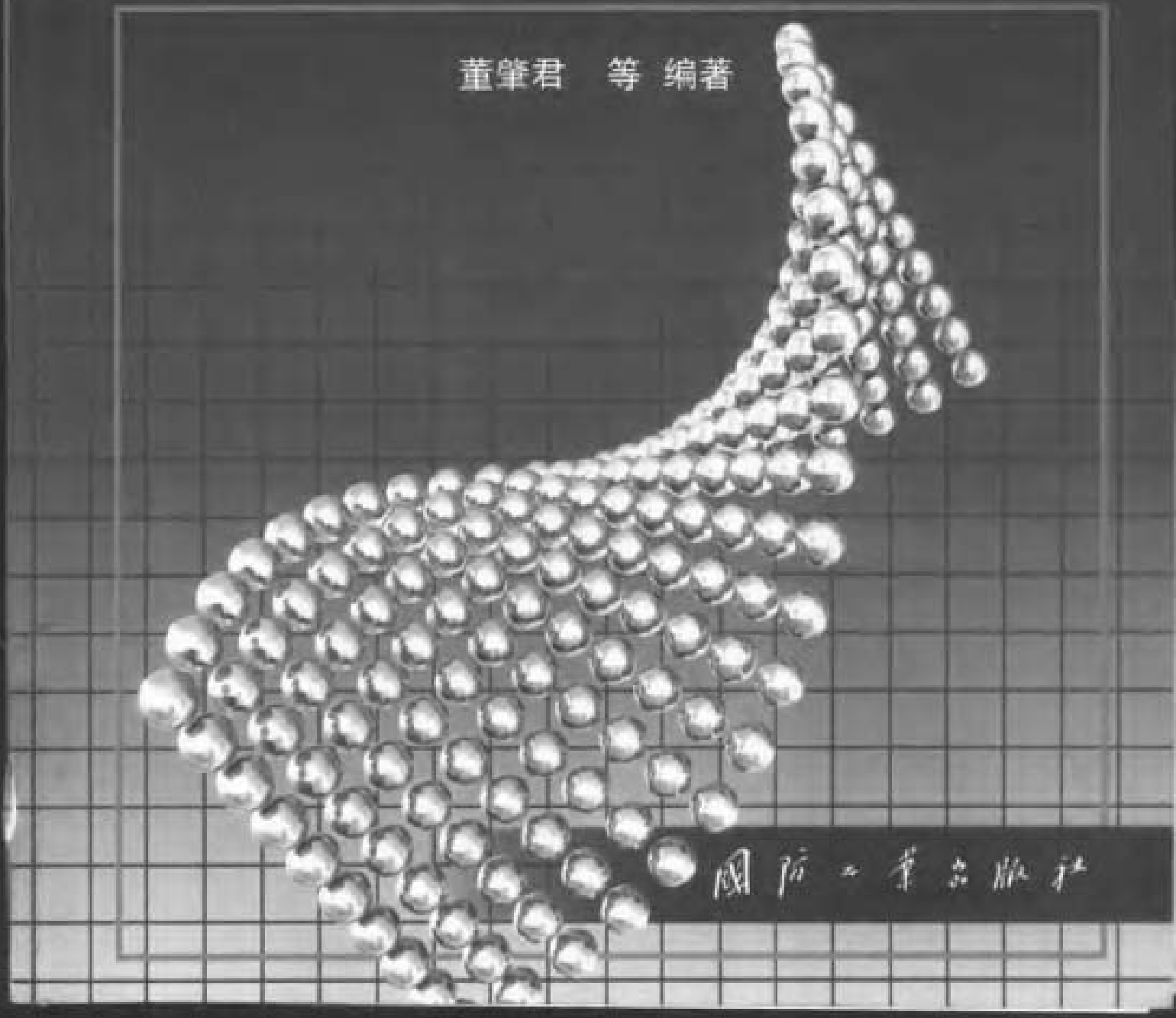
系统工程与运筹学

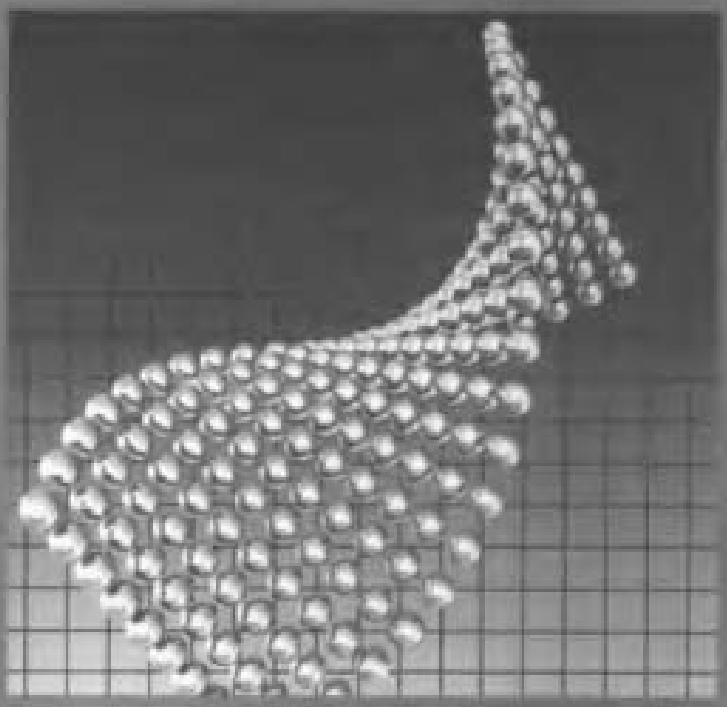
XITONG GONGCHENG YU YUNCHOUXUE

董肇君 等 编著



国防工业出版社

责任编辑　张文峰　肖志力



ISBN 7-118-03043-0

9 787118 030433 >

ISBN 7-118-03043-0/F·105
定价：38.00 元

系统工程与运筹学

董肇君　等编著

国防工业出版社

·北京·

内 容 简 介

本书立足于整体指导局部的系统思想，把辩证唯物论与现代科学技术相结合，把定性分析与定量分析相结合。重点介绍了系统科学与系统工程的基本原理与方法，并结合经济管理专业实践从建立模型和模型求解角度介绍了运筹学的基本方法。

本书可作为高等工科院校经济管理专业本科生和其他专业研究生的教学参考书，也可作为工程技术人员、管理干部培训班教材和自学参考书。

图书在版编目(CIP)数据

系统工程与运筹学 / 董肇君等编著. —北京：国防工业出版社，2003.1

ISBN 7-118-03043-0

Ⅰ.系… Ⅱ.董… Ⅲ.①系统工程②运筹学
Ⅳ.①N945②022

中国版本图书馆 CIP 数据核字(2002)第 099697 号

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号)

(邮政编码 100044)

涿中印刷厂印刷

新华书店经售

*

开本 787×1092 1/16 印张 26½ 617 千字

2003 年 1 月第 1 版 2003 年 1 月北京第 1 次印刷

印数：1—3000 册 定价：38.00 元

(本书如有印装错误，我社负责调换)

前　言

系统工程是20世纪50年代发展起来的新兴工程技术科学，是现代化管理的基础科学之一。我国自20世纪70年代开始研究，在经济管理以及工农业生产各个领域已取得丰硕成果。它不仅用于宏观决策，也用于微观决策，是从事规划设计、经营管理、科学研究不可缺少的理论和工具，其发展令人瞩目

系统工程立足于整体指导局部的思想，把辩证唯物论与现代科学技术相结合，把定性分析与定量分析相结合，给人一种认识世界、改造世界的崭新方法，是我国四化建设中不可缺少的一门方法性科学。

运筹学是系统工程的基础理论之一，是为系统工程提供数学理论和方法的科学。它是在系统思想指导下采用建立数学模型和模型求解等定量化方法为系统决策服务的技术科学，在我国属数学科学。

系统工程与运筹学是密不可分的，系统工程强调系统思想，运筹学强调数学方法，二者相辅相成可使问题得到较合理的解决。

本书尽量回避运筹学中有关数学方法的论证，从系统思想出发，以“问题导向”，由浅入深地介绍数学模型的建立和求解方法，力求通俗与实用。

本书共十二章，内容包括：系统科学方法论与系统；系统科学与系统工程；系统工程的重要方法——模型化；系统工程常用预测方法和模型；投入产出综合平衡模型；静态线性系统最优化模型；非线性静态系统最优化模型及求解方法；网络最优化方法；系统试验和模拟；系统决策；网络计划技术；随机服务系统；动态系统基础等。全书由董肇君主编，其中，系统工程常用预测方法与模型、投入产出综合平衡模型由符启勋同志编写、静态线性最优化模型(6.4节、6.5节、6.6节)由唐智娟同志编写，其他章节由董肇君编写。

本书可作为经济管理专业本科生和其他专业研究生必修与选修课的教学参考书，亦可作为各类现代化管理学习班、培训班教材，尚可供有关工程技术人员、管理干部自学和参考。

在本书编写过程中参阅了大量资料和著作，吸收了同行们辛勤劳动的成果，在此向同行们表示感谢。

由于编者理论水平有限，书中谬误之处在所难免，敬请广大读者批评指正。

编　者

2002年8月于天津

目　录

第 1 章　系统科学方法论与系统 …… 1

1.1　系统科学方法论 …… 1

1.2　系统的概念和特征 …… 6

1.3　系统的分类 …… 8

第 2 章　系统科学与系统工程 …… 10

2.1　系统科学体系 …… 10

2.2　系统工程 …… 14

2.3　系统工程的基础理论和工具 …… 19

2.4　系统工程的研究方法和步骤 …… 22

2.5　系统工程的应用和发展 …… 27

第 3 章　系统工程的重要方法——模型化 …… 36

3.1　模型与模型化 …… 36

3.2　系统模型化的基本理论 …… 39

3.3　模型化的程序和常用建模方法 …… 42

第 4 章　系统工程常用预测方法和模型 …… 46

4.1　预测科学 …… 46

4.2　定性预测技术 …… 48

4.3　定量预测技术 …… 51

4.4　带定性变量的线性回归预测模型 …… 68

4.5　判别分析预测模型与综合评审 …… 72

4.6　时间序列预测 …… 79

4.7　灰色预测——GM(1,1)模型 …… 82

第 5 章　投入产出综合平衡模型 …… 88

5.1　部门间投入产出综合平衡模型 …… 88

5.2　部门间投入产出综合平衡模型的应用 …… 98

5.3　企业投入产出模型 …… 115

第 6 章　静态线性系统最优化模型 …… 119

6.1　最优化及最优化模型的建立方法 …… 119

6.2　系统线性规划模型 …… 121

6.3　国民经济综合平衡最优化模型 …… 137

6.4 线性规划的求解方法 …… 146
6.5 灵敏度分析和参数规划 …… 168
6.6 对偶规划及影子价值 …… 181
6.7 整数规划 …… 184
6.8 模糊线性规划及其应用 …… 186
6.9 灰色线性规划及其应用 …… 194
6.10 Lingo 软件简介 …… 198

第 7 章 非线性静态系统最优化模型及求解方法 …… 203
7.1 非线性系统最优化模型 …… 203
7.2 无约束非线性规划问题求解方法及原理 …… 213
7.3 有约束非线性规划的求解 …… 225

第 8 章 网络最优化方法 …… 237
8.1 网络和图的基本概念 …… 237
8.2 一笔画和最小部分树问题 …… 239
8.3 最短路径问题 …… 243
8.4 最大流问题 …… 250
8.5 最小费用最大流 …… 257

第 9 章 系统试验和模拟 …… 260
9.1 系统的试验设计和分析 …… 261
9.2 模拟技术——蒙特卡罗法 …… 273
9.3 系统动态学模拟原理 …… 282
9.4 系统动态学模型 …… 287

第 10 章 系统决策 …… 294
10.1 决策树 …… 294
10.2 风险决策和不确定决策 …… 299
10.3 多目标决策 …… 303
10.4 目标规划法 …… 309
10.5 对策分析 …… 314

第 11 章 网络计划技术 …… 319
11.1 系统管理的网络技术 …… 319
11.2 网络图的编制 …… 321
11.3 关键线路法(CPM) …… 326
11.4 CPM 网络的优化 …… 334
11.5 计划评审技术(PERT) …… 343
11.6 决策关键线路法(DCPM) …… 349
11.7 图解评审技术(GERT) …… 355

第 12 章　随机服务系统 …… 365
12.1　随机服务系统的组成和特性…… 366
12.2　定长服务系统…… 370
12.3　生灭过程…… 371
12.4　普阿松输入、负指数分布服务系统分析(M/M/D)…… 374
12.5　随机服务系统的费用优化…… 381

第 13 章　动态系统基础 …… 385
13.1　动态系统一般知识…… 385
13.2　动态系统最优化理论和方法简介…… 402
参考文献…… 416

第1章　系统科学方法论与系统

1.1　系统科学方法论

人类在与自然界作斗争的过程中，靠自己的聪明才智逐步地认识、掌握自然界发展和变化的规律，并不断地向大自然索取所需要的衣、食、住、行等物品。然而人类在改造自然界的同时，也在改造自己。他们在失败的挫折中吸取教训，在成功的喜悦中总结经验，使自己的头脑更加聪明，并增强了认识世界和改造世界的能力。在向科学技术进军的旅途中，他们经过长期的积累，逐步从中提炼出一套认识和研究自然界的一般性方法。这种方法在实践中不断深化、充实、丰富和提高，就形成了科学方法论。科学方法论的出现对科学技术的发展起到了重要的作用。

1.1.1　系统科学方法论的发展

科学技术进步的过程反映了人类认识世界、改造世界的过程，也体现出科学方法论的发展过程。从方法论角度研究，科学方法论的发展大体上经历了古代、近代、现代三个相互联系的时期。

1.1.1.1　古代方法论

自人类出现以来，他们无时无刻不在同自然界打交道。在科学技术相当落后的古代，人们为了从自然界得到足够的生产和生活资料，在与自然界作斗争的过程中总结出一套把世界当作一个整体来研究的方法。这种研究方法在古希腊和古中国的一些著作中都有阐述。如古希腊辩证法奠基人之一的赫拉克利特在《论自然界》一书中说过："世界是包括一切的整体"；被马克思称为伟大思想家的亚里士多德在《工具篇》一书中曾指出，事物的生灭变化是由物料因、形式因、动力因、目的因引起的，并得出"整体大于它的各部分的总和"的论断而成为古代方法论的代表著。公元前6世纪—5世纪之间，我国春秋末期思想家老子强调自然界的统一性；南宋陈亮的"理一分殊"思想称"理一"为天地万物的整体，"分殊"是这一整体的每一事物的功能并试图从整体角度说明部分与整体的关系。这说明了古代人类采用"从统一的物质本原出发，把世界当作统一体"的方法来研究世界的。

古代这种研究事物的方法虽然强调对自然界的整体性和统一性的认识，但由于当时科学技术很不发达，尚缺乏对组成自然界这一整体的各个细节的认识能力，因此对自然界整体性和统一性的认识也是不完全的。由于当时自然现象总的联系还不能从细节的认识上得到证实，所以往往只得用理想、幻想或猜测的联系代替尚未证实的联系，这就造成在整体这一概念上和整体组成部分之间联系的刻划上的直观性、模糊性和思辨性。因而在处理事物时，逻辑方法占主导地位。形象地说，这个时期可称为"只见森林，不见树木"的

时期，这也正是当时的客观历史条件所决定的。

1.1.1.2 近代科学方法论

随着人类的进步和科学技术的发展，人们越来越感到古代对自然界的研究方法并不能进一步深刻地揭示出事物发展变化规律，这是由于自然现象总的联系还没有在细节上得到证明的缘故。

19世纪下半叶，近代科学开始兴起，力学、天文学、物理学、化学、生物学等科目从混为一体的哲学中分离出来，产生了自然科学。自然科学的诞生、发展和进一步分化，开辟了实验科学的新阶段，该时期科学研究的主要任务是通过实验、解剖和观察，收集资料、积累经验、进行资料的整理和加工，以期弄清自然界的细节。这在当时的科学技术水平下，只有把事物分成若干独立部分，分门别类地独立地进行分析才能办到。因此在实验、解剖和观察的基础上的分析方法就成为这个时期的主导方法。培根的著作《新工具》、迪卡尔的著作《方法论》就很自然地成为近代科学方法论的经典著作，并指导着该时期科学研究工作和生产实践。这两部著作对推动近代科学的发展起到了相当重要的作用。

近代科学方法论在促进自然科学知识的深化上、在材料的整理和加工上、在研究自然界的具体方法和手段上都高于古代方法论，这对当时自然科学的发展起到了很大的推动作用。但从方法论角度看，该时期的科学方法论尚存在不足之处，即在深入细致地考察自然界细节的同时，由于学科越分越细，以致忽略了对整体的研究，而造成“只见树木，不见森林”的状况，往往以对事物局部的较深刻的认识取代了对事物整体的认识而产生片面性和局限性。这正如恩格斯所说：“就是以这些障碍堵塞了自己从了解部分到了解整体、到洞查普遍联系的道路。”

19世纪下半叶，自然科学已经取得了很大成就，自然界的细节已能够得到较深刻的认识，尤其是能量守恒、细胞学和进化论的发现，更进一步地揭示出客观世界的普遍联系，这为人类重新综合地研究客观世界这一整体奠定了基础。马克思和恩格斯在上述“三论”的基础上总结出一套认识世界的新方法——唯物辩证法，并提出物质世界是由相互联系、相互依赖、相互制约、相互作用的事物和过程所形成的统一整体的认识世界的一般方法。由于马克思和恩格斯是在充分揭示出物质世界内在普遍联系的基础上提出世界的整体性和统一性这一概念的，因此它区别并高于以逻辑方法和思辩为主导的古代方法。又由于它是以“自然科学本身所提供的事实，以近乎系统的形式描绘出自然界相互联系的清晰图画”的，故它又区别并高于15世纪—19世纪上半叶的近代科学方法论。这样，马克思和恩格斯所创立的科学方法论就成为20世纪以来人们认识世界和改造世界的最先进的方法论。

1.1.1.3 现代系统科学方法论

20世纪初，人们在马克思和恩格斯所揭示出的认识世界的一般方法的指引下，进一步明确了系统概念的内涵，尤其是系统论、控制论和信息论出现以后，系统的概念与现代科学技术有机地结合起来，使系统的概念由定性转化为定量、由经验上升为理论，从而形成了一套以马克思主义唯物辩证法为基础的、既有理论指导、又有科学方法并拥有先进计算手段的系统科学方法论。它是从整体上最优地解决各类系统问题的锐利武器，是沟通马克思主义哲学和系统科学的桥梁，是对20世纪30年代以来的科学方法进行系统、科学地概括和总结，是我们当前处理系统问题的基本方法论。

从方法论的发展可以看到，人类认识现实世界的过程，是一个不断深化的过程，是在真理的长河中，逐步前进的。在古代，既少理论依据，又缺乏观测和实验手段，所以对许多事物，往往只能睹其外貌。随着科学和技术的进步，近代科学开始兴起，开创了实验科学的新阶段，使人们对事物的构成部分（个体）有了深刻地了解，但认识是不断深化的，在对部分有了更多规律性地了解之后，再回过头来利用所得到的信息，从关联人手探索出一套认识事物整体的方法，从而使人类认识世界、改造世界的能力达到较高的水平，这是科学发展的必然结果。系统科学方法论所创立的认识世界、改造世界的新方法，必将使人类社会开始一个新的时代——系统时代。

1.1.2 系统科学方法论的特征和组成

1.1.2.1 系统科学方法论的特征

系统科学方法论是在系统论、控制论和信息论基础上创立的，它是唯物辩证法的具体化，是架设在对客观事物和过程高度抽象的唯物辩证法与客观事物和过程具体运行实践之间的桥梁。作为20世纪以来认识世界、改造世界的基本方法论，它具有如下特征：

（1）整体性。整体性是系统科学方法论的基本出发点，它为人们从整体上研究客观事物提供了有效方法。该方法论要求人们始终把研究对象作为一个整体来看待，认为世界上事物和过程都不是其组成部分杂乱无章的偶然堆积，而是一个合乎规律的、由各要素组成的有机体。这一整体的性质与规律只存在于各组成要素间的相互联系、相互依赖、相互制约和相互作用之中。而各组成部分孤立的特征的总和并不能反映整体的特征。

（2）综合性。综合是相对分析而言的。15世纪—19世纪上半叶的方法论主要是以分析为特征的。当时科学本身的发展在客观上需要把整体分解成部分加以精确地研究。然而由于事物本身是其组成部分相互联系的整体，科学的发展本身又要求揭示不同物质运动形式的共同属性和规律，这就需要采用综合的方法。综合方法是把系统的各组成部分、各部分的结构和性能、各部分的联系、历史发展等因素联系起来加以考察，从中找出共同性和规律性的方法。因此系统科学方法论的关键是综合。

（3）定量化和最优化。定量化是系统科学方法论与传统方法论的主要区别之一。运用系统科学方法论处理系统问题时，总是尽量采用各种数学语言和数学工具使系统得到较精确的定量描述，以反映系统发展变化的规律。在定性分析指导下的定量化可克服单纯定性分析的缺点，为更进一步深刻地认识事物和过程提供捷径。最优化是定量化的方法之一，它可依据需要，确定系统最优目标而得到最大效益。它是系统科学方法论中方法的核心。

（4）信息化。系统科学方法论以信息论为其基础理论之一。因此，它在处理系统问题时特别强调信息的重要作用，这一点区别于只着眼于物质、能量，而忽视信息的传统方法。该方法论在处理系统问题时撇开物质与能量的具体形态，而把任何系统当作一个信息的传输和加工的系统，并认为只有信息流才能使系统维持正常的、有目的的运动，强调信息流对系统的支配、调节和控制作用。

（5）人—机方式。系统科学方法论是以人利用电子计算机作为处理系统问题的基本方式。在处理系统问题的过程中，人始终处于主导地位。当前，系统越来越复杂，需要处理的信息量越来越大，仅靠人脑进行加工是不够的，必须借助于人脑的延伸物——计算机

的帮助。因此人一机方式就自然地成为系统科学方法论处理系统问题的基本方式。

1.1.2.2 系统科学方法论的组成

系统科学的方法论是系统方法、反馈方法、信息方法相互关联而构成的方法体系。

1. 系统方法

系统方法是指按照事物本身的系统性，把被研究对象作为系统，始终从整体与部分、整体与环境、部分与部分之间的相互联系、相互依赖、相互制约、相互作用的关系上精确、综合地考察和研究，以使系统处于最优运行状态的一种方法。其特点是整体性、综合性和最优化。系统分析方法、模型化方法、最优化方法和模拟实验方法等是其经常使用的工具。

2. 信息方法

信息方法是指运用信息的观点，把系统看作借助于信息的获取、传递、加工、处理而实现其有目的性的运动的一种研究方法。其特点是把信息的概念作为分析和处理问题的基础，它完全撇开研究对象的具体运动形态，把系统的有目的的运动抽象为一个信息变换过程。该过程如图 1.1.1 所示。

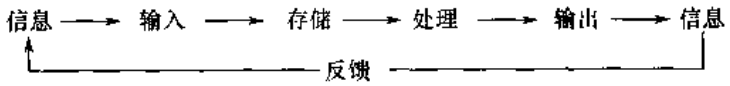


图 1.1.1 系统的信息处理过程

由图 1.1.1 可见，信息方法强调信息流的处理和加工过程，尤其是反馈信息流的存在能使系统实现预期的目标。

当前，由于经济管理体制等原因，我国在应用信息方法时有很多不足之处，主要表现在以下几方面：

(1) 信息流流动速度过低；

(2) 信息失真；

(3) 信息积累和存储不足；

(4) 信息通道不畅等。

为了有效地利用信息方法必须将加快信息流的流动速度、提高信息的可靠性、疏通信息通道、注意积累资料等工作作为首要任务加以解决。当前 Internet 技术的发展使人们获得信息的方式与途径更加广阔，进一步建设与完善“信息高速”公路成为我国当务之急。

3. 反馈方法

反馈方法是指用系统活动成果的信息调整系统的活动使之实现系统目标的一种方法，也就是在系统控制的过程中，收集行动效果的信息，并将其与系统目标相比较，从而提供控制信息的方法。其特点是在系统控制过程中始终运用反馈的概念。

反馈的概念包括以下三方面内容：

(1) 向既定目标的行动；

(2) 对行动效果的了解；

(3) 行动效果与既定目标的比较。反馈方法不仅对工程控制系统有重要作用，在经

济管理系统中的作用也是很大的。如现代化管理方法中的滚动计划法实质上就是一个个反馈的过程。系统的反馈过程如图 1.1.2 所示。

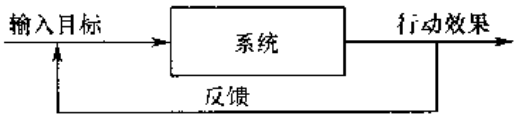


图 1.1.2　反馈过程

应用反馈方法的关键因素仍是信息的流动速度，保证信息处理过程的通畅是及时得到反馈信息的基础，因此信息方法和反馈方法是紧密相连的。

系统方法、信息方法和反馈方法是相辅相成的，在处理系统问题时它们总是交织在一起的，这也正是系统科学方法论的特点。

总的来说，系统方法是以系统论为基础的，信息方法是以信息论为基础的，反馈方法是以控制论为基础的，而三者的有机结合就构成了系统科学方法论。

1.1.3　系统科学方法论的指导思想——系统思想

系统科学方法论的指导思想是系统思想。系统思想就是系统思维方法，它是指唯物辩证法所体现的物质世界普遍联系及整体性的思想，是“以近乎系统的形式描绘出自然界相互联系的清晰图画”的思维方法。

系统思维方法与传统思维方法不同，传统思维方法是在以 15 世纪—19 世纪上半叶的方法论为基础的，它把了解世界当作尽可能彼此孤立的部分的总和，它习惯上把被研究对象分成若干独立部分，并把复杂对象的行为定义为各独立部分特性的简单相加，也就是采用“简单分解，简单相加”原理。按照这种思维方法，在处理问题时，先把被研究对象分解成若干无关联的部分，然后单独对各部分进行深入细致地研究，最后把研究的结果汇总，总结出研究对象的总体特性。这种以局部为基点，探索事物发展变化规律性的方法是片面的，甚至是错误的。其原因就在于它忽视了事物各组成部分之间的联系，忽视了部分与整体的差别，缺乏综合，缺乏辩证法。

系统思维方法是进行分析和综合的辩证思维工具。它把被研究对象当作一个完整、统一的整体来研究它的性质，也就是说，整体具有部分所没有的一些特性，也正是这些特性的存在而区别于部分。“整体大于部分之和”的非加和原理是系统思想的基本点。以这种思维方法研究问题，总是把被研究对象当作一个系统，从系统整体出发来研究系统内各组成部分、各部分之间的联系和系统与环境的关系。为了最大限度地利用整体特性，系统思想要求从不同角度把系统的概念贯通、综合，使系统的局部目标服从整体目标。这就是系统思维方法的一条原则。在此原则下，系统思想的另一特点就是从各部分的关联中探讨系统的整体特性，从而正确地认识整体。利用分析和综合的方法是系统思想处理问题的基本方法。

两种思维方法在表面上看好像只是形式上的不同，但事实上却有本质的差别，用两种方法处理同一问题所得结果是不同的。

两种思维方法是不同的，其区别见表 1.1.1。

表 1.1.1 系统思维方法与传统思维方法的区别

	系统思维方法	传统思维方法
出发点	整体	部分
研究基点	各组成部分之间的关系	部分
研究顺序	从整体到局部或由上级到基层，再由局部向整体综合	从局部到整体，由基层到上级
局部目标和整体目标的关系	局部目标必须服从整体目标，整体目标是局部目标的综合	以局部利益为基础形成局部目标，将局部目标叠加，形成整体目标。
考虑系统状态	动态地研究事物全过程	静态地研究事物的单一过程

随着科学技术的发展，人类社会活动不断大型化和复杂化，出现了许多庞大而复杂的系统，如社会经济系统、生态系统、管理系统等。这些系统有一些共同的特点：

(1) 规模越来越大，结构越来越复杂；

(2) 需要从时间、性能、费用、可靠性、可维修性等方面进行综合评价；

(3) 不确定因素越来越多和要求准确性越来越高的矛盾日益加深；

(4) 需要处理的信息量越来越大，信息的作用也越来越强；

(5) 解决问题需要各学科协同作战等。因此人们在实践中要求系统思想不仅能定性而且还能定量，以便从整体出发，协调部分的活动且把各部分的活动综合成技术上先进、经济上合理，并能协调运转的系统，取得最大效益。

20 世纪中叶，现代科学技术的发展为系统思想的定量化提供了一套具有数学理论、能够定量处理系统各组成部分相互联系的科学方法和强有力的计算工具——电子计算机，这就使系统思想由思维方法上升为理论——系统科学方法论。由此可见，系统科学方法论是以系统思想为指导，把系统思维方法和现代科学技术结合起来的产物，它既为我们提供思维方法，又为我们提供具体的理论和工具，因而是处理系统问题的基本方法论。

1.2 系统的概念和特征

系统一词由来已久，可一直追溯到古希腊唯物主义者德谟克利特(约公元前 540 年—480 年)的一本没有留传下来的著作《宇宙大系统》。但系统一词真正得到广泛使用还是 20 世纪 40 年代以后的事情，其内涵也在不断地发展和完善，并逐渐开始趋于统一。

1.2.1 系统的概念

1.2.1.1 系统的定义

关于系统的准确定义，国内外有不同的说法，主要有以下几种：

(1) 一般系统理论创始人冯·贝塔朗费认为："系统是相互作用的诸要素的综合体"；

(2) 韦氏大辞典(Weberster 大辞典)中系统(System)被解释为有组织或被组织的整体，被组合的整体所形成的各种概念和原理的综合，以有规则地相互作用、相互依赖的形式组成的诸要素的集合；

(3) 钱学森教授把系统定义为："极其复杂的研制对象，即由相互作用和相互依赖的

若干组成部分组合成的具有特定功能的有机整体，而且这个'系统'本身又是它所从属的一个更大系统的组成部分。"

为了适应系统工程学研究的需要，本书中所指的系统是由相互联系、相互依赖、相互制约、相互作用的若干部分，按照一定的方式，为了一定的目的组合而成的存在于特定环境之中并具有一定功能的有机整体。这个整体本身又是它所从属的更大整体的组成部分。

1.2.1.2 系统的属性

从系统的定义看，系统应具备以下六个基本属性：

(1) 整体性。系统作为由若干相互作用和相互联系的部分有机组合的、有一定结构和功能的整体，其本质特征是有机的整体性。系统整体性首先是系统目标的整体性；其次是系统功能的整体性，即组成系统的各部分的功能必须服从系统整体的功能，系统功能不等于各组成部分功能的简单相加，确定对系统的评价准则时，必须以系统整体为基础；再次是系统规律的整体性，系统整体的规律不是各组成部分规律的叠加。总之一切系统都是整体的，是组成部分与环境相互作用的整体，是各组成部分之间相互联系、相互作用、相互依赖、相互制约所形成的整体。

(2) 有序性(层次性)。凡系统都有结构，结构都是有序的。系统的有序性主要体现在系统的层次上。由于系统的各组成部分在系统中所处的地位不同，而形成了不同的层次，该层次关系决定了系统内物质、能量和信息的流动，从而使系统能够作为一个整体发挥较高的功能和效率。

(3) 集合性。系统都是由两个或两个以上可识别的部分(或子系统)所构成的多层次整体。作为子系统，是系统不可缺少的一部分。

(4) 关联性。系统各组成部分(子系统)之间按照一定的方式相互联系、相互依赖、相互制约、相互作用的性质叫系统的关联性。通过系统的关联性可揭示出系统整体特性和整体与部分的关系。

(5) 目的性。系统各组成部分按照统一的目的组织起来的性质叫系统的目的性。任何系统，尤其是人造系统都具有特定的功能，其组成都具有一定的目的并且有达成目的的手段。作为系统的一个组成部分都有为系统目的服务的一面，同时作为不同于其他组成部分又有维护自身利益的一面，因此研究确定系统目的和子系统目的之间的关系，保证各子系统在系统总目的的指导下，协同配合，分工合作，在完成各子系统目的的同时达成系统的目的是研究系统目的性的主要内容。

(6) 环境适应性。任何一个系统总处于特定的环境之中并与环境不断地进行物质、能量、信息的交换。系统离不开环境，而且必须适应环境的变化，否则系统将不能存在。这种系统随着环境的变化而存在的性质叫系统的环境适应性。

根据系统的定义和属性，我们可得出如下三条结论：

(1) 整体中每一个部分的性质或行为将影响整体的性质和行为；

(2) 每一部分的性质或行为以及影响整体的途径依赖于其它一个或几个部分的性质或行为；

(3) 每一部分对整体都不具备独立的影响，所以系统不能分出独立的子系统。

1.2.1.3 系统的运行模式

从系统的结构特点和与环境的关系方面研究，系统都按以下模式运行：

(1) 任何系统都由环境输入物质、能量和信息，经系统处理后向环境输出物质、能量和信息，因此系统均具有将输入转化为输出的功能。系统由输入、处理、输出三部分组成，并可用图 1.2.1 的模式表示。

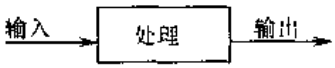


图 1.2.1 系统模式图

(2) 系统内部都有物质、能量、信息三种流的流动。系统本身的运动过程就是对这三种流的处理过程，我们过去往往只重视物质、能量两种流的管理，而忽视了信息流的管理，这是造成工作被动的原因之一。从对系统的组织管理角度研究，信息流是至关重要的。

(3) 系统都有反馈和自适应能力。系统都靠信息的反馈控制调整自身的运行，以适应环境并实现目标。

(4) 系统都有一严密的层次结构。生物系统、工程系统和非工程系统均如此。

1.3 系统的分类

系统广泛地存在于自然界，为了便于研究，需对系统进行分类。如何对系统进行分类取决于研究目的。依研究目的，选择不同的分类标志对系统分类可界定研究对象的属性。

1.3.1 按单一标志分类

1. 按组成部分的属性分

(1) 自然系统。由自然物组成的系统叫自然系统，如太阳系统等。

(2) 人造系统。由人或社会集团按某种目的建立的系统叫人造系统，如电力系统等。

(3) 复合系统。人造系统和自然系统复合而成的系统叫复合系统，如气象预报系统等。

其中第(2)、(3)两类系统是本书的主要研究对象。

2. 按与环境的关系分

(1) 封闭系统。与环境很少发生物质、能量、信息交换的系统叫封闭系统。实际上，世界上根本不存在封闭系统，只是为了研究方便，有时人为地把一个系统当作封闭系统处理。

(2) 开放系统。与环境经常发生能量、物质、信息交换的系统叫开放系统。系统与环境所进行的物质、能量、信息的交换，可影响系统的结构、功能及其发展。一旦系统与环境的联系切断，就会对系统的稳定产生不利影响，而使系统遭到破坏。

3. 按所处的状态分

(1) 静态系统。系统的要素不随时间变化的系统叫静态系统。事实上这种系统是不存在的，只是为了研究方便，人为地假设系统处于相对平衡或静止的状态。

(2) 动态系统。系统的要素随时间变化的系统叫动态系统。

4. 按形态分

(1) 实体系统。由实物构成的系统叫实体系统。如机械系统等。

(2) 概念系统。由概念、原理、方法等非物质组成的系统叫概念系统。如软件系统等。

5. 按系统的规模分

(1) 小型系统。所包含的部分(子系统)数量较少或变量数较少的系统属小型系统。一般来讲,这类系统各部分之间的关联较简单、如工艺简单、产品品种较少的企业。

(2) 中型系统。所包含的部分(子系统)较多或描述系统状态的变量数较多,各部分之间的关联比较复杂的系统属中型系统。如中型企业、地、县一级的区域系统等。

(3) 大型系统。所包含的影响因素相当多、组成部分(子系统)或描述系统状态的变量数很多、结构层次复杂,关联复杂的系统属大型系统。如国民经济系统、生态经济系统等。

(4) 超大型系统。包含的部分(子系统)或描述系统状态的变量数相当多、结构层次极其复杂以致应用常规的方法和手段无法完成其分析和优化,必须采取一些独特的处理方法的系统属超大型系统。如生物圈系统、人体系统等。

此外,还有很多其他分类方法,如按具体研究对象可把系统分为不同的对象系统,如教育系统、林业系统、电力系统等,在此就不一一赘述了。

1.3.2 综合分类

综合分类是按研究目的将某些分类标志加以组合形成组合分类标志,并以该组合分类标志对系统进行分类的一种分类方法。如将规模与环境关系两个标志组合进行分类可将系统划分为大型开放系统、中型开放系统、小型封闭系统等。

第 2 章　系统科学与系统工程

2.1　系统科学体系

2.1.1　现代系统科学体系

钱学森教授于 1981 年构造设计出现代科学体系如图 2.1.1 所示。

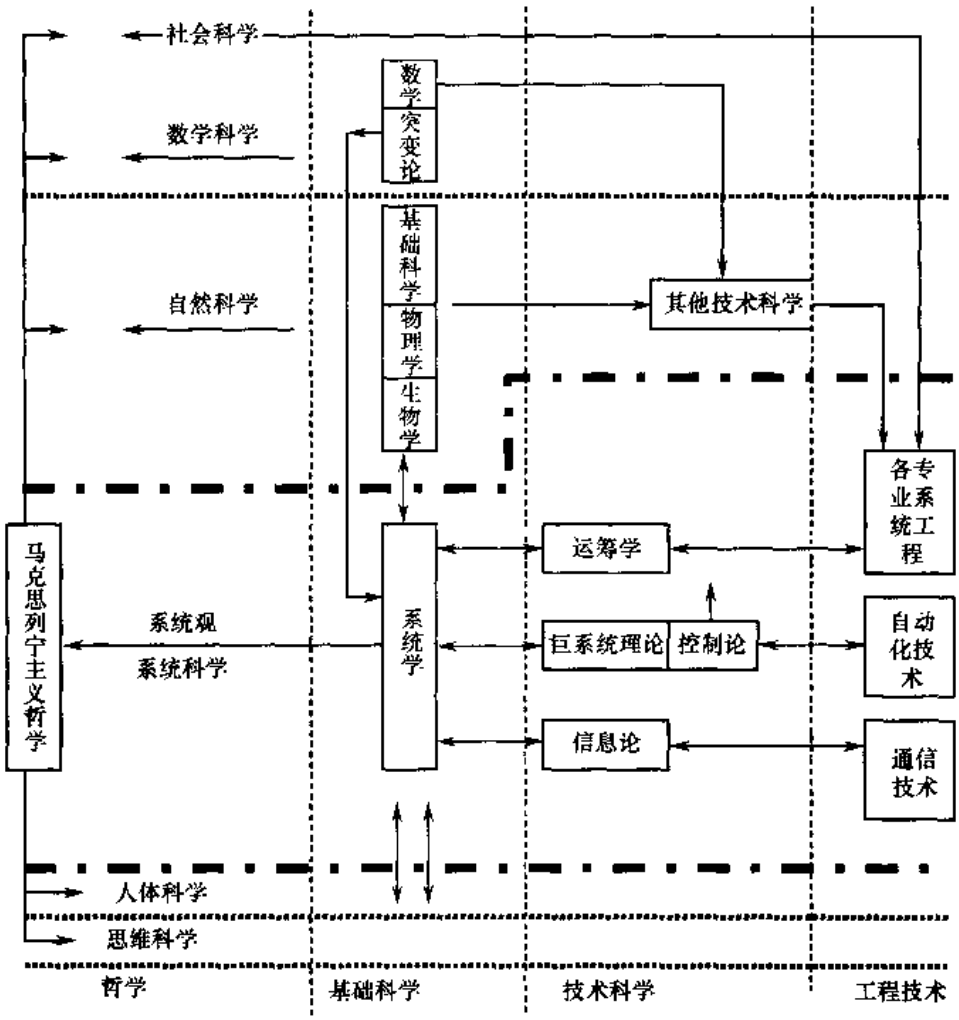


图 2.1.1　系统科学体系

按钱学森教授的构想，现代科学技术体系纵向可分为四个层次：哲学、基础科学、技术科学和工程技术。横向可由自然科学、社会科学、数学科学、系统科学、人体科学和思维科学组成。系统科学是在自然科学、社会科学和数学科学之外正在形成的一个新的学科体

系。钱学森教授认为,在系统科学中,直接与改造客观世界的社会实践相联系的是一类新的工程学——系统工程、自动化技术和通信技术,它们属工程技术层次的科学,这类工程技术的共同基础理论是运筹学、控制论和信息论,它们属技术科学层次的科学。系统科学也像自然科学一样有基础科学,构筑这门基础科学的材料一方面来自工程技术实践中提炼的技术科学——运筹学、控制论和信息论,另一方面来自自然科学和数学科学中的系统理论,如冯·贝塔朗费的一般系统理论和理论生物学,普里戈金的系统远离热平衡态的耗散结构理论,哈肯的协同学、托姆的突变论、艾肯的生命自组织超循环理论等,把上述两方面的材料,融汇贯通,综合发展就可能建立起系统科学的基础科学及适应于一切系统的一般理论——系统学。系统科学从各门系统工程到运筹学、控制论和信息论,再到系统学就形成了一个体系——系统科学。从马克思主义哲学到系统学的桥梁可以称为"系统观"或"系统论",它将成为辩证唯物主义的一个组成部分。

2.1.2 系统理论

在系统科学中,构成系统学的几个基本理论是:一般系统理论、运筹学、控制论、信息论、协同论、突变论和耗散结构等。

这里仅对组成系统学的几个系统理论做一简要介绍。

2.1.2.1 一般系统理论

该理论是美籍奥地利生物学家冯·贝塔朗费创立的逻辑和数学领域的科学,其目的是为了要建立适用于系统的一般原则。一般系统理论起源于机体论,它是一种与机械论相对立的生物学理论。冯·贝塔朗费指出机械论的三个基本错误是:

(1) 简单分解,简单相加的原理;

(2) 机械观点;

(3) 被动反应观点。

冯·贝塔朗费针对"简单分解、简单相加原理",提出了"非加和"原理和"整体大于部分之和"的概念。

他认为:把有机体分解为各要素,并简单相加以说明机体的属性是错误的,把孤立的各组成部分的活动性质和活动方式简单的相加,不能说明高一级水平的活动性质和活动方式。如果了解各组成部分之间存在的全部关系,则高一级水平的活动就能从各组成部分中推导出来。因此,要了解事物的整体特性,就要了解各组成部分,更要了解它们之间的关系。这与我们以前介绍的系统思想是一致的。

冯·贝塔朗费针对机械论的错误观点,提出适用于系统的一般原则为:

(1) 整体性原则;

(2) 相互关联原则;

(3) 有序性原则;

(4) 动态原则。

他所提出的整体性原则就是一切有机体都是一个整体——系统;相互关联原则就是系统各组成部分之间都是相互联系、相互依赖、相互制约、相互作用的;有序性原则就是一切有机体都是按严格的等级组织起来的;动态原则就是这一切有机体本身都处于积极的运动状态。

冯·贝塔朗费从理论生物学的角度总结了系统思想，运用类比、同构的方法建立起开放系统的一般系统论，提出生命现象的有组织性、有目的性和有序性。如果我们把研究对象当作一个系统——有机整体，就完全可以应用贝塔朗费的系统论。

2.1.2.2 耗散结构学说

耗散结构学说是比利时物理学家普里戈金于1970年创立的，它是一种非平衡系统理论。普里戈金认为，非平衡可为有序之源，不可逆过程可导致称为耗散结构的新型物态。他把这种远离平衡态的、稳定的、有序的结构称为耗散结构，回答了开放系统如何由无序走向有序的问题，他因此而荣获了诺贝尔奖。

1. 平衡、稳定和有序的概念

马克思认为："平衡是经常不断地消除经常的不协调"；恩格斯认为："相对静止即平衡"；毛泽东说："所谓平衡就是矛盾暂时的相对的统一"。因此，平衡具有协调、按比例、静止、统一等含义，可定义为：系统各组成部分间保持相互协调、保持即定的某种比例而体现出的系统整体状态，称平衡态，简称平衡。一般而言，它是人们追求的目标。用数学语言可描述为：反映系统特征量的边际值为零的状态

$$\frac{\mathrm{d}w}{\mathrm{d}t}=0 \quad 或 \quad \frac{\mathrm{d}w}{\mathrm{d}k_j}=0$$

式中 $\mathrm{d}w$——系统整体特征的变化率；

$\mathrm{d}k_j$——第 j 要素的变化率；

$\mathrm{d}t$——时间的变化率；

在经济系统中，总供给等于总需求、收入等于支出，均可称为平衡，否则称为非平衡。

稳定与平衡密切相关，按控制论的观点，如果系统处于平衡态，在一小的扰动下使其偏离平衡态，在一即定的时间内系统靠自身的力量仍能保持其平衡态，则该平衡态是稳定的，称稳定的平衡态，否则是不稳定的平衡态。

有序是指系统各组成部分有严格的秩序和结构，而无序则是组成部分杂乱无章，没有一定的秩序和结构。系统有序与无序是相对的，一般用用"熵"来度量系统的有序程度。熵越大系统无序程度越高，反之系统有序程度越高。

2. 耗散结构

耗散结构的概念是对应于平衡结构而提出的，它的严格定义是来自于物理学中的非平衡态热力学。平衡态热力学主要以平衡系统的稳定有序结构为研究对象，解决整个系统由稳定有序结构变成不稳定的相对无序的结构，或变成稳定的有序程度较低的结构问题。关于非平衡态以及从无序到有序的变化机制只能由耗散结构学说来解释。

人们通常认为，系统若处于一种混乱无序的非平衡状态是不可能出现一种稳定有序结构的。但普里戈金通过长期的研究后指出：远离平衡态的开放系统在外界条件变化达到某一阈值时，量变可能引起质变，系统通过与外界不断地交换能量和物质，自动产生一种自组织现象，组成系统的各子系统会形成一种互相协同的作用，可能使系统从原来的无序状态转变为一种稳定有序的结构，从而实现系统由无序向有序、由较低的有序向较高的有序转化。这种非平衡态下的稳定有序结构称之为耗散结构，它是一种动态的稳定有序结构。

3. 耗散结构学说的基本观点

耗散结构学说的基本观点为：耗散结构是系统的客观表现。其产生条件为：

(1) 系统必须是与环境进行能量、物质和信息交换的开放系统，并不断地引入足够大的负熵流；

(2) 系统必须是远离平衡态的非平衡系统；

(3) 系统内各要素之间存在非线性的相互作用。

一个系统熵的变化 ds 由二项构成：一项是系统内的熵产生 $d_i s$；一项是系统与外界进行能量交换引起的熵流 $d_e s$，可表示为：

$$ds = d_i s + d_e s$$

当系统是孤立系统或封闭系统时，因与外界环境不存在能量、物质、信息的交换，故 $d_e s = 0$，此时，如系统处于非平衡态，按热力学第二定律，非平衡态总是使系统不可逆地、自发地趋于熵取最大值的平衡态，$d_i s$ 将逐渐增大，ds 随之增大，因总熵不断增加，系统内有序逐渐变为无序，系统最终必定走向平衡；此时，如系统处于平衡态，$d_i s$ 已达极大值，$d_e s = 0$，总熵 ds 也达到最大值，系统处于完全无序状态。

当系统开放且处于非平衡态时，在与环境进行能量、物质、信息的交换过程中产生熵流 $d_e s$，熵流可正，可负，可为零，如果 $d_e s < 0$，且其绝对值大于熵产生 $d_i s$ 时，才可能有 $ds < 0$，才能使 ds 为负，系统总熵才能减少，系统才能由无序走向有序。这种负熵流的引入将离不开系统与环境的交流和远离平衡态两个条件，这两个条件就是系统由稳定的无序产生不稳定的有序，再到稳定有序的必要条件。

从数学的角度看，系统的某种特定状态对应于系统动力学方程的某个特解，在系统由稳定无序向稳定有序转变过程中的行为所对应的特解应既可能有不稳定特解，又可能有稳定特解，此情况下的系统动力学方程必须是非线性的，所以耗散结构的形成又必须以系统内各要素间的非线性为充分条件。

用耗散结构理论研究经济系统，可以这样理解：当经济系统是开放的且处于非平衡态时，意味着各经济部门发展不平衡且向平衡发展，此时熵产生大于熵流。如果与外界交流较少时，系统将因熵增大而逐渐走向平衡。但当某些发展快的部门从外界获得足够的资金和技术，便会迅速发展壮大，系统中的其他部门一部分可能消亡，另一部分则会在它们的带动和影响下也逐渐发展，从而使整个系统生机勃勃，这是因为引入的负熵的绝对值大于熵产生使 $ds < 0$ 的结果，是系统走向有序的表现。按经济学中边际效益递减规律，当投入的资金和技术达到一定限度之后，再投入资金将使这些部门的效益减少，致使这些部门投入减少，导致负熵增大，当负熵的绝对值等于熵产生时，$ds = 0$，此时，系统开始由有序向无序迈进；当负熵的绝对值小于熵产生时，$ds > 0$，系统就更加无序。如果熵产生达到最大值，则该经济系统就完全无序，走向死亡。但经济系统不会走向终结，因为在系统处于非平衡态时，只要是开放的，必定会产生新的投资热点，萌生出新的经济增长点，使系统向更高的有序前进。

2.1.2.3 协同学

协同学，又称协合学，它是德国物理学家哈肯在 20 世纪 70 年代后期建立起来的一种非平衡系统论，它是以研究不同学科间存在着的共同特征为目的的一门横向学科，也是许多学科形成的一门边缘学科。它借助耗散结构理论的一些研究成果，成功地解释了系统的局部与整体，系统由简单到复杂，从低级到高级，由无序到有序稳定发展过程中最本质的东西，即协同作用。

哈肯认为:必须从具体地分析各种可能形成非平衡有序结构的系统行为入手,建立共同的数学模型,并对其进行动力学和统计学两方面的考察,才能认识非平衡开放系统的稳定有序结构形成的条件、特征及其规律。

按协同学的观点:系统均是物质的且可分为封闭系统和开放系统两类。非平衡的封闭系统只能自发地趋于平衡。趋于封闭的平衡也就是趋于无序、趋于简单。一个开放的系统既可处于平衡态,也可处于非平衡态。处于平衡态的开放系统在一定条件下可呈现出有序结构,称为静的有序。处于非平衡态的开放系统在一定条件下也能出现宏观有序结构,称为动的有序。开放性是产生有序结构的必要条件,而非线性是产生有序的基础,只有协同性才是产生有序性的直接原因。

系统的主要特征是协同性,而层次性、结构稳定性、有序性都与协同性有关。因此,协同学在系统论创新过程中发挥了重大的作用。系统的协同性是系统整体性的理论依据,它可以拓展到各种复杂系统,它说明整个自然界是由无限多个不同的物质形态组成的相互联系的统一体。根据统一性原则,认为可以利用已知系统揭示的规律性去认识和揭示未知的各种复杂系统的规律性。因此,我们在研究系统时,只需模拟一个已知的复杂系统的特征所得到的基本理论、规律、方法就可将它推广到未知的复杂系统中去。系统的这种属性,称之为系统的规范性。

此外,由协同学的研究引出系统论的协同性原则可定义为一个开放系统内部各于系统之间的协调同步的非线性作用。由此可知,协同性与有序性恰好构成一个因果关系,也即协同性是有序性的原因,而有序性则是协同性的结果。一个开放系统,由于协同性所导致的有序性是一种"活"的有序性,或称"动"的有序性。正是这种动态性,才使系统得以发展。

总之,系统论的很多特性与原则,都是从协同性的研究中获得解释的,因此,系统论的协同性、规范性、自组织性(即系统由自身矛盾运动能够自身走向有序结构,它是系统通过信息的反馈作用在与环境交换物质和能量的过程中,不断地调整自己行为和活动的结果)等都是从近代科学研究新成果中引出的。这不但丰富了系统论的研究内容,而且为我们发展和创新系统论打开了思路,提供了新的方法。

协同学主要是从物理学角度出发运用现代科学技术的最新成果和现代数学理论提出了多维相空间的理论,研究了系统的组织状态,提出系统必定走向一有序结构,该结构就是系统的客观目标,不管从空间的哪一点开始,终归要走到这个代表有序结构的点,系统只有在此目标上才是稳定的,否则就不稳定。该理论解决了复杂系统如何由无序走向有序,且证明了不仅开放系统如此,封闭系统也如此;不仅平衡系统如此,即使非平衡态也如此。可见哈肯的系统论比普里戈金提出的耗散结构更前进了一步。

2.2 系统工程

2.2.1 系统工程的定义

现代科学技术为系统思想的定量化创造了条件,它为系统思想定量化提供了数学理论和强有利的计算工具——电子计算机,并推动了系统科学的发展。到20世纪60年代,系统思想的定量化已发展成既有理论指导,又有科学方法和实践内容的新的工程技术学

科——系统工程。

系统工程作为一门学科问世以来，仅仅50多年的时间，在各行各业、各个领域都得到了广泛的应用，收到了良好的效果，同时系统工程的实践也促进了本学科的继续发展与完善。无庸置疑，系统工程已成为当前最有前途的学科之一。但是系统工程毕竟是一门非常年青的学科，它的理论和方法尚需在实践中进一步发展与完善。到目前为止，关于系统工程的定义和研究的内容，国内外学者仍齐说不一，原因在于：

(1) 系统工程的理论和方法是在自然科学、社会科学和数学科学向纵深发展时产生一些需要协同解决的问题的情况下产生的，从事不同专业的人，出于专业兴趣，对系统工程有不同的理解；

(2) 由于系统工程是现代科学技术的产物，它综合地运用各学科的先进成果去解决面临的问题，因此很难划清系统工程的学科界限。

由于以上原因，从事不同专业的人为系统工程所作的定义也各不相同。

我们这里首先介绍一下钱学森教授对系统工程定义的说明。钱学森教授在《系统思想与系统工程》一文中指出："20世纪40年代以来，国外对定量化系统方法的实际应用相继取了许多不同的名称：运筹学(OPERATIONS RESEARCH)、管理科学(MANAGEMENT SCIENCE)、系统工程(SYSTEMS ENGINEERING)、系统分析(SYSTEMS ANALYSIS)、系统研究(SYSTEMS RESEARCH)，还有费用效果分析(COST EFFECTIVENESS ANALYSIS)等等。他们所谓运筹学，指目的在于增加现有系统效率的分析工作；所谓管理科学，指大企业的经营管理技术；所谓系统工程，指设计新系统的科学方法；所谓系统分析，指对若干可供选择的执行特定任务的系统方案进行选择比较；如果上述选择比较着重在成本费用方面，即所谓费用效果分析；所谓系统研究，指拟制新系统的实现程序。现在看来，由于历史原因形成的这些不同名称，混淆了工程技术与其理论基础即技术科学的区别，用词不够妥当，认识也不够深刻。国外曾有人试图给这些名词的涵义以精确区分，但未见取得成功。"

"用定量化的系统方法处理大型复杂系统的问题，无论是系统的组织建立，还是系统的经营管理，都可以统一地看成是工程实践。工程这个词18世纪在欧洲出现的时候，本来专指作战兵器的制造和执行服务于军事目的的工作。从后一涵义引伸出一种更普遍的看法：把服务于特定目的的各项工作的总体称为工程，如水利工程，机械工程……，如果这个特定的目的是系统的组织建立或者是系统的经营管理，就可以统统看成是系统工程。国外称运筹学、管理科学、系统分析、系统研究以及费用效果分析的工程实践内容，均可以用系统的概念统一归入系统工程；国外所称运筹学、管理科学、系统分析、系统研究以及费用效果分析的数学理论和算法，可以统一地看成是运筹学。"

综合归纳上述意见，我们认为：系统工程是一门新兴的工程技术学科，是应用科学。它不仅定性，而且定量地为系统的规划设计、试验研究、制造使用和管理控制提供科学方法的方法论科学。它的最终目的是使系统运行在最优状态。

系统工程的研究内容具有以下特点：

(1) 系统工程不同于一般的工程技术学科。一般工程技术学科，如水利工程、机械工程等都与形成实物实体的对象有关，国外称这类工程为"硬"工程，而系统工程的研究对象除了这类"硬"工程之外，尚包括这种工程的组织与经营管理一类国外称之为"软"科学的

各种内容。

(2) 系统工程涉及各种学科、各个领域的各种内容。把一个工程项目作为系统来处理,必须涉及其他工程技术的内容。如果把一般工程学科作为一条代表专业的纵线,则系统工程则是跨越各条纵线的一条横线。它通过横向的综合,提出解决问题的方法和步骤,因此它是跨越不同学科的综合性科学。

如建筑工程是专业技术工程学科,它主要研究建筑设计和施工技术,而建筑系统工程则是综合社会、经济、生态及其他工程技术系统等进行城市的规划、设计与管理、建筑物的设计、施工组织等问题,开展以规划、设计、施工及管理为主线,以社会学、经济学、生态经济学、美学、工业工程学、电子学等为基础的综合地、系统地研究,以实现城市系统社会、经济、生态效益的统一和优化。

(3) 系统工程在研究问题时,概念、原则、方法是主要的、本质的,而在系统工程中应用的具体数学方法和计算技术是处理和解决系统问题的手段和工具,是为系统工程的概念、原则和方法服务的。这个观点是辩证唯物论的系统观的重要原则之一。

(4) 任何系统都是人、设备和过程的有机组合,其中人是最主要的因素。因此在应用系统工程的方法处理系统问题时,要以人为中心。

2.2.2 基本定律和基本观点

应用系统工程的方法解决问题,必须运用下述基本定律和基本观点。

2.2.2.1 基本定律——系统性能、功效不守恒定律

系统性能、功效不守恒定律是指:当系统发生变化时,物质、能量守恒,但性能和功效不守恒,且不守恒性是普遍的和无限的。

定律中系统的变化是指:

(1) 由子系统合成系统(由部分组成整体);

(2) 系统分解成子系统(整体分解成部分);

(3) 系统内部结构的变化。

"不守恒"是指:

(1) 产生新的性能和功效;

(2) 原有性能、功效的增强、减弱或消失。

"普遍性"是指:任何系统都具有这种性质。

"无限性"是指:低级层次子系统向高层次系统发展是无限的。这种无限性体现出低层次子系统构成高层次系统时,系统性能、功效的复杂变化,即高层次系统具备低层次系统的基本属性,同时又产生低层次系统不具备的新的属性。

该定律成立的主要依据是:

(1) 由物质不灭定律和能量守恒定律可知,系统内物质、能量和信息在流动的过程中物质是不灭的、能量是守恒的,而反映系统性能和功效的信息,因可受干扰而失真、放大或缩小,以至湮灭,故是不守恒的;

(2) 由于系统的变化,系统内物质、能量、信息在时间上和空间上发生叠加、互补和抵消,从而改变了系统的性能和功效。

这一定律在进行系统研究时,在处理系统与子系统、子系统与子系统、系统与环境的

关系并使之定量化时是很重要的。

2.2.2.2 系统工程处理问题的基本观点

1. 整体性观点

所谓整体性观点即全局性观点或系统性观点，也就是在处理问题时，采用以整体为出发点，以整体为归宿的观点。这种观点的要点是：

(1) 处理问题时需遵循从整体到部分进行分析，再从部分到整体进行综合的途径，首先要确定整体目标，并从整体目标出发，协调各组成部分的活动；

(2) 组成系统的各部分处于最优状态，系统未必处于最优状态；

(3) 整体处于最优状态，可能要牺牲某些部分的局部利益和目标；

(4) 不完善的子系统，经过合理的整合，可能形成性能完善的系统。

系统是由很多子系统相互关联而成的，而研究系统的目的是为了达成系统的整体目标。以简单分解，简单相加的方法，从部分着手研究问题，必定会影响全局，使我们离开辩证法，陷入形而上学。

这里还必须强调的是：系统的整体性包含时间的整体性和空间的整体性两个方面，这是系统的时空观。六七十年代提出的"边设计、边施工、边生产"这一貌似正确，实际错误的方法，就是形而上学时空观的具体体现。

2. 综合性的观点

所谓综合性的观点就是在处理系统问题时，把对象的各部分、各因素联系起来加以考查，从关联中找出事物规律性和共同性的研究方法。这种方法可以避免片面性和主观性。

阿波罗登月计划总指挥韦伯曾指出，当前科学技术的发展有两种趋势：一是向纵深发展，学科日益分化；二是向整体方向发展，搞横的综合。阿波罗计划中没有一项新发明的自然科学理论和技术，都是现成科学的运用，关键在于综合，综合是最大的科学。系统工程就是指导综合研究的理论和方法。韦伯的这段话说明了综合性的观点是系统工程处理问题时的基本观点。

要做到综合地处理问题，应从以下几方面着手：

(1) 要应用多学科知识解决问题。避免片面性，只靠一个人是不行的。因此在处理系统问题时，必须组成由多方人士参加的"专家组"，由一种"T"字形人才——系统工程师去组织。

(2) 要用综合性指标来评价系统，系统综合评价指标体系如图 2.2.1 所示。

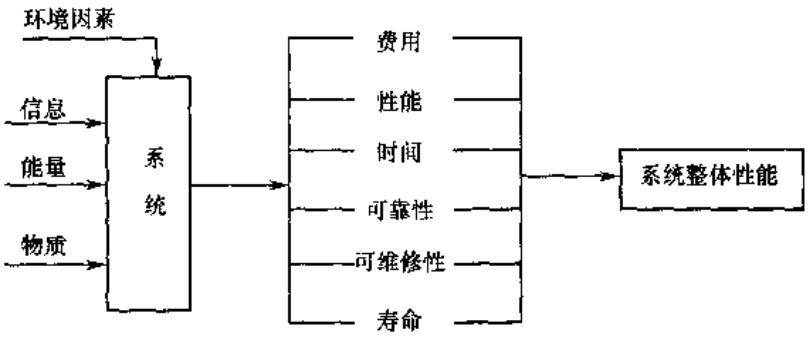


图 2.2.1 系统综合评价体系

(3) 不被表面现象迷惑,注意研究潜在因素对系统的影响。如某地区在200年前,森林、农田、牧场构成一协调的生态系统,在此系统中,林丰、草旺、田肥,从事各种生产经营活动,各得其所。为了发展农业,毁林、毁牧开荒,只经100多年的时间,这片肥沃的土地就变成了沙丘累累的荒芜地带。这就是对系统的开发缺乏综合性的后果。"文革"期间,我国南方地区围湖造田,一些林区、牧区毁林开荒、毁牧开荒所引起的后果都是缺乏综合性的例证。

3. 科学性的观点

所谓科学性的观点就是要准确、严密、有充足科学依据地去论证一个系统发展和变化的规律性。不仅要定性,而且必须定量地描述一个系统,使系统处于最优运行状态。

马克思曾明确指出,一种科学,只有当他成功地应用了数学的时候,才能达到完善的地步。数学方法已成为解决系统工程问题的主要方法。在强调采用定量方法的同时,有以下两个问题必须引起我们的注意:

(1) 必须在定性分析的基础上进行定量分析,定量分析必须以定性分析为前提。过去我们善于应用定性的分析方法。只进行定性分析,不能准确地说明一个系统,只有进行了定量分析之后,对系统的认识才能达到一定的深度,结论才能令人信服。然而没有定性分析作指导,定量分析就失去了依据,就会成为"数学游戏"。因此,我们强调要摆正定性分析和定量分析的辩证关系,在处理问题时,一定要在定性分析的基础上应用数学方法,建立模型,进行优化,从面达到系统最优化的目的。

如我们在安排生产计划时,可在各种资源的限制下制定一个使利润达到最大值的生产计划。这就需要在约束组成、确定评价目标等方面进行定性分析,然后在定性分析的基础上应用数学规划等工具,建立模型,完成该项任务。

(2) 合理处理最优和满意的关系。在处理系统问题时,使系统达到最优比较困难,在个别情况下,"最优"有时不被人理解和不愿意接受,因此有时利用满意的概念会使问题得到圆满的解决。从数学上的最优,过渡到情意上的满意是西蒙的一大发现。因此我们在处理问题时,要处理好满意和最优的关系。这一原则也是不违背科学性的观点的,因为寻求满意解也是科学。

4. 关联性的观点

所谓关联性的观点是指从系统各组成部分的关联中探索系统的规律性的观点。我们曾指出,一个系统是由很多因素相互关联而成的,正是这些关联决定了系统的整体特性。因此在处理系统时,必须努力找出系统各组成部分之间的关系,并设法用明确的方式描述这些关系的性质,来揭示和推断系统整体特征,也只有抓住这些联系,用数学、物理、经济学的各种工具建立关系模型才能定量和定性地解决系统问题。不然对一些复杂的问题会感到无从下手。难怪钱伟长教授在介绍系统工程时曾风趣地称系统工程为"关系学"。

如美籍苏联经济学家列昂节夫在研究国民经济系统时,就是抓住各物质生产部门之间的联系并使其定量化,从而以投入产出模型揭示国民经济总体的发展变化规律。

揭示系统各组成部分之间的关联是靠分析和观察实现的,切忌凭空臆造和估计。

如达尔文曾发现英国有一种三叶草与村子中猫的数量有关,通过观察发现,三叶草靠土蜂传粉,田鼠以土蜂为生,猫又是鼠的天敌,因此构成了一串称之为食物链的联系:三叶草——土蜂——田鼠——猫。猫多、鼠少,则蜂多、三叶草就茂盛,反之,猫少、鼠多、土蜂

就少，三叶草必然就少。把该关系定量化，即可得出猫与三叶草的关系。就可通过控制猫的多少，实现对三叶草的控制。这一简单的生态系统就是从关联入手解决的。可见关联性的观点在解决系统问题中的重要作用。

5. 实践性的观点

实践性的观点就是要勇于实践，勇于探索，要在实践中丰富和完善以及发展系统工程学理论。

系统工程不是束之高阁的空头理论，也不是玄妙的数学游戏，它是来源于实践并指导实践的理论和方法，只有在实践中、在改造自然界的斗争中系统工程才会大有作为并得到迅速的发展。采用“问题导向”，摒弃“方法导向”是系统工程实践的主要方法。

为了推广系统工程的方法，实践性是很重要的，只有系统工程的广泛实践，才能使人们认识和了解系统工程的作用，才能促进系统工程的应用和发展。

坚持实践性的观点，促进系统工程的发展，可使这门新兴的学科在我国的四化建设中发挥更大的作用。

2.3 系统工程的基础理论和工具

系统工程是综合地运用各种学科的科学成就、为系统的规划设计、试验研究、制造使用和管理控制提供科学方法的工程技术，它是在运筹学、控制论和计算科学广泛实践的基础上，应用系统方法去解决其实践内容的工程技术。按钱学森教授所建立的系统科学体系(见图 2.1.1)系统工程的基础理论是运筹学、控制论和信息论等组成的一类技术科学以及为其提供计算方法的计算科学。

2.3.1 系统工程的基础理论

2.3.1.1 运筹学

运筹学是 20 世纪 40 年代发展起来的一门科学，它是管理系统的人为了获得系统运行的最优解而使用的一种科学方法。

这里所说的运筹学，沿用了第二次世界大战时出现的名词——运筹学。当时的运筹学包含了一些我们今天所说的军事系统工程的内容，故又叫军事运筹学。现在的运筹学属技术科学，它不包括军事系统工程的内容，只包括系统工程的特有的数学理论：线性规划、非线性规划、博弈论、排队论、库存论、动态规划等内容。

目前，有些人把运筹学和系统工程混淆起来，甚至有人认为二者是一个内容，这是错误的。造成这一错误认识的原因是对运筹学概念的理解，运筹学这一概念，国外与国内有不同的解释，国外的运筹学有点与国内系统工程相似的内容，因此，被称为“狭义系统工程”，而国内运筹学是按钱学森教授的说法，将其定义为数学方法与理论。由系统科学体系可见，运筹学和系统工程是既有联系，又有区别的两类不同层次的科学。我们可以这样来理解两类学科的联系、区别和涵义：

(1) 运筹学是从系统工程中提炼出来的基础理论，属于技术科学；系统工程是运筹学的实践内容属工程技术。

(2) 运筹学在国外被称为狭义系统工程，与国内的运筹学内涵不同，它解决具体的

“战术问题”;系统工程侧重于研究战略性的“全局问题”。

(3) 运筹学只对已有系统进行优化;系统工程从系统规划设计开始就运用优化的思想。

(4) 运筹学是系统工程的数学理论,是实现系统工程实践的计算手段,是为系统工程服务的;系统工程是方法论,着重于概念、原则、方法的研究,只把运筹学作为手段和工具使用。

常用的运筹学方法包括以下几种:

(1) 数学规划。数学规划是在某一组约束条件下,寻求某一函数(目标函数)的极值问题的一种方法。如果约束条件用一组线性等式或不等式表示,目标函数是线性函数时,就是线性规划。线性规划是求解这类问题的理论和方法,它在企业经营管理、生产计划的安排、人员物资的分配,交通运输计划的编制等方面有广泛的应用,是目前理论上比较成熟、实践中应用较广的一种运筹学方法。

如果在所考虑的数学规划问题中,约束条件或目标函数不完全是线性的,则称为非线性规划。在实践工作中所遇到的大量问题一般都是非线性问题,用线性规划是难以解决的,这也正是线性规划的局限性。非线性规划是解决这类问题的理论和方法。这种方法在理论上不如线性规划成熟,但随着科学的发展和电子计算机的普及,非线性规划将越来越重要,它能比线性规划更准确、更严密地解决问题。

(2) 动态规划。这种方法是动态条件下,解决多阶段决策过程最优化的一种数学方法,它可使多维或多级问题变成一串每级只有一个变量的单级问题。适用于解决多阶段的生产规划、运输及经营决策等问题。

目前,动态规划还没有一套一般算法,只有一些特殊的解法。

(3) 库存论。物资管理是经营管理的主要内容之一。该理论主要研究在什么时间、以多大数量组织进货使得存储费用和补充采购的总费用最少。库存问题包括静态库存模型和概率型库存模型。其中静态库存模型实质上是无约束非线性规划模型的一种。

(4) 排队论。排队论是研究服务系统工作过程的一种数学理论和方法,是研究随机聚散的理论。它通过个别随机服务现象的统计研究,找出反映这些现象的平均特性,从而改进服务系统的工作状况。

(5) 网络分析和网络计划。研究网络图中点和线关系的一般规律的理论,叫网络分析。它是应用图论的基本知识解决生产、管理等方面问题的一种方法。网络计划是用网络图的形式解决生产计划的安排、控制问题的一种管理方法。常用的网络计划方法有关键线路法(CPM),计划评审技术(PERT)以及决策关键线路法(DCPM)、图解评审技术(GERT)等。

(6) 决策论。决策论应用于经营决策。它是根据系统的状态、可选取的策略以及选取这些策略对系统所产生的后果等对系统进行综合的研究,以便选取最优决策的一种方法。

(7) 对策论。对策论又称博弈论,是研究竞争现象的数学理论与方法。最早产生于第二次世界大战,用于军事对抗,后来扩展到各种竞争性活动。在竞争活动中,由于竞争各方有各自不同的目标和利益,它们必须研究对手可能采取的各种行动方案,并力争制定和选择对自己最有利的行动方案。对策论就是研究竞争中是否存在最有利的方案及如何寻找该方案的数学理论与方法。

2.3.1.2 控制论

控制论是20世纪40年代中期发展起来的一门技术科学，它是自动控制、电子技术、无线通信、生物学、生理学、统计力学等学科相互渗透的产物，它应用反馈的方法研究控制系统如何根据环境的变化来决定和调整自己行动的问题。它除了研究工程控制系统以外，还研究组织的管理系统，进行动态分析和动态控制、动态优化等问题。

控制论是1948年美国科学家维纳创立的。它的两个基本观点是：

(1) 一切生命和无生命系统都是信息系统；

(2) 一切生命和无生命系统都是反馈系统。

这就消除了生命系统与无生命系统的控制界限，提出了任何科学方法可以同样适用于有生命系统和无生命系统，其中包括社会系统和经济系统。

控制论可以分为三代：第一代为古典控制论；第二代为现代控制论；第三代为大系统理论。古典控制论研究单输入、单输出系统，以输出输入方程(传递函数)研究系统的运动规律。现代控制论引入了状态变量，它从系统状态变化的角度说明一个系统的运动，故可解决高阶问题。大系统理论主要解决大规模的复杂系统问题，因为这类系统无法用传统的方法来解决，因此需要采用一些特殊的方法如分解协调法来解决。

按控制论在各学科的应用，又有工程控制论、生物控制论，经济控制论和社会控制论等分支。

2.3.1.3 计算科学

计算科学是20世纪40年代以来伴随着电子计算机的发展而出现的一门新兴的科学。它主要研究为各种模型提供精确或近似解法的理论。其应用称计算技术，它为系统思想的定量化开拓了广阔的前景，成为系统工程不可缺少的一门基础理论。

在现代数学中，任何一个一般性的计算方法和计算程序等均可叫算法。一切可设想的数学问题其算法均分两大类：一是无算法；二是有算法。在有算法的问题中又分为有效算法和无效算法两类，无效算法是指计算时间与问题规模的增大呈指数曲线增大，有效算法是指计算时间只随问题规模的增大呈多项式函数曲线的增长。

计算科学就是为系统工程探索各类数学模型的算法和有效算法问题。

系统工程在实践的过程也提出一些算法课题来充实和丰富计算科学的内容。

系统工程这些基础理论在实践中是各有侧重的。运筹学主要处理静态问题，控制论主要解决动态问题，二者相辅相成，在计算科学的协助下成为系统工程的不可缺少的理论基础，指导着系统工程的实践。

2.3.2 系统工程的得力工具——计算机

仅有理论和方法，没有得力的工具，系统工程的实践是不可能的。计算机科学的发展为系统工程的实践提供了得力的工具，使复杂的计算成为可能。

计算机的发明是人类历史上的一次大的技术革命，以前的技术革命主要是减轻体力劳动，把人类从繁重的体力劳动中解放出来。计算机的出现则是把人类从繁重的脑力劳动中解放出来，因此意义更加深远。

在进行系统的规划设计、试验研究、制造使用和经营管理中，计算、数据处理、逻辑判断和模拟的工作量相当大，没有计算机是不可能进行的。系统工程发展到目前水平是和

电子计算机的发展分不开的，同时系统工程的发展也推动了计算机科学的发展。

当前计算机正向两个方向——超大型化和微型化方向发展，尤其是计算机微型化，使其进入各行各业，成为我们日常工作中的必备工具，这将进一步推动系统工程的实践。

在现代化企业的管理中，计算机已成为不可缺少的工具，它是信息处理和控制的中心，因此也可以说系统工程是计算机在企业管理中推广和应用的科学。

为了更好地了解计算机在系统工程中的作用，现研究人怎样利用计算机获取和处理信息，并在处理系统问题中发挥作用的。

一般在系统中计算机有如下功能：

(1) 数据的采集；

(2) 数据的处理；

(3) 信息的存储；

(4) 信息的检索；

(5) 信息的显示；

(6) 信息的传输；

(7) 系统的模拟和实验；

(8) 系统的控制等。

随着计算机科学的发展，以计算机网络为基础的管理信息系统将成为系统工程的重要领域之一。Internet / Intranet 和可视化技术的发展又为系统工程的应用提供广阔的前景。

人类与自然界的斗争已取得了丰硕成果并积累了丰富的经验，自然科学作为人类与大自然作斗争的经验总结已取得了相当程度的胜利，把自然科学的成果应用到社会科学中去同样也会取得胜利，系统工程就是这样一门科学。它除了把运筹学、控制论和计算科学作为自己的基础理论之外，每个专业的系统工程还有自己特殊的专业基础理论，如工程系统工程的专业基础是工程设计与施工，企业系统工程的基础理论是管理经济学、管理学等。

2.4 系统工程的研究方法和步骤

2.4.1 系统工程的一般研究方法

进行任何一项工作所使用的方法，总是取决于它的指导思想。系统工程的研究方法也同样取决于它的指导思想——系统思想。综合系统方法、反馈方法和信息方法就构成了系统工程处理问题的基本方法。

由于系统工程是人、设备和过程有机地、有秩序地组合于一定环境之中的工程技术，因此，系统工程就必须在系统思想的指导下，不仅要研究系统的组成部分，还必须研究各部分之间的关联；不仅要研究单一过程，而且还要研究事物发展的全过程；不仅要考虑技术因素，还要研究考虑社会、经济、环境、心理、生态等各种因素；不仅要研究工程问题，而且还必须研究其组织和管理问题；不仅考虑当前情况，进行静态的研究，还必须考虑长远的发展和变化，进行动态的研究；不仅考虑单一方案，还必须综合地考虑各种方案；不仅考虑最优解，还必须因地制宜地考虑满意解或次优解；不仅要研究物质、能量的流动，而且更重要的是研究信息的流动；不仅要考虑系统内在的联系，而且要考虑与系统有关的环境因

素和人的因素等。

总之,我们需要从以下4个方面着手研究一个系统:

(1) 系统是各组成部分的有机组合,是各子过程的有机组合。单独研究子系统和子过程并不能揭示系统运动的规律性,只有在确定子系统之间、子过程之间的互相联系的情况下,研究子系统和子过程,才能达到系统整体最优化的目的。

(2) 事物总是在不断发展和变化的,因此,不能孤立地静止地研究系统,必须动态地研究和探索系统发展和变化的规律性。

(3) 没有比较就没有鉴别,只有在多方案的分析和比较下,才能识别优劣,故在多方案的论证中选择最优的或满意的方案是系统工程处理问题的主要方法。

(4) 系统工程特点之一是数学理论和行为科学的统一,要充分考虑人在系统中的作用和地位,要充分发挥人的因素的作用,是系统工程处理问题时不可忽略的。

系统工程在处理问题时,经常采用数学模型和模拟的方法:

(1) 数学模型是在对系统进行分析的基础上,根据研究的目的,在一定的假设条件下用抽象的数学语言,概括系统的内在规律,从而确定系统状态的一种方法。

(2) 模拟是通过对系统的分析,用物理模型或数学模型对系统进行定量地描述,以体现系统发展变化规律性的一种方法。

这两种方法,我们在今后的章节中将做详细地介绍。

2.4.2 霍耳三维结构空间

在从事系统工程的研究中,长期以来逐渐形成了一套科学的工作方法和步骤,按照这样的方法和步骤,一般就可以了解一个系统,对于复杂的系统,按照这样的步骤就不会感到无从下手。然而,不同的研究对象因特点各不相同,所以采用的方法和步骤也不尽相同。为了研究适用于系统的一般方法和步骤,现介绍一般系统的处理程序,对于具体的问题,只能根据系统思想的基本原则、立场和观点,灵活地运用。

霍耳(Hall)把系统工程的研究方法和步骤用三维的迪卡尔坐标系来表示,我们把该坐标系统称为三维结构空间。

霍耳把工作进程或工作阶段叫做时间维;把在系统各阶段中的思维过程叫做逻辑维;把每个思维过程中所及的专业知识叫做专业维。这就组成了包括时间、逻辑、专业的三维结构空间(见图2.4.1)。

2.4.2.1 工作阶段——时间维

霍耳把任一系统由规划设计起到更新为止的整个寿命周期划分为7个阶段:

(1) 规划阶段;

(2) 初步设计阶段;

(3) 系统开发阶段(研制阶段);

(4) 生产阶段;

(5) 安装阶段;

(6) 运行阶段;

(7) 更新阶段。

这7个阶段构成了系统的生命周期,即系统由孕育诞生到被新系统所代替的全过程。

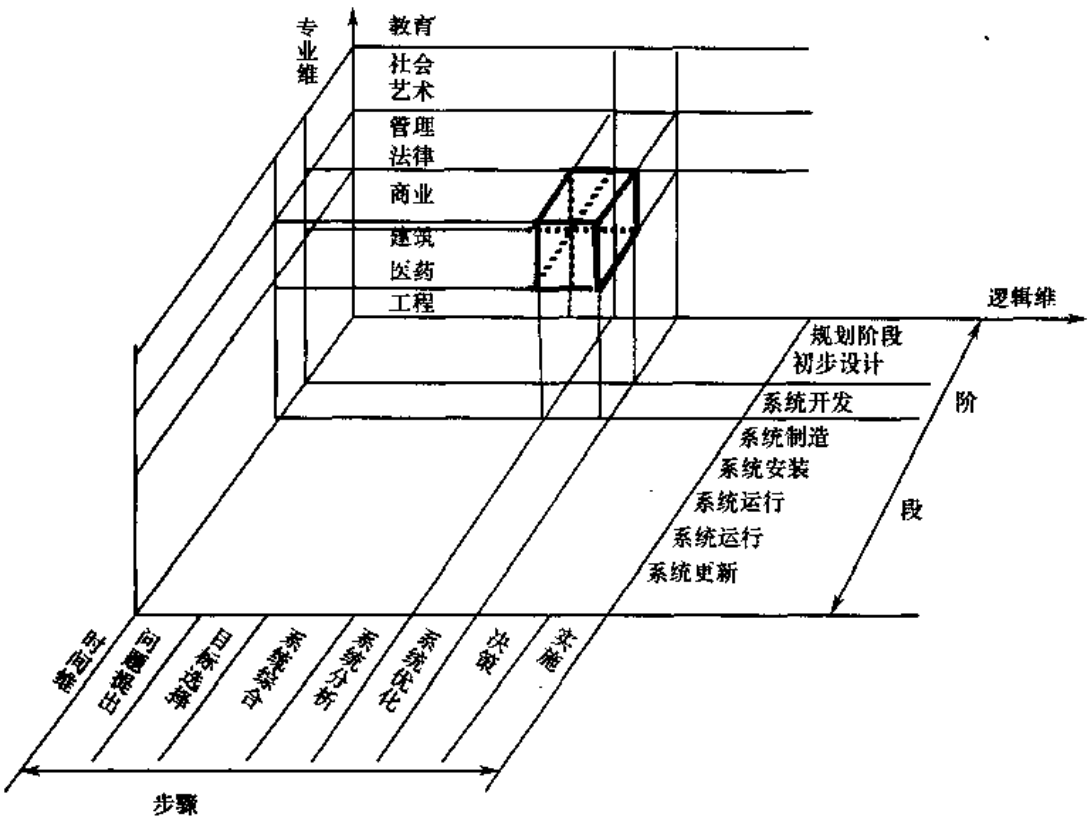


图 2.4.1 霍耳系统工程三维结构

2.4.2.2 思维过程——逻辑维

霍耳把每工作阶段的思维过程进一步地划分为 7 个步骤：

(1) 问题的阐述；

(2) 目标的选择；

(3) 系统的综合；

(4) 系统的分析；

(5) 最优化；

(6) 决策；

(7) 计划实施。

逻辑维包括了对任一阶段的全部思维过程。

1. 问题的阐述

该过程主要包括对系统性质的认识，了解系统的环境、目的、系统的各组成部分及其联系等。该阶段就是把系统的一切情况阐述清楚，属于定性地对系统进行研究的阶段。该步对任一系统，任一工作阶段都是必须首先进行的，它是为进一步分析和研究奠定基础的关键阶段。问题阐述的清楚与否，将直接影响系统的分析、综合、优化，以至今后的决策。如日本曾对打开坤表的销路进行研究，他们通过调查分析发现妇女戴表的目的，不仅是为了计时，而且讲时髦和漂亮是很重要的目的。因此改变了习惯上的从计时准确、坚固耐用方面对手表功能进行改进的作法，而在便宜、漂亮和式样新颖上下功夫，结果打开了销路。

2. 目标的选择

在问题阐述之后，面临的问题就是制定改进方案。在制定改进方案之前，要确定评价方案优劣的标准，以便以此为根据对方案进行选择。目标选择的正确与否是至关重要的。霍耳曾指出："选择一个正确的目标比选择一个正确的系统重要得多，选择一个错误的目标，等于解决一个错误的问题，选择一个错误的系统，只不过是选择一个非最优化的系统"。霍耳的这段话指出了选择目标的重要性。如第二次世界大战时期，由于德军的轰炸，使商船受到威胁，有人建议在商船上安装高射炮来保护商船，但有人认为，商船安装高射炮，飞机击落率仅为4%，故反对安装。但有人提出，以飞机击落率为目标是错误的，商船安装高射炮的目的是保护商船，应以商船被击沉率为目标。经统计，商船安装高射炮后，击沉率由原来的25%下降到10%，故以安装为宜。可见目标选择的重要性。

3. 系统的综合

在目标选定之后即可拟定方案。该阶段需根据问题的性质和所确定的目标提出几套方案并确定每套方案的参数，以便从中选择较好的方案。在提出方案的时候要打破传统观念，要集思广益，群策群力，尽量挖掘实现系统的所有可能方案。

4. 系统的分析

系统的分析是对系统综合中所有备选方案进行模型化的过程，即把每个方案的参数与目标结合起来建立模型的过程，是通过比选，进行方案选择的过程。

5. 最优化

最优化过程是处理模型的过程，是对所建立的模型求解，为决策者提供决策依据的过程。该过程经常使用各种单目标或多目标的最优化方法。

6. 决策

一般地说，最优化过程即可得到最优化方案作为决策。但事实上，任何定量过程都不会十全十美，在处理问题时，涉及大量人的因素、社会因素和各种不确定因素，在当前科学技术条件下，有些因素尚不能准确定量，因此要求决策者发挥自己的决策能力，在充分考虑定性因素的情况下，参考最优解，做出最后决定。决策过程正是体现定性因素在处理系统问题时的重要作用的过程。

7. 计划实施

决策之后，需把方案的详细实施步骤和内容，变成切实可行的计划，然后下达执行。

2.4.2.3 专业维

霍耳把系统工程处于某阶段、某一思维过程中所涉及的专业知识按照定量化的难易程度由下至上排列，其顺序是工程、医药、建筑工程、商业、法律、经济管理、社会和艺术等专业知识。

霍耳的三维结构空间是当前比较清楚地说明系统工程研究方法和步骤的模式。在所有工作阶段都包含着不同的思维过程和专业知识，因此以一向量(工作阶段、思维过程、专业知识)表示出系统工程所处的任一位置，就是霍耳三维结构空间作为系统工程的方法和步骤的本质。

2.4.3 三阶段法

除霍耳归纳的三维空间之外，系统工程的研究方法和步骤还有日本开发的三阶段法。

该方法把系统的研究分为三个阶段，每个阶段又分为前后两个子阶段，即：

1．系统开发阶段

(1) 系统开发的计划阶段；

(2) 系统开发的实施阶段。

2．系统建造阶段

(1) 系统建造设计阶段；

(2) 系统建造实施阶段。

3．系统运行阶段

各阶段的目的、性质、主要活动、主要活动成果及其活动依据等见表 2.4.1。

表 2.4.1　三阶段法

阶段	系统开发阶段		系统建造阶段		系统运行阶段
	开发规划阶段	开发实施阶段	实施计划阶段	建造实施阶段	
各阶段的目的和性质	讨论系统开发对象及开发必要性，制订开发规划方针和规划计划书，以明确开发系统的目的、目标和要求事项。 制定说明书和开发计划书	根据系统的要求、说明书和开发计划书，分析系统的功能、环境条件、费用、效果、实现可能性等，并编制系统设计书、建造基本计划书、实施设计计划书	根据系统设计书、实施设计计划书编制建造设计书，建造实施说明书	根据建造说明书和建造实施计划书，进行最合理的、有效的系统有关设备的制造，同时讨论运行方法和维护方法	根据已给定的运行、维护方法以及有效地、合理地运用和维护，同时根据运行和维护结果，讨论对系统的修改和调整
各阶段的主要活动	1．调查研究； 2．开发规划方针和计划的决定； 3．系统要求说明书的编制； 4．系统开发规划书的编制	1．掌握上一阶段成果； 2．系统的功能、环境条件等的分析讨论。 3．系统设计书，制造基本计划，实施设计计划书等的编制	1．掌握上一阶段成果； 2．建造设计书的编制； 3．建造实施计划书的编制	1．系统有关设备的制造； 2．系统运行维护方法的编制； 3．培训运行维护人员	1．系统的运行维护和管理； 2．掌握系统运行情况； 3．运行维护结果报告的编制
各阶段的最后成果	1．系统的要求说明； 2．系统开发计划书	1．系统设计书； 2．制造基本计划书； 3．实施设计计划书	1．建造设计书； 2．建造实施计划书	1．建造系统的硬件； 2．已经确定的系统设计与建造设计； 3．运行维护说明书；运行维护管理说明书； 4．设备运行计划书	1．系统运行结果； 2．系统的维护与保养结果
保证各阶段活动的依据	1．开发规划方针； 2．开发规划计划书	1．系统的要求说明书； 2．系统的开发计划书	1．系统设计书； 2．建造基本计划书； 3．实施设计计划书	1．建造设计书； 2．制造实施设计书	1．已经确定的系统设计和建造设计； 2．运行维护说明书；运行维护管理说明书； 3．设备运行计划书

注：该表引自《系统工程学》，王众托编。

2.5 系统工程的应用和发展

系统工程是20世纪60年代以后才开始兴起的一门科学，它是以系统思想为指导解决社会实践问题的一种方法。然而系统的概念却由来已久，我国古代就有很多用系统思想处理和解决系统问题的实例，他们的成就为系统工程树立了很好的榜样。但由于当时科学技术并不发达，因此还不能叫做系统工程，只能叫系统思想的应用。

2.5.1 系统思想的应用

2.5.1.1 都江堰水利工程

举世闻名的四川都江堰水利工程是战国时期秦国太守李冰父子二人主持修建的。都江堰由渔咀岷江分水工程、飞沙堰分洪排沙工程和宝并口引水工程等三大主体工程巧妙结合而成，这三大主体工程和120个附属渠堰工程形成了相互联系的有机整体，缺一不可。没有渔咀岷江分水工程就不能把岷江水分为内江和外江，没有宝并口的束水作用，就不会形成涡流，泥沙就过不了飞沙堰，宝并口就要淤塞，内江之水就无法流入川西大平原。正因为都江堰这一整体发挥了三个孤立部分所发挥不了的作用，才能分导岷江急流，化害为利，灌溉500多万亩良田。都江堰的规划、设计和施工的科学水平和创见用今天的系统方法衡量也是毫不逊色的。

2.5.1.2 皇宫修复工程

皇宫修复工程是宋《梦溪笔谈》中记载的一个小故事。说的是北宋真宗时期，皇宫失火，丁谓主持修复。其中清理废墟、取土烧砖、运输建筑材料是三大主要工程，在当时的交通技术情况下，这些工作是相当繁重的。为了在最短的时间内，用最少的劳动消耗修复皇宫，丁谓提出一个一举三得的方案：在皇宫前大道挖沟取土烧砖，解决了取土烧砖问题；引汴河水入沟，由汴河从水路运入木材、石料等建筑材料，解决运输建筑材料问题；修好皇宫后，把碎砖废瓦等回填入沟，修复原来的大道，解决清理废墟问题。该方案3个环节环环紧扣，缺一不可，真可谓“一举三役济”“省费以万计”，又快又省地完成了重建皇宫的任务。这一举三得的施工方案不正是系统思想的结晶吗。

上述两个实例是我国古代劳动人民运用系统思想解决实际问题的范例，同时也证明了系统思想由来已久。

2.5.2 近代系统思想的应用

2.5.2.1 泰勒创立的科学管理法

被誉为科学管理之父的泰勒，通过研究把产品生产过程分为三个方面进行分析：

(1) 工序分析，解决了工序组成的合理性问题；

(2) 动作分析，解决完成该工序的合理动作问题，如工作地的布置、两手动作的配合等，在研究相同工序上工作的若干工人动作的基础上，总结出一套最合理的动作，拍成电影，重新训练工人，从而提高了劳动生产率；

(3) 时间分析，解决了动作时间消耗的合理性问题，标定出各种工作的时间定额。

泰勒的研究方法体现出把产品生产过程作为一个整体，从整体到部分的研究方法，他

从整体的合理组成部分以及部分之间的关系着手，从各种方案的组合中选出最优方案，这就是最优化思想。因此泰勒的科学管理方法是系统思想在管理中的应用。

2.5.2.2 曼哈顿计划

曼哈顿计划是美国研制第一颗原子弹的计划。

1940年，爱因斯坦等著名科学家向美国政府建议研制原子弹，美国总统罗斯福采纳后，请理论生物学家奥本海默来组织领导这项工作。他动员了25000名各种专业的科学家和工程师，组织12000名工人进行全面合作，用了三年半的时间完成了这项工作。组织这么多人，进行这种史无前例工作是很困难的。他在执行计划的过程中，从整体出发，把研究课题分解成大量的小课题，组织很多小组负责各个课题的研究工作，同时他非常注重各课题之间的联系和它们之间的等级关系，使全部课题组合起来能达到整个计划的最优结构。在生产原子弹燃料这一中心项目上，奥本海默组织大家研究讨论，提出了六种方案，经过激烈的讨论，相持不下。他确定一个原则，首先要保证按时完成任务，其他皆是次要的。因此他根据可靠性理论中可靠性低的原件可组成可靠性高的系统的原理，决定六种方案同时试验，结果按时得到了生产原子弹所需的铀，于1944年5月第一颗原子弹爆炸成功。这说明奥本海默在组织这项计划时，时刻从系统的总目标出发来处理问题。

这项计划是大规模组织起来完成一项军事科研生产任务的著名实例。在整个计划的实施过程中，奥本海默始终从整体出发，为实现整体目标，协调各部分的行动，最后胜利地完成了任务。

当时，还没有系统工程一词，但奥本海默的工作却体现出系统方法的本质。国外都认为这是系统工程方法成功的起点。

2.5.3 现代系统工程的应用[①]

2.5.3.1 阿波罗登月计划

阿波罗登月计划是美国继曼哈顿计划、北极星计划之后，在大型项目研制上运用系统工程取得成功的一个实例。阿波罗计划的全部任务分别由地面、空间和登月三部分组成，是一个复杂庞大的工程计划。它不仅涉及到火箭技术、电子技术和冶金、化工等多种技术，还需要了解宇宙空间的物理环境以及月球本身的构造和形状。完成这个计划，除了要考虑每一部分间的配合与协调外，还要估算各种未知因素可能带来的种种影响。这项计划涉及到40多万人，研制的零件有几百万件，耗资3百亿美元，历时11年之久。这个计划的成功，关键在于整个组织管理过程采用了系统工程的方法和步骤，其实施过程如下：

1. 建立组织管理机构，明确职责分工

为了完成登月活动，首先确定所需的组织形式和管理原则。国家宇航局设立了阿波罗计划办公室主管全部工作，在该局附属的3个研究中心分别设立了阿波罗项目办公室，受该局设计办公室的领导，负责分管的任务。

计划和项目办公室的职能有：

(1) 项目的计划和控制；

(2) 系统工程；

---

① 本部分应用的实例摘自《全国管理科学研究班教材》(内部发行)。

(3) 可靠性和质量保证;

(4) 试验;

(5) 操作实施。并分别设立主管部门。

为适合于研究和研制工作的情况,整个管理工作分成5个方面:确定计划的基本要求,量测性能、分析和评价、控制和指导变化、动作和反馈。在整个管理过程中和5个职能部门内,经常考虑和处理的变量是工程进度、成本费用和技术性能。

2. 制定和选择方案

计划办公室成立后,首先为实现登月选择飞行方案,提供了三个备选方案:

(1) 直接飞行,使用新型运载火箭;

(2) 地球轨道交汇,使用土星运载火箭分别发射载人航天飞行器和液氧贮箱;

(3) 地球轨道交会,使土星运载火箭一次发射载人航天飞行器和登月舱。

对三个方案分别在技术因素、工作进度、成本费用和研制难易度等方面权衡利弊。结果认为第三个方案能确保在最短期间内,最经济地完成了阿波罗的全部目标。

3. 组织管理过程

飞行方案确定之后就开始其他的计划和管理工作。

(1) 确定该计划的基本要求。阿波罗计划采用了"工作细分结构"的系统分析方法。把整个计划由上而下逐级分成项目、系统、分系统、任务、分任务等6个层次。这样做的优点是:可确定计划的所有分支细目及其相互关系,明确哪个部门负责什么工作;可作为绘制计划评审网络图的基础,保证所有分支细目都包括在工作进度内;可作为编制预算的基础,把预算要求、实际成本费用与具体工作成果三者联系起来。总之,它为管理人员把整个计划的进度、财务和技术三方面的要求联成一个整体提供了共同的基础。

为了明确工程进度的要求,阿波罗计划采用了计划评审技术,以形成各主承包商与政府之间的进度管理系统。

(2) 性能监测。即按照阿波罗计划的基本要求,审查整个计划的进展情况。为了及时掌握整个计划的情况,要经常取得各职能部门的质量管理和进度的信息,在各职能部门制定了报告的具体要求,建立了"管理信息与控制系统"和一系列"设计审查及产品检查"制度。"管理信息与控制系统"将所需要的各种性能的信息绘成100张左右的图表,经过整理分析后,每月上报计划负责人,以使他能洞察全局,集中力量抓薄弱环节。在阿波罗计划的全部过程中,连续进行各种审查和检查,如初步设计审查、关键设计审查、设计检定审查、首次产品结构检查、飞行合格检定等。

(3) 分析评价。主要工作是对性能数据进行评价并归纳出必须及时采取行动的问题,向计划负责人报告。阿波罗计划中应用了成本相关分析法等,以便保持各个工作部门相互平衡和保证完成主要目标。成本相关分析法就是分析工作进度与成本费用之间的关系,看花的钱是否得到预期工作量,绘出今后费用增长率变化的曲线,以估计完成计划所需的总费用,它可用来比较几个主要承包商完成的工作量,找出哪些环节上费用未能产生预期的工作效果。

(4) 对系统的控制。研究计划往往会发生变化,因此必须进行系统地控制,以保证全部计划的实现。凡是影响最终产品的形式、装配、功能等的变化,都必须由规定的某一级决策机构批准。如在结构上的改变,分别由办公室的结构控制委员会批准。为了控制成

本，每个季度审查各个部分的实际成本，划出超支动向，查出薄弱环节，必要时调整财务计划。阿波罗计划由于强调财务管理和成本估计做得较好，财务支出始终保持相对稳定，平均每年实际成本比预计成本增长不到1%。

(5) 保证整个管理结构在日常活动中履行各种指令、规程、程序并为下一轮性能测量提供反馈数据资料，目的是让各级负责人明确自己的职责范围，做到事事有人负责，把实干的人选安排到合适的岗位上，并赋以权力。在日常活动中召开计划审查会，以促进管理行动的有效执行。

总之，阿波罗计划的成功是依靠所采用的管理系统和管理工具——计算机，而更主要的是采取了从整体出发并面向整个系统的综合管理方式。

2.5.3.2 韩国国民经济的模型

1964年—1965年，韩国政府准备采取措施打破经济停滞局面，决定进行一系列货币、财政和贸易政策的改革，并在第二个五年计划实现这些意愿与措施。韩国当时的经济情况很糟糕，主要问题是国内积累很少，外汇不足。当完成计划纲要时，预计年增长率可望超过8%，为了稳妥起见，最后确定为7%，但强调尽可能超过这个数字。

韩国得到了国际发展署的支持，该署派顾问参与模型的制定，其工作步骤如下：

1. 组织措施

为了更好地协调计划的制定和执行情况，韩国政府进行了改组，成立了经济计划部，集中管理计划与预算工作，由一名副总理领导，计划分析工作由该部与国际发展署共同完成。

2. 目标的制定

计划工作的目标是：

(1) 确定社会的经济目标；

(2) 计算一条能快速实现这些目标的可行的增长道路；

(3) 找出主要的约束条件，并设法减轻这些约束；

(4) 制定详细的投资计划和公共政策，保证经济沿着理想的道路发展。

计划工作是连续进行的，每年还需做许多具体计划，以不断审查与修改计划。鉴于当时韩国人普遍感到贫困无望，因此，用简明方法制定改革目标并辅以技术分析。这在当时是很重要的一个措施。

计划工作的目标是制定协调一致的规划和政策，而不是最优化的规划、政策。

3. 制定模型

采用的主要定量的工具是：

(1) 一个投入—产出平衡模型；

(2) 一个中期的大范围经济模型；

(3) 一个近期的稳定化模型；

(4) 一个用于钢铁与石化部门的综合线性规划模型；

(5) 一个作地区性平衡用的线性规划模型。

投入—产出模型包括43个国内生产部门、34个进口部门。投入—产出系数是以1965年价格为基础，并根据技术预测做了修改。结构系数是以抽样调查数据为基础，加上了详细的生产成本、固定资产、能力、输出等数据。这些数据由许多部门的“工业委员

会"提供,委员会成员包括工程师、财政分析人员和经济学家。委员会的职责是审查投入—产出预测。例如,钢铁部门的生产估算是由消耗部门来审查;钢铁工业的投入、需求是由每个部门的专家来审查。

预先确立了一套准则,以便对6个方案进行选择,每个方案都是根据不同的假设设计的。

4. 五年计划的战略

第二个五年计划战略的主要内容有四项:扩大出口、国内资本的动员、有效地使用人力与技术以及经济持续稳定。

中期的投入—产出模型计算结果说明,出口与国内节约是增长的约束条件。计划人员相信阻碍快速扩大出口的惟一因素在于供应方面;投资计划对于实现出口、就业与技术战略是极为重要的。同时还制定了一项财政计划保证财政稳定,并制定更为严格的节约约束条件。投资分配与财政调节是第二个五年计划中头三个年度资源总预算分析工作的焦点。

五年计划为每个部门制定了任务,第一个年度资源预算中列出:纸张生产,本部门投资即将大量增长,除了正在进行的计划外,极需新的投资;有色金属,有关有色金属矿石的冶炼计划的建议已经足够了,但是加工设备的计划建议还很需要。这些都是根据对投入—产出模型的预计和已经做出的重大投资进行了详细研究、计算和修改后才做出的。为了完成43个部门的投资估算和150个生产单位的预测工作,需要做出不寻常的努力来收集新的技术数据。

在第二个年度资源总预算了中提出了多种资金流动方案以供选择,它们对于货币政策和外部变数都做了不同假设。每项计划都有附件说明了通货膨胀的影响。这些研究的基本工具是稳定化模型。

5. 经济成果

五年计划及其有关的分析与讨论是促使韩国经济高速发展的重要因素,计划说明:

(1) 高水平的外援是必要的;

(2) 外贸的重要点转变了,从减少进口到扩大出口,重点放在劳动力密集的工业;

(3) 年度计划预算方法有助于将资源用在卡脖子的地方,如电力、水泥等;

(4) 贸易、投资、货币政策的相互依赖关系已被政府当局列为决策工作中的一部分。

在五年计划期间及其以后时期,韩国经济每年保持10%的增长率(实际增长率),外汇(美元)储备每年增长40%(1973年—1974年石油提价以前)。

2.5.3.3 墨西哥对农业问题进行的系统研究

墨西哥政府与世界银行共同用五年时间(1970年—1974年),制定和运用农业部门的计划模型。整个工作分为三个阶段。第一阶段到1972年为止,集中力量研究基本方法并检验计算模型的示范性数据,1973年出版了一本关于墨西哥农业问题的书。农业模型命名为"雨神",后来成为整个计划的代号。

第二阶段开始时,墨西哥政府扩大工作范围,邀请世界银行的人士参加合作,继续研究方法问题及其在农业方面的应用。在此阶段政府发布的各有关文件都是以此模型分析为依据,其中的一个文件协助埃切维里亚政府制定了农业指导方针。

第三个阶段的主要工作是由总统府的一个工作组进行的，并继续与世界银行进行合作，“雨神”模型及其子模型在墨西哥政府和世界银行的计算机上运算了3年多时间。后来罗马的食物与农业组织邀请世界银行及各国系统分析人员推广这种方法，制定一个更为广泛的计划，以便用于世界其他地方。

1．问题与模型

农业问题十分复杂，许多因素相互影响，产品定价政策能够大大影响就业率、收入分配和外贸平衡问题。墨西哥各地的生态特征各不相同，农产品丰富多采，遇到了复杂的农业政策问题。1968年—1970年，墨西哥官员怀疑农业能否继续满足国内消耗需要和提供国家的大部分外汇。他们看到开垦荒地和搞水利，代价昂贵。农业开始出现长期的增长缓慢情况，同时他们为农业能否吸收大量新劳力表示忧虑。他们提出了下述问题：农业能发展多快？就业可能性如何？改变定价政策、各种谷物的构成和生产技术，将发生什么影响？

为此他们拟制了“雨神”模型，模型描述了各地的生产条件以及国内外农业产品价格和贸易量对墨西哥农业的影响。在生产方面，模型给出了各地区的小经济范围的农场生产价格和输入数据。在市场方面，利用了估算的消费者需求量、进出口和运费等参量，同时利用线性规划描绘出了合适的市场稳定点。

模型中生产问题，基本上是利用了农业方面的资料。经济问题，如谷物供应函数是靠模型求解的，而不是模型的一个输入。这样，“雨神”模型就有助于把工程数据转换成经济关系，发挥了系统分析与运筹学的主要优点。

模型包括了33种产品并反映了它们供求的相互影响。在墨西哥的大多数农作物区都种植多种谷物，因此鼓励种植某种谷物肯定会影响其他谷物产品，线性规划法很适合于用来解决这类相互影响的问题。即能在有限的土地和其他资源限制下，对各种谷物和技术进行最优选择。如果把“雨神”模型内的有关“地区生产”的子模型独立出来，就能对地区性的方案进行更为仔细的选择。

“雨神”及其子模型是较为详尽的，但又很灵活，做些修改就能进行不同种类的分析，例如，模型在谷物每年的耕种周期内，能按月给出使用土地、灌溉、劳动力方面的资料，但有时也能每两周给出水的利用数据。

因此，“雨神”模型能同时求解农业方面许多总的特性，如总产量、价格、劳动收入、外贸等，甚至结合谷物、技术和地区方面的数据及其互相作用等，给出更为具体的输出数据。

从数学意义上说，“雨神”是个最优化模型。然而，它只是为了保证模拟某些问题，如市场的最佳的竞争和垄断问题。

2．具体的用途

“雨神”模型是在总统府的一个工作组和后来成立的农业协调委员会共同管理下，用于下列工作：

(1) 制定总的农业政策，包括产量、农业人数、外贸和有关的投资等问题；

(2) 制定玉米、小麦等谷物的价格政策；

(3) 制定有关项目的价格，如农业机械、水的价格等；

(4) 出口战略；

(5) 对灌溉工程进行鉴定。

模型给出的某些数据，在政策讨论的过程中起到了最重要的作用。例如模型数据说明如果国民经济总产值增长率目标是8%的话，农业增长率必须超过5%。对可耕地，灌溉工程的必须增长率也做了相应的估算，同时表明这样的增长率还不能吸收新的进入农业生产的劳动力，为此：

(1) 提高农业机械价格(相对于农业劳动力价格)，以减缓劳力转移；

(2) 鼓励出口花费劳力较多的谷物；

(3) 紧缩畜牧区，以便腾出土地耕种谷物，因为每亩谷物地能吸收更多劳力。

为了增加出口，计算结果说明，必须改变出口的谷物品种。最后还计算了谷物对价格变化的影响，以及价格变化对其他谷物的影响，从而协助政府制定了一份新的谷物价格表。

必须指出的是，墨西哥政府中如果没有训练有素和积极热心的人士，"雨神"模型就不能得到如此的应用。

2.5.3.4 印度巴罗达市公共汽车

1. 线路与时刻表拟定问题的提出

印度的巴罗达是一座发展迅速的城市，城市面积78km^2，居民50万，在市郊正在兴建一所炼油厂，一所化肥厂以及一个石油化工联合企业，预计城市的未来发展将更快。虽然城市正在不断扩大，把愈来愈多的工业单位圈了进来，但其工商业中心仍在旧城区，那里居住着70%的居民。

古吉拉特政府陆运公司经营着巴罗达的公共汽车事业，整个车队有100辆车，行驶路程达3000km。1972年，这些公共汽车每日载容量为138 000名，为可容载客量的47%。过了几年，该公司发现，路线行程增加很快，但乘客量与可容量之比不断下降，而高峰与平均运输量之差增加很快。如何把有限的车辆部署好，特别是在高峰阶段，就成为一个很严重的问题。

1972年该公司将此问题提交给一家私人印度咨询公司的运筹学小组，要求研究现有路线、高峰运输的合理化计划，以及利用现有车队来拟制时刻表和发车次数。

2. 目标的确定

最后确定的目标是拟制一套方法来制定、试验和鉴定各种行车时刻表，从而使公司有利可图。同时还决定这份时刻表是一份满意的而不是最佳的时期表，因为最佳化研究不仅费时，在许多情况下也得不偿失。

3. 模型的拟制

公共汽车时刻表的问题一般分为如下几个方面：

(1) 确定需求量：即每天不同时间各站上下的乘客量；

(2) 制定行驶线路：根据需求量，运用合理化计算，定出路线；

(3) 制定行驶的频率：即每条建议的路线的行驶频率；

(4) 调配车辆：根据路线和频率，调配车辆；

(5) 确定车上工作人员。

寻求一个最优解方案要求同时对上述项目综合调查研究，但是一个满意解方案只需分别研究这些项目后即能得出。运筹小组与公司双方同意采取后者。

4. 收集资料

把汽车站分为三类:一类站为整天都有大量乘客上车;二类站为只在高峰时期大量乘客上车;三类站为其余的车站。车站的选择和分类是与公司工作人员合作进行的。

利用了两天时间对每个站的乘客需求量作了详细的观察,每隔5min把排队人数作一统计并记录下来,并记下每辆车的到达时间、终点站等资料和乘客的目的地。

5. 资料的使用

将原始资料归纳后,就能得到每站要求上车的乘客量(以每小时为基础),各站下车人数。研究成果是一个起点—终点矩阵,也就是一张双程行车表,说明从各起点到终点每小时内要求运走的人数。

6. 制定公共汽车路线

对上述图表进行研究后,标出载客容量繁重的各段线路,并结合其他各段情况,制定公共汽车路线。

7. 确定发车频率

这是使用一台小型电子计算机确定的,其过程如下:输入计算机的有行驶路线、各站之间的行驶时间、上下车需求量。路线确定后再计算每隔5min每站乘客总和。然后根据下列准则确立路线:

(1) 路线的载各量必须在最低水平以上,使载客量与可容载客量之比达到可接受的水平;

(2) 乘客等车时间必须小于公司规定的服务标准。

这种方法将给出过多的路线,因为一条线上乘客可以乘坐许多路的汽车。因此使用了一台大型计算机来消除这些重复因素。第一次计算是一种概算,第二次给出了详细计划。

8. 制定时刻表

有了路线和旅程表后,就必须调度车辆,以确定各线路所需的公共汽车数,这也是用计算机完成的,其要求是把闲置时间和跑空车的现象减至最低程度。

9. 模拟

使用了各站的随机需求量来模拟系统的工作情况,以检验模型的性能。

10. 结论

这不是一个最优解,但对该公司有重大意义,因为它比较容易重新制定时刻表,并能满足主要的要求。计算机的初步计算说明能获得较大的好处,载客量与可容载客量之比可望增加11%。其他的经验表明还可增加公共汽车利用率,车辆可以减少10%,巴罗达市采纳了本研究的建议,并获得预期的结果。

2.5.4 系统工程的应用范围

系统工程的应用范围非常广泛,大至国家系统、社会系统、产业系统、各种工业系统、各种服务系统等,小至企业的产品开发、经营计划、生产管理、库存管理等。可以说,它能应用于解决一切部门复杂而又困难的项目的规划设计问题,管理控制问题以及生产运行问题。

日本学者秋山穰和西川智登二人将系统工程的应用范围概括总结为表2.5.1,由表可见,系统工程在各行各业都有用武之地。

表 2.5.1　系统工程的应用范围

应用范围		应用例子
自然对象系统	宇　宙	宇宙开发、宇宙飞行、通信卫星等
	气象、灾害	天气预报、地震预报、防灾、台风、洪水、震灾对策、人工气象开发等
	土地、资源	土地开发、海洋开发、资源开发、能源开发、太阳能开发、地热开发、潮力开发、治山治水、河流开发、农业灌溉、水库流量控制、土地利用、造田、环境保护
	农、林、渔业	农业资源、林业资源、渔业资源、人工农业等
人体系统	生理、病理	生理分析、生理模拟、病理分析、病理模拟、病理情报检索等
	脑、神经、心理	思考模型模拟、自动翻译、人工智能、机器人研究、控制论模型、心理适应诊断、职业病研究
	医　疗	自动诊断、自动施疗、物理治疗、自动调剂、医疗工程、医院情报管理、医院管理、医疗保险、假手足、人工内脏等
产业系统	技术开发	新技术开发、新产品开发、技术情报管理、原子能利用、最优设计、最优控制、过程模拟、自动设计、自动制图等
	工业设施	发电厂设备、钢铁厂设备、化工设备、过程自动化、机械自动化、自动仓库、工业机器人等
	网络系统	电力网、配管分配、安全回路、控制回路、道路计划、情报网
	服务系统	铁道航空的座席预约、旅店剧院预约、银行联机系统、自动售票、情报服务等
	交通控制	航空管制、铁道自动运行，道路交通管理、新交通系统等
	经营管理	经营系统、经营模拟、经营组织、经营预测、需要预测、经营计划、生产管理、资财管理、在库管理、销售管理、财务管理、车辆分配管理、经营情报系统、事物工作自动化等
社会系统	国际系统	防卫协调、国际能源问题、粮食问题、国际资源问题、国际环境保护、国际情报网、发展中国家的开发等
	国家行政	经济预测、经济计划、公共事业计划、金融政策、保卫、治安警察、外交情报、经济情报服务、司法情报、行政管理、邮政、职业介绍等
	地区社会	地区规划、城市规划、防灾对策、垃圾处理、地区生活情报系统、公用计划、老年人、残废人对策、地区医药
	文化教育	自动广播、计算机辅助教学、文化教育情报服务、教育计划、自动检字、自动印刷、自动编辑等

第 3 章　系统工程的重要方法——模型化

3.1　模型与模型化

系统是由许多单元相互联系所组成的整体，系统的特性可由这些单元之间的联系推导出来。为了掌握系统发展变化的规律性，必须根据系统的目的，抓住系统各单元之间的联系，进行系统地考察与研究，其中最容易、最方便的方法就是模型化。因此模型化的方法成为系统工程最重要的方法之一。

3.1.1　模型

模型有各种各样的定义，其中较为普遍的定义为：模型是为了了解系统的结构和行为通过抽象、归纳、演绎、类比等方法，用适当的表现形式描述出来的仿制品。

我们日常生活和工作中经常使用模型，如建筑模型、汽车模型等实体系统的仿制品（放大或缩小的模型），它可以帮助我们了解建筑造型、汽车式样等；教学中使用的原子模型，可帮助学生形象的理解原子的结构；经济分析中所使用的文字、符号、图表、曲线等可为分析者提供经济活动运行状况及特征等信息。它们虽然描述形式各异，但都具有共同的特点：

（1）它们都是被研究对象的模仿和抽象；

（2）它们都是由与研究目的有关的、反映被研究对象某些特征的主要因素构成的；

（3）反映被研究对象各部分之间的关联，体现系统的整体特征。

由模型定义及特点可见，模型首先必须与所研究系统“相似”，也就是“像”，这种相似不是指形状上的“相似”，而是指本质上的“相似”；其次，模型必须有一定的描述形式，描述形式可以是形状的放大或缩小，但更普遍的是文字、符号、图表等；再次则是必须采用一套有科学依据的方法来描述。采用什么样的方法、怎样描述才能得到与所研究系统相似的模型则是系统模型化的内容。

3.1.2　模型化

系统的模型化就是建立系统模型。它是把系统各单元之间相互关联的信息，用数学、物理及其他方法进行抽象，使其与系统有相似结构或行为并体现系统这一完整统一整体的科学方法。

模型化之所以成为系统工程的重要方法，其原因在于：

（1）系统工程的研究对象是工程技术、社会、经济和生命等诸因素交织在一起的人、设备和过程的统一体，其中很多因素是难以定量的，因此需要应用计算机进行模拟分析，实现计算机的模拟就必须建立模型。

(2) 经济性是评价系统的重要指标之一,应用模型化的方法可达到少花钱、多办事的目的,从而实现经济上的节约。模型化的方法还可在各种不同的系统参数中选择最优的参数,在各种不同的方案中选择最优方案而不必对实际系统进行各种实验和调整,从而达到以较少的费用可靠地实现系统最优化的目的。

(3) 安全可靠。某些系统的实验和运行蕴藏着危险性,这使系统的实际实验和研究难度加大。用模型化的方法可避免各种危险而提出各种可靠的数据,为决策提供依据。

(4) 可对不能进行实际实验的系统进行研究。某些系统,如生态系统、国民经济系统、社会系统是不允许实验的,为了探索这类系统的运行规律,只能靠模型化的方法。

模型化是建立模型的过程,该过程如图 3.1.1 所示。

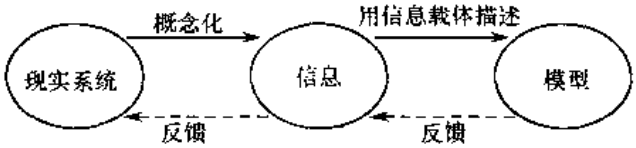


图 3.1.1 模型化过程示意图

由图可见,模型化的过程是对现实系统进行分析和观察,通过概念化获取信息,这是对系统的认识过程。对获取的信息经加工、处理,进一步深化认识后抽象出模型并用确定的形式进行描述,这是提高认识的过程。由于对系统的认识是逐步提高的,因此模型化的过程是认识—提高—再认识—再提高的过程。图 3.1.1 中的反馈即体现了这种再认识—再提高的过程。通过上述模型化的过程就可建立一个既反映现实系统的结构或行为,又能指导系统运行的模型,所以模型是源于现实系统又高于现实系统的人类思维的外在表现形式。

3.1.3 系统模型的分类

系统是多种多样的,对相同的系统可用不同的方法从不同的侧面按不同的研究目的建立不同形式的模型,因此必须对系统模型进行分类。系统模型分类方法很多,现介绍以下几种常用的系统模型分类方法。

3.1.3.1 按形态分

按形态可将模型分为实体模型和抽象模型。

1. 实体模型

把实体系统的功能和结构,以原型作为要素进行描述使其与系统原型基本相似的模型。它是系统原型几何尺寸的放大或缩小,从而体现系统的某些特性的模型。如教学用的原子模型、汽车模型、地形的沙盘模型等。但不是所有系统都可得到实体模型的,只是一些具有实物实体的系统才能建立实物模型。这类模型有直观、形象的特点,故又称形象模型。

2. 抽象模型

用概念、原理、方法等非物质形态对系统进行描述得到的模型。如用数学方法描述的模型、用逻辑关系描述的框图、用类比方法描述的类比模型等。这类模型的特点是从模型表面上已看不出系统原型的形象,模型只反映系统的本质特征,只是与系统在本质上相

似。它是经人类的思维活动在对系统原形的认识—提高—再认识—再提高的基础上高度抽象的产物，它是系统工程中经常使用的模型。这种模型又可分成以下几类：

(1) 数学模型。数学模型是用数学方法描述的系统模型。它是以字母、数字和各种数学符号对系统结构、特性以及内在联系进行数学抽象而建立的模型。它的主要特点是可通过模型的求解，即通过数学运算而得出系统运行的规律、特点及结构等，是系统工程中最常用的模型。如国民经济综合平衡模型、随机服务系统模型、可靠性模型、最优化模型等。

(2) 逻辑模型。逻辑模型是用图描述的体现系统运行逻辑关系或系统组成部分间逻辑关系的模型。它是系统组成部分的逻辑关系的抽象。其主要特点是明显的显示系统各部分之间联系，既可用于定性分析，又可进行定量计算或指示系统运行程序。如网络计划法中的 CPM 网络图、某种算法的计算机程序框图、计算机结构原理图、结构模型图等。

(3) 模拟模型。模拟模型是用物理属性描述的系统模型，它是用一个容易实现控制或求解的系统代替或近似描述一个不容易实现控制或求解的系统模型。因此这类模型既可用实体形式抽象，又可用数学形式抽象。用数学方法抽象体现系统物理属性变化的模型又叫数学模拟模型。前者如力—电压相似系统，后者如系统动态学模型等。

(4) 分析模型。分析模型指用曲线、图、表描述系统结构和特点的模型。它用线、图、表等形式鲜明地表示系统的结构、变化趋势等，一般用于系统状况分析之中，如直方图、变动曲线、雷达图等。

3.1.3.2 其他分类方法

(1) 按对象划分，系统模型可分为经济模型、社会模型、生态模型、工程模型等。

(2) 按研究问题的出发点划分，系统模型又分为宏观模型、微观模型等。

(3) 按用途划分，系统模型又分为预测模型、结构模型、过程模型、决策模型、性能模型、组织模型、行为模型、最优化模型等。

总之，系统模型可按不同标志进行分类，但无论按什么分类标志进行分类，在模型的实际应用中都必须符合系统的目的，都必须依据研究目的选择一种或几种模型。

3.1.4 数学模型的特点及分类

由于数学模型是系统工程中应用最多的模型，因此对数学模型的特点进一步阐述如下：

(1) 数学模型具有高度的抽象性，因此是较高层次的模型，它是其他一些模型的基础。

(2) 具有明确性，它可明确地描述各因素、各变量之间的关系。

(3) 可通过求解得出直观认识无法解决的问题。

(4) 可处理多变量、关系复杂的问题。

(5) 具有较高的精度。

(6) 适应性强。

(7) 调整性能好，可根据需要随时修改参数和变量之间的关系。

(8) 分析速度快，可用计算机实现自动化计算。

(9) 便于交流、便于信息的存储和加工。

其缺点则是缺乏直观性、对一些无法定量或难以准确定量的问题描述困难，缺乏形象性和实时感以及存在一些模型的建立和求解需较深的理论知识等问题。虽然如此，数学模型已成为人们解决问题的不可缺少的重要工具。

数学模型的分类如图 3.1.2 所示。

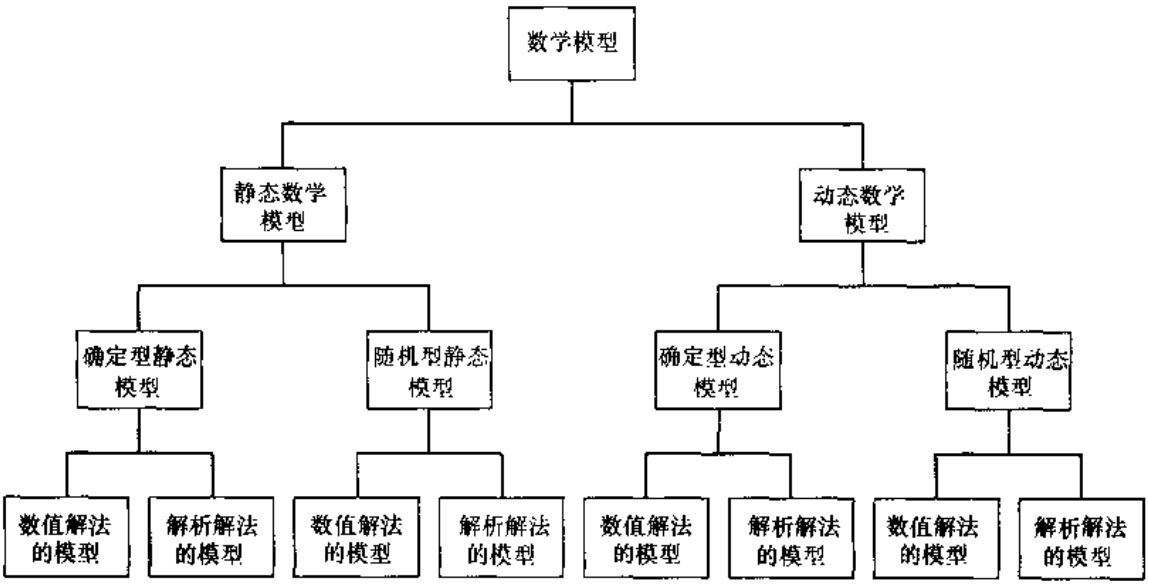


图 3.1.2 数学模型的分类图

3.2 系统模型化的基本理论

系统模型化的基本理论有“黑箱”理论、“白箱”理论、“灰箱”理论和数理统计与分析理论四种。

3.2.1 “黑箱”理论

对内部结构和行为不清楚的系统，依据可控因素的输入所引起的可观测因素变化的各种实验数据来确定系统运行规律，从而建立系统模型的理论和方法称之为“黑箱”理论。

“黑箱”理论即是将系统当作未知的“黑箱”，通过实验的方法得出系统运行规律的方法。它通常用输出输入方程(传递函数)来描述系统。

3.2.2 “白箱”理论

对系统内部结构和行为清楚的系统，应用各种已知的科学知识进行描述从而建立系统模型的理论和方法称之为“白箱”理论。

“白箱”理论即将系统当作一个已知的“白箱”，通过系统输入引起系统状态的变化，进而使系统输出变化的规律来描述系统的方法。它通常用状态方程描述一个系统。

3.2.3 “灰箱”理论

对系统内部结构和行为的主要部分清楚，其他部分不清楚的系统，采用已知的科学知

识建立模型，然后通过实验对所建的模型进行补充和修正，从而建立系统模型的理论和方法称之为“灰箱”理论。

“灰箱”理论是将“白箱”理论和“黑箱”理论相结合而建立模型的一种方法。该方法有较大的实用价值。

由“黑箱”到“灰箱”再到“白箱”，是人们认识系统的客观规律。一切系统都是由未知到了解一部分再到完全了解。至目前为止，一无所知的“黑箱”是很少的，尤其是系统工程所研究的系统。然而完全清楚的系统也是极少的，绝大部分系统都是介于“黑箱”和“白箱”之间的“灰箱”，因此近些年来科学家们都致力于研究“灰箱”建模理论，其中以邓聚龙教授所开发的灰色系统理论最为突出。他提出的五步建模思想与方法是灰色系统建立模型的重要方法论。该方法体现出建立系统模型时，由定性到定量、由粗到细、由灰到白的全过程。

邓聚龙教授所提出的五步建模思想与方法如下：

第一步 思想开发，建立语言模型

研究系统首先要明确目的、目标、要求和条件。而这些问题的明确首先要有思想的开发，然后将思想开发后的结果，用准确精练的语言进行描述，建立语言模型。

第二步 明确关系，建立网络模型

在语言模型的基础上，进行因素分析、原因与结果的辨识，以及关系的归纳分析。然后将原因与结果作为一个整体用方框图表示，就得到一个单元网络或一个环节的框图，如图 3.2.1 所示。

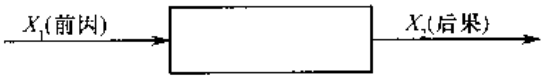


图 3.2.1 原因结果框图

在图 3.2.1 中，前因亦称该环节的输入，以 X_1 表示，后果亦称该环节的输出，以 X_2 表示。但系统是多个因素(变量)相互关联组成的，作为后果的因素可能是下一个环节的前因，作为前因的因素也可能是上一个环节的后果，有时还会相互穿插，交替影响。为此就可能得到图 3.2.2 所示的多环节框图，即网络图。

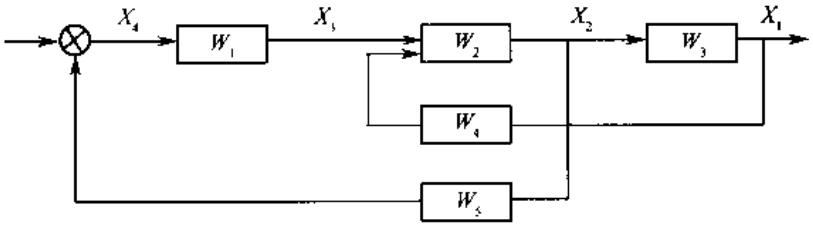


图 3.2.2 网络图

第三步 建立初步的定量关系，得出量化模型

得到网络模型后，搜集前因与后果间的数量关系。如 X_1 与 X_2 之间有比例关系，即

$$X_1 = kX_2$$

则在图 3.2.1 的框中填上 k，这就是量化模型。

第四步 研究原因与结果间的动态关系建立动态模型

上述量化模型只能说明因果间的简单量化关系，它不能说明原因作用于环节上以后，结果如何发展变化。前因如果随时间变化，后果又如何变，是增长还是衰减，变得快还是变得慢。显然这些问题的回答得依赖 X_1 与 X_2 的时间序列数据，通过输出输入的时间序列数据可以建立它们之间的发展变化关系，这称为动态模型。

第五步 模型的改进，建立优化模型

上述动态模型如果动态品质不能令人满意，则应采取适当的措施，改变系统参数、结构或加入新的环节，作这种优化处理后的模型称为优化模型。

完成上述五个步骤即完成了一个灰色系统的建模。关于灰色系统的建模理论和方法请参阅有关专著。

3.2.4 数理统计和分析的理论

对属于“黑箱”但又不能进行实验的系统，采用数理统计和分析的方法，应用统计规律建立系统模型的理论和方法称为数理统计和分析的理论。应用数理统计和分析的理论和方法建立系统模型是系统工程中最常用的方法之一，它最常用于系统预测模型的建立。

3.2.5 建立模型的原则

无论采用哪种理论建立系统模型，都必须符合以下原则：

(1) 系统模型是现实系统的代表而不是系统的本身。建立模型时，要抓住系统的本质行为、各部分之间的普遍联系，建立一个比系统简洁得多的又能反映系统基本特征而不是全部特征的模型。

(2) 系统模型要符合一定的假设条件。任何模型都要有假设条件，关键在于假设条件要尽量符合实际。假设条件依系统的研究目的而定，一般情况下，满足一定环境，为了特定目的的模型与系统全部特征并不吻合。因此，合理的假设是处理系统的重要前提，也是模型适用范围的界限。

(3) 模型的规模、难度要适当。所谓模型规模即指模型的大小，一般以阶次来反映，“大”的系统可建较大规模的模型，“小”的系统可建小些的模型。建立模型的目的是为了研究系统的特性，因此模型的规模应根据研究目的而定，只要能达到研究目的，应尽可能建立小规模的模型，这样可减少处理模型的工作量。

所谓模型的难度是指求解模型所应用的理论的深浅程度。所需理论较深，处理难度大，反之则小。因此所建模型的难度也应依据系统的研究目的，尽可能建立难度小的模型。

建立模型要注意“防止掉入过于简单的陷阱，又要防止陷入过于复杂的泥潭”。

(4) 模型要有代表性，要有指导意义。为建模型而建模型是模型化的最大禁忌之一，模型化的目的是处理系统，因此所建模型必须代表系统的普遍特性，要应用由特殊到一般的原理建立一适用面广，有指导意义的模型。

(5) 模型要保证足够的精度。因模型是系统的代表，故在建模型时要把反映系统本质的因素包含在模型之中，而把非本质因素排除在模型之外且使其不影响系统的特征或影响甚小。这就要求模型所反映的本质与系统的本质特征误差很小，即保证足够的精度。

(6) 尽量采用标准化的模型和借鉴已有成功经验的模型。建立模型是一件复杂的创造性劳动,标准化的模型和有成功经验的模型中凝集了前人和同行的心血和劳动,采用标准化模型和借鉴并发展有成功经验的模型既可节约劳动,又可丰富模型化的理论和方法。

3.3 模型化的程序和常用建模方法

3.3.1 模型化的程序

由模型化的定义可知,系统的模型化就是建立系统模型的过程,该过程也是一个系统。实现该系统的程序可用图 3.3.1 表示。

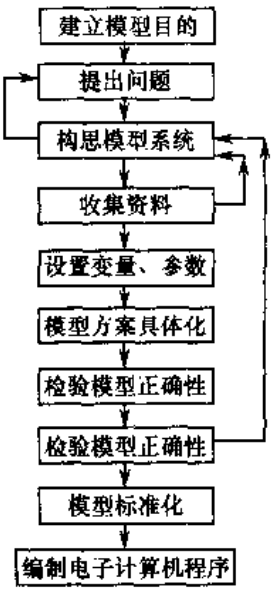


图 3.3.1 系统模型化程序

第一步 根据系统的目的,提出建立模型的目的。建立模型必须目的明确,它应明确回答"为什么建立模型?"等一类问题。如建立统计核算自动化系统中统计报表子模型,其目的就是实现统计报表自动化。

第二步 根据建立模型的目的,提出要解决的具体问题。该步应明确回答"解决哪些问题?"之类的提问,也就是将建模目的具体化。提出问题实质上是对系统中影响建模目的的各种要素进行详细分析的过程。如要实现统计报表自动化,就必须详细分析报表种类、核算过程、核算方法、数据来源等,从而提出需解决的问题。

第三步 根据所提出的问题,构思要建立的模型类型、各类模型之间的关系等,即构思所要建立的模型系统。

为了达到建模目的,解决所提出的问题,一般要建立几个模型(个别情况可建一个模型),因此该阶段需回答"建一些什么样的模型?"、"它们之间的关系是什么?"等问题。如统计报表自动化模型中需建总产值统计模型和劳动生产率核算模型等,二者的关系是只有总产值计算之后,才能计算劳动生产率等,这样就构成一模型体系。

该步与问题提出阶段是一个反复修正的过程,问题的提出是构思模型系统的基础,而构思的模型系统又可补充问题的提出,这样多次反馈,则使问题提出更全面、模型结构更合理。

第四步 根据所构思的模型体系,收集有关资料。为了实现所构思的模型,必须根据模型的要求收集有关资料,该步主要应回答"模型需要哪些资料?"等问题。如总产值核算模型需价格、商品产品数量和种类,劳动生产率模型需职工数等。

该步与构思的模型体系也有反馈关系,有时,构思的模型所需的资料很难收集,这就需要重新修改模型,进而可能影响到问题的提出等。这样经过几次反馈即可收集建模所需的资料。

第五步 设置变量和参数。变量和参数是构思模型时提出的,参数是在资料的收集,加工、整理后得出的,该步只是给以定义,一般要用一组符号表示,并整理成数据表和参数表的形式。该步需回答"需要哪些变量和参数?"。

第六步 模型具体化。模型具体化就是将变量和参数按变量之间的关系和模型之间

的关系连接起来，用规定的形式进行描述。它应回答“模型的形式是什么?”之类问题。如总产值模型的形式是：

$$\Pi = \sum_{i-1}^{n} P_i Q_i$$

式中 Π—— 总产值；

P_i—— i 类产品的价格；

Q_i—— i 类产品的产量；

n—— 产品的种类。

第七步 检验模型的正确性。模型正确与否将直接影响建模目的。该步应回答“模型正确吗?”一类问题。检验模型的正确性应先从各模型之间的关系开始，研究所构成的模型体系是否能实现建模目的；而后研究每个模型是否正确地反映所提出的问题。一般检验方法是试算。如试算不正确，则应重新审查所构思的模型系统，从中找出问题。因此它与构思模型又构成反馈。

第八步 将模型标准化。模型标准化是很重要的，一般情况下模型要对同类问题有指导意义，因此需具有通用性。该步需回答“该模形通用性如何?”等问题。如统计报表自动化系统模型应在一个行业、部门内通用才有实用价值。

第九步 根据标准化的模型编制计算机程序，使模型运行。该步需回答“计算时间短吗?”、“占用内存少吗?”等问题。

完成上述步骤，系统模型化才结束。

3.3.2 常用建模方法

系统模型化是对系统整体而言的。由上述程序可见，系统模型化最终落实到对具体提出的问题建立模型，因此本节介绍几种常见的建立模型的具体方法。

3.3.2.1 直接法

该方法是按系统的性质和范围，通过直接分析的方法，应用已知的科学知识建立模型，它是应用“白箱”理论建立模型的方法。

如某公司欲建立总产值计算模型，就需找出影响总产值计算的因素——价格和产量，根据总产值的经济意义：以不变价格和商品产品产量之积为总产值，即可确定两因素之间的关系，因产量是影响总产值的可变因素，故将其设为变量，从而建立如下模型：

$$\Pi = \sum_{i-1}^{n} P_{i0} Q_i$$

式中 P_{i0}——i 产品的不变价格；

Q_i——i 产品的产量；

n——产品种类数；

Π——总产值。

3.3.2.2 数理统计分析法

该方法用数理统计学理论，根据对所占有资料的分析与计算，最后得出模型。常用方法有：线性回归、非线性回归、旋转回归、数量化理论等。

如为预测我国胶合板生产量，需找出胶合板生产量变化的规律。经统计分析，胶合板

生产量随时间的推移呈增长趋势，若仅限于研究胶合板随时间变化的规律性，则可利用现有统计资料采用回归分析的方法建立模型。依据统计资料，经回归分析建立的我国胶合板生产量的预测模型为：

$$Y = 11.10 + 1.537t$$

式中　Y——胶合板产量预测值(万 m^3)；

t——年(1968 年 $t=0$)

3.3.2.3　模拟分析法

用一个容易实现试验和计算的系统代替与其有相似特性的、进行数量描述与试验有困难的系统的方法叫模拟分析法。

如力—电压相似系统就是用容易实验的电振荡系统模拟不容易实验的机械振荡系统的实例。图 3.3.2 可用一容易实现实验的电振荡系统图 3.3.3 来代替，电振荡系统可作为机械振荡系统的模拟模型。因机械振荡系统的数学模型为

$$P - m\frac{d^2x}{dt^2} + f\frac{dx}{dt} + kx$$

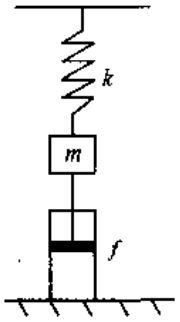


图 3.3.2　机械振荡系统

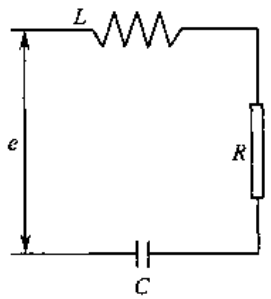


图 3.3.3　电振荡系统

该系统实验较困难，而电振荡系统模型为

$$e = L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q$$

且容易试验，通过对电振荡系统的试验，即可按表 3.3.1 所列的对应关系确定机械振荡系统的参数。

表 3.3.1　机械振荡系统与电振荡系统参数对照表

机械振荡系统		电振荡系统		机械振荡系统		电振荡系统	
力	P	电压	e	弹簧刚度	k	电容倒数	$\frac{1}{C}$
质量	m	电感	L	位移	x	电量	q
阻尼系数	f	电阻	R	速度	$\frac{dx}{dt}$	电流	$\frac{dq}{dt}$

3.3.2.4　模拟法

某公司准备建一临时材料库向 n 个工程项目供应材料，如何选择建设位址才能使总运输费用最省。

该问题可采用求不均匀物体重心的方法，按如下方法建立模拟模型：

按工程项目相对位置在一木盘上钻孔，将若干条线绳的一端连在一小铁环上，另一端分别穿过钻好的小孔，并按项目所需材料的比例挂上相应的砝码，这样小铁环就平衡在一个固定的位置(见图 3.3.4)。

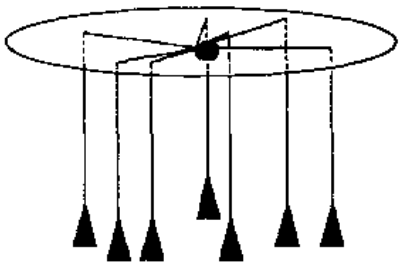


图 3.3.4　选址模拟模型

该模型就是应用求不均匀物体重心这一物理原理解决不能实验或不容易实验的系统问题的实例。

第4章　系统工程常用预测方法和模型

4.1　预测科学

嫦娥奔月是我国古代劳动人民对飞往月球、探索月球奥秘的一种幻想，体现出古人对未知事物的探求，但这不是预测。

三国时期，诸葛亮“未出茅庐已定三分天下”，是他根据当时的社会形势对社会发展规律的认识，是在分析天时、地利、人和的基础上，对社会发展的一种推测，这虽算预测但不能叫科学预测。

我们所说的预测是科学预测，是一门既有理论指导，又有科学方法的一种认识世界的工具，是一门广泛应用于社会、经济、科学、技术等方面的一门新科学。它是随科学技术的发展而产生，随科学技术的进步而发展的。

4.1.1　预测的概念

在调查研究的基础上对事物的未来进行科学的分析，研究其发展变化的规律性叫做预测分析；预测分析中所采用的方法和手段，称为预测技术。前者是预测理论，后者是预测方法，总称为预测的理论和方法。把预测理论和预测方法作为一个整体来研究的科学叫预测科学，简称为预测。

用预测的理论和方法研究技术经济问题的发展变化规律、预见未来，称技术经济预测；同理，用预测的理论和方法研究经济、社会、技术等的发展变化规律，则分别叫做经济预测、社会预测、技术预测等。

把系统作为预测对象，了解系统发展变化的规律性，预测系统未来则称为系统预测。系统预测是系统工程的重要内容，是系统规划设计、经营管理和系统决策的基础。

4.1.2　预测原理

预测是在调查研究的基础上对事物未来发展变化的规律进行研究的理论和方法的总称，预测的基本原理是：

(1) 整体性原理。事物是由若干部分相互关联而成的有机整体，事物发展变化的过程也是一个有机整体。因此，以整体性为特征的系统理论是预测的基本理论。

(2) 可知性原理。由于事物发展过程的统一性，即事物发展的过去、现在和将来是一个统一的整体，所以人类不但可以认识预测对象的过去和现在，而且也可以通过过去到现在的发展规律，推测将来的发展变化。

(3) 可能性原则。预测对象的发展有各种各样的可能性，预测是对预测对象发展的各种可能性的一种估计。如果认为预测是必然结果，则失去了预测的意义。

(4) 相似性原理。把预测对象与类似的已知事物的发展变化规律进行类比,可以对预测对象进行描述。

(5) 反馈原理。预测未来的目的是为了更好地指导当前,因此应用反馈原理不断地修正预测才会更好地指导当前工作,为决策提供依据。

4.1.3 预测的步骤

科学的预测是在广泛调查研究的基础上进行的,涉及方法的选择、资料的收集、数据的整理、建立预测模型、利用模型预测和对预测结果进行分析等一系列工作。

总的来说,预测的步骤为:

(1) 确定目标。该阶段的内容为确定预测对象、提出预测目的和目标,明确预测要求等。

(2) 选择预测方法。预测方法很多,到目前为止,各类预测方法不下几百种。因此应根据预测的目的和要求,考虑预测工作的组织情况,合理地选择效果较好的、既经济又方便的一种或几种预测方法。

(3) 收集和分析数据。该阶段根据预测目标和所选择预测方法的要求去收集所需原始数据。原始数据是进行预测的重要依据,所收集原始数据的质量和可靠性将直接影响预测的结果。对原始数据的要求是数据量足、质量高,只有这样,才能贴切地反映事物的规律。因此收集足够数量的可靠性高的数据是这阶段的任务。

数据的分析和整理是发现系统发展变化规律性和系统各组成部分内存联系的关键,是建立预测模型的根据,因此要选择合适的数据处理方法。

(4) 建立预测模型。建立预测模型是预测的关键工作,它取决于所选择的预测方法和所收集到的数据。建立模型的过程分为建立模型和模型的检验分析两个阶段。目前一部分人建立预测模型时,只建模型,不进行检验,这样的预测是不令人信服的。只有通过检验的模型,才能用于预测。

(5) 模型的分析。模型的分析是指对系统内部、外部的因素进行评定,找出使系统转变的内部因素和客观环境对系统的影响,以分析预测对象的整体规律性。

(6) 利用模型预测。所建立的模型是在一定假设条件下得到的,因此也只适用于一定条件和一定预测期限。如果将其推广到更大范围,就要利用分析、类比、推理等方法来确定模型的适用性。只有在确认模型符合预测要求时,才可利用模型进行预测。

(7) 预测结果的分析。利用预测模型所得的预测结果并不一定与实际情况符合。因为在建立模型时,往往有些因素考虑不周或因资料缺乏以及在处理系统问题时的片面性等使预测结果与实际情况偏离较大,故需从两个方面进行分析:① 用多种预测方法预测同一事物,将预测结果进行对比分析、综合研究之后加以修正和改进;② 应用反馈原理及时用实际数据修正模型,使预测模型更完善。

在组织预测的过程中,需注意以下几点:

① 加强预测工作的领导;

② 依靠集体的力量;

③ 协调预测人员的认识;

④ 注意积累资料;

⑤ 采用先进的计算技术和工具——计算机。

4.1.4 预测技术和预测模型

随着预测科学的发展,预测科学的应用领域越来越广,预测方法越来越多。据估计,目前预测方法达150多种,其中较常用有的十几种。

4.1.4.1 预测技术分类

用不同的分类原则对预测技术的分类如下:

(1) 按预测技术的属性,预测技术可分为:定性预测技术;定量预测技术;定时预测技术。

(2) 按预测对象,预测技术可分为:科学预测;社会预测;经济预测;市场预测等。

(3) 按预测方式,预测技术可分为:直观性预测;探索性预测;目标预测;反馈预测等。

(4) 按预测时间,预测技术可分为:短期预测;中期预测;长期预测;未来预测。

4.1.4.2 预测模型及分类

预测模型是预测的核心,建立预测模型是预测技术的核心。

(1) 预测模型按变量之间的关系可分为:因果关系模型、时间关系模型和结构关系模型等;

(2) 按变量形式又可分为线性预测模型和非线性模型;

(3) 按变量的数量,预测模型可分为一元模型和多元模型等;

(4) 按变量的性质,预测模型可分为定量因子模型、含定性因子模型和定性因子模型。

4.2 定性预测技术

定性预测技术是以预测者的经验为基础,判断发展趋势、探讨发展变化规律的方法。它适用于缺乏数据资料的情况下对事物的预测。定性预测的优点是方法简便、灵活。实践中,有时即使有充足的数量资料,也采用定性预测技术,其原因是把定性预测的结论与定量预测的结果相比较可以提高预测的准确性,同时在定性预测的指导下进行定量预测尚可起到定量预测起不到的作用。

常用的定性预测技术有德尔菲法、交叉概率法、专家意见法等。

4.2.1 德尔菲法

德尔菲法(Delphi)是美国兰德公司研究人员赫尔马(O. Helmet)和达尔奇(N. Dalkey)于20世纪40年代开发的一种预测方法。目前该方法已广泛用于军事预测、人口预测、医疗卫生保健预测、经营和需求预测、教育预测以及方案评价的决策分析等领域。

该方法的过程是预测机构或人员预先选定与预测问题有关的专家10人~15人,采用信件往来的方式与其建立联系,将他们的意见进行整理、综合、归纳后再匿名反馈给各位专家再次征求意见,按这种方式多次反复,直至使专家们的意见趋于一致为止,最后得出预测结论。

该方法的具体步骤为:

(1) 选择专家。专家人数的确定依据所预测问题的复杂性和所需知识面的宽窄，一般以 10 人～15 人为宜。所选择的专家彼此不发生联系，只用书信的方式与预测人员直接发生联系。

(2) 编制并邮寄“专家应答表”。首次交往需向专家介绍预测的目的，提供现有的相关资料，并邮寄“专家应答表”。为避免浪费专家的时间，“专家应答表”应力求简练，只需专家用“是”、“否”之类简单词句或符号回答或给予简单的评分。

(3) 分析整理“专家应答表”。收集专家的意见和反应，整理“专家应答表”，进行综合、分析、归纳等工作。

(4) 与专家反复交换意见。将整理、分析、归纳和综合的结果反馈给各专家并进一步提供有关资料，让专家修订自己的意见，填写“专家应答表”，如此反复进行直至得出预测结论。

(5) 将最终预测结论函告各专家并致谢。

采用德尔菲法整理专家所提供的资料，有时需将定性资料转化为定量数据，定量数据一般采用中位数、上、下四分位数来反映预测结果及其分散程度。

由定性资料转化为定量数据的方法是对预测中的每个因素给定一个分值 c_j，对应投票的专家数为 B_j，则均值和方差分别为

$$E=\frac{\sum c_jB_j}{\sum B_j};\qquad D=\frac{\sum (c_j-E_j)^2B_j}{\sum B_j}$$

如某家具厂对某种新家具的销售状况采用德尔菲法进行预测，第一轮预测资料整理结果见表 4.2.1。由计算可见，第一轮的预测结果为新产品基本畅销。

表 4.2.1　第一轮预测资料整理结果

销售状况	分值	人数	E	D
十分畅销	3	9		
基本畅销	2	28	2.05	0.4164
一般	1	5		
较差	0	1		

采用德尔菲法的好处是：

(1) 可以消除召开专家讨论会所出现的随声附和、崇拜专家、固持己见和有顾虑等弊病；

(2) 可使意见迅速集中。

这种方法是在假设预测项目的各因素之间无交互作用的前提下进行的，因此有一定的局限性，在使用该方法时，必须注意该点。

4.2.2　主观概率法

主观概率即是某人对某事件发生可能性的主观估计值。对同一事物来说，不同的人因知识、阅历、看问题的角度不同等原因造成对问题的估计值也不同，这就是主观因素在起作用。主观概率法就是在调查专家主观概率的基础上，寻求最佳主观估计的科学方法。

如果要预测某一事件发生的可能性，先调查一组专家的主观概率，然后加权平均即得某事件发生的概率，即

$$P=\frac{\sum_j P_j B_j}{\sum_j B_j}$$

式中 P ——事件发生概率的预测值；

P_j——第 j 种概率分级；

B_j——选第 j 种概率分级为主观概率的专家数。

4.2.3 交叉概率法

交叉概率法是对在交互影响因素作用下的事物进行预测的一种定性预测技术。

很多事物的发生或发展对其他事物将产生各种各样的影响，根据各事物之间的相互影响研究事物发生的概率，并用以修正专家的主观概率。从而对事物的发展作出较客观的评价则是该方法的基本思想。

该方法的步骤为：

(1) 确定各事物之间的影响关系；

(2) 确定各事物之间的影响程度；

(3) 计算某事物发生时对其他事物发生概率的影响；

(4) 分析其他事件对该事件的影响；

(5) 确定修正后的主观概率。

现举例说明该预测方法。

现以美国能源政策评价预测分析来说明交叉概率法的使用。经简化，影响美国能源政策的因素有：

E_1——用煤炭代替石油，其概率 $P_1=0.3$；

E_2——降低国内石油价格，其概率 $P_2=0.4$；

E_3——控制空气、水源的质量标准，其概率 $P_3=0.3$。

这些因素之间的关系见表 4.2.2：

表中向上的箭头表示正方向的交叉影响，它表明该事件的发生将促进另一事件发生之概率。而箭头向下，则表明负的影响。说明该事件发生将抑制或消除另一事件发生的概率。“—”号表示两事件无明显关系或相互间没有影响。

根据上表列出的矩阵，可求出其中各因素相互影响程度数值，用以修正发生概率，作出预测。

E_i 事件发生后，其余事件 E_j 发生的概率可按下式调整：

$$P'_j=P_j+KS(1-P_j) \tag{4.2.1}$$

式中 P_j ——E_i 事件发生前，t 时间 E_j 事件发生的概率；

P'_j——E_i 事件发生后，t 时间 E_j 发生的概率；

K ——说明 E_i 发生对 E_j 的影响方向，若 E_i 对 E_j 的影响为正，则取 $K=1$，若 E_i 对 E_j 影响为负则取 $K=-1$，若无影响取 $K=0$；

S——表明 E_i 发生对 E_j 的影响程度，$0<S<1$，随影响程度由小到大，S 取值由 0 到 1 逐渐加大。

事件 E_i 发生后，E_j 发生概率的调整如图 4.2.1 所示。

表 4.2.2　相互影响矩阵表

事件	事件发生概率	对其他诸事件的影响		
		E_1	E_2	E_3
E_1	0.3	—	↑	↑
E_2	0.4	↓	—	—
E_3	0.3	↓	↓	—

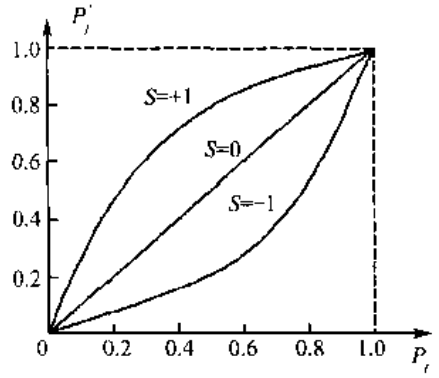


图 4.2.1　概率调整

事件 E_i 发生后对其余事件 E_j 的影响程度--般可用专家会议或专家调查法加以确定。

4.3　定量预测技术

我们所研究的系统是复杂的，只靠经验对其进行定性预测是不够的，还必须从数量上研究系统的变化，因此还需研究定量预测技术。

回归预测是经常使用的一种定量预测方法，是研究变量之间相互关系的数理统计分析方法。工作中经常使用的经验公式，大多是用某种回归分析方法得到的。本节主要介绍各种回归分析模型。

4.3.1　一元线性回归模型

回归预测模型的一般形式是：

$$Y=f(X)$$

如果 $f(X)$ 为一元线性函数形式时，模型变为

$$Y=a+bX$$

如果当参数 a、b 已知时，给定 X 的值即可确定 Y 的值。在笛卡儿坐标系中，该式可用一条斜率为 b、截距为 a 的直线表示，因此这种形式的回归分析又称一元线性回归。

进行预测是用已占有的数据，分析系统发展变化的规律性。根据所占有的若干组数据 $(X_i,Y_i)(i=1,2,\cdots,n)$，计算出系数 $\hat{a}$ 和 $\hat{b}$，就得出该事物发展变化的规律性 $\hat{Y}=\hat{a}+\hat{b}X$，这就是所要确定的预测模型。

一元线性回归模型在经济管理中的应用有两类，一类是反映因果关系的模型，另一类是反映时间序列的模型，虽然方法相同，但所适用的问题不同，预测时应用的原理也不同，前者是因果对应，后者是趋势外推。

下面以某公司 1988 年—2001 年产品产量的数据（见表 4.3.1），说明建立一元线性预测模型的方法。

表 4.3.1　某公司 1988 年—2001 年产品产量数据　　(单位:万 t)

年　份	1988	1989	1990	1991	1992	1993	1994
t	0	1	2	3	4	5	6
产量(Y)	10.59	17.67	17.07	17.21	18.24	18.84	17.62
年　份	1995	1996	1997	1998	1999	2000	2001
t	7	8	9	10	11	12	13
产量(Y)	19.21	18.44	20.85	25.22	29.24	32.99	35.11

4.3.1.1　一元线性回归模型的简易求法

预测方程为：
$$\hat{Y}=\hat{a}+\hat{b}X \tag{4.3.1}$$

求该方程参数的简易方法是：

(1) 将 n 组数据(X_i,Y_i)平均地分为两组(分组数决定于需确定的参数个数)，并分别代入式(4.3.1)；

(2) 把各组方程相加，得到二个以 $\hat{a}$、$\hat{b}$ 为变量的线性方程；

(3) 解此线性方程组可求出 $\hat{a}$、$\hat{b}$ 的值，代入式(4.3.1)，即是所求预测方程。

根据表 4.3.1 的数据并将其分为二组，代入式(4.3.1)后得：

1988 年 $X=0$　$10.59=\hat{a}$

1989 年 $X=1$　$14.67=\hat{a}+\hat{b}$

1990 年 $X=2$　$17.07=\hat{a}+2\hat{b}$

1991 年 $X=3$　$17.21=\hat{a}+3\hat{b}$

1992 年 $X=4$　$18.24=\hat{a}+4\hat{b}$

1993 年 $X=5$　$18.84=\hat{a}+5\hat{b}$

1994 年 $X=6$　$17.62=\hat{a}+6\hat{b}$

$$114.24=7\hat{a}+21\hat{b}$$

1995 年 $X=7$　$19.21=\hat{a}+7\hat{b}$

1996 年 $X=8$　$18.44=\hat{a}+8\hat{b}$

1997 年 $X=9$　$20.85=\hat{a}+9\hat{b}$

1998 年 $X=10$　$25.22=\hat{a}+10\hat{b}$

1999 年 $X=11$　$29.24=\hat{a}+11\hat{b}$

2000 年 $X=12$　$32.99=\hat{a}+12\hat{b}$

2001 年 $X=13$　$35.11=\hat{a}+13\hat{b}$

$$181.06=7\hat{a}+70\hat{b}$$

以 $\hat{a}$、$\hat{b}$ 为变量的方程组为：
$$\begin{cases}7\hat{a}+70\hat{b}=181.06\\7\hat{a}+21\hat{b}=114.24\end{cases}$$

解该方程组得：　$\hat{a}=12.23$；　$\hat{b}=1.3637$

预测方程为：　$\hat{Y}=12.23+1.3637X$

4.3.1.2　最小二乘法

对于回归方程 $\hat{Y}=\hat{a}+\hat{b}X$，将所占有的数据 $x_j(j=1,2,\cdots,n)$代入后得：

$$\hat{Y}_i=\hat{a}+\hat{b}X_i$$

令
$$Y_i-\hat{Y}_i=e_i$$

式中　e_i——所占有数据 Y_i 与预测值 $\hat{Y}_i$ 的误差。

为了防止误差正、负抵消，采用误差平方和最小作为确定参数 $\hat{a}$、$\hat{b}$ 的准则，这种确定参数 $\hat{a}$、$\hat{b}$ 的方法叫最小二乘法。

依据最小二乘法原理得：

$$\min \quad Q=\sum_{i=1} e_i^2=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2=\sum_{i=1}^{n}(Y_i-\hat{a}-\hat{b}X_i)^2 \tag{4.3.2}$$

使 Q 最小，驻点可由式(4.3.2)的微分为零来确定，由$\frac{dQ}{d\hat{a}}=0,\frac{dQ}{d\hat{b}}=0$得到方程组

$$\begin{cases}-2\sum_{i=1}^{n}(Y_i-\hat{a}-\hat{b}X_i)=0\\-2\sum_{i=1}^{n}(Y_i-\hat{a}-\hat{b}X_i)X_i=0\end{cases} \tag{4.3.3}$$

解该方程组得

$$\hat{a}=Y-\hat{b}\bar{X} \tag{4.3.4}$$

$$\hat{b}=\frac{\sum_{i=1}^{n}X_iY_i-\bar{X}\sum_{i=1}^{n}Y_i}{\sum_{i=1}^{n}X_i^2-\bar{X}\sum_{i=1}^{n}X_i} \tag{4.3.5}$$

式中

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i;\quad \bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$$

参数 $\hat{b}$ 还可写成如下形式：

$$\hat{b}=\frac{n\sum_{i=1}^{n}X_iY_i-\sum X_i\sum Y_i}{n\sum_{i=1}^{n}X_i^2-(\sum_{i=1}^{n}X_i)^2}\quad 或\quad \hat{b}=\frac{\sum_{i=1}^{n}X_iY_i-n\overline{XY}}{\sum_{i=1}^{n}X_i^2-n\bar{X}^2} \tag{4.3.6}$$

为了计算方便，可用列表的方法进行计算，上例的计算表见表 4.3.2。

表 4.3.2 计算表

序号	X_i	Y_i	X_i^2	Y_i^2	X_iY_i
1	0	10.59	0	112.1481	0.00
2	1	14.97	1	224.1009	14.97
3	2	17.07	4	291.3849	34.14
4	3	17.21	9	296.1841	51.63
5	4	18.24	16	332.6976	72.96
6	5	18.84	25	354.9456	94.20
7	6	17.62	36	310.4644	105.72
8	7	19.21	49	369.0241	134.47
9	8	18.44	64	340.0336	147.52
10	9	20.85	81	434.7225	187.65
11	10	25.22	100	636.0483	252.20
12	11	29.24	121	854.9776	321.64
13	12	32.99	144	1088.3400	395.88
14	13	35.11	169	1232.7120	456.43
合计	91	295.60	819	6877.7840	2269.41

用最小二乘法计算的参数 $\hat{a}$、$\hat{b}$ 为：$\hat{a}=11.10$，$\hat{b}=1.537$

故预测模型为：$Y=11.10+1.537X$

由式(4.3.5)知，如果 $\sum_i X_i=0$，

则
$$\hat{b}=\frac{\sum_i X_i Y_i}{\sum_i X_i^2} \tag{4.3.7}$$

$$\hat{a}=\overline{Y} \tag{4.3.8}$$

当数据满足 $\sum_i X_i=0$ 时，尤其是时间序列将自变量 t 经适当变换可使 $\sum_i t_i=0$，运用式(4.3.7)和式(4.3.8)，可使计算简化。

值得指出的是，用一元线性回归的方法建立预测模型只适用于散点图近似呈直线分布的情况，也就是当 X、Y 两个变量有大致的线性关系时才能应用，否则预测误差较大，是不适用的。

4.3.1.3 线性回归预测模型检验过程及预测精度

由上述的讨论可知，一元线性回归方程在某种程度上揭示了两个变量间的线性相关关系。但在应用线性回归的计算公式时会发现，并不需要预先假设两个变量之间一定具有线性相关关系，也就是说，对任意给定的 N 组数据都可根据公式确定一条直线而得出预测方程。这样一来，需解决这条直线能否反映出所研究系统的变化规律问题，精确地说，需研究这条直线是否有实际使用价值。

我们曾指出，只有当两个变量之间有大致的线性关系时，用该方法所得到的预测模型才是适用的。能否用一个数量指标来评价两变量大致呈线性关系的程度，以决定用回归分析所得到的数学模型与研究系统的规律相符呢？又怎样确定预测模型的预测精度呢？这就是线性回归检验和精度分析需解决的问题。

1．相关系数

我们把评价 X，Y 两个变量之间线性关系密切程度的数量指标叫相关系数，并以 r 表示，其计算公式为

$$r=\frac{L_{XY}}{\sqrt{L_{XX}L_{YY}}} \tag{4.3.9}$$

式中
$$L_{XX}=\sum_i (X_i-\overline{X})^2=\sum_i X_i^2-n\overline{X}^2 \tag{4.3.10}$$

$$L_{YY}=\sum_i (Y_i-\overline{Y})^2=\sum_i Y_i^2-n\overline{Y}^2 \tag{4.3.11}$$

$$L_{XY}=\sum_i (X_i-\overline{X})(Y_i-\overline{Y})=\sum_i X_iY_i-n\overline{XY} \tag{4.3.12}$$

L_{XX} 叫 X 的离差平方和，它是反映自变量 X 波动的一个指标，L_{XX} 越大，X 的波动越大，反之越小；L_{YY} 叫 Y 的离差平方和，它是反映变量 Y 波动的一个指标，L_{YY} 越大，Y 的波动越大，反之越小；L_{XY} 叫做 X、Y 的离差乘积和。

由于
$$\sqrt{L_{XX}}\geq\sum_i (X_i-\overline{X})$$

$$\sqrt{L_{YY}}\geq\sum_i (Y_i-\overline{Y})$$

$$\sqrt{L_{XX}L_{YY}} \geq \sum_i (X_i - \bar{X})(Y_i - \bar{Y}) = L_{XY}$$

故 r 的取值范围为 $-1 \leqslant r \leqslant 1$。

为了说明 r 如何反映两个变量之间线性关系的密切程度，现研究 $\hat{b}$ 与 r 的关系。由式(4.3.6)知

$$\hat{b} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} = \frac{\sum_i X_i Y_i - n\overline{XY}}{\sum_i X_i^2 - n\bar{X}^2} = \frac{L_{XY}}{L_{XX}} \tag{4.3.13}$$

将式(4.3.13)的分子、分母同乘$\sqrt{L_{YY}}$，并把 $L_{XX} = \sqrt{L_{XX} \cdot L_{XX}}$代入得：

$$\hat{b} = \frac{L_{XY}\sqrt{L_{YY}}}{\sqrt{L_{XX}L_{YY}} \cdot \sqrt{L_{XX}}} = r\sqrt{\frac{L_{YY}}{L_{XX}}} \tag{4.3.14}$$

或

$$r^2 = \hat{b}^2 \frac{L_{XX}}{L_{YY}} \tag{4.3.15}$$

由式(4.3.14)和式(4.3.15)可见：

(1) 当 $r=0$ 时，$b=0$，则回归直线是一条与 X 轴平等的直线，说明 X 的变化与 Y 无关，其几何表示如图 4.3.1 所示。

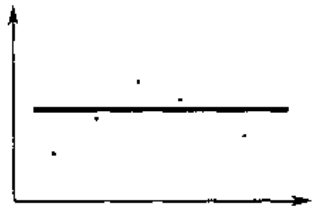


图 4.3.1　X 的变化与 Y 无关

(2) 当 $r^2=1$ 即 $|r|=1$ 时，$\hat{b}^2 = \frac{L_{YY}}{L_{XX}}$或 $b^2 L_{XX} = L_{YY}$

这时我们考查回归的误差平方和 Q：

$$Q = \sum_i (Y_i - \hat{Y}_i)^2 = \sum_i (Y_i - \hat{a} - \hat{b}X_i)^2$$

把$\hat{a} = \bar{Y} - \hat{b}\bar{X}$ 代入：

$$Q = \sum_i (Y_i - Y + \hat{b}X - \hat{b}X_i)^2 = \sum_i [(Y_i - \bar{Y}) - \hat{b}(X - \bar{X})]^2 =$$

$$\sum_i [(Y_i - \bar{Y})^2 + \hat{b}^2(X_i - \bar{X})^2 - 2\hat{b}(Y_i - \bar{Y})(X_i - \bar{X})] =$$

$$L_{YY} + \hat{b}^2 L_{XX} - 2\hat{b}L_{XY}$$

由式(4.3.13)知，　$L_{XY} = \hat{b}L_{XX}$

$\therefore Q = L_{YY} + \hat{b}^2 L_{XX} - 2\hat{b}^2 L_{XX} = L_{YY} - \hat{b}^2 L_{XX}$

因为当$|r|=1$ 时，$\hat{b}^2 L_{XX} = L_{YY}$，故 $Q=0$。

由此可知，当 $r=1$ 时，$Q=0$，即所有点(x_i, y_i)均在回归直线上，称完全正相关；当 $r=-1$时，称完全负相关。其几何意义如图 4.3.2、图 4.3.3 所示。

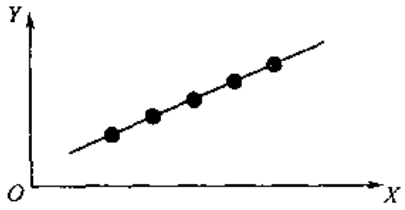


图 4.3.2 完全正相关

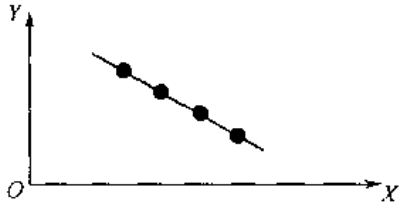


图 4.3.3 完全负相关

当 $-1<r<1$ 时，X 与 Y 之间存在着一定的线性相关关系。当 $r>0$ 时，$\hat{b}>0$，Y 随 X 的增大而呈增加趋势，此时称正相关；当 $r<0$ 时，$\hat{b}<0$，Y 随 X 的增大呈减小趋势，此时称负相关；r 的绝对值越大时，散点越集中在回归直线附近，反之散点离回归直线越远、越分散。当 $-1<r<1$ 时的几何意义如图 4.3.4、图 4.3.5 所示。

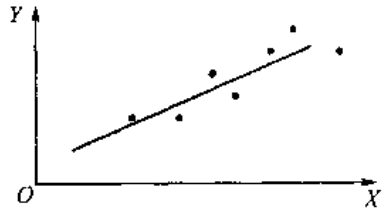


图 4.3.4 正相关

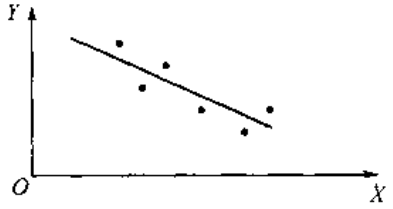


图 4.3.5 负相关

由上述分析可见，指标 r 可衡量两变量的线性相关程度。但 r 只为我们提供了相对比较的评价依据，若进行绝对评价，则显得依据不足。因此为进行绝对评价，必须进行显著性检验。

2. 显著性检验

进行显著性检验，实际上相当于规定一个合理的、认为能满足使用要求的指标界限，并用该指标界限对系统预测模型的适用性进行绝对评价。

由式(4.3.9)可知，r 值的大小取决于 X_i、Y_i 和数据数量 n。因此，为任何系统都规定一个统一的标准值是不能反映不同情况的差异的，也是不合理的。显著性检验就是依据所占有的数据量及其分布情况、变量个数等条件，确定一个合理的标准作为评价指标。

常用的显著性检验有三种：t 检验、F 检验、r 检验。

(1) t 检验。t 检验的意义是检验回归方程中参数 b 的估计值 $\hat{b}$，在某一显著性水平下(通常选为 0.05)是否为零。该检验是在假设 $\hat{b}=0$ 的情况下进行的。如果 $\hat{b}$ 为零，则说明 Y 与 X 的变化无关。因此该方法根据占有数据的多少(样本数 n)查 $t_{1-\frac{\alpha}{2}}(n-2)$ 的分布表，确定 t 的临界值 t_α，与根据实际问题计算的 t 分布值 t 进行比较，如果当 $t>t_\alpha$，则说明原假设不成立，相关显著，回归方程有实用价值。否则原假设成立，可认为 $\hat{b}$ 在所确定的显著性水平下为零，即 $b=0$，这时，回归方程无实用价值。

t 的计算公式为：

$$t=\frac{\hat{b}}{S}\sqrt{L_{XX}} \tag{4.3.16}$$

式中，S 为 Y 的均方差，$S=\sqrt{\dfrac{\sum(Y_i-\bar{Y})^2}{n-2}}=\sqrt{\dfrac{L_{XX}L_{YY}-L_{XY}^2}{(n-2)L_{XX}}}$

(2) F 检验。F 检验的意义与 t 检验相同，只不过是查 $F_{1-a}(1,n-2)$表确定 F 的临界值 F_a。

F 的计算公式为
$$F=(n-2)\frac{r^2}{1-r^2} \tag{4.3.17}$$

当 $F>F_a$ 时，否定原假设，变量相关显著。

(3) r 检验。为了使用方便，由公式：$(n-2)\frac{r^2}{1-r^2}>F_{1-a}(1,n-2)$

反求出 r 的临界值 r_a，即可通过 r 的大小直接判断显著性。

当
$$|r|\geq\sqrt{\frac{1}{\frac{n-2}{F_{1-a}(1,n-2)}+1}}=r_a$$
时，两变量相关显著。

将 $\sqrt{\frac{1}{\frac{n-2}{F_{1-a}(1,n-2)}+1}}$ 编成表，即是检验相关系数的临界值 r_a 表。

上例中，相关系数 $r=0.9162$

应用 t 检验法：$t_{0.975}=2.179$，$t=7.863$，$t>t_{0.975}$，故回归效果显著。

应用 F 检验法：$F_{0.95}(1,14-2)=4.75$，$F=62.73$，$F>F_{0.95}(1,12)$，故回归效果显著。

应用 r 检验：$r=0.9162$，$r_a=0.5760$，$r>r_a$，故回归效果显著。

3. 方差分析

为了估计预测精度需对预测模型作方差分析。

应用预测模型 $\hat{Y}=\hat{a}+\hat{b}X$，当 $X=X_0$ 时，求出的预测值 $\hat{Y}_0$ 只是实际 Y_0 的期望值，且该估计是无偏估计。由数理统计知

其方差为
$$D(\hat{Y}-Y)=\sigma^2\left[1+\frac{1}{n}+\frac{(X-\bar{X})^2}{\sum_i(X_i-\bar{X})^2}\right]$$

因为 $\hat{\sigma}$ 是 σ 的无偏估计，所以可用 $\hat{\sigma}$ 代替 σ，由于 $\hat{\sigma}^2=\frac{1}{n-2}\sum_i(\hat{Y}_i-\hat{a}-\hat{b}X_i)^2$ 且 Y 落在 $(\hat{Y}-\delta,\hat{Y}+\delta)$ 内的概率为 $1-\alpha$，即
$$P(\hat{Y}-\delta,\hat{Y}+\delta)=1-\alpha \tag{4.3.18}$$
$$\therefore\delta^2=F_{1-a}(1,n-2)\cdot\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(X-\bar{X})^2}{\sum_i(X_i-\bar{X})^2}\right] \tag{4.3.19}$$

或
$$\delta^2=t_{1-\frac{a}{2}}(n-2)\cdot\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(X-\bar{X})^2}{\sum_i(X_i-\bar{X})^2}\right] \tag{4.3.20}$$

由 δ 的计算公式可知，δ 的大小取决于数据组数(样本数) n 和 X 的大小。当 n 大时，δ 值小，预测精度高，反之则低；在数据组数一定且在 $X=\bar{X}$ 时，δ 最小；若 X 越远离

$\bar{X}$，δ 越大，则预测误差越大。其几何意义见图 4.3.6。由此可得出提高线性回归预测精度的方法为：① 增加数据组数；② 使预测期限尽量接近 $\bar{X}$。

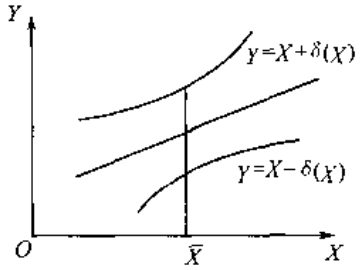


图 4.3.6　预测精度示意图

在实际工作中增加占有的数据量，需增加预测费用和时间，因此要以系统思想确定合理的预测精度和期限，达到以最省的预测费用取得最好的预测效果的目的。

现总结一下线性回归的方法和步骤：

(1) 整理占有的数据 $(X_i, Y_i)(i=1,2,\cdots,n)$。

(2) 运用 $\hat{Y}=\hat{a}+\hat{b}\bar{X}$ 和 $\hat{b}=\dfrac{L_{XY}}{L_{XX}}$ 求出 $\hat{a}$ 和 $\hat{b}$ 得到预测方程：$\hat{Y}=\hat{a}+\hat{b}X$。

(3) 进行检验：①求出相关系数 r；②选择 t 检验、F 检验或 r 检验法对预测模型的显著性进行检验。

(4) 利用模型进行预测，并用

$$\delta^2=F_{1-\alpha}(1,n-2)\cdot\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(X-\bar{X})^2}{\sum_i(X_i-\bar{X})^2}\right]$$

或

$$\delta^2=t_{1-\frac{\alpha}{2}}(n-2)\cdot\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(X-\bar{X})^2}{\sum_i(X_i-\bar{X})^2}\right]$$

确定置信区间。

目前实际工作中很多人只进行到第三步的①为止，这是不够的。因为在未完成以后各步计算时，相关系数只是一个相对的评价指标。只有完成上述全部过程，才基本上可以对预测对象的变化规律有一明确认识，才能真正做好预测工作。这也正是我们在回归检验中费了这么多笔墨的原因。

4.3.2　一元非线性回归模型

在实践中，经常遇到两个变量之间的关系呈非线性关系的情况。一般情况下，非线性函数都可通过变量代换的方法或应用泰勒级数展开而变成一元线性和多元线性函数。对于用变量代换的方法将其化为一元线性关系的问题，可采用一元线性回归的方法，对于用泰勒级数将其化为多元线性关系的问题，可用多元回归分析法去解决。

4.3.2.1　化一元非线性函数为线性函数

在一元函数中可化曲线方程为直线方程的问题有很多。在进行该项工作之前，较困难的是确定要配合的曲线类型。

确定曲线类型的方法一般为：

(1) 根据理论分析以及过去所积累的经验,确定 X、Y 之间的函数类型;

(2) 在数据量不大的情况下,作出散点图,观察散点的分布,确定函数类型;

(3) 采用多种曲线模型进行回归分析后,进行相对比较分析,从中选择一个较好形式的模型作为预测模型。

下面介绍将特殊的曲线方程化成直线方程的变量代换方法。

1. 双曲线函数

双曲线函数的一般形式为 $$\frac{1}{Y}=a+\frac{b}{X} \tag{4.3.21}$$

令 $\frac{1}{Y}=Y^*$,$\frac{1}{X}=X^*$,代入式(4.3.21)得

$$Y^*=a+bX^* \tag{4.3.22}$$

式(4.3.22)为一元线性函数形式,所以预测模型为

$$\hat{Y}^*=\hat{a}+\hat{b}X^*$$

2. 指数函数

指数函数的一般形式为 $$Y=Ae^{bX} \tag{4.3.23}$$

对式(4.3.23)取对数后得

$$\ln Y=\ln A+bX \tag{4.3.24}$$

令 $\ln Y=Y^*$,$\ln A=a$,代入式(4.3.24)则得出化为线性函数的预测模型

$$\hat{Y}^*=\hat{a}+\hat{b}X$$

3. 指数函数

指数函数的一般形式为 $$Y=Ae^{-\frac{b}{X}} \tag{4.3.25}$$

对式(4.3.25)取对数得

$$\ln Y=\ln A+\frac{b}{X}$$

令 $\ln Y=Y^*$,$\ln A=a$,$\frac{1}{X}=X^*$,代入上式,得出化成一元线性函数模型的形式为

$$Y^*=a+bX^* \tag{4.3.26}$$

4. 对数函数

对数函数的一般形式为 $$Y=a+b\ln X \tag{4.3.27}$$

令 $\ln X=X^*$,代入式(4.3.27),得线性模型为

$$Y^*=a+bX^* \tag{4.3.28}$$

5. 幂函数

幂函数的一般形式为 $$Y=AX^b \tag{4.3.29}$$

对式(4.3.29)取对数得

$$\ln Y=\ln A+b\ln X$$

令 $\ln Y=Y^*$,$\ln A=a$,$\ln X=X^*$,代入上式得线性模型为

$$Y^*=a+bX^* \tag{4.3.30}$$

6. S 曲线

S 曲线的一般形式为 $$Y=\frac{1}{a+be^{-X}} \tag{4.3.31}$$

对式(4.3.3.1)取倒数得

$$\frac{1}{Y}=a+b\mathrm{e}^{-X}$$

令 $$\frac{1}{Y}=Y^*,\mathrm{e}^{-X}=X^*$$

代入上式得线性模型为

$$Y^*=a+bX^* \tag{4.3.32}$$

现以某地区总产值与投资的资料(见表 4.3.3)来研究二者之间的相关关系。

表 4.3.3 总产值和投资额 (单位:亿元)

年 份	1960	1965	1970	1975	1980	1985	1990	1995
总产值(X_i)	2.9	9.2	7.2	12	16	37	49	52
总投资(Y_i)	0.49	2.01	2.03	6.16	4.29	5.89	7.8	7.57

采用指数函数模型 $Y=A\mathrm{e}^{bX}$。

将总投资 Y_i 取对数,按线性回归法计算得:

$Y^*=0.165+0.042X$, $r=0.8116$, $F=11.57$, $F_a=5.99$, $F>F_a$ 故回归显著。

$\because \hat{a}=0.165$, $\therefore \ln A=0.165$, $A=1.18$

$$\hat{b}=b=0.042$$

故预测模型为:$Y=1.18\mathrm{e}^{0.042X}$。

需要指出的是,把一元非线性问题化为线性问题确定预测模型的方法,剩余平方和、剩余标准差和相关系数与化成线性关系以后的形式不同,其原因是所求出的方程使变换以后的 Y^* 与相应回归线上的点之差最小,而不是变换前的 Y 与回归线上相应点之差最小。本例中是以 $\min Q=\sum_i(\ln Y_i-\ln \hat{Y}_i^*)^2$ 为准则进行的回归,所以 $Q=\sum_i(Y_i-\hat{Y}_i)^2$ 应以 $\hat{Y}=1.18\mathrm{e}^{0.042X}$ 为基准进行计算。

如本例中应按以下公式计算预测模型 $Y=1.18\mathrm{e}^{0.042X}$ 的剩余平方和:$Q=\sum\delta^2=\sum(Y_i-\hat{Y}_i)^2$,剩余标准差:$S=\sqrt{\frac{Q}{N-2}}$;相关指数:$R^2=1-\frac{Q}{L_{YY}}$。

由于曲线类型很多且有些很相似,故在实际工作中最好的办法是用计算机按不同函数配合曲线,从中选择一个比较好的,即比较 Q、R^2 和 S(Q、S 小者为好,R^2 大者为好)。

4.3.2.2 化一般一元非线性函数为线性函数

任何一元非线性函数都可通过数学分析(泰勒级数展开,取近似值)的方法化成以下形式

$$Y=a+b_1X+b_2X^2+\cdots+b_nX^n \tag{4.3.33}$$

通过变量代换的方法,可将式(4.3.33)化成一个多元线性模型。

令 $X_1=X$, $X_2=X^2$, $\cdots$, $X_n=X^n$,代入式(4.3.33)得

$$Y=a+b_1X_1+b_2X_2+\cdots+b_nX_n \tag{4.3.34}$$

这就是多元线性模型的一般形式,它可采用多元线性回归的方法建立模型。

4.3.3 多元线性回归模型

以前讨论的回归预测方法，只涉及两个变量。但在实际问题中，影响因变量的因素往往不止一个，此时必须探索在各种因素综合作用下系统变化的规律性，以预测系统的变化，为系统决策提供依据。解决这类问题的方法之一就是多元线性回归分析。

多元线性回归分析的原理与一元线性回归分析基本相同，不同点只是计算复杂、分析方法的理论较深。现以二元回归分析为例，得出多元线性回归分析的一般方法。

4.3.3.1 二元线性回归分析

某公司木器厂准备根据影响生产成本的两个重要因素——劳动量和木材消耗量，建立成本预测模型，资料见表 4.3.4。

表 4.3.4 某公司木器厂有关资料

月	总成本 Y/千元	劳动量 X_1/kh	木材耗用量 X_2/km³
1	3.1	3.9	2.4
2	2.6	3.6	2.1
3	2.9	3.8	2.3
4	2.7	3.9	1.9
5	2.8	3.7	1.9
6	3.0	3.9	2.1
7	3.2	3.8	2.4

现考虑的自变量有两个：一个是劳动量，以 X_1 表示；另一个是木材消耗量，以 X_2 表示。若家具生产总成本以 Y 表示，现研究当 X_1、X_2 改变时，Y 的变化规律。我们采用线性回归模型

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 \tag{4.3.35}$$

作为预测模型。

因该方程在三维空间中可用一个平面表示，故称它为 Y 对 X_1、X_2 的回归平面，$\hat{b}_0$、$\hat{b}_1$、$\hat{b}_2$ 称为回归系数。

仍用最小二乘法，以误差平方和最小作为回归分析的准则，即求

$$\min \quad Q = \sum_{i=1}^{n} (Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i})^2 \tag{4.3.36}$$

由微分学知，当 $\dfrac{\partial Q}{\partial \hat{b}_0} = \dfrac{\partial Q}{\partial \hat{b}_1} = \dfrac{\partial Q}{\partial \hat{b}_2} = 0$ 时，Q 有极小值。

故得方程组：

$$\begin{cases} \dfrac{\partial Q}{\partial \hat{b}_0} = -2 \sum\limits_{i} (Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i}) = 0 \\ \dfrac{\partial Q}{\partial \hat{b}_1} = -2 \sum\limits_{i} (Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i}) X_{1i} = 0 \\ \dfrac{\partial Q}{\partial \hat{b}_2} = -2 \sum\limits_{i} (Y_i - \hat{b}_0 - \hat{b}_1 X_{1i} - \hat{b}_2 X_{2i}) X_{2i} = 0 \end{cases} \tag{4.3.37}$$

化简后得：

$$\begin{cases} n\hat{b}_0 + \sum_i X_{1i}\hat{b}_1 + \sum_i X_{2i}\hat{b}_2 = \sum_i Y_i \\ \sum_i X_{1i}\hat{b}_0 + \sum_i X_{1i}^2\hat{b}_1 + \sum_i X_{1i}X_{2i}\hat{b}_2 = \sum_i X_{1i}Y_i \\ \sum_i X_{2i}\hat{b}_0 + \sum_i X_{1i}X_{2i}\hat{b}_1 + \sum_i X_{2i}^2\hat{b}_2 = \sum_i X_{2i}Y_i \end{cases} \tag{4.3.38}$$

式(4.3.38)称为正规方程组，解式(4.3.38)的第一个方程可得

$$\hat{b}_0 = \bar{Y} - \hat{b}_1\bar{X}_1 - \hat{b}_2\bar{X}_2 \tag{4.3.39}$$

其中，$\bar{Y} = \frac{1}{n}\sum_i Y_i$，$\bar{X}_1 = \frac{1}{n}\sum_i X_{1i}$，$\bar{X}_2 = \frac{1}{n}\sum_i X_{2i}$

把 $\hat{b}_0$ 代入式(4.3.38)的后两个方程，并整理得

$$\begin{cases} l_{11}\hat{b}_1 + l_{12}\hat{b}_2 = l_{10} \\ l_{21}\hat{b}_1 + l_{22}\hat{b}_2 = l_{20} \end{cases} \tag{4.3.40}$$

用矩阵形式描述为

$$\begin{pmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{pmatrix} \cdot \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} l_{10} \\ l_{20} \end{pmatrix} \quad \text{或} \quad L\hat{B} = L_0$$

其中，
$$l_{kj} = \sum_i^n (X_{ki} - \bar{X}_k)(X_{ji} - \bar{X}_j) \qquad (k = 1,2;\ j = 1,2) \tag{4.3.41}$$

$$l_{k0} = \sum_i^n (X_{ki} - \bar{X}_k)(Y_i - \bar{Y}) \qquad (k = 1,2) \tag{4.3.42}$$

该方程也叫正规方程，它形式简单，求解也非常容易。因矩阵 $\begin{pmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{pmatrix}$ 是一对称非奇异矩阵，故 $\hat{\boldsymbol{B}} = \boldsymbol{L}^{-1}\boldsymbol{L}_0$ 即为正规方程(4.3.40)的解。

由式(4.3.40)解出 $\hat{b}_1$、$\hat{b}_2$ 后，代入式(4.3.39)即求出 $\hat{b}_0$。

求出 $\hat{b}_0$、$\hat{b}_1$、$\hat{b}_2$ 后得出预测方程

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2$$

通过上述推导可知，建立二元线性预测模型的关键是形成正规方程。在形成正规方程后就把二元线性回归分析归结为求解线性方程组的解的问题了。

现以上例说明建立预测模型的方法和步骤：

1. 求正规方程组的系数(即建立 $\boldsymbol{L}$ 和 $\boldsymbol{L}_0$ 矩阵)

$$l_{11} = \sum_i (X_{1i} - \bar{X}_1)(X_{1i} - \bar{X}_1) = \sum_i (X_{1i} - \bar{X}_1)^2 = 0.08$$

$$l_{12} = \sum_i (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = 0.03$$

$$l_{21} = \sum_i (X_{2i} - \bar{X}_2)(X_{1i} - \bar{X}_1) = 0.03$$

$$l_{22} = \sum_i (X_{2i} - \bar{X}_2)(X_{2i} - \bar{X}_2) = \sum_i (X_{2i} - \bar{X}_2)^2 = 0.277$$

$$l_{10} = \sum_i (X_{1i} - \bar{X}_1)(Y_i - \bar{Y}) = 0.08$$

$$l_{20}=\sum_{i}(X_{2i}-\bar{X}_2)(Y_i-\bar{Y})=0.21$$

∴正规方程为

$$\begin{cases}0.08\hat{b}_1+0.03\hat{b}_2=0.08\\0.03\hat{b}_1+0.277\hat{b}_2=0.21\end{cases}$$

或

$$\begin{bmatrix}0.08 & 0.03\\0.03 & 0.277\end{bmatrix}\cdot\begin{bmatrix}\hat{b}_1\\\hat{b}_2\end{bmatrix}=\begin{bmatrix}0.08\\0.21\end{bmatrix}$$

2. 求解正规方程和 $\hat{b}_0$

方程组求解结果为:$\hat{b}_1=0.7461$,$\hat{b}_2=0.6769$,代入式(4.3.39)后,得出 $\hat{b}_0=1.3956$。

3. 建立模型

根据所求系数,预测模型为

$$\hat{Y}=-1.3956+0.7461X_1+0.6769X_2$$

当给定 X_1、X_2,即给定劳动量和木材消耗量时,可预测木器生产的总成本。如 $X_1=4$、$X_2=3$ 时,总成本 Y 的预测值为

$$\hat{Y}=-1.3956+0.7461\times4+0.6769\times3=3.7(千元)。$$

4.3.3.2 多元线性回归分析

多元线性回归分析是研究系统在多个因素的影响下,建立系统总体变化规律预测模型的方法。现在二元回归分析的基础上将其推广,即可得到多元线性回归分析的一般方法。

多元线性回归模型的一般形式是

$$\hat{Y}=\hat{b}_0+\hat{b}_1X_1+\hat{b}_2X_2+\cdots+\hat{b}_mX_m \tag{4.3.43}$$

该方程是 $m+1$ 维空间的一个超平面,所以它为 Y 对 X 的回归超平面方程。

由式(4.3.39)可知,如果当自变量数由两个增加到 m 个时,b_0 的求法可推广到

$$b_0=\bar{Y}-\hat{b}_1\bar{X}_1-\hat{b}_2\bar{X}_2-\cdots-\hat{b}_m\bar{X}_m=\bar{Y}-\sum_{k=1}^{m}\hat{b}_k\bar{X}_k \tag{4.3.44}$$

正规方程组的系数,即 L 矩阵的元素 L_{kj} 为

$$l_{kj}=\sum_{i=1}^{m}(X_{ki}-\bar{X}_k)(X_{ji}-\bar{X}_j)$$

正规方程的常数项,即 L_0 向量为

$$l_{k0}=\sum_{i=1}^{n}(X_{ki}-X_k)(Y_i-\bar{Y})$$

求解正规方程组,即可确定 $\hat{b}_1,\hat{b}_2,\cdots,\hat{b}_m$,代入式(4.3.44),求出 $\hat{b}_0$,从而得到预测模型:

$$\hat{Y}=\hat{b}_0+\hat{b}_1X_1+\hat{b}_2X_2+\cdots+\hat{b}_mX_m$$

4.3.3.3 多元线性回归模型的检验

多元线性回归模型也和一元线性回归分析一样,需要检验模型的适用性。检验的基本思想是,首先对整个多元回归模型进行显著性检验。因多元线性回归分析中自变量的数量多,因此,在总的显著性分析的基础上,还需分析哪些因素对因变量的影响大,哪些因

素影响小,哪些可以忽略不计,最后再确定它的区间估计值。

1. 全相关系数

多元分析与一元分析相比,一元分析是多元分析在变量数为 $m=1$ 时的特殊情况。因此,其分析的公式也有相似之处。

多元分析的平方和分解公式仍为

$$L_{YY}=U+Q(\text{回归平方和与剩余平方和})$$

其中,
$$U=\sum_i(\hat{Y}_i-\overline{Y})^2;Q=\sum_i(Y_i-\hat{Y}_i)^2 \tag{4.3.45}$$

经推导得:

$$U=\sum_i^m \hat{b}_k l_{k0} \tag{4.3.46}$$

$$Q=L_{YY}-U=L_{YY}-\sum_k^m l_{k0}\hat{b}_k \tag{4.3.47}$$

剩余标准差:

$$S_Y=\sqrt{\frac{Q}{n-m-1}}=\sqrt{\frac{1}{n-m-1}(L_{YY}-\sum_k^m l_{k0}\hat{b}_k)} \tag{4.3.48}$$

相关系数:

$$R^2=\frac{U}{L_{YY}},\ \text{即}\ R=\sqrt{1-\frac{Q}{L_{YY}}} \tag{4.3.49}$$

这里 R 是衡量整个回归效果的数量指标,称全相关系数,其含意与一元线性回归分析中相关系数 r 完全一致。

2. 回归显著性检验(用 F 检验)

若回归无意义,即 $\hat{b}_1=\hat{b}_2=\cdots=\hat{b}_m=0$

则
$$F=\frac{n-m-1}{m}\cdot\frac{R^2}{1-R^2}<F_{1-a}(m,n-m-1) \tag{4.3.50}$$

否则 $F>F_{1-a}(m,n-m-1)$。

由此式可见当 $m=1$ 时,就是一元线性回归时的公式(4.3.17)。

在整个回归显著的基础上,多元分析还必须对多元回归系数进行分析,也就是判断每个自变量与因变量的显著性。因为在多元回归分析中经常出现 Y 与自变量总体有相关关系,但 Y 与某个个别自变量 X_i 可能无关,也就是说,X_i 对 Y 并不起作用或它的作用已被其他 X_j 所代替。进行这步分析可确定出对 Y 影响较大的因素和影响较小或没有影响的因素,以决定应该把哪些因素包含在预测模型中,把哪些因素从所确定的模型中剔除,以便建立一个更简单、更适用、更方便的预测模型。

其检验采用 t 检验法:

$$t_k=\frac{\hat{b}_k}{\sqrt{C_{kk}}\cdot S_Y} \tag{4.3.51}$$

其中,C_{kk} 为正规方程系数矩阵 L 的逆矩阵 L^{-1} 中对角线上的元素。如果

$$|t_k|>t_1-\frac{a}{2}(n-m-1)$$

则认为 X_k 对 Y 有作用，否则无作用，应予以剔除。

从多元线性回归模型中剔除某个影响不大的因素 X_k（有时需要研究剔除某个影响比较大的因素），是在 $\hat{b}_k=0$ 的假设下，由 t 检验确定的，但是决不能从预测模型中以 $\hat{b}_k$ 较小为理由简单地删除 $\hat{b}_kX_k$ 项。因为当剔除某一因素后，相当于从原始数据中去掉了这一因素所对应的所有数据，这必导致回归模型的变化，故需重新配合方程。为了避免复杂的计算，数理统计学为我们提供了一个不需重新配合方程的计算公式和算法。其算法如下：

设剔除因素之前的回归模型为

$$\hat{Y}=\hat{b}_0+\hat{b}_1X_1+\hat{b}_2X_2+\cdots+\hat{b}_mX_m$$

剔除因素后，新回归模型为

$$\hat{Y}=\hat{b}_0^*+\hat{b}_1^*X_1+\hat{b}_2^*X_2+\cdots+\hat{b}_{i-1}^*X_{i-1}+\hat{b}_{i+1}^*X_{i+1}+\cdots+\hat{b}_m^*X_m$$

则

$$\hat{b}_j^*=\hat{b}_j-\frac{C_{ij}}{C_{ii}}\hat{b}_i \qquad j=1,2,\cdots,m;\quad j\neq i \tag{4.5.52}$$

$$\hat{b}_0^*=\bar{Y}-\sum_{j\neq i}\hat{b}_j^*\bar{X}_j$$

式中 C_{ij}——$\boldsymbol{L}^{-1}$矩阵中相应元素。

当因素剔除后，回归方程变成了新形式，回归平方和也发生了变化。我们把剔除某因素后，回归平方和所发生的变化叫偏回平方和。

原回归平方和 $U=\sum_k\hat{b}_kl_{k0}$，剔除因素后的回归平方和 $U'=\sum_{k\neq i}\hat{b}_k^*l_{k0}$，则剔除第 i 个因素 X_i 的偏回归平方和为 $U_i=U-U'$。由数理统计学知：

$$U_i=U-U'=\frac{\hat{b}_k^2}{C_{ii}} \tag{4.3.53}$$

由上述可知，回归平方和 U 体现出所有变量对 Y 的影响，而偏回归平方和 U_i 则体现出第 i 个因素单独对 Y 的影响。如果因素 X_i 对 Y 的影响大，则 U_i 大，反之则小。

式(4.3.52)、式(4.3.53)提供了在一次回归之后，重新进行回归计算的方便条件，可使计算工作大大简化。

在剔除因素得到新模型之后，再经过检验，直到所有剩下的因素都通过 t 检验为止，即可确定回归系数的区间估计。

其计算公式为

$$d_k=t_{1-\frac{\alpha}{2}}(n-m-1)\sqrt{C_{kk}\cdot S_y} \tag{4.3.54}$$

区间为：

$$(\hat{b}_k-d_k,\hat{b}_k+d_k) \tag{4.3.55}$$

在 $X_1,X_2,\cdots,X_m$ 给出确定值之后，根据 $\hat{Y}=\hat{b}_0+\hat{b}_1X_1+\hat{b}_2X_2+\cdots+\hat{b}_mX_m$ 可预测出 Y 值，但尚需估计误差范围。一般采用剩余标准差 S_y 的 2 倍，即可保证 $\alpha=0.05$ 的置信度。

故 Y 的估计误差为 $2S_y$，区间为 $(Y-2S_y,Y+2S_y)$。

现举例说明多元回归预测的方法和步骤。

某建筑公司认为平均造价与工程结构有关，现对 13 个建筑公司高层建筑造价进行调查得表 4.3.5，现该公司准备根据该资料建立快速报价估算模型。

表 4.3.5 影响造价因素调查表

工程编号	机械化施工所占的比重/%	主体结构工程量所占的比重/%	基础工程量所占的比重/%	装饰工程量所占的比重/%	平均造价/万元
1	10	68	8	12	109.4
2	11	66	9	12	113.3
3	21	47	4	26	115.9
4	1	40	23	34	83.8
5	1	31	22	44	72.5
6	3	71	17	6	102.7
7	2	54	18	22	93.1
8	7	26	6	60	78.5
9	1	29	15	52	74.3
10	11	56	8	20	104.3
11	11	31	8	47	87.6
12	7	52	6	33	95.9
13	11	55	9	22	109.2
平均值	7.462	48.154	117.69	30.000	95.423

设每千平方米的平均造价为 Y,影响平均造价的因素为:机械化施工量所占的比重 X_1,主体结构工程量所占的比重 X_2,基础工程量的所占比重 X_3,装饰工程量的所占比重 X_4。

该公司平均造价预测模型的一般形式为

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \hat{b}_4 X_4$$

其模型建立过程为:

(1) 求 L 矩阵,建立正规方程组。根据

$$l_{ki} = \sum_i (X_{ki} - \bar{X})(X_{ji} - \bar{X}_j) \text{ 和 } l_{k0} = \sum_i (X_{ki} - \bar{X}_k)(Y_i - \bar{Y})$$

得:
$$\boldsymbol{L} = \begin{bmatrix} 415.23 & 251.08 & -372.62 & -290.00 \\ 251.08 & 2905.69 & -166.54 & -3041.00 \\ -372.62 & -166.54 & 492.31 & 38.00 \\ -290.00 & -3041.00 & 38.00 & 3362.00 \end{bmatrix}; \boldsymbol{L}_0 = \begin{bmatrix} 775.96 \\ 2292.95 \\ -618.23 \\ -2481.70 \end{bmatrix}$$

正规方程组为

$$\begin{aligned} 415.23\hat{b}_1 + 251.08\ \hat{b}_2 - 372.62\hat{b}_3 - \quad 290\hat{b}_4 &= 775.96 \\ 251.08\hat{b}_1 + 2905.69\hat{b}_2 - 166.54\hat{b}_3 - 3041.00\hat{b}_4 &= 2292.95 \\ -372.62\hat{b}_1 - 166.54\ \hat{b}_2 + 492.31\hat{b}_3 + \quad 38.00\hat{b}_4 &= -618.23 \\ -290.00\hat{b}_1 - 3041.00\hat{b}_2 + 38.00\ \hat{b}_3 + 3362.00\hat{b}_4 &= -2481.70 \end{aligned}$$

(2) 求解正规方程组和求 $\hat{b}_0$。求解结果为

$\hat{b}_1 = 1.5511$, $\hat{b}_2 = 0.5101$, $\hat{b}_3 = 0.1019$, $\hat{b}_4 = -0.1441$;

根据 $\hat{b}_0 = \bar{Y} - \sum_k \hat{b}_k \bar{X}_k$,求得 $\hat{b}_0$ 为

$$\hat{b}_0 = 62.4062$$

(3) 建立预测模型。预测模型为

$$Y = 62.4052 + 1.5511X_1 + 0.5101X_2 + 0.1019X_3 - 0.1441X_4$$

(4) 回归显著性检验

$$L_{YY} = \sum_i (Y_i - \bar{Y})^2 = 2715.76 \qquad U = \sum_k \hat{b}_k l_{k0} = 2667.84$$

全相关系数 $R^2 = \dfrac{U}{L_{YY}}$ 为 $R^2 = 0.9824, R = 0.9911$,

采用 F 检验,求临界值 F_α 和 F。

$$F_{1-\alpha}(m, n-m-1) = F_{0.95}(4,8) = 3.84$$

$$F = \frac{n-m-1}{m} \cdot \frac{R^2}{1-R^2} = 111.3, F > F_\alpha$$

故回归显著。

(5) 各回归系数检验。剩余平方和 Q 和剩余标准差 S_r 为

$$Q = L_{YY} - U = 47.92, S_r = \sqrt{\frac{Q}{n-m-1}} = 2.447。$$

L^{-1} 矩阵对角线上元素 C_{kk} 为

$$C_{11} = 0.09276, C_{22} = 0.08761, C_{33} = 0.09526, C_{44} = 0.08408。$$

将上述数据代入式(4.3.51)得

$$t_1 = \frac{\hat{b}_1}{\sqrt{C_{11}} \cdot S_y} = 2.081, t_2 = \frac{\hat{b}_2}{\sqrt{C_{22}} \cdot S_y} = 0.7043$$

$$t_3 = \frac{\hat{b}_3}{\sqrt{C_{33}} \cdot S_y} = 0.1349, t_4 = \frac{\hat{b}_4}{\sqrt{C_{44}} \cdot S_y} = -0.2031$$

查 t 分布表,当 $\alpha = 0.1$ 时 $t_{0.95}(n-m-1) = 1.860$,只有 b_1 显著不为零。

(6) 剔除因素。因 $|t_i|$ 中,$t_3 = 0.1349$ 最小,故将 X_3 剔除。按式(4.3.52)

$$\hat{b}_1^* = \hat{b}_1 - \frac{C_{13}}{C_{33}}\hat{b}_3 = 1.4519$$

$$\hat{b}_2^* = \hat{b}_2 - \frac{C_{23}}{C_{33}}\hat{b}_3 = 0.4161$$

$$\hat{b}_4^* = \hat{b}_4 - \frac{C_{43}}{C_{33}}\hat{b}_3 = -0.2365$$

$$\hat{b}_0^* = Y - \sum_{j=k} \hat{b}_j^* \bar{X}_j = 71.6482$$

故新回归方程(预测模型)为

$$\hat{Y} = 71.6482 + 1.4519X_1 + 0.4161X_2 - 0.2365X_4$$

偏回归平方和 U_3 为 $U_3 = \dfrac{\hat{b}_3^2}{C_{33}} = 0.1090$

新预测模型的回归平方和 $U' = U - U_3$ 为

$$U' = 2667.84 - 0.1090 = 2667.73$$

可见 X_3 的影响很小。如此可一直进行到所有因素均显著为止。但需注意的是每剔除一个因素之后，都要重新进行回归系数的检验。

4.4 带定性变量的线性回归预测模型

以前所研究的回归预测方法都有一个共同的特点，这就是影响系统发展变化的因素都可用明确的数量指标描述，如劳动量、材料消耗、产值等。

但实际问题中，除了这类问题之外，还有些不能定量或不能准确定量的因素，如季节、气候等。这些因素虽然难以定量描述，但却可以清楚地分级，如季节可分为春、夏、秋、冬，气候可分好、坏，性别可分男、女等。我们把这类因素叫定性因素。

在分析系统变化规律时，有时定性因素比定量因素对系统的影响还显著，如季节因素对建筑施工的影响就很大，在研究建筑系统时，不考虑这些因素，就不能正确地反映系统变化的规律。因此在进行系统预测时，必须同时考虑定量因素和定性因素。

用以前所述的回归方法分析带定性因素的预测是很困难的。随着数理统计理论的发展，出现了一门专门研究这类问题的方法和理论——数量化理论。研究带定性变量的线性回归问题是数量化理论的一个组成部分，本节简要地介绍处理这类问题的基本理论和方法，深入细致的内容请参阅有关数量化理论的专著。

4.4.1 基本概念

(1) 基准变量：数量化理论把多元分析中的因变量称为基准变量。

(2) 说明变量：数量化理论把多元分析中的自变量叫做说明变量。

(3) 项目和类目：说明变量中的定性变量叫项目，定性变量的分级叫类目。如影响建筑业的生产的因素中季节性是定性变量，叫做项目，而把它分成的春、夏、秋、冬四季称为四个类目。

(4) 反应：第 i 个样本的第 j 个项目值叫反应，记为 $\delta_i(j,k)$，且

$$\delta_i(j,k)=\begin{cases}1 & \text{当第 } i \text{ 样本取第 } j \text{ 项目的第 } k \text{ 类目时}\\ 0 & \text{当第 } i \text{ 样本不取第 } j \text{ 项目的第 } k \text{ 类目时}\end{cases}$$

(5) 反应表：将所占有的资料经整理排列后所得到的表格叫反应表。

(6) 反应矩阵：把反应表中的反应 $\delta_i(j,k)$ 写成矩阵形式，所得矩阵 $\boldsymbol{X}=\{\delta_i(j,k)\}$ 叫反应矩阵。

反应表一般形式见表 4.4.1。

表 4.4.1 反应表

	1	2	…	m
	1 2…r_1	1 2…r_2	1	1 2…r_m
1	$\delta_1(1,1)\delta_1(1,2)\cdots\delta_1(1,r_1)$	$\delta_1(2,1)\delta_1(2,2)\cdots\delta_1(2,r_2)$	…	$\delta_1(m,1)\delta_1(m,2)\cdots\delta_1(m,r_m)$
2	$\delta_2(1,1)\delta_2(1,2)\cdots\delta_2(1,r_1)$	$\delta_2(2,1)\delta_2(2,2)\cdots\delta_2(2,r_2)$	…	$\delta_2(m,1)\delta_2(m,2)\cdots\delta_2(m,r_m)$
⋮	⋮	⋮	…	
n	$\delta_n(1,1)\delta_n(1,2)\cdots\delta_n(1,r_1)$	$\delta_n(2,1)\delta_n(2,2)\cdots\delta_n(2,r_2)$	…	$\delta_n(m,1)\delta_n(m,2)\cdots\delta_n(m,r_m)$

反应矩阵的一般形式为

$$\boldsymbol{X}=\begin{bmatrix}\delta_1(1,1)\ \delta_1(1,2)\cdots\delta_1(1,r_1)\cdots\delta_1(m,1)\ \delta_1(m,2)\cdots\delta_1(m,r_m)\\ \delta_2(1,1)\ \delta_1(1,2)\cdots\delta_2(1,r_1)\cdots\delta_2(m,1)\ \delta_2(m,2)\cdots\delta_2(m,r_m)\\ \cdots\qquad\qquad\cdots\qquad\qquad\cdots\\ \delta_n(1,1)\ \delta_n(1,2)\cdots\delta_n(1,r_n)\cdots\delta_n(m,1)\ \delta(m,1)\cdots\delta_n(m,r_n)\end{bmatrix}$$

如对影响道路施工成本的因素进行分析，得表 4.4.2。

表 4.4.2　影响道路施工成本的因素

项目及类目 / 样本	成本 千元/m	机械化施工所占比重 X_1/%	取土距离 X_2 /m	季节		施工条件		
				冬 X_{31}	夏 X_{32}	好 X_{41}	中 X_{42}	差 X_{43}
1	5	0.7	300	1	0	1	0	0
2	5.1	0.6	400	1	0	0	1	0
3	7	0.2	600	0	1	1	0	0
4	2	0.4	700	0	1	0	0	1
5	8	0 3	300	1	0	0	1	0
6	8	0.4	900	1	0	1	0	0
7	9	0.5	500	0	1	0	0	1
注：路宽以 20m 为标准。								

由反应表可见，隶属于一个项目的反应之和为 1，因此反应矩阵是奇异矩阵，可以证明，反应矩阵 $\boldsymbol{X}$ 的秩为

$$\text{rank}\quad \boldsymbol{X}=\sum_{j=1}^{m} r_j - m + 1$$

式中　m——定性因子数；

r_j——第 j 个定性因子的类目数。

4.4.2　基本方法

带定性因子的线性回归分析步骤如下：

(1) 整理所占有的数据，列出反应表，得到反应矩阵。

(2) 建立模型。

设所研究的系统有 m 个说明变量且第 j 个项目有 r_j 个类目。

数量化理论假设，基准变量 Y_i 与各项目、类目的反应 $\delta_i(j,k)$ 之间有如下关系：

$$Y_i=\sum_{j=1}^{m}\sum_{k=1}^{r_j}\delta_i(j,k)\cdot b_{jk}+e_i \tag{4.4.1}$$

式中　b_{jk}——依赖于第 j 项目第 k 类目的待定常数；

Y_i、$\delta_i(j,k)$——第 i 个样本的基准变量值和反应；

e_i——误差。

把上式写成矩阵形式为

$$\boldsymbol{Y}=\boldsymbol{XB}+\boldsymbol{e} \tag{4.4.2}$$

式中

$$\boldsymbol{Y}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix};\quad \boldsymbol{X}=|\delta_i(j,k)|\quad ;\quad \boldsymbol{B}=\begin{bmatrix} b_{11} \\ \vdots \\ b_{1r_1} \\ b_{21} \\ \vdots \\ b_{2r_2} \\ b_{m1} \\ \vdots \\ b_{mr_m} \end{bmatrix};\qquad \boldsymbol{e}=\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

误差平方和 Q 为

$$Q=\sum_{i=1}^{n} e_i^2=\boldsymbol{E}^{\mathrm{T}}\boldsymbol{E}=(\boldsymbol{Y}-\boldsymbol{XB})^{\mathrm{T}}(\boldsymbol{Y}-\boldsymbol{XB})=$$
$$(\boldsymbol{Y}^{\mathrm{T}}-\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})(\boldsymbol{Y}-\boldsymbol{XB})=$$
$$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}-\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{XB}-\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{XB}$$

$\because\quad \boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y})^{\mathrm{T}}=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{XB}\qquad (\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{XB}\text{ 为数量})$

$$Q=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}-2\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{XB}$$

用最小二乘法以 $\dfrac{\partial Q}{\partial B}=0$ 确定参数 B。

由矩阵微分法得正规方程：

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{XB}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} \tag{4.4.3}$$

由于 $\boldsymbol{X}$ 是奇异矩阵，故 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 也是奇异的，其逆不存在。数量化理论证明了 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 的秩为 $\sum_j r_j-(m-1)$，且用 $\boldsymbol{B}$ 的一组特解作预测可反映系统变化规律。

因此采用如下解法：

删除第 j 项目第 1 类目所对应的方程（$j=2,3,\cdots,m$），并令 $b_{j1}=0$。因此共删掉 $m-1$ 个方程，使新正规方程的系数矩阵成为非奇异矩阵，故可惟一地求出一组解，这组解称为正规方程组的特解，记为 B^0：

$$(\boldsymbol{B}^0)^{\mathrm{T}}=(b_{11}^0,\ b_{12}^0,\cdots,\ b_{1r_1}^0,0,\ b_{22}^0,\cdots,\ b_{2r_2}^0,\cdots,0,\ b_{m_2}^0,\cdots,\ b_{mr_m}^0)$$

由此可得到预测方程为

$$\boldsymbol{Y}=\sum_{j=1}^{m}\sum_{k=1}^{r_j}\delta_i(j,k)\hat{b}_{jk}^0 \tag{4.4.4}$$

以上讨论为说明变量全部是定性因子的情况。如果说明变量中尚有定量因子时，把定量因子的定量数据列入反应矩阵中，按同样的方法求解正规方程，即可求得参数 B。

现以表 4.4.2 的数据为例，说明带定性变量问题的线性回归方法。

该例中：

$$\boldsymbol{X}=\begin{bmatrix}0.7 & 5 & 1 & 0 & 1 & 0 & 0\\ 0.6 & 4 & 1 & 0 & 0 & 1 & 0\\ 0.2 & 6 & 0 & 1 & 1 & 0 & 0\\ 0.4 & 7 & 0 & 1 & 0 & 0 & 1\\ 0.3 & 3 & 1 & 0 & 0 & 1 & 0\\ 0.4 & 9 & 1 & 0 & 1 & 0 & 0\\ 0.5 & 5 & 0 & 1 & 0 & 0 & 1\end{bmatrix};\qquad \boldsymbol{Y}=\begin{bmatrix}5.0\\ 5.1\\ 7.0\\ 12.0\\ 8.0\\ 8.0\\ 9.0\end{bmatrix};$$

去掉第二项目的第一类目后，按式(4.4.3)，正规方程为

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}=\begin{bmatrix}0.7 & 0.6 & 0.2 & 0.4 & 0.3 & 0.4 & 0.5\\ 5 & 4 & 6 & 7 & 3 & 9 & 5\\ 1 & 1 & 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 1 & 1 & 0 & 0 & 1\\ 0 & 1 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}0.7 & 5 & 1 & 0 & 0 & 0\\ 0.6 & 4 & 1 & 0 & 1 & 0\\ 0.2 & 6 & 0 & 1 & 0 & 0\\ 0.4 & 7 & 0 & 1 & 0 & 1\\ 0.3 & 3 & 1 & 0 & 1 & 0\\ 0.4 & 9 & 1 & 0 & 0 & 0\\ 0.5 & 5 & 0 & 1 & 0 & 1\end{bmatrix}=$$

$$\begin{bmatrix}1.55 & 16.9 & 2 & 1.1 & 0.9 & 0.9\\ 16.9 & 241 & 21 & 18 & 7 & 12\\ 2 & 21 & 4 & 0 & 2 & 0\\ 1.1 & 18 & 0 & 3 & 0 & 2\\ 0.9 & 7 & 2 & 0 & 2 & 0\\ 0.9 & 12 & 0 & 2 & 0 & 2\end{bmatrix}$$

$$\begin{bmatrix}1.55 & 16.9 & 2 & 1.1 & 0.9 & 0.9\\ 16.9 & 241 & 21 & 18 & 7 & 12\\ 2 & 21 & 4 & 0 & 2 & 0\\ 1.1 & 18 & 0 & 3 & 0 & 2\\ 0.9 & 7 & 2 & 0 & 2 & 0\\ 0.9 & 12 & 0 & 2 & 0 & 2\end{bmatrix}^{-1}\cdot\begin{bmatrix}b_1\\ b_2\\ b_{31}\\ b_{32}\\ b_{42}\\ b_{43}\end{bmatrix}=\begin{bmatrix}22.86\\ 312.4\\ 26.1\\ 28.0\\ 13.1\\ 21\end{bmatrix}$$

求解结果为

$\boldsymbol{B}=(-9.6621,0.2129,10.3231,7.6545,-0.1707,0.0000,5.9159)^{\mathrm{T}}$

预测模型为

$$\hat{Y}=-9.6621X_1+0.212X_2+10.3231X_{31}+7.6545X_{32}-0.1707X_{42}+5.9159X_{43}$$

如现在道路施工中机械化比重为0.6、取土距离为400m、施工条件较差，在夏季进行作业时，其每延米施工成本的预测值应为

$$\hat{Y}=-9.6621\times0.6+0.212\times4+10.3231\times0+7.6545\times1-0.1707\times0+5.9159\times1$$
$$=8.6247$$

这样即解决了含定性因子的系统预测问题。

4.5 判别分析预测模型与综合评审

我们所研究的系统是复杂的，实践中对系统发展变化的估计有时不需要以定量的方式描述，而只需以程度来评价。如系统发展的前景好与坏等。这就是说，系统内部各因子（自变量）可以是定性的，也可以是定量的，而系统的变化规律则以定性的方式体现。判别分析就是解决这类问题的方法，它也是数量化理论研究的问题之一。

4.5.1 判别分析

4.5.1.1 二级判别分析的基本思想

当一个系统的发展变化只有两个方向的时候，根据对占有资料的分析，确定系统某状态下发展方向的方法叫二级判别。如果系统存在多种发展方向，根据对占有资料的分析，确定系统某状态下的发展方向叫多级判别。二级判别是多级判别的基础。因此现以二级判别分析为例，介绍判别分析预测的基本思想。

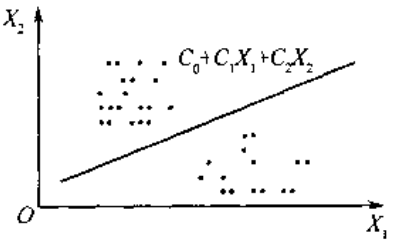


图 4.5.1 销售状况散点图

下面举例说明二级判别的基本思想。

某房地产商准备开发一新型住宅，为确定消费群体状况，以售价和消费水平作为分析影响销售情况的因素，现把消费者购买意向按有意向和无意向二种情况分别汇总整理，得出表 4.5.1，并以消费水平为 X_1，价格水平为 X_2，做出预计销售状况散点图见图 4.5.1。

表 4.5.1 消费者购买意向

组别		消费水平 X_1（月收入）	价格水平 X_2/（元/m²）
有意向	A_1	$X_{1,1}^{(A)}$	$X_{2,1}^{(A)}$
	A_2	$X_{1,2}^{(A)}$	$X_{2,2}^{(A)}$
	…	…	…
	A_{n1}	$X_{1,n1}^{(A)}$	$X_{2,n2}^{(A)}$
无意向	B_1	$X_{1,1}^{(B)}$	$X_{2,1}^{(B)}$
	B_2	$X_{1,2}^{(B)}$	$X_{2,2}^{(B)}$
	…	…	…
	B_{n2}	$X_{1,n1}^{(B)}$	$X_{2,n2}^{(B)}$

从散点图看，有意向组和无意向组基本可用一条直线 $L: C_0 + C_1X_1 + C_2X_2$ 将其分成两部分，位于上方的点基本是有意向点，位于直线下方的点基本上是无意向点。用数学描述，则 $C_0 + C_1X_1 + C_2X_2 > 0$ 或 $C_1X_1 + C_2X_2 > -C_0$ 为有意向；$C_0 + C_1X_1 + C_2X_2 < 0$

或 $C_1X_1+C_2X_2<-C_0$ 为无意向。

如果能求出这条直线方程即可解决该产品在 X_1、X_2 的各种水平下的预计销售状况，也就是说 $C_1X_1+C_2X_2$ 可作为判别 X_1、X_2 为不同水平时是有意向还是无意向的函数，$-C_0$ 是判别的准则。

$$Y=C_1X_1+C_2X_2 \tag{4.5.1}$$

称之为判别函数。

$$Y_0=C_0 \tag{4.5.2}$$

称之为判据。

如果通过对占有数据的分析，确定出 C_0、C_1、C_2 的值，则对于给定的 X_1 和 X_2，就可求出 Y_0，当 $Y>Y_0$ 时，可判给定水平为有意向，否则判为无意向。这就解决了所提出的问题，预测了系统的发展。

由上述可见，判别预测的关键就是根据什么原则处理所占有的资料和如何求出 C_0、C_1 和 C_2。

4.5.1.2 最优判别准则、判别函数和判据的确定

因为影响系统发展方向的因素很多，为了判别系统的发展方向，就必须拟定一个综合评价指标，也就是借助于它把各因素统一起来。

设影响系统的因素有 p 个，则对系统的综合评价指标可表示为

$$Y=C_1X_1+C_2X_2+\cdots+C_pX_p \tag{4.5.3}$$

现把系统发展变化的方向分为二组，一组称为 A 组，一组称为 B 组，A 组的综合评价指标以 $\overline{Y}(A)$ 表示，B 组的综合评价指标以 $\overline{Y}(B)$ 表示。现有属 A 组的数据 n_1 组，属 B 组的数据 n_2 组(见表 4.5.2)。故各组的综合评价指标分别为

表 4.5.2 分组数据表

组 别	序 号	X_1	X_2	…	X_p
A 组	1	$X_{1,1}^{(A)}$	$X_{2,1}^{(A)}$	…	$X_{p,1}^{(A)}$
	2	$X_{1,2}^{(A)}$	$X_{2,2}^{(A)}$	…	$X_{p,2}^{(A)}$
	⋮	…	…	…	…
	n_1	$X_{1,n_1}^{(A)}$	$X_{2,n_1}^{(A)}$	…	$X_{p,n_1}^{(A)}$
B 组	1	$X_{1,1}^{(B)}$	$X_{2,1}^{(B)}$	…	$X_{p,1}^{(B)}$
	2	$X_{1,2}^{(B)}$	$X_{2,2}^{(B)}$	…	$X_{p,2}^{(B)}$
	⋮	…	…	…	…
	n_2	$X_{1,n_2}^{(B)}$	$X_{2,n_2}^{(B)}$	…	$X_{p,n_2}^{(B)}$

$$\overline{Y}(A)=\frac{1}{n_1}\sum_{i=1}^{n_1}Y_i^{(A)}$$

$$\overline{Y}(B)=\frac{1}{n_2}\sum_{i=1}^{n_2}Y_i^{(B)} \tag{4.5.4}$$

其中，

$$Y_i^{(A)}=\sum_{j=1}^{p}C_jX_{ji}^{(A)}$$

$$Y_i^{(B)} = \sum_{j=1}^{p} C_j X_{ji}^{(B)}$$

从两组的综合评价指标来看，如果两组指标的差距越大则判别越容易，同时如果组内综合评价指标差距越小，就说明它们越集中在 $\bar{Y}$ 的附近，也有利于判别，根据这个道理，费歇尔提出了以

$$\max L = \frac{(\bar{Y}(A) - \bar{Y}(B))^2}{\sum_{i=1}^{n_1}(Y_i^{(A)} - \bar{Y}(A))^2 + \sum_{i=1}^{n_2}(Y_i^{(B)} - \bar{Y}(B))^2} \tag{4.5.5}$$

作为确定分类的原则，该原则称费歇尔原理。

由式(4.5.5)可知，$L \to \max$ 是在 $(\bar{Y}(A) - Y(B))^2 \to \max$ 和 $\sum_{i=1}^{n_1}(Y_{ii}^{(A)(A)} - \bar{Y}(A))^2 + \sum_{j=1}^{n_2}(Y_i^{(B)} - \bar{Y}(B))^2 \to \min$ 时实现的，因此，费歇尔原则是以组间差距最大，组内差距最小作为判别原则的，这也正符合上述分析中所提出的组间差距越大，组内差距越小越有利于判别的原则。其几何意义可用图 4.5.2 表示。

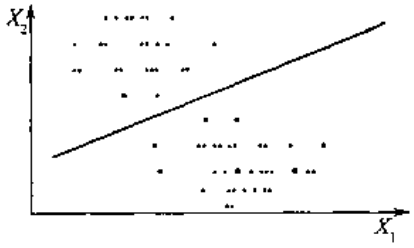


图 4.5.2　费歇尔原理示意图

因为 L 是 C_j 的函数，故应用微分法，当 $\frac{\partial L}{\partial C_j} = 0, j = 1,2,\cdots p$ 时，即可求出使 L 达最大值的 C_j。

对式(4.5.5)求导后，经推导得出 C_j 是下述方程组的解：

$$\begin{cases} S_{11}C_1 + S_{12}C_2 + \cdots + S_{1p}C_p = \bar{X}_1^{(A)} - \bar{X}_1^{(B)} \\ S_{21}C_1 + S_{22}C_2 + \cdots + S_{2p}C_p = \bar{X}_2^{(A)} - \bar{X}_2^{(B)} \\ \cdots \\ S_{p1}C_1 + S_{p2}C_2 + \cdots + S_{pp}C_p = \bar{X}_p^{(A)} - \bar{X}_p^{(B)} \end{cases} \tag{4.5.6}$$

其中，
$$S_{ij} = \sum_{k=1}^{n_1}(X_{ik}^{(A)} - \bar{X}_i^{(A)})(X_{jk}^{(A)} - \bar{X}_j^{(A)}) + \sum_{k=1}^{n_2}(X_{ik}^{(B)} - \bar{X}_i^{(B)})(X_{jk}^{(B)} - \bar{X}_j^{(B)})$$

$$S_{kk} = \sum_{i=1}^{n_1}(X_{ki}^{(A)} - \bar{X}_k^{(A)})^2 + \sum_{i=1}^{n_2}(X_{ki}^{(B)} - \bar{X}_k^{(B)})^2$$

求解式(4.5.6)即可确定 $C_j (j = 1,2,\cdots,p)$，从而确定判别函数

$$Y = C_1X_1 + C_2X_2 + \cdots + C_pX_p = \sum_{j=1}^{p} C_jX_j \tag{4.5.7}$$

判据的选择方法是以 A 组和 B 组数据量的大小为权数，计算其综合评价指标的加权平均数 Y_c，其计算公式如下：

$$Y_c=\frac{n_1\overline{Y}(A)+n_2\overline{Y}(B)}{n_1+n_2} \tag{4.5.8}$$

当 $X_1=X_1^0, X_2=X_2^0, \cdots, X_p=X_p^0$ 时，代入判别函数得：

$$Y_o=C_1X_1^0+C_2C_2^0+\cdots+C_pX_p^0$$

如果 $Y_o>Y_c$ 且$\overline{Y}(A)>\overline{Y}(B)$，则属 A 组；

$\overline{Y}(A)<\overline{Y}(B)$，则属 B 组；

如果 $Y_o<Y_c$ 且$\overline{Y}(A)>\overline{Y}(B)$，则属 B 组；

$\overline{Y}(A)<\overline{Y}(B)$，则属 A 组。

这样就解决了这类系统的预测问题。

4.5.1.3 判别分析检验

为考查判别分析的可靠性，需根据费歇尔的判别原则，考虑 A 组数据和 B 组数据在 $\overline{Y}(A)$和 $\overline{Y}(B)$的概率来分析。假设 A、B 组各因素的指标均服从正态分布，则 A 组指标在 $\overline{Y}(A)$处的概率密度大，B 组在 $\overline{Y}(B)$处的概率密度大，如果 $\overline{Y}(A)$与 $\overline{Y}(B)$的距离较大，则在 Y_c 处的概率密度小，反之则较大。当预测指标接近 Y_c 时，判误的可能性增大，因此需对判别模型进行检验，以便以一定的置信度确定其可靠性。

检验方法为：(1) 根据 $F_{1-\alpha}(P, n_1+n_2-P-1)$，确定临界值 F_α。

(2) 按下式计算 F 值

$$F_\alpha=\frac{n_1n_2(n+n_2-P-1)}{P(n_1+n_2)(n_1+n_2-2)}\cdot D^2 \tag{4.5.9}$$

其中，$D^2=(\overline{X}(A)-\overline{X}(B))^{\mathrm{T}}S^{-1}(\overline{X}(A)-\overline{X}(B))$

$$\overline{X}(A)=(\overline{X}_1(A), \overline{X}_2(A), \cdots, \overline{X}_p(A))^{\mathrm{T}}$$

$$\overline{X}(B)=(\overline{X}_1(B), X_2(B), \cdots, \overline{X}_p(B))^{\mathrm{T}}$$

S 为式(4.5.6)的系数矩阵。

(3) 比较 F 和 F_α，如果 $F>F_\alpha$，则判别效果良好，否则判别效果较差。

4.5.1.4 多级判别

在实践中，往往系统发展变化的方向不止两个，而是多个，如销售状况可能存在畅销、一般和滞销三种，质量分级为好、中、差三种等。实现这类问题的判别称为多级判别。

根据二级判别的原理和方法，采用逐级分级判别的方法可实现多级判别。

其步骤为：

(1) 将问题先分成两大类，建立二级判别模型，求出判据；

(2) 在上述的每类中，再分成两类，建立二级判别模型，求出判据；

(3) 重复(2)的作法，直到满足分级需要为止；

(4) 对预测对象先按第一级判别函数和判据确定从属的子类，一直到确定出明确归属为止。

4.5.1.5 判别分析的应用

某市郊区镇进行土地定级工作，依据土地定级规程并结合本镇的具体情况，拟采用网格法，以繁华度、交通便捷度和文化氛围三个指标作为土地质量评价指标(见表 4.5.3)。

网格设置为100m×100m,并以网格中心至商、服中心的距离作为评价繁华度的因素,以距主要交通枢纽站的距离作为评价交通便捷度的因素,以距主要文化设施的距离作为评价文化氛围的因素。现选择12块典型宗地进行详细调查,并按表4.5.4将实际距离转换成分值。现依据典型宗地的分值和定级状况建立判别分析模型对该镇土地定级。

表4.5.3 土地质量评价指标

级 别	繁华度分值	交通便捷度分值	文化氛围分值
Ⅰ	9 $X_{11}^{(A)}$	8 $X_{21}^{(A)}$	7 $X_{31}^{(A)}$
Ⅰ	7 $X_{12}^{(A)}$	6 $X_{22}^{(A)}$	6 $X_{32}^{(A)}$
Ⅰ	10 $X_{13}^{(A)}$	7 $X_{23}^{(A)}$	8 $X_{33}^{(A)}$
Ⅰ	8 $X_{14}^{(A)}$	4 $X_{24}^{(A)}$	5 $X_{34}^{(A)}$
Ⅰ	9 $X_{15}^{(A)}$	9 $X_{25}^{(A)}$	3 $X_{35}^{(A)}$
Ⅰ	8 $X_{16}^{(A)}$	6 $X_{26}^{(A)}$	7 $X_{36}^{(A)}$
Ⅰ	7 $X_{17}^{(A)}$	5 $X_{27}^{(A)}$	6 $X_{37}^{(A)}$
平均值	$\bar{X}_1^{(A)}=8.29$	$\bar{X}_2^{(A)}=6.43$	$\bar{X}_3^{(A)}=6.00$
Ⅱ	4 $X_{11}^{(B)}$	4 $X_{21}^{(B)}$	4 $X_{31}^{(B)}$
Ⅱ	3 $X_{12}^{(B)}$	6 $X_{22}^{(B)}$	6 $X_{32}^{(B)}$
Ⅱ	6 $X_{13}^{(B)}$	3 $X_{23}^{(B)}$	3 $X_{33}^{(B)}$
Ⅱ	2 $X_{14}^{(B)}$	4 $X_{24}^{(B)}$	5 $X_{34}^{(B)}$
Ⅱ	1 $X_{15}^{(B)}$	2 $X_{25}^{(B)}$	2 $X_{35}^{(B)}$
平均值	$\bar{X}_1^{(B)}=3.2$	$\bar{X}_2^{(B)}=3.8$	$\bar{X}_3^{(B)}=4.0$

表4.5.4 距离—分值转换表

距离/m	分值	距离/m	分值
1000以内	1	3000～3500	6
1000～1500	2	3500～4000	7
1500～2000	3	4000～4500	8
2000～2500	4	4500～5000	9
2500～3000	5	5000～5500	10

步骤1 根据式(4.5.6)求 $\boldsymbol{S}$ 矩阵:

$$\boldsymbol{S}=\begin{bmatrix} 22.229 & 8.3429 & 2 \\ 8.3429 & 26.5143 & 6 \\ 2 & 6 & 26 \end{bmatrix}$$

步骤2 解方程组,

$$\begin{cases} 22.22\ C_1 + 8.3429\ C_2 + 2\ C_3 = 5.09 \\ 8.3429C_1 + 26.5143C_2 + 6\ C_3 = 2.63 \\ 2C_1 + \quad 6C_2 + 26C_3 = 2 \end{cases}$$

得 $C_1 = 0.217, C_2 = 0.018, C_3 = 0.056$

步骤 3 建立判别函数：

$$Y = 0.217X_1 + 0.018X_2 + 0.056X_3$$

步骤 4 求判据：

$$Y_C = \frac{n_1 \overline{Y}^{(A)} + n_2 \overline{Y}^{(B)}}{n_1 + n_2} = \frac{7 \times 2.25 + 5 \times 0.99}{12} = 1.725$$

步骤 5 检验(略)。

步骤 6 预测。如某宗地数据为：

$$X_1 = 9; X_2 = 5; X_3 = 8$$

则 $Y_0 = 0.217 \times 9 + 0.018 \times 5 + 0.056 \times 8 = 2.491$

因 $2.491 > 1.725$，且 $\overline{Y}^{(A)} > \overline{Y}^{(B)}$，故该宗地属于Ⅱ级。

这样即可解决多种评价指标下的判断准则问题。该方法除可用于判断预测外还可用于自动化设备和计算机自动检测系统。

4.5.2 综合评审

综合评审是系统决策的预备工作，是应用模糊集合的概念、用定量化的方法预测系统发展可能性的一种方法，特别适用于多因子或多目标的系统。

现举例说明评审方法。

某房地产开发商对新开发的产品——嘉和花园进行市场预测，预测拟从房型、区位、价格三方面了解不同类型家庭的反应，以得出不同类型家庭的需求，为确定该产品的目标市场提供决策依据。假设消费者的反应可分为非常欢迎、欢迎、不太欢迎、不欢迎四种，现采用综合评审方法进行预测。

综合评审方法是在分别考虑预测对象的单一指标的基础上了解消费者的反应后得到预测结论的。

现仅考虑房型，而不考虑区位和价格，得到某类消费者的反应为 r_{1j}：

$$r_{1j} = (0.2, 0.7, 0.1, 0)$$

r_{1j}的含义为消费者对房型所持的态度向量，$r_{11} = 0.2$ 为有 20% 的消费者持非常欢迎的态度，$r_{12} = 0.7$ 为有 70% 的消费者持欢迎的态度；$r_{13} = 0.1$ 为有 10% 的消费者持不太欢迎的态度；$r_{14} = 0$ 为没有消费者对房型不欢迎。

同理，经调查亦可得到仅考虑区位、不考虑房型和价格，以及仅考虑价格、不考虑房型和区位的某类消费者反应 r_{2j}、r_{3j}，假设 r_{2j}、r_{3j} 如下：

$$r_{2j} = (0, 0.4, 0.5, 0.1)$$

$$r_{3j} = (0.2, 0.3, 0.4, 0.1)$$

由此得到消费者的反应矩阵 $\boldsymbol{R} = \{r_{ij}\}$。

$$\boldsymbol{R}=\begin{Bmatrix}0.2 & 0.7 & 0.1 & 0\\ 0 & 0.4 & 0.5 & 0.1\\ 0.2 & 0.3 & 0.4 & 0.1\end{Bmatrix}$$

召集某类家庭成员讨论，并采用AHP法尚可确定房型、区位和价格三因素的重要性。如知识分子家庭认为三要素的重要性为：A=(0.2,0.5,0.3)，则 $\boldsymbol{AR}=\boldsymbol{B}$ 为知识分子家庭对该产品的综合评审向量。

$$\boldsymbol{B}=\boldsymbol{AR}=(0.2\quad 0.5\quad 0.3)\cdot\begin{bmatrix}0.2 & 0.7 & 0.1 & 0\\ 0 & 0.4 & 0.5 & 0.1\\ 0.2 & 0.3 & 0.4 & 0.1\end{bmatrix}=(0.1\quad 0.43\quad 0.39\quad 0.08)$$

$\boldsymbol{B}$ 元素的含义为：10% 的知识分子家庭非常欢迎该产品；43% 的人欢迎；39% 的人不太欢迎，而 8% 的人不欢迎。

如果将非常欢迎定为 100 分，欢迎定为 80 分，不太欢迎定为 60 分，不欢迎定为 40 分，则评审结果为：

$$\boldsymbol{K}=\boldsymbol{BP}$$

式中 $\boldsymbol{P}$——评审等级评分。

$$\boldsymbol{K}=(0.1\quad 0.43\quad 0.39\quad 0.08)\cdot(100\quad 80\quad 60\quad 40)^{\mathrm{T}}=71$$

这说明，该新产品对知识分子家庭处于欢迎与不太欢迎之间。

总结该例，对综合评审的描述如下：

设评审准则单元为

$E=\{e_1、e_2、\cdots、e_n\}$（相当于上例中的房型、区位、价格等）

评审评语集合为

$V=\{v_1、v_2、\cdots、v_m\}$（相当于上例中的非常欢迎、欢迎、不太欢迎、不欢迎等结论评语）

令 r_{ij} 为用 i 评审准则衡量评审对象后得出评语 v_j 的可能性（一般 $0\leqslant r_{ij}\leqslant 1$）。

固定 i，用单一因子 e_i 衡量所有对象，得出 r_{ij}，当改变 i 且 i 遍历 $i=1,2,\cdots,n$ 后得到反应矩阵 $\boldsymbol{R}$，分别确定评审准则的重要性系数和评语相应的等级分数，按 $\boldsymbol{K}=\boldsymbol{ARP}$ 即可确定评审结论。

综合评审的另一种算法是将布尔代数的运算推广到一般代数运算中（即“∪”运算，类似于代数运算中的加法，称为“逻辑加”，为参与运算的数字中取大者；“∩”运算类似于代数运算中的乘法，称为“逻辑乘”，为参与运算的数字中取小者；）对上式进行运算：

$$\begin{aligned}\boldsymbol{B}=\boldsymbol{AR}&=(0.2\quad 0.5\quad 0.3)\begin{Bmatrix}0.2 & 0.7 & 0.1 & 0\\ 0 & 0.4 & 0.5 & 0.1\\ 0.2 & 0.3 & 0.4 & 0.1\end{Bmatrix}\\&=(0.2\quad 0.4\quad 0.5\quad 0.1)\end{aligned}$$

将其规范化（即 $\sum_j b_j=1$）：

$$\boldsymbol{B}=(0.17\quad 0.34\quad 0.40\quad 0.09)$$

这说明该产品对知识分子家庭不太欢迎占 40%，欢迎次之占 34%。然后按一般代数运算规则计算评审结论：$\boldsymbol{K}=\boldsymbol{BP}$

$$\boldsymbol{K}=(0.17 \quad 0.34 \quad 0.40 \quad 0.09)\cdot(100 \quad 80 \quad 60 \quad 40)^{\mathrm{T}}=71.8$$

结论与上法基本相同。

4.6 时间序列预测

回归分析是用于变量之间呈相关关系时的一种建立预测模型的方法。如果原始数据是一随时间变化的数列，即时间序列，采用回归分析法时自变量为时间 t，所建立模型形式为 $y=f(t)$。采用趋势外推原理则可对未来进行预测。但回归分析只是时间序列分析的一种方法，由时间序列本身特性所决定，它还有其特定的分析方法。

4.6.1 时间序列分析的内容

4.6.1.1 时间序列的形成

时间序列是由以下四种情况合成的结果：

(1) 长期趋势的变化 X_t，序列随时间呈现的倾向性变化；

(2) 季节性周期变化 S_t，序列在一年中随季节呈现有规律性的周期性变化；

(3) 循环变化 C_t，序列以不固定的周期呈现出的波动性变化；

(4) 随机变化 ε_t，各种不确定因素作用下的无规则变化。

4.6.1.2 时间序列模型

时间序列模型分为加法模型和比例模型两类。

1. 加法模型

加法模型理论认为时间序列是长期趋势 X_t、季节性变化 S_t、循环变化 C_t 以及随机变化 ε_t 四种变化的叠加，故模型形式为

$$y_t = X_t + S_t + C_t + \varepsilon_t$$

2. 比例模型

比例模型理论认为，时间序列的形成是以趋势变化 X_t 为主干，其他变化均是对趋势变化的修正，故模型形式为

$$y_t = X_t \cdot S_t \cdot C_t \cdot \varepsilon_t$$

4.6.1.3 时间序列分析的内容

时间序列分析的内容见表 4.6.1。

表 4.6.1 时间序列分析的内容

内　容	方法类别	方　法
趋势分析	移动平均法	简易移动平均法； 加权平均法； 趋势修正移动性系数法； 二次移动平均数法
	指数平滑法	指数平滑法； 二次指数平滑法； 三次指数平滑法

（续）

内　容	方法类别	方　法
季节性变化分析	季节性分析法	简易季节性分析法； 周期图分析法
随机变化分析 （周期分析、随机分析）	平稳随机序列分析法	自回归分析(AR)； 移动平均分析(MA)； 自回归移动平均分析(ARMA)

4.6.2 趋势预测

4.6.2.1 移动平均数法

时间序列 $x_1, x_2, \cdots, x_t, \cdots$，有

$\bar{x}_t = \dfrac{x_t + x_{t-1} + \cdots + x_{t-N+1}}{N}$，$t \geqslant N$，　　N 为移动平均的期数。

$\bar{x}_t$ 为时间序列的移动平均数序列，记为 $\{\bar{x}, t \geqslant N\}$。

移动平均数序列与原时间序列相比，前者比后者平滑，它是滤除了原序列的某些干扰后的结果，因此更能体现出原序列的趋势变化。

按趋势递推原理，以 $\bar{x}_t$ 作为 $t+1$ 期的预测值即 $y_{t+1} = x_t$，

可得预测模型：$y_{t+1} = x_t = \dfrac{x_t + x_{t+1} + \cdots + x_{t-N+1}}{N}$

4.6.2.2 加权移动平均数法

用移动平均数进行预测是将各期数据的重要性等同对待，如果考虑各期数据的重要性，对每个序列值乘以加权因子，则时间序列的加权平均值序列为

$$x_t = \frac{a_0 x_t + a_1 x_{t-1} + \cdots + a_{N-1} x_{t-N+1}}{N} =$$

$$\omega_0 x_t + \omega_1 x_{t-1} + \cdots + \omega_{t-n} x_{t-n+1}$$

式中 ω_i 为加权因子，应满足 $\sum_{i=1}^{N-1} \omega_i = 1$。

以 $\bar{x}_t$ 作为下一期预测值，即 $y_{t+1} = \bar{x}_t$

则预测模型为　$y_{t+1} = \omega_0 x_t + \omega_1 x_{t-1} + \cdots + \omega_{N-1} x_{t-N+1}$

该模型既可体现对原始数据的平滑，又考虑了原序列各期值的重要性程度，预测结果一般比只考虑趋势的移动平均数法更接近实际。

由上述模型可见，预测值 y_{t+1} 是由 N 期数据按一定比例组成的，一般情况下，近期数据对预测值的影响大，ω 应选较大的值，历史上远期数据对预测值的影响小，ω 应选较小的值。

现以某房地产开发公司某年商品房销售数据为例，用两种预测方法预测下年度 1 月份销售量（见表 4.6.2）。

表 4.6.2　移动平均数法和加权移动平均数法预测

时间/月	1	2	3	4	5	6	7	8	9	10	11	12	1
销售量	18	15	12	20	25	24	26	25	28	26	29	26	
移动平均法 $N=3$				15.0	15.7	19.0	23.0	25.0	25.0	26.3	26.3	27.7	27.0
移动平均法 $N=4$					16.3	18.0	20.3	23.8	25.0	25.8	26.3	27.0	27.3
加权移动平均数法				14.0	16.5	21.2	23.7	25.2	25.2	26.7	26.5	27.8	27.0
注：加权移动平均法中 $N=3$，$a_0=1.5$，$a_1=1$，$a_2=0.5$。													

4.6.2.3　修正移动平均数法

当时间序列呈现增长或减少趋势时，采用移动平均数法将产生滞后现象。产生滞后的原因为：

假设时间序列呈线性增长趋势，则方程为

$$y_t = a + bt$$

当 t 增加至 $t+N$ 时，序列值为 $y_{t+N} = a + bt + Nb$

但采用移动平均数法预测时，预测值为

$$y_{t+N} = \frac{1}{N}\sum_{i=1}^{N} y_{t+i} = a + bt + \frac{(1+2+\cdots+N)}{N}b = a + bt + \frac{N+1}{2}b$$

二者之差为　$Nb - \frac{N+1}{2}b = \frac{N-1}{2}b$

故在 $t+N$ 期，移动平均数法的预测值滞后了 $\frac{N-1}{2}b$。

为了消除移动平均数法预测所产生的滞后，应在移动平均数法预测值的基础上，以 $\frac{N-1}{2}b$ 为修正量对移动平均数预测模型进行修正，故得修正移动平均数法预测模型：

$$y_t = \bar{x}_t + \frac{N-1}{2}b_t$$

$$y_{t+k} = y_t + kb_t$$

式中　y_t——第 t 期预测值；

$\bar{x}_t$——第 t 期移动平均数；

b_t——第 t 期平均增长量，应取移动平均数计算期内的平均增长量。

b_t 可按线性回归公式 $b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$ 计算，为使其计算简化，可使 $\sum_i t_i = 0$，而使用简化公式 $b = \frac{\sum t_i y_i}{\sum_i t_i^2}$。

如上例，预测下年度 1 月份和 5 月份销量：

当 $N=3$ 时，$\bar{x}_1=27$　$b_1=\dfrac{-1\times26+0\times29+1\times26}{1+0+1}=0$

$$y_1=27+\frac{3-1}{2}\times0=27 \qquad y_{1+4}=27+4\times0=27$$

当 $N=4$ 时，$\bar{x}_1=27.3$　$b_1=\dfrac{(-3\times28-1\times26+1\times29+3\times26)\times2}{(-3)^2+(-1)^2+(1)^2+(3)^2}=-0.3$

$$y_1=27.3+\frac{3-1}{2}\times0.3=27 \qquad \hat{y}_{1+4}=27-0.3\times4=25.8$$

（注意：当 $N=4$ 时，为使 $\sum_i t_i=0$，选 $t_1=-3$　$t_2=-1$　$t_3=1$　$t_4=3$，间隔为2，故计算 b_1 时应乘以2。）

4.6.2.4　指数平滑法

对时间序列 x_t，若预测值按

$$y_t=\alpha x_t+(1-\alpha)y_{t-1} \quad 或 \quad y_t=y_{t-1}+\alpha(x_{t-1}-y_{t-1})$$

计算，则该预测法叫指数平滑法，其中 α 为平滑系数且 $0\leqslant\alpha\leqslant1$。

由上式可见，当期预测值是由当期实际值和上期预测值按比例构成的或是由上期预测值与上期预测误差的修正值构成的。

把上式展开，将有助于对该方法的深刻理解：

$$\begin{aligned}y_t&=\alpha x_t+(1-\alpha)y_{t-1}=\alpha x_t+(1-\alpha)[\alpha x_{t-1}+(1-\alpha)y_{t-2}]=\\&\alpha x_t+\alpha(1-\alpha)x_{t-1}+(1-\alpha)^2y_{t-2}=\\&\alpha x_t+\alpha(1-\alpha)x_{t-1}+(1-\alpha)^2[\alpha x_{t-2}+(1-\alpha)y_{t-3}]=\\&\alpha x_t+\alpha(1-\alpha)x_{t-1}+\alpha(1-\alpha)^2x_{t-2}+(1-\alpha)^3y_{t-3}=\cdots=\\&\alpha x_t+\alpha(1-\alpha)x_{t-1}+\alpha(1-\alpha)^2x_{t-2}+\alpha(1-\alpha)^3x_{t-3}+\cdots\end{aligned}$$

由展开式可见：

（1）如 $\alpha=1$ 预测值取当期实际值，$\alpha=0$ 取时间序列的初始值。

（2）预测值是由时间序列值按一定比例构成的，因 $0\leqslant\alpha\leqslant1$，故近期数据占的比重大，距预测期远的数据比重小，当数据量很大时，初始数据对预测值的影响甚微，α 取值大，近期数据占的比重越大，α 取值小，近期数据占的比重越小。

（3）α 值的大小影响预测值。α 大，更贴近原序列，但滞后小；α 小，更平滑，但滞后大。

（4）预测值实质是历史数据的加权平均数，且权数按指数变化，因此该方法叫指数平滑法，是一种特殊的加权移动平均数法。

4.7　灰色预测——GM(1,1)模型

4.7.1　灰色预测的基本原理

时间序列预测是采用趋势预测原理进行的，然而时间序列预测存在以下问题：

（1）时间序列变化趋势不明显时，很难建立起较精确的预测模型；

（2）它是在系统按原趋势发展变化的假设下进行预测的，因而未考虑对未来变化产生影响的各种不确定因素。

为克服上述缺点，邓聚龙教授引入了灰色因子的概念，采用“累加”和“累减”的方法创立了灰色预测理论。

4.7.1.1 GM(1,1)模型的基本原理

当一时间序列无明显趋势时，采用累加的方法可生成一趋势明显的时间序列。

如时间序列 $X^{(0)}=\{32,38,36,35,40,42\}$ 的趋势并不明显，但将其元素进行“累加”所生成的时间序列 $X^{(1)}=\{32,70,106,141,181,223\}$，则是一趋势明显的数列，按该数列的增长趋势可建立预测模型并考虑灰色因子的影响进行预测，然后采用“累减”的方法进行逆运算，恢复原时间序列，得到预测结果，这就是灰色预测的基本原理。

4.7.1.2 灰色预测的类型

灰色预测是基于灰色预测模型 GM(1,1)的预测，按其应用的对象可有四种类型：

(1) 数列预测。这类预测是针对系统行为特征值的发展变化所进行的预测。

(2) 突变预测。这类预测是针对系统行为的特征值超过某个阈值的异常值将在何时出现的预测。

(3) 季节突变预测。若系统行为的特征有异常值出现或某种事件的发生是在一年中的某个特定的时区，则该预测为季节性突变预测。

(4) 拓扑预测。这类预测是对一段时间内系统行为特征数据波形的预测。

4.7.1.3 GM(1,1)模型的建立方法和步骤

设：原始时间序列为 $X^{(0)}=\{x^{(0)}(1),\ x^{(0)}(2),\cdots,\ x^{(0)}(n)\}$

其累加生成序列为 $X^{(1)}=\{x^{(1)}(1),\ x^{(1)}(2),\cdots,\ x^{(1)}(n)\}$

按累加生成序列建立的微分方程模型为

$$\frac{\mathrm{d}X^{(1)}}{\mathrm{d}t}+aX^{(1)}=u$$

其解的离散描述形式为

$$X^{(1)}(t+1)=\left(X^{(0)}(1)-\frac{u}{a}\right)e^{-at}+\frac{u}{a}$$

确定了参数 a 和 u 后，按此模型递推，即可得到预测的累加数列，通过检验后，再累减即得到预测值。

其步骤如下：

(1) 由原始序列 $X^{(0)}$ 按下式计算累加生成序列 $X^{(1)}(t)$

$$X^{(1)}(i)=\sum_{m=1}^{n}X^{(0)}(m)$$

(2) 按 $X^{(1)}$，采用最小二乘法按下式确定模型参数

$$\hat{\boldsymbol{a}}=\begin{pmatrix}a\\u\end{pmatrix}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_N$$

式中

$$\boldsymbol{B}=\begin{bmatrix}-\frac{1}{2}(X^{(1)}(1)+X^{(1)}(2)) & 1\\ -\frac{1}{2}(X^{(1)}(2)+X^{(1)}(3)) & 1\\ \cdots & \cdots\\ -\frac{1}{2}(X^{(1)}(n-1)+X^{(1)}(n)) & 1\end{bmatrix};\ \boldsymbol{Y}_N=\begin{bmatrix}X^{(0)}(2)\\X^{(0)}(3)\\\cdots\\X^{(0)}(n)\end{bmatrix}$$

(3) 建立预测模型,求出累加序列

$$X^{(1)}(t+1)=(X^{(0)}(1)-\frac{u}{a})e^{-at}+\frac{u}{a}$$

(4) 采用残差分析法进行模型检验。

(5) 根据系统未来变化,确定预测值上下界,即按下式确定灰平面:

上界 $X^{(1)}_{max}(n+t)=X^{(1)}(n)+t\sigma_{max}$

下界 $X^{(1)}_{min}(n+t)=X^{(1)}(n)+t\sigma_{min}$

(6) 用模型进行预测。利用上述模型预测是利用累加生成序列 $X^{(1)}$ 的预测值,利用累减生成法将其还原,即可以得到原始序列 $X^{(0)}$ 的预测值,如满足灰因子条件则完成预测。

4.7.1.4 模型检验

GM(1,1)模型通常采用残差检验法。所谓残差检验法是指按所建模型计算出累加序列,再按累减生成法还原,还原后将其与原始序列 $X^{(0)}$ 相比较,求出两序列的差值即为残差,通过计算相对精度以确定模型精度的一种方法。

如果相对精度均满足要求精度,则模型通过检验;

如果不满足要求精度,可通过上述残差序列建立残差 GM(1,1)模型对原模型进行修正。

残差模型 GM(1,1),可提高原模型的精度,共有两种方式:

(1) 当用累加生成序列的残差建立 GM(1,1)残差模型时,其残差序列为

$$\varepsilon^{(0)}(t)=\hat{X}^{(1)}(t)-X^{(1)}(t)$$

其累加生成 GM(1,1)模型为

$$\varepsilon^{(1)}(t+1)=(\varepsilon^{(0)}(1)-\frac{u_\varepsilon}{a_\varepsilon})e^{-a_\varepsilon t}+\frac{u_\varepsilon}{a_\varepsilon}$$

其导数即为对模型 $\hat{X}(1)$ 的修正项:

$$\delta(t-i)(-a_\varepsilon)(\varepsilon^{(0)}(1)-\frac{u_\varepsilon}{a_\varepsilon})e^{-a_\varepsilon t}$$

式中 $$\delta(t-i)=\begin{cases}1 & \text{当 } t\geqslant i \text{ 时}\\ 0 & \text{当 } t<i \text{ 时}\end{cases}$$

修正后的模型为

$$\hat{X}^{(1)}(t+1)=(X^{(0)}(1)-\frac{u}{a})e^{-at}+\frac{u}{a}+\delta(t-i)(-a_\varepsilon)(\varepsilon^{(0)}(1)-\frac{u_\varepsilon}{a_\varepsilon})e^{-a_\varepsilon t}$$

或 $$\hat{X}^{(0)}(t+1)=-a(X^{0}(1)-\frac{u}{a})e^{-at}+\delta(t-i)(-a_\varepsilon)^2(\varepsilon^{(0)}(t)-\frac{u_\varepsilon}{a_\varepsilon})e^{-a_\varepsilon t}$$

(2) 当用还原模型的残差序列建立 GM(1,1)模型时,残差序列为

$$q^{(1)}(t)=\hat{X}^{(0)}(1)-X^{(0)}(t)$$

其累加生成模型为

$$q^{(1)}(t+1)=(q^{(0)}(1)-\frac{u_q}{a_q})e^{-a_q t}+\frac{u_q}{a_q}$$

对模型的修正项求其导数形式

$$\delta(t-i)(-a_q)(q^{(0)}(1)-\frac{u_q}{a_q})e^{-a_q t}$$

式中
$$\delta(t-i)=\begin{cases}1 & \text{当 } t\geqslant I \text{ 时}\\ 0 & \text{当 } t<I \text{ 时}\end{cases}$$

修正后的模型为 $\hat{X}^{(1)}(t+1)$的导数和 $q^{(1)}(t+1)$的导数之和,即

$$\hat{X}^{(0)}(t+1)=-a(X^{(0)}(1)-\frac{u}{a})e^{-at}+\delta(t-i)(-a_q)(q^{(0)}(1)-\frac{u_q}{a_q})e^{-a_q t}$$

或
$$\hat{X}^{(1)}(t+1)=(X^{(0)}(1)-\frac{u}{a})e^{-at}+\frac{u}{a}+(q^{(0)}(1)-\frac{u_q}{a_q})e^{-a_q t}+\frac{u_q}{a_q}$$

综上所述,GM(1,1)模型实质上是采用线性化方法建立的一种指数预测模型。因此,当系统呈指数变化时,预测精度较高。

4.7.2 GM(1,1)预测模型的应用

4.7.2.1 数列预测模型

某公司销售额数据见表 4.7.1。

表 4.7.1 某公司销售额

年份	1995	1996	1997	1998	1999	2000
销售额	434.5	470.5	527.5	571.4	626.4	685.2

现建立 GM(1,1)预测模型并预测 2001 年、2002 年销售额。

初始时间序列:

$$X^{(0)}=\{434.5,470.5,527.6,571.4,626.4,685.2\}$$

第一步 求累加生成数列

$$X^{(1)}=\{434.5,905,1432.6,2004,2630.4,3315.6\}$$

第二步 用最小二乘法求参数 $\hat{a}=(a,u)^{\mathrm{T}}$

$$\boldsymbol{B}=\begin{bmatrix}-\frac{1}{2}(X^{(1)}(1)+X^{(1)}(2)) & 1\\ -\frac{1}{2}(X^{(1)}(2)+X^{(1)}(3)) & 1\\ -\frac{1}{2}(X^{(1)}(3)+X^{(1)}(4)) & 1\\ -\frac{1}{2}(X^{(1)}(4)+X^{(1)}(5)) & 1\\ -\frac{1}{2}(X^{(1)}(5)+X^{(1)}(6)) & 1\end{bmatrix}=\begin{bmatrix}-2973.0 & 1\\ -669.75 & 1\\ -1168.8 & 1\\ -1718.3 & 1\\ -2317.2 & 1\end{bmatrix}$$

$$Y_N=(470.5,527.6,571.4,626.4,685.2)^{\mathrm{T}}$$

代入 $\hat{\boldsymbol{a}}=(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Y}_N$ 得

$$\hat{\boldsymbol{a}}=\begin{pmatrix}-0.0916\\414.0736\end{pmatrix}$$

因 $X^{(1)}(1)=434.5$ 得

$$X^{(1)}(t+1)=\left(X^{(1)}(1)-\frac{u}{a}\right)\mathrm{e}^{-at}+\frac{u}{a}=$$

$$4953.04815\mathrm{e}^{0.0916t}-4518.54815$$

第三步 检验。

检验结果见表 4.7.2,由表可见精度较高,模型可用。

表 4.7.2 检验数表

年份	按模型计算数据 $X^{(1)}$	还原数据 $X^{(0)}$	原始数据 $X^{(0)}$	绝对误差	相对误差/%
1995	434.5	434.5	434.5	0	0
1996	909.8	475.3	470.5	-4.8	1.0
1997	1430.8	521.0	527.6	6.6	1.25
1998	2001.7	570.9	571.4	0.5	0.08
1999	2627.5	625.8	626.4	0.6	0.095
2000	3313.3	685.8	685.2	-0.6	0.087

第四步 建立灰平面。

假设该公司生产受生产能力的限制,每年销售额的增长量不超过 70 万元,但不低于 20 万元,该公司最高生产能力的销售额为 800 万元,最低为 600 万元,故灰平面为

上界:$X^{(1)}_{\max}(t+k)=X^{(1)}(t)+k\delta_{\max}$

当 $t=6$ 时,$X^{(1)}_{\max}(6+k)=X^{(1)}(6)+k\delta_{\max}$

$$X^{(1)}_{\max}(6+k)=3315.6+70K$$

下界:$X^{(1)}_{\min}(t+k)=X^{(1)}(t)+k\delta_{\min}=3315.6+20K$

第五步 预测 2001 年、2002 年销售额。

2001 年: $\hat{X}^{(1)}(6+1)=4953.04815\mathrm{e}^{0.0916\times6}-4518.54815=4062.9$

$$\hat{X}^{(0)}(7)=4062.9-3313.3=749.6$$

2002 年: $\hat{X}^{(1)}(7+1)=4953.04815\mathrm{e}^{0.0916\times7}-4518.54815=4886.1$

$$\hat{X}^{(0)}(8)=4886.1-(3313.3+749.6)=823.2$$

由预测值可见,$X^{(0)}(8)=823.2$,比 $X^{(0)}(7)$高 73.6 万元,超过增长限度,故应取增长值的最高限 749.6+70=819.6 万元,但该值超过该公司最大生产能力,故最终预测值为 800 万元。

4.7.2.2 灾变预测

某企业生产用原料属受自然灾害影响较大的农产品。一般来说,自然灾害的发生有其偶然性,但对历史数据的整理,仍可发现一定的规律性。为确保生产不受自然灾害的影

响,该企业希望了解影响原料供应的规律性并提前作好原料储备,所收集数据见表 4.7.3 并规定每亩平均收获量小于 320kg 时为欠收年份,将影响原料的正常供应,现应用灰色灾变预测来预测下次发生欠收的年份。

表 4.7.3 历年原料收获统计表

年份	1984	1985	1986	1987	1988	1989	1990	1991	1992
收获量/kg	390.6	412	320	559	380	542	553	310	561
年份	1993	1994	1995	1996	1997	1998	1999	2000	
收获量/kg	300	632	540	406.2	314	576	587	318	

第一步 将上表中年份用序号替换,并找出收获量小于 320kg 的年份序号形成初始序列 $\omega^{(0)}$。

本例初始序列 $\omega^{(0)}=\{3,8,10,14,17\}$

累加生成序列 $\omega^{(1)}=\{3,11,21,35,52\}$

第二步 按 $\omega^{(1)}$建 GM(1,1)模型。

$$\boldsymbol{B}=\begin{bmatrix} -\frac{1}{2}(\omega^{(1)}(1)+\omega^{(1)}(2)) & 1 \\ -\frac{1}{2}(\omega^{(1)}(2)+\omega^{(1)}(3)) & 1 \\ -\frac{1}{2}(\omega^{(1)}(3)+\omega^{(1)}(4)) & 1 \\ -\frac{1}{2}(\omega^{(1)}(4)+\omega^{(1)}(5)) & 1 \end{bmatrix}=\begin{bmatrix} -7 & 1 \\ 16 & 1 \\ -28 & 1 \\ -43.5 & 1 \end{bmatrix}$$

$$Y_N=(8,10,14,17)^{\mathrm{T}}$$

$$\hat{\boldsymbol{a}}=(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_N=(-0.25361 \quad 6.258339)^{\mathrm{T}}$$

模型为 $\omega^{(1)}(t+1)=(\omega^{(0)}(1)-\frac{u}{a})\mathrm{e}^{-at}+\frac{u}{a}=27.67702\mathrm{e}^{0.25361t}-24.67702$

当 $t=5$ 时,$\hat{\omega}^{(1)}(6)=73.6848$

$$\hat{\omega}^{(0)}(6)=73.6848-52=21.6848$$

下次发生收获量小于 320kg 的年份为:2004 年—2005 年,因按年份序号预测应为 21.6848,即 21 号或 22 号,现最后序号 17 对应 2000 年,故 21 − 17 = 4,22 − 17 = 5,即四五年后将出现收获量小于 320kg 的可能。

第5章　投入产出综合平衡模型

投入产出综合平衡模型，又叫投入产出模型，它是投入产出法的核心。

投入产出法是一种经济数学方法，它必须以正确的经济理论为指导。我国采用的投入产出法与资本主义国家所采用的投入产出法不同，它是在列昂节夫提出的投入产出法的基础上经改造后形成的，是在马克思主义经济理论指导下，从我国客观实际出发，通过编制投入产出表和建立相应的经济数学模型，利用现代化数学方法和电子计算机，综合考察和分析我国国民经济各部门，各地区、各时期的数量依存关系和综合比例的一种方法。

静态投入产出法的核心是静态投入产出综合平衡模型，这是系统静态模型中较有代表性的、反映非工程系统特性的平衡模型，它对加深理解系统静态模型的建立方法和应用系统思想全面综合地考察问题，尤其是在定量地处理复杂的经济现象方面有很重要作用，因此本章将就投入产出法的基本原理、投入产出模型的建立和求解以及投入产出模型的应用等做一系统的介绍。

5.1　部门间投入产出综合平衡模型

部门间投入产出综合平衡模型是利用现代数学方法和电子计算机，从数量方面考察国民经济各部门之间生产与分配的关系，反映和分析国民经济的综合比例以及部门比例的一种经济数学模型。它是对部门间所存在的直接和间接联系的数学模拟，是经济平衡和计划管理的重要工具。

5.1.1　投入、产出和部门之间的联系

“投入”是指社会产品在生产过程中所消耗的劳动力、劳动手段和劳动对象等；“产出”是指社会产品生产后分配的去向和数量，它包括生产消耗、消费、出口、增加储备、基本建设等。

国民经济系统中“投入”与“产出”的关系如图5.1.1所示。

按社会主义再生产理论，国民经济各部门的投入按生产要素划分，包括劳动对象、劳动手段和劳动力。劳动手段是劳动得以发挥作用的装备物，它一般可服务于多个生产周期，在实物形态上逐渐磨损，一次更新，在价值上一次垫付，随其磨损程度逐渐以固定资产折旧的形式转移到产品中去。劳动对象是劳动作用其上的被加工物，在一个生产周期期内，在实物上被消耗，价值全部转移到产品中去。二者之和构成了产品的物化劳动，即生产资料转移部分。劳动力即劳动能力，是指活的人体中存在的、每当人生产某种使用价值时就运用的体力和智力的总和。劳动力和生产资料的结合，组成生产过程，它形成价值并使价值增值。这种价值的增值就是劳动者新创造的价值。它是由必要劳动 V(劳动报酬)和社会劳动 M(以利

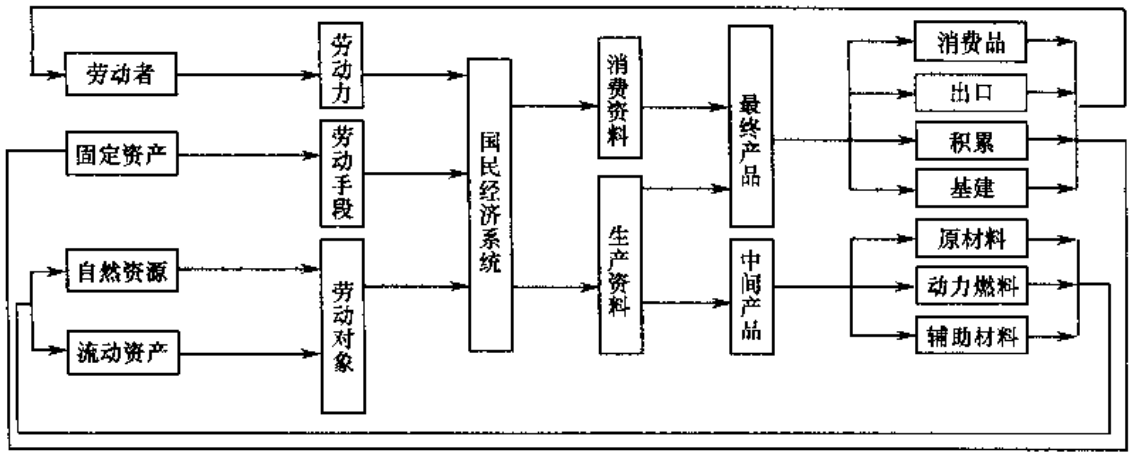


图 5.1.1 “投入”与“产出”的关系

润和税金的形式体现)两部分组成的,它构成了产品的活劳动。产出包括生产资料和消费资料两部分,进一步又可按用途分为最终产品和中间产品。最终产品是指生产领域已最终加工完毕可供社会消费和使用的产品,即各部门的总产品中扣除生产性消耗后不参加本时期周转的那部分产品,是本时期生产领域的最终成果,是活劳动所创造价值的总和。它包括消费品,增加库存和净出口等。中间产品是指在生产领域中尚需进一步加工的产品,是为生产最终产品服务的,当生产出最终产品后,中间产品的实物形态消失并转化为最终产品,它是社会产品物化劳动的总和,包括原材料、燃料动力、辅助材料等。

国民经济各部门在国民经济中既是生产者,又是消费者。作为生产者,它“产出”产品,供其他部门消费,把产品按需要分配出去;作为消费者,对它“投入”其他部门的产品供其消费,通过消费过程把产品生产出来。因此,“投入”和“产出”是相互依赖的,没有投入就不能产出,没有产出,国民经济的生产也不能进行。国民经济各部门的这种投入和产出的相互依赖关系,就构成了国民经济各部门之间在产品生产和分配上的相当复杂的技术经济联系。从投入角度研究,这种联系不仅体现在国民经济各部门对其他部门产品的直接消耗上,而更重要的是体现在对其他部门的产品的间接消耗上,直接消耗和间接消耗的总和才能真正地反映国民经济各部门之间的技术经济联系。

假设国民经济系统由五个部门组成,一般情况下,其结构模型如图 5.1.2 所示。系统

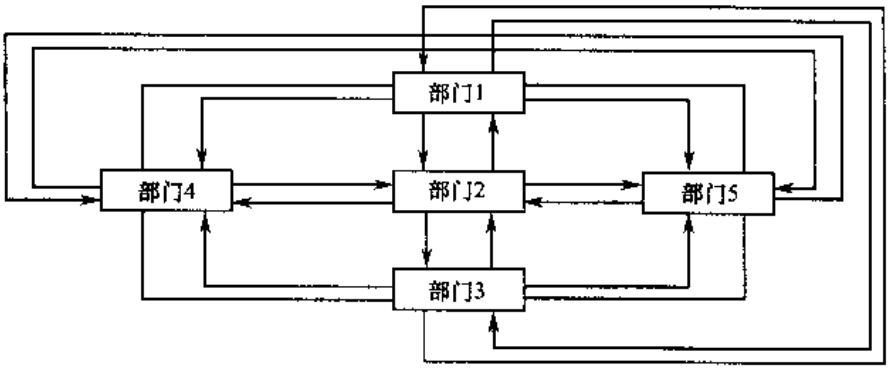


图 5.1.2 部门之间的关系

各部门之间直接联系的经济涵义是部门 i 在生产中直接消耗 j 部门的产品，这种消耗为直接消耗。它是比较容易发现和掌握的，也是容易定量描述的。由结构模型解析法知，系统各单元之间除直接联系之外，还有 i 部门消耗 j 部门产品，j 部门又消耗 k 部门产品，从而引起 i 部门通过 j 部门间接地消耗 k 部门产品等，依此类推就构成了部门之间的复杂联系。这种部门 i 通过部门 j 对其他各部门产品的消耗为间接消耗，这种消耗是研究国民经济系统时不易掌握、不易定量描述的。部门间投入产出综合平衡模型就是以这种直接消耗和间接消耗所反映的技术经济联系为依据，揭示并确定国民经济系统发展的内在规律。

5.1.2 价值型投入产出表

根据国民经济各部门既是生产者，又是消费者这一特点，将各部门生产需要的各种投入和所生产产品的分配去向有规律地排列在一张纵横交错的棋盘式表格中，就形成了一张反映国民经济各部门之间投入、产出关系的表格，该表格叫投入产出表。如果该表是以实物单位计量的，叫实物型投入产出表；如以货币单位计量，则叫价值型投入产出表。因价值型的投入产出表比实物型投入产出表更能反映国民经济各部门间的联系，故本节主要介绍价值型投入产出表。

5.1.2.1 价值型投入产出表的结构

价值型投入产出表的结构见表 5.1.1。

表 5.1.1 价值型投入产出表 （单位：亿元）

		中间产品										最终产品												社会总产品
		农业	冶金工业	机械工业	纺织、缝纫皮革工业	其他工业	建筑业	交通运输、邮电业	商业	合计	固定资产更新改造大修	消费			积累						进出口		合计	
												个人消费	社会消费	合计	固定资产			流动资产		合计	进口(-)	出口(+)	进出口差额	
															生产性	非生产性	小计	生产性	非生产性	小计				
劳动对象消耗	农业	第Ⅰ部分									第Ⅱ部分													
	冶金工业																							
	纺织、缝纫、皮革工业																							
	其他工业																							
	建筑业																							
	交通运输、邮电业																							
	商业																							
	合计																							

（续）

		中间产品										最终产品												社会总产品
		农业	冶金工业	机械工业	纺织、缝纫皮革工业	其他工业	建筑业	交通运输、邮电业	商业	合计	固定资产更新改造大修	消费			积累							进出口		合计	
												个人消费	社会消费	合计	固定资产			流动资产			合计	进口(-)	出口(+)	进出口差额	
															生产性	非生产性	小计	生产性	非生产性	小计					
	固定资产折旧																								
新创造价值	劳动报酬	第Ⅲ部分										第Ⅳ部分													
	社会纯收入 利润税金																								
	社会纯收入 其他																								
	社会纯收入 小计																								
	合计																								
社会总产品																									

价值型投入产出表中的部门是按“产品类别”划分的产品部门，即同一产品部门生产类型相近的产品。本章所指的国民经济部门均指产品部门。

价值型投入产出表中，主栏和宾栏的产品部门按相同顺序排列。主栏包括生产资料转移价值和活劳动新创造价值两大部分，它反映了社会产品的价值构成；宾栏包括中间产品和最终产品两大部分，它反映了社会产品的分配和使用去向。主栏和宾栏的两大部分相互交叉构成了相互联系的四个部分。

第Ⅰ部分反映了国民经济各部门在产品生产和消耗上的技术联系，揭示出部门之间的投入产出关系；第Ⅱ部分反映了各部门产品最终使用的比例和构成，体现出各部门的最终成果；第Ⅲ部分反映了各部门提取的折旧价值和各部门创造的国民收入价值及构成；第Ⅳ部分反映国民收入的再分配的某些因素，由于该部分较复杂，编表时一般省略。

5.1.2.2 价值型投入产出表的平衡关系

设：X_{ij}——第 j 部门在生产中一年所消耗的 i 部门产品数量；

X_i——第 i 部门一年中所生产的社会总产品量；

Y_i——第 i 部门一年中所提供的最终产品总量；

D_j——第 j 部门一年所提取的折旧额；

V_j——第 j 部门一年内所支付的劳动报酬额；

M_j——第 j 部门劳动者一年内为社会劳动所创造的价值，即社会纯收入或利、税额；

N_j——第 j 部门一年内所创造的国民收入，$N_j = M_j + V_j$。

按投入产出表四个部分的经济涵义，可得如下四种平衡关系：

(1) 根据投入产出表中行所体现的某部门产品一部分作为其他部门的中间产品，另一部分作为为社会提供的最终产品这一经济关系得出反映各部门产品分配去向和数量的平衡关系：

中间产品 + 最终产品 = 总产品

用公式描述为

$$\sum_j X_{ij} + Y_i = X_i \qquad (i = 1,2,\cdots,n) \tag{5.1.1}$$

(2) 根据投入产出表中列所体现某部门产品价值是由生产资料转移价值(物化劳动)和新创造价值(活劳动)所构成的这一经济关系,得出反映产品价值形成过程的平衡关系:

生产资料转移价值 + 新创造价值 = 总产值

用公式描述为

$$\sum_i X_{ij} + D_j + N_j = X_j \qquad (j = 1,2,\cdots,n) \tag{5.1.2}$$

(3) 根据同一物质生产部门的总产出量和总投入量相等的经济关系,得出部门产品分配量之和与该部门生产过程中活劳动和物化劳动消耗量之和相等的平衡关系,用公式描述为

$$\sum_i X_{ij_0} + D_{j_0} + N_{j_0} = \sum_j X_{i_0 j} + Y_{i_0} \qquad (i_0 = j_0) \tag{5.1.3}$$

(4) 根据各部门为社会所提供的最终产品总和是各部门新创造价值与所提折旧的总和,即第Ⅱ部分与第Ⅲ部分总和相等这一经济关系,得出国民收入与折旧额之和(国民生产总值)与最终产品之和相等的平衡关系,用公式可描述为

$$\sum_j D_j + \sum_j N_j = \sum_i Y_i \tag{5.1.4}$$

这四种平衡关系是国民经济各部门之间的基本平衡关系,是建立部门间投入产出综合平衡模型的基础。

5.1.3 部门间投入产出综合平衡模型

5.1.3.1 投入产出综合平衡模型的技术经济参数

部门间投入产出综合平衡模型是在投入产出表的基础上,计算出各种技术经济参数后建立的模型,由投入产出表可以计算如下技术经济参数:

1. 直接消耗系数

直接消耗系数的经济涵义是 j 部门生产单位产品所消耗的 i 部门产品数量,以 a_{ij} 表示,其计算公式为

$$a_{ij} = \frac{X_{ij}}{X_j} \qquad (i,j = 1,2,\cdots,n) \tag{5.1.5}$$

用矩阵表示则为

$$\boldsymbol{A} = \boldsymbol{Q}\hat{\boldsymbol{X}}^{-1} \tag{5.1.6}$$

式中

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix};\boldsymbol{Q} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{n1} & X_{n2} & \cdots & X_{nn} \end{bmatrix};\hat{\boldsymbol{X}}^{-1} = \begin{bmatrix} \frac{1}{X_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{X_2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \frac{1}{X_n} \end{bmatrix}$$

A 矩阵为直接消耗系数矩阵，用以说明国民经济各部门的技术联系。

2. 直接折旧系数

直接折旧系数的经济涵义是某部门生产单位产品所提取的折旧，以 a_{Dj} 表示，其计算公式为

$$a_{Dj}=\frac{D_j}{X_j} \qquad (j=1,2,\cdots,n) \tag{5.1.7}$$

3. 劳动报酬系数

劳动报酬系数的经济涵义是 j 部门生产单位产品必要劳动创造的价值或 j 部门生产单位产品需付出的劳动报酬量，以 a_{vj} 表示，计算公式为

$$a_{vj}=\frac{V_j}{X_j} \qquad (j=1,2,\cdots,n) \tag{5.1.8}$$

4. 社会纯收入系数

社会纯收入系数的经济涵义是 j 部门生产单位产品为社会创造的纯收入数量，以 a_{mj} 表示，计算公式为

$$a_{mj}=\frac{M_j}{X_j} \qquad (j=1,2,\cdots,n) \tag{5.1.9}$$

5. 国民收入系数

国民收入系数又叫净产值系数，它的经济涵义为 j 部门生产单位产品所创造的国民收入量，以 a_{Nj} 表示，计算公式为

$$a_{Nj}=\frac{N_j}{X_j} \qquad (j=1,2,\cdots,n) \tag{5.1.10}$$

6. 劳动消耗系数

设 L_j 为第 j 部门年均劳动力人数，则劳动消耗系数的经济涵义为 j 部门生产单位产品所需的劳动力数量，以 a_{Lj} 表示，计算公式为

$$a_{Lj}=\frac{L_j}{X_j} \qquad (j=1,2,\cdots,n) \tag{5.1.11}$$

a_{Dj}、a_{Vj}、a_{Mj} 和 a_{Nj} 四个系数用矩阵形式可描述为

$$\boldsymbol{A}_d=\boldsymbol{Q}_d\hat{\boldsymbol{X}}^{-1} \tag{5.1.12}$$

式中

$$\boldsymbol{A}_d=\begin{bmatrix} a_{D1} & a_{D2} & \cdots & a_{Dn} \\ a_{V1} & a_{V2} & \cdots & a_{Vn} \\ a_{M1} & a_{M2} & \cdots & a_{Mn} \\ a_{N1} & a_{N2} & \cdots & a_{Nn} \end{bmatrix} \qquad \boldsymbol{Q}_d=\begin{bmatrix} D_1 & D_2 & \cdots & D_n \\ V_1 & V_2 & \cdots & V_n \\ M_1 & M_2 & \cdots & M_n \\ N_1 & N_2 & \cdots & N_n \end{bmatrix}$$

5.3.1.2 投入产出综合平衡模型

由各系数的经济涵义和投入产出表可知：

$$\sum_i a_{ij}+a_{Dj}+a_{Nj}=1 \qquad (j=1,2,\cdots,n)$$

部门间投入产出综合平衡模型的建立过程如下：

根据式(5.1.5)得

$$X_{ij} = a_{ij}X_j$$

将其代入式(5.1.1)即得到按行建立的投入产出模型：

$$\sum_i a_{ij}X_j + Y_i = X_i \qquad (i = 1,2,\cdots,n)$$

移项得

$$X_i - \sum_i a_{ij}X_j = Y_i \qquad (i = 1,2,\cdots,n) \tag{5.1.13}$$

写成矩阵形式：

$$\boldsymbol{X} - \boldsymbol{AX} = \boldsymbol{Y}$$

$$(\boldsymbol{I} - \boldsymbol{A})X = \boldsymbol{Y} \tag{5.1.14}$$

式中

$$(\boldsymbol{I} - \boldsymbol{A}) = \begin{bmatrix} 1 - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & 1 - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & 1 - a_{nn} \end{bmatrix}; X = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{bmatrix}; \boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{bmatrix}$$

$(\boldsymbol{I} - \boldsymbol{A})$矩阵称为列昂节夫矩阵。在$(\boldsymbol{I} - \boldsymbol{A})$中，纵列说明每种产品投入与产出的关系；负号表示投入，矩阵每一列都说明为生产单位产品所要投入的各种产品数量；正号表示产出，主对角线上的元素均为正元素，它表示各部门扣除自身消耗后的净产出。$(\boldsymbol{I} - \boldsymbol{A})$矩阵把 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的关系描述出来，揭示出社会总产品和最终产品的相互关系。

将 $X_{ij} = a_{ij}X_j$ 和 $D_j = a_{Dj}X_j$ 代入式(5.1.2)即得到按列建立的投入产出模型：

$$\sum_i a_{ij}X_j + a_{Dj}X_j + N_j = X_j \qquad (j = 1,2,\cdots,n)$$

整理后得：

$$[I - (\sum_i a_{ij} + a_{Dj})]X_j = N_j \qquad (j = 1,2,\cdots,n) \tag{5.1.15}$$

写成矩阵形式为

$$[\boldsymbol{I} - \hat{\boldsymbol{C}}]\boldsymbol{X} = \boldsymbol{N} \tag{5.1.16}$$

式中

$$[\boldsymbol{I} - \hat{\boldsymbol{C}}] = \begin{bmatrix} 1 - (\sum_i a_{i1} + a_{D1}) & 0 & \cdots & 0 \\ 0 & 1 - (\sum_i a_{i2} + a_{D2}) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 - (\sum_i a_{in} + a_{Dn}) \end{bmatrix}$$

$$\boldsymbol{N} = (N_1 \quad N_2 \quad \cdots \quad N_n)^{\mathrm{T}}$$

$\hat{\boldsymbol{C}}$ 矩阵对角线上的各元素表示生产单位产品的生产资料转移价值(物化劳动量)，故

也称为物耗系数。$(\boldsymbol{I}-\hat{\boldsymbol{C}})$矩阵则表示生产单位产品新创造的价值量，它揭示出社会总产值与国民收入之间的函数关系。

式(5.1.13)和式(5.1.15)或式(5.1.14)和式(5.1.16)就是投入产出模型的基本形式。

5.1.4 完全消耗系数

完全消耗系数的经济涵义是第 j 部门生产单位最终产品对各部门产品的直接消耗和间接消耗量的总和，以 b_{ij} 表示，其计算公式为

$$b_{ij}=a_{ij}+\sum_{k} b_{ik}\cdot a_{kj} \tag{5.1.17}$$

式中　k——中间产品部门。

$\sum b_{ik}\cdot a_{kj}$——通过 k 种中间产品而形成的第 j 种产品对第 i 种产品的间接消耗。

如单位木材产品对钢材的完全消耗系数 $b_{钢,木}$ 为单位木材产品对钢材的直接消耗量 $a_{k,木}$ 与木材生产对钢材的间接消耗量 $\sum b_{钢,k}\cdot a_{k,木}$ 之和，而木材生产对钢材的间接消耗则体现在木材生产过程中所消耗的各种产品对钢材的需求上。如生产单位木材产品对煤的需求为 $a_{煤,木}$，煤的生产又需钢，且生产单位煤产品对钢的完全消耗为 $b_{钢,煤}$，则木材通过煤对钢的需求(间接消耗)为 $b_{钢,煤}\cdot a_{煤,木}$，把木材生产中所直接消耗的各种产品(以 k 表示)对钢材的间接消耗加总即得到生产单位木材产品对钢材的全部间接消耗，即 $\sum b_{钢,k}\cdot a_{k,木}$。所以单位木材产品对钢材的完全消耗量为

$$b_{钢,木}=a_{钢,木}+\sum b_{钢,k}\cdot a_{k,木}$$

将式(5.1.17)写成矩阵形式为

$$\boldsymbol{B}=\boldsymbol{A}+\boldsymbol{BA}$$

移项并整理得：　$\boldsymbol{B}=\boldsymbol{A}(\boldsymbol{I}-\boldsymbol{A})^{-1}=[\boldsymbol{I}-(\boldsymbol{I}-\boldsymbol{A})](\boldsymbol{I}-\boldsymbol{A})^{-1}$

按矩阵乘法展开得：　$\boldsymbol{B}=(\boldsymbol{I}-\boldsymbol{A})^{-1}-\boldsymbol{I}$ (5.1.18)

式中　$(\boldsymbol{I}-\boldsymbol{A})^{-1}$称为列昂节夫逆阵。

由式(5.1.18)可见，完全消耗系数矩阵与列昂节夫逆阵的关系相当密切，它们在数值上仅仅主对角线上的元素差“1”，其他皆相同，因此人们称列昂节夫逆阵为完全需要系数。

列昂节夫逆阵的经济涵义是生产单位最终产品对社会总产品的完全消耗，系数中包含了该单位最终产品，而完全消耗系数 $\boldsymbol{B}$ 不包括单位最终产品，仅反映生产单位最终产品对各种中间产品的完全消耗。无论是$(\boldsymbol{I}-\boldsymbol{A})^{-1}$，还是 $\boldsymbol{B}$，都反映完全消耗这一概念，只有在明确其经济涵义的情况下才能合理地应用，才不至混淆。

完全消耗系数均为正值，在数值上都比直接消耗系数大，有些产品的完全消耗系数要比直接消耗系数大得多，因此应用完全消耗系数进行经济预测和经济决策比用直接消耗系数准确得多。因此，国民经济如不计算产品之间存在的大量间接消耗关系，也就谈不上真正的综合平衡。表 5.1.2 列出苏联 1966 年计算的部分工业产品对电力的直接消耗系数和完全消耗系数。由表可见，完全消耗系数与直接消耗系数之比最高竟达 7.5 倍，如在国民经济的计划中忽略了间接消耗肯定会给国民经济的发展带来不良的后果。

表 5.1.2　部分工业品对电力的消耗系数

产　品	直接消耗系数	间接消耗系数	完全消耗系数	完全消耗系数与直接消耗系数之比
原　铝	19677.2	2967.2	21644.4	1.2
铁合金	4480.4	928.1	5408.5	1.2
精煤和煤砖	22.4	10.1	32.5	1.4
精钢	758.5	4958.7	5717.2	7.5
小汽车	1438.4	1942.0	3380.4	2.4
电力机车	260.7	391.3	652.0	2.5
货车车箱	4.3	14.0	18.3	4.3
金属切削机床	3.4	6.7	10.1	3.0
挖土机	15.8	18.5	34.3	2.2
织布机	1.9	1 7	3.6	1.4
压缩机	2.5	4.8	7.3	2.9
钢	67.1	156.7	223.8	3.3

5.1.5　模型的求解

部门间投入产出模型的基本形式是：

$$(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}=\boldsymbol{Y}$$

$$(\boldsymbol{I}-\hat{\boldsymbol{C}})\boldsymbol{X}=\boldsymbol{N}$$

这两个模型都是 n 个方程、$2n$ 个变量的线性方程组，前者揭示了总产品和最终终产品的关系，后者揭示了总产值和国民收入之间的关系。当已知各部门总产品 $\boldsymbol{X}$ 时，即可求出最终产品和国民收入值。

如果当最终产品 $\boldsymbol{Y}$ 已知时，方程组 $(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}=\boldsymbol{Y}$ 的解为

$$\boldsymbol{X}=(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y} \tag{5.1.19}$$

从而可确定社会总产品的数量。

如果国民收入 $\boldsymbol{N}$ 已知时，方程组 $(\boldsymbol{I}-\hat{\boldsymbol{C}})\boldsymbol{X}=\boldsymbol{N}$ 的解为

$$\boldsymbol{X}=(\boldsymbol{I}-\hat{\boldsymbol{C}})^{-1}\boldsymbol{N} \tag{5.1.20}$$

或可直接写成如下形式

$$X_j=\frac{N_j}{1-(\sum_i a_{ij}+a_{Dj})}=\frac{N_j}{a_{Nj}} \qquad (j=1,2,\cdots,n) \tag{5.1.21}$$

从而确定了社会总产品量。

5.1.6　投入产出表及投入产出模型实例

现举例说明投入产出表及投入产出模型。

设国民经济系统是由农业、工业、建筑业、运输邮电和商业饮食业五个部门组成，其投入产出表见表 5.1.3。

表 5.1.3　部门投入产出表(价 值 型)　　(单位:亿元)

产出 / 投入		中间产品						最终产品							社会总产值
		农业	工业	建筑	运输邮电	商业饮食	合计	固定资产更新改造	消费	固定资产积累		流动资金积累		合计	
										生产性	非生产性	生产性	非生产性		
劳动对象消耗	农业	320	150	50		200	720	5	1000	50	150	25	50	1280	2000
	工业	200	2500	300	150	100	3250	150	50	400	50	100	50	800	4050
	建筑	20	100	50	20	20	210	200	150	300	30	20	–	700	910
	运输邮电	15	250	25	10	20	320	100	50	200	10	100	15	475	795
	商业饮食	15	100	50	30		195	20	250	30	10	–	5	315	510
	合计	570	3100	475	210	340	4695	475	1500	980	250	245	120	3570	8265
固定资产折旧		60	300	20	50	20	450								
新创造价值	劳动报酬 V	1000	200	50	35	50	1335								
	社会劳动 M	370	450	365	500	100	1785								
	国民收入 N	1370	540	415	535	150	3120								
社会总产品		2000	4050	910	795	510	8265								

由表 5.1.3 计算的直接消耗系数见表 5.1.4。

表 5.1.4　直接消耗系数表

	农业	工业	建筑	运输邮电	商业饮食
农业	0.1600	0.0370	0.0549	0	0.3922
工业	0.1000	0.6173	0.3297	0.1887	0.1961
建筑	0.0100	0.0247	0.0549	0.0252	0.0392
运输邮电	0.0075	0.0617	0.0275	0.0126	0.0392
商业饮食	0.0075	0.0247	0.0549	0.0377	0

由表 5.1.3 计算的其他各种参数见表 5.1.5。

表 5.1.5　其他各种参数表

	农业	工业	建筑	运输邮电	商业饮食
a_{Dj}	0.300	0.0741	0.0220	0.0629	0.0392
a_{Vj}	0.5000	0.0494	0.0549	0.0440	0.0980
a_{Mj}	0.1850	0.1111	0 4011	0.6289	0.1961
a_{Nj}	0.6850	0.1605	0.4560	0.6729	0.2941

由表 5.1.3 计算的完全消耗系数如表 5.1.6。

表 5.1.6　完全消耗系数表

	农业	工业	建筑	运输	商业饮食
农业	0.2181	0.1717	0.1630	0.0572	0.5286
工业	0.3665	1.8668	1.0840	0.6051	0.7747
建筑	0.0243	0.0861	0.0951	0.0471	0.0714
运输	0.0337	0.1862	0.1031	0.0546	0.0953
商业饮食	0.0208	0.0838	0.0920	0.0577	0.0306

投入产出模型的基本形式如下：

(1) 按行建立的模型。

$$\begin{bmatrix} 0.8400 & -0.0370 & -0.0549 & 0 & 0.3922 \\ -0.1000 & 0.3827 & -0.3297 & -0.1887 & -0.1961 \\ -0.0100 & -0.0250 & 0.9451 & -0.0252 & -0.0392 \\ -0.0075 & -0.0617 & -0.0275 & 0.9874 & -0.0392 \\ -0.0075 & -0.0247 & -0.0549 & -0.0377 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix}$$

(2) 按列建立的模型。

$$\begin{bmatrix} 0.6850 & 0 & 0 & 0 & 0 \\ 0 & 0.1605 & 0 & 0 & 0 \\ 0 & 0 & 0.4560 & 0 & 0 \\ 0 & 0 & 0 & 0.6729 & 0 \\ 0 & 0 & 0 & 0 & 0.2941 \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix}$$

5.2　部门间投入产出综合平衡模型的应用

5.2.1　投入产出模型的基本假设与求解条件

投入产出模型是应用经济数学方法对国民经济系统中各子系统间的相互联系进行抽象而得出的反映国民经济系统内在规律性的模型。第 3 章中曾指出，任何系统的模型化，都必须有一定的假设条件，而在一定假设条件下得到的模型不适合于假设条件之外的情况，这对投入产出模型也同样适用。在建立投入产出模型时也有一定的假设条件，由于这些假设条件造成了模型在使用中的局限性，因此必须清楚地了解这些假设条件才能更好地运用投入产出模型。

5.2.1.1　投入产出模型的基本假设

1. 国民经济各部门投入与产出是惟一线性关系

由投入产出模型知，部门投入产出之间呈线性关系，且用一组线性方程组表示。产出

量随投入量的增加而成比例地增长是该模型的基本特征。如果系统中某部门投入量与产出量不成比例，则不符合线性叠加原理，也就不能直接使用投入产出模型。

2、同一物质生产部门，投入结构必须相同

该假设是指国民经济各部门必须以相同的投入结构生产同一类产品。

投入产出表是以"产品"进行部门分类的，产品部门要求按产品的经济用途和产品的生产消耗构成及工艺技术条件上的同类性来划分，据此，把有相同消耗构成的同类产品汇总成一个部门就形成了产品部门。但同一产品在不同地区，不同企业，因工艺条件、消耗水平、原材料需求等的差别而导致投入结构不同。把这样一类产品，作为同一"产品部门"用同一投入结构表示其生产过程，以产品的社会平均投入水平和投入结构代替个别地区、企业的投入水平和结构，会使问题更具有普遍性，且容易处理。这就是投入产出法在处理部门关系时的根本出发点。

如果部门内各产品投入水平和结构差别相当大，应用投入产出模型会产生很大的误差，因此在应用中必须注意该假设，使产品部门的划分合乎假设条件。

3. 产品分类的有限性

社会总产品的种类是相当多的，很难在投入产出表中一一列出，即使能列出，也会造成投入产出表编制和求解上的困难，这对经济平衡和计划管理是不必要的。因此以有限的产品部门进行社会产品分类，或舍掉或合并，而形成投入产出表的部门数。

根据有限的产品部门所计算的完全消耗系数自然会因产品部门划分的有限性而受影响，因此当部门数较少时，完全消耗系数的误差相对大，反之相对小。

认真理解上述三条假设是更好地应用投入产出模型的前提。

5.2.1.2 模型求解的基本条件

投入产出模型的求解要求满足一定的条件。现将在正常经济活动中由经济过程的客观实际决定的条件叫经济条件；把模型求解时必须满足的数学处理上的要求称之为数学条件。

1. 模型求解的经济条件

(1) 各部门社会总产品量必须大于零，即 $X_j>0$。该条件是保证国民经济各部门的生产活动有意义的前提。

(2) 直接消耗系数 $a_{ii}<1$。a_{ii}为产品 i 生产过程中对本部门产品的消耗。$a_{ii}<1$ 是保证本部门生产有意义的前提。如果 $a_{ii}\geqslant 1$，则说明本部门所生产的产品不足以补偿本期自身生产的消耗，这样该部门就失去了存在的价值。

(3) 直接消耗系数矩阵必须满足条件：$0\leqslant a_{ij}<1$，矩阵 $\boldsymbol{A}$ 为非负矩阵。

由直接消耗系数的经济意义知，$a_{ij}>1$ 说明生产成果小于生产消耗，投入大于产出，由于部门本期生产的产品不足以补偿生产的消耗，生产也就失去了意义。当生产产品 j 时，不需投入产品 i，则 $a_{ij}=0$，故 a_{ij}的范围为 $0\leqslant a_{ij}<1$。

(4) 直接消耗系数的列总计 $\sum_i a_{ij}<1$。

$\sum_i a_{ij}<1$ 的经济涵义为生产资料的转移价值必须小于 1，这是由于任何产品的生产必须有活劳动参加的缘故。$\sum_i a_{ij}+a_{Dj}+a_{Nj}=1$ 是单位 j 产品在生产中所消耗的活劳动

和物化劳动占总产品比重之和，且 $a_{Dj}>0, a_{Nj}>0$，故 $\sum_i a_{ij}<1$。

2. 投入产出模型求解的数学条件

(1) $(I-A)^{-1}$必须存在。由模型$(I-A)X=Y$ 知，当给定外生变量 Y 以后，该方程组解存在的充分且必要条件为$(I-A)$为非奇异矩阵，即$(I-A)^{-1}$存在。

(2) $(I-\hat{C})^{-1}$存在。由模型$(I-\hat{C})X=N$ 知，当 N 作为外生变量给定时，X 有惟一一组解的充要条件是矩阵$(I-\hat{C})$非奇异，即$(I-\hat{C})^{-1}$存在。因$(I-\hat{C})$为对角形矩阵，故只要 $1-(\sum_i a_{ij}+a_{Dj})\neq 0$，即$(\sum_i a_{ij}+a_{Dj})\neq 1$ 即可。

线性代数理论可以证明，$(I-A)^{-1}$和$(I-\hat{C})^{-1}$在满足经济条件下是存在的，因此只要所建立的模型符合经济条件，模型就是可以求解的。

5.2.2 投入产出模型的应用

5.2.2.1 分析国民经济各种比例关系

1. 计算分析两大部类的比例及扩大再生产规模

按照马克思的社会再生产理论，社会生产分为生产资料生产和消费资料生产两大部类，因此，社会总产品也相应分成生产资料和消费资料。为了研究两大部类的比例关系及其变化的规律性，确定社会扩大再生产的规模，必须根据产品的实际经济用途对产品进行分类。用传统的计划方法很难作到这点，而应用投入产出表却可把两大部类清楚地划分出来。

根据价值型投入产出表(见表 5.1.3)，按产品的分配去向分类，中间产品、固定资产更新改造产品、用于生产性积累的产品以及进口产品中用于生产资料的部分产品等应属于第Ⅰ部类产品；而消费品、用于非生产性积累的产品及进口产品中用于消费的部分产品应属于第Ⅱ部类产品。对投入产出表 5.1.3 按上述分类稍加整理即可得出两大部类产品计算表(见表 5.2.1)。

表 5.2.1 两大部类产品计算表

	第Ⅰ部类					第Ⅱ部类			
	中间产品	更新改造	生产性固定资金积累	生产性流动资金积累	合计	消费	非生产性固定资金积累	非生产性流动资金积累	合计
农　业	720	5	50	25	800	1000	150	50	1200
工　业	3250	150	400	100	3900	50	50	50	150
建　筑	210	200	300	20	730	150	30	–	180
运输邮电	320	100	200	100	720	50	10	15	75
商业饮食	195	20	30	0	245	250	10	5	265
合计	4695	475	980	245	6395	1500	250	120	1870

由表 5.2.1 可见，社会总产品中第Ⅰ部类的产品为 6395 亿元，第Ⅱ类产品为 1870 亿元。

两大部类中 C、V、M 等价值流量应按下式计算：

$$\left.\begin{aligned} C_{\mathrm{I}} &= \sum_j X_j^{\mathrm{I}} \cdot C_j \quad , \quad C_{\mathrm{II}} = \sum_j X_j^{\mathrm{II}} \cdot C_j \\ V_{\mathrm{I}} &= \sum_j X_j^{\mathrm{I}} \cdot a_{Vj} \quad , \quad V_{\mathrm{II}} = \sum_j X_j^{\mathrm{II}} \cdot a_{Vj} \\ M_{\mathrm{I}} &= \sum_j X_j^{\mathrm{I}} \cdot a_{Mj} \quad , \quad M_{\mathrm{II}} = \sum_j X_j^{\mathrm{II}} \cdot a_{Mj} \end{aligned}\right\} \tag{5.2.1}$$

写成矩阵形式为：

$$\begin{pmatrix} C_{\mathrm{I}} & C_{\mathrm{II}} \\ V_{\mathrm{I}} & V_{\mathrm{II}} \\ M_{\mathrm{I}} & M_{\mathrm{II}} \end{pmatrix} = \begin{pmatrix} C_1 & C_2 & \cdots & C_n \\ a_{V1} & a_{V2} & \cdots & a_{Vn} \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} X_1^{\mathrm{I}} & X_1^{\mathrm{II}} \\ X_2^{\mathrm{I}} & X_2^{\mathrm{II}} \\ \vdots & \vdots \\ X_n^{\mathrm{I}} & X_n^{\mathrm{II}} \end{pmatrix} \tag{5.2.2}$$

式中 C_I，C_{II}——第Ⅰ、Ⅱ部类的生产资料转移价值；

V_I，V_{II}——第Ⅰ、Ⅱ部类必要劳动新创造的价值；

M_I，M_{II}——第Ⅰ、Ⅱ部类社会劳动所创造的价值；

X_j^{I}，X_j^{II}——j 生产部门第Ⅰ、Ⅱ部类的产品数量。

将数据代入式(5.2.2)得

$$\begin{pmatrix} C_{\mathrm{I}} & C_{\mathrm{II}} \\ V_{\mathrm{I}} & V_{\mathrm{II}} \\ M_{\mathrm{I}} & M_{\mathrm{II}} \end{pmatrix} = \begin{pmatrix} 0.3150 & 0.8395 & 0.5440 & 0.3271 & 0.7059 \\ 0.5000 & 0.0494 & 0.0549 & 0.0440 & 0.0980 \\ 0.1850 & 0.1111 & 0.4011 & 0.6289 & 0.1961 \end{pmatrix} \cdot \begin{pmatrix} 800 & 1200 \\ 3900 & 1150 \\ 730 & 180 \\ 720 & 75 \\ 245 & 265 \end{pmatrix} = \begin{pmatrix} 4332 & 813 \\ 688 & 647 \\ 1375 & 410 \end{pmatrix}$$

且

$$C_{\mathrm{I}} + C_{\mathrm{II}} = 5145$$

$$V_{\mathrm{I}} + V_{\mathrm{II}} = 1335$$

$$M_{\mathrm{I}} + M_{\mathrm{II}} = 1785$$

根据两大部类的实物构成和价值构成可以计算出两大部类的比例关系及其他一些指标。计算得出的两大部类分析见表 5.2.2。

表 5.2.2 两大部类分析表

指标		社会总产品 X/亿元	生产资料转移价值 C/亿元	必要劳动创造价值 V/亿元	社会劳动新创造价值 M/亿元	国民收入 N/亿元	有机构成 $\frac{C}{V}$	工资盈利率 $\frac{M}{V}$	成本盈利率 $\frac{M}{C+V}$
社会总产品		8265	5145	1335	1785	3120	3.85	1.337	0.2755
第Ⅰ部类	数　量	6395	4332	688	1375	2063	6.30	1.99	0.2739
	比重/%	77.37	84.2	51.54	77	66.12	–	–	–

（续）

指标		社会总产品 X /亿元	生产资料转移价值 C/亿元	必要劳动创造价值 V/亿元	社会劳动新创造价值 M/亿元	国民收入 N/亿元	有机构成 $\frac{C}{V}$	工资盈利率 $\frac{M}{V}$	成本盈利率 $\frac{M}{C+V}$
第Ⅱ部类	数　量	1870	813	647	410	1057	1.26	0.634	0.2808
	所占的比重/%	22.63	15.8	48.46	23	83.88	–	–	–

根据计算的结果和马克思主义扩大再生产理论，得出扩大再生产的规模为

$$I(V+M)-\text{Ⅱ}(C)=6888+1375-813=1250\text{ 亿元}$$

2. 分析确定积累和消费的比例

积累和消费的比例关系是国民经济中最重要、最综合的比例关系之一，它体现了发展生产和改善人民生活，长远利益和目前利益，国家、集体、个人利益之间的关系，同时，它还与国民经济的其他许多比例和指标有着密切的联系。

应用投入产出表和投入产出模型可以很直观地反映出积累与消费的比例及构成。由于国民生产总值$(N+D)$与最终产品总值相等，社会总产品的分配使用实质上是最终产品的分配与使用。因此，对表 5.1.3 进行整理可得积累与消费计算表(见表 5.2.3)。

表 5.2.3　积累与消费计算表

	国民收入生产额/亿元	国民收入使用额						
		合计/亿元	积累		消费		部门构成	
			数量	比重/%	数量	比重/%	积累	消费
农　业	1370	1275	275	21.57	1000	78.43	17.24	86.67
工　业	650	650	600	95.31	50	7.69	37.62	3.33
建　筑	415	500	350	70.00	150	30.00	21.94	10.00
运输邮电	535	375	325	86.67	50	13.33	20.38	3.33
商业饮食	150	295	45	15.25	250	84.75	2.82	16.67
合　计	3120	3095	1595	51.53	1500	48.47	100.00	100.00

表中的国民收入生产额是指国民经济各部门所创造的国民收入量，即 $V_j+M_j=N_j$ $(j=1,2,\cdots,n)$。

国民收入使用额为最终产品中积累和消费的总额，其中积累包括生产性积累和非生产性积累；消费包括用于人民个人消费和社会消费的最终产品。

积累比重和消费比重一栏分别是指各部门提供的积累和消费量占本部门国民收入使用额的比重。

部门构成是指各部门所提供的积累和消费量占总积累和消费量的比重，也就是单位积累和消费所需各部门提供的产品量。

当积累和消费再详细分类时，还可同样分析积累内部的比例结构和消费内部的比例结构。

由表 5.2.3 可见，积累和消费之间的关系已完全由投入产出表明显地揭示出来。

3. 分析国民经济各部门的比例及其各部门对国民经济的影响

国民经济各部门的比例是分析国民经济结构和了解各部门在整个经济系统中的地位和作用的重要依据。

由于投入产出表是以产品部门为单位进行统计计量的，同时又指明了各产品部门的产品去向及投入来源，因此可通过投入产出表对社会产品、国民收入生产的主要比例关系进行综合分析，从多方面找出部门间的内在联系。

(1) 直接消耗系数和完全消耗系数。投入产出模型中的直接消耗系数和完全消耗系数可比较深入地反映每一部门与其他部门的内在联系和相互依存关系，特别是通过完全消耗系数，可以揭示出部门间的种种间接联系，这是用传统的方法不能做到的。

(2) 通过投入产出表第Ⅰ部分的部门间流量矩阵 Q 可以说明各部门间在生产中的相互依赖程度，从而也可说明它们在国民经济中的地位和作用。如各项中间消耗占总消耗的比重就反映了各部门之间的相互依赖程度，以表 5.1.3 的资料计算的各项中间消耗占总消耗的比重见表 5.2.4。

表 5.2.4　中间消耗比重表

	农　业	工　业	建　筑	运输邮电	商业饮食	合　计
农　业	6.82	3.19	1.06	0	4.26	15.33
工　业	4.26	53.25	6.39	3.19	2.13.	69.22
建　筑	0.42	2.13	1.06	0.430	0.43	4.47
运输邮电	0.32	5.32	0.53	0.21	0.43	6.82
商业饮食	0.32	2.13	1.06	0.64	0	4.15
合　计	12.14	66.03	10.12	4.47	7.24	100.00

(3) 部门间联系程度系数。利用流量矩阵 Q 或直接消耗系数矩阵 A 还可反映从投入角度研究的某产品部门与其他部门的联系程度，该联系程度可用联系程度系数 r_{ij} 表示。其计算公式如下：

$$r_{ij}=\frac{X_{ij}}{\sum_i X_{ij}} \quad 或 \quad r_{ij}=\frac{a_{ij}}{\sum_i a_{ij}} \tag{5.2.3}$$

以表 5.1.3 资料计算的联系程度系数见表 5.2.5。

表 5.2.5　联系程度系数表

	农　业	工　业	建　筑	运输邮电	商业饮食
农　业	0.561	0.048	0.105	0	0.588
工　业	0.351	0.806	0.632	0.714	0.294
建　筑	0.035	0.032	0.105	0.095	0.059
运输邮电	0.026	0.081	0.053	0.048	0.059
商业饮食	0.026	0.032	0.105	0.143	0
合　计	1.000	1.000	1.000	1.000	1.000

(4) 影响系数。国民经济各部门对国民经济总体的影响是不同的,利用投入产出模型可确定各部门对国民经济总体的影响程度,第 j 个部门对国民经济的影响程度可用影响系数 r_j 表示,其计算公式为

$$r_j = \frac{\sum_i b_{ij} + 1}{\frac{1}{n}\sum_j \sum_i b_{ij} + 1} \tag{5.2.4}$$

r_j 的分子是生产单位 j 种最终产品对各种产品的完全需要量的总和,即是 $(\boldsymbol{I}-\boldsymbol{A})^{-1}$ 矩阵中列的合计值;分母是从国民经济总体角度研究的生产一个单位最终产品对各种产品完全需要量的平均值,即 $(\boldsymbol{I}-\boldsymbol{A})^{-1}$ 矩阵中各列合计数的平均值。二者之比:当 $r_j=1$ 时,说明 j 部门生产单位最终产品对各种产品的完全需要量等于国民经济平均水平;当 $r_j>1$ 时,说明 j 部门生产单位最终产品对各种产品的完全需要量大于国民经济平均水平,该部门的发展将需要更多的物资资源,因而体现出对国民经济的影响较大;当 $r_j<1$ 时,说明 j 部门生产单位最终产品对各种产品的完全需要量小于国民经济平均水平,故对国民经济的影响较小。上例的影响系数见表 5.2.6。

表 5.2.6 影响系数表

部 门	农 业	工 业	建 筑	运输邮电	商业饮食
影响系统 r_j	0.69778	1.4242	1.0645	0.7643	1.0490

由表 5.2.6 可见工业部门对国民经济的影响最大,农业部门影响最小。

5.2.2.2 产品价格模型和产品价格变动的影响

1. 产品价格模型

马克思的劳动价值理论是制定各种产品价格的理论基础。价格是商品价值的货币表现。产品的价值是由产品的社会必要劳动时间决定的,在不考虑供需关系的影响时,产品价值应与价格相等。

由投入产出表列所体现的产品价值形成过程可知,产品价值由生产资料转移价值 $(\sum_i X_{ij} + D_j)$ 和活劳动新创造价值 $(V_j + M_j)$ 组成的,而单位产品的价值——价格可用投入产出表计算的各种参数表示为

$$P_i = \sum_k a_{ki}p_k + a_{Di} + a_{Vi} + a_{Mi};\ i=1,2,\cdots,n; \tag{5.2.5}$$

用矩阵表示为 $\boldsymbol{P} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{P} + \boldsymbol{A}_D + \boldsymbol{A}_V + \boldsymbol{A}_M$

整理得 $$\boldsymbol{P} = [(\boldsymbol{I}-\boldsymbol{A})^{-1}]^{\mathrm{T}}(\boldsymbol{A}_D + \boldsymbol{A}_V + \boldsymbol{A}_M) \tag{5.2.6}$$

式中 $\boldsymbol{P} = (P_1, P_2, \cdots, P_n)^{\mathrm{T}}$ $\boldsymbol{A}_D = (a_{D1}, a_{D2}, \cdots, a_{Dn})^{\mathrm{T}}$

$\boldsymbol{A}_V = (a_{V1}, a_{V2}, \cdots, a_{Vn})^{\mathrm{T}}$ $\boldsymbol{A}_M = (a_{M1}, a_{M2}, \cdots a_{Mn})^{\mathrm{T}}$

式(5.2.6)即为产品的价格模型。

2. 分析某部门或某些产品部门价格变动对其他部门的影响

国民经济各部门之间的联系是复杂的,如果其中一个或几个部门价格发生变化,势必影响其他部门的价格或社会纯收入值(税金和利润)。因此需分析:

(1) 当某部门或某些部门价格变动时,其他部门为保持本部门的社会纯收入不变化

(即利润额不受影响),价格应如何调整;

(2) 当某些部门价格变动时,其他部门价格不变,各部门社会纯收入将如何变化。

了解上述两方面问题可为制定价格政策提供依据,是利用经济杠杆调整各部门之间关系的关键,是进行价格体制改革,使价格体制趋于合理的基础。投入产出模型,可模拟各种情况达到所提出的目的。

(1)模拟分析某部门或某些部门价格变化所引起的其他部门产品价格的变化。

设第 k 个部门价格变化 ΔP_k,则其他 $n-1$ 个部门为保持本部门社会纯收入不变,价格调整幅度应按下式计算:

$$\begin{bmatrix} \Delta P_1 \\ \Delta P_2 \\ \cdots \\ \Delta P_{k-1} \\ \Delta P_{k+1} \\ \cdots \\ \Delta P_n \end{bmatrix} = \left[\begin{bmatrix} 1-a_{1,1} & -a_{1,2} & \cdots & -a_{1,k-1} & a_{1,k+1} & \cdots & -a_{1,n} \\ -a_{2,1} & 1-a_{2,2} & \cdots & -a_{2,k-1} & -a_{2,k+1} & \cdots & -a_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -a_{k-1,1} & -a_{k-1,2} & \cdots & 1-a_{k-1,k-1} & -a_{k-1,k+1} & \cdots & -a_{k-1,n} \\ -a_{k+1,1} & -a_{k+1,2} & \cdots & -a_{k+1,k-1} & 1-a_{k+1,k+1} & \cdots & -a_{k+1,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -a_{n,1} & -a_{n,2} & \cdots & -a_{n,k-1} & -a_{n,k+1} & \cdots & 1-a_{n,n} \end{bmatrix}^{-1} \right]^{\mathrm{T}} \begin{bmatrix} a_{k,1} \\ a_{k,2} \\ \vdots \\ a_{k,k-1} \\ a_{k,k+1} \\ \vdots \\ a_{k,n} \end{bmatrix} \cdot \Delta P_k \tag{5.2.7}$$

式中 $\begin{bmatrix} a_{k,1} \\ a_{k,2} \\ \vdots \\ a_{k,k-1} \\ a_{k,k+1} \\ \vdots \\ a_{k,n} \end{bmatrix} \cdot \Delta P_k$ 为 K 部门提价对其余 $n-1$ 个部门产品价格的直接影响,再乘上 $[(\boldsymbol{I}-\boldsymbol{A})^{-1}]^{\mathrm{T}}$ 即是全部影响。

式(5.2.7)尚可利用完全消耗系数矩阵 $\boldsymbol{B}$ 或$(\boldsymbol{I}-\boldsymbol{A})^{-1}$矩阵中第 k 行元素来计算,其公式为

$$\begin{bmatrix} \Delta P_1 \\ \Delta P_2 \\ \cdots \\ \Delta P_{k-1} \\ \Delta P_{k+1} \\ \cdots \\ \Delta P_n \end{bmatrix} = \begin{bmatrix} b_{k1} \\ b_{k2} \\ \vdots \\ b_{k,k-1} \\ b_{k,k+1} \\ \vdots \\ b_{k,n} \end{bmatrix} \cdot \frac{\Delta P_k}{1+b_{kk}} \text{ 或 } \begin{bmatrix} \Delta P_1 \\ \Delta P_2 \\ \cdots \\ \Delta P_{k-1} \\ \Delta P_{k+1} \\ \cdots \\ \Delta P_n \end{bmatrix} = \begin{bmatrix} A_{k,1} \\ A_{k,2} \\ \vdots \\ A_{k,k-1} \\ A_{k,k+1} \\ \vdots \\ A_{k,n} \end{bmatrix} \cdot \frac{\Delta P_k}{A_{kk}} \tag{5.2.8}$$

式中 A_{kj}为$(\boldsymbol{I}-\boldsymbol{A})^{-1}$矩阵中第 k 行的元素,$j \neq k$。

该式的经济涵义是:$A_{k1}, A_{k2}, \cdots, A_{k,k-1}, A_{k,k+1}, \cdots, A_{k,n}$为$(\boldsymbol{I}-\boldsymbol{A})^{-1}$矩阵中第 k

行元素(不包括第 k 列),也就是完全消耗系数 b_{ki},它表明各部门生产单位最终产品对 k 产品的完全消耗量;A_{kk}为完全需要系数,是生产单位 k 产品对 k 种社会总产品的完全需要量 $b_{k,k+1}$,它也是提供单位 k 产品所需要的全部 k 产品量。$\frac{1}{A_{kk}}\cdot\Delta P_k$ 表示考虑了 k 产品生产对自身的完全消耗后的价格变化率。通俗地说,k 部门只有生产 A_{kk} 个 k 产品,才能生产出一个单位最终产品,而消费者消费单位最终产品 k,相当于消费了 A_{kk} 个 k 产品,因此$\frac{\Delta P_k}{A_{kk}}$就是从国民经济角度考虑完全消耗后的单位产品价格变动率。以该价格变动率乘以该部门生产单位产品对 k 产品的完全消耗量,即得到因第 k 部门价格变动对各部门的影响。

如果当多个部门价格变动时,其他各部门价格的变化可在式(5.2.8)的基础上经类推得出。

设有 n_1 个部门价格发生变动,不失一般性,令 n_1 个部门为投入产出表中最后 n_1 个部门,则计算公式为

$$\begin{bmatrix}\Delta P_1\\ \Delta P_2\\ \cdots\\ \Delta P_{n-n_1}\end{bmatrix}=\begin{bmatrix}A_{n_1,1} & A_{n_1+1,1} & \cdots & A_{n,1}\\ A_{n_1,2} & A_{n_1+2,2} & \cdots & A_{n,2}\\ \cdots & \cdots & \cdots & \cdots\\ A_{n_1,n-n_1} & A_{n_1+2,n-n_1} & \cdots & A_{n,n-n_1}\end{bmatrix}\cdot\begin{bmatrix}A_{n_1,n_1} & \cdots & A_{n_1,n}\\ A_{n_1+1,n_1} & \cdots & A_{n_1+1,n}\\ \cdots & \cdots & \cdots\\ A_{n,n_1} & \cdots & A_{n,n}\end{bmatrix}^{-1}\cdot\begin{bmatrix}\Delta P_{n_1}\\ \Delta P_{n_2}\\ \cdots\\ \Delta P_{n_n}\end{bmatrix}\tag{5.2.9}$$

上例中,假设农业部门和运输邮电部门价格分别上涨 5% 和 10%,其他部门为保持本部门社会纯收入不变,价格相应的调整幅度计算如下:

$$\begin{bmatrix}\Delta P_2\\ \Delta P_3\\ \Delta P_5\end{bmatrix}=\begin{pmatrix}0.1717 & 0.1862\\ 0.1630 & 0.1031\\ 0.5285 & 0.0953\end{pmatrix}\cdot\begin{pmatrix}1.2181 & 0.0572\\ 0.0337 & 1.0546\end{pmatrix}^{-1}\cdot$$

$$\begin{pmatrix}0.05\\ 0.10\end{pmatrix}=\begin{pmatrix}0.0237\\ 0.0156\\ 0.0283\end{pmatrix}=\begin{pmatrix}2.37\%\\ 1.56\%\\ 2.83\%\end{pmatrix}$$

(2) 模拟分析当某些部门产品价格变动时,其他部门价格不变,各部门社会纯收入的变化。

该模拟分析对我国价格体制改革有相当重要的作用。我国目前的价格体制存在很多不合理之处,然而立刻全面调整,使其达到合理,必然会对国民经济产生很大的冲击,给国民经济带来很大损失。因此,只能有步骤地对产品价格进行逐步调整。为保证市场稳定和人民生活水平不下降,国家规定当某些部门价格变动时其他产品价格要相对稳定,这就必然使各部门国民收人生产额发生变化。这种模拟分析可预测价格变动对国民收入的影

响,从而为建立合理的价格体系奠定基础。

从投入产出表知,各种参数之间存在如下关系:

$$\sum_i a_{ij} + a_{Dj} + a_{Vj} + a_{Mj} = 1 \qquad (j = 1, 2, \cdots, n)$$

设 P_{ij}——生产单位 j 产品时,所需 i 产品的价值;

P_i——单位 i 产品的价值;

ΔP_i——单位 i 产品价值的变化幅度。

故直接消耗系数又可表示为

$$a_{ij} = \frac{P_{ij}}{P_j} \tag{5.2.10}$$

当 i 产品和 j 产品均以不同的价格变动率变动时,a_{ij} 也必然发生变化,如变化后的直接消耗系数为 a_{ij}^*,则 a_{ij} 与 a_{ij}^* 的关系为

$$a_{ij}^* = \frac{P_{ij}(1+\Delta P_i)}{P_j(1+\Delta P_j)} = a_{ij}\frac{1+\Delta P_i}{1+\Delta P_j} \tag{5.2.11}$$

设产品调价后工资和折旧不调整,劳动报酬系数和折旧系数也将发生变化,如变化后的劳动报酬系数和折旧系数为 a_{Vj}^* 和 a_{Dj}^*,则 a_{Vj} 与 a_{Vj}^*,a_{Dj} 与 a_{Dj}^* 之间的关系为

$$a_{Vj}^* = a_{Vj}^*\frac{1}{1+\Delta P_j};\ a_{Dj}^* = a_{Dj}\frac{1}{1+\Delta P_j} \tag{5.2.12}$$

调价后,投入产出表各系数之间仍有如下关系:

$$\sum_i a_{ij}^* + a_{Dj}^* + a_{Dj}^* + a_{Mj}^* = 1 \tag{5.2.13}$$

将式(5.2.11)、式(5.2.12),代入式(5.2.13)得

$$\sum_i a_{ij}\frac{1+\Delta P_i}{1+\Delta P_j} + a_{Dj}\frac{1}{1+\Delta P_j} + a_{Vj}\frac{1}{1+\Delta P_j} + a_{Mj}^* = 1$$

移项并整理得

$$a_{Mj}^* = 1 - \frac{1}{1+\Delta P_j}\left[\sum_i a_{ij}(1+\Delta P_i) + a_{Dj} + a_{Vj}\right] \tag{5.2.14}$$

又因为

$$M_j^* = a_{Mj}^* \cdot X_j^* = a_{Mj}^* \cdot X_j(1+\Delta P_j) \tag{5.2.15}$$

所以,社会纯收入部分的增长率为

$$\frac{M_j^*}{M_j} - 1 = \frac{a_{Mj}^* X_j(1+\Delta P_j)}{a_{Mj} \cdot X_j} - 1 = \frac{a_{Mj}^*(1+\Delta P_j)}{a_{Mj}} - 1 \tag{5.2.16}$$

将式(5.2.14)代入式(5.2.16)并整理得:

$$\frac{M_j^*}{M_j} - 1 = \frac{\Delta P_j - \sum_i a_{ij}\Delta P_i}{a_{Mj}} \tag{5.2.17}$$

现以上例说明该模拟分析方法。

设农业产品价格上涨 5%,工业产品价格下降 3%,其他部门价格不变,分析各部门社会纯收入的变化。

用式(5.2.14)、式(5.2.15)、式(5.2.17)可得表 5.2.7。由表 5.2.7 可见，当农产品价格上涨 5%、工业产品下降 3% 时，各部门社会纯收入均发生变化，社会纯收入总共增长了 38.78 亿元，总物价水平下降 0.26%，由此可模拟分析当某些部门价格变化时，引起的社会纯收入、社会总产品的变化及物价总水平的变化，为决策提供信息。

表 5.2.7　各部门价格变化对社会纯收入的影响

指　标	农业 5%	工业 3%	建筑	运输邮电	商业饮食	合计
a_{Mj}^*	0.219	0.1005	0.4083	0.6346	0.1823	—
a_{Mj}	0.185	0.1111	0.4011	0.6289	0.1961	—
X_j^*	2100.0	3928.50	910.00	795.00	510.00	8243.5
X_j	2000.0	4050.00	910.00	795.00	510.00	8265.0
M_j^*	459.90	394.81	371.55	504.51	92.97	1823.74
M_j	370.00	450.00	365.00	500.00	100.00	1785.0
$M_j^* - M_j$	89.90	−55.19	6.55	4.51	−7.03	38.74
$\left(\frac{M_j^*}{M_j}-1\right)\cdot 100\%$	34.03	−12.26	1.79	0.90	−7.03	2.17
$\left(\frac{X_j^*}{X_j}-1\right)\cdot 100\%$	5	−3	0	0	0	−0.26

用类似方法还可确定价格变化对劳动报酬系数的影响以及工资应做如何调整，其计算公式为

$$a_{Vj}^* = 1 - \frac{1}{1+\Delta P_j}\left[\sum_i a_{ij}(1+\Delta P_j) + a_{Dj} + a_{Mj}\right] \tag{5.2.18}$$

$$\frac{V_j^*}{V_j} - 1 = \frac{p_j - \sum_i a_{ij}\Delta P_i}{a_{Vj}} \tag{5.2.19}$$

在制定价格政策前，反复应用上述方法分析、测算、调整即可模拟价格变动的各种影响，从各种方案中选择一个较好的方案，即可作为决策。

3. 分析税收、利润、工资、折旧的变动对产品价格的影响

税收、利润等是重要的经济杠杆，它可以调整各部门之间的关系，协调各部门之间的利益，是处理国家、部门之间关系的有力工具。因此研究这些因素的变动对国民经济的影响，可为国民经济管理提供很多重要信息。

工资和产品价格是影响人民生活水平的主要因素。研究工资的增长对价格的影响是处理国民经济再分配环节的重要依据。当前利用投入产出模型模拟分析工资、税收、利润、价格等因素对国民经济的影响，已成为国民经济管理的重要方法。

令　P'——折旧、工资、税金和利润变动后的产品价格；

ΔA_D——各部门折旧系数增量；

ΔA_V——各部门劳动报酬系数增量；

ΔA_M——各部门社会纯收入系数增量。

由价格模型可得各部门价格增量 ΔP 为：

$$\Delta P = P' - P = [(I - A)^{-1}]^{\mathrm{T}} \cdot [\Delta A_D + \Delta A_V + \Delta A_M] \tag{5.2.20}$$

该式可确定当各部门折旧、劳动报酬、利润变动时，相应各部门价格的变动。

如上例中，当农业部门劳动报酬增长 10%，工业部门社会纯收入提高 5%，运输邮电部门折旧提高 5% 时，各部门价格的变动为

$$\begin{bmatrix}\Delta P_1\\ \Delta P_2\\ \Delta P_3\\ \Delta P_4\\ \Delta P_5\end{bmatrix} = \begin{bmatrix}1.2181 & 0.1717 & 0.1630 & 0.0572 & 0.5286\\ 0.3665 & 2.8668 & 1.0840 & 0.6051 & 0.1747\\ 0.0243 & 0.0861 & 1.0951 & 0.0471 & 0.0714\\ 0.0337 & 0.1862 & 0.1031 & 1.0546 & 0.0953\\ 0.0208 & 0.0838 & 0.0920 & 0.0577 & 1.0306\end{bmatrix}^{\mathrm{T}} \cdot (\Delta \boldsymbol{A}_D + \Delta \boldsymbol{A}_V + \Delta \boldsymbol{A}_M)$$

将 $\Delta \boldsymbol{A}_D = \begin{bmatrix}0\\0\\0\\a_{D4}\times 5\%\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0.003145\\0\end{bmatrix}$；$\Delta \boldsymbol{A}_V = \begin{bmatrix}a_{V1}\times 10\%\\0\\0\\0\\0\end{bmatrix} = \begin{bmatrix}0.05\\0\\0\\0\\0\end{bmatrix}$；

$$\Delta \boldsymbol{A}_M = \begin{bmatrix}0\\a_{M2}\times 5\%\\0\\0\\0\end{bmatrix} = \begin{bmatrix}0\\0.005556\\0\\0\\0\end{bmatrix}$$

代入上式得

$$\begin{bmatrix}\Delta P_1\\ \Delta P_2\\ \Delta P_3\\ \Delta P_4\\ \Delta P_5\end{bmatrix} = \begin{bmatrix}1.2181 & 0.1717 & 0.1630 & 0.0572 & 0.5286\\ 0.3665 & 2.8668 & 1.0840 & 0.6051 & 0.1747\\ 0.0243 & 0.0861 & 1.0951 & 0.0471 & 0.0714\\ 0.0337 & 0.1862 & 0.1031 & 1.0546 & 0.0953\\ 0.0208 & 0.0838 & 0.0920 & 0.0577 & 1.0306\end{bmatrix}^{\mathrm{T}} \cdot \begin{bmatrix}0.05\\0.005556\\0\\0.003145\\0\end{bmatrix} = \begin{bmatrix}0.06305\\0.02510\\0.01450\\0.00954\\0.03104\end{bmatrix} = \begin{bmatrix}6.305\%\\2.510\%\\1.450\%\\0.954\%\\3.104\%\end{bmatrix}$$

5.2.3 投入产出法是进行国民经济综合平衡的得力工具

投入产出模型的基本形式是：

$$(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}=\boldsymbol{Y}$$

$$(\boldsymbol{I}-\hat{\boldsymbol{C}})\boldsymbol{X}=\boldsymbol{N}$$

两个模型都是 n 个方程 $2n$ 个变量的方程组，因此只要确定 n 个外生变量，方程都可有惟一的一组解。根据模型的这一特点，可用于国民经济综合平衡工作。

1. 以最终产品 $\boldsymbol{Y}$ 为外生变量，进行国民经济的综合平衡

当 $\boldsymbol{Y}$ 已知时，由模型 $(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}=\boldsymbol{Y}$，可求出社会总产品量 $\boldsymbol{X}$，即

$$\boldsymbol{X}=(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y}$$

当与生产能力和进口相平衡后，可按下式进行物力平衡：

$$\boldsymbol{Q}=\boldsymbol{A}\hat{\boldsymbol{X}}$$

从而确定部门间的流量。

按 $\boldsymbol{N}=(\boldsymbol{I}-\hat{\boldsymbol{C}})\boldsymbol{X}$，或 $\boldsymbol{Q}_v=\boldsymbol{A}_v\hat{\boldsymbol{X}}$ 求出各部门生产的国民收入值及价值流量，并用 $\sum_i Y_i=\sum_j N_j$ 进行财力平衡。

按 $\boldsymbol{L}=\boldsymbol{A}_L\hat{\boldsymbol{X}}$，确定生产所需劳动力数量，可与劳动力资源进行人力平衡，从而完成国民经济的综合平衡工作，进而可编制计划期投入产出表，完成国民经济计划工作。

2. 以社会总产品 $\boldsymbol{X}$ 为外生变量进行国民经济的综合平衡

当社会总产品 $\boldsymbol{X}$ 已知时，由 $\boldsymbol{Y}=(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}$、$\boldsymbol{N}=(\boldsymbol{I}-\hat{\boldsymbol{C}})\boldsymbol{X}$ 可确定出国民经济各部门的最终产品，与需要量平衡后，按 $\boldsymbol{Q}=\boldsymbol{A}\hat{\boldsymbol{X}}$ 可确定部门间的流量，完成物力平衡；按 $\sum_j N_j=\sum_i Y_i$ 进行财力平衡；按 $\boldsymbol{L}=\boldsymbol{A}_L\hat{\boldsymbol{X}}$ 进行人力平衡，从而可编制计划期投入产出表，完成制定国民经济计划的平衡和计算工作。

3. 以国民收入 $\boldsymbol{N}$ 为外生变量进行国民经济的综合平衡

当国民收入值 $\boldsymbol{N}$ 已知时，由 $\boldsymbol{X}=(\boldsymbol{I}-\hat{\boldsymbol{C}})^{-1}\boldsymbol{N}$ 可确定社会总产品量；进行生产能力与进口平衡后，可按 $(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{X}=\boldsymbol{Y}$ 确定各部门所能提供的最终产品量；在最终产品与社会需求平衡后，可按 $\boldsymbol{Q}=\boldsymbol{A}\hat{\boldsymbol{X}}$ 进行物力平衡；按 $\boldsymbol{L}=\boldsymbol{A}_L\hat{\boldsymbol{X}}$ 进行人力平衡，从而完成国民经济计划的综合平衡工作。

4. 以部分最终产品和部分社会总产品为外生变量进行国民经济综合平衡

以部分最终产品、部分社会总产品作为外生变量(共 n 个)进行国民经济综合平衡的方法是以第一种方法为基础，结合第二种方法对模型进行变型所形成的一种较为实用的国民经济综合平衡方法。外生变量的选择可组合成很多形式，其模型形式为

$$\begin{bmatrix} X_m \\ C_m \\ X_k \\ C_{n-(2m+k)} \end{bmatrix}=\begin{bmatrix} A_{11} & \cdots & A_{14} \\ A_{21} & \cdots & A_{24} \\ \cdots & \cdots & \cdots \\ A_{41} & \cdots & A_{44} \end{bmatrix}\begin{bmatrix} Y_m \\ C'_m \\ C'_k \\ Y_{n-(2m+k)} \end{bmatrix} \tag{5.2.21}$$

式中 A_{ij}——$(\boldsymbol{I}-\boldsymbol{A})^{-1}$矩阵的分块矩阵；

X_m，Y_m——产品生产量未知数，最终产品产量未知数；

C_m，C'_m——产品产量预测值，最终产品量预测值；

X_k，C'_k——产品生产未知数，最终产品量预测值；

$C_{n-(2m+k)}$，$Y_{n-(2m+k)}$——产品生产量预测值，最终点站产品量未知数。

当 $2m = n$ 时上述模型则变为

$$\begin{bmatrix} X_m \\ C_{n-m} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} Y_m \\ C_{n-m} \end{bmatrix} \tag{5.2.22}$$

或

$$\begin{bmatrix} C_m \\ X_{n-m} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} Y_m \\ C_{n-m} \end{bmatrix} \tag{5.2.23}$$

上述模型求解后可以确定全部的 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$，进而可进行国民经济的综合平衡工作并编制出计划期的投入产出表为制定国民经济计划服务。

5.2.4 经济效益分析和经济政策研究

所谓经济效益，一般是指以最小的消耗取得最大的经济效果。宏观经济效益是指从整体上，从整个国民经济全局出发，使消耗最少，效果最大。从生产成果和消耗对比角度分析考核经济效益的指标有：社会总产品与劳动消耗之比；国民收入与劳动消耗之比；国民收入与资源消耗之比等，总之把投入产出表第Ⅰ、Ⅱ、Ⅲ部分所提供的资料相互对比就可以非常灵活地从不同角度以不同指标评价经济效益。

上例的部分经济效益分析指标见表 5.2.8。

表 5.2.8 国民经济效益指标

	$\frac{N}{X}$	$\frac{N}{C}$	$\frac{M}{C+V}$	$\frac{X}{V}$	$\frac{N}{V}$	$\frac{C}{V}$
社会总产品	0.378	0.6064	1.2754	6.19	2.34	3.85
农　业	0.685	2.1750	1.2267	2.00	1.37	0.63
工　业	0.161	0.1912	1.1250	20.25	3.25	11.33
建　筑	0.456	0.8384	1.6697	18.20	8.30	9.90
运输邮电	0.673	2.0579	2.6949	22.71	15.29	7.43
商业饮食	0.294	0.4167	1.4166	10.20	3.00	7.20

运用这些资料可以找出国民经济发展的薄弱环节，提高宏观经济效益。

投入产出法尚可进行政策的研究和评价，如研究能源政策，就业政策，投资政策，环境保护政策等。在进行相应研究时，应根据投入产出法的基本原理，制定专用模型。

现以环境污染投入产出模型[①] 为例说明投入产出模型在经济政策研究中的应用。

这个模型与通常的投入产出模型的区别，首先表现在它的部门分类上，除了通常的生产部门以外，在主栏中增加了 m 个污染部门，在宾栏中增加了 m 个相应的消除污染部门。污染部门与消除污染部门的名称与分类是相同的。它们可以按照污染及消除污染的种类不同加以划分。由于污染的种类很多，不可能全部列入投入产出表中，一般都选择危害比较严重的正采取措施加以消除的部分。模型的基本结构见表 5.2.9。

① 该模型选自《经济数学方法与模型》，陈锡康等著。

表 5.2.9　投入产出环境污染模型的基本结构

		计量单位	中间产品										最终产品领域产生的污染	总计
			生产部门					消除污染部门						
			1	2	…	n	合计	1	2	…	m	合计		
生产部门	1		A_{11}	A_{12}	…	A_{1n}	$\sum_{j=1}^{n} X_{1j}$	E_{11}	E_{12}	…	E_{1m}	$\sum_{j=1}^{n} E_{1j}$	Y_1	X_1
	2		A_{21}	A_{22}	…	A_{2n}	$\sum_{j=1}^{n} X_{2j}$	E_{21}	E_{22}	…	E_{2m}	$\sum_{j=1}^{n} E_{2j}$	Y_2	X_2
	⋮													
	n		A_{n1}	A_{n2}	…	A_{nn}	$\sum_{j=1}^{n} X_{nj}$	E_{n1}	E_{n2}	…	E_{nm}	$\sum_{j=1}^{n} E_{nj}$	Y_n	X_n
	合计													
污染产生部门	1		P_{11}	P_{12}	…	P_{1n}	$\sum_{j=1}^{n} P_{1j}$	F_{11}	F_{12}	…	F_{1m}	$\sum_{j=1}^{n} F_{1j}$	R_1	Q_1
	2		P_{21}	P_{22}	…	P_{2n}	$\sum_{j=1}^{n} P_{2j}$	F_{21}	F_{22}	…	F_{2m}	$\sum_{j=1}^{n} F_{2j}$	R_2	Q_2
	⋮													
	m		P_{m1}	P_{m2}	…	P_{mn}	$\sum_{j=1}^{n} P_{mj}$	F_{m1}	F_{m2}	…	F_{mm}	$\sum_{j=1}^{n} F_{mj}$	R_m	Q_m
	合计													

模型中符号的涵义如下：

X_{ij}——第 j 个生产部门所消耗的第 i 个产生部门的产品数量；

P_{ij}——第 j 个生产部门在产生过程中所产生的第 i 种污染物的数量；

E_{ij}——第 j 个消除污染部门在消除污染过程中所消耗的第 i 个生产部门产品的数量；

F_{ij}——第 j 个消除污染部门在消除污染过程所产生的第 i 种污染物的数量，例如城市中废料化灰工厂在处理垃圾的同时又造成了空气污染；

R_i——最终需求领域所产生的第 i 种污染物的数量，如居民的垃圾、污水等；

X_j——第 j 部门的产量；

Y_j——第 j 部门的最终产品数量；

Q_j——以实物量计算的第 j 种污染物的总量；

S_j——以实物量计算的第 j 种污染物的消除量；

在这个模型中有如下关系式：

(1) 各部门产品的生产与使用平衡方程。

$$\sum_{j=1}^{n} X_{ij} + \sum_{j=1}^{m} E_{ij} + Y_i = X_i \qquad (i = 1,2,\cdots,n) \tag{5.2.24}$$

(2) 各种污染物的形成方程。

$$\sum_{j=1}^{n} P_{ij} + \sum_{j=1}^{m} F_{ij} + R_i = Q_i \qquad (i=1,2,\cdots,n) \tag{5.2.25}$$

在上述方程中，除了需引入直接消耗系数 a_{ij} 以外，尚需引入污染的产生系数 p_{ij}、消除污染部门的消耗系数 e_{ij} 和消除污染部门的污染产生系数 f_{ij}，其中：

$$p_{ij} = \frac{P_{ij}}{X_j} \qquad \begin{pmatrix} i=1,2,\cdots,m \\ j=1,2,\cdots,n \end{pmatrix} \tag{5.2.26}$$

P_{ij} 表示第 j 个生产部门单位产值所产生的第 i 种污染物的数量。

$$e_{ij} = \frac{E_{ij}}{S_j} \qquad \begin{pmatrix} i=1,2,\cdots,m \\ j=1,2,\cdots,n \end{pmatrix} \tag{5.2.27}$$

它表示消除单位第 j 种污染物所消耗的第 i 个生产部门的产品数量。

$$f_{ij} = \frac{F_{ij}}{S_j} \qquad \begin{pmatrix} i=1,2,\cdots,m \\ j=1,2,\cdots,n \end{pmatrix} \tag{5.2.28}$$

它表示消除单位第 j 种污染物所产生的第 i 种污染物的数量。

把上述系数代入式(5.2.24)和式(5.2.25)，得到

$$\sum_{j=1}^{n} a_{ij}X_j + \sum_{j=1}^{m} e_{ij}S_j + Y_i = X_i \qquad (i=1,2,\cdots,n) \tag{5.2.29}$$

$$\sum_{j=1}^{n} P_{ij}X_j + \sum_{j=1}^{m} f_{ij}S_j + R_i = Q_i \qquad (i=1,2,\cdots,n) \tag{5.2.30}$$

分别以 $\boldsymbol{A}=(a_{ij})_{n\times n}$、$\boldsymbol{P}=(p_{ij})_{m\times n}$、$\boldsymbol{E}=(e_{ij})_{n\times m}$、$\boldsymbol{F}=(f_{ij})_{m\times m}$ 表示直接消耗系数矩阵、污染的产生系数矩阵、消除污染部门的消耗系数矩阵和消除污染部门的污染产生系数矩阵，并令：

$$\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathrm{T}} \qquad \boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_n)^{\mathrm{T}}$$
$$\boldsymbol{S}=(S_1,S_2,\cdots,S_m)^{\mathrm{T}} \qquad \boldsymbol{Q}=(Q_1,Q_2,\cdots,Q_m)^{\mathrm{T}}$$
$$\boldsymbol{R}=(R_1,R_2,\cdots,R_m)^{\mathrm{T}}$$

则上述方程组可写为

$$\boldsymbol{AX}+\boldsymbol{ES}+\boldsymbol{Y}=\boldsymbol{X} \tag{5.2.31}$$

$$\boldsymbol{PX}+\boldsymbol{FS}+\boldsymbol{R}=\boldsymbol{Q} \tag{5.2.32}$$

以 $\hat{\boldsymbol{a}}$ 表示 $m\times m$ 对角矩阵，其对角线上的元素 a_j 表示第 j 种污染物的消除比例，因而：

$$S=\hat{a}Q \tag{5.2.33}$$

以 $\boldsymbol{U}=(U_1,U_2,\cdots,U_m)$ 表示各种污染物的净增加量向量，则：

$$\boldsymbol{U}=\boldsymbol{Q}-\boldsymbol{S}=\boldsymbol{Q}-\hat{\boldsymbol{a}}\boldsymbol{Q}=(\boldsymbol{I}-\hat{\boldsymbol{a}})\boldsymbol{Q} \tag{5.2.34}$$

据此可把式(5.2.31)和式(5.2.32)改写为

$$\boldsymbol{AX}+\boldsymbol{F}\hat{\boldsymbol{a}}\boldsymbol{Q}+\boldsymbol{Y}=\boldsymbol{X} \tag{5.2.35}$$

$$\boldsymbol{PX}+\boldsymbol{F}\hat{\boldsymbol{a}}\boldsymbol{Q}+\boldsymbol{R}=\boldsymbol{Q} \tag{5.2.36}$$

现讨论以下三种情况：

(1) 当 $\hat{a}=0$,即社会不进行消除污染的活动,这时我们有:

$$\boldsymbol{AX}+\boldsymbol{Y}=\boldsymbol{X} \tag{5.2.37}$$

$$\boldsymbol{PX}+\boldsymbol{R}=\boldsymbol{Q} \tag{5.2.38}$$

当给定最终产品 $\boldsymbol{Y}$ 及最终需求领域所产生的污染物数量 $\boldsymbol{R}$ 以后,就可以计算出各部门的产量及污染的产生量如下:

$$\boldsymbol{X}=(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y} \tag{5.2.39}$$

$$\boldsymbol{Q}=\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y}+\boldsymbol{R} \tag{5.2.40}$$

这时 $\boldsymbol{L}_0=\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}$ 就表示污染物的完全产生系数矩阵。污染物的完全产生系数表示为生产单位最终产品所直接和间接产生的各种污染物的数量。

(2) 当 $\hat{a}=\boldsymbol{I}$,即各部门所产生的污染物全部被消除,因 $\boldsymbol{S}=\boldsymbol{Q}_0$ 而。这时:

$$\boldsymbol{AX}+\boldsymbol{EQ}+\boldsymbol{Y}=\boldsymbol{X} \tag{5.2.41}$$

$$\boldsymbol{PX}+\boldsymbol{FQ}+\boldsymbol{R}=\boldsymbol{Q} \tag{5.2.42}$$

当给定 $\boldsymbol{Y}$ 和 $\boldsymbol{R}$ 以后,我们也可以计算出各部门的产量以及污染物的产出量,即求解该方程组得:

$$\begin{aligned}\boldsymbol{X}=&(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y}+(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{E}[(\boldsymbol{I}-\boldsymbol{F})-\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{E}]^{-1}\cdot\\&[\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y}+\boldsymbol{R}]\end{aligned} \tag{5.2.43}$$

$$\boldsymbol{Q}=[(\boldsymbol{I}-\boldsymbol{F})-\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{E}]^{-1}[\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y}+\boldsymbol{R}] \tag{5.2.44}$$

这时污染物的完全产生系数,可由以下分块矩阵 $\boldsymbol{L}_1$ 的第二行元素表示。①

$$\boldsymbol{L}_1=\begin{bmatrix}\boldsymbol{I}-\boldsymbol{A} & -\boldsymbol{E}\\ -\boldsymbol{P} & (\boldsymbol{I}-\boldsymbol{F})\end{bmatrix}^{-1}=$$

$$\begin{bmatrix}(I-A)^{-1}+(I-A)^{-1}E[(I-F)-P(I-A)^{-1}E]^{-1}P(I-A)^{-1} & (I-A)^{-1}E[(I-F)-P(I-A)^{-1}E]^{-1}\\ [(I-F)-P(I-A)^{-1}E]^{-1}P(I-A)^{-1} & [(I-F)-P(I-A)^{-1}E]^{-1}\end{bmatrix} \tag{5.2.45}$$

也就是说 $[(\boldsymbol{I}-\boldsymbol{F})-\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{E}]^{-1}\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}$ 是各生产部门污染物的完全产生系数矩阵,并且有

$$\begin{bmatrix}X\\ Q\end{bmatrix}=L_1\begin{bmatrix}Y\\ R\end{bmatrix} \tag{5.2.46}$$

在第一种情况下完全没有进行消除污染的活动,这显然不符合环境保护的要求,实际情况也并非如此。第二种情况下污染完全被消除是一种理想的情况,目前也不可能做到。实际情况是所消除掉的各种污染物占所产生的污染物的一个较小的比例,所以我们应该讨论更一般的情况。

(3) 当 $0<a_1<1$ 时,给定 $\boldsymbol{Y}$ 和 $\boldsymbol{R}$ 值后,由式(5.2.35)和式(5.2.36)可以解出

$$\begin{aligned}\boldsymbol{X}=&(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y}+(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{E}\hat{a}[(\boldsymbol{I}-\boldsymbol{F}\hat{a})-\\&\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{E}\hat{a}]^{-1}[\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y}+\boldsymbol{R}]\end{aligned} \tag{5.2.47}$$

① 第一行与考虑污染的情况下各生产部门产品的完全消耗系数矩阵相对应。

$$\boldsymbol{Q}=[(\boldsymbol{I}-\boldsymbol{F}\hat{\boldsymbol{a}})-\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{E}\hat{\boldsymbol{a}}]^{-1}[\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{Y}+\boldsymbol{R}] \tag{5.2.48}$$

这时污染物的完全产生系数同样可以由分块矩阵的第二行表示

$$\boldsymbol{L}=\begin{bmatrix} I-A & -E\hat{a} \\ -P & (I-F\hat{a}) \end{bmatrix}^{-1}=$$
$$\begin{bmatrix} (I-A)^{-1}+(I-A)^{-1}E\hat{a}[(I-F\hat{a})-P(I-A)^{-1}E\hat{a}]^{-1}P(I-A)^{-1} & \\ [(I-F\hat{a})-P(I-A)^{-1}E\hat{a}]^{-1}P(I-A)^{-1} & \end{bmatrix}$$
$$\begin{bmatrix} (I-A)^{-1}E\hat{a}[(I-F\hat{a})-P(I-A)^{-1}E\hat{a}]^{-1} \\ [(I-F\hat{a})-P(I-A)^{-1}E\hat{a}]^{-1} \end{bmatrix} \tag{5.2.49}$$

$$\begin{bmatrix} X \\ Q \end{bmatrix}=L\begin{bmatrix} Y \\ R \end{bmatrix} \tag{5.2.50}$$

同时可以计算出污染的消除量 S 和污染的净增加量 U：

$$S=\hat{a}[(I-F\hat{a})-P(I-A)^{-1}E\hat{a}]^{-1}[P(I-A)^{-1}Y+R] \tag{5.2.51}$$

$$U=(1-\hat{a})[(I-F\hat{a})-P(I-A)^{-1}E\hat{a}]^{-1}[P(I-A)^{-1}\boldsymbol{Y}+R] \tag{5.2.52}$$

因为式(5.2.37)和式(5.2.38)是式(5.2.35)和式(5.2.36)当 $\hat{a}=0$ 时的特殊情况，所以在式(5.2.47)和式(5.2.48)中，令 $\hat{a}=0$，即可得出式(5.2.39)和式(5.2.40)。同样式(5.2.41)和式(5.2.42)是式(5.2.35)和式(5.2.36)中令 $\hat{a}=I$ 时的特殊情况，所以在式(5.2.47)、式(5.2.48)和式(5.2.49)中，令 $\hat{a}=I$，即可得出式(5.2.43)、式(5.2.44)和式(5.2.45)。

该模型可从数量上确定污染量和污染消除量以及污染增加量等，如将其转化为价值型模型尚可解决消除污染费用及其对各部门产品价格的影响等问题，可为国家制定环境保护政策提供决策依据。

5.3 企业投入产出模型

企业投入产出模型是投入产出法在企业中的应用。企业中采用投入产出法是现代化管理的要求，是实现企业全面计划管理的一个具体途径。

企业投入产出模型是运用现代数学方法研究企业的生产过程，进行生产与消费之间平衡的一种科学方法。它以在企业管理中使用电子计算机为背景，帮助我们查明生产过程中所存在的基本数量关系和数量界限，达到细致、准确、快速的要求；它可提供一些传统方法无法得到的重要信息和技术经济指标。

5.3.1 企业投入产出表的结构

价值型企业投入产出表见表 5.3.1。由表 5.3.1 可见，价值型投入产出表是由以下六个部分组成：

(1) 第Ⅰ部分是由企业的 n 种产品或劳务组成的(有时也可由企业的子部门组成)，即由横行 n 个本企业内部资源与纵列 n 种本企业内部消耗交叉组成的棋盘式部分，它由同名称同次序的产品组成，反映了企业内各产品之间的生产联系、各种产品与劳务之间的

生产技术性联系；

(2) 第Ⅱ部分是本企业内部资源与最终产品栏相交组成的，反映本企业最终产品的最终使用情况；

(3) 第Ⅲ部分为外购资源与企业内部消耗交叉组成的，它反映外购资源的使用；

(4) 第Ⅳ部分是外购资源与最终产品相交叉组成的，它反映企业的转手外销业务；

(5) 左下角为第Ⅴ部分，它反映固定费用(企业管理费，其中包括折旧、大修理和管理人员工资等)和新创造价值(生产人员工资和税金、利润等)；

(6) 右下角为第Ⅵ部分，企业投入产出表为空白。

表 5.3.1 企业投入产出表结构 (价值型)

		本企业内部消耗					最终产品			总产量
		1	2	…	n	合计	外销	储备	合计	
本企业自有资源	1	X_{11}	X_{12}	…	X_{1n}	$\sum_j X_{1j}$			Y_1	X_1
	2	X_{21}	X_{22}	…	X_{2n}	$\sum_j X_{2j}$			Y_2	X_2
	⋮	⋮		…	⋮				⋮	⋮
	n	X_{n1}	X_{n2}	…	X_{nn}	$\sum_j X_{nj}$			Y_n	X_n
	计	$\sum_i X_{i1}$	$\sum_i X_{i2}$	…	$\sum_i X_{in}$	$\sum_{ij} X_{ij}$			$\sum_i Y_i$	$\sum_i X_i$
外购资源	1	h_{11}	h_{12}	…	h_{1n}	$\sum_j h_{1j}$				
	2	h_{21}	h_{22}	…	h_{2n}	$\sum_j h_{2j}$				
	⋮	⋮		…	⋮					
	n	h_{n1}	h_{n2}	…	h_{nn}	$\sum_j h_{nj}$				
	计	$\sum_i h_{i1}$	$\sum_i h_{i2}$	…	$\sum_i h_{in}$	$\sum_{ij} h_{ij}$				
企业管理费		S_1	S_2	…	S_n	$\sum_j S_j$				
劳动工资		V_1	V_2	…	V_n	$\sum_j V_j$				
工厂成本税金、利润		M_1	M_2	…	M_n	$\sum_j M_j$				
总产品		X_1	X_2	…	X_n	$\sum_j X_j$				

5.3.2 企业投入产出模型

5.3.2.1 企业投入产出表所反映的平衡关系

(1) 表 5.3.1 的第Ⅰ、Ⅲ、Ⅴ三部分表明了纵列各产品对横行各要素的投入情况，即生产消耗的构成，包括企业内部资源、外购资源、设备和劳动的投入，即所谓产品的生产成

本，加上税金和利润，即是各种产品的生产总值。它体现了产品的价值形成过程和生产消耗过程。用公式可描述为

$$\sum_{i=1}^{n} X_{ij} + \sum_{i=1}^{k} h_{ij} + S_j + Z_j = X_j \tag{5.3.1}$$

式中 X_{ij}——年内第 j 种产品生产所需要的第 i 种产品数量；

h_{ij}——年内第 j 种产品生产所消耗的第 i 种外购资源数量；

S_j——第 j 种产品一年中的企业管理费；

Z_j——第 j 种产品一年中新创造的价值，$Z_j = M_j + V_j$。

(2) 表 5.3.1 的第Ⅰ、Ⅱ两部分表明企业自产产品按用途的分配使用情况，用来补偿已消耗的生产资料或作为最终产品，故有如下内部资源分配方程：

$$\sum_{j} X_{ij} + Y_j = X_j \qquad (j = 1,2,\cdots,n) \tag{5.3.2}$$

在此基本平衡关系的基础上，引入企业消耗结构——内部资源消耗系数和外购资源消耗系数。

设内部资源消耗系数和外购资源消耗系数分别为 a_{ij} 和 d_{ij}，则计算公式为

$$a_{ij} = \frac{X'_{ij}}{x_j} \qquad (i,j = 1,2,\cdots,n) \tag{5.3.3}$$

$$d_{ij} = \frac{h_{ij}}{x_j} \qquad (i = 1,2,\cdots,n) \tag{5.3.4}$$

式中 X'_{ij}——一年中 j 产品生产对 i 种内部资源的实际消耗量，$X'_{ij} = X_{ij} - \Delta X_{ij}$。

当 $\Delta X_{ij} > 0$ 时，$X'_{ij} > X_{ij}$，表示动用库存；

$\Delta X_{ij} = 0$ 时，$X'_{ij} = X_{ij}$，表示库存不变；

$\Delta X_{ij} < 0$ 时，$X'_{ij} < X_{ij}$，表示增加库存。

由式(5.3.3)和式(5.3.4)可知，在假设 $\Delta X_{ij} = 0$ 时，

$$X_{ij} = a_{ij}X_j;\ h_{ij} = d_{ij}X_j \tag{5.3.5}$$

将其代入式(5.3.1)、式(5.3.2)得

$$\sum_{j} a_{ij}X_j + Y_i = X_i \qquad (i = 1,2,\cdots,n) \tag{5.3.6}$$

$$\sum_{i} a_{ij}X_j + \sum d_{ij}X_j + W_j = X_j \qquad (j = 1,2,\cdots,n) \tag{5.3.7}$$

式中 $W_j = S_j + Z_j$——企业管理费和新创造价值的总和。

整理后用矩阵形式表示为

$$(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X} = \boldsymbol{Y} \tag{5.3.8}$$

$$(\boldsymbol{I} - \hat{\boldsymbol{C}})\boldsymbol{X} = \boldsymbol{W} \tag{5.3.9}$$

其中

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

$$\hat{\boldsymbol{C}}=\begin{bmatrix}\sum_{i}^{n}a_{i1}+\sum_{i}^{k}d_{i1} & 0 & \cdots & 0\\ 0 & \sum_{i}^{n}a_{i2}+\sum_{i}^{k}d_{i2} & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & \sum_{i}^{n}a_{in}+\sum_{i}^{k}d_{in}\end{bmatrix}$$

$$\boldsymbol{X}=\begin{bmatrix}X_1\\X_2\\\vdots\\X_n\end{bmatrix};\qquad \boldsymbol{Y}=\begin{bmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{bmatrix};\qquad \boldsymbol{W}=\begin{bmatrix}W_1\\W_2\\\vdots\\W_n\end{bmatrix}$$

式(5.3.8)体现了最终产品和总产品之间的关系，$(\boldsymbol{I}-\boldsymbol{A})^{-1}$为内部资源完全需要系数，它反映了生产单位最终产品对总产品的需求，它与内部资源完全消耗系数仍有如下关系：

$$(\boldsymbol{I}-\boldsymbol{A})^{-1}=\boldsymbol{B}+\boldsymbol{I} \tag{5.3.10}$$

式(5.3.9)体现了企业总产品和企业管理费、税金利润之间的关系。由于企业管理费基本为一固定量，故该式所反映的基本是总产品与利润的关系。$\hat{C}$ 中的元素为企业生产单位产品所消耗的内部资源和外购资源量。外购资源的完全消耗系数 G 按下式计算：

$$\boldsymbol{G}=\boldsymbol{D}(\boldsymbol{I}-\boldsymbol{A})^{-1} \tag{5.3.11}$$

根据企业投入产出表尚可依投入产出原理计算工资系数、劳动消耗系数、利润系数等经济参数，为企业经济活动分析创造条件。

5.3.3 价值型企业投入产出模型的应用

价值型企业投入产出模型是企业现代化管理的有力工具，其应用主要为以下几个方面：

(1) 制定企业计划。企业实行全面计划管理，制定计划是重要环节。应用企业投入产出模型可进行各种计划的计算。如利用式(5.3.8)在以最终产品为外生变量时，可计算总产品计划量；根据总产品计划量，按 $h_{ij}=d_{ij}X_{ij}$ 可编制物资供应计划，按 $(\boldsymbol{I}-\hat{\boldsymbol{C}})\boldsymbol{X}=\boldsymbol{W}$ 可计算各种产品的利润计划量等。

(2) 进行企业经济活动分析和进行经济效益评价。由于企业投入产出表分类较细，它综合地反映了企业生产活动，因此可以从各个角度，用不同的指标分析企业活动，如 $\frac{M}{V}$ 即为各种产品的工资盈利率、$\frac{M}{C}$ 为资金盈利率等，所要计算的各种指标一目了然。通过对投入产出表的分析和评价，尚可揭示出生产的薄弱环节、存在的问题，为制定今后生产方针提供信息。

(3) 可分析外部市场变化对企业的影响。如外部原材料市场发生变化，某些物资由长线变为短线，对企业生产计划有何影响，原材料价格的变化对盈利额的影响以及产品价格应如何调整等。

第 6 章　静态线性系统最优化模型

系统，尤其是包含社会因素、组织管理因素的系统都处于不断地运动状态。系统的特性、系统的各组成部分及他们之间的联系都随时间的推移而变化，它们都是动态系统。对这种系统进行定量描述，因影响系统特性的各要素是时间的函数，故在系统分析时会产生很多描述和计算上的困难。为了能够比较简单地对系统进行定量描述，经常用系统处在相对静止状态或平衡状态来代替系统本来的复杂运动状态。我们把系统处在这种相对静止或平衡的状态叫做静态，描述系统这种状态的模型叫做静态模型。

上一章中介绍的平衡模型就是静态模型的一种。它所揭示的只是系统各组成部分之间的相互联系和比例关系，而未反映系统达到目标的程度。使用该模型可提出若干符合部门之间比例关系的计划方案，但它们不一定是使系统处于最优运行状态的方案。在处理系统问题时，经常需要得到一个既符合各部分之间制约关系，又能使某一评价指标达到最大或最小的方案，采用平衡模型显然是不能胜任的。因此需要研究一种在一定限制条件下使系统处于最优运行状态的理论和方法——系统的最优化方法。

6.1　最优化及最优化模型的建立方法

6.1.1　最优化

最优化是系统方法处理问题的基本方法之一，是从整体出发，实现系统最优运行状态的根本保证，是系统工程处理问题的关键环节。

系统最优化是在系统目标分析、环境分析和系统预测的基础上通过建立数学模型、数学模型的求解而实现系统的定量化，并为系统运行在最优状态提供科学决策依据的过程和方法蹬总称。

系统总是处在一定的环境条件之中，研究系统的目的，无非是使处在一定环境条件限制下的系统在按某些目标评价时达到最优状态，因此最优化过程是指得到系统在一定限制条件下达到评价目标极大值(极小值)方案的过程。该过程一般包括：

(1) 从系统思想出发对系统评价目标的定性和定量分析；

(2) 对系统约束条件的定性和定量分析；

(3) 建立系统模型；

(4) 系统模型求解；

(5) 对求解结果进行分析和系统因素变化时对求解结果影响的分析。

在系统最优化的过程中，以定性分析为指导，把系统目标、约束条件用数学形式进行描述，建立数学模型并求解的方法叫最优化方法，应用最优化方法所建立的模型叫最优化模型。

最优化方法是最优化过程的关键和核心，其中应用最广泛的最优化方法就是数学规划。

数学规划是研究系统在一定约束条件下达到某一评价目标最大(或最小)的一种决策方法。其关键是从系统思想出发，在定性分析的指导下建立数学规划模型。

数学规划模型的一般形式是：

满足约束条件：
$$G_i(\boldsymbol{X}_j) \leqslant b_i$$
$$\boldsymbol{D}_j \leqslant \boldsymbol{X}_j \leqslant D_j \tag{6.1.1}$$

使评价目标：$\max(\min)\quad Z = f(\boldsymbol{X}_j)$ (6.1.2)

式(6.1.1)是系统必须满足的限制条件的数学描述，它通常是由等式或不等式组成，一般称之为约束条件，约束条件在模型中一般以 s.t. 表示模型中可简写成 s.t.，s.t. 是英文 subject to 的缩写，其涵义为“受 …… 的限制”。

式(6.1.2)是系统评价目标的数学描述，称之为目标函数。

建立系统最优化模型的过程实质上是通过定性分析，把环境对系统的限制和系统的评价目标用促使系统状态变化的因素(即变量)来描述的过程。根据对系统目标函数和约束条件的描述形式的差别，数学规划模型可分为线性规划模型和非线性规划模型等。本章主要介绍各种线性最优化模型的建立、求解和求解结果的分析方法。

6.1.2 数学规划模型的建立方法和步骤

数学规划模型的建立方法和步骤为：

(1) 问题的提出。该步是建立数学模型的开始和基础，其主要内容为：根据要达到的目的，提出所解决的问题，确定评价系统的指标；根据系统所处的环境，确定所调查和所需掌握的材料，并对材料进行分类、整理，找出影响系统评价指标的因素。

(2) 编制原始数据表。该步是建立模型的依据，其内容是按所提出的问题，对分类、整理后的资料进行编撰，找出数据之间的联系，发现系统的规律，编出反映系统各因素之间平衡关系的原始数据表。

(3) 设置变量和各种符号。

(4) 根据系统目标分析的结论，建立目标函数。该步的作法一般为：将系统评价目标用简练的文字描述，即建立语言模型并用变量和各种参数将文字描述转变为数学描述。

(5) 确定约束条件。在问题提出的基础上，充分考虑系统环境因素和系统内在因素，将系统的各种限制条件和平衡关系用简练的文字描述，并用变量和各种参数以不等式或等式的形式进行数学描述。

建立模型是复杂的创造过程，它不仅是一门技术，而且是一门艺术。所建立的模型能否反映提出问题的实质是模型成功与否的标志。模型的繁简，关键就在于能否准确地反映所提出问题的实质。一般情况下，建立模型的准则为：在准确地反映所提出问题的前提下，建立简单明了、通俗易懂的模型。切忌故弄玄虚，搞数学游戏，把本来已清楚的问题复杂化；同时也要注意把复杂系统简单化的倾向，也就是在建立模型时，要注意防止掉进过于简单的陷并和陷入过于复杂的泥潭。

6.2 系统线性规划模型

当系统最优化模型的目标函数 $f(\boldsymbol{X})$ 和约束条件 $g_i(\boldsymbol{X})$ 均为变量 $\boldsymbol{X}$ 的线性函数时，数学规划模型就是线性规划模型。

线性规划模型是数学规划中研究最早、发展较快、应用较广且比较成熟的一个分支。它可解决如下两个方面的问题：

(1) 一个系统如何统筹安排，才能以尽量少的人力、物力、财力等资源，达到系统的目标；

(2) 一个系统如何在人力、财力、物力等资源一定的情况下，达到系统的最大(最小)目标。这两个问题实质上是一个问题的两个侧面，它们都是寻求系统的某一目标最优的问题。

6.2.1 运输问题

运输问题是一类线性规划问题的总称。它是由在产地和销地之间安排调运方案开始的，故称为运输问题。但这类模型所解决的不仅仅是编制产品的最优调运方案，而是与该问题有类似特点的所有问题，因此运输模型是这类问题的总称。

现以产品运输计划的编制为例，说明该模型的形式和建立方法。

1. 问题的提出

我国木材产品产区主要有 4 个，分布在东北、西北、西南、东南，各地区生产的木材品种和数量各不相同。所生产的各种木材除本地区使用外尚需满足其他地区的需求，现需根据各生产地的木材生产量和各需求地的需求量，编制木材调运计划，使木材运输费用最省。

编制木材运输计划需考虑供需关系，供需关系无外乎有以下 3 种：

(1) 总产量和总需求量相等。这种情况称之为平衡运输问题。其特点是生产地生产的木材必须全部运出，否则需求地的需求量就得不到满足；从另一方面看，需求地的需求量必须如数满足，否则将造成生产地产品积压或影响其他地区需求。

(2) 总产量大于需求量。这是不平衡型运输问题。其特点是各地需求量完全可以满足，同时某些产地的部分产品必须储存起来，如果以运输费用最省为目标，运输计划的安排就必须同时考虑各地区之间的调运量和存储量。

(3) 总需求量大于总产量。这也是一种不平衡的运输问题，其特点是所生产的产品必须全部运出才能最大限度地满足需求，然而即使全部运出产品，也满足不了需求，这样以运输费用最省来安排运输计划就需考虑少向哪些需求地供应的问题。

2. 编制原始数据表

本问题中 4 个产地作为主栏，6 个需求地作为宾栏。主栏和宾栏的相交处体现出产地和需求地的运输关系，将调查或计算的单位产品运输费用填入相交处的小框中，将生产地的生产量和需求地的需求量填入相应的栏内，则完成了原始数据表的编制工作。

本例所得的原始数据表见表 6.2.1。

表 6.2.1　木材调运原始数据表　　(单位:万 m^3)

	东北	西北	东南	西南	华北	华南	产量
东北	1	3	5	7	2	4	1000
西北	3	1	5	3	2	4	500
东南	5	5	1	5	3	2	700
西南	7	3	5	1	3	3	800
需求量	700	400	700	200	600	400	3000

3. 设变量和设置符号

设 x_{ij} 为第 i 个生产地向第 j 个需求地运输的木材数量,因已知各产地向需求地的运输单价、各产地的生产量和各需求地的需求量,故设置符号:

a_i——第 i 个生产地的生产量($i = 1,2,3,4$);

b_j——第 j 个需求地的需求量($j = 1,2,\cdots,6$);

c_{ij}——第 i 个生产地向第 j 个需求地供应单位木材的运费。

4. 建立目标函数

该系统的目标是节约运输费用,其评价指标为总运输费用。$C_{ij} \cdot X_{ij}$ 为第 i 生产地向第 j 需求地的运输费用,加总即得总费用,故总运费为

$$\sum_j \sum_i C_{ij} \cdot X_{ij} \tag{6.2.1}$$

系统目标是使运输费用最省,故目标函数为

$$Z = \sum_h \sum_i C_{ij} \cdot X_{ij} \to \min$$

5. 确定约束条件

(1) 平衡型运输问题($\sum_i a_i = \sum_j b_j$)

① 生产地的木材必须全部运出,即每个生产地运往各需求地的木材必须与生产量相符。该约束条件反映了行的平衡关系,数学描述为

$$\sum_j X_{ij} = a_i \qquad (i = 1,2,3,4) \tag{6.2.2}$$

② 需求地的需求量必须得到满足,即所有生产地向第 j 个需求地供应的木材量必须与其需求量相等。该约束条件反映了列的平衡关系,数学描述为

$$\sum_i X_{ij} = b_j \qquad (j = 1,2,\cdots,6) \tag{6.2.3}$$

③ 运输量不能为负值,数学描述为

$$X_{ij} \geqslant 0 \qquad (i = 1,2,3,4; j = 1,2,\cdots,6) \tag{6.2.4}$$

将原始数据代入式(6.2.1) ~ 式(6.2.4)即得到平衡型运输问题的线性规划模型。

如果木材产地有 m 个,需求地有 n 个,则式(6.2.1) ~ 式(6.2.4)中的 i,j,就改写成

$i = 1,2,\cdots,m; j = 1,2,\cdots,n$，这就是运输问题模型的一般形式

$$\min \quad Z = \sum_j \sum_i C_{ij} \cdot X_{ij}$$

s.t.为：

$$\sum_j X_{ij} = a_i \qquad (i = 1,2,\cdots,m)$$

$$\sum_i X_{ij} = b_j \qquad (j = 1,2,\cdots,n)$$

$$X_{ij} \geqslant 0 \qquad (i = 1,2,\cdots,m \quad j = 1,2,\cdots,n)$$

(2) 产大于需的不平衡运输问题($\sum_i a_i > \sum_j b_j$)

① 生产地的木材不一定全部运出，即第 i 个生产地运往各需求地的木材之和可能少于生产量，最多等于生产量，其数学描述为

$$\sum_j X_{ij} \leqslant a_i \qquad (i = 1,2,3,4) \tag{6.2.5}$$

②、③ 约束条件与平衡型运输模型相同，这种情况的运输模型的一般形式为

目标函数为：

$$Z = \sum_j \sum_i C_{ij} \cdot X_{ij} \to \min$$

$$\sum_j X_{ij} \leqslant a_i \qquad (i = 1,2,\cdots,m)$$

s.t.为：

$$\sum_i X_{ij} = b_j \qquad (j = 1,2,\cdots,n)$$

$$X_{ij} \geqslant 0 \qquad (i = 1,2,\cdots,m \quad j = 1,2,\cdots,n)$$

(3) 产小于需的不平衡运输问题($\sum_i a_i < \sum_j b_j$)

①、③ 约束与平衡型运输模型相同。

需求地的需求量不一定完全满足，即各生产地运往同一需求地的木材数量小于或等于需求量，其数学描述为

$$\sum_i X_{ij} \leqslant b_j \qquad (j = 1,2,\cdots,6) \tag{6.2.6}$$

这种情况下运输模型的一般形式为

目标函数：

$$Z = \sum_j \sum_i C_{ij} \cdot X_{ij} \to \min$$

$$\sum_j X_{ij} = a_i \qquad (i = 1,2,\cdots,m)$$

s.t.为：

$$\sum_i X_{ij} \leqslant b_j \qquad (j = 1,2,\cdots,n)$$

$$X_{ij} \geqslant 0 \qquad (i = 1,2,\cdots,m \quad j = 1,2,\cdots,n)$$

由平衡型运输模型和不平衡型运输模型的差别可见，条件不同，模型的形式就不同。它们虽然只是等号和不等号的差别，但所反映问题的实质是大不相同的。因此，在建立模型的时候，必须对每个符号和不等号都要进行仔细地研究与推敲，使模型反映的实质与所提出的问题一致。

6.2.2 生产能力配置问题

在上述运输问题中，仅考虑到生产地和需求地之间的运输费用，而未考虑到各地区的

木材生产费用。如果每个生产企业的生产能力为 W_i，且 $\sum_i W_j \geqslant \sum_j b_j$ 时，就必须兼顾运输费用和生产费用来统筹安排各地生产量和运输量。在原平衡型运输模型的基础上经扩充就可得到生产能力配置模型。如果当 $\sum_j W_j < \sum_j b_j$ 时，为了使生产满足需求，需要考虑运输费用、生产费用、基建投资来统筹安排各地区的生产量、运输量和在某些地区扩大生产能力的问题。该问题也可以在运输问题的基础上经扩充得到。

下面就在平衡型运输模型的基础上，经过扩充得出上述两种情况的生产能力配置模型。

6.2.2.1 当 $\sum_j W_j \geqslant \sum_j b_j$ 时的生产能力配置模型

1. 问题的提出

设有 m 个木材加工企业，生产能力分别为 W_i，共有需求单位 n 个，其需求量为 b_j，当 $\sum_i W_j \geqslant \sum_j b_j$ 时，如何统筹安排生产量和运输量，使生产费用和运输费用的总和最小。

2. 原始数据

见表 6.2.2。

表 6.2.2 生产配置模型原始数据

需求 / 生产	1	2	…	n	产量	生产能力	生产成本
1	X_{11}	X_{12}	…	X_{1n}	X_1	W_1	C_1
2	X_{21}	X_{22}	…	X_{2n}	X_2	W_2	C_2
…	…	…	…	…	…	…	…
M	X_{m1}	X_{m2}	…	X_{mn}	X_m	W_m	C_m
需求量	b_1	b_1	…	b_n			

3. 设变量和设置符号

设：X_{ij} —— 第 i 个木材加工厂供应第 j 个需求单位的木材运输量；

X_i —— 第 i 个木材加工厂的生产量。

已知：C_{ij} —— 第 i 个木材加工厂向第 j 个需求单位运输单位产品的运价；

C_i —— 第 i 个木材加工厂的生产成本。

4. 建立目标函数

由平衡型运输模型知，运输总费用为 $\sum_j \sum_i C_{ij}X_{ij}$，各单位的生产费用为 C_iX_i，总生产费用为 $\sum_i C_iX_i$，问题的评价指标是生产和运输费用的总和，故目标函数为

$$Z = \sum_j \sum_i C_{ij}X_{ij} + \sum_i C_iX_i \rightarrow \min \tag{6.2.7}$$

5. 约束条件

(1) 各木材加工厂的生产量不能超过生产能力：

$$X_i \leqslant W_i \qquad (i = 1,2,\cdots,m) \tag{6.2.8}$$

(2) 各木材加工厂的生产量必须全部运出：

$$\sum_j X_{ij} = X_i \qquad (i = 1,2,\cdots,m) \tag{6.2.9}$$

(3) 各需求单位的需求量必须得到满足：

$$\sum_i X_{ij} = b_j \qquad (j = 1,2,\cdots,n) \tag{6.2.10}$$

(4) 各木材加工厂的产品运出量和生产量非负：

$$X_{ij} \geqslant 0; \qquad X_i \geqslant 0 \qquad (i = 1,2,\cdots,m; \quad j = 1,2,\cdots,n) \tag{6.2.11}$$

式(6.2.7) ～ 式(6.2.11) 即是$\sum_i W_i \geqslant \sum_j b_j$时的线性规划模型。

6.2.2.2 当$\sum_i W_i < \sum_j b_j$时的生产能力配置模型

1. 问题的提出

由于总生产能力小于总需求量，为了满足需求，必须增加木材加工厂的生产能力。在增加生产能力时，可采用扩建原有企业的方法，也可采用新建木材加工厂的方法，或同时扩建或新建。因此，需根据现有生产能力、生产成本、投资效益、资源条件等因素综合考虑，统筹安排新建、扩建的生产能力、各企业的生产量、产品的调运量，使运输费用、基本建设投资和生产费用总额最少。

2. 原始数据

见表 6.2.3。

表 6.2.3 原始数据

	1	2	…	n	生产量	现生产能力	单位产品基建投资	生产成本	资源
1	X_{11}	X_{12}	…	X_{1n}	X_1	W_1	K_1	C_1	H_1
2	X_{21}	X_{22}	…	X_{2n}	X_2	W_2	K_2	C_2	H_2
…	…	…	…	…	…	…	…	…	…
m'	$X_{m'1}$	$X_{m'2}$	…	$X_{m'n}$	X_m	$W_{m'}$	$K_{m'}$	$C_{m'}$	$H_{m'}$
需求量	b_1	b_2	…	b_n					

3. 设变量和设置符号

设：X_{ij} —— 第 i 个木材加工地向第 j 个需求地的木材供应量；

X_i —— 第 i 个木材加工地的生产量；

Z_i —— 第 i 个木材加工地新、扩建的生产能力。

已知：K_i —— 第 i 个木材加工地新、扩建生产能力的单位基本建设投资额；

W_i —— 第 i 个木材加工地现有生产能力；

C_i —— 第 i 个木材加工地的生产成本；

H_i —— 第 i 个木材加工地其他资源的限制，如电力等；

E —— 投资效果系数。

4. 建立目标函数

因对问题的评价是按生产费用、运输费用和基建投资费用最小进行的，所以在建立目标函数时，必须考虑各项费用的可比性。该问题中，基建投资是一次性支出，与一年内的生产费用和运输费用无可比性，因而不能直接相加，必须借助于投资效果系数将基建投资换算成每年的支出，才能使各项费用具有可比性，这种经换算后的总费用称为“换算费用”。因此，采用换算费用作为系统的评价指标是合理的，以换算费用最小作为目标，其数学描述如下：

$$\min \quad Z = \sum_j \sum_i C_{ij}X_{ij} + \sum_i C_iX_i + E\sum_i K_iZ_i \tag{6.2.12}$$

式中 E—— 投资效果系数，是投资回收期的倒数，在此则表示平均每年投资占投资总额的比例。

5. 约束条件

(1) 生产能力限制，即实际生产量应小于或等于该地现有生产能力和新、扩建生产能力的总和。

$$X_i \leqslant W_i + Z_i \qquad (i = 1,2,\cdots,m') \tag{6.2.13}$$

(2) 各木材加工厂生产能力不能超过该地区资源的限制，即现有生产能力和新、扩建生产能力之和必须小于该木材加工厂每年可得到的资源量。

$$W_i + Z_i \leqslant H_i \qquad (i = 1,2,\cdots,m') \tag{6.2.14}$$

(3) 各木材加工厂的生产量必须全部运出。

$$\sum_j X_{ij} = X_i \qquad (i = 1,2,\cdots,m') \tag{6.2.15}$$

(4) 各需求单位的木材需求量必须全部满足。

$$\sum_i X_{ij} = b_j \qquad (j = 1,2,\cdots,n) \tag{6.2.16}$$

(5) 新、扩建生产能力、生产量、木材运出量为非负。

$$Z_i \geqslant 0; X_i \geqslant 0; X_{ij} \geqslant 0 \tag{6.2.17}$$

$$(j = 1,2,\cdots,m'\ ; i = 1,2,\cdots,n)$$

6.2.2.3 考虑原料供应费用的生产能力配置模型

上述模型属于加工地靠近原料产地的生产能力配置模型。由于未考虑原料供应费用，求解结果不适于远离原料产地的加工企业。对于远离原料产地的大型加工企业，在建立模型时必须考虑原料供应费用。

1. 问题的提出

现有 m 个胶合板厂，有 n 个胶合板需求单位，现有的胶合板总生产能力小于需求量，需按生产费用、基建投资、产品运输费用、原料运输费用统筹安排投资方向，以达到在满足需求的情况下，使总换算费用最省。

该问题属两阶段问题，第一阶段是由林业局向胶合板厂供应原木；第二阶段是胶合板厂向用户供应胶合板。为了满足胶合板厂的需求，必须有足够的原木资源作保证。

现假设该条件已满足且由 k 个林业局向 m 个胶合板厂供应胶合板材。

2. 原始数据

见表 6.2.4。

表 6.2.4　原始数据

胶板厂 / 林业局	1	2	…	m	原木拥有量	需求地胶合板厂	1	2	…	n	生产量	生产能力	限制	扩建生产能力	基建投资	单位成本
1			(q_{ki})		G_1	1	C_{ij}				X_1	W_1	H_1	Z_1	K_1	C_1
2		(Y_{ki})			G_2	2		X_{ij}			X_2	W_2	H_2	Z_2	K_2	C_2
…					…	…					…	…	…	…	…	…
k					G_k	m					X_m	W_m	H_m	Z_m	K_m	C_m
原木需求量	r_1X_1	r_2X_2	…	r_mX_m	…	…	b_1	b_2	…	b_n						

3. 设变量和设置符号

设：Y_{ki} —— 第 k 个林业局向 i 个胶合板厂供应的木材数量；

X_{ij} —— 第 i 个胶合板厂向第 j 个需求单位供应的胶合板数量；

X_i —— 第 i 个胶合板厂的生产量；

Z_i —— 第 i 个胶合板生产地新、扩建的生产能力。

已知：W_i —— 第 i 个胶合板生产厂现有生产能力；

C_i —— 第 i 个胶合板产地的单位生产成本；

K_i —— 第 i 个胶合板产地新、扩建单位生产能力的单位基本建设投资额；

H_i —— 对胶合板厂新、扩建的限制条件(如水源、电力等)；

q_{ki} —— 第 k 个林业局向第 i 个胶合板产地供应单位原木的单位运价；

r_i —— 第 i 个胶合板产地的单位产品木材消耗量；

C_{ij} —— 第 i 个胶合板厂向第 j 个胶合板需求地供应单位胶合板的运费；

G_k —— k 原料产地的胶合板材产量；

b_j —— 第 j 个胶合板需求地的需求量。

4. 建立目标函数

总换算费用 = 原料运输费用 + 胶合板运输费用 + 胶合板生产费用 + 投资的换算费用。

其数学表达式为

$$\min \quad Z = \sum_k \sum_i q_{ki} \cdot y_{ki} + \sum_j \sum_i C_{ij} \cdot X_{ij} + \sum_i C_i X_i + F \sum_i K_i \cdot Z_i \qquad (6.2.18)$$

5. 约束条件

(1) 原料供应方面的限制：原料产地运出的胶合板材数量应小于其生产量。

$$\sum_i Y_{ki} \leqslant G_k \qquad (k = 1,2,\cdots,n) \qquad (6.2.19)$$

(2) 胶合板生产地的原料需求量应得到满足。

$$\sum_k Y_{ki} = r_i \cdot x_i \qquad (i = 1,2,\cdots,m) \qquad (6.2.20)$$

(3) 各胶合板生产地的生产量应小于或等于现有生产能力和新、扩建生产能力的总和。

$$X_i \leqslant W_i + Z_i \qquad (i = 1,2,\cdots,m) \tag{6.2.21}$$

(4) 各胶合板生产地的总生产能力不能超过某些资源的限制，如水、电等。

$$W_i + Z_i \leqslant H_i \qquad (i = 1,2,\cdots,m) \tag{6.2.22}$$

(5) 胶合板生产地的生产量应全部运出。

$$\sum_j X_{ij} = X_i \qquad (i = 1,2,\cdots,m) \tag{6.2.23}$$

(6) 各需求地的胶合板需求量应得到满足。

$$\sum_i X_{ij} = b_j \qquad (j = 1,2,\cdots,n) \tag{6.2.24}$$

(7) 原料运输量、胶合板生产量、产品运输量和生产能力的新、扩建规模应为非负。

$$Y_{ki} \geqslant 0; X_i \geqslant 0; X_{ij} \geqslant 0; Z_i \geqslant 0 \tag{6.2.25}$$

$$(k = 1,2,\cdots,k; i = 1,2,\cdots,m; j = 1,2,\cdots,n)$$

上述模型在考虑新建生产能力时，只把拟选的新建厂地址作为生产地处理并令其现有生产能力 $W_i = 0$ 即可。

6.2.3 资源利用模型

1. 问题的提出

某建筑材料厂现有 m 种资源且每种资源拥有量为定值。现该厂生产 n 种产品，在尽可能充分利用各种资源的情况下，如何安排生产任务能使企业利润最大。

2. 原始数据

见表 6.2.5。

表 6.3.5 某建筑厂原始数据

产品 / 资源	1	2	…	n	拥有量
1	a_{11}	a_{12}	…	a_{1n}	a_1
2	a_{21}	a_{22}	…	a_{2n}	a_2
…	…	…	…	…	…
m	a_{m1}	a_{m2}	…	a_{mn}	a_m
单位利润	C_1	C_2	…	C_N	
市场需求上限	$\overline{K}_1$	$\overline{K}_2$	…	$\overline{K}_n$	
市场需求下限	$\underline{K}_1$	$\underline{K}_2$	…	$\underline{K}_n$	

3. 设变量和设置符号

设：X_j —— 各种产品的计划生产量。

已知：C_j —— 第 j 种产品的单位利润；

a_i —— 第 i 种资源的拥有量；

$\overline{K}_j$、$\underline{K}_j$ ——j 产品产量的上下限；

a_{ij} —— 生产单位 j 产品所消耗的 i 种资源量。

4. 建立目标函数

以生产利润作为评价指标，应追求最大利润。C_jX_j 为第 j 种产品的利润，总利润为 $\sum_j C_jX_j$，故目标函数为

$$\max \qquad Z = \sum_j C_jX_j \tag{6.2.26}$$

5. 约束条件

(1) 各种资源尽量充分利用，即各种资源消耗量必须小于或等于其拥有量

$$\sum_j a_{ij}X_j \leqslant a_i \qquad (i = 1,2,\cdots,m) \tag{6.2.27}$$

(2) 第 j 种产品的生产量不能低于市场需求的下限，也不能超过市场需求的上限

$$\underline{K}_j \leqslant X_j \leqslant \overline{K}_j \qquad (j = 1,2,\cdots,n) \tag{6.2.28}$$

(3) 各产品的生产量非负。

$$X_j \geqslant 0 \qquad (j = 1,2,\cdots,n) \tag{6.2.29}$$

6.2.4 下料问题和配套问题

6.2.4.1 下料问题

合理下料是企业合理利用原材料的主要措施。利用线性规划模型可以比较好地解决提高原材料利用率问题。解决材料利用问题的模型之一是下料模型，也叫裁剪模型。它可用于服装裁剪以节约布料，用于金属下料以节约金属材料等，因此是一种用途较广的模型。

1. 问题的提出

某建筑公司制材厂为提高锯材出材率，对锯材原木按材种、等级、径级、材长、稍度、弯曲度等标志分为 m 类，并对每类原木的下锯方案进行了合理的研究，画出下料图谱，同时，按计划的品种、规格得出每类原木的下料方案。在木材品种、规格计划量已经确定的情况下，如何安排进锯的原木种类和数量才能既完成计划任务又使原木的消耗量最少。

2. 原始数据表

见表 6.2.6。

表 6.2.6 某建筑公司制材厂原始数据

产品 / 原木类型	1	2	…	n	原木拥有量 / 根
1	a_{11}	a_{12}	…	a_{1n}	a_1
2	a_{21}	a_{22}	…	a_{2n}	a_2
…	…	…	…	…	…
m	a_{m1}	a_{m2}	…	a_{mn}	a_m
任务量上限	$\overline{b}_1$	$\overline{b}_2$	…	$\overline{b}_n$	
任务量下限	$\underline{b}_1$	$\underline{b}_2$	…	$\underline{b}_n$	

3. 设变量和设置符号

设：X_i —— 为完成计划所需的第 i 种原木的根数。

已知：a_{ij} —— 第 i 种类型原木按下料图所能生产的第 j 种锯材数量；

r_i —— 第 i 种类型原木的单株材积。

4. 建立目标函数

为完成锯材生产任务，r_iX_i 为第 i 类原木的消耗量(将根数换算成材积)，$\sum_i r_ix_i$ 为全部原木的消耗量，故目标函数为

$$Z = \sum_i r_iX_i \rightarrow \min \tag{6.2.30}$$

5. 约束条件为

(1) 生产量必须在计划规定范围内，即不能高于需求量的上限，也不能低于需求量的下限。

$$\underline{b}_1 \leqslant \sum_i a_{ij}X_j \leqslant \bar{b}_1 \qquad (j = 1,2,\cdots,n) \tag{6.2.31}$$

(2) 各类原木消耗量不能超过拥有量。

$$X_i \leqslant a_i \qquad (i = 1,2,\cdots,m) \tag{6.2.32}$$

(3) 原木消耗量为非负整数。

$$X_i \geqslant 0 \text{ 的整数} \qquad (i = 1,2,\cdots,m) \tag{6.2.33}$$

该模型适用于原料资源丰富的制材企业，对于建筑企业的木材加工车间，因受原料数量限制，只能在给定原木材种、规格的基础上，合理安排下料方案，达到既完成生产任务，又提高木材利用率的目的。这种情况就需要对每种原木拟定几套下料方案，分别确定每种原木的各种进锯方案的下料数量。该情况的原始数据表见表 6.2.7。

表 6.2.7　原木的原始数据

		1	2	…	n	原木拥有量
1	1	$a_{1\ 11}$	$a_{1\ 12}$	…	$a_{1\ 1n}$	D_1
	2	$a_{1\ 21}$	$a_{1\ 22}$	…	$a_{1\ 2n}$	
	…	…	…	…	…	
	l_1	$a_{1\ l11}$	$a_{1\ l12}$	…	$a_{1\ l1n}$	
2	1	$a_{2\ 11}$	$a_{2\ 12}$	…	$a_{2\ 1n}$	D_2
	2	$a_{2\ 21}$	$a_{2\ 22}$	…	$a_{2\ 2n}$	
	…	…	…	…	…	
	l_2	$a_{2\ l21}$	$a_{2\ l22}$	…	$a_{2\ l2n}$	
…	…	…	…	…	…	…
m	1	$a_{m\ 11}$	$a_{m\ 12}$	…	$a_{m\ 1n}$	D_m
	2	$a_{m\ 21}$	$a_{m\ 22}$	…	$a_{m\ 2n}$	
	…	…	…	…	…	
	l_m	$a_{m\ lm1}$	$a_{m\ lm2}$	…	$a_{m\ lmn}$	
任务量下限		$\bar{b}_1$	$\bar{b}_2$	…	$\bar{b}_n$	
任务量上限		$\underline{b}_1$	$\underline{b}_2$	…	$\underline{b}_n$	

设：X_{ik} —— 第 i 种原木按第 k 种方案下料的根数。

已知：r_i —— 第 i 类原木的单株材积；

a_{ikj} —— 第 i 种原木，按第 k 种下料方案，生产第 j 种锯材产品的数量；

D_i —— 第 i 种原木拥有量；

l_i —— 第 i 种原木的下料方案数。

目标函数为：

$$Z = \sum_{i} \sum_{k=1}^{l_i} r_i \cdot X_{ik} \to \min \tag{6.2.34}$$

s.t.为：① 产品生产量在计划规定的范围内

$$\underline{b}_j \leqslant \sum_{i} \sum_{k=1}^{l_i} a_{ikj} \cdot X_{ik} \leqslant b_j \qquad (j = 1,2,\cdots,n) \tag{6.2.35}$$

② 各种原木的消耗量不能超过现有的原木量

$$\sum_{k=1}^{l_i} X_{ik} \leqslant D_i \qquad (i = 1,2,\cdots,m) \tag{6.2.36}$$

③ 各类原木、各种进锯方案的下料原木根数非负且为整数

$$X_{ik} \geqslant 0 \text{ 的整数} \tag{6.2.37}$$

下料模型尚广泛地用于解决建筑项目钢筋、塑钢窗、玻璃的下料等问题。

6.2.4.2 配套问题

1. 问题的提出

产品配套问题是指当所生产的产品需配套使用时的一种下料计划安排方法。如某建筑公司木制品厂生产门、窗、家具等产品，需不同规格板方材 m 种，且这些板方材必须配套使用，才能组装成一个成品门窗，其比例为 $p_1 : p_2 : \cdots : p_m$。当板方材原料规格为 n 种且原料数量一定时，如何安排下料能使所生产的门窗数量最多。

2. 原始数据

见表 6.2.8。

表 6.2.8 配套问题原始数据

零件 / 原料	1	2	…	m	拥有量
1	a_{11}	a_{12}	…	a_{1m}	a_1
2	a_{21}	a_{22}	…	a_{2m}	a_2
…	…	…	…	…	…
n	a_{n1}	a_{n2}	…	a_{nm}	a_n
比例	p_1	p_2	…	p_m	

3. 设变量和设置符号

设： X_i —— 第 i 种原料的下料数量。

已知：a_{ij} —— 第 i 种原料可加工第 j 种构件的数量；

a_i —— 第 i 种原料的拥有量；

$p_1 : p_2 : \cdots : p_m$ —— 为各种构件的配套比。

4. 建立目标函数

因为 $\sum_i a_{ij} X_i (j = 1,2,\cdots,m)$ 为第 j 种构件的数量，它与构成门窗比例之商就是门窗

的套数。因此，门窗套数为

$$\frac{\sum_i a_{i1}X_i}{p_1} = \frac{\sum_i a_{i2}X_i}{p_2} = \cdots = \frac{\sum_i a_{im}X_i}{p_m}$$

其中任一个都可体现出门窗套数的多少，故目标函数为

$$Z = \frac{\sum_i a_{i1}X_i}{p_1} \to \max \tag{6.2.38}$$

5. 约束条件

(1) 保证门窗生产的配套关系

$$\frac{\sum_i a_{i1}X_i}{p_1} = \frac{\sum_i a_{i2}X_i}{p_2} = \cdots = \frac{\sum_i a_{im}X_i}{p_m} \tag{6.2.39}$$

(2) 原料的消耗量不能超过拥有量

$$X_j \leqslant a_i \qquad (i = 1,2,\cdots,n) \tag{6.2.40}$$

(3) 原料下料数量非负：

$$X_i \geqslant 0 \qquad (i = 1,2,\cdots,n) \tag{6.2.41}$$

6.2.4.3 另一类配套问题

另一类配套问题是从设备分配角度来研究的。

1. 问题提出

某建筑设备厂生产卷扬机，卷扬机由 n 种零件组成，考虑配件供应，各零件的生产量保证 $p_1 : p_2 : \cdots : p_n$ 的比例关系。现工厂有 m 种设备，每种设备 a_i 台，不同设备生产不同零件时效率不同(如不能生产，效率为零)，如何安排设备，才能在充分利用设备的情况下，使卷扬机生产量最大。

2. 原始数据

见表 6.2.9。

表 6.2.9 建筑设备厂卷扬机原始数据

零件 / 设备	1	2	…	n	拥有量
1	a_{11}	a_{12}	…	a_{1n}	a_1
2	a_{21}	a_{22}	…	a_{2n}	a_2
…	…	…	…	…	…
m	a_{m1}	a_{m2}	…	a_{mn}	a_m
比例	p_1	p_2	…	p_n	

3. 设变量和设置符号

设： X_{ij} —— 第 i 种设备加工第 j 种零件的台数。

已知： $p_1 : p_2 : \cdots : p_n$ 为零件的配套比；

a_i —— 第 i 种设备的拥有量；

a_{ij} —— 第 i 种设备加工第 j 种零件的效率。

4. 建立目标函数

绞盘机生产量最多，意味着配套数最多，故目标函数为

$$\max \qquad Z = \frac{\sum_i a_{ij} X_{ij}}{p_j} \qquad (\text{其中 } j \in N,\ N = 1,2,\cdots,n) \tag{6.2.42}$$

5. 约束条件

(1) 设备要尽量充分利用

$$\sum_j X_{ij} \leqslant a_i \qquad (i = 1,2,\cdots,m) \tag{6.2.43}$$

(2) 产品要配套

$$\frac{\sum_i a_{i1} X_{i1}}{p_1} = \frac{\sum_i a_{i2} X_{i2}}{p_2} = \cdots = \frac{\sum_i a_{in} X_{in}}{p_n} \tag{6.2.44}$$

(3) 变量非负

$$X_{ij} \geqslant 0 \qquad (i = 1,2,\cdots,m;\ j = 1,2,\cdots,n) \tag{6.2.45}$$

6.2.5 生产任务分配问题

生产任务的分配基本分两种形式，一种是生产任务在各基层单位的分配，如总厂的生产任务在各分厂间的分配，各分厂的生产任务在各车间的分配等；另一种是生产任务确定之后，制定生产日历计划等。

6.2.5.1 生产任务在基层单位的分配

1. 问题提出

某构件公司向所属的 n 个构件厂分配各种构件的生产任务，如何根据各构件厂的生产能力，生产成本、资源量等条件分配生产任务，才能在完成全公司任务的同时，使全公司生产总成本最低。

2. 原始数据

见表 6.2.10。

表 6.2.10 某构件公司构件厂原始数据

产品 / 构件厂	1	2	…	m	资源量
1	C_{11}, W_{11} X_{11}, K_{11}	C_{12}, W_{12} X_{12}, K_{12}	…, … …, …	C_{1m}, W_{1m} X_{1m}, K_{1m}	b_1
…	…, … …, …	…, … …, …	…, … …, …	…, … …, …	…
n	C_{n1}, W_{n1} X_{n1}, K_{n1}	C_{n2}, W_{n2} X_{n2}, K_{n2}	…, … …, …	C_{nm}, W_{nm} X_{nm}, K_{nm}	b_n
生产任务	a_1	a_2	…	a_m	

3. 设变量和设置符号

设：X_{ij} —— 第 i 个构件厂生产第 j 种产品的任务数。

已知：C_{ij} —— 第 i 个构件厂生产第 j 种产品的成本；

W_{ij} —— 第 i 个构件厂生产第 j 种产品的生产能力；

b_i —— 第 i 个构件厂资源的限制(如水泥、钢材的供应量);

K_{ij} —— 第 i 个构件厂生产第 j 种产品资源(水泥、钢材等)的消耗定额;

a_j —— 构件公司第 j 种产品任务量。

4. 建立目标函数

$$Z = \sum_j \sum_i C_{ij} X_{ij} \rightarrow \min \tag{6.2.46}$$

5. 约束条件

(1) 保证完成公司的各项生产任务

$$\sum_i X_{ij} \geqslant a_j \qquad (j = 1,2,\cdots,m) \tag{6.2.47}$$

(2) 各构件厂相应产品生产能力的限制

$$X_{ij} \leqslant W_{ij} \qquad (i = 1,2,\cdots,m;j = 1,2,\cdots,m) \tag{6.2.48}$$

(3) 资源的限制

$$\sum K_{ij} \cdot X_{ij} \leqslant b_i \qquad (i = 1,2,\cdots,n) \tag{6.2.49}$$

(4) 变量的非负限制

$$X_{ij} \geqslant 0 \qquad (i = 1,2,\cdots,n;j = 1,2,\cdots,m) \tag{6.2.50}$$

6.2.5.2 制定日历计划问题

1. 问题的提出

因构件生产是季节性较强的生产,不同月份各种生产的作业成本不同,结合企业的生产实际,统筹安排各月生产量使总成本最低是制定年度生产计划的任务。

假设:为使生产计划稳妥地完成,该厂要求"十一"前完成计划生产量的 80%,且 4 月份 ~ 6 月份黄金季节要完成某些品种产品任务的 70%。如何根据各月的生产成本、构件需求及全年生产能力等条件合理安排各月生产量使总生产成本最低。

2. 原始数据

见表 6.2.11。

表 6.2.11 原始数据

产品 \ 月份	1	2	3	…	11	12	任务量
1	W_{11} C_{11}	W_{12} C_{12}	W_{13} C_{13}	… …	$W_{1,11}$ $C_{1,11}$	$W_{1,12}$ $C_{1,12}$	a_1
…	… …	… …	… …	… …	… …	… …	… …
n	W_{n1} C_{n1}	W_{n2} C_{n2}	W_{n3} C_{n3}	… …	$W_{n,11}$ $C_{n,11}$	$W_{n,12}$ $C_{n,12}$	a_n

3. 设变量和设置符号

设: X_{is} ——s 月份 i 类产品的计划生产量。

已知:W_{is} ——i 类产品 s 月的生产能力或资源限制;

C_{is} ——s 月 i 种产品的生产成本。

4. 建立目标函数

现以年生产总成本最低为目标：

$$Z = \sum_{i}\sum_{s} C_{is}X_{is} \to \min \tag{6.2.51}$$

5. 约束条件

(1) 必须完成各种产品全年生产任务。

$$\sum_{s=1}^{12} X_{is} \geqslant a_i \qquad (i = 1,2,\cdots,n) \tag{6.2.52}$$

(2) 各月份产品产量不能超过本月生产能力或资源的限制。

$$X_{is} \leqslant W_{is} \qquad (i = 1,2,\cdots,n;s = 1,2,\cdots,12) \tag{6.2.53}$$

(3) 第二季度某些产品产量要超过总任务量的 70% 以上。

$$\sum_{s=4}^{6} X_{is} \geqslant 0.7a_i \qquad (依照生产产品确定) \tag{6.2.54}$$

(4) 10 月份前要完成各种产品任务量的 80% 以上。

$$\sum_{s=1}^{9} X_{is} \geqslant 0.8\sum_{i} a_i \tag{6.2.55}$$

(5) 变量非负。

$$X_{is} \geqslant 0 \qquad (i = 1,2,\cdots,n;s = 1,2,\cdots,12) \tag{6.2.56}$$

6.2.6 人员分配问题

1. 问题提出

某工组有 n 名工人，现准备从事 n 项工作，每种工作只能由一个人承担，每人只能承担一项工作。当不同的人从事各种工作的效益不同时，如何分配人员，才能使全工组总收益最大。

2. 原始数据

见表 6.2.12。

表 6.2.12 原始数据

工人＼工作	1	2	…	…	n
1	C_{11}	C_{12}	…	…	C_{1n}
2	C_{21}	C_{22}	…	…	C_{2n}
…	…	…	…	…	…
n	C_{n1}	C_{n2}	…	…	C_{nn}

3. 设变量和设置符号

设：X_{ij}——第 i 个人从事第 j 种工作，且

$$X_{ij} = \begin{cases} 0, & 当第\ i\ 个人不从事第\ j\ 种工作时 \\ 1, & 当第\ i\ 个人从事第\ j\ 种工作时 \end{cases}$$

已知：C_{ij}——第 i 个人从事第 j 种工作的收益。

4. 建立目标函数

现以总收益最大为目标

$$Z = \sum_{j}\sum_{i} C_{ij}X_{ij} \rightarrow \max \tag{6.2.57}$$

5. 约束条件

(1) 每个人只能从事一项工作。

$$\sum_{j} X_{ij} = 1 \quad (i = 1,2,\cdots,n) \tag{6.2.58}$$

(2) 每项工作只能由一个人承担。

$$\sum_{i} X_{ij} = 1 \quad (j = 1,2,\cdots,n) \tag{6.2.59}$$

(3) 每个人从事某项工作只有做与不做两种可能。

$$X_{ij} = \begin{cases} 0, \text{当第 } i \text{ 个人不从事第 } j \text{ 种工作时} \\ 1, \text{当第 } i \text{ 个人从事第 } j \text{ 种工作时} \end{cases} \tag{6.2.60}$$

6.2.7 经济系统建立目标函数和约束条件应注意的问题

建立线性规划模型时，目标可根据系统的要求来确定，约束方程可依据环境和内部条件对系统的限制来确定。然而，目标与约束并不是孤立的，而是密切相关的，尤其是经济管理系统，正确处理目标函数与约束条件的关系尤为重要。

6.2.7.1 目标的选择

1. 可供经济管理系统选择的系统目标

(1) 运营效果最大；

(2) 运营耗费最小；

(3) 运营效率最高；

(4) 换算费用最低；

(5) 满意程度最高；

(6) 与即定目标差距最小等。

2. 选择目标函数应注意的问题

(1) 在选择目标函数时，深刻理解所选指标的内涵是十分重要的，因为各种指标均有一定的选用领域与范围，均是从某一角度对系统的评价，单纯地追求一种指标使其达到最大或最小，可能造成系统运行的畸形，使结果脱离实际。如总产值指标是评价一个经济系统工作总量的指标，该指标包含了物化劳动的转移价值，如选择该项指标作为目标函数，可能会造成单纯追求产值而忽视经济效益的倾向。

(2) 在众多可供选择的目标函数中，依研究目的确定主要指标和次要指标是目标函数选择中的第二个应考虑的问题。主要指标反映了与研究目的直接相关的，对系统影响较大的指标，应该予以重视，但次要指标对主要指标有补充、限制等作用，也是不能忽略的，忽略次要指标，以主要指标最大或最小为目标函数，也可能造成系统运行畸形，而得出片面的结论。

(3) 目标函数组成要素中的量纲和数量级问题是经常被忽视的问题之一。在经济管理系统中，不同量纲的数据，无可加性，简单相加无实际经济意义，因此不能作为目标函数。而相同量纲，但数量级不同的数据，即使有可比性也不能作为目标函数，因为这样组成的目标函数，数量级小的部分作用甚小，达不到预期效果。如用生产费用与运输费用之和最小为目标，如果生产费用为亿元级，而运输费用为百万元级，二者之和是亿元级数据，此

时目标中运输费用仅占1%，其实质等于未考虑运输费用。

(4) 尽量采用综合指标作为目标函数。当系统有多种评价指标时，为更确切地反映系统目标，可按各指标的重要性采用加权方法建立综合指标作为目标函数。如采用利润这一综合性指标，比单独采用产值最大或成本最低作为目标要好。

(5) 注意处理线性与非线性目标函数的关系。经济系统目标往往是非线性的，建立线性目标函数是假设变量间呈线性关系。如果非线性函数在某区间内可用线性函数替代或经适当简化将其化为线性函数又不影响精度时，该线性目标函数在有限区间内是可用的，否则只能用非线性规划模型解决。这就是说线性规划有其假设前提。

6.2.7.2 约束条件的确定

1. 系统约束条件的分类

可供经济管理系统选择的约束条件有以下几类：

(1) 自然资源类约束条件，如森林、矿产储藏量、土地面积等；

(2) 企业资源类约束条件，如职工人数、生产能力、设备数量、工时、能源等；

(3) 生产特性类约束条件，如工序的平衡与工艺衔接，生产的特殊规律等；

(4) 社会、法律及基本经济规律类约束，如市场需求、技术经济条件、法律规定等；

(5) 其他类。

2. 建立约束条件时应注意的问题

约束条件是为实现系统目标而对系统的限制。因此建立约束条件时必须：

(1) 依据系统目标，研究限制条件。如以总产值为目标，则可能产生忽视经济效益的缺点，为实现系统目标，必须考虑该指标的缺陷，在约束条件中对缺陷加以限制。如，在约束中加入成本上限的限制则可弥补总产值指标的不足。以次要指标作为约束条件，可弥补仅以主要指标为目标的缺陷，在约束中考虑低数量级指标，在目标中只考虑高数量级指标，可解决不同数量级指标的组合问题等。

(2) 根据所研究系统的具体情况，有些自然资源和企业资源可能不对系统构成实质限制，可省略该约束以减少约束数量，简化计算。如劳动力约束，在我国可能不是约束，但在人力资源短缺的国家或地区，则可能构成约束。

(3) 结构性约束。结构是经济系统的重要特征，结构性失衡往往是影响系统稳定性的重要因素，忽略结构性约束，将对系统造成不良后果，如劳动力总量在我国的一些地区不构成实质性约束，但不等于劳动力构成对系统无影响，如技术人才、性别比例等可能是系统的重要约束，因此在建立约束条件时，既要考虑总量，又要考虑其结构的限制。

总之，在建立线性规划模型时，应综合考虑目标与约束条件的关系，反复推敲，切忌将二者孤立起来，单独研究。一般情况下，应先确定目标，然后再依据目标确定约束条件，所建的模型才更可靠。

6.3 国民经济综合平衡最优化模型

国民经济综合平衡模型仅解决了部门之间相互协调、按比例发展的问题，并未解决国民经济以最小的消耗达到最优发展的问题。因此，以国民经济综合平衡模型为基础，合理地安排国民经济各部门之间的比例关系，使国民经济以最小的消耗达到最优发展是制定经济和

社会发展计划的主要任务，建立国民经济综合平衡最优化模型就是解决该问题的途径之一。

从理论上讲，建立如下模型即可实现国民经济计划的最优化：

$$\begin{aligned} \min \quad & Z = A_L X \\ \text{s.t.} \quad & (I - A)X = Y \\ & X \geqslant 0 \end{aligned} \tag{6.3.1}$$

式中 A_L——劳动消耗系数，其他符号见第 5 章。

该模型是以投入产出表的基本平衡关系为约束，以劳动消耗量最小为目标实现的国民经济计划最优化。但仔细研究会发现，当约束$(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X} = \boldsymbol{Y}$，以最终产品 Y 作为外生变量时，X 就惟一地被确定了，劳动量也就随之惟一地被确定了，这与国民经济综合平衡模型并无不同。如果不预先确定最终产品，则可能造成某些最终产品的最优值为零，使模型失去经济意义，因此需研究国民经济综合平衡最优化模型的建立方法。

6.3.1 对各产品部门中不同生产方式进行选择的最优化模型

国民经济综合平衡模型的产品部门中，同类产品的生产方式是不同的，因而消耗也不相同。如果将模型中的产品部门按生产方式展开，则 $\boldsymbol{A}$ 矩阵变为：

$$\boldsymbol{A} = \left\{\begin{array}{cccc|cccc|c|cccc} a_{11}^1 & a_{11}^2 & \cdots & a_{11}^r & a_{12}^1 & a_{12}^2 & \cdots & a_{12}^r & \cdots & a_{1n}^1 & a_{1n}^2 & \cdots & a_{1n}^r \\ a_{21}^1 & a_{21}^2 & \cdots & a_{21}^r & a_{22}^1 & a_{22}^2 & \cdots & a_{22}^r & \cdots & a_{2n}^1 & a_{2n}^2 & \cdots & a_{2n}^r \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1}^1 & a_{n1}^2 & \cdots & a_{n1}^r & a_{n1}^1 & a_{n2}^2 & \cdots & a_{n2}^r & \cdots & a_{nn}^1 & a_{nn}^2 & \cdots & a_{nn}^r \end{array}\right\}$$

式中 r——生产方式的类型数。

这样$(\boldsymbol{I} - \boldsymbol{A})$的第 j 列可写成$(I_j - A_j^*)$，即：

$$(\boldsymbol{I}_j - \boldsymbol{A}_j^*) = \left\{\begin{array}{cccc} -a_{1j}^1 & -a_{1j}^2 & \cdots & -a_{1j}^r \\ -a_{2j}^1 & -a_{2j}^2 & \cdots & -a_{2j}^r \\ \cdots & \cdots & \cdots & \cdots \\ 1-a_{jj}^1 & 1-a_{jj}^2 & \cdots & 1-a_{jj}^r \\ \cdots & \cdots & \cdots & \cdots \\ -a_{nj}^1 & -a_{nj}^2 & \cdots & -a_{nj}^r \end{array}\right\}$$

由此投入产出模型的平衡关系可写成：

$$\sum_{j=1}^{n}\sum_{s=1}^{r}(I_j - A_j^s)X_j^s \geqslant Y$$

根据这种思想，可建立第 j 部门第 s 种生产方式有生产能力限制的最优计划模型，以达到在国民经济综合平衡的基础上以最小的劳动消耗选择不同生产方式的目的。

模型的一般形式如下：

$$Z = \sum_j \sum_s a_s X_j^* \to \min \tag{6.3.2}$$

s.t. (1) 平衡关系，最终产品需求得到满足

$$\sum_j \sum_s (I_j - A_j^s)X_j^s \geqslant Y \tag{6.3.3}$$

(2) 第 j 部门第 s 种生产能力的限制

$$\boldsymbol{X}_j^s \leqslant d_j^s \qquad (i = 1,2,\cdots,n;s = 1,2,\cdots,r) \tag{6.3.4}$$

(3) 非负条件

$$\boldsymbol{A}_j^* \geqslant 0 \qquad (j = 1,2,\cdots,n;s = 1,2,\cdots,r) \tag{6.3.5}$$

该模型当已知 Y 时，即可得出最优解。在实际应用中，尚可根据实际需要加入其他约束。

6.3.2 考虑除平衡关系以外限制因素的最优化模型

为了使国民经济获得最优发展，除了平衡和比例关系之外，还有很多其他的限制条件，如自然资源、劳动力、资金等。考虑这些因素限制的最优化模型为

目标：　　max　　$Z = \sum_i Y_i$ 　　(6.3.6)

约束：

(1) 部门间产品的平衡关系

$$(I - A)X \geqslant Y \tag{6.3.7}$$

(2) 劳动力资源的限制

$$A_L X \leqslant L \tag{6.3.8}$$

(3) 某种短缺资源的限制

$$DX \leqslant W \tag{6.3.9}$$

(4) 某种自然资源的限制

$$CX \leqslant K \tag{6.3.10}$$

(5) 资金的限制

$$EX \leqslant F \tag{6.3.11}$$

(6) 变量非负

$$X \geqslant 0,\ Y \geqslant 0 \tag{6.3.12}$$

式中　A_L —— 产品的劳动消耗系数；

L —— 劳动力资源量；

D —— 对某种短缺资源的完全消耗系数；

W —— 短缺资源的拥有量；

C —— 对某种自然资源的完全消耗系数；

K —— 自然资源计划期可供使用量；

E —— 产品生产的资金占用系数；

F —— 计划期的资金拥有量。

6.3.3 某些部门生产能力有限制时的国民经济最优计划模型

国民经济的发展要求有计划、按比例，一些部门的发展必然会促进其他部门的发展，但也受其他部门的制约。当国民经济部门划分成有生产能力限制和无生产能力限制二种类型时，平衡关系 $AX + Y \leqslant X$ 可写成如下分块形式

$$\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \cdot \left[\begin{array}{c} X^{(1)} \\ \hline X^{(2)} \end{array}\right] + \left[\begin{array}{c} Y^{(1)} \\ \hline Y^{(2)} \end{array}\right] = \left[\begin{array}{c} X^{(1)} \\ \hline X^{(2)} \end{array}\right]$$

式中 A_{11} —— 有生产能力限制的产品部门对有生产能力限制的产品部门的消耗子矩阵；

A_{21} —— 无生产能力限制的产品部门对有生产能力限制的产品部门的消耗子矩阵；

A_{12} —— 有生产能力限制的产品部门对无生产能力限制的产品部门的消耗子矩阵；

A_{22} —— 无生产能力限制的产品部门对无生产能力限制的产品部门的消耗子矩阵。

平衡关系可进一步改写成

$$\boldsymbol{A}_{11}\boldsymbol{X}^{(1)} + \boldsymbol{A}_{12}\boldsymbol{X}^{(2)} + \boldsymbol{Y}^{(1)} \leqslant \boldsymbol{X}^{(1)}$$

$$\boldsymbol{A}_{21}\boldsymbol{X}^{(1)} + \boldsymbol{A}_{22}\boldsymbol{X}^{(2)} + \boldsymbol{Y}^{(2)} \leqslant \boldsymbol{X}^{(2)}$$

故最优化模型[①]为

目标： max $$Z = \sum_i Y_i^{(1)} + \sum_i Y_i^{(2)} \tag{6.3.13}$$

s.t. (1) 平衡关系

$$\boldsymbol{A}_{11}\boldsymbol{X}^{(1)} + \boldsymbol{A}_{12}\boldsymbol{X}^{(2)} + \boldsymbol{Y}^{(1)} \leqslant \boldsymbol{X}^{(1)} \tag{6.3.14}$$

$$\boldsymbol{A}_{21}\boldsymbol{X}^{(1)} + \boldsymbol{A}_{22}\boldsymbol{X}^{(2)} + \boldsymbol{Y}^{(2)} \leqslant \boldsymbol{X}^{(2)} \tag{6.3.15}$$

(2) $X^{(1)}$ 类产品的生产能力限制

$$\boldsymbol{X}^{(1)} \leqslant \boldsymbol{W} \tag{6.3.16}$$

(3) 变量非负

$$\boldsymbol{X}_i^{(1)} \geqslant 0 \qquad \boldsymbol{X}_i^{(2)} \geqslant 0$$

$$\boldsymbol{Y}_i^{(1)} \geqslant 0 \qquad \boldsymbol{Y}_i^{(2)} \geqslant 0 \tag{6.3.17}$$

6.3.4 国民经济综合平衡实用最优化模型

6.3.4.1 逾期使用产品和最终产品的划分

国民经济综合平衡模型的基本关系为 $\boldsymbol{AX} + \boldsymbol{Y} = \boldsymbol{X}$，其基本涵义为本期生产消耗和最终产品之和等于本期社会总产品，或本期所生产的产品全部用于本期生产性消耗和作为最终产品。这就明显地存在一个假设，即本期所生产的产品在本期内即可全部用于消耗。但事实上，国民经济各部门的产品中有相当一部分产品本期不能提供本期使用。如本期生产的粮食产品，只有一部分用于本期，另一部分只能用于下期，这就造成基本平衡关系 $\boldsymbol{AX} + \boldsymbol{Y} = \boldsymbol{X}$ 不能真正反映国民经济各部门的平衡。因此需考虑产品的逾期使用问题，对上式进行修正。

现将国民经济各部门的产品分为两类：一类是由于某种原因，生产过程较长并比较固定，而本期生产主要是供下期使用的产品，这类产品称为逾期使用产品；另一类是本期生产主要供本期使用的产品，这类产品称之为非逾期使用产品。

假定国民经济前 K 个部门是逾期使用产品部门，则其本期生产消耗和供本期使用的最终产品是上期所生产的产品，也就是上期所生产的产品一部分用作本期的生产消耗，一

① 该模型根据《经济数学与模型》一书编写

部分作为本期的最终产品(而不是本期生产的最终产品)。

在此假设下,前 K 个部门的平衡关系变为:

$$a_{i1}\boldsymbol{X}_1 + a_{i2}\boldsymbol{X}_2 + \cdots + a_{in}\boldsymbol{X}_n + \boldsymbol{Z}_i = \boldsymbol{X}_i^0 \quad (i = 1,2,\cdots,k) \tag{6.3.18}$$

式中 $\boldsymbol{Z}_i$ —— 为 i 部门供本期使用的最终产品;

$\boldsymbol{X}_i^0$ —— 为 i 部门上期的总产品。

$\boldsymbol{Y}_i - \boldsymbol{Z}_i = \boldsymbol{X}_i - \boldsymbol{X}_i^0 = \Delta\boldsymbol{X}_i$ 为本期不能使用的产品增量,它是本期生产的最终产品与可供本期使用的最终产品之差。

其他部门的平衡关系为

$$a_{i1}\boldsymbol{X}_1 + a_{i2}\boldsymbol{X}_2 + \cdots + a_{in}\boldsymbol{X}_n + \boldsymbol{Z}_i = \boldsymbol{X}_i \quad (i = k+1,k+2,\cdots,n) \tag{6.3.19}$$

其中,$\boldsymbol{Z}_i = \boldsymbol{Y}_i$,即非逾期使用产品中可供本期使用的最终产品与本期生产的最终产品相等。

该修正后的平衡方程用矩阵形式表示为

$$\boldsymbol{AX} + \boldsymbol{Z} = \tilde{\boldsymbol{X}} \tag{6.3.20}$$

式中

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \qquad \boldsymbol{X} = \begin{Bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{Bmatrix}$$

$$\boldsymbol{Z} = \begin{Bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{Bmatrix} \qquad \tilde{\boldsymbol{X}} = \begin{Bmatrix} X_1^0 \\ X_2^0 \\ \vdots \\ X_k^0 \\ X_{k+1} \\ X_{k+2} \\ \vdots \\ X_n \end{Bmatrix}$$

上式是在假设逾期使用产品全部在第二年使用,而非逾期使用产品全部用于本期的情况下得出的。实际上,逾期使用产品中仍有部分可在本期使用,非逾期使用产品中尚有一部分本期不能使用。因此 $\tilde{X}$ 实际上可由 X^0 和 X 两部分各乘以一个逾期使用率表示,即

$$\tilde{\boldsymbol{X}} = \hat{\boldsymbol{S}}\boldsymbol{X}^0 + (\boldsymbol{I} - \hat{\boldsymbol{S}})\boldsymbol{X} \tag{6.3.21}$$

式中

$$\hat{\boldsymbol{S}} = \begin{bmatrix} S_1 & 0 & \cdots & 0 \\ 0 & S_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & S_n \end{bmatrix}$$

S_i 为逾期使用率,它表示上期产品逾期使用的百分比($0 \leqslant \boldsymbol{S} \leqslant 1$),$\hat{S}X^0$ 表示上期产品的逾期使用量;$(I - \hat{S})X$ 表示本期产品可供本期使用的数量;$\hat{S}X^0$ 与 $(I - \hat{S})\boldsymbol{X}$ 之和为本期实际可使用的产品数量,即 $\tilde{X}$。

因而平衡关系可写为

$$\boldsymbol{AX} + \boldsymbol{Z} = \hat{\boldsymbol{S}}\boldsymbol{X}^0 + (\boldsymbol{I} - \hat{\boldsymbol{S}})\boldsymbol{X} \tag{6.3.22}$$

本期可供使用的最终产品按使用方向又进一步划分为生产性投资(由基本建设$\boldsymbol{BW}$[①]和增加生产部门流动资金$\boldsymbol{Z}_1$两部分构成)、非生产性投资(积累)$\boldsymbol{Z}_2$、居民个人消费$\boldsymbol{Z}_3$、社会集体消费$\boldsymbol{Z}_4$、出口$\boldsymbol{Z}_5$、进口$\boldsymbol{Z}_6$。故国民经济综合平衡模型的基本平衡关系最后可写成

$$\boldsymbol{AX} + \boldsymbol{BW} + \boldsymbol{Z}_1 + \boldsymbol{Z}_2 + \boldsymbol{Z}_3 + \boldsymbol{Z}_4 + \boldsymbol{Z}_5 - \boldsymbol{Z}_6 = \hat{\boldsymbol{S}}\boldsymbol{X}^0 + (\boldsymbol{I} - \hat{\boldsymbol{S}})\boldsymbol{X} \tag{6.3.23}$$

式中

$$\boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \quad \boldsymbol{W} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{bmatrix} \quad \boldsymbol{Z}_1 = \begin{bmatrix} Z_{11} \\ Z_{21} \\ \vdots \\ Z_{n1} \end{bmatrix}$$

$$\boldsymbol{Z}_2 = \begin{bmatrix} Z_{12} \\ Z_{22} \\ \vdots \\ Z_{n2} \end{bmatrix} \quad \boldsymbol{Z}_3 = \begin{bmatrix} Z_{13} \\ Z_{23} \\ \vdots \\ Z_{n3} \end{bmatrix} \quad \boldsymbol{Z}_4 = \begin{bmatrix} Z_{14} \\ Z_{24} \\ \vdots \\ Z_{n4} \end{bmatrix} \quad \boldsymbol{Z}_5 = \begin{bmatrix} Z_{15} \\ Z_{25} \\ \vdots \\ Z_{n5} \end{bmatrix} \quad \boldsymbol{Z}_6 = \begin{bmatrix} Z_{16} \\ Z_{26} \\ \vdots \\ Z_{n6} \end{bmatrix}$$

6.3.4.2 国民经济综合平衡最优化模型的约束分析

国民经济综合平衡最优化模型的约束,总的来说可分为三组。

1. 生产要素约束组

它表明生产要素对扩大生产规模的限制。生产要素由劳动对象、劳动手段和劳动力组成,任何生产过程均受三者的限制。

(1) 劳动对象约束。生产部门的劳动对象包括两种,一种是自然资源如森林、矿山等,另一种则是本部门和其他部门的产品。因此自然资源的限制和国民经济各部门间的基本平衡关系就是劳动对象约束条件,其形式如下:

① 基本平衡关系约束

$$\boldsymbol{AX} + \boldsymbol{BW} + \boldsymbol{Z}_1 + \boldsymbol{Z}_2 + \boldsymbol{Z}_3 + \boldsymbol{Z}_4 + \boldsymbol{Z}_5 - \boldsymbol{Z}_6 = \hat{\boldsymbol{S}}\boldsymbol{X}^0 + (\boldsymbol{I} - \hat{\boldsymbol{S}})\boldsymbol{X} \tag{6.3.24}$$

② 现有自然资源的限制

$$X_{it} < h_{it}(t = 1,2,\cdots,p) \tag{6.3.25}$$

式中 i_t —— 依靠自然资源的生产部门;

h_{it}——i_t 部门可供当期使用的自然资源。

(2) 劳动手段约束。劳动手段主要指生产能力而言。生产能力对部门生产的限制是劳动手段约束的集中表现。部门生产能力是由劳动手段的实物构成决定的,因此以部门本期产品所占用的固定资产小于或等于本期所拥有的固定资产即可反映出生产能力的限制。

设:K_i 为 i 部门固定资产占用系数。其涵义为 i 部门生产单位产品所占用的固定资产

---

① 因不同部门基本建设投资实物构成不同,为了反映各部门所进行的基本建设数量与用作基本建设的最终产品之间的联系,有必要进行详细评价,表示本年度对部门的基建投资额,为部门单位基建投资对产品的完全需要量,则各部门在基建投资方面对部门产品的需要量为 L 用矩阵表示则为 $\boldsymbol{BW}$。

数量，其计算公式为：

$$K_k = \frac{\phi_i}{X_i} \quad (i = 1,2,\cdots,n) \tag{6.3.26}$$

式中 ϕ_i ——i 部门占用的固定资产；

X_i ——i 部门的总产品。

所以，i 部门产品生产的固定资产占用量为：

$$\phi_i = K_i \cdot X_i \quad (i = 1,2,\cdots,n) \tag{6.3.27}$$

i 部门固定资产的拥有量采用部门平均水平，它由两部分组成，一部分为上期结余，一部分是基本建设投资的动用部分。

设：ϕ_i^0 ——i 部门上期末固定资产结余额；

C_i ——i 部门基建投资动用系数；

W_i ——i 部门基建投资量；

$C_i W_i$ ——i 部门基建投资动用额。

所以，i 部门本期固定资产拥有量为

$$\phi_i^0 + \frac{1}{2} C_i W_i$$

劳动手段约束可描述为：i 部门本期生产所占用的固定资产要少于或等于部门所拥有的固定资产，即

$$K_i X_i < \phi_i^0 + \frac{1}{2} C_i W_i \tag{6.3.28}$$

用矩阵表示为

$$\hat{\boldsymbol{K}}\boldsymbol{X} < \boldsymbol{\phi}^0 + \frac{1}{2}\hat{\boldsymbol{C}}\boldsymbol{W} \tag{6.3.29}$$

式中

$$\hat{\boldsymbol{K}} = \begin{bmatrix} K_1 & 0 & \cdots & 0 \\ & K_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & K_n \end{bmatrix} \quad \boldsymbol{\phi}^0 = \begin{bmatrix} \phi_1^0 \\ \phi_2^0 \\ \vdots \\ \phi_n^0 \end{bmatrix} \quad \hat{\boldsymbol{C}} = \begin{bmatrix} C_1 & 0 & \cdots & 0 \\ & C_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & C_n \end{bmatrix} \quad \boldsymbol{W} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix}$$

(3) 劳动力约束。劳动力对国民经济发展的限制体现在三个方面：① 劳动力总量的限制；② 劳动力分布造成的对部门劳动力需求的限制；③ 专门人才的限制。

设：a_{Li} 为第 i 部门劳动消耗系数，其计算公式为

$$a_{Li} = \frac{L_i}{X_i} = \frac{1}{T_i}$$

式中 L_i ——i 部门的劳动力数量；

T_i ——i 部门的劳动生产率。

故劳动力约束可描述为：

劳动力总量的限制为

$$\sum_i a_{Li} X_i \leqslant \sum_i L_i \tag{6.3.30}$$

部门劳动力的限制为

$$a_{Li}X_i \leqslant L_i \tag{6.3.31}$$

专门人才的限制为

$$S_i^{(s)} \cdot a_{Li} \cdot X_i \leqslant L_i^{(s)} \tag{6.3.32}$$

式中 $S_i^{(s)}$——第 i 部门第 s 种专门人才的比重系数；

$L_i^{(s)}$——i 部门现有的第 s 种专门人才。

2．消费约束组

它表明为满足人民对消费资料的需求而对扩大再生产规模的限制和要求。

(1) 居民消费的供需平衡。国民经济最终产品所能提供的居民消费量为 z_3。居民收人主要用于购买消费品、非商品性开支和储蓄。因此居民收入扣除非商品开支和储蓄后应与国民经济所能提供的消费品量相平衡。

设生产性部门和非生产性部门职工的平均工资为 g_1，农业人口的平均收人为 g_2，d 为非生产部门的职工数，如果前 m 个部门为农业部门，则消费约束为

$$g_1\left(\sum_{m+1}^{n} a_{Li}X_i + d\right) + g_2\left(\sum_{i=1}^{m} a_{Li}X_i\right) - u \leqslant \sum_{i}^{n} Z_{i3} \tag{6.3.33}$$

式中 u——非商品性开支和储蓄额。

(2) 主要商品的平衡(如商品粮等)。商品粮是指农业部门提供的粮食产品扣除农业人口消费之后的那部分产品，其使用方向为供应非农业人口消费和各种生产性消耗、储备、出口等。

设：V——农业人口的需求量；

β——非农业人口与非农业职工所占的比重；

α——实际供应水平；

m_i——商品率；

d——非物质生产部门的职工数。

考虑逾期使用的商品粮平衡为

$$\alpha\beta\left(\sum_{i=m}^{n} aL_iX_i + d\right) + v = \sum_{i=1}^{m} m_i[s_iX_i^0 + (1 - S_i)X_i] \tag{6.3.34}$$

3．积累与消费比例约束组

它表明了为保证合理的积累和消费比例对扩大再生产规模的限制。

积累和消费的比例是国民经济发展的重要比例关系，它可通过积累基金和消费基金在国民收人中的比例来反映。如果规定积累的最低水平 r_1 和消费的最低水平 r_2，则约束可表示为

$$\text{积累基金} / \text{国民收入} \geqslant r_1；\text{消费基金} / \text{国民收人} \geqslant r_2$$

将上式具体化就是积累和消费的比例约束。

在社会总产品中扣除生产性消耗和折旧即是国民收人：

$$\sum_j N_j = \sum_i X_i - \sum_j \sum_i a_{ij}X_j - \sum_j f_j\left(\phi_j^0 + \frac{1}{2}C_jW_j\right) \tag{6.3.35}$$

式中 f_j——国民经济第 j 部门的综合折旧率。

生产性积累(增加的固定资产 $\sum_i \sum_j b_{ij}W_j$、增加的流动资金 $\sum_i Z_{i1}$ 和储备 $\sum_i S_i(X_i - X_i^0)$) 和非生产性积累 $\sum Z_{i3}$ 两部分构成积累基金：

$$\sum_i Y_{il} = \sum_i \sum_j b_{ij} W_j - \sum_i f_i\left(\phi_i^0 + \frac{1}{2} C_i W_i\right) + \sum_i Z_{i1} + \sum_i S_i (X_i - X_i^0) + \sum_i Z_{i2} \tag{6.3.36}$$

式中 Y_{il}——国民经济第 i 个部门的积累量。

消费基金由居民消费和社会消费两部分组成：

$$\sum_i y'_{il} = \sum_i Z_{i3} + \sum_i Z_{i4} \tag{6.3.37}$$

因此，积累约束为

$$\sum_i \sum_j b_{ij} W_j - \sum_i f_i\left(\phi_i^0 + \frac{1}{2} C_i W_i\right) + \sum_i Z_{i1} + \sum_i S_i (X_i - X_i^0) + \sum_i Z_{i2} \geqslant r_2\left[\sum_i X_i - \sum_i \sum_j a_{ij} X_j - \sum_j f_j\left(\phi_i^0 + \frac{1}{2} C_i W_i\right)\right] \tag{6.3.38}$$

消费约束为

$$\sum_i Z_{i3} + \sum_i Z_{i4} \geqslant r_2\left[\sum_i X_i - \sum_i \sum_j a_{ij} X_j - \sum_j f_j\left(\phi_j^0 + \frac{1}{2} C_j W_j\right)\right] \tag{6.3.39}$$

为了保证社会总产值与最终产品的平衡，要求进出口必须平衡，故又有进出口平衡约束

$$\sum_i Z_{i5} = \sum_i Z_{i6} \tag{6.3.40}$$

最后应设置非负约束，即

$$X_i \geqslant 0, W \geqslant 0, Z_Q \geqslant 0, Z_2 \geqslant 0, Z_3 \geqslant 0, Z_4 \geqslant 0, Z_5 \geqslant 0, Z_6 \geqslant 0 \tag{6.3.41}$$

6.3.4.3 国民经济综合平衡最优化模型的目标函数

国民经济系统是多目标系统，需依据经济规律建立综合评价指标体系。将评价指标具体化，大体可得以下几种：

(1) 国民收入最大；

(2) 社会最终产品最多；

(3) 社会总产品数量最多；

(4) 消费品数量最多；

(5) 在其他部门生产量保持一定水平的前提下，使某些部门生产量最大等。

通常对多目标问题的处理方法是考虑各目标的重要性采用评分等方法为各目标确定一定的权数后，将其化为单目标来处理，具体原则和方法见多目标决策部分。

现采用其中一种指标作为目标函数，如：

(1) 采用国民收入最大的目标函数：

$$\max \quad Z = \sum_i X_i - \sum_i \sum_j a_{ij} X_j - \sum_j f_j\left(\phi_j^0 + \frac{1}{2} C_j W_j\right) \tag{6.3.42}$$

(2) 采用最终产品最多的目标函数：

$$\max \quad Z = \sum_i X_i - \sum_i \sum_j a_{ij} X_j \tag{6.3.43}$$

将前面三种约束与目标函数相结合就构成了对国民经济系统实用的线性规划模型。当然在实践中，尚需考虑各种其他情况对约束进行增减或修改，使其更符合实际，这样才能为国民经济管理提供更科学的决策依据。

6.4 线性规划的求解方法

6.4.1 求解运输问题的特殊方法——表上作业法

运输问题是一类特殊的线性规划问题，其模型的一般形式为

$$\begin{aligned} \min \quad & Z = \sum_i \sum_j C_{ij} X_{ij} \\ \text{s.t.} \quad & \sum_j X_{ij} = a_i (i = 1,2,\cdots,m) \\ & \sum_i X_{ij} = b_j (j = 1,2,\cdots,n) \end{aligned} \tag{6.4.1}$$

$$X_{ij} \geqslant 0 \quad i = 1,2,\cdots,m \quad j = 1,2,\cdots,n$$

由式(6.4.1)的约束条件可见，变量的系数均为0和1，且系数矩阵为以下特殊形式

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 1 & \cdots & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix}$$

鉴于这种类型的线性规划问题可采取特殊解法，故单独作以介绍。

6.4.1.1 平衡型运输问题的求解方法

表上作业法的基本思想是：在确定一组可行的运输方案之后，以此方案为基础进行调整，比较调整前、后的费用，决定调整方案的取舍，最后得到最优方案。

现举例说明表上作业法求解运输问题的步骤。

某建筑公司有三个钢材供应站，钢材拥有量分别为700t、400t、900t，准备向四个建筑工地供应钢材，各工地的钢材需要量为300t、600t、500t、600t，各供应站向各工地的钢材调运单位运价列于表6.4.1的小方框中，现需制定运输费用最省的调运计划。

表6.4.1 各供应站向各工地的钢材调运单位运价

工地 / 供应站	1	2	3	4	钢材拥有量
1	3	11	400 3	300 10	700
2	300 1	9	100 2	8	400
3	7	600 4	10	300 5	900
钢材需求量	300	600	500	600	2000

该问题是典型的运输问题，且 $\sum_i a_i = \sum_j b_j = 2000$，为平衡型运输问题。用表上作业法求解的步骤为：

1．确定初始可行方案

确定初始方案的方法很多，较常用的有最小元素法、行最小元素法、列最小元素法、西北角法等，其中最小元素法较好。

用最小元素法编制初始方案的基本原则是就近供应，即由运价表中选最小元素 $\min\{c_{ij}\}$ 确定相对应的供需地点 A_i 和 B_j，以供应地 A_i 的拥有量 a_i 和需求地 B_j 的需求量 b_j 的最小值作为调运量 X_{ij}（$X_{ij} = \min\{a_i, b_j\}$），直至编制一完整的调运方案为止。

以表 6.4.1 中的运价为例子说明具体方法。

在运价表中，运费最低的元素为第二行第一列的“1”，即 $\min\{c_{ij}\} = C_{21} = 1$，从而确定供应地 A_2 向需求地 B_1 调运钢材，调运量为 $X_{21} = \min\{400, 300\} = 300$。这样做的经济意义是：按最低的运价首先由 A_2 供应站满足 B_1 工地的需求。因 A_2 供应站的钢材拥有量大于 B_1 工地的需求量，所以选择 300 作为调运量，表示 B_1 工地的全部需求由 A_2 供应站供应。既然 B_1 工地的需求全部被满足，故应勾掉 B_1 所在的列，以表示其他供应站不能再向 B_1 工地供应钢材。A_2 供应站向 B_1 工地供应后，剩余量为 $400 - 300 = 100$，故 A_2 供应站的拥有量 a_2 变成 100，这样就完成一次选择。再按上述办法继续进行，每确定一个 X_{ij}，勾掉一行或一列，直到全部的行和列被勾掉为止，即得一个初始调运方案。本例所得到的初始方案见表 6.4.1。

在用最小元素法确定初始方案时，当遇到供应地拥有量 a_i 与需求地需求量 b_j 相等的情况时，则需同时勾掉第 i 行和第 j 列。因这时第 i 个供应地已再无货可供，第 j 个工地已满足。但需在第 i 行或第 j 列的某个位置添“0”，并在以后的计算中把它当作一个实在的运输量，以满足确定一个 X_{ij} 只勾掉一行或一列的需要。

在确定初始方案之后，需检查是否满足 X_{ij} 的个数与供应地和需求地数量之和减“1”这一条件，即 $(m + n - 1)$，如满足，则称初始方案为初始可行方案，否则虽满足供需条件，但不满足运算要求，故不能称其为初始可行方案。

一般按上述最小元素法所得到的方案均为初始可行方案。

2．初始可行方案的检验

初始可行方案是否为最优方案，需进行检验。检验的目的有二：

(1) 确定是否为最优方案；

(2) 如不是最优方案，为方案的调整提供依据。因此，进行初始可行方案的检验是表上作业方法的关键。

对初始可行方案的检验方法有两种，一种是闭回路法，一种是位势法。闭回路法在进行方案检验时，比较复杂，本节主要介绍位势法。

位势法的方法为：

对初始可行方案的每个供应地和需求地都按下式确定其位势值：

$$U_i + V_j = C_{ij}, \text{ 且 } U_1 = 0$$

式中 U_i —— 第 i 个供应地的位势；

V_j —— 第 j 个需求地的位势；

C_{ij} ——第 i 供应地向第 j 需求地供应物资的单位运费。

位势的涵义可理解为对应于初始可行方案，供需双方各自支付的单位运费值，U_i 表示在初始可行方案下，第 i 个供应地运出单位物资时，所支付的运费量，V_j 表示第 j 个需求地收到单位物资时，所支付的运费值；$U_i + V_j = C_{ij}$ 为供应者和需求者所支付的运费之和，它应与运输部门实收的运费 C_{ij} 相等。简单地说，运输部门所实收的运费，分别由供需双方的位势共同负担。

由上式可见，当 $U_1 = 0$ 时，可通过初始可行方案逐一求出 U_i 和 V_j。

令 Z_{ij} 为初始可行方案下 i、j 两地对应的检验数，且

$$Z_{ij} = C_{ij} - (U_i + V_j)$$

若 $Z_{ij} = 0$，则表明由 i 地向 j 地调运单位物资时，运输部门实收运费与供需双方所支付的运费相等，对应于初始可行方案的 $Z_{ij} = 0$。

若 $Z_{ij} > 0$，则表明由 i 地向 j 地调运单位物资时，运输部门实收运费高于供需双方所支付的运费之和，初始可行方案是较便宜的方案。

若 $Z_{ij} < 0$，则表明由 i 地向 j 地调运单位物资时，运输部门的实收运费低于供需双方所支付的运费之和，初始可行方案应以增加由 i 向 j 的调运量为方向进行调整。

由上述可见，Z_{ij} 可作为评价方案优劣的标准，从而得出如下判断准则：

(1) 若所有 $Z_{ij} \geqslant 0$ 时，初始可行方案为最优方案；

(2) 当 Z_{ij} 至少有一个小于零时，该方案不是最优方案，且应以最小的 Z_{ij} 作为方案调整的依据。

现用位势法检验上述初始可行方案。

(1) 求位势：

令 $U_1 = 0$，按 $Z_{ij} = C_{ij} - (U_i + V_j)$ 求出各供应地和需求地的位势 U_i 和 V_j。

根据 $U_1 + V_3 = C_{13}$ 得：$V_3 = C_{13} - U_1 = 3 - 0 = 3$；

根据 $U_3 + V_4 = C_{34}$ 得：$U_3 = C_{34} - V_4 = 5 - 10 = -5$；

根据 $U_2 + V_3 = C_{23}$ 得：$U_2 = C_{23} - V_3 = 2 - 3 = -1$；

根据 $U_3 + V_2 = C_{32}$ 得：$V_2 = C_{32} - U_3 = 4 - (-5) = 9$；

根据 $U_1 + V_4 = C_{14}$ 得：$V_4 = C_{14} - U_1 = 10 - 0 = 10$；

根据 $U_2 + V_1 = C_{21}$ 得：$V_1 = C_{21} - U_2 = 1 - (-1) = 2$。

(2) 将按 $Z_{ij} = C_{ij} - (U_i + V_j)$ 计算的判断数，填入表6.4.2的无调运量的空格中(因有调运量的格中 Z_{ij} 均为零，故不计算)。

$$Z_{11} = C_{11} - (U_1 + V_1) = 3 - (0 + 2) = 1$$

$$Z_{12} = C_{12} - (U_1 + V_2) = 11 - (0 + 9) = 2$$

$$Z_{22} = C_{22} - (U_2 + V_2) = 9 - (-1 + 9) = 1$$

$$Z_{24} = C_{24} - (U_2 + V_4) = 8 - (-1 + 10) = -1$$

$$Z_{11} = C_{11} - (U_1 + V_1) = 3 - (0 + 2) = 1$$

$$Z_{31} = C_{31} - (U_3 + V_1) = 7 - (-5 + 2) = 10$$

$$Z_{33} = C_{33} - (U_3 + V_3) = 10 - (-5 + 3) = 12$$

由表6.4.2中检验数可见，该初始可行方案不是最优方案。且应由第二个供应站向第

表 6.4.2　判断数

	B_1	B_2	B_3	B_4	木材拥有量	U_i
A_1	3 (1)	11 (2)	400 3 D	300 10 C −	700	0
A_2	1 300	9 (1)	100 2 − A	8 B + (−1)	400	−1
A_3	7 (10)	4 600	10 (12)	300 5	900	−5
木材需求量	300	600	500	600	2000	
V_j	2	9	3	10		

四个工地增加钢材供应量。

3. 初始可行方案的改进

Z_{ij} 为负值是指由 i 地向 j 地调运单位物资与初始可行方案相比运输费用的减少值。因此应对初始可行方案进行调整，尽量多地增加由 i 地向 j 地的供应量。为了保证供需平衡，需寻找一条改变调运的路线，即寻找一条闭合回路。所谓闭合回路是指顶角为直角的任意多边形所组成的回路。该多边形的特点是其顶角除对应于 Z_{ij} 为负值的格外，其余顶角都对应于初始可行方案。如本例中闭合回路中所对应的多边形为 $ABCD$（见表 6.4.2）。B 点所对应的是 $Z_{ij}=-1$ 的空格，即初始可行方案 $X_{24}=0$，其余三个点分别对应于调运量为 100、300、400 的格，即 $X_{23}=100$，$X_{13}=400$，$X_{14}=300$，如果向 X_{24} 增加一个单位的调运量，为了保证行和列的平衡，X_{23} 必须减少一个单位、X_{13} 必须增加一个、X_{14} 必须减少一个单位的调运量。为了清楚地表明这种为了保持平衡所需要调整的增加或减少关系，在 Z_{ij} 为负值的闭回路顶点标“+”号，然后沿闭回路按顺时针或逆时针的方向，以该“+”号为起点，“−”号和“+”相间地对其他顶点进行标注，“−”号则表示所对应的调运量应减少，“+”号则表示相对应的调运量应增加。需用调整的调运量则为所有“−”号调运量中最小的调运量，如本例中的调整调运量为 100（见图 6.4.1）。调整后的初始可行方案为：X_{24} 由零增加到 100，X_{13} 由 400 增加到 500，X_{14} 由 300 减少到 200，而 X_{23} 由 100 变为零而退出原初始可行方案。调整后又得到一新的可行方案（见表 6.4.3），该方案与初始可行方案相比，运输费用下降了 $1\times100=100$（万元）。

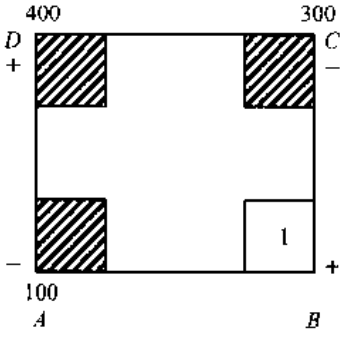


图 6.4.1

对于该可行方案再应用位势法得出检验数，重复上述步骤即可在有限步内得到运输问题的最优解。

表 6.4.3　可行方案

加工厂 / 林业局	B_1	B_2	B_3	B_4	木材拥有量
A_1	3	11	500 3	200 10	700
A_2	300 1	9	2	100 8	400
A_3	7	600 4	10	300 5	900
木材需求量	300	600	500	600	2000

6.4.1.2　非平衡运输问题的求解方法

任何非平衡运输问题均可化成平衡型运输问题来处理。

1. $\sum_i a_i > \sum_j b_j$型非平衡型问题

由于供应地的拥有量大于需求量，因此有的产地的产品不能全部运出，需设置仓库存储。如果将仓库当作需求者，且需求量为 $\sum_i a_i - \sum_j b_j$，因就地存储不需运费，故设运费为零，这样就将这类不平衡问题化成了平衡型问题。

假设上例中第二个供应站的拥有量增加了 200 万 m^3，如需求量不变，则问题变为不平衡问题。化成平衡型问题见表 6.4.4，应用上述表上作业法即可求解。

表 6.4.4　拥有量大于需求量情况的平衡表

工地 / 供应站	B_1	B_2	B_3	B_4	仓库	钢材拥有量 /t
A_1	3	11	3	10	0	700
A_2	1	9	2	8	0	600
A_3	7	4	10	5	0	900
钢材需求量 /t	300	600	500	600	200	2200

2. $\sum_i a_i < \sum_j b_j$型非平衡型问题

由于需求量大于拥有量，故有的需求地的需求量不能全部得到满足。为了全部满足各需求地的需求量，必须增加供应地的拥有量。现假设存在一虚拟的供应地，其拥有量为 $\sum_j b_j - \sum_i a_i$，即可满足供应，使不平衡变为平衡。但由于该拥有量是假设的，实际是不存在的，故由它向各需求地的单位产品调运费用为很大的值，这在安排调运方案时，就必然首先满足费用低，最后满足费用高的地区，也就是先将真实供应地的物资调出，最后才是假设供应地的物资调出。这就实现了把不平衡问题化为平衡问题的变化，使问题得到解

决。在表 6.4.1 的基础上，假设 B_1 工地需求量增加 200t，引入虚拟供应站 A_4，它向各工地运输钢材的单位费用为 M(很大的数)，则得表 6.4.5 所列的平衡表，从而实现了非平衡问题向平衡问题的转化，应用表上作业法即可求解。

表 6.4.5　需求量大于生产量情况的平衡表

加工厂 / 林业局	B_1	B_2	B_3	B_4	木材拥有量 /t
A_1	3	11	3	10	700
A_2	1	9	2	8	400
A_3	7	4	10	5	900
A_4	M	M	M	M	200
木材需求量 /t	500	600	500	600	2200

6.4.2　匈牙利法

分配问题模型的基本形式是

目标函数为：min
$$Z = \sum_i \sum_j C_{ij} X_{ij}$$

s.t.
$$\sum_j X_{ij} = 1 \quad (i = 1,2,\cdots,n)$$
$$\sum_i X_{ij} = 1 \quad (j = 1,2,\cdots,n)$$
$$X_{ij} = \begin{cases} 1, & i \text{ 从事 } j \text{ 工作时} \\ 0, & i \text{ 不从事 } j \text{ 工作时} \end{cases}$$

由上述模型可见，分配问题是运输问题当 $a_i = 1$、$b_i = 1$ 时的特殊情况，这类问题除用表上作业法求解之外尚可采用一种特殊方法 —— 匈牙利法求解。

匈牙利法是从系数矩阵 $\{C_{ij}\}$ 出发来确定最优分配方案的方法。由于 X_{ij} 的一行或一列中只有一个“1”，因此从矩阵 $\{C_{ij}\}$ 的一行或一列中减去任一常数 K，则目标函数也相应减少 K，这样以 Z 为目标的最优解和以 $Z - K$ 为目标的最优解必然是相同的。故匈牙利法就是从 $\{C_{ij}\}$ 出发，从矩阵 $\{C_{ij}\}$ 的一行或一列中分别减去该行或该列的最小元素后，得到与 $\{C_{ij}\}$ 同最优解的系数矩阵 $\{b_{ij}\}$，再以 $\{b_{ij}\}$ 为基础按系数最小的原则选取分配方案的一种方法。

现举例说明其计算步骤。

某建筑公司有五个工程队，现准备去五个工地作业，由于工程队的设备、人力各不相同，五个工地的条件也各不相同，因此每个工程队在不同的工地工作所需的作业时间也不相同，设每个工程队在每个工地工作的计划时间见表 6.4.6，安排哪个工程队去哪个工地进行作业，才能使企业作业的总时间最少。

表 6.4.6　计划时间

工程队 \ 工地	1	2	3	4	5
1	13	7	10	7	10
2	8	10	6	6	6
3	7	15	10	11	10
4	14	14	6	6	9
5	4	9	6	9	6

步骤一　在系数矩阵的每行和每列中减去最小元素，使矩阵的每行和每列至少有一个零元素。

本例的变换如下：

$$\begin{pmatrix}13&7&10&7&10\\8&10&6&6&6\\7&15&10&11&10\\14&14&6&6&9\\4&9&6&9&6\end{pmatrix}\begin{matrix}-7\\-6\\-7\\-6\\-4\end{matrix}\xrightarrow{1}\begin{pmatrix}6&0&3&0&3\\2&4&0&0&0\\0&8&3&4&3\\8&8&0&0&3\\0&5&2&5&2\end{pmatrix}\xrightarrow{2}\begin{pmatrix}6&⓪&3&ø&3\\2&4&ø&⓪&ø\\⓪&8&3&4&3\\8&8&⓪&ø&3\\ø&5&2&5&2\end{pmatrix}$$

步骤二　按零元素最少的行优先进行分配。由于零元素所对应的分配关系所消耗的时间最少，故在零元素最少的行优先选择零元素并将其用 ○ 圈起，零元素一旦选定，则说明该工程队不能再在其他工地工作，也说明该工地不能再安排其他工程队，故应划掉与该零元素同行和同列的其他零元素，按此法一直到进行不下去为止。本例的选择结果见$\xrightarrow{2}$。

步骤三　如果选中的零元素个数与矩阵维数相同，则选中的零元素所对应的分配关系即是最优方案；如果选取中的零元素个数小于矩阵维数，则说明有的工程队未分配工作、有的工地未安排工程队。因此需要进行调整，调整方法为：

① 在无 ○ 号的行标 √ 号；

② 在有 √ 号行上的零元素所在的列标 √ 号；

③ 再在标 √ 号的列中有 ○ 号的零元素所对应的行标 √；

④ 重复直至进行不下去为止；

⑤ 对无 √ 号的行和有 √ 号的列画直线，在直线未覆盖的元素中选择最小元素；

⑥ 从有 √ 号的行中减去该元素，在有 √ 号的列上加上该元素，则得到一个新矩阵。

本例结果见$\xrightarrow{3}$；

步骤四　在此新矩阵的基础上重复第二至第四步即可得出最优解，见$\xrightarrow{4}$。

$$\xrightarrow{3}\begin{pmatrix}ø&ø&ø&ø&ø\\2&4&0&0&0\\0&8&3&4&3\\8&8&0&0&3\\0&5&2&5&2\end{pmatrix}\begin{matrix}\\\\√\\\\√\end{matrix}\begin{matrix}\\\\-2\\\\-2\end{matrix}\xrightarrow{4}\begin{pmatrix}8&⓪&3&0&3\\4&4&ø&0&⓪\\⓪&6&1&2&1\\10&8&ø&⓪&3\\ø&3&⓪&3&ø\end{pmatrix}$$

本例的最优解是：第一工程队分配去第二工地，工期 7 天；第二工程队分配去第五工地，工期 6 天；第三工程队分配去第一工地，工期 7 天；第四工程队分配去第四工地，工期 6 天；第五工程队分配去第三工地，工期 6 天；全部作业共需 32 天。

这里需要指出的是，用匈牙利法求解分配问题的条件是：系数矩阵$\{C_{ij}\}$的元素非负，因此求解最大目标的分配问题不能采用改变系数符号的办法将其变为最小目标来求解，即 $\max Z = \sum_i \sum_j C_{ij}X_{ij}$ 不能用 $Z' = -Z = -\sum_i \sum_j C_{ij}X_{ij} \to \min$ 来解。

为了求解最大目标的分配问题，可定义一新矩阵 $\boldsymbol{B}$，其元素按下式计算：

$$b_{ij} = M - C_{ij}$$

式中 M——矩阵$\{C_{ij}\}$中的最大元素。

因为：

$$\sum_i \sum_j b_{ij}X_{ij} = \sum_i \sum_j (M - C_{ij})X_{ij} = \sum_i \sum_j MX_{ij} - \sum_i \sum_j C_{ij}X_{ij} = nM - \sum_i \sum_j C_{ij}X_{ij}$$

由上式可见，只有当 $\sum_i \sum_j C_{ij}X_{ij}$ 取大值时，$\sum_i \sum_j b_{ij}X_{ij}$ 才能取最小值。因此 $\sum_i \sum_j b_{ij}X_{ij}$ 取最小值的解，就是 $\sum_i \sum_j C_{ij}X_{ij}$ 取最大值的解。

通过这种方法就可解决变最大目标为最小的目标的分配问题。

6.4.3 线性规划的标准形式

前面所介绍的各种线性规划模型，由于问题的繁简程度及问题的性质不同，所建的模型也各不相同，这为求解线性规划带来很大困难。因此提出，应采用一种标准形式的线性规划模型，通过对其求解方法的研究，解决任何形式线性规划的求解问题。

线性规划的标准形式是：

$$\begin{aligned} \min \quad & Z = C_1X_1 + C_2X_2 + \cdots + C_nX_n \\ \text{s.t.} \quad & a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n = b_1 \\ & a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n = b_2 \\ & \cdots \\ & a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n = b_m \\ & X_1 \geqslant 0 \quad X_2 \geqslant 0 \quad \cdots \quad X_n \geqslant 0 \end{aligned} \tag{6.4.2}$$

用矩阵形式表示为

$$\begin{aligned} \min \quad & Z = CX \\ \text{s.t.} \quad & AX = b \\ & X \geqslant 0 \end{aligned}$$

其中：

$$\boldsymbol{C} = (C_1, C_2, \cdots, C_n)$$

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \qquad \boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \qquad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

$\boldsymbol{A}$ 为 $m \times n$ 阶实数矩阵，$\boldsymbol{C}$ 和 $\boldsymbol{X}$ 均为 n 维向量，$\boldsymbol{b}$ 为 m 阶大于零的列向量。

由于系统的复杂性，所建立的模型并非都是上述标准形式，因此需把任一形式的线性规划模型变为标准形式，再应用统一的方法来求解。

为把任一线性规划模型化为标准型，可用以下方法。需从目标函数、约束条件、变量和约束值几方面研究。

6.4.3.1 目标函数

建立模型时，目标函数有最大(max)和最小(min)两种。在标准型中规定用最小(min)的 Z 表示目标函数，这在处理上与目标最大并无本质区别，因为当目标要求最大时，可通过把目标函数两端同乘“-1”的方法化为最小目标，即 $\min Z' = \max(-Z)$。

6.4.3.2 约束

在标准型中，规定约束为等式。一个系统的约束很难全用等式表示，但它们都可以化为等式。

设约束：$$a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n \geqslant b_1$$

在不等式左端减去一个非负的变量 X_{n+1}，总可以使其变为等式

$$a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n - X_{n+1} = b_1$$

同理，如约束为 $$a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n \leqslant b_1$$

在不等式左端加上一个非负的变量 X_{n+1}，也可以使其变为等式

$$a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + X_{n+1} = b_1$$

我们称这种将不等式化为等式而加上或减去的非负变量 X_{n+1} 为松弛变量。一切不等式约束均可通过引入松弛变量的方法化为等式约束。

6.4.3.3 变量

标准型中要求所有变量必须非负。在实际中有很多变量，如温度、距离等均可出现负值，因此，如果变量 X_k 不满足非负条件，可引入两个非负的变量 X'_k 和 X''_k，使 $X_k = X'_k - X''_k$ 成立，则 X_k 的任何情况，均可由此式满足，同时也满足了非负要求。

6.4.3.4 约束值

标准型的约束值规定为 $b_i \geqslant 0$。在某些实际问题中，b_i 可能为负值，为了化为标准形式，在该约束方程两端同乘“-1”，则可使 b_i 满足 $b_i \geqslant 0$ 的条件。

经过上述处理，任何线性规划问题都可化成标准形式。这样在研究线性规划求解问题时，可专门研究标准形式。只要标准形式可以求解，其他问题也均可化成标准形式求解。上述变换的实质是以增加变量为代价实现用统一方法求解线性规划的。

6.4.4 线性规划基本定义、定理及图解法

在介绍线性规划的解法之前，首先介绍几个定义和定理：

考虑线性规划模型

$$\begin{aligned} \min \quad & Z = CX \\ \text{s.t.} \quad & AX = b \\ & X \geqslant 0 \end{aligned}$$

定义：

(1) 可行解:凡满足约束条件 $AX = b, X \geqslant 0$ 的解叫可行解。

(2) 可行区:可行解的集合叫可行区。

(3) 基矩阵:约束条件 $AX = b$ 中,A 是 $m \times n$ 阶矩阵,它的秩为 m,选其中任意 m 列所形成的非奇异 $m \times m$ 子矩阵 B,称之为基矩阵。

(4) 基解:$AX = b$ 中,令不与基矩阵 B 的列相对应的 $n - m$ 个变量为零,所形成的方程组为基方程组,变量为基变量,所得的解称为基矩阵 B 的基解。

(5) 基可行解:凡满足非负条件的基解为基可行解。

(6) 退化基解:基解中有一个或多于一个基变量为零时,这个解称为退化基解,否则为非退化基解。

(7) 退化基可行解:若基可行解中有一个或多于一个基变量为零时,这个解称为退化基可行解。

(8) 最优解:使目标函数 Z 达到最小值的基可行解叫最优解。

可行解、基解、基可行解和最优解之间的关系可用图 6.4.2 的集合关系说明。

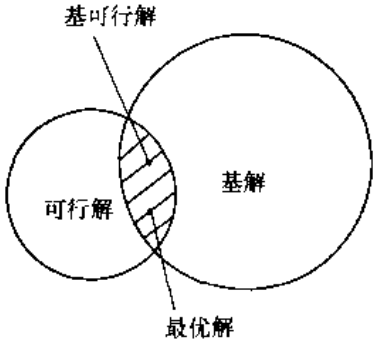


图 6.4.2　集合关系

定理一　(解的存在定理):对应于标准形式的线性规划,① 若存在可行解,则必存在一个基可行解;② 若存在一个最优可行解,则必存在一个最优基可行解。

定理二　(极点和基解的等值性定理):令 K 是满足约束条件的所有 n 维向量 $\boldsymbol{X}$ 所组成的凸多胞形,那么向量 $\boldsymbol{X}$ 为 $\boldsymbol{K}$ 的一个极点的充分且必要条件是 $\boldsymbol{X}$ 为约束方程的基可行解。

上述两个定理为求解线性规划问题所提供的指导思想是只要找到约束条件所组成的凸多胞形的极点(即基可行解),就可从中找到最优解。

在了解这些基本概念和定理后,即可研究线性规划问题的求解方法。首先根据定义和定理,用图解法求解线性规划问题。

这种方法只适用于两个变量的线性规划模型,它是采用笛卡儿坐标系进行的,其指导思想是根据约束方程确定可行区及其顶点,再从顶点找出最优解,并且无需把模型变成标准型。

现举例说明如何用图解法求解线性规划问题。

[例 1]　某工厂生产两种产品,每种产品材料(钢材、电力)消耗定额、单位利润及材料的资源量见表 6.4.7。如何安排生产,才能使利润最大。

表 6.4.7　单位消耗、资源拥有量、单位利润表

资源 \ 产品	产品 1	产品 2	资源量
钢材	20	50	1000
电力	20	10	600
单位利润	250	500	

制定产品生产计划的模型为：

$$\begin{aligned} \max \quad & Z = 250X_1 + 500X_2 \\ \text{s.t.} \quad & 20X_1 + 50X_2 \leqslant 1000 \\ & 20X_1 + 10X_2 \leqslant 600 \\ & X_1 \geqslant 0, X_2 \geqslant 0 \end{aligned}$$

现将约束方程画在笛卡儿坐标系上(见图 6.4.3)，得出可行区 $OBCD$，它是个凸多边形。再画 $Z = 250X_1 + 500X_2$，$Z = C$(C 为常数) 的等 Z 值线。当 $Z = 0$ 时，直线通过原点，随着 Z 值的增加，直线向右上方平移；这样就得到一簇线，它的 Z 值由小到大。当 $Z = 11250$ 时，直线正好通过 C 点，当 $Z > 11250$ 时，直线已脱离可行区，即不再与凸多边 $OBCE$ 相交，因此最大值为 11250，对应于图中点 $C(25,10)$，其坐标 $X_1 = 25, X_2 = 10$，即为最优解。该厂应生产 25 个产品 1，10 个产品 2，才能使利润最多，其值为 11250 元。

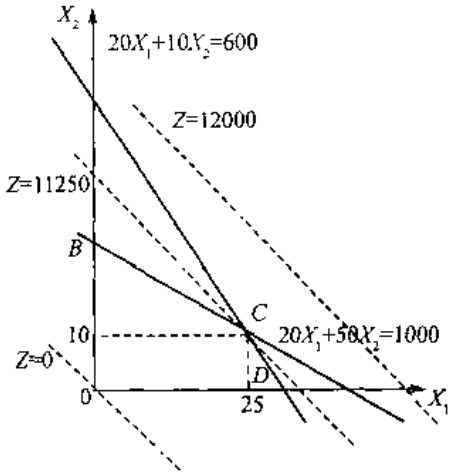


图 6.4.3　图解法

由上述可知，图解法就是先用约束方程找出可行区，再用等 Z 值线确定使目标函数达到极大或极小的顶点，该顶点的坐标就是最优解。

由图解法可知，当等值线与可行区的一条边相平行时，为一特殊情况。这种情况下，当 Z 值线平移时，不是在可行区的某顶点，而是在可行区的某条边脱离可行区，这样一来，这条边上的所有点都是最优点，此时线性规划有无穷多组最优解。

简单、清楚、且不需把规划模型变成标准形式是图解法的优点。

只适用于两个变量，且做图准确程度不易保证，实用价值很小，则是其致命缺点。但图解法作为理解线性规划问题求解过程的工具，其意义是重要的。因此必须深刻理解图解法的涵义，为学习线性规划实用解法 —— 单纯形法奠定基础。

在解含多个变量的线性规划问题时，图解法则无能为力了，现介绍按定义和定理来求解的过程。

仍用上例，并引入松弛变量将其化为标准型：

$$\begin{aligned}\min \quad & Z' = -Z = -250X_1 - 500X_2 + 0X_3 + 0X_4 \\ \text{s.t.} \quad & 20X_1 + 50X_2 + X_3 = 1000 \\ & 20X_1 + 10X_2 + X_4 = 600 \\ & X_1, X_2, X_3, X_4 \geqslant 0\end{aligned}$$

模型中矩阵

$$\boldsymbol{A} = \begin{pmatrix} 20 & 50 & 1 & 0 \\ 20 & 10 & 0 & 1 \end{pmatrix}$$

为 2×4 矩阵，其秩为 2。先取基矩阵 $\boldsymbol{B}$，$\boldsymbol{B}$ 的取法有 6 种：

$$\boldsymbol{B}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

基变量为　X_3, X_4；

基方程为　$X_3 + 0 = 1000$，

$0 + X_4 = 600$；

基解为　$X_3 = 1000, X_4 = 600$；

非基解　$X_1 = 0, X_2 = 0$；

因 $X_1, X_2, X_3, X_4 \geqslant 0$，故这组解为基可行解。

$$\boldsymbol{B}_2 = \begin{pmatrix} 50 & 1 \\ 10 & 0 \end{pmatrix}$$

基变量为　X_2, X_3；

基方程为　$50X_2 + X_3 = 1000, 10X_2 + 0 = 600$；

基解为　$X_2 = 60, X_3 = -2000$。

非基解为　$X_1 = X_4 = 0$，因 $X_3 = -2000 < 0$，故这组解不是基可行解。

再继续取

$$\boldsymbol{B}_3 = \begin{pmatrix} 20 & 50 \\ 20 & 10 \end{pmatrix} \qquad \boldsymbol{B}_4 = \begin{pmatrix} 20 & 1 \\ 20 & 0 \end{pmatrix}$$

$$\boldsymbol{B}_5 = \begin{pmatrix} 50 & 0 \\ 10 & 0 \end{pmatrix} \qquad \boldsymbol{B}_6 = \begin{pmatrix} 20 & 0 \\ 20 & 1 \end{pmatrix}$$

计算结果见表 6.4.8(包括把基可行解代入目标函数所得的 Z 值)。

表 6.4.8　计算结果

序号	基本解	非基本解	Z 值	是否是基可行解	对应点
1	$X_3 = 1000$ $X_4 = 600$	$X_1 = 0$ $X_2 = 0$	0	是	O
2	$X_2 = 60$ $X_3 = -2000$	$X_1 = 0$ $X_4 = 0$	30000	不是	E

（续）

序号	基本解	非基本解	Z 值	是否是基可行解	对应点
3	$X_1 = 25$ $X_2 = 10$	$X_3 = 0$ $X_4 = 0$	11250	是	C
4	$X_1 = 30$ $X_3 = 400$	$X_2 = 0$ $X_4 = 0$	7500	是	D
5	$X_2 = 20$ $X_4 = 400$	$X_1 = 0$ $X_3 = 0$	10000	是	B
6	$X_1 = 50$ $X_4 = -400$	$X_2 = 0$ $X_3 = 0$	12500	不是	F

因最优解必定是基可行解，故由表可见，它只能在 O、B、C、D 四点中选择，C 点所对应的基解 $X_1 = 25, X_2 = 10, Z = 11250$ 是该问题的最优解，与图解法的结果一致。

用这种方法求解线性规划，可克服图解法的缺点，可处理两个以上变量的线性规划问题，但在约束较少、变量也较少的时候该方法才适用，其原因是该方法必须全部求出基解，而基解的数目等于 m、n 的组合数，即

$$\binom{n}{m} = \frac{n!}{(n-m)!m!}$$

当 n、m 较大时，基解数相当大，如 $n = 10, m = 5$，共有

$$\binom{10}{5} = \frac{10!}{5! \cdot 5!} = 252$$

种组合形式，基解共有 252 组，因此当 m、n 较大时，用该方法几乎无法计算。虽然如此，该方法仍是我们理解线性规划定义、定理和线性规划求解方法的最基本方法。

6.4.5 单纯形法

单纯形法是丹西格(G.B.Dantzig)1947 年提出的求解线性规划问题的一般方法。到目前为止，它仍是求解线性规划问题时使用最普遍的一种方法。单纯形法的基本思想是：根据线性规划的标准型，从约束集合的一个基可行解过渡到另一个基可行解，且不断减少目标函数值，从而得到最优解，使目标达到极小的一种方法。

为了牢固地掌握单纯形法，先从线性方程组的求解谈起。

由初等线性变换可知，把一个线性方程组中的任何一个方程 E_t，乘以一个不为零的常数 K，方程组的另一个方程 E_i 用 $E_i + KE_t$ 代替后，所得的方程组与原方程组同解。应用初等变换的这一性质求解线性方程组时，最好的方法是主消元法。该方法为了使方程组的系数矩阵变成单位矩阵，可采用如下一套线性变换：

(1) 令 $i = 1$；

(2) 为使第 i 个方程 E_i 的第 i 个变量 X_i 的系数为“1”，用第 i 个变量的系数 a_{ii} 遍除第 i 个方程后，得出方程 E'_i；

(3) 其他方程 $E_K(k \neq i)$ 用 $E_k - \frac{a_{kj}}{a_{ij}}E_i = E_k - a_{ki}E''$ 代替，消掉方程 E_K 中的变量 X_i；

(4) 如果 $i = n$，则结束；否则 $i - i + 1$ 转(2)。

如方程组：$X_1 + 2X_2 + 3X_3 = 6$

$2X_1 + 3X_2 + X_3 = -1$

$3X_1 + X_2 + 2X_3 = 7$

应用上述变换方法，即得

$$\begin{pmatrix} 1 & 2 & 3 & 6 \\ 2 & 3 & 1 & -1 \\ 3 & 1 & 2 & 7 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & -1 & -5 & -13 \\ 0 & -5 & -7 & -11 \end{pmatrix} \longrightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 & 6 \\ 0 & 1 & 5 & 13 \\ 0 & -5 & -7 & -11 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -7 & -20 \\ 0 & 1 & 5 & 13 \\ 0 & 0 & 18 & 54 \end{pmatrix} \longrightarrow$$

$$\begin{pmatrix} 1 & 0 & -7 & -20 \\ 0 & 1 & 5 & 13 \\ 0 & 0 & 1 & 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{pmatrix}$$

明显地求出方程的解：

$$X_1 = 1; X_2 = -2; X_3 = 3。$$

这种线性变换方法使方程组的求解变得容易和明显。在这种方法的启发和提示下，求解线性规划时，也可采用类似的方法。但线性规划与方程组不同，线性规划的约束方程组一般含 n 个变量，有 m 个方程(且设 $n > m$)，因而在应用上述变换时需将消元法的定义扩大，即在 n 个变量中有 m 个具有如下性质：(1) 每个变量只在一个方程中出现，且系数为"1"；(2) 任意一个方程都不出现 m 个变量中的两个。这样线性规划的约束方程组就变成如下形式：

$$\begin{aligned} X_1 \quad &+ \tilde{a}_{1,m+1}X_{m+1} + \tilde{a}_{1,m+2}X_{m+2} + \cdots + \tilde{a}_{1,n}X_n = \tilde{b}_1 \\ X_2 \quad &+ \tilde{a}_{2,m+1}X_{m+1} + \tilde{a}_{2,m+2}X_{m+2} + \cdots + \tilde{a}_{2,n}X_n = \tilde{b}_2 \\ &\cdots \\ X_m &+ \tilde{a}_{m,m+1}X_{m+1} + \tilde{a}_{m,m+2}X_{m+2} + \cdots + \tilde{a}_{m,n}X_n = \tilde{b}_n \end{aligned}$$

用矩阵表示：

$$\tilde{\boldsymbol{A}}\boldsymbol{X} = \tilde{\boldsymbol{b}}$$

式中

$$\tilde{\boldsymbol{A}} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \tilde{a}_{1,m+1} & \cdots & \tilde{a}_{1,n} \\ 0 & 1 & \cdots & 0 & \tilde{a}_{2,m+1} & \cdots & \tilde{a}_{2,n} \\ 0 & 0 & \cdots & 0 & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & \tilde{a}_{m,m+1} & \cdots & \tilde{a}_{m,n} \end{pmatrix}$$

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \qquad \tilde{\boldsymbol{b}} = \begin{pmatrix} \tilde{b}_1 \\ \tilde{b}_2 \\ \vdots \\ \tilde{b}_m \end{pmatrix}$$

如果取 $\tilde{A}$ 矩阵的前 m 列为基矩阵，则 $X_1, X_2, \cdots, X_m$ 为基变量，令 $X_{m+1}, X_{m+2}, \cdots, X_n$ 为零，显然 $X_1 = \tilde{b}_1, X_2 = \tilde{b}_2, \cdots, X_m = \tilde{b}_m$ 为线性规划的一组基解，如果所有 $X_j \geqslant 0 (j = 1, 2, \cdots, m)$，这组基解就是基可行解。如果将某个基变量和非基变量互换位置，再经过上述线性变换，就可得到另一组基可行解。如此下去，就可得到一组又一组的基可行解。在上述变换的基础上，为了由一组基可行解变换到另一组基可行解，通常不需按上述变换从头开始，只需按以下步骤即可：

(1) 选择第 r 个方程的第 s 个变量所在项 $a_{rs}X_s (a_{rs} \neq 0)$，把 $a_{rs}X_s$ 项叫做中心元素；

(2) 为使 $a_{rs}X_s$ 项的系数为"1"，用 $1/a_{rs}$ 遍乘第 r 个方程后，代替第 r 个方程；

(3) 为了使其余方程中不含未知量 X_s，对于 $i = 1, 2, \cdots, (r-1), (r+1) \cdots, m$ 来说，方程 E_i 用 $E_i - \frac{a_{is}}{a_{rs}}E_r$ 代替。

这就使一个非基变量成为基变量，使原来的基变量变为非基变量，这种使一个基变量变成非基变量的变化过程，叫旋转变换，它是单纯形法的基础。

将线性规划模型的目标函数

$$Z = C_1X_1 + C_2X_2 + \cdots + C_nX_n$$

改写成：

$$-Z + C_1X_1 + C_2X_2 + \cdots + C_nX_n = 0$$

的形式，并与约束方程合并，构成一个方程组，该方程组叫增广方程组。

对增广方程组

$$\begin{gathered}
a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n = b_1 \\
a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n = b_2 \\
\cdots \\
a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n = b_m \\
-Z + C_1X_1 + C_2X_2 + \cdots + C_nX_n = 0
\end{gathered}$$

实行旋转变换，可得如下形式：

$$\begin{gathered}
X_1 \quad + \tilde{a}_{1,m+1}X_{m+1} + \tilde{a}_{1,m+2}X_{m+2} + \cdots + \tilde{a}_{1,n}X_n = \tilde{b}_1 \\
X_2 \quad + \tilde{a}_{2,m+1}X_{m+1} + \tilde{a}_{2,m+2}X_{m+2} + \cdots + \tilde{a}_{2,n}X_n = \tilde{b}_2 \\
\cdots \\
X_m + \tilde{a}_{m,m+1}X_{m+1} + \tilde{a}_{m,m+2}X_{m+2} + \cdots + \tilde{a}_{m,n}X_n = \tilde{b}_n \\
-Z + \tilde{C}_{m+1}X_{m+1} + \tilde{C}_{m+2}X_{m+2} + \cdots + \tilde{C}_nX_n = \tilde{Z}
\end{gathered}$$

该形式叫增广标准型。

当令 $X_{m+1} = X_{m+2} = \cdots = X_n = 0$ 时，$X_1 = \tilde{b}_1, X_2 = \tilde{b}_2, \cdots, X_m = \tilde{b}_m$ 为一组基解。

$Z = -\tilde{Z}$ 为目标值；当 $\tilde{b}_1, \tilde{b}_2, \cdots, \tilde{b}_m \geqslant 0$ 时，该组基解为基可行解，它对应于约束集合可行区多胞体的一个极点。

这组基可行解是否为最优解？如果不是，又应怎样由这组基可行解寻找另一组基可行解，使目标值 Z 下降，从而找到最优解呢？现给出最优解的判断定理。

最优解的判断定理为：如果增广标准型的目标函数方程中的所有系数 $\tilde{C}_{m+1}, \tilde{C}_{m+2}, \cdots, \tilde{C}_n$ 均非负，即所有 $\tilde{C}_j \geqslant 0, j = m+1, m+2, \cdots, n$，则该基可行解为最优解(极小解)，

其中 $\tilde{C}_j$ 叫做指标因子，Z 是总指标。

简单地说，就是目标函数中对应于所有非基变量的系数 $\tilde{C}_j \geqslant 0$，则基可行解是惟一的最优解，Z 是极小值；反之，任意一个 $\tilde{C}_j < 0$，则基可行解不是最优解，Z 也不是极小值。

为了深刻地理解判定定理，现直观地证明该定理。

将增广标准型中目标函数方程

$$-Z + \tilde{C}_{m+1}X_{m+1} + \tilde{C}_{m+2}X_{m+2} + \cdots + \tilde{C}_nX_n = \tilde{Z}$$

改写成

$$Z = -\tilde{Z} + \tilde{C}_{m+1}X_{m+1} + \tilde{C}_{m+2}X_{m-2} + \cdots + \tilde{C}_nX_n = -\tilde{Z} + \sum_{j=m+1}^{n} \tilde{C}_jX_j$$

式中 X_j——非基变量。

当令其为零时，才能得到基可行解和极小值。如果其中某个 X_j 不为零，不失一般性，设 $X_{m+1} \neq 0$，由线性规划标准型中所规定的变量非负性知，X_{m+1} 必然大于零，即为正值。

如果 $\tilde{C}_{m+1} > 0$，则 $\tilde{C}_{m+1}X_{m+1} > 0$，$-\tilde{Z} + \tilde{C}_{m+1}X_{m+1} = Z' > Z$；

如果 $\tilde{C}_{m+1} < 0$，则 $\tilde{C}_{m+1}X_{m+1} < 0$，$-\tilde{Z} + \tilde{C}_{m+1}X_{m+1} = Z' < Z$。

由此可见，当 $\tilde{C}_j \geqslant 0$ 时，X_j 的增大会使 Z 增加，所以 X_j 只有为零，才能使 Z 最小，这时 $X_j = 0$ 的假设是对的，所以基可行解是最优解；当 $\tilde{C}_j < 0$ 时，X_j 的增加会使 Z 减少，所以，这种情况下的基可行解不是最优解，Z 也不是最小值。

用该定理可清楚地判断所求的解是否为最优解，如果不是最优解，尚为下一步工作指出了方向，即找出可以调整的未知量——增广标准型中系数 $\tilde{C}_j$ 为负值的变量。因为只有它的变化才能使目标下降。通过对该定理的证明可清楚地说明线性规划标准形式中规定非负条件的必要性，如无非负条件，判定理是不能成立的。

现用上例说明怎样应用单纯形法求解线性规划问题：

$$\begin{aligned} \max \quad & Z = 250X_1 + 500X_2 \\ \text{s.t.} \quad & 20X_1 + 50X_2 \leqslant 1000 \\ & 20X_1 + 10X_2 \leqslant 600 \\ & X_1 \geqslant 0, X_2 \geqslant 0 \end{aligned}$$

引入松弛变量，得增广标准型：

$$20X_1 + 50X_2 + X_3 = 1000 \tag{6.4.3}$$

$$20X_1 + 10X_2 + X_4 = 600 \tag{6.4.4}$$

$$-250X_1 - 500X_2 - Z' = 0 \tag{6.4.5}$$

显然基可行解为 $X_3 = 1000, X_4 = 600, X_1 = X_2 = 0$，目标 $Z' = 0$。因为 X_1, X_2 是非基变量，且其系数为负，所以 $Z' = 0$ 不是最优解，也就是 X_1, X_2 还可以由“0”增大而使 Z' 减少。

现先考查 X_2 的变化。单从目标函数看，X_2 增大的越多，则 Z' 减少的越多。但 X_2 的增大是受限制的。如当 $X_1 = 0, X_2 > 0$ 时，式(6.4.3)、式(6.4.4) 变为：

$$20X_2 + X_3 = 1000 \rightarrow X_3 = 1000 - 50X_2$$

$$10X_2 + X_4 = 600 \rightarrow X_4 = 600 - 10X_2$$

当 X_2 增大到 1000/50 = 20 时，$X_3 = 0$，X_2 再增大，则 $X_3 < 0$，违反了非负条件，所以 X_2 只能由零增大到20；X_2 的增加受 X_3 的限制，同时也受 X_4 的限制，当 X_2 增加到600/10 = 60 时，使 $X_4 = 0$。二者相比较，X_3 的限制比 X_4 的限制要严厉。因此，X_2 只能由"0"增加到20，这是由式(6.4.3)中的 $\theta_1 = 1000/50$ 和式(6.4.4)中的 $\theta_2 = 600/10$ 取最小值决定的。

选 θ 值小的那个方程的 X_2 项作为中心元素进行旋转变换，这个过程叫做换基，即 X_2 进入基矩阵以代替该方程中原来的基变数 X_3 的过程，简称 X_2"进基"、X_3"出基"。

变换后得：

$$\begin{aligned} \frac{2}{5}X_1 + X_2 + \frac{1}{50}X_3 &= 20 \\ 6X_1 \qquad - \frac{1}{5}X_3 + X_4 &= 400 \\ -50X_1 \qquad + 10X_3 - Z' &= 10000 \end{aligned}$$

这时 $X_2 = 20, X_4 = 100, X_1 = X_3 = 0, Z' = -10000$，显然比换基前的目标值小得多。

由此可得出结论，如果得到一组基可行解，经上述换基旋转变换后，所得到的基解也一定是基可行解，其原因是选择进基元素的原则为 $\min\{\theta_i\}$，这时只能使出基元素降为零，而不能变为负值，所以这组解仍为基可行解。

由于非基变量 X_1 的系数 $\tilde{C}_1 = -50$，仍小于零，故 X_1 的增加仍可以使 Z' 进一步下降，按上述分析方法确定进基元素和出基元素。因为，$\theta_1 = \frac{20}{2/5} = 50, \theta_2 = \frac{400}{16} = 25$，所以 X_1 进基，X_4 出基，旋转变换后得

$$\begin{aligned} X_2 + \frac{1}{40}X_3 - \frac{1}{40}X_4 &= 10 \\ X_1 \qquad - \frac{1}{80}X_3 + \frac{1}{16}X_4 &= 25 \\ -Z' + \frac{75}{8}X_3 + \frac{25}{8}X_4 &= 11250 \end{aligned}$$

基可行解为：$X_1 = 25, X_2 = 10, X_3 = X_4 = 0$，$Z' = -11250$。因为 X_3、X_4 的系数 $\tilde{C}_j$ 全部大于零，故 $Z' = -11250$ 是极小值，$Z' = -Z, Z = 11250$ 是极大值。

该算法与上述两种算法的结果是相同的，且只经过二次旋转变化，一共计算了3个极点，就求出了该线性规划的最优解，把计算工作量由计算6个极点减少为计算3个极点，从而减少了计算次数，缩短了计算时间。如果变量数量很多，该方法的优越性就更为明显。可见单纯形法是一种较实用的求解线性规划问题的好方法。

这就是单纯形法求解线性规划问题的基本步骤和原理。

有两种情况必须说明：

(1) 在选择旋转中心时，如果把式(6.4.3)、式(6.4.4)的 $50X_1, 10X_2$ 改成负值，则当 $X_1 = 0$ 时，

$$\begin{aligned} X_3 &= 1000 + 50X_2 \\ X_4 &= 600 + 10X_2 \end{aligned}$$

此时 X_2 为任何值，X_3、X_4 始终大于零，所以当 $\tilde{C}_j < 0$ 的列中 $\bar{a}_{ij}$ 全为负或零时，该线性规划的解无界。

(2) 在有退化解时，假设式(6.4.3)$20X_1 + 50X_2 + X_3 = 0$，如 $X_1 = 0, X_3 = -50X_2$，X_2 增加时，任何 X_2 均使 X_3 为负，故 X_2 只能为零，这时，尽管 Z 不是最优，X_2 也不能再增长，只能进行其他的旋转变换。

现在把非最优基可行解改进过程的一般步骤叙述如下：

如果有 $\tilde{C}_j < 0$，假设不考虑退化情况($\tilde{b}_j = 0$)，则总可以用旋转变换由一组基可行解过渡到另一组使目标函数值更小的基可行解。如果有多个 $\tilde{C}_j < 0$，则要选择最小的 $\tilde{C}_j$ 所对应的变量 X_s(对 Z 影响最大的一个) 即：

$$\tilde{C}_s = \min\{C_j \mid C_j < 0\}$$

这样可以使计算次数减少，节约计算时间。

选定 X_s 以后，使它变成基变量，也就是使它从零增加到下列基变量中某个为零为止

$$X_1 = \tilde{b}_1 - \tilde{a}_{1s}X_s;$$
$$X_2 = \tilde{b}_2 - \tilde{a}_{2s}X_s$$
$$\vdots$$
$$X_m = \tilde{b}_m - \tilde{a}_{ms}X_s$$

如果 $\tilde{a}_{is}$ 均小于零，则 X_s 可以无限增大，由此可以得出下述的解无界定理：

定理：如果在增广标准形中，对某些 s 来说，所有系数 $\tilde{a}_{is}$ 均非正，而 $\tilde{C}_s$ 为负，所构成的可行解使 Z 值无下界。

如果至少有一个 $\tilde{a}_{si}$ 为正，则 X_s 不能无限增大，所达到的值为

$$X_s = \frac{\tilde{b}_r}{\tilde{a}_{rs}} = \min\left[\frac{b_i}{\tilde{a}_{is}}\right] \qquad (\tilde{a}_{is} > 0)$$

这时变量 X_r 成为非基变量(出基)，X_s 变成基变量(进基)。经过以 $\tilde{a}_{is}X_s$ 为中心元素的线性变换，可得到新的标准形，再进行上述循环。由于任何有解的线性规划，其约束所形成的凸多胞形的极点数为有限个，而且从一个极点向另一个极点转移的时候，问题向最优的方向前进而抛掉一些极点，因此如果每次的迭代不出现退化，则在有限步内必定达到最优解。单纯形法的几何意义可用图 6.4.4 表示。

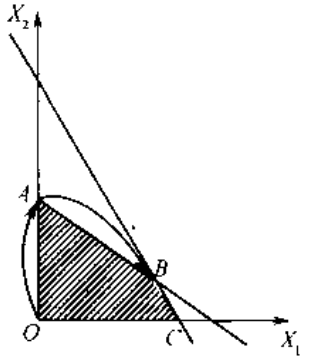


图 6.4.4　单纯形法的几何意义

6.4.6　人工变量法

单纯形原理是在假设增广标准型中 $\tilde{b}_i \geqslant 0$ $(i = 1,2,\cdots,m)$ 的情况下，才得到 $X_i = \tilde{b}_i (i = 1,2,\cdots,m), X_j = 0 (j = m+1, m+2, \cdots, n)$ 这样一组基可行解的。

约束均为“$\leqslant$”类型不等式的线性规划模型，在引入松弛变量后，可直接得到初始基

可行解，进而应用单纯形法可求出最优解。然而，当模型的约束包含等式和"≥"类型不等式时，就很难一下子找出基可行解，因而也就不能应用单纯形法。为了仍能应用单纯形法，需要一种容易求得初始基可行解的方法，这就是人工变量法。

人工变量法的思想是：在初始基可行解不容易求得的情况下，在每个约束方程中加入一个人为的非负变量 —— 人工变量，得到一个包括人工变量在内的基可行解，利用单纯形法，求出最优解，当人工变量全部为零时，所求得的最优解就是原问题的解。

下面就介绍如何引入人工变量和通过人工变量求解线性规划问题的方法。

在约束方程的标准型中加入人工变量：

$$\begin{aligned}
a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + X_{n+1} &= b_1 \\
a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + X_{n+2} &= b_2 \\
&\vdots \\
a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + X_{n+m} &= b_m \\
C_1X_1 + C_2X_2 + \cdots + C_nX_n - Z &= 0 \\
X_j \geqslant 0 \ (j &= 1,2,\cdots,n,n+1,n+2,\cdots,n+m)
\end{aligned}$$

其中，X_{n+1}，X_{n+2}，…，X_{n+m} 是人工变量。这样一来，令 $X_1 = X_2 = \cdots = X_n = 0$，$X_{n+1}$，$X_{n+2}$，…，$X_{n+m}$ 则构成了一组基变量，且 $X_{n+1} = b_1$，$X_{n+2} = b_2$，…，$X_{n+m} = b_m$，$Z = 0$。因为线性规划标准型中规定 $b_i \geqslant 0$，所以这组解是基可行解，因而可应用单纯形法求解了。引入人工变量，虽然可方便地得到初始基可行解，但由于人工变量的引入破坏了原约束的平衡关系，可能导致与原问题不同解。由非负条件知，人工变量 $X_j \geqslant 0$ $(j = n+1, n+2, \cdots, n+m)$，故只有当人工变量 $X_j = 0$ 时，所得最优解才与原问题同解。为了保证所有人工变量为零，现假设一个目标函数：

$$\min \quad W = X_{n+1} + X_{n+2} + \cdots + X_{n+m} = \sum_{j=n+1}^{n+m} X_j$$

如果 W 的极小值为零，则人工变量必然全部为零。如果 W 的极小值不为零，则无法使人工变量为零，原问题无解。

因此引入人工变量后，要首先用单纯形法求 W 的极小值，以判断人工变量是否为零，线性规划问题是否有解。如果人工变量均为零，可去掉所有人工变量和目标函数 W，而得出原问题的基可行解，继续应用单纯形法，即可求出最优解。

综上所述，引入人工变量后的形式是：

$$\begin{aligned}
a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + X_{n+1} &= b_1 \\
a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + X_{n+2} &= b_2 \\
&\vdots \\
a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + X_{n+m} &= b_m \\
C_1X_1 + C_2X_2 + \cdots + C_nX_n - Z &= 0 \\
-W + X_{n+1} + X_{n+2} + \cdots + X_{n+m} &= 0
\end{aligned}$$

整理后得增广标准形为：

$$\begin{aligned}
a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + X_{n+1} &= b_1 \\
a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + X_{n+2} &= b_2
\end{aligned}$$

$$\cdots$$

$$a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + \qquad X_{n+m} = b_m$$

$$-Z + C_1X_1 + C_2X_2 + \cdots + C_nX_n \qquad = 0$$

$$-W - \sum_i^n a_{i1}X_1 - \sum_i a_{is}X_s - \cdots - \sum_i a_{in}X_n \qquad = -\sum_i^m b_i$$

应用单纯形法对此式进行旋转变换即可求解。

该方法又叫二阶段法，即第一阶段先求 W 目标(第二目标)的极小值，当 W 目标的极小值为零时，再进入第二阶段，求原目标(第一目标)的极小值，第一阶段计算的目的是引导出原目标的基可行解。

另一种求解方法叫大 M 法，该方法不必设置第二目标而在原目标上加上人工变量，其系数为很大的数 M(如果为最大目标，其系数为很小的数 $-M$)，其目的是使引入的人工变量对目标不产生影响。

线性规划求解过程见图 6.4.5。

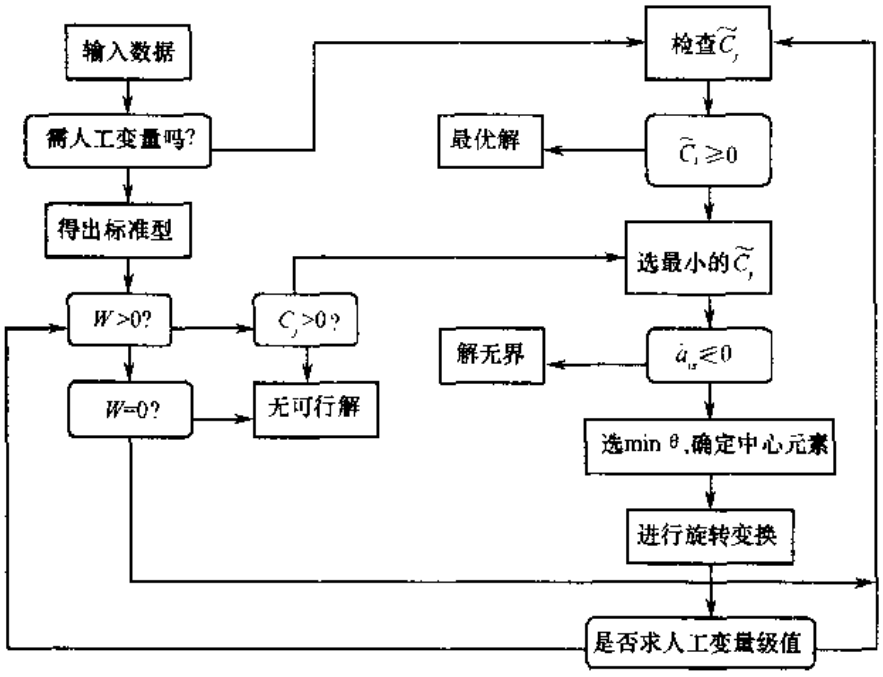


图 6.4.5　单纯形法计算框图

单纯形法计算很麻烦，且容易出错，所以当变量稍多时，可根据单纯形原理编制程序用计算机进行计算，既快、又准确、可以收到很好的效果。也可应用现成的求解线性规划的软件求解。

现举例说明用单纯形法求解线性规划问题的全过程。

[例 2]

$$\min \quad Z = 4X_1 + X_2 + X_3$$

$$\text{s.t.} \quad 2X_1 + X_2 + 2X_3 = 4$$

$$3X_1 + 3X_2 + X_3 = 3$$

$$X_1 \geqslant 0, X_2 \geqslant 0, X_3 \geqslant 0$$

(1) 引入人工变量 X_4, X_5，得到标准型：

$$\begin{aligned}
2X_1 + X_2 + 2X_3 + X_4 &= 4\\
3X_1 + 3X_2 + X_3 + X_5 &= 3\\
4X_1 + X_2 + X_3 - Z &= 0\\
-W + X_4 + X_5 &= 0\\
X_1, X_2, X_3, X_4, X_5 &\geqslant 0
\end{aligned}$$

(2) 进行旋转变换得：

$$\begin{aligned}
2X_1 + X_2 + 2X_3 + X_4 &= 4\\
\boxed{3X_1} + 3X_2 + X_3 + X_5 &= 3\\
5X_1 + X_2 + X_3 - Z &= 0\\
-5X_1 - 4X_2 - 3X_3 - W &= -7
\end{aligned}$$

初始基可行解为：$X_1 = X_2 = X_3 = 0, X_4 = 4, X_5 = 3$；目标值 $Z = 0$；人工变量目标的极小值为 $W = 7$。

(3) 进行旋转变换求 W 的极小值：

因为 X_1 的系数为最小的负数，故 X_1 进基，而 $\frac{3}{3} = 1 < \frac{4}{2} = 2$，故 X_5 出基。$3X_1$ 为旋转中心，经旋转变换得：

$$\begin{aligned}
-X_2 + \boxed{\frac{4}{3}X_3} + X_4 - \frac{2}{3}X_5 &= 2\\
X_1 + X_2 + \frac{1}{3}X_3 + \frac{1}{3}X_5 &= 1\\
-3X_2 - \frac{1}{3}X_3 - \frac{4}{3}X_5 - Z &= -4\\
X_2 - \frac{4}{3}X_3 + \frac{5}{3}X_5 - W &= -2
\end{aligned}$$

X_3 进基，X_4 出基，以 $\frac{4}{3}X_3$ 为旋转中心，再进行放置变换得：

$$\begin{aligned}
-\frac{3}{4}X_2 + X_3 + \frac{4}{3}X_4 - \frac{1}{2}X_5 &= \frac{3}{2}\\
X_1 + \boxed{\frac{5}{4}X_2} - \frac{1}{4}X_4 + \frac{1}{2}X_5 &= \frac{1}{2}\\
-\frac{13}{4}X_2 + \frac{1}{4}X_4 - \frac{3}{2}X_5 - Z &= -\frac{7}{2}\\
X_4 + X_5 - W &= 0
\end{aligned}$$

此时最优解为 $X_1 = \frac{1}{2}$，$X_3 = \frac{3}{2}$，$X_2 = X_4 = X_5 = 0$；目标值为：$Z = \frac{7}{2}$；人工变量之和为 $W = 0$，满足引入人工变量的条件，该问题有解。去掉人工变量和人工变量之和所构成的目标函数 W，得到不含人工变量的基可行解。由此可见，引入人工变量的目的仅仅是为了得到原问题的基可行解。

(4) 应用单纯形原理，求原问题最优解：

因目标函数非基变量的系数中仍有负值，故 X_2 进基，X_1 出基，以 $\frac{5}{4}X_2$ 为旋转中心再

进行旋转变换得：

$$\frac{3}{5}X_1 + X_3 = \frac{9}{5} \qquad \frac{4}{5}X_1 + X_2 = \frac{2}{5} \qquad \frac{13}{5}X_1 - Z = -\frac{11}{5}$$

最优解为：$X_1 = 0, X_2 = \frac{2}{5}, X_3 = \frac{9}{5}$；目标值为：$Z = \frac{11}{5}$。

一个线性规划问题，应用电子计算机可以很快地求出最优解。但对于变量不十分多，而又无计算机的情况，往往不采用上述形式的解法，而采用上述原理在表格上进行计算，该表格叫单纯形表。用单纯形表计算比较清楚，不易出错。当前，单纯形表的形式比较多，但原理均相同。

上例可把系数列成表 6.4.9 ～ 表 6.4.12 进行计算。

表 6.4.9　系数

$C_Z \rightarrow$				4	1	1	0	0
$C_W \rightarrow$				0	0	0	1	1
变量				非基变量			基变量	
				X_1	X_2	X_3	X_4	X_5
C_B^W	C_B^Z	$-W$	-7	-5	-4	-3	0	0
		$-Z$	0	4	1	1	0	0
1	0	X_4	4	2	1	2	1	0
1	0	X_5	3	③	3	1	0	1

表 6.4.10　系数

$C_Z \rightarrow$				0	1	1	0	4
$C_W \rightarrow$				1	0	0	1	0
变量				非基变量			基变量	
				X_5	X_2	X_3	X_4	X_1
C_B^W	C_B^Z	$-W$	-2	$\frac{5}{3}$	1	$-\frac{4}{3}$	0	0
		$-Z$	-4	$-\frac{4}{3}$	-3	$-\frac{1}{3}$	0	0
1	0	X_4	2	$-\frac{2}{3}$	1	$\boxed{\frac{4}{3}}$	1	0
0	4	X_1	1	$\frac{1}{3}$	1	$\frac{1}{3}$	0	1

单纯形表的计算方法是：将所有已知数字填入表格后，按上述方法找出旋转中心并用圈圈起，如表6.4.9 中的 ③。进行旋转变换后得表 6.4.10。表 6.4.10 中 W、Z 两行数字可计算出，其计算方法是：W、Z 行的第一列表示目标值，是 C_B^W 和 C_B^Z 两列数分别与基变量的值之积的和并取负值，即 $C_B^W X_B$、$C_B^Z X_B$，该两行的其他列为判断数 $C_j - \sum_i C_i a_{ij}$。如表 6.4.10 中 $-W = -(1\times 2 + 0\times 1) = -2$；$-Z = -(0\times 2 + 4\times 1) = -4$；非基变量 X_5 对应的目标 W 的判断数为：$1 - \left[1\times\left(-\frac{2}{3}\right) + 0\times\frac{1}{3}\right] = \frac{5}{3}$；目标 Z 的判断数为：$0 - \left[0\times\left(-\frac{2}{3}\right) + 4\times\frac{1}{3}\right] = -\frac{4}{3}$；其他算法均相同。这样用四张表即可完成计算。求解结果

表 6.4.11　系数

$C_Z \rightarrow$				0	1	0	1	4
$C_W \rightarrow$				1	0	1	0	0
变量				非基变量			基变量	
				X_5	X_2	X_4	X_3	X_1
C_B^W	C_B^Z	$-W$	0	1	0	1	0	0
		$-Z$	$-\frac{7}{2}$	$-\frac{3}{2}$	$-\frac{13}{4}$	$\frac{1}{4}$	0	0
0	1	X_3	$\frac{3}{2}$	$-\frac{1}{2}$	$-\frac{3}{4}$	$\frac{3}{4}$	1	0
0	4	X_1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$-\frac{1}{4}$	0	1

表 6.4.12　系数

$C_Z \rightarrow$				4	1	1
变量				非基变量	基变量	
				X_1	X_3	X_2
	C_B^Z	$-Z$	$-\frac{11}{5}$	$\frac{13}{5}$	0	0
	1	X_3	$\frac{9}{5}$	$\frac{3}{5}$	1	0
	1	X_2	$\frac{2}{5}$	$\frac{4}{5}$	0	1

与上法相同。

6.5　灵敏度分析和参数规划

6.5.1　单纯形法的数学描述

上述单纯形法是在线性变换的基础上用代数方法描述的，其目的是借助于线性变换加深对单纯形法的理解。为了便于对线性规划模型的求解进行分析，现用矩阵理论对单纯形法进行描述。

对于线性规划标准型：

$$\begin{aligned} \min \quad & \boldsymbol{Z} = \boldsymbol{CX} \\ \text{s.t.} \quad & \boldsymbol{AX} = \boldsymbol{b} \\ & \boldsymbol{X} \geqslant 0 \end{aligned} \tag{6.5.1}$$

设 $\boldsymbol{A}$ 为 $m \times n$ 阶非奇异矩阵，$\boldsymbol{B}$ 为 $\boldsymbol{A}$ 矩阵中一个非奇异的 $m \times m$ 阶基矩阵，利用矩阵分块原理将 $\boldsymbol{A}$、$\boldsymbol{X}$、$\boldsymbol{C}$ 写成分块形式：

$$\begin{aligned} \min \quad & \boldsymbol{Z} = (\boldsymbol{C}_B \quad \boldsymbol{C}_N) \cdot \begin{pmatrix} \boldsymbol{X}_B \\ \boldsymbol{X}_N \end{pmatrix} \\ \text{s.t.} \quad & (\boldsymbol{B} \quad \boldsymbol{A}_N) \cdot \begin{pmatrix} \boldsymbol{X}_B \\ \boldsymbol{X}_N \end{pmatrix} = \boldsymbol{b} \end{aligned} \tag{6.5.2}$$

$$\begin{pmatrix} X_B \\ X_N \end{pmatrix} \geqslant 0$$

式(6.5.2)中的约束条件可写成：

$$\boldsymbol{B}X_B + \boldsymbol{A}_N X_N = b$$

两边同左乘 B^{-1} 并整理得：

$$\boldsymbol{X}_B = \boldsymbol{B}^{-1}b - \boldsymbol{B}^{-1}\boldsymbol{A}_N\boldsymbol{X}_N \tag{6.5.3}$$

目标函数可写成：

$$\boldsymbol{Z} = (\boldsymbol{C}_B \quad \boldsymbol{C}_N) \cdot \begin{pmatrix} \boldsymbol{X}_B \\ \boldsymbol{X}_N \end{pmatrix} = \boldsymbol{C}_B\boldsymbol{X}_B + \boldsymbol{C}_N\boldsymbol{X}_N$$

将式(6.5.3)代入得：

$$\begin{aligned} \boldsymbol{Z} &= \boldsymbol{C}_B(\boldsymbol{B}^{-1}b - \boldsymbol{B}^{-1}\boldsymbol{A}_N\boldsymbol{X}_N) + \boldsymbol{C}_N\boldsymbol{X}_N = \\ &\boldsymbol{C}_B\boldsymbol{B}^{-1}b - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}_N\boldsymbol{X}_N + \boldsymbol{C}_N\boldsymbol{X}_N = \\ &\boldsymbol{C}_B\boldsymbol{B}^{-1}b + (\boldsymbol{C}_N - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}_N)\boldsymbol{X}_N \end{aligned} \tag{6.5.4}$$

由式(6.5.3)、式(6.5.4)可见，当 $\boldsymbol{X}_N = 0$ 时，

$$\boldsymbol{X}_B = \boldsymbol{B}^{-1}b;\boldsymbol{Z} = \boldsymbol{C}_B\boldsymbol{B}^{-1}b$$

如果 $\boldsymbol{B}^{-1}b \geqslant 0$，则这组解为基可行解；在目标函数中，$\boldsymbol{X}_N$ 的系数$(\boldsymbol{C}_N - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}_N) \geqslant 0$，可使 $\boldsymbol{Z}$ 取得出小值，即 $\boldsymbol{Z} = \boldsymbol{C}_B\boldsymbol{B}^{-1}b$。所以

$$(\boldsymbol{C}_N - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}_N) \geqslant 0 \tag{6.5.5}$$

为解的最优性判断准则。又因为：

$$\begin{aligned} \boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} &= (\boldsymbol{C}_B \quad \boldsymbol{C}_N) - \boldsymbol{C}_B\boldsymbol{B}^{-1}(\boldsymbol{B} \quad \boldsymbol{A}_N) = \\ &(\boldsymbol{C}_B \quad \boldsymbol{C}_N) - (\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{B} \quad \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}_N) = \\ &(\boldsymbol{C}_B \quad \boldsymbol{C}_N) - (\boldsymbol{C}_B\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}_N) = \\ &(0 \quad \boldsymbol{C}_N - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}_N) \end{aligned}$$

所以 $\boldsymbol{C}_N - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}_N$ 只是$\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}$ 中非基变量的系数。而 $\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}$ 中基变量的系数均为零。这样一来包括基变量系数在内的最优解判断准则可用下式描述：

$$\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} \geqslant 0 \tag{6.5.6}$$

如果 $\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}$ 中第 j 列元素小于零，则令 X_j 进基，选

$$\min\{\theta\} = \min\left\{\frac{(B^{-1}b)_i}{(B^{-1}P_j)_i} \;\middle|\; B^{-1}P_j > 0\right\} = \frac{(B^{-1}b)_l}{(B^{-1}P_j)_l} \tag{6.5.7}$$

其中，$\boldsymbol{P}_j$ 为$\boldsymbol{A}$ 矩阵中第j 列元素的行所对应的基变量$\boldsymbol{X}_j$ 为出基变量。这就可得到一个新基。对 $\boldsymbol{B}^{-1}\boldsymbol{A}$ 为$\boldsymbol{A}$ 矩阵、$\boldsymbol{B}^{-1}\boldsymbol{b}$ 为$\boldsymbol{b}$ 向量所构成的线性规划问题，再以该新基进行上述计算，可在有限步内求出使目标函数达到极小值的最优解。其最优解所对应的基为最优基，因此当 B 为最优基时，必然有 $\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} \geqslant 0$，否则解无界。

单纯形法的矩阵描述和线性变换描述有如表 6.5.1 所列的对应关系。

表 6.5.1　单纯形法矩阵描述和线性变换描述对照表

	矩阵描述	线性变换描述
基　解	$\boldsymbol{B}^{-1}\boldsymbol{b}$	$\bar{b}$
目标值	$\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b}$	$\bar{Z}$
判断准则	$\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}\geqslant 0$	$\bar{C}_j\geqslant 0$
变换后的约束系数矩阵	$\boldsymbol{B}^{-1}\boldsymbol{A}$	$\bar{a}_{ij}=\begin{cases}1 & i=j\\ 0 & i,j=1,2,\cdots,m \quad \text{且} j\neq i\\ \bar{a}_{ij} & i=1,2,\cdots,m \quad j=m+1,m+2,\cdots,n\end{cases}$

通过对单纯形法的矩阵描述，可简化变换的叙述，为灵敏度分析和参数规划的介绍创造方便条件。

6.5.2　灵敏度分析

灵敏度分析是对线性规划求解结果进行分析的一种方法。单纯形法是在参数 A、C、b 确定的情况下求得最优解的，然而在实践中，参数 A、C、b 都受各种客观条件的限制且带有一定的随机性，当客观条件发生变化时，它们可能会发生变化。如 C 可能因市场的变动而变化，b 可能会因资源条件的变化而波动，A 会因技术进步的影响而改变等。那么，这些条件的变化会使所得到的最优解发生什么变化呢？如果欲使最优解不变，这些因素的变动范围又是多少呢？这就是灵敏度分析要解决的问题之一。

单纯形法是在变量数和约束条件数确定的条件下求解的。但实践中会经常出现新情况，这些新情况的出现会使变量数增加，也可能会增加约束条件，那么这些情况的出现对系统的影响如何呢？这也是灵敏度分析的问题之一。

下面分别就上述几种问题进行研究。

6.5.2.1　对目标函数系数的灵敏度分析

目标函数某个系数的变化是否对已求得的最优解发生影响取决于判断数 $\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}$ 是否仍满足判定准则。当系数的变化仍能满足判断准则时，最优解不变，否则最优解改变。因此目标函数系数的变化对最优解的影响主要应从判断准则来分析。

以矩阵表示的最优解判断准则为

$$\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}\geqslant 0$$

设目标函数第 i 个变量的系数 C_i 增加 ΔC_i，将 $C_i+\Delta C_i$ 代入上式，即可确定满足最优解判定条件的变动范围 ΔC_i。

现以 6.4 节中例 1 说明这种情况下的灵敏度分析。

该例最优基 $\boldsymbol{B}=\begin{pmatrix}50 & 20\\ 10 & 20\end{pmatrix}$；$\boldsymbol{C}_B=(-500 \quad -250)$；

$$\boldsymbol{B}^{-1}\boldsymbol{A}=\begin{pmatrix}0 & 1 & \frac{1}{40} & -\frac{1}{40}\\ 1 & 0 & -\frac{1}{80} & \frac{1}{16}\end{pmatrix} \tag{6.5.8}$$

如果市场发生变化影响到 X_1 产品的利润，现研究使最优解不变的 C_1 取值的范围。

将 $C' = C_1 + \Delta C_1 = -(250 + \Delta C_1)$ 代入式(6.5.6)

$$\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} = [-(250 + \Delta C_1) \quad -500 \quad 0 \quad 0] - [-500 \quad (250 + \Delta C_1)] \begin{pmatrix} 0 & 1 & \frac{1}{40} & -\frac{1}{40} \\ 1 & 0 & -\frac{1}{80} & \frac{1}{16} \end{pmatrix} =$$

$$\left(0 \quad 0 \quad \frac{75}{8} - \frac{\Delta C_1}{80} \quad \frac{25}{8} + \frac{\Delta C_1}{16}\right) \geqslant 0$$

$$\frac{75}{8} - \frac{\Delta C_1}{80} \geqslant 0 \qquad \Delta C_1 \leqslant 750$$

$$\frac{25}{8} + \frac{\Delta C_1}{16} \geqslant 0 \qquad \Delta C_1 \geqslant -50$$

从而得出，当 C_1 的变动范围为 $-50 \leqslant \Delta C_1 \leqslant 750$ 时最优解不变。

用图解法来说明该问题(见图 6.4.3)，目标函数系数变化导致目标函数等值线的斜率发生变化，当 X_1 产品的利润达 $250 + 750 = 1000$ 时，目标函数的等值线与电力约束线平行，当 X_1 产品的利润降到 $250 - 50 = 200$ 时，目标函数的等值线与钢材约束线平行，故利润在二者之间最优解不变。当 X_1 产品的利润超过 1000 元时，最优点由 C 移到 D，当 X_1 产品的利润低于 200 元时，最优点由 C 转移到 B。因此，这类问题的灵敏度分析实质上是确定目标函数等值线或等值面在最优点附近的变动范围问题。

6.5.2.2 约束条件常数项变动的灵敏度分析

线性规划模型约束条件中的常数项通常表示某种资源量的限制。实践中各种资源经常因某些不确定因素的影响而变动，当资源量变化的时候，最优解通常也跟着发生改变。然而，确定最优基不变情况下的资源量的变动范围和相应的最优解对系统控制是有重要意义的。对约束条件常数项的灵敏度分析就是在不改变基变量的情况下，确定某些资源量的变动范围及相应最优解的一种方法。

因为 b 的变动并不影响判断准则，只影响最优解 $\boldsymbol{B}^{-1}\boldsymbol{b}$，所以当 $\boldsymbol{b}_i$ 增加 Δb_i 时，若使最优基不变，只需使基可行解仍满足 $\boldsymbol{B}^{-1}\boldsymbol{b} \geqslant 0$ 即可，也就是仍旧保持 $\boldsymbol{B}^{-1}\boldsymbol{b}$ 为基可行解。

如上例中，原最优基为

$$\boldsymbol{B} = \begin{pmatrix} 50 & 20 \\ 10 & 20 \end{pmatrix} \qquad \boldsymbol{B}^{-1} = \begin{pmatrix} -\frac{1}{40} & -\frac{1}{40} \\ -\frac{1}{80} & \frac{1}{16} \end{pmatrix}$$

如果当 b_1 的变动量为 Δb_1 时，为使最优基不变应满足：

$$\boldsymbol{B}^{-1}\boldsymbol{b} = \begin{pmatrix} -\frac{1}{40} & -\frac{1}{40} \\ -\frac{1}{80} & \frac{1}{16} \end{pmatrix} \cdot \begin{pmatrix} 100 + \Delta b_1 \\ 600 \end{pmatrix} \geqslant 0$$

整理得不等式：

$$10 + \frac{1}{40}\Delta b_1 \geqslant 0$$

$$25 - \frac{1}{800}\Delta b_1 \geqslant 0$$

解不等式得：$-400 \leqslant \Delta b_1 \leqslant 2000$

这说明当钢材限制在 $600 \leqslant b_1 \leqslant 3000$ 时，X_1、X_2 仍为最优基，其最优解由 $\boldsymbol{B}^{-1}\boldsymbol{b}$ 所决定。这就解决了由资源量的变动范围估计最优基是否变化以及不必重新计算即可确定最优解的问题。

用图解法来说明该问题则是约束条件中常数项的变化，导致可行区的变化，在钢材约束的常数项由1000下降到600时，最优点由 C 点转移到 D 点，若钢材资源量超过该范围，最优基就发生了变化，即只生产甲产品或只生产乙产品。

6.5.2.3 约束条件系数矩阵中某些元素变化的灵敏度分析

如果约束条件系数矩阵中非基变量的某元素发生变化，在所有判断数仍满足判断准则时，基矩阵不变，故最优解仍为 $\boldsymbol{B}^{-1}\boldsymbol{b}$。因此通过计算判定条件 $\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} \geqslant 0$ 就可确定 A 矩阵中非基变量的系数变动范围。

如6.4节例2中，X_1 为非基变量，设 a_{11} 的变动为 $a_{11}+\Delta a_{11}$，则：

$$\boldsymbol{C}_B\boldsymbol{B}^{-1} = (1 \quad 1)\begin{pmatrix} \frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{5} & \frac{2}{5} \end{pmatrix} = \begin{pmatrix} \frac{2}{5} & \frac{1}{5} \end{pmatrix}$$

$$\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} = (4 \quad 1 \quad 1) - \begin{pmatrix} \frac{2}{5} & \frac{1}{5} \end{pmatrix}\begin{pmatrix} 2+\Delta a_{11} & 1 & 2 \\ 3 & 3 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} \frac{13}{5} - \frac{2}{5}\Delta a_{11} & 0 & 0 \end{pmatrix} \geqslant 0$$

由此得不等式：

$$\frac{13}{5} - \frac{2}{5}\Delta a_{11} \geqslant 0,\ \Delta a_{11} \leqslant 6.5$$

即当 a_{11} 的变动小于6.5时，即可保证最优解不变。

如果约束矩阵中基变量的系数变化，$\boldsymbol{B}^{-1}$ 则随之变化，最优解 $\boldsymbol{B}^{-1}\boldsymbol{b}$ 也随之变化。但为使最优基不变，仍需使判断数满足判别条件，即使 $\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} \geqslant 0$ 成立，从而可确定系数的变动范围。但由于这种计算很复杂，一般常给定确定的变动量来计算判断数是否满足判别条件或将变动系数的变量作为一个新增加的变量处理。

6.5.2.4 增加新变量的灵敏度分析

在线性规划模型中，如果增加一个新变量，判断该新变量对最优解是否有影响或确定变量为何值时才能对最优解产生影响的问题是对线性规划增加新变量的灵敏度分析问题。

线性规划模型增加一个新变量 $\boldsymbol{X}_K$ 等于在约束矩阵 $\boldsymbol{A}$ 中增加一列 $\boldsymbol{P}_K$，目标函数系数向量增加一个元素 $\boldsymbol{C}_K$。如该新增加的变量对最优基 $\boldsymbol{B}$ 无影响，则判断数仍需满足 $\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} \geqslant 0$，更确切地说，新增变量应为非基变量，其判断数应大于等于零，即

$$\boldsymbol{C}_K - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_K \geqslant 0$$

如果当 $\boldsymbol{C}_K - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_K < 0$，$\boldsymbol{X}_K$ 可成为基变量而替换原基变量。故解不等式 $\boldsymbol{C}_K - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{P}_K < 0$ 就可求出 $\boldsymbol{X}_K$ 成为基变量的范围。

如6.4节例1中，现准备生产一种新产品3，该产品材料消耗量为30kg/件，电力消耗量为30(°)/件，利润额为350元/件。现考查该产品是否有利投产；如不利投产，当利润为

何值时，才有利投产。

对于原问题，$X_1 = 25, X_2 = 10$ 为最优解，现令该新产品产量为 X_5，当引入新变量之后，$X_1 = 25, X_2 = 10, X_5 = 0$ 仍为基可行解。但该基可行解是否为最优解必须依据 $\boldsymbol{C} - \boldsymbol{C_B B^{-1} A} \geqslant 0$ 来判断，因原问题只增加了一个变量，其他因素未变，故计算判断数只需计算 $\boldsymbol{C_5} - \boldsymbol{C_B B^{-1} P_5}$（$\boldsymbol{P_5}$ 为 $\boldsymbol{A}$ 矩阵中新增加的第五列）。

将数据代入得：

$$-350 - (-500 \quad -250) \cdot \begin{pmatrix} \frac{1}{40} & -\frac{1}{40} \\ -\frac{1}{80} & \frac{1}{16} \end{pmatrix} \cdot \begin{pmatrix} 30 \\ 30 \end{pmatrix} = 25 > 0$$

所以最优解为：$X_1 = 25, X_2 = 10, X_5 = 0$。这说明引入新变量并不改变最优解，也说明该产品不利投产。

只有在 $C_5 - C_B B^{-1} P_5 < 0$ 时，引入的新变量才改变原问题的解。因此，当满足式 $C_K < C_B B^{-1} P_K$ 时，新变量才能成为基变量，求解该不等式求出 C，即可确定新变量成为基变量的范围。

本例中 $C_5 < -375$，即当新产品的单位利润超过 375 元 / 件时，才有利于投产。

6.5.2.5 增加新约束的灵敏度分析

一个系统经常因客观条件的变化需增加一些约束。因此需分析当客观条件变化时对系统的影响。分析增加新的约束条件对系统最优解的影响是另一种灵敏度分析。

一个线性规划模型在增加一个约束方程之后，原基矩阵就不再是单位阵了，因此就必须对该约束实行线性变换使基矩阵成为单位阵。如果经线性变换后，新增约束的松弛变量或人工变量是基变量，则原问题的最优解不变，否则需按对偶单纯形法进行变换得到引入新约束后的最优解。

如在 6.4 节例 1 中增加一项工时约束，即产品 1 生产需 10h/ 件，产品 2 生产为 15h/ 件，现有工时 400h，现研究增加该约束后最优解将如何变化。

新增加的约束方程为：

$$10X_1 + 15X_2 + X_5 = 400$$

在原最优解的基础上引入新约束后的系数为：

	X_1	X_2	X_3	X_4	X_5	b
	0	1	$\frac{1}{40}$	$-\frac{1}{40}$	0	10
	1	0	$-\frac{1}{80}$	$\frac{1}{16}$	0	25
	10	15	0	0	1	400
$-Z$	0	0	$\frac{75}{8}$	$\frac{25}{8}$	0	11250

经变换得新约束中的松弛变量为基变量，故新增加的工时约束对原最优解无影响。

6.5.3 参数规划

在进行系统分析的时候，影响系统的各种因素往往随客观条件的变化而变化。如果客

观条件的变化可估计时，应用灵敏度分析可确定这种变化对系统的影响，然而有时客观条件的变化不容易估计，这就需要了解客观条件在某个范围内时系统的变化。如在制定生产计划时，由于某些原材料的价格发生变化，而使以利润为目标的目标函数系数 C 随某个参数（原材料价格变化的百分比等）而变化，原材料的资源量因某些条件的变化增加或减少，使得约束条件中右端的常数项也随某个参数变化，这时我们关心的是参数在某个范围内时系统的最优解是什么，以便在资源和利润等客观因素一旦确定了的情况下不必用单纯形法进行复杂的计算就能很方便地得出最优解，很快地为决策提供信息，这对加快决策速度，提高决策能力，节约计算时间是很有益的。解决这类问题的方法就是参数规划法。

所谓参数规划法是指确定系统模型中参数变化的范围及最优解的数学规划法，它与灵敏度分析不同，灵敏度分析是在已确定的条件下求出最优解，并研究当各种因素变化时系统的变化或研究系统在确定状态下，各因素的变化范围，而参数规划是在不确定变动的情况下求出各种情况的最优解，以便得到当不确定因素一旦有确定值时系统的最优解。因此二者解决的问题实质上是一个问题的两个侧面。由于参数规划比较复杂，本节主要介绍单参数的参数规划问题。

6.5.3.1 目标函数含有参数的线性规划问题

在一组确定的约束条件下，系统目标函数的系数，在不确定因素的影响下可能随某一参数在某个范围内的变化。为解决这类问题，原系统模型可描述为

$$\begin{aligned} \min \quad & \boldsymbol{Z} = (\boldsymbol{C} + \lambda \boldsymbol{C}^*)\boldsymbol{X} = \boldsymbol{C}\boldsymbol{X} + \lambda \boldsymbol{C}^* \boldsymbol{X} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{X} = \boldsymbol{b} \\ & \boldsymbol{X} \geqslant 0 \end{aligned}$$

其中，λ 为实数参数，$\boldsymbol{C}^* = (C_1^*, C_2^*, \cdots, C_n^*)$ 为 λ 的系数向量。

由目标函数可见，该问题的目标是由不带参数的 $\boldsymbol{C}\boldsymbol{X}$ 和带参数 $\lambda \boldsymbol{C}^* \boldsymbol{X}$ 的两部分组成，其求解方法可把 λ 当作已知值来处理。

由单纯形法知，若 $\boldsymbol{B}$ 为基矩阵，则 $\boldsymbol{X}_B = \boldsymbol{B}^{-1}\boldsymbol{b}$ 为基可行解，当 $\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} \geqslant 0$ 时，这组解为最优解。

对于该参数规划问题，判别准则可写成：

$$\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A} + \lambda(\boldsymbol{C}^* - \boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{A}) \geqslant 0 \tag{6.5.9}$$

由式(6.5.9)可见，它由两部分组成，一部分不带参数，一部分带参数。

由于判断数中含有参数，在参数未给定的情况下，无法直接判断问题的最优解，故将判断准则用参数 λ 表示。

由式(6.5.9)可得如下关系：

当 $[\boldsymbol{C}^* - \boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{A}]_j > 0$ 时，
$$\lambda \geqslant \frac{[\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}]_j}{[\boldsymbol{C}^* - \boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{A}]_j} \tag{6.5.10}$$

当 $[\boldsymbol{C}^* - \boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{A}]_j < 0$ 时，
$$\lambda \leqslant \frac{[\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}]_j}{[\boldsymbol{C}^* - \boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{A}]_j} \tag{6.5.11}$$

由式(6.5.10)、式(6.5.11)可知，当 λ 同时满足该两个不等式，即在二者的交集中时，$\boldsymbol{X} = \boldsymbol{B}^{-1}\boldsymbol{b}$ 为最优解，这样最优解的判断准则就可用 λ 来表示。

如果令

$$\underline{\lambda}_B = \begin{cases} \max\left\{-\dfrac{[\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}]_j}{[\boldsymbol{C}^*-\boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{A}]_j}\middle|(\boldsymbol{C}^*-\boldsymbol{C}_B^*\boldsymbol{b}^{-1}\boldsymbol{A})_j>0\right\} \\ -\infty \qquad\qquad (\boldsymbol{C}^*-\boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{A})_j \leqslant 0 \end{cases}$$

$$\bar{\lambda}_B = \begin{cases} \min\left\{-\dfrac{[\boldsymbol{C}-\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A}]_j}{[\boldsymbol{C}^*-\boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{A}]_j}\middle|(\boldsymbol{C}^*-\boldsymbol{C}_B^*\boldsymbol{b}^{-1}\boldsymbol{A})_j<0\right\} \\ \infty \qquad\qquad (\boldsymbol{C}^*-\boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{A})_j \geqslant 0 \end{cases} \tag{6.5.12}$$

则 λ 在区间$[\underline{\lambda}_B,\bar{\lambda}_B]$有最优解 $X_B = B^{-1}b$。

$\underline{\lambda}_B$ 和$\bar{\lambda}_B$为基B 的下特征数和上特征数。区间$[\underline{\lambda}_B,\bar{\lambda}_B]$为最优区间,区间内每个 λ 对应的解都是最优解 $\boldsymbol{B}^{-1}\boldsymbol{b}$,目标函数极值为:

$$\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b} + \lambda\boldsymbol{C}_B^*\boldsymbol{B}^{-1}\boldsymbol{b}(\text{为 } \lambda \text{ 的函数})$$

如果 λ 不在区间$[\underline{\lambda}_B,\bar{\lambda}_B]$内,则说明判断数不满足判断准则,基变量不是最优解,需要进行换基迭代来寻求最优解。

当 $\lambda > \bar{\lambda}_B$ 时,λ 所对应的非基变量进基,相对应的基变量出基,按单纯形法的旋转变换进行换基迭代。

当 $\lambda < \underline{\lambda}_B$ 时,λ 所对应的非基变量进基,相对应的基变量出基,也按旋转变换进行迭代直到所有 λ 都在区间$[\underline{\lambda}_B,\bar{\lambda}_B]$内为止。

现举例说明这种参数规划的计算方法:

假设 6.4 节例 1 中,甲乙两种产品的利润随原材料的价格的变化而变化,设 λ 为原材料价格的变动幅度,且第一种产品利润与原材料价格的关系是 $250+\lambda$,第二种产品利润与原材料价格的关系是 $500+2\lambda$(原材料价格上涨时,λ 为负;反之为正)。该问题用一般的线性规划不能求解,但应用参数规划可确定 λ 在某一范围时的最优解。

问题模型为

$$\begin{aligned} \max\quad & Z = (250+\lambda)X_1 + (500+2\lambda)X_2 \\ \text{s.t.}\quad & 20X_1 + 50X_2 \leqslant 1000 \\ & 20X_1 + 10X_2 \leqslant 600 \\ & X_1, X_2 \geqslant 0 \end{aligned}$$

变成参数规划的标准形式为

$$\begin{aligned} \min\quad & Z' = -(250+\lambda)X_1 - (500+2\lambda)X_2 = \\ & \qquad -250X_1 - 500X_2 - \lambda X_1 - 2\lambda X_2 \\ \text{s.t.}\quad & 20X_1 + 50X_2 + X_3 = 1000 \\ & 20X_1 + 10X_2 + X_4 = 600 \\ & X_1, X_2, X_3, X_4 \geqslant 0 \end{aligned}$$

用单纯形表求解(见表 6.5.2)。

表 6.5.2　用单纯形表求解

$\boldsymbol{X}$	X_1	X_2	X_3	X_4
$\boldsymbol{C}$	−250	−500	0	0
$\boldsymbol{C}^*$	−1	−2	0	0

（续）

λ				-250	$\overline{-250}$		
C_B^*	C_B	Z' Z'_λ	0 0	-250 1	-500 -2		
0 0	0 0	X_3 X_4	1000 600	20 20	⑤⓪ 10	1 0	0 1

$\because C^* - C_B^* B^{-1} A < 0$，　$\therefore \underline{\lambda}_B = -\infty$

$\bar{\lambda}_B = \min\left\{\left(-\frac{-250}{-1}\right), -\left(\frac{-500}{-2}\right)\right\} = -250$，故 λ 的最优区间为：$(-\infty, -250]$，λ 在该区间内有最优解为 $X_3 = 1000, X_4 = 600$，最小值为 $Z' = 0 + 0\lambda = 0$。

如 λ 不在区间内，即 $\lambda > -250$，超过上限 $\bar{\lambda}_B$，则 λ_B 所对应变量的判断数为负，故应进基，本例 X_2 中为进基元素，选 $\min\theta = \min\left\{\frac{1000}{50}, \frac{600}{10}\right\} = 20$ 所在行的基变量出基，即 X_3 出基，以 50 为旋转中心，进行旋转变换后得单纯形表 6.5.3。

表 6.5.3　单纯形表

X				X_1	X_2	X_3	X_4
C				-250	-500	0	0
C^*				-1	-2	0	0
λ				$\overline{-250}$		$\underline{-250}$	
C_B^*	C_B	Z' Z'_λ	-10000 -40	-50 $-\frac{1}{5}$	0 0	10 $\frac{2}{50}$	0 0
-2 0	-500 0	X_2 X_4	20 400	$\frac{2}{5}$ ⑯	1 0	$\frac{1}{50}$ $-\frac{1}{5}$	0 1

$$\underline{\lambda}_B = \max\left\{-\frac{10}{\frac{2}{50}}\right\} = -250 \quad \bar{\lambda}_B = \min\left\{-\frac{-50}{-\frac{1}{5}}\right\} = -250$$

所以最优区间为：$[-250, -250]$ 此时最优解为：$X_2 = 20, X_4 = 400$，极小值为 $Z' = -10000 + (-40) \times (-250) = 0$，这说明生产所得到的收入全部弥补材料涨价的损失。

当 $\lambda > -250$ 时，即超过上限时，变量 X_1 的判断数为负，故 X_1 进基，相应的基变量 X_4 出基，16 为旋转中心，进行旋转变换的结果见表 6.5.4。

表 6.5.4　旋转变换结果

X				X_1	X_2	X_3	X_4
C				-250	-500	0	0
C^*				-1	-2	0	0
λ						$\overline{-250}$	$\underline{-250}$
C_B^*	C_B	Z Z'_λ	-11250 -45	0 0	0 0	$\frac{75}{8}$ $\frac{3}{80}$	$\frac{25}{8}$ $\frac{1}{80}$
-2 -1	-500 -250	X_2 X_1	10 25	0 1	1 0	$\frac{1}{40}$ $-\frac{1}{80}$	$-\frac{1}{40}$ $\frac{1}{16}$

$$\underline{\lambda_B} = \max\left\{-\frac{\frac{75}{8}}{\frac{3}{80}}, -\frac{\frac{25}{8}}{\frac{1}{80}}\right\} = \max\{-250, -250\} = -250$$

$$\overline{\lambda_B} = \infty$$

所以最优区间为:$[-250, \infty)$ 最优解为:$X_2 = 10, X_1 = 25$,最小值为 $Z' = -11250 - 45\lambda$。

这样就把 λ 的变动范围分为三个区间(实际上为两个区间):$(-\infty, -250]$、$[-250, -250]$、$[-250, \infty)$,同时确定了 λ 在每个区间所以应的最优解和目标的最小值。当 $\lambda \leqslant -250$ 时,以不生产为好,当 $\lambda > -250$ 时,最优解不变,仍为:$X_1 = 10, X_2 = 25$,但目标值却取决于 λ 的大小。

如该例子中,假设原材料价格为100元,现上涨2倍,则 x_1 的单位利润下降300元,x_2 单位利润下降600元($\lambda = -300$)以不生产为好,如价格上涨一倍($\lambda = -200$),则仍可按最优计划安排生产,所得利润最多为:

$$Z = -(-11250 + 45 \times 200) = 2250$$

由此可见,应用参数规划可把取值范围划成几个区段并确定每个区段的最优解和最优值,当给定实际值后可很方便地确定其最优解和最小值而不必重新进行计算,这对系统决策是很有利的。

6.5.3.2 约束常数项含参数的线性规划问题

约束右侧的常数项,通常是某类资源的限制,由于资源的变化可能也会引起拥有量随某个参数变化,因此其一般形式可描述为:

$$\begin{aligned} \min \quad & \boldsymbol{Z} = \boldsymbol{CX} \\ \text{s.t.} \quad & \boldsymbol{AX} = \boldsymbol{b} + \lambda \boldsymbol{b}^* \\ & \boldsymbol{X} \geqslant 0 \end{aligned} \tag{6.5.13}$$

其中,λ 为实数参数,$\boldsymbol{b}^* = (b_1^*, b_1^*, \cdots, b_m^*)^{\mathrm{T}}$。

对该问题用单纯形法求解,则对应于基矩阵 $\boldsymbol{B}$ 的解的一般形式为 $\boldsymbol{B}^{-1}\boldsymbol{b}$,故该问题的基解为

$$\boldsymbol{X}_B = \boldsymbol{B}^{-1}(\boldsymbol{b} + \lambda \boldsymbol{b}^*) = \boldsymbol{B}^{-1}\boldsymbol{b} + \lambda \boldsymbol{B}^{-1}\boldsymbol{b}^* \tag{6.5.14}$$

若

$$\boldsymbol{B}^{-1}\boldsymbol{b} + \lambda \boldsymbol{B}^{-1}\boldsymbol{b}^* \geqslant 0 \tag{6.5.15}$$

X_B 才是基可行解,其目标值为

$$\boldsymbol{Z} = \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b} + \lambda \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b}^* \tag{6.5.16}$$

由式(6.5.14)、式(6.5.16)可知,该问题的基可行解和目标值均为 λ 的函数,故需用 λ 来判断该组基解是否为基可行解。

由式(6.5.15)知:

当 $(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i > 0$ 时,$\lambda \geqslant -\dfrac{(\boldsymbol{B}^{-1}\boldsymbol{b})_i}{(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i}$

当 $(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i < 0$ 时,$\lambda \leqslant -\dfrac{(\boldsymbol{B}^{-1}\boldsymbol{b})_i}{(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i}$

$$
\text{令 } \underline{\lambda}_B = \begin{cases} \max\left\{ -\dfrac{(\boldsymbol{B}^{-1}\boldsymbol{b})_i}{(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i} \,\middle|\, (\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i > 0 \right\} \\ -\infty, \text{当} (\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i \leqslant 0 \quad (i = 1,2,\cdots,m) \text{时} \end{cases}
$$

$$
\bar{\lambda}_B = \begin{cases} \min\left\{ -\dfrac{(\boldsymbol{B}^{-1}\boldsymbol{b})_i}{(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i} \,\middle|\, (\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i < 0 \right\} \\ \infty, \text{当} (\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i \geqslant 0 \quad (i = 1,2,\cdots,m) \text{时} \end{cases} \tag{6.5.17}
$$

若使 $\boldsymbol{B}^{-1}\boldsymbol{b} + \lambda\boldsymbol{B}^{-1}\boldsymbol{b} \geqslant 0$ 成立，则 λ 应满足：$\underline{\lambda}_B \leqslant \lambda \leqslant \bar{\lambda}_B$，这就把判断基解是否为基可行解的问题代换成用 λ 来判断了。

$\underline{\lambda}_B$ 与 $\bar{\lambda}_B$ 分别为 B 的下特征数与上特征数，区间 $[\underline{\lambda}_B, \bar{\lambda}_B]$ 为最优区间，$\boldsymbol{X}_B = \boldsymbol{B}^{-1}\boldsymbol{b} + \lambda\boldsymbol{B}^{-1}\boldsymbol{b}$ 为最优解，$Z = \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b} + \lambda\boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{b}$ 为目标极小值。

如果 λ 不在区间 $[\underline{\lambda}_B, \bar{\lambda}_B]$ 内，则说明基解不是基可行解，因此需进行换基迭代。

当 $\lambda > \bar{\lambda}_B$ 时，因 $\lambda > \bar{\lambda}_B = \dfrac{(\boldsymbol{B}^{-1}\boldsymbol{b})_K}{(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_K}((\boldsymbol{B}^{-1}\boldsymbol{b}^*)_i < 0)$，则 $\boldsymbol{X}_K = (\boldsymbol{B}^{-1}\boldsymbol{b})_K - \lambda(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_K < 0$，这时 λ 所对应的基变量 $\boldsymbol{X}_K$ 应出基，而进基元素按下式选择：

$$
\min\left\{ \frac{(\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A})_{kj}}{(\boldsymbol{B}^{-1}\boldsymbol{A})_{kj}} \,\middle|\, (\boldsymbol{B}^{-1}\boldsymbol{A})_{kj} < 0 \right\} \tag{6.5.18}
$$

其原因是：从第 K 个约束行看，当 $\lambda < \bar{\lambda}_B$ 时，由于 $\boldsymbol{X}_K = (\boldsymbol{B}^{-1}\boldsymbol{b})_K + \lambda(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_K < 0$，它不是基可行解，如果 $(\boldsymbol{B}^{-1}\boldsymbol{A})_{Kj} > 0$，则无论非基变量 X_j 如何由零增大，由于 $X_K = [(\boldsymbol{B}^{-1}\boldsymbol{b})_K - \lambda(\boldsymbol{B}^{-1}\boldsymbol{b}^*)_K] - (\boldsymbol{B}^{-1}\boldsymbol{A})_{Kj}X_K$，所以不会使 $X_K \geqslant 0$，因此只有 $(\boldsymbol{B}^{-1}\boldsymbol{A})_{Kj} < 0$ 的元素所对应的非基变量 X_j 才能使 X_K 由负增大到零，即由基变量变为非基变量，至于在 $\left\{ -\dfrac{(\boldsymbol{C} - \boldsymbol{C}_B\boldsymbol{B}^{-1}\boldsymbol{A})_{kj}}{(\boldsymbol{B}^{-1}\boldsymbol{A})_{kj}} \,\middle|\, (\boldsymbol{b}^{-1}\boldsymbol{A})_{kj} < 0 \right\}$ 中选择最小元素作为旋转中心是由下节对偶问题决定的，在此只要记住就行了。

由上述可知，如果第 K 行的全部值均为正数，则是解无界的情况。

当 $\lambda < \underline{\lambda}$ 时，仍按上述道理确定进基元素、出基元素和旋转中心。

现举例说明这种参数规划的求解方法。

$$
\begin{aligned}
\min \quad & Z = X_1 + 3X_2 + 3X_3 - 5X_4 + X_5 + 3X_6 \\
\text{s.t.} \quad & X_1 + 2X_3 + X_4 - X_6 = -1 + 3\lambda \\
& X_2 + X_3 + X_5 + X_6 = -2 + \lambda \\
& X_4 + X_5 + 2X_6 = -3 + 2\lambda \\
& X_j \geqslant 0, j = 1,2,\cdots,6
\end{aligned}
$$

解：引入人工变量 X_7、X_8、X_9，并化为标准型得：

$$
\begin{aligned}
\min \quad & W = X_7 + X_8 + X_9 \\
\text{s.t.} \quad & -Z + X_1 + 3X_2 + 3X_3 - 5X_4 + X_5 + 3X_6 = 0 \\
X_7 \quad & -X_1 - 2X_3 - X_4 + X_6 = 1 - 3\lambda \\
X_8 \quad & -X_2 - X_3 - X_5 - X_6 = 2 - \lambda \\
X_9 \quad & -X_4 - X_5 - 2X_6 = 3 - 2\lambda \\
& X_j \geqslant 0, j = 1,2,\cdots,9
\end{aligned}
$$

$-Z$、X_7、X_8、X_9 为基变量，其单纯形表如下：

变量						$-Z$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
C_W													1	1	1
C_z							1	3	3	−5	1	3	0	0	0
λ	C_B^Z	C_B^W	$-W$	−6	6		1	1	3	2	2	2			
			$-Z$	0	0	1	1	3	3	−5	1	3	0	0	0
1/3	0	1	X_7	1	−3		⊖1		−2	−1		3	1	0	0
2	0	1	X_8	2	−1			−1	−1		−1	−1		1	0
3/2	0	1	X_9	3	−2					−1	−1	−2	0	0	1

$$\underline{\lambda}_B = -\infty, \bar{\lambda}_B = \min\left\{-\frac{2}{-3}, -\frac{2}{-1}, -\frac{3}{-2}\right\} = \frac{1}{3}$$

最优区间为$(-\infty, \frac{1}{3})$，$W = 6 - 6\lambda$，在此区间 $W \neq 0$，故无可行解，因此确定当 $\lambda > \frac{1}{3}$ 时，人工变量的目标值 W 可否为零。

因 $\lambda > \frac{1}{3}$ 时，其对应基变量 X_7 为出基变量，进基元素为

$$\min\left\{-\frac{1}{-1}, -\frac{3}{-2}, -\frac{2}{-1}\right\} = \min\{1, 1.5, 2\} = 1$$

X_1 为进基变量，旋转中心为 −1，经旋转变换得下表：

变量						$-Z$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
C_W													1	1	1
C_Z							1	3	3	−5	1	3	0	0	0
λ	C_B^Z	C_B^W	$-W$	−5	3		0	1	1	1	2	3	1	0	0
			$-Z$	1	−3	1	0	3	1	−6	1	4	1	0	0
1/3	1	0	X_1	−1	3		1	0	2	1	0	−1	−1	0	0
2	0	1	X_8	2	−1		0	−1	−1	−1	−1	−1	0	1	0
3/2	0	1	X_9	3	−2		0	0	0	⊖1	−1	−1	0	0	1

$$\underline{\lambda}_B = \max\left\{-\frac{-1}{3}\right\} = \frac{1}{3}, \bar{\lambda}_B = \min\left\{-\frac{2}{-1}, -\frac{3}{-2}\right\} = \frac{3}{2}$$

最优区间为$\left[\frac{1}{3}, \frac{3}{2}\right]$，$W = 5 - 3\lambda$，在该区间内 $W \neq 0$，故需在 $\lambda > \frac{3}{2}$ 的范围内确定 W 可否为零。

因 $\lambda > \frac{3}{2}$，故 X_9 出基，X_4 为进基元素，经旋转变换得下表：

变量						$-Z$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
C_W													1	1	1
C_Z							1	3	3	-5	1	3	0	0	0
λ	C_B^Z	C_B^W	$-W$ $-Z$	-2 -17	1 9	0 1	0 0	1 3	1 1	0 0	1 7	1 16	1 1	0 0	0 0
-2	1	0	X_1	2	1		1	0	2	0	-1	-3	-1	0	1
$\bar{2}$	0	1	X_8	2	-1		0	-1	⊖1	0	-1	-1	0	1	0
$\underline{\frac{3}{2}}$	-5	0	X_4	-3	2		0	0	0	1	1	2	0	0	-1

$$\underline{\lambda}_B = \max\left\{-\frac{2}{1}, -\frac{-3}{2}\right\} = \frac{3}{2}, \bar{\lambda}_B = \min\left\{-\frac{2}{-1}\right\} = 2$$

最优区间为：$\left[\frac{3}{2}, 2\right]$，$W = 2 - \lambda$，当 $\lambda = 2$ 时，$W = 0$，原区间有基可行解。因基中有人工变量 X_8，且 X_8 对应行有负值，故使 X_8 出基，选：

$\min\left\{-\frac{1}{-1}, -\frac{1}{-1}, -\frac{1}{-1}, -\frac{1}{-1}\right\} = 1$ 所对应的变量 X_3 进基，进行旋转变换得下表：

变量					$-Z$	X_1	X_2	X_3	X_4	X_5	X_6
C						1	3	3	-5	1	3
λ	C_B	$-Z$	-15	8	1	0	2	0	0	6	15
$\bar{6}$	1	X_1	6	-1		1	⊖2	0	0	-3	-5
$\underline{2}$	3	X_3	-2	1		0	1	1	0	1	1
$\frac{3}{2}$	-5	X_4	-3	2		0	0	0	1	1	2

$$\underline{\lambda}_B = \max\left\{-\frac{-2}{1}, -\frac{-3}{2}\right\} = 2, \bar{\lambda}_B = \min\left\{-\frac{6}{-1}\right\} = 6$$

最优区间为：$[2,6]$，最优解为：$X_1 = 6 - \lambda$，$X_3 = -2 + \lambda$，$X_4 = -3 + 2\lambda$，$X_j = 0 (j = 2,5,6)$，目标函数极小值 $Z = 15 - 8\lambda$。当 $\lambda > 6$，即超过上限时，X_1 变为非基可行解，故 X_1 出基，X_2 进基。经旋转变换得下表：

变量					$-Z$	X_1	X_2	X_3	X_4	X_5	X_6
C						1	3	3	-5	1	3
λ	C_B	$-Z$	-9	7	1	1	0	0	0	3	10
$\bar{6}$	3	X_2	-3	$\frac{1}{2}$		$-\frac{1}{2}$	1	0	0	$\frac{3}{2}$	$\frac{5}{2}$
-2	3	X_3	1	$\frac{1}{2}$		$\frac{1}{2}$	0	1	0	$-\frac{1}{2}$	$-\frac{3}{2}$
$\frac{3}{2}$	-5	X_4	-3	2		0	0	0	1	1	2

$$\underline{\lambda}_B = \max\left\{-\frac{-3}{\frac{1}{2}}, -\frac{1}{\frac{1}{2}}, -\frac{-3}{2}\right\} = 6, \qquad \bar{\lambda}_B = \infty$$

最优区间为$[6, \infty)$,最优解为:$X_2 = -3 + \frac{1}{2}\lambda, X_3 = 1 + \frac{1}{2}\lambda, X_4 = -3 + 2\lambda, X_j = 0(j = 1,5,6)$,对应的目标函数值为:$Z = 9 - 7\lambda$。

通过该例可以看出,λ 的取值范围被分为两个区间:$[2,6]$、$[6,\infty]$。当 λ 取值在区间 $[2,6]$ 内时,最优解为:$X_1 = 6 - \lambda, X_3 = -2 + \lambda, X_4 = -3 + 2\lambda, X_2 = X_5 = X_6 = 0$,目标的极小值为:$Z = 15 - 8\lambda$。如果 λ 的实际值为 3,则 $X_1 = 3, X_3 = 1, X_4 = 3, X_2 = X_5 = X_6 = 0, Z = -9$。当 λ 取值为$[6,\infty]$之间时,最优解为:$X_2 = -3 + \frac{1}{2}\lambda, X_3 = 1 + \frac{1}{2}\lambda, X_4 = -3 + 2\lambda, X_1 = X_5 = X_6 = 0$,目标函数值为:$Z = 9 - 7\lambda$,如果实际值为 $\lambda = 8$,则最优解为 $X_2 = 1, X_3 = 5, X_4 = 13, X_1 = X_5 = X_6 = 0$,目标函数值 $Z = -47$。

约束右端的常数项含有参数时,它的对偶规划是目标函数的系数含参数的问题,因此也可通过对偶理论,化成目标函数的系数含参数的问题去解。

6.6 对偶规划及影子价值

应用线性规划处理问题经常出现以下情况:

(1) 所建模型变量并不多,但约束条件却很多。求解这类问题时,由于引入松弛变量和人工变量数较多,导致 A 矩阵的规模急骤增大。如两个变量,10 个约束的线性规划模型,如果都是大于等于约束,则需引入松弛变量和人工变量 20 个,从而使 A 矩阵的阶次由 10×2 增大为 10×22,这使计算工作量增大若干倍,因此需寻找一种处理这类问题的简便方法。

(2) 在处理问题时,经常需从不同角度来研究。如某建材厂生产两种产品,其单位产品的消耗量及单位利润见表 6.6.1,现欲安排生产计划,可建如下模型:

$$\begin{aligned} \max \quad & Z = 250X_1 + 500X_2 \\ \text{s.t.} \quad & 2X_1 + 2.5X_2 \leqslant 1000 \\ & 0.5X_1 + X_2 \leqslant 600 \\ & X_1 \geqslant 0, X_2 \geqslant 0 \end{aligned} \tag{6.6.1}$$

表 6.6.1 单位产品的消耗量仅单位利润

	甲产品	乙产品	拥有量
原料 A	2	2.5	1000
原料 B	0.5	1	600
单位利润	250	500	

如果从另一个角度研究,现将原料出售,又不低于产品生产所获得的利润,两种原料的最低售价应为多少才合算。

设：Y_1、Y_2 为两种原料的最低售价，模型应为

$$\begin{aligned} \min \quad & S = 1000Y_1 + 600Y_2 \\ \text{s.t.} \quad & 2Y_1 + 0.5Y_2 \geqslant 250 \\ & 2.5Y_1 + Y_2 \geqslant 500 \\ & Y_1 \geqslant 0, Y_2 \geqslant 0 \end{aligned} \tag{6.6.2}$$

同一组数据，从不同的角度研究可建不同的模型。那么，两个模型之间有什么区别和联系，又各反映什么经济本质呢?这又是一个需要探讨的问题。

(3) 在研究问题中，经常需要分析某种资源的增加或减少对目标值的影响程度。有些资源的增加并不影响目标值，这类资源是长线资源，而某些资源的增加，对目标值影响很大，这种资源是较珍贵的资源，称为短线资源。为了确定资源的长短程度，也需要一种评价方法。

本节研究的线性规划对偶理论为解决上述三方面问题，提供了一套完整的理论和方法。

6.6.1 线性规划的对偶理论

现以第二个问题中所提出的两个模型为例来研究两个模型之间的内在联系。

从两个模型看，第一个模型的目标要求最大，而第二个模型的目标要求最小；前者的变量数和后都的约束数相等；第一个模型的 C 是第二个模型的 b，第一个模型的 b 是第二个模型的 C；第二个模型的 A 矩阵是第一个模型 A 矩阵的转置。

如果把第一个模型叫原问题，第二个问题叫对偶问题，并将这种对应关系推而广之，则得如表 6.6.2 所列的对应关系。

表 6.6.2 对应关系

原问题(或对偶问题)		对偶问题(或原问题)	
目标函数 $\min Z$		目标函数 $\max Z$	
约束条件	约束条件数 m 个	变量	对偶变量数 m 个
	约束条件为 $\geqslant$		对偶变量 $Y_j \geqslant 0$
	约束条件为 $=$		对偶变量 Y_j 为自由变量
变量	变量 X_i 数为 n 个	约束条件	约束条件为 n 个
	变量 X_i 为自由变量		约束条件为 $=$
	变量 $X_j \geqslant 0$		约束条件为 $\leqslant$
约束的系数矩阵为 A		约束的系数矩阵为 A^T	
约束常数项为 b		约束常数项为 C	
指标因数为 C		指标因数为 b	

由对应关系可见，原问题和对偶问题是互为对偶的，即把其中任一个称为原问题，另一个则是它的对偶问题。

运筹学中对偶理论指出：

(1) 一对对偶问题，是一个问题的两个侧面面，其目标是一致的，即若原问题有最优解，那么对偶问题也有最优解，且目标函数值相等；

(2) 原问题的检验数 $C - C_B B^{-1} A$，对应于对偶问题的一组基解，基 B 为最优基，则最优基下的检验数对应于对偶问题的最优解；

(3) 对偶问题的最优解是原问题的最优基下的检验数。

由上述对偶理论知，上述第一个问题可通过求解它的对偶问题使模型简化。模型为两个变量 10 个约束的问题，其对偶问题为 2 个约束 10 个变量的问题，加上松弛变量也只有 12 个变量，A 矩阵为 2×12 阶，比 10×22 阶要小得多。这就是说，原问题和对偶问题中求解其中的一个就等于求解了另一个，因此，遇到第一种情况，可通过求解其对偶问题使其方便简单地得到解决。

第二个问题则不必建立两个模型，而只在建立并求解一个模型的同时，通过对偶关系解决另一问题。

第三个问题则不需求解对偶规划模型，只需在求解原问题时保留其最优基下的判断数即可确定影子价格，解决资源的长短线问题。

6.6.2 影子价格

由单纯形法知，目标函数 $Z = C_BB^{-1}b$，当 b 增加一个单位时，Z 增加 C_BB^{-1}，C_BB^{-1} 称为单纯形乘子。因为它体现了资源增加一个单位时，目标函数的增长量，起到了资源参考价格的作用。因此又称为影子价格。影子价格在国外又称为机会成本、会计价格、隐涵价格，最优计划价格、完全竞争条件下的市场价格以及最优分工协作方案的实现价格等。它是经济管理中相当重要的参数之一。

由判断准则 $C - C_BB^{-1}A \geqslant 0$ 知，在最优基时只有非基变量所对应的判断数才大于零，基变量的判断数均为零。因此对原问题来说，松弛变量不为零，则它一定是基变量，且在基矩阵中，因基变量的判断数为零，故根据对偶原理，它的影子价格为零，这种资源的增长不会使目标值增加，故它是长线资源。如果松弛变量为非基变量，其值为零，且判断数大于零，这说明系统取最优解时，该资源已用尽，其数量的增加可使目标函数值增加，它的影子价格就是它所对应的判断数(因资源松弛变量对应的检验数在 $C - C_BB^{-1}A$ 中一般为 $-C_BB^{-1}$)。

影子价格在国民经济管理中有重要意义。我国实行社会主义市场经济体制，目前我国的物价政策尤其是垄断性行业使某些商品或服务的价格与影子价格不符，这就造成企业追求最大利润的目标与国家追求国民收入最大的目标不一致。如当一种产品长线时，其影子价格为零，但该产品的销售价格不变，企业多生产还可增加产值和利润，反之短线产品影子价格很高，但销售价格较低，这就使得企业的最优目标与社会最优目标相矛盾，因而即使上级有最优计划，而下级却因利害冲突而缺乏积极性。为了改变经济系统子部门所追求的目标与系统目标不一致的情况，采用影子价格实行管理是很必要的。这样就可以使子部门根据单位最大利益所制定的计划与系统整体的最优计划相符，既可协调上下级之间的关系，又可调动下级的积极性，使其主动、自觉地为系统最优计划服务。

6.6.3 对偶单纯形法

在单纯形法的标准型中曾规定 $b \geqslant 0$，其目的是为了保证引入人工变量之后可以方便地得到一组基可行解。由于对偶问题中的 b 是原问题中的 C，对偶问题中的 C 是原问题中的 b，根据这种对应关系可得一种求解线性规划问题的方法——对偶单纯形法。

对偶单纯形法是以对偶理论为基础，从一组基解过渡到另一组基解最后得到基可行解的一种方法。对某类问题它不必引入人工变量，也不必规定 $b \geqslant 0$，因而可使计算简化。

对偶单纯形法的计算步骤如下：

(1) 将模型化为

$$\begin{aligned} \min \quad & Z = CX \\ \text{s.t.} \quad & AX = b \\ & X \geqslant 0 \end{aligned}$$

的形式并得出增广标准形。

(2) 求初始基解(判断数均大于零的基解，可不是基可行解)。

(3) 如果 $\tilde{b}_i \geqslant 0$，则得最优解，计算结束，否则转(4)。

(4) 选 $\tilde{b}_r = \min\{\tilde{b}_i \mid \tilde{b}_i < 0\}$ 行。

(5) 如果所有 $a_{rj} \geqslant 0$，则问题解无界，否则转(6)。

(6) 选 $\tilde{C}_s = \min\left\{-\dfrac{\tilde{C}_j}{\tilde{a}_{rj}} \mid \tilde{a}_{rj} < 0\right\}$ 列，得旋转中心 a_{rs}。

(7) 进行旋转变换，转(3)。

现将单纯形法和对偶单纯形法的对应关系列于表 6.6.3。

表 6.6.3　对应关系

步骤	单纯形法	对偶单纯形法
最优解判断准则	$\tilde{C}_i \geqslant 0$ (所有 i)	$\tilde{b}_i \geqslant 0$ (所有 i)
选择进基元素	若 $\tilde{C}_s = \min\{\tilde{C}_j \mid \tilde{C}_j < 0\}$ 则 X_s 进基	若 $\dfrac{\tilde{C}_s}{\tilde{a}_{rs}} = \min\{\dfrac{\tilde{C}_j}{-\tilde{a}_{rj}} \mid \tilde{a}_{rj} < 0\}$ 则 X_s 进基
选择出基元素	若 $\dfrac{\tilde{b}_r}{\tilde{a}_{rs}} = \min\left\{\dfrac{\tilde{b}_i}{\tilde{a}_{is}} \mid \tilde{a}_{is} > 0\right\}$ 则 X_r 出基	若 $\tilde{b}_{rs} = \min\{\tilde{b}_i \mid \tilde{b}_i < 0\}$ 则 X_r 出基
解无界	$\tilde{a}_{is} \leqslant 0 (i = 1,2,\cdots,m)$	$\tilde{a}_{rj} \geqslant 0 (j = 1,2,\cdots,n)$
旋转中心	$\tilde{a}_{rs}$	$\tilde{a}_{rs}$

应用对偶单纯形法求解的例题请参看 6.5.3 节参数规划部分。

6.7　整 数 规 划

整数规划是模型中变量有整数限制的特殊线性规划。通常认为，这类线性规划，只要在不考虑整数约束条件下求解，再将非整数解“四舍五入”化成整数即可得到整数规划的解，表面上看，似乎合情合理，但实质上是错误的，即使正确，也是偶合。其原因在于：

(1) 全部“五入”，必然造成资源不足，不满足约束条件，这组解一定不是可行解；

(2) 全部舍弃小数部分，可得到可行解，但未必是最优解；

(3) 部分“五入”，部分“四舍”，也可能出现上述两种情况之一结果。因此，不能用“取整”的方法求解整数规划问题。

现介绍分枝定界法求解整数规划的基本原理和步骤。

用分枝定界法求解整数规划的基本原理是分枝和定界。所谓分枝即是在无整数约束的线性规划问题的可行区中，去掉某一非整解的小数部分，将可行区分为两部分(即两

枝),再在两个部分中寻找整数解,依次逐级分枝,直到最优解为止。所谓定界是指在分枝过程中依据每次分枝确定最优解的上、下界限,求解过程只在该上下界限范围内分枝并逐渐缩小上界限,最终求出最优解。

该方法的步骤为:

设求解的整数规划为问题 A,与之相对应的不考虑整数约束的一般线性规划为问题 B,求解 B 可能得到以下情况之一:

(1) B 无可行解,则 A 亦无可行解;

(2) B 有最优解且符合问题 A 的整数约束,则 B 的最优解为 A 的最优解;

(3) B 有最优解,但不符合 A 的整数约束,则其最优解为 $\overline{X}$。

因此只研究第三种情况。

第一步 分枝。在 B 的最优解中任选一不符合整数条件的变量 x_j,其值为 b_j,构造两个约束;

$x_j \leqslant [b_j]_{整数}$; $x_j \geqslant [b_j]_{整数} + 1$

这两个约束则去掉了可行区中$([b_j]_{整数}, [b_j]_{整数} + 1)$部分,而将可行区分为两个部分,即两枝,从而得到两个后继的一般线性规划;

定界。以每个后继线性规划为一分枝,求解问题,并找出各问题最大的最优值取代 $\bar{Z}$,从符合整数条件的各分枝中找出最大的最优值作为 $\underline{Z}$,如无符合取整条件的最优解,则 $\underline{Z} = 0$;

第二步 比较与剪枝,各分枝的目标函数值中若有小于 $\underline{Z}$ 者,则剪掉该枝,若大于 $\underline{Z}$ 且不符合整数条件,则重复第一步,直到 $\underline{Z} = Z^*$ 止。

现举例说明其算法。

某企业构件厂生产甲、乙两种钢筋混凝土构件,原材料单耗、原材料拥有量及单位利润见表 6.7.1。如可安排生产,才能使企业获利润最多。

表 6.7.1 某企业构件厂构件数据

	甲	乙	拥有量
钢材	20	50	1450
水泥	20	10	650
单位利润	250	500	

因钢筋混凝土构件生产量只能是整数,故该问题是整数规划问题,其模型为

$$\begin{aligned} \max \quad & Z = 250x_1 + 500x_2 \\ \text{s.t.} \quad & 20x_1 + 50x_2 \leqslant 1450 \\ & 20x_1 + 10x_2 \leqslant 650 \\ & x_1 \geqslant 0 \quad x_2 \geqslant 0 \quad \text{且均为整数} \end{aligned}$$

按分枝定界法计算步骤,该问题为 A,与之相对应的无整数约束的线性规划问题为 B。B 的解为

$$x_1 = 22.5 \quad x_2 = 20 \quad Z = 15625 \quad \underline{Z} = 0 \quad \bar{Z} = 15625$$

因 $x_1 = 22.5$ 为非整数,引入 $x_1 \leqslant 22$ 和 $x_1 \geqslant 23$ 两个约束,将可行区分为两个部分,得两个一般线性规划

$$B_1 \quad \begin{aligned} \max \quad & Z = 250x_1 + 500x_2 \\ \text{s.t.} \quad & 20x_1 + 50x_2 \leqslant 1450 \\ & 20x_1 + 10x_2 \leqslant 650 \\ & x_1 \geqslant 23, x_2 \geqslant 0 \end{aligned} \qquad B_2 \quad \begin{aligned} \max \quad & Z = 250x_1 + 500x_2 \\ \text{s.t.} \quad & 20x_1 + 50x_2 \leqslant 1450 \\ & 20x_1 + 10x_2 \leqslant 650 \\ & 0 \leqslant x_1 \leqslant 22, x_2 \geqslant 0 \end{aligned}$$

求解得 B_1: $x_1 = 23$ $x_2 = 19$ $Z = 15250$

B_2: $x_1 = 22$ $x_2 = 20.2$ $Z = 15250$

B_1 的解为整数，目标 Z 大于$\underline{Z}$，故$\underline{Z} = 15250$

B_2 的解存在非整数，故 $\bar{Z} = 15600$

对 B_2 进行分枝得一般线性规划 B_{21}，B_{22}：

B_{21} $\max Z = 250x_1 + 500x_2$

s.t. $20x_1 + 50x_2 \leqslant 1450$

$20x_1 + 10x_2 \leqslant 650$

$0 \leqslant x_1 \leqslant 22$

$x_2 \geqslant 21$

B_{22} $\max Z = 250x_1 + 500x_2$

s.t. $20x_1 + 50x_2 \leqslant 1450$

$20x_1 + 10x_2 \leqslant 650$

$0 \leqslant x_1 \leqslant 22$

$0 \leqslant x_2 \leqslant 20$

求解 B_{21} 得：$x_1 = 22$ $x_2 = 20$ $Z = 15500$

求解 B_{22} 得：$x_1 = 20$ $x_2 = 21$ $Z = 15500$

因 B_{21}，B_{22} 均为整数解，且目标值高于上一分枝的$\underline{Z}$，故

$$\underline{Z} = 15500 = Z^*$$

该问题的分枝定界过程见图 6.7.1。

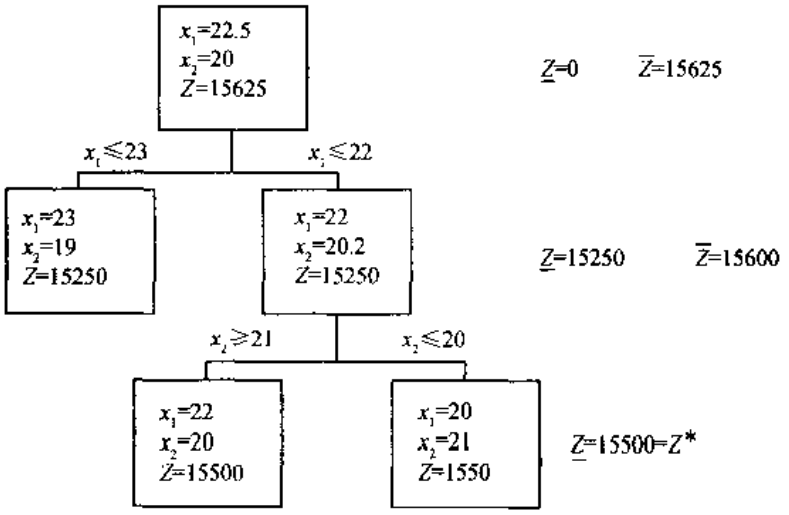


图 6.7.1 分枝定界图

分枝定界法求解整数规划问题仍很麻烦，用 6.10 节介绍的 Lingo 软件可方便地求解各种类型的整数规划。

6.8 模糊线性规划及其应用

数学规划是系统最优化方法中普遍采用的方法。但该方法是在目标函数和约束条件均为确定值条件下的最优化方法，虽然其中线性规划可通过灵敏度分析确定其模型各种参量在变动范围内的最优解，但在建模时并未考虑这些参量的模糊性。因此数学理论又提出了一种用于系统规划的方法 —— 模糊规划法。

模糊规划是应用模糊数学的基本原理，在一般数学规划模型的基础之上对目标或约束条件或目标和约束条件模糊化后得到规划模型进行系统规划的一种方法。本节主要介绍模糊线性规划。

6.8.1 模糊线性规划及求解的基本原理

6.8.1.1 模糊线性规划及其分类

一般线性规划的基本形式为

$$\begin{aligned} \min \quad & Z = CX \\ & AX = B \\ & X \geqslant 0 \end{aligned}$$

如将线性规划中的线性目标 CX 和线性约束条件 $AX \leqslant B$ 模糊化，可得以下几类模糊线性规划：

(1) 第一类，目标函数是明确的，而约束条件是模糊的。即

目标函数为：$\min \quad Z = CX$

约束条件为：$AX \lesssim B$

$X \geqslant 0$

式中 $\lesssim$ —— 模糊约束，$AX \lesssim B$ 表示 AX 大约小于 B 这一模糊约束。线性规划模型中 B 一般是某种资源或生产能力、市场需求等限制量，因此上式可体现在限制量 B 的基础上的某一变动范围。

(2) 第二类，目标函数和约束条件均是模糊的。其形式为

目标函数为：$CX \lesssim Z_0$

约束条件为：$AX \lesssim B$

$X \geqslant 0$

式中 Z_0—— 决策时提出的目标期望值。

(3) 第三类，约束条件明确而目标函数是模糊的。其形式为

目标函数为：$CX \lesssim Z_0$

约束条件为：$AX = B$

$X \geqslant 0$

现举例说明三类模糊规划模型。

某工厂生产甲、乙两种产品。其原材料、燃料、动力的单位消耗、单位利润等资料见表 6.8.1。

表 6.8.1 原材料、燃料、动力、单位消耗和单位利润表

	甲	乙	资源拥有量
钢材	3	2	2500
煤	1	0.5	900
电	10	7	1000
单位利润	20	15	

根据表中资料可得线性规划模型为：

设甲、乙两种产品的产量为：X_1, X_2

目标函数为：max　　$Z = 20X_1 + 15X_2$

约束条件为：s.t.

$$\begin{aligned} 3X_1 + 2X_2 &\lesssim 2500 \\ X_1 + 0.5X_2 &\lesssim 900 \\ 10X_1 + 7X_2 &\lesssim 10000 \\ X_1 \geqslant 0,\ X_2 &\geqslant 0 \end{aligned}$$

如果将上述条件加以改变，即资源拥有量分别在 2500、900 和 10000 的基础上可增加，但不超过 10%，由此可得如下第一类模糊规划：

$$\begin{aligned} \max\quad & Z = 20X_1 + 15X_2 \\ \text{s.t.}\quad & 3X_1 + 2X_2 \lesssim 2500 \\ & X_1 + 0.5X_2 \lesssim 900 \\ & 10X_1 + 7X_2 \lesssim 10000 \\ & X_1 \geqslant 0,\ X_2 \geqslant 0 \end{aligned}$$

如在此模型基础上再确定预期目标利润值 $Z_0 = 25000$ 元或稍少些，即可得第二类模糊规划模型：

$$\begin{aligned} 20X_1 + 15X_2 &\lesssim 25000 \\ 3X_1 + 2X_2 &\lesssim 2500 \\ X_1 + 0.5X_2 &\lesssim 900 \\ 10X_1 + 7X_2 &\lesssim 10000 \\ X_1 \geqslant 0,\ X_2 &\geqslant 0 \end{aligned}$$

如在线性规划模型的基础上只确定预期目标利润值 $Z_0 = 25000$ 元或稍少些，则可得第三类模糊规划：

$$\begin{aligned} 20X_1 + 15X_2 &\lesssim 25000 \\ 3X_1 + 2X_2 &\leqslant 2500 \\ X_1 + 0.5X_2 &\leqslant 900 \\ 10X_1 + 7X_2 &\leqslant 10000 \\ X_1 \geqslant 0,\ X_2 &\geqslant 0 \end{aligned}$$

6.8.1.2　第一类模糊线性规划的求解方法

第一类模糊线性规划的标准形式为

$$\begin{aligned} \max\quad & Z = CX \\ \text{s.t.}\quad & AX \lesssim B \\ & X \geqslant 0 \end{aligned}$$

其中 $B = (b_1, b_2, \cdots, b_m)^T$，各分量 b_i 的增量为 d_i，即包括最大增量的 B 向量为 $B_i \fallingdotseq (b_1 + d_1, b_2 + d_2, \cdots, b_m + d_m)^T$。

该类模糊规划的求解方法如下：

第一步 选择隶属函数 $\mu_i(d), i = 1,2,\cdots,m$。

隶属函数是模糊数学中的基本概念。其精确的定义为：设给定论域 V，所谓 V 上的一个模糊子集 A 是指对于任意的 $X \in V$，都有一个确定的数 $\mu_A(X) \in [0,1]$，用该数表示 X 属于 A 的程度。

映射：

$$\mu_A : V \to [0,1]$$
$$X \to \mu_A(X)$$

称为 A 的隶属函数，数 $\mu_A(X)$ 叫做 V 中元素 X 对模糊子集的隶属度。

通俗地说，当 X 明确地属于模糊子集 A，则隶属度 $\mu_A(X) = 1$；明确地不属于该集合，则隶属度 $\mu_A(X) = 0$，如不明确地知道 X 是否属于该集合，则可根据 X 接近 A 集合的程度用 $[0,1]$ 之间的一个数来表示其属于该集合的程度，该数叫隶属度，描述隶属度计算过程的函数叫隶属函数。

第一类模糊规划中一般选严格递减函数作为隶属函数。最简单的严格递减函数是一条斜率为负的直线（见图 6.8.1）。

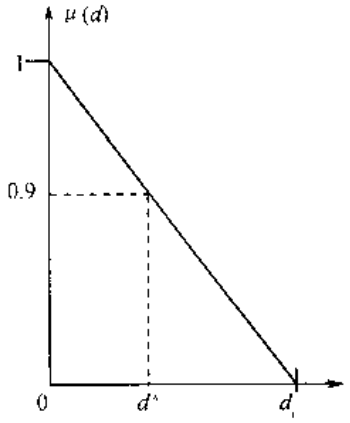


图 6.8.1 隶属函数

由图 6.8.1 可见，当 $d \in [0,d_i]$ 时，$\mu_i(d_i) \in [0,1]$。当 $d \leqslant 0$ 时，$\mu_i(d_i) = 1$，当 $d = d_i$ 时，$\mu_i(d_i) = 0$。用公式描述则为：

$$\mu_i(d_i) = \begin{cases} 1 & d < 0 \\ 1 - \dfrac{1}{d_i} d & 0 \leqslant d < d_i \\ 0 & d > d_i \end{cases}$$

第二步 求 B^0 下一般线性规划的最优解和极值 Z_0^*。

第三步 按下式将目标函数模糊化。

$$\mu_F(X) = \frac{CX}{Z_0^*}$$

第四步 给定截水平 λ_k 和精度 $\varepsilon > 0$，计算 d_i^λ 和 B^λ，从而形成以 $\mu_F(X)$ 为目标，以 B^λ 为约束右端项的一般线性规划。求解该线性规划得其最优解 X^* 和最优值 Z_k^*

d_i^λ 和 B^λ 的计算公式如下：

$$d_i^\lambda = \max\{d_i \mid \mu_i(d) \geqslant \lambda\}$$
$$B^\lambda = B + d_i^\lambda$$

第五步 按 $\varepsilon_k = \lambda_k - Z_k^*$ 计算 ε_k，如果 $A \mid \varepsilon_k \mid \leqslant \varepsilon_0$，则 X^* 为该模糊线性规划的最优解，将其代入原目标得该模糊线性规划极值，λ_k 则为该模糊线性规划的最优截水平。如果 $\mid \varepsilon_k \mid > \varepsilon_0$，则需按 $\lambda_{k+1} = \lambda_k - r_k\varepsilon_k$ 再确定截水平，重复第四步计算，直至得到最优解为止。该步中 r_k 的选择原则为

$$0 < r_k \leqslant 1, \text{并使 } 0 \leqslant r_{k+1} \leqslant 1$$

6.8.1.3 第二类模糊线性规划的求解方法

第二类模糊线性规划的标准形式为

$$CX \lesssim Z_0$$
$$AX \lesssim B$$
$$X \geqslant 0$$

Z_0 为决策人的目标期望值，其他与第一类模糊线性规划模型相同。

该类模糊线性规划的求解方法如下：

第一步 将模糊目标和模糊约束结合起来，写成如下形式：

$$\begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{bmatrix} \lesssim \begin{bmatrix} Z_0 \\ b_1 \\ b_2 \\ \cdots \\ b_m \end{bmatrix}$$

用矩阵描述则为

$$A'X \lesssim B'$$

第二步 确定隶属函数 $\mu(X)$。

隶属函数的确定原则为

- 当 $(A'X) \leqslant b_i'$ 时，$\mu_i(X) = 1$；
- 当 $b_i \leqslant (A'X)_i \leqslant b_i + d_i$ 时，$0 < \mu_i(X) < 1$；
- 当 $(A'X) > b_i + d_i$ 时，$\mu_i(X) = 0$

根据上述原则，将隶属函数取为

$$\mu_i(X) = \begin{cases} 1 & (A'X)_i \leqslant b_i \\ \dfrac{b_i + d_i - (A'X)_i}{d_i} & 0 < (A'X)_i < b_i + d_i \\ 0 & (A'X)_i \geqslant b_i + d_i \end{cases}$$

当然也可将 $\mu_i(X)$ 取其他函数形式。

第三步 根据最优判决条件，建立一般线性规划模型。

由模糊数学基本理论可知，最优判决条件为

$$\max_{X \geqslant 0} \left\{ \min_{0 \leqslant i \leqslant m} \left(\frac{b_i + d_i - (A'X)_i}{d_i} \right) \right\}$$

令

$$\lambda = \min_{0 \leqslant i \leqslant m} \left(\frac{b_i + d_i - (A'X)_i}{d_i} \right)$$

根据上述最优判决条件可得如下一般线性规划模型：

$$
\begin{aligned}
\max \quad & \lambda \\
\text{s.t.} \quad & \lambda \leqslant (b_0 + d_0 - c_1x_1 - c_2x_2 - \cdots - c_nx_n)/d_0 \\
& \lambda \leqslant (b_1 + d_1 - a_{11}x_1 - a_{12}x_2 - \cdots - a_{1n}x_n)/d_1 \\
& \qquad \cdots \\
& \lambda \leqslant (b_m + d_m - a_{m1}x_1 - a_{m2}x_2 - \cdots - a_{mn}x_n)/d_m \\
& \lambda \geqslant 0, X_j \geqslant 0 \quad (j = 1,2,\cdots,n)
\end{aligned}
$$

将其写成规范形式有：

$$
\begin{aligned}
\max \quad & \lambda \\
\text{s.t.} \quad & d_0\lambda + c_1x_1 + c_2x_2 + \cdots + c_nx_n \leqslant b_0 + d_0 \\
& d_1\lambda + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \leqslant b_1 + d_1 \\
& \qquad \cdots \\
& d_m\lambda + a_{m1}x_1 - a_{m2}x_2 - \cdots - a_{mn}x_n \leqslant b_m + d_m \\
& \lambda \geqslant 0, X_j \geqslant 0 \quad (j = 1,2,\cdots,n)
\end{aligned}
$$

第四步 求解上述一般线性规划即可得到最大隶属度 λ 和模型的最优解。

由上述可见，第二类模糊规划的求解实质上是把它变成求隶属度最大的一般线性规划问题。

6.8.2 模糊线性规划的应用

模糊线性规划的建模方法和原则与一般线性规划并无本质区别，只是在求解方法上有较大的不同。以前所介绍的所有类型模型均可将其转化为模糊线性规划模型。如果约束条件右端项无变动余地，只需将 d_i 设为零即可。因此本节主要通过一简单例子说明其求解方法。

6.8.2.1 第一类模糊线性规划的应用

某建筑公司木制品厂生产甲、乙两种实木门窗，已知甲种门窗每套可获利润 2 千元，乙种每项套获利 1 千元，甲生产资金占用量每套为 4 千元，乙每套 5 千元，企业现有生产资金 200 千元，通过挖掘潜力有可能再增加 5 千元。甲乙两种门窗的市场需求预测结果为：甲 40 套 ～ 45 套，乙 30 套 ～ 35 套。现决定最优生产方案以使企业两种门窗的生产获最大利润。

根据上述问题提出，建立如下模糊线性规划模型：

设 该企业生产甲产品 X_1 套，乙产品 X_2 套。

则

$$
\begin{aligned}
\max \quad & Z = 2X_1 + X_2 \\
\text{s.t.} \quad & 4X_1 + 5X_2 \lesssim 200 \\
& X_1 \lesssim 40 \\
& X_2 \lesssim 30 \\
& X_1 \geqslant 0, X_2 \geqslant 0
\end{aligned}
$$

求解该模型：

第一步 选择隶属函数 $\mu_i(d_i)$。

因该问题各模糊约束右端项各分量的增长均为 5，故 $B = (200,40,30)^T, B' = (200 + 5, 40 + 5, 30 + 5)^T$ 隶属函数均取直线递减形式，则

$$\mu_i(d) = \begin{cases} 1 - 0.2d, & 0 \leqslant d \leqslant 5 \\ 0 & d > 5 \end{cases}$$

第二步 用单纯形法求解一般线性规划模型。

$$\begin{aligned} \max \quad & Z = 2X_1 + X_2 \\ \text{s.t.} \quad & 4X_1 + 5X_2 \leqslant 205 \\ & X_1 \leqslant 45 \\ & X_2 \leqslant 35 \\ & X_1 \geqslant 0, X_2 \geqslant 0 \end{aligned}$$

其最优解为：$X_1^* = 45; X_2^* = 5; Z_0^* = 95$

第三步 将目标函数模糊化。

$$\mu_F(X) = \frac{1}{95}(2X_1 + X_2) = 0.021X_1 + 0.0105X_2$$

第四步 令 $\lambda_1 = 0.9, \varepsilon_0 = 0.01$，计算 d_i^λ, B^λ 并求解以 B^λ 为约束右端项的一般线性规划模型。

$$d_1^\lambda = d_2^\lambda = d_3^\lambda = \max\{d_i \mid \mu_i(d) \geqslant \lambda\}$$

$$d^\lambda = \max\{1 - 0.2d \geqslant 0.9\} = \max\{d \leqslant 0.5\} = 0.5$$

$$B^\lambda = (200 + 0.5, 40 + 0.5, 30 + 0.3) = (200.5, 40.5, 30.5)$$

一般线性规划模型为

$$\begin{aligned} \max \quad & 0.021X_1 + 0.0105X_2 \\ \text{s.t.} \quad & 4X_1 + 5X_2 \leqslant 200.5 \\ & X_1 \leqslant 40.5 \\ & X_2 \leqslant 30.5 \\ & X_1 \geqslant 0, X_2 \geqslant 0 \end{aligned}$$

求解得：$X_1^* = 40.5; X_2^* = 7.7; Z_1^* = 0.934$

第五步 计算 $\varepsilon_1 = \lambda_1 - Z_1^*$，判断最优解。

$$\varepsilon_1 = 0.9 - 0.934 = -0.034$$

$$|\varepsilon_1| = 0.034 > 0.01 = \varepsilon_0$$

故上述模型的解不是最优解，需再选截水平 λ_0。

$$\lambda_2 = \lambda_1 - r_1\varepsilon_1, \text{选 } r_1 = 1, \text{则}$$

$$\lambda_2 = 0.9 + 0.034 \times 1 = 0.934$$

第六步 计算 d_1^λ 和 B^λ 并求解一般线性规划。

$$d_\lambda = \max\{d_i \mid \mu_i(d) \geqslant \lambda\} = \max\{1 - 0.2d \geqslant 0.934\} = \max\{d \leqslant 0.33\} = 0.33$$

$$B^\lambda = (200.33, 40.33, 30.33)$$

一般线性规划模型为

$$\begin{aligned} \max \quad & 0.021X_1 + 0.0105X_2 \\ \text{s.t.} \quad & 4X_1 + 5X_2 \leqslant 200.33 \\ & X_1 \leqslant 40.33 \end{aligned}$$

$$X_2 \leqslant 30.33$$

$$X_1 \geqslant 0,\ X_2 \geqslant 0$$

其解为：$X_1^* = 40.33, X_2^* = 7.805, z_2^* = 0.931$

第七步 求 ε_2 并判断是否得到最优解

$$\varepsilon_2 = \lambda_2 - Z_2^* = 0.934 - 0.931 = -0.003 < \varepsilon_0$$

故最优水平 $\lambda^* = 0.934, X_1^* = 40.3, X_2^* = 7.802$，

最优值为：88.462。

6.8.2.2 第二类模糊线性规划的应用

某企业生产两种产品，甲产品每件可获利 30 元，乙产品可获利 40 元，两种产品在 A、B、C、D 各工序加工工时消耗和市场需求见表 6.8.2。

表 6.8.2 市场需求

	A	B	C	D	需求量
甲	15	30	20	35	300 ~ 400
乙	15	10	10	15	400 ~ 500

A 工序加工工时为 12000 工时，最多可达 13000 工时；B 工序加工工时为 16750 工时，最少可达 15750 工时；C 工序加工工时为 12500 工时，最多可达 13500 工时；D 工序加工工时最多为 22000 工时，最少为 21000 工时。企业确定的该月份目标利润在(30000 ~ 31000)元之间，如何搭配两种产品的生产才能实现企业目标。

设 甲、乙两品种产品产量为 X_1 和 X_2。

依题意可得如下模糊规划模型：

$$\begin{aligned}
30X_1 + 40X_2 &\lesssim 31000 \quad d_0 = -1000\\
15X_1 + 15X_2 &\lesssim 12000 \quad d_1 = 1000\\
30X_1 + 10X_2 &\lesssim 16750 \quad d_2 = -1000\\
20X_1 + 10X_2 &\lesssim 12500 \quad d_3 = 1000\\
35X_1 + 15X_2 &\lesssim 22000 \quad d_4 = -1000\\
X_1 &\lesssim 300 \quad d_5 = 100\\
X_2 &\lesssim 400 \quad d_6 = 100\\
X_1 \geqslant 0,\ X_2 &\geqslant 0
\end{aligned}$$

将上述模型化为标准模糊线性规划形式：

$$\begin{aligned}
-30X_1 - 40X_2 &\lesssim -31000 \quad d_0 = 1000\\
15X_1 + 15X_2 &\lesssim 12000 \quad d_1 = 1000\\
-30X_1 - 10X_2 &\lesssim -16750 \quad d_2 = 1000\\
20X_1 + 10X_2 &\lesssim 12500 \quad d_3 = 1000\\
-35X_1 - 15X_2 &\lesssim -22000 \quad d_4 = 1000\\
X_1 &\lesssim 300 \quad d_5 = 100\\
X_2 &\lesssim 400 \quad d_6 = 100\\
X_1 \geqslant 0,\ X_2 &\geqslant 0
\end{aligned}$$

根据第二类模糊线性规划的求解方法，其隶属函数分别为

$$\mu_0(X) = (1000 - 31000 + 30X_1 + 40X_2)/1000$$

$$\mu_1(X) = (1000 + 12000 - 15X_1 - 15X_2)/1000$$
$$\mu_2(X) = (1000 - 16750 + 30X_1 + 10X_2)/1000$$
$$\mu_3(X) = (1000 + 12500 - 20X_1 - 10X_2)/1000$$
$$\mu_4(X) = (1000 - 22000 + 35X_1 + 15X_2)/1000$$
$$\mu_5(X) = (100 + 300 - X_1)/100$$
$$\mu_6(X) = (100 + 400 - X_2)/100$$

由第二类模糊规划的最优判决条件,得如下一般线性规划模型:

$$\begin{aligned}
&\max \quad \lambda \\
&\text{s.t.} \quad \lambda \leqslant (1000 - 31000 + 30X_1 + 40X_2)/1000 \\
&\qquad \lambda \leqslant (1000 + 12000 - 15X_1 - 15X_2)/1000 \\
&\qquad \lambda \leqslant (1000 - 16750 + 30X_1 + 10X_2)/1000 \\
&\qquad \lambda \leqslant (1000 + 12500 - 20X_1 - 10X_2)/1000 \\
&\qquad \lambda \leqslant (1000 - 22000 + 35X_1 + 15X_2)/1000 \\
&\qquad \lambda \leqslant (100 + 300 - X_1)/100 \\
&\qquad \lambda \leqslant (100 + 400 - X_2)/100 \\
&\qquad \lambda \geqslant 0, X_1 \geqslant 0, X_2 \geqslant 0
\end{aligned}$$

整理得

$$\begin{aligned}
&\max \quad \lambda \\
&\text{s.t.} \quad 1000\lambda - 30X_1 - 40X_2 \leqslant -30000 \\
&\qquad 1000\lambda + 15X_1 + 15X_2 \leqslant 13000 \\
&\qquad 1000\lambda - 30X_1 - 10X_2 \leqslant -15750 \\
&\qquad 1000\lambda + 20X_1 + 10X_2 \leqslant 13500 \\
&\qquad 1000\lambda - 35X_1 - 15X_2 \leqslant -21000 \\
&\qquad 100\lambda + X_1 \leqslant 400 \\
&\qquad 100\lambda + X_2 \leqslant 5000 \\
&\qquad \lambda \geqslant 0, X_1 \geqslant 0, X_2 \geqslant 0
\end{aligned}$$

求解该一般线性规划问题,得

$$\lambda = 0.25, X_1 = 375, X_2 = 475$$

企业可实现利润 $30 \times 375 + 40 \times 475 = 30250$ 元。

6.9 灰色线性规划及其应用

6.9.1 灰色线性规划及求解的基本原理

6.9.1.1 基本概念及原理

灰色线性规划是灰色系统最优化的重要方法之一。

灰色系统是指系统中部分信息清楚,部分信息不清楚的系统。系统中清楚的信息称为

白色信息，不精确的信息称为黑色信息，而介于二者之间的的信息称为灰色信息。这类系统通常用灰色元素 $\otimes$（简称灰元）、灰矩阵 ——$\otimes A$ 和用各种数学运算符号表示其关系的灰色方程来描述。

含有灰色元素的线性规划称为灰色线性规划。灰色线性规划中经常用有界灰元，即 $\otimes \in [\underline{c}, \infty]$ 或 $\otimes \in [-\infty, \bar{c}]$ 表示。将该类灰元按某种准则、规则、概念进行量化或以其上界或下界值作为估计值的方法称为灰元白化，简称白化，灰元白化后产生的数值为白化值。

根据灰色规划中灰元的白化方法，灰色规划分为预测型和漂移型灰色规划两种类型。两种类型灰色规划求解的基本原理均是将灰色线性规划中的灰元白化，从而得到一普通线性规划，对该线性规划求解后，按某些原则进行处理从而得到该灰色规划的最优解。

6.9.1.2 预测型灰色规划模型及求解

预测型灰色线性规划是指约束条件中存在灰元且约束值中某一量是由时间序列描述的。其基本形式为：

$$\begin{aligned} \max \quad & \boldsymbol{Z} = \boldsymbol{CX} \\ \text{s.t.} \quad & \otimes \boldsymbol{AX} \leqslant \boldsymbol{B}(t) \\ & \boldsymbol{X} \geqslant 0 \end{aligned}$$

式中 $\boldsymbol{B}(t)$—— 时间序列。

该类灰色规划的求解方法是将 $\otimes A$ 中灰元素按上限值白化，再按 $GM(1,1)$ 模型对 $B_t(t)$ 进行预测，确定 t 时间的约束值，从而得到一普通线性规划，求解该线性规划，则求出其最优解。

6.9.1.3 漂移型灰色规划模型及求解

当目标函数和约束条件中均含有界灰元时，该类线性规划称漂移型灰色线性规划。其基本描述形式为

$$\begin{aligned} \max \quad & Z = \otimes \boldsymbol{CX} \\ \text{s.t.} \quad & \otimes \boldsymbol{AX} \leqslant \boldsymbol{B} \\ & \boldsymbol{X} \geqslant 0 \end{aligned}$$

其中，$\otimes(C_i) \in [\underline{C}_i, \overline{C}_i]$；$\otimes(a_{ij}) \in [\underline{a}_{ij}, \bar{a}_{ij}]$

令 $0 \leqslant \alpha \leqslant 1$，可按下式将灰元 $\otimes(c_i)$ 和 $\otimes(a_{ij})$ 白化得到白化值。

$$\widetilde{\otimes}(C_i) = \underline{C}_i + \alpha(\overline{C}_i - \underline{C}_i)$$

$$\widetilde{\otimes}(a_{ij}) = \underline{a}_{ij} + \alpha(\bar{a}_{ij} - \underline{a}_{ij})$$

其中，α 称为白化系数。

当 $\alpha = 1$ 时，白化值取上限；当 $\alpha = 0$ 时，白化值取下限；当 $\alpha \in (0,1)$ 时，则白化值将在灰元上限与下限之间漂移，故记为 $\widetilde{\otimes}(c_i)$ 和 $\widetilde{\otimes}(a_{ij})$。

在上述灰色规划模型中，如果约束条件的白化值取上限，即 $\alpha = 1$，则满足约束条件的 X 取最小值，记为 $X_{\min}$；反之则取最大值，记为 $X_{\max}$。目标函数中灰元取上限，($\alpha = 1$)，则目标函数中变量有最大的权，反之有最小的权。

如果当约束条件中白化系数 $\alpha = 0$，目标函数中白化系数 $\alpha = 1$，则该灰色线性规划必有最大值；反之必有最小值，分别记为 $Z_{\max}$ 和 $Z_{\min}$。

如果在目标函数和约束条件中，α 均取相同的值，即按一致性取数原则，该灰色线性规划的极值必然在 Z_{min} 和 Z_{max} 之间，故下式成立：

$$Z_{min} \leqslant Z \leqslant Z_{max}。$$

设 C_α —— 目标函数中灰元取 α 时的变量系数；

X_α —— 对应于灰元取 α 时的变量的解。

则定义

$$\mu_\alpha = \frac{C_\alpha^T X_\alpha}{Z_{max}}$$

为白化系数取 α 时的可信度，当 μ_α 超过某一预先规定的标准时，即得到该问题的最优解。

其求解步骤为：

(1) 对约束条件取白化系数 $\alpha = 0$，对目标函数白化系数取 $\alpha = 1$，将灰色线性规划变为一般线性规划。求解该线性规划，得出最优解和极值 Z_{max}。

(2) 给出 α 值，按一致性取数原则，将灰色规划变为一般线性规划，求解得出最优解和极值 Z_α。

(3) 给定可信度 μ 的预期值 μ^*，并计算 μ_α：

$$\mu_\alpha = \frac{Z_\alpha}{Z_{max}}$$

(4) 如果 $\mu_\alpha \geqslant \mu^*$，则 Z_α 为该灰色规划的极值，X_α 为其最优解。否则再选 α 按上步计算，直至使 $\mu_\alpha \geqslant \mu^*$ 为止。

由上述可见，灰色线性规划的求解，最终仍归结为一般线性规划的求解。只是以可信度是否满足要求作为选择最优解的标准。

6.9.2 灰色线性规划的应用

某建筑公司木材加工厂地板生产车间生产两种规格木地板，由于从原木外观很难准确确定各种地板块生产的原木消耗，因而销售木地板所得利润也难以准确估计，经研究历史生产状况，得如下表资料：

	甲产品	乙产品	资源量
原木消耗量(m^3)	1 ~ 21	4 ~ 10	360
电力消耗 /kW	4	5	198
工时消耗 /h	3	10	300
单位利润 / 千元	1 ~ 7	4 ~ 12	

设：甲、乙两种地板块的产量分别为 X_1 和 X_2。

由上述资料得灰色线性规划模型如下：

$$\begin{aligned} \max \quad & Z = \otimes(c_1)X_1 + \otimes(c_2)X_2 \\ \text{s.t.} \quad & \otimes(a_{11})X_1 + \otimes(a_{12})X_2 \leqslant 360 \\ & 4X_1 + 5X_2 \leqslant 198 \\ & 3X_1 + 10X_2 \leqslant 300 \end{aligned}$$

$$X_1 \geqslant 0,\ X_2 \geqslant 0$$

式中 $\otimes(c_1) \in [1,7], \otimes(c_2) \in [4,12]$,

$\otimes(a_{11}) \in [1,21], \otimes(a_{12}) \in [4,10]$

求解该漂移型线性规划。

(1) 构造漂移模型:

$$\widetilde{\otimes}(c_1) = c_1 + \alpha(\bar{c}_1 - \underline{c}_1) = 1 + \alpha\,(7-1) = 1 + 6\alpha$$

$$\widetilde{\otimes}(c_2) = c_2 + \alpha(\bar{c}_2 - \underline{c}_2) = 4 + \alpha\,(12-4) = 4 + 8\alpha$$

$$\widetilde{\otimes}(a_{11}) = a_{11} + \alpha\,(\bar{a}_{11} - \underline{a}_{11}) = 1 + \alpha\,(21-1) = 1 + 20\alpha$$

$$\widetilde{\otimes}(a_{12}) = a_{12} + \alpha\,(\bar{a}_{12} - \underline{a}_{12}) = 4 + \alpha\,(10-4) = 4 + 6\alpha$$

对目标函数取 $\alpha = 1$,对约束条件取 $\alpha = 0$ 建立一般线性规划模型并求解。

当 $\alpha = 1$ 时,目标函数变为

$$\max \quad Z = 7X_1 + 12X_2$$

当 $\alpha = 0$ 时,约束条件变为

$$\begin{aligned} X_1 + 4X_2 &\leqslant 360 \\ 4X_1 + 5X_2 &\leqslant 198 \\ 3X_1 + 10X_2 &\leqslant 300 \\ X_1 \geqslant 0,\ X_2 &\geqslant 0 \end{aligned}$$

求解该线性规划模型得

$$X_1 = 19.2, X_2 = 24.24, Z_{\max} = 425.28$$

(2) 设 $\mu^* = 0.85$,并令 $\alpha = 0.5$,求 μ_α。

当 $\alpha = 0.5$ 时,按一致取数原则:

$$\otimes(c_1) = 1 + 6 \times 0.5 = 4 \qquad \otimes(a_{11}) = 1 + 20 \times 0.5 = 11$$

$$\otimes(c_2) = 4 + 8 \times 0.5 = 8 \qquad \otimes(a_{21}) = 4 + 6 \times 0.5 = 7$$

则一般线性规划模型为

$$\begin{aligned} \max \quad & Z_{\alpha=0.5} = 4X_1 + 8X_2 \\ \text{s.t.} \quad & 11X_1 + 7X_2 \leqslant 360 \\ & 4X_1 + 5X_2 \leqslant 198 \\ & 3X_1 + 10X_2 \leqslant 300 \\ & X_1 \geqslant 0,\ X_2 \geqslant 0 \end{aligned}$$

求解该线性规划问题得:

$$X_1 = 16.85, X_2 = 24.94, Z_{\alpha=0.5} = 266.92$$

故 $\mu_{\alpha=0.5} = \dfrac{Z_\alpha}{Z_{\max}} = \dfrac{266.92}{425.28} = 0.627$

(3) 判断最优解。

因 $\mu^* > \mu_{\alpha=0.5}$,故不满足最优条件。

(4) 取 $\alpha = 1$ 再按以上步骤计算,此时一般线性规划模型如下:

$$\begin{aligned}\max \quad & Z_{a=1} = 7X_1 + 12X_2 \\ \text{s.t.} \quad & 21X_1 + 10X_2 \leqslant 360 \\ & 4X_1 + 5X_2 \leqslant 198 \\ & 3X_1 + 10X_2 \leqslant 300 \\ & X_1 \geqslant 0,\ X_2 \geqslant 0\end{aligned}$$

求解得：$X_1 = 3.34, X_2 = 29, Z_{a=1} \approx 371.38$

$$\mu_{a=1} = \frac{371.328}{425.28} = 0.873$$

$$L^* < L_\Delta = 1$$

故得最优解为

$X_1 = 3.34, X_2 = 29, Z^* = 371.38$。

6.10 Lingo 软件简介

Lingo软件是由美国 Lindo system 公司开发的求解线性规划、整数规划和非线性规划的通用软件。可由 www.lindo.com 网站下载试用版。对形式简单的模型，可直接输入模型求解；对复杂的模型可采用该软件提供的简单的语句进行描述，现简要介绍该软件所提供的语言并举例说明其用法。

6.10.1 Lingo 语言简述

Lingo 所提供的模型描述语言对模型的描述以 Model：开始，以 End 结束，中间由三段组成，第一段为设置段，由 Sets：开始，至 Endsets 为止；第二段为数据段，由 Data：开始，至 Enddata 为止；第三段为目标和约束段，由模型的约束或目标开始，由 End 结束。为使模型易读，语言用“!”作为注释语句的开始，每条语句均需以“;”结束，语言字母不分大小写。

6.10.1.1 设置段

模型中向量、矩阵的名称及阶次是由设置段完成的。该段由 Sets：开始，由 Endsets 结束，中间为设置语句，设置语句的语法格式为：

Setsname / 清单 | 条件 /：[属性 1]，[属性 2]，…，[属性 N]；

Setsname 为向量或矩阵名；清单为表示向量阶次的指示数或向量元素的名称，可用数字、字母表示；条件是选择清单的依据；属性是具有向量特性的变量和常量。如语句 Cangky/1，2，3，4 /：a；中 Cangky 表示仓库向量，1,2,3,4 表示共有四个仓库，1,2,3,4 为仓库名；a 为四个仓库的物资拥有量，它与 Cangky 是同阶的向量。a(1) 为第 1 仓库的拥有量，a(i) 为第 i 个仓库的拥有量。

向量、变量、常量名必须以字母开始，不能用汉字。清单的描述方法有多种，如“Cangky / jia，yi，bing，ding/：”、“Cangky / 1 . . 4 /”与“Cangky/1，2，3，4 /：”是等价的。

两个以上的向量可以组成矩阵，其语句形式为：

Setsname(Setsmame1, Setsname2, …) :X,C;

如向量 Cangky / 1..4 / : a;和 Kehy / 1..5 / : b;可形成矩阵:

Ck - Kehy (Cangky, Kehy): X,C;

其中 Ck - Kehy 为矩阵名,它由 Cangky (仓库) 和 Kehy(客户) 两个向量组成的表示仓库与客户间联系的矩阵,X 和 C 为与该矩阵有相同属性的变量或常量。

4 个供应单位、5 个需求单位的平衡型运输模型设置语句如下:

```
Sets:
    Cangky / 1..4 / : a;
    Kehy / 1..5 / : b;
    Ck - Kh ( Cangky, Kehy ): X,C;
Endsets
```

6.10.1.2 数据段

数据段是设置模型中常量的段。在设置段中并未指明哪个量是常量,哪个量是变量,如果某向量或矩阵在设置段中设置了属性,在数据段中又被赋值,则为常量,否则即是变量。

数据段由 Data:开始,至 Enddata 结束,中间即是数据赋值语句,其格式是

常量名 = 数据 1,数据 2,…,… ;

如运输模型中的数据赋值语句如下:

```
Data:
    a = 1000, 2000, 3000, 4000;        ! 四个仓库的拥有量;
    b = 2000, … ;                      ! 五个客户的需求量;
    c = … ;                            ! i 仓库向 j 客户的单位运费用矩阵;
    …                                  ! X 未赋值,说明是变量;
    Enddata
```

6.10.1.3 约束与目标函数段

该段用以描述模型的目标函数和约束条件。对于形式简单的模型可进行输入,对于描述上相对复杂的模型可采用 Lingo 语句描述。

用 Lingo 语句描述目标和约束时,除用注释语句说明外,尚可在语句前加上约束名和目标,其形式为[ST1] , [object] 等。

对于表达式复杂的约束和目标来说,循环和累加是主要运算,如运输模型的约束条件和目标函数为

$$\text{s.t} \quad \sum_j X_{ij} = a_i \qquad (I = 1, 2, 3, \cdots, m)$$

$$\sum_i X_{ij} = b_j \qquad (j = 1, 2, \cdots, n)$$

$$\min \quad Z = \sum_i \sum_j C_{ij} \cdot X_{ij}$$

均可用循环语句和累加语句完成其描述。

循环语句的一般描述形式为:

@ For(设置名(列表) | [条件]:循环体);

累加语句的一般描述形式为：

@ Sum(设置名(列表)|[条件]:累加体);

循环体和累加体中均可嵌入循环和累加语句。

上述运输模型约束和目标函数可描述为：

```
@ For ( Cangky (i): [ST1]
      @ Sum ( Kehy ( j ): X (i,j) ) = A (i) );      !拥有量约束;
@ For ( kehy ( j ): [ST2]
      @ Sum ( Cangky ( i ): X( i,j ) ) = B ( j ) ); !需求量约束;
[ obj ]  Min = @ Sum ( Cangky (i):
      @ Sum (kehy ( j ): C ( i,j )* X ( i, j )));
```

整数规划和(0,1)规划模型可对变量加以限制，Lingo 提供了对变量进行限制的函数，常用的有：

@ Gin(变量名) —— 整数变量函数；

@ Bin(变量名) —— 二进数变量函数(01 规划用)；

@ Bnd(下界,变量名,上界)—— 变量带上下界的函数。

如运输问题中要求变量 X 为整数时，该线性规划即为整数规划。在约束方程中应用整数变量函数即可：

```
@ For ( Cangky (i):@ For ( Kehy ( j ):
      [ST] @ Gin ( X ( i,j ) ) );  !整数约束;
```

如运输模型中客户需求量有上下限约束，在设置语句中将 b 改为 bs,bx(需求量的上限和下限)两个常量向量，则上述[ST]约束可改写为：

```
@ For ( Kehy ( j ):
      @ Sum ( Cangky (i ): @ Bnd ( b(j), X( i, j ), b(j) ) ) );
```

6.10.2 软件应用举例

6.10.2.1 运输问题

设有运输问题见表 6.10.1。

表 6.10.1 运输问题

	1	2	3	4	5	拥有(a_i)
甲	3	2	5	4	6	1000
乙	7	3	2	1	5	2000
丙	3	1	2	5	8	3000
丁	2	5	6	7	9	4000
需求量上限	2000	1500	2500	4500	1500	12000
需求量下限	1500	1000	2000	2500	1000	8000

用 Lingo 语句描述如下：

```
MODEL：需求有上下限的运输模型例
SETS：
```

```
cangku/1..4/:a; !四个仓库,拥有量 a;
kehu /1..5/:bx,bs,z; !五个客户,需求量上限为 bs,下限为 bx;
ck_kh(cangku,kehu):x,c; !单位运费矩阵为 c(i,j), 决策变量为 x(i,j);
ENDSETS
DATA:
a = 1000,2000,3000,4000;
bs = 2000,1500,2500,4500,1500;
bx = 1500,1000,2000,2500,1000;
c = 3,2,5,4,6,
    7,3,2,1,5,
    3,1,2,5,8,
    2,5,6,7,9;
ENDDATA
@for(cangku(i):
    [st1]@sum(kehu(j):x(i,j)) <= a(i));!拥有量限制;
@for(kehu(j):
    [st2]@sum(cangku(i):x(i,j)) = z(j));!需求量;
@for(kehu(j):@bnd(bx(j),z(j),bs(j)));!需求量上下限限制;
[obj]min = @sum(cangku(i):@sum(kehu(j):c(i,j) * x(i,j)));
END
```

求解结果为:

```
obj = 19500
z = 1500,1000,2000,2500,1000
x =    0,    0,    0,    0, 1000,
       0,    0,    0, 2000,    0,
       0, 1000, 2000,    0,    0,
    1500,    0,    0,  500,    0,
```

6.10.2.2 生产问题

某公司用三种原料生产三种产品,其需求上下限、单位利润、公司材料拥有量及单位产品原料消耗量见表 6.10.2。该公司应如何安排生产才能使公司获利最大。

表 6.10.2 原始数据表

	产品甲	产品乙	产品丙	拥有量
原料 A	5	8	2	500
原料 B	7	2	3	500
原料 C	6	5	7	600
单位利润	7	8	5	
需求上限	60	100000	30	
需求下限	40	20	0	

设 x_j 为第 j 种产品的产量；a_i 为 i 原料的拥有量；bs_j 为 j 产品需求量的上限；bx_j 为 j 产品需求量的下限；c_{ij} 为 j 产品对 i 原料的单位消耗；c_j 为 j 产品的单位利润。

该问题的线性规划模型为

$$\begin{aligned} \max \quad & z = \sum_{1}^{3} C_j \cdot X_j \\ \text{s.t.} \quad & \sum_{j} C_{ij} \cdot X_j \leqslant a_i \quad (i = 1,2,3) \\ & bx_j \leqslant X_j \leqslant bs_j \quad (i = 1,2,3) \\ & X_J \geqslant 0 \end{aligned}$$

用 LINGO 语言编写的程序如下：

```
MODEL:
SETS:
chp/1..3/:x,bs,bx,c,xql;
yl/1..3/:a;
chpyl(yl,chp):cij;
ENDSETS
DATA:
bs = 60,100000,30;
bx = 40,20,0;
c = 7,8,5;
a = 500,500,600;
cij = 5,8,2,
    7,2,3,
    6,5,7;
ENDDATA
@for(yl(i):
        @sum(chp(j):cij(i,j) * x(j)) < = a(i));
@for(chp(j):
        xql(j) = @sum(yl(i):
                    cij(i,j) * x(j)));
@for(chp(j):
    @BND( bx(j),xql(j) ,bs(j)));
[obj]max = @sum(chp(j):c(j) * x(j));
END
```

求解结果为：

目标值：514.1667，$X_1 = 3.33$，$X_2 = 59.79$，$X_3 = 2.5$。

第7章　非线性静态系统最优化模型及求解方法

7.1　非线性系统最优化模型

线性规划模型有很大的局限性，它只适用于变量间呈现线性关系的系统。然而实际系统往往不能用线性函数表示其目标，也不能用线性关系表示其约束。因此，研究非线性系统最优化模型的建立和求解方法是很必要的。

非线性最优化模型是目标函数 $f(X)$ 或约束条件 $g_i(X)\geqslant 0$ 中的任何一个为非线性函数的模型，应用这种模型解决的问题叫非线性规划问题。

非线性规划模型的一般形式为

$$\begin{aligned} &\min \quad Z = F(X) \\ &\text{s.t.} \quad G(X)\geqslant 0 \end{aligned} \tag{7.1.1}$$

非线性规划模型按约束条件的数量可分为有约束和无约束两种。有约束条件的非线性规划模型按约束条件的形式又分为等式约束和不等式约束两种。

非线性规划模型中有些常用的特殊形式，如当约束条件为线性，而目标为二次函数时，这种非线性规划叫二次规划，其一般形式为

$$\begin{aligned} &\min \quad \boldsymbol{Z} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{X} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{X} + \boldsymbol{K} \\ &\text{s.t.} \quad \boldsymbol{A}\boldsymbol{X}\geqslant \boldsymbol{B} \\ &\qquad \boldsymbol{X}\geqslant 0 \end{aligned} \tag{7.1.2}$$

式中　$\boldsymbol{Q}$—— 正定或半正定对称矩阵。

当目标函数和约束条件是多元多项式时，这种规划叫几何规划，其一般形式为

目标函数：
$$\min \quad F(X) = \sum_{i}^{m} C_i\left[\prod_{j=1}^{n} X_j \cdot p_{ij}\right]$$

$$\text{s.t.} \quad G(X) = \sum a_i\left[\prod_{j-1}^{n} X_j \cdot q_{ij}\right] \tag{7.1.3}$$

各种不同形式的模型，可解决不同类型的问题，也有不同的求解方法，本节主要介绍各种模型的建立方法。

7.1.1　最优选址问题

已有若干现成的设施，现欲建立一服务设施，如何选择建设地址，使付出的代价最小，例如一个居民区，把学校、商店、幼儿园选在什么位置使群众最方便，一个林区木材加工厂设在什么位置才能使各林业局所提供的木材运费最低等都是最优选址问题，它们可用无

约束的非线性规划模型来解决。如果学校、幼儿园不能建在繁华的闹市区或主要街道旁，加工厂不能建在生产易燃、易爆产品的工厂附近，这就构成了对问题的某种限制，这类问题可用有约束的非线性规划模型来解决。

现举例说明选址问题模型的建立方法。

某公司准备建一临时木材加工点，向各工地供应门窗和木料，现需确定加工点建在什么位置，才能使加工点向各工地供应材料的费用最低。

设加工点向第 i 个工地的原料供应量为 w_i；

单位原材料的运费为 β(元 /m^3 · km)。

采用笛卡儿坐标系，设木材加工点位置的坐标为 $p(x,y)$，各工地位置的坐标为 $p_i(x_i,y_i)$(见图 7.1.1)，则第 i 个工地与木材加工点的距离可由解析几何的两点间距离公式求得

$$d_i = \sqrt{(x-x_i)^2 + (y-y_i)^2}$$

木材加工点向第 i 个工地供应原料的运费为

$$C_i = w_i \cdot d_i \cdot \beta$$

加工点向各工地供应原料的总运费为

$$C = \sum_i c_i = \beta \sum_i w_i d_i = \beta \sum_i w_i \sqrt{(x-x_i)^2 + (y-y_i)^2}$$

因此，该问题的目标函数为

$$\min \quad C = \beta \sum w_i \sqrt{(x-x_i)^2 + (y-y_i)^2} \tag{7.1.4}$$

根据该模型，选择适当的 x、y 就可使 C 达最小。

由数学分析知，求函数极小值的必要条件为

$$\frac{\partial C}{\partial x} = 0 \qquad \frac{\partial C}{\partial y} = 0$$

对 C 求偏导得

$$\frac{\partial C}{\partial x} = \beta \sum_i \frac{w_i}{d_i}(x - x_i) = 0 \tag{7.1.4a}$$

$$\frac{\partial C}{\partial y} = \beta \sum_i \frac{w_i}{d_i}(y - y_i) = 0 \tag{7.1.4b}$$

解该联立方程得

$$X = \frac{\sum_i \frac{w_i x_i}{d_i}}{\sum_i \frac{w_i}{d_i}} \tag{7.1.4c}$$

$$Y = \frac{\sum_i \frac{w_i x_i}{d_i}}{\sum_i \frac{w_i}{d_i}} \tag{7.1.4d}$$

由式(7.1.4c)和(7.1.4d)知，d_i 仍为 x、y 的函数，故该解不是以显函数形式给出的，可采用迭代法求解。其迭代过程为：预先给定 $x^{(0)}$、$y^{(0)}$，代入式(7.1.4c)和式(7.1.4d)的右端，得到 $x^{(1)}$、$y^{(1)}$，再代入式(7.1.4c)和式(7.1.4d)的右端得到 $x^{(2)}$、$y^{(2)}$，…，如此计算下去，直到计算出 $x^{(n+1)}$、$y^{(n+1)}$，使 $\| p(x^{(n+1)}, y^{(n+1)}) - p(x^{(n)}, y^{(n)}) \| \leqslant \varepsilon$（$\varepsilon$ 为预先确定的精度）为止，即可得到满足一定精度要求的木材加工厂位置坐标。

如果在某个位置不允许建立加工点，可将位置以一约束条件表示。如在原点附近是一个半径为 R 的水库，且有一条河(见图 7.1.2)，因此加工点不能建在河流和水库上，该限制条件可描述为

$$x^2 + y^2 \geqslant R^2$$
$$x + y \neq 1$$

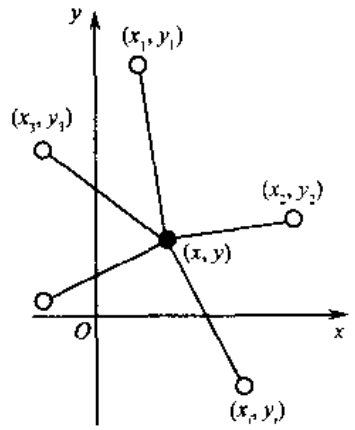


图 7.1.1　各工地位置的坐标(一)

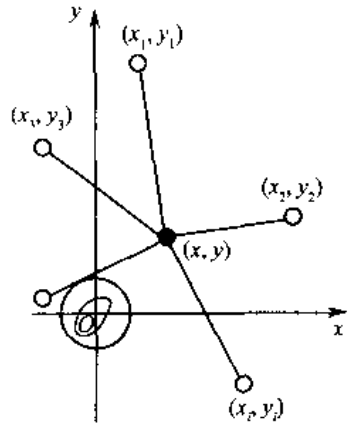


图 7.1.2　各工地位置坐标(二)

因此得有约束的非线性规划模型为

$$\min \quad C = \beta \sum_i w_i d_i$$

s.t.　(1) 加工点不能建在水库上：

$$x^2 + y^2 \geqslant R^2 \tag{7.1.5}$$

(2) 加工点不能建在河流中：

$$x + y \neq 1$$

7.1.2　最大体积问题

从边长 1 m 的正方形铁皮上切下相等四个方块，将各边折成一个无盖的铁盒，问怎样裁法才能使其容积最大？

该问题在数学分析中是一个典型的求函数极值的问题，实质上该问题是一典型的有约束的非线性规划。

设剪下方块的边长为 X，则模型为(见图 7.1.3)

目标函数：

$$\max \quad V = X(1-2X)^2$$
$$\text{s.t.} \quad 0 < X < 0.5$$

再如，假设现有长 1.5 m，厚 1 cm 的箱板材 10 m^3，如制成包装箱，可制多少个多大规

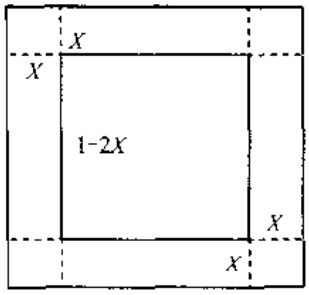


图 7.1.3　正方形铁皮

格的包装箱,才能使所装货物最多?

设制成木箱的规格为 $X \times Y \times Z$(长 × 宽 × 高),所制木箱个数为 n,则模型为

$$\max \quad V = nXYZ \tag{7.1.6}$$

s.t.　① 木材尽量充分利用

$$(0.02XZ + 0.02YZ + 0.02XY) \cdot n \leqslant 10 \tag{7.1.7}$$

② 长和宽不能超过材长

$$X \leqslant 1.5$$
$$Y \leqslant 1.5$$

③ 木箱个数、长度、宽度、高度非负

$$n \geqslant 0 \quad X \geqslant 0 \quad Y \geqslant 0 \quad Z \geqslant 0$$

7.1.3　最优生产批量问题

最优生产批量问题又叫非线性盈亏分析。

在企业的经营管理中,经常需要以最大利润来安排生产,因此需要确定生产批量与利润之间的关系。

众所周知,利润为销售收入与总成本费用之差,总成本费用为固定成本与变动成本之和。

设生产量等于销售量,且均为 Q,销售收入为 R,固定成本为 b_0。

当销售收入与销售量呈二次函数关系(见图 7.1.4)时,可描述为

$$R = a_1Q + a_2Q^2 \tag{7.1.8}$$

如变动成本与产量之间的关系也为二次函数关系,则

$$C_1 = b_1Q + b_2Q^2$$

总费用为 $C = b_0 + b_1Q + b_2Q^2$

利润 P 与销售收入之间的关系为

$$P = R - C = a_1Q + a_2Q^2 - b_0 - b_1Q - b_2Q^2$$

整理得

$$P = (a_1 - b_1)Q + (a_2 - b_2)Q^2 - b_0$$

以最大利润为目标的生产批量模型为

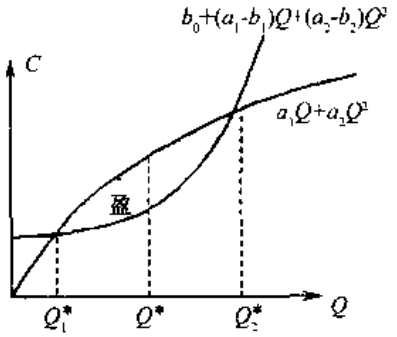


图 7.1.4　$C-Q$ 关系图

$$\max \quad P(Q) = (a_1 - b_1)Q + (a_2 - b_2)Q^2 - b_0 \tag{7.1.9}$$

这是一无约束非线性规划模型，应用数学分析方法即可求解。

由$\frac{dP}{dQ} = 0$得

$$Q^* = \frac{a_1 - b_1}{2(a_2 - b_2)} \tag{7.1.10}$$

式中 Q^*——最佳生产批量。

最大利润值为

$$P_{max} = (a_2 - b_2)Q^{*2} + (a_1 - b_1)Q^* - b_0 \tag{7.1.11}$$

如令$P = 0$，则可得到盈亏平衡点Q_1^*和Q_2^*。Q_1^*和Q_2^*是二次方程式$(a_2 - b_2)Q^{*2} + (a_1 - b_1)Q^* - b_0 = 0$的解，区间$[Q_1^*, Q_2^*]$是企业的盈利区间。

所建立的非线性规划模型可为企业经营决策提供科学的依据。

7.1.4 库存问题

应用非线性规划研究生产物资供应系统，可得各种库存模型。

7.1.4.1 确定型不允许缺货的存储系统模型(一)

对于生产周期较短的系统，需一次备足一段时间内生产所需的物资，当库存被消耗后，必须立即得到补充，以维持生产的正常进行，这就是不允许缺货的存储系统。研究该系统的目的是适当安排进货批量，使库存系统单位时间内所支出的费用最小。

该模型有以下假设：

(1) 当所需物资不足时，因缺货所造成的损失(缺货费用)为∞；

(2) 当库存量降低到零时，立即可以得到补充；

(3) 生产对物资的需求是连续的、均匀的、且单位时间的物资消耗量为R；

(4) 每次进货量Q不变；

(5) 物资存贮的单位费用不变，用h表示。

根据上述假设，现分析一个存储周期内与存储有关的各种费用和存储量之间的关系。

一个存储周期是由进货、生产消耗到再进货为止的全部时间，是两次进货所间隔的时间t。在t时间内，系统的总费用C由下式决定：

$$C = T_1 + T_2 \tag{7.1.12}$$

式中 T_1——物资存储费用；

T_2——组织进货费用。

根据假设，进货量Q应满足t时间内的生产需求，所以

$$Q = Rt \tag{7.1.13}$$

t时间内的平均储量为$\triangle OQt$的面积(见图7.1.5)，亦即

$$\bar{Q} = \int_0^t (Q - R\tau)d\tau = \frac{1}{2}Rt^2 \tag{7.1.14}$$

所以存储费用为

$$T_1 = Q \cdot h = \frac{1}{2}Rt^2h \tag{7.1.15}$$

组织进货费用T_2由两部分组成，一部分是固定费用C_1，它是与进货数量无关的费用，如旅费等；一部分是与进货数量有关的变动费用，如运费等，它与进货量成正比，单位

变动费用若以 C_2 表示，则变动费用为 C_2Q。所以组织进货费用为

$$T_2 = C_1 + C_2Q = C_1 + C_2Rt \tag{7.1.16}$$

总费用为

$$C = T_1 + T_2 = \frac{1}{2}Rt^2h + C_1 + C_2Rt \tag{7.1.17}$$

单位时间总费用为

$$T(t) = \frac{C}{t} = \frac{1}{2}Rth + \frac{C_1}{t} + C_2R \tag{7.1.18}$$

由式(7.1.18)可见，单位时间的总费用是存储周期的函数，是由直线 $\frac{1}{2}Rth$ 和曲线 $\frac{C_1}{t} + C_2R$ 合成的结果，如图 7.1.6 所示。

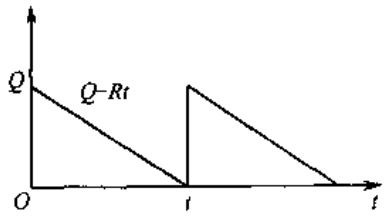


图 7.1.5　$Q-t$ 关系

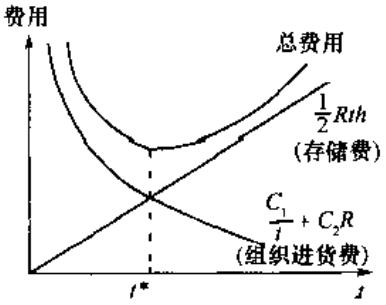


图 7.1.6　总费用与存储费用、进货费用的关系

以单位时间费用最小为目标函数建立的存储模型如下：

$$\min \qquad T(t) = \frac{1}{2}Rth + \frac{C_1}{t} + C_2R \tag{7.1.19}$$

对式(7.1.19)求微分并令其为零，即可得出最佳存储周期 t^*：

$$t^* = \sqrt{\frac{2C_1}{Rh}} \tag{7.1.20}$$

从而得出最优进货批量 Q^* 和单位时间的最小费用 $T(t^*)$ 为

$$\begin{aligned} Q^* &= Rt^* = \sqrt{\frac{2RC_1}{h}} \\ T(t^*) &= \sqrt{2RC_1h} + C_2R \end{aligned} \tag{7.1.21}$$

由式(7.1.20)、式(7.1.21)可见，最佳存储周期、最佳进货批量与 C_2 无关，它只由物资消耗速度、单位物资的存储费用和组织进货的固定费用决定。

如某建筑公司，每天需水泥 100 t，每次组织进货的固定费用为 300 元，水泥的存储费用为 0.1 元 /d·t，则最佳存储周期为

$$t^* = \sqrt{\frac{2C_1}{Rh}} = \sqrt{\frac{2 \times 300}{100 \times 0.1}} \approx 8 \text{ d}$$

最佳进货批量为

$$Q^* = Rt^* = 100 \times 8 = 800 \text{ t}$$

不考虑可变进货费用的单位最小费用为

$$T(t) = \sqrt{2 \times 100 \times 300 \times 0.1} = 77.46 \text{ 元}$$

7.1.4.2 确定型不允许缺货的存储系统模型(二)

该模型是针对生产周期相对长的物资供应系统所建立的模型。它与上述模型的差别在于为了节约费用,不需一次存足一个周期内生产所需的物资,而是以一定的速度进货,边进货边消耗,且为使生产不停顿,要求进货速度大于消耗速度。

设 P 为进货速度,且 $P > R$,由图 7.1.7 可见,$\triangle OZt$ 的面积为一个存储周期内的平均存储量,即

$$\overline{Q} = \int_0^{t_1} (P - R)\tau d\tau + \int_{t_1}^{t_2} (Q - R\tau) d\tau \tag{7.1.22}$$

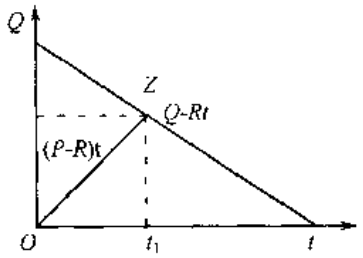


图 7.1.7 确定型不允许缺货的存储系统模型(二)

因为以 P 为进货速度、以 t_1 为进货时间的全部进货量 Q 必须满足以 R 为消耗速度、以 t 为消耗时间的全部需求,所以有:

$$Pt_1 = Rt$$

$$t_1 = \frac{R}{P}t \tag{7.1.23}$$

将式(7.1.23)代入式(7.1.22)得平均储量 $\overline{Q}$ 为

$$\overline{Q} = \int_0^{\frac{R}{P}t} (P - R)\tau d\tau + \int_{\frac{R}{P}t}^{t} (Q - R\tau) d\tau$$

积分后得

$$\overline{Q} = \frac{P - R}{2P} Rt^2 \tag{7.1.24}$$

平均存储费用为

$$T_1 = \overline{Q}h = \frac{P - R}{2P} Rt^2 h \tag{7.1.25}$$

进货费用为

$$T_2 = C_1 + C_2 Q = C_1 + C_2 Rt \tag{7.1.26}$$

一个周期的总费用为

$$C = T_1 + T_2 = \frac{P - R}{2P} Rt^2 h + C_1 + C_2 Rt \tag{7.1.27}$$

单位时间的总费用为

$$T(t) = \frac{C}{t_2} = \frac{P - R}{2P} Rth + \frac{C_1}{t} + C_2 R \tag{7.1.28}$$

故该系统的目标函数为

$$\min \quad T(t) = \frac{P-R}{2P}Rth + \frac{C_1}{t} + C_2R \tag{7.1.29}$$

用微分求极值的方法求解该模型，令$\frac{dT(t)}{dt} = 0$得方程

$$\frac{P-R}{2P}Rh - \frac{C_2}{t^2} = 0$$

解该方程得最佳存储周期 t_2^* 和最佳批量 Q^* 为

$$t^* = \sqrt{\frac{2PC_1}{Rh(P-R)}} \tag{7.1.30}$$

$$Q^* = \sqrt{\frac{2PC_1R}{h(P-R)}} \tag{7.1.31}$$

由 t^* 和 Q^* 可见，当时进货速度远远大于物资的消耗速度，即 $P >> R$ 时，则：$t^* = \sqrt{\frac{2C_1}{Rh}}$，$Q^* = \sqrt{\frac{2C_1R}{h}}$，即变为生产周期短的最佳周期和最佳批量公式。

7.1.4.3 确定型允许缺货的存储系统模型(一)

上述两个模型的假设为不允许缺货，即缺货产生的损失为 ∞。但某些生产过程可因缺少原材料而停产，因缺乏物资保证而导致停产的损失称之为缺货损失费用，设缺少单位物资的损失费用为 h_1，则对生产周期较短的系统，其模型建立过程(见图 7.1.8) 如下：

一个存储周期的总费用 C 为存储费用 T_1、进货费用 T_2 和缺货损失费用 T_3 之和，即

$$C = T_1 + T_2 + T_3 \tag{7.1.32}$$

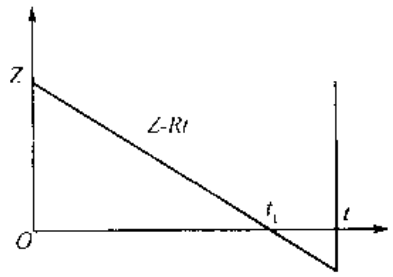


图 7.1.8 确定型允许缺货的存储系统模型(一)

$\triangle OZt_1$ 的面积为平均储量(见图 7.1.8)。

$$\overline{Q} = \int_0^t (Z - Rt)d\tau \tag{7.1.33}$$

由于一次进货量 Z 可满足 t_1 时间的需求，故：$Z = Rt_1 \qquad t_1 = \frac{Z}{R}$

将其代入式(7.1.33) 得

$$\overline{Q} = \int_0^{\frac{Z}{R}} (Z - Rt)dt = \frac{Z^2}{2R} \tag{7.1.34}$$

故存储费用 T_1 为

$$T_1 = \overline{Q} \cdot h = \frac{Z^2}{2R}h \tag{7.1.35}$$

$\triangle t_1tZ'$ 的面积为平均缺货量 $\overline{Q}'$。它可按下式计算

$$\bar{Q}' = \left|\int_{t_1}^{t}(Z - R\tau)\mathrm{d}\tau\right| = \left|\int_{\frac{Z}{R}}^{t}(Z - R\tau)\mathrm{d}\tau\right| = \frac{1}{2R}(Rt - Z)^2 \tag{7.1.36}$$

故缺货损失费用 T_2 为

$$T_2 = \frac{(Rt - Z)^2}{2R}h_1 \tag{7.1.37}$$

仅考虑固定进货费用时，

$$T_3 = C_1 \tag{7.1.38}$$

将式(7.1.35)、式(7.1.37)、式(7.1.38)代入式(7.1.32)，得一个存储周期的总费用 C 为

$$C = \frac{Z^2}{2R}h + C_1 + \frac{(Rt - Z)^2}{2R}h_1$$

单位时间费用 $T(t,Z)$ 为

$$T(t,Z) = \frac{Z^2}{2Rt}h + \frac{C_2}{t} + \frac{(Rt - Z)^2}{2Rt}h_1 \tag{7.1.39}$$

故，该系统的模型为

$$\min \quad T(t,Z) = \frac{Z^2}{2Rt}h + \frac{C_2}{t} + \frac{(Rt - Z)^2}{2Rt}h_1 \tag{7.1.40}$$

由偏微分法求极值得

$$t^* = \sqrt{\frac{2(h + h_1)C_1}{hh_1R}} \tag{7.1.41}$$

$$Z^* = \sqrt{\frac{2Rh_1C_1}{h(h + h_1)}} \tag{7.1.42}$$

$$t_1^* = \sqrt{\frac{2hC_1h_1R}{h + h_1}} \tag{7.1.43}$$

当 $h_1 \to \infty$ 时，$\frac{h_1}{h + h_1} = 1$，上式变为不允许缺货模型。

7.1.4.4 确定型允许缺货的存储系统模型(二)

对于允许缺货且生产周期较长的存储系统，依照上述方法即可建立其模型。

由图 7.1.9 可见，此时平均储量 $\bar{Q}$ 为：

$$\bar{Q} = \int_0^{t_1}(P - R)t\mathrm{d}t + \int_{t_1}^{t_2}(Q - Rt)\mathrm{d}t = \frac{PZ^2}{2R(P - R)}$$

平均缺货量 $\bar{Q}$ 为

$$\bar{Q}' = \left|\int_{t_2}^{t_3}(Q - Rt)\mathrm{d}t + \int_{t_3}^{t}[(P - R)t - Q]\mathrm{d}t\right| = \frac{1}{2P}(P - R)Rt^2 - Zt + \frac{PZ^2}{2R(P - R)}$$

其中：$t_1 = \frac{Z}{P - R}$；$t_2 = \frac{PZ}{R(P - R)}$；$t_3 = t - \frac{R(t - t_2)}{P}$

总费用 C 为

$$C = T_1 + T_2 + T_3 = \frac{PZ^2h}{2R(P - R)} + C_1 + \frac{h_1}{2P}(P - R)Rt^2 - Zth_1 + \frac{PZ^2h_1}{2R(P - R)}$$

单位时间的总费用为

$$T(t,Z)=\frac{C_1}{t}+\frac{PZ^2h}{2R(P-R)t}+\frac{(P-R)Rth_1}{2P}-Zh_1+\frac{PZ^2h_1}{2R(P-R)t}$$

使 $T(t,Z)\to\min$,则是该系统模型。

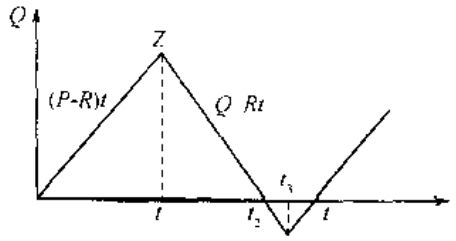


图 7.1.9　确定型允许缺货的存储系统模型(二)

用微分法求解后,得最佳存储周期和批量的计算公式如下:

$$t^*=\sqrt{\frac{2PC_1(h+h_1)}{R(P-R)hh_1}} \tag{7.1.44}$$

$$Z^*=\sqrt{\frac{2C_1Rh_1(P-R)}{Ph(h+h_1)}} \tag{7.1.45}$$

$$Q^*=Rt^*=\sqrt{\frac{2PC_1(h+h_1)}{(\frac{P}{R}-1)hh_1}} \tag{7.1.46}$$

当 $h_1\to\infty$、$P>>R$ 时,模型即化为不允许缺货且生产周期较短的模型。

7.1.5　资源分配问题

在系统内,任何一种资源在向子系统分配时,都要按择优分配的原则进行。但当评价函数呈非线性时,只有达到一定的分配比例,才能使系统达到最优。因此系统内分配资源且评价函数呈非线性的问题是一非线性规划问题。

如某电力公司有 n 个发电厂,该公司计划发电量为 m,如何根据各发电厂的生产成本、合理安排生产任务,才能使全公司发电生产成本最低。

众所周知,任一发电厂的生产成本都与其发电量呈非线性关系,假设对每个电厂进行研究后得出其发电成本 C 与其发电量之间的关系 $C_i=f_i(x)$,当 $f(x)$ 为非线性函数,则该电力公司生产任务的分配应以下式为目标:

$$\min\quad C=\sum_{i=1}^{n}C_i=\sum_{i}^{n}f_i(x) \tag{7.1.47}$$

假设现仅考虑必须完成上级限定的电力生产任务,则约束条件为

$$\sum_{i=1}^{n}x_i=m \tag{7.1.48}$$

这就建立了资源分配的非线性有约束最优化模型。

7.1.6　非线性曲线拟合问题

回归分析法是预测时经常使用的一种方法,而回归分析法一般均采用最小二乘法确定参数。最小二乘法的本质是以回归曲线上的点与实测点的误差平方和最小为目标的一

个非线性规划问题。在回归分析中的最小二乘法是无约束的非线性规划问题，它可应用微分求极值的方法求解，但有时在处理系统预测问题时，尚需考虑某些限制，这样就需建立一个有约束的非线性规划模型来解决曲线拟合问题。

如已知某企业的生产函数形式为

$$Y = A_1 + A_2^{-A_3 t} \tag{7.1.49}$$

现对该企业进行观测得一组资料(t_i, y_i) $(i = 1,2,\cdots,n)$，若已知模型中参数 A_1 和 A_2 的关系为

$$A_1 + A_2 = 1, \text{且 } A_3 > 0$$

则所得建立的模型为

$$\begin{aligned} \min \quad & f(A_1, A_2, A_3) = \sum_{i=1}^{n} (Y_i - A_1 - A_2^{-A_3 t_i})^2 \\ \text{s.t.} \quad & A_1 + A_2 = 1 \\ & A_3 > 0 \end{aligned} \tag{7.1.50}$$

按该模型所得的曲线即可符合实际需求。

总之，非线性规划模型的建立比较灵活，能尽可能地符合实际情况，应用范围较广。

线性规划可看作是非线性规划的一种特殊情况，即几何规划中 $P = 1$ 的情况。因此，线性规划也可认为是非线性规划的应用。

随着科学技术的发展，尤其是数学和计算技术的发展，非线性规划已成为系统规划设计和管理控制等工作中应用最广泛的一种最优化方法。

7.2 无约束非线性规划问题求解方法及原理

系统模型建立之后，求解则是关键。对于无约束的非线性规划，目标函数的极值就是问题的最优解；对于有等式约束的非线性规划问题，目标的极值需在约束的交点中寻找，同时要求变量数多于约束数，否则问题无解；对于有不等式约束的非线性规划，目标函数的极值需在约束构成的可行区内寻找。

非线性规划的求解要比线性规划的求解难得多，到目前为止，尚没有一种像求解线性规划单纯型法那样的适用于各种问题的一般算法。

无约束非线性规划模型的一般形式是：

$$\min \quad Z = f(X), X = (X_1, X_2, \cdots, X_n)^{\mathrm{T}} \text{且 } X \in E^n$$

式中 E^n——n 维欧氏空间。

该目标函数的最优点 $X = X^*$，$f(X)$ 的最优值为 $f(X^*)$。

如果模型的目标函数为

$$\max \quad Z = f(X)$$

则可通过乘以“-1”的方法，使其变为

$$\min \quad Z' = -Z = -f(X)$$

从而满足一般形式的要求。

7.2.1 解析法

当无约束非线性规划模型有显式解析表达式且连续可微时，应用函数求极值的基本原理，利用一阶导数和二阶导数，通过求驻点来确定最优解的方法叫解析法，也称间接法。

7.2.1.1 函数极小值的定义

定义 1 对于 $X^* \in \Omega$（Ω 为可行集），如果存在着一个 $\varepsilon > 0$，使得所有与 X^* 的距离小于 ε 的 X（即 $X \in \Omega$ 且 $|X - X^*| < \varepsilon$）都满足不等式 $f(X) \geqslant f(X^*)$，则称 X^* 为函数 $f(X)$ 在 Ω 上的一个相对极小点（或局部极小点）；都满足 $f(X) > f(X^*)$ 且 $X \neq X^*$，则 X^* 是函数 $f(X)$ 在 Ω 上的严格相对极小点。

定义 2 点 $X^* \in \Omega$，对于所有的 $X \in \Omega$ 使得 $f(X) \geqslant f(X^*)$，则 X^* 是 $f(X)$ 函数在 Ω 上的全局最小点，若 $f(X) > f(X^*)$ 且 $X \neq X^*$，则称 X^* 为 $f(X)$ 在 Ω 上的严格全局最小点。

定义的几何解释如图 7.2.1、图 7.2.2 所示。

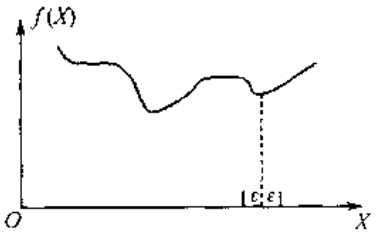


图 7.2.1 局部最小点

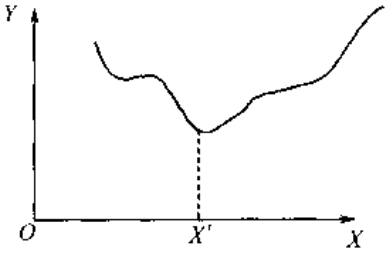


图 7.2.2 严格极小点

7.2.1.2 凸函数的概念

定义：在凸集 Ω 上的函数 $f(X)$，如果对每一对 X_1、X_2，且 X_1、$X_2 \in \Omega$，及每一个 α，$0 \leqslant \alpha \leqslant 1$，有 $f(\alpha X_1 + (1-\alpha)X_2) \leqslant \alpha f(X_1) + (1-\alpha)f(X_2)$ 成立，则函数 $f(X)$ 是凸的。如果 $f(\alpha X_1 + (1-\alpha)X_2) < \alpha f(X_1) + (1-\alpha)f(X_2)$，则函数 $f(X)$ 是严格凸的。

凸函数的几何意义如图 7.2.3 所示。

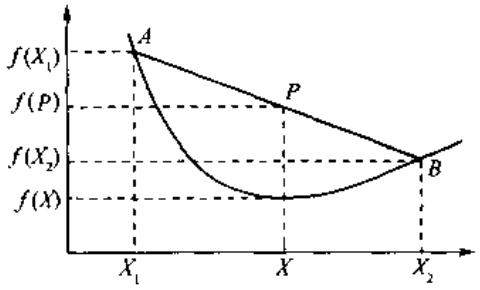


图 7.2.3 凸函数

由图 7.2.3 可知，在函数 $f(X)$ 上任取两点 A 和 B，连接点 A 和 B 得直线 AB。若 A、B 两点的坐标分别为 $A(X_1, f(X_1))$、$B(X_2, f(X_2))$，则直线 AB 上任意一点 $P(X, f(P))$ 可用以下方程描述：

$$X = \alpha X_2 + (1 - \alpha)X_1$$
$$f(P) = \alpha f(X_2) + (1 - \alpha)f(X_1) \qquad 0 \leqslant \alpha \leqslant 1$$

其中，α 为比例系数，当 $\alpha = 0$ 时，$X = X_1$，$f(X) = f(X_1)$，即 P 位于 A 点，当 $\alpha = 1$ 时，$X = X_2$，$f(X) = f(X_2)$，P 位于 B 点；当 $0 < \alpha < 1$ 时，P 点在线段 AB 上，$f(X_1) < f(P) < f(X_2)$。

因 $f(X) = f(\alpha X_2 + (1 - \alpha)X_1)$，故有：

$$f(X) \leqslant f(P) \text{ 或 } f(X) < f(P)$$

即 $$f(\alpha X_2 + (1 - \alpha)X_1) \leqslant \alpha f(X_2) + (1 - \alpha)f(X_1)$$

或 $$f(\alpha X_2 + (1 - \alpha)X_1) < \alpha f(X_2) + (1 - \alpha)f(X_1)$$

这说明，如果函数 $f(X)$ 为凸函数，则直线 AB 永远在曲线的上部或曲线上，否则不是凸函数。

7.2.1.3 求函数极值的条件

1. 必要条件(求驻点)

对一元函数 $f(X)$ 来说，$\dfrac{\mathrm{d}f(X)}{\mathrm{d}X} = 0$，

对二元函数 $f(X_1, X_2)$ 来说，$\dfrac{\partial f(X_1, X_2)}{\partial X_1} = 0$，$\dfrac{\partial f(X_1, X_2)}{\partial X_2} = 0$。

对多元函数 $f(X)$ 来说，$\dfrac{\partial f}{\partial X_1} = \dfrac{\partial f}{\partial X_2} = \cdots = \dfrac{\partial f}{\partial X_n} = 0$，

即 $$\nabla f = \left(\frac{\partial f}{\partial X_1}, \frac{\partial f}{\partial X_2}, \cdots, \frac{\partial f}{\partial X_n}\right)^{\mathrm{T}} = 0$$

2. 充分条件

若函数 $f(X)$ 为凸函数，则函数有相对极小值，若函数 $f(X)$ 为凹函数，则函数有相对极大值。

若函数 $f(X)$ 为严格凸函数，则函数在 $\nabla f(x) = 0$ 处有全局最小值；若函数 $f(X)$ 是严格凹函数，则函数在 $\nabla f(x) = 0$ 处有全局最大值。

7.2.1.4 判定函数凸性的条件

1. 用梯度判断函数凸性

若函数 $f(X)$ 为凸集 Ω 内的可微函数，则 $f(X)$ 为凸函数的充要条件为：对任意 X 有：

$$f(X + \Delta X) \geqslant f(X) + \nabla^{\mathrm{T}} f(X)\Delta X$$

其中，$X \in \Omega$，$X + \Delta X \in \Omega$

2. 由二阶偏导数矩阵判断函数凸性

若 $f(X)$ 是凸集内的二次可微函数，则 $f(X)$ 为凸函数的充要条件是二阶偏导数矩阵 $\nabla^2 f(X)$ 处处半正定，$f(X)$ 为 Ω 内严格凸函数的充要条件为二阶偏导数矩阵 $\nabla^2 f(X)$ 处处正定。

利用二阶偏导数矩阵判断函数的凸性，实质上就是判断 $\nabla^2 f(X)$ 的正定性。

二阶偏导数矩阵 $\nabla^2 f(X)$ 一般称为海赛矩阵，其形式为

$$\boldsymbol{H} = \nabla^2 f(X) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial X_1^2} & \dfrac{\partial^2 f}{\partial X_1 \partial X_2} & \cdots & \dfrac{\partial^2 f}{\partial X_1 \partial X_n} \\ \dfrac{\partial^2 f}{\partial X_2 \partial X_1} & \dfrac{\partial^2 f}{\partial X_2^2} & \cdots & \dfrac{\partial^2 f}{\partial X_3 \partial X_n} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial^2 f}{\partial X_n \partial X_1} & \dfrac{\partial^2 f}{\partial X_n \partial X_2} & \cdots & \dfrac{\partial^2 f}{\partial X_n^2} \end{bmatrix}$$

海赛矩阵是一实对称矩阵。矩阵 $\boldsymbol{H}$ 正定是指除 $\boldsymbol{X} = 0$ 以外，对所有 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X} > 0$，则 $\boldsymbol{H}$ 正定，若 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X} \geqslant 0$，则半正定，若 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X} < 0$，则 H 负定；若 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X} \leqslant 0$，则 $\boldsymbol{H}$ 半负定；若对某些 $\boldsymbol{X}$，有 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X} > 0$，对另外的一些 $\boldsymbol{X}$，有 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{X} < 0$，则 $\boldsymbol{H}$ 为不定。

其具体判定方法有二种：

(1) 根据实对称矩阵的特征值确定其正定性。

若 $n \times n$ 实对称矩阵 $\boldsymbol{H}$ 为正定，则特征值 $\lambda_i > 0$ $(i = 1,2,\cdots,n)$；

若 $n \times n$ 实对称矩阵 $\boldsymbol{H}$ 为半正定，则特征值 $\lambda_i \geqslant 0$ $(i = 1,2,\cdots,n)$；

若 $n \times n$ 实对称矩阵 $\boldsymbol{H}$ 为负定，则特征值 $\lambda_i < 0$ $(i = 1,2,\cdots,n)$；

若 $n \times n$ 实对称矩阵 $\boldsymbol{H}$ 为半负定，则特征值 $\lambda_i \leqslant 0$ $(i = 1,2,\cdots,n)$；

若 $n \times n$ 实对称矩阵 $\boldsymbol{H}$ 为不定，则特征值 λ_i 既有 $\lambda_i > 0$ 存在，又有 $\lambda_j < 0 (j \neq i)$ 存在。

特征值 λ_i 为 $|\lambda \boldsymbol{I} - \boldsymbol{H}| = 0$ 的解，$|\lambda \boldsymbol{I} - \boldsymbol{H}|$ 为特征多项式，是矩阵 $\lambda \boldsymbol{I} - \boldsymbol{H}$ 的行列式。

(2) 根据赛维斯特定理判定。

$n \times n$ 阶实对称矩阵正定的充要条件为方阵 $\boldsymbol{H}$ 左上角各阶主子行列式均大于零，即

$\det \quad H_i > 0$

若 $\det \quad H_i > 0$，且 $\det \quad \boldsymbol{H} = 0$，则 $\boldsymbol{H}$ 半正定；

若当 $i = 1,3,5,\cdots$ 时，$\det \quad H_i < 0$；

$i = 2,4,6,\cdots$ 时，$\det \quad H_i > 0$，则 $\boldsymbol{H}$ 负定。

当 $\det \quad \boldsymbol{H} = 0$，且 $\quad i = 1,3,5,\cdots,\det \quad H_i \leqslant 0$；

$i = 2,4,6,\cdots,\det \quad H_i \geqslant 0$，则 $\boldsymbol{H}$ 为半负定。

综上所述，无约束非线性规划问题的解析解法可归结为

(1) 求驻点：通过 $\nabla f(X) = 0$，求出驻点 X^*；

(2) 判断函数的凸性，以确定极值属性。

7.2.2 迭代法(直接法)

当目标函数较复杂或无显式表达式时，因无法用解析法求出驻点，采用建立迭代模型直接求得最优解的方法叫迭代法，亦称直接法。

迭代法分为两类，一类是区间消去法，如斐波那契法(Fibonacci)、0.618 法(黄金分割法)等；另一类是多维搜索法，如坐标轮换法、方向加速法、单纯形法等。

迭代法的基本思想是先给定一初始点 X_0，寻找一适当方向，并在该方向上向前迈进适当步长，使目标下降，达到新的一点，再将新点当作 X_0，重复上述步骤，直至达到最优解

为止，其迭代公式的一般形式为

$$\boldsymbol{X}^{(k+1)} = \boldsymbol{X}^{(k)} + \lambda^{(k)}\boldsymbol{P}^{(k)} \tag{7.2.1}$$

式中 $\boldsymbol{P}^{(k)}$——第 k 次迭代的方向向量；

$\lambda^{(k)}$——第 k 次迭代的步长。

7.2.2.1 成功失败法

成功失败法的基本思想是在$[a,b]$内求单峰函数 $f(X)$ 的极小值时，从预先给定的初始点 X_0 出发，按预先确定的步长 λ，利用目标函数 $f(X)$ 给出的信息，在区间$[a,b]$内搜索，若搜索一次，目标值下降，则搜索成功，这时以成功点为 X_0，以大的步伐（通常以 $\lambda' = 2\lambda$）前进。若目标函数不下降，则搜索失败，此时改为在 X_0 点以小步后退（通常选 $\lambda' = -\frac{1}{4}\lambda$）。这样反复搜索，直至得到满足一定精度的最优解为止。

其成功失败法计算框图如图 7.2.4 所示。

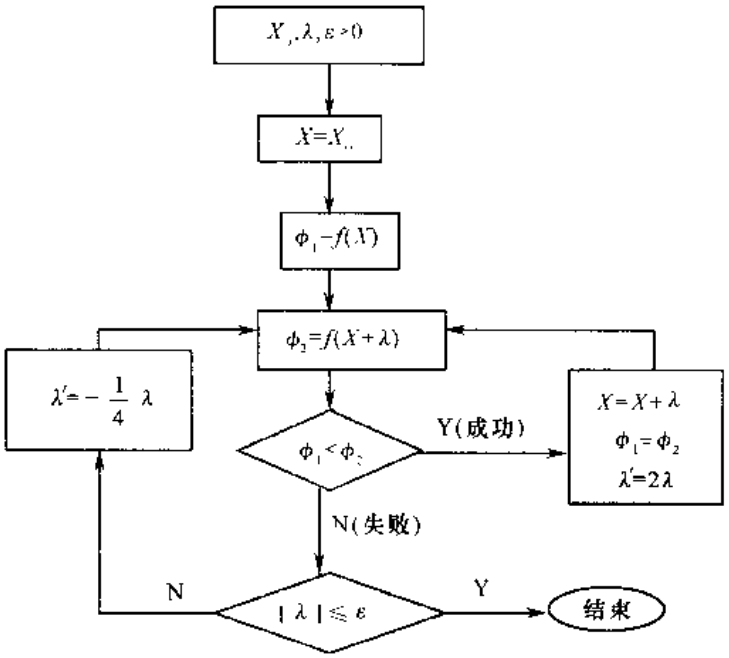


图 7.2.4 成功失败法计算框图

7.2.2.2 斐波那契法

设函数 $f(X)$ 在区间$[a,b]$内为下单峰函数，且有惟一的相对极小点 X^*，若在区间内任取 a_1、b_1 两点，计算 $f(a_1)$ 和 $f(b_1)$，则① 当 $f(a_1) > f(b_1)$ 时，X^* 在$[a_1,b]$内，如图 7.2.5 所示；② 当 $f(a_1) < f(b_1)$ 时，X^* 在$[a,b_1]$内，如图 7.2.6 所示。

如出现第一种情况，下一步可在$[a_1,b]$中寻找 X^*，即去掉区间$[a,a_1]$，把区间$[a,b]$缩小为$[a_1,b]$；如出现第二种情况，下一步则在$[a,b_1]$中寻找 X^*，即去掉区间$[b_1,b]$，把区间$[a,b]$缩小为$[a,b_1]$，如点 a_1、b_1 选择恰当，则可把区间$[a,b]$缩到最小限度。如将 a_1、b_1 选在中点附近，区间$[a,b]$可缩小到原长度的一半左右（第三种情况），如图 7.2.7 所示。在区间$[a_1,b]$或$[a,b_1]$中再选一点 b_2 或 a_2，则又可将该区间缩短，如此下去，就可将区间一段一段地缩短，最后找到最优解。这就是斐波那契法的基本思想。

为了在一定的计算次数下，得到满足一定精度要求的最优解，现分析选点个数（计算

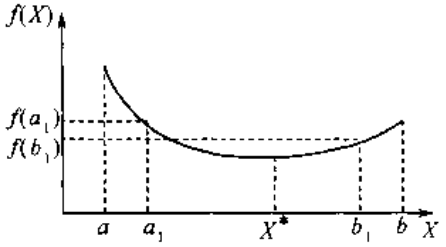


图 7.2.5　第一种情况

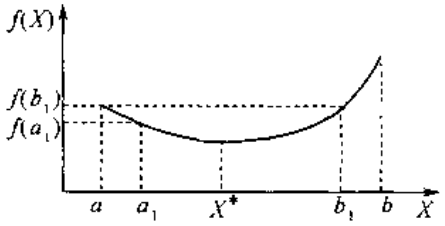


图 7.2.6　第二种情况

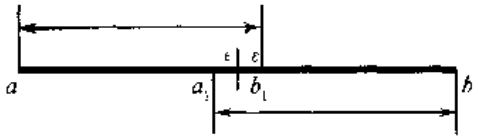


图 7.2.7　第三种情况

函数值次数）与区间压缩程度之间的关系。

设选点个数为 n，区间长度为 L，则：

当 $n = 0$，即不选取点时，区间长度不变，为原区间长度 $L/1$；

当 $n = 1$，即只选一点时，区间长度不变，仍为原区间长度 $L/1$；

当 $n = 2$，即选两点时，区间长度最大可压缩为原长度的一半，即 $L/2$；

当 $n = 3$，即选三点时，区间长度最大可压缩为原长度的 $1/3$，即 $L/3$；

当 $n = 4$，即选四点时，区间长度最大可压缩为原长度的 $1/5$，即 $L/5$。

如此可一直推到 n 较大的情况。现把选点个数 n 和表示区长度压缩倍数的分母值联系起来，可得到如下数列：

$$F_0 = 1, F_1 = 1, F_2 = 2, F_3 = 3, F_4 = 5, \cdots, \text{即} \{F_k\} (k = 0, 1, 2, \cdots, n),$$

该数列称斐波那契数列，它体现了选点个数与区间压缩程度之间的关系，其通式为

$$\begin{cases} F_{k+2} = F_{k+1} + F_k \\ F_0 = 1, F_1 = 1 \end{cases} \tag{7.2.2}$$

按式(7.2.2)可得出在选点数确定的情况下，区间被压缩的程度。如 $F_{20} = 10946$，则说明当计算 20 次函数值（选 20 个点）时，就可将原区间长度 L 压缩为 $\frac{L}{10946}$（即 $0.00009L$）。

这样就可将求目标函数 $f(X)$ 在区间 $[a, b]$ 内满足相对精度 δ 的极小值问题化为计算若干次目标函数值的问题，也就是可按下式确定选取点个数：

$$F_n > \frac{1}{\delta} \tag{7.2.3}$$

如区间 $[a, b]$ 长度为 L，$\delta = 1\%$，则 $F_{11} = 144 > 100$。这就是说，适当选择 11 个点来计算目标函数值，就可将原区间 L 缩短为 $\frac{1}{144}L$，从而满足精度要求。

如果给定绝对精度 η，可按 $\delta = \frac{\eta}{L}$ 化为相对精度来确定斐波那契数 F_n。

如区间长度 $L=10$,要求满足绝对精度 $\eta=0.01$ 的极小点,则相对精度 δ 和斐波那契数 F_n 为

$$\delta=\frac{0.01}{10}=0.001,\qquad F_{16}=1597>1000$$

即选 16 点可得到满足精度要求的最优解。

现定义两个相邻斐波那契数之比 F_{n-1}/F_n 为收缩率,它反映了第 n 次计算目标函数值时区间收缩的比率。其实质是:设区间长度为 L,选 n 个计算点可将区间 L 压缩为 L/F_n,选择 $n-1$ 个计算点可将区间长度压缩为 L/F_{n-1}。在连续选择计算点时,是在计算第 $n-1$ 个点的基础上计算第 n 个点的,$\dfrac{L/F_n}{L/F_{n-1}}=\dfrac{F_{n-1}}{F_n}$ 则表示在第 $n-1$ 次计算的基础上又缩短的比率,也可以说是第 n 次计算后所剩余的区间长度在 $n-1$ 次计算后剩余长度中所占的比例。

上述收缩率的涵义可由斐波那契数列经推导得出:

$$\because\quad F_n=F_{n-1}+F_{n-2}$$

两端同乘区间长度得 $\qquad F_nL=F_{n-1}L+F_{n-2}L$

同除 F_n 得 $\qquad L=\dfrac{F_{n-1}}{F_n}L+\dfrac{F_{n-2}}{F_n}L$

由此可见,区间长度由两部分组成,每次去掉的部分为区间长度的 F_{n-2}/F_n 倍,剩余部分为区间长度 L 的 F_{n-1}/F_n 倍。因此收缩率的意义是原区间长度的收缩比率。

根据收缩率的涵义即可得到连续选点位置的迭代公式

$$\lambda_k=b_{k-1}+\frac{F_{n-k}}{F_{n-k+1}}(a_{k-1}-b_{k-1})=a_{k-1}+\frac{F_{n-k-1}}{F_{n-k+1}}(b_{n-1}-a_{n-1})$$

$$\lambda'_k=a_{k-1}+\frac{F_{n-k}}{F_{n-k+1}}(b_{k-1}-a_{k-1})=b_{k-1}+\frac{F_{n-k-1}}{F_{n-k+1}}(a_{n-1}-b_{n-1})\qquad(7.2.4)$$

当 $k=1$ 时

$$\lambda_1=b_0+\frac{F_{n-1}}{F_n}(a_0-b_0)=a_0+\frac{F_{n-2}}{F_n}(b_0-a_0)$$

$$\lambda'_1=a_0+\frac{F_{n-1}}{F_n}(b_0-a_0)=b_0+\frac{F_{n-2}}{F_n}(a_0-b_0)$$

当 $k=n-1$ 时,$\lambda_{n-1}=\lambda'_{n-1}=\dfrac{1}{2}(a_{n-2}+b_{n-2})$

因 $f(\lambda_{n-1})=f(\lambda'_{n-1})$,无法比较大小,故令:

$$\lambda'_{n-1}=a_{n-2}+(\frac{1}{2}+\varepsilon)(b_{n-2}-a_{n-2})$$

选 $\qquad \min\ \{\lambda_{n-1},\lambda^*_{n-1}\}=X^*$

该方法的计算框图如图 7.2.8 所示。

7.2.2.3 0.618 法

斐波那契法的优点在于可根据预先确定的精度,确定计算的次数。缺点是每次计算都需确定收缩率,且收缩率是不等的。由斐波那契数计算的收缩率序列 $\left[\dfrac{F_{n-1}}{F_n}\right]$ 可知,当

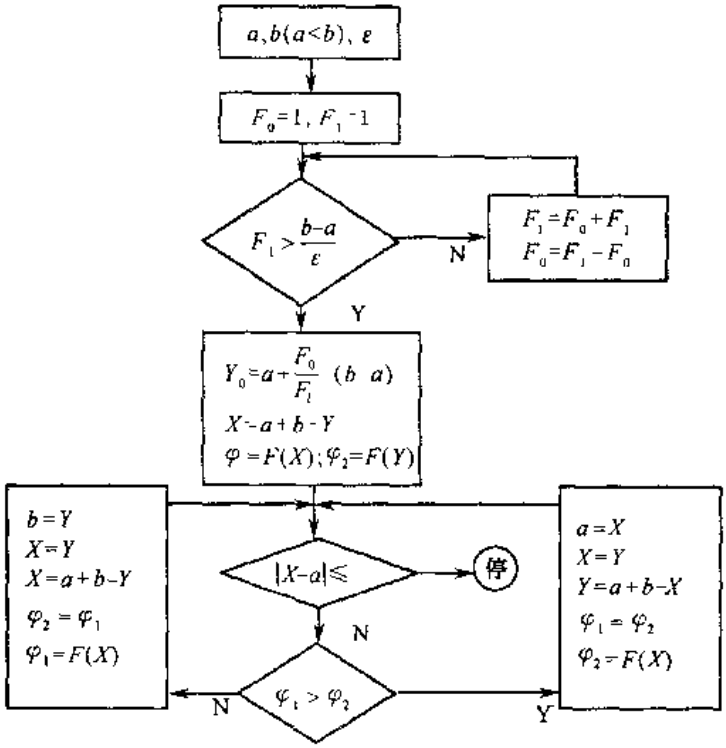


图 7.2.8　斐波那契法计算框图

$n \to \infty$ 时，$\left[\frac{F_{n-1}}{F_n}\right] \to 0.618$，这样可使收缩率由变数变为固定数，使计算简便。

由斐波那契数之间的关系也可推出该结论

$$F_{n+1} = F_n + F_{n-1}$$

同除 F_n 后得

$$\frac{F_{n+1}}{F_n} = 1 + \frac{F_{n+1}}{F_n}$$

如果当 $n \to \infty$ 时，F_{n-1}/F_n 的极限存在，且为 λ，则

$$\frac{1}{\lambda} = 1 + \lambda$$

整理得

$$\lambda^2 + \lambda - 1 = 0$$

解方程得

$$\lambda = 0.61802398 \approx 0.618$$

同理可证明

$$\lim_{n\to\infty} \frac{F_{n-1}}{F_{n+1}} = \lim_{n\to\infty} \frac{F_{n-1}}{F_n} \cdot \frac{F_n}{F_{n+1}} \approx 0.382$$

由此可见 0.618 法是斐波那契法在收缩率为常量时的特例，当给定相对精度 δ 时，则 $\lambda^{n-1} \leqslant \delta$ 可计算出计算次数，如 $\delta = 0.01$，则 $n \geqslant 10.5$，计算 11 次即可求出满足精度要求的最优值。

其计算过程如图 7.2.9 所示。

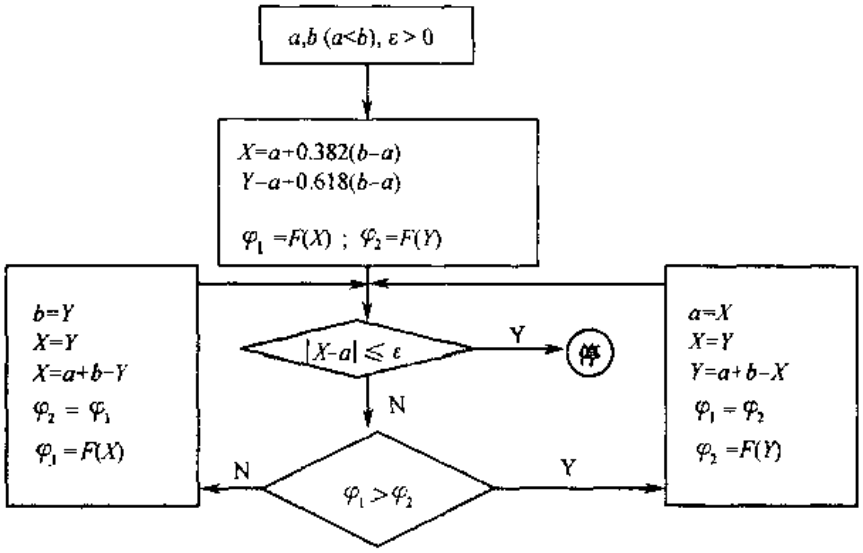


图 7.2.9　0.618 法计算框图

7.2.2.4　坐标轮换法

上述三种迭代方法对于一元函数是特别有效的，因此也称一维搜索或线性搜索。对多元函数，应用一维搜索原理也可解决，其中使用较多的是坐标轮换法。

坐标轮换法是以一维搜索为基础，轮流按坐标轴方向搜索最优解的一种方法。其基本思想是：对于任一初始点 $X^{(0)}$，先在其第一个坐标轴方向 e^1 上采用一维搜索的方法，求得该方向的最优解，再以该最优解为初始点在另一坐标轴方向上重复上述步骤，当所有坐标轮换一次之后得到 $X^{(1)}$ 点，再从 e^1 开始继续轮换，直到得到最优解为止。图 7.2.10 是一二维情况的坐标轮换过程，其计算框图如图 7.2.11 所示。

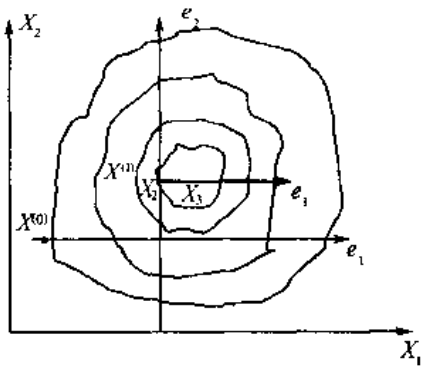


图 7.2.10　坐标轮换

7.2.3　解析 — 迭代法

解析 — 迭代法亦称间接 — 直接法，是以解析法为基础的一种迭代法，是上述二种方法的结合。它吸收了迭代法计算方便，便于利用电子计算机和解析法能准确地确定最优解的优点，利用迭代法对复杂的解析模型（对模型导数方程的求解等）进行迭代，从而确定最优解。该法只能应用于目标函数以显式表示且可导的模型。

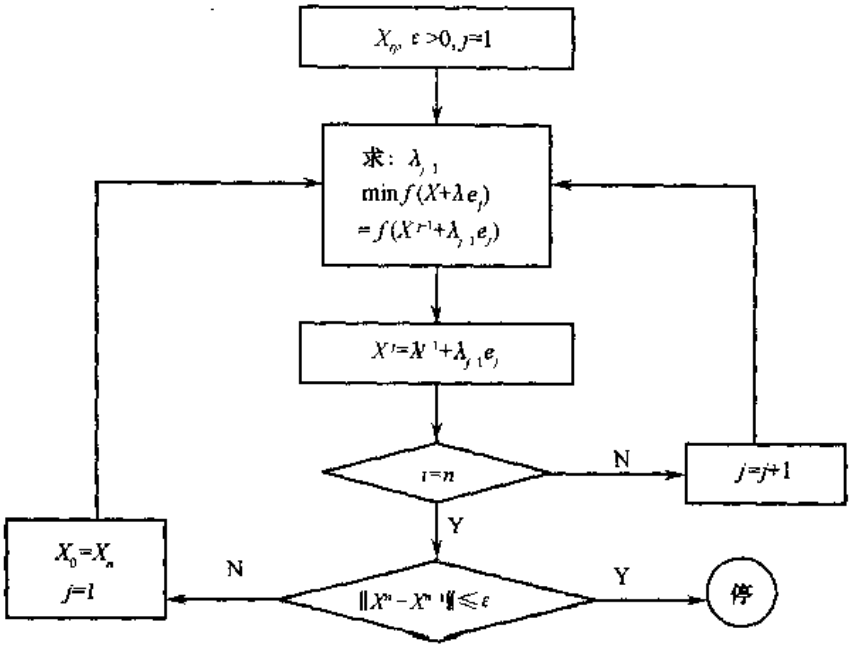


图 7.2.11　坐标轮换法计算框图

7.2.3.1　牛顿法

1. 古典牛顿法

现研究牛顿法求方程近似解的迭代过程。由图7.2.12可见，函数 $f(x)$ 与 X 轴相交于 X^*，X^* 为方程 $f(x)=0$ 的根。

如果给定一点 X_0，则 $f(X_0)$ 处的导数为 BC，交 X 轴于 X_1（C 点），如再求 $f(X_1)$（D 点）处的导数，则交 X 轴于 X_2（E 点），$X_0, X_1, \cdots$ 所组成的数列 $\{X_n\}$ 将逐渐趋近于 X^*，其迭代公式可由 $\triangle CDE$ 的关系得出。

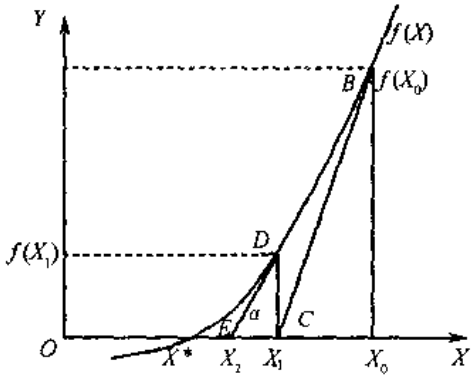


图 7.2.12　牛顿法

设 C 点的 X 坐标为 X_K，则 D 点的 Y 坐标为 $f(X_K)$，E 点的 X 坐标为 X_{K+1}。

$$\because \tan\alpha = \frac{CD}{CE} = f'(X_K) = \frac{f(X_K)}{X_K - X_{K+1}}$$

$$\therefore X_{k+1} = X_K - \frac{f(X_K)}{f'(X_K)} \tag{7.2.5}$$

对以 min $f(X)$ 为目标函数的非线性规划来说，为求其最小值，必先求其驻点，即$\nabla f = 0$，如∇f较复杂，直接求解较困难，把∇f当作$f(X)$即可应用上述迭代公式求出驻点。但因∇f 函数本身是$f(X)$的一阶导数，故式(7.2.5)应变成如下形式：

$$X_{K+1} = X_K - H^{-1}(X_K)\nabla f(X_K) \tag{7.2.6}$$

当给定初始值 X_0 时，代入上式即可求出 X_1，将 X_1 代入上式，即可求出 X_2，如此继续下去，即得一数列$\{X_n\}$，如果它是收敛的，则必收敛于驻点。

因该方法的基本出发点是求驻点，属解析法，但求驻点的过程却用迭代法，因此称为解析—迭代法。其计算框图如图 7.2.13 所示。

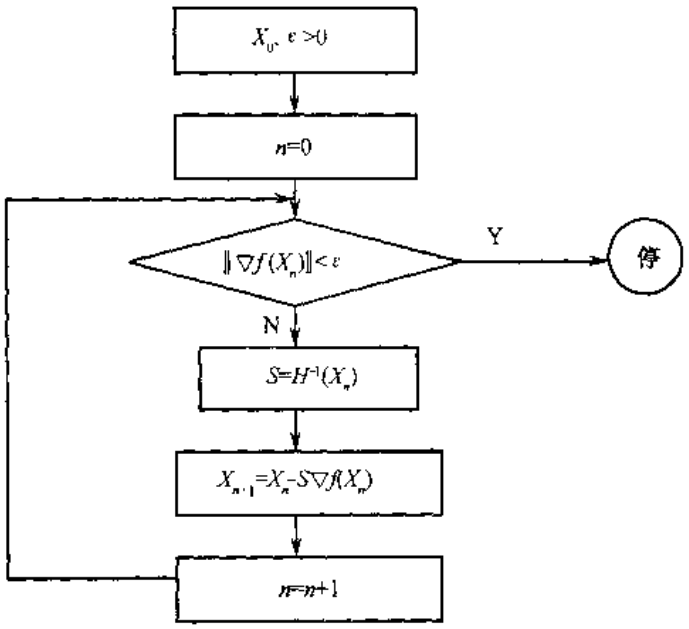


图 7.2.13　古典牛顿法计算框图

2. 广义牛顿法

广义牛顿法是在古典牛顿法的基础上，应用一维搜索确定最优步长，从而加快搜索速度的一种方法。

古典牛顿法的迭代公式中，$-H^{-1}(X_K)\nabla f(X_K)$ 仅是步长为"1"时在 X_K 点的一个搜索方向。这可由函数在 X_K 点沿 Z 方向以单位步长到达 $X_K + \lambda Z$ 时，目标函数值 $f(X_K + \lambda Z)$ 的泰劳展开式证明。

$$\because \quad f(X_K + \lambda Z) \approx f(X_K) + f'(X_K)Z + \frac{1}{2}f''(X_K)Z^2$$

最优方向 Z，可通过 $f'(X_K + \lambda Z)$ 求得，即

$$f'(X_K + \lambda Z) = f'(X_K) + \lambda f''(X_K)Z = 0$$

$$\lambda Z = -H^{-1}(X_K)\nabla f(X_K)$$

$\because \lambda = 1$

$\therefore Z = -H^{-1}(X_K)\nabla f(X_K)$

是 λ 为单位步长时的最佳搜索方向。

为了使目标在 X_K 处一步到达 Z 方向的最优点，需要确定最优步长 λ_K^*，其确定方法就是采用一维搜索：

$$\min_{\lambda \geqslant 0} f(X_K + \lambda Z_K) = f(X_K + \lambda_K^* Z_K)$$

因此广义牛顿法的迭代公式是：

$$X_{n+1} = X_n + \lambda_K^* H^{-1}(X_K) \nabla f(X_K) \tag{7.2.7}$$

其计算框图如图 7.2.14 所示。

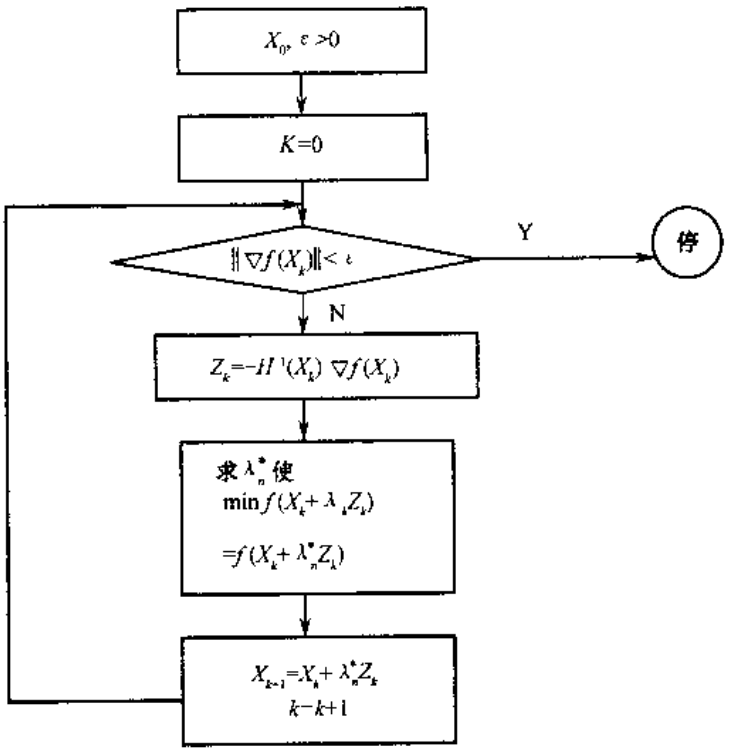


图 7.2.14　广义牛顿法计算框图

7.2.3.2　最速下降法

非线性目标函数的梯度为

$$\nabla f(X) = \frac{\partial f}{\partial X} = \left[\frac{\partial f}{\partial X_1}, \frac{\partial f}{\partial X_2}, \cdots, \frac{\partial f}{\partial X_n}\right]^{\mathrm{T}}$$

由高等数学知，梯度与函数的等值线是垂直的，即梯度方向是函数的法线方向。因此沿梯度方向，$f(X)$ 的值增长最快，反之沿负梯方向 $f(X)$ 的值下降最快。这可通过 $f(X_n) - f(X_n - \lambda Z) = -\lambda f'(X_n)Z - 0(\lambda)$，略去高阶无穷小后，当 Z 与 X_n 处的负梯度方向 $-f(X_n)$ 一致时，使函数值下降最多而证实。

因此最速下降法的基本思想是：给定初始点 X_0，按使目标下降最快的方向 —— 负梯度方向迈出最优步长，即一步达到该方向的最小点，再以此点为初始点，重复上述步骤，直至得到最优解为止。其计算框图如图 7.2.15 所示。

7.2.3.3 共轭梯度法

共轭梯度法实质上是对梯度法的修正，使梯度法的搜索方向改为共轭方向。

所谓共轭方向是指正定矩阵 A，如果满足 $P_1^{\mathrm{T}}AP_0 = 0$，则 P_1 和 P_0 关于 A 共轭。如果 $A = I$，则 P_1 和 P_0 正交，可见共轭是正交的推广。

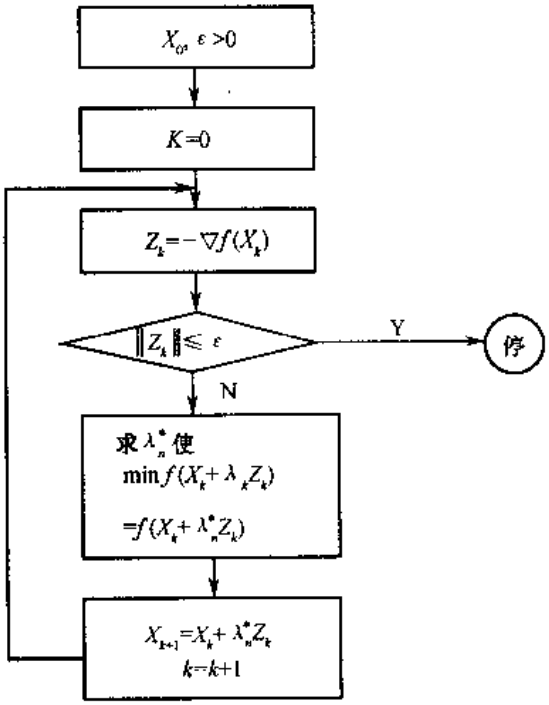


图 7.2.15 最速下降法计算框图

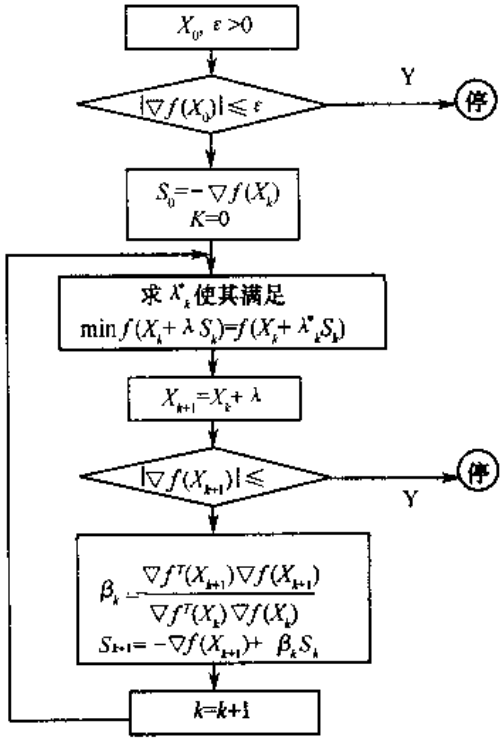


图 7.2.16 共轭方向法计算框图

所研究的目标函数总可以通过泰劳展开而用二次函数来近似，因此 A 是已知的。如果把初始点 X_0 的负梯度方向作为 P_0，则总可找到其共轭方向 P_1。

以共轭方向 P_1 作为搜索方向，并在此搜索方向上选择最优步长，即可到达该方向的最小点，再以此点为初始点，以此方向为初始方向，确定新的共轭方向和最优步长，又可找到这个方向的最小点，如此即可在有限步内到达最优点，这就是共轭梯度法的基本思想。

共轭梯度法中的搜索方向是一个与前一步搜索方向有关的量，其迭代步骤如图 7.2.16 所示。

无约束非线性规划的求方法很多，本节介绍的仅是其中比较常用的和基本的几种。不同的非线性规划问题，采用不同的求解方法，可收到不同的效果，有的方法计算时间较短，有的方法计算时间较长，因此如用某种方法计算费时较多，则应考虑更换其他方法。

7.3 有约束非线性规划的求解

7.3.1 有等式约束的非线性规划问题

这类非线性规划模型的一般形式是：

$$\left.\begin{aligned}\min \quad & f(X_1,X_2,\cdots,X_n)\\ \text{s.t.} \quad & g_1(X_1,X_2,\cdots,X_n)=0\\ & g_2(X_1,X_2,\cdots,X_n)=0\\ & \cdots\\ & g_m(X_1,X_2,\cdots,X_n)=0\end{aligned}\right\} \tag{7.3.1}$$

用向量表示为

$$\begin{aligned}\min \quad & F(X)\\ \text{s.t.} \quad & G(X)=0\end{aligned} \tag{7.3.2}$$

为了解决这个问题,数学已为我们研究出一套成型的方法——拉格朗日乘子法。

对于有等式约束的非线性规划模型引入拉格朗日乘子构成了拉格朗日函数,可把有约束问题变成无约束问题。有等式约束非线性规划的拉格朗日函数如下:

$$L(X,\lambda)=f(X)+\sum_{i=1}^{m}\lambda_i g_i(X) \tag{7.3.3}$$

原模型共有 n 个变量,引入拉格朗日乘子后,连同拉格朗日乘子共有 $n+m$ 个变量。因此建立拉格朗日函数把有约束问题化为无约束问题实质上是用比原问题高 m 维的代价换来去掉 m 个约束的效果。

求解拉格朗日函数,即得到原问题的最优解。因拉格朗日函数是无约束问题,故可采用上节所述的各种方法,现仅介绍解析法。

解析法(间接法)求解有等式约束非线性规划的步骤如下:

(1) 建立等式约束的非线性规划模型;

(2) 引入拉格朗日乘子,构成以 X_i、λ_i 为变量的拉格朗日函数;

$$L(X,\lambda)=f(X)+\sum_{i-1}\lambda_i g_i(X)$$

(3) 按求极值的必要条件求偏导数,得到以 X_i、λ_i 为未知量的 $m+n$ 个方程:

$$\frac{\partial L}{\partial X_j}=\frac{\partial f}{\partial X_j}+\sum_i\lambda_i\frac{\partial g_i(X)}{\partial X_i}=0 \quad (j=1,2,\cdots,n)$$

$$\frac{\partial L}{\partial \lambda_j}=g_i(X)=0 \qquad (i=1,2,\cdots,m)$$

(4) 判断海赛矩阵的正定性或根据实际问题判断极值性质;

(5) 求解第三步得到的方程组,得出 X_1^*、X_2^*、$\cdots$、X_n^*;λ_1^*、λ_2^*、$\cdots$、λ_m^* 并代入 $f(X)$,得目标的极小值。

现以资源分配问题为例说明有等式约束非线性规划问题的解析解法。

某电力公司有三个发电厂,设发电量分别为 X_1、X_2、X_3,发电成本与发电量的关系分别为 $\frac{1}{2}X_1^2$、X_2^2、$\frac{3}{2}X_3^2$,如限定总发电量为 11 万 kW,如何安排生产,才能使全公司总成本最低。

(1) 建立模型:

$$\min \quad f(X) = \frac{1}{2}X_1^2 + X_2^2 + \frac{3}{2}X_3^2$$
$$\text{s.t.} \quad X_1 + X_2 + X_3 = 11$$

(2) 引入拉格朗日乘子,建立拉格朗日函数:

$$L(X,\lambda) = \frac{1}{2}X_1^2 + X_2^2 + \frac{3}{2}X_3^2 + \lambda(X_1 + X_2 + X_3 - 11)$$

(3) 求偏导数:

$$\frac{\partial L}{\partial X_1} = X_1 + \lambda = 0 \qquad \frac{\partial L}{\partial X_2} = 2X_2 + \lambda = 0$$
$$\frac{\partial L}{\partial X_3} = 2X_3 + \lambda = 0 \qquad \frac{\partial L}{\partial \lambda} = X_1 + X_2 + X_3 - 11 = 0$$

(4) 该问题必存在极小值,故判断海赛矩阵正定性从略;

(5) 求解上述四个方程组成的方程组,其求解结果为

$$X_1 = -\lambda \qquad X_2 = -\frac{\lambda}{2}$$
$$X_3 = -\frac{\lambda}{3} \qquad \lambda = -6$$
$$\therefore \quad X_1^* = 6, X_2^* = 3, X_3^* = 2;$$
$$f(X^*) = \frac{1}{2} \times 6^2 + 3^2 + \frac{3}{2} \times 2^2 = 33$$

由该例子可见,虽然第一电厂生产成本最低,但全部由它生产并不合算,只有采用模型求解所确定的最优解,才是最经济的方案。

7.3.2 有不等式约束的非线性规划问题

有不等式约束非线性规划的一般形式是:

$$\begin{aligned} &\min \quad f(X) \\ &\text{s.t.} \quad h_j(X) \leqslant 0 \ (j = 1,2,\cdots,h) \\ &\qquad g_i(X) - 0 \quad (i = 1,2,\cdots,m) \end{aligned} \tag{7.3.4}$$

7.3.2.1 拉格朗日乘子法

拉格朗日乘子法是通过引入松弛变量把不等式约束变成等式约束,再引入拉格朗日乘子将其变为无约束问题的拉格朗日函数,最后应用求解无约束问题的各种方法求解。但由于引入松弛变量,使求解变得复杂,需对拉格朗日乘子和松弛变量加以讨论。

为了把不等式 $h_j(X) \leqslant 0$ 变成等式,引入松弛变量 Q_j,因非线性规划中无变量的非负条件,故采用 Q_j^2 作为松弛变量,得

$$h_j(X) + \theta_j^2 = 0$$

引入松弛变量后的非线性规划模型是:

$$\begin{aligned} &\min \quad f(X) \\ &\text{s.t.} \quad h_j(X) + \theta_j^2 = 0 \ (j = 1,2,\cdots,h+m) \end{aligned} \tag{7.3.5}$$

引入拉格朗日乘子后得拉格朗日函数为

$$L(X,\lambda,\theta) = f(X) + \sum_j \lambda_j (h_j(X) + \theta_j^2) \tag{7.3.6}$$

按求函数极值的必要条件：

$$\frac{\partial L}{\partial X_i} = \frac{\partial f}{\partial X_i} + \sum_j \lambda_j \cdot \frac{\partial h_j(X)}{\partial X_j} = 0 \tag{7.3.6a}$$

$$\frac{\partial L}{\partial \lambda_j} = h_j(X) + \theta_j^2 = 0 \tag{7.3.6b}$$

$$\frac{\partial L}{\partial \theta_j} = 2\lambda_j \theta_j = 0 \tag{7.3.6c}$$

由式(7.3.6c)可见，λ_j^*、θ_j^* 有以下三种情况：

(1) $\lambda_j^* = 0 \quad \theta_j^* \neq 0$

(2) $\lambda_j^* \neq 0 \quad \theta_j^* \neq 0$

(3) $\lambda_j^* = 0 \quad \theta_j^* \neq 0$

现分别对该三种情况加以讨论：

(1) 当 $\lambda_j^* = 0$、$\theta_j^* \neq 0$ 时。由式(7.3.6a)可知，$\frac{\partial L}{\partial X_i} = \frac{\partial f}{\partial X_i} = 0$，即 $\nabla f = 0$，这说明所有约束不起作用；这正是无约束极值的必要条件；

(2) 当 $\lambda_j^* \neq 0$、$\theta_j^* = 0$ 时，由式(7.3.6b)可见，$h_j(X) = 0$，这就说明最优解只能在约束的边界上，又因 $\lambda_j \neq 0$，所以这是所有约束均为等式时的情况，最优解不满足 $\nabla f(X^*) = 0$。

(3) 当 $\lambda_j^* = 0$、$\theta_j^* = 0$ 时，由式(7.3.6a)可见，$\frac{\partial f(X)}{\partial X_i} = 0$，由式(7.3.6b)可见 $h_j(X) = 0$，这说明极值点既在约束边界上，又满足 $\nabla f(X^*) = 0$，即约束边界恰好经过无约束最优点。

现举例说明这三种情况：

设有约束非线性规划模型为

$$\begin{aligned} &\min \quad f(X) = (X-a)^2 + b \\ &\text{s.t.} \quad X \leqslant c \end{aligned}$$

引入松弛变量 Q 后的拉格朗日函数为

$$L(X,\lambda,\theta) = (X-a)^2 + b + \lambda(X - c + \theta^2)$$

求驻点得：

① $$\frac{\partial L}{\partial X} = 2(X-a) + \lambda = 0$$

② $$\frac{\partial L}{\partial \lambda} = X - c + \theta^2 = 0$$

③ $$\frac{\partial L}{\partial \theta} = 2\lambda\theta = 0$$

讨论：(1) 当 $\lambda_j^* = 0$、$\theta_j^* \neq 0$ 时，$X^* = a$(见图 7.3.1)。

(2) 当 $\lambda_j^* \neq 0$、$\theta_j^* = 0$ 时，$X^* = c$(见图 7.3.2)。

(3) 当 $\lambda_j^* = 0$、$\theta_j^* = 0$ 时，$X^* = a = c$(见图 7.3.3)。

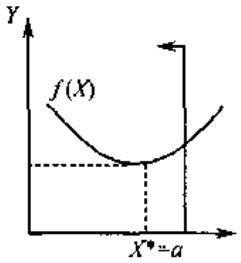


图 7.3.1　讨论(1)

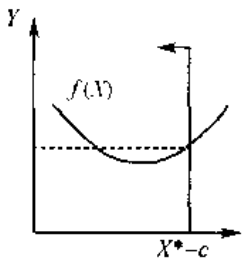


图 7.3.2　讨论(2)

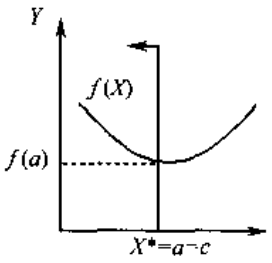


图 7.3.3　讨论(3)

7.3.2.2　库恩—图克定理

上述方法通过引入松弛变量把不等式约束化为等式约束，从而建立拉格朗日函数求解模型。但当不等式约束很多时，由于引入松弛变量很多，使求解的分析变得很困难，因此下面介绍求解有不等式约束非线性规划问题的完整理论——库恩—图克定理。

定理 1　对非线性规划问题

$$\begin{aligned} &\min \quad f(X) \\ &\text{s.t.} \quad g_i(X) \leqslant 0 \quad (i = 1,2,\cdots,m) \end{aligned} \tag{7.3.7}$$

设 $f(X)$ 及 $g_i(X)$ 均为可微函数，定义：

$$L(X,\lambda) = f(X) + \sum_i \lambda_i g_i(X)$$

假设 λ_i 存在，则最优点必满足以下条件：

$$\left.\begin{aligned} &(1)\ \frac{\partial f(X^*)}{\partial X_j} + \sum \lambda_i^* \frac{\partial g_i(X^*)}{\partial X_j} = 0 \quad (j = 1,2,\cdots,n) \\ &\quad \text{或} \nabla_X L(X^*,\lambda^*) = 0 \\ &(2)\ g_i(X^*) \leqslant 0 \quad (i = 1,2,\cdots,m) \\ &\quad \text{或} \nabla_\lambda L(X^*,\lambda^*) \leqslant 0 \\ &(3)\ \lambda_i^* g_i(X^*) = 0 \quad (i = 1,2,\cdots,m) \\ &\quad \text{或} (\lambda^*)^{\mathrm{T}} G(X^*) = 0 \\ &(4)\ \lambda^* \geqslant 0 \end{aligned}\right\} \tag{7.3.8}$$

其中，$\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m)^{\mathrm{T}}$；

$G(X) = (g_1(X), g_2(X), \cdots, g_m(X))^{\mathrm{T}}$；

λ_1 为库恩—图克乘子；

(1) ~ (4) 称库恩—图克条件。

对于非线性规划的一些特殊情况，其库恩—图克条件如定理 2 和定理 3。

定理 2　对于下述非线性规划：

$$\begin{aligned} &\min \quad f(X) \\ &\text{s.t.} \quad X_i \geqslant 0 \quad (i = 1,2,\cdots,n) \end{aligned} \tag{7.3.9}$$

X^* 为最优解的必要条件是：

$$\left.\begin{array}{ll}(1) & f(X^*)\geqslant 0\\ (2) & (X^*)^T f(X)=0\\ (3) & X^*\geqslant 0\end{array}\right\} \tag{7.3.10}$$

定理 3 对于下述非线性规划

$$\begin{aligned}&\min \quad f(X)\\ &\text{s.t.} \quad g_i(X)\leqslant 0 \quad (i=1,2,\cdots,m)\\ &\qquad X_j\geqslant 0 \quad (j=1,2,\cdots,n)\end{aligned} \tag{7.3.11}$$

X^* 为最优解的必要条件为：

$$\begin{cases}(1)\ \dfrac{\partial f(X^*)}{\partial X_j}+\sum\limits_{i=1}^{m}\lambda_i^*\dfrac{\partial g_i(X^*)}{\partial X_j}\geqslant 0\\ (2)\ g_i(X^*)\leqslant 0\\ (3)\ X_j^*\left[\dfrac{\partial f(X^*)}{\partial X_j}+\sum\limits_{i}\lambda_i^*\dfrac{\partial g_i(X^*)}{\partial X_j}\right]=0\\ (4)\ \lambda_i^* g_i(X^*)=0\\ (5)\ X_j^*\geqslant 0\\ (6)\ \lambda_i\geqslant 0\end{cases} \tag{7.3.12}$$

库恩 — 图克条件是非线性规划领域中最重要的理论成果，它适用于求解一般的非线性规划，是确定某点为极值的必要条件。若 $f(X)$ 和 $g_i(X)$ 均为凸函数，则库恩 — 图克条件为 $f(X)$ 取极值的必要且充分条件。

二阶充分条件 对非线性规划：

$$\begin{aligned}&\min \quad f(X)\\ &\text{s.t.} \quad g_i(X)\leqslant 0 \quad (i=1,2,\cdots,m)\end{aligned}$$

满足库恩 — 图克条件的 X^* 为极小点的充分条件是：当任意非零向量 Z 满足：

$$\begin{aligned}&Z^{\mathrm{T}}g_i(X^*)\leqslant 0\text{ 时，使}\\ &Z^{\mathrm{T}}\left[\nabla^2 f(X)+\sum\lambda_i*\nabla^2 g_i(X^*)\right]Z>0\end{aligned} \tag{7.3.13}$$

成立。

式中 $\nabla^2 f(X^*)$、$\nabla^2 g_i(X^*)$—— 是函数 $f(X)$、$g(X)$ 在 X^* 处的海赛矩阵。

现举例说明用库恩 — 图克条件求解有不等式约束的非线性规划问题。

$$\begin{aligned}&\min \quad f(X)=X_1^2+6X_1X_2-4X_1-2X_2\\ &\text{s.t.} \quad g_1(X)=X_1^2+2X_2-1\leqslant 0\\ &\qquad g_2(X)=X_2-X_1-\frac{1}{2}\leqslant 0\end{aligned}$$

由库恩 — 图克条件得：

(1) $2X_1+6X_2-4+2\lambda_1X_1-\lambda_2=0$

$6X_1-2+2\lambda_1+\lambda_2=0$

(2) $X_1^2+2X_2-1\leqslant 0$

$$X_2 - X_1 - \frac{1}{2} \leqslant 0$$

(3) $\lambda_1(X_1^2 + 2X_2 - 1) = 0$

$$\lambda_2(X_2 - X_1 - \frac{1}{2}) = 0$$

(4) $\lambda_1 \geqslant 0$, $\lambda_2 \geqslant 0$

现分别讨论以下四种情况：

第一种：$\lambda_1 = 0$, $\lambda_2 = 0$

该情况使库恩—图克条件(3)、(4)满足。

解(1)得：$X_1 = \frac{1}{3}$, $X_2 = \frac{5}{9}$ 代入(2)得：

$$X_1^2 + 2X_2 - 1 = \frac{2}{9} > 0$$

即不满足 $g_1(X) \leqslant 0$，故该解不是最优解。这种情况实质是由 $\nabla f(X)$ 得到的，是无约束问题的解。

第二种：$\lambda_1 \geqslant 0$, $\lambda_2 \geqslant 0$

由(3)知 $\lambda_1 \geqslant 0$, $\lambda_2 \geqslant 0$ 时，$g_1(X) = 0$, $g_2(X) = 0$，即是等式约束的情况，解在约束的交点上。二者联立并求解得

$$\begin{cases} X_1 = 0 \\ X_2 = \frac{1}{2} \end{cases} \qquad \begin{cases} X_1 = -2 \\ X_2 = -\frac{3}{2} \end{cases}$$

代入条件(1)得

$$\begin{cases} \lambda_1 = 1.5 \\ \lambda_2 = -1 < 0 \end{cases} \qquad \begin{cases} \lambda_1 = -\frac{31}{2} < 0 \\ \lambda_2 = 45 \end{cases}$$

均不符合库恩—图克条件(4)，故该组解也不是最优解。

第三种：$\lambda_1 = 0$, $\lambda_2 \geqslant 0$

由(3)知，当 $\lambda_2 \geqslant 0$ 时，$g_2(X) = 0$，即

$$X_2 - X_1 - \frac{1}{2} = 0$$

由条件(1)得

$$2X_1 + 6X_2 - 4 - \lambda_2 = 0$$

$$6X_1 - 2 + \lambda_2 = 0$$

三方程联立求解得

$$X_1 = \frac{3}{14}, X_2 = \frac{10}{14}, \lambda_2 = \frac{10}{14}$$

代入 $g_1(X)$ 得

$$X_1^2 + 2X_2 - 1 = \frac{93}{196} > 0$$

不满足库恩—图克条件，故该解也不是最优解。

第四种：$\lambda_1 \geqslant 0$, $\lambda_2 = 0$

依照第三种情况，得方程组：

$$\begin{cases} 2X_1 + 6X_2 - 4 + 2\lambda_1 X_1 = 0 \\ 6X_1 - 2 + 2\lambda_1 = 0 \\ X_1^2 + 2X_2 = 1 \end{cases}$$

求解得 $X_1 = X_2 = -1 - \sqrt{2} \qquad \lambda_1 = 4 + 3\sqrt{2}$

代入 $g_2(X)$ 得 $X_2 - X_1 - \frac{1}{2} = -\frac{1}{2} < 0$

该组解全部满足库恩—图克条件，故是最优解。即 $\lambda_1^* = 4 + 3\sqrt{2}, \lambda_2^* = 0, X_1^* = X_2^* = -1 - \sqrt{2}$。

用二阶充分条件检验，为使 $Z_1 g_1(X) \leqslant 0, Z_2 g_2(X) \leqslant 0$ 成立，必使 $Z_1 > 0, Z_2 > 0$。将数据代入式(7.3.13)得

$$(Z_1 \quad Z_2)\left[\begin{pmatrix} 2 & 6 \\ 6 & 0 \end{pmatrix} + (4 + 3\sqrt{2})\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}\right]\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

$$= (10 + 6\sqrt{2})Z_1^2 + 12Z_1Z_2 > 0$$

故这组解是最小解。

7.3.2.3 拉格朗日乘子和库恩—图克乘子的经济涵义——影子价格

库恩—图克条件实质上是拉格朗日乘子法的推广。其区别在于库恩—图克乘子有确定的符号，而拉格朗日乘子可正可负，但它们的经济涵义是相同的。

在线性规划中曾指出单纯形乘子是资源的影子价格，在非线性规划中，拉格朗日乘和库恩—图克乘子也是相应资源的影子价格。

由 $L(X,\lambda) = f(X) + \sum_i \lambda_i g_i(X)$ 可知，当只有一个约束时，X^* 为最优解的必要条件之一是：

$$\frac{\partial L}{\partial X_j} = \frac{\partial f(X)}{\partial X_j} + \lambda \frac{\partial g(X)}{\partial X_j} = 0 \qquad (j = 1,2,\cdots,n)$$

由此可得 $$\lambda = -\frac{\partial f(X)/\partial g X_j}{\partial g(X)/\partial X_j}$$

该式体现出约束的变化使目标变动的比率，即边际值。如果 λ_i 为零，则说明约束中资源数量的变动不会使目标变化；λ 为正，则说明资源量的变化与目标的变化方向相反；λ 为负，说明资源量的变化与目标的变化方向相同。资源量的变化使目标值变化的大小取决于 λ 值的大小，因此 λ 起到了“价格”的作用。

7.3.2.4 罚函数法

拉格朗日乘子法和应用库恩—图克定理求解非线性规划问题均属解析方法(直接法)。

罚函数法是目前应用较广泛的一种求解有约束非线性规划问题的数值解法。

罚函数法的基本思想是把有约束问题变为无约束问题来求解。但考虑到约束对目标函数的影响程度而在构成无约束函数的时候，在目标函数上加一惩罚项，该项为破坏约束而规定一高代价，惩罚项的系数确定了惩罚的严厉程度，同时也确定了无约束问题与原问题的近似程度，当该系数趋近无穷大时，就可得到原问题的解。

对非线性规划问题

$$\begin{aligned} &\min \quad f(X) \\ &\text{s.t.} \quad g_i(X) \leqslant 0 \quad (i = 1,2,\cdots,m) \end{aligned} \tag{7.3.14}$$

定义下式为无约束问题的罚函数：

$$\min \quad P(X,\mu) = f(X) + \sum_i \mu_i p_i(x) \tag{7.3.15}$$

其中 μ_i 为正常数，$P(x)$ 为连续函数，当且仅当 X 满足约束时，$p_i(x) = 0$；否则 $p_i(x) > 0$。$\mu_i p_i(x)$ 称为惩罚项，μ_i 为惩罚因子。

由式(7.3.15)知，如果 $\mu_i = 0$，$p_i(x) \neq 0 (i = 1,2,\cdots,m)$，对于罚函数 $P(X,\mu)$，其最小解由 $\partial f/\partial X$ 决定，相当于不考虑约束，只求原目标函数极值的无约束问题。因 $p_i(x) \neq 0$，X 不满足约束，$P(X,\mu)$ 的极小值不是原问题的极小值。如果 $\mu_i > 0$，$p_i(x) \neq 0 (\mu_i, p_i(x) > 0)$，这时虽然考虑到约束，但因 X 不满足约束，$P(X,\mu)$ 的最小值仍不是原问题的最小值。只有当 $\mu_i > 0$，$p_i(x) = 0$ 时，$\mu_i p_i(x) = 0$，$P(X,\mu)$ 的最小值才是原问题 $f(x)$ 的极小值，罚函数的最优解才是原问题的最优解。

因 $\mu_i > 0$，在 X 不满足约束时，$p_i(x) > 0$，二者之积 $\mu_i p_i(x) > 0$；又由式(5.10.15)知，当且仅当 $\mu_i p_i(x) = 0$ 时，$P(X,\mu)$ 才是原问题的极小值，所以在 $\mu_i > 0$ 时，为使 $P(X,\mu)$ 成为原问题的极小值，只有 $p_i(x) = 0$。在 $\mu_i > 0$ 且为某一确定值 μ_k，X 不满足约束且 $p_i(x)$ 为一确定值 p_k 时，二者之积也为一确定值 C。如欲使 C 不变，μ_i 增大，$p_i(x)$ 必须相应减小；如果 μ_i 趋近无穷大，$p_i(x)$ 必趋近于“0”，$p_i(x)$ 趋近于零表示 X 由不满足约束逐渐向满足约束趋近，罚函数 $P(X,\mu)$ 的最优解也就由 $f(x)$ 决定了。如果 μ_i 为一相对大的数，$p_i(x)$ 则为与零有一定距离的数，μ 越大，$p_i(x)$ 与零的距离越小，当距离小于规定的精度时，罚函数 $P(X,\mu)$ 的最优解可近似地作为原问题的最优解。这样一来 μ_i 就为 X 不满足约束的一种惩罚，μ_i 越大，惩罚越严厉，就越使 X 趋近于约束集合，当 μ_i 趋近于无穷大时，X 就满足约束。因此惩罚因子 μ_i 体现出对 X 不满足约束的惩罚。

由此我们可以得出结论：

(1) 罚函数 $P(X,\mu)$，当惩罚因子 $\mu_i \to \infty$ 时的最优解为原问题的最优解。

(2) 当惩罚因子 μ_i 为一较大值 μ_k 时，罚函数 $P(X,\mu)$ 的最优解为原问题的近似解，其近似程度决定于 μ_k 的大小。

(3) 惩罚因子 μ_i 的增大，使 X 逐渐满足约束，仅当 $\mu_i \to \infty$ 时，X 才满足约束。

应用上述三条结论就可以应用罚函数法处理有约束非线性规划问题。

罚函数法中惩罚因子 μ_i 的作用是使最优解由约束集合的外部逐渐向约束集合内趋近，因此又叫外点法。其经济解释为：若把目标函数看成“价格”，而约束看作某种规定，要求在某种规定范围内购买价格最便宜的东西，且对规定制定一种罚款制度，如果符合“规定”，罚款为零；反之，处以罚款。购货人员所付出的总代价是价格与罚款的总和。所以，应以总代价最小为最终目标，从而得到无约束问题。当把罚款规定得很苛刻时，违反规定需要支付相当多的罚款，这总是不合算的，这就迫使采购物资时，尽量符合规定，使罚款为零。在数学上则表现为惩罚因子 μ_i 很大时，无约束问题(罚函数)的最优解应该满足约束条件，从而得到有约束问题的最优解。

现分两种情况介绍罚函数法。

1. 有等式约束的罚函数法

有等式约束的非线性规划为：

$$\begin{aligned} &\min \quad f(X) \\ &\text{s.t.} \quad g_i(X) = 0 \quad (i = 1,2,\cdots,m) \end{aligned}$$

构建如下形式的罚函数：

$$\min \quad P(X,\mu) = f(X) + \sum_i \mu_i g_i^2(X) \tag{7.3.16}$$

式中 $g_i^2(X)$—— 标准二次罚函数。

如果用解析法求解罚函数 $P(X,\mu)$ 的极小值，先把 μ_i 当作已知量，得到的最优解 $X^*(\mu_i)$，再令 $\mu_i \to \infty$ 即可得到原问题的最优解。

如

$$\begin{aligned} &\min \quad f(X) = 2X_1^2 + X_2^2 \\ &\text{s.t.} \quad X_2 = 1 \end{aligned}$$

则罚函数

$$P(X,\mu) = 2X_1^2 + X_2^2 + \mu(X_2 - 1)^2$$

$$\frac{\partial P}{\partial X_1} = 4X_1 = 0$$

$$\frac{\partial P}{\partial X_2} = 2X_2 + 2\mu(X_2 - 1) = 0$$

解该方程组得

$$X_1 = 0 \qquad X_2 = \frac{\mu}{1+\mu}$$

当 $\mu \to \infty$ 时，$X_2 = 1$，所以原问题的最优解为

$$X_1^* = 0,\ X_2^* = 1,\ \text{目标值}\ f(X^*) = 1$$

表 7.3.1 说明当 μ 逐渐增大时，X_2 由约束外的点 $X_2 = 0$ 逐渐趋于约束集合的情况。

表 7.3.1 μ 与 X_2 的关系

μ	0	1	2	3	4	5	6	…	10	…	100	…	1000	…
X_2	0	0.5	0.67	0.75	0.8	0.83	0.857	…	0.909	…	0.909	…	0.9990	…

由表 7.3.1 可见，当 $X_1 = 0, X_2 = 0$ 时，是目标函数在无约束时的最优解，最优点在约束集合外，随着 μ 的增大，它逐渐趋于约束集合 $X_2 = 1$，从而得到原问题的最优解：$X_1^* = 0,\ X_2^* = 1$。

采用罚函数法求解非线性规划问题，一般用迭代法。迭代法步骤如下：

(1) 取 $\mu_1 > 0$ 作为惩罚因子的初始值，允许误差 $\varepsilon > 0$，令 $k = 1$；

(2) 求 μ 为确定值 μ_k 时，罚函数的极小值：

$$\min \quad P(X,\mu_k) = f(X) + \sum_{i=1}^{m} \mu_{ik} g_i^2(X) = P(X^{(k)},\mu_k)$$

(3) 如果 $g_i(X(k)) \leqslant \varepsilon (i = 1,2,\cdots,m)$，则迭代结束，$X^* = X(k)$，否则令 $\mu_k < \mu_{k+1}$，$k = k + 1$ 转 2。

2. 有不等式约束的非线性规划问题

$$\begin{aligned} &\min f(X) \\ &\text{s.t.} \ g_i(X) \leqslant 0 \quad (i = 1,2,\cdots,m) \end{aligned}$$

所构造的罚函数为

$$\min \quad P(X,\mu) = f(X) + \sum_{i=1} \lambda_i g_i^2(X)\mu_i(g_i(X)) \tag{7.3.17}$$

式中 $\mu_i(g_i(x))$——内部罚函数，

$$\mu_i(g_i(x)) = \begin{cases} 1 & \text{当 } X \text{ 不满足约束时，} \\ 0 & \text{当 } X \text{ 满足约束时。} \end{cases}$$

在求该罚函数极小值时，首先需研究约束是否可满足，如 X 无明显满足约束的条件，则可一律设其不满足约束并令 $\mu_i(g_i(x)) = 1$，而后采用求解等式约束的方法求解。

如：

$$\begin{aligned} \min \quad & f(X) = -X_1 \\ \text{s.t.} \quad & (X_1 - 4)^3 + (X_2 - 3) \leqslant 0 \\ & (X_1 - 4)^3 - X_2 + 3 \leqslant 0 \\ & X_1 \geqslant 0,\ X_2 \geqslant 0 \end{aligned}$$

罚函数为

$$\begin{aligned} P(X,u,\mu) = & -X_1 + \mu_1[(X_1 - 4)^3 + (X_2 - 3)]^2 u_1 + \\ & \mu_2[(X_1 - 4)^3 - (X_2 - 3)]^2 u_2 - \mu_3 X_1^2 u_3 - \mu_4 X_2^2 u_4 \end{aligned}$$

如在第一象限求解，则 $X_1 \geqslant 0$，$X_1 \geqslant 0$，故 $u_3 = u_4 = 0$，$u_1 = u_2 = 1$，令 $\mu_1 = \mu_2$，则罚函数 $P(X,\mu)$ 为

$$\begin{aligned} P(X,\mu) = & -X_1 + \mu_1[(X_1 - 4)^3 + (X_2 - 3)]^2 + \\ & \mu_2[(X_1 - 4)^3 - (X_2 - 3)]^2 \end{aligned}$$

用解析法求解得

$$X_1 = \sqrt{\frac{1}{12\mu}} + 4; X_2 = 3$$

当 $\mu \to \infty$ 时，$X_1^* = 4$，$X_2^* = 3$，最优解为 $f(X) = -4$。

7.3.2.5 碰壁法

碰壁法又叫内点法或障碍法。外点法总是假设 X 不满足约束，通过惩罚因子的增大，使 X 逐渐趋于约束集合。障碍法也是罚函数法的一种，但与上法不同，它总假设 X 在约束集内且距约束边界很远，如在约束集内寻找最优解时，惩罚很小，当 X 靠近约束边界，满足某一约束时，给予高的惩罚，阻止 X 越出边界，因此，对非线性规划：

$$\begin{aligned} \min \quad & f(X) \\ \text{s.t.} \quad & g_i(X) \geqslant 0 \quad (i = 1,2,\cdots,m) \end{aligned}$$

有无约束障碍函数 $P(X,\mu)$ 为

$$P(X,\mu) = f(X) + \sum_i \mu_i \frac{1}{g_i(X)}$$

或

$$P(X,\mu) = f(X) + \sum_i \mu_i \lg(g_i(X)) \tag{7.3.18}$$

式中 $\mu_i \dfrac{1}{g_i(X)}$ 或 $\mu_i \lg(g_i(X))$——障碍项；

μ_i——障碍因子。

当障碍函数 $P(X,\mu)$ 的最优解 $X^*(\mu)$ 中，$\mu \to 0$ 时，即得原问题的解。

由式(7.3.18)可见,当 X 靠近约束边界,使 $g_i(X)\to 0$ 时,$\frac{1}{g_i(X)}$ 或 $\lg(g_i(X))$ 均趋近于无穷大,这样则使 $P(X,\mu)$ 很大,现欲求 $P(X,\mu)$ 的极小值,故仅当 $\mu_i\frac{1}{g_i(X)}$ 或 $\mu_i\lg(g_i(X))$ 为零,即只有当 $\mu_i\to 0$ 时,才能保证 X 不越出约束边界,才能使障碍函数取极小值。由于该方法,一直在约束集合内寻找最优解,故亦称内点法。

如:

$$\begin{aligned}\min\quad & f(X)=\frac{1}{3}(X_1+1)^3+X_2\\ \text{s.t.}\quad & g_1(X)=X_1-1\geqslant 0\\ & g_2(X)=X_2\geqslant 0\end{aligned}$$

建立的障碍函数为

$$\min\quad P(X,\mu)=\frac{1}{3}(X_1+1)^3+X_2+\frac{\mu}{X_1-1}+\frac{\mu}{X_2}$$

用解析法,则

$$\frac{\partial P}{\partial X_1}=(X_1+1)^2-\frac{\mu}{(X_1-1)^2}=0$$

$$\frac{\partial P}{\partial X_2}=1-\frac{\mu}{X_2^2}=0$$

解得

$$X_1=\sqrt{1+\mu},\ X_2=\sqrt{\mu}$$

当 $\mu\to 0$ 时,$X_1^*=1,X_2^*=0$,最优解 $f(X^*)=\frac{8}{3}$。

用迭代法求解可仿照外点法,不同的是在每次迭代中,障碍因子逐渐减小。

有约束非线性规划问题的求解比较复杂,因此在应用内点法或非外点法时,总假设有连续的一阶、二阶偏导数,然而当 μ 值较大时,会引起二阶偏导数不连续或迭代求解收敛过慢,故有时需采用一些其他方法来处理。一般情况下,用 Lingo 软件可很方便地求解。

第 8 章　网络最优化方法

在实际生产和生活中，一些系统的结构模型可清楚地表示系统内部物流、能流和信息流的传递数量和方向，且其单元之间的关系四通八达，构成一网状图形。鉴于这类网状图形可应用图论知识采取一些特殊的最优化方法，因此本章将介绍这类系统的最优化。

8.1　网络和图的基本概念

8.1.1　网络的基本概念和术语

人们在从事各项活动中，经常用点和连接各点的线表示一个系统各单元及之间的某种特定联系。这种用点和线表示的系统各单元及其之间联系的示意图称之为“图”。图 8.1.1 即是一交通示意图，它表示了内蒙古大兴安岭地区各镇的地理位置和交通联系。

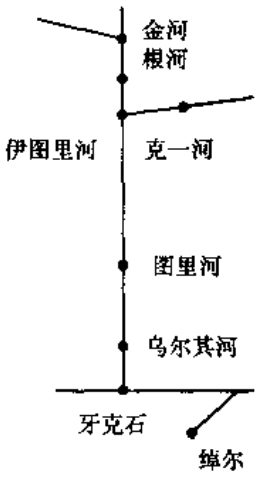


图 8.1.1　内蒙古大兴安岭地的各镇的地理位置和交通联系

交通图、通信网都是图的一种。如果在图的线上标有某种数量指标，则这种图叫网络，通常也称为图。本章所研究的图均指这种带有数量指标的图——网络。

图按其线是否具有方向性又分为有向图和无向图。如输电线路只能由发电厂输往用电地，这种图是有向图；而电话网中任一条线可往返地通信，这种图是无向图。

现介绍图论中经常使用的术语：

(1) 顶点、弧(边)。用点表示组成系统的单元叫图的顶点，用线表示的两单元之间的

特定的关系叫弧或边。一般以 a_i 表示图的顶点，以 $a(i,j)$ 表示由 a_i 出发到 a_j 的一条弧，a_i 称为弧的起点，a_j 称为弧的终点，$a(i,j)$ 称为弧的长。如果 a_i 到 a_j 或 a_j 到 a_i ($i \neq j$) 之间存在一条弧，则称两个顶点是相邻的。

(2) 链、路。如果图中顶点 a_i 到 a_j 之间有二条以上的弧连接，则称之为一条链，在有向图中由一个顶点到另一个顶点之间有同向的弧连接，则 a_i 到 a_j 构成一条路。

如果一个顶点 a_i 有到达 a_i 的一条链，则这条链叫圈；在有向图中弧的方向相同的圈叫回路。如图 8.1.2 所示，$a_2-a_3-a_5-a_8-a_7$ 是一条 a_2 由到 a_7 的链，而 $a_2-a_4-a_3-a_5-a_6$ 则是一条由 a_2 到 a_6 的路。而 $a_5-a_6-a_7-a_8-a_5$ 则构成一个圈，而 $a_3-a_2-a_4-a_3$ 则是一个回路。

(3) 连通图。一个图中，任意两顶点 a_i 和 a_j 之间至少有一条链存在，则该图为连通图，否则为不连通的图。

8.1.2 树的概念和术语

(1) 树。连通且不含圈的图叫树(见图 8.1.3)

(2) 树的性质。树的任意两顶点之间存在且仅存在一条链，如果去掉树中任意一条弧，则图是不连通图；如果用弧把任意不相邻的顶点连接起来，仅存在一个圈，再去掉该圈的另一条弧，又得到另一棵树。因此，该性质可以用一口诀来说明："加一弧，得一圈；去一弧，破一圈"。

(3) 树形图。树形图是满足以下条件的有向树(见图 8.1.4)：

① 图中一顶点 a_0 不是任何弧的终点，该点称为树的根；

② 每一顶点($a_i \neq a_0$)是惟一一条弧的终点。

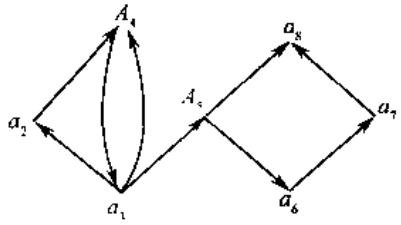


图 8.1.2 链、路图

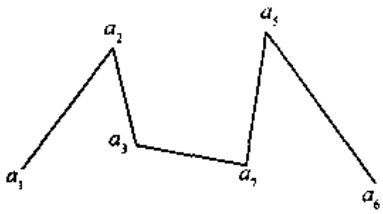


图 8.1.3 树

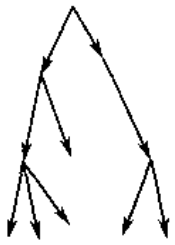


图 8.1.4 树形图

(4) 部分树。在一连通图中，包括所有顶点的一棵树叫部分树。对于给定的图，属于部分树的弧叫树的内边，不属于部分树的弧，叫树的外边。

如果给定连通图，求得一棵部分树使得树的内边长度总和最小，这棵部分树叫最小部分树。

8.2 一笔画和最小部分树问题

8.2.1 一笔画和奇偶点图上作业法

8.2.1.1 一笔画问题

一笔画问题又叫中国邮递员问题。

一个邮递员在他所负责的投递范围内投递书报，每次都必须走遍所有街道，任务完成后返回邮局，怎样走法路线最短呢？

普鲁士哥尼斯堡(现称加里宁格勒)的普雷瓦河中有两个小岛，连接两岸与小岛共架设了七座桥(见图 8.2.1)，人们提出从任一地点出发，可否不重复地走过七座桥返回原地的问题。经过长时期的争论与试验，均未得到解决。

如果将上述两个问题抽象成图的语言就是：给定一个无向的连通图，怎样才能使每条边至少出现一次，达到边长的总和最小。这类问题就叫一笔画问题。

很明显，如果图中有一个圈，能使每条边正巧出现一次，则边长总和必定是最小的，一笔画算法就是解决这类问题的算法。该问题是瑞典数学家欧拉(Euler)首先解决的，因此又叫欧拉问题。

欧拉在研究哥尼斯堡七桥问题时将图 8.2.1 抽象成图 8.2.2 的形式，并指出：从某点出发不重复地走过七座桥返回原地是不可能的。1736 年欧拉提出以下两个定理：

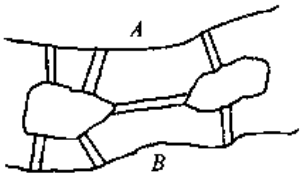


图 8.2.1 小岛间的七座桥

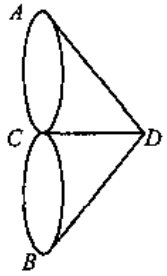


图 8.2.2 图 8.2.1 的抽象描述

定理一 若图 G 的每个顶点所关联的边数是偶数条，则图 G 是一笔圈图；若除链的端点以外其余每个顶点所关联的边数是偶数条，则图是一笔链图。

定理二 如果一个图的奇点数为 2，则此图是一笔画图，但只能从一个奇点出发到达另一个奇点为止。

一个图的顶点可以分成两类，一类是点的关联边数为偶数条，这类点称为偶点；否则称奇点。由定理一知，只有当奇点数为零时，图是一笔画图，即可由任一点出发，不重复地走遍所有边，返回出发点，得到一条最短的线路，图 8.2.3 就是这类图。由定理二知，图 8.2.4 也是一笔画图，因含有二个奇点，所以它只能由 $a_4(a_3)$ 点出发不重复地走遍所有边回到 $a_3(a_4)$。

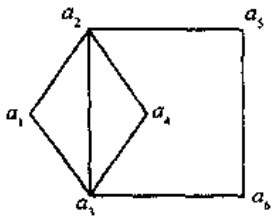


图 8.2.3 一条最短的线路

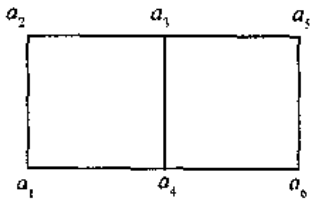


图 8.2.4 一笔画图

8.2.1.2 奇偶点图上作业法

在实践中,一笔画图是很少的。对非一笔画图来说,要从图的一点出发经过所有边后返回该点,则必须在某些边上重复。如欲使所通过的路程最短,就需研究在哪些边上重复才能使所重复的路程最短,解决该问题的方法之一就是奇偶点图上作业法。

一个图中,既然有奇点,必然成对出现。如欲消灭图中所有奇点,就必须引入新边。

引入新边消灭奇点的方法很多,现需解决的问题是如何在引入新边的各种方法中选择一种使其长度最小的方案。

为了迅速地确定引入新边的最优方案,现将引入的新边称之为重复边,并给出以下判断定理:

在最优方案中,图的每一边上至多有一条重复边,并且图中每个圈上重复边的总长度不大于该圈总长的一半。

按照该判定定理,可先任意选择一方案并判断其最优性,如不是最优方案,则可让重复边在圈的另一些边上重复,再按定理进行判断并调整,直至满足判定条件为止。

现举例说明奇偶点图上作业法。

某工地值班人员每天需对所管辖的范围进行巡视,其巡视路线如图 8.2.5(a)所示,如何确定最优巡视路线?

图 8.2.5(a)中有 4 个奇点,因此必须引入重复边。现引入(a_1 a_2)、(a_1 a_4)、(a_3 a_9)、(a_8 a_9)得到图 8.2.5(b)。它是一笔画图,其总长度为

$$(5+5+6+9+2+3+4+3+4+4+4+4)+(5+9)+(4+3)=74$$

由于圈 $a_1a_2a_5a_4a_1$ 中重复边长为 $5+9=14$,超过总圈长$(5+9+6+4)=24$ 的一半,此路线不满足定理,因此应在此圈的另外两条边上引重复边,代替已引入的两条重复边,得图 8.2.5(c),这时总长降到:$53+10+7=70$。

由 a_3、a_2、a_5、a_8、a_9、a_6、a_3 所构成的圈上重复边的长度为

$$6+3+4=13$$

全圈长为

$$5+6+4+3+4+2=24$$

显然该圈不满足定理,再将重复边调整到该圈的非重复边上,得图 8.2.5(d)。

该图总边长为

$$53+7+8=68$$

所有圈中重复边长均不超过所在圈长的一半,符合判定定理,该图为最优方案。因该图是一笔画图,故值班员可在任一地出发不重复地走遍所有边返回出发地,所走的路程最少。

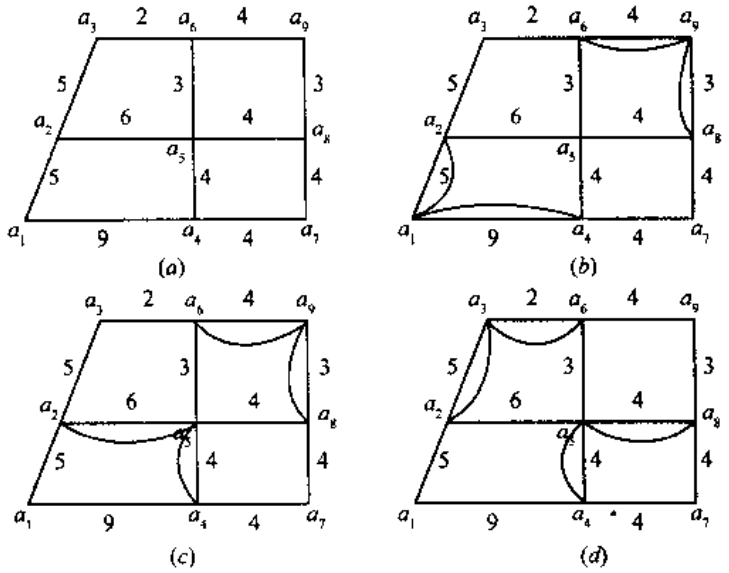


图 8.2.5　奇偶点图上作业法举例

一笔画问题的应用很广，如道路设计中踏察路线的确定、工作地的安排、工件加工路线的设计、旅游线路的安排等很多工作，均可抽象成这类网络来解决。

8.2.2　最小部分树问题

8.2.2.1　最小部分树定理

当且仅当树 T 的每条外边满足

$$l_{ij} \geqslant \max\{l_{ii_1}, l_{i_1i_2}, \cdots, l_{i_ki_j}\} \tag{8.2.1}$$

时，树 T 为最小部分树。

式中　　l_{ij}——$a(ij)$的长度；

$\{l_{ii_1}, l_{i_1i_2}, \cdots, l_{i_ki_j}\}$——为树中由 a_i 到 a_j 的惟一一条链。

通俗地说，上述定理的涵义为：树的任意一条外边 l_{ij} 比所对应的链中最长的边还长；也可以说，最小部分树是由图中去掉圈中最长的边构成的。

与此相反，还可得到最大树定理：

当且仅当树 T 的每条外边 l_{ij} 满足：

$$l_{ij} \leqslant \min\{l_{ii_1}, l_{i_1i_2}, \cdots, l_{i_ki_j}\} \tag{8.2.2}$$

时，树 T 为最大树。

由此定理，我们可以解决一类在既定的图中寻找一棵最小部分树的问题。

8.2.2.2　最小部分树算法

如某小区准备在六栋楼房间架设电话线，各楼房的位置如图 8.2.6(a)所示，如何架线可使各楼房间均能通话且电话线路最短。

算法一　逐次地从图中未选的边中选最短的边，并使所选的边不构成圈，直至得到一棵最小部分树。

该问题中，依次选择 a_2a_3、a_2a_4、a_4a_5、a_4a_6、a_2a_1，如图 8.2.6(b)所示。

架设电话线的最短长度为

$$1+2+3+4+5=15$$

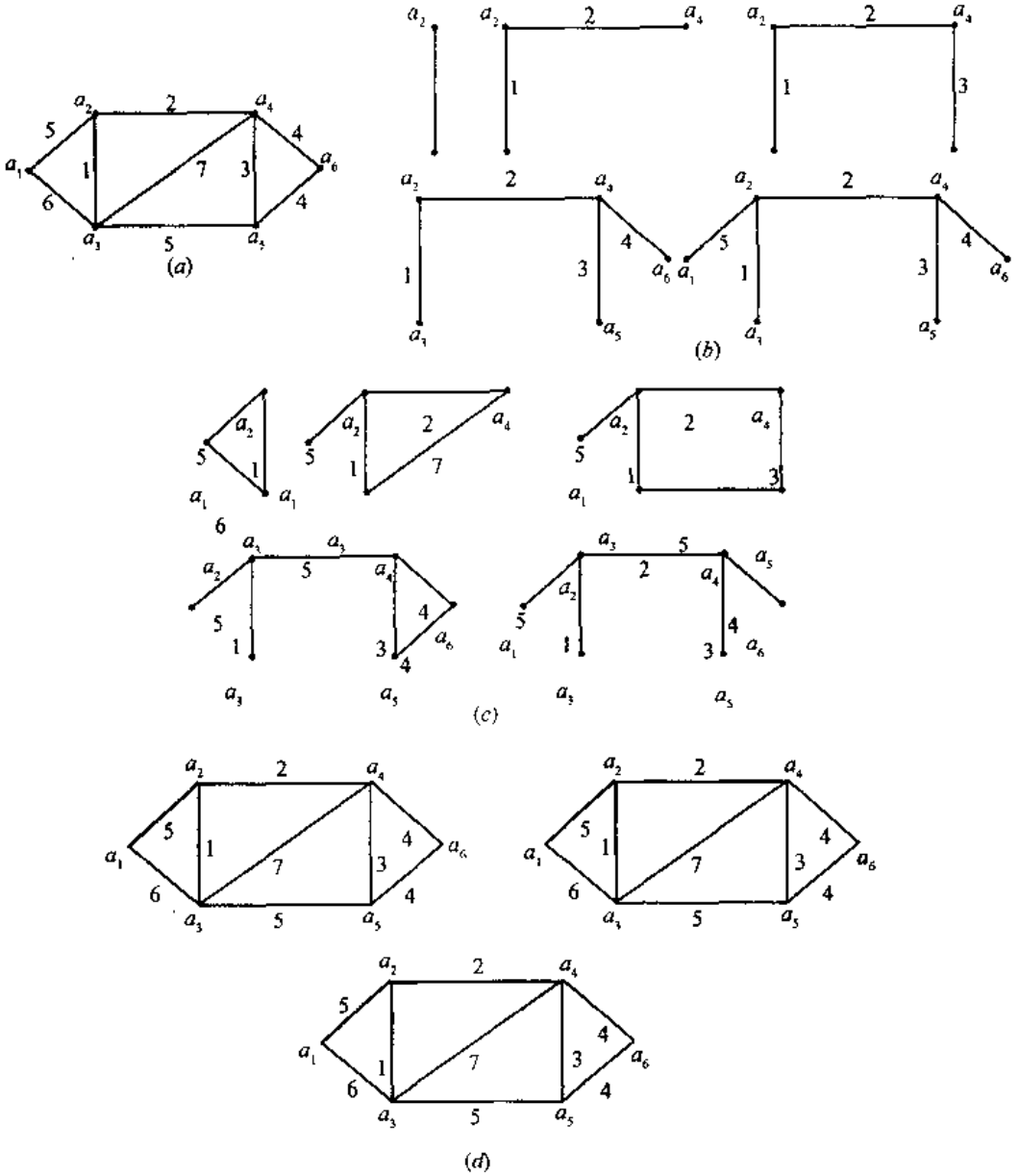
算法二 逐次地从 G 中还未去掉的边中，去掉最长的边，并使每次去边后剩下的图仍是连通图。

这种方法与上述方法正好相反(见图 8.2.6(b))，故不赘述。

算法三(丢边破圈法) 在 G 中任选一圈，去掉最大的一条边，将余下的边再与图中其他边组成的圈中去掉最大的边，如此，直到得到的一棵部分树为止。

对上例，先选圈 $a_1\,a_2\,a_3\,a_1$，去掉最长的边($a_3\,a_1$)；再选圈 $a_2\,a_4\,a_3\,a_2$，去掉最长的边($a_3\,a_4$)；在圈 $a_4\,a_5\,a_3\,a_2$ 中去掉($a_3\,a_5$)；在圈 a_4, a_5, a_6, a_4 中去掉(a_4, a_6)或(a_5, a_6)即得到一棵部分树，且为最小部分树，见图 8.2.6(c)。

算法四 从图 G 中任选一棵部分树 T，加上 T 的一条外边形成一个圈，然后用丢边破圈法去掉一条长边，如此下去，直到得到最小部分树。



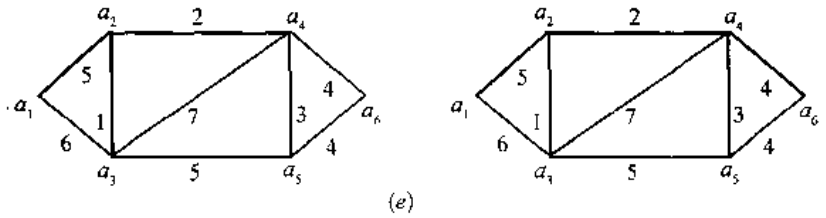


图 8.2.6 最小部分树算法举例

仍用图 8.2.6(a)来说明。任选一棵树如图 8.2.6(d),在树中加入一条外边(a_2 a_4),构成圈 a_3 a_4 a_2 a_3,去掉最长的边(a_3 a_4),再加入外边则均大于或等于该树的内边,故得到最小部分树。

算法五(生成法) 这种算法是使树在由小到大逐步生成的过程中,在可能生长的方向,取最小的树枝,则生成的部分树是最小部分树。

仍用图 8.2.6(a)来说明:任取一点 a_1 为起始点,它的可能生长方向为:向 a_2 或 a_3 方向,$l_{12}=5$,$l_{13}=6$,选 $\min\{5,6\}=5$,故向 a_2 方向生长,用黑线把 a_1a_2 加粗,这就生成了一条树枝。在此树枝的基础上,可能生长方向有三个,即由 a_1 向 a_3,a_2 向 a_3,a_2 向 a_4,其中 $l_{13}=6$,$l_{23}=1$,$l_{24}=2$,因此树枝的生长为由 a_2 到 a_3,把 a_2a_3 加粗,这样就又生成一条树枝。在此两条树枝的基础上,树仍可按三个方向生长,$a_2\to a_4$;$a_3\to a_4$;$a_2\to a_5$ 选 $a_2\to a_4$,这样一直继续下去,直到不能再进行为止。则粗线所表示的树即是最小部分树。见图 8.2.6(e)。

由以上各种算法可见,计算结果都是相同的。

8.3 最短路径问题

最小部分树问题的特点是在图中得到一个包含所有顶点的一棵树,使其边长之和最小。

所谓最短路径问题是在图中求从某一固定点 S 到另一固定点 T 的路,使其长度为所有由 S 到 T 的路径中最短的一条。与最小部分树问题的不同点在于不一定包含图的所有顶点。

8.3.1 两固定顶点间的最短路径问题(狄克斯特拉法)

如图 8.3.1 所示,假设出发点是①,终点是④,从①出发,到④有各种不同的走法,现欲寻找一条由①到达④的最短路径。

由图可知,由节点①到节点④共有 3 条路线:

(1) ①→②→④,距离是 6+4=10;

(2) ①→③→②→④,距离是 2+1+4=7;

(3) ①→③→④,距离是 2+8=10;

路线①→③→②→④最短,距离为"7"。

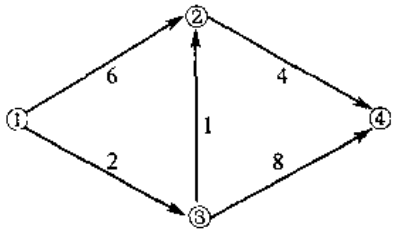


图 8.3.1 ①到达④的最短路径

这种寻找最短路径的方法是列出由①点到④点的全

部路径，通过比较找出距离最短的路径，这种方法叫穷举法。这种方法准确可靠，但只适用于小网络，不适用于大型网络。现应用树的生成原理，采用生成法解最短路径问题。

现举例说明该方法。工地道路网如图 8.3.2(a)所示，工地值班员准备从 a_1 点出发去 a_8 点，怎样走，距离最近呢?

由 a_1 出发，在 a_1 点标"①"，圈内的"0"表示与点 a_1 的最短距离。标记后，用树的生成法一步一步向外试探，由 a_1 点可能生长方向为 $a_1 \to a_2$、$a_1 \to a_4$，因 $l_{12}=5$，$l_{14}=9$，$\min\{5,9\}=5$；故 a_2 点 a_1 到点的最短距离为 5，在 a_2 点标记"⑤"，并加深 a_1a_2，如图 8.3.2(b)所示。

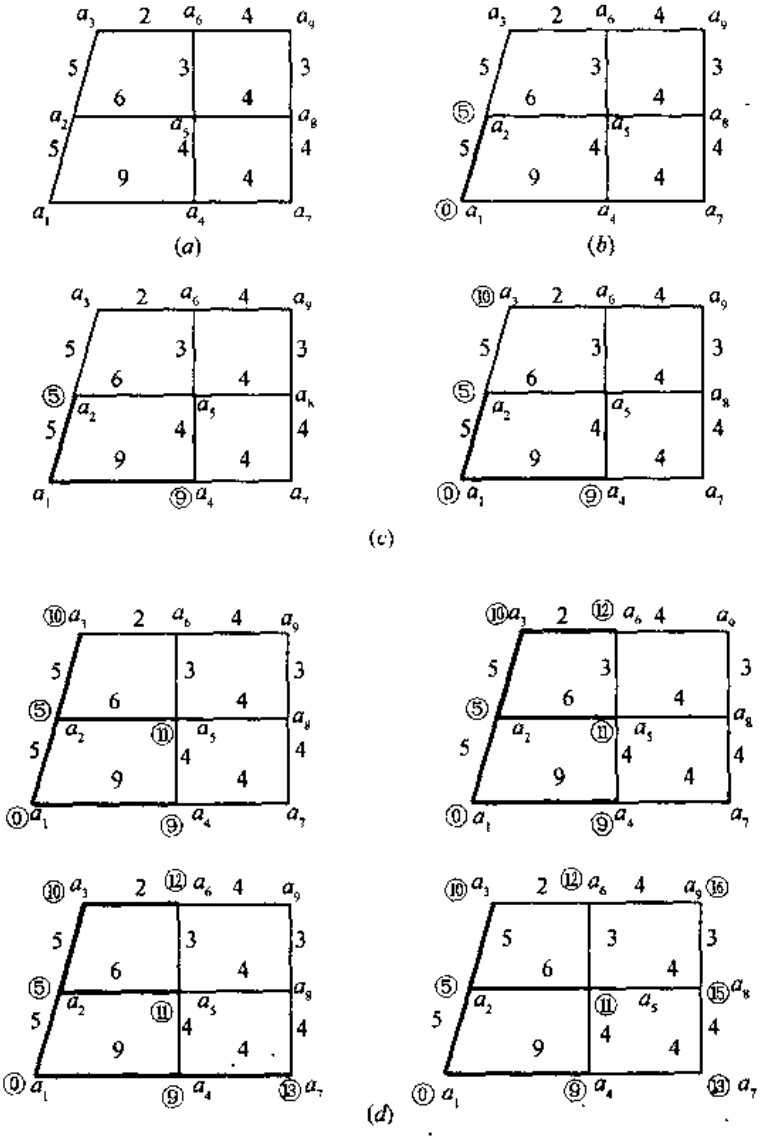


图 8.3.2　工地道路网各点至 a_1 点的最短路径

在 a_1a_2 这条树枝的基础上，再在其可能的生长方向中选择距离 a_1 最近的点作为生长方向，可能的生长线有 $a_1 \to a_4$、$a_2 \to a_3$、$a_2 \to a_5$，在这三条生长线中，每条都有一点在已生成的树枝内，并已标上距 a_1 点的最短距离，而另一点都是未标记与 a_1 最短距离的点，为了选择一条到出发点 a_1 距离最短的路径，必须用树内已标记的距离加上这条可能生长

线的长度，即 $l_{14}=0+9=9$，$l_{23}=5+5=10$，$l_{25}=5+6=11$，选择 $\min\{9,10,11\}=9$，即 a_1a_4 作为生长线，在 a_4 处标记“⑨”并加粗 a_1a_4，得图 8.3.2(c)。

重复上述过程，则得图 8.3.2(d)。由 a_1 出发到达 a_8 的最短路径为 $a_1\rightarrow a_2\rightarrow a_5\rightarrow a_8$，长度为 15，同时还得出图中各点到 a_1 的最短距离，即各顶点的标记数和路程。

该方法是狄克斯特拉(Dijkstra)于 1959 年提出的，因此又叫狄克斯特拉法。该方法除可在图上标记之外，对于大型网络，需建立数学模型用计算机来解，其算法如下：

(1) 列出网络矩阵 A，其元素 a_{ij} 为：

$$a_{ij}=\begin{cases}a(i,j), & \text{当 } a_i \text{ 与 } a_j \text{ 相邻时}\\ 0, & \text{当 } i=j \text{ 时}\\ \infty, & \text{当 } a_i \text{ 与 } a_j \text{ 不相邻时}\end{cases}$$

(2) 定义 $d(j)$为第 j 个顶点的标记数，表示第 j 个顶点到点 s 的最短距离。当 $j=s$ 时，$d(j)=0$，否则 $d(j)=\infty$。令 Y 为最后标记的顶点。

(3) 开始计算：令 $Y=s$。

(4) 对每个未标记的点 j，按下式重新计算 $d(j)$值：

$$d(j)=\min\{d(j),d(y)+a(y,j)\}$$

(5) 按下式选择标记点 k：

$\min\{d(j)\}=d^*(k)$，$j\in$未标记的点。标记 $d(j)$所对应的点和边(y,k)并令 $y=k$。

(6) 如果 $\min\{d(j)\}=\infty$，$j\in$未标记的点，则停止，输出无最短路径。

如果 T 已标记，即 $k=T$ 则结束，输出最短路径，否则转(4)。

8.3.2 边长有负值网络问题——福特法

上述算法只适用于边长 $a(i,j)\geqslant 0$ 的网络，对边长有负值的网络是不适用的。

求解边长有负值的网络在思路上与上法基本相同，但由于边长存在负值，采用上述方法可能出现先标记的数字大于后标记数字的情况。尤其是出现负回路的情况将使标记法无法进行。因此必须在计算方法上加以改变。

现介绍有负边长网络最短路径的计算方法。该方法的基本要点是：确定 s 点到 T 点的最短路径必须象上法那样，在确定 s 点到中间各点最短距离的基础上进行，即确定 s 点到网络各点的最短距离。如果网络有 n 个顶点，则由 s 至某点的最短距离必须由 s 出发，直接到达该点和经由一个顶点到达该点、经由二个顶点…、经由 $n-1$ 个顶点到达该点的最小值，中间经过 i 个顶点的最短距离，又是在中间经过 $i-1$ 个顶点的基础上进行的，因此算法必须从确定直接到达每个顶点的最短距离开始，依次计算经由一个顶点、二个顶点…、$n-1$ 个顶点到达该顶点的最短距离。如确定由 s 经由一个顶点到达 2 顶点的最短距离，必须在已知 s 直接到达各顶点最短距离的基础上，由路径 $s-s-2$、$s-1-2$、$s-2-2$、$s-3-2$、…$s-n-2$ 中选最小值来确定(画横线部分是 s 直接到达各顶点的距离)，如此，只要知道 s 点到各顶点的距离，就可递推得出点 s 经由一个顶点、经由二个顶点…、经由 $n-1$ 个顶点到达各顶点的最短距离。通过选择最小值时的对应关系即可确定路径。

该算法用数学语言描述如下：

设：(1)有 n 个顶点，且每对顶点间均有边存在(如无边存在，则设边长为∞)；

(2) 由 s 到 j 的最短距离为 $l_{sj}(j=1,2,\cdots,n)$，当 $j=T$ 时为所求路径的最短距离；

(3) $l_{sj}^{(i)}$ 为 s 径由 i 个顶点到达 j 点的最短距离 $(j=1,2,\cdots,n)$；

(4) k 为最短路径上的中间点。

则
$$l_{sj}=\min\{l_{sk}+a(k,j)\} \tag{8.3.1}$$

$l_{sk}(k=1,2,\cdots,n-1)$ 可由下式递推得出

$$\begin{gathered} l_{sj}^{(i)}=a(s,j) \\ l_{sj}^{(i)}=\min\{l_{sk}^{(i-1)}+a(s,j)\} \\ i=2,\cdots,n-1 \end{gathered} \tag{8.3.2}$$

如此则得出由 s 至网络各顶点的最短距离，最短路径在每次选择中确定。

1. 具体算法

步骤 1 列出网络矩阵 $\boldsymbol{A}$，其元素 a_{ij} 为

$$a_{ij}=\begin{cases} a(i,j), & \text{当 } a_i \text{ 与 } a_j \text{ 相邻时} \\ 0, & \text{当 } i=j \text{ 时} \\ \infty, & \text{当 } a_i \text{ 与 } a_j \text{ 不相邻时} \end{cases}$$

该矩阵的行表示 i 顶点直接到达各顶点的距离，列表示各顶点到达 j 顶点的距离。

步骤 2 确定 $l_{sj}^{(1)}$，它是 $\boldsymbol{A}$ 矩阵中 s 行的元素，表示出发点 s 到达各顶点的距离。

步骤 3 按 $l_{sj}^{(i)}=\min\{l_{sk}^{(i-1)}+a(s,j)\}$ 计算 $l_{sj}^{(i)}$，并标记路径。当 $l_{sj}^{(i)}=l_{sj}^{(i-1)}$ 时，计算结束。$L_{sj}^{(i)}$ 表示 s 经 i 个顶点到达 j 的最短距离。

步骤 4 根据 $l_{sj}^{(i)}$ 及标记，反向确定最短路径。

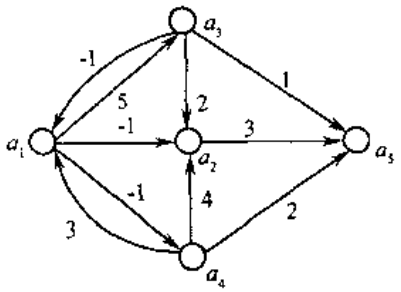


图 8.3.3 网络矩阵

2. 举例

该算法可在表上进行，现举例说明其算法。

步骤 1 建立图 8.3.3的网络矩阵。

步骤 2 确定 $l_{sj}^{(1)}$，它是网络矩阵中第一行（a_1 顶点对应的行），见表 8.3.1。

表 8.3.1 网络矩阵表

	a_1	a_2	a_3	a_4	a_5	$l_{sj}^{(1)}$	$l_{sj}^{(2)}$	$l_{sj}^{(3)}$
a_1	0	-1	5	-1	∞	0	0	0
a_2	∞	0	∞	∞	3	-1	-1(1-2)	-1(1-2)
a_3	-1	2	0	∞	1	5	5(1-3)	5(1-3)
a_4	3	4	∞	0	2	-1	-1(1-4)	-1(1-4)
a_5	∞	∞	∞	∞	0	∞	1(4-5)	1(4-5)

步骤 3 按 $l_{sj}^{(i)}=\min\{l_{sk}^{(i-1)}+a(s,j)\}$ 计算 $l_{sj}^{(i)}$。该步在表上的计算方法是以 $l_{sj}^{(1)}$ 列与各顶点所对应列 a_j 的元素相加，从中取最小值，记在 $l_{sj}^{(2)}$ 列对应于 a_j 的行的位置。如 $l_{sj}^{(2)}$ 列中的第三行元素 5，是 $l_{sj}^{(1)}$ 列的元素分别与 a_3 列元素相加的最小值，它是由 a_1—a_3 这一路径决定的，故在其后标出(1-3)以备确定路径时使用。$l_{sj}^{(3)}$ 用 $l_{sj}^{(2)}$ 列元素分别与

各列元素相加取最小值后填在相应的位置上，依此类推，一直计算到 $l_{sj}^{(3)}$ 和 $l_{sj}^{(2)}$ 完全相等时为止。本例计算到 $l_{sj}^{(3)}$ 与 $l_{sj}^{(0)}$ 相同，故计算结束。

步骤 4 确定由 a_1 出发到各顶点的最短路径。假设现欲求 a_1 至 a_5 的最短路径，由表 8.3.1 中 $l_{sj}^{(3)}$ 可知，其最短距离为 1，其路径由其后的标记(4－5)知，是由 a_4—a_5；由 a_4 所对应行的 $l_{sj}^{(2)}$ 列中的标记知是由 a_1—a_4，故最短路径为 a_1—a_4—a_5，长度为 1。

该方法除可解决含有负边长的网络之外，也可用于解决全部是正边长的网络。如图 8.3.2 所示的网络，用该方法计算的结果见表 8.3.2。

表 8.3.2 计算结果

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	$l_{sj}^{(1)}$	$l_{sj}^{(2)}$	$l_{sj}^{(3)}$	$l_{sj}^{(4)}$
a_1	0	5	∞	9	∞	∞	∞	∞	∞	0	0	0	0
a_2	∞	0	5	∞	6	∞	∞	∞	∞	5	5(1－2)	5(1－2)	5(1－2)
a_3	∞	∞	0	∞	∞	2	∞	∞	∞	∞	10(2－3)	10(2－3)	10(2－3)
a_4	∞	∞	∞	0	4	∞	4	∞	∞	9	9(1－4)	9(1－4)	9(1－4)
a_5	∞	∞	∞	∞	0	3	∞	4	∞	∞	11(2－5)	11(2－5)	11(2－5)
a_6	∞	∞	∞	∞	∞	0	∞	∞	4	∞	∞	12(3－6)	12(3－6)
a_7	∞	∞	∞	∞	∞	∞	0	4	∞	∞	13(4－7)	13(4－7)	13(4－7)
a_8	∞	∞	∞	∞	∞	∞	∞	0	3	∞	∞	15(5－8)	15(5－8)
a_9	∞	∞	∞	∞	∞	∞	∞	∞	0	∞	∞	∞	16(6－9)

由表 8.3.2 中 $l_{sj}^{(4)}$ 一列数字可见，与生成法各顶点的标记数是完全相同的。其路径由 $l_{sj}^{(4)}$ 一列反向寻找即可确定(有负边长时，路径的确定必须按顶点向 $l_{sj}^{(i-1)}$ 依次倒推得出)。

根据该步骤可编制程序用计算机进行计算。

最短路径问题还有很多解法，如可找到由出发点到终点的任一棵树，应用最小部分树原理或丢边破圈法等去解决，有兴趣的同志请参阅有关资料。

8.3.3 任意一对顶点之间的最短路径问题——福劳德算法

网络图中任意一对顶点之间的最短路径问题比上述的最短路径问题更有普遍性。该问题可应用上述方法分别将图中每个顶点当作 S，每个顶点当作 T，通过反复计算来确定。但这样做，工作量将相当大，因此采用一种新算法——福劳德算法。

设网络顶点编号为 $1,2,\cdots,N$，令 d_{ij}^m(D^m 矩阵的元素)为第 i 个顶点到第 j 个顶点的一条最短路径的长度，该路中只允许前 m 个顶点作为中间顶点；如不存在这种路，则 $d_{ij}^m=\infty$。由 d_{ij}^m 涵义知：$d_{ii}^0=0$，d_{ij}^0 为两相邻顶点间的弧长，d_{ij}^N 为所要确定的 ij 两点间的最短距离。

1. 福劳德算法的步骤

(1) 确定 $\boldsymbol{D}^0$ 矩阵，d_{ij}^0 为 i 点到 j 点的最短弧长，如 i、j 两点间无弧，则 $d_{ij}^0=\infty$，当 $i=j$ 时，$d_{ij}^0=0$；

(2) 对 $m=1,2,\cdots,N$，依次由 $\boldsymbol{D}^{m-1}$ 应用下式确定 $\boldsymbol{D}^m$：

$$d_{ij}^m=\min\{d_{im}^{m-1}+d_{mj}^{m-1},d_{ij}^{m-1}\}$$

如果选出的最小元素是由 $d_{im}^{m-1}+d_{mj}^{m-1}$ 决定的，则记下 m；如二者相等或取决于 d_{ij}^{m-1}，则不记录；

(3) 直到计算出 $\boldsymbol{D}^m$ 为止。

2. 举例

某工地有四个重点防火点，道路和防火点位置如图8.3.4所示，现安全员需确定在任一防火点到其他防火点的最短路径。

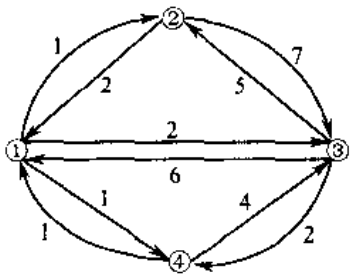


图 8.3.4 防火点位置

应用福劳德算法：

(1) 写出 $\boldsymbol{D}^0$：

$$\boldsymbol{D}^0=\begin{array}{c}\\1\\2\\3\\4\end{array}\begin{array}{c}\begin{array}{cccc}1&2&3&4\end{array}\\\begin{bmatrix}0&1&2&1\\2&0&7&\infty\\6&5&0&2\\1&\infty&4&0\end{bmatrix}\end{array}$$

(2) 根据 $\boldsymbol{D}^0$ 列表求出 $\boldsymbol{D}^1$ 和所对应的路径(计算见表 8.3.3)如下：

$$\boldsymbol{D}^1=\begin{array}{c}\begin{bmatrix}0&1&2&1\\2&0&4&3\\6&5&0&2\\1&2&3&0\end{bmatrix}\end{array}\begin{array}{c}\begin{array}{c}1\\2\\3\\4\end{array}\\\begin{bmatrix}&&1&1\\&&&\\1&1&&\end{bmatrix}\end{array}\begin{array}{c}\\1\\2\\3\\4\end{array}\begin{array}{c}\\\begin{array}{cccc}1&2&3&4\end{array}\end{array}$$

表 8.3.3 计算

$d_{ij}^1=\min\{d_{i2}^0+d_{2j}^0,d_{ij}^0\}$	中间点	$d_{ij}^1=\min\{d_{i2}^0+d_{2j}^0,d_{ij}^0\}$	中间点
$d_{11}^1=d_{11}^0=0$		$d_{31}^1=\min\{6+0,6\}=6$	
$d_{12}^1=\min\{0+1,1\}=1$		$d_{32}^1=\min\{6+1,5\}=5$	
$d_{13}^1=\min\{0+2,2\}=2$		$d_{33}^1=\min\{6+2,0\}=0$	
$d_{14}^1=\min\{0+1,1\}=1$		$d_{34}^1=\min\{6+1,2\}=2$	
$d_{21}^1=\min\{2+0,2\}=2$		$d_{41}^1=\min\{1+0,1\}=1$	
$d_{22}^1=\min\{2+1,0\}=0$		$d_{42}^1=\min\{1+1,\infty\}=2$	1
$d_{23}^1=\min\{2+2,7\}=4$	1	$d_{43}^1=\min\{1+2,4\}=3$	1
$d_{24}^1=\min\{2+1,\infty\}=3$	1	$d_{44}^1=\min\{1+1,0\}=0$	

$\boldsymbol{D}^1$ 表示中间点为 1 时的各点间的最短距离；对应的路径表示经过中间点 1 的最短路径，如 $d_{12}^1=2$ 表示路径④－①－②对应的距离为 2。

(3) 由 $\boldsymbol{D}^1$ 按上法求 $\boldsymbol{D}^2$，此时中间点为两个，即 $m=2$，公式为

$$d_{ij}^2=\min\{d_{i2}^1+d_{2j}^1,d_{ij}^1\}$$

计算得

$$\boldsymbol{D}^2=\begin{pmatrix}0&1&2&1\\2&0&4&3\\6&5&0&2\\1&2&3&0\end{pmatrix}$$

因中间点无选择，故对应路径与上同。注意，$\boldsymbol{D}^2$ 为经过两个点，而不是经第二个顶点，因为 $d_{ij}^2=\min\{d_{i2}^1+d_{2j}^1,d_{ij}^1\}$ 中 d_{i2}^1 已经为经过一个中间点的值。

(4) 再由 $\boldsymbol{D}^2$ 计算 $\boldsymbol{D}^3$，此时公式为

$$d_{ij}^3=\min\{d_{i3}^2+d_{3j}^2,d_{ij}^2\}$$

计算得 $\boldsymbol{D}^3$ 为

$$\boldsymbol{D}^3=\begin{pmatrix}0&1&2&1\\2&0&4&3\\6&5&0&2\\1&2&3&0\end{pmatrix}$$

对应路径仍同上。

(5) 最后由 $\boldsymbol{D}^3$ 求 $\boldsymbol{D}^4$，公式为

$$d_{ij}^4=\min\{d_{i4}^3+d_{4j}^3,d_{ij}^3\}$$

计算得 $\boldsymbol{D}^4$ 和相应的路径为

$$\boldsymbol{D}^4=\begin{pmatrix}0&1&2&1\\2&0&4&3\\6&4&0&2\\1&2&3&0\end{pmatrix}$$

其路径为：

$$\begin{array}{c}1\\2\\3\\4\end{array}\begin{pmatrix}1&2&3&4\\&&1&1\\4&4&&\\1&1&&\end{pmatrix}$$

由 $\boldsymbol{D}^4$ 可见，网络任两点间的距离均求出，所对应的路径也可确定，如点3与点2之间的最短距离为4，其路径为③—④—①—②，即由路径矩阵知，先由第3行第2列对应的4，决定先由③至④，再查该矩阵的第4行第二列④到②的路径为④—①，再查①—②的路径，因①—②对应元素为零，故可直接到达，从而确定③至②的路径为：③—④—①—②。

8.3.4 最短路径问题的应用

最短路径问题在生产中有很重要的作用，不少问题都可抽掉其实际意义，借助于图的概念建立网络模型而化为最短路径问题。现举例说明最短路径问题在设备更新中的应用。

某企业用最短路径法制定设备更新计划。每年年初企业都需决定购买新设备，还是继续使用旧设备。购买新设备，需支付购置费，继续使用旧设备，需支付一笔维修费，现需

制定一个五年计划,使企业在设备更新和维修中所支付的费用最少。

若已估出设备在五年中每年年初的价格和使用不同年限的设备维修费用(见表8.3.4)就可用网络模型来求解。

表8.3.4 设备价格和维修费用预计表 (单位:万元)

年份	1	2	3	4	5
价格	1.1	1.1	1.2	1.2	1.3
维修费	0.5	0.6	0.8	1.1	1.8

根据表8.3.3的资料可得历年费用支出矩阵 $\boldsymbol{D}$

$$\boldsymbol{D}=\begin{pmatrix}1.6 & 2.2 & 3.0 & 4.1 & 5.9\\ 0 & 1.6 & 2.2 & 3.0 & 4.1\\ 0 & 0 & 1.7 & 2.3 & 3.1\\ 0 & 0 & 0 & 1.7 & 2.3\\ 0 & 0 & 0 & 0 & 0\end{pmatrix}$$

式中 d_{ij}——第 i 年购进的设备使用到第 j 年所支出的总费用。

其网络模型如图8.3.5所示。

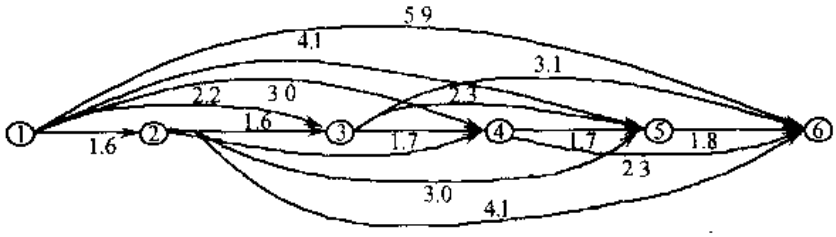


图8.3.5 设备更新与维修问题的网络模型

应用生成法求解,得最短路径,见图8.3.6,由图可知,当第一年购进新设备,应于第三年初或第四年初再购新设备,其他年份应继续使用旧设备为最优方案,所支出的总费用为5.3万元。

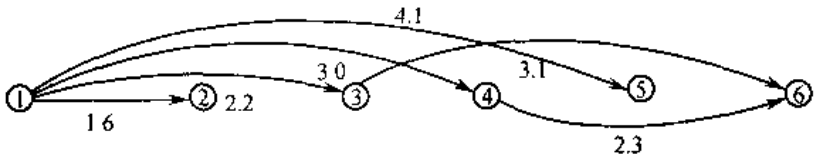


图8.3.6 最短路径

8.4 最大流问题

在实践中除经常遇到最短路径问题之外,还有一类很重要的问题,这就是求一个网络系统的最大通过能力,也就是本节要研究的网络最大流问题。

如我国东北是木材、煤炭、石油等产品的主要产区,如何在有限的运输能力下,使得这些物质尽可能多地运往祖国各地;在通过能力一定的情况下,在一定的时间内如何使甲乙

两地通过的人流、物流、能流和信息最多，这类问题都是网络最大流问题。

在介绍网络最大流问题前，先做几个假设并规定一些符号的意义。

8.4.1 基本假设和符号

(1) 把出发点称为源点，以 S 表示；终点称为汇点，以 T 表示；其余点为中间点，以 i 表示。

(2) 假设源点有无限多的物质、能量、信息及人力等待运出，而汇点则有无限大的仓库来存储这些物资、能量、信息及人力等。

(3) 假设中间点不存储任何物资、能量和信息等，只起传递作用。

(4) 由 i 点到 j 点的最大传递数量规定为线路的容量，以 $C(i,j)$ 表示。

(5) 规定 $X(i,j)$ 为由 i 点到 j 点的实际通过流量。

8.4.2 基本的定理和概念

8.4.2.1 能流守恒定律

依据上述假设，仿照电学中著名的克希荷夫定律，可得到能流守恒定律。

因为 $\sum X(i,j), j\in S(i)$，(其中 $S(i)$ 为可以直接到达 j 点的点的集合)表示流出 i 点的总的流量。$\sum X(j,i), j\in R(i)$，(其中 $R(i)$ 为可以直接到达 i 点的点的集合)表示流入 i 点的总流量。由于克希荷夫定律中流入节点的电流之和与流出节点的电流之和相等的原理与上述假设中的中间点只起传递作用相似，因此有：

$$\sum_{j\in S(i)} X(i,j) - \sum_{j\in R(i)} X(j,i) = \begin{cases} f, & \text{当 } i = S \text{ 时} \\ 0, & \text{当 } i \neq S, i \neq T \text{ 时} \\ -f, & \text{当 } i = T \text{ 时} \end{cases} \tag{8.4.1}$$

其中 f 表示流量。由于式(8.4.1)与克希荷夫定律相似，故叫做能流守恒方程。

由假设(4)知：$0\leqslant X(i,j)\leqslant C(i,j)$，即实际流量为非负且小于或等于线路的容量。因此其数学规划模型为

$$\max \quad f$$

$$\text{s.t.} \quad \sum_{j\in S(i)} X(i,j) - \sum_{j\in R(i)} X(j,i) = \begin{cases} f, & \text{当 } i = S \text{ 时} \\ 0, & \text{当 } i \neq S, i \neq T \text{ 时} \\ -f, & \text{当 } i = T \text{ 时} \end{cases}$$

$$X(i,j)\leqslant C(i,j)$$

$$X(i,j)\geqslant 0$$

该问题用单纯形法可得到最优解，故最大流问题本质上是线性规划问题，但当网络比较大时，约束方程的数量和变量的数量很多，线性规划模型很大，求解很不方便，用图的特性可使问题的求解变得容易。

8.4.2.2 割集及割集容量

设 G 是一连通图，满足下列条件的边的集合叫割集：

(1) 把这些边全部从图 G 中去掉，图 G 则成为不连通的两部分；

(2) 把该边集合的任何子集从图 G 中去掉，图 G 仍是连通的。

现将割集中各支路容量的和称为割集容量。

8.4.2.3 最大流—最小割定理

定理：网络中从源点 S 到汇点 T 的最大流量等于把 S 和 T 分开的最小的割集容量。

这个定理的意义是明确的。一般说来，从 S 到 T 的净流量，也就是由割集分开的两个集合 S 和 T 之间的净流量

$$f=\sum_{i\in S,j\in S}X(i,j)-\sum_{i\in S,j\in\bar{S}}X(j,i)$$

式中 S——S 点所在的集合；

$\bar{S}$——T 点所在的集合。

由于 $$0\leqslant X(i,j)\leqslant C(i,j),f>0,$$

所以 $$f=\sum_{i\in S,j\in S}X(i,j)-\sum_{i\in S,j\in\bar{S}}X(j,i)\leqslant\sum_{i\in S,j\in\bar{S}}C(i,j) \tag{8.4.2}$$

显然只有当 $\sum C(i,j)=f$ 时，才能取极大值。也就是只有当 $\sum_{i\in S,j\in\bar{S}}X(j,i)=0$，且 $X(i,j)=C(i,j)$时，才能使 $f=\sum_{i\in S,j\in S}C(i,j)$。

由上述分析可见，只有 S 集到$\bar{S}$ 集中各支路的流$X(i,j)$饱和，同时由 $\bar{S}$ 集到S 集支路流量为零时，网络才能达到最大容量。

上述推理为处理最大流问题打开了思路并提供了方法。

8.4.3 应用“最大流—最小割定理”的穷举法

现研究图 8.4.1 的网络。应用“最大流—最小割定理”，找出所有把 S 与 T 分开的割集，并计算其割集容量见表 8.4.1。

表 8.4.1 割集容量

s	$\bar{S}$	割 集	割集容量
s	2,3,4,5,6,T	$(s,2),(s,3)$	22
s,2	3,4,5,6,T	$(s,3),(2,4),(2,5)$	20
s,3	2,4,5,6,T	$(s,2),(3,4),(3,2),(3,6)$	27
s,2,3	4,5,6,T	$(2,5),(2,4),(3,4),(3,6)$	21
s,2,3,4	5,6,T	$(2,5),(4,5),(4,6),(4,T),(3,6)$	23
s,3,6	2,4,5,T	$(S,2),(3,2),(3,4),(6,T)$	32
s,2,3,4,5,6	T	$(4,T),(5,T),(6,T)$	23
s,2,4,5	3,6,T	$(s,3),(4,6),(4,T),(5,T)$	32
s,3,4,5	2,5,T	$(s,2),(3,2),(4,5),(4,T),(6,T)$	36
s,2,5	3,4,6,T	$(s,2),(2,4),(5,T)$	24
s,2,3,4,5	6,T	$(3,6),(4,6),(4,T),(5,T)$	22

这里值得注意的是，有向图的割集是有向的，只有从 S 指向$\bar{S}$ 的支路饱和，$\bar{S}$ 指向S 的支路为零时的流量才是最大的流量。因此表中令 $\bar{S}$ 由指向S 的支路为零，故由 $\bar{S}$ 到S 的支路都未列在表 8.4.1 的割集之中。

由表可见：

(1) 最小的割集容量为 20，即该网络系统的最大流量为 20。

(2) 欲提高网络的通过能力，其关键在于$(s,3),(2,4),(2,5)$三条支路，如不提高这

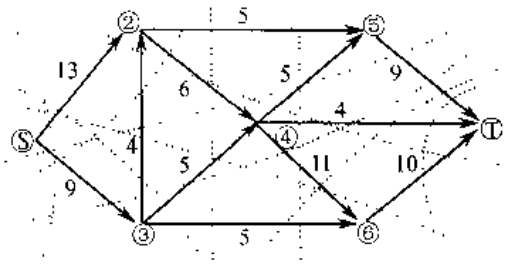


图 8.4.1 网络图

三条支路的允许通过能力，由 S 到 T 的流量是不会改变的，这三条支路就像瓶颈一样限制了通过流量的大小。因此，最大流问题不只帮助我们确定通过的最大流量，而且还为我们提出了提高通过能力的途径，达到使用有限的资源收到最大效果的目的。

8.4.4 标记法

1. 步骤

采用穷举法处理最大流问题虽然准确，但太繁。因此用上述推导得出的结论构造出标记法。标记法先假设网络各支路中有一满足流量守恒方程的初始流量，该初始流量叫“可行流”。可行流的选择方法有 2 种：①为了计算方便，可将可行流全部选为零，因为零也是符合流量守恒方程的；②为了加快计算速度，通常把可行流选择成使其中某些主要支路饱和。在该可行流的基础上，再试探着对其进行扩充，一次一次地试探，一次次地扩充，直到再无法扩充为止，就找到了该网络的最大流。其步骤如下：

第一步 先给源点以标记$(-,\infty)$，其中“$-$”表示开始，“∞”依据假设条件，表示进入 S 的流量可无穷大，这时 S 成为已标记未检查的点。

第二步 用已标记未检查的点 i 对有联系的未标记的点 j 进行检查。

(1) 首先检查从已标记点出发的支路：

① 如果初始流量小于最大容量，即 $X(i,j)<C(i,j)$，则 j 点标$(i^+,\varepsilon(j))$，$(i^+,\varepsilon(j))$表示可以由 i 点向 j 点增加的流量为$\varepsilon(j)$，其中 $\varepsilon(j)=\min\{\varepsilon(i),\ C(i,j)-X(i,j)\}$；$\varepsilon(j)$表示 i 点能够向 j 点提供的数量和支路能进一步扩充的流量中最小的一个，如图 8.4.2(a)所示。

② 如果 $X(i,j)=C(i,j)$，则不标记。完成上述检查后，由 i 点出发的支路检查完

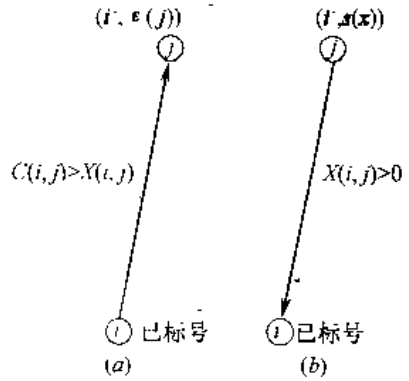


图 8.4.2 标记法

毕，j 点成为已标记未检查的点。

(2) 检查流向节点的支路；

① 如果 $X(i,j)>0$，则 j 标记$(i^-,\varepsilon(j))$，表示 i 向 j 可以少提供的流量为$\varepsilon(j)$，实质上也等于 j 向 i 多提供$\varepsilon(j)$，其中 $\varepsilon(j)=\min\{\varepsilon(i), X(i,j)\}$，$\varepsilon(j)$表示 i 可以少提供的数量和流量中最小的一个(图 8.4.2(b))。

② 当 $\varepsilon(i)<X(j,i)$，则说明 i 点不能提供足够流量使 $X(j,i)=0$，因此还要有一部分流量流向 i。

③ $X(i,j)=0$ 不标记。

通过对 i 点的检查，使 j 点成为已标记、未检查的点，而 i 点则由已标记未检查的点变为已标记、已检查的点。这样一直进行到 T 点也标记$(n^+,\varepsilon(T))$为止。

第三步 从 T 反向查找可以扩充的路径，并确定扩充量。其方法是由 T 点的标记$(n^+,\varepsilon(T))$，倒推到 n 点，再按 n 点的标记倒推到另一点 $n-1$，最后一直倒推到源点 S 为止。就可找到一条由 S 到 T 的可以扩充的路径。在此扩充线路上标记相应的值$\varepsilon(j)$，并选 $\varepsilon^*(j)=\min\{\varepsilon(j)\}$作为扩充量。

第四步 根据扩充路径上标记的扩充流量。对支路(i,j)扩充时，如果 j 的标记为 i^+，初始流量 $X(i,j)$就增加 $\varepsilon^*(j)$；如果标记为 i^-，初始流量 $X(i,j)$就减少 $\varepsilon^*(j)$，这就完成了一次扩充，得到一个新的可行流。

第五步 重复以上步骤，一直到无扩充路径为止，即得到该网络的最大流。

2. 举例说明

某工地需大量砾石，采石场与工地间的公路网及运输能力如图 8.4.3 所示，现需确定该公路网可运输砾石的最大数量，以便安排运输计划。

采用标记法，假设已预先给定初始可行流并用括号标在线路旁。

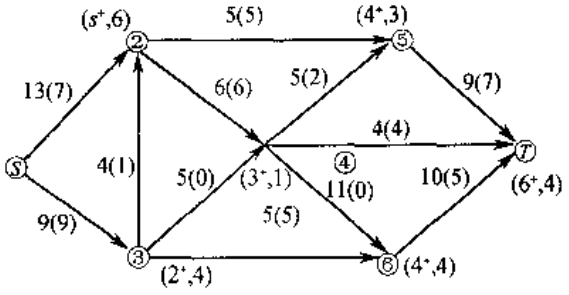


图 8.4.3 公路网及运输能力

第一步 S 点标记$(-,\infty)$。

第二步 对已标记、未检查的点 S 进行检查。先看由 i 出发的支路$(S,2)$、$(S,3)$。因 $C(S,2)=13$，$X(S,2)=7$，$C(S,2)>X(S,2)$，则 2 点应标记$(S^+,\varepsilon(2))$，$\varepsilon(2)=\min\{\infty,C(S,2)-X(S,2)\}=\min\{\infty,6\}=6$，所以 2 点标记$(S^+,6)$，它表明由 S 可增加 6 单位流量到点 2。

$C(S,3)=9$，$X(S,3)=9$ 为饱和支路；无指向 S 点的支路，故 S 点已检查完，成为已标记、已检查的点。

点 2 成为已标记未检查的点，对其进行检查。先看由 2 点出发的两条支路(2,4)、(2,5)，$C(2,5)=X(2,5)$，$C(2,4)=X(2,4)$，均为饱合支路；再看流向点 2 的支路(3,2)，因

$X(3,2)>0$，$\varepsilon(3)=\min\{\varepsilon(2),X(3,2)\}=\min\{6,4\}=4$，所以点 3 标记$(2^-,4)$，表明点 2 可以向点 3 提供 4，或点 3 可以少向点 2 提供 4 个单位。2 点成为已标记、已检查的点，点 3 成为已标记未检查的点。

再继续按第二步检查点 3，支路(3,4)中，$C(3,4)=5$，$X(3,4)=0$，$C(3,4)>X(3,4)$；支路(3,6)中，$C(3,6)=X(3,6)$为饱合的，故点 3 检查完毕，点 4 又成为已标记未检查的点，其标记为$(3^+,4)$。

对点 4 进行检查：

$C(4,5)=5$，$X(4,5)=2$，$C(4,5)>X(4,5)$，$\varepsilon(5)=\min\{4,5-2\}=3$，5 点标记为$(4^+,3)$；$C(4,6)=4$，$X(4,6)=0$，$C(4,6)>X(4,6)$，$\varepsilon(6)=\min\{4,4\}=4$，点 6 标记为$(4^+,4)$，$C(4,T)=4$，$X(4,T)=4$，属饱和支路，点 4 检查完毕。

对点 5 进行检查：

$C(5,T)=9$，$X(5,T)=7$，$C(5,T)>X(5,T)$。$\varepsilon(T)=\min\{2,9-7\}=2$，$T$ 点标记$(5^+,2)$。

这样，即可确定扩充路径。但点 6 尚未检查，且它也与 T 有关，因此需在检查完点 6 之后，再确定 T 点的标记。

对点 6 进行检查：

$C(6,T)=10$，$X(6,T)=5$，$C(6,T)>X(6,T)$，$\varepsilon(T)=\min\{4,10-5\}=4$，$T$ 应标记$(6^+,4)$。

此时 T 点出现了两个标记，这两个标记说明，如果从点 5 扩充，可使 T 增加的流量数为 2，而从 6 点扩充，可使 T 增加 4，我们当然希望 T 点增加的越多越好，因此 T 的标记应取$(6^+,4)$。

第三步 寻找扩充路径并确定扩充量：

$T\to6\to4\to3\to2\to S$ 为扩充线路并按实际的箭头方向标出，在箭线上标出扩充值如图 8.4.4 所示。

由图 8.4.4 可见，虽然 S 向 2 点可以提供的增量为 6，但由于以后的线路只允许通过 4 个单位砾石，因此只能在所有可能的扩充量中选一个最小的量进行扩充，即 $\varepsilon^*(j)=\min\{\varepsilon(j)\}=\min\{6,4,4,4,4,\}=4$。

把扩充后的数字填在图上，再按上述步骤重新开始扩充，直到进行不下去为止，即得到最大流，如图 8.4.5 所示。

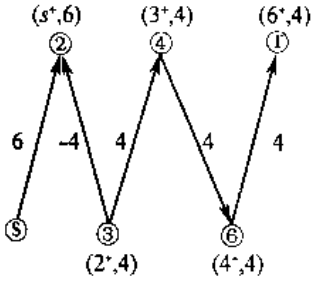


图 8.4.4 扩充值

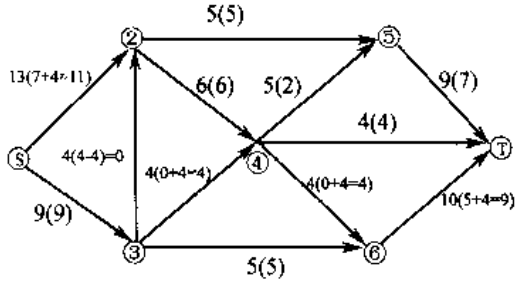


图 8.4.5 最大流

该公路网络的最大流为源点发出 11+9=20，汇点收到 9+4+7=20。该采石场每天只能运出 20 单位砾石。如欲提高公路运输能力，因最后计算是在点 2 停止的，故由点 2 出发的路是“瓶颈”支路。

8.4.5 几种特殊情况的处理

8.4.5.1 多源点和多汇点的处理

在实践中经常遇到多个源点和多个汇点的情况，如采石场有很多个，而需砾石的单位并非一家。对于这种多源点和多汇点的情况，只要经过适当处理，也可用上述方法解决。

通常采用引入超源点和超汇点的办法来解决这类问题。超源点就是虚拟一个源点，使它与所有的源点相连；超汇点就是虚拟一个汇点，使它与所有的汇点相连。这就构成了单源点、单汇点的情况，这样就可采用上述方法确定最大流量了。

现举例说明多源点、多汇点的处理方法。

甲、乙、丙三个城市，工厂区在丙城，工人上班从甲城到丙城可走两条路，一条是从甲城到丙城，一条是从甲城经乙城到丙城。每天由早 7:00 至 9:00 为运输高峰，每个工人不早于 7:00 离开甲城，不晚于 9:00 到达丙城，车辆平均行驶时间为：甲到乙 60min，甲到丙 60min，乙到丙 30min，在甲、乙二城规定的发车时间间隔为 30min，每次运输工人数：甲到丙 5000 人，甲到乙 2000 人，乙到丙 3000 人，甲、乙、丙城有任意多的车辆，求在 7:00 到 9:00之间到达丙城的人数最多是多少？该问题可抽象成图 8.4.6。

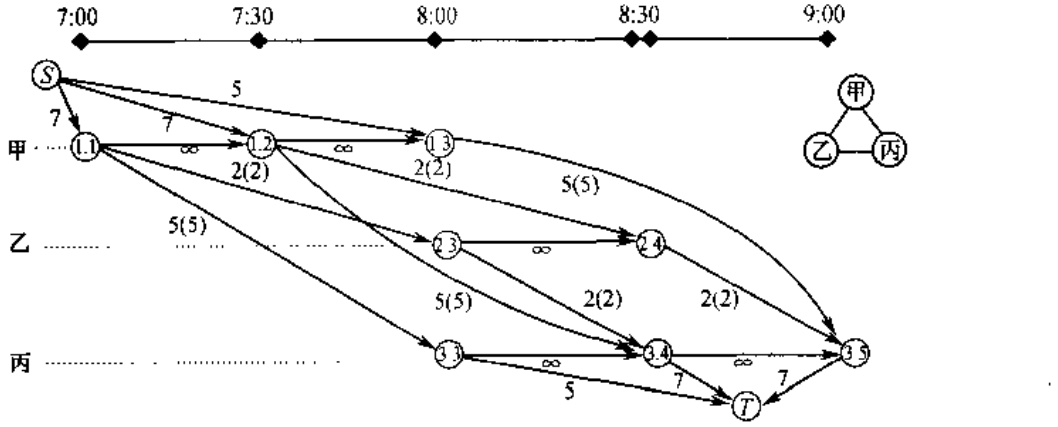


图 8.4.6 甲、乙、丙三个城市线路网络模型

由图 8.4.6 的网络模型可见一共能有 19000 人在 9:00 前到达丙城。引入一个超源点 S 和超汇点 T，并假设工人是由 S 点出发到 T 点终止，就解决了该问题。

从求最大流的过程中可以看到，如果每条弧的容量为整数，并且可行流也是整数，那么求得的最大流也一定是整数。由于最大流的这一性质，可把某些要求解为整数的实际问题或理论问题，化成最大流问题处理。如运输量不以吨计算，而是以船、车去考虑，则应用最大流方法可得到一个最大的整数解。

8.4.5.2 无向网络的处理

以上研究的是有向图的最大流问题。在实际问题中，尚存在着无向网络的最大流问题。

对于无向网络，同样可以应用上述方法，只不过需对无向网络做些规定。

对于无向图，每一条边(i,j)都有一容量$C(i,j)=C(j,i)$，该容量可认为是满足条件：$0\leqslant X(i,j)\leqslant C(i,j)$的由$i$到$j$的流量$X(i,j)$，也可以是满足条件$0\leqslant X(j,i)\leqslant C(j,i)$的从$j$到$i$的流量。因此，可把一条边看成两个方向相反的弧$(i,j)$及$(j,i)$，并令其容量均为$C(i,j)$，如图8.4.7所示。

这样把一无向网络化成有向网络后，就可应用最大流的解法求解了。但在解的过程中要注意变化后的可行流，在某一对弧(i,j)与(j,i)上，可能不满足$X(i,j)+X(j,i)\leqslant C(i,j)$，这是由于$X(i,j)\geqslant 0, X(j,i)\geqslant 0$，也就是$(i,j)$这条弧上出现了对流的缘故，所以在替换后的有向网络上求出最大流，然后消灭它在某些边上的对流现象，就得到原来无向网络的最大流。

8.4.5.3 顶点有容量网络的处理

网络中的一些顶点具有容量，即点a_i具有一个容量$C(a_i)$，要求流进a_i的总容量不超过$C(a_i)$。例如码头、车站、仓库的存储能力和吞吐能力有限等。这类点具有容量的问题可以化成弧具有容量的问题去解决。具体做法是：把点a_i变成两个点a'_i和a''_i，并增加一段弧(a'_i,a''_i)，令其容量为$C(a_i)=C(a'_i,a''_i)$。而把指向a_i的弧变为指向a'_i，把由a_i出发的弧改为由出发(见图8.4.8)。然后再按上述方法求解。

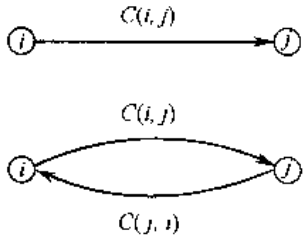


图8.4.7 无向网络的处理

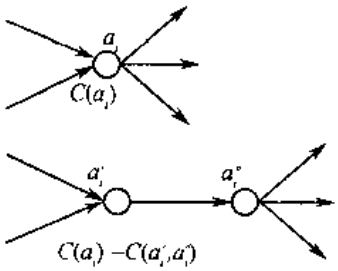


图8.4.8 顶点有容量网络的处理

8.5 最小费用最大流

在最大流问题的基础上，如果考虑每条弧上的运输费用，则提出最小费用最大流问题。该问题的数学表达式为

约束条件：

$$0\leqslant X(i,j)\leqslant C(i,j)$$

$$\sum_{j\in S(i)}X(i,j)-\sum_{j\in R(i)}X(j,i)=\begin{cases}-f, & i=S\\ 0, & i\neq S, i\neq T\\ f, & i=T\end{cases}$$

$$f=f^*, f^* \text{ 为最大流}$$

使

$$\sum_i\sum_j d(i,j)\cdot X(i,j)\to \min$$

现应用最短路径法求最小费用最大流。

该方法的基本思想是：当网络有最大流，且有若干个方案时，如果把网络的弧长用单位费用表示，则应用最短路径法可求出由S到T运输单位物资费用最少的一条链，

并得到运输单位物资到各顶点时的最少费用。按最小费用链取最大流可得一初始方案。

当流量增加时，为了使费用增加最少，需对初始方案进行扩充。假设一个单位的流量在初始方案中由 S 到 T 只能流过一条非饱和链，故需在所有非饱和链中选择一条费用增加最少的链作为对初始方案的扩充。

其作法为：在初始方案的基础上，构造一个新网络，该网络的顶点仍为原网络的各顶点，而每条弧(i,j)都变成方向相反的两条弧，弧长定义为

$$l(i,j)=\begin{cases} d(i,j), & \text{当 } X_{i,j}^{(k-1)} < C(i,j) \text{ 时} \\ \infty, & \text{当 } X_{(i,j)}^{k-1} = C(i,j) \text{ 时} \end{cases}$$

$$l(j,i)=\begin{cases} -d(i,j), & \text{当 } X_{i,j}^{(k-1)} > 0 \text{ 时} \\ \infty & \text{当 } X_{i,j}^{(k-1)} = 0 \text{ 时} \end{cases} \tag{8.5.1}$$

上式的含义是如果弧上有流量且饱和，则因其不能再有流量通过，故去掉正向弧，因其可减少流量，故应加上反向弧；如不饱和，因其尚可通过流量或减少流量，故应保留正向弧，加上反向弧；如果无流量，因流量不能减少，则不必加反向弧，只保留正向弧。

对此新网络用最短路径法寻找最短路径就是一条费用增加最小的链，使其饱和就完成了对初始方案的一次扩充，重复上述步骤，即可得到最小费用最大流。

现举例说明算法。

某工地与采砂场间道路的容许流量为 $C(i,j)$，单位运价为 $d(i,j)$，工地与采砂场之间的网络如图 8.5.1(a)所示(弧中左边的数字为 $d(i,j)$，右面的数字为 $C(i,j)$，现需确定怎样运输才能使运输的砂料最多且运费最省。

第一步 用最短路径法，以 $d(i,j)$为弧长确定由 S 到 T 的最短路径(最小费用链)并令其饱合，得初始流量(见图 8.5.1(b))。

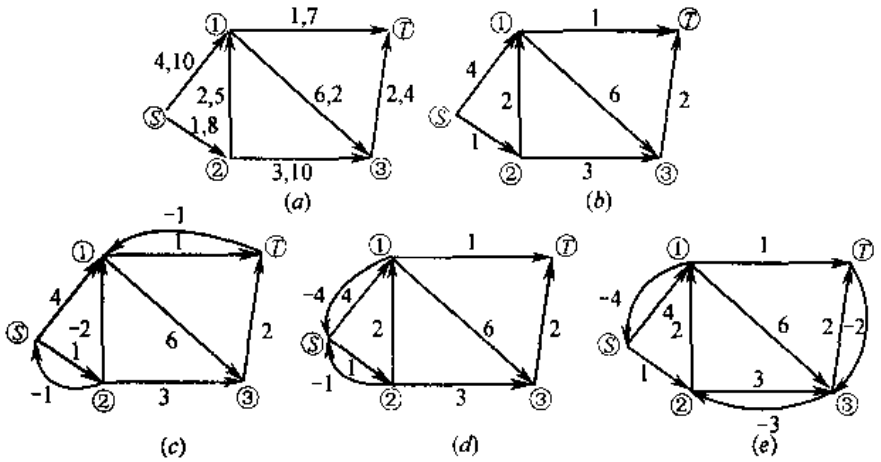
所得最小费用链为 $S-2-1-T$

流量为 $\min\{C(S,2);C(2,1);C(1,T)\}=5$

初始方案为 $X(S,2)=5$，$X(2,1)=5$，$X(1,T)=5$

第二步 按图 8.5.1(b)和式(8.5.1)形成新网络如图 8.5.1(c)所示。

第三步 对图 8.5.1(c)用最短路径法求出由 S 到 T 的最短路径，得出最小费用链。



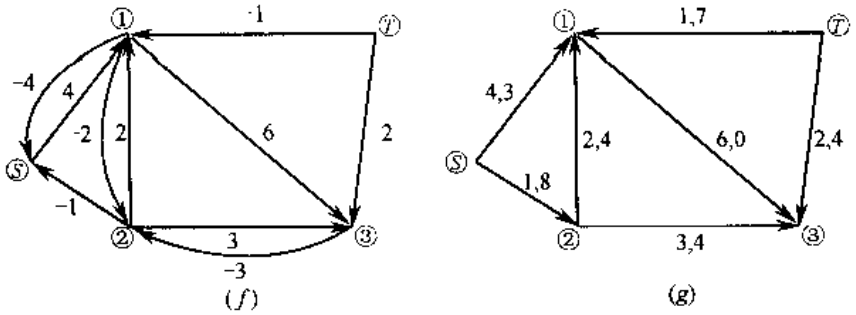


图 8.5.1　某工地与采砂场网络模型

最小费用链为

$$S-1-T$$

流量为

$$\min\{C(S,1),C(1,T)-X(1,T)\}=2$$

得新初始方案为

$$S(S,1)=2,X(S,2)=5,X(2,1)=5,X(1,T)=7$$

第四步　在此方案的基础上再按第二步的方法形成新网络(图 8.5.1(d)),最小费用链为

$$S-2-3-T$$

流量为

$$\min\{C(S,2)-X(S,2),C(2,S),C(3,T)\}=3$$

新初始方案为

$$X(S,1)=2,X(S,2)=8,X(2,1)=5,X(1,T)=7,X(2,3)=3,X(3,T)=3$$

第五步　在此方案基础上,得到新网络(图 8.5.1(e))。

最小费用链为 $S-1-2-3-T$,其中①—②为一反向流量,说明该段弧上的正向流量减少,才能得到最小费用链。

流量为

$\min\{C(S,1)-X(S,1),X(2,1),C(2,3)-X(2,3),C(3,T)-X(3,T)\}=1$。

初始方案为

$X(S,1)=3,X(S,2)=8,X(1,T)=7,X(2,3)=4,X(3,T)=4,X(2,1)=4$。

第六步　再按上述步骤进行,得新网络(见图 8.5.1(f)),由 S 到 T 的最短路径为 ∞,运算结束。

第七步　所得最小费用最大流,如图 8.5.1(g)所示。图中每条弧的左边为单位运价,右边为实际流量。

最大流量为

$$f^*=7+4=11$$

最小费用为

$$4\times3+1\times8+2\times4+1\times7+3\times4+2\times4=55$$

由计算可知,该采砂场共可运出砂料 11 单位,最小费用为 55。

第9章　系统试验和模拟

在系统决策之前，如果采用各种最优化方法，通过对系统进行抽象、建立数学模型和模型求解分析等过程得到最优决策，这诚然是最好的。然而实际并不是这样，各种最优化方法都有其局限性，各种定量化技术都有其不足，因此在决策之前采用一些补充方法可为决策提供更多的信息。

各种定量化技术的局限性表现在以下几个方面：

(1) 各种定量化技术都有假设，实际系统往往与假设都有一定的差距，致使系统的实际运行与模型求解结果不同。

(2) 很多影响系统的因素具有随机性，致使其准确定量产生困难。虽然概率论和数理统计为我们提供了一套科学的方法，但在随机因素的影响下，系统的决策可能失误。

(3) 影响系统的因素中有一部分定性因子，它们无法或很难准确定量，因此必须对定量化的结果进行讨论和分析，这种讨论和分析有时带有主观性，致使决策失误。

(4) 影响系统的因素很多，为了清楚地反映系统的状况，不得不忽略掉某些“所谓”次要因素来建立模型，有时有些系统受某些潜在因素的影响，而这些潜在影响并未被人类所认识，所建模型也未考虑，从而导致模型未反映系统的本质特征，造成决策失误。

(5) 有些数学模型很复杂，求解很困难，只能采取一些近似或简化的解法，影响求解精度。

鉴于定量化方法有上述局限性，仅采用某些定量化的技术作为系统决策的依据是不够的，还必须在定量方法的基础上对系统进行实际试验。如何用最少的试验次数，取得最好的试验成果，则是系统试验研究的内容。

对于不允许试验或试验需很高代价的系统，只能借助于电子计算机采用现代化的模拟技术，使系统运行，以揭示实际系统的运行规律，为决策提供更充足的依据。如何建立这种模型并在计算机上实现其实际运行是系统模拟所需解决的问题。

总之，在系统决策之前应综合分析最优化模型的求解结果和试验、模拟结果才能使决策的依据更充分，因此系统试验和模拟就成为系统决策的重要依据之一，试验和模拟方法也就成为系统工程的常用方法之一。

试验和模拟是两类性质的工作。系统试验是指在缩小的规模和范围内对影响系统运行的内、外因素进行分析、测定，从而得出系统变化规律、确定系统满意运行方案的一种方法。系统模拟一般指对那些不能进行实际试验或即使能进行试验，但需大量资金、时间或可能造成严重后果的系统，通过建立模拟模型，在计算机上体现系统运行状况、反映系统运行规律，为系统决策提供依据的一种方法。

9.1 系统的试验设计和分析

把系统试验工作作为系统来研究，就必须以系统思想为指导，采用系统工程的方法进行。其中对系统试验进行规划设计和分析是最重要的工作之一。

对系统试验进行规划设计，取决于试验的目的。一般情况下，确定各要素对系统目标影响的显著性、确定系统运行的最佳方案则是系统试验的主要目的。以该目的为出发点的系统试验方法是正交设计。

正交设计是利用一套规格化的正交表来科学地挑选试验条件，合理地安排试验，从中推断出系统最佳运行条件的一种方法，同时通过正交设计的分析尚可提出比试验结果好得多的信息，为系统的决策提供方便的条件和依据。

9.1.1 正交设计

9.1.1.1 正交设计的基本术语

正交设计中常用的术语有：指标、因子和水平。

正交设计把试验设计要考虑的结果和评价准则称为指标，一般以 y_i 表示第 i 次试验的指标值，$\bar{y}$ 表示总平均指标；把对试验结果和对评价指标可能产生影响且在试验中明确了条件加以对比的因素称为因子，一般以大写字母表示；把每个因子在试验中的具体条件称为因子的水平，简称水平，一般以表示因子的大写字母加上脚标来表示。

如在研究城市绿化系统经济效益的时候，需要确定最佳城防林营造方案。据资料分析，影响城防林经济效益的因素有树高、林带宽、树种组成。现需通过调查或试验确定该三个因素是否都对城防林经济效益有显著影响，它们怎样组合才能使所营造的城防林经济效益最佳。选择的试验或调查条件为：树高：10m、15m、20m；林带宽：5m、10m、15m；树种组成：纯阔叶林、纯针叶林、针阔混交林。在这种情况下，城防林经济效益的大小就是指标；树高、林带宽、树种组成为因子；树高的三个条件：10m、15m、20m 为该因子的三个水平；林带宽的试验条件：5m、10m、15m 为该因子的三个水平；树种组成的三种情况：纯阔叶林、纯针叶林、针阔混交林为该因子的三个水平。

9.1.1.2 正交表

正交试验是按正交表进行试验设计的。正交表的符号为

$$L_n(K^j)$$

式中 L——正交表符号；

n——试验次数；

K——最高水平数；

j——列数。

正交表是根据任一列所含各水平的个数和任意两列可能出现的数对数相同的原则编制的，$L_9(3^4)$和 $L_8(2^7)$两个正交表见表 9.1.1 和表 9.1.2。

上例可选 $L_9(3^4)$正交表，即采用三种水平四列进行九次试验的正交表。

在选择正交表时，一般采取 $j>m$ 的原则（m 为因子个数），其原因是在很多情况下，

表 9.1.1 $L_8(2^7)$正交表

试验号 \ 列号	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

表 9.1.2 $L_9(3^4)$正交表

试验号 \ 列号	1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

注:任意两列的交互作用出现于另外二列。

各因子之间有相互影响,而正交表中有反映这种交互作用的列。

每张正交表都有一张交互作用列表,该表是确定正交表中体现各因子交互作用对指标产生影响的列的表格。表 9.1.3 是 $L_8(2^7)$ 正交表的交互作用列表。表中第一行第二列的元素"3",是指第一列因子和第二列因子对指标的交互作用体现在第三列,也就是把第一、二列因子的交互作用当作一个因子来考虑,且该因子的位置应在第三列。这样在安排正交设计表头时,第一列为 A、第二列为 B、第三列则为 A、B 两因子的交互作用 $A\times B$,而不应该安排 C 因子。因此,在考虑各因子的交互作用时,所选正交表的列数必须大于所分析因子数。

从正交表的结构看,用正交设计进行的试验安排具有均衡分散性和整齐可比性,因此可以用最少的试验次数得到各因子各水平合理组合的最佳方案。

如上例中,我们选 $L_9(3^4)$ 正交表,只进行九次试验,就能得到三因子的三个水平组合成的 27 个方案的信息,通过分析还可得出比简单比较 27 个方案更多的信息,因此正交

表 9.1.3 $L_8(2^7)$交互作用列表

列号 / 列号	1	2	3	4	5	6	7
	(1)	3	2	5	4	7	6
		(2)	1	6	7	4	5
			(3)	7	6	5	4
				(4)	1	2	3
					(5)	3	2
						(6)	1

设计可以最少的“投入”得到最多的“产出”，是一种效益较高的试验方法。

正交设计除可进行试验的安排之外，在可能的条件下尚可用于调查研究和资料收集的设计。如在确定城市绿化系统经济效益时，对已营造的各种类型城防林进行调查，如有可能，按正交设计的方法进行调查选点，可减少大量的调查工作量和数据处理工作量。

当正交表选定之后，为了消除系统误差，各因子的水平号应作随机化处理，如采用抽签法确定各水平的对应号，而后按正交表规定的试验号进行试验。

上例中假设三个因子无交互作用，且抽签号与安排的顺序相同，则可按表 9.1.4 所列的试验号进行试验。试验结果的指标值列于最末一栏。

表 9.1.4 试验结果

水平 / 列号 / 试验号	A	B	C	D	效益 y_i/（百元/亩）
1	10	5	阔		1.72
2	10	10	针		1.82
3	10	15	混		1.80
4	15	5	阔		1.62
5	15	10	针		1.83
6	16	15	混		1.98
7	20	5	阔		1.59
8	20	10	针		1.60
9	20	15	混		1.81
注：数据是假设的。					

9.1.2 正交设计分析

9.1.2.1 正交设计分析程序

正交设计分析是从正交表的试验数据出发来分析各因子水平对指标的影响，从而确定最佳组合方案的过程。其分析程序如图 9.1.1 所示。对上例经初步分析可得表9.1.5。

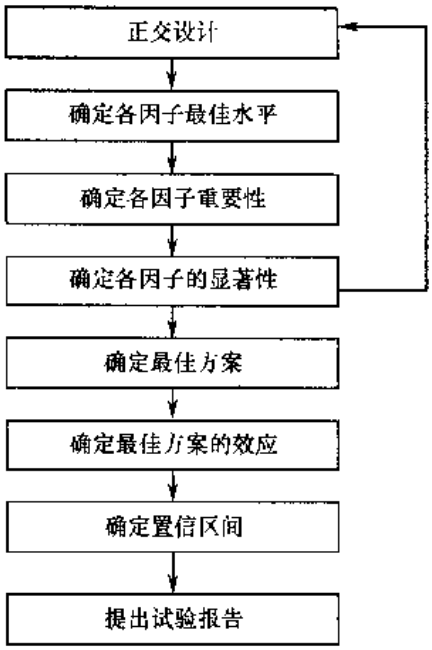


图 9.1.1 正交设计分析程序

表 9.1.5 初步分析结果

列号 / 水平 / 试验号	A(树高)	B(带宽)	C(主要树种)	D	效益 y_i/(百元/亩)
1	10	5	阔		$Y_1=1.72$
2	10	10	针		$Y_2=1.82$
3	10	15	混		$Y_3=1.80$
4	15	5	阔		$Y_4=1.92$
5	15	10	针		$Y5=1.83$
6	15	15	混		$Y_6=1.98$
7	20	5	阔		$Y_7=1.59$
8	20	10	针		$Y_8=1.60$
9	20	15	混		$Y_9=1.81$
M_1	$M_{11}=5.34$	$M_{12}=5.23$	$M_{13}=5.30$	$M_{14}=5.36$	$\sum_i Y_i = 16.07$
M_2	$M_{21}=5.73$	$M_{22}=5.25$	$M_{23}=5.55$	$M_{24}=5.39$	
M_3	$M_{31}=5.00$	$M_{32}=5.59$	$M_{33}=5.22$	$M_{34}=5.32$	$\bar{Y} = 1.786$
$\bar{y}_1$	$\bar{y}_{11}=1.78$	$\bar{y}_{12}=1.743$	$\bar{y}_{13}=1.767$	$\bar{y}_{14}=1.787$	
$\bar{y}_2$	$\bar{y}_{21}=1.91$	$\bar{y}_{22}=1.750$	$\bar{y}_{23}=1.850$	$\bar{y}_{24}=1.797$	
$\bar{y}_3$	$\bar{y}_{31}=1.69$	$\bar{y}_{32}=1.863$	$\bar{y}_{33}=1.740$	$\bar{y}_{34}=1.773$	
极差 R_1	$R_1=0.72$	$R_2=0.36$	$R_3=0.33$	$R_4=0.07$	
方差 S_1	$S_1=0.08896$	$S_2=0.02729$	$S_3=0.01976$	$S_4=0.00082$	$S_{总}=0\ 13682$

表中数据 M_{ij}、Y_{ij}、R_j、S_j 是在表 9.1.4 试验数据的基础上分析得到的。

现以城市绿化系统经济效益试验和调查的资料为例说明正交设计的分析过程和原理。

9.1.2.2 正交设计分析

1. 因子最佳水平的确定

因子最佳水平是指每个因子的各水平中使指标达最佳的水平。为确定因子的最佳水

平，必须确定该因子各水平对指标的影响。为了排除其他因子的影响，采用分类的方法。如确定树高因子的最佳水平，可将九次试验分为三组，分组情况见表9.1.6。

表9.1.6 分组情况

因素 / 试验号	A	B	C	效益平均值
1、2、3	全为 A_1(10m)	B_1、B_2、B_3 各1次	C_1、C_2、C_3 各1次	$\bar{y}_{11}=1.78$
4、5、6	全为 A_2(15m)	B_1、B_2、B_3 各1次	C_1、C_2、C_3 各1次	$\bar{y}_{21}=1.91$
7、8、9	全为 A_3(20m)	B_1、B_2、B_3 各1次	C_1、C_2、C_3 各1次	$y_{31}=1.69$

由表可见，这三组试验数据体现了树高因子(A)对指标的影响，而排除了林带宽和树种组成的影响，这是由于林带宽和树种组成对树高各种水平的影响均相等的缘故。

设第 i 次试验的指标值为 Y_i，第 k 因子第 J 水平的指标平均值为 $\bar{y}_{jk}$，第 k 因子的水平数为 W_k，则

$$\bar{y}_{jk}=\frac{\sum_{i\in k,j}Y_i}{W_k}=\frac{M_{jk}}{W_k}\qquad j=1,2,\cdots,w_j \tag{9.1.1}$$

式中 M_{jk}为 k 因子 j 水平的指标合计值，$M_{jk}=\sum_{i\in k,j}Y_i$。

本例中：$\bar{y}_{11}=\frac{1}{3}(1.72+1.82+1.80)=1.78$

$\bar{y}_{21}=\frac{1}{3}(1.92+1.83+1.98)=1.91$

$\bar{y}_{31}=\frac{1}{3}(1.59+1.60+1.87)=1.69$

$\bar{y}_{21}$最大说明排除林带宽和树种组成的影响后，三种树高中，树高为15m的 A_2 水平效果最好，该因子的最佳水平为 A_2 水平。其计算可按下式进行：

$$\bar{y}_{jk}^{*}=\max\{\bar{y}_{jk}\} \tag{9.1.2}$$

式中 $\bar{y}_{jk}^{*}$——k 因子 j 水平的平均指标值，k 因子 j 水平为最佳水平。

按式(9.1.1)，同理可确定其他各因子各水平的平均值及最佳水平，计算结果见表9.1.7。

表9.1.7 各因子各水平的指标平均值

指标平均值 / 因子 / 各水平平均值	A	B	C	D
$\bar{y}_1$	$\bar{y}_{11}=1.78$	$y_{12}=1.743$	$\bar{y}_{13}=1.764$	$\bar{y}_{14}=1.787$
$\bar{y}_2$	$\bar{y}_{21}=1.91$	$\bar{y}_{22}=1.75$	$\bar{y}_{23}=1.850$	$\bar{y}_{24}=1.797$
$\bar{y}_3$	$\bar{y}_{31}=1.667$	$y_{32}=1.863$	$\bar{y}_{33}=1.740$	$y_{34}=1.773$

由表可见，表中每列的最大值即是对应最佳水平的指标平均值。

2. 因子重要性分析

各因子对指标的影响是不同的，其重要性也各不相同。为了评价各因子的重要性，需

拟定一评价指标。通常采用离差平方 S 和或极差 R 作为评价指标。

由表 9.1.7 中每列的数字可见，各数据间的差距越大，说明该因子各水平相差悬殊，对指标的影响大，反之则小。因此以离差平方和或极差可粗略的揭示出各因子的重要性。

计算 k 因子离差平方和时需考虑其水平数，其计算公式为

$$S_k = W_k \sum_{j}^{W_k} (\bar{y}_{jk} - \bar{y})^2 \quad (k = 1,2,\cdots) \tag{9.1.2}$$

式中 S_k——k 因子的离差平方和；

W_k——因子的水平数；

$\bar{y}$——总平均指标值，$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$。

k 因子的极差按下式计算：

$$R_k = \max\{M_{jk}\} - \min\{M_{jk}\} \tag{9.1.3}$$

式中 M_{jk}——k 因子 j 水平的指标合计值；

$$M_{jk} = \sum_{i \in km_j} y_i$$

本例中 S_k 和 R_k 的计算结果见表 9.1.8。

表 9.1.8 S_k 和 R_k 计算结果

因子 / R_k、S_k	A	B	C	D
R_k	0.73	0.36	0.33	0.07
S_k	0.08896	0.02729	0.01976	0.00082

由表可见 $S_A > S_B > S_C > S_D$，故树高最重要，林带宽次之，树种组成再次。

正交表的第四列同样可计算出离差平方和和极差平均值。该例的作用相当于我们把未考虑的因子列于第四列，如果该列的离差平方和很大，说明在分析影响指标的因子时，忽略了有更大影响的因子。由此可见，正交设计考虑了我们所未考虑到的事情，可为我们提供更重要的信息。如果该列离差平方和很小，则可认为是随机误差平方和。

3. 因子显著性的检验

因子的重要性只说明该因子相对其他因子的重要程度，而未说明该因子对指标影响的显著程度。如果某因子对指标的作用不显著，则可排除该因子而使决策简化。

应用正交表进行显著性检验比较容易。因为试验的总离差平方和可分解为因子的离差平方和和随机误差平方和，由于总离差平方和为 $S = \sum_k S_k$，且每个因子均有 r 个水平，故自由度为 $r-1$。

设 S_k 相对独立，则当因子作用不显著时，

$$F = \frac{S_k / f_k}{S_{误} / f_{误}} \sim F(S_k, f_{误}) \tag{9.1.5}$$

式中 S_k——k 因子离差平方和；

$S_{误}$——误差的离差平方和；

f_k——k 因子的自由度；

$f_{误}$——误差的自由度。

当计算出的 F 值大于临界值 $F_{1-\alpha}(S_k, f_{误})$ 时，k 因子在 α 水平下作用显著，否则作用不显著。

$S_{误}$ 为误差的离差平方和，在正交表中未列入因子的列的 S 值在相对小的情况下均可认为是 $S_{误}$，另外，正交表中已列入因子的列中 S 值小于误差列的 S 时，也应当作误差处理。因此，误差离差平方和和相应的自由度应按下式计算：

$$S'_{误} = \sum_{j \in S_{误} \cup S_k} S_j \qquad f'_{误} = \sum_{j \in S_{误} \cup S_k} f_j$$

$F = \dfrac{S_k / f_k}{S'_{误} / f'_{误}}$ 即可用作不考虑交互作用时的判断值，将其与临界值相比较即可确定各因子的显著性。上例显著性的 F 检验见表 9.1.9。

表 9.1.9　显著性的 F 检验

	S_k	f_k	F	显著性
S_1	0.08896	2	108.49	显著
S_2	0.02729	2	33.28	较显著
S_3	0.01976	2	24.10	较显著
S_4	0.00082	2	—	
$F_{0.95}(2,2)=19$　　$F_{0.90}(2,2)=91.01$				

由表可见，A、B、C 三因子都显著地影响城防林的经济效益，其中 A 因子有极显著的影响。

4. 确定最佳方案

经显著性检验之后，可确定对指标有显著影响的因子、排除对指标影响不显著的因子。在此基础之上可选择与确定最佳方案。

最佳方案的确定方法是选择对指标有显著影响的因子中的最佳水平，对于对指标无显著影响的因子可不考虑，或根据实际情况决定。

如上例中的三个因子均对经济效益有显著影响，故其最佳组合方案为树高 15m(水平)、林带宽 15m(水平)、树种组成为针叶林的城防林。故在条件可能的条件下，应尽可能营造这种类型的城防林，会收到较高的经济效益。

从上例中得到的最佳方案可见，该方案是未进行试验的方案。这说明，正交设计试验方法不仅能确定所试验方案的情况，而且还可通过试验数据的分析，推断出各因子水平组合中的最佳方案。本例中就是通过对九个方案试验数据的分析，推断出未进行试验的最佳方案，这正体现出正交试验的优越性。

5. 效应和平均效应

为了确定正交设计所得最佳方案的指标值，就必须确定显著因子水平的效应。所谓效应即是评价因子取各水平时的指标 $\bar{y}_{jk}$ 与指标总平均值之差，它反映了因子取某一水平时，相对于指标总平均值的增加量。其计算公式为

$$a_{jk} = \bar{y}_{jk} - y \tag{9.1.6}$$

本例的效应矩阵见表 9.1.10，表中第一列的数字说明，树高取水平 1，可使经济效益

比平均水平降低 0.006 百元;取水平 2 可使经济效益提高 0.124 百元;取水平 3 可使经济效益降低 0.119 百元。其他列的涵义类推。

表 9.1.10 效应矩阵

因子 / 水平	A	B	C
水平 1	-0.006	-0.003	-0.019
水平 2	0.124	-0.036	0.064
水平 3	-0.119	0.077	-0.046

设 A 因子的各水平为 a_i,B 因子的各水平为 b_i,C 因子的各水平为 C_i。在不考虑交互作用的情况下,最佳方案的指标值可用平均效应或工程平均来确定,即用总平均值 $\bar{y}$ 加上显著因子所取定水平的效应。其计算公式为

$$\mu_{优} = \mu_a A^* B^* C^* = \bar{y} + a^* + b^* + c^* \tag{9.1.7}$$

式中 $\mu_{优}$——最佳方案的平均效应;

a^*, b^*, c^*——A、B、C 各因子的最佳水平所对应的效应。

本例中,最佳方案的工程平均为

$$\mu_{优} = \mu A_2 B_3 C_2 = y + a_2 + b_3 + c_2 = 1.786 + 0.124 + 0.077 + 0.064 = 2.051$$

即树高 15m、带宽 15m 的针叶城防林经济效益平均值为 2.051(百元/亩)。

置信区间按下式确定:

$$\delta = \sqrt{F_{1-\alpha}(1, \tilde{f}_{误}) \frac{\tilde{S}_{误}}{\tilde{f}_{误} \cdot n_e}} \tag{9.1.8}$$

式中 $\tilde{S}_{误}$——$S_{误}$ 与不显著因子自由度之和;

$\tilde{f}_{误}$——$f_{误}$ 与不显著因子自由度之和;

$F_{1-\alpha}(1, \tilde{f}_{误})$——$F$ 的临界值;

n_e——有效重复度,且 $n_e = \dfrac{\text{数据个数}}{1 + \text{显著因子自由度之和}}$。

本例中 $\tilde{S}_{误} = 0.00082$;$\tilde{f}_{误} = 2$;$F_{0.95}(1,2) = 200$;$n_e = \dfrac{9}{1+6} = \dfrac{9}{7}$;

$$\delta = \sqrt{200 \times \frac{0.00082 \times 7}{2 \times 9}} = 0.235$$

故在 $\alpha = 0.05$ 水平下置信区间为

$$(\mu - \delta, \mu + \delta) = (2.051 - 0.235, 2.051 + 0.235) = (1.798, 2.306)$$

即该方案的平均指标值在 1.798~2.306 的范围之内。

9.1.3 交互作用的显著性检验、最佳方案、平均效应及置信区间

上例所研究的问题是在树高、林带宽和树种组成之间无交互作用情况下的正交设计。但实际生产中往往各因子之间的交互作用对指标有很大影响,甚至有两上因素单独对指标无显著影响,但其交互作用却对指标有显著影响的情况,因此需研究这种情况下的正交设计。

在介绍正交表时曾指出，正交表都附有一个交互作用列表(见表 9.1.3)，根据该表可得出代表因子交互作用的列，把该列作为一个因子处理可按上法进行正交设计的分析，但在进行显著性检验、确定最佳方案、平均效应以及置信区间时，却稍有不同。

9.1.3.1 交互作用显著性的 F 检验

把交互作用列当作一个因子考虑进行显著性检验的基本公式仍是

$$F = \frac{S_k / f_k}{S_{误} / f_{误}} \tag{9.1.9}$$

当 $F \geqslant F_\alpha$ 时，因子是显著的，否则是不显著的。

如某建筑公司木材加工厂研究不同材种的木材加工工艺时，确定如下因子及水平：

A：树种 A_1 = 针叶，A_2 = 阔叶。

B：工艺 B_1 = Ⅰ，B_2 = Ⅱ。

C：锯条类型 C_1 = 甲，C_2 = 乙。

D：传送带类型 D_1 = Ⅰ，D_2 = Ⅱ。

现以劳动生产率为评价指标，选正交表 $L_8(2^7)$。经 8 次试验得表9.1.11。

将 A 与 B 两因子的交互作用当作因子进行方差分析，并进行显著性检验，见表 9.1.12。

由表可得出如下结论：A、B 两因子并不显著(树种、工艺单独对劳动生产率影响不显著)，而锯条类型、树种与工艺的交互作用两因子对生产率的影响较大。

表 9.1.11 经 8 次试验结果

试验号	A	B	A×B	C			D	劳动生产率/(m^3/人)
	1	2	3	4	5	6	7	
1	1	1	1	1	1	1	1	115
2	1	1	1	2	2	2	2	160
3	1	2	2	1	1	2	2	145
4	1	2	2	2	2	1	1	155
5	2	1	2	1	2	1	2	140
6	2	1	2	2	1	2	1	155
7	2	2	1	1	2	2	1	100
8	2	2	1	2	1	1	2	125
M_1	575	570	500	500	540	535	525	=1,095
M_2	520	525	595	595	555	560	570	$\bar{y}$=136,875
R_j	55	45	95	95	15	25	45	
S_j	378.1	253.1	1128.1	1128.1	28.1	78.1	253.1	

表 9.1.12 显著性检验

方差名称	S	f	S	F	显著性
A	378.1	1	378.1	7.1	
B	253.1	1	253.1	4.8	
C	1128.1	1	1128.1	21.2	*
D	253.1	1	253.1	4.8	
A×B	1128.1	1	1128.1	21.2	*
误	106.2	2	53.1		

9.1.3.2 确定最佳方案及平均效应

在多因素且有交互作用的试验中，最佳方案的确定是困难的和复杂的。

任一交互作用的两因子 A、B 水平，可按正交表的原理互相搭配成一套方案。如 A、B 均为二水平，相当于 AB 交互作用时共有四个水平，见表 9.1.13。

表 9.1.13 交互影响效应

试验号	A、B	指标和	平均指标	与 $\bar{y}$ 的差 $[ab]_{ij}$	交互影响效应 $(ab)_{ij}$
1、2	1 1	275	137.5	0.625	-11.875
3、4	1 2	300	150	13.125	11.875
5、6	2 1	295	147.5	10.625	11.875
7、8	2 2	225	112.5	-24.375	-11.875
注：表中数字来源于原正交表中包含该组合的试验数据。					

其中与 $\bar{y}$ 之差一列表示 A、B 两因素不同水平组合对指标的影响，即 A、B 各水平不同组合的效应，以 $[ab]_{ij}$ 表示。如果从中去掉 A、B 两因素单独对指标的影响（A、B 各水平的单独效应），则剩余的即是 A、B 交互作用的影响（以 $(ab)_{ij}$ 表示），即

$$(ab)_{ij} = [ab]_{ij} - (a_i + b_j) \tag{9.1.10}$$

a_i、b_j 分别是 A 因素 i 水平和 B 因素 j 水平的效应。应用以前的方法计算结果见表 9.1.14。

表 9.1.14 计算结果

	B_1	B_2	总和	效应 a_i
A_1	115+160=275	145-155=300	575	6.875
A_2	140+155=295	100+125=225	520	-6.875
总和	570	525	$T=1095$	
效应 b_j	5.625	-5.625		

因此 $(ab)_{11}=0.625-(6.875+5.625)=-11.875$

$(ab)_{12}=13.125-(6.875-5.625)=11.875$

$(ab)_{21}=10.625-(-6.875+5.625)=11.875$

$(ab)_{22}=24.375-(-6.875-5.625)=-11.875$

当交互作用影响显著时，有如下四种情况：

(1) 两因子都不显著；

(2) A 因子显著，B 因子不显著；

(3) A 因子不显著，B 因子显著；

(4) 两因子都显著。

当为(1)时，选 $(ab)_{ij}$ 中最大者，本例属该种情况，因 $(ab)_{12}=(ab)_{21}$，故选 A_1B_2 或

A_2B_1 均可；当为(2)时，选 A 的最优水平，再选 B 与之搭配最好的水平；当为(3)时，选 B 的最优水平，再选 A 与之搭配的最好水平；当为(4)时，选 $(ab)_{ij}$ 中最大者。

由上述计算得出本例最佳方案为 $A_1B_2C_2D_1$ 或 $A_2B_1C_2D_1$，其平均效应(工程平均)为显著因子的效应与总平均值之和：

$$\mu_{优} = \bar{y} + (ab)_{12} + c_2 = 136.875 + 11.875 + 11.875 = 160.625$$

9.1.3.3 置信区间的确定

置信区间的确定方法仍是将交互因子作为一个因子处理，其自由度为 $r-1$，且按

$$\delta = \sqrt{F_{1-a}(1,\tilde{f}_{误})\frac{\tilde{S}_{误}}{\tilde{f}_{误}\cdot n_e}} \tag{9.1.11}$$

计算

式中 $\tilde{S}_{误}$——是随机误差 $S_{误}$ 与所有不显著因子和所有不显著交互作用的离差平方之和；

$\tilde{f}_{误}$——上述离差平方和相对应的自由度；

n_e——有效重复度，且 $n_e = \dfrac{\text{数据个数}}{1+\text{显著因子及显著交互作用自由度之和}}$。

本例中：

$$\tilde{S}_{误} = S_{误} + S_A + S_B + B_C + S_D = 990.5$$

$$\tilde{f}_{误} = f_{误} + f_A + f_B + f_C + f_D = 5$$

$$n_e = \frac{8}{1+1+1} = \frac{8}{3}$$

$$\delta = \sqrt{F(1,5)\frac{\tilde{S}_{误}}{\tilde{f}_{误}\cdot n_e}} = \sqrt{6.61\frac{3\times 990.5}{5\times 8}} = 22.16$$

置信区间为

$$(\mu_{优}-\delta,\mu_{优}+\delta) = (160.625-22.16,160.625+22.16) = (138.47,182.75)$$

9.1.4 常用表头最佳设计

由交互作用列表可见，安排有交互作用的正交表表头是复杂的事情，安排不当，会影响分析工作的进行，也影响分析结果的准确性。因此需将不可忽略的交互作用安排在与单独因子不同的列上，以免使单独因子的效应被混淆。在因素多、试验次数少的情况下，发生混淆是必然的，但安排适当可尽量减少这种混杂。表 9.1.15～表 9.1.17 是常用正交表表头的最佳设计方案。

表 9.1.15 $L_8(2^7)$最佳设计表

因子数 \ 实施 \ 列号		1	2	3	4	5	6	7
3	1	A	B	AB	C	AC	BC	
4	1/2	A	B	AB CD	C	AC BD	BC AD	D

表 9.1.16　$L_{16}(2^{15})$最佳设计表

因子数 \ 列号	实施	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	1	A	B	AB	C	AC	BC	DE	D	AD	BD	CE	CD	BE	AE	E
5	$\frac{1}{2}$	A	B	AB	C	AC	BC	DE	D	AD	BD	CE	CD	BE	AE	E
6	$\frac{1}{4}$	A	B	AB DE	C	AC DF	BC EF		D	AD BF CF	BD AE CE	E	CD AF	F		CE BF
7	$\frac{1}{8}$	A	B	AB DE FG	C	AC DF EG	BC EF DG		D	AD BE CF	BD AE CG	E	CD AF BG	F	G	CE BF AG
8	$\frac{1}{16}$	A	B	AB DE FG CH	C	AC DF EG B H	BC EF DG AH		D	AD BE CF CH	BD AE CG FH	E	CD AF BG EH	F	G	CE BF AG DH

表 9.1.17　$L_{27}(2^{13})$最佳设计表

因子数 \ 列号	实施	1	2	3	4	5	6	7	8	9	10	11	12	13
3	1	A	B	AB	AB	C	AC	AC	BC			BC		
4	$\frac{1}{3}$			AB CD	AB	C	AC BC	AC	BC AD	D	AD	BC	BD	CD

9.1.5　水平数不等的正交试验

实践中，问题总是各种各样的。上述方法都是因子水平数相等的正交试验。但在试验中经常因某些原因，如受试验条件限制等使某些因素不能取更多的水平，因此需解决不等水平的试验问题。

解决不等水平的试验方法有两种：① 直接选用不等水平的正交表；② 拟水平法。

不等水平的正交表叫混合型正交表，如一因子有四个水平、三个因子有二个水平，可选用 $L_8(4\times2^4)$混合型正交表(见表 9.1.18)。

表 9.1.18　$L_8(4\times2^4)$混合型正交表

试验号 \ 因子	A	B	C	D
1	1	1	1	1
2	1	2	2	2
3	2	1	1	2
4	2	2	2	1

（续）

因子 / 试验号	A	B	C	D
5	3	1	2	1
6	3	2	1	2
7	4	1	2	2
8	4	2	1	1

其计算分析方法与上述相同。

如在现成的混合型正交表中找不到现成的正交表，或即使能找到但需试验的次数过多，则需采用拟水平法。

所谓拟水平法即是将所有因子水平数人为地用该因子的其他水平代替而使其相等的方法。

如三因子的试验中，A、B 两因子均三水平，而 C 因子只能有二水平，可将该因子的两个水平中认为较好的一个水平当作第三水平，而使各因子的水平数相等，从而可采用 $L_9(3^4)$ 正交表。

9.1.6 多指标试验

上述试验方法均属单指标试验，即以一个指标对试验进行评价。而实践中经常采用多指标进行评价，为解决该问题，必须将多个评价指标化成单一指标。具体化法可采用系统决策中将多目标化为单目标的一些方法，详细内容请参阅本书多目标决策部分。

9.2 模拟技术——蒙特卡罗法

蒙特卡罗法是一种随机模拟技术，是系统辅助决策和决策检验的重要工具之一，是系统工程解决问题的一种重要手段。

蒙特卡罗法是运用一系列随机数来模拟某项概率分布，采取反复采样的方法在计算机上使系统运行，从而为系统决策提供依据或对系统决策进行检验的一种方法。

9.2.1 蒙特卡罗法的基本原理

采用随机数的方法解决问题是蒙特卡罗法的基本点。现举例说明蒙特卡罗法应用随机数解决问题的过程。

例 1 当函数 $f(x)$ 很复杂时，求图 9.2.1 中阴影部分的面积。

阴影部分的面积为

$$S = \int_0^a f(x)\mathrm{d}x$$

当函数 $f(x)$ 很复杂时，求该定积分的值是很困难的。现应用蒙特卡罗法采用随机数来求解。

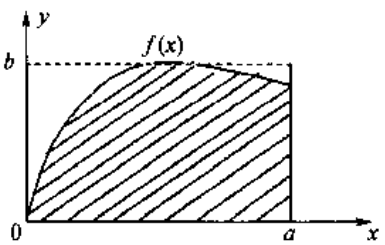


图 9.2.1 求阴影部分面积

设:在区间$[0,a]$内$f(x)>0$,且$f(x)$的最大值为b。

由图可见,阴影部分的面积包含在由边长a和b组成的矩形内。在矩形内产生随机点并确定是否落在影部分。显然,若所提供的随机点均匀地分布在矩形内,则阴影部分的面积与矩形面积之比应等于落在阴影部分随机点的个数n与随机点总数N之比,即

$$\frac{\int_0^a f(x)\mathrm{d}x}{ab} = \frac{n}{N} \tag{9.2.1}$$

由式(9.2.1)知,阴影部分面积(积分值)为

$$\int_0^a f(x)\mathrm{d}x = \frac{abn}{N} \tag{9.2.2}$$

当随机点的数量足够大时,阴影部分的面积可用矩形面积ab乘以n/N来估计,N的数量越大,估计越准确。由此可见,该问题求解的关键是产生随机数和判定随机点是否落在阴影部分之内。其解决方法是:在x轴的$[0,a]$区间内随机地选取一点x作为x_0,在y轴的$[0,b]$区间随机地选取y作为y_0,若$y_0 \leqslant f(x_0)$,则随机点(x_0, y_0)落在阴影部分内,否则落在阴影部分外。如此反复进行,即可产生一个又一个的随机点,记录所发生的随机点总数N和落在阴影部分的随机点数n即可得出n/N,从面求出积分值。

例 2 某工地搅拌机在1min~10min搅拌一车混凝土,力工在1min~6min内完成其送料工作,现分析力工的工作状况。

力工的工作取决于到搅拌混凝土的时间和送料所需的时间,但二者均是随机量。现用标有数字1~10的卡片和一个骰子分别模拟搅拌混凝土的时间间隔和送料时间,随机抽取一张卡片,其标号表示搅拌混凝土的时间,掷一下骰子,其点数则表示送料的工作时间,如此可模拟力工的工作过程,表9.2.1为模拟结果。

表9.2.1 力工工作状况模拟表

原木	混凝土搅拌的时间/min	混凝土搅拌的时刻	送料时间/min	送料时刻		混凝土在系统中的全部时间	力工空闲时间/min
				开始	结束		
1	—	8:00	1	8:00	8:01	1	0
2	3	8:03	4	8:03	8:07	4	2
3	7	8:10	4	8:10	8:14	4	3
4	3	8:13	2	8:14	8:16	3	0
5	9	8:22	1	8:22	8:23	1	6
6	10	8:32	5	8:32	8:37	5	9
7	6	8:38	4	8:38	8:42	4	1
8	8	8:46	6	8:46	8:52	6	4
9	8	8:54	1	8:54	8:55	1	2
10	8	9:02	3	9:02	9:05	3	7
11	7	9:09	5	9:09	9:14	5	4
12	3	9:12	5	9:14	9:19	7	0
13	8	9:20	3	9:20	9:23	3	1

（续）

原木	混凝土搅拌的时间/min	混凝土搅拌的时刻	送料时间/min	送料时刻		混凝土在系统中的全部时间	力工空闲时间/min
				开始	结束		
14	4	9:24	6	9:24	9:30	6	1
15	4	9:28	1	9:30	9:31	3	0
16	7	9:35	1	9:35	9:36	1	4
17	1	9:36	6	9:36	9:42	6	0
18	6	9:42	1	9:42	9:43	1	0
19	7	9:49	2	9:49	9:51	2	6
20	6	9:55	2	9:55	9:57	2	5
总　计						68	55

由表可见，在早 8:00 到 9:57 的 117min 内，力工共运送 20 车搅拌料，纯工作时间 68min，空闲时间 55min，平均每运送一车需 3.4min，该力工工时利用率为

$$\eta = \frac{68}{117} = 53\%$$

可见该力工的工作是轻松的，工时利用率不高，工时浪费较大。

处理该问题的关键在于产生搅拌时间和力工工作时间的随机数。如果所产生的随机数符合实际情况，则模拟结果是可靠的。

由上述两个例子可见，蒙特卡罗法可解决确定型问题（如例 1），也可解决随机型问题（如例 2）。该方法的优点是：简单，可用于解决复杂的问题；一般不需要连续性、可微性等条件且特别适用于求解高维问题。其缺点是计算工作量大、精度不高。因此能用传统的物理试验和数学方法处理的问题，采用这种方法只能作为决策的参考。这种方法对解决无法应用数学方法或传统的物理试验的问题是很有效的。

由上述两个例子可见，蒙特卡罗法解决问题的过程大体如下：

（1）建立适当的概率模型（随机事件、随机过程等）；

（2）进行模拟；

（3）对模拟结果进行统计处理，以得出问题的解。

9.2.2 随机数的产生

蒙特卡罗法解决问题的基本点是利用随机数。例 2 中的随机数是用抽卡片和掷骰子的方法得到的。采用这种方法产生随机数的随机性是不可靠的。为了高效率、高速度地产生随机数，就必须使用计算机，因此计算机就成为系统模拟的重要工具。

用计算机产生的随机数通常是伪随机数。伪随机数是指在产生随机数的过程中，第 $i+1$ 个随机数是根据第 i 个随机数按一定的公式计算得到的，因此容易在计算机上产生。它的较严密的定义为：在计算机上用算法得到的“统计性质”上近似于[0,1]上均匀分布的数，称为伪随机数，它区别于真正从[0,1]上均匀分布中抽样所取得的随机数。

常用的随机数及产生办法如下。

9.2.2.1 均匀分布随机数及产生方法

均匀分布随机数是最简单、最常用的随机数，它是产生其他任何分布随机数的基础。

均匀分布随机数是在[0,1]区间内，以同样的概率产生的一系列随机数，其概率密度和分布函数为

$$f(x)=\begin{cases}1, & 0\leqslant x\leqslant 1\\ 0, & \text{其他}\end{cases} \tag{9.2.3}$$

$$F(x)=\begin{cases}x, & 0\leqslant x\leqslant 1\\ 0, & x>1\end{cases} \tag{9.2.4}$$

在计算机内，均匀分布的随机数通常是由伪随机数发生器产生的，在无随机数发生器的计算机中可用以下方法编一简单程序得到或将均匀分布随机数存入磁盘随时调用。

1. 平方取中法

任选一大于32位[①]的$2n$位(即偶数位)的整数x_0作为初始值(种子)，然后将其平方得到一$4n$位的整数，取出中间的$2n$位用为第一个随机数x_1；再把x_1平方取出中间的$2n$位作为第二个随机数x_2，如此不断地重复下去就可得到$x_1,x_2,\cdots$这一均匀分布的伪随机数列。如：

$$x_0=2158 \qquad x_0^2=04656964$$

$$x_1=6569 \qquad x_1^2=43151761$$

$$x_2=1517 \qquad x_2^2=02301289$$

$$x_3=3012 \qquad \cdots$$

这样一直计算下去就得到一个四位数的伪随机数列，这种产生随机数的方法只要用一简单程序就可实现。但该方法的缺点是循环周期短，较易退化。

2. 乘积取中法

选$2n$位的整数x_0和x_1作为初始值，取其乘积中间的$2n$位作为x_2，然后在x_1和x_2的乘积中以同样的方法得到x_3，如此不断地重复下去就可得到一个伪随机数列。如：

$$x_0=1234 \qquad x_1=5678$$

$$x_0\cdot x_1=07006652 \qquad x_2=0066$$

$$x_1\cdot x_2=00374748 \qquad x_3=3747$$

$$\cdots$$

这种伪随机数列也可在计算机上用一简单程序实现，其优点是循环周期较长，均匀分布的特点较好，但较易退化。

3. 混合同余法

同余运算又称模运算，一般以$A(\text{mod } B)$表示，其运算结果是整数A与整数B相除的余数。如$7(\text{mod } 3)=1$。

混合同余法一般按下述递推公式计算随机数：

$$x_{n+1}=(a\cdot x_n+b)(\text{mod } M)$$

---

① 这里所指的32位数是二进制数

其中 a,b,M 以及初值 x_0 都为正整数。

如果,$b=0$,则:$x_{n+1}=ax_n(\mathrm{mod}\ M)$

此时随机数仅由 ax_n 和 M 相除的余数决定,因此叫乘同余法,乘同余法是计算机中最常采用的方法,很多计算机都具有同余运算功能,因此用计算机产生这种随机数很方便。

一个理想的伪随机数发生器,应能产生均分布的随机数列且各数字之间相关性差,数列循环周期长。因此,在二进制计算机中,M 一般取 2^w(其中 w 为计算机字长),a 取 2^l+1、2^l+2^m+1 一类数字可产生较好的均分布的伪随机数。

9.2.2.2 给定分布的随机数

模拟工作中经常遇到给定分布随机变量的模拟,根据随机变量的特性,可分为离散随机变量模拟和连续随机变量模拟两种。

1. 离散型随机变量的模拟

设 Z 为离散型随机变量,已知

$$P\{Z=a_i\}=P_i$$

其中,$P_i\geqslant 0$,$\sum_i p_i=1$。现欲以[0,1]区间上均匀分布随机数 $x_1,x_2,\cdots$ 为基础产生该随机变量的随机数。其产生方法为

$$Z_i=\begin{cases}a_1;\ \text{当}\ 0\leqslant x_k\leqslant p_1\\ a_2;\ \text{当}\ p_1\leqslant x_k<p_1+p_2\\ \cdots\\ a_i;\ \text{当}\ p_1+p_2+\cdots+p_{i+1}\leqslant x_k<p_1+p_2+\cdots+p_i\\ \cdots\end{cases}\qquad (k=1,2,\cdots)$$

这种方法实质上是将整个区间划分成几段,[0,1]区间均匀分布随机数出现在哪个区间段内,就等于发生了哪个事件。

2. 连续型随机变量的模拟

对于连续型随机变量 Z,其分布函数为 $F(\mu)$,即 $p\{Z<\mu\}=F(\mu)$,如果 $F(\mu)$连续且严格单调上升,易知 $X=F(Z)$为[0,1]区间的均匀分布,如 X 为[0,1]区间内均匀分布的随机变量,则

$$Z=F^{-1}(X)$$

便是以 $F(\mu)$为分布函数的随机变量。其中 $F^{-1}(\mu)$为 $F(\mu)$的反函数。

如已得一组[0,1]上均匀分布的随机数$\{x_1,x_2,\cdots\}$,可由:

$Z_k=F^{-1}(x_k)$ $(k=1,2,\cdots)$ 求得一组以 $F(\mu)$为分布函数的随机数。

如指数分布的密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geqslant 0\\ 0, & x<0\end{cases}$$

其分布函数为

$$F(x)=\begin{cases}1-e^{-\lambda x}, & x\geqslant 0\\ 0, & x<0\end{cases}$$

若 Z 为指数分布的随机变量,则

$x=1-e^{-\lambda t}$为[0,1]上均匀分布的随机变量，$Z=-\frac{1}{\lambda}\ln(1-x)$即为指数分布的随机变量，因 x 为[0,1]区间均匀分布的随机变量，故 $1-x$ 也是[0,1]上的均匀分布的随机变量，现仍以 x 表示，则

$$Z=-\frac{1}{\lambda}\ln x$$

为指数分布的随机变量。

9.2.2.3 正态分布随机变量的产生方法

正态分布是使用较多的分布，因其分布函数较复杂，在模拟精度不高的情况下，可采用近似方法。该方法首先由[0,1]区间均匀分布的随机数得到 $N(0,1)$分布的随机数，再按：$X=\mu+\sigma Z$ 得到$N(\mu,\sigma)$分布的随机数。

由[0,1]区间均匀分布随机数得到 $N(0,1)$分布随机数的方法是：

设 $x_1,x_2,\cdots,x_n$ 是[0,1]区间上均匀分布的随机变量，则其均值为 1/2，方差为 1/12。

根据中心极限定理得

$$Z=\frac{\sum_{i=1}^{n}(x_i-\frac{1}{2})}{n\cdot\frac{1}{12}}$$

在 n 较大时，其分布近似于 $N(0,1)$分布。

当 $n=12$ 时，有

$$Z=\sum_{i=1}^{12}(X_i-\frac{1}{2})=$$

$$x_1+x_2+\cdots+x_{12}-6=x_1+x_2+\cdots+$$

$$x_6-(1-x_7)-\cdots-(1-x_{12})$$

因 x_i 和 $1-x_i$ 均是[0,1]区间的均匀分布，所以可取

$$Z=x_1+x_2+\cdots+x_6-x_7-\cdots-x_{12}$$

这一作法在一般使用中精度是足够的。

如果模拟精度要求较高，则可直接应用以下公式在计算机中产生 $N(0,1)$分布的随机数：

$$Z_1=\sqrt{-2\ln x_1}\cos 2\pi x_2$$

$$Z_2=\sqrt{-2\ln x_1}\sin 2\pi x_2$$

其中 Z_1 和 Z_2 是相互独立的 $N(0,1)$分布的随机变量，x_1、x_2 是[0,1]区间均匀分布的随机数。这样，给出二个均匀分布的随机数，就得到二个正态分布的随机数，如此反复，即可得到一正态分布的随机数列。

9.2.3 蒙特卡罗法的应用

9.2.3.1 合理库存量的确定

确定型的库存问题，可建立一非线性规划模型(无约束)确定合理的库存量，随机型库存则比较困难。应用蒙特卡罗法可确定随机库存问题的最佳方案。

某建筑公司木制品厂以木材为生产原料。经分析，一年内木制门、窗的社会需求呈正态分布，故原料需求也呈正态分布，经统计其均值 $\mu=100m^3$，方差 $\sigma=30m^3$，若木材存储费用为 200 元/周，因缺货造成的损失为 100 元/m^3，现应用蒙特卡罗法确定合理的库存量。

$N(100,30)$随机数表如下：

99	167	123	83	95	87	95	63	131	71
91	87	139	99	51	87	55	107	107	139
95	99	91	103	103	111	99	99	99	103
67	151	123	75	75	99	163	151	111	123
87	115	47	59	103	147	151	55	59	67

现分别以库存量为 $s=60, a=70, s=80, s=90, s=100, s=110$ 六种方案进行模拟。

假设仓库存储量为 $60m^3$，第一周需求量为 $99m^3$，缺货为 $99-60=39m^3$，缺货损失 $c_2=100\times39=3900$ 元，依次选取随机数作为需求量，计算到第十五周，需求量为 $51m^3$，仓库积压量为 $60-51=9m^3$，此时存储费用为：$c_1=9\times200=1800$ 元，按照该方法一直计算 50 周为止，即模拟 50 周的库存变化情况。

对其他五种方案也按此法计算，计算结果列于表 9.2.2。

表 9.2.2 库存量模拟结果 （单位：千元）

周次	存储量 / 需求量	$s=60$		$s=70$		$s=80$		$s=90$		$s=100$		$s=110$	
		c_1	c_2	c_1	c_2	c_1	c_2	c_1	c_2	c_1	c_2	c_1	c_2
1	99		3.0		2.9		1.9		9.0	0.2		2.2	
2	167		10.7		9.7		8.7		7.7		6.7		5.7
3	123		6.3		5.3		4.3		3.7		2.3		1.3
4	83		2.3		1.3		0.3	1.4		3.4		5.4	
5	95		3.5		2.5		1.5		0.5	1.0		3.0	
6	87		2.7		1.7		0.7	0.6		2.6		4.6	
7	95		3.5		2.5		1.5		0.5	1.0		3.0	
8	63		0.3	1.4		3.4		5.4		7.4		9.4	
9	131		7.1		6.1		5.1		4.1		3.1		2.1
10	71		1.1		0.1	1.8		3.8		5.8		7.8	
11	91		3.1		2.1		1.1		0.1	1.8		3.8	
12	87		2.7		1.7		0.7	0.6		2.6		4.6	
13	139		7.9		6.9		5.9		4.9		3.9		2.9
14	99		3.9		2.9		1.9		0.9	0.2		2.2	
15	51	1.8		3.8		5.8		7.8		9.8		11.8	
16	87		2.7		1.7		0.9	0.6		2.6		4.6	
17	55	1.0		3.0		5.0		7.0		9.0		11.0	
18	107		4.7		3.7		2.7		1.7		0.7	0.6	
19	107		4.7		3.7		2.7		1.7		0.7	0.6	
20	139		7.9		6.9		5.9		4.9		3.9		2.9
21	95		3.5		2.5		1.5		0.5	1.0		3.0	
22	99		3.9		2.9		1.9		0.9	0.2		2.2	
23	91		3.1		2.1		1.1		0.1	1.8		3.8	
24	103		4.3		3.3		2.3		1.3		0.3	1.4	
25	103		4.3		3.3		2.3		1.3		0.3	1.4	
26	111		5.1		4.1		3.1		2.1		1.1		0.1
27	99		3.9		2.9		1.9		0.9	0.2		2.2	

（续）

周次	存储量 / 需求量	$s=60$		$s=70$		$s=80$		$s=90$		$s=100$		$s=110$	
		c_1	c_2	c_1	c_2	c_1	c_2	c_1	c_2	c_1	c_2	c_1	c_2
28	99		3.9		2.9		1.9		0.9	0.2		2.2	
29	99		3.9		2.9		1.9		0.9	0.2		2.2	
30	103		4.3		3.3		2.3		1.3		0.3	1.4	
31	67		0.7	0.6		2 6		4.6		6.6		8.6	
32	151		9.1		8.1		7.1		6.1		5.1		4.1
33	123		6.3		5.3		4.3		3.3		2.3		1.3
34	75		1.5		0.5	1 0		3.0		5.0		7.0	
35	75		1.5		0.5	1.0		3.0		5.0		7.0	
36	99		3.9		2.9		1.9		0.9	0.2		2.2	
37	163		10.3		9.3		8.3		7.3		6.3		5.3
38	151		9.1		8.1		7.1		6.1		5.1		4.1
39	111		5.1		4.1		3.1		2.1		1.1		0.1
40	123		6.3		5.3		4.3		3.3		2.3		1.3
41	87		2.7		1 7		0.7	0.6		2.6		4.6	
42	115		5.5		4.5		3.5		2.5		1.5		0.5
43	47	2.6		4.6		6.6		8.6		10.		12.6	
44	59	0.2		2.2	1		4.2		6.2		6		10.2
45	103		4.3		3.3		2.3		1.3	8.2	0.3	1.4	
46	147		8.7		7.3		6.7		5.7		4.7		3.7
47	151		9.1		8.1		7.1		6.1		5.1		4.1
48	55	1.0		3.0		5.0		7.0				11.0	
49	59	0.2		2.2		4 2		6.2		9.0		10.2	
50	67		0.7	0.6		2.6		4.6		8.2		8.6	
										6.6			
合计		68	2040	214	1609	432	1222	710	861	1130	571	1778	395

将表 9.2.2 模拟结果整理得表 9.2.3。

表 9.2.3 模拟结果

库存量/m^3	存储费/千元	缺货损失/千元	总费用/千元
60	6.8	204.0	210.8
70	21.4	160.9	182.3
80	43.2	122.2	165.4
90	71.3	86.1	157.1(最小)
100	113.0	57.1	170.1
110	177.8	39.5	217.3

由表 9.2.3 可见，当库存量为 $90m^3$ 时，该木制品厂原料库存比较经济。

9.2.3.2 风险决策的模拟

某企业开发一种新产品，该产品的盈利售价、可变成本和销售量三个因素的影响。新产品进入市场的售价有三种可能：(1)每件 40 元的概率为0.3；(2)每件 50 元的概率为 0.5；(3)每件 60 元的概率为0.2。因原材料价格的变动使该产品可变成本发生变化，其变化亦有三种可能：(1)每件可变成本 20 元的概率为 0.1；(2)为 30 元的概率为 0.6；(3)为 40 元的概率为 0.3。预测该产品的销售量也有三种可能：(1)销售 3000 件的概率为 0.2；(2)4000 件的概率为 0.4；(3)5000 件的概率为 0.4。现用蒙特卡罗法进行模拟。

由于售价、可变成本、销售量均为离散变量，故需应用[0,1]区间上均匀分布的随机数形成符合该分布率的随机数。

当随机变量在[0,0.3]区间时，售价为 40 元；在[0.3,0.8]区间时，售价为 50 元；在[0.8,1]区间内时，售价为 60 元。

当随机变量在[0,0.1]区间时，可变成本为 20 元，在[0.1,0.7)区间时，为 30 元；在[0.7,1]区间时，为 40 元。

当随机变量在[0,0.2]区间时，销售量为 3000 件，在[0.2,0.6]区间销售量为 4000 件，在[0.6,1]区间内时销售量为 5000 件。

若固定成本为 50000 元，利润则按下式计算：

$$\Pi = (p - p_1) \cdot Q - p_0$$

式中 Π——利润额；

p——售价；

p_1——可变成本；

Q——销量；

p_0——固定成本。

模拟结果见表 9.2.4。

表 9.2.4 模拟结果

序号	售价/千元		可变成本/千元		销售量/件		利润/千元
	随机数	模拟量	随机数	模拟量	随机数	模拟量	
1	0.893	60	0.079	20	0.939	5000	150
2	0.274	40	0.513	30	0.243	4000	−10
3	0.379	50	0.267	30	0.411	4000	30
4	0.208	40	0.183	30	0.795	5000	0
5	0.306	50	0.942	40	0.879	5000	0
6	0.286	40	0.870	40	0.704	5000	−50
7	0.846	60	0.450	30	0.289	4000	70
8	0.057	40	0.464	30	0.188	3000	−20
9	0.028	40	0.567	30	0.972	5000	0
10	0.285	4	0.981	40	0.751	5000	−50
11	0.965	060	0.119	30	0.332	4000	70
12	0.346	50	0.006	20	0.154	3000	40
13	0.633	5	0.402	30	0.864	5000	50
14	0.704	050	0.253	30	0.156	3000	10
15	0.196	40	0.569	30	0.146	3000	−20
16	0.783	50	0.845	40	0.123	3000	10
17	0.418	50	0.055	20	0.738	5000	100
18	0.152	40	0.663	30	0.171	3000	−20
19	0.760	50	0.801	40	0.405	4000	−10
20	0.908	6	0.879	40	0.322	4000	30
21	0.603	050	0.915	40	0.954	5000	0
22	0.749	50	0.306	30	0.645	5000	50
23	0.097	40	0.508	30	0.706	4000	−10
24	0.893	60	0.963	40	0.946	5000	50
25	0.104	40	0.028	20	0.942	5000	50

经 25 次模拟得平均利润为 20800 元，如继续模拟可使平均利润的精度提高，可为决策提供依据。

9.3 系统动态学模拟原理

对自然科学家或工程师来说，进行试验已司空见惯，他们可通过试验来探索和发现事物发展变化规律。但是，以社会为研究对象的社会科学家却只能靠社会调查和分析。可以设想，如果社会科学工作者、战略家、管理者在进行社会、企业重大变革之前，能像自然科学家那样通过“社会科学试验室”对各种政策方案进行实际试验之后再做出决策，对社会科学的研究会有多么重大的意义。

为了适应该需要，20 世纪 50 年代美国麻省理工学院斯隆管理学院的福雷斯特教授融控制理论、系统论、信息论、计算机模拟技术、管理科学及决策论等学科的知识为一体，开发了系统动态学(SYSTEM DYNAMICS)，简称 SD。

该方法是一种用计算机对社会大系统进行模拟的方法，是用计算机研究发展战略与策略的一种方法，因此，该方法被誉为“战略与策略试验室”，福雷斯特也被誉为当代的伽里略。

系统动态学是建立在系统论、自动控制理论和信息论基础上的，它依靠系统理论来分析系统的结构和层次、依靠自动控制理论中的反馈原理对系统进行调节、依靠信息论中信息传递原理来描述系统，并采用电子计算机对系统动态行为进行模拟。它最适合于分析和研究复杂的社会经济系统。

9.3.1 系统因果关系分析

要想了解系统的动态行为，必须依据系统理论研究系统的结构和层次以及系统各组成部分之间的关系。系统动态学认为，系统各组成部分之间的关系是一种因果关系，这种因果关系决定了系统内存的运行规律。系统动态学还认为，任何系统都有是反馈系统，即系统的输出对输入产生影响，也就是系统的行为受系统过去行为的影响。因此只要将系统各组成部分之间的因果关系(包括反馈关系)应用信息传递的原理进行描述，即可揭示系统的动态行为。因此系统动态学方法是以确定系统因果关系为基础的。

1. 因果关系图

系统动态学中因果关系是指系统各要素(子系统)之间的原因和结果关系，通常以因果关系图表示。

因果关系图是用箭线表示的系统内各子系统之间因果关系的图，图 9.3.1 表示系统内 A 子系统对 B 子系统的作用，其中箭线叫关系键，A 是原因，B 是结果。

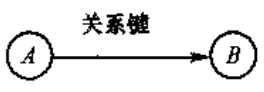


图 9.3.1 A 对 B 的作用

因果关系分成正因果关系和负因果关系，分别以“+”号和“-”号标在关系键旁。

正因果关系是指 A、B 变化方向一致，负因果关系是指 A、B 变化方向相反，因此关系键根据所标符号不同又称正键和负键。无论是正键还是负键都只是一种逻辑关系，无

时间意义，也无计算意义。

图 9.3.2 表示环境质量和人的寿命的正因果关系；图 9.3.3 表示水污染和农产品产量的负因果关系。

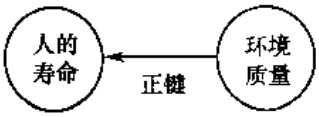


图 9.3.2 正因果关系

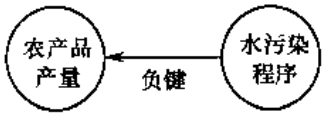


图 9.3.3 负因果关系

2. 因果反馈环

所谓因果反馈环是指两个以上的因果关系键首尾串联而成的封闭环形。从总体上讲，无法确定谁是起点、谁是终点，即无法确定谁是原因、谁是结果。如图 9.3.4 为一因果反馈环。

反馈环中关系键可正、可负，它们组成的反馈环分为正反馈环和负反馈环。如图 9.3.5(a)所示，投资增加，产量提高，产量提高，又会促使投资进一步增加，这样一个因素的增加或减少经反馈后会使该因素进一步增加或减少的反馈环叫正反馈环，它具有自我加强的功能。产量的增加会影响价格，使价格下降，价格的上涨会使产量增加。这种反馈环中的一个因素增加(减少)，经反馈后使该因素减小(增加)的反馈环是负反馈环，如图 9.3.5(b)，它具有自我调整和稳定的作用。

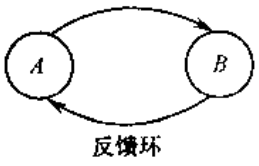


图 9.3.4 因果反馈环

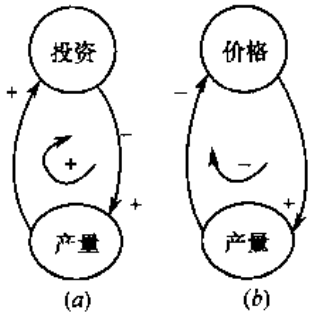


图 9.3.5 正反馈环和负反馈环

当一个反馈环中因素较多、关系键较多时，需判断反馈环的性质。为使判断方便起见，现给出因果反馈环的分析方法：

(1) 若环中各键均为正键或环中有偶数条负键时，该环为正反馈环；

(2) 若环中有奇数条负键，则环为负反馈环。

9.3.2 负反馈环分析

因负反馈环有自我调整和稳定的效果，因此若使系统处于稳定状态可通过负反馈环来实现。

现以一库存系统为例说明负反馈环的作用。

某企业目标库存量 MB 为 6000，计划人员依据目前库存决定进货量，进货速度由调整期决定。设目前实际库存量为 1000，调整期为 5 周，现模拟该系统是如何达到目标库存量的。

根据目标库存量(MB)、调整期(TZ)、实际库存量(KC)、进货速度(DH)之间的关系可得图 9.3.6 的因果反馈图和以下方程：

$$DH_t = (MB - KC_t)/TZ$$

根据该方程得模拟表 9.3.1。

表 9.3.1　库存系统模拟表

时间	库存变化量	库存量(KC)	库存差($MB-KC$)	进货速度(DH)
0	0	1000	5000	1000
2	2000	3000	3000	600
4	1200	4200	1800	360
6	720	4920	1080	216
8	432	5352	648	129.6
10	259	5611	389	78
12	156	5767	233	47
14	93	5860	140	28
16	56	5916	84	17
18	34	5950	50	10
20	20	5970	30	6
22	12	5982	18	3.6
24	7.2	5989.2	10.8	2.1

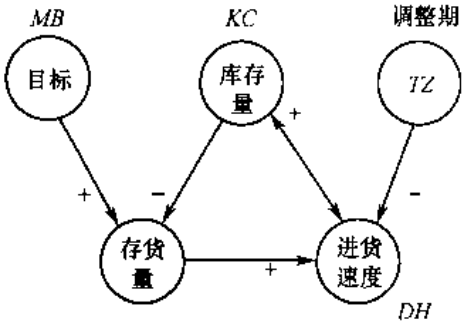


图 9.3.6　库存系统因果关系图

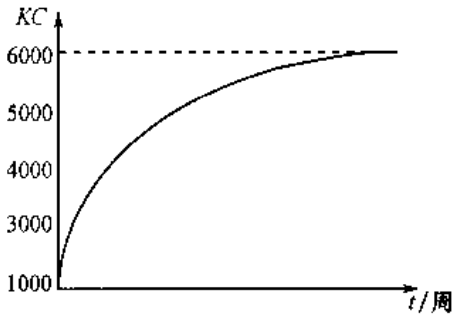


图 9.3.7　库存变化曲线

由表 9.3.1 可见，该系统由目前状态需经 24 周的时间才能趋近于目标库存量。其库存变化可用图 9.3.7 的曲线表示。

如果改变目标库存量或调整期，可得各种不同方案的模拟结果，从各种不同的模拟方案中选择一个较好的方案即可作为决策。

9.3.3　正反馈环分析

与负反馈环相反，正反馈环由于具有自我强化作用而使系统逐步背离原始状态而呈发散状态。

如某企业依靠增加投资扩大生产规模而提高产值。如果投资量为产值的 10%，产值

的增加值为投资的二倍，在企业当前产值为 100 万元时，研究该企业投资额和产值的变动情况。

该企业的发展受产值和投资额的影响，且二者构成一因果反馈环，如图 9.3.8 所示。由图 9.3.8 可见，产值的增加使投资额增加，投资额的增加又促使产值增加，该因果反馈环为一正反馈环。该正反馈环可按下式进行模拟：

$$DH_t = 10\% KC_t$$

$$KC_{t+1} = KC_t + 2DH_t$$

式中 DH_t——第 t 年投资增长量；

KC_t——第 t 年产值。

该正反馈环的模拟结果见表 9.3.2。由模拟结果可见，正反馈环所反映的系统行为是一个持续增长的过程，其增长曲线如图 9.3.9。如果不加以抑制，该过程永远不会终止。

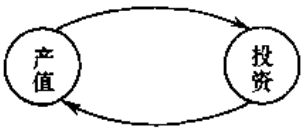


图 9.3.8 产值与投资因果反馈环

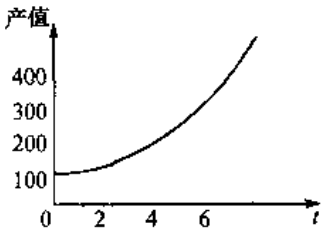


图 9.3.9 产值增长曲线

表 9.3.2 正反馈环模拟结果

时间/年	产值变化量	产值	年投资增加量
0	-	100.0	10.0
1	20.0	120.0	12.0
2	24.0	144.0	14.4
3	28.8	172.8	17.3
4	34.6	207.4	20.7
5	41.5	248.9	24.9
6	49.8	298.7	29.9
7	59.8	358.5	35.9
8	71.8	430.3	43.0
9	86.0	516.3	51.6
10	103.2	619.5	62.0

实际上这种增长过程不会无休止地进行下去，它总要受到其他因素或条件的制约而稳定在某一状态。如某一产品投放市场其需求量总是趋于某一上限的增长过程，开始投放市场时，产品深受欢迎，产品销售量很大，得到的资金也多，从而有条件扩大生产规模，进一步增加产量，这与上例基本一致。但过一段时间后，由于某种原因(或是需求已近饱和，或是生产规模扩大受到限制，或是原材料的约束等)使系统不能按上述形式发展下去而达到某一稳定水平，这是系统中其他因素的抑制作用所造成的。因此需研究正负反馈环所组成的系统。

9.3.4 正负反馈环组合的分析

某城市人口受出生量和死亡量及迁入、迁出数量的影响，其因果关系图是由正、负反馈环构成的，如图 9.3.10 所示。该人口系统的动态行为取决于人口的增长量和减少量，其实质是取决于增长量与人口数所构成的正反馈环和人口数与减少量构成的负反馈环之间的相互作用。当正反馈环的强度高于负反馈环的强度时，人口呈增长态势，当负反馈环的强度高于正反馈环时，人口呈某种稳定态势，当正负反馈环强度相同时，则人口维持初始状态。

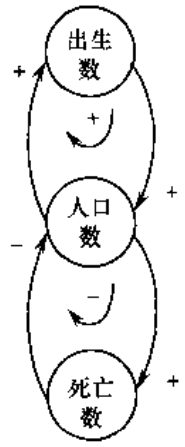


图 9.3.10 正负反馈环系统

根据上述道理，可通过增长量和减少量控制城市人口数量，通过模拟可揭示出各种人口政策的效果。

设当前该城市人口数量为 100 万人，采用当前一对夫妻一个孩子的生育政策，近五年，放开高学历、高职称人员的迁入等政策，建立的模拟模型如下：

$$RK_{t+1} = RK_t + RK_t^*(R_1 - R_2)$$

式中 RK_t——t 年的人口数；

R_1——人口增长率，$R_1=0.9\%$（前五年），$R_1=0.7\%$（五年后）；

R_2——人口减少率，$R_2=0.3\%$。

模拟结果见表 9.3.3

表 9.3.3 某城市人口数量模拟数据表

年份	增长量/万人	减少量/万人	人口数/万人
2000	—	—	100
2001	0.9	0.3	100.6
2002	0.9054	0.3018	101.2522
2003	0.9113	0.3038	101.86
2004	0.9167	0.3055	102.471
2005	0.9222	0.3074	103.0857
2006	0.7216	0.3093	103.4980
2007	0.7245	0.3105	103.9120
2008	0 7274	0.3117	104.3277
2009	0.7303	0.3130	104.7450
2010	0.7332	0.3142	105.1640

由表 9.3.3 可见采用这种人口控制政策，10 年后该城市人口增长了 5 万人，共新出生和外地迁入人口 8.2 万人，死亡和迁出人口 3.1 万。第五年后人口增长速度减缓。

通过对上述三种反馈环的分析可见，系统动态学方法是采用一组差分方程来描述系统动态行为的，且描述方法是分别对每对因果关系进行的，因此系统动态学方法具有“积木式”灵活性的特点。也正是由于该方法的这种特点比应用微分方程法（状态方程法）更

简便、更灵活地反映系统的动态行为，因此有广泛的适用性。但需指出的是，系统动态学方法是对连续系统的模拟，是采用了分块离散化的描述方法，决不是对离散系统的模拟。

9.4 系统动态学模型

系统动态学模型包括两部分内容，一是反映系统各组成部分关系的流图；二是由流图抽象出的反映系统动态过程的方程式。

9.4.1 系统流图

由于社会系统的复杂性，使人们无法用准确的语言和文字描述其行为和结构。人们曾以方程式的形式描述系统，然而它又很难直观地描述系统的结构。为了更好地掌握系统的构造和动态特性以及便于人们对系统特性的讨论和沟通，系统动态学采用一套独特的方法来描述系统，这就是流图。

系统动态学将系统当作信息反馈系统，并将该信息反馈系统的所有组成部分及其关系、各组成部分的状态以及对系统状态的控制用规定的符号和方法进行描述，经这种描述后所得到的图即是系统流图。它是在系统因果关系图基础上绘制的。

系统流图使用符号的规定及其涵义如下。

1. 积累(Level)变量

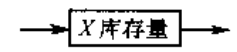


图 9.4.1 积累符号

积累变量是系统内部流的堆积，是系统状态的描述。在某个时间间隔内积累的变动量等于这个时间间隔内输入流速与输出流速之差的积累。系统内任意流动的“物”都存在积累，都可用一基本的模式计算。在流图中用矩形表示积累变量，矩形左上角为变量的名称，中间是变量的意义，指向矩形的箭线为输入流，离开矩形的箭线为输出流。以 X 为变量名的库存量可用图 9.4.1 的符号表示，其输入流为入库物资，输出流为出库物资。

积累变量的计算模式为

$$L = L_0 + DT(R_1 - R_2)$$

式中 L——积累变量的当前观测值；

L_0——积累变量的前次观测值；

DT——观测时间间隔；

R_1, R_2——输入流和输出流的流速。

2. 流速(Rate)变量

积累变量是系统活动的结果所表现出的状态，是系统状态变量。而流速是系统活动的象征，它是积累和积累之间单位时间流过的流量。每当流速发生变化，积累必然发生变化。系统的流速通常可人为调整，因此具有决策性质，体现了决策过程。决策目标、已观测的条件、目标与观测条件之差，这种差异引起的行动等都影响流速。

流速一般以图 9.4.2 的符号描述，符号中 y 为变量名，中间是变量的意义。

以出库量和入库量为流速，以库存量为积累，则库存关系可用图 9.4.3 表示。

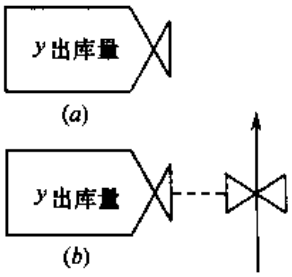


图 9.4.2　流速符号



图 9.4.3　库存模型

3. 流

流图中共有以下四种流：

(1) 物流，以符号——→表示；

(2) 订货流，以符号—○—○—→表示；

(3) 资金流，以符号—$—$—→表示；

(4) 信息流，以符号……→表示。

4. 辅助变量

辅助变量是设置在积累和流速之间信息通道上的变量。当流速表达式很复杂时，可用辅助变量描述其中一部分，使流速表达式简化。其符号为一圆圈，圈内注明变量名及意义，如图 9.4.4 所示。

5. 信息的取出

信息来自信息源。流图中用一小圆圈标在信息源上，并以信息流的符号表示从信息源取出信息及其流向。图 9.4.5 为由不同信息源取出信息的画法。

6. 常数

在系统模拟中数值不变的参数叫常数。它通常以一段实线表示，如图 9.4.6 所示。常数可直接输入给流速，也可通过辅助变量输入给流速。



图 9.4.4　辅助变量符号

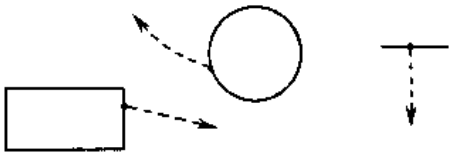


图 9.4.5　由不同信息源取出信息的符号

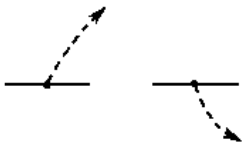


图 9.4.6　常数符号

7. 源与漏

流由系统之外流入或由系统流向环境，则外部环境称为源和漏。图 9.4.7 为源与漏的符号。

8. 其他

流图中还有一般函数、宏函数等的描述。

一般函数表示法如图 9.4.8 所示，宏函数表示法如图 9.4.9 所示，转移函数表示法如图 9.4.10 所示，其意义见 DYNAMO 模拟语言。

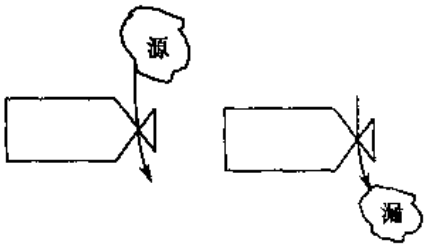


图 9.4.7　源与漏的符号

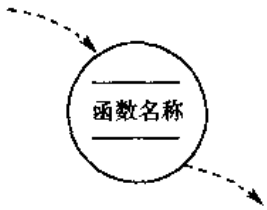


图 9.4.8　函数符号

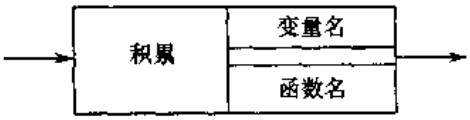


图 9.4.9　宏函数表示法

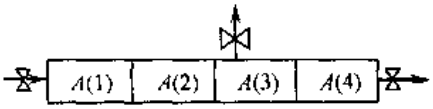


图 9.4.10　转移函数表示法

9.4.2　流图设计

系统流图设计是在因果关系图的基础上进行的，是系统分析过程的结果，流图设计过程如下：

(1) 确定系统边界；

(2) 确定系统构成要素及反馈环分析；

(3) 确定各反馈环中积累变量和流速变量；

(4) 确定积累变量与流速的关系。

现举例说明流图的设计。

某生态试验站研究羊与草构成的草原生态系统，通过研究发现羊和草之间有相互依赖关系：羊吃草，若羊的数量增加，草的数量减少，最终导致羊减少，草的数量增加，又将引起羊的数量增加。这种相互作用的结果可达到某种生态平衡，现研究羊与草的动态关系及最终的稳定状态。

首先确定系统边界。因系统是由羊与草组成的，故系统边界取羊与草。

其次确定羊与草之间的因果关系，因果关系图可由上述观察分析得出。

系统边界与因果关系图如图 9.4.11 所示。

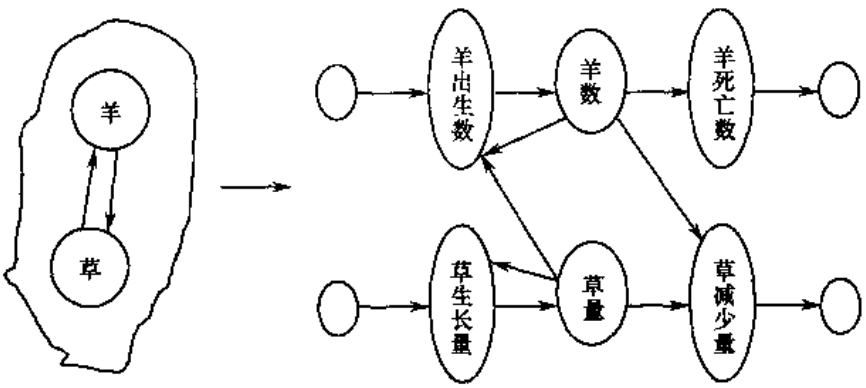


图 9.4.11　羊与草的因果关系图

再次，确定积累变量和流速变量。

系统中仅有二个反馈环，故有二个积累变量——羊的数量和草的数量，其变量名设为 X_1 和 X_2；决定积累变量状态 X_1 的量是羊的出生数和死亡数，它们是决定 X_1 的流速，故为流速变量，令其分别为 R_1 和 R_2；决定积累变量状态 X_2 的量是草的生产量和草的减少量，故它们也为流速变量，令其为 K_1、K_2。

最后根据因果关系图画出流图草图以便仔细研究流图的积累和流速结构。

积累结构比较简单，引起积累变化的仅有输入流和输出流。

流速变量结构较复杂，因流速变量实质是一决策过程，因此必须确定影响其决策的因素。

本例中影响 R_2 和 K_1 的因素较少。羊数的减少量取决于羊的积累量（羊数）和死亡率，因此需由 X_2 引出积累量的信息和观测出死亡率；草的生长量取决于现有草的数量和草的生产率，因此也需由 X_2 引出积累量的信息和观测出草的生长率。

R_1 和 K_2 则较复杂。羊的增长量取决于草的减少量，即羊所吃掉草的数量和羊、草的积累量以及每吨草所能养活（繁殖）的羊数。因此应由积累变量 X_1 和 X_2 引出有关信息和确定繁殖常量、草的减少量等信息。

K_2 是草的减少量，即羊吃掉的草量，它取决于羊的数量、草的数量和每只羊所消耗的草量（比率），故应由 X_1、X_2 引出有关信息并观测羊所消耗草量占总草量的比重，由此可得流图（见图 9.4.12）。图中 C_1、C_2 是羊的繁殖常数和死亡常数；Z_1、Z_2 是草的繁殖常数和草的减少常数。

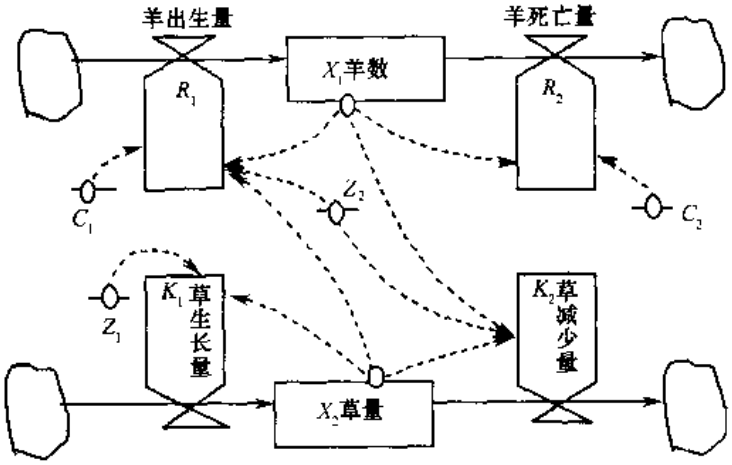


图 9.4.12　流图

9.4.3　方程式的建立

上述流图仅能简明地描述系统各要素间的逻辑关系和系统构造，但不能表明它们之间的数量关系，也就是说仅仅依靠流图还不能定量的描述系统的动态行为，建立方程式即可解决该问题。

系统动态学模型的方程式是根据系统流图建立的一连串方程式，它实质上是将反映系统动态的连续模型用一系列离散模型来代替，从而模拟连续系统的一种模型。

为了反映系统的动态特性，所建的方程组必须有时间标志，系统动态学规定：当前时刻以 K 表示，若模拟时间间隔为 DT，则 K 时刻的前一 DT 时刻为 J，后一 DT 时刻为 L，这样，JK 则表示 K 的前一时间间隔，KL 表示 K 时刻的后一时间间隔。系统动态学还规定，在积累变量和流速变量后必须标注时间，以体现时间的概念。如 $X_1.K$、$X_2.K$ 均表示 K 时刻(当前时刻)积累变量 X_1 和 X_2 的值，$X_1.J$、$X_2.J$ 和 $X_1.L$、$X_2.L$ 分别表示 K 时刻前和 K 时刻后的积累变量 X_1 和 X_2 的值。流速是反映某时间间隔内积累变量的变化，故流速变量后应标注一个时间间隔，如 $R_1.JK$ 表示 JK 时间间隔内的流速。

根据流图建立方程式时，必须对每个积累变量和每个流速变量建立方程式。

积累变量方程式有一严格的表述：

$$L.K = L.J + DT * (R_1.JK - R_2.JK) \tag{9.4.1}$$

式中 DT——模拟时间间隔；

$L.K$、$L.J$——K 和 J 时刻的积累量；

$R_1.JK$、$R_2.JK$——输入流速和输出流速；

对于流速变量则无统一的模式，必须根据控制流速的各因素之间的关系进行描述。

现以上述羊、草构成的草原生态系统为例，以图 9.4.12 为基础建立方程式。

1. 积累变量 X_1——羊的数量

当前羊的数量应为前期羊的数量与前期到当前一段时间内的增量之和，其方程式为

$$X_1.K = X_1.J + DT * (R_1.JK - R_2.JK) \tag{9.4.2}$$

2. 积累变量 X_2——草的数量

当前草的数量为前期草的数量和到当前草的增长量之和，即

$$X_2.K = X_2.J + DT * (K_1.JK - K_2.JK) \tag{9.4.3}$$

3. 草的生长量——K_1

草在 KL 期间的生长量为当前草量与生产率之积或符合某种函数关系，即

$$K_1.KL = Z_1 * X_2.K \text{ 或 } K_1.KL = f(X_2.K) \tag{9.4.4}$$

4. 羊的死亡量——R_2

羊在 KL 期间的死亡量为当前羊的数量和死亡率的积或符合某种函数关系，即

$$R_2.KL = C_2 * X_1.K \text{ 或 } R_2.KL = f(X_1.K) \tag{9.4.5}$$

5. 草减少量——K_2

草的减少量为每只羊的食草量占总草量的比率与草量和羊数之积，即

$$K_2.KL = Z_2 * X_1.K * X_2.K \tag{9.4.6}$$

这是因为 $Z_2 * X_1.K$ 为现有羊所吃草的数量占草量的比重，故 $Z_2 * X_1.K * X_2.K$ 为草的减少量，即是羊吃掉的草量。

6. 羊的增加量——R_1

羊的增加量取决于草的数量、羊的数量、每消耗一吨草所能养活(繁殖)的羊数以及草的减少常数，即

$$R_1.KL = C_1 * Z_2 * X_1.K * X_2.K \tag{9.4.7}$$

这是因为后三个数之积为草的减少量，C_1 为每减少 1t 草(羊吃 1t 草)可繁殖的羊数，故上式反映了羊的出生量。

有了上述方程式后，给定初始值、各种常数和模拟时间间隔即可进行动态模拟。

如当　　$C_1=0.5$，$Z_2=0.01$，$X_1=0.5$，$X_2=5$

$$K_1=f(X_2.K)=-0.1(X_2.K)^2+1.5(X_2.K)$$

$$R_2=f(X_1.K)=e^{0.1(x_1.K)}-1$$

该系统的模拟结果如见表 9.4.1。

表 9.4.1　系统模拟表

时间/年	羊数 X_1/万只	草数 X_2/万 t	羊出生数 R_1/万只	羊死亡数 R_2/万只	草生长量 K_1/万 t	草减少量 K_2/万 t
0	0.5000	5.000	0.0125	0.0513	5.000	0.025
1	0.4612	9.975	0.023	0.0472	5.012	0.046
2	0.437	14.941	0.032	0.0447	0.088	0.065
3	0.424	14.964	0.032	0.0430	0.054	0.063
4	0.413	14.955	0.031	0.0420	0.067	0.062
5	0.402	14.940	0.030	0.0410	0.090	0.060
6	0.391	14.970	0.029	0.0400	0.045	0.059
7	0.380	14.960	0.028	0.0380	0.060	0.056
8	0.370	14.960	0.027	0.038	0.060	0.055
9	0.359	14.970	0.027	0.037	0.060	0.052
10	0.349	14.964	0.026	0.033	0.060	0.052
11	0.339	14.970	0.025	0.035	0.045	0.051
12	0.329	14.960	0.025	0.033	0.06	0.049
13	0.321	14.970	0.024	0.033	0.045	0.048
14	0.312	14.970	0.023	0.032	0.045	0.047
15	0.303	14.97	0.023	0.030	0.045	0.045
16	0.296	14.97	0.022	0.030	0.045	0.044
17	0.288	14.97	0.022	0.029	0.045	0.043
18	0.281	14.97	0.021	0.028	0.045	0.042
19	0.274	14 97	0.021	0.028	0.045	0.041
20	0.267	14.98	0.020	0.028	0.030	0.040
21	0.259	14.99	0.019	0.026	0.045	0.039

由上述例题可见，系统动态学模型的基础是绘制流图，然后则是根据流图建立方程式，建立方程式中的关键和难点又是确定流速的结构。

另外还需指出的是，模拟的时间间隔对模拟结果的影响很大，因此要适当选择模拟时间间隔。

从上例的模拟结果可见，该系统 20 年后间的数量基本稳定在 2.5 万只左右，而第五年左右开始，基本是自然死亡使羊的总量每年递减，草量基本维持平衡。将羊数和草量的变动用曲线描述可得其变化趋势(见图 9.4.13 和图 9.4.14)。

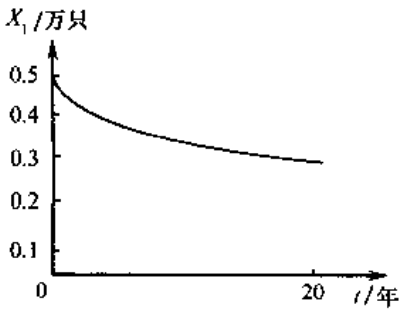


图 9.4.13　羊数变化曲线

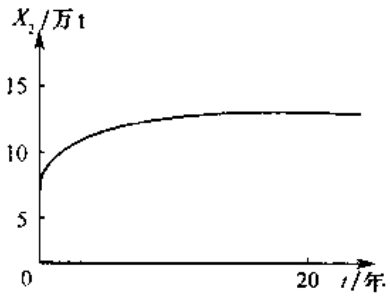


图 9.4.14　草的变化曲线

这种模拟方法还可考虑各种延迟的因素，故在处理问题时是很方便的。

第 10 章　系统决策

系统决策是指在一定的条件下，根据系统的状态在可采取的各种策略中，依据系统目标选取一个最优策略并付诸实施的过程。

以前所介绍的最优化方法和模拟试验方法都是定量地研究系统决策的方法，它们是进行决策的主要依据之一。这些决策方法都有一个共同的特点，这就是在分析问题时未考虑不确定因素，也就是研究对象为确定型输入和输出的系统。然而很多系统都受随机因素的干扰，因此在选择方案时，决策者尚需根据自己的经验、知识和智慧，考虑随机因素进行主观判断和逻辑判断，进而做出决策，以达到预期的目的，这种以经验为基础的决策叫经验决策。

科学的决策不同于经验决策，它是在对系统进行科学分析的基础上，运用科学思维方法，采用科学的决策技术而做出有科学依据决策的过程。本章着重介绍决策过程中经常使用的决策技术。

10.1　决 策 树

10.1.1　概率树

应用树来表示一系列连续发生的事件，以树枝表示事件发生的概率，则这棵树叫概率树。

例如，在相同条件下，做若干次重复试验，成功 S 次的概率服从二项式分布。若成功 S 的概率为 P，失败 F 的概率为 q，并二事件为互斥事件，现做 3 次试验，可得 8 种结果，这 8 种结果出现的概率可用概率树清楚地表示出来。

以树枝表示事件发生的概率，以△表示事件的结果，以○表示试验，则概率树如图 10.1.1 所示。

由图 10.1.1 可见，每种结果实现的概率可用树的顶点到该结果所构成的链的所有弧的概率积来表示，用该方法即可确定连续发生事件的概率。

某城市气象部门，每天都发出火灾危险信号，以提醒居民，信号分为三种：

(1) 白色信号(W)：表示空气湿度高，不易发生火灾。

(2) 黄色信号(Y)：表示天气干燥，需要注意防火。

(3) 红色信号(R)：表示天气相当干燥，需特别注意防火。

若当天发出一种信号，第二天发出另种信号的概率见表 10.1.1，现需确定今天发出白色信号，3 天后发出各种信号的概率为多少。这个问题应用概率树很容易解决。

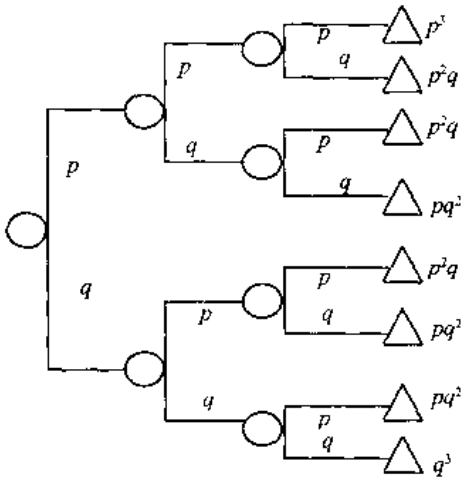


图 10.1.1　概率树

表 10.1.1　发出信号的概率

	W	Y	R
W	$\frac{2}{3}$	$\frac{1}{3}$	
Y	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
R	$\frac{2}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

3 天后的信号是今天信号经过 3 次变化的所有可能中的一种，其概率树如图 10.1.2 所示。3 天后出现：

(1) 白色信号的概率：

$$P_{(W)}=P(X_1)+P(X_6)+P(X_8)+P(X_{11})=\frac{641}{1080}\approx 0.5935$$

(2) 黄色信号的概率：

$$P_{(Y)}=P(X_2)+P(X_4)+P(X_7)+P(X_9)+P(X_{12})=\frac{677}{2160}\approx 0.3134$$

(3) 红色信号的概率：

$$P_{(R)}=P(X_5)+P(X_{10})+P(X_{13})=\frac{67}{710}\approx 0.0931$$

这样就得出了 3 天以后出现各种信号的可能性。

10.1.2　决策树

在概率树的树枝上加上各事件相应的费用、成本等数量指标，则表示出由于采取不同方案的不同支出，这种树叫决策树。

如市政工程公司承包一公路的修建任务，合同规定 8 月 1 日开工，8 月 31 日交工。超期 5 天内，每天罚款 400 元，由第六天开始每天罚款 600 元。按工程公司所掌握的气象资料，从 8 月 16 日以后，天气要发生变化，只有 40% 的可能遇到好天气，可以按期完工，50% 要遇到小雨，会使工期延长 5 天，而有 10% 的可能会遇大雨使工期延长 10 天。如果

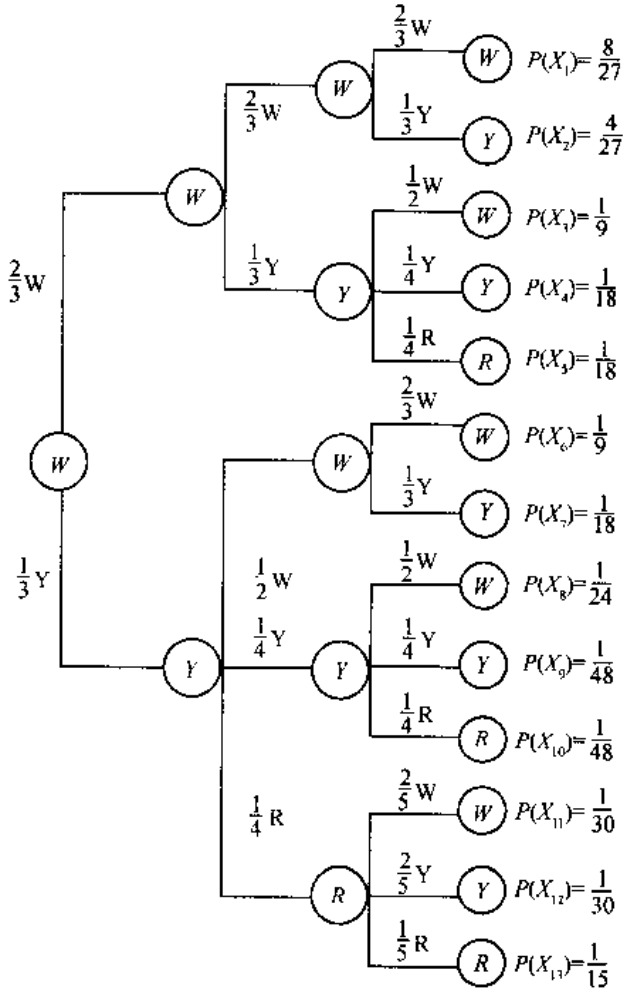


图 10.1.2　火灾信号变化的概率树

采用加班加点的方法，可使工程提前到 16 日前完成，但每天要付出 75 元的加班费；而采取应急措施，每延期一天需付出 200 元的应急费用。如遇到小雨则提前一天的可能是 50%，提前二天的可能是 30%，提前三天的可能是 20%；如遇大雨提前二天的可能是 70%，提前三天的可能是 20%，提前四天的可能是 10%。各种情况所需的追加费用是：

不采取应急措施，且遇小雨：$400\times5=2000$ 元

采取应急措施，且遇小雨：

提前一天所应多支出的用：$1600+800=2400$ 元

提前二天所应多支出的费用：$1200+600=1800$ 元

提前三天所应多支出的费用：$800+400=1200$ 元

不采取应急措施，且遇大雨：$2000+3000=5000$ 元

采取应急措施，且遇大雨：

提前二天应多支出的费用为：$1600+(2000+1800)=5400$ 元

提前三天应多支出的费用为：$1400+(2000+1200)=4600$ 元

提前四天应多支出的费用为：$1200+(2000+600)3800$ 元

采取加班的办法:75×15=1125 元。

因此,可得到一个带成本费用的概率树,即决策树,如图 10.1.3 所示。

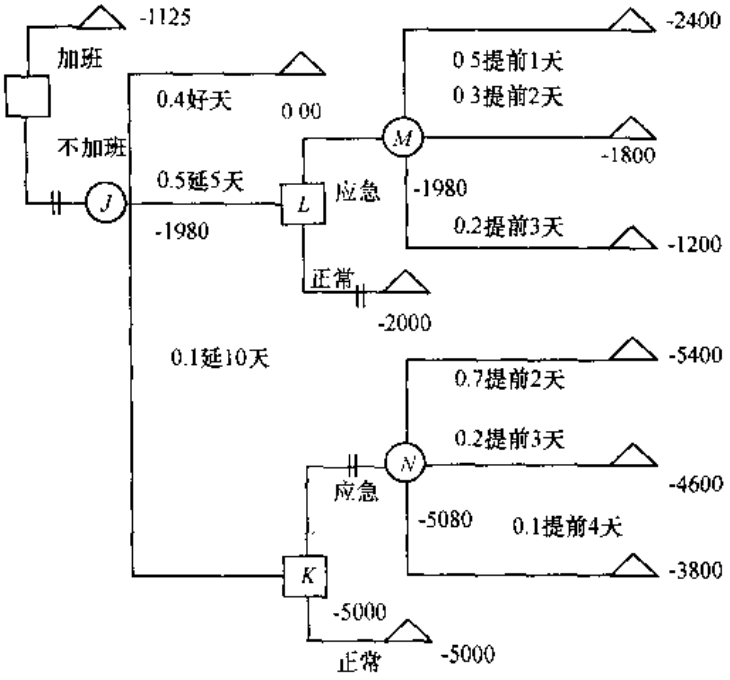


图 10.1.3 决策树

10.1.3 应用决策树做决策分析

得到带费用指标的决策树并不是我们的目的,而我们的目的是应用这种决策树进行分析以选择一个比较理想的方案。

10.1.3.1 价值期望法

这是一种决策者只根据客观情况进行决策的办法,这种方法叫价值期望法。

该方法的基本思想就是确定方案节点的价值期望值,并以此为根据进行决策。

所谓价值期望值是指以概率为权数的价值的加权平均值。

$$E(i) = \sum P_{ik}C_{ik} \tag{10.1.1}$$

表示第 i 个概率节点的价值期望值,其中 ik 是第 i 个概率节点的第 k 个事件,C_{ik} 是 i 概率节点的第 k 个事件相应的费用。因此图 10.1.3 中所有用圆圈表示的概率节点都可计算出价值期望值:

$$E(M)=0.5\times(-2400)+0.3\times(-1800)+0.2\times(-1200)=-1980 \text{ 元}$$

$$E(N)=0.7\times(-5400)+0.2\times(-4600)+0.1\times(-3800)=-5080 \text{ 元}$$

在决策节点 L 点和 K 点,如果以支出为负,收入为正,则

$$E(j) = \max\{E(i)\} \tag{10.1.2}$$

因此:

$$E(L)=\max\{-5080,-5000\}=-5000$$

$$E(K)=\max\{-2000,-1980\}=-1980$$

故 j 节点：$E(j)=0.4\times0+0.5\times(-1980)+0.1\times(-5000)=-1490$

A 节点：$E(A)=\max\{-1125,-1490\}=-1125$

因此，比较好的方案就是一开始就采用加班加点的方法使工期提前到 8 月 16 完成，只需多花费 1125 元的加班费，而其他方案需支出更高的代价。因此选该方案比其他方案都优越。

应用决策树可以直接在图上进行，其步骤如下：

(1) 从树的梢部开始，计算最近概率节点的价值期望值；

(2) 比较由同一决策点出发的各概率节点的价值期望值，选其成本最低或价值最高的决策方案，并标于该决策点上；

(3) 这样一级一级地向树根部推进，最后得到根部决策点的价值期望值；

(4) 根据决策的价值期望值，做出决策，一级一级地向后推，找出所选择的决策和所需的费用。

这种方法完全是由数理统计的计算而得出决策的，而在计算中应用了期望值的概念，如果选中了一条概率型的决策，在某次执行中由于事件的随机性并不一定得到与计算相同的结果，只有在多次反复的执行中，才能得到这一平均值。因此，这种方法只是供决策者参考的依据之一。

10.1.3.2 效用决策法

网络值期望值法决策除概率的估计值之外不掺杂任何个人的主观见解，完全依据客观条件进行决策。但由于决策者所处的地位不同、决策角度以及价值观念不同，通常都把上述方法当做参考。

图 10.1.4 所示的决策树，如按上述方法应选择 A 方案，这样不是得到 100000 元的利润，就是亏损 20000 元，而且亏损 20000 元的可能性很大。

对一个资本雄厚的决策者来说，宁愿选 A；但对一个资本很少的决策者来说虽然他渴望得到 100000 元的利润，但却承受不了 20000 元亏损的打击，因此它只好选择旱涝保收的 B 方案。

这就说明，在决策过程中，决策者要依据自己的价值准则进行决策，要把自己的实际情况和科学方法相结合。而价值期望法只是客观地反映了平均水平，而不能反映决策者的意志，为了在决策中反映决策者的意志应采用效用理论。

为了体现决策者的主观作用，可把费用、价值等指标转换为“效用”。效用是决策者价值观念的反映。每个决策者都有自己的“效用”函数，这个函数反映出决策者的冒险程度和主观意识。“效用”函数用曲线表示，叫效用曲线。

如图 10.1.4 的决策树，可利用图 10.1.5 的效用曲线得到图 10.1.6 的决策树。按照价值期望法即可得出符合资本少的人的利益的方案，因此这种方法体现了决策者的主观见解。

确定效用曲线是不容易的，它是随时间和条件、随决策者而不同的，故只能在自己多年经验的基础上，采用自己认为合适的“效用”曲线。

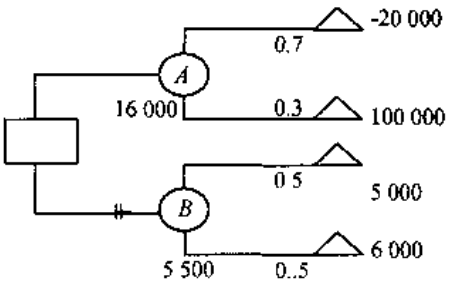


图 10.1.4 决策树

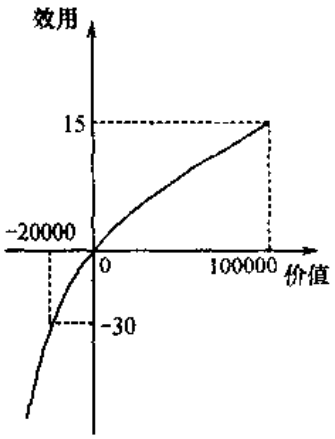


图 10.1.5　效用曲线

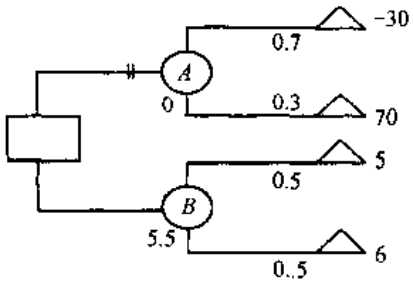


图 10.1.6　利用效用曲线得到的决策树

10.2　风险决策和不确定决策

非确定型决策分为风险决策和不确定决策两种。

10.2.1　风险决策

所谓风险决策是指在不确定因素的概率已知的情况下，无论选择哪种决策都要承担一定风险的决策。这种决策通常必须具备以下条件：

(1) 决策者有明确的目标；

(2) 有二种以上决策者无法控制的不确定因素；

(3) 有二个以上方案可供决策者选择；

(4) 可估计出不同方案在各种不确定因素下的损益值；

(5) 可估计出各种不确定因素出现的概率。

风险决策的主要方法是收益矩阵法。收益矩阵法的基本原理与决策树法相同，只是决策过程中采用损益矩阵的形式。

所谓损益矩阵是以决策方案为主栏，各种不确定因素为宾栏、各方案与各不确定因素对应的损益为元素的矩阵，其一般形式见表 10.2.1。

表 10.2.1　损益矩阵

因素 / 概率 / 收益值 / 方案	θ_1	θ_2	…	θ_m
	P_1	P_2	…	P_m
A_1	a_{11}	a_{12}	…	a_{1m}
A_2	a_{21}	a_{22}	…	a_{2m}
⋮	…	…	…	…
A_N	an_1	a_{n2}	…	a_{nm}

如某建筑公司在组织施工时,由气象部门得到降水状况预报为:0.4 的概率为无雨,0.5 的概率为小雨,0.1 的概率为大雨。现该公司准备了三套生产安排方案:A_1、A_2 和 A_3。三种方案在三种天气所对应的损益矩阵见表 10.2.2。

表 10.2.2　三种方案在三种天气所对应的损益矩阵

因素及概率 / 方案	无雨	小雨	大雨	期望值
	0.4	0.5	0.1	
A_1	150	−70	−100	15
A_2	0	100	110	61
A_3	80	100	70	89

用损益矩阵采用期望值法进行决策的步骤为:

(1)按 $E(A_i)=\sum a_{ij}p_j$ 计算出各方案的期望值 $E(A_i)$:

$$E(A_1)=150\times0.4+(-70)\times0.5+(-100)\times0.1=15$$

$$E(A_2)=100\times0.5+110\times0.1=61$$

$$E(A_3)=80\times0.4+100\times0.5+70\times0.1=89$$

(2)选 $\max\{E(A_i)\}$ 所对应的方案作为决策方案

$\max\{15,61,89\}=89$,对应于 A_3 方案,故选 A_3 方案为决策方案。

该方法可应用效用理论将损益矩阵元素用效用代替,以体现决策者的价值观和决策能力。

损益矩阵法可用决策树法代替,但决策树法所解决的问题有时用损益矩阵法求解会更复杂和麻烦。

10.2.2　不确定型决策

不确定型决策与风险决策的区别仅在于对不确定因素的发生概率不能做出估计。因此这种决策有完全的不确定性,故称不确定型决策。这种决策方法也是从损益矩阵出发的,所不同的是损益矩阵中没有概率一行,其一般形式见表 10.2.3。

表 10.2.3　不确型决策

因素 / 收益 / 方案	θ_1	θ_2	…	θ_j	…	θ_m
A_1	C_{11}	C_{12}	…	C_{1j}	…	C_{1m}
A_1	C_{21}	C_{22}	…	C_{2j}	…	C_{2m}
⋮	⋮	⋮	…	⋮	…	⋮
A_i	C_{i1}	C_{i2}	…	C_{ij}	…	C_{im}
⋮	⋮	⋮	…	⋮	…	C_{im}
A_n	C_{n1}	C_{n2}	…	C_{nj}	…	C_{nm}

根据不同要求,决策者可采用如下决策方法和原则进行决策。

10.2.2.1　小中取大法则

这种法则的思想是从各种方案在最不利条件下的收益中选取收益最大的一个方案。

即

$$\max_{A}\{\min_{\theta}\{C(A_i,\theta_j)\}\} \tag{10.2.1}$$

也就是在最不利的条件下得到一个最大的收益。如上例中，假设无气象资料，如选 A_1 方案，则最坏是损失 100 万元；如选 A_2 方案最坏是没收益，即收益为零；如选 A_3 方案最最坏收益 70，即

$$\min\{150,-70,-100\}=-100$$
$$\min\{0,100,110\}=0$$
$$\min\{80,100,70\}=70$$

在三者中选最大者：$\max\{-100,0,70\}=70$。因此应选 A_3 方案。在损益矩阵上直接进行计算，见表 10.2.4。

表 10.2.4　小中取大法则计算结果

因素 / 收益 / 方案	无雨 θ_1	小雨 θ_2	大雨 θ_3	本方案中最小收益 $\min C(A_{A_i},\theta_j)$
A_1	150	−70	−100	−100
A_2	0	100	110	0
A_3	80	100	70	70

这种方法是保守的、悲观的，但是最可靠的。因此，这种方法一般适于要求可靠性高的情况，如预测洪水、地震等进行方案选择时多采用此方法，目的是留有充分的余地，减少损失。

10.2.2.2　大中取大法则

这种方法与小中取大法则正好相反，它的主导思想是宁可承担最大损失的风险，也要为达到最大收益而冒险，因此，它总是在一切如意的情况下进行决策的。其方法是在各种方案中选出最大收益，然后再从中选择最大值，该值所对应的方案即是决策方案。

用数学语言可描述为

$$\max_{A}\{\max_{\theta}\{C(A_i,\theta_j)\}\} \tag{10.2.2}$$

其计算见表 10.2.5

表 10.2.5　大中取大法则计算结果

因素 / 收益 / 方案	无雨 θ_1	小雨 θ_2	大雨 θ_3	本方案中最大值
A_1	150	−70	−100	150
A_2	0	100	110	110
A_3	80	100	70	100

这种方法选择 A_1 方案。可见如果实现此方案，可得到最大收益，但这种决策方法带有很大的冒险性，不是收获最大，就是损失最大，因此，不适于要求可靠性高的决策。

10.2.3　比例法

比例法是上述两种方法的折衷。它的思想是把连接每一方案的最大值点和最小值点

的直线分成比例线段，用比例系数的变动来确定其收益，然后在所有方案的收益中选择收益值最大的方案。

现简单介绍一种线段方程的表示方法。

线段 ab 上任一点 L，可用下式表示

$$\alpha a+(1-\alpha)b=L \qquad 0\leqslant\alpha\leqslant 1$$

比例系数 α 把线段分成两段，分点为 L。当 $\alpha=0$ 时，分点 L 在 b 点上；当 $\alpha=1$ 时，则分点在 a 上；如果 $0<\alpha<1$，则分点 L 在线段 ab 上。

根据这个道理，我们依下式确定每一方案的收益值供决策：

$$H_i=\alpha V_{大}+(1-\alpha)V_{小} \qquad 0\leqslant\alpha\leqslant 1$$

式中 H_i——供决策的方案收益；

$V_{大}$——该方案中最大收益；

$V_{小}$——是该方案中最小收益；

α——比例系数。

因此，上例中如选 $\alpha=0.5$ 时

$$H_1=0.5\times 150+0.5\times(-100)=20$$

$$H_2=0.5\times 110+0.5\times 0=55$$

$$H_3=0.5\times 100+0.5\times 70=85$$

按 $\max\{H_i\}$ 来确定方案，即确定选 A_3 方案。

如选 $\alpha=0.1$ 时，则

$$H_1=0.1\times 150+0.9\times(-100)=-75$$

$$H_2=0.1\times 110+0.9\times 0=11$$

$$H_3=0.1\times 100+0.9\times 70=73$$

$\max\{H_i\}=\max\{-75,11,73\}=73$ 应选择 A_3 方案。

由该方法可见，当 $\alpha=0$ 时，则是小中取大法；当 $\alpha=1$ 时，是大中最大法。因此，上述两种方法都是这种方法的特例。比例法是一种特殊的平均法。该方法中 α 的选择是重要的，它体现出决策者的冒险程度。

10.2.4 平均值法

平均值法是用每一方案收益的算术平均值作为该方案供决策的收益值(实质上是把各种收益出现的概率视为相等)，然后再取其最大值作为决策方案。可表示为

$$\max\left\{\frac{1}{n}\sum_{j=1}^{n}C(A_i,\theta_l)\right\} \qquad (i=1,2,\cdots,n)$$

上例中：

$$\frac{1}{3}\times[150+(-70)+(-100)]=-6.67$$

$$\frac{1}{3}\times[0+100+110]=70$$

$$\frac{1}{3}\times[80+100+70]=83.33$$

可见按此法应选 A_3 方案。

10.2.5 最小后悔值法

顾名思义,这种方法就是在方案选择之后,在客观情况与主观愿望相违背时,使后悔最小的一种方法。为了达到后悔最小的目的,假定每种方案的最大收益是理想的,如果没选择该方案将后悔,则后悔的数值是其他方案与本方案最大收益的差,所选的方案应是在每种方案中最大的后悔值中选择一最小的,也即是后悔值最小,故该方法又称最大最小法。

上例中用最小后悔值法计算结果见表 10.2.6,结果选中 A_3 方案。

表 10.2.6 最小后悔值法计算结果

方案 \ 天气 θ	无雨 θ_1	小雨 θ_2	大雨 θ_3	最大后悔值
A_1	150 150=0	100+70=−170	110+100=210	210
A_2	150−0=150	100−100=0	110−110=0	150
A_3	150−80=70	100−100=0	110−70=40	70

以上介绍的是几种常用的不确定型决策方法,因侧重面不同,决策者的立场不同,因而选择的方法也不同,从而也就得出不同的决策,收到不同的效果,故在选用这些方法时,不能死搬硬套某种方法,应该领会各种方法的思想,依据实际情况,分析社会及自然条件去决策。对于一些有特殊要求的事件,要特别慎重考虑,要充分地估计最不利的因素,而使意外的损失最小。决策问题是一个政策性很强、涉及面很广的问题,决不能掉以轻心,草率从事。在能够收集到资料的情况下,要尽力把材料收集齐全而采用数理统计和分析的方法去处理,一般能收到比较满意的结果。

10.3 多目标决策

在实际生产与生活中,对一事物的决策不能仅仅依靠一个指标,而必须同时考核多种目标。如在生活中,到商店去挑选商品总是要求物美、价廉;评价一个企业,需根据各种经济技术指标进行综合评定。

10.3.1 多目标决策的基本概念

一般情况下,评价系统的各种目标都是从不同侧面反映系统的,很难用其中一个指标来代替其他。因此,这类问题一般无最优解可取。故在处理多目标问题时,多采用非劣解和选优解两个概念。

10.3.1.1 非劣解和选优解

在处理有 n 个目标 $f_1(X),\cdots,f_n(X)$,$X=\{X_1,X_2,\cdots,X_n\}$,使得所有的目标都最大,即 $\max f_1(X)$, $\max f_2(X)$, $\cdots$, $\max f_n(X)$ 的问题中,总存在着非劣解 X^*,对于所有的 X,当 $i=1,2,\cdots,n(i\neq j)$ 时,都有 $f_i(X)\leqslant f_j(X^*)$,至少存在一个 j,使 $f_j(X)>$

$f_i(X^*)$，则 X^* 叫非劣解。通俗地说，就是能使一些目标达到极大值，而对另一些目标则不一定能达极大值的 X^* 叫做非劣解。

在所有非劣解中按照一定的法则选择一个比较好的解，作为决策，这个解称为“选优解”。

10.3.1.2 “选优解”的确定

确定非劣解的过程是决策的分析过程，而在非劣解中选择“选优解”的过程是决策的决定过程。多目标决策都是由分析过程和决定过程两部分组成的。分析过程一般由分析者即系统分析人员完成，而决定阶段都是由决策者即领导者来完成。虽然两个阶段可明显地分开，但分析者必须提供足够的信息供领导决策才能得到更好的效果。一般采用如下方法：

(1) 决策者和分析者共同商定出一种方法和原则，按此原则和方法直接求“选优解”；

(2) 分析者只提供非劣解，由决策者决定选优解；

(3) 决策者和分析者不断相互交换意见，不断改进非劣解，直到最后得出决策者满意的“选优解”为止。

第一种方法简单明了，只要确定了原则，问题就解决了，但原则的确定不是十分容易的事情。

第三种方法是比较可取的方法，然而需要作深入细致的思想工作和相互协调的分析工作，使分析者和决策者意见逐步趋于一致，这种方法决策的时间是较长的。

10.3.2 多目标决策的具体方法

上面介绍了处理多目标决策问题的原则问题，但在具体处理问题中，必须根据这些原则再加上一些数学方法，才能使多目标决策问题定量化，才能更有利于决策。现介绍几种常用的多目标决策的定量方法。

10.3.2.1 化多目标为单目标法

在多目标问题中求出满足全部目标且使其都最优的 X 是不可能的。然而利用一些数学方法，经过一定的处理，变多目标为单目标，就可利用所学过的处理单目标最优化问题的方法去解决。

基于这种思想，根据问题的不同性质，采用不同的方法，即可把复杂的多目标问题化为单目标问题。

1. 加权平均和法

当目标函数 $f_1(X)$，$f_2(X)$，…，$f_n(X)$ 都要求最小(或最大)时，可引入加权乘子构成新的目标函数：

$$\min \quad V(X) = \sum_{i=1}^{n} \lambda_i f_i(X) \tag{10.3.1}$$

加权乘子 λ_i 的确定直接影响到决策的结果，因此，选择 λ_i 要有充分的经验或用统计调查的方法得出。一般常采用 AHP 法或价值工程中的评分法、比率法等。该方法要求各目标的数量级和量纲相同，所形成的新目标有一定的经济意义。

2. 数学规划法

这种方法是从多目标中选择一个最重要的目标 $f_i(X)$，使它满足最大或最小，而使其

他的所有目标满足 $f' \leqslant f(X) \leqslant f''$，从而构成了一个以重要目标 $f(X)$ 为单目标，以其余目标为约束的一个数学规划问题，即

$$\begin{aligned} &\max \quad f(X) \\ &\text{s.t.} \quad f'_i \leqslant f_i(X) \leqslant f''_i, \qquad i = 2,3,\cdots,n \end{aligned} \tag{10.3.2}$$

例如某建筑公司以产值、成本、劳动生产率、能源消耗水平作为评价指标。在评价时，可以把上述指标化成以产值为主指标，对其他指标都给予一定限制的数学规划问题，从而可得到如下数学规划：

$$\begin{aligned} \max \quad & f_1(X) && \text{产值最高} \\ \text{s.t.} \quad & f_2(X) \leqslant b_1 && \text{成本小于规定值} \\ & f_3(X) \geqslant b_2 && \text{劳动生产率高于一定值} \\ & f_4(X) \leqslant b_3 && \text{能源消耗低于一定水平} \\ & \cdots \\ & AX = b && \text{原问题约束} \end{aligned}$$

求解后就可以得到一个比较理想的决策。

3. 目的规划法

这种方法的基本思想是为所有目标确定一个预期达到的目标值 $f_i^*(X)$，使做出的决策与该值越接近越好，如完全符合此目标，则是最优解；如不符合，则以离差平方和的大小来衡量其偏离预期值的程度，从而把目标函数 $f_1(X)$，$f_2(X)$，$\cdots$，$f_n(X)$ 化成

$$\min \quad V(X) = \sum_{i=1}^{n} [f_i(X) - f_i^*(X)]^2$$

按这种思想还可根据目标的重要程度而采用加权目的规划法，即

$$\min \quad V(X) = \sum_{i=1}^{n} \lambda_i [f_i(X) - f_i^*(X)]^2 \tag{10.3.3}$$

现举例说明如何用目的规划法求解多目标问题。

例 评价某单位工作业绩的三个指标是成本、产值和污染，成本函数为 $f_1(X) = 2X$，产值函数为 $f_2(X) = 4X$，污染函数为 $f_3(X) = X$。先假设预期成本为 2，产值为 16，污染为 0.5，则 $X = 1$，$X = 4$，$X = 0.5$ 可分别使各自的目标达到预期值。可见，不存在使三个指标同时达到预期值的 X，不存在最优解。故利用目的规划法建立评价函数：

$$\min \quad V(X) = \sum_{i=1}^{3} [f_i(X) - f_i^*(X)]^2 = (2X-2)^2 + (4X-16)^2 + (X-0.5)^2$$

求 $V(x)$ 的极小值：

$$\frac{\partial V(X)}{\partial X} = 4(2X-2) + 8(4X-16) + 2(X-0.5) = 42X - 137 = 0$$

$$X = 137/42, f_1(X) = 137/21, f_2(X) = 274/21, f_3(X) = 137/42$$

则 $X = 3.262$ 即是该问题的选优解。

再如：某建筑公司木制品厂，生产两种产品，一种粗制品，一种精制品，精制品供出口，以求外汇，因此产量越多越好，两种产品用料相同，均为 $1m^3$/件，但精制品却是粗制品用工量的 2 倍（粗制品用工量为一工日），精制品单价为 0.4 千元，粗制品为 0.3 千元，该厂现有 $400m^3$ 木材，500h 工时，现要求出口越多越好，且工厂的产值越高越好。应如何安排

生产?

首先建立该问题的目标函数。令 X_1、X_2 分别为两种产品产量,则模型为

$$\begin{aligned} \max \quad & f_1(X) = 0.4X_1 + 0.3X_2 \\ \max \quad & f_2(X) = X_1 \\ \text{s.t.} \quad & X_1 + X_2 \leqslant 400 \\ & 2X_1 + X_2 \leqslant 500 \\ & X_1 \geqslant 0, X_2 \geqslant 0 \end{aligned}$$

现采用目的规划法。现分别求出对应于每个目标函数的最优值,并以此最优值作为预期值。

先按模型

$$\begin{aligned} \max \quad & f_1(X) = 0.4X_1 + 0.3X_2 \\ \text{s.t.} \quad & X_1 + X_2 \leqslant 400 \\ & 2X_1 + X_2 \leqslant 500 \\ & X_1 \geqslant 0, X_2 \geqslant 0 \end{aligned}$$

求得 $X_1 = 100, X_2 = 300, f_1^0(X) = 130$ 千元

再按模型

$$\begin{aligned} \max \quad & f_2(X) = X_1 \\ \text{s.t.} \quad & X_1 + X_2 \leqslant 400 \\ & 2X_1 + X_2 \leqslant 500 \\ & X_1 \geqslant 0, X_2 \geqslant 0 \end{aligned}$$

求得 $X_1 = 250$, $X_2 = 0, f_2^0(X) = 250$ 个

以这两个最优值为预期值建立新的目标函数 $V(X)$:

$$\min \quad V(X) = [(0.4X_1 + 0.3X_2) - 130]^2 + (X_1 - 250)^2$$

这就把两个目标的问题化成了单目标问题。求解单目标有约束的非线性规划问题:

$$\begin{aligned} \min \quad & V(X) = [(0.4X_1 + 0.3X_2) - 130]^2 + (X_1 - 250)^2 \\ \text{s.t.} \quad & X_1 + X_2 \leqslant 400 \\ & 2X_1 + X_2 \leqslant 500 \\ & X_1 \geqslant 0, X_2 \geqslant 0 \end{aligned}$$

即可得到尽量充分利用原材料和工时的最优解。应用罚函数法得

$$X_1 = 244, X_2 = 12$$

$$f_1(X) = 101.2, f_2(X) = 244$$

这种把多目标问题化成单目标问题的目标规划法的几何解释是一以预期值为圆心的空间球体到可行集面的最小距离即是目标值,切点就是该问题的解。二维情况见图 10.3.1。

这是因为新目标函数是一个圆的方程的缘故。

$f_1^*(X)$和 $f_2^*(X)$是两个预期值,而 V 则是以 $f_1^*(X)$、$f_2^*(X)$为圆心、以 $f_1(X)$和$f_2(X)$为变量的圆,因此,在等 V 线上交于可行区边缘的点,则是问题的解。同理也可将该道理推广到三维以至 n 维空间。

4. 费用效果分析法

通常情况下,系统目标 $f_1(X), \cdots, f_n(X)$可分为两大类,一类是费用型目标,如成本、

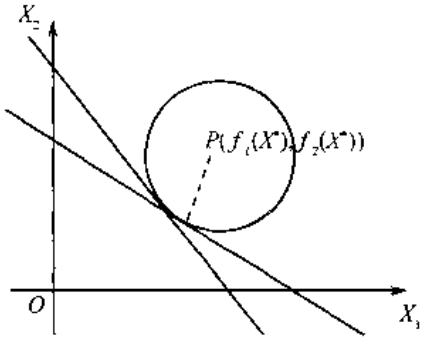


图 10.3.1　化多目标为单目标的几何解释

费用等;另一类是效果型目标,如利润、产值等。不失一般性,设前m个目标为费用型目标,后$n-m$个为效果型目标,若前者以$C(X)$表示,后者以$\mu(X)$表示,从经济效益最大的角度研究,则以最小的费用得到最大的效果可作为评价系统的主要指标。

如用绝对效益指标可将该多目标问题化作如下单目标:

$$\max \quad V(X)=\mu(X)-C(X)=\sum_{i=m+1}^{n} f_i(X)-\sum_{j=1}^{m} f_j(X)$$

如用相对效益指标,可得如下单目标:

$$\max \quad V(X)=\frac{\mu(X)}{C(X)}=\frac{\sum_{i=m+1}^{n} f_i(X)}{\sum_{j=1}^{m} f_j(X)}$$

5. 满意度法

若系统有多个目标,且每个目标都可确定一上、下界$f_i^H(X)$和$f_i^L(X)$,且$f_i^H(X)$是希望达到的目标;$f_i^H(X)$和$f_i^L(X)$的差表示该目标允许变动的范围,该范围越小,说明越靠近所追求的目标,用数学描述可表示为

$$\min \quad f_i^H(X)-f_i^L(X)$$

$f_i(x)$与$f_i^L(x)$的差表示实际目标值与下界的变动范围,该范围越大,表明该范围越靠近所追求的目标,用数学描述可表示为

$$\max \quad f_i(X)-f_i^L(X)$$

这样,按费歇尔原则可定义一目标的满意度λ_i:

$$\lambda_i=\frac{f_i(X)-f_i^L(X)}{f_i^H(X)-f_i^L(X)}$$

若对所有目标来说可采用总满意度最大作为决策,故可将多目标化为如下单目标:

$$\max \quad V(X)=\sum_{i}^{m}\lambda_i=\sum_{i=1}^{m}\frac{f_i(X)-f_i^L(X)}{f_i^H(X)-f_i^L(X)}$$

同时增加如下约束:

$$f_i^L(X)\leqslant f_i(X)\leqslant f_i^H(X)$$

若对一些系统,一些目标达$f_i^H(X)$,另一目标达$f_i^L(X)$会影响系统正常运行,这时各目标需同等满意度,该情况可令$\lambda_1=\lambda_2=\cdots=\lambda_n=\lambda$,上法即变为同等满意度法。

$$\because \lambda_i = \frac{f_i(X) - f_i^L(X)}{f_i^H(X) - f_i^L(X)}$$

$$\therefore f_i(X) - f_i^L(X) = \lambda[f_i^H(X) - f_i^L(X)]$$

则多目标问题化为

$$\begin{aligned} \max \quad & \lambda \\ \text{s.t.} \quad & f_i(X) - f_i^L(X) = \lambda[f_i^H(X) - f_i^L(X)] \\ & \lambda \geqslant 0 \end{aligned}$$

把多目标化为单目标的方法多用于方案的评价和选择，然而在企业的经营管理中也有很大的实用价值。在衡量一个企业经营管理水平的时候，往往由于目标多而无法互相比较，因此在评价过程 经常把多目标化为单目标，以更客观的反映各企业经营效果的差别，使得评价工作有所遵循。因此应深入理解这种方法的本质，以更有效地进行企业间或企业内的经营管理、进行系统的评价。

10.3.2.2 目标分层法

在多目标问题中，每个目标的重要性是各不相同的，在处理多目标问题的时候，首先要分清各目标的重要性。目标重要性的划分随问题的不同而不同，如有的企业以产量为主要目标，有的企业以成本为主要目标等。有时候，目标重要性的划分要由一定的历史时期的一定任务而定。但不论怎样，各种目标总可根据其重要性的不同，而划分成不同的层次，因此，根据目标可划分层次的特点，得到一种解决多目标问题的方法，这种方法叫目标分层法。

目标分层法的主要思想是：把所有的目标按其重要性的顺序排列起来，然后求出第一重要目标的最优解集合 R_1，在此 R_1 集合中再求第二位重要目标的最优解集合 R_2，依次作下去，直到把全部目标求完为止，则满足最后一个目标的最优解，就是该多目标问题的解。

这种思想的数学语言表述如下：

设已按重要性排好顺序的目标 $f_1(X)$，$f_2(X)$，$\cdots$，$f_n(X)$

可按

$$\begin{aligned} & f_1(X^{(1)}) = \min f_1(X) && X \in R_0 \\ & f_2(X^{(2)}) = \min f_2(x) && X \in R_1 \\ & R_1 = \{X \mid f_1(X^{(1)}) = \min f_1(X)\} && x \in R_0 \\ & \cdots \\ & f_n(X^{(n)}) = \min f_n(X) && X \in R_{n-1} \\ & R_i = \{X \mid f_i(X^{(i)}) = \min f_i(X)\} && X \in R_{i-1} \quad (i = 1,2,\cdots,n) \end{aligned} \tag{10.3.4}$$

求满足 $f_n(X^{(n)}) = \min f_n(X)$ 的解 $X^{(n)}$ 则为多目标问题 $f_i(X)(i = 1,2,\cdots,n)$ 的解。

这种方法的几何解释如图 10.3.2 所示，即第一位重要目标在 R_0 范围内求解后得到 R_1 集；而第二位目标在 R_1 集合中求解后，得到 R_2 集；依此类推，最后收缩到中间的最优点，可见它是对所有目标都基本可以满足的解，即是该多目标问题的解。

采用这种方法时，如果出现前面目标的解集 R_i 是一个点集或空集，后面的目标就无法在其中求解，因此，这时不能应用此法。为了适应这种情况，在数学上采用了“宽容”的方法，所谓宽容的方法就是将 R_i 在适当的范围内加以“宽容”，“宽容”就是把 R_i 的范围

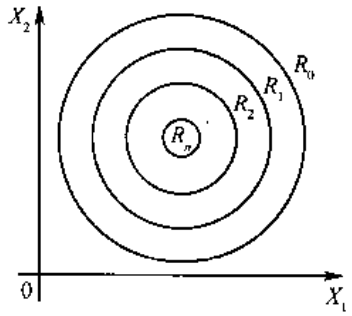


图 10.3.2　目标分层法的几何解释

适当扩大，使 R_i 集由点集或空集而变成一个小的“大范围”。范围的大小，主要依赖宽容量 δ 的大小。

10.3.2.3　功效系数法

每个目标都具有自己所特有的特征。如有的目标要求越大越好，如劳动生产率指标；有的要求越小越好，如成本指标；也有要求适中为佳的，如可靠性指标。如果把目标的特性用其特性曲线来表示，并引入功效系数的概念，可方便地将多目标化为单目标。功效系数一般以 d 表示，它是表示目标满足程度的参数。当 $d=1$ 时，表示对目标最满意，而当 $d=0$ 表示对目标最不满意。且 $0\leqslant d\leqslant 1$，按上述方法可得出不同目标函数和功效系数之间的变化关系，如图 10.3.3 所示。

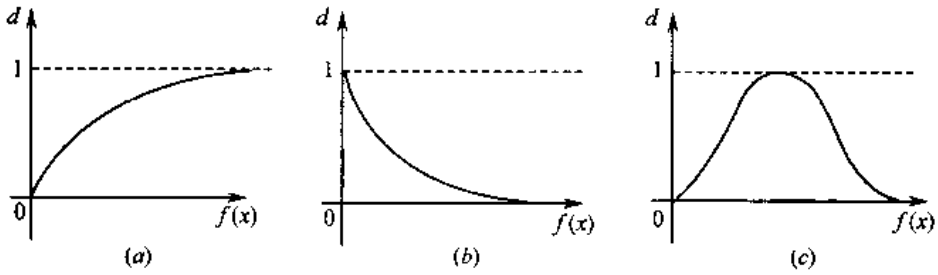


图 10.3.3　目标函数和功效系数的关系

(a) 目标越大越好；(b) 目标越小越好；(c) 目标适中为好。

当目标函数的这种特性曲线已知以后，对于任一多目标问题，当给定一组 x，即可以得到一组相应的 d，则根据各目标的 d 值构成一评价函数(几何平均数)：

$$\max \quad D=\sqrt[p]{d_1\cdot d_2\cdots d_p}=D(x) \tag{10.3.5}$$

其中 p 是目标函数种类数(或个数)。

当 $D=1$ 时为所有的目标函数都处在最满意的情况；当 $D=0$ 时则相反。因此，由不同的 X 值即可确定不同的 D 值，也得到不同的满意程度，进而反映出目标的不同功效来。

作为一个综合的目标 D，我们总是要求它越大越好，因此逐步调正变量，则可使 D 达到最大值，而达到多目标决策的目的。

10.4　目标规划法

以上所述的各种化多目标为单目标的方法均有其适用的条件。如受量纲、数量级等

限制，其中一些方法还会生成非线性目标，使求解变得复杂。因此，人们探索出一种克服上述缺点的化多目标为单目标方法，这就是目标规划法。目标规划法也是一种化多目标为单目标的方法，但与其他化多目标为单目标的方法不同，它不受量纲、数量级等条件的限制，也不会使原线性多目标问题变为非线性问题，使求解简化，同时，该方法可更接近管理者决策的实际，因此，得到广泛地应用。

10.4.1 目标规划模型

管理者在决策之前往往先制定一些原则，然后按原则进行决策。仔细分析这些原则会发现，各项原则之间可能存在矛盾，可能与资源相冲突，要实现最优化是不可能的。但问题总是要解决的，既然达不到最优，就应寻找最大限度地"满足"原则要求的方案作为决策，使管理者达到"最大的"满意。

目标规划法是求一组变量的值，在一组资源约束和目标约束下，实现管理目标与实际目标之间的偏差最小的一种方法。应用目标规划法解决多目标决策问题时首先要建立目标规划模型。目标规划模型由变量、约束和目标函数组成，其模型结构如下：

1. 变量

目标规划模型中的变量除决策变量外还有偏差变量：d^+，d^-。正偏差变量 d^+ 表示决策值超过目标值的部分，负偏差变量 d^- 表示决策值未达到目标的部分，因决策值不可能既超过目标值，同时又达不到目标值，因此恒有 $d^+ \cdot d^- \geqslant 0$ 和 $d^+ \cdot d^- = 0$。

2. 约束

目标规划模型中约束分为目标约束、绝对约束和非负约束三种，绝对约束又称资源约束，是指问题必须严格满足的条件；目标约束是将多个管理目标引入偏差变量后当作约束条件处理的约束。约束的一般形式为

$$\sum_j C_{ij}X_j + d_i^- - d_i^+ = g_i \qquad (i = 1,2,\cdots)$$

式中 g_i——第 i 个目标约束的目标值；

C_{ij}——目标约束中决策变量的参数；

d^+、d^-——以目标值 g_i 为标准而设置的偏差变量。

建立约束需注意的问题是：

(1) 对于绝对约束，g_i 则为资源限制值，上式中不加 d^+、d^-。

(2) 非负约束是指偏差变量非负，d^+、$d^- \geqslant 0$，至于决策变量是否要求非负，依具体问题要求决定。

(3) 在目标规划约束中，凡已列入目标约束的资源约束，不应再列入资源约束。

(4) 如果有明显的目标要求，可在 d^+ 和 d^- 中只选一个。

3. 目标函数

目标规划中的目标函数是求管理目标与实际实现目标的偏差最小。由于目标规划是解决多目标决策问题的，因此不能简单地对各管理目标偏差求和，必须按管理目标的重要性确定优先权因子 P_n 和权数 W_K。

优先权因子是对不同层次的管理目标设置的，对有 n 个管理目标的系统，有：

$$P_1 \gg P_2 \gg \cdots \gg P_n$$

其中“≫”符号表示远远大于的意思，是按目标分层法的思想，在第一目标实现的基础上再考虑第二目标，依次类推。

权数 w_K 是为同一层次中的管理子目标设置的。在同一层次管理目标中可能有多个管理子目标，为区分这些管理子目标的重要程度，以 w_K 作为权数（不一定要求 $\sum_K w_K = 1$）。

目标规划的目标函数是通过在目标约束中引入的偏差变量描述的，但应选取正负偏差中的哪一个进入目标函数，需考虑管理目标的要求，一般遵循下列原则：

(1) 要求目标准确实现。此时即是要求负偏差、正偏差均最小，因此应将二者均列入目标函数

$$Z_{\min} = d_i^+ + d_i^-$$

(2) 要求目标只能超过。负偏差越小越能体现这一管理目标要求，因此应将负偏差列入目标函数

$$Z_{\min} = d_i^-$$

(3) 要求目标不能突破。正偏差越小越可体现这一管理要求，因此应将正偏差列入目标函数

$$Z_{\min} = d_i^+$$

综上所述，目标规划目标函数是由优先权因子、权数和偏差变量组成，其一般形式为

$$Z_{\min} = \sum_K P_K(w_1 d_i^+ + w_2 d_i^-)$$

10.4.2 目标规划应用例

某构件公司商品混凝土车间生产能力为 20t/h，每天工作 18h，现有 2 个施工现场分别需要商品混凝土 A 为 150t，商品混凝土 B 为 100t，两种混凝土的构成、单位利润及企业所拥有的原料见表 10.4.1，现管理部门提出：

表 10.4.1 原料消耗、拥有量 R 单位利润表

	A	B	拥有资源量
水泥/t	0.35	0.25	50t
砂/t	0.55	0.65	130t
单位利润/元	100	80	

(1) 充分利用生产能力；

(2) 加班不超过 2h；

(3) 产量尽量满足两工地需求；

(4) 力争实现利润 2 万元/天。

试建立目标规划模型拟定一个满意的生产计划。

[解]

1. 确定变量

设 X_1、X_2 分别为两种商品混凝土的产量。

2. 约束条件

(1) 目标约束：

P_1 级：要求生产能力充分利用，即要求剩余工时越小越好。

$$x_1+x_2+d_1^- - d_1^+ = 160(\text{t}) \quad \text{其中要求 } d_1^- \to 0$$

P_2 级：要求可以加班，但每日不超过 2h，日产量不能超过 200t。

$$x_1+x_2+d_2^- - d_2^+ = 200(\text{t}) \quad \text{其中要求 } d_2^+ \to 0$$

P_3 级：两个工地需求尽量满足，但不能超过需求。

$$x_1+d_3^- = 150(\text{t}) \quad \text{其中要求：} d_3^- \to 0$$

$$x_2+d_4^- = 100(\text{t}) \quad d_4^- \to 0$$

因需求量不能超过其需要，故 $d_3^+, d_4^+ = 0$

P_4 级：目标利润超过 2 万元。

$$100x_1+80x_2+d_5^- - d_5^+ = 20000(\text{元}), \text{其中要求 } d_5^- \to 0$$

(2) 资源约束：

① 水泥需求不超过现有资源

$$0.35x_1+0.25x_2 \leqslant 50$$

② 砂需求不超过现有资源

$$0.55x_1+0.6x_2 \leqslant 130$$

(3) 非负约束：

$$x_1 \geqslant 0,\ x_2 \geqslant 0, d_i^-, d_i^+ \geqslant 0 \quad (i=1,2,\cdots,5)$$

3. 目标函数

依目标约束中的要求，第三层目标中有 2 个子目标，其权数可依其利润多少的比例确定，即 100:80，简化为 5:4，故 $W_1=5, W_2=4$。故目标函数为

$$Z_{\min} = P_1d_1^- + P_2d_2^+ + P_3(5d_3^- + 4d_4^-) + P_4d_5^-$$

整理得该问题的目标规划模型为

$$Z_{\min} = P_1d_1^- + P_2d_2^+ + P_3(5d_3^- + 4d_4^-) + P_4d_5^-$$

约束：

$$\begin{gathered}
x_1 + x_2 + d_1^- - d_1^+ = 160 \\
x_1 + x_2 + d_2^- - d_2^+ = 200 \\
x_1 + d_3^- = 150 \\
x_2 + d_4^- = 100 \\
100x_1 + 80x_2 + d_5^- - d_5^+ = 20000 \\
0.35x_1 + 0.25x_2 \leqslant 50 \\
0.55x_1 + 0.6x_2 \leqslant 130 \\
x_1 \geqslant 0, x_2 \geqslant 0, d_i^- d_i^- \geqslant 0 \quad (i = 1,2,\cdots,5)
\end{gathered}$$

10.4.3 目标规划求解

10.4.3.1 目标规划的图解法

用图解法求解目标规划只能求解有 2 个变量的目标规划问题，没有多大实用价值，但

它可揭示出求解目标规划的基本思想,对理解目标规划有重要意义。现说明图解法求解目标规划的过程。

10.4.2 例中如果不考虑资源约束,则目标规划模型为

$$Z_{\min} = P_1 d_1^- + P_2 d_2^+ + P_3(5d_3^- + 4d_4^-) + P_4 d_5^-$$

$$\text{s.t.} \quad x_1 + x_2 + d_1^- - d_1^+ = 160 \tag{10.4.1}$$

$$x_1 + x_2 + d_2^- - d_2^+ = 200 \tag{10.4.2}$$

$$x_1 + d_3^- = 150 \tag{10.4.3}$$

$$x_2 + d_4^- = 100 \tag{10.4.4}$$

$$100x_1 + 80x_2 + d_5^- - d_5^+ = 20000 \tag{10.4.5}$$

$$x_1 \geqslant 0,\ x_2 \geqslant 0,\ d_i^- d_i^- \geqslant 0$$

该模型用图解法求解的步骤为:

(1) 不考虑偏差变量,将式(10.4.1)~式(10.4.5)目标约束画在直角坐标系的第一象限;

(2) 在 P_1 目标和 P_2 目标中标出偏差方向,d_1^-、d_1^+ 和 d_2^-、d_2^+,式(10.4.1)和式(10.4.2)目标实现,x_1、x_2 的取值范围在四边形 $ABCD$ 之中(见图 10.4.1);

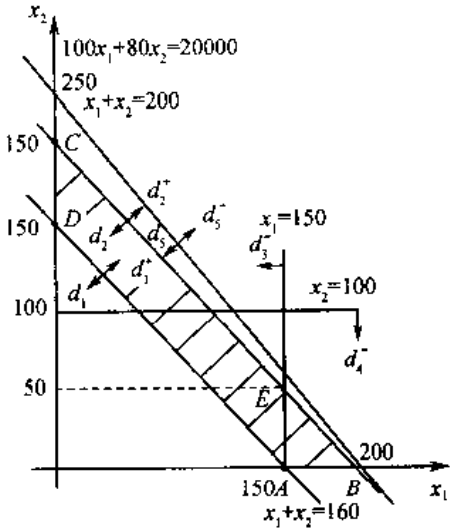


图 10.4.1 目标规划的图解法

(3) 第三目标有 2 个子目标,先考虑权数较大的子目标,在其上标出偏差方向 d_3^-,因 d_3^- 进入目标,意味着 $d_3^- = 0$ 是最好的,故 x_1、x_2 的取值在直线 EA 中,在 EA 上再考虑第 2 个子目标,显然 $d_4^- = 0$ 时的 E 点较好;

(4) 在最后一个管理目标标注偏差变量 d_5^-、d_5^+,因此时解空间只是一个点 E,则选 E 点,其坐标为(150,50)。

此时,由式(10.4.1)和式(10.4.2)可知:$d_1^+ = 40$;$d_2^+ = d_2^- = 0$;需加班 2h,加班产量 40t。

由式(10.4.3)和式(10.4.4)可知,X_1 产量为 150t,X_2 产量为 50t,d_4^- 为 50t,距要求

差 50t。

由式(10.4.5)可知,利润为 19000 元,$d_5^- = 1000$,距要求差 1000 元。

如果考虑资源约束,将坐标资源约束画在图上,然后再画目标约束,解法的思想是相同的。

值得指出的是,管理目标的优先权因子和权数值是很重要的。如果加权因子或权数值改变;会改变管理目标顺序,使求解结果发生变化,这是因为目标规划法是按目标分层原理设计的,其求解是先求出第一目标的解空间,然后再在该解空间中求第二目标的解空间,依此类推最后得到问题的解。因此,采用目标规划法在建立模型时,必须严格按管理目标层次进行。

10.4.3.2 用单纯形法求解目标规划

目标规划法模型结构虽然有加权因子和权数值,但与线性规划并无本质区别,因此可用单纯形法求解。在用单纯形法求解目标规划时,可将优先权因子和权数值当作偏差变量的系数处理,只是检验数一行中含有不同层次的加权因子,即

$$C_j - Z_j = \sum a_{kj} P_K \qquad j = 1,2,\cdots,M;K = 1,2,\cdots k$$

将该检验数行按优先权因子排成 K 行,依优先权因子的系数按判断准则逐层次判断,如果 P_1 层系数全为正,则判断 P_2 层,依次类推,直至全部判断数为正。其步骤如下:

(1) 建立初始单纯形表,表中检验数行按优先权因子排成 K 行,令 $K=1$。

(2) 检查该行是否存在负数,且对应的 $K-1$ 行的系数为零,若有,则选其中最小者所对应的变量为进基变量,转(3),若无转(5)。

(3) 按 $\theta_i = \min\left\{\dfrac{b_i}{a_{ij}} \middle| a_{ij} \geqslant 0\right\}$ 选出基变量,如果存在多个相同的 θ_i,则选优先权层次较高的变量为出基变量,确定旋转中心。

(4) 按单纯形法进行旋转变换。

(5) 当 $n = K$ 时,计算结束,表中的解为满意解,否则 $K = K+1$,转(2)。

在求解过程中可能出现当 K 层判断数全为正,计算到 $K+1$ 时,$K+1$ 层判断数全为正,而 K 层又出现负判断数的情况,这说明第 K 层目标与第 $K+1$ 层目标有冲突,当然应首先满足第 K 层。这就是(2)中选择进基变量时,规定检查判断数为负的同时对应 $K-1$ 行系数为零的道理。这也是求解目标规划的单纯形法与只根据判断数为负确定进基变量的线性规划单纯形法的不同点。

10.5 对策分析

对策分析是与决策分析截然不同的两类问题,又叫对策论或博弈论。

决策是决策人对不同自然环境和情况而做出的策略选择。而对策则不然,它的研究对象是具有竞争活动能力的双方,而且这种活动的成败纯粹依赖于竞争双方发挥自己的主观能动作用和所采取的策略。对策研究的范围很广泛,从简单的掷骰子、下棋到两军对垒、企业竞争都是对策论研究的范畴。

对策论把参与竞争的各方叫局中人,它们都有各自的决策权,局中人至少是两个,只有两个局中人的对策叫二人对策,如下棋等,如果有多个局中人则叫多人对策,如打桥牌

等。

每个局中人所采取的策略在一局对策中是一套完整的方案，而且这种策略是可供局中人选择的多个方案中的一个。在一局对策中，可供选用的方案既可是有限的，也可以是无限的，这将由对策问题的性质来确定。如果在一局对策中有胜负之分，局中人一得一失，得失相等则叫零和。

本节主要研究这种“二人零和对策”。

我国古代齐王和田忌赛马的故事，就是对策问题的最好例子。战国时期齐王和田忌赛马，各从自己的上等马、中等马、下等马中选出一匹进行比赛，每输一局，输银千两，齐王的马都比田忌的马好，但田忌的谋士出了一个主意，让田忌的下等马与齐王的上等马赛，而用上等马和中等马分别与齐王的中等马和下等马赛，这样田忌非但没输，反而赢了一千两。由这个故事可知，每个局中人所采取的策略是非常重要的。

如果局中人 A 有 m 种策略，$i=1,2,\cdots,m$；局中人 B 有 n 种策略，$j=1,2,\cdots,n$。当 A 采用第 i 种策略时，B 采用第 j 种策略时 B 支付给 A C_{ij}，可得到反映得失情况的矩阵，叫支付矩阵，见表 10.5.1。

表 10.5.1　支付矩阵

支付 B / A	1	2	…	j	…	n
1	C_{11}	C_{12}	…	C_{1j}	…	C_{1n}
2	C_{21}	C_{22}	…	C_{2j}	…	C_{2n}
…	…	…	…	…	…	…
i	C_{i1}	C_{i2}	…	C_{ij}	…	C_{in}
…	…	…	…	…	…	…
m	C_{m1}	C_{m2}	…	C_{mj}	…	C_{mn}

齐王和田忌赛马故事的支付矩阵见表 10.5.2。

表 10.5.2　齐王和田忌赛马的支付矩阵

齐王策略 \ 田忌策略	上中下	上下中	中上下	中下上	下中上	下上中
上中下	3	1	1	1	1	−1
上下中	1	3	1	1	−1	1
中上下	1	−1	3	1	1	1
中下上	−1	1	1	3	1	1
下中上	1	1	−1	1	3	1
下上中	1	1	1	−1	1	3

支付矩阵的元素 C_{ij} 表示竞争双方的支付量，当 $C_{ij}>0$ 时，表示局中人 B 支付给局中人 A，当 $C_{ij}=0$ 时，表示局中人 B 支付给局中人 A 的量为零；当 $C_{ij}<0$ 时，表示局中人 A 支付给局中人 B 或局中人 B 支付给局中人 A 一负量。

对策研究的目的是为了研究对策问题的内在运动规律，从而为局中人确定一种最优的策略。

最优策略中有一种特例叫最优纯策略，我们先从这个问题谈起。

如有一支付矩阵见表 10.5.3。局中人 A 有四种策略，B 有三种策略，在双方互不保密的情况下，当 A 采取第一种策略时，B 为了得到最大收益必然采取第三种策略，同理当 A 采取第二种策略，则 B 采取第二种，当 A 采取第三种策略，B 采取第三种；当 A 采取第四种策略，B 采取第一种。这样得到一列由于 A 采取不同的策略，B 所支付的数(-8，2-9，-3)，从这一列数中可见，A 只有选第二种策略，才是最合算的。

表 10.5.3　支付矩阵

B / A	1	2	3	最小支付数
1	-7	1	-8	-8
2	6	2	4	2
3	16	-1	-9	-9
4	-3	0	5	-3
最大收入数	16	2	5	

同理，当 B 采取不同策略时，A 为了得到最大收益，必采用相应的策略来对付，如 B 选第一策略，A 必选第三策略，B 选第二种策略，则 A 必选第二种，B 选第三种，A 选第四种；这就得出一行由于 B 取不同策略，A 所得到的收益数，B 为了使 A 得到最少，必选用第二策略。

用数学式表示，则

$$\begin{aligned}&\text{对 } A:\min_j C_{ij} \quad \text{然后}\max_i\{\min_j C_{ij}\}=V\\&\text{对 } B:\max_i C_{ij} \quad \text{然后}\min_j\{\max_i C_{ij}\}=V'\cdot\\&\text{当}\max_i\{\min_j C_{ij}\}=\min_j\{\max_i C_{ij}\}\end{aligned}\tag{10.5.1}$$

即 $V=V'$ 的策略叫最优纯策略。这种策略表明，对策双方即使在不保密的情况下，谁也不会得到更多的便宜，这就是最优纯策略的特点。

然而并不是所有的对策问题都有最优纯策略，上面举的齐王和田忌赛马的例子就不存在最优纯策略。齐王和田忌都为了赢得对方的银子，就都要在保密的情况下采取自己相应的策略，冒险地竞争。

对于不存在最优纯策略的对策，引入混合策略的概念，即每个局中人的决策，不是决定哪一个纯策略，而是决策用多大的概率去取每个纯策略。如果田忌取各种策略的概率各为 1/6，则

$S_{田}=\left\{\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right\}$ 是田忌的一种混合策略。

一般情况下，A 有 m 个策略，B 有 n 个策略，则

$$S_A=\{X=(X_1,X_2,\cdots,X_m);\quad X_i\geqslant 0;\sum_{i=1}^{m}X_i=1\}$$

$$S_B=\{Y=(Y_1,Y_2,\cdots,Y_n);Y_j\geqslant 0;\sum_{j=1}^{n}Y_j=1\}$$

分别是 A 和 B 的混合策略。其中 X_i 是 A 取某种策略的概率，Y_j 是 B 取某种策略的概率。

这时 B 向 A 的支付值就要以期望值来表示：

$$E(X_i,Y_j)=\sum_{i-1}^{m}\sum_{j-1}^{n}X_iC_{ij}Y_j=XCY^{\mathrm{T}} \tag{10.5.2}$$

该式表示了 A 只有 X_i 的可能取 i 策略，B 只有 Y_j 可能取 j 策略，B 向 A 支付 C_{ij} 的可能性只有 X_iY_j。如田忌只有 1/6 的可能取上中下策略，而齐王也有 1/6 的可能取上中下策略，则田忌支付给齐王 $C_{ij}=3$ 的可能性只有 $1/6\times 1/6=1/36$，支付额的期望值为 83.3 两。

这里不加证明地引用一个定理，即极大极小定理。

定理：对于任何以 $[C_{ij}]$ 表示的支付矩阵，都有：

$$\max_{x\in S_A}\min_{y\in S_B}E(x,y)=\min_{y\in S_B}\max_{x\in S_A}E(x,y)=V \tag{10.5.3}$$

这是一个和最优纯策略相似的公式，不同点是它采用了数学期望值。

由此，我们总可以找到一种最优的混合策略：x^*、y^*，使得

$$\min_{y\in S_B}E(x^*,y)=\max_{x\in S_A}E(x,y^*) \tag{10.5.4}$$

即当 A 采用混合策略 x^* 时，B 必然相应采取混合策略 y^*，无论二者如何变化，谁也不会得到大于 V（期望值）的收益。

现介绍图解法求混合策略。

由表 10.5.4 的支付矩阵，显然有：

$$\max_i\min_j C_{ij}=8$$

$$\min_j\max_i C_{ij}=10$$

表 10.5.4　支付矩阵

		Y_1 1	Y_2 2	B 最小支付
X_1	1	11	5	5
X_2	2	8	10	8
A 最大收入数		11	10	

不存在最优纯策略，现在研究 A、B 应各取哪种混合策略。

对于这种仅有两种策略的情况，采用图解法用图 10.5.1 来计算：它的纵坐标是当 A 采取不同混合策略时的收益，横轴表示所取的各种混使策略的概率即 X_i，因为当 $X_1=0$ 时，X_2 必定为 1，同理，当 $X_2=0$ 时，X_1 必定为 1，且 $X_1+X_2=1$。图中 $B(1)$ 表示 B 采取第一种策略，A 采取不同策略时的收益，$B(2)$ 表示当 B 采取第二种策略，A 采取不同混

合策略时的收益，由图可知，当 A 采取 K 以右的混合策略时，B 为了使 A 得到更少的收益，必须取 $B(2)$的策略，当 A 采取 K 点以左的混合策略时，B 为了使 A 收到最少的收益，故必采取 $B(1)$的策略，因此 A 为了得到最大的收益，一定采取 K 点所对应的策略(混合策略)。

K 点的坐标可以通过解析几何求两条直线交点的方法求解或用平面几何的方法求解。

同理，也可以得到 B 所采取的混合策略。见图 10.5.2，即当 B 采取 K 左侧的策略时，A 为了得到更大的收益(相对应于 B 的支付)而采用 $A(2)$的策略，而当 B 采取 K 右侧的策略时，A 必定采用 $A(1)$的策略，因此，B 为了支付最小必须采取 K 时的策略。

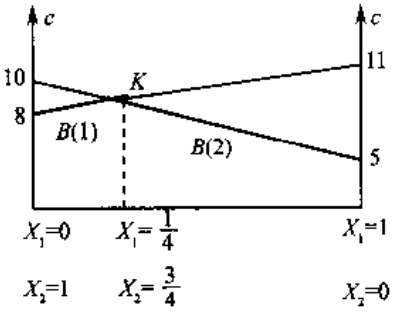


图 10.5.1　图解法

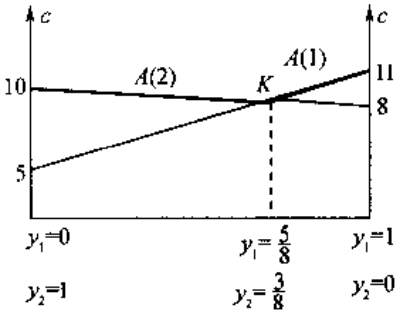


图 10.5.2　B 所采取的混合策略

第 11 章　网络计划技术

当系统决策之后，面临的问题就是制定计划和组织实施。网络计划技术就是在制定计划和组织实施过程中采用的一种科学方法。本章主要对目前管理中常用的网络计划方法做一系统的介绍。

11.1　系统管理的网络技术

网络计划技术是系统管理的重要工具之一，是系统工程常用的管理技术。它是利用网络图对计划任务的进度、费用及其组成部分之间的相互关系进行计划、检查和控制，以使系统协调运转的科学方法。由于这种方法的主要特点是统筹安排，为了突出这一特点，我国把各种不同的网络计划方法统称为统筹法。

管理的对象都可作为一个系统来研究。任一系统就其内容来说，各有特点，但却都具有共性，那就是必须按系统单元内在的时间和空间联系，把物质、能量和信息有机地组织起来，在最短的时间内，以最少的消耗实现系统的目标，取得最大的效益。为了更有效地管理系统运行，目前广泛使用的统筹方法有以下几种。

11.1.1　甘特图法

该方法是科学管理的奠基人泰勒的学生、美国福兰克兵工厂顾问甘特于 20 世纪 40 年代开发的一种计划与管理技术。它以时间为横坐标，以工序为纵坐标，以线条的长短表示一项工作或作业的开始和完成时刻以及工作的进展情况。由于它以条形图进行系统计划与管理，故又称横道图、条形图等，图 11.1.1 即是一份简化的现场准备作业的条形图。甘特图的最大特点是简单明了、容易绘制、使用方便。最大缺点是：

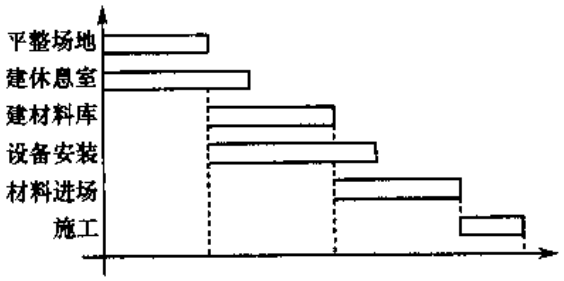


图 11.1.1　甘特图

(1)不能反映各项工作之间错综复杂的联系和制约的分工协作关系；

(2)不能反映系统中哪些工作是主要的、关键性的生产联系和工序，反映不出全局的关键所在，不利于最合理地管理整个系统，可能造成在不重要的工作上投入过分的资源，却忽视了影响大局的关键性工作，虽然忙得团团转，却徒劳无功，致使系统目标难以实现。

由于甘特图法的这些缺点，造成其使用上的局限性。随着科学和技术的发展，系统的规模越来越大，应用甘特图法已无法满足需要，这就需要探讨一种新的计划管理方法。

20世纪50年代以来,各国科学家都在进行这方面的探索,自从关键路线法和计划评审技术出现以后,在计算机广泛应用的基础上使系统的计划与管理进入一个新的阶段。

甘特图法虽有缺点,但目前仍有很大的实用价值,它与关键线路法等配合使用会收到更好的效果。

11.1.2 关键线路法(CPM法)

关键线路法一般称CPM法,CPM是英文Critical Path method的缩写。

该方法是美国杜邦公司为建造新工厂从事计划与管理的研究而提出的,并在1958年的建厂工作中发挥了很大作用,使工程工期提前2个月,初步显示出其优越性。而后,杜邦公司不仅把CPM法应用于大型工程,而且也应用于小型工程和维修工程,都同样收到了良好的效果。美国加泰迪克公司在47项工程中使用CPM法,平均节约时间22%,节约资金15%,其效果是显著的。

该方法以网络图的形式表示各工序之间在时间和空间上的相互关系以及各工序的工期,通过时间参数的计算,确定关键线路和总工期,从而制定出系统计划并指示出系统管理的关键所在。

该方法问世后,立刻引起世界各国的重视,很多国家引入该方法都收到了良好的效果。1961年由华罗庚教授将该方法引入我国并推广到各行各业,派生出一些新的方法,如时间—费用网络等,使其内容更加丰富。

11.1.3 计划评审技术(PERT)

计划评审技术又叫PERT,PERT是英文Program Evaluation and Review Teacnique的缩写,也译作计划协调技术。

PERT是由美国海军特种计划局和洛克希德公司、汉米尔顿公司于1958年1月联合开发的一种新的计划管理方法。它的首次应用使美国"北极星"导弹潜艇工程的工期由原计划的10年缩短为8年。由于它的成功,自1962年起美国政府规定一切新开发的工程项目必须采用这种方法。我国的很多大型开发性工程也都应用了这种方法,收到了良好的效果。

这种方法与CPM法既有联系又有区别。其联系是二者的网络图形和计算方法基本相同,区别是研究对象和研究目的不同。

从研究对象看,PERT主要侧重研究新开发系统。CPM主要应用于已开发过的有一定经验的系统。

从研究目的看,PERT主要用于研究系统各项工作安排情况的评价和审查,而CPM法主要研究完成任务的工期和关键工作。这主要体现在使用CPM法时,各工序的执行时间可根据定额等确定,为一确定值,而对开发性工程因无工作经验,各工序的执行时间受各种因素影响,很难估计出准确值,故带有随机性。

从计算方法看,一般情况下可将PERT网络中各工序的具有随机性的工期化成确定型的工期,从面使PERT网络变为CPM网络,从该点看,如果认为确定型问题是随机问题的特例的话,CPM网络是PERT网络在工期不受随机因素干扰时的特例。

11.1.4 决策关键路法(DCPM)

DCPM 是英文 Decision Critical Path Method 的缩写,是决策关键线路法的意思。

任何一项工作都可用多种方案进行计划和管理,因此需要在多种方案的比较中进行决策。然而 CPM 和 PERT 二种方法虽然先进,但却只能考虑完成任务的一种方案,对不同方案必须画出不同的网络图,不仅增大绘制网络图的工作量,而且,在比较和分析时也不能一目了然。为了克服这种局限性,又开发了 DCPM。

DCPM 的特点在于把 CPM 网络和决策理论结合起来,在同一张网络图上表示完成同一任务的各种不同方案,通过特定的计算方法,可得出完成该项任务所对应的最小费用方案及其关键路线等,为管理者决策提供了方便。

11.1.5 图解评审技术(GERT)

GERT 是英文 Graphical evaluation and Review Technique 的缩写,意思是图解评价和调整技术,因此称为图解评审技术或图解协调技术。

上述各种方法中各工序均属确定型工序,但实践中存在随机工序,如检验工序后的返修工序,这类工序应用上述各种方法很难解决,GERT 是一种广义网络计划法,是解决这类问题的一种方法。

GERT 的特点是在网络图中同时考虑工序和工序执行时间的随机性,因此它可用于各种系统。它除可将 CPM、PERT 网络作为特例之外,尚可用于系统的模拟,因此该方法有广阔的应用前景。

为了加速我国四个现代化的步伐,提高计划管理水平,推广网络计划法有重要的现实意义。我们相信这些科学的计划与管理方法必须在我国的四化建设中发挥巨大作用。

11.2 网络图的编制

网络图又叫统筹图,它是计划任务及其组成部分相互关系的综合反映,是进行计划、管理和计算的基础。

11.2.1 网络图的组成

网络图是针对一项任务编制的。任务是指一项有目的、有开始和结束标志的许多相互关联且有不同指标要求的工作所组成的事物。

任何一项任务都是由工作或工序组成的,现将为完成某项任务在工艺和组织管理上相互独立的活动称之为工作或工序,它包括人的各式各样的相互协调的劳动。

任务和工作(工序)是相对存在的,一项任务由工作组成,但将该工作当作一项任务,它又由很多小的工作组成,因此任务是可逐级分解的。

当把一项任务分解成工作的时候,必须考虑其流程特性。流程是指一项任务随时间的推移而逐渐展开的过程。流程特性就是指任务在展开的过程中工作与工作之间紧密关联的特点。按流程特性,工作都有紧前工作和紧后工作。所谓紧前工作是指一项工作开始之前必须完成的工作;紧后工作是指一项工作完成之后紧接着能进行的工作。紧前工

作和紧后工作是相对的，一项工作是其紧后工作的紧前工作，又是其紧前工作的紧后工作，因此可应用流程特性将一项任务分解成工作，并表示出各工作间的逻辑关系。

一项工作的开始和结束时刻的标志称为事项。事项为两项工作的交接点，又称节点，它是前项工作结束、后项工作开始的标志。任一事项都具有瞬时性，因其体现工作间的衔接过程且与完成的时间相比是短暂的，所以不消耗资源，不占用时间和空间。网络图中的第一个事项称为起始事项，最后事项为结束事项。其余事项为中间事项，中间事项起承上启下的作用。事项还必须具有易检性，即必须具有容易检验和判断的标准。

11.2.2 网络图及编制

各种网络计划方法都需绘制网络图，因此，编制网络图是制定计划的基础和开始。编制网络图是在对任务进行分析和分解的基础上，按规定的画法画出网络图的过程。该过程分为两步。

11.2.2.1 任务的分解和分析

任务分解和分析的工作内容为：把一项任务分解成工作（或工序），分析并确定各个工作在工艺和组织方面的相互联系及相互制约关系；确定工作的先后顺序，找出各工作的所有紧前、紧后和与它平行的工作；确定该工序的工期等。为了明显清晰起见，需列成“工作逻辑关系表”（见表 11.2.1）。该表是编制网络图的基础和关键，是编制计划的开始。因此，必须深入调查研究，与有关技术部门相互配合，密切协作，与生产部门协商，不断修改，才能客观地正确地反映出任务的结构和内在联系，才能得到符合客观实际的“工作逻辑关系表”。

表 11.2.1 工作逻辑关系表

工序名称	紧前工作	紧后工作	工期	调整期	…

11.2.2.2 画网络图

1. 网络图的表示法

在网络计划法中，网络图有单代号和双代号两种表示法。两种表示方法各有利弊，现分别介绍如下：

（1）单代号网络图：它用○表示一个工作，用箭头线表示各工作之间的联系。在○内用横线分开的上下两个数字表示工作名称和工期，如图 11.2.1 所示。它表示了第六号工作的紧前工作是第五号工作，而两工作的工期分别是 2 天和 3 天。

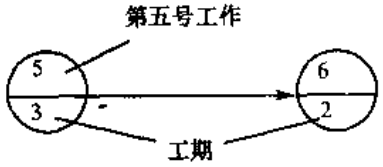


图 11.2.1 单代号网络图表示法图

表 11.2.2 工作与紧后工作

工作	紧后工作
A	D,E,F
B	D,E,F
C	D,E,F

单代号网络图中,有时为了简化描述,可人为地引入一个没有工作内容和工期的虚拟的工作,该工作叫虚工作。根据表 11.2.2 得图 11.2.2(a),引入虚工作后 K 的网络图(见图 11.2.2(b)),由图可见在单代号表示法中虚工作的简化作用是明显的。

(2) 双代号网络图:这种表示法用○表示事项,以箭头线表示工作。在箭头线上面标工作名称,在箭线下面标工期。图 11.2.3 所示的工作中,ⓘ 表示 5 号工作的开始事项,ⓙ 表示 5 号工作的结束事项,由 ⓘ 到 ⓙ 的箭线则表示 5 号工作,箭线上面的 5 表示工作的名称,下面的数字 3 表示该工作的工期。

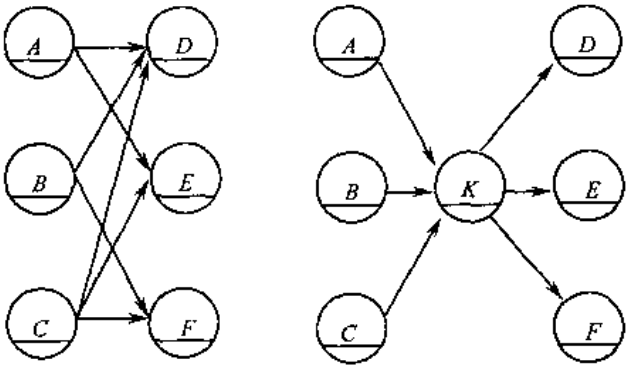


图 11.2.2 单代号网络图虚工作

在双代号网络图中,为了正确地表示各工作之间的逻辑关系,有时尚需要引入虚工作,在双代号网络图中,所谓虚工作是指没有具体工作内容,不消耗资源和时间,只表示各工作之间的逻辑关系的以虚箭头线表示的工作,表 11.2.3 表示的逻辑关系可用引入虚工作的网络图(图 11.2.4(b))表示。

表 11.2.3

工作	紧后工作	工期
A	B	3
B	-	5
C	B,D	8
D	-	6

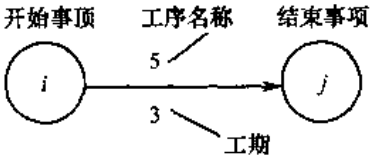


图 11.2.3 双代号网络图

对于同一任务,在绘制网络图时,由于采用的方法不同,所得网络图的形式也不同,如图 11.2.4 所示。

2. 绘制网络图的基本规则(双代号网络图)

(1) 只允许有一个总的开工事项和总的完工事项。如果一项任务出现多个开工事项或完工事项,可引入虚工作予以消除。如某项目准备作业的网络图见图 11.2.5;

(2) 所有箭线方向必须由左到右,时间必须由现在到将来,网络图中不允许出现回路;

(3) 事项号码必须由小到大;

(4) 进入某事项的工作虽然可有多条,但由同一事项进入该事项的工作只能有一条。如遇此情况,必须引入虚工序予以消除,如图 11.2.6 所示;

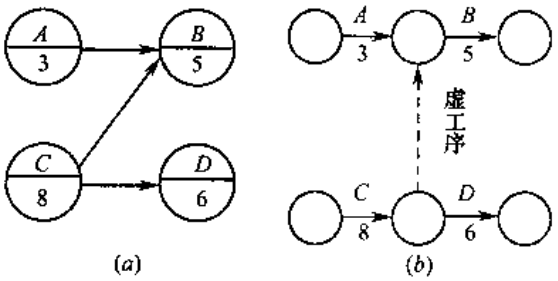


图 11.2.4　单代号与双代号网络图

(a)单代号；(b)双代号。

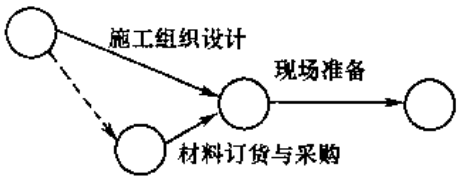


图 11.2.5　某项目准备作业的网络图

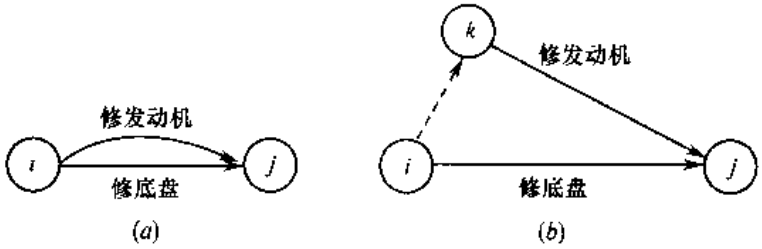


图 11.2.6　由同一事项进入该事项的工作路径

(a)不正确；(b)正确。

(5) 箭线尽量避免交叉。

3. 网络图的几种基本形式(双代号)

网络图中各工作逻辑关系见表 11.2.4。

4. 网络图的简化

对一组有共同开始和结束事项的工作可以进行简化，如图11.2.7 和图 11.2.8 所示。

掌握上述基本原则和基本逻辑关系，就可以绘制出任何一项任务的网络图。

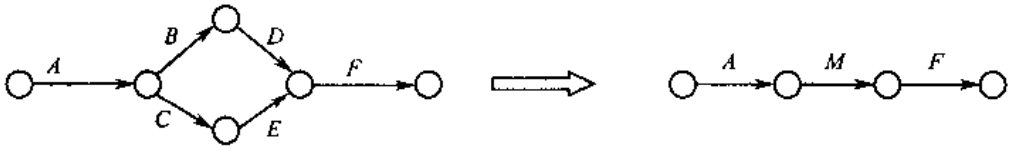


图 11.2.7　网络图的简化(一)

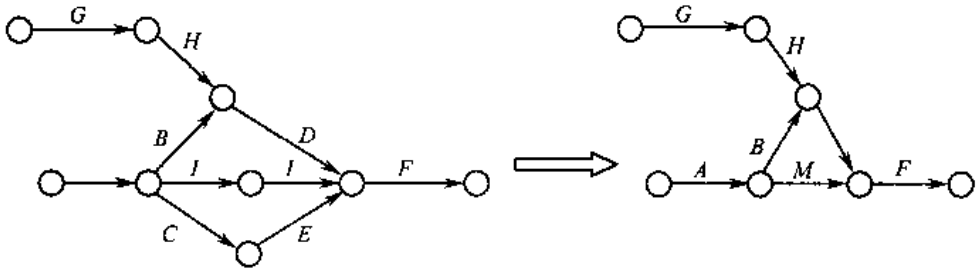
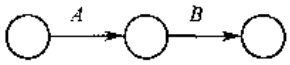
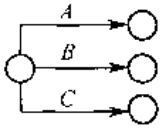
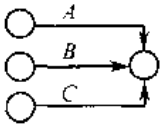
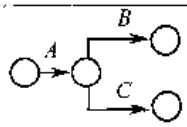
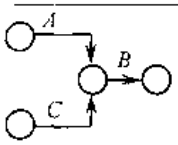
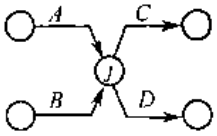
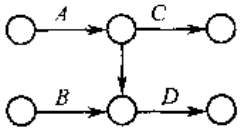
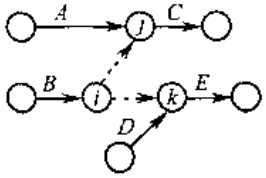
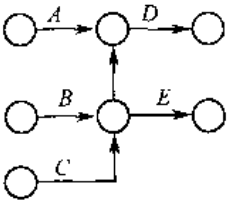
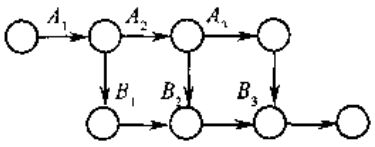


图 11.2.8　网络图的简化(二)

表 11.2.4　双代号网络图中各工作逻辑关系表示方法

序号	工作之间的逻辑关系	网络图中表示方法	说　明
1	有 A、B 两项工作按照依次施工方式进行		B 工作依赖着 A 工作，A 工作约束着 B 工作的开始
2	有 A、B、C 三项工作同时开始工作		A、B、C 三项工作称为平行工作
3	有 A、B、C 三项工作同时结束		A、B、C 三项工作称为平行工作
4	有 A、B、C 三项工作只有在 A 完成后，B、C 才能开始		A 工作制约着 B、C 工作的开始。BC 为平行工作
5	有 A、B、C 三项工作 C 工作只有在 A、B 完成后才能开始		C 工作依赖着 A、B 工作。A、C 为平行工作
6	有 A、B、C、D 四项工作只有当 A、B 完成后 C、D 才能开始		通过中间事件 j 正确地表达了 A、B、C、D 之间的关系
7	有 A、B、C、D 四项工作 A 完成后 C 才能开始 A、B 完成后 D 才开始		D 与 A 之间引入了逻辑连接(虚工作)只有这样才能正确表达它们之间的约束关系
8	有 A、B、C、D、E 五项工作 A、B 完成后 C 开始，B、D 完成后 E 开始		虚工作 ij 反映出 C 工作受到 B 工作的约束；虚工作 jk 反映出 E 工作受到 B 工作的约束
9	有 A、B、C、D、E 五项工作 A、B、C 完成后 D 才能开始，B、C 完成后 E 才能开始		这是前面序号 1、5 情况通过虚工作连接起来，虚工作表示 D 工作受到 B、C 工作制约
10	A、B 两项工作分三个施工段，平行施工		每个工种工程建立专业工作队，不同工种之间用逻辑搭接关系表示

11.2.2.3 网络图的标号

为统一起见，采用删掉箭杆法对网络图的事项进行编号，即把由已编号事项出发的所有箭线去掉，显露出来的事项按由小到大的顺序编号。但编号中需留有余地，以备修改之用，如图 11.2.9 所示。

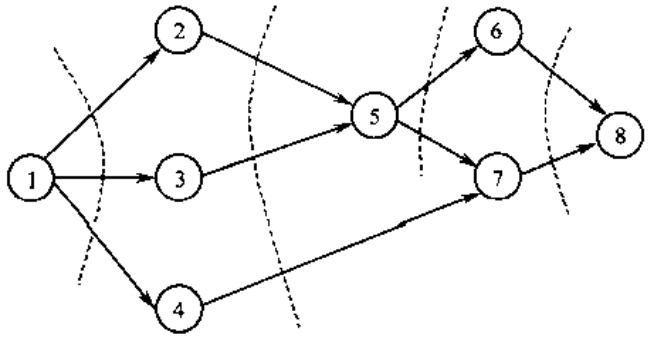


图 11.2.9 双代号网络图的编号

11.3 关键线路法(CPM)

在网络最优化中，已研究了多类网络的算法，而 CPM 网络图的计算类似于求最长路径。但因它是一种带时间参数的网络图，又有自己的规律和特点。现介绍关键线路法的几种常用算法。

11.3.1 分析法

CPM 网络的最基本计算方法是分析法。该方法揭示了 CPM 网络的内在规律，其他算法都是根据这种算法的原理进行的。因此，了解这种算法的原理，则可处理一切 CPM 网络。

11.3.1.1 关键路线、路长和时差

假设某任务的网络图(如图 11.3.1 所示)由事项⓪开始到事项⑤为止，中间连续不断的路有 5 条，把每条路所含工作的工期之和称路长，则得到表 11.3.1。

由表 11.3.1 可见，⓪→①→②→④→⑤决定了整个任务的工期。这条路上的各工作的工期如果延长，总工期必定延长；该路上某工作工期缩短，则总工期必提前，因此我们把这种决定任务总工期的路叫关键线路，把关键线路的路长和其余线路的路长之差叫线路时差。线路时差是松动时间，是与关键线路相比能够提供多少机动时间而不影响总工期的时间。

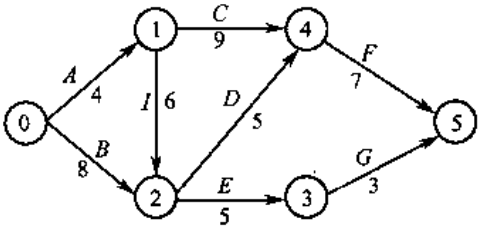


图 11.3.1 某任务的网络图

表 11.3.1　路长和时差

路	路长	线路时差
⓪→①→④→⑤	4+9+7=20	2
⓪→①→②→④→⑤	4+6+5+7=22	0
⓪→①→②→③→⑤	4+6+4+3=17	5
⓪→②→③→⑤	8+4+3=15	7
⓪→②→④→⑤	8+5+7=20	2

线路的时差以下式计算：

$$F(L_n)=L_{cp}-L_n \tag{11.3.1}$$

式中　$F(L_n)$——线路时差；

L_{cp}——关键路线路长；

L_n——第 n 条线路的路长。

11.3.1.2　网络时间参数的确定

1. 事项(i)的最早开工时刻 $t_{ES}(i)$

只有进入 i 事项的全部工作完成之后，由 i 事项开始的工作才能进行，这个时刻叫 i 事项的最早开始时刻。该时刻受紧前工作制约，并按下式计算(见图 11.3.2)：

$$t_{ES}(i)=\max\{t_{ES}(k)+T(k,i)\} \tag{11.3.2}$$

其中　$t_{ES}(i)$——第 i 事项最早开始时刻；

$t_{ES}(k)$——第 k 事项最早开始时刻。

$T(k,i)$——(k,I)工作的工期。

$k=0$ 时，$t_{ES}(k)=0$；

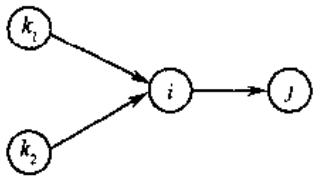


图 11.3.2　网络时间参数的确定

现以图 11.3.1 为例进行计算：

事项①的最早开始时刻 $t_{ES}(1)=\max\{0+4\}=4$

事项②的最早开始时刻 $t_{ES}(2)=\max\{0+8,4+6\}=10$

依此类推可以计算出所有事项的最早开始时刻。

2. 工作的最早开始时刻和最早完工时刻

由事项的衔接性知，一项工作的开始时刻与该工作开始事项的最早开始时刻是相同的，因此有

$$t_{ES}(i,j)=t_{ES}(i) \tag{11.3.3}$$

也可以写成　$t_{ES}(i,j)=\max\{t_{ES}(k,i)+T(k,i)\}$　(11.3.4)

工作的最早完工时刻 t_{EF} 为

$$t_{EF}(i,j)=t_{ES}(i,j)+T(i,j) \tag{11.3.5}$$

式中　$t_{EF}(i,j)$——(i,j)工作的最早完工时刻。

3. 事项(j)的最晚完工时刻

当某任务在预定的最早时刻完成时，所有进入该事项的工作最晚必须完成的时刻，也就是由该事项出发的所有工作在保证任务于某一预定的最早时刻完成时，必须能开始进行的时刻叫事项的最晚完工时刻，该时刻以 $t_{LF}(j)$ 表示，其计算公式如下(见图 11.3.3)：

$$t_{LF}(j)=\min\{t_{LF}(k)-T(j,k)\} \tag{11.3.6}$$

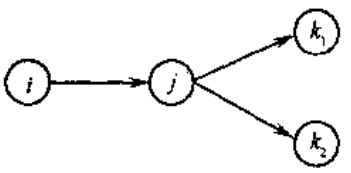


图 11.3.3　最晚完工时刻

以图 11.3.1 为例的计算如下：

⑤ 事项的最晚完工时刻为 22(与⑤事项的最早开始时刻相同)；

④ 事项的最晚完工时刻为 $t_{LF}(4)=\min\{22-7\}=15$；

③ 事项的最晚完工时刻为 $t_{LF}(3)=\min\{22-3\}=19$；

② 事项的最迟完工时刻为 $t_{LF}(2)=\min\{15-5,19-4\}=10$；

① 事项的最迟完工时刻为 $t_{LF}(1)=\min\{15-9,10-6\}=4$；

⓪事项的最迟完工时刻为 $t_{LF}(0)=\min\{4-4,10-3\}=0$。

可见它受紧后工序的影响，是由后向前推算的。

4. 工作的最晚完工时刻和最晚开工时刻

工作的最晚完工时刻等于工作结束事项的最晚完工时刻，又等于所有紧后事项的最晚完工时刻和工期之差中最小的一个，同时，也是所有紧后工作的最迟完成时刻与工期之差的最小一个。由图 11.3.4 可见，其计算公式为

$$t_{LF}(i,j)=t_{LF}(j)=\min\{t_{LF}(k)-T(j,k)\}=\min\{t_{LF}(j,k)-T(j,k)\} \tag{11.3.7}$$

工作最晚必须开工时刻为

$$t_{LS}(i,j)=t_{LF}(i,j)-T(i,j) \tag{11.3.8}$$

5. 工作总时差和关键线路

一项工作的最晚结束时刻与最早结束时刻之差或最晚开始时刻与最早开始时刻之差为工作的总时差。它表示该项工作的宽裕时间，用公式表示为

$$R(i,j)=t_{LS}(i,j)-t_{ES}(i,j)=t_{LF}(i,j)-t_{EF}(i,j) \tag{11.3.9}$$

$$R(i,j)\geqslant 0$$

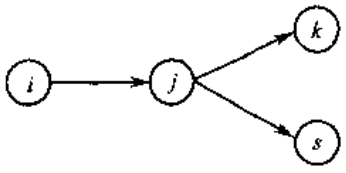
仍用上例计算，得

$$R(0,1)=0,\ R(0,2)=2,\ R(1,2)=0,\ R(2,3)=5,\cdots$$

只有总时差为零的工作，才是关键线路上的工作。本例中关键线路上的工作为 0—1—2—4—5。

6. 工作的单时差

工作的总时差表示一项工作的宽裕时间，然而在某些情况下，这种时差是不能单独使用的。如图 11.3.1 的例中，③—⑤工作的总时差 $R(3,5)=t_{LS}((3,5)=t_{ES}(3,5)=19-14=5$，②—③工作的总时差 $R(2,3)=t_{LS}(2,3)-t_{ES}(2,3)=15-10=5$ 或 $R(2,3)=t_{LF}(2,3)-t_{EF}(2,3)=19-14=5$，两工作的总时差均由 19－14 决定的。如果②—③工作使用了这段宽裕时间，则③—⑤工作就无宽裕时间可用，故引入一个表示可仅供本工作使用的宽裕时间——单时差。单时差是在紧后工作最早开始的情况下，本工作可单独使用的时差，它是不能存储的，是紧后工作最早开始时刻与本工作最早结束时刻之差(图 11.3.5)。



11.3.4 最晚完工和最晚开工时刻

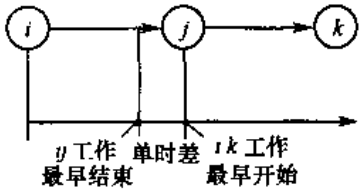


图 11.3.5 单时差

其计算公式为

$$r(i,j)=t_{ES}(j,k)-t_{EF}(i,j) \tag{11.3.10}$$

由式(11.3.10)可知 $r(i,j)\geqslant 0$。

单时差和总时差的区别在于前者只能本工作使用,后者有可能存储供紧后工作使用。

7. 干扰时差

既然在某些情况下总时差是不能单独使用的宽裕时间,而单时差是各项工作可单独使用的宽裕时间,那么二者之间的关系是:

$$R(i,j)\geqslant r(i,j)$$

现定义

$$I(i,j)=R(i,j)-r(i,j) \tag{11.3.11}$$

为干扰时差,它体现出本工作可以不使用而留给紧后工作使用的可存储的宽裕时间。

8. 单时差和线路时差的关系

由单时差的定义知,单时差和线路时差的关系为

$$F(L_n)=\sum_{k=1}^{m} r_k(i,j) \tag{11.3.12}$$

式中 k——某条线路上的工作;

m——某条线路上的工作数。

因此,无论是使用总时差还是使用单时差,在网络图的一条线路上总不能超过线路时差。

当线路时差为零时,该线路为关键线路,则

$$F(L_n)=\sum_{k=1}^{m} r_k(i,j)=0$$

$\because\quad r(i,j)\geqslant 0$

$\therefore\quad r_k(i,j)=0\quad (k=1,2,\cdots,m)$

$\because\quad F(L_n)=0$

$\therefore\quad I_k(i,j)=0\quad (k=1,2,\cdots,m)$

由式(11.3.11)知 $R_k(i,j)=0(k=1,2,\cdots,m)$。

由上述推导可见,网络图中关键线路的各工作的单时差和总时差均为零,但总时差为零的工作必定是关键线路,而单时差为零的工作则不然。因此,$R_K(i,j)=0$ $(k=1,2,\cdots,m)$是该线路为关键线路的充要条件。

上述即为确定关键线路的分析法,其计算步骤归纳如下:

(1) 求事项的最早开始时刻 $t_{ES}(i)$;

(2) 求工作的最早结束时刻 $t_{EF}(i,j)$;

(3) 令终点事项的最早开始时刻为总工期;

(4) 求事项的最晚完工时刻 $t_{LF}(i)$;

(5) 求工作的最晚开始时刻 $t_{LS}(i,j)$;

(6) 确定总时差 $R(i,j)$,单时差 $r(i,j)$;

(7) $R(i,j)=0$ 的工作所连成的线路为关键线路。

11.3.2 图算法

这种算法是一种简单、明了、方便的算法,但只适用于 50 个工作以下的小网络。如果超过 50 个工作,则因网络图复杂而不方便,容易出错,故不易采用。图算法的实质是把分

析法每步的计算结果用不同的符号标在图上。仍以图 11.3.1 为例说明。用上述原理在图上算出数据，以□表示事项最早开工时刻，以△表示事项的最晚完工时刻，各事项上△和□内数字相等即标志着总时差为零。把这样的点连接起来就是关键线路，然后计算出各种时差并标在图上，从而完成了整个计算，如图 11.3.6 所示。

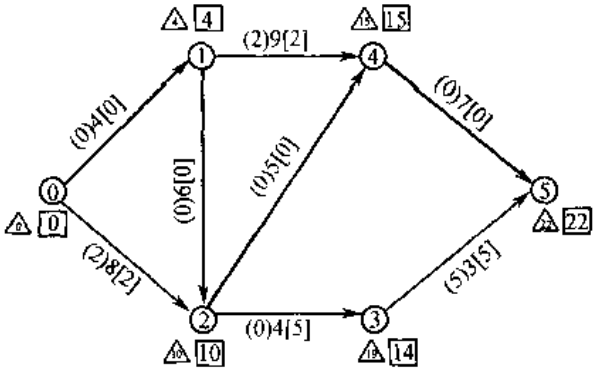


图 11.3.6 图算法

其中，( )内是单时差；[ ]内是总时差。

11.3.3 表算法

这种算法适用于大于 50 个工作的情况，100 个左右工作比较适宜。它的原理是采用分析法在表格上进行计算，这就避免了图算法所造成的混乱，在表格上可以清楚地计算出时差和关键线路。

该方法先编制一个合适的表格(见表 11.3.2)并把已知数据填入表内。在填数据时，工作的顺序要由小到大，即首先是 i 由小到大，在相同的 i 中，j 由小到大。

表 11.3.2 时间参数计算表

工序号		工期	最早开工	最早完工	最迟开工	最迟完工	总时差	单时差	关键线路
i	j	$T(i,j)$	t_{ES}	t_{EF}	t_{LS}	t_{LF}	$R(i,j)$	$r(i,j)$	
1	2	3	4	5=3+4	6=7-3	7	8=6-4	9	10
0	1	4	0	4	0	4	0	0	√
0	2	8	0	8	2	10	2	2	
1	2	6	4	10	4	10	0	0	√
1	4	9	4	13	6	15	2	2	
2	3	4	10	14	15	19	5	0	
2	4	5	10	15	10	15	0	0	√
3	5	3	14	17	19	22	5	5	
4	5	7	15	22	15	22	0	0	√

在计算各工作的最早开始时刻和最早结束时刻时，确定 t_{ES} 是关键。表算法确定最早开始时刻和最早结束时刻的方法是：

(1) 以零事项开始的工作，$t_{ES}(0)=0$；

(2) 按 $t_{EF}(0,j)=0+T(0,j)$ 计算这类工作的最早结束时刻；

(3) 以非零事项开始的工作，其最早开始时间为 $t_{ES}(i,j)=\max\{t_{ES}(k,i)+T(k,i)\}$ 因此需由上至下在已计算出的最早结束时间中查找与 i 事项号码相同的最早结束时间作为 i

事项的最早开始时间，这时有两种情况：①所有以 j 为结束事项的工作的最早结束时间均相同或只有一个，则 i 事项的最早开始时刻均为该值；②如不相同，则 i 事项的最早开始时刻应从中选一最大值，这样则体现出 i 事项的最早开始时刻是由紧前工作所决定这一条件；

(4) 其他工作的最早结束时刻，按 $t_{EF}(i,j)=t_{ES}(i,j)+T(i,j)$ 计算，即第三栏加第四栏。由此，第四栏和第五栏可求出。

然后求 $t_{LS}(i,j)$ 和 $t_{LF}(i,j)$，由于总工期即是最后工作的最早完工时刻，并已由上步算出，因此填入第七栏的相应位置上，由下向上按 $t_{LS}(i,j)=t_{LF}(i,j)-T(i,j)$ 即可求出第六栏和第七栏的所有数据，其中 $t_{LF}(i,j)$ 是由该步计算中由进入同一事项的 t_{LS} 中选一最小值所确定。

总时差则是第七栏和第五栏的差或第六栏与第四栏的差，即 $t_{LF}(i,j)-t_{EF}(i,j)=R(i,j)$ 或 $t_{LS}(i,j)-t_{ES}(i,j)=R(i,j)$。

单时差则根据 $r(i,j)=t_{ES}(j,k)-t_{EF}(i,j)$ 计算。如确定(0,2)工作的单时差 r，则由(2,3)工作或(2,4)工作的 $t_{ES}(2,3)=t_{ES}(2,4)=10$ 减去(0,2)工作的 $t_{EF}(0,2)=8$，故得 $r(0,2)=2$。

最后根据 $R(i,j)$ 为零的工作顺序，确定关键线路。时差的运用可根据 $R(i,j)$ 和 $r(i,j)$ 考虑。

本例题则得 0－1、1－2、2－4、4－5、即 0－1－2－4－5 为关键线路。

11.3.4 矩阵法

由表算法可见，当工作数量多时仍很麻烦且易出错，尤其是单时差的计算。当任务由100项～200项左右工作组成，又无计算机时，采用矩阵法比较快。

这种算法是以事项为基础的，有 n 个事项的网络图，则构成一个 $n\times n$ 的矩阵，见表11.3.3。该矩阵的行元素表示从 i 事项出发的所有工作所对应的工期，列元素表示以事项 j 为终点的所有工作对应的工期。

表 11.3.3 矩 阵 表

i \ j	t_{LF}	0	4	10	19	15	22
t_{ES}	事项	0	1	2	3	4	5
0	0	0	4	8	∞	∞	∞
4	1	∞	0	6	∞	9	∞
10	2	∞	∞	0	4	5	∞
15	3	∞	∞	∞	0	∞	3
15	4	∞	∞	∞	∞	0	7
22	5	∞	∞	∞	∞	∞	0
时	差	0	0	0	5	0	0

其计算步骤是：

(1) 把所有工期填入矩阵内，如两事项间无工作，则填∞，在 $i=j$ 的位置填零。

(2) 按 $t_{ES}(i)=\max\{t_{ES}(k)+T(ki)\}$，求出 $t_{ES}(i)$，填入相应的位置。

设开始点事项 $t_{ES}(0)=0$，则 $t_{ES}(i)=0+4=4$，$t_{ES}(2)=\max\{0+8,4+6\}=10$，即是由事项 $j=2$ 的一列中有数字的元素分别与所对应的 i 事项的 $t_{ES}(i)$ 相加，从中找到最大

值，填入相应位置。如此继续，则 $t_{ES}(i)$一列可填满。

(3) 按 $t_{LF}(j)=\max\{t_{LF}(k)-T(jk)\}$，求出 $t_{LE}(j)$，填入相应位置。由于事项⑤的 $t_{ES}(5)$即是事项⑤的 $t_{LF}(5)$，故 $t_{ES}(5)=t_{LF}(5)=22$。则 $t_{LF}(4)=22-7=15$；$t_{LF}(3)=22-3=19$，$t_{LF}(2)=\min\{19-4,15-5\}=10$，也就是 j 所对应的行中有数字的元素，与 t_{LF}相减后取最小值填入相应的位置。这样 t_{LF}一行可填满。

(4) 按 $R(i)=t_{LF}(i)-t_{ES}(i)$计算时差。对于事项的总时差可以这样理解：事项是两个时刻的汇集，如果两个时刻重合，则总时差为零，即表示该事项没有间隔时间；如果不为零，则说明该事项有间隔时间，这个时间则是与该事项有联系的宽裕时间。

因此时差为零的事项是关键路线：0—1—2—4—5 而③事项则有 5 天宽裕时间。

这种算法在所有笔算法中是最快、最方便的一种，在无计算机的情况下，是一种很适用的方法。

11.3.5 计算机算法

在计划比较大且网络图比较复杂的情况下，用上述各算法均比较困难，加之任何计划在执行的过程中，都需要经常地进行调整，而且必须在一定的时间内得出新调整后的计划，这样采用上述办法则不能满足工作的需要。计算机由于计算速度快，而且一旦软件准备就绪，随时可以调用，这对计划调整非常便利，因此用计算机进行网络的计算并对任务进行管理是一好方法。目前，成熟的软件较多，如：MP，P3 等。

计算机算法的根本原理仍是以分析法为基础，计算时间参数的公式，即是编制程序的数学模型。

计算机可进行如下工作：

(1) 根据给定的逻辑关系表，计算各时间参数；

(2) 确定网络的关键线路；

(3) 画出标有关键线路、工作时差的横道图和网络图；

(4) 进行资源、工期调整等。

由于计划本身是一种预测，因此在编制计划和计划执行中，都会产生偏离，采用其他方法时，必须重新做起，而应用计算机可很方便地进行调整与控制，调整可以分为在编制计划中的调整和执行过程中的调整两种，前者称静态调整，后者称为动态调整。

(1) 静态调整。当网络图编制完后，通过计算得出总工期，关键线路和横道图以及资源利用状况，若满足实际要求，则证明计划是可行的，否则需要进行调整。这种调整就需对网络图、工期和任务的分解等工作重新综合考虑，重新计算做出调整。

(2) 动态调整。在网络图已确定的情况下，在计划执行的过程中，由于受到意外因素的影响，而使计划与实际发生偏离所进行的调整叫动态调整。动态调整是在计划执行过程中进行，因此，根据已执行完的网络结构和参数对尚未执行的计划做修改。调整的方法是根据已执行的情况，对工艺流程进行改变，以促使总工期提前或通过调入资源、加班、加点或引进新技术等方法使未执行部分得到一个新的加快工期，使总工期下降。程序框图如图 11.3.7 所示。

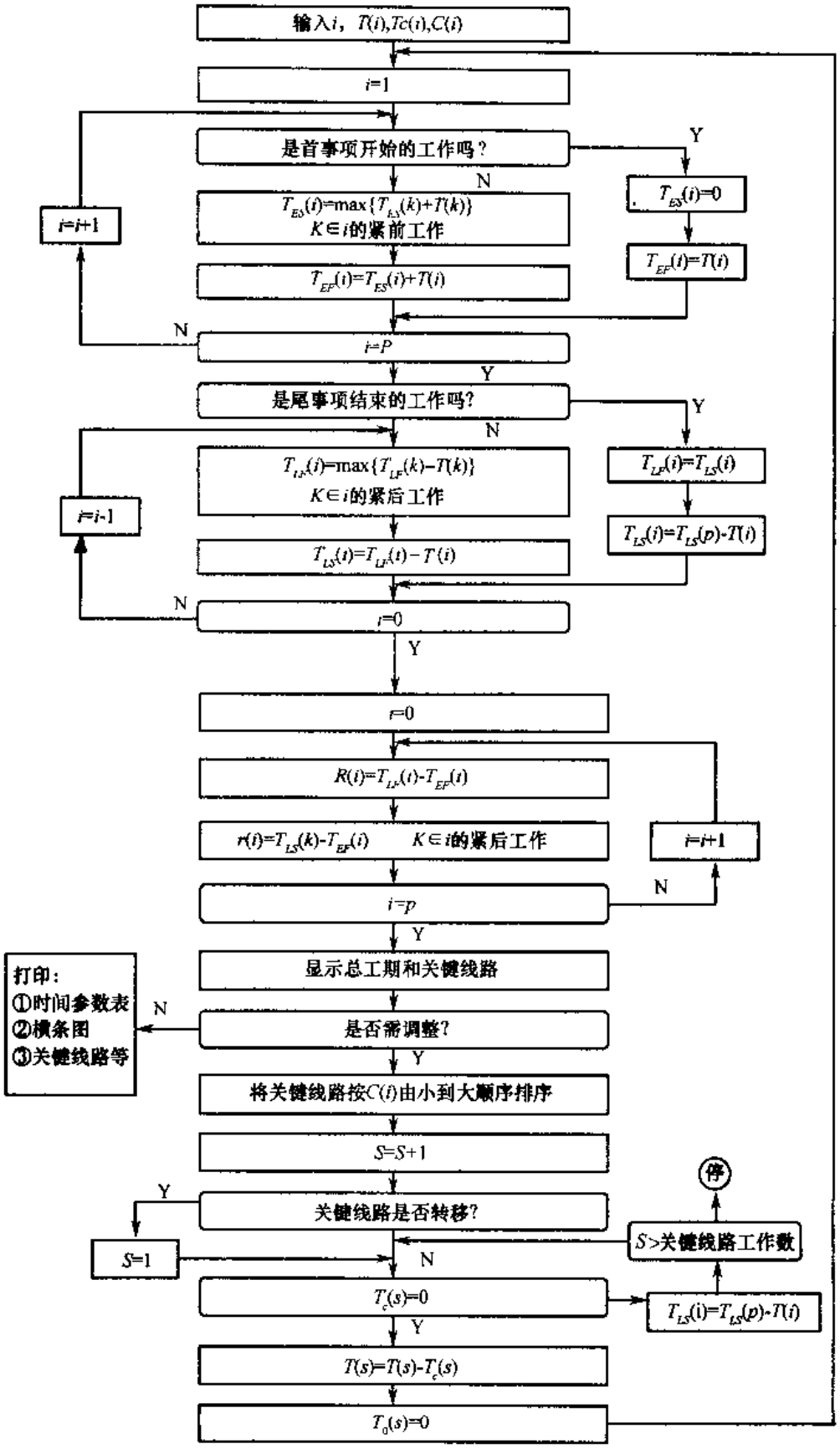


图 11.3.7　CPM 网络计算与调整框图

11.4 CPM 网络的优化

最优计划是以最短的时间(或在一定的时间内)用最少的人力、财力、物力得到最大效果的计划。网络优化就是应用网络图所反映的客观规律,通过网络图的各种安排,制订一最优计划的过程。

网络图的优化主要有时间优化、时间—费用优化、流程优化和资源优化等。

11.4.1 时间的优化

时间优化主要解决如何缩短总工期的问题。

任何网络图,经计算都可得到完成任务的总工期 $T_{总}$,该总工期与规定的工期 $T_{规}$ 之间有如下三种关系:

(1) $T_{总} < T_{规}$,这种情况说明任务的要求并不紧迫,关键线路还可以延长,可投入更少的人、财、物力资源,使得资源的消耗强度降低。

(2) $T_{总} = T_{规}$,这种情况是较合适的计划,无需修改。

(3) $T_{总} > T_{规}$,这种情况说明计划总工期不能满足实际需要,必须对网络图进行修改和调整。

工期的长短主要取决于关键线路,如果关键线路上的工序工期延长,总工期必定延长,关键线路上的工序工期缩短,总工期相应缩短,但工期缩短是有前提的,这就是说只有在关键线路不发生转移的前提下,总工期才随着关键线路上工序工期的缩短而缩短。如果关键线路发生转移,则工期的长短取决于新的关键路线。因此,这种情况下,工期的缩短是受限制的。现主要介绍 $T_{总} > T_{规}$ 时,压缩工期的方法。

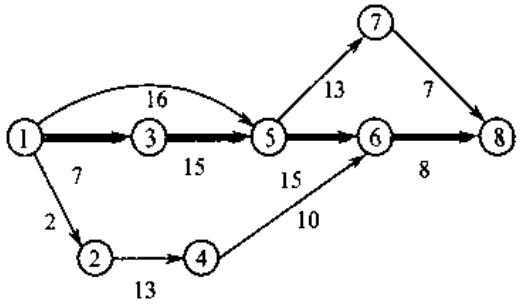


图 11.4.1 网络图

11.4.1.1 消除负时差法

如图 11.4.1 所示的网络图,关键线路 1—3—5—6—8 决定了总工期 $T_{总}$ = 45 天。如果要求在 40 天内必须完成该任务,则必须对网络图进行调整。但把关键线路中某一工序缩短 5 天,总工期并不一定缩短 5 天,如把⑤—⑥工序缩短 5 天,总工期并不能缩短到 40 天,而只能缩短 3 天,这是由于关键线路转移到 1—3—5—7—8 而造成的。但采取一些办法,总可以找出缩短工期的途径。

把由上图计算出的总时差列于表 11.4.1 中。

表 11.4.1 总时差

工序	总时差	关键路	工序	总时差	关键路
1—2	12		4—6	12	
1—3	0	√	5—6	0	√
1—5	6		5—7	3	
2—4	12		6—8	0	√
3—5	0	√	7—8	3	

表 11.4.2 总时差的变化

工序	总时差	关键路	工序	总时差	关键路
1—2	7		4—6	7	
1—3	−5	√	5—6	−5	√
1—5	1		5—7	−2	
2—4	7		6—8	−5	√
3—5	−5	√	7—8	−2	

如果把总工期定为40天，也就是将事项⑧的最晚并工时刻定为40，那么总时差的变化见表11.4.2，由表可见，关键线路的时差变为-5，因此关键线路又可定义为：包括负数在内的最小宽裕时间工序所组成的线路。次关键线路可定义为：最负总时差和次负总时差所组成的线路。

图11.4.2为工序⑤—⑥的工期缩短5天后的网络图，它的总工期为42天，不满足工期要求，因为仍存在着负的时差，即次关键线路成为关键线路，由此可知，在调整中只要使网络图不存在负时差，即可达到真正缩短工期的目的。在此，采用了图11.4.3的调整方法。即在③—⑤工序缩短2天，在⑤—⑥工序缩短3天。

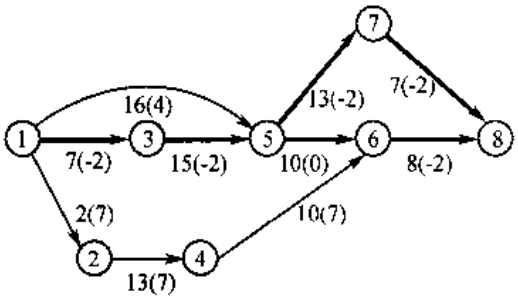


图11.4.2　工期缩短的网络图

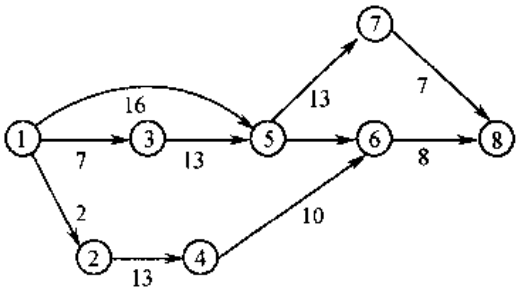


图11.4.3　调整后的网络图

消除负时差法的具体方法是：

(1) 按要求工期计算出各工序的总时差，并以负时差确定关键线路和次关键线路。

(2) 如仅关键线路的时差为负时，按调整费用最少的原则选择一个或几个工序进行调整，直至使负时差为零。

(3) 如存在次关键线路，则说明为达到预定工期，在某些关键线路上调整可能使关键线路转移。此时，若关键线路和次关键线路存在公共工序，且公共工序的调整费用最低，则可在公共工序上调整，使总时差为零。如不存在公共工序或公共工序调整费用较高，则需在关键和次关键线路上分别调整一些工序。如将关键线路和次关键线路的非公共工序部分称为平行路，则两平行路时差之差可称为平行路的有效松弛量，显然，该有效松弛量为次负时差与最负时差之差。以有效松弛量为调整量调整关键线路上调整费用最少的非公共工序，则可得到两条关键线路。

(4) 继续以新有效松弛量为调整量调整两条关键线路中费用较少的工序，直至使总时差全部为零为止。

(5) 继续以新有效松弛量为调整量调整两条关键线路中费用较少的工序，直至使总时差全部为零为止。

两条以上关键线路工期的调整原则是在关键线路中选择调整费用最低的工序，如是公共工序，可直接压缩；若不是公共工序，则需分别考虑公共工序和非公共工序的调整费用来决定。

11.4.1.2　改串行工作为并行工作

改串行作业为并行作业是一种常用的且简单的缩短工期的方法，该方法把可分段进行的工作分段进行。即变顺序进行的工作为并行工作。如挖地基和浇注混凝土两道工序，本应在挖完地基之后再进行浇注混凝土的工作，如图11.4.4所示。但这种工作可交叉进行，如当地基挖一半时，即可边挖地基边浇注，因此一段时间内两工作可并行，从而节

约了时间，如将网络图改成图 11.4.5 的形式，总工期可由 6+6=12 天缩短为 3+3+3=9 天，总工期缩短 3 天。如果这类工序在关键线路上，即可以采用这种办法缩短工期。

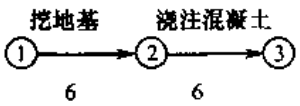


图 11.4.4 串行工作

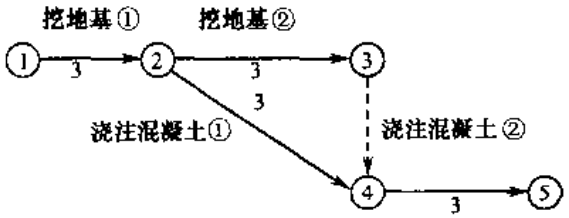


图 11.4.5 改串行工作为并行工作

11.4.2 时间—费用优化

任务的时间费用优化主要解决以下问题：

(1) 在限定工期内寻找调整费用最小的方案；

(2) 寻找总调整费用最小的工期。这两类问题的基本思想是把线性规划和 CPM 网络结合起来，即把网络结构和限定的工期作为约束条件，把调整所需的费用作为目标，应用线性规划的方法确定工序调整时间，进而编制出任务计划。

11.4.2.1 限定工期，以最小调整费用为目标的线性规划方法

消除负时差法是以时间为主，在调整中兼顾费用的调整方法。以最小费用为目标的线性规划法与上述方法不同，该方法是在详细计算调整费用变化的基础上进行工序调整，确定各工序开工时间以达到预定工期。

现将包括 m 个工序 n 个节点的 CPM 网络图记为 $R(m,n-1)$，并令第 k 工序的开始事项、结束事项、正常工期、压缩后的工期、调整费用分别为 P_k、S_k、a_k、b_k 和 $C_k(k=1,2,\cdots,m)$。

如果 $R(m,n-1)$的各工序均按正常工期 a_k 工作，则总工期必定最长，此时工期以 T_{max}表示。如果各工序均按压缩后的工期 b_k 工作，则总工期必最短，此时工期以 T_{min}表示。第 k 工序的调整范围为 a_k-b_k，如设第 k 工序的实际作业时间为 t_k，则第 k 工序的结束事项 S_k 和开始事项 P_k 的发生时刻不能小于工序的作业时间，即 $t(S_k)-t(P_k)\geqslant t_k$，如限定工期为 $T(T_{min}\leqslant T\leqslant T_{max})$，问题可用下述线性规划模型描述：

目标函数：
$$Z=\sum_{k=1}^{m}C_k(a_k-t_k)\rightarrow\min \tag{11.4.1}$$

s.t. (1) 网络结构限制
$$t(S_k)-t(P_k)\geqslant t_k \qquad (k=1,2,\cdots,m) \tag{11.4.2}$$

(2) 工期限制：
$$t(n-1)-t(0)=T \tag{11.4.3}$$

(3) 工序时间限制：
$$b_k\leqslant t_k\leqslant a_k \qquad (k=1,2,\cdots,m) \tag{11.4.4}$$

(4) 非负条件
$$t(p_k)\geqslant 0,t_k\geqslant 0\ (k=1,2,\cdots,m;\ p_k=1,2,\cdots,n-1) \tag{11.4.5}$$

式中 $t(p_k)$——k 工序开始事项发生时刻；

$t(s_k)$——k 工序结束事项发生时刻。

令 $a_k - t_k = \tau_k$，$t(0)=0$，且引入松弛变量，上述模型变为

目标函数：
$$Z=\sum_{k=1}^{m} C_k\tau_k \to \min \tag{11.4.6}$$

s.t. (1) 将式(11.4.2)与 $\tau_k = a_k - t_k$ 相加后引入松弛变量即可得网络结构限制：
$$t(S_k) - t(P_k) + \tau_k - \sigma_k = a_k \quad (k=1,2,\cdots,m) \tag{11.4.7}$$

(2) 因 $t(0)=0$，工期限制为：
$$t(n-1)=T \tag{11.4.8}$$

(3) 工序时间限制改写成：
$$\tau_k \leqslant a_k - b_k \tag{11.4.9}$$

(4) 非负条件：
$$\tau_k \geqslant 0,\ t(S_k)\geqslant 0,\ t(p_k)\geqslant 0,\ \sigma_k \geqslant 0 \tag{11.4.10}$$

式中 σ_k——松弛变量，其物理意义为宽裕时间。

在 $R(m,n-1)$中，不可压缩工序的特征为
$$b_k = t_k,\ \tau_k = 0$$

故这类工序在约束式(11.4.7)中的形式是：
$$t(S_k) - t(P_k) - \sigma_k = a_k \tag{11.4.11}$$

以 $n-1$ 为结束事项的工序，因 $t(S_k)=T$，故该类工序在约束式(11.4.7)中的形式为
$$t(P_k) - \tau_k + \sigma_k = T - a_k \tag{11.4.12}$$

将上述模型写成矩阵形式为
$$Z = C^{\mathrm{T}}W \to \min \tag{11.4.13}$$

s.t.
$$R(m,n-1)t - W - \sigma = a \tag{11.4.14}$$
$$t(n-1)=T \tag{11.4.15}$$
$$W \leqslant D \tag{11.4.16}$$
$$W\geqslant 0,\ t\geqslant 0,\ \sigma \geqslant 0 \tag{11.4.17}$$

其中
$$C^{\mathrm{T}}=(C_1,C_2,\cdots,C_m)$$
$$a=(a_1,a_2,\cdots,a_m)^{\mathrm{T}}$$
$$W=(\tau_1,\tau_2,\cdots,\tau_m)^{\mathrm{T}}$$
$$b=(b_1,b_2,\cdots,b_m)^{\mathrm{T}}$$
$$\sigma=(\sigma_1,\sigma_2,\cdots,\sigma_m)^{\mathrm{T}}$$
$$t=(t_{(1)},t_{(2)},\cdots,t_{(n-1)})^{\mathrm{T}}$$
$$D=a-b=(d_1,d_2,\cdots,d_m)^{\mathrm{T}}$$
$R(m,n-1)$为网络矩阵①

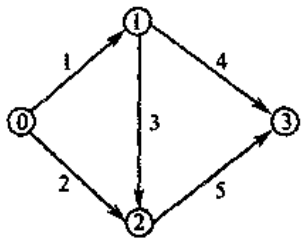


图 11.4.6 网络图

① 网络矩阵是指网络图中工序与节点关系的矩阵，$R(m,n-1)$网络的网络矩阵为 $m\times n$ 阶矩阵，矩阵中第 j 列第 k 行元素 $R(k,j)$为
1，当 k 工作指向 j 事项时；
$R(k,j)=0$，当 k 工作与 j 事项无关时；
-1，当 k 工作离开 j 事项时。
例中的 $R(m,n-1)$为按上述规则形成网络矩阵后，去掉第一列后的矩阵，其原因为 $t(0)=0$。

现举例说明该模型。

将如图 11.4.6 所示网络图的数据列于表 11.4.3。

表 11.4.3　网络图的数据表

工序 k	开始事项 P_k	结束事项 S_k	正常工期 a_k	压缩后工期 b_k	单位费用 C_k
1	0	1	10	5	3
2	0	2	8	6	4
3	1	2	5	3	2
4	1	3	9	4	5
5	2	3	4	2	1

这是一 R (5,3)网络。

目标函数：$Z=3\tau_1+4\tau_2+2\tau_3+5\tau_4+\tau_5\rightarrow\min$

s.t.

(1) 网络结构限制：

$$\begin{bmatrix}1&0&0\\0&1&0\\-1&1&0\\-1&0&1\\0&-1&1\end{bmatrix}\begin{bmatrix}t_{(1)}\\t_{(2)}\\t_{(3)}\end{bmatrix}+\begin{bmatrix}\tau_1\\\tau_2\\\tau_3\\\tau_4\\\tau_5\end{bmatrix}-\begin{bmatrix}\sigma_1\\\sigma_2\\\sigma_3\\\sigma_4\\\sigma_5\end{bmatrix}=\begin{bmatrix}10\\8\\5\\9\\4\end{bmatrix}$$

(2) 工期限制：

$$t(3)=T(T\text{ 为预先给定的工期})$$

(3) 工序时间限制：

$$\begin{bmatrix}\tau_1\\\tau_2\\\tau_3\\\tau_4\\\tau_5\end{bmatrix}\leqslant\begin{bmatrix}d_1\\d_2\\d_3\\d_4\\d_5\end{bmatrix}=\begin{bmatrix}5\\2\\2\\5\\2\end{bmatrix}$$

(4) 非负限制：

$$\tau_k\geqslant0,t(i)\geqslant0\qquad k=1,2,\cdots,5;$$

$$\sigma_k\geqslant0\qquad i=1,2,3$$

求解该线性规划模型即可得到 τ_k，由 $a_k-\tau_k$ 即可确定各工序的实际作业时间。$t(i)$为各节点的发生时刻。由此可得到满足既定工期并使调整费用最省的调整方案。

11.4.2.2　确定最小费用工期的方法

1. 传统方法

实践中常需探讨以最小调整费用为目标求最优工期的方法。

工期和费用之间一般呈非线性关系，若准确地确定时间—费用曲线是件不容易的事情，其传统方法如下。

生产费用可分为直接费用和间接费用两种，采取增加资源的方法缩短工期，其直接费用变化曲线如图 11.4.7(a)所示，由于压缩了工期，使间接费用减少，间接费用变化曲线如图 11.4.7(b)所示，总费用的变化为二者的合成，如图 11.4.7(c)所示。

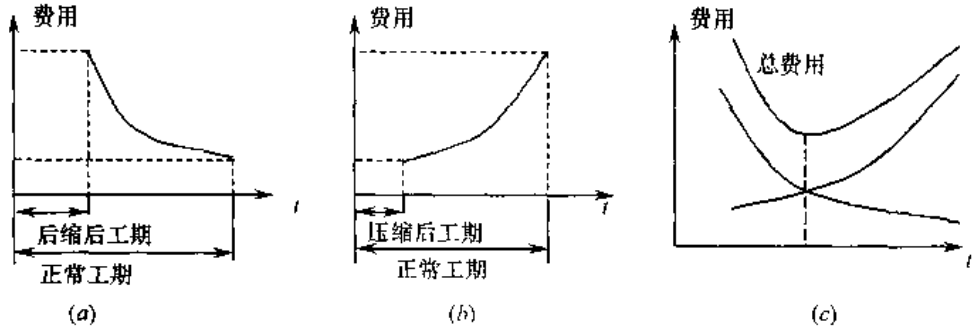


图 11.4.7　时间—费用曲线

因直接费用和间接费用与时间的变化规律一般均呈非线性关系，为了处理方便，现假设其为直线，故各工序压缩一天的直接费用 q 为

$$q=\frac{\text{压缩工期后费用}-\text{正常工期费用}}{\text{正常完工工期}-\text{压缩后工期}}$$

总工期每压缩一天的间接费用以生产管理费用的平均值 d 计算。

这样即可按不同压缩方案描出总费用变化曲线，找出最小点。从而可确定最小费用工期。

现举例说明该算法。

图 11.4.8 所示网络的各项费用如表 11.4.4。

关键线路为 0—3—5。

在关键线路上压缩工期：

方案①　正常工期(60 天)的费用为

$$12800+60\times200=24800(\text{元})$$

方案②　将工序 G 压缩 10 天，关键线路未转移，总工期为 50 天，费用为

$$12800+50\times200+10\times150=24300(\text{元})$$

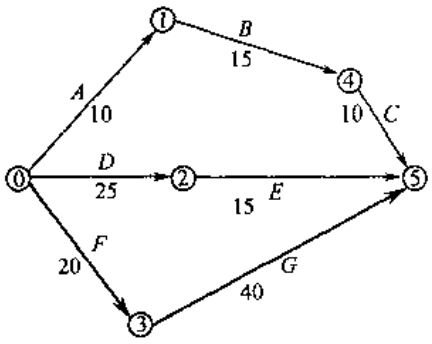


图 11.4.8　网络图

表 11.4.4　图 11.4.8 所示网络的各项费用

工序	正常完工直接费用	压缩工期的单位费用
A	800	—
B	3000	18
C	2000	120
D	1500	5 日内 60，5 日外 100
E	3000	120
F	500	—
G	2000	150
	间接费用 200/日	

方案③　再将关键工序 G 压缩 10 天，关键线路虽然转移，但出现两条关键线路，总工期为 40 天，费用为

$$12800+20\times150+40\times200=23800(\text{元})$$

方案④　同时压缩 G 和 D 两工序，G 压缩 15 天，D 压缩 5 天，关键线路为 35 天，费

用为

$$12800+150\times25+60\times5+35\times200=23850(元)$$

依此类推即可画出时间—费用曲线，找出最小点，最小费用工期为 40 天。

由此例可见，该方法的弊端是：

(1) 确定压缩方案，计算关键线路等工作是复杂的，而所确定的压缩方案又不是最优方案。若得到同一工期的最优压缩方案，还需计算该工期所对应的大量方案，对于大型网络显然是行不通的；

(2) 因每得一个工期和该工期的相应费用即得到时间——费用曲线上的一点，因此选择方案数量的多少将直接影响时间费用曲线的精度，若选点多，精度可提高，但计算工作量加大，因此该方法只适用于小型简单网络。

2. 用线性规划法确定最小费用工期

应用 14.4.2.1 节中所建立的线性规划模型，可通过改变 T 的方法得到一条时间——费用曲线。其方法是让 T 在 T_{max} 和 T_{min} 之间变化，每选一个 T 则得到一相应的最小费用的调整方案，由此可得出不同工期与所对应最小费用的变化曲线。

如果将该线性规划模型与其他最优化方法结合起来会直接搜索出最小费用工期。

11.4.3 流程优化

同一问题采用不同的安排方法将有不同的完成时间，因此需进行流程的优化，排序问题是流程优化的方法之一，现介绍简单的排序问题。

11.4.3.1 几种产品分别在二种设备上加工的排序问题

某建筑公司现有 A、B、C 三种规格原木需分别在大、小二台带锯机上加工，原木必须先在大带锯上加工后，再在小带锯机上加工，根据各规格原木的数量和其在不同带锯上加工的效率，得加工时间见表 11.4.5，现对该任务进行流程优化。

该任务的流程安排方法共有六种。ABC、ACB、CAB、CBA、BAC、BCA。如原木规格很多则方案更多。每种流程安排方法可得一个网络图，现需从各种流程安排中选择一种最优的方案。

设各规格原木在大带锯上的加工时间分别为 $a_1,a_2,\cdots,a_n$，在小带锯上的加工时间为 $b_1,b_2,\cdots,b_n$，则排序问题的步骤为：

(1) 在 $a_1,a_2,\cdots,a_n;b_1,b_2,\cdots,b_n$ 中选最小元素；

(2) 若最小元素为 a_i，则将 i 排在第一位，并从原木集合中去掉该类；

表 11.4.5 加工时间

(单位：天)

锯机 / 原木	大	小
A	3	4
B	4	2
C	6	5

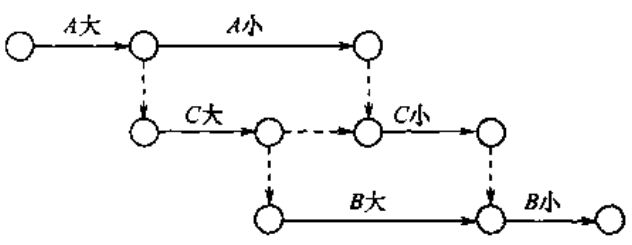


图 11.4.9 流程网络图

(3) 若最小者为 b_i,则将 i 排在最后一位,并从原木集合中去掉该类;

(4) 对其余各类原木重复(1)~(2),直至原木集合为空集为止。

如上例按此方法得加工顺序为 ACB,工期为 16 天,网络图如图 11.4.9 所示。

11.4.3.2 二种产品在不同设备上的排序问题

某建筑公司用长 1m 和 2m 的两种规格原木在各种机床上生产门、窗,两种规格的原木都需经过大带锯下料、小带锯下料、刨光、开榫四道工序,两种原料占用设备时间不同。现假设两种原木占用设备的时间见表 11.4.6。

现采用图 11.4.10 所示的示意图,图中斜线位置为干扰区,即两种规格原料需在一个设备加工的情况。

表 11.4.6 两种原木占用设备的时间

工序	2m	1m
a	5	2
b	3	3
c	1	2
d	2	1

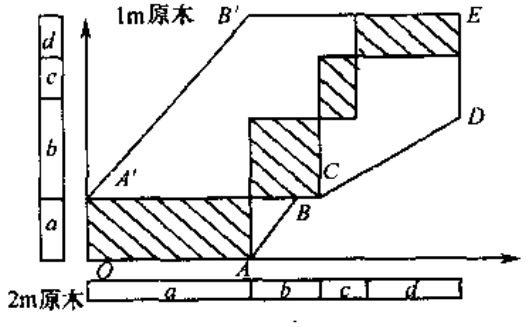


图 11.4.10 示意图

假设先加工 1m 长的原木,则 OA' 表示该原木在 a 工序加工,$A'B'$ 为一 45°斜线,它与干扰区不相交,说明可连续加工,与另一规格原木不争设备。其网络图如 11.4.11 所示。

再假设先加工 2m 长原木,则 OA 表示 2m 原木在 a 工序加工,1m 米原木等待,AB 表示 2m 原木在 b 工序加工,此时 1m 原木在 a 工序加工后,等待一天(BC),CD 表示 2m 原木在 c、d 二工序加工,1m 原木在 b 工序加工,DE 表示 1m 原木在 c、d 工序加工,因此可得网络图 11.4.12。

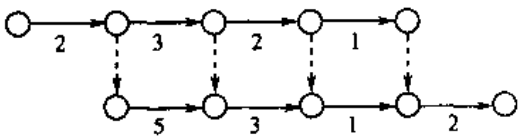


图 11.4.11 先加工 1m 原木网络图

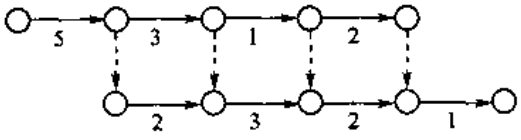


图 11.4.12 先加工 2m 原木网络图

由上述两网络图可见,第一方案工期为 13(天),第二方案为 14 天,第一方案优于第二方案。相差 1 天是由于第二方案 1m 原木在 BC 段(b 工序)等待一天的缘故。由图 11.4.10 可见,应选择等待时间短的方案,所得结果与网络图计算结果相同。

11.4.4 资源优化

在实践中,资源的数量是有限的,如果工作安排不当,可能导致任务不能如期完成。资源优化就是解决在资源量一定的情况下,使资源利用更合理(均衡)的问题。

合理安排资源的原则是:

(1) 当资源受限制时,在不推迟或少推迟工期的前提下,合理安排和充分利用资源;

(2) 首先保证各关键工序的资源需求量;

(3) 利用非关键工序的时差,调整其开工和完工时刻,使资源达到均衡利用。

当前进行资源优化的方法主要是试探法。由于该方法比较麻烦、不易量化且需有丰富的经验,故现建议一种可量化的调整资源利用强度的方法。

衡量生产资源的利用是否均衡、合理,需引入方差的概念。

如果资源量是时间的函数,在连续状态下其方差为

$$U = \frac{1}{T}\int_0^T (F_{(t)} - F_m)^2 \mathrm{d}t \tag{11.4.18}$$

式中 T——工期;

$F_{(t)}$——为 t 时刻的资源量;

F_m ——为资源的平均值,且

$$F_m = \frac{1}{T}\int_0^T F_{(t)} \mathrm{d}t \tag{11.4.19}$$

将式(11.4.19)代入式(11.4.18)得

$$U = \frac{1}{T}\int_0^T F_{(t)}^2 \mathrm{d}t - F_m^2 \tag{11.4.20}$$

在离散状态下的方差为

$$U=(F_1-F_m)^2+(F_2-F_m)^2+\cdots+(F_n-F_m)^2= F_1^2+F_2^2+\cdots+F_n^2-nF_m^2 \tag{11.4.21}$$

假设 k_j 工作所需资源量为 f_{kj},如果 k_j 工作向后推迟一天,则必然影响整个资源的分布。

设第 i 天所需资源量为 F_i,第 $j+1$ 天为 F_{j+1}。当 k_j 工作向后推迟一天时,第 i 天所需资源量减少 f_{kj},变成 $F'_i = F_i - f_{kj}$;第 $j+1$ 天的资源增加 f_{kj},变成 $F'_{i+1} = F_{i+1} + f_{kj}$;第 $i+1$ 天到 j 天所需的资源量仍保持不变(见图 11.4.13);此时总方差的变化为

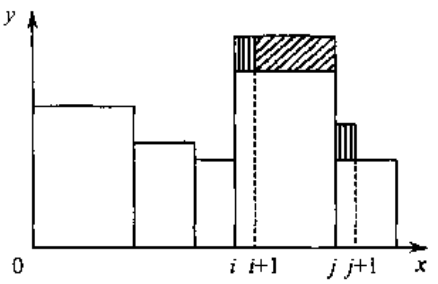


图 11.4.13 资源优化

$$\Delta U=[(F_{j+1}+f_{kj})^2-F_{j+1}^2]-[(F_j-f_{kj})^2-F_j^2]$$

整理得

$$\Delta U=2f_{kj}[F_{j+1}-(F_j-f_{kj})]$$

由图 11.4.13 可知,因 $f_{kj}\geqslant 0$,所以

$$\begin{aligned} &\text{当} \quad [F_{j+1}-(F_j-f_{kj})]<0 \text{ 时},\Delta U<0 \\ &\text{当} \quad [F_{j+1}-(F_j-f_{kj})]=0 \text{ 时},\Delta U=0 \\ &\text{当} \quad [F_{j+1}-(F_j-f_{kj})]>0 \text{ 时},\Delta U>0 \end{aligned} \tag{11.4.22}$$

$\Delta U=0$ 说明 k_j 工序推迟一天,资源利用与推迟前相同;$\Delta U>0$,说明 k_j 工作推迟一天以后,方差比以前更大,资源利用更不均衡了;当 $\Delta U<0$ 时,说明 k_j 工作推迟一天以后,资源利用较以前均衡了。因此 ΔU 即可作为评价某项工作向后推移一天时,资源利用变化情况的指标。

如 k_j 工作向后推移二天,则可写成

$$\Delta U = F_{j+1} - (F_j - f_{kj}) + F_{j+2} - (F_{j+1} - f_{kj}) \tag{11.4.23}$$

按此规律可写出某工作向后推移 n 天的判别式。

在执行某项任务时，应注意关键工序，而在研究资源利用时需利用非关键线路上各工序的宽裕时间，因此资源调整方法和步骤如下：

(1) 首先利用带关键线路和有时差标记的横条图，将非关键工序放置最早开工时刻，画出相应的资源分布图。

(2) 由后向前在非关键线路上确定调整工序并在全部可能的调整范围内进行调整，确定较好的开工时刻，直至所有非关键线路上的工序进行完为止。

(3) 再按第二步计算，直至找不出更均衡的方案为止。

值得指出的是第一次全部调整完后，并不等于调整工作结束。因其中某一工序调整之后，其他工序的变化将对它产生影响，因此必须按第二步重新调整，直至得到较满意的方案为止。一些软件(如 P3)已按该原理设计了程序，可方便地调用。

11.5 计划评审技术(PERT)

在 CPM 法中，各工序的工期是确定的，因此它只适用于已经定型或有足够经验的任务。在很多大型开发性工程或无足够经验的任务中，因无充分、准确的资料，故各工序的工期只能由估计得出。把估计的工期作为 CPM 网络的工期，即构成 PERT 网络。应用 PERT 网络制定计划的方法叫计划评审技术，即 PERT 法。

现将 PERT 网络的常用的处理方法简介如下。

11.5.1 根据平均作业时间确定完成任务总工期及概率的方法

该方法认为：PERT 网络和 CPM 网络是本质相同而处理对象不同的两种方法，即认为 CPM 是 PERT 的一种特例。

根据由特殊到一般的规律，只要掌握 CPM 和 PERT 的不同点即可根据 CPM 法处理 PERT 网络，从而可处理这类开发性工程。

该方法对工作工期的估计一般采用“三时估计法”。所谓“三时估计法”就是对该任务的每项工作事先估计出最乐观完成时间 t_0、最可能完成时间 t_m 和最悲观完成时间 t_p。最乐观完成时间 t_0 是指完成该工作时，一切如所期望的那样，很顺利地完成所需要的时间；最可能完成时间 t_m 是指如果多次重复地进行该项工作，在该时间完成该项工作的概率最大。如果采用单一时间估计法，该时间就是“计划工期”。最悲观完成时间 t_p 是指在一切不顺利、一切不如意的情况下完成该工作所需要的时间。

任何一项工作，都可估计出以上三种时间，因此这是一种工作执行时间具有随机性的问题。

通过数学方法将这类随机问题化成确定型问题之后，即可采用 CPM 法解决 PERT 网络了。

分别令 $t_0 = a$、$t_m = m$、$t_p = b$，则该工作的平均工期为

$$t_E = \frac{1}{6}(a + 4m + b) \tag{11.5.1}$$

方差和均方差为

$$V=\frac{1}{36}(b-a)^2 \tag{11.5.2}$$

$$\sigma=\frac{1}{6}(b-a) \tag{11.5.3}$$

式(11.5.1)中 t_E 实质是一加权平均数,是假设工作的可能完成时间权数为 4,乐观时间和悲观时间的权数为 1。

式(11.5.2)和式(11.5.3)表示工作时间与平均值的离散程度,V 和 σ 值大,则说明平均值的代表性差;反之,则代表性强。

以 t_{Ei} 作为 CPM 网络的工期,则可采用 CPM 法解决 PERT 网络,确定出关键线路和总工期。

总工期为

$$T_{cp}=\sum_{i=1}^{k}\frac{1}{6}(a_i+4m_i+b_i) \tag{11.5.4}$$

式中 k——关键线路上的工作数。

该总工期的总方差为

$$\sum V=\frac{1}{36}\sum_{i=1}^{k}(b_i-a_i)^2 \tag{11.5.5}$$

该总工期的总均方差为

$$\sum\sigma=\frac{1}{6}\sqrt{\sum_{i=1}^{k}(b_i-a_i)^2} \tag{11.5.6}$$

如果完成任务时间 t 服从参数为$(T_{cp},\sum\sigma^2)$的正态分布,则其概率密度函数为

$$P(t_k)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{(t_k-T_{cp})^2}{2\sigma^2}} \tag{11.5.7}$$

分布函数为

$$F(T_k)=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int_0^{\frac{T_k-T_{cp}}{\sqrt{2\pi\sigma^2}}}e^{-\frac{t^2}{2}}dt=\Phi\left(\frac{T_k-T_{cp}}{\sqrt{\sum V}}\right)=\Phi(\lambda) \tag{11.5.8}$$

式中 T_k——任务规定完工工期。

由式(11.5.8)知,当 $T_k=T_{cp}$ 时,$F(T_k)=\frac{1}{2}$。这说明当用平均工期计算出的总工期作为计划工期时,其完成概率为 0.5(见图 11.5.1)。

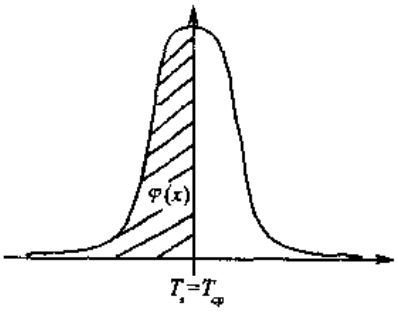


图 11.5.1 完成概率

如设 $\frac{T_k-T_{cp}}{\sqrt{\sum V}}=\lambda$,则根据正态分布表(见表 11.5.1)可确定在 T_k 完成任务的概率 $F(T_k)$。

λ 称计划难易程度系数。当 $\lambda\leqslant-3$ 时,表明任务在规定的日期内很难完成;当 λ 在区间$(-3,-0.5)$时,表示需尽很大努力才能完成任务;当 λ 在区间$(-0.5,0.5)$时,表示任务在规定期限内完成任务可能性很大;当 λ 在区间$(0.5,3)$时,表示容易按期

完成计划；当 $\lambda \geqslant 3$ 时，表示任务在指定日期内完成很容易，计划留有余地太大，比较保守。

λ 为负值的计划比较先进，是较紧的计划；λ 为正值的计划是留有余地的计划。

由上述分析可知，CPM 法和 PERT 法的区别是：

(1) 工作工期的确定方法不同。CPM 法是确定型工期，PERT 法是随机型工期。

(2) 总工期的确定方法不同。CPM 法是确定型的总工期；而 PERT 法根据计划难易程度系数确定指定日期完成任务的概率，计算出的平均总工期 T_{cp} 仅有 0.5 的概率能够实现。

(3) 解决问题的侧重点不同。CPM 法主要解决完成任务的关键工作；而 PERT 法主要解决在指定日期完成任务的可能性，而对任务进行计划、评价和审查。

该方法是当前最广泛采用的方法。该方法的缺点是总工期以关键线路上工作的平均工期确定的，这就导致在确定关键线路时并未考虑工作执行时间的随机性，故用它计算的在指定日期完成任务的概率是不可靠、不合理的，只能作为编制计划的参考。

表 11.5.1　求任务在指定日期 T_k 完成的概率 $P(T_k)$——拉普拉斯函数表

λ	$P(T_k)$	λ	$P(T_k)$	λ	$P(T_k)$	λ	$P(T_k)$
-5.00	0.000	-2.28	0.012	-1.71	0.044	-1.28	0.100
-4.50	0.000	-2.26	0.012	-1.70	0.045	-1.27	0.102
-4.00	0.000	-2.24	0.013	-1.69	0.046	-1.26	0.104
-3.80	0.000	-2.22	0.013	-1.68	0.047	-1.25	0.106
-3.60	0.000	-2.20	0.014	-1.67	0.048	-1.24	0.108
-3.40	0.000	-2.18	0.015	-1.66	0.049	-1.23	0.109
-3.20	0.001	-2.16	0.015	-1.65	0.050	-1.22	0.111
-3.00	0.001	-2.14	0.016	-1.64	0.051	-1.21	0.113
-2.98	0.001	-2.12	0.017	-1.63	0.052	-1.20	0.115
-2.96	0.002	-2.10	0.018	-1.62	0.053	-1.19	0.117
-2.94	0.002	-2.08	0.019	-1.61	0.054	-1.18	0.119
-2.92	0.002	-2.06	0.020	-1.60	0.055	-1.17	0.121
-2.90	0.002	-2.04	0.021	-1.59	0.056	-1.16	0.123
-2.88	0.002	-2.02	0.022	-1.58	0.057	-1.15	0.125
-2.86	0.002	-2.00	0.023	-1.57	0.058	-1.14	0.127
-2.84	0.002	-1.99	0.023	-1.56	0.059	-1.13	0.129
-2.82	0.002	-1.98	0.024	-1.55	0.061	-1.12	0.131
-2.80	0.003	-1.97	0.024	-1.54	0.062	-1.11	0.134
-2.78	0.003	-1.96	0.025	-1.53	0.063	-1.10	0.136
-2.76	0.003	-1.95	0.026	-1.52	0.064	-1.09	0.138
-2.74	0.003	-1.94	0.026	-1.51	0.066	-1.08	0.140
-2.72	0.003	-1.93	0.027	-1.50	0.067	-1.07	0.142

（续）

λ	$P(T_k)$	λ	$P(T_k)$	λ	$P(T_k)$	λ	$P(T_k)$
−2.70	0.004	−1.92	0.027	−1.49	0.068	−1.06	0.145
−2.68	0.004	−1.91	0 028	−1.48	0.069	−1.05	0.147
−2 66	0.004	−1.90	0.029	−1.47	0.071	−1.04	0.149
−2.64	0.004	−1.89	0.029	−1.46	0.072	−1.03	0.152
−2.62	0.004	−1.88	0.030	−1.45	0.074	−1.02	0.154
−2.60	0.005	−1.87	0.031	−1.44	0.075	−1 01	0.156
−2.58	0.005	−1.86	0 031	−1.43	0.076	−1.00	0.159
−2.56	0.005	−1.85	0.032	−1.42	0.078	−0.99	0.161
−2.54	0.006	−1.84	0.033	−1.41	0.079	−0.98	0.164
−2.52	0.006	−1.83	0.034	−1.40	0.081	−0.97	0.166
−2.50	0.006	−1 82	0.034	−1.39	0.082	−0.96	0 169
−2.48	0.007	−1.81	0.035	−1.38	0.084	−0.95	0.171
−2.46	0.007	−1.80	0.036	−1.37	0.086	−0.94	0.174
−2.44	0.007	−1.79	0.037	−1.36	0.087	−0.93	0.176
−2.42	0.008	−1.78	0.038	−1.35	0.089	−0.92	0.179
−2.40	0.008	−1.77	0 038	−1.34	0.090	−0.91	0.181
−2.38	0.009	−1.76	0.039	−1.33	0.092	−0.90	0.184
−2.36	0.009	−1.75	0.040	−1.32	0.093	0.89	0.187
−2.34	0.010	−1.74	0.041	−1.31	0.095	−0.88	0.189
−2.32	0.010	−1.73	0.042	−1.30	0.097	−0.87	0.192
−2.30	0.010	−1.72	0.043	−1.29	0.099	−0.86	0.195
−0.85	0.198	−0.42	0.337	0.01	0.504	0.44	0.670
−0.84	0.201	−0.41	0.341	0.02	0.508	0.45	0.674
−0.83	0.203	−0.40	0.345	0.03	0.512	0.46	0.677
−0.82	0.206	−0.39	0.348	0.04	0.516	0.47	0.681
−0.81	0.209	−0.38	0.352	0.05	0.520	0.48	0.684
−0.80	0.212	−0.37	0.356	0.06	0.524	0.49	0.688
−0.79	0.215	−0.36	0.359	0.07	0.528	0.50	0.692
−0.78	0.218	−0.35	0.363	0.08	0.532	0.51	0.695
−0.77	0.221	−0.34	0.367	0.09	0.536	0.52	0.699
−0.76	0.224	−0.33	0.371	0.10	0.540	0.53	0.702
−0.75	0.227	−0.32	0.375	0.11	0.544	0.54	0.705
−0.74	0.230	−0.31	0.378	0.12	0.548	0.55	0.709
−0.73	0.233	−0.30	0.382	0.13	0.552	0.56	0.712
−0.72	0.236	−0.29	0.386	0.14	0.556	0.57	0.716
−0.71	0.239	−0.28	0 390	0.15	0.560	0.58	0.719

（续）

λ	$P(T_k)$	λ	$P(T_k)$	λ	$P(T_k)$	λ	$P(T_k)$
−0.70	0.242	−0.27	0.394	0.16	0.564	0.59	0.722
−0.69	0.245	−0.26	0.397	0.17	0.568	0.60	0.726
−0.68	0.248	−0.25	0.401	0.18	0.571	0.61	0.729
−0.67	0.251	−0.24	0.405	0.19	0.575	0.62	0 732
−0.66	0.255	−0.23	0.109	0.20	0.579	0.63	0.736
−0.65	0.258	−0.22	0.413	0.21	0.583	0.64	0.739
−0.64	0.261	−0.21	0.417	0.22	0.587	0.65	0.742
−0.63	0.264	−0.20	0.421	0.23	0.591	0.66	0.745
−0.62	0.268	−0.19	0.425	0.24	0 595	0.67	0.749
−0.61	0.271	−0.18	0.429	0.25	0.599	0.68	0.752
−0.60	0.274	−0.17	0 433	0.26	0.603	0.69	0.755
−0.59	0.278	−0.16	0.436	0.27	0.606	0.70	0.758
−0.58	0.281	−0.15	0.440	0.28	0.610	0.71	0.761
−0.57	0.284	−0.14	0.444	0.29	0.614	0.72	0.764
−0.56	0.288	−0.13	0.448	0.30	0.618	0.73	0.767
−0.55	0.291	−0.12	0.452	0.31	0.622	0.74	0.770
−0.54	0.295	−0.11	0.456	0.32	0.626	0.75	0.773
−0.53	0.298	−0.10	0.460	0.33	0.629	0.76	0.776
−0.52	0.302	−0.09	0.464	0.34	0.633	0.77	0.779
−0.51	0.305	−0.08	0.468	0.35	0.639	0.78	0.782
−0.50	0.309	−0.07	0.472	0.36	0.641	0.79	0.785
−0.49	0.312	−0.06	0.476	0.37	0.644	0.80	0.788
−0.48	0.316	−0.05	0.480	0.38	0.648	0.81	0.791
−0.47	0.319	−0.04	0.484	0.39	0.652	0.82	0.794
−0.46	0.323	−0.03	0.488	0.40	0.656	0.83	0.797
−0.45	0.326	−0.02	0.492	0.41	0.659	0.84	0.800
−0.44	0.330	−0.01	0.496	0.42	0.663	0.85	0.802
−0.43	0.334	0.00	0.500	0.43	0.667	0.86	0.805
0.87	0.808	1.30	0.903	1.73	0.958	2.3	0.990
0.88	0.811	1.31	0.905	1.74	0.959	2.3	0.990
0.89	0.813	1.32	0.907	1.75	0.960	2.4	0.991
0.90	0.816	1.33	0.908	1.76	0.961	2.4	0.991
0.91	0.819	1.34	0.910	1.77	0.962	2.4	0.992
0.92	0.821	1.35	0.912	1.78	0.963	2.4	0.992

（续）

λ	$P(T_k)$	λ	$P(T_k)$	λ	$P(T_k)$	λ	$P(T_k)$
0.93	0.824	1.36	0.913	1.79	0.963	2.4	0.993
0.94	0.826	1.37	0.915	1.80	0.964	2.5	0.993
0.95	0.829	1.38	0.916	1.81	0.965	2.5	0.993
0.96	0.832	1.39	0.918	1.82	0.966	2.5	0.994
0.97	0.834	1.40	0.919	1.83	0.966	2.5	0.994
0.98	0.837	1.41	0.921	1.84	0.967	2.5	0.995
0.99	0.839	1.42	0.922	1.85	0.968	2.6	0.995
1.00	0.841	1.43	0.924	1.86	0.969	2.6	0.995
1.01	0.844	1.44	0.925	1.87	0.969	2.6	0.995
1.02	0.846	1.45	0.927	1.88	0.970	2.6	0.996
1.03	0.849	1.46	0.928	1.89	0.971	2.6	0.996
1.04	0.851	1.47	0.929	1.90	0.971	2.7	0.996
1.05	0.853	1.48	0.931	1.91	0.972	2.7	0.996
1.06	0.855	1.49	0.932	1.92	0.973	2.7	0.997
1.07	0.858	1.50	0.933	1.93	0.973	2.7	0.997
1.08	0.860	1.51	0.935	1.94	0.974	2.7	0.997
1.09	0.862	1.52	0.936	1.95	0.974	2.8	0.997
1.10	0.864	1.53	0.937	1.96	0.975	2.8	0.997
1.11	0.867	1.54	0.938	1.97	0.976	2.8	0.997
1.12	0.869	1.55	0.939	1.98	0.976	2.8	0.998
1.13	0.871	1.56	0.941	1.99	0.977	2.8	0.998
1.14	0.873	1.57	0.942	2.00	0.977	2.9	0.998
1.15	0.875	1.58	0.943	2.02	0.978	2.9	0.998
1.16	0.877	1.59	0.944	2.04	0.979	2.9	0.998
1.17	0.879	1.60	0.945	2.06	0.980	2.9	0.998
1.18	0.881	1.61	0.946	2.08	0.981	2.9	0.998
1.19	0.883	1.62	0.947	2.10	0.982	3.0	0.999
1.20	0.885	1.63	0.948	2.12	0.983	3.0	0.999
1.21	0.887	1.64	0.950	2.14	0.984	3.0	0.999
1.22	0.889	1.65	0.951	2.16	0.985	3.2	0.999
1.23	0.891	1.66	0.952	2.18	0.985	3.4	1.000
1.24	0.893	1.67	0.953	2.20	0.986	3.6	1.000
1.25	0.894	1.68	0.954	2.22	0.987	3.8	1.000
1.26	0.896	1.69	0.955	2.24	0.988	4.0	1.000
1.27	0.898	1.70	0.955	2.26	0.988	4.5	1.000
1.28	0.900	1.71	0.956	2.28	0.989	5.0	1.000
1.29	0.902	1.72	0.957	2.30	0.989		

11.5.2 根据作业的标准差确定关键线路的方法

由于工序的作业时间采用“三时估计法”，所以由开始事项到结束事项的所有线路完成时间都在$\sum_{i=1}^{k} t_{Ei}$的上下波动，这种波动可由标准差评价，标准差(均方差)越大，线路完成时间的波动就越大，因此关键线路应考虑均方差的大小来确定。

任一线路在概率为1的条件下完成任务的时间 T_L 可由下式确定：

$$T_L = T_E + K\sigma \tag{11.5.9}$$

式中 σ——均方差；$K=3.9$；

对于整个网络图，即对该项任务而言，关键线路应取决于 $\max\{T_L\}$。

以该关键线路为中心安排其他各项工作并随时进行调整，可得一较完善的计划。

11.5.3 根据各线路在指定日期内完成任务的概率确定关键线路

根据规定任务完成时间可按 $\lambda = \dfrac{T_k - T_{cp}}{\sum \sigma}$ 确定各条线路的完成概率，如果实现概率越小，则说明该线路越关键，从而可确定关键线路和总工期。该方法综合了上述两种方法的优点，是一种比较好的方法，但对大型网络有计算复杂的缺点。

现举例说明上述三种方法的区别。

根据图11.5.2所示的网络图及参数计算结果见表11.5.2。

表11.5.2 计算结果

线路	按 t_E	按 $T_k=52$ 天	$P=1$ 时完成任务工期
Ⅰ	48	0.966	57
Ⅱ	50	0.992	54
Ⅲ	46	0.977	58
关键线路	Ⅱ(50)	Ⅰ(0.966)	Ⅲ(58)

由表11.5.2可见，三种方法所得结果是不同的，这也正说明了PERT网络的主要目的是评价和审查工作的执行情况，而不特别注重确定最后完成任务的时间和关键线路。因此用PERT网络解决问题时，根据工作执行情况，随时分析和研究各工作是重要内容，计划评审的名称也来源于此。

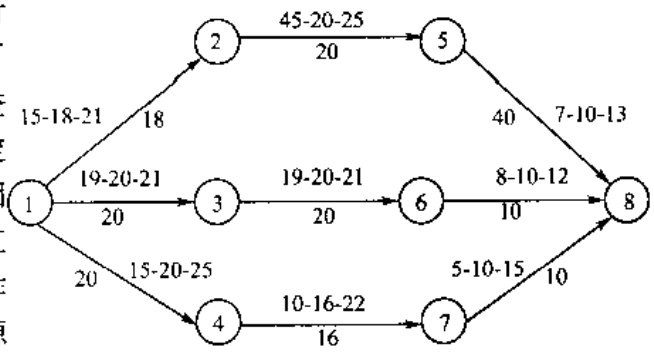


图11.5.2 网络图

11.6 决策关键线路法(DCPM)

CPM法和PERT法已广泛地用于各行各业，这两种方法主要着重于完成任务的时间计划。虽然可以通过时间—费用优化解决完成任务的最小费用和最小费用工期问题，但

主要解决在确定的工序下完成任务的时间—费用问题。

完成任何项任务都可采用不同的工艺方案，而不同的工艺方案都对应有不同的工期和费用。因此综观全局，在各种不同的工艺方案和费用组合中寻找最小费用工期，则是决策进行面临的一项主要任务。采用 CPM 法和 PERT 法虽可解决该问题，但其最大缺点是每一工艺方案都分别对应一套网络图，这对解决上述问题来说很不方便，有时甚至是不可能的。

决策关键线路法就是将 CPM 法与决策理论相结合，在同一张网络图上表示出各种不同工艺过程或方案，为决策提供决策信息的一种方法。

11.6.1 DCPM 法的基本原理

一般来说，对一项任务进行分析时，首先淘汰那些费用高而工期长的工艺方案，采用费用小而工期短的方案。因此计划采用的各种不同工艺方案都具有如下特点：

(1) 工期相对长，则费用相对低，为了缩短工期，就必须采用费用较高的工艺方案；

(2) 一项任务的各项工作中可改变工艺方案的工作是有限的。

鉴于上述特点可得如下结论：

(1) 采用工期较长的工艺方案完成任务，则费用必最低，采用工期较短的方案，则费用较高；各工作按不同工艺方案的组合，其工期必在最长和最短期间变动，费用相应在最低和最高之间变动，这种工期和费用之间的关系一般呈非线性，如图 11.6.1 中 p_{01} 所示。

(2) 虽然可改变工艺过程的工作数有限，但整个任务所组成的方案却很多，因此确定曲线 p_{01} 很困难。

(3) 在加强经济管理的今天，当超过规定的完成任务期限时，可规定罚款(罚款率为 p)，提前完成任务时，给予奖励(奖金率为 r)，由此可得奖惩曲线 p_2。决策应同时考虑不同方案的总费用($p_{01}+p_2$)，即以总费用 $p=p_{01}+p_2$ 来选择一套工艺方案和工期作为决策。

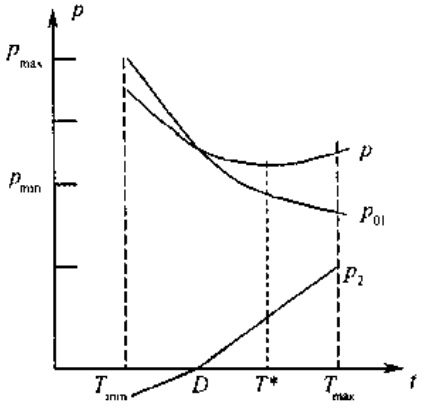


图 11.6.1 时间—费用关系

由图 11.6.1 可见必定存在一种工艺方案在工期 T^* 完成任务且费用 p 最小。

DCPM 法就是根据该原理，寻求该种确定最小费用、最小费用工期以及相应工艺方案的一种方法。

11.6.2 DCPM 网络图

DCPM 和 CPM、PERT 法一样，需把一项任务分解成若干工作，但该法与 CPM、PERT 法有以下不同：

(1) DCPM 法在工序分解和分析时，要考虑实现每一工作的各种不同工艺方案和各工作在不同工艺下的相互关系。

(2) DCPM 法除考虑各种不同工艺方案中各工作的时间参数外，尚需研究各工作所对应的费用。

(3) CPM 法只能采用单代号网络图，且节点分为“和”型节点(以○表示)和“或”型节

点(以△表示)。

"和"型节点表示"和"型工作,所谓"和"型工作就是在完成一项任务的过程中,无论选取那种工艺方案都必须完成的工作。

"或"型节点表示"或"型工作,所谓"或"型工作就是在完成一项任务的过程中,在某种工艺方案时执行其中某一项工作而不执行其他各工作,这类工作叫"或"型工作。

DCPM 网络规定,体现决策过程的"或"型节点以 $S_i=\{S_{i1},S_{i2},,\cdots,S_{in}\}$ 表示,其中 $S_{ij}(j=1,2,\cdots,n)$ 为决策工作,j 以工期由长到短,费用由低到高的顺序排列。

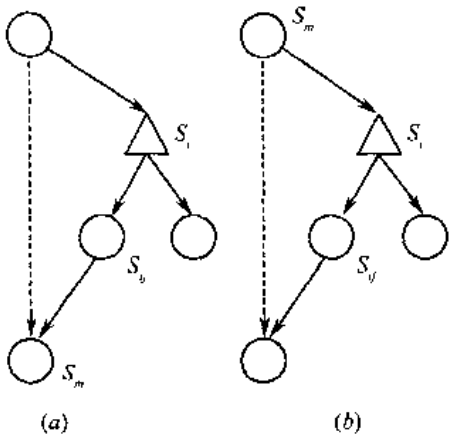


图 11.6.2 网络图

如果采用决策变量 d_{ij}:

$$d_{ij}=\begin{cases}1, S_{ij}\text{工作必须执行}\\0, S_{ij}\text{工作不执行}\end{cases}\qquad \text{且}\sum_j d_{ij}=1,(i=1,2,\cdots,m)$$

将 d_{ij} 乘以 S_{ij},则可将"或"型节点变为一"和"型节点,这样就相当于把一张 DCPM 网络图变成一张 CPM 网络图,改变 d_{ij} 即可得到不同工艺方案所对应的 *CPM* 网络图。

由于引入了决策变量,且当 $d_{ij}=0$ 的工作不起作用时,存在 S_{ij} 工作与其关联工作的接续问题。如图 11.6.2(a)中 S_m 为 S_{ij} 的紧后工作,如果 $d_{ij}=0$,S_{ij} 不执行,则 S_m 被断开,因此需将 S_m 与 S_{ij} 的紧前工作联接起来。同理图 11.6.2(b)中 S_m 为 S_{ij} 的紧前工作,如果 S_{ij} 当不执行时,必须把 S_m 与 S_{ij} 的紧后工序连接起来以保证网络的连续性。

(4) 在"和"型节点上方的左侧标该工作的工期,右侧标该工作的费用,在"或"型节点的上方左、右两侧面均标记"0",决策工作的时间和费用参数按"和"型工作标注。

11.6.3 DCPM 网络的算法

11.6.3.1 整数规划法

设 P_0 为"和"型工作费用,$P_0=\sum C_i$;

P_1 为"或"型工作费用,$P_1=\sum_i^L \sum_j C_{ij}d_{ij}$;

P_2 为奖罚费用,$P_2=\begin{cases}f(T), & \text{当 } T>D \text{ 时}\\0, & \text{当 } T=D \text{ 时}\\f_1(T), & \text{当 } T<D \text{ 时}\end{cases}$

整数规划模型为

目标函数: $P=P_0+P_1+P_2\rightarrow\min$

约束:

①"和"型工作

$$t_E(j)\geqslant t_E(i)+t_i \qquad (i \text{ 属 } j \text{ 工作的紧前工作})$$

②对"或"型工作

$$t_E(k)\geqslant(t_E(i)+t_{ij})\cdot d_{ij} \qquad (j=1,2,\cdots,n)$$

k 属“和”型工作集合；i 属“或”型工作集合；ij 属 k 的紧前工作集合。

③决策变量之和应等于 1

$$\sum_j d_{ij}=1 \qquad (i=1,2,\cdots,m)$$

11.6.3.2 试探法

由于 DCPM 网络中，“和”型节点在任一方案中都是必须执行的，因此其费用之和在整个决策中是固定不变的，现以 P_0 表示。而“或”型节点则随决策不同各工序执行的费用以及完成的时间也不同，因此“或”型节点费用之和 P_1 为任务工期 T 的函数，即 $P_1(T)$；完成任务的工程费用 P_{01} 亦为工期 T 的函数，故有

$$P_{01}=P_0+P_1(T)$$

因奖罚费用 P_2 也与完成任务时间有关，故 P_2 亦为工期 T 的函数，即 $P_2(T)$。

完成任务的总费用为

$$P(T)=P_0+P_1(T)+P_2(T)$$

该算法是从完成任务的最大工期 T_{max} 开始，以最小的费用增长量在关键线路上调整决策使完成任务的工期减少，从而得到 $P_{01}(T)$、$P_2(T)$ 和 $P(T)$，最后选择 $\min\{P(T)\}$ 的决策并确定网络参数。

其步骤为：

(1) 计算 P_0；

(2) 把所有“或”型节点放在工期最长、费用最小的决策上，计算 P_1；

(3) 求出关键线路和工期 T，计算 $P_2(T)$、$P(T)$；

(4) 若 $T=T_{min}$ 转(5)，否则转(6)；

(5) 选 $\min\{P(T)\}$ 得出最小费用工期及所对应的工艺方案，计算结束；

(6) 在所有满足 $\Delta T>0$ 的决策组合中，选使 P_1 增加最少的决策，转(3)。

其中 $\Delta T=T-T'$，T 与 T' 分别表示关键线路上每步压缩前后的工期。

由上述算法可知，该方法是从 T_{max} 开始到 T_{min} 为止，中间每步都在费用上升最小的前提下使总工期缩短的。

该算法的核心是第 6 步，即在关键线路上压缩工期使费用增加最少，要做到这点，必须考虑关键线路上的各种决策。现将关键线路上对“或”型节点的决策分为进压缩决策和退压缩决策两种。所谓进压缩决策是指决策过程中采用工期较短、费用较高的工作或方案达到压缩总工期的决策；所谓退压缩决策是指当某一“或”型节点采用工期较短、费用较高的工作或方案，而在另一些“或”型节点中采用较长工期、较低费用的工作或方案代替工期较短、费用较高的工作而压缩总工期的决策。这就是说，要达到压缩工期的目的，必须同时考虑进压缩和退压缩两种情况，而且不是以 ΔT 最小而以费用增加最少进行压缩。第(6)步的计算就是在所有进压缩和退压缩的组合中选择一个使费用 P_1 增加最小的方案来计算 $P_{01}(T)$ 和 $P(T)$ 的。因此 $P(T)$ 的计算是以 $P_1(T)$ 增加最少为目标，即

$$P(T)=P_0+\min\{P_1(T)+P_2(T)\}$$

在进行某次压缩时，如出现两条以上关键线路时，必须像消除负时差法一样，在各关键线路上同时压缩才能达到压缩总工期的目的，如果几条关键线路有公共工序，则要权衡压缩公共工序和分别压缩各条非公共工序所支出的费用，并从中选择使 P_1 增加最少的决

策。一般情况，优先考虑公共工序是有益的。

现举例说明 DCPM 法的计算过程和方法。

某任务网络图如图 11.6.3 所示。该任务规定 96 天完成，每延期一天罚款 50 元，每提前一天奖励 10 元，应如何安排该任务。

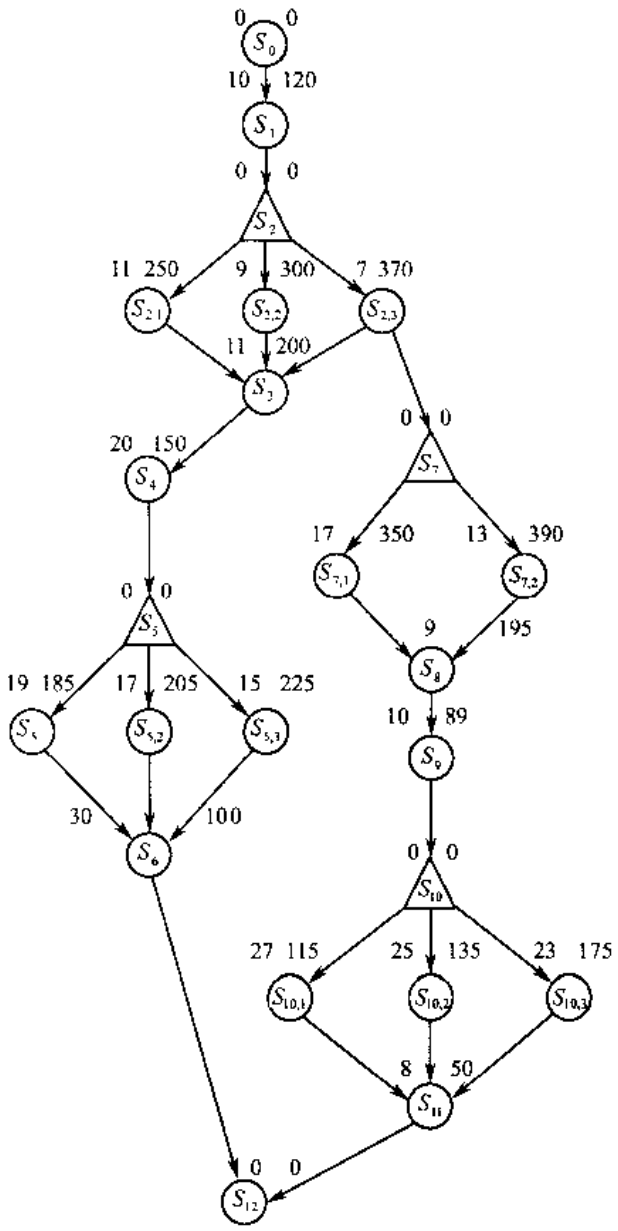


图 11.6.3 某任务网络图

(1) 计算 $P_0 = 904$(元)；$T_{max} = 103$(天)；$T_{min} = 93$(天)。

当 T_{max}时，关键线路为 S_0—S_1—S_2—S_7—S_8—S_9—S_{10}—S_{11}—S_{12}

总费用为 $P = p_0 + p_1 + p_2 = 904 + 900 + 300 = 2104$(元)

(2) 关键线路上“或”型节点为 S_2、S_7、S_{10}。

S_2 节点由 $S_{2,1} \rightarrow S_{2,2}$，费用增加 20 元，工期压缩 2 天；

S_7 节点由 $S_{7,1} \rightarrow S_{7,2}$，费用增加 40 元，工期压缩 4 天；

S_{10}节点由 $S_{10,1} \rightarrow S_{10,2}$，费用增加 20 元，工期压缩 2 天。

故选 $S_{10,2}$，作为决策。

(3) 计算关键线路，此时关键线路为 2 条，工期为 101 天，总费用为

$$904 + 920 + 250 = 2074(\text{元})$$

(4) 在两条关键线路上压缩工期，如果在公共工序上压缩，即 S_2 点由 $S_{21} \rightarrow S_{22}$，费用增加 50 元，总工期压缩二天，如在两条关键线路上分别压缩，则：一种方案为

S_5 节点由 $S_{51} \rightarrow S_{52}$，费用增加 20 元；工期压缩 2 天；

S_7 节点由 $S_{71} \rightarrow S_{72}$，费用增加 40 元；工期压缩 4 天；

S_{10}节点由 $S_{10,2} \rightarrow S_{10,1}$，费用减少 20 元；工期延长 2 天。

这种压缩是 S_7 节点的进压缩和 S_{10}点的退压缩与 S_5 的进压缩，使总工期由 101 天减少到 99 天。工序调整费用共增加 $20 + 40 - 20 = 40$ 元。

另一种方案为

S_5 节点由 $S_{51} \rightarrow S_{52}$，费用增加 20 元，工期压缩 2 天；

S_{10}节点由 $S_{10,2} \rightarrow S_{10,3}$，费用增加 40 元，工期压缩 2 天。

工期压缩二天，费用增加 60 元，故以费用最小选择决策，即 $P_1 = \{50, 40, 60\} = 40$(元)，

选择两条关键线路分别压缩的第一种方案作为决策。此时总工期 99 天,总费用为 2014 元。

(5) 按此方法继续进行即可得到工期由 $T_{max}=103$,到工期 $T_{min}=93$ 的各方案的总费用,从中选择最小费用所对应的工期即为最小费用工期。计算过程见表 11.6.1。

表 11.6.1 计算过程

步骤	决 策	P_0	P_1	P_2	P	T
0		904	900	350	2154	103
1	$S_{10,2}$	904	920	250	2074	101
2	$S_{52};S_{72};S_{10,1}$	904	960	150	2014	99
3	$S_{52};S_{10};S_{72}$	904	1000	50	1954	97
4	$S_{22};S_{53};S_{7,2};S_{10,2}$	904	1050	-10	1944	96
5	$S_{22};S_{53};S_{72};S_{10,2}$	904	1120	-30	1994	93

由计算可见,该网络通过五次计算即求出最小费用工期,而该问题实质共有 54 种方案,这种简化主要是由每次计算都是选择费用增长最少的工作,抛掉了一些较差方案,因此该方法是在优中选优。

这里需说明,DCPM 网络是以最小费用确定工期的,其决策结果可能出现超过规定工期的情况,这说明虽然超过工期需支付罚金,但从经济角度研究是合算的,至于是否采用该决策,决策者需从各方面考虑决定。

如图 11.6.4 所示的网络图,规定工期为 18 天,超过工期每天罚款 8 元,每提前一天得奖金 4 元,其最小费用工期的计算见表 11.6.2。

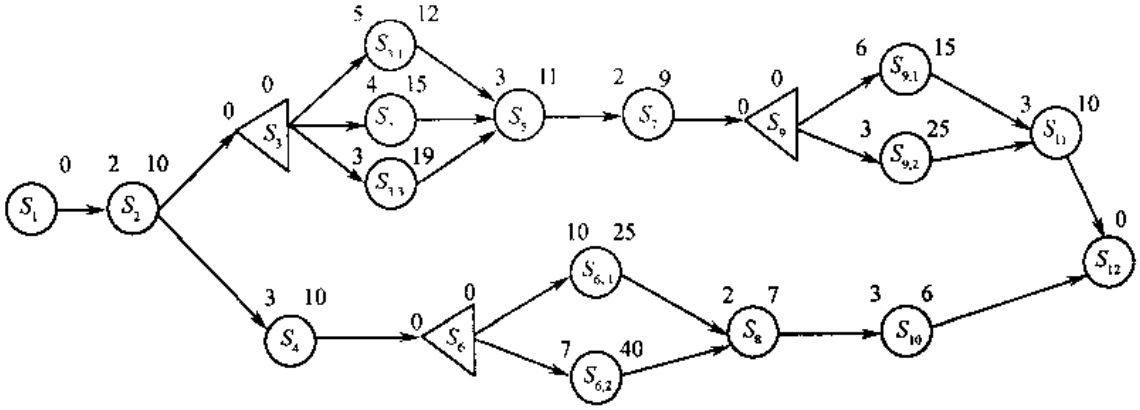


图 11.6.4 网络图

表 11.6.2 最小费用工期的计算

步骤	决 策	P_0	P_1	P_2	P	T
0	$S_{3,2};S_{6,1};S_{9,1};$	66	52	24	142	21
1	$S_{3,2};S_{9,1};S_{6,1};$	66	55	16	137	20
2	$S_{6,2};S_{3,3};S_{9,1};$	66	74	8	148	19
3	$S_{6,2};S_{3,1};S_{9,2};$	66	77	0	143	18
4	$S_{6,2};S_{3,3};S_{9,2};$	66	80	-4	142	17

可见最优决策为 $S_{3,2}S_{9,2}S_{6,1}$，总工期为 20 天，总费用为 137。从经济角度研究，完成该任务罚款是合算的。

11.7 图解评审技术(GERT)

GERT 是 1966 年在执行阿波罗登月计划时开发的一种新技术。它是综合网络理论、概率论和自动控制技术中的信号流图法而成的一种广义网络计划法。它不仅适用于工作执行时间是确定型的 CPM 网络，也适用于执行工作时间是随机型的 PERT 网络，除此而外，它还适用于执行的工作是随机型的情况，因此被称为广义网络计划法。从某种意义上讲，CPM、PERT 都可作为 GERT 的一种特例。因此它比 CPM、PERT 有更广泛的适用性。

11.7.1 GERT 网络图

在编制 GERT 网络图时，所进行的系统分解和分析工作与 CPM、PERT 基本相同。但因其可反映工序的随机性，故尚需收集各工作执行的概率 P_i 和反映其完成时间 t 分布情况的参数及分布函数。如对某工作，需收集的资料为：紧前工作、紧后工作、该工作发生的概率 $P_i=0.8$，执行时间 t 为服从参数 $\mu=5,\sigma=8$ 的正态分布等资料。

GERT 网络可根据系统输入和输出的各种情况，按表 11.7.1 所列的符号表示节点。

表 11.7.1 符号表

输出 \ 输入	异或	或	与
确定型			
随机型			

“异或”输入指通向某节点的任一支路实现，则该节点即可实现，但在给定时间内只能有一条支路实现。如某一时间内，工厂所需原料可能由船运，也可能由火车运，且只能有一种方式，该情况的网络图可用“异或”节点(见图 11.7.1)。

“或”输入是指通向该节点的任一支路实现，该节点则实现，但节点实现时间取决于通向该节点的各工作中时间最短的支路。如采用船和火车运输原料，不论是船还是火车，只要原料到达，即可开始生产，其网络图见图 11.7.2。

“与”输入是指通向该节点的所有支路实现，该节点才实现，节点实现的时间是指向该节点的所有工作中时间最长者。CPM、PERT 网络中的节点均系这类节点；确定型输出是指由该点出发的所有支路均以概率 $P=1$ 实现；随机型输出是指由该节点出发的支路各依概率 P_1 实现，实现的时间参数 t 服从某种分布。

GERT 网络规定工作用箭头线表示，箭头线上面标工作名称，下面标(P,t)，P 表示该工作实现的概率，t 为实现该工作的服从某种分布的时间参数。如图 11.7.3 所示的网络图，对于确定型输出的工序 $P_i=1,i=1,2,\cdots,n$；对于随机型输出，离开该节点的所有工作实现的概率之和为“1”，即

$$\sum_i P_i = 1$$

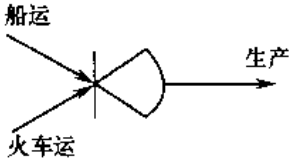


图 11.7.1 “异或”节点

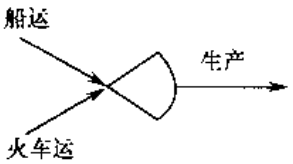


图 11.7.2 “或”输入

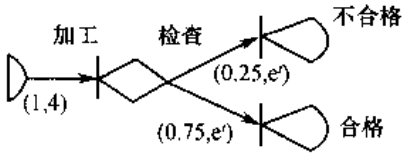


图 11.7.3 网络图

11.7.2 网络图的分析计算

11.7.2.1 信号流图及简化

1. 信号流图

信号流图是自动控制理论中用图论知识反映一组线性方程组以表示信息传递关系的一种方法，是一种表示线性方程组的图，它描述信息从一点向另一点的传递，并表明各信息之间的相互关系。

2. 信号流图的术语和基本概念

(1) 节点：用以表示信息或信息量的点。

(2) 传输：两节点间的增益。

(3) 支路：两节点间的定向线段。

(4) 输入节点：只有输出支路的节点。

(5) 输出节点：只有输入支路的节点。

(6) 混合节点：既有输入又有输出支路的节点。

(7) 通路：相同方向支路所构成的路径。

(8) 回路：起点和终点为同一点的通路。

(9) 不接触回路：无公共节点的回路。

(10) 前向通路：在由输入节点到输出节点的一条通路上，通过任何节点不多于一次的通路。

(11) 前向通路增益：前向通路中各支路传输之积。

3. 信号流图的性质

(1) 支路表示一个信号对另一信号的关系，信号只能沿着箭头方向通过。

(2) 节点可以把所有输入支路的信号叠加并把总的信号传递到所有输出支路。

(3) 各节点可用单位传输引出构成输出节点。

4. 信号流图与线性方程组

信号流图的一个节点表示一个变量或信号，而支路表示一个变量或信号对另一变量或信号的关系，因而，线性关系 $x_2 = ax_1$ 可用信号流图 $\bullet \xrightarrow[x_1]{a} \underset{x_2}{\bullet}$ 表示，其中 a 为传输或增益。这样即可把任一线性方程组用一信号流图表示。

例 1 方程组：

$$X_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + b_1u_1$$
$$X_2 = a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + b_2u_2$$

$$X_3 = a_{31}X_1 + a_{32}X_2 + a_{33}X_3 + b_3u_3$$

可用信号流图表示，如图 11.7.4 所示。

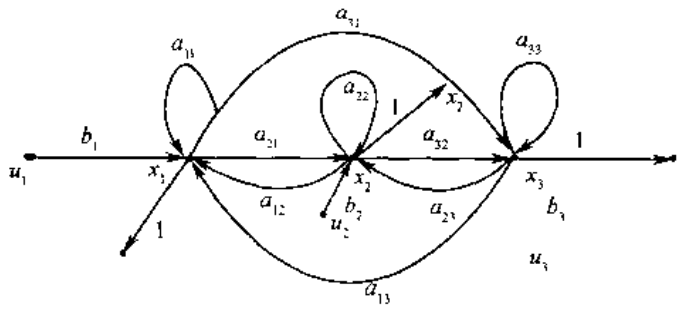


图 11.7.4　例 1 的信号流图

例 2　信号流图也可化成一方程组，如图 11.7.5 所示的信号流图可表示为一方程组：

$$x_2 = ax_1$$
$$x_3 = bx_2 + ex_4$$
$$x_4 = dx_3$$
$$x_5 = fx_2 + cx_3$$
$$x_6 = gx_5$$

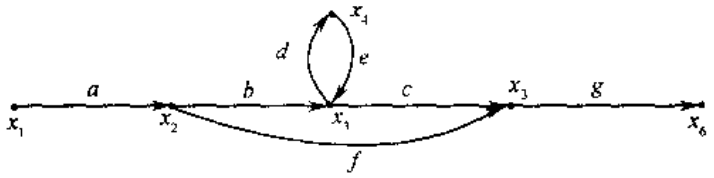


图 11.7.5　例 2 的信号流图

5. 信号流图的简化

根据信号流图的性质，可把复杂的信号流图化简而得到二节点间的总传输。

(1) 用串联支路的总传输等于各支路传输之积可消掉中间节点，如图 11.7.6(a)可表示为图 11.7.6(b)。

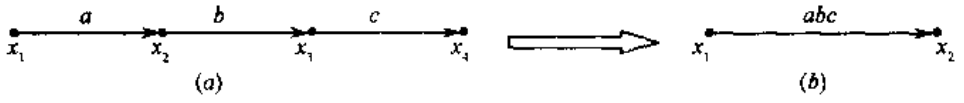


图 11.7.6　串联支路

(2) 用并联支路的总传输等于各支路传输之和可变并联支路为单一支路，如图 11.7.7(a)可表示为图 11.7.7(b)。

(3) 可消掉混合节点，如图 11.7.8(a)可表示为图 11.7.8(b)。

(4) 消掉回路，信号流图中的回路可通过消掉混合节点后消掉。如图 11.7.9(a)可通过消掉混合节点的方法变为图 11.7.9(b)的形式，而后化成图 11.7.9(c)的形式。

该结论可按如下方法导出：

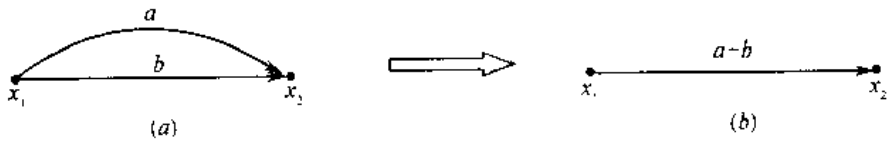


图 11.7.7　并联支路

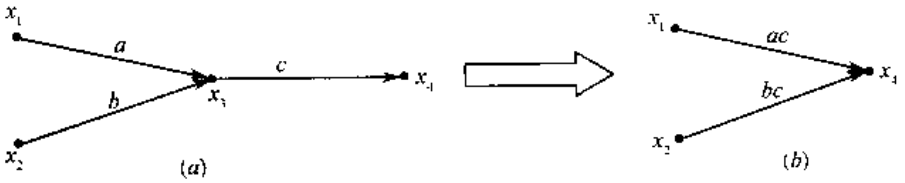


图 11.7.8　消掉混合节点

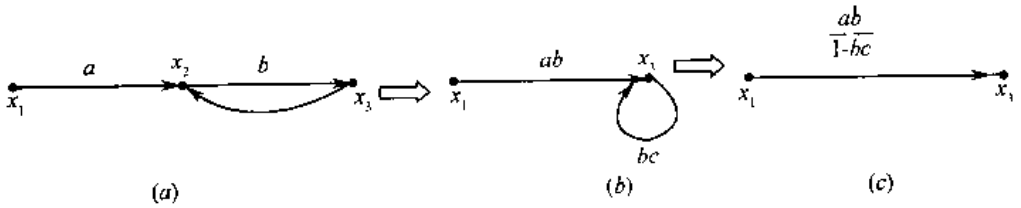


图 11.7.9　消掉回路

$\because x_3 = bx_2,\ x_2 = ax_1 + cx_3$

$\therefore x_3 = b(ax_1 + cx_3) = abx_1 + bcx_3$，得图 11.7.9(b)。

由上式移项得：

$x_3 - bcx_3 = abx_1, (1-bc)\ x_3 = abx_1, x_3 = \frac{ab}{1-bc}x_1$，得图 11.7.9(c)。

因此任一信号流图都可化为 $x_1 \xrightarrow{a} x_2$ 这种最简单的形式，而得到任两节点间的总传输。

例 2 中，如要确定 x_1 和 x_6 之间的关系可用上述方法化简(见图 11.7.10)。

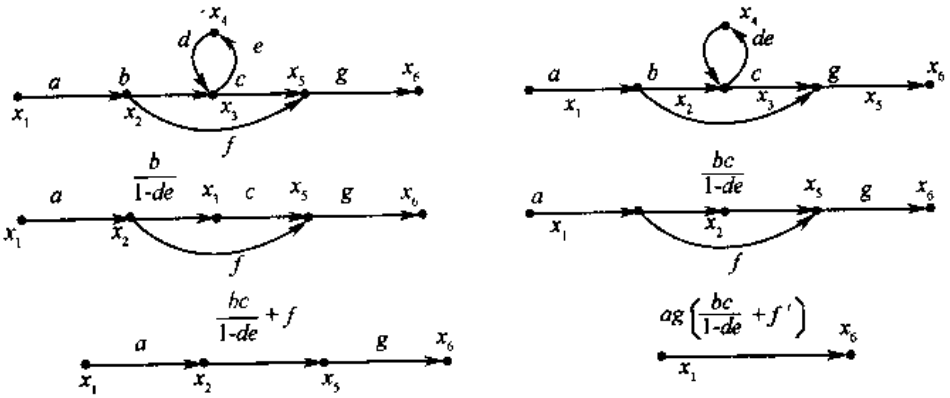


图 11.7.10　信号流图的化简

6. 梅森公式

除可对信号流图进行化简确定两节点之间的关系外，尚可通过计算确定两节点间的总传输 P，其计算公式为梅森公式：

$$P=\frac{1}{\Delta}\sum P_k\Delta_k$$

式中 P_k——第 k 条前向通路的增益；

Δ——流图特征式，为线性方程组系数行列式。$\Delta=1-$所有不同回路的增益之和+每两个不接触回路增益积之和+每三个不接触回路增益积之和+…，即：

$$\Delta=1-\sum_a L_a+\sum_{b,c}L_bL_c+\sum_{d,e,f}L_dL_eL_f+\cdots$$

Δ_k——除去与第 k 条前向通路相接触的回路的流图中第 k 条前向通路的特征式余子式。

P 的计算是在从输入节点到输出节点间全部可能的路上进行。

如上例中，前向通路有两条：$P_1=abcg$，$P_2=afg$

特征式：$\Delta=1-ed$，$\Delta_1=1$，$\Delta_2=1-ed$

$$\therefore P=\frac{1}{1-de}[abcg+(1-ed)abg]=\frac{abcg}{1-ed}+$$

$$\frac{afg}{1-ed}(1-ed)=a\cdot g\left(\frac{bc}{1-ed}+f\right)$$

与流图化简的结果相同。

11.7.2.2 GERT 网络图的分析和计算

在把一实际问题设计成 GERT 网络图之后，即可着手进行网络图的分析和计算。因研究对象是一随机问题，因此就不像 CPM 和 PERT 法那样把执行任务的工期、各工序的安排等作为研究的主要目的，而是把任务由开始到结束可能出现的结果及其概率作为研究的主要目的。

1. 具有“异或”输入节点的网络图

对有“异或”节点的网络图，各节点的输出有确定型和随机型两种，对于确定型输出，它有固定的工期 t 和实现的概率 $P=1$；对于随机型输出，则各工作的实现概率为 P，完成时间 t 则是某一概率密度 $f(t)$。因各工序的执行时间是一随机变量，各工序的叠加则是各随机变量求和或求积的运算，该和与积的密度函数的计算是相当困难的，因此我们引入矩母函数（或特征函数、拉普拉斯变换等）对各工序（含确定型输出）的执行时间 t 进行矩母函数变换：

$$M_t(s)=E(e^{st})$$

即 $$M_t(s)=\int_t e^{st}f(t)\mathrm{d}t \qquad (t\text{ 为连续变量})；$$

或 $$M_t(s)=\sum_t e^{st}f(t) \qquad (t\text{ 为离散变量})。$$

对确定型输出的工序，其执行时间为常量 $t=t_0$。因 t_0 为非随机变量，故 $\int_t f(t)\mathrm{d}t=1$，

$$M_{t_0}(s) = \int_t e^{st_0} f(t)\mathrm{d}t = e^{st_0}\int_t f(t)\mathrm{d}t = e^{st_0}$$

GERT 网络图中各工序有二个参数：P 和 $f(t)$，变换后则为 P 和 $M_t(s)$，现以 P 与 $M_t(s)$的积作为该工序的参数，叫等效函数。如以 $W_k = P_k M_{tk}(s)$表示第 k 工序的等效函数，则可利用信号流图化简的方法把任一有异或节点的网络图化简，而求出由输入节点到输出节点的总传输。

根据概率论的知识，如果输入到输出的串行各工序时间之和为：$T = \sum_i t_i$ 即总工期），当各工序执行时间是相互独立的随机变量时，总工期 T 亦是随机变量，其矩母函数为串行各工序执行时间 t_i 的矩母函数之积，即

$$M_T(s) = \prod_i M_{ti}(s)$$

所以
$$W_E(s) = PM_T(s) = \prod_k M_k(s) \cdot P_k = \prod_k W_k(s)$$

故
$$P = \frac{\prod_k W_k(s)}{M_T(s)} = \frac{W_E(s)}{M_T(s)}$$

由矩母函数性质知，当 $S=0$ 时，

$$M(s) = E\{e^{st}\}|_{s=0} = 1$$

是未实现变换以前的情况，所以

$$P = W_E(0)$$

从而得出由输入节点到输出节点的实现概率。

将 $P = W_E(0)$，代入 $P = \frac{W_E(s)}{M_T(s)}$，得矩母函数为

$$M_T(s) = \frac{W_E(s)}{W_E(0)}$$

由输入节点到输出节点所需的平均时间 $t = E(t_i)$可由矩母函数与矩的关系导出。该关系为：矩母函数 $M_T(s)$的一阶导数在 $S=0$ 时的值为一阶矩（原点矩）$E(t_i)$，所以，平均完成时间 t 为

$$E(t_i) = \left.\frac{dM_r(s)}{ds}\right|_{s=0}$$

现以某生产过程为例说明有异或节点的 GERT 网络的处理办法。

在一条生产线上，零件由生产线开端开始加工，假设加工过程为 4h，在进行下一步加工之前需对工件进行检查，其中 25% 的零件不合格，需返修，其他零件可输入下工序加工。设检验时间（含等待）服从参数为 $\mu=1$ 的负指数分布；返修加工需 3h，再检验后有 30% 的零件报废，70% 可输入下一工序。设返修检验时间服从参数为 $\mu=0.5$ 的负指数分布；在第二道工序中有 60% 的零件需加工 10h，40% 的工件需加工 14h，而后进行最终检验，每件需 1h，结果约 95% 合格入库，5% 不合格报废，试分析该生产线的生产情况。

第一步 依题意画 GERT 网络图（见图 11.7.11）。

第二步 把各工序的时间参数进行矩母函数变换。

负指数分布 $\mu e^{-\mu t}$的矩母函数为 $M(s) = \left(1 - \frac{s}{\mu}\right)^{-1}$，本例中变换结果见图 11.7.11。

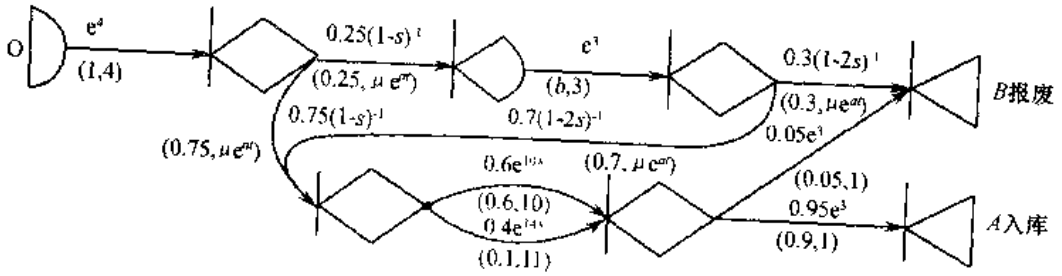


图 11.7.11 GERT 网络图

第三步 求等效函数。

$$W_E(s)=\prod_k W_k(s)=\prod_k M_{tk}(s)\cdot P_k$$

设 $W_{E(OA)}(s)$和 $W_{E(OB)}(s)$分别为零件由开始加工到成品和废品的总传输。采用化简流图计算总传输的方法，把 GERT 网络化成由开始加工到形成成品和由开始加工到产品废品的串联前向通路(见图 11.7.12)

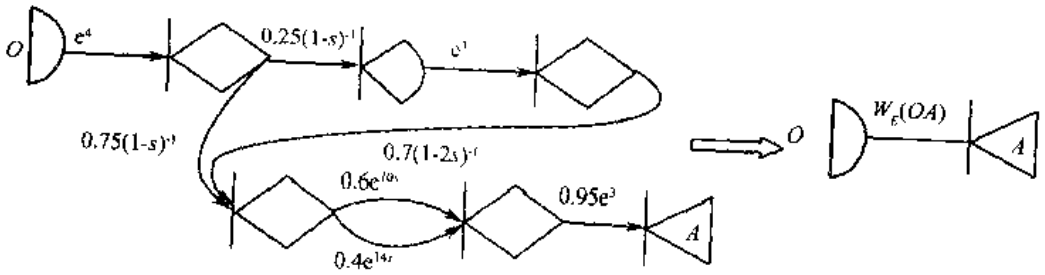


图 11.7.12 GERT 网络图的化简

$$W_{E(OA)}(s)=e^{4s}[0.25(1-s)^{-1}\cdot e^{3s}\cdot 0.7(1-2s)^{-1}+0.75(1-s)^{-1}]\cdot[e^{10s}\cdot 0.6+0.4\cdot e^{14s}][0.95e^{s}]$$

当 $s=0$ 时，$W_{E(OA)}(s)=0.87875$

因只有两种结果，故 $W_{E(OB)}(s)=0.12125$

由此步分析可知，该生产线的成品率约 87.875%，废品率约为 12.125%。

第四步 求平均时间 $t=E(t_i)$

因

$$M_T(s)=\frac{W_E(s)}{W_E(o)}=\frac{1}{0.87875}W_{E(OA)}(s)$$

故该生产线出成品的平均时间为

$$\left.\frac{dM_T(s)}{ds}\right|_{s=0}=18.55(\text{h})$$

同理亦可求出出废品的平均时间。

2. 具有“与”节点或“或”节点的网络

由于异或节点可进行严密的分析，因此在网络图的设计中尽可能把问题抽象成异或节点。对某些具有“与”节点或“或”节点的网络图通过适当地改造总可将其变换成异或节点。

(1) 对于只有一个输入边的节点，无论是“与”节点还是“或”节点，都与“异或”节点等效。

(2) 对于有二个以上输入的“与”节点的网络图(见图 11.7.13)可引入有确定型输出的 S 节点，用(S,1)和(S,2)两工作连接 1、2 两节点。该工作的参数为 $P=1, t=0$，则网络可

改成图 11.7.13(c)的形式,因节点③为"与"节点,故只有当 a、b 两工作同时实现时,节点③才能实现,因此节点③实现的可能性为 P_aP_b,节点③不实现的概率为 $1-P_aP_b$,节点③实现的时间为 $t_E=\max\{t_a,t_b\}$,如此即可将节点③分成二个"异或"节点 3 和$\bar{3}$,网络图即改成图 11.7.13(e)的形式。图中 t 由②、①两节点不指向节点③的支路的时间分布决定。

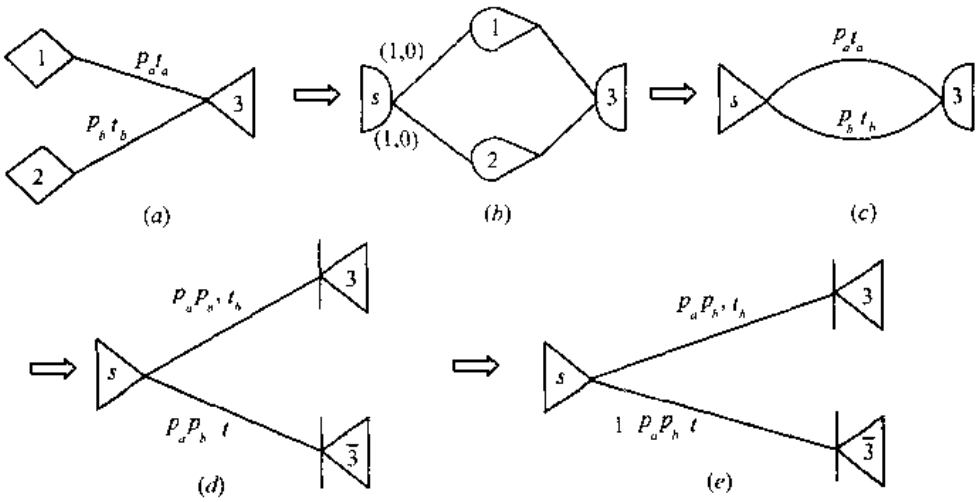


图 11.7.13 "与"节点的网络图

对有两个以上输入的或节点的网络图也可用同样的方法分析。

图 11.7.14(a)为有二个输入的或节点的网络图,对或型节点③,其实现的可能性有三种:(1)a 实现,b 不实现,其概率为 $p_a-p_{anb}=p_a-p_a\cdot p_b$($p_{anb}$为 a、b 同时实现的概率),实现时间为 t_a;(2)b 实现,a 不实现,其概率为 $p_b\cdots p_{bna}=p_b-p_ap_b$,实现时间为 t_b,(3)a、b 都实现,其概率为 $p_a\cdot p_b$,实现时间为 $t_{ab}=\min\{t_a,t_b\}$,节点③不实现的可能性为除去以上三种情况的值:$1-(p_a+p_b-p_a\cdot p_b)$,实现时间由由①、②节点不指向③节点的工作时间分布决定。由此图 11.7.14(a)所示的网络图,可改成图 11.7.14(b)的形式。

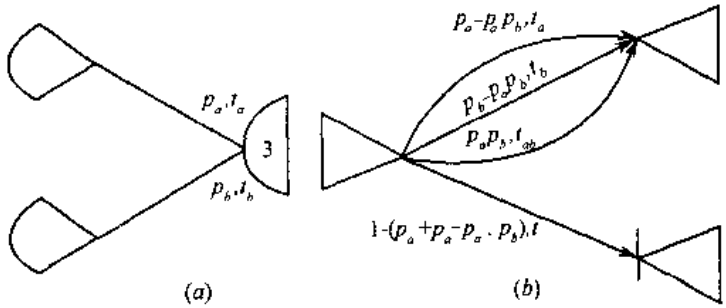


图 11.7.14 二个输入的"或"节点的网络图

对于复杂的网络图,这种变换是困难的,因此从网络图设计开始就尽量少用"或"和"与"节点为好。

3. 有费用指标为 GERT 网络图

有些网络图中工作的参数除概率和时间外还有实现该工作的费用,即有(P,t,c)三个参数。此时在网络图运算前可采用矩母函数 $M_{t,c}(s_1s_2)=e^{s_1t+s_2t}$进行变换得 $W_E(s_1,s_2)$,当 $s_1=s_2=0$ 时,$W(0,0)=P$,

$$\left.\frac{\partial M_t(s_1,0)}{\partial s_1}\right]_{s_1=0} = \bar{t};\ \left.\frac{\partial M_T(0,s_2)}{\partial s_2}\right]_{s_2=0} = \overline{C}$$

现举例说明有“异或”节点且有费用指标的 GERT 网络图的分析方法。

现准备进行土地资源调查，完成任务的方法有两种，一种是采用先进的遥感技术；另一种是传统方法，传统方法中又可采用小规模抽样、大规模抽样和其他方法三种方案。进行该工作的程序是进行各种方案的可行性研究，如否决方案，则写可行性报告；如采纳某种方案则需经实施、资料整理、精度分析工作；如精度分析符合要求则调查结束，如精度不符合要求，则调查失败。在安排工作前需对各种可能发生的情况进行参数估计并根据人力、物力、财力情况进行分析做出决策。

该问题的 GERT 网络分析方法的程序如下：

(1) 对工作流程进行系统分析，画出框图。

(2) 绘制网络图，并估计参数。

在研究方案时，有 70% 的人建议采用传统调查方案，30% 的人建议采用先进的调查方法。讨论需 3 天，费用 400 元。

在讨论传统方案时，共提出四种方法：

(1) 14% 的人建议采用普查的方法，研究 2 天，费用 300 元，得出结论为：人力、物力、财力不足，必导致失败；

(2) 24% 的人建议采用大规模抽样调查，讨论 1 天，费用为 150 元，得出结论为：调查需 50 天，费用为 10 万元；

(3) 24% 的人建议采用小规模抽样调查，讨论 1 天，费用为 150 元，得出结论为：调查需 30 天，费用为 5 万元；

(4) 38% 的人提出将(3)、(4)方案结合考虑，为保证精度，在大规模抽样的基础上，采用小规模抽样复查的办法，讨论 2 天，费用 300 元，得出结论为：调查需 60 天，费用为 13 万元。

在讨论采用先进技术的调查方案时，40% 的人认为费用太高，且技术复杂，不可行。60% 的人认为可行，讨论进行 1 天，支出费用 150 元，且得出需 10 天时间，费用需 15 万元，精度可完全满足要求。

对采用传统方案中的(2)、(3)两方法时，需进行精度检查，需时间 5 天，费用 1 万元，且 30% 的可能失败，70% 的可能成功。

根据上述讨论，得出系统分析图与网络图(见图 11.7.15 和图 11.7.16)。

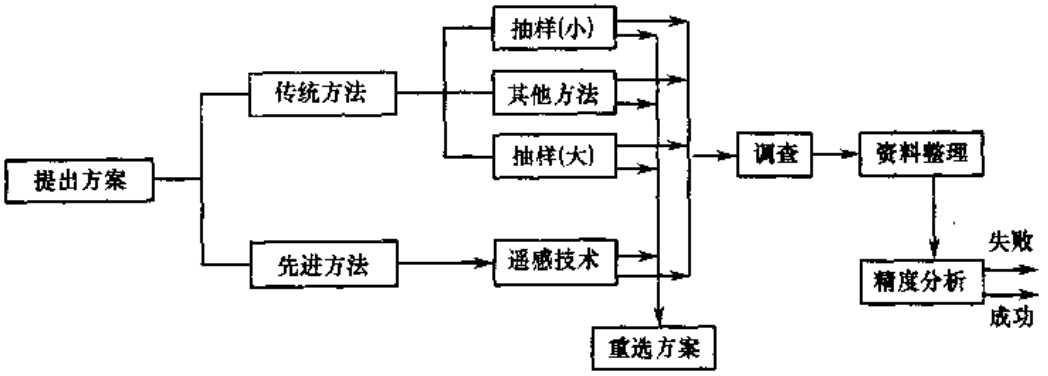


图 11.7.15 GERT 系统分析图

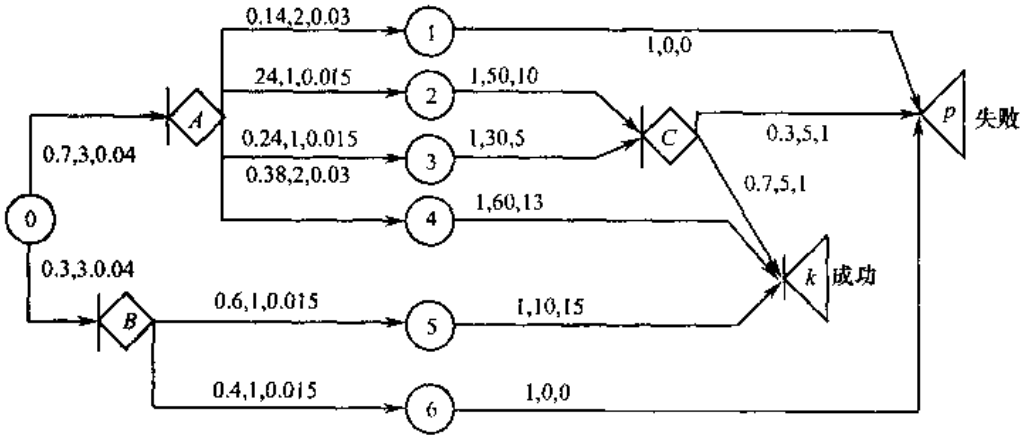


图 11.7.16 GERT 网络图

4、GERT 网络图计算

将上述网络图中参数按 $M_{t,s}(s_1,s_2)=e^{s_1+s_2}$ 化为矩母函数后，依网络图简化原理得出等效函数

$$W_{E(OK)}(s_1s_2)=0.7e^{3s_1+0.04s_2}[(0.24e^{51s_1+10.015s_2}+0.24e^{31s_1+5.015s_2})0.7e^{5s_1+s_2}$$
$$0.38e^{62s_1+13.03s_2}]+0.3e^{3s_1+0.04s_2}+0.6e^{s_1+0.015s_2}e^{10s_1+15s_2}$$

当 $s_1=s_2=0$ 时，

$$W_{E(OK)}(0,0)=0.7\times[(0.24+0.24)\times0.7+0.38]+0.3\times0.6=0.6812$$

同理 $$W_{E(OP)}(0,0)=0.3188$$

由上述计算可知，由讨论开始到工作结束，成功的可能性为 0.6812，失败的可能性为 0.3188。

对 $W_{E(OK)}(s_1,s_2)$ 进行整理得

$$W_{E(OK)}(s_1,s_2)=0.1176e^{56s_1+11.055s_2}+0.1176e^{39s_1+6.055s_2}+$$
$$0.3905e^{65s_1+13.07s_2}+0.2642e^{14s_1+15.055s_2}$$

故 $M_T(s_1,s_2)$ 为

$$M_T(s_1,s_2)=0.1726e^{59s_1+11.055s_2}+0.1726e^{39s_1+6.055s_2}+$$
$$0.3905e^{65s_1+13.07s_2}+0.2642e^{14s_1+15.055s_2}$$

平均完成时间 $\bar{t}$ 为

$$\bar{t}=\frac{\partial M_T(s_1,0)}{\partial s_1}\bigg|_{s_1=0}=0.1726\times59+0.1726\times39+0.3905\times65+$$
$$0.2642\times14=454961\approx46(\text{天})$$

完成任务的平均费用 $\bar{C}$ 为

$$\bar{C}=\frac{\partial M_T(0,s_2)}{\partial s_2}\bigg|_{s_2=0}=0.1726\times11.055+0.1726\times6.055+$$
$$0.3905\times13.07+0.2642\times15.055=12.035(\text{万元})$$

进一步可求出二阶矩(方差)，从而可确定出工期和费用的上下限，供决策参考。

第 12 章　随机服务系统

随机服务系统又叫排队论，它是运筹学的一个组成部分。它是研究排队现象和拥挤现象的一门科学，是研究系统随机聚集现象的理论。它是考虑随机因素进行排队系统的规划设计、试验研究和管理控制的科学方法，是系统工程不可缺少的基础理论之一。

在生产和日常生活中经常遇到排队现象。如上下班乘公共汽车，当汽车未到达时，乘客就排成一队候车是排队现象，到医院看病挂号也是排队现象。为了研究方便，我们将等待服务的对象称为“顾客”，而将提供服务的设备或人员称为“服务员”或“服务台”，并将“顾客”和“服务员”组成的系统称之为服务系统。服务系统中服务员和顾客可以同时是人，也可以同时是物，也可以一方是人，另一方是物。如医院中顾客是患者，服务员是医生；公共汽车是服务员，乘客是顾客等。

实践中，顾客什么时候来，按什么规律来，事先往往不能准确估计，而服务员进行服务的时间也随时机与条件的变化而变化，因而造成服务系统的状态是随机的，为了强调服务系统的随机性，一般称之为随机服务系统。

为了说明随机服务系统的研究内容，现举例说明如下。

有一个加油站，经统计，该加油站平均 20min 来一辆车加油。平均 15min 可为一辆车加完油，这就构成了一个以加油站为服务员、以前来加油的车辆为顾客的服务系统。

如果从上班开始，每隔 20min 来一辆车加油，15min 后加完油并离开加油站，则该加油站的工作是均衡的。这种服务系统是确定型服务系统。

然而实际生产中并非如此。从车辆到达方面研究，往往是上班后不久加油的车辆较多，下班前加油车辆较少，并不像确定型服务系统那样有秩序；从车辆加油时间方面研究，由于车辆大小和型号各不相同，所以加油时间也不相同，也不像确定型系统那样有秩序。这必然造成加油站有时排队、有时空闲。这就必须研究：

(1) 到达该系统的车辆排队的可能性有多大；

(2) 平均排队的车辆数是多少；

(3) 平均等待时间多少；

(4)加油站的运行状况，包括服务员的忙闲情况、改进系统的对策等。

如果排队的顾客过多，排队的时间过长，是不利的，增加服务员可减少排队顾客数和等待时间。但增加过多，又会造成服务设施的闲置和浪费，因此系统的管理者必须在二者之间进行平衡，做到既提高服务质量，又降低服务成本，取得最佳的效果。为此，必须研究和探索这类系统的运行规律，在理论上提出解决方法，为系统决策提供依据。

应用排队论的理论和方法研究上述例子，可得如下结论：该系统只有 25% 的可能不排队，75% 的可能需排队；一般情况下有二辆车在排队；一辆车由进入系统到离开系统大约需一小时。可见该加油站并不像确定型系统那样均衡地工作。上述数据可为系统决策

提供可靠的依据。

本章仅就随机服务系统的一般理论和某些简单的系统做一简要介绍，以说明处理随机服务系统的基本原理、方法和步骤。

12.1 随机服务系统的组成和特性

各类不同的服务系统，尽管形式和内容不同，然而都是由输入、排队和服务三部分组成的。现分别介绍如下。

12.1.1 输入过程及特性

输入过程是指各类不同的“顾客”按什么规律进入系统。这些“顾客”可能是待修的机器、医院的患者、打电话的用户、待装待卸的车辆等，它们都按一定的规律陆续到达系统，等待服务。

12.1.1.1 定长输入

顾客有规律地按一定的时间间隔到达系统，这种输入为定长输入。它是一种确定型输入，是随机服务系统的特例。如果系统每隔时间 a 到达一名顾客，则单位时间到达的顾客数为 $1/a$。正常生产的流水线均属此例。

12.1.1.2 普阿松输入

顾客到达系统的时间间隔服从普阿松分布的输入叫普阿松输入，又称简单流。

所谓简单流是指满足以下条件的输入流：

(1) 平稳性。在区间$[a,a+t]$内有 k 个顾客到达的概率与区间起点无关，只与 k 有关。通俗地说就是系统在任意的时间间隔 t 内到达 k 个顾客的概率是相等的，不因所取区间位置的不同而改变。

(2) 无后效性。不相交的区间内到达的顾客数是相互独立的，也就是在两个不同的区间内到达的顾客之间无联系。如理发馆中先到达的顾客和后到达的顾客是无关系的，它们都以相同的被服务者的身份出现。

(3) 普通性。令 $\psi(t)$表示长度为 t 的区间内至少到达两个顾客的概率，则当 $t=0$ 时，$\psi(t)=0$，也就是在某一时刻不能同时到达两个顾客。

(4) 有限性。在任意有限的区间内到达有限个顾客的概率为“1”。用 $U_k(t)$表示简单流在长度为 t 的时间内顾客到达的概率，则

$$\sum_{k=0}^{\infty} U_k(t)=1 \qquad U_0(t)\neq 1$$

其涵义为不能恒无顾客到达。

凡符合上述四个条件的输入流叫简单流，也叫普阿松流。它是一种用途较广、最容易处理的输入流，很多问题都可抽象为这种形式的输入。如修理工和机器所组成的系统，出故障的机器就可以认为是按简单流出现的。概率论提供了成熟的理论，可作为建立普阿松输入系统数学模型的准备。

对于简单流，长度为 t 的时间间隔内到达 k 个顾客的概率 $U_k(t)$服从普阿松分布，其密度函数为

$$U_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}\quad(k=1,2,\cdots)\tag{12.1.1}$$

式中 λ——大于零的常数。

因为普阿松分布是描述离散变量的，而对于简单流输入(普阿松输入)，顾客的到达也是离散的，因此，这种分布能够描述问题的一般性质。

由式(12.1.1)可见，给定时间和顾客数可得到该时间内到达顾客数的概率。

如在 $t=1,k=3$ 时，$U_3(1)=e^{-\lambda}\frac{(\lambda)^3}{3!}=\frac{\lambda^3}{3!}e^{-\lambda}$

大于零的常数 λ 代表单位时间内到达的顾客数量的平均值，也叫输入强度。它是由随机变量 t 的数学期望得到的，是在$[0,t]$时间间隔内到达 k 个顾客的加权平均值：

$$E(t)=\sum_{k=1}^{\infty}k\cdot U_k(t)=\sum_{k=1}^{\infty}k\cdot\frac{(\lambda t)^k}{k!}e^{-\lambda t}=$$

$$\sum_{k=1}^{\infty}\frac{(\lambda t)^k}{(k-1)!}\cdot e^{-\lambda t}=\lambda t e^{-\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda t)^{k-1}}{(k-1)!}$$

令 $k-1=\omega$，则

$$E(t)=\lambda t e^{-\lambda t}\sum_{\omega=0}^{\infty}\frac{(\lambda_t)^{\omega}}{\omega!}$$

$\because$ $$e^{a}=\sum_{i=0}^{\infty}\frac{a^i}{i!}$$

$\therefore$ $$E(t)=\lambda t e^{-\lambda t}\cdot e^{\lambda t}=\lambda t$$

当 $t=1$ 时，$E(t)=\lambda$，因此它表示了单位时间内到达顾客数的平均值。

普阿松输入的顾客到达是离散的，但时间却是连续的，也就是时间间隔 $T_i=t_it_{i-1}(i=1,2,3,\cdots)$服从负指数分布，它的概率密度为 $\lambda e^{-\lambda t}$，其数学期望：

$$E(t)=\int_0^{\infty}t\lambda e^{-\lambda t}dt=\int_0^{\infty}-t de^{-\lambda t}=\frac{1}{\lambda}$$

所以在简单流输入时，顾客到达的平均间隔长度为 $1/\lambda$，即每到达一个顾客的平均时间为 $1/\lambda$。

12.1.1.3 爱尔朗输入

这种输入是在 t 时间内到达 k 个顾客的概率服从爱尔朗分布，其概率密度为

$$P_k(t)=\frac{\mu k(\mu k t)^{k-1}}{(k-1)!}e^{\mu k t}\qquad t>0\tag{12.1.2}$$

其数学期望为

$$E(t)=\frac{1}{\mu}$$

爱尔朗分布还存在于这样一类排队系统中：有 k 个串联的服务员，每个服务员的服务时间相互独立，服从相同的负指数分布(参数为 μk)，当一个顾客依次通过 k 个服务员总共需要的服务时间就服从 K 阶爱尔朗分布。当 $k=1$ 时，它即成为负指数输入。$k=\infty$ 时，即为确定型输入。

12.1.1.4 一般输入

顾客在时间 t 内到达的概率服从任意的分布 $A(t)$($A(t)$为任意函数)，这种输入叫**一般输入**。

上述四种输入中,前三种输入都第四种输入的特例。

12.1.1.5 成批到达输入

这种输入是到达的所有顾客以上述各种形式输入,不同的只是每次到达不像简单流那样具有普通性,而是到达一批顾客,而顾客数又是随机的。

12.1.2 排队规则及特性

排队规则是指到达的顾客按怎样规定的次序等候服务。如上汽车按到达的先后次序上车,出库的物资按先到先出或后到先出的顺序出库等,各种不同类型的服务系统都有自己规定的次序。

12.1.2.1 损失制

当顾客到达时,所有服务台均被占用,则该顾客就离去,永不再来。如电话系统就属于损失制,当一次呼叫不通,则此次呼叫作废,这次呼叫则永远消失(注意:如果再次呼叫被认为与上次呼叫无关)。

12.1.2.2 排队制(等待制)

当顾客到达时,所有服务台均被占用,他们就排队等待服务。其排队方式有以下几种:

1. 单服务台

(1) 先到先服务。按到达的时间顺序进行排队,这种排队是通常遇到的大多数情况;

(2) 后到先服务。即按到达的时间顺序出后向前服务。如仓库内后到的货物放在最上而,必须先付出;通讯卫星最后收到的信息往往最有用,必须先处理等;

(3) 随机服务。当服务台有空时,在顾客中按随机原则挑选一个顾客服务;

(4) 优选权服务。当服务员有空时,优先对某个顾客服务或服务员中断服务而让位于有优先权的顾客。如医院中的急诊、重要物资的装卸、重点设备的维修都要优先于一般患者、一般物资和一般设备。

2. 多服务台的服务

当系统有 n 个服务台时,排队形式可有多种。常见的是在每个服务台前排成一队或排成公共一队。当服务台有空时,按顺序进行服务。

12.1.2.3 混合制

(1) 排队长度(队长)有限。当顾客到达时,若队长已等于规定长度时,顾客离去;若小于规定长度时,则排队。系统不存在超过队长的状态。如医院挂号已满,就不再排队了。

(2)等待时间有限。顾客在队中排队超过时间 t 时,则离去。

12.1.3 服务机构及特性

服务机构是指同一时刻有多少服务设备可供服务,即可同时接待几位顾客。如:医院有几位医生看病;几个修理工去处理出故障的设备等都构成了服务机构。医生处理病人的时间、修理机器的时间等就是服务机构对顾客的服务时间。

12.1.3.1 定长服务

每个顾客的服务时间一定,且为 a,此时服务时间的密度函数是:

$$A(t)=\begin{cases}1, t\geqslant a\\0, t<a\end{cases} \tag{12.1.3}$$

如流水线上相同的工件,加工时间是相同的,即属定长服务。

12.1.3.2 负指数服务

每个顾客的服务时间相互独立,具有相同的负指数分布,其密度函数为

$$A(t)=\begin{cases}\mu e^{\mu t}, t\geqslant 0\\0, \quad t<0\end{cases} \tag{12.1.4}$$

式中 μ——一个大于零的常数,为单位时间服务的顾客数,μ^{-1}为平均服务时间。

12.1.3.3 爱尔朗服务

每个顾客的服务时间具有相同的爱尔朗分布,其密度为

$$A(t)=\frac{k\mu(k\mu t)^{k-1}}{(k-1)!}e^{k\mu t} \qquad t\geqslant 0, \mu>0 \tag{12.1.5}$$

当 $k=1$ 时,则为负指数服务,当 $k=\infty$ 时,为定长服务。μ 为单位时间服务的顾客数,μ^{-1}为平均服务时间。

12.1.3.4 一般分布

所有顾客服务时间是相互独立的,其密度函数 $A(t)$ 是 t 的任意函数。

此外还有多服务台共同服务、服务时间依赖于排队长度等服务模式。

对于一个随机服务系统,要评价它的效益,必须拟定评价指标。对于不同的系统,可拟定不同的评价指标,从不同的角度去评价。因此,必须根据系统的需要建立评价指标体系。

评价随机服务系统的指标有很多,但主要有:

(1) 等待时间,即顾客由到达系统开始到服务结束为止所需要的时间。这是顾客最关心的时间。

(2) 忙期,即服务台连续工作的时间,这是服务员最关心的指标。

(3) 队长,即在系统中顾客的数目。这是顾客、服务员和系统设计人员都非常关心的指标,尤其是系统设计和管理人员可利用该指标选择各种参数和进行决策。

12.1.4 服务系统分类的表示法

为了一目了然地分清服务系统的种类,采用一定的分类标记是必要的。按肯德尔的三要素标记法,随机服务系统的分类记号以"□/□/□"表示,其中第一位表示到达规律,即输入类型;第二位表示服务类型;第三位表示服务台个数,如图 12.1.1 所示。

前两位由以下符号表示:

D:表示定长输入和定长服务;

M:表示普阿松输入和负指数服务;

G:表示一般输入和一般服务;

E_k:表示爱尔朗输入和爱尔朗服务。

服务台个数以数字表示。

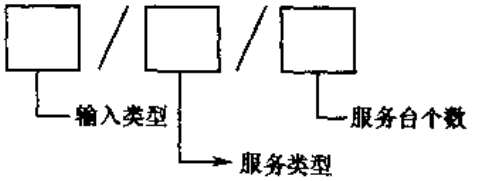


图 12.1.1 服务系统表示法

如 M/M/3,表示普阿松输入、负指数服务、三个服务台的随机服务系统。M/G/1 表示普阿松输入、一般服务、一个服务台的随机服务

系统。

在 1971 年一次关于排队论符号标准化会议上决定，将 Kendall 符号扩充成为

X/Y/Z/A/B/C

形式，其中前三项意义不变，而 A 处填写系统容量限制 N，B 处填写顾客源数目 m，C 处填写服务规则，如先到先服务 FCFS，后到后服务 LCFS 等。并约定如略去后三项，即指 X/Y/Z/∞/∞/FCFS 的情形。

12.2 定长服务系统

定长服务系统是确定型的服务系统，是随机服务系统的一个特例。

如果一个服务台，顾客到达时间和服务时间都是定长的，则这类系统叫定长或确定型服务系统。根据顾客到达时间和服务时间的不同，共有三种情况：

(1) 顾客到达时间大于服务时间，服务台有空闲。

(2) 顾客到达时间等于服务时间，服务台没有空闲。

(3) 顾客到达时间小于服务时间，服务台没有空闲，且无法完成任务，排队长度将无限地增加。

这三种情况都是明显的，这里我们主要研究当顾客到达时间大于服务时间，而且在预先已有顾客排队的情况下，何时才能消除排队。

现举例说明这类系统的处理方法。

生产线的某工位，每隔 30min 到达一个部件，加工该部件需 20min，且在上班前已有 3 个部件等候加工，在上班后多长时间才能消除排队现象。

显然，上班后要先加工已排队的部件，而在加工它们的过程中，还继续到达部件，因此排队要继续下去。由于顾客到来的时间大于服务时间，终有一个部件到来时不需排队，而从该部件到达以后再到达的部件就不需排队了。现用图 12.2.1 来分析该系统。

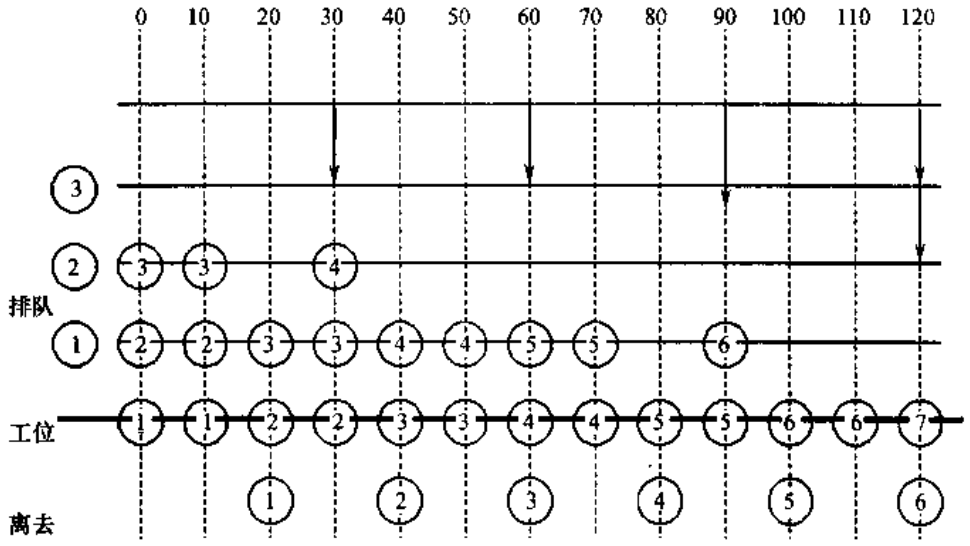


图 12.2.1 定长排队系统

由图 12.2.1 可见，当第七个部件到达时，可不必排队而直接加工，我们称这以后的状态为稳定状态。

从该例的分析可得出服务系统消除预先排队状况的一般规律。

如果定长服务系统预先有 i 个顾客排队，顾客到达时间间隔为 a，服务时间为 b，且 $a>b$，经过 T 时间后可消除排队，A 为在 T 时间内到达的新顾客数，则

$$T=(A+i)b \tag{12.2.1}$$

$$T\leqslant(A+1)a \tag{12.2.2}$$

式(12.2.1)为消除排队的服务时间，式(12.2.2)为消除排队的到达时间，也就是超过以后再到达的顾客就不再排队了。

解此不等式组得：

$$(A+i)b\leqslant(A+1)a,\ A\geqslant\frac{ib-a}{a-b},$$

因为到达的顾客数 A 必须是整数，故

$$A(t)=\begin{cases}\dfrac{ib-a}{a-b} & \text{当}\dfrac{ib-a}{a-b}\text{为整数时}\\ \left[\dfrac{ib-a}{a-b}\right]+1 & \text{当}\dfrac{ib-a}{a-b}\text{为非整数时}\\ 0 & \text{当 } ib<a \text{ 时}\end{cases} \tag{12.2.3}$$

这样由 A 即可确定出消除排队的时间。如上例中 $a=30,b=20,i=3$，

$$A=\frac{3\times20-30}{30-20}=3$$

$T=(3+3)\times20=120$，即 120min 后消除排队。

计算结果与图 12.2.1 的结果是一致的。

12.3 生 灭 过 程

12.3.1 生灭过程定义

丹麦数学家爱尔朗把封闭系统热分子的扩散比作电话系统电话呼叫的生灭。一次呼叫，相当于热分子透进封闭系统，而一次呼叫的结束，则好像热分子从封闭系统扩散出来。这样他把热力学系统统计平衡模型借用过来而建立了电话呼叫的统计平衡模型。根据这种统计平衡的假设，呼叫的“生”和“灭”，使系统的状态改变，从而得出下述的生灭过程。它是处理随机服务系统的基础。

生灭过程即假设一个系统有有限个状态 $0,1,\cdots,k$ 或可数个状态 $0,1,\cdots,N(t)$ 为系统所处的状态，在任一时刻 t，系统处于 i 状态，则在 $(t,t+\Delta t)$ 内，系统由 i 状态转移到 $i+1$ 状态的概率为 $\lambda_i\Delta t+O(\Delta t)$（$\lambda_i$ 是大于零的常数），而由 i 状态转移到 $i-1$ 状态的概率为 $\mu_i\Delta t+O(\Delta t)$（$\mu_i$ 是大于零的常数），并在 $(t,t+\Delta t)$ 内，发生两次或两次以上转移的概率为 $O(\Delta t)$，这样一个系统随时间变化的过程 $N(t)$，就是生灭过程，图 12.3.1 即是生灭过程状态转移的示意图。

一个服务系统，顾客的到来和离开都会引起系统状态的改变，因此，顾客到来和服务

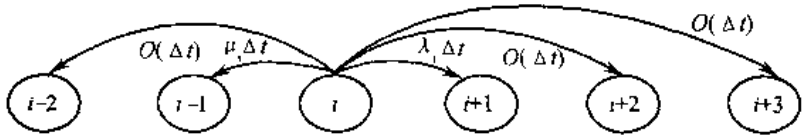


图 12.3.1 生灭过程

的过程就是一个系统状态变化的生灭过程。

这里的 λ_i 是输入强度，即是单位时间到达的顾客数，μ_i 是服务率，是单位时间的服务人数。

12.3.2 生灭过程微分方程

生灭过程的状态转移图(见图 12.3.1)只能定性地分析一个系统的变化，只有进行定量描述才能更深刻地解释和处理服务系统，才能揭示出服务系统排队的本质，才能为搞好系统的规划设计、管理和控制提供可靠的依据。

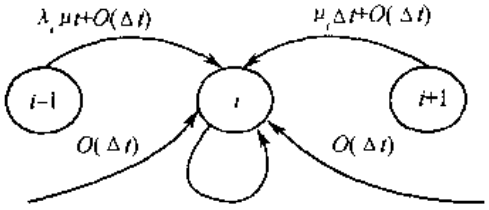


图 12.3.2

以 $p_i(t)$ $(i=1,2,\cdots)$为系统在 t 时刻处于 i 状态的概率，现研究系统在 $t+\Delta t$ 时刻处在 i 状态的概率。

由图 12.3.2 可见：

(1) 系统在 t 时刻处在 i 状态的概率为 $p_i(t)$，在 $t+\Delta t$ 时刻仍处在 i 状态的概率为 $p'_i(t+\Delta t)$，则 $p'_i(t+\Delta t)$是系统在 $t+\Delta t$ 时刻不转移到 $i+1$ 状态的概率 $1-(\lambda_i\Delta t+O(\Delta t))$、不转移到 $i-1$ 状态的概率 $1-(\mu_i\Delta t+O(\Delta t))$和系统在 t 时刻处在 i 状态的概率 $p_i(t)$的积决定的(是由该三个独立事件同时发生决定的)。用公式表示为

$$p'_i(t+\Delta t)=p_i(t)\cdot[1-(\mu_i\Delta t+O(\Delta t))]\cdot[1-(\lambda_i\Delta t+O(\Delta t))]$$

略去无穷小项得

$$p'_i(t+\Delta t)=p_i(t)\cdot[1-(\mu_i\Delta t-\lambda_i\Delta t)]$$

(2) 系统在 t 时刻处在 $i-1$ 状态的概率是 $P_{i-1}(t)$，在 $t+\Delta t$ 时刻转移到 i 状态的概率为 $p''_i(t+\Delta t)$，是由系统在 t 时刻处在 $i-1$ 状态的概率 $P_{i-1}(t)$和在 $t+\Delta t$ 时刻转移到 i 状态的概率 $\lambda_{i-1}\Delta t+O(\Delta t)$及不转移到其他状态的概率 $1-(\mu_{i-1}\Delta t+O(\Delta t))$的积决定的。即

$$p''_i(t+\Delta t)\lambda_{i-1}=p_{i-1}(t)\cdot[1-(\mu_{-1i}\Delta t+O(\Delta t))]\cdot[\Delta t+O(\Delta t)]$$

略去无穷小项得

$$p''_i(t+\Delta t)=p_{i-1}(t)\lambda_{i-1}\Delta t$$

(3) 系统在 t 时刻处在 $i+1$ 状态的概率为 $P_{i+1}(t)$，在时 $t+\Delta t$ 时刻转移到 i 状态的概率为 $p'''_i(t+\Delta t)$，同理可得：

$$p'''_i(t+\Delta t)=p_{i-1}(t)\mu_{i+1}\Delta t$$

(4) 由生灭过程定义可知,由所有其他状态转来的概率为无穷小,故忽略不计。

由全概率公式知,由此四种状态构成的 i 状态的概率为

$$p_i(t+\Delta t)=p'_i(t+\Delta t)+p''_i(t+\Delta t)+p'''_i(t+\Delta t)$$

$$p_i(t+\Delta t)=p_i(t)(1-\lambda_i\Delta t-\mu_i\Delta t)+p_{i-1}(t)\lambda_{i-1}\Delta t+p_{i+1}(t)\mu_{i+1}\Delta t$$

移项后同除 Δt 以得

$$p=\frac{(t+\Delta t)-p_i(t)}{\Delta t}=p_i(t)(\lambda_i+\mu_i)+p_{i-1}(t)\lambda_{i-1}+p_{i-1}(t)\mu_{i+1}$$

令 $\Delta t\to 0$,则得

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t}=p_{i-1}(t)\lambda_{i-1}-p_i(t)(\lambda_i+\mu_i)+p_{i+1}(t)\mu_{i+1}\quad(i=0,1,\cdots)\tag{12.3.1.a}$$

在有限情况下:

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t}=-\lambda_0 p_0(t)+p_1(t)\mu_1\quad(i=0)\tag{12.3.1.b}$$

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t}=-\lambda_{k-1}p_{k-1}(t)+p_k(t)\mu_k\quad(i=k)\tag{12.3.1.c}$$

式(12.3.1.b)、式(12.3.1.c)是因为它们处在首尾两个状态,再向前或向后转移的可能性为零的缘故。

如果当系统有可数个状态,即 $i=\infty$ 时,则式(12.3.1.c)就不存在了。

式(12.3.1)的三个方程就叫做生灭过程的微分方程,也叫状态方程,它反映了系统的动态特性,是生灭过程的定量数学描述,是研究随机服务系统的一个很重要的方程。

12.3.3 生灭过程稳态方程

生灭过程 $N(t)$ 在某状态的概率是时间的函数,但其随着时间的增大而趋于稳定,即 $\lim\limits_{t\to\infty}p_i(t)=p_j>0$,这就是生灭过程的极限定理。

在实际应用中,当极限 $\lim\limits_{t\to\infty}p_j(t)$ 存在的时候,常常把它当作任一时刻系统处于状态 j 的概率。因此,应用生灭过程的稳态形式来考察生灭过程,则有 $\frac{\mathrm{d}p_i(t)}{\mathrm{d}t}=0$,从而式(12.3.1)可写成

$$\begin{cases}\lambda_{i-1}p_{i-1}+\mu_{i-1}p_{i-1}-(\lambda_i+\mu_i)p_i=0 & (0<i<k)\quad(12.3.2.a)\\ \mu_1 p_1-\lambda_0 p_0=0 & (i=0)\quad(12.3.2.b)\\ \lambda_{k-1}p_{k-1}-\lambda_k p_k=0 & (i=k)\quad(12.3.2.c)\\ p_0+p_1+\cdots+p_k=1 & (12.3.2.d)\end{cases}$$

此方程组称为生灭过程的稳态方程。其中,式(12.3.2.d)是全部状态的概率和。之所以引入该方程是因为前三个方程所组成的方程组是线性相关的。求解此方程组即可得系统处于各种状态的概率。

因为 $i=k$ 是可数状态的特殊形式,因此可数状态 $i=1,2,\cdots,\infty$,除包括 $i=1,2,\cdots,k$ 以外,还要包括一些项,即

$$\begin{cases} \lambda_{i-1}p_{i-1} + \mu_{i-1}p_{i-1} - (\lambda_i + \mu_i)p_i = 0 & (i > 0) \\ \mu_1 p_1 - \lambda_0 p_0 = 0 & (i = 0) \\ p_0 + p_1 + \cdots + p_k + \cdots = 1 \end{cases} \quad (12.3.3)$$

解此方程组，很明显，当 $i=0$ 时，由式(12.3.2.b)知

$$p_1 = \frac{\lambda_0}{\mu_1} p_0$$

当 $i=1$ 时，由式(12.3.2.a)知

$$p_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0$$

依此类推，可得到

$$p_3 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} p_0$$

$$p_k = \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k} p_0 \quad (12.3.4)$$

$\because$ $p_0 + p_1 + \cdots + p_k + \cdots = 1$

$\therefore$ $p_0\left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots + \frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu\mu_2\cdots\mu_k} + \cdots\right) = 1$

$$p_0 = 1\Big/\left(1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots + \frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu\mu_2\cdots\mu_k} + \cdots\right)$$

$$= 1\Big/\left(1 + \sum_{i=1}^{\infty} \frac{\lambda_{j-1}\lambda_{j-2}\cdots\lambda_0}{\mu_j\mu_{j-1}\cdots\mu_1}\right) \quad (12.3.5)$$

当为有限个状态时，

$$p_0 = 1\Big/\left(1 + \sum_{i=1}^{k} \frac{\lambda_{j-1}\lambda_{j-2}\cdots\lambda_0}{\mu_j\mu_{j-1}\cdots\mu_1}\right) \quad (12.3.6)$$

当$\frac{\lambda_{j-1}}{\mu_j} < 1$ 的时候，p_0 的分母是收敛的，因此，可求出 p_0。当 p_0 求出后，按 12.3.4 式即可求出其余的各种状态了。

该模型描述了生灭过程，它是研究随机服务系统的重要数学工具之一。

12.4 普阿松输入、负指数分布服务系统分析(M/M/D)

由于随机服务系统的动态情况非常复杂，现只研究比较简单的处于平衡状态的情况。在平衡状态中涉及到一般分布也难于处理，因此只介绍应用比较普遍而又容易处理的几类系统：普阿松输入和负指数分布服务的系统(爱尔朗输入和分布服务的系统与此种情况类似，可参照此法去处理。)

普阿松输入、负指数服务系统的代号是 M/M/D，这类系统必须满足以下四个条件：

(1) 顾客到达服从普阿松分布，时间间隔服从负指数分布，λ 为单位时间到达的平均顾客数，$\frac{1}{\lambda}$为顾客到达的平均时间间隔。

(2) 排成一队。

(3) 先到先服务。

(4) 顾客服务服从负指数分布，μ 为服务台单位时间平均服务的顾客数，μ^{-1}为平均的服务时间间隔。

12.4.1 单服务台系统(M/M/1)

12.4.1.1 单服务台系统的稳态解

系统有可数个状态，且单位时间平均到达的顾客数为 λ，单位时间平均服务顾客数为 μ，因此，根据生灭过程可以得到图 12.4.1 的状态图。

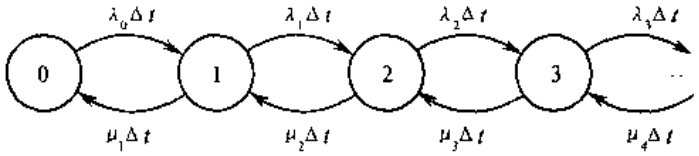


图 12.4.1 M/M/1 系统状态图

由于输入是简单流，故由输入而产生的由一个状态向另一个状态的转移的概率是相同的，即 $\lambda_0=\lambda_1=\lambda_2=\cdots=\lambda$。又由于是单服务台，故由服务而产生的由一个状态转移到另一个状态的概率也是相同的，即 $\mu_0=\mu_1=\mu_2=\cdots=\mu$。

假设 $\lambda<\mu$，且令$\frac{\lambda}{\mu}2\rho$，应用式(12.3.4)得

$$p_1=\frac{\lambda}{\mu}p_0=\rho\ p_0;p_2=\frac{\lambda^2}{\mu^2}p_0=\rho^2p_0,\cdots,p_k=\rho^kp_0;$$

$\because\ p_0+\ p_1+\cdots+\ p_k+\cdots=1$

$\therefore\ p_0+\rho p_0+\cdots+\rho^kp_0+\cdots=1$

$\therefore\ p_0=1/1+\rho+\rho^2+\cdots+\rho^k+\cdots=1$，由于 $1+\rho+\rho^2+\cdots+\rho^k+\cdots$是以 ρ 为公比的等比级数，且 $\rho<1$，故

$$1+\rho+\rho^2+\cdots+\rho^k+\cdots=\frac{1}{1-\rho}$$

$$\therefore\qquad p_0=\frac{1}{\frac{1}{1-\rho}}=1-\rho;p_1=\rho(1-\rho);p_2=\rho^2(1-\rho);\cdots \tag{12.4.1}$$

12.4.1.2 单服务台系统的评价指标

研究处理随机服务系统，主要目的是确定评价服务质量的指标，服务质量评价指标及确定方法如下：

1. 系统处在各种状态的概率 p_i

该指标可说明系统的忙闲程度，也可以使顾客清楚地知道进入系统被服务的各种可能性。我们仍以本章开始的例子用式(12.4.1)进行计算。

每 20min 到来一辆车，即 $\lambda=3$ 辆/h，每 15min 为一辆车加完油，即 $\mu=4$ 辆/h。

$$\because\qquad \rho=\frac{\lambda}{\mu}=\frac{3}{4}=0.75$$

$$\therefore\qquad p_0=1-\rho=0.25\qquad\qquad p_3=\rho^3(1-\rho)=0.1055$$

$$p_1=\rho(1-\rho)=0.1875 \qquad p_4=\rho^4(1-\rho)=0.0791$$
$$p_2=\rho^2(1-\rho)=0.1406 \qquad p_5=\rho^5(1-\rho)=0.0593\cdots$$

这说明顾客进入系统,只有 25% 的可能不排队,排一辆车的可能是 18.75%,…。

2. 平均顾客数(系统中的顾客数)L

平均顾客数是指系统中顾客的平均人数,其中包括正在被服务的顾客。它可以为系统分析和设计人员提供改进系统和设备设置及场地等信息。

这里的平均顾客数是顾客和顾客出现概率的加权平均数。即

$$L=\sum_{k=0}^{\infty}k\cdot p_k=0\cdot p_0+1\cdot p_1+2\cdot p_2+\cdots+lp_1+\cdots=$$
$$\rho p_0\frac{\mathrm{d}}{\mathrm{d}\rho}\int_0^{\rho}(1+2\rho+3\rho^2+\cdots+l\rho^{l-1}+\cdots)\mathrm{d}\rho=$$
$$\rho p_0+2\rho^2p_0+\cdots+l\rho^lp_0+\cdots=$$
$$\rho p_0(1+2\rho+3\rho^2+\cdots+l\rho^{l-1}+\cdots)=$$
$$\rho p_0\frac{\mathrm{d}}{\mathrm{d}\rho}(\rho+\rho^2+\rho^3+\cdots+\rho^l+\cdots)=$$
$$\rho p_0\frac{\mathrm{d}}{\mathrm{d}\rho}\left(\frac{\rho}{1-\rho}\right)=\frac{\rho}{1-\rho}=\frac{\lambda}{\mu-\lambda}$$

所以,系统中的顾客数 L 为

$$L=\frac{\lambda}{\mu-\lambda} \tag{12.4.2}$$

上例中 $\lambda=3,\mu=4,\rho=\frac{3}{4}$,则系统中平均顾客数

$$L=\frac{\rho}{1-\rho}=\frac{\frac{3}{4}}{1-\frac{3}{4}}=3$$

因此,系统设计者可以考虑按平均三个顾客来安排人力、物力、设备等。

3. 平均排队顾客数 L_q

该指标为纯排队顾客数,不包括正在被服务的顾客。明显的,以

$$L_q=L-L_s \tag{12.4.3}$$

来计算,其中 L_q 是正在排队的顾客平均数,L_s 是正被服务顾客平均数。

该指标可为系统分析者和设计者提供队长情况,以决定是否需要改善服务以及提供排队场地。

需要注意的是 L_s 是平均的正在被服务的顾客数,而决非实际正在被服务的顾客数,L_s 是被服务顾客及其概率的加权平均数:

$$\begin{aligned}L_s&=0\cdot p_0+1\cdot p_1+\cdots+1\cdot p_k+\cdots=\\&p_1+p_2+p_3+\cdots+p_k+\cdots=\\&1-p_0=1-(1-\rho)=\rho=\frac{\lambda}{\mu}\end{aligned} \tag{12.4.4}$$

故

$$L_q=L-L_s=\frac{\lambda}{\mu-\lambda}-\frac{\lambda}{\mu}=\frac{\lambda^2}{\mu(\mu-\lambda)} \tag{12.4.5}$$

上例中　　$L_s=\rho=0.75$　　$L_q=3-0.75=2.25$

该系统平均排队车辆是2.25，而不是3－1＝2，因此，系统设计人员可根据此数据设计停车场地。

4．顾客在系统中的平均停留时间 W

该指标是每个顾客最关心的，也是系统分析和设计人员应该重点考虑的问题。顾客都希望在系统中停留时间越短越好，而要达到此目的，必须增加设备和改善服务。因此，系统分析和设计人员应权衡考虑，而选择一最优的方案。

因为$\frac{1}{\lambda}$表示了每到达一个顾客的平均时间间隔，L 是系统中平均顾客数，因此，平均停留时间为：

$$W=\frac{1}{\lambda}=\frac{1}{\mu-\lambda} \tag{12.4.6}$$

上例中 $W=\frac{L}{\lambda}=\frac{1}{4-3}=1(\mathrm{h})$，说明每辆车自进加油站到离开加油站，大约需要1h。因此，顾客对该系统是不会太满意的，系统的设计和分析人员应该提出改善服务的建议和方案。

5．系统中多于 k 个顾客的概率

这项指标是用于判断系统中出现 k 个以上顾客的可能性，提供该信息的目的是使系统管理人员在出现 k 个以上顾客时应采取措施的依据。

系统中出现多于 k 个顾客的概率是“1”和少于等于 k 个顾客的概率之差，即：

$$L(n>k)=1-(p_0+p_1+p_2+p_3+\cdots+p_k)=$$
$$1-(1-\rho)-\rho(1-\rho)-\cdots-\rho^k(1-\rho)=\rho^{k+1}$$

因此
$$L(n>k)=\rho^{k+1}=\left(\frac{\lambda}{\mu}\right)^{k+1} \tag{12.4.7}$$

在上例中
$$L(n>0)=\rho^1=0.75$$
$$L(n>1)=\rho^2=0.5625$$
$$L(n>2)=\rho^3=0.4219\cdots$$

6．顾客在系统中停留的时间超过 T_0 的概率

这个指标是为了在系统设计中保证一个可靠的精度，当顾客在系统中停留时间超过 T_0 时，可能造成损失而影响系统的效益。因此，系统设计和分析人员必须对该指标引起重视。下面不加证明地给出计算公式

$$p(T>T_0)=\mathrm{e}^{-\mu\left(1-\frac{\lambda}{\mu}\right)T_0} \tag{12.4.8}$$

上例中，如果某辆车按在系统中停留1h的计划前去加油，否则将影响去执行任务，那么，该辆车来不及执行任务的可能性是：

$$p(T>1)=\mathrm{e}^{-4\left(1-\frac{3}{4}\right)\times 1}=\mathrm{e}^{-1}=0.368$$

由计算可见，影响执行任务的可能性是很大的。该指标在工程设计中和系统设计中的用处还在于为了保证任务的按时完成，往往预先规定一概率，而确定其进入系统的时间。

如上例中，规定来不及去执行任务的可能性只允许10%，那么，它应该提前多长时间

去加油呢?

由公式反推却可得出

$$e^{-4\left(1-\frac{3}{4}\right)T_0}=10\%\text{，则 }T_0=-\ln 0.1=2.3026$$

可见应该提前 140min 去加油。

这个例子说明该项指标的重要性和实用价值。

还有一些其他指标如用于纯排队所用的时间 W_q 等,将根据系统的不同要求而定。一般的,上述六个指标即可说明一个 M/M/1 系统。

12.4.2 多服务台系统(M/M/S)

12.4.2.1 多服务台系统评价指标

多服务台和单服务台的差别只在于服务情况不同,其输入与单服务台相同,$\lambda_1=\lambda_2=\cdots=\lambda$ 为单位时间平均到达的顾客数量。然而服务确要分为两种情况;一种是当系统中顾客少于服务台个数;另一种是系统中顾客等于或大于服务台个数(见图 12.4.2)。

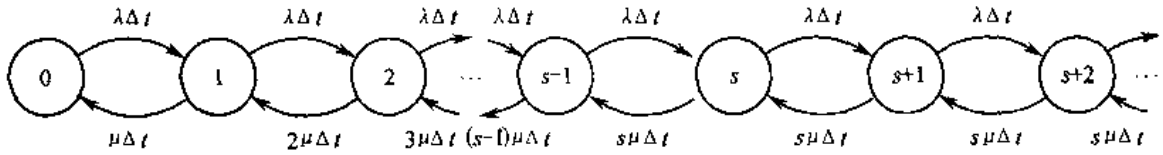


图 12.4.2 多服务台系统

当顾客数小于服务台个数时,所有进入系统的顾客,立刻被服务,所以当有 s 个服务台时,在 s 状态以前各状态,单位时间服务的顾客数为 $\mu,2\mu,3\mu,\cdots,(s-1)\mu$,只有在 s 状态以后,全部服务台均被占用,在单位时间内服务的顾客数为 $S\mu$。这样,按生灭过程的稳态解就可以研究各种指标。

1. 各种状态下的概率

$$p_1=\frac{\lambda_0}{\mu_1}p_0,p_2=\frac{\lambda_0\lambda_1}{\mu_1\mu_2}p_0=\frac{\lambda^2}{2\mu^2}p_0,p_3=\frac{\lambda_3}{2\times 3\mu^3}p_0,\cdots$$

$$p_s=\frac{\lambda^s}{s!\ \mu^s}p_0,p_{s+1}=\frac{\lambda_s}{\mu_{s+1}}p_s=\frac{\lambda^{s+1}}{s\cdot s!\ \mu^{s+1}}p_0,p_{s+2}=\frac{\lambda^{s+2}}{s^2\cdot s!\ \mu^{s+2}}p_0,\cdots$$

由于 $\sum_{i=0}^{\infty}p_i=1$ 得

$$p_0\left\{\sum_{n=0}^{s-1}\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)+\frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s\left[\frac{1}{1-\frac{\lambda}{\mu s}}\right]\right\}=1$$

则

$$p_0=1\Big/\left\{\sum_{n=0}^{s-1}\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)+\frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s\left[\frac{1}{1-\frac{\lambda}{\mu s}}\right]\right\}\tag{12.4.9}$$

p_0 求出后,各状态下的概率 $p_1,p_2,\cdots,p_s,p_{s+1},\cdots$即可求了。

2. 顾客到来需排队的概率

这是求全部服务台已被占满，即大于或等于服务台个数的顾客到来的概率。

$$p(n \geqslant s)=\sum_{n=s}^{\infty} p_n=\frac{\lambda^s}{s!\mu^s}p_0+\frac{\lambda^{s+1}}{s!s\cdot\mu^{s+1}}p_0\cdots=$$

$$p_0\frac{\lambda^s}{s!\mu^s}\left(1+\frac{\lambda}{s\mu}+\frac{\lambda}{s^2\mu^s}+\cdots\right)= \tag{12.4.10}$$

$$\frac{p_0}{1-\frac{\lambda}{s\mu}}\cdot\frac{1}{s!}\cdot\left(\frac{\lambda}{\mu}\right)^s$$

3．平均排队的顾客数

这是求系统中服务台全部被占用需要等待服务而纯排队的顾客数，它是超过服务台的顾客数和其对应的概率的加权平均数：

$$L_q=\sum_{k=0}^{\infty}kp_{s+k}=1p_{s+1}+2p_{s+2}+3p_{s+3}+\cdots+\cdots=$$

$$\frac{\lambda^{s+1}}{s\cdot s!\mu^{s+1}}p_0+\frac{2\lambda^{s+2}}{s^2\cdot s!\cdot\mu^{s+2}}p_2+\cdots= \tag{12.4.11}$$

$$\frac{p_0\cdot\lambda^{s+1}}{s\cdot s!\mu^{s+1}}\cdot\frac{1}{\left(1-\frac{\lambda}{s\mu}\right)^2}$$

4．系统中平均顾客数

$$L=L_q+\frac{\lambda}{\mu} \tag{12.4.12}$$

5．每个顾客在系统中平均停留时间

$$W=L/\lambda \tag{12.4.13}$$

6．顾客在系统中停留时间大于 T_0 的概率

$$p(t>T_0)=\mathrm{e}^{-\mu T_0}\left\{1+\frac{\left(\frac{\lambda}{\mu}\right)^s\cdot p_0\left[1-\mathrm{e}^{-\mu T_0\left(s-1-\frac{\lambda}{\mu}\right)}\right]}{s!\left(1-\frac{\lambda}{\mu s}\right)\left(s-1-\frac{\lambda}{\mu}\right)}\right\} \tag{12.4.14}$$

12.4.2.2　多服务台系统与单服务台系统比较

下面通过实例说明多服务台各指标的计算并与单服务台系统相比较。

某砂场有甲、乙、丙 3 个装车小组，每小组平均 10min 装一车，面空车平均到达 15 辆/h，现研究 3 个小组独立作业和联合作业哪种组织形式更好。

1．独立作业

先看独立作业。由于 3 个小组独立工作，则相当于 3 个单服务台，空车相当于顾客，则每小组每小时平均到达$\frac{15}{3}=5$辆，即 $\lambda=5$ 辆/h，$\mu=6$ 辆/h。

(1) 各种状态的概率：

$$p_0=1-\rho=1-\frac{5}{6}=\frac{1}{6}=0.1667$$

$$p_1=\frac{5}{6}\times0.1667=0.1389$$

$$p_2=\frac{5}{6}\times0.1389=0.1158$$

(2) 砂场平均停留车辆数量 L：

$$L=\frac{\lambda}{\mu-\lambda}=\frac{5}{6-5}=5(\text{辆})$$

(3) 平均等待的空车数量 L_q：

$$L_q=L-\frac{\lambda}{\mu}=5-\frac{5}{6}=4.1667(\text{辆})$$

(4) 空车在砂场的平均停留时间 W：

$$W=L/\lambda=5/5=1(\text{h})$$

2. 联合作业

如果 3 个小组联合作业就相当于一个多服务台的系统，

$$\lambda=15\ \text{辆/h},\ \mu=6\ \text{辆/h},\ s=3。$$

(1) 各种状态下的概率：

$$p_0=1/\left[1+\frac{\lambda}{\mu}+\frac{1}{2}\cdot\left(\frac{\lambda}{\mu}\right)^2+\frac{1}{6}\cdot\left(\frac{\lambda}{\mu}\right)^3\left(1-\frac{\lambda}{3\mu}\right)\right]=0.04495$$

$$p_1=\frac{\lambda}{\mu}p_0=0.1123$$

$$p_2=0.1403$$

(2) 平均等待的空车数量 L_q：

$$L_q=\frac{p_0}{s\cdot s!}\left(\frac{\lambda}{\mu}\right)^{s+1}\left[\frac{1}{1-\frac{\lambda}{s\mu}}\right]^2=3.5112$$

(3) 砂场中等待的车辆数量 L：

$$L=L_q+\frac{\lambda}{\mu}=3.5112+\frac{15}{6}=6.0112$$

(4) 车辆在砂场平均停留时间 W：

$$W=L/\lambda=6.0112/15=0.4(\text{h})$$

两个方案进行对比的结果，3 个小组分散与联合作业的劳动强度相同，都是 0.1667 的概率无车装(联合作业$\frac{1}{3}(3\times p_0+2\times p_1+p_2)=0.1667$)，但是联合作业却使空车的停留时间由 1h 降低到 0.4h。可见联合作业比分散作业有优越性。

12.4.3 有限队长、单服务台系统(M/M/1 有限队长)

有限队长是指系统中纯排队的顾客数不能超过规定的数量，超过则拒绝服务。

对于有限队长为 k 的单服务台系统，它的输入与输出可用下式描述：

$$\lambda_i=\begin{cases}\lambda,\text{当 } i=1,2,\cdots,k \text{ 时}\\0,\text{当 } i=k+1,k+2,\cdots\text{时}\end{cases}$$

$$\mu_i=\begin{cases}\mu,\text{当 } i=1,2,\cdots,k \text{ 时}\\0,\text{当 } i=k+1,k+2,\cdots\text{时}\end{cases}$$

对它的分析仍以生灭过程的稳态解为基础，由式(12.3.6)得：

(1) 各种状态下的概率：

$$p_0=\frac{1}{1+\sum_{i=1}^{k}\left(\frac{\lambda}{\mu}\right)^i}=\frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} \tag{12.4.15}$$

$$p_i=\left(\frac{\lambda}{\mu}\right)^i p_0 \qquad i\leqslant k \tag{12.4.16}$$

(2) 系统中顾客的平均数 L：

$$L=\frac{\lambda}{\mu-\lambda}-\frac{(k+1)\left(\frac{\lambda}{\mu}\right)^{k+1}}{1-\left(\frac{\lambda}{\mu}\right)^{k+1}} \tag{12.4.17}$$

(3) 平均排队的顾客数 L_q：

$$L_q=L-L_s=L-(1-p_0) \tag{12.4.18}$$

(4) 系统有效输入 $\lambda_{有效}$：

由于是有效队长，当到达顾客数超过系统所允许的顾客容量 k 时，则顾客离去，这将造成损失。因此，这种系统也称为损失制系统。在计算 L_q 时，不能应用公式 $L_q=L-\frac{\lambda}{\mu}$，必须对其加以修正，从而引入有效输入系数 $\lambda_{有效}$。即 $L_q=L-\frac{\lambda_{有效}}{\mu}$。

在损失制的系统中 λ 有可能大于 μ，即 $\frac{\lambda}{\mu}>1$，但由于采用了损失制而能在一定的时间完成任务。

由于 $L_q=L-\frac{\lambda}{\mu}$，所以 $\lambda_{有效}=\mu(L-L_q)=\mu\{L-[L-(1-p_0)]\}$

$$\lambda_{有效}=\mu(1-p_0) \tag{12.4.19}$$

(5) 顾客在系统中的平均停留时间 W：

$$W=L/\lambda_{有效} \tag{12.4.20}$$

(6) 损失率：

由于队长有限而排不上队的顾客即为系统的损失，故损失率为系统容量为 k 时的概率。

$$p_k=\left(\frac{\lambda}{\mu}\right)^k p_0 \tag{12.4.21}$$

12.5 随机服务系统的费用优化

研究随机服务系统的目的之一就是以最省的费用达到最好的服务质量。

众所周知，欲减少顾客排队和排队时间，无外乎是增加服务人员和服务设施或提高服务人员水平两条途径。无论采用哪条途径都会相应地增加服务机构的费用。但如不改善服务质量，顾客排队过多，等待时间过长会使一部分顾客离去而影响收益，造成损失。因此有必要研究服务系统处在什么样的状况才能使系统最优的问题。

由于服务系统类型比较多，且大部情况都比较复杂，本节主要介绍 M/M/1 和 M/M/S 系统的优化模型。

12.5.1 M/M/1 系统优化模型

为改进 M/M/1 系统的服务质量，只能由服务率 μ 入手，确定最佳服务率 μ^*。

设 C_s——服务机构单位时间的费用；

C_w——每个顾客在系统中停留单位时间的费用。

服务系统总成本 C 则为单位时间的服务成本和顾客在系统中停留单位时间的费用之和的期望值：

$$C = C_s\mu + C_w + L_q \tag{12.5.1}$$

将 $L_q = \frac{\lambda}{\mu - \lambda}$ 代入上式得

$$C = C_s\mu + C_w \frac{C_w\lambda}{\mu - \lambda} \tag{12.5.2}$$

由式(12.5.2)可见，系统总费用由两部分组成，前部分是服务费用，它与服务率成正比；后部分与服务率和到达率有关，在顾客到达规律确定的条件下，与服务率成反比，可用图 12.5.1 说明系统总费用和服务率 μ 的关系。

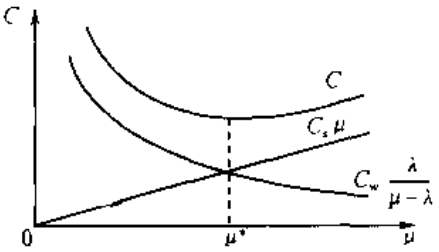


图 12.5.1 费用与服务率的关系

为使服务系统处于最佳状态，可通过式(12.5.2)求微分的方法确定最优服务率

$$\frac{dc}{d\mu} = C_s - C_w \frac{\lambda}{(\mu - \lambda)^2} = 0$$

$$(\mu - \lambda)^2 = \frac{C_w\lambda}{C_s}$$

$$\because \mu > \lambda \qquad \therefore \mu - \lambda > 0$$

$$\mu^* = \lambda + \sqrt{\frac{C_w\lambda}{C_s}} \tag{12.5.3}$$

对于队长有限的 M/M/1 系统，最优服务率由下述方程的解确定

$$\rho^{N+1}\frac{N - (N+1)\rho + \rho^{N+1}}{(1 - \rho^{N+1})^2} = \frac{C_s}{G} \tag{12.5.4}$$

式中 $\rho = \frac{\lambda}{\mu}$；

G——为单位顾客服务的收入；

N——系统中最大顾客数。

12.5.2 M/M/S 系统优化模型

M/M/S 系统中确定最优的服务台数是主要的优化问题。

设系统单位时间总费用为 C，则

$$C = C'_s s + C_w \cdot L \tag{12.5.5}$$

式中 C'_s——每个服务台单位时间的成本。

系统单位时间总费用 C 是以 s 为变量的非线性函数，应用非线性规划即可求出 s(需

加入 s 为整数的约束)。

由于服务台数 s 是离散变量,亦可采用边际分析法确定最佳服务台数。

C 是 s 的函数,最佳服务台数 s^*,使 $C(s^*)$取最小值,即

$$C(s^*)\leqslant C(s^*+1)$$
$$C(s^*)\leqslant C(s^*-1) \tag{12.5.6}$$

将式(12.5.5)代入得

$$C'_s s^* + C_w \cdot L(s^*) \leqslant C'_s(s^*+1) + C_w \cdot L(s^*+1)$$
$$C'_s s^* + C_w \cdot L(s^*) \leqslant C'_s(s^*-1) + C_w \cdot L(s^*-1)$$

化简得

$$L(s^*) - L(s^*+1) \leqslant \frac{C'_s}{C_w} \leqslant L(s^*-1) - L(s^*) \tag{12.5.7}$$

由式(12.5.7)可知,当依次求出 $s=1,2,\cdots$的 L 值,并计算相邻 L 值之差,则得到一系列的区间,当$\frac{C'_s}{C_w}$落在某区间内时,即可确定为最优的服务台个数。

现举例说明 M/M/1 和 M/M/S 系统的费用优化方法。

在 12.4.2.2 节的例子中,采用分组作业方式,如果每小组工作 1h 需支付 10 元工资,每停工 1h,损失 10 元,该小组的工作效率应是多少才能使系统总费用最低。

这是单服务台求最优服务率的问题。

因 $\lambda = 5$ 辆/h, $C_s = 10$ 元, $C_w = 10$ 元,

故 $$\mu^* = \lambda + \sqrt{\frac{C_w}{C_s}\lambda} = 5 + \sqrt{\frac{10}{10}\cdot 5} = 5 + 2.24 = 7.24(\text{辆})$$

该小组工作效率应为每小时装 7.24 辆。

如该砂场准备改进装车设备,设计指标为:48 辆/h,每个装车台可装车 25 辆/h,每辆车装车成本为 4 元,车辆等待平均损失 6 元/h,该砂场应设几个装车台才能使总费用最少。

这是多服务台系统求最优服务台数量的问题。表面上需 2 个装车台即可装 50 辆车/h,似乎问题很简单。现分析计算确定最佳服务台数量。

因 $C'_s = 4$ 元, $C_w = 6$ 元, $\lambda = 48$, $\mu = 25$, $\rho = \frac{\lambda}{\mu} = 1.92$

设装车台个数为 s,则由式(12.4.12)知

$$L = \frac{p_0 \lambda^{s+1}}{s \cdot s! \mu^{s+1}} \cdot \left\{\frac{1}{1-\frac{\lambda}{s\mu}}\right\} + \frac{\lambda}{\mu}$$

$$= \frac{(1.92)^{s+1}}{s \cdot s!} \cdot \left\{\frac{1}{1-\frac{48}{s \cdot 25}}\right\}^2 p_0 + 1.92$$

其中 $$p_0 = \left\{\sum_{n=0}^{s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right) + \left(\frac{1}{s!}\right)\left(\frac{\lambda}{\mu}\right)^s \left\{\frac{1}{1-\frac{\lambda}{s\mu}}\right\}\right\}^{-1}$$

令 $s=1,2,3,4,5$,并依次代入上式,得下表:

装车台数 s	平均车辆数 L	$L(s)-L(s+1)\sim L(s-1)-L(s)$	总费用 C
1	∞		∞
2	24.490	21.845～∞	154.94
3	2.645	0.582～21.845	27.87
4	2.063	0.111～0.582	28.38
5	1.952		31.71

$\frac{C_s'}{C_w}=\frac{4}{6}=0.667$,落在(0.582,21.845)区间内,故应设3个装车台可使系统总费用最低。

第 13 章　动态系统基础

以前各章介绍的均是假设系统处在相对静止或平衡状态时所采用的方法，也就是假设系统各组成部分及它们之间的联系不随时间的推移而变化。然而自然界和社会中的所有系统，其组成要素及相互之间的联系都是随时间的推移而变化的，都是动态系统。

一个动态系统，在不同的控制下，将产生不同的行为。我们研究动态系统的目的就是要探讨系统发展变化的内在规律，制定一套合适的控制措施，达到某评价指标的最大(小)值，使系统按人类的要求运行。然而，什么样的系统、在什么条件下才能按人的意图运行，取决于系统的原理和其内在的规律性，只有掌握系统及其各组成部分的属性，按其相互的联系所表现的规律性制定相应的措施，才能更好地使系统为人类服务。这就是动态系统所研究的内容。

研究动态系统的基本理论是古典控制论、现代控制论和大系统理论。它们是属技术科学范畴的，是系统工程的基础理论。

由于动态系统的处理需要较深的数学理论，因此本章将就动态系统的一般处理方法作一简要介绍，作为进一步研究的入门。

13.1　动态系统一般知识

动态系统不论是机械的、工程的，还是经济的、社会的、生态的，都可应用微分方程描述其发展变化的规律性。这种用微分方程描述的反映系统动态特征的数学表达式叫系统动态模型。建立反映系统动态特性的数学模型是研究和分析一个动态系统的重要方法之一，也是最关键、最重要的环节之一。

13.1.1　系统动态模型的分类

系统动态数学模型按解决问题目的可分一般动态模型和最优化动态模型两种。

一般动态模型是用微分方程(组)或差分方程(组)来反映系统动态行为的模型。它是研究动态系统内在运行规律和在控制条件下系统如何运行的重要工具，是进行动态系统决策的主要依据。

动态最优化模型是按系统运行内在规律性，使系统某些性能指标达到最优的模型。

上述两种模型按时间取值形式不同又都分成连续型和离散型模型两类。

连续型模型是指时间变量 t，在某一时间范围内可连续取值的模型。该类模型的特点是模型各变量的导数存在，故可用微分方程描述。而离散模型则时间变量 t 仅取某时间范围内的整数值，系统变量的导数不存在，只能用差分方程描述系统的行为。

连续模型和离散模型按变量之间的关系可分为线性和非线性模型；按变量系数是否

随时时间的变化分为时变和定常模型;按常数项是否为零又分为齐次和非齐次模型等,系统模型的分类见表 13.1.1。

表 13.1.1 系统动态模型分类表

		一般动态模型		最优化动态模型	
		连续型	离散型	连续型	离散型
线性模型	定常系统	线性连续定常模型	线性离散定常模型	线性连续定常最优化模型	线性离散定常最优化模型
	时变系统	线性连续时变模型	线性离散时变模型	线性连续时变最优化模型	线性离散时变最优化模型
非线性模型	定常系统	非线性连续定常模型	非线性离散定常模型	非线性连续定常最优化模型	非线性离散定常最优化模型
	时变系统	非线性连续时变模型	非线性离散时变模型	非线性连续时变最优化模型	非线性离散时变最优化模型

由于非线性模型和时变模型的内容涉及较深的数学理论,因此本章主要介绍线性定常系统,其目的是使读者初步掌握处理动态系统的一般方法,作为研究动态系统的入门。

13.1.2 连续动态系统模型的表达形式

13.1.2.1 微分方程法

微分方程法是描述系统动态特性的基本方法。著名的多马增长模型即是这种模型。

多马对投资量变化的分析得出如下结论:

(1) 投资量的变化将影响生产能力,从企业角度研究,企业产出量的变化应与投资量成正比,故有

$$\frac{dB}{dt} = EI \tag{13.1.1}$$

式中 $\frac{dB}{dt}$——产出量(以价值表示)的变动率;

E ——投资效益系数;

I ——投资量。

(2) 生产能力的增长受社会需求量变化的影响,从社会需求角度研究,投资量的变动与需求量的变动成正比,故有

$$\frac{dY}{dt} = \frac{dI}{dt} \cdot \frac{1}{s} \tag{13.1.2}$$

式中 $\frac{dY}{dt}$——需求量的变动率;

$\frac{dI}{dt}$ ——投资量的变动;

s ——储蓄边际值。

(3) 生产能力的增长速度总是要与需求的增长速度相吻合,因此有

$$\frac{dB}{dt} = \frac{dY}{dt} \tag{13.1.3}$$

将式(13.1.1)和式(13.1.2)代入式(13.1.3)得

$$\frac{1}{s} \cdot \frac{dI}{dt} = EI$$

整理得

$$\frac{1}{I}\cdot\frac{\mathrm{d}I}{\mathrm{d}t}=sE \quad 或 \quad \frac{1}{I}\dot{I}=sE \tag{13.1.4}$$

式(13.1.4)表示的微分方程就是反映投资量和需求量平衡关系的微分方程,它反映出该投资系统历年投资量的变化规律。该规律可通过求解该微分方程得到。

应用高等数学知识,该微分方程的求解过程和结果如下:

$$\int\frac{1}{I}\frac{\mathrm{d}I}{\mathrm{d}t}\mathrm{d}t=\int SE\,\mathrm{d}t$$

$$\ln I=sEt+C$$

$$I(t)=\mathrm{e}^{sEt}\cdot\mathrm{e}^{c}=A\mathrm{e}^{sEt}$$

令 $t=0$,则 $I(0)=A$,根据初始条件确定出常数 A,从而得出历年投资的变化规律为

$$I(t)=I(0)\cdot\mathrm{e}^{sEt} \qquad t>0 \tag{13.1.5}$$

可见该投资系统的变化呈指数函数增长。

这就是用微分方程的方法所建立的动态系统数学模型。这种建立模型的方法是反映系统动态特性的最基本方法。

13.1.2.2 传递函数法

直接求解用微分方程法表示的系统动态模型,只能在很简单的情况下才能做到。对于单输入、单输出系统,线性微分方程的一般形式为

$$a_0y^{(n)}+a_1y^{(n-1)}+\cdots+a_ny=b_0u^{(m)}+b_1u^{(m-1)}+\cdots+b_mu \tag{13.1.6}$$

该方程反映了系统的输入 u 和输出 y 的关系,它是一高阶微分方程,其求解是很困难的。

现把反映单输入、单输出系统的线性微分方程叫输出输入方程,可借助于一种数学工具——拉普拉斯变换使求解变得容易,现将输出输入方程经拉普拉斯变换后所得到的方程叫传递函数。下面分别介绍拉普拉斯变换法和传递函数。

1. 拉普拉斯变换和反变换

在高等数学中解高阶线性微分方程的一种简便的运算方法就是把线性微分方程转换为复变数的代数方程,这就是拉普拉斯变换。

拉普拉斯变换规定:

$f(t)$——时间 t 的函数,且当 $t<0$ 时,$f(t)=0$;

s——复变数;

L——拉普拉斯运算符号,放在某量前表示该量用拉普拉斯积分:$\int_0^\infty \mathrm{e}^{-st}\,\mathrm{d}t$ 进行变换。$F(t)$的拉普拉斯变换:

$$F(s)=L[f(t)]=\int_0^\infty f(t)\mathrm{e}^{-st}\,\mathrm{d}t \tag{13.1.7}$$

拉普拉斯变换有如下性质:

当 $L[f(t)]=F(s)=\int_0^\infty \mathrm{e}^{-st}f(t)\mathrm{d}t$ 时,

(1) $L[f_1(t)+f_2(t)]=F_1(s)+F_2(s)$;

(2) $L[af(t)]=aF(s)$;

(3) $L[f(t-C)]=e^{-Cs}F(s)$, $t>C$, C 为正常数;

(4) $L[e^{at}f(t)]=F(s-C)$, C 为常数;

(5) $Lf'(t)=sF(s)-f(0)$, (0)为函数的初值;

(6) $Lf''(t)=s^2F(s)-sf(0)-f'(0)$;

(7) $L[\int_0^t f(t)dt]=\frac{F(s)}{s}$;

(8) $L[\int f(t)dt]=\frac{F(s)}{s}+\frac{1}{s}[\int f(t)dt]t=0$;

(9) $L[tf(t)]=-F'(s)$。

根据拉普拉斯变换的定义和上述性质可推导出一些常用函数的拉普拉斯变换。

例1 函数 $f(t)=Ae^{at}$, $t\geqslant 0$, A、a 均为常数

则 $F(s)=L[f(t)]=\int_0^\infty Ae^{-at}e^{-st}dt=A\int_0^\infty e^{-(a+s)t}dt=\frac{A}{a+s}$($s$ 的实部大于 $-a$)

例2 函数 $f(t)=\begin{cases}0, & t<0\\ 1, & t\geqslant 0\end{cases}$

则 $$F(s)=L[f(t)]=\int_0^\infty e^{-st}dt=\frac{1}{s}$$

例3 函数 $f(t)=t$, $t\geqslant 0$,

则 $$F(s)=L[f(t)]=\int_C^\infty f(t)e^{-st}dt=\int_0^\infty te^{-st}dt=\frac{1}{s^2}。$$

在数学手册中通常把常用函数的拉普拉斯变换列成表,而不必自行推导,直接查找即可使用,常用函数的拉普拉斯变换如下:

$$L_C=\frac{C}{s},\ C\text{ 为常数},$$

$$Lt=\frac{1}{s^2}$$

$$Lt^n=\frac{n!}{s^{n+1}},\ n\text{ 为正整数},$$

$$Lt^{-\frac{1}{2}}=\sqrt{\frac{\pi}{s}}$$

$$Lt^{\frac{1}{2}}=\frac{1}{2}\sqrt{\pi}\cdot s^{-\frac{3}{2}}$$

$$Le^{Ct}=\frac{1}{s-C}$$

$$Le^{-Ct}=\frac{1}{s+C}$$

$$Lt^ne^{Ct}=\frac{n!}{(s-C)^{n+1}},\ n\text{ 为正整数},$$

$$L\cos Ct=\frac{s}{s+C^2}$$

$$Lt\cos Ct=\frac{s^2-C^2}{(s^2+C^2)^2}$$

$$L\mathrm{e}^{at}\cos Ct=\frac{s-a}{(s-a)^2+C^2}$$

$$L s\sin Ct=\frac{C}{s^2+C^2}$$

$$L t\sin Ct=\frac{2Cs}{(s^2+C^2)^2}$$

$$L\mathrm{e}^{at}\sin Ct=\frac{C}{(s-a)^2+C^2}$$，为常数。

由拉普拉斯变换推导成时间表达式的运算叫拉普拉斯反变换。它是拉普拉斯变换的逆运算，以 L^{-1}表示，即 $L^{-1}[F(s)]=f(t)$。

用积分表示则是 $f(t)=\frac{1}{2\pi j}\int_{C-j\omega}^{C+j\omega}F(s)\mathrm{e}^{st}\mathrm{d}s$，$t>0$。

可见直接积分是相当困难的，因此，一般的办法都是运用拉普拉斯变换表。

对于较简单的拉普拉斯变换可直接查表，但对于较复杂的形式，必须经过变化使其符合表中的形式后再应用变换表，一般采用部分分式展开的方法。

如果拉普拉斯变换 $F(s)$可分解成 n 个符合变换表形式的分量：

$$F(s)=F_1(s)+F_2(s)+\cdots+F_n(s)$$

则

$$L^{-1}[F(s)]=L^{-1}[F_1(s)]+L^{-1}[F_2(S)]+\cdots+L^{-1}[F_n(s)]$$

如果拉普拉斯变换 $F(s)$是一分式，即 $F(s)=\frac{B(s)}{A(s)}$，且 $A(s)$的阶次高于 $B(s)$的阶次，则可应用待定系数法将其变成一种简单的形式

$$F(s)=\frac{B(s)}{A(s)}=\frac{k(s+Z_1)(s+Z_2)\cdots(s+Z_m)}{(s+P_1)(s+P_2)\cdots(s+P_m)}=\frac{a_1}{s+P_1}+\frac{a_2}{s+P_2}+\cdots+\frac{a_n}{s+P_n}$$

然后求反变换

$$L^{-1}[F(s)]=L^{-1}\left(\frac{a_1}{s+P_1}\right)+L^{-1}\left(\frac{a_2}{s+P_2}\right)+\cdots+L^{-1}\left(\frac{a_n}{s+P_n}\right)$$

这种情况必须在已知 $A(s)$的根的情况下，也就是必须在能把 $A(s)$分解成因式的情况下才能使用。如求 $F(s)=\frac{s^2+2s+3}{(s+1)^3}$的拉普拉斯变换

$$F(s)=\frac{a_1}{(s+1)^3}+\frac{a_2}{(s+1)^2}+\frac{a_3}{(s+1)}=\frac{s^2+2s+3}{(s+1)^3}$$

则

$$a_1+a_2(s+1)+a_3(s+1)^2=s^2+2s+3$$

当$s=-1$时，$a_1=2$

$s=0$时，$a_2+a_3=1$

$s=1$时，$a_2+2a_3=2$，解得 $a_2=0$，$a_3=1$

即

$$F(s)=\frac{3}{(s+1)^3}+\frac{1}{(s+1)}$$

$$\therefore f(t)=L^{-1}[F(s)]=L^{-1}\left[\frac{2}{(s+1)^3}\right]+L^{-1}\left[\frac{1}{(s+1)}\right]=t^2\mathrm{e}^{-t}+\mathrm{e}^{-t}=(t^2+1)\mathrm{e}^{-t}$$

应用拉普拉斯变换可解线性微分方程，其方法就是应用拉普拉斯变换把微分方程中

的每一项进行变换，得到一个以 s 为变量的代数方程，解出的解再经过拉氏反变换即是微分方程的解。

2. 传递函数

传递函数是应用拉普拉斯变换所确定的系统在初始状态为零是时的输出和输入之比，以 $T(s)$ 表示。

将式(13.1.6)进行拉氏变换后，可得传递函数的一般形式：

$$T(s)=\frac{Y(s)}{u(s)}=\frac{b_0s^m+b_1s^{m-1}+\cdots+b_{m-1}s+b_m}{a_0s^n+a_1s^{n-1}+\cdots+a_{n-1}s+a_n} \tag{13.1.8}$$

由式(13.1.8)可见

$$Y(s)=T(s)\cdot u(s)$$

经拉氏反变换即可得到输入和输出的关系。

用传递函数求解微分方程并确定系统动态特性的步骤如下：

(1) 求出系统的传递函数；

(2) 将系统输入进行拉普拉斯变换；

(3) 按 $Y(s)=T(s)\cdot u(s)$ 求出输入和输出的拉普拉斯变换；

(4) 由系统输出的拉普拉斯变换经反变换后求出系统的输出。

传递函数法是古典控制论的主要工具，它主要解决单输入单输出的线性定常系统。

13.1.2.3 状态空间法

状态空间法是用状态变量描述系统的一种方法。系统状态又叫状态变量，是表示系统特性的一组变量。系统的每一状态对应于欧式空间中的一点，该空间就是状态空间，状态空间的维数等于决定系统状态的独立变量的个数。

系统状态随时间的变化是有规律的，描述该变化规律的方程叫状态方程，它一般是一阶矩阵微分方程——一阶微分方程组。

状态空间法用较严格的数学语言可描述为：所谓状态变量就是确定系统状态的最小的一组变量 $x_1(t),x_2(t),\cdots,x_n(t)$，它满足：

(1) 在任何时刻 $t=t_0$，这种变量的值 $x_1(t_0),x_2(t_0),\cdots,x_n(t_0)$ 都表示系统在 t_0 时刻的状态。

(2) 当系统在 $t\geqslant t_0$ 的输入和上述初始状态确定时，状态变量能完全表示出系统将来的行为。

这种状态变量与系统的输入之间的一阶微分方程组就是状态方程，而反映状态变量与系统输出关系的代数方程就是输出方程。

状态方程和输出方程的一般形式是：

$$\dot{X}=f(x,t,u) \tag{13.1.9}$$

$$y=g(x,u) \tag{13.1.10}$$

式(13.1.9)是状态方程，式(13.1.10)是输出方程，其中 x 是状态变量；u 是系统的输入，又称控制变量；y 是系统的输出。

状态方程和输出方程的线性标准形式为

$$\dot{X}=AX+Bu \tag{13.1.11}$$

$$y=CX+Du \tag{13.1.12}$$

13.1.2.4 各种描述方法之间的转换

状态空间法是现代控制论描述动态系统的主要方法，它可解决多输入多输出系统的描述，然而各种描述方法的基础都是微分方程法，因此研究各种描述方法间的转换可揭示出古典控制论和现代控制论之间的联系。现分别介绍各种描述方法间的转换。

1. 由高阶微分方程求状态方程

现仅就输入函数不含导数的情况来说明。

这时系统的运动方程为

$$y^{(n)}+a_1y^{(n-1)}+\cdots+a_{n-1}\dot{y}+a_ny=u \tag{13.1.13}$$

根据微分方程的理论，如 $y(0),\dot{y}(0),\cdots,y^{(n-1)}(0)$ 及 $t\geqslant 0$ 的输入 $u(t)$ 已知时，则系统未来的运动状态完全确定，现取 $y(t),\dot{y}(t),\cdots,y^{(n-1)}(t)$ 作为系统的状态变量，将这些变量记为

$$X_1=y,\quad X_2=\dot{y},\quad \cdots,\quad X_n=y^{(n-1)} \tag{13.1.14}$$

则 $\dot{X}_1=X_2,\quad \dot{X}_2=X_3,\quad \cdots,\quad \dot{X}_{n-1}=X_n,$

$$\dot{X}_n=u-a_1y^{(n-1)}-a_2y^{(n-2)}-\cdots-a_ny=u-a_1X_n-a_2X_{n-1}-\cdots-a_nX_1 \tag{13.1.15}$$

将其写成矩阵形式，得状态方程为

$$\dot{\boldsymbol{X}}=\boldsymbol{AX}+\boldsymbol{Bu}$$

输出方程为

$$\boldsymbol{Y}=\boldsymbol{CX}$$

其中，$\boldsymbol{X}=\begin{bmatrix}X_1\\X_2\\\vdots\\X_n\end{bmatrix}$ $\boldsymbol{A}=\begin{bmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\\vdots&\vdots&\vdots&\cdots&\vdots\\0&0&0&\cdots&1\\-a_n&-a_{n-1}&-a_{n-2}&\cdots&-a_1\end{bmatrix}$ $\boldsymbol{B}=\begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}$

$\boldsymbol{Y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}$ $\boldsymbol{C}=(1\quad 0\quad 0\cdots 0)$

如一系统的微分方程为

$$\dddot{y}+6\ddot{y}+11\dot{y}+6y=6u$$

令 $X_1=y,\quad X_2=\dot{y},\quad X_3=\ddot{y}$

则 $\dot{X}_1=X_2\quad \dot{X}_2=X_3$

$$\dot{X}_2=\ddot{y}=6u-6y-11\dot{y}-6\ddot{y}$$

代入输出输入方程得

$$\dot{X}_3=6u-6y-11\dot{y}-6\ddot{y}$$

$$\dot{X}_2=X_3$$

$$\dot{X}_1=X_2$$

$$y=X_1$$

即

$$\dot{\boldsymbol{X}}=\begin{bmatrix}\dot{X}_1\\ \dot{X}_2\\ \dot{X}_3\end{bmatrix}=\begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 1\\ -6 & -11 & -6\end{bmatrix}\begin{bmatrix}X_1\\ X_2\\ X_3\end{bmatrix}+\begin{bmatrix}0\\ 0\\ 6\end{bmatrix}u$$

也可写成 $$\dot{\boldsymbol{X}}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}u$$

其中

$$\dot{\boldsymbol{X}}=\begin{bmatrix}\dot{X}_1\\ \dot{X}_2\\ \dot{X}_3\end{bmatrix}\quad \boldsymbol{A}=\begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 1\\ -6 & -11 & -6\end{bmatrix}\quad \boldsymbol{B}=\begin{bmatrix}0\\ 0\\ 6\end{bmatrix}$$

输出方程为 $$y=X_1$$

写成矩阵形式为 $$y=\boldsymbol{C}\boldsymbol{X}$$

其中， $$\boldsymbol{C}=(1\quad 0\quad 0)\qquad \boldsymbol{X}=(\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3)^{\mathrm{T}}$$

2. 由传递函数求状态方程

通过传递函数，采用类似电路并联的方法也可以求出状态方程。

仍用上例： $$\dddot{y}+6\ddot{y}+11\dot{y}+6y=6u$$

则传递函数 $T(s)$ 为

$$T(s)=\frac{6}{(s+1)(s+2)(s+3)}=\frac{3}{s+1}-\frac{6}{s+2}+\frac{3}{s+3}$$

因为 $$T(s)=\frac{Y(s)}{u(s)}$$

所以 $$Y(s)=\frac{3}{s+1}u(s)-\frac{6}{s+2}u(s)+\frac{3}{s+3}u(s)$$

可见系统的输出是由三部分组成的，相当于图 13.1.1。

由于 $$Y(s)=X_1+X_2+X_3$$

故有：

$$X_1=\frac{3}{s+1}u(s)$$

$$X_2=\frac{-6}{s+2}u(s)$$

$$X_3=\frac{3}{s+3}u(s)$$

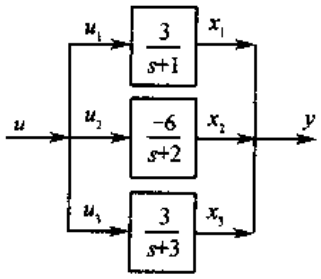


图 13.1.1 系统的输出

所以 $X_1(s+1)=3u(s)$，$X_2(s+2)=-6u(s)$，$X_3(s+3)=3u(s)$

进行拉普拉斯反变换得到

$$\dot{X}_1=-X_1+3u$$

$$\dot{X}_2=-2X_2-6u$$

$$\dot{X}_3=-3X_3+3u,$$

写成标准形式

$$\dot{\boldsymbol{X}}=\begin{bmatrix}\dot{X}_1\\ \dot{X}_2\\ \dot{X}_3\end{bmatrix}=\begin{bmatrix}-1 & 0 & 0\\ 0 & -2 & 0\\ 0 & 0 & -3\end{bmatrix}\begin{bmatrix}X_1\\ X_2\\ X_3\end{bmatrix}+\begin{bmatrix}3\\ -6\\ 3\end{bmatrix}u$$

输出方程 $Y=X_1+X_2+X_3$，写成标准形式为

$$\boldsymbol{Y}=(1\quad 1\quad 1)\cdot\begin{bmatrix}X_1\\X_2\\X_3\end{bmatrix}$$

由传递函数采用类似电路串联的方法仍可得到状态方程和输出方程。

如上述的传递函数：

$$T(s)=\frac{6}{(s+1)(s+2)(s+3)}=\frac{6}{s+1}\cdot\frac{1}{s+2}\cdot\frac{1}{s+3}$$

$$\because\qquad T(s)=\frac{Y(s)}{u(s)}\quad 或\quad Y(s)=\frac{6u(s)}{s+1}\cdot\frac{1}{s+2}\cdot\frac{1}{s+3}$$

由串联图 13.1.2 可知前一级的输出是后一级的输入，

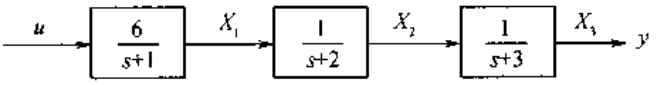


图 13.1.2　串联图

故　　$X_1=\dfrac{6}{s+1}u(s)$，　$X_2=\dfrac{1}{s+2}X_1$，　$X_3=\dfrac{1}{s+3}X_2$，　$Y=X_3$

即　　$X_1(s+1)=6u$，$X_2(s+2)=X_1$，$X_3(s+3)=X_2$，$Y=X_3$

进行拉普拉斯反变换得到状态方程和输出方程：

$$\dot{\boldsymbol{X}}=\begin{bmatrix}\dot{X}_1\\\dot{X}_2\\\dot{X}_3\end{bmatrix}=\begin{bmatrix}-1&0&0\\1&-2&0\\0&1&-3\end{bmatrix}\cdot\begin{bmatrix}X_1\\X_2\\X_3\end{bmatrix}+\begin{bmatrix}6\\0\\0\end{bmatrix}u$$

$$\boldsymbol{Y}=(0\quad 0\quad 1)\cdot\begin{bmatrix}X_1\\X_2\\X_3\end{bmatrix}$$

从上面的运算可以看出当单输入、单输出的传递函数一定时所得的状态方程和输出方程因所采用的方法不同而不同，这也正说明系统在相同的输入和输出要求下，可用不同结构去实现，这是符合客观事实的。

3．由状态方程求传递函数

一单输入、单输出系统，其状态方程为

$$\dot{\boldsymbol{X}}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}u$$

$$\boldsymbol{Y}=\boldsymbol{C}\boldsymbol{X}+Du$$

式中　$X(t)$——n 维状态向量；

$u(t)$和 $y(t)$——分别为输入和输出函数；

$\boldsymbol{A}$——$n\times n$ 阶系数矩阵；

$\boldsymbol{B}$——$n\times 1$ 阶控制矩阵；

$\boldsymbol{C}$——$1\times n$ 阶输出矩阵；

D——数量。

对上述二方程进行拉普拉斯变换，得

$$sX(s)=AX(s)+Bu(s)\qquad(13.1.16)$$

$$Y(s)=CX(s)+Du(s) \tag{13.1.17}$$

从方程(13.1.16)得

$$sX(s)-AX(s)=[sI-A]X(s)=Bu(s)$$

两边同左乘$[sI-A]^{-1}$,得

$$X(s)=[sI-A]^{-1}Bu(s)$$

将其代入式(13.1.17)得

$$Y(s)=C(sI-A)^{-1}B\,u(s)+Du(s)$$

由此得传递函数为

$$T(s)=\frac{Y(s)}{u(s)}=s(sI-A)^{-1}B+D \tag{13.1.18}$$

例:系统动态方程为

$$\dot{\boldsymbol{X}}=\begin{bmatrix}\dot{X}_1\\ \dot{X}_2\\ \dot{X}_3\end{bmatrix}=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ -5 & -3 & -2\end{bmatrix}\cdot\begin{bmatrix}X_1\\ X_2\\ X_3\end{bmatrix}+\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}u$$

$$\boldsymbol{Y}=\left(\frac{3}{2}\quad 1\quad \frac{1}{2}\right)\cdot\begin{bmatrix}X_1\\ X_2\\ X_3\end{bmatrix}$$

其传递函数求法为

(1) 求$(sI-A)^{-1}$:

$$(sI-A)^{-1}=\begin{bmatrix}s & -1 & 0\\ 0 & s & -1\\ 5 & 3 & s+2\end{bmatrix}^{-1}=\frac{1}{s^3+2s^2+3s+5}\begin{bmatrix}s^2+2s+3 & s+2 & 1\\ -5 & s(s+2) & s\\ -5s & -(3s+5) & s^2\end{bmatrix}$$

(2) 代入式(13.1.18)求出 $T(s)$:

$$T(s)=\frac{\left(\frac{3}{2}\quad 1\quad \frac{1}{2}\right)}{s^3+2s^2+3s+5}\begin{bmatrix}s^2+2s+3 & s+2 & 1\\ -5 & s(s+2) & s\\ -5s & -(3s+5) & s^2\end{bmatrix}\cdot\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}=$$

$$\frac{\left(\frac{3}{2}\quad 1\quad \frac{1}{2}\right)}{s^3+2s^2+3s+5}\begin{bmatrix}1\\ s\\ s^2\end{bmatrix}=\frac{s^2+2s+3}{2s^3+4s^2+6s+10}$$

由上述可见,对单输入单输出系统,古典控制论和现代控制论是密切联系的,现代控制论在描述系统动态方面较古典控制论优越。

在多输入多输出的情况下应用传递函数对系统进行更深刻的分析则无能为力了,只有用状态方程和输出方程,才能清楚地描述。

13.1.3 离散动态系统模型的表达形式

在连续函数的数学分析中,当自变量 x 有一增量 Δx 时,函数有一增量 Δy,当 $\Delta x\to 0$ 时,$\frac{\Delta y}{\Delta x}$的极限就是该函数的一阶导数。

我们所研究的动态系统是以时间和状态为变量的,因此连续模型可用微分方程描述。

然而对于离散系统，变量 t 只能取某一时期内的整数，而且时间 t 的变化 $\Delta t=1$。由于函数不连续，导数不存在，加之 $\Delta t=1$，故只能用另一种形式来描述系统的变化，这种形式就是差分。

所谓差分是指函数在离散时间下因变量的增量 Δy 与自变量增量 Δt 的商，即$\frac{\Delta y}{\Delta t}$，因 $\Delta t=1$，故函数的增量 Δy 以下式表示：

$$\Delta y(t)=y(t+1)-y(t) \tag{13.1.19}$$

式中 $\Delta y(t)$——该函数的一阶差分；

$y(t+1)$——该函数 $t+1$ 期的函数值；

$y(t)$ ——t 期的函数值。

由差分的涵义可知，它与连续时间变量下的导数有相同的意义。

13.1.3.1 差分方程

差分方程是描述离散动态系统的基本方法。著名的蜘蛛网模型就是一差分方程。

设在完全市场经济条件下的某产品，其生产量由制定计划时的市场价格决定，而该产品供应市场的时间需待生产完成之后。因而市场供应量 $Q_s(t+1)$取决于制定计划时的市场价格 $P(t)$，其关系如下：

$$Q_s(t+1)=S[P(t)] \text{ 或 } Q_s(t)=S[P(t-1)]$$

为使 t 时刻市场供需平衡，故有

$$Q_d(t)=Q_s(t) \tag{13 1.20}$$

因此需研究供给函数和需求函数以确定供应量和需求量随时间变化的规律。现假设供给函数和需求函数如下：

$$Q_s(t)=-\gamma+\delta(t-1) \tag{13.1.21}$$

$$Q_d(t)=\alpha\beta(t) \tag{13.1.22}$$

将式(13.1.21)和式(13.1.22)代入式(13.1.20)得

$$\beta P(t)-\delta P(t-1)=\alpha+\gamma$$

将其向后推移一个时期，即

$$P(t+1)+\frac{\delta}{\beta}P(t)=\frac{\alpha+\gamma}{\beta} \tag{13.1.23}$$

这就是反映完全市场经济状况下为维持供需平衡关系的价格变动的一阶差分方程。

经求解(见差分方程求解方法)可得出价格的变动规律为

$$P(t)=\left(P(\theta)-\frac{\alpha+\gamma}{\beta+\delta}\right)\left(-\frac{\delta}{\beta}\right)^t+\frac{\alpha+\gamma}{\beta+\delta} \tag{13.1.24}$$

式中 $P(\theta)$——初始价格。

由式(13.1.24)可见，价格的变动取决于供给率 δ 和需求率 β。当 $\delta>\beta$ 时，价格围绕均衡价格上下波动且呈发散状态(波动幅度越来越大)；当 $\delta<\beta$ 时，价格围绕均衡价格上下波动且呈阻尼振荡形式，最后稳定于均衡价格$\frac{\alpha+\gamma}{\beta+\delta}$。其变化可用图 13.1.3($a$)和 13.1.3($b$)表示，因价格变动线呈蜘蛛网状，故叫蜘蛛网模型。

对线性定常系统来说，其差分方程的一般形式为

$$y(t+n)+a_1y(t+n-1)+a_2y(t+n-2)+\cdots+a_ny(t)=$$

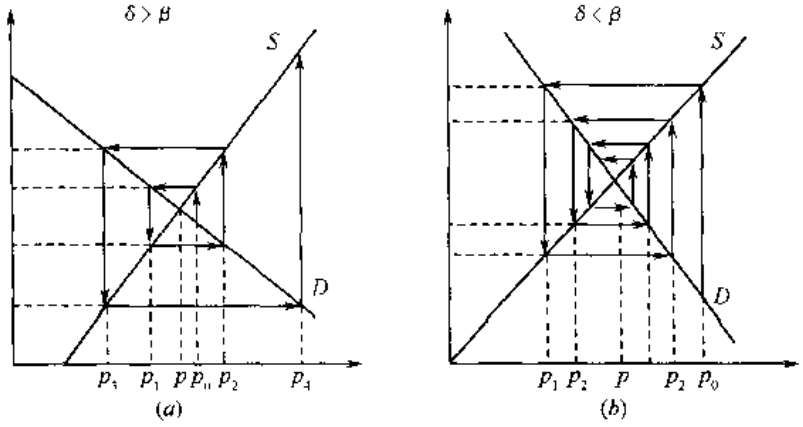


图 13.1.3　蛛网模型

$$bu(t+m)+b_1u(t+m-1)+\cdots+b_nu(t) \tag{13.1.25}$$

13.1.3.2　离散系统的状态方程

离散系统状态方程的一般形式为

$$X(t+1)=AX(t)+Bu(t) \tag{13.1.26}$$

$$Y(t)=CX(t)+Du(t) \tag{13.1.27}$$

式(13.1.26)是差分状态方程,式(13.1.27)是差分输出方程。

与连续系统相似,也可由差分方程推导出状态方程和输出方程,其步骤和方法如下:

(1) 确定差分状态变量;

(2) 确定标准形式的差分状态方程和输出差分方程。

如离散系统差分方程为

$$y(t+3)+2y(t+2)+5y(t+1)+6y(t)=u(t)$$

设:
$$X_1(t)=y(t)$$
$$X_2(t)=X_1(t+1)=y(t+1)$$
$$X_3(t)=X_2(t+1)=y(t+2)$$

为一组状态变量,则

$$X_1(t+1)=X_2(t)$$
$$X_2(t+1)=X_3(t)$$
$$X_3(t+1)=u(t)-6X_1(t)-5X_2(t)-2X_3(t)$$

输出方程为

$$Y(t)=X_1(t)$$

写成矩阵形式为

$$\boldsymbol{X}(t+1)=\boldsymbol{A}\boldsymbol{X}(t)+\boldsymbol{B}u(t)$$
$$\boldsymbol{Y}(t)=\boldsymbol{C}\boldsymbol{X}(t)$$

式中
$$\boldsymbol{X}(t+1)=\begin{bmatrix}X_1(t+1)\\X_2(t+1)\\X_3(t+1)\end{bmatrix},\boldsymbol{A}=\begin{bmatrix}0&1&0\\0&0&1\\-6&-6&-2\end{bmatrix},\boldsymbol{B}=\begin{bmatrix}0\\0\\1\end{bmatrix}$$
$$\boldsymbol{C}=(1\quad 0\quad 0),D=0$$

13.1.4 微分方程组的解法

由上述可见，连续线性定常系统的动态行为均可由一微分方程描述，且均为可化成状态方程的形式。因此研究动态系统运行的规律，必须研究其解法。

微分方程组的求解需要很深的数学知识，这里我们只能简单地说明一下各类方程组解的形式和有关微分方程组解法的一些基本定义和概念，更深的有关微分方程组的解法和知识请参看有关书籍。

1. 纯量方程

$$\dot{X}(t)=aX(t) \tag{13.1.28}$$

由高等数学可知，纯量方程(13.1.28)的解为

$$X(t)=\mathrm{e}^{at}X(0) \tag{13.1.29}$$

$X(0)$是当 $t=0$ 时的初始值。

2. 齐次微分方程组

$$\dot{\boldsymbol{X}}(t)=\boldsymbol{A}\boldsymbol{X}(t) \tag{13.1.30}$$

式中 $\boldsymbol{A}$——常量矩阵；

$\boldsymbol{X}(t)$——向量。

当定义

$$\mathrm{e}^{At}=I+At+\frac{1}{2}A^2t^2+\cdots+\frac{1}{k!}A^kt^k+\cdots \tag{13.1.31}$$

时，微分方程组的解为

$$X(t)=\mathrm{e}^{At}X(0)\text{或 }X(t)=\mathrm{e}^{At}X(t_0) \tag{13.1.32}$$

$X(0)$是当 $t=0$ 时的初始值，$X(t_0)$是当 $t=t_0$ 时的值。

e^{At}可通过定义所给出的级数和求得，它也是一个与 $\boldsymbol{A}$ 阶数相同的矩阵。

根据解的惟一性定理，方程组(13.1.30)有且仅有一个满足初始条件 $t=0$ 或 $t=t_0$ 的解，即式(13.1.32)。

但当 t 为不同值时，由式(13.1.32)表示的方程组(13.1.30)的解也不同，这些解向量所组成矩阵叫方程组的解矩阵。

如果 $X_1,X_2,\cdots,X_n$ 是线性无关的，即当 $C_1,C_2,\cdots,C_n$ 不全为零时，$C_1X_1+C_2X_2+\cdots+C_nX_n\neq 0$，则这些解向量所组成的解矩阵叫基本解矩阵，以 $\boldsymbol{\Phi}(t)$表示：

$$\boldsymbol{\Phi}(t)=(X_1(t),X_2(t),\cdots,X_n(t))$$

若 $\boldsymbol{\Phi}(t)$是方程组 $\dot{\boldsymbol{X}}=\boldsymbol{A}\boldsymbol{X}(t)$的基本解矩阵，则方程组的任意一组解为

$$\varphi(t)=\boldsymbol{\Phi}(t)\cdot\boldsymbol{C} \tag{13.1.33}$$

式中 $\boldsymbol{C}$——任意非零向量。如果 $\boldsymbol{C}$ 是初始条件向量，则 $\varphi(t)$为方程组满足初始条件的解。

当方程组 $\dot{\boldsymbol{X}}=\boldsymbol{A}\boldsymbol{X}(t)$中 $\boldsymbol{A}$ 矩阵各元素是常量时，此系统称为定常系统。则 e^{At}是此定常系统的基本解矩阵，也称转移矩阵：

$$\boldsymbol{\Phi}(t)=\mathrm{e}^{At}\quad\text{且当 }t=0\text{ 时，}\boldsymbol{\Phi}(0)=\boldsymbol{I}$$

从形式上看，方程组 $\dot{\boldsymbol{X}}=\boldsymbol{A}\boldsymbol{X}(t)$的解已经解决，且解为 $\boldsymbol{\Phi}(t)=\mathrm{e}^{At}$。然而 $\boldsymbol{\Phi}(t)$是一矩阵，矩阵的各元素并不是很容易求得的，因此，再给出可直接计算的解的形式：即当且仅当 λ 是 A 的特征值，η 是对应于 λ 的特征向量，也就是 λ 是 $|\lambda I-A|=0$ 的解，η 是对

就于$(\lambda I-A)\cdot\eta=0$的非零解时：

$$\boldsymbol{\Phi}(t)=e^{\lambda t}\cdot\boldsymbol{\eta} \tag{13.1.34}$$

为式(13.1.30)的解。

因此，线性方程组$\dot{\boldsymbol{X}}=\boldsymbol{A}\boldsymbol{X}(t)$，当矩阵$\boldsymbol{A}$有$n$个特征值$\lambda_1,\lambda_2,\cdots,\lambda_n$和相对应的特征向量时，基本解矩阵可表示为

$$\boldsymbol{\varphi}(t)=(e^{\lambda_1 t}\eta_1,e^{\lambda_2 t}\eta_2,\cdots,e^{\lambda_n t}\eta_n)$$

这样一来，方程组满足初始条件的解可用：

$$\boldsymbol{\varphi}(t)=\boldsymbol{\Phi}(t)\cdot C_0=(e^{\lambda_1 t}\eta_1,e^{\lambda_2 t}\eta_2,\cdots,e^{\lambda_n t}\eta_n)\cdot\boldsymbol{C}_0 \tag{13.1.35}$$

或用$\boldsymbol{\varphi}(t)=e^{\lambda t}\cdot\boldsymbol{C}_0$表示，其中$\boldsymbol{C}_0$是初始条件向量。

如果A有重根，即当$(\lambda I-A)\cdot\eta=0$，e为重根个数时，可以证明$(\lambda_i I-A)k\cdot\eta_i=0$，$k\geqslant e$，也就是在计算$e^{At}$时，只取无穷级数的前$e-1$项，而自$e$项以后的所有项都为零，这就是使一无穷级数的计算问题变为有限的形式。

因C可惟一地分解成：$C=\eta_1+\eta_2+\cdots+\eta_k$，方程组的任意解为

$$\varphi(t)=e^{At}\sum_{i=1}^{k}\eta_i=\sum_{i=1}^{k}e^{\lambda_i t}\cdot e^{(A-\lambda_i I)t}\cdot\eta_i=$$

$$\sum_{i=1}^{k}e^{\lambda_i t}\cdot\left[I+t(A-\lambda_i I)+\frac{1}{2}t^2(A-\lambda_i I)^2+\cdots+\frac{1}{(l-1)!}t^{l-1}(A-\lambda I)^{l-1}\right]\eta_i$$

如果$C=C_0$，则得到我们所需要的结论：

$$\varphi(t)=\sum_{i=1}^{k}e^{\lambda_i t}\cdot\left[\sum_{j=0}^{\eta_i-1}\frac{t^j}{j!}(A-\lambda_i I)^j\right]\cdot\eta_i \tag{13.1.36}$$

式中　η_i——λ_i的重数。

下面通过一个例题说明定常动态系统求解的过程。

例：求方程组

$$\begin{aligned}\dot{X}_1&=3X_1-X_2+X_3\\ \dot{X}_2&=2X_1+X_3\\ \dot{X}_3&=X_1-X_2+2X_3\end{aligned}$$

满足初始条件$C_0=\begin{bmatrix}C_{01}\\C_{02}\\C_{03}\end{bmatrix}$的解。

解：

首先求λ：

$$\because\quad A=\begin{bmatrix}3&-1&1\\2&0&1\\1&-1&2\end{bmatrix};\quad P(\lambda)=|\lambda I-A|=0$$

解得$(\lambda-1)(\lambda-2)^2=0$，$\lambda_1=1,n_1=1,\lambda_2=2,n_2=2$(重数为2)。

其次求对应于λ_1和λ_2的特征向量η_1和η_2。

当$\lambda_1=1$时，$(\lambda_I-A)\cdot\eta_1=0$

则
$$\begin{bmatrix}-2 & 1 & -1\\ -2 & 1 & -1\\ -1 & 1 & -1\end{bmatrix}\begin{bmatrix}\eta_{11}\\ \eta_{12}\\ \eta_{13}\end{bmatrix}=0 \quad \eta_1=\begin{bmatrix}0\\ 1\\ 1\end{bmatrix} \quad a=\begin{bmatrix}0\\ a\\ a\end{bmatrix}$$

当 $\lambda_2=2,\eta_2=2$ 时$(\lambda I-A)^2\cdot\eta_2=0$

$$\begin{bmatrix}-1 & 1 & -1\\ -2 & 2 & -1\\ -1 & 1 & 0\end{bmatrix}^2\begin{bmatrix}\eta_{21}\\ \eta_{22}\\ \eta_{23}\end{bmatrix}=0 \quad 即 \begin{bmatrix}0 & 0 & 0\\ -1 & 1 & 0\\ -1 & 1 & 0\end{bmatrix}\begin{bmatrix}\eta_{11}\\ \eta_{12}\\ \eta_{13}\end{bmatrix}=0$$

则 $\eta_{21}=\eta_{22}$，

$$\eta_2=\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}\cdot\gamma+\begin{bmatrix}1\\ 1\\ 0\end{bmatrix}\cdot\beta=\begin{bmatrix}\beta\\ \beta\\ \gamma\end{bmatrix}$$

再次：求出特征向量与初始向量的关系：

如果 $\boldsymbol{C}=\boldsymbol{C}_0=\begin{bmatrix}C_{01}\\ C_{02}\\ C_{03}\end{bmatrix}$，则解即是满足初始向量的解，

由此得：

$$\begin{bmatrix}\beta\\ \alpha+\beta\\ \alpha+\gamma\end{bmatrix}=\begin{bmatrix}C_{01}\\ C_{02}\\ C_{03}\end{bmatrix}$$

即
$$C_{01}=\beta,C_{02}=\alpha+\beta,C_{03}=\alpha+\gamma$$

求出
$$\alpha=-C_{01}+C_{02},\beta=C_{01},\gamma=C_{03}-C_{02}+C_{01}$$

$$\therefore \quad \eta_1=\begin{bmatrix}0\\ a\\ a\end{bmatrix}=\begin{bmatrix}0\\ -C_{01}+C_{02}\\ -C_{01}+C_{02}\end{bmatrix},\quad \eta_2=\begin{bmatrix}\beta\\ \beta\\ \gamma\end{bmatrix}=\begin{bmatrix}C_{01}\\ C_{01}\\ C_{03}-C_{02}+C_{01}\end{bmatrix}$$

最后根据式(13.1.22)求出满足 C_0 的解 $\varphi(t)$。

$$\varphi(t)=\sum_{i=1}^{2}e^{\lambda_i t}\left[\sum_{j=0}^{\eta_i-1}\frac{t^j}{j!}(A-\lambda_i I)^j\right]\eta_i=\sum_{i=1}^{2}e^{\lambda_i t}[I+t(A-\lambda_i I)]\eta_i=$$

$$e^t\cdot\eta_2+e^{2t}[I+t(A-2I)]\eta_2=$$

$$e^t\begin{bmatrix}0\\ -C_{01}+C_{02}\\ -C_{01}+C_{02}\end{bmatrix}+e^{2t}\left\{\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}+t\begin{bmatrix}1 & -1 & 1\\ 2 & -2 & 1\\ 1 & -1 & 0\end{bmatrix}\right\}\cdot\begin{bmatrix}C_{01}\\ C_{01}\\ C_{03}-C_{02}+C_{01}\end{bmatrix}$$

经矩阵运算并整理得

$$\boldsymbol{\varphi}(t)=\begin{bmatrix}(1+t)e^{2t} & -te^{2t} & te^{2t}\\ -e^t+(1+t)e^{2t} & e^t-e^{2t} & te^{2t}\\ -e^t+e^{2t} & e^t-e^{2t} & te^{2t}\end{bmatrix}\cdot\begin{bmatrix}C_{01}\\ C_{02}\\ C_{03}\end{bmatrix}$$

显然，$\boldsymbol{\varphi}(t)$是满足初始条件 $C_0=\begin{bmatrix}C_{01}\\ C_{02}\\ C_{03}\end{bmatrix}$ 的解。

3. 非齐次微分方程组

$$\dot{\boldsymbol{X}}(t)=\boldsymbol{A}X(t)+\boldsymbol{B}u(t) \tag{13.1.37}$$

当 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 矩阵的各元素都是常数的时候，则为定常系统，此非齐次方程组在区间 $[t_0,t]$ 内的解为

$$X(t)=\mathrm{e}^{A(t-t_0)}X(t_0)+\int_0^t \mathrm{e}^{A(t-\tau)}Bu(\tau)\mathrm{d}\tau \tag{13.1.38}$$

当 $t=0$ 时

$$X(t)=\mathrm{e}^{At}X(0)+\int_0^t \mathrm{e}^{A(t-\tau)}Bu(\tau)\mathrm{d}\tau$$

式(13.1.38)又叫状态转移方程。

4. 线性时变系统

$$\dot{\boldsymbol{X}}(t)=\boldsymbol{A}(t)X(t)+\boldsymbol{B}(t)u(t) \tag{13.1.39}$$

当矩阵 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的元素是时间的函数时，此系统称为时变系统。这类问题比较复杂，到目前为止尚得不出解的通式，因此这里就不多讨论了，只给出非齐次时变系统解的形式(并不能按此得出解的结果)：

$$X(t)=\Phi(t,t_0)(t_0)+\int_0^t \Phi(t,\tau)b(\tau)u(\tau)\mathrm{d}\tau \tag{13.1.40}$$

13.1.5 差分状态方程的求解方法

一阶线性定常差分方程求解的最简单方法是迭代法。迭代法的基本思想是在给定初值的条件下，按照差分的递推关系求出其解。

如差分方程 $y(t+1)=y(t)+2$ 的求解过程为：

(1) 令 $t=0,1,2,\cdots,k,\cdots$

(2) 列出对应的递推关系：

$$\begin{aligned}
&y(1)=y(0)+2\\
&y(2)=y(1)+2=y(0)+2+2=y(0)+2\times 2\\
&y(3)=y(2)+2=y(0)+2\times 2+2=y(0)+3\times 2\\
&\cdots\\
&y(k)=y(k-1)+2=\ y(0)+k\times 2\\
&\cdots
\end{aligned}$$

如果给定初始值 $y(0)$，则可推出任一时刻的 $y(t)$。

本例的特点是可推导出求 $y(t)$ 的通式，但复杂系统有时不易求出简单的通式，这时由给定的 $y(0)$，即可求出 $y(1)$，由 $y(1)$ 可求 $y(2)$ …依此方法递推，即可求出任一时刻的 y 值，这用计算机计算求解是很方便的。

13.1.5.1 迭代法求解差分状态方程

差分状态方程的一般形式为

$$\begin{aligned}
X(t+1)&=AX(t)+Bu(t)\\
Y(t)&=CX(t)+Du(t)
\end{aligned} \tag{13.1.41}$$

令方程(13.1.41)中 $t=0,1,2,\cdots,k,\cdots$ 得

$$X(1)=AX(0)+Bu(0)$$

$$X(2)=AX(1)+Bu(1)=A(AX(0)+Bu(0))+Bu(1)=$$
$$A^2X(0)+ABu(0)+Bu(1)$$
$$X(3)=AX(2)+Bu(2)=A(A2X(0)+Abu(0)+Bu(1))+Bu(2)=$$
$$A^3X(0)+A^2Bu(0)+ABu(1)+Bu(2)$$
$$\cdots$$
$$X(k)=A^kX(0)+A^{k-1}Bu(0)+A^{k-2}Bu(1)+\cdots+ABu(k-2)+Bu(k-1)=$$
$$A^kX(0)+\sum_{i=0}^{k-1}A^{k-i-1}Bu(i) \qquad (13.1.42)$$
$$\cdots$$

由式(13.1.42)可见,差分状态方程解的通式由两部分组成,一部分是初始状态$X(0)$的组合;另一部分是控制变量$u(i)(i=1,2,\cdots,k-1)$的组合。这就是说,系统的状态转移是由$\boldsymbol{A}^k$决定的,因此$\boldsymbol{A}^k$是差分状态方程的转移矩阵:

$$\Phi(k)=\boldsymbol{A}^k$$

它满足:
$$\Phi(k+1)=A\cdot\Phi(k)\quad \Phi(0)=I \qquad (13.1.43)$$

利用转移矩阵,差分状态方程的通解又可写成:

$$X(k)=\Phi(k)X(0)+\sum_{i=0}^{k-1}\Phi(k-i-1)Bu(i)=$$
$$\Phi(k)X(0)+\sum_{j=0}^{k-1}\Phi(j)Bu(k-j-1) \qquad (13.1.44)$$

将式(13.1.44)代入式(13.1.41)得输出方程为

$$Y(k)=C\Phi(k)X(0)+C\sum_{j=0}^{k-1}\Phi(j)Bu(k-j-1)+Du(k) \qquad (13.1.45)$$

13.1.5.2 Z变换法求解差分状态方程

在求解微分方程时,引入了拉普拉斯变换。与求解微分方程相对应,在求解差分方程时引入一种新的方法——Z变换。

Z变换又称离散拉普拉斯变换,Z变换的变量定义为

$$Z=e^{sT} \qquad (13.1.46)$$

式中 s——拉普拉斯变换变量;

T——采样周期、即差分时间间隔,我们所研究的情况T为1。

设离散函数$X(t)$,则其Z变换为

$$X(Z)=Z[X(t)]=\sum_{k=0}^{\infty}X(t)Z^{-k} \qquad (13.1.47)$$

Z变换和拉普拉斯变换一样,可通过查表的方法求得离散函数的Z变换。

Z变换也有反变换,其符号为Z^{-1},反变换也可通过查阅反变换表求得。

利用Z变换可求解线性差分方程,其方法与拉普拉斯变换一样,是将差分方程变成以Z为变量的代数方程,求解后再经反变换求出原问题的解。其求解差分方程的过程与用拉氏变换求解连续线性微分方程基本相同。

如差分方程:
$$X(t+2)+3X(t+1)+2X(t)=0$$

进行Z变换得(根据Z变换的性质):

$$Z^2X(Z)-Z^2X(0)-ZX(1)+3ZX(Z)-3ZX(0)+2X(Z)=0$$

如给定 $X(0)=0, X(1)=1$,则

$$Z^2X(Z)-Z+3ZX(Z)+2X(Z)=0$$

$$X(Z)=\frac{Z}{Z^2+3Z+2}=\frac{Z}{Z+1}-\frac{Z}{Z+2}$$

$$Z^{-1}[X(Z)]=Z^{-1}\left[\frac{Z}{Z+1}-\frac{Z}{X+2}\right]$$

查 Z 变换表得

$$X(t)=(-1)^t+(-2)^t$$

用 Z 变换求解离散状态方程的方法与解差分方程的方法相同。对离散状态方程:

$$X(t+1)=AX(t)+Bu(t)$$

进行 Z 变换得

$$ZX(Z)-ZX(0)=AX(Z)+Bu(Z)$$

移项并整理得

$$(ZI-A)X(Z)=ZX(0)+Bu(Z)$$

同时左乘$(ZI-A)^{-1}$得

$$X(Z)=(ZI-A)^{-1}(ZX(0)+Bu(Z))=(ZI-A)^{-1}ZX(0)+(ZI-A)^{-1}Bu(Z)$$

反变换得

$$X(t)=Z^{-1}[(ZI-A)^{-1}Z]X(0)+Z^{-1}[(ZI-A)^{-1}Bu(Z)] \quad (13.1.48)$$

该式即为该差分状态方程的解,其转移矩阵 $\boldsymbol{\Phi}(t)$为

$$\boldsymbol{\Phi}(t)=Z^{-1}[(ZI-A)^{-1}Z]=\boldsymbol{A}^k$$

控制变量的组合项为

$$Z^{-1}[(ZI-A)^{-1}Bu(Z)]=\sum_{i=0}^{k-1}A^{k-i-1}Bu(i)$$

由此可得到离散动态系统的运行规律。

13.2 动态系统最优化理论和方法简介

动态模型可清楚地描述系统的动态行为,这对我们了解系统发展变化的规律性有重要意义。然而按照系统发展变化的规律性,对其实施有效的控制使其达到某项评价指标的最优值对系统的决策有更加重要的意义。因此需研究动态系统的优化理论和方法。

由于动态最优化理论需较深的数学知识,因此本节只对最优化基本理论和方法做简要的介绍。

动态系统最优化又叫最优控制。对连续系统来说,最优控制即是求满足状态方程

$$\dot{X}=f[X(t),u(t),t] \quad (13.2.1)$$

使指标泛函:

$$J=\theta[X(t),t]+\int_0^{t_1}L[X(\tau),u(\tau),\tau]\mathrm{d}\tau \quad (13.2.2)$$

达极小值的最优控制 $u^*(t)$的问题。

对离散系统来说,最优控制是求满足离散状态方程

$$X(t+1)=f[X(t),u(t),t] \tag{13.2.3}$$

使指标泛函：

$$J=\theta[X(N),N]+\sum_{K=0}^{N-1}L[X(K),u(K),K] \tag{13.2.4}$$

达最小的控制序列 $u_K^*,K=0,1,\cdots,N-1$ 的问题。

解决最优控制问题的基本理论和方法有变分法、动态规划法和极大值原理，为了避开泛函分析的数学理论，我们主要介绍动态规划原理和极大值原理。

13.2.1 动态规划

动态规划是一种分步最优化方法，是一多阶段最优决策方法。

所谓多阶段决策是指按时间或空间顺序将决策过程分为若干阶段，若确定出其中任一阶段的状态转移规律，则整个系统的状态转移规律就被惟一地确定，这种寻找最优控制的方法叫分阶段决策法，这种方法的基本理论是贝尔曼提出的最优化原理。

最优化原理不仅可解决离散系统的最优化，而且还可解决连续系统的最优化问题。

13.2.1.1 贝尔曼最优化原理

1. 连续系统最优化原理

对以式(13.2.1)、式(13.2.2)表示的系统，当 $X(t_0)=X_0$ 时，若 $X^*(t)$、$u^*(t)$是最优轨道和最优控制，在$[t_0,t_f]$内任一时刻 t_1，$X^*(t)$、$u^*(t)$仍然是以 $X(t_1)$为初始状态的最优轨道和最优控制。

2. 离散系统最优化原理

对以式(13.2.3)和式(13.2.4)表示的系统，当 $K=0$ 时，$X(K)=X(0)$为初始状态，如果 $X^*(K)$和 $u^*(K)(K=0,1,\cdots,N-1)$是最优状态序列和最优控制序列，则无论 $X(K)(0\leqslant K\leqslant N-1)$及前 K 段控制规律如何，后部子过程的控制规律 $u^*(K),u^*(K+1),u^*(K+2),\cdots,u^*(N-1)$对 $X(K)$来说也必定是最优控制序列。

无论是连续系统还是离散系统，贝尔曼最优化原理都说明：当从 t_0 到 t_f 的最优轨道和最优控制确定以后，无论 $t_0\leqslant t\leqslant t_f$ 是如何转移到$X(t_1)$状态的，一旦 $X(t_1)$已知，则从 t_1 到 t_f 的控制仍是最优控制，也就是说，最优轨道的后一段轨道仍是最优轨道。简单地说，也就是整体最优，则必定后一段最优。这正像甲乙两地间的最短路程确定之后，无论以什么方法、从什么地点出发，凡到达这条最短路程上的任意一点后，只有沿这条路走才最短。因此动态规划就是根据这一原理由后向前递推解决问题的。图 13.2.1 就是该原理的示意图。

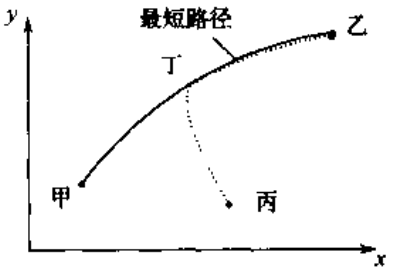


图 13.2.1 连续系统化原理示意图

应用最优化原理可以减少最优决策的计算量，将 N 阶段总体最优决策问题化为 N 个一阶段问题来处理是动态规划方法的本质。

贝尔曼的最优化原理是求解动态规划的必要条件，但不是充分且必要条件。最优化定理才是求解动态规划的充要条件。最优化定理如下：

动态规划的最优性定理：设阶段数为 n 的多阶段决策过程，其阶段编号为 $k=0,1,\cdots,n-1$。

允许策略 $p_{0,n-1}^{*}=(u_0^{*},u_1^{*},\cdots,u_{n-1}^{*})$ 是最优策略的充要条件是对任一个 k，$0<k<n-1$ 和 $s_0\in S_0$ 有

$$V_{0,n-1}(s_0,p_{0,n-1}^{*})=\underset{p_{0,k-1}\in P_{0,k-1}(s_0)}{\text{opt}}\{V_{0,k-1}(s_0,p_{0,k-1})+\underset{p_{k,n-1}\in P_{k,n-1}(\tilde{s}_k)}{\text{opt}}V_{k,n-1}(\tilde{s}_k,p_{k,n-1})\}$$

式中 $p_{0,n-1}=(p_{0,k-1},p_{k,n-1})$，$\tilde{s}_k=T_{k-1}(s_{k-1},u_{k-1})$，它是由给定的初始状态 s_0 和子策略 $p_{0,k-1}$ 所确定的 k 段状态。当 V 是效益函数时，opt 取 max；当 V 是损失函数时，opt 取 min。

二者的区别在于最优化原理是在无论过去处于何种状态和采取何种决策，其后部子过程总是最优的，而最优化定理则在于后部最优必须与以前的状态与决策相关而构成整体最优。

13.2.1.2 动态规划方程

根据最优化原理可推导出求解动态系统最优控制的一般方法。现分离散系统和连续系统两种情况介绍。

1. 离散系统动态规划方程

设可分段离散系统状态方程为

$$X(K+1)=f[X(K),u(K),K] \tag{13.2.5}$$

式中 f——n 维向量函数；

$X(K+1)$——n 维状态向量在 $K+1$ 时刻的向量值；

$u(K)$——m 维控制向量在 K 时刻的向量值；

$\boldsymbol{X}(0)$——为 $K=0$ 时的初始向量。

则离散系统最优控制问题的数学模型为

$$\min\quad J=J_0[X(N)]+\sum_{K=0}^{N-1}L[X(K),u(K)] \tag{13.2.6}$$

$$\text{s.t.}\quad g[X(K),u(K)]\leqslant 0 \tag{13.2.7}$$

$$X(K+1)=f[X(K),u(K)] \tag{13.2.8}$$

式中 $J_0[X(N)]$——$X(N)$ 状态的指标泛函值。

现欲求最优控制序列 $u^{*}(0),u^{*}(1),\cdots,u^{*}(N-1)$ 使系统在约束限制下由初始状态 $X(0)$ 转移到终止状态 $X(N)$ 的指标泛函 $j_N\to\min$。

根据最优化原理现将寻找 N 阶段总体最优决策变为 N 个一阶段的最优决策问题来处理。

如果第一阶段选择该阶段的最优决策 $u^{*}(0)$ 使系统由 $X(0)$ 状态转移到 $X^{*}(1)$ 状态，即 $X^{*}(1)=f[X(0),u^{*}(0)]$，则第二阶段需选择该阶段的最优决策 $u^{*}(1)$ 使系统由 $X^{*}(1)$ 状态转移到 $X^{*}(2)=f[X(1),u^{*}(1)]$，最后第 N 阶段需选择该阶段的最优决策 $u^{*}(N-1)$ 使系统由 $X^{*}(N-1)$ 转移到 $X^{*}(N)=f[X(N-1),u^{*}(N-1)]$，这则是传统方法。

根据最优化原理，无论该系统第一阶段 $X(0)$、$u(0)$ 如何选取，第二阶段以后的控制序列 $\{u(K)\}(K=1,2,\cdots,N-1)$ 一定是 $N-1$ 阶段的最优控制序列，它必定使泛函指标

的后 $N-1$ 项与 $J_0[X(N)]$之和最小，这样就在求 N 阶段的指标泛函时，嵌入了后部子过程的最优指标泛函，就把 N 阶段决策问题化为 N 个单一阶段的决策，其嵌入过程就是递推求解的过程，现简述如下：

由初始状态 $X(0)$转移到终止状态 $X(N)$的指标泛函可按下式计算：

$$J_N^*[X(0)] = \min_{u(0)}\{L[X(0),u(0)] + J_{N-1}^*[X(1)]\} \tag{13.2.9}$$
$$X(1) = f[X(0),u(0)]$$

其中，$J_N^*[X(0)]$为 N 阶段总体决策的最优指标泛函，它是由初始状态 $X(0)$向 $X(1)$状态转移的控制向量 $u(0)$和后部子过程指标泛函决定的，如果 $J_{N-1}^*[X(1)]$已知，则选择 $u^*(0)$即可得出总体最优控制中的 $u^*(0)$，这是一单一阶段决策问题。

再根据最优化原理：

$$J_{N-1}^*[X(1)] = \min\{L[X(1),u(1)] + J_{N-2}^*[X(2)]\} \tag{13.2.10}$$
$$X(2) = f[X(1),u(1)]$$

按上述分析，可通过选择 $u^*(1)$而得出 $J_{N-1}^*[X(1)]$，依此类推，得出各阶段中选择 $u(K)$的通式

$$J_{N-K}^*[X(K)] = \min_{u(k)}\{L[X(K),u(K)] + J_{N-(k+1)}^*[X(k+1)]\} \tag{13.2.11}$$
$$X(K) = f[X(K-1),u(K-1)]$$

当 $K=N-2$ 和 $K=N-1$ 时，得

$$J_2^*[X(N-2)] = \min_{u(N-2)}\{L[X(N-2),u(N-2)] + J_1^*[X(N-1)]\} \tag{13.2.12}$$
$$X(N-1) = f[X(N-2),u(N-2)]$$

$$J_1^*[X(N-1)] = \min_{u(N-1)}\{L[X(N-1),u(N-1)] + J_0^*[X(N)]\} \tag{13.2.13}$$
$$X(N) = f[X(N-1),u(N-1)]$$

由上述可知，从最后一阶段开始，因 $J_0[X(0)]$已知，可由 $u(N-1)$中确定该阶段的最优控制 $u^*(N-1)$而得到 $J_1^*[X(N-1)]$，又因 $J_1^*[X(N-1)]$已知，又可按式(13.2.12)在 $u(N-2)$中选择最优控制 $u^*(N-2)$，得到 $J_2^*[X(N-2)]\cdots$，最后，在第 N 步由于 $J_{N-1}^*[X(1)]$已知，可由 $u(0)$中选择最优控制 $u^*(0)$确定 $J_N^*[X(0)]$，这样 $u^*(N),u^*(N-1),\cdots,u^*(2),u^*(1),u^*(0)$，分别由各阶段单独得出，所构成的控制序列使指标泛函达最小，从而完成了多阶段最优决策问题。式(13.2.11)就是离散系统的贝尔曼动态规划方程，是求解离散系统动态最优模型的递推公式。

如某水泥厂第 K 季度的生产量为 $X(K)$，生产费用为 $aX^2(K)$，其各季度订货量为 $S(K)$，则第 K 季度库存的增加量为 $X(K)-S(K)=\Delta Z(K)$，若单位产品的存储费用为 b，则第 $K+1$ 季度的库存将由下式决定：

$$Z(K+1)=Z(K)+\Delta Z(K)=Z(K)+X(K)-S(K)$$

若使一年中生产量等于销售量(订货量)，即年初和年末的库存均为零，则有：

$$Z(1)=0 \qquad Z(5)^{①}=0$$

现欲确定全年的最优生产计划，使之既能完成全年订货任务，又使生产费用和存贮费用之和最小。

该问题的模型如下：

$$\begin{aligned} \min \quad & J=\sum_{K-1}^{N}[aX^2(K)+bZ(K)] \\ \text{s.t.} \quad & Z(K+1)=Z(K)+X(K)-S(K) \\ & Z(1)=0 \qquad Z(5)=0 \end{aligned} \tag{13.2.14}$$

其中状态变量为 $Z(K)$，即库存量，控制变量为库存增量 $\Delta Z(K)=X(K)-S(K)$，初始状态 $Z(1)=0$，终止状态 $Z(5)=0$。式(13.2.14)即为生产库存系统的最优控制数学模型。

现假设 $a=0.005$，$b=1$，各季度订货量见表 13.2.1，则指标泛函的具体形式为

$$J=\sum_{k-1}^{4}[0.005X^2(K)+Z(K)]$$

表 13.2.1　各季度订货量

季　度	1	2	3	4
订货量 $S(K)$	600	700	500	1200

按照贝尔曼动态规划方程，其求解应从第四季度开始，确定该阶段的最优控制，其通式为：

$$J^*_{N-K+1}[Z(K)]=\min_{\Delta Z(K)}\{0.005X^2(K)+Z(K)+J^*_{N-(K+1)+1}[Z(K+1)]\} \tag{13.2.15}$$

$$Z(K+1)=Z(K)+X(K)-S(K)$$

(由于该问题指标泛函中 $K=1,2,3,4$，故通式中 J^* 的脚标为 $N-K+1$ 和 $N-(K+1)+1$)

(1) 当 $K=4$ 时，式(13.2.15)变为

$$J_1^*[Z(4)]=\min_{\Delta Z(4)}\{0.005X^2(4)+Z(4)+J_0^*[Z(5)]\}$$

$$Z(5)=Z(4)+X(4)-S(4)$$

将已知数据代入得最优控制 $\Delta Z=-Z(4)$，即

$$X^*(4)=1200-Z(4)$$

将其代入 $J_1^*[Z(4)]$ 得

$$J_1^*[Z(4)]=\min_{\Delta Z(4)}\{0.005[1200-Z(4)]^2+Z(4)\}=$$

$$0.005Z^2(4)-11Z(4)+7200 \tag{13.2.16}$$

(2)当 $K=3$ 时，式(13.2.15)变为

$$J_2^*[Z(3)]=\min_{\Delta Z(3)}\{0.005X^2(3)+Z(3)+J_1^*[Z(4)]\}=$$

$$\min_{\Delta Z(3)}\{0.005X^2(3)+Z(3)+0.005Z^2(4)-11Z(4)-7200\}$$

$$Z(4)=Z(3)+X(3)-S(3)=Z(3)+X(3)-500$$

① Z(5)为下年年初，本年度第四季末。

将 $Z(4)$代入 $J_2^*[Z(3)]$得

$$J_2^*[Z(3)] = \min_{\Delta Z(3)}\{0.005X^2(3) + Z(3) + 0.005[Z(3) + X(3) - 500]^2 - 11[Z(3) + X(3) - 500] + 7200\}$$

选择适当的 $X(3)$可使右边大括号内函数取最小值,故应用微分方法可求出 $X^*(3)$:

$$\frac{d\{\cdot\}}{dX(3)} = 0.02X(3) - 16 + 0.01Z(3) = 0 \quad ①$$

$$X^*(3) = 800 - 0.5Z^2(3)$$

代入 $J_2^*[X(3)]$得

$$J_2^*[Z(3)] = 7550 - 7Z(3) + 0.0025Z(3)$$

(3)当 $K=2$ 时,式(13.2.15)变为

$$J_3^*[Z(2)] = \min_{\Delta Z(2)}\{0.005X^2(2) + Z(2) + J_2^*[Z(3)]\} = \min_{\Delta Z(2)}\{0.005X^2(2) + Z(2) + 7550 - 7Z(3) + 0.005Z^2(3)\}$$

$$Z(3) = Z(2) + X(2) - S(2) = Z(2) + X(2) - 700$$

$$\therefore \quad J_3^*[Z(2)] = \min_{\Delta Z(2)}\{0.005X^2(2) + Z(2) + 7550 - 7[Z(2) + X(2) - 700] + 0.025[Z(2) + X(2) - 700]^2\}$$

同 $K=3$ 时一样,令 $\frac{d\{\cdot\}}{dX(2)} = 0$ 得到该阶段的最优控制 $X^*(2)$:

$$\frac{d\{\cdot\}}{dX(2)} = 0.015X(2) + 0.005Z(2) - 10.5 = 0$$

$$X^*(2) = 700 - \frac{1}{3}Z(2)$$

代入 $J_3^*[Z(2)]$得

$$J_3^*[Z(2)] = 10000 - 6Z(2) + \frac{0.005}{3}Z^2(2)$$

(4)当 $K=1$,式(13.2.15)变为

$$J_4^*[Z(1)] = \min_{\Delta Z(1)}\{0.005X^2(1) + Z(1) + J_3^*[Z(2)]\} = \min_{\Delta Z(1)}\{0.005X^2(1) + 10000 - 6Z(2) + \frac{0.005}{3}Z^2(2)\}$$

$$Z(2) = Z(1) + X(1) - S(1) = Z(1) + X(1) - 600 = X(1) - 600$$

代入 $J_4^*[Z(1)]$得

$$J_4^*[Z(1)] = \min_{\Delta Z(1)}\{0.005X^2(1) + 10000 - 6[X(1) - 600] + \frac{0.005}{3}[X(1) - 600]^2\}$$

$$\frac{d\{\cdot\}}{dX(1)} = 0.01X(1) - 6 + \frac{0.005}{3}[X(1) - 600] = \frac{0.04}{3}X(1) - 8 = 0$$

$$X^*(1) = 600(\mathrm{m}^3)$$

代入 $J_4^*[Z(1)] = 11800$

由状态方程 $Z(t+1) = Z(t) + X(t) - S(t)$可求出 $Z(K)$,

$$\because Z^*(1) = 0, X^*(1) = 600(\mathrm{m}^3)$$

---

① $\{\cdot\}$表示 $J_2^*[Z(3)]$式中大括号内的表达式.

$$\therefore Z^*(2)=Z^*(1)+X^*(1)-S(1)=0+600-600=0$$

$$X^*(2)=700-\frac{1}{3}Z^*(2)=700-0=700(\mathrm{m}^3)$$

$$Z^*(3)=Z^*(2)+X^*(2)-S(2)=0+700-700=0$$

$$X^*(3)=800-0.5Z^*(3)=800(\mathrm{m}^3)$$

$$Z^*(4)=Z^*(3)+X^*(3)-S(3)=0+800-500=300(\mathrm{m}^3)$$

$$X^*(4)=1200-Z^*(4)=1200-300=900(\mathrm{m}^3)$$

因此得出最优生产计划为：

第一季度生产 $600\mathrm{m}^3$、第二季度生产 $700\mathrm{m}^3$、第三季度生产 $800\mathrm{m}^3$(库存 $300\mathrm{m}^3$)、第四季度生产 $900\mathrm{m}^3$，可使总费用最低为 11800(百元)。

如果各季度均按订货量安排生产，虽然可减少产品的存货费用，但总费用却增加 900(百元)，因这时生产总费用为：$\sum_{k=1}^{4}0.005X^2(K)=12700$(百元)。

由上例可见，利用动态最优化原理可使多阶段问题化为单一阶段问题解决且可使计算简化。

2. 连续系统动态规划方程

连续动态规划方程又叫哈密顿—雅可比—贝尔曼方程，它是由贝尔曼最优化原理推导出的求最优指标泛函的偏微分方程。

设连续系统的状态方程为

$$\dot{X}=f[\boldsymbol{X}(t),u(t),t] \tag{13.2.17}$$

式中 $\boldsymbol{X}(t)$——n 维状态向量，当 $t_0=0$ 时，$X(t_0)=X(0)$为系统的初始状态，则连续系统最优控制的最优化模型为

$$\begin{aligned}&\min \quad J=\theta[X(t_f),t_f]+\int_{t_0}^{t_f}L[X(\tau),u(\tau),\tau]\mathrm{d}\tau\\&\text{s.t.} \quad \dot{X}=f[X(t),u(t),t]\\&\qquad X(t_0)=X(0)\end{aligned} \tag{13.2.18}$$

式中 $\theta[X(t_f),t_f]$——仅取决于最终状态的目标泛函。

仿照离散系统

$$J^*[X(0)]=\min_u\int_{t_0}^{t_f}L[X(\tau),u(\tau),\tau]\mathrm{d}\tau$$

对$[t_0,t_f]$中任一时刻 t，由其 t_f 时的最优控制 $u^*(t)$可使后部子过程的指标泛函达最小，即

$$J^*[X(t)]=\int_t^{t_f}L[X(\tau),u^*(\tau),\tau]\mathrm{d}t\to\min$$

因此，

$$J^*[X(0)]=\min_u\left[\int_{t_0}^{t}L[X(\tau),u(\tau),\tau]\mathrm{d}\tau+\int_t^{t_f}L[X(\tau),u^*(\tau),\tau]\mathrm{d}\tau\right]=$$

$$\min_u\left\{\int_{t_0}^{t}L[X(\tau),u(\tau),\tau]\mathrm{d}\tau+J^*[X(t)]\right\} \tag{13.2.19}$$

如果 $J^*[X(t)]$已知，则 $J^*[X(0)]$将仅由 $\int_{t_0}^{t}L[X(\tau),u(\tau),\tau]\mathrm{d}\tau$ 决定，也就是从

中选择一最优控制 $u^*[X(0)]$即可实现指标泛函的极小值。故关键需知道求 $J^*[X(t)]$的方法，因 $J^*[X(t)]=\int_t^{t_f}L[X(\tau),u^*(\tau),\tau]\mathrm{d}\tau$。

根据最优化原理再将其分为两段：t 到 $t+\Delta t$ 和 $t+\Delta t$ 到 t_f。这样 $J^*[X(t)]$又可写成：

$$J^*[X(t)]=\min_u\left\{\int_{t_0}^{t+\Delta t}L[X(\tau),u(\tau),\tau]\mathrm{d}\tau+\int_{t+\Delta t}^{t_f}L[X(\tau),u^*(\tau),\tau]\mathrm{d}\tau\right\}=$$

$$\min_u\left\{\int_t^{t+\Delta t}L[X(\tau),u(\tau),\tau]\mathrm{d}\tau+J^*[X(t+\Delta t)]\right\}\tag{13.2.20}$$

因 $J^*[X(t)]$既是 $X(t)$的函数，又是 t 的函数，故将 $J^*[X(t+\Delta t)]$在 $J^*[X(t)]$处按泰勒级数展开，略去无穷小项后得

$$J^*[X(t+\Delta t)]=J^*[X(t)]+\frac{\partial J^*[X(t)]}{\partial t}\cdot\Delta t+\left[\frac{\partial J^*[X(t)]}{\partial X(t)}\right]^{\mathrm{T}}\frac{\mathrm{d}X}{\mathrm{d}t}\cdot\Delta t=$$

$$J^*[X(t)]+\frac{\partial J^*[X(t)]}{\partial t}\cdot\Delta t+\left[\frac{\partial J^*[X(t)]}{\partial X(t)}\right]^{\mathrm{T}}f[X(t),u(t),t]\cdot\Delta t\tag{13.2.21}$$

又∵ $\int_t^{t+\Delta t}L[X(\tau),u(\tau),\tau]\mathrm{d}\tau=L[X(t+a\Delta t),u(t+a\Delta t),t+a\Delta t]\cdot\Delta t$

$0\leqslant a\leqslant 1$，当 Δt 较小时：

$$\int_t^{t+\theta}L[X(\tau),u(\tau),\tau]\mathrm{d}t=L[X(t),u(t),t]\cdot\Delta t\tag{13.2.22}$$

将式(13.2.21)和式(13.2.22)代入式(13.2.20)得

$$J^*[X(t)]=\min_u\left\{L[X(t),u(t),t]\cdot\Delta t+J^*[X(t)]+\frac{\partial J^*[X(t)]}{\mathrm{d}t}\cdot\Delta t\right.$$

$$\left.+\left[\frac{\partial J^*[X(t)]}{\partial X(t)}\right]^{\mathrm{T}}\cdot f[X(t),u(t),t]\cdot\Delta t\right\}$$

因 $J^*[X(t)]$和$\frac{\partial J^*[X(t)]}{\partial t}\cdot\Delta t$ 与选择该阶段的控制 $u(t)$无关，故可提到括号外，然后两边消掉 $J^*[X(t)]$，同除 Δt 并令 $\Delta t\to 0$，得到

$$-\frac{\partial J^*[X(t)]}{\partial t}=\min_u\left\{L[X(t),u(t)t]+\left[\frac{\partial J^*[X(t)]}{\partial X(t)}\right]^2\cdot f[X(t),u(t)t]\right\}\tag{13.2.23}$$

该式则是连续系统的动态规划方程，叫贝尔曼方程。

令

$$\frac{\partial J^*[X(t)]}{\partial t}=\lambda,L[X(t),u(t)t]+\lambda^{\mathrm{T}}f[X(t),u(t)t]=H(X,u,\lambda,t)$$

则式(13.2.23)可表示为

$$\frac{\partial J^*[X(t)]}{\partial^* t}=\min|H(X,u,\lambda,t)|$$

这样就把求解指标泛函极小值的问题化成求函数 $H[X(t),u(t),\lambda(t),t]$的最小值问题。

现将 $$H(X,u,\lambda,t)=L[X(t),u(t)t]+\lambda^{\mathrm{T}}f[X(t),u(t)t]\tag{13.2.24}$$

称之为哈密顿函数，它是求连续动态系统最优控制的基础。由哈密顿函数是 $u(t)$和 $X(t)$的函数，故对其求导并令其为零即可得出 $u^*(t)$和 $X^*(t)$，它们均为 λ 的函数，将其代入式(13.2.23)得

$$\frac{\partial J^*[X(t)]}{\partial t} + L\{X^*(t),u^*(t),t\} + \lambda^{\mathrm{T}} f[X^*(t),u^*(t),t] = 0 \tag{13.2.25}$$

该式即为哈密顿—雅可比—贝尔曼方程，它是求解连续系统动态规划的基本方程。如果该方程可解，则可求出 $u^*(t)$，进而可确定 $X^*(t)$和 $J^*[X(t)]$。

综上所述，连续系统动态规划问题的求解步骤为：

(1) 根据指标泛函和状态方程，构造哈密顿函数：

$$H(X,u,\lambda,t) = L[X(t),u(t),t] + \lambda^{\mathrm{T}} f[X(t),u(t),t]$$

其构造法类似于求解非线性规划时引入拉格朗日乘子一样，它是由指标泛函和状态方程借助于 λ^{T} 构成的。因该方程中 $L(X,u,t)$和 $f(X,u,t)$已知，故可求 $u^*(t)$和 $X^*(t)$，它们均是 λ 的函数。

(2) 将求得的 $u^*(t)$和 $X^*(t)$代入哈密顿—雅可比—贝尔曼方程：

$$\frac{\partial J^*[X(t)]}{\partial t} + L\{X^*(t),u^*(t),t\} + \lambda^{\mathrm{T}} f[X^*(t),u^*(t),t] = 0$$

因式中 $\lambda = \dfrac{\partial J^*[X(t)]}{\partial t}$，且 $u^*(t)$和 $X^*(t)$均是 $\dfrac{\partial J^*[X(t)]}{\partial t}$的函数，因此由该偏微分方程可解出 $J[X(t)]$。

(3) 将求得的 $J[X(t)]$代入第一步求出的 $u^*(t)$和 $X^*(t)$中即得到 $u^*(t)$和 $X^*(t)$。

(4) 利用初始条件，指标泛函的极小值可求出。

求解连续系统的动态规划问题的关键是求解偏微分方程——哈密顿—雅可比—贝尔曼方程，该方程用解析法一般很难求解，只能借助于计算机采用数值解法。

13.2.2 极大值原理

极大值原理是苏联庞特里亚金等人于 20 世纪 50 年代初提出的解决最优控制问题的方法，因此也称庞特里亚金极大值原理。

连续系统的庞特里亚金极大值原理为

对以：

$$\min \quad J = \int_{t_0}^{t_f} L[X(t),u(t),t]\mathrm{d}t \tag{13.2.26}$$

为目标泛函，以

$$\dot{X}(t) = f[X(t),u(t),t] \tag{13.2.27}$$

为状态方程的动态控制系统，在初始状态和终止状态为

$$X(t_0) = X(0), N[X(t_f,t_f)] = 0 \tag{13.2.28}$$

时，从所有容许控制中确定一最优控制 $u^*(t)$使指标泛函达极小的必要条件是

对区间$[t_0,t_f]$内每个 t，以 u 为变量的哈密顿函数 H：

$$H(X(t),u(t),\lambda(t),t) = -L[X(t),u(t)t] + \lambda^{\mathrm{T}} f[X(t),u(t)t]$$

必在 $u=u(t)$处取极大值，其中 $\lambda(t)$为伴随变量，它是伴随方程

$$\dot{\lambda}=-\frac{\partial H}{\partial X}$$

的解。

因该定理的中心是求哈密顿函数 H 的极大值，故称极大值原理。

如将哈密顿函数改写成类似于非线性规划中引入拉格朗日乘子的形式，则可将上述必要条件改写成：

对区间$[t_0,t_f]$内每个 t，以 u 为变量的哈密顿函数：

$$H[X(t),u(t),\lambda(t),t]=L[X(t),u(t)t]+\lambda^{\mathrm{T}}f[X(t),u(t)t] \tag{13.2.29}$$

必在 $u=u(t)$处取极小值，其中，$\lambda(t)$为伴随变量，它是伴随方程：

$$\dot{\lambda}=-\frac{\partial H}{\partial X}$$

的解。

因该定理的中心问题是求哈密顿函数的极小值，故也叫极小值原理，它与极大值原理是一致的。为了便于记忆，使指标泛函求极小和哈密顿函数求极小一致，故采用后者较方便，通常也把它叫做极大值原理，而不叫极小值原理。

对线性定常系统来说，庞特里亚金极大值原理是使指标泛函达极小的充分且必要条件。

根据极大值原理，求解最优控制问题的步骤为：

(1) 根据指标泛函和状态方程建立哈密顿函数：

$$H[X(t),u(t),\lambda(t),t]=L[X(t),u(t)t]+\lambda^{\mathrm{T}}f[X(t),u(t)t]$$

(2) 求哈密顿函数极小值，即令

$$\frac{\partial H}{\partial X}=0,\text{得出 } u^*[X(t),\lambda(t),t] \tag{13.2.30}$$

(3) 求解伴随方程：

$$\dot{\lambda}=-\frac{\partial H}{\partial X}$$

$$\dot{X}=\frac{\partial H}{\partial \lambda}$$

考虑边界条件，可得出 $X^*(t)$和 $\lambda^*(t)$。

(4) 将 $X^*(t)$、$\lambda^*(t)$代入 $u^*[X(t),\lambda(t),t]$，则得到最优控制 $u^*[X^*(t),\lambda^*(t),t]$。

(5) 将 $X^*(t)$，$u^*(t)$代入目标泛函，则得到最佳目标值。

值得指出的是，上述步骤是针对一般情况而言的，具体解决问题时，应根据模型的情况灵活解决，现以最优积累率模型为例说明极大值原理的应用。

在进行国民经济综合平衡时，确定合适的积累率是一重要指标，它涉及到国民经济的发展、社会再生产的规模及人民生活水平的提高。总结建国以来进行国民经济综合平衡的经验，有些专家指出25%左右为宜，其依据是多年来，当积累率在25%左右时，国民收入和消费水平都增长较快；当积累率较高时，降低了经济发展速度，影响了人民生活水平的提高。这是根据历史的经验得到的结论，当然也有专家持不同看法。然而合理的积累

率应该是多少？从理论上可否推算出来则需在定性分析的基础上，借助于经济数学模型，尤其是考虑时间因素的动态模型来解决。

最优积累率模型是龚德思等同志建立的，其模型的建立方法及求解过程如下：

1．该问题的状态方程和约束条件

国民收入分为积累基金和消费基金两大部分。积累率是指积累基金与国民收入之比。在社会主义计划经济中，积累率可以有计划地进行控制，这是本模型把积累率取作控制变量的经济背景。

设：$x(t)$为时刻 t 的国民收入，状态变量；

$u(t)$为时刻 t 的积累率，控制变量；

$\alpha(t)$为时刻 t 的积累效果系数，即单位积累基金在单位时间内产生的国民收入的增长额。

根据上述定义，可得

$$\frac{\mathrm{d}x}{\mathrm{d}t}=\dot{x}=\alpha(t)x(t)u(t) \tag{13.2.31}$$

式(13.2.31)的经济含义是国民收入乘以积累率为积累基金，再乘以积累效果系数，就是单位时间内国民收入的增长额。

积累效果系数 $\alpha(t)$受到积累基金使用方向、劳动生产率和管理水平等一系列因素的影响。但是，这些因素本身又受到积累率的制约，因此，积累率是影响积累效果系数的主要因素。根据历年的统计资料分析，我国的积累效果系数 $\alpha(t)$与积累率 $u(t)$近似于直线关系，且直线斜率为负，由统计资料，得 $\alpha(t)$的近似关系为

$$\alpha(t)=a\alpha-bu(t) \tag{13.2.32}$$

式中：$a=0.958$、$b=1.897$，代入状态方程，得

$$\dot{x}=[au(t)-bu^2(t)]x(t) \tag{13.2.33}$$

设初始状态的国民收入为 x_0，即

$$x(t)|_{t_0}=x_0 \tag{13.2.34}$$

因为积累率是积累基金占国民收入的比例，所以有

$$0\leqslant u(t)\leqslant 1 \tag{13.2.35}$$

根据问题的经济性质，国民收入应为正增长，即式(13.2.31)应非负，则由积累效果系数 $\alpha(t)$的定义有：

$$\alpha(t)\geqslant 0$$

于是由式(13.2.32)得

$$0\leqslant u(t)\leqslant\frac{a}{b}=0.505$$

式(13.2.33)是状态方程，式(13.2.34)是初始条件，式(13.2.35)是控制变量的约束条件。

2．性能指标及解的结果

根据经济学界对这个问题的讨论，可以在不同意义上提出不同的性能指标。

第一类指标，主要考虑国民收入，如要求在给定的计划期$[0,T]$内，国民收入增长总额最大，这种情况下的性能指标是单位时间内国民收和增长额的积分：

$$\max \quad J=\int_0^T x(t)\,\mathrm{d}t=\max\int_0^T a(t)u(t)x(t)\mathrm{d}t$$

或写成

$$\max \quad J=\int_0^T [au(t)-bu^2(t)]x(t)\mathrm{d}t \tag{13.2.36}$$

利用庞特里亚金极大值原理,构造哈密顿函数为

$$H=(au-bu^2)x+\lambda(au-bu^2)x \tag{13.2.37}$$

式中 λ——伴随变量。

由正则方程

$$\dot{\lambda}=-\frac{\partial H}{\partial x}=-(au-bu^2)-\lambda(au-bu^2) \tag{13.2.38}$$

由终端时间 T 固定,状态 x 自由的条件

$$\lambda(t)|_{t=T}=0 \tag{13.2.39}$$

$u^*(t)$若为最优控制,由极大值原理的必要条件

$$H(x(t),u^*(t),\lambda^*(t))=\max_{u\in U}H(x(t),u^*(t),\lambda^*(t))$$

式中 U——容许控制集合;

$x^*(t)$和 $\lambda(t)$——状态方程式(13.2.33)、式(13.2.34)和伴随方程式(13.2.38)、式(13.2.39)对应于 $u^*(t)$的解。

对任一固定的 $t\in[0,T]$,将 $H(x(t),u,\lambda^*(t))$视为 u 的单变量函数,有

$$\frac{\partial H}{\partial u}=(x+\lambda x)(a-2bu)=0$$

因此

$$u=\frac{a}{2b}=\frac{0.958}{2\times 1.897}=0.253$$

因为 $u=0.253$ 在控制变量约束 $0\leqslant u(t)\leqslant\frac{a}{b}=0.505$ 范围内,故认为最优积累率应为 0.253。将 $u=\frac{a}{2b}$代入式(13.2.33):

$$\dot{x}(t)=\frac{a^2}{4b}x(t) \tag{13.2.40}$$

解式(13.2.40),并以式(13.2.34)为初始条件,得

$$x(t)=x_0\mathrm{e}^{\frac{a^2}{4b}t}\qquad t\in[0,T] \tag{13.2.41}$$

即国民收入将按指数增长。$[0,T]$内国民收入增长总额最大,为 $J=x_0(\mathrm{e}^{\frac{a^2}{4b}t}-1)$。

第二类指标,要求在给定计划期$[0,T]$内,积累基金增长总额最大,以对国民经济的长远发展有利,这种情况下的性能指标为式(13.2.31),它是单位时间内国民收入增长额乘以积累率 $u(t)$之后再积分,即

$$\max \quad J=\int_0^T u(t)x(t)\mathrm{d}t=\int_0^T a(t)u^2(t)x(t)\mathrm{d}t$$

或写成

$$J=\int_0^T [au^2(t)-bu^3(t)]x(t)\mathrm{d}t \tag{13.2.42}$$

式中 T——已知末时刻，积分结果表示在计划期内积累基金增长总额。

以式(13.2.42)作为性能指标，以式(13.2.33)～式(13.2.35)作为约束条件，就可以解出最优积累率。

我们直接给出结果：

$$u^*(t)=\frac{1}{3b}(a-b\lambda+\sqrt{a^2+ab\lambda+b^2\lambda^2}) \tag{13.2.43}$$

式中 λ 为该问题的伴随变量，满足如下的伴随方程和终值条件

$$\dot{\lambda}=-(u^*+\lambda)[au^*-b(u^*)^2] \tag{13.2.44}$$

$$\lambda(T)=0 \tag{13.2.45}$$

而且，由式(13.2.43)～式(13.2.45)确定的最优积累率 $u^*(t)$ 为时间 t 的单调递增函数，如图 13.2.2 所示，并且满足不等式：

$$\frac{a}{2b}<u^*(t)\leqslant\frac{2a}{3b}=u_{上}\quad t\in[0,T]$$

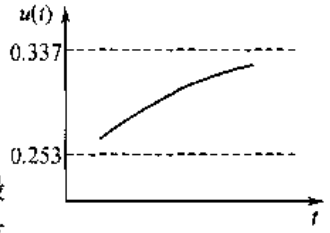


图 13.2.2 积累基金增长最快时的 $u^*(t)$

本例可算出

$0.253<u^*(t)\leqslant 0.337$

上述结果表明，为使计划期[0，T]内积累基金增长总额最大，最优积累率应按图 13.2.2 的曲线，在限定的范围内逐年有所增加，但不能超过上限 0.337。这就说明，即使单纯从多积累的角度出发，积累率也不能任意提高。因此，我们将

$$u_{上}=\frac{2a}{3b} \tag{13.2.46}$$

称为合理积累率的上限。

第三类指标，在给定的计划期[0，T]内，要求消费基金增长总额最大，以最大限度地提高人民的物质文化生活水平。消费基金等于国民收和减去积累基金。在这类指标下的性能指标可写成

$$\max\quad J=\int_0^T(1-u)(au-bu^2)x\mathrm{d}t \tag{13.2.47}$$

式中 T——计划期末，这样表示的性能指标可以描述消费基金增长总额。

以式(13.2.47)作为性能指标，以式(13.2.33)～式(13.2.35)作为约束条件，利用庞特里亚金极大值原理，就可解出最优积累率，我们直接给出结果

$$u^*(t)=\frac{1}{3b}[a+b(1+\lambda)-\sqrt{a^2-ab(1+\lambda)+b^2(1+\lambda)^2}] \tag{13.2.48}$$

式中 λ——伴随变量，满足如下的伴随方程和终值条件

$$\dot{\lambda}=-(1-u^*+\lambda)[au^*-b(u^*)^2] \tag{13.2.49}$$

$$\lambda(T)=0 \tag{13.2.50}$$

由式(13.2.48)～式(13.2.50)确定的最优积累率 $u^*(t)$ 为时间 t 的单调递减函数，并且满足不等式

$$u_{下}=\frac{1}{3b}[a+b-\sqrt{a^2-ab+b^2}]\leqslant\frac{a}{2b}\quad t\in[0,T] \tag{13.2.51}$$

本例，

$$0.213 \leqslant u^*(t) < 0.253$$

从图 13.2.3 所示的 $u^*(t)$最优积累率的曲线可以看出，为使计划期$[0, T]$内消费基金增长总额最大，最优积累率 $u^*(t)$应按一定规律逐年有所减少，但不能低于下限 0.213。这就是说，即使单纯从多消费的角度出发，积累率也不能任意降低。因此，我们将 $u_{下}=0.213$ 称为合理积累率的下限。

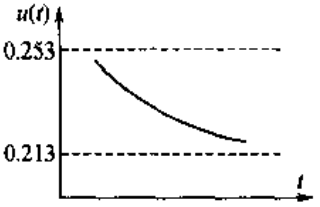


图 13.2.3 消费基金增长总额最大时间 $u^*(t)$

实际上，国民经济的决策不可能只考虑某个单一的目标。根据问题的经济性质，一般应兼顾国民收入的增长，积累基金的增长和消费基金的增长各项指标，可以采取每项指标的乘一个加权系数再取和来表示综合性指标，约束条件仍是式(13.2.33)～式(13.2.35)，这时求出的最优积累率的变化范围为

$$0.23 \leqslant u^*(t) \leqslant 0.28$$

这个结果同人们的经验判断，积累率在 25% 左右的数据是一致的。

根据最优控制原理，利用计算机，不但可以计算出各年的积累率，而且可以计算出相应的国民收入等。

这个模型有两个特点：一是前提简单，没有太多的假设条件；二是计算方便，对统计资料的要求也少，仅为历年的积累率和国民收入的增加额两批数据，这是任何一个国家最起码的统计资料。本模型的不足之处在于模型简单，不得不忽略许多因素，这对结果的准确性会带来一定的影响。

对离散系统，庞特里亚金极大值原理为

以

$$\min \quad J = \sum_{K=0}^{N-1} L[X(K), u(K), K] \tag{13.2.52}$$

为指标泛函，以

$$X(K+1) = f[X(k), u(K), K] \tag{13.2.53}$$

为差分状态方程的动态控制系统，在所有容许控制中确定一最优控制序列 $u(K)(K=0, 1, \cdots, N-1)$，使指标泛函达极小值的必要条件为离散哈密顿函数：

$$H[X(k), u(K), K] = L[X(k), u(K), K] + \lambda_{k+1}^T f[X(k), u(K), K] \tag{13.2.54}$$

取极小值，其中 $\boldsymbol{\lambda}$ 为伴随向量，它是伴随方程：

$$\lambda_K = \frac{\partial H[X(K), u(K), \lambda(K), K]}{\partial K(K)} \tag{13.2.55}$$

的解。

离散系统最优控制的求解方法与连续系统基本相同，只是其伴随方程为离散方程，由 $\dfrac{\partial H(K)}{\partial X(K)} = \lambda(K)$可得出 λ 的递推公式：

$$\lambda_{K+1} = g[X(k), \lambda(K), K] \tag{13.2.56}$$

由 $\dfrac{\partial H(K)}{\partial u(K)} = 0$ 可得 $u(K)$与 $\lambda(K+1)$的关系式，代入状态方程得 X 的递推公式

$$X(K+1) = f[X(K), \lambda(K+1), K] \tag{13.2.57}$$

从而可求出 $X(K)$和 $\lambda(K)$。

参考文献

1 钱学森．论系统工程．长沙：湖南科技出版社，1983

2 汪应洛．系统工程导论．北京：机械工业出版社，1982

3 王众托．系统工程学．北京：国防工业出版社，1980

4 《运筹学》教材编写组．运筹学(修订版)．北京：清华大学出版社，1999

5 陈锡康等．经济数学方法与模型．北京：中国财政经济出版社，1982

6 马促书等．数学规划讲义．北京：中国人民大学出版社，1981

7 董肇君等．林业系统工程．哈尔滨：黑龙江科技出版社，1994

8 邓聚龙．灰色预测与决策．武汉：华中工学院出版社，1986

9 许仁忠．模糊数学及其在经济管理中的应用．四川：西南财经大学出版社，1987

10 E·米涅卡[美]．网络和图的最优化算法．李家澄等译．北京：中国铁道出版社，1984

11 徐光辉．随机服务系统．北京：科学出版社，1980

12 邵汉光等．投入产出法(部门联系平衡法)概论．北京：中国人民大学出版社，1983

13 钟契夫等．投入产出原理及其应用．北京：中国社会科学出版社，1982