

石油化工实用 最优化计算方法

徐亦方 等著

SHIYOUHUAGONG
SHIYONGZUIYOUHUA
JISUANFANGFA

石油大学出版社
SHIYODAXUECHUBANSHE

石油化工实用最优化计算方法

徐亦方 潘建华 陶 克著

石油大学出版社

内 容 简 介

本书从实用角度介绍了石油化工生产中常用的最优化算法, 其中 Armijo 法、序贯二次规划, Watchdog 技术, Steward 通路路程等都是近年来较新的算法。本书还介绍使用各种算法的技巧, 对使用中发生的一些问题, 从现象、原因到解决方法都进行了介绍和讨论。书后附有一些较新的不易获得的源程序和实际优化问题的习题。

本书可作为石油化工专业研究生的教材参考书, 也可供其它专业研究生、大学生等参考。同时本书也是现场实际工作者的一本有用的工具书。

石油化工实用最优化计算方法

徐亦方 潘建华 陶克 编著

石油大学出版社

山东省 东营市

蓬莱东方印刷厂印刷

全国新华书店发行

开本 787×1092 1/16 15 印张 239 千字

1990 年 12 月第 1 版 1990 年 12 月第 1 次印刷

印数 1-1500 册

ISBN 7-5636-0124-4/TE·25

定 价: 3.41 元(压膜)

代 序

近数十年来，最优化方法已深入到许多工程学科领域，在应用过程中都获得较大的经济效益，反过来又促进了最优化方法的发展。

本书汇集了在石油化工领域中较广泛实用的一些较新的算法，加以系统阐述，在国内最优化书籍中尚少见到，其内容也具一定深度，可供研究生作为基本教材，亦可供石油化工科技工作人员作为参考用书。

本书作者在化工最优化领域具有较好的实践经验，本书主编作者曾在国外进修，从事本领域的教学与应用研究十余年，承担了本书的统编工作，作出了可以期望的贡献。

最优化方法正在发展，与石油化工的结合与应用具有广阔的前景，希望本书的出版，将对多学科的结合和石油化工生产应用方面，均能起到促进作用。

沈 复

前 言

最优化是一门既有严格理论基础又有直观推断规则的科学。它有很强的应用背景，在现实生活许多领域中都在发挥着重要的作用。从实用的角度看，要掌握这门科学，需要有一定的实践经验。首先对各种算法要通过典型算例的试验来掌握其性能，其次要学习从现实问题中如何抽象出数学模型。当然，理论基础，特别是对那些基本定理的掌握，也是重要的。

本书就是基于上面的认识编写的。我们力求将目前在石油化工领域得到较广泛应用的算法加以介绍。并且把一些较新的算法，其中有些算法只散见于各种杂志，也编入本书。

本书适用于石油加工专业和应用化学专业的研究生，课时为 60 学时，这些内容是在近六年中给研究生讲课的讲稿的基础上编写而成的。从我们授课的经验看，要对学生的水平做适当的估计，由此决定内容取舍，较难章节都加了星号。如果数学基础较弱，则对内容要补充一些基础数学，并可删去一些内容，如对二次规划可以只讲线性互补法或者只讲有效集方法。

各种文献和专著所使用的符号都不完全相同，这是最优化文献中的一个特点。为便于读者阅读各种软件的使用说明书，本书各章的符号仍依循原文献或专著，而不强求统一。这会给初学者带来一些困难，但学习不同表达方式，对读者总是有益的。

本书由潘建华执笔写第三、第四章，陶克写第五章，其他各章由徐亦方编写并统编全书。在成书过程中，齐玉民同志参加了部分工作。

一九九〇年六月

目 录

绪 论	1
第 一 章 数学基础	5
第 二 章 直线搜索	22
第 三 章 无约束极小问题的计算方法	32
第 四 章 最小二乘问题的计算方法*	52
第 五 章 线性规划	58
第 六 章 二次规划	84
第 七 章 整数规划	101
第 八 章 有约束问题的最优性条件	110
第 九 章 序贯二次规划法*	117
第 十 章 可变容差法*	126
第十一章 罚函数法	133
第十二章 动态规划	141
第十三章 最优化算法实践	149
综合习题	163
附录 程序	170

绪 论

石油化工过程通常可以用数学模型描述,模型中包括各种参数和变量。对于可以选取定值的我们称为设计参数,例如设备参数,过程参数等。这些参数的选取常根据经验选定。凡是通过一定函数关系可以求解的变量称为过程变量,例如通过物料平衡,能量平衡和相平衡等,我们可以求出过程中的温度、压力、流量等变量的解。还有一类变量称为设计变量(也称决策变量、控制变量),这类变量是可以由我们在一定范围内选取的,选取的恰当与否对过程的技术经济指标有重要的影响。这类变量个数相当于模型的自由度,其选值的过程称为最优化问题求解过程。自由度大容易达到最优设计(或最优决策),但求解较困难。通常我们将设计变量1~10个的问题称为小型设计问题,将10~50个的问题称为中型设计问题,将大于50个变量的问题称为大型设计问题。

组成最优化问题的模型分为两部分:即目标函数部分和约束条件部分。目标函数可以是单目标也可以为多目标,本书主要介绍单目标。多目标问题可以用权将几个单目标加起来形成一个单目标问题,也可以用其他方法将多目标问题转换为单目标问题。约束条件又可分为上下限约束(也称边界约束)和性能约束(有时也称资源约束),此外,还常规定变量为非负或限制其类型(如整数等)。

满足所有约束的解空间称为可行域,在可行域中能使目标函数最优的解称最优解。最优解可以在边界上,也可以不在边界上。

最优化问题分类:

没有约束条件的问题称为无约束最优化问题,记为:

$$\min f(x) \quad x \in E^n \quad (0.1)$$

目标函数和约束条件都是线性函数的问题称为线性规划(Linear Programming),记为:

$$\begin{aligned} (\text{LP}) \quad & \min c^T x \quad x \in E^n \\ & \text{s.t.} \quad Ax \leq b \\ & \quad \quad x \geq 0 \end{aligned} \quad (0.2)$$

*变量类型限定为整型数的称为整数规划(Integer Programming), 常遇到的是线性整数规划, 记为:

$$\begin{aligned}
 (\text{IP}) \quad & \min c^T x \quad x \in E^n \\
 \text{s.t.} \quad & Ax \leq b \\
 & x \geq 0 \quad (\text{整数})
 \end{aligned} \tag{0.3}$$

其中变量限定为 0 或 1 整数的问题, 称为 0/1 整数规划, 记为:

$$\begin{aligned}
 (0/1) \quad & \min c^T x \quad x \in E^n \\
 \text{s.t.} \quad & Ax \leq b \\
 & x = (0,1) \quad \text{或写成 } x = 0V1
 \end{aligned} \tag{0.4}$$

目标函数为二次函数, 约束条件为线性函数的问题称为二次规划(Quadratic Programming), 记为:

$$\begin{aligned}
 (\text{QP}) \quad & \min \frac{1}{2} x^T B x + c^T x \quad x \in E^n \\
 \text{s.t.} \quad & Ax \leq b \\
 & x \geq 0
 \end{aligned} \tag{0.5}$$

一般的非线性规划通式记为:

$$\begin{aligned}
 (\text{NLP}) \quad & \min f(x) \quad x \in E^n \\
 \text{s.t.} \quad & g(x) \leq 0 \quad g: E^n \rightarrow E^p \\
 & h(x) = 0 \quad h: E^n \rightarrow E^l
 \end{aligned} \tag{0.6}$$

在最优化问题的一般形式中只是取极小, 因为极大问题都可以通过下式转换为极小问题:

$$\max f(x) = -\min\{-f(x)\} \tag{0.7}$$

为了简便, 今后只研究求极小问题。

各类最优化问题的难易程度如图 0.1 所示:

最优化算法进展情况见 0.2 图。由图可见, 一般非线性规划问题设计变量加约束条件之和约在 100 以内, 目前才比较有把握求解。而线性规划问题的规模(设计变量加约束条件之和)可达数万。

在石油化工领域内最优化问题类型有生产计划问题, 换热网络匹配问题, 多参数公式拟合问题, 流程设计最优化等问题。许多求解方程组的问题也可以

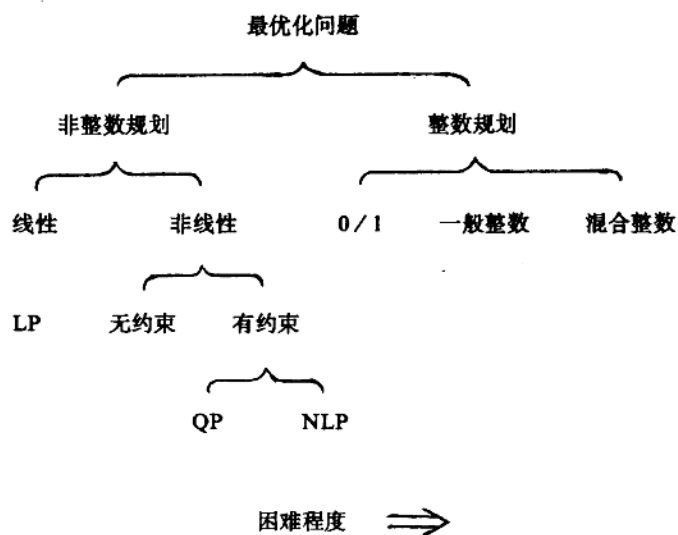


图 0.1 各类优化问题的难易程度

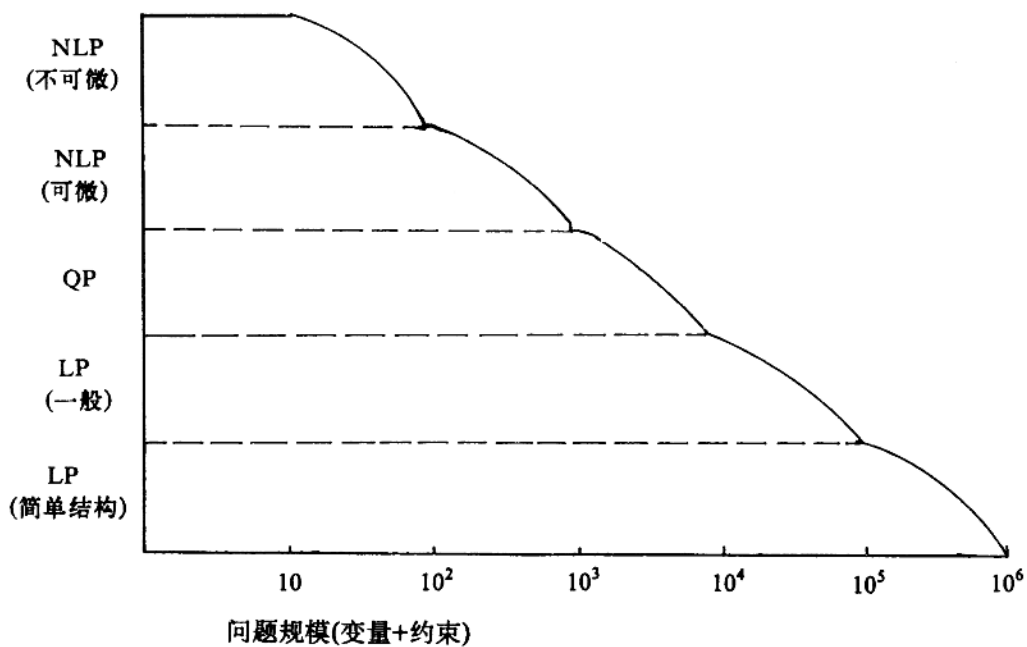


图 0.2 数学规划可解问题的规模概况

用与最优化类似的方法解决。近 20 余年来利用最优化方法解决了许多生产问题，取得不少经济效益。此外，最优化算法近年来也有不少进展，使最优化问题可解的规模不断扩大，算法的稳定性也不断提高。

本书除介绍各种算法的基本原理外，较重视实用技巧，并且尽可能补充加入一些近年来新发展的方法。许多实例都是取自我们科研工作的结果。

通过本课程我们希望读者能掌握各种算法的选择和一些基本技巧，并且在算法失败时能够分析其原因和采取一些措施来克服这些困难。

第一章 数学基础

§ 1.1 多元函数和等值线图

本节开始我们将讨论多元单值函数的一些特性。设函数为 $f(x)$, $x \in E^n$ 。对任意一个多元函数, 下列等式 $z=f(x)$ 定义为 $n+1$ 维空间中的一个面。当 $n=2$ 时, $z=f(x_1, x_2)$ 表示三维空间 (x_1, x_2, z) 中的一个面。若 z 取定值 c , 则 $f(x_1, x_2)=c$ 是 $z=c$ 平面上的曲线。对于一组 c 值, 我们将其相应的一组曲线绘在一张平面图上, 就构成函数 $f(x_1, x_2)$ 的一张等值线图。图 1.1 表示的是下列函数的等值线图, $f(x_1, x_2)=e^{x_1}(4x_1^2+2x_2^2+4x_1x_2+2x_2+1)$ 。 c 的取值为 0.2, 0.4, 0.7, 1.0, 1.7, 1.8, 2, 3, 4, 5, 6 和 20。可以看出: 不同值的等值线不相交, 这是因为这里讨论的是单值函数; 此外, 极小点不止一个, 这种情况在我们处理实际问题时常会遇到; 还有, 在极小点附近, 等值线近似于一个同心椭圆族, 这一现象将在讨论 Taylor 展开式时说明。

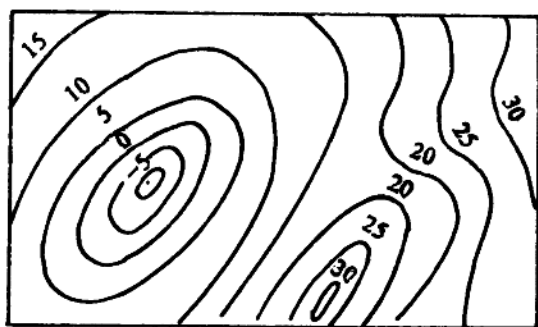


图 1.1 等值线图

§ 1.2 梯度和 Hesse 阵

与一元函数相似, 设多元函数 $f(x)$ 在 x_0 处可微, 则 $f(x)$ 在该点关于各变量的一阶偏导数存在, 记为

$$\frac{\partial f(x_0)}{\partial x_1}, \frac{\partial f(x_0)}{\partial x_2}, \dots, \frac{\partial f(x_0)}{\partial x_n}$$

我们定义 $f(x)$ 在 x 处的梯度 $\nabla f(x)$ 为

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right]^T \quad (1.1)$$

附带说明一下, 在本书中除了特别指明以外, 向量都是指列向量。这样 $\nabla f(x)$ 就表示一个列向量。梯度也可以称为函数 $f(x)$ 关于向量 x 的一阶导数。

梯度有下列性质:

性质 1: 函数在某点的梯度若不为零, 则其方向是过该点的等值面的法线方向。

证: 设过点 x_0 的等值面方程为

$$f(x) = f(x_0) = \gamma_0 \quad (1.2)$$

又设 $x_1 = x_1(\theta)$, $x_2 = x_2(\theta)$, $\dots$, $x_n = x_n(\theta)$ 是过点 x_0 同时又完全在等值面 $f(x_0)$ 上的任意一条光滑曲线 L 的方程, 其中 θ 是参数, 点 x_0 所对应的参数是 θ_0 。代入上式, 有

$$f(x_1(\theta), x_2(\theta), \dots, x_n(\theta)) = \gamma_0 \quad (1.3)$$

两端同时在 θ_0 处关于 θ 求导数, 根据复合函数微分法得到

$$\frac{\partial f(x_0)}{\partial x_1} x'_1(\theta_0) + \frac{\partial f(x_0)}{\partial x_2} x'_2(\theta_0) + \dots + \frac{\partial f(x_0)}{\partial x_n} x'_n(\theta_0) = 0 \quad (1.4)$$

而向量 $t(\theta_0)$

$$t(\theta_0) = [x'_1(\theta_0), x'_2(\theta_0), \dots, x'_n(\theta_0)]^T$$

正是曲线 L 在 x_0 处的切向量。于是上式可写成

$$\nabla f(x_0)^T t(\theta_0) = 0 \quad (1.5)$$

此式说明函数 $f(x)$ 在 x_0 处的梯度方向 $\nabla f(x_0)$ 与过该点在等值面上的任一条曲线

L 的切线正交。也就是说梯度方向为 x_0 处等值面的法线方向。

性质 2: 梯度方向是函数变化率最大的方向, 并且是指向函数值增大的方向。

先说明方向导数的定义。

[定义] 设 $f: E^n \rightarrow E^1$ 在 x_0 处可微, p 是固定不变的向量, 且是任意正数, e 是方向 p 上的单位向量, 则称:

$$\frac{\partial f(x_0)}{\partial p} = \lim_{t \rightarrow 0^+} \frac{f(x_0 + te) - f(x_0)}{t} \quad (1.6)$$

为函数 $f(x)$ 在点 x_0 处沿方向 p 的方向导数。从方向导数的定义可以看出若方

向导数 $\frac{\partial f(x_0)}{\partial p} < 0$, 则沿方向 p 函数值下降。

[定理] 设 $f: E^n \rightarrow E^1$ 在 x_0 处可微, 则

$$\frac{\partial f(x_0)}{\partial p} = \nabla f(x_0)^T e \quad (1.7)$$

证: 由 $f(x_0 + te) = f(x_0) + t \nabla f(x_0)^T e + O(t \|e\|)$

代入方向导数的定义, 则

$$\begin{aligned} \frac{\partial f(x_0)}{\partial p} &= \lim_{t \rightarrow 0^+} \frac{f(x_0) + t \nabla f(x_0)^T e + O(t \|e\|) - f(x_0)}{t} \\ &= \nabla f(x_0)^T e \end{aligned}$$

由此定理, 我们得出: 若任意方向 p , 有

$$\nabla f(x_0)^T p > 0 \quad (1.8)$$

则 p 方向是函数 $f(x)$ 在 x_0 处的函数上升方向。

现令 $p = \nabla f(x_0)$, 所以

$$\nabla f(x_0)^T \nabla f(x_0) > 0 \quad (1.9)$$

即梯度方向指向函数值增加的方向。这就证明了性质 2 的后半部分。

又由

$$\begin{aligned} \nabla f(x_0)^T e &= \|\nabla f(x_0)\|_2 \|e\|_2 \cos \beta \\ &= \|\nabla f(x_0)\|_2 \cos \beta \end{aligned} \quad (1.10)$$

式中 β 是 $\mathbf{p}$ 与 $\nabla f(\mathbf{x}_0)$ 的夹角。当取两个向量的夹角为零时，也就是说取

$\mathbf{p} = \nabla f(\mathbf{x}_0)$ 时， $\frac{\partial f(\mathbf{x}_0)}{\partial \mathbf{p}}$ 取最大值。这就证明了性质 2 的前半部分。

从上面的讨论可以得出，要使函数下降最快，也就是说要使方向导数小于 0 且取最负值，我们只需取 $\mathbf{p} = -\nabla f(\mathbf{x}_0)$ 即可。

下面讨论多元单值函数 $f(\mathbf{x})$, $\mathbf{x} \in E^n$ 关于 $\mathbf{x}$ 的二阶导数。

首先考虑向量的向量值函数的导数：

设 $\mathbf{g}: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, 且 $\mathbf{x}_0 \in D$, 如果 $\mathbf{g}(\mathbf{x})$ 的所有分量 $g_1(\mathbf{x}), \dots, g_m(\mathbf{x})$ 在 $\mathbf{x}_0$ 点都可微。若 $\mathbf{g}(\mathbf{x})$ 在点 $\mathbf{x}_0$ 处可微，则对于任意的 n 维向量 $\mathbf{p}$ 都有

$$\lim_{\|\mathbf{p}\| \rightarrow 0} \frac{g_i(\mathbf{x}_0 + \mathbf{p}) - g_i(\mathbf{x}_0) - \nabla g_i(\mathbf{x}_0)^T \mathbf{p}}{\|\mathbf{p}\|} = 0 \quad (1.11)$$

$i = 1, 2, \dots, m$

因为向量的极限是通过它所有分量的极限来定义的，所以(1.11)等价于

$$\lim_{\|\mathbf{p}\| \rightarrow 0} \frac{\mathbf{g}(\mathbf{x}_0 + \mathbf{p}) - \mathbf{g}(\mathbf{x}_0) - \nabla \mathbf{g}(\mathbf{x}_0)^T \mathbf{p}}{\|\mathbf{p}\|} = \mathbf{0} \quad (1.12)$$

其中

$$\nabla \mathbf{g}(\mathbf{x}_0) = \begin{bmatrix} \frac{\partial g_1(\mathbf{x}_0)}{\partial x_1} & \frac{\partial g_1(\mathbf{x}_0)}{\partial x_2} & \dots & \frac{\partial g_1(\mathbf{x}_0)}{\partial x_n} \\ \frac{\partial g_2(\mathbf{x}_0)}{\partial x_1} & \frac{\partial g_2(\mathbf{x}_0)}{\partial x_2} & \dots & \frac{\partial g_2(\mathbf{x}_0)}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_m(\mathbf{x}_0)}{\partial x_1} & \frac{\partial g_m(\mathbf{x}_0)}{\partial x_2} & \dots & \frac{\partial g_m(\mathbf{x}_0)}{\partial x_n} \end{bmatrix} \quad (1.13)$$

称为向量值函数 $\mathbf{g}(\mathbf{x})$ 在点 $\mathbf{x}_0$ 处的导数，也称向量值函数 $\mathbf{g}(\mathbf{x})$ 在点 $\mathbf{x}_0$ 处的 Jacobi 阵。

设 $m=n$, 并且 $\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x})$, 其中 $f: E^n \rightarrow E^1$, 假定它具有二阶连续偏导数。这时,

$$g_1(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_1}, \quad g_2(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_2}, \quad \dots, g_n(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_n}$$

由(1.13)得到

$$\nabla^2 f(x) = \nabla[\nabla f(x)] = \begin{bmatrix} \frac{\partial^2 f(x_0)}{\partial x_1^2} & \frac{\partial^2 f(x_0)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x_0)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \quad (1.14)$$

这就是 n 元函数 $f(x)$ 关于 x 的二阶导数, 称为 $f(x)$ 的 Hesse 阵。

因此, 多元函数 $f(x)$ 的一阶导数是它的梯度 $\nabla f(x)$, 二阶导数是它的 Hesse 阵 $\nabla^2 f(x)$ 。在最优化中这是两个常用的概念。

当 $f(x)$ 的所有二阶偏导数连续时, 有

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i}, \quad i, j = 1, 2, \cdots, n$$

因此, Hesse 阵在这种情况下是对称的。

§ 1.3 有关梯度和 Hesse 阵的几个公式

以下几个特殊的计算公式是常用的

(1) 若 $f(x) = c$ (常数), 则 $\nabla f(x) = 0$, 即

$$\nabla c = 0 \quad (1.15)$$

(2) $\nabla(b^T x) = b$

$$(1.16)$$

证: 设 $b = [b_1, b_2, \cdots, b_n]^T$, $x = [x_1, x_2, \cdots, x_n]^T$, 则

$$b^T x = \sum_{i=1}^n b_i x_i$$

于是 $\nabla(b^T x)$ 的第 j 个分量是

$$\frac{\partial}{\partial x_j} (b^T x) = \frac{\partial}{\partial x_j} \left(\sum_{i=1}^n b_i x_i \right) = b_j, \quad j = 1, 2, \dots, n$$

所以 $\nabla(b^T x) = b$

$$(3) \nabla(x^T x) = 2x \quad (1.17)$$

(4) 若 Q 是对称方阵, 则

$$\nabla(x^T Q x) = 2Qx \quad (1.18)$$

(5) 设 c 是分量全为常数的 n 维向量, 0 是 $n \times n$ 阶零矩阵, 则

$$\nabla c = 0 \quad (1.19)$$

(6) $\nabla x = I$, 其中 x 是 n 维向量, I 是 $n \times n$ 阶单位阵. (1.20)

(7) $\nabla(Qx) = Q$, 其中 Q 是 $n \times n$ 阶矩阵 (1.21)

(8) 设 $\Phi(t) = f(x_0 + tp)$, 其中 $f: E^n \rightarrow E^1$, $\Phi: E^1 \rightarrow E^1$, 则

$$\Phi'(t) = \nabla f(x_0 + tp)^T p \quad (1.22)$$

$$\Phi''(t) = p^T \nabla^2 f(x_0 + tp) p \quad (1.23)$$

证: 由

$$\Phi(t) = f(x_1^{(0)} + tp_1, x_2^{(0)} + tp_2, \dots, x_n^{(0)} + tp_n)$$

对 t 求导, 得

$$\Phi'(t) = \sum_{i=1}^n \frac{\partial f(x_0 + tp)}{\partial x_i} p_i = \nabla f(x_0 + tp)^T p$$

以上是由多元函数复合函数求导式得出的.

再对 t 求一次导数

$$\begin{aligned} \Phi''(t) &= \sum_{i=1}^n \frac{d}{dt} \left[\frac{\partial f(x_0 + tp)}{\partial x_i} \right] p_i \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(x_0 + tp)}{\partial x_j \partial x_i} p_j p_i \\ &= p^T \nabla^2 f(x_0 + tp) p \end{aligned}$$

§ 1.4 多元函数的 Taylor 展开式

多元函数的 Taylor 展开式在最优化方法中是导出算法和分析问题的基本公式。

[定理] 设 $f: E^n \rightarrow E^1$ 有二阶连续偏导数, 则

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(\bar{x}) p \quad (1.24)$$

其中 $\bar{x} = x + \theta p$, $0 < \theta < 1$.

证: 设 $\Phi(t) = f(x+tp)$, 于是

$$\Phi(0) = f(x), \quad \Phi(1) = f(x+p)$$

按一元函数 Taylor 展开式把 $\Phi(t)$ 在 $t=0$ 点展开, 得到

$$\Phi(t) = \Phi(0) + \Phi'(0)t + \frac{1}{2} \Phi''(\theta t)t^2 \quad (1.25)$$

其中 $0 < \theta < 1$. 由

$$\Phi'(0) = \nabla f(x)^T p$$

$$\Phi''(\theta) = p^T \nabla^2 f(x+\theta p) p$$

代入(1.25), 得

$$\Phi(t) = \Phi(0) + \nabla f(x)^T p \cdot t + \frac{1}{2} p^T \nabla^2 f(x+\theta p) p \cdot t^2$$

令 $t=1$, 得到

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+\theta p) p$$

又因为 $\nabla^2 f(x)$ 的每一个分量都是连续的, 所以有

$$\frac{\partial^2 f(x+\theta p)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \delta_{ij} \quad i, j = 1, 2, \dots, n$$

当 $\|p\| \rightarrow 0$ 时, $\delta_{ij} \rightarrow 0$. 于是 $\frac{1}{2} p^T \nabla^2 f(\bar{x}) p$ 可写成

$$\frac{1}{2} p^T \nabla^2 f(\bar{x}) p = \frac{1}{2} p^T \nabla^2 f(x) p + O(\|p\|^2)$$