



国家“九五”重点图书
出版规划项目

走向科学的明天丛书

精益求精的最优化

李惠玲 施洪亮 刘开峰 刘玲 著

广西教育出版社

图书在版编目(CIP)数据

精益求精的最优化/李惠玲等著. —南宁:广西教育出版社, 2000. 4

(走向科学的明天丛书)

ISBN 7-5435-2977-7

I. 精... II. 李... III. 最优化-普及读物
IV. 0224-49

中国版本图书馆 CIP 数据核字 (2000) 第 19553 号

走向科学的明天丛书

精益求精的最优化

李惠玲 施洪亮 刘开峰 刘玲 著



广西教育出版社出版

南宁市鲤湾路 8 号

邮政编码: 530022 电话: 5850219

本社网址 <http://www.gep.com.cn>

读者电子信箱 master@gep.com.cn

全国新华书店经销 广西民族印刷厂印刷

*

开本 850×1168 1/32 4.75 印张 插页 4 96 千字

1999 年 12 月第 1 版 2000 年 8 月第 3 次印刷

印数: 9 001—14 000 册

ISBN 7-5435-2977-7/G·2261 定价: 9.00 元

如发现印装质量问题, 影响阅读, 请与承印厂联系调换

序

在世纪之交,我们这套《走向科学的明天丛书》问世了。这是一套面向青少年朋友的大型科普读物,是为了补充学校教育之不足,从数学、物理学、化学、天文学、地球科学和生命科学六大基础科学的历史发展、当前的成就、未来的璀璨远景,分类展示给读者。

本世纪末,有一股反科学的逆流,认为科学的时代已经过去。例如美国的约翰·霍根,他写了一本书《科学的终结》,他说:“科学(尤其是纯科学)已经终结,伟大而又激动人心的科学发现时代已一去不复返了!”与此同时,法国当代女巫伊丽莎白·泰西埃也写了一本畅销书《占星术——21 世纪的科学》,再加上那些“世纪末”的谣言和形形色色的邪教,把社会搅得似乎有点混乱。

然而,科学永远是照亮世界的火炬,光芒所至,一切邪魔歪道都会原形毕露。这套《走向科学的明天丛书》也正是告诉大家,21 世纪的科学非但不会终结,还将会有更大的发展。

为什么《走向科学的明天丛书》还是从数、理、化、天、地、生这老的六大基础科学讲起? 因为我们不能割断人类认识客

观世界的历史,这是人类认识绝对真理的长河中的一个非常重要的环节,近代科学和未来的科学都是在这个基础上发展起来的,边缘科学、前沿科学……我们都在科学的明天中讲到了。有人不顾客观的科学发展的历史事实,主观地想把科学体系打乱,从而建立个人的“新科学体系”,这样只能把科学搞乱,给伪科学以钻空子的机会。

在80年代初期,科普界曾有过一场争论,那就是有人说知识的科普已经过时,科普的任务是普及科学思想和科学方法,而这个任务将由科学文艺(主要是科幻小说)来完成。我们说科学基础知识与科学思想和科学方法是刀与刃的关系,抛弃科学基础知识,科学思想和科学方法就成了无刀之刃,只是幻想与空话。科学基础知识越深厚,科学之剑也就越坚实,砥砺出来的剑刃也就无坚不摧。我们推出这套《走向科学的明天丛书》,也就是想让每一位读者都能得到这柄坚实的剑,而砥砺剑刃则需要读者们自己的努力了。

这套丛书的编写是在一批老科普作家支持下集体完成的,他们多年来在教育 and 科研第一线工作,如今大多已年近花甲或年过花甲,但为了科普事业的发展,他们仍然在百忙之中创作了这批精彩的科普作品,我们应该向他们表示衷心感谢。

最后,要特别感谢广西教育出版社,正是在编辑们的精心设计和组织下,这套《走向科学的明天丛书》才能与读者早日见面。

郭正谊

1999年8月20日

致青少年朋友

无论哪个国家,数学和本国语文都是学生的主课。这两科构成了人们最基本的文化素养。数学,则是最具国际性的学科。到20世纪中叶,世界各国的数学课程大体上是相同的。算术、几何、代数、三角成为人类基础教育的主题。这一切,古希腊的学者都已完成了,埃及、巴比伦、印度、阿拉伯和古代中国的数学家也都有特定的贡献。因此,现在的中小学数学内容,是人类的共同财富。

17世纪牛顿和莱布尼兹发明微积分,20世纪冯·诺依曼等发明了计算机,使得数学文明发生了巨大的变化。数学,以更新的面貌推动着社会前进。20世纪的数学发展,抵得上过去的几百年。现在的标准数学学科分类,有96门大学科,几百门小学科。现在世界上已经找不到能够通晓整个数学的数学家了。

一方面是中小学数学内容相对不变,另一方面是现代数学内容飞速发展。这二者间如何协调?除了加强中小学数学教育改革之外,对公众进行数学普及是一件大事。在“科教兴国”的今天,没有数学的普及是不可想像的。

2000年,被国际数学家联合会确定为数学年,目的是让公众了解数学。

广西教育出版社出版《走向科学的明天丛书》,其中包括的数学学科一套共6册。数学学科的内容是如此广泛,一套科普丛书是不可能介绍完全的。我们只选择一些重要的、比较熟悉的部分向读者做一些力所能及的介绍,希望本丛书能帮助读者对当代数学及其前景作一管窥。

数学全套书由以下6册组成:

1.《数学的明天》,这是我个人对现代数学的一些感受,主要由一些新闻和故事组成,期望从整体上看看数学。

2.《集合与面积》,这是一本涉及无限的书,是现代数学的精华部分,我们做了简单的描述。主要作者是李惠玲和金家樑。苏明剑、刘珊、吴作章写了一些初稿。

3.《精益求精的最优化》,反映了人类期望用数学方法求得精确控制经济、管理、军事以及生产过程的愿望。李惠玲是主要作者。参加写作的有施洪亮、刘开峰和刘玲。

4.《大千世界的随机现象》,主要介绍概率和数理统计的简单内容。当“降水概率”在电视屏幕上出现时,高中毕业生却全然没有在课堂上听说过“概率”,这是数学教育的悲哀。对这本书,我花了一些力气,刘萍作了一些整理工作。初稿由张东鸿、李雪峰、何君等完成。

5.《组合数学方兴未艾》,对未来会有重大发展的一门学科做了介绍。计算机是“离散”的,数据是离散的。“组合”爆炸是一个现实问题。书中谈了许多中国古代数学家的贡献。此书由王春萍、张建国写初稿。

6.《平面几何定理的机器证明》,是一本介绍当代中国数学家成就的普及性书籍。由江西师范大学孙熙椿教授撰写,

预想会受到读者的关注。

这套书是1998年动手写的。我拟订了提纲。开始时由华东师范大学数学系一批研究生写初稿,开过几次会。有些同学很用心,写得不错。不过毕竟第一次写作,需要改动之处甚多,有许多则完全是另起炉灶。我在匆忙中找李惠玲教授帮忙。她费了许多工夫。

由于匆忙,参与的人多,错误之处在所难免,对此心中十分忐忑,诚恳地希望读者原谅。

数学不像有些人宣传的那样,存在“数学危机”。数学在一日千里地前进。祝愿中国数学繁荣发展,尽早实现“21世纪数学大国”梦想。

张奠宙

1999年12月

写在前面的话

优中选优,好上加好,精益求精乃是人类昨天、今天和明天的一个永恒的话题。随着生产方式的进步,优化问题日益突出。如果说,19世纪以前已经有了优化问题,那么只是在20世纪才形成了独立的最优化数学理论。

有一句广告词是“没有最好,只有更好”。这是指在一门学问、一个事业乃至一种产品的历史发展过程中,总是从一个高峰到另一个高峰,永无止境。但是,每一个高峰在当时当地的条件下,都是最好的,达到了“最优”状态。数学工作的一个重要任务,就是在数量上刻画各种“最优”,为人类的利益服务。

做任何事情,都有一个效益问题。“多”、“快”、“好”、“省”这四个字,概括了工农业生产中的“最优化”要求。反映在数学上便产生了最优规划论、最优策略论、最优控制论以及最优流程论等学科,并陆续归纳出如人事指派、物资调度、运输计划、金融策略、投资组合、经济调控、节约下料、仓库存储、优选试验、统筹安排等数学研究课题,这些统称为“最优化”的理论,构成了20世纪数学发展的一个亮点。

最优化数学理论,古已有之,于今为盛。古代先人对某些事情的数学道理虽然说不清楚,但在实践中已经加以了运用。

举例来说,挑水的水桶为什么是圆柱形的?方形水桶做起来不是更方便吗?殊不知这这也是一个典型的数学最优化问题。从数学上可以证明,在周长一定的条件下,圆的面积最大。于是,在高度相同的柱体中,圆柱体的体积最大。也就是说,高度一样的长方体、棱柱体、圆柱柱体等形状的水桶,以圆柱体水桶所装的水最多。当然,圆柱形水桶还有美观的艺术特征以及不易损坏的好处,这些就离开数学了。

由这一简单的例子可见,优化是一种数学意识。本书的安排是从一些例子谈起,通过中学里常见的优化问题,领会华罗庚先生推广优选法、统筹法的良苦用心,然后由浅入深,逐步介绍线性规划、微积分方法以及目前还在发展着的各种优化的理论和技术。21 世纪的信息交换、生产效益、服务安排等,将会更加“量化和优化”。可以毫不夸张地说,缺乏“优化意识”的公民,是难以在未来社会上立足的。

愿读者在向大自然作斗争、为人类社会作贡献的历程中,以不断进取的精神和意识,完成最好,获得最快,达到最优!

李惠玲

1999 年 12 月

目
录

序	
致青少年朋友	
写在前面的话	
最优化——人类智慧的结晶	1
☞ 追求最优——从一些例子谈起	1
☞ 旗落帆升——堵丁柱	3
☞ 中学数学中的极值问题	8
☞ 莫斯科大学入学试题中的优化问题	14
☞ 采购的窍门——库存策略	18
统筹法和优选法	27
☞ 华罗庚的愿望	27
☞ 协调技术——统筹法大意	30
☞ 0.618 与优选法	36
☞ 进一步研究	44
线性规划	49
☞ 历史概况与经济意义	49
☞ 求解思路与图上作业法	58
☞ 0—1 规划	68
☞ 其他问题	75
微积分方法	79
☞ 从二次函数的极值谈起	79
☞ 曲线的切线和极值	82
☞ 微分方法求极值的推广	90
优化问题的未来发展	100
☞ 非线性规划	100
☞ 随机规划	108

☞ 动态规划..... 120

☞ 多目标规划..... 129

☞ 组合优化..... 133

最优化

——人类智慧的结晶



追求最优

——从一些例子谈起

中国古代有许多精彩的最优化决策，田忌赛马的故事更是尽人皆知的。齐王与他的大将田忌赛马时约定：各自出上、中、下三种等级的马各一匹，进行三场对抗赛，双方每输一场要付给对方黄金一千两。由于田忌的马比齐王同等级的马都略逊一筹，如果双方都用同等级的马进行比赛，齐王肯定会赢得全部三场比赛。田忌的军师孙臆替田忌出了个主意：用自己的下等马

和国王的上等马比赛,而用自己的上等马和国王的中等马比赛,用中等马和国王的下等马比赛。第一场比赛,齐王的马以极大的优势取得胜利,但紧接着的第二、第三场比赛,田忌的马都取得胜利。这样,齐王倒反输了一千两黄金。实际上,齐王在三类马的对比中,占得了局部最优,但全局不一定最优。局部的优化是总体优化的必要条件,但不是充分条件!

这一例子也说明,优化问题并不神秘,它们大量存在于生活之中。优化决策不涉及设备添置、大量资金和人力的投入等,只是一种智力的角逐。田忌的马并没有增加实力,成功的只是策略。在知识经济时代,这种以知识和智力为基础的数学学科,受到人们的重视,自不待言。

最优化问题也出现在大自然的“优胜劣汰”的自然选择之中。天工造物,使人惊叹,生物进化,积数亿年的优胜劣汰,繁衍至今,创造了许多最美、最经济的最优结构。人体的结构、动植物的结构和生长规律,无不体现最优原理。让我们来看一看蜂房的结构。

蜂窝结构令建筑师、工程师和经济师们倾倒。蜂窝的形状如图 1,单个蜂房(图 2)的横断面是正六边形,房底由三个菱形拼成(图 3)。蜂窝由一个个蜂房连接而成。截面为正六边形的蜂房是最省料的结构。在面积一定的平面图形中,圆

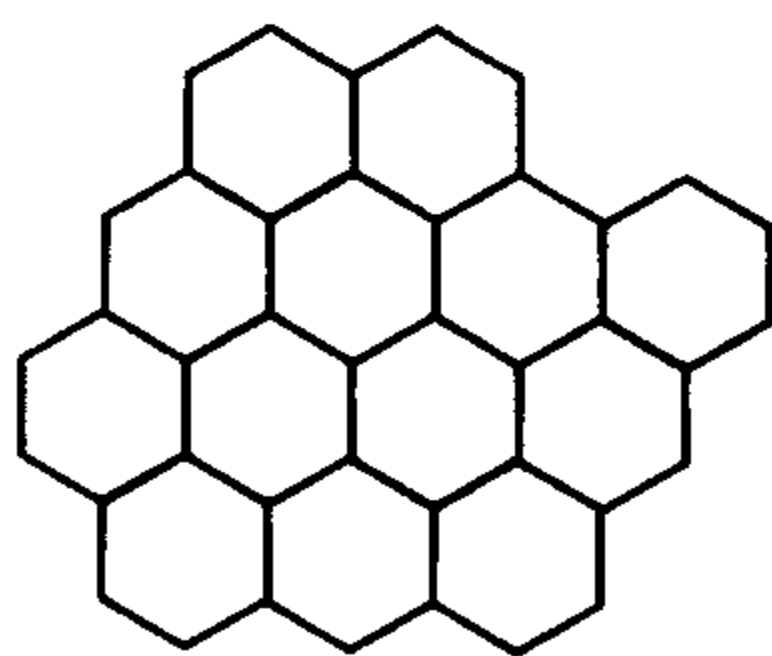


图 1

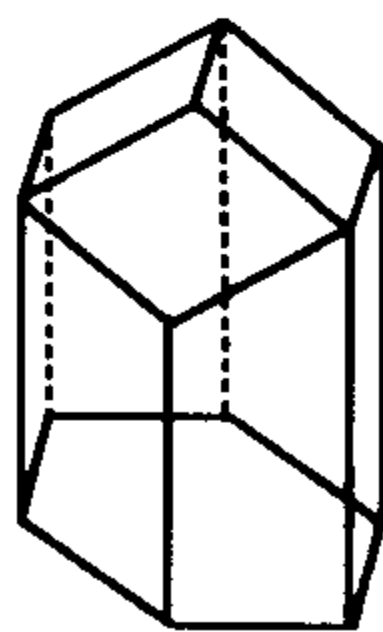


图 2

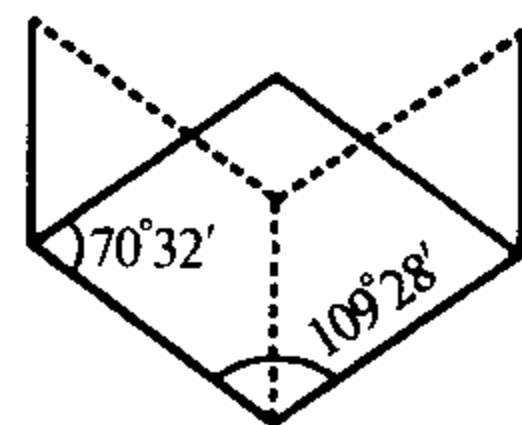


图 3

的周长最短(即最省料),但圆不能铺满整个平面,能铺满整个平面的多边形只有三种:正三角形、正方形和正六边形,在这三种图形中,正六边形的周长最短(即最省料)。蜂房底部由三个菱形拼成,菱形的钝角是 $109^{\circ}28'$,锐角是 $70^{\circ}32'$,这种结构也是最省料的。由于蜂窝结构具有省料且质量轻的特点,使要求最大限度地减轻自身质量的航空零部件及一些建筑构件,常采用蜂窝结构,如有些直升机的旋翼后段件就采用了蜂窝结构。物竞天择,适者生存,优胜劣汰,大自然的奥秘,令人类世代探索不尽,永无止境。

实际上,大自然并不懂得“最优化”,蜜蜂不可能有数学知识。但是,省工省料的要求却是蜜蜂生存所必需的。人类的最优化知识和理论,不仅能顺应大自然的规律,而且能够主动加以运用。充分发挥人的主观能动性,在人类活动中发挥更大的效益,当是“最优化”理论的鲜明特色之一。

最优化理论在工程技术领域的重要性容易了解。火箭的最佳发射时间,人造卫星的最佳工作姿态,钢铁厂的最佳轧钢控制,铁路交通的最佳运行图,现代农业的最佳管理体制……处处都有最优化理论的踪迹。

旗落帆升——堵丁桂

美国北卡罗来纳州中部有三所大学——北卡罗来纳州大学、北卡罗来纳大学和公爵大学。它们分布在一个正三角形的三个顶点上,还有一些公司依附在这个三角区域内,称为“研究三角”。1966年,“研究三角”向电话公司申请一张“私家网络”连接这三所大学,因而需要构造一个总长最短的线路

网连接各大学,作为电话公司收费的依据。对于呈正三角形的“研究三角”,只要任取正三角的两边即可(图 4)。不妨假设正三角形的边长为 1 个单位,则该网络的总长为 2 个单位,“研究三角”照此付款,电话公司也满

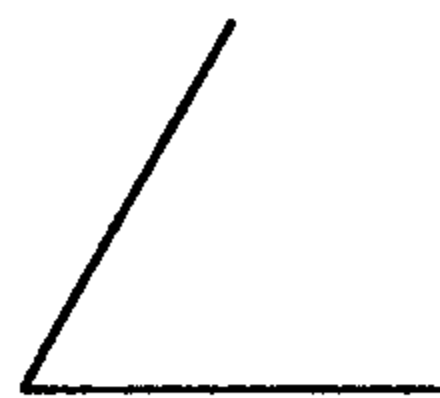


图 4

意。它仍可用公共的网络线路,不需要另铺设线路。作为电话公司,如何在若干城镇铺设通讯线路,使总费用最小,且让交费与收费双方都满意? 这是通讯线路网络优化问题,是电话公司所必须解决的。

为了反映一些复杂的关系,人们常常在纸上用点、线等符号画出各种各样的示意图,如铁路交通图、电话线路图、自来水管分布图、航空线图,这些图表示各地点之间的通路关系。不仅如此,一些“活动”过程也可以用图表示,如一个学院举行系际的球赛,就可用点表示系队,两队间的比赛用连线表示,比赛的全过程就可用一张图来表示;管理系统中各部门之间的领导和被领导关系亦可用图表示。为了反映图所表示的内容,可根据实际问题的需要,在点或线旁标上数(有时称权数),这些数在不同场合有不同的含义,如两点间的距离、比赛时间、日期……这种用连线表示对象之间的某种特定关系的图能使复杂的事情具体化、简单化。一般情况下,对图中相对点的位置及线的曲直不作要求,如图 5 与图 6 所示的关系是相同的。在形形色色的图中,有一种图很像一棵树,图中所有的点都必须通过线连成一片,且不含圈(图 7),不妨称它为“树”。但是图 8 不是树,因为点 1,2,3 形成圈,如果去掉圈中的任一边便是树。对于电话公司,它要研究的问题是不仅要使通话点之间能通话(允许通过其他通话点),而且要使电话线最短。

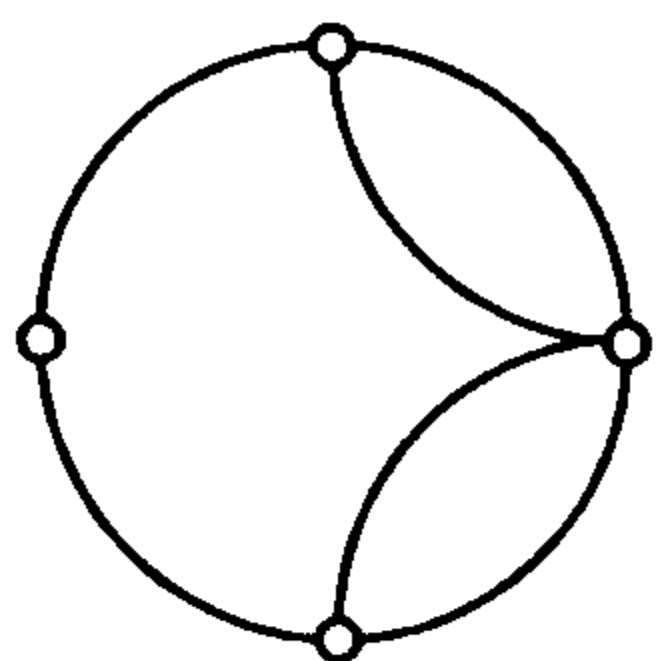


图 5

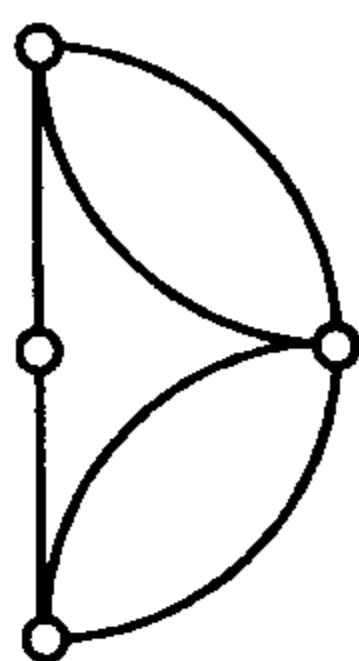


图 6

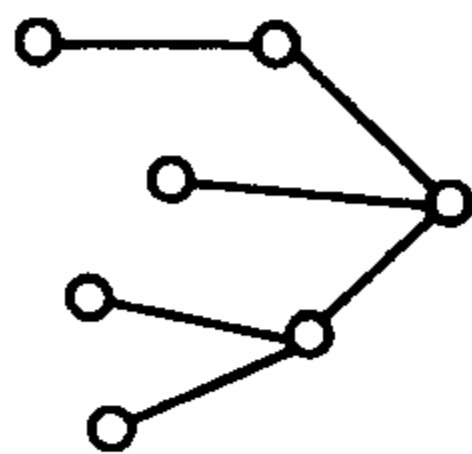


图 7

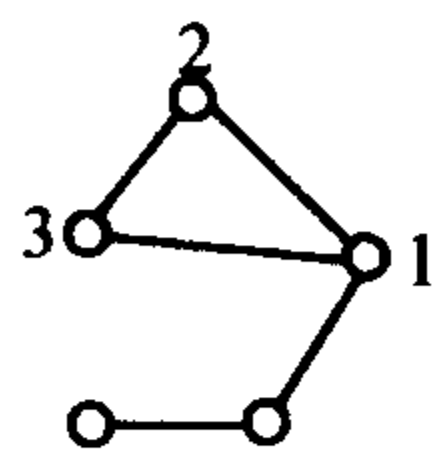


图 8

1959 年,贝尔研究所的克鲁斯克尔提出的最小生成树构造法(避圈法)成为电话公司网络线路的算法依据。最小生成树构造方法如下:在 n 个点中,先连接最近的两个点,然后在余下的可能连线中选择最短且不与已有的连线构成圈的连接线,按此方法继续构造,直到有 $n - 1$ 条连接线为止。显然,“研究三角”取两条边作为收费依据是合理的,它符合最小生成树构造法。时过 1 年,“研究三角”的私家网络在正三角形的中心位置添了一个“新据点”。电话公司以为这会增加网络线路,可以多收费,可是按最小生成树构造法计算,结果只有 $\sqrt{3}$ 个单位,比原来还短,几乎是打了个八七折 $\left(\frac{\sqrt{3}}{2} \approx 0.866 \dots\right)$ 。电话公司疑窦大起,派人实地勘察,发现该地区并没有设第四所大学,正三角形区域中心是一片沼泽,在中心点矗立着一根旗杆。就是这根旗杆戳了电话公司一个大窟窿,不但钱少收了,而且还要从三所大学铺设线路到沼泽地中心的旗杆处,真是“赔了夫人又折兵”。情况是否会越来越糟? 贝尔研究所数学中心做了大量的分析和实验,作出“折扣不会大于 $\frac{\sqrt{3}}{2}$ ”的猜想。这一猜想在提出后的 20 年里都未得到证明,成为数学难题,难题随“旗”招展。1987 年,我国青年数学家堵丁柱解决了这一难题,北卡罗来纳州“研究三角”的沼泽地里的那面“旗”在“飘扬”了 20 年后终于

“徐徐落下”。

“研究三角”私家网络问题是一个求极值问题。假设 A 、 B 、 C 三个城市分布在正三角形的三个顶点，欲在三个城市间铺设通讯电缆，怎样才能使网络最短？设正三角形的边长为 1，图 9、图 10、图 11 是三种铺线方案，它们的总长分别为：

$$l_1 = 2, l_2 = 1 + \frac{\sqrt{3}}{2}, l_3 = 3 \cdot \frac{2 \cdot \sqrt{3}}{3 \cdot 2} = \sqrt{3}。$$

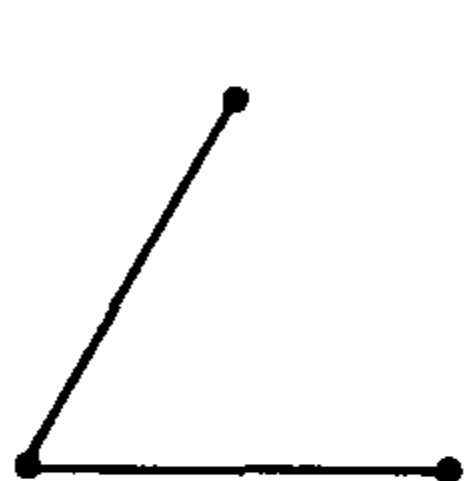


图 9

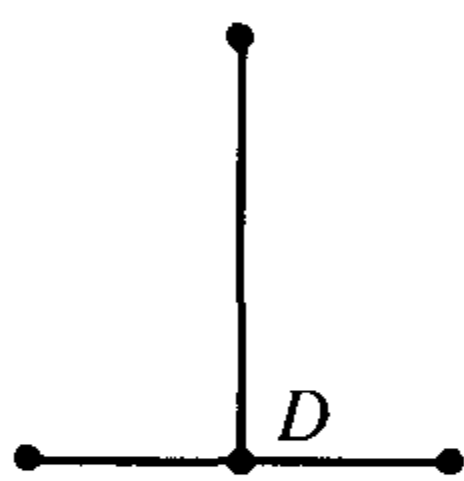


图 10

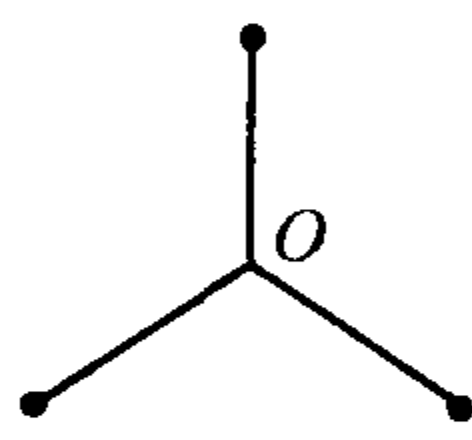


图 11

显然，第三种方案为最短线路。是否还有能使线路更短的方案？对电话公司来说，是否会有更糟的结果？堵丁柱证明了以下定理：允许在已有图上的城市以外添设电缆连结点（如图 10 和图 11 中的 D 点和 O 点），那么图上的最短线将满足：

$$\frac{\text{最短线}}{\text{原计算线}} \leq \frac{\sqrt{3}}{2}$$

这个问题被称为斯坦纳问题。斯坦纳问题还可推广到凸四边形的情形。

设 A 、 B 、 C 、 D 为四个工厂所在地， $AB = a$ (km)， $BC = \frac{2}{3}a$ (km)， $CD = \frac{1}{2}a$ (km)， $\angle ABC = 45^\circ$ ， $\angle BCD = 120^\circ$ ，要找一个点 H 设置供应站，使它到四个工厂的距离总和 $HA + HB + HC + HD$ 最小。不难证明，对角线 AC 、 BD 的交点 H_0 即为所求。在凸四边形内任选一个不同于 H_0 的点 H' （图 12），有：

$$H'A + H'B + H'C + H'D > H_0A + H_0B + H_0C + H_0D$$

(因 $H'A + H'C > AC = H_0A + H_0C$

$$H'B + H'D > BD = H_0B + H_0D$$

两式相加便得结果)

$$\text{故 } (HA + HB + HC + HD)_{\min}$$

$$= H_0A + H_0B + H_0C + H_0D$$

$$= AC + BD$$

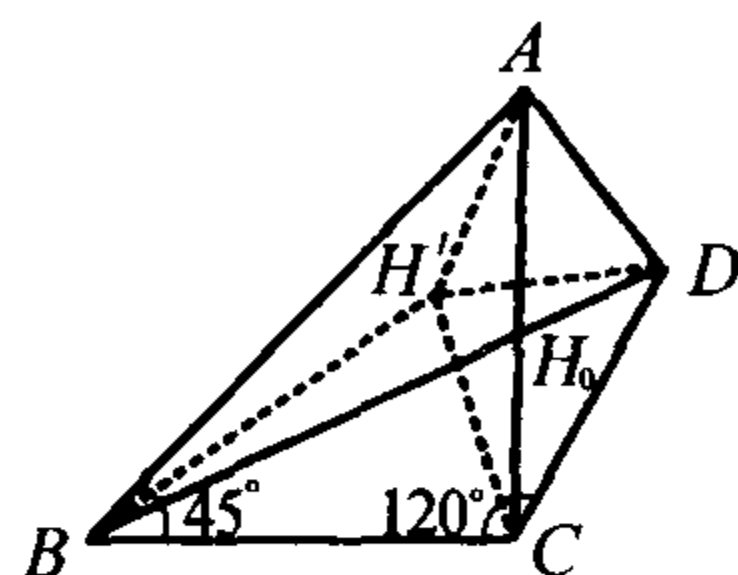


图 12

$$\begin{aligned} \text{且 } AC + BD &= \sqrt{a^2 + \left(\frac{2}{3}a\right)^2 - 2 \cdot a \cdot \frac{2a}{3} \cdot \frac{\sqrt{2}}{2}} \\ &\quad + \sqrt{\left(\frac{a}{2}\right)^2 + \left(\frac{2}{3}a\right)^2 + 2 \cdot \frac{a}{2} \cdot \frac{2}{3}a \cdot \frac{1}{2}} \\ &= \sqrt{13 - 6\sqrt{2}} \cdot \frac{a}{3} + \frac{\sqrt{37}}{6}a \approx 1.72a \end{aligned}$$

供应站应设在 H_0 , 它到各厂距离之和最小(约 $1.72a$)。斯坦纳问题系列除前面列举的外, 还有: 所有内接于圆的三角形中, 等边三角形有最大的面积; 在所有外切于圆的三角形中, 等边三角形的面积最小等。它们都可用几何方法求解。

堵丁柱, 1948 年出生于黑龙江省齐齐哈尔市。1967 年, 他以优异的成绩考入齐齐哈尔市重点中学。1978 年, 他考入东北重机学院。同年夏天, 中国科学院恢复招收研究生, 当年报考应用数学所者达 2000 多人, 而录取名额仅有 24 个, 堵丁柱从众多报考者中脱颖而出, 被应用数学所录取。1981 年, 堵丁柱以优异的成绩获得硕士学位后进入美国加州大学圣塔芭芭拉分校学习, 3 年后以全 A 成绩获得博士学位。其时适逢陈省身教授主持的、设在柏克利的“数学中心”招收博士后, 要在 70 多名全美国各大学的新博士后申请人中挑选 6 名, 堵丁柱又被选中。1 年后, 他应美国著名的麻省理工学院应用

数学系主任 Kleitnau 之邀,任该系助理教授。1987 年,他辞去麻省理工学院的教职,毅然返回祖国,中国科学院破格聘任他为研究员。他和程民德、吴文俊、谷超豪、杨乐、冯康等数学家共同发起召开“二十一世纪中国数学展望学术讨论会”,筹划组合最优化的国际会议。他是中国科学院 1989 年获青年一等研究奖的惟一青年数学家。

中学数学中的极值问题

中学的数学课程名目不少,最初登场的是代数和几何。代数研究数,几何研究形。对于刚入中学的初中生,形的直观富有吸引力,它揭示了日常生活中似乎微不足道的小事背后的“大道理”。“抄近路”,人人都会,却不一定明白其理由。“两点之间线段最短”,“三角形两边之和大于第三边”,“直线外一点到直线各点的连线中垂线最短”,这些“大道理”令人豁然开朗,也是中学生最早遇到的极值问题。

2000 多年前,在几何发达的古希腊,人们就知道利用图形对称性解决诸如“在河岸上取一点 C ,使它到 A 、 B 两村路程之和最短”这一类问题(图 13)(只需找出 A 的关于河岸 EF 的对称点,便可确定 C 的位置)。许多极值问题可以用初等数学方法求解。例如,一块长 25cm、宽 18cm 的铁板,去掉一个和三边相切的圆后,要在剩下的铁板中剪出一个最大的圆,其直径是多少(图 14)? 设所求圆的半径为 x ,利用勾股定理可得方程:

$$(9+x)^2 = (9-x)^2 + (16-x)^2$$

解得 $x = 4\text{cm}$ 。即剩下的铁板能剪出的最大圆的直径为

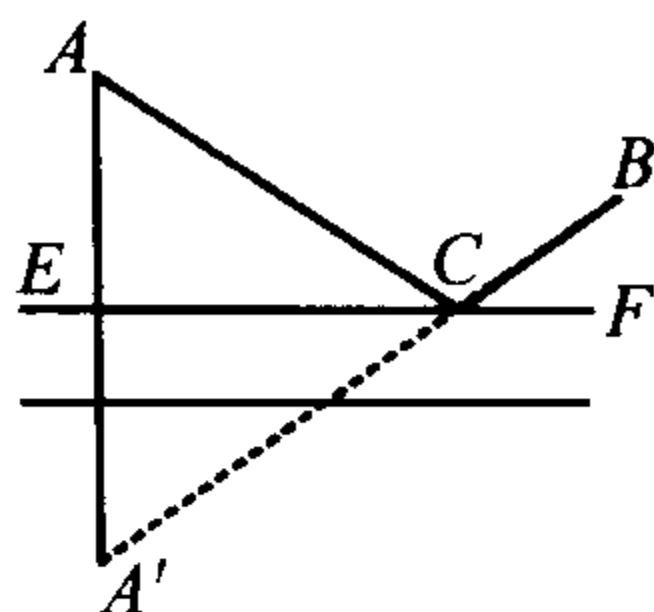


图 13

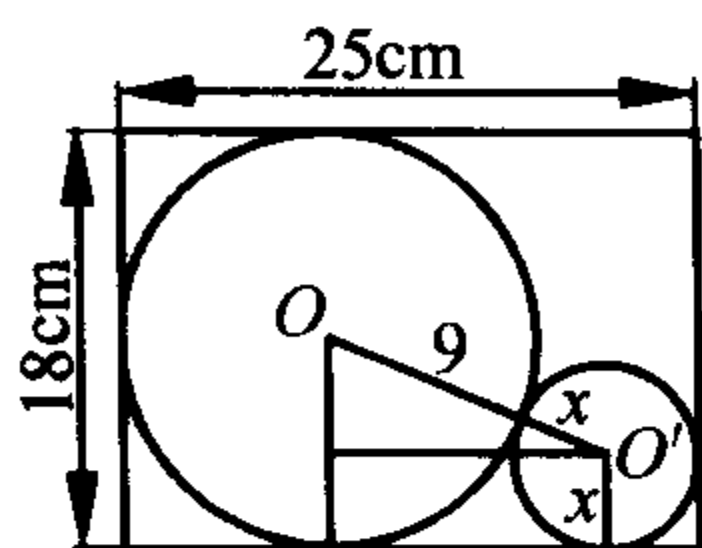


图 14

8cm。用几何方法寻求最优解常显得精巧无比, 乐趣无穷。1471 年, 德国数学家米勒提出了塑像问题: 假定有一个塑像高 h 英尺, 立在一个高 p 英尺的底座上。一个人注视着这个塑像并朝它走去, 这个人的水平视线离地 a 英尺(图 15)。问此人应站在离塑像基底多远处, 才能使塑像看上去最大? 要使塑像看上去最大, 视角必须最大, 问题转化为在水平视线 EF 上求使视角最大的点。只要过 AB 作一个圆与 EF 相切, 切点 M 即为所求。因为 EF 上的其他点都在圆外, 它们对塑像的视角都比 M 点的视角小。

有些极值问题可转化为仅含正弦函数或余弦函数的形式, 然后利用三角函数的有界性求极值。例如, 钳工要从一块小于半圆的扇形钢板中切割出一块面积最大的矩形钢板(图 16), 使废料最少, 应如何切割?

设扇形 OAB 的半径为 R , 中心角为 2α , $CDEF$ 为内接矩

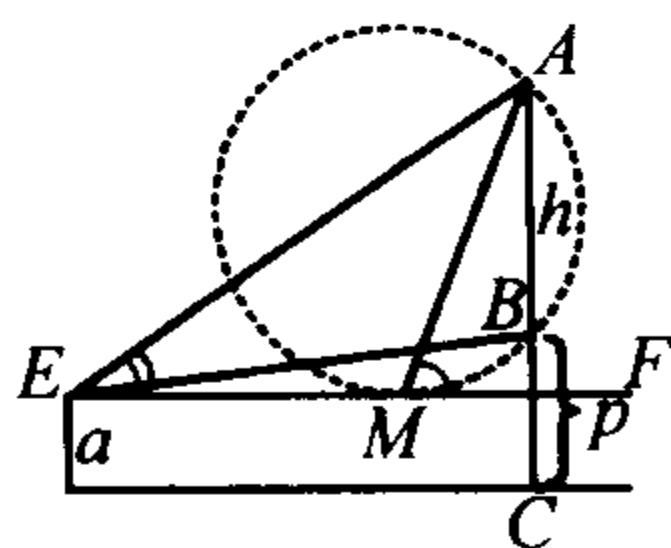


图 15

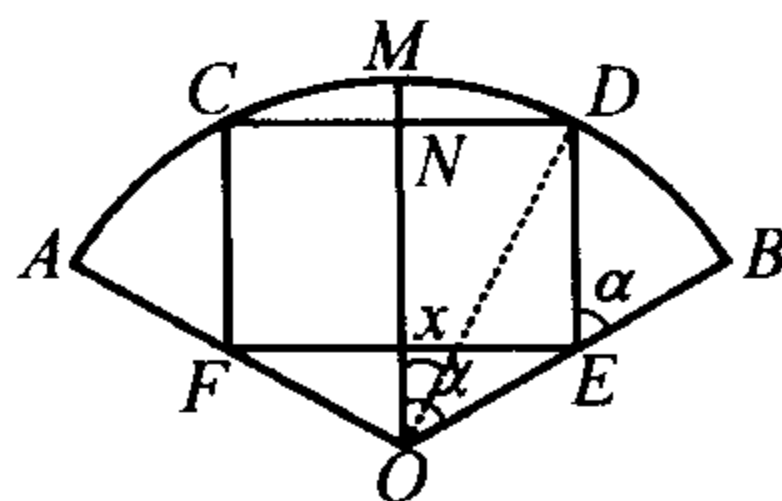


图 16

形, M 为 AB 的中点, $\angle BOM = \alpha$, 设 $\angle DOM = x$, 由正弦定理可得 $\frac{DE}{\sin \angle DOE} = \frac{OD}{\sin \angle DEO}$, 即 $\frac{DE}{\sin(\alpha - x)} = \frac{R}{\sin(180^\circ - \alpha)}$, $DE = \frac{R \sin(\alpha - x)}{\sin \alpha}$, 矩形面积 $S = CD \cdot DE = 2R \cdot \sin x \cdot R \cdot \frac{\sin(\alpha - x)}{\sin \alpha}$, 欲使废料最小, 应使 S 最大。由上式可见, 当 $\sin x \cdot \sin(\alpha - x)$ 最大时, 矩形面积 S 才能最大。

$$\because \sin x \cdot \sin(\alpha - x) = \frac{1}{2} [\cos(2x - \alpha) - \cos \alpha],$$

当 $\cos(2x - \alpha)$ 最大, 即 $x = \frac{\alpha}{2}$ 时, S 取得极大值。可见, 只要将扇形的弧四等分, 从第一分点开始画内接矩形即可。内接矩形的最大面积为:

$$S_{\text{极大}} = \frac{R^2(1 - \cos \alpha)}{\sin \alpha} = R^2 \cdot \operatorname{tg} \frac{\alpha}{2}$$

由点、线、面组成的几何图形蕴含着许多令人眼花缭乱的极值问题, 上述几个例子都是借助于几何图形, 用几何或三角函数方法直接解决的极值问题。随着代数学的发展和笛卡尔坐标的建立, 形与数的结合更紧密, 变量的极值可归化为函数的极值。极值是特定问题中变量的常量, 用函数求极值使许多不易用几何直观表示的问题得以解决。在中学教材中, 除直接用几何和三角方法求极值外, 求函数的极值尚有以下三种类型:

第一类是求二次函数 $f(x) = ax^2 + bx + c$ 的极值。因为

$$f(x) = ax^2 + bx + c = a \left(x + \frac{b}{2a} \right)^2 + \frac{4ac - b^2}{4a}, (a \neq 0)$$

当 $a > 0$ 时, 它的图形是开口朝上的抛物线, 有极小值。顶点的函数值 $f\left(-\frac{b}{2a}\right) = \frac{4ac - b^2}{4a}$ 就是函数 $f(x)$ 的极小值; 当

$a < 0$ 时, 它的图形是开口朝下的抛物线, 有极大值。顶点的函数值 $f\left(-\frac{b}{2a}\right) = \frac{4ac - b^2}{4a}$ 是函数 $f(x)$ 的极大值。极值是局部概念, 最值是全局概念, 当函数在所讨论的范围内只有一个极值时, 这个极值必是最值。由于二次函数只有顶点一个极值, 所以二次函数的极大(或极小)值就是最大(或最小)值(图 17、图 18)。例如, 圆锥 $V-ACB$ 的底半径为 r , 高为 h , 圆

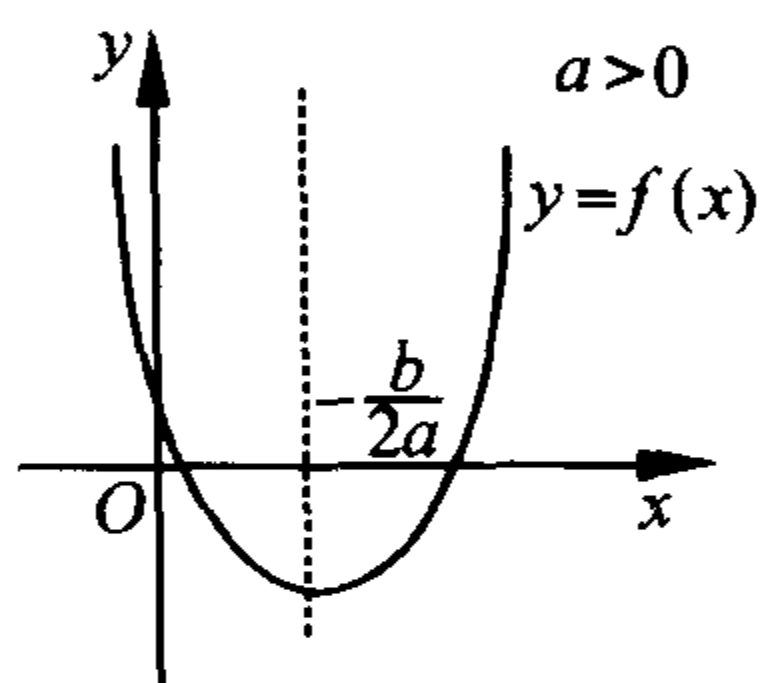


图 17

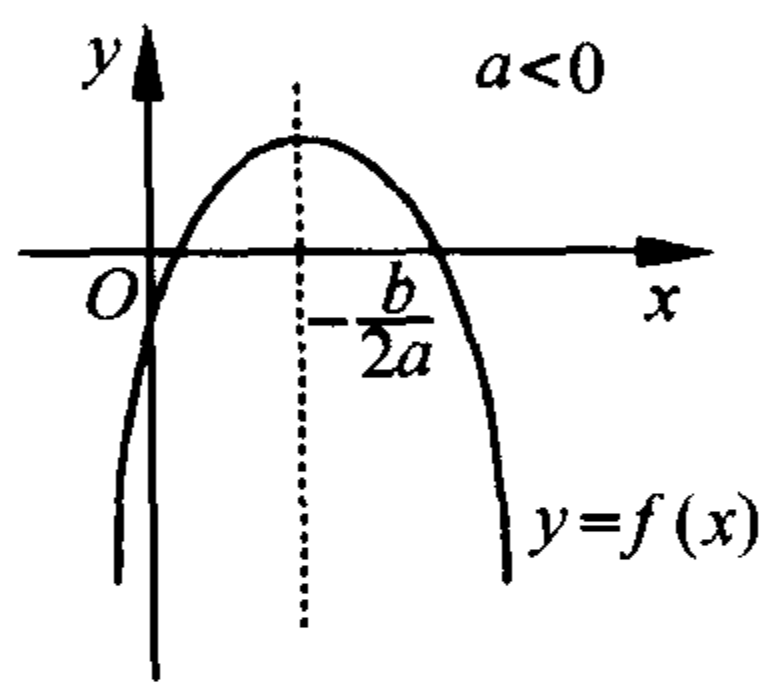


图 18

柱 $DEFG$ 内接于圆锥中, 圆柱底面半径多大时, 圆柱的侧面积最大?(图 19)

设圆柱的底面半径为 x , 它的侧面积为 y , 则

$$y = 2\pi x \cdot ED$$

由于 $\triangle BDE \sim \triangle BOV$,

$$\therefore \frac{ED}{VO} = \frac{DB}{OB}, \frac{ED}{h} = \frac{r-x}{r},$$

$$ED = \frac{h}{r}(r-x).$$

$$\text{故 } y = \frac{2\pi h}{r}x(r-x) = -\frac{2\pi h}{r}\left(x - \frac{r}{2}\right)^2 + \frac{1}{2}\pi hr.$$

根据二次函数的性质, 因为 $-\frac{2\pi h}{r} < 0$, 当 $x = \frac{r}{2}$ 时, y 有

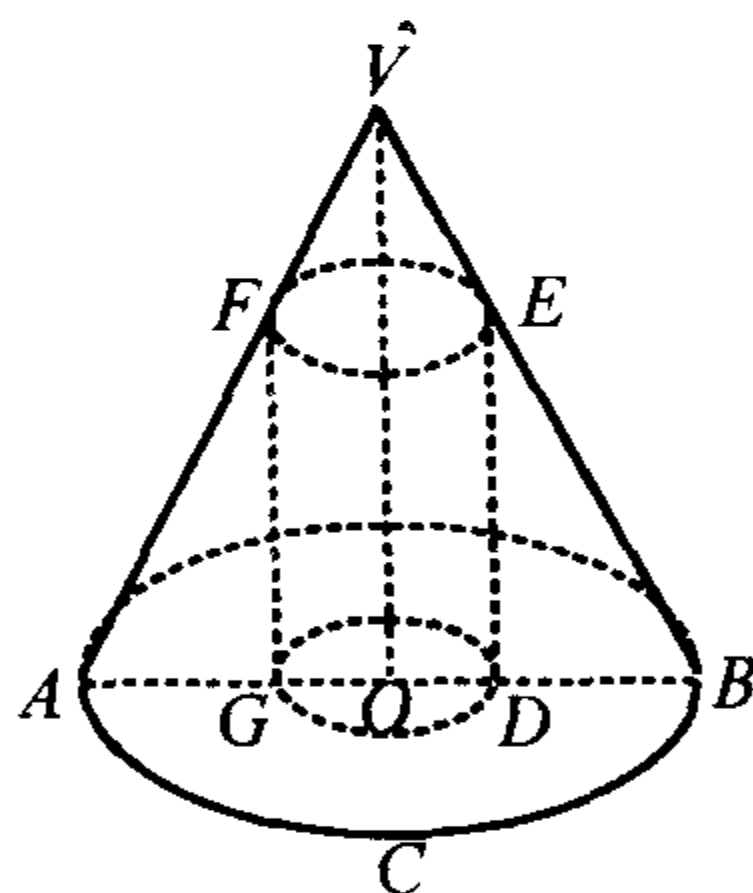


图 19

最大值 $\frac{1}{2}\pi hr$ 。即当圆柱的半径为圆锥底圆半径的一半时，其侧面积最大。

又如，某大炮以初速度 v 、仰角 θ 发射炮弹，它的水平分速度和垂直向上分速度分别为 $v_x = v\cos\theta$, $v_y = v\sin\theta$ (空气阻力略去)(图 20)，它的弹道方程为 $y = x\tg\theta - \frac{gx^2}{2v^2\cos^2\theta}$, y 是 x 的二次函数，它的轨迹是开口朝下的抛物线。根据一元二次函数的性质，可求得炮弹的最大高度和最大射程。若 $\theta = 45^\circ$ ，初速度 $v = 800$ 米/秒时，炮弹的最大高度可达 16000 米，最大射程为 64000 米。

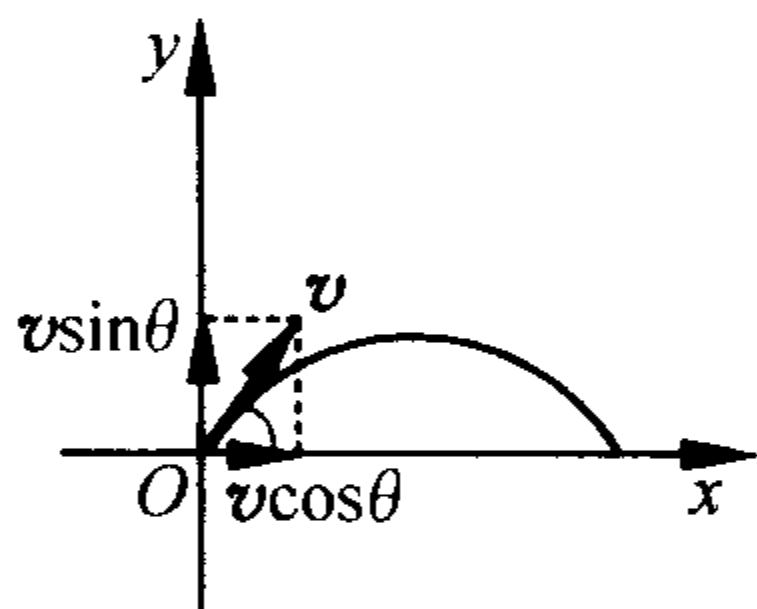


图 20

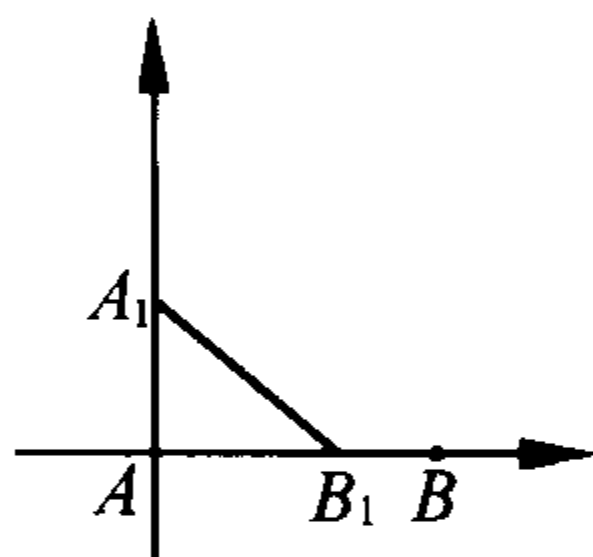


图 21

再如， A 、 B 两船相距 10 千米，如两船同时出发， A 船以时速 a 千米向正北行驶， B 船以时速 b 千米向正西行驶。两船何时相距最近？假定 t 小时后 A 船行至 A_1 ， B 船行至 B_1 (图 21)，则 $|AA_1| = at$ ， $|AB_1| = 10 - bt$ ，此时两船的距离 $|A_1B_1| = \sqrt{AA_1^2 + AB_1^2} = \sqrt{(at)^2 + (10 - bt)^2} = \sqrt{(a^2 + b^2)t^2 - 20bt + 100}$ 。根据二次函数的性质可知，当 $t = \frac{10b}{a^2 + b^2}$ 时，两船相距最近， $|A_1B_1| = \frac{10a}{\sqrt{a^2 + b^2}}$ 。类似的极值问题还可以举出很多。

第二类是通过一元二次方程解的判别式求极值。有些函

数不是一元二次函数,但可通过变换将它归结为以 x 为变量、 y 为系数的二次方程 $f_1(y)x^2 + f_2(y)x + f_3(y) = 0$, 然后利用判别式求极值。由于 x 是实数,所以该二次方程的判别式 $\Delta = f_2(y)^2 - 4f_1(y) \cdot f_3(y) \geq 0$, 此不等式的解便是 y 的取值范围,从而可得 y 的临界值。当对应的 x 是在函数定义域范围之内时, y 的临界值就是所求的极值。

例如,求 $y = (x + 2) + \sqrt{-x^2 + 10x - 21}$ 的极值。

可将原式有理化后得

$$2x^2 - (6 + 2y)x + (y^2 - 4y + 25) = 0,$$

$$\text{且 } y - (x + 2) \geq 0.$$

因 x, y 均为实数,故 $\Delta \geq 0$,

$$\text{即 } \begin{cases} (6 + 2y)^2 - 8(y^2 - 4y + 25) \geq 0, \\ y \geq x + 2 \geq 5, \end{cases}$$

解得 $5 \leq y \leq 7 + 2\sqrt{2}$, 因定义域为 $3 \leq x \leq 7$, 当 $y = 7 + 2\sqrt{2}$ 时, $x = 5 + \sqrt{2}$ 是在定义域中,故 y 的极大值为 $7 + 2\sqrt{2}$ 。当 $y = 5$ 时, $x = 3$ 是在定义域中,故 y 的极小值为 5。

第三类是通过变形,利用基本不等式求函数的极值(最值)。用于求极值的常用不等式有:

$$a + b \geq 2\sqrt{ab} \quad (a, b > 0), \text{ 当且仅当 } a = b \text{ 时, } a \cdot b \text{ 有最大值, } a + b \text{ 有最小值} \dots\dots\dots (1)$$

$$\frac{a + b + c}{3} \geq \sqrt[3]{abc} \quad (a, b, c > 0), \text{ 当且仅当 } a = b = c \text{ 时, } a \cdot b \cdot c \text{ 有最大值, } a + b + c \text{ 有最小值} \dots\dots\dots (2)$$

例如,欲制造一个容积为 V (定值)的带盖长方体容器,如何选择尺寸最省料?

设长方体容器的长、宽、高分别为 x, y, z , 则体积 $V = x \cdot y \cdot z$, 全面积 $S = 2(xy + yz + zx)$, 利用不等式 (2)

有：

$$xy + yz + zx \geq 3 \sqrt[3]{x^2 y^2 z^2} = 3 \sqrt[3]{V^2} (\text{定值})$$

当且仅当 $xy = yz = zx$ 时，等号成立。即 $x = y = z$ 时， $S = 2(xy + yz + zx)$ 有最小值 $6 \sqrt[3]{V^2}$ 。所以取边长为 $\sqrt[3]{V}$ 的正方体时，容器的全面积最小，即最省料。

综上所述，中学数学涉及的求极值问题的类型，除几何或三角函数外，主要是有显式表达式的代数函数（如不含指数、对数的函数）。许多实际问题中的极值都可用中学教材中的初等方法解决，在 17 世纪微积分学创立以前，所有极值问题都只能用初等方法解决。

莫斯科大学入学试题中的优化问题

长期以来，大学入学考试试题以纯数学概念和运算技巧为主。随着现代生产和科学技术的发展，数学已渗透到各个领域，数学教育除了培养学生的数学思维能力外，还强调数学应用能力，要求学生有解决实际问题的能力。因此，数学应用、数学建模日益受到重视，大学入学考试的命题要求也随之而变化。由于人类的大量活动无不寻求最优，因此优化问题成为培养学生数学应用能力和解决实际问题能力的主要内容。俄罗斯莫斯科大学是世界上享有声誉的一流高校。1994 年，莫斯科大学一些系的入学考试试题中涉及的最优化问题的实践性是很强的。该校经济系的入学试题中的优化问题是：一项工作任务，需要生产甲零件 3000 个、乙零件 5000 个。生产车间有 100 名工人，每人每单位时间能生产 3

个甲零件。若生产乙零件，则每个工人每单位时间能生产 5 个零件。问该怎样安排工人生产才能使任务尽快完成？此题实质上是要求在最短时间内完成任务。两种不同零件分给 100 名工人，可以有許多分配方案。若要最快完成，则必須所有工人都同时完成各自的任務，沒有人处于停工状态。这个问题较简单，车间调度人員只需要口算便可得出結果。甲零件所需的总工时为 $\frac{3000}{3} = 1000$ （单位时间），同理，乙零件所需总工时亦为 1000（单位时间）。因此，甲、乙两种零件各指派 50 名工人便可同时完成任务，不会有人处于停工状态。这个问题求解虽易，却能检验学生理论联系生产实际的能力。

同年，莫斯科大学数学力学系的人學试题共有 3 个优化问题，其中一题出现在预考试题中，另两题出现在主要考试题的口试试题中。预考试题为：在约束 $SA \leq 4, SB \geq 7, SC \geq 9, AB = 5, BC \leq 6, AC \leq 8$ 下，求三棱锥 $S-ABC$ 的体积的最大值。这道几何题是个数学规划问题（在给定的条件下，求实值函数的极大或极小值），它的目标函数是三棱锥 $S-ABC$ 的体积 V ，求最大值，约束条件有 6 个，可通过三角函数求解（图 22）。设 $S-ABC$ 是满足约束条件的棱锥，取面 ABS 作为它的底面。根据余弦定理：

$$\begin{aligned}\cos \angle SAB &= \frac{AS^2 + AB^2 - BS^2}{2AB \cdot AS} \leq \frac{-8}{40} \\ &= -\frac{1}{5}, \\ \therefore \sin \angle SAB &\leq \frac{2\sqrt{6}}{5}.\end{aligned}$$

由于棱锥的高不会超过它的侧棱，

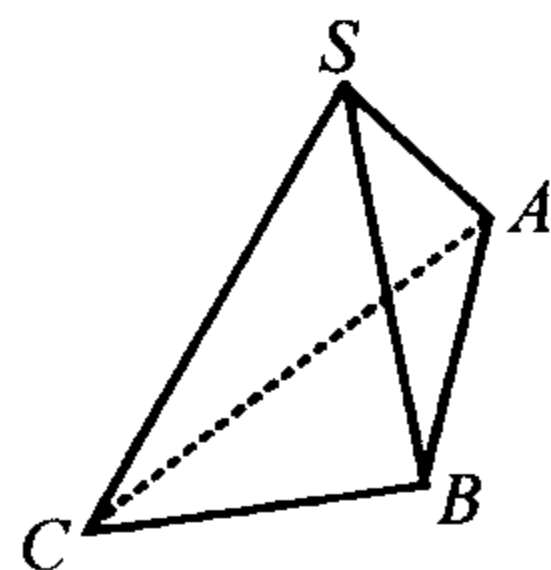


图 22

$$\begin{aligned}
 \therefore V_{S-ABC} &\leq \frac{1}{3} BC \cdot S_{\triangle BAS} \\
 &= \frac{1}{6} BC \cdot AB \cdot SA \cdot \sin \angle SAB \\
 &\leq \frac{1}{6} \cdot 6 \cdot 5 \cdot 4 \cdot \frac{2\sqrt{6}}{5} = 8\sqrt{6}.
 \end{aligned}$$

故三棱锥的最大体积为 $8\sqrt{6}$ 。

□ 试试题中的求最值题为：

(1) 对于每个 $a > 0$, 在条件 $2|x| + |x - 3a| \leq 9a$ 下, 求量 $|x - 2a|$ 的最大值;

(2) 对于每个 $a > 0$, 求量 $x^2 - |x + a| + x - a$ 在区间 $[-3, 3]$ 上的最小值。

这两题的格式亦已规划化, 解题技巧和难度显然大于经济系入学试题中的优化题。但结合实际不那么直接, 不需要考生自己建立数学模型。

题(1)中的目标函数 $y = |x - 2a|$, 约束条件为

$$g(x) = 2|x| + |x - 3a| \leq 9a, \text{ (图 23)}$$

由题可见, 在区间 $[-2a, 4a]$ 上, $g(x) \leq 9a$ 成立, 函数 $y = |x - 2a|$ 在 $x = -2a$ 上有最大值 $4a$ 。

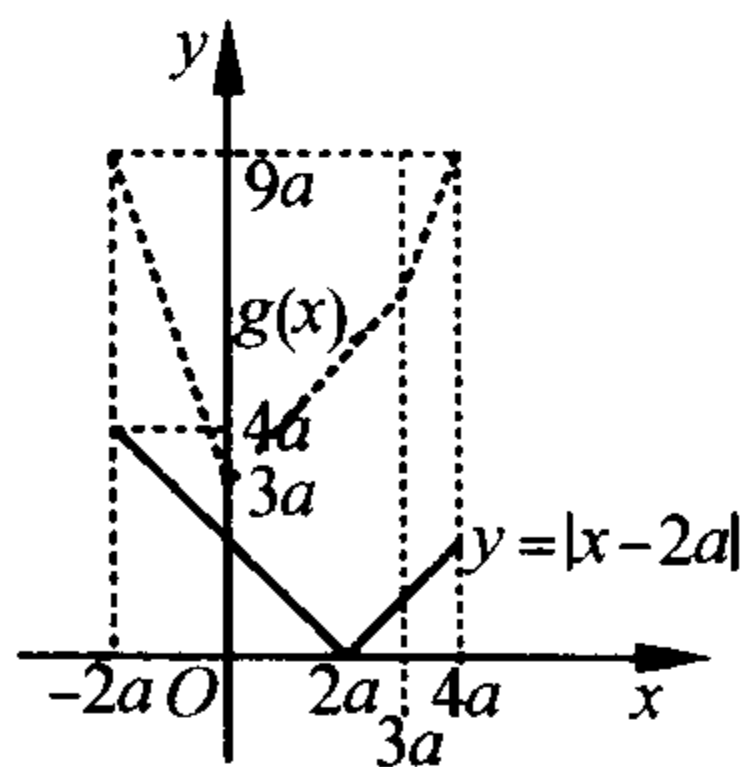


图 23

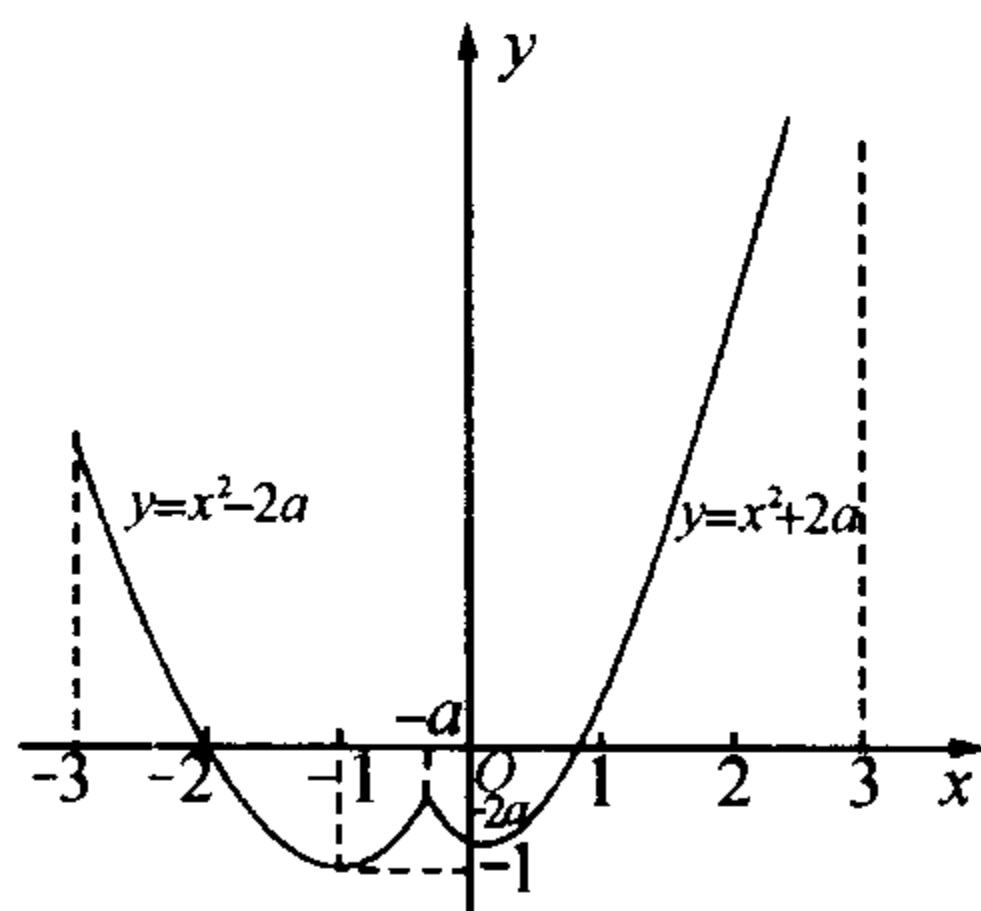


图 24

题(2)中,目标函数 $y = x^2 - |x + a| + x - a = \begin{cases} x^2 + 2x(x \leq -a), \\ x^2 - 2a(x \geq -a), \end{cases}$ 约束条件为 $-3 \leq x \leq 3$ 。由图 24 可见,

它的最小值或者等于左支抛物线的最小值(-1),或者等于右支抛物线的最小值($-2a$),比较后取其较小者。

我国自 1978 年恢复高考以来,入学考试数学试题中曾有一段时间没有联系实际的试题。自 1993 年开始,我国大学入学考试数学试题中恢复了数学应用题,虽然在内容和实践性上不像莫斯科大学经济系试题那么直接,但改变了清一色的纯数学概念和运算技巧题的状况,它标志着中学数学教学开始从传统知识传授模式转向以激励学生学习,以学生为主体的实践模式,不限于数学概念和技巧的理解和运用,还注意培养学生分析问题、正确运用逻辑推理,将实际问题转化为数学问题的能力。早在 20 世纪 60 年代初,以华罗庚教授为代表的一批数学家和数学教育家曾向中学生介绍解决实际问题的建模思想,在各种数学竞赛中出现了许多优化的实际问题,如 1962 年北京市高三数学竞赛第二试题中的“储油行车”问题。1978 年,全国中学数学竞赛恢复以后,又为学生提供了一些数学建模知识,如“洗衣服的数学”、“人口控制方案”、“生物数学趣谈”、“十人打水问题”等。此外,华罗庚杯数学竞赛中涉及的以基本传统模型为样板的简化编拟的竞赛题,有最短路问题、售货员问题、选址问题、下料问题等,对培养学生的创造性思维和综合、分析、解决问题的能力十分有利,也极大地激起学生学习数学的兴趣。第五届华罗庚杯竞赛中有一个口答题:某人从金坛市出发,去扬州、常州、苏州、杭州各一次,最后返回金坛,已知各市之间的路费如下表(见下页)所示,请你为他设计一条路

费最省的路线。

	金 坛	常 州	扬 州	苏 州	杭 州
金 坛	0	30	40	50	60
常 州	30	0	15	25	30
扬 州	40	15	0	15	25
苏 州	50	25	15	0	15
杭 州	60	30	25	15	0

一名参赛的初一学生仅用时 3 分就正确解答了这道试题。一年一次的全国大学生建模竞赛活动也于 1992 年开始。这些活动有极强的导向作用,在很大程度上培养了学生综合运用知识和解决实际问题的能力,有助于学生广泛猎取知识和跨学科思维能力的培养。大、中学生的建模竞赛试题多数是最优化问题,趣味性和实用性很强,参赛学生深感受益匪浅。然而,能参加竞赛的学生毕竟是少数。“要培养学生分析问题、解决问题的能力”,这是我国中学数学教学大纲中早有明文规定的。因此,数学教学中对全体学生进行类似数学建模竞赛的培训是十分必要的,这样能提高全体学生的数学应用能力。

采购的窍门——库存策略

在商品世界中,谁能以最小的代价提供或得到最优最多的商品,谁就是赢家胜者。精明的家庭主妇心中自有一套采购的窍门,对自家需要什么,需要多少,采购方式和时间,存贮多少,心中有数。经她的统筹安排和巧妙的采购,能使一个经

济并不富裕的家庭丰衣足食。

对易变质的蔬菜、鱼肉、鲜果等时效性极强的食品,主妇们每天采购,不作贮存,而且常在收市前采购,通常能降价10%~30%。而来去匆匆的上班族却只能赶早市采购,每次要多买一些,作些贮存,以减少采购次数,节约时间。这两种不同的采购策略,对一个家庭的开支显然是大不相同的。以一家三口为例,计划期 T 为一个月(按30天计算)作比较,假设每家各类鲜活食品总消耗量 Q 为60千克,平均单价为 $k=8$ 元/千克,进货及消耗为均匀的,每月采购次数为 n ,每次购

货量 $q = \frac{Q}{n}$,每次进货费用(如交通费等)

为 C_2 , $C_2=0.2$ 元/次。由于均匀消耗,则

平均库存量为 $\frac{1}{2}q$ (图25),单位食品单位

时间库存费为 C_1 , $C_1=0.4$ 元/千克·天(如电费、冰箱折旧等),则总费用 $E(t) = \text{购货款} + \text{进货费} + \text{库存费} = Q \cdot k +$

$C_2 \cdot n + \frac{1}{2}q \cdot T \cdot C_1$ 。 $Q \cdot k$ 为定值,假设主妇每月采购30

次,上班族每月采购10次,即使不考虑价格差异,主妇每月的总费用与上班族的总费用之差为:

$$\begin{aligned} & C_2(30-10) + (0 - \frac{1}{2} \times \frac{60}{10} \times 30 \times C_1) \\ & = 0.2 \times 20 - \frac{1}{2} \times \frac{60}{10} \times 30 \times 0.4 = 4 - 36 = -32 \text{ 元} \end{aligned}$$

主妇要比上班族少花费32元。如果考虑主妇以降价15%购货,则购货款可减少 $60 \times 8 \times 0.15 = 72$ 元,这样每月总费用可节约 $32 + 72 = 104$ 元。对于一般工薪阶层来说,这可是一笔不少的“收益”。对于商家,降价销售鲜活食品,不仅可

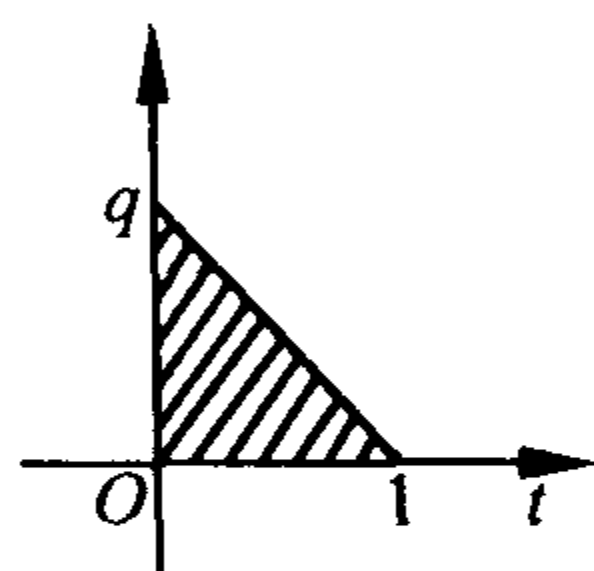


图 25

以避免因食品变质而造成经济损失,而且可节省存贮费用,经济上仍是合算的,何乐而不为。

近几年,批发打折之风盛行,除批发商店外,一些零售商店亦愿意批发销售。人们对于不易变质、可贮存的日用品,如牙膏、肥皂、洗衣粉、卫生纸等物品,愿意以批发价适量购进,适当贮存。这既不需库存费,且价格一般可降低 10% ~ 20%。对商家而言,让利批发销售决不吃亏。对于商家或生产厂家,从理论上讲,零库存最有利,这样可以减少库存费,不需占用流动资金,但难以实现。如果库存过少,虽可节省库存费用,但可能由于意外事故(如运输障碍等)而造成供不应求,导致停工待料或顾客转移,或失去时节而耽误生产,甚至造成重大影响,如电厂设备损坏,不能及时更换备件而造成停电,产生不可估量的损失。然而,库存过多,不仅过多占用流动资金或因贮存过久造成损耗,还会占用仓库和因需保持一定的保管人员而增加库存费用。因此,如何控制物资的库存量,采取合理的购货策略,既能保障供应,又能减少库存费用,是商家和厂家必须认真研究的问题。早在 1915 年,哈里斯就已对商业中的库存问题作了研究,建立了一个简单的数学模型,得到最优解。1918 年,威尔逊又重新得到哈里斯公式,但他们均未能引起人们的注意。到了 20 世纪 50 年代,库存理论才发展成为运筹学的一个分支。典型的库存问题是不断通过“供—存—销”三个环节满足需要。最简单的情况是需求量是均匀的,即需求率 R 为常数,不许缺货,一旦缺货能立即得到补充(能瞬时进货)。如果一个订货周期 t 只订货一次($n = 1$),订货量为 Q , $Q = Rt$,则:

$$\begin{aligned} \text{总耗资 } S(t) &= \text{货款} + \text{进货费用} + \text{库存费} \\ &= k \cdot Q + C_2 \cdot n + \frac{1}{2} C_1 Q t \end{aligned}$$

$$= kRt + C_2 + \frac{1}{2} C_1 R t^2$$

一个订货周期内的平均耗资为：

$$\frac{S(t)}{t} = kR + \frac{C_2}{t} + \frac{1}{2} C_1 R t$$

我们的目标是要在平均费用 $\frac{S(t)}{t}$ 最小的原则下，确定最佳订货策略。由于在总耗资计算公式中第一项货款 kQ 为常量，在考虑最小值时可不予考虑，只需考虑如何使后两项之和 $C(t) = \frac{C_2}{t} + \frac{1}{2} C_1 R t$ 最小。问题转化为求函数 $y = \frac{C_2}{t} + \frac{1}{2} C_1 R t$ 的最小值。可将此函数式转化为 t 的一元二次方程：

$$\frac{1}{2} C_1 R t^2 - y t + C_2 = 0 \quad (*)$$

因 t 是实数，故判别式 $\Delta \geq 0$ ，即：

$$(-y)^2 - 4 \cdot \frac{1}{2} C_1 \cdot R \cdot C_2 \geq 0$$

因 $y > 0$ ，故取 $y \geq \sqrt{2RC_1C_2}$ ， y 的最小值 $y_0 = \sqrt{2RC_1C_2}$ ，代入 $(*)$ ，得相应的 t 值为 $t_0 = \frac{\sqrt{2RC_1C_2}}{RC_1} = \sqrt{\frac{2C_2}{RC_1}}$ ，这就是最

佳订货周期。相应的订货量 $Q_0 = R t_0 = \sqrt{\frac{2C_2 R}{C_1}}$ ，这就是库存论中著名的经济订货批量公式，也称最佳批量公式或威尔逊公式，简记为 E.O.Q。 Q_0 称为经济订货批量或最佳批量。以 t_0 为订货周期，订货批量为 Q_0 时费用最小，最小费用 $\min C(t) = C(t_0) = y_0 = \sqrt{2C_1C_2R}$ 。从图 26 可见，费用曲线

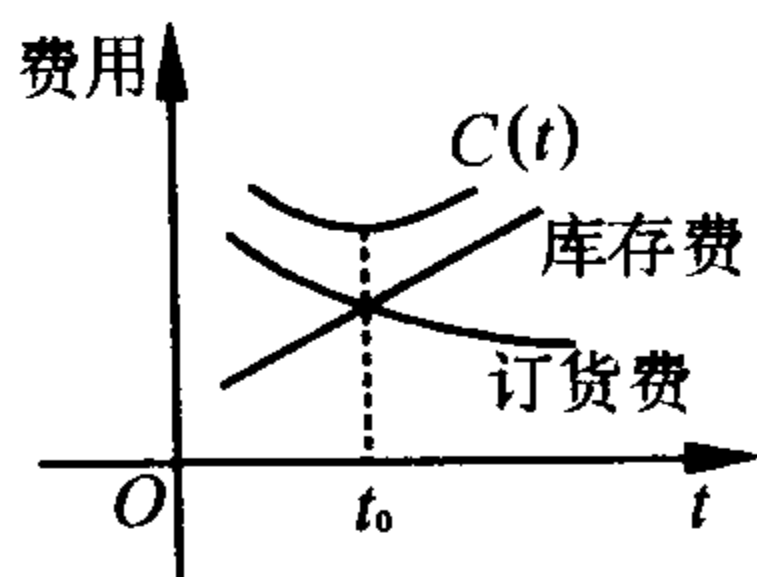


图 26

$C(t)$ 的最低点 $C(t_0)$ 的横坐标 t_0 正好与订货费曲线 $\frac{C_2}{t}$ 和库存费曲线 $\frac{1}{2}C_1Rt$ 交点的横坐标一致。

下面我们来看一个实际应用的例子。

某商店每月需要商品 100 件,即 $R = 100$ 件/月,每次进货费 C_2 为 5 元,商品到达后先进仓库,每件商品每月库存费为 0.4 元。如果不允许缺货,且能瞬时进货,应该怎样订货使耗资最小? 根据上面的公式,它的最佳订货周期 $t_0 = \sqrt{\frac{2 \times 5}{0.4 \times 100}} = 0.5$ (月),即每月订货 2 次。最佳订货量 $Q_0 = \sqrt{\frac{2 \times 5 \times 100}{0.4}} = 50$ (件),即每次订货量为 50 件。平均最小费用为 $C(t) = \sqrt{2 \times 0.4 \times 5 \times 100} = 20$ 元/月,即每月最小费用为 20 元。其库存状态如图 27。

上述库存问题是在资金不受限制的情况下进行的,如果资金或库房面积有限,仅允许每次订货 20 件,则一个月内应订货 5 次,此时库存状态如图 28,平均库存量为 10 件/天,进货费为 $5 \times 5 = 25$ 元/月,库存费为 $0.4 \times 10 = 4$ 元/月,总平均费用为 $25 + 4 = 29$ 元/月,显然比最小平均费用多 9 元,但节约了流动资金(即可以少付贷款利息)。所以,不同的问题应有不同的库存策略。

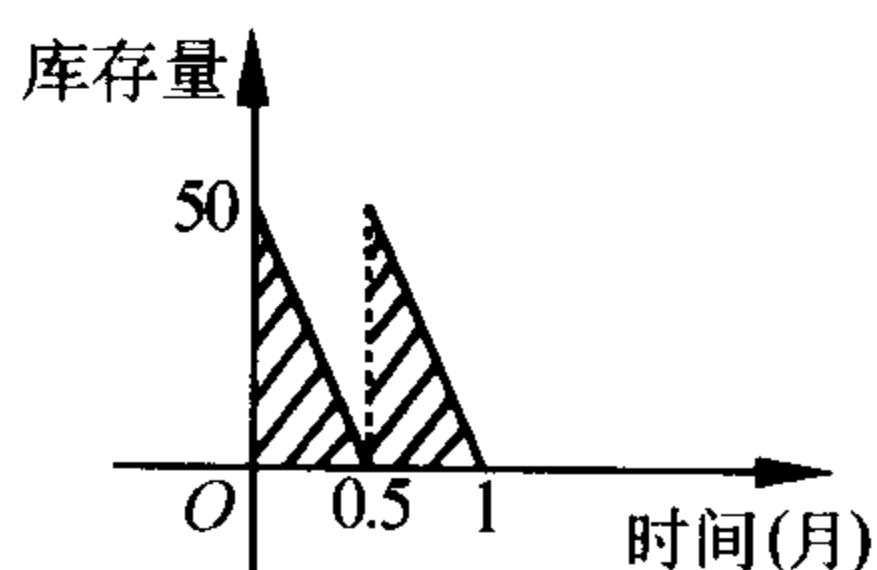


图 27

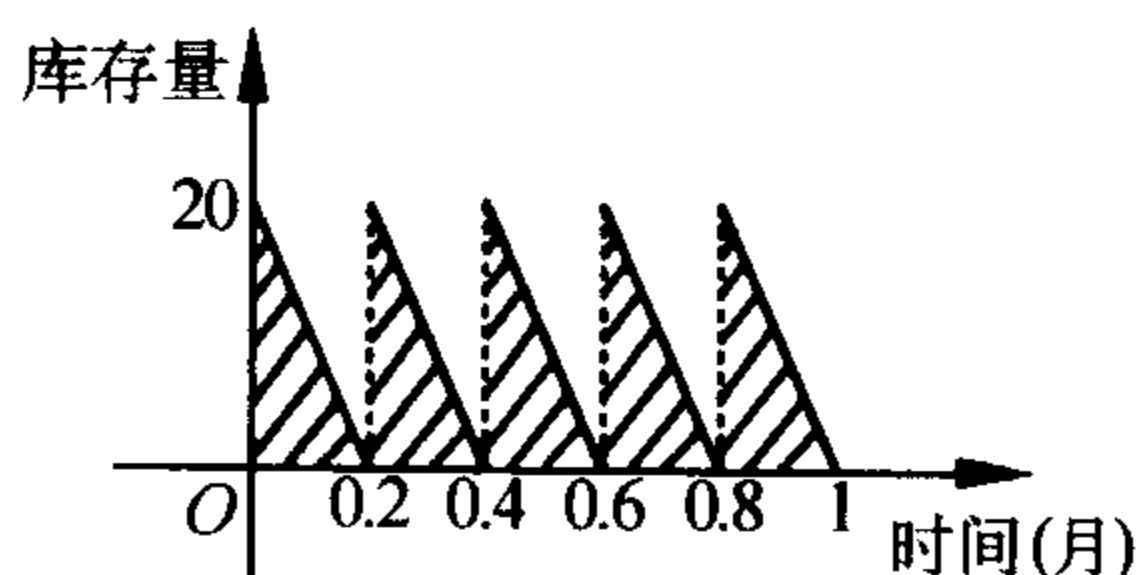


图 28

在实际生产经营中，还会遇到不许缺货、当库存量为零时立即开始生产，而生产需要一定时间，在时间 t' 内生产批量 Q ，生产速度 $P = \frac{Q}{t'}$ 必须大于需求速度

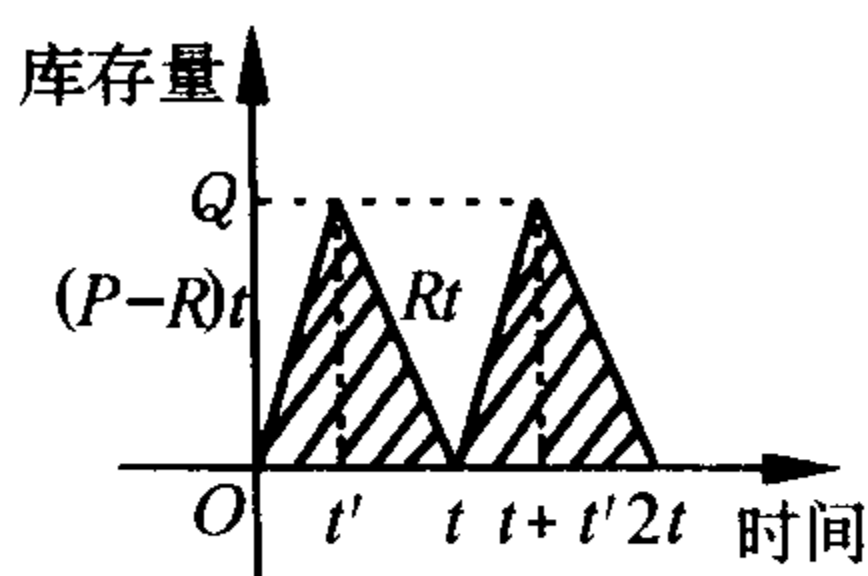


图 29

R ，才能保证不缺货，其他情况与前面相同。这种情况的库存状态如图 29。在时间区间 $[0, t']$ 中进行生产，库存量以 $(P - R)$ 的速度增加，直到库存量为 Q 才停止生产，然后在时间区间 $[t', t]$ 中库存量又以 R 的速度减少，直到库存量为零，又开始生产，进入另一周期。在总费用最小的原则下确定库存策略，求最佳生产周期 t_0 、最佳生产批量 Q_0 和最佳生产时间 t' 。由于在时间 t' 内生产的批量 $Q = Pt'$ 应满足这一周期内的需求量 $Q = Rt$ ，即 $Pt' = Rt$ ，故 $t' = \frac{Rt}{P}$ 。在周期时间 t 内的总库存量等于图 29 中带阴影的三

角形面积 $\frac{1}{2}t(P - R)t'$ ，平均库存量为 $\frac{1}{2}(P - R)t' = \frac{1}{2}(P - R)\frac{Rt}{P}$ 。若不考虑货款，时间 t 内平均总费用 $C(t)$

$= \frac{C_2}{t} + \frac{1}{2}C_1(P - R)\frac{Rt}{P}$ 。为求 $C(t)$ 的最小值，将函数式 $y = \frac{C_2}{t} + \frac{1}{2}C_1(P - R)\frac{Rt}{P}$ 转化为 t 的二次方程：

$$\frac{1}{2}C_1(P - R)\frac{R}{P}t^2 - yt + C_2 = 0$$

利用二次方程有实根 t 的判别式 $\Delta \geq 0$ ，即：

$$(-y)^2 - 4 \cdot \frac{1}{2}C_1(P - R)\frac{R}{P} \cdot C_2 \geq 0$$

$$\text{则 } y \geq \sqrt{\frac{2C_1C_2(P-R)R}{P}} \quad (y > 0),$$

$$y \text{ 的最小值 } y_0 = \sqrt{\frac{2C_1C_2R(P-R)}{P}},$$

$$\text{相应的最佳周期 } t_0 = \sqrt{\frac{2C_2P}{C_1R(P-R)}},$$

$$\text{最佳生产时间为 } t'_0 = \frac{Rt_0}{P} = \sqrt{\frac{2C_2 \cdot R}{C_1P(P-R)}},$$

$$\text{此时平均最小费用 } \min C(t) = y_0 = \sqrt{\frac{2C_1C_2R(P-R)}{P}}.$$

在某些情况下，允许缺货，当库存量降到零时，还可再等一段时间再订货，这样可以减少订货次数，少付进货费，同时库存费也相应减少，即使产生缺货损失，但综合考虑可降低总费用而不会造成不良影响。这种情况的库存状态如图 30。用类似方法可建立库存策略数学模型，确定最佳订货周期 t_0 ：

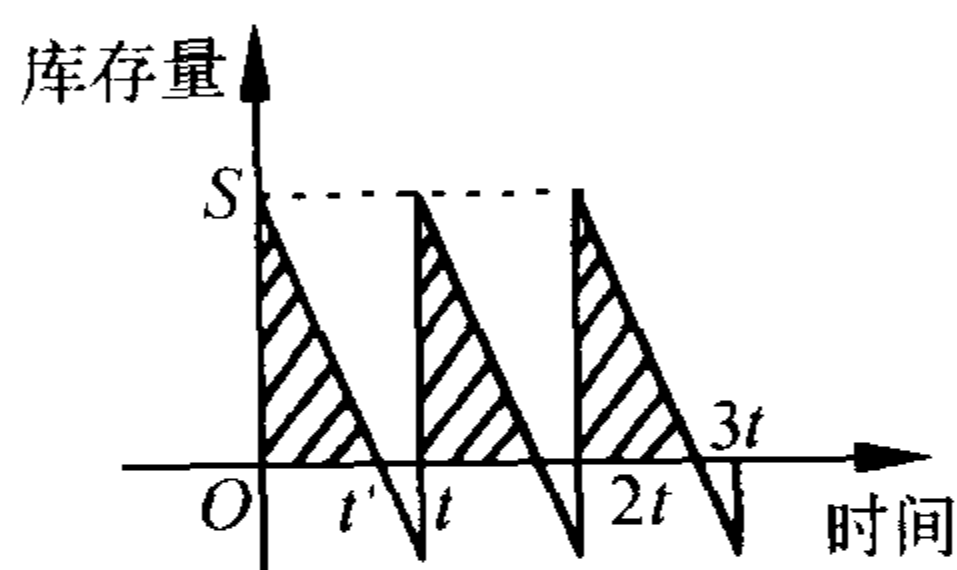


图 30

其中 C_3 为单位时间内缺货的损失费。

$$t_0 = \sqrt{\frac{2(C_1 + C_3)C_2}{C_1C_3R}}$$

其中 C_3 为单位时间内缺货的损失费。

$$\text{最佳库存量为 } S_0 = \sqrt{\frac{2C_2C_3R}{C_1(C_2 + C_3)}},$$

$$\text{时间 } t_0 \text{ 内最大缺货量为 } Rt_0 - S_0 = \sqrt{\frac{2C_1 \cdot C_2 \cdot R}{C_3(C_1 + C_3)}},$$

$$\text{最小平均费用 } \min C(t, S) = C(t_0, S_0) = \sqrt{\frac{2C_1C_2C_3R}{C_1 + C_3}}.$$

在考虑库存策略时,还可能会遇到一些不确定的因素。例如著名的“报童问题”:

报童每天从邮局订购报纸零售,而每天的销售量却是一个随机变量,如果订得过多卖不完时,就要遭受经济损失;如果订得过少,供不应求,收入也会减少。报童必须根据以往的销售情况确定每天的订量,使自己的收入达到最大(或者损失最小)。使收入达到最大与损失最小实质上是一致的,只是考虑问题的角度不同,可根据具体情况任选一种。若从损失的期望值最小的角度考虑,造成报童损失的原因有二:一是报纸售不出去造成损失;二是供不应求,失去销售机会造成损失。假设报童的订报数为 Q ,报纸的需求量为 r 的概率可凭经验或统计资料确定为 $P(r)$ 。报童每售一张报纸获利 a 元,如果卖不出去,每张报纸要赔 b 元,则当供大于求时,因售不完造成的损失期望值为 $\sum_{r=0}^Q b(Q-r)P(r)$,当供不应求时,因失去销售机会造成的损失期望值为 $\sum_{r=Q+1}^{\infty} a(r-Q)P(r)$,因此,报童损失的期望值 $C(Q)$ 为:

$$C(Q) = b \sum_{r=0}^Q (Q-r)P(r) + a \sum_{r=Q+1}^{\infty} (r-Q)P(r)$$

要求 Q 的值,使 $C(Q)$ 达到最小时,此时的 Q 即为最佳订货量。通过计算,可求得当 Q 满足 $\sum_{r=0}^Q P(r) = \frac{a}{a+b}$ 时,可使损失 $C(Q)$ 达到最小。

报童问题是具有代表性的简单单周期随机库存问题。报童问题的数学模型对于商店订购季节性商品或易腐烂变质的商品都有意义。因为季节性商品和易腐烂商品的销售情况对经济收益的影响事先难以预测。简单单周期模型的特征是对

于订货量的决策在整个需求过程中只进行一次,当库存耗尽时没有进货问题。同时规定两次订货之间没有联系,看做是独立的订货。

有些库存问题虽不能用上述模型求解,却可用运筹学的其他方法解决。

统筹法和优选法

华罗庚的愿望

华罗庚,这是个熟悉的名字,它鼓舞着成千上万的青少年在崎岖的科学道路上攀登。他的“天才在于积累,聪明在于勤奋”的教诲,鼓舞着人们坚定地迈向成才之路。

他教导青年人在学习上要“拳不离手,曲不离口”,提倡“弄斧必到班门”,敢于暴露自己的缺点找差距,提倡数学科学工作者要敢于与大数学家较量,找高手“下棋”。中国象棋有个规矩是“观棋不语真君子,举子无悔大丈夫”,华罗庚在一次演讲中却提

倡“观棋不语非君子,互相帮助;落子有悔大丈夫,改正缺点”。20世纪50年代,他针对中国科学院数学研究所的青年数学工作者只顾发表同一水平论文的偏向,及时指出:“不仅要速度,还要有加速度。”提倡“不仅要出科研成果,还要不断提高质量。”他在给故乡金坛中学的赠言中写道:“人说不到黄河心不死,我说到了黄河心更坚。”他的一生贯穿着不断拼搏,不断奋进的精神,直到老年仍奋斗不止。1979年,已近70高龄的华老仍表示“树老易空,人老易松,科学之道,戒之以空,戒之以松,我愿一辈子从实以终”,这是他向衰老抗争的决心书。他在第二次因心肌梗塞住院时,仍坚持工作,并提出“我的哲学不是尽量延长生命,而是尽量多做工作”的要求。有位记者曾采访他:“您最大的愿望是什么?”他不假思索地回答:“工作到最后一天。”1985年6月12日,华老在日本东京大学刚作完演讲,人们还在热烈鼓掌,鲜花还未送到他手中,这位数学巨人就因心脏病突发而去世。他实现了自己的诺言和愿望。

华罗庚是自学成才的典范。1910年11月12日,他出生在江苏省金坛县一个叫“乾生泰”的小店主家。1925年初中毕业后,由于家境困难,他考入由黄炎培主办的上海中华职业商校学习。后又因家庭经济日益困难而在毕业前一年辍学,回家帮助父亲经营小店。此时,他对数学产生了浓厚兴趣,经常从曾留法学物理的王维克老师处借数学书看,《大代数》、《解析几何》、《微积分》等都是他利用零星时间读完的。王维克十分欣赏华罗庚的数学才能。1930年,年仅19岁的华罗庚发表了题为《苏家驹之代数的五次方程式解法不能成立的理由》的论文,他的数学才能引起了当时清华大学数学系系主任熊庆来的注意。经熊庆来推荐,华罗庚于1931年到清华大学任数学系助理。仅有初中学历的华罗庚凭着坚强的毅力,

仅用了一年半的时间就学完数学系的全部课程,用英文写了3篇论文,先后在国外数学杂志上发表。1933年,他的数学水平与能力已为大家所认识,被破格提升为助教。一年后,他开始登台讲授微积分。同年,他担任了中华文化教育基金会董事乙种研究员,1935年提升为教员。在清华园,他除全面掌握了高等数学基础知识外,还直入数论前沿,致力于“华林”问题及有关问题的研究。他得到来华讲学的法国数学家阿达玛和美国数学家维纳的好评和指点,开始进入堆垒数论主流的研究。1936年,他以访问学者的身份赴英国剑桥大学进修,写了十几篇论文,他的关于“塔内问题”的研究被誉为“华氏定理”。当时许多人劝他考博士学位,这对于他已是探囊取物之事,而他却说“学位于我如浮云”。1937年,抗日战争爆发,充满爱国激情的华罗庚毅然回国,任西南联大教授。他除完成专著《堆垒数论》外,还研究群论、矩阵几何学、自守函数与多复变函数论等,从数论向多领域开拓,并取得卓越的成就。1946年7月,他在美国普林斯顿高级研究院作研究工作,后又在普林斯顿大学讲授数论。1948年春,他出任伊利诺伊大学终身教授。1949年,他举家回国,1950年2月,执教于清华大学,并负责筹备中国科学院数学研究所,1952年任数学研究所第一任所长。1955年,任中国科学院数理化学部委员,并担任数理化学部副主席,在此期间他完成了多部专著。

1965年后,他将主要精力用于在工业战线普及和应用数学方法上,亲自到工厂基层推广“优选法”和“统筹法”(简称“双法”)。在近20年的时间里,他的足迹遍及20多个省、市、自治区的工厂、矿山和企业。他亲自讲课,普及数学知识,取得良好的经济效益。他的“统筹方法平话及补充”及“优选法”是我国运筹学研究与应用成果的标志。他是中国最早为把数

学理论研究和生产实践紧密结合作出巨大贡献的科学家。他把数学方法创造性地应用于国民经济领域,并培养了一批为国民经济服务的科技队伍。华罗庚对数学应用的重视,不只反映在他亲自率队下基层推广“双法”上,他对数学的认识是全面的,既有深度又有广度,既重视理论研究又重视数学对生产实践和国民经济的作用。他自筹建数学研究所时起,就强调数学所的工作不只是偏重纯数学研究,还应注意数学知识的普及和应用。从理论到实践,从科学院专业数学人员到基层工程技术人员和中小學生,都在他的视野之中。1980年,身为中国科学院副院长的华罗庚仍兼任数学研究所和应用数学研究所所长职务。同年,在第四届国际数学教育大会上华罗庚应邀作一小时的大会报告,主题就是在群众中推广应用数学,推广统筹法、优选法。20世纪80年代,法国南锡大学、香港中文大学、美国伊利诺伊大学先后授予他荣誉博士称号,并被美国科学院、第三世界科学院、德国巴伐利亚科学院选为院士。华罗庚一生献身数学,是中国现代史上杰出的数学家,他是我国解析函数论、典型群、矩阵几何、自守函数论和多复变函数论等多方面研究的创始人与开拓者,他的学术成就达到很高的水平,正如数学家们所赞扬的:“他可以被选为任何学术团体的成员或任何科学院院士”——贝特曼;“他是这个时代的领袖数学家之一”——哈伯斯塔姆;“他的工作范围广阔,使他堪称世界名列前茅的数学家之一”——舍恩费尔斯。

协调技术——统筹法大意

1965年,华罗庚教授在全国推广统筹法和优选法时,曾

用一个简单的例子介绍统筹法的含义。例子是说,家中来了客人,想泡茶招待,可是开水没有了。于是主人要洗水壶(需1分)、灌水(1分)、烧水(15分)、洗茶杯(2分)、拿茶叶(1分)、泡茶(1分)。怎样安排这些工序,使客人尽快喝到茶水?通常有两种方案:一种是先将一切准备工作(洗壶、洗茶杯、拿茶叶)做好后再烧水(图 31);另一种是洗好水壶后烧水,在烧水过程中洗茶杯、拿茶叶(图 32)。

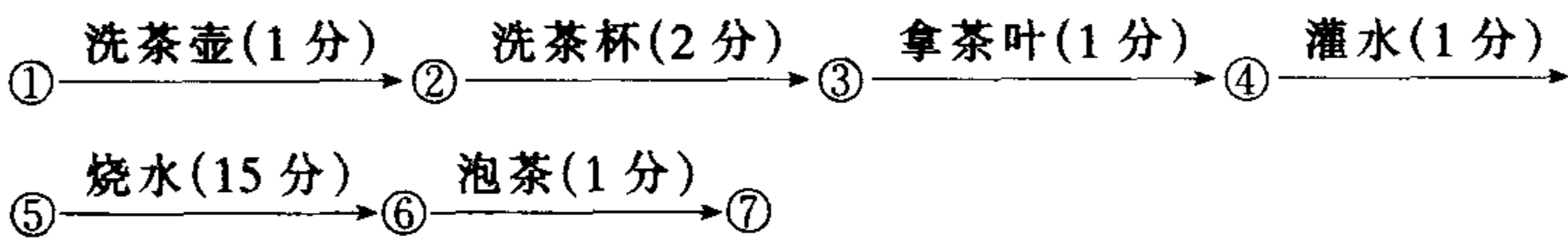


图 31

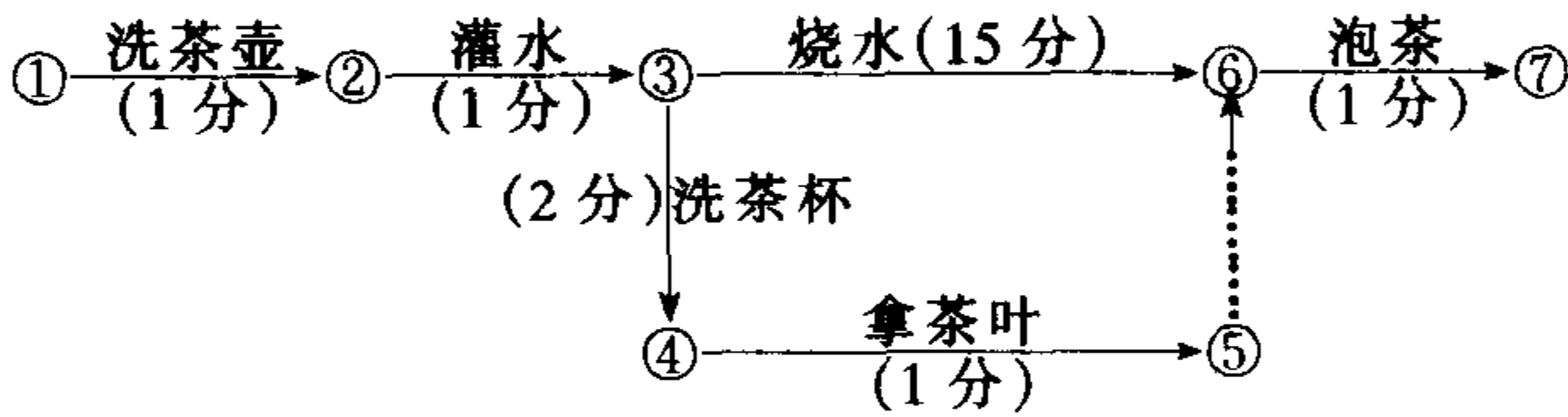


图 32

这两个示意图的工作顺序和时间都十分明确,第一种方案需要 21 分,客人才能喝到茶水,而第二种方案只需要 18 分,客人就能喝到茶水,比第一种方案少 3 分。

这种用箭线表示某项任务各工序之间关系的示意图常称为工序流程图,它可以将一项复杂的工作任务(不妨称为“工程”)全面具体展示出来。工序流程图中箭线表示工序,工序名称和所需时间标在箭杆上,两工序的交接点(称结点)用圆圈表示,它表示前一工序完工和后一工序开工,并按开工先后从左到右编号,整个图中只有一个总开工点和一个总完工点,图中虚箭线仅表示工序之间的逻辑关系,而实际上是不耗时间不费资源的“虚工序”,如图 32 中的⑤……→⑥。

长期以来,在生产规划、组织和管理上,一直沿用横道图方法安排工程进度。横道图是在一项工程的每个工序后面画一条水平线,用它表示工序进度的开始和结束时间。这种图虽简单直观,但不能反映各工序之间错综复杂和相互制约的关系,也不能反映哪些工序是主要的、关键的,不利于从工程全局考虑合理地组织生产、控制生产过程。而工序流程图不仅全面反映工序之间的关系,还可找出关键路线和关键工序,调整工序关系,合理安排,优化资源分配,以便顺利完成或提前完成任务。这种工序流程图深受管理人员和技术人员的欢迎,人们称它为网络图,是 20 世纪 50 年代才开始广为使用的。

关键路线是指对完成整个工程影响最大,从总开工点出发,到达总完工点所需时间总和最大的路线,是整个工程的主要矛盾线。图 32 中,共有两条路线:一条是① $\xrightarrow{1\text{分}}$ ② $\xrightarrow{1\text{分}}$ ③ $\xrightarrow{15\text{分}}$ ⑥ $\xrightarrow{1\text{分}}$ ⑦,需 18 分;另一条是① $\xrightarrow{1\text{分}}$ ② $\xrightarrow{1\text{分}}$ ③ $\xrightarrow{2\text{分}}$ ④ $\xrightarrow{1\text{分}}$ ⑤ $\cdots$ ⑥ $\xrightarrow{1\text{分}}$ ⑦,需 6 分。所以,前一条为关键路线。关键路线的长度(总时间)就是整个工程的完工期,是完成整个工程所需要的最少时间。关键路线上的工序是关键工序。因此,管理者要采取各种措施,努力使关键工序按时完成或提前完成,以确保整个工程按时或提前完成。

为了加快工程进度,通过对网络图的分析,有些工序可以同时进行或交叉进行,不同的工序同时进行称为平行作业,如图 32 中的洗茶杯、拿茶叶和烧开水。有些工序需要较长时间才能完成,只要条件允许,可以将相邻几个工序分批分期交替进行,这种方法称交叉作业法。例如建造高楼时,砌墙与搭脚手架这两个工序可以分段交替进行(图 33):砌第一段墙与搭

第一段脚手架同时进行；砌第二段墙时可利用已完工的第一段脚手架，同时开始搭第二段脚手架；砌第三段墙时便可使用已完工的第二段脚手架……画图时可用虚箭线表示其关系。

实际使用工序流程图时，除标出工序的完工期外，有时还要求标出每道工序完工的日期，甚至时、分，以便掌握进度。在制订计划时，有些项目的工序时间由于是新开发工序或使用了新工艺、新材料、新技术等原因无资料可参考，难以确定，只能给个估计值。方法是先给出三个估计值：最快可能完成时间 A 、最慢可能完成时间 B 、最可能完成时间 M ，然后按公式 $T = \frac{A + 4M + B}{6}$ 计算平均值，作为该工序时间，并将 $A-M-B$ 及 T 值标在流程图上。例如，图 34 表示某一工程项目。

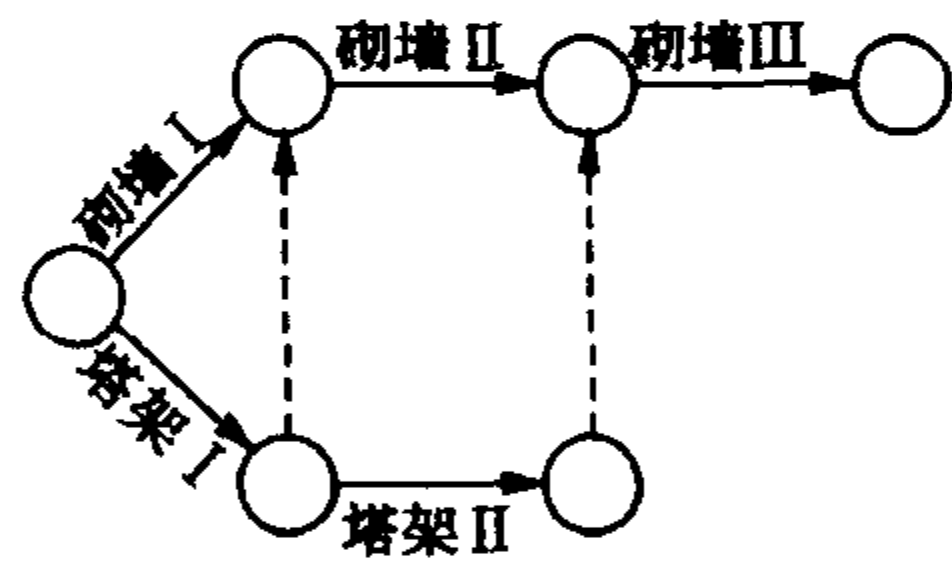


图 33

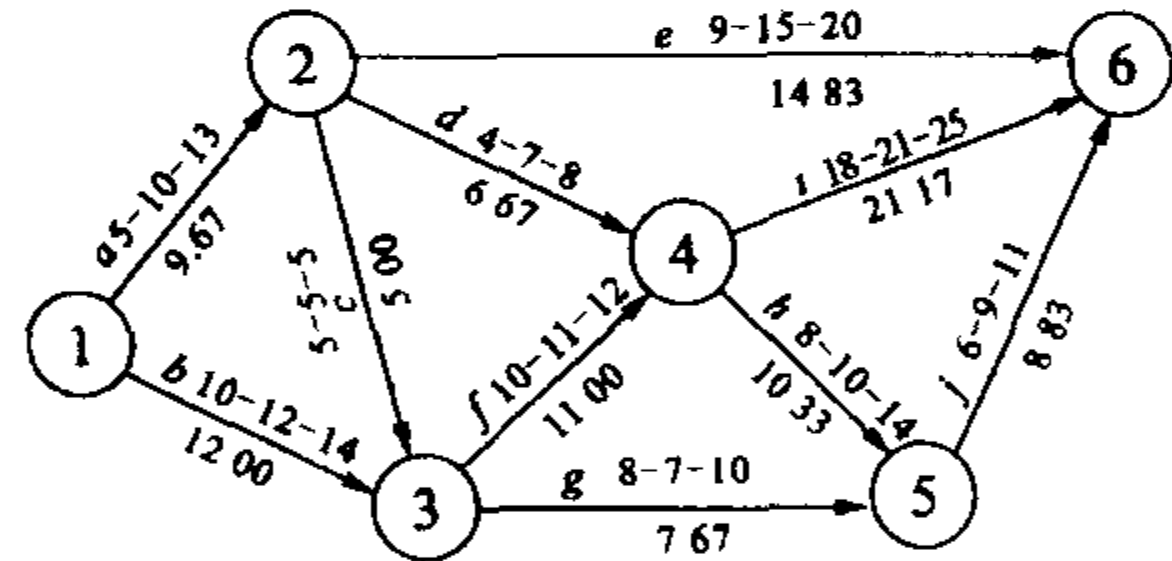


图 34

该工程项目的各工序时间估计值如下表：

工序名称	最短时间 A	最可能时间 M	最长时间 B	工序估计时间 T
a	5	10	13	9.67
b	10	12	14	12.00
c	5	5	5	5.00
d	4	7	8	6.67

续表

工序名称	最短时间 <i>A</i>	最可能时间 <i>M</i>	最长时间 <i>B</i>	工序估计时间 <i>T</i>
<i>e</i>	9	15	20	14.83
<i>f</i>	10	11	12	11.00
<i>g</i>	8	7	10	7.67
<i>h</i>	8	10	14	10.33
<i>i</i>	18	21	25	21.17
<i>j</i>	6	9	11	8.83

该项工程从总开工点到总完工点的路线共 9 条：

- ①→②→⑥ $9.67 + 14.83 = 24.50$
- ①→②→④→⑥ $9.67 + 6.67 + 21.17 = 37.51$
- ①→②→④→⑤→⑥
 $9.67 + 6.67 + 10.33 + 8.83 = 35.50$
- ①→②→③→④→⑥
 $9.67 + 5.00 + 11.00 + 21.17 = 46.84$
- ①→②→③→④→⑤→⑥
 $9.67 + 5.00 + 11.00 + 10.33 + 8.83 = 44.83$
- ①→②→③→⑤→⑥
 $9.67 + 5.00 + 7.67 + 8.83 = 31.17$
- ①→③→④→⑥ $12.00 + 11.00 + 21.17 = 44.17$
- ①→③→④→⑤→⑥
 $12.00 + 11.00 + 10.33 + 8.83 = 42.16$
- ①→③→⑤→⑥ $12.00 + 7.67 + 8.83 = 28.50$

由此可知,关键路线应是①→②→③→④→⑥,关键工序为 a 、 c 、 f 、 i ,工程预计完工时间为 46.84 天。

由于工序估计值是不确定量,故整个工程完工估计值 T 也是不确定量。这种靠估计时间得到的流程图称为非肯定流程图。

对一个工程项目用网络图进行分析,找出关键路线和关键工序,便于人们对这些工程的进度进行控制,使项目顺利或提前完成,以取得最优效果的科学方法,称为统筹法,也叫网络法。网络法是运筹学的一个分支。对复杂而庞大的项目,可以分解成几个分项目,每个分项目又可分解成许多小项目,要视其复杂程度设计工序流程图。一张工序流程图可以画得很细,将所有细小的工作都体现在图上,也可以画得概括些、综合些,突出大的协调关系,而把那些分项目及小项目另画工序流程图。这种综合性的工序流程图称为一级流程图,有关分项目及小项目的流程图称为二级流程图和三级流程图。分级画流程图,既有利于集中指挥,全面协调,又有利于分级管理,充分发挥各级管理人员的作用。统筹即协调,国外流行的计划评审法、关键路线法等都是通过建立网络图实现的。计划评审法也称为计划协调技术,是计划管理的一种方法,它是 1958 年美国海军部在研制北极星核潜艇计划过程中创造的,以数理统计学为基础,以网络分析为主要内容,以电子计算机为手段的新的计划管理方法,常用于大型、复杂的计划和工程。关键路线法是 1956 年美国杜邦公司提出的,其目的是以最少的时间完成给定的任务或工程。这些方法都体现了统筹协调的意思。1965 年,华罗庚教授开始在全国推广这些新的科学管理方法时,将其统一命名为统筹法。

0.618 与 优选法

0.618,一个美妙而又神奇的数字!被人们称为黄金数。爱神维纳斯和女神雅典娜的雕像下身长与全身长之比正是 0.618,人体的自身美也遵循 0.618。据大量调查统计,现今女性,下身的高度平均只占全身高的 0.58。为了达到 0.618,



图 35

不少女子穿上高跟鞋,芭蕾舞演员起舞时踮起脚尖;报幕员出台亮相时,不是站在舞台中央,而是偏向一方。画家则更懂得运用 0.618。早在 16 世纪欧洲文艺复兴时期,德国画家丢勒的名作(图 35),画面上仅有一双手,但手的中心位置却不在画面中央,而是在靠左下五分之三(接近 0.618)处。画家都懂得画的主体不宜放在画面正中的道理。右面三组图画(图 36)让人们选择,人们几乎无一例外地都选择右边的。建筑师们也很在行,窗户与房屋的矩形结构,其边长之比常选 0.618。在雅典,用于

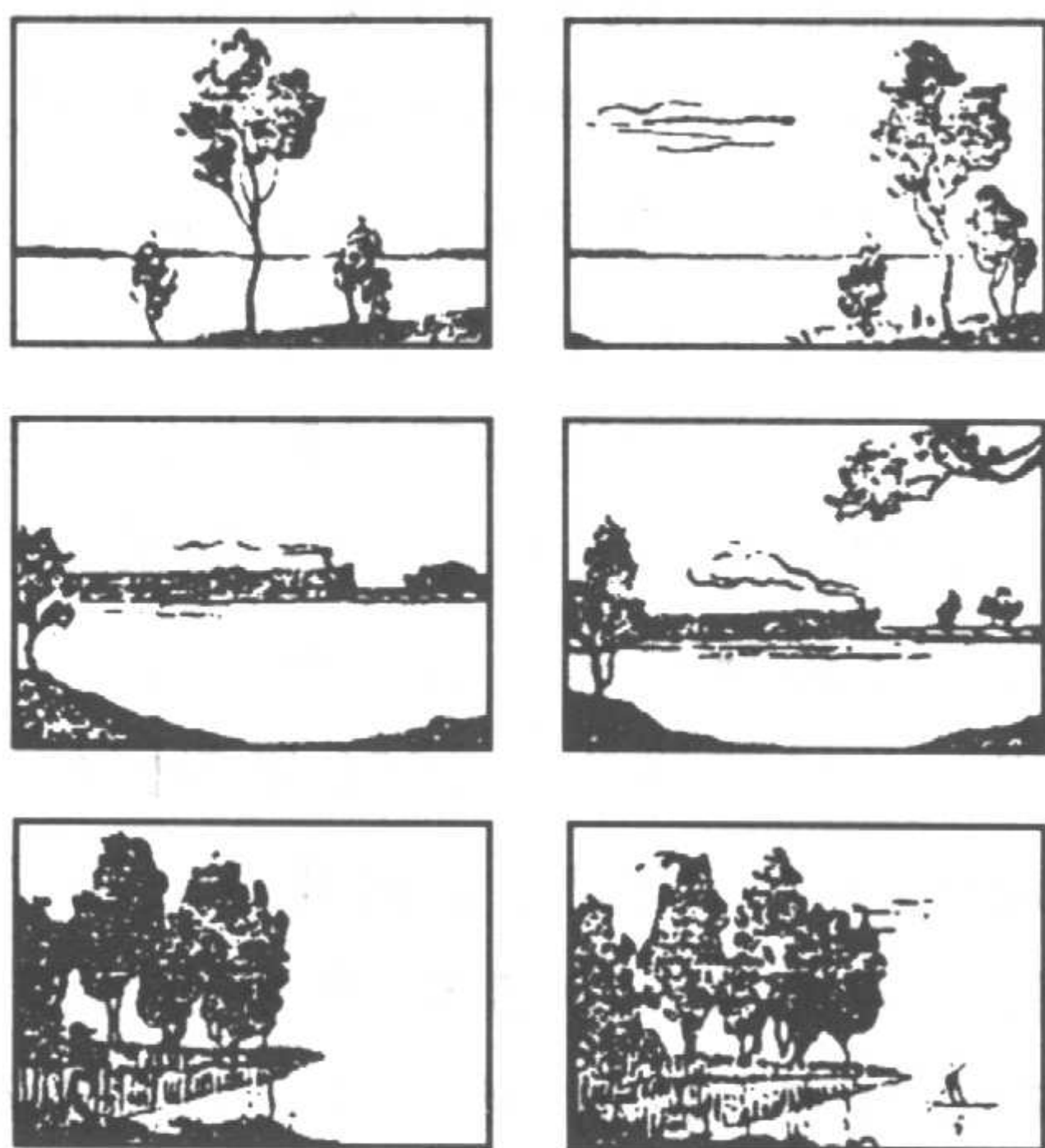


图 36

祭雅典娜女神的巴台农神殿就是按 0.618 的比例建造的。0.618 这个数被艺术家和建筑师们广泛地应用于他们的作品之中。

19 世纪,德国心理学家弗希纳曾经做过一次心理测试。他召开一次“矩形展览会”,会上展出了精心制作的各种矩形,要求参观者投票选择各自认为最美的矩形。结果有四种矩形入选,它们的宽与长之比分别是 $5:8 (=0.625)$ 、 $8:13 (=0.615)$ 、 $13:21 (=0.619)$ 、 $21:34 (=0.618)$ 。似乎人们对美的肯定近似于一个定数。无怪人们称 0.618 为“美的密码”。

具有如此魅力的 0.618,其真面目和由来如何? 公元前 5 世纪,哲学家毕达哥拉斯认为:“凡是美的东西,都具有共同的特性,这就是部分与部分和部分与整体之间的协调一致。”假定线段 AB 上有一个分点 C ,要使部分与部分和部分与整体之间协调,必须有 $CB:AC = AC:AB$ (图 37),即:

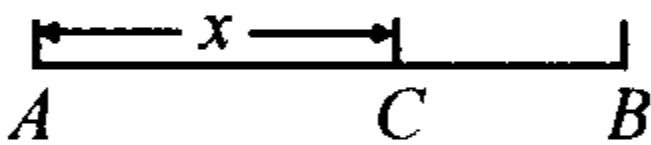


图 37

$$\frac{\text{一部分}}{\text{另一部分}} = \frac{\text{另一部分}}{\text{整体}}$$

不妨令 $AB = 1, AC = x (0 < x < 1)$, 则 $\frac{1-x}{x} = \frac{x}{1}, x^2 = 1 - x, x^2 + x - 1 = 0$, 解得 $x_1 = \frac{\sqrt{5}-1}{2} \approx 0.618033988 \cdots$ (舍去 $x_2 = \frac{-\sqrt{5}-1}{2} = -1.618033988 \cdots$)。

线段 AB 的这一分割称为黄金分割, C 点称为黄金分割点, 比值 $\frac{\sqrt{5}-1}{2}$ 称为黄金分割比, 记为 $W = \frac{\sqrt{5}-1}{2} \approx 0.618$ 。公元前 300 年左右, 欧几里得就用圆规直尺找出线段 AB 的黄金分割点(图 38): 过 B 作 AB 的垂线 MB , 取 $BM = \frac{AB}{2} =$

$\frac{1}{2}$ (设 $AB = 1$), 联结 AM , 在 AM 上截取 $DM = BM = \frac{1}{2}$, 再

在 AB 上截取 $AC = AD$, 由于 $\frac{AC}{AB} = \frac{AD}{AB} = \frac{AM - \frac{AB}{2}}{AB} =$

$$\frac{\sqrt{1^2 + \left(\frac{1}{2}\right)^2} - \frac{1}{2}}{1} = \frac{\sqrt{5} - 1}{2} = W \approx 0.618, \text{ 因此, 点 } C \text{ 就是 } AB$$

的黄金分割点。自欧几里得以后直到中世纪, 有许多数学家研究黄金分割比, 发现它有许多奇妙的性质。开普勒对黄金分割比作过很高的评价。他认为几何学有两大宝藏: 一个是毕达哥拉斯定理 (即勾股定理), 另一个就是黄金分割。宽与长之比等于黄金分割比的矩形称为“黄金矩形”。一个黄金矩

形 (图 39), 它的宽为 a , 长为 b , 则 $a:b = \frac{\sqrt{5}-1}{2} = W$ 。在该黄金矩形中以 a 为一边作一正方形, 剩下一个矩形, 它的宽与长

$$\text{之比等于 } \frac{b-a}{a} = \frac{b}{a} - 1 = \frac{1}{\frac{a}{b}} - 1 = \frac{1}{W} - 1 = \frac{2}{\sqrt{5}-1} - 1 = \frac{\sqrt{5}-1}{2}$$

$= W$ 。

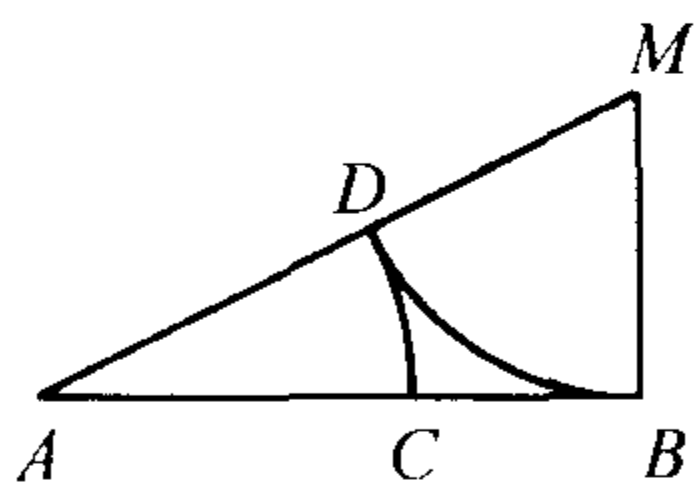


图 38

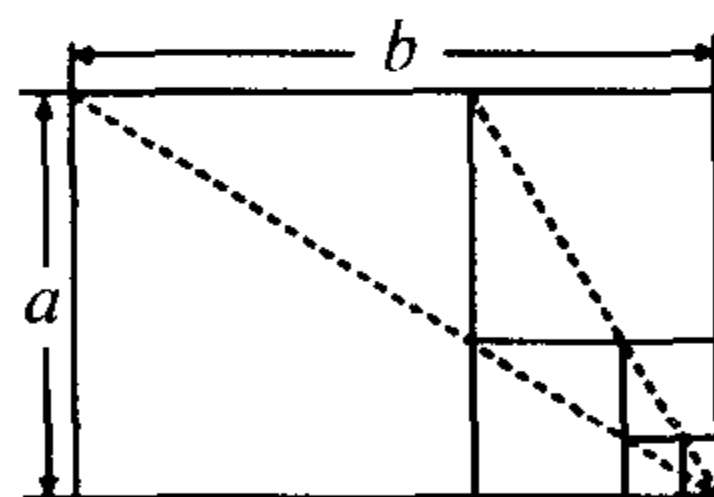
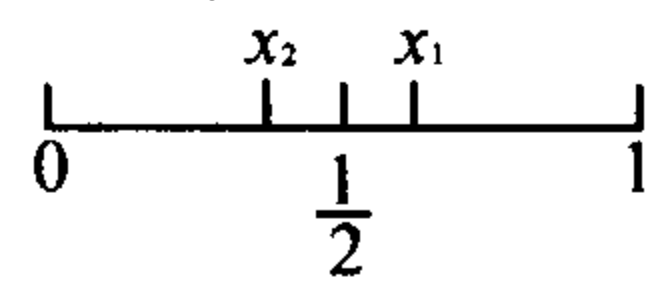


图 39

这说明剩下的矩形仍是黄金矩形。在剩下的矩形中, 以 $b-a$ 为边长作正方形后, 再在剩下的矩形中, 作一个边长为 $a - (b-a) = 2a-b$ 的正方形……以此类推, 可将黄金矩形分割成无限个等比递缩的正方形之和。从图上可见, 在黄金

矩形中,各正方形的角点形成两条直线,一条正是大黄金矩形的对角线,另一条是小黄金矩形的对角线。黄金矩形在造型艺术中有美学价值,比其他矩形容易引起美感,古希腊人就曾把黄金矩形引入他们的工艺品中。除黄金矩形外,还有黄金三角形(即顶角为 36° 的等腰三角形,其底与腰之比为黄金分割比)、黄金椭圆(短半轴与长半轴之比为黄金分割比,它的面积与半径为 $c = \sqrt{a^2 - b^2}$ 的焦点圆的面积相等)、黄金长方体(其长:宽:高 = $\frac{1}{W}:1:W$,它的表面积与外接球的表面积之比等于 $W:\pi$)。这些黄金图形都是建筑师和艺术家们所青睐的。黄金比简直是天赐的比例,一个漂亮的无理数。正如英国理论物理学家狄拉克说的:“上帝是位了不起的数学家,他用极高超的数学建造了宇宙。”英国著名物理学家琼斯宣称:“宇宙最伟大的建筑师原来是位纯数学家。”

0.618 不仅是“美的密码”,也是“优化天使”。在生产实践和科学试验中,为了得到优质、高产、低耗的效果,要寻求出主要因素的最佳点。而许多问题的目标与因素之间没有明确的数学表达式,或表达式很复杂,难以用函数求极值的方法解决,只能通过试验找出有关因素的最佳点。这种通过试验选优的方法称为优选法。如果目标所涉及的因素只有一个,就称单因素优选法,涉及的因素多于一个则称多因素优选法。通过试验进行优选的方法很多,就单因素优选法而言,常用的有 0.618 法、分数法、对分法、抛物线法等,其中以 0.618 法最受青睐,它能以较少的试验次数较快地找到最佳点。0.618 法的要点是先取试验点所在范围的



0.618 比例处做第一次试验,即第一点 $x_1 = \text{小} + (\text{大} - \text{小}) \times 0.618$ (图 40),再

图 40

在第一点关于试验范围的对称点 x_2 处做第二次试验,即第二点 $x_2 = \text{小} + (\text{大} - \text{小}) \times 0.382 = \text{小} + \text{大} - [\text{小} + (\text{大} - \text{小}) \times 0.618] = \text{小} + \text{大} - \text{中间试验点}$ (即简称为“加两头减中间”)。然后比较这两点的试验结果,去掉“坏”点外侧的部分,例如若 x_1 比 x_2 好,则去掉 $[0, x_2]$ 区间,再对留下部分继续做试验。再取中间试验点的对称点进行试验,作比较后决定取舍,逐步缩小试验范围。应用 0.618 法,每次试验可以去掉试验范围的 0.382,留下原试验范围的 0.618。试验范围随试验次数增加而缩小,第 n 次的试验范围为原来的 $(0.618)^{n-1}$,试验范围能迅速缩小,故能以较少的试验次数迅速找到最佳点。例如,某厂为了得到最好的零件表面质量,进行酞酐含量试验,这是一个单因素试验。试验范围是 100~200 克/升。用 0.618 法选优,

第一试验点酞酐含量为:
 $100 + (200 - 100) \times 0.618 = 161.8$ 克/升
 第二试验点酞酐含量为:
 $100 + (200 - 161.8) = 138.2$ 克/升
 比较两点的试验结果,发现第一试验点零件的表面质量比第二试验点好,因此去掉 100~138.2 这一范围,新试验范围变为 138.2~200 克/升,第三试验点是第一试验点在新试验范围中的对称点,继续试验 (图 41)。

第三试验点酞酐含量为 $200 + 138.2 - 161.8 = 176.4$ 克/升,比较第三试验点与第一试验点的试验结果,发现第三试验点零

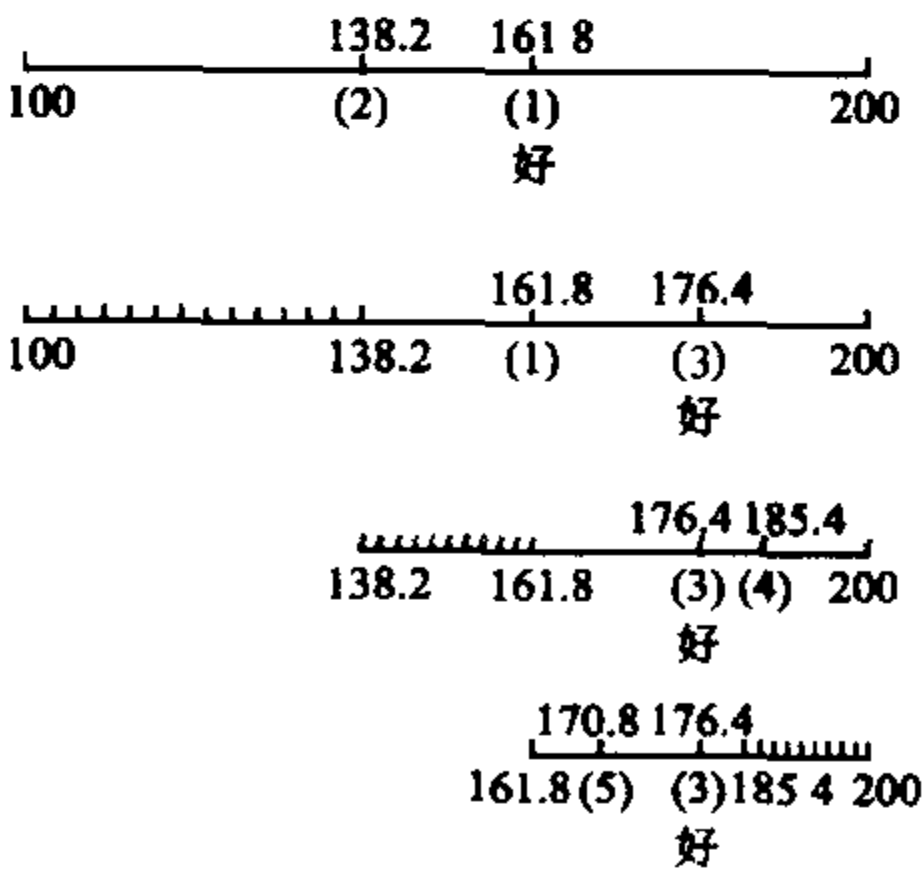


图 41

件的表面质量比第一试验点好,因此去掉范围 138.2 ~ 161.8,试验范围缩小为 161.8 ~ 200 克/升。

第四试验点酪酐含量为 $200 + 161.8 - 176.4 = 185.4$ 克/升,比较第四点与第三点的试验结果,还是第三点好,故去掉 185.4 ~ 200 克/升这一段,试验范围缩小到 161.8 ~ 185.4 克/升。

第五试验点酪酐含量为 $161.8 + 185.4 - 176.4 = 170.8$ 克/升,比较第五点与第三点的试验结果,还是第三点好,所以确定第三试验点是最佳点,酪酐含量应为 176.4 克/升,零件的表面质量最好。

0.618 法十分简便,只要记住“加两头减中间”或找对称点,便能掌握。如今,0.618 法已为广大工人、技术人员和管理干部所熟悉。

单因数优选法中的分数法、对分法与 0.618 法的不同点在于试验点的选择方法不同。分数法犹如 0.618 法的孪生兄弟,它适用于试验范围是一些不连续点、间隔不等的点组或试验点只能取一些特定数,只允许做一定次数试验的情况。分数法是按分数数列

$$1, \frac{1}{2}, \frac{2}{3}, \frac{3}{5}, \frac{5}{8}, \frac{8}{13}, \dots, \frac{F_n}{F_{n+1}}, \dots \quad (*)$$

(其中 $F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, F_5 = 8, \dots, F_n = F_{n-2} + F_{n-1}, \dots$) 的顺序安排试验点。例如,当只允许做一次试验(n

$= 1$) 时,就取试验范围的 $\frac{1}{2} \left(= \frac{F_1}{F_2} \right)$ 处做试验,其精确度为 $\frac{1}{2}$;

若只允许做两次试验($n = 2$) 时,选试验范围的 $\frac{2}{3} \left(= \frac{F_2}{F_3} \right)$ 处

做第一次试验,再选试验范围的 $\frac{1}{3}$ 处(正是第一次试验点的对

称点)做第二次试验,比较两次试验的结果,取优者为最佳点,其精确度为 $\frac{1}{3}$;若只允许做3次试验($n=3$)时,则选试验范围的 $\frac{3}{5}\left(=\frac{F_3}{F_4}\right)$ 处做第一次试验,再选择试验范围的 $\frac{2}{5}$ 处(它是 $\frac{3}{5}$ 的对称点)作第二次试验,比较这两个试验点的好坏,去掉“坏”点的外侧一段范围,缩小了试验范围,再找出“好”点在新范围中的对称点(即原试验范围的 $\frac{1}{5}$ 处或 $\frac{4}{5}$ 处)做第三次试验,并与留下的“好”点作比较,选其优者为最佳点,其精确度为 $\frac{1}{5}$ 。依此类推可知,当只允许做 n 次试验时,第一次试验应取在试验范围的 $\frac{F_n}{F_{n+1}}$ 处,然后按“加两头减中间”(即找对称点)的办法继续做试验,直到做完 n 次试验,便可得最佳点,其精确度为 $\frac{1}{F_{n+1}}$ 。例如,调试某设备线路时,需要选用一个电阻,调试者手中只有7个电阻,它们的阻值按大小排列分别为:(1)0.5k Ω , (2)1k Ω , (3)1.3k Ω , (4)2k Ω , (5)3k Ω , (6)5k Ω , (7)5.5k Ω 。由于阻值是离散的,如果用0.618法,没有合适的电阻。为此,可虚设两个点0.8,然后用分数法以 $\frac{5}{8}$ 做第一次试验。即取第五个电阻(3k Ω)做第一次试验,以后按“加两头减中间”法取第三个电阻(1.3k Ω)做第二次试验,比较两次试验结果,去掉“坏”点,继续做试验。共做了4次试验,即可找到较好的电阻。有时遇到试验范围中的点数不够分数的分母数时,也可分组试验,只是要适当缩小试验的范围。

0.618法与分数法的共同特点是从第二次试验开始不断缩小试验范围,但缩小率却不相同。取2到10个试验点的缩

小率作比较,结果为:

试验点数	2	3	4	5	6	7	8	9	10
分数法	0.333	0.2	0.125	0.077	0.048	0.029	0.018	0.011	0.0069
0.618 法	0.382	0.236	0.145	0.090	0.054	0.034	0.021	0.013	0.0081

可见 0.618 法的缩小率比分数法的缩小率大,所以 0.618 法比分数法能更快找到最佳点。由于分数数列(*)是由数列 $1, 2, 3, 5, \dots, F_n = F_{n-2} + F_{n-1}, \dots$ 构造而成,这个数列称为斐波那契数列,因此,分数法也称斐波那契法。可以证

明, $\lim_{n \rightarrow \infty} \frac{F_n}{F_{n+1}} = \frac{\sqrt{5}-1}{2} \approx 0.618$, 即分数数列(*)以黄金比为极

限,这说明分数数列中的项 $\frac{F_n}{F_{n+1}}$ 是黄金比的近似数。分数法

与 0.618 法真是殊途同归,就像是一对孪生兄弟。虽然

0.618 法和分数法是在 1953 年才由美国数学家基弗提出的,

但是,它们所涉及的斐波那契数列和黄金比都有着悠久的历史。

古希腊时,黄金比已被视为“贵比黄金”。斐波那契数列

是在 1202 年由意大利数学家莱翁纳多提出来的。这个数列

与一个有趣的“兔子繁殖”问题相关:假定一对刚出生的小兔子,

一个月后就长成大兔子,再过一个月便能生下一对小兔子,

此后每个月都能生

一对小兔子(图 42)。

一年内若不发生死亡

事故,问一对刚出生的

兔子,一年内能繁殖成

多少对兔子? 逐月推

算,从 1 月到 12 月,兔

子的对数应分别为:

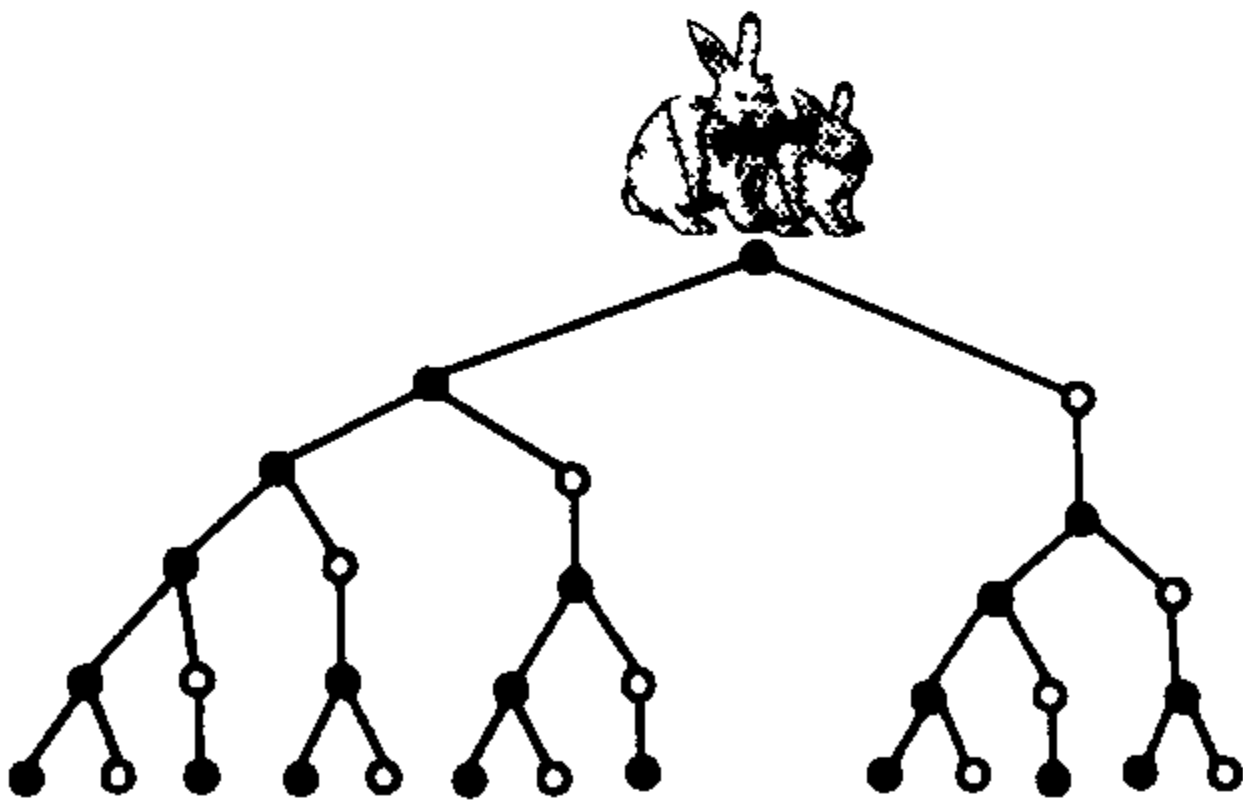


图 42

1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233。

这就是说,一对刚出生的兔子,一年内能繁殖成 233 对。一年内兔子每月的对数构成的数列正是斐波那契数列的前 12 项。

关于斐波那契数列有许多有趣的话题,它可以构造一些极为有趣的智力游戏。例如,一个魔术师,拿着一块边长为 13cm 的正方形地毯,要求他的朋友将它分成 4 块,再将它们拼成一个长 21cm、宽 8cm 的长方形。这位朋友对魔术师的算术水平之差表示惊讶,因为这一拆一拼会造成 1cm^2 面积的损失,这 1cm^2 的面积哪里去了(图 43)? 你知道吗?

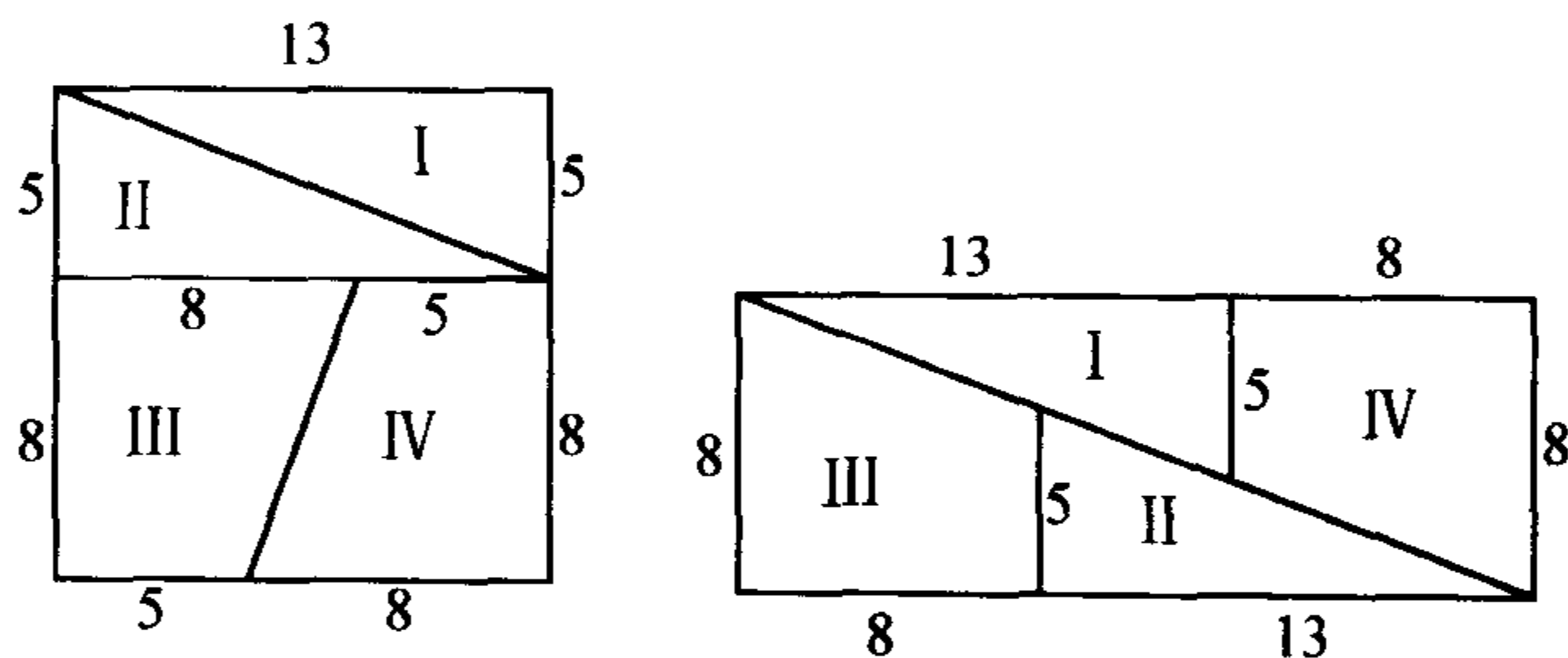


图 43

此外,自然界有许多生长变化过程都与 0.618 和斐波那契数列相关。例如,葵花上的反螺旋数目、松树节(枝丫)的数目都是相邻的斐波那契数。真是“天赐的比例”,妙不可言!

进一步研究

求优是人类的天性,在与大自然的斗争中,为了生存和发展,人们掌握了许多优化方法。在自给自足的经济环境下,要

依靠自己的力量丰衣足食。民以食为天,加工和储藏粮食是基本技能,于是家家都备有筛子:谷类用粗眼筛,粉类用细眼筛。走过建筑工地,总能见到用金属筛网分离砂子中的块石;在磨料厂,人们用直径大小不等的筛网将磨料分级,如此等等,不胜枚举。筛子的功能是优化批量实物个体的体积,达到整体优化的目的。筛选法虽古老,但是“筛选”的思想却具有普遍意义,永放光彩。层层筛选寻找优者,是当今各领域选优的基本思想。从打谷场上利用风力(风车或自然风)扬起谷物分离秕子,到高校入学考试、招聘人才,无不体现筛选的优化思想。不同性质的问题有不同的优化目标,因而采用不同的筛选方法。实物过筛以体积为标准,风力扬谷物以质量为标准,入学考试、人才招聘以成绩为标准。可见“筛选法”是比统筹法和 0.618 法更为直接和直观的优化方法。

还有一种经常使用的直接选优方法是模拟法,通过物理模型或实物模型进行模拟试验,得到最优方案。例如,工厂要建一个仓库为几个生产车间供应材料,仓库建在哪里才能使运费最小? 已知生产车间地点为 $P_i(x_i, y_i), i = 1, 2, \dots, n$, 假设运费近似等于距离与运量之乘积,在一定时期内各车间材料需求量为 W_i , 要确定仓库地点 $P(x, y)$, 使总运费最小, 即:

$$\min \sum_{i=1}^n W_i \sqrt{(x_i - x)^2 + (y_i - y)^2}$$

从表面上看,这个问题的目标(总运费)函数不太复杂,求这个函数的最小值还算容易。采用避开目标函数直接求优更为方便。做一个带有坐标刻度的平板,在相应的车间位置 $P_i(x_i, y_i)$ 处钻个孔,过每个孔穿一条细绳,使绳能在孔中滑动,绳的下端吊一个按比例缩小了的质量 W'_i , 垂在板的下

方,绳的另一端与同一个小圆环相连。小圆环最后停留的位置便是仓库的近似地点。

物理模型在实际中的应用非常广泛,如航空器的外形与它在空气中飞行时的受力关系密切,怎样的外形才能使飞行的阻力最小?通常利用风洞模拟太空环境对模型进行试验,对航空器的外形进行优选。自计算机模拟应用以来,许多复杂问题的选优得以又快又好地解决。

实践中常会碰到结构性质不清楚而仅有描述功能的数据的情况,这类问题就无法根据问题的性质进行选优,只能通过数据分析进行选优。例如,车间生产线突然出现异常现象,产品质量不稳定(出现超差),如何根据超差数据找出原因,控制生产过程,优化产品质量?又如,已知两个集体的功能数据,如两个班级学生的成绩、两个球队队员的身高等,如何判定它们的优劣?通常用平均值判别。如果两个集体的功能平均值相等时,又该如何区分其优劣?不妨设两个集体的成员的功能数据分别为 $x_1, x_2, \dots, x_n$ 及 $y_1, y_2, \dots, y_m$, 它们的平均值

为 $\bar{x}$ 与 $\bar{y}$, $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, $\bar{y} = \frac{\sum_{j=1}^m y_j}{m}$ 。因为 $\bar{x} = \bar{y}$, 每个成员与平均值的偏差分别为 $x_i - \bar{x}$ 和 $y_j - \bar{y}$, 由于这些偏离值有正有负, 总的偏离值不能简单相加, 为了防止正负值相抵消, 可用平均

偏离平方和 $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$, $\frac{\sum_{j=1}^m (y_j - \bar{y})^2}{m}$ 分别表示这两个集体各自的偏离程度, 称为方差, 记为 σ 。

σ 越大说明该集体的功能数值越分散, 功能参差不齐, σ 越小说明该集体的功能很均匀, 多集中在平均值附近, 所以 σ 越小越好。这种通过数据分析寻优的方法属数理统计学范畴。数理统计学是专门研究

那些事先性质不明,要靠数据分析去解决的各种问题的学问。

统筹法是通过网络图全面直观地反映工序、时间、路长、费用、物流等因素之间的逻辑关系,通过网络图的分析寻找最优方案。这种方法最早是由美国杜邦公司在兰德公司配合下于 1956 年提出的,目的是以最少的时间完成任务。国外称之为关键路线法或计划评审法,曾用于阿波罗登月计划,并取得显著效果,于是为人们所重视。20 世纪 60 年代这种通过网络分析寻优的方法不断发展,从最短路线到最关键路线、最大流量、最小费用等一连串问题的提出及其解法的问世,令人们耳目一新。它们的算法特点是利用网络分析,逐步搜索,搜索一次计算一次,逐次迭代直至得到最优结果。计算机的广泛应用使网络分析成为生机勃勃的数学分支。

单因素优选法的方法很多,除 0.618 法和分数法外,还有对分法、对生法、分批试验法等,它们的区别是选择试验点的方法不同,如对分法是每次取试验范围的中点做试验,而且必须能判断每次试验结果是“好”还是“坏”,以决定留下哪一半范围继续做试验。对于试验范围大,试验次数多的情况,采用分批试验法以加快试验进度,可减小试验代价。这些不同的优选方法有着两个共同的特点:一是具有序贯性,必须根据前面的试验结果安排后面的试验,试验范围随着试验次数的增加而不断缩小;二是搜索性,事先不能确定那个方向是好是坏,犹如瞎子走路,先用拐杖试探一下,能前进就向前迈步,遇到阻力不能前进就退回,另选方向试探,这种方法称为“一维搜索”。一维搜索也不限于前述的几种方法。多因素的优选法要比单因素优选法复杂,在理论上和方法上也不如单因素优选法成熟。对于多因素问题的优选,不同方法之间难以比较优劣,往往只能通过实际例子做相对比较。对多因素的优

选问题,往往是采用降维法:先固定一些因素,将问题转化为某个单因素问题进行优选,然后再对其他因素逐一进行优选。例如,解双因素问题,可先固定一个因素Ⅰ,用单因素方法对因素Ⅱ进行优选,取得最佳点,然后将因素Ⅱ固定,再对因素Ⅰ进行优选。由于固定某个因素的方法不同,产生许多不同的双因素优选方法,如纵横对折法、从好点出发法、平行线法、双因素爬山法、陡度法、矩阵法等,这些方法还可以混合使用。类似地,对三因素问题可用降维法,每次固定一个或两个因素,将问题转化为双因素或单因素问题。由于固定因素的方法不同而产生不同的优选方法,如平行平面法、平分平面法、从好点出发的平行平面法等。在多因素问题优选时,首先要找出影响指标的主要因素,即抓住主要矛盾。如果是多目标多因素问题,同样应找出影响各个目标的主要因素,根据实际情况,确定试验范围,选择优化方法。

筛选法、模拟法、数据分析法与统筹法、0.618法、分数法有共同的特点,即不依赖数学模型,直接求优。它们适用于那些不易建立数学模型或即使能建立数学模型但不易求解的问题。直接求优的方法往往是从实践活动中摸索得到,然后上升为理论,甚至独树一帜,成为应用数学的一个新的分支。从网络法(统筹法)到图论,从单因素到多因素,永无止境,最优化问题永远是最富时代气息的课题。

线性规划



历史概况与经济意义

在日常工作中,遇到重大问题需要决策时,总要做一番规划:如何利用已有的人、财、物等条件,取得最好的工作效果。有些人习惯于凭经验作决定,对重大问题常导致差之毫厘,失之千里的后果。经验是靠不住的,只有当经验上升为理论时才具有可靠性。最可靠的方法是将问题量化,利用数学方法找出好的方案。先将与问题有关的已知条件及要解决的目标都转化为数学式子,通过数学运算,求出最好的

解决方案。这种依靠数学方法作出规划的形式常称为数学规划。

生产单位如何根据生产设备能力安排各类生产计划(如年度计划按月分配、车间或工段的生产安排等);一种需要数种原料合成的产品,在保证质量要求的前提下,如何配料使成本最低;一笔可用于投资的基金,在有数种不同投资方案时,如何选择获利最多的方案;来自不同产地的资源如何调运到各需求地,从而使总运费最少;在保证营养需求的原则下如何配置食品,使费用最省……在国民经济中这类问题比比皆是。

例如,某工厂生产 A、B 两种轴类产品,都要经过车与磨两道工序,在现有车床、磨床设备条件和规定的工时定额要求下,如何安排生产使获利最大? 车床工段和磨床工段每天可利用的工时数分别为 60 和 40,工时定额及每件产品的利润如表:

应该如何安排这两种产品的生产任务?

	工时定额(时/件)	
	产品 A	产品 B
车 工	5	10
磨 工	4	4
利润(元/件)	10	15

假设 A 产品的日生产量为 x_1 (件), B 产品的日生产量为 x_2 (件),目标是所获得的总利润 $Z = 10x_1 + 15x_2$ 最大,即 $\max Z = 10x_1 + 15x_2, x_1 \geq 0, x_2 \geq 0$,限制条件是:

车床工段所占用的工时不超过 60,即:

$$5x_1 + 10x_2 \leq 60,$$

磨床工段所占用的工时不超过 40,即:

$$4x_1 + 4x_2 \leq 40。$$

这就是一个数学规划。

在纺织行业中,原料成本在产品成本中所占的比重很大,如何合理配料,降低原料成本,对降低产品成本有极其重要的意义。棉纺织成品常标有 32 支、80 支等各种棉纱标号,不同支数的棉纱有不同的质量标准。棉纱的拉力的质量指标是将拉力乘以试验棉纱支数的结果为标准,支数越大,质量越高。棉纱的棉结杂质是以单位面积中出现的疙瘩和杂质黑点粒数为标准,粒数越少质量越好。例如,21 支的棉纱的质量指标是拉力不能低于 2350,棉结杂质粒数不能超过 95 粒。假设某厂要生产“21 支”纱,现有 $C_1, C_2, \dots, C_8$ 八种原棉,已知它们的质量指标及价格(如下表),如何配棉,才能使所生产的 21 支纱成本最低?

品 种 项 目	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	21 支纱要求
拉力指标	2321	2574	2100	2196	2576	2300	2756	2660	≥ 2350
棉结杂质(粒)	80	80	75	60	80	80	100	90	≤ 95
混用上限	≤ 20	≤ 20	≤ 20	≤ 20	≤ 40	≤ 10	≤ 10	≤ 10	
原棉单价 (元/担)	124.96	139.80	139.80	155.99	139.80	139.80	147.89	139.80	最低

这个问题中待确定的量是配棉比例(百分数) $x_1, x_2, \dots, x_8$,这些变量必须满足:

(1) 拉力指标

$$\begin{aligned}
 &2321x_1 + 2574x_2 + 2100x_3 + 2196x_4 + 2576x_5 \\
 &\quad + 2300x_6 + 2756x_7 + 2660x_8 \geq 2350
 \end{aligned}$$

(2) 棉结杂质

$$\begin{aligned}
 &80x_1 + 80x_2 + 75x_3 + 60x_4 + 80x_5 + 80x_6 \\
 &\quad + 100x_7 + 90x_8 \leq 95
 \end{aligned}$$

(3) 配比要求

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 = 100\%$$

(4) 原棉混用上限及非负值为

$$0 \leq x_1 \leq 0.2, 0 \leq x_2 \leq 0.2, 0 \leq x_3 \leq 0.2, 0 \leq x_4 \leq 0.2, 0 \leq x_5 \leq 0.4, 0 \leq x_6 \leq 0.1, 0 \leq x_7 \leq 0.1, 0 \leq x_8 \leq 0.1$$

要求目标是成本函数最小,即:

$$\begin{aligned} \min C = & 124.96x_1 + 139.8x_2 + 139.8x_3 + 155.99x_4 \\ & + 139.8x_5 + 139.8x_6 + 147.89x_7 + 139.8x_8 \end{aligned}$$

这个数学规划的变量和约束条件都比前例多,且约束条件中的不等号方向不尽相同,但目标函数式及限制条件中的函数式都是一次式。

在国民经济领域和人民生活中,涉及的物资千千万万,由于自然条件限制,物资的产地与需求地往往不一致,绝大多数物资需要调剂。如何使产地的物资不积压,需求地又得到满足,做到产销平衡,即总供应量等于总需求量,同时还要使运费最低,以降低成本。这类问题既普遍又重要,被称为运输问题。例如,有两苹果产地 A_1, A_2 , 产量分别为 23 吨、27 吨,要供应 B_1, B_2, B_3 三个城市,这三个城市的需求量分别为 17 吨、18 吨、15 吨,从产地 A_1, A_2 到所需地 B_1, B_2, B_3 的运价(元/吨)如表。在产销平衡的前提下,如何调运才能使总运费最低? 设产地 A_i 到需地 B_j 的运量为 x_{ij} , $i = 1, 2, j = 1, 2, 3$ 。这些变量必须满足产销平衡的要求,即:

$$x_{11} + x_{12} + x_{13} = 23 \text{ (} A_1 \text{ 的总产量)}$$

$$x_{21} + x_{22} + x_{23} = 27 \text{ (} A_2 \text{ 的总产量)}$$

运价 产地 \ 需地	需地		
	B_1	B_2	B_3
A_1	50	60	70
A_2	60	110	160

$$x_{11} + x_{21} = 17(B_1 \text{ 的需求量})$$

$$x_{12} + x_{22} = 18(B_2 \text{ 的需求量})$$

$$x_{13} + x_{23} = 15(B_3 \text{ 的需求量})$$

要求总运费最小：

$$\min Z = 50x_{11} + 60x_{12} + 70x_{13} + 60x_{21} + 110x_{22} + 160x_{23}$$

这种产销平衡运输问题还可以推广到更一般的情况，即有 m 个产地 $A_1, A_2, \dots, A_m$ ，它们的产量分别为 $a_1, a_2, \dots, a_m$ 。有 n 个销售地 $B_1, B_2, \dots, B_n$ ，它们的需求量分别为 $b_1, b_2, \dots, b_n$ 。在产销平衡的前提下使总运费最小，即 $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$ 。假设从产地 A_i 运到销地 B_j 的单位运价为 C_{ij} ，如下表：

产销平衡表

运价 销地 产地	B_1	B_2	B_3	...	B_n	产地 A_i 产量
A_1	C_{11}	C_{12}	C_{13}	...	C_{1n}	a_1
A_2	C_{21}	C_{22}	C_{23}	...	C_{2n}	a_2
A_3	C_{31}	C_{32}	C_{33}	...	C_{3n}	a_3
$\vdots$ A_m	C_{m1}	C_{m2}	C_{m3}	...	C_{mn}	a_m
需求量 b_j	b_1	b_2	b_3	...	b_n	

寻找使总运价最小的调运方案。若以 x_{ij} 表示从产地 A_i 调运到销地 B_j 的货量，则此问题类似于苹果调运问题，它的规划形式为：

求目标函数(总运费)最小，即：

$$\begin{aligned}\min Z = & C_{11}x_{11} + C_{12}x_{12} + \cdots + C_{1n}x_{1n} + C_{21}x_{21} + \cdots \\ & + C_{2n}x_{2n} + \cdots + C_{m1}x_{m1} + \cdots + C_{mn}x_{mn}\end{aligned}$$

这些变量要满足的条件为：

$$x_{i1} + x_{i2} + \cdots + x_{in} = a_i \quad (i = 1, 2, \cdots, m)$$

$$x_{1j} + x_{2j} + \cdots + x_{mj} = b_j \quad (j = 1, 2, \cdots, n), x_{ij} \geq 0$$

运输问题与轴类产品加工问题、配棉问题所涉及的目标函数和约束条件都是变量的一次函数式,不同点是运输问题的约束条件都是等式,求目标函数的最小值。

不难发现,前述三个例子所描述的数学模型的共性有三:一是模型的基本结构相同,都是在给定的限制条件下,求某个指标的最优解。二是基本结构中都包含变量、约束条件和目标函数三个因素。变量的多少取决于问题复杂程度及对问题控制的粗细程度。要求控制的程度越细,变量个数就越多,也越能反映实际情况。每个变量代表要回答的某个小问题,它可以取实数。模型中的约束条件以等式或不等式表示,它反映实际问题的复杂程度,是实现目标的限制因素。目标函数是对要决策问题目标的数学描述,目标函数的最大值或最小值是一个求极值问题。三是目标函数及约束条件的函数式都是变量的一次式,即线性函数,这是很重要的特点。线性函数具有相加性和比例性两个特点。相加性是指数学模型中包含了多项活动,总的活动结果是各项活动结果之和。例如目标函数的总收益是全部单项活动的收益之和,资源总利用率是各种资源利用率之和。比例性是指目标函数或约束条件中变量的“收益”是与变量值成比例,即每个活动的总“收益”是单位数量乘每个单位的“收益”。如前述例子中生产轴类的利润与生产量是成线性比例关系。显然,具有这些共性的数学规划可归纳为如下的形式:

$$\begin{aligned} & \min(\text{或 max}) \quad c_1x_1 + c_2x_2 + \cdots + c_nx_n \\ & \text{约束条件} \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leq b_1 (\text{或} \geq b_1) \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \leq b_2 (\text{或} \geq b_2) \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \leq b_m (\text{或} \geq b_m) \\ x_1, x_2, \cdots, x_n \geq 0 \end{cases} \end{aligned}$$

这种数学规划称为线性规划。如果目标函数是求最小值,约束函数为等式,这时线性规划的形式为:

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

这种形式称为线性规划的标准形式。描述产销平衡的运输问题是标准形式的线性规划。但在实际工作中,往往出现产销不平衡的情况,或寻求目标函数极大值等问题,这些问题的线性规划都是非标准形式。对于非标准形式的线性规划可以转化为标准形式。求目标函数极大值的规划 $\max Z = c_1x_1 + c_2x_2 + \cdots + c_nx_n$, 可改写成等价形式 $\min(-Z) = -c_1x_1 - c_2x_2 - \cdots - c_nx_n$, 便是极小化形式。对约束条件是不等式的, 如 $x_{i1} + x_{i2} + \cdots + x_{in} \leq a_i$, 可引入一个新的未知量(称松弛变量) $x_{n+i} \geq 0$, 将不等式改写成:

$$x_{i1} + x_{i2} + \cdots + x_{in} + x_{n+i} = a_i$$

类似的, 可将不等式 $x_{1j} + x_{2j} + \cdots + x_{nj} \geq b_j$ 改写成:

$$x_{1j} + x_{2j} + \cdots + x_{nj} - x_{n+j} = b_j, x_{n+j} \geq 0$$

松弛变量可理解为变量所代表的点 x 到约束边界的一段“距离”。对于产销不平衡的运输问题,它的约束条件是不等式,当供应量大于需求量时,约束条件中会出现 $x_{i1} + x_{i2} + \cdots + x_{in} \leq a_i$,可引入松弛变量 $x_{n+i} \geq 0$,使 $x_{i1} + x_{i2} + \cdots + x_{in} + x_{n+i} = a_i$ 。引入松弛变量的意义是为了求得供求平衡,可虚设一个需求点,这相当于在供应地(产地 A_i)设立一个仓库,将部分多余的产品储存起来,其储存量 x_{n+i} 等于供应量 a_i 与需求量 $\sum_{j=1}^n x_{ij}$ 之差。这部分储存量是不需要运输的,它的运费为零,不会影响总运费(目标函数值)的最小值。当供应量小于需求量时,约束条件中会出现:

$$x_{1j} + x_{2j} + \cdots + x_{mj} \leq b_j$$

可引入松弛变量 $x_{m+j} \geq 0$,使

$$x_{1j} + x_{2j} + \cdots + x_{mj} + x_{m+j} = b_j$$

这相当于在需求点 B_j 虚设一个仓库,将供不应求的差额 x_{m+j} 另找门路补充在仓库中储存。由于仓库设在需求点上,也不需要运费,其运费为零,故不会影响总运费(目标函数)的最小值。既然非标准形式的线性规划问题都可以转化为标准形式,所以只要掌握标准形式的求解方法便可解决一般线性规划问题。

线性规划研究的问题主要有两大类:一类是一项任务确定后,如何统一安排,做到以最少的人力、物力完成任务;另一类是在一定量的人力、物力条件下,如何安排和使用,以发挥最大的效益。这两类问题是一个问题的两个方面,即寻求整个问题的某种指标的最优解。如今,线性规划已渗透到经济活动的各个领域,诸如生产组织、资源配置、交通运输、城市规划、工农业布局、服务网点、设备利用、仓库储备、环境优化、国

防建设等。

线性规划的研究最早始于 20 世纪 30 年代苏联学者康托罗维奇和希奇柯克等人研究生产组织和运输等问题。1947 年,美国联邦空军审计所的数学顾问丹齐格开发了一个数学规划,用于制定部署、训练、后勤保障的方案,并于 1948 年出版了《线性结构的规划》一书,从理论上为线性规划奠定了基础。1948 年夏,库普曼和丹齐格提出了求解线性规划的有效方法——单纯形法。至今,单纯形法仍在线性规划中占有重要地位。20 世纪 50 年代后,线性规划进一步与图论、网络理论以及一些数学基础学科(如拓扑学)联系起来,同时也发展了一些专门用来解决特殊线性规划问题的解法。随着计算机的发展,计算能力的提高,线性规划的应用范围也日益扩大,各种新的算法不断出现,如内点法、大步长跟踪算法、鞍面算法等。在 1951 年,国际水平只能解约束条件为 10 个方程的线性规划问题,到了 1963 年,就能解 1000~10000 个方程的线性规划问题。1956 年,解一个 67 个方程的线性规划问题要 1 小时,而到 1963 年只需要 28 秒。现在,不仅解题规模大、速度快,而且已有线性规划的专门程序,求解时只要通过计算机调用程序,输入数据,很快便可得出结果。正如 1982 年年近七旬的丹齐格教授在第 11 届数学规划大会上所说:“给 70 个人分配 70 项不同的任务,共有 $70!$ 种方案,是天文数字,要从中找出最优方案,即使用每秒能运算 10 亿次的大型计算机处理,也要从 150 年前开始直到太阳熄灭才会有结果。如果用单纯形法的软件,在电子计算机上计算,只需要几秒钟便可得出结果。”1984 年,美国贝尔实验室的数学家卡玛卡把射影几何原理用于大规模的线性规划问题,取得重大突破,成功地用于美国电话电报公司改建太平洋沿岸 20 个国家

的庞大电话网计划,求出了涉及 4.2 万个因素设计的最小投资数,曾引起轰动。

线性规划主要应用于企业规划、工农业生产、管理决策及经济领域,自 20 世纪 50 年代开始受到人们普遍重视,产生了巨大的经济效益。1975 年,瑞典皇家科学院把经济学的诺贝尔奖授予苏联的康托罗维奇和美国的库普曼,以奖励他们对资源最优分配理论的贡献。

求解思路与图上作业法

线性规划作为数学规划的一种类型,其特点是目标函数和约束方程都是线性函数。众所周知,线性函数的图象最简单最直观,这不仅为线性规划的求解提供了既直观又可想像的图解方法,也为线性规划的一般解法(单纯形法)提供了清晰的思路。线性函数的特性(相加性和比例性)使某些问题(如运输问题)甚至不需要通过数学模型而直接采用图上作业求解。

为了说明线性规划一般解法(单纯形法)的求解思路,以两个变量的线性规划图解为例。

$$\begin{aligned} \text{例 1} \quad & \max \quad 2x_1 + 5x_2 \\ & \text{约束条件} \begin{cases} x_1 + 2x_2 \leq 8 \\ x_1 \leq 4 \\ x_2 \leq 3 \\ x_1 \geq 0, x_2 \geq 0 \end{cases} \end{aligned}$$

显然,满足约束条件的点就是图 44 中的阴影区域(即凸多边形 OABCD)。目标函数为 $Z = 2x_1 + 5x_2$, 当 Z 取不同

值时,它代表斜率为 $-\frac{2}{5}$ 的平行线族。求最优解就是要在阴影区域中寻找使 Z 值最大的点。不妨平移直线 $x_2 = -\frac{2}{5}x_1$,让它扫过阴影区域,阴影区域中最后与直线接触的点 $C(2,3)$ 就是使 $z = 2x_1 + 5x_2$ 的值最大的点,即 $x_1 = 2, x_2 = 3$ 为最优解,相应的目标函数取最大值为19。

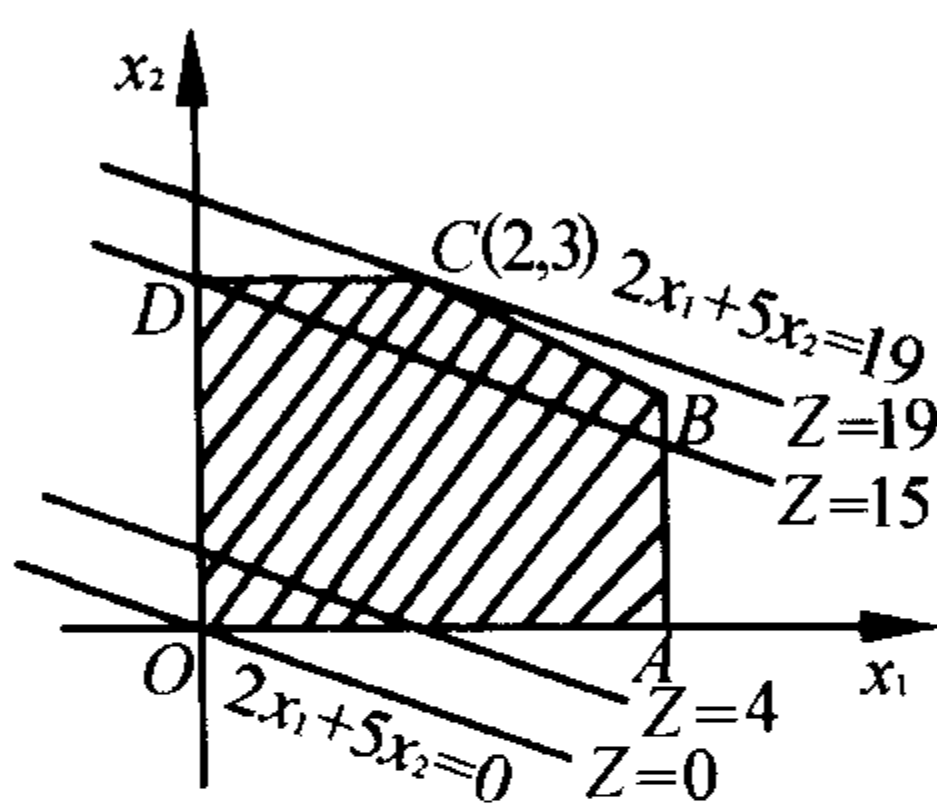


图 44

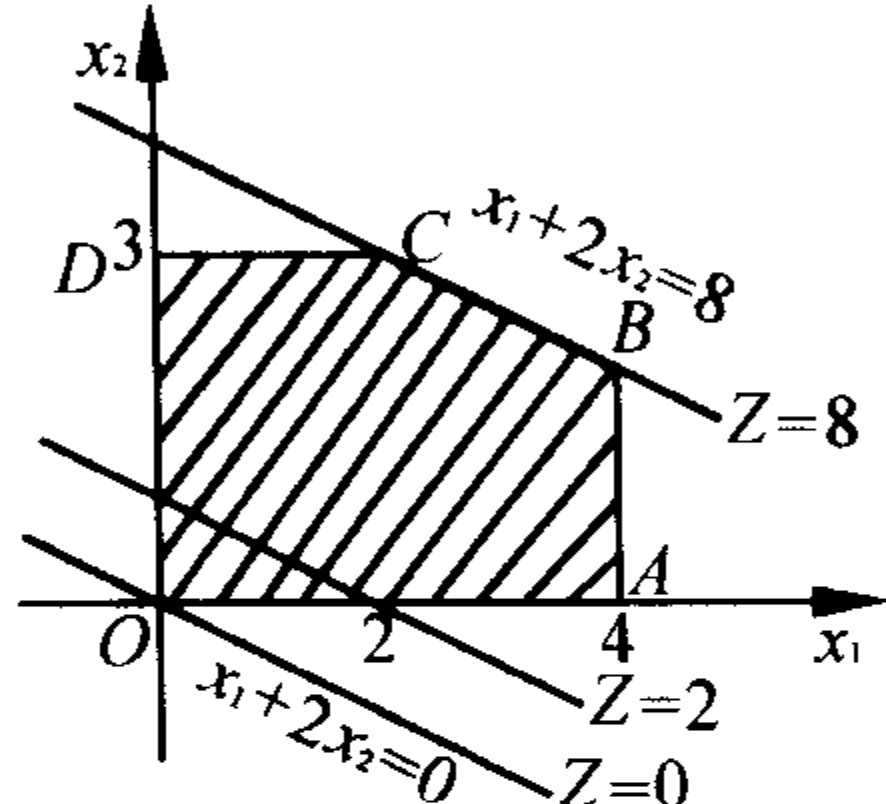


图 45

例 2 $\max \quad x_1 + 2x_2$
约束条件(与例 1 同)

满足约束条件的点的集合仍是阴影区域($OABCD$)(图 45),求最优解时只需平移直线 $x_2 = -\frac{1}{2}x_1$ ($Z = 0$),当该平移直线扫过阴影区域时与多边形的边 BC 重合,这说明线段 BC 上的点满足 $x_1 + 2x_2 = 8$,都是最优解,相应的目标函数的最大值为 8,这个线性规划有无穷多个解。

例 3 $\max \quad x_1 + x_2$
约束条件 $\begin{cases} x_1 - x_2 \geq 1 \\ -x_1 + 2x_2 \leq 0 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$

满足约束条件的点的集合为阴影区域 $ABCD$ (图 46),是

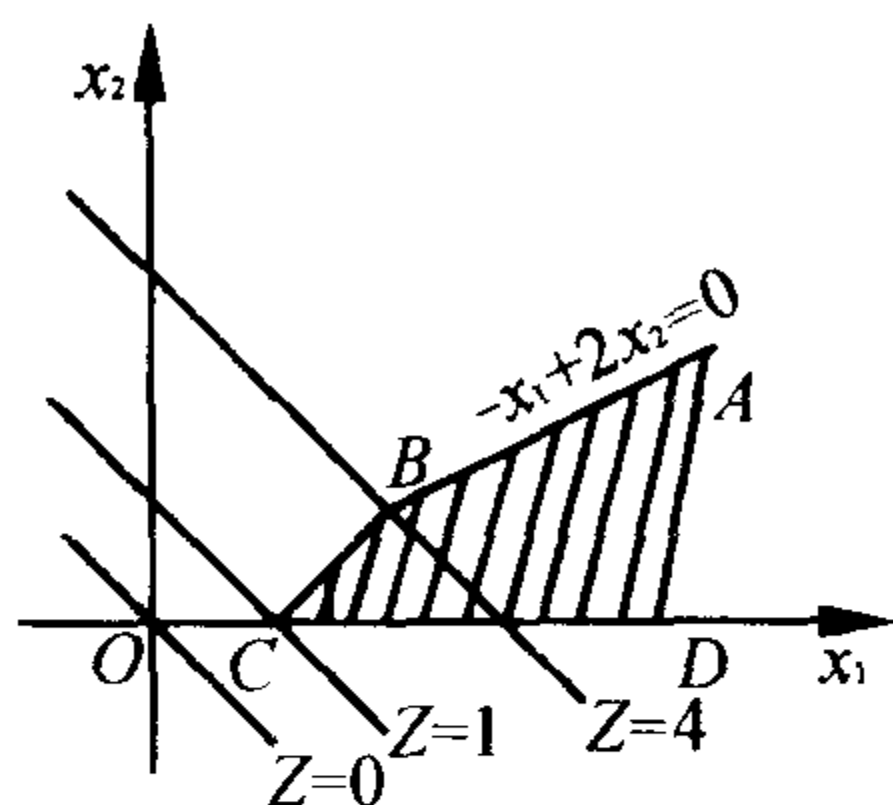


图 46

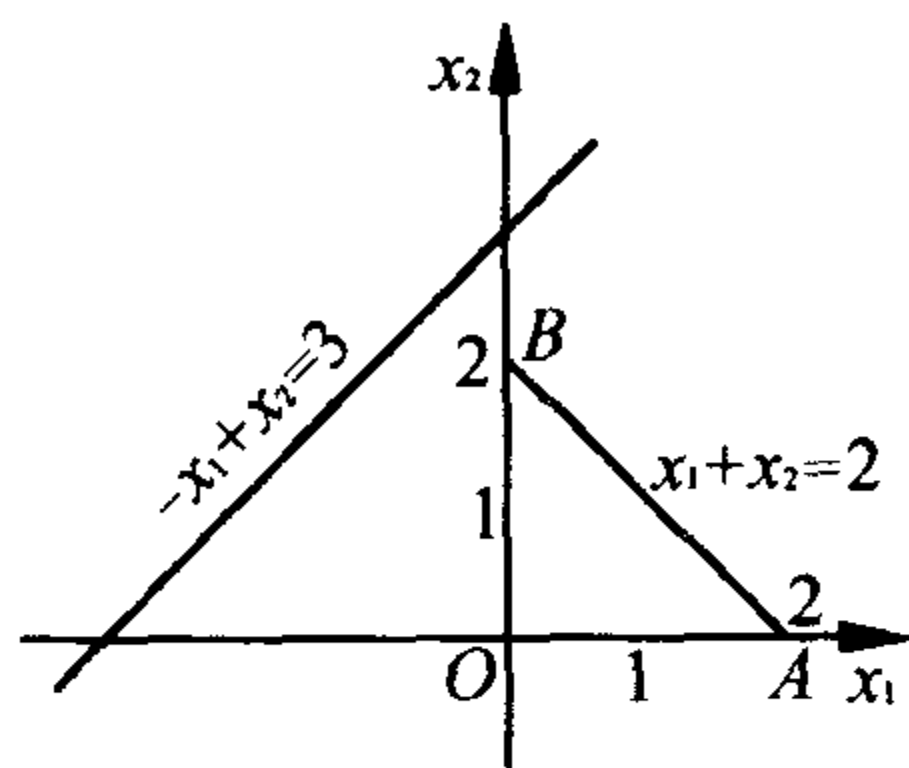


图 47

无界区域。因此,目标函数 $x_1 + x_2$ 的值不可能有上界,取不到最大值,这个线性规划是无解的。

例 4 $\min x_1 + x_2$

$$\text{约束条件} \begin{cases} -x_1 + x_2 \geq 3 \\ x_1 + x_2 \leq 2 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

由于约束条件是不相容的,所以满足约束条件的点集是空集(图 47),故此问题不可能有解。

从上述 4 个两个变量的线性规划的图解情况可得出如下启示:线性规划问题不一定都有解;如果有解,可能有惟一解或无穷多解。由于约束函数都是线性的(两个变量的线性函数是直线),可证明满足约束条件的点集若不是空集则一定是凸多边形。又由于目标函数是线性的,它有最优解时一定可以在凸多边形的顶点取得(不只是在顶点),因此,含两个变量的线性规划求解时,可先求出满足约束条件的点集,然后计算每个顶点的目标函数值,选其最优者即为所求的解。例 1 及例 2 中满足约束条件的点集是同一个凸多边形 $OABCD$,顶点坐标分别为 $O(0,0)$, $A(4,0)$, $B(4,2)$, $C(2,3)$, $D(0,3)$ 。例 1 各顶点的目标函数值为: $Z_O = 0$, $Z_A = 8$, $Z_B = 18$, $Z_C =$

19, $Z_D = 15$, C 为最优解, 即 $x_1 = 2, x_2 = 3$, 相应的最大目标函数值为 19。例 2 中各顶点的目标函数值分别为 $Z_O = 0, Z_A = 4, Z_B = 8, Z_C = 8, Z_D = 6$, 故 B 和 C 为最优解, 由于约束条件及目标函数都是线性的, 图象为直线, 故线段 BC 上的所有点都是最优解。所以两个变量线性规划的解可从满足约束条件的凸多边形顶点中去寻找。已经证明, 这个结论对多于两个变量的线性规划问题也是有效的, 这也是线性规划的一般解法——单纯形法的基础。由于线性规划的约束条件是有限个的, 因此解的允许域的顶点数目和边界条数都是有限的, 可以从顶点中选出一个作为初始点, 由一个顶点转移到另一个顶点, 经过有限次转移必可得出结果。显然, 当变量增加时, 顶点的数量也增加, 寻找顶点也需花一番功夫计算。许多实际问题中变量和约束方程的个数很多, 计算量十分大, 这正是电子计算机的用武之地。

交通运输被称为国民经济的大动脉, 它不仅要求物资运送流畅, 还要求省时省钱, 它是经济发展的“先行官”, 也是线性规划发展的“先行官”。运输问题是一类特殊的线性规划问题, 除线性规划一般解法外, 还有特殊的解法。我国最早应用线性规划、发展线性规划理论的正是运输问题。20 世纪 50 年代, 东北是我国重工业和交通运输最发达的地区, 是煤、粮食生产基地。如何合理调运煤、粮食到各地, 是新中国面临的重要课题之一。当时, 东北有个物资调运小组, 他们在实践中创造了一种不同于单纯形法, 而且极易为人们掌握的“图上作业法”, 有效地解决了煤、粮食的合理调运。1958 年, 中国科学院数学研究所的万哲先在和大家讨论的基础上, 最终对图上作业法作了严格的理论证明, 并在全国推广应用, 为我国运输业作出了卓越的贡献。图上作业法方法简单、直观易懂、便

于推广,比传统的代数方法有更大的优越性。我国是从 20 世纪 50 年代末期开始推广运筹学的,作为运筹学主要分支的线性规划得到了广泛的应用,尤其是交通运输及生产计划等部门,在运用线性规划后取得很大的经济效益。

图上作业法适用于在一定交通线上,使用同一种交通工具调运某种物资,要使总运费最小的运输问题。由于运费与吨数、千米数成正比,所以运费最小便是重量、里程数最小。图上作业法的步骤是:先作物资流向图,再在图上调配,使运输距离最短。作流向图的方法是:以圆圈表示产地或仓库(发货点),发货数量标在圆圈内,以方框表示需求地(收货点),收货量标在方框内;两点间的连线表示交通线,两点间距离标在交通线旁边;用箭线“ $\longrightarrow$ ”表示物资流动方向,画在前进方向交通线的右侧。物资通过的数量(加上圆括号)标在箭线前进方向的右边,这就构成一张物资调运流向交通图(图 48)。从流向图可看到,相反流向的

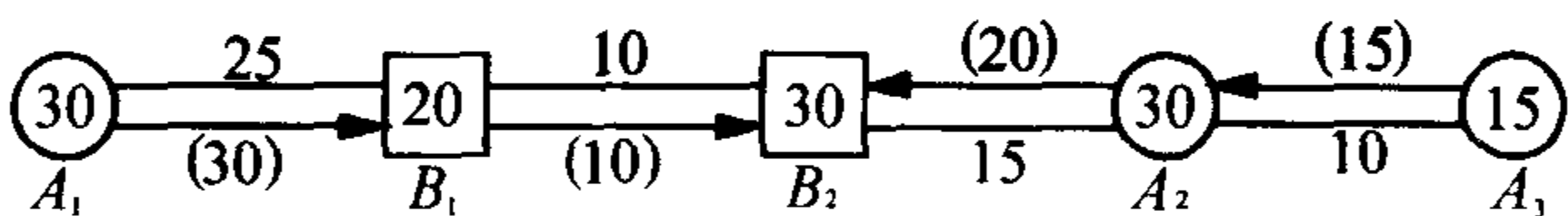


图 48

流量正好标记在交通线的异侧。从各发货点调到各收货点的物资调运方案可以有很多,为寻找运输费用最小的方案,必须避免“对流”和“迂回”的运输,使总距离最短。避免“对流”就是要避免同一物资在同一线路上往返运输。图 49 中有 15 吨物资从 A_1 经 A_2 到 B_2 ,又有 15 吨物资从 A_2 经过 A_1 最后运到 B_1 ,可见 A_1A_2 段有 15 吨物资往返运输,这是不合理的,应调整流向图,改为图 50,这样就避免了对流运输,可节约 $2 \times 15 \times 40 = 1200$ (吨·千米) 的运费。

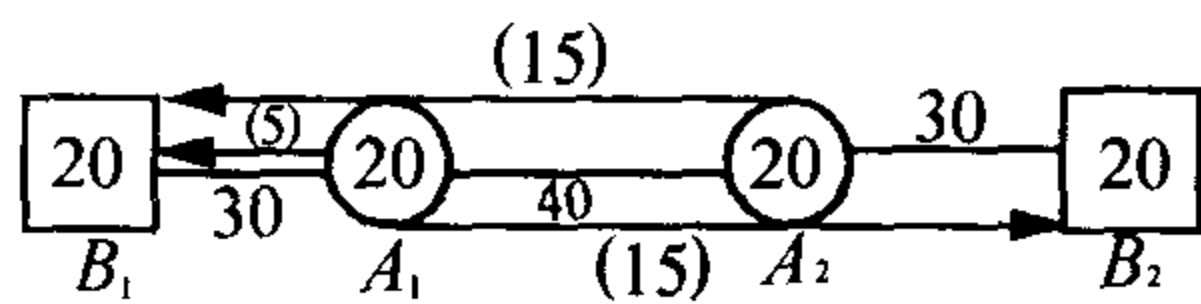


图 49

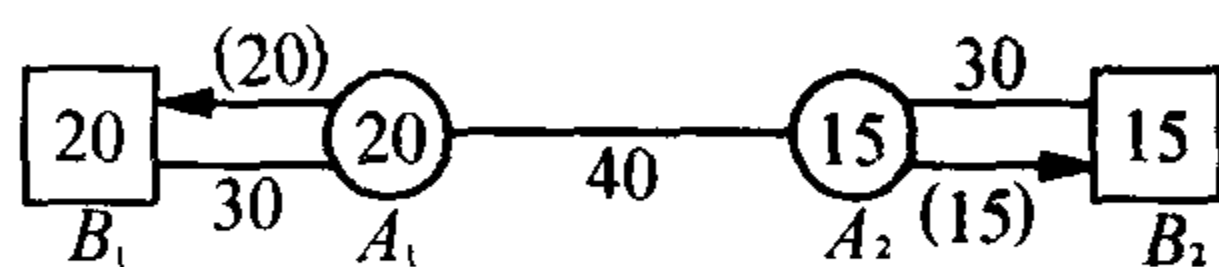


图 50

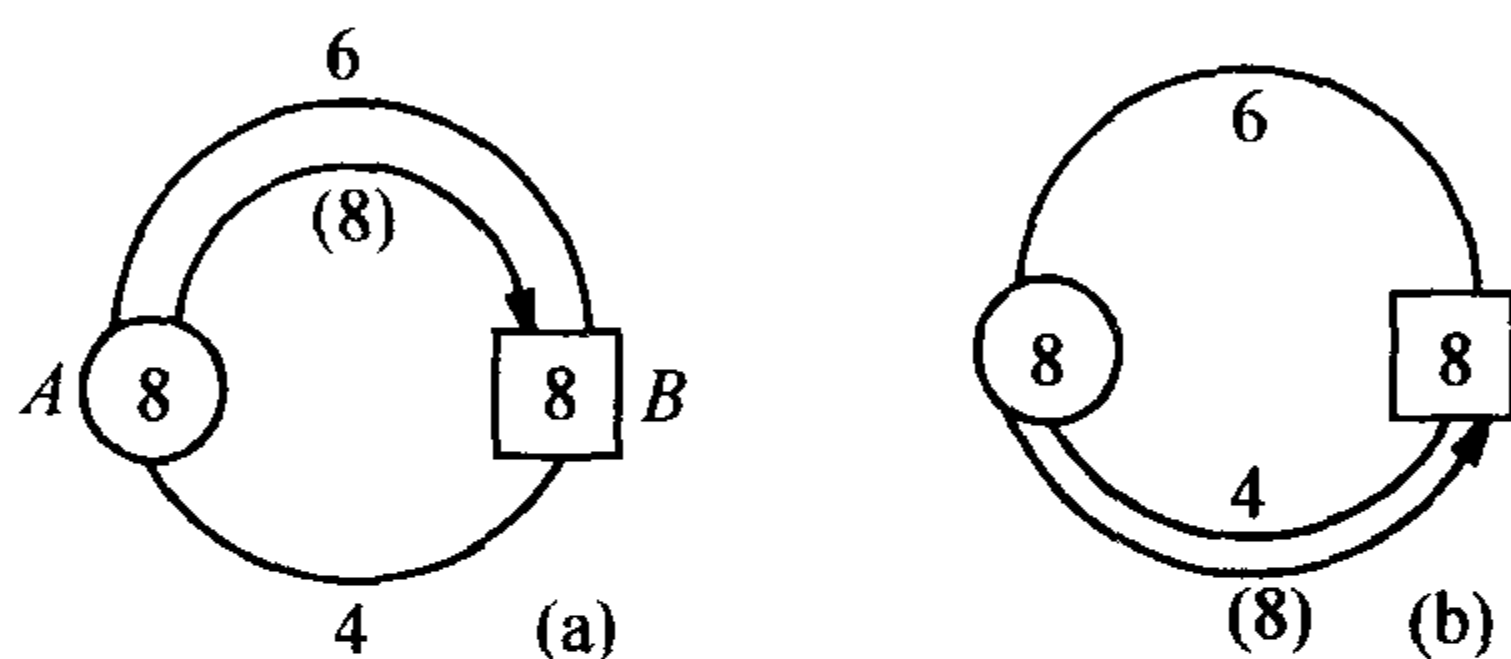


图 51

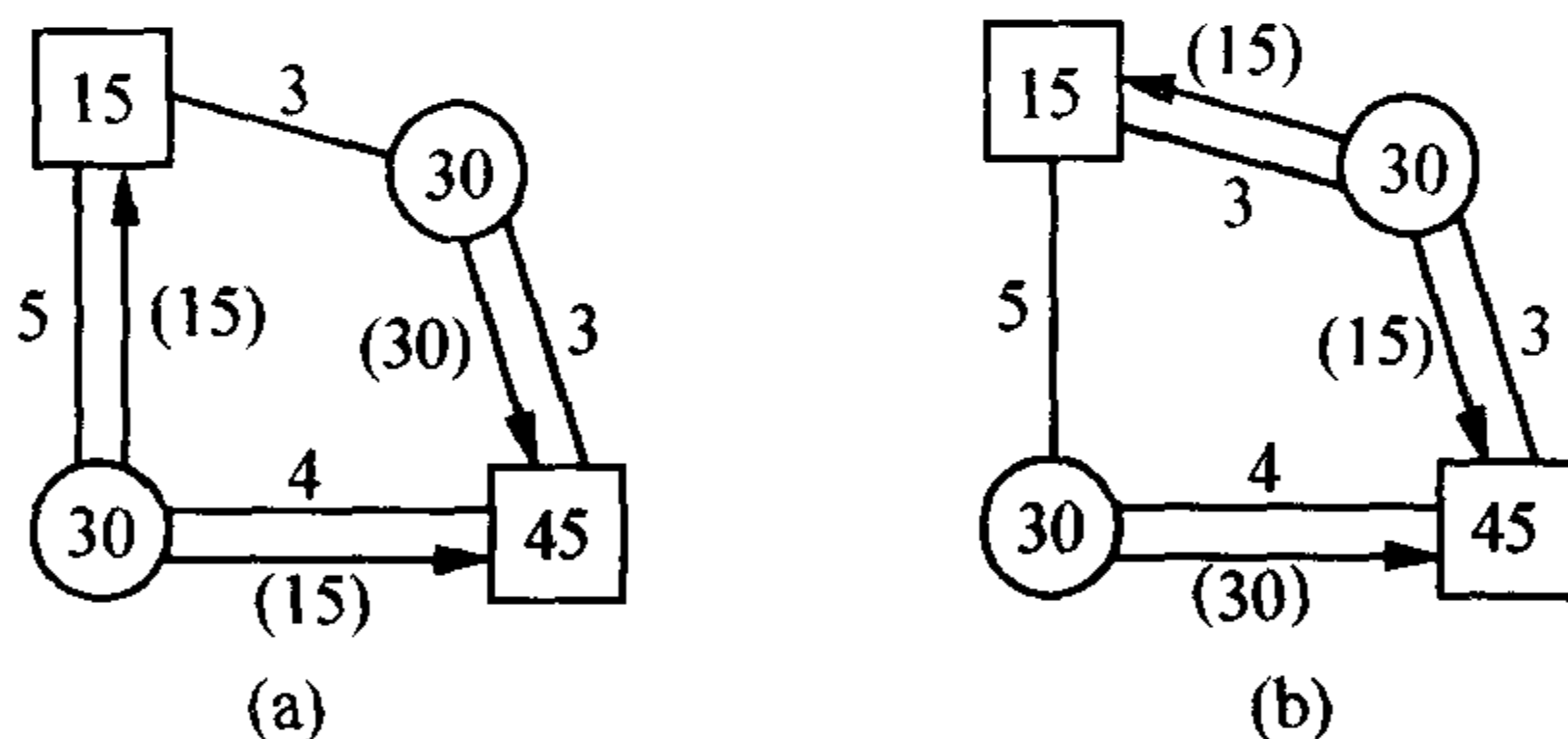


图 52

当交通图成环路时,相反流向的物资可分别在环路的内侧(图 51)和外侧标记流向(图 52)。在内侧的流向称内侧流向,在外侧的流向称外侧流向。若内(或外)侧流向的总长度大于整个环路长的一半时,就称为“迂回”运输,如图 51(a)中内侧物流路长“6”大于环路总长“10”的一半;图 52(a)中内侧物流路长“3 + 5 = 8”大于环路总长“15”的一半,属于“迂回”运输,是不合理的,应分别调整为图 51(b)和图 52(b)。调整后

的流向图可分别节约：

图 51, $8 \times 6 - 8 \times 4 = 16$ (吨·千米)

图 52, $(15 \times 5 + 15 \times 4 + 30 \times 3) - (15 \times 3 + 15 \times 3 + 30 \times 4) = 15$ (吨·千米)

一个物资调运方案中,如果物资流向图中没有“对流”和“迂回”,就是最优方案,就是运费最小的方案。因此,图上作业法是通过调整运输路线,先找出没有“对流”的方案,然后找出没有“迂回”的方案,从而得到最优方案。

求解运输问题,还有一种原始方法叫“表上作业法”,也叫“阶梯方法”,它是将解一般线性规划问题的单纯形法用到运输问题上,并结合运输特点而设计的。表上作业法由丹齐格于 1951 年首先提出,后又经过不少人的研究和改进,至今仍是解运输问题的主要方法之一。它是通过在表上调整,求得最优解。先确定一个初始方案,按最小元素法逐步调整表上数据,直到最优解。例如,要从 A_1, A_2 两地调运商品到 B_1, B_2, B_3 三地销售。它的物资平衡表及单位运价如表 1、表 2:

表 1 物资平衡表 (吨)

销地 产地	B_1	B_2	B_3	发货量
A_1	x_{11}	x_{12}	x_{13}	110
A_2	x_{21}	x_{22}	x_{23}	130
销 量	80	60	100	

表 2 运价表 (万元)

销地 产地	B_1	B_2	B_3
A_1	2	3	4
A_2	3	4	6

注: x_{ij} 表示从 A_i 调运到 B_j 的货量

先从表 2 中找出运价最小者“2”,它是 A_1 运到 B_1 的单位运价。由于这个运价最小, A_1 的物资应先满足 B_1 , A_1 的产量为 110 吨,而 B_1 的需求量为 80 吨,满足 B_1 后 A_1 还剩 30 吨,因此,在平衡表上 A_1 行与 B_1 列交点处填上 80,货量 110 改为

30(表 3),并在运价表中划去 B_1 列(表 4)。再从表 4 中找出运价最小元素“3”,即 A_1 运到 B_2 的运价为 3。因 B_2 的需求量是 60 吨,因此先将 A_1 剩下的 30 吨运给 B_2 ,再从 A_2 调运 30 吨给 B_2 , A_2 还剩 100 吨,此时,平衡表为表 5。运价表中划去 B_2 所在列(表 6)。最后,只需考虑 B_3 的需求,已没有价格可选择。将 A_2 剩下的 100 吨运给 B_3 ,平衡表为表 7。此时总运费 $S = 80 \times 2 + 30 \times 3 + 30 \times 4 + 100 \times 6 = 970$ (万元)

表 3 (吨)

销地 \ 产地	B_1	B_2	B_3	货量
A_1	80			30
A_2				130
销 量	80	60	100	

表 4 (万元)

销地 \ 产地	B_1	B_2	B_3
A_1	2	3	4
A_2	3	4	6

表 5 (吨)

销地 \ 产地	B_1	B_2	B_3	货量
A_1	80	30		
A_2		30		100
销 量	80	60	100	

表 6 (万元)

销地 \ 产地	B_1	B_2	B_3
A_1	2	3	4
A_2	3	4	6

表 7 (吨)

销地 \ 产地	B_1	B_2	B_3	货量
A_1	80	30		110
A_2		30	100	130
销 量	80	60	100	

表 8 (吨)

销地 \ 产地	B_1	B_2	B_3	货量
A_1	80	25	5	110
A_2		35	95	130
销 量	80	60	100	

由于运价与吨·千米数成正比,因此产地到销售地的最小运价意味着相应的距离是最短的。所以,最小元素法实质上反映了“就近供应”的原则。由于最后一次调配时,价格不能选择,该分配方案是否最优未定,若按表 8 的方案调运,它的总运费 S 为:

$$S = 80 \times 2 + 25 \times 3 + 5 \times 4 + 35 \times 4 + 95 \times 6 = 965 \text{ (万元)}$$

比前一个方案少 5 万元。

为了进一步确定所得的方案是否最优,还需要作进一步的检验,如果不是最优方案,就要作进一步调整。检验和调整的方法如下:第一步,首先画出每个空格的闭回路。从一个空格出发,沿水平或垂直方向前进,遇到一个有数字的格子时,就按与前进方向垂直的方向转向前进,前进时又会遇到一个有数字的格子,再按上述方法转向,经若干次后,必回到原来出发的空格,形成一条由线段组成的闭回路。每一空格子都可以作出一条闭回路。表 7 中,仅在 x_{13} 及 x_{21} 处有两个空格子,它们的闭回路分别是: $x_{13} \rightarrow x_{23} \rightarrow x_{22} \rightarrow x_{12} \rightarrow x_{13}$ 和 $x_{21} \rightarrow x_{22} \rightarrow x_{12} \rightarrow x_{11} \rightarrow x_{21}$,如表 9。

表 9 (吨)

产地 \ 销地	B_1	B_2	B_3	产量
A_1	80 2 x_{11} 	30 3 x_{12} 	 4 x_{13} 	110
A_2	 3 x_{21} 	30 4 x_{22} 	100 6 x_{23} 	130
销 量	80	60	100	

第二步,计算每一空格的检验数。在第 i 行第 j 列的空格 x_{ij} 的检验数 λ_{ij} 的计算公式为: λ_{ij} = 闭回路中第奇数次拐角点运

价总和减去闭回路中第偶数次拐角点运价的总和(为了计算方便,将表 2 中的运价记在表 9 格子的右上角,并加以边框)。按此公式计算, $\lambda_{13} = (6 + 3) - (4 + 4) = 1$, $\lambda_{21} = (2 + 4) - (3 + 3) = 0$ 。当空格的检验数 $\lambda_{ij} > 0$ 时,说明该方案非最优解,需要调整。此例中 $\lambda_{13} > 0$,故此方案非最优解,需要调整。调整的方法是:取第一个出现正检验数的空格闭回路上的第奇数次拐角点中最小运量作为调整数,将这条回路上所有奇数拐角点的运量都减去这个调整数,而这条回路上的偶数拐角点的运量都加上这个调整数,如例中的 $\lambda_{13} > 0$,调整数为 30,调整后的平衡表如表 10。

表 10 (吨)

产地 \ 销地	B ₁	B ₂	B ₃	产量
A ₁	80 2	0 3	30 4	
A ₂	3	60 4	70 6	
销 量	80	60	100	

注意:调整后所得的运量为零时,应标上数“0”,新方案的检验数为: $\lambda_{12} = (4 + 4) - (3 + 6) = -1 < 0$, λ_{21} 仍为零,两个检验数都不是正数,已是最优方案,即 $x_{11} = 80, x_{12} = 0, x_{13} = 30, x_{21} = 0, x_{22} = 60, x_{23} = 70$ 。总运费为:

$$S = 80 \times 2 + 30 \times 4 + 60 \times 4 + 70 \times 6 = 940 \text{(万元)}$$

新方案比原方案节约 30 万元。原方案只是一个近似最优解。

表上作业法是解运输问题的一种方法,是将解线性规划的单纯形法结合运输特点而设计的。由于计算机的应用,这个方法可用来解成百上千个产地和销地的运输问题。

0—1 规划

一个旅行者外出远游,必须携带必要的日常用品,但由于负重能力的限制只能携带少量几种日用品,选带那些日用品才能使使用价值最大?这个问题中的决策对象是日常用品。为了优化牲畜品种,要在原有不同品种的牲畜中选择两个品种进行交配,在新品种符合质量指标的前提下选择哪两个品种交配的费用最小?这个问题的决策对象是牲畜。某班接到一项任务,该任务只允许数人参加,如何选派才能最快完成任务?这个问题中的决策对象是人。一项综合工程需建在不同的几个地方,在可供选择的地点中如何选址才能使综合效益最佳?这个问题中的决策对象是地点。类似的问题还可举出很多,这类问题都是规划问题,它们的决策对象都面临两种状态:选中与选不中、派与不派、设与不设、要与不要,二选一。这类问题中的不同决策对象用变量 x_i 表示时,变量 x_i 只能取两个值,为方便计算,习惯上以 0 和 1 两个数值表示,选中的对象取值为 1,不被选中者取值为 0,即
$$x_i = \begin{cases} 1(\text{选中}) \\ 0(\text{选不中}) \end{cases}$$

许多实际问题都具有这种特点。

例 1 背包问题。

一个旅行者,为了准备必要的物品,要在背包里装一些最有用的东西。他共有 m 件物品,按物品有用的程度定出物品的“价值”,第 i 件的价值记为 c_i ,第 i 件物品的重量为 a_i ,而他最多只能带 b 千克的物品。那么他应选带哪些物品,使总重量不超过 b 且使用价值最大?

由于物品只能整件带,变量 x_i 表示第 i 件物品“带”或“不带”,目标函数为 $c_1x_1 + c_2x_2 + \cdots + c_mx_m$,规定:

$$x_i = \begin{cases} 1, & \text{当带第 } i \text{ 件物品时} \\ 0, & \text{当不带第 } i \text{ 件物品时} \end{cases}$$

问题要求目标函数最大,即 $\max \sum_{i=1}^m c_ix_i$,约束条件为:

$$\begin{cases} \sum_{i=1}^m a_ix_i \leq b \\ x_i = 0 \text{ 或 } 1 \end{cases}$$

例 2 指派问题。

车间里有 4 位师傅 A_i ($i = 1, 2, 3, 4$), 他们都能独立完成四项不同的任务 B_j ($j = 1, 2, 3, 4$), 由于他们各有特长, 完成四项任务所需的时间不同, 右边的表中列出了他们完成每项工作所需

时 间 师 傅	任 务 (时)				
		B_1	B_2	B_3	B_4
A_1		3	14	10	5
A_2		10	4	12	10
A_3		9	14	15	13
A_4		7	8	11	9

的时间。规定每项工作安排一人, 且仅一人。问怎样安排才能使四项任务完成的总时间最少?

设变量 x_{ij} 表示安排 A_i 去做 B_j 项工作, 并规定:

$$x_{ij} = \begin{cases} 1 & (\text{安排 } A_i \text{ 做 } B_j \text{ 工作}) \\ 0 & (\text{不安排 } A_i \text{ 做 } B_j \text{ 工作}) \end{cases}$$

每个人都有一项工作: $\sum_{j=1}^4 x_{ij} = 1, (i = 1, 2, 3, 4)$

每项工作都有一人做: $\sum_{i=1}^4 x_{ij} = 1, (j = 1, 2, 3, 4)$

目标是花费的总时间最少, 即

$$\begin{aligned} \min T = & 3x_{11} + 14x_{12} + 10x_{13} + 5x_{14} + 10x_{21} + 4x_{22} \\ & + 12x_{23} + 10x_{24} + 9x_{31} + 14x_{32} + 15x_{33} + 13x_{34} \end{aligned}$$

$$+ 7x_{41} + 8x_{42} + 11x_{43} + 9x_{44}$$

$$\text{约束条件} \begin{cases} \sum_{j=1}^4 x_{ij} = 1 \\ \sum_{i=1}^4 x_{ij} = 1 (i, j = 1, 2, 3, 4) \\ x_{ij} = 0 \text{ 或 } 1 \end{cases}$$

例 3 选址问题。

某公司准备在市东、西、南三个区建立门市部,有 7 个点 $A_i (i = 1, 2, \dots, 7)$ 可供选择。要求:

(1)在东区, A_1, A_2, A_3 三个点中至多选两个;

(2)在西区, A_4, A_5 两个点中至少选一个;

(3)在南区, A_6, A_7 两个点中至少选一个。

如果选用 A_i 点,设备投资费用为 b_i 元,每年可获利润 c_i 元,而投资总额不得超过 B 元。问如何选点可使年利润为最大?

若以变量 x_i 表示 A_i 是否被选用的量,规定:

$$x_i = \begin{cases} 1 (\text{当 } A_i \text{ 被选用}) \\ 0 (\text{当 } A_i \text{ 不被选用}) \end{cases}$$

则该问题的数学规划为:

$$\begin{aligned} \max \quad & \sum_{i=1}^7 c_i x_i \\ \text{约束条件} \quad & \begin{cases} x_1 + x_2 + x_3 \leq 2 \\ x_4 + x_5 \geq 1 \\ x_6 + x_7 \geq 1 \\ x_i = 0 \text{ 或 } 1 (i = 1, 2, \dots, 7) \\ \sum_{i=1}^7 b_i x_i \leq B \end{cases} \end{aligned}$$

例 4 选项问题。

某公司有机会对三个项目进行投资,根据该公司的预算,

前两年每年能投资 6 万元,后两年每年能投资 7 万元,而这两个项目每年所需的投资额及利润如表。怎样选择投资项目可获得最大利润?

投 资 年 度	项目所需投资额(万元)			投资额 (万元)
	B_1	B_2	B_3	
第一年	0	4	2	6
第二年	5	1	2	6
第三年	4	2	3	7
第四年	5	2	3	7
利 润	17	10	16	

该公司对每个项目 $B_i (i=1,2,3)$ 仅有两种选择:投与不投,可用变量 x_i 表示,并规定:

$$x_i = \begin{cases} 1(\text{对 } B_i \text{ 投资}) \\ 0(\text{对 } B_i \text{ 不投资}) \end{cases}$$

由于每年投资额的限制,故有:

$$\text{约束条件} \begin{cases} 4x_2 + 2x_3 \leq 6 \\ 5x_1 + x_2 + 2x_3 \leq 6 \\ 4x_1 + 2x_2 + 3x_3 \leq 7 \\ 5x_1 + 2x_2 + 3x_3 \leq 7 \\ x_i = 0 \text{ 或 } 1, (i=1,2,3) \end{cases}$$

目标函数是利润 $L = 17x_1 + 10x_2 + 16x_3$, 求最大值。

上述 4 例都是变量只能取 0 或 1 的线性规划,这种线性规划称为 0—1 规划。0—1 规划是一种特殊的线性规划,它反映的实际问题涉及面很广。有 m 个变量的 0—1 规划,它的解至多只有 2^m 个。如果变量数目不多,可列举出所有可能解,再检验它是否符合约束条件,然后从中选出最优解,这种方法称完全枚举法。如例 4 中共有 3 个变量, $m=3$, $x_i (i=1,2,3)$ 取 0 或 1,数组 (x_1, x_2, x_3) 有 2^3 个方案,分别为 $(0,0,0), (1,0,0), (0,1,0), (0,1,1), (1,1,0), (0,0,1),$

$(1,0,1), (1,1,1)$ 。先检查这些数组是否符合约束条件。其中满足约束条件的有：

$$\begin{array}{ll} (0,0,0) & L=0 \\ (1,0,0) & L=17 \\ (0,1,0) & L=10 \\ (0,0,1) & L=16 \\ (0,1,1) & L=26 \\ (1,1,0) & L=27 \end{array}$$

比较这些解的优劣,使 L 最大的解是 $(1,1,0)$,即最优解为 $x_1=1, x_2=1, x_3=0$,此时 $L=27$ 。完全枚举法虽简单,但当变量较多时,计算十分麻烦,必须寻找适合多变量0—1规划的解法。

有一种专门用来解 0—1 规划的方法,称为匈牙利法,它是比较方便的。匈牙利法是通过一系列简约矩阵求得最优解,故也称为简约矩阵法。它是在匈牙利数学家库尼格证明了有关理论的基础上建立起来的。它通过对分配问题所给出的 $n \times n$ 阶矩阵进行一系列简化,使矩阵的每一行每一列都各含有一个 0 元素,然后根据这 n 个 0 元素(不惟一)即可得最优解。以例 2 为例,题中的矩阵为

$$\begin{pmatrix} 3 & 14 & 10 & 5 \\ 10 & 4 & 12 & 10 \\ 9 & 14 & 15 & 13 \\ 7 & 8 & 11 & 9 \end{pmatrix}$$

如果矩阵中第一行减去它的最小元素 3,该矩阵变换成:

$$\begin{pmatrix} 3-3 & 14-3 & 10-3 & 5-3 \\ 10 & 4 & 12 & 10 \\ 9 & 14 & 15 & 13 \\ 7 & 8 & 11 & 9 \end{pmatrix}$$

此时目标函数 T 变成 T' , 变换后的目标函数为:

$$T' = T - 3(x_{11} + x_{12} + x_{13} + x_{14}) = T - 3(\because x_{11} + x_{12} + x_{13} + x_{14} = 1)$$

显然, T' 与原目标函数 T 仅差一个常数, T 与 T' 能同时取最小(或最大)值, 即 T' 与 T 同解。因此, 可以对矩阵的每行(或列)作加减变换, 直到每一行每一列都含有 0 元素为止, 都不会改变问题的最优解, 只改变目标函数的值。例如,

$$\begin{matrix} \begin{pmatrix} 3 & 14 & 10 & 5 \\ 10 & 4 & 12 & 10 \\ 9 & 14 & 15 & 13 \\ 7 & 8 & 11 & 9 \end{pmatrix} & \xrightarrow{\text{每行减去本行的最小数}} & \begin{pmatrix} 0 & 11 & 7 & 2 \\ 6 & 0 & 8 & 6 \\ 0 & 5 & 6 & 4 \\ 0 & 1 & 4 & 2 \end{pmatrix} \\ & & \begin{matrix} \text{第三列} \\ \text{减去 4,} \\ \text{第四列} \\ \text{减去 2} \end{matrix} \xrightarrow{\hspace{1cm}} \begin{pmatrix} 0 & 11 & 3 & 0 \\ 6 & 0 & 4 & 4 \\ 0 & 5 & 2 & 2 \\ 0 & 1 & 0 & 0 \end{pmatrix} \end{matrix}$$

最后这个矩阵的每一行每一列都含有 0 元素, 矩阵经过这些变换后, 只影响目标函数的值, 而不会改变最优解。现在可按这个矩阵进行任务分派, 为了使目标函数达到最小, 可以从 0 元素最少的行开始指派, 每行每列只指派一个 0 元素, 指派到的 0 记为 0^* , 同行同列的其他 0 元素不再指派, 记为 Φ 。由此可见, 指派方式不是惟一的。按这种方法, 前面的矩阵可变换为:

$$\begin{pmatrix} 0 & 11 & 3 & 0 \\ 6 & 0^* & 4 & 4 \\ 0 & 5 & 2 & 2 \\ 0 & 1 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} \Phi & 11 & 3 & 0 \\ 6 & 0^* & 4 & 4 \\ 0^* & 5 & 2 & 2 \\ \Phi & 1 & 0 & 0 \end{pmatrix} \rightarrow$$

$$\begin{pmatrix} \Phi & 11 & 3 & 0 \\ 6 & 0^* & 4 & 4 \\ 0^* & 5 & 2 & 2 \\ \Phi & 1 & 0^* & \Phi \end{pmatrix} \rightarrow \begin{pmatrix} \Phi & 11 & 3 & 0^* \\ 6 & 0^* & 4 & 4 \\ 0^* & 5 & 2 & 2 \\ \Phi & 1 & 0^* & \Phi \end{pmatrix}$$

凡是 0^* 处的 x_{ij} 取 1, 其他元素 x_{ij} 取 0, 即指派结果为:

$$(x_{ij}) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

矩阵中 $x_{14} = 1$, 表示安排 A_1 做 B_4 工作; $x_{22} = 1$, 表示安排 A_2 做 B_2 工作; $x_{31} = 1$, 表示安排 A_3 做 B_1 工作; $x_{43} = 1$, 表示安排 A_4 做 B_3 工作。

在这种安排下, 目标函数(总时间) T 为最小。

$$\begin{aligned} \min T &= 3 \cdot 0 + 14 \cdot 0 + 10 \cdot 0 + 5 \cdot 1 + 10 \cdot 0 + 4 \cdot 1 + 12 \cdot 0 + 10 \cdot 0 \\ &\quad + 9 \cdot 1 + 14 \cdot 0 + 15 \cdot 0 + 13 \cdot 0 + 7 \cdot 0 + 8 \cdot 0 + 11 \cdot 1 + 9 \cdot 0 \\ &= 5 + 4 + 9 + 11 \\ &= 29(\text{时}) \end{aligned}$$

匈牙利算法比较方便, 不需要特殊工具, 仅靠纸与笔就可进行。当变量很多时, 可通过计算机运算。现实世界中还存在着许多具有组合特征的、涉及“是”或“非”决策的最优化问题, 这类问题也可归结为 0—1 规划。如果线性规划中的变量仅限于取整数时, 这样的线性规划称为整数规划。显然, 0—1 规划是整数规划的特殊情况。因此, 整数规划的求解方法都可用于 0—1 规划。

其他问题

20 世纪 30 年代末,自康托罗维奇开始研究线性规划以来,线性规划在方法上和理论上得到迅速发展。20 世纪 30 年代后期,单纯形法的提出使线性规划的求解得到有效的解决。特殊类型的线性规划常有特殊的解法,如解 0—1 规划的匈牙利法。有些线性规划问题的全部变量都必须取整数值(称为全整数规划),有些线性规划则只要求一部分变量取整数值(称为混合型线性规划)。在现实世界中,整数规划(线性或非线性)有着广泛的应用,如行囊问题、下料问题、工序安排问题等都可归结为整数规划。在约束条件及变量数目都差不多的情况下,整数规划的计算量要比一般线性规划大得多。许多数学家为整数规划提供了有效的解法。1958 年,高沫莱首先提出系统地解整数规划的一种方法,叫割平面法。20 世纪 60 年代,兰德、道尔奇、李特尔等人又提出了分支定界法。这两种方法用来解规模不大的实际问题均能取得较好的效果。割平面法的基础仍然是解一般线性规划,先不考虑变量的整数要求,如果所得最优解不是整数,可通过增加约束条件(称为割平面方程),缩小允许解域,使被切掉的那部分解域不包含任何整数解。然后在缩小了的允许解域中,再按一般线性规划的算法求解,直至找到整数解为止。分支定界法也称“枚举法”,是解整数规划及其他离散最优化问题的一种方法。具体做法是按一般规划问题求解,当最优解不是整数时,可增加约束条件,将原问题分解成几个分支(称“分支”),先求出某分支上的最优解,并将相应的目标函数值作为其他分支取舍

的“界”。如此一次一次地分解(“分支”)、“定界”,逐步逼近最优解。

例如,求解整数规划 $\max \quad 40x_1 + 90x_2$

$$\text{约束条件: } \begin{cases} 9x_1 + 7x_2 \leq 56 \\ 7x_1 + 20x_2 \leq 70 \\ x_1 \geq 0, x_2 \geq 0, x_1, x_2 \text{ 为整数} \end{cases}$$

先按一般线性规划求解,可得最优解为:

$x_1 = 4.809, x_2 = 1.817$ (保留三位小数)目标函数最大值为 355.890,由于此解 x_1, x_2 不是整数,不符合要求。因此,可将解域 S_1 (图 53)分解成两部分。只要从所得的解 $x_1 = 4.809$ 或 $x_2 = 1.817$ 中任选一个,如选 $x_1 = 4.809$,最靠近它的整数为 4 和

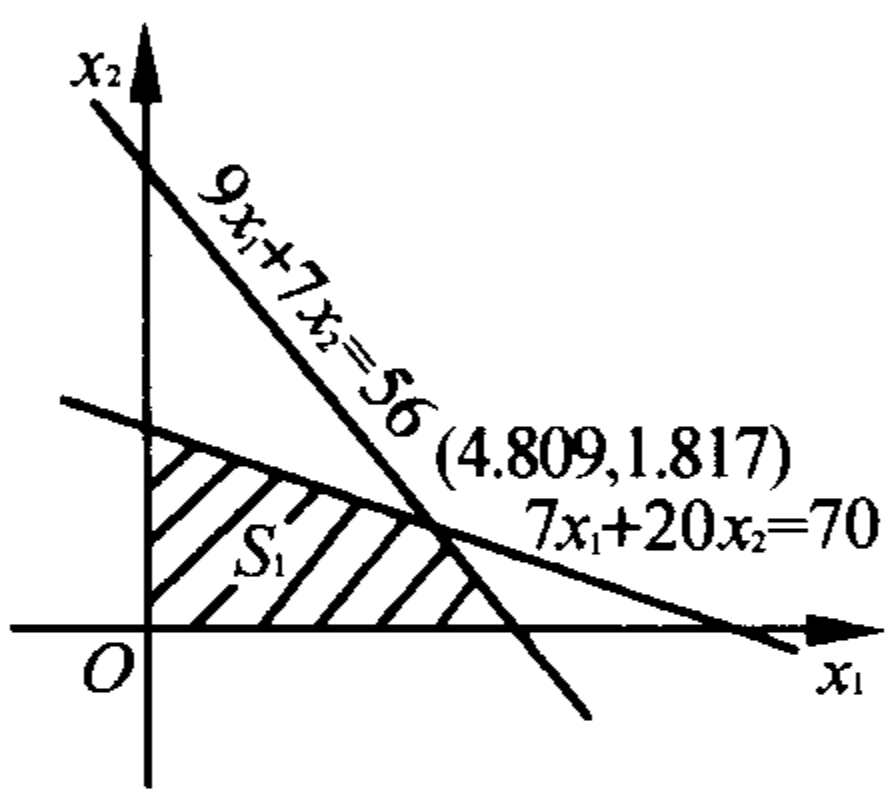


图 53

5,故可添加约束条件: $x_1 \leq 4$ 和 $x_1 \geq 5$ 。将原来的整数规划分成两个新的整数规划:

(I) $\max \quad 40x_1 + 90x_2$ 约束条件为: $9x_1 + 7x_2 \leq 56$ $7x_1 + 20x_2 \leq 70$ $x_1 \leq 4$ $x_1 \geq 0, x_2 \geq 0$ 为整数	(II) $\max \quad 40x_1 + 90x_2$ 约束条件为: $9x_1 + 7x_2 \leq 56$ $7x_1 + 20x_2 \leq 70$ $x_1 \geq 5$ $x_1 \geq 0, x_2 \geq 0$ 为整数
---	--

(I)的允许解域为 S_2 , (II)的允许解域为 S_3 。原问题被分成两支(这就是“分支”),允许解域中已抛弃掉 $4 < x_1 < 5$ (图 54)这部分。可先解其中一个分支(I),将分支(II)暂放

一边。得最优解为：

$$x_1 = 4.000, \quad x_2 = 2.100$$

目标函数最大值为 349.000。 x_1 已是整数，符合要求，但 $x_2 = 2.100$ 仍不符合整数的要求，现再以 x_2 为依据进行分支。因最靠近 2.100 的整数是 2 和 3，于是再将 (I) 分成两个分支，成为两个新的整数规划：

(I - I) max $40x_1 + 90x_2$ 约束条件： $9x_1 + 7x_2 \leq 56$ $7x_1 + 20x_2 \leq 70$ $x_1 \leq 4, \quad x_2 \leq 2$ $x_1 \geq 0, \quad x_2 \geq 0$ 为整数	(I - II) max $40x_1 + 90x_2$ 约束条件： $9x_1 + 7x_2 \leq 56$ $7x_1 + 20x_2 \leq 70$ $x_1 \leq 4, \quad x_2 \geq 3$ $x_1 \geq 0, \quad x_2 \geq 0$ 为整数
---	--

(I - I) 的允许解域为 S_3 , (I - II) 的允许解域为 S_4 。(I) 的允许解域中已抛弃掉 $2 < x_2 < 3$ (图 55) 这部分。解 (I - I) 得最优解为 $x_1 = 4.000, \quad x_2 = 2.000$, 是整数解，目标函数最大值为 340.000。此解 ($x_1 = 4, \quad x_2 = 2$) 不仅是 (I - I) 的最优整数解，自

然也是原问题的整数解，但不一定是原问题的最优整数解，因为尚未对 (II) 及 (I - II) 等分支求解，在这些分支中也许还有更优的整数解。所以，还要用类似方法对 (II) 及 (I - II) 求

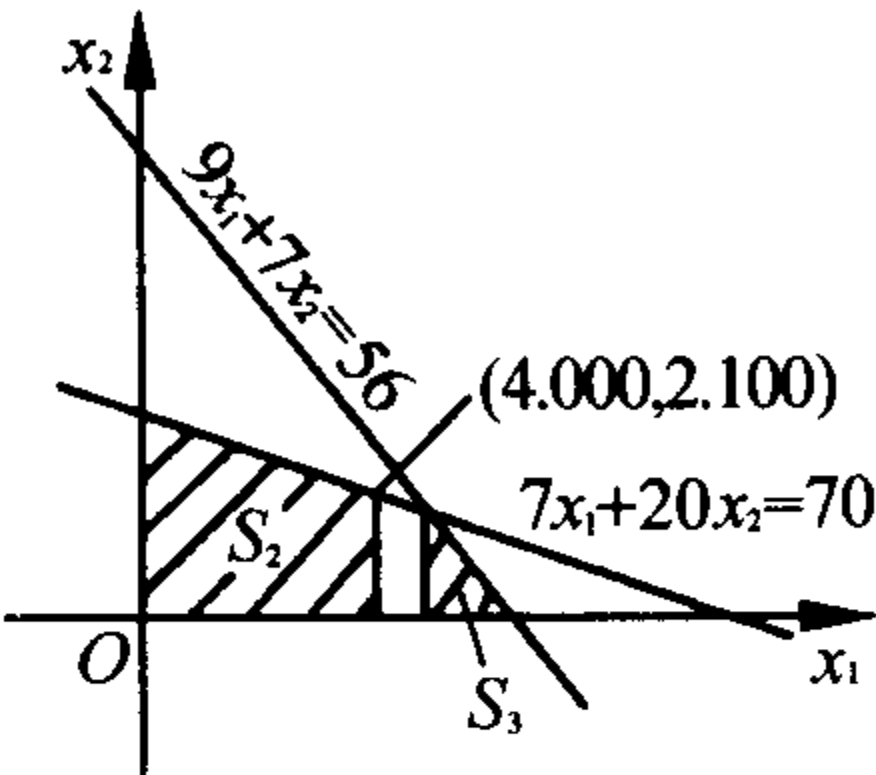


图 54

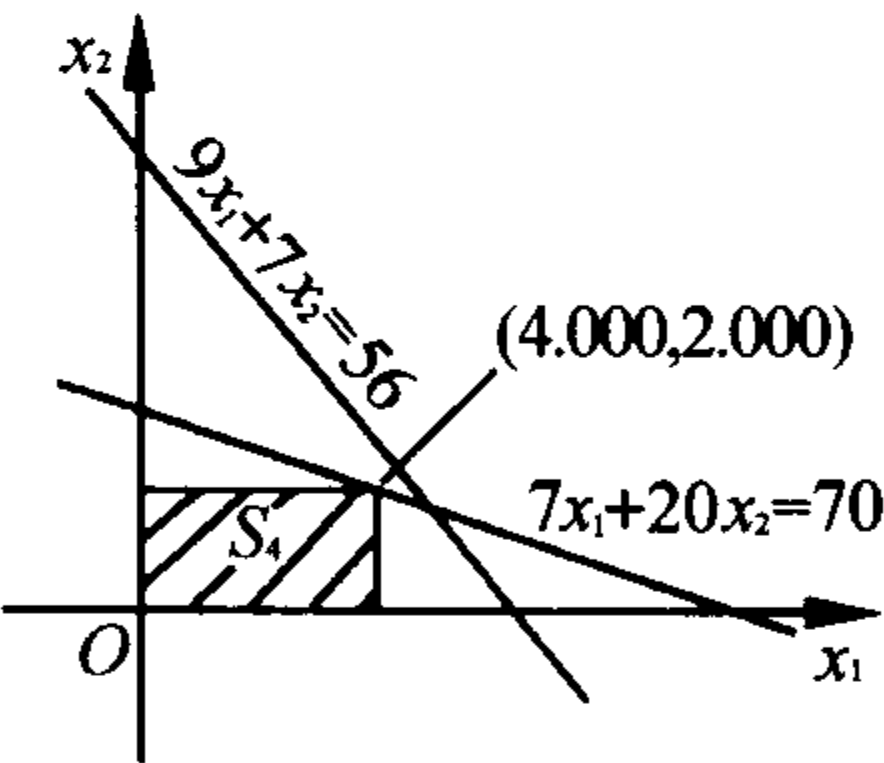


图 55

解。此时,目标函数值 340 可作为(Ⅱ)及(Ⅰ-Ⅱ)求解时目标函数的界(这就是“定界”)。目标函数值优于 340 的留下,作为新的“界”,否则弃之。继续不断“分支”、“定界”,直至原问题取得最优整数解为止。经计算,该问题的最优整数解是 $x_1 = 4$, $x_2 = 2$, 目标函数最大值是 340。

由于不同方法所得的“解”的序列收敛速度(逼近最优解的速度)是不同的,因此有时用不同的方法配合使用,可取得更好效果。整数规划是运筹学中发展最快的分支之一,它的新理论与新解法不断出现,它与其他数学分支(如图论、群论等)的互相渗透也日益显著。

线性规划问题的数学模型是描述实际问题的抽象数学形式,它反映了客观事物数量之间的规律性。建立数学模型时,可粗可细,越细越接近真实,但求解也越复杂,因此,要抓住问题的本质,使数学模型既简单又能比较真实地反映实际情况。有些大型的线性规划问题,除了少数约束条件外,其约束条件和变量可分成若干相互独立的组,各组的约束条件和所含的变量各不相同,构成若干独立的子规划及一个连结这些子规划的主规划。先解主规划,然后利用主规划的解构造每个子规划的目标函数。再将子规划的解加到主规划上。重复这个过程,经有限次运作后,可得到问题的最优解,或断定问题无最优解。

微积分方法



从二次函数的极值谈起

17 世纪,笛卡尔的解析几何学诞生,使数学从繁琐的几何证明和令人费解的代数规则中解脱出来。几何学与代数学的紧密结合,使几何的直观性与代数刻画数量的抽象性融于一体。一方面,几何图形可以用简洁的代数符号表示;另一方面,借助几何的直观性展现抽象的代数语言的意义,让人们感受数学勾画的宇宙壮丽美景。

二次函数的极值问题是大家熟悉的最优化问题。它的目标函数是二次函数

$f(x) = ax^2 + bx + c (a \neq 0)$ 。当 $a > 0$ 时,其最优解为 $x = -\frac{b}{2a}$,目标函数 $f(x)$ 的最小值为 $\frac{4ac - b^2}{4a}$;当 $a < 0$ 时,其最优解 $x = -\frac{b}{2a}$,目标函数 $f(x)$ 的最大值为 $\frac{4ac - b^2}{4a}$ 。目标函数 $f(x) = ax^2 + bx + c$ 的图象是一条抛物线,由于它只有一个顶点,所以它的极(小或大)值就是最(小或大)值。 $f(x)$ 在顶点附近的变化情况如下:

自变量 x 二次项系数	$x < -\frac{b}{2a}$	$x = -\frac{b}{2a}$	$x > -\frac{b}{2a}$
$a > 0$	$f(x)$ 递减 ↘	极小值	$f(x)$ 递增 ↗
$a < 0$	$f(x)$ 递增 ↗	极大值	$f(x)$ 递减 ↘

可见,当自变量增大时,若函数从递减转为递增时,必有极小值;函数从递增转为递减时,必有极大值。如果函数的增减性质不改变,将不可能产生极大(或极小)值。例如:目标函数为一次函数 $f(x) = ax + b (a \neq 0)$,它的图象是一条直线,无极大(小)值可言。如果是线段,则端点的函数值便是极大(或极小)值,即最优解在端点。三次函数 $f(x) = x^3 + 1$ (图 56),它在定义域 $(-\infty, +\infty)$ 上为递增函数,故无极值。函数 $f(x)$

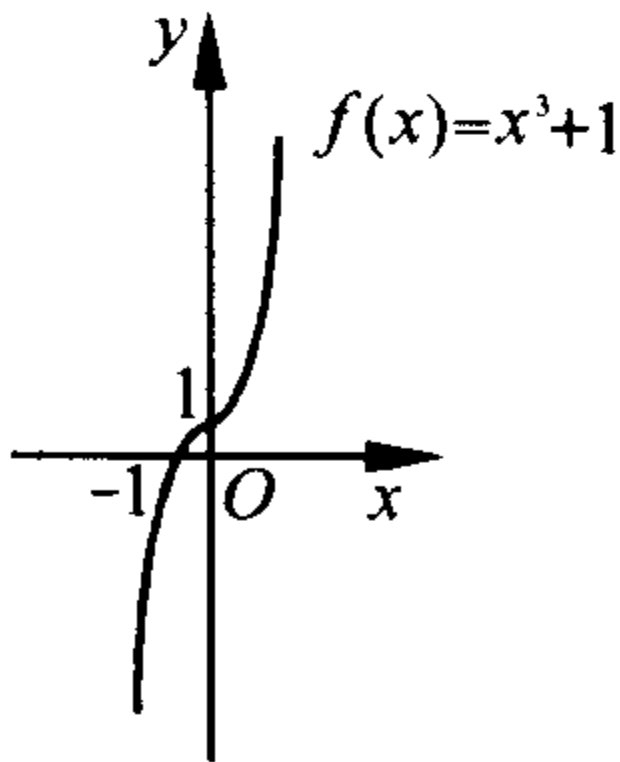


图 56

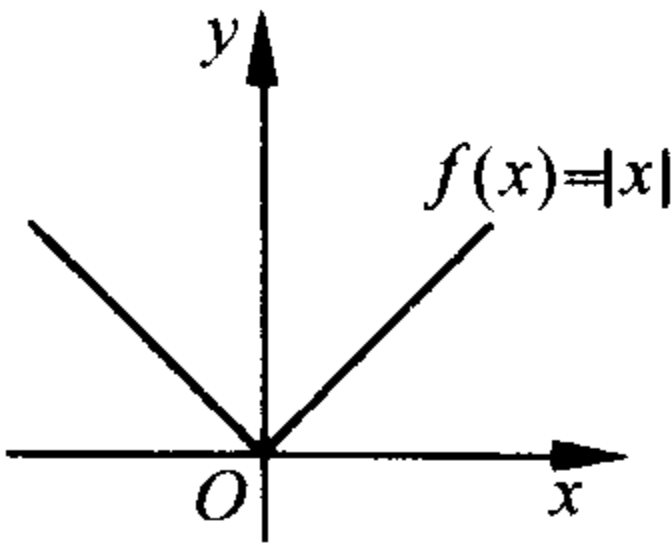


图 57

$$= |x| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases} \quad \text{它的图象}$$

如图 57, 由两条线段组成。当 $x < 0$ 时, $f(x)$ 递减; 当 $x > 0$ 时, $f(x)$ 递增。因此, 最优解为 $x = 0$, 目标函数有极小值 $f(x) = f(0) = 0$ 。而三次函数 $f(x) = x^3 - x^2 - x + 1$ 的图象如图 58。

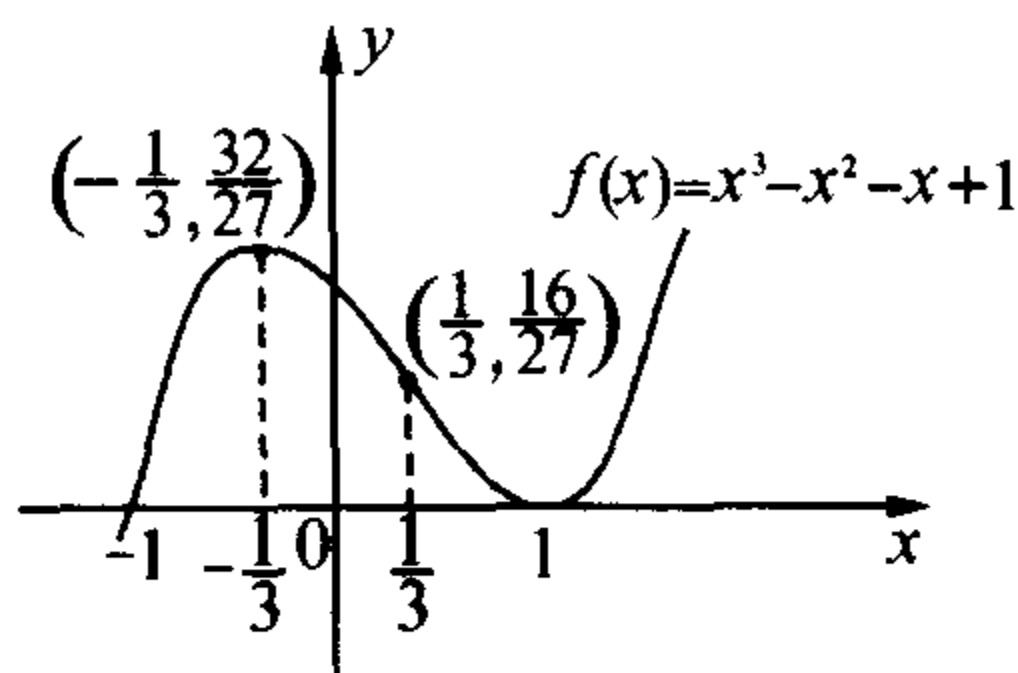


图 58

其函数变化情况如下表:

x	$x < -\frac{1}{3}$	$x = -\frac{1}{3}$	$-\frac{1}{3} < x < 1$	$x = 1$	$x > 1$
$f(x)$	递增 ↗	极大值 $\frac{32}{27}$	递减 ↘	极小值 0	递增 ↗

从表中看出, 它有一个极大值和一个极小值。

由此可见, 三次函数不同于二次函数 $y = ax^2 + bx + c$, 二次函数一定有一个且只有一个极大(或极小)值, 而三次函数可能没有极大(小)值, 也可能有多于一个极大(小)值。从图象上可以看到, 一次函数的图象是一条直线; 二次函数和三次函数的图象都是曲线。三次函数的曲线比二次函数的曲线要柔软。曲线的柔软性与函数是否具有保凸(或保凹)性有关。保凸(或保凹)性是指函数在定义域上不改变凸(或凹)的性质(图 59、图 60)。

所谓凸(凹)函数是指对函数 $f(x)$ 定义域中的任意两个点 $x_1, x_2 (x_1 < x_2)$, 都有 $f\left(\frac{x_1 + x_2}{2}\right) \geq (\leq) \frac{f(x_1) + f(x_2)}{2}$ 成立, 即 $[x_1, x_2]$ 中任何一点的函数值不大于割线 AB 的相应

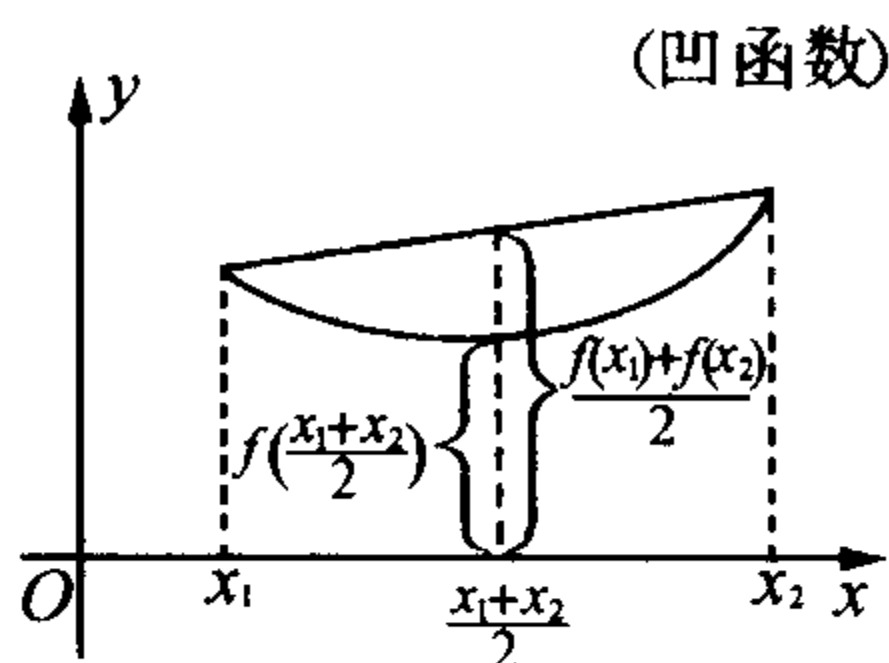


图 59

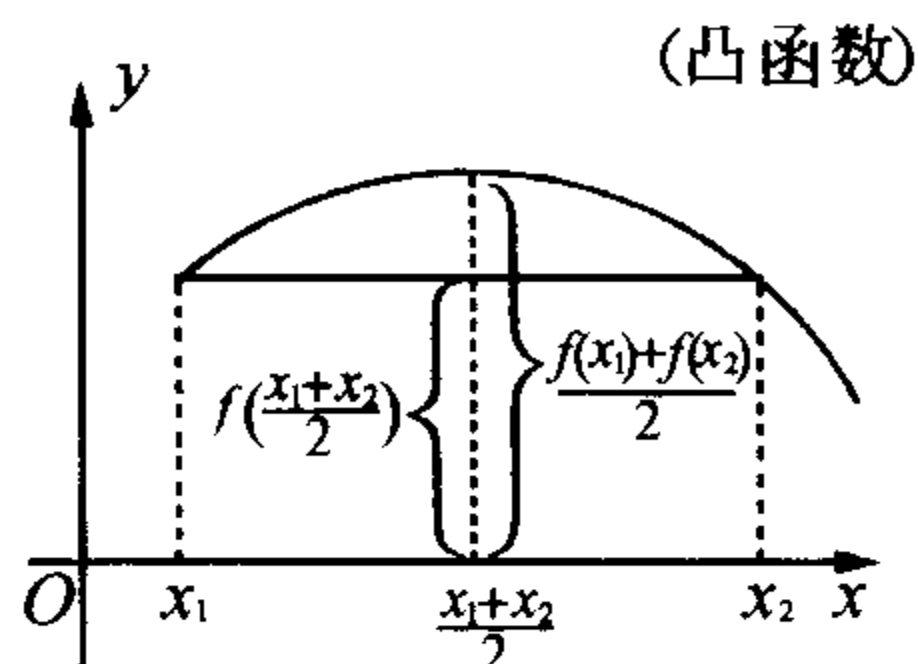


图 60

的纵坐标,其几何意义是弧 AB 在弦 AB 的上方。显然,若 $f(x)$ 是凸函数,则 $-f(x)$ 必是凹函数。凸(或凹)函数仅具有一个顶点。二次函数是凸(或凹)函数,它的图象(抛物线)具有保凸性。因此,它在定义域内有一个且仅有一个顶点,有一个极大(或极小)值。三次函数则不具有保凸性,它的凸(凹)向要改变。如 $f(x) = x^3 + 1$ 在 $x = 0$ 时改变了凸向, $f(x) = x^3 - x^2 - x + 1$ 在 $x = \frac{1}{3}$ 处改变了凸向,因此,三次函数可能没有极值,也可能有,但至多只有两个极值。微分法可以对更广泛一类函数的极值作出判断。

曲线的切线和极值

17 世纪初期,当时数十位世界级数学家曾费尽心机琢磨以下四类科学问题:

(1) 运动物体的瞬时速度和加速度问题。

运动着的物体在每一时刻都应该有速度,可是在瞬时刹那,移动的距离和时间都是 0,而 $\frac{0}{0}$ 是无意义的。怎么计算瞬时速度?

(2) 曲线的切线问题。

光线通过透镜的角度是光线与镜面曲线的法线的夹角。法线垂直于切线,如何求切线或法线?

(3) 函数的最大值和最小值问题。

炮弹的射程依赖于炮筒对地面的发射角,获得最大射程的发射角是多少? 炮弹获得最大高度的发射角又是多少? 行星离太阳的最远、最近距离怎样计算等等。

(4) 曲线的长度及所围面积问题。

如何计算在给定时间内行星移动的距离? 如何计算一般曲线所围区域的面积? 如何求行星作用于另一物体的引力等问题。

这四类问题在同一历史时期出现,正是社会生产技术的发展对数学提出的要求。在这四类问题中,切线问题是最基本的,它与其他问题的关系最密切。当时人们对切线的意义还不清楚,古希腊人把切线定义为“和曲线只有一个接触点,且位于曲线一边的直线”,这种定义对较复杂的曲线并不适用。后来有人将几何与力学联系起来定义切线,认为炮弹的水平速度 PQ 和垂直速度 PR 的合速度的方向(平行四边形对角线)就是 P 点的切线(如图 61),并用这种切线概念解决了不少古希腊人的定义不能解决的问题。然而,令人遗憾的是它毕竟是物理学上的概念,不是数学领域的概念,数学家无法“容忍”。1637 年,费马在他的《求最大值和最小值的方法》手稿中,提出求曲线 $f(x)$ 在 P 点的切线(图 62)的方法:因为 $\triangle TQP \sim \triangle PRT_1$, 有 $TQ = \frac{e \cdot f'(x)}{f(x+e) - f(x)}$, 用 e 除以分子分母,然后令 $e = 0$, 即得 TQ 的值,从而确定 T 的位置, PT 即为切线。这种涉及极限的思想与现在的微分学中切线的概

念相一致,不过当时表达得很不清楚。此外,笛卡尔等人曾从不同角度相继给出了不同的求切线的方法。

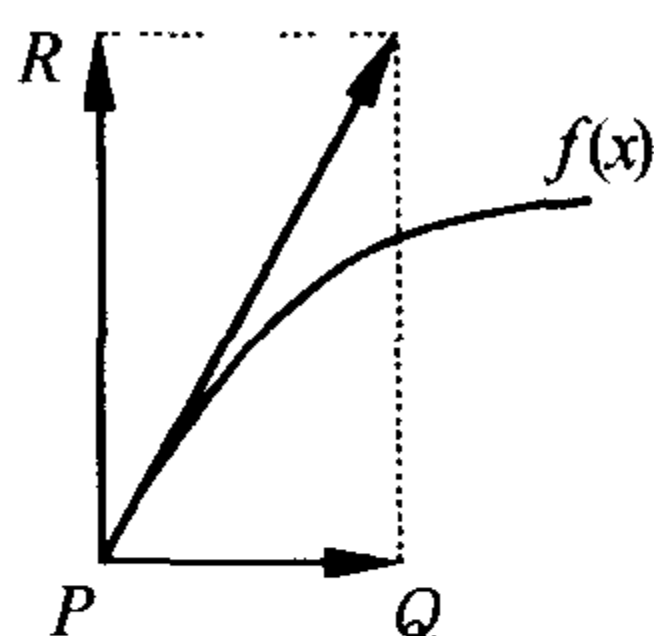


图 61

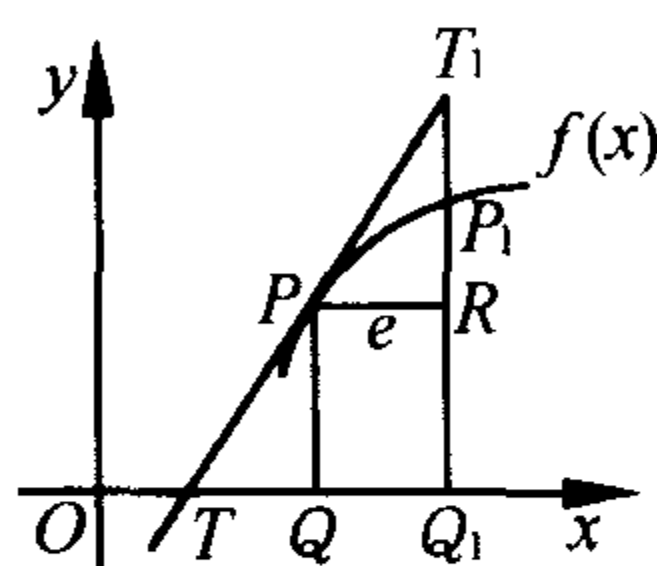


图 62

上述四类科学问题的研究和解决,最终导致微积分的创立。在众多的数学家中,两位数学大师牛顿和莱布尼兹达到顶峰,成为微积分学的创始人。微分学中的第一个基本概念是与切线相关的导数概念。函数 $f(x)$ 在 $x = x_0$ 的导数是指

$x \rightarrow x_0$ 时, $\frac{f(x) - f(x_0)}{x - x_0}$ 的极限存在,并记为 $f'(x_0)$,即:

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0)$$

由图 63 可见,比值 $\frac{f(x) - f(x_0)}{x - x_0}$ 正

是割线 P_0P 的斜率,当 $x \rightarrow x_0$ 时, P 点沿曲线移动到 P_0 ,此时割线 P_0P 变成 P_0 处的切线 P_0T ,而

$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$ 便是 P_0 处的切

线的斜率 K ,即 $K = f'(x_0) = \tan \alpha$,

因而 P_0 点的切线方程为 $y - f(x_0)$

$= f'(x_0)(x - x_0)$ 。微分学中求切线的方法与费马求切线的方法基本一致,只是费马没有直接利用极限概念而且显得含

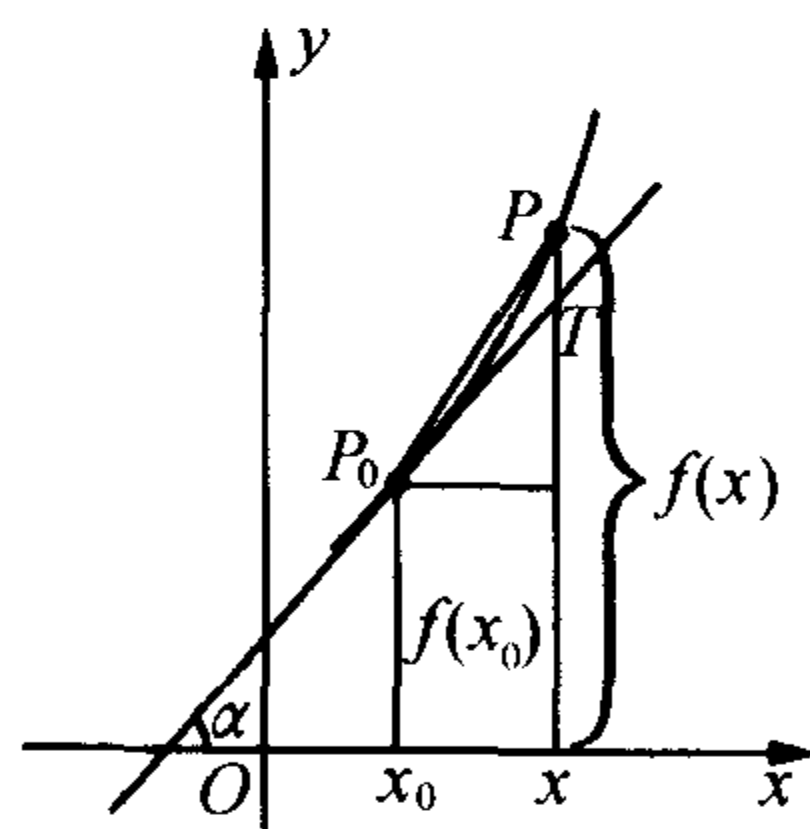


图 63

糊。从物理学上,运动物体在 $t = t_0$ 时的瞬时速度 $v(t_0)$ 正是路程函数 $s(t)$ 在 t_0 时刻的导数:

$$v(t_0) = \lim_{t \rightarrow t_0} \frac{s(t) - s(t_0)}{t - t_0} = s'(t_0)$$

其中比值 $\frac{s(t) - s(t_0)}{t - t_0}$ 表示在 t 与 t_0 时间内物体的平均速度,瞬时速度是当 $t \rightarrow t_0$ 时平均速度的极限值,类似可得到加速度 $a(t_0) = v'(t_0)$ 。由此可见,微分学中的导数概念将四类科学问题中的前两类问题的关系刻画得入木三分,后两类问题也是微积分学创立的主要原动力,它们之间有着一脉相承的关系。求函数的最大值和最小值问题是从开普勒观测酒桶的形状开始的。他发现酒商求酒桶体积的方法不精确,在他的《测量酒桶体积的新科学》(1615年)中,他证明了在所有内接于球面的、具有正方形底的正平行六面体中,立方体的容积最大。他还注意到,当越来越接近最大体积时,对应尺寸的变化所产生的体积变化越来越小。费马在他的《求最大值和最小值的方法》中,也描述了函数在极大值(或极小值)附近时的这一特性。例如,要在已知长为 a 的

直线段上求一点,使被这点分成的两部分
线段组成的矩形最大(图 64)。他不仅得

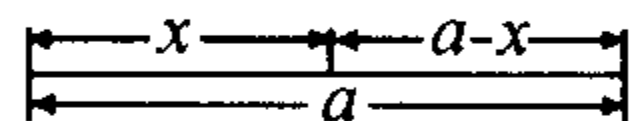


图 64

到所求的分点应在 $\frac{a}{2}$ 处(即正方形面积最大),而且发现当 x 接近 $\frac{a}{2}$ 时,矩形的面积 $x(a - x)$ 变化缓慢。他认为这一规律具有普遍意义。对于一个函数 $f(x)$,如果 $f(x)$ 在 x_0 处取得极大(小)值时,它在 x_0 附近的函数值变化是很小的,当 x 有同一增量时,函数在 x_0 处附近的改变量 $f(x) - f(x_0)$ 要比其他地方小得多。因此,函数曲线的切线在极值处越来越

接近水平状态,当处在极值点 $x = x_0$ 时,切线呈水平状态,即切线的斜率 $K = 0$,故有:

$$f'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = 0$$

正是切线斜率的变化刻画了函数值变化,才使微分法求极值成为顺理成章的。

对函数定义域中任意的点 x ,当 $f'(x) > 0$ 时,切线与 x 轴的夹角 α 小于 $\frac{\pi}{2}$,函数曲线呈上升势态,即 $f(x)$ 单调增加;当 $f'(x) < 0$ 时,切线与 x 轴的夹角 α 大于 $\frac{\pi}{2}$,函数曲线呈下降势态,即 $f(x)$ 单调递减,所以,切线斜率 $K = f'(x)$ 在 x_0 两侧取不同符号时,函数在 x_0 处必能取得极值(极大值或极小值)。判别 $f(x_0)$ 是极大值或极小值的方法如下:

x	$x < x_0$	$x = x_0$	$x > x_0$
$f'(x)$	> 0	$= 0$	< 0
$f(x)$	递增 ↗	极大值 $f(x_0)$	递减 ↘

x	$x < x_0$	$x = x_0$	$x > x_0$
$f'(x)$	< 0	$= 0$	> 0
$f(x)$	递减 ↘	极小值 $f(x_0)$	递增 ↗

例如,要求函数 $f(x) = x^3 - 12x$ 的极值,可先求 $f'(x) = 3x^2 - 12 = 3(x - 2)(x + 2)$,当 $x = \pm 2$ 时, $f'(\pm 2) = 0$,当 $x \in (-\infty, -2)$, $f'(x) > 0$, $f(x)$ 必单调增加;当 $x \in (-2, 2)$, $f'(x) < 0$, $f(x)$ 单调减小;当 $x \in (2, +\infty)$, $f'(x) > 0$, $f(x)$ 单调增加(图 65)。

$f(x) = x^3 - 12x$ 的变化情况如下:

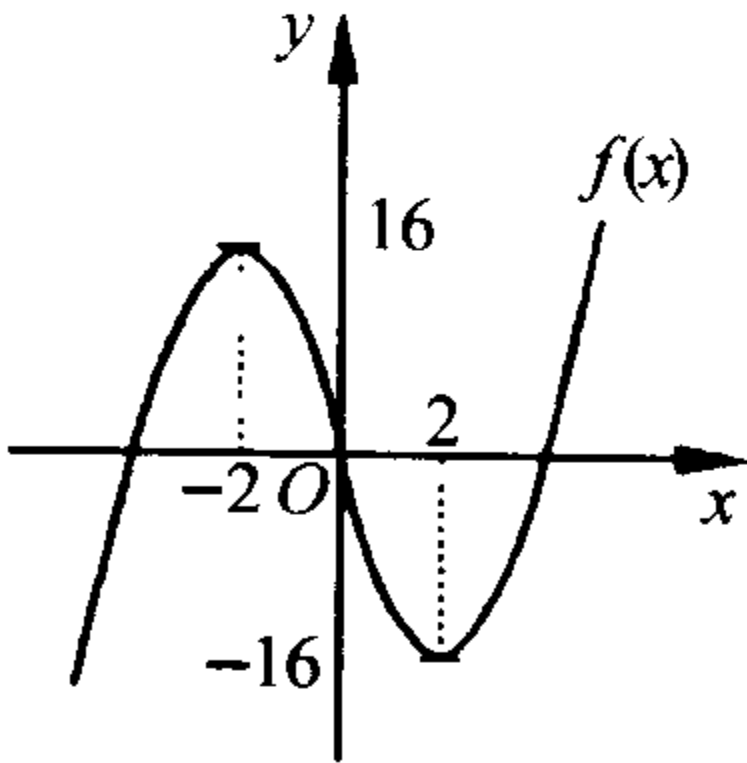


图 65

x	$(-\infty, -2)$	-2	$(-2, +2)$	2	$(2, +\infty)$
$f'(x)$	>0	0	<0	0	>0
$f(x)$	递增 ↗	极大值 $f(-2)=16$ 有水平切线	递减 ↘	极小值 $f(2)=-16$ 有水平切线	递增 ↗

函数 $f(x)$ 的极大值为 $f(-2)=16$, 极小值为 $f(2)=-16$ 。显然, 对于可求导数的函数 $f(x)$, 取得极大(小)值的点 x_0 必有 $f'(x_0)=0$ (即有水平切线)。但并不是说使 $f'(x_0)=0$ 的点都能取得极值, 例如, $f(x)=x^3$, 它有导数 $f'(x)=3x^2$, 且 $f'(0)=0$, 可是当 $x=0$ 时, $f(0)$ 不是极值 (如图 66), 因为当 $x \neq 0$ 时, $f'(x) > 0$, 不改变符号, 所以 $f(0)$ 不是极值。只有当函数 $y=f(x)$ 在 x_0 附近向上凸时 (即向下凹时), 在 x_0 处函数才有极大值 $f(x_0)$; 当函数 $y=f(x)$ 在 x_0 附近向下凸 (向上凹) 时, 在 x_0 处函数才有极小值 $f(x_0)$ 。对于存在二阶导数的函数, 其图形的凸凹还可以用二阶导数 $f''(x)$ 来判断: $f''(x) > 0$ 表示函数图象向下凸 (向上凹); $f''(x) < 0$ 表示函数图象向上凸 (向下凹)。所以, 对二阶导数存在的函数可用下面的列表判别有无极值:

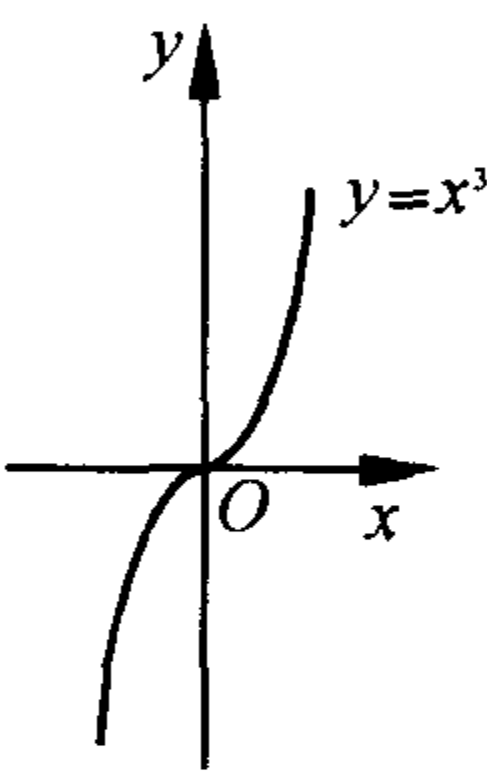


图 66

x	$< x_0$	x_0	$> x_0$
$f'(x)$		0	
$f''(x)$	< 0	< 0	< 0
$f(x)$	↖ 向上凸	极大值 $f(x_0)$	↖ 向上凸

x	$< x_0$	x_0	$> x_0$
$f'(x)$		0	
$f''(x)$	> 0	> 0	> 0
$f(x)$	↗ 向下凸	极小值 $f(x_0)$	↗ 向下凸

前例 $f(x) = x^3 - 12x$, $f'(x) = 3x^2 - 12$, $f''(x) = 6x$, $f'(\pm 2) = 0$, 当 $x < 0$ 时, $f''(x) < 0$, 图象向上凸; 当 $x > 0$ 时, $f''(x) > 0$, 图象向下凸, 故 $f(2)$ 为极小值, $f(-2)$ 为极大值。

对于那些导数不存在的函数, 如

$$f(x) = |x| = \begin{cases} x & (x \geq 0) \\ -x & (x < 0) \end{cases} \quad (\text{图 57})$$

它在 $x = 0$ 处虽有极小值 $f(0) = 0$, 但该点 $(0, f(0))$ 无切线, 即该点的导数不存在, 不能使用上述判别法, 要具体分析判别。由于函数的极值是一个局部概念, 是与邻近的函数值相比较而言的, 因此, 求函数 $f(x)$ 的最大(或最小)值时, 必须先求出所有极大值和极小值以及区间端点的函数值, 将这些函数值作比较后才能确定函数 $f(x)$ 在区间上的最大值和最小值(图 67)。对某些实际问题或某些函数(如

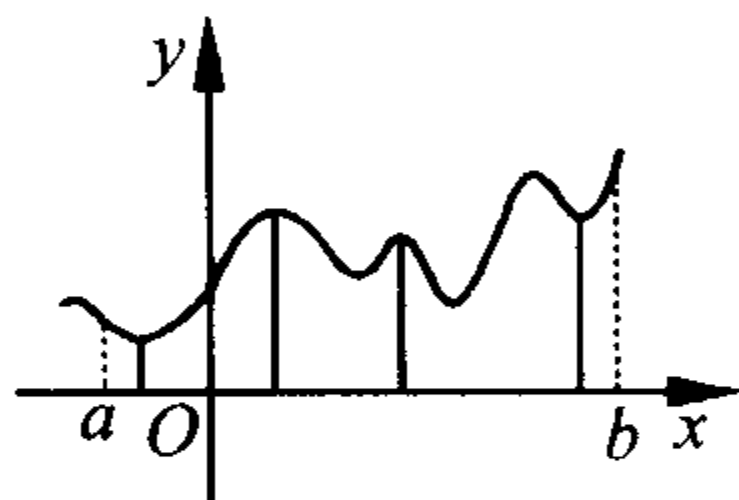


图 67

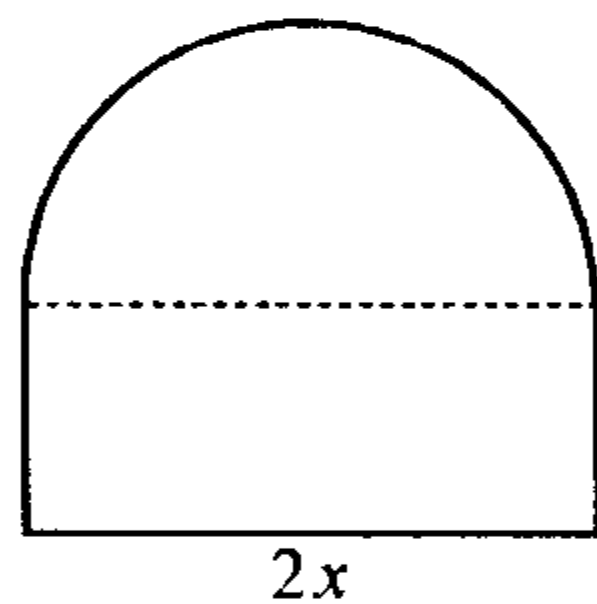


图 68

二次函数), 在它的定义范围内, 只有一个最大(小)值时, 那么当求到一个极大(小)值, 则这个极大(小)值必为最大(小)值。

例如, 建造引水隧道, 隧道的截面由矩形加半圆组成(图 68)。要求隧道截面周长为 15 米, 如何选取尺寸才能使隧道截面积最大?

若以 $2x$ 表示矩形底长, 则半圆的半径为 x , 矩形的宽必为 $\frac{15 - \pi x - 2x}{2}$, 隧道的截面积 $S(x) = 2x \cdot \frac{15 - \pi x - 2x}{2} + \frac{\pi x^2}{2} = 15x - \left(2 + \frac{\pi}{2}\right)x^2$, 求 x 使 $S(x)$ 极大。因为 $S'(x) =$

$15 - (\pi + 4)x$, 令 $S'(x) = 0$, 解得 $x = \frac{15}{\pi + 4}$ 。当 $x < \frac{15}{\pi + 4}$ 时, $S'(x) > 0$, $S(x)$ 递增; 当 $x > \frac{15}{\pi + 4}$ 时, $S'(x) < 0$, $S(x)$ 递减, 即在 $x = \frac{15}{\pi + 4}$ 两侧 $S'(x)$ 改变了符号, 故 $x = \frac{15}{\pi + 4}$ 时, $S(x)$ 有极大值。这个极大值必为最大值, 所以当半圆半径为 $x = \frac{15}{\pi + 4}$ 米, 矩形底长为 $2 \times \frac{15}{\pi + 4}$ 米时, 隧道截面积最大。

又如, 某厂每月生产 q 吨产品的总成本 $C(q)$ 为 $\frac{1}{3}q^3 - 7q^2 + 111q + 40$ (万元), 每月销售这些产品的总收入为 $R(q) = 100q - q^2$ (万元), 每月产量 q 为多少时才能获利最大?

由于每月的利润为:

$$\begin{aligned} L(q) &= R(q) - C(q) \\ &= 100q - q^2 - \left(\frac{1}{3}q^3 - 7q^2 + 111q + 40 \right) \\ &= -\frac{1}{3}q^3 + 6q^2 - 11q - 40 \end{aligned}$$

$L'(q) = -q^2 + 12q - 11$, 令 $L'(q) = 0$, 解得 $q_1 = 1$, $q_2 = 11$ 。即 $L'(1) = 0$, $L'(11) = 0$, 且 $L'(q)$ 与 $L(q)$ 随 q 的变化情况如下:

q	< 1	1	$(1, 11)$	11	> 11
$L'(q)$	< 0	0	> 0	0	< 0
$L(q)$	递减 ↘	极小值 $L(1)$	递增 ↗	极大值 $L(11)$	递减 ↘

故 $q = 11$ 时, $L(q)$ 最大。即当每月生产 11 吨时, 利润最大为

$$L(11) = -\frac{1}{3} \times 11^3 + 6 \times 11^2 - 11 \times 11 - 40 = 121\frac{1}{3} \text{ (万元)}$$

以上两个例子也可以用二阶导数方法求解。

在日常生活中,许多问题都可以归结为求极值问题,在经济领域中尤为普遍。消费者按一定标准求最优值,企业和商业部门则寻求一定的产量(或库存)、价格,以获得最大利润和付出最小成本。

牛顿和莱布尼兹创立的微积分学使人们豁然开朗,进入了新的天地。微分法为求极值提供了一整套完整的算法,它不仅简便,而且适用于所有可微函数。

微分方法求极值的推广

我们已知道,一元二次函数的极值是在它的抛物线图象的顶点取得。不难想像,如果函数的图象是抛物面,它的极值也应在抛物面的顶点取得。形如起伏的山峦、沙漠、海洋等曲面,它们的极值也应在它们的顶峰或深谷取得。这些曲面的数学表达式如何? 如何求这些曲面的极值?

如果把曲面看成是由空间中的点组成,为了表示空间中的点,必须先建立空间直角坐标系。为此,在直角坐标平面 xOy 上,过原点引出 z 轴,它垂直于 xOy 平面,如图 69,从而构成空间直角坐标系。空间中的任一点 M 能惟一确定三个有序数 x, y, z ; 反之,任给三个有序数 x, y, z 必可确定空间惟一点 M , 这样空间中的点与三个有序数建立了一一对应关系,这三个有序数 x, y, z 称为点 M 的坐标。在空间直角坐标

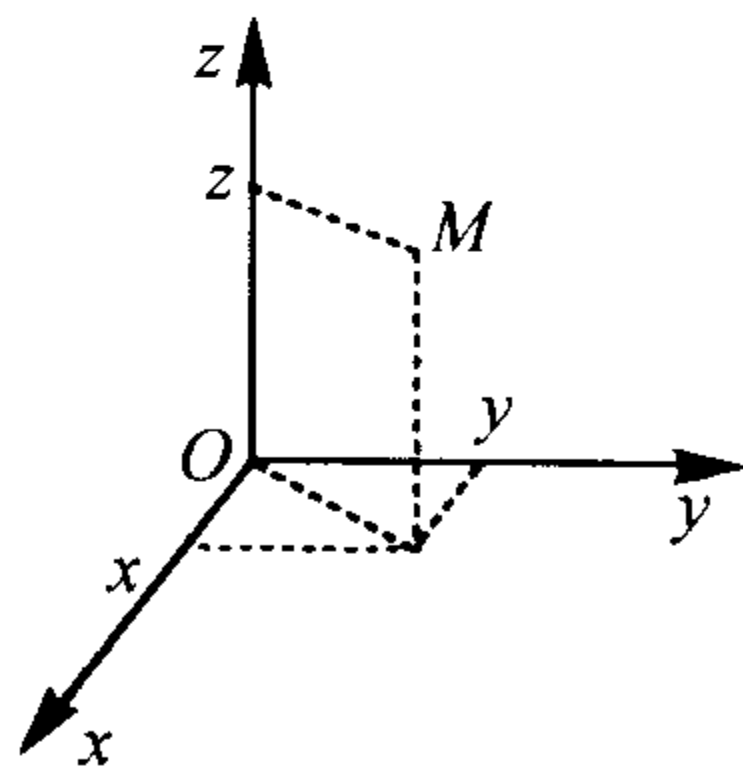


图 69

系中有许多性质和特性是与平面直角坐标系类似的。例如，平面上两点 $A(x_1, y_1)$, $B(x_2, y_2)$ 的距离 $|AB| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ ，空间中的两点 $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$ 的距离 $|P_1 P_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$ 。平面上以原点为中心， R 为半径的上（下）半圆为 $y = \sqrt{R^2 - x^2}$ ($y = -\sqrt{R^2 - x^2}$)。在空间中，则有类似的上（下）半球面为 $z = \sqrt{R^2 - x^2 - y^2}$ ($z = -\sqrt{R^2 - x^2 - y^2}$)，如图 70。平面直角坐标系中过原点的抛物线为 $y = x^2$ ，在空间直角坐标系中有相应的抛物面 $z = x^2 + y^2$ (图 71) 等等。可是，球面、抛物面都可以用 z 关于变量 x, y 的函数来表示，如果对 xOy 平面上某一区域 D 中的每一点 (x, y) 都能按一定的对应关系确定相应的函数值 z ，就称变量 z 是变量 x, y 的二元函数，即 $z = f(x, y)$ ，变量 x, y 称为自变量， x, y 的取值范围 D 称为函数 $Z = f(x, y)$ 的定义域。二元函数 $z = f(x, y)$ 表示曲面 (图 72)。关于一元函数极值的一些概念和求法类似可推广到二元函数。二元函数 $z = f(x, y)$ 在某点 $P_0(x_0, y_0)$ 的极值是相对于点 P_0 附近的点而言的。二元函数 $z = f(x, y)$ 在点

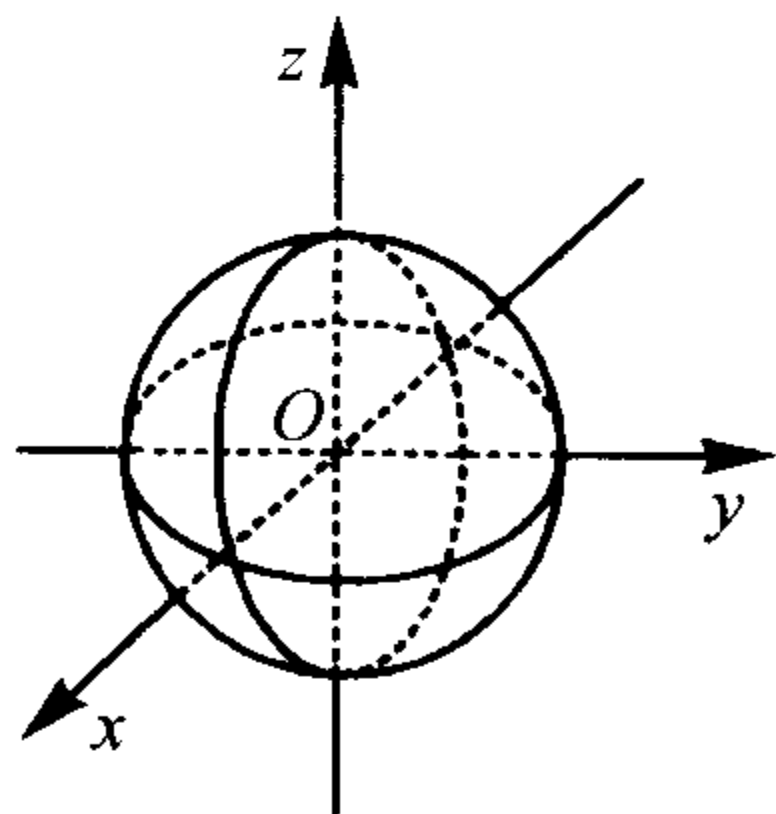


图 70

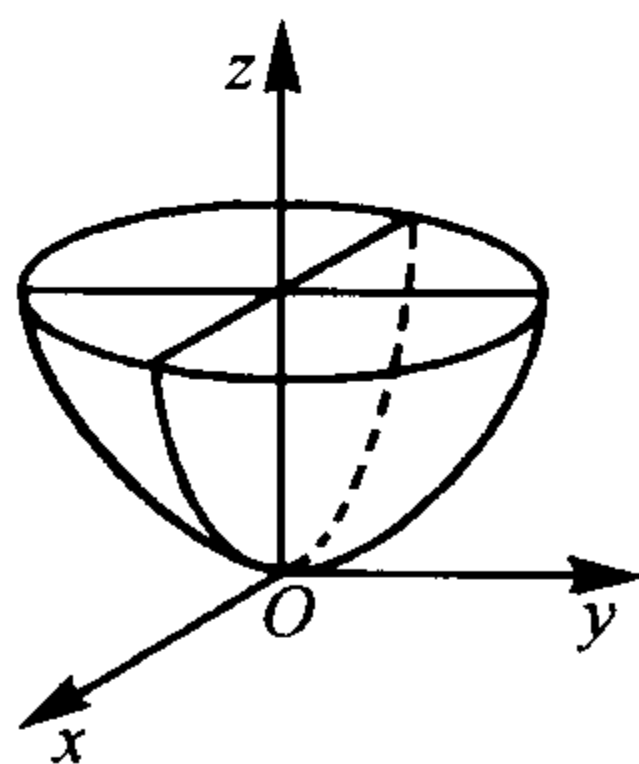


图 71

$P_0(x_0, y_0)$ 有极大(或极小)值是指 $f(x, y)$ 在 P_0 点的值 $f(x_0, y_0)$ 大于(或小于)它在 P_0 周围(如以 P_0 为圆心的小圆)各点的值, 即 $f(x, y) \leq f(x_0, y_0)$ [或 $f(x, y) \geq f(x_0, y_0)$], 则称函数 $z = f(x, y)$ 在点 P_0 处有极大(或极小)值 $f(x_0, y_0)$ 。对于光滑的曲面, 当在 $P_0(x_0, y_0)$ 处有极小值(图 72)或有极大值(图 73), 曲面在 P_0 点的切平面必平行于 xOy 平面, 曲面的切平面方程为 $Z - f(x_0, y_0) = f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0)$, 其中 $f'_x(x_0, y_0)$,

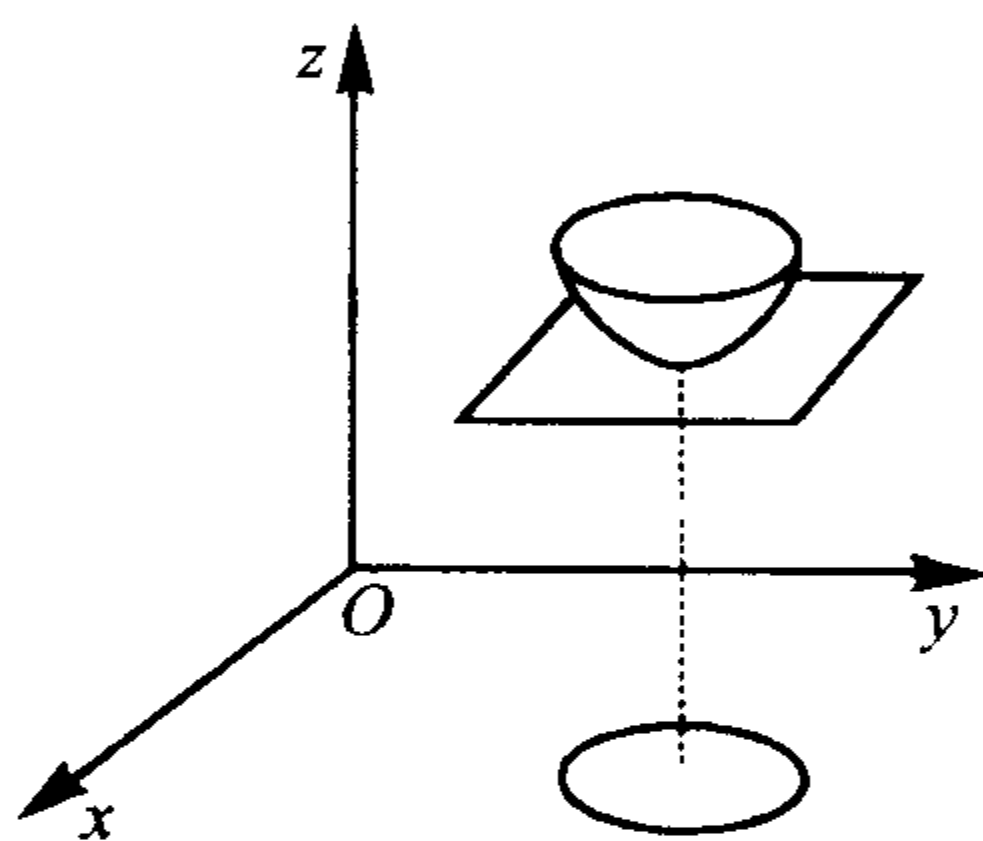


图 72

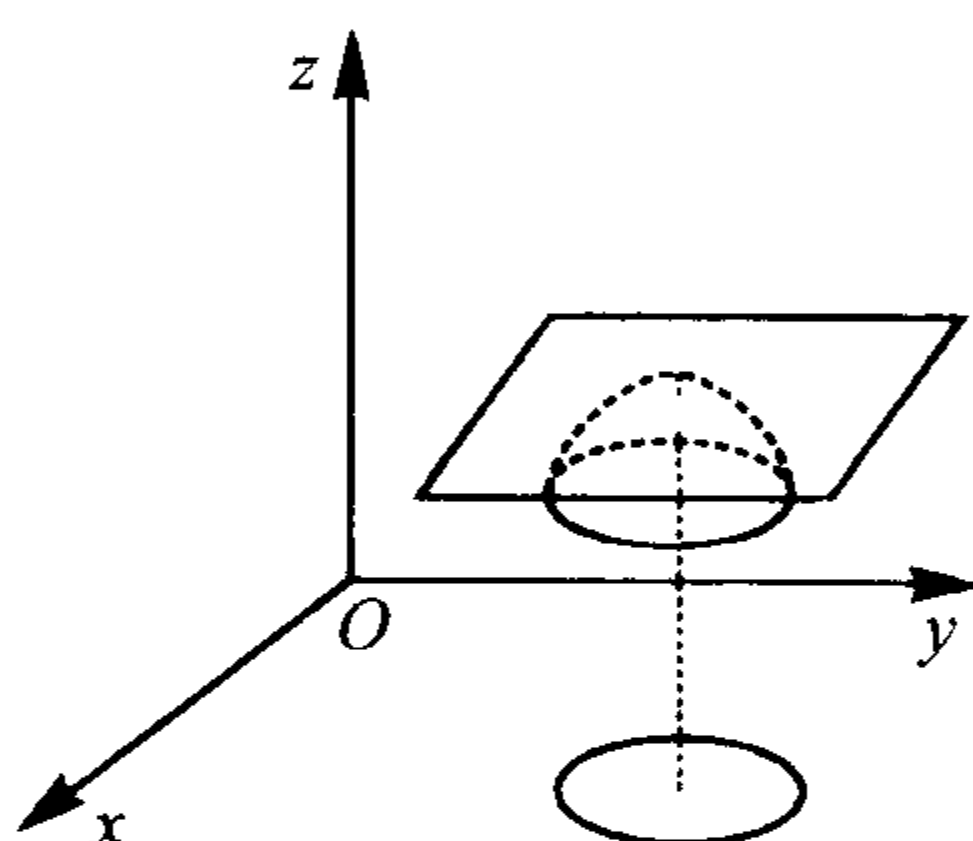


图 73

$f'_y(x_0, y_0)$ 分别是 $f(x, y)$ 在 P_0 处对 x, y 的偏导数。偏导数 $f'_x(x, y)$ 是当 y 取为常量时, 函数 $z = f(x, y)$ 成了 x 的一元函数而对变量 x 所求的导数。类似地, $f'_y(x, y)$ 是当 x 取常量时, 函数 $z = f(x, y)$ 成了 y 的一元函数, 再对变量 y 所求的导数。由切平面方程可知, 当有极大(小)值 $f(x_0, y_0)$ 时, 必须 $f'_x(x_0, y_0) = 0, f'_y(x_0, y_0) = 0$ 。

类似于一元函数, 求二元函数极值时, 可先求函数对 x, y 的偏导数, 并令

$$\begin{cases} f'_x(x, y) = 0 \\ f'_y(x, y) = 0 \end{cases}$$

解出 x, y 的值 (x_0, y_0) 。在点 $P_0(x_0, y_0)$ 处是否能取得极值, 则要进一步判别。可将 $f(x_0, y_0)$ 与点 (x_0, y_0) 周围的函数值作比较, 或进一步通过二阶偏导数判定:

记 $A = f''_{xx}(x_0, y_0), B = f''_{xy}(x_0, y_0), C = f''_{yy}(x_0, y_0)$, 计算 $AC - B^2$ 的值,

$$\text{当 } AC - B^2 \begin{cases} > 0 \text{ 时, } f(P_0) \text{ 为极值。当 } A < 0 \text{ 时, } f(P_0) \\ & \text{为极大值; 当 } A > 0 \text{ 时, } f(P_0) \text{ 为极小值。} \\ < 0 \text{ 时, } P_0 \text{ 不是极值点。} \\ = 0 \text{ 时, 不能确定。} \end{cases}$$

如果 $f'_x(x, y) = 0, f'_y(x, y) = 0$, 仅有惟一解, 而又已知问题确有极值, 那么这个解 (x_0, y_0) 便是取极值的点。

例如, 某农户要用钢板做一个容积为 2 米³ 的长方体贮粮柜。粮柜的长、宽、高的尺寸为多少时, 用料最省? 这是一个下料问题, 目标是求柜子的表面积 A 最小。设柜子的长、宽分别为 x 米和 y 米, 其高 z 必为 $\frac{2}{xy}$ 米, 柜子的表面积为 $A = 2\left(xy + \frac{2}{y} + \frac{2}{x}\right)$ 。 A 是 x, y 的二元函数, 要求 A 最小时的尺寸 x, y, z , 先求 A 对 x, y 的偏导数:

$$\begin{cases} A'_x = 2\left(y - \frac{2}{x^2}\right) \\ A'_y = 2\left(x - \frac{2}{y^2}\right) \end{cases} \text{ 并令 } \begin{cases} A'_x = 0 \\ A'_y = 0 \end{cases} \text{ 即 } \begin{cases} 2\left(y - \frac{2}{x^2}\right) = 0 \\ 2\left(x - \frac{2}{y^2}\right) = 0 \end{cases}$$

解得 $x = \sqrt[3]{2}, y = \sqrt[3]{2}$, 即长、宽均为 $\sqrt[3]{2}$, 此时高必为 $\sqrt[3]{2}$ 。由解的惟一性可知, 当长 = 宽 = 高 = $\sqrt[3]{2}$ 时, 函数取得最小值, 即柜子的边长为 $\sqrt[3]{2}$ 米的正方形为最优解, 此时用料为 $f(x, y, z) = 6\sqrt[3]{4}$ (米²), 是最省的。

一般情况下, 使 $f'_x(x, y) = f'_y(x, y) = 0$ 的点函数不一

定能取得极值。例如 $z = y^2 - x^2$, 它的 $f'_x(x, y) = -2x$, $f'_y(x, y) = 2y$, $f'_x(0, 0) = f'_y(0, 0) = 0$, 但在点 $(0, 0)$ 函数不能取得极大(小)值(图 74)。由于其图象如马鞍状, 故称点 $(0, 0)$ 为鞍点。

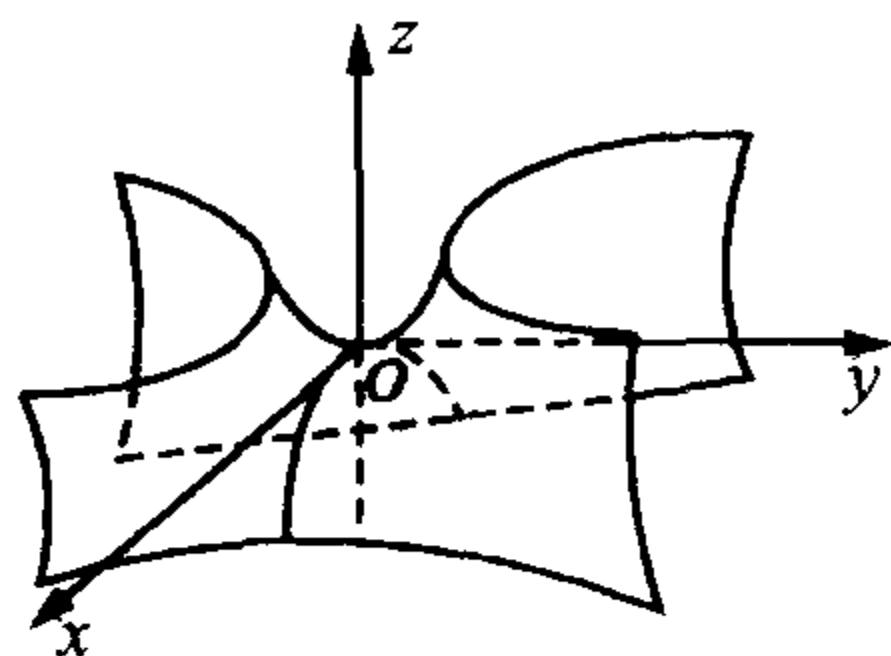


图 74

求一元函数和二元函数的极值的微分法可以推广到更为一般的情况, 对 n 元函数 $f(x_1, x_2, \dots, x_n)$ 的极值问题亦有相应的微分法。微分法求极值的特点是利用函数图象在顶点的切线或切面直接求极值, 因此它要求函数在顶点有切线或切面。还有一些求极值的方法是通过逐步搜索求得最优点。从一个初始点出发, 向不同的方向搜索, 寻找有利的方向前进, 如此反复, 逐步逼近最佳点。由于确定搜索的方式不同而产生各种不同方法, 如最速下降法、牛顿法等。

最优化通常有两种情况——求最大或最小, 如效益最大、质量最优或成本最小、耗资最小等。由于函数 $f(x)$ 的极大值与 $-f(x)$ 的极小值是等价的, 因此, 求函数极大值可转化为求 $-f(x)$ 的极小值。通常的最优化问题常以求目标函数 $f(x)$ 的极小值为标准形式, 常称为极小化问题。有些求函数的极值问题是附有约束条件的, 称为有约束条件的最优化问题。有约束条件的最优化问题通常设法转化成无约束条件形式后再求解, 所以无约束条件的最优化问题是最基本的。对于带约束条件 $g_j(x) = 0, j = 1, 2, \dots, m$ 的最优化问题, 有两种处理方法: 消元法(降维法)和拉格朗日乘子法(升维法)。

消元法是把约束条件中的某些变量表示成另一些变量的函数, 代入目标函数, 使目标函数的变量个数减少(即降维)。

仍以做粮柜为例,目标函数是用料(表面积) A , $A = 2(xy + yz + zx)$,是三维函数,它的约束条件为 $xyz - 2 = 0$,从约束条件中解出 $z = \frac{2}{xy}$,代入目标函数,使目标函数变成 $A = 2\left(xy + \frac{2}{y} + \frac{2}{x}\right)$,目标函数从三维函数降为二维函数,问题转化为无约束条件的最优化问题。一般地,对 n 维目标函数带有 m 个约束等式的优化问题,可先解约束条件联立方程组,把 n 个变量中的 m 个变量用其余 $n - m$ 个变量表示,再代入目标函数,这样就消去了 m 个变量,使目标函数从 n 维降为 $n - m$ 维,原问题就转化为无约束条件的最优化问题。这种消元法虽然简单,但实际使用时困难很大,甚至无法实现。因为当约束方程组复杂时,往往解不出来,即使解出来,也常使降维后的目标函数很繁,难于处理。所以此法实用意义不大。

拉格朗日乘子法是解有约束条件最优化问题的一种经典方法,它通过构造拉格朗日函数,将原问题转化为无约束条件的最优化问题。

对带约束条件 $h_j(x) = 0, j = 1, 2, \dots, m$ 的最优化问题 $\min f(x)$,可先构造拉格朗日函数:

$$L(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j h_j(x)$$

其中 λ_j 是待定常数,称为拉格朗日乘数。然后用微分法求解。

例如,表面积为 a^2 的长方体中,求体积最大者。设长方体的长、宽、高分别为 x, y, z ,问题是要在约束条件

$$2(xy + yz + zx) = a^2, x, y, z > 0$$

下求体积函数的最大值 $\max V = xyz$ 。用拉格朗日乘子法,需先将约束条件转化为约束等式:

$$h(x, y, z) = 2(xy + yz + zx) - a^2 = 0$$

再构造拉格朗日函数 $L(x, y, z, \lambda)$:

$$\begin{aligned} L(x, y, \lambda) &= xyz + \lambda h(x, y, z) \\ &= xyz + 2\lambda \left(xy + yz + zx - \frac{a^2}{2} \right) \end{aligned}$$

变量由三个上升为四个(升维),问题转化成无约束问题。分别对 x, y, z, λ 求偏导数,并令它们等于 0:

$$\begin{cases} L'_x(x, y, z, \lambda) = yz + 2\lambda(y + z) = 0 \\ L'_y(x, y, z, \lambda) = xz + 2\lambda(x + z) = 0 \\ L'_z(x, y, z, \lambda) = xy + 2\lambda(x + y) = 0 \\ L'_\lambda(x, y, z, \lambda) = 2\left(xy + yz + zx - \frac{a^2}{2}\right) = 0 \end{cases}$$

得解为 $x = y = z = \frac{\sqrt{6}}{6}a$ 。

所以,在表面积为 a^2 的条件下,棱长为 $\frac{\sqrt{6}}{6}a$ 的正方体的体积

最大,最大的体积 $V = xyz = \frac{\sqrt{6}}{36}a^3$ 。

在实际问题中,存在更广泛的一类函数,它的函数值不是取实数,函数的自变量也不是实数,而是函数。这种函数不是实变量函数,而是函数的“函数”,常称为泛函数,如求 $J(y)$

$= \int_{x_1}^{x_2} F(x, y, y')dx$ 的极值。这里 $y = f(x)$ 是要求的未知

函数,选择 $y(x)$ 使 $J(y)$ 最优,这类极值问题称为泛函极值。研究这类极值的方法称为变分法。变分法的发展最早来自三大问题——最速降线问题、等周问题和测地线问题的讨论。

最速降线是指不在同一铅垂线上的两个定点 P_1 和 P_2 ,问质点从 P_1 沿怎样的曲线下降到 P_2 所用的时间最短。在

16 世纪以前,几乎所有的人都认为连接两点的直线段是最短的,路线最短所用的时间也必是最少的。17 世纪初期,伽利略提出异议,他认为按自由落体原理很可能是过 P_1 和 P_2 的圆弧。1696 年 6 月,约翰·伯努利在《教师学报》上提出这个问题,作为向欧洲数学家的挑战。半年后问题仍未解决,伯努利于 1697 年元旦再次发表著名的公告,向“全世界最能干的数学家”挑战。这个问题的形式是求 $y(x)$,使求表示下降时间的积分

$$J(y) = \frac{1}{\sqrt{2g}} \int_{x_1}^{x_2} \sqrt{\frac{1 + [y'(x)]^2}{y(x) - \alpha}} dx$$

最小(图 75)。其中 g 为重力加速度, $\alpha = y_1 - \frac{v_1^2}{2g}$, v_1 为初速度。正确的答案是连结 P_1 与 P_2 的上凹旋轮线:

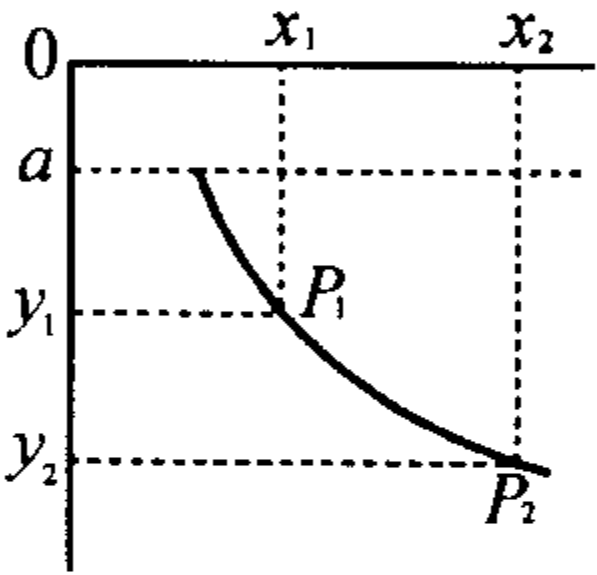


图 75

$$x = R(\theta - \sin\theta), y = R(\theta - \cos\theta), \theta \text{ 为参数。}$$

牛顿、莱布尼兹及伯努利兄弟都求得正确解答。令人吃惊的是,惠更斯对钟摆的研究结果竟与最速降线吻合。摆锤沿着摆线弧摆动,不论其摆幅大小,做一次完全摆动,所用的时间是完全相同的,故称摆线为等时曲线。等时曲线也是最速降线。

等周问题是指在给定周长的所有封闭平面曲线中,求它所围的面积最大者。这个问题可追溯到古希腊以前的时代。有个故事,说的是古代有一位公主叫狄多,离开自己的父亲泰雅王,历尽艰险到了北非的地中海海滨定居。她向当地土著购买一块“不大于一张犍牛皮所能围起来的”土地。精明的公主将犍牛皮切割成很细的皮条,再将这些牛皮条一条一条连接起来。怎样使牛皮条所围的土地面积最大? 由于靠海的边

沿不需要用牛皮绳围,于是公主用牛皮绳靠海岸围了一个半圆。传说狄多在那里成为迦太基人的奠基人和第一位女王。1697年,雅各布·伯努利提出了一个包含几种情形的相当复杂的等周问题,向他的弟弟约翰·伯努利挑战。如果解答正确,约翰可得50个金币的奖励。约翰最初给出了几种解法,但都是错误的。1700年,雅各布自己得出一个正确的答案,答案为圆。兄弟俩各为自己解法的正确性而争论。1718年,约翰在此基础上做了进一步研究,得出一个精确的、形式上漂亮的等周问题的一般解法。等周问题就是要求 $x(t)$ 和 $y(t)$,

使得弧长 $L = \int_{t_1}^{t_2} \sqrt{[x'(t)]^2 + [y'(t)]^2} dt$ 为给定常数时,

面积积分 $J = \int_{t_1}^{t_2} (xy' - x'y) dt$ 取得极大值。

测地线问题(也称短程线问题)是指曲面上两点间长度最短的路径。对于最短路线问题我们有过许多幻想和趣题。上小学时,听老师说,美国在地球的另一面,中国是黑夜时美国正是白

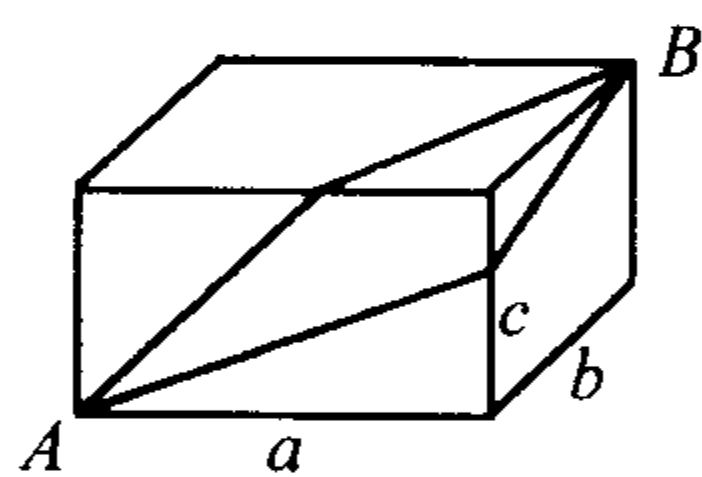


图 76

天,心想若是在地球上开个隧道,直通美国,该多方便啊!因为连结两点的线段是最短的路线。一只蜘蛛在房间一个角落A上,一只苍蝇在它的对角线的另一端B(图76),房间的长、宽、高分别为 a, b, c 。蜘蛛应沿怎样的路线爬行,才能最快捉到苍蝇?类似的,苍蝇为了防止蜘蛛的袭击,要查看房间各个面上是否有蜘蛛埋伏,检查一遍然后回到原地,苍蝇至少要爬多长的路?这些都是简单的测地线问题。1697年,约翰提出了在凸曲面上求两点间的最短弧问题。1698年,约翰的哥哥雅各布解决了锥面和旋转面上的测地线问题。1728年,约

翰又用雅各布的方法取得另外几类曲面的最短线问题的解。

一般测地线问题的形式是 $J = \int_{x_1}^{x_2} f(x, y, y') dx$, 是要求从 (x_1, y_1) 到 (x_2, y_2) 的路线 $y(x)$, 使 J 达到极大或极小。

约翰在最速降线、等周问题及测地线问题上取得丰硕的成果, 使他成为变分法的先驱之一。这三个问题的提出, 引起众多数学家的兴趣, 除伯努利兄弟外, 牛顿、莱布尼兹、洛比达等人都以各自的方法参与作答, 为变分法奠定了基础。19 世纪中叶, 在弹性体的研究中提出的多重积分的极值问题, 曾引起数学界的兴趣, 最后由阿尔采拉、希尔伯特和托内利等人解决了。以上这些问题均属古典变分问题, 它们的解是在定义域内部取得的。到 20 世纪 40 年代, 布利斯及麦克沙恩研究了泛函极值在定义域边界上取得的非古典泛函极值, 这类极值问题已涉及最优控制理论。

优化问题的未来发展



非线性规划

在现实世界中,虽然线性规划问题广泛存在,但是许多实际问题难以用线性函数描述,它的结构也不是由直线、平面等几何形体构成。非线性的现象大量存在,有些问题看上去很简单,但并不是线性函数所能描述的,连一个长方体的容积和侧面积都不是长、宽、高的线性函数。一块边长为 60 厘米的正方形铝板,欲将其做成一个底为正方形的无盖长方体容器,只需在铝板的四角剪去一个相等的小正方形,然后

将各边折弯后焊接即可。为使容积最大,应剪去多大的小正方形?显然,剪去的小正方形边长 x 就是容器的高,容器的容积 $V = x(60 - 2x)^2$, x 的取值范围应是 $0 \leq x \leq 30$, 它的目标函数为容积最大 $\max x(60 - 2x)^2$ (图 77), 这是单变量 x 的三次函数, 不是线性函数。这个问题中材料一定, 求最大容积, 如果换一个角度, 例如要求容积一定, 使材料费用最省的底为正方形的长方体无盖容器, 该如何选尺寸。容器四壁的材料每平方米重 2 千克, 价格为 10 元, 底面材料每平方米重 1.5 千克, 价格为 15 元, 容积规定为 12 米³, 要求总重量不超过 40 千克, 选取怎样的尺寸能使材料费用最小? 该容器底面边长为 x_1 , 容器高为 x_2 , 则所需费用 $z = 15x_1^2 + 40x_1x_2$, 问题是求目标函数的最小值, 即

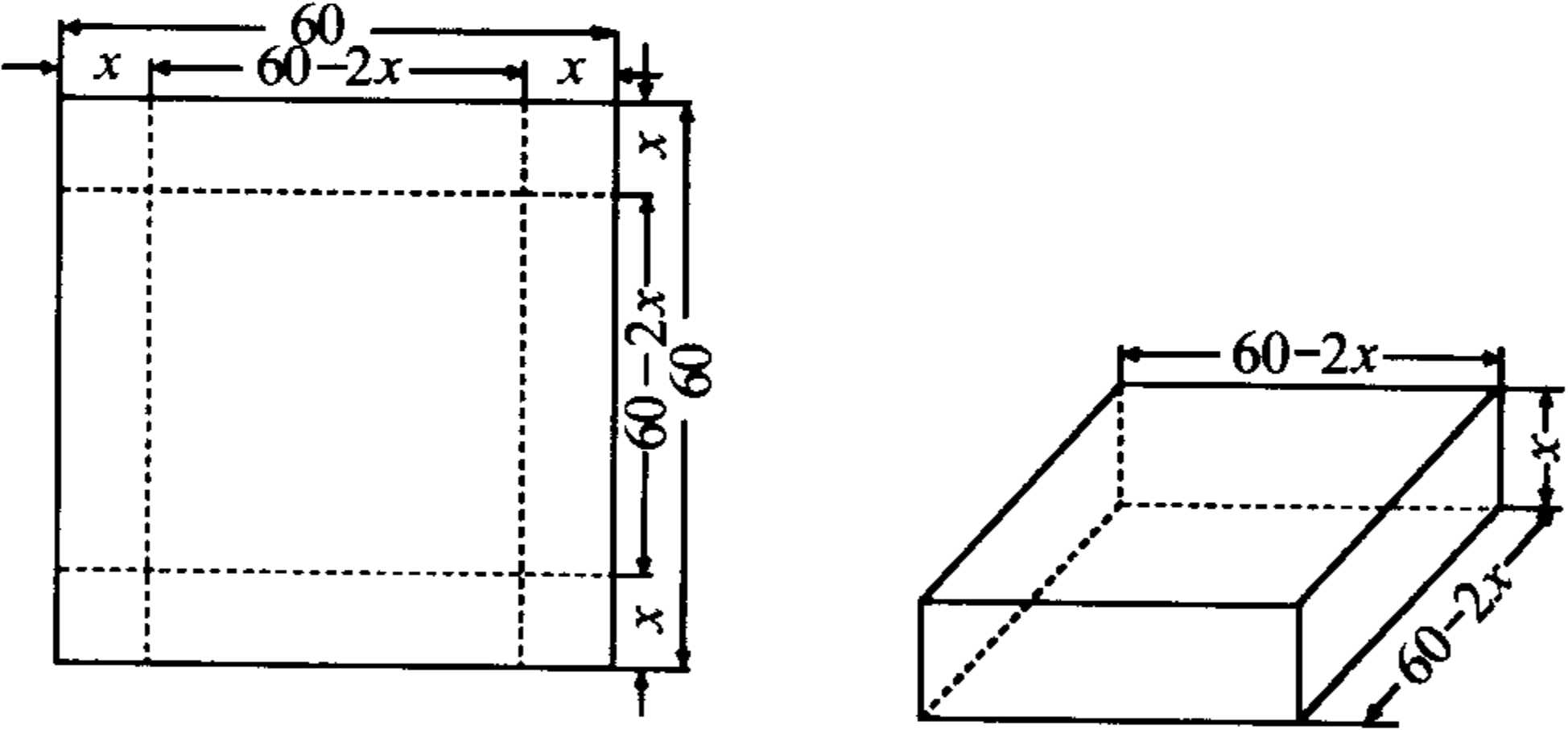


图 77

$$\begin{aligned} \min \quad & 15x_1^2 + 40x_1x_2 \\ \text{约束条件} \quad & \begin{cases} x_1^2 \cdot x_2 = 12 \\ 2 \cdot 4x_1 \cdot x_2 + 1.5x_1^2 \leq 40 \\ x_1 > 0, x_2 > 0 \end{cases} \end{aligned}$$

它的目标函数及两个约束条件函数都是两个变量非线性函数。在现实世界中, 非线性问题比比皆是, 远远多于线性问题。例如, 某公司共有 100 个用户, 它们的位置为 (a_i, b_i) ,

其中 $i = 1, 2, \dots, 100$ 。第 i 个用户的货物需求量为 A_i ，公司向用户送货的运价为每千米 B_i 元。公司仓库应建在何处才能使货物运送到各用户的总运费最小？假设仓库的位置是 (x, y) ，则向第 i 个用户供货的运费为 $A_i \cdot B_i \sqrt{(x - a_i)^2 + (y - b_i)^2}$ ，求 (x, y) 使得总运费

$$\sum_{i=1}^{100} A_i B_i \sqrt{(x - a_i)^2 + (y - b_i)^2} \text{ 最小, 它的目标函数也是非}$$

线性的。再如，某部门在一段时间内可用于投资的总金额为 a 亿元，可选择的项目有 n 个，其编号为 $1, 2, 3, \dots, n$ ，已知若对第 j 项投资 a_j 亿元，可获利为 c_j 亿元。应如何投资才能获得最佳的经济效益？该题的变量为 $x_1, x_2, \dots, x_n$ ，规定：

$$x_j = \begin{cases} 1 & (\text{对第 } j \text{ 项投资}) \\ 0 & (\text{不对第 } j \text{ 项投资}) \end{cases} \quad (j = 1, 2, \dots, n)$$

求 x_j 的值使目标函数(利润率)最大。即：

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

$$\text{约束条件 } \sum_{j=1}^n a_j x_j \leq a \quad (j = 1, 2, \dots, n)$$

它的目标函数亦为非线性的。

凡是数学模型中含有非线性函数的规划问题就称为非线性规划问题。由于非线性规划的复杂性，致使至今还没有一个适合任何形式的非线性规划问题的有效算法。目前只能对不同类型的问题给出不同的方法，并且只有当目标函数具有较好的特性时，才能得到比较满意的解。非线性规划问题亦可分两类，一类是无约束条件的非线性规划，另一类是有约束条件的非线性规划。由于等式约束条件 $g_j(x) = 0$ 等价于两

$$\text{个不等式} \begin{cases} g_j(x) \geq 0 \\ -g_j(x) \geq 0 \end{cases}$$

所以非线性规划的标准形式可写成：

$$\min f(x)$$

$$\text{约束条件 } g_j(x) \geq 0 (j=1, 2, \dots, l)$$

$$\text{其中 } x = (x_1, x_2, \dots, x_n)$$

对无约束条件的非线性规划，利用微分法求解时常常要解非线性方程组 $\frac{\partial f}{\partial x_i} = 0, i = 1, 2, \dots, n$ 。这是相当难解的，有许多实际问题往往很难求解或者根本求不出目标函数的偏导数。因此，求解非线性规划问题一般采用数值方法，按极小化要求最常用的方法是下降迭代法。所谓迭代，就是从一点 $x^{(0)}$ 出发，按某种规则（即算法）求出后继点 $x^{(1)}$ ，再从 $x^{(1)}$ 出发按同样的规则求出 $x^{(1)}$ 的后继点 $x^{(2)}$ ，如此类推，得到一系列点 $x^{(1)}, x^{(2)}, \dots, x^{(k)}, \dots$ 所谓下降，就是在每次迭代中要使后继点的目标函数值有所减少，即 $f[x^{(k+1)}] < f[x^{(k)}]$ 。可见下降迭代法就是逐步找后继点，使相应的目标函数值逐步减小，直到点列 $\{x^{(k)}\}$ 不断趋向目标函数极小值的点。迭代过程是一种搜索过程，在搜索过程中，从点 $x^{(k)}$ 出发必须确定搜索方向 d_k 和每步的步长 λ_k ，才能得到新的点（后继点） $x^{(k+1)}$ ， $x^{(k+1)} = x^{(k)} + \lambda_k d_k$ （图 78），使 $f(x^{(k)} + \lambda_k d_k) = \min f(x_k + \lambda d_k)$ ，搜索方向 d_k 由事先选定的规则（算法）确定。由于确定搜索方向的算法有多种，不同的算法产生了不同的寻优方法。例如，多元函数的求

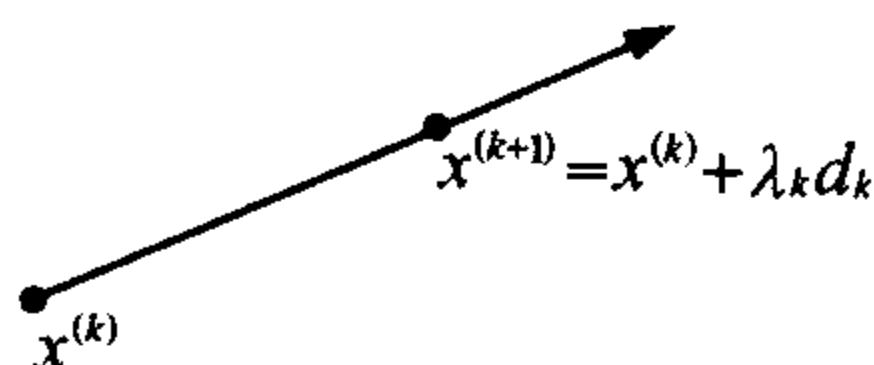


图 78

解过程中, 方向的选择可选择梯度方向。 n 元函数 $f(x)$ 的梯度方向 $f_x(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$ 是函数值增加最快的方向。负梯度方向 $-f_x(x) = \left(-\frac{\partial f(x)}{\partial x_1}, -\frac{\partial f(x)}{\partial x_2}, \dots, -\frac{\partial f(x)}{\partial x_n} \right)$ 是函数值下降最快的方向。

以二元函数 $u = f(x_1, x_2)$ 为例, 它的梯度记为 $\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right)$, 即以偏导数为坐标的向量。梯度方向是函数值增加最快的方向, 它是曲面等高线的法线方向(图 79)。因此, 负梯度方向 $\left(-\frac{\partial f}{\partial x_1}, -\frac{\partial f}{\partial x_2} \right)$ 是函数值减小最快的方向, 故称负梯度方向 $\left(-\frac{\partial f}{\partial x_1}, -\frac{\partial f}{\partial x_2} \right)$ 为最速下降方向。

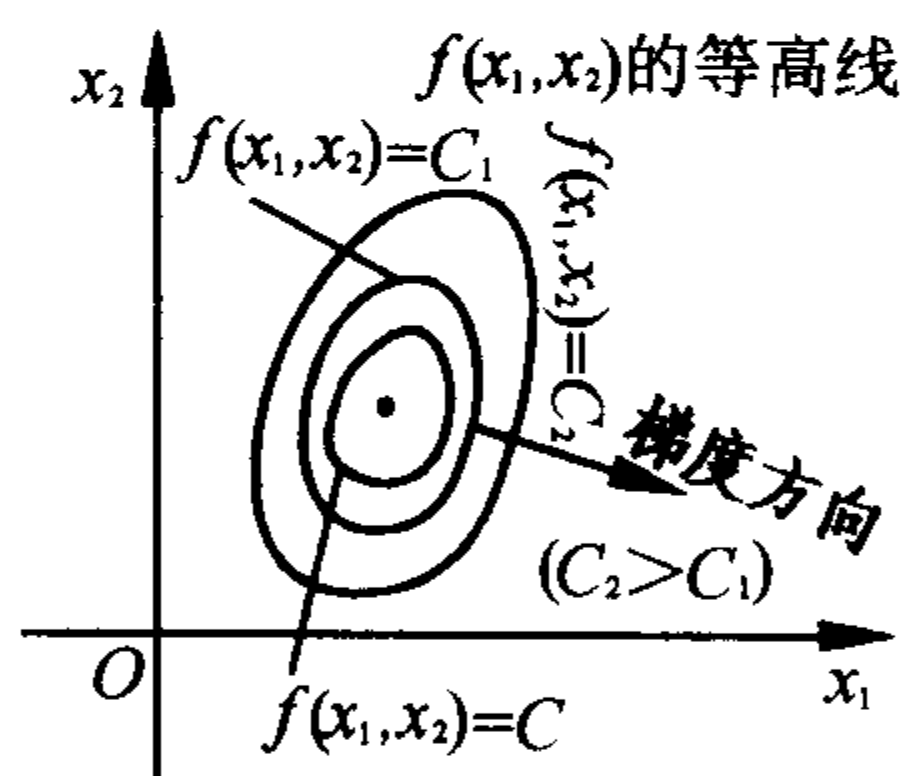


图 79

二元函数的几何图形为曲面(图 80)。而 $f(x_1, x_2) = c$ (常数)表示曲面 $u = f(x_1, x_2)$ 与平面 $u = c$ 的交线 l , l 在 $x_1 O x_2$ 平面上的投影称为等高线, 等高线的形状由曲面 $u = f(x_1, x_2)$ 决定; 反之亦然。当 $u = f(x_1, x_2)$ 为极小值 $\min f(x_1, x_2)$ 时, 等高线收缩为一点 (x_1^*, x_2^*) , 相应的函数值 $f(x_1^*, x_2^*)$ 为函数 $u = f(x_1, x_2)$ 的极小值 $\min f(x_1, x_2)$ 。求步长 λ_k 时, 由于 x_k 和 d_k 已知, 函数 $f(x_k + \lambda d_k)$ 是单变量

λ 的函数,按单变量函数求极值的方法便可得到 λ_k , 使得 $f(x_k + \lambda_k d_k) = \min f(x_k + \lambda d_k)$ 。

单变量函数寻优的方法很多,除微分法外还有 0.618 法、分数法、牛顿法、抛物线逼近法等。0.618

法和分数法适用于单峰函数,使用时必须先确定目标函数的单峰区间,然后再按 0.618 法、分数法求得最优解,它们原理简单,计算容易,效果也好,在实际中得到广泛的应用。牛顿

法(也叫切线法)利用公式 $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$ 迭代。它的几何意义是过点 $(x_k, f'(x_k))$ 作曲线 $f'(x)$ 的切线,与 x 轴交于点 $(x_{k+1}, 0)$ (如图 81),该切线的斜率显然是 $f''(x_k)$,

$f''(x_k) = \frac{f'(x_k)}{x_k - x_{k+1}}$,故 $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$ 。若 x_{k+1} 不满足

精度时,可过点 $(x_{k+1}, f'(x_{k+1}))$ 再作曲线 $f'(x)$ 的切线与 x 轴交于点 $(x_{k+2}, 0)$,如

此继续,得到一系列 $f'(x)$ 的切线与 x 轴的交点逼近 x^* , $f'(x^*) = 0$,这个点列 $\{x_k\}$ 逼近最优解的速度很快。显然,牛顿法要求目标函数有一阶、二阶导数,而且使用时要注意选好初

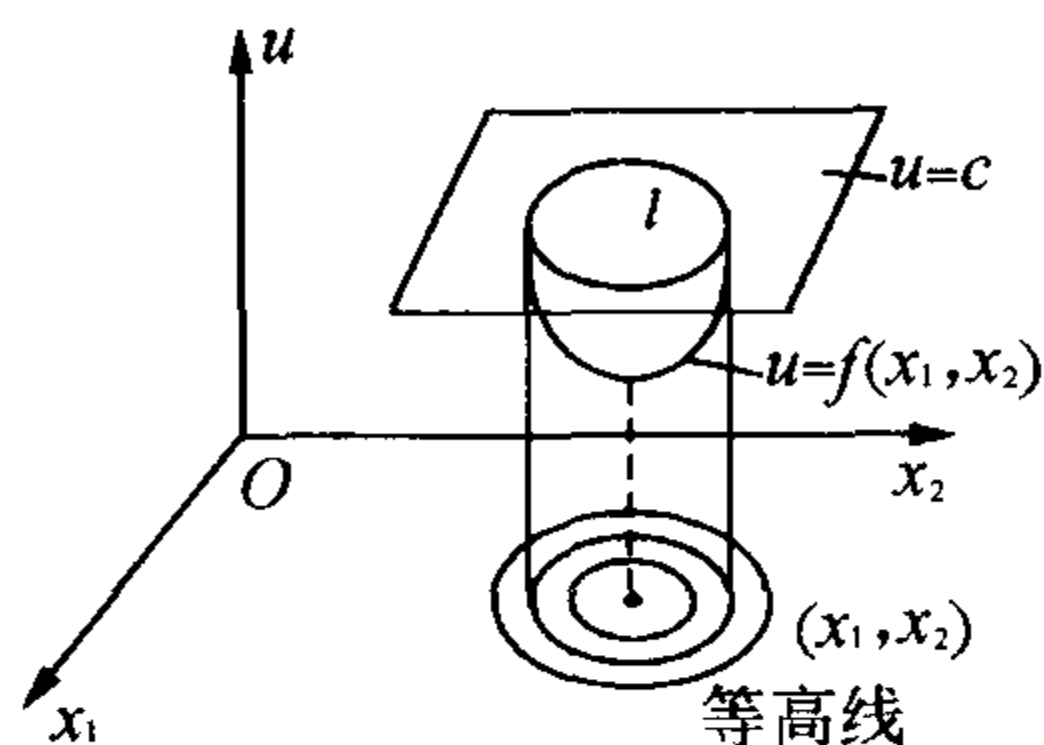


图 80

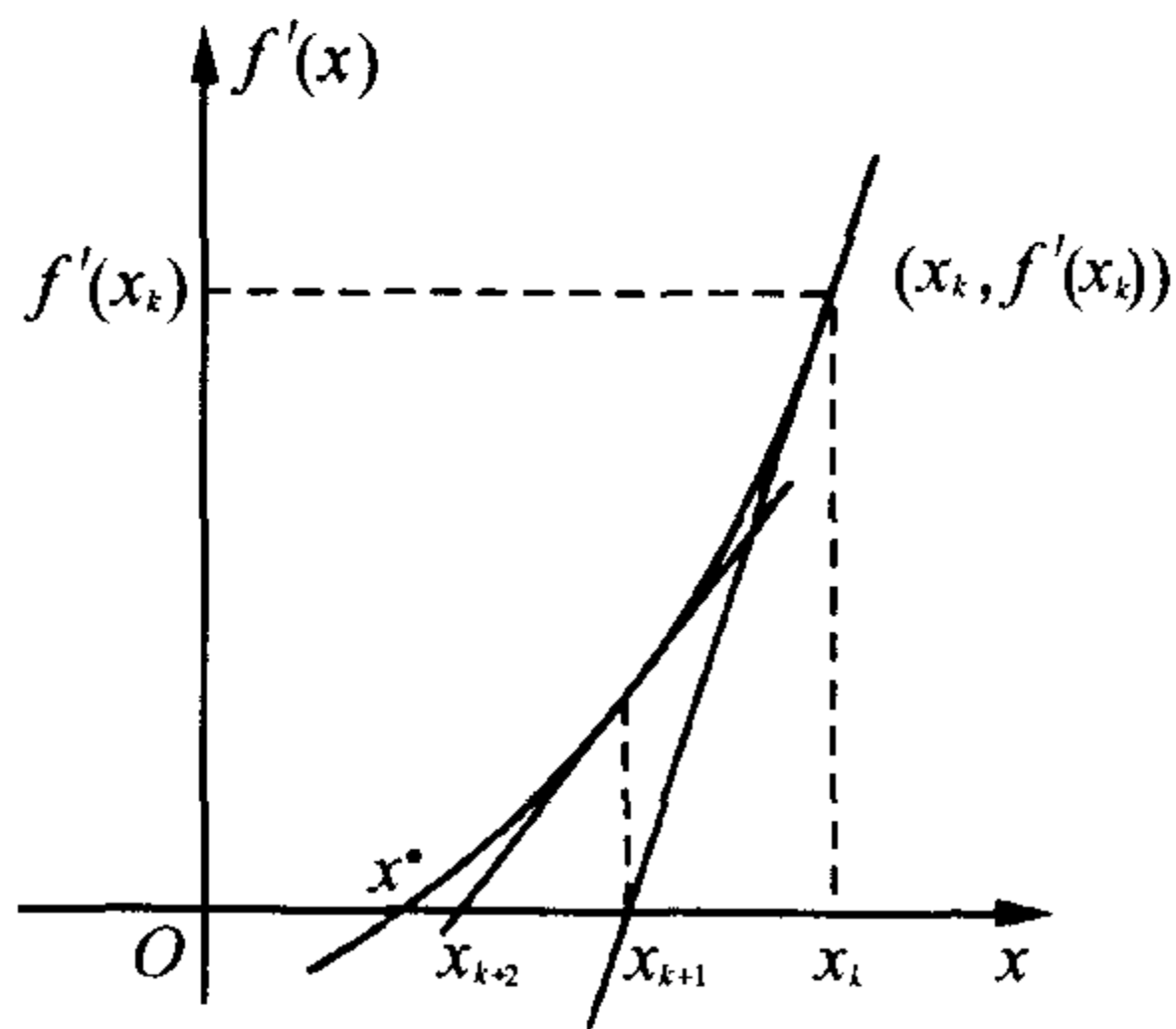


图 81

始点。有些函数难以求导,使用牛顿法受到限制。抛物线逼近法是在目标函数 $f(x)$ 的极小值附近,用抛物线[二次三项式 $\varphi(x) = a_2 x^2 + a_1 x + a_0$]逼近目标函数。方法是先选三个点 $x_1, x_2, x_3 (x_1 < x_2 < x_3)$, 使相应的函数值为 $f(x_1) > f(x_2), f(x_2) < f(x_3)$, 要求抛物线在这三点与目标函数 $f(x)$ 重合, 即

$$\begin{cases} \varphi(x_1) = a_2 x_1^2 + a_1 x_1 + a_0 = f(x_1) \\ \varphi(x_2) = a_2 x_2^2 + a_1 x_2 + a_0 = f(x_2) \\ \varphi(x_3) = a_2 x_3^2 + a_1 x_3 + a_0 = f(x_3) \end{cases}$$

因为 $\varphi'(x) = 2 a_2 x + a_1$, 令 $\varphi'(x) = 0$, 得 $x = -\frac{a_1}{2a_2}$, 由方程组消去 a_0 便可解得 a_1, a_2 及 x ,

$$x = \frac{1}{2} \frac{(x_2^2 - x_3^2)f(x_1) + (x_3^2 - x_1^2)f(x_2) + (x_1^2 - x_2^2)f(x_3)}{(x_2 - x_3)f(x_1) + (x_3 - x_1)f(x_2) + (x_1 - x_2)f(x_3)}$$

令 $x_k = x$, 从 x_1, x_2, x_3, x_k 中选择相应函数值最小的点作为新的 x_2 , 然后在 x_2 的左右各取一点, 组成新的三点, 重复上述过程, 可得 $f(x)$ 的极小值。

寻找单变量函数 $f(x)$ 最优解存在的区间可用外推内插法。以求函数极小值为例, 只需寻找三个点 $x_1, x_2, x_3, x_1 < x_2 < x_3$, 满足 $f(x_1) > f(x_2), f(x_2) < f(x_3)$ (两头大中间小) 即可。例如用外推内插法求函数 $f(x) = x^2 - 6x + 2$ 在 $[0, 10]$ 上的极值点存在区间。先外推: 取初始点 $x_1 = 1$, 初始步长 $h_0 = 0.1$,

$$\begin{array}{ll} x_1 = 1, & f(x_1) = -3 \\ x_2 = x_1 + h_0 = 1.1, & f(x_2) = -3.39 \\ x_3 = x_2 + 2 h_0 = 1.3, & f(x_3) = -4.11 \\ x_4 = x_3 + 4 h_0 = 1.7, & f(x_4) = -5.31 \end{array}$$

$$x_5 = x_4 + 8 h_0 = 2.5, \quad f(x_5) = -6.75$$

$$x_6 = x_5 + 16 h_0 = 4.1 \quad f(x_6) = -5.79$$

至此已有三个点 x_4, x_5, x_6 ($x_4 < x_5 < x_6$) 的函数值满足两头大中间小的要求, $f(x_5) < f(x_4), f(x_5) < f(x_6)$, 极小点一定在 $[x_4, x_6]$ 上, 因此可舍去区间 $[0, 10]$ 上的其他部分, 再在 $[x_5, x_6]$ 上内插一个点 x_7 , 令 $x_7 = \frac{1}{2}(x_5 + x_6) = 3.3, x_5 < x_7 < x_6$, 且 $f(x_7) = -6.91 < f(x_5)$, 故取三点 x_5, x_6, x_7 ($x_5 < x_7 < x_6$), 它们的函数值也符合两头大中间小的要求, 因此函数的极小值必在 $[x_5, x_6]$ 上, 即在 $[2.5, 4.1]$ 上。通过不断外推内插, 极值点存在的区间可以不断缩小。非线性单变量函数的极值问题是求无约束极值及非线性规划极值的基础。

无约束条件下的多变量函数寻优可利用变量轮换法, 它将各坐标轴方向看做有利的搜索方向, 轮流按各坐标轴方向搜索最优点。给定一个出发点 x_k , 当按第 i 个坐标轴 x_i 的方向搜索时, 其他 $n-1$ 个坐标均保持给定的状态不变, 这样就将多变量寻优问题转化为一系列单变量函数的寻优问题。依次沿各坐标轴方向进行一维搜索, 直到 n 个坐标轴方向全部优选完毕, 得到一个好点 x_{k+1} , 完成一次迭代。继续这一过程, 直到所得的新点满足精度要求为止。变量轮换法的基本原理简单, 它适用于各变量间无联系, 或坐标轴方向搜索比较容易的情况。前面提到的最速下降法是目标函数值下降最快的方向, 它适用于具有一阶连续偏导数的目标函数。此外, 还有单纯形搜索法、多变量牛顿法、共轭梯度法等。各种寻优方法各有其特点和适用范围, 针对具体问题选择合适的寻优方法, 或组合使用几种方法, 以求效果最好, 效率最高, 这是我们所追求的目标。

随机规划

排队现象十分普遍,也使人心烦。上车排队、购物排队、买票排队、看病排队、交费排队、换牌照排队、车辆船只等待装卸排队,凡此种种,似乎都因拥挤造成。还有许多无形的排队现象,如打电话占线、因跑道占用飞机在机场上空盘旋等待着陆、机器发生故障等待修理、因线路不足而引起争夺卫星传送权、大型水库因水位太高等待排水、计算机终端用户等待计算……这些有形和无形的排队现象,影响着国民经济的发展和人们的生活质量。无论是有形的还是无形的排队现象,其共同特点是它们的出现是随机的,并要求接受某种服务,都是由特定功能的“服务站”和“顾客”构成。车辆、商店、售票处、医院、码头、机场、水库、电话总机、维修站等都是“服务站”,购物者、病人、打电话的人、等待着陆的飞机、水位、终端用户等都是“顾客”。因此,如何优化服务系统、减少排队现象是研究随机服务系统理论的目的。当“服务站”繁忙时,“顾客”必须排队,顾客以怎样的方式排队又是不确定的。有些顾客排队等候,有些顾客暂时离去过一会再来,有些顾客当排队一定时间后离去,也有些顾客放弃服务不再返回。研究随机服务的理论称排队论。任何服务系统都是为某种目的而建立的,有一定服务要求。服务能力太小,不能满足顾客的要求,导致服务质量差;但是服务能力过大,“服务台”数目过多,虽可避免排队,却会造成设备利用率不高而导致不必要的浪费,这是另一种经济损失。如何合理地设计和控制服务系统、优化服务系统,使它能最经济最有效地满足顾客要求是一门大学问,它是

20 世纪下半叶的研究课题之一,并成为数学规划的一个分支。由于服务系统顾客到达的随机性和服务时间的随机性,必须用随机变量来描述它,用随机工具来处理。因此,反映这类最优化的数学规划称为随机规划。随机规划有多种类型,这里主要介绍随机服务系统和随机决策。

不难发现,任何随机服务系统都由顾客到达情况、服务规则、服务机制三部分组成。顾客到达情况称为系统输入,如顾客是单个到达还是成批到达,到达的时间间隔是确定的还是不确定的,顾客之间的关系是独立的还是相关的,最常见的顾客到达方式如电话呼叫情况,据长期资料统计,人们发现时间 t 内出现 k 个电话的概率(可能性) $U_k(t)$ 可按公式

$$U_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad k = 0, 1, 2, \dots$$

计算(其中 λ 表示单位时间内顾客到达的平均数),这种输入方式称为泊松输入,或称泊松分布。服务规则是指顾客到达后按怎样方式排队及服务站的服务规则,如顾客到达后是排队等候或是立即离去(如电话占线),或是排队等候一定时间后再离去;服务的次序规则是否先到先服务,或是有优先服务规定,或定时服务等。服务机制是指服务站的数目和特性,服务台的结构是串联的还是并联的,或是混合型的,在同一时刻能为多少顾客提供服务,顾客是单个接受服务还是成批接受服务,服务时间多长。最常见的是顾客服务时间相互独立,单位时间内被服务的顾客平均数为 μ ,每个顾客平均服务时间为 $\frac{1}{\mu}$,则服务时间遵循参数为 μ ($\mu > 0$) 的指数函数 $B(x)$:

$$B(x) = \begin{cases} 1 - e^{-\mu x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

由于随机服务系统的输入方式、服务规则及服务机构都有不

同类型,因此,组成的随机服务系统类型也不同,通常用记号 $A/B/n$ 表示,其中 A 表示系统输入(顾客到达)的规律, B 表示服务时间的规则, n 表示并列的服务台数目。例如 $M/M/n$ 表示泊松输入/指数分布/ n 个服务站,其中 M 代表输入为泊松分布、服务时间为指数分布。对于 $M/M/1$ 系统,

主要服务质量指标有:服务强度 $\rho = \frac{\lambda}{\mu}$,它刻画了服务机构的

效率,即服务台忙的概率。队长 $L_s = \frac{\rho}{1-\rho}$,反映系统中顾客

数的平均值。排队长 $L_q = L_s - \rho = \frac{\rho^2}{1-\rho}$,反映等待服务的顾

客平均数。每个顾客在系统中逗留的平均时间 $W_s = \frac{L_s}{\lambda}$,顾客

排队等待的平均时间 $W_q = \frac{L_q}{\lambda}$ 。例如,某火车站售票处只有

一个售票窗口,每小时到达窗口购买火车票的顾客数遵循泊松分布,平均为 15 人。每个顾客购票时间遵循指数分布,平均为 3 分,服务台数 $n=1$ 。则平均到达率 $\lambda=15$ 人/时,平均

服务率 $\mu = \frac{60}{3} = 20$ 人/时,服务强度 $\rho = \frac{\lambda}{\mu} = 0.75$,队长

$L_s = \frac{\rho}{1-\rho} = 3$ (人),排队长 $L_q = L_s - \rho = 2.25$ (人),每个顾客

平均逗留时间 $W_s = \frac{L_s}{\lambda} = \frac{3}{15} = 0.2$ (时),即 12 分,每个顾客排

队平均等待时间 $W_q = \frac{L_q}{\lambda} = 0.15$ (时),即 9 分。此外,还可

以计算顾客到达后不需排队(即系统中没有顾客)的概率 P_0

$= 1 - \rho = 1 - 0.75 = 0.25$;队中已有一个顾客排队的概率 P_1

$= \rho P_0 = 0.1875$;队中已有两个顾客排队的概率 $P_2 = \rho P_1$

$= \rho^2 P_0 = 0.1406$;队中已有 k 个顾客排队的概率 $P_k = \rho^k P_0$,进

一步还可以计算出顾客到达后必须等待两个以上顾客的概率 $P(n > 2) = 1 - P(n \leq 2) = 1 - P_0 - P_1 - P_2 = 1 - 0.25 - 0.1875 - 0.1406 = 0.4219$ 等等。该随机服务系统每个顾客买票平均等待时间为 9 分,不算太长,车站服务台的服务强度(即设备利用率)达 75%,尚属良好。该系统的运行不算太差,旅客和车站双方都可以接受,但仍不能使旅客满意,应继续改进,以减少排队现象。

电话系统一直是排队论传统的应用场所,也是排队论的发源地和研究动力所在地。20 世纪初,美国贝尔电话公司发明了自动电话,以满足工商业发展所带来的通讯需求。自动电话系统的发明提出了一个不可回避的问题,即通话线路与用户呼叫的数量关系。要打电话方便,就得线路足够多,但线路太多会降低使用效率,造成不必要的浪费。如何处理服务质量与设备利用率这一对矛盾? 1909 年,丹麦数学家埃尔朗对此进行了研究并发表了一篇题为《概率与电话通话理论》的文章。1915 年,他在物理学家吉布斯的统计平衡概念的启发下,运用比拟的思维,借用了统计平衡概念,建立了生灭模型(称为埃尔朗公式),解决了这一著名的问题。他将一个电话呼叫的发生比作一个热分子渗透进封闭系统,而将一个电话呼叫的结束,比作一个热分子从封闭系统扩散出去。他借用热力学的统计平衡模型,建立了电话呼叫的统计平衡方程。

假设 λ 表示单位时间内电话呼叫的次数, μ 表示单位时间内通话完毕的次数, P_n 表示电话占线 n 次的概率。他得到的关系式为:

$$P_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} P_0, \quad n = 0, 1, 2, \dots$$

显然, $\sum_{n=0}^{\infty} P_n = 1$, 进而得到 $P_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} e^{-\frac{\lambda}{\mu}} = \frac{\rho^n}{n!} e^{-\rho}$, $n = 0, 1, 2, \dots$ 其中 $\rho = \frac{\lambda}{\mu}$ 是通话时间内的平均呼叫次数。这就是前面提及的泊松输入。

如果有 c 条电话线路, 当顾客到达后, 发现 c 条线路全被占用时, 就离开系统而不再接受服务, 电话公司因而失掉顾客, 这种系统称为损失制系统, 记为 $M/M/c$ 。它的数量指标公式为:

$$\rho = \frac{\lambda}{c\mu}$$

$$\begin{aligned} \text{损失率 } P_c &= \frac{(c\rho)^c}{c!} \left[\sum_{k=0}^c \frac{(c\rho)^k}{k!} \right]^{-1} \\ &= \frac{(c\rho)^c}{c!} e^{-c\rho} \left[\sum_{k=0}^c \frac{(c\rho)^k}{k!} e^{-c\rho} \right]^{-1} \end{aligned}$$

此式称为埃尔朗损失公式, 它的优点是可利用泊松分布求 P_c 。此时, 服务台(线路)平均占用个数 N 的算式为 $N = c\rho(1 - P_c)$ 。

例如, 电话订票是极为普遍的手段。假设电话订票处输入的顾客订票状态是服从泊松分布, 到达率 $\lambda = 12$ 次/时, 交谈的时间服从指数分布, 平均交谈时间为 5 分。若要使订票者打不通电话的概率小于或等于 0.05, 订票处至少应设几部电话?

若以分为时间单位, 则:

$$\lambda = \frac{12}{60} = 0.2 \text{ 次/分}, \quad \mu = \frac{1}{5} \text{ 次/分},$$

$$c\rho = \frac{\lambda}{\mu} = 1 \text{ (在这种状态下, 总能达到统计平衡状态, 不}$$

需要对 λ 、 μ 加以限制)

要求 c , 使 $P_c = \frac{1}{c!} \left[\sum_{k=0}^c \frac{1}{k!} \right]^{-1} \leq 0.05$,

查泊松分布表(参数=1), 可得

$$P_4 = \frac{0.017}{0.996} = 0.017,$$

$$P_3 = \frac{0.061}{0.981} = 0.062,$$

故 $P_4 < 0.05 < P_3$, 因此电话订票处至少要装四部电话, 才能使打不通电话的概率小于或等于 0.05。此时, 平均占用电话机的台数 N 为:

$$N = c\rho(1 - P_c) = 1 - P_c = 1 - P_4 = 0.983$$

电话系统的设计、运行依据埃尔朗公式计算通话线路的数量和损失率。埃尔朗的这一成果至今仍闪耀着异彩。

排队论在交通运输、生产线、贮存、水库可靠性等方面都有应用。近些年来, 作为连续时间的经典排队论有了发展, 对带有服务员休假、服务台可修等各种特殊机制的随机服务系统进行了深入研究, 发现了一些新的现象, 并取得新的研究成果。近 20 多年来, 计算机的应用为排队论开辟了新的理论和应用领域, 如分时系统、多程序系统、存贮器设计等。有些无法用解析方法处理的复杂系统, 借助计算机的模拟进行定量分析。

日常生活中经常遇到的另一类随机规划是随机决策, 所谓决策就是在几种不同的行动方案中作出抉择, 也是规划的一种方式。人们在工作 and 生活中遇到问题时, 由于解决的方法往往不止一种, 需要作出决策, 选择一种最好的解决方法。这些问题大可至国家的政治、经济、军事等方面的策略, 小可至工厂企业或个人家庭生活琐事。例如, 上班前天气突然转阴, 要不要带伞? 此事可能出现的状态共 4 种, 如果对各种可

能的结局给以评分如右表：

对此各人有不同的决策,有稳扎稳打思想的人认为带伞时出现最坏的结果是-0.5分,不带伞时出现最

评分 带伞 否	天气	下雨	不下雨
带伞		1	-0.5
不带伞		-1	1

坏的结果是-1分,从中选优者,决定带伞。另一种人对气象情况比较熟悉,平时注意天气变化,他先作判断,认为虽然阴天,但下雨的概率不大,假设下雨概率为0.3,不下雨的概率为0.7,若带伞,得分为 $0.3 \times 1 - 0.7 \times 0.5 = -0.05$,不带伞的得分为 $0.3 \times (-1) + 0.7 \times 1 = 0.4$,于是他决定不带伞。决策的任务是为决策者提供优化的或满意的方案,并对可能的结果进行分析。决策问题的特点是多目标性、决策影响的长期性、后果的不确定性、决策的序贯性和可供选用的决策的多样性。决策问题自古有之,齐王与田忌赛马的故事已为人们所熟悉。齐王出马有6种方案:(上、中、下)、(上、下、中)、(中、上、下)、(中、下、上)、(下、上、中)、(下、中、上)。田忌出马同样有这6种方案。双方搭配便有36种比赛格局,齐王赢得三千两黄金的格局有6种,赢得一千两黄金的格局有24种,只有6种格局齐王反输一千两黄金(见下表)。

齐王赢得矩阵

齐王策略 齐王赢得	田忌策略	b_1 (上、中、下)	b_2 (上、下、中)	b_3 (中、上、下)	b_4 (中、下、上)	b_5 (下、上、中)	b_6 (下、中、上)
a_1 (上、中、下)		3	1	1	1	-1	1

续表

齐王策略 齐王 赢 得	田忌策略	b_1 (上、中、下)	b_2 (上、下、中)	b_3 (中、上、下)	b_4 (中、下、上)	b_5 (下、上、中)	b_6 (下、中、上)
a_2 (上、下、中)		1	3	1	1	1	-1
a_3 (中、上、下)		1	-1	3	1	1	1
a_4 (中、下、上)		-1	1	1	3	1	1
a_5 (下、上、中)		1	1	-1	1	3	1
a_6 (下、中、上)		1	1	1	-1	1	3

因此,田忌赢的概率仅为 $1/6$ 。孙臆分析了对方的情况,认为齐王仍会采取(上、中、下),建议田忌采取(下、上、中)的策略,从而把握了主动权,取得胜利。

随着社会的进步,决策问题越来越复杂,一个好的决策通常要经过多次反复才能完成。特别是在现代管理中,决策更显重要,寻求最优决策是最优化管理的重要方面。决策的种类主要可分为确定型决策、风险决策、不确定型决策三种。确定型决策只存在一种确定状态,而风险型和不确定型决策中都存在两个或两个以上的不可控制状态,具有随机性。风险型决策是指在各种不可控制状态的发生概率可以预先估计的情况下作决策;而不确定型决策是指在各种状态的发生概率无法确定、无法估计的情况下作决策,是凭“主观概率”(专家头脑中的经验)作预测,误差大。风险型决策和不确定型决策虽都属随机型决策,但它们的求解途径却不

相同。

风险型决策所使用的方法是概率论及随机过程理论。它所面对的不确定状态,需要定量地给出它的概率,可通过对事情的分析,或对历史资料数据的处理,或通过数学模型计算等方式,得到随机变量的概率。按最大概率准则的基本思想,将风险型决策化为确定型决策来解决。因为概率大的事情发生的可能性也大,选取概率大的一种状态进行决策,其余的状态则忽略不管,这样就变成确定型决策。

不确定型决策问题,对各种状态的出现毫无把握或无先例可循,只能靠主观意向决策,常采用:(1)乐观法(大中取大法),即选取最大效益方案作为决策方案;(2)悲观法(小中取大法),即从各种最坏情况中,选择结局最好的方案;(3)最小最大后悔法,尽可能减小后悔,该方案是先找出各个方案的最大后悔值,然后选择最大后悔值为最小的方案,作为最优方案。

例如,工厂要确定下一年度产品的生产量,根据经验和市场调查估计,不同生产规模的销路概率及每种生产方案获利情况如下表:

批 量 销 路 概 率 (万 元)	销 路 概 率		
	好(S_1) 0.3	中(S_2) 0.5	差(S_3) 0.2
大批量(A_1)	20	12	8
中批量(A_2)	16	16	10
小批量(A_3)	12	12	12

这是对三种产品销路、三种生产方案进行的抉择,属风险型决策,目标是工厂获利最大。按最大概率准则, $P(S_2)=0.5$ 最大,于是就对 S_2 状态进行决策。在 S_2 状态下,获利最大的应是中批量生产。最大概率准则使用方便,应用较广,尤其是对某一状态发生的概率显著大于其他状态,且益损值又差不多的情况,效果较好。如果各状态的发生概率很接近,而益损值差别较大时,使用该准则的效果就不好。在这种情况下,可用益损期望值准则求解。益损期望值准则的基本思想是把每个方案的益损值看做离散型随机变量,它可能取的值就是各种状态下的益损值。求出各种方案的益损期望值,并进行比较,选取最优者。仍以前例为例,以 $E(A_i)$ 表示方案 A_i 的益损期望值,则各方案的益损期望值分别为:

方案 A_1 的益损期望值

$$E(A_1) = 20 \times 0.3 + 12 \times 0.5 + 8 \times 0.2 = 13.6$$

方案 A_2 的益损期望值

$$E(A_2) = 16 \times 0.3 + 16 \times 0.5 + 10 \times 0.2 = 14.8$$

方案 A_3 的益损期望值

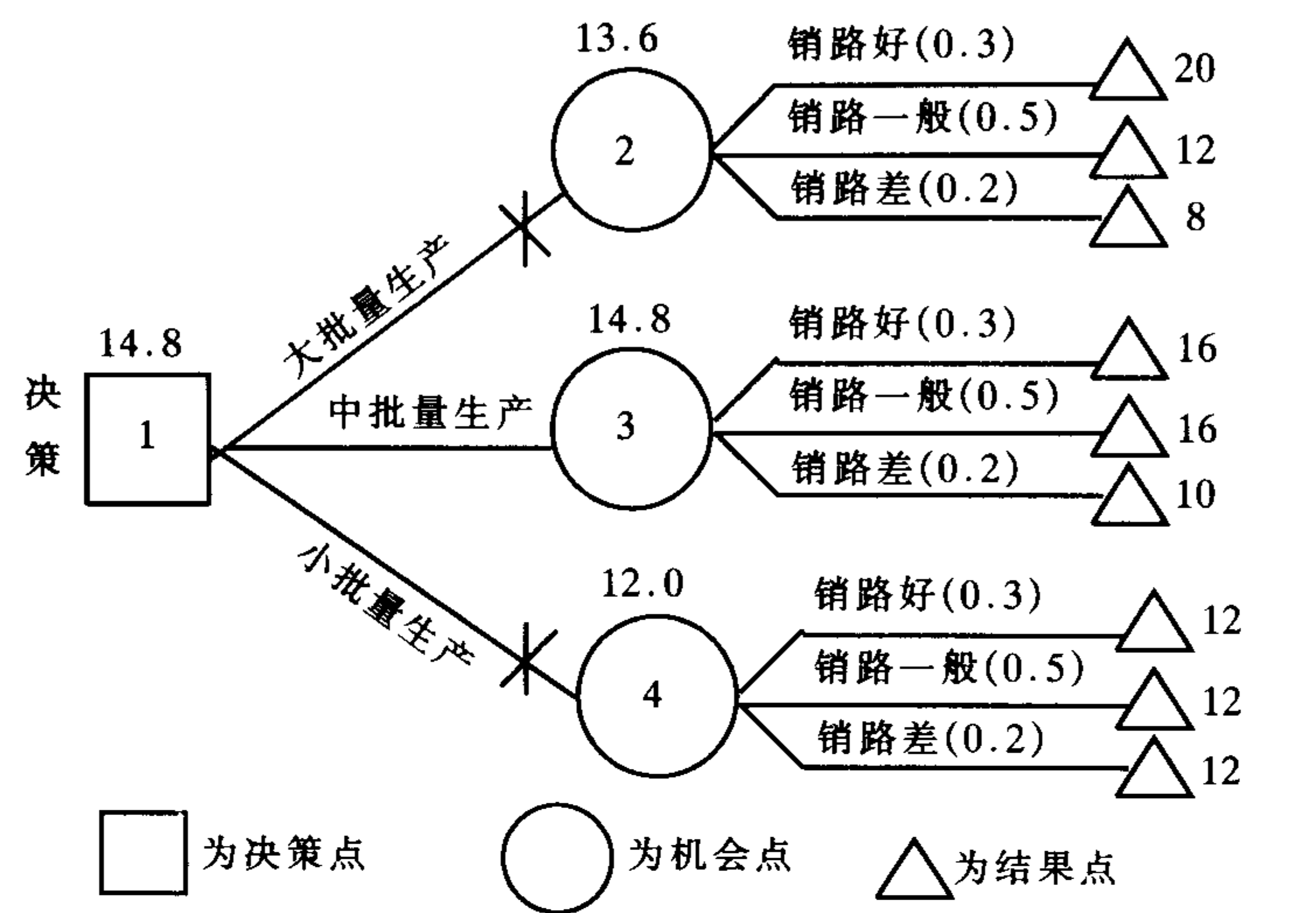
$$E(A_3) = 12 \times 0.3 + 12 \times 0.5 + 12 \times 0.2 = 12.0$$

可见益损期望值(获利期望值) $E(A_2)=14.8$ 最大,即中批量生产可使工厂获利最大。

由于益损期望值准则利用了统计规律,因此在多次决策中,成功的决策占大多数,是一种有效的决策准则。

另外,有些问题可通过矩阵运算得到最优决策方案(称矩阵法)。矩阵法为使用计算机进行决策创造了条件,对计算量大的决策问题,矩阵法更显出其优越性。有些问题可通过决策树法得到最优决策方案(称决策树法)。决策树法直观明了,可使决策者周密思考各有关因素,有顺序、有步骤地进行

决策,特别是对于复杂的多级决策问题,其优越性格外明显。
如前例的决策树法如下:



将所有决策点和机会点自左至右编号。在机会点上方标出相应方案的益损期望值,比较后选出最优的(本例中取最大)期望值,标在决策点上方,对落选的方案,在相应的直线上打“×”号。这是一个单级决策问题,一次决策便可得到最优方案。目前,随机策略的研究多集中在多目标决策和多级决策方面,相应的软件开发也日益受到重视。

对于不确定型决策问题,除了无法确定各种状态发生的概率外,其他都与风险型决策相同。所以,关键是如何确定概率。一般可采取以下几种方法进行决策:

(1) 等可能性准则。认为各种状态的概率是等可能性的。如果决策问题有几种状态 $S_1, S_2, \dots, S_n$, 不妨认为 $P(S_1) = P(S_2) = \dots = P(S_n) = \frac{1}{n}$ 。然后按风险型决策的益

损期望值准则进行决策。这个等可能性准则是法国数学家拉普拉斯首先提出的,故又称拉普拉斯准则。例如,某厂生产某种产品,有三种生产方案,其利润如下表。根据等可能性准则

有 $P(S_1) = P(S_2) = \frac{1}{2}$ 。

益损期望值为:

$$\begin{aligned} E(A_1) &= (20 - 2) \times 0.5 = 9 \\ E(A_2) &= (15 + 2) \times 0.5 = 8.5 \\ E(A_3) &= (10 + 6) \times 0.5 = 8 \end{aligned}$$

显然, A_1 是最优方案。

(2) 乐观准则。乐观准则的基本思想是对情况抱乐观态度,将每个方案 A_k 的最优益损值 $u(A_k)$ 算出后作比较,选择最优的益损值,其相应的方案 A_k 作为最优决策方案。上例中:

利 润 方 案	状 态		
		S_1 畅销	S_2 滞销
A_1		20	-2
A_2		15	2
A_3		10	6

$$\begin{aligned} u(A_1) &= \max\{20, -2\} = 20 \\ u(A_2) &= \max\{15, 2\} = 15 \\ u(A_3) &= \max\{10, 6\} = 10 \end{aligned}$$

$\max\{20, 15, 10\} = 20$, 即 A_1 为最优方案。

(3) 悲观准则。其基本思想是对客观情况抱悲观态度。如果决策目标获利最大,则把每个方案 A_K 的最小益损值 $v(A_K)$ 求出,作比较,选出最大者,则相应的方案为最优决策。如前例,按悲观准则决策如下:

$$v(A_1) = \min\{20, -2\} = -2$$

$$v(A_2) = \min\{15, 2\} = 2$$

$$v(A_3) = \min\{10, 6\} = 6$$

$\max\{-2, 2, 6\} = 6$, 即 A_3 为最优方案。

(4) 乐观悲观混合准则。该准则的特点是对客观情况的估计既不总是乐观,也不总是悲观。为此引入一个乐观系数 α ($0 \leq \alpha \leq 1$), 当 $\alpha = 1$, 即为乐观准则; 当 $\alpha = 0$, 即为悲观准则。混合准则算法是先根据乐观程度给出 α 值, 再对每个方案 A_k 计算它的益损值 $w(A_k)$:

$$w(A_k) = \alpha \max_{1 \leq j \leq n} (a_{kj}) + (1 - \alpha) \min_{1 \leq j \leq n} (a_{kj}), k = 1, 2, \dots, m,$$

其中 a_{kj} 为第 k 个方案在第 j 个状态下的益损值

然后在 $w(A_1), w(A_2), \dots, w(A_m)$ 中取得优者, 相应的方案即为最优决策方案。

对于不确定型决策问题, 采用不同的准则进行决策, 所得的结果可能不完全一致, 难以判断哪一个准则的好坏, 也没有统一的评价标准, 往往由决策者所持的看法而定。如果要使不确定型决策问题得到合理的解决, 就要设法确定各种状态发生的概率, 使之成为风险决策。

动态规划

我国民间流传一个两人玩的游戏: 两人从 1 开始轮流报数, 从 1 报到 100, 规定每人至少报一个数, 至多报 5 个连续的数, 最先报到 100 的人为胜者。人们经过分析发现, 要成为胜者必须掌握“良机”。由于对方最多可连续报 5 个数, 因此要抢到“100”就必须抢到“94”, 要抢到“94”就必须抢到“88”,

依次是 82, 76, $\cdots$, 10, 4, 这一系列“制胜点”的第一个数是 4, 谁先报到 4, 谁就能最后抢到“100”。可见第一个报数的人只要每次报“制胜点”便稳操胜券。这个“制胜点”是从结果“100”出发, 反向推理, 逐步向起点方向推进作出每一步的决策, 这种反向推理方法对取得最优结果很有效。古罗马有一个故事, 说的是古希腊有位国王要为二八妙龄的公主择婿, 不顾女儿已有心上人, 执意要通过一种传统仪式择婿。方法是先由公主在自己认为合适的求婚者中选出 10 人, 让他们围着公主站成一圈, 然后由公主根据自己的意愿挑选任何一个人作为起点, 并按顺时针方向逐个数, 当数到 17, 第 17 个人即被淘汰, 退出圈子, 然后继续从 1 开始再数到 17, 这第 17 个人又被淘汰, 如此继续下去, 直到剩下最后一个人即为女婿。公主苦苦思索, 最后一个人原先应在什么位置? 她用 10 个金币做试验, 发现最后留下的一枚金币竟是原来起数金币的第三枚。于是在择婿仪式上, 公主从“心上人”的前两位起数, 终于如愿以偿。这就是著名的“计子问题”。这种反向推理的方法就是从结果出发, 逐步往前追溯得到最优方案。因此可见, 对那些可分为不同阶段, 而每个阶段情况不断变化的动态问题, 不妨使用反向推理的方法求出最佳方案。

在工程技术、工农业生产、国民经济及军事等部门, 有许多事情由许多层次组成, 在时间上或空间上都可划分为不同阶段, 各阶段彼此相互联系, 要求在各阶段作出决策。而随着时间的推移, 事情呈运动状态, 如何在规定的标准下, 使整个决策过程达到最好的结果, 这就是多阶段决策过程的最优化问题, 这类最优化问题称为动态规划问题。例如, 最短路线问题是典型的动态规划问题。图 82 是一个路线网, 网上有 $A, B_1, B_2, B_3, C_1, C_2, C_3, D_1, D_2$ 和 E 10 个地点、通路和

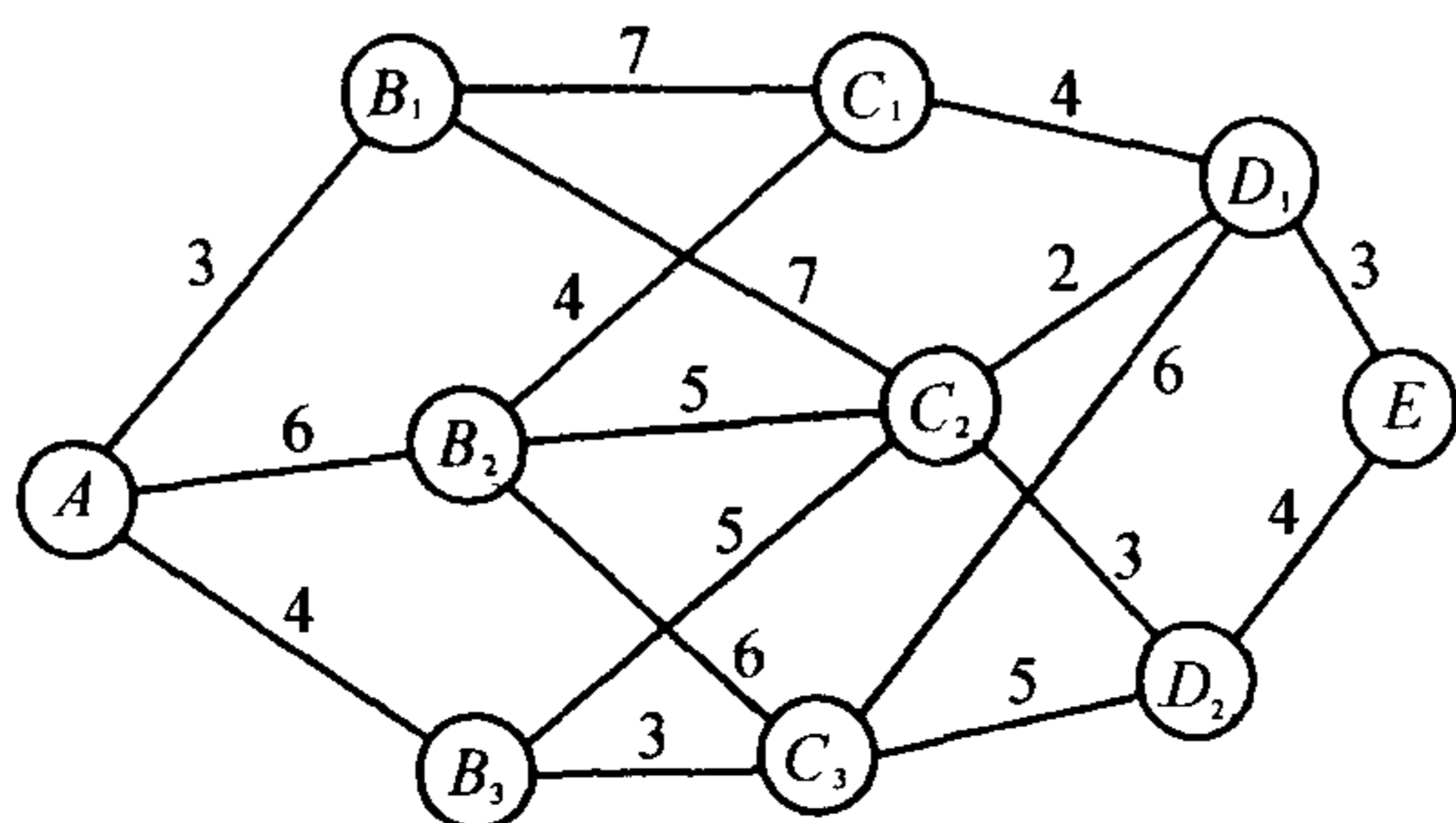


图 82

两地间的距离。如果要从 A 到 E 铺设地下管道,应选择那条路线,可使管道总长度最短? 这是一个多阶段的决策过程最优化问题。从 A 出发,必须经过 B 段、 C 段、 D 段才能到达 E 。它可分为四个决策阶段:第一阶段为 $A \rightarrow B$,第二阶段为 $B \rightarrow C$,第三阶段为 $C \rightarrow D$,第四阶段为 $D \rightarrow E$ 。各阶段可选择不同的路线铺设地下管道,不同路线的管道总长度也不同。如果使用穷举法,从 A 到 E 共有 12 条不同的路线,将这 12 条路线一一列出,分别计算其长度,加以比较,便可得出最优解:从 A 经 B_3, C_2, D_1 到 E 为最短路线,总长度为 14。所谓最短路线,是指各条路线中总长度最短的路线。最短路线并不意味着每一阶段的路线都是最短。如本例中最短路线 $A-B_3-C_2-D_1-E$,它的第一阶段 A 到 B_3 比从 A 到 B_1 长,而 $A-B_1$ 却不在所选的最短路线中。当阶段很多,每个阶段上又有很多的选择时,用穷举法求最短路线十分麻烦。若用动态规划方法求解,就较为方便。动态规划的方法是将全过程分为 n 个阶段,上例中 $n=4$ 。第 k 阶段的起点 x 到 E 的最优(最短)路长记为 $f_k(x)$ ($k=1,2,3,4$)。最后阶段($k=4$)只有 D_1, D_2 到终点 E 的路各一条,它们的路长分别为

$d_4(D_1, E) = 3, d_4(D_2, E) = 4$, 故 $f_4(D_1) = d_4(D_1, E) = 3$,
 $f_4(D_2) = d_4(D_2, E) = 4$ 。

向前推移, 考虑第三阶段情况, 即 $k = 3$ 时, 该阶段的出发点有 C_1, C_2, C_3 , 它们到 E 的可能路线的最短路长分别为:

$$f_3(C_1) = d_3(C_1, D_1) + f_4(D_1) = 4 + 3 = 7,$$

仅 1 条最短路 $C_1 \rightarrow D_1 \rightarrow E$;

$$f_3(C_2) = \min \begin{cases} d_3(C_1, D_1) + f_4(D_1) = 2 + 3 = 5 \\ d_3(C_2, D_2) + f_4(D_2) = 3 + 4 = 7 \end{cases} = 5,$$

只有 1 条最短路 $C_2 \rightarrow D_1 \rightarrow E$;

$$f_3(C_3) = \min \begin{cases} d_3(C_3, D_1) + f_4(D_1) = 6 + 3 = 9 \\ d_3(C_3, D_2) + f_4(D_2) = 5 + 4 = 9 \end{cases} = 9,$$

有 2 条最短路线 $C_3 \rightarrow D_1 \rightarrow E$ 和 $C_3 \rightarrow D_2 \rightarrow E$ 。

再向前推移, 考虑第二阶段情况。 $k = 2$ 时, 该阶段上的出发点有 B_1, B_2, B_3 三个点, 类似地可求出 B_i 到 E 的最短路线:

$$f_2(B_1) = \min \begin{cases} d_2(B_1, C_1) + f_3(C_1) = 7 + 7 = 14 \\ d_2(B_1, C_2) + f_3(C_2) = 7 + 5 = 12 \end{cases} = 12,$$

有 1 条最短路线 $B_1 \rightarrow C_2 \rightarrow D_1 \rightarrow E$;

$$f_2(B_2) = \min \begin{cases} d_2(B_2, C_1) + f_3(C_1) = 4 + 7 = 11 \\ d_2(B_2, C_2) + f_3(C_2) = 5 + 5 = 10 \\ d_2(B_2, C_3) + f_3(C_3) = 6 + 9 = 15 \end{cases} = 10,$$

有 1 条最短路线 $B_2 \rightarrow C_2 \rightarrow D_1 \rightarrow E$;

$$f_2(B_3) = \min \begin{cases} d_2(B_3, C_2) + f_3(C_2) = 5 + 5 = 10 \\ d_2(B_3, C_3) + f_3(C_3) = 3 + 9 = 12 \end{cases} = 10,$$

有 1 条最短路线 $B_3 \rightarrow C_2 \rightarrow D_1 \rightarrow E$ 。

继续向前推移, 考虑第一阶段情况。 $k = 1$ 时, 该阶段上的出发点仅 A 一点, 类似地可考虑 A 到 E 的最短路线:

$$f_1(A) = \min \begin{cases} d_1(A, B_1) + f_2(B_1) = 3 + 12 = 15 \\ d_1(A, B_2) + f_2(B_2) = 6 + 10 = 16 \\ d_1(A, B_3) + f_2(B_3) = 4 + 10 = 14 \end{cases} = 14,$$

有一条最短路线 $A \rightarrow B_3 \rightarrow C_2 \rightarrow D_1 \rightarrow E$ 。

该问题的最优解即为 $A \rightarrow B_3 \rightarrow C_2 \rightarrow D_1 \rightarrow E$, 其长度为 14, 这一结果与穷举法的结果一致。

动态规划方法的基本思想是将一个多阶段决策过程的最优化问题看做一个反向的连续递推过程。从终点开始, 由后向前逐段地向起点方向寻找最优策略。在第 k 阶段求第 k 个子过程的最优策略时, 要利用第 $k+1$ 个子过程的最优化结果, 决策过程随空间(或时间)逐步推移, 对问题的每个阶段进行取优。动态规划的最优化原理, 不仅要求目标可分, 而且要求决策过程状态可分。它包含着对多阶段决策过程中的状态的特殊要求, 要求前面各阶段的状态不影响后面过程的发展。也就是说, 如果某阶段的状态已给定, 则在这一阶段以后过程的发展不受前面各阶段状态的影响(称无后效性), 这是动态规划方法求最优策略的基本要求。否则, 就不能用动态规划方法求解, 如著名的货郎担问题。货郎必须去 n 个城镇售货, 每个城镇去一次, 且不许重复, 最后回到原来的出发地。由于在某一阶段, 他后面的路线与到达该城镇以前的走法有关, 已去过的城镇不能再去, 前面各阶段的状态影响着后面的过程, 这时就不能用动态规划方法求解。

动态规划问题既可以是空间上划分阶段, 也可以是时间上划分阶段, 如机器负荷分配问题。某厂 5 年计划初期拥有机器 n 台, 机器可以在高负荷和低负荷两种状态下工作。每年年初要重新分配完好的机器在高、低两种不同负荷下进行生产, 使 5 年内产品的总产量最大。机器在高负荷下生产, 产

量虽高,但损坏也严重,因此,将所有完好的机器都投入高负荷生产,不一定能使总产量最大。由于每年年初重新分配一次机器,故可分为 5 个阶段,每年年初为一个阶段开始,必须求出经前一年生产后还有多少完好的机器,并确定在高、低两种负荷下生产的机器台数。要选择最优方案,使 5 年内的总产量达到最大,这是多阶段决策最优问题。该问题无法用穷举法求解,因为它可以是非整数解(例如,0.5 台可以理解为这台机器在高、低负荷下各生产半年),它的变量可以视为连续变量,决策方案有无限多个,穷举不尽,只能用动态规划方法求解。

动态规划不仅可以用来解决离散的或连续的,确定性的或随机性的决策过程的最优化问题,也可用于某些非时序最优化问题。它不仅可以解决有时间或空间层次的动态问题,还可以解决没有时间或空间因素的“静态问题”。它在系统工程、自动控制、资源分配、生产计划、存贮问题、复合系统的可靠性问题、设备更新问题、排序问题等领域都有广泛的应用。有些静态最优化问题用线性规划或非线性规划求解,但由于结构特殊,往往可以看成是一个多阶段决策的最优化问题,利用动态规划的方法求解显得更简单更容易。例如,复合系统可靠性问题也可用动态规划方法解决。某工作系统由 A, B, C 三个部件串联而成,三个部件的工作互相独立。各部件正常工作的概率 A 为 0.7, B 为 0.8, C 为 0.6。现有购买部件的经费 10 万元,各部件的单价是 A 为 2 万元, B 为 3 万元, C 为 1 万元。每个部件都应考虑备件,三个部件应各购买多少个才能使系统正常工作的概率达到最大? 如果将问题看做多阶段决策过程的最优化问题,就可用动态规划的方法解决。整个决策过程可分成三个阶段,第一阶段对部件 A 的数量作

决策,第二阶段对部件 B 的数量作决策,第三阶段对部件 C 的数量作决策。决策变量 U_k (第 k 个部件的数量)只能取非负整数, $k=1,2,3$ 。计算结果最优策略为:购买 A 部件 2 个, B 部件 1 个, C 部件 3 个,系统的可靠性程度可以达到 0.682。

在实践活动中,常会碰到多阶段不确定型决策问题,它是一种带有随机性的多阶段决策问题,称为多阶段的随机决策问题,是动态规划的另一分支。例如,某部门打算在 5 周内采购一批原料,估计未来 5 周内的价格可能有三种,每种价格的概率如右表,未来的 5 周中,每一周应按什么价格采购才是最佳决策(总采购费最小)?

价格	450	470	500
概率	0.25	0.35	0.40

这个随机性的多阶段决策问题可化为动态规划问题求解。以一周为一个阶段,共 5 个阶段,第 k 阶段即为第 k 周, $k=1,2,3,4,5$, x_k 表示第 k 周原料的价格, $f_k(x_k)$ 表示当假定第 $k-1$ 周末采购、第 k 周内价格为 x_n 时,按最优策略进行采购时所花的最低价钱。从最后阶段往前推移,先考虑 $k=5$ (即第五周)的情况,若前四周未采购,则必须在第五周采购,无选择余地地按价购进,它的费用为:

$$f_5(x_5)=\begin{cases} 450 & \text{当 } x_5=450 \\ 470 & \text{当 } x_5=470 \\ 500 & \text{当 } x_5=500 \end{cases}$$

现考虑 $k=4$,即第四周的情况。假定前三周末采购,第四周可能采购,也可能不采购,究竟是否采购要看第五周采购的费用期望值是多。第五周采购的期望值为:

$$\begin{aligned} &0.25 \times f_5(450) + 0.35 \times f_5(470) + 0.40 \times f_5(500) \\ &= 0.25 \times 450 + 0.35 \times 470 + 0.40 \times 500 = 477 \end{aligned}$$

因此,第四周的采购费用:

$$f_4(x_4) = \min(x_4, 477) = \begin{cases} 450 & \text{当 } x_4 = 450 \\ 470 & \text{当 } x_4 = 470 \\ 477 & \text{当 } x_4 = 500 \end{cases}$$

这说明,在第四周若价格为 450 或 470 时,就应采购,若价格为 500 时,不应采购。现再考虑第三周的情况。假定前两周未采购,用类似方法计算后两周采购的费用期望值为:

$$\begin{aligned} & 0.25 \times f_4(450) + 0.35 \times f_4(470) + 0.4 \times f_4(500) \\ &= 0.25 \times 450 + 0.35 \times 470 + 0.40 \times 477 = 467.80 \end{aligned}$$

$$f_3(x_3) = \min(x_3, 467.8) = \begin{cases} 450 & \text{当 } x_3 = 450 \\ 467.8 & \text{当 } x_3 = 470 \text{ 或 } 500 \end{cases}$$

这说明,若第三周的价格为 450 就采购,若价格为 470 或 500 时,就不采购。类似地考虑第二周的情况。先计算以后几周的最优策略采购的费用期望值为:

$$\begin{aligned} & 0.25 \times f_3(450) + 0.35 \times f_3(470) + 0.40 \times f_3(500) \\ &= 0.25 \times 450 + 0.35 \times 467.8 + 0.40 \times 467.8 \\ &= 463.35 \end{aligned}$$

$$\begin{aligned} f_2(x_2) &= \min(x_2, 463.35) \\ &= \begin{cases} 450 & \text{当 } x_2 = 450 \\ 463.35 & \text{当 } x_2 = 470 \text{ 或 } 500 \end{cases} \end{aligned}$$

这说明,若第二周的价格为 450 时就采购,否则就不采购。最后考虑第一周的情况,第一周是否采购要看后几周按最优决策的费用期望值:

$$\begin{aligned} & 0.25 \times f_2(450) + 0.35 \times f_2(470) + 0.40 \times f_2(500) \\ &= 0.25 \times 450 + 0.35 \times 463.35 + 0.4 \times 463.35 \\ &= 460.01 \end{aligned}$$

$$f_1(x_1) = \min(x_1, 460.01)$$

$$= \begin{cases} 450 & \text{当 } x_1 = 450 \\ 460.01 & \text{当 } x_1 = 470 \text{ 或 } 500 \end{cases}$$

这说明,只有当第一周的价格为 450 时才采购,否则就不采购。

综上所述,最优采购方案是:当第一、第二、第三周的价格为 450 时才采购,否则不采购;在第四周内,若市场价格为 450 或 470 时就应采购,否则不采购;到第五周时,只有一种方案,即按市场价采购。

在国民经济的一些领域,如城市规划、交通运输、人口增长、土地与水资源利用、工业和农业生产等问题常呈现时间和空间上的阶段性和周期性,对这类问题可利用动态规划结合其他数学方法建立数学模型进行处理,取得优化方案。如城市交通的分配虽是静态的,而实际上同一道路区间的交通量在不同时段变化很大。运用动态规划进行描述和处理已有许多成功的例子,如县级经济作为国民经济结构的基本单位是十分重要的,它是一个庞大、多层次、开放的复杂系统,连锁反应强,处于动态变化过程。四川省曾对丘陵大县的经济发展作了系统研究和数量分析,建立了由三个模型群(功能分析预前模型群、决策分析模型群、子系统结构功能分析与结构优化模型群)组成的模型体系,每个模型群中又包含若干子模型,既有多目标规划,又有动态规划,通过这些模型的求解,设计了县级经济发展的总体规划。从这些事例可见,动态规划能很好地解决一些较复杂的实际问题。对于阶段数目不定或无限离散决策过程的动态规划可用函数迭代法或策略迭代法,这两种方法都是从选“好”的指标最优函数(或初始策略)出发,利用动态规划的函数方程进行迭代,得出一个指标最优的函数序列(或策略序列),使这一序列收敛于最优解。求解动

态规划还有一种数值方法,称微分动态规划法,它是将函数空间的牛顿方法思想与动态规划方法的基本思想结合起来,既可求解离散决策过程的最优化问题,也可用来求解连续决策过程的最优化问题。从变分法角度看,它是一种二阶变分法。

多目标规划

人们在寻优过程中要考虑的目标往往有好几个。择偶要求才貌双全,是以才与貌作为主要目标,其实还有次要目标,如体质、家庭状况、职业等。有些人购物要求价廉物美,以价格与质量作为主要目标,而将维修、品牌等放在次要位置,而有些人则宁愿花高价买名牌。一个河道水利开发工程要考虑发电、防洪、灌溉、航运、养殖、环保、旅游等诸多目标,必须根据实际情况确定哪些是主要目标,哪些是次要目标,对次要目标稍为“忽视”,列为限制条件。一个企业所追求的目标有产量、质量、销售量、利润、资产增值率等诸多因素。这些问题中的多个目标往往不能兼得,甚至有些目标间还可能互相矛盾。因此,要根据实际情况抓住几个主要目标,而次要目标则稍加“怠慢”,降为限制条件,以保证主要目标达到最优水平。而主要目标的重要性也是不尽相同的,如何使综合效应最佳,这就是多目标优化要解决的。显然,多目标问题要比单目标问题复杂得多。

机械设计中的几个设计指标要同时都达到最优,则需建立多个目标函数。例如,齿轮变速箱的设计通常有下列几方面的要求:

- (1) 各传动轴间的轴心距尽可能小,使变速箱结构紧凑,

体积较小；

(2) 各齿轮体积总和尽可能小,使材料耗费少,以降低成本；

(3) 齿轮的最大圆周速度尽可能低,使变速箱运转噪声小；

(4) 传动效率尽可能高,以节省能源。

按照这些要求,可分别建立目标函数 $f_1(x)$, $f_2(x)$, $f_3(x)$, $f_4(x)$ 。要求这几个目标函数都达到最优,这是一个多目标规划：

$$\min f(x) = \min[f_1(x), f_2(x), f_3(x), f_4(x)]。$$

最优化问题中的目标函数多于一个的数学规划称为多目标规划,或称为多指标最优化、多准则最优化。

多目标规划中,目标函数 $f_1(x)$, $f_2(x)$, $\dots$, $f_p(x)$ 的度量单位各不相同,它们的量纲不同,因此,对带量纲的目标函数要作规范化处理。处理方法是:将带量纲的目标函数记为 $\bar{f}_i(x)$,作函数

$$f_i(x) = \frac{\bar{f}_i(x)}{\bar{d}_i}, \text{ 其中 } \bar{d}_i = \left| \min_{x \in R} \bar{f}_i(x) \right|, \quad (i = 1, 2, \dots, p)$$

则 $f_i(x)$ 便是不带量纲的。如果多目标函数规划问题中的目标函数已作规范化处理,那么它的标准形式为：

$$V-\min[f_1(x), f_2(x), \dots, f_p(x)], x \in R$$

$$\text{约束条件: } g_i(x) \geq 0 \quad (i = 1, 2, \dots, m)$$

“V—min”表示对各个 $f_j(x)$ ($j = 1, 2, \dots, p$) 都越小越好,以此符号区别于单目标规划。如果目标函数 $f_j(x)$ 和约束函数 $g_i(x)$ 均为线性函数时,就称为多目标线性规划问题。如果 $f_j(x)$ 和 $g_i(x)$ 中有一个或多个为非线性函数时,就称此规划

问题为多目标非线性规划问题。

多目标规划由于目标函数重要性不同,甚至互相有冲突,往往不存在绝对最优解(即对每个目标都是最优的解),而且解与解之间不易直接比较好坏,没有一个明确而合适的标准。因此,需要按目标的重要性排定等级,确定各目标值,给出权数,将多目标规划转化为单目标规划。通常用构造评价函数的方法将多目标规划转化为单目标规划。构造评价函数的常用方法有线性加权和法、平方加权法、极大极小法、乘除法等。线性加权和法是根据各目标的重要程度给出一组权数,每个目标 $f_i(x)$ 给一个权数 $\lambda_i \geq 0, i = 1, 2, \dots, p$, 且 $\sum_{i=1}^p \lambda_i = 1$ 。将原问题转化为单目标规划问题:

$$\min \sum_{i=1}^p \lambda_i f_i(x)$$

$$\text{约束条件: } g_j(x) \geq 0 (j = 1, 2, \dots, m)$$

平方和加权法则体现“自报公议”原则,即先对每个目标函数确定一个最优值的下界 $f_i(0)$, $\min_{x \in R} f_i(x) \geq f_i(0), i = 1, 2, \dots, p$ (“自报”),然后给出一组权数 $\lambda_1, \lambda_2, \dots, \lambda_p, \lambda_i \geq 0$,

$$\sum_{i=1}^p \lambda_i = 1 \text{ (“公议”)}$$

再将多目标规划转化为单目标规划:

$$\min \sum_{i=1}^p \lambda_i [f_i(x) - f_i(0)]^2$$

$$\text{约束条件 } g_j(x) \geq 0 (j = 1, 2, \dots, m)$$

极大极小法(min—max 法)基于“在最不利情况下找出一个最有利的策略”的考虑。将多目标规划转化为单目标规划:

$$\min_{x \in R} [\max_{1 \leq i \leq p} \lambda_i f_i(x)], \text{ 其中 } \lambda_i (\geq 0) \text{ 为权数}$$

$$\sum_{i=1}^p \lambda_i = 1$$

约束条件: $g_j(x) \geq 0 (j = 1, 2, \dots, m)$

乘除法是把求最大(max)的目标函数彼此相乘作分子,把求最小(min)的目标函数彼此相乘作分母,以这样的乘除形式的函数作为总评价目标函数,求最大。

以上这些方法都是日常工作中使用很广的方法,在企业和各管理部门已较普遍运用。还可以用别的方法构造出其他类型的评价函数。在构造评价函数时,如何确定一组合理的权数,准确反映各目标的重要程度是比较困难的,通常是凭经验直接给定,或按某种规定算出。

遇到多目标、多准则、多方案的复杂问题,可根据其目标、准则与方案的相互关系,建立递阶层次结构,先确定各方案相对于子目标的重要性的权数,加以综合处理。此外,还可用分层序列法将多个目标依其重要程度逐一排队,加以协调综合处理。对于互相矛盾的目标,采取适当匹配增加满意程度,确定最优解。对多目标规划问题,要得到绝对最优解的可能性是很小的,尤其是各分目标的优化互相矛盾时更难得到,只有求得非劣解或弱非劣解时才有意义。20世纪70年代开始,多目标规划的研究越来越受到人们的重视,它与经济学、决策分析、系统理论、控制论以及对策论等有密切的联系,在实际问题中有广泛的应用。虽然国内外对多目标问题进行了许多研究,提出了一些解决方法,但比起单目标优化问题,在理论上和算法上还不完善,也不系统。

组 合 优 化

在日常生活、工作和建设上常会遇到路线问题。一个货郎要到若干城镇售货,如何选择路线,使每个城镇恰好经过一次,最后回到出发地而使总路程最短;一个医疗队下乡,一个艺术团到基层慰问演出,一个慰问团到灾区各居民点送温暖,要求每个居民点都恰好经过一次,最后回到出发地,如何选择路线才能使总行程最短;由于战略需要,部队要撤出某地区,离开前必须炸掉几座主要桥梁以阻挡敌人,要求走过一座桥后立即炸掉该桥,而使总路程最短;在城市管道、电线铺设上经常采用没有回路的树形结构,如何架设电话线使任何两地能直接通话,而又要使架设的总路线或铺设费用最小。以上各例都是求最短路线的。

在运输问题上不仅要求路程短,还要使运量最大。例如要从几个城市通过铁路调运物资到灾区几个居民点,由于各段铁路线的运输总量是有限的,如何选择运输方案使总路程最短而运送的物资总量最大。无论是最短路问题还是最大流量问题,若用点表示地点,用点与点的连线表示两地的线路,这些问题都可由点与线构成网络图来描述。网络图能反映所要解决的问题中各方面的关系,如输入与输出的因果关系、上下级关系、时间先后的关系、数量多少关系等。通过网络图了解和控制整个结构和功能,便于从总体角度考虑问题,达到整体最优,即将寻求最优化的实际问题转化为网络最优化问题。网络最优化的方法很多,不同的方法适应不同性质的具体问题。由于网络理论不需要高深的数学基础,且较直观,易被人

们所掌握。近 20 年来,网络优化的理论和方法迅速发展,成为“图论”的一个分支,它在信息论、控制论、工程技术、交通运输、化学键、社会集团结构等众多领域已广泛应用。工程和生产中的电网络、交通网络、运输网络、工序网络等知识已为广大工程技术人员和管理人员所掌握。他们对常用的网络优化方法已轻车熟路,利用网络知识简化复杂过程,得到最优方案,取得了巨大的经济效益。不难发现,利用网络解决的最优化问题,它的变量可以是离散的,求解过程的算法特点是走一步算一步,每步计算时,点与路线都按某种规则在更迭,逐步迭代直至得到最优方案,每步计算结果都是点与路线组合的结果。这种特点在线性规划、整数规划、0—1 规划、动态规划及搜索技术等方法上已有呈现,这类问题的数学规划可称为组合规划。它的算法繁多,直观有趣,前面介绍过的如 0.618 法、分数法、正交拉丁方法、单纯形法、分支定界法、割平面法等都体现了搜索的思想,前进时一旦发现此路“不通”(不比原来好),就立即回头,改换方向,直到满意为止。这种求解思路用于一些游戏性的问题十分有效。例如,要在一个每行每列均为 4 个方格的棋盘上布 4 个棋子,要求任意两个棋子不在同一行和同一列上,也不在对角线上,求符合这种布局的方案数。再如,一个有 33 个方格的棋盘(图 83),开始时仅空着中央一个方格,棋子在棋盘上移动的规则是:可以跨越水平或垂直方向的相邻棋子进入另一侧相邻的空格,并“吃掉”该相邻的棋子。要求找到一种步骤,使得全盘只剩下一个棋子留在棋盘中央(即第 17 格子)。这类问题若用列举法一一检验十分困难,而用组合优化的搜索技术就很容易解决。搜索技术在实际应用中已取得不少令人瞩目的成果,如 20 世纪 60 年代,美国寻找在大西洋失踪的核潜艇“打谷号”和“蝎子号”以

及在地中海失踪的氢弹,都是应用搜索技术取得成功的。

魔方是一种 20 世纪 70 年代问世,80 年代在我国风靡一时的玩具。它是由 26 个棱长为 1.9 厘米的小立方体构成(图 84),这些小立方体能自由地绕一个同样大小的中心块转动,边块和角块可以分别转至任何其他边块和角块的位置。为了能看清这些小方块移动的位置,

		1	2	3		
		4	5	6		
7	8	9	10	11	12	13
14	15	16		18	19	20
21	22	23	24	25	26	27
		28	29	30		
		31	32	33		

图 83

这些小方块表面涂有不同的颜色。玩魔方的基本要求是当各个面上的颜色打乱以后,用尽可能少的动作使之恢复原状,这要在约 300 多亿种可能的变化中寻找满意的方案,谈何容易。

2000 多年前中国玩具“六通”的神奇不亚于魔方。传说春秋战国时期的能工巧匠鲁班,为了测试儿子的智力,做了一个由六块大小一样、中段有不同镂空的正四棱柱形木块组成的“六通”(图 85)。一天晚上,鲁班当着儿子的面拆了“六通”,要求儿子在次日黎明前将它组装好。这是要在约 300 万种不同的组合方案中寻找一种成功的方案,亦非易事。这类智力玩具在国内外流传甚多。我国民间流传的三国棋,借喻“三国演义”中“智算华容”,关羽放走曹操的故事而设计的(图

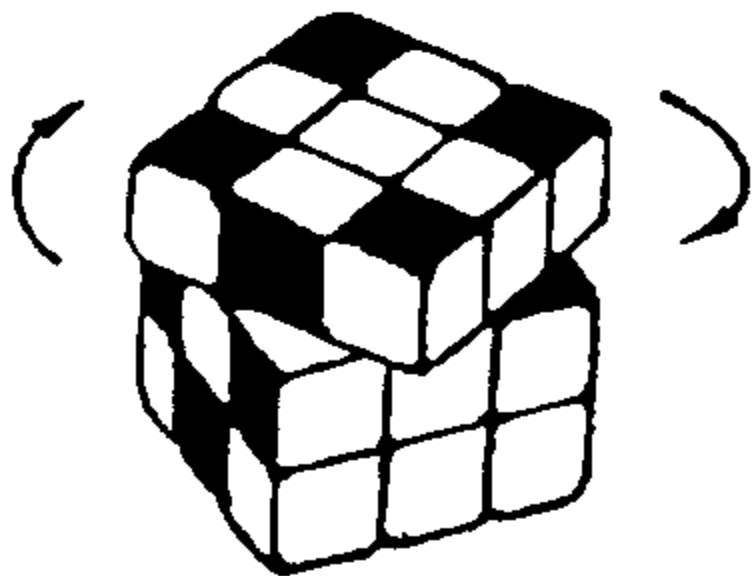


图 84

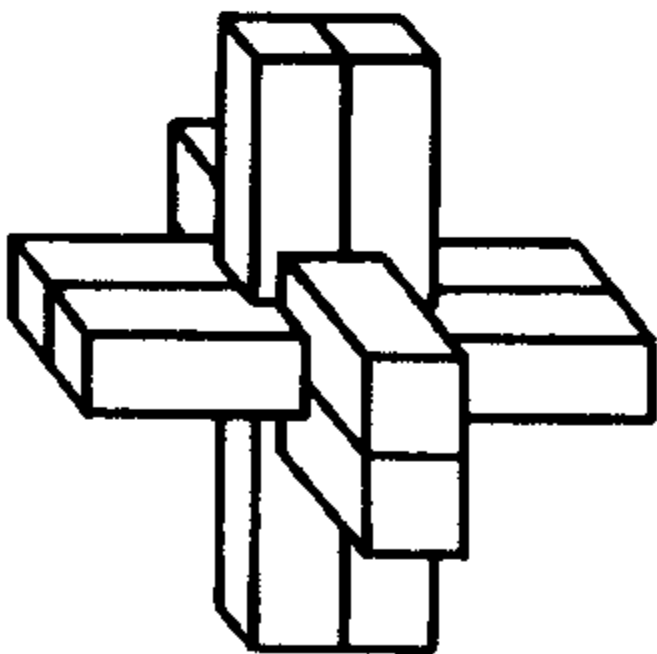


图 85

86)。游戏要求不离开盒子,移动盒中的矩形块,最后使曹操从缺口处逃出。曹操和关羽各应走什么路线才能以最少的步数使曹操能逃命? 最少的步数是多少?

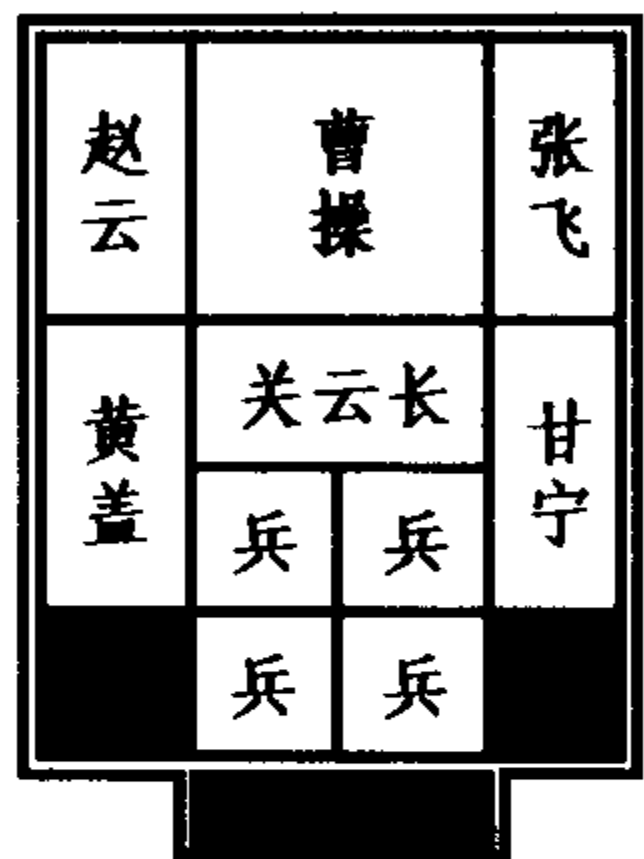


图 86

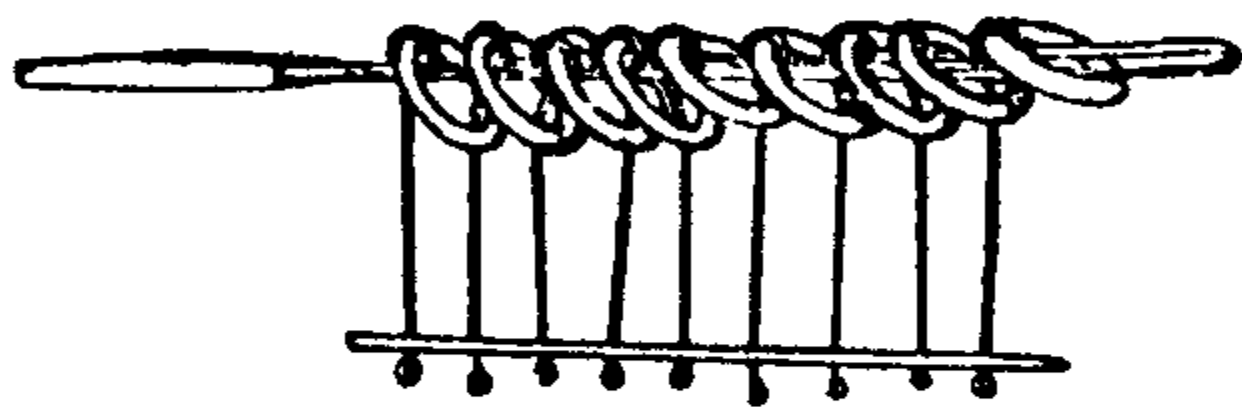


图 87

老年人大概曾玩过我们祖先流传下来的一种叫做“九连环”的玩具,它由 9 个相同的金属环、9 条连杆,一根活动柄和一个剑形框套组成(图 87),玩的目标是要将框套上的 9 个金属环脱下来(或将已与框套脱开的金属环重新套在框套上),至少要多少次动作才能成功? 这类求最优方案的组合玩具在古代就有,而它的理论沉睡久远,只是到了 20 世纪最后 20 年才被人们重视,得以发展,成为世人瞩目的新学科。

组合优化有赖于组合算法,组合数学的发展不断为组合优化提供新的算法。组合数学的起源可追溯到公元前 22 世纪,我国古代传说中的大禹在治水时,在洛水里浮现一只大龟,龟背上刻有一个 3 阶幻方(现称“洛书”),纵的 3 列、横的 3 行及两对角线上 3 个数字的和都等于 15。“洛书”传到印度后成为吉祥的象征,至今仍有许多印度少女将“洛书”图样背在胸前,作为护身符。由于“洛书”共有 9 个数,被汉代徐岳称为“九宫算”。幻方之“神秘”是变幻莫测、奥妙无穷的(图 88)。

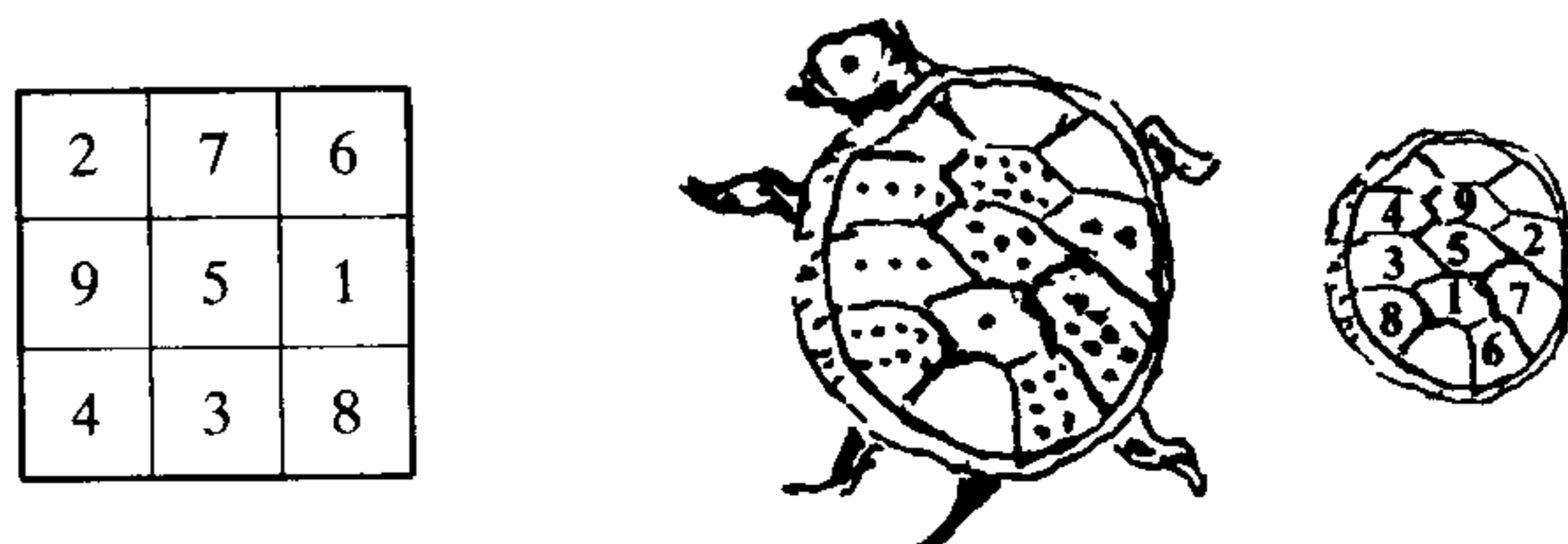


图 88

1977 年,“旅行者”号宇宙飞船上天时,带去“地球之音”,其中有包括八达岭长城雄姿在内的 115 张照片、35 种自然音响、60 种语言问候语和 27 支世界名曲。华罗庚教授则认为,为了沟通不同星球的信息,最好带两个图形,一个是“洛书”图,一个是勾股定理图(图 89)。

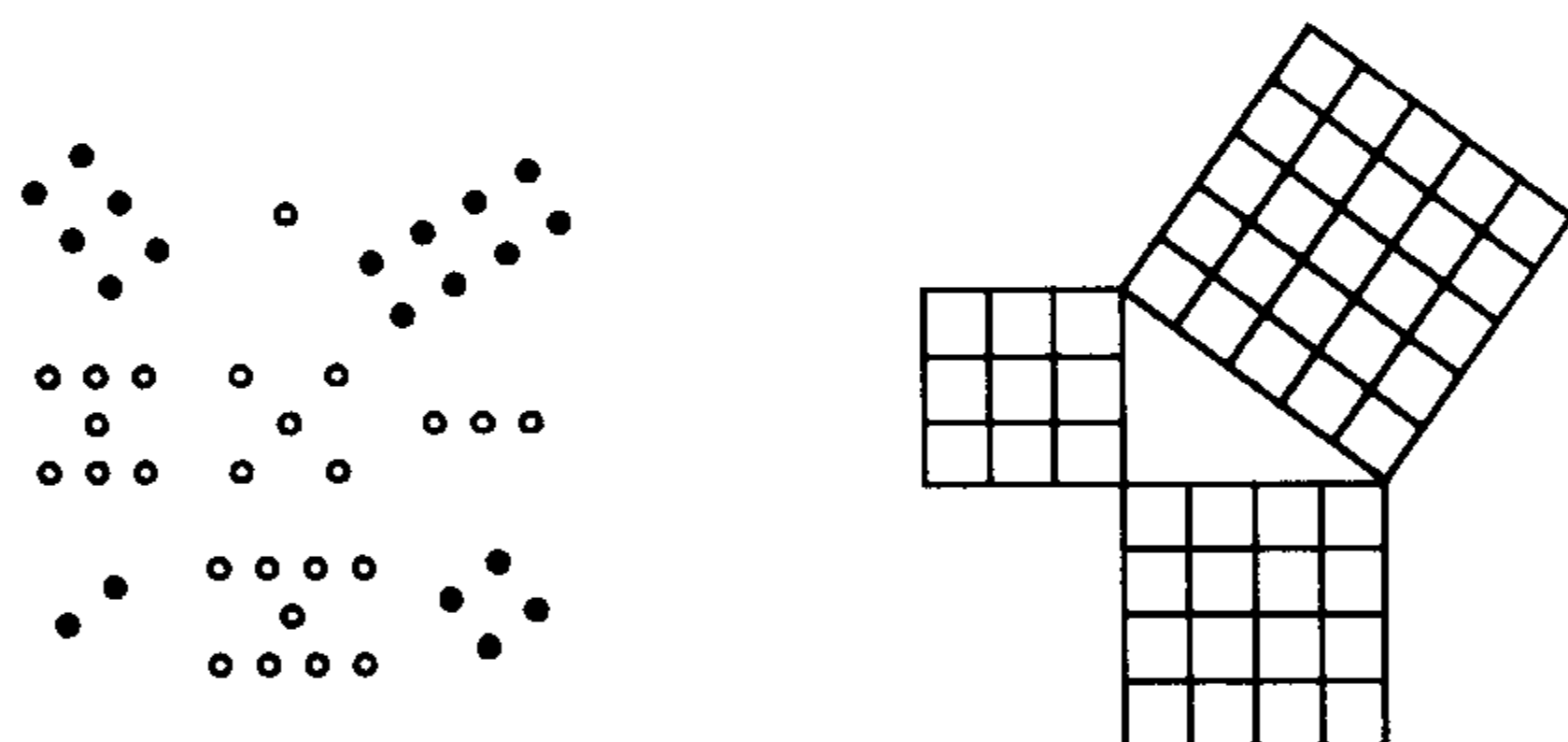


图 89

组合数学中的许多著名问题已广为流传,如哥尼斯堡七桥问题、哈密尔顿回路问题、兔子繁殖问题、夫妻问题、四色问题、鸽笼问题、36 个军官问题、寇克满 15 个女生问题等等。这些问题不仅趣味无穷,而且各代表一类实际问题。电子计算机的发展和应用使组合数学(尤其是图论)犹如一片被开垦的处女地,焕发着青春活力,异军突起,其势不亚于 20 世纪 50 年代的计算方法。组合数学不仅自身崛起,而且在理论上

与数论、代数学、概率与统计、函数论等有密切关系。在应用方面,已渗透到国防工业、物理、化学、生物、计算机科学、空间技术、信息编码、物质结构、遗传工程、实验设计和人工智能等领域,真是“后生可畏,前途无量”!