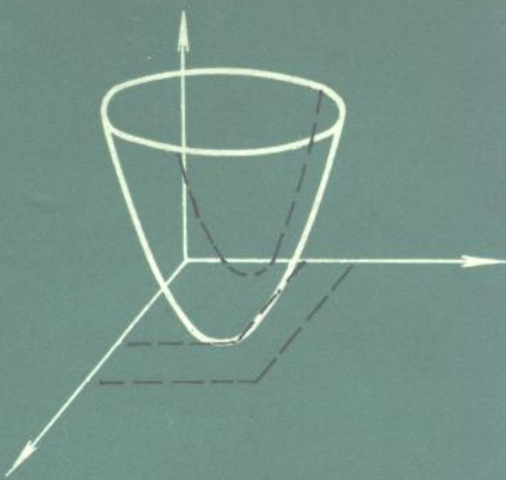
非线性最优化引论

INTRODUCTION TO NONLINEAR OPTIMIZATION

[美] D. A. Wismer　R. Chattergy 合著



北京工业学院出版社

内 容 简 介

本书内容涉及到了非线性最优化的各个领域，但对深度做了恰当的选择；着力于讲解各种方法的思想，而并不追求于数学的严谨性；并且多用例题及有典型意义的题目来阐述概念。图文并茂、直观性很强。不仅适宜于学生和初学者，即使教师和熟悉该学科内容的读者也会从本书的内容中得益。

本书由多人分章翻译：刘宝光(一，附录)，陈菊英（二），周瑞珍(三)，赵广琴（四，五)，丁丽娟（六，七)，邓乃扬（八，九），王新民(十，十一)，马国瑜(十二，十三)。另外，许金伟也参加了部分译稿的整理工作。

INTRODUCTION TO
NONLINEAR OPTIMIZATION
A Problem Solving Approach
Elsevier North-Holland, Inc., New York, 1978

非线性最优化引论

〔美〕D.A.Wismer
R.Chattergy 合著

邓乃扬 刘宝光 等译

北京工业学院出版社出版
新华书店北京发行所发行 各地新华书店经售
通县向阳印刷厂印刷

787×1092毫米 32开本 14.375印张 321千字
1987年6月第一版 1987年6月第一次印刷
印数：1—5,000 册
统一书号：13434·56 定价. 2.35 元

序　言

十多年来，作者讲授非线性最优化，对象是工程和商业行政系科的大学生、研究生和进修生。这一题材本质上涉及到许多学科，但还没有一种共认的理论。在这许多有关学科中，有些学科有性质十分不同的多种数学理论体系，基础在微积分，但也涉及许多近代的理论成果；有大量的、种类繁多的计算算法，以及理论、计算和应用的无穷尽的专门例子。

我们的目的，是在不要求广泛的数学预备知识的前提下，将这些各式各样的材料组织在一个引论性的教程中。为了达到这一目标，开始时我们使用过各种现成的教材，以广泛的注释、例题和习题作为补充。在教学过程中，这些补充材料逐渐取代着教材，而教材则逐渐降到补充材料的地位，最后产生了初步满足我们需要的这本教科书。

本书的主要目标之一，是通过广泛地使用例题、示例性的题解和附有答案的补充题，使读者对于非线性最优化问题的提法和解法有直观和深入的了解。

在这本教科书中，我们对于组成非线性最优化课题的大多数主要领域，提供了一个简明的讨论。这些领域包括：

源自微积分的古典无约束最优化方法。

使用 Lagrange 乘子加进约束项而产生的鞍点。

Kuhn-Tucker 理论、对偶性，以及它与古典方法的联系。

一维搜索算法和计算结果。

约束和无约束问题的梯度搜索算法。

惩罚函数和有关的计算方法。

共轭方向算法和二次收敛性。

有二次目标函数和线性约束问题的专门理论和计算方法。

线性规划概念和解法概述。

只取整数值解的问题的表述和求解。

多阶段决策问题及其动态规划解法。

处理高维问题的分解和谱系方法。

另外，书中共有122个例题，153道示例性的附有解答的习题和136道补充题，补充题的答案附在书的最后。

本书内容的这一取舍，是想通过非线性最优化理论基础的研究，与实际要求获得计算解这二者之间找出一个适当的处理办法。有时，这两个领域是不能区分的；而本书从头至尾，始终通过大量地使用数值例子、示例性的有解答的习题，达到使读者对内容直观而深入地了解。这些例子和有解答的习题作为一种手段，用以扩展在各章中只简要讨论过的内容，以及使用那些材料去求解有关的问题。

我们想本书可以用作关于非线性最优化课题的介绍性的教科书。除此之外，它也可以作为更高级的非线性最优化教程的一个有用的补充，大量的例子和有解答的习题对于学生和教师都是有帮助的。

作者向帮助过我们的许多学生表示感谢。这些年来，他们帮助形成了本书，并且为许多已解的习题和例题付出了劳动。特别要提到以前的助教P.Bansal博士，对他杰出的努力致以谢意。

David A. Wismer

R. Chattergy

目　录

第一章 最优化问题

在可能的条件下，如何以最好的方式做事情，就是最优化所探讨的问题。不言而喻，问题的这种解决办法是人们十分想得到的，近年来受到越来越大的重视。当然，很久以来数学家们就在研究求得最优解的方法，或许早在Descartes和Fermat就开始了，他们在十七世纪，甚至在Newton建立微积分之前，就研究过这样的问题。

最优化问题的解如果不随时间变化，则称这类问题为**静态最优化**或**非线性规划**。一个非线性规划问题，用一组代数方程或超越方程来描述，这组方程称为数学模型，或简称为模型。问题的解当然仅仅对于它的模型来说是最优的。因此，对于一个实在的系统，为了能得到近于最优的解，或者甚至仅仅为求得可用的解(常常是如此)，所取的模型**必须密切地近似真实系统的性能**。

描述对应于任意指定解的利润或利益的式子，叫做**目标函数**，也常称作成本函数或性能指标函数。当系统模型和目标函数都是线性式时，就是**线性规划问题**。

如果最优化问题的解是时间的函数，则称这类问题为**动态最优化**。这时，问题的数学模型，通常由一组微分方程和一个目标函数组成，后者可以是一个积分。描述这类问题的解法的数学理论，从几个彼此相容的观点发展起来，产生了诸如变分法、动态规划、以及Pontryagin极大值原理等名目。对于连续的时间变数，如果仅考虑它的若干个离散的

点，动态最优化问题可以用非线性规划近似地叙述和求解。

对于大多数有实际意义的最优化问题，它的解必须满足描述该实在系统的模型中所给出的限制，并且使目标函数取值最大或最小。这类问题，对静态和动态两种情形，均称为**约束最优化问题**。约束问题中，系统的数学模型代表约束。按照各约束式中是出现等号(=)、还是不等号(≥，≤)又可以将约束进一步分成**等式约束**和**不等式约束**。在有些情况下不出现约束，问题仅由目标函数构成，称为**无约束最优化问题**。

一、 数学预备知识

在开始叙述正文之前，先规定一些记号，并且给出某些数学预备知识。

本书所涉及的函数几乎都是向量变数的可微函数。于是除非另有说明，否则$f(\mathbf{x})$都是表示f是一个实值函数，$\mathbf{x}$则是变数的n维组。因此

$$\mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

以及

$$\mathbf{x}^T=[x_1,\ x_2\cdots x_n]$$

如果需要指出f是不止一次可微的，将专门指明。如果$\mathbf{x}$是在某一特定点计值，则用下标或上标表示。即如$\mathbf{x}_0$、$\mathbf{x}_1$或$\mathbf{x}^*$表明n维变数$\mathbf{x}$是一组特定的数。一般地说，我们保留$\mathbf{x}^*$用来代表最优解。

1.4 n维欧氏空间 由实数的**所有的** n 元组所构成的集合，记作 E^n，称为 n 维空间。在 E^n 中，如果通常的向量加法规则(平行六面体法则)和标量同向量的乘法规则都适用，并且对于 E^n 中任意二向量x和y，有用下式定义的**距离**

$$\|\mathbf{x}-\mathbf{y}\|=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2}$$

则称 E^n 为 **n 维欧氏空间**①。当y=0，有

$$\|\mathbf{x}\|=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}$$

称为 x 的范数或大小。

如无特别指明，则 $\mathbf{x}\in E^n$。E^1 是整个实数轴，E^2 是平面，E^3 是我们生活于其中的空间等等。对于 $n>3$ 的情形，E^n 较难想象。

两个向量 $\mathbf{x}\in E^n$、$\mathbf{y}\in E^n$ 的**内积**(或**点积**)，是将它们的对应元素相乘后再相加所得的标量，即

$$\mathbf{x}^T\mathbf{y}=\sum_{i=1}^{n}x_i y_i$$

如果二向量的内积为零，即 $\mathbf{x}^T\mathbf{y}=0$，则称他们是**正交的**。

由内积和欧氏范数的性质得知，对于任意的 $\mathbf{x}$、$\mathbf{y}\in E^n$，有

$$|\mathbf{x}^T\mathbf{y}|\leqslant\|\mathbf{x}\|\|\mathbf{y}\|$$

其中等号当且仅当 $\mathbf{x}=\alpha\mathbf{y}$ 时成立，α 是标量。这一重要的关系式称为**Schwarz 不等式**。

向量组 $\mathbf{x}_1$，$\mathbf{x}_2$，$\cdots\mathbf{x}_n$，如果关系式

$$\sum_{i=1}^{n}\alpha_i\mathbf{x}_i=0$$

①欧氏空间的严格定义可见有关的线性代数书籍——译者

总是蕴函 $\alpha_i=0$，$i=1, 2, \cdots, n$，则称这些向量是**线性无关**的。

线性向量空间 E^n 中的向量集合 S，如果(1)S 的所有向量线性无关，(2)E^n 中的每一个向量都可以表示成 S 中各向量的线性组合，则称S是E^n的**一组基**。

一个向量集合，如果E^n 中的任一向量都可以表示成该集合中向量的线性组合，则称它生成向量空间 E^n。例如，向量 $\mathbf{x}_1=(1, 0, 0)$，$\mathbf{x}_2=(0, 1, 0)$，$\mathbf{x}_3=(0, 0, 1)$，和 $\mathbf{x}_4=(1, 1, 0)$生成E^3，但它们不构成一组基，但是，上述前三个向量构成一组基，而且生成E^3。

2. **凸性** 凸集和凸函数的概念在最优化中是特别重要的。为了建立这些概念，我们首先定义**直线**。

过E^n 中二不同点$\mathbf{x}_1$、$\mathbf{x}_2$ 的直线定义作集合

$$X=\{\mathbf{x} \mid \mathbf{x}=\lambda\mathbf{x}_2+(1-\lambda)\mathbf{x}_1,\ \text{所有}\lambda\},$$

其中 λ 是标量。如果 $0\leqslant\lambda\leqslant1$，则$X$表示连结$\mathbf{x}_1$ 和 $\mathbf{x}_2$的**线段**。

例1-1 在图1-1中给出了当$\mathbf{x}\in E^2$ 时，上述集合X定义

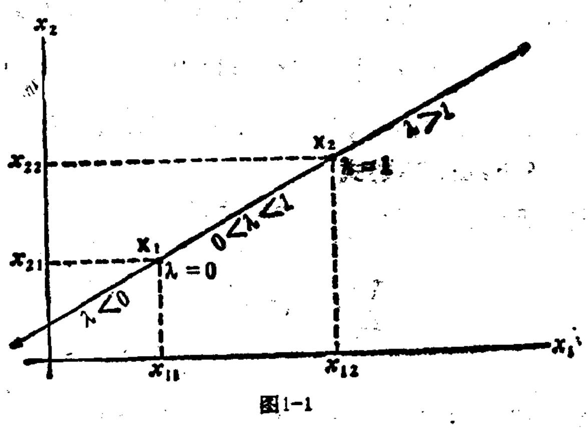


图1-1

一条直线的情形。

集合 X，如果对于其中的任意二点x_1和x_2，连结两点的线段也在此集合中，则称X是**凸的**。

例1-2　图1-2中(a)、(b)所给出的两个集合是凸的，(c)、(d)给出的集合不是凸的。

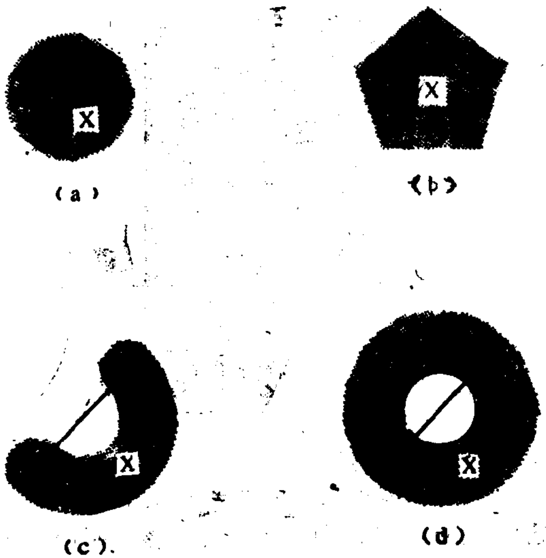


图 1-2

E^n 中的**超平面**定义作下述点集

$$X=\{\mathrm{x}\mid \mathrm{c}^T\mathrm{x}=z\}$$

其中$\mathrm{c}\neq 0$，是n维向量，z是常数。一个超平面将E^n分割为两个**闭半空间**，即

$$X=\{\mathrm{x}\mid \mathrm{c}^T\mathrm{x}\geqslant z\} \quad 和 \quad X=\{\mathrm{x}\mid \mathrm{c}^T\mathrm{x}\leqslant z\}$$

当不等式严格成立时，X是**开半空间**。

例1-3　超平面

$$X=\{\mathrm{x}\mid 2x_1+x_2=3\}$$

将E^2分为半空间X_1和X_2，见图1-3。如果$z=0$，则超平面通过原点。

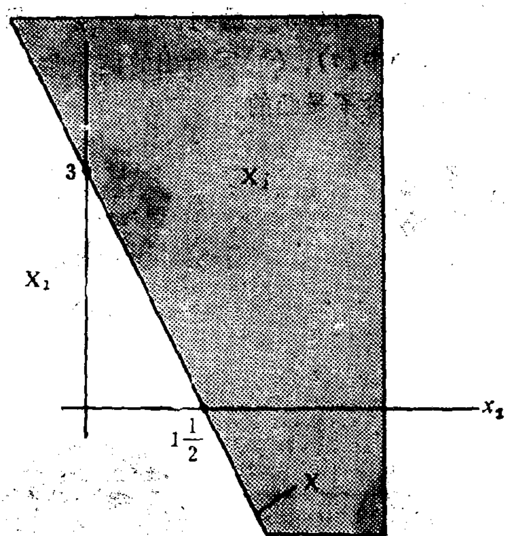


图 1-3

［解］ 在E^n中给定超平面$\mathbf{c}^T\mathbf{x}=z$，向量$\mathbf{c}$称为该超平面的**法向量**。这就是说，向量$\mathbf{c}$正交于该超平面内的任意向量。

例1-4 对于超平面$X=\{\mathbf{x}|-2x_1+x_2=0\}$，

$$\mathbf{c}=\begin{bmatrix}-2\\1\end{bmatrix}，向量\mathbf{x}=\begin{bmatrix}1\\2\end{bmatrix}在超平面内。这时$$

$$\mathbf{c}^T\mathbf{x}=(1)(-2)+(2)(1)=0$$

二向量正交(垂直，见图1-4)。

［解］ 一类重要的超平面称为**支撑超平面**。凸集X，给定一个边界点$\mathbf{w}$，超平面$\mathbf{c}^T\mathbf{x}=z$如果满足：$\mathbf{c}^T\mathbf{w}=z$，并且X的所有点都属于这一超平面的一个闭半空间，则称此超平面

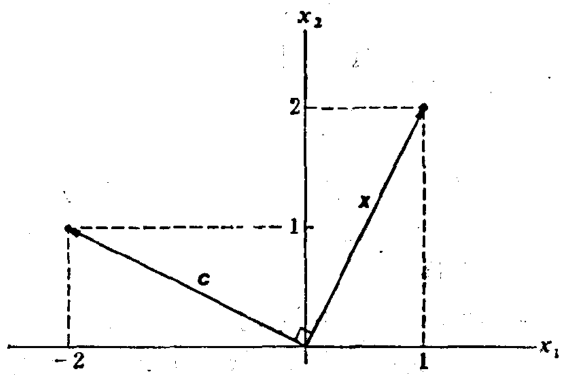


图 1-4

为X在点w的支撑超平面。也就是说，对于所有的$\mathbf{x}\in X$，$\mathbf{c}^T\mathbf{x}\geqslant z$，或者，对于所有的$\mathbf{x}\in X$，$\mathbf{c}^T\mathbf{x}\leqslant z$。

例 1-5 在E^2中，凸集的支撑超平面是该集合在某个边界点的切线；在E^3中，则是在某个边界点的切平面(见图1.5)。

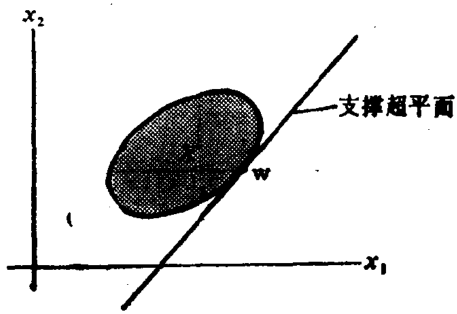


图 1-5

［解］ 函数$f(\mathbf{x})$，如果对于E^n中的凸集X的任意二点$\mathbf{x}_1$和$\mathbf{x}_2$，对于所有的λ，$0\leqslant\lambda\leqslant1$，都有

$$f(\lambda x_2+(1-\lambda)x_1)\leqslant\lambda f(x_2)+(1-\lambda)f(x_1)$$

则称$f(x)$为凸集X上的**凸函数**。

如果上述定义中，当$0<\lambda<1$及$x_1\neq x_2$时，成立严格的不等式，则称函数是**严格凸的**。

类似地，将不等式反向，以同样方式可以定义**凹函数**。线性函数在整个E^n中既是凸的又是凹的。但它既不是严格凸的，也不是严格凹的。

例1-6 考虑如图 1-6所示的一元函数。对于所有的λ，$0\leqslant\lambda\leqslant1$，函数曲线在连结二点$(x_1,\ f(x_1))$、$(x_2,f(x_2))$的线段的下方，因此函数是凸的；事实上，也是严格凸的。

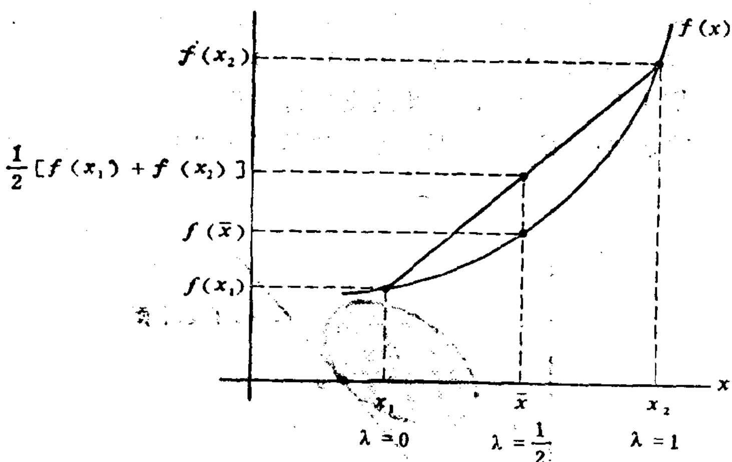


图 1-6

［证］ 从定义可见,如果$f(x)$是(严格)凸的,那么$-f(x)$是(严格)凹的,反之亦然。这样一来,仅发展(比如说)凸函数的理论就够了，因为可以简单地将任意凹函数转化为凸函数。

例1-7 在图1-7中，函数 $f(x)$ 是凸的，而 $-f(x)$ 是凹的。凸函数 $f(x)$ 的最小值点 x^* 也就是凹函数 $-f(x)$ 的最大值点。但这两个函数在那点的函数值是不同的，而是 $\min f(x) = -\max[-f(x)]$。这一命题在一般情形也成立。

［证］ 注意，凸函数的定义中仅要求函数在一凸集上有定义，既不要求函数 $f(x)$ 是连续的，也不要求是可微的。

可微函数 $f(x)$ 的**梯度**是由偏导数组成的 n 维向量

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

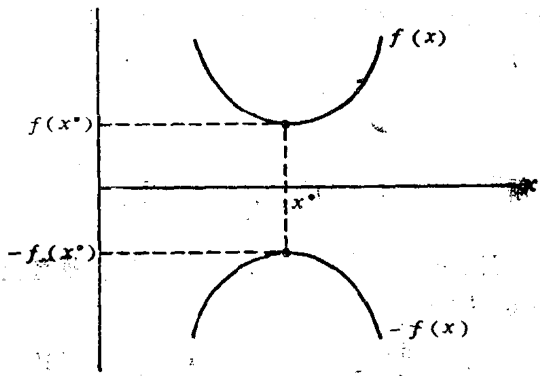


图 1-7

如果 $f(x)$ 还具有二阶偏导数，其Hesse **矩阵**是由二阶偏导数组成的 $n\times n$ 对称矩阵

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