(\boldsymbol{x})\\&=f(\boldsymbol{x})+\frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^T\nabla f(\boldsymbol{x}),\end{aligned}\tag{1.4.22}$$

将（1.4.22）式代入（1.4.21）式，有

$$\frac{1}{2}f(\boldsymbol{x})+\frac{1}{2}f(\boldsymbol{y})>f(\boldsymbol{x})+\frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^T\nabla f(\boldsymbol{x}).\tag{1.4.23}$$

对（1.4.23）式进行化简，得到（1.4.18）式.

"充分性".

(1) $\forall\,\boldsymbol{x},\boldsymbol{y}\in D$, $f(\boldsymbol{y})\geqslant f(\boldsymbol{x})+(\boldsymbol{y}-\boldsymbol{x})^T\nabla f(\boldsymbol{x})$, $\forall\,\alpha\in[0,1]$，令 $\boldsymbol{z}=\alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y}$，因此有

$$f(\boldsymbol{x})\geqslant f(\boldsymbol{z})+(\boldsymbol{x}-\boldsymbol{z})^T\nabla f(\boldsymbol{z}),\tag{1.4.24}$$

$$f(\boldsymbol{y})\geqslant f(\boldsymbol{z})+(\boldsymbol{y}-\boldsymbol{z})^T\nabla f(\boldsymbol{z}),\tag{1.4.25}$$

在（1.4.24）式上乘 α，在（1.4.25）式上乘（$1-\alpha$），两式相加得到

$$\alpha f(\boldsymbol{x})+(1-\alpha)f(\boldsymbol{y})\geqslant f(\boldsymbol{z})=f(\alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y}),$$

所以 f 为凸函数.

(2) $\forall x, y \in D, f(y) > f(x) + (y-x)^T \nabla f(x)$,其证明过程与前面相同,只需将(1.4.24)式和(1.4.25)式中的"$\geqslant$"换成"$>$"即可.

可微凸函数的几何意义:对于单变量可微凸函数,其曲线仅次于其切线的上方(见图 1.4.5).

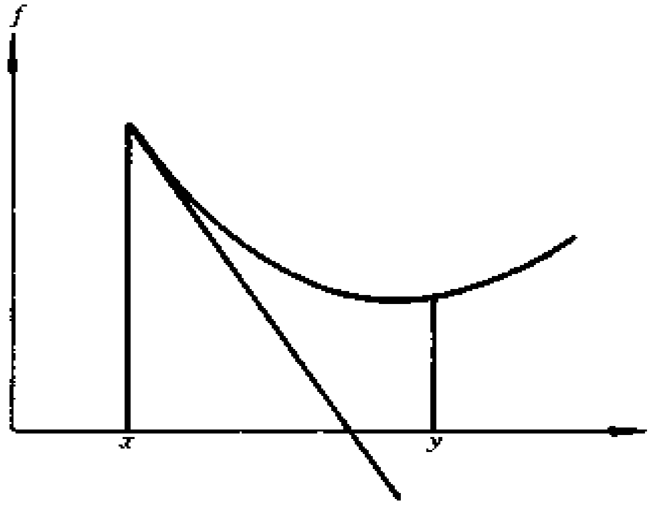


图 1.4.5 可微凸函数的几何意义

再讨论二次可微凸函数的性质.

定理 1.4.6 设 $D \subset R^n$ 为非空开凸集,$f(x)$ 在 D 上二次可微,则 $f(x)$ 为 D 上的凸函数的充分必要条件是:$\forall x \in D$,$\nabla^2 f(x)$半正定.若$\forall x \in D, \nabla^2 f(x)$为正定矩阵,则$f(x)$为 D 上的严格凸函数.

证明 先证明定理的第一部分.

"充分性".

设$\nabla^2 f(x)(x \in D)$半正定,$\forall x, y \in D$,由二阶 Taylor 展式

$$\begin{aligned} f(y) &= f(x) + (y-x)^T \nabla f(x) + \\ &\quad \frac{1}{2}(y-x)^T \nabla^2 f(x + \theta(y-x))(y-x) \\ &\geqslant f(x) + (y-x)^T \nabla f(x). \end{aligned} \tag{1.4.26}$$

注意到 (1.4.26) 式中 $\theta\in(0,1)$, $x+\theta(y-x)=\theta y+(1-\theta)x\in D$, 由定理1.4.5, $f(x)$ 是凸函数.

"必要性".

设 $f(x)$ 是 D 上的凸函数, 由定理1.4.5, $\forall x, y\in D$, 有

$$f(y)\geqslant f(x)+(y-x)^T\nabla f(x). \tag{1.4.27}$$

对于任意的 $x\in D$, 任意的 $d\neq 0$, $d\neq R^n$, 由于 D 是开集, $\exists\delta>0$,使得当 $\alpha\in(0,\delta)$ 时, $x+\alpha d\in D$. 应用 (1.4.27) 式得到

$$f(x+\alpha d)\geqslant f(x)+\alpha d^T\nabla f(x), \tag{1.4.28}$$

另一方面, 由二阶 Taylor 展开式, 得到

$$f(x+\alpha d)=f(x)+\alpha d^T\nabla f(x)+\frac{1}{2}\alpha^2 d^T\nabla^2 f(x)d+o(\alpha^2), \tag{1.4.29}$$

比较 (1.4.28) 式和 (1.4.29) 式, 得到

$$\frac{1}{2}\alpha^2 d^T\nabla^2 f(x)d+o(\alpha^2)\geqslant 0, \tag{1.4.30}$$

在 (1.4.30) 式两端同除 α^2, 并令 $\alpha\to 0^+$, 得到

$$d^T\nabla^2 f(x)d\geqslant 0,$$

即$\nabla^2 f(x)$半正定.

再证第二部分.

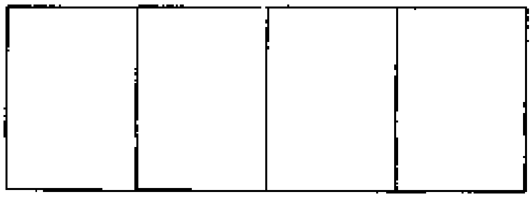
由于$\nabla^2 f(x+\theta(y-x))$正定,所以(1.4.26)式改为严格大于号即可.

习题一

1.1 设经验模型为 $y=\beta_0+\beta_1x_1+\beta_2x_2$, 且已知 N 给数据 (x_{1i}, x_{2i}), y_i, $i=1,2,\cdots,N$. 今欲选择 β_0、β_1 和 β_2, 使按模型计算出的值与实测值偏离的平方和最小, 试导出相应的最

优化问题.

1.2 某单位拟建一排四间的猪舍，平面布置如下图所示.由于资金及材料的限制，围墙和隔墙的总长度不能超过 40 米.为使猪舍面积最大，问如何选择长、宽尺寸？试建立相应的最优化问题.



1.3 用图解法求解无约束问题

$$\min \quad f(\boldsymbol{x})=x_1^2+4x_2^2-4x_1-8x_2.$$

1.4 用图解法求解约束问题

$$\min \quad f(\boldsymbol{x}) = x_1^2 + x_2^2,$$

$$\text{s.t.} \quad c(\boldsymbol{x}) = (x_1 - 1)^3 - x_2^2 = 0.$$

1.5 求下列函数的梯度和 Hesse 矩阵：

(1) $f(\boldsymbol{x})=3x_1x_2^2+4e^{x_1x_2}$；

(2) $f(\boldsymbol{x})=x_1^{x_2}+\ln(x_1x_2)$；

(3) $f(\boldsymbol{x})=x_1^2+x_2^2+x_3^2$；

(4) $f(\boldsymbol{x})=\ln(x_1^2+x_1x_2+x_2^2)$.

1.6 设 $D\subset R^n$ 为凸集，A 为 $m\times n$ 阶矩阵，试证集合

$$A(D) = \{y \mid y=\boldsymbol{A}\boldsymbol{x},\ \boldsymbol{x}\in D\}$$

是凸集.

1.7 证明两个凸集的交集是凸集. 问：两个凸集的并集是否为凸集？证明或举出反例.

1.8 证明凸函数的和函数是凸函数.

1.9 判断下列函数是否为凸函数或凹函数.

(1) $f(\boldsymbol{x})=x_1^2+2x_1x_2-10x_1+5x_2$；

(2) $f(\boldsymbol{x})=-x_1^2+2x_1x_2-5x_2^2+10x_1-10x_2$；

(3) $f(\boldsymbol{x})=x_1x_2+2x_1^2+x_2^2+2x_3^2-6x_1x_3$.

1.10　设 f 为可微凸函数. 证明 f 为线性函数的充分必要条件是: f 即是凸函数又是凹函数.

1.11　设 $D\subset R^n$ 为凸集, f ($\boldsymbol{x}$) 是定义在 D 上的凸函数的充要条件是: $\forall\, m\geqslant 2$, $\alpha_i\geqslant 0$, $i=1, 2, \cdots, m$, $\sum_{i=1}^{m}\alpha_i=1$, $\boldsymbol{x}^{(i)}\in D$, $i=1, 2, \cdots, m$, 恒有

$$f\Big(\sum_{i=1}^{m}\alpha_i\boldsymbol{x}^{(i)}\Big)\leqslant\sum_{i=1}^{m}\alpha_i f\ (\boldsymbol{x}^{(i)}).$$

1.12　证明不等式. 若 $x_i\geqslant 0$, $i=1, 2, \cdots, n$, 有

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i}\geqslant\frac{1}{n}\sum_{i=1}^{n}\sqrt{x_i}.$$

第二章　线性规划

§2.1 引　　言

本书在第二章和第三章介绍线性规划问题. 线性规划是数学规划的一个重要分支，历史比较悠久，理论比较成熟，方法较为完善. 线性规划的思想最早可以追溯到1939年，当时的苏联数学家、经济学家 Л. В. Канторовнч（康特罗维奇）在《生产组织与计划中的数学方法》一书中提出了类似线性规划的模型，以解决下料问题和运输问题，并给出了"解决乘数法"的求解方法. 然而，他们的工作长期未被人们知道. 由于战争的需要，美国的经济学家 T. C. Koopmans（库普曼斯）重新独立的研究运输问题，并很快看到了线性规划在经济学中应用的意义. 在这之后，线性规划也被人们广泛地用于军事、经济等各方面. 由于他们在这方面的突出贡献，康特罗维奇和库普曼斯联合得到1975年诺贝尔经济学奖.

对线性规划贡献最大的是美国数学家 G. B. Dantzig（丹捷格），他在1947年提出了求解线性规划的**单纯形法**（Simplex Method），并同时给出了许多很有价值的理论，为线性规划奠定了理论基础. 在1953年，丹捷格又提出了改进单纯形法，1954年 Lemke（兰姆凯）提出了**对偶单纯形法**（dual Simplex method）.

在1976年，R. G. Bland 提出避免出现循环的方法后，使线性规划的理论更加完善. 但在1972年，V. Klee 和 G. Minmty 构造了一个例子，发现单纯形法的迭代次数是指数次运算，

不是好算法——并不是多项式算法（多项式算法被认为是好算法），这对单纯形法提出了挑战.

1979 年、前苏联青年数学家 Л. Г. Фачиян（哈奇扬）提出了一种新算法——椭球算法，是一个多项式算法. 这一结果在全世界引起了极大轰动，被认为是线性规划理论上的历史突破. 然而在实际计算中，椭球算法的计算量与单纯形算法差不多，因此，椭球算法并不实用.

1984 年，在美国的贝尔实验室工作的印度数学家 N. Karmarkar（卡玛卡尔）又提出了一个多项式算法——Karmarkar 算法，Karmarkar 算法本质上属于内点法，这种算法不仅在理论上优于单纯形法，而且也显示出对求解大规模实际问题的巨大潜力.

此外，1980 年前后，形成求解线性规划的**有效集法**（active set method)，在理论上有效集法与单纯形法是等价的，但解决问题的侧重点不同，因此各有优缺点，起着相互补充的作用.

本书仅介绍单纯形法和修正单纯形法以及对偶单纯形法.

§ 2.2 线性规划的数学模型

2.2.1 数学模型

在第一章的概论中，我们已经介绍了线性规划问题的基本情况，这里我们对线性规划问题作详细的介绍.

【例 2.2.1】 运输问题.

设有某种物资需要从 m 个产地 A_1，A_2，…，A_m 运到 n 个销地 B_1，B_2，…，B_n，其中每个产地的产量为 a_1，a_2，…，a_m，每个销地的销量为 b_1，b_2，…，b_n，并且满足产销平衡，即 $\sum\limits_{i=1}^{m} a_i = \sum\limits_{j=1}^{n} b_j$. 设从产地 A_i 到销地 B_j 的运费单价为 c_{ij}（$i=$

$1, 2, \cdots, m, j=1, 2, \cdots, n$)，问如何调运使总运费最少?

解 设从产地 A_i 到销地 B_j 的运量为 x_{ij} 个单位，则总运费为 $\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}$，并且满足

$$\sum_{j=1}^{n}x_{ij}=a_i, i=1,2,\cdots,m(\text{第 } i \text{ 个产地的产量为} a_i);$$

$$\sum_{i=1}^{m}x_{ij}=b_j, j=1,2,\cdots,n(\text{第 } j \text{ 个销地的销量为} b_j);$$

$$x_{ij}\geqslant 0, i=1,2,\cdots,m, j=1,2,\cdots,n(\text{自然要求}).$$

因此运输问题的数学表达形式为

$$\begin{aligned}
\min\quad & \sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij},\\
\text{s. t.}\quad & \sum_{j=1}^{n}x_{ij}=a_i,\ i=1,2,\cdots,m,\\
& \sum_{i=1}^{m}x_{ij}=b_j,\ j=1,2,\cdots,n,\\
& x_{ij}\geqslant 0,\ i=1,2,\cdots,m,\ j=1,2,\cdots,n.
\end{aligned} \tag{2.2.1}$$

具有 (2.2.1) 式的线性规划问题称为**运输问题** (transportation problem). 运输问题 (2.2.1) 的求解可用后面讲到的求解一般线性规划的方法. 另外，还可用特殊的方法来求解，这里就不多做介绍了.

下面我们给出线性规划的标准形式

$$\begin{aligned}
\min\quad & c_1x_1+c_2x_2+\cdots+c_nx_n,\\
\text{s. t.}\quad & a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1,\\
& a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2,\\
& \vdots\\
& a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m,\\
& x_1\geqslant 0,\ x_2\geqslant 0,\ \cdots,\ x_n\geqslant 0,
\end{aligned} \tag{2.2.2}$$

写成紧凑格式

$$
\begin{aligned}
\min \quad & \sum_{j=1}^{n} c_j x_j , \\
\text{s. t.} \quad & \sum_{j=1}^{n} a_{ij} x_j = b_j , \quad i = 1, 2, \cdots, m , \qquad (2.2.3) \\
& x_j \geqslant 0, \quad j = 1, 2, \cdots, n ,
\end{aligned}
$$

和矩阵形式

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^T \boldsymbol{x} , \\
\text{s. t.} \quad & A\boldsymbol{x} = \boldsymbol{b} , \qquad (2.2.4) \\
& \boldsymbol{x} \geqslant 0,
\end{aligned}
$$

其中

$$
\boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},
$$

$$
\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.
$$

2.2.2 将一般线性规划化为标准形式

对于一般线性规划问题，目标函数有可能求极小，也有可能求极大．除等式约束外，还有可能有不等式约束．对于每个变量 x_j 不一定都有非负限制．对于一个一般的线性规划问题，我们首先将它化为标准形式．

（1）若目标是求极大．$\max \quad \boldsymbol{c}^T\boldsymbol{x}$ 等价于 $\min - \boldsymbol{c}^T\boldsymbol{x}$．

（2）若约束条件是不等式约束．

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

等价于

$$\begin{cases}\sum_{j=1}^{n} a_{ij}x_j + x_{n+i} = b_i, \\ x_{n+i} \geqslant 0,\end{cases}$$

此时称 x_{n+i}为**松驰变量**（slack variable）．不等式约束

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

等价于

$$\begin{cases}\sum_{j=1}^{n} a_{ij}x_j - x_{n+i} = b_i, \\ x_{n+i} \geqslant 0,\end{cases}$$

此时称 x_{n+i}为**剩余变量**（surplus variable）或称为松弛变量．

(3) 若某个 x_j 无限制．引进两个非负变量 $x'_j \geqslant 0$，$x''_j \geqslant 0$，令 $x_j = x'_j - x''_j$ 代入目标函数和约束方程中，化为非负限制．

§2.3 线性规划的基本性质

定义 2.3.1 考虑标准线性规划问题（2.2.4）．若 $\boldsymbol{x}$ 满足方程 $A\boldsymbol{x} = \boldsymbol{b}$，$\boldsymbol{x} \geqslant 0$．则称 $\boldsymbol{x}$ 为**可行解**（feasible solution），称

$$D = \{\boldsymbol{x} \mid A\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geqslant 0\} \qquad (2.3.1)$$

为可行解集．

2.3.1 解线性规划的图解法

为了便于理解后面所讲的理论，在这里我们先介绍一下线性规划的几何意义——求解线性规划的图解法．

【例 2.3.1】 用图解法求解线性规划问题：

$$
\begin{aligned}
\min \quad & -2x_1-5x_2, \\
\text{s.t.} \quad & x_1+2x_2\leqslant 8, \\
& x_1\leqslant 4, \\
& x_2\leqslant 3, \\
& x_1,x_2\geqslant 0.
\end{aligned}
$$

解 首先画出可行解集（见图 2.3.1）. 事实上，可行解集 D 是半空间的交集. 再画出等高线，

$$-2x_1-5x_2=k.$$

注意到 k 取最小值的点在方程 $x_1+2x_2=8$ 和 $x_3=3$ 的交点处，联立方程，得到解 $x_1=2$，$x_2=3$，此时目标值为 $f^*=-19$.

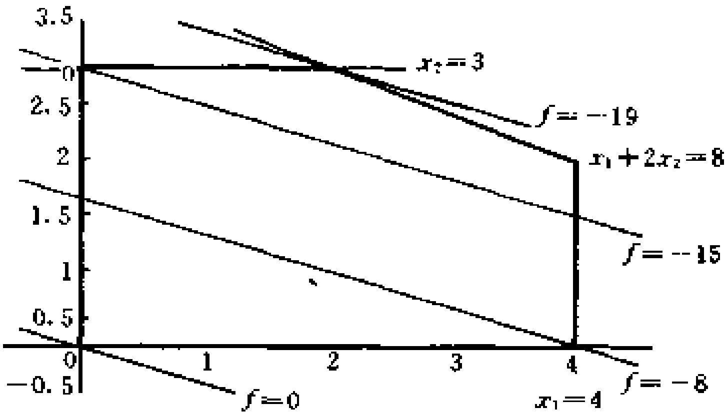


图2.3.1 例2.3.1的几何意义

本题的特点是：可行解集有界，有最优解且最优解唯一.

【例 2.3.2】 用图解法求解线性规划问题：

$$
\begin{aligned}
\min \quad & -x_1-2x_2, \\
\text{s.t.} \quad & x_1+2x_2\leqslant 8, \\
& x_1\leqslant 4, \\
& x_2\leqslant 3, \\
& x_1,x_2\geqslant 0.
\end{aligned}
$$

解 可行解集 D 与例 2.3.1 相同，画出等高线，并注意到等高线 $-x_1-2x_2=-8$ 与方程 $x_1+2x_2=8$ 是重叠的. 因此点（2，3）与点（4.2）之间的线段上的点均是最优解.

本题的特点是：可行解集有界，有无穷个最优解.

【例 2.3.3】 用图解法求解线性规划问题：

$$\begin{aligned} \min \quad & 2x_1 + 2x_2, \\ \text{s.t.} \quad & x_1 - x_2 \geqslant 1, \\ & -x_1 + 2x_2 \leqslant 0, \\ & x_1, x_2 \geqslant 0. \end{aligned}$$

解 先画出可行解集 D（见图 2.3.2），再画出等高线，联立方程$\begin{cases} x_2 = 0, \\ x_1 - x_2 = 1, \end{cases}$得到最优点 $x^* = (1, 0)^T$，$f^* = 2$.

本题的特点是：可行解集无界，但有最优解.

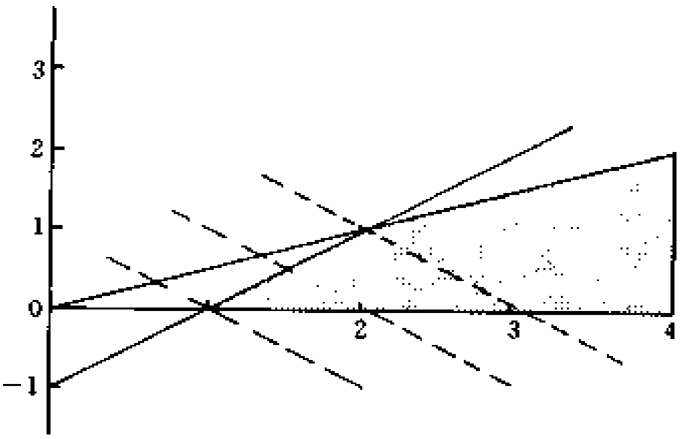


图 2.3.2

【例 2.3.4】 用图解法求解线性规则问题：

$$\begin{aligned} \min \quad & -2x_1 - 2x_2, \\ \text{s.t.} \quad & x_1 - x_2 \geqslant 1, \\ & -x_1 + 2x_2 \leqslant 0, \\ & x_1, x_2 \geqslant 0. \end{aligned}$$

解 注意到可行解集无界，当 x_1、x_2 增大时，f 减少，因此无有限最优解，也称为无最优解.

本题的特点是：可行解集无界，无最优解.

【例 2.3.5】 用图解法求解线性规划问题：

$$\begin{aligned}\min\quad & 2x_1+2x_2,\\ \text{s.t.}\quad & x_1-x_2\leqslant -1,\\ & x_1+x_2\leqslant -1,\\ & x_1,x_2\geqslant 0.\end{aligned}$$

解 考虑区域 $D_1=\{x\mid x_1-x_2\leqslant -1,\ x_1+x_2\leqslant -1\}$ 和区域 $D_2=\{x\mid x_1\geqslant 0,\ x_2\geqslant 0\}$（见图 2.3.3）. 由于 $D_1\cap D_2=\phi$，因此问题无可行解，当然也没有最优解.

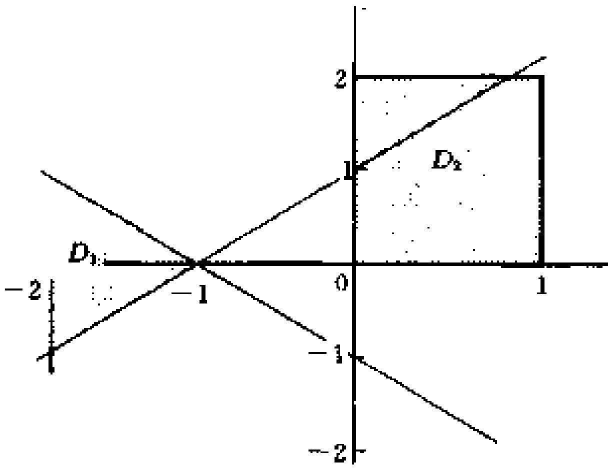


图 2.3.3

本题的特点是：无可行解，即无最优解.

综上所述，我们可以得到以下结论：

(1) 有的线性规划问题有一个最优解，有的有无穷个最优解，有的没有最优解.

(2) 若线性规划问题的最优解存在，则最优解必在可行解集 D 的某个“顶点”处达到.

(3) 若某两个顶点是最优解，则这两个顶点的联线上任取一点都是最优解.

2.3.2 线性规划的基本概念

考虑线性规划问题

$$
\begin{aligned}
\min \quad & \boldsymbol{c}^T\boldsymbol{x} \\
\text{s.t.} \quad & A\boldsymbol{x} = \boldsymbol{b}, \\
& \boldsymbol{x} \geqslant 0,
\end{aligned}
\tag{2.3.2}
$$

其中 $A=A_{m\times n}$，$m<n$，$\boldsymbol{c}\in R^n$，$\boldsymbol{b}\in R^m$，并假设 rank（A）$=m$，即 A 是行满秩的.

设

$$
A = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_m, \boldsymbol{p}_{m+1}, \cdots, \boldsymbol{p}_n],
$$

其中

$$
\boldsymbol{p}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}
\tag{2.3.3}
$$

是矩阵 A 的第 j 列.

由于 rank（A）$=m$，则存在 m 列向量线性无关，为了方便起见，不妨设 A 前的 m 列向量 $\boldsymbol{p}_1$，$\boldsymbol{p}_2$，…，$\boldsymbol{p}_m$ 线性无关，记

$$
A = [B, N],
\tag{2.3.4}
$$

则 $B=B_{m\times m}$且 det（B）$\neq 0$，即 B 是非奇异矩阵.

显然，矩阵 A 的任一列均可以由矩阵 B 的 m 列线性组合表示出来.

定义 2.3.2 称 $B=$（$\boldsymbol{p}_1$，$\boldsymbol{p}_2$，…，$\boldsymbol{p}_m$）为线性规划的**基矩阵**（basic matrix）或简称为**基**（basis），称 $\boldsymbol{p}_1$，$\boldsymbol{p}_2$，…，$\boldsymbol{p}_m$ 为基向量，与其向量对应的 x_1，x_2，…，x_m 称为**基变量**（basic variable），记作 $\boldsymbol{x}_B=$（x_1，x_2，…，x_m）^{T}，称 A 中其余的

$n-m$ 个列向量 $\boldsymbol{p}_{m+1}$，$\boldsymbol{p}_{m+2}$，…，$\boldsymbol{p}_n$ 为非基向量，记 $N=(\boldsymbol{p}_{m+1}, \boldsymbol{p}_{m+2}, \cdots, \boldsymbol{p}_n)$ 称为**非基矩阵**（nonbasic matrix），称与非基向量对应的变量 x_{m+1}，x_{m+2}，…，x_n 为**非基变量**（nonbasic variable），记 $\boldsymbol{x}_N=(x_{m+1}, x_{m+2}, \cdots, x_n)^T$.

考虑约束方程组

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{2.3.5}$$

作分块处理，令

$$A = [B, N], \qquad \boldsymbol{x} = \begin{bmatrix} x_B \\ x_N \end{bmatrix},$$

代入（2.3.5）式，得到

$$B\boldsymbol{x}_B + N\boldsymbol{x}_N = \boldsymbol{b}, \tag{2.3.6}$$

即

$$\boldsymbol{x}_B = B^{-1}\boldsymbol{b} - B^{-1}N\boldsymbol{x}_N, \tag{2.3.7}$$

当非基变量 $\boldsymbol{x}_N=0$ 时，基变量 $\boldsymbol{x}_B=B^{-1}b$.

定义 2.3.3 称

$$x = \begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{bmatrix} = \begin{bmatrix} B^{-1}\boldsymbol{b} \\ 0 \end{bmatrix} \tag{2.3.8}$$

为线性规划对应于基 B 的**基本解**（basic solution）.

由于矩阵 A 有 n 个列向量，从中任取 m 个线性无关的列向量就构成线性规划的一个基，而每个基对应一个基本解，因此基本解的个数最多有 $C_n^m=\dfrac{n!}{m!\,(n-m)!}$ 个，即基本解的个数是有限个.

对于任一个基本解，至少有 $n-m$ 个分量为零，因此，一个基本解至多有 m 个非零分量.

定义 2.3.4 称非零分量个数小于 m 的基本解为**退化的基本解**（degenerate basic solution）；否则称为**非退化的基本解**（nondegenerate basic solution）. 如果线性规划问题的所有基本解

都是非退化的，则称线性规划问题是非退化的；否则称为退化的.

显然，非退化基本解非零分量的个数恰好为 m.

【例 2.3.6】 求线性规划问题

$$
\begin{aligned}
\min \quad & -2x_1 - 5x_2, \\
\text{s.t.} \quad & x_1 + 2x_2 \leqslant 8, \\
& x_1 \leqslant 4, \\
& x_2 \leqslant 3, \\
& x_1, x_2 \geqslant 0
\end{aligned}
$$

所有基及对应的基本解.

解 引进松弛变量 x_3、x_4、x_5，将问题化成标准形式：

$$
\begin{aligned}
\min \quad & -2x_1 - 5x_2, \\
\text{s.t.} \quad & x_1 + 2x_2 + x_3 = 8, \\
& x_1 + x_4 = 4 \\
& x_2 + x_5 = 3, \\
& x_1, x_2, \cdots, x_5 \geqslant 0,
\end{aligned}
$$

其中

$$
A = \begin{bmatrix} 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 8 \\ 4 \\ 3 \end{bmatrix}.
$$

由于

$$
\det(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3) = \begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} \neq 0,
$$

所以 $\boldsymbol{p}_1$、$\boldsymbol{p}_2$、$\boldsymbol{p}_3$ 线性无关，从而可以构成一组基

$$
B^{(1)} \doteq (\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3),
$$

基变量为 $\boldsymbol{x}_B = (x_1, x_2, x_3)^T$，非基变量是 $\boldsymbol{x}_N = (x_4, x_5)^T$.

我们有

$$B^{(1)}=\begin{bmatrix}1&2&1\\1&0&0\\0&1&0\end{bmatrix},\ B^{(1)^{-1}}=\begin{bmatrix}0&1&0\\0&0&1\\1&-1&-2\end{bmatrix},$$

$$B^{(1)^{-1}}b=\begin{bmatrix}0&1&0\\0&0&1\\1&-1&-2\end{bmatrix}\begin{bmatrix}8\\4\\3\end{bmatrix}=\begin{bmatrix}4\\3\\-2\end{bmatrix}.$$

因此，$\boldsymbol{x}_B=(4,\ 3,\ -2)^T$，$\boldsymbol{x}_N=(0,\ 0)^T$. 基本解为 $\boldsymbol{x}^{(1)}=(4,\ 3,\ -2,\ 0,\ 0)^T$.

类似地，我们可以求出其它基本解，并列表如下：

基	基变量	基本解	可行否	函数值
$B^{(1)}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3)$	$\boldsymbol{x}_B^{(1)}=(x_1,x_2,x_3)^T$	$\boldsymbol{x}^{(1)}=(4,3,-2,0,0)^T$	否	—
$B^{(2)}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_4)$	$\boldsymbol{x}_B^{(2)}=(x_1,x_2,x_4)^T$	$\boldsymbol{x}^{(2)}=(2,3,0,2,0)^T$	可行	−19
$B^{(3)}=(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_5)$	$\boldsymbol{x}_B^{(3)}=(x_1,x_2,x_5)^T$	$\boldsymbol{x}^{(3)}=(4,2,0,0,1)^T$	可行	−18
$B^{(4)}=(\boldsymbol{p}_1,\boldsymbol{p}_3,\boldsymbol{p}_5)$	$\boldsymbol{x}_B^{(4)}=(x_1,x_3,x_5)^T$	$\boldsymbol{x}^{(4)}=(4,0,4,0,3)^T$	可行	−8
$B^{(5)}=(\boldsymbol{p}_1,\boldsymbol{p}_4,\boldsymbol{p}_5)$	$\boldsymbol{x}_B^{(5)}=(x_1,x_4,x_5)^T$	$\boldsymbol{x}^{(5)}=(8,0,0,-4,3)^T$	否	—
$B^{(6)}=(\boldsymbol{p}_2,\boldsymbol{p}_3,\boldsymbol{p}_4)$	$\boldsymbol{x}_B^{(6)}=(x_2,x_3,x_4)^T$	$\boldsymbol{x}^{(6)}=(0,3,2,4,0)^T$	可行	−15
$B^{(7)}=(\boldsymbol{p}_2,\boldsymbol{p}_4,\boldsymbol{p}_5)$	$\boldsymbol{x}_B^{(7)}=(x_2,x_4,x_4)^T$	$\boldsymbol{x}^{(7)}=(0,4,0,4,-1)^T$	否	—
$B^{(8)}=(\boldsymbol{p}_3,\boldsymbol{p}_4,\boldsymbol{p}_5)$	$\boldsymbol{x}_B^{(8)}=(x_3,x_4,x_5)^T$	$\boldsymbol{x}^{(8)}=(0,0,8,4,3)^T$	可行	0

表 2.3.1　例 2.3.6 的全部基本解

去掉松驰变量，得到的点是 $\boldsymbol{x}^{(1)}=(4,\ 3)$，$\boldsymbol{x}^{(2)}=(2,\ 3)$，$\boldsymbol{x}^{(3)}=(4,\ 2)$，$\boldsymbol{x}^{(4)}=(4,\ 0)$，$\boldsymbol{x}^{(5)}=(8,\ 0)$，$\boldsymbol{x}^{(6)}=(0,\ 3)$，$\boldsymbol{x}^{(7)}=(0,\ 4)$，$\boldsymbol{x}^{(8)}=(0,\ 0)$（见图 2.3.4）.

定义 2.3.5　若 $\boldsymbol{x}$ 是可行解又是基本解，则称 $\boldsymbol{x}$ 为**基本可行解**（basic feasible solution），此时称 B 为可行基.

事实上，基本可行解就是满足非负条件的基本解，即基本可行解的充分必要条件是：

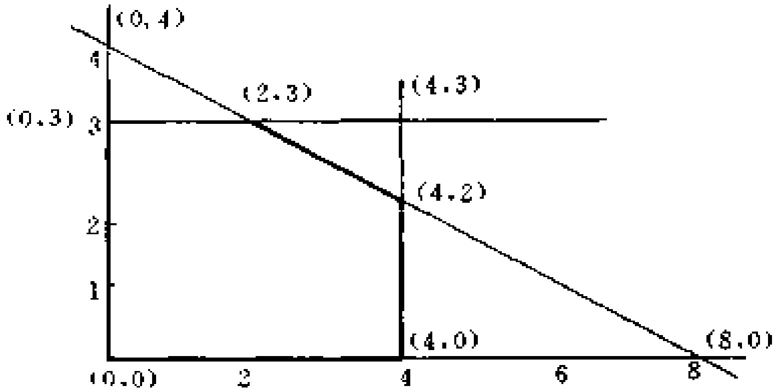


图 2.3.4

$$x_B = B^{-1}b \geqslant 0,\ x_N = 0. \tag{2.3.9}$$

在例 2.3.6 中，$x^{(2)}$、$x^{(3)}$、$x^{(4)}$、$x^{(6)}$、$x^{(8)}$是基本可行解.

使目标函数值达到最小值的可行解称为线性规划的**最优解**(optimal solution).

定义 2.3.6 设

$$D = \{x \mid Ax = b, x \geqslant 0\}, \tag{2.3.10}$$

如果$\exists x^* \in D$，$\forall x \in D$，恒有

$$c^T x \geqslant c^T x^*, \tag{2.3.11}$$

则称 x^* 为线性规划的最优解.

定义 2.3.7 若 x^* 是最优解又是基本可行解，则称 x^* 为**最优基本可行解** (optimal basic feasible solution).

在例 2.3.6 中，最优解就是最优基本可行解，即 $x^* = (2, 3, 0, 2, 0)^T$，其最优目标函数值为 $f^* = -19$.

2.3.3 线性规划的基本性质

引理 2.3.1 设 $\bar{x} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n)^T$ 是线性方程组 Ax

$=b$ 的一个解，则 $\bar{x}$ 是基本解的充分必要条件是：$\bar{x}$ 的非零分量 $\bar{x}_{i_1}$，$\bar{x}_{i_2}$，…，$\bar{x}_{i_r}$ 对应的向量 $\bar{p}_{i_1}$，$\bar{p}_{i_2}$，…，$\bar{p}_{i_r}$ 线性无关.

证明 "必要性"，由定义直接得到.

"充分性"，因为 rank（A）$=m$，因此有 $r\leqslant m$. 故存在向量 p_{ir+1}，p_{ir+2}，…，p_{im}，使得 p_{i1}，p_{i2}，…，p_{im} 线性无关，构成基向量.

定理 2.3.1 如果线性规划问题有可行解，则一定有基本可行解.

证明 设 $\bar{x}$ 是线性规划问题的一个可行解，不妨设其非零分量为 $\bar{x}_1$，$\bar{x}_2$，…，$\bar{x}_r$. 若 $\bar{x}$ 不是基本可行解，则其对应的向量 p_1，p_2，…，p_r 线性相关，即存在着 δ_1，δ_2，…，δ_r 不全为0，使得

$$\sum_{i=1}^{r}\delta_i p_i = 0. \tag{2.3.12}$$

令 $\delta_{r+1}=\delta_{r+2}=\delta_n=0$，$\boldsymbol{\delta}=(\delta_1, \delta_2, \cdots, \delta_n)^T$，取相当小的正数 α，使得

$$\bar{x}_i \pm \alpha\delta_i \geqslant 0,\ i = 1,2,\cdots,n. \tag{2.3.13}$$

取 $x=\bar{x}\pm\alpha\delta$，显然有

$$\begin{aligned} Ax &= \sum_{i=1}^{n} p_i x_i = \sum_{i=1}^{n} pi(\bar{x}_i \pm \alpha\delta_i) = \sum_{i=1}^{n} p_i\bar{x}_i \pm \alpha\sum_{i=1}^{n} p_i\delta_i \\ &= \sum_{i=1}^{n} p_i\bar{x}_i = A\bar{x} = b, \end{aligned} \tag{2.3.14}$$

所以，在满足不等式

$$\bar{x}_i + \alpha\delta_i \geqslant 0,\ \bar{x}_i - \alpha\delta_i \geqslant 0,\ i = 1,2,\cdots,r$$

的同时，$\bar{x}+\alpha\boldsymbol{\delta}$ 和 $\bar{x}-\alpha\boldsymbol{\delta}$ 是两个可行解. 可以选择 α，使上述诸式中至少有一个取等号. 这时就得到一个可行解 $\bar{x}+\alpha\boldsymbol{\delta}$ 或 $\bar{x}-\alpha\boldsymbol{\delta}$，它的非零分量数至少比 $\bar{x}$ 少一个. 如果这个可行解还不是基本可行解，我们就可以仿照上法继续作下去…，由于只有一个

非零分量对应的列向量一个是线性无关的，因此必存在基本可行解.

定理 2.3.2 如果线性规划问题有最优解，则一定有最优基本可行解.

证明 设 $\bar{\boldsymbol{x}}$ 是最优解，若 $\bar{\boldsymbol{x}}$ 不是基本可行解，则由定理 2.3.1 的证明过程可知，存在向量 $\boldsymbol{\delta}\neq 0$ 和常数 $\alpha>0$，使得 $\bar{\boldsymbol{x}}+\alpha\boldsymbol{\delta}$ 和 $\bar{\boldsymbol{x}}-\alpha\boldsymbol{\delta}$ 是可行解，并且有

$$f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{\delta})=f(\bar{\boldsymbol{x}})+f(\alpha\boldsymbol{\delta})\geqslant f(\bar{\boldsymbol{x}}), \tag{2.3.15}$$

$$f(\bar{\boldsymbol{x}}-\alpha\boldsymbol{\delta})=f(\bar{\boldsymbol{x}})-f(\alpha\boldsymbol{\delta})\geqslant f(\bar{\boldsymbol{x}}), \tag{2.3.16}$$

因此有，$f(\alpha\boldsymbol{\delta})=0$. 由此得到

$$f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{\delta})=f(\bar{\boldsymbol{x}}-\alpha\boldsymbol{\delta})=f(\bar{\boldsymbol{x}}), \tag{2.3.17}$$

按照定理 2.3.1 的证明方法继续作下去，最后得到最优基本可行解且目标函数值仍等于 $f(\bar{\boldsymbol{x}})$.

引理 2.3.2 可行解集 $D=\{\boldsymbol{x}\mid A\boldsymbol{x}=\boldsymbol{b},\ \boldsymbol{x}\geqslant 0\}$ 是一个凸集.

证明 由凸集的定义显然.

定理 2.3.3 $\bar{\boldsymbol{x}}$ 是可行解集 D 的极点（顶点）的充分必要条件是：$\bar{\boldsymbol{x}}$ 是基本可行解.

证明 "必要性". 设 $\bar{\boldsymbol{x}}$ 是可行解集 D 的极点，我们只需证明 $\bar{\boldsymbol{x}}$ 是基本解. 由定理 2.3.1 的证明过程可知，若 $\bar{\boldsymbol{x}}$ 不是基本可行解，则存在向量 $\boldsymbol{\delta}\neq 0$，使得 $\bar{\boldsymbol{x}}+\alpha\boldsymbol{\delta}$ 和 $\bar{\boldsymbol{x}}-\alpha\boldsymbol{\delta}$ 是可行解且

$$\bar{\boldsymbol{x}}=\frac{1}{2}[(\bar{\boldsymbol{x}}+\alpha\boldsymbol{\delta})+(\bar{\boldsymbol{x}}-\alpha\boldsymbol{\delta})]$$

与 $\bar{\boldsymbol{x}}$ 是极点矛盾.

"充分性". 设

$$\bar{\boldsymbol{x}}=\begin{bmatrix}\boldsymbol{x}_B\\ \boldsymbol{x}_N\end{bmatrix}=\begin{bmatrix}B^{-1}\boldsymbol{b}\\ 0\end{bmatrix},$$

假设存在点 $\boldsymbol{x}^{(1)}$、$\boldsymbol{x}^{(2)}$ 和 $\alpha\in(0,1)$，使得

$$\bar{\boldsymbol{x}} = \alpha \boldsymbol{x}^{(1)} + (1-\alpha)\boldsymbol{x}^{(2)}, \tag{2.3.18}$$

记

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} \boldsymbol{x}_B^{(1)} \\ \boldsymbol{x}_N^{(1)} \end{bmatrix}, \quad \boldsymbol{x}^{(2)} = \begin{bmatrix} \boldsymbol{x}_B^{(2)} \\ \boldsymbol{x}_N^{(2)} \end{bmatrix}, \tag{2.3.19}$$

则 (2.3.18) 式表示为

$$B^{-1}\boldsymbol{b} = \alpha \boldsymbol{x}_B^{(1)} + (1-\alpha)\boldsymbol{x}_B^{(2)}, \tag{2.3.20}$$

$$0 = \alpha \boldsymbol{x}_N^{(1)} + (1-\alpha)\boldsymbol{x}_N^{(2)}, \tag{2.3.21}$$

由于 α，$(1-\alpha) > 0$，$\boldsymbol{x}_N^{(1)}$，$\boldsymbol{x}_N^{(2)}$，$\geqslant 0$，由 (2.3.21) 式得到 $\boldsymbol{x}_N^{(1)} = 0$，$\boldsymbol{x}_N^{(2)} = 0$．又由于 $\boldsymbol{x}^{(1)}$、$\boldsymbol{x}^{(2)}$ 都是可行点，可由线性方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 得到

$$\boldsymbol{x}_B^{(1)} = B^{-1}\boldsymbol{b} - B^{-1}N\boldsymbol{x}_N^{(1)} = B^{-1}\boldsymbol{b}, \tag{2.3.22}$$

$$\boldsymbol{x}_B^{(2)} = B^{-1}\boldsymbol{b} - B^{-1}N\boldsymbol{x}_N^{(2)} = B^{-1}\boldsymbol{b}, \tag{2.3.23}$$

即 $\bar{\boldsymbol{x}} = \boldsymbol{x}^{(1)} = \boldsymbol{x}^{(2)}$ 是 D 的极点．

推论 2.3.1　如果线性规划问题的可行解集 D 非空，则 D 至少有一个顶点，至多有有限个顶点．

推论 2.3.2　如果线性规划问题有最优解，则最优解必在某个顶点处达到．

§ 2.4　单纯形方法

求解线性规划的单纯形方法（Simplex Method）是 G.B.Dantzig（丹捷格）在 1947 年首次提出来的．该方法不仅理论上完善，算法简单，而且适用于各种类型的线性规划问题，是一种求解线性规划问题的最常用方法．

由线性规划的基本性质可知，最优点一定在某个顶点处达到，因此，我们只需考虑顶点处的函数值的情况．

在介绍单纯形法之前，我们需作一些假定．

(1) $A = A_{m\times n}$，且 $m < n$，即行数少于列数，否则线性规

划问题的约束方程组 $A\boldsymbol{x}=\boldsymbol{b}$，只有唯一解或无解．

（2）rank（A）$=m$，即矩阵 A 是行满秩的，并且不考虑退化与循环情况．

2.4.1 引例

【例 2.4.1】 考虑线性规划问题

$$\begin{aligned}
\min\quad & -2x_1-5x_2,\\
\text{s.t.}\quad & x_1+2x_2\leqslant 8,\\
& x_1\leqslant 4,\\
& x_2\leqslant 3,\\
& x_1,x_2\geqslant 0.
\end{aligned}\tag{2.4.1}$$

引进松驰变量 x_3、x_4、x_5，将问题化为标准形式

$$\begin{aligned}
\min\quad & -2x_1+5x_2+0x_3+0x_4+0x_5,\\
\text{s.t.}\quad & x_1+2x_2+x_3=8,\\
& x_1+x_4=4,\\
& x_2+x_5=3,\\
& x_1,x_2,x_3,x_4,x_5\geqslant 0,
\end{aligned}\tag{2.4.2}$$

此形式称为线性规划的典式，这里约束方程的系数矩阵为

$$A=\begin{bmatrix}1&2&1&0&0\\1&0&0&1&0\\0&1&0&0&1\end{bmatrix}=[\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3,\boldsymbol{p}_4,\boldsymbol{p}_5],$$

$$\text{rank}(A)=3.$$

我们先求初始基本可行解．事实上，$B=[\boldsymbol{p}_3,\ \boldsymbol{p}_4,\ \boldsymbol{p}_5]=I$，$B^{-1}\boldsymbol{b}=(8,\ 4,\ 3)^T$ 是一个基本可行解．

由约束方程，我们得到

$$\begin{cases}x_3=8-x_1-2x_2,\\ x_4=4-x_1,\\ x_5=8-x_2,\end{cases}\tag{2.4.3}$$

将 (2.4.3) 式代入目标函数，得到

$$f = -2x_1 - 5x_2 + 0x_3 + 0x_4 + 0x_5 = -2x_1 - 5x_2, \tag{2.4.4}$$

取 $x_1=0$，$x_2=0$，得到 $x_3=8$，$x_4=4$，$x_5=3$，因此，基本可行解为 $\boldsymbol{x}=(0, 0, 8, 4, 3)^T$，其目标值为 $f=0$.

该点是否为最优点呢？我们看一下 (2.4.4) 式，当 x_1、x_2 增大时，目标函数值 f 随着 x_1、x_2 的增大而减少，所以它不是最优点. 因此，我们要增大 x_1、x_2 的值，哪一个值增大，使函数值下降更明显呢？由 (2.4.4) 式表明是 x_2，因此，我们令 $x_1=0$，$x_2=\theta$，代入 (2.4.3) 式，得到

$$\begin{cases} x_3 = 8 - 2\theta, \\ x_4 = 4, \\ x_5 = 3 - \theta, \end{cases} \tag{2.4.5}$$

由可行性，得到

$$3 - \theta \geqslant 0, 8 - 2\theta \geqslant 0, \tag{2.4.6}$$

即

$$\theta = \min\left\{\frac{3}{1}, \frac{8}{2}\right\} = 3. \tag{2.4.7}$$

将 $\theta=3$ 代入 (2.4.5) 式，得到新的点 $\boldsymbol{x}=(0, 3, 2, 4, 0)^T$. 注意，这是一个新的基本可行解. 此时，基为 $B=(\boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{p}_4)$，基变量为 $\boldsymbol{x}_B=(x_2, x_3, x_4)^T=(3, 2, 4)^T$，非基变量为 $\boldsymbol{x}_N=(x_1, x_5)^T=(0, 0)^T$. 那么，此点是否为最优点呢？我们还要考虑线性规划的典式.

我们仍然需要将基变量表示成非基变量的函数

$$\begin{cases} x_2 = 3 - x_5, \\ x_3 = 2 - x_1 + 2x_5, \\ x_4 = 4 - x_1, \end{cases} \tag{2.4.8}$$

目标函数也表示成非基变量的函数

$$f = -15 - 2x_1 + 5x_5, \tag{2.4.9}$$

因此，其典式为

$$\begin{aligned} \min \quad & -2x_1 + 0x_2 + 0x_3 + 0x_4 + 5x_5 - 15, \\ \text{s.t.} \quad & x_2 + x_5 = 3, \\ & x_1 + x_3 - 2x_5 = 2, \\ & x_1 + x_4 = 4, \\ & x_1, x_2, x_3, x_4, x_5 \geqslant 0. \end{aligned} \tag{2.4.10}$$

此时，只要增加 x_1 的值，目标函数值还可以下降，因此，$\boldsymbol{x} = (0,\ 3,\ 2,\ 4,\ 0)^T$ 不是最优解．类似地，令 $x_5 = 0$，$x_1 = \theta$，并由可行性得到

$$\theta = \min\left\{\frac{2}{1}, \frac{4}{1}\right\} = 2. \tag{2.4.11}$$

将 $x_1 = \theta = 2$ 代入（2.4.10）式，得到新的基本可行解 $\boldsymbol{x}^{(3)} = (2,\ 3,\ 0,\ 2,\ 0)^T$．它是否是最优解呢？我们再考虑该点的典式（由类似前面的方法推导得到）：

$$\begin{aligned} \min \quad & 0x_1 + 0x_2 + 2x_3 + 0x_4 + x_5 - 19, \\ \text{s.t.} \quad & x_1 + x_3 - 2x_5 = 2, \\ & x_2 + x_3 + x_5 = 3, \\ & -x_3 + x_4 = 2, \\ & x_1, x_2, x_3, x_4, x_5 \geqslant 0. \end{aligned} \tag{2.4.12}$$

注意到 x_3、x_5 的系数均为正，因此此点为最优解．

从上述例子可以看到，求解线性规划问题需要解决以下几个问题：

（1）求初始基本可行解．即找出可行基 B，使得 $B^{-1}\boldsymbol{b} \geqslant 0$，即典式中右端项大于等于 0．

（2）判别基本解是否为最优解时，要考查典式中目标函数的系数是否大于等于 0．

(3) 要利用典式从一个基本可行解，得到另一个基本可行解．

从上述过程可以看到，此方法太麻烦，也不易找出规律．因此我们设法找出一种简单、方便、可行而且便于计算机实践的方法，这就是单纯形法．

2.4.2 基本可行解的改进

考虑线性规划问题

$$\begin{aligned} \min \quad & f = \boldsymbol{c}^T\boldsymbol{x}, \\ \text{s.t.} \quad & A\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant 0. \end{aligned} \tag{2.4.13}$$

设 rank $(A) = m$，$A = [B, N]$，$B = B_{m\times m}$ 为非奇异矩阵，并且 $B^{-1}\boldsymbol{b} \geqslant 0$，即 B 是可行基，此时，基本可行解为

$$x = \begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{bmatrix} = \begin{bmatrix} B^{-1}\boldsymbol{b} \\ 0 \end{bmatrix}, \tag{2.4.14}$$

目标函数值为

$$f = \boldsymbol{c}^T\boldsymbol{x} = [\boldsymbol{c}_B^T, \boldsymbol{c}_N^T]\begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{bmatrix} = \boldsymbol{c}_B^T\boldsymbol{x}_B + \boldsymbol{c}_N^T\boldsymbol{x}_N = \boldsymbol{c}_B^T B^{-1}\boldsymbol{b}. \tag{2.4.15}$$

现在讨论如何从这个基本可行解出发，得到一个新的基本可行解，而且使目标函数值下降．

设 $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{bmatrix}$ 是任意的非退化的基本可行解，由约束条件得到

$$B\boldsymbol{x}_B + N\boldsymbol{x}_N = \boldsymbol{b}, \tag{2.4.6}$$

我们得到了基变量关于非基变量的表达式

$$\boldsymbol{x}_B = B^{-1}\boldsymbol{b} - B^{-1}N\boldsymbol{x}_N, \tag{2.4.17}$$

那么目标函数关于非基变量的表达式为

$$
\begin{aligned}
f = \boldsymbol{c}^T\boldsymbol{x} &= [\boldsymbol{c}_B^T, \boldsymbol{c}_N^T]\begin{bmatrix}\boldsymbol{x}_B \\ \boldsymbol{x}_N\end{bmatrix} = \boldsymbol{c}_B^T\boldsymbol{x}_B + \boldsymbol{c}_N^T\boldsymbol{x}_N \\
&= \boldsymbol{c}_B^T(B^{-1}\boldsymbol{b} - B^{-1}N\boldsymbol{x}_N) + \boldsymbol{c}_N^T\boldsymbol{x}_N \\
&= \boldsymbol{c}_B^T B^{-1}\boldsymbol{b} - (\boldsymbol{c}_B^T B^{-1}N - \boldsymbol{c}_N^T)\boldsymbol{x}_N,
\end{aligned} \tag{2.4.18}
$$

由此，我们得到了线性规划问题的典式

$$
\begin{aligned}
\min \quad & f = \boldsymbol{c}_B^T B^{-1}\boldsymbol{b} - (\boldsymbol{c}_B^T B^{-1}N - \boldsymbol{c}_N^T)\boldsymbol{x}_N, \\
\text{s.t.} \quad & \boldsymbol{x}_B + B^{-1}N\boldsymbol{x}_N = B^{-1}\boldsymbol{b}, \\
& \boldsymbol{x}_B, \boldsymbol{x}_N \geqslant 0.
\end{aligned} \tag{2.4.19}
$$

令

$$
\lambda_j = \boldsymbol{c}_B^T B^{-1}\boldsymbol{p}_j - c_j, j = m+1, m = 2, \cdots, n,
$$

$$
B^{-1}N = \begin{bmatrix} \alpha_{1m+1} & \alpha_{1m+2} & \cdots & \alpha_{1n} \\ \alpha_{2m+1} & \alpha_{2m+2} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{mm+1} & \alpha_{mm+2} & \cdots & \alpha_{mn} \end{bmatrix}, \tag{2.4.20}
$$

$$
B^{-1}\boldsymbol{b} = [\beta_1, \beta_2, \cdots, \beta_m]^T,
$$

则线性规划问题（2.4.19）可以写成

$$
\begin{aligned}
\min \quad & f = \boldsymbol{c}_B^T B^{-1}\boldsymbol{b} - \sum_{j=m+1}^{n} \lambda_j x_j, \\
\text{s.t.} \quad & x_i + \sum_{j=m+1}^{n} \alpha_{ij}x_j = \beta_i, \ i = 1, 2, \cdots, m, \\
& x_j \geqslant 0, \qquad j = 1, 2, \cdots, n.
\end{aligned} \tag{2.4.21}
$$

称 λ_i 为检验数．当 $\lambda_j > 0$ 时，x_j 增加，则目标函数值下降，由此我们可以得到改进的基本可行解．

为了便于计算，我们构造单纯形表．

线性规划问题（2.4.19）等价于线性规划问题

$$
\begin{aligned}
\min \quad & f, \\
\text{s.t.} \quad & f + 0^T x_B + (c_B^T B^{-1} N - c_N^T) x_N = c^T B^{-1} b, \\
& x_B + B^{-1} N x_N = B^{-1} b, \\
& x_B, x_N \geqslant 0.
\end{aligned}
\tag{2.4.22}
$$

我们将线性规划问题（2.4.22）约束条件的系数列成表格，得到单纯形表：

	f	x_B	x_N	RHS
f	1	0^T	$c_B^T B^{-1} N - c_N^T$	$c_B^T B^{-1} b$
x_B	0	I	$B^{-1} N$	$B^{-1} b$

(2.4.23)

表中 RHS 是指右端项（right-hand-side）. 上表的具体表达式为：

	f	x_1	$\cdots$	x_r	$\cdots$	x_m	$\cdots$	x_j	$\cdots$	x_k	$\cdots$	RHS
f	1	0	$\cdots$	0	$\cdots$	0	$\cdots$	λ_j	$\cdots$	λ_k	$\cdots$	f^0
x_1	0	1	$\cdots$	0	$\cdots$	0	$\cdots$	α_{1j}	$\cdots$	α_{1k}	$\cdots$	β_1
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_r	0	0	$\cdots$	1	$\cdots$	0	$\cdots$	α_{rj}	$\cdots$	α_{rk}	$\cdots$	β_r
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_m	0	0	$\cdots$	0	$\cdots$	1	$\cdots$	α_{mj}	$\cdots$	α_{mk}	$\cdots$	β_m

上表反映了线性规划问题在当前情况下的全部信息. 当 $x_N = (x_{m+1}, x_{m+2}, \cdots, x_n)^T = (0, 0, \cdots, 0)^T$ 时，$x_B = B^{-1} b = (\beta_1, \beta_2, \cdots, \beta_m)^T$，$f^0 = c_B^T B^{-1} b$. 其检验数为 $\lambda_j = c_B^T B^{-1} p_j - c_j$.

当 $\lambda_j > 0$ 时，x_j 增加，目标函数值下降. 为使目标函数下降的快，应找最大的 λ_j，因此令 $\lambda_k = \max\ \{\lambda_j \mid j = m+1,\ m+$

2，…，$n\}$，这样取 $x_k=\theta$，$x_j=0$，$j=m+1$，$m+2$，…，n，$j\neq k$. 由问题（2.4.21）中的约束条件和非负限制，得到

$$x_i=\beta_i-\alpha_{ik}\theta\geqslant 0,\quad i=1,2,\cdots,m. \qquad (2.4.25)$$

当 $\alpha_{ik}\leqslant 0$ 时，（2.4.25）式显然成立；而当 $\alpha_{ik}>0$ 时，只有当 $\theta\leqslant\beta_i/\alpha_{ik}$时，（2.4.25）式才成立. 因此得到

$$\theta=\min\left\{\frac{\beta_i}{\alpha_{ik}}\mid\alpha_{ik}>0\right\}=\frac{\beta_r}{\alpha_{rk}}, \qquad (2.4.26)$$

此时有

$$x_i=\beta_i-\alpha_{ik}\frac{\beta_r}{\alpha_{rk}}\geqslant 0,\ i=1,2,\cdots,m, \qquad (2.4.27)$$

特别当 $i=r$ 时，$x_r=0$，而非基变量中的 $x_k=\beta_r/\alpha_{rk}$. 这样 x_k 变为基变量，x_r 变为非基变量. 当线性规划问题是非退化的，则有 $f=f^0-\dfrac{\lambda_k\beta_r}{\alpha_{rk}}<f^0$，即目标函数值严格下降.

引理2.4.1 若 $B=(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_r,\cdots,\boldsymbol{p}_m)$ 是线性规划的一组基，则

$$\overline{B}=(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_{r-1},\boldsymbol{p}_k,\boldsymbol{p}_{r+1},\cdots,\boldsymbol{p}_m)$$

也是线性规划的一组基.

证明 因为 $B^{-1}B=I$，所以

$$B^{-1}\overline{B}=\begin{bmatrix}1 & \cdots & \alpha_{1k} & \cdots & 0\\ \vdots & & \vdots & & \vdots\\ 0 & \cdots & \alpha_{rk} & \cdots & 0\\ \vdots & & \vdots & & \vdots\\ 0 & \cdots & \alpha_{mk} & \cdots & 1\end{bmatrix}, \qquad (2.4.28)$$

这里 α_{ik}是$B^{-1}\boldsymbol{p}_k$ 的第 i 个分量. 令

$$E = \begin{bmatrix} 1 & \cdots & -\alpha_{1k}/\alpha_{rk} & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & -\alpha_{r-1k}/\alpha_{rk} & \cdots & 0 \\ 0 & \cdots & 1/\alpha_{rk} & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & -\alpha_{r+1k}/\alpha_{rk} & \cdots & 0 \\ 0 & \cdots & -\alpha_{mk}/\alpha_{rk} & \cdots & 1 \end{bmatrix}, \qquad (2.4.29)$$

因此，$EB^{-1}\overline{B} = I$，即$\overline{B}^{-1} = EB^{-1}$. $\overline{B}$ 是线性规划问题的一组基. 由此我们得到下述定理.

定理 2.4.1 按上述方法得到的点是新的改进的基本可行解.

下面的问题是如何得到新的基本可行解对应的单纯形表. 由引理 2.4.1 的证明可知$\overline{B}^{-1} = EB^{-1}$，即左乘一个初等矩阵. 这等价于在单纯形表上作**转轴运算** (pivoting)，即作初等变换，将主元素 α_{rk}化为 1，第 k 列的其它元素化为 0. 这样我们就得到新的单纯形表：

	f	x_1	$\cdots$	x_r	$\cdots$	x_m	$\cdots$	x_j	$\cdots$	x_k	$\cdots$	RHS
f	1	0	$\cdots$	$-\frac{\lambda_k}{\alpha_{rk}}$	$\cdots$	0	$\cdots$	$\lambda_j - \frac{\lambda_k}{\alpha_{rk}}\alpha_{rj}$	$\cdots$	0	$\cdots$	$f^0 - \frac{\lambda_k}{\alpha_{rk}}\beta_r$
x_1	0	1	$\cdots$	$-\frac{\alpha_{1k}}{\alpha_{rk}}$	$\cdots$	0	$\cdots$	$\alpha_{1j} - \frac{\alpha_{1k}}{\alpha_{rk}}\alpha_{rj}$	$\cdots$	0	$\cdots$	$\beta_1 - \frac{\alpha_{1k}}{\alpha_{rk}}\beta_r$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_{r-1}	0	0	$\cdots$	$-\frac{\alpha_r - 1k}{\alpha_{rk}}$	$\cdots$	0	$\cdots$	$\alpha_{r-1j} - \frac{\alpha_{r-1k}}{\alpha_{rk}}\alpha_{rj}$	$\cdots$	0	$\cdots$	$\beta_{r-1} - \frac{\alpha_{r-1k}}{\alpha_{rk}}\beta_r$
x_r	0	0	$\cdots$	$\frac{1}{\alpha_{rk}}$	$\cdots$	0	$\cdots$	$\frac{\alpha_{rj}}{\alpha_{rk}}$	$\cdots$	1	$\cdots$	$\frac{\beta_r}{\alpha_{rk}}$
x_{r+1}	0	0	$\cdots$	$-\frac{\alpha_r - 1k}{\alpha_{rk}}$	$\cdots$	0	$\cdots$	$\alpha_{r+1j} - \frac{\alpha_{r+1k}}{\alpha_{rk}}\alpha_{rj}$	$\cdots$	0	$\cdots$	$\beta_{r+1} - \frac{\alpha_{r+1k}}{\alpha_{rk}}\beta_r$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_m	0	0	$\cdots$	$-\frac{\alpha_{mk}}{\alpha_{rk}}$	$\cdots$	1	$\cdots$	$\alpha_{mj} - \frac{\alpha_{mk}}{\alpha_{rk}}\alpha_{rj}$	$\cdots$	0	$\cdots$	$\beta_m - \frac{\alpha_{mk}}{\alpha_{rk}}\beta_r$

(2.4.30)

2.4.3 最优性条件与无界情况

设 B 是基本可行基，即 $B^{-1}\boldsymbol{b}\geqslant 0$. 我们可将线性规划问题 (2.2.4) 化成 (2.4.21)，即

$$\min \quad f = \boldsymbol{c}^T B^{-1}\boldsymbol{b} - \sum_{j=m+1}^{n}\lambda_j x_j,$$

$$\text{s.t.} \quad x_i + \sum_{j=m+1}^{n}\alpha_{ij}x_j = \beta_i,\ i = 1,2,\cdots,m,$$

$$x_j \geqslant 0,\ j = 1,2,\cdots,n.$$

当 $\lambda_j\leqslant 0$, $j=m+1$, $m+2$, $\cdots$, n 时，当 x_j ($j=m+1$, $m+2$, $\cdots$, n) 增大，目标函数 f 也随着增大. 因此，此基本可行解是最优解. 我们有如下定理：

定理 2.4.2 设 B 为线性规划的一个可行基，若对所有的检验数

$$\lambda_j = \boldsymbol{c}_B^T B^{-1}\boldsymbol{p}_j - c_j \leqslant 0,\ j\in R, \tag{2.4.31}$$

(其中 R 是非基变量指标集) 则此基本可行解为线性规划问题最优解.

【例 2.4.1】 用单纯形法求解线性规划问题：

$$\min \quad -2x_1 - 5x_2 + 0x_3 + 0x_4 + 0x_5,$$

$$\text{s.t.} \quad x_1 + 2x_2 + x_3 = 8$$

$$x_1 + x_4 = 4,$$

$$x_2 + x_5 = 3,$$

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

解 列初始单纯形表：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	2	5^*	0	0	0	0
x_3	0	1	2	1	0	0	8
x_4	0	1	0	0	1	0	4
x_5	0	0	$\boxed{1}$	0	0	1	3

注意到 $\lambda_2 = \max\{\lambda_j \mid j \in R\} = \max\{2, 5\} = 5$，$\theta = \max\{\frac{8}{2}, \frac{3}{1}\} = \frac{3}{1}$. 因此以 α_{32} 为中心作转轴运算，x_2 进基，x_5 出基，得到第二张单纯形表：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	2^*	0	0	0	-5	-15
x_3	0	$\boxed{1}$	0	1	0	-2	2
x_4	0	1	0	0	1	0	4
x_2	0	0	1	0	0	1	3

类似地得到第三张单纯形表：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	0	0	-2	0	-1	-19
x_1	0	1	0	1	0	-2	2
x_4	0	0	0	-1	1	2	2
x_2	0	0	1	0	0	1	3

此时，$\lambda_3=-2<0$，$\lambda_5=-1<0$，得到最优解 $x^*=(2,3,0,2,0)^T$，$f^*=-19$.

【例 2.4.2】 用单纯形法求解线性规划问题：

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

解 列初始单纯形表：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	1	2^*	0	0	0	0
x_3	0	1	2	1	0	0	8
x_4	0	1	0	0	1	0	4
x_5	0	0	$\boxed{1}$	0	0	1	3

第二张单纯形表：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	1^*	0	0	0	-2	-6
x_3	0	$\boxed{1}$	0	1	0	-2	2
x_4	0	1	0	0	1	0	4
x_2	0	0	1	0	0	1	3

第三张单纯形表：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	0	0	-1	0	0^*	-8
x_1	0	1	0	1	0	-2	2
x_4	0	0	0	-1	1	$\boxed{2}$	2
x_2	0	0	1	0	0	1	3

此时，$\lambda_3=-1<0$，$\lambda_5=0$，得到最优解 $x^*=(2,3,0,2,0)^T$，$f^*=-8$. 注意到 $\lambda_5=0$，我们可以继续作转轴运算，得到第四张单纯表：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	0	0	-1	0	0	-8
x_1	0	1	0	0	1	0	4
x_5	0	0	0	$-\frac{1}{2}$	$\frac{1}{2}$	1	1
x_2	0	0	1	$\frac{1}{2}$	$-\frac{1}{2}$	0	2

得到另一个最优解 $x^*=(4,2,0,0,1)^T$，而最优目标值仍为 $f^*=-8$.

从上述两个例子可以看出，当 $\lambda_j<0$，$\forall j\in R$ 时，线性规划问题的最优解是唯一的. 若 $\exists j\in R$，使得 $\lambda_j=0$，线性规划问题有无穷多个最优解.

下面我们讨论无界情况.

当 $\lambda_k>0$ 时，若 $\alpha_{ik}\leqslant 0$，$i=1,2,\cdots,m$，则 $\forall\,\theta\geqslant 0$，有

$$x_i=\beta_i-\alpha_{ik}\theta\geqslant 0, i=1,2,\cdots,m,$$

此时，$f=f^0-\lambda_k\theta\to-\infty$，当 $\theta\to+\infty$，由此得到：

定理 2.4.3　设 B 为线性规划的一个可行基，若 $\exists k \in R$，使得 $\lambda_k > 0$ 且 $\alpha_{ik} \leqslant 0$，$i = 1, 2, \cdots, m$，则此线性规划问题无有限最优解（称为无最优解）.

【例 2.4.3】 用单纯形法求解线性规划问题：

$$\begin{aligned} \min \quad & -x_1 - 3x_2, \\ \text{s.t.} \quad & x_1 - 2x_2 \leqslant 4, \\ & -x_1 + x_2 \leqslant 3, \\ & x_1, x_2 \geqslant 0. \end{aligned}$$

解　引进松弛变量 x_3、x_4，将问题化为标准形式

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

列初始单纯形表：

	f	x_1	x_2	x_3	x_4	RHS
f	1	1	3^*	0	0	0
x_3	0	1	-2	1	0	4
x_4	0	-1	$\boxed{1}$	0	1	3

第二张单纯形表：

	f	x_1	x_2	x_3	x_4	RHS
f	1	4^*	0	0	-3	-9
x_3	0	-1	0	1	2	10
x_2	0	-1	1	0	1	3

此时，$\lambda_1=4>0$，而 $\alpha_{11}=-1<0$，$\alpha_{21}=-1<0$，此线性规划问题无有限最优解.

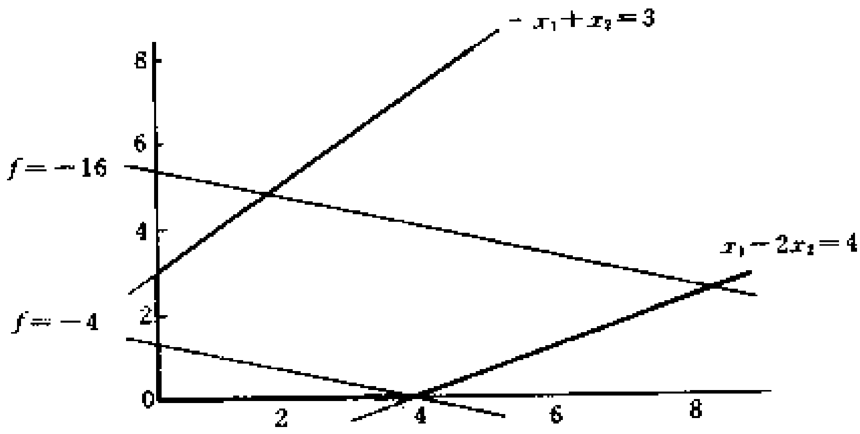


图 2.4.1　例 2.4.3 的几何意义

在 $\alpha_{ik}\leqslant 0$，$i=1,2,\cdots,m$ 的条件下，我们构造出可行集 $D=\{\boldsymbol{x}\mid A\boldsymbol{x}=\boldsymbol{b},\ \boldsymbol{x}\geqslant 0\}$ 的一个极方向：

$$\boldsymbol{d}=\begin{bmatrix}-\overline{\boldsymbol{p}}_k\\ \boldsymbol{e}_k\end{bmatrix}, \tag{2.4.32}$$

这里 $\overline{\boldsymbol{p}}_k=\begin{bmatrix}\alpha_{1k}\\ \alpha_{2k}\\ \vdots\\ \alpha_{mk}\end{bmatrix}=B^{-1}p_k$，$\boldsymbol{e}_k\in R^{n-m}$ 的单位向量.

事实上，由（1.4.8）式知，$\boldsymbol{d}$ 是 D 的一个方向. 下面证明它是一个极方向. 若存在两个方向 $\boldsymbol{d}^{(1)}$、$\boldsymbol{d}^{(2)}$，使得

$$\boldsymbol{d}=\alpha_1\boldsymbol{d}^{(1)}+\alpha_2\boldsymbol{d}^{(2)},\ \alpha_1,\alpha_2>0,$$

则得到

$$\boldsymbol{e}_k=\alpha_1\boldsymbol{d}_N^{(1)}+\alpha_2\boldsymbol{d}_N^{(2)},$$

所以存在着 α，使得

$$\boldsymbol{d}_N^{(1)}=\alpha\boldsymbol{d}_N^{(2)}, \tag{2.4.33}$$

由于 $\boldsymbol{d}_N^{(1)}$、$\boldsymbol{d}_N^{(2)}$ 是 D 的方向，因此满足

$$d_B^{(i)} = B^{-1}Nd_N^{(i)},\ i = 1,2, \tag{2.4.34}$$

结合 (2.4.33) 式和 (2.4.34) 式得到 $\boldsymbol{d}^{(1)} = \alpha \boldsymbol{d}^{(2)}$，即 $\boldsymbol{d}$ 是一个极方向.

进一步，我们可以得到：d 满足

$$\boldsymbol{c}^T\boldsymbol{d} = \boldsymbol{c}_B^T\boldsymbol{d}_B + \boldsymbol{c}_N^T\boldsymbol{d}_N = -\boldsymbol{c}_B^TB^{-1}\boldsymbol{p}_k + c_k < 0, \tag{2.4.35}$$

即沿该极方向前进，目标函数值下降. 这一点反映了无有限最优解的几何意义.

在例 2.4.3 中，它的一个极方向为

$$\boldsymbol{d} = (1,1,1,0)^T.$$

2.4.4 算法

算法 2.4.1 (从基本可行解开始的单纯形法)

(1) 选取初始基本可行基 B，列出初始单纯形表：

	f	x_B	x_N	RHS
f	1	0^T	$c_B^TB^{-1}N - c_N^T$	$c_B^TB^{-1}b$
x_B	0	I	$B^{-1}N$	$B^{-1}b$

(2) 令 $\lambda_k = \max\{\lambda_j \mid j \in R\}$ (R 为非基变量指标集). 若 $\lambda_k \leqslant 0$，则停止计算 (此时得到最优解).

(3) 若 $\alpha_{ik} \leqslant 0$，$i = 1, 2, \cdots, m$，则停止计算 (此时线性规划问题无有限最优解)；否则确定 θ，

$$\theta = \min\left\{\frac{\beta_i}{\alpha_{ik}} \mid \alpha_{ik} > 0\right\} = \frac{\beta_r}{\alpha_{rk}},$$

然后以 α_{rk} 为中心进行转轴运算，即置

$$\alpha_{ij} = \alpha_{rj}/\alpha_{rk},\quad j = 1,2,\cdots,n,\ j \neq k,$$

$$\beta_r = \beta_r/\alpha_{rk},\quad \alpha_{rk} = 1,$$

$$\lambda_j = \lambda_j - \lambda_k \alpha_{rj}, \quad j = 1,2,\cdots,n,$$
$$\alpha_{ij} = \alpha_{ij} - \alpha_{ik}\alpha_{rj}, i = 1,2,\cdots,m, \ i \neq r, \quad j = 1,2,\cdots,n,$$
$$\beta_i = \beta_j - \alpha_{ik}\beta_r, \quad i = 1,2,\cdots,m, i \neq r,$$
$$f = f - \lambda_k \beta_r,$$
然后转（2）.

2.4.5 从任意点开始的单纯形法——求解线性规划的两阶段方法

在前面介始的单纯形法中，需要从初始基本可行基开始，即要求 $B^{-1}b \geqslant 0$；否则算法无法实现．如何求出一个基本可行基呢？这对于一般线性规划问题是困难的．在这里我们介始一种求基本可行基的方法——**人工变量**（artificial variable）方法．

考虑线性规划问题

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

并假设 $b_i \geqslant 0$，$i = 1, 2, \cdots, m$（若某个 $b_{i_0} < 0$，对第 i_0 行乘上 -1）．写成矩阵形式

$$\begin{aligned}
\min \quad & f = \boldsymbol{c}^T\boldsymbol{x}, \\
\text{s.t.} \quad & A\boldsymbol{x} = \boldsymbol{b}, \\
& \boldsymbol{x} \geqslant 0,
\end{aligned} \tag{2.4.36}$$

其中 $\boldsymbol{b} \geqslant 0$.

现在寻找基本可行基．从前面的例子可以看出，当右端项大于等于 0 时，只要方程组系数矩阵中存在一个单位矩阵，也就是说存在典式，就得到了基本可行解．

引进人工变量 x_{n+1}, x_{n+2}, $\cdots$, x_{n+m}，考虑辅助线性规划问题

$$
\begin{aligned}
\min \quad & x_0 = x_{n+1} + x_{n+2} + \cdots + x_{n+m}, \\
\text{s.t.} \quad & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + x_{n+1} = b_1, \\
& a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + x_{n+2} = b_2, \\
& \qquad\qquad \vdots \\
& a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + x_{n+m} = b_m, \\
& x_1, x_2, \cdots, x_n, x_{n+1}, x_{n+2}, \cdots, x_{n+m} \geqslant 0.
\end{aligned} \tag{2.4.37}
$$

写成矩阵形式

$$
\begin{aligned}
\min \quad & x_0 = \boldsymbol{e}^T \boldsymbol{x}_a \\
\text{s.t.} \quad & A\boldsymbol{x} + \boldsymbol{x}_a = \boldsymbol{b}, \\
& \boldsymbol{x}, \boldsymbol{x}_a \geqslant 0,
\end{aligned} \tag{2.4.38}
$$

其中 $\boldsymbol{x}_a = (x_{n+1}, x_{n+2}, \cdots, x_{n+m})^T$，$\boldsymbol{e} = (1, 1, \cdots, 1)^T \in R^m$.

显然，当 $\boldsymbol{x}=0$，$\boldsymbol{x}_a=\boldsymbol{b}$ 是辅助线性规划问题（2.4.38）的基本可行解，即 $(x_1, x_2, \cdots, x_n, x_{n+1}, x_{n+2}, \cdots, x_{n+m})^T = (0, 0, \cdots, 0, b_1, b_2, \cdots, b_m)^T$ 是辅助线性规划问题（2.4.37）的基本可行解.

事实上，辅助线性规划问题的系数矩阵为

$$
[A, I] = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_n, \boldsymbol{p}_{n+1}, \boldsymbol{p}_{n+2}, \cdots, \boldsymbol{p}_{n+m}]
= \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\
a_{21} & a_{22} & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 1
\end{bmatrix}.
$$

由于 $\boldsymbol{b} \geqslant 0$，基本可行基为 $B=I$，故辅助线性规划问题本身就是典式.

下面我们对辅助线性规划问题的求解情况进行讨论.

辅助线性规划问题（2.4.38）等价于

$$\begin{aligned}
\min \quad & x_0, \\
\text{s.t.} \quad & x_0 - \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}_a = 0, \\
& A\boldsymbol{x} + \boldsymbol{x}_a = \boldsymbol{b}, \\
& \boldsymbol{x}, \boldsymbol{x}_a \geqslant 0,
\end{aligned}$$

得到单纯形表：

	x_0	$\boldsymbol{x}$	$\boldsymbol{x}_a$	RHS
x_0	1	$\boldsymbol{0}^T$	$-\boldsymbol{e}^T$	0
$\boldsymbol{x}_a$	0	A	I	$\boldsymbol{b}$

(2.4.39)

我们称为原始单纯形表，这张表还不能利用．注意到基变量对应的检验数应为 0，因此我们还需将单纯形表（2.4.39）的第一行基变量的检验数化为 0，经初等变换，就得到初始单纯形表：

	x_0	$\boldsymbol{x}$	$\boldsymbol{x}_a$	RHS
x_0	1	$\boldsymbol{e}^TA$	$\boldsymbol{0}^T$	$\boldsymbol{e}^T\boldsymbol{b}$
$\boldsymbol{x}_a$	0	A	I	$\boldsymbol{b}$

由于辅助线性规划问题有基本可行解，且目标函数值下界为 0，因此辅助线性规划问题一定有最优基本可行解．现针对辅助线性规划的最优解的情况进行讨论．

（1）若辅助线性规划问题的最优目标函数值大于 0，则原线性规划问题无可行解．

事实上，若线性规划问题（2.4.36）有可行解 $\bar{\boldsymbol{x}}$，则令 $\boldsymbol{x}=\bar{\boldsymbol{x}}$，$\boldsymbol{x}_a=0$ 为辅助线性规划问题（2.3.38）的可行解，且目标函数值为 0，这与最优目标函数值大于 0 矛盾．

【例 2.4.4】 考虑线性规划问题：

$$\begin{aligned} \min \quad & -3x_1+4x_2, \\ \text{s.t.} \quad & x_1+x_2 \leqslant 4, \\ & 2x_1+3x_2 \geqslant 18, \\ & x_1,x_2 \geqslant 0. \end{aligned}$$

解 引进松弛变量 x_3、x_4，化为标准形式

$$\begin{aligned} \min \quad & -3x_1+4x_2, \\ \text{s.t.} \quad & x_1+x_2+x_3=4, \\ & 2x_1+3x_2-x_4=18, \\ & x_1,x_2,x_3,x_4 \geqslant 0. \end{aligned}$$

引进人工变量 x_5（因 x_3 对应的列已是单位向量，所以不需要引进人工变量 x_6），得到的辅助线性规划为

$$\begin{aligned} \min \quad & x_5, \\ \text{s.t.} \quad & x_1+x_2+x_3=4, \\ & 2x_1+3x_2-x_4+x_5=18, \\ & x_1,x_2,x_3,x_4,x_5 \geqslant 0. \end{aligned}$$

其原始单形表为：

	x_0	x_1	x_2	x_3	x_4	x_5	RHS
x_0	1	0	0	0	0	-1	0
x_3	0	1	1	1	0	0	4
x_5	0	2	3	0	-1	1	18

将基变量的检验数化为 0，作初等变换得到初始单纯形表：

	x_0	x_1	x_2	x_3	x_4	x_5	RHS
x_0	1	2	3^*	0	−1	0	18
x_3	0	1	$\boxed{1}$	1	0	0	4
x_5	0	2	3	0	−1	1	18

第二张单纯形表：

	x_0	x_1	x_2	x_3	x_4	x_5	RHS
x_0	1	−1	0	−3	−1	0	6
x_2	0	1	1	1	0	0	4
x_5	0	−1	0	−3	−1	1	6

此时检验数均小于等于 0，得到辅助线性规划问题的最优解 $\boldsymbol{x}^* = (0, 4, 0, 0, 6)^T$，最优目标函数值 $x_0^* = 6 > 0$. 因此原线性规划问题无可行解.

(2) 若辅助线性规划问题的最优目标函数值等于 0，则分两种情况讨论：

(a) 若辅助线性规划问题的最优基中不含有人工变量，则得到原线性规划问题的基本可行解.

不妨设辅助线性规划问题的最优单纯形表为：

	x_0	x_B	x_N	x_a	RHS
x_0	1	0^T	0^T	$-e^T$	0
x_B	0	I	$B^{-1}N$	B^{-1}	$B^{-1}b$

(2.4.40)

此时去掉人工变量所对应的列得到原线性规划的原始单纯形表：

	x_0	x_B	x_N	RHS
f	1	$-c_B^T$	$-c_N^T$	0
x_B	0	I	$B^{-1}N$	$B^{-1}b$

(2.4.41)

再作初等变换，将原始单纯形表中基变量对应的检验数化为 0，就得到原线性规划问题的初始单纯形表：

	f	x_B	x_N	RHS
f	1	0^T	$c_B^TB^{-1}N-c_N^T$	$c_B^TB^{-1}b$
x_B	0	I	$B^{-1}N$	$B^{-1}b$

(2.4.42)

这样就得到原线性规划问题的一个基本可行解．然后从这个基本可行解出发继续求解．

【例 2.4.5】 求解线性规划问题：

$$
\begin{aligned}
\min\quad & x_1-2x_2,\\
\text{s.t.}\quad & x_1+x_2\geqslant 2,\\
& -x_1+x_2\geqslant 1,\\
& x_2\leqslant 3,\\
& x_1,x_2\geqslant 0.
\end{aligned}
$$

解 引进松弛变量 x_3、x_4、x_5，化为标准形式

$$
\begin{aligned}
\min\quad & x_1-2x_2,\\
\text{s.t.}\quad & x_1+x_2-x_3=2,\\
& -x_1+x_2-x_4=1,\\
& x_2+x_5=3,
\end{aligned}
$$

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

引进人工变量 x_6、x_7，得到的辅助线性规划问题

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

其原始单形表为：

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
x_0	1	0	0	0	0	0	−1	−1	0
x_6	0	1	1	−1	0	0	1	0	2
x_7	0	−1	1	0	−1	0	0	1	1
x_5	0	0	1	0	0	1	0	0	3

作初等变换将基变量对应的检验数化为 0，得到初始单纯形表：

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
x_0	1	0	2*	−1	−1	0	0	0	3
x_6	0	1	1	−1	0	0	1	0	2
x_7	0	−1	$\boxed{1}$	0	−1	0	0	1	1
x_5	0	0	1	0	0	1	0	0	3

经计算得到

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
x_0	1	2^*	0	-1	1	0	0	-2	1
x_6	0	$\boxed{2}$	0	-1	1	0	1	-1	1
x_2	0	-1	1	0	-1	0	0	1	1
x_5	0	1	0	0	1	1	0	-1	2

和

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
x_0	1	0	0	0	0	0	-1	-1	0
x_1	0	1	0	$-\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
x_2	0	0	1	$-\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$
x_5	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$

得到辅助线性规划的最优解，且最优目标函数值为 0.

第二步我们来求原线性规划问题的解. 由于 x_6、x_7 是人工变量且不在基变量中，去掉 x_6、x_7 相对应的列，构造出原线性规划问题的原始单纯形表：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	-1	2	0	0	0	0
x_1	0	1	0	$-\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
x_2	0	0	1	$-\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{3}{2}$
x_5	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{3}{2}$

作初等变换，将基变量对应的检验数化为0，得到初始单纯形表：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	0	0	$\frac{1}{2}$	$\frac{3}{2}^*$	0	$-\frac{5}{2}$
x_1	0	1	0	$-\frac{1}{2}$	$\boxed{\frac{1}{2}}$	0	$\frac{1}{2}$
x_2	0	0	1	$-\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{3}{2}$
x_5	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{3}{2}$

经计算得到

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	-3	0	2^*	0	0	-4
x_4	0	2	0	-1	-1	0	1
x_2	0	1	1	-1	0	0	2
x_5	0	-1	0	$\boxed{1}$	0	1	1

和

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	-1	0	0	0	-2	-6
x_4	0	1	0	0	1	1	2
x_2	0	0	1	0	0	1	3
x_3	0	-1	0	1	0	1	2

最后得到原线性规划的最优解 $\boldsymbol{x}^* = (0,\ 3,\ 2,\ 2,\ 0)^T$，最优

目标函数值为 $f^* = -6$.

(b) 若辅助线性规划问题的最优基中含有人工变量（不妨设为 x_{n+r}），我们分两种情况考虑：

(i) 若 $a_{rj}=0$，$\forall j \in R$（非基变量非人工变量指标集），则第 r 个方程是多余方程.

设 J 为人工变量中非基变量指标集，因此有

$$x_{n+r} = \beta_r - \sum_{j \in R} \alpha_{rj} x_j - \sum_{j \in J} \alpha_{rn+j} x_{n+j} \qquad (2.4.43)$$

注意到 $\beta_r = 0$（因为此时目标函数值为 0）和 $\alpha_{rj}=0$，$\forall j \in R$，则（2.4.43）式变为

$$x_{n+r} = -\sum_{j \in J} \alpha_{rn+j} x_{n+j}. \qquad (2.4.44)$$

另一方面，$\boldsymbol{x}$、$\boldsymbol{x}_a$ 是辅助线性规划问题的可行解，得到

$$x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij} x_j, \qquad (2.4.45)$$

特别当 $i=r$ 时，有

$$x_{n+r} = b_r - \sum_{j=1}^{n} a_{ij} x_j, \qquad (2.4.46)$$

将（2.4.46）式和（2.4.45）式代入（2.4.44）式两端，得到

$$b_r - \sum_{j=1}^{n} a_{rj} x_j = -\sum_{j \in I} a_{ij} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right), \qquad (2.4.47)$$

即第 r 行可以由其它行线性表出，第 r 行方程是多余的.

【例 2.4.6】 求解线性规划问题：

$$\begin{aligned}
\min \quad & -x_1 + 2x_2 - 3x_3, \\
\text{s.t.} \quad & x_1 + x_2 + x_3 = 6, \\
& -x_1 + x_2 + 2x_3 = 4, \\
& 2x_2 + 3x_3 = 10, \\
& x_3 \leqslant 2, \\
& x_1, x_2, x_3 \geqslant 0.
\end{aligned}$$

解 引进松弛变量 x_4，化为标准形式：

$$\begin{aligned}
\min\quad & -x_1+2x_2-3x_3,\\
\text{s.t.}\quad & x_1+x_2+x_3=6,\\
& -x_1+x_2+2x_3=4,\\
& 2x_2+3x_3=10,\\
& x_3+x_4=2,\\
& x_1,x_2,x_3,x_4\geqslant 0.
\end{aligned}$$

引进人工变量 x_5、x_6、x_7，得到的辅助线性规划问题

$$\begin{aligned}
\min\quad & x_5+x_6+x_7,\\
\text{s.t.}\quad & x_1+x_2+x_3+x_5=6,\\
& -x_1+x_2+2x_3+x_6=4.\\
& 2x_2+3x_3+x_7=10,\\
& x_3+x_4=2,\\
& x_1,x_2,\cdots,x_7\geqslant 0,
\end{aligned}$$

其原始单形表为：

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
x_0	1	0	0	0	0	-1	-1	-1	0
x_5	0	1	1	1	0	1	0	0	6
x_6	0	-1	1	2	0	0	1	0	4
x_7	0	0	2	3	0	0	0	1	10
x_4	0	0	0	1	1	0	0	0	2

作初等变换，将基变量对应的检验数化为 0，得到初始单纯形表：

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
x_0	1	0	4	6*	0	0	0	0	20
x_5	0	1	1	1	0	1	0	0	6
x_6	0	−1	1	2	0	0	1	0	4
x_7	0	0	2	3	0	0	0	1	10
x_4	0	0	0	$\boxed{1}$	1	0	0	0	2

经计算得到：

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
x_0	1	0	4*	0	−6	0	0	0	8
x_5	0	1	1	0	−1	1	0	0	4
x_6	0	−1	$\boxed{1}$	0	−2	0	1	0	0
x_7	0	0	2	0	−3	0	0	1	4
x_3	0	0	0	1	1	0	0	0	2

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
x_0	1	4*	0	0	2	0	−4	0	8
x_5	0	$\boxed{2}$	0	0	1	1	−1	0	4
x_6	0	−1	1	0	−2	0	1	0	0
x_7	0	2	0	0	1	0	−2	1	4
x_3	0	0	0	1	1	0	0	0	2

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
x_0	1	0	0	0	0	−2	−2	0	0
x_1	0	1	0	0	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	0	2
x_2	0	0	1	0	$-\frac{3}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	2
x_7	0	0	0	0	0	−1	−1	1	0
x_3	0	0	0	1	1	0	0	0	2

得到辅助线性规划的最优解 $x=(2,2,2,0,0,0,0)^T$，最优目标函数值为0. 但此时基变量中含有人工变量 x_7. 考虑元素 α_{ij}，$j\in R$，注意到 $\alpha_{34}=0$（x_5、x_6 是人工变量不考虑），因此第3个方程是多余方程. 去掉多余方程，再去掉相应的人工变量，进入第二阶段.

	f	x_1	x_2	x_3	x_4	RHS
f	1	1	-2	3	0	0
x_1	0	1	0	0	$\frac{1}{2}$	2
x_2	0	0	1	0	$-\frac{3}{2}$	2
x_3	0	0	0	1	1	2

经初等变换得到初始单纯形表：

	f	x_1	x_2	x_3	x_4	RHS
f	1	0	0	0	$-\frac{3}{2}$	-4
x_1	0	1	0	0	$\frac{1}{2}$	2
x_2	0	0	1	0	$-\frac{3}{2}$	2
x_3	0	0	0	1	1	2

得到最优解 $x^*=(2,2,2)^T$，$f^*=-4$.

事实上，由线性规划的约束条件可直接看出第三个约束是多余的，是由前二个约束相加得到的.

(ⅱ) 若存在 $k\in R$，使得 $\alpha_{rk}\neq0$，则以 α_{rk} 为中心作转轴运算. 然后重复上述工作.

这样做，x_k 进基，x_{n+r} 出基，由于 $\beta_r=0$，因此目标函数

值不会增加，但基中的人工变量减少一个.

【例 2.4.7】 求解线性规划问题：

$$
\begin{aligned}
\min \quad & x_1 + 2x_2 + 3x_3, \\
\text{s.t.} \quad & x_1 - x_2 - x_3 \geqslant 4, \\
& x_2 - x_3 \geqslant 2, \\
& x_1 + x_2 + 3x_3 \leqslant 8, \\
& x_1, x_2, x_3 \geqslant 0.
\end{aligned}
$$

解 引进松弛变量 x_4、x_5、x_6，化为标准形式

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

引进人工变量 x_7、x_8，得到的辅助线性规划问题

$$
\begin{aligned}
\min \quad & x_7 + x_8, \\
\text{s.t.} \quad & x_1 - x_2 - x_3 - x_4 + x_7 = 4, \\
& x_2 - x_3 - x_5 + x_8 = 2, \\
& x_1 + x_2 + 3x_3 + x_6 = 8, \\
& x_1, x_2, \cdots, x_8 \geqslant 0.
\end{aligned}
$$

经计算得到如下单纯形表：

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	RHS
x_0	1	1^*	0	−2	−1	−1	0	0	0	6
x_7	0	$\boxed{1}$	−1	−1	−1	0	0	1	0	4
x_8	0	0	1	−1	0	−1	0	0	1	2
x_6	0	1	1	3	0	0	1	0	0	8

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	RHS
x_0	1	0	1^*	-1	0	-1	0	-1	0	2
x_1	0	1	-1	-1	-1	0	0	1	0	4
x_8	0	0	1	-1	0	-1	0	0	1	2
x_6	0	0	$\boxed{2}$	4	1	0	1	-1	0	4

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	RHS
x_0	1	0	0	-3	$-\frac{1}{2}$	-1	$-\frac{1}{2}$	$-\frac{1}{2}$	0	0
x_1	0	1	0	1	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	6
x_8	0	0	0	-3	$-\frac{1}{2}$	$\boxed{-1}$	$-\frac{1}{2}$	$\frac{1}{2}$	1	0
x_2	0	0	1	2	$\frac{1}{2}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	0	2

注意到此时 $\alpha_{23}\neq 0$，$\alpha_{24}\neq 0$，$\alpha_{25}\neq 0$ 和 $\alpha_{26}\neq 0$，以它们之一（不妨以 α_{25}为中心）为中心作转轴运算，得到

	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	RHS
x_0	1	0	0	0	0	0	0	-1	-1	0
x_1	0	1	0	1	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	6
x_5	0	0	0	3	$\frac{1}{2}$	1	$\frac{1}{2}$	$-\frac{1}{2}$	-1	0
x_2	0	0	1	2	$\frac{1}{2}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	0	2

得到辅助线性规划的最优解，再求解原线性规划问题.

小结：

对于线性规划问题

$$\begin{aligned} \min \quad & f = c^T x, \\ \text{s.t.} \quad & Ax = b, (b \geqslant 0) \\ & x \geqslant 0, \end{aligned} \tag{2.4.48}$$

考虑辅助线性规划问题

$$\begin{aligned} \min \quad & x_0 = e^T x_a, \\ \text{s.t.} \quad & Ax + x_a = b, \\ & x, x_a \geqslant 0. \end{aligned} \tag{2.4.49}$$

若辅助线性规划（2.4.49）的最优目标函数值 $\min x_0 > 0$，则原线性规划（2.4.48）无可行解；否则进行如下讨论：

若最优基中无人工变量，则去掉人工变量得到原线性规划（2.4.48）的基本可行解；否则不妨设人工变量 x_{n+r} 是基变量，讨论如下：

若 $a_{rj}=0$，$\forall j \in R$（非基变量非人工变量指标集），则第 r 个方程是多余方程. 去掉第 r 个方程和人工变量 x_{n+r} 对应的列，继续进行计算；否则存在 $k \in R$，使得 $a_{rk} \neq 0$，以 a_{rk} 为中心作转轴运算，这样，x_k 进基，x_{n+r} 出基，继续进行计算.

通常可以把两阶段放在一起，即

	f	x_0	x_1	$\cdots$	x_n	x_{n+1}	$\cdots$	x_{n+m}	RHS
f	1	0	$-c_1$	$\cdots$	$-c_n$	0	$\cdots$	0	0
x_0	0	1	0	$\cdots$	0	-1	$\cdots$	-1	0
x_{n+1}	0	0	a_{11}	$\cdots$	a_{1n}	1	$\cdots$	0	b_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_{n+m}	0	0	a_{m1}	$\cdots$	a_{mn}	0	$\cdots$	1	b_m

(2.4.50)

来进行求解，这里就不再重复了.

2.4.6 大 M 方法

考虑线性规划问题

$$P: \quad \begin{aligned} \min \quad & \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & A\boldsymbol{x}=\boldsymbol{b},\ (\boldsymbol{b}\geqslant 0) \\ & \boldsymbol{x}\geqslant 0. \end{aligned}$$

引进人工变量 $\boldsymbol{x}_a$，考虑线性规划问题

$$P(M): \quad \begin{aligned} \min \quad & \boldsymbol{c}^T\boldsymbol{x}+M\boldsymbol{e}^T\boldsymbol{x}_a, \\ \text{s.t.} \quad & A\boldsymbol{x}+\boldsymbol{x}_a=\boldsymbol{b}, \\ & \boldsymbol{x},\ \boldsymbol{x}_a\geqslant 0, \end{aligned}$$

其中 M 是一个充分大的数.

当 $\boldsymbol{x}=0$、$\boldsymbol{x}_a=\boldsymbol{b}$ 是$P(M)$的一个基本可行解，因此我们从基本可行解出发，对$P(M)$求解，分以下情况进行讨论.

1．$P(M)$有最优解

(1) 若 $(\boldsymbol{x}^*, 0)$ 是$P(M)$的最优解，则 $\boldsymbol{x}^*$ 是 P 的最优解.

事实上，若 $\boldsymbol{x}$ 是 P 的可行解，则 $(\boldsymbol{x}, 0)$ 是$P(M)$的可行解，因此有

$$\boldsymbol{c}^T\boldsymbol{x}+0\geqslant \boldsymbol{c}^T\boldsymbol{x}^*+0, \tag{2.4.51}$$

所以 $\boldsymbol{x}^*$是 P 的最优解.

【例 2.4.8】 用大 M 方法求解线性规划问题

$$\begin{aligned} \min \quad & x_1-2x_2, \\ \text{s.t.}\ & x_1+x_2\geqslant 2, \\ & -x_1+x_2\geqslant 1, \\ & x_2\leqslant 3, \\ & x_1,x_2\geqslant 0. \end{aligned}$$

解 引进松弛变量 x_3、x_4、x_5 和人工变量 x_6、x_7，得到

$P(M)$问题

$$\begin{aligned}\min\quad & x_1-2x_2+Mx_6+Mx_7,\\ \text{s.t.}\quad & x_1+x_2-x_3+x_6=2,\\ & -x_1+x_2-x_4+x_7=1,\\ & x_2+x_5=3,\\ & x_1,\ x_2,\ \cdots,\ x_7\geqslant 0.\end{aligned}$$

其原始单纯形表为：

	f	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
f	1	-1	2	0	0	0	$-M$	$-M$	0
x_6	0	1	1	-1	0	0	1	0	2
x_7	0	-1	1	0	-1	0	0	1	1
x_5	0	0	1	0	0	1	0	0	3

初始单纯形表为：

	f	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
f	1	-1	$2+2M^*$	$-M$	$-M$	0	0	0	$3M$
x_6	0	1	1	-1	0	0	1	0	2
x_7	0	-1	$\boxed{1}$	0	-1	0	0	1	1
x_5	0	0	1	0	0	1	0	0	3

经计算得到：

	f	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
f	1	$1+2M^*$	0	$-M$	$2+M$	0	0	$-2-2M$	$-2+M$
x_6	0	$\boxed{2}$	0	-1	1	0	1	-1	1
x_2	0	-1	1	0	-1	0	0	1	1
x_5	0	1	0	0	1	1	0	-1	2

	f	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
f	1	0	0	$\frac{1}{2}$	$\frac{3}{2}^*$	0	$-\frac{1}{2}-M$	$-\frac{3}{2}-M$	$-\frac{5}{2}$
x_1	0	1	0	$-\frac{1}{2}$	$\boxed{\frac{1}{2}}$	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$
x_2	0	0	1	$-\frac{1}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{2}$
x_5	0	0	0	$\frac{1}{2}$	$\frac{1}{2}$	1	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{3}{2}$

	f	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
f	1	-3	0	2^*	0	0	$-2-M$	$-M$	-4
x_4	0	2	0	-1	1	0	1	-1	1
x_2	0	1	1	-1	0	0	1	0	2
x_5	0	-1	0	$\boxed{1}$	0	1	-1	0	1

	f	x_1	x_2	x_3	x_4	x_5	x_6	x_7	RHS
f	1	-1	0	0	0	-2	$-M$	$-M$	-6
x_4	0	1	0	0	1	1	0	-1	2
x_2	0	0	1	0	0	1	0	0	3
x_3	0	-1	0	1	0	1	-1	0	1

最优解为 $\boldsymbol{x}^* = (0, 3, 1, 2, 0, 0, 0)^T$，$f^* = -6$.

注意：用大 M 方法得到的计算过程与两阶段法的计算过程完全相同.

（2）若（$\boldsymbol{x}^*$，$\boldsymbol{x}_a^*$）是$P(M)$的最优解且 $\boldsymbol{x}_a^* \neq 0$，则问题 P

无可行解.

若 P 存在可行解 $\boldsymbol{x}$，则 $(\boldsymbol{x}, 0)$ 是 $P(M)$ 的可行解，则有

$$\boldsymbol{c}^T\boldsymbol{x}^* + M\boldsymbol{e}^T\boldsymbol{x}_a^* \leqslant \boldsymbol{c}^T\boldsymbol{x}, \qquad (2.4.52)$$

这与 M 可以充分大矛盾.

【例 2.4.9】 用大 M 方法求解线性规划问题

$$\begin{aligned} \min \quad & -x_1 - 3x_2 + x_3, \\ \text{s.t.} \quad & x_1 + x_2 + 2x_3 \leqslant 4, \\ & -x_1 + x_3 \geqslant 4, \\ & x_3 \geqslant 3, \\ & x_1, \ x_2, \ x_3 \geqslant 0. \end{aligned}$$

解 引进松驰变量 x_4、x_5、x_6 和人工变量 x_7、x_8，得到 $P(M)$ 问题

$$\begin{aligned} \min \quad & -x_1 - 3x_2 + x_3 + Mx_7 + Mx_8, \\ \text{s.t.} \quad & x_1 + x_2 + 2x_3 + x_4 = 4, \\ & -x_1 + x_3 - x_5 + x_7 = 4, \\ & x_3 - x_6 + x_8 = 3, \\ & x_1, \ x_2, \ \cdots, \ x_8 \geqslant 0. \end{aligned}$$

原始单纯形表为：

	f	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	RHS
f	1	1	3	-1	0	0	0	$-M$	$-M$	0
x_4	0	1	1	2	1	0	0	0	0	4
x_7	0	-1	0	1	0	-1	0	1	0	4
x_8	0	0	0	1	0	0	-1	0	1	3

经计算得到：

	f	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	RHS
f	1	$1-M$	3	$-1+2M^*$	0	$-M$	$-M$	0	0	$7M$
x_4	0	1	1	$\boxed{2}$	1	0	0	0	0	4
x_7	0	-1	0	1	0	-1	0	1	0	4
x_8	0	0	0	1	0	0	-1	0	1	3

	f	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	RHS
f	1	$\frac{3}{2}-2M$	$\frac{7}{2}-M$	0	$\frac{1}{2}-M$	$-M$	$-M$	0	0	$2+3M$
x_3	0	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$	0	0	0	0	2
x_7	0	$-\frac{3}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	-1	0	1	0	2
x_8	0	$-\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	0	-1	0	1	1

当 M 充分大时，检验数均小于 0，达到最优解，但由于 $x_7=2$ 和 $x_8=1$，因此，原问题无可行解.

2．$P(M)$无有限最优解

(1) 若存在 $\lambda_k>0$ 且 $a_{ik}\leqslant 0$（$i=1,2,\cdots,m$），并且人工变量 $x_a^*=0$，则原线性规划 P 无有限最优解.

事实上，由于（x^*，0）是$P(M)$的可行解，则 x^* 是 P 的可行解. 由定理2.4.3后的讨论可知，存在 $D=\{\boldsymbol{x}\mid A\boldsymbol{x}+\boldsymbol{x}_a=\boldsymbol{b},\ \boldsymbol{x}\geqslant 0,\ \boldsymbol{x}_a\geqslant 0\}$ 的极方向

$$(\boldsymbol{d},\boldsymbol{d}_a)\geqslant 0, A\boldsymbol{d}+\boldsymbol{d}_a=0, \tag{2.4.53}$$

且

$$\boldsymbol{c}^{\mathrm{T}}\boldsymbol{d}+M\boldsymbol{e}^{\mathrm{T}}\boldsymbol{d}_a<0. \tag{2.4.54}$$

由于 $d_a \geqslant 0$，M 充分大，（2.4.54）式蕴涵着 $\boldsymbol{d}_a = 0$. 因此（2.4.53）式和（2.4.54）式变为

$$A\boldsymbol{d} = 0, \boldsymbol{d} \geqslant 0, \boldsymbol{c}^T\boldsymbol{d} < 0. \qquad (2.4.55)$$

令 $\boldsymbol{x} = \boldsymbol{x}^* + \alpha\boldsymbol{d}$，则 x 满足

$$A\boldsymbol{x} = A\boldsymbol{x}^* + \alpha A\boldsymbol{d} = A\boldsymbol{x}^* = \boldsymbol{b}, \qquad (2.4.56)$$

$$\boldsymbol{x}^* + \alpha\boldsymbol{d} \geqslant 0, \qquad (2.4.57)$$

和

$$\boldsymbol{c}^T\boldsymbol{x} = \boldsymbol{c}^T\boldsymbol{x}^* + \alpha\boldsymbol{c}^T\boldsymbol{d} \to +\infty, \qquad (2.4.58)$$

当 $\alpha \to +\infty$时，所以原问题无有限最优解.

【例 2.4.10】 用大 M 方法求解线性规划问题：

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

解 引进人工变量 x_5、x_6，得到$P(M)$问题

$$\begin{aligned} \min \quad & -x_1 - x_2 + Mx_5 + Mx_6 \\ \text{s.t.} \quad & x_1 - x_2 - x_3 + x_5 = 1, \\ & -x_1 + x_2 + 2x_3 - x_4 + x_6 = 1, \\ & x_1,\ x_2,\ \cdots,\ x_6 \geqslant 0. \end{aligned}$$

运算过程如下：

	f	x_1	x_2	x_3	x_4	x_5	x_6	RHS
f	1	1	1	0	0	$-M$	$-M$	0
x_5	0	1	-1	-1	0	1	0	1
x_6	0	-1	1	2	-1	0	1	1

	f	x_1	x_2	x_3	x_4	x_5	x_6	RHS
f	1	1	1	M^*	$-M$	0	0	$2M$
x_5	0	1	-1	-1	0	1	0	1
x_6	0	-1	1	$\boxed{2}$	-1	0	1	1

	f	x_1	x_2	x_3	x_4	x_5	x_6	RHS
f	1	$1+\frac{1}{2}M^*$	$1-\frac{1}{2}M$	0	$-\frac{1}{2}M$	0	$-\frac{1}{2}M$	$\frac{3}{2}M$
x_5	0	$\boxed{\frac{1}{2}}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{3}{2}$
x_3	0	$-\frac{1}{2}$	$\frac{1}{2}$	1	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$

	f	x_1	x_2	x_3	x_4	x_5	x_6	RHS
f	1	0	2^*	0	1	$-M-2$	$-M-1$	-3
x_1	0	1	-1	0	-1	2	1	3
x_3	0	0	0	1	-1	1	1	2

此时，$x_5=0$，$x_6=0$，当 x_2 增大时，有 $f\to-\infty$，问题无有限最优解.

(2) 若存在 $\lambda_k>0$ 且 $a_{ik}\leqslant 0$（$i=1,2,\cdots,m$），并且人工变量 $x_a^*\neq 0$，则原线性规划 P 无可行解.

证明：为了方便起见，我们假设得到如下单纯形表：

$$
\begin{array}{c|c|ccc|ccccc|ccc|c}
 & f & x_1 & \cdots & x_p & \cdots & x_j & \cdots & x_k & \cdots & x_{n+p+1} & \cdots & x_{n+m} & \text{RHS} \\
\hline
f & 1 & 0 & \cdots & 0 & \cdots & \lambda_j & \cdots & \lambda_k & \cdots & 0 & \cdots & 0 & f \\
\hline
x_1 & 0 & 1 & \cdots & 0 & \cdots & \alpha_{1j} & \cdots & \alpha_{1k} & \cdots & 0 & \cdots & 0 & \beta_1 \\
\vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & & \vdots & & \vdots & \ddots & \vdots & \vdots \\
x_p & 0 & 0 & \cdots & 1 & \cdots & \alpha_{pj} & \cdots & \alpha_{pk} & \cdots & 0 & \cdots & 0 & \beta_p \\
\hline
x_{n+p+1} & 0 & 0 & \cdots & 0 & \cdots & \alpha_{p+1j} & \cdots & \alpha_{p+1k} & \cdots & 1 & \cdots & 0 & \beta_{p+1} \\
\vdots & \vdots & \vdots & \ddots & \vdots & & \vdots & & \vdots & & \vdots & \ddots & \vdots & \vdots \\
x_{n+m} & 0 & 0 & \cdots & 0 & \cdots & \alpha_{mj} & \cdots & \alpha_{mk} & \cdots & 0 & \cdots & 1 & \beta_m \\
\hline
\end{array}
$$

因此得到

$$x_{n+i} = \sum_{j\in R}\alpha_{ij}x_j + \sum_{j\in J}\alpha_{in+j}x_{n+j} = \beta_i, \qquad (2.4.59)$$
$$i = p+1, p+2, \cdots, m,$$

其中 R 是非人工变量非基变量指标集，J 是人工变量非基变量指标集．对于 $i = p+1$，$p+2$，…，m 求和，并交换加法次序得到

$$\sum_{i=p+1}^{m} x_{n+i} = \sum_{j\in R}\Big(\sum_{i=p+1}^{m}\alpha_{ij}\Big)x_j + \sum_{j\in J}\Big(\sum_{i=p+1}^{m}\alpha_{in+j}\Big)x_{n+j}$$
$$= \sum_{i=p+1}^{m}\beta_i, \qquad (2.4.60)$$

如果 P 有可行解 $\bar{x}$，则 $(x, x_a) = (\bar{x}, 0)$ 是 $P(M)$ 的可行解，代入（2.4.60）式得到

$$\sum_{j\in R}\Big(\sum_{i=p+1}^{m}\alpha_{ij}\Big)\bar{x}_j = \sum_{i=p+1}^{m}\beta_i. \qquad (2.4.61)$$

下面证明 $\sum\limits_{i=p+1}^{m}\alpha_{ij}\leqslant 0$，$\forall j\in R$.

对于 $j=k$ 时，命题显然成立，因为 $\alpha_{ik}\leqslant 0$，$i=1,2,\cdots,m$. 当 $j\neq k$ 时，由检验数的定义，注意到人工变量对应的系数是 M，因此有

$$\begin{aligned}\lambda_j &= c_B^T B^{-1} p_j - c_j \\ &= \sum_{i=1}^{p} c_i\alpha_{ij} + M\left(\sum_{i=p+1}^{m}\alpha_{ij}\right) - c_j \\ &\leqslant \lambda_k,\end{aligned} \tag{2.4.62}$$

注意 M 充分大，若 $\sum\limits_{j=p+1}^{m}\alpha_{ij}>0$，则（2.4.62）式矛盾. 因此 $\sum\limits_{i=p+1}^{m}\alpha_{ij}\leqslant 0$.

由于 $\beta_i>0$，$\sum\limits_{j=p+1}^{m}\alpha_{ij}\leqslant 0$（$i=p+1,p+2,\cdots,m$），$\overline{x}_j\geqslant 0$（$j\in R$），所以（2.4.61）式不可能成立，因此，$P$ 无可行解.

【例 2.4.11】 用大 M 方法求解线性规划问题：

$$\begin{aligned}\min\quad & -x_1-x_2, \\ \text{s.t.}\quad & x_1-x_2\geqslant 1, \\ & -x_1+x_2\geqslant 1, \\ & x_1,\ x_2\geqslant 0.\end{aligned}$$

解 引进松驰变量 x_3、x_4 和人工变是 x_5、x_6，得到$P(M)$问题：

$$\begin{aligned}\min\quad & -x_1-x_2+Mx_5+Mx_6 \\ \text{s.t.}\quad & x_1-x_2-x_3+x_5=1, \\ & -x_1+x_2-x_4+x_6=1, \\ & x_1,\ x_2,\ \cdots,\ x_6\geqslant 0.\end{aligned}$$

运算过程如下：

	f	x_1	x_2	x_3	x_4	x_5	x_6	RHS
f	1	1	1	0	0	$-M$	$-M$	0
x_5	0	1	-1	-1	0	1	0	1
x_6	0	-1	1	0	-1	0	1	1

	f	x_1	x_2	x_3	x_4	x_5	x_6	RHS
f	1	1^*	1	$-M$	$-M$	0	0	$2M$
x_5	0	$\boxed{1}$	-1	-1	0	1	0	1
x_6	0	-1	1	0	-1	0	1	1

	f	x_1	x_2	x_3	x_4	x_5	x_6	RHS
f	1	0	2^*	$1-M$	$-M$	-1	0	$2M-1$
x_1	0	1	-1	-1	0	1	0	1
x_6	0	0	0	-1	-1	1	1	2

此时，$x_6=2$，而 $\lambda_2>0$．原问题无可行解．

事实上，方程 $x_1-x_2\geqslant 1$ 与方程 $-x_1+x_2\geqslant 1$ 矛盾．

§2.5　改进单纯形法

本节将讨论改进单纯形法，其主要目的是减少算法的运算量，为此我们先回顾一下从基本可行解开始的单纯形法．对于线性规划问题

$$\begin{aligned}\min\quad & \boldsymbol{c}^T\boldsymbol{x},\\ \text{s.t.}\quad & A\boldsymbol{x}=\boldsymbol{b}, \qquad\qquad (2.5.1)\\ & \boldsymbol{x}\geqslant 0,\end{aligned}$$

考虑从基本可行解开始的单纯形法，设初始单纯形表如下：

	f	x_1	$\cdots$	x_r	$\cdots$	x_m	$\cdots$	x_j	$\cdots$	x_k	$\cdots$	RHS
f	1	0	$\cdots$	0	$\cdots$	0	$\cdots$	λ_j	$\cdots$	λ_k	$\cdots$	f^0
x_1	0	1	$\cdots$	0	$\cdots$	0	$\cdots$	α_{1j}	$\cdots$	α_{1k}	$\cdots$	β_1
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_r	0	0	$\cdots$	1	$\cdots$	0	$\cdots$	α_{rj}	$\cdots$	α_{rk}	$\cdots$	β_r
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_m	0	0	$\cdots$	0	$\cdots$	1	$\cdots$	α_{mj}	$\cdots$	α_{mk}	$\cdots$	β_m

(2.5.2)

若以 α_{rk} 为中心作转轴运算，其计算是在一个 $(m+1)\times(n+1)$ 的表上进行．而事实上，有些计算是没有必要的．

回顾一下单纯形法的计算过程．计算过程需要计算 λ_j、α_{ij}、β_i 和目标函数值 f，其计算公式如下：

$$\lambda_j = \boldsymbol{c}_B^T B^{-1} \boldsymbol{p}_j - c_j,\quad j \in R,\quad f = \boldsymbol{c}_B^T B^{-1} \boldsymbol{b},$$

$$\overline{\boldsymbol{p}_j} = \begin{bmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \vdots \\ \alpha_{mj} \end{bmatrix} = B^{-1} p_j,\quad \overline{\boldsymbol{b}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} = B^{-1}\boldsymbol{b}.$$

为了便于计算，令 $\boldsymbol{\pi} = B^{-T}\boldsymbol{c}_B$，或 $\boldsymbol{\pi}^T = \boldsymbol{c}_B^T B^{-1}$．因此，只要知道 B^{-1}、$\boldsymbol{\pi}$、$\overline{\boldsymbol{b}}$ 和 $\boldsymbol{c}_B^T\overline{\boldsymbol{b}}$，就可以得到单纯形表的全部信息，因此构造改进单纯形表为：

基变量	RHS
$\boldsymbol{\pi}^T$	$\boldsymbol{c}_B^T\overline{\boldsymbol{b}}$
B^{-1}	$\overline{\boldsymbol{b}}$

(2.5.3)

此时，检验数的计算为 $\lambda_j = \boldsymbol{\pi}^T\boldsymbol{p}_j - c_j$，$j \in R$．若存在 $\lambda_k > 0$，需

要计算 $\bar{\boldsymbol{p}}_k = B^{-1}\boldsymbol{p}_k$. 若 $\boldsymbol{p}_k \leqslant 0$，则停止计算（无有限最优解）；否则作转轴计算. 在改进单纯形表中增加一列 $\begin{bmatrix} \lambda_k \\ \bar{\boldsymbol{p}}_k \end{bmatrix}$，得到：

基变量	RHS		x_k
$\boldsymbol{\pi}^T$	$\boldsymbol{c}_B^T \bar{\boldsymbol{b}}$		λ_k
B^{-1}	$\bar{\boldsymbol{b}}$		$\bar{\boldsymbol{p}}_k$

(2.5.4)

以 a_{rk} 为中心作转轴运算，得到新的 B^{-1}.

算法 2.5.1（改进单纯形算法）

(1) 取初始可行基 B，计算 $\boldsymbol{\pi}^T = \boldsymbol{c}_B^T B^{-1}$，$\bar{\boldsymbol{b}} = B^{-1}\boldsymbol{b}$，列单纯形表：

	基变量	RHS
f	$\boldsymbol{\pi}^T$	$\boldsymbol{c}_B^T \bar{\boldsymbol{b}}$
$\boldsymbol{x}_B$	B^{-1}	$\bar{\boldsymbol{b}}$

(2) 计算

$$\lambda_j = \boldsymbol{\pi}^T \boldsymbol{p}_j - c_j, \quad j \in R.$$

令 $\lambda_k = \max\{\lambda_j \mid j \in R\}$，若 $\lambda_k \leqslant 0$，则停止计算（$\boldsymbol{x}_B = \bar{\boldsymbol{b}}$，$\boldsymbol{x}_N = 0$ 最优解）；

(3) 计算

$$\bar{\boldsymbol{p}}_k = B^{-1}\boldsymbol{p}_k,$$

若 $\bar{\boldsymbol{p}}_k \leqslant 0$，则停止计算（无有限最优解）；否则增加一列 $\begin{bmatrix} \lambda_k \\ \bar{\boldsymbol{p}}_k \end{bmatrix}$，得到：

	基变量	RHS		x_k
f	$\boldsymbol{\pi}^T$	$\boldsymbol{c}_B^T\overline{\boldsymbol{b}}$		λ_k
$\boldsymbol{x}_B$	B^{-1}	$\overline{\boldsymbol{b}}$		$\overline{\boldsymbol{p}}_k$

确定

$$\theta = \min\left\{\frac{\beta_i}{\alpha_{ik}} \mid \alpha_{ik} > 0\right\} = \frac{\beta_r}{\alpha_{rk}},$$

然后以 α_{rk}为中心进行转轴运算，设 $\boldsymbol{\pi}^T=(\pi_1,\ \pi_2,\ \cdots,\ \pi_m)$，$\overline{\boldsymbol{b}}=(\beta_1,\ \beta_2,\ \cdots,\ \beta_m)^T$，$B^{-1}=(b_{ij})_{m\times m}$，$\overline{\boldsymbol{p}}_k=(\alpha_{1k},\ \alpha_{2k},\ \cdots,\ \alpha_{mk})^T$，置

$$b_{rj} = \frac{b_{rj}}{\alpha_{rk}},\ j = 1,2,\cdots,m,$$

$$\beta_r = \frac{\beta_r}{\alpha_{rk}},$$

$$b_{rj} = b_{rj} - \alpha_{rk}b_{rj},\ i = 1,2,\cdots,m,\ i \neq r,\ j = 1,2,\cdots,m,$$

$$\beta_i = \beta_i - \alpha_{ik}\beta_r,\ i = 1,2,\cdots,m,\ i \neq r,$$

$$\pi_j = \pi_j - \lambda_k b_{rj},\ j = 1,2,\cdots.m,$$

$$f = f - \lambda_k\beta_r,$$

然后转（2）.

注意到，在改进单纯形法中运算是在 $(m+1)\times(m+1)$ 阶表上进行运算，而单纯形法需要在 $(m+1)\times(n+1)$ 的表上运算. 当 $n \gg m$ 时，改进单纯形法要比单纯形法运算量少得多.

【例 2.5.1】 用改进单纯形法求解线性规划问题：

$$\begin{aligned}
\min\quad & -x_1-2x_2+x_3-x_4-4x_5+2x_6, \\
\text{s.t.}\quad & x_1+x_2+x_3+x_4+x_5+x_6 \leqslant 6, \\
& 2x_1-x_2-2x_3+x_4 \leqslant 4, \\
& x_3+x_4+2x_5+x_6 \leqslant 4,
\end{aligned}$$

$x_1, x_2, \cdots, x_6 \geqslant 0.$

解 引进松弛变量 x_7、x_8、x_9，将问题化为标准形式，其基为 $B = (\boldsymbol{p}_7, \boldsymbol{p}_8, \boldsymbol{p}_9) = I$，$\boldsymbol{\pi}^T = \boldsymbol{c}_B^T B^{-1} = (0, 0, 0)$ 和 $\overline{\boldsymbol{b}} = \boldsymbol{b} = (6, 4, 4)^T$.

第一步，列初始单纯形表：

	基变量			RHS
f	0	0	0	0
x_7	1	0	0	6
x_8	0	1	0	4
x_9	0	0	1	4

计算 λ_j：

$$\lambda_1 = \pi^T p_1 - c_1 = -c_1 = 1,$$
$$\lambda_2 = \pi^T p_2 - c_2 = -c_2 = 2,$$
$$\lambda_3 = \pi^T p_3 - c_3 = -c_3 = -1,$$
$$\lambda_4 = \pi^T p_4 - c_4 = -c_4 = -1,$$
$$\lambda_5 = \pi^T p_5 - c_5 = -c_5 = 4,$$
$$\lambda_6 = \pi^T p_6 - c_6 = -c_6 = -2,$$

这里 $\lambda_5 = \max\{\lambda_j \mid j \in R\} = 4$. 计算 $\overline{\boldsymbol{p}}_5 = B^{-1}\boldsymbol{p}_5 = \boldsymbol{p}_5 = \begin{bmatrix}1\\0\\2\end{bmatrix} \not\leqslant 0$.

增加一列：

	基变量			RHS	x_5
f	0	0	0	0	4
x_7	1	0	0	6	1
x_8	0	1	0	4	0
x_9	0	0	1	4	$\boxed{2}$

传转轴运算，得到：

	基变量			RHS
f	0	0	-2	-8
x_7	1	0	$-\frac{1}{2}$	4
x_8	0	1	0	4
x_5	0	0	$\frac{1}{2}$	2

第二步. 此时 $\boldsymbol{\pi}^T=(0,0,-2)$，计算 λ_j：

$$\lambda_1=\boldsymbol{\pi}^T\boldsymbol{p}_1-c_1=(0,0,-2)\begin{bmatrix}1\\2\\0\end{bmatrix}+1=1,$$

$$\lambda_2=\boldsymbol{\pi}^T\boldsymbol{p}_2-c_2=(0,0,-2)\begin{bmatrix}1\\-1\\0\end{bmatrix}+2=2,$$

$$\lambda_3=\boldsymbol{\pi}^T\boldsymbol{p}_3-c_3=(0,0,-2)\begin{bmatrix}1\\-2\\1\end{bmatrix}-1=-3,$$

$$\lambda_4=\boldsymbol{\pi}^T\boldsymbol{p}_4-c_4=(0,0,-2)\begin{bmatrix}1\\1\\1\end{bmatrix}+1=-1,$$

$$\lambda_6=\boldsymbol{\pi}^T\boldsymbol{p}_6-c_6=(0,0,-2)\begin{bmatrix}1\\0\\1\end{bmatrix}+2=-4,$$

$$\lambda_9=\boldsymbol{\pi}^T\boldsymbol{p}_9-c_9=(0,0,-2)\begin{bmatrix}0\\0\\1\end{bmatrix}-0=-2.$$

所以 $\lambda_2=\max\{\lambda_j|j\in R\}=2$. 计算

$$\overline{p}_2 = B^{-1}p_2 = \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \not\leq 0,$$

增加一列，得到：

	基变量			RHS	x_2
f	0	0	-2	-8	2
x_7	1	0	$-\frac{1}{2}$	4	$\boxed{1}$
x_8	0	1	0	4	-1
x_5	0	0	$\frac{1}{2}$	2	0

作转轴运算，得到：

	基变量			RHS
f	-2	0	-1	-16
x_2	1	0	$-\frac{1}{2}$	4
x_8	1	1	$-\frac{1}{2}$	8
x_5	0	0	$\frac{1}{2}$	2

第三步．计算 λ_j：

$$\lambda_1 = \pi^T p_1 - c_1 = (-2, 0, -1)\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + 1 = -1,$$

$$\lambda_3=\boldsymbol{\pi}^T\boldsymbol{p}_3-c_3=(-2,0,-1)\begin{bmatrix}1\\-2\\1\end{bmatrix}-1=-4,$$

$$\lambda_4=\boldsymbol{\pi}^T\boldsymbol{p}_4-c_4=(-2,0,-1)\begin{bmatrix}1\\1\\1\end{bmatrix}+1=-2,$$

$$\lambda_6=\boldsymbol{\pi}^T\boldsymbol{p}_6-c_6=(-2,0,-1)\begin{bmatrix}1\\0\\1\end{bmatrix}-2=-5,$$

$$\lambda_7=\boldsymbol{\pi}^T\boldsymbol{p}_2-c_2=(-2,0,-1)\begin{bmatrix}1\\0\\0\end{bmatrix}-0=-2,$$

$$\lambda_9=\boldsymbol{\pi}^T\boldsymbol{p}_9-c_9=(-2,0,-1)\begin{bmatrix}0\\0\\1\end{bmatrix}-0=-1,$$

所以 $\lambda_j\leqslant 0$，$j\in R$，得到最优解 $\boldsymbol{x}^*=(0,4,0,0,2,0,0,8,0)^T$，$f^*=-16$.

习 题 二

2.1 将下列线性规划问题化成标准形式.

$$\begin{aligned}\max\quad & f=x_1+2x_2+5x_3,\\ \text{s.t.}\quad & x_1+x_2-x_3\leqslant 10,\\ & 2x_1+x_2+2x_3\geqslant -1,\\ & x_1,\ x_2\geqslant 0,\\ & x_3\ \text{无限制}.\end{aligned}$$

2.2 用图解法求解下列线性规划问题：

(1) $\min\quad f=-x_1+2x_2,$

$$\begin{aligned} \text{s.t.} \quad & -x_1+x_2\leqslant 2, \\ & x_1+2x_2\leqslant 6, \\ & x_1,\ x_2\geqslant 0. \end{aligned}$$

(2)
$$\begin{aligned} \max \quad & f=-x_1+2x_2, \\ \text{s.t.} \quad & -x_1+x_2\leqslant 2, \\ & x_1+2x_2\leqslant 6, \\ & x_1,\ x_2\geqslant 0. \end{aligned}$$

(3)
$$\begin{aligned} \max \quad & f=3x_1+6x_2, \\ \text{s.t.} \quad & -x_1+x_2\leqslant 2, \\ & x_1+2x_2\leqslant 6, \\ & x_1,\ x_2\geqslant 0. \end{aligned}$$

(4)
$$\begin{aligned} \max \quad & f=-x_1+2x_2, \\ \text{s.t.} \quad & -x_1+x_2\leqslant 2, \\ & x_1,\ x_2\geqslant 0. \end{aligned}$$

(5)
$$\begin{aligned} \min \quad & f=3x_1+6x_2, \\ \text{s.t.} \quad & x_1-x_2\leqslant -2, \\ & x_1+x_2\leqslant -5, \\ & x_1,\ x_2\geqslant 0. \end{aligned}$$

2.3 求下列线性规划问题的全部基本解.

$$\begin{aligned} \min \quad & f=-x_1+2x_2, \\ \text{s.t.} \quad & -x_1+x_2\leqslant 2, \\ & x_1+2x_2\leqslant 6, \\ & x_1,\ x_2\geqslant 0. \end{aligned}$$

2.4 考虑线性规划问题

$$\begin{aligned} \min \quad & \boldsymbol{c}^T\boldsymbol{x},\ (\boldsymbol{c}\neq 0) \\ \text{s.t.} \quad & A\boldsymbol{x}\leqslant \boldsymbol{b}, \\ & \boldsymbol{x}\geqslant 0. \end{aligned}$$

若 x^0 满足 $Ax^0 < b$，$x^0 > 0$，证明：x^0 不是该问题的最优解.

2.5 用单纯形法求解习题 2.2 的 (1)、(2)、(3)，并与图解法作比较.

2.6 用单纯形法求解下列线性规划问题：

(1)
$$
\begin{aligned}
\max\quad & f = 5x_1 + 4x_2, \\
\text{s.t.}\quad & x_1 + 2x_2 \leqslant 6, \\
& 2x_1 - x_2 \leqslant 4, \\
& 5x_1 + 3x_2 \leqslant 15, \\
& x_1,\ x_2 \geqslant 0.
\end{aligned}
$$

(2)
$$
\begin{aligned}
\max\quad & f = 3x_1 + 2x_2, \\
\text{s.t.}\quad & 2x_1 - 3x_2 \leqslant 3, \\
& -x_1 + x_2 \leqslant 5, \\
& x_1,\ x_2 \geqslant 0.
\end{aligned}
$$

2.7 已知线性规划问题的初始单纯形表和当前单纯形表如下，求出 a 到 l 的值.

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	a	1	-2	0	0	0
x_4	0	b	c	d	1	0	6
x_5	0	-1	3	e	0	1	1

初始单纯形表

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	0	7	j	k	l	-9
x_1	0	g	2	-1	$\frac{1}{2}$	0	f
x_5	0	h	i	1	$\frac{1}{2}$	1	4

当前单纯形表

2.8　下列单纯形表是求解极小化问题 $\min -5x_1-3x_2$ 得到的当前单纯形表，其中 x_3、x_4 是松驰变量，不等式形式为“$\leqslant$”.

	f	x_1	x_2	x_3	x_4	RHS
f	1	b	-1	f	g	-10
x_3	0	c	0	1	$\frac{1}{5}$	2
x_1	0	d	e	0	1	a

(1) 求未知量 a 到 g.

(2) 求 B^{-1}.

(3) 该表是否为最优单纯形表.

2.9　证明：设线性规划问题是非退化的，若检验数 $\lambda_j<0$，$\forall j\in R$，则该线性规划问题的最优解是唯一的.

2.10　证明：设 $\boldsymbol{x}^{(1)}$，$\boldsymbol{x}^{(2)}$，…，$\boldsymbol{x}^{(k)}$ 是线性规划问题的全部最优基本可行解，则最优解的集合可表示为

$$S^*=\left\{\boldsymbol{x}^* \mid \boldsymbol{x}^*=\sum_{i=1}^{k}\alpha_i\boldsymbol{x}^{(i)},\alpha_i\geqslant 0,\sum_{i=1}^{k}\alpha_i=1\right\}.$$

2.11　用二阶段法求解下列线性规划问题：

(1)
$$\begin{aligned}\min\quad & f=3x_1-2x_2+x_3,\\ \text{s.t.}\quad & 2x_1+x_2+x_3\leqslant 3,\\ & 3x_1+x_2+x_3=2,\\ & x_1+x_3=3,\end{aligned}$$

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

(2) $\max\ f = 2x_1 - x_2 + x_3,$

s.t. $x_1 + x_2 - 2x_3 \leqslant 8,$

$4x_1 - x_2 + x_3 \geqslant 2,$

$2x_1 + 3x_2 - x_3 \geqslant 4,$

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

(3) $\min\ f = x_1 - x_2 - 2x_3,$

s.t. $x_1 - 2x_2 = 2,$

$x_1 - 3x_2 - x_3 = 1,$

$x_1 - x_2 + x_3 = 3,$

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

(4) $\min\ f = x_1 - x_2 + x_3,$

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

$x_1 + x_3 = 2,$

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

2.12 在二阶段法中求解辅助线性规划问题可以用来检查是否有多余结束. 考虑如下三个约束

$$x_1 - x_2 \geqslant 2,$$
$$2x_1 + 3x_2 \geqslant 4,$$
$$3x_1 + 2x_2 \geqslant 6,$$

注意到第三个约束是前两个约束之和. 问求解辅助线性规划问题时, 能否发现它是多余约束吗? 试讨论之.

2.13 用大 M 方法求解下列线性规划问题:

(1) $\min\ f = 3x_1 - 2x_2 + 5x_3,$

s.t. $x_1 + 2x_2 + x_3 \geqslant 5,$

$-3x_1 + x_2 - x_3 \leqslant 4,$

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

(2) $\min \quad f=x_1-2x_2+2x_3+x_4,$

$$\begin{aligned}\text{s.t.}\quad & x_1+x_2+x_3+x_4=4,\\ & x_1+x_2-x_3-x_4=6,\\ & x_1,\ x_2,\ x_3,\ x_4\geqslant 0.\end{aligned}$$

(3) $\max \quad f=x_1+x_2,$

$$\begin{aligned}\text{s.t.}\quad & x_1+x_2\geqslant 1,\\ & x_1-x_2\leqslant 1,\\ & -x_1+x_2\leqslant 1,\\ & x_1,\ x_2,\ \geqslant 0.\end{aligned}$$

(4) $\min \quad f=-x_1+x_2,$

$$\begin{aligned}\text{s.t.}\quad & x_1-x_2\leqslant -2,\\ & -2x_1+x_2\leqslant -2,\\ & x_1,\ x_2,\ \geqslant 0.\end{aligned}$$

2.14 能否出现这种情况，大 M 问题（P（M）问题）无有限最优解，而原问题（P 问题）的最优解有界.

2.15 设原问题（P 问题）的可行解集

$$D=\{\boldsymbol{x}\mid A\boldsymbol{x}=\boldsymbol{b},\boldsymbol{x}\geqslant 0\}$$

有界，问是否出现大 M 问题（P（M）问题）的可行解集

$$D_a=\{(\boldsymbol{x},\boldsymbol{x}_a)\mid A\boldsymbol{x}+\boldsymbol{x}_a=\boldsymbol{b},\boldsymbol{x},\boldsymbol{x}_a\geqslant 0\}$$

无界的情况，试用大 M 算法的求解过程讨论这个问题.

2.16 用改进单纯形法求解下列线性规划问题：

(1) $\min \quad f=-2x_2+x_3,$

$$\begin{aligned}\text{s.t.}\quad & -x_1+2x_2-x_3\leqslant 4,\\ & x_1+x_2+x_3\leqslant 9,\\ & 2x_1-x_2-x_3\leqslant 5,\\ & x_1,\ x_2,\ x_3\geqslant 0.\end{aligned}$$

(2) $\min \quad f=x_1+6x_2-7x_3+x_4+x_5,$

$$
\text{s.t.} \quad \begin{aligned}
& x_1 - \frac{3}{4}x_2 + 2x_3 - \frac{1}{4}x_4 = 5, \\
& \frac{1}{4}x_2 + 3x_3 - \frac{3}{4}x_4 + x_5 = 6, \\
& x_1,\ x_2,\ x_3,\ x_4,\ x_5 \geqslant 0.
\end{aligned}
$$

第三章　线性规划的对偶问题

对偶是线性规划中最重要的概念之一，对于每一个线性规划问题（称为原始问题）都有另一个线性规划问题（称为对偶问题）与之相对应，原始问题与对偶问题之间有着密切的关系，它们是从不同角度来描述实质上相同的问题.

本章主要介绍线性规划的对偶问题，相应的对偶理论以及对偶单纯形算法.

§3.1 对 偶 问 题

3.1.1 对偶问题的提出

【例 3.1.1】 某工厂打算生产 m 种产品，每种产品的产量至少为 b_i（$i=1, 2, \cdots, m$）个单位. 生产这 m 种产品需要 n 种原料，每种原料的单位成本为 c_j（$j=1, 2, \cdots, n$），并且每一个单位的第 j 种原料可以生产第 i 种产品 a_{ij} 个单位，问如何安排生产使成本最低?

解 设第 j 种原料的需要量为 x_j 个单位,则总成本为 $\sum_{j=1}^{n} c_j x_j$. 第 i 种产品的产量为 $\sum_{j=1}^{n} a_{ij} x_j$. 所以得到如下线性规划问题:

$$
\begin{aligned}
\min\quad & \sum_{j=1}^{n} c_j x_j, \\
\text{s.t.}\quad & \sum_{j=1}^{n} a_{ij} x_j \geqslant b_i, \quad i=1, 2, \cdots, m, \\
& x_j \geqslant 0, \qquad\qquad j=1, 2, \cdots, n.
\end{aligned}
$$

其矩阵形式为

$$\begin{aligned}\min\quad & \boldsymbol{c}^T\boldsymbol{x}, \\ \text{s.t.}\quad & A\boldsymbol{x}\geqslant \boldsymbol{b}, \\ & \boldsymbol{x}\geqslant 0,\end{aligned}\tag{3.1.1}$$

称问题（3.1.1）为**原始问题**（Primal problem）.

现从另一角度讨论这个问题. 设第 i 种产品的单价为 y_i，问如何定价使产品的总产值最大?

设第 i 种产品的产量仍为 b_i，那么总产值为 $\sum\limits_{i=1}^{m} b_i y_i$. 若 a_{ij} 的意义不变，那么 $\sum\limits_{i=1}^{m} a_{ij} y_i$ 表示使用第 j 种原料的隐含价格. 自然第 j 种原料的隐含价格不能超过它的实际价格 c_j，即

$$\sum_{i=1}^{m} a_{ij} y_i \leqslant c_j, (j = 1,2,\cdots,n)$$

因此，得到另一个线性规划问题

$$\begin{aligned}\max\quad & \sum_{i=1}^{m} b_i y_i, \\ \text{s.t.}\quad & \sum_{i=1}^{m} a_{ij} y_i \leqslant c_j, \quad j=1,\ 2,\ \cdots,\ n, \\ & y_i \geqslant 0, \qquad\quad i=1,\ 2,\ \cdots,\ m,\end{aligned}$$

其矩阵形式为

$$\begin{aligned}\min\quad & \boldsymbol{b}^T\boldsymbol{y}, \\ \text{s.t.}\quad & A^T\boldsymbol{y}\geqslant \boldsymbol{c}, \\ & \boldsymbol{y}\geqslant 0,\end{aligned}\tag{3.1.2}$$

称问题（3.1.2）是问题（3.1.1）的**对偶**（dual），或称为**对偶问题**（dual problem）. 它是从不同角度对同一个问题进行描述的. 由于问题（3.1.2）是由问题（3.1.1）导出的，因此也称问题（3.1.1）为对偶的标准形式.

3.1.2 对偶问题的一般形式

1. 对偶问题的对偶

$$
\begin{array}{ll} \min & c^T x, \\ \text{s.t.} & Ax \geqslant b, \\ & x \geqslant 0. \end{array}
\xrightarrow{\text{对偶}}
\begin{array}{ll} \max & b^T y, \\ \text{s.t.} & A^T y \leqslant c, \\ & y \geqslant 0. \end{array}
$$

$$
\xrightarrow{\text{等价}}
\begin{array}{ll} \min & (-b)^T y, \\ \text{s.t.} & -A^T y \geqslant -c, \\ & y \geqslant 0. \end{array}
\xrightarrow{\text{对偶}}
\begin{array}{ll} \max & (-c)^T x, \\ \text{s.t.} & (-A^T)^T x \leqslant -b, \\ & x \geqslant 0. \end{array}
$$

$$
\xrightarrow{\text{等 价}}
\begin{array}{ll} \min & c^T x, \\ \text{s.t.} & Ax \geqslant b, \\ & x \geqslant 0. \end{array}
$$

由此得到如下定理.

定理 3.1.1 对偶问题的对偶是原始问题.

2. 线性规划标准形式的对偶

考虑线性规划的标准形式

$$
\begin{array}{ll} \min & c^T x, \\ \text{s.t.} & Ax = b, \\ & x \geqslant 0, \end{array} \tag{3.1.3}
$$

将它化为不等式问题

$$
\begin{array}{ll} \min & c^T x, \\ \text{s.t.} & Ax \geqslant b, \\ & -Ax \geqslant -b, \\ & x \geqslant 0, \end{array}
$$

其对偶问题为

$$
\begin{array}{ll} \max & b^T y' + (-b)^T y'', \\ \text{s.t.} & A^T y' - A^T y'' \leqslant c, \\ & y' \geqslant 0, \ y'' \geqslant 0. \end{array}
$$

令 $y = y' - y''$，得到

$$
\begin{array}{ll} \max & b^T y, \\ \text{s.t.} & A^T y \leqslant c, \end{array} \tag{3.1.4}
$$

$\boldsymbol{y}$ 无限制.

3. 混合形式问题的对偶考虑线性规划问题

$$\begin{aligned}
\min \quad & \boldsymbol{c}^T\boldsymbol{x}, \\
\text{s.t.} \quad & A^{(1)}\boldsymbol{x} \geqslant \boldsymbol{b}^{(1)}, \\
& A^{(2)}\boldsymbol{x} = \boldsymbol{b}^{(2)}, \\
& A^{(3)}\boldsymbol{x} \leqslant \boldsymbol{b}^{(3)}, \\
& \boldsymbol{x} \geqslant 0,
\end{aligned} \tag{3.1.5}$$

将它写成线性规划的标准形式

$$\begin{aligned}
\min \quad & \boldsymbol{c}^T\boldsymbol{x}, \\
\text{s.t.} \quad & A^{(1)}\boldsymbol{x} - \boldsymbol{x}_s = \boldsymbol{b}^{(1)}, \\
& A^{(2)}\boldsymbol{x} = \boldsymbol{b}^{(2)}, \\
& -A^{(3)}\boldsymbol{x} - \boldsymbol{x}_t = -\boldsymbol{b}^{(3)}, \\
& \boldsymbol{x},\ \boldsymbol{x}_s,\ \boldsymbol{x}_t \geqslant 0,
\end{aligned}$$

由标准形式的对偶得到该问题的对偶问题

$$\begin{aligned}
\max \quad & \boldsymbol{b}^{(1)^T}\boldsymbol{y}^{(1)} + \boldsymbol{b}^{(2)^T}\boldsymbol{y}^{(2)} - \boldsymbol{b}^{(3)^T}\boldsymbol{y}^{(3)}, \\
\text{s.t.} \quad & A^{(1)^T}\boldsymbol{y}^{(1)} + A^{(2)^T}\boldsymbol{y}^{(2)} - A^{(3)^T}\boldsymbol{y}^{(3)} \leqslant \boldsymbol{c}, \\
& -\boldsymbol{y}^{(1)} \leqslant 0, \\
& -\boldsymbol{y}^{(3)} \leqslant 0, \\
& \boldsymbol{y}^{(1)},\ \boldsymbol{y}^{(2)},\ \boldsymbol{y}^{(3)} \text{无限制},
\end{aligned}$$

化简为

$$\begin{aligned}
\max \quad & \boldsymbol{b}^{(1)^T}\boldsymbol{y}^{(1)} + \boldsymbol{b}^{(2)^T}\boldsymbol{y}^{(2)} - \boldsymbol{b}^{(3)^T}\boldsymbol{y}^{(3)}, \\
\text{s.t.} \quad & A^{(1)^T}\boldsymbol{y}^{(1)} + A^{(2)^T}\boldsymbol{y}^{(2)} - A^{(3)^T}\boldsymbol{y}^{(3)} \leqslant \boldsymbol{c}, \\
& \boldsymbol{y}^{(1)},\ \boldsymbol{y}^{(3)} \geqslant 0,\ \boldsymbol{y}^{(2)} \text{无限制}.
\end{aligned} \tag{3.1.6}$$

§ 3.2 线性规划的对偶理论

本节介绍与对偶有关的基本理论.

定理 3.2.1　若 $\boldsymbol{x}$、$\boldsymbol{y}$ 分别是原始问题和对偶问题的可行解，则它们对应的目标函数满足

$$\boldsymbol{c}^T\boldsymbol{x} \geqslant \boldsymbol{b}^T\boldsymbol{y}. \tag{3.2.1}$$

证明　考虑对偶的标准形式. 由约束条件，有 $\boldsymbol{y}^T(\boldsymbol{b}-A\boldsymbol{x})\geqslant 0$ 和 $(\boldsymbol{c}^T-\boldsymbol{y}^TA)\boldsymbol{x}\geqslant 0$，因此

$$\boldsymbol{c}^T\boldsymbol{x} \geqslant \boldsymbol{y}^TA\boldsymbol{x} \geqslant \boldsymbol{b}^T\boldsymbol{y}.$$

推论 3.2.1　若存在 $\boldsymbol{x}^0$ 和 $\boldsymbol{y}^0$ 分别是原始问题和对偶问题的可行解，并且满足 $\boldsymbol{c}^T\boldsymbol{x}^0=\boldsymbol{b}^T\boldsymbol{y}^0$，则 $\boldsymbol{x}^0$ 和 $\boldsymbol{y}^0$ 分别是原始问题和对偶问题的最优解.

证明　考虑对偶的标准形式，只需证明 x^0 是原始问题的最优解.

设 x 是原始问题的可行解，由定理 3.2.1，有

$$\boldsymbol{c}^T\boldsymbol{x} \geqslant \boldsymbol{b}^T\boldsymbol{y}^0 = \boldsymbol{c}^T\boldsymbol{x}^0.$$

推论 3.2.2　若原始问题和对偶问题之一的目标函数值无界，则另一个问题不可行.

证明　若对偶问题的目标函数值无上界，则原始问题不可行；否则，设 $\boldsymbol{x}^0$ 是原始问题的可行解，$\boldsymbol{y}$ 是对偶问题的可行解，由定理 3.2.1，有 $\boldsymbol{c}^T\boldsymbol{x}^0\geqslant\boldsymbol{b}^T\boldsymbol{y}$，与对偶问题无上界矛盾.

定理 3.2.2　若原始问题与对偶问题之一有最优解，则两问题均有最优解，并且它们的最优目标函数值相等.

证明　设原始问题

$$\begin{aligned} \min\quad & \boldsymbol{c}^T\boldsymbol{x}, \\ \text{s.t.}\quad & A\boldsymbol{x}=\boldsymbol{b}, \\ & \boldsymbol{x}\geqslant 0 \end{aligned} \tag{3.2.2}$$

有最优解 $\boldsymbol{x}^*$，则检验数 $\lambda_j=\boldsymbol{c}_B^TB^{-1}\boldsymbol{p}_j-c_j\leqslant 0$，即

$$\boldsymbol{c}_B^TB^{-1}A-\boldsymbol{c}^T\leqslant 0. \tag{3.2.3}$$

令 $\boldsymbol{y}^*=\boldsymbol{\pi}=B^{-T}\boldsymbol{c}_B$，即 $\boldsymbol{y}^{*T}=\boldsymbol{c}_B^TB^{-1}$，因此，

$$A^T\boldsymbol{y}^*\leqslant\boldsymbol{c} \tag{3.2.4}$$

是对偶问题的可行解．并且有

$$b^T y^* = y^{*T} b = c_B^T B^{-1} b = c^T x^*, \qquad (3.2.5)$$

由推论 3.2.1，y^* 是对偶问题的最优解．

由上述定理，可以得到如下结论：

(1) 原始问题和对偶问题均有最优解 x^* 和 y^*，并且 $c^T x^* = b^T y^*$．

(2) 若一个问题无最优解，则另一个问题不可行．

(3) 原始问题和对偶问题都不可行．

它们之间的关系式可由下表表示：

原始问题	对应关系	对偶问题
有最优解	$\longleftrightarrow$	有最优解
无界解	$\longleftrightarrow$	不可行
	⤢	
不可行	$\longleftrightarrow$	无界解

定理 3.2.3　（松驰互补定理）设 x^*、y^* 分别是原始问题 (3.1.1) 与对偶问题 (3.1.2) 的可行解，则 x^*、y^* 为原始问题 (3.1.1) 与对偶问题 (3.1.2) 最优解的充分必要条件是：

$$y^{*T}(Ax^* - b) = 0,$$

$$(c^T - y^{*T}A)x^* = 0. \qquad (3.2.6)$$

证明　因为 x^*、y^* 是可行解，则有

$$b^T y^* = y^{*T} b \leqslant y^{*T} A x^* \leqslant c^T x^*. \qquad (3.2.7)$$

“充分性”，若 x^*、y^* 是最优解，则 $b^T y^* = c^T x^*$，由

(3.2.7) 式，可导出 (3.2.6) 式.

"必要性"，若 (3.2.6) 式成立，显然 $\boldsymbol{c}^T\boldsymbol{x}^* = \boldsymbol{b}^T\boldsymbol{y}^*$，因此，$\boldsymbol{x}^*$、$\boldsymbol{y}^*$ 是最优解.

推论 3.2.1 设 $\boldsymbol{x}^*$、$\boldsymbol{y}^*$ 分别是原始问题 (3.1.3) 与对偶问题 (3.1.4) 的可行解，则 $\boldsymbol{x}^*$、$\boldsymbol{y}^*$ 为原始问题 (3.1.3) 与对偶问题 (3.1.4) 最优解的充分必要条件是：

$$(\boldsymbol{c}^T - \boldsymbol{y}^{*T}A)\boldsymbol{x}^* = 0. \tag{3.2.8}$$

有了松弛互补定理，可以利用对偶问题求出原始问题的最优解.

【例 3.2.1】 写出下列线性规划问题的对偶问题，并用图解法求解.

$$\begin{aligned} \min \quad & f = 2x_1 + 3x_2 + 5x_3 + 2x_4 + 3x_5, \\ \text{s.t.} \quad & x_1 + x_2 + 2x_3 + x_4 + 3x_5 \geqslant 4, \\ & 2x_1 - 2x_2 + 3x_3 + x_4 + x_5 \geqslant 3, \\ & x_1,\ x_2,\ x_3,\ x_4,\ x_5 \geqslant 0. \end{aligned}$$

解 写出对偶问题

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

用图解法求解对偶问题 (见图 3.2.1)，得到 $\boldsymbol{y}^* = \left(\frac{4}{5}, \frac{3}{5}\right)^T$. 在对偶问题中，第 2，3，4 个方程中不等号成立，由松驰互补定理得到，$x_2^* = x_3^* = x_4^* = 0$. 再由于 $y_1^* > 0$，$y_2^* > 0$，所以原始问题中两个约束均等号成立，求解方程组得到 $x_1^* = 1$，$x_5^* = 1$

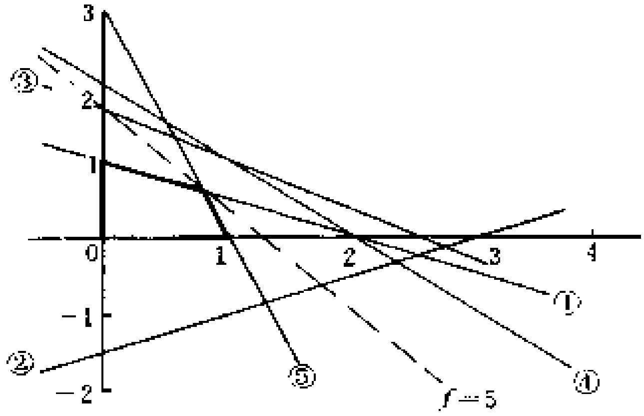


图 3.2.1

和 $f^*=5$.

§3.3 对偶单纯形法

对偶单纯形法是 Lemke（兰姆凯）在 1954 年提出的，单纯形法的基本思想是在可行解集的顶点上进行迭代，而对偶单纯形法的基本思想是在对偶规划的可行解集上进行迭代．对于某些线性规划问题用对偶单纯形法求解要比单纯形法求解简单．

对于线性规划标准形式的对偶问题（3.1.4），若 $\boldsymbol{y}$ 满足 $A^T\boldsymbol{y}\leqslant\boldsymbol{c}$，则称 $\boldsymbol{y}$ 是对偶可行解，称集合 $\{\boldsymbol{y}\mid A^T\boldsymbol{y}\leqslant\boldsymbol{c}\}$ 为对偶可行解集．

设 B 是一个基（不一定是可行基），其检验数为 $\lambda_j=\boldsymbol{c}_B^TB^{-1}\boldsymbol{p}_j-c_j$，若 $\lambda_j\leqslant 0$（$j=1, 2, \cdots, n$），则有

$$\boldsymbol{c}_B^TB^{-1}A-\boldsymbol{c}^T\leqslant 0. \tag{3.3.1}$$

令 $\boldsymbol{y}^T=\boldsymbol{c}_B^TB^{-1}$，则 $\boldsymbol{y}$ 是对偶问题的可行解，此时称 B 是对偶可行基．

定义 3.3.1 若基本解 $\begin{bmatrix}\boldsymbol{x}_B\\ \boldsymbol{x}_N\end{bmatrix}=\begin{bmatrix}B^{-1}\boldsymbol{b}\\ 0\end{bmatrix}$ 对应的检验数 $\lambda_j\leqslant 0$，

$\forall j \in R$，则称该基本解 $\begin{bmatrix} \boldsymbol{x}_B \\ \boldsymbol{x}_N \end{bmatrix} = \begin{bmatrix} B^{-1}\boldsymbol{b} \\ 0 \end{bmatrix}$ 为原始问题的正则解．

考虑单纯形表：

	f	x_1	$\cdots$	x_r	$\cdots$	x_m	$\cdots$	x_j	$\cdots$	x_k	$\cdots$	RHS
f	1	0	$\cdots$	0	$\cdots$	0	$\cdots$	λ_j	$\cdots$	λ_k	$\cdots$	f^0
x_1	0	1	$\cdots$	0	$\cdots$	0	$\cdots$	α_{1j}	$\cdots$	α_{1k}	$\cdots$	β_1
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_r	0	0	$\cdots$	1	$\cdots$	0	$\cdots$	α_{rj}	$\cdots$	α_{rk}	$\cdots$	β_r
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_m	0	0	$\cdots$	0	$\cdots$	1	$\cdots$	α_{mj}	$\cdots$	α_{mk}	$\cdots$	β_m

(3.3.2)

并假设 $\boldsymbol{x}$ 是正则解，即 $\lambda_j \leqslant 0$ $(j \in R)$．

若 $\beta_i \geqslant 0$，$i=1, 2, \cdots, m$，则 $\boldsymbol{x}$ 是基本可行解．由最优性条件，$\boldsymbol{x}$ 是最优解．

若存在 $\beta_r < 0$，则在第 r 行选择元素 α_{rk}，并以它为中心作转轴运算，使得到的新的 β'_r，并满足 $\beta'_r > 0$．因此，要求 $\alpha_{rk} < 0$．

若 $\forall j \in R$，$\alpha_{rj} \geqslant 0$，由单纯形表（3.3.2）得到

$$x_r = \beta_r - \sum_{j \in R} \alpha_{rj} x_j \leqslant \beta_r < 0 \tag{3.3.3}$$

与 x_r 的非负性矛盾．因此线性规划问题无可行解．

除要求 $\alpha_{rk} < 0$ 之外，α_{rk} 还应满足

$$\frac{\lambda_k}{\alpha_{rk}} = \min\left\{\frac{\lambda_j}{\alpha_{rj}} \mid \alpha_{rj} < 0, j \in R\right\}, \tag{3.3.4}$$

这是因为需要保证在转轴之后，新的检验数 λ'_j 满足

$$\lambda'_j = \lambda_j - \frac{\lambda_k}{\alpha_{rk}} \alpha_{rj} \leqslant 0, \tag{3.3.5}$$

事实上，当 $\alpha_{rj}\geqslant 0$ 时，（3.3.5）式显然成立. 当 $\alpha_{rj}<0$ 时，（3.3.5）式等价于

$$\frac{\lambda_j}{\alpha_{rj}}\geqslant\frac{\lambda_k}{\alpha_{rk}}, \tag{3.3.6}$$

因此，（3.3.4）式保证了（3.3.5）式成立.

算法 3.3.1 （从正则解开始的对偶单纯形法）

（1）选取初始对偶可行基 B（即 $\boldsymbol{x}$ 是正则解），列出初始单纯形表：

	f	$\boldsymbol{x}_B$	$\boldsymbol{x}_N$	RHS
f	1	0^T	$\boldsymbol{c}_B^TB^{-1}N-\boldsymbol{c}_N^T$	$\boldsymbol{c}_B^TB^{-1}\boldsymbol{b}$
$\boldsymbol{x}_B$	0	I	$B^{-1}N$	$B^{-1}\boldsymbol{b}$

（2）计算

$$\beta_r=\min\{\beta_i\mid i=1,2,\cdots,m\},$$

若 $\beta_r\geqslant 0$，则停止计算$\left(\boldsymbol{x}=\begin{bmatrix}B^{-1}\boldsymbol{b}\\0\end{bmatrix}\text{为最优解}\right)$.

（3）若 $\alpha_{rj}\geqslant 0$，$j\in R$，则停止计算（此时线性规划问题无可行解）；否则确定 k，使得

$$\frac{\lambda_k}{\alpha_{rk}}=\min\left\{\frac{\lambda_j}{\alpha_{rj}}\mid \alpha_{rj}<0,j\in R\right\},$$

然后以 α_{rk} 为中心进行转轴运算，即置

$$\alpha_{rj}=\frac{\alpha_{rj}}{\alpha_{rk}},\quad j=1,2,\cdots,n,\ j\neq k,$$

$$\beta_r=\frac{\beta_r}{\alpha_{rk}},\quad \alpha_{rk}=1,$$

$$\lambda_j=\lambda_j-\lambda_k\alpha_{rj},\quad j=1,2,\cdots,n,$$

$$\alpha_{ij}=\alpha_{ij}-\alpha_{ik}\alpha_{rj},\quad i=1,2,\cdots,m,i\neq r,j=1,j=1,2,\cdots,n,$$

$$\beta_i=\beta_i-\alpha_{ik}\beta_r,\quad i=1,2,\cdots,m,\ i\neq r,$$

$$f = f - \lambda_k \beta_r,$$

然后转（2）.

【例 3.3.1】 用对偶单纯形法求解线性规划问题：

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

解 引进松弛变量 x_4、x_5，化为标准形式

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

取 $B =$ （$\boldsymbol{p}_4$，$\boldsymbol{p}_5$） $= I$，列表计算：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	-2	-3	-4	0	0	0
x_4	0	-1	-2	-1	1	0	-3
x_5	0	$\boxed{-2}$	1	-3	0	1	-4^*

$\beta_2 = \min\{-3,\ -4\} = -4$. 选择 $\alpha_{21} = -2$ 为中心作转轴运算，得到：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	0	-4	1	0	-1	4
x_4	0	0	$\boxed{-\frac{5}{2}}$	$\frac{1}{2}$	1	$\frac{1}{2}$	-1^*
x_1	0	1	$-\frac{1}{2}$	$\frac{3}{2}$	0	$-\frac{1}{2}$	2

$\beta_1 = \min\{-1, 2\} = -1$. 作转轴运算得到：

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	0	0	$-\frac{9}{5}$	$-\frac{8}{5}$	$-\frac{1}{5}$	$\frac{28}{5}$
x_2	0	0	1	$-\frac{1}{5}$	$-\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$
x_1	0	1	0	$\frac{7}{5}$	$-\frac{1}{5}$	$\frac{2}{5}$	$\frac{11}{5}$

得到最优解 $x^* = \left(\frac{11}{5}, \frac{1}{5}, 0, 0, 0\right)^T$, $f^* = \frac{28}{5}$.

【例 3.3.2】 用对偶单纯形法求解线性规划问题：

$$\begin{aligned} \min\quad & x_1 + 2x_2, \\ \text{s.t.}\quad & x_1 + x_2 \leqslant 4, \\ & 2x_1 + 3x_2 \geqslant 18, \\ & x_1, x_2 \geqslant 0. \end{aligned}$$

解 引进松弛变量 x_3, x_4，化为标准形式

$$\begin{aligned} \min\quad & x_1 + 2x_2, \\ \text{s.t.}\quad & x_1 + x_2 + x_3 = 4, \\ & -2x_1 - 3x_2 + x_4 = -18, \\ & x_1, x_2, \cdots, x_4 \geqslant 0. \end{aligned}$$

计算过程如下：

	f	x_1	x_2	x_3	x_4	RHS
f	1	-1	-2	0	0	0
x_3	0	1	1	1	0	4
x_4	0	$\boxed{-2}$	-3	0	1	-18^*

	f	x_1	x_2	x_3	x_4	RHS
f	1	0	$-\frac{1}{2}$	0	$-\frac{1}{2}$	9
x_3	0	0	$\boxed{-\frac{1}{2}}$	1	$\frac{1}{2}$	-5^*
x_1	0	1	$\frac{3}{2}$	0	$-\frac{1}{2}$	9

	f	x_1	x_2	x_3	x_4	RHS
f	1	0	0	-1	-1	14
x_2	0	0	1	-2	-1	10
x_1	0	1	0	3	1	-6*

此时，$a_{2j}\geqslant 0$，$j\in R=\{3,4\}$，所以线性规划问题无可行解.

§3.4 第一个正则解的求法

在有些情况下，有明显的正则解，例如

$$\begin{aligned} \min \quad & c^T x,(c \geqslant 0), \\ \text{s.t.} \quad & Ax \leqslant b,(\text{若} \geqslant \text{两端同乘} -1) \\ & x \geqslant 0. \end{aligned} \tag{3.4.1}$$

引进松驰变量 x_s，将问题（3.4.1）化成标准形式

$$\begin{aligned} \min \quad & c^T x, \\ \text{s.t.} \quad & Ax + x_s = b, \\ & x\geqslant 0. \end{aligned} \tag{3.4.2}$$

此时，$x=0$、$x_s=b$ 是问题（3.4.2）的一个正则解.

在一般情况下，若已知问题的一个基本解，不妨设其典式为

$$\begin{aligned} \min \quad & f = f_0 - \sum_{j\in R}\lambda_j x_j, \\ \text{s.t.} \quad & x_i = \beta_i - \sum_{j\in R}\alpha_{ij}x_j, \\ & x_j\geqslant 0,\ j=1,2,\cdots,n, \end{aligned} \tag{3.4.3}$$

若 $\lambda_j\leqslant 0$，$\forall j\in R$，则该基本解为正则解. 若存在 $\lambda_j>0$，$j\in R$，则该基本解不是正则解. 在这种情况下，增加一个变量 x_0 和一个约束

$$x_0 + \sum_{j \in R} x_j = M, \tag{3.4.4}$$

其中 M 是一个充分大的数．考虑问题

$$\begin{aligned}
\min \quad & f = f_0 - \sum_{j \in R} \lambda_j x_j, \\
\text{s.t.} \quad & x_0 + \sum_{j \in R} x_j = M, \\
& x_i = \beta_i - \sum_{j \in R} \alpha_{ij} x_j, \\
& x_j \geqslant 0, \ j = 1, \ 2, \ \cdots, \ n,
\end{aligned} \tag{3.4.5}$$

称问题（3.4.5）为问题（3.4.3）的扩充问题．不妨设问题（3.4.5）的单纯形表为：

	f	x_0	x_1	$\cdots$	x_r	$\cdots$	x_m	$\cdots$	x_j	$\cdots$	x_k	$\cdots$	RHS
f	1	0	0	$\cdots$	0	$\cdots$	0	$\cdots$	λ_j	$\cdots$	λ_k	$\cdots$	f^0
x_0	0	1	0	$\cdots$	0	$\cdots$	0	$\cdots$	1	$\cdots$	$\boxed{1}$	$\cdots$	M
x_1	0	0	1	$\cdots$	0	$\cdots$	0	$\cdots$	α_{1j}	$\cdots$	α_{1k}	$\cdots$	β_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_r	0	0	0	$\cdots$	1	$\cdots$	0	$\cdots$	α_{rj}	$\cdots$	α_{rk}	$\cdots$	β_r
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_m	0	0	0	$\cdots$	0	$\cdots$	1	$\cdots$	α_{mj}	$\cdots$	α_{mk}	$\cdots$	β_m

(3.4.6)

令 $\lambda_k = \max\{\lambda_j \mid j \in R\} > 0$，以第 0 行第 k 列的 1 作为主元进行转轴运算，得到新的单纯形表：

	f	x_0	x_1	$\cdots$	x_r	$\cdots$	x_m	$\cdots$	x_j	$\cdots$	x_k	$\cdots$	RHS
f	1	$-\lambda_k$	0	$\cdots$	0	$\cdots$	0	$\cdots$	$\lambda_j-\lambda_k$	$\cdots$	0	$\cdots$	$f^0-\lambda_k M$
x_k	0	1	0	$\cdots$	0	$\cdots$	0	$\cdots$	1	$\cdots$	1	$\cdots$	M
x_1	0	$-\alpha_{1k}$	1	$\cdots$	0	$\cdots$	0	$\cdots$	$\alpha_{1j}-\alpha_{1k}$	$\cdots$	0	$\cdots$	$\beta_1-\alpha_{1k}M$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_r	0	$-\alpha_{rk}$	0	$\cdots$	1	$\cdots$	0	$\cdots$	$\alpha_{rj}-\alpha_{rk}$	$\cdots$	0	$\cdots$	$\beta_r-\alpha_{rk}M$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_m	0	$-\alpha_{mk}$	0	$\cdots$	0	$\cdots$	1	$\cdots$	$\alpha_{mj}-\alpha_{mk}$	$\cdots$	0	$\cdots$	$\beta_m-\alpha_{mk}M$

(3.4.7)

因为 $\lambda_j-\lambda_k\leqslant 0$，$j\in R$，$j\neq k$，$-\lambda_k\leqslant 0$，所以得到扩充问题(3.4.5)的正则解.

【例 3.4.1】 考虑如下线性规划问题：

$$\begin{aligned}
\min\quad & -x_4+x_5,\\
\text{s.t.}\quad & x_1-x_4+4x_5=-5,\\
& x_2+x_4-3x_5=1,\\
& x_3-2x_4+5x_5=-1,\\
& x_1,\ x_2,\ \cdots,\ x_5\geqslant 0.
\end{aligned}$$

求它的扩充问题的正则解.

解 该问题没有明显的正则解，增加一个变量 x_0 和一个约束

$$x_0+x_4+x_5=M,$$

得到扩充问题，列出单纯形表：

	f	x_0	x_1	x_2	x_3	x_4	x_5	RHS
f	1	0	0	0	0	1^*	-1	0
x_0	0	1	0	0	0	$\boxed{1}$	1	M
x_1	0	0	1	0	0	-1	4	-5
x_2	0	0	0	1	0	1	-3	1
x_3	0	0	0	0	1	-2	5	-1

$\lambda_4=1>0$，作转轴运算，得到

	f	x_0	x_1	x_2	x_3	x_4	x_5	RHS
f	1	-1	0	0	0	0	-2	$-M$
x_4	0	1	0	0	0	1	1	M
x_1	0	1	1	0	0	0	5	$M-5$
x_2	0	-1	0	1	0	0	-4	$-M+1$
x_3	0	2	0	0	1	0	7	$2M-1$

为扩充问题的正则解.

对于扩充问题，可以得到如下结论：

(1) 扩充问题无可行解，则原问题无可行解；

(2) 扩充问题有最优解 $(x_0^*,\ \boldsymbol{x}^*)$，则 $\boldsymbol{x}^*$ 是原问题的可行解. 如果将 $\boldsymbol{x}^*$ 代入目标函数 $f(\boldsymbol{x})$ 中，$f(\boldsymbol{x}^*)$ 与 M 无关，则 $\boldsymbol{x}^*$ 是原问题的最优解；否则原问题无有限最优解.

受篇幅限制，这里就不再证明和举例了.

习 题 三

3.1 给出下面问题的对偶问题：

$$\begin{aligned}
\min\quad & 2x_1+3x_2-5x_3,\\
\text{s.t.}\quad & x_1+x_2-x_3+x_4\geqslant 5,\\
& 2x_1+x_3\leqslant 4,\\
& x_2+x_3+x_4=6,\\
& x_1\leqslant 0,\\
& x_2,\ x_3\geqslant 0,\\
& x_4\ \text{无限制}.
\end{aligned}$$

3.2 构造一个例子，其原始问题和对偶问题均没有可行解，用以说明：一个问题不可行并不意味着另一个问题无界.

3.3 考虑线性规划问题

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

其中 $m=n$，$\boldsymbol{b}=\boldsymbol{c}$ 和 $A^T=A$. 证明：如果存在 $\bar{\boldsymbol{x}}$，使得 $A\bar{\boldsymbol{x}}=\boldsymbol{b}$，则 $\bar{\boldsymbol{x}}$ 是最优解.

3.4 用对偶理论证明：如果线性规划问题

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

有最优解，则线性规划问题

$$\begin{aligned}
\min\quad & \boldsymbol{c}^T\boldsymbol{x},\\
\text{s.t.}\quad & A\boldsymbol{x}=\boldsymbol{b}',\\
& \boldsymbol{x}\geqslant 0
\end{aligned}$$

不可能无界，无论 b' 取何值.

3.5 用对偶理论证明 Farkas 定理：设

系统 Ⅰ： $\quad A\boldsymbol{x}=\boldsymbol{b}, \boldsymbol{x}\geqslant 0;$

系统 Ⅱ： $\quad A^T\boldsymbol{y}\leqslant 0, \boldsymbol{b}^T\boldsymbol{y}>0,$

则两系统有且仅有一个有解.

提示：考虑线性规划问题与它的对偶问题

$$\begin{array}{llcll} \min & 0^T\boldsymbol{x}, & & \min & \boldsymbol{b}^T\boldsymbol{y}, \\ \text{s.t.} & A\boldsymbol{x}=\boldsymbol{b}, & \text{和} & \text{s.t.} & A^T\boldsymbol{y}\leqslant 0, \\ & \boldsymbol{x}\geqslant 0. & & & \boldsymbol{y}\ \text{无限制}. \end{array}$$

3.6 考虑线性规划问题

$$\begin{array}{ll} \max & 10x_1+24x_2+20x_3+20x_4+25x_5, \\ \text{s.t.} & x_1+x_2+2x_3+3x_4+5x_5\leqslant 19, \\ & 2x_1+4x_2+3x_3+2x_4+x_5\leqslant 57, \\ & x_1,\ x_2,\ x_3,\ x_4,\ x_5\geqslant 0. \end{array}$$

(1) 写出相应的对偶问题，并验证 $\boldsymbol{y}=(4,\ 5)^T$ 是对偶问题的可行解.

(2) 用 (1) 的信息求出原始问题和对偶问题的最优解.

3.7 写出下列线性规划的对偶问题并用图解法求出它们的解：

(1)
$$\begin{array}{ll} \min & f=x_1+3x_2+5x_3+6x_4, \\ \text{s.t.} & x_1+2x_2+3x_3+x_4\geqslant 2, \\ & -2x_1+x_2-x_3+3x_4\leqslant -3, \\ & x_1,\ x_2,\ x_3,\ x_4\geqslant 0. \end{array}$$

(2)
$$\begin{array}{ll} \max & f=3x_1+x_2+4x_3, \\ \text{s.t.} & 6x_1+3x_2+5x_3\leqslant 25, \\ & 3x_1+4x_2+5x_3\leqslant 20, \\ & x_1,\ x_2,\ x_3\geqslant 0. \end{array}$$

3.8 下面的单纯形表是某个线性规划问题的最优解，其中x_4、x_5是松驰变量，约束的类型为“$\leqslant$”.

	f	x_1	x_2	x_3	x_4	x_5	RHS
f	1	0	-4	0	-4	-2	-40
x_3	0	0	$\frac{1}{2}$	1	$\frac{1}{2}$	0	$\frac{5}{2}$
x_1	0	1	$-\frac{1}{2}$	0	$-\frac{1}{6}$	$\frac{1}{3}$	$\frac{5}{2}$

(1) 写出原问题；

(2) 写出原问题的对偶问题；

(3) 由上表得到对偶问题的最优解.

3.9 用对偶单纯形法求解下列线性规划问题：

(1)
$$\begin{aligned}\min\quad & f=x_1+3x_2+5x_3+6x_4,\\ \text{s.t.}\quad & x_1+2x_2+3x_3+x_4\geqslant 2,\\ & -2x_1+x_2-x_3+3x_4\leqslant -3,\\ & x_1,\ x_2,\ x_3,\ x_4\geqslant 0.\end{aligned}$$

(2)
$$\begin{aligned}\max\quad & f=-4x_1-6x_2-18x_3,\\ \text{s.t.}\quad & x_1+3x_3\geqslant 3,\\ & x_2+2x_3\geqslant 5,\\ & x_1,\ x_2,\ x_3\geqslant 0.\end{aligned}$$

(3)
$$\begin{aligned}\min\quad & f=-x_1+x_2+x_3,\\ \text{s.t.}\quad & -x_1+x_2-2x_3\geqslant 1,\\ & -x_1-2x_2+x_3\geqslant 1,\\ & x_1,\ x_2,\ x_3\geqslant 0.\end{aligned}$$

(4)
$$\begin{aligned}\min\quad & f=3x_1+5x_2-x_3+2x_4-4x_5,\\ \text{s.t.}\quad & x_1+x_2+x_3+3x_4+x_5\geqslant 6,\end{aligned}$$

$$-x_1-x_2+2x_3+x_4-x_5\geqslant 3,$$
$$x_1,\ x_2,\ x_3,\ x_4,\ x_5\geqslant 0.$$

3.10 证明：若在对偶单纯形表中，$\exists\beta_r<0$，且 $\alpha_{rj}\geqslant 0$，$\forall j\in R$，则对应的对偶问题无上界.

3.11 证明：若扩充问题无可行解，则原问题无可行解.

第四章　无约束最优化问题的一般结构

将在第四章、第五章、第六章和第七章介绍无约束最优化问题

$$\min f(\boldsymbol{x}),\ \boldsymbol{x} \in R^n \tag{4.0.1}$$

的求解方法．在本章主要讨论无约束问题的最优性条件、下降算法及算法收敛性等概念．

§4.1　无约束问题的最优性条件

4.1.1　无约束问题的局部解

第一章介绍了无约束问题的全局解，但全局解实际上是非常难找到的，这里介绍无约束问题局部解的概念．

定义 4.1.1　若存在 $\boldsymbol{x}^* \in R^n$，$\varepsilon > 0$，$\forall \boldsymbol{x} \in R^n$，使当 $\|\boldsymbol{x}-\boldsymbol{x}^*\| < \varepsilon$ 时，恒有

$$f(\boldsymbol{x}) \geqslant f(\boldsymbol{x}^*), \tag{4.1.1}$$

则称 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 的**局部极小点**（local minimum point），或称 $\boldsymbol{x}^*$ 是无约束问题的**局部解**（local minimum solution）．若对 $\boldsymbol{x} \in R^n$，使当 $0 < \|\boldsymbol{x}-\boldsymbol{x}^*\| < \varepsilon$ 时，恒有

$$f(\boldsymbol{x}) > f(\boldsymbol{x}^*), \tag{4.1.2}$$

则称 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 的严格局部极小点，或称 $\boldsymbol{x}^*$ 是无约束问题的严格局部解．

局部解也称为无约束问题的最优解．

【例 4.1.1】　考察无约束问题

$$\min f(\boldsymbol{x}) = x_1^4 - 2x_1^2x_2 + x_1^2 - 2x_1x_2 + 2x_2^2 + \frac{9}{2}x_1 - 4x_2 + 4$$

的局部解.

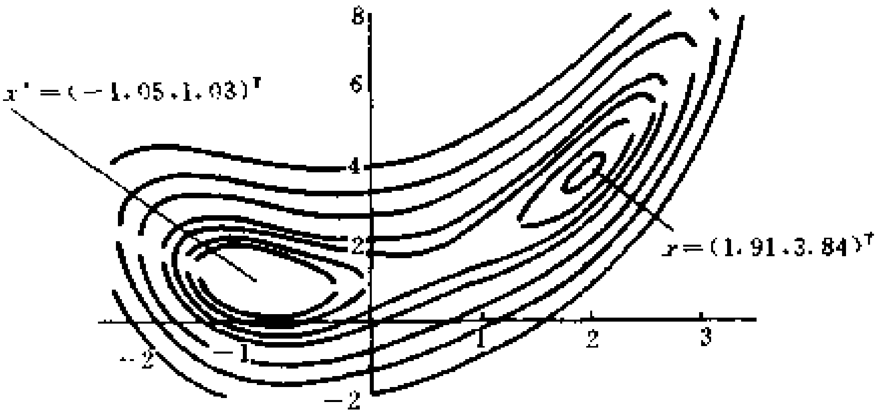


图 4.1.1　局部解的几何意义

该问题的目标函数的等高线如图 4.1.1 所示. 由于 $f(\boldsymbol{x}^*)=0.51$, $f(\bar{\boldsymbol{x}})=0.99$, 所以 $\boldsymbol{x}^*$ 是全局极小点, 但就局部范围来说, $\boldsymbol{x}^*$ 和 $\bar{\boldsymbol{x}}$ 均是局部极小点, 即问题的局部解.

用定义 4.1.1 检验无约束问题的局部解是十分困难的, 因此需要给出它的必要和充分条件.

4.1.2　无约束问题的局部解的必要条件

首先给出无约束问题局部解的**一阶必要条件** (first-order necessary condition).

定理 4.1.1　(无约束问题局部解的一阶必要条件) 设 $f(x)$ 具有连续的一阶偏导数, 若 $\boldsymbol{x}^*$ 是无约束问题 (4.0.1) 的局部解, 则

$$\nabla f(\boldsymbol{x}^*) = 0. \tag{4.1.3}$$

证明　对任意的 $\boldsymbol{d} \neq 0$, $\boldsymbol{d} \in R^n$, 构造单变量函数

$$\phi(\alpha) = f(\boldsymbol{x}^* + \alpha \boldsymbol{d}), \tag{4.1.4}$$

因为 x^* 是无约束问题的局部解，因此，$\alpha=0$ 是一元函数 $\phi(\alpha)$ 的局部极小点，由一元函数极小点的必要条件，有

$$0=\phi'(0)=\langle\nabla f(x^*),d\rangle=d^T\nabla f(x^*),$$

$$\forall d\neq 0, d\in R^n, \tag{4.1.5}$$

由 d 的任意性，所以，(4.1.3) 式成立.

一阶必要条件表明，在无约束问题的局部解 x^* 处，任何方向上的方向导数均为 0，即在 x^* 处的切平面是水平的.

定理 4.1.1 的逆命题不成立，即梯度为 0 的点不一定是局部解. 这种情况由一元函数就可以得到验证. 但此类点也是很重要的点，我们称梯度为 0 的点为**稳定点**.

定理 4.1.2 （无约束问题局部解的**二阶必要条件**（second-order necessary condition)）设 $f(x)$ 具有连续的二阶偏导数，若 x^* 是无约束问题 (4.0.1) 的局部解，则

(1) $$\nabla f(x^*)=0; \tag{4.1.6}$$

(2) $$\nabla^2 f(x^*) \text{ 半正定}. \tag{4.1.7}$$

证明 只需证明结论 (2). 对于任意的 $d\neq 0$，$d\in R^n$，由于 x^* 是局部解，则存在 $\epsilon>0$，使当 $0<\alpha<\epsilon$ 时，有

$$f(x^*+\alpha d)\geqslant f(x^*). \tag{4.1.8}$$

另一方面，由二阶 Taylor 展式

$$f(x^*+\alpha d)=f(x^*)+\alpha d^T\nabla f(x^*)$$
$$+\frac{1}{2}\alpha^2 d^T\nabla^2 f(x^*)d+o(\alpha^2), \tag{4.1.9}$$

并结合 (4.1.6) 式和 (4.1.8) 式得到

$$\frac{1}{2}\alpha^2 d^T\nabla^2 f(x^*)d+o(\alpha^2)\geqslant 0, \tag{4.1.10}$$

在 (4.1.10) 式两端同除 α^2，并令 $\alpha\to 0+$，得到

$$d^T\nabla^2 f(x^*)d\geqslant 0, \forall d\in R^n, \tag{4.1.11}$$

即 $\nabla^2 f(x^*)$ 半正定.

定理 4.1.2 表明，在局部解 x^* 处的二阶方向导数是正的．下面讨论无约束问题的局部解的充分条件．

4.1.3 无约束问题的局部解的充分条件

定理 4.1.3 （无约束问题局部解的**二阶充分条件**（second-order sufficient condition））设 $f(\boldsymbol{x})$ 具有连续的二阶偏导数，若在 x^* 处满足：

(1) $$\nabla f(\boldsymbol{x}^*) = 0 \tag{4.1.12}$$

(2) $$\nabla^2 f(\boldsymbol{x}^*) \text{ 正定}, \tag{4.1.13}$$

则 x^* 是无约束问题（4.0.1）的严格局部解．

证明 对于 $\boldsymbol{x}^* \in R^n$，$\boldsymbol{x} \neq \boldsymbol{x}^*$，令 $\boldsymbol{d} = \dfrac{\boldsymbol{x}-\boldsymbol{x}^*}{\|\boldsymbol{x}-\boldsymbol{x}^*\|}$，所以 $\|\boldsymbol{d}\| = 1$．那么 $\boldsymbol{x} = \boldsymbol{x}^* + \alpha\boldsymbol{d}$，$\alpha = \|\boldsymbol{x}-\boldsymbol{x}^*\|$，由二阶 Taylor 展开式

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha\boldsymbol{d}^T\nabla f(\boldsymbol{x}^*) + \frac{1}{2}\alpha^2\boldsymbol{d}^T\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} + o(\alpha^2) \tag{4.1.14}$$

和（4.1.12）式，得到

$$f(x) - f(x^*) = \frac{1}{2}\alpha^2\boldsymbol{d}^T\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} + o(\alpha^2), \tag{4.1.15}$$

只需证明，当 $0<\alpha<\epsilon$ 时，（4.1.15）式右端大于 0．由于函数 $F(\boldsymbol{d}) = \boldsymbol{d}^T\nabla^2 f(x^*)\boldsymbol{d}$ 是闭球 $\{\boldsymbol{d} \mid \|\boldsymbol{d}\| = 1\}$ 上的连续函数，因此存在 $r>0$，使得

$$\boldsymbol{d}^T\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} \geqslant r, \tag{4.1.16}$$

因此，

$$\frac{1}{2}\boldsymbol{d}^T\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} + \frac{o(\alpha^2)}{\alpha^2} \geqslant \frac{1}{2}r + \frac{o(\alpha^2)}{\alpha^2} > 0, \tag{4.1.17}$$

由于$\frac{o(\alpha^2)}{\alpha^2}\to 0$，$(\alpha\to 0)$ $r>0$. 因此存在 $\varepsilon>0$，当 $0<\alpha<\varepsilon$ 时，(4.1.17) 式成立. 即 (4.1.15) 式右端大于 0. 因此，x^* 是严格局部解.

思考：为什么不直接说 (4.1.15) 式右端大于 0.

下面引进一个新概念.

定义 4.1.2　若 G 是 $n\times n$ 阶正定对称矩阵，称函数

$$f(x)=\frac{1}{2}\boldsymbol{x}^TG\boldsymbol{x}+\boldsymbol{r}^T\boldsymbol{x}+\delta \tag{4.1.18}$$

为正定二次函数.

【例 4.1.2】 证明：若目标函数为正定二次函数，则相应的无约束问题有严格局部解，且此局部解也是全局解.

证明　由 (4.1.18) 式得到

$$\nabla f(\boldsymbol{x})=G\boldsymbol{x}+\boldsymbol{r}, \tag{4.1.19}$$

则方程 $G\boldsymbol{x}+\boldsymbol{r}=0$ 的解 $\boldsymbol{x}^*=-G^{-1}\boldsymbol{r}$ 存在且唯一. 并且 $\nabla^2 f(\boldsymbol{x})=G$ 正定. 由定理 4.1.3，$\boldsymbol{x}^*$ 是严格局部解. 下面证明 $\boldsymbol{x}^*$ 是全局解，由 Taylor 展开式

$$\begin{aligned}f(\boldsymbol{x})&=f(\boldsymbol{x}^*)+(\boldsymbol{x}-\boldsymbol{x}^*)^T\nabla f(\boldsymbol{x}^*)\\&\quad+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^T\nabla^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)\\&=f(\boldsymbol{x}^*)+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^TG(\boldsymbol{x}-\boldsymbol{x}^*),\end{aligned} \tag{4.1.20}$$

且 G 是正定对称矩阵，因此

$$f(\boldsymbol{x})-f(\boldsymbol{x}^*)=\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^TG(\boldsymbol{x}-\boldsymbol{x}^*)\geqslant 0,\forall \boldsymbol{x}\in R^n, \tag{4.1.21}$$

所以 $\boldsymbol{x}^*$ 是全局解.

在今后所讲的算法中，从理论上讲，往往只能保证所得到的点满足一阶必要条件，不能保证满足二阶充分条件. 换句话说，

我们只能得到稳定点，只有在一定的条件下（如 $f(x)$是凸函数），稳定点才是无约束问题的局部解（或全局解）.

定理 4.1.4　设目标函数$f(x)$是连续可微的凸函数，则 x^* 是无约束问题的全局解的充分必要条件是：

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

推论　设$f(x)$是连续可微的凸函数，则无约束问题的任一局部解均是全局解.

定理及推论的证明留作习题.

§4.2　无约束问题的一般下降算法

4.2.1　最速下降法

先考查二维情况．设其目标函数的等高线如图 4.2.1 所示．如何来求最优点 x^*呢？很自然地想法是从某一点出发，以“最快”的速度达到极小点 x^*．那么，什么方向是函数下降最快的方向呢？由第一章的预备知识可知，是最速下降方向，即负梯度方向.

设 $x = x(t)$为点 x 所要走的曲线，其中 t 为参数，那么

$$\frac{dx(t)}{dt} = \left(\frac{dx_1(t)}{dt}, \frac{dx_2(t)}{dt}, \cdots, \frac{dx_n(t)}{dt}\right)^T$$

表示曲线 $x = x(t)$ 的切线方向，因此有如下关系式

$$\begin{cases} \dfrac{dx(t)}{dt} = -\nabla f(x(t)), \\ x(t_1) = x^{(1)}, \end{cases} \tag{4.2.1}$$

这里 t_1 是初始时刻，$x^{(1)}$是初始位置.

问题（4.2.1）是一个一阶非线性常微分方程组．可以证明：问题（4.2.1）的解$x(t)$存在，且当 $t\to\infty$时，有 $x(t)\to x^*$，即得到无约束问题(4.0.1)的最优解.

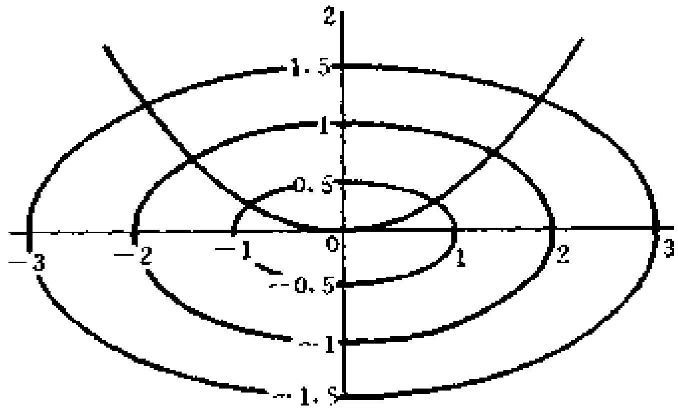


图 4.2.1

看一个简单的例子.

【例 4.2.1】 按上述方法求解无约束问题

$$\min f(\boldsymbol{x}) = x_1^2 + 2x_2^2.$$

解 $\nabla f(\boldsymbol{x})=(2x_1,4x_2)^T$,由(4.2.1)式得到

$$\begin{cases}\dfrac{\mathrm{d}x_1}{\mathrm{d}t} = -2x_1,\\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = -4x_2\end{cases}$$

解微分方程组得到

$$\begin{cases}x_1(t) = c_1e^{-2t},\\ x_2(t) = c_2e^{-4t},\end{cases}$$

当 $t\to+\infty$时, $\boldsymbol{x}(t)\to(0,0)^T=\boldsymbol{x}^*$.

$\boldsymbol{x}(t)$趋于 $\boldsymbol{x}^*$的路径是一条抛物线

$$x_2(t) = cx_1(t)^2.$$

事实上，微分方程组（4.2.1）的求解并不像例 4.2.1 那么方便，有时则无法得到解析表达式，因此，求解方程组（4.2.1）的方法并不是一个行之有效的方法.

下面探讨有效的方法. 一种比较直观的方法是将上述方法得到的曲线离散化，即在 $\boldsymbol{x}^{(1)}$处，计算梯度$\nabla f(\boldsymbol{x}^{(1)})$，这样得到

方向 $d^{(1)}=-\nabla f(x^{(1)})$，沿 $d^{(1)}$方向前进，得到 $x^{(2)}$，再计算 $\nabla f(x^{(2)})$，令 $d^{(2)}=-\nabla f(x^{(2)})$，沿 $d^{(2)}$方向前进，得到 $x^{(3)}$，如此下去……．具体过程见图 4.2.2．

所谓沿 $d^{(k)}$方向前进得到 $x^{(k+1)}$，是指求一元函数极小，即求解一维问题

$$\min_{\alpha}\phi(\alpha)=f(x^{(k)}+\alpha d^{(k)}), \tag{4.2.2}$$

得到最优步长 α_k，置

$$x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}, \tag{4.2.3}$$

上述过程称为一维搜索，称 $d^{(k)}$为搜索方向．若 α_k 为$\phi(\alpha)$的精确极小点，则称一维搜索是精确的．

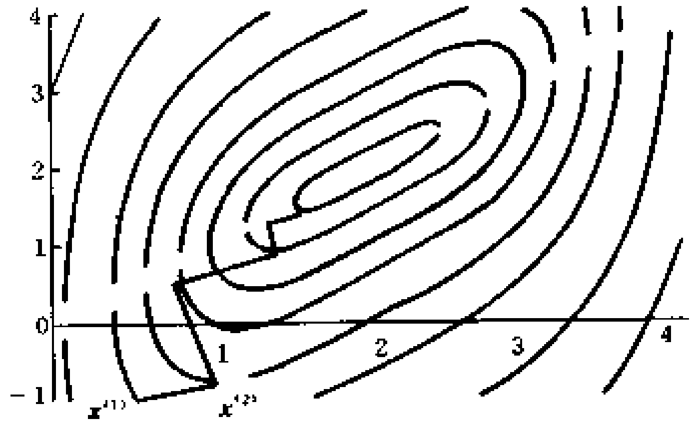


图 4.2.2　最速下降法的几何意义

这样就得到相应的算法．

算法 4.2.1（最速下降法）

（1）取初始点 $x^{(1)}$，置 $k=1$．

（2）若$\nabla f(x^{(k)})=0$，则停止计算（$x^{(k)}$为无约束问题的最优解）；否则置

$$d^{(k)}=-\nabla f(x^{(k)}).$$

（3）一维搜索．求解一维问题

$$\min_{\alpha}\phi(\alpha)=f(x^{(k)}+\alpha d^{(k)}),$$

得 α_k，置

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}.$$

(4) 置 $k=k+1$，转 (2).

【例 4.2.2】 用最速下降法求解无约束问题

$$\min f(x) = \frac{x_1^2}{a} + \frac{x_2^2}{b}, \quad (a>0, b>0) \qquad (4.2.4)$$

取初始点 $x^{(1)} = (a, b)^T$.

解 由于

$$\nabla f(x) = \left(\frac{2}{a}x_1, \frac{2}{b}x_2\right)^T,$$

因此，$\nabla f(x^{(1)}) = (2, 2)^T$，取 $d^{(1)} = -\nabla f(x^{(1)}) = (-2, -2)^T$，作一维搜索，构造函数

$$\phi(\alpha) = f(x^{(1)} + \alpha d^{(1)}) = \frac{(a-2\alpha)^2}{a} + \frac{(b-2\alpha)^2}{b},$$

对 $\phi(\alpha)$ 求导数，得到

$$\phi'(\alpha) = -4\left(1-\frac{2}{a}\alpha\right) - 4\left(1-\frac{2}{b}\alpha\right),$$

解方程 $\phi'(\alpha)=0$，得最优步长

$$\alpha_1 = \frac{ab}{a+b},$$

故

$$\begin{aligned} x^{(2)} &= x^{(1)} + \alpha_1 d^{(1)} \\ &= (a, b)^T + \frac{ab}{a+b}(-2, -2)^T \\ &= \left(\frac{a(a-b)}{a+b}, \frac{b(b-a)}{a+b}\right)^T, \end{aligned}$$

若 $a=b$，则 $x^{(2)} = (0, 0)^T$，$\nabla f(x^{(2)}) = (0, 0)^T$，停止计算，$x^{(2)}$ 为无约束问题的最优解.

若 $a \neq b$，则再进行一次迭代，得到

$$x^{(3)} = \left(\frac{a(a-b)^2}{(a+b)^2}, \frac{b(b-a)^2}{(a+b)^2}\right)^T,$$

如此下去，可得到

$$x^{(k+1)} = \left(\frac{a(a-b)^k}{(a+b)^k}, \frac{b(b-a)^k}{(a+b)^k}\right)^T, \qquad (4.2.5)$$

当 $k\to\infty$时，$x^{(k)}\to (0,0)^T$，得到无约束问题（4.2.4）的最优解.

4.2.2 一般下降算法

上一小节讨论了最速下降法，现对最速下降法进行分析.

定理 4.2.1 若一维搜索是精确的，则最速下降法产生的相邻两次的搜索方向是相互正交的，即

$$\langle d^{(k+1)}, d^{(k)}\rangle = 0,\ k = 1,2,\cdots,$$

证明 令 $\phi(\alpha) = f(x^{(k)} + \alpha d^{(k)})$，所以 $\phi'(\alpha) = \langle \nabla f(x^{(k)} + \alpha d^{(k)}), d^{(k)}\rangle$. 由于一维搜索是精确的，故有

$$\begin{aligned} 0 = \phi'(\alpha_k) &= \langle \nabla f(x^{(k)} + \alpha_k d^{(k)}), d^{(k)}\rangle \\ &= \langle \nabla f(x^{(k+1)}), d^{(k)}\rangle \\ &= -\langle d^{(k+1)}, d^{(k)}\rangle, \end{aligned}$$

即 $d^{(k+1)}$与 $d^{(k)}$正交.

定理 4.2.1 表明，最速下降法相邻的两次迭代的前进方向是相互垂直的，因而整个行进路径呈锯齿形，因此从全局来看，收敛较慢.

回过头来再看一下例 4.2.2，若取 $a = 99$，$b = 1$，则由(4.2.5) 式得到

$$x^{(k+1)} = \left(99\left(\frac{98}{100}\right)^k, \left(-\frac{98}{100}\right)^k\right)^T,$$

当 $k\to\infty$时，$x^{(k)}\to 0$ 的速度很慢.

从前面的分析可以得到这样一个结论：最速下降法不最速.

换句话说，最速下降方向未必是最好的搜索方向．应该考虑其它的下降方向作为搜索方向．

再进一步对算法 4.2.1 进行分析，停止计算的条件（通常称为终止准则）是 $\nabla f(\boldsymbol{x}^{(k)})=0$ 也不够合理，其含义是当算法达到稳定点时停止计算；否则就继续计算下去．而在例 4.2.2 中，当 $a\neq b$ 时，只有当 $k\to\infty$ 时，才能达到稳定点．也就是说，计算将永远进行下去，永远不停机．这显然是不合理的．因此在实际计算中应采用较弱条件作为终止准则，如在点 $x^{(k)}$ 处的梯度向量接近于 0，即

$$\|\nabla f(\boldsymbol{x}^{(k)})\|<\epsilon, \tag{4.2.6}$$

其中式中的 ϵ 是预先给定的精度要求．

当然还有其它的终止准则，如

$$\|\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}\|<\epsilon, \tag{4.2.7}$$

$$|f(\boldsymbol{x}^{(k+1)})-f(\boldsymbol{x}^{(k)})|<\epsilon, \tag{4.2.8}$$

本书的终止准则主要以（4.2.6）式为准．

通过以上分析，给出一般下降算法．

算法 4.2.2　（一般下降算法）

（1）取初始点 $\boldsymbol{x}^{(1)}$，置精度要求 ϵ，置 $k=1$．

（2）若在点 $\boldsymbol{x}^{(k)}$ 处满足某个终止准则（如（4.2.6）），则停止计算（$\boldsymbol{x}^{(k)}$ 为无约束问题的最优解）；否则适当地选择 $\boldsymbol{x}^{(k)}$ 处的搜索方向 $\boldsymbol{d}^{(k)}$．

（3）一维搜索．求解一维问题

$$\min_{\alpha}\phi(\alpha)=f(\boldsymbol{x}^{(k)}+\alpha\boldsymbol{d}^{(k)}),$$

得 α_k，置

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\alpha_k\boldsymbol{d}^{(k)}.$$

（4）置 $k=k+1$，转（2）．

显然，若使算法 4.2.2 切实可行，必需首先解决以下三个问

题：

(1) 如何确定某点处的搜索方向？

(2) 如何进行一维搜索？

(3) 如何确定当前点的终止准则？

这三个问题留在以后的各章中去解决．在下一章将介绍一维搜索的方法．接着在后面的两章中将介绍确定搜索方向的方法，这一部分内容是无约束最优化方法的重点，每一种确定搜索方向的方法就决定了一种算法．对于第三个问题——终止准则，一般是与算法相联系的．

§4.3 算法的收敛性

4.3.1 算法的收敛性

这一节主要介绍有关算法的基本概念．

在前面讨论的下降算法是一类迭代方法，即从任意的初始点 $\boldsymbol{x}^{(1)}$ 出发，构造出点列 $\{\boldsymbol{x}^{(k)}\}$，并满足

$$f(\boldsymbol{x}^{(k)}) > f(\boldsymbol{x}^{(k+1)}),\ k = 1,2,\cdots, \tag{4.3.1}$$

但这个条件并不能保证序列 $\{\boldsymbol{x}^{(k)}\}$ 达到或收敛到无约束问题的最优解．

所谓收敛，是指序列 $\{\boldsymbol{x}^{(k)}\}$ 或它的一个子列（不妨仍记为 $\{\boldsymbol{x}^{(k)}\}$）满足

$$\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}^*, \tag{4.3.2}$$

这里 $\boldsymbol{x}^*$ 是无约束问题的局部解．

但是，通常要获得 (4.3.2) 这样强的结果是困难的，往往只能证明 $\{\boldsymbol{x}^{(k)}\}$ 的任一聚点的稳定点，或者证明更弱的条件

$$\lim_{k\to\infty} \inf \| \nabla f(\boldsymbol{x}^{(k)}) \| = 0. \tag{4.3.3}$$

（这种情况也称为收敛）

若对于某些算法来说，只有当初始点 $x^{(1)}$ 充分靠近极小点 x^* 时，才能保证序列 $\{x^{(k)}\}$ 收敛到 x^*，则称这类算法为局部收敛. 反之，若对任意的初始点 $\{x^{(1)}\}$，产生的序列 $\{x^{(k)}\}$ 收敛到 x^*，则称这类算法为全局收敛.

4.3.2 算法的收敛速率

如果算法产生的序列 $\{x^{(k)}\}$ 虽然收敛到 x^*，但收敛的太“慢”，以致于在计算机允许的时间内仍得不到满意的结果，那么，这类算法也称不上收敛. 因此，算法的收敛速率是一个十分重要的问题.

这里简单地介绍一个收敛速率的有关概念.

设序列 $\{x^{(k)}\}$ 收敛到 x^*，若极限

$$\lim_{k\to\infty}\frac{\| x^{(k+1)}-x^*\|}{\| x^{(k)}-x^*\|}=\beta \tag{4.3.4}$$

存在，当 $0<\beta<1$ 时，则称 $\{x^{(k)}\}$ 为线性收敛；当 $\beta=0$，则称 $\{x^{(k)}\}$ 为超线性收敛；当 $\beta=1$ 时，则称 $\{x^{(k)}\}$ 为次线性收敛. 因为次线性收敛的收敛速度太慢，一般不考虑它.

若存在某个 $p\geqslant 1$，有

$$\lim_{k\to\infty}\frac{\| x^{(k+1)}-x^*\|}{\| x^{(k)}-x^*\|^p}=\beta<+\infty, \tag{4.3.5}$$

则称 $\{x^{(k)}\}$ 为 p 阶收敛. 当 $p>1$ 时，p 阶收敛必为超线性收敛，但反之不一定成立.

在最优化算法中，通常考虑线性收敛、超线性收敛和二阶收敛. 如果说一个算法是线性（超线性或二阶）收敛的，是指算法产生的序列（在最坏情况下）是线性（超线性或二阶）收敛的.

4.3.3 算法的二次终止性

上面谈到的收敛性和收敛速率能够较为准确地刻划出算法的

优劣程度．但使用起来比较困难．特别是证明一个算法是否收敛或具有什么样的收敛速率，需要很强的理论知识．在这里给出一较为简单地判断算法优劣的评价标准——算法的二次终止性．

定义 4.3.1 若某个算法对于任意的正定二次函数，从任意的初始点出发，都能经有限步迭代达到其极小点，则称该算法具有二次终止性．

为什么用算法的二次终止性来作为判断算法优劣的标准呢？其原因有二．

(1) 正定二次目标函数具有某些好的性质，因此一个好的算法应能够在有限步内达到其极小点．

(2) 对于一个一般的目标函数，若在其极小点处的 Hesse 矩阵$\nabla^2 f(\boldsymbol{x}^*)$正定，由 Taylor 展开式得到

$$\begin{aligned} f(\boldsymbol{x}) = f(\boldsymbol{x}^*) &+ \nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \\ &+ \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T \nabla^2 f(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \\ &+ o(\|\boldsymbol{x} - \boldsymbol{x}^*\|^2), \end{aligned} \tag{4.3.6}$$

即目标函数 $f(\boldsymbol{x})$在极小点附近与一个正定二次函数相近似，因此可以猜想，对于正定二次函数好的算法，对于一般目标函数也应具有较好的性质．

因此，在后面我们用算法是否具有二次终止性来作为一个算法好坏的评价标准，即若算法具有二次终止性，则认为是好算法；否则认为算法的计算效果较差．在例 4.2.2 中，当 $a \neq b$ 时，最速下降法不可能在有限步内达到无约束问题的极小点，即最速下降法不具有二次终止性．从这一点来看，最速下降法不是一个好算法．

二次终止性是一个很重要的性质，后面将要讲到的许多算法总是根据它而设计出来的．

习题四

4.1 根据无约束问题局部解的必要条件和充分条件求解下列无约束问题：

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

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

(3) $\min f(\boldsymbol{x})=2x_1^2+2x_1x_2+4x_1x_3+3x_2^2+2x_2x_3+5x_3^2+4x_1-2x_2+3x_3$

4.2 根据无约束问题局部解的必要条件和充分条件求解无约束问题：

$$\min f(\boldsymbol{x}) = 2x_1^3 - 3x_1^2 - 6x_1x_2(x_1 - x_2 - 1).$$

4.3 设 G 是半正定矩阵，证明：若 $\bar{x}$ 满足 $\bar{\boldsymbol{x}}^TG\bar{\boldsymbol{x}}=0$，则 $\bar{\boldsymbol{x}}$ 满足 $G\bar{\boldsymbol{x}}=0$.

4.4 证明定理 4.1.4 及推论.

4.5 若去掉定理 4.1.4 推论中的可微条件，问其结论是否成立?，或举出反例，或给出相应的证明.

4.6 设

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^TG\boldsymbol{x} + \boldsymbol{r}^T\boldsymbol{x} + \delta$$

是正定二次函数，证明：一维问题

$$\min_{\alpha}\phi(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)})$$

的最优步长为

$$\alpha_k = -\frac{\nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d}^{(k)}}{\boldsymbol{d}^{(k)T}G\boldsymbol{d}^{(k)}}.$$

4.7 用最速下降法求解无约束问题

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

取初始点 $\boldsymbol{x}^{(1)}=(1, 1)^T$，迭代四步.

4.8 用最速下降法求解无约束问题

$$\min f(\boldsymbol{x}) = 10x_1^2 + x_2^2,$$

取初始点 $\boldsymbol{x}^{(1)} = (0.1,\ 1)^T$，试验证其收敛速度是线性的.

4.9 设 $u^{(k)} = k^{-2}$，$v^{(k)} = 2^{-k}$，$w^{(k)} = k^{-k}$，$x^{(k)} = \alpha^{2^k}$ $(0<\alpha<1)$，证明：

(1) 序列 $\{u^{(k)}\}$ 的收敛阶为 1，但不是线性收敛；

(2) 序列 $\{v^{(k)}\}$ 线性收敛，且收敛阶为 1；

(3) 序列 $\{w^{(k)}\}$ 超线性收敛，且收敛阶为 1；

(4) 序列 $\{x^{(k)}\}$ 二阶收敛.

第五章　一维搜索

一维搜索又称**线搜索**（Line Search），是指单变量函数最优化，即求一维问题

$$\min_{\alpha} \quad \phi(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}) \qquad (5.0.1)$$

最优解 α_k 的数值方法.

如果求得的 α_k 使目标函数沿 $\boldsymbol{d}^{(k)}$ 方向达到极小，即使得

$$f(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}) = \min_{\alpha} f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$$

或

$$\phi(\alpha_k) = \min_{\alpha} \phi(\alpha), \qquad (5.0.2)$$

则称该一维搜索为**精确一维搜索**（Exact Line Search），称 α_k 为最优步长. 如果存在 α_k，使得

$$f(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}) < f(\boldsymbol{x}^{(k)}) \text{或} \ \phi(\alpha_k) < \phi(0), \qquad (5.0.3)$$

则称该一维搜索为**非精确一维搜索**（Inexact Line Search）.

本章在前二节介绍最常用的精确一维搜索算法，在第三节简单介绍有关的非精确一维搜索算法.

§5.1　试探法

试探法的基本思想是陆续取一些试探点，逐次比较 $\phi(\alpha)$ 在这些点的函数值，完成一维搜索的任务.

5.1.1　搜索区间与单峰函数（单谷函数）

一维搜索的主要结构是：首先确定搜索区间，再用某种方法缩小这个区间，从而得到所需的最优解.

定义 5.1.1 设 α^* 是 $\phi(\alpha)$ 的极小点，若存在闭区间 $[a, b]$，使得 $\alpha^* \in [a, b]$，则称 $[a, b]$ 是 $\phi(\alpha)$ 的搜索区间.

确定搜索区间的一种简单的方法是进退法，其基本思想是从某一点出发，按一定的步长，确定函数值呈“高—低—高”的三点. 如果一个方向不成功，就退回来，再沿相反的方向寻找. 具体算法如下：

算法 5.1.1（确定搜索区间的进退法）

（1）取初始步长 α，置初始值 $\mu_3=0$，$\phi_3=\phi(\mu_3)$，并置 $k=0$.

（2）置

$$\mu=\mu_3+\alpha, \phi=\phi(\mu) \text{ 和 } k=k+1.$$

（3）如果 $\phi<\phi_3$，则置

$$\mu_2=\mu_3, \phi_2=\phi_3, \mu_3=\mu, \phi_3=\phi \text{ 和 } \alpha=2\alpha, k=k+1,$$

转（2）；

（4）如果 $k=1$，则置

$$\mu_2=\mu, \phi_2=\phi, \text{和 } \alpha=-\alpha,$$

转（2）；否则置

$$\mu_1=\mu_2, \phi_1=\phi_2, \mu_2=\mu_3, \phi_2=\phi_3, \mu_3=\mu, \phi_3=\phi,$$

并令

$$a=\min\{\mu_1, \mu_3\}, b=\max\{\mu_1, \mu_3\},$$

停止计算（搜索区间为 $[a, b]$）.

仅仅知道 $\phi(\alpha)$ 的搜索区间是不够的，本节所介绍的算法还要求函数在搜索区间上是**单峰函数**（或单谷函数，Unimodal Function）. 这里给出单峰函数的定义.

定义 5.1.2 设函数 $\phi(\alpha)$ 在区间 $[a, b]$ 内存在极小点 α^*，$\alpha^* \in (a, b)$. 如果对于任意的 α_1、α_2，满足：

(1) 当 $\alpha_1 < \alpha_2 \leqslant \alpha^*$ 时，有 $\phi(\alpha_1) \geqslant \phi(\alpha_2)$；

(2) 当 $\alpha^* \leqslant \alpha_1 < \alpha_2$ 时，有 $\phi(\alpha_1) \leqslant \phi(\alpha_2)$.

则称 $\phi(\alpha)$ 在区间 $[a, b]$ 上是单峰函数.

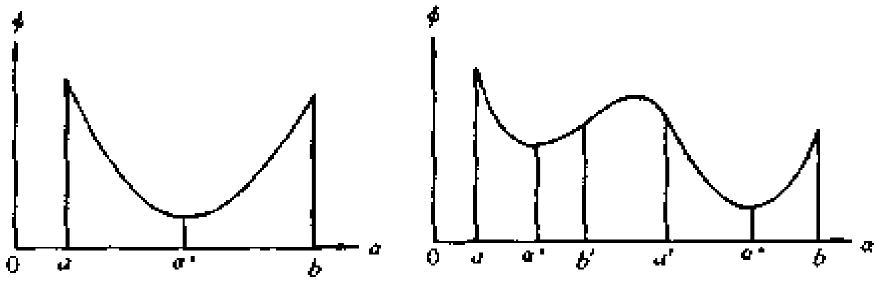


图 5.1.1　单峰函数和多峰函数

通常认为函数在由算法 5.1.1 得到的搜索区间上是单峰（单谷）函数.

5.1.2　0.618 法

0.618 法又称为**黄金分割法**（Golden Section Method），是用于在单峰函数区间上求极小的一种方法. 其基本思想是通过取试探点和进行函数值比较，使包含极小点的搜索区间不断减少，当区间长度缩短到一定程度时，就得到函数极小点的近似值.

0.618 是一元二次方程

$$\tau^2 + \tau - 1 = 0 \tag{5.1.1}$$

的根

$$\tau = \frac{\sqrt{5} - 1}{2} \tag{5.1.2}$$

的近似值.

先在搜索区间 $[a, b]$ 上确定两个试探点，其中左试探点为

$$\alpha_l = a + (1 - \alpha)(b - a), \tag{5.1.3}$$

右试探点为

$$\alpha_r = a + \tau(b - a), \tag{5.1.4}$$

再分别计算这两个试探点的函数值 $\phi_l = \phi(\alpha_l)$，$\phi_r = \phi(\alpha_r)$. 由单峰函数的性质，若 $\phi_l < \phi_r$，则区间 $[\alpha_r, b]$ 内不可能有极小点，因此去掉区间 $[\alpha_r, b]$，令 $a' = a$，$b' = \alpha_r$，得到一个新的搜索区间，若 $\phi_l > \phi_r$，则区间 $[a, \alpha_l]$ 内不可能有极小点，去掉区间 $[a, \alpha_l]$，令 $a' = \alpha_l$，$b' = b$，得到一个新的搜索区间.

类似上面的步骤，在区间 $[a', b']$ 内再计算两个新的试探点

$$\alpha'_l = a' + (1 - \tau)(b' - a'), \tag{5.1.5}$$

$$\alpha'_r = a' + \tau(b' - a'), \tag{5.1.6}$$

再比较函数值，再确定新的区间，如此下去，⋯.

在上述方法中，似乎每次迭代中需要计算两个试探点及它们的函数值，这里取 τ 值和其它值并没有什么不同，真是这样吗？下面对新的试探点进行分析：

(1) 若 $\phi_l < \phi_r$，则去掉区间 $[\alpha_r, b]$，那么新的右试探点为

$$\alpha'_r = a' + \tau(b' - a') = a + \tau(\alpha_r - a) = a + \tau^2(b - a),$$

注意到 τ 是方程 (5.1.1) 的根，因此有

$$\alpha'_r = a + \tau^2(b - a) = a + (1 - \tau)(b - a) = \alpha_l, \tag{5.1.7}$$

即原区间的左试探点.

(2) 若 $\phi_l > \phi_r$，则去掉区间 $[a, \alpha_l]$，那么新的左试探点为

$$\begin{aligned}\alpha'_l &= a' + (1 - \tau)(b' - a') = \alpha_l + (1 - \tau)(b - \alpha_l) \\ &= a + (1 - \tau)(b - a) + \tau(1 - \tau)(b - a)\end{aligned}$$

$$= a + (1 - \tau^2)(b - a) = a + \tau(b - a) \\ = \alpha_r, \tag{5.1.8}$$

即原区间的右试探点.

因此，在上述计算过程中，只需要计算一个新试探点和一个点的函数值.

事实上，0.618 法除第一次需要计算两个试探点外，其余各步每次只需计算一个试探点和它的函数值，这大大提高了算法的效率. 因此就得到相应的算法.

算法 5.1.2 (0.618 法)

(1) 置初始搜索区间 $[a, b]$，并置精度要求 ϵ，并计算左右试探点

$$\alpha_l = a + (1 - \tau)(b - a),\ \alpha_r = a + \tau(b - a),$$

其中 $\tau = \frac{\sqrt{5} - 1}{2}$，及相应的函数值

$$\phi_l = \phi(\alpha_l),\ \phi_l = \phi(\alpha_r).$$

(2) 如果 $\phi_l < \phi_r$，则置

$$b = \alpha_r,\ \alpha_r = \alpha_l,\ \phi_r = \phi_l,$$

并计算

$$\alpha_l = a + (1 - \tau)(b - a),\ \phi_l = \phi(\alpha_l);$$

否则置

$$a = \alpha_l,\ \alpha_l = \alpha_r,\ \phi_l = \phi_r,$$

并计算

$$\alpha_r = a + \tau(b - a),\ \phi_r = \phi(\alpha_r).$$

(3) 若 $|b - a| \leqslant \epsilon$，做：如果 $\phi_l < \phi_r$，则置 $\mu = \alpha_l$；否则置 $\mu = \alpha_r$，停止计算（μ 作为问题的解）. 否则解 (2).

【例 5.1.1】 用 0.618 法求解一维问题

$$\min \phi(\alpha) = e^\alpha - 5\alpha$$

在区间 [1, 2] 内的极小点，计算 4 步.

解 $a=1$, $b=2$, $\alpha_l=1.382$, $\alpha_r=1.618$, $\phi_l=-2.927$, $\phi_r=-3.047$, 所以 $\phi_l>\phi_r$, 去掉区间 $[1, \alpha_l]$. 详细计算结果见表 5.1.1.

表 5.1.1 迭代 4 次的计算结果

迭代次数	a	b	α_l	α_r	ϕ_l	ϕ_r	$\lvert b-a\rvert$
1	1	2	1.382*	1.618*	-2.927*	-3.047*	1.00
2	1.382	2	1.618	1.764*	-3.047	-2.984*	0.618
3	1.382	1.764	1.528*	1.618	-3.031*	-3.047	0.332
4	1.528	1.764	1.618	1.674*	-3.047	-3.037*	0.236

表中带 * 号的值是新计算出的.

经 n 次计算后，最终小区间的长度为 $\tau^{n-1}(b-a)$，这里 $b-a$ 为初始区间的长度. 由于每次迭代极小区间的收缩比为 τ，故 0.618 法的收敛速率是线性的，收敛比为 $\tau=\dfrac{\sqrt{5}-1}{2}$.

5.1.3 Fibonacci（斐波内奇）方法

现在不要求每次迭代极小区间的收缩比不变，而希望在试验点个数相同的情况下，找出一种选取试验点的最佳策略，使得最终的极小区间的长度达到最小，换句话说，如果规定试验点的个数为 n，且最终的极小区间长度为 1，问如何选取这 n 个点，使得原始极小区间的长度最大?

令 L_n 表示试验点数为 n、最终极小区间长度为 1 时，原始区间 $[a, b]$ 的最大可能长度，我们来找 L_n 的一个上界，设 α_l 为左试探点、α_r 为右试探点，如果极小点 α^* 位于区间 $[a, \alpha_l]$，则在此区间内至多还可以有 $n-2$ 个试验点（因为已计算过 α_l 和 α_r），因此 $\alpha_l-a\leqslant L_{n-2}$. 另一方面，如果极小点 α^* 位

于区间 $[\alpha_l, b]$ 内，则包括 α_r 在内，还可以作 $n-1$ 个试验点，所以 $b-\alpha_l \leqslant L_{n-1}$. 因此 $b-a=(b-\alpha_l)+(\alpha_l-a) \leqslant L_{n-2}+L_{n-1}$，故有如下关系式：

$$L_n \leqslant L_{n-2}+L_{n-1},$$

显然，不计算函数值和仅计算一点处的函数值都不能使极小区间缩小，即 $L_o=L_1=1$.

由前面的讨论得到，如果原始区间长度满足递推关系

$$F_n = F_{n-2}+F_{n-1}, \qquad (5.1.9)$$

$$F_0 = F_1 = 1, \qquad (5.1.10)$$

则 F_n 将是最大原始区间的长度. (5.1.9) 式和 (5.1.10) 式确定了唯一的一个自然数列 $\{F_k\}$：

n	0	1	2	3	4	5	6	7	8	9	10	…
F_n	1	1	2	3	5	8	13	21	34	55	89	…

称为 Fibonacci 数，相应的试探方法称为 Fibonacci 方法或分数法.

下面给出 Fibonacci 数列 $\{F_k\}$ 的通项表达式.

令 $F_k=r^k$，代入 (5.1.9) 式得到

$$r^2 = 1 + r \text{ 或 } r^2 - r - 1 = 0, \qquad (5.1.11)$$

它的两个根为

$$r_1 = \frac{1+\sqrt{5}}{2}, \quad r_2 = \frac{1-\sqrt{5}}{2},$$

因此，差分方程 (5.1.9) 的通解有如下的形式：

$$F_k = Ar_1^k + Br_2^k. \qquad (5.1.12)$$

由初始条件 (5.1.10) 式得到

$$\begin{cases} A + B = 1, \\ r_1A + r_2B = 1, \end{cases} \qquad (5.1.13)$$

其方程组的解为

$$A=\frac{1}{\sqrt{5}}r_1,\ B=-\frac{1}{\sqrt{5}}r_2,$$

所以 F_k 的通项表达式为

$$F_k=\frac{1}{\sqrt{5}}\left\{\left(\frac{1+\sqrt{5}}{2}\right)^{k+1}-\left(\frac{1-\sqrt{5}}{2}\right)^{k+1}\right\}. \tag{5.1.14}$$

注意到这里有一个有趣的现象，F_k 是自然数但其表达式确是无理数.

现在讨论如何选取试验点来实现一维搜索，由关系式 (5.1.9) 得到

$$\frac{F_{n-2}}{F_n}+\frac{F_{n-1}}{F_n}=1,\ n=2,3,\cdots, \tag{5.1.15}$$

Fibonacci 方法的基本思想与 0.618 法相同. 在搜索区间 $[a,b]$ 上，先取左、右试验点

$$\alpha_l=a+\frac{F_{n-2}}{F_n}(b-a), \tag{5.1.16}$$

$$\alpha_r=a+\frac{F_{n-1}}{F_n}(b-a), \tag{5.1.17}$$

比较函数值 $\phi(\alpha_l)$和 $\phi(\alpha_r)$，重新确定搜索区间.

(1) 若$\phi(\alpha_l)<\phi(\alpha_r)$，去掉区间 $[\alpha_r,b]$，令 $a'=a$，$b'=\alpha_r$，再计算新的试探点

$$\begin{aligned}\alpha'_r&=a'+\frac{F_{n-2}}{F_{n-1}}(b'-a')=a+\frac{F_{n-2}}{F_{n-1}}(\alpha_r-a)\\&=a+\frac{F_{n-2}}{F_{n-1}}\cdot\frac{F_{n-1}}{F_n}(b-a)\\&=a+\frac{F_{n-2}}{F_n}(b-a)=\alpha_l\end{aligned} \tag{5.1.18}$$

和

$$\alpha'_l = a' + \frac{F_{n-3}}{F_{n-1}}(b' - a').$$

(2) 若$\phi(\alpha_l) > \phi(\alpha_r)$，去掉区间 $[a, \alpha_r]$，令 $a' = \alpha_l$，$b' = b$，再计算新的试探点

$$\begin{aligned}
\alpha'_l &= a' + \frac{F_{n-3}}{F_{n-1}}(b' - a') = a_l + \frac{F_{n-3}}{F_{n-1}}(b - a_l) \\
&= a + \frac{F_{n-2}}{F_n}(b - a) + \frac{F_{n-3}}{F_{n-1}}\left(1 - \frac{F_{n-2}}{F_n}\right)(b - a) \\
&= a + \frac{F_{n-2}}{F_n}(b - a) + \frac{F_{n-3}}{Fn}(b - a) \\
&= a + \frac{F_{n-1}}{F_n}(b - a) = \alpha_r,
\end{aligned} \tag{5.1.19}$$

和

$$\alpha'_r = a' + \frac{F_{n-2}}{F_{n-1}}(b' - a').$$

所以 Fibonacci 方法与 0.618 法一样，除第一次外，以后每次只计算一个点处的函数值.

Fibonacci 方法每次保留的区间长度为$\frac{F_{k-1}}{F_k}$，因此若计算 n 个试验点，最终的区间长度为

$$\frac{F_{n-1}}{F_n} \cdot \frac{F_{n-2}}{F_{n-1}} \cdots \frac{F_2}{F_3} \cdot \frac{F_1}{F_2}(b - a) = \frac{1}{F_n}(b - a),$$

因此，任给一精度要求 ϵ，要求最终的区间长度小于 ϵ，即有

$$\frac{1}{F_n}(b - a) < \epsilon, \tag{5.1.20}$$

那么选择 n 满足

$$F_n > \frac{b-a}{\varepsilon}. \tag{5.1.21}$$

当 $n=2$ 时，由于 $F_0=F_1=1$，因此 α_l 与 α_r 相重合．当 $n=1$ 时，区间收缩比是 1，不缩减．因此需要增加一个量 δ，称为分离间隔，于是得到 Fibonacci 方法．

算法 5.1.3 （Fibonacci 方法）（分数法）

（1）置初始搜索区间 $[a, b]$，并置精度要求 ε，选取分离间隔 $\delta<\varepsilon$，求最小正整数 n，使得 $F_n>\frac{b-a}{\varepsilon}$，并计算左右试探点

$$\alpha_l = a + \frac{F_{n-2}}{F_n}(b-a),\ \alpha_r = a + \frac{F_{n-1}}{F_n}(b-a)$$

及相应的函数值

$$\phi_l = \phi(\alpha_l), \phi_r = \phi(\alpha_r).$$

（2）置 $n=n-1$.

（3）如果 $\phi_l<\phi_r$，则置

$$b = \alpha_r,\ \alpha_r = \alpha_l,\ \phi_r = \phi_l.$$

如果 $n>2$，则计算

$$\alpha_l = a + \frac{F_{n-2}}{F_n}(b-a),\ \phi_l = \phi(\alpha_l);$$

否则计算

$$\alpha_l = \alpha_r - \delta,\ \phi_l = \phi(\alpha_l).$$

（4）$(\phi_l \geqslant \phi_r)$ 置

$$a = \alpha_l,\ \alpha_l = \alpha_r,\ \phi_l = \phi_r$$

如果 $n>2$，则计算

$$\alpha_r = a + \frac{F_{n-1}}{F_n}(b-a),\ \phi_r = \phi(\alpha_r);$$

否则计算

$$\alpha_r = \alpha_l + \delta, \ \phi_r = \phi(\alpha_r).$$

(5) 若 $n=1$，做：如果 $\phi_l < \phi_r$，则置 $\mu = \alpha_r$；否则置 $\mu = \alpha_r$，停止计算（μ 作为问题的极小点）．否则转 (2)．

【例 5.1.2】 用 Fibonacci 方法求解一维问题

$$\min \phi(\alpha) = e^{\alpha} - 5\alpha$$

在区间 [1, 2] 内的极小点，取 $n=4$，$\delta=0.01$．

解 计算结果见表 5.1.2．

表 5.1.2 Fibonacci 方法（$n=4$）的计算结果

迭代次数	a	b	α_l	α_r	ϕ_l	ϕ_r	$\lvert b-a \rvert$
1	1.000	2.000	1.400*	1.600*	-2.945*	-3.047*	1.000
2	1.400	2.000	1.600	1.800*	-3.047	-2.950*	0.600
3	1.400	1.800	1.590*	1.600	-3.046*	-3.047	0.400
4	1.590	1.800	1.600	1.610*	-3.047	-3.047*	0.210

其中带 * 号的值是新计算出的．

现在来分析 Fibonacci 方法与 0.618 法的效率．可以证明

$$\tau^n < \frac{1}{F_n} < \tau^{n-1}, \tag{5.1.22}$$

(5.1.22) 式右端的不等式表明：若同样计算 n 次目标函数值，则 Fibonacci 方法略优于 0.618 法．而左端的不等式又表明，它不如多计算一次函数值的 0.618 法．由算法 5.1.3 可知，Fibonacci 方法还需要确定计算次数 n 和做其它的工作，因此在实际应用中，Fibonacci 方法不如 0.618 法．

5.1.4 二分法

在本节的最后，稍微提一下最简单的试探法——二分法．二分法本质上是求方程

$$\phi'(\alpha) = 0 \tag{5.1.23}$$

的根．注意到我们的目标是求 $\phi(\alpha)$的极小点，因此在选择初始区间 $[a, b]$ 时应满足：$\phi'(a)<0$ 和 $\phi'(b)>0$,其余的方法与求非线性方程的二分法完全相同，这里就不再重复了．

§5.2 插值法

插值法是一类重要的一维搜索方法，其基本思想是利用搜索区间上某点的信息构造插值多项式（通常不超过三次）$\hat{\phi}(\alpha)$,逐步用$\hat{\phi}(\alpha)$的极小点来逼近$\phi(\alpha)$的极小点 α^*．当 $\phi(\alpha)$有比较好的解析性质时，插值法比试探法（如 0.618 法）的效果好．

本节介绍三种较为常见的插值法．

5.2.1 三点二次插值法

已知函数在三点 μ_1、μ_2、μ_3（$\mu_1<\mu_2<\mu_3$）处的函数值为 ϕ_1、ϕ_2、ϕ_3，并满足

$$\phi_1 > \phi_2, \phi_3 > \phi_2,$$

即三点满足“两端高中间低”．这三个点可由算法 5.1.1（经微小改动）得到．这个条件是为了保证构造的函数$\hat{\phi}(\alpha)$在区间 $[\mu_1, \mu_3]$ 内有极小点．

由数值分析的知识，得到过三个点 (μ_1, ϕ_1)、(μ_2, ϕ_2)、(μ_3, ϕ_3) 的二次插值公式为

$$\begin{aligned}\hat{\phi}(\alpha) = & \phi_1 \frac{(\alpha-\mu_2)(\alpha-\mu_3)}{(\mu_1-\mu_2)(\mu_1-\mu_3)} \\ & + \phi_2 \frac{(\alpha-\mu_1)(\alpha-\mu_3)}{(\mu_2-\mu_1)(\mu_2-\mu_3)} \\ & + \phi_3 \frac{(\alpha-\mu_1)(\alpha-\mu_2)}{(\mu_3-\mu_1)(\mu_3-\mu_1)},\end{aligned} \tag{5.2.1}$$

对（5.2.1）式求导数，并求解方程

$$\bar{\phi}'(\alpha) = 0,$$

得到 $\bar{\phi}$（α）的极小点

$$\mu = \frac{\phi_1(\mu_1^2 - \mu_3^2) + \phi_2(\mu_3^2 - \mu_1^2) + \phi_3(\mu_1^2 - \mu_2^2)}{2[\phi_2(\mu_2 - \mu_3) + \phi_2(\mu_3 - \mu_1) + \phi_3(\mu_1 - \mu_2)]}, \tag{5.2.2}$$

用 μ 作为 α^* 的估计值，并计算 μ 处的函数值 $\phi = \phi(\mu)$.

第一次的近似结果往往不够理想，需要作进一步的近似. 现已得到四个点（μ_1，ϕ_1）、（μ_2，ϕ_2）、（μ_3，ϕ_3）和（μ，ϕ），如何选取三个点呢？仍然按照最初的原则，选取满足“两端高中间低”的三个点，这样就得到了相应的算法.

算法 5.2.1　（三点二次插值法）

（1）取初始点 $\mu_1 < \mu_2 < \mu_3$，计算 $\phi_i = \phi(\mu_i), i = 1, 2, 3$，并且满足 $\phi_1 > \phi_2$，$\phi_3 > \phi_2$，置精度要求 ε.

（2）计算

$$A = 2[\phi_1(\mu_2 - \mu_3) + \phi_2(\mu_3 - \mu_1) + \phi_3(\mu_1 - \mu_2)],$$

若 $A = 0$，则置 $\mu = \mu_2$，$\phi = \phi_2$，停止计算（输出 μ，ϕ 的信息）.

（3）计算

$$\mu = [\phi_1(\mu_2^2 - \mu_3^2) + \phi_2(\mu_3^2 - \mu_1^2) + \phi_3(\mu_1^2 - \mu_2^2)]/A,$$

若 $\mu < \mu_1$ 或 $\mu > \mu_3$（$\mu \notin (\mu_1, \mu_3)$），则置 $\mu = \mu_2$，$\phi = \phi_2$，停止计算（输出 μ，ϕ 的信息）.

（4）计算 $\phi = \phi$（μ）. 若 $|\mu - \mu_2| < \varepsilon$，则停止计算（$\mu$ 作为极小点）.

（5）如果 $\mu \in (\mu_2, \mu_3)$，则：

$$\begin{cases} \text{若 } \phi < \phi_2\text{，则置 } \mu_1 = \mu_2,\ \phi_1 = \phi_2,\ \mu_2 = \mu,\ \phi_2 = \phi; \\ \text{否则置 } \mu_3 = \mu,\ \phi_3 = \phi; \end{cases}$$

否则（$\mu \in (\mu_1, \mu_2)$）：

$$\begin{cases} \text{若 } \phi < \phi_2\text{，则置 } \mu_3 = \mu_2,\ \phi_3 = \phi_2,\ \mu_2 = \mu,\ \phi_2 = \phi; \\ \text{否则置 } \mu_1 = \mu,\ \phi_1 = \phi. \end{cases}$$

(6) 转 (2).

【例 5.2.1】 用三点二次插值法求解一维问题

$$\min\phi(\alpha) = \alpha^3 - 2\alpha + 1,$$

取初始点 $\mu_1=0$, $\mu_2=1$, $\mu_3=3$, $\varepsilon=10^{-2}$.

解 $\phi_1=1$, $\phi_2=0$, $\phi_3=22$, 满足 $\phi_1>\phi_2$, $\phi_3>\phi_2$, 表 5.2.1 列出它的计算结果.

表 5.2.1 例 5.2.1 的计算结果

迭代次数	μ_1	μ_2	μ_3	ϕ_1	ϕ_2	ϕ_3	$\bar{\mu}$	$\bar{\phi}$	$\lvert\mu_2-\bar{\mu}\rvert$
1	0	1	3	1	0	22	0.625	−0.006	0.375
2	0	0.625	1	1	−0.006	0	0.808	−0.089	0.183
3	0.625	0.808	1	−0.006	−0.089	0	0.815	−0.089	0.007

5.2.2 两点二次插值法

若已知$\phi(\alpha)$在两点 μ_1、μ_2 处的函数值 $\phi_i=\phi(\mu_i)(i=1,2)$ 和一点 μ_1 处的导数值 $\phi_1'=\phi'(\mu_1)$,并且 $\phi_1'<0$. 现构造二次函数$\hat{\phi}(\alpha)$,为了使 $\hat{\phi}(\alpha)$ 在区间 $[\mu_1, \mu_2]$ 上有极小点, ϕ_1、ϕ_2 和 ϕ_1' 还应满足

$$\phi_2 > \phi_1 + \phi'_1(\mu_2-\mu_1). \tag{5.2.3}$$

设 $\hat{\phi}(\alpha)$ 具有如下形式:

$$\hat{\phi}(\alpha) = A(\alpha-\mu_1)^2 + B(\alpha-\mu_1) + C, \tag{4.2.4}$$

其中 A、B、C 为待定常数.

由插值条件

$$\hat{\phi}(\mu_1)=\phi_1, \hat{\phi}(\mu_2)=\phi_2, \hat{\phi}'(\mu_1)=\hat{\phi}'_1 \tag{5.2.5}$$

得到

$$A=\frac{\phi_2-\phi_1-\phi'_1(\mu_2-\mu_1)}{(\mu_2-\mu_1)^2}, B=\phi'_1, C=\phi_1, \tag{5.2.6}$$

将（5.2.6）式代入（5.2.4）式，并求解方程

$$\hat{\phi}'(\alpha) = 0,$$

得到

$$\mu = \mu_1 - \frac{\phi_1'(\mu_2 - \mu_1)^2}{2[\phi_2 - \phi_1 - \phi_1'(\mu_2 - \mu_1)]}. \tag{5.2.7}$$

将 μ 作为 $\phi(\alpha)$ 的极小点 α^* 的近似值.

类似于三点二次插值法，如果第一次插值结果不够理想，再继续构造插值函数来求 α^* 的近似值. 由此得到如下算法.

算法 5.2.2（两点二次插值法）

（1）取初始点 μ_1，初始步长 α 和步长缩减因子 ρ，置精度要求 ϵ，并计算

$$\phi_1 = \phi(\mu_1), \ \phi'_1 = \phi'(\mu_1).$$

（2）若 $\phi_1' \leqslant 0$，则置 $\alpha = |\alpha|$；否则置 $\alpha = -|\alpha|$.

（3）计算

$$\mu_2 = \mu_1 + \alpha \text{ 和 } \phi_2 = \phi(\mu_2).$$

（4）如果 $\phi_2 \leqslant \phi_1 + \phi_1'(\mu_2 - \mu_1)$，则置 $\alpha = 2\alpha$，转（3）.

（5）计算

$$\mu = \mu_1 - \frac{\phi_1'(\mu_2 - \mu_1)^2}{2[\phi_2 - \phi_1 - \phi_1'(\mu_2 - \mu_1)]},$$

$$\phi = \phi(\mu), \phi' = \phi'(\mu).$$

（6）若 $|\phi'| \leqslant \epsilon$，则停止计算（$\mu$ 作为极小点）；否则置

$$\mu_1 = \mu, \ \phi_1 = \phi, \ \phi_1' = \phi', \ \alpha = \rho\alpha,$$

转（2）.

在算法 5.2.2 中，通常取 $\alpha = 1$，$\rho = \frac{1}{10}$.

【例 5.2.2】 用两点二次插值法求解一维问题

$$\min \phi(\alpha) = \alpha^3 - 2\alpha + 1.$$

取初始点 $\mu_1 = 0$，初始步长 $\alpha = 1$，$\epsilon = 10^{-3}$.

解　表 5.2.2 列出全部计算结果.

表 5.2.2　例 5.2.2 的计算结果

迭代次数	μ_1	ϕ_1	ϕ_1'	α	μ_2	ϕ_2	μ	$\lvert\phi'\rvert$
1	0.000	1.000	-2.000	1.000	1.000	0.000	1.000	1.000
2	1.000	0.000	1.000	-0.100	0.900	-0.071	0.828	0.055
3	0.828	-0.088	0.055	0.010	0.818	-0.089	0.817	0.000

5.2.3　两点三次插值法

已知两点 μ_1、μ_2 处的函数值 $\phi_i=\phi(\mu_i)(i=1,2)$ 和它们的导数值 $\phi_i'=\phi'(\mu_i)(i=1,2)$，由 Himiter 插值公式可以构造出一个三次插值多项式 $\tilde{\phi}(\alpha)$. 由方程 $\tilde{\phi}'(\alpha)=0$ 得到 $\tilde{\phi}(\alpha)$ 的极小点

$$\mu=\mu_1+(\mu_2-\mu_1)\left(1-\frac{\phi'_2+w+z}{\phi'_2-\phi'_1+2w}\right), \quad (5.2.8)$$

其中

$$z=\frac{3(\phi_2-\phi_1)}{\mu_2-\mu_1}-\phi'_1-\phi'_2, \quad (5.2.9)$$

$$w=\operatorname{sgn}(\mu_2-\mu_1)\cdot\sqrt{z^2-\phi'_1\cdot\phi'_2}, \quad (5.2.10)$$

下面给出相应的算法.

算法 5.2.3　（两点三次插值法）

（1）置初始值 μ_1 和初始步长 α 和步长缩减因子 ρ，置精度要求 ϵ，并计算 $\phi_1=\phi(\mu_1),\phi_1'=\phi'(\mu_1)$.

（2）如果 $\phi_1'>0$，则置 $\alpha=-|\alpha|$；否则置 $\alpha=|\alpha|$.

（3）置 $\mu_2=\mu_1+\alpha$，并计算 $\phi_2=\phi(\mu_2)$，$\phi_2'=\phi'(\mu_2)$.

（4）如果 $\phi_1'\cdot\phi_2'>0$，则置

$$\alpha=2\alpha,\mu_1=\mu_2,\phi_1=\phi_2,\phi'_1=\phi'_2,$$

转（3）.

（5）计算

$$z=\frac{3(\phi_2-\phi_1)}{\mu_2-\mu_1}-\phi'_1-\phi'_2,$$

$$w=\mathrm{sign}(\mu_2-\mu_1)\cdot\sqrt{z^2-\phi'_1\cdot\phi'_2},$$

$$\mu=\mu_1+(\mu_2-\mu_1)\left(1-\frac{\phi'_2+w+z}{\phi'_2-\phi'_1+2w}\right),$$

和

$$\phi=\phi(\alpha),\ \phi'=\phi'(\mu).$$

（6）如果$|\phi'|<\varepsilon$，则停止计算（μ作为问题的极小点）；否则置

$$\alpha=\rho\alpha,\ \mu_1=\mu,\ \phi_1=\phi,\ \phi'_1=\phi',$$

转（2）.

在算法 5.2.3 中，通常取$\alpha=1$，$\rho=\frac{1}{10}$.

§5.3　非精确一维搜索方法

在前面两节中所介绍的方法是试图求一维问题$\min\limits_{\alpha}\phi(\alpha)$的极小点，而对于无约束问题的整体来看，有时并不要求一定得到极小点，而只需要有一定的下降量即可．这样做的优点是减少每次的一维搜索时间，使整体效果最好，这就是为什么要讨论非精确一维搜索方法.

非精确一维搜索的基本思想是求μ，使得$\phi(\mu)<\phi(0)$，但不希望μ值过大，过大会引起点列$\{x^{(k)}\}$产生大幅度的摆动，也不希望μ值过小，过小会使点列$\{x^{(k)}\}$在达不到x^*之前而没有进展.

5.3.1 Goldstein 方法

Goldstein 方法是 Goldstein 在 1967 年提出来的，预先指定两个参数 β_1 和 β_2（精度要求），满足 $0<\beta_1<\beta_2<1$，用下面两个不等式来限定步长 μ，即

$$\phi(\mu)\leqslant\phi(0)+\mu\beta_1\phi'(0),\tag{5.3.1}$$

$$\phi(\mu)\geqslant\phi(0)+\mu\beta_2\phi'(0),\tag{5.3.2}$$

(5.3.1) 式和 (5.3.2) 式的几何意义见图 5.3.1.

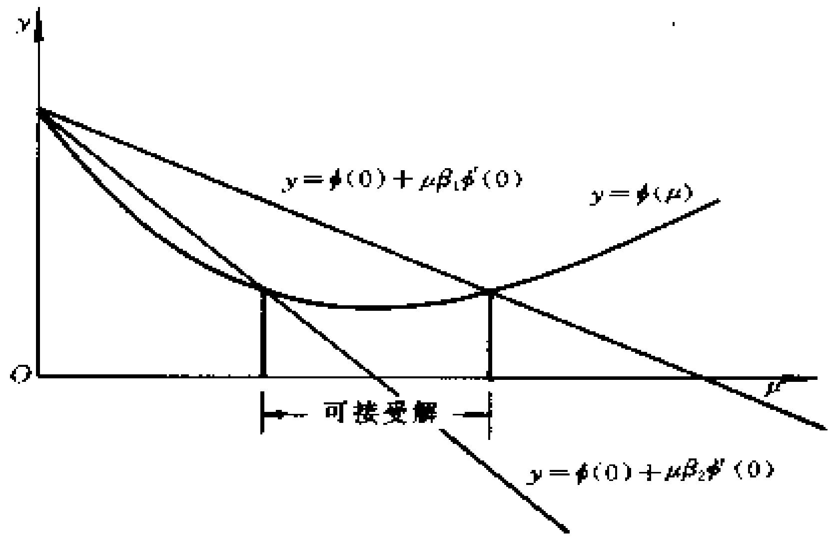


图 5.3.1

从图来看，μ 值在 $y=\phi(\mu)$ 图形夹于直线 $y=\phi(0)+\beta_1\phi'(0)\mu$ 和直线 $y=\phi(0)+\beta_2\phi'(0)\mu$ 之间，直线 $y=\phi(0)+\beta_1\phi'(0)\mu$ 控制 μ 值不要过大，而直线 $y=\phi(0)+\beta_2\phi'(0)\mu$ 是控制 μ 值不要过小. 由此得到相应的算法.

算法 5.3.1 （Goldstein 方法）

(1) 取初始试探点 μ，置 $\mu_{\min}=0$，$\mu_{\max}=+\infty$（充分大的数）. 置精度要求 $0<\beta_1<\beta_2<1$.

(2) 如果 $\phi(\mu)>\phi(0)+\beta_1\phi'(0)\mu$，则置

$$\mu_{\max}=\mu;$$

否则如果 $\phi(\mu)\geqslant\phi(0)+\beta_2\phi'(0)\mu$，则停止计算（$\mu$ 作为非精确

一维搜索步长)；

否则置

$$\mu_{\min} = \mu.$$

(3) 如果 $\mu_{\max} < +\infty$ (有限)，则置

$$\mu = \frac{1}{2}(\mu_{\min} + \mu_{\max});$$

否则置

$$\mu = 2\mu_{\min}.$$

(4) 转 (2).

5.3.2 Armijo 方法

Armijo 方法是 Goldstein 方法的一种变形，是 Armijo 在 1969 年提出来的.

预先取定一大于 1 的数 M 和 $0<\beta_1<1$，μ 的选取使得

$$\phi(\mu) \leqslant \phi(0) + \mu\beta_1\phi'(0) \tag{5.3.3}$$

成立，而 $M\mu$ 不成立. 通常 M 为 2 到 10 之间.

5.3.3 Wolfe-Powell 方法

Wolfe-Powell 方法是在 1969 年到 1976 年期间，由 Wolfe 和 Powell 提出的. 预先指定参数 β_1 和 β_2，且 $0<\beta_1<\beta_2<1$，使得步长 μ 满足

$$\phi(\mu) \leqslant \phi(0) + \beta_1\phi'(0)\mu, \tag{5.3.4}$$

$$\phi'(\mu) \geqslant \beta_2\phi'(0), \tag{5.3.5}$$

(5.3.5) 式有时写成

$$|\phi'(\mu)| \leqslant \beta_2 |\phi'(0)|. \tag{5.3.6}$$

Wolfe-Powell 方法的优点是：在可接受解中包含了最优解 α^*，而 Goldstein 方法确不能保证这一点. Wolfe-Powell 方法的几何意义见图 5.3.2.

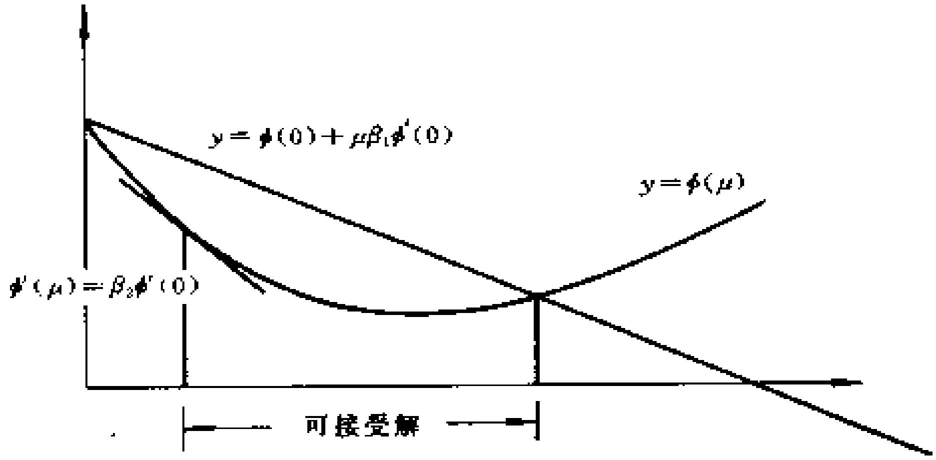


图 5.3.2

习题五

5.1 考虑问题

$$\min\phi(\alpha) = (\alpha + 1)^2,$$

取初始点 $\mu_3=0$，步长 $\alpha=0.2$，用确定搜索区间的进退法求出 $\{\mu_1, \mu_2, \mu_3\}$.

5.2 用 0.618 法求解一维问题

$$\min\phi(\alpha) = (\alpha^2 - 1)^2$$

在区间 [0, 2] 内的一个解，取精度要求 $\epsilon=0.1$.

5.3 证明：Fibonacci 序列满足

$$F_{n-2}^2 - F_{n-1} \cdot F_{n-3} = (-1)^n.$$

5.4 求极限 $\lim\limits_{n\to\infty}\dfrac{F_{n-1}}{F_n}$.

5.5 用 Fibonacci 方法求解

$$\min\phi(\alpha) = (\alpha^2 - 1)^2$$

在区间 [0, 2] 的解，取 $n=5$，$\delta=0.01$.

5.6 设 ϕ (α) 是二次函数，其二次项的系数为正数，记

$$\phi_1 = \phi(\bar{\alpha} - h), \ \phi_2 = \phi(\bar{\alpha}), \ \phi_3 = \phi(\bar{\alpha} + h),$$

其中$\bar{\alpha}$和h是给定的常数，且$h \neq 0$，求证：$\phi(\alpha)$的极小值为

$$\phi^* = \phi_2 - \frac{(\phi_1 - \phi_3)^2}{8(\phi_1 - 2\phi_2 + \phi_3)}.$$

5.7 用三点二次插值法求解一维问题

$$\min \phi(\alpha) = e^{\alpha} - 5\alpha,$$

取$\mu_1 = 1$，$\mu_2 = 2$，$\mu_3 = 3$（计算三步）.

5.8 用两点二次插值法求解一维问题

$$\min \phi(\alpha) = e^{\alpha} - 5\alpha,$$

取$\mu_1 = 1$，步长$\alpha = 1$（计算三步）.

5.9 写出 Armijo 方法的计算步骤.

5.10 写出 Wolfe-Powell 方法的计算步骤.

第六章　使用导数的最优化方法

本章和下一章介绍求解无约束最优化问题

$$\min f(\boldsymbol{x}),\quad \boldsymbol{x}\in R^n$$

的最优化方法．通常把求解无约束问题的算法大致分为两类：一类在计算过程中要用到目标函数的导数值，这类方法在本章介绍；另一类仅用到目标函数的函数值不必计算导数，这类方法称为直接方法，将在下章讨论．

§6.1　Newton 法

6.1.1　Newton 法

若 x^* 是无约束问题的局部解，则 x^* 满足

$$\nabla f(\boldsymbol{x})=0,\tag{6.1.1}$$

因此，可以通过求解方程组（6.1.1）来得到无约束最优化问题解．注意到方程组（6.1.1）是非线性的，处理起来比较困难，因此考虑它的一个线性逼近．选取初始点 $\boldsymbol{x}^{(1)}$（作为 $\boldsymbol{x}^*$ 的第一次近似），在 $\boldsymbol{x}^{(1)}$处线性展开，略去高阶部分得到

$$\nabla f(\boldsymbol{x})\approx\nabla f(\boldsymbol{x}^{(1)})+\nabla^2 f(\boldsymbol{x}^{(1)})(\boldsymbol{x}-\boldsymbol{x}^{(1)}).\tag{6.1.2}$$

令（6.1.2）式右端为 0，即

$$\nabla f(\boldsymbol{x}^{(1)})+\nabla^2 f(\boldsymbol{x}^{(1)})(\boldsymbol{x}-\boldsymbol{x}^{(1)})=0.\tag{6.1.3}$$

求解线性方程组（6.1.3）得到

$$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}-(\nabla^2 f(\boldsymbol{x}^{(1)}))^{-1}\nabla f(\boldsymbol{x}^{(1)}),$$

作为 $\boldsymbol{x}^*$ 的第二次近似．

如果 $\boldsymbol{x}^{(2)}$的精度不够，可以在 $\boldsymbol{x}^{(2)}$处将$\nabla f(\boldsymbol{x})$ 展开，求解

相应的线性方程组，得到 $\boldsymbol{x}^{(3)}$，如此下去，可以得到序列 $\{\boldsymbol{x}^{(k)}\}$，并且满足如下迭代公式

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - (\nabla^2 f(\boldsymbol{x}^{(k)}))^{-1}\nabla f(\boldsymbol{x}^{(k)}), k = 1,2,\cdots, \tag{6.1.4}$$

称（6.1.4）式为 Newton 迭代公式.

为了便于计算，将（6.1.4）式改为

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{d}^{(k)}, \tag{6.1.5}$$

其中 $d^{(k)}$是线性方程组

$$\nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{d} = -\nabla f(\boldsymbol{x}^{(k)}) \tag{6.1.6}$$

的解. 通常称（6.1.6）式为 Newton 方程.

由上述推导过程，我们得到如下算法.

解法 6.1.1（Newton 法）

(1) 取初始点 $\boldsymbol{x}^{(1)}$，置精度要求 ϵ，置 $k=1$.

(2) 如果 $\|\nabla f(\boldsymbol{x}^{(k)})\| \leqslant \epsilon$，则停止计算（$\boldsymbol{x}^{(k)}$作为无约束问题的解）；否则求解线性方程组

$$\nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{d} = -\nabla f(\boldsymbol{x}^{(k)}),$$

得到 $\boldsymbol{d}^{(k)}$.

(3) 置

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{d}^{(k)}, k = k + 1,$$

转 (2).

【例 6.1.1】 用 Newton 法求解无约束问题

$$\min f(\boldsymbol{x}) = 4x_1^2 + x_2^2 - x_1^2 x_2,$$

分别取初始点为 $\boldsymbol{x}_A = (1,\ 1)^T$，$\boldsymbol{x}_B = (3,\ 4)^T$，$\boldsymbol{x}_C = (2,\ 0)^T$，精度要求 $\epsilon = 10^{-3}$.

解

$$f(\boldsymbol{x}) = 4x_1^2 + x_2^2 - x_1^2 x_2,$$

$$\nabla f(\boldsymbol{x}) = (8x_1 - 2x_1x_2,\ 2x_2 - x_1^2)^T,$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 8 - 2x_2 & -2x_1 \\ -2x_1 & 2 \end{pmatrix}.$$

(1) 取 $\boldsymbol{x}^{(1)}=\boldsymbol{x}_A=(1,1)^T$，得到$\nabla f(\boldsymbol{x}^{(1)})=(6,1)^T$，
$\nabla^2 f(\boldsymbol{x})=\begin{pmatrix}6 & -2\\ -2 & 2\end{pmatrix}$，解线性方程组

$$\nabla^2 f(\boldsymbol{x}^{(k)})\ \boldsymbol{d}=-\nabla f(\boldsymbol{x}^{(k)}),$$

即

$$\begin{pmatrix}6 & -2\\ -2 & 2\end{pmatrix}\begin{pmatrix}d_1\\ d_2\end{pmatrix}=-\begin{pmatrix}6\\ 1\end{pmatrix},$$

得到 $d_1=-1.75$，$d_2=-2.25$，即 $\boldsymbol{d}^{(1)}=(-1.75,-2.25)^T$，所以

$$\begin{aligned}\boldsymbol{x}^{(2)}&=\boldsymbol{x}^{(1)}+\boldsymbol{d}^{(1)}=(1,1)^T+(-1.75,-2.25)^T\\&=(-0.75,-1.25)^T.\end{aligned}$$

再进入第二轮计算. 经过 5 次迭代，$\boldsymbol{x}^{(k)}$收敛到问题的极小点 $\boldsymbol{x}^*=(0,0)^T$，计算结果见表 6.1.1.

表 6.1.1　$\boldsymbol{x}^{(1)}=(1,1)^T$ 的计算结果

k	$\boldsymbol{x}^{(k)}$	$f(\boldsymbol{x}^{(k)})$	$\nabla f(\boldsymbol{x}^{(k)})$	$\Vert\nabla f(\boldsymbol{x}^{(k)})\Vert$	$\nabla^2 f(\boldsymbol{x}^{(k)})$	
1	1.0000	4.0000	6.0000	6.0828	6.0000	-2.0000
	1.0000		1.0000		-2.0000	2.0000
2	-0.7500	4.5156	-7.8750	8.4495	10.5000	1.5000
	-1.2500		-3.0625		1.5000	2.0000
3	-0.1550	0.1273	-1.2911	1.3388	8.3300	0.3100
	-0.1650		-0.3540		0.3100	2.0000
4	-0.0057	0.0003	-0.0459	0.0511	8.0222	0.0115
	-0.0111		-0.0223		0.0115	2.0000
5	-0.0000	0.0000	-0.0001	0.0001	8.0000	0.0000
	-0.0000		-0.0000		0.0000	2.0000

(2) 取 $\boldsymbol{x}^{(1)}=\boldsymbol{x}_B=(3,4)^T$，计算步骤同 (1)，最后 $\boldsymbol{x}^{(k)}$ 收敛到 $(2\sqrt{2},4)^T$ 目标函数的鞍点，计算结果见表 6.1.2.

表 6.1.2　$x^{(1)}=(3,4)^T$ 的计算结果

k	$x^{(k)}$	$f(x^{(k)})$	$\nabla f(x^{(k)})$	$\Vert\nabla f(x^{(k)})\Vert$	$\nabla^2 f(x^{(k)})$
1	3.0000 4.0000	16.0000	0.0000 -1.0000	1.0000	0.0000　-6.0000 -6.0000　2.0000
2	2.8333 4.0000	16.0000	0.0000 -0.2078	0.0278	0.0000　-5.6667 -5.6667　2.0000
3	2.8284 4.0000	16.0000	0.0000 0.0000	0.0000	0.0000　-5.6569 -5.6569　2.0000

(3) 取 $x^{(1)}=x_C=(2,0)^T$，得到 $\nabla^2 f(x)=\begin{pmatrix}8 & -4\\ -4 & 2\end{pmatrix}$，Hesse 矩阵奇异，无法进行下一步计算.

例 6.1.1 表明用 Newton 法求解无约束问题会出现以下几种情况：①收敛到极小点，②收敛到鞍点，③Hesse 矩阵奇异，无法继续计算.

6.1.2　修正 Newton 法（阻尼 Newton 法）

在介绍修正 Newton 法之前，先分析一下 Newton 法的优缺点.

Newton 法的优点：

(1) Newton 法产生的点列 $\{x^{(k)}\}$ 若收敛，则收敛速度快——具有二阶收敛速率（证明略）.

(2) Newton 法具有二次终止性. 事实上，设

$$f(x)=\frac{1}{2}x^TGx+r^Tx+\delta,$$

任取 $x^{(1)}$，则 $\nabla f(x^{(1)})=Gx^{(1)}+r$，$\nabla^2 f(x^{(1)})=G$，因此

$$\begin{aligned}x^{(2)}&=x^{(1)}-(\nabla^2 f(x^{(1)}))^{-1}\nabla f(x^{(1)})\\&=x^{(1)}-G^{-1}(Gx^{(1)}+r)=-G^{-1}r\end{aligned}$$

$$= \boldsymbol{x}^*,$$

即算法一步达到极小点.

Newton 法的缺点:

(1) 可能会出现在某步迭代时目标函数值上升. 即存在 k, 使得 $f(\boldsymbol{x})^{(k+1)} > f(\boldsymbol{x}^{(k)})$. 如在例 6.1.1 中, 取 $\boldsymbol{x}^{(1)} = (1, 1)^T$, $f(\boldsymbol{x}^{(1)}) = 4$, 由 Newton 法得到的 $\boldsymbol{x}^{(2)} = (-0.75, -1.25)^T$, 此时, $f(\boldsymbol{x}^{(2)}) = 4.5156$ 函数值上升.

(2) 当 $\boldsymbol{x}^{(1)}$ 距 $\boldsymbol{x}^*$ 较远时, 产生的点列 $\{\boldsymbol{x}^{(k)}\}$ 可能不收敛; 或者收敛到鞍点; 或者 $\nabla^2 f(\boldsymbol{x}^{(k)})$ 奇异无法计算. 在例 6.1.1 中, 取初始点 $\boldsymbol{x}^{(1)} = (3, 4)^T$ 时, 产生的点列收敛到鞍点. 取初始点 $\boldsymbol{x}^{(1)} = (2, 0)^T$ 时, $\nabla^2 f(\boldsymbol{x}^{(1)})$ 奇异无法计算.

(3) 需要计算 Hesse 矩阵, 计算量大. 因为需要计算 $\frac{1}{2}n(n+1)$ 个函数值.

修正 Newton 法是针对缺点 (1) 提出改进方案的. 为了避免函数值上升, 在算法中增加一维搜索策略, 即将 $\boldsymbol{d}^{(k)}$ 作为搜索方向 (此时称 $\boldsymbol{d}^{(k)}$ 为 Newton 方向), 而不是作为增量, 由此得到如下算法.

算法 6.1.2 (修正 Newton 法)

(1) 取初始点 $\boldsymbol{x}^{(1)}$, 置精度要求 ε, 置 $k=1$.

(2) 如果 $\|\nabla f(\boldsymbol{x}^{(k)})\| \leqslant \varepsilon$, 则停止计算 ($\boldsymbol{x}^{(k)}$ 作为无约束问题的解); 否则求解线性方程组

$$\nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{d} = -\nabla f(\boldsymbol{x}^{(k)}),$$

得到 $\boldsymbol{d}^{(k)}$.

(3) 一维搜索. 求解一维问题

$$\min \phi(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}),$$

得 α_k, 置

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}.$$

(4) 置 $k=k+1$，转 (2).

由于采用了一维搜索，限制了 $f(\boldsymbol{x}^{(k)}+\alpha\boldsymbol{d}^{(k)})$ 上升，因此也称为阻尼 Newton 法.

修正 Newton 法保持了 Newton 法的优点，如算法的二次终止性等，克服了目标函数值上升的缺点，由于增加了一维搜索，对缺点 (2) 有所改善，但不能完全克服. 对于缺点 (3) 没有任何改进，因为修正 Newton 法仍需要计算 Hesse 矩阵. 有关不计算 Hesse 矩阵的算法，我们将在下两节讨论.

6.1.3 关于 Newton 法的进一步讨论

Newton 法和修正 Newton 法的主要困难在于：当 Hesse 矩阵 $\nabla^2 f(\boldsymbol{x}^{(k)})$ 不正定，得到的 Newton 方向不是下降方向，这样算法可能达不到极小点.

一种修正方案是将 Newton 方向与最速下降方向相结合. 设 $\boldsymbol{d}_N^{(k)}$ 是 Newton 方向，即 $\boldsymbol{d}_N^{(k)}$ 满足

$$\nabla^2 f(\boldsymbol{x}^{(k)})\ \boldsymbol{d}_N^{(k)}=-\nabla f(\boldsymbol{x}^{(k)})$$

或
$$\boldsymbol{d}_N^{(k)}=-[\nabla^2 f(\boldsymbol{x}^{(k)})]^{-1}\nabla f(\boldsymbol{x}^{(k)}),$$

这里假设 $\nabla^2 f(\boldsymbol{x}^{(k)})$ 非奇异. 设 θ_k 是 $d^{(k)}$ 与 $-\nabla f(\boldsymbol{x}^{(k)})$ 之间的夹角. 当然希望 θ_k 不要大于 $\frac{\pi}{2}$，即存在正数 $0<\eta<1$，使得 $\cos\theta_k\geqslant\eta$.

Goldstein 和 Price 在 1967 年提出的修正方案是令

$$\boldsymbol{d}^{(k)}=\begin{cases}\boldsymbol{d}_N^{(k)}, & \text{若 } \cos\theta_k\geqslant\eta,\\ -\nabla f(\boldsymbol{x}^{(k)}), & \text{其它}.\end{cases}\tag{6.1.7}$$

这样得到的 $\boldsymbol{d}^{(k)}$ 总是下降方向，并满足 $\cos\theta_k\geqslant\eta$.

另一种修正方案是当 Hesse 矩阵非正定时，选择负曲率 (Negative Curvature) 下降方向. 这类方法是由 Fiacco 和

McCormick 在 1968 年最早提出的.

定义 6.1.1 若 $\boldsymbol{d}^{(k)}$满足

$$\boldsymbol{d}^{(k)T}\nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{d}^{(k)} < 0, \tag{6.1.8}$$

则称 $\boldsymbol{d}^{(k)}$是点 $\boldsymbol{x}^{(k)}$的负曲率方向.

当$\nabla^2 f(\boldsymbol{x}^{(k)})$ 非正定时，负曲率方向一定存在. 事实上，当$\nabla^2 f(\boldsymbol{x}^{(k)})$ 非正定时，一定存在负特征值和对应的特征向量 $\boldsymbol{u}$，则 $\boldsymbol{u}$ 满足（6.1.8）式，因此一种直观地求负曲率下降方向的方法是令

$$\boldsymbol{d}^{(k)} = -\operatorname{sgn}(\boldsymbol{u}^T\nabla f(\boldsymbol{x}^{(k)}))\boldsymbol{u}. \tag{6.1.9}$$

当然还有比（6.1.9）式更好的负曲率方向.

还有一种方法是将 Newton 方程（6.1.6）改为

$$G^{(k)}\boldsymbol{d} = -\nabla f(\boldsymbol{x}^{(k)}), \tag{6.1.10}$$

其中 $G^{(k)} = \nabla^2 f(\boldsymbol{x}^{(k)}) + \nu_k I$. 当 ν_k 大于$\nabla^2 f(\boldsymbol{x}^{(k)})$ 的最小特征值时，$G^{(k)}$正定，这样由（6.1.10）式得到的搜索方向是下降方向.

关于这些算法的细节和其它的修正方案可参见其它的参考文献.

§6.2 共轭梯度法

上节介绍的 Newton 法虽然具有收敛速度快和二次终止性的优点，但还有产生的点列可能不收敛和计算量大等缺点，修正 Newton 法虽然克服了 Newton 法的某些缺点，但计算量大这一缺点仍没有克服. 针对这一问题我们将在本节和下一节讨论的算法中予以克服，这些算法是以正定二次函数为基础，仍适用一般的可微函数，并对于一般可微函数具有如下性质：

（1）产生的搜索方向是下降方向；

（2）不必计算 Hesse 矩阵，只计算目标函数值和梯度值；

(3) 算法具有二次终止性.

6.2.1 正交方向和共轭方向

首先考虑一类特殊的正定二次函数

$$f(w)=\frac{1}{2}w^Tw+r^Tw+\delta. \qquad (6.2.1)$$

所以$\nabla f(w)=w+r$，令$\nabla f(w)=0$，得到$f(w)$的极小点$w^*=-r$，由$f(w)$在w^*处的Taylor展开式得到

$$\begin{aligned}f(w)&=f(w^*)+\frac{1}{2}(w-w^*)^T(w-w^*)\\&=f(w^*)+\frac{1}{2}\|w-w^*\|^2.\end{aligned} \qquad (6.2.2)$$

(6.2.2)式表明$f(w)$的等高面是一族(超)球面. 特别对于$n=2$时，其等高线是一族圆. 任取初始点$w^{(1)}$，沿两个相互正交的方向进行精确一维搜索，得到点$w^{(3)}=w^*$，即为问题的最优解(见图6.2.1).

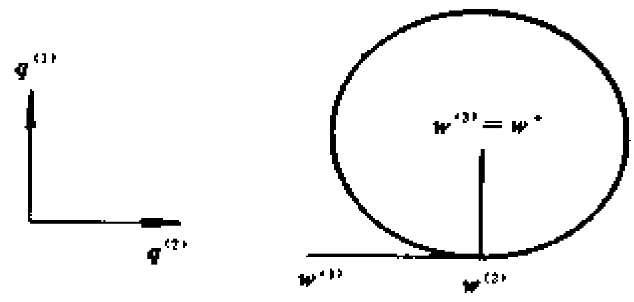


图6.2.1 正交方向与特殊的正定二次函数(二维)

当$n=3$时也有类似的结果，设$q^{(1)}$、$q^{(2)}$、$q^{(3)}$是三个相互正交的非零向量，则任取初始点$w^{(1)}$，依次沿它们进行精确一维搜索，达到极小点$w^{(4)}=w^*$(见图6.2.2). 特别注意到，$w^{(2)}$是过$w^{(1)}$沿方向l生成的直线上的极小点，$w^{(3)}$是过$w^{(1)}$由

$\{q^{(1)}, q^{(2)}\}$ 张成的平面 π 上的极小点，$w^{(4)}$是整个空间上的极小点.

现在将上述结果推广到 n 维情况.

定义 6.2.1 若 R^n 中的 k（$k \leqslant n$）个向量 $q^{(1)}, q^{(2)}, \cdots, q^{(k)}$两两正交，即

$$\langle q^{(i)}, q^{(j)} \rangle = 0, \quad i \neq j, \quad 1 \leqslant i, \ j \leqslant k, \tag{6.2.3}$$

则称它们为 k 个正交方向，若又满足

$$q^{(i)} \neq 0, \quad i = 1, 2, \cdots, k, \tag{6.2.4}$$

则称为 k 个非零的正交方向.

可以猜想，任取一初始点 $w^{(1)}$，依次沿 n 个非零的正交方向$q^{(1)}, q^{(2)}, \cdots, q^{(n)}$作精确一维搜索得到最优点 $w^{(n+1)} = w^*$. 并且 $w^{(2)}$是过 $w^{(1)}$沿 $q^{(1)}$方向的直线 l 上的极小点，$w^{(3)}$是过 $w^{(1)}$由 $\{q^{(1)}, q^{(2)}\}$ 张成平面 π 上的极小点，……，$w^{(k+1)}$是过 $w^{(1)}$由 $\{q^{(1)}, q^{(2)}, \cdots, q^{(k)}\}$ 张成的线性流形 $\overline{W}_k$ 上的极小点. 下面来证明这一点.

定理 6.2.1 设目标函数为

$$f(w) = \frac{1}{2} w^T w + r^T w + \delta,$$

$\{q^{(1)}, q^{(2)}, \cdots, q^{(k)}\}$ 是 k（$k \leqslant n$）个非零的正交方向. 从任意的初始点 $w^{(1)}$出发，依次沿 $\{q^{(1)}, q^{(2)}, \cdots, q^{(k)}\}$ 作精确一维搜索，得到 $\{w^{(2)}, w^{(3)}, \cdots, w^{(k+1)}\}$，则 $w^{(k+1)}$ 是 $f(w)$ 在线性流形

$$\overline{W}_k = \left\{ w = w^{(1)} + \sum_{i=1}^{k} \alpha_i q^{(i)} \mid -\infty < \alpha_i < +\infty \right\} \tag{6.2.5}$$

上的唯一极小点. 特别当 $k = n$ 时，则 $w^{(n+1)}$是$f(w)$ 在整个空间上的唯一极小点.

证明 由于$f(w)$ 是特殊的正定二次函数，它在线性流形

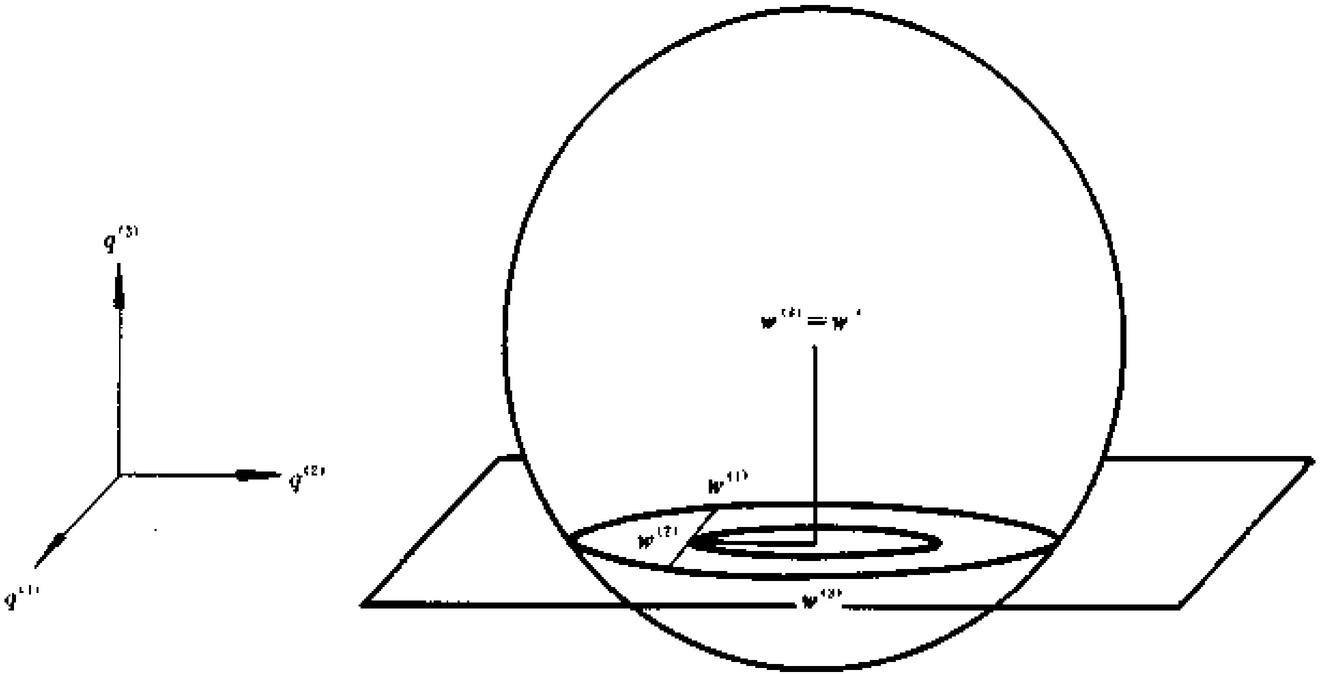


图 6.2.2　正交方向与特殊的正定二次函数(三维)

$\overline{W}_k$ 上有唯一的极小点 $\hat{w}^{(k)}$，并满足

$$\langle \nabla f(\hat{w}^{(k)}),\ q^{(i)} \rangle = 0,\ i = 1,\ 2,\ \cdots,\ k. \tag{6.2.6}$$

注意 $\hat{w}^{(k)}$是线性流形 w_k 上的点，应有

$$\hat{w}^{(k)} = w^{(1)} + \sum_{i=1}^{k} \hat{\alpha}_i q^{(i)}, \tag{6.2.7}$$

因此得到

$$\nabla f(\hat{w}^{(k)}) = \hat{w}^{(k)} + r = w^{(1)} + r + \sum_{i=1}^{k} \hat{\alpha}_i q^{(i)}.$$

由（6.2.6）式得到

$$\left(w^{(1)} + r + \sum_{j=1}^{k} \hat{\alpha}_j q^{(j)} \right)^T q^{(i)} = 0,$$

即

$$(w^{(1)} + r)^T q^{(i)} + \hat{\alpha}_i \| q^{(i)} \|^2 = 0,$$

所以

$$\hat{\alpha}_i = -\frac{1}{\| q^{(i)} \|^2} (w^{(1)} + r)^T q^{(i)}. \tag{6.2.8}$$

另一方面，按定理条件得到

$$w^{(k+1)} = w^{(1)} + \sum_{i=1}^{k} \alpha_i q^{(i)},$$

其中 α_i 为最优步长．由习题 4.6 的结论，则 α_i 为

$$\begin{aligned} \alpha_i &= -\frac{1}{\| q^{(i)} \|^2}(w^{(i)} + r)^T q^{(i)} \\ &= -\frac{1}{\| q^{(i)} \|^2}\left(w^{(1)} + r + \sum_{j=1}^{i-1} \alpha_j q^{(j)} \right)^T q^{(i)} \\ &= -\frac{1}{\| q^{(i)} \|^2}(w^{(1)} + r)^T q^{(i)}, \end{aligned}$$

表达式与（6.2.8）式相同，即 $w^{(k+1)} = \hat{w}^{(k)}$

下面考虑一般正定二次函数

$$f(x)=\frac{1}{2}x^TGx+r^Tx+\delta, \tag{6.2.9}$$

引进变换

$$w=\sqrt{G}x, \tag{6.2.10}$$

则有

$$\begin{aligned} f(x)&=\frac{1}{2}(\sqrt{G}x)^T(\sqrt{G}x)+r^T\sqrt{G}^{-1}\sqrt{G}x+\delta \\ &=\frac{1}{2}w^Tw+\tilde{r}^Tw+\delta=\tilde{f}(w). \end{aligned}$$

$\tilde{f}(w)$ 是特殊正定二次函数. 在 W 空间下的正交方向，在 X 空间下是什么方向呢？设 $q^{(1)}$与 $q^{(2)}$正交，作变换

$$q^{(i)}=\sqrt{G}d^{(i)} \quad 或 \quad d^{(i)}=\sqrt{G}^{-1}q^{(i)}, \tag{6.2.11}$$

所以

$$0=\langle q^{(1)}, q^{(2)}\rangle=\langle\sqrt{G}d^{(1)}, \sqrt{G}d^{(2)}\rangle=(d^{(1)})^TGd^{(2)},$$

由此给出如下定义.

定义 6.2.2 设 G 为 $n\times n$ 阶正定对称矩阵. 若 $d^{(1)}$、$d^{(2)}$ 满足

$$(d^{(1)})^TGd^{(2)}=0, \tag{6.2.12}$$

则称 $d^{(1)}$、$d^{(2)}$关于 G 共轭. 若 $d^{(1)}$, $d^{(2)}$, …, $d^{(k)}$ $(k\leqslant n)$ 两两关于 G 共轭，即

$$(d^{(i)})^TGd^{(j)}=0,\ i\neq j,\ i,\ j=1,\ 2,\ \cdots,\ k \tag{6.2.13}$$

则称 $d^{(1)}$, $d^{(2)}$, …, $d^{(k)}$为 G 的k 个共轭方向. 若 $d^{(i)}\neq 0$, $i=1, 2, \cdots, k$，则称为 G 的k 个非零共轭方向.

特别，当 $G=I$ 时，共轭方向就是正交方向.

由共轭方向的几何意义，可以把定理 6.2.1 作如下推广.

定理 6.2.2 （扩展子空间定理）设目标函数为

$$f(x)=\frac{1}{2}x^TGx+r^Tx+\delta,$$

$\{d^{(1)}, d^{(2)}, \cdots, d^{(k)}\}$ 是 G 的 k ($k \leqslant n$) 个非零的共轭方向，从任意的初始点 $x^{(1)}$ 出发，依次沿 $\{d^{(1)}, d^{(2)}, \cdots, d^{(k)}\}$ 作精确一维搜索，得到 $\{x^{(2)}, x^{(3)}, \cdots, x^{(k+1)}\}$，则 $x^{(k+1)}$ 是 $f(x)$ 在线性流形

$$X_k = \left\{x = x^{(1)} + \sum_{i=1}^{k} \alpha_i d^{(i)} \mid -\infty < \alpha_i < +\infty\right\}$$

上的唯一极小点. 特别当 $k = n$ 时，则 $x^{(n+1)}$ 是 $f(x)$ 在整个空间上的唯一极小点.

推论 在定理6.2.2的假设下，有

$$\langle \nabla f(x^{(k+1)}), d^{(i)} \rangle = 0, i = 1, 2, \cdots, k, \qquad (6.2.14)$$

证明 因为 $x^{(k+1)}$ 是线性流形 X_k 上的极小点，所以也是 $d^{(k)}$ 方向上的极小点，因此 (6.2.14) 式成立.

6.2.2 共轭梯度法的推导

共轭梯度法 (Conjugate Graient Method) 是最著名的共轭方向方法，它首先由 Hestenes 和 Stiefel 在 1952 年提出来作为解线性方程组的方法. 由于解线性方程组等价于极小化一个正定二次函数，故 1964 年 Fletcher 和 Reeves 提出了求解无约束问题的共轭梯度法.

设

$$f(x) = \frac{1}{2} x^T G x + r^T x + \delta,$$

其中 G 是正定对称矩阵. 由扩展子空间定理可知，若 $d^{(1)}, d^{(2)}, \cdots, d^{(n)}$ 为 n 个 G 共轭方向，那么从任意的初始点 $x^{(1)}$ 出发，至多作 n 次精确一维搜索，就可以得到目标函数唯一的极小点. 现根据这一思想导出对于正定二次函数的共轭梯度法.

任取初始点 $x^{(1)}$，若 $\nabla f(x^{(1)}) = 0$，则停止计算，$x^{(1)}$ 作为无约束问题的极小点. 当 $\nabla f(x^{(1)}) \neq 0$ 时，令

$$d^{(1)} = -\nabla f(x^{(1)}), \tag{6.2.15}$$

(因为此时还没有得到其它信息) 然后沿 $d^{(1)}$方向进行一维搜索，得到点 $x^{(2)}$. 若$\nabla f(x^{(2)}) \neq 0$ 时，令

$$d^{(2)} = -\nabla f(x^{(2)}) + \beta_1^{(2)} d^{(1)}, \tag{6.2.16}$$

并且使 $d^{(1)}$、$d^{(2)}$满足

$$(d^{(1)})^T G d^{(2)} = 0, \tag{6.2.17}$$

即 $d^{(1)}$、$d^{(2)}$关于 G 共轭. 将 (6.2.16) 式代入 (6.2.17) 式，可得到 $\beta_1^{(2)}$ 的表达式

$$\beta_1^{(2)} = \frac{(d^{(1)})^T G \nabla f(x^{(2)})}{(d^{(1)})^T G d^{(1)}}. \tag{6.2.18}$$

将 (6.2.18) 式得到的 $\beta_1^{(2)}$ 代入 (6.2.16) 式，这样得到的 $d^{(2)}$是与 $d^{(1)}$关于 G 共轭. 再从 $x^{(2)}$出发，沿 $d^{(2)}$作一维搜索，得到 $x^{(3)}$. 如此下去，……，假设在 $x^{(k)}$处，$\nabla f(x^{(k)}) \neq 0$，构造 $x^{(k)}$处的搜索方向 $d^{(k)}$如下：

$$d^{(k)} = -\nabla f(x^{(k)}) + \beta_1^{(k)} d^{(1)} + \beta_2^{(k)} d^{(2)} + \cdots + \beta_{k-1}^{(k)} d^{(k-1)}. \tag{6.2.19}$$

下面来确定系数 $\beta_i^{(k)}$ ($i = 1, 2, \cdots, k-1$). 由于要求构造的搜索方向是关于 G 共轭的，即满足

$$\begin{aligned}
0 &= (d^{(i)})^T G d^{(k)} \\
&= -(d^{(i)})^T G \nabla f(x^{(k)}) + \sum_{j=1}^{k-1} \beta_j^{(k)} (d^{(i)})^T G d^{(j)} \\
&= -(d^{(i)})^T G \nabla f(x^{(k)}) + \beta_i^{(k)} (d^{(i)})^T G d^{(i)}, i = 1,2,\cdots,k-1,
\end{aligned}$$

所以

$$\beta_i^{(k)} = \frac{(d^{(i)})^T G \nabla f(x^{(k)})}{(d^{(i)})^T G d^{(i)}}. \tag{6.2.20}$$

将 (6.2.20) 式得到的 $\beta_i^{(k)}$ 代入 (6.2.19) 式，这样得到的 $d^{(k)}$是与 $d^{(1)}$，$d^{(2)}$，…，$d^{(k-1)}$关于 G 共轭. 又由推导过程可知，前面已得到的搜索方向 $d^{(1)}$，$d^{(2)}$，…，$d^{(k-1)}$已是 G 的

$k-1$ 个共轭方向，所以 $\boldsymbol{d}^{(1)}$，$\boldsymbol{d}^{(2)}$，…，$\boldsymbol{d}^{(k)}$是 G 的k 个共轭方向.

由扩展子空间定理，当 $k=n$ 时，得到 n 个非零的G 共轭方向，$\boldsymbol{x}^{(n+1)}$为整个空间上的唯一极小点.

6.2.3 计算公式的简化

上面导出的共轭方向的计算公式（6.2.19）式和（6.2.20）式，只能用于正定二次函数．而且随着 k 的增大，存储量增多，因此上述两式不能用于一般的目标函数．下面利用正定二次函数的性质，对（6.2.19）式和（6.2.20）式进行化简，得到适用于一般可微的目标函数的计算公式.

由于构造的搜索方向是非零的 G 共轭方向，由定理 6.2.2 的推论，得到

$$\nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d}^{(i)}=0,\quad i=1,\ 2,\ \cdots,\ k-1.$$

结合（6.2.19）式，有

$$\begin{aligned}&\nabla f(\boldsymbol{x}^{(k)})^T\nabla f(\boldsymbol{x}^{(i)})\\&=\nabla f(\boldsymbol{x}^{(k)})^T(-\boldsymbol{d}^{(i)}+\beta_1^{(i)}\boldsymbol{d}^{(1)}+\cdots+\beta_{i-1}^{(i)}\boldsymbol{d}^{(i-1)})\\&=0,i=1,2,\cdots,k-1.\end{aligned}\tag{6.2.21}$$

下面计算 $\beta_i^{(k)}$（$i=1,\ 2,\ \cdots,\ k-1$）．为了方便起见，引进记号

$$\boldsymbol{s}^{(i)}=\boldsymbol{x}^{(i+1)}-\boldsymbol{x}^{(i)}=\alpha_i\boldsymbol{d}^{(i)},\tag{6.2.22}$$

所以

$$G\boldsymbol{s}^{(i)}=G\boldsymbol{x}^{(i+1)}-G\boldsymbol{x}^{(i)}=\nabla f(\boldsymbol{x}^{(i+1)})-\nabla f(\boldsymbol{x}^{(i)})\tag{6.2.23}$$

并由（6.2.21）式，有

$$\begin{aligned}(\boldsymbol{d}^{(i)})^TG\nabla f(\boldsymbol{x}^{(k)})&=\nabla f(\boldsymbol{x}^{(k)})^TG\boldsymbol{d}^{(i)}\\&=\frac{1}{\alpha_i}\nabla f(\boldsymbol{x}^{(k)})^TG\boldsymbol{s}^{(i)}\end{aligned}$$

$$= \frac{1}{\alpha_i}\nabla f(\boldsymbol{x}^{(k)})^T(\nabla f(\boldsymbol{x}^{(i+1)}) - \nabla f(\boldsymbol{x}^{(i)}))$$
$$= 0,\ i = 1,2,\cdots,k-2. \tag{6.2.24}$$

因此，(6.2.20) 式简化为

$$\beta_i^{(k)} = 0, \quad i = 1,\ 2,\ \cdots,\ k-2, \tag{6.2.25}$$

故 (6.2.19) 式可以写成

$$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}) + \beta_{k-1}\boldsymbol{d}^{(k-1)}, \tag{6.2.26}$$

（此时可不要 β_{k-1}的上标）其中

$$\beta_{k-1} = \frac{(\boldsymbol{d}^{(k-1)})^T G\nabla f(\boldsymbol{x}^{(k)})}{(\boldsymbol{d}^{(k-1)})^T G\boldsymbol{d}^{(k-1)}}. \tag{6.2.27}$$

再化简 β_{k-1}，类似于 (6.2.24) 式的推导过程，可以得到

$$(\boldsymbol{d}^{(k-1)})^T G\nabla f(\boldsymbol{x}^{(k)})$$
$$= \frac{1}{\alpha_{k-1}}\nabla f(\boldsymbol{x}^{(k)})^T(\nabla f(\boldsymbol{x}^{(k)}) - \nabla f(x^{(k-1)})) \tag{6.2.28}$$

和

$$(\boldsymbol{d}^{(k-1)})^T G\boldsymbol{d}^{(k-1)}$$
$$= \frac{1}{\alpha_{k-1}}(\boldsymbol{d}^{(k-1)})^T(\nabla f(\boldsymbol{x}^{(k)}) - \nabla f(\boldsymbol{x}^{(k-1)})), \tag{6.2.29}$$

注意到定理 6.2.2 的推论和 (6.2.26) 式有

$$\begin{aligned}
&(\boldsymbol{d}^{(k-1)})^T(\nabla f(\boldsymbol{x}^{(k)}) - \nabla f(\boldsymbol{x}^{(k-1)})) \\
&= -(\boldsymbol{d}^{(k-1)})^T\nabla f(\boldsymbol{x}^{(k-1)}) \\
&= (\nabla f(\boldsymbol{x}^{(k-1)}) - \beta_{k-2}\boldsymbol{d}^{(k-2)})^T\nabla f(\boldsymbol{x}^{(k-1)}) \\
&= \nabla f(\boldsymbol{x}^{(k-1)})^T\nabla f(\boldsymbol{x}^{(k-1)}),
\end{aligned} \tag{6.2.30}$$

结合 (6.2.28) 式、(6.2.29) 式和 (6.2.30) 式，得到

$$\beta_{k-1}=\frac{\nabla f(\boldsymbol{x}^{(k)})^T(\nabla f(\boldsymbol{x}^{(k)})-\nabla f(\boldsymbol{x}^{(k-1)}))}{\nabla f(\boldsymbol{x}^{(k-1)})^T\nabla f(\boldsymbol{x}^{(k-1)})},\tag{6.2.31}$$

或

$$\beta_{k-1}=\frac{\|\nabla f(\boldsymbol{x}^{(k)})\|^2}{\|\nabla f(\boldsymbol{x}^{(k-1)})\|^2}.\tag{6.2.32}$$

这样就得到了用于一般可微函数的共轭梯度法. 其搜索方向构造如下:

$$\begin{cases}\boldsymbol{d}^{(1)}=-\nabla f(\boldsymbol{x}^{(1)}),\\ \boldsymbol{d}^{(k)}=-\nabla f(\boldsymbol{x}^{(k)})+\beta_{k-1}\boldsymbol{d}^{(k-1)},\end{cases}\tag{6.2.33}$$

由(6.2.31)式和(6.2.33)式构造的计算公式称为 PRP(Polak-Ribiere-Polyak, 1969)公式, 相应的方法称为 PRP 算法. 由(6.2.32)式和(6.2.33)式构造的公式称为 FR(Fletcher-Reeves, 1964)公式, 相应的方法称为 FR 算法.

算法 6.2.1(FR 算法或 PRP 算法)

(1) 取初始点 $\boldsymbol{x}^{(1)}$, 置精度要求 ϵ, 置 $k=1$.

(2) 如果 $\|\nabla f(\boldsymbol{x}^{(k)})\|\leqslant\epsilon$, 则停止计算($\boldsymbol{x}^{(k)}$作为无约束问题的解); 否则置

$$\boldsymbol{d}^{(k)}=-\nabla f(\boldsymbol{x}^{(k)})+\beta_{k-1}\boldsymbol{d}^{(k-1)},$$

其中

$$\beta_{k-1}=\begin{cases}0, & k=1,\\ \dfrac{\|\nabla f(\boldsymbol{x}^{(k)})\|^2}{\|\nabla f(\boldsymbol{x}^{(k-1)})\|^2}, & k>1,\end{cases}$$

或

$$\beta_{k-1}=\begin{cases}0, & k=1,\\ \dfrac{\nabla f(\boldsymbol{x}^{(k)})^T(\nabla f(\boldsymbol{x}^{(k)})-\nabla f(\boldsymbol{x}^{(k-1)}))}{\nabla f(\boldsymbol{x}^{(k-1)})^T\nabla f(\boldsymbol{x}^{(k-1)})}, & k>1.\end{cases}$$

(3) 一维搜索. 求解一维问题

$$\min\phi(\alpha)=f(\boldsymbol{x}^{(k)}+\alpha\boldsymbol{d}^{(k)}),$$

得 α_k，置

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\alpha_k\boldsymbol{d}^{(k)},$$

(4) 置 $k=k+1$，转 (2).

【例 6.2.1】 用共轭梯度法（FR 算法）求解无约束问题

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

取 $\boldsymbol{x}^{(1)}=(0,0)^T$.

解

$$\nabla f(\boldsymbol{x})=\begin{pmatrix}3x_1-x_2-2\\x_2-x_1\end{pmatrix}=\begin{pmatrix}g_1\\g_2\end{pmatrix},$$

$$\nabla^2 f(\boldsymbol{x})=\begin{pmatrix}3&-1\\-1&1\end{pmatrix}=G,$$

一维搜索步长为

$$\alpha_k=\frac{(\boldsymbol{d}^{(k)})^T\nabla f(\boldsymbol{x}^{(k)})}{(\boldsymbol{d}^{(k)})^TG\boldsymbol{d}^{(k)}}=\frac{d_1g_1+d_2g_2}{3d_1^2+d_2^2-2d_1d_2}. \qquad (6.2.34)$$

取 $\boldsymbol{x}^{(1)}=(0,0)^T$，则 $\nabla f(\boldsymbol{x}^{(1)})=(-2,0)^T$，$\boldsymbol{d}^{(1)}=-\nabla f(\boldsymbol{x}^{(1)})=(2,0)^T$. 由 (6.2.34) 式得到步长 $\alpha_1=\frac{2^2}{3\cdot 2^2}=\frac{1}{3}$，因此

$$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\alpha_1\boldsymbol{d}^{(1)}=(0,0)^T+\frac{1}{3}(2,0)^T=\left(\frac{2}{3},0\right)^T.$$

再计算第二轮循环：因为 $\nabla f(\boldsymbol{x})^{(2)}=\left(0,-\frac{2}{3}\right)^T$，有

$$\beta_1=\frac{\|\nabla f(\boldsymbol{x}^{(2)})\|^2}{\|\nabla f(\boldsymbol{x}^{(1)})\|^2}=\frac{\left(\frac{2}{3}\right)^2}{2^2}=\frac{1}{9},$$

因此

$$\boldsymbol{d}^{(2)}=-\nabla f(\boldsymbol{x}^{(2)})+\beta_1\boldsymbol{d}^{(1)}=\left(0,\frac{2}{3}\right)^T+\frac{1}{9}(2,0)^T$$

$$= \left(\frac{2}{9}, \frac{2}{3}\right)^T.$$

步长为

$$\alpha_2 = -\frac{\frac{2}{9}\cdot 0 + \frac{2}{3}\cdot\left(-\frac{2}{3}\right)}{3\left(\frac{2}{9}\right)^2 + \left(\frac{2}{3}\right)^2 - 2\cdot\frac{2}{9}\cdot\frac{2}{3}} = \frac{3}{2},$$

由此得到

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} + \alpha_2 \boldsymbol{d}^{(2)} = \left(\frac{2}{3}, 0\right)^T + \frac{3}{2}\left(\frac{2}{9}, \frac{2}{3}\right)^T = (1, 1)^T.$$

计算梯度$\nabla f(\boldsymbol{x}^{(3)}) = (0,0)^T$, $\| \nabla f(\boldsymbol{x}^{(3)}) \| = 0$. 所以 $\boldsymbol{x}^{(3)} = (1,1)^T$ 为最优解，最优目标函数值为 $f(\boldsymbol{x}^{(3)}) = -1$.

另外，还有两个计算 β_{k-1}的公式

$$\beta_{k-1} = \frac{\nabla f(\boldsymbol{x}^{(k)})^T(\nabla f(\boldsymbol{x}^{(k)}) - \nabla f(\boldsymbol{x}^{(k-1)}))}{\boldsymbol{d}^{(k-1)T}(\nabla f(\boldsymbol{x}^{(k)}) - \nabla f(\boldsymbol{x}^{(k-1)}))},$$

称为 Growder-Wolfe 公式，和

$$\beta_{k-1} = \frac{\nabla f(\boldsymbol{x}^{(k)})^T \nabla f(\boldsymbol{x}^{(k)})}{\boldsymbol{d}^{(k-1)T} \nabla f(\boldsymbol{x}^{(k-1)})},$$

称为 Dixon 公式.

6.2.4 共轭方向的下降性和算法的二次终止性

共轭梯度法虽然是针对正定二次函数导出的，但仍适用于一般可微函数. 这里首先需要解决一个问题：共轭梯度法产生的搜索方向是否是下降方向？

定理 6.2.3 设 $f(\boldsymbol{x})$具有连续的一阶偏导数，并假设一维搜索是精确的，考虑用共轭梯度法求解无约束问题，若 $\nabla f(\boldsymbol{x}^{(k)}) \neq 0$，则搜索方向 $\boldsymbol{d}^{(k)}$是 $\boldsymbol{x}^{(k)}$处的下降方向.

证明 只需证明$\nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{d}^{(k)} < 0$. 注意到 $\boldsymbol{x}^{(k)}$是 $\boldsymbol{d}^{(k)}$方向上的精确极小点，有$\nabla f(\boldsymbol{x}^{(k)})^T \boldsymbol{d}^{(k-1)} = 0$，因此

$$\nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)})^T\nabla f(\boldsymbol{x}^{(k)}) + \beta_{k-1}\nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d}^{(k-1)}$$
$$= -\|\nabla f(\boldsymbol{x}^{(k)})\|^2 < 0. \tag{6.2.35}$$

定理 6.2.4 若一维搜索是精确的，则共轭梯度法具有二次终止性.

证明 考虑用共轭梯度法求解正定二次目标函数，若 $\nabla f(\boldsymbol{x}^{(k)})=0$，$(k\leqslant n)$，则 $x^{(k)}$ 为无约束问题的最优解；否则由算法得到的搜索方向 $\boldsymbol{d}^{(1)}$，$\boldsymbol{d}^{(2)}$，…，$\boldsymbol{d}^{(n)}$ 是共轭的，由扩展子空间定理可知，$\boldsymbol{x}^{(n+1)}$ 是无约束问题的最优解.

由定理 6.2.4 可知，对于正定二次函数，共轭梯度法至多 n 步终止. 如果算法在 n 步没有终止，则说明目标函数不是正定二次函数，或者说目标函数没有进入一个正定二次函数的区域，因此这种共轭已没有意义，此时搜索方向应重新开始，即令

$$\boldsymbol{d}^{(k)} = -\nabla f(\boldsymbol{x}^{(k)}).$$

算法每 n 步重新开始一次，称为 n 步重新开始策略. 实际计算表明，n 步重新开始的 FR 算法要优于原 FR 算法.

对于 PRP 算法，当 $\nabla f(\boldsymbol{x}^{(k)})\approx\nabla f(\boldsymbol{x}^{(k-1)})$ 时，有 $\beta_{k-1}\approx 0$，因此 $\boldsymbol{d}^{(k)}\approx-\nabla f(\boldsymbol{x}^{(k)})$，即自动重新开始. 试验表明，对于大型问题，PRP 算法优于 FR 算法.

§ 6.3 变 度 量 法

变度量法 (Variable Metric Method) 也称为拟 Newton 法 (Quasi-Newton Method)，是求解无约束最优化问题最有效的算法之一，它不需要求二阶 Hesse 矩阵，而只利用一阶导数来构造二阶信息的近似矩阵，从而该算法有较好的收敛性质.

6.3.1 变度量法的导出

在修正 Newton 法中，需要求解 Newton 方程

$$\nabla^2 f(\boldsymbol{x}^{(k)})d = -\nabla f(\boldsymbol{x}^{(k)})$$

来确定搜索方向. 但这一做法的主要缺点是需要计算 Hesse 矩阵, 计算最大. 要克服上述缺点, 最直观想法是: 不计算 Hesse 矩阵 $\nabla^2 f(\boldsymbol{x}^{(k)})$, 而用一个 "近似" 矩阵 $B^{(k)}$ 来代替 $\nabla^2 f(\boldsymbol{x}^{(k)})$. 其中 $B^{(k)}$ 是根据迭代过程中的某些信息得到的. 因此用方程组

$$B^{(k)}d = -\nabla f(\boldsymbol{x}^{(k)}) \tag{6.3.1}$$

的解 $\boldsymbol{d}^{(k)}$ 来作为搜索方向, 即用方程 (6.3.1) 来代替 Newton 方程. 这就得到一类新的算法, 称此类算法为变度量法. 由于 $B^{(k)}$ "近似" 于 $\nabla^2 f(\boldsymbol{x}^{(k)})$, 可以猜想, 该算法应有较好的收敛速率, 并且克服了计算量大的缺点. 那么问题的关键是如何构造矩阵 $B^{(k)}$? 首先, $B^{(k)}$ 不可能在数值上近似 $\nabla^2 f(\boldsymbol{x}^{(k)})$, 因为我们的目标是不计算 Hesse 矩阵 $\nabla^2 f(\boldsymbol{x}^{(k)})$. 那么, $B^{(k)}$ 只能在性质上近似 $\nabla^2 f(\boldsymbol{x}^{(k)})$. 因此, 先研究一下 Hesse 矩阵 $\nabla^2 f(\boldsymbol{x}^{(k)})$ 的性质.

考虑 $\nabla f(\boldsymbol{x})$ 在 $x^{(k+1)}$ 处的 Taylor 展开式

$$\nabla f(\boldsymbol{x}) \approx \nabla f(\boldsymbol{x}^{(k+1)}) + \nabla^2 f(\boldsymbol{x}^{(k+1)})(\boldsymbol{x} - \boldsymbol{x}^{(k+1)}), \tag{6.3.2}$$

取 $\boldsymbol{x} = \boldsymbol{x}^{(k)}$, 得到

$$f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)}) \approx \nabla^2 f(\boldsymbol{x}^{(k+1)})(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}). \tag{6.3.3}$$

记

$$\boldsymbol{s}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}, \boldsymbol{y}^{(k)} = \nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)}).$$

所以 (6.3.3) 式改写为

$$\nabla^2 f(\boldsymbol{x}^{(k+1)})\boldsymbol{s}^{(k)} \approx \boldsymbol{y}^{(k)}. \tag{6.3.4}$$

由于要求 $B^{(k)}$ 近似 $\nabla^2 f(\boldsymbol{x}^{(k)})$, 即 $B^{(k+1)}$ 近似于 $\nabla^2 f(\boldsymbol{x}^{(k+1)})$, 也就是说 $B^{(k+1)}$ 就满足 (6.3.4) 式, 并将 "≈" 改为 "=", 得到

$$B^{(k+1)}s^{(k)}=y^{(k)}, \tag{6.3.5}$$

称（6.3.5）式为拟 Newton 方程，并且假设 $B^{(k+1)}$是对称矩阵.

由（6.3.5）式并不能唯一地确定 $B^{(k+1)}$，因此还需要附加一些条件. 假设前面已得到的矩阵 $B^{(k)}$，并且矩阵 $B^{(k)}$已与 $\nabla^2 f(\boldsymbol{x}^{(k)})$ "近似". 一种自然的想法就是 $B^{(k+1)}$是由 $B^{(k)}$经过修正得到的，即

$$B^{(k+1)}=B^{(k)}+\Delta B^{(k)}, \tag{6.3.6}$$

而且修正项 $\Delta B^{(k)}$具有"简单"的形式. 当然，简单可以有不同的理解，这里"简单"是指矩阵的秩少. 当然秩最少是 1，因此首先导出秩 1 的修正公式.

设 $\Delta B^{(k)}=\boldsymbol{u}\boldsymbol{v}^T$，即

$$B^{(k+1)}=B^{(k)}+\boldsymbol{u}\boldsymbol{v}^T, \tag{6.3.7}$$

将（6.3.7）式代入拟 Newton 方程（6.3.5)，得到

$$\boldsymbol{y}^{(k)}=B^{(k+1)}\boldsymbol{s}^{(k)}=B^{(k)}\boldsymbol{s}^{(k)}+(\boldsymbol{v}^T\boldsymbol{s}^{(k)})\boldsymbol{u}, \tag{6.3.8}$$

所以

$$\boldsymbol{y}^{(k)}-B^{(k)}\boldsymbol{s}^{(k)}=(\boldsymbol{v}^T\boldsymbol{s}^{(k)})\boldsymbol{u}, \tag{6.3.9}$$

只需取 $\boldsymbol{u}=\boldsymbol{y}^{(k)}-B^{(k)}s^{(k)}$，$\boldsymbol{v}$ 满足

$$\boldsymbol{v}^T\boldsymbol{s}^{(k)}=1, \tag{6.3.10}$$

则（6.3.9）式成立. 由 $B^{(k)}$的对称性，并希望 $B^{(k+1)}$对称，即 $\Delta B^{(k)}$对称，因此只需 $\boldsymbol{v}$ 与 $\boldsymbol{u}$ 共线，即

$$\boldsymbol{v}=\lambda\boldsymbol{u}=\lambda(\boldsymbol{y}^{(k)}-B^{(k)}\boldsymbol{s}^{(k)}), \tag{6.3.11}$$

将（6.3.11）式代入（6.3.10）式，得到

$$1=\lambda(\boldsymbol{y}^{(k)}-B^{(k)}\boldsymbol{s}^{(k)})^T\boldsymbol{s}^{(k)},$$

所以

$$\lambda=\frac{1}{(\boldsymbol{y}^{(k)}-B^{(k)}\boldsymbol{s}^{(k)})^T\boldsymbol{s}^{(k)}}. \tag{6.3.12}$$

结合（6.3.11）式和（6.3.7）式得到

$$B^{(k+1)}=B^{(k)}+\frac{(y^{(k)}-B^{(k)}s^{(k)})(y^{(k)}-B^{(k)}s^{(k)})^T}{(y^{(k)}-B^{(k)}s^{(k)})^T s^{(k)}},\tag{6.3.13}$$

称 (6.3.13) 式为对 $B^{(k)}$的对称秩 1 修正公式. 由此可导出相应的变度量法.

算法 6.3.1 (对称秩 1 修正公式的变度量法)

(1) 取初始点 $\boldsymbol{x}^{(1)}$, 置初始矩阵 $B^{(1)}$ ($=I$), 置精度要求 ϵ, 置 $k=1$.

(2) 如果 $\|\nabla f(\boldsymbol{x}^{(k)})\|\leqslant\epsilon$, 则停止计算 ($\boldsymbol{x}^{(k)}$作为无约束问题的解); 否则求解线性方程组

$$B^{(k)}\boldsymbol{d}=-\nabla f(\boldsymbol{x}^{(k)}),$$

得到 $\boldsymbol{d}^{(k)}$.

(3) 一维搜索. 求解一维问题

$$\min\phi(\alpha)=f(\boldsymbol{x}^{(k)}+\alpha\boldsymbol{d}^{(k)}),$$

得 α_k, 置

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\alpha_k\boldsymbol{d}^{(k)}.$$

(4) 修正 $B^{(k)}$, 记

$$\boldsymbol{s}^{(k)}=\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)},\quad \boldsymbol{y}^{(k)}=\nabla f(\boldsymbol{x}^{(k+1)})-\nabla f(\boldsymbol{x}^{(k)}),$$

置

$$B^{(k+1)}=B^{(k)}+\frac{(y^{(k)}-B^{(k)}s^{(k)})(y^{(k)}-B^{(k)}s^{(k)})^T}{(y^{(k)}-B^{(k)}s^{(k)})^T s^{(k)}}.$$

(5) 置 $k=k+1$, 转 (2).

6.3.2 几种常用的变度量法

前面介绍了对称秩 1 修正公式的变度量法. 但秩 1 公式有两个主要缺点.

(1) 要求 $(\boldsymbol{y}^{(k)}-B^{(k)}\boldsymbol{s}^{(k)})^T\boldsymbol{s}^{(k)}\neq 0$, 若出现 $(\boldsymbol{y}^{(k)}-B^{(k)}\boldsymbol{s}^{(k)})^T\boldsymbol{s}^{(k)}=0$, 则无法进一步计算.

(2) 不能保证正定性的传递，即 $B^{(k)}$正定，由秩 1 公式得到的 $B^{(k+1)}$不一定正定. 只有当 $(\boldsymbol{y}^{(k)}-B^{(k)}\boldsymbol{s}^{(k)})^T\boldsymbol{s}^{(k)}>0$，才能保证 $B^{(k+1)}$正定. 而这个条件往往在算法中很难实现，即使有 $(\boldsymbol{y}^{(k)}-B^{(k)}\boldsymbol{s}^{(k)})^T\boldsymbol{s}^{(k)}>0$，它也可能很小，会产生较大的舍入误差，数值计算效果不好. 这样使秩 1 校正公式在应用中受到限制.

为了克服秩 1 公式的缺点，考虑对称秩 2 修正公式. 设

$$\Delta B^{(k)}=\boldsymbol{u}^{(1)}(\boldsymbol{v}^{(1)})^T+\boldsymbol{u}^{(2)}(\boldsymbol{v}^{(2)})^T, \qquad (6.3.14)$$

即

$$B^{(k+1)}=B^{(k)}+\boldsymbol{u}^{(1)}(\boldsymbol{v}^{(1)})^T+\boldsymbol{u}^{(2)}(\boldsymbol{v}^{(2)})^T, \qquad (6.3.15)$$

将 (6.3.15) 式代入拟 Newton 方程得到

$$B^{(k)}\boldsymbol{s}^{(k)}+((\boldsymbol{v}^{(1)})^T\boldsymbol{s}^{(k)})\boldsymbol{u}^{(1)}+((\boldsymbol{v}^{(2)})^T\boldsymbol{s}^{(k)})\boldsymbol{u}^{(2)}=\boldsymbol{y}^{(k)}, \qquad (6.3.16)$$

与秩 1 公式的推导类似，只需取 $\boldsymbol{u}^{(i)}$，$\boldsymbol{v}^{(i)}$ $(i=1,2)$ 满足

$$\begin{cases}\boldsymbol{u}^{(1)}=\boldsymbol{y}^{(k)},\\ (\boldsymbol{v}^{(1)})^T\boldsymbol{s}^{(k)}=1,\end{cases}\qquad \begin{cases}\boldsymbol{u}^{(2)}=-B^{(k)}\boldsymbol{s}^{(k)},\\ (\boldsymbol{v}^{(2)})^T\boldsymbol{s}^{(k)}=1,\end{cases} \qquad (6.3.17)$$

则 (6.3.16) 式成立. 再由 $B^{(k)}$的对称性和类似于秩 1 公式的推导过程，得到

$$B^{(k+1)}=B^{(k)}-\frac{B^{(k)}\boldsymbol{s}^{(k)}(\boldsymbol{s}^{(k)})^TB^{(k)}}{(\boldsymbol{s}^{(k)})^TB^{(k)}\boldsymbol{s}^{(k)}}+\frac{\boldsymbol{y}^{(k)}(\boldsymbol{y}^{(k)})^T}{(\boldsymbol{y}^{(k)})^T\boldsymbol{s}^{(k)}}, \qquad (6.3.18)$$

称 (6.3.18) 式为对 $B^{(k)}$的对称秩 2 修正公式，也称为 BFGS 公式 (Brogden-Fletcher-Goldfarb-Shanno 公式). 该公式是他们在 1970 年提出来的，相应的算法称为 BFGS 算法.

算法 6.3.2 (BFGS 算法)

(1) 取初始点 $\boldsymbol{x}^{(1)}$，置初始矩阵 $B^{(1)}$ $(=I)$，置精度要求 ε，置 $k=1$.

(2) 如果 $\|\nabla f(\boldsymbol{x}^{(k)})\|\leqslant\varepsilon$，则停止计算 ($\boldsymbol{x}^{(k)}$作为无约束问题的解)；否则求解线性方程组

$$B^{(k)}\boldsymbol{d} = -\nabla f(\boldsymbol{x}^{(k)}),$$

得到 $\boldsymbol{d}^{(k)}$.

(3) 一维搜索. 求解一维问题

$$\min\phi(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)}),$$

得 α_k，置

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}.$$

(4) 修正 $B^{(k)}$，记

$$\boldsymbol{s}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)},\quad \boldsymbol{y}^{(k)} = \nabla f(\boldsymbol{x}^{(k+1)}) - \nabla f(\boldsymbol{x}^{(k)}),$$

置

$$B^{(k+1)} = B^{(k)} - \frac{B^{(k)}\boldsymbol{s}^{(k)}\ (\boldsymbol{s}^{(k)})^T B^{(k)}}{(\boldsymbol{s}^{(k)})^T B^{(k)}\boldsymbol{s}^{(k)}} + \frac{\boldsymbol{y}^{(k)}\ (\boldsymbol{y}^{(k)})^T}{(\boldsymbol{y}^{(k)})^T\boldsymbol{s}^{(k)}}.$$

(5) 置 $k=k+1$，转 (2).

【例 6.3.1】 用变度量法（BFGS 算法）求解无约束问题

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

取 $\boldsymbol{x}^{(1)} = (0,\ 0)^T$.

解

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix},\quad \nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} = G,$$

一维搜索步长为

$$\alpha_k = -\frac{(\boldsymbol{d}^{(k)})^T\nabla f(\boldsymbol{x}^{(k)})}{(\boldsymbol{d}^{(k)})^T G\boldsymbol{d}^{(k)}} = -\frac{\boldsymbol{d}_1 g_1 + \boldsymbol{d}_2 g_2}{3\boldsymbol{d}_1^2 + d_2^2 - 2d_1d_2}.$$

(6.3.19)

取 $\boldsymbol{x}^{(1)} = (0,\ 0)^T$，则 $\nabla f(\boldsymbol{x}^{(1)}) = (-2,0)^T$，$B^{(1)} = I$，所以

$$\boldsymbol{d}^{(1)} = -(B^{(1)})^{-1}\nabla f(\boldsymbol{x}^{(1)}) = -\nabla f(\boldsymbol{x}^{(1)}) = (2,0)^T,$$

$$\alpha_1 = \frac{2^2}{3.2^2} = \frac{1}{3},$$

因此

$$x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = (0,\ 0)^T + \frac{1}{3}(2,\ 0)^T = \left(\frac{2}{3},\ 0\right)^T.$$

再计算第二轮循环：因为$\nabla f(x^{(2)}) = \left(0,\ -\frac{2}{3}\right)^T$，有

$$s^{(1)} = x^{(2)} - x^{(1)} = \left(\frac{2}{3},\ 0\right)^T - (0,\ 0)^T = \left(\frac{2}{3},\ 0\right)^T,$$

$$\begin{aligned} y^{(1)} &= \nabla f(x^{(2)}) - \nabla f(x^{(1)}) \\ &= \left(0, \frac{2}{3}\right)^T - (-2, 0)^T = \left(2, \frac{2}{3}\right)^T, \end{aligned}$$

$$(s^{(1)})^T B^{(1)} s^{(1)} = \left(\frac{2}{3}, 0\right) \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} = \frac{4}{9},$$

$$(y^{(1)})^T s^{(1)} = \left(2, -\frac{2}{3}\right) \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} = \frac{4}{3},$$

$$B^{(1)} s^{(1)} (s^{(1)})^T B^{(1)} = \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix} \left(\frac{2}{3}, 0\right) = \begin{pmatrix} \frac{4}{9} & 0 \\ 0 & 0 \end{pmatrix},$$

$$y^{(1)} (y^{(1)})^T = \begin{pmatrix} 2 \\ -\frac{2}{3} \end{pmatrix} \left(2, -\frac{2}{3}\right) = \begin{pmatrix} 4 & -\frac{4}{3} \\ -\frac{4}{3} & \frac{4}{9} \end{pmatrix},$$

所以

$$\begin{aligned} B^{(2)} &= B^{(1)} - \frac{B^{(1)} s^{(1)} (s^{(1)})^T B^{(1)}}{(s^{(1)})^T B^{(1)} s^{(1)}} + \frac{y^{(1)} (y^{(1)})^T}{(y^{(1)})^T s^{(1)}} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{9}{4} \begin{pmatrix} \frac{4}{9} & 0 \\ 0 & 0 \end{pmatrix} + \frac{3}{4} \begin{pmatrix} 4 & -\frac{4}{3} \\ -\frac{4}{3} & \frac{4}{9} \end{pmatrix} \\ &= \begin{pmatrix} 3 & -1 \\ -1 & \frac{4}{3} \end{pmatrix}. \end{aligned}$$

解方程组

$$B^{(2)}\boldsymbol{d}=-\nabla f\ (\boldsymbol{x}^{(2)}),$$

即

$$\begin{pmatrix}3 & -1\\ -1 & \frac{4}{3}\end{pmatrix}\begin{pmatrix}\boldsymbol{d}_1\\ \boldsymbol{d}_2\end{pmatrix}=\begin{pmatrix}0\\ \frac{2}{3}\end{pmatrix},$$

得 $\boldsymbol{d}_1=\frac{2}{9}$，$\boldsymbol{d}_2=\frac{2}{3}$，即 $\boldsymbol{d}^{(2)}=\left(\frac{2}{9},\ \frac{2}{3}\right)^T$. 步长为

$$\alpha_2=-\frac{\frac{2}{9}\cdot 0+\frac{2}{3}\cdot\left(-\frac{2}{3}\right)}{3\left(\frac{2}{9}\right)^2+\left(\frac{2}{3}\right)^2-2\cdot\frac{2}{9}\cdot\frac{2}{3}}=\frac{3}{2},$$

由此得到

$$\boldsymbol{x}^{(3)}=\boldsymbol{x}^{(2)}+\alpha_2\boldsymbol{d}^{(2)}=\left(\frac{2}{3},\ 0\right)^T+\frac{3}{2}\left(\frac{2}{9},\ \frac{2}{3}\right)^T=\ (1,\ 1)^T,$$

计算梯度 $\nabla f(\boldsymbol{x}^{(3)})\ =\ (0,\ 0)^T$，$\|f(\boldsymbol{x}^{(3)})\ \|=0$. 所以 $\boldsymbol{x}^{(3)}=\ (1,\ 1)^T$ 为最优解，最优目标函数值为 $f\ (\boldsymbol{x}^{(3)})\ =-1$.

下面从另一角度去构造变度量法. 将 Newton 方程改写为

$$\boldsymbol{d}^{(k)}=-(\nabla^2 f(\boldsymbol{x}^{(k)}))^{-1}\nabla f(\boldsymbol{x}^{(k)}),$$

一种直观的想法是不去构造近似于 $\nabla^2 f(\boldsymbol{x}^{(k)})$ 的矩阵 $B^{(k)}$，而是构造近似于 $(\nabla^2 f(\boldsymbol{x}^{(k)}))^{-1}$ 的矩阵 $H^{(k)}$. 这样计算 $\boldsymbol{d}^{(k)}$ 不需要求解线性方程组，而直接计算

$$\boldsymbol{d}^{(k)}=-H^{(k)}\nabla f(\boldsymbol{x}^{(k)}). \tag{6.3.20}$$

类似于 $B^{(k)}$ 的推导过程，相应的拟 Newton 方程为

$$H^{(k+1)}\boldsymbol{y}^{(k)}=\boldsymbol{s}^{(k)}. \tag{6.3.21}$$

类似地，可以构造出对 $H^{(k)}$ 的对称秩 1 修正公式

$$H^{(k+1)}=H^{(k)}+\frac{(\boldsymbol{s}^{(k)}-H^{(k)}\boldsymbol{y}^{(k)})\ (\boldsymbol{s}^{(k)}-H^{(k)}\boldsymbol{y}^{(k)})^T}{(\boldsymbol{s}^{(k)}-H^{(k)}\boldsymbol{y}^{(k)})^T\boldsymbol{y}^{(k)}} \tag{6.3.22}$$

和相应的对 $H^{(k)}$ 的对称秩 2 修正公式

$$H^{(k+1)}=H^{(k)}-\frac{H^{(k)}\boldsymbol{y}^{(k)}\ (\boldsymbol{y}^{(k)})^{T}H^{(k)}}{(\boldsymbol{y}^{(k)})^{T}H^{(k)}\boldsymbol{y}^{(k)}}+\frac{\boldsymbol{s}^{(k)}\ (\boldsymbol{s}^{(k)})^{T}}{(\boldsymbol{s}^{(k)})^{T}\boldsymbol{y}^{(k)}},\tag{6.3.23}$$

称（6.3.23）式为 DFP 公式（Daviden-Fletcher-Powell 公式）. 该公式是由 Daviden 在 1959 年提出来的，后由 Fletcher 和 Powell 在 1963 年予以简化，相应的算法称为 DFP 算法.

算法 6.3.3（DFP 算法）

（1）取初始点 $\boldsymbol{x}^{(1)}$，置初始矩阵 $H^{(1)}$（$=I$），置精度要求 ε，置 $k=1$.

（2）如果 $\|\nabla f(\boldsymbol{x}^{(k)})\|\leqslant\varepsilon$，则停止计算（$\boldsymbol{x}^{(k)}$作为无约束问题的解）；否则计算搜索方向

$$\boldsymbol{d}^{(k)}=-H^{(k)}\nabla f(\boldsymbol{x}^{(k)}).$$

（3）一维搜索. 求解一维问题

$$\min\phi(\alpha)=f(\boldsymbol{x}^{(k)}+\alpha\boldsymbol{d}^{(k)}),$$

得 α_k，置

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\alpha_k\boldsymbol{d}^{(k)}.$$

（4）修正 $H^{(k)}$，记

$$\boldsymbol{s}^{(k)}=\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)},\boldsymbol{y}^{(k)}=\nabla f(\boldsymbol{x}^{(k+1)})-\nabla f(\boldsymbol{x}^{(k)}),$$

置

$$H^{(k+1)}=H^{(k)}-\frac{H^{(k)}\boldsymbol{y}^{(k)}(\boldsymbol{y}^{(k)})^{T}H^{(k)}}{(\boldsymbol{y}^{(k)})^{T}H^{(k)}\boldsymbol{y}^{(k)}}+\frac{\boldsymbol{s}^{(k)}(\boldsymbol{s}^{(k)})^{T}}{(\boldsymbol{s}^{(k)})^{T}\boldsymbol{y}^{(k)}}.$$

（5）置 $k=k+1$，转（2）.

【例 6.3.2】 用变度量法（DBF 算法）求解无约束问题

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

取 $\boldsymbol{x}^{(1)}=(0,\ 0)^T$.

解

$$\nabla f(\boldsymbol{x})=\begin{pmatrix}3x_1-x_2-2\\x_2-x_1\end{pmatrix}=\begin{pmatrix}g_1\\g_2\end{pmatrix}.$$

$$\nabla^2 f(\boldsymbol{x})=\begin{pmatrix}3 & -1\\ -1 & 1\end{pmatrix}=G,$$

一维搜索步长为

$$\alpha_k=-\frac{(\boldsymbol{d}^{(k)})^T\nabla f(\boldsymbol{x}^{(k)})}{(\boldsymbol{d}^{(k)})^T G\boldsymbol{d}^{(k)}}=-\frac{d_1g_1+d_2g_2}{3d_1^2+d_2^2-2d_1d_2}.$$

取 $\boldsymbol{x}^{(1)}=(0,0)^T$,则$\nabla f(\boldsymbol{x}^{(1)})=(-2,0)^T$，$H^{(1)}=I$，所以

$$\boldsymbol{d}^{(1)}=-H^{(1)}\nabla f(\boldsymbol{x}^{(1)})=-\nabla f(\boldsymbol{x}^{(1)})=(2,0)^T,$$

$$\alpha_1=\frac{2^2}{3\cdot 2^2}=\frac{1}{3},$$

因此

$$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\alpha_2\boldsymbol{d}^{(1)}=(0,0)^T+\frac{1}{3}(2,0)^T=\left(\frac{2}{3},0\right)^T.$$

再计算第二轮循环：因为$\nabla f(\boldsymbol{x}^{(2)})=\left(0,-\frac{2}{3}\right)^T$，有

$$\boldsymbol{s}^{(1)}=\boldsymbol{x}^{(2)}-\boldsymbol{x}^{(1)}=\left(\frac{2}{3},0\right)^T-(0,0)^T=\left(\frac{2}{3},0\right)^T,$$

$$\begin{aligned}\boldsymbol{y}^{(1)}&=\nabla f(\boldsymbol{x}^{(2)})-\nabla f(\boldsymbol{x}^{(1)})\\&=\left(0,-\frac{2}{3}\right)^T-(-2,0)^T\\&=\left(2,-\frac{2}{3}\right)^T,\end{aligned}$$

$$(\boldsymbol{y}^{(1)})^T H^{(1)}\boldsymbol{y}^{(1)}=\left(2,-\frac{2}{3}\right)\begin{pmatrix}2\\ -\frac{2}{3}\end{pmatrix}=\frac{40}{9},$$

$$(\boldsymbol{s}^{(1)})^T\boldsymbol{y}^{(1)}=\left(\frac{2}{3},0\right)\begin{pmatrix}2\\ -\frac{2}{3}\end{pmatrix}=\frac{4}{3},$$

$$H^{(1)}\boldsymbol{y}^{(1)}(\boldsymbol{y}^{(1)})^T H^{(1)}=\begin{pmatrix}2\\ -\frac{2}{3}\end{pmatrix}\left(2,-\frac{2}{3}\right)=\begin{pmatrix}4 & -\frac{4}{3}\\ -\frac{4}{3} & \frac{4}{9}\end{pmatrix},$$

$$s^{(1)}(s^{(1)})^T = \begin{bmatrix} \frac{2}{3} \\ 0 \end{bmatrix} \left(\frac{2}{3}, 0\right) \begin{bmatrix} \frac{4}{9} & 0 \\ 0 & 0 \end{bmatrix},$$

所以

$$H^{(2)} = H^{(1)} - \frac{H^{(1)} y^{(1)} (y^{(1)})^T H^{(1)}}{(y^{(1)})^T H^{(1)} y^{(1)}} + \frac{s^{(1)}(s^{(1)})^T}{(s^{(1)})^T y^{(1)}}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{9}{40} \begin{bmatrix} 4 & -\frac{3}{4} \\ -\frac{4}{3} & \frac{4}{9} \end{bmatrix} + \frac{3}{4} \begin{bmatrix} \frac{4}{9} & 0 \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{13}{30} & \frac{3}{10} \\ \frac{3}{10} & \frac{9}{10} \end{bmatrix},$$

得到

$$d^{(2)} = -H^{(2)} \nabla f(x^{(2)}) = \begin{bmatrix} \frac{13}{30} & \frac{3}{10} \\ \frac{3}{10} & \frac{9}{10} \end{bmatrix} \begin{bmatrix} 0 \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{5} \\ \frac{3}{5} \end{bmatrix},$$

步长为

$$\alpha_2 = -\frac{\frac{2}{3} \cdot \left(-\frac{3}{5}\right)}{3\left(\frac{1}{5}\right)^2 + \left(\frac{3}{5}\right)^2 - 2 \cdot \frac{1}{5} \cdot \frac{3}{5}} = \frac{5}{3},$$

由此得到

$$x^{(3)} = x^{(2)} + \alpha_2 d^{(2)} = \left(\frac{2}{3},\ 0\right)^T + \frac{5}{3}\left(\frac{1}{5},\ \frac{3}{5}\right)^T = (1,\ 1)^T.$$

计算梯度 $\nabla f(x)^{(3)} = (0,\ 0)^T$，$\| f(x)^{(3)} \| = 0$. 所以 $x^{(3)} = (1,\ 1)^T$ 为最优解，最优目标函数值为 $f(x)^{(3)} = -1$.

前面介绍了两种变度量法——BFGS 算法和 DFP 算法. 对于正定二次函数，两种算法具有相同的效果. 但对于一般可微函

数（非正定二次函数）两者效果并不相同．一般认为，BFGS 算法在收敛性质和数值计算方面均优于 DFP 算法．另一方面，在计算过程中，DFP 算法不必求解线性方程组，这一点又优于 BFGS 算法．能否让 BFGS 算法在求搜索方向时也不求解线性方程组呢？那么，首先要解决的问题是找出由 $(B^{(k)})^{-1}$ 到 $(B^{(k+1)})^{-1}$ 的修正公式．

定理 6.3.1 （Sherman-Morrison 公式）设 A 是 n 阶可逆矩阵，$\boldsymbol{u}$ 和 $\boldsymbol{v}$ 均为 n 维向量，若 $1+\boldsymbol{v}^TA^{-1}\boldsymbol{u}\neq 0$，则扰动后的矩阵 $A+\boldsymbol{u}\boldsymbol{v}^T$ 也可逆，且

$$(A+\boldsymbol{u}\boldsymbol{v}^T)^{-1}=A^{-1}-\frac{A^{-1}\boldsymbol{u}\boldsymbol{v}^TA^{-1}}{1+\boldsymbol{v}^TA^{-1}\boldsymbol{u}}. \tag{6.3.24}$$

证明 直接在（6.3.24）式两端左乘 $(A+\boldsymbol{u}\boldsymbol{v}^T)$ 即可．

考虑 $H^{(k)}=(B^{(k)})^{-1}$，$H^{(k+1)}=(B^{(k+1)})^{-1}$，对 BFGS 公式用两次 Sherman-Morrison 公式，就可导出对 $H^{(k)}$ 的 BFGS 公式

$$H^{(k+1)}=\left(I-\frac{\boldsymbol{s}^{(k)}(\boldsymbol{y}^{(k)})^T}{(\boldsymbol{y}^{(k)})^T\boldsymbol{s}^{(k)}}\right)H^{(k)}\left(I-\frac{\boldsymbol{s}^{(k)}(\boldsymbol{y}^{(k)})^T}{(\boldsymbol{y}^{(k)})^T\boldsymbol{s}^{(k)}}\right)^T+\frac{\boldsymbol{s}^{(k)}(\boldsymbol{s}^{(k)})^T}{(\boldsymbol{y}^{(k)})^T\boldsymbol{s}^{(k)}}. \tag{6.3.25}$$

利用（6.3.25）式，可以得到对 $H^{(k)}$ 的 BFGS 算法．

算法 6.3.3 （对 $H^{(k)}$ 的 BFGS 算法）

(1) 取初始点 $\boldsymbol{x}^{(1)}$，置初始矩阵 $H^{(1)}$ $(=I)$，置精度要求 ϵ，置 $k=1$．

(2) 如果 $\|\nabla f(\boldsymbol{x}^{(k)})\|\leqslant\epsilon$，则停止计算（$\boldsymbol{x}^{(k)}$ 作为无约束问题的解）；否则计算搜索方向

$$\boldsymbol{d}^{(k)}=-H^{(k)}\nabla f(\boldsymbol{x}^{(k)}).$$

(3) 一维搜索．求解一维问题

$$\min\phi(\alpha)=f(\boldsymbol{x}^{(k)}+\alpha\boldsymbol{d}^{(k)}),$$

得 α_k，置

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}.$$

(4) 修正 $H^{(k)}$，记

$$s^{(k)} = x^{(k+1)} - x^{(k)}, y^{(k)} = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}),$$

置

$$H^{(k+1)} = \left(I - \frac{s^{(k)}(y^{(k)})^T}{(y^{(k)})^T s^{(k)}}\right) H^{(k)} \left(I - \frac{s^{(k)}(y^{(k)})^T}{(y^{(k)})^T s^{(k)}}\right) + \frac{s^{(k)}(s^{(k)})^T}{(y^{(k)})^T s^{(k)}}.$$

(5) 置 $k = k+1$，转 (2).

§6.4 变度量法的基本性质

在这一节中主要讨论变度量法的基本性质，这里主要讨论两个性质：

(1) 对于一般可微函数，变度量法产生的搜索方向是下降方向，即搜索方向的下降性；

(2) 变度量法具有算法的二次终止性.

6.3.1 算法的下降性

定理 6.4.1 设矩阵 $B^{(k)}$（或 $H^{(k)}$）是正定对称矩阵，且 $(s^{(k)})^T y^{(k)} > 0$，则由 BFGS 公式（或 DFP 公式）构造的 $B^{(k+1)}$（或 $H^{(k+1)}$）是正定对称的.

证明 仅对 BFGS 公式进行证明即可.

由于 $(s^{(k)})^T y^{(k)} > 0$，蕴含着 $s^{(k)} \neq 0$，再由 $B^{(k)}$ 的正定性，得到

$$(s^{(k)})^T B^{(k)} s^{(k)} > 0, \tag{6.4.1}$$

因此 BFGS 公式 (6.3.18) 有意义. 对任意的 $x \in R^n$，考察

$$x^TB^{(k+1)}x = x^TB^{(k)}x - \frac{(x^TB^{(k)}s^{(k)})^2}{(s^{(k)})^TB^{(k)}s^{(k)}} + \frac{(x^Ty^{(k)})^2}{(y^{(k)})^Ts^{(k)}}$$

$$= \frac{(x^TB^{(k)}x)((s^{(k)})^TB^{(k)}s^{(k)}) - (x^TB^{(k)}s^{(k)})^2}{(s^{(k)})^TB^{(k)}s^{(k)}}$$

$$+ \frac{(x^Ty^{(k)})^2}{(y^{(k)})^Ts^{(k)}}. \tag{6.4.2}$$

由广义 Cauchy-Schwarz 不等式（见习题 6.12），知

$$(x^TB^{(k)}x)((s^{(k)})^TB^{(k)}s^{(k)}) \geqslant (x^TB^{(k)}s^{(k)})^2, \tag{6.4.3}$$

且等式成立的充分必要条件是：x 与 $s^{(k)}$共线，即存在着 $\lambda \neq 0$，使得 $x=\lambda s^{(k)}$. 因此，(6.4.2) 式等式右端大于等于 0. 当 $x=\lambda s^{(k)}$时，(6.4.2) 式等于右端的 第一项为 0，而第二项大于 0；否则第一项大于 0，而第二项大于等于 0，因此 $x^TB^{(k+1)}x>0$. 即 $B^{(k+1)}$正定，对称性是显然的.

推论 若 $d^{(k)}$是下降方向，且一维搜索是精确的，并设 $B^{(k)}$（或 $H^{(k)}$）是正定对称矩阵，则由 BFGS 公式（或 DFP 公式）构造的 $B^{(k+1)}$（或 $H^{(k+1)}$）是正定对称的.

证明 只需证明 $(s^{(k)})^Ty^{(k)}>0$. 注意到一维搜索是精确的，有 $s^{(k)T}\nabla f(x^{(k+1)})=0$ 和步长 $\alpha_k>0$，有

$$\begin{aligned}(s^{(k)})^Ty^{(k)} &= (s^{(k)})^T(\nabla f(x^{(k+1)}) - \nabla f(x^{(k)})) \\ &= -(s^{(k)})^T\nabla f(x^{(k)}) \\ &= -\alpha_k(d^{(k)})^T\nabla f(x^{(k)}) > 0.\end{aligned} \tag{6.4.4}$$

定理 6.4.2 若用 BFGS 算法（或 DFP 算法）求解无约束问题. 设初始矩阵 $B^{(1)}$（或 $H^{(1)}$）是正定对称矩阵，且一维搜索是精确的，若 $\nabla f(x^{(k)})\neq 0$，则产生的搜索方向 $d^{(k)}$是下降方向.

证明 仅对 BFGS 公式进行证明. 用数学归纳法证明：

(1) $B^{(k)}$正定；

(2) $\boldsymbol{d}^{(k)T}\nabla f(\boldsymbol{x}^{(k)}) < 0$.

当 $k=1$ 时，由条件知 $B^{(1)}$正定，因此

$$\nabla f(\boldsymbol{x}^{(1)})^T\boldsymbol{d}^{(1)} = -\nabla f(\boldsymbol{x}^{(1)})^T(B^{(1)})^{-1}\nabla f(\boldsymbol{x}^{(1)})<0.$$

假设当 $k=m$ 时，命题为真. 即 $B^{(m)}$ 正定，$\boldsymbol{d}^{(m)T}\nabla f(\boldsymbol{x}^{(m)}) < 0$，由定理 6.4.1 的推论得到，$B^{(m+1)}$是正定对称矩阵，因此有

$$\nabla f(\boldsymbol{x}^{(m+1)})^T\boldsymbol{d}^{(m+1)} = -\nabla f(\boldsymbol{x}^{(m+1)})^T(B^{(m+1)})^{-1}\nabla f(\boldsymbol{x}^{(m+1)})<0,$$

因此，当 $k=m+1$ 时，命题成立.

由定理 6.4.1 和定理 6.4.2 得到，若用变度量法一般函数极小时，若每步迭代一维搜索是精确的，或满足 $(\boldsymbol{s}^{(k)})^T\boldsymbol{y}^{(k)}>0$，则有两种可能：要么停止在某一个稳定点；要么产生一个严格递减的序列 $\{f(\boldsymbol{x}^{(k)})\}$. 若目标函数满足一定条件，可以证明，由变度量法产生的点列 $\{\boldsymbol{x}^{(k)}\}$ 收敛到 $f(\boldsymbol{x})$ 的极小点 $\boldsymbol{x}^*$，其收敛速率是超线性的（证明略）.

6.3.2 搜索方向的共轭性和算法的二次终止性

定理 6.4.3 若用变度量法（BFGS 算法或 DFP 算法）求解正定二次函数

$$\min f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{r}^T\boldsymbol{x} + \delta,$$

若一维搜索是精确的，假设已进行了 m 次迭代，则：

(1) 搜索方向 $\boldsymbol{d}^{(1)}$，$\boldsymbol{d}^{(2)}$，…，$\boldsymbol{d}^{(m)}$是 m 个非零的 G 共轭方向；

(2) 对于 $j=1, 2, \cdots, m$，有

$$B^{(m+1)}\boldsymbol{s}^{(j)} = \boldsymbol{y}^{(j)} \quad (H^{(m+1)}\boldsymbol{y}^{(j)} = \boldsymbol{s}^{(j)}), \qquad (6.4.5)$$

且当 $m=n$ 时，有

$$B^{(n+1)} = G, \ (H^{(n+1)} = G^{-1}).$$

证明 由算法知 $d^{(1)}$, $d^{(2)}$, …, $d^{(m)}$ 是非零的. 先证明共轭性. 用数学归纳法证明

$$\begin{cases}(s^{(i)})^T G s^{(j)}=0, & 1\leqslant j<i\leqslant m, \\ B^{(m+1)} s^{(j)}=y^{(j)} & 1\leqslant j\leqslant m.\end{cases} \tag{6.4.6}$$

当 $m=2$ 时, 有

$$\begin{aligned}(s^{(2)})^T G s^{(1)} &= (s^{(2)})^T y^{(1)} = \alpha_2(d^{(2)})^T y^{(1)} \\ &= -\alpha_2 \nabla f(x^{(2)})^T (B^{(2)})^{-1} y^{(1)} \\ &= -\alpha_2 \nabla f(x^{(2)})^T s^{(1)} = 0,\end{aligned} \tag{6.4.7}$$

(6.4.7) 式中最后一个等号是由精确一维搜索得到的. 我们有

$$B^{(3)} s^{(2)} = y^{(2)} \text{(拟 Newton 方程)},$$

$$B^{(3)} s^{(1)} = B^{(2)} s^{(1)} - \frac{(s^{(2)})^T B^{(2)} s^{(1)}}{(s^{(2)})^T B^{(2)} s^{(2)}} B^{(2)} s^{(2)} + \frac{(y^{(2)})^T s^{(1)}}{(y^{(2)})^T s^{(2)}} y^{(2)}. \tag{6.4.8}$$

由 (6.4.7) 式和拟 Newton 方程 $B^{(2)} s^{(1)}=y^{(1)}$, 可知

$$(s^{(2)})^T B^{(2)} s^{(1)} = (s^{(2)})^T y^{(1)}=0,$$

$$(y^{(2)})^T s^{(1)} = (s^{(2)})^T G s^{(1)}=0,$$

故 (6.4.8) 式为

$$B^{(3)} s^{(1)} = y^{(1)},$$

命题为真.

假设当 $m=k-1$ 时, (6.4.6) 式成立, 即

$$\begin{cases}(s^{(i)})^T G s^{(j)}=0, \ 1\leqslant j<i\leqslant k-1, \\ B^{(k)} s^{(j)}=y^{(j)}, \ 1\leqslant j\leqslant k-1.\end{cases} \tag{6.4.9}$$

当 $m=k$ 时, 只需证明

$$\begin{cases}(s^{(k)})^T G s^{(j)}=0, \ j=1, 2, \cdots k-1, \\ B^{(k+1)} s^{(j)}=y^{(j)}, \ j=1, 2, \cdots, k\end{cases} \tag{6.4.10}$$

即可, 由归纳法假设和扩展子空间定理的推论, 可以得到

$$(s^{(k)})^T G s^{(j)} = (s^{(k)})^T y^{(j)} = \alpha_2 (d^{(k)})^T y^{(j)}$$
$$= -\alpha_k \nabla f(x^{(k)})^T (B^{(k)})^{-1} y^{(j)}$$
$$= -\alpha_k \nabla f(x^{(k)})^T s^{(j)} = 0, j = 1, 2, \cdots, k-1, \tag{6.4.11}$$

和

$$B^{(k+1)} s^{(j)} = B^{(k)} s^{(j)} - \frac{(s^{(k)})^T B^{(k)} s^{(j)}}{(s^{(k)})^T B^{(k)} s^{(k)}} B^{(k)} s^{(k)} + \frac{(y^{(k)})^T s^{(j)}}{(y^{(k)})^T s^{(k)}} y^{(k)}. \tag{6.4.12}$$

由（6.4.11）式和归纳法假设，有

$$(s^{(k)})^T B^{(k)} s^{(j)} = (s^{(k)})^T y^{(j)} = 0,$$
$$(y^{(k)})^T s^{(j)} = (s^{(k)})^T G s^{(j)} = 0,$$

因此（6.4.12）式为

$$B^{(k+1)} s^{(j)} = y^{(j)}, \quad j = 1, 2, \cdots, k-1.$$

注意到当 $j = k$ 时，上式为拟 Newton 方程，因此（6.4.9）式成立．命题为真．

下面证明定理的最后一部分 $B^{(n+1)} = G$.

由于 $s^{(1)}$，$s^{(2)}$，…，$s^{(n)}$ 是非零的 G 共轭方向，因此线性无关．由（6.4.5）式得到

$$B^{(n+1)} [s^{(1)} s^{(2)} \cdots s^{(n)}] = [y^{(1)} y^{(2)} \cdots y^{(n)}] = G [s^{(1)} s^{(2)} \cdots s^{(n)}], \tag{6.4.13}$$

所以

$$B^{(n+1)} = G,$$

另一组公式可以同理证明．

定理 6.4.3 说明，用变度量法求解正定二次函数时，有 $B^{(n+1)} = G$（或 $H^{(n+1)} = G^{-1}$），最终在数值上与 Hesse 矩阵相同．从另一个角度说明，变度量法作 n 步运算相当于一步 Newton 法．

下面证明变度量法具有二次终止性.

定理 6.4.4 若一维搜索是精确的，则变度量法（BFGS 算法或 DFP 算法）具有二次终止性.

证明 考虑用 BFGS 算法或 DFP 算法求解正定二次目标函数，若 $\nabla f(\boldsymbol{x}^{(k)})=0\ (k\leqslant n)$，则 $x^{(k)}$ 为无约束问题的最优解；否则由算法得到的搜索方向 $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(n)}$ 是共轭的，由扩展子空间定理可知，$\boldsymbol{x}^{(n+1)}$ 是无约束问题的最优解.

类似于共轭梯度法的分析，对于变度量法也可以采用 n 步重新开始策略.

变度量法是无约束最优化算法中最为有效的方法之一. 在一定条件下，算法具有超线性收敛性，并且还具有其它的一些性质，这里就不介绍了.

本节介绍了两种最基本的变度量法——BFGS 算法和 DFP 算法. 关于变度量法还有其它若干类算法，如 Broyden 矩阵族、Huang 矩阵族和 Dennis 矩阵族算法，有兴趣的读者可以参见有关参考书.

§6.5 非线性最小二乘问题

在实际问题中，常常会碰到一种特殊类型的函数求极值问题，如目标函数为平方和形式

$$f(x)=\frac{1}{2}\sum_{i=1}^{m} r_i^2(\boldsymbol{x}),\ \boldsymbol{x}\in R^n,\ m\geqslant n, \qquad (6.5.1)$$

通常 $r_i(\boldsymbol{x})(i=1, 2\cdots, m)$ 是非线性函数，因此称问题 (6.5.1) 为非线性最小二乘问题. 在第一章介绍的曲线拟合问题就是典型的非线性最小二乘问题.

虽然，非线性最小二乘问题属于无约束最优化问题，可以用前面几节介绍的算法求解，但由于问题 (6.5.1) 具有特殊的形

式，因此针对这种特殊形式，本节介绍一些更加行之有效的方法.

为了今后讨论方便，引进一些记号. 令

$$\boldsymbol{r}(\boldsymbol{x})=(r_1(\boldsymbol{x}),r_2(\boldsymbol{x}),\cdots,r_m(\boldsymbol{x}))^T, \tag{6.5.2}$$

因此，目标函数（6.5.1）可改写为

$$f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{r}(\boldsymbol{x})^T\boldsymbol{r}(\boldsymbol{x}). \tag{6.5.3}$$

定义 r（x）的 Jacobi 矩阵 J 为

$$J(x)=\begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \cdots & \frac{\partial r_1}{\partial x_n} \\ \frac{\partial r_2}{\partial x_1} & \frac{\partial r_2}{\partial x_2} & \cdots & \frac{\partial r_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_m}{\partial x_1} & \frac{\partial r_m}{\partial x_2} & \cdots & \frac{\partial r_m}{\partial x_n} \end{bmatrix}=\begin{bmatrix} \nabla r_1(\boldsymbol{x})^T \\ \nabla r_2(\boldsymbol{x})^T \\ \vdots \\ \nabla r_m(\boldsymbol{x})^T \end{bmatrix}, \tag{6.5.4}$$

那么 $f(\boldsymbol{x})$ 的梯度可以写成

$$\nabla f(\boldsymbol{x}) = \sum_{i=1}^{m}\nabla r_i(\boldsymbol{x})r_i(\boldsymbol{x}) = J(\boldsymbol{x})^T\boldsymbol{r}(\boldsymbol{x}), \tag{6.5.5}$$

Hesse 矩阵可以写成

$$\begin{aligned}\nabla^2 f(\boldsymbol{x}) &= \sum_{i=1}^{m}\nabla r_i(\boldsymbol{x})\nabla r_i(\boldsymbol{x})^T + \sum_{i=1}^{m} r_i(\boldsymbol{x})\nabla^2 r_i(\boldsymbol{x}) \\ &= J(\boldsymbol{x})^T J(\boldsymbol{x}) + \sum_{i=1}^{m} r_i(\boldsymbol{x})\nabla^2 r_i(\boldsymbol{x}).\end{aligned} \tag{6.5.6}$$

6.5.1 Gauss—Newton 法

在 Newton 法中，需要求解 Newton 方程

$$\nabla^2 f(\boldsymbol{x}^{(k)})\boldsymbol{d} = -\nabla f(\boldsymbol{x}^{(k)}),$$

由（6.5.6）式可知，这里需要计算 $r_i(\boldsymbol{x})$（$i=1, 2, \cdots, m$）的 Hesse 矩阵，计算量大. 为了简化计算，略去（6.5.6）式中的第二项，即求解方程组

$$J(\boldsymbol{x}^{(k)})^T J(\boldsymbol{x}^{(k)})\boldsymbol{d} = -J(\boldsymbol{x}^{(k)})^T \boldsymbol{r}(\boldsymbol{x}^{(k)}), \tag{6.5.7}$$

得 $\boldsymbol{d}^{(k)}$，然后置

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{d}^{(k)}, \tag{6.5.8}$$

这样就得到 Gauss-Newton 法.

算法 6.5.1　（Gauss-Newton 法）

（1）取初始点 $\boldsymbol{x}^{(1)}$，置精度要求 ε，置 $k=1$.

（2）如果 $\| J(\boldsymbol{x}^{(k)})^T \boldsymbol{r}(\boldsymbol{x}^{(k)}) \| \leqslant \varepsilon$，则停止计算（$\boldsymbol{x}^{(k)}$ 作为无约束问题的解）；否则求解线性方程组

$$J(\boldsymbol{x}^{(k)})^T J(\boldsymbol{x}^{(k)})\boldsymbol{d} = -J(\boldsymbol{x}^{(k)})^T r(\boldsymbol{x}^{(k)}),$$

得到 $\boldsymbol{d}^{(k)}$.

（3）置

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{d}^{(k)}, \quad k = k+1,$$

转（2）.

在许多实际问题中，当局部解 $\boldsymbol{x}^*$ 对应的目标函数值 $f(\boldsymbol{x}^*)$ 接近于 0，这时，如果当迭代点 $\boldsymbol{x}^{(k)}$ 接近 $\boldsymbol{x}^*$ 时，$\| r(\boldsymbol{x}^{(k)}) \|$ 较小，或曲线 $r_i(1\boldsymbol{x})$ 接近直线（$\nabla^2 r_i(\boldsymbol{x}) \approx 0$）时，采用 $\nabla^2 f(\boldsymbol{x}^{(k)}) \approx J(\boldsymbol{x}^{(k)})^T J(\boldsymbol{x}^{(k)})$ 的 Gauss-Newton 法可望有较好的效果；但当 $\| r(\boldsymbol{x}^{(k)}) \|$ 较大时，或曲线 $r_i(\boldsymbol{x})$ 的曲率较大时，作 $\nabla^2 f(\boldsymbol{x}^{(k)})$ 近似略去不容忽视的项 $\sum\limits_{i=1}^{m} r_i(\boldsymbol{x})\nabla^2 r_i(\boldsymbol{x})$，因而难于期待 Gauss-Newton 法会有较好的效果.

显然，矩阵 $J(\boldsymbol{x}^{(k)})^T J(\boldsymbol{x}^{(k)})$ 是半正定矩阵. 当 Jacobi 矩阵 $J(\boldsymbol{x}^{(k)})$ 为列满秩时，矩阵 $J(\boldsymbol{x}^{(k)})^T J(\boldsymbol{x}^{(k)})$ 是正定矩阵，因此由方程组(6.5.7)得到的 $\boldsymbol{d}^{(k)}$ 是 $f(\boldsymbol{x})$ 的下降方向. 但仍不能保证有

$f(x^{(k+1)})<f(x^{(k)})$，因此可采用类似于修正 Newton 法的方法，增加一维搜索策略.

6.5.2 Levenberg-Marquardt 方法

在 Gauss-Newton 法中，当 $J(x^{(k)})$的各列是线性相关的，或接近线性相关，换句话说，矩阵 $J(x^{(k)})^TJ(x^{(k)})$ 奇异或是病态的. 那么求解线性方程组（6.5.7）会出现一些困难，克服该困难的方法之一是将线性方程组（6.5.7）改为

$$(J(x^{(k)})^TJ(x^{(k)})+\nu I)d=-J(x^{(k)})^Tr(x^{(k)}), \tag{6.5.9}$$

其中 $\nu>0$ 是迭代过程中需要调整的参数.

Levenberg-Marquardt 方法的关键是如何调整参数 ν. 为此先介绍两个定理.

定理 6.5.1 若 $\boldsymbol{d}(\nu)$是方程(6.5.9)的解，则 $\|\boldsymbol{d}(\nu)\|^2$ 是 ν 的连续下降函数，且当 $\nu\to+\infty$时，有 $\|\boldsymbol{d}(\nu)\|\to 0$.

证明 因为 $J(x^{(k)})^TJ(x^{(k)})$ 是对称半正定矩阵，因此存在正交阵 $P^{(k)}$，使得

$$\begin{aligned}&(P^{(k)})^T(J(x^{(k)})^TJ(x^{(k)}))P^{(k)}\\&=\Lambda^{(k)}=\operatorname{diag}(\lambda_1^{(k)},\lambda_2^{(k)},\cdots,\lambda_n^{(k)}).\end{aligned} \tag{6.5.10}$$

将（6.5.9）式左乘 $P^{(k)}$，得到

$$\begin{aligned}&(P^{(k)})^T(J(x^{(k)})^TJ(x^{(k)})+\nu I)P^{(k)}(P^{(k)})^{-1}d(\nu)\\&=-(P^{(k)})^TJ(x^{(k)})^Tr(x^{(k)}),\end{aligned} \tag{6.5.11}$$

由(6.5.10)式，(6.5.11)式化简为

$$(\Lambda^{(k)}+\nu I)(P^{(k)})^{-1}\boldsymbol{d}(\nu)=-(P^{(k)})^TJ(x^{(k)})^Tr(x^{(k)}). \tag{6.5.12}$$

因为 $\Lambda^{(k)}$的对角元素 $\lambda_1^{(k)}$，$\lambda_2^{(k)}$，…，$\lambda_n^{(k)}$ 是 $J(x^{(k)})^TJ(x^{(k)})$的特征值，有 $\lambda_i^{(k)}\geqslant 0$，$i=1,2,\cdots,n$. 所以对一切 $\nu>0$，有 $\lambda_i^{(k)}+\nu>0$，$i=1,2,\cdots,n$，因此矩阵

$(\Lambda^{(k)}+\nu I)^{-1}$存在，方程（6.5.12）的解为

$$d(\nu) = -P^{(k)}(\Lambda^{(k)}+\nu I)^{-1}(P^{(k)})^T J(x^{(k)})^T r(x^{(k)}),\tag{6.5.13}$$

所以

$$\begin{aligned}\|d(\nu)\|^2 &= d(\nu)^T d(\nu)\\ &= r(x^{(k)})^T J(x^{(k)})P^{(k)}(\Lambda^{(k)}+\nu I)^{-1}(P^{(k)})^T \cdot\\ &\quad P^{(k)}(\Lambda^{(k)}+\nu I)^{-1}(P^{(k)})^T J(x^{(k)})^T r(x^{(k)})\\ &= r(x^{(k)})^T J(x^{(k)})P^{(k)}((\Lambda^{(k)}+\nu I)^2)^{-1}(P^{(k)})^T J(x^{(k)})^T r(x^{(k)}).\end{aligned}\tag{6.5.14}$$

令

$$u^{(k)} = (P^{(k)})^T J(x^{(k)})^T r(x^{(k)}),\tag{6.5.15}$$

于是（6.5.14）式改写为

$$\begin{aligned}\|d(\nu)\|^2 &= (u^{(k)})^T((\Lambda^{(k)}+\nu I)^2)^{-1}u^{(k)}\\ &= (u_1^{(k)},u_2^{(k)},\cdots,u_n^{(k)})\begin{pmatrix}\frac{1}{(\lambda_1^{(k)}+\nu)^2} & & & \\ & \frac{1}{(\lambda_2^{(k)}+\nu)^2} & & \\ & & \ddots & \\ & & & \frac{1}{(\lambda_n^{(k)}+\nu)^2}\end{pmatrix}\begin{pmatrix}u_1^{(k)}\\ u_2^{(k)}\\ \vdots\\ u_n^{(k)}\end{pmatrix}\\ &= \sum_{i=1}^{n}\frac{(u_i^{(k)})^2}{(\lambda_i^{(k)}+\nu)^2},\end{aligned}\tag{6.5.16}$$

由于 $u_i^{(k)}$，$\lambda_i^{(k)}$（$i=1,2,\cdots,n$）与参数 ν 无关，因此 $\|d(\nu)\|^2$ 是 ν 的连续下降函数，且当 $\nu\to+\infty$ 时，有 $\|d(\nu)\|\to 0$.

定理6.5.1表明，由 Levenberg-Marquardt 方法得到的$d(\nu)$满足$\|d(\nu)\|\leqslant\|d(0)\|$（$\nu>0$），并且 ν 越大，其模就越小. 从几何直观上来看，当$J(x)^{(k)T}J(x^{(k)})$ 接近奇异时，由线性方程组（6.5.7）得到的 $d^{(k)}$的模就相当大. 而采用 Levenberg-Marquardt 方法后，当增大 ν，$\|d^{(k)}\|$ 减少，这样就限制了

$\|d^{(k)}\|$，使得 $\|d^{(k)}\|$ 不会过大．因此，Levenberg-Marquardt 方法也称为阻尼最小二乘法．

定理 6.5.2 若 $d(\nu)$ 是方程(6.5.9)的解，则 $d(\nu)$ 是 $f(x)$ 在 $x^{(k)}$ 处的下降方向，且当 $\nu\to+\infty$ 时，$d(\nu)$ 的方向与 $-J(x^{(k)})^Tr(x^{(k)})$ 方向一致．

证明 因为 $\nu>0$，矩阵 $J(x^{(k)})^TJ(x^{(k)})+\nu I$ 正定，所以有

$$\begin{aligned}&\nabla f(x^{(k)})^Td(\nu)\\&=-r(x^{(k)})^TJ(x^{(k)})^T(J(x^{(k)})^TJ(x^{(k)})+\nu I)^{-1}J(x^{(k)})^Tr(x^{(k)})\quad<0,\end{aligned}$$

所以 $d(\nu)$ 是下降方向．

现在证明第二个结论．记

$$\Psi(\nu)=-\frac{d(\nu)^TJ(x^{(k)})^Tr(x^{(k)})}{\|d(\nu)\|\cdot\|J(x^{(k)})^Tr(x^{(k)})\|},\tag{6.5.17}$$

表示 $d(\nu)$ 与 $-J(x)^{(k)T}r(x)^{(k)}$ 夹角的方向余弦．由定理 6.5.1 的推导过程，(6.5.13) 式、(6.5.15) 式和 (6.5.16) 式得到

$$\Psi(\nu)=\left(\sum_{i=1}^{n}\frac{(u_i^{(k)})^2}{\lambda_i^{(k)}+\nu}\right)\left(\sum_{i=1}^{n}\frac{(u_i^{(k)})^2}{(\lambda_i^{(k)}+\nu)^2}\right)^{-\frac{1}{2}}\left(\sum_{i=1}^{n}(u_i^{(k)})^2\right)^{-\frac{1}{2}},\tag{6.5.18}$$

于是有

$$\lim_{\nu\to+\infty}\Psi(\nu)=1,$$

因此当 ν 充分大时，$d(\nu)$ 的方向与 $-J(x^{(k)})^Tr(x^{(k)})$ 的方向相一致．

定理 6.5.2 表明，当 ν 充分大时，Levenberg-Marquardt 方法产生的搜索方向 $d^{(k)}$ 与负梯度方向相一致．

下面给出参数 ν 的调整方法．首先定义一个二次函数

$$q(d)=f(x^{(k)})+(J(x^{(k)})^Tr(x^{(k)}))^Td+\frac{1}{2}d^T(J(x^{(k)})^TJ(x^{(k)}))d,\tag{6.5.19}$$

其增量为

$$
\begin{aligned}
\Delta q^{(k)} &= q(\boldsymbol{d}^{(k)}) - q(0) \\
&= (J(\boldsymbol{x}^{(k)})^T r(\boldsymbol{x}^{(k)}))^T \boldsymbol{d}^{(k)} + \frac{1}{2}(\boldsymbol{d}^{(k)})^T (J(\boldsymbol{x}^{(k)})^T J(\boldsymbol{x}^{(k)}))d^{(k)}.
\end{aligned} \tag{6.5.20}
$$

再考虑目标函数的增量

$$
\Delta f^{(k)} = f\ (\boldsymbol{x}^{(k)} + \boldsymbol{d}^{(k)})\ - f\ (\boldsymbol{x}^{(k)})\ = f\ (\boldsymbol{x}^{(k+1)})\ - f\ (\boldsymbol{x}^{(k)}), \tag{6.5.21}
$$

定义 γ_k 为两增量之比，即

$$
\gamma_k = \frac{\Delta r^{(k)}}{\Delta q^{(k)}} = \frac{f(\boldsymbol{x}^{(k+1)}) - f(\boldsymbol{x}^{(k)})}{(J(\boldsymbol{x}^{(k)})^T r(\boldsymbol{x}^{(k)}))^T \boldsymbol{d}^{(k)} + \frac{1}{2}(\boldsymbol{d}^{(k)})^T (J(\boldsymbol{x}^{(k)})^T J(\boldsymbol{x}^{(k)}))\boldsymbol{d}^{(k)}}. \tag{6.5.22}
$$

当 γ_k 接近于 1 时，表示 $\Delta f^{(k)}$ 接近 $\Delta q^{(k)}$，即 $f(\boldsymbol{x}^{(k)} + \boldsymbol{d}^{(k)})$接近于 $q(\boldsymbol{d}^{(k)})$,也就是说,$f(\boldsymbol{x})$在 $\boldsymbol{x}^{(k)}$附近接近一个二次函数，那么用 Gauss-Newton 法求解最小二乘问题效果较好．换句话说，用 Levenberg-Marquardt 方法求解时参数 ν 应取得小一些．反过来，当 γ_k 接近于 0 时，表明 $\Delta f^{(k)}$与 $\Delta q^{(k)}$近似的程度不好，或者说 $\boldsymbol{d}^{(k)}$不应取得过大，应减少 $\boldsymbol{d}^{(k)}$的模长．定理 6.5.1 表明，此时应增大参数 ν，来限制 $\boldsymbol{d}^{(k)}$的模．当 γ_k 不接近于 0，也不接近于 1 时，认为参数 ν 选取适当，此时 ν 不作调整．

通常 γ_k 的临界值为 0.25 和 0.75，由此得到如下算法．

算法 6.5.2 （Levenberg-Marquardt 方法）

(1) 取初始点 $\boldsymbol{x}^{(1)}$和 ν_1 ($\neq 0$)，置精度要求 ϵ，置 $k=1$．

(2) 如果 $\| J(\boldsymbol{x}^{(k)})^T \boldsymbol{r}(\boldsymbol{x}^{(k)}) \| \leqslant \epsilon$,则停止计算($\boldsymbol{x}^{(k)}$作为无约束问题的解)；否则求解线性方程组

$$
(J(\boldsymbol{x}^{(k)})^T J(\boldsymbol{x}^{(k)}) + \nu_k I)\boldsymbol{d} = -J(\boldsymbol{x}^{(k)})^T \boldsymbol{r}(\boldsymbol{x}^{(k)}),
$$

得到 $\boldsymbol{d}^{(k)}$．

(3) 置

$$x^{(k+1)}=x^{(k)}+d^{(k)},$$

并计算

$$\gamma_k=\frac{f(x^{(k+1)})-f(x^{(k)})}{(J(x^{(k)})^T r(x^{(k)}))^T d^{(k)}+\frac{1}{2}(d^{(k)})^T(J(x^{(k)})^T J(x^{(k)}))d^{(k)}}.$$

(4) 若 $\gamma_k<0.25$，则置 $\nu_{k+1}=4\nu_k$；否则如果 $\gamma_k>0.75$，则置 $\nu_{k+1}=\nu_k/2$；否则置 $\nu_{k+1}=\nu_k$.

(5) 置 $k=k+1$，转 (2).

这里就不详细讨论初始参数 ν_1 的选取，通常选择一个较小的正数就可以了.

习 题 六

6.1 用 Newton 法求解无约束问题

$$\min f(x)=2x_1^2-2x_1x_2+x_2^2+2x_1-2x_2,$$

取初始点 $x^{(1)}=(0,0)^T$.

6.2 用 Newton 法求解无约束问题

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

取初始点 $x^{(1)}=(0,3)^T$，精度要求 $\varepsilon=10^{-3}$.

6.3 证明：修正 Newton 法具有二次终止性.

6.4 设 G 为正定对称矩阵，求证：若 $s^{(1)}, s^{(2)}, \cdots, s^{(k)}$ 为 G 的 k 个非零的共轭方向，则 $s^{(1)}, s^{(2)}, \cdots, s^{(k)}$ 线性无关.

6.5 若 G 为三对角阵

$$G=\begin{pmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix},$$

并定义

$$s^{(i)}=(1,2,\cdots,i,0,0,\cdots,0)^T,\quad i=1,2,\cdots,n,$$

证明：$s^{(1)}$，$s^{(2)}$，…，$s^{(n)}$是 G 的 n 个共轭方向.

6.6 设 G 是 n 阶正定对称矩阵，$p^{(1)}$，$p^{(2)}$，…，$p^{(n)}$是 R^n 中的 n 个线性无关向量，试证：如下定义的向量 $d^{(1)}$，$d^{(2)}$，…，$d^{(n)}$，

$$d^{(1)}=p^{(1)},$$

$$d^{(k+1)}=p^{(k+1)}-\sum_{i=1}^{k}\frac{(p^{(k+1)})^TGd^{(i)}}{(d^{(i)})^TGd^{(i)}}d^{(i)},$$

$$k=1,2,\cdots,n-1$$

是 n 个非零 G 共轭向量.

6.7 设 $f(x)$是正定二次函数

$$f(x)=\frac{1}{2}x^TGx+r^Tx+\delta,$$

试证：若 $\bar{x}$ 与 $\hat{x}$ 分别是 $f(x)$在两条平行方向 d 的直线上的极小点，则方向 $u=\bar{x}-\hat{x}$ 与方向 d 关于 G 共轭.

6.8 用共轭梯度法(FR 算法或 PRP 算法)求解无约束问题

$$\min f(x)=2x_1^2-2x_1x_2+x_2^2+2x_1-2x_2,$$

取 $x^{(1)}=(0,0)^T$.

6.9 设 $f(x)$ 是正定二次函数

$$f(x)=\frac{1}{2}x^TGx+r^Tx+\delta,$$

若 $s^{(1)},s^{(2)},\cdots,s^{(n)}$是 G 的 n 个非零共轭方向，证明 $f(x)$的极小点 x^* 可以表示成

$$x^*=-\sum_{k=1}^{n}\frac{r^Ts^{(k)}}{(s^{(k)})^TGs^{(k)}}s^{(k)}.$$

6.10 分别用 BFGS 算法和 DFP 算法求解无约束问题

$$\min f(x)=2x_1^2-2x_1x_2+x_2^2+2x_1-2x_2,$$

取 $x^{(1)}=(0,0)^T$.

6.11 导出对 $H^{(k)}$ 的对称秩 1 修正公式，得到相应的变度量法，并求解无约束问题

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

取 $x^{(1)}=(0,0)^T$，$H^{(1)}=I$.

6.12 证明广义 Cauchy-Schwarz 不等式：设 A 为正定对称矩阵，则

$$(x^TAx)(y^TAy)\geqslant(x^TAy)^2,$$

且等号成立的充分必要条件是：x 与 y 共线.

6.13 考虑用 DFP 算法求解正定二次目标函数

$$f(x)=\frac{1}{2}x^TGx+r^Tx+\delta,$$

若一维搜索是精确的，证明：

$$G^{-1}=\sum_{k=1}^{n}\frac{s^{(k)}(s^{(k)})^T}{(y^{(k)})^Ts^{(k)}},$$

其中 $s^{(k)}=x^{(k+1)}-x^{(k)}$，$y^{(k)}=\nabla f(x^{(k+1)})-\nabla f(x^{(k)})=Gs^{(k)}$.

6.14 考虑用对 $B^{(k)}$ 的对称秩 1 修正公式得到的变度量法求解正定二次目标函数

$$f(x)=\frac{1}{2}x^TGx+r^Tx+\delta,$$

若得到的 $s^{(1)}$，$s^{(2)}$，…，$s^{(n)}$ 线性无关，则：

(1) 对于 $k=1,2,\cdots,n$，有

$$B^{(k+1)}s^{(j)}=y^{(j)},\quad j=1,2,\cdots,k.$$

(2) $B^{(n+1)}=G$.

6.15 设 $B^{(k)}$ 是由对 $B^{(k)}$ 的对称秩 1 修正公式(6.3.13)得到的，$H^{(k)}$ 是对 $H^{(k)}$ 的对称秩 1 修正公式(6.3.22)得到的. 若 $H^{(k)}=B^{(k)-1}$，用 Sherman-Morrison 公式证明：

$$H^{(k+1)}=B^{(k+1)-1}.$$

6.16　用 Gauss-Newton 法求解最小二乘问题

$$\min f(\boldsymbol{x})=\frac{1}{2}\left[(x_2-x_1^2)^2+(1-x_1)^2\right],$$

取 $\boldsymbol{x}^{(1)}=(0,\ 0)^T$，迭代二步（计算到 $\boldsymbol{x}^{(3)}$）.

6.17　用 Levenberg-Marquardt 方法求解最小二乘问题

$$\min f(\boldsymbol{x})=\frac{1}{2}\left[(x_1-x_1^2)^2+(1-x_1)^2\right],$$

取 $\boldsymbol{x}^{(1)}=(0,\ 0)^T$，$\nu_1=0.1$，迭代二步（计算到 $\boldsymbol{x}^{(3)}$）.

第七章　直 接 方 法

本章介绍求解无约束最优化问题

$$\min f(x), x \in R^n$$

的直接方法. 该类方法仅用到目标函数的函数值，而不必计算导数值.

§7.1　Powell 方法

7.1.1　坐标轮换法

最简单的直接方法是**坐标轮换法**（Cyclic Coordinate Method），也称为**交替方向法**（Alternating Directions Method），它的基本原理是任取初始点 $x^{(1)}$，沿各个坐标轴方向进行一维搜索，即

$$x^{(1)} \Rightarrow x_{11} \xrightarrow{e_1} x_{12} \xrightarrow{e_2} x_{13} \longrightarrow \cdots \longrightarrow x_{1n} \xrightarrow{e_n} x_{1n+1},$$

$$x_{1n+1} \Rightarrow x_{21} \xrightarrow{e_1} x_{22} \xrightarrow{e_2} x_{23} \longrightarrow \cdots \longrightarrow x_{2n} \xrightarrow{e_n} x_{2n+1},$$

$$x_{2n+1} \Rightarrow x_{31} \xrightarrow{e_1} x_{32} \xrightarrow{e_2} x_{33} \longrightarrow \cdots$$

其中 e_i 是各坐标轴方向. 当 x_{in+1}满足某一终止准则时，停止计算.

坐标轮换法简单，但其效率太低，应提出某些改进方案，在作完每一轮搜索后，增加了一个方向 $u^{(i)} = x_{in+1} - x_{i1}$，再沿 $u^{(i)}$作一次一维搜索，即

$$x^{(1)} \Rightarrow x_{11} \xrightarrow{e_1} x_{12} \xrightarrow{e_2} x_{13} \longrightarrow \cdots \longrightarrow$$

$$x_{1n} \xrightarrow{e_n} x_{1n+1} \xrightarrow{u^{(1)} = x_{1n+1} - x_{11}} x_{1n+2},$$

$$x_{1n+2} \Rightarrow x_{21} \xrightarrow{e_1} x_{22} \xrightarrow{e_2} x_{23} \longrightarrow \cdots \longrightarrow$$

$$x_{2n} \xrightarrow{e_n} x_{2n+1} \xrightarrow{u^{(2)} = x_{2n+1} - x_{21}} x_{2n+2},$$

$$x_{2n+2} \Rightarrow x_{31} \xrightarrow{e_1} x_{32} \xrightarrow{e_2} x_{33} \longrightarrow \cdots$$

从直观上来看（见图 7.1.1），增加搜索方向 $u^{(1)}$，将提高算法效率，但为什么在第二轮的循环中，不继续使用 $u^{(1)}$ 呢？如果使用 $u^{(1)}$ 效果不好，那么为什么在第一轮中增加 $u^{(1)}$ 作为搜索方向呢？因此，取定坐标轴方向的方法是不可取的，应从 $n+1$ 个方向（前面的 n 个方面和一个新增加的方向）中，选取 n 个"好的"方向，作为下一轮的搜索方向，因此得到如下算法.

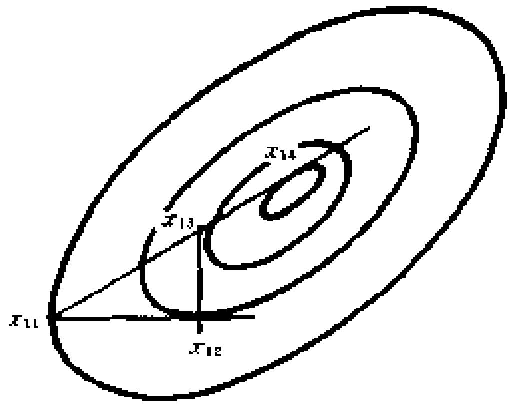


图 7.1.1

算法 7.1.1

(1) 取初始点 $x^{(1)}$，置 $d^{(i)} = e_i$，$i = 1, 2, \cdots, n$（其中 e_i 是 n 个坐标轴方向）.

(2) 从 $x^{(1)}$ 出发，依次沿 $d^{(i)}$（$i = 1, 2, \cdots, n$）和 u 作一维搜索，即

$$x^{(1)} \xrightarrow{d^{(1)}} x^{(2)} \xrightarrow{d^{(2)}} x^{(3)} \longrightarrow \cdots \longrightarrow$$

$$x^{(n)} \xrightarrow{d^{(n)}} x^{(n+1)} \xrightarrow{u = x^{(n+1)} - x^{(1)}} x^{(n+2)}.$$

(3) 若 $x^{(n+2)}$ 满足某一终止准则，则停止计算；否则从 $\{d^{(1)}, d^{(2)}, \cdots, d^{(n)}, u\}$ 选择好的 n 个方向作为新的搜索方向 $d^{(1)}, d^{(2)}, \cdots, d^{(n)}$.

(4) 置 $x^{(1)} = x^{(n+2)}$，转 (2).

只有给出好的 n 个搜索方向的判别准则，算法 7.1.1 才能实现. 下面讨论好的搜索方向的判别准则.

7.1.2 正交程度和共轭程度

前面提到搜索方向好坏的标准问题，对于一般函数而言，很难给出一个好坏的判别标准. 但对于正定二次函数，由扩展子空间定理可知，共轭方向是好的方向. 那么，越接近共轭方向越是好方向. 为了讲清这个问题，先来讨论正交程度的概念.

对于特殊正定二次函数

$$f(\boldsymbol{w}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \boldsymbol{r}^T\boldsymbol{w} + \delta, \tag{7.1.1}$$

正交方向是好方向. 当 n 个方向不是正交方向，那么又如何定义它的好坏呢? 先看一下二维情况.

设 $\boldsymbol{q}^{(1)}$、$\boldsymbol{q}^{(2)}$为单位向量，当 $\boldsymbol{q}^{(1)}$、$\boldsymbol{q}^{(2)}$正交时，其构成的正方形面积为 1. 当 $\boldsymbol{q}^{(1)}$、$\boldsymbol{q}^{(2)}$不正交时，其构成的平行四边形面积越接近于 1，则 $\boldsymbol{q}^{(1)}$、$\boldsymbol{q}^{(2)}$越接近于正交. 反之就越接近于平行. 因此自然会想到用平行四边形面积来定义二个向量的正交程度，即

$$\delta = \| \boldsymbol{q}^{(1)} \times \boldsymbol{q}^{(2)} \|.$$

设 $\boldsymbol{q}^{(1)} = (q_{11}, q_{21})^T$，$\boldsymbol{q}^{(2)} = (q_{12}, q_{22})^T$，则有

$$\delta = | q_{11}q_{22} - q_{12}q_{21} | = | \det(\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}) |, \tag{7.1.2}$$

即用 $\boldsymbol{q}^{(1)}$、$\boldsymbol{q}^{(2)}$构成的行列式的绝对值来表示它们的正交程度.

再讨论三维情况，设 $\boldsymbol{q}^{(1)}$、$\boldsymbol{q}^{(2)}$、$\boldsymbol{q}^{(3)}$为单位向量，当 $\boldsymbol{q}^{(1)}$、$\boldsymbol{q}^{(2)}$、$\boldsymbol{q}^{(3)}$两两正交时，其构成的六面体的体积为 1；否则小于 1. 因此用 $\boldsymbol{q}^{(1)}$、$\boldsymbol{q}^{(2)}$、$\boldsymbol{q}^{(3)}$构成的六面体的体积，即混合积的绝对值

$$\delta = | (\boldsymbol{q}^{(1)} \times \boldsymbol{q}^{(2)}) \cdot \boldsymbol{q}^{(3)} | = | \det(\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{q}^{(3)}) | \tag{7.1.3}$$

来定义它们的正交程度.

可以想象，对于 n 维情况，可以用 n 个单位向量的行列式的绝对值定义正交程度.

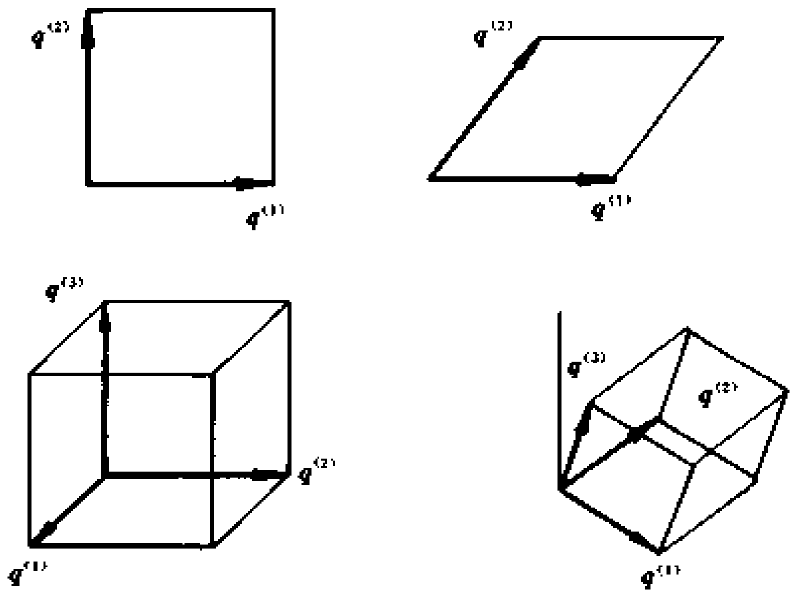


图 7.1.2　二维和三维正交程度的几何意义

定义 7.1.1　设 $\boldsymbol{q}^{(1)}$，$\boldsymbol{q}^{(2)}$，…，$\boldsymbol{q}^{(n)}$是 R^n 中的 n 个向量，其正交程度定义为

$$\delta = \delta(\boldsymbol{q}^{(1)},\boldsymbol{q}^{(2)},\cdots,q^{(n)}) = \begin{cases} 0, \exists\, \boldsymbol{q}^{(i)} = 0, \\ \dfrac{|\det(\boldsymbol{q}^{(1)},\boldsymbol{q}^{(2)},\cdots,\boldsymbol{q}^{(n)})|}{\prod\limits_{i=1}^{n} \|\boldsymbol{q}^{(i)}\|}, \text{其它}. \end{cases} \tag{7.1.4}$$

定理 7.1.1　由定义 7.1.1 定义的正交程度满足

$$\delta(\boldsymbol{q}^{(1)},\boldsymbol{q}^{(2)},\cdots,\boldsymbol{q}^{(n)}) \leqslant 1, \tag{7.1.5}$$

且等号成立的充分必要条件是：$\boldsymbol{q}^{(1)}$，$\boldsymbol{q}^{(2)}$，…，$\boldsymbol{q}^{(n)}$是 n 个非零的正交方向.

证明　对于某个 $\boldsymbol{q}^{(i)}=0$，命题显然成立. 考虑 $\boldsymbol{q}^{(i)}\neq 0$，$i=1, 2,\cdots, n$. 不妨设 $\|\boldsymbol{q}^{(i)}\|=1$，$i=1, 2, \cdots, n$. 所以

$$\delta = |\det(\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \cdots, \boldsymbol{q}^{(n)})|.$$

令

$$Q = (\boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \cdots, \boldsymbol{q}^{(n)}),$$

则

$$Q^TQ = \begin{pmatrix} (\boldsymbol{q}^{(1)})^T\boldsymbol{q}^{(1)} & (\boldsymbol{q}^{(1)})^T\boldsymbol{q}^{(2)} & \cdots & (\boldsymbol{q}^{(1)})^T\boldsymbol{q}^{(n)} \\ (\boldsymbol{q}^{(2)})^T\boldsymbol{q}^{(1)} & (\boldsymbol{q}^{(2)})^T\boldsymbol{q}^{(2)} & \cdots & (\boldsymbol{q}^{(2)})^T\boldsymbol{q}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ (\boldsymbol{q}^{(n)})^T\boldsymbol{q}^{(1)} & (\boldsymbol{q}^{(n)})^T\boldsymbol{q}^{(2)} & \cdots & (\boldsymbol{q}^{(n)})^T\boldsymbol{q}^{(n)} \end{pmatrix}, \tag{7.1.6}$$

由于 Q 非奇异，所以 Q^TQ 是正定对称矩阵. 由线性代数知识（见习题 7.1）有

$$|\det(Q)|^2 = \det(Q^TQ) \leqslant \prod_{i=1}^{n} (\boldsymbol{q}^{(i)})^T\boldsymbol{q}^{(i)} = 1,$$

且等号成立的充分必要条件是：Q^TQ 为对角阵. 即 $\boldsymbol{q}^{(1)}$，$\boldsymbol{q}^{(2)}$，$\cdots$，$\boldsymbol{q}^{(n)}$正交. 命题成立.

定理 7.1.1 表明定义 7.1.1 是合理的. 因此可以用定义 7.1.1 来度量 n 个方向的正交程度.

下面来讨论共轭程度. 对于一般正定二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^TG\boldsymbol{x} + \boldsymbol{r}^T\boldsymbol{x} + \delta,$$

由上一章的知识得到，令

$$\boldsymbol{q}^{(i)} = \sqrt{G}\boldsymbol{d}^{(i)},$$

则正交方向对应于 G 共轭方向，因此可以类似地定义 n 个方向的共轭程度.

定义 7.1.2 设 G 为 n 阶正定对称矩阵，$\boldsymbol{d}^{(1)}$、$\boldsymbol{d}^{(2)}$，$\cdots$，$\boldsymbol{d}^{(n)}$是 R^n 中的 n 个向量，其 G 共轭程度定义为

$$\Delta = \Delta(\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(n)}) =$$

$$\begin{cases} 0, & \exists\, \boldsymbol{d}^{(i)} = 0, \\ \dfrac{|\det(\sqrt{G})| \cdot |\det(\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(n)})|}{\prod\limits_{i=1}^{n} \sqrt{(\boldsymbol{d}^{(i)})^{T} G \boldsymbol{d}^{(i)}}}, & \text{其它}. \end{cases} \tag{7.1.7}$$

也有类似的定理.

定理 7.1.2　由定义 7.1.2 定义的 G 共轭程度满足

$$\Delta(\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \cdots, \boldsymbol{d}^{(n)}) \leqslant 1, \tag{7.1.8}$$

且等号成立的充分必要条件是：$\boldsymbol{d}^{(1)}$，$\boldsymbol{d}^{(2)}$，…，$\boldsymbol{d}^{(n)}$是 n 个非零的 G 共轭方向.

7.1.3　Powell 直接方法

在算法 7.1.1 中，曾提出了一种方案，就是在 $n+1$ 个方向中，找出 n 个好方向. 在前一小节中，给出了好方向的标准，即其共轭程度越接近于 1 越好. 下面将这两种方法结合起来，得到 Powell直接方法. 以二维情况为例，导出 Powell 直接方法.

设最初的两个方向 $\boldsymbol{d}^{(1)}$、$\boldsymbol{d}^{(2)}$，从 $\boldsymbol{x}^{(1)}$出发，沿方向 $\boldsymbol{d}^{(1)}$进行一维搜索得到 $\boldsymbol{x}^{(2)}$，由 $\boldsymbol{x}^{(2)}$出发沿方向 $\boldsymbol{d}^{(2)}$进行一维搜索得到 $\boldsymbol{x}^{(3)}$. 因此确定了一个新的方向 $\boldsymbol{u} = \boldsymbol{x}^{(3)} - \boldsymbol{x}^{(1)}$，再从 $\boldsymbol{x}^{(1)}$出发沿方向 $\boldsymbol{u}$ 进行一维搜索得到 $\boldsymbol{x}^{(4)}$. 见图 7.1.3.

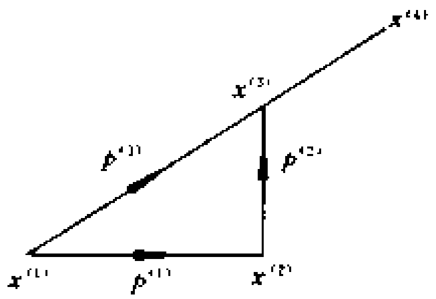


图 7.1.3

下面讨论的问题是：如何从 $\boldsymbol{d}^{(1)}$、$\boldsymbol{d}^{(2)}$和 $\boldsymbol{u}$ 中，确定两个好方向？也就是说，找两个共轭程度最大的方向.

为了方便起见，分别取 $\boldsymbol{d}^{(1)}$、$\boldsymbol{d}^{(2)}$、$\boldsymbol{u}$ 上的方向 $\boldsymbol{p}^{(1)}$、$\boldsymbol{p}^{(2)}$、$\boldsymbol{p}^{(3)}$，满足

$$(\boldsymbol{p}^{(i)})^T G \boldsymbol{p}^{(i)} = 1,\ i = 1,2,3, \tag{7.1.9}$$

现计算 $\boldsymbol{p}^{(1)}$、$\boldsymbol{p}^{(2)}$、$\boldsymbol{p}^{(3)}$ 中任意两个方向的共轭程度. 由定义 7.1.2 得到

$$\Delta = \Delta(\boldsymbol{p}^{(1)}, \boldsymbol{p}^{(2)}) = |\det(\sqrt{G})| \cdot |\det(\boldsymbol{p}^{(1)}, \boldsymbol{p}^{(2)})|, \tag{7.1.10}$$

$$\Delta_1 = \Delta(\boldsymbol{p}^{(2)}, \boldsymbol{p}^{(3)}) = |\det(\sqrt{G})| \cdot |\det(\boldsymbol{p}^{(2)}, \boldsymbol{p}^{(3)})|, \tag{7.1.11}$$

$$\Delta_2 = \Delta(\boldsymbol{p}^{(3)}, \boldsymbol{p}^{(1)}) = |\det(\sqrt{G})| \cdot |\det(\boldsymbol{p}^{(3)}, \boldsymbol{p}^{(1)})|, \tag{7.1.12}$$

由于 $\boldsymbol{x}^{(3)}$是从 $\boldsymbol{x}^{(1)}$出发，沿方向 $\boldsymbol{p}^{(1)}$、$\boldsymbol{p}^{(2)}$进行一维搜索得到的，因此 $\boldsymbol{x}^{(3)} - \boldsymbol{x}^{(1)}$可以表示成 $\boldsymbol{p}^{(1)}$、$\boldsymbol{p}^{(2)}$的线性组合，即

$$\boldsymbol{x}^{(3)} - \boldsymbol{x}^{(1)} = \boldsymbol{x}^{(3)} - \boldsymbol{x}^{(2)} + \boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)} = \alpha_2 \boldsymbol{p}^{(2)} + \alpha_1 \boldsymbol{p}^{(1)}. \tag{7.1.13}$$

另一方面，$\boldsymbol{x}^{(3)} - \boldsymbol{x}^{(1)}$可以写成

$$\boldsymbol{x}^{(3)} - \boldsymbol{x}^{(1)} = \alpha_3 \boldsymbol{p}^{(3)}, \tag{7.1.14}$$

因此得到

$$\boldsymbol{p}^{(3)} = \frac{\alpha_1}{\alpha_3}\boldsymbol{p}^{(1)} + \frac{\alpha_2}{\alpha_3}\boldsymbol{p}^{(2)}, \tag{7.1.15}$$

将（7.1.15）式代入（7.1.11）式和（7.1.12）式分别得到：

$$\begin{aligned}\Delta_1 &= |\det(\sqrt{G})| \cdot \left|\det\left(\boldsymbol{p}^{(2)}, \frac{\alpha_1}{\alpha_3}\boldsymbol{p}^{(1)} + \frac{\alpha_2}{\alpha_3}\boldsymbol{p}^{(2)}\right)\right| \\ &= \left|\frac{\alpha_1}{\alpha_3}\right| |\det(\sqrt{G})| \cdot |\det(\boldsymbol{p}^{(1)}, \boldsymbol{p}^{(2)})| \\ &= \left|\frac{\alpha_1}{\alpha_3}\right| \Delta,\end{aligned} \tag{7.1.16}$$

$$\Delta_2 = |\det(\sqrt{G})| \cdot \left|\det\left(\frac{\alpha_1}{\alpha_3}\boldsymbol{p}^{(1)} + \frac{\alpha_2}{\alpha_3}\boldsymbol{p}^{(2)}, \boldsymbol{p}^{(1)}\right)\right|$$

$$= \left| \frac{\alpha_2}{\alpha_3} \right| \Delta, \tag{7.1.17}$$

因此，比较其共轭程度的大小就可转化为比较系数绝对值$|\alpha_1|$、$|\alpha_2|$、$|\alpha_3|$的大小.

首先要解决的问题是计算出 α_1、α_2 和 α_3 的值.

由于 $\boldsymbol{x}^{(2)}$是从 $\boldsymbol{x}^{(1)}$出发，经精确一维搜索得到的，因此

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \alpha_1 \boldsymbol{p}^{(1)},$$

因为 α_1 为一维函数$\phi(\alpha) = f(\boldsymbol{x}^{(1)} + \alpha \boldsymbol{p}^{(1)})$的极小点，所以有

$$0 = \phi'(\alpha) = \nabla f(\boldsymbol{x}^{(2)})^T \boldsymbol{p}^{(1)}, \tag{7.1.18}$$

因此，由$f(\boldsymbol{x})$在 $\boldsymbol{x}^{(2)}$处的 Taylor 展开式，并注意到（7.1.9）式，有

$$\begin{aligned} f(\boldsymbol{x}^{(1)}) &= f(\boldsymbol{x}^{(2)}) + \nabla f(\boldsymbol{x}^{(2)})^T(\boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)}) + \frac{1}{2}(\boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)})^T G(\boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)}) \\ &= f(\boldsymbol{x}^{(2)}) - \alpha_1 \nabla f(\boldsymbol{x}^{(2)})^T \boldsymbol{p}^{(1)} + \frac{1}{2}\alpha_1^2 (\boldsymbol{p}^{(1)})^T G \boldsymbol{p}^{(1)} \\ &= f(\boldsymbol{x}^{(2)}) + \frac{1}{2}\alpha_1^2, \end{aligned} \tag{7.1.19}$$

所以

$$\alpha_1^2 = 2[f(\boldsymbol{x}^{(1)}) - f(\boldsymbol{x}^{(2)})]. \tag{7.1.20}$$

同理可得

$$\alpha_2^2 = 2[f(\boldsymbol{x}^{(2)}) - f(\boldsymbol{x}^{(3)})]. \tag{7.1.21}$$

关于 α_3 的计算就不能同理可得了，因为 $\boldsymbol{x}^{(3)}$不是 $\boldsymbol{p}^{(3)}$方向上的极小点，换句话说，α_3 不是函数$\phi(\alpha) = f(\boldsymbol{x}^{(1)} + \alpha \boldsymbol{p}^{(3)})$的极小点.

注意到 $\boldsymbol{x}^{(4)}$是从 $\boldsymbol{x}^{(1)}$出发，沿方向 $\boldsymbol{u}$ 作精确一维搜索得到的极小点，即

$$\boldsymbol{x}^{(4)} = \boldsymbol{x}^{(1)} + \bar{\alpha} \boldsymbol{u}, \tag{7.1.22}$$

其中 $\bar{\alpha}$ 是函数$\phi(\alpha)=f(\boldsymbol{x}^{(1)}+\alpha\boldsymbol{u})$上的极小点，这样得到

$$\boldsymbol{x}^{(4)}=\boldsymbol{x}^{(1)}+\bar{\alpha}\alpha_3\boldsymbol{p}^{(3)}, \tag{7.1.23}$$

其中 $\bar{\alpha}\alpha_3$ 是函数$\phi(\alpha)=f(\boldsymbol{x}^{(1)}+\alpha\boldsymbol{p}^{(3)})$上的极小点. 由前面的推导过程得到

$$(\bar{\alpha}\alpha_3)^2=2[f(\boldsymbol{x}^{(1)})-f(\boldsymbol{x}^{(4)})], \tag{7.1.24}$$

即

$$\alpha_3^2=\frac{2[f(\boldsymbol{x}^{(1)})-f(\boldsymbol{x}^{(4)})]}{\bar{\alpha}^2}. \tag{7.1.25}$$

由（7.1.16）式和（7.1.17）式，当$|\alpha_1|>|\alpha_3|$时，有 $\Delta_1>\Delta$，即共轭程度增加，因此应选用 $\boldsymbol{p}^{(2)}$、$\boldsymbol{p}^{(3)}$作下一轮的搜索方向，也就是说，用 $\boldsymbol{p}^{(3)}$替换 $\boldsymbol{p}^{(1)}$.

由（7.1.20）式和（7.1.25）式可知

$$|\alpha_1|>|\alpha_3|$$

等价于

$$2[f(\boldsymbol{x}^{(1)})-f(\boldsymbol{x}^{(2)})]>\frac{2[f(\boldsymbol{x}^{(1)})-f(\boldsymbol{x}^{(4)})]}{\bar{\alpha}^2}$$

等价于

$$|\bar{\alpha}|>\left(\frac{f(\boldsymbol{x}^{(1)})-f(\boldsymbol{x}^{(4)})}{f(\boldsymbol{x}^{(1)})-f(\boldsymbol{x}^{(2)})}\right)^{\frac{1}{2}}, \tag{7.1.26}$$

即当（7.1.26）式成立，用 $\boldsymbol{u}$ 替换方向$\boldsymbol{d}^{(1)}$.

同理可得，当

$$|\bar{\alpha}|>\left(\frac{f(x^{(1)})-f(x^{(4)})}{f(x^{(2)})-f(x^{(3)})}\right)^{\frac{1}{2}} \tag{7.1.27}$$

成立时，用 u 替换方向$d^{(2)}$.

在选用（7.1.26）式或（7.1.27）式前，自然要选择$|\alpha_1|$和$|\alpha_2|$中较大的一个，即下降量较大的一个作为判别项.

将上述推导过程推广到 n 维情况，选择下降量最大的方向 $\boldsymbol{d}^{(m)}$，其下降量为 μ，即

$$\mu = \max\{f(\boldsymbol{x}^{(i)}) - f(\boldsymbol{x}^{(i+1)}) \mid i = 1,2,\cdots,n\}$$
$$= f(\boldsymbol{x}^{(m)}) - f(\boldsymbol{x}^{(m+1)}),$$

判别条件为

$$|\bar{\alpha}| > \left(\frac{f(\boldsymbol{x}^{(1)}) - f(\boldsymbol{x}^{(n+2)})}{\mu}\right)^{\frac{1}{2}}, \qquad (7.1.28)$$

当（7.1.28）式成立时，用 $\boldsymbol{u}$ 替换 $\boldsymbol{d}^{(m)}$.

另外，最初的 n 个方向也不一定取坐标轴方向，只需满足线性无关即可．由此导出 Powell 直接方法．

算法 7.1.2（Powell 直接方法）

(1) 取初始点 $\boldsymbol{x}^{(1)}$，置 n 个线性无关的向量 $\boldsymbol{d}^{(1)}$，$\boldsymbol{d}^{(2)}$，$\cdots$，$\boldsymbol{d}^{(n)}$，置精度要求 ε.

(2) 对于 $i=1$，2，$\cdots$，n，求解一维问题

$$\min\phi_i(\alpha) = f(\boldsymbol{x}^{(i)} + \alpha\boldsymbol{d}^{(i)}),$$

得 α_i，置

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)} + \alpha_i\boldsymbol{d}^{(i)}.$$

(3) 令

$$\mu = \max\{f(\boldsymbol{x}^{(i)}) - f(\boldsymbol{x}^{(i+1)}) \mid i = 1,2,\cdots,n\}$$
$$= f(\boldsymbol{x}^{(m)}) - f(\boldsymbol{x}^{(m+1)}).$$

(4) 置 $\boldsymbol{u} = \boldsymbol{x}^{(n+1)} - \boldsymbol{x}^{(1)}$．求解一维问题

$$\min\phi(\alpha) = f(\boldsymbol{x}^{(1)} + \alpha\boldsymbol{u}),$$

得 $\bar{\alpha}$，置

$$\boldsymbol{x}^{(n+2)} = \boldsymbol{x}^{(1)} + \bar{\alpha}\boldsymbol{u}.$$

(5) 若 $\|\boldsymbol{x}^{(n+2)} - \boldsymbol{x}^{(1)}\| \leqslant \varepsilon$，则停止计算（$\boldsymbol{x}^{(n+2)}$作为问题的解）.

(6) 若

$$|\bar{\alpha}| > \left(\frac{f(\boldsymbol{x}^{(1)}) - f(\boldsymbol{x}^{(n+2)})}{\mu}\right)^{\frac{1}{2}},$$

则用 $\boldsymbol{u}$ 替换 $\boldsymbol{d}^{(m)}$，即置

$$d^{(i)} = d^{(i+1)}, \qquad i = m, m+1, \cdots, n-1,$$

$$d^{(n)} = u.$$

(7) 置 $x^{(1)} = x^{(n+2)}$，转 (2).

【例 7.1.1】 用 Powell 直接方法求解无约束问题

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

取 $x^{(1)} = (2, 2)^T$，$d^{(1)} = (1, 0)^T$，$d^{(2)} = (-1, 1)^T$，计算一轮.

解

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 8x_2 \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix},$$

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

一维搜索步长为

$$\alpha_k = -\frac{(d^{(k)})^T \nabla f(x^{(k)})}{(d^{(k)})^T G d^{(k)}} = -\frac{d_1 g_1 + d_2 g_2}{2d_1^2 + 8d_2^2}.$$

取 $x^{(1)} = (2, 2)^T$，$d^{(1)} = (1, 0)^T$，得 $\alpha_1 = -2$. 所以

$$x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = (2,2)^T - 2(1,0)^T = (0,2)^T.$$

再取 $d^{(2)} = (-1, 1)^T$，得 $\alpha_2 = -\frac{8}{5}$，所以

$$x^{(3)} = x^{(2)} + \alpha_2 d^{(2)} = (0,2)^T - \frac{8}{5}(-1,1)^T = \left(\frac{8}{5}, \frac{2}{5}\right)^T.$$

计算下降量

$$f(x^{(1)}) - f(x^{(2)}) = 4,$$

$$f(x^{(2)}) - f(x^{(3)}) = 12.8,$$

所以，$\mu = 12.8$，$m = 2$，计算

$$u = x^{(3)} - x^{(1)} = \left(\frac{8}{5}, \frac{2}{5}\right)^T - (2,2)^T = \left(-\frac{2}{5}, -\frac{8}{5}\right)^T,$$

---

* 为了计算方便，这里仍然需要计算函数的导数值。

得到 $\bar{\alpha}=\frac{17}{13}$，因此

$$x^{(4)}=x^{(1)}+\bar{\alpha}u=(2,2)^T+\frac{17}{13}\left(-\frac{2}{5},-\frac{8}{5}\right)^T=\left(\frac{96}{65},-\frac{6}{65}\right)^T,$$

作判别，计算

$$\left(\frac{f(x^{(1)})-f(x^{(4)})}{\mu}\right)^{\frac{1}{2}}=1.1787,$$

有

$$|\bar{\alpha}|>\left(\frac{f(x^{(1)})-f(x^{(4)})}{\mu}\right)^{\frac{1}{2}},$$

因此用 u 替换 $d^{(2)}$. 第二轮使用的搜索方向为

$$d^{(1)}=(1,0)^T,$$

$$d^{(2)}=\left(-\frac{2}{5},-\frac{8}{5}\right)^T.$$

§7.2 模式搜索方法

在这一节和下一节讨论一类直接搜索算法，利用这种算法建立的迭代点移动不需要使用一维搜索技巧.

模式搜索方法（Pattern Search Method）是 R. Hooke（霍克）和 T. A. Jeeves（乔维斯）于 1961 年提出的，因此也称为 Hook-Jeeves 方法，此方法有明显的几何意义，为介绍这种方法，从求一个二元函数的极小点谈起. 这相当于寻找某个曲面的最低点，或者形象地说，相当于从一座山岭的某处出发，设法走到附近某一盆地的最低点，怎样才能尽快达到这一目标呢？很显然，如果能找到一条山谷，沿山谷行进是最好的方法.

模式搜索方法就是根据上述思想设计的. 它由两部分组成，一个是探测移动，另一个是模式移动.

7.2.1 探测移动

这种移动是在某个已知点附近，沿坐标轴 e_i 方向进行探测，其目的是获得一个更小值的点.

取初始点 $x^{(1)}$，探测移动的具体过程如下：

记 $t_1^{(1)}=x^{(1)}$，先在坐标方向 e_1 上作探测. 设 α 是固定步长. 如果 $f(t_1^{(1)}+\alpha e_1)<f(t_1^{(1)})$,则沿 e_1 方向探测成功，置 $t_2^{(1)}=t_1^{(1)}+\alpha e_1$；否则探测失败，考虑相反的方向. 如果 $f(t_1^{(1)}-\alpha e_1)<f(t_1^{(1)})$,则沿 $-e_1$ 方向探测成功，置 $t_2^{(1)}=t_1^{(1)}-\alpha e_1$；否则探测失败，置 $t_2^{(1)}=t_1^{(1)}$.

再沿 e_2 方向进行探测. 如果 $f(t_2^{(1)}+\alpha e_2)<f(t_2^{(1)})$,则置 $t_3^{(1)}=t_2^{(1)}+\alpha e_2$;否则若 $f(t_2^{(1)}-\alpha e_2)<f(t_2^{(1)})$,则置 $t_3^{(1)}=t_2^{(1)}-\alpha e_2$;否置 $t_3^{(1)}=t_2^{(1)}$.

继续作下去，最后得到 $t_{n+1}^{(1)}$.

7.2.2 模式移动

在探测移动后，若 $f(t_{n+1}^{(1)})<f(x^{(1)})$,则作模式移动,即令

$$\begin{aligned} x^{(2)} &= t_{n+1}^{(1)}, \\ t_1^{(2)} &= x^{(2)}+(x^{(2)}-x^{(1)}). \end{aligned} \tag{7.2.1}$$

7.2.3 算法的基本思想

算法的基本思想是将探测移动和模式移动有机的结合在一起. 设已得到 $x^{(k)}$和 $t_1^{(k)}$，从 $t_1^{(k)}$ 出发，沿各个坐标轴作探测移动得到 $t_2^{(k)}$，$t_3^{(k)}$，…，$t_{n+1}^{(k)}$.

如果 $f(t_{n+1}^{(k)})<f(x^{(k)})$,则作模式移动,即令

$$\begin{aligned} x^{(k+1)} &= t_{n+1}^{(k)}, \\ t_1^{(k+1)} &= x^{(k+1)}+(x^{(k+1)}-x^{(k)}), \end{aligned} \tag{7.2.2}$$

置 $k=k+1$，重复上述计算，

如果

$$f(t_{n+1}^{(k)}) \geqslant f(x^{(k)}), \tag{7.2.3}$$

分两种情况讨论.

(1) 若$t_1^{(k)} \neq x^{(k)}$,则 $t_1^{(k)}$ 是由上一次的模式移动得到的,而(7.2.3)式表明该模式移动使目标函数值上升,即模式移动失败,应将 $t_1^{(k)}$ 退回到 $x^{(k)}$处,即令 $t_1^{(k)}=x^{(k)}$,重复上述计算.

(2) 若 $t_1^{(k)}=x^{(k)}$，这说明上次的模式移动失败且在 $t_1^{(k)}$ 周围的探测移动全部失败. 这是由于步长 α 过大的原因引起的，应减少步长，再重复上述计算.

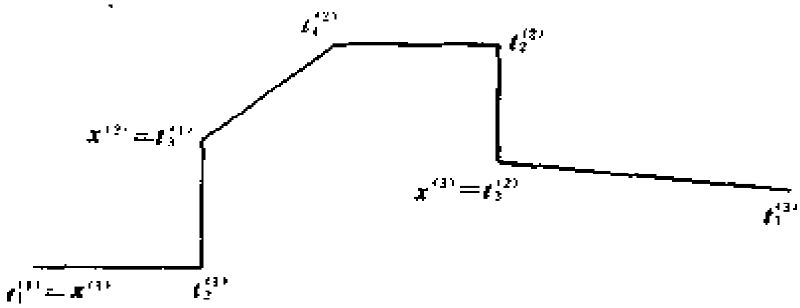


图 7.2.1

终止准则：

当步长 α 充分小，就认为 $x^{(k)}$在极小点附近了，终止计算. 因此终止准则为

$$\alpha < \varepsilon, \tag{7.2.4}$$

其中 ε 是预先给定的精度.

7.2.4 算法

由上述算法的基本思想，得到如下算法.

算法 7.2.1 (模式搜索法)

(1) 取初始点 $x^{(1)}$，初始步长 $\alpha>0$，置精度要求 ε. 置 $t_1=x^{(1)}, k=1$.

(2) 对于 $i=1, 2, \cdots, n$，做：如果 $f(\boldsymbol{t}_i+\alpha\boldsymbol{e}_i)<f(\boldsymbol{t}_i)$，则置 $\boldsymbol{t}_{i+1}=\boldsymbol{t}_i+\alpha\boldsymbol{e}_i$；否则若 $f(\boldsymbol{t}_i-\alpha\boldsymbol{e}_i)<f(\boldsymbol{t}_i)$，则置 $\boldsymbol{t}_{i+1}=\boldsymbol{t}_i-\alpha\boldsymbol{e}_i$；否则置 $\boldsymbol{t}_{i+1}=\boldsymbol{t}_i$.

(3) 若 $f(\boldsymbol{t}_{n+1})<f(\boldsymbol{x}^{(k)})$，则置

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{t}_{n+1},$$

$$\boldsymbol{t}_1=\boldsymbol{x}^{(k+1)}+(\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}),$$

置 $k=k+1$，转 (2).

(4) 若 $\boldsymbol{t}_1\neq\boldsymbol{x}^{(k)}$，则置 $\boldsymbol{t}_1=\boldsymbol{x}^{(k)}$，转 (2).

(5) 若 $\alpha<\varepsilon$，则停止计算；否则置 $\alpha=\alpha/2$，转 (2).

【例 7.2.1】 用模式搜索法求解无约束问题

$$\min f(\boldsymbol{x})=(x_1-1)^2+5(x_1^2-x_2)^2,$$

取初始点 $\boldsymbol{x}^{(1)}=(2, 0)^T$，步长 $\alpha=\frac{1}{2}$.

解 令 $\boldsymbol{t}_1=\boldsymbol{x}^{(1)}=(2, 0)^T$，$f(\boldsymbol{t}_1)=81$. 先在 $\boldsymbol{e}_1$ 方向上作探测移动：

$$\boldsymbol{t}_1+\alpha\boldsymbol{e}_1=(2,0)^T+\frac{1}{2}(1,0)^T=\left(\frac{5}{2},0\right)^T,$$

$$f(\boldsymbol{t}_1+\alpha\boldsymbol{e}_1)=197\frac{9}{16}>f(\boldsymbol{t}_1),$$

$$\boldsymbol{t}_1-\alpha\boldsymbol{e}_1=(2,0)^T-\frac{1}{2}(1,0)^T=\left(\frac{3}{2},0\right)^T,$$

$$f(\boldsymbol{t}_1-\alpha\boldsymbol{e}_1)=25\frac{9}{16}<f(\boldsymbol{t}_1),$$

所以

$$\boldsymbol{t}_2=\boldsymbol{t}_1-\alpha\boldsymbol{e}_1=\left(\frac{3}{2},0\right)^T,\ f(\boldsymbol{t}_2)=25\frac{9}{16}.$$

再作 $\boldsymbol{e}_2$ 方向上的探测移动：

$$\boldsymbol{t}_2+\alpha\boldsymbol{e}_2=\left(\frac{3}{2},0\right)^T+\frac{1}{2}(0,1)^T=\left(\frac{3}{2},\frac{1}{2}\right)^T,$$

$$f(t_2 + \alpha e_2) = 15\frac{9}{16} < f(t_2),$$

所以

$$t_3 = t_3 + \alpha e_2 = \left(\frac{3}{2}, \frac{1}{2}\right)^T, \ f(t_3) = 15\frac{9}{16}.$$

探测移动完成.

因为 $f(t_3) < f(x^{(1)}) = 81$,作模式移动,令

$$\begin{cases} x^{(2)} = t_3 = \left(\frac{3}{2}, \frac{1}{2}\right)^T, \\ t_1 = x^{(2)} + (x^{(2)} - x^{(1)}) = 2x^{(2)} - x^{(1)} \\ \quad = 2\left(\frac{3}{2}, \frac{1}{2}\right)^T - (2,0)^T = (1,1)^T. \end{cases}$$

第一轮计算完成.

作第二轮计算. 从 t_1 出发作探测移动, 对于 e_1 方向有

$$f(t_1) = 0, \ f(t_1 + \alpha e_1) = 8\frac{1}{16} > f(t_1),$$

$$f(t_1 - \alpha e_1) = 3\frac{1}{16} > f(t_1),$$

所以 $t_2 = t_1 = (1, 1)^T$. 对于 e_2 方向有

$$f(t_2) = 0, f(t_2 + \alpha e_2) = 1\frac{1}{4} > f(t_2),$$

$$f(t_2 - \alpha e_2) = 1\frac{1}{4} > f(t_2),$$

所以 $t_3 = t_2 = (1,1)^T, f(t_3) = 0$.

因为 $f(t_3) < f(x^{(2)}) = 15\frac{9}{16}$,再作模式移动

$$\begin{cases} x^{(3)} = t_3 = (1,1)^T, \\ t_1 = x^{(3)} + (x^{(3)} - x^{(2)}) = 2x^{(3)} - x^{(2)} \\ \quad = 2(1,1) - \left(\frac{3}{2}, \frac{1}{2}\right)^T = \left(\frac{1}{2}, \frac{3}{2}\right)^T. \end{cases}$$

第二轮计算完成.

作第三轮计算. 从 t_1 出发作探测移动, 对于 e_1 方向有

$$f(t_1)=8\frac{1}{16},\ f(t_1+\alpha e_1)=1\frac{1}{4}<f(t_1),$$

所以 $t_2=t_1+\alpha e_1=\left(1,\ \frac{3}{2}\right)^T$. 对于 e_2 方向有

$$f(t_2)=1\frac{1}{4},\ f(t_2+\alpha e_2)=5, f(t_2-\alpha e_2)=0<f(t_2),$$

所以 $t_3=t_2-\alpha e_2=(1,\ 1)^T$, $f(t_3)=0$.

此时, $f(t_3)=f(x^{(3)})$, 而 $t_1\neq x^{(3)}$, 说明第二轮中的探测移动失败, 将 t_1 退回到 $x^{(3)}$ 处, 即取 $t_1=x^{(3)}=(1,1)^T$, 再作探测移动. 我们发现全部探测移动失败, 减少步长 $\alpha=\frac{1}{2}\Big/2=\frac{1}{4}$.

再次作探测移动, 仍然失败, 再次减少步长, 最后得到最优点 $x^{(3)}=(1,\ 1)^T$.

§7.3 单纯形调优法

定义 7.3.1 在 n 维空间中由一组不在同一超平面的 $n+1$ 个点组成的几何图形称为单纯形. 当这些点的距离都相等时, 称为正单纯形.

在二维空间中, 单纯形为三角形, 在三维空间中, 单纯形为四面体.

单纯形调优法 (Simplex Method) 的基本思想是比较一般单纯形的 $n+1$ 个顶点的目标函数值, 并在迭代过程中逐渐地把单纯形向最优点移动. 这种方法最初是由 W. Spendley, G. R. Hext 和 F. R. Himsworth 设计的 (1962 年), 随后为 J. Nelder, R. Mead 所改进 (1964 年).

单纯形法的移动是通过称为反射、收缩和扩大的三种运算来实现的.

考虑无约束问题

$$\min f(\boldsymbol{x}), \quad \boldsymbol{x} \in R^n,$$

设 $\boldsymbol{x}^{(0)}$, $\boldsymbol{x}^{(1)}$, …, $\boldsymbol{x}^{(n)}$是构成单纯形的 R^n 中的 $n+1$ 个顶点，并定义

$$f(\boldsymbol{x}^{(h)}) = \max\{f(\boldsymbol{x}^{(i)}) \mid i = 0,1,\cdots,n\}, \tag{7.3.1}$$

$$f(\boldsymbol{x}^{(l)}) = \min\{f(\boldsymbol{x}^{(i)}) \mid i = 0,1,\cdots,n\}. \tag{7.3.2}$$

7.3.1 反射

由于 $\boldsymbol{x}^{(h)}$是单纯形中目标函数最大的顶点，则我们期望把它反射到对面 $\boldsymbol{x}^{(r)}$，得到较小函数值的点，构造出一个新的单纯形，使单纯形向函数值小的方向移动.

在作反射之前，先求出除 $\boldsymbol{x}^{(h)}$外的其余所有顶点的重心，即

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{\substack{i=0\\ i\neq h}}^{n} \boldsymbol{x}^{(i)}, \tag{7.3.3}$$

则反射点为

$$\boldsymbol{x}^{(r)} = \bar{\boldsymbol{x}} + \alpha(\bar{\boldsymbol{x}} - \boldsymbol{x}^{(h)}), \tag{7.3.4}$$

称 α 为反射系数，通常取 $\alpha = 1$.

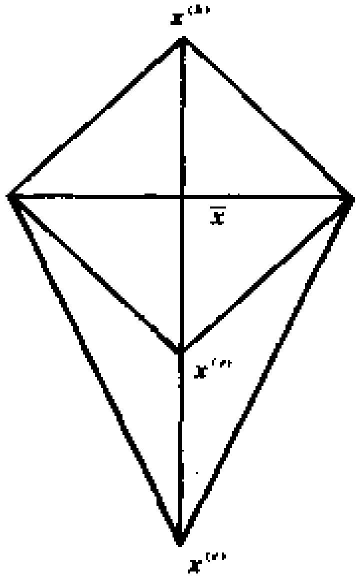


图 7.3.1

7.3.2 扩大

在得到反射点 $\boldsymbol{x}^{(r)}$后，将点 $\boldsymbol{x}^{(r)}$处的函数值 $f(\boldsymbol{x}^{(r)})$与 $f(\boldsymbol{x}^{(l)})$作比较. 若 $f(\boldsymbol{x}^{(r)}) < f(\boldsymbol{x}^{(l)})$，则反射成功. 并表明，当 $\boldsymbol{x}$ 向 $\boldsymbol{x}^{(r)}$方向移动时，目标函数 $f(\boldsymbol{x})$有可能进一步下降. 因此可以将 $\boldsymbol{x}^{(r)}$扩大到 $\boldsymbol{x}^{(e)}$，即令

$$x^{(e)} = \bar{x} + \gamma(x^{(r)} - \bar{x}), \tag{7.3.5}$$

称 γ 为扩大系数，通常取 $\gamma=2$.

如果 $f(x^{(e)})<f(x^{(l)})$,则表示扩大成功.用 $x^{(e)}$替换 $x^{(h)}$,得到一个新的单纯形.若 $f(x^{(e)})\geqslant f(x^{(l)})$,则表示扩大失败,此时用 $x^{(r)}$替换 $x^{(h)}$,得到一个新的单纯形.(见图 7.3.1).

7.3.3 收缩

若 $f(x^{(r)})\geqslant f(x^{(l)})$，考虑次最大值 $\max\{f(x^{(i)}\mid i=0,1,\cdots,n,i\neq h\}$，若

$$\max\{f(x^{(i)})\mid i=0,1,\cdots,n,i\neq h\}\geqslant f(x^{(r)})\geqslant f(x^{(l)}),$$

则仍用 $x^{(r)}$替换 $x^{(h)}$，得到一个新的单纯形.

若$f(x^{(r)})>\max\{f(x^{(i)})\mid i=0,1,\cdots,n,i\neq h\}$ 时，若用 $x^{(r)}$替换 $x^{(h)}$，会出现循环. 因此进行收缩，即令

$$x^{(c)} = \bar{x} + \beta(x^{(h')} - \bar{x}), \tag{7.3.6}$$

称 β 为收缩系数，通常取 $\beta=\frac{1}{2}$. 其中 $x^{(h')}$定义为

$$f(x^{(h')}) = \min\{f(x^{(h)}),f(x^{(r)})\}. \tag{7.3.7}$$

若$f(x^{(c)})<f(x^{(h')})$,则用 $x^{(c)}$替换 $x^{(h)}$，得到一个新的单纯形；否则进行缩边.（见图 7.3.2).

7.3.4 缩边

若$f(x^{(c)})>f(x^{(h')})$,则进行缩边，即置

$$x^{(i)} = x^{(l)} + \frac{1}{2}(x^{(i)} - x^{(l)}),\quad i=0,1,\cdots,n,i\neq l, \tag{7.3.8}$$

得到一个新的单纯（见图 7.3.3).

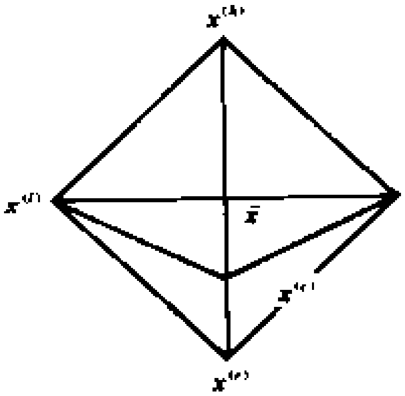


图 7.3.2

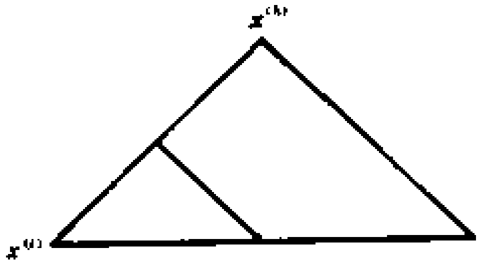


图 7.3.3

7.3.5 终止准则

当函数在当前单纯形的 $n+1$ 个顶点的目标函数值的标准差小于指定精度 ϵ 时，即

$$Q=\left\{\sum_{i=0}^{n}\frac{[f(x^{(i)})-f(\bar{x})]^2}{n+1}\right\}^{\frac{1}{2}}\leqslant\varepsilon, \qquad (7.3.9)$$

则停止计算.

7.3.6 算法

算法 7.3.1（单纯形调优法）

(1) 取初始点 $x^{(0)}$，初始步长 λ，置

$$x^{(i)}=x^{(0)}+\lambda e_i,\quad i=1,2,\cdots,n,$$

置 $\alpha=1$，$\beta=\frac{1}{2}$，$\gamma=2$，置精度要求 ε.

(2) 求 $x^{(h)}$、$x^{(l)}$，满足

$$f(x^{(h)})=\max\{f(x^{(i)})\mid i=0,1,\cdots,n\},$$
$$f(x^{(l)})=\min\{f(x^{(i)})\mid i=0,1,\cdots,n\}.$$

计算重心

$$\bar{x}=\frac{1}{n}\sum_{\substack{i=0\\ i\neq h}}^{n}x^{(i)}$$

和反射点

$$x^{(r)} = \bar{x} + \alpha(\bar{x} - x^{(h)}).$$

(3) 若$f(x^{(r)}) < f(x^{(l)})$,则计算扩大点

$$x^{(e)} = \bar{x} + \gamma(x^{(r)} - \bar{x}).$$

若$f(x^{(e)}) < f(x^{(l)})$,则置

$$x^{(h)} = x^{(e)};$$

否则置

$$x^{(h)} = x^{(r)},$$

然后转 (6).

(4) 若$f(x^{(r)}) < \max\{f(x^{(i)}) \mid i = 0, 1, \cdots, n, i \neq h\}$,则置

$$x^{(h)} = x^{(r)},$$

然后转 (6).

(5) 若$f(x^{(r)}) < f(x^{(h)})$,则置

$$x^{(h)} = x^{(r)}.$$

计算收缩点

$$x^{(c)} = \bar{x} + \beta(x^{(h)} - \bar{x}),$$

若$f(x^{(c)}) < f(x^{(h)})$,则置

$$x^{(h)} = x^{(c)};$$

否则进行缩边,置

$$x^{(i)} = x^{(l)} + \frac{1}{2}(x^{(i)} - x^{(l)}), \quad i = 0, 1, \cdots, n, i \neq l.$$

(6) 如果

$$\left\{\sum_{i=0}^{n} \frac{[f(x^{(i)}) - f(\bar{x})]^2}{n+1}\right\}^{\frac{1}{2}} \leqslant \varepsilon,$$

则停止计算($f(x^{(i)}), i = 0, 1, 2, \cdots, n$ 中的最小值点作为最优点);否则转 (2).

【例 7.3.1】 用单纯形调优法求解无约束问题

$$\min f(\boldsymbol{x}) = (x_1 - 1)^2 + 5(x_1^2 - x_2)^2,$$

取初始点 $\boldsymbol{x}^{(0)} = (2, 0)^T$，初始步长 $\lambda = \frac{1}{2}$，计算到单纯形三个初始点全部被替代为止.

解 $\boldsymbol{x}^{(0)} = (2, 0)^T$，$\lambda = \frac{1}{2}$，所以

$$\boldsymbol{x}^{(1)} = (2,0)^T + \frac{1}{2}(1,0)^T = \left(\frac{5}{2},0\right)^T,$$

$$\boldsymbol{x}^{(2)} = (2,0)^T + \frac{1}{2}(0,1)^T = \left(2,\frac{1}{2}\right)^T.$$

计算目标函数值

$$f(\boldsymbol{x}^{(0)}) = 81, f(\boldsymbol{x}^{(1)}) = 197\frac{9}{16}, f(\boldsymbol{x}^{(2)}) = 62\frac{1}{4},$$

因此，$l = 2$，$h = 1$. 计算重心

$$\overline{\boldsymbol{x}} = \frac{1}{2}(\boldsymbol{x}^{(0)} + \boldsymbol{x}^{(2)}) = \frac{1}{2}\left[(2,0)^T + \left(2,\frac{1}{2}\right)^T\right] = \left(2,\frac{1}{4}\right)^T.$$

反射点

$$\begin{aligned} \boldsymbol{x}^{(r)} &= \overline{\boldsymbol{x}} + \overline{\boldsymbol{x}} - \boldsymbol{x}^{(h)} = 2\overline{\boldsymbol{x}} - \boldsymbol{x}^{(1)} \\ &= 2\left(2,\frac{1}{4}\right)^T - \left(\frac{5}{2},0\right)^T = \left(\frac{3}{2},\frac{1}{2}\right)^T, \end{aligned}$$

$$f(\boldsymbol{x}^{(r)}) = 15\frac{9}{16} < f(\boldsymbol{x}^{(l)}).$$

反射成功，再计算扩大点

$$\begin{aligned} \boldsymbol{x}^{(e)} &= \overline{\boldsymbol{x}} + 2(\boldsymbol{x}^{(r)} - \overline{\boldsymbol{x}}) = 2\boldsymbol{x}^{(r)} - \overline{\boldsymbol{x}} \\ &= 2\left(\frac{3}{2},\frac{1}{2}\right)^T - \left(2,\frac{1}{4}\right)^T = \left(1,\frac{3}{4}\right)^T, \end{aligned}$$

$$f(\boldsymbol{x}^{(e)}) = \frac{5}{16} < f(\boldsymbol{x}^{(l)}).$$

扩大成功，置 $\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(e)} = \left(1, \frac{3}{4}\right)^T$，得到新的单纯形.

进行第二轮计算.

$$x^{(0)} = (2,0)^T,\ x^{(1)} = \left(1, \frac{3}{4}\right)^T,\ x^{(2)} = \left(2, \frac{1}{2}\right)^T,$$

$$f(x^{(0)}) = 81,\ f(x^{(1)}) = \frac{5}{16},\ f(x^{(2)}) = 62\frac{1}{4},$$

所以，$l=1$，$h=0$. 计算重心

$$\bar{x} = \frac{1}{2}(x^{(1)} + x^{(2)}) = \frac{1}{2}\left[\left(1, \frac{3}{4}\right)^T + \left(2, \frac{1}{2}\right)^T\right] = \left(\frac{3}{2}, \frac{5}{8}\right)^T.$$

反射点

$$\begin{aligned} x^{(r)} &= \bar{x} + \bar{x} - x^{(h)} = 2\bar{x} - x^{(0)} \\ &= 2\left(\frac{3}{2}, \frac{5}{8}\right)^T - (2,0)^T = \left(1, \frac{5}{4}\right)^T, \end{aligned}$$

$$f(x^{(r)}) = \frac{5}{16} = f(x^{(l)}) < f(x^{(2)}).$$

反射失败，但它小于次最大值点，所以置 $x^{(0)} = x^{(r)} = \left(1,\ \frac{5}{4}\right)^T$，得到新的单纯形.

进行第三轮计算.

$$x^{(0)} = \left(1, \frac{5}{4}\right)^T,\ x^{(1)} = \left(1, \frac{3}{4}\right)^T,\ x^{(2)} = \left(2, \frac{1}{2}\right)^T,$$

$$f(x^{(0)}) = \frac{5}{16},\ f(x^{(1)}) = \frac{5}{16},\ f(x^{(2)}) = 62\frac{1}{4},$$

所以，$h=2$，$l=0$（或 $l=1$）. 计算重心

$$\bar{x} = \frac{1}{2}(x^{(0)} + x^{(1)}) = \frac{1}{2}\left[\left(1, \frac{5}{4}\right)^T + \left(1, \frac{3}{4}\right)^T\right] = (1,1)^T.$$

反射点

$$\begin{aligned} x^{(r)} &= \bar{x} + \bar{x} - x^{(h)} = 2\bar{x} - x^{(2)} \\ &= 2(1,1)^T - \left(2, \frac{1}{2}\right)^T = \left(0, \frac{3}{2}\right)^T, \end{aligned}$$

$$f(x^{(r)}) = 12\frac{1}{4} > f(x^{(l)}) = f(x^{(0)}) = f(x^{(1)}).$$

反射失败，并且大于次最大值点，所以进行收缩. 因为 $f(x^{(r)})$

$< f(x^{(h)})$，令

$$x^{(h')} = x^{(r)} = \left(0, \frac{3}{2}\right)^T, \ f(x^{(h')}) = 12\frac{1}{4}.$$

计算收缩点

$$x^{(c)} = \bar{x} + \frac{1}{2}(x^{(h')} - \bar{x}) = \frac{1}{2}(x^{(h')} + \bar{x})$$

$$= \frac{1}{2}\left[\left(0, \frac{3}{2}\right)^T + (1,1)^T\right] = \left(\frac{1}{2}, \frac{5}{4}\right)^T,$$

$$f(x^{(c)}) = 5\frac{1}{4} < f(x^{(h')}).$$

收缩成功，置 $x^{(2)} = x^{(c)} = \left(\frac{1}{2}, \frac{5}{4}\right)^T$，得到新的单纯形.

到此三个初始点全部被替换，得到

$$x^{(0)} = \left(1, \frac{5}{4}\right)^T, \ x^{(1)} = \left(1, \frac{3}{4}\right)^T, \ x^{(2)} = \left(\frac{1}{2}, \frac{5}{4}\right)^T,$$

$$f(x^{(0)}) = \frac{5}{16}, \ f(x^{(1)}) = \frac{5}{16}, \ f(x^{(2)}) = 5\frac{1}{4},$$

最优解为 $\left(1, \frac{5}{4}\right)^T$ 或 $\left(1, \frac{3}{4}\right)^T$. （实际上，本题的最优解为 $x^* = (1,1)^T, f(x^*) = 0$）

习 题 七

7.1 设 A 是 n 阶正定对称矩阵，则

$$\det(A) \leqslant \prod_{i=1}^{n} a_{ii},$$

且等号成立的充分必要条件是：A 为对角阵.

7.2 计算下列各方向组的正交程度：

(1) $d^{(1)} = (1, 0)^T$，$d^{(2)} = (1, 1)^T$；

(2) $d^{(1)} = (1, 0)^T$，$d^{(2)} = (-1, 1)^T$；

(3) $d^{(1)} = (1, 0)^T$，$d^{(2)} = (-1, 0)^T$；

(4) $\boldsymbol{d}^{(1)}=(1,0)^T$, $\boldsymbol{d}^{(2)}=(0,1)^T$.

7.3 设

$$G=\begin{pmatrix}1 & 0\\ 0 & 4\end{pmatrix},$$

计算7.3题中各方向组的 G 共轭程度.

7.4 用Powell直接方法求解无约束问题

$$\min f(\boldsymbol{x}) = x_1^2 + 2x_2^2,$$

取 $\boldsymbol{x}^{(1)}=(2,2)^T$, $\boldsymbol{d}^{(1)}=(1,-1)^T$, $\boldsymbol{d}^{(2)}=(1,1)^T$, 问第二轮应选择哪两个方向作为新的搜索方向.

7.5 考虑求解 $\min f(\boldsymbol{x}), \boldsymbol{x}\in R^n$ 的Powell基本算法:

(1) 取初始点 $\boldsymbol{x}^{(1)}$和 n 个坐标轴方向 $\boldsymbol{d}^{(1)}$, $\boldsymbol{d}^{(2)}$, …, $\boldsymbol{d}^{(n)}$, 置精度要求 ε, 置 $\boldsymbol{z}^{(1)}=\boldsymbol{y}^{(1)}=\boldsymbol{x}^{(1)}$, 置 $k=1$, $j=1$.

(2) 对于 $i=1, 2, \cdots, n$, 求解一维问题

$$\min\phi(\alpha) = f(\boldsymbol{z}^{(i)} + \alpha\boldsymbol{d}^{(i)}),$$

得 α_i, 置

$$\boldsymbol{z}^{(i+1)} = \boldsymbol{z}^{(i)} + \alpha_i\boldsymbol{d}^{(i)}.$$

(3) 令 $\boldsymbol{u}=\boldsymbol{z}^{(n+1)}-\boldsymbol{z}^{(1)}$, 并求解一维问题

$$\min\phi(\alpha) = f(\boldsymbol{z}^{(n+1)} + \alpha\boldsymbol{u}),$$

得 $\hat{\alpha}$, 置

$$\boldsymbol{y}^{(j+1)} = \boldsymbol{z}^{(n+1)} + \hat{\alpha}\,\boldsymbol{u}.$$

(4) 如果 $j<n$, 则置

$$\boldsymbol{d}^{(l)} = \boldsymbol{d}^{(l+1)}, l = 1,2,\cdots,n-1, \boldsymbol{d}^{(n)} = \boldsymbol{u},$$

$$\boldsymbol{z}^{(1)} = \boldsymbol{y}^{(j+1)}, j = j+1,$$

转(2); 否则置 $\boldsymbol{x}^{(k+1)}=\boldsymbol{y}^{(n+1)}$, 转(5).

(5) 如果 $\|\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)}\|\leqslant\varepsilon$, 则停止计算; 否则置 $\boldsymbol{z}^{(1)}=\boldsymbol{y}^{(1)}=\boldsymbol{x}^{(k+1)}$, $j=1$, $k=k+1$, 转(2).

(a) 假设 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^TG\boldsymbol{x}+\boldsymbol{r}^T\boldsymbol{x}$, G 是正定对称矩阵, 证明:

若 n 个坐标轴方向全部被替换后，得到的搜索方向 $\boldsymbol{d}^{(1)}$，$\boldsymbol{d}^{(2)}$，…，$\boldsymbol{d}^{(n)}$是关于 G 共轭的，若进一步假设，$\boldsymbol{d}^{(i)}\neq 0$，$i=1$，2，…，$n$，则 Powell 基本算法具有二次终止性.

(b) 用 Powell 基本算法求解

$$\min f(\boldsymbol{x}) = (x_1 + x_2)^2 + (x_1 - 1)^2,$$

取初始点 $\boldsymbol{x}^{(1)} = (2, -1)^T$.

(c) 用 Powell 基本算法求解

$$\min f(\boldsymbol{x}) = (x_1 - x_2 + x_3)^2 + (-x_1 + x_2 + x_3)^2 + (x_1 + x_2 - x_3)^2,$$

取初始点 $\boldsymbol{x}^{(1)} = \left(\frac{1}{2}, 1, \frac{1}{2}\right)^T$，在求解过程中你会发现它达不到最优解 $(0, 0, 0)^T$，请解释其原因.

7.6 考虑正定二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x},$$

其中 G 为二阶矩阵，若 $\boldsymbol{x}^{(1)}\neq 0$，设从 $\boldsymbol{x}^{(1)}$出发沿最速下降方向作精确一维搜索得到点 $\boldsymbol{x}^{(2)}$，再从 $\boldsymbol{x}^{(2)}$出发沿最速下降方向作精确一维搜索得到点 $\boldsymbol{x}^{(3)}$，试证明：$\boldsymbol{x}^{(3)}$与 $\boldsymbol{x}^{(1)}$的连线通过原点.并讨论这一结论用于加速最速下降法的可能性.

7.7 用模式搜索法求解

$$\min f(\boldsymbol{x}) = x_1^2 + 2x_2^2,$$

取初始点 $\boldsymbol{x}^{(1)} = (-2, -1)^T$，初始步长 $\alpha = \frac{1}{2}$，计算 10 个～15 个点.

7.8 用单纯形调优法求解

$$\min f(\boldsymbol{x}) = x_1^2 + 2x_2^2,$$

取初始点 $\boldsymbol{x}^{(0)} = (-2, -1)^T$，初始步长 $\lambda = \frac{1}{2}$，计算到单纯形的三个初始点全部被取代为止.

第八章 约束问题的最优性条件

从本章起开始讨论约束最优化问题

$$
\begin{aligned}
\min \quad & f(x), \quad x \in R^n, \\
\text{s.t.} \quad & c_i(x) = 0, i \in E = \{1,2,\cdots,l\}, \\
& c_i(x) \leqslant 0, i \in I = \{l+1, l+2, \cdots, l+m\}.
\end{aligned} \tag{8.0.1}
$$

第一章已对约束最优化问题（8.0.1）作了简单的介绍，在这一章主要介绍约束问题的局部解及其必要条件和充分条件.

§8.1 约束问题局部解的概念

8.1.1 局部解的概念

在第一章中已介绍了约束问题的全局解的概念. 与无约束问题相类似，求出约束问题的全局解是非常困难的，这里介绍约束问题局部解的概念.

定义 8.1.1 对于一般约束最优化问题（8.0.1），记其可行域为

$$D = \{x \mid c_i(x) = 0, i \in E, c_i(x) \leqslant 0, i \in I\}, \tag{8.1.1}$$

若对 $x^* \in D$，存在 $\epsilon > 0$，使当 $x \in D$ 且 $\| x - x^* \| \leqslant \epsilon$ 时，总有

$$f(x) \geqslant f(x^*),$$

则称 x^* 为约束问题（8.0.1）的局部解，或简称 x^* 为最优解. 若当 $x \in D$ 且 $0 < \| x - x^* \| \leqslant \epsilon$，总有

$$f(\boldsymbol{x}) > f(\boldsymbol{x}^*),$$

则称 $\boldsymbol{x}^*$ 为约束问题的严格局部解.

【例 8.1.1】 试确定约束问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = x_2^2 + x_2^2, \\ \text{s.t.} \quad & c_1(\boldsymbol{x}) = (x_1 - 1)^2 - x_2^2 - 4 = 0, \\ & c_2(\boldsymbol{x}) = x_2 - 1 \leqslant 0 \end{aligned}$$

的局部解.

由图 8.1.1 可以得到问题的局部解 $\boldsymbol{x}^* = (-1, 0)^T$ 和 $\hat{\boldsymbol{x}}^* = (3, 0)^T$. 其中 $\boldsymbol{x}^*$ 是全局解.

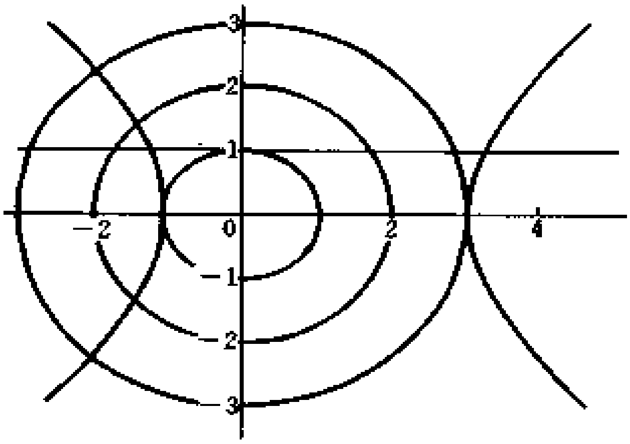


图 8.1.1 约束问题的局部解

8.1.2 约束问题局部解的一阶必要条件

前面介绍了局部解的概念，下面从几何直观给出约束最优化问题的一阶必要条件，在下一节再给出严格的证明.

1. 等式约束问题

考虑两个变量一个等式约束的简单情况：

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = f(x_1, x_2), \\ \text{s.t.} \quad & c(\boldsymbol{x}) = c(x_1, x_2) = 0. \end{aligned} \tag{8.1.2}$$

从图 8.1.2 可以看出，若 $\boldsymbol{x}^*$ 是约束问题的局部解，则在

x^* 处目标函数 $f(x)$ 的梯度 $\nabla f(x^*)$ 和约束函数 $c(x)$ 的梯度 $\nabla c(x^*)$ 共线，即存在 λ^*，使得

$$\nabla f(x^*) + \lambda^* \nabla c(x^*) = 0,$$

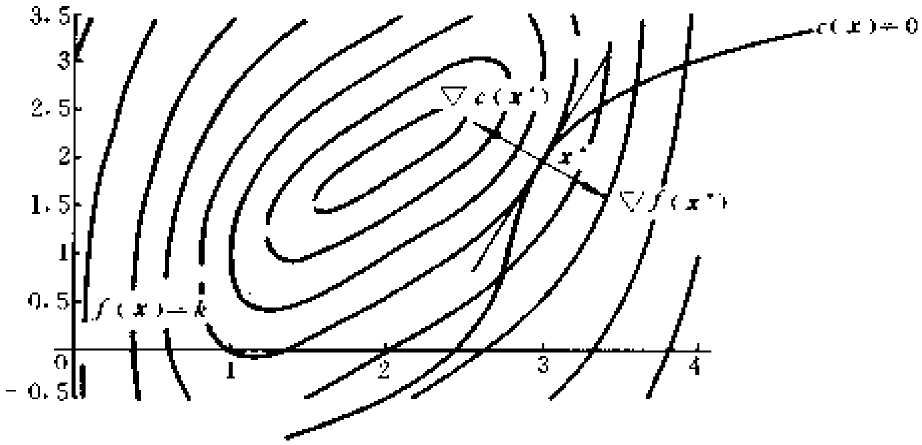


图 8.1.2　等式约束问题局部解的几何意义（二维）

写成分量形式

$$\begin{cases} \dfrac{\partial f}{\partial x_1} + \lambda^* \dfrac{\partial c}{\partial x_1} = 0, \\ \dfrac{\partial f}{\partial x_2} + \lambda^* \dfrac{\partial c}{\partial x_2} = 0, \end{cases}$$

和满足约束条件

$$c(x^*) = 0.$$

这一点与数学分析方面的内容完全相同.

为了方便起见，引入 Lagrange 函数

$$L(x, \lambda) = f(x) + \lambda c(x),$$

由此得到，若 x^* 是约束问题（8.1.2）的局部解，则存在 λ^*，使得

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \lambda^* \nabla c(x^*) = 0, \\ c(x^*) = 0 \quad (\text{相当于 } \nabla_\lambda L(x^*, \lambda^*) = 0). \end{cases}$$

注意，这里 Lagrange 函数相当于无约束问题中目标函数所起的

作用，即在局部解处其梯度为0.

再讨论一个三维变量两个等式约束的约束问题：

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = f(x_1, x_2, x_3), \\ \text{s.t.} \quad & c_1(\boldsymbol{x}) = c_1(x_1, x_2, x_3) = 0, \\ & c_2(\boldsymbol{x}) = c_2(x_1, x_2, x_3) = 0. \end{aligned} \tag{8.1.3}$$

类似地画出可行区域和等高面（见图8.1.3）.

可行域 D 是曲面 $c_1(\boldsymbol{x})=0$ 和 $c_2(\boldsymbol{x})=0$ 的交线，若 $\boldsymbol{x}^*$ 为约束问题（8.1.3）的局部解，设在 $\boldsymbol{x}^*$ 处曲线 D 的切线方向为 T，由梯度的性质得到，T 与 $\nabla c_1(\boldsymbol{x}^*)$ 和 $\nabla c_2(\boldsymbol{x}^*)$ 正交.

另一方面，因为 $\boldsymbol{x}^*$ 是 $f(\boldsymbol{x})$ 在 D 上的极小点，因此等高面 $f(\boldsymbol{x}) = f(\boldsymbol{x}^*)$ 与可行域 D 相切，即 T 与 $\nabla f(\boldsymbol{x}^*)$ 正交. 由此得到 $\nabla f(\boldsymbol{x}^*)$、$\nabla c_1(\boldsymbol{x}^*)$ 和 $\nabla c_2(\boldsymbol{x}^*)$ 共面，即存在 λ_1^*、λ_2^*，使得

$$\nabla f(\boldsymbol{x}^*) + \lambda_1^* \nabla c_1(\boldsymbol{x}^*) + \lambda_2^* \nabla c_2(\boldsymbol{x}^*) = 0,$$

同时，$\boldsymbol{x}^*$ 还满足

$$c_1(\boldsymbol{x}^*) = 0, \quad c_2(\boldsymbol{x}^*) = 0.$$

引入 Lagrange 函数

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \lambda_1 c_1(\boldsymbol{x}) + \lambda_2 c_2(\boldsymbol{x}),$$

若 $\boldsymbol{x}^*$ 是约束问题（8.1.3）的局部解，则存在 $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*)^T$，使得

$$\begin{cases} \nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \nabla f(\boldsymbol{x}^*) + \lambda_1^* \nabla c_1(\boldsymbol{x}^*) + \lambda_2^* \nabla c_2(\boldsymbol{x}^*) = 0, \\ c(\boldsymbol{x}^*) = 0 \quad (\text{即 } c_1(\boldsymbol{x}^*) = 0, c_2(\boldsymbol{x}^*) = 0). \end{cases}$$

2. 不等式约束问题

仍然从简单的情况出发，给出不等式约束问题的一阶必要条件.

考虑不等式约束问题：

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = f(x_1, x_2), \\ \text{s.t.} \quad & c(\boldsymbol{x}) = c(x_1, x_2) \leqslant 0, \end{aligned} \tag{8.1.4}$$

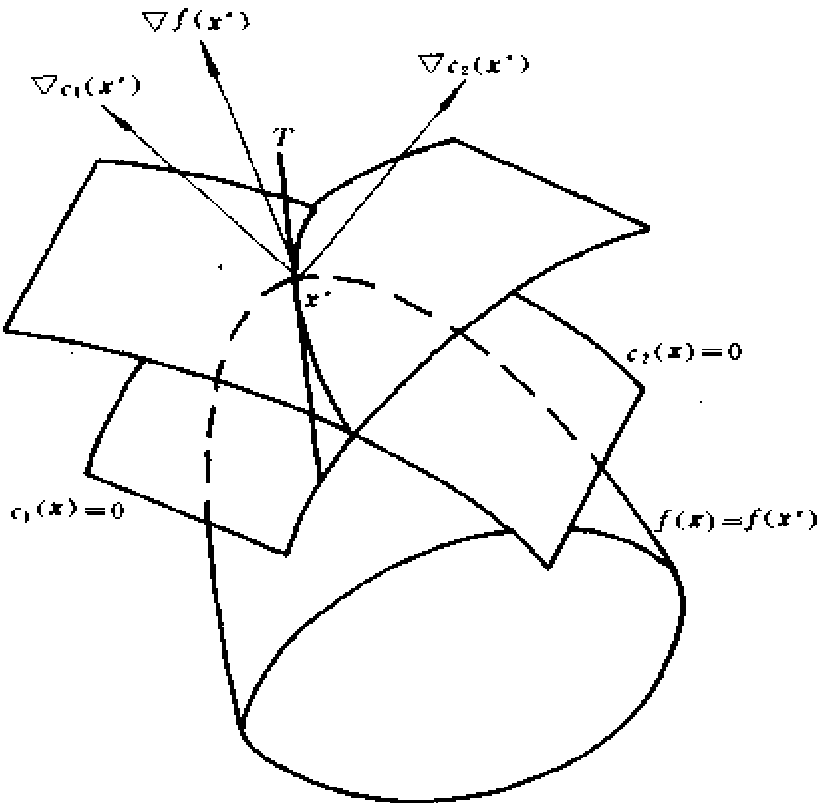


图 8.1.3　等式约束问题局部解的几何意义（三维）

借助于几何图形来讨论一阶必要条件. 分两种情况进行讨论.

(1) $c(x^*)=0$，此时 x^* 在约束的边界上，因此 x^* 也是等式约束问题 (8.1.2) 的局部解，由等式约束的一阶必要条件，即存在 λ^*，使得

$$\nabla f(x^*) + \lambda^* \nabla c(x^*) = 0.$$

考虑到 $\nabla f(x^*)$ 与 $\nabla c(x^*)$ 的方向相反，因此有 $\lambda^* \geqslant 0$，见图 8.1.4. 因此，一阶必要条件为

$$\begin{aligned} &\nabla f(x^*) + \lambda^* \nabla c(x^*) = 0, \\ &\lambda^* \geqslant 0, \\ &c(x^*) = 0. \end{aligned} \tag{8.1.5}$$

(2) $c(x^*)<0$，表明 x^* 在约束的内部，因此 x^* 是无约束问题 $\min\{f(x) \mid x \in R^n\}$. 换句话说，约束 $c(x) \leqslant 0$ 不起作用，

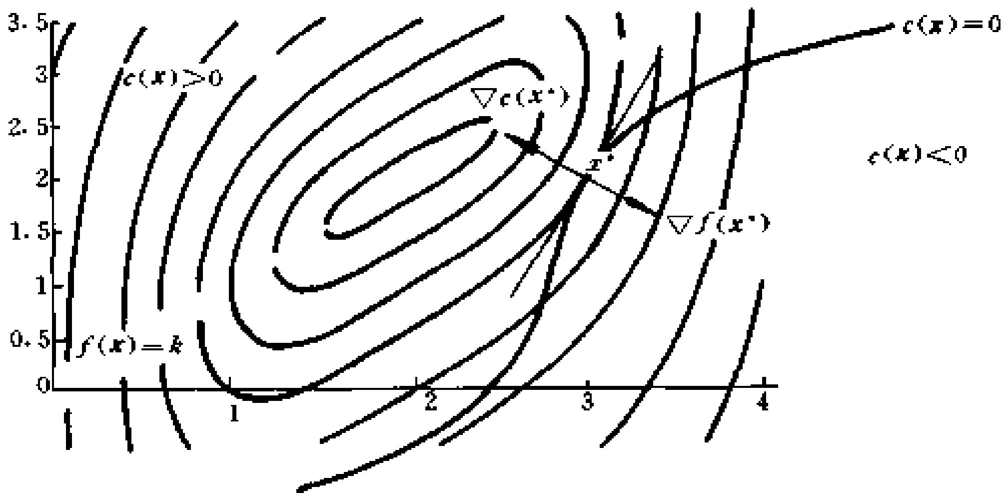


图 8.1.4　不等式约束问题局部解的几何意义($c(x^*)=0$)

即是非有效的. 因此由无约束问题的一阶必要条件,有

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

因此,其一阶必要条件可写成

$$\begin{aligned} &\nabla f(x^*) + \lambda^* \nabla c(x^*) = 0, \\ &\lambda^* = 0, \\ &c(x^*) < 0. \end{aligned} \tag{8.1.6}$$

将(8.1.5)式和(8.1.6)式写成统一的形式. 引进 Lagrange 函数

$$L(x,\lambda) = f(x) + \lambda c(x),$$

若 x^* 是约束问题(8.1.4)的局部解,则存在 λ^* 满足

$$\begin{aligned} &\nabla_x L(x^*,\lambda^*) = \nabla f(x^*) + \lambda^* \nabla c(x^*) = 0, \\ &c(x^*) \leqslant 0, \\ &\lambda^* \geqslant 0, \\ &\lambda^* c(x^*) = 0. \end{aligned} \tag{8.1.7}$$

3. 一般约束问题

从前面的几何直观, 似乎可以得到如下结论: 若 x^* 是约束问题 (8.0.1) 的局部解, 则存在 $\lambda^* = (\lambda_1^*, \lambda_2^*, \cdots,$

$\lambda_{l+m}^*)^T$，使得

$$\begin{aligned}
&\nabla_x L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*) = \nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{l+m}\lambda_i^*\nabla c_i(\boldsymbol{x}^*) = 0,\\
&c_i(\boldsymbol{x}^*) = 0,\quad i\in E = \{1,2,\cdots,l\},\\
&c_i(\boldsymbol{x}^*) \leqslant 0,\quad i\in I = \{l+1,l+2,\cdots,l+m\},\\
&\lambda_i^* \geqslant 0,\quad i\in I = \{l+1,l+2,\cdots,l+m\},\\
&\lambda_i^* c_i(\boldsymbol{x}^*) = 0,\quad i\in I = \{l+1,l+2,\cdots,l+m\}.
\end{aligned}\tag{8.1.8}$$

其中$L(x,\lambda)$为 Lagrange 函数

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{l+m}\lambda_i c_i(\boldsymbol{x}).\tag{8.1.9}$$

§8.2 约束问题局部解的必要条件

上节介绍了约束问题局部解的概念，并从直观上给出了局部解的一阶必要条件，本节将给出严格的数学证明.

8.2.1 Farkas 引理

定义 8.2.1 设集合 $C\subset R^n$ 非空，如果 $\forall \boldsymbol{x}\in C$，$\lambda\geqslant 0$，都有

$$\lambda\boldsymbol{x}\in C,$$

则称 C 为 R^n 的一个**锥**（Cone）. 若 C 是锥，又是凸集，则称 C 是一个**凸锥**（Convex Cone）.

【例 8.2.1】 设集合 $S\subset R^n$ 非空，证明：

$$C(S) = \{\lambda\boldsymbol{x}\mid \boldsymbol{x}\in S,\lambda\geqslant 0\}\tag{8.2.1}$$

是一个锥，称为 S 的生成锥. 若 S 是凸集，则 $C(S)$是凸锥（习题 8.1）.

【例 8.2.2】 设 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，…，$\boldsymbol{a}_m \in R^n$，证明：

$$C = \left\{ \sum_{i=1}^{m} \lambda_i \boldsymbol{a}_i \mid \lambda_i \geqslant 0, i = 1,2,\cdots,m \right\} \tag{8.2.2}$$

是凸锥（习题 8.2）.

定理 8.2.1 设 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，…，$\boldsymbol{a}_m$ 和 $\boldsymbol{w} \in R^n$，C 是由 (8.2.2) 式定义的. 若 $\boldsymbol{w} \notin C$，则存在超平面 $\boldsymbol{d}^T\boldsymbol{x} = 0$，分离 C 和 $\boldsymbol{w}$，即

$$\begin{aligned} &\boldsymbol{d}^T\boldsymbol{x} \leqslant 0, \qquad \forall \boldsymbol{x} \in C, \\ &\boldsymbol{d}^T\boldsymbol{w} > 0. \end{aligned} \tag{8.2.3}$$

证明 显然 C 是闭凸集. 由第一章的定理 1.4.3 得到，存在 $\boldsymbol{d} \in R^n$，$\boldsymbol{d} \neq 0$ 和 $\alpha \in R$，使得

$$\boldsymbol{d}^T\boldsymbol{x} \leqslant \alpha, \qquad \forall \boldsymbol{x} \in C,$$

$$\boldsymbol{d}^T\boldsymbol{w} > \alpha.$$

因为 C 是锥，有 $0 \in C$，因此 $\alpha \geqslant 0$，即 $\boldsymbol{d}^T\boldsymbol{w} > 0$. 下面证明

$$\boldsymbol{d}^T\boldsymbol{x} \leqslant 0, \quad \forall \boldsymbol{x} \in C.$$

若存在 $\bar{\boldsymbol{x}} \in C$，使得 $\boldsymbol{d}^T\bar{\boldsymbol{x}} > 0$，则对一切 $\lambda \geqslant 0$，$\lambda\bar{\boldsymbol{x}} \in C$，因此有

$$\lambda \boldsymbol{d}^T \bar{\boldsymbol{x}} \leqslant \alpha$$

当 $\lambda \to \infty$ 时，不等式左端趋于无穷，这与 α 是固定的数矛盾.

定理 8.2.2（Farkas 引理） 设 $\boldsymbol{a}_1$，$\boldsymbol{a}_2$，…，$\boldsymbol{a}_m$ 和 $\boldsymbol{w} \in R^n$.

系统Ⅰ：存在 $\boldsymbol{d}$ 满足

$$\begin{aligned} &\boldsymbol{a}_i^T\boldsymbol{d} \leqslant 0, \qquad i = 1,2,\cdots,m, \\ &\boldsymbol{w}^T\boldsymbol{d} > 0; \end{aligned} \tag{8.2.4}$$

系统Ⅱ：存在非负常数 λ_1，λ_2，…，λ_m，使得

$$\boldsymbol{w} = \sum_{i=1}^{m} \lambda_i \boldsymbol{a}_i, \tag{8.2.5}$$

则两系统有且仅有一个有解.

证明 分两种情况来讨论.

(1) 若系统Ⅱ有解，则系统Ⅰ无解.

设系统Ⅱ有解，即存在 $\lambda_1,\lambda_2,\cdots,\lambda_m$ 且 $\lambda_i\geqslant 0, i=1,2,\cdots,m$，使得

$$w = \sum_{i=1}^{m}\lambda_i a_i,$$

若系统Ⅰ有解，则有

$$0 < w^T d = \sum_{i=1}^{m}\lambda_i a_i^T d \leqslant 0. \tag{8.2.6}$$

矛盾.因此系统Ⅰ无解.

(2)若系统Ⅱ无解，则系统Ⅰ有解.

设系统Ⅱ无解，构造集合

$$C = \left\{ v \mid v = \sum_{i=1}^{m}\lambda_i a_i, \lambda_i \geqslant 0, i = 1,2,\cdots,m \right\},$$

显然 C 是闭凸锥.系统Ⅱ无解表明，$w \notin C$，由定理 8.2.1，则存在 d 满足

$$d^T x \leqslant 0, \quad \forall x \in C,$$
$$d^T w > 0.$$

特别地，取 $x = a_i, i = 1,2,\cdots,m$，则系统Ⅰ有解.

Farkas 引理是由 Farkas 在 1902 年给出的，它有多种叙述形式，我们在这里给出的是最常见的形式之一.

推论 1 设 $a_1,a_2,\cdots,a_m$ 和 $w\in R^n$.

系统Ⅰ：存在 d 满足

$$\begin{aligned} & a_i^T d \leqslant 0, \quad i = 1,2,\cdots,m, \\ & d_j \geqslant 0, \quad j = 1,2,\cdots,n, \\ & w^T d > 0, \end{aligned} \tag{8.2.7}$$

系统Ⅱ：存在非负常数 $\lambda_1,\lambda_2,\cdots,\lambda_m$，使得

$$w \leqslant \sum_{i=1}^{m}\lambda_i a_i \tag{8.2.8}$$

则两系统有且仅有一个有解(留作习题).

推论 2 设 $a_1,a_2,\cdots a_{l+m}$ 和 $w\in R^n$.

系统 Ⅰ:存在 d 满足

$$\begin{aligned} &a_i^Td = 0, \quad i = 1,2,\cdots,l, \\ &a_i^Td \leqslant 0, \quad i = l+1,l+2,\cdots,l+m, \\ &w^Td > 0, \end{aligned} \tag{8.2.9}$$

系统 Ⅱ:存在常数 $\lambda_1,\lambda_2,\cdots,\lambda_{l+m}$ 且 $\lambda_i\geqslant 0, i=l+1,l+2,\cdots,l+m$,使得

$$w = \sum_{i=1}^{l+m}\lambda_i a_i, \tag{8.2.10}$$

则两系统有且仅有一个有解(留作习题).

8.2.2 约束问题局部解的一阶必要条件

考虑一般约束问题

$$\begin{aligned} \min \quad & f(x), && x\in R^n, \\ \text{s.t.} \quad & c_i(x) = 0, && i\in E = \{1,2,\cdots,l\}, \\ & c_i(x) \leqslant 0, && i\in I = \{l+1,l+2,\cdots,l+m\}, \end{aligned} \tag{8.2.11}$$

这里我们假设 $f(x)$、$c_i(x)(i=1,2,\cdots,l+m)$是连续可微函数.

设 x^* 是约束问题(8.2.11)的可行点,$\{x^{(k)}\}$是约束问题(8.2.11)的可行点列,并满足 $x^{(k)}\to x^*(k\to\infty)$且 $x^{(k)}\neq x^*$,记

$$x^{(k)} = x^* + \delta_k d^{(k)}, \tag{8.2.12}$$

其中 $d^{(k)}$有固定模长,而 $\delta_k>0$,由于 $x^{(k)}\to x^*$,因此有 $\delta_k\to 0$.(例如,$d^{(k)} = \dfrac{x^{(k)}-x^*}{\|x^{(k)}-x^*\|}$,$\delta_k = \|x^{(k)}-x^*\|$)

若$\{d^{(k)}\}$有极限,即 $d^{(k)}\to d$,则称$\{x^{(k)}\}$为方向可行点列,$d^{(k)}$为方向序列,而称 d 为 x^* 处的**可行方向**(Feasible Direction).

记

$$\mathscr{F}^* = \mathscr{F}(x^*) = \{d \mid d \text{ 是 } x^* \text{ 处的可行方向}\} \tag{8.2.13}$$

为全体可行方向的集合,因此,由定义得知,若 $d\in\mathscr{F}^*$ 的充分必要

条件是:存在可行点列 $\boldsymbol{x}^{(k)}=\boldsymbol{x}^*+\delta_k\boldsymbol{d}^{(k)}$, $\|\boldsymbol{d}^{(k)}\|=\|\boldsymbol{d}\|$, $\boldsymbol{d}^{(k)}\to\boldsymbol{d}$ 和 $\delta_k\to 0$.

显然,$\mathscr{F}^*\cup\{0\}$ 为一锥,称为 $\boldsymbol{x}^*$ 处的**切锥**(Tangent Cone).

【例 8.2.3】 考虑可行域

$$D=\{\boldsymbol{x}=(x_1,x_2)^T\mid x_2-x_1^3\leqslant 0,\ -x_2\leqslant 0\},\tag{8.2.14}$$

设 $\boldsymbol{x}^*=(0,0)^T$, $\{\boldsymbol{x}^{(k)}\}$ 为曲线 $x_2=x_1^3$ 上满足 $x_2^{(k)}>0$ 的一个趋于 $\boldsymbol{x}^*$ 的点列,则

$$\boldsymbol{d}^{(k)}=\frac{\boldsymbol{x}^{(k)}-\boldsymbol{x}^*}{\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\|}\to(1,0)^T,$$

故 $\boldsymbol{d}=(1,0)^T\in\mathscr{F}^*$.

定义 8.2.2 设 $\hat{\boldsymbol{x}}$ 是一般约束问题(8.2.11)的可行点,当 $i\in I$ 时,对某个约束有 $c_i(\hat{\boldsymbol{x}})=0$,则称约束 $c_i(\boldsymbol{x})\leqslant 0$ 为 $\hat{\boldsymbol{x}}$ 处的**有效约束**(Active Constraints),若 $c_i(\hat{\boldsymbol{x}})<0$,则称约束 $c_i(\boldsymbol{x})\leqslant 0$ 为 $\hat{\boldsymbol{x}}$ 处的非有效约束.

定义有效约束指标集

$$I(\hat{\boldsymbol{x}})=\{i\mid c_i(\hat{\boldsymbol{x}})=0,i\in I\},\tag{8.2.15}$$

简称 $I(\hat{\boldsymbol{x}})$ 为 $\hat{\boldsymbol{x}}$ 处的**有效集**(Active Set).

设 $\boldsymbol{x}^*$ 是约束问题(8.2.11)的可行点,定义

$$\begin{aligned}F^*=F(\boldsymbol{x}^*)=\{&\boldsymbol{d}\mid\boldsymbol{d}\neq 0,\boldsymbol{d}^T\nabla c_i(\boldsymbol{x}^*)=0,\\&i\in E,\boldsymbol{d}^T\nabla c_i(\boldsymbol{x}^*)\leqslant 0,i\in I(\boldsymbol{x}^*)\}.\end{aligned}\tag{8.2.16}$$

显然,$F^*\cup\{0\}$ 为一锥,称为 $\boldsymbol{x}^*$ 处的线性化锥.

引理 8.2.1 $\mathscr{F}^*\subset F^*$.

证明 设 $\boldsymbol{d}\in\mathscr{F}^*$,则存在 $\boldsymbol{x}^{(k)}=\boldsymbol{x}^*+\delta_k\boldsymbol{d}^{(k)}$ 是可行点,且 $\delta_k\to 0$, $\boldsymbol{d}^{(k)}\to\boldsymbol{d}$. 由 Taylor 展开式

$$c_i(\boldsymbol{x}^{(k)})=c_i(\boldsymbol{x}^*)+\delta_k\boldsymbol{d}^{(k)T}\nabla c_i(\boldsymbol{x}^*)+o(\delta_k),\tag{8.2.17}$$

当 $i\in E$ 时，$c_i(\boldsymbol{x}^{(k)})=0$，$c_i(\boldsymbol{x}^*)=0$，所以(8.2.17)式化简为

$$\delta_k \boldsymbol{d}^{(k)T}\nabla c_i(\boldsymbol{x}^*)+o(\delta_k)=0, \tag{8.2.18}$$

在(8.2.18)式两端同除 δ_k，并令 $k\to\infty$，得到

$$\boldsymbol{d}^T\nabla c_i(\boldsymbol{x}^*)=0. \tag{8.2.19}$$

当 $i\in I(\boldsymbol{x}^*)$时，$c_i(\boldsymbol{x}^{(k)})\leqslant 0$，$c_i(\boldsymbol{x}^*)=0$，所以(8.2.17)式化简为

$$\delta_k \boldsymbol{d}^{(k)T}\nabla c_i(\boldsymbol{x}^*)+o(\delta_k)\leqslant 0, \tag{8.2.20}$$

在(8.2.20)式两端同除 δ_k，并令 $k\to\infty$，得到

$$\boldsymbol{d}^T\nabla c_i(\boldsymbol{x}^*)\leqslant 0, \tag{8.2.21}$$

因此，$\boldsymbol{d}\in F^*$.

命题反之并不成立. 如可行域取例 8.2.3 中的 D，可行点 $\boldsymbol{x}^*=(0,0)^T$，则$\nabla c_1(\boldsymbol{x}^*)=(0,1)^T$，$\nabla c_2(\boldsymbol{x}^*)=(0,-1)^T$，故

$$\begin{aligned}F^* &= \{\boldsymbol{d}=(d_1,d_2)^T \mid \boldsymbol{d}\neq 0,\boldsymbol{d}^T\nabla c_1(\boldsymbol{x}^*)\leqslant 0,\\ &\qquad \boldsymbol{d}^T\nabla c_2(\boldsymbol{x}^*)\leqslant 0\}\\ &= \{\boldsymbol{d}=(d_1,d_2)^T \mid d_1\neq 0,d_2=0\}\\ &= \{(d_1,0)^T \mid d_1\neq 0\},\end{aligned}$$

显然，$\boldsymbol{d}=(-1,0)^T\in F^*$，但由(8.2.14)式确定的可行域 D 对于所有的可行点 $\boldsymbol{x}^{(k)}$ 有 $x_1^{(k)}\geqslant 0$，故不存在这样的点列 $\boldsymbol{x}^{(k)}=\boldsymbol{x}^*+\delta_k\boldsymbol{d}^{(k)}=\delta_k\boldsymbol{d}^{(k)}$，使得 $\boldsymbol{d}^{(k)}\to\boldsymbol{d}=(-1,0)^T$，因此，$\boldsymbol{d}\notin\mathscr{F}^*$.

因此，要使 $F^*=\mathscr{F}^*$，需要对约束附加条件，通常称任何一个保证 $F^*=\mathscr{F}^*$ 成立的条件为**约束限制条件**(Constraint Qualification).

在预备知识中谈过，若方向 $\boldsymbol{d}$ 的一阶方向导数小于 0，则该方向是下降方向，记

$$\mathscr{D}^*=\mathscr{D}(\boldsymbol{x}^*)=\{\boldsymbol{d}\mid \boldsymbol{d}^T\nabla f(\boldsymbol{x}^*)<0\}. \tag{8.2.22}$$

定理 8.2.3 若 $\boldsymbol{x}^*$ 是约束问题(8.2.11)的局部解，则

$$\mathscr{F}^*\cap\mathscr{D}^*=\phi. \tag{8.2.23}$$

证明 任取 $\boldsymbol{d}\in\mathscr{F}^*$，则存在可行点列 $\boldsymbol{x}^{(k)}=\boldsymbol{x}^*+\delta_k\boldsymbol{d}^{(k)}$，并且 $\delta_k\to0$ 和 $\boldsymbol{d}^{(k)}\to\boldsymbol{d}$. 由 Taylor 展式

$$f(\boldsymbol{x}^{(k)})=f(\boldsymbol{x}^*)+\delta_k\boldsymbol{d}^{(k)T}\nabla f(\boldsymbol{x}^*)+o(\delta_k), \tag{8.2.24}$$

因为 $\boldsymbol{x}^*$ 是局部解，存在着 K，当 $k\geqslant K$ 时，有 $f(\boldsymbol{x}^{(k)})\geqslant f(\boldsymbol{x}^*)$，由(8.2.24)式得到

$$\delta_k\boldsymbol{d}^{(k)T}\nabla f(\boldsymbol{x}^*)+o(\delta_k)\geqslant0, \tag{8.2.25}$$

在(8.2.25)式两端同除 δ_k，并令 $k\to\infty$，得到

$$\boldsymbol{d}^T\nabla f(\boldsymbol{x}^*)\geqslant0, \tag{8.2.26}$$

故 $\boldsymbol{d}\notin\mathscr{D}^*$.

定理 8.2.4 （约束问题局部解的一阶必要条件）设约束问题(8.2.11)中 $f(\boldsymbol{x}),c_i(\boldsymbol{x})(i=1,2,\cdots,l+m)$ 具有连续的一阶偏导数，若 $\boldsymbol{x}^*$ 是约束问题(8.2.11)的局部解，并且在 $\boldsymbol{x}^*$ 处约束限制条件成立（即 $F^*=\mathscr{F}^*$），则存在常数 $\lambda^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_{l+m}^*)^T$，使得

$$\begin{aligned}
&\nabla_x L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)=\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{l+m}\lambda_i^*\nabla c_i(\boldsymbol{x}^*)=0,\\
&c_i(\boldsymbol{x}^*)=0,\quad i\in E=\{1,2,\cdots,l\},\\
&c_i(\boldsymbol{x}^*)\leqslant0,\quad i\in I=\{l+1,l+2,\cdots l+m\},\\
&\lambda_i^*\geqslant0,\quad i\in I=\{l+1,l+2,\cdots l+m\},\\
&\lambda_i^*c_i(\boldsymbol{x}^*)=0,\quad i\in I=\{l+1,l+2,\cdots l+m\},
\end{aligned} \tag{8.2.27}$$

其中 $L(\boldsymbol{x},\boldsymbol{\lambda})$ 为 Lagrange 函数

$$L(\boldsymbol{x},\boldsymbol{\lambda})=f(\boldsymbol{x})+\sum_{i=1}^{l+m}\lambda_ic_i(\boldsymbol{x}). \tag{8.2.28}$$

证明 因为 $\boldsymbol{x}^*$ 是约束问题(8.2.11)的局部解，由定理 8.2.3 得到 $\mathscr{F}^*\cap\mathscr{D}^*=\phi$，再由约束限制条件得到 $F^*\cap\mathscr{D}^*=\phi$，令 $\boldsymbol{a}_i=$

$\nabla c_i(\boldsymbol{x}^*), i \in E \cup I(x^*), \boldsymbol{w} = -\nabla f(\boldsymbol{x}^*)$,因此系统

$$
\begin{aligned}
&\boldsymbol{a}_i^T \boldsymbol{d} = 0, \qquad i \in E, \\
&\boldsymbol{a}_i^T \boldsymbol{d} \leqslant 0, \qquad i \in I(x^*), \\
&\boldsymbol{w}^T \boldsymbol{d} > 0
\end{aligned}
$$

无解,由定理 8.2.2 的推论 2,存在 $\lambda_i^*(i \in E \cup I(x^*))$,使得

$$
\begin{aligned}
&\boldsymbol{w} = \sum_{i \in E \cup I(\boldsymbol{x}^*)} \lambda_i^* \boldsymbol{a}_i, \\
&\lambda_i^* \geqslant 0, \quad i \in I(\boldsymbol{x}^*),
\end{aligned}
\tag{8.2.29}
$$

令

$$\lambda_i^* = 0, i \in I \setminus I(\boldsymbol{x}^*).$$

故命题成立.

由于这一定理是 Kuhn 和 Tucker(1951)给出的,因此称上述一阶必要条件为 **Kuhn-Tucker 条件**(Kuhn-Tucker Conditions),或简称为 K-T 条件.称满足一阶必要条件的点为 Kuhn-Tucker 点,或简称 K-T 点.由于函数(8.2.28)的思想可追溯到 Lagrange (1760-1761),因此称为 Lagrange 函数,称 $\boldsymbol{\lambda}^*$ 为 $\boldsymbol{x}^*$ 处的 Lagrange 乘子(Lagrange Multiplies).

后人发现 Karush 早在 1939 年就发现了类似地最优性条件,所以一阶必要条件也称为 Karush-Kuhn-Tucker 条件,或 K-K-T 条件,称满足一阶必要条件的点为 K-K-T 点.为与以前出版的教科书保持一致,本书仍沿用从前的叫法,即 K-T 条件和 K-T 点.

请注意,在定理 8.2.4 中增加了约束限制条件,若无此条件,则局部解不一定是 K-T 点.

事实上,考虑约束问题

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) = x_1, \\
\text{s.t.} \quad & c_1(\boldsymbol{x}) = x_2 - x_1^3 \leqslant 0, \\
& c_2(\boldsymbol{x}) = -x_2 \leqslant 0,
\end{aligned}
$$

因为 $x_1^3 \geqslant x_2 \geqslant 0$，所以 $\boldsymbol{x}^* = (0,0)^T$ 是约束问题的局部解. 由例 8.2.3 知 $F^* \neq \mathscr{F}^*$，约束限制条件不成立.

考虑 Lagrange 函数

$$\begin{aligned} L(\boldsymbol{x},\boldsymbol{\lambda}) &= f(\boldsymbol{x}) + \lambda_1 c_1(\boldsymbol{x}) + \lambda_2 c_2(\boldsymbol{x}) \\ &= x_1 + \lambda_1(x_2 - x_1^3) - \lambda_2 x_2, \end{aligned}$$

因此，对一切 $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T$，

$$\nabla_x L(\boldsymbol{x}^*, \boldsymbol{\lambda}) = \begin{pmatrix} 1 \\ \lambda_1 - \lambda_2 \end{pmatrix} \neq 0,$$

所以 $\boldsymbol{x}^*$ 不是 K-T 点.

8.2.3 约束限制条件

前面提到，凡能保证 $F^* = \mathscr{F}^*$ 成立的任何一组条件均称为约束限制条件，若约束问题(8.2.11)的约束在局部解 $\boldsymbol{x}^*$ 处满足某种约束限制条件，则在 $\boldsymbol{x}^*$ 处 K-T 条件成立. 约束限制条件有许多种，这里只介绍最简单、最常用的两种.

先介绍一个引理.

引理 8.2.2 设 $c_i(\boldsymbol{x})(i=1,2,\cdots,l)$ 具有连续的一阶偏导数，考虑集合

$$S = \{\boldsymbol{x} \mid c_i(\boldsymbol{x}) = 0, i = 1,2,\cdots,l\}, \tag{8.2.30}$$

若 $\boldsymbol{x}^* \in S$，$\nabla c_i(\boldsymbol{x}^*)(i=1,2,\cdots,l)$ 线性无关，对于任意的 $\boldsymbol{d} \in M$，其中

$$M = \{\boldsymbol{d} \mid \boldsymbol{d}^T \nabla c_i(\boldsymbol{x}^*) = 0, i = 1,2,\cdots,l\}, \tag{8.2.31}$$

则存在曲线 $\boldsymbol{x} = \boldsymbol{x}(t)$，$t^* \leqslant t \leqslant \beta$，满足

$$\begin{aligned} & c_i(\boldsymbol{x}(t)) \equiv 0, \qquad i = 1,2,\cdots,l, \\ & \boldsymbol{x}(t^*) = x^*, \\ & \frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}\Big|_{t=t^*} = \boldsymbol{d}. \end{aligned} \tag{8.2.32}$$

证明 由篇幅的限制，略去定理的证明.

定理 8.2.5 若在约束问题(8.2.11)的局部解 $\boldsymbol{x}^*$ 处下述两条件之一成立:

(1)$c_i(\boldsymbol{x})(i\in E\cup I(\boldsymbol{x}^*))$是线性函数;

(2)$\nabla c_i(\boldsymbol{x}^*)(i\in E\cup I(\boldsymbol{x}^*))$线性无关;

则在 $\boldsymbol{x}^*$ 处有

$$F^* = \mathscr{F}^*.$$

证明 情况(1)显然,只须证明情况(2).

设 $\boldsymbol{d}\in F^*$,则 $\boldsymbol{d}$ 满足$\nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d}=0, i\in E, \nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d}\leqslant 0, i\in I(\boldsymbol{x}^*)$. 考虑集合

$$I_0(\boldsymbol{x}^*) = \{i \mid \nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d} = 0, i \in I(\boldsymbol{x}^*)\}, \tag{8.2.33}$$

因此,当 $i\in E\cup I_0(\boldsymbol{x}^*)$时,$\nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d}=0$. 构造集合

$$S^- = \{\boldsymbol{x} \mid c_i(\boldsymbol{x}) = 0, i \in E \cup I_0(\boldsymbol{x}^*)\}, \tag{8.2.34}$$

$$M^- = \{\boldsymbol{d} \mid \nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d} = 0, i \in E \cup I_0(\boldsymbol{x}^*)\}, \tag{8.2.35}$$

由定理条件和引理 8.2.2,则存在曲线 $\boldsymbol{x}=\boldsymbol{x}(t), t^*\leqslant t\leqslant\beta$,满足 $c_i(\boldsymbol{x}(t))\equiv 0, i\in E\cup I_0(\boldsymbol{x}^*)$和 $\boldsymbol{x}(t^*)=\boldsymbol{x}^*, \left.\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}\right|_{t=t^*}=\boldsymbol{d}$.

下面证明 $\boldsymbol{x}(t)$是可行点.

当 $i\in I\setminus I(\boldsymbol{x}^*)$,有 $c_i(\boldsymbol{x}^*)<0$,即 $c_i(\boldsymbol{x}(t^*))<0$,由连续性可知,存在 β',当 $t^*\leqslant t\leqslant\beta'$ 时,有 $c_i(\boldsymbol{x}(t))\leqslant 0$. 当 $i\in I(\boldsymbol{x}^*)\setminus I_0(\boldsymbol{x}^*)$时,由于$\nabla c_i(\boldsymbol{x}^*)^T d<0$,存在 β'',当 $t^*\leqslant t\leqslant\beta''$ 时,有

$$\begin{aligned} c_i(\boldsymbol{x}(t)) &= c_i(\boldsymbol{x}(t^*)) + \nabla_x c_i(\boldsymbol{x}(t^*)) \left.\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}\right|_{t=t^*}(t-t^*) \\ &\quad + o(|t-t^*|) \\ &= c_i(\boldsymbol{x}^*) + \nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d}(t-t^*) + o(|t-t^*|) \end{aligned}$$

$$\leqslant c_i(\boldsymbol{x}^*) = 0. \tag{8.2.36}$$

取 $\alpha = \min\{\beta, \beta', \beta''\}$，则当 $t^* \leqslant t \leqslant \alpha$ 时，有 $c_i(\boldsymbol{x}(t)) = 0, i \in E$，$c_i(\boldsymbol{x}(t)) \leqslant 0, i \in I$，即 $x(t)$是可行点.

令 $\boldsymbol{x}^{(k)} = \boldsymbol{x}(t_k)$，其中 $t_k \to t^*$，$\boldsymbol{d}^{(k)} = \dfrac{\boldsymbol{x}^{(k)} - \boldsymbol{x}^*}{t_k - t^*}$，$\delta_k = t_k - t^*$，所以 $\boldsymbol{x}^{(k)} = \boldsymbol{x}^* + \delta_k \boldsymbol{d}^{(k)} \in D$，当 $k \to +\infty$时，有

$$\boldsymbol{d}^{(k)} = \frac{\boldsymbol{x}^{(k)} - \boldsymbol{x}^*}{t_k - t^*} \to \frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}\Big|_{t=t^*} = \boldsymbol{d} \text{ 和 } \delta_k = t_k - t^* \to 0. \tag{8.2.37}$$

所以 $\boldsymbol{d} \in \mathscr{F}^*$.

【例 8.2.4】 求约束问题

$$\begin{aligned}
\min \quad & f(\boldsymbol{x}) = x_1^2 + x_2, \\
\text{s.t.} \quad & c_1(\boldsymbol{x}) = x_1^2 + x_2^2 - 9 \leqslant 0, \\
& c_2(\boldsymbol{x}) = x_1 + x_2 - 1 \leqslant 0
\end{aligned}$$

的 K-T 点.

解 由 Kuhn-Tucker 条件得到

$$2x_1 + 2\lambda_1 x_1 + \lambda_2 = 0, \quad ①$$

$$1 + 2\lambda_1 x_2 + \lambda_2 = 0, \quad ②$$

$$x_1^2 + x_2^2 - 9 \leqslant 0, \quad ③$$

$$x_1 + x_2 - 1 \leqslant 0, \quad ④$$

$$\lambda_1 \geqslant 0, \quad ⑤$$

$$\lambda_2 \geqslant 0, \quad ⑥$$

$$\lambda_1(x_1^2 + x_2^2 - 9) = 0, \quad ⑦$$

$$\lambda_2(x_1 + x_2 - 1) = 0, \quad ⑧$$

分几种情况进行讨论：

(1) $\lambda_1 = 0, \lambda_2 = 0$，与②矛盾.

(2) $\lambda = 0, \lambda_2 \neq 0$，由①②得到 $\lambda_2 = -1$，与⑥矛盾.

(3)$\lambda_1 \neq 0, \lambda_2 = 0$，由①②⑦得到

$$\begin{cases} (1+\lambda_1)x_1 = 0, \\ 1 + 2\lambda_1 x_2 = 0, \\ x_1^2 + x_2^2 = 9. \end{cases}$$

解方程组得到

$$x_1 = 0, \quad x_2 = -3, \quad \lambda_1 = \frac{1}{6},$$

因此，K-T 点为 $\boldsymbol{x}^* = (0, -3)^T$，相应的乘子为 $\boldsymbol{\lambda}^* = \left(\frac{1}{6}, 0\right)^T$.

(4)$\lambda_1 \neq 0, \lambda_2 \neq 0$，由⑦⑧得到

$$\begin{cases} x_1^2 + x_2^2 = 9, \\ x_1 + x_2 = 1, \end{cases}$$

解方程组得到

$$x_1 = \frac{1 \pm \sqrt{17}}{2}, \quad x_2 = \frac{1 \mp \sqrt{17}}{2},$$

将$(x_1, x_2)^T$ 代入①②得到 $\lambda_1 < 0$，因此该点不是 K-T 点.

综上所述，约束问题有唯一的 K-T 点 $\boldsymbol{x}^* = (0, -3)^T$，相应的乘子为 $\boldsymbol{\lambda}^* = \left(\frac{1}{6}, 0\right)^T$. 由几何直观可以看出，$\boldsymbol{x}^*$ 是约束问题的最优解.

与无约束最优化问题一样，满足一阶必要条件的点并不一定是极小点. 考虑例子

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = x_2, \\ \text{s.t.} \quad & c(\boldsymbol{x}) = -x_1^2 - x_2 \leqslant 0, \end{aligned}$$

K-T 点应满足方程组

$$-2\lambda x_1 = 0, \qquad ①$$

$$1 - \lambda = 0, \qquad ②$$

$$-x_1^2 - x_2 \leqslant 0, \qquad ③$$

$$\lambda \geqslant 0, \qquad ④$$

$$\lambda(x_1^2 + x_2) = 0, \qquad ⑤$$

由②得到 $\lambda=1$，由⑤得到 $x_1^2+x_2=0$，再由①得到 $x_1=0$，因此 K-T 点为 $\boldsymbol{x}^*=(0,0)^T$，相应的乘子为 $\lambda^*=1$. 但 $\boldsymbol{x}^*=(0,0)^T$ 不是约束问题的局部解.

下面我们讨论一般约束问题(8.2.11)的二阶充分条件.

§8.3 约束问题局部解的充分条件

在上节对约束问题局部解的一阶必要条件给予了严格的证明，本节讨论约束问题局部解的二阶充分条件，类似于一阶必要条件的分析过程，我们先从几何直观给出约束问题的局部解的二阶充分条件，然后再给出严格的数学证明.

8.3.1 简单约束问题的二阶充分条件

为了从几何直观上给出约束问题局部解的二阶充分条件，先考虑简单的约束问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x})=f(x_1,x_2), \\ \text{s.t.} \quad & c(\boldsymbol{x})=c(x_1,x_2)=0. \end{aligned} \qquad (8.3.1)$$

为便于讨论，考虑无约束问题

$$\min \quad f(\boldsymbol{x}) = f(x_1,x_2), \qquad (8.3.2)$$

其一阶必要条件是

$$\nabla f(\boldsymbol{x}^*) = 0, \qquad (8.3.3)$$

二阶充分条件是：

①$\nabla f(\boldsymbol{x}^*)=0$；

②目标函数 $f(x)$ 在 $\boldsymbol{x}^*$ 处 Hesse 矩阵 $\nabla^2 f(\boldsymbol{x}^*)$ 正定.

而对于约束问题(8.3.1)，其一阶必要条件是

$$\nabla_x L(x^*,\lambda^*) = \nabla f(x^*) + \lambda^* \nabla c(x^*) = 0,$$
$$c(x^*) = 0, \tag{8.3.4}$$
其中
$$L(x,\lambda) = f(x) + \lambda c(x)$$
是 Lagrange 函数.

比较(8.3.4)式和(8.3.3)式发现,对于约束问题(8.3.1),Lagrange 函数 $L(x,\lambda)$起着无约束问题(8.3.2)中的目标函数 $f(x)$的作用,即一阶必要条件在 x^* 处关于 x 的梯度等于 0. 再进一步考察无约束问题(8.3.2)的二阶充分条件,可以猜想,对于约束问题(8.3.1)的二阶充分条件是:①$\nabla_x L(x^*,\lambda^*)=0$,②Lagrange 函数 $L(x,\lambda)$在(x^*,λ^*)处的 Hesse 矩阵$\nabla_x^2 L(x^*,\lambda^*)$正定.

这个结论是正确的,但这个结论要求的条件太强了.因为是考虑约束问题,因此不必要求 Hesse 矩阵$\nabla_x^2 L(x^*,\lambda^*)$正定,也就是说,不必要求任何方向上的二阶方向导数大于 0,而只在某些方向上二阶方向导数大于 0 即可.请看下面的情况.

画出曲面 $y=L(x,\lambda^*)$,它是一个马鞍面(见图 8.3.1),其中 x^* 是马鞍面的鞍点,并假设约束 $c(x)=0$ 是一直线,马鞍面在直线方向上二阶方向导数大于 0,而在与其垂直的方向上二阶方向导数小于 0,并注意到,在可行域内 $L(x,\lambda^*)=f(x)+\lambda^* c(x)=f(x)$,因此,$x^*$ 仍是约束问题的局部解.

当 $c(x)=0$ 为曲线时,只要在 x^* 处的切线方向 d 满足二阶方向导数大于 0,则仍能保证 x^* 是约束问题(8.3.1)的局部解.因此对于约束问题(8.3.1),局部解的二阶充分条件应为:

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

(2)对一切 $d\in M$,其中
$$M = \{d \mid d \neq 0, \nabla c(x^*)^T d = 0\}, \tag{8.3.5}$$
有
$$d^T \nabla_x^2 L(x^*,\lambda^*) d > 0. \tag{8.3.6}$$

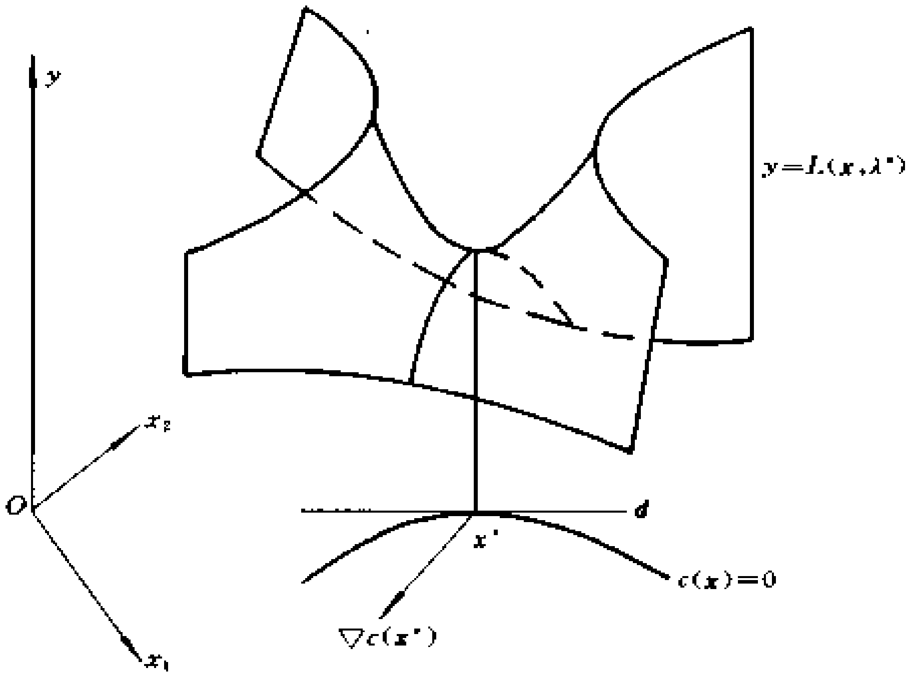


图 8.3.1　简单约束问题局部解的二阶充分条件的几何意义

下面对约束问题局部解的二阶充分条件给出严格的证明.

8.3.2　约束问题局部解的二阶充分条件

定理 8.3.1　(约束问题局部解的二阶充分条件)考虑一般约束问题(8.2.11),设 $f(\boldsymbol{x})$、$c_i(\boldsymbol{x})$($i\in E\cup I$)具有连续的二阶偏导数,若存在 $\boldsymbol{x}^*$ 满足下列条件:

(1)K-T 条件成立.即存在 $\boldsymbol{\lambda}^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_{l+m}^*)^T$ 使得

$$
\begin{aligned}
&\nabla_x L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)=\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{l+m}\lambda_i^*\nabla c_i(\boldsymbol{x}^*)=0,\\
&c_i(\boldsymbol{x}^*)=0,\quad i\in E=\{1,2,\cdots,l\},\\
&c_i(\boldsymbol{x}^*)\leqslant 0,\quad i\in I=\{l+1,l+2,\cdots l+m\},\\
&\lambda_i^*\geqslant 0,\quad i\in I=\{l+1,l+2,\cdots l+m\},\\
&\lambda_i^*c_i(\boldsymbol{x}^*)=0,\quad i\in I=\{l+1,l+2,\cdots l+m\},
\end{aligned}
\tag{8.3.7}
$$

且 λ_i^* 和 $c_i(\boldsymbol{x}^*)(i\in I)$ 不同时为 0(称为严格松弛互补条件).

(2)对于任意的 $\boldsymbol{d}\in M$,有

$$\boldsymbol{d}^T\nabla_x^2L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\boldsymbol{d}>0, \tag{8.3.8}$$

其中

$$M=\{\boldsymbol{d}\mid \boldsymbol{d}\neq 0,\nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d}=0, i\in E\cup I(\boldsymbol{x}^*)\}, \tag{8.3.9}$$

$I(\boldsymbol{x}^*)$ 是 $\boldsymbol{x}^*$ 处的有效约束指标集,则 $\boldsymbol{x}^*$ 是约束问题(8.2.11)的严格局部解.

证明 反证法.若 $\boldsymbol{x}^*$ 不是约束问题的严格局部解,则存在可行点列 $\{\boldsymbol{x}^{(k)}\}$, $\boldsymbol{x}^{(k)}\to\boldsymbol{x}^*$,使得

$$f(\boldsymbol{x}^{(k)})\leqslant f(\boldsymbol{x}^*). \tag{8.3.10}$$

令

$$\boldsymbol{x}^{(k)}=\boldsymbol{x}^*+\delta_k\boldsymbol{d}^{(k)}, \tag{8.3.11}$$

其中

$$\delta_k=\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\|,\quad \boldsymbol{d}^{(k)}=\frac{\boldsymbol{x}^{(k)}-\boldsymbol{x}^*}{\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^*\|}, \tag{8.3.12}$$

因此, $\|\boldsymbol{d}^{(k)}\|=1$ 和 $\delta_k\to 0$.

因为 $\boldsymbol{d}^{(k)}$ 有界,必有收敛子列,不妨仍记为 $\boldsymbol{d}^{(k)}$,即 $\boldsymbol{d}^{(k)}\to\boldsymbol{d}$.由 Taylor 展开式和(8.3.10)式得到

$$0\geqslant f(\boldsymbol{x}^{(k)})-f(\boldsymbol{x}^*)=\delta_k\nabla f(\boldsymbol{x}^*)^T\boldsymbol{d}^{(k)}+o(\delta_k), \tag{8.3.13}$$

在(8.3.13)式两端同除 δ_k,并令 $k\to\infty$,得到

$$\nabla f(\boldsymbol{x}^*)^T\boldsymbol{d}\leqslant 0. \tag{8.3.14}$$

用类似的方法可以得到

$$\nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d}=0,\quad i\in E, \tag{8.3.15}$$

$$\nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d}\leqslant 0,\quad i\in I(\boldsymbol{x}^*). \tag{8.3.16}$$

(8.3.16)表明,①对一切 $i\in I(\boldsymbol{x}^*)$, $\nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d}=0$,或者,②存在 $q\in I(\boldsymbol{x}^*)$,使得 $\nabla c_q(\boldsymbol{x}^*)^T\boldsymbol{d}<0$.下面证明①、②两种情

况均不会出现.

若①成立,则 $d\in M$. 考虑 Lagrange 函数

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

在 x^* 处的 Taylor 展开式

$$L(\boldsymbol{x}^{(k)},\boldsymbol{\lambda}^*) = L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*) + \delta_k\nabla_x L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)^T\boldsymbol{d}^{(k)} + \frac{1}{2}\delta_k^2\boldsymbol{d}^{(k)T}\nabla_x^2 L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\boldsymbol{d}^{(k)} + o(\delta_k^2). \tag{8.3.17}$$

注意到

$$\begin{aligned} L(\boldsymbol{x}^{(k)},\boldsymbol{\lambda}^*) &= f(\boldsymbol{x}^{(k)}) + \sum_{i=1}^{l+m}\lambda_i^* c_i(\boldsymbol{x}^{(k)}) \\ &= f(\boldsymbol{x}^{(k)}) + \sum_{i\in I(\boldsymbol{x}^*)}\lambda_i^* c_i(\boldsymbol{x}^{(k)}) \\ &\leqslant f(\boldsymbol{x}^{(k)}) \end{aligned} \tag{8.3.18}$$

和

$$L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*) = f(\boldsymbol{x}^*) + \sum_{i=1}^{l+m}\lambda_i^* c_i(\boldsymbol{x}^*) = f(\boldsymbol{x}^*), \tag{8.3.19}$$

因此,(8.3.17)式可写成

$$\begin{aligned} 0 &\geqslant f(\boldsymbol{x}^{(k)}) - f(\boldsymbol{x}^*) \geqslant L(\boldsymbol{x}^{(k)},\boldsymbol{\lambda}^*) - L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*) \\ &= \frac{1}{2}\delta_k^2\boldsymbol{d}^{(k)T}\nabla_x^2 L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\boldsymbol{d}^{(k)} + o(\delta_k^2), \end{aligned} \tag{8.3.20}$$

在(8.3.20)式两端同除 δ_k^2,并令 $k\to\infty$,得到

$$\boldsymbol{d}^T\nabla_x^2 L(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)\boldsymbol{d} \leqslant 0, \tag{8.3.21}$$

与(8.3.8)式矛盾.

再假设(2)成立. 由一阶必要条件和(8.3.15)式得到

$$\nabla f(\boldsymbol{x}^*)^T\boldsymbol{d} = -\sum_{i=1}^{l+m}\lambda_i^* c_i(\boldsymbol{x}^*)^T\boldsymbol{d}$$

$$= - \sum_{i \in I(x^*)} \lambda_i^* \nabla c_i(\boldsymbol{x}^*)^T \boldsymbol{d}$$

$$\geqslant - \lambda_q^* \nabla c_K(\boldsymbol{x}^*)^T \boldsymbol{d} > 0, \qquad (8.3.22)$$

与(8.3.14)式矛盾.

【例 8.3.1】 用约束问题局部解的一阶必要条件和二阶充分条件求解约束问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = x_1^2 - x_2^2 - 4x_2, \\ \text{s.t.} \quad & c(\boldsymbol{x}) = x_2 = 0. \end{aligned}$$

解 由约束问题的一阶必要条件得到

$$\begin{cases} 2x_1 = 0, \\ -2x_2 - 4 + \lambda = 0, \\ x_2 = 0, \end{cases}$$

求解得到 K-T 点 $\boldsymbol{x}^* = (0,0)^T$ 和相应的乘子 $\lambda^* = 4$. 下面验证二阶充分条件. Lagrange 函数在$(\boldsymbol{x}^*, \lambda^*)$处的 Hesse 矩阵为

$$\nabla_x^2 L(\boldsymbol{x}^*, \lambda^*) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix},$$

不是正定矩阵. 考虑集合

$$\begin{aligned} M &= \{\boldsymbol{d} \mid \boldsymbol{d} \neq 0, \nabla c(\boldsymbol{x}^*)^T \boldsymbol{d} = 0\} \\ &= \{(\alpha, \beta)^T \mid (\alpha, \beta)^T \neq 0, \beta \neq 0\} \\ &= \{(\alpha, 0)^T \mid \alpha \neq 0\}, \end{aligned}$$

对于 $\boldsymbol{d} \in M$,有

$$\boldsymbol{d}^T \nabla_x^2 L(\boldsymbol{x}^* \lambda^*) \boldsymbol{d} = (\alpha, 0) \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} \alpha \\ 0 \end{pmatrix} = 2\alpha^2 > 0,$$

因此, $\boldsymbol{x}^*$ 是约束问题的局部解.

§8.4 Lagrange 乘子的意义

本节讨论与约束问题局部解 $\boldsymbol{x}^*$ 相应的 Lagrange 乘子的意

义.

考虑一般约束问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \quad x \in R^n, \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) = 0 \ i \in E = \{1, 2, \cdots, l\}, \\ & c_i(\boldsymbol{x}) \leqslant 0, i \in I = \{l+1, l+2, \cdots, l+m\}, \end{aligned} \tag{8.4.1}$$

对问题(8.4.1)约束的右端项进行扰动,得到扰动问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \quad x \in R^n, \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) = \varepsilon_i \ i \in E = \{1, 2, \cdots, l\}, \\ & c_i(\boldsymbol{x}) \leqslant \varepsilon_i, i \in I = \{l+1, l+2, \cdots, l+m\}. \end{aligned} \tag{8.4.2}$$

令 $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{l+m})^T$,记扰动问题(8.4.2)的最优解为 $\boldsymbol{x}^*(\boldsymbol{\varepsilon})$,相应的 Lagrange 乘子为 $\boldsymbol{\lambda}^*(\boldsymbol{\varepsilon})$,特别地,当 $\boldsymbol{\varepsilon} = 0$ 时,有 $\boldsymbol{x}^*(0) = \boldsymbol{x}^*$, $\boldsymbol{\lambda}^*(0) = \boldsymbol{\lambda}^*$. 我们的目标是计算在 $\boldsymbol{\varepsilon} = 0$ 处 $\nabla_{\boldsymbol{\varepsilon}} f(\boldsymbol{x}^*(\boldsymbol{\varepsilon}))$ 的值.

这里并不给出严格的证明,而是通过对简单情况的讨论(一个等式约束或一个不等式约束),来分析其计算过程,得到相应的结论.

首先讨论等式约束情况

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \\ \text{s.t.} \quad & c(\boldsymbol{x}) = 0. \end{aligned} \tag{8.4.3}$$

设最优解为 $\boldsymbol{x}^*$,相应的乘子为 λ^*. 其扰动问题为

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \\ \text{s.t.} \quad & c(\boldsymbol{x}) = \varepsilon, \end{aligned} \tag{8.4.4}$$

其最优解为 $\boldsymbol{x}^*(\varepsilon)$,相应的乘子为 $\lambda^*(\varepsilon)$,并且假设

$$\boldsymbol{x}^*(0) = \boldsymbol{x}^*, \quad \lambda^*(0) = \lambda^*,$$

计算

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} f(\boldsymbol{x}^*(\varepsilon)) \mid_{\varepsilon=0} = \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^*(\varepsilon))^T \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \boldsymbol{x}^*(\varepsilon) \mid_{\varepsilon=0}$$

$$= \nabla f(\boldsymbol{x}^*)^T \left[\frac{\mathrm{d}}{\mathrm{d}\epsilon} \boldsymbol{x}^*(\epsilon) \right]_{\epsilon=0}. \tag{8.4.5}$$

由扰动问题的约束条件，得到

$$c(\boldsymbol{x}^*(\varepsilon)) \equiv \varepsilon. \tag{8.4.6}$$

在(8.4.6)式两端对 ε 求导数，得到

$$1 = \frac{\mathrm{d}}{\mathrm{d}\epsilon} c(\boldsymbol{x}^*(\epsilon)) |_{\epsilon=0} = \nabla_x c(\boldsymbol{x}^*(\epsilon))^T \frac{\mathrm{d}}{\mathrm{d}\epsilon} \boldsymbol{x}^*(\epsilon) |_{\epsilon=0}$$

$$= \nabla c(\boldsymbol{x}^*)^T \left[\frac{\mathrm{d}}{\mathrm{d}\epsilon} \boldsymbol{x}^*(\epsilon) \right]_{\epsilon=0}. \tag{8.4.7}$$

由约束问题的一阶必要条件

$$\nabla f(\boldsymbol{x}^*) + \lambda^* \nabla c(\boldsymbol{x}^*) = 0,$$

并与(8.4.7)式相比较，(8.4.5)式可写成

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} f(\boldsymbol{x}^*(\epsilon)) |_{\epsilon=0} = \nabla_x f(\boldsymbol{x})^T \left[\frac{\mathrm{d}}{\mathrm{d}\epsilon} \boldsymbol{x}^*(\epsilon) \right]_{\epsilon=0}$$

$$= -\lambda^* \nabla c(\boldsymbol{x}^*)^T \left[\frac{\mathrm{d}}{\mathrm{d}\epsilon} \boldsymbol{x}^*(\epsilon) \right]_{\epsilon=0} = -\lambda^*. \tag{8.4.8}$$

再讨论不等式约束情况

$$\begin{aligned} &\min \quad f(\boldsymbol{x}), \\ &\text{s.t.} \quad c(\boldsymbol{x}) \leqslant 0. \end{aligned} \tag{8.4.9}$$

设最优解为 $\boldsymbol{x}^*$，相应的乘子为 λ^*. 其扰动问题为

$$\begin{aligned} &\min \quad f(\boldsymbol{x}), \\ &\text{s.t.} \quad c(\boldsymbol{x}) \leqslant \varepsilon, \end{aligned} \tag{8.4.10}$$

其最优解为 $\boldsymbol{x}^*(\varepsilon)$，相应的乘子为 $\lambda^*(\varepsilon)$. 并且假设

$$\boldsymbol{x}^*(0) = \boldsymbol{x}^*, \quad \lambda^*(0) = \lambda^*.$$

现分两种情况讨论.

(1) $c(\boldsymbol{x}^*) = 0$，即 $c(\boldsymbol{x}) \leqslant 0$ 是 $\boldsymbol{x}^*$ 处的有效约束. 不难设想当 $|\varepsilon|$ 很小时，有 $c(\boldsymbol{x}^*(\varepsilon)) = \varepsilon$，即 $\boldsymbol{x}^*(\varepsilon)$ 仍在约束 $c(\boldsymbol{x}) \leqslant \varepsilon$ 的边界上，也就是说，约束 $c(\boldsymbol{x}) \leqslant \varepsilon$ 仍是有效约束，因此 $\boldsymbol{x}^*(\varepsilon)$ 也是等式约束问题(8.4.4)的局部解，由前所述，有

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}f(\boldsymbol{x}^*(\varepsilon))\mid_{\varepsilon=0}=-\lambda^*.$$

(2)$c(\boldsymbol{x}^*)<0$,即 $c(\boldsymbol{x})\leqslant 0$ 是 $\boldsymbol{x}^*$ 处的非有效约束.此时 $\boldsymbol{x}^*$ 是无约束问题 $\min f(\boldsymbol{x})$的局部解,因此当$|\varepsilon|$很小时,$\boldsymbol{x}^*$ 也是扰动问题(8.4.10)的局部解.因此有

$$\boldsymbol{x}^*(\epsilon)=\boldsymbol{x}^*,\quad f(\boldsymbol{x}^*(\epsilon))=f(\boldsymbol{x}^*),$$

所以

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}f(\boldsymbol{x}^*(\epsilon))\mid_{\epsilon=0}=0=-\lambda^*.$$

两种情况可以写成统一的表达式,将简单情况推广到一般情况,得到如下定理.

定理 8.4.1 设 $f(\boldsymbol{x})$、$c_i(\boldsymbol{x})(i\in E\cup I)$具有连续的二阶偏导数,$\boldsymbol{x}^*$ 是约束问题(8.4.1)的局部解,并在 $\boldsymbol{x}^*$ 处满足约束限制条件和二阶充分条件,$\boldsymbol{\lambda}^*$ 是相应的 Lagrange 乘子.进一步假设 $\boldsymbol{x}^*(\boldsymbol{\epsilon})$是扰动问题(8.4.2)的局部解,$\lambda^*(\epsilon)$是相应的乘子,则

$$\nabla_{\boldsymbol{\varepsilon}}f(\boldsymbol{x}^*(\boldsymbol{\varepsilon}))\mid_{\boldsymbol{\varepsilon}=0}=-\boldsymbol{\lambda}^*.\tag{8.4.11}$$

证明 (略)

【例 8.4.1】 某企业预算以 2 千元作为广告费,根据以往的经验,若以 x_1 千元作广播广告,x_2 千元作报纸广告,销售金额为 $-2x_1^2-10x_2^2-8x_1x_2+18x_1+34x_2$(单位:千元),试问:

(1)如何分配 2 千元广告费?

(2)广告费预算作微小改变的影响如何?

解 第一个问题是求极值问题,相应的最优化问题为

$$\begin{aligned}\min\quad & f(\boldsymbol{x})=2x_1^2+10x_2^2+8x_1x_2-18x_1-34x_2,\\ \text{s.t.}\quad & x_1+x_2-2=0,\\ & x_1\geqslant 0,x_2\geqslant 0.\end{aligned}$$

由一阶必要条件得到

$$4x_1+8x_2-18+\lambda_1-\lambda_2=0,\tag{①}$$

$$8x_1 + 20x_2 - 34 + \lambda_1 - \lambda_3 = 0, \qquad ②$$

$$x_1 + x_2 = 2, \qquad ③$$

$$x_1, x_2 \geqslant 0, \qquad ④$$

$$\lambda_2, \lambda_3 \geqslant 0, \qquad ⑤$$

$$\lambda_2 x_1 = 0,\ \lambda_3 x_2 = 0. \qquad ⑥$$

分情况进行讨论：

(1)$\lambda_2=0,\lambda_3=0$,求解①②③得到 $x_1=1,x_2=1,\lambda_1=6$.

(2)$\lambda_2=0,\lambda_3\neq0$,求解①②③⑥得到 $x_1=2,x_2=0$ 和 $\lambda_3<0$,与⑤矛盾.

(3)$\lambda_2\neq0,\lambda_3=0$,求解①②③⑥得到 $x_1=0,x_2=2$ 和 $\lambda_2<0$,与⑤矛盾.

(2)$\lambda_2\neq0,\lambda_3\neq0$,由⑥得到 $x_1=0,x_2=0$,与③矛盾.

因此,K-T 点为 $\boldsymbol{x}^*=(1,1)^T$,相应的乘子为 $\boldsymbol{\lambda}^*=(6,0,0)^T$. 注意到目标函数是凸的,所以 x^* 是最优解.

第二个问题:广告费作微小改动.考虑扰动问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x})=2x_1^2+10x_2^2+8x_1x_2-18x_1-34x_2, \\ \text{s.t.} \quad & x_1+x_2-2=\epsilon, \\ & x_1\geqslant0,\ x_2\geqslant0, \end{aligned}$$

由定理 8.4.1,有

$$\frac{\mathrm{d}f(\boldsymbol{x}^*(\epsilon))}{\mathrm{d}\epsilon}\Big|_{\epsilon=0}=-6,$$

当 ϵ 增加时,$f(\boldsymbol{x}^*(\epsilon))$下降,即 $-f(\boldsymbol{x}^*(\epsilon))$上升,也就是说,当广告费增加后,销售金额也随着增加.并由上式可知,销售金额的增加大约是广告费增加的 6 倍,可见适当增加广告费的预算是有利的.

具体计算如下：

若广告费由 2 千元增加到 2.1 千元,增加了 0.1 千元,而最优

解由 $x^*=(1,1)^T$ 变为 $x^*(0.1)=(1.15,0.95)^T$，相应的销售金额由 $-f(x^*)=32$ 千元增加到 $-f(x^*(0.1))=32.59$ 千元，销售金额增加了 0.59 千元，接近广告费预算增加的 6 倍.

习 题 八

8.1 设集合 $S\subset R^n$ 是非空凸集，证明：S 的生成锥 $C(S)$ 是凸锥.

8.2 设 $a_1,a_2,\cdots,a_m\in R^n$，证明：

$$C=\{\sum_{i=1}^{m}\lambda_i a_i \mid \lambda_i\geqslant 0, i=1,2,\cdots,m\}$$

是凸锥.

8.3 证明定理 8.2.2 的推论 1.

8.4 证明定理 8.2.2 的推论 2.

8.5 考虑约束

$$c_1(x)=x_1^2-x_2^2-x_3=0,$$
$$c_2(x)=x_1-x_2^2+x_3^2=0.$$

问：约束的可行点是否满足约束限制条件？

8.6 对于约束问题

$$\begin{aligned}\min\quad & f(x)=5-x_1-x_2^2,\\ \text{s.t.}\quad & c_1(x)=-x_1\leqslant 0,\\ & c_2(x)=-x_2\leqslant 0,\\ & c_3(x)=x_1^2+x_2^2-1\leqslant 0,\end{aligned}$$

考察 $x^*=(1,0)^T$ 是否为约束问题的 K-T 点，是否为约束问题的极小点.

8.7 考虑约束问题

$$\begin{aligned}\min\quad & f(x)=\frac{1}{2}a(x_1-2)^2-x_1-x_2,\\ \text{s.t.}\quad & c_1(x)=x_2^2-x_1\leqslant 0,\end{aligned}$$

$$c_2(\boldsymbol{x}) = x_1 + x_2 - 2 \leqslant 0.$$

(1)求出所有的使两个约束均为有效约束的可行点；

(2)求出 α 的取值范围，使得 α 在该范围取值时，由(1)得到的可行点满足一阶必要条件.

8.8　利用等式约束问题的一阶必要条件求解约束问题

$$\begin{aligned}\min \quad & f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^p, \\ \text{s.t.} \quad & c(\boldsymbol{x}) = \sum_{i=1}^{n} x_i - a = 0,\end{aligned}$$

其中 $p>1, a>0$.

8.9　给定问题

$$\begin{aligned}\min \quad & f(\boldsymbol{x}) = \sum_{i=1}^{n} f_i(x_i), \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1, \\ & x_i \geqslant 0, i = 1,2,\cdots,n.\end{aligned}$$

其中 $f_i(i=1,2,\cdots,n)$是可微函数. 设 $\boldsymbol{x}^*$ 是该约束问题的局部解，证明：存在 μ^*，使得

$$\begin{aligned} f'_i(x_i^*) &= \mu^*, \quad x_i^* > 0, \\ f'_i(x_i^*) &\geqslant \mu^*, \quad x_i^* = 0. \end{aligned}$$

8.10　给定问题

$$\begin{aligned}\min \quad & f(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{c_i}{x_i}, \\ \text{s.t.} \quad & \sum_{i=1}^{n} a_i x_i = b, \\ & x_i \geqslant 0, i = 1,2,\cdots,n,\end{aligned}$$

其中 $a_i, c_i(i=1,2,\cdots,n), b>0$.

(1)写出 Kuhn-Tucker 条件；

(2)证明目标函数的最优值为

$$f(x^*) = \frac{1}{b}\left(\sum_{i=1}^{n} \sqrt{a_i c_i}\right)^2.$$

8.11　用 Kuhn-Tucker 条件证明不等式

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geqslant \left(\prod_{i=1}^{n} x_i\right)^{1/n},$$

其中 $x_1, x_2, \cdots, x_n \geqslant 0$.(提示:考虑下列约束问题之一

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} x_i, \\ \text{s.t.} \quad & \prod_{i=1}^{n} x_i = 1, \\ & x_i \geqslant 0, i = 1,2,\cdots,n, \end{aligned}$$

$$\begin{aligned} \max \quad & \prod_{i=1}^{n} x_i, \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = 1, \\ & x_i \geqslant 0, i = 1,2,\cdots,n.) \end{aligned}$$

8.12　考虑约束问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \quad x \in R^n \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) \leqslant 0, \quad i = 1,2,\cdots,m, \end{aligned}$$

其中 $f(\boldsymbol{x}), c_i(\boldsymbol{x})(i=1,2,\cdots,n)$是凸函数,称为凸规划(Convex Programming)问题.

(1)证明:凸规划问题的可行域

$$D = \{\boldsymbol{x} \mid c_i(\boldsymbol{x}) \leqslant 0, i = 1,2,\cdots,n\}$$

是凸集.

(2)设 $f(\boldsymbol{x})$、$c_i(\boldsymbol{x})(i=1,2,\cdots,n)$有连续的一阶偏导数,若 $\boldsymbol{x}^*$ 是 K-T 点,则 $\boldsymbol{x}^*$ 是凸规划问题的全局解.

8.13 考虑线性规划问题

$$\begin{aligned} \min \quad & \boldsymbol{c}^T\boldsymbol{x}, \\ \text{s.t.} \quad & A\boldsymbol{x} \leqslant \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant 0, \end{aligned}$$

试写出它的一阶必要条件,并证明其一阶必要条件也是充分条件.

8.14 试证:约束问题

$$\begin{aligned} \min \quad & \| A\boldsymbol{x} - \boldsymbol{b} \|^2, \\ \text{s.t.} \quad & \boldsymbol{x} \geqslant 0 \end{aligned}$$

是凸规划问题,并证明其最优解 x^* 满足不等式方程组

$$(A^TA)\boldsymbol{x} \geqslant A^T\boldsymbol{b}.$$

8.15 证明约束问题的二阶必要条件:考虑约束问题(8.0.1),设 $f(\boldsymbol{x})$、$c_i(\boldsymbol{x})(i \in E \cup I)$具有连续的二阶偏导数,若 $\boldsymbol{x}^*$ 是约束问题(8.0.1)的局部解,且在 $\boldsymbol{x}^*$ 处 $\nabla c_i(\boldsymbol{x}^*)(i \in E \cup I(\boldsymbol{x}^*))$线性无关,则存在向量 $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, \cdots, \lambda_{l+m}^*)^T$,使得:

(1)Kuhn-Tucker 条件成立;

(2)对任意的 $\boldsymbol{d} \in M = \{\boldsymbol{d} \mid \nabla c_i(\boldsymbol{x}^*)^T\boldsymbol{d} = 0, i \in E \cup I(x^*)\}$ 均有

$$\boldsymbol{d}^T \nabla_x^2 L(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{d} \geqslant 0,$$

其中 $L(\boldsymbol{x}, \boldsymbol{\lambda})$为约束问题的 Lagrange 函数.

8.16 考虑约束问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = x_1^2 + 2x_1 + x_2^4, \\ \text{s.t.} \quad & c(\boldsymbol{x}) = x_1x_2 - x_1 = 0. \end{aligned}$$

(1)验证 $\boldsymbol{x}^* = (0,0)^T$ 满足局部解的一阶必要条件.

(2)试问 $\boldsymbol{x}^* = (0,0)^T$ 是否满足局部解的二阶充分条件?

(3)试问 $\boldsymbol{x}^* = (0,0)^T$ 是否为约束问题的局部解(或全局解).

8.17 用约束问题局部解的一阶必要条件和二阶充分条件求解约束问题

$$\min \quad f(\boldsymbol{x}) = x_1 x_2,$$
$$\text{s.t.} \quad c(\boldsymbol{x}) = x_1^2 + x_2^2 - 1 = 0$$

的最优解和相应的乘子,并用图解法验证你的结论.

8.18 考虑约束问题

$$\min \quad f(\boldsymbol{x}) = x_1 x_2,$$
$$\text{s.t.} \quad c(\boldsymbol{x}) = x_1^2 + x_2^2 - R^2 = 0.$$

利用 8.17 题的结果,问如何调整 R($R<1$ 还是 $R>1$),使该问题的最优目标函数值小于 $f(\boldsymbol{x}^*)$(8.17 题的结果).

第九章　二次规划问题

在讨论一般约束问题的求解之前，先讨论一类最简单的约束问题——凸二次规划问题的求解方法．这不仅因为二次规划本身的重要性，而且它在一般约束最优化问题的求解中，起着十分重要的作用．

§9.1　二次规划的基本概念和基本性质

考虑约束问题

$$
\begin{aligned}
\min\quad & f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^TG\boldsymbol{x}+\boldsymbol{r}^T\boldsymbol{x},\quad \boldsymbol{x}\in R^n,\\
\text{s.t.}\quad & c_i(\boldsymbol{x})=\boldsymbol{a}_i^T\boldsymbol{x}-b_i=0,\quad i\in E=\{1,2,\cdots,l\},\\
& c_i(\boldsymbol{x})=\boldsymbol{a}_i^T\boldsymbol{x}-b_i\leqslant 0,\quad i\in I=\{l+1,l+2,\cdots,l+m\},
\end{aligned}
\tag{9.1.1}
$$

其中 G 为 $n\times n$ 阶对称矩阵，$\boldsymbol{r}$、$\boldsymbol{a}_i(i\in E\cup I)$ 为 n 维向量，$b_i(i\in E\cup I)$ 为纯量，称问题(9.1.1)为**二次规划**(Quadratic Programming)问题．若 G 为(正定)半正定矩阵，则称问题(9.1.1)为(严格)**凸二次规划**(Convex Quadratic Programming)．

定理 9.1.1　$\boldsymbol{x}^*$ 是(严格)凸二次规划的全局解的充分必要条件是：$\boldsymbol{x}^*$ 是 K-T 点，即存在 $\boldsymbol{\lambda}^*=(\lambda_1^*,\lambda_2^*,\cdots,\lambda_{l+m}^*)^T$，使得

$$
\begin{aligned}
& G\boldsymbol{x}^*+\boldsymbol{r}+\sum_{i=1}^{l+m}\lambda_i^*\boldsymbol{a}_i=0,\\
& \boldsymbol{a}_i^T\boldsymbol{x}^*-b_i=0,\quad i\in E,\\
& \boldsymbol{a}_i^T\boldsymbol{x}^*-b_i\leqslant 0,\quad i\in I,\\
& \lambda_i^*\geqslant 0,\quad i\in I,
\end{aligned}
\tag{9.1.2}
$$

$$\lambda_i^*(\boldsymbol{a}_i^T\boldsymbol{x}^* - b_i) = 0, i \in I.$$

证明 “必要性”,由约束问题的一阶必要条件,得证.

“充分性”,设 $\boldsymbol{x}^*$ 是 K-T 点,考虑 $\forall \boldsymbol{x} \neq \boldsymbol{x}^*, \boldsymbol{x} \in D$,这里

$$D = \{\boldsymbol{x} \mid \boldsymbol{a}_i^T\boldsymbol{x} = b_i, i \in E, \boldsymbol{a}_i^T\boldsymbol{x} \leqslant b_i, i \in I\}, \tag{9.1.3}$$

由于 G 半正定(正定),因此有

$$\begin{aligned} f(\boldsymbol{x}) - f(\boldsymbol{x}^*) &= \nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T G(\boldsymbol{x} - \boldsymbol{x}^*) \\ &\overset{(>)}{\geqslant} \nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \\ &= -\sum_{i \in E}\lambda_i^*\boldsymbol{a}_i^T(\boldsymbol{x} - \boldsymbol{x}^*) - \sum_{i \in I(\boldsymbol{x}^*)}\lambda_i^*\boldsymbol{a}_i^T(\boldsymbol{x} - \boldsymbol{x}^*) - \\ &\quad \sum_{i \in I \setminus I(\boldsymbol{x}^*)}\lambda_i^*\boldsymbol{a}_i^T(\boldsymbol{x} - \boldsymbol{x}^*) \\ &\geqslant 0, \end{aligned} \tag{9.1.4}$$

因此,$\boldsymbol{x}^*$ 是(严格)全局解.

推论 (严格)凸二次规划问题的局部解均是全局解.

定理 9.1.2 若 $\boldsymbol{x}^*$ 是凸二次规划(9.1.1)的全局解,则 $\boldsymbol{x}^*$ 是如下等式约束二次规划问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{r}^T\boldsymbol{x}, \boldsymbol{x} \in R^n, \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) = \boldsymbol{a}_i^T\boldsymbol{x} - b_i = 0, i \in E \cup I(\boldsymbol{x}^*) \end{aligned} \tag{9.1.5}$$

的全局解.

证明 若 $\boldsymbol{x}^*$ 是问题(9.1.1)的全局解,则 $\boldsymbol{x}^*$ 是问题(9.1.1)的 K-T 点,也是问题(9.1.5)的 K-T 点,由定理 9.1.1 及推论,则 $\boldsymbol{x}^*$ 是问题(9.1.5)的全局解.

§9.2 等式约束二次规划问题

本节讨论只有等式约束的二次规划问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{r}^T\boldsymbol{x}, \\ \text{s.t.} \quad & A^T\boldsymbol{x} = \boldsymbol{b}, \end{aligned} \tag{9.2.1}$$

其中 $A = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m)$, $m < n$ 且 $\mathrm{rank}(A) = m$,即矩阵是列满秩的.

定理 9.2.1　设 G 是半正定(正定)矩阵,则 $\boldsymbol{x}^*$ 是约束问题(9.2.1)的全局解,$\boldsymbol{\lambda}^*$ 为相应的乘子的充分必要条件是:$\boldsymbol{x}^*$、$\boldsymbol{\lambda}^*$ 是线性方程组

$$\begin{pmatrix} G & A \\ A^T & 0 \end{pmatrix}\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{r} \\ \boldsymbol{b} \end{pmatrix} \tag{9.2.2}$$

的解.

证明　考虑问题(9.2.1)的 Lagrange 函数

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{r}^T\boldsymbol{x} + \boldsymbol{\lambda}^T(A^T\boldsymbol{x} - \boldsymbol{b}), \tag{9.2.3}$$

则问题(9.2.1)的 Kuhn-Tuchker 条件等价于线性方程组(9.2.2).由定理 9.1.1,命题成立.

定理 9.2.1 表明,求解等式约束二次规划问题,可转化为求解线性方程组的问题,我们先看一个例子.

【例 9.2.1】　求解凸二次规划问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = x_1^2 + x_2^2 + x_3^2, \\ \text{s.t.} \quad & x_1 + 2x_2 - x_3 = 4, \\ & x_1 - x_2 + x_3 = -2. \end{aligned}$$

解　因为

$$G=\begin{pmatrix}2&0&0\\0&2&0\\0&0&2\end{pmatrix},\quad \boldsymbol{r}=\begin{pmatrix}0\\0\\0\end{pmatrix},$$

$$A^T=\begin{pmatrix}1&2&-1\\1&-1&1\end{pmatrix},\quad \boldsymbol{b}=\begin{pmatrix}4\\-2\end{pmatrix},$$

因此,线性方程组为

$$\begin{pmatrix}2&0&0&1&1\\0&2&0&2&-1\\0&0&2&-1&1\\1&2&-1&0&0\\1&-1&1&0&0\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\\\lambda_1\\\lambda_2\end{pmatrix}=\begin{pmatrix}0\\0\\0\\4\\-2\end{pmatrix},$$

求解方程组得到

$$x_1^*=\frac{2}{7},\ x_2^*=\frac{10}{7},\ x_3^*=-\frac{6}{7},\ \lambda_1^*=-\frac{8}{7},\ \lambda_2^*=\frac{4}{7}.$$

因此,二次规划的最优解及相应的乘子为

$$\boldsymbol{x}^*=\left(\frac{2}{7},\frac{10}{7},-\frac{6}{7}\right)^T,\quad \boldsymbol{\lambda}^*=\left(-\frac{8}{7},\frac{4}{7}\right)^T.$$

直接用定理 9.2.1 求解等式约束问题的最优解时,问题的维数由 n 变成了 $n+m$,问题的维数增大了许多,一种克服上述缺点的方法是变量消去法.

9.2.1 直接消去法

首先对矩阵 A 作分块,设

$$A=\begin{pmatrix}A_B\\A_N\end{pmatrix}\quad 且\ A_B\in R^{m\times m}\ 非奇异,\tag{9.2.4}$$

则相应的分块有

$$\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_B\\\boldsymbol{x}_N\end{pmatrix},\quad G=\begin{pmatrix}G_{BB}&G_{BN}\\G_{NB}&G_{NN}\end{pmatrix},\quad \boldsymbol{r}=\begin{pmatrix}\boldsymbol{r}_B\\\boldsymbol{r}_N\end{pmatrix},\tag{9.2.5}$$

因此，等式约束问题(9.2.1)可以写成

$$\min \quad f(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x}_B^T G_{BB}\boldsymbol{x}_B + \boldsymbol{x}_B^T G_{BN}\boldsymbol{x}_N + \boldsymbol{x}_N^T G_{NB}\boldsymbol{x}_B + \boldsymbol{x}_N^T G_{NN}\boldsymbol{x}_N) + \boldsymbol{r}_B^T\boldsymbol{x}_B + \boldsymbol{r}_N^T\boldsymbol{x}_N, \tag{9.2.6}$$

$$\text{s.t.} \quad A_B^T\boldsymbol{x}_B + A_N^T\boldsymbol{x}_N = b.$$

考虑问题(9.2.6)的约束条件，由于 A_B 非奇异，可将 $\boldsymbol{x}_B$ 表示成 $\boldsymbol{x}_N$ 的函数，即消去 $\boldsymbol{x}_B$，得到

$$\boldsymbol{x}_B = A_B^{-T}\boldsymbol{b} - A_B^{-T}A_N^T\boldsymbol{x}_N. \tag{9.2.7}$$

将(9.2.7)式代入问题(9.2.6)中的目标函数，得到相应的无约束问题，其目标函数为

$$\begin{aligned}\hat{f}(\boldsymbol{x}_N) = {} & \frac{1}{2}\boldsymbol{x}_N(G_{NN} - A_NA_B^{-1}G_{BN} \\ & - G_{NB}A_B^{-T}A_N^T + A_NA_B^{-1}G_{BB}A_B^{-T}A_N^T)\boldsymbol{x}_N \\ & + \boldsymbol{b}^TA_B^{-1}(G_{BN} - G_{BB}A_B^{-T}A_N^T)\boldsymbol{x}_N \\ & + (\boldsymbol{r}_N^T - \boldsymbol{r}_B^TA_B^{-T}A_N^T)\boldsymbol{x}_N \\ & + \frac{1}{2}\boldsymbol{b}^TA_B^{-1}G_{BB}A_B^{-T}\boldsymbol{b} + \boldsymbol{r}_B^TA_B^{-T}\boldsymbol{b},\end{aligned} \tag{9.2.8}$$

令

$$\begin{aligned}\hat{G}_N &= G_{NN} - A_NA_B^{-1}G_{BN} - G_{NB}A_B^{-T}A_N^T + A_NA_B^{-1}G_{BB}A_B^{-T}A_N^T, \\ \hat{\boldsymbol{r}}_N &= \boldsymbol{r}_N - A_NA_B^{-1}\boldsymbol{r}_B + (G_{NB} - A_NA_B^{-1}G_{BB})A_B^{-T}\boldsymbol{b}, \\ \hat{\boldsymbol{\delta}} &= \frac{1}{2}\boldsymbol{b}^TA_B^{-1}G_{BB}A_B^{-T}\boldsymbol{b} + \boldsymbol{r}_B^TA_B^{-T}\boldsymbol{b},\end{aligned} \tag{9.2.9}$$

则相应的无约束问题为

$$\min \quad \hat{f}(\boldsymbol{x}_N) = \frac{1}{2}\boldsymbol{x}_N^T\hat{G}_N\boldsymbol{x}_N + \hat{\boldsymbol{r}}_N^T\boldsymbol{x}_N + \hat{\delta}. \tag{9.2.10}$$

若 $\hat{G}_N$ 是正定对称矩阵，则问题(9.2.10)有唯一解

$$\boldsymbol{x}_N^* = -\hat{G}_N^{-1}\hat{\boldsymbol{r}}_N, \tag{9.2.11}$$

由(9.2.7)式可以得到问题(9.2.1)的最优解

$$x^* = \begin{pmatrix} x_B^* \\ x_N^* \end{pmatrix} = \begin{pmatrix} A_B^{-T}b + A_B^{-T}A_N^T\hat{G}_N^{-1}\hat{r}_N \\ -\hat{G}_N^{-1}\hat{r}_N \end{pmatrix}. \quad (9.2.12)$$

由于相应的乘子 λ^* 满足

$$Gx^* + r + A\lambda^* = 0,$$

即

$$\begin{pmatrix} G_{BB} & G_{BN} \\ G_{NB} & G_{NN} \end{pmatrix}\begin{pmatrix} x_B^* \\ x_N^* \end{pmatrix} + \begin{pmatrix} r_B \\ r_N \end{pmatrix} + \begin{pmatrix} A_B \\ A_N \end{pmatrix}\lambda^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

所以

$$\lambda^* = -A_B^{-1}(G_{BB}x_B^* + G_{BN}x_N^* + r_B). \quad (9.2.13)$$

定理 9.2.2 若等式约束问题(9.2.1)中的 G 是正定对称矩阵,则相应的无约束问题(9.2.10)中的 $\hat{G}_N$ 也是正定对称矩阵.

证明 构造矩阵

$$F = \begin{pmatrix} -A_B^{-T}A_N^T \\ I \end{pmatrix}, \quad (9.2.14)$$

并且 $\mathrm{rank}(F) = n - m$,因此

$$\hat{G}_N = F^TGF = (-A_NA_B^{-1}, I)\begin{pmatrix} G_{BB} & G_{BN} \\ G_{NB} & G_{NN} \end{pmatrix}\begin{pmatrix} -A_B^{-T}A_N^T \\ I \end{pmatrix}, \quad (9.2.15)$$

由于 F 是列满秩的,并且 G 正定,因此,$\hat{G}_N$ 也是正定的,对称性显然.

定理 9.2.2 表明,对于等式约束的严格凸二次规划问题,可以用直接消去法得到原问题的最优解.

【例 9.2.2】 用直接消去法求解凸二次规划问题

$$\min \quad f(x) = x_1^2 + x_2^2 + x_3^2,$$

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

$$x_1 - x_2 + x_3 = -2.$$

解 将约束写成

$$\begin{aligned} x_1 + 2x_2 &= 4 - x_3, \\ x_1 - x_2 &= -2 - x_3, \end{aligned} \tag{9.2.16}$$

求解方程组(9.2.16)得到

$$x_1 = 0 - \frac{1}{3}x_3, \quad x_2 = 2 + \frac{2}{3}x_3. \tag{9.2.17}$$

将(9.2.17)代入目标函数 $f(x)$中，得到

$$\hat{f}(x_3) = \frac{14}{9}x_3^2 + \frac{8}{3}x_3 + 4, \tag{9.2.18}$$

求解无约束问题(9.2.18)，得到

$$x_3^* = -\frac{6}{7},$$

代入(9.2.17)式中，得到约束问题的最优解

$$\boldsymbol{x}^* = \left(\frac{2}{7}, \frac{10}{7}, -\frac{6}{7},\right)^T.$$

注意到乘子 $\boldsymbol{\lambda}^*$满足方程

$$G\boldsymbol{x}^* + \boldsymbol{r} + A\boldsymbol{\lambda}^* = 0,$$

因此有

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1^* \\ \lambda_2^* \end{pmatrix} = -\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{2}{7} \\ \frac{10}{7} \\ -\frac{6}{7} \end{pmatrix}.$$

求解前两行得到

$$\lambda_1^* = -\frac{8}{7}, \quad \lambda_2^* = \frac{4}{7}.$$

这个结果与例 9.2.1 相同.

9.2.2 广义消去法

直接消去法简单、直观,它的不足之处是:A_B 可能接近一奇异矩阵,由(9.2.12)式求解 x^* 将会使数值不稳定.直接消去法的一个直接推广是**广义消去**(Generalized Elimination)法.

令 S 和 Z 分别为 $n\times m$ 和 $n\times(n-m)$ 矩阵,满足

$$A^TS = I,\quad A^TZ = 0, \tag{9.2.19}$$

且 $(S:Z)$ 为非奇异矩阵.由(9.2.19)式,得到

$$\boldsymbol{x} = S\boldsymbol{b} \tag{9.2.20}$$

是方程

$$A^T\boldsymbol{x} = \boldsymbol{b} \tag{9.2.21}$$

的一个可行解.现在讨论方程(9.2.21)的通解,设 $\boldsymbol{d}$ 是齐次方程 $A^T\boldsymbol{d}=0$ 的解,则方程(9.2.21)的通解为

$$\boldsymbol{x} = S\boldsymbol{b} + \boldsymbol{d}. \tag{9.2.22}$$

由(9.2.19)式知,Z 的列构成 A^T 的零空间的一组基,因此 $\boldsymbol{d}$ 可以表示成

$$\boldsymbol{d} = Z\boldsymbol{y} = \sum_{i=1}^{n-m} y_i z_i, \tag{9.2.23}$$

其中 $z_1,z_2,\cdots,z_{n-m}$ 为矩阵 Z 的列,$y_1,y_2,\cdots,y_{n-m}$ 为向量 $\boldsymbol{y}$ 的分量.因此,方程(9.2.21)的通解可以写成

$$\boldsymbol{x} = S\boldsymbol{b} + Z\boldsymbol{y}. \tag{9.2.24}$$

这样,由(9.2.24)式来消去问题的约束,化为一个无约束问题,即

$$\begin{aligned}\min\quad \hat{f}(\boldsymbol{y}) = {} & \frac{1}{2}\boldsymbol{y}^T(Z^TGZ)\boldsymbol{y} + (\boldsymbol{r} + GS\boldsymbol{b})^TZ\boldsymbol{y} \\ & + \frac{1}{2}(2\boldsymbol{r} + GS\boldsymbol{b})^TS\boldsymbol{b}.\end{aligned} \tag{9.2.25}$$

若 Z^TGZ 是正定的,则问题(9.2.25)有唯一的极小点,并可通过求解线性方程组

$$(Z^TGZ)\boldsymbol{y} = -Z^T(GS\boldsymbol{b} + \boldsymbol{r}) \tag{9.2.26}$$

得到其解 $\boldsymbol{y}^*$. 将 $\boldsymbol{y}^*$ 代入(9.2.24)式,得到约束问题(9.2.1)的最优解

$$x^* = Sb + Zy^* = Sb - Z(Z^TGZ)^{-1}Z^T(GSb + r). \tag{9.2.27}$$

下面计算相应的乘子,由一阶必要条件有 $A\lambda^* = -(Gx^* + r)$,左乘 S^T 得到

$$\begin{aligned}\lambda^* &= -S^T(Gx^* + r)\\ &= -S^T(G - GZ(Z^TGZ)^{-1}Z^TG)Sb\\ &\quad - (S^T - S^TGZ(Z^TGZ)^{-1}Z^T)r.\end{aligned} \tag{9.2.28}$$

因此,一旦确定了 S 和 Z,就可计算出 x^* 和 λ^*. 特别,若取

$$Z = \begin{pmatrix} -A_B^{-T}A_N^T \\ I \end{pmatrix}, \quad S = \begin{pmatrix} A_B^{-T} \\ 0 \end{pmatrix}, \quad y = x_N, \tag{9.2.29}$$

则广义消去法就是前面讲过的直接消去法.

一种最重要的取法是基于矩阵 A 的 QR 分解,将矩阵 A 分解为

$$A = Q\begin{pmatrix} R \\ 0 \end{pmatrix} = (Q_1, Q_2)\begin{pmatrix} R \\ 0 \end{pmatrix} = Q_1R, \tag{9.2.30}$$

其中 Q 为 $n \times n$ 阶正交阵,R 为 $m \times m$ 阶上三角阵,Q_1 和 Q_2 分别为 $n \times m$ 和 $n \times (n-m)$ 阶矩阵. 此时取

$$S = Q_1R^{-T}, \quad Z = Q_2, \tag{9.2.31}$$

满足条件(9.2.19),且$(S:Z)$非奇异.

下面讨论最优解的计算方法. 注意 $Sb = Q_1R^{-T}b$,先解方程

$$R^Tv = b, \tag{9.2.32}$$

得到 v,因此,$Sb = Q_1v$.

再解方程组(9.2.26),得到 y^*,所以 $x^* = Sb + Zy^*$. 乘子 λ^* 可由方程

$$R\lambda^* = -Q_1^T(Gx^* + r) \tag{9.2.33}$$

确定.

仍考虑例 9.2.2,用广义消去法求解.这里 A 的 QR 分解为

$$A=\begin{pmatrix}1&1\\2&-1\\-1&1\end{pmatrix}=\begin{pmatrix}\frac{1}{\sqrt{6}}&\frac{4}{\sqrt{21}}&\frac{1}{\sqrt{14}}\\\frac{2}{\sqrt{6}}&-\frac{1}{\sqrt{21}}&-\frac{2}{\sqrt{14}}\\-\frac{1}{\sqrt{6}}&\frac{2}{\sqrt{21}}&-\frac{3}{\sqrt{14}}\end{pmatrix}\begin{pmatrix}\sqrt{6}&-\frac{\sqrt{6}}{3}\\0&\frac{\sqrt{21}}{3}\\0&0\end{pmatrix},$$

故可取

$$S=\frac{1}{14}\begin{pmatrix}5&8\\4&-2\\-1&4\end{pmatrix},\quad Z=\begin{pmatrix}1\\-2\\-3\end{pmatrix},$$

于是向量 $Sb=\frac{1}{7}(2,10,-6)^T$,而由于 $\boldsymbol{r}=0$, $GS\boldsymbol{b}=2S\boldsymbol{b}$,所以有 $Z^T(GS\boldsymbol{b}+\boldsymbol{r})=0$,得到 $\boldsymbol{y}^*=0$,因而有 $\boldsymbol{x}^*=S\boldsymbol{b}=\frac{1}{7}(2,10,-6)^T$.由方程 $R\boldsymbol{\lambda}^*=-Q_1^T(G\boldsymbol{x}^*+\boldsymbol{r})$得到 $\boldsymbol{\lambda}^*=\left(-\frac{8}{7},\frac{4}{7}\right)^T$.请注意,若采用求解方程组的方法求出 $S\boldsymbol{b}$,则在实际计算中并不要求直接求出 S.

下面给出矩阵 S 和 Z 的一般格式.任取 $n\times(n-m)$阶矩阵,使得矩阵$(A:V)$是非奇异的,令其逆分解形式为

$$(A:V)^{-1}=\begin{pmatrix}S^T\\Z^T\end{pmatrix},\tag{9.2.34}$$

其中 S 和 Z 分别为 $n\times m$ 和 $n\times(n-m)$矩阵.显然 S 和 Z 满足(9.2.19)式.若取

$$V=\begin{pmatrix}0\\I\end{pmatrix},\tag{9.2.35}$$

则有

$$(A:V)^{-1}=\begin{pmatrix}A_B & 0\\ A_N & I\end{pmatrix}^{-1}=\begin{pmatrix}A_B^{-1} & 0\\ -A_NA_B^{-1} & I\end{pmatrix}$$
$$=\begin{pmatrix}S^T\\ Z^T\end{pmatrix},\tag{9.2.36}$$

此时,该方法是直接消去法.另外,如果取

$$V = Q_2,$$

这时,

$$(A:V)^{-1}=(Q_1R:Q_2)^{-1}=\begin{pmatrix}R^{-1}Q_1^T\\ Q_2^T\end{pmatrix},\tag{9.2.37}$$

得到的方法就是由 QR 分解得到的.

§9.3 有效集法

本节讨论凸二次规划的有效集法.有效集法的基本思想是通过求解有限个等式约束二次规划问题来得到一般约束二次规划问题(9.1.1)的最优解.

由定理 9.1.2 可知,若 x^* 是约束问题(9.1.1)的全局解,则 x^* 是等式约束问题(9.1.5)的全局解.因此,只要能确定出 x^* 处的有效约束指标集 $I(x^*)$,通过求解等式约束问题(9.1.5)就可得到一般约束问题的最优解.

9.3.1 有效集法的基本步骤

已知 $x^{(t)}$是一般约束问题(9.1.1)的可行点,确定相应的有效约束指标集

$$I(x^{(t)})=\{i\mid a_i^Tx^{(t)}=b_i,i\in I\},\tag{9.3.1}$$

并假设 $a_i(i\in E\cup I(x^{(t)}))$线性无关.

考虑等式约束问题

$$\min \quad f(x) = \frac{1}{2}x^T G x + r^T x,$$
$$s.t. \quad a_i^T x - b_i = 0, \quad i \in E \cup I(x^{(1)}), \tag{9.3.2}$$

求解等式约束问题(9.3.2),其最优解为 $\bar{x}^{(1)}$,相应的乘子为 $\lambda^{(1)}$. 下面分几种情况进行讨论:

(1)若 $\bar{x}^{(1)} \neq x^{(1)}$,由于 $\bar{x}^{(1)}$是问题(9.3.2)的最优解,因此有

$$f(\bar{x}^{(1)}) < f(x^{(1)}), \tag{9.3.3}$$

再分两种情况讨论.

(i)若 $\bar{x}^{(1)}$是原问题(9.1.1)的可行点,此时取 $x^{(2)} = \bar{x}^{(1)}$. 若 $x^{(2)}$在不等式约束的内部,则 $x^{(2)}$处的有效约束个数不变,即 $I(x^{(2)}) = I(x^{(1)})$. 若 $x^{(2)}$位于某一不等式约束的边界上(不妨设是第 p 个约束),则在 $x^{(2)}$处的有效约束个数增加一个,即 $I(x^{(2)}) = I(x^{(1)}) + \{p\}$. 重复上一轮计算.

(ii)若 $\bar{x}^{(1)}$不是原问题(9.1.1)的可行点,构造方向

$$d^{(1)} = \bar{x}^{(1)} - x^{(1)}, \tag{9.3.4}$$

由于 $x^{(1)}$是可行点,这表明从 $x^{(1)}$点出发,沿方向 $d^{(1)}$前进,在达到 $\bar{x}^{(1)}$之前,一定能遇到某约束的边界. 在什么情况下会出现这种情况呢? 我们再作进一步的分析.

由于 $\bar{x}^{(1)}$是问题(9.3.2)的最优解,因此,$\bar{x}^{(1)}$满足等式约束和 $x^{(1)}$处的有效约束,问题只能出现在那些非有效约束上. 令

$$x = x^{(1)} + \alpha d^{(1)}, \tag{9.3.5}$$

考虑 $i \in I \setminus I(x^{(1)})$,要求

$$a_i^T x - b_i = a_i^T x^{(1)} - b_i + \alpha a_i^T d^{(1)} \leqslant 0, \tag{9.3.6}$$

当 α 增加,只有当 $a_i^T d^{(1)} > 0$ 时,才有可能破坏约束条件(9.3.6). 因此

$$\alpha \leqslant \frac{b_i - a_i^T x^{(1)}}{a_i^T d^{(1)}}, \quad a_i^T d^{(1)} > 0,$$

即得到

$$\alpha_1 = \min\left\{\frac{b_i - a_i^T x^{(1)}}{a_i^T d^{(1)}} \mid a_i^T d^{(1)} > 0, i \notin I(x^{(1)})\right\}$$

$$= \frac{b_p - a_p^T x^{(1)}}{a_p^T d^{(1)}}, \tag{9.3.7}$$

此时，

$$x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} \tag{9.3.8}$$

是问题(9.1.1)的可行点，并且满足

$$f(x^{(2)}) < f(x^{(1)}).$$

由推导过程可知，在 $x^{(2)}$处的有效约束指标集 $I(x^{(2)})$满足

$$I(x^{(2)}) = I(x^{(1)}) + \{p\}. \tag{9.3.9}$$

重复上一轮计算.

(2)若 $\bar{x}^{(1)} = x^{(1)}$，它表明 $\bar{x}^{(1)}$无进展，仍分两种情况讨论.

(i)若存在 $q \in I(x^{(1)})$，并且 $\lambda_q^{(1)} < 0$. 在问题(9.3.2)中的约束 $a_q^T x - b_q = 0$ 上加一个扰动 $\varepsilon_q < 0$，即得到扰动约束 $a_q^T x - b_q = \varepsilon_q$. 求解相应的扰动问题，设其最优解为 $\bar{x}^{(1)}(\varepsilon)$，由§8.4 中乘子的意义，可知

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon_q} f(\bar{x}^{(1)}(\varepsilon))\bigg|_{\varepsilon_q = 0} = -\lambda_q^{(1)} > 0,$$

这表明当 ε_q 由 0 开始减少时，最优目标值也随着下降，换句话说，在等式约束问题(9.3.2)中当去掉约束 $a_q^T x - b_q = 0$，则可得到一个更好的点. 因此，令

$$x^{(2)} = \bar{x}^{(1)} = x^{(1)}, \quad I(x^{(2)}) = I(x^{(1)}) - \{q\}. \tag{9.3.10}$$

重复上一轮计算.

(ii)若 $\lambda_i^{(1)} \geqslant 0$, $\forall i \in I(x^{(1)})$，此时，$\bar{x}^{(1)} = x^{(1)}$是 K-T 点. 由定理 9.1.1 可知，$x^{(1)}$是二次规划的最优解.

9.3.2 等式约束问题的化简

在有效集法中，需要求解若干个等式约束二次规划问题(9.3.2)，现对问题(9.3.2)进行化简. 设 $\boldsymbol{x}^{(k)}$是一般二次规划的可行点，令

$$\boldsymbol{x} = \boldsymbol{x}^{(k)} + \boldsymbol{d}, \tag{9.3.11}$$

则

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{r}^T\boldsymbol{x} \\ &= \frac{1}{2}(\boldsymbol{x}^{(k)} + \boldsymbol{d})^T G(\boldsymbol{x}^{(k)} + \boldsymbol{d}) + \boldsymbol{r}^T(\boldsymbol{x}^{(k)} + \boldsymbol{d}) \\ &= \frac{1}{2}\boldsymbol{d}^T G\boldsymbol{d} + \nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d} + f(\boldsymbol{x}^{(k)}), \end{aligned} \tag{9.3.12}$$

而对于等式约束和有效约束，有

$$c_i(\boldsymbol{x}) = \boldsymbol{a}_i^T\boldsymbol{x} - b_i = \boldsymbol{a}_i^T\boldsymbol{x}^{(k)} - b_i + \boldsymbol{a}_i^T\boldsymbol{d} = 0,$$

因此，

$$\boldsymbol{a}_i^T\boldsymbol{d} = 0, \quad i \in E \cup I(x^{(k)}). \tag{9.3.13}$$

结合(9.3.12)式和(9.3.13)式，等式约束二次规划问题化简为

$$\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{d}^T G\boldsymbol{d} + \nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d} \\ \text{s.t.} \quad & \boldsymbol{a}_i^T\boldsymbol{d} = 0, \quad i \in E \cup I(\boldsymbol{x}^{(k)}). \end{aligned} \tag{9.3.14}$$

由此得到求解一般约束二次规划的有效集法.

9.3.3 有效集法

下面给出求解凸二次规划的有效集法.

算法 9.3.1(有效集法)

(1)取初始可行点 $\boldsymbol{x}^{(1)}$，即 $\boldsymbol{x}^{(1)}$满足

$$\boldsymbol{a}_i^T\boldsymbol{x}^{(1)} - b_i = 0, i \in E,$$

$$\boldsymbol{a}_i^T\boldsymbol{x}^{(1)} - b_i \leqslant 0, i \in I,$$

确定 $\boldsymbol{x}^{(1)}$处的有效约束指标集

$$I(\boldsymbol{x}^{(1)}) = \{i \mid \boldsymbol{a}_i^T\boldsymbol{x}^{(1)} - b_i = 0, i \in I\},$$

置 $k=1$.

(2)求解等式二次规划问题

$$\begin{aligned}
\min \quad & \frac{1}{2}\boldsymbol{d}^TG\boldsymbol{d} + \nabla f(\boldsymbol{x}^{(k)})^T\boldsymbol{d} \\
\text{s.t.} \quad & \boldsymbol{a}_i^T\boldsymbol{d} = 0, \quad i \in E \cup I(\boldsymbol{x}^{(k)}),
\end{aligned}$$

得到 $\boldsymbol{d}^{(k)}$.

(3)若 $\boldsymbol{d}^{(k)} = 0$,则计算相应的乘子 $\boldsymbol{\lambda}^{(k)}$. 若 $\lambda_i^{(k)} \geqslant 0$, $\forall i \in I(\boldsymbol{x}^{(k)})$,则停止计算($\boldsymbol{x}^{(k)}$为一般二次规划(9.1.1)的最优解,$\boldsymbol{\lambda}^{(k)}$为相应的乘子);否则求

$$\lambda_q^{(k)} = \min\{\lambda_i^{(k)} \mid i \in I(\boldsymbol{x}^{(k)})\},$$

并置 $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)}$, $I(\boldsymbol{x}^{(k+1)}) = I(\boldsymbol{x}^{(k)}) - \{q\}$, $k=k+1$,转(2).

(4)($\boldsymbol{d}^{(k)} \neq 0$)计算

$$\begin{aligned}
\hat{\alpha}_k &= \min\left\{\frac{b_i - \boldsymbol{a}_i^T\boldsymbol{x}^{(k)}}{\boldsymbol{a}_i^T\boldsymbol{d}^{(k)}} \mid \boldsymbol{a}_i^T\boldsymbol{d}^{(k)} > 0, i \notin I(\boldsymbol{x}^{(k)})\right\} \\
&= \frac{b_p - \boldsymbol{a}_p^T\boldsymbol{x}^{(k)}}{\boldsymbol{a}_p^T\boldsymbol{d}^{(k)}},
\end{aligned}$$

取 $\alpha_k = \min\{\hat{\alpha}_k, 1\}$,置

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}.$$

如果 $\alpha_k = \hat{\alpha}_k$,则置

$$I(\boldsymbol{x}^{(k+1)}) = I(\boldsymbol{x}^{(k)}) + \{p\};$$

否则置

$$I(\boldsymbol{x}^{(k+1)}) = I(\boldsymbol{x}^{(k)}),$$

置 $k=k+1$ 转(2).

【例 9.3.1】 用有效集法求解二次规划问题

$$\min \quad f(\boldsymbol{x}) = x_1^2 + x_2^2 - 2x_1 - 4x_2,$$

$$\text{s.t.}\quad c_1(\boldsymbol{x}) = x_1 + x_2 - 1 \leqslant 0,$$
$$c_2(\boldsymbol{x}) = -x_1 \leqslant 0,$$
$$c_3(\boldsymbol{x}) = -x_2 \leqslant 0,$$

取初始点 $\boldsymbol{x}^{(1)} = (0,0)^T$.

解 计算 $\nabla f(\boldsymbol{x}) = (2x_1 - 2, 2x_2 - 4)^T$. 因为 $\boldsymbol{x}^{(1)} = (0,0)^T$，所以 $I(\boldsymbol{x}^{(1)}) = \{2,3\}$，$\nabla f(\boldsymbol{x}^{(1)}) = (-2,-4)^T$，相应的等式约束问题为

$$\begin{aligned} \min \quad & d_1^2 + d_2^2 - 2d_1 - 4d_2, \\ \text{s.t.} \quad & -d_1 = 0, \\ & -d_2 = 0. \end{aligned}$$

由等式约束问题的一阶必要条件得到

$$\begin{pmatrix} 2 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 0 \\ 0 \end{pmatrix},$$

求解方程组得到 $d_1 = 0, d_2 = 0, \lambda_2 = -2, \lambda_3 = -4$，即 $\boldsymbol{d}^{(1)} = (0,0)^T$，$\boldsymbol{\lambda}^{(1)} = (0,-2,-4)^T$，所以 $\boldsymbol{x}^{(1)} = (0,0)^T$ 不是最优解.

再进行第二轮计算. 取 $\boldsymbol{x}^{(2)} = (0,0)^T$，$I(\boldsymbol{x}^{(2)}) = I(\boldsymbol{x}^{(1)}) - \{3\} = \{2\}$，相应的等式二次规划问题为

$$\begin{aligned} \min \quad & d_1^2 + d_2^2 - 2d_1 - 4d_2, \\ \text{s.t.} \quad & -d_1 = 0, \end{aligned}$$

相应的一阶必要条件为

$$\begin{pmatrix} 2 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 0 \end{pmatrix},$$

求解方程组得到 $d_1 = 0, d_2 = 2$，即 $\boldsymbol{d}^{(2)} = (0,2)^T$.

计算

$$\hat{\alpha}_2 = \frac{b_1 - a_1^T x^{(2)}}{a_1^T d^{(2)}} = \frac{1-0}{2} = \frac{1}{2},$$

$$\alpha_2 = \min\left\{\frac{1}{2}, 1\right\} = \frac{1}{2},$$

所以

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} + \alpha_2 \boldsymbol{d}^{(2)} = (0,0)^T + \frac{1}{2}(0,2)^T = (0,1)^T,$$

由于 $\alpha_2 = \hat{\alpha}_2 = \frac{1}{2}$，所以在 $\boldsymbol{x}^{(3)}$处的有效约束力 $I(\boldsymbol{x}^{(3)}) = I(\boldsymbol{x}^{(2)}) + \{1\} = \{1,2\}$.

下面进行第三轮计算. 在 $\boldsymbol{x}^{(3)}$处$\nabla f(\boldsymbol{x}^{(3)}) = (-2,-2)^T$，相应的等式二次规划问题为

$$\begin{aligned} \min \quad & d_1^2 + d_2^2 - 2d_1 - 2d_2, \\ \text{s.t.} \quad & d_1 + d_2 = 0, \\ & -d_1 = 0, \end{aligned}$$

由等式约束问题的一阶必要条件得到

$$\begin{bmatrix} 2 & 0 & 1 & -1 \\ 0 & 2 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix},$$

求解方程组得到 $d_1 = 0, d_2 = 0, \lambda_1 = 2, \lambda_2 = 0$，即 $\boldsymbol{d}^{(3)} = (0,0)^T$，$\boldsymbol{\lambda}^{(3)} = (2,0,0)^T$，因此，$\boldsymbol{x}^{(3)} = (0,1)^T$ 是最优解，$\boldsymbol{\lambda}^{(3)} = (2,0,0)^T$ 是相应的乘子.

在实际计算中，初始点 $\boldsymbol{x}^{(1)}$的选取并非易事，需要用类似于线性规划求初始基本可行点的方法构造辅助问题，得到初始可行点. 若二次规划是严格凸的，且是非退化的，则有效集法经有限步运算，可求出二次规划的最优解(证明略). 当 G 是非正定矩阵时，按

照上述方法计算，有可能得不到最优解．若想求出最优解，需对算法作必要的改动，这里就不深入讨论了．

§9.4 对偶问题

9.4.1 对偶二次规划问题

本节讨论严格凸二次规划问题

$$\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{x}^T G \boldsymbol{x} + \boldsymbol{r}^T \boldsymbol{x}, \\ \text{s.t.} \quad & A^T \boldsymbol{x} \leqslant \boldsymbol{b}, \end{aligned} \tag{9.4.1}$$

其中 G 是正定对称矩阵，$A=(a_1,a_2,\cdots,a_m)$，$\boldsymbol{b}=(b_1,b_2,\cdots,b_m)^T$．由最优解的一阶必要条件，存在 $\boldsymbol{\lambda}^*\geqslant 0$，在 $\boldsymbol{x}^*$ 处满足

$$G\boldsymbol{x}^* + \boldsymbol{r} + A\boldsymbol{\lambda}^* = 0,$$

因此，

$$\boldsymbol{x}^* = -G^{-1}(A\boldsymbol{\lambda}^* + \boldsymbol{r}). \tag{9.4.2}$$

由于 $\boldsymbol{x}^*$ 是可行点，因此有

$$A^T\boldsymbol{x}^* - \boldsymbol{b} \leqslant 0,$$

引进松弛变量 $\boldsymbol{y}^*\geqslant 0$，得到

$$A^T\boldsymbol{x}^* + \boldsymbol{y}^* - \boldsymbol{b} = 0. \tag{9.4.3}$$

将(9.4.2)式代入(9.4.3)得到

$$-A^T G^{-1}(A\boldsymbol{\lambda}^* + \boldsymbol{r}) + \boldsymbol{y}^* - \boldsymbol{b} = 0,$$

因此有

$$(A^T G^{-1} A)\boldsymbol{\lambda}^* - \boldsymbol{y}^* + A^T G^{-1}\boldsymbol{r} + \boldsymbol{b} = 0. \tag{9.4.4}$$

令

$$D = A^T G^{-1} A, \quad \boldsymbol{d} = A^T G^{-1}\boldsymbol{r} + \boldsymbol{b},$$

(9.4.4)式可写成

$$D\boldsymbol{\lambda}^* - \boldsymbol{y}^* + \boldsymbol{d} = 0, \tag{9.4.5}$$

并有

$$\boldsymbol{\lambda}^* \geqslant 0, \quad \boldsymbol{y}^* \geqslant 0, \tag{9.4.6}$$

因此,(9.4.5)式和(9.4.6)式是下面二次规划

$$\begin{aligned} \min \quad & \frac{1}{2}\boldsymbol{\lambda}^T D\boldsymbol{\lambda} + \boldsymbol{d}^T\boldsymbol{\lambda}, \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geqslant 0 \end{aligned} \tag{9.4.7}$$

的 K-T 条件,$\boldsymbol{y}^*$ 为 $\boldsymbol{\lambda}^*$ 处相应的乘子.二次规划问题(9.4.7)要比原二次规划问题(9.4.1)简单的多,通常称问题(9.4.7)是二次规划问题(9.4.1)的对偶问题.因此,可通过求解问题(9.4.7)来得到原二次规划问题(9.4.1)的最优解.

9.4.2 Hilderth-D'Espo 方法

该方法是以解二次规划问题(9.4.7)为基础的,再由(9.4.2)式得到原问题(9.4.1)式的最优解.若没有 $\boldsymbol{\lambda} \geqslant 0$ 的限制,则 $\boldsymbol{\lambda}^*$ 满足

$$D\boldsymbol{\lambda} + \boldsymbol{d} = 0, \tag{9.4.8}$$

注意 D 是半正定矩阵,特别当 A 为列满秩时,D 为正定矩阵,但无论哪种情况,D 的对角元素均是正的,因此可用求解线性方程组的迭代法求解方程组(9.4.8).

我们采用 Jacobi 迭代格式求解方程组(9.4.8).假设已知 $\lambda^{(k)}$,其迭代格式为

$$\begin{aligned} \hat{\lambda}_i &= -\sum_{\substack{j=1 \\ j\neq i}}^{m} \frac{d_{ij}}{d_{ii}}\lambda_j^{(k)} - \frac{d_i}{d_{ii}}, i = 1,2,\cdots,m, \\ \lambda_i^{(k+1)} &= \max\{0,\hat{\lambda}_i\}, i = 1,2,\cdots,m, \end{aligned} \tag{9.4.9}$$

其中 $D=(d_{ij})_{m\times m}$,$\boldsymbol{d}=(d_1,d_2,\cdots,d_m)^T$.

算法终止准则为

$$\|\boldsymbol{\lambda}^{(k+1)} - \boldsymbol{\lambda}^{(k)}\| \leqslant \varepsilon. \tag{9.4.10}$$

当然,在述算法中,也可以采用 Gauss-Seidel 迭代格式求解方程组(9.4.8).

【例 9.4.1】 Hilderth-D'Espo 方法求解二次规划问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = \frac{1}{2}((x_1 - 2)^2 + (x_2 - 2)^2), \\ \text{s.t.} \quad & 0 \leqslant x_1 \leqslant 1, \\ & 0 \leqslant x_2 \leqslant 1. \end{aligned}$$

解 由于

$$G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{r} = \begin{pmatrix} -2 \\ -2 \end{pmatrix},$$

$$A = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

因此

$$D = A^T G^{-1} A = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{d} = A^T G^{-1} \boldsymbol{r} + \boldsymbol{b} = \begin{pmatrix} -1 \\ -1 \\ 2 \\ 2 \end{pmatrix},$$

得到的线性方程组为

$$\begin{cases} \lambda_1 - \lambda_3 - 1 = 0, \\ \lambda_2 - \lambda_4 - 1 = 0, \\ -\lambda_1 + \lambda_3 + 2 = 0, \\ -\lambda_2 + \lambda_4 + 2 = 0, \end{cases}$$

因此，Jacobi 迭代格式为

$$\begin{cases} \hat{\lambda}_1 = \lambda_3^{(k)} + 1, \\ \hat{\lambda}_2 = \lambda_4^{(k)} + 1, \\ \hat{\lambda}_3 = \lambda_1^{(k)} - 2, \\ \hat{\lambda}_4 = \lambda_2^{(k)} - 2 \end{cases}$$

和

$$\lambda_i^{(k+1)} = \max\{0, \hat{\lambda}_i\}, \quad i = 1,2,3,4,$$

取 $\boldsymbol{\lambda}^{(0)} = (0,0,0,0)^T$，则有 $\hat{\lambda}_1 = 1, \hat{\lambda}_2 = 1, \hat{\lambda}_3 = -2, \hat{\lambda}_4 = -2$，故 $\boldsymbol{\lambda}^{(1)} = (1,1,0,0)^T$.

再进行第二轮计算，得到 $\hat{\lambda}_1 = 1, \hat{\lambda}_2 = 1, \hat{\lambda}_3 = -1, \hat{\lambda}_4 = -1$，因此 $\boldsymbol{\lambda}^{(2)} = (1,1,0,0)^T$. 由于 $\| \boldsymbol{\lambda}^{(2)} - \boldsymbol{\lambda}^{(1)} \| = 0$，停止计算. 得到

$$\boldsymbol{\lambda}^* = (1,1,0,0)^T,$$

代入(9.4.2)式，得到

$$\boldsymbol{x}^* = -G^{-1}(A\boldsymbol{\lambda}^* + \boldsymbol{r}) = (1,1)^T,$$

所以原二次规划问题的最优解为 $\boldsymbol{x}^* = (1,1)^T$，相应的乘子为 $\boldsymbol{\lambda}^* = (1,1,0,0)^T$.

Hilerth-D'Espo 方法的缺点是效率低，当迭代不收敛时，无法得到原问题的最优解.

习 题 九

9.1 求解下列等式约束二次规划问题：

(1) $$\begin{aligned} \min \quad & f(\boldsymbol{x}) = x_1^2 + x_2^2, \\ \text{s.t.} \quad & c(\boldsymbol{x}) = x_1 + x_2 - 1 = 0. \end{aligned}$$

(2) $\min \quad f(\boldsymbol{x}) = \frac{3}{2}x_1^2 + x_2^2 + \frac{1}{2}x_3^2 - x_1x_2 - x_2x_3 + x_1 + x_2 + x_3,$

s.t.　$c(\boldsymbol{x})=x_1+2x_2+x_3-4=0.$

9.2　用直接消去法求解 9.1 题.

9.3　用有效集法求解下列凸二次规划问题：

(1)
$$
\begin{aligned}
\min\quad & f(\boldsymbol{x}) = x_1^2+4x_2^2-2x_1-x_2,\\
\text{s.t.}\quad & x_1+x_2\leqslant 1,\\
& x_1,x_2\geqslant 0.
\end{aligned}
$$

(2)
$$
\begin{aligned}
\min\quad & f(\boldsymbol{x}) = x_1^2+4x_2^2,\\
\text{s.t.}\quad & x_1-x_2\leqslant 1,\\
& x_2\leqslant 1,\\
& x_1+x_2\geqslant 1.
\end{aligned}
$$

9.4　x^* 是二次规划问题

$$
\begin{aligned}
\min\quad & f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x}+\boldsymbol{r}^T\boldsymbol{x},\\
\text{s.t.}\quad & A^T\boldsymbol{x} = \boldsymbol{b}
\end{aligned}
$$

的全局解的充分必要条件是：x^*，λ^* 满足

$$
\begin{aligned}
\boldsymbol{\lambda}^* &= -(A^TG^{-1}A)^{-1}(A^TG\boldsymbol{r}+\boldsymbol{b}),\\
\boldsymbol{x}^* &= -G^{-1}(\boldsymbol{r}+A\boldsymbol{\lambda}^*).
\end{aligned}
$$

9.5　用 Hilderth-D'Espo 方法求解习题 9.3.

9.6　用 Hilderth-D'Espo 方法求解二次规划问题

$$
\begin{aligned}
\min\quad & f(\boldsymbol{x}) = \frac{3}{2}x_1^2+\frac{1}{2}x_2^2+x_1x_2-15x_1-7x_2,\\
\text{s.t.}\quad & x_1+x_2\leqslant 3,\\
& 2x_1-x_2\leqslant 4,\\
& x_2\leqslant 2,\\
& x_1,x_2\geqslant 0,
\end{aligned}
$$

你将发现 $\boldsymbol{\lambda}^{(k)}$ 不收敛.

第十章　可行方向法

本章介绍的方法是通过在可行域内直接搜索最优解来求解约束最优化问题，该方法是从可行点出发，沿可行方向前进寻找最优点，因此称为**可行方向法**（Method of Feasible Directions）. 为了便于讨论，本章仅讨论求解线性约束问题

$$
\begin{aligned}
\min\quad & f(\boldsymbol{x}),\quad \boldsymbol{x}\in R^n,\\
\text{s.t.}\quad & \boldsymbol{a}_i^T x - b_i = 0,\ i\in E=\{1,2,\cdots,l\},\\
& \boldsymbol{a}_i^T x - b_i \leqslant 0,\ i\in I=\{l+1,l+2,\cdots,l+m\}
\end{aligned}
\tag{10.0.1}
$$

的可行方向法. 这些方法也可推广到非线性约束问题.

§10.1　可行方向法

最早的可行方向法是 G.Zoutendijk 在 1960 年首先提出的，因此这里介绍的可行方向法称为 Zoutendijk 可行方向法.

下面给出可行方向的定义，注意它与第八章中可行方向定义的差别.

定义 10.1.1　设 $\overline{\boldsymbol{x}}$ 是约束问题（10.0.1）的可行点，D 是约束问题（10.0.1）的可行域，即

$$D=\{\boldsymbol{x}\mid \boldsymbol{a}_i^T\boldsymbol{x}-b_i=0,\ i\in E,\ \boldsymbol{a}_i^T\boldsymbol{x}-b_i\leqslant 0,\ i\in I\},\tag{10.1.1}$$

若 $\boldsymbol{d}\neq 0$，$\boldsymbol{d}\in R^n$，存在 $\delta>0$，使当 $\alpha\in(0,\delta]$ 时，有

$$\boldsymbol{x}+\alpha\boldsymbol{d}\in D,$$

则称 $\boldsymbol{d}$ 为 $\bar{x}$ 处的一个可行方向.

下面给出可行方向的一个充分必要条件.

定理 10.1.1 设 $\bar{x}$ 是约束问题（10.0.1）的可行点，则 $\boldsymbol{d}$ 为 $\bar{x}$ 处的可行方向的充分必要条件是：

$$\begin{aligned} &\boldsymbol{a}_i^T\boldsymbol{d}=0, \quad i\in E, \\ &\boldsymbol{a}_i^T\boldsymbol{d}\leqslant 0, \quad i\in I(\bar{\boldsymbol{x}}), \end{aligned} \tag{10.1.2}$$

这里 $I(\bar{\boldsymbol{x}})$ 是 $\bar{\boldsymbol{x}}$ 处的有效约束指标集.

证明 “必要性”. 设 $\bar{\boldsymbol{x}}$ 是可行点. $\boldsymbol{d}$ 是 $\bar{\boldsymbol{x}}$ 处的可行方向，则存在 $\delta>0$，使当 $\alpha\in(0,\delta]$ 时有 $\bar{\boldsymbol{x}}+\alpha\boldsymbol{d}\in D$，即满足

$$\begin{aligned} &\boldsymbol{a}_i^T(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i=0, \ i\in E, \\ &\boldsymbol{a}_i^T(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i\leqslant 0, \ i\in I. \end{aligned} \tag{10.1.3}$$

因为 $\bar{\boldsymbol{x}}\in D$，当 $i\in E\cup I(\bar{x})$时，有 $\boldsymbol{a}_i^T\bar{\boldsymbol{x}}-b_i=0$. 因此（10.1.2）式成立.

“充分性”. 设 $\bar{\boldsymbol{x}}$ 是可行点，$\boldsymbol{d}$ 满足（10.1.2）式，$\forall\alpha$ 有

$$\boldsymbol{a}_i^T(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i=\boldsymbol{a}_i^T\bar{\boldsymbol{x}}-b_i+\alpha\boldsymbol{a}_i^T\boldsymbol{d}=0, \ i\in E, \tag{10.1.4}$$

$\forall\alpha\geqslant 0$ 有

$$\boldsymbol{a}_i^T(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i=\boldsymbol{a}_i^T\bar{\boldsymbol{x}}-b_i+\alpha\boldsymbol{a}_i^T\boldsymbol{d}\leqslant 0, \ i\in I(\bar{\boldsymbol{x}}). \tag{10.1.5}$$

当 $i\in I\setminus I(\bar{\boldsymbol{x}})$，由于 $\boldsymbol{a}_i^T\bar{\boldsymbol{x}}-b_i<0$，只有当 $\boldsymbol{a}_i^T\boldsymbol{d}>0$ 时，约束

$$\boldsymbol{a}_i^T(\bar{\boldsymbol{x}}+\alpha\boldsymbol{d})-b_i=\boldsymbol{a}_i^T\bar{\boldsymbol{x}}-b_i+\alpha\boldsymbol{a}_i^T\boldsymbol{d}\leqslant 0 \tag{10.1.6}$$

才有可能遭到破坏，因此只需取

$$\delta=\min\left\{\frac{b_i-\boldsymbol{a}_i^T\bar{\boldsymbol{x}}}{\boldsymbol{a}_i^T\boldsymbol{d}} \mid \boldsymbol{a}_i^T\boldsymbol{d}>0, \ i\notin I(\bar{\boldsymbol{x}})\right\},$$

当 $\alpha\in(0,\delta]$ 时，有（10.1.6）式成立.

定义 10.1.2 设$\bar{x}$是约束问题（10.0.1）的可行点，若 $\boldsymbol{d}$ 是$\bar{x}$处的可行方向，又是$\bar{x}$处的下降方向，则称 $\boldsymbol{d}$ 是$\bar{x}$处的可行下降方向.

前面已讲过，若 $\boldsymbol{d}$ 满足$\nabla f(\bar{x})^T\boldsymbol{d}<0$,则 $\boldsymbol{d}$ 是$\bar{x}$处的下降方向.由定理 10.1.1 可知,若 $\boldsymbol{d}$ 满足：

$$\boldsymbol{a}_i^T\boldsymbol{d}=0,\ i\in E, \tag{10.1.7}$$

$$\boldsymbol{a}_i^T\boldsymbol{d}\leqslant 0, i\in I(\bar{x}), \tag{10.1.8}$$

$$\nabla f(\bar{x})^T\boldsymbol{d}<0, \tag{10.1.9}$$

则 $\boldsymbol{d}$ 是 $\bar{x}$ 处的可行下降方向.

因此，我们的目标是寻找满足（10.1.7）式～（10.1.9）式的 $\boldsymbol{d}$，得到可行下降方向. 这样，从 $\bar{x}$ 出发，沿方向 $\boldsymbol{d}$ 前进，得到一个新的可行点 $\hat{x}$，而在 $\hat{x}$ 满足

$$f(\hat{x})<f(\bar{x}).$$

从下降算法的角度来看,自然希望在满足(10.1.7)式和(10.1.8)式的前题下,$\nabla f(\bar{x})^Td$ 越小越好，这就引出一个线性规划问题

$$\begin{aligned}\min\quad & \nabla f(\bar{x})^T\boldsymbol{d},\\ \text{s.t.}\quad & \boldsymbol{a}_i^T\boldsymbol{d}=0,\ i\in E,\\ & \boldsymbol{a}_i^T\boldsymbol{d}\leqslant 0,\ i\in I(\bar{x}).\end{aligned} \tag{10.1.10}$$

那么，线性规划（10.1.10）的最优解正是我们所期望的.但这同时又会出现一个问题，设 $\hat{\boldsymbol{d}}$ 是线性规划（10.1.10）的最优解，并且满足$\nabla f(\bar{x})^T\hat{\boldsymbol{d}}<0$,令 $\boldsymbol{d}=\alpha\hat{\boldsymbol{d}}$，则 d 满足约束条件，但当 $\alpha\to+\infty$时，有$\nabla f(\bar{x})^T\boldsymbol{d}\to-\infty$，即线性规划问题(10.1.10）无有限最优解. 这给线性规划问题（10.1.10）的求解带来了困难.

为解决这一问题，就必须对 d 或目标函数$\nabla f(\bar{x})^T\boldsymbol{d}$ 加以某些限制，称这些限制为**规范约束**（Normalization Constraint）. 根

据规范约束的不同，可以得到不同的规划问题. 在这里给出三种规划问题.

问题Ⅰ：

$$
\begin{aligned}
\min \quad & \nabla f(\bar{x})^T \boldsymbol{d}, \\
\text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{d} = 0, \qquad i \in E, \\
& \boldsymbol{a}_i^T \boldsymbol{d} \leqslant 0, \qquad i \in I(\bar{x}), \\
& -1 \leqslant d_i \leqslant 1, \quad i = 1,2,\cdots,n.
\end{aligned}
\tag{10.1.11}
$$

问题Ⅱ：

$$
\begin{aligned}
\min \quad & \nabla f(\bar{x})^T \boldsymbol{d}, \\
\text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{d} = 0, \qquad i \in E, \\
& \boldsymbol{a}_i^T \boldsymbol{d} \leqslant 0, \qquad i \in I(\bar{x}), \\
& \boldsymbol{d}^T \boldsymbol{d} \leqslant 1.
\end{aligned}
\tag{10.1.12}
$$

问题Ⅲ：

$$
\begin{aligned}
\min \quad & \nabla f(\bar{x})^T \boldsymbol{d}, \\
\text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{d} = 0, \qquad i \in E, \\
& \boldsymbol{a}_i^T \boldsymbol{d} \leqslant 0, \qquad i \in I(\bar{x}), \\
& \nabla f(\bar{x})^T \boldsymbol{d} \geqslant -1.
\end{aligned}
\tag{10.1.13}
$$

在上述三个规划问题中，问题Ⅰ和问题Ⅲ是线性规划问题，因此可用求解线性规划问题的算法求解，问题Ⅱ虽然是一个非线性规划问题，但具有特殊的形式，因此也可用特殊的方法求解.

定理 10.1.2 设 $\bar{x}$ 是约束问题（10.0.1）的可行点，则 $\bar{x}$ 成为约束问题（10.0.1）的 K-T 点的充分必要条件是：问题Ⅰ或问题Ⅱ或问题Ⅲ的最优目标函数值为 0.

证明 以问题Ⅰ为例，其它两个问题同理.

“充分性”. 设问题Ⅰ的最优目标函数值为 0，由对满足问题

Ⅰ的约束条件的 $\boldsymbol{d}$，均有$\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{d}\geqslant 0$，即系统

$$\begin{aligned}&\boldsymbol{a}_i^T\boldsymbol{d}=0,\ i\in E,\\&\boldsymbol{a}_i^T\boldsymbol{d}\leqslant 0,\ i\in I(\bar{\boldsymbol{x}}),\\&-\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{d}>0\end{aligned}$$

无解，由定理 8.2.2（Farkas 引理）的推论 2，存在着$\bar{\lambda}_i(i\in E\cup I(\bar{\boldsymbol{x}}))$，使得

$$\begin{aligned}&-\nabla f(\bar{\boldsymbol{x}})=\sum_{i\in E\cup I(\bar{x})}\bar{\lambda}_i\boldsymbol{a}_i,\\&\bar{\lambda}_i\geqslant 0,\quad i\in I(\bar{\boldsymbol{x}}).\end{aligned}\tag{10.1.14}$$

令$\bar{\lambda}_i=0, i\in I\setminus I(\bar{\boldsymbol{x}})$，则有

$$\begin{aligned}&\nabla f(\bar{\boldsymbol{x}})+\sum_{i=1}^{l+m}\bar{\lambda}_i\boldsymbol{a}_i=0,\\&\bar{\lambda}_i\geqslant 0,\quad i\in I,\\&\bar{\lambda}_ic_i(\bar{\boldsymbol{x}})=0,\quad i\in I,\end{aligned}$$

即 $\bar{\boldsymbol{x}}$ 是 K-T 点.

“必要性”. 设 $\bar{\boldsymbol{x}}$ 是 K-T 点，则有

$$\nabla f(\bar{\boldsymbol{x}})=-\sum_{i=1}^{l+m}\bar{\lambda}_i\boldsymbol{a}_i,$$

若问题Ⅰ的目标函数值小于 0，则有

$$\begin{aligned}0>\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{d}&=-\sum_{i\in E}\bar{\lambda}_i\boldsymbol{a}_i^T\boldsymbol{d}-\sum_{i\in I(\bar{x})}\bar{\lambda}_i\boldsymbol{a}_i^T\boldsymbol{d}-\sum_{i\in I\setminus I(\bar{x})}\bar{\lambda}_i\boldsymbol{a}_i^T\boldsymbol{d}\\&=-\sum_{i\in I(\bar{x})}\bar{\lambda}_i\boldsymbol{a}_i^T\boldsymbol{d}\geqslant 0,\end{aligned}\tag{10.1.15}$$

矛盾. 因此问题Ⅰ的最优目标函数值为 0.

由定理 10.1.2 可知，若问题Ⅰ或问题Ⅱ或问题Ⅲ的最优目标函数值为 0，则 $\bar{\boldsymbol{x}}$ 是 K-T 点，停止计算. 若最优目标函数值小于 0，则 $\boldsymbol{d}$ 是 $\bar{\boldsymbol{x}}$ 处的可行下降方向，那么沿 $\boldsymbol{d}$ 方向进行一维搜索，得到下一个可行点 $\hat{\boldsymbol{x}}$，并且有$f(\hat{\boldsymbol{x}})<f(\bar{\boldsymbol{x}})$.

一维搜索问题

在可行方向法中的一维搜索与无约束问题中的一维搜索是有区别的. 在无约束问题中，一维搜索只需求该方向上的极小点，或者有一定的下降量的点，而在可行方向法中，一维搜索除使目标函数值下降外，还要保证其点在可行域内，即实际上在一个区间上求一维极小. 下面给出搜索区间的具体讨论.

设 $\boldsymbol{x}^{(k)}$是约束问题（10.0.1）的可行点，$\boldsymbol{d}^{(k)}$是 $\boldsymbol{x}^{(k)}$处的可行下降方向，令

$$\boldsymbol{x} = \boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}, \tag{10.1.16}$$

考虑约束条件，注意 $\boldsymbol{d}^{(k)}$是问题Ⅰ或问题Ⅱ或问题Ⅲ的解，因此有

$$\begin{aligned}
&\boldsymbol{a}_i^T\boldsymbol{x} - b_i = \boldsymbol{a}_i^T\boldsymbol{x}^{(k)} - b_i + \alpha\boldsymbol{a}_i^T\boldsymbol{d}^{(k)} = 0, i \in E, \forall \alpha,\\
&\boldsymbol{a}_i^T\boldsymbol{x} - b_i = \boldsymbol{a}_i^T\boldsymbol{x}^{(k)} - b_i + \alpha\boldsymbol{a}_i^T\boldsymbol{d}^{(k)} \leqslant 0, i \in I(\boldsymbol{x}^{(k)}), \forall \alpha \geqslant 0,
\end{aligned} \tag{10.1.17}$$

因此，可能被破坏的约束是 $x^{(k)}$处的非有效约束

$$\boldsymbol{a}_i^T\boldsymbol{x} - b_i = \boldsymbol{a}_i^T\boldsymbol{x}^{(k)} - b_i + \alpha\boldsymbol{a}_i^T\boldsymbol{d}^{(k)} \leqslant 0, i \notin I(x^{(k)}) \tag{10.1.18}$$

中 $\boldsymbol{d}^{(k)}$满足 $\boldsymbol{a}_i^T\boldsymbol{d}^{(k)}>0$ 的那些约束. 为保证可行性，一维搜索步长应满足

$$\alpha \leqslant \min\left\{\frac{b_i - \boldsymbol{a}_i^T\boldsymbol{x}^{(k)}}{\boldsymbol{a}_i^T\boldsymbol{d}^{(k)}} \mid \boldsymbol{a}_i^T\boldsymbol{d}^{(k)} > 0, i \notin I(\boldsymbol{x}^{(k)})\right\},$$

因此取

$$\alpha_{\max} = \min\left\{\frac{b_i - \boldsymbol{a}_i^T\boldsymbol{x}^{(k)}}{\boldsymbol{a}_i^T\boldsymbol{d}^{(k)}} \mid \boldsymbol{a}_i^T\boldsymbol{d}^{(k)} > 0, i \notin I(\boldsymbol{x}^{(k)})\right\}, \tag{10.1.19}$$

若集合 $\{a_i^T d^{(k)}>0,\ i\notin I(x^{(k)})\}=\phi$，则令

$$\alpha_{\max}=+\infty, \tag{10.1.20}$$

因此，一维搜索问题为

$$\begin{aligned} \min\quad & \phi(\alpha)=f(x^{(k)}+\alpha d^{(k)}), \\ \text{s.t.}\quad & 0\leqslant\alpha\leqslant\alpha_{\max}. \end{aligned} \tag{10.1.21}$$

综上所述，给出相应的算法.

算法 10.1.1（Zoutendijk 可行方向法）

（1）取初始可行点 $x^{(1)}$，即 $x^{(1)}$满足

$$a_i^T x^{(1)}-b_i=0,\ i\in E,$$

$$a_i^T x^{(1)}-b_i\leqslant 0,\ i\in I,$$

置 $k=1$.

（2）确定 $x^{(k)}$处的有效约束指标集

$$I(x^{(k)})=\{i\mid a_i^T x^{(k)}-b_i=0, i\in I\}.$$

（3）求解线性规划子问题

$$\begin{aligned} \min\quad & \nabla f(x^{(k)})^T d, \\ \text{s.t.}\quad & a_i^T d=0,\quad i\in E, \\ & a_i^T d\leqslant 0,\quad i\in I(x^{(k)}), \\ & -1\leqslant d_i\leqslant 1,\quad i=1,2,\cdots,n. \end{aligned}$$

得到 $d^{(k)}$.

（4）若$\nabla f(x^{(k)})^T d^{(k)}=0$，则停止计算（$x^{(k)}$为 K-T 点）；否则求解一维问题

$$\begin{aligned} \min\quad & \phi(\alpha)=f(x^{(k)}+\alpha d^{(k)}), \\ \text{s.t.}\quad & 0\leqslant\alpha\leqslant\alpha_{\max}, \end{aligned}$$

其中

$$\alpha_{\max}=\begin{cases}\min\left\{\dfrac{b_i-a_i^T x^{(k)}}{a_i^T x^{(k)}}\mid a_i^T d^{(k)}>0,\ i\notin I(x^{(k)})\right\},\\ +\infty,\ \{a_i^T d^{(k)}>0,\ i\notin I(x^{(k)})\}=\phi,\end{cases}$$

得到 α_k，置

$$x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}.$$

(5) 置 $k=k+1$，转 (2).

在实际计算中，算法第 (4) 步 $\nabla f(x^{(k)})^T d^{(k)}=0$ 可改为 $|\nabla f(x^{(k)})^T d^{(k)}|\leqslant\epsilon$，其中 ϵ 是预先指定的精度。

【例 10.1.1】 用可行方向法求解线性约束问题

$$\begin{aligned}\min\quad & f(x)=2x_1^2-2x_1x_2+2x_2^2-4x_1-6x_2,\\ \text{s.t.}\quad & a_1^T x-b_1=x_1+x_2-2\leqslant 0,\\ & a_2^T x-b_2=x_1+5x_2-5\leqslant 0,\\ & a_3^T x-b_3=-x_1\leqslant 0,\\ & a_4^T x-b_4=-x_2\leqslant 0,\end{aligned}\tag{10.1.22}$$

取初始点 $x^{(1)}=(0,0)^T$.

解

$$\nabla f(x)=(4x_1-2x_2-4,\ -2x_1+4x_2-6)^T,$$

当 $x^{(1)}=(0,0)^T$ 时，$\nabla f(x^{(1)})=(-4,-6)^T$. 有效约束指标集为 $I(x^{(1)})=\{3,4\}$. 所以线性规划子问题为

$$\begin{aligned}\min\quad & -4d_1-6d_2,\\ \text{s.t.}\quad & -d_1\leqslant 0,\\ & -d_2\leqslant 0,\\ & -1\leqslant d_1\leqslant 1,\\ & -1\leqslant d_2\leqslant 1,\end{aligned}$$

因此线性规划子问题化简为

$$\begin{aligned} \min \quad & -4d_2-6d_2, \\ \text{s.t.} \quad & 0\leqslant d_1\leqslant 1, \\ & 0\leqslant d_2\leqslant 1. \end{aligned} \tag{10.1.23}$$

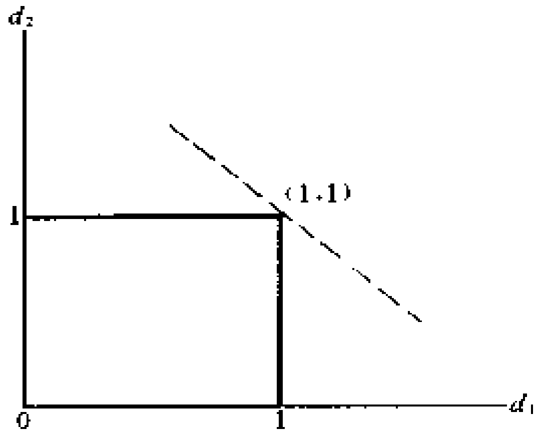


图 10.1.1

用图解法求解问题（10.1.23）（见图 10.1.1），得到 $d_1=1$，$d_2=1$，即 $\boldsymbol{d}^{(1)}=(1,1)^T$，其目标值为 −10．所以 $\boldsymbol{d}^{(1)}$是可行下降方向．

下面作一维搜索．考虑目标函数

$$\begin{aligned} \phi(\alpha) &= f(\boldsymbol{x}^{(1)}+\alpha\boldsymbol{d}^{(1)}) \\ &= 2(0+\alpha)^2-2(0+\alpha)(0+\alpha) \\ &\quad +2(0+\alpha)^2-4(0+\alpha)-6(0+\alpha) \\ &= 2\alpha^2-10\alpha, \end{aligned}$$

再考虑约束条件，由于

$$\boldsymbol{a}_1^T\boldsymbol{d}^{(1)}=2>0, \qquad \boldsymbol{a}_2^T\boldsymbol{d}^{(1)}=6>0,$$

因此

$$\alpha_{\max}=\min\left\{\frac{2-0}{2},\ \frac{5-0}{6}\right\}=\frac{5}{6}.$$

求解一维问题

$$\begin{aligned} \min \quad & \phi(\alpha)=2\alpha^2-10\alpha, \\ \text{s.t.} \quad & 0\leqslant\alpha\leqslant\frac{5}{6}, \end{aligned}$$

得到 $\alpha_1=\frac{5}{6}$．置

$$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\alpha_1\boldsymbol{d}^{(1)}=(0,0)^T+\frac{5}{6}(1,1)^T=\left(\frac{5}{6},\frac{5}{6}\right)^T.$$

再进行第二轮计算．

在点 $\boldsymbol{x}^{(2)}=\left(\frac{5}{6},\frac{5}{6}\right)^T$ 处，梯度 $\nabla f(\boldsymbol{x}^{(2)})=\left(-\frac{7}{3},-\frac{13}{3}\right)^T$. 有效约束指标集为 $I(\boldsymbol{x}^{(2)})=\{2\}$. 相应的线性规划子问题为

$$\begin{aligned}\min\quad & -\frac{7}{3}d_1-\frac{13}{3}d_2,\\ \text{s.t.}\quad & d_1+5d_2\leqslant 0,\\ & -1\leqslant d_1\leqslant 1,\\ & -1\leqslant d_2\leqslant 1,\end{aligned}$$

用图解法求解（见图 10.1.2）得到 $\boldsymbol{d}^{(2)}=\left(1,\ -\frac{1}{5}\right)^T$. 其目标值为 $-\frac{22}{15}$. 作一维搜索，考虑目标函数

$$\phi(\alpha)=f(\boldsymbol{x}^{(2)}+\alpha\boldsymbol{d}^{(2)})=\frac{62}{25}\alpha^2-\frac{22}{15}\alpha-\frac{125}{8}$$

和约束条件

$$\alpha_{\max}=\min\left\{\frac{2-\frac{5}{6}-\frac{5}{6}}{\frac{4}{5}},\ \frac{0+\frac{5}{6}}{\frac{1}{5}}\right\}=\min\left\{\frac{1/3}{4/5},\ \frac{5/6}{1/5}\right\}=\frac{5}{12},$$

求解一维问题

$$\begin{aligned}\min\quad & \phi(\alpha)=\frac{62}{25}\alpha^2-\frac{22}{15}\alpha-\frac{125}{8},\\ \text{s.t.}\quad & 0\leqslant\alpha\leqslant\frac{5}{12},\end{aligned}$$

得到 $\alpha_2=\frac{55}{186}$. 置

$$\boldsymbol{x}^{(3)}=\boldsymbol{x}^{(2)}+\alpha_2\boldsymbol{d}^{(2)}=\left(\frac{35}{31},\ \frac{24}{31}\right)^T.$$

再进行第三轮计算.

在点 $\boldsymbol{x}^{(3)}=\left(\frac{35}{31},\frac{24}{31}\right)^T$ 处，梯度 $\nabla f(\boldsymbol{x}^{(3)})=$

$\left(-\frac{32}{31},-\frac{160}{31}\right)^T$，有效约束指标集仍为 $I(x^{(2)})=\{2\}$，相应的线性规划子问题为

$$\begin{aligned}\min\quad & -\frac{32}{31}d_1-\frac{160}{31}d_2,\\ \text{s.t.}\quad & d_1+5d_2\leqslant 0,\\ & -1\leqslant d_1\leqslant 1,\\ & -1\leqslant d_2\leqslant 1,\end{aligned}$$

用图解法求解（见图 10.1.3）得到 $d^{(3)}=\left(1,\ -\frac{1}{5}\right)^T$，其目标值为 0. 因此 $x^{(3)}$ 是 K-T 点，由于目标函数是凸的，因此 $x^{(3)}$ 是最优解.

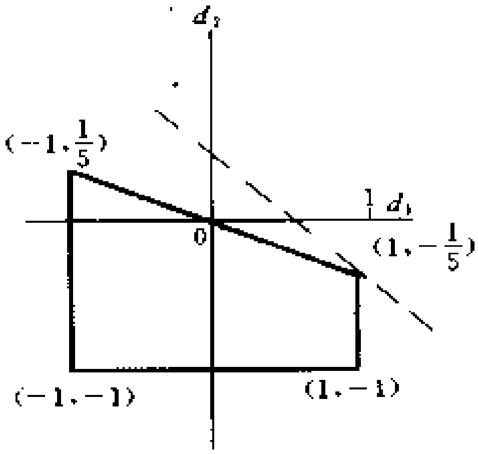


图 10.1.2

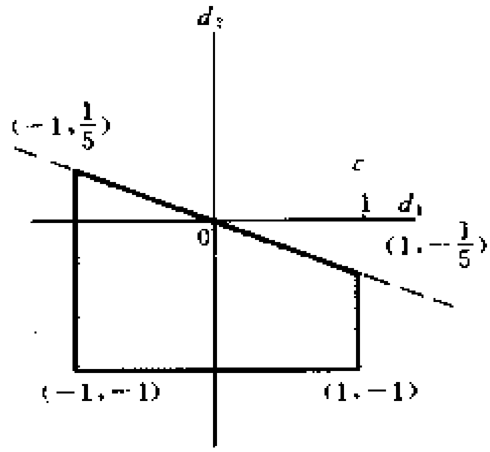


图 10.1.3

将上述计算结果列表如下：

表 10.1.1　例 10.1.1 的计算结果

k	$x^{(k)}$	$f(x^{(k)})$	$\nabla f(x^{(k)})$	$I(x^{(k)})$
1	$(0,0)^T$	0	$(-4,-6)^T$	$\{3,4\}$
2	$\left(\frac{5}{6},\frac{5}{6}\right)^T$	-6.94	$\left(-\frac{7}{3},-\frac{13}{3}\right)^T$	$\{2\}$
3	$\left(\frac{35}{31},\frac{24}{31}\right)^T$	-7.16	$\left(-\frac{32}{31}-\frac{160}{31}\right)^T$	$\{2\}$

续表

k	$d^{(k)}$	$\nabla f(x^{(k)})^T d^{(k)}$	α_{max}	α_k
1	$(1,\ 1)^T$	-10	$\frac{5}{6}$	$\frac{5}{6}$
2	$\left(1,\ -\frac{1}{5}\right)^T$	$-\frac{22}{15}$	$\frac{5}{12}$	$\frac{55}{186}$
3	$\left(1,\ -\frac{1}{5}\right)^T$	0		

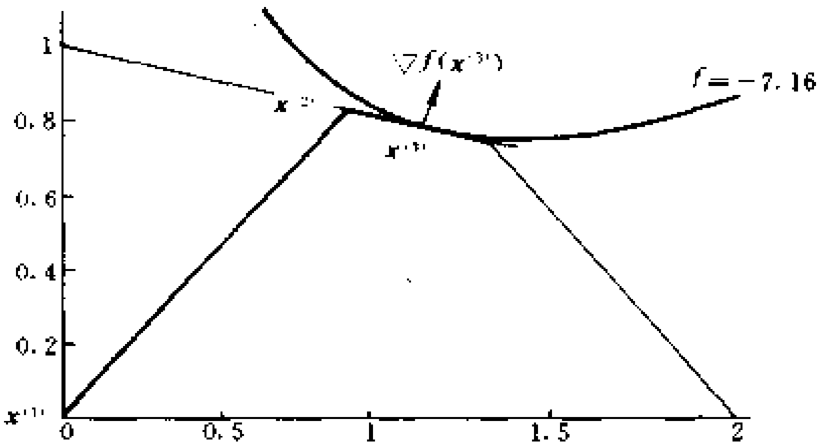


图 10.1.4

§10.2 投影梯度法

对于无约束问题，任取一点，若其梯度不为 0，则沿负梯度方向前进，总可以找到一个新的使函数值下降的点，这就是最速下降法．对于约束问题，如果再沿负梯度方向前进，可能是不可行的，因此需将负梯度方向投影到可行方向上去，这就是**投影梯度法**（Projected Gradient Method）的基本思想．投影梯度法是可行方向法的一种．

10.2.1 投影矩阵

定义 10.2.1 若 $n\times n$ 阶矩阵 P 满足

$$P^T=P \text{ 和 } P^2=P, \tag{10.2.1}$$

则称 P 为**投影矩阵**（Projection Matrix）.

定理 10.2.1 设 P 是 $n\times n$ 阶矩阵，则下述论断成立：

(1) 若 P 是投影矩阵，则 P 为半正定矩阵；

(2) P 是投影矩阵的充分必要条件是：$I-P$ 为投影矩阵.

(3) 设 P 是投影矩阵. 令 $Q=I-P$. 记

$$L=\{Px\,|\,x\in R^n\} \text{ 和 } L^{\perp}=\{Qx\,|\,x\in R^n\}, \tag{10.2.2}$$

证明 L 和 $L^{\perp}$ 是正交的线性空间，且任一点 $x\in R^n$，可唯一地表示成 $p+q$，其中 $p\in L$，$q\in L^{\perp}$.

证明

(1) 设 P 是投影矩阵，则有

$$x^TPx=x^TP^TPx=(Px)^T(Px)=\|Px\|^2\geqslant 0, \quad \forall x\in R^n, \tag{10.2.3}$$

所以 P 是半正定矩阵.

(2) 令 $Q=I-P$，只需证明若 P 是投影矩阵，则 Q 也是投影矩阵.

因为 P 是投影矩阵，则 P 是对称等幂矩阵，因此有

$$\begin{aligned} Q^T&=(I-P)^T=(I-P)=Q, \\ Q^2&=(I-P)(I-P)=I-2P+P^2=I-P=Q. \end{aligned} \tag{10.2.4}$$

(3) 显然 L 与 $L^{\perp}$ 是线性空间，且有

$$P^TQ=P^T(I-P)=P-P^2=0, \tag{10.2.5}$$

因此，当 $u=Px\in L$，$v=Qy\in L^{\perp}$ 时，有

$$u^Tv=x^TP^TQy=0, \tag{10.2.6}$$

因此 L 与 $L^{\perp}$ 相互垂直. 设 $x\in R^n$ 为任意一点，有

$$x=Ix=(P+Q)x=Px+Qx=p+q, \tag{10.2.7}$$

其中 $p=Px$，$q=Qx$. 下面证明表达表是唯一的. 设

$$x = p' + q',$$

所以

$$p' + q' = p + q, \tag{10.2.8}$$

即

$$p' - q = q - q', \tag{10.2.9}$$

由于 p，$p' \in L$，所以 $p' - p \in L$，同理，$q - q' \in L^{\perp}$. 由 (10.2.9) 式得到 $p' - p \in L^{\perp}$ 和 $q - q' \in L$. 因此有

$$\| p' - p \|^2 = (p' - p)^T (p' - p) = 0, \tag{10.2.10}$$

即 $p' = p$. 同理可得 $q = q'$.

10.2.2 投影梯度法

定理 10.2.2 设 $\bar{x}$ 是约束问题 (10.0.1) 的可行点，且 $f(x)$ 具有连续的一阶偏导数，若 P 是投影矩阵，且 $P\nabla f(\bar{x}) \neq 0$，则

$$d = -P\nabla f(\bar{x}) \tag{10.2.11}$$

是 $\bar{x}$ 处的下降方向. 此外，若

$$N = (a_{i_1}, a_{i_2}, \cdots, a_{i_r}) \tag{10.2.12}$$

为列满秩矩阵，其中 $i_j \in E \cup I(\bar{x})$，令 P 具有如下形式

$$P = I - N(N^TN)^{-1}N^T, \tag{10.2.13}$$

则 d 是 $\bar{x}$ 处的可行下降方向.

证明 因为 $P\nabla f(\bar{x}) \neq 0$，则有

$$\begin{aligned} \nabla f(\bar{x})^T d &= -\nabla f(\bar{x})^T P \nabla f(\bar{x}) \\ &= -\nabla f(\bar{x})^T P^T P \nabla f(\bar{x}) \\ &= -\| P\nabla f(\bar{x}) \|^2 < 0, \end{aligned} \tag{10.2.14}$$

因此 d 是下降方向.

若 P 具有 (10.2.13) 式的形式，则有

$$N^TP = N^T - N^TN(N^TN)^{-1}N^T = 0, \tag{10.2.15}$$

因此有

$$a_i^T d = - a_i^T P \nabla f(\bar{x}) = 0, i \in E \cup I(\bar{x}), \tag{10.2.16}$$

d 是可行方向. 因此 d 是可行下降方向.

由定理 10.2.2 可知，在 $P\nabla f(\bar{x}) \neq 0$ 的条件下，可以得到 $\bar{x}$ 处的可行下降方向，可类似于 § 10.1 中的一维搜索方法，得到下一个可行点. 当 $P\nabla f(\bar{x}) = 0$ 时，情况又将如何呢? 请看下一个定理.

定理 10.2.3 设 $\bar{x}$ 是约束问题（10.0.1）的可行点，且 $f(x)$ 具有连续的一阶偏导数，若 P 是由（10.2.12）式和（10.2.13）式确定的投影矩阵，并满足

$$P\nabla f(\bar{x}) = 0, \tag{10.2.17}$$

记

$$\lambda = -(N^T N)^{-1} N^T \nabla f(\bar{x}), \tag{10.2.18}$$

若 $\lambda_i \geqslant 0$，$i \in I(\bar{x})$，则 $\bar{x}$ 是 K-T 点. 若存在 $q \in I(\bar{x})$，使得 $\lambda_q < 0$，并记 $\bar{N}$ 为矩阵 N 中去掉 λ_q 对应的列后得到的矩阵，并令 $\bar{P} = I - \bar{N}(\bar{N}^T \bar{N})^{-1} \bar{N}^T$，$d = -\bar{P}\nabla f(\bar{x})$，则 d 是 $\bar{x}$ 处的可行下降方向.

证明 由（10.2.13）式、（10.2.17）式和（10.2.18）式，有

$$\begin{aligned} 0 &= P\nabla f(\bar{x}) \\ &= (I - N(N^T N)^{-1} N^T)\nabla f(\bar{x}) \\ &= \nabla f(\bar{x}) - N(N^T N)^{-1} N^T \nabla f(\bar{x}) \\ &= \nabla f(\bar{x}) + N\lambda, \end{aligned} \tag{10.2.19}$$

即

$$\nabla f(\bar{x}) + \sum_{i \in E \cup I(\bar{x})} \lambda_i a_i = 0, \tag{10.2.20}$$

若 $\lambda_i \geqslant 0, i \in I(\bar{x})$，则令 $\lambda_i = 0, i \notin I(\bar{x})$，所以 $\bar{x}$ 是 K-T 点.

若存在 $q \in I(\bar{x})$，使得 $\lambda_q < 0$，证明 $d = -\bar{P}\nabla f(\bar{x})$ 是可行下

降方向.

首先证明 $\boldsymbol{d}\neq 0$. 若 $\boldsymbol{d}=0$, 则由 $\overline{P}\nabla f(\overline{\boldsymbol{x}})=0$,类似地导出,存在

$$\boldsymbol{\mu}=-(\overline{N}^T\overline{N})^{-1}\overline{N}^T\nabla f(\overline{\boldsymbol{x}}),$$

使得

$$\nabla f(\overline{\boldsymbol{x}})+\sum_{\substack{i\in E\cup I(\overline{\boldsymbol{x}})\\ i\neq q}}\mu_i\boldsymbol{a}_i=0. \tag{10.2.21}$$

用 (10.2.20) 式减 (10.2.21) 式, 得到

$$\lambda_q\boldsymbol{a}_q+\sum_{\substack{i\in E\cup I(\overline{\boldsymbol{x}})\\ i\neq q}}(\lambda_i-\mu_i)\boldsymbol{a}_i=0. \tag{10.2.22}$$

由于 $\lambda_q<0$, 这与 N 为列满秩矩阵矛盾.

再证明 d 是可行方向, 类似于定理 10.2.2 的证明可知

$$\boldsymbol{a}_i^T\boldsymbol{d}=0,\ i\in E\cup I(\overline{\boldsymbol{x}}),\ i\neq q, \tag{10.2.23}$$

并由 (10.2.20) 式和 (10.2.23) 式, 有

$$\begin{aligned}\boldsymbol{a}_q^T\boldsymbol{d}&=\left(-\frac{1}{\lambda_q}\nabla f(\overline{\boldsymbol{x}})-\frac{1}{\lambda_q}\sum_{\substack{i\in E\cup I(\overline{x})\\ i\neq q}}\lambda_i\boldsymbol{a}_i\right)^T\boldsymbol{d}\\&=-\frac{1}{\lambda_q}\nabla f(\overline{\boldsymbol{x}})^T\boldsymbol{d}\\&=\frac{1}{\lambda_q}\nabla f(\overline{\boldsymbol{x}})^T\overline{P}\nabla f(\overline{\boldsymbol{x}})\\&=\frac{1}{\lambda_q}\|\overline{P}\nabla f(\overline{\boldsymbol{x}})\|^2<0,\end{aligned} \tag{10.2.24}$$

所以 $\boldsymbol{d}$ 是 $\overline{\boldsymbol{x}}$ 处的可行方向. 由定理 10.2.2 可知, $\boldsymbol{d}\neq 0$ 是下降方向.

由定理 10.2.2 和定理 10.2.3, 可以给出投影梯度法的具体算法. 由于投影梯度法是 J. B. Rosen 在 1960 年提出来的, 因此也称为 Rosen 投影梯度法.

算法 10.2.1 (Rosen 投影梯度法)

（1）取初始可行点 $\boldsymbol{x}^{(1)}$，即 $\boldsymbol{x}^{(1)}$满足

$$\boldsymbol{a}_i^T\boldsymbol{x}^{(1)}-b_i=0,\ i\in E,$$

$$\boldsymbol{a}_i^T\boldsymbol{x}^{(1)}-b_i\leqslant 0,\ i\in I,$$

置 $k=1$.

（2）确定 $x^{(k)}$处的有效约束指标集

$$I(\boldsymbol{x}^{(k)})=\{i\mid \boldsymbol{a}_i^T\boldsymbol{x}^{(k)}-b_i=0,\ i\in I\}.$$

（3）若 $l=0$（无等式约束）且 I（$\boldsymbol{x}^{(k)}$）$=\phi$，则令

$$\boldsymbol{d}^{(k)}=-\nabla f(\boldsymbol{x}^{(k)});$$

否则令

$$N^{(k)}=(\boldsymbol{a}_{i_1},\ \boldsymbol{a}_{i_2},\ \cdots,\ \boldsymbol{a}_{i_r}),\quad i_j\in E\cup I(\boldsymbol{x}^{(k)}),$$

$$P^{(k)}=I-N^{(k)}(N^{(k)T}N^{(k)})^{-1}N^{(k)T},$$

$$\boldsymbol{d}^{(k)}=-P^{(k)}\nabla f(\boldsymbol{x}^{(k)}).$$

（4）若 $\boldsymbol{d}^{(k)}=0$，则做以下工作：

若 $l=0$ 且 $I(\boldsymbol{x}^{(k)})=\phi$，则停止计算（$\boldsymbol{x}^{(k)}$为 K-T 点）；否则计算

$$\boldsymbol{\lambda}^{(k)}=-(N^{(k)T}N^{(k)})^{-1}N^{(k)T}\nabla f(\boldsymbol{x}^{(k)}).$$

若 $\lambda_i^{(k)}\geqslant 0,\ i\in I(\boldsymbol{x}^{(k)})$，则停止计算（$\boldsymbol{x}^{(k)}$为 K-T 点）；否则令

$$\lambda_q^{(k)}=\min\{\lambda_i^{(k)}\mid i\in I(\boldsymbol{x}^{(k)})\},$$

置

$$N^{(k)}=(N^{(k)}\text{中去掉}\ \lambda_q^{(k)}\ \text{对应的列}\ \boldsymbol{a}_q),$$

$$P^{(k)}=I-N^{(k)}\ (N^{(k)T}N^{(k)})^{-1}N^{(k)T},$$

$$\boldsymbol{d}^{(k)}=-P^{(k)}\nabla f(\boldsymbol{x}^{(k)}).$$

（5）（$\boldsymbol{d}^{(k)}\neq 0$）求解一维问题

$$\begin{aligned}&\min\quad \phi(\alpha)=f(\boldsymbol{x}^{(k)}+\alpha\boldsymbol{d}^{(k)}),\\&\text{s.t.}\quad 0\leqslant\alpha\leqslant\alpha_{\max},\end{aligned}$$

其中

$$\alpha_{\max}=\begin{cases}\min\left\{\dfrac{b_i-\boldsymbol{a}_i^T\boldsymbol{x}^{(k)}}{\boldsymbol{a}_i^T\boldsymbol{d}^{(k)}}\,\middle|\,\boldsymbol{a}_i^T\boldsymbol{d}^{(k)}>0,\ i\notin I(\boldsymbol{x}^{(k)})\right\},\\ +\infty,\quad \{\boldsymbol{a}_i^T\boldsymbol{d}^{(k)}>0, i\notin I(\boldsymbol{x}^{(k)})\}=\phi,\end{cases}$$

得到 α_k，置

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}+\alpha_k\boldsymbol{d}^{(k)}.$$

(6) 置 $k=k+1$，转 (2).

在实际计算中，算法第 (4) 步 $\boldsymbol{d}^{(k)}=0$ 可改为 $\|\boldsymbol{d}^{(k)}\|\leqslant\epsilon$，其中 ϵ 是预先指定的精度.

【例 10.2.1】 用梯度投影法求解线性约束问题

$$\begin{aligned}\min\quad & f(\boldsymbol{x})=2x_1^2-2x_1x_2+2x_2^2-4x_1-6x_2,\\ \text{s.t.}\quad & \boldsymbol{a}_1^T\boldsymbol{x}-b_1=x_1+x_2-2\leqslant 0,\\ & \boldsymbol{a}_2^T\boldsymbol{x}-b_2=x_1+5x_2-5\leqslant 0,\\ & \boldsymbol{a}_3^T\boldsymbol{x}-b_3=-x_1\leqslant 0,\\ & \boldsymbol{a}_4^T\boldsymbol{x}-b_4=-x_2\leqslant 0,\end{aligned}$$

取初始点 $\boldsymbol{x}^{(1)}=(0,\ 0)^T$.

解

$$\nabla f(\boldsymbol{x})=(4x_1-2x_2-4,\ -2x_1+4x_2-6)^T,$$

当 $\boldsymbol{x}^{(1)}=(0,\ 0)^T$ 时，$\nabla f(\boldsymbol{x}^{(1)})=(-4,\ -6)^T$. 有效约束指标集为 $I(\boldsymbol{x}^{(1)})=\{3,\ 4\}$，则

$$N^{(1)}=\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix},$$

$$P^{(1)}=I-N^{(1)}(N^{(1)T}N^{(1)})^{-1}N^{(1)T}=I-I=0.$$

计算时注意到，$N^{(1)}$是非奇异的，有 $N^{(1)}(N^{(1)T}N^{(1)})^{-1}N^{(1)T}=I$，因此

$$P^{(1)}\nabla f(\boldsymbol{x}^{(1)})=(0,0)^T,$$

计算

$$\lambda^{(1)} = -(N^{(1)T}N^{(1)})^{-1}N^{(1)T}\nabla f(x^{(1)})$$
$$= -N^{(1)-1}\nabla f(x^{(1)}) = (-4,-6)^T,$$
$$\lambda_4^{(1)} = \min\{\lambda_i^{(1)} \mid i \in I^{(1)}\} = -6.$$

去掉第 4 个约束，得到

$$N^{(1)} = \begin{pmatrix} -1 \\ 0 \end{pmatrix},\ N^{(1)T}N^{(1)} = 1,$$

所以

$$P^{(1)} = I - N^{(1)}(N^{(1)T}N^{(1)})^{-1}N^{(1)T}$$
$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \end{pmatrix}(-1,0) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$
$$d^{(1)} = -P^{(1)}\nabla f(x^{(1)}) = -\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -4 \\ -6 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}.$$

下面作一维搜索. 考虑目标函数

$$\phi(\alpha) = f(x^{(1)} + \alpha d^{(1)}) = 72\alpha^2 - 36\alpha,$$

再考虑约束条件，由于 $a_1^T d^{(1)} = 6 > 0$，$a_2^T d^{(1)} = 30 > 0$，因此

$$\alpha_{\max} = \min\left\{\frac{2-0}{6},\ \frac{5-0}{30}\right\} = \frac{1}{6},$$

求解一维问题

$$\begin{aligned} \min \quad & \phi(\alpha) = 72\alpha^2 - 36\alpha, \\ \text{s.t.} \quad & 0 \leqslant \alpha \leqslant \frac{1}{6}, \end{aligned}$$

得到 $\alpha_1 = \frac{1}{6}$. 置

$$x^{(2)} = x^{(1)} + \alpha_1 d^{(1)} = (0,0)^T + \frac{1}{6}(0,6)^T = (0,1)^T.$$

再进行第二轮计算.

在点 $x^{(2)} = (0,1)^T$处，梯度$\nabla f(x^{(2)}) = (-6,-2)^T$. 有效约束指标集为 $I(x^{(2)}) = \{2,3\}$，则

$$N^{(2)}=\begin{pmatrix}1 & -1\\ 5 & 0\end{pmatrix},$$

$$P^{(2)}=I-N^{(2)}(N^{(2)T}N^{(2)})^{-1}N^{(2)T}=I-I=0,$$

$$P^{(2)}\nabla f(\boldsymbol{x}^{(2)})=(0,0)^T,$$

计算乘子

$$\begin{aligned}\boldsymbol{\lambda}^{(2)}&=-(N^{(2)T}N^{(2)})^{-1}N^{(2)T}\nabla f(\boldsymbol{x}^{(2)})\\&=-N^{(2)-1}\nabla f(\boldsymbol{x}^{(2)})\\&=-\begin{pmatrix}0 & \frac{1}{5}\\ -1 & \frac{1}{5}\end{pmatrix}\begin{pmatrix}-6,\\ -2\end{pmatrix}=\begin{pmatrix}\frac{2}{5}\\ -\frac{28}{5}\end{pmatrix},\end{aligned}$$

$$\lambda_3^{(2)}=-\frac{28}{5}.$$

去掉第 3 个约束，得到

$$N^{(2)}=\begin{pmatrix}1\\ 5\end{pmatrix},\ N^{(1)T}N^{(1)}=26,$$

所以

$$\begin{aligned}P^{(2)}&=I-N^{(2)}(N^{(2)T}N^{(2)})^{-1}N^{(2)T}\\&=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}-\frac{1}{26}\begin{pmatrix}1\\ 5\end{pmatrix}(1,5)=\begin{pmatrix}\frac{25}{26} & -\frac{5}{26}\\ -\frac{5}{26} & \frac{1}{26}\end{pmatrix},\end{aligned}$$

$$\boldsymbol{d}^{(2)}=-P^{(2)}\nabla f(\boldsymbol{x}^{(2)})=-\begin{pmatrix}\frac{25}{26} & -\frac{5}{26}\\ -\frac{5}{26} & \frac{1}{26}\end{pmatrix}\begin{pmatrix}-6\\ -2\end{pmatrix}=\begin{pmatrix}\frac{70}{13}\\ -\frac{14}{13}\end{pmatrix}.$$

由于一维搜索与模长无关，因此，为了计算方便，取

$$\boldsymbol{d}^{(2)}=(5,\ -1)^T.$$

下面作一维搜索．考虑目标函数

$$\phi(\alpha)=f(\boldsymbol{x}^{(2)}+\alpha\boldsymbol{d}^{(2)})=62\alpha^2-28\alpha-4,$$

再考虑约束条件，由于 $a_1^T d^{(2)}=4>0$、$a_4^T d^{(2)}=1>0$，因此

$$\alpha_{\max}=\min\left\{\frac{2-1}{4},\ \frac{0+1}{1}\right\}=\frac{1}{4},$$

求解一维问题

$$\min\quad \phi(\alpha)=62\alpha^2-28\alpha-4,$$

$$\text{s.t.}\quad 0\leqslant\alpha\leqslant\frac{1}{4},$$

得到 $\alpha_1=\frac{7}{31}$. 置

$$\begin{aligned}x^{(3)}&=x^{(2)}+\alpha_2 d^{(2)}\\&=(0,\ 1)^T+\frac{7}{31}(5,\ -1)^T\\&=\left(\frac{35}{31},\ \frac{24}{31}\right)^T.\end{aligned}$$

最后进行第三轮计算.

在点 $x^{(3)}$处，梯度$\nabla f(d^{(3)})=\left(-\frac{32}{31},-\frac{160}{31}\right)^T$,有效约束指标集为 $I(d^{(3)})=\{2\}$,于是

$$N^{(3)}=\begin{pmatrix}1\\5\end{pmatrix},\quad P^{(3)}=\begin{bmatrix}\frac{25}{26}&-\frac{5}{26}\\-\frac{5}{26}&\frac{1}{26}\end{bmatrix},\quad P^{(3)}\nabla f(x^{(2)})=\begin{pmatrix}0\\0\end{pmatrix},$$

计算乘子

$$\lambda^{(3)}=-\frac{1}{26}(1,\ 5)\begin{bmatrix}-\frac{32}{31}\\-\frac{160}{31}\end{bmatrix}=\frac{32}{31},$$

所以 $x^{(3)}=\left(\frac{35}{31},\ \frac{24}{31}\right)^T$ 为最优解，相应的乘子为 $\lambda^*=\left(0,\ \frac{32}{31},\ 0,\ 0\right)^T$.

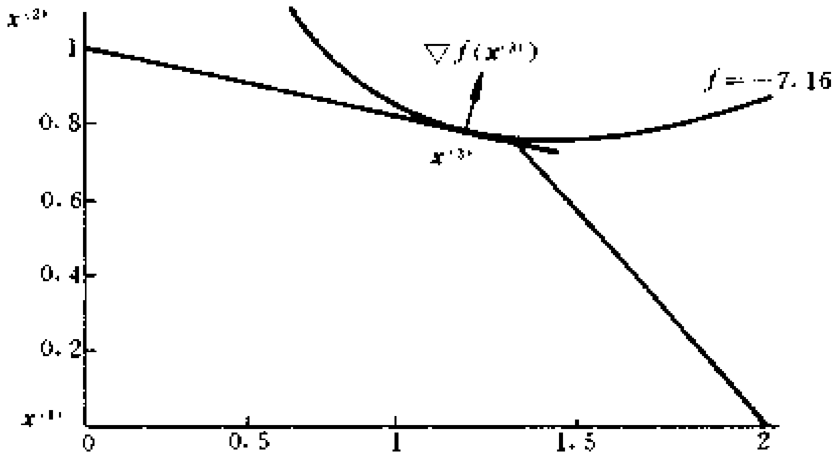


图 10.2.1

将上述计算结果列表如下：

表 10.2.1　例 10.2.1 的计算结果

k	$x^{(k)}$	$f(x^{(k)})$	$\nabla f(x^{(k)})$	$I(x^{(k)})$	$N^{(k)}$
1	$(0,0)$	0	$(-4,-6)$	$\{3,4\}$ $\{3\}$	$\begin{pmatrix}-1 & 0\\0 & -1\end{pmatrix}$ $\begin{pmatrix}-1\\0\end{pmatrix}$
2	$(0,1)$	-4	$(-6,-2)$	$\{2,3\}$ $\{2\}$	$\begin{pmatrix}1 & -1\\5 & 0\end{pmatrix}$ $\begin{pmatrix}1\\5\end{pmatrix}$
3	$\left(\frac{35}{31},\frac{24}{31}\right)$	-7.16	$\left(-\frac{32}{31}-\frac{160}{31}\right)$	$\{2\}$	$\begin{pmatrix}1\\5\end{pmatrix}$

k	$P^{(k)}$	$d^{(k)}$	$\lambda^{(k)}$	α_{max}	α_k
1	$\begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}$ $\begin{pmatrix}0 & 0\\0 & 1\end{pmatrix}$	$(0,\ 0)$ $(0,\ 6)$	$(-4,\ -6)$ —	— $\frac{1}{6}$	— $\frac{1}{6}$

续表

k	$P^{(k)}$	$d^{(k)}$	$\lambda^{(k)}$	$\alpha_{\max}$	α_k
2	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	(0, 0)	$\left(\frac{2}{5}, -\frac{28}{5}\right)$	—	—
	$\begin{bmatrix} \frac{25}{26} & -\frac{5}{26} \\ -\frac{5}{26} & \frac{1}{26} \end{bmatrix}$	(5, −1)	—	$\frac{1}{4}$	$\frac{7}{31}$
3	$\begin{bmatrix} \frac{25}{26} & -\frac{5}{26} \\ -\frac{5}{26} & \frac{1}{26} \end{bmatrix}$	(0, 0)	$\frac{32}{31}$		

10.2.3 投影矩阵 $P^{(k)}$和 $(N^{(k)T}N^{(k)})^{-1}$的计算

投影梯度法在求解线性约束问题的过程中，会出现两种情况：第一种是当 $\alpha_k=\alpha_{\max}$时，在 $x^{(k+1)}$处增加一个有效约束，因此需要重新计算投影矩阵和相应的乘子向量；第二种情况是当 $P\nabla f(x^{(k)})=0$,且乘子 $\lambda_q^{(k)}<0$，需要去掉第 q 个约束，也需要计算投影矩阵．这两种情况有一个特点：就是在有效约束集合中增加或减少一个约束，如果按投影矩阵公式直接计算的话，计算量大．那么一种自然的想法，就是能否找到减少计算量的计算方法．

首先讨论分块矩阵的求逆公式．设

$$A=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},\quad A^{-1}=\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},\tag{10.2.25}$$

其中 A_{11}、A_{22}、B_{11}、B_{22}为方阵，且 A_{11}与 B_{11}、A_{22}与 B_{22}的阶数相同．为了简单起见，对于不同阶和单位阵均记为 I．由于 $AA^{-1}=I$，故有

$$A_{11}B_{12}+A_{12}B_{22}=0,\tag{10.2.26}$$

$$A_{21}B_{12}+A_{22}B_{22}=I. \qquad (10.2.27)$$

由（10.2.26）式得到

$$B_{12}=-A_{11}^{-1}A_{12}B_{22}. \qquad (10.2.28)$$

将（10.2.28）式代入（10.2.27）式，得到

$$-A_{21}A_{11}^{-1}A_{12}B_{22}+A_{22}B_{22}=I,$$

化简得

$$(A_{22}-A_{21}A_{11}^{-1}A_{12})\ B_{22}=I, \qquad (10.2.29)$$

记

$$A_0=(A_{22}-A_{21}A_{11}^{-1}A_{12}), \qquad (10.2.30)$$

由（10.2.29）式得到

$$B_{22}=A_0^{-1}, \qquad (10.2.31)$$

代入（10.2.28）式得到

$$B_{12}=-A_{11}^{-1}A_{12}A_0^{-1}. \qquad (10.2.32)$$

同样，由 $A^{-1}A=I$，得到

$$B_{11}A_{11}+B_{12}A_{21}=I, \qquad (10.2.33)$$

$$B_{21}A_{11}+B_{22}A_{21}=0, \qquad (10.2.34)$$

由（10.2.34）式和（10.2.31）式得到

$$B_{21}=-B_{22}A_{21}A_{11}^{-1}=-A_0^{-1}A_{21}A_{11}^{-1}, \qquad (10.2.35)$$

由（10.2.33）式和（10.2.32）式得到

$$B_{11}=A_{11}^{-1}-B_{12}A_{21}A_{11}^{-1}=A_{11}^{-1}+A_{11}^{-1}A_{12}A_0^{-1}A_{21}A_{11}^{-1}. \qquad (10.2.36)$$

因此，若已知矩阵 A_{11} 的逆，就可计算出矩阵 A 的逆矩阵．这样做比直接求 A^{-1} 计算量少得多．

反之，若已知矩阵 A^{-1}，由（10.2.33）式和（10.2.34）式得到

$$B_{11}=A_{11}^{-1}-B_{12}A_{21}A_{11}^{-1}, \qquad (10.2.37)$$

$$B_{21}+B_{22}A_{21}A_{11}^{-1}=0. \qquad (10.2.38)$$

将（10.2.38）式代入（10.2.37）式，得到

$$B_{11}=A_{11}^{-1}+B_{12}B_{22}^{-1}B_{21},$$

即

$$A_{11}^{-1}=B_{11}-B_{12}B_{22}^{-1}B_{21}. \tag{10.2.39}$$

下面讨论投影矩阵 P 和矩阵 $(N^TN)^{-1}$的计算.

（1）不妨设

$$N=(\boldsymbol{a}_1,\ \boldsymbol{a}_2\cdots,\ \boldsymbol{a}_{p-1}), \tag{10.2.40}$$

$$P=I-N(N^TN)^{-1}N^T, \tag{10.2.41}$$

现增加一个约束 p，设

$$\widetilde{N}=[N,\ \boldsymbol{a}_p] \tag{10.2.42}$$

$$\widetilde{P}=I-\widetilde{N}(\widetilde{N}^T\widetilde{N})^{-1}\widetilde{N}^T. \tag{10.2.43}$$

下面求出计算 $\widetilde{P}$ 和 $(\widetilde{N}^T\widetilde{N})^{-1}$的计算公式，令

$$A=\widetilde{N}^T\widetilde{N}=\begin{pmatrix}N^T\\ \boldsymbol{a}_p^T\end{pmatrix}(N,\boldsymbol{a}_p)=\begin{pmatrix}N^TN & N^T\boldsymbol{a}_p\\ \boldsymbol{a}_p^TN & \boldsymbol{a}_p^T\boldsymbol{a}_p\end{pmatrix}$$

$$=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}, \tag{10.2.44}$$

和

$$(\widetilde{N}^T\widetilde{N})^{-1}=B=\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix}, \tag{10.2.45}$$

由（10.2.30）式～（10.2.36）式，得到

$$\begin{aligned}A_0&=(A_{22}-A_{21}A_{11}^{-1}A_{12})\\&=a_p^Ta_p-a_p^TN(N^TN)^{-1}N^Ta_p\\&=\boldsymbol{a}_p^T(I-N(N^TN)^{-1}N^T)\boldsymbol{a}_p\\&=\boldsymbol{a}_p^T\boldsymbol{P}\boldsymbol{a}_p,\end{aligned} \tag{10.2.46}$$

因此，

$$B_{22}=A_0^{-1}=(\boldsymbol{a}_p^T\boldsymbol{P}\boldsymbol{a}_p)^{-1}, \tag{10.2.47}$$

$$B_{12} = -A_{11}^{-1}A_{12}A_0^{-1} = -\frac{(N^TN)^{-1}N^T\boldsymbol{a}_p}{\boldsymbol{a}_p^TP\boldsymbol{a}_p}, \tag{10.2.48}$$

$$B_{21} = -A_0^{-1}A_{21}A_{11}^{-1} = -\frac{\boldsymbol{a}_p^TN(N^TN)^{-1}}{\boldsymbol{a}_p^TP\boldsymbol{a}_p}, \tag{10.2.49}$$

$$\begin{aligned} B_{11} &= A_{11}^{-1} + A_{11}^{-1}A_{12}A_0^{-1}A_{21}A_{11}^{-1} \\ &= (N^TN)^{-1} + \frac{(N^TN)^{-1}N^T\boldsymbol{a}_p\boldsymbol{a}_p^TN(N^TN)^{-1}}{\boldsymbol{a}_p^TP\boldsymbol{a}_p}. \end{aligned} \tag{10.2.50}$$

由（10.2.47）式～（10.2.50）式可计算出 $(\tilde{N}^T\tilde{N})^{-1}$，并经整理得到

$$\begin{aligned} \tilde{P} &= I - \tilde{N}(\tilde{N}^T\tilde{N})^{-1}\tilde{N}^T \\ &= I - (N, \boldsymbol{a}_p)\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}\begin{pmatrix} N^T \\ \boldsymbol{a}_p^T \end{pmatrix} \\ &= P - \frac{P\boldsymbol{a}_p\boldsymbol{a}_p^TP}{\boldsymbol{a}_p^TP\boldsymbol{a}_p}. \end{aligned} \tag{10.2.51}$$

由（10.2.51）式可直接计算 $\tilde{P}$，其计算量相当小.

（2）当 $\lambda_p<0$ 时，$\tilde{N}$ 是从 N 中去掉 λ_q 对应的列 $\boldsymbol{a}_q$，不妨设

$$N = [\tilde{N}, \boldsymbol{a}_q],$$

此时，

$$N^TN = \begin{pmatrix} \tilde{N}^T\tilde{N} & \tilde{N}^T\boldsymbol{a}_q \\ \boldsymbol{a}_q^T\tilde{N} & \boldsymbol{a}_q^T\boldsymbol{a}_q \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

由于

$$(N^TN)^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

已知，则由（10.2.39）式得到矩阵 $\tilde{N}^T\tilde{N}$ 的计算公式，然后再计算出相应的 $\tilde{P}$.

§ 10.3 既约梯度法

既约梯度法（Reduced Gradient Method）是由 Wolfe 在 1963 年提出的一种可行方向法，因此也称为 Wolfe 既约梯度法．它的基本思想是将求解线性规划的单纯形法推广到求解非线性规划问题．

考虑线性约束问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \quad \boldsymbol{x} \in R^n, \\ \text{s.t.} \quad & A\boldsymbol{x} = \boldsymbol{b}, \\ & \boldsymbol{x} \geqslant 0, \end{aligned} \tag{10.3.1}$$

其中 $A = A_{m \times n}$，$m < n$，$\operatorname{rank}(A) = m$．进一步假设，$A$ 的任意 m 列均是线性无关的．

设 x 是可行点，不妨设

$$A = [B,\ N], \tag{10.3.2}$$

其中 $B = B_{m \times m}$是可逆矩阵．此时相应的分解为

$$\boldsymbol{x}^T = (\boldsymbol{x}_B^T,\ \boldsymbol{x}_N^T), \tag{10.3.3}$$

其中 $\boldsymbol{x}_B > 0$ 称为基变量，$\boldsymbol{x}_N \geqslant 0$ 称为非基变量．梯度$\nabla f(\boldsymbol{x})$也作相应的划分

$$\nabla f(\boldsymbol{x})^T = (\nabla_B f(\boldsymbol{x})^T, \nabla_B f(\boldsymbol{x})^T). \tag{10.3.4}$$

设 $\boldsymbol{d}$ 是约束问题（10.3.1）的可行方向，则 $\boldsymbol{d}$ 就满足

$$A\boldsymbol{d} = 0, \tag{10.3.5}$$

$$d_j \geqslant 0, \quad (\text{当 } x_j = 0 \text{ 时}) \tag{10.3.6}$$

对 $\boldsymbol{d}$ 作相应的划分，令

$$\boldsymbol{d}^T = (\boldsymbol{d}_B^T,\ \boldsymbol{d}_N^T), \tag{10.3.7}$$

由（10.3.2）式，（10.3.5）式化为

$$B\boldsymbol{d}_B + N\boldsymbol{d}_N = 0, \tag{10.3.8}$$

因此，

$$d_B = -B^{-1}Nd_N. \tag{10.3.9}$$

进一步，若 d 满足

$$\nabla f(x)^T d < 0, \tag{10.3.10}$$

则 d 是 x 处的下降方向. 将（10.3.10）式改写成

$$\nabla_B f(x)^T d_B + \nabla_N f(x)^T d_N < 0, \tag{10.3.11}$$

并将（10.3.9）式代入（10.3.11）式，得到

$$(\nabla_N f(x)^T - \nabla_B f(x)^T B^{-1}N)d_N < 0. \tag{10.3.12}$$

令

$$\begin{aligned} r^T &= (r_B^T, r_N^T) \\ &= \nabla f(x)^T - \nabla_B f(x)^T B^{-1}A \\ &= (\nabla_B f(x)^T, \nabla_N f(x)^T) - \nabla_B f(x)^T B^{-1}(B, N) \\ &= (0, \nabla_N f(x)^T - \nabla_B f(x)^T B^{-1}N), \end{aligned} \tag{10.3.13}$$

称 r 为目标函数 $f(x)$ 在 x 处的**既约梯度**(Reduced Gradient).

由上述推导可知，d 满足

$$d_B = -B^{-1}Nd_N, \tag{10.3.14}$$

$$d_j \geqslant 0, \ (\text{当 } x_j = 0) \tag{10.3.15}$$

$$r_N^T d_N < 0, \tag{10.3.16}$$

则 d 是 x 处的可行下降方向. 因此，可以按照如下方法选择 d_N. 设 d_j 是 d_N 的分量，令

$$d_j = \begin{cases} -r_j, & \text{当 } r_j \leqslant 0, \\ -x_j r_j, & \text{当 } r_j > 0, \end{cases} \quad j \in \mathcal{N}, \tag{10.3.17}$$

其中 $\mathcal{N}$ 表示非基变量指标集.

定理 10.3.1 设 x 是可行点，$f(x)$ 具有连续的一阶偏导数，令 $x = \begin{pmatrix} x_B \\ x_N \end{pmatrix}$，$x_B > 0$，$A = [B, N]$，且 B^{-1} 存在，$r^T =$

$\nabla f(\boldsymbol{x})^T - \nabla_B f(\boldsymbol{x})^T B^{-1} A$, $\boldsymbol{d} \in R^n$ 是按(10.3.17)式和(10.3.14)式构造的.如果 $\boldsymbol{d} \neq 0$,则 $\boldsymbol{d}$ 是可行下降方向.进一步,$\boldsymbol{d} = 0$ 的充分必要条件是:$\boldsymbol{x}$ 是 K-T 点.

证明 显然由(10.3.17)式构造的 d_j 满足(10.3.15)式,因此 $\boldsymbol{d}$ 是可行方向.

下面证明 $\boldsymbol{d}$ 是下降方向.即证明 $\boldsymbol{d}$ 满足(10.3.16)式.由(10.3.17)式得到

$$\begin{aligned}\boldsymbol{R}_N^T \boldsymbol{d}_N &= \sum_{j \in N} r_j d_j = \sum_{r_j \leqslant 0} r_j d_j + \sum_{r_j > 0} r_j d_j \\ &= -\sum_{r_j \leqslant 0} r_j^2 - \sum_{r_j > 0} x_j r_j^2 \leqslant 0,\end{aligned} \qquad (10.3.18)$$

因为 $\boldsymbol{d} \neq 0$, $\boldsymbol{d}_N \neq 0$,因此存在 $r_j < 0$ 或 $x_j > 0$ 且 $r_j > 0$.因此(10.3.18)式中小于号严格成立.

最后证明定理的第二部分.

约束问题(10.3.1)的局部解的 K-T 条件为,存在 $\boldsymbol{u}$、$\boldsymbol{v}$ 使得

$$\nabla f(\boldsymbol{x}) + A^T \boldsymbol{u} - \boldsymbol{v} = 0, \qquad (10.3.19)$$

$$\boldsymbol{v} \geqslant 0, \qquad (10.3.20)$$

$$\boldsymbol{v}^T \boldsymbol{x} = 0. \qquad (10.3.21)$$

对 $\boldsymbol{v}$ 作划分,令 $\boldsymbol{v}^T = (\boldsymbol{v}_B^T, \boldsymbol{v}_N^T)$,由(10.3.20)式和 $\boldsymbol{x} \geqslant 0$ 可知(10.3.21)式等价于

$$\boldsymbol{v}_B^T \boldsymbol{x}_B = 0 \text{ 和 } \boldsymbol{v}_N^T \boldsymbol{x}_N = 0. \qquad (10.3.22)$$

注意到 $\boldsymbol{x}_B > 0$,因此 $\boldsymbol{v}_B = 0$.这样(10.3.19)式化为

$$\nabla_B f(\boldsymbol{x})^T + \boldsymbol{u}^T B = 0, \qquad (10.3.23)$$

$$\nabla_N f(\boldsymbol{x})^T + \boldsymbol{u}^T N - \boldsymbol{v}_N^T = 0. \qquad (10.3.24)$$

由(10.3.23)式可知,

$$\boldsymbol{u}^T = -\nabla_B f(\boldsymbol{x})^T B^{-1}, \qquad (10.3.25)$$

将(10.3.25)式代入(10.3.24)式,得到

$$\boldsymbol{u}_N^T = \nabla_N f(\boldsymbol{x})^T - \nabla_B f(\boldsymbol{x})^T B^{-1} N = \boldsymbol{r}_N^T, \tag{10.3.26}$$

因此 K-T 条件转化为

$$\boldsymbol{r}_N \geqslant 0, \tag{10.3.27}$$

$$\boldsymbol{r}_N^T \boldsymbol{x}_N = 0. \tag{10.3.28}$$

当 $\boldsymbol{d}=0$ 时，即 $\boldsymbol{d}_N=0$，因此有

$$r_j = 0, \tag{10.3.29}$$

或者

$$x_j r_j = 0 \text{ 且 } r_j > 0, \tag{10.3.30}$$

即（10.3.27）式和（10.3.28）式成立．因此 $\boldsymbol{x}$ 是 K-T 点．

反过来，若 $\boldsymbol{x}$ 是 K-T 点，则（10.3.27）式和（10.3.28）式成立．由（10.3.17）式得 $d_j=0$．即 $\boldsymbol{d}_N=0$．因此 $\boldsymbol{d}_B=0$，从而 $\boldsymbol{d}=0$．

一维搜索

类似于§10.1 的讨论，这里讨论既约梯度法中一维搜索的搜索区间．由（10.3.17）式可知，当 $x_j=0$ 时，有 $d_j \geqslant 0$．因此不会破坏 $\boldsymbol{x}$ 的可行性．只有当 $x_j>0$ 且 $d_j<0$ 时，可行性才有可能被破坏．因此取

$$\alpha_{\max} = \min\left\{ -\frac{x_j}{d_j} \,\middle|\, d_j<0 \text{ 且 } x_j>0 \right\},$$

当 $\{d_j<0 \text{ 且 } x_j>0\} = \phi$ 时，取 $\alpha_{\max} = +\infty$．

算法 10.3.1（既约梯度法）

（1）取初始可行点 $\boldsymbol{x}^{(1)}$，即 $\boldsymbol{x}^{(1)}$满足

$$A\boldsymbol{x}^{(1)} = \boldsymbol{b},$$

$$\boldsymbol{x}^{(1)} \geqslant 0,$$

置 $k=1$．

（2）构造指标集

$$J_k = \{ j \mid x_j \text{ 是 } \boldsymbol{x}^{(k)} \text{ 的 } m \text{ 个最大分量之一} \},$$

并计算

$$B = (a_j \mid j \in J_k),$$

$$N = (a_j \mid j \notin J_k),$$

$$\boldsymbol{r}_N^T = \nabla_N f(\boldsymbol{x}^{(k)})^T - \nabla_B f(\boldsymbol{x}^{(k)})^T B^{-1} N,$$

$$d_j^{(k)} = \begin{cases} -r_j, & \text{当 } r_j \leqslant 0, \\ -x_j r_j, & \text{当 } r_j > 0, \end{cases} \quad j \notin J_k,$$

$$\boldsymbol{d}_B^{(k)} = -B^{-1} N \boldsymbol{d}_N^{(k)}.$$

(3) 若 $\boldsymbol{d}^{(k)} = 0$，则停止计算（$\boldsymbol{x}^{(k)}$作为最优解）；否则求解一维问题

$$\begin{aligned} \min \quad & \phi(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}), \\ \text{s.t.} \quad & 0 \leqslant \alpha \leqslant \alpha_{\max}, \end{aligned}$$

其中

$$\alpha_{\max} = \begin{cases} \min\left\{ \dfrac{-x_j^{(k)}}{d_j^{(k)}} \,\middle|\, d_j^{(k)} < 0 \text{ 且 } x_j^{(k)} > 0 \right\}, \\ +\infty, \qquad \{ d_j^{(k)} < 0 \text{ 且 } x_j^{(k)} > 0 \} = \phi, \end{cases}$$

得到 α_k，置

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}.$$

(4) 置 $k = k+1$，转 (2).

在实际计算中，算法第 (3) 步 $\boldsymbol{d}^{(k)} = 0$ 可改为 $\| \boldsymbol{d}^{(k)} \| \leqslant \epsilon$，其中 ϵ 是预先指定的精度.

【例 10.3.1】 用既约梯度法求解线性约束问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = 2x_1^2 - 2x_1 x_2 + 2x_2^2 - 4x_1 - 6x_2, \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 2, \\ & x_1 + 5x_2 + x_4 = 5, \\ & x_1,\ x_2,\ x_3,\ x_4 \geqslant 0, \end{aligned}$$

取初始点 $\boldsymbol{x}^{(1)}=(0,0,2,5)^T$.

解

$$\nabla f(\boldsymbol{x})=(4x_1-2x_2-4,-2x_1+4x_2-6,0,0)^T,$$

因为 $\boldsymbol{x}^{(1)}=(0,0,2,5)^T$，所以

$$J_1=\{3,4\},$$

$$B=[a_3,a_4]=\begin{pmatrix}1&0\\0&1\end{pmatrix},$$

$$N=[a_1,a_2]=\begin{pmatrix}1&1\\1&5\end{pmatrix},$$

$$\nabla f(\boldsymbol{x}^{(1)})^T=(-4,-6,0,0),$$

$$\nabla_B f(\boldsymbol{x}^{(1)})^T=(0,0),$$

因此

$$\begin{aligned}\boldsymbol{r}^T&=\nabla f(\boldsymbol{x}^{(1)})^T-\nabla_B f(\boldsymbol{x}^{(1)})^T B^{-1}A\\&=(-4,-6,0,0)-(0,0)\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}1&1&1&0\\1&5&0&1\end{pmatrix}\\&=(-4,-6,0,0),\end{aligned}$$

所以

$$\boldsymbol{r}_N^T=(-4,-6).$$

注意，既约梯度类似于求解线性规划的单纯形法中的目标行的计算，并且有 $\boldsymbol{r}_B=0$. 列出表格如下：

		x_1	x_2	x_3	x_4
$\boldsymbol{x}^{(1)}$		0	0	2	5
$\nabla f(\boldsymbol{x}^{(1)})$		-4	-6	0	0
$\nabla_B f(\boldsymbol{x}^{(1)})=\begin{pmatrix}0\\0\end{pmatrix}$	x_3	1	1	1	0
	x_4	1	5	0	1
$\boldsymbol{r}$		-4	-6	0	0

$r_1=-4<0$，$r_2=-6<0$，得到

$$\boldsymbol{d}_N=\begin{pmatrix}d_1\\d_2\end{pmatrix}=\begin{pmatrix}4\\6\end{pmatrix},$$

$$\boldsymbol{d}_B=\begin{pmatrix}d_3\\d_4\end{pmatrix}=-B^{-1}N\boldsymbol{d}_N=-\begin{pmatrix}1&1\\1&5\end{pmatrix}\begin{pmatrix}4\\6\end{pmatrix}=\begin{pmatrix}-10\\-34\end{pmatrix},$$

因此，搜索方向为

$$\boldsymbol{d}^{(1)}=(4,\ 6,\ -10,\ -34)^T.$$

求解一维问题，这里

$$\alpha_{\max}=\min\left\{\frac{-2}{-10},\ \frac{-5}{-34}\right\}=\frac{5}{34},$$

$$\phi(\alpha)=56\alpha^2-52\alpha,$$

因此

$$\alpha_1=\alpha_{\max}=\frac{5}{34},$$

$$\boldsymbol{x}^{(2)}=\boldsymbol{x}^{(1)}+\alpha_1\boldsymbol{d}^{(1)}=\left(\frac{10}{17},\ \frac{15}{17},\ \frac{9}{17},\ 0\right)^T.$$

下面作第二轮计算，在 $\boldsymbol{x}^{(2)}$处，有

$$J_2=\{1,2\},$$

$$B=[a_1,a_2]=\begin{pmatrix}1&1\\1&5\end{pmatrix},$$

$$N=[a_3,a_4]=\begin{pmatrix}1&0\\0&1\end{pmatrix},$$

$$\nabla f(\boldsymbol{x}^{(2)})^T=\left(-\frac{58}{17},\ -\frac{62}{17},0,0\right),$$

得到

		x_1	x_2	x_3	x_4
$\boldsymbol{x}^{(2)}$		$\frac{10}{17}$	$\frac{15}{17}$	$\frac{9}{17}$	0
$\nabla f(\boldsymbol{x}^{(2)})$		$-\frac{58}{17}$	$-\frac{62}{17}$	0	0
$\nabla_B f(\boldsymbol{x}^{(2)})=\begin{pmatrix}-\frac{58}{17}\\-\frac{62}{17}\end{pmatrix}$	x_1	1	1	1	0
	x_2	1	5	0	1

对上表分别以 α_{11}、α_{22}为中心作两次转轴运算，将$\nabla f(\boldsymbol{x}^{(2)})$中对应的基变量的元素为化 0，补到最后一行得到既约梯度 $\boldsymbol{r}$（有关的证明略），由此有

		x_1	x_2	x_3	x_4
$\boldsymbol{x}^{(2)}$		$\frac{10}{17}$	$\frac{15}{17}$	$\frac{9}{17}$	0
$\nabla f(\boldsymbol{x}^{(2)})$		$-\frac{58}{17}$	$-\frac{62}{17}$	0	0
$\nabla_B f(\boldsymbol{x}^{(2)})=\begin{pmatrix}-\frac{58}{17}\\-\frac{62}{17}\end{pmatrix}$	x_1	1	0	$\frac{5}{4}$	$-\frac{1}{4}$
	x_2	0	1	$-\frac{1}{4}$	$\frac{1}{4}$
$\boldsymbol{r}$		0	0	$\frac{57}{17}$	$\frac{1}{17}$

此时 $r_3=\frac{57}{17}>0$、$r_4=\frac{1}{17}>0$，得到

$$\boldsymbol{d}_N=\begin{pmatrix}-x_3r_3\\-x_4r_4\end{pmatrix}=\begin{pmatrix}-\frac{9}{17}\cdot\frac{57}{17}\\0\cdot\frac{1}{17}\end{pmatrix}=\begin{pmatrix}-\frac{513}{289}\\0\end{pmatrix},$$

$$\boldsymbol{d}_B = -B^{-1}N\boldsymbol{d}_N = -\begin{bmatrix} \frac{5}{4} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{bmatrix}\begin{bmatrix} -\frac{513}{289} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{2565}{1165} \\ -\frac{513}{1156} \end{bmatrix},$$

因此，搜索方向为

$$\boldsymbol{d}^{(2)} = \left(\frac{2565}{1156}, -\frac{513}{1156}, -\frac{513}{289}, 0\right)^T.$$

求解一维问题，这里

$$\alpha_{\max} = \min\left\{\frac{-\frac{15}{17}}{-\frac{513}{1156}}, \frac{-\frac{9}{17}}{-\frac{513}{289}}\right\} = \frac{17}{57},$$

$$\phi(\alpha) = 12.21\alpha^2 - 5.95\alpha - 6.435.$$

因此

$$\alpha_2 = \frac{68}{279}.$$

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} + \alpha_2\boldsymbol{d}^{(2)} = \left(\frac{35}{31}, \frac{24}{31}, \frac{3}{31}, 0\right)^T.$$

再作第三轮计算．在 $x^{(3)}$ 处，$J_3 = \{1, 2\}$，B、N 同上．$\nabla f(\boldsymbol{x}^{(3)}) = \left(-\frac{32}{31}, -\frac{160}{31}, 0, 0\right)^T$，经转轴计算得到：

		x_1	x_2	x_3	x_4
$\boldsymbol{x}^{(3)}$		$\frac{35}{31}$	$\frac{24}{31}$	$\frac{3}{31}$	0
$\nabla f(\boldsymbol{x}^{(3)})$		$-\frac{32}{31}$	$-\frac{160}{31}$	0	0
$\nabla_B f(\boldsymbol{x}^{(3)}) = \begin{bmatrix} -\frac{32}{31} \\ -\frac{160}{31} \end{bmatrix}$	x_1	1	0	$\frac{5}{4}$	$-\frac{1}{4}$
	x_2	0	1	$-\frac{1}{4}$	$\frac{1}{4}$
$\boldsymbol{r}$		0	0	0	$\frac{32}{31}$

此时 $r_3=0$，$r_4=\frac{32}{31}$，所以 $d_3=\frac{3}{31}\cdot 0=0$，$d_4=0\cdot\frac{32}{31}=0$. 因此，$\boldsymbol{d}_N=0$，$\boldsymbol{d}_B=0$，即 $\boldsymbol{d}=0$，所以 $\boldsymbol{x}^{(3)}$是最优点.

将上述计算结果列表如下：

表 10.3.1　例 10.3.1 的计算结果

k	$\boldsymbol{x}^{(k)}$	$f(\boldsymbol{x}^{(k)})$	$\boldsymbol{r}^{(k)}$	$\boldsymbol{d}^{(k)}$	$\alpha_{\max}$	α_k
1	$(0,0,2,5)$	0.0	$(-4,-6,0,0)$	$(4,6,-10,-34)$	$\frac{5}{34}$	$\frac{5}{34}$
2	$\left(\frac{10}{17},\frac{15}{17},\frac{9}{17},0\right)$	-6.436	$\left(0,0,\frac{57}{17},\frac{4}{17}\right)$	$\left(\frac{2565}{1156},-\frac{513}{1156},-\frac{513}{289},0\right)$	$\frac{17}{57}$	$\frac{68}{278}$
3	$\left(\frac{35}{31},\frac{24}{31},\frac{3}{31},0\right)$	-7.16	$\left(0,0,0,\frac{32}{31}\right)$	$(0,0,0,0)$	—	—

习　题　十

10.1　用可行方向法求解约束问题

$$\begin{aligned}\min\quad & f(\boldsymbol{x})=x_1^2+x_1x_2+2x_2^2-12x_1-18x_2,\\ \text{s.t.}\quad & -3x_1+6x_2\leqslant 9,\\ & -2x_1+x_2\leqslant 1,\\ & x_1,x_2\geqslant 0;\end{aligned}$$

取初始点 $\boldsymbol{x}^{(1)}=(0,\ 0)^T$.

10.2　考虑不等式约束问题

$$\begin{aligned}\min\quad & f(\boldsymbol{x}),\qquad \boldsymbol{x}\in R^n,\\ \text{s.t.}\quad & c_i(\boldsymbol{x})\leqslant 0,\quad i\in I=\{1,2,\cdots,m\},\end{aligned}$$

其中 $f(\boldsymbol{x})$、$c_i(\boldsymbol{x})$，$(i\in I)$具有连续的偏导数，设 $\bar{x}$ 是约束问题

的可行点，若在 $\bar{x}$ 处 d 满足

$$\nabla f(\bar{x})^T d < 0,$$
$$\nabla c_i(\bar{x})^T d < 0, \quad i \in I(\bar{x}),$$

则 d 是 $\bar{x}$ 处的可行下降方向.

10.3 考虑不等式约束问题

$$\begin{aligned} \min \quad & f(x), \qquad x \in R^n, \\ \text{s.t.} \quad & c_i(x) \leqslant 0, \quad i \in I = \{1, 2, \cdots, m\}, \end{aligned}$$

其中 $f(x)$、$c_i(x)(i \in I)$ 具有连续的偏导数. 设 $\bar{x}$ 是约束问题的可行点,构造线性规划问题

$$\begin{aligned} \min \quad & z, \\ \text{s.t.} \quad & \nabla f(\bar{x})^T d - z \leqslant 0, \\ & \nabla c_i(\bar{x})^T d - z \leqslant 0, \quad i \in I(\bar{x}), \\ & -1 \leqslant d_i \leqslant 1, \quad i = 1, 2, \cdots, n. \end{aligned}$$

设线性规划问题的最优解为 $(\bar{d}, \bar{z})$. 证明:

(1) 若 $\bar{z} < 0$, 则 $\bar{d}$ 是 $\bar{x}$ 处的可行下降方向.

(2) 若 $\bar{z} = 0$, 且在 $\bar{x}$ 处 $\nabla c_i(\bar{x})(i \in I(\bar{x}))$ 线性无关,则 $\bar{x}$ 是不等式约束问题的 K-T 点.

10.4 利用习题 10.3 的结论，构造求解不等式约束问题

$$\begin{aligned} \min \quad & f(x), \qquad x \in R^n, \\ \text{s.t.} \quad & c_i(x) \leqslant 0, \quad i \in I = \{1, 2, \cdots, m\} \end{aligned}$$

的可行方向法. 试写出相应的算法步骤.

10.5 考虑变量有上下界的约束问题

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & a_i \leqslant x_i \leqslant b_i, \quad i = 1, 2, \cdots, n \end{aligned}$$

设 $\bar{x}$ 是可行点, $\nabla_i = \dfrac{\partial f(x)}{\partial x_i}\Big|_{x=\bar{x}}$. 考虑用可行方向法的计算过程得到的可行下降方向.

(1) 设

$$d_i=\begin{cases}-1, & 如果\ \overline{x}_i>a_i\ 且\ \nabla_i\geqslant 0,\\ 1, & 如果\ \overline{x}_i<b_i\ 且\ \nabla_i<0,\\ 0, & 其它.\end{cases}$$

证明：$\boldsymbol{d}=(d_1,d_2\cdots,d_n)^T$ 是子问题Ⅰ的最优解.

(2) 设

$$d_i=\begin{cases}\dfrac{-\nabla_i}{\sqrt{\sum_{j\in J}\nabla_i^2}}, & 如果\ i\in J,\\ 0, & 其它,\end{cases}$$

其中 $J=\{j\mid(\overline{x}_i>a_i\ 且\ \nabla_i\geqslant 0\})$ 或者 $(\overline{x}_i<b_i\ 且\ \nabla_i<0)\}$,

证明：$\boldsymbol{d}=(d_1,d_2\cdots,d_n)^T$ 是子问题Ⅱ的最优解.

(3) 利用 (1) 和 (2) 求解有上下界的约束优化问题

$$\begin{aligned}\min\quad & 2x_1^2-x_1x_2+3x_2^2-3x_1-2x_2,\\ x.t.\ & -3\leqslant x_1\leqslant 0,\\ & -4\leqslant x_2\leqslant 1,\end{aligned}$$

取初始点 $\boldsymbol{x}^{(1)}=(-3,-4)^T$，并比较两方法得到的轨迹.

10.6 用投影梯度法求解习题 10.1.

10.7 用投影梯度法求解约束问题

$$\begin{aligned}\min\quad & 4x_1^2+x_2^2-32x_1-34x_2,\\ \text{s.t.}\quad & x_1\leqslant 2,\\ & x_1+2x_2\leqslant 6,\\ & x_1,x_2\geqslant 0,\end{aligned}$$

取初始点 $\boldsymbol{x}^{(1)}=(0,0)^T$.

10.8 考虑不等式约束

$$\boldsymbol{a}_i^T\boldsymbol{x}-b_i\leqslant 0,\ i=1,2,\cdots,m,$$

设 $\hat{\boldsymbol{x}}$ 是可行点，$I(\hat{\boldsymbol{x}})$为有效约束，令

$$N=[\boldsymbol{a}_{i_1}, \boldsymbol{a}_{i_2}, \cdots, \boldsymbol{a}_{i_r}], \quad i_j \in I(\bar{x}),$$

$$P = I - N(N^T N)^{-1} N^T,$$

试对下列情况给出相应的几何解释：

(1) $P\nabla f(\hat{x}) = 0$,

(2) $P\nabla f(\hat{x}) = \nabla f(\hat{x})$,

(3) $P\nabla f(\hat{x}) \neq 0$.

10.9 设 $N = \begin{bmatrix} 1 & -1 \\ 2 & 1 \\ -1 & 0 \\ 0 & 1 \end{bmatrix}$，用递推方法（（10.2.51）式）求出投影矩阵 P.

10.10 对于等式约束二次规划问题

$$\min \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G \boldsymbol{x} + r^T \boldsymbol{x},$$

$$\text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b},$$

其中 G 为 $n \times n$ 阶正定对称矩阵，A 为 $m \times n$（$m < n$）阶矩阵，且 $\text{rank}(A) = m$. 考虑如下算法：

(1) 取初始可行点 $\boldsymbol{x}^{(1)}$，初始正定对称矩阵 $H^{(1)}$. 置精度要求 ε，置 $k=1$.

(2) 取 $\| P^{(k)} \nabla f(\boldsymbol{x}^{(k)}) \| \leqslant \varepsilon$，则停止计算；否则置

$$\boldsymbol{d}^{(k)} = -P^{(k)} \nabla f(\boldsymbol{x}^{(k)}),$$

其中

$$P^{(k)} = H^{(k)} - H^{(k)} A^T (A H^{(k)} A^T)^{-1} A H^{(k)}.$$

(3) 一维搜索. 求解一维问题

$$\min \quad \phi(\alpha) = f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}),$$

$$\text{s.t.} \quad \alpha \geqslant 0,$$

得 α_k，置

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}.$$

(4) 如果 $k<n$，修正 $H^{(k)}$，置

$$H^{(k+1)}=H^{(k)}-\frac{H^{(k)}\boldsymbol{y}^{(k)}\ (\boldsymbol{y}^{(k)})^TH^{(k)}}{(\boldsymbol{y}^{(k)})^TH^{(k)}\boldsymbol{y}^{(k)}}+\frac{\boldsymbol{s}^{(k)}\ (\boldsymbol{s}^{(k)})^T}{(\boldsymbol{s}^{(k)})^T\boldsymbol{y}^{(k)}},$$

其中

$$\boldsymbol{s}^{(k)}=\boldsymbol{x}^{(k+1)}-\boldsymbol{x}^{(k)},\boldsymbol{y}^{(k)}=\nabla f(\boldsymbol{x}^{(k+1)})-\nabla f(\boldsymbol{x}^{(k)}),$$

然后置 $k=k+1$，转 (2).

(5) 置 $\boldsymbol{x}^{(1)}=\boldsymbol{x}^{(k+1)}$，置 $k=1$，转 (2).

假设在上述算法中，一维搜索是精确的.

(a) 证明：$\boldsymbol{d}^{(k)}$是可行方向，$\boldsymbol{x}^{(k+1)}$是可行点.

(b) 证明：$P^{(k)}(H^{(k)})^{-1}P^{(k)}=P^{(k)}$，所以当$P^{(k)}\nabla f(\boldsymbol{x}^{(k)})\neq 0$ 时，$\boldsymbol{d}^{(k)}$是下降方向.

(c) 若 $P^{(k)}\nabla f(\boldsymbol{x}^{(k)})=0$，则 $\boldsymbol{x}^{(k)}$是原问题的 K-T 点.

(d) 若算法产生了 $\boldsymbol{d}^{(1)}$，$\boldsymbol{d}^{(2)}$，…，$\boldsymbol{d}^{(l)}$，证明：

$$\boldsymbol{d}^{(i)T}G\boldsymbol{d}^{(j)}=0,\quad 1\leqslant j<i\leqslant l,$$
$$H^{(l+1)}\boldsymbol{y}^{(i)}=\boldsymbol{s}^{(i)},\quad i=1,\ 2,\ \cdots,\ l.$$

(e) 算法至多 $n-m$ 步终止.

10.11 设 A 为 $m\times n$ 阶矩阵，$m<n$，$\operatorname{rank}(A)=m$，证明 $\bar{\boldsymbol{d}}$ 为问题

$$\begin{aligned}&\min\quad \|-\nabla f(\boldsymbol{x})-\boldsymbol{d}\|,\\&\text{s.t.}\quad A\boldsymbol{d}=\boldsymbol{0}\end{aligned}$$

的最优解的充要条件是：$\bar{\boldsymbol{d}}$ 是 $-\nabla f(\boldsymbol{x})$ 在 A 的零空间上的投影.

10.12 用既约梯度法求解约束问题

$$\begin{aligned}\min\quad & x_1^2+x_1x_2+2x_2^2-6x_1-14x_2,\\ \text{s.t.}\quad & x_1+x_2+x_3=2,\\ & -x_1+2x_2\leqslant 3,\\ & x_1,\ x_2,\ x_3\geqslant 0.\end{aligned}$$

10.13 在既约梯度法中，如果用如下方法构造 $\boldsymbol{d}_N$ 中的分

量d，如下：

$$d_i=\begin{cases}-r_i & \text{如果}\ \bar{x}_i>0\ \text{或}\ r_i\leqslant 0,\\ 0, & \text{其它},\end{cases}\quad i\in A,$$

证明：

（1）若 $\boldsymbol{d}\neq 0$，则 $\boldsymbol{d}$ 是$\bar{\boldsymbol{x}}$处的下降方向；

（2）$\boldsymbol{d}=0$ 的充分必要条件是：$\bar{\boldsymbol{x}}$为约束问题的 K-T 点.

第十一章　乘　子　法

求解约束最优化问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}), \quad \boldsymbol{x} \in R^n, \\ \text{s.t.} \quad & c_i(\boldsymbol{x}) = 0, \quad i \in E = \{1,2,\cdots,l\}, \\ & c_i(\boldsymbol{x}) \leqslant 0, \quad i \in I = \{l+1, l+2, \cdots, l+m\} \end{aligned} \tag{11.0.1}$$

有两类方法．一类是由第九章、第十章介绍的方法，它们利用约束问题本身的性质，直接求解约束问题．另一类方法是将约束问题转化为一系列无约束问题，通过求解一系列无约束最优化问题，来得到约束问题的最优解．这类方法称为**序列无约束极小化方法**（Sequerical Unconstrained Minimization Technique）．本章介绍的惩罚函数法和乘子法属于后种方法．

§11.1　惩罚函数法

11.1.1　罚函数法

最早的**罚函数**（Penalty function）法是由 Courant 在 1943 年提出来的，其基本思想是：对不满足约束条件的点进行惩罚，通过求解多个罚函数的极小得到约束问题的最优解．

先给一个简单的例子来阐明罚函数法的基本思想．考虑一个变量仅有一个不等式约束的问题：

$$\begin{aligned} \min \quad & f(x) = x^2, \\ \text{s.t.} \quad & c(x) = x + 1 \leqslant 0, \end{aligned} \tag{11.1.1}$$

其可行域为 $(-\infty, -1]$，最优解为 $x^* = -1$．

现将问题（11.1.1）转化为无约束问题，也就是说，希望构造相应的无约束问题，而该无约束问题的解恰好是约束问题的最优解．从直观上看，要做到这一点就必须增大在非可行域处的目标函数值，即对非可行域处的目标函数加以惩罚．

现构造惩罚函数，一种较为简单而又能保证函数具有连续的偏导数的方法为：

$$P(x,\sigma)=\begin{cases} f(x), & c(x)\leqslant 0, \\ f(x)+\frac{\sigma}{2}[c(x)]^2, & c(x)>0, \end{cases}$$

$$=\begin{cases} x^2, & x+1\leqslant 0, \\ x^2+\frac{\sigma}{2}(x+1)^2, & x+1>0, \end{cases} \tag{11.1.2}$$

其无约束问题

$$\min P(x,\sigma) \tag{11.1.3}$$

的最优解为

$$\overline{x}(\sigma)=-\frac{\sigma}{2+\sigma}, \tag{11.1.4}$$

当 $\sigma\to+\infty$ 时，有

$$\overline{x}(\sigma)\to -1=x^*, \tag{11.1.5}$$

即约束问题的最优解．罚函数的几何意义见图 11.1.1.

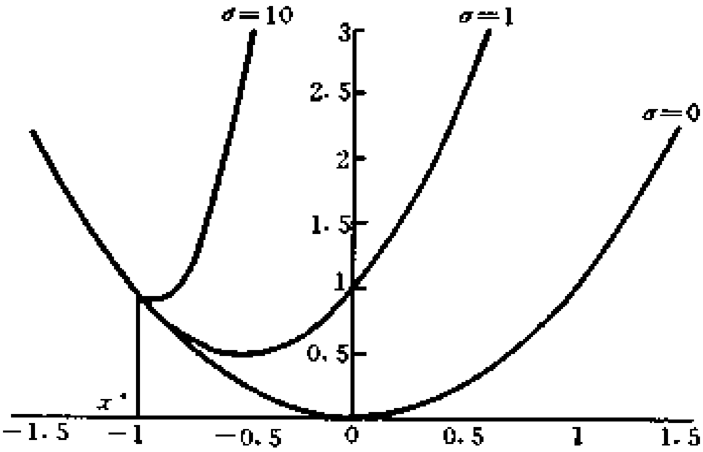


图 11.1.1　罚函数的几何意义

因此，对于不等式约束问题

$$\begin{aligned}&\min \quad f(\boldsymbol{x}), \quad \boldsymbol{x} \in R^n, \\ &\text{s.t.} \ c(\boldsymbol{x}) \leqslant 0,\end{aligned} \tag{11.1.6}$$

其罚函数定义为

$$\begin{aligned}P(\boldsymbol{x},\sigma) &= \begin{cases} f(\boldsymbol{x}), & c(\boldsymbol{x}) \leqslant 0, \\ f(\boldsymbol{x}) + \dfrac{\sigma}{2}[c(\boldsymbol{x})]^2, & c(\boldsymbol{x}) > 0 \end{cases} \\ &= f(\boldsymbol{x}) + \frac{\sigma}{2}(\max\{0, c(\boldsymbol{x})\})^2.\end{aligned} \tag{11.1.7}$$

对于等式约束问题

$$\begin{aligned}&\min \quad f(\boldsymbol{x}), \quad \boldsymbol{x} \in R^n, \\ &\text{s.t.} \quad c(\boldsymbol{x}) = 0,\end{aligned} \tag{11.1.8}$$

其罚函数定义为

$$P(\boldsymbol{x},\sigma) = f(\boldsymbol{x}) + \frac{\sigma}{2}(c(\boldsymbol{x}))^2, \tag{11.1.9}$$

即不满足等式方程就加以惩罚.

由上面的分析可以得到，对于一般约束问题（11.0.1），其罚函数定义为

$$\begin{aligned}P(\boldsymbol{x},\sigma) &= f(\boldsymbol{x}) + \frac{\sigma}{2}\Big(\sum_{i\in E}[c_i(\boldsymbol{x})]^2 + \sum_{i\in I}[\max\{0, c_i(\boldsymbol{x})\}]^2\Big) \\ &= f(\boldsymbol{x}) + \frac{\sigma}{2}S(\boldsymbol{x}),\end{aligned} \tag{11.1.10}$$

其中

$$S(\boldsymbol{x}) = \Big(\sum_{i\in E}[c_i(\boldsymbol{x})]^2 + \sum_{i\in I}[\max\{0, c_i(\boldsymbol{x})\}]^2\Big),$$

称 σ 为惩罚因子，称 $\frac{\sigma}{2}S(\boldsymbol{x})$ 为惩罚项.

可以期望，当 σ 充分大时，无约束问题

$$\min P(\boldsymbol{x},\sigma) = f(\boldsymbol{x}) + \frac{\sigma}{2}S(\boldsymbol{x}) \tag{11.1.11}$$

的最优解 $\bar{x}=\bar{x}(\sigma)$接近约束问题的最优解 x^*.

因此，可得到如下算法.

算法 11.1.1 (罚函数法)

(1) 选择序列 $\{\sigma_k\}$，σ_k 递增趋于 $+\infty$，取初始点 $x^{(0)}$，置精度要求 ϵ，置 $k=1$.

(2) 以 $x^{(k-1)}$为初始点，求解无约束问题

$$\min P(x,\sigma_k) = f(x) + \frac{\sigma_k}{2}S(x), \qquad (11.1.12)$$

得到最优解 $x^{(k)}$.

(3) 若 $\sigma_k S(x^{(k)}) \leqslant \epsilon$，则停止计算（$x^{(k)}$作为约束问题的最优解）；否则置

$$k = k+1,$$

转 (2).

【例 11.1.1】 用罚函数法求解约束问题

$$\begin{aligned} \min \quad & f(x) = \frac{1}{2}\left(x_1^2 + \frac{1}{3}x_2^2\right), \\ \text{s.t.} \quad & c(x) = x_1 + x_2 - 1 = 0. \end{aligned} \qquad (11.1.13)$$

解

$$\begin{aligned} P(x,\sigma) &= f(x) + \frac{\sigma}{2}c^2(x) \\ &= \frac{1}{2}\left(x_1^2 + \frac{1}{3}x_2^2\right) + \frac{\sigma}{2}(x_1 + x_2 - 1)^2, \end{aligned}$$

所以

$$\frac{\partial P}{\partial x_1} = x_1 + \sigma(x_1 + x_2 - 1),$$

$$\frac{\partial P}{\partial x_2} = \frac{1}{3}x_2 + \sigma(x_1 + x_2 - 1).$$

令$\nabla_x P(x,\sigma)=0$,得到

$$\bar{x}_1(\sigma) = \frac{\sigma}{1+4\sigma}, \quad \bar{x}_2(\sigma) = \frac{3\sigma}{1+4\sigma}.$$

令 $\sigma\to\infty$，得到

$$x^*=\left(\frac{1}{4},\ \frac{3}{4}\right)^T.$$

无论是问题（11.1.1）还是问题（11.1.13），其相应的无约束问题的最优解均在可行域的外部. 而对于一般约束问题，除非它的最优解 x^* 也是一个无约束问题的最优解，通常 x^* 总位于可行域的边界上，因此，采用罚函数法得到的无约束问题的最优解 $x^{(k)}$均位于可行域的外部，因此，罚函数法也称为外罚函数法.

11.1.2 罚函数法的收敛性质

对于罚函数法，有如下性质.

定理 11.1.1 若序列 $\{\sigma_k\}$ 递增趋于 $+\infty$，$x^{(k)}$是无约束问题（11.1.12）的全局最优解，则：

（1）序列$\{P(x^{(k)},\sigma_k)\}$非减；

（2）序列$\{S(x^{(k)})\}$非增；

（3）序列$\{f(x^{(k)})\}$非减.

证明 先证（1）.

设 $\sigma_k<\sigma_{k+1}$、$x^{(k)}$和 $x^{(k+1)}$分别是无约束问题$P(x,\sigma_k)$和 $P(x,\sigma_{k+1})$的全局最优解，因此有

$$\begin{aligned}P(x^{(k)},\sigma_k)&\leqslant P(x^{(k+1)},\sigma_k)\\&=f(x^{(k+1)})+\frac{\sigma_k}{2}S(x^{(k+1)})\\&\leqslant f(x^{(k+1)})+\frac{\sigma_{k+1}}{2}S(x^{(k+1)})\\&=P(x^{(k+1)},\sigma_{k+1}),\end{aligned}\tag{11.1.14}$$

所以$\{P(x^{(k)},\sigma_k)\}$非减.

再证明（2）. 由于

$$P(x^{(k+1)},\sigma_{k+1})\leqslant P(x^{(k)},\sigma_{k+1}),\tag{11.1.15}$$

和

$$P(\boldsymbol{x}^{(k)},\sigma_k)\leqslant P(\boldsymbol{x}^{(k+1)},\sigma_k), \tag{11.1.16}$$

将（11.1.15）式和（11.1.16）式相加，经整理得到

$$\begin{aligned}&P(\boldsymbol{x}^{(k+1)},\sigma_{k+1})-P(\boldsymbol{x}^{(k+1)},\sigma_k)\\&\leqslant P(\boldsymbol{x}^{(k)},\sigma_{k+1})-P(\boldsymbol{x}^{(k)},\sigma_k),\end{aligned} \tag{11.1.17}$$

即

$$\frac{\sigma_{k+1}-\sigma_k}{2}S(\boldsymbol{x}^{(k+1)})\leqslant\frac{\sigma_{k+1}-\sigma_k}{2}S(\boldsymbol{x}^{(k)}),$$

所以

$$S(\boldsymbol{x}^{(k+1)})\leqslant S(\boldsymbol{x}^{(k)}),$$

即$\{S(\boldsymbol{x}^{(k)})\}$非增.

由（11.1.16）式得到

$$f(\boldsymbol{x}^{(k)})+\frac{\sigma_k}{2}S(\boldsymbol{x}^{(k)})\leqslant f(\boldsymbol{x}^{(k+1)})+\frac{\sigma_k}{2}S(\boldsymbol{x}^{(k+1)}),$$

再由结论（2）得到

$$f(\boldsymbol{x}^{(k)})-f(\boldsymbol{x}^{(k+1)})\leqslant\frac{\sigma_k}{2}(S(\boldsymbol{x}^{(k+1)})-S(\boldsymbol{x}^{(k)}))\leqslant 0,$$

即$\{f(\boldsymbol{x}^{(k)})\}$非减.

定理 11.1.2 设 $f(\boldsymbol{x})$、$c_i(\boldsymbol{x})(i\in E\cup I)$具有连续的一阶偏导数，$\boldsymbol{x}^*$是约束问题（11.0.1）的全局最优解，惩罚因子 $\{\sigma_k\}$递增趋于$+\infty$，若 $\boldsymbol{x}^{(k)}$是罚函数$P(\boldsymbol{x},\sigma_k)$的全局解，则 $\{\boldsymbol{x}^{(k)}\}$的任一聚点必是约束问题（11.0.1）的全局解.

证明 设 $\tilde{\boldsymbol{x}}$ 是 $\{\boldsymbol{x}^{(k)}\}$ 的一个聚点，则存在收敛子列，不妨仍设为 $\{\boldsymbol{x}^{(k)}\}$. 因为 $\boldsymbol{x}^{(k)}$是罚函数$P(\boldsymbol{x},\sigma_k)$的全局解，因此有

$$P(\boldsymbol{x}^{(k)},\sigma_k)\leqslant P(\boldsymbol{x}^*,\sigma_k)=f(\boldsymbol{x}^*)+\frac{\sigma_k}{2}S(\boldsymbol{x}^*), \tag{11.1.18}$$

由于 x^* 是约束问题的可行点，因此

$$S(x^*)=\sum_{i\in E}[c_i(x^*)]^2+\sum_{i\in I}[\max\{0,c_i(x^*)\}]^2=0,$$

因此，

$$P(x^{(k)},\sigma_k)\leqslant f(x^*), \qquad (11.1.19)$$

即

$$f(x^{(k)})+\frac{\sigma_k}{2}S(x^{(k)})\leqslant f(x^*). \qquad (11.1.20)$$

由于 $\sigma_k\to+\infty$，（11.1.20）式表明，$S(x^{(k)})\to 0$. 即 $S(\tilde{x})=0$，因此 $\tilde{x}$ 是可行点.

再证明，$\tilde{x}$ 是全局最优解. 由（11.1.20）式，并注意到 $S(x^{(k)})\geqslant 0$，因此有 $f(x^{(k)})\leqslant f(x^*)$. 令 $k\to\infty$，得到 $f(\tilde{x})\leqslant f(x^*)$. 另一方面，由于 $\tilde{x}$ 是可行点，x^* 是全局最优解，则有 $f(x^*)\leqslant f(\tilde{x})$. 因此，

$$f(\tilde{x})=f(x^*). \qquad (11.1.21)$$

证毕.

由（11.1.20）式和（11.1.21）式知，当 $k\to\infty$ 时，有 $\frac{\sigma_k}{2}S(x^{(k)})\to 0$. 这就是为什么终止准则为 $\sigma_k S(x^{(k)})\leqslant\epsilon$ 的原因.

定理 11.1.3 设 $f(x)$、$c_i(x)$（$i\in E\cup I$）具有连续的一阶偏导数，约束问题(11.0.1)的全局最优解存在，惩罚因子 $\{\sigma_k\}$ 递增趋于 $+\infty$，若 $x^{(k)}$ 是罚函数 $P(x,\sigma_k)$ 的全局解，且 $x^{(k)}\to x^*$（$k\to\infty$），在 x^* 处 $\nabla c(x^*)$（$i\in E\cup I(x^*)$）线性无关，则 x^* 是约束问题(11.0.1)的全局解，且

$$\lim_{k\to\infty}\sigma_k c_i(x^{(k)})=\lambda_i^*,\quad i\in E, \qquad (11.1.22)$$

$$\lim_{k\to\infty}\sigma_k\max\{0,c_i(\boldsymbol{x}^{(k)})\} = \lambda_i^*, \quad i\in I, \quad (11.1.23)$$

其中 $\boldsymbol{\lambda}^*$ 是 $\boldsymbol{x}^*$ 处的 Lagrange 乘子.

证明 由定理 11.1.2 可知，$\boldsymbol{x}^*$ 是约束问题的全局解，只需证明定理的第二部分，即（11.1.22）式和（11.1.23）式成立.

因为 $\boldsymbol{x}^{(k)}$ 是罚函数 $P(\boldsymbol{x},\sigma_k)$ 的全局解，则有

$$\nabla_x P(\boldsymbol{x}^{(k)},\sigma_k) = 0, \quad (11.1.24)$$

即

$$\nabla f(\boldsymbol{x}^{(k)}) + \sigma_k\Big(\sum_{i\in E}c_i(\boldsymbol{x}^{(k)})\nabla c_i(\boldsymbol{x}^{(k)})$$

$$+\sum_{i\in I}\max\{0,c_i(\boldsymbol{x}^{(k)})\}\nabla c_i(\boldsymbol{x}^{(k)})\Big) = 0. \quad (11.1.25)$$

当 $i\in I/I(\boldsymbol{x}^*)$ 时，有 $c_i(\boldsymbol{x}^*)<0$，则存在 K，当 $k>K$ 时，有 $c_i(\boldsymbol{x}^{(k)})<0$，因此

$$\lim_{k\to\infty}\sigma_k\max\{0,c_i(\boldsymbol{x}^{(k)})\} = 0 = \lambda_i^*, \quad i\in I/I(\boldsymbol{x}^*),$$

因此，（11.1.25）式改为

$$\nabla f(\boldsymbol{x}^{(k)}) + \sigma_k\Big(\sum_{i\in E}c_i(\boldsymbol{x}^{(k)})\nabla c_i(\boldsymbol{x}^{(k)})$$

$$+\sum_{i\in I(\boldsymbol{x}^*)}\max\{0,c_i(\boldsymbol{x}^{(k)})\}\nabla c_i(\boldsymbol{x}^{(k)})\Big) = 0. \quad (11.1.26)$$

从约束问题的一阶必要条件知

$$\nabla f(\boldsymbol{x}^*) + \sum_{i\in E\cup I(\boldsymbol{x}^*)}\lambda_i^*\nabla c_i(\boldsymbol{x}^*) = 0,$$

且 $\nabla c(\boldsymbol{x}^*)(i\in E\cup I(\boldsymbol{x}^*))$ 线性无关，因此当 $k\to\infty$ 时，有

$$\sigma_k c_i(\boldsymbol{x}^{(k)}) \to \lambda_i^*, \quad i\in E,$$

$$\sigma_k\max\{0,c_i(\boldsymbol{x}^{(k)})\} \to \lambda_i^*, \quad i\in I.$$

定理 11.1.3 表明，令 $\lambda_i^{(k)} = \sigma_k c_i(\boldsymbol{x}^{(k)})$，$i\in E$，$\lambda_i^{(k)} = \sigma_k\max\{0,c_i(\boldsymbol{x}^{(k)})\}$，$i\in I$ 作为约束问题的 Lagrange 乘子的近似值.

11.1.3 罚函数法的病态性质

罚函数法的优点是将约束问题化成一系列无约束问题，可用求解无约束问题的算法得到约束问题的最优解. 但是当惩罚因子 σ_k 充分大后，函数$P(x,\sigma_k)$通常是一个病态函数，即其 Hesse 矩阵$\nabla_x^2 P(x^{(k)},\sigma_k)$的条件数非常大，这给无约束问题的求解带来了困难.

考虑等式约束问题

$$\begin{aligned} \min \quad & f(x), x \in R^n, \\ \text{s.t.} \quad & c_i(x) = 0, i \in E = \{1,2,\cdots,l\}, \end{aligned} \tag{11.1.27}$$

其罚函数为

$$P(x,\sigma_k) = f(x) + \frac{\sigma_k}{2}\Big(\sum_{i\in E} c_i^2(x)\Big). \tag{11.1.28}$$

在 $x^{(k)}$处的 Hesse 矩阵为

$$\begin{aligned} \nabla_x^2 P(x^{(k)},\sigma_k) = & \nabla^2 f(x^{(k)}) + \sum_{i\in E} \sigma_k c_i(x^{(k)}) \nabla^2 c_i(x^{(k)}) \\ & + \sigma_k \sum_{i\in E} \nabla c_i(x^{(k)}) \nabla c_i(x^{(k)})^T \\ = & \nabla_x^2 L(x^{(k)},\lambda^{(k)}) + \sigma_k \sum_{i\in E} \nabla c_i(x^{(k)}) \nabla c_i(x^{(k)})^T. \end{aligned} \tag{11.1.29}$$

注意到矩阵$\sum_{i\in E} \nabla c_i(x^{(k)}) \nabla c_i(x^{(k)})^T$ 的秩为 l，所以当 $\sigma_k \to \infty$时，$\nabla_x^2 P(x^{(k)},\sigma_k)$中有 l 个特征值趋于∞，而其余的特征值有界，因此条件数趋于∞.

看一个具体的例子.

考虑约束问题

$$\begin{aligned} \min \quad & f(x) = x_1^2 + x_2^2, \\ \text{s.t.} \quad & c(x) = x_1 + 1 = 0, \end{aligned}$$

其罚函数为

$$P(\boldsymbol{x},\sigma) = x_1^2 + x_2^2 + \frac{\sigma}{2}(x_1 + 1)^2.$$

Hesse 矩阵为

$$\nabla_x^2 P(\boldsymbol{x},\sigma) = \begin{pmatrix} 2+\sigma & 0 \\ 0 & 2 \end{pmatrix},$$

该矩阵的条件数为

$$\mathrm{cond}(\nabla_x^2 P(\boldsymbol{x},\sigma)) = \frac{2+\sigma}{2} \to \infty.(\sigma \to \infty)$$

由上述分析可以看到，在使用罚函数法时，会出现这样的矛盾，选取很大的 σ_1，或者 σ_k 增加的很快，可以使算法收敛的快，但是很难精确地求解相应的无约束问题；选取很小的 σ_1，或者缓慢地增加 σ_k，可以保持 $x^{(k)}$ 与 $P(\boldsymbol{x},\sigma_{k+1})$ 的极小点接近，使求解相应的无约束问题变得容易，但收敛太慢，效果很差．因此，在实际计算过程中，要根据实际情况适当调整 σ_1 的选取和 σ_k 的增加速度．一种建议是选取 $\sigma_k = 0.1 \times 2^{k-1}$.

11.1.4 内罚函数法（障碍罚函数）

前面讲过的罚函数法是外罚函数法，即每步得到的 $x^{(k)}$ 均在可行域的外部，只能近似地满足约束条件．这对于某些实际问题，这样的最优解是不能接受的．本小节讨论内罚函数法，也称为**障碍罚函数法**（Barrier Function Methods）或闸罚函数法，其基本思想是在约束区域的边界上设置障碍，使函数的极小点在可行域的内部．

考虑约束问题

$$\begin{aligned} &\min \quad f(\boldsymbol{x}), \boldsymbol{x} \in R^n, \\ &\text{s.t.} \quad c_i(\boldsymbol{x}) \leqslant 0, i \in I = \{1,2,\cdots,m\}, \end{aligned} \tag{11.1.29}$$

其内罚函数为

$$P(\boldsymbol{x}, r) = f(\boldsymbol{x}) + rB(\boldsymbol{x}), \qquad (11.1.30)$$

这里仍称 $rB(\boldsymbol{x})$为惩罚项,称 $B(\boldsymbol{x})$为障碍函数.当 $\boldsymbol{x}$ 从可行域的内部趋于边界时,$B(\boldsymbol{x})$趋于∞.通常有两种方式选择 $B(\boldsymbol{x})$,一种是由 Carrall 在 1961 年提出的倒数障碍函数

$$B(\boldsymbol{x}) = \sum_{i=1}^{m} \frac{-1}{c_i(\boldsymbol{x})}, \qquad (11.1.31)$$

另一种是由 Frisch 在 1955 年提出的对数障碍函数

$$B(\boldsymbol{x}) = -\sum_{i=1}^{m} \ln[-c_i(\boldsymbol{x})], \qquad (11.1.32)$$

由于障碍函数在可行域的边界上无界，所以对于障碍罚函数法的初始点一定要在可行域的内部，即初始点 $\boldsymbol{x}^{(0)}$满足

$$c_i(\boldsymbol{x}^{(0)}) < 0, i = 1,2,\cdots,m.$$

设 $\overline{\boldsymbol{x}} = \boldsymbol{x}(r)$是无约束问题

$$\min \quad P(\boldsymbol{x}, r) = f(\boldsymbol{x}) + rB(\boldsymbol{x})$$

的全局最优解，$\boldsymbol{x}^*$是约束问题（11.1.29）的全局最优解，在满足一定条件下，可以证明

$$\lim_{r\to 0}\overline{\boldsymbol{x}}(r) = \boldsymbol{x}^*. \qquad (11.1.33).$$

对于两种障碍函数，是哪种形式的的效果好呢？我们从一个简单的例子讨论一下．考虑问题

$$\begin{aligned} \min \quad & f(x) = -x, \\ \text{s.t.} \quad & c(x) = x + 1 \leqslant 0, \end{aligned} \qquad (11.1.34)$$

取

$$B(x) = -\frac{1}{x+1}, \qquad (11.1.35)$$

则内罚函数为

$$P(x, r) = -x - \frac{r}{x+1},$$

由

$$\frac{\partial P}{\partial x} = -1 + \frac{r}{(x+1)^2} = 0,$$

得到

$$\bar{x} = \bar{x}(r) = -1 - \sqrt{r}. \qquad (11.1.36)$$

当 $r \to 0$ 时，有 $\bar{x}(r) \to x^*$.

若取

$$B(x) = -\ln[-x-1], \qquad (11.1.37)$$

则内罚函数为

$$P(x,r) = -x - r\ln[-x-1],$$

由

$$\frac{\partial P}{\partial x} = -1 - \frac{r}{x+1} = 0,$$

得到

$$\bar{x} = \bar{x}(r) = -1 - r, \qquad (11.1.38)$$

当 $r \to 0$ 时,有 $\bar{x}(r) \to x^*$.

内罚函数的几何意义见图 11.1.2.

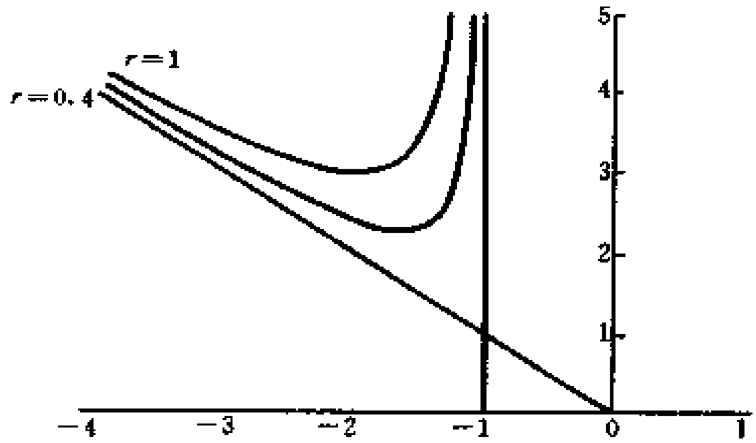


图 11.1.2 内罚函数的几何意义

比较（11.1.36）式和（11.1.38）式，显然由（11.1.38）式得到的 $\bar{x}(r) \to x^*$ 的速度比由(11.1.36)式得到的 $\bar{x}(r) \to x^*$

速度快.这个结论对于一般情况也成立,因此常常选用(11.1.32)式作为障碍函数.

下面给出具体的算法.

算法 11.1.2（障碍罚函数法）

(1) 选择序列 $\{r_k\}$, r_k 递减趋于 0, 取初始点 $\boldsymbol{x}^{(0)} \in D_0$, 这里

$$D_0 = \{\boldsymbol{x} \mid c_i(\boldsymbol{x}) < 0, \quad i \in I\},$$

置精度要求 ε, 置 $k = 1$.

(2) 以 $\boldsymbol{x}^{(k-1)}$ 为初始点, 求解无约束问题

$$\min \quad P(\boldsymbol{x}, r_k) = f(\boldsymbol{x}) + r_k B(\boldsymbol{x}),$$

其中

$$B(\boldsymbol{x}) = \sum_{i=1}^{m} \frac{-1}{c_i(\boldsymbol{x})} \text{ 或 } B(\boldsymbol{x}) = -\sum_{i=1}^{m} \ln[-c_i(\boldsymbol{x})],$$

得到最优解 $\boldsymbol{x}^{(k)}$.

(3) 若 $r_k B(\boldsymbol{x}^{(k)}) \leqslant \varepsilon$,则停止计算($\boldsymbol{x}^{(k)}$作为约束问题的最优解);否则置

$$k = k + 1,$$

转 (2).

使用算法 11.1.2 求解约束问题时, 应注意以下几个问题:

(1) D_0 非空, 即可行域的内部非空;

(2) 在算法第二步求解无约束问题时, 作一维搜索过程中, 要保证得到的点不出约束的边界. 这是因为由 (11.1.31) 式确定的$B(\boldsymbol{x})$在 $c_i(\boldsymbol{x}) = 0$ 处函数无定义,而由(11.1.32)式确定的 $B(\boldsymbol{x})$在 $c_i(\boldsymbol{x}) \geqslant 0$ 处无定义.

(3) 由于无约束问题的最优解 $\boldsymbol{x}^{(k)}$在可行域的内部, 因此在求解无约束问题中, 其终止准则可采用通常的终止准则.

内罚函数虽然克服了外罚函数的缺点, 但内罚函数不能处理等式约束问题. 若要处理等式约束问题, 需要将外罚函数与内罚

函数结合起来使用，称为混合罚函数.

内罚函数与外罚函数法一样，当 r 充分小后，内罚函数 $P(x,r)$ 也是病态的，也同样会出现求解无约束问题的困难.

§11.2 等式约束问题的乘子法

上节讲到（外）罚函数法的主要缺点在于需要惩罚因子趋于无穷大，才能得到约束问题的极小点，这会使罚函数的 Hesse 矩阵变得病态，给无约束问题的数值求解带来很大困难. 为克服这一缺点，Hestenes 和 Powell 于 1964 年各自独立地提出乘子法. 这一节和下一节将介绍乘子法的基本内容.

11.2.1 改进罚函数法的设想

考虑一个简单的等式约束问题

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) = f(x_1,x_2), \\ \text{s.t.} \quad & c(\boldsymbol{x}) = c(x_1,x_2) = 0, \end{aligned} \tag{11.2.1}$$

其罚函数问题为

$$\min \quad P(\boldsymbol{x},\sigma) = f(\boldsymbol{x}) + \frac{\sigma}{2}c^2(\boldsymbol{x}), \tag{11.2.2}$$

设 $\boldsymbol{x}^*$ 是（11.2.1）的全局解，考察 $\boldsymbol{x}^*$ 是否为无约束问题（11.2.2）的全局解（局部解）. $P(\boldsymbol{x},\sigma)$ 在 $\boldsymbol{x}^*$ 处的梯度为

$$\nabla_x P(\boldsymbol{x},\sigma) = \nabla f(\boldsymbol{x}^*) + \sigma c(\boldsymbol{x}^*)\nabla c(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*).$$

在通常的意义下，$\nabla f(\boldsymbol{x}^*) \neq 0$，也就是说，$\boldsymbol{x}^*$ 不是问题（11.2.2）的全局（局部）解，因此我们不能期望选择一个固定的 σ，求解一个无约束问题得到约束问题的最优解.

现在自然提出一个问题：是否存在这样的函数 $\phi(\boldsymbol{x},\sigma)$，使得 $\boldsymbol{x}^*$ 恰好是无约束问题

$$\min \quad \phi(\boldsymbol{x},\sigma) \tag{11.2.3}$$

的全局或局部解，以期望求解一个无约束问题得到约束问题的最优解，这样可大大提高算法效率.

由无约束问题的二阶充分条件，若函数 $\phi(x,\sigma)$在 x^* 处满足

$$\nabla_x\phi(x^*,\sigma) = 0, \tag{11.2.4}$$

$$\nabla_x^2\phi(x^*,\sigma) \text{ 正定}, \tag{11.2.5}$$

则 x^* 是无约束问题（11.2.3）的严格局部解.

现在构造满足条件（11.2.4）和（11.2.5）的函数$\phi(x,\sigma)$.

什么函数在 x^* 处的梯度为 0 呢？自然会想到 Lagrange 函数

$$L(x,\lambda) = f(x) + \lambda c(x), \tag{11.2.6}$$

在（x^*，λ^*）处的梯度等于 0，那么 Lagrange 函数 $L(x,\lambda)$在(x^*,λ^*)处是否满足条件（11.2.5）呢？一般情况下是不满足的. 由约束问题的二阶充分条件可知：Lagrange 函数的 Hesse 矩阵在 x^* 处只有切空间上正定，即

$$d^T\nabla_x^2L(x^*,\lambda^*)d > 0, \quad \forall d \in M,$$

其中

$$M = \{d \mid d^T\nabla c(x^*) = 0\},$$

并不能保证在整个空间上$\nabla_x^2L(x^*,\lambda^*)$正定. 为了便于理解，请先看一下 Lagrange 函数 $L(x,\lambda^*)$的几何图形.

从图 11.2.1 可以看出，Lagrange 函数在 d 方向上二阶方向导数大于 0，而在曲线的法向方向∇c（x^*）上，二阶方向导数小于 0. 为了使构造的函数在法向方向上二阶方向导数大于 0，仍然使用罚函数法的基本思想，在法向方向上加以惩罚，也就是说，当 x 离开d 方向时，即$c(x)\neq 0$ 时，对 Lagrange 函数加以惩罚，这样得到函数

$$\phi(x,\sigma) = f(x) + \lambda^* c(x) + \frac{\sigma}{2}c^2(x), \tag{11.2.7}$$

可以期望，当 σ 充分大时，有$\nabla_x^2\phi(x^*,\sigma)$正定.

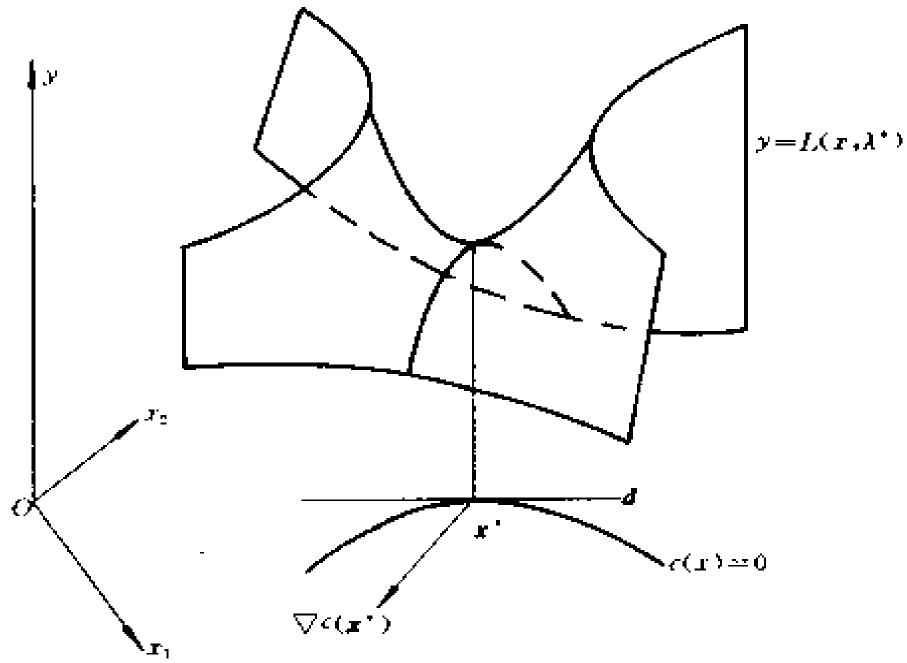


图 11.2.1 Lagrange 函数 L (x, λ^*) 的几何图形

引理 11.2.1 设 W 是 $n\times n$ 阶矩阵，a 为 n 阶向量，若对一切 $\boldsymbol{d}$ 满足 $\boldsymbol{d}\neq 0$，$\boldsymbol{a}^T\boldsymbol{d}=0$，均有 $\boldsymbol{d}^T\boldsymbol{W}\boldsymbol{d}>0$，则存在 $\sigma^*>0$，使当 $\sigma\geqslant\sigma^*$ 时，矩阵 $\boldsymbol{W}+\sigma\boldsymbol{a}\boldsymbol{a}^T$ 正定.

证明 考虑集合

$$K=\{\boldsymbol{d}\mid\|\boldsymbol{d}\|=1\},$$

只需证明，$\forall\,\boldsymbol{d}\in K$，当 $\sigma\geqslant\sigma^*$ 时，有

$$\boldsymbol{d}^T(\boldsymbol{W}+\sigma\boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{d}>0. \tag{11.2.8}$$

事实上，$\forall\,z\neq 0$，则 $\boldsymbol{d}=\dfrac{\boldsymbol{z}}{\|\boldsymbol{z}\|}\in K$，则 $\boldsymbol{z}^T(\boldsymbol{W}+\sigma\boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{z}>0$ 与 $\boldsymbol{d}^T(\boldsymbol{W}+\sigma\boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{d}>0$ 等价.

令

$$K'=\{\boldsymbol{d}\mid\boldsymbol{d}^T\boldsymbol{W}\boldsymbol{d}\leqslant 0,\boldsymbol{d}\in K\},$$

若 $K'=\phi$，则 $\forall\,\boldsymbol{d}\in K$，有 $\boldsymbol{d}^T\boldsymbol{W}\boldsymbol{d}>0$，因此 $\forall\,\sigma>0$，有 $\boldsymbol{d}^T(\boldsymbol{W}+\sigma\boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{d}\geqslant d^TWd>0$. 因此假设 $K'\neq\phi$. 当 $\boldsymbol{d}\in K/K'$ 时，有 $\boldsymbol{d}^T\boldsymbol{W}\boldsymbol{d}>0$，因此 $\forall\,\sigma>0$，有 $\boldsymbol{d}^T(\boldsymbol{W}+\sigma\boldsymbol{a}\boldsymbol{a}^T)\boldsymbol{d}\geqslant\boldsymbol{d}^T\boldsymbol{W}\boldsymbol{d}>0$.

下面考虑 $d \in K'$. 由于 K' 是有界闭集，则函数 $d^T W d$ 与 $(a^T d)^2$ 在 K' 上取到极小值. 不妨设 $(d^{(1)})^T W d^{(1)}$ 和 $(a^T d^{(2)})^2$ 分别为函数的极小值，并且 $a^T d^{(2)} \neq 0$；否则由定理条件，有 $(d^{(2)})^T W d^{(2)} > 0$ 与 $d^{(2)} \in K'$ 矛盾. 因此取

$$\sigma^* > \frac{(d^{(1)})^T W d^{(1)}}{(a^T d^{(2)})^2} > 0, \tag{11.2.9}$$

当 $\sigma \geqslant \sigma^*$ 时，有

$$\begin{aligned} d^T (W + \sigma a a^T) d &= d^T W d + \sigma (a^T d)^2 \\ &\geqslant (d^{(1)})^T W d^{(1)} + \sigma (a^T d^{(2)})^2 > 0. \end{aligned}$$

因此，$\forall d \in K$，(11.2.8) 式成立.

定理 11.2.1 设 x^*、λ^* 满足约束问题 (11.2.1) 的二阶充分条件，则存在 $\sigma^* > 0$，使当 $\sigma \geqslant \sigma^*$ 时，x^* 是无约束问题

$$\min_x \phi(x, \sigma) = f(x) + \lambda^* c(x) + \frac{\sigma}{2} c^2(x) \tag{11.2.10}$$

的严格局部解. 反之，若 $x^{(k)}$ 是 $\phi(x, \sigma_k)$ 的极小点，并且 $c(x^{(k)}) = 0$，则 $x^{(k)}$ 是约束问题 (11.2.1) 的局部解.

证明 设 x^*、λ^* 满足约束问题 (11.2.1) 的二阶充分条件. 因为

$$\phi(x, \sigma) = f(x) + \lambda^* c(x) + \frac{\sigma}{2} c^2(x),$$

所以

$$\begin{aligned} \nabla_x \phi(x^*, \sigma) &= \nabla f(x^*) + \lambda^* \nabla c(x^*) + \sigma c(x^*) \nabla c(x^*) = 0, \\ \nabla_x^2 \phi(x^*, \sigma) &= \nabla^2 f(x^*) + \lambda^* \nabla^2 c(x^*) + \sigma \nabla c(x^*) \nabla c(x^*)^T \\ &= \nabla_x^2 L(x^*, \lambda^*) + \sigma \nabla c(x^*) \nabla c(x^*)^T. \end{aligned}$$

取 $W = \nabla_x^2 L(x^*, \lambda^*)$，$a = \nabla c(x^*)$. 由引理 11.2.1，则存在 σ^*，当 $\sigma \geqslant \sigma^*$ 时，$\nabla_x^2 \phi(x^*, \sigma)$ 正定，所以 x^* 是无约束问题 (11.2.10) 的严格局部解.

反过来,若 $x^{(k)}$ 是 $\phi(x,\sigma_k)$ 的极小点,并且 $c(x^{(k)})=0$,则当 $\|x-x^{(k)}\|\leqslant\varepsilon$ 时,有

$$\begin{aligned}\phi(x,\sigma_k) &\geqslant \phi(x^{(k)},\sigma_k)\\ &= f(x^{(k)})+\lambda^* c(x^{(k)})+\frac{\sigma_k}{2}c^2(x^{(k)})\\ &= f(x^{(k)}),\end{aligned} \tag{11.2.11}$$

特别当 x 是约束问题(11.2.1)的可行点时,有

$$f(x)\geqslant f(x^{(k)}),$$

所以 $x^{(k)}$ 是约束问题(11.2.1)的局部解.

11.2.2 乘子法的基本思想

前面讲到,当构造出函数 $\phi(x,\sigma)$ 后,可以通过求解一个无约束问题得到约束问题的最优解. 但事实上,并不能做到这一点. 因为 $\phi(x,\sigma)$ 中的 λ^* 未知. 在得到 x^* 之前,我们是无法知道它的. 克服上述困难的方法是用参数 λ 代替 λ^*,得到增广 Lagrange 函数(也称为乘子罚函数)

$$\phi(x,\lambda,\sigma)=f(x)+\lambda c(x)+\frac{\sigma}{2}c^2(x). \tag{11.2.12}$$

考虑其相应的无约束问题

$$\min\phi(x,\lambda,\sigma), \tag{11.2.13}$$

其最优解为 $\bar{x}=\bar{x}(\lambda,\sigma)$.

由前所述(见定理 11.2.1),只要当 σ 充分大(不一定趋于 ∞),就有

$$\lim_{\lambda\to\lambda^*}\bar{x}(\lambda,\sigma)=x^*. \tag{11.2.14}$$

这样做,虽然达不到我们的目的——求一个无约束问题得到约束问题的最优解,但可以保证只要当 σ 大到一定量后,调整参数 λ 使之趋于 λ^*,就能得到约束问题的最优解. 这就是乘子法的基本思想.

从另一个角度考虑问题，对于任意的 λ，$\phi(\boldsymbol{x},\lambda,\sigma)$是约束问题

$$\begin{aligned}&\min \quad f(\boldsymbol{x})+\lambda c(\boldsymbol{x}),\\&\text{s.t.} \quad c(\boldsymbol{x})=0\end{aligned} \tag{11.2.15}$$

的惩罚函数，而问题（11.2.15）与问题（11.2.1）等价．在满足一定的条件下，由罚函数的收敛定理可知，

$$\lim_{\sigma\to\infty}\overline{\boldsymbol{x}}(\lambda,\sigma)=\boldsymbol{x}^*. \tag{11.2.16}$$

从这种角度来看，若固定 λ（特别取 $\lambda=0$），则乘子法就是罚函数法．结合（11.2.14）式，乘子法的效率要高于罚函数法．

乘子迭代公式

由（11.2.14）式知，当 $\lambda\to\lambda^*$ 时，有$\overline{\boldsymbol{x}}(\lambda,\sigma)\to x^*$，但此时 λ^* 未知，那么参数 λ 将如何调整呢？

在给定 $\lambda^{(k)}$、σ_k 后，求解无约束问题

$$\min \phi(\boldsymbol{x},\lambda^{(k)},\sigma_k), \tag{11.2.17}$$

其最优解为 $\boldsymbol{x}^{(k)}$．由无约束问题的一阶必要条件有

$$\begin{aligned}&\nabla_x\phi(\boldsymbol{x}^{(k)},\lambda^{(k)},\sigma_k)\\&=\nabla f(\boldsymbol{x}^{(k)})+\lambda^{(k)}\nabla c(\boldsymbol{x}^{(k)})+\sigma_k c(\boldsymbol{x}^{(k)})\nabla c(\boldsymbol{x}^{(k)})\\&=0.\end{aligned} \tag{11.2.18}$$

当 σ_k 充分大时，由（11.2.16）式可知

$$\boldsymbol{x}^{(k)}\approx\boldsymbol{x}^*,\nabla f(\boldsymbol{x}^{(k)})\approx\nabla f(\boldsymbol{x}^*),\nabla c(\boldsymbol{x}^{(k)})\approx\nabla c(\boldsymbol{x}^*),$$

因此有

$$\nabla f(\boldsymbol{x}^*)+[\lambda^{(k)}+\sigma_k c(\boldsymbol{x}^{(k)})]\nabla c(\boldsymbol{x}^*)\approx 0. \tag{11.2.19}$$

而在 x^*处，由约束问题的一阶条件有

$$\nabla f(\boldsymbol{x}^*)+\lambda^*\nabla c(\boldsymbol{x}^*)=0,$$

所以有

$$\lambda^*\approx\lambda^{(k)}+\sigma_k c(\boldsymbol{x}^{(k)}),$$

这样得到乘子迭代公式.

$$\lambda^{(k+1)} = \lambda^{(k)} + \sigma_k c(x^{(k)}). \tag{11.2.20}$$

终止准则

下面解决乘子法的最后一个问题，算法的终止准则.

若 $x^{(k)}$ 是无约束问题（11.2.17）的局部解，并且满足 $c(x^{(k)})=0$,由(11.2.18)式得到

$$\nabla f(x^{(k)}) + \lambda^{(k)} \nabla c(x^{(k)}) = 0, \tag{11.2.21}$$

因此，$x^{(k)}$是约束问题（11.2.1）的 K-T 点，$\lambda^{(k)}$为相应的乘子. 类似于定理 11.2.1 后半部分的证明过程，知 $x^{(k)}$为约束问题（11.2.1）的局部解，停止计算. 因此其终止准则为

$$| c(x^{(k)}) | \leqslant \epsilon,$$

其中 ϵ 是指定的精度要求.

11.2.3 一般等式约束问题的乘子法

考虑等式约束问题

$$\begin{aligned} &\min \quad f(x), \qquad x \in R^n, \\ &\text{s.t.} \quad c_i(x) = 0, \quad i \in E = \{1,2,\cdots,l\}, \end{aligned} \tag{11.2.22}$$

构造函数

$$\phi(x,\sigma) = f(x) + \sum_{i=1}^{l} \lambda_i^* c_i(x) + \frac{\sigma}{2} \sum_{i=1}^{l} c_i^2(x), \tag{11.2.23}$$

这里 λ^*是 x^*处的 Lagrange 乘子.

对于一般约束问题，构造的 ϕ（x，σ）有类似地结论.

定理 11.2.2 设 x^*、λ^*满足约束问题（11.2.22）的二阶充分条件，则存在 $\sigma^*>0$，使当 $\sigma \geqslant \sigma^*$ 时，x^* 是无约束问题 $\min\limits_{x} \phi$（x，σ)的严格局部解. 反之，若 $x^{(k)}$是 $\phi(x,\sigma_k)$的极小点,并且$c_i(x^{(k)})=0, i \in E$,则 $x^{(k)}$是约束问题(11.2.22)的局部解.

与简单情况相同，在求解无约束问题时，先构造增广 Lagrange 函数

$$\phi(x,\lambda,\sigma) = f(x) + \sum_{i=1}^{l} \lambda_i c_i(x) + \frac{\sigma}{2}\sum_{i=1}^{l} c_i^2(x), \tag{11.2.24}$$

在取定 $\boldsymbol{\lambda}^{(k)}$、$\sigma_k$ 后，求解无约束问题

$$\min \phi(x,\boldsymbol{\lambda}^{(k)},\sigma_k), \tag{11.2.25}$$

其最优解为 $x^{(k)}$.

类似于简单问题的推导，其乘子迭代公式为

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} + \sigma_k c_i(x^{(k)}), \quad i \in E = \{1,2,\cdots,l\}, \tag{11.2.26}$$

其终止准则为

$$\left(\sum_{i=1}^{l}[c_i(x^{(k)})]^2\right)^{\frac{1}{2}} \leqslant \varepsilon. \tag{11.2.27}$$

这样就得到相应的算法. 该算法是由 Hestenes 和 Powell 几乎同时独立提出来的（1969 年）. 因此称为 Hestenes-Powell 乘子法，或简称 *HP* 乘子法.

算法 11.2.1（HP 乘子法）

(1) 选择序列 $\{\sigma_k\}$，σ_k 递增趋于 $+\infty$，取初始点 $x^{(0)}$，初始乘子 $\boldsymbol{\lambda}^{(1)}$（$\boldsymbol{\lambda}^{(1)}=0$），置精度要求 ε，置 $k=1$.

(2) 以 $x^{(k-1)}$为初始点，求解无约束问题

$$\min\phi(x,\boldsymbol{\lambda}^{(k)},\sigma_k) = f(x) + \sum_{i=1}^{l} \lambda_i^{(k)} c_i(x) + \frac{\sigma_k}{2}\sum_{i=1}^{l} c_i^2(x),$$

得到最优解 $x^{(k)}$.

(3) 乘子迭代公式. 置

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} + \sigma_k c_i(x^{(k)}), \quad i = 1,2,\cdots,l.$$

(4) 若 $x^{(k)}$满足

$$\left(\sum_{i=1}^{l}[c_i(x^k)]^2\right)^{\frac{1}{2}} \leqslant \varepsilon,$$

则停止计算（$x^{(k)}$为约束问题的最优解，$\lambda^{(k)}$为相应的乘子）；否则置

$$k = k+1,$$

转（2）.

11.2.4 乘子法的效率

从理论上对乘子法的效率进行分析和比较是较为困难的．限于篇幅，这里仅从一简单的例子予以说明．

考虑用乘子法求解例 11.1.1

$$\min \quad f(x) = \frac{1}{2}\left(x_1^2 + \frac{1}{3}x_2^2\right),$$

$$\text{s.t.} \quad c(x) = x_1 + x_2 - 1 = 0,$$

其增广 Lagrange 函数为

$$\begin{aligned}\phi(x,\lambda,\sigma) &= f(x) + \lambda c(x) + \frac{\sigma}{2}c^2(x)\\ &= \frac{1}{2}\left(x_1^2 + \frac{1}{3}x_2^2\right)\\ &\quad + \lambda(x_1 + x_2 - 1) + \frac{\sigma}{2}(x_1 + x_2 - 1)^2.\end{aligned}$$

求偏导数

$$\frac{\partial\phi}{\partial x_1} = x_1 + \lambda + \sigma(x_1 + x_2 - 1) = 0,$$

$$\frac{\partial\phi}{\partial x_2} = \frac{1}{3}x_2 + \lambda + \sigma(x_1 + x_2 - 1) = 0,$$

得到无约束问题$\min\ \phi(x,\lambda,\sigma)$的最优解

$$\bar{x}(\lambda,\sigma) = \left(\frac{\sigma-\lambda}{1+4\sigma}, \frac{3(\sigma-\lambda)}{1+4\sigma}\right)^T. \tag{11.2.28}$$

当 λ 固定，$\sigma \to +\infty$时，由（11.2.28）式得到

$$\bar{x}(\lambda, \sigma) \to \left(\frac{1}{4}, \frac{3}{4}\right)^T = x^*.$$

讨论 σ 固定、$\lambda \to \lambda^*$ 的情况.

由乘子迭代公式有

$$\hat{\lambda} = \lambda + \sigma c(\bar{x}(\lambda, \sigma)) = \lambda - \sigma \frac{1 + 4\lambda}{1 + 4\sigma} = \frac{\lambda - \sigma}{1 + 4\sigma},$$

令 $\hat{\lambda} \to \lambda^*$，$\lambda \to \lambda^*$，解方程得到 $\lambda^* = -\frac{1}{4}$，因此由（11.2.28）式得到，当 $\lambda \to -\frac{1}{4}$时，有

$$\bar{x}(\lambda, \sigma) \to \left(\frac{1}{4}, \frac{3}{4}\right)^T = x^*,$$

由此我们看到，当 $\lambda \to \lambda^*$ 和 $\sigma \to +\infty$时都有 $\bar{x}$（λ，σ）$\to x^*$.

下面将乘子法与罚函数法作一下比较．对于罚函数的最优解（相当于 $\lambda = 0$）有

$$\bar{x}(\sigma) = \left(\frac{\sigma}{1 + 4\sigma}, \frac{3\sigma}{1 + 4\sigma}\right)^T. \qquad (11.2.29)$$

我们取 $\sigma_k = 0.1 \times 2^{k-1}$，$k = 1, 2, \cdots$，$\lambda^{(1)} = 0$，乘子迭代公式为

$$\lambda^{(k+1)} = \lambda^{(k)} + \sigma_k c(x^{(k)}),$$

乘子法得到的点列为

$$x^{(k)} = \left(\frac{\sigma_k - \lambda^{(k)}}{1 + 4\sigma_k}, \frac{3(\sigma_k - \lambda^{(k)})}{1 + 4\sigma_k}\right)^T,$$

由罚函数法得到的点列为

$$x^{(k)} = \left(\frac{\sigma_k}{1 + 4\sigma_k}, \frac{3\sigma_k}{1 + 4\sigma_k}\right)^T.$$

其具体计算结果见表 11.2.1.

表 11.2.1　例 11.2.1 的计算结果

k	$x^{(k)}$（罚函数法）	$x^{(k)}$（乘子法）	k	$x^{(k)}$（罚函数法）
01	(0.07143, 0.21429)	(0.07143, 0.21429)	11	(0.24939, 0.74817)
02	(0.11111, 0.33333)	(0.15079, 0.45238)	12	(0.24970, 0.74909)
03	(0.15385, 0.46154)	(0.21184, 0.63553)	13	(0.24985, 0.74954)
04	(0.19048, 0.57143)	(0.24092, 0.72275)	14	(0.24992, 0.74977)
05	(0.21622, 0.64865)	(0.24877, 0.74632)	15	(0.24996, 0.74989)
06	(0.23188, 0.69565)	(0.24991, 0.74973)	16	(0.24998, 0.74994)
07	(0.24060, 0.72180)	(0.25000, 0.74999)	17	(0.24999, 0.74997)
08	(0.24521, 0.73563)	(0.25000, 0.75000)	18	(0.25000, 0.74999)
09	(0.24758, 0.74275)		19	(0.25000, 0.74999)
10	(0.24879, 0.74636)		20	(0.25000, 0.75000)

由表 11.2.1 可以看出，乘子法的收敛速度要比罚函数法快得多．乘子法在第 8 步与罚函数法第 20 步得到的结果相同．并注意到 $\sigma_{20}=0.1\times 2^{19}=52428.8$，而 $\sigma_8=0.1\times 2^7=12.8$．相比之下，乘子法不会有过大的惩罚因子．

从这个例子可粗略地看出乘子法优于罚函数法．

§ 11.3　一般约束问题的乘子法

本节讨论一般约束问题的乘子法，考虑约束问题

$$
\begin{aligned}
\min\quad & f(x),\quad x\in R^n,\\
\text{s.t.}\quad & c_i(x)=0, i\in E=\{1,2,\cdots,l\},\\
& c_i(x)\leqslant 0, i\in I=\{l+1,l+2,\cdots,l+m\},
\end{aligned}
\tag{11.3.1}
$$

由上节讨论的等式约束问题的乘子法，只需将一般约束问题（11.3.1）化为等式约束问题