通·识·教·育·丛·书

数学赏析

Enjoy Mathematics

向隆万◎著

上海交通大學出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS
北京大学出版社
PEKING UNIVERSITY PRESS

内 容 提 要

本书前三章与数学基础有关，包括数学的对象——数与空间、数学常用的形式逻辑方法、公理化体系、有限与无限等；第四章和第五章介绍微积分的基本概念、理论和方法，包括导数、微分、定积分、不定积分等；第六章介绍优化问题。本书还提供了两个选修材料："唐诗格律的形式体系"和"微积分在经济问题中的若干应用"。

本书取材也有一定特色。如从"科学计数法"、"二进制"、"准确数与近似数"等不同角度介绍"数"；又如讲"空间"，强调笛卡尔坐标引出解析几何的革命性作用，还介绍了高维和分数维空间的意义。在阐述"公理化体系"时，并未停留于数学的逻辑严格性，而是指出它是人类认识世界的重要思想方法，即使在人文社会科学中，也有强大的生命力。对于"导数"和"定积分"这两个微积分最核心的概念，用了相对冗长的篇幅描述历史背景，以及阐明"局部以直代曲，再取极限"的思想方法；在"优化问题"中，不仅介绍了微分学求极值的方法，也指出微积分的局限；特别是以崇敬的心情简介了当年数学大师华罗庚先生大力推广的"优选法"和"统筹方法"。

为了突出数学思想的魅力和文化底蕴，本书适当减弱了严格的数学推导；但本书又不是科普通俗读物，希望读者除了会"赏"，还要能"析"；保持了一些数学论证，希望读者能学会解决一些实际问题，并能享受其中的乐趣。

图书在版编目(CIP)数据

数学赏析/向隆万著. —上海：上海交通大学出版社，2012
(通识教育丛书)
ISBN 978-7-313-08207-7

Ⅰ.①数… Ⅱ.①向… Ⅲ.①数学—青年读物 Ⅳ.①O1-49

中国版本图书馆 CIP 数据核字(2012)第 039478 号

数学赏析

向隆万 著

上海交通大学出版社
北京大学出版社 出版发行

(上海市番禺路 951 号 邮政编码 200030)

电话：64071208 出版人：韩建民

常熟市文化印刷有限公司印刷 全国新华书店经销

开本：710mm×1000mm 1/16 印张：20.25 字数：323 千字

2012 年 4 月第 1 版 2012 年 4 月第 1 次印刷

ISBN 978-7-313-08207-7/O 定价：43.00 元

总序

通识教育再认识

自19世纪初美国博德学院的帕卡德教授最早明确提出通识教育概念以降，世界各国通识教育呈现出跌宕不息的动态过程，相关的研究、讨论与实践不断深入。我国在20世纪90年代中期以来，针对高等教育过分强调专业教育而忽视综合素质培养的状况，在加强学生素质教育及西方通识教育理念本土化方面，进行了有益的探索，对通识教育的认识在深化，共识在提高。然而，时至今日，对通识教育的本质及其作用等若干关键问题的认识，或似是而非，或语焉不详，需要进一步厘清。

一、通识教育是人本教育

讲通识教育是人本教育，至少包含“育人为本”和“以人为本”两层含义。纵观近现代大学的发展历程，虽然大学越来越多地承担着诸如科技创新、服务社会、文化引领等诸多功能，但是，培养人始终是大学的基本功能。相对于研究机构和产业部门等其他社会组织，大学存在的终极理由和根本使命是培养人，就是要在受教育者年轻而又最具可塑性的时候教育他们，塑造他们。

通常，国外将大学功能概括为“Teaching，Research，Service”，其实三者都是为人才培养服务的。简单地视“Teaching”为“人才培养”极不妥当，将人才培养、科学研究和服务社会三大功能等量齐观，更是错上加错。三种功能绝非并列关系，而是主从关系，是“一体两翼”。人才培养是“体”，科学研究和服务社会是从人才培养这个根本使命和核心功能中派生出来的“两翼”，体之不存，翼将焉附？

我们不能在教育功能多元化中，迷失育人这一本然价值。现在，论及大学办学理念者，几乎言必称洪堡。殊不知，以强调科学研究和学术自由而著称的他，并不是就学术论学术，而是围绕培养学生而提出。洪堡认为，只有将科研和教学结合起来，才有利于学生形成良好的思维方式和高尚品格。当年，蔡元培就任北京大学校长发表演说时，开宗明义地宣告："请以三事为诸君告：一曰抱定宗旨；二曰砥砺德行；三曰敬爱师友。"显然，三事实为一事，就是育人。哈佛大学前校长劳伦斯·H·萨默尔斯亦深刻指出，"对一所大学来说，再没有比培养人更重要的使命。假如大学都不能承载这一使命，我看不出社会上还有哪家机构能堪当此任。假如我们葬送了人文教育的薪火相传，一切将覆水难收。"近期，国家和上海市的《中长期教育改革和发展规划纲要》均高扬育人为本的旗帜，将其作为核心理念贯穿文本始终。

然而，诚如《失去灵魂的卓越》一书作者哈瑞·刘易斯指出的那样，实际运行的大学已经忘记了更重要的教育学生的任务，学术追求替代了大学教育。现在，大学里高深研究和教书育人存在于两个完全不同的平面上，究其原因，前卡耐基教学促进基金会主席欧内斯特·博耶一语道破：前者是愉悦、成名和奖励之源，后者却或多或少地成为大学不愿承担的负荷，只是用来维持其存在的堂而皇之的理由。大学尤其是研究型大学，过去常常忽视本科生教育，现在依然如此，普遍存在的重科研轻教学、重科技轻人文、重知识轻心智等倾向，并未得到根本改观。鉴于此，教育界众多有识之士大声疾呼，要回归大学本质，重振本科教育。在高等教育大众化的今天，大学教育代表着一个民族、一个国家的未来和希望。如果十八九岁二十几岁年轻人的教育出了问题，吞下恶果的终将是整个国家和社会。

人本教育的第二层含义是以人为本。以人为本需要探究"以什么人为本"和"以人的什么为本"这两个基本问题。进一步深究，"以什么人为本"又需要回答：是以学生为本还是以教师为本？是以全体学生为本，还是以某些或某一（几）类学生为本？长期以来，大学的焦点从学生转到了教师，学生的主体地位未能得到充分体现，各大学引以为傲的，是拥有世界知名的教授和原创性的科学研究。即便宣称是以学生为本或给予学生足够的关注，教育资源也未能公平惠及所有学生，以牺牲普通学生的正常教育为代价，换取一些所谓优异学生的超常教育和过度教育并非个案。这不仅严重背离了孔子"有教无类"的教育主张，也与马克思、

恩格斯在《共产党宣言》中提出的“每一个人的自由发展是一切人自由发展的条件”的著名观点，以及一个更高级的社会形式应“以每个人的全面而自由的发展为基本原则”背道而驰。众所周知，每一个学生都是独有的生命个体，在先天禀赋、家庭背景、成长环境、知识掌握、兴趣爱好、主观努力、学业成绩、专业技能等方面客观存在各种差异，学校要正视并尊重学生的差别，恪守、秉持并践行“一切为了学生，为了一切学生”，“为了每一个学生的终身发展”的理念，让每个学生都拥有成功梦想的机会。

关于“以人的什么为本”，更是见仁见智，莫衷一是。随着商业社会竞争的日益激烈和就业形势的日渐严峻，让学生用最短的时间掌握最多的知识和技能，成为教育活动的不懈追求，以“专业技能”为本，便顺理成章并大行其道。通识教育应以“完善人格”为本，即以“精神成人”而非“专业成才”为本，亦即以人的行为养成、道德认知、情感体验、理想信念、心灵攀登和全面发展为本，着力把学生培养成有个人修养、有社会担当、有人文情怀、有科学精神、有历史眼光、有全球视野的完整人。

其实，早在古罗马时期，思想家西塞罗就认为，教育的目标不仅是培养具有某些专门技能的人，教育的崇高目标，应当是培养使其他德行相形见绌的真正的拥有至善人格的人。1945 年，哈佛委员会在著名红皮书《自由社会中的通识教育》中同样明确地提出，通识教育着眼于学生身体、道德和智力的和谐发展，致力于把学生培养成为知识全面、视野广阔、教养博雅和人格完整的人。我国著名教育家潘光旦一针见血地指出，“教育的理想是在发展整个的人格”。蔡元培先生亦精辟论述到：“教育者，养成人格之事业也。使仅仅为灌注知识、练习技能之作用，而不贯之以理想，则是机械之教育，非所以施于人类也。”可以说，强调教育的本质乃是培养健全的人，是古今中外前辈先贤们深邃的通识教育思想精要所在。

二、通识教育是自由教育

穷源溯流，通识教育（General education）的理论渊源，可以追溯到古希腊的博雅教育或自由教育（Liberal arts）。亚里士多德最早提出自由教育思想，他认为自由教育既不立足于实用，也不立足于需求，而是为了心灵的自由；通过发展理性，提升智慧及道德水平，实现人的身心和谐发展。当时，博雅指称人类心灵中的成就，同时包括艺术及知识。而博雅教育就是广博知识及洞察力的教育，是

真正能抓得住真理及美的教育，是造就博大风雅、博学文雅、博闻儒雅、博古典雅、举止优雅、志趣高雅之谦谦君子的教育。

1828 年，耶鲁大学在其发表的报告中提出，大学的目的在于提供心灵训练和教养，充实具有知识的心灵。英国红衣主教和教育家纽曼进一步发展了这种思想，他在《大学的理想》一书中，不仅系统论述了自由教育思想，而且明确提出，对受教育者而言，大学教育就是自由教育。现代通识教育以适应社会要求、满足学生兴趣和维系文化传承为其内核，其要义是对自由与人文传统的继承。个体藉着知识、智慧、善意与爱，在精神上摆脱物质的束缚，在生活中摆脱各种利害，不为物役，不以物喜，不以己悲，从而获得真正自由。通识教育鼓励反省求真，追求心灵的成长和人性内在的精神解放，在真正的学习和探究中，展现个体的潜能，体悟生命的意义，诠释生活的真谛，实现对功利的超拔，对自我的超越。

从词源角度讲，虽然在不同的历史时期，人们对自由教育中 liberal 一词的认知大相径庭，如将 liberal 理解为文雅的（genteel）、书面的（bookish）、解放的（liberating）、符合绅士身份的（becoming a gentleman）、高贵的（noble）等多种意思，但大多数理解还是关乎自由。其中，最常见的是将 liberal 解释为“自由的”（free），如康德、汉娜·阿伦特、汉斯-格奥尔格·伽达默尔、罗伯特·赫钦斯等，或解释为“使人自由的”（make man free），如古罗马政治家、哲学家塞涅卡等。实际上，自由一直是西方居支配性地位的一种观念。在西方传统中，自由具有最高价值，是一切人文科学和教育的核心。自由不仅是民主、科学、理性、正义、良知、宽容等普遍价值的元价值，也是人文学科最基本的价值支点。裴多菲的诗“生命诚可贵，爱情价更高；若为自由故，两者皆可抛”就是对“不自由毋宁死”的明证。德国哲学家、诠释学创始人、时任柏林大学校长施莱尔马赫曾言：“大学的目的并不在于教给学生一些知识，而在于为其养成科学的精神，而这种科学精神无法靠强制，只能在自由中产生”。1987 年时任耶鲁大学校长的施密德特，在迎新典礼上慷慨陈词：“一所大学似乎是孕育自由思想并能最终自由表达思想的最糟糕同时又是最理想的场所”；“自由的探求才会及时更正谬误，代替愚昧，才能改变偶尔因我们感情用事而认为世界是分离的、虚构的和骗人的偏见。”在我国，学术大师陈寅恪的“独立之精神，自由之思想”，与西方的这种传统和倡导遥相呼应，并日渐成为中国知识分子共同追求的学术精神、价值取向和人生理想。

自由教育作为通识教育的一大鲜明特征，不仅体现为对心灵自由和精神解

放的追求，还体现为对批判性思维的崇尚。在新韦氏词典里，批判性思维是指“以审慎分析判断为特点，并在最严格意义上隐含着客观判断的尝试而定褒贬优劣”。人类的思考有其内在缺陷，经常陷于偏颇、笼统、歧义、自欺、僵化和促狭之中，不自觉地倾向自我（和社会）中心主义、人类中心主义、西方中心主义或某某中心主义。既有的知识系统，不管创造它们的先贤圣哲多么睿智，其中的片面、寡陋、扭曲、非理性、傲慢甚至偏见都在所难免。通识教育并非共识教育或认同教育，学生要敢于质疑、反思、检讨、追问、解构乃至颠覆，不仅从学理逻辑的角度审视，还要关切知识理性背后的正义性和善意性，发展各种知性美德。此外，批判性思维还要体现苏格拉底“未加审视的生活不值一过”的原则，秉持古希腊“自省生活”的理想，不断提高个体自我感悟和向内反省的智慧。要充分认识到，若放任自流，许多未加审视的生活加在一起，会使这个世界因黑白颠倒、是非混淆、美丑不分、正义不彰而危机四伏。

受教育者主体地位的确立，是自由教育的前提。通识教育作为摆脱各种奴役成为自由自主之人的教育，必须让学生真正成为学习的主人，培养学习兴趣、激发学习动力是“自由教育”的要点。潘光旦认为，人的教育是“自由的教育”，以自我为对象。自由的教育是“自求”的，不是“受”和“施”的，教师只应当有一个责任，就是在学生自求的过程中加以辅助，而不是喧宾夺主。只有这样，教育才能真正进入“自我”状态，学生才能通过“自求”至“自得”进而成为“自由的人”，也就是上面谈及的“至善”境界中的完整人。黎巴嫩著名诗人、艺术家、哲理散文家纪伯伦说得好：“真正有智慧的老师不会仅仅传授知识给任何学生，他会传授更珍贵的东西：信念和热忱。真正的智者不会手把手地带学生进入知识的殿堂，只会带学生走向自身能够理解的那扇门”。

在现实教育中，教育机构和教师的主导作用发挥得很充分，学生的主体地位和主动性却体现得严重不足，这种情况从基础教育一直延续到高等教育。原本应是“养成”教育的通识教育，变成为“开发”教育，被开发、被培养、被教育、被教化、被塑造、被拔尖等不一而足，课业负担重，学习兴趣缺乏，创新意识不足，几成常态，学生只是消极被动地参与其中，体会不到学习的乐趣。针对这些情形，教育要切实承担起责任，注重激发和调动学生内在的激情、兴趣、好奇心和探索冲动，要像中国近代教育家陶行知强调的那样，解放学生的“头脑、双手、眼睛、嘴巴、空间和时间”，使他们能想、能干、能看、能谈，不受任何禁锢地学习和发展。

自由、自在、自觉地阅读经典，是通识教育的良方。芝加哥大学、哈佛大学、哥伦比亚大学、耶鲁大学、斯坦福大学、牛津大学、剑桥大学、香港中文大学等著名学府，十分重视学生对经典文本的研读。深入阅读柏拉图、亚里士多德、莎士比亚、康德等西方经典和儒家等华夏经典，以及《可兰经》、《源氏物语》等非西方经典，目的在于培养学生内在的价值尺度、精神品格、独立意识和批判精神，帮助学生养成健全有力的人格。学生自在地徜徉于浩渺的知识海洋，漫步于辉煌的精神家园，穆然深思，精研奥理；悠然遐想，妙悟游心；享受愉悦，怡养性情。通过与先贤对话，与智者神交，从中感悟人类思想的深度、力度、高度和厚度，领略历代硕儒的闳博哲思和学理旨趣，体味铮铮君子的人生情怀和胸襟气象，修得个体生命的丰盈圆融。

三、通识教育的无用之用

“用”可分为“有用之用”和“无用之用”。在很多人看来，所谓有用就是可产生功利的、现实的、物质的、实在的和直接的效用、功用或好处。由于深受经世致用思维和实用主义思潮的影响，特别是市场经济条件下，大学教育过分强调与市场接轨和需求导向，过分追求学以致用和实用理性，过分信奉使用价值而非价值本身，过度渲染只有过得“富有”才有可能“富有价值”，过分注重工具理性，严重忽视价值理性，人被“物化”已是当今不争的事实。

通识教育本身不是一个实用性、专业性、职业性的教育，也不直接以职业作准备为依归。基于功利性的价值取向，通识教育似乎无用，然而，相对于“有用有所难用”的专业教育，通识教育却“无用无所不用”。通识教育充分体现老子“有之以为利，无之以为用”的思想，充分体现罗素“从无用的知识与无私的爱的结合中更能生出智慧”的论断，其“无用之用”主要体现为：

一是彰显人的目的性，回到“人之为人”的根本问题(essential questions)上，使人活得更明白、更高贵和更有尊严。如前所述，通识教育是一种人本教育，强调培养的是全人而不是工具人、手段人。康德有句名言“人是目的，不是手段”，这一命题深刻表达了人的价值与尊严。现在经常讲“这个有什么用”，其实就是把自己当手段，谋求市场上能有(效)用。通识教育不追求“学以致用”，更看重“学以致知”和“学以致省”。大学是理想的存在，是道德高地，是社会的良心，是人类的精神家园，大学教育是知识、能力和价值观三位一体的教育，与专业教育

相比，通识教育侧重于价值观的塑造，更突出精神品格和价值诉求，关切所做每件事情背后的动机、价值和意义，思考专业知识层面之上的超越性问题和事关立命安身的终极性问题，对伦理失落、精神颓废、生活浮华和自我自利保持起码的警觉和反省能力；对物质主义、拜金主义、享乐主义和无边消费主义等种种时弊，以及低俗、庸俗和媚俗等现象，保持清醒的认知和足够的张力，自觉抵制浑浑噩噩的市侩生活。通识教育倡导持之以恒地用知识、智慧、美德丰富与涵养自身，力戒见识短浅、视野狭窄和能力空洞，推崇仰望星空，瞭望彼岸，保持一种超越的生活观，像海德格尔所言“诗意地栖居”。

二是有助于打好人生底色，完善人格，滋养成为合格公民的素养。通识教育引导学生形成正确的世界观、人生观和荣辱观，使学生获得对世界与人生的本质意义广泛而全面的理解，形成于己于国都可持续发展的生活方式，培养诚信、善良、质朴、感恩、求真、务实等道德品质，引导学生认识生命，珍惜生命、热爱生活，崇尚自尊、自爱、自信、自立、自强和自律，养成开阔的视野、阳光的心态、健全的心智和完善的人格。通识教育还帮助学生思考生态环境与生命伦理问题，促进学生树立善待环境、敬畏生命、推己及人、服务社会的理念，构建生命与自我、与自然、与他人、与社会的和谐关系。通识教育突出民主法治、公平正义、权利义务的理念，帮助学生树立责任、程序、宪政等意识，培养他们成为合格公民的素质。同时，通识教育有助于学生找到与自身禀赋相匹配的爱好和兴趣，有助于锤炼在多元化社会和全球化环境生活的能力，为即将展开的职业生涯打下坚实的根基。

三是有助于形成知识的整体观和通透感。通识教育是关于人的生活的各个领域知识和所有学科准确的一般性知识的教育，是把有关人类共同生活最深刻、最基本的问题作为教育要素的教育，恰如杜威所言：“教育必须首先是人类的，然后才是专业的”。通识教育致力于破除传统学科领域的壁垒，贯通中西，融会古今，综合全面地了解知识的总体状况，帮助学生建构知识的有机关联，实现整体把握，培养学生贯通科学、人文、艺术与社会之间经络的素养，避免知识的碎片化，避免因过早偏执于某一学科而导致的学术视角狭隘，防止一叶障目的片面，盲人摸象的偏见，鼠目寸光的短视及孤陋寡闻的浅薄，力图博学多识，通情达理，通权达变，融会贯通、思辨精微乃至出神入化，力争“究天人之际，通古今之变，成一家之言”。

《易经》的“君子多识前言往行”、《中庸》的“博学之、审问之、慎思之、明辨之、笃行之”、《老子》的“执大象，天下往”、《淮南子》的“通智得而不劳”、《论衡》的“博

览古今为通人”，孔子的“君子不器”、荀子的“学贯古今，博通天人；以浅持博，以古持今，以一持万”、王充的“切忌守信一学，不好广观”、颜之推的“夫学者，贵博闻”，以及陈澹然的“不谋万世者，不足谋一时；不谋全局者，不足谋一域”，这些响彻人间千百年的箴言，无不说明通识教育中“通”（通晓、通解、明白、贯通）和“识”（智慧、见识、器识）的极端重要性，“博闻，择其善而从之”，讲的就是越趋于广博、普通的知识，越有助于人的理智、美德的开发及全面修养。需要注意的是，博学不能“杂而无统”（朱熹），“每件事都知道一点，但有一件事知道得多一些”（约翰·密尔）。通识教育应当将博与专统一起来，各学科专业知识的简单叠加，无助于学生形成通透、系统的知识体系。

四是有助于发展智能素质。教育的目标不仅要“授人以鱼”，更重要的是“授人以渔”。纽曼认为，自由（通识）教育之所以胜过任何专业教育，是因为它使科学的、方法的、有序的、原理的和系统的观念进入受教育者的心灵，使他们学会思考、推理、比较和辨析。接受过良好通识教育的学生，其理智水平足以其胜任任何一种职业。通识教育注重弘扬人文精神和科学精神，陶冶性情，崇尚真理，发展学生的理性、良知和美德。通过向学生展示人文、艺术、社会科学、自然科学和工程技术等领域知识及其演化流变、陈述阐发、分析范式和价值表达，帮助学生扩大知识面，构建合理的知识结构，强化思维的批判性和独立性，进而转识成智，提升学生的洞察、选择、整合、迁移和集成创新能力，尤其能提升学生有效思考的能力、清晰沟通的能力、作出明确判断的能力和辨别一般性价值的能力，这些比掌握一门具体的专业技能更本质更重要，并能产生最大的溢出效应。

四、“通”、“专”之辨

在教育教学实践中，尽管通识和专业、教育理想和社会需求间存在矛盾和冲突，但在认知理念和培养原则上应该明确，通识教育和专业教育都很重要，不能简单地讲孰重孰轻，更不能将它们对立和割裂。这二者之间应是相辅相成、相得益彰的关系，是体与用、道与术的关系，是传承与创新、坚守与应变的关系，下分述之。

首先，通识教育与专业教育都不可或缺，它们作为一对范畴，共同构成高等教育的全部内容。一方面，专业教育是大学教育之必需。这是因为，从科技演化趋势层面看，当今知识和科技发展表现出两个鲜明的向度：一是各学科领域之间的交叉融合越来越强，综合集成的要求越来越迫切；另一趋势则是学科学术越

来越专，专业分工越来越细，尤其是进入网络时代，知识和资讯爆发性增长，客观上要求从“广而泛”转向“专而精”，若术无专攻，则难以立足。从国家和社会发展层面看，中国作为一个后发新兴经济体，建设与发展任务十分繁重，亟需大批各行各业的专业人才，以服务于富国强民的国家战略。从教育机构义务角度看，当大学接受一名学生时，就当然地负有为学生提升能力的责任。当今高等教育已不再是精英教育，而是大众教育，大众教育需要紧密结合社会实践和市场需求。专业教育可以让学生尽快进入某一专业领域，在较短时间内习得具有胜任力的专业知识，学生将来无论是进职场就业，还是到研究生院某个专业深造，都由此而具备竞争力。从学生最现实的角度考量，通过专业教育学生掌握安身立命的谋生技能和本领。

另一方面，通识教育是大学教育之必然。上文已谈及，现代科技发展两个向度之一，就是知识领域或专业领域间的融通贯通。然而，专业教育容易使人单一片面，甚或成为局限在过于狭窄的专业领域中的工作机器，按米兰·昆德拉的说法，“专门化训练的发展，容易使人进入一个隧道，越往里走就越不能了解外面的世界，甚至也不了解他自己。”更糟糕的是，一直以来专业教育深受工具理性支配，在很大程度上已经沦为一种封闭性的科学教条，成为现代工业生产体系的一个环节，促进人心灵成长的价值几近泯灭。通识教育强调价值性、广博性与贯通性，正好可以纠偏矫正，观照专业教育。尽管在不同时代、不同国家和地区通识教育产生的具体社会背景不尽相同，但相同之处都是对过分专业化的一种反动，其指向是革一味偏重专业之弊。此外，如前所述，通识教育的“通”不仅指称在学科领域和专业领域的“通”，更是为人和为学的“通”。为此，恰如“寻找灵性教育”的小威廉姆·多尔所言，就是要确立科学（逻辑、推理）、艺术（文化、人文）和精神（伦理、价值观、生命、情感等）三大基石，并在科学、艺术和精神之间进行关键性整合互动，还要在更大的时空和更广泛的社会实践中，不断提升“每个人全面而自由发展”的生命价值。为人为学之通，既是通识教育的题中之义，更是大学教育的灵魂。

第二，通识教育和专业教育是相辅相成、相得益彰的关系。通识与专业，或广博与专精，抑或古人眼里的“博”与“约”是辩证关系，专而不通则盲，通而不专则空。它们密不可分，互为前提，相互依存，相互促进。不通，则知识狭窄，胸襟狭隘，思路不广，头脑闭塞，往往就事论事，盲目不知其所以。同时，缺乏多学科、多领域知识的启迪与支撑，“专”也没有基础；反之，不专，则博杂不精，一知半解，

浮光掠影，空泛浅薄。何况知识浩如烟海，汗牛充栋，且人生有涯，知识无限，若滥学无方，将一事无成。所以，需在专中求通，通中求专，专通结合，博约互补。既要遵循学术自有的分类和流变，又要注重整体关联和宏观把握，在掌握各种专门技能和领域知识的同时，拥有宽厚的基础和综合的素质。在培养学生上，宜采用“通—专—通”的动态模式，即学生刚入学时不分专业，先进入文理学院或书院接受通识教育；接下来，高年级本科生和研究生在此基础上进行宽口径的专业教育。之后，他们接受更高一个层次的通识教育，在新的起点和更厚实的基础上再进一步聚焦专业学习，如此循环往复，螺旋推进。

第三，通识教育与专业教育是体与用、道与术的关系。前面已经指出，通识教育是关乎人的根本问题的教育，旨在引导学生形成正确的世界观、人生观和价值观，在有限的人生中充分发挥天赋良能和生命潜能。有鉴于此，通识教育具有基础性、本体性和深刻性，故应以通识为体，专业为用。同时，通识教育又是人格养成和悟道的教育，涵养人格知、情、志三维度中的“情”和“志”，以及领悟万术之源、众妙之门的“道”，要仰仗生活底蕴和文化自觉的培植，而通识教育正是培植这种底蕴和自觉的重要手段之一。其实，孔子早就提出“君子不器”的重要思想。他认为，君子无论是做学问还是从政，都应该博学多识，才能统揽全局，领袖群伦；才不会像器物一样，只能作有限目的之用。陶行知亦提出“生活即教育”的生活教育理论，并毕生践行。梅贻琦在他《大学一解》一文中更是明确表达“通识，一般生活之准备也；专识，特种事业之准备也。通识之用，不止润身而已，亦所以自通于人也。信如此论，则通识为本，而专识为末”，“大学教育应在通而不在专，社会所需要者，通才为大，而专家次之。”他掷地有声地指出：“以无通才为基础之专家临民，其结果不为新民，而为扰民。”孔子的思想、梅贻琦的观点和陶行知践行的理论意义深远，至今仍闪烁着智慧的光芒，照亮通识教育的复兴之路。

第四，通识教育与专业教育是传承与创新的关系，是坚守与应变（或罗盘与地图）的关系。统计研究揭示，最近十年内科学技术的成就，超过了人类历史上以往所有成就的总和，十年间知识已翻了一番。抽样调查表明，一个大学毕业生离校五年以后，其所学知识一半已经陈旧，十年以后可能大部分陈旧。文献计量研究亦表明，一些基础学科文献的半衰期为 8～10 年，而工程技术和新兴学科的半衰期约 3～5 年。实际上，早在科学技术还不十分发达的 1949～1965 年间，美国已有八千种职业消失，同时又出现了六千种新的职业。诚然，当今知识更新的周期越

来越短，科技升级换代的频率越来越快，专业教育必须不断创新，以变应变，才能应对迅速变化的世界，才能因应“今天的教师，用昨天的知识，教明天的学生”的悖论。

在这个日新月异的时代，通识教育却要传承亘古不变的真善美，坚守世世代代本色生活的价值与意义，追问世界根底的本原和终极，反省历久弥新的伦理和人生。通过通识教育，保证千百年来的文明薪火相传，永恒绵延；同时，守望人类文明共同体，确立代际“最大公约数”。因此，通识教育不是什么新、什么前沿就学什么，恰恰相反，通识教育课程中没有流行或时尚的东西，不包含那些尚未经过岁月涤荡和历史检验的材料。芝加哥大学就明确规定，凡是活着的人的言论，不得放进通识教育课程。通识教育深谙罗曼·罗兰那句“很快就不流行的叫流行，很快就不时尚的叫时尚”的名言，以及与时俱进必与时俱迁的道理，在开放、多元、多样和多变中，保持坚守传承的品格和追问反省的本性，以惯看秋月春风的淡定和浪花淘尽英雄的从容，确保不在滚滚红尘中迷失，不被汹涌潮流所裹挟，就像罗盘，永远锁定方向，指针北斗。

以上是对通识教育的一些粗浅认识。近年来，上海交通大学在新一轮教育思想大讨论的推动下，不断深化对通识教育理念的认识，成立了校通识教育指导委员会与通识教育教材建设委员会，依据人文学科、社会科学、自然科学与工程技术、数学与逻辑这四个模块，初步形成了通识核心课程体系，并明确将出版通识教育系列丛书，作为加快推进通识教育的重要抓手。在最近两年多的时间里，交大通识教育指导委员会的多位专家和丛书的作者与交大出版社和北大出版社保持密切接触和沟通，其间，北大出版社社长、总编等多次赴交大沟通出版事宜，交大出版社领导和编辑也多次赴北大出版社进行接洽。众所周知，交大出版社以理工科著作和教材出版见长，北大出版社在人文、社科方面实力超群，两家优秀出版社强强联合，联袂推出这套丛书，可谓珠联璧合。衷心希望这套丛书能得到广大读者的认可和喜爱。

徐 飞 博士

上海交通大学战略学教授、博导

上海交通大学通识教育指导委员会主任

2011 年 3 月

前言

我从事大学数学教学四十多年，开设过数学课程十余门，也参与编写过若干教材，但是编写通识教材《数学赏析》却是首次尝试。

数学的重要性是不言而喻的，从幼儿园到大学，数学都是必修课。但是，相当一段时期以来，在功利主义思潮的影响下，大学数学只是强调工具的传授，对数学知识背后深厚的思想基础和精神价值却大大忽视了。本书希望在这方面作一些探索。

在编写过程中，遇到前所未有的挑战。如何选材与行文，以吸引读者特别是人文社科读者对数学的兴趣？如何将数学与人文艺术交融？如何兼顾科学性与可读性？似无章可循；另一方面，编写这本教材又给我带来莫大的愉悦。因不受现有框架约束，便可旁征博引，信马由缰。苏东坡说过："作文如行云流水，初无定质，但常行于所当行，常止于不可不止"，现在有了一点体会。

本书共分六章。前三章与数学基础有关，包括数学的对象——数与空间、数学常用的形式逻辑方法、公理化体系、有限与无限等。这部分着重介绍公理化逻辑体系，同时穿插了证明费马大定理、哥德巴赫猜想和四色问题的故事，赞扬了人们为攀登科学高峰而不懈努力的精神。第四章和第五章介绍微积分的基本概念、理论和方法，包括导数、微分、定积分、不定积分等。第六章介绍优化问题。

本书不同于通常的教材，为了突出数学思想的魅力和文化底蕴，适当减弱了严格的数学推导；但本书又不是科普通俗读物，希望读者除了会"赏"，还要能"析"。在不少例题中，数学论证非常详尽，希望读者能真正学会解决一些实际问题，并能享受其中的乐趣。特别在极限微积分部分，包含了不少数学推理，甚至保留了极限的逻辑定义。实际上，微积分的力量，正在于能够定量解决过去无法企及的变化问题。如果完全脱离数学公式，很难领略微积分的实质与妙处。当然，要找到恰当的平衡，非常不易，作者只是尝试。对于并不想深入了解理论与方法的读者，不妨跳过一些数学推理；对于希望进一步了解并掌握微积分的读

者,本书可以作为参考和入门。

本书取材也有一定特色。比如讲“数”,没有重复中学学过的“自然数→有理数→无理数→实数→复数”的发展过程,而是从不同角度介绍“科学计数法”、“二进制”、“准确数与近似数”等;又如讲“空间”,强调笛卡尔坐标引出解析几何的革命性作用,还介绍了高维和分数维空间的意义。在阐述“公理化体系”时,并未停留在数学的逻辑严格性,而是指出它是人类认识世界的重要思想方法,即使在人文社会科学中,也有强大的生命力。对于“导数”和“定积分”这两个微积分最核心的概念,用了相对冗长的篇幅描述历史背景,以及阐明“局部以直代曲,再取极限”的思想方法;在“优化问题”中,不仅介绍了微分学求极值的方法,也指出微积分的局限;特别是以崇敬的心情简介了当年数学大师华罗庚先生大力推广的“优选法”和“统筹方法”。

本书力图介绍数学理论与方法的来龙去脉,“既见树木,又见森林”;同时注意图文并茂,除数学插图外,还加入了一些人物肖像和书刊景物图片,供读者欣赏。每章最后提出一些思考与练习题,并附上参考文献或网站,便于有兴趣的读者进一步查阅。

书法有王、颜、欧、柳等各种“体”,京剧有梅、程、尚、荀等不同“派”,好的数学教材似乎也应当是个性化的。希望这本《数学赏析》写出一些新意。

本书提供两个选修材料:“唐诗格律的形式体系”和“微积分在经济问题中的若干应用”。

直到20世纪初,学写格律诗仍是中国文人的必修课目。老一辈政治家、军事家、科学家和文学家一样,留下了许多脍炙人口的诗篇。其实,唐诗格律也是一种形式体系,基于“声韵”与“对仗”的“公理”,以“0-1”对应“平—仄”,根据“粘对”规则,即成体系。作为选修内容,希望为学生文化素养的提高而添砖加瓦。三十年前,作者在美国访问进修时,年轻气盛,曾在哈佛大学“燕京社”和普林斯顿大学东亚文学系做过讲座,效果尚好;2008年、2009年和2010年作者连续三年在上海交通大学新生研讨课中选讲,学生表现出浓厚的兴趣,不少学生的绝句律诗习作达到相当的水准。

从定量角度研究经济问题,从而找出规律,是经济工作者和各类管理人员的重要任务。以经济建设为中心已确定为我国发展的基本路线。从事任何领域工作的人,特别是各级领导者都应当对经济和管理具备一定的定量分析知识。我们

选择了几个经济管理问题的概念和案例，用微积分方法研究，也作为选修内容。

既然是选修，教师完全可以略去；有兴趣的读者可以自学，既可以当时学，也不妨以后学。

在编写过程中，受益于许多名家和文科数学教学先行者，特别是母校复旦大学老师李大潜院士和国内同行张奠宙、顾沛、张顺燕、谭永基等教授卓越工作的启示。上海交通大学同事乐经良教授和沈灏教授给予许多建议、鼓励和帮助；物理系尤峻汉教授仔细校阅了与近代物理有关的内容；青年诗人程羽黑博士对隋唐以来诗韵的演变提出了不少建议；同事肖柳青副教授和我进行了有益的探讨，她还提供了“优化问题”和“微积分在经济问题中的若干应用”的部分素材；青年同事刘小军老师在编排版面和绘制插图方面花费了辛勤的劳动；上海市欧美同学会青年干事陈卓君和刘炎麟帮助制作图片和校对清样；付梓之前上海交通大学外语学院施旻副教授对本书英文提要认真审阅和润色在此一并表示感谢。上海交通大学通识教育指导委员会主任徐飞教授和教务处田冰雪、鲁莉、杨西强等同志，出版社郁金豹、易文娟等同志对本书的编写出版始终给予关心和帮助。正是他们的鼓励和合作，鞭策我完成本书的编写。

作者也曾计划将“线性代数”和“概率统计”的若干内容编入书中，既限于篇幅，更因作者积淀不够而作罢。总之，我痛感缺乏撰写通识教材的学养，更谈不上经验。初稿虽成，难免忐忑；取材未必恰当，疏误在所难免。从哲学和文化的高度看数学，讲数学，更有“如履薄冰”之感。希望同行，特别是使用本书的老师和广大同学们提出批评建议，以便不断修正提高。

书成之日，百感交集，得小诗一首，以记此刻心情。

挑灯四载求通识，溯古推今甘苦尝。
思绪层层丝结茧，悟怀点点墨成章。
上穷碧落星辰迹，下觅红尘优选方。
妙处难言须赏析，书终回首已枯肠。

向隆万

2012 年春于上海交通大学

目 录

Contents

ABSTRACT

Mathematics, as we might have known it, is not simply a technical system or some principles or formulas to be applied to many different fields, but, perhaps more importantly, a part of history, culture, ideology, and philosophy. Under the inspiration or idea of Liberal Arts Education, the author tries to focus on the latter in this book.

The book consists of two parts. Part one is about the foundation of mathematics, which includes three chapters: "Number and Space", "Mathematical Proof and Axiomatic System" and "Infinity and Limit"; Part two mainly introduces calculus, which also consists of three chapters: "Derivatives and Differentials", "Integrals", and "Optimization Problems". Also provided here for the readers are two optional chapters, "The Formal Pattern of Tang-poems" and "Applications of Calculus to Business and Economics".

Some materials selected in this book are arranged in an unusual way, for example, the stories of people making hundreds of years' efforts in order to prove Fermat Last Theorem, Goldbach Conjecture, Four Color Map, etc. are recounted. As extensions of human abstract thinking, Minkowsky-Einstein 4D space-time and Mandelbrot's fractal geometry are also discussed.

A relatively long chapter is taken to elaborate on the subject of calculus, including its origin, substance and application. It may be necessary to understand why calculus is considered as one of the most glorious achievements of mankind, and to know how great Newton and Leibniz and many other mathematicians once were.

When the optimal problems are discussed, calculus shows its powerful

function. However, nothing is perfect. The limitation of calculus is also pointed out, and some useful direct methods are briefly mentioned, including golden section method, PERT net, linear programming, etc.

Although the rigor proof of some theorems and formulas may be weakened due to the length of the book, the complete mathematical deduction, including Weierstrass's definition of limit, are also kept in this book. Whether readers are able to understand it or not, however, they should admire it as the strict way of reasoning that makes mathematics possible to play a key role in the modern science and technology.

绪论

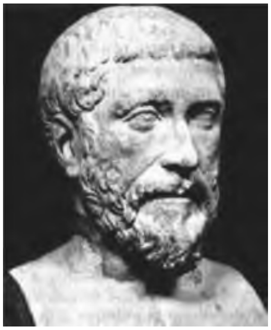
怎样赏析数学

0.1 绪　　言

这本《数学赏析》主要是为文科学生写的通识教材，也可以作为理工科学生的参考书。同时，为没有学过高等数学，却对数学有兴趣的不同年龄、不同学历和工作岗位的读者提供一份课余和工余的阅读资料。

“文科学生为什么要学数学？”如果把这个问题向几千年前的思想家、哲学家、教育家请教，他们或许会做如下回答：

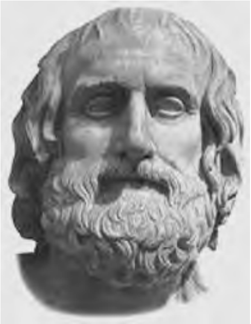
第一，数学是宇宙的特性，甚至是宇宙起源。古希腊哲学家毕达哥拉斯(Pythagoras，572-497 B.C.)学派认为，从音乐和声到日月星辰，“万物皆数”，“数乃宇宙的要素”[0-1]；中国思想家、道家鼻祖老子(571-471 B.C.)更认为：“道生一，一生二，二生三，三生万物”[0-2]。



毕达哥拉斯　572-497 B.C.

老子　571-471 B.C.

孔子 551-479 B.C.



柏拉图 427-347 B.C.

第二,数学是文化。逻辑推理和抽象能力是有教养人的必备素养。中国儒家创始人孔子(551-479 B.C.)是举世公认的教育家。他在曲阜讲学的六门课程是“礼、乐、射、御、书、数”,合称“六艺”[0-3],其中的“数”就是数学。古希腊哲学家柏拉图(Plato, 427-347 B.C.)在他的名著《Republic(理想国)》中说:“打仗的人必须学习数的技巧,否则他就不知道如何布置他的部队;哲学家也要学,因为他必须跳出茫如大海的万变现象而抓住真正的实质。…… 算术有很伟大而崇高的作用,它迫使灵魂用抽象的数来进行推理,而厌弃在辩论中引入可见和可捉摸的对象。”[0-4]他甚至在雅典办的学院门口高挂“不懂几何者不得入内”的牌子。希腊哲学家普洛克洛斯(Proclus Lycaeus, 412-485)说过,毕达哥拉斯等古希腊学者“将几何学研究变成了一种通识教育,从头检验这门科学的基本原理,用一种非物质的理性方式探索其定理。”[0-5]

第三,数学经世济用。数学的起源当来自实际应用。相传由周文王所作的《周易》中有“上古结绳而治,后世圣人易之以为书契,百官以治,万民以察”[0-6]之说。在埃及、两河流域和中国的考古发现中都有先民原始数字的记载。如果说,以欧几里得(Euclid, 330-275 B.C.)的《Elements(几何原本)》[0-7]为代表的古希腊学者更重视数学的逻辑推理体系的话,那么以《周髀算经》(约 300-100 B.C.)[0-8]和《九章算术》(约 50-100)[0-9]为代表的古代中国学者更重视数学在实际问题中的应用和数学方法的探求。

既然如此,现在为什么会提出文科学生学数学的问题呢?随着社会的发展,几千年来,自然科学、人文与社会科学以及数学本身都得到巨大的发展。古人有可能同时精通文科和理科,如张衡(78-139)既能制作候风地动仪,又能吟诵《西

欧几里得《几何原本》(300 B.C.)

《周髀算经》(约 300－100 B.C.)

京赋》;直到欧洲文艺复兴时期,达·芬奇(Leonardo da Vinci, 1452-1519)既是画家、音乐家,又是工程师、发明家,同时在艺术和科学上达到当时的顶峰。但是,自此之后,学科分得越来越细,越来越深。一个人精力有限,不要说文理难以得兼,就是在一门学科中,也只能在某一分支深入研究。1994 年,英裔美国数学家怀尔斯(Andrew Wiles, 1953-)证明了提出 350 年的"费马大定理(Fermat's Last Theorem)"[0-10],获得菲尔兹奖。据说全世界能看懂他证明的人不超过 100 个。

张衡 78-139

达·芬奇 1452-1519

怀尔斯 1953-

另一方面,某些人文社会学科数量关系极其复杂,很难建立数学模型;即使建立模型,也无法求解。因此,长期以来,数学在人文艺术学科中,很难发挥直接作用。如果说,自然科学和工程技术领域强调逻辑思维,注重定量分析;那么,人文艺术领域的学者更依赖形象思维,擅长于定性分析,确实有过文学家、历史学

家“数学考零分”的轶事。

应当指出，数学工作者忽视对数学教育与教学的研究，忽视对数学的宣传与普及也是文科教学与数学渐行渐远的原因。20世纪著名德裔美国数学家柯朗(Richard Courant，1888-1972)在与罗宾(H. Robbins，1915-2001)合写的名著《What is Mathematics?（什么是数学）》[0-11]一书中尖锐指出：“两千多年来，人们一直认为每一个受教育者都必须具备一定的数学知识。但是今天，数学教育的传统地位却陷入了严重的危机之中。而且遗憾的是，数学工作者要对此负一定的责任。数学教学有时竟演变成空洞的解题训练。这种训练虽然可以提高形式推导的能力，却不能导致真正的理解与深入的独立思考。数学研究已出现一种过分专门化和过于强调抽象的趋势，而忽视了数学的应用以及与其他领域的联系。”

“天下大事，合久必分，分久必合”。20世纪80年代以来，文科大学生要学数学的呼声又强了起来。主要原因是随着数学和计算机技术的迅速发展，数学的应用领域越来越广，逐步渗透到人文社会学科。

首先是经济学。

马克思 1818-1883

恩格斯 1820-1895

经济现象中显然蕴含着数量关系。马克思在经济学研究中就运用了不少数学。他在1873年给恩格斯的信中说：“为了分析危机，我不止一次地想计算经济行为不规则曲线的升和降，并曾想用数学公式得出危机的主要规律。”[0-12]但是经济现象非常复杂，当时的数学工具还不够完备，所以长期以来，人们停留在定性研究的阶段。我国大学的“经济学”专业，也多作为文科来招生和培养。

凯恩斯 1883-1946

萨缪尔森 1915-

20 世纪 30 年代以来，这种现象发生了变化。经济学和数学相结合，无论在理论上还是实践上都导致前所未有的重要成果。1936 年英国经济学家凯恩斯(John Maynard Keynes，1883-1946)发表名著《The General Theory of Employment，Interest and Money(就业、利息和货币通论)》，创立宏观经济学。然而他首先是数学家，早在 1921 年他就发表了数学专著《Treatise on Probability(概率专论)》[0-13]。由于经济对社会发展的重要性，1969 年首设诺贝尔经济学奖，到 2011 年已评出 69 位得主。有趣的是，诺贝尔并没有设数学奖，但是在经济学奖获得者中确有不少数学家。例如：首届得主弗里希(Ragnar Frisch，1895-1973)和丁帕尔根(Jan Tinbergen，1903-1994)是计量经济学的创始人；1970 年得主萨缪尔森(Paul A. Samuelson，1915-2009)用严格数学理论总结数理经济学；1972 年得主阿罗(Kenneth J. Arrow，1921-)是数学博士。他基于四条公理推导出著名的“不可能定理”，指出“不可能每个人都完全自由，也不可能实现完全的自由经济”，震动了经济界；1973 年得主列昂节夫(Wassili Leontief，1906-1999)提出的“投入产出矩阵方法”至今为各国沿用；1975 年得主是苏联数学家康托诺维奇(Л·В·Канторович，1912-1986)。又如 1994 年得主纳什(John F. Nash Jr.，1928-)是“纳什均衡(Nash equilibrium)”的提出者，以他的生平经历为基础的传记电影《A Beautiful Mind(美丽心灵)》荣获 2002 年奥斯卡 4 项大奖[0-14]；1997 年并列得主默顿(Robert C. Merton，1944-)在建立和求解描述期权的偏微分方程中作出奠基性工作；2005 年由欧曼(Robert J. Aumann，1930-)和谢林(Thomas C. Schelling，1921-)共同获奖，他们运用博弈论(game theory)解释各式各样的冲突与合作，包括贸易纷争、帮派犯罪、政

治决策、劳资谈判等。欧曼是以色列裔美籍学者，他提出“无限重复博弈(infinitely repeated games)”理论，以数学为工具，分析在冲突发生时，己方与对方所拥有的各种选择；2007年得主赫维茨(Leonid Hurwicz，1917－2008)，马斯金(Eric Maskin，1950－)和梅耶森(Roger B. Myerson，1951－)也基于博弈论，提出激励机制设计理论。这样的例子不胜枚举。现在，“经济数学”[0-15]、“计量经济学”[0-16]、“金融数学”[0-17]、“精算数学”[0-18]等已成为应用数学的重要分支。马克思曾经强调说：“一门科学只有在成功地应用数学时，才算达到了真正完善的地步。”[0-19]可以说，经济学是传统社会科学领域中“成功地应用数学”的第一门学科。

阿罗 1921-

纳什 1928-

《美丽心灵》海报

接着是管理科学。现在，运筹学、概率统计、博弈论、系统仿真技术已成为其不可或缺的工具。

不仅如此，在人文社会科学中数学也逐渐发挥着作用。如数学和语言学结合，形成了“数理语言学”[0-20]，为机器翻译、密码破译、语言识别作出重大贡献。又如中外学者用统计推断和模糊数学等工具，判断考古发掘的文物年代，甄别文献的真伪等。在政治科学中，也出现了“数理政治学”[0-21]和“数理人口学”[0-22]的分支。在绘画音乐等艺术领域中，数学也发挥出日益重要的作用。借助电脑绘画、作曲等软件的问世和不断升级，都基于数学模型的建立与改进。

如同古代思想家那样，把数学看作思维方式和文化精神的观念也有所回归。1989年美国国家研究委员会发表专题报告：《人人关心数学教育的未来》，强调“数学提供了有特色的思考方式，……数学能使我们更好地了解我们生活在其中的充满信息的世界。”[0-23]为此，提出了各学科加强数学教学的

建议。

改革开放以来，我国正处于从计划经济向市场经济转轨的过程。一个文科大学毕业生，可能会改行，从事经济管理或行政管理工作。未来的各级领导人包括国家领导人中，很可能有相当数量从文科毕业生中涌现。从这个意义看，文科学生具备一定的数学素养也是很重要的。

我国的教育工作者非常关注文科大学生的数学教育。教育部直属的大学数学与统计教学委员会对文科数学教学的必要性和基本要求正在进行研究；近十年来，不少高校的同行已经编写出不少各具特色的文科数学教材[0-24]～[0-26]；作为文化素质教育的教材[0-27]和通识教材[0-28]也陆续问世。本书也是一种尝试。

"文科学生怎样学数学？"这是个仁者见仁、智者见智的问题。笔者认为，"赏析"二字或可概括。首先要"赏"，即学会欣赏。正像理工科学生应会欣赏文学艺术那样，文科学生也要会欣赏数学，欣赏数学的美，欣赏千百年来专业和业余数学家如何为攀登一座又一座数学高峰所作出的努力。这就要求教材写出思想，写出趣味，既有理论，又有应用。另一方面，既然是教材，就不同于科普讲座，学生还是要学到东西，这就要求"析"，即学会"分析"。这一点和理工科学生欣赏艺术不完全相同。理工科学生不一定要自己会画，自己作曲，自己演奏乐器，能欣赏就行；文科学生学数学，虽然学时不多，却应当掌握一些基本的数学知识和方法，并能学会初步应用，进而体会到数学中逻辑推理的力量和唯物辩证法的精神。

0.2　人文社会科学和文学艺术中数学的应用举例

下面列举人文社会科学中的几个数学问题。

0.2.1　优化问题

促进国民经济又好又快发展，加快建设节约型社会，这是我国长期发展的指导方针。如何使得成本最小，效益最大，时间最省，这就是优化问题，数学可以发挥重大作用。

【例 0-1】　某企业亏损 1050000 元。经产业结构调整，确定只生产两种支柱产品，设其产量分别为 x 和 y 件，成本函数包括固定成本 250000 元以及可变

成本 $x^2+2xy+y^2$ 元。设两种产品的销售价格为 P_1 元和 P_2 元，根据市场调查，两种产品的需求函数分别是

$$x=2650-P_1,\ y=1040-0.25P_2$$

如何确定两种产品的生产水平，使得总利润最大？如果生产销售周期为 1 年，试问一年后能否扭亏为盈？

解：不难建立利润 L 对两种产品数量的依赖关系：

$$\text{总利润}\ L=\text{总销售收入}\ R-\text{总成本}\ C$$

$$R=xP_1+yP_2=x(2650-x)+y(4160-4y)$$

$$C=x^2+2xy+y^2+250000$$

所以

$$\begin{aligned}L&=x(2650-x)+y(4160-4y)-(x^2+2xy+y^2+250000)\\&=-2x^2-2xy-5y^2+2650x+4160y-250000\end{aligned}$$

从给定条件看，有如下约束：$0\leqslant x\leqslant 2650$，$0\leqslant y\leqslant 1040$，而且 x 和 y 是整数。

不妨试凑，若取 $x=2650$，$y=0$，或取 $x=0$，$y=1040$，都得到 $L<0$，显然不可取；若取 $x=300$，$y=200$，得到 $L=877000$ 元；用“穷举法”，一共有$2651\times1041=2759691$ 种可能。但是学了微分学方法，很容易得到，当 $x=505$ 件，$y=315$ 件时，总利润 L 最大，达到 1074325 元，可以扭亏为盈。（具体解法见第六章 6.1.4）

【例 0-2】 A，B 两个产地分别生产同一批规格产品 12 千吨和 8 千吨。而甲、乙、丙三地分别需要该产品 8 千吨、6 千吨和 6 千吨。各产销地之间每千吨的运价表如表 0.1 所示。怎样确定运输方案，使得总运费最少？

表 0.1

万　元	到　甲	到　乙	到　丙
从 A	4	5	6
从 B	5	2	4

解： 设从 A 运到甲地 x 千吨，运到乙地 y 千吨，运到丙地 $12-x-y$ 千吨，从 B 运到甲地 $8-x$ 千吨，运到乙地 $6-y$ 千吨，运到丙地 $8-(8-x)-(6-y)=x+y-6$ 千吨，则总运费

$$\begin{aligned} F &= 4x+5y+6(12-x-y)+5(8-x)+2(6-y)+4(x+y-6) \\ &= -3x+y+100 \end{aligned}$$

约束条件为：$x\geqslant 0$，$y\geqslant 0$；$12-x-y\geqslant 0$，$8-x\geqslant 0$，$6-y\geqslant 0$，$x+y-6\geqslant 0$。

这里 x 和 y 都是实数，有无限种可能性，“穷举法”就更难应用了。学了线性规划后，可以求得当 $x=8$，$y=0$ 时，总运费 $F=76$ 万元最省。（具体解法见第六章 6.2.3）

0.2.2　问卷调查中敏感问题如何设计

在社会科学研究中，常用问卷调查法。如何设计问卷，有不少学问。过于简略，达不到调查目的；过于繁琐，被调查人会感到厌倦而敷衍了事，同样达不到调查目的。特别对于一些敏感问题，如果提问不当，例如“你作过弊吗？”，“你是否逃过税？”，“你是否喜欢我教的这门课？”，往往会使被调查人难堪，回答言不由衷，甚至拒绝回答。

笔者曾看到一份调查归国留学人员的问卷。有一道“五选二”的选择题是：“你回国的主要动因是：(A) 报效祖国的情怀；(B) 中国提供发展机遇；(C) 与家人团聚；(D) 工资待遇更高；(E) 文化与饮食习惯更适应。”这就使人很难选择。其实多数留学人员归国是多种因素综合的。如果不选(A)，有“思想落后”之嫌；如果选(D)，则受“唯利是图”之责。尽管是不记名调查，但当调查显示多数人未选(A)，有人选(D)，也会使被调查人感到压力。

借助数学中概率统计方法，可以巧妙设计敏感问题的问卷。

【例 0-3】 调查某大型企业 200 位高管人员逃税的比例。有理由假定，大多数人不愿意承认逃税。如果问“你是否逃过税？”显然难得到真实结果。准备一副扑克牌，让这 200 位高管随机抽。规定抽到黑桃、红心和方块时，回答“问题 1：你逃过税吗？”（“是”或“非”）；抽到梅花时回答“问题 2：你从不逃税吗？”（“是”或“非”）。在无记名回答过程中调查者并不知道被调查者抽到什么牌，被

调查者也不必解释。调查结果表明，有 60 张问卷的回答“是”。怎样通过回收问卷来估计逃税的比例？

解： 假定逃税的比例为 $a\%$，回答“是”的概率为：

$$P(\text{是}) = \frac{3}{4}a + \frac{1}{4}(1-a) = \frac{60}{200}$$

于是解出 $a=0.1$，即大约有 10% 的高管逃税。为什么被调查者愿意真实回答？因为没有顾虑，调查者无法从结果中判断谁逃税。如果把抽到问题 1 的概率记为 p，那么抽到问题 2 的概率就是 $1-p$，于是由 $pa+(1-p)(1-a)=\frac{m}{n}$，解出

$$a = \frac{1}{2p-1}\left(\frac{m}{n}-1+p\right), \quad p \neq \frac{1}{2}$$

这里 n 是调查对象的人数，m 是回答“是”的人数。

上面两个问题是同一件事的相反提法，还是可能引起被调查者的戒备；此外，上述方法有一个前提，即抽到“问题 1”和“问题 2”的概率不能都等于 $\frac{1}{2}$，否则解不出 a。

可以把“问题 2”换为与“问题 1”毫不相关的问题，也可以抽到两个问题的概率都等于 $\frac{1}{2}$。

【例 0-4】 某年级调查 160 名大学生无故缺课的人数。规定抽到黑桃和红桃的学生回答“问题 1：你曾无故缺课吗？”（“是”或“非”）；抽到方块和梅花的学生回答“问题 2：你的学生证号码最后一位是奇数吗？”（“是”或“非”）。结果有 55 人回答“是”。

假定“曾无故缺课”的比例为 $a\%$，由于“学生证号码最后一位是奇数”的概率是 $\frac{1}{2}$，故回答“是”的概率为：$P(\text{是}) = \frac{1}{2}a + \frac{1}{2}\times\frac{1}{2} = \frac{55}{160}$，解出 $a=\frac{3}{16}=18.75\%$，即大约有近 19% 的学生曾无故缺课。这里抽到两个问题的

概率都等于 $\frac{1}{2}$，依然可解。

注意：问题 2 回答“是”的概率必须已知，记为 $b\left(这里\ b=\frac{1}{2}\right)$，则由 $pa+(1-p)b=\frac{m}{n}$ 解出

$$a=\frac{1}{p}\left(\frac{m}{n}-(1-p)b\right)$$

0.2.3　文学作品真伪的鉴定

基于数理统计方法，通过电脑鉴别文学作品真伪已经成为有力手段。20 世纪 70 年代中期，英国剑桥大学的两位师生用此法侦破伪造莎士比亚作品的奇案而震动了西方文学界。自莎士比亚逝世后，经常有人“发现”莎翁“新作”，真假难辨，从而牟利。这两位师生对一家出版商出版的莎士比亚“新作品”，用计算机对它的修辞和结构与公认的莎士比亚作品进行分析和比较。发现某些修辞和用语莎翁从不采用，某些句型与结构也同莎翁惯用方式不同。他们将用计算机分析出来的大量证据向法院起诉，使伪造莎士比亚作品的出版商哑口无言，只得承认作伪。这项研究成果引起了国内外学者的重视，纷纷利用计算机来分析研究别的古典文学作品。20 世纪 80 年代，中国学者开始用此方法研究《红楼梦》的作者，取得可喜成果。

《红楼梦》著作权的问题多年来一直争论不休。目前红学家们普遍的看法是：前八十回为曹雪芹所作，后四十回为高鹗所续。

1980 年到 1986 年，美国威斯康星大学华裔学者陈炳藻先生先后发表多篇论文，他将《红楼梦》一百二十回本按顺序编成三组，每组四十回。并将《儿女英雄传》作为第四组进行比较研究，从每组中任取 8 万字，分别挑出名词、动词、形容词、副词、虚词这五种词，运用数理语言学，通过计算机程序对这些词进行编排、统计、比较和处理，进而找出各组相关程度。结果发现《红楼梦》前八十

回与后四十回所用的词汇正相关程度达 78.57%，而《红楼梦》与《儿女英雄传》所用词的正相关程度是 32.14%。由此推断得出前八十回与后四十回的作者均为曹雪芹一人的结论。

但是，不少人认为陈取的样本不足，方法也值得改进。1987 年，正在威斯康星大学访问的复旦大学数学系李贤平教授的工作引人注目。他运用计算机技术中的模式识别法和统计学家使用的探索性数据分析法，把《红楼梦》一百二十回本作为一个整体，以 47 个虚字为识别特征，对它们在书中各回的出现频率进行统计分析。他在计算机前工作了数百小时，绘制了 300 多张图纸，通过计算机将使用频率绘成图纸，再根据图纸反映出的表明不同创作风格的星云状和阶梯状图形，提出了震惊红学界的《红楼梦》成书过程新观点，证明了《红楼梦》各回写作风格具有不同的类别，各部分实际上是由不同作者在不同时期里完成的。李贤平认为："《红楼梦》前八十回是曹雪芹据《石头记》增删而成，其中插入了他早年著的《金瓶梅》式小说《风月宝鉴》，并增写了具有深刻内涵的许多内容。《红楼梦》后四十回是曹家亲友在曹雪芹全书尚未完成就突然去世之后，搜集整理原稿并加工补写而成。程伟元将全稿以活字版印刷刊行。高鹗校勘异文补遗订讹"。他的这一看法否定了被红学界一直视为曹雪芹作前八十回，高鹗续后四十回的定论[0-29]。

0.2.4 音乐中的数学与电脑辅助作曲

人们对数学与音乐之间联系的研究有着悠久的历史。最早可以追溯到公元前六世纪，当时毕达哥拉斯学派把音乐解释为宇宙的普遍和谐，并阐明了弦振动产生的乐音与弦长的关系。比如两根弦长比为 2∶1，则短弦与长弦音高差 8 度；若 3∶2，则短弦音比长弦音高 5 度，也就是说，谐声是由长度成整数比的同样绷紧的弦发出的。毕氏理论长期在西方音乐界占据了统治地位。欧几里得也写过音乐方面的著作。欧洲文艺复兴后，许多大数学家如牛顿（Newton, 1642-1727）、欧拉（Euler, 1707-1783）都研究过音乐，撰写过论文；特别是傅立叶（Fourier, 1768-1830）证明了所有的乐声，不管是器乐还是声乐，都可以用数学式来表达和描述，而且证明了这些数学式是简单的周期正弦函数的和[0-30]。

欧拉　1707-1783

牛顿　1642-1727

傅立叶　1768-1830

中国明代数学家、律学家、历学家朱载堉(1536-1611)创建了十二平均律。这是音乐学和音乐物理学的一大革命,也是世界科学史上的一大发明。朱载堉为明太祖朱元璋九世孙,他在总结前人乐律理论的基础上,通过精密计算和科学实验,成功地发现十二平均律的等比数列规律,称其为“密率”,在其《律学新说》卷一中,他概述了十二平均律的计算方法:“创立新法:置一尺为实,以密率除之,凡十二遍。”在《律吕精义·内篇》卷一中,他对十二平均律做了描述:“盖十二律黄钟为始,应钟为终,终而复始,循环无端。……是故各律皆以黄钟……为实,皆以应钟倍数 1.059463……为法除之,即得其次律也。”[0-31]

朱载堉　1536-1611

(为了阅读方便,引文中用阿拉伯数字代替了原文中的汉字数字。)完全与现代音乐中通用的十二平均律相合。用现代数学符号表达:如果某根弦长为 L_0,振动发出一个音(例如 C),则高一个半音($C^{\#}$)的弦长为

$$L_1 = \frac{L_0}{\sqrt[12]{2}} \approx 0.9439L_0$$

高 i 个半音的弦长为

$$L_i = \frac{L_0}{\sqrt[12]{2^i}} \approx (0.9439)^i L_0$$

($i=2, 3, \cdots, 11$, 对应于 D, $D^{\#}$, E, F, $F^{\#}$, G, $G^{\#}$, A, $A^{\#}$, B)

容易看到，下一个 C 音的弦长为 $L_{12}=\dfrac{L_0}{\sqrt[12]{2^{12}}}=\dfrac{L_0}{2}$，正好是原弦长的一半。

在朱载堉的时代，计算工具相当落后。他硬是在自制的 81 档双排大算盘，开两次平方再开 1 次立方，求出 $\sqrt[12]{2}$ 之值，精确到 25 位有效数字。相传这一关键数字，被传教士通过丝绸之路传到欧洲，德国伟大作曲家巴赫(Johann S. Bach, 1685-1750)根据它制造出了世界上第一架钢琴。德国物理学家赫尔姆霍茨(Helmholtz, 1821-1894)说过："在中国人中，据说有一个王子叫朱载堉，他在旧派音乐家的大反对中，倡导七声音阶。把八度分成十二个半音以及变调的方法，也是这个有天才和技巧的国家发明的。"


巴赫　1685-1750


赫尔姆霍茨　1821-1894


李约瑟　1900-1995

朱载堉多才多艺。除乐律上的突出贡献外，他首创"舞学"，为舞学制定了大纲，奠定了理论基础，绘制大量舞谱和舞图；他对天文历法开拓了新的领域，求出了计算回归年长度值的公式。1986 年，专家们用现代高科技的测量手段对朱载堉的计算结果进行了验证，惊奇地发现，朱载堉计算的 1554 年的长度值与我们今天计算的仅差 17 秒钟！他还是我国历史上第一个精确计算出北京地理位置(北纬 39°56′，东经 116°20′)的人。正因为如此，英国学者李约瑟(Joseph T. M. Needham, 1900-1995)不仅把朱载堉称之为"世界上第一个平均律数字的创建人"，更称他为"中国文艺复兴式的圣人"。

朱载堉还擅长写曲，特别是对人世间的丑恶现象，嬉笑怒骂，进行鞭挞。例如有一首《黄莺儿·戒得志》："君子失时不失相，小人得志肚儿胀。昨日无钱去做贼，今日有奶便呼娘；真臭物，实荒唐。君不见，街前骡子学马走，到底还是驴

儿样！”足可与关汉卿媲美！

中国古代音律学着重研究弦律、管律和钟律，文献之多，在世界文化史中写下辉煌篇章[0-32]。与弦、管、钟相比，更清晰表达乐理的乐器是钢琴。在钢琴的键盘上，从一个 C 键到下一个 C 键就是音乐中的一个八度音程（如图 0.1 所示）。其中共包括 13 个键，有 8 个白键和 5 个黑键，而 5 个黑键分成 2 组，一组有 2 个黑键，一组有 3 个黑键。

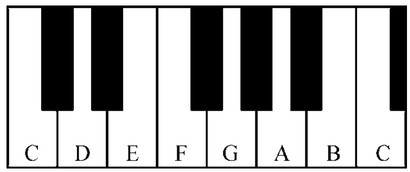


图 0.1

实验测出，1、2、3、4、5、6、7、$\dot{1}$等音阶是由等比数列规定的。再来看图 0.1，显然这个八度音程被黑键和白键分成了 12 个半音，由于下一个 C 键发出乐音的振动次数（即频率）是第一个 C 键振动次数的 2 倍，设相邻两个音频之比为 r，显然满足 $r^{12}=2$，解出 $r\approx 1.05946$。物理测量出中央 C 的频率为 261.6 Hz，从中央 C 往上 8 度的 12 个音高频率测量结果（取整数）如表 0.2 所示。

表 0.2

序　号	1	2	3	4	5	6	7	8	9	10	11	12	13
音名	C	$C^{\#}$	D	$D^{\#}$	E	F	$F^{\#}$	G	$G^{\#}$	A	$A^{\#}$	B	C
频率/Hz	262	277	294	311	330	350	370	392	415	440	466	494	523
琴键	白	黑	白	黑	白	白	黑	白	黑	白	黑	白	白

不难验证，前后两个序号音频（相差半音）之比恰为 $r\approx 1.06$；序号 13 的 C 音频恰为中央 C 音频的两倍。

物理实验显示，敲击音叉后振动发声，频率表示音高，振幅表示音量，随着时间的变化，形成一个正弦波。傅立叶发现，自然界的声音可以视为一系列不同频率、不同振幅、不同相位的正弦波的叠加。其中最低的频率称为“基音频率”，决定了音高；其他频率称为“泛音频率”，决定了音色。一个乐器发出的声音，其泛音频率通常是基音频率的整数倍。例如一个小号发出上面表上的 A 音，其基音频率当然就是 440 Hz，其泛音频率有 880 Hz、1320 Hz、1760 Hz、……等。

反过来，由一个纯粹的图像出发，我们只要对它进行适当的分段，形成适当的小节，并在曲线上选取适当的点作为音符的位置所在，那么就可以作出一节节

希林格 1895-1943

的乐曲。20世纪20年代美国哥伦比亚大学的数学和音乐教授约瑟夫·希林格(Joseph Schillinger，1895-1943)曾经把《纽约时报》的一条起伏不定的商务曲线描述在坐标纸上，然后把这条曲线的各个基本段按照适当的、和谐的比例和间隔转变为乐曲，最后在乐器上进行演奏，结果发现这竟然是一首曲调优美、与巴赫的音乐作品极为相似的乐曲！这位教授毕生致力于发展一套名为“基于数学的艺术”的大师教材(master text entitled The Mathematical Basis of the Arts)。他逝世后这套大师教材被称为“希林格音乐作曲系统(The Schillinger System of Musical Composition)”。根据这一系统的准则，所有的音乐杰作都可以转变为数学公式。许多作曲家深受其影响[0-32]。20世纪美国最出色的作曲家之一乔治·格什温(George Gershwin，1898-1937)在创作著名歌剧《Porgy and Bess(波吉与贝丝)》时，综合了爵士音乐、古典音乐和黑人民歌等不同风格，就运用了这一作曲系统。电脑问世并迅速普及后，在希林格系统基础上，涌现和发展了电脑作曲方法和软件，成为方便的作曲辅助工具[0-33]。

格什温 1898-1937

音乐和数学常有巧妙的结合。注意图0.1中的13个键，其中8个白键，5个黑键，3个黑键为1组，2个黑键为另1组。这里出现的数字1、1、2、3、5、8、13恰好就是著名的斐波那契数列$f_n=1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, \cdots\cdots$中的前几个数。

斐波那契 1170-1240

800年前，意大利数学家斐波那契(Fibonacci，1170-1240)提出了著名的“兔子生兔子的问题”[0-34]：有一个人把一对小兔子放在四面围着的地方，假定：① 每对小兔子长大到能生育需1个月；② 每个月一对大兔子生下一对小兔子；③ 兔子不死。问题是，一年后共有多少对兔子？

显然，开始时为 1 对小兔子；一个月后，还是 1 对，不过长成了大兔子。两个月后，有 2 对（一对大，一对小）；3 个月后，有 3 对（2 大 1 小）；总之，以后每个月的兔子对数可分成两部分：一部分是前一个月的兔子对数；另一部分是新生的兔子对数，正好等于前两个月的兔子对数。即

$$f_1=1,\ f_2=1,\ f_3=2,\ f_4=3,\ \cdots,\ f_n=f_{n-2}+f_{n-1}\quad (n\geqslant 3)$$

显然，一年后共有 $f_{13}=233$ 对兔子。

斐波那契数与音乐的结合还有许多例子。1964 年，美国数学界的一本学术季刊《斐波那契季刊》引证出，巴赫著名的《Fugue（赋格曲）》之所以优美，就因为他为之倾注了毕生心血的和声，是一组斐波那契数列[0-35]。英国一个室内乐团以演绎巴赫篇章而著名，其名称就叫“斐波那契乐团”。

斐波那契数列还有个奇妙的性质：前项与后项之比值逐渐逼近“黄金分割数”

$$\begin{aligned}\frac{f_{n-1}}{f_n}&=0.5,\ 0.666,\ 0.6,\ 0.625,\ 0.6154,\ \cdots\rightarrow\frac{\sqrt{5}-1}{2}\\&=0.618033988\cdots\end{aligned}$$

这一性质使得斐波那契数在优化技术中发挥重要作用[0-36]。人们早就发现黄金分割在古典建筑和绘画中的神奇作用。几乎每一座欧洲古典建筑都是对于黄金分割的颂歌。

斐波那契数列的发现和应用说明了一个重要现象：数学和其他基础科学的某些偶然发现，当时未见得有多少意义，若干年代后却发现有重大应用价值。

0.3　文理兼通的中外数学家

1981 年诺贝尔医学生理奖获得者罗杰·史贝尼（Roger W. Sperry，1913-　）揭示，人脑的左右半球有不同的功能[0-37]。数学物理与逻辑思维由人的大脑左半球处理，而文学音乐与形象思维则由人的大脑右半球控制。历史上有不少人两者都得到均衡发展，和谐统一。

例如19世纪英国童话《Alice's Adventure in Wonderland(爱丽丝梦游奇境记)》的作者刘易斯·卡洛尔(Lewis Carroll,1832-1898)其实是一位数学家。他是给爱女讲故事时构思出这篇名著的。当时维多利亚女王极为赏识,颁发懿旨:卡洛尔的下一著作必须首先让她御览。想不到下一部新作竟是一篇关于行列式的数学论文,使她哭笑不得。[0-38]

中国文理兼通的数学家不乏其人。例如上海交通大学数学系教授、安徽师范大学数学系主任、《中学数学教学》第一任主编雷垣(1912-2002),1931年获大同大学理学士学位后,在他中学同窗好友傅雷的影响下,又考入上海音乐学院。之后到美国密歇根大学继续攻读数学,兼修音乐。他是钢琴大师傅聪的启蒙教师之一。他创作了许多音乐作品。图0.2所示是他在20世纪40年代为归国留学生的团体欧美同学会创作的会歌《谊歌》,中英文歌词和谱曲均出自他一人之手,可见其不凡的中英文文学修养和高超的音乐造诣。

图0.2

苏步青　1902-2003

又如原复旦大学校长、中科院院士苏步青教授(1902-2003)，当代杰出的数学家、教育家，在科研和教学上取得了令世人叹服的光辉业绩。早在做研究生时，他就发现四次(三阶)代数锥面，被学术界誉称为“苏锥面”；以后在“射影曲线论”、“射影曲面论”、“高维射影空间共轭网理论”、“一般空间微分几何学”和“计算几何”等方面都取得世界同行公认的成就。苏步青不仅是数学大师，还是一位文学大师，称得上是写作大家和诗人。他从小酷爱古诗文，13 岁学写诗。读初小时常骑在牛背上诵读《千家诗》等。几十年来，他与诗为伴，与诗书同行。苏步青不仅读诗，更有作诗兴趣，一生中写了近千首诗作。在他 96 岁高龄时出版的《苏步青业余诗词钞》[0-39]，共收近体诗 444 首，词 60 首，由苏老手写影印，其中 1931-1949 年早期作品 191 首，包括词 47 首。从中我们可以领略苏老 60 年间的学术生涯和诗书技艺折射的光芒，富有时代气息。


業餘詩詞鈔　浙江平陽蘇步青撰
《前篇　詩鈔》
感懷寄友　一九三一年由日歸國船上
渡口雲煙海鳥飛，江邊春色認依稀。十年海上君休笑，贏得鬢絲和布衣。
村居　一九三八年
江村落木初，修竹蔭孤廬。麥秀風千畝，牛歸雨一蓑。客談天下事，潮送故人書。亂定知何日，

0.4 思考与练习

1. 通过阅读参考资料，试图提出一两个人文社会科学或音乐艺术中的数学问题(不必解出，当然能解更好)。

2. 找一个问题用两种方法进行问卷调查，可以是已知答案的问题(例如班级中女同学的比例，或在江苏、浙江上中学的同学的比例等)，从而比较调查精度。

3. 通过网站搜索用数学研究音乐的一两项新成果。

4. 中国“导弹之父”钱学森先后在上海交通大学和美国麻省理工学院学习

工程,在数学物理和力学领域都有极高造诣。他之所以能在科学上取得突出成就,与他有一位做艺术家的妻子不无关系。钱夫人蒋英是中央音乐学院的声乐教授,是著名军事理论家蒋百里的女儿,蒋百里曾促成过哲学家罗素和诗人泰戈尔的访华。蒋英从小就跟随父亲游历欧洲,并在德国柏林音乐学院留学十年,研究德国艺术歌曲。与蒋英结合以来,钱学森早已习惯了在琴声与歌声的伴奏下读书、思考,有时他也会哼唱一曲,他说他在科学技术上的很多新思路、新想法,都是在和蒋英谈音乐谈艺术的过程中产生出来的灵感。钱学森主张,搞科学研究的人应该拥有艺术的广阔思维方法,这样可以避免死心眼,想问题能宽一点,活一点。试通过网站或其他资料找到钱老讲话的原文,并发表感想。

5. 你认为文科学生赏析数学的意义如何?对这门课的教学方法有何建议?

参考文献或网站

[0-1] 莫里斯·克莱因(Morris Kline):《古今数学思想(Mathematical Thought from Ancient to Modern Times)》(江泽涵,张理京,张锦炎,申又枨,朱学贤,钱敏平,邓东皋,丁同仁,刘西垣,叶其孝,庄圻泰,万伟勋,石生明,张顺燕,姜伯驹,孙树本,章学诚,程民德,张恭庆,聂灵沼,吴光磊译),上海科学技术出版社(2005).

[0-2] 老子:《道德经》第四十二章,河南人民出版社(2008).

[0-3] 《辞源》,商务印书馆(1986).

[0-4] 柏拉图:《理想国(Republic)》(段至诚译),中国对外翻译出版公司出版(2006).

[0-5] 波耶(C. B. Boyer):《微积分概念发展史(The History of the Calculus and its Conceptual Development)》(唐生译),复旦大学出版社(2007).

[0-6] 《周易》(郭彧译注),中华书局(2006).

[0-7] 欧几里得:《几何原本(Elements)》,http://aleph0.clarku.edu.

[0-8] 《周髀算经》(江晓原,谢筠译注),辽宁教育出版社(1996).

[0-9] 张苍:《九章算术》(曾海龙 译解),重庆大学出版社(2006).

[0-10] 爱德华兹(H. M. Edwards):《费马大定理(Fermat's Last Theorem)》,科学出版社(2011).

[0-11] 柯朗,罗宾(R. Courant, H. Robbins),《什么是数学?对思想和方法的基本研究(What is Mathematics? An Elementary Approach to Ideas and Methods)》(左平 张饴慈译),复旦大学出版社(2007).

[0-12] 中共中央马克思恩格斯列宁斯大林著作编译局:《马克思恩格斯全集》第33

卷,人民出版社(1973).

[0-13] 斯基德尔斯基(R. Skidelsky):《凯恩斯传(John Maynard Keynes)》(相蓝欣 储英译),三联书店(2006).

[0-14] 电影《美丽心灵(A Beautiful Mind)》,梦工厂(Dream Works Picture)(2002).

[0-15] 赵胜民:《经济数学》,科学出版社(2005).

[0-16] 张寿,于清:《计量经济学》,上海交通大学出版社(1984).

[0-17] 吴岚,黄海:《金融数学引论》, 北京大学出版社(2005).

[0-18] 鲍尔斯(N. L. Bowers),《精算数学(Actuarial Mathematics)》(余跃年 郑韫瑜译),上海科技出版社(1996).

[0-19] 中共中央马克思恩格斯列宁斯大林著作编译局:《回忆马克思》,人民出版社(2005).

[0-20] 方立:《数理语言学》,汉语教学网 www.pop.com.cn.

[0-21] 卢卡斯(W. F. Lucas):《应用数学丛书第 2 卷:政治及有关模型(Modules in Applied Mathematics,Vol II:Political & Related Models)》(王国秋 刘德铭译),国防科技大学出版社(1996).

[0-22] 内森·凯菲茨(N. Keyfitz):《应用数理人口学(Applied Mathematical Demography)》(郑真真 顾大男 任强译),华夏出版社(2000).

[0-23] 美国数学科学教育委员会(Mathematical Sciences Education Board):《人人关心数学教育的未来(Everyone is Concerned About the Future of Mathematical Teaching)》,世界图书出版公司(1993).

[0-24] 姚孟臣等:《大学文科高等数学》,高等教育出版社(1997).

[0-25] 华宣积,谭永基,徐惠平:《文科高等数学》,复旦大学出版社(2000).

[0-26] 汪国柄:《大学文科数学》,清华大学出版社(2005).

[0-27] 顾沛:《数学文化》,高等教育出版社(2008).

[0-28] 谭永基,俞红:《现实世界的数学视角与思维》,复旦大学出版社(2010).

[0-29] 李贤平:《红楼梦成书新说》,《复旦学报》社会科学版(第 5 期),复旦大学出版社(1987).

[0-30] 克莱因(M. Kline):《西方文化中的数学(Mathematics in Western Culture)》(张祖贵译),复旦大学出版社(2005).

[0-31] 胡企平:《中国传统管律文化通论》,上海音乐出版社(2003).

[0-32] http://en.wikipedia.org/wiki/Joseph_Schillinger.

[0-33] 张弛:《电脑作曲教程》,科学出版社(2006).

[0-34] 斐波那契(Fibonacci):《算盘书(Liber abaci)》,http://mathworld. wolfram. com/ Fibonacci Number. html

[0-35] 《斐波那契数及艺术、建筑和音乐中的黄金分割(Fibonacci Numbers and Golden Section in Art,Architecture and music)》,www. mcs. surrey. ac. uk..

[0-36] 华罗庚:《优选学》,科学出版社(1981).

[0-37] 史贝尼(R. W. Sperry):《神经科学研究项目:两半侧脑的分工(Lateral Specialization in the Surgically Separated hemispheres, Thirs Neurosciences Study Program)》, Cambridge: MIT(1974).

[0-38] 埃克塞雷(C. E. Eckersley):《基本英语(Essential English)》,Longman 出版社(1970).

[0-39] 苏步青:《苏步青业余诗词钞》,群言出版社(1994).

第一章

数与空间

1.1 数的故事

从数学诞生之日起,数学的第一个对象就是“数”。中小学数学已经介绍过自然数、小数、分数、有理数、无理数、实数、虚数、复数等不同类型的数。如何准确而有效地记录一个数,人类花了数千年进行探索。从中国殷墟甲骨以及埃及、印度、希腊、两河流域等国家和地区发现的文物中都有对数字的记载。各种记数方法看起来只是符号不同,对数学的发展却有极其重要的意义。现在世界通用的“阿拉伯数字”其实是印度人首先采用的,由于经过阿拉伯地区传入欧洲,才被欧洲人这样称呼。“阿拉伯数字”确实简洁有效。例如下面的算式

$$12345 \times 67890 = 838102050$$

用汉字描述为:“一万二千三百四十五乘以六万七千八百九十等于八亿三千八百一十万二千零五十”;用英文描述则是:“twelve thousands and three hundreds and forty five times sixty seven thousands and eight hundreds and ninety equals eight hundred and thirty eight millions and one hundred and two thousands and fifty”。如果写一篇数学或科学技术文章用到大量数字及运算,阿拉伯数字的优越性就更明显了。数学符号的简洁程度对数学的发展也至关重要。以最简单的四则运算符号为例,+号和−号最早用于运输箱子重量的超亏,1841 年才首次出现于德国数学家魏德曼(J. Widmann, 1460-?)一份手稿中;×号是英国人奥德雷特(W. Oughtred, 1574-1660)于 1631 年首创,实际上是个倾斜的

＋号，表示另一种加法；÷号则由瑞士人雷恩(J. H. Rahn)于 1659 年提出。这四个符号沿用至今。哲学上人们常讨论“内容”与“形式”的关系。一般说来，内容比形式更重要；但在一定条件下，形式会对内容起决定性作用。《数学符号史》[1-1]一书系统介绍了人类社会几千年来的数学符号史，作者认为，数学符号像一颗颗耀眼的明珠，镶嵌在数学的宏伟殿堂上，表明数学的概念、运算、关系和推理，使数学思维过程准确、概括、简明，从而更容易揭示数学对象的本质。

1.1.1　科学记数法

20 世纪科学技术迅猛发展，人们研究对象的数字尺度差别很大。比如物理学既研究分子、原子、电子、夸克，又研究太阳系、银河系、河外星系。为了用数字统一描述，人们采用了以十进位为基础用数量级表示的科学记数法。规则如下：

$$A = a \times 10^n \tag{1.1}$$

其中 A 是任意正数，a 是一个满足 $1 \leqslant a < 10$ 的数，其数字排列与 A 相同；n 是“数量级”，也称“幂”或“指数”，取法如下：

当 $1 \leqslant A < 10$，$n = 0$；

当 $A \geqslant 10$，$n = A$ 的位数减 1；

当 $0 < A < 1$，$n = A$ 的小数点后第一个非零数字前 0 的个数，包括小数点前那个 0 在内。例如：$9 = 9 \times 10^0$；$12345 = 1.2345 \times 10^4$；$0.00567 = 5.67 \times 10^{-3}$。

表 1.1 中列出了有专门名称的前 6 个大数，每个数都是它前一个的 1000 倍。超过万亿的专名极少使用。每秒数一个数，要花一个多星期才能数到一百万，半辈子才能数到十亿。就算你与宇宙同寿，也数不到一百亿亿(a quintillion)。

表 1.1

中文数字(英文)	阿拉伯数字	幂形式	数到此数所需的时间(每秒 1 次，持续不断)
一(One)	1	10^0	1 秒
千(Thousand)	1000	10^3	17 分钟
百万(Million)	1000000	10^6	12 天
十亿(Billion)	1000000000	10^9	32 年

续 表

中文数字(英文)	阿拉伯数字	幂形式	数到此数所需的时间 (每秒1次,持续不断)
万亿(Trillion)	1000000000000	10^{12}	32000年(超过地球文明的历史)
千万亿(Qudrillion)	1000000000000000	10^{15}	3200万年(超过地球上有人类的历史)
百亿亿(Quintillion)	1000000000000000000	10^{18}	320亿年(超过宇宙的年龄)

更大的数叫做 sextillion(10^{21})、septillion(10^{24})、octillion(10^{27})、nonillion(10^{30})和 decillion(10^{33})。

一旦掌握了科学记数方法,就能方便地表示庞大的数。例如一勺土里所含微生物的大致数目(10^8);地球上所有沙滩上的沙粒数目(可能是10^{20});地球上所有生命的数量(10^{29});地球上所有生命所含原子的数目(10^{41});太阳里面的原子核数目(10^{57});整个宇宙中的基本粒子(电子,质子,中子)总数(10^{80})[1-2]。大家知道,美国有一家著名的搜索引擎公司"谷歌(google)"[1-3]。1 google$=10^{100}$,超过整个宇宙中的基本粒子总数。该公司取此域名,正显示其海量的搜索能力。同样,对于极小的数也可方便地表达。例如:1纳米(nm)$=10^{-9}$ m;1个氢原子质量$=1.674\times10^{-27}$ kg;1个电子质量$=9.10\times10^{-31}$ kg;1顶夸克的衰变时间$=10^{-24}$秒;等等。

交通大学杰出校友、中国"导弹之父"钱学森院士是享誉海内外的科学巨匠,当之无愧的"两弹一星"元勋。他在晚年运用系统科学思想,在许多学科和领域提出了一系列重要创见。他曾按空间尺度提出"五观"的宇宙物质结构[1-4]。


钱学森 1911-2009

(1)"渺观":$<10^{-34}$ m;

(2)"微观":$10^{-34}\sim10^{-8}$ m;

(3)"宏观":$10^{-8}\sim3\times10^{17}$ m;

(4)"宇观":$3\times10^{17}\sim10^{29}$ m;

(5)"胀观":$>10^{29}$ m

1.1.2 二进位制记数

可能因为人类长10个手指,所以绝大多数民族记数采用十进位制。宋朝人

王应麟编写的中国古代儿童启蒙书《三字经》[1-5]中，就有十进位的口诀："知某数，识某文。一而十，十而百。百而千，千而万"。1.1 式就基于十进位制。当然也有少数例外。中国曾采用 1 斤＝16 两；英国至今沿用 1 英尺(foot)＝12 英寸(inches)，1 码(yard)＝3 英尺，1 英里(mile)＝5280 英尺，1 加仑(gallon)＝4.55 升(liter)，1 磅(pound)＝16 盎司(ounce)等英制单位。实际上，可以用任何自然数作为进位基数。二进制就是除十进制外最重要的进位制。

莱布尼茨 1646-1716

追根溯源，最早的二进位制出自中国的八卦。18 世纪德国哲学家、数学家莱布尼茨(Leibniz, 1646－1716)从他的传教士朋友鲍威特寄给他的拉丁文译本《易经》中，读到了八卦的组成结构，惊奇地发现只有两个基数：0 和 1，即"阴爻╍ "和"阳爻━"。于是，他把 17 世纪在欧洲已经萌芽的二进位制进行了完善。他形象地用 1 表示上帝，用 0 表示虚无，上帝从虚无中创造出所有的实物，恰好在数学中用 1 和 0 表示了所有的数。

从理论上分析，在所有可能的进位制中二进位制的基数最少，一共只有两个：0，1。更大的数则逢二进一，任何数都用一行 0 和 1 表示。而十进位需要 10 个基数：0，1，2，3，4，5，6，7，8，9，更大的数则逢十进一。表 1.2 是若干十进制数的二进制表示。

表 1.2

十进制	0	1	2	3	4	5	6	7	8	9	10	16
二进制	0	1	10	11	100	101	110	111	1000	1001	1010	10000

在二进位制中的加法和乘法也特别简单。

(1) 二进制加法：$0+0=0$，$0+1=1$，$1+0=1$，$1+1=0$(进位为 1)；

(2) 二进制乘法：$0\times0=0$，$1\times0=0$，$0\times1=0$，$1\times1=1$

二进制数字和十进制数字可以方便地互换。

十进位制的数有十个数字 0、1、2、…、9，进位的规则是"逢十进一"。十进制数的一般形式为

$$A = a_1\times10^{n-1}+a_2\times10^{n-2}+\cdots+a_n\times10^0 \tag{1.2}$$

(a_1，…，a_n 取 0，1，…，9；n 取正整数)

二进位制的数只有两个数字 0、1，它的进位规则是“逢二进一”，2 是二进位制的进位单位。同十进位制的数一样，二进位制的数可以比较大小，它可以进行加、减、乘、除四则运算。由于计算机的计算与记忆元件只有“开”和“关”两种状态，因此，计算机上通用的是二进位制[1-6]。二进位制的数一般形式为：

$$A = a_1 \times 2^{n-1} + a_2 \times 2^{n-2} + \cdots + a_n \times 2^0 \tag{1.3}$$

（a_1，⋯，a_n 取 0，1；n 取正整数）

将十进制的数化为二进制的数，只要不断地用 2 去除，直到商为 0 为止。得到的逐个余数，就是二进制的数字，把它们依次排出，就得到与十进制数相等的二进制数。例如：

$$\begin{array}{r|l}
2 & 5 \quad \cdots 1(\text{第 1 位余数}) \\ \hline
2 & 2 \quad \cdots 0(\text{第 2 位余数}) \\ \hline
 & 1 \quad \cdots 1(\text{第 3 位余数})
\end{array}$$

$\therefore\ 5_{10} = 101_2$

$$\begin{array}{r|l}
2 & 26 \quad \cdots 0(\text{第 1 位余数}) \\ \hline
2 & 13 \quad \cdots 1(\text{第 2 位余数}) \\ \hline
2 & 6 \quad \cdots 0(\text{第 3 位余数}) \\ \hline
2 & 3 \quad \cdots 1(\text{第 4 位余数}) \\ \hline
 & 1 \quad \cdots 1(\text{第 5 位余数})
\end{array}$$

$\therefore\ 26_{10} = 11010_2$

将二进制的数化为十进制的数，只要将二进制数的每个数字，依次乘以 2 的正整数次幂，然后求和，就可得到与它相等的十进制数。例如：

$101_2 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 4 + 1 = 5$；

$11010_2 = 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 18 + 8 + 2 = 26$。

【例 1-1】　某商店为提高销售效率，将 61 件商品分成 6 箱事先装好，使得顾客购买时，不管买几件商品都不需要打开包装就能满足要求，问每箱应事先放几件商品？

分析：问题是如何在 1-61 之间选 6 个数，使其和为 61，且又能通过求和，表示出 1-61 之间的各个数。将 6 位的二进制数（由于每箱都不能空，所以每位上数字都是 1），化为十进制数，可得到各箱应装的商品数。

解：$111111_2 = 2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 2^0 = 32 + 16 + 8 + 4 + 2 + 1 = 63_{10}$。

注意到 $63>61$，故第 6 箱内不装入 32 件，而装 $32-(63-61)=30$(件)。

答：各箱中应放入的商品数，分别是 1 件、2 件、4 件、8 件、16 件、30 件。

如果需要的商品数小于 30 件，可以从前面 5 个盒子中，挑选若干个盒子就可满足；如果需要的商品数大于或等于 30 件，可先取第 6 个盒子，其余的由前 5 个盒子中，挑选若干个盒子来补足。

随着电子计算机的广泛应用，二进位制的优越性更加突出。因为电子计算机是用电子元件的不同状态来表示的不同的数。对于十进位制，要求元件能准确地变化出十种状态，这在技术上非常难实现；而二进位制只需要两种状态就能实现。通常一个开关只有"开"和"关"两种状态，可以分别表示 0 和 1；那么一个开关的两种状态就可以表示一个二进位制的数，称为"比特(bit)"，记为 b；n 个开关就可以表示 n 个二进位制数。这样运算起来就非常方便。计算机中把 8 个比特称为 1 个"字节(Byte)"，记为 B，即 1 B＝8 b，作为存储的基本单位。在计算机、数码照相机等数字化家用电器中还常见"KB(千字节)"、"MB(兆字节)"、"GB(千兆字节)"等字样，其意义如下：

$$1\ \mathrm{KB}=1024\ \mathrm{B}\approx 10^3\ \mathrm{B}$$

$$1\ \mathrm{MB}=1048576\ \mathrm{B}\approx 10^6\ \mathrm{B}=10^3\ \mathrm{KB}$$

$$1\ \mathrm{GB}=1073741824\ \mathrm{B}\approx 10^9\ \mathrm{B}=10^3\ \mathrm{MB}$$

美国学者尼葛庞余帝(Nicholas Negroponte，1943－)在其名著《Being Digital(数字化生存)》[1-7]中，将物质世界和信息世界作了有趣的比较：原子构成大千物质世界，而"比特"则构成海量信息世界。

1.1.3 准确数与近似数

1) 准确数与近似数

无论是科学研究还是日常生活中，我们遇到的数有两种。一种是准确数，例如"物质有 4 态：固态、液态、气态和等离子态"、"1 个水分子由 2 个氢原子和 1 个氧原子构成"、"掷钱币出现正面的概率是 0.5"、"我们班级有 32 名学生"、"本学期我修 6 门课"、"我班为抗震救灾捐献善款共 34220 元"等，其中的 4、1、2、0.5、32、6、34220 都是准确数。另一种是近似数，例如"地球的平均半径为 6371

千米，极地半径为 6357 千米、赤道半径为 6378 千米”、“中国有 13 亿人口”、“本市明天的最高温度为摄氏 32 度，最低温度为摄氏 24 度”、“国内外为汶川地震捐款总数达 467 亿元人民币”等，其中的 6371 千、6357 千、6378 千、13 亿、32、21、467 亿都是近似数。

有的准确数虽然理论上有意义，但实际应用中只能用近似数去逼近。数学中遇到不少无理数，例如“边长为 1 的正方形对角线长度为$\sqrt{2}$”、“单位圆的面积为 π”等，这里的$\sqrt{2}$和 π 是准确数。但是在生产实践中，人们只能用有理数来近似。例如：

$$\sqrt{2} \approx 1.4142136, \pi \approx \frac{22}{7} \approx 3.143(\text{“疏率”}),$$

$$\text{或}\ \pi \approx \frac{355}{113} \approx 3.141593(\text{“密率”})$$

不同情况下，取近似的程度也不同。例如做化学实验，天平称量要求保留小数点后 4 位数字；台秤称量要求保留小数点后 1 位数字；滴定管读数要求保留小数点后 2 位。又如经济统计，随规模不同，近似程度也不同。例如按中国国家统计局发布的数据，2010 年中国国内生产总值(GDP)为 397983 亿元，约为 3.98×10^{13} 元；同年上海的 GDP 为 17165.98 亿元，约为 1.72×10^{12} 元；上海闵行区的 GDP 为 1364.37 亿元，约为 1.36×10^{11} 亿元。这里，数量级是关键，取 3 位有效数字就可以了。

为了分析和计算的方便，在不少情况下，人们宁愿选取近似数。

【例 1-2】 大海航行的水手要知道他所在位置与海平线的距离。设 R 为地球半径，H 为甲板或桅杆上水手眼睛到海平面的垂直距离，如何估计到海平线的距离 D？(图 1.1)

解：由勾股定理，不难建立关系式：$R^2 + D^2 = (R+H)^2 = R^2 + 2RH + H^2$，化简得

$$D = \sqrt{2RH + H^2} \tag{1.4}$$

因为地球半径 $R = 6371$ 千米，而 H 不会超过 100 米，所以

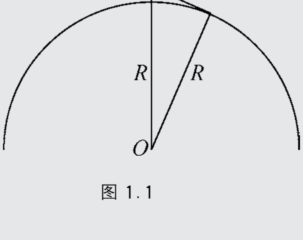


图 1.1

$$D=\sqrt{2RH+H^2}=\sqrt{2RH\left(1+\frac{H}{2R}\right)}\approx\sqrt{2RH}$$

代入 $R=6371$，$D\approx\sqrt{2RH}=112.88\sqrt{H}$（千米）。

由于航海常用海里表示距离，1 海里＝1.852 千米，故 $D\approx\frac{112.88}{1.852}\sqrt{H}=60.95\sqrt{H}$（海里），这里 H 的单位是千米。实际上水手用米来估计 H 更方便，不妨记为 $h=1000H$，则

$$D\approx 60.95\sqrt{\frac{h}{1000}}=1.927\sqrt{h}\approx 2\sqrt{h} \tag{1.5}$$

例如若 $h=6$ 米，则 $H=0.006$ 千米，按 1.4 式算出 $D=8.74$ 千米＝4.72 海里；按 1.5 式算出 $D\approx 4.89$ 海里，虽然有些误差，后者容易记忆得多。

2）绝对误差与相对误差

设将一个准确数记为 A，其近似数记为 A^*，则称

$$E=|A-A^*| \tag{1.6}$$

为“绝对误差”；

$$E_r=\frac{E}{A}\times 100\% \tag{1.7}$$

为“相对误差”。

【例 1-2】中，用 1.5 式代替 1.4 式计算距离的绝对误差和相对误差：

$$E=|4.72-4.89|=0.17\text{ 海里；}E_r=\frac{0.17}{4.72}\times 100\%=3.6\%$$

注意，绝对误差是有量纲的，量纲不同，数字也不同。比如上例的绝对误差以海里为单位，如果改为以千米为单位，$E=0.17\times 1.852=0.315$ 千米。但相对误差是无量纲的，无论绝对误差采用何种量纲，其相对误差是不变的。

人们往往并不知道准确数，但可以知道绝对误差的界限：

$$E=|A-A^*|\leqslant M \tag{1.8}$$

其中 M 就称为“绝对误差界”，简称为“误差界”。各种测量仪器标定的“精度”就

是指测量结果的误差界。

3）近似数的有效数字及其运算

有效数字是由全部准确数字和最后一位（只能是一位）不确定数字组成，它们共同决定了有效数字的位数。有效数字位数的多少反映了测量的准确度，在测定准确度允许的范围内，数据中有效数字的位数越多，表明测定的准确度越高。例如 $\pi=3.14159265\cdots$，前述的“疏率”有 4 位有效数字，“密率”有 7 位有效数字。

刘徽 265-316

中国古代数学家在圆周率的研究中有突出贡献。三国时期的刘徽（265-316）发明“割圆术”，用圆的内接正 n 多边形的周长与直径之比来近似圆周率。例如取 $n=48$，得到 $\pi\approx3.139$；他还用一个分数来近似：

$$\pi\approx\frac{157}{50}\approx3.140$$

后人称之为“徽率”。前面提到的“疏率”和“密率”则是南北朝人祖冲之（429-500）的创造。

祖冲之 429-500

注意：在数字前面的“0”起定位作用，不是有效数字；数字中间的“0”都是有效数字；数字后面的“0”，也是有效数字。如 0.015 有 2 位有效数字，0.0150 有 3 位有效数字，而 0.001005 有 4 位有效数字。

对于准确数，其有效数字可视为无限多位。

有效数字的运算规则如下：

（1）加减法：当几个近似数相加或相减时，它们的和或差保留几位有效数字，应以小数点后位数最少（即绝对误差最大）的数为依据。如：

$$3.25+40.1+573.045\approx3.3+40.1+573.0=616.4$$

（2）乘除法：对几个近似数进行乘除运算时，它们的积或商的有效数字位数，应以其中相对误差最大的（即有效数字位数最少的）那个数为依据，可以多保留一位有效数字。如：

$$10.3\times1.5693\approx10.3\times1.57=16.17$$

经济学中有“木桶效应”，即由若干木条箍成的木桶，最短一根木条决定了木桶盛水的深度。很类似，由若干近似数进行运算时，近似程度最低的那个数决定了整个运算结果的近似程度。

1.2 坐标与曲线(平面解析几何)

1.2.1 代数的发展

“数”和“形”是数学的主要对象。但是直到18世纪之前，人们对“数”和“形”的研究几乎是互相独立的。

研究“形”的学科是几何。由于古希腊学者的杰出贡献，特别是欧几里得关于平面几何公理体系的建立，使得几何的研究水平在很长的历史时期中遥遥领先。可以说，从希腊时代到16世纪，几何统治着数学。另一方面，几何研究的进步非常缓慢，以至于今天的中学教材，仍沿用着欧氏几何的基本框架。

研究“数”的学科主要是代数和数论。由于数论中的不少猜想尚未证明，至今还吸引着许多数学工作者或业余爱好者的兴趣；特别是20世纪后半叶以来，数论在近似计算和编码破密上发挥了重要作用，重新焕发出光芒。尽管如此，数论仍未收入中学、大学的基础课教材中。应当说，以解方程为中心的“代数”却是研究“数”的主干。人们为了求解3次、4次以及5次以上的代数方程，进行了千百年的努力。这方面，中国古代学者作出了卓越贡献。

与以证明定理为中心的希腊古典数学不同，中国古代数学以创造算法特别是各种解方程的算法为主线。从线性方程组到高次多项式方程，乃至不定方程，中国古代数学家创造了一系列先进的算法(中国数学家称之为“术”)，他们用这些算法去求解相应类型的代数方程，从而解决导致这些方程的各种各样的科学和实际问题。法国科学院通讯院士加布里尔(Pierre Gabriel，1933-1986)教授在他撰写的教科书[1-8]中就称解线性方程组的消元法为“张苍法”。张苍(256-152 B.C.)是西汉初年历算学方面的突出代表，他在历法、算学方面取得了很大的成就：① 他提出和制订了一套比较完整的关于度、量、衡方面的理论，他把算学研究成果直接用于国计民生。② 在采用历法方面，张苍提倡采用《颛顼历》。③ 增订、删补《九章算术》[1-9]。元代朱世杰(生卒不详)所写的两部名著《算学

启蒙》三卷(1299)和《四元宝鉴》三卷(1303)是我国数学发展的重要里程碑。前者创立了代数加法和乘法的正负法则;后者把天元术推广为“四元术”(四元高次联立方程解决),而欧洲到1775年才提出同样的解法。他的工作使中国的代数学在当时达到了顶峰[1-10]。

古代埃及人、印度人和阿拉伯人对代数的发展也作出了卓越的贡献。讲到代数,不能不提到古希腊学者丢番图(Diophantus, 246-330)的成就。他首创用符号代表数,并提出“平方”、“立方”等概念及符号。在他的著作中研究了大量的代数方程,被后人称为“代数之父”。相传在他的墓碑上刻着如下的铭文:“坟中安葬着丢番图,多么令人惊讶,他忠实地记录了所经历的道路。上帝给予的童年占六分之一,又过了十二分之一,两颊长胡,再过七分之一,点燃结婚的蜡烛,五年之后天赐贵子,可怜迟到的宁馨儿,享年仅及其父之半,便进入坟墓,悲伤只有用数论研究去弥补,又过四年,他也走完了人生的旅途。”[0-1]这就是一道一元一次代数方程题,设丢番图活了 x 岁,则由墓志铭,可建立方程

$$\frac{x}{6}+\frac{x}{12}+\frac{x}{7}+5+\frac{x}{2}+4=x$$

答案是 $x=84$, 他活了84岁!

1.2.2 解析几何产生的背景

如前所述,直到16世纪,数学的发展比较缓慢,这是和社会和生产的发展相对缓慢是一致的。中国古代数学有辉煌的成就。除前面提到的墨子、张苍、刘徽、祖冲之、朱世杰外,还有不少在当时领先于世界的数学家,例如:

秦九韶纪念堂

杨辉三角形

秦九韶(1202-1261)提出解不定方程的“大衍求一术”(现被数学界命名为“中国剩余定理”)和高次方程的数值解法;

沈括 1030-1094

沈括(1030-1094)在其名著《梦溪笔谈》中用“隙积术”求级数之和;

杨辉(13 世纪)在《详解九章数学》中发展了“隙积术”,并首次记载了二项式系数,被称为“杨辉三角形”;等等。[1-10]

但是,以小农经济为基础的中国封建社会发展非常缓慢,占统治地位的儒家思想并不重视数学,认为是“六艺之末”,于仕途无补。长达近两千年的科举制度,并未将数学列为考试内容。这就使得中国的数学和其他科学一样,逐渐由先进到落后。

中世纪的欧洲受到罗马教廷的控制,由于生产水平和思想禁锢,数学发展也较缓慢。13 世纪中叶开始,出现了资本主义生产方式的萌芽,家庭手工业转化为工场手工业,航海探险不仅开拓殖民地,掠夺海外原料和廉价劳动力,也促进了造船业的发展。天文、造船迫切需要科学技术的支持;而人们对地球的认识更是产生了对宗教教义的怀疑。伟大的变革终于在欧洲诞生。

欧洲文艺复兴带来了思想大解放。[1-11]

意大利艺坛三杰达·芬奇(Da Vinci, 1452-1519)、米开朗琪罗(Michelangelo, 1475-1564)和拉菲尔(Raphael, 1483-1520)在绘画雕塑中一扫中世纪宗教作品的沉沉死气,无论是绘画《蒙娜丽莎》、《圣母》,还是雕塑《大卫》,都注入了人性的光辉。

蒙娜丽莎

圣母

大卫

波兰天文学家哥白尼(Copernicus，1473-1543)提出“日心说”，向教会提出挑战[1-12]；

意大利物理学家伽利略(Galileo，1564-1642)发明望远镜观测月球和行星，通过实验否定了亚里士多德的某些理论[1-13]；

英国医生哈维(William Harvey，1578-1657)提出人体血液循环论，否定了关于血液是“圣父、圣子、圣灵三位一体灵气”的教义；等等。

哥白尼 1473-1543

伽利略 1564-1642

哈维 1578-1657

甚至宗教本身也经历了重大改革，其代表人物是德国教士马丁·路德(Martin Luther，1483-1546)创立的路德教派。他先后将圣经的旧约、新约从希腊文、希伯来文译成了德文。德国大诗人海涅(Heinrich Heine，1797-1856)认为，马丁·路德对圣经的翻译创造了德语。

1520年，路德发表了被称为宗教改革三大论著的《致德意志贵族公开书》、《教会被囚于巴比伦》和《基督徒的自由》[1-14]，更是唤起民众摧毁罗马教廷神权统治的有力思想武器。思想解放在数学界必然带来很大冲击。

马丁·路德 1483-1546

海涅 1797-1856

笛卡儿(Rene Descartes,1596－1650)是法国17世纪最伟大的学者之一[1-15]。有趣的是,他被后世学者评为:“第一个杰出的近代哲学家,是近代生物学的奠基人,是第一流的物理学家,但只偶然地是个数学家”[0-1]。也许正因为他是哲学家,能跳出具体学科细节的束缚;正因为他在自然科学许多领域多有贡献,所以特别注重把科学成果付之应用。由于天文航海及生物研究之需,笛卡儿曾参与望远镜和显微镜的设计,这就要研究几何问题。他痛感当时占数学统治地位的几何方法有两大弊端。一是过于依赖对图形的绘制和度量,很难准确;二是过于依赖精巧的逻辑思维,往往一个几何问题一个方法,效率太低。能否用代数的方法计算几何量?能否用代数方程描述几何图形及证明几何性质?这就是催生解析几何的背景,笛卡儿是当之无愧的奠基人。

笛卡儿 1596-1650

1.2.3 直角坐标系的引入

1) 点与坐标的“一一对应”关系

现在,中学数学都会介绍直角坐标系又称为“笛卡儿坐标系”,就是对首创者笛卡儿的纪念。用两条垂直的有向直线(“横轴”和“纵轴”,通常也称 x 轴和 y 轴)相交,交点称为“原点”(通常用 O 表示,即 Origin 的第一个字母)。平面上任一点可以由一对有序实数(称为“横坐标”和“纵坐标”)表示;反之,任一对有序实数也唯一确定平面上的一个点(图1.2(a)):

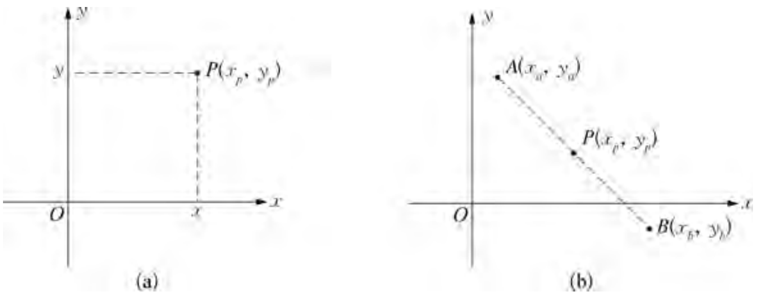


图1.2

$$点\ P \Leftrightarrow 坐标(x, y) \tag{1.9}$$

这样,求两点的距离就不必用尺度量,而可以精确计算了(图 1.2(b))。

$$A, B\text{ 两点的距离} = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} \tag{1.10}$$

这里 (x_a, y_a) 和 (x_b, y_b) 分别是点 A 和点 B 的坐标。

求 A,B 两点的中点 P 也很方便:

$$x_p = \frac{x_a + x_b}{2},\ y_p = \frac{y_a + y_b}{2} \tag{1.11}$$

这里(x_p, y_p)是中点 P 的坐标(图 1.2(b))。

中学已学过直线与圆的方程,下面简单复习和说明。

2) 直线与圆的方程

任一直线都"一一对应"于一个二元一次代数方程

$$Ax + By + C = 0 \quad (A, B\text{ 不全为 }0) \tag{1.12}$$

这里(x, y)称为"动坐标"。当 $B\neq 0$,式 1.12 可以改写为

$$y = kx + b \tag{1.13}$$

其中的 k 和 b 有明确的几何意义。记 α 为直线的"倾角",则

$$k = \tan\alpha \tag{1.14}$$

称为直线的"斜率",而 b 是直线与纵轴交点的坐标,称为"截距"(图 1.3(a)),因而式 1.13 称为"斜截式"。

当 $B=0$,式 1.12 可以改写为 $x = C$,是一条垂直的直线(图 1.3(b))。

两点可以连成直线。过任意两点 $P_1(x_1, y_1)$ 和 $P_2(x_2, y_2)$ 的直线方程为

$$\frac{x - x_2}{x_1 - x_2} = \frac{y - y_2}{y_1 - y_2} \tag{1.15}$$

式 1.15 称为"两点式"(图 1.3(c))。显然,两点可以确定斜率

$$k = \frac{y_2 - y_1}{x_2 - x_1} \tag{1.16}$$

如果已知直线的斜率 k,并过一点 $P(x_0, y_0)$,则此直线方程为

$$y - y_0 = k(x - x_0) \tag{1.17}$$

式 1.17 称为"点斜式"(图 1.3(d))。

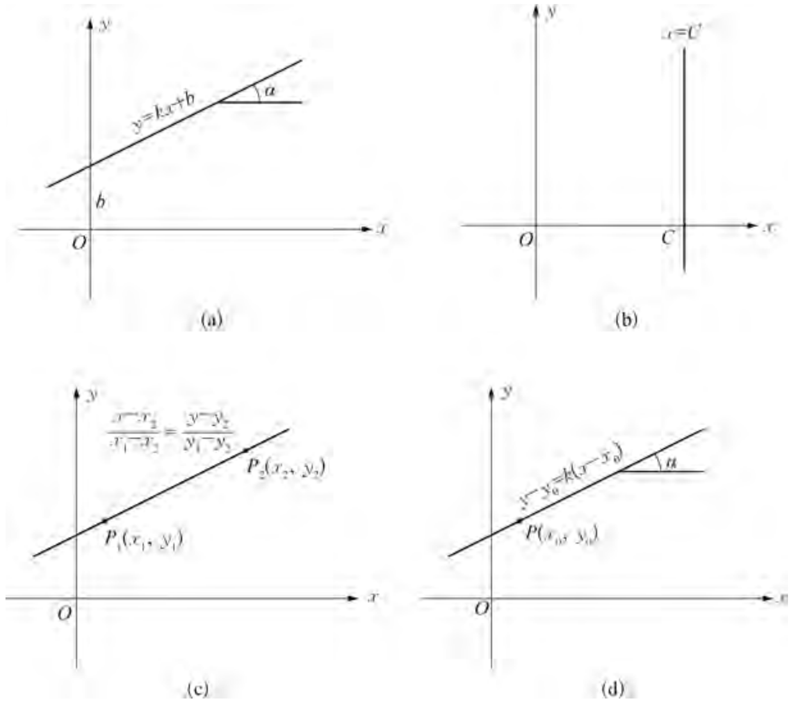


图 1.3

圆的方程不难从圆的定义而得。设圆心坐标为 (x_0, y_0)，半径长为 R，圆周上任一点的动坐标为 (x, y)，则由距离公式 1.10 得到圆的公式为

$$(x-x_0)^2+(y-y_0)^2=R^2 \tag{1.18}$$

3）动坐标的革命性意义

形式上 1.10 式和 1.18 式不过是二元一次方程和二元二次方程，但是它们和以前的代数方程有本质的差别。已往人们求解代数方程，求的是有限而确定的根。比如对一元二次方程求两个根，一元三次方程求三个根，一元 n 次方程求 n 个根。对单独的二元一次方程或二元二次方程称为“不定方程”，一般情况下，被认为没有意义。对一个不定方程，只有求其整数解时，才引起人们的兴趣。例如费马大定理就是求三元 n 次方程

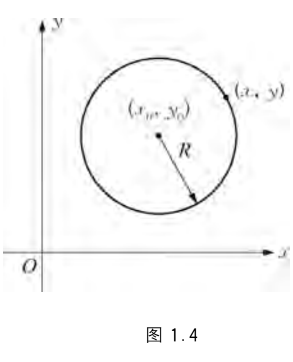


图 1.4

$$x^n+y^n=z^n \tag{1.19}$$

的整数解。当 $n=2$ 时，有无数解，例如(3，4，5)、(5，12，13)、(99，4900，4901)以及它们的倍数都是解；但当 $n\geqslant 3$，无整数解。

通过动坐标的引入，笛卡儿赋予不定方程以全新的意义。由于直线或圆上的动坐标(x, y)有无限组，这就把变量引入了数学。对此恩格斯有很高的评价："数学中的转折点是笛卡儿的变数。有了变数，辩证法进入了数学，有了变数，微分和积分也就立刻成为必要的了。"[1-16] 辩证法就体现在，原来几乎不相关的"形"与"数"，或者说"几何"与"代数"通过动坐标统一起来。既可以把几何问题化为代数方程，把测量化为计算；又可以考察代数方程的几何意义，通过直观的图形更好理解和解决问题。因此，坐标几何又称为"解析几何"。

【例 1-3】 证明三角形三条中线交于一点。

这是平面几何的一道经典难题，但是用解析几何的方法，就非常容易解决。

在笛卡儿坐标系上考虑任意三角形OAB。为方便计，设O为原点，OA在轴上，且$|OA|=1$；设点B坐标为(p, q)(图 1.5)。

由 1.11 式，三边中点的坐标分别为 $L\left(\frac{1}{2}, 0\right)$，$M\left(\frac{1+p}{2}, \frac{q}{2}\right)$，$N\left(\frac{p}{2}, \frac{q}{2}\right)$。再由两点式 1.14，得到三条中线的方程为

$$OM: y=\frac{q}{1+p}x \tag{1.20}$$

$$AN: y=\frac{q}{p-2}x-\frac{q}{p-2} \tag{1.21}$$

$$BL: y=\frac{2q}{2p-1}x-\frac{q}{2p-1} \tag{1.22}$$

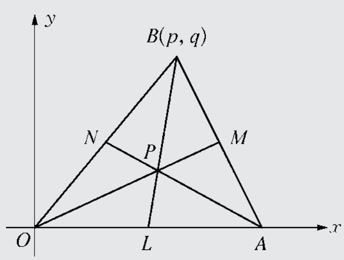


图 1.5

解联立方程 1.20 和 1.21，求出的交点 P 的坐标为 $\left(\frac{1+p}{3}, \frac{q}{3}\right)$，它也满足 1.22 式，这就"计算"出三角形三条中线交于一点。

反过来，利用坐标，把一组成对的数字表示为若干点，并连成曲线或折线，这也是科学研究和日常生活中常用的方法。

【例 1-4】 据中国、日本、美国官方公布的统计数字，2005-2010 年三国国民生产总值(GDP)如表 1.3 所示，单位为亿美元[1-17]。

表 1.3

国家 \ 年份	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
中国	22343	26801	34022	43026	49964	59847
日本	45522	43626	44370	48870	50330	54589
美国	124220	133989	140618	143691	141191	146578

建立直角坐标系，横轴为年份 t，纵轴为 GDP，每年的一对数据(t, GDP)就对应平面上一个点。图 1.6 将每国的 6 个点连成折线，可以形象地看到三国的经济总体实力及其增长趋势。

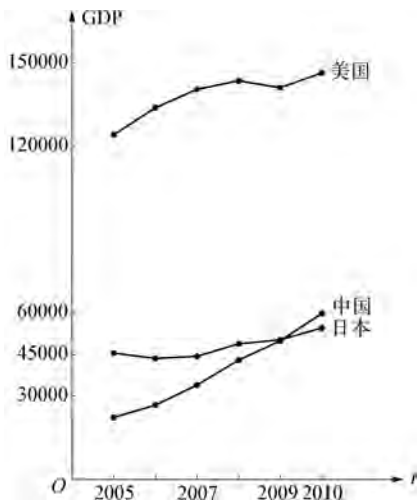


图 1.6

在微积分的学习中，常常考察函数、导数等数学概念的几何意义，这一切都建立在解析几何的基础之上。

1.3 空间与曲线曲面

1.3.1 二维欧氏空间

前面介绍解析几何中的坐标平面数学上称为“二维欧式空间”。“维

(dimension)”表示变量活动的自由度。平面上任一点可以有 x 和 y 两个自由活动方向，所以平面是二维空间。而直线上任一点只有沿此直线一个自由活动方向，所以直线是一维欧氏空间。所谓“欧氏空间”是指此空间是“平直”的。也有非欧氏空间，一条曲线(例如圆)上任一点也只有沿此曲线一个自由活动方向，所以也是一维空间，但它不是欧氏空间，因为“弯曲”而不“平直”，比如曲线上两点的中点坐标并不符合 1.11 式。

1.3.2 圆锥曲线与二元二次方程

前面我们提到，平面上任何直线都是一个二元一次方程，反之，任何一个二元一次方程 1.12 都在坐标平面上表示一条直线。我们又提到，任何一个圆，都可表示为特殊形式的二元二次方程。那么，任何一个二元二次方程

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \tag{1.23}$$

都能表示为圆吗？它还能表示一些什么曲线？笛卡儿发现，式 1.23 可以包含所有的圆锥曲线！

阿波罗尼奥斯 262-192 B.C.

人们对圆锥曲线的认识很早。欧几里得已经提到并有一些研究成果。不过对圆锥曲线有系统研究的当推另一位古希腊数学家阿波罗尼奥斯(Apollonius, 262-192 B.C.)[0-1]。他的名著《Conic Section(圆锥曲线论)》总结了前人的知识，并予以系统化。正是他首先为圆锥曲线取了 3 个名称：椭圆、抛物线、双曲线。

设一正圆锥的轴线竖直，底面水平，锥面与底面的夹角为 α。现用一平面与锥面相截，设此平面与圆锥底面的夹角为 β，则平面与圆锥面的交线称为“圆锥曲线”(图 1.7)：

情况 1：当 $0 \leqslant \beta = \beta_1 < \alpha$，交线为“椭圆”；

情况 2：当 $\beta = \beta_2 = \alpha$，交线为“抛物线”；

情况 3：当 $\alpha < \beta = \beta_3 \leqslant \dfrac{\pi}{2}$，交线为“双曲线”，有上下两支。

圆锥曲线还有一种定义：

(1) 椭圆：到两个定点(称为“焦点”)的距离之和等于定长(定长必须大于两个定点间的距离)的动点的轨迹；

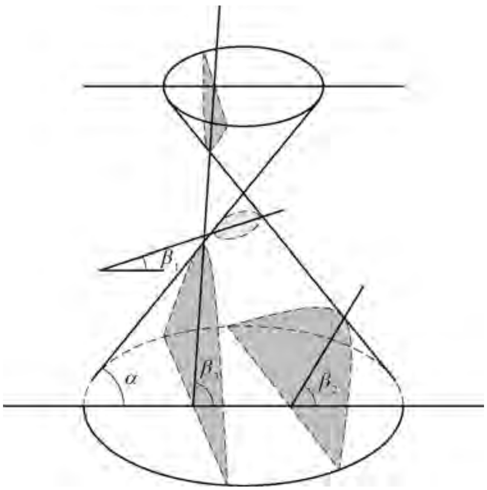


图 1.7

（2）双曲线：到两个定点（称为“焦点”）的距离的差的绝对值为定值（定值必须小于两个定点的距离）的动点轨迹；

（3）抛物线：到一个定点（称为“焦点”）和一条定直线（称为“准线”）的距离相等的动点轨迹。

其统一定义是：到定点的距离与到定直线的距离的比 e 是常数的点的轨迹叫做圆锥曲线。当 $0<e<1$ 时为椭圆；当 $e=1$ 时为抛物线；当 $e>1$ 时为双曲线。

按上述定义，可以推出三种曲线的标准方程[1-18]。仍设(x, y)为曲线上的动坐标，则在笛卡尔直角坐标系中，三种曲线的方程及对应的图形如图 1.8 所示。

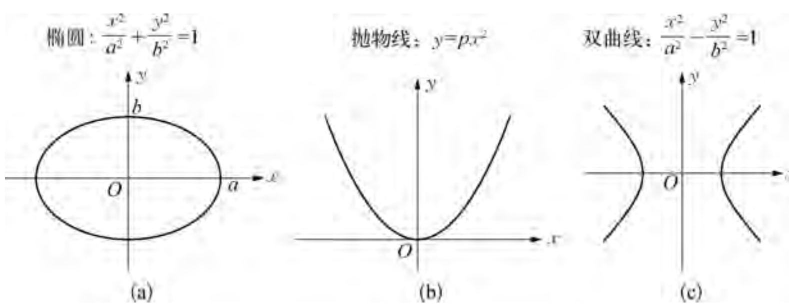


图 1.8

显然，这三种圆锥曲线都可以表达为特殊的二元二次方程，都是式 1.23 的特例。

圆锥曲线有重要的光学性质：

（1）椭圆：点光源在一个焦点上，光线经椭圆反射，必通过另一个焦点。

（2）双曲线：点光源在一个焦点上，反射光线与另一焦点到反射点的连线在同一条直线上。

（3）抛物线：点光源在焦点上，反射光线相互平行且垂直于准线。这个性质有许多应用。从手电筒到探照灯，都把光源放在焦点上，通过旋转抛物面，反射

出平行光束；而太阳灶则相反，通过旋转抛物面把平行的太阳光发射到一个点，产生很高的热量。

1.3.3 三维欧氏空间与曲面

1）三维欧氏空间

我们生活的空间有东西、南北、上下三个方向，这正是三维欧氏空间的原型。用三条垂直的有向直线（“横轴”、“纵轴”和“竖轴”，通常也称 x 轴、y 轴和 z 轴）相交，交点也称为“原点”。空间中任一点可以由三个有序实数（称为“横坐标”、“纵坐标”和“竖坐标”）表示；反之，任三个有序实数也唯一确定空间中的一个点（图 1.9）。

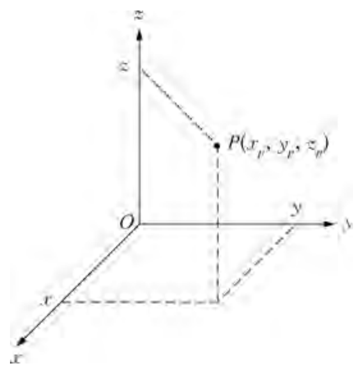


图 1.9

$$点 \Leftrightarrow 坐标(x,\ y,\ z) \tag{1.24}$$

许多二维空间的公式可以推广到三维。例如两点的距离公式为

$$A,\ B\ 两点的距离 = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2 + (z_b - z_a)^2} \tag{1.25}$$

这里 $(x_a,\ y_a,\ z_a)$ 和 $(x_b,\ y_b,\ z_b)$ 分别是点 A 和点 B 的坐标。求 A,B 两点的中点 P 也很方便：

$$x_p = \frac{x_a + x_b}{2},\ y_p = \frac{y_a + y_b}{2},\ z_p = \frac{z_a + z_b}{2} \tag{1.26}$$

这里 $(x_p,\ y_p,\ z_p)$ 是中点 P 的坐标。

2）平面

类似于二维空间中的直线，三维空间中任一平面都“一一对应”于一个三元一次代数方程

$$Ax + By + Cz + D = 0 \quad (A,\ B,\ C\ 不全为\ 0) \tag{1.27}$$

这里 $(x,\ y,\ z)$ 称为“动坐标”。

3）球面

类似于二维空间中的圆，三维空间中半径为 R，球心为 $(x_0,\ y_0,\ z_0)$ 的球面

上动坐标(x, y, z)满足三元二次方程

$$(x-x_0)^2+(y-y_0)^2+(z-z_0)^2=R^2 \tag{1.28}$$

1.4 空间概念的发展与抽象

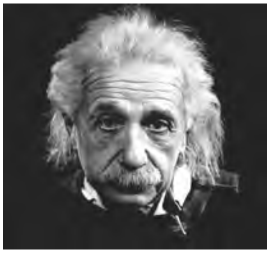
虽然人类生活在三维空间,但是通过思维、类比和抽象,空间概念有了很大的拓广,突破了一、二、三维的限制。比如用 5 种颜色调成的色彩,可以看作 5 维空间的点;又如为了确定“神舟 7 号”在空间中的位置和姿态,把飞船看作刚体,需要 6 个参数(力学中称为“自由度”),即 3 个平动参数,3 个转动参数,我们可以看作一个 6 维空间,其中每个点有 6 个坐标,每个坐标对应一个参数[1-19]。又如绪论 0.2.3 节提到的《红楼梦》作者鉴定,研究者把《红楼梦》一百二十回本作为一个整体,以 47 个虚字为识别特征,对它们在书中各回的出现频率进行统计分析。那么每一回可以看作 47 维空间的一个点,如果前 80 回的 80 个点密集地聚集在一起,而后 40 回的 40 个点与前 80 个点有较大的分离,就可判定两者的作者不像是同一人[0-29]。

有两类超乎直观想象,又有重大应用的空间,即四维时空和分数维空间,它们是 20 世纪科学家的重大贡献。

1.4.1 爱因斯坦-闵可夫斯基四维时空

很久以前,人们就从日常经验中建立了朴素的“时间”和“空间”的概念。他们认为,“空间”就像一个大空箱子,里面装着宇宙万物。即使把所有物质拿走,这个大空箱子依然存在。后来牛顿把这种独立于物质之外的空间叫做“绝对空间”。“时间”则被想象成不停流动的长河,就是平常说的“光阴如流水”。它也是和物质没有任何关系的独立存在,而且和空间无关。牛顿把它称作“绝对时间”。这种质朴的时间空间概念如此简明直观,对它的数学描述也就很简单:只要在“绝对空间”中建立一个三维的笛卡尔坐标系,其中任何一组坐标(x, y, z)就决定了物体在空间的位置。另外再单独建立一个一维的时间轴 t,其上任何一点就表示了一个确定的时刻。三维空间坐标系和一维时间坐标系是分别建立的,彼此独立,时间和空间互不相关。

20世纪初，科学巨匠爱因斯坦（Albert Einstein，1879-1955）根据实验证实的“光速不变性”（即不同惯性参考系中测得的光速居然都相同，都是 c，而且和光源是否运动无关），认识到物理学的“相对性原理”的普遍性，从而建立了“相对论”[1-20]。相对论其实就是关于时间和空间的新理论。按照相对论时空观，不存在独立于运动物质之外的、绝对的时间和空间。用物理学的话说：没有指定参考系，空间和时间就失去明确的定义。而参考系并不神秘，任何运动的物体如地球、火车、轮船等等，都可以作为参考系。指定一个运动物作参考系后，尺长和时钟快慢才有明确定义，对时间和空间的定量度量才有可能，时间空间才有了明确意义。不同参考系中的尺长和时钟快慢彼此不同，甚至连两个不同地点的时钟是否“同步”运行都各有不同说法。由此不难理解，为什么不同参考系中测得的光速竟是相同的。所有这些结论都与我们熟悉的绝对时空观念大相径庭。这是因为在日常经验中，我们接触的都是低速环境。这时，人们的感觉是尺长与时钟走的快慢都是绝对的，和选用哪个参考系无关。由此产生错觉：时间和空间都是绝对的。



爱因斯坦　1879-1955

相对论进一步指出，时间和空间彼此互相依存，关系密切。实际上，这更符合我们的日常经验。例如经常谈到某月某日在某地发生了一个事件，这正好说明，只有同时标明时间和空间地点才能明确规定一个事件，缺一不可。为了更准确地描述物理事件和物理过程，爱因斯坦的一位好友闵可夫斯基（Hermann Minkowski，1864-1909）推广了原来的三维笛卡儿坐标系，在长宽高基础上又增加了一条时间轴 t。这就是闵可夫斯基四维时空[1-21]。其中任何一点都对应一组坐标 (x, y, z, t)，它明确标定了空间地点和时间，故可用它代表一个事件。闵可夫斯基四维时空的引进是对相对论的一大贡献，由此还得到一系列新的有用的物理结果。

闵可夫斯基　1864-1909

下面介绍闵可夫斯基如何把三维欧几里得空间的距离公式推广到四维时空。

由1.25式，三维欧几里得空间的距离平方为

$$\Delta s^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 \tag{1.29}$$

它是坐标变换不变量，因为从老坐标(x, y, z)变换到新坐标(x', y', z')时，两点间距保持不变。即

$$\begin{aligned}\Delta s^2 &= (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 \\ &= (x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2\end{aligned} \tag{1.30}$$

闵可夫斯基在他的四维时空中引进了类似于三维空间中"距离"的量，叫做"时空间隔"，或简称"间隔"。记作

$$\Delta s^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2 \tag{1.31}$$

它也是坐标变换不变量(用物理的话说，是参考系变换不变量)。当从一个惯性参考系S变到另一个新惯性参考系S'时，四维坐标系相应改变。四维时空坐标当然作相应变换：即$(x, y, z, t) \to (x', y', z', t')$。在参考系变换时，光速是不变的，可以用一个数学式表示。设在老参考系S中，光在t_1时刻从(x_1, y_1, z_1)点出发(称为"事件1")，然后在t_2时刻到达(x_2, y_2, z_2)点(称为"事件2")。由于光速是c，故有

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2 = 0 \tag{1.32}$$

当换到新参考系S'观察光的发射和到达这两个事件1和2时，由于光速不变，同样有

$$(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2 - c^2(t'_2 - t'_1)^2 = 0 \tag{1.33}$$

所以有

$$\begin{aligned}&(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2 - c^2(t'_2 - t'_1)^2 \\ &= (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2 = 0\end{aligned} \tag{1.34}$$

1.34式就是光速不变在四维时空中的数学表达式。著名的相对论坐标变换(洛伦兹变换)以及通过论证得到的相对论新的时空观都是由此导出的。

上述的事件1和2是通过光讯号传播联系在一起的。一般而言，世上的任何两事件也可以用其他非光过程联系(例如火车的出发和到达)，甚至两件事彼此毫无联系。此时1.32式或1.33式中的平方和就不一定等于零。也可能大于零或小于零。但是可以证明，两个坐标系(参考系)中写出的这个平方和仍然相

等，是坐标变换不变量。即

$$
\begin{aligned}
&(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2 - c^2(t'_2 - t'_1)^2 \\
&= (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2 \qquad (1.35)
\end{aligned}
$$

闵可夫斯基把这个坐标变换不变式（来自光速不变）叫作“时空间隔”。为了和三维笛卡儿坐标系中的两点距离式1.29作类比，他把“时空间隔”也记作：

$$\Delta s^2 \equiv (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2 \qquad (1.36)$$

闵可夫斯基希望1.36式更像三维空间中的距离平方的表示式，引入如下坐标：

$$x_1 = x,\ x_2 = y,\ x_3 = z;\ x_4 = ict \quad (\text{其中 } i = \sqrt{-1} \text{ 是虚数单位，} i^2 = -1) \qquad (1.37)$$

这样，时间轴 t 就成为虚轴，并且带有因子光速 c，1.36式就真正表为4个平方之和：

$$\Delta s^2 \equiv (x_1 - x'_1)^2 + (x_2 - x'_2)^2 + (x_3 - x'_3)^2 + (x_4 - x'_4)^2 \qquad (1.38)$$

其实，1.36式给出的“时空间隔”比笛卡儿坐标系中的距离意义要复杂。前三个平方之和联系着空间距离，而第4个平方项联系着时间；既包含了两事件的空间距离，也包含了时间间隔。

1.36式常常表示成更加简单的形式如下：

$$\Delta s^2 = x^2 + y^2 + z^2 - c^2t^2 \qquad (1.39)$$

这相当于把事件1看成在 $t=0$ 时刻发生在坐标原点$(0, 0, 0)$，事件2则看成 t 时刻发生在点(x, y, z)。

当1.36式大于零时，表示前三项平方和（空间项）为主，即两点空间距离很远，在 t_2-t_1 时刻，光不能从一点传播到另一点。所以称为“类空间隔”。这么称呼的另一个理由是：当1.36式大于零，总可以找到一个参考系（坐标系），在其中观察时，事件1和2是“同时发生”，即此参考系中，有 $t=0$，那么此系中，就的确表示了两事件之间的空间距离；当1.36式小于零，时间项（负项）为主，在t_2-t_1 时刻光传播的距离大于两点的空间距离，所以称为“类时间隔”，或者说，此时总可以找到一参考系，其中考察的两事件1和2是在空间同一地点发生的，即此系中有 $x=0$，$y=0$，$z=0$。故此时 Δs 就的确表示了两事件之间的时间间隔了。最后，当

1.36式平方和恰为零,则称作“类光间隔”。从 t_1 到 t_2,光恰好从一点传播到另一点,它无疑表示:这两物理事件是由光传播(或可以用光传播)联系在一起的。总之,根据 $\Delta s^2=0$, >0 或 <0,可以把四维时空划分出“类光区域”、“类空区域”以及“类时区域”。在类时区域中的任何两点代表的两事件(其中事件1已固定在原点),总是有因果联系,或者可以用有因果联系的两事件联系在一起;而在类空区域中任何一点,它与处在原点的事件1无论如何不会有因果联系。这样,世上发生在任何一处的事件,它和处在原点的事件1之间是否有因果联系就一目了然了。

闵可夫斯基的四维时空是表达爱因斯坦“狭义相对论”的有力数学工具,也为日后爱因斯坦建立“广义相对论”打下良好基础。“广义相对论”是相对论的引力理论,其中对引力场的描写很有趣,采用了黎曼几何中的数学技巧。爱因斯坦认为引力中心(例如太阳)周围的时空的欧几里得平直性遭到破坏,发生了“时空弯曲”。平常说的太阳对地球的引力效应,可以看成是一种几何效应,地球会绕日运行,仅仅是由于太阳周边时空变得弯曲所致。从而在“广义相对论”中,爱因斯坦用“弯曲”的四维时空(黎曼几何)代替了闵可夫斯基四维“平直”的时空(欧几里得几何)。用浅显的话说:四维弯曲时空(即引力场)的特点是:其中每一个时空点处都有自己的尺和钟,各点的尺长和钟的快慢都不同。所以时间、空间不均匀了,“弯曲”了。

1.4.2 分数维空间

普通几何学研究的对象,一般都具有整数的维数。比如,零维的点、一维的线、二维的面、三维的立体,乃至前面介绍的四维的时空。在20世纪70年代末80年代初,产生了新兴的分形几何学(fractal geometry),空间具有不一定是整数的维,而存在一个分数维数。这是几何学的新突破,引起了数学家和自然科学者的极大关注。

曼德布罗特 1924-

客观事物有它自己的特征长度,要用恰当的尺度去测量。例如用尺来测量京沪铁路的长度,似乎太短;而用尺来测量毛发的直径,又嫌太长,从而产生了特征长度。还有的事物很难设定不变的特征尺度,必须同时考虑从小到大的许许多多尺度,如自然界中的湍流现象、静室中缭绕的轻烟、云彩的边界、山川的轮廓等。海岸线的长度也很难用特定的尺度测量。诞生于波兰

的美籍法国数学家曼德布罗特(B. B. Mandelbrot，1924-　)于1967年在《科学》杂志上发表了题为《英国的海岸线有多长?》的著名论文[1-22]。海岸线作为曲线，其特征是极不规则、极不光滑，呈现极其蜿蜒复杂的变化。用不同尺度测量，长度会有很大变化，如果用公里作测量单位，从几米到几十米的一些曲折会被忽略；改用米来做单位，测得的总长度会增加，但是一些厘米量级以下的就不能反映出来(图1.10)。

图 1.10

另外，在没有建筑物或其他东西作为参照物时，在空中拍摄的100公里长的海岸线与放大了的10公里长海岸线的两张照片，看上去可能很相似。事实上，具有自相似性的形态广泛存在于自然界中，如：连绵的山川、飘浮的云朵、岩石的断裂口、雪花、树冠、树叶等，曼德布罗特把这些部分与整体以某种方式相似的形体称为“分形(fractal)”。1975年，他创立了分形几何学(fractal geometry)。在此基础上，形成了研究分形性质及其应用的科学，称为分形理论(fractal theory)。

分维，又称分数维，作为分形的定量表征和基本参数，是分形理论的又一重要原则。在数学上，把欧氏空间的几何对象(点、线、面)连续地拉伸、压缩、扭曲，维数也不变，这就是拓扑维数。然而，这种传统的维数观受到了挑战。曼德布罗特曾描述过一个绳球的维数：从很远的距离观察这个绳球，可看作一点(零维)；从较近的距离观察，它充满了一个球形空间(三维)；再近一些，就看到了绳子(一维)；再向微观深入，绳子又变成了三维的柱，三维的柱又可分解成一维的纤维。那么，介于这些观察点之间的中间状态又如何呢？显然，并没有绳球从三维对象变成一维对象的确切界限。

数学家豪斯道夫(Felix Hausdoff，1868-1942)在1919年提出了连续空间的概念，也就是空间维数是可以连续变化的，它可以是整数也可以是分数，称为豪斯道夫维数，记作 D。[1-23]

豪斯道夫 1868-1942

设 L 是某客体沿其每个独立方向皆扩大的倍数，K 为得到的新客体是原客体的倍数，则

$$K = L^D \tag{1.40}$$

比如，一个长方形，长宽都扩大到原来的 3 倍，即 $L=3$，其面积是原来的 9 倍，即 $K=9$，故 $D=2$；而一个立方体，若长宽高都扩大到原来的 2 倍，其体积是原来的 8 倍，即 $K=8$，故 $D=3$。这与我们熟知的长方形是二维，立方体是三维完全一致。对 1.40 式两边以相同底取对数，就得到

$$D = \frac{\log K}{\log L} \tag{1.41}$$

显然，在一般情况下 D 是一个分数。下面我们介绍一条分数维曲线——柯克曲线(Koch curve)，然后应用 1.41 式求出其维数。柯克(Helge von Koch，1870-1924)是瑞典数学家，早在 1904 年他构造出一组类似雪花的曲线：首先画一个线段，然后把它平分成三段，去掉中间那一段并用两条等长的线段代替。这样，原来的一条线段就变成了四条小的线段。用相同的方法把每一条小的线段的中间三分之一替换为等边三角形的两边，得到了 16 条更小的线段。然后继续对 16 条线段进行相同的操作，并无限地迭代下去。图 1.11 是这个图形前五次迭代的过程，可以看到这样的分辨率下已经不能显示出第五次迭代后图形的所有细节了。

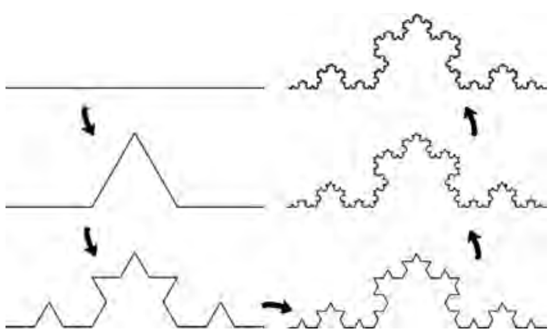


图 1.11

注意：整个线条的长度每一次都变成了原来的 4/3。如果最初的线段长为 1，那么第一次操作后总长度变成了 4/3，第二次操作后总长增加到 16/9，第 n 次操作后长度为 $(4/3)^n$。毫无疑问，操作无限进行下去，这条曲线将达到无限长。

当把三条这样的曲线头尾相接组成一个封闭图形时(图 1.12),有趣的事情发生了。这个雪花一样的图形有着无限长的边界,但是它的面积却是有限的。柯克曲线是最早提出的分形图形之一。正如前面提到的分形特点:如果把它分成若干部分,每一个部分都和原来的形状一样,只是大小不同。这样的图形叫做“自相似”。一条柯克曲线中包含有无数大小不同的柯克曲线。你可以对这条曲线的尖端部分不断放大,但你所看到的始终和最开始一样。它的复杂性不随尺度减小而消失。另外值得一提的是,这条曲线是一条连续的,但处处不光滑的曲线。曲线上的任何一个点都是尖点。

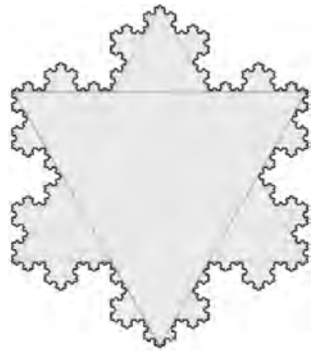


图 1.12

如果初始线段长度为 1,若测量长度放大到 3,4 条边长为 1 的折线的长度为 4,即放大到 4 倍。每迭代一次,测量长度放大到前次的 3 倍,新的折线段长度都是前次的 4 倍。按 1.41 式,柯克曲线的维数为

$$D=\frac{\log 4}{\log 3}=1.26186\cdots \tag{1.42}$$

当然,自然界的海岸线并不具备严格数学意义下的自相似,但可近似计算。根据曼德布罗特的计算,英国海岸线的维数为 1.25,南非的海岸线维数是 1.02。有了分维,海岸线的长度就确定了。

分形几何与分数维问世 30 多年来,不仅在数学物理的研究中发挥了作用,而且应用领域涉及生物、化学、医药、金融、贸易等许多方面[1-24]。

图 1.13 列出几个典型分形图像:

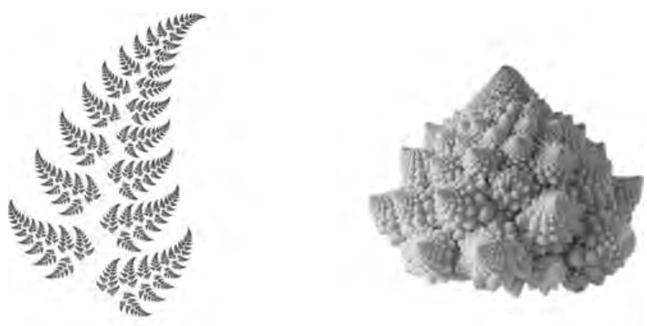


图 1.13

1.5 思考与练习

1. 数的表示

(1) 用科学计数法表示下列各数：

(a) 280003.56 (b) 0.00808

(2) 找出几个很大和很小的物理量，用科学计数法表示。

(3) 将下列10进位数写为2进位数：

(a) 507 (b) 36.08

(4) 将下列2进位数写为10进位数：

(a) 1001011 (b) 1001.011

(5) 有3个近似数12.345，1.7，0.8567和两个准确数30和0.956，求这5个数的平均值。

2. 平行截线的中点轨迹

(1) 平面几何告诉我们，给定两条相交直线 L_1，L_2，如果一组平行直线与这两条直线相交，则每条直线与 L_1，L_2 的两个交点的中点共直线。能用解析几何的方法验证吗？

(2) 如果一组平行直线与给定的抛物线相交，交点的中点仍在同一直线上吗？试用解析几何方法证明；

(3) 如果一组平行直线与给定的椭圆或双曲线相交，交点的中点仍在同一直线上吗？你能证明吗？

3. 高维空间与分维

(1) 你能举出现实问题中一两个高维(>3维)空间的例子吗？

(2) 你对爱因斯坦-闵可夫斯基四维时空如何理解？

(3) 你能举出几个分形现象吗？能找到分形或分维的一两个应用吗？

参考文献或网站

[1-1] 徐品方，张红：《数学符号史》，科学出版社(2006).

[1-2] 卡尔·萨根(Carl Sagan)：《数以十亿(Billions and Billions)》(碧声译)，《三思科学》电子杂志 No.8(2002).

[1-3] 谷歌的含义是什么？(what is meaning of Google?)www. google. com.

[1-4] 钱学森：关于马克思主义哲学和文艺学美学方法论和几个问题,《文艺研究》(1986).

[1-5] 王应麟：《三字经》,长春出版社(2007).

[1-6] 徐安东,王群慧,叶强：《大学计算机文化》,上海交通大学出版社(2006).

[1-7] 尼葛庞余帝(N. Negroponte)：《数字化生存(Being Digital)》(胡泳、范海译),海南出版社(1997).

[1-8] 加布里尔(Gabriel P.)：《矩阵,几何,线性代数(Matrizen, Geometry, Linear Algebra)》,Birkhaeuser(1996).

[1-9] 张苍：《九章算术》(白话译本),重庆大学出版社(2006).

[1-10] 杜石然,李俨：《中国古代数学简史》,中华书局(1963).

[1-11] 沃尔夫林(Heinrich Wolfflin)：《意大利文艺复兴艺术导论 (An Introduction to the Italian Renaissance)》(潘耀昌,陈平译),中国人民大学出版社(2001).

[1-12] 李旭甫：《哥白尼传》,经济日报出版社(2005).

[1-13] 库兹涅佐夫：《伽利略传》(中译本),商务印书馆(2001).

[1-14] 《路德文集》(中译本),上海三联书店(2005).

[1-15] 尚建新：《笛卡儿传》,河北人民出版社(1997).

[1-16] 恩格斯：《自然辩证法(Dialectics of Nature)》,人民出版社(1971).

[1-17] 百度：中国日本美国 GDP 统计,www. baidu. com.

[1-18] 樊映川：《高等数学讲义》,高等教育出版社(1964).

[1-19] 卢德馨：《大学物理学》,高等教育出版社(1998).

[1-20] 爱因斯坦：《狭义与广义相对论浅说(Relativity, the Special and the General Theory(A Popular Exposition))》(杨润殷译),北京大学出版社(2008).

[1-21] 纳博(Nabor, G. L.)：《闵可夫斯基时空几何(The geometry of Minkowski spacetime)》,Dover Publication(2003).

[1-22] 曼德布罗特(Mandelbort,B. B.)：《英国海岸线有多长？统计自相似和分数维。科学(How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. Science)》, New Series, Vol. 156, No. 3775. (1967).

[1-23] 维基百科：豪斯多夫维 www. zh. wikipedia. org.

[1-24] 法尔考纳(Falconer, K.)：《分形几何——数学基础及其应用(Fractal Geometry-Mathematical Foundations and Applications)》(曾文曲译),人民邮电出版社(2007).

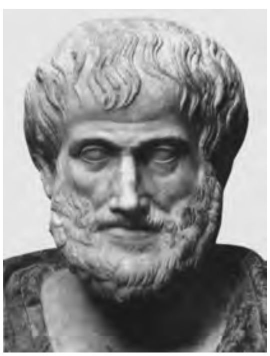
第二章

数学证明与公理化体系

2.1 形式逻辑简介

人们常说，数学是一门严密的科学。之所以严密，就是所有结论都要经过证明。而数学证明的主要工具是形式逻辑。这里我们对形式逻辑先做一个简单介绍。

形式逻辑产生于古希腊，人们公认其奠基人是亚里士多德（Aristotle，384-322 B. C.）。亚氏是柏拉图的学生，但创立了不同的哲学体系。“吾爱吾师，吾更爱真理”就是他的名言。亚里士多德研究的学问，包括了现在学科的哲学、物理学、生物学、天文学、大气科学、心理学、逻辑学、伦理学、政治学、艺术美学，几乎涵盖了所有的领域[2-1]。



亚里士多德　384－322 B.C.

在《Organon（工具论）》一书中，亚里士多德开创了形式逻辑这门学科，提出了四个公理：同一律、不矛盾律、排中律、充分理由律，并确定了概念、判断、推理等逻辑形式。他最先提出归纳和演绎两种方法，强调以数学与逻辑推理来证明科学的原理。亚里士多德当年所使用的许多专有名词，至今仍被教科书所使用。他开创的这套方法大大促进了科学的发展。在他的影响下，数学家欧几里得写出名著《原理》[0-7]，其主要结果和方法也仍为今日大学、中学的教科书所采用。

几乎与希腊学者同时，中国在百家争鸣的春秋战国时期，也曾经出现研究逻

辑的学派。在墨子(468 - 376 B. C.)[2-2]、公孙龙(325 - 250 B. C.)[2-3]和荀子(298-238 B. C.)[2-4]的著作中闪耀着逻辑思考的光辉。可惜中国古代数学偏重求解实际问题的方法而不重视逻辑推理,特别是汉代帝王“罢黜百家,独尊儒术”之后,中国的逻辑学未能得到很好发展。


墨子 468 - 376 B.C.


公孙龙 325 - 250 B.C.


荀子 298 - 238 B.C.

形式逻辑的主要内容是“概念”、“判断”和“推理”,在数学中对应着“定义”、“定理”和“证明”。

2.1.1 概念——反映事物特有属性的思维形式

任何学科都包含许多概念。比如文学中有“诗歌”、“小说”、“戏剧”等概念;音乐中有“声乐”、“器乐”等概念;数学中有“算术”、“几何”、“代数”、“微积分”、“概率统计”等概念。人们日常生活中,“父母”、“子女”、“桌椅板凳”、“柴米油盐酱醋茶”都是概念。

1) 概念的“内涵”与“外延”

概念的“内涵”是指概念所反映的事物的特有属性。

概念的“外延”是指具有概念所反映特性的所有事物。

概念的内涵与概念的外延是互相制约的。概念的内涵确定了,其外延也跟着确定。例如“人”的内涵是“能制造和使用生产工具的动物”;“工人”的内涵是“从事生产劳动和以工资收入为生活来源的人”;“女工人”的内涵是“女性的工人”。显然,“人”的外延比“工人”大而内涵却不如“工人”多;“女工人”的内涵比“工人”多,外延却更小。数学中的“复数”、“实数”、“有理数”、“整数”、“自然数”这几个概念也很相似。成立如下规律:

概念的内涵与外延成反比。即如果概念 A 的内涵比概念 B 的内涵多，则 A 的外延比 B 的外延小；反之亦然。

没有任何外延的概念称为“虚假概念”。例如“热爱和平的战犯”；“-10 K（绝对温度）的物质”；“$x^2+4=0$ 的实数解”；等等。

2）概念间的关系

如果把一个概念看作一个集合，用空集记号 $\varnothing$ 表示虚假概念，则任何两个概念 A 和 B 之间的关系都可用图 2.1 表示。

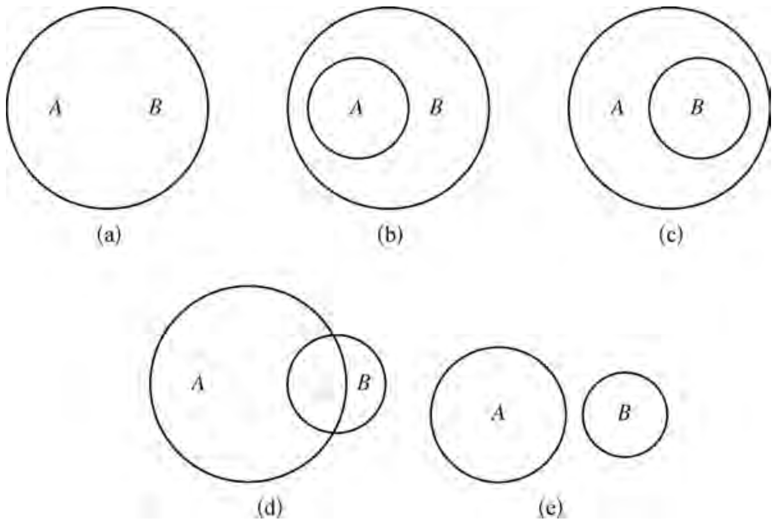


图 2.1

（1）全同关系 $A=B$（图 2.1(a)）。例如“《蜀道难》的作者”和“唐代诗人李白”全同；“中华人民共和国首都”和“北京市”全同；“相似三角形”与“两只对应角相等的三角形”全同。

（2）从属关系即蕴含关系。$A\subset B$，称为“B 是 A 的种概念”或“A 是 B 的属概念”，也即“A 蕴含于 B”或“B 蕴含 A”（图 2.1(b)）。例如“女工人”是“工人”的属概念，“工人”是“人”的属概念；“自然数 N”蕴含于“整数 Z”蕴含于“有理数 Q”蕴含于“实数 R”蕴含于“复数 C”；等。同理可以定义 $B\subset A$（图2.1(c)）。

（3）交叉关系。$A\cap B\neq\varnothing$（图 2.1(d)）。例如“工人”和“妇女”交叉；“整数”和“正数”交叉。

(4) 全异关系。$A\cap B=\varnothing$(图 2.1(e))。所谓"风马牛不相及"就是指"风"和"马"和"牛"都是全异关系。"圆周率"和"有理数"也是全异关系。

注意：如果两个概念 A 和 B 全异，即 $A\cap B=\varnothing$，又都是概念 U 的下属关系，即 $A\subset U$ 和 $B\subset U$，再可分为两种情况：

(1) A 和 B"不相容"(或"反对")：$A\cup B\subset U$(图 2.2(a))。

(2) A 和 B"互斥"(或"矛盾")：$A=\bar{B}$，或 $B=\bar{A}$(图 2.2(b))。

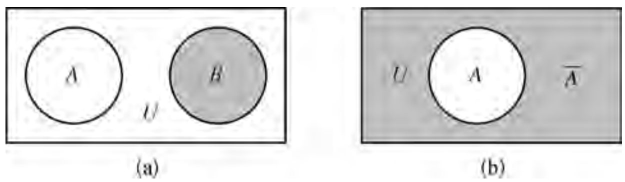


图 2.2

例如："中国人"和"美国人"不相容，但并不是互斥，因为"人类"中还有"俄国人"、"印度人"等既非"中国人"，又非"美国人"；但"中国人"和"非中国人"就是互斥关系。前述"圆周率"和"有理数"也是不相容，但非互斥，因为与"有理数"全异的实数还有 $\sqrt{2}$，$\sqrt{3}$，e 等；但"无理数"和"有理数"在实数范围内就是互斥的。

2.1.2　定义——揭示概念内涵的逻辑方法

科学的概念必须有清晰的内涵，从而确定其外延。例如："哲学是关于自然科学和社会科学的总结与概括"、"圆是到定点等距的平面动点轨迹"，等等。定义的一般形式是：

$$Ds\text{ 是 }Dp \tag{2.1}$$

其中 Ds 代表被定义项，Dp 代表定义项。

数学中的概念必须有严格的定义。一个概念的定义不一定唯一。例如圆也可以按墨子的定义为"圆，一中等长也"[2-5]，或用解析几何术语定义为"平面直角坐标系中满足方程 $(x-a)^2+(y-b)^2=R^2$ 的动点 (x,y) 集合"。但是同一概念的不同定义必须全同，可以互推，数学上称为"等价"。

【例 2-1】　中学学过数列。常用的等差数列和等比数列是用通项来定义

的。等差数列：

$$a_n = a_1 + (n-1)d \tag{2.2}$$

其中 a_1 是数列的首项；d 是公差。等比数列：

$$a_n = a_1 q^{n-1} \tag{2.3}$$

其中 a_1 是数列的首项；q 是公比。

定义项中不能直接或间接地包括被定义项。例如以下定义是不正确的："哲学是研究哲学的学问"、"生命就是有生命的物质的生理现象"，"圆是圆形物体的数学抽象"，等等。

循环定义是指用此事物说明彼事物，又用彼事物说明此事物。如：定义线段为"线段是直线中的一段"；定义直线为"直线是线段向两端无限扩展"。这是必须避免的。

数学中有时会用到"递归定义"来定义一个基本事物，在定义中一步一步引导出规则。如"数学表达式"可以如下定义：

（1）一个数或一个变量是数学表达式；

（2）用一对括弧将数学表达式括起来也是数学表达式；

（3）用"＋"、"－"、"×"、"/"、"$\sqrt{\quad}$"等运算符连接的两个数学表达式还是数学表达式。根据上面定义，可以看出：$3x^2y+\sqrt{\dfrac{x-y}{2xy}}$ 是数学表达式。

第一讲中提到的斐波那契数列也是通过递归方式定义的。即

$$f_1 = 1,\ f_2 = 2,\ f_n = f_{n-2} + f_{n-1} \quad (n \geqslant 3) \tag{2.4}$$

随着科学的发展，一个概念的定义可能会改变。例如物理中关于"热"和"光"的定义就改变和更新了许多次。又如平面几何讲到"圆"时，对"切线"的定义是"与圆有一个交点的直线"；但是对于一般曲线的切线，这个定义就不恰当，而必须用微积分知识来定义了。汉字简洁优美，有时会带来二义性，一句话按字面理解，可能是一个概念的定义，也可能是一个概念的属性。中国古代逻辑学家公孙龙曾提出"白马非马"的命题，按常规理解，"白马"当然是"马"，说"白马非马"岂不是说"张三非人"吗？当时就引起学术界的辩论。相传孔子的六世孙孔穿，找上门去辩论，结果被公孙龙驳得无以应对，吃了败仗。其实，这就是概念定

义与属性的混淆。普通人说“白马是马”，意思是“白马是马的一种”，或者按集合论的语言，“白马是马这个集合的子集”，也就是说，“白马”是“马”的属概念。但是作为“白马”的定义，说“白马是马”就不正确了。公孙龙正是从概念的逻辑关系出发，指出：“马者，所以命形也；白者，所以命色也。命色者非命形也。故曰：‘白马非马’。”[2-3]

在学术论战和法庭辩论等场合，必须防止概念的混淆或偷换。第二次世界大战结束不久，1946 年 5 月 4 日，十一个同盟国在东京设立远东国际军事法庭，审判 28 名日本甲级战犯。中国被日本侵略最久，受害最大，中国检察组理所当然地承担了最重的起诉任务。在起诉和辩论过程中，他们经常驳斥被告律师歪曲定义、偷换概念的伎俩，以下就是两个案例。

开庭不久，被告律师以中日直到 1941 年 12 月 9 日才宣战的事实，辩称在此之前，不存在战争，自然谈不上战争罪。这样，“九一八事变”、“卢沟桥事变”、“南京大屠杀”都可一笔勾销！显然这是歪曲了战争的定义。从中外词典或法学教科书中都可找到战争的定义，虽叙述有详略，但核心意思都是指“国家或集团之间的武装斗争。”[2-6]中国检察官向哲浚（1892-1987）在 1946 年 5 月 14 日的发言中予以痛斥：“这是一个关于战争的正确定义的问题。从 1931 年9 月 18 日起，日本就在中国采取了军事行动，杀死了数以万计的中国人，包括士兵和平民。十四年前，即 1937 年 7 月 7 日，日本在卢沟桥发动了战争，一晚就杀死了几百人。随后，日本出兵中国各地，杀死了数以百万计的中国士兵、儿童、妇女和手无寸铁的无助的平民。这些是全世界都知道的事实，如果这不是战争，我想问，那什么是战争？”[2-7]中国代表团还根据侵华日军于 1928 年制造“皇姑屯事件”，炸死张作霖的事实，指出一个国家用阴谋手段杀害另一个国家的政府首脑，就是战争行为。国际法庭采纳了中国代表团的意见，确定 1928 年 1 月 1 日为对日本战犯起诉的起始日。[2-8]

向哲浚　1892-1987

又如在被告土肥原贤二和板垣征四郎的个人辩护阶段，中国检察官首席顾问倪征　（1906-2004）提出一份日本关东军的《奉天特务机关报》，该报首页盖有

土肥原的名章。其中一页载有“华南人士一闻土肥原和板垣之名，有谈虎色变之慨”等语。土肥原确实是“中国通”，当时在报告中引用中国成语，炫耀他们在中国的“战功”，想不到成为日后的罪证。不料土肥原的美籍辩护律师华伦(F. N. Warren)见势不妙，竟跳出来向法庭说，这个文件讲的是一只老虎，与案情无关，法庭应拒绝接纳作为证据。这是典型的偷换概念，因为“谈虎色变”中的“虎”已由“凶恶的老虎”引申为“令人害怕憎恶的人或事”。倪征𣋉当即戳穿他的拙劣表现。他说，“谈虎色变”是一句中国成语，意思是中国人一谈起土肥原、板垣两人，有如提到猛虎，足见这两人的凶恶。全场肃静的法庭这时哄堂大笑[2-9]。

倪征𣋉 1906-2004

正是由于中国和其他各同盟国代表团的共同努力，正义得到伸张，东条英机、土肥原贤二、板垣征四郎等七名罪大恶极的战犯被押上历史的绞刑架[2-10]。

2.1.3 判断——对事物有所肯定或否定的思维形式

1) 命题的“真”与“假”

逻辑的主要作用是判断一个命题是否成立，非真即假。数学的重要功能也是证明一个命题是否成立。

“命题”必须是一个陈述句，而不能是疑问句或感叹句。通常用字母 P、Q 等表示命题。例如：用 P 表示“所有人在法律面前平等”；用 Q 表示“两条平行线的同位角不相等”；等等。

一个命题有两种可能的取值：“真”(记为 T，True 的第一字母)或“假”(记为 F，False 的第一字母)，两者必居其一，也只居其一。前面两个例子中，“所有人在法律面前平等”是真，可以记为 $P \in T$，而在欧几里得几何中，“两条平行线的同位角不相等”是错的，即 $Q \in F$，或者 $Q \notin T$。

2) A、E、I、O 四种直言命题

直言命题是人们思维中最常见的命题，也是数学中出现最多的命题。主要

有以下四类：

（1）全称肯定判断（A 判断）："所有的 S 都是 P"。

例如："一切罪犯都要受法律惩罚"；"所有的物质都由原子组成"；"所有自然数都是有序的"；等等。

（2）全称否定判断（E 判断）："所有的 S 都不是 P"。

例如："所有法律都不能与宪法相违"；"所有无机物都不具有生命特征"；"所有自然数都不是负数"；等等。

（3）特称肯定判断（I 判断）："有的 S 是 P"。

例如："有的小说是描写历史故事的"；"有的金属是液态的"；"有的自然数是素数"；等等。

（4）特称否定判断（O 判断）："有的 S 不是 P"。

例如："有的小说不是描写历史故事的"；"有的金属非固态"；"有的自然数不是素数"；等等。

两个肯定判断由拉丁字 Affirms（肯定）的元音字母 A 和 I 表示，两个否定判断则由拉丁字 Nego（否定）的元音字母 E 和 O 表示。

日常生活和科学技术中还有两种直言命题很常见：

（5）单称肯定判断："某一个确定的 S 是 P"。

例如："祖冲之是中国古代数学家"，"π 是无理数"，等等。

（6）单称否定判断："某一个确定的 S 不是 P"。

例如："李世民不是中国古代数学家"，"π 不是有理数"，等等。

由于主词是"某一个确定的 S"，这是个只有一个元素的集合，所以我们有时把"单称肯定判断"和"单称否定判断"分别归于"A 判断"和"E 判断"。

直言命题都由量词（Classifier）、主词（Subject）、系词（Copula）和谓词（Predicate）四部分组成。

数学中常用两个逻辑符号表示量词的全称和特称。

$\forall$：表示"所有的"、"一切的"（英文 All 或 Any 的第一个字母 A 上下颠倒）；

$\exists$：表示"有的"、"存在着"（英文 Exist 第一个字母 E 左右颠倒）。

主词和谓词比较清楚；系词"是"或"不是"在中文中有时不出现，例如"一切罪犯都要受法律惩罚"＝"一切罪犯是要受法律惩罚的"，"有的金属非固态"＝

“有的金属不是固态”,等等。

3) 四种直言命题的关系

不妨把主词 S 和谓词 P 看作两个概念,将图 2.1 中的 A 和 B 换为 S 和 P,则四种直言判断可以表述为:

$$
\begin{aligned}
&A\text{ 判断}: \forall x \in S \rightarrow x \in P \quad \text{或} \quad S \subseteq P\{(1),(2)\}\\
&E\text{ 判断}: \forall x \in S \rightarrow x \notin P \quad \text{或} \quad S \subseteq \bar{P}\{(5)\}\\
&I\text{ 判断}: \exists x \in S \rightarrow x \in P \quad \text{或} \quad S \cap P \neq \varnothing \{(1),(2),(3),(4)\}\\
&O\text{ 判断}: \exists x \in S \rightarrow x \in \bar{P} \quad \text{或} \quad S \cap \bar{P} \neq \varnothing \{(3),(4),(5)\}
\end{aligned} \tag{2.5}
$$

根据 S 和 P 两个概念外延的关系及 2.5 式,A、E、I、O 四种判断逻辑上有着密切的联系。

(1) $A \subset I$, A 真 I 必真;$\bar{I} \subset \bar{A}$, I 假 A 必假。

(2) $E \subset O$, E 真 O 必真;$\bar{O} \subset \bar{E}$, O 假 E 必假。

(3) $A \cap E = \varnothing$, A、E 不同真;但 $\bar{A} \cap \bar{E} = \{(3), (4)\}$,$A$、$E$ 可同假。

(4) $A \cap O = \varnothing$, $\bar{A} \cap \bar{O} = \varnothing$, A、O 不同真也不同假。

(5) $E \cap I = \varnothing$, 但 $\bar{E} \cap \bar{I} = \varnothing$, E、I 不同真也不同假。

(6) $I \cap O = \{(4)\} \neq \varnothing$, I、O 可同真;但 $\bar{I} \cap \bar{O} = \varnothing$, I、O 不可同假。

用逻辑术语表达,A 和 O 及 E 和 I 是“矛盾”关系(即“互斥”关系);A 和 E 及 I 和 O 是“反对”关系(即“不相容”关系);A 和 I 及 E 和 O 是“差等”关系(即“从属”关系)。关于 A、E、I、O 四种判断逻辑联系的详细讨论参阅[2-5]。

A 和 O 或 E 和 I 的“矛盾”关系常用来推翻一个全称肯定判断或全称否定判断。比如一件罪证,便足以驳斥一个犯罪嫌疑人声称自己一切活动都奉公守法的谎言。这也是法律事务中特别注重人证物证的原因。

1937 年 12 月 13 日,侵华日军攻占南京,疯狂的侵略者进行了长达六周的血腥大屠杀。然而,狡猾的日本军方封锁消息,在失败前夕更是尽力销毁罪证,所以国际舆论甚至日本国民都不知道日军在南京犯下的滔天大罪。东京审判期间,日本辩护团竟然矢口否认屠杀南京市民的事实,并妄称屠杀中国军人俘虏是误传,甚至反诬对外国权益和财产的侵犯可能是中国军队干的。为弄清“南京大屠杀”真相,远东国际军事法庭设立独立单元进行调查。中国检察组早有准备,

事先经过艰难取证，向法庭提供了包括中国人和美国人在内的 13 人的宣誓证词；并由三位死里逃生的中国证人和一位美籍医生罗伯特·威尔逊和一位美籍神父约翰·马基登台作证，马基神父曾拍摄长达 105 分钟的纪录影片，真实史料震惊了法庭和世界。南京大屠杀的元凶，时任日军司令的松井石根终于被判死刑。[2-10]

4）假言判断

断定某一命题成立是另一命题成立的条件的判断，称为“假言判断”。即：

$$\text{“如果 } P\text{，那么 } Q\text{”，或 } P \Rightarrow Q \tag{2.6}$$

例如：“如果一个人具备逻辑知识，他的思维会更清晰正确”，“如果两个三角形的对应角相等，则它们相似”，等等。

2.1.4 数学定理

1）数学定理举例

数学中的判断多为假言判断，即在给定条件下推出某个或某些结论，并多以定理或猜想的形式出现。下面举出几个著名的定理或猜想。

【例 2-2】 勾股定理。设直角三角形的两直角边长为 a，b，斜边长为 c，则有

$$a^2 + b^2 = c^2 \tag{2.7}$$

这个定理现在已列入初中教材。最早是谁发现的？西方数学史家通常归功于公元前六世纪古希腊哲学家毕达哥拉斯，并把满足式 2.7 的一组正整数称为“毕达哥拉斯数”。但是按《周髀算经》的记载，早在公元前十一世纪西周开国时代，有个名叫商高的人已经明确指出了“勾三股四弦五”的关系，比毕氏早五百年。中国古称直角边为“勾”与“股”，称斜边为“弦”或“径”，这就是“勾股定理”命名的由来，也有人纪念商高而命名为“商高定理”[2-11]。不过考古发现，人们对勾股定理特例的认识可能更早。公元前 1700 年巴比伦泥板上已经刻有不少组毕达哥拉斯数，如：(45，60，75)，(65，72，97)，(119，120，169)，(1679，2400，2929)，等等；甚至有人认为公元前 3000 年英国石庙在建造时就用到毕达哥拉斯数。看来勾股定理是人类认识世界、探求真理的共同创造。实际上，存在无数组“毕达哥拉斯数”。

【例 2-3】 费马(Pierre de Fermat，1601-1665)大定理

$$x^n + y^n = z^n \quad \text{当 } n > 2 \text{ 没有整数解} \tag{2.8}$$

当 $n=1$,x 和 y 取任何整数值,都可找到 z 对应的整数值,也就是说有无数组正整数解。

当 $n=2$,式 2.8 就是勾股定理 2.7。(x, y, z)也有无数组整数解,例如任取两个奇数 m 和 n，且 $m > n$，则满足

$$a = mn, \ b = \frac{m^2 - n^2}{2}, \ c = \frac{m^2 + n^2}{2} \tag{2.9}$$

的 (a, b, c) 都是毕达哥拉斯数。

从古希腊以后,许多人企图寻找 $n > 2$ 时式 2.8 的整数解,均无结果。费马在 1637 年给友人的信中写道:“不可能将一个立方数写成两个立方数之和,或者将一个四次幂数写成两个四次幂之和。一般地,对任何一个数,其幂次大于 2,就不可能写成同幂次的另两数之和。对此我得到了一个真正奇妙的证明,可惜空白太小无法写下来。”费马死后人们始终没有找到他那“奇妙的证明”,以至于费马到底是否真正证出 2.8 式成为千古之谜[0-10]。

【例 2-4】 哥德巴赫(Christin Goldbach,1690-1764)猜想[2-12]。1742 年德国数学教师哥德巴赫慕名给当时的大数学家欧拉写信,提出关于正整数和素数之间关系的两个推测:

猜想 1:如果一个偶数大于等于 6,那么它一定可以表示为两个奇素数之和。

例如:$6 = 3 + 3$，$8 = 3 + 5$，$10 = 5 + 5$，$12 = 5 + 7$，…

猜想 2:如果一个奇数大于等于 9,那么它一定可以表示为三个奇素数之和。

例如:$9 = 3+3+3$，$11 = 3+3+5$，$13 = 3+3+7$，$15 = 3+5+7$，…

前面提到过,素数是除 1 及本身之外无其他自然数可整除的自然数。素数集合为无限集:$\{2, 3, 5, 7, 11, 13, \cdots\}$,除了 2 以外,都是奇素数。

猜想 2 是猜想 1 的推论。因为一个奇数减去一个奇素数一定是偶数,如果猜想 1 成立,猜想 2 自然也成立。

【例 2-5】 代数基本定理：n 次代数方程

$$a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0 \quad (a_0 \neq 0) \tag{2.10}$$

在复数域内有且仅有 n 个根。例如：

$$n = 1, \quad a_0x + a_1 = 0 \quad (a_0 \neq 0) \Rightarrow x = \frac{a_1}{a_0}$$

$$n = 2, \quad a_0x^2 + a_1x + a_2 = 0 \quad (a_0 \neq 0) \Rightarrow x_{1,2} = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0a_2}}{2a_0} \tag{2.11}$$

2) 定理的四种形式及其关系

正定理：如果 P，那么 Q。即 $P \Rightarrow Q$

逆定理：如果 Q，那么 P。即 $Q \Rightarrow P$

否定理：如果 $\overline{P}$，那么 $\overline{Q}$。即 $\overline{P} \Rightarrow \overline{Q}$

逆否定理：如果 $\overline{Q}$，那么 $\overline{P}$。即 $\overline{Q} \Rightarrow \overline{P}$

正定理及其逆否定理同真同假；但是正定理的真假，却不能断定逆定理和否定理的真假。

如果把 P 和 Q 也看作集合，那么 $P \Rightarrow Q$ 相当于 $P = Q$ 或 $P \subset Q$。当 $P = Q$，定理的四种形式同时成立；但当 $P \subset Q$，$\overline{Q} \subset \overline{P}$，即逆否定理成立，而 $Q \subset P$（逆定理）和 $\overline{P} \subset \overline{Q}$（否定理）不成立。

【例 2-6】 正定理：如果两个三角形全等⇒对应角相等(T)；

逆定理：对应角相等⇒两个三角形全等(F)，因为对应角相等的两个三角形可能仅仅相似而不全等；

否定理：两个三角形不全等⇒对应角不相等(F)，两个不全等但相似的三角形对应角还是相等；

逆否定理：对应角不相等⇒两个三角形不全等(T)。

如果把上述正定理中的“全等”换为“相似”，那么四种形式的定理都真。

3) 充分条件，必要条件和充分必要条件

若 $P \Rightarrow Q$，P 真 Q 也真，或有 P 必有 Q，称 P 为 Q 的“充分条件”；

Q 假 P 必假，或无 Q 必无 P，称 Q 为 P 的“必要条件”。

若 $P \Rightarrow Q$，同时 $Q \Rightarrow P$，则称 P 为 Q 的“充分必要条件”，简称“充要条件”，

或者说 P 与 Q “等价”，记为 $P \Leftrightarrow Q$。

在日常生活中有许多例子。例如在人才招聘告示上常见“必须××岁以下具有大专学历”等字样。这里“××岁以下具有大专学历”就是录用的必要条件，因为不具备此条件，不可能被录用；但不是充分条件。因为即使具备此条件，仍不一定被录用。又如“经名师指点”是“学业进步”的充分条件，却不是必要条件，因为也有人自学成才。

充分条件，必要条件和充分必要条件在数学上更是常见。例如：

$x=2 \Rightarrow x^2=4$，$x=2$ 是 $x^2=4$ 的充分条件，却非必要，因为 $x=-2$ 也可推出 $x^2=4$。【例 2-6】中“三角形全等”是“对应角相等”的充分条件而不是必要条件；“对应角相等”是“三角形全等”的必要条件而不是充分条件。而“三角形相似”却是“对应角相等”的充要条件。

2.1.5 推理——通过已知判断得出未知判断的思维过程

推理能力是人和高级灵长类动物的主要差别之一，任何学科以及日常生活中都少不了。

中国古典诗歌充满着浪漫和夸张，却也常有蕴含着推理的精彩作品。宋代大政治家王安石的诗作中常有理性思维。例如：“墙角数枝梅，凌寒独自开。遥知不是雪，为有暗香来。”[2-13]因为梅花有香而雪无香，所以从“暗香”推理出墙角是梅而“不是雪”。彭漪涟教授的专著《古诗词中的逻辑》[2-14]有许多精彩的论述和例证。

王安石　1021-1086

福尔摩斯

柯南道尔　1859-1930

文学作品中将推理发挥到极致的当然是侦探小说。看过《福尔摩斯探案集》[2-15]的读者，都为小说作者柯南道尔（Arthur Conan Doyle，1859-1930）的推

理描述所折服。在第一篇《血字的研究》中，福尔摩斯说："推断和分析的科学也像其他技艺一样，只有经过长期和耐心的钻研才能掌握；初学的人，不妨先从掌握较浅显的问题入手。比如遇到了一个人，一瞬之间就要辨识出这人的历史和职业。"他还"历时不到一秒钟"就推断出初次见面的华生是参加过阿富汗战争的军医。他的推理过程是这样的："这一位先生，具有医务工作者的风度，但却是一副军人气概。那么，显见他是个军医。他是刚从热带回来，因为他脸色黝黑，但是，从他手腕的皮肤黑白分明看来，这并不是他原来的肤色。他面容憔悴，这就清楚地说明他是久病初愈而又历尽了艰苦。他左臂受过伤，现在动作起来还有些僵硬不便。试问，一个英国的军医在热带地方历尽艰苦，并且臂部负过伤，这能在什么地方呢？自然只有在阿富汗了。"

1）三段论的推理形式

"三段论"是用得最多的逻辑推理形式，有如下规则：

（1）由三个判断组成。第一个称为"大前提"，第二个称为"小前提"，第三个称为"结论"；

（2）一共包含也只能包含三个概念，每个概念在两个判断中各出现一次；

（3）在大小前提中出现而在结论中不出现的概念称为"中项"。

例如："一切科学规律是不以人们意志为转移的，逻辑学的规律是科学规律，所以逻辑学的规律是不以人们意志为转移的。"

这里的三句话就分别是"大前提"、"小前提"和"结论"。其中只有三个概念："科学规律"（出现在两个前提中各一次，即"中项"），"逻辑学的规律"（出现在小前提和结论中各一次），"不以人们意志为转移"（出现在大前提和结论中各一次）。

"三段论"推理中，M 是中项，S 和 P 分别是结论中的主词项和谓词项。根据两个前提中 S、P 和 M 的位置，可以分为 4 个"格"；每个格又可按前提中 M、S 和 P 的位置分为许多"式"。最常用的是下面的格式：

大前提：	M——P	$M\subseteq P$（A 判断）
小前提：	S——M	$S\subseteq M$（A 判断）
——————	即	—————— （2.12）
结　论：	S——P	$S\subseteq P$（A 判断）

在式 2.12 的推理格式中，大小前提和结论都是 A（全称肯定）判断。由于三

句话每句都可以有 A、E、I、O 这四种判断类型，共有 $4^3=64$ 种“式”。不过大多不符合推理的要求。例如：大前提：“书法展览中有的作品是隶书”(I)，

小前提：“张三的作品参加了书法展览”(A)，

结　论：“张三的作品是隶书”(A)。

这个推理显然不成立，因为张三的作品可以不是隶书，而是篆书、楷书，或行书、草书。读者不妨尝试用其他的判断类型，确定能否有效推理。

“三段论”推理在逻辑上称为“演绎推理”。大前提通常是一般规律，小前提指出具体对象，从而推出结论，这是“从一般到特殊”的过程。只要前提为真，结论也一定为真；反过来，如果推出错误的结论，只有两种可能：或者是前提出错，或者是推理违规。

推理根据的判断叫“前提”。整个数学就是以为数不多的公理作为前提，按照严格的推理形式，证明出一个又一个判断，称为“定理”。已经证明正确的定理又作为前提，再推出新的结论。

在政治法律领域，推理非常重要。在远东国际军事法庭的法官座次排定过程中，中国法官梅汝璈(1904-1973)通过有理有利有节的推理，捍卫了中国的荣誉。[2-16]东京审判共有 11 位法官，庭长韦勃是澳大利亚人，自然坐在法官席的正中。其左右应当是哪国法官？一种自然推理是：“大前提：法官座次按在日本投降书上签字次序排列”，“小前提：美国、中国在日本投降书上签字为前两名”，“结论：美国和中国法官应坐在庭长两侧”。但是庭长却希望让美国和英国法官居中，于是提出种种不同的“大前提”，例如“按联合国安全理事会五强顺序”、“按国名顺序”、“按法官年资为序”，等等。目的是排斥中国法官居中，这些理由都经不起推敲而未被接纳。梅汝璈最后说：“我看依照日本投降书上受降签字的次序安排坐席最为合理。”他幽默地“建议”：换一个“大前提”：“我们不妨找一个体重测量器来看看个人的体重是多少，然后按照它来安排席次，体重者居中，体轻者居旁。”梅汝璈郑重指出：“中国是受日本侵略最惨烈、抗战最久、牺牲最大的国家，在审判日本战犯的国际法庭里她应有的席位竟会被

梅汝璈　1904-1973

降低到一贯只知向日本投降的英国之下，这是不可思议的事情！”在他的力争下，中国法官终于坐在法官席上庭长的左侧，参加了长达近三年的审判。

东京审判法官席(右5为梅汝璈)

2）归纳推理

逻辑中，还有一类“归纳推理”，是一个“从特殊到一般”的过程。

生活中人们常用类比推出某些结论。

例如，某单位人事部门发现，连续几年的新职工中，某几所大学的毕业生表现较好，于是得出结论：今后就从这几所大学招聘毕业生。

科学发现更是常出于对观测或实验现象的归纳。例如开普勒(Johannese Kepler，1571-1630)的天体运动定律，门捷列夫(Dmitri Mendeleev，1834-1907)的元素周期表等。

开普勒　1571-1630

门捷列夫　1834-1907

白居易　772-846

甚至在诗歌中常用的“比兴”手法，实质上也是归纳推理。例如白居易(772-846)有一首诗[2-17]：

赠君一法决狐疑，不用钻龟与祝蓍。

试玉要烧三日满，辨材须待七年期。

周公恐惧流言日，王莽谦恭未篡时。
向使当初身便死，一生真伪复谁知？

他通过“试玉”、“辨材”两种对“物”的判断以及对“周公”和“王莽”这两个人的评价，归纳出“对人、对事要得到全面的认识，都要经过时间的考验，从整个历史去衡量、去判断”的结论。

数学中的推理就是证明，我们后面将专门介绍。

日常生活中，常会遇到似是而非的推理，甚至很有学问的科学工作者也会犯逻辑错误。例如经常有人指责中医药“不科学”，甚至说成是“伪科学”，主张“立即停止缺乏科学理论、违背科学精神、没有安全保障的中医中药研究”。他们的推理大略有以下几种：

推理一 大前提:“作为科学，医药必须写出分子式”；
小前提:“许多中药写不出分子式”；
结论:“中药不是科学。”

推理二 大前提:“中医是科学，必定能治好各种病”；
小前提:“我母亲得病，看中医没治好，死了”；
结论：由反证法，“中医不是科学”。

推理三 大前提:“科学必须让人看得懂”；
小前提:“我能看懂量子力学、相对论，却看不懂中医的阴阳五行理论”；
结论:“中医理论不是科学。”

请读者思考和讨论，上述推理是否合乎逻辑。

由于逻辑方法的重要性，我国国务院学位委员会办公室组织的硕士学位入学资格考试中包括“逻辑推理能力测试”[2-19]，不少大企业对求职者也会进行逻辑能力的面试。下面是几个典型例题。

【例 2-7】 一公司面临减少员工数量。董事会决定首先解雇效率较低的员工，而不是解雇工龄短的员工。这个决定基于哪个假定？

A. 有能比较准确判定员工效率的方法；

B. 每个人的工作效率都不相同；

C. 最有经验的员工是最好的员工；

D. 薪酬越高的员工效率越高。

正确答案是A。只有建立一套能准确判定员工效率的方法或准则，才能实施董事会的决定。

【例2-8】 某手机店失窃，赵、钱、孙、李四人涉嫌被调查。4人笔录如下：

赵："窃贼是孙"；

钱："窃贼是李"；

孙："如果我作案，那么李是主犯"；

李："我没有偷。"

这四人只有一人说假话，判定以下哪项为真？

A. 赵说假话，窃贼是钱；

B. 钱说假话，窃贼是孙；

C. 孙说假话，窃贼是孙；

D. 李说假话，窃贼是孙和李。

正确答案是D。因为钱和李的笔录矛盾，两者必有一假，故只有B或D为真；若钱说假话。其他三人都讲真话，但孙和李的笔录矛盾，所以B为假；若李说假话，那么李偷了手机，其他三人都说真话，则除李之外，孙也参与作案，因而D为真，即窃贼是孙和李，且李是主犯，孙是从犯。

【例2-9】 赵、钱、孙、李、周在假日准备参加三项活动：看电影、打球、旅游，但每人只能参加其中一项，且遵循下列条件：

(1) 赵、钱、孙三人参加的活动皆不相同；

(2) 只有两人打球；

(3) 李、孙参加不同活动；

(4) 如果赵、周其中一人去看电影，另一人也去看电影。

作出以下六个判断：

1. 下面哪一项准确列出了赵—钱—孙—李—周五人可以分别参加的活动？

A. 影—球—球—旅—影；

B. 影—旅—球—球—影；

C. 球—旅—影—球—影；

D. 球—旅—影—球—旅。

2. 若周去打球，则下面除了哪一项之外都可能正确？

A. 钱看电影；　B. 赵打球；　C. 赵旅游；　D. 李打球。

3. 下面哪两个人可以一起看电影？

A. 钱李；　B. 赵李；　C. 钱周；　D. 孙钱。

4. 下面除哪一项外一定错误？

A. 只有钱旅游；　B. 只有孙旅游；

C. 只有李旅游；　D. 只有周旅游。

5. 若赵打球，则所有可以看电影的人是

A. 钱；　B. 钱李；　C. 孙周；　D. 孙钱李。

6. 若把“只有两人打球”的条件改为“只有三人打球”，则下面哪一人必须参加除打球之外的某一项活动？

A. 赵；　B. 钱；　C. 孙；　D. 李。

正确答案：1. D；2. D；3. A；4. B；5. D；6. C。请读者自行验证。

2.2 数学证明方法

数学命题要成立，必须经过逻辑上严格证明。反过来，只有经过证明的数学命题才能称之为定理，否则只能是猜想。下面介绍四种常用的数学证明方法。

2.2.1 “三段论”直接证明

这是最常见的证明方法。

【例 2-10】 大前提：n 次代数方程(2.10 式)在复数域有且只有 n 个根(代数基本定理)；

小前提：$x^3 - x^2 + 4x - 4 = 0$ 是 3 次代数方程；

结　论：$x^3 - x^2 + 4x - 4 = 0$ 在复数域有且只有 3 个根

这里的大前提是代数基本定理。有意思的是，迄今为止，这个在代数学中起着基础作用的定理尚无纯代数方法的证明。代数基本定理的第一个严格证明通常认为是德国数学家高斯于 1799 年给出的（哥廷根大学的博士论文），现有 200 多种证法。特别是用复变函数的知识，人们得到一个优美的证明。[2-18]

值得注意的是，根据代数基本定理，只能证明上述方程有 3 个根，却不能求出这 3 个根。事实上，本题可以通过因式分解，

$$x^3 - x^2 + 4x - 4 = (x-1)(x^2+4) = (x-1)(x-2i)(x+2i) = 0$$

求出 3 个根为

$$x_1 = 1,\ x_2 = 2i,\ x_3 = -2i$$

其中 $i = \sqrt{-1}$ 是虚数单位。

从古至今，人们为了寻找一般的求根方法，可以说前仆后继。

当 $n=1,\ 2$，式 2.11 给出了求根公式。这两个公式已无法考证出于何时何人之手，现在是中学数学教材中的基本内容之一。

当 $n > 2$，人们还想找出类似于式 2.11 的求根公式，却经历了漫长的历程。

1535 年意大利数学天才塔塔利亚（Tartaglia N.，约 1499-1557）在一场数学擂台赛中遇到一个挑战性问题：“一块蓝宝石卖了 500 金币，所得利润是其成本的立方根，求其利润。”

设利润为 x 金币，容易建立 x 满足的方程：

$$x^3 + x = 500$$

塔塔利亚很快求出结果，获得胜利。另一位意大利人卡尔丹（Girolamo Cardano，1501-1576）在立下保密的誓言后，向塔塔利亚请教，塔氏以诗歌形式透露了他解上述 3 次方程求根的秘密：

卡尔丹　1501-1576

立方共诸物，和要写右边，巧设两个数，差值同右和；
此法要牢记，再定两数积：诸物三之一，还把立方计；
既知差与积，两数算容易，复求立方根，相减题答毕。

卡尔丹经过钻研，在1545年出版了数学专著《大术》，包括三次方程的求根公式。

考虑

$$x^3+px+q=0 \quad (p,\ q\in R) \tag{2.13}$$

（对于更一般的形式 $ay^3+by^2+cy+d=0$（$a\neq 0$），作代换 $x=y+b/3a$ 就可化为上述形式）

根据一元一次方程和一元二次方程的求解公式，塔塔利亚猜想一元三次方程应有如下形式：

$$x=\sqrt[3]{A}+\sqrt[3]{B}=A^{1/3}+B^{1/3} \tag{2.14}$$

将上式取立方：

$$\begin{aligned}x^3&=A+3A^{2/3}B^{1/3}+3A^{1/3}B^{2/3}+B\\&=A+B+3(AB)^{1/3}(A^{1/3}+B^{1/3})\\&=A+B+3(AB)^{1/3}x\end{aligned}$$

则得 $x^3-3(AB)^{1/3}x-(A+B)=0$，与原方程2.13比较：

$$p=-(AB)^{1/3},\ q=-(A+B)$$

根据韦达(Francois Viete，1540-1603)定理，A 和 B 是一元二次方程 $y^2+qy-(p/3)^3=0$ 的根，由式2.14，就得到原方程的一个根，这就是著名的卡尔丹公式。当时只能求一组实根，又过了大约200年后，随着人们对虚数认识的加深，1732年，瑞士数学家欧拉找到了一元三次方程三个根的完整表达式：

$$x_1=\sqrt[3]{A}+\sqrt[3]{B},\ x_2=\sqrt[3]{A}\omega+\sqrt[3]{B}\omega^2,\ x_3=\sqrt[3]{A}\omega^2+\sqrt[3]{B}\omega \tag{2.15}$$

其中：

$$\omega=\frac{-1+i\sqrt{3}}{2},\ A=-\frac{q}{2}+\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3},\ B=-\frac{q}{2}-\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}$$

当 $n=4$，很快由卡尔丹的学生费拉里(Ludovico Ferrari，1522-1565)推出了求根公式。考虑

$$x^4+bx^3+cx^2+dx+e=0 \tag{2.16}$$

移项变为 $x^4+bx^3=-cx^2-dx-e$，两边加上 $(bx/2)^2$，左边为完全平方

$$(x^2+bx/2)^2=(b^2/4-c)x^2-dx-c$$

两边再加 $(x^2+bx/2)y+y^2/4$，化为

$$(x^2+bx/2+y/2)^2=(b^2/4-c+y)x^2+(by/2-d)x+y^2/4 \tag{2.17}$$

y 是参数，取任何值两边都相等。设法使上式右边也成为完全平方，即令

$$\Delta=(by/2-d)^2-(b^2/4-c+y-e)=0 \tag{2.18}$$

这是关于 y 的二次方程，解出 y 的两个根，就可由式 2.17 得到 x 的四个根。

建立卡尔丹公式和费拉里公式都遵循三段论推理。比如“等量加等量仍然相等”是公理，二次方程求根公式和韦达定理是已经证明的结论，在推理过程中都作为大前提。但是这两个公式都很复杂，以至于现在绝大多数中学大学都不介绍。在工程实际问题中遇到三次或四次代数方程求根时，除能方便因式分解的特殊情况外，人们更多会采用近似方法借助计算机求根。

伽罗华　1811-1831

至于 $n\geqslant 5$，人们又花了两百年，始终未找到哪怕是很复杂的求根公式。数学家开始产生逆向思维，莫非此时干脆不存在用代数式表达的求根公式？天才的法国数学家伽罗华(Evariste Galois，1811-1831)完美地解决了这个问题。他证明了五次以上的代数方程不存在一般的求根公式，并由此创造了今天称为“群论”的新的数学分支。用此方法，他还解决了三大作图问题中的两个：“不能三等分任意角”和“体倍增不可能”。可惜伽罗华少年气盛，决斗身亡，年仅 20 岁！[0-1]

2.2.2　反证法

基于“三段论”的直接证明并不容易。比如“等量加等量仍相等”是一条尽人

皆知的公理,可是在证明一个结论的过程中加什么“等量”却大有讲究。比如上面介绍证明费拉里公式的过程中先后加了两次“等量”:$(bx/2)^2$ 和 $(x^2+bx/2)y+y^2/4$,非常巧妙,人类想出这两个“等量”,至少花了一千年!

反证法其实也是基于“三段论”:先假定待证明的结论不成立,作为前提,然后推出与已知命题矛盾的命题。因为“三段论”推理不可能错,那么错误结论只可能产生与假设的前提出错,从而反证待证结论的正确。当直接证明不易时,反证法可能会意想不到的便捷。

【例 2-11】 证明 $\sqrt{2}$ 是无理数。

假定 $\sqrt{2}$ 是有理数 m/n,其中 m 和 n 是整数并且无公约数,即不可能都是偶数。

$$\sqrt{2}=\frac{m}{n}\Rightarrow 2=\frac{m^2}{n^2}\Rightarrow m^2=2n^2,\ m\text{ 是偶数,记 } m=2k$$

$$(2k)^2=2n^2\Rightarrow 2k^2=n^2$$

于是 n 也是偶数,与假设矛盾。所以证明了 $\sqrt{2}$ 不是有理数而是无理数。

2.2.3 举反例方法

数学定理的结论大多为全称肯定判断 A:“所有的 S 都是 P”,其矛盾判断是特称否定判断 O:“有的 S 不是 P”。数学证明中,只要找到一个反例便足以推翻结论,因而“举反例”是数学证明的常用方法。瑞士数学家欧拉是最多产的数学家之一。他 28 岁瞎了右眼,59 岁瞎了左眼,却未影响他的创造力和洞察力。他提出了许多数学猜想,有的至今还未证明,也有的猜想被否定。下面就介绍一个被反例推翻的欧拉猜想。

【例 2-12】 欧拉研究过费马大定理,并且证明了当 $n=3$ 时式 2.8 成立。他虽未最终证明,却对费马大定理的正确性深信不疑。1769 年他将 $a^3+b^3=c^3$ 无整数解推广为

$$a^4+b^4+c^4=d^4 \tag{2.19}$$

$$a^5+b^5+c^5+d^5=e^5 \tag{2.20}$$

…………

$$\text{一般，}\quad a_1^n + a_2^n + \cdots + a_{n-1}^n = a_n^n\ (n \geqslant 3) \tag{2.21}$$

都没有整数解。这就是欧拉提出的许多猜想之一，称为“幂次求和猜想”。

200 多年后，人们借助计算机，举出了反例，否定了上述猜想。

美国人兰德(Leon J. Lander)和帕金(Thomas R. Parkin)于 1967 年找到式 2.20 的一组整数解：

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5$$

美国人艾尔基斯(Noam D. Elkies)于 1988 年找到式 2.19 的一组整数解：

$$2682440^4 + 15365639^4 + 18796760^4 = 20615673^4$$

举反例并不容易，前面提到的费马大定理，自从提出后，始终有人企图举反例推翻，均未成功，最后英籍美国数学家怀尔斯还是凭借代数几何的现代方法正面证明，经历了 358 年[0-10]。

【例 2-4】提到的哥德巴赫猜想是哥德巴赫在 1742 年提出的。欧拉接到哥德巴赫的信后，曾经验证了 100 以内的偶数均满足猜想，但未证明。至今 270 年了，仍未得证。人们企图用越来越大的偶数来验证，最好能找到反例而否定。表 2.1 是部分验证记录。

表 2.1

最大偶数的数量级	实验人与年代
1×10^4	Desboves (1885)
1×10^5	Pipping (1938)
1×10^8	Stein and Stein (1965ab)
2×10^{10}	Granville et al. (1989)
4×10^{11}	Sinisalo (1993)
1×10^{14}	Deshouillers et al. (1998)
4×10^{14}	Richstein (1999，2001)
2×10^{16}	Oliveira e Silva (Mar. 24，2003)
6×10^{16}	Oliveira e Silva (Oct. 3，2003)
2×10^{17}	Oliveira e Silva (Feb. 5，2005)
3×10^{17}	Oliveira e Silva (Dec. 30，2005)

尽管对大到 3×10^{17} 的偶数仍未找到反例，却不能作为哥德巴赫猜想的证明。20 世纪 20 年代，人们感到直接证明猜想太难，提出一个较弱的命题：

“一个大于等于 6 的偶数可以表示为两项之和，第一项由 a 个素数相乘，第二项由 b 个素数相乘。简记为 $(a+b)$ 问题。”表 2.2 就是世界各国数学家攀登高峰的历程。

表 2.2

年	$a+b$	数学家	国籍
1920	9+9	Brun	挪威
1924	7+7	Rademacher	德国
1932	6+6	Estermann	英国
1937	4+9	Ricci	意大利
1941	4+4	Byxwrao	苏联
1956	3+4	王元	中国
1957	3+3	Vinogradov	苏联
1957	2+3	王元	中国
1962	1+4	潘承洞	中国
1962	1+3	王元	中国
1966	1+2	陈景润	中国


陈景润 1933-1996

我国数学家陈景润(1933-1996)在 1966 年的成果(1+2)，离证明哥德巴赫猜想(1+1)仅一步之遥，45 年来竟无人超越。数学家在证明猜想的过程中提出种种新方法，即使未能证明猜想，往往也促进数学理论的发展。现在常有人缺乏基本的数学知识，也不掌握数学推理的逻辑方法，却贸然宣布自己证明了哥德巴赫猜想；还有人企图用哲学的一般原理“说明”哥德巴赫猜想成立；这都是不可取的。

2.2.4 数学归纳法

数学中的许多结论是先通过归纳而构成“猜想”，然后再予以严格证明。

【例 2-13】 下述猜想是否成立？当 k 是奇数时，n^k-n 能被 k 整除，这里 n 是正整数。

当 $k=3$，$n^3-n=n(n-1)(n+1)$。对任意正整数 n，如果是 3 的倍数，则 n^3-n 能被 n 整除；如果是 3 的倍数加 1，则 n^3-n 能被 $n-1$ 整除；如果是 3 的倍数减 1，则 n^3-n 能被 $n+1$ 整除。总之，n^3-n 都能被 3 整除；

当 $k=5$，$n^5-n=n(n^4-1)=n(n^2+1)(n-1)(n+1)$。对任意正整数 n，如果是 5 的倍数，一定以 0 或 5 为个位数，n^5-n 能被 n 除尽；如果以 1 或 6 结尾，n^5-n 一定能被 $n-1$ 除尽；如果以 4 或 9 结尾，n^5-n 一定能被 $n+1$ 除尽；如果以 2 或 8 结尾，则 n^2+1 以 5 结尾；如果以 3 或 7 结尾，则 n^2+1 以 0 结尾。总之，n^5-n 都能被 5 整除；

当 $k=7$ 时，可以类似证明，上述猜想也成立。

能否由此归纳出当 $k=$ 任意奇数时上述猜想都成立？答案是否定的。可以举出反例，例如取 $k=9$，$n=2$，$2^9-2=510$，不能被 9 整除。不过，如果把猜想中的“当 k 是奇数”换为“当 k 是素数”，结论就正确，这是另一个以费马命名的定理：“费马小定理”。费马小定理和许多数论中得到的结论在提出时并不知道有什么应用，现在却成为信息安全的数学基础[2-19]。

虽然一般的归纳法在数学上并不严格，“数学归纳法”在数学意义上却是和演绎推理同样严格的推理法。其方法如下：

设 $p(n)$ 是一个与自然数 n 有关的数学命题。如果能够证明：

(1) $p(1)$成立；　(2.22)

(2) 若 $p(k)$成立(这里 $k>1$)，可以推出 $p(k+1)$也成立；　(2.23)

那么 $p(n)$对所有自然数 n 都成立。

【例 2-14】 微积分中会遇到下面的等式：

$$1^2+2^2+3^2+\cdots+n^2=\frac{1}{6}n(n+1)(2n+1) \tag{2.24}$$

证明：（1）$n=1$，左式$=1^2=1$，右式$=\dfrac{1}{6}\times1\times2\times3=1$，上式成立；

（2）设当 $n=k>1$ 时 2.24 式成立，即 $1^2+2^2+3^2+\cdots+k^2=\dfrac{1}{6}k(k+1)(2k+1)$

当 $n=k+1$，左式$=1^2+2^2+3^2+\cdots+k^2+(k+1)$

$$
\begin{aligned}
&=\frac{1}{6}k(k+1)(2k+1)+(k+1)^2\\
&=\frac{1}{6}(k+1)(k(2k+1)+6(k+1))\\
&=\frac{1}{6}(k+1)(2k^2+7k+6)=\frac{1}{6}k(k+1)(k+3)(2k+3)\\
&=\frac{1}{6}(k+1)((k+1)+1)(2(k+1)+1)\\
&=\text{右。}
\end{aligned}
$$

所以式 2.24 对所有自然数 n 成立。

数学归纳法可以略加推广，式 2.23，2.24 改为

（1）$p(m)$成立；$(m>1)$ （2.25）

（2）若 $p(k)$成立 $(k>m)$，可以推出 $p(k+1)$也成立； （2.26）

那么 $p(n)$对所有大于或等于 m 的自然数 n 都成立。

【例 2-15】 证明任意 n 边的凸多边形的内角和

$$\sum^{n}\alpha_i=2(n-2)\pi;\quad n\geqslant3 \tag{2.27}$$

其中 α_1，α_2，…，α_n 分别是 n 个内角。

这里要说明两点：所谓凸多边形是指该多边形内任意两点相连的直线段都在多边形内。三角形、矩形、平行四边形、梯形、所有正多边形都是凸多边形；但锦旗形状的五边形、五角星（实际上是十边形）形就不是凸多边形。

2.27 式中的 $\sum$ 是求和符号，α_i 是通项，i 是哑指标，1 和 n 分别是下限和上限，即 i 的起点和终点。实际上，把 2.27 式展开，即 $\sum^{n}\alpha_i=\alpha_1+\alpha_2+\cdots+\alpha_n$；同样，2.24 式左端也可写为 $\sum^{n}i^2$。

下面用推广的数学归纳法证明式 2.27：

(1) 当 $n=3$，三角形内角和等于 $180°$，即 $\sum^{3}\alpha_i=2\pi$，2.27 式显然成立；

(2) 设 $n=k$，2.27 式成立，即 k 边凸多边形的内角和为 $\sum^{k}\alpha_i=2(k-2)\pi$，而 $k+1$ 边凸多边形，相当于一个 k 边凸多边形加一个三角形，其内角和

$$\sum^{k+1}\alpha_i=\sum^{k}\alpha_i+2\pi=2(k-2)\pi+2\pi=2(k-1)\pi=2((k+1)-2)\pi$$

式 2.27 成立。证毕。

以上四种证明方法可以称为“纸笔证明法”，延续了几千年。

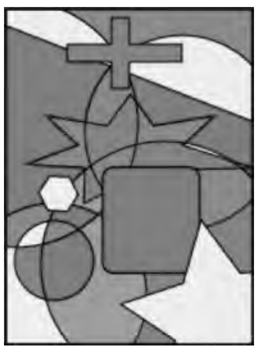
电子计算机诞生后，用人脑进行的大量而繁琐的数学计算逐渐为电脑所取代；人们也尝试用电脑完成复杂的数学证明。事实上，前面提到的对“幂次求和猜想”的否定和对哥德巴赫猜想的探索过程，已经借助电脑的高速运算功能。真正成功用计算机直接证明的例子是对“四色问题”的解决。而机器证明取得最新成果的则是我国数学家吴文俊院士。

2.3 数学定理的机器证明

2.3.1 四色猜想的机器证明

“四色猜想”(The Four Color Conjecture)也称“四色定理(The Four Color Theorem)”，和费马大定理及哥德巴赫猜想并称近代数学三大难题。1852 年英国人弗兰西斯·格斯里(Francis Guthrie, 1831-1899)来到一家科研单位搞地图着色工作时，发现了一种有趣的现象：“看来，每幅地图都可以用四种颜色着色，使得有共同边界的国家着上不同的颜色。”这个结论能不能从数学上加以

严格证明呢？这就是著名的“四色猜想”。他把这个发现告诉他的兄弟弗雷德里克·格思里（Fredrick Guthrie），弗雷德里克是伦敦学院大学（University College London）的学生，他向他的数学老师德·摩根（De Morgan，1806-1871）请教，德·摩根无法给出答复，他就向数学界同行通报。剑桥大学教授凯雷（Arthur Cayley，1821-1895）接到德·摩根的求助信后，于1878年6月13日向伦敦数学会（London Mathematical Society）提出这个问题，又在1879年皇家地理学会（Royal Geographical Society）发表论文《On the colouring of maps（论地图着色）》，这样，四色猜想逐渐为人所知。



四色绘地图

德·摩根 1806-1871

1879年6月17日凯雷的学生、英国律师坎普（Alfred B. Kempe，1849-1922）在著名的《Nature（自然）》杂志上宣称已经证明了四色猜想，同年他把论文在《American Journal of Mathematics（美国数学学报）》上发表。坎普的成果获得数学界的喝彩，他被选为皇家学会成员；他提出的“坎普链（Kempe chains）”方法受到普遍赞誉。然而好景不长，11年后，希沃德（Percy J. Heawood，1861-1955）指出坎普的证明有误，坎普自己也承认其中的错误。于是“四色定理”又回到“四色猜想”。希沃德花了将近60年研究四色猜想，进展不大。不过他带来了副产品，他同时研究了在其他曲面上地图着色问题，对所需颜色数量得到了以他命名的“希沃德估计（Heawood estimate）”。数学史表明，常常“有心栽花花不发，无心插柳柳成荫”。

进入20世纪后，人们把四色问题和“图论（graph theory）”结合起来。两位哈佛大学教授作出很大贡献。哈斯勒尔·惠特尼（Hassler Whitney，1907-1989）是位文理兼通的学者，曾做过社区乐团的团长。他把地图上的一个区域看

作图上的一个“顶点(vertex)”,两相邻区域的边界看作连接两个顶点的“边(edge)”,这些边都不交叉。四色问题就等价于一个图论问题:“无论图上的顶点有多少个,只需四种颜色表示顶点,使相邻顶点不同色。”美国数学家乔治·伯克霍夫(George Birkhoff, 1884-1944)曾任美国数学会会长,他提出把一个复杂图简化的方法。这样,大大加速了研究进程。人们开始了证明不超过 N 个区域可以用四色着色的征程(见表 2.3);对四色问题的研究中心也由英国转移到美国。

表 2.3

N	22	25	27	35	39	95
数学家	Birkhoff	Franclin	Reynolds	Winn	Ore 和 Stemple	Mayer
年	1920	1922	1926	1940	1970	1976

图的顶点(或地图区域)数目可以无限多,又如何验证?幸运的是,人们发现,通过化简,可以归结为有限个顶点。1969 年希区(Heesch)借助于坎普链,证明至多大约 8900 个顶点;1976 年阿佩尔(Kenneth I. Appel, 1932-)和哈肯(Wolfgang Haken, 1928-)减少到 1936 个顶点,以后进一步减少到 1476 个顶点。即使如此,用人脑验证仍然不可能。好在这时电子计算机已经有了长足的进步,他们在伊利诺大学的两台计算机上运算了 1200 个小时,作了 100 亿次判断,终于证明了四色猜想,轰动了世界[2-20]。人们认识到电脑不仅可以代替人脑进行数值计算,也可以代替人脑进行推理证明。阿佩尔和哈肯的工作如果由人脑完成,可能需要数百年,显然是不可能的。不过,这一证明至今并不被所有的数学家接受,因为它不能由人工直接验证,人们还必须对计算机编译的正确性以及运行这一程序的硬件设备充分信任。至今人们还是偏爱“纸笔证明法”,有人甚至认为:“一个好的数学证明应当像一首诗——而这纯粹是一本电话簿!”

2.3.2　机器证明与吴方法

机器证明的思想可以回溯到 17 世纪。当时已有机械构造的计算机问世,可以用机器代替繁琐而重复的人力计算。传统的欧氏几何证明定理的方法要求对个别的定理寻求个别的证法,而且每一证明总是要求某种新的、往往是奇巧的想法,笛卡儿和莱布尼兹都设想,能否制造某种计算机代替复杂的人脑证明,即用机器证明几何定理。20 世纪初德国数学家大卫·希尔伯特(David Hilbert,

1862-1943)明确地提出了公理系统的机械化判定问题,即把几何定理变为代数关系式之间的定理。这些关系式或者是多项式等式,或者是多项式不等式。他把几何定理的机器证明问题归结为下面三个主要步骤:

第一步,从几何的公理系统出发,引进数系统与坐标系统,使任意几何定理的证明问题成为纯代数问题,这是几何的代数化与坐标化;

第二步,将几何定理假设部分的代数关系式进行整理,然后依确定步骤验证定理终结部分的代数关系式是否可以从假设部分已整理成序的代数关系式中推出,这是几何的机械化;

第三步,依据第二步中的确定步骤编成程序,并在计算机上实施,以得出定理是否成立的最后结论。

如果一门几何可以找到这样三步(事实上只要前面两步即可)来完成定理的证明,就说这门几何可以机械化,并把可以机械化的这一结论称为机械化定理。

希尔伯特 1862-1943

代数关系式间的定理证明的机械化问题表达起来比较明确,但机器证明并不一定容易。首先,由于解决这些代数问题的计算量往往过大,使人望而却步;其次,因为代表几何关系而出现的那些代数关系式往往杂乱无章,使人无从着手。由于计算机的出现,把杂乱无章的代数关系式整理得井然有序,使计算机得以发挥其威力,便成为整个问题的关键。

吴文俊(1919-)是我国当代的著名数学家,1940年毕业于上海交通大学。1945年他进入中央研究院数学研究所,由数学家陈省身(1911-2004)加以指导,学习拓扑学。1947年,他根据中法互换留学生的协定,来到法国的斯特拉斯堡深造。他在嘉当(Elie Cartan,1865-1951)和爱勒斯曼(Charles Ehresmann,1905-1979)的指导下,获得了法国国家博士学位;1951年,他回到祖国。

吴文俊在拓扑学方面的成就集中在“复形在欧氏空间中的实现问题”上,这是拓扑学发展中的一个重要课题。他利用一种拓扑性质——示嵌类,提供了解决这类问题的一条途径;“吴类(Wu Class)”已是拓扑工作者的常识,写入教科书


嘉当　1869-1951


吴文俊　1919-


陈省身　1911-2004

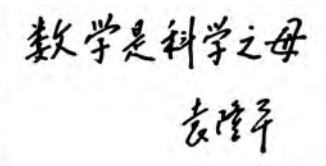
和词典。1956年，新中国第一次评选国家自然科学奖，吴文俊以“示性类和示嵌类”的研究荣获一等奖。那次全国只有三位科学家获一等奖，另两位是“中国火箭之父”钱学森和数学家华罗庚。吴文俊也很重视应用。他把拓扑学知识用于无线电工程的线路板设计问题，取得了良好的效果。

20世纪70年代以来，吴文俊研究计算机证明的问题。他继承并发展了中国古代的数学思想，在机器证明定理上取得重大突破。他认为，源于西方的公理化思想和源于中国的机械化思想，对于数学的发展都发挥了巨大作用，理应兼收并蓄。如今，计算机科学被认为是算法的科学。以算法为核心的机械化思想，既传统又前瞻，将为信息时代数学科学的创新发挥重大作用。

吴文俊研究的几何定理的机器证明旨在寻求一般性的方法，不仅适用于个别定理，而且适用于某一类型的定理，甚至可以说是某一种几何的所有定理。只要依照他所述的方法机械地进行，在有限步之后，就可对整个一类定理得到统一的证真或证伪，而无分难易。

1977年吴文俊创立了初等几何定理证明的机械化方法，国际上称“吴方法”，首次实现了高效的几何定理的机器证明。[2-21]、[2-22]“吴方法”也可用于几何定理的自动发现和未知关系的自动推导。吴文俊的开创性成果，打破了国际自动推理界在几何定理自动证明研究中长期徘徊不前的局面，也使我国在这一领域处于领先地位。由于吴先生的杰出贡献，1998年，他获得国际自动推理界的最高奖“Herbrand（赫伯兰）奖”。此奖每两年颁发一次，每届最多授予一人，迄今为止，获奖者均为自动推理界的领袖人物。2000年，吴文俊与“杂交水稻之父”袁隆平（1930-　）荣获国家最高科技奖。

袁隆平 1930-



中国工程院院士、2000年度国家最高科学技术奖获得者、杂交水稻育种专家袁隆平先生的题词

数学机械化方法的应用领域极其广阔,它可以为数学和其他领域的研究提供工具,为计算推理提供一种强有力的工具。在数学研究中的应用,可以把数学家从繁重的脑力劳动中解放出来,从而推动学科发展。近代数学史上第一次由中国人开创的这一新领域,吸引了各国数学家前来学习,原来手工计算上千项要几天功夫的证明,现在用计算机1秒钟就可以完成。在物理、机器人学、计算机视觉以及促进现代数学研究等重大高科技的前沿领域“吴方法”也实现了成功的应用。数学机械化研究的兴起,是中国当代数学发展中一个引人瞩目的具有中国传统特色的新里程碑。

吴文俊书画:“梦回故园”

吴文俊工作之余,寄情笔墨丹青,书画作品也有很高水平。

2.4 公理化体系

什么是公理化体系?我们先考虑“三段论”推理。“三段论”推理的前提必须正确,否则结论当然不真。【例2-15】用到平面几何中“三角形内角和等于180°”作为前提;为了证明“三角形内角和等于180°”,又用到“平行线与直线相交,内错角相等和同位角相等”作为前提;为了证明“内错角相等”和“同位角”相等,再要用到“过直线外一点只能引一条直线和已知直线平行”作为前提。那么,是否有最原始的命题无法从其他前提推出呢?合乎逻辑的回答是肯定的。数学上把原始的不需证明也无法证明的命题称为“公理(Axiom)”,由为数不多的“公理”和若干基本概念即“定义”出发,通过推理,证明出一系列结论即“定理”,就组成一

个“公理化体系”。构成体系的公理必须满足“独立性”(即不能从一条公理推出另一条公理)、“相容性”(即公理之间不能互相矛盾)和“完备性”(即这个体系中所有定理都能由这些公理推出)。

2.4.1 欧几里得几何的定义、公理与公设

欧几里得是古希腊最有影响的数学家之一。他写的《Elements(原本)》[0-7]是公理化体系的典范。全书共13卷,首先给出“点”、“线”、“面”、“平行直线”等基本概念的定义。例如:“点是没有部分的东西”、“线是没有宽度的长度”、“面是只有长度和宽度的东西”、“平行直线是这样一些直线,它们在同一平面内;无限延长后在两个方向都不会相交”,等等。

欧几里得 325—265 B.C.

接着欧几里得提出5条公理和5条公设(hypothesis)。他遵循亚里士多德的规定:“公理”是适合一切科学的真理,必须一望而知。他采用了5条公理:

公理1 和一个量相等的量彼此相等;

公理2 等量加等量,总量仍相等;

公理3 等量减等量,余量仍相等;

公理4 彼此重合的东西是相等的;

公理5 整体大于部分。

他提出的“公设”则只适用于几何:

公设1 从任一点到任一点作直线是可能的;

公设2 把有限直线不断延长是可能的;

公设3 以任一点为中心和任一距离为半径作圆是可能的;

公设4 所有直角彼此相等;

公设5 若一直线与两直线相交,且同侧所交两内角之和小于两直角。则两直线无限延长后必交于该侧的一点。

虽然《原本》的主要对象是几何学，但他提出的公理化方法却对数学的其他分支，甚至对人类的思维方法都有极大影响。据说《原本》从古到今各个版本的发行量仅次于《圣经》。

我国明代大学者徐光启(1562-1633)于1607年和意大利传教士利玛窦(Matteo Ricci，1552-1610)两人共同译完《原本》前六卷，即平面几何部分。

徐光启 1562-1633

利玛窦 1552-1610

徐光启为官廉洁，又以勤奋著称。《明史》说他“从西洋人利玛窦学天文、历算、火器，尽其术。遂遍习兵机、屯田、盐策、水利诸书”[2-23]。的确，他一生涉足的科学门类极其广泛，在农业、军事、天文、气象、水利、建筑、机械制造、测量、制图、医学、音乐和会计等领域，成就斐然，多有著作惠世。例如，他精通天文历法，完成了《崇祯历书》的编定。清朝入关，顺治把《崇祯历书》改名为《西洋新历》，颁布全国，施行至今。又如，他编写了著名的《农政全书》，是中国历史上最为完整的农业百科全书。[2-24]

徐光启在翻译《原本》过程中，克服了许多困难。他曾对利玛窦说：“一物不知，儒者之耻。”针对当时“几何对修身齐家治国平天下何用”的责难，他坚定地回答：“无用之用，众用所基”，他认为，欧几里得几何“似而晦，实则明；似而繁，实则简；似至难，实则易”，他的预言“窃百年之后，必人人习之”已成现实。徐光启首创的许多译名都很精彩。“几何”既是拉丁文对应词“geometria”的音译，又有汉语的“求索”之意；又如“点”、“线”、“直线”、“曲线”、“平行线”、“角”、“直角”、“锐角”、“钝角”、“三角形”、“四边形”等译名，不但在我国一直沿用至今，并且还影响了日本、朝鲜各国。[2-25]徐光启的故乡是上海的徐家汇，离上海交通大学一箭之遥，如今是上海科学文化的中心之一，也是典型的历史文化保护区。

2.4.2 公理化的影响

如前所述，由于欧几里得几何逻辑严密，公理化方法在数学的其他分支和自然科学甚至社会科学中都有很大影响。

例如对于算术，意大利数学家，也是符号逻辑的奠基人皮亚诺（Giuseppe Peano，1858-1932）提出5条公理[2-26]：

公理1　1属于非空数集 N；（注意：这里的“1”只是一个符号，也有人用“zero”表示）

公理2　N 的每个数 a 有后继数 a'；

公理3　1不是任何数的后继数；

公理4　若 $a' = b'$，则 $a = b$；

公理5　（归纳公理）若数集 M 是数 N 的子集合，即 $M \subset N$，若1属于 M，当 a 属于 M，必有 a' 属于 M，则 $M = N$。

皮亚诺　1858-1932

柯尔莫哥洛夫　1903-1987

前面介绍的数学归纳法就基于上述公理，2.22式与公理1和公理3有关，2.23式则与公理2与特别是公理5有关。

其他还可以举出不少例子，如对于集合论，德国数学家策梅洛（Ernst Zermelo，1871-1953）提出了7条公理；对于概率论，前苏联数学家柯尔莫哥洛夫（А. Н. Колмогоров，1903-1987）提出3条公理[2-27]，这里不一一介绍。

公理化方法对物理的理论研究，有着很大的影响。物理中的定律（Law）或基本假设相当于数学中的公理，是演绎推理的基础，本身却不必经过逻辑证明。

例如经典力学中的牛顿三定律；狭义相对论中的“相对性”和“光速不变性”；量子力学的冯·诺依曼公理等等[1-19]。

公理化方法对社会科学的研究也有深远影响。

形式逻辑本身就有四条公理，早在古希腊时期就由亚里士多德提出：

（1）同一律：在同一思维过程中概念和论题必须保持同一，不能转移或偷换。

（2）不矛盾律（也称为“矛盾律”）：任何思想不能既是真实的又是虚假的，也就是说，两个不相容的命题不可能同真，必有一假。

（3）排中律：两个互相否定的判断不能同假，必有一真。

（4）充足理由律：正确命题必须有充足理由。

荷兰哲学家斯宾诺莎（Baruch Spinoza，1632-1677）将公理化方法用于哲学研究，引起思想界的震动。斯宾诺莎与笛卡尔、莱布尼兹同时，在其名著《Ethic（伦理学）》[2-28]一书中几乎完全模仿欧几里得几何学方式来书写。比如第一部分是“论神”。先定义了8个“界说”的概念（详细定义此处从略）：① 自因；② 自类有限；③ 实体；④ 属性；⑤ 样式，分殊；⑥ 神；⑦ 自由，必然；⑧ 永恒；然后给出七条公理（我国哲学家贺麟先生译为“公则”）：


斯宾诺莎 1632-1677


贺麟 1902-1992

（1）一切事物不是在自身内，就必定是在他物内。

（2）一切事物，如果不能通过他物而被认识，就必定通过自身而被认识。

（3）如果有确定原因，则必定有结果相随，反之，如果无确定的原因，则决无结果相随。

（4）认识结果有赖于认识原因，并且也包含了认识原因。

（5）凡两物间无共同之点，则这物不能借那物而被理解，换言之，这物的概念不包含那物的概念。

（6）真观念必定符合它的对象。

（7）凡是可以设想为不存在的东西，则它的本质不包含存在。

然后提出 36 个“命题”，每一个“命题”都有“证明”。比如：“命题一　实体按其本性必先于它的分殊。证明：按界说三和界说五此理自明。”又如：“命题三　凡是彼此之间没有共同之点的事物。这物不能为那物的原因。证明：假如两物之间没有共同之点，则（公则五）这物不能借另一物而被理解，所以（公则四）这物不能为那物的原因。此证。”

斯宾诺莎把哲学中形而上的命题“证明”出来，甚至关于“上帝-神”的“存在”也必须证明，真是有着非凡的勇气。斯宾诺莎并不排斥上帝，但他把上帝和宇宙看成是一回事而不是有形的神。斯宾诺莎的上帝不仅仅包括了物质世界，还包括了精神世界。他认为人的智慧是上帝智慧的组成部分。上帝通过自然法则来主宰世界，所以物质世界中发生的每一件事都有其必然性。

斯宾诺莎的思想对后世有很大影响。有人问爱因斯坦是否信神，他这样回答：“我相信斯宾诺莎的神，而这位神显示在一切生物的和谐里。”他又说：“有一种超越一切的力量，支持着全宇宙的科学法则和自然界的运行变化。如果我们将这种力量称为上帝，那我就要向这位上帝低头。”[2-29]

《几何原本》中严密的逻辑思维也启发了中国的思想家。戊戌变法主将梁启超（1873-1929）就盛赞《几何原本》“字字精金美玉，是千古不朽的著作”。而戊戌变法领袖康有为（1858-1927）更是根据《几何原本》将公理体系推衍至社会政治领域，发表了自己关于未来的理想社会的蓝图。在他的早期著作《实理公法全书》和名动天下的著作《大同书》里得到了充分发挥。[2-30]

其中“实理”相当于公理。他提出了四条：

康有为 1858－1927

梁启超 1873－1929

(1) 人各合天地原质以为人。

(2) 人各具一魂,故有知识,所谓智也。然灵魂之性,各个不同。

(3) 人之始生,便具爱恶二质。及其长也,与人相接时,发其爱质,则必有益于人。发其恶质,则必有损于人。

(4) 人之始生,有信而无诈,诈由习染而有。

康有为在对人性及人的本质之判断的这四条“实理”的基础上概括出了人是天生自由而平等的,人性尊严应当得到制度之保护的自然权利思想。康有为还提出许多相当于“定理”的“公法”,是根据“实理”合乎逻辑地推导出来的关于确保“实理”得以贯彻的各项基本原则。即使在今天仍然难以想象,对近代中国的命运至关重要的戊戌变法,居然与欧几里得的几何学产生关联。

2.4.3 非欧几何与相对论

从欧几里得的《原本》问世之日起,就有人对“第五公设”即“平行公设”的独立性表示怀疑。无数人试图通过其他公理和公设来推导出这条公设,两千年来,总不成功。德国数学家高斯(Carl Friedrich Gauss, 1777-1855)领悟到第五公设应当是独立于其他公设的,或许将第五公设换成另一个与其矛盾的公设,也能推出一个几何体系。但是欧几里得几何长期占统治地位,慎重的高斯并未发表他的研究成果。19世纪另外两位并不知名的数学家,对欧几里得几何体系发出了公开挑战。一位是俄国的罗巴切夫斯基(Н. И. Лобачевский, 1792-1856),一位

是匈牙利的波约伊(János Bolyai，1802-1860)。他们用一个与第五公设矛盾的命题作为不证自明的公设，前四条公设相同，仍按形式逻辑规则，推出一整套几何命题。

高斯　1777-1855

罗巴切夫斯基　1792-1856

波约伊　1802-1860

欧几里得的平行公设有一个等价表述是：

“在平面上过直线外一点，只能作此直线的一条平行线”

罗巴切夫斯基的替代公设为：

“在平面上过直线外一点，至少可以作此直线的两条平行线”

他推出的系统称为“罗氏几何”，又称“双曲几何”；后来，德国数学家黎曼(Georg Friedrich Bernhard Riemann，1826-1866)又用与上面两条公设都矛盾的公设来替代：

“在平面上过直线外一点，不能作出此直线的平行线”

他推出的系统称为“黎曼几何”，又称“椭圆几何”。“罗氏几何”和“黎曼几何”统称“非欧几何”。

由于罗氏公设和黎曼公设似乎与人们在现实世界中的感受不同，非欧几何一度被认为是数学家的“游戏”，没有实际意义。但是黎曼坚信非欧几何可以在现实中找到应用背景。他用微分几何方法将三种几何统一处理，他认为，三种几


黎曼 1826-1866

何的区别，在于它们的高斯曲率不同，因而适用于不同空间。例如对于二维空间，欧式几何在平面上成立；罗氏几何对于一种“伪球面”成立；黎曼几何则对于球面成立。[2-31]

研究非欧几何，需要比较深的数学知识。为了帮助理解，让我们考察一个球面。球面上过两点的“直线”如何定义？和平面相似，定义为“过两点最短的连线”。球面上就是“过两点并以球心为圆心的大圆”。从上海坐飞机到旧金山，飞机并不是沿纬线自西向东飞，而是朝北绕北极再向南飞，这就是沿大圆路线，距离最短。由于所有大圆都相交，所以黎曼公设成立。

非欧几何的生命力突出地表现在为爱因斯坦的“相对论”提供了数学框架。正如牛顿力学对应于三维欧氏几何空间，狭义相对论对应于四维时空“双曲几何空间”，而广义相对论对应于四维时空的“椭圆几何空间”。[2-31]


孙中山 1866-1925

公理化体系对人们思想的影响是深远的。人们经常自觉或不自觉地应用某些“公理”指导或约束自己的思想和行为。例如“人之初，性本善”还是“人之初，性本恶”？其实就是两个对立的“公理”；不同的信仰也是不同的“公理”。似乎不必去争论是非，而是看由这样的公理推出的结论是否合理。中国几千年的封建社会都把“皇帝是天子”作为“公理”，从而推出“普天之下，莫非王土”、“君要臣死，臣不得不死”、“君主世代相传”等“判断”。但是它们阻碍了社会的发展，1911 年孙中山先生领导的辛亥革命推翻满清王朝，赶走了末代皇帝，人们用行动否定了那些封建“公理”！

2.4.4 公理化体系的局限

由于以欧几里得几何为发端的公理化方法逻辑严密，影响越来越大。针对欧几里得几何逻辑上的缺陷，希尔伯特进行了梳理提高，在其名著《Foundations of Geometry（几何学基础）》[2-32]中，提出从公理出发的纯粹演绎系统，形成近代

公理化思想体系。他认为,数学的每一个分支,都可以从一些简单的事实出发,用严格的逻辑推理的办法,推演出结论来;数学的任务就是逻辑推理。1900 年在法国巴黎召开的国际数学家会议上,法国数学家、物理学家庞加莱(Henri Poincare,1854-1912)宣布:"数学的严格性,看来直到今天才可以说是实现了。"[2-33]

庞加莱　1854-1912

罗素　1872-1970

但是,这种乌托邦式的梦想很快破灭了。1902 年英国著名哲学家、数学家罗素(Bertrand Arthur William Russell,1872-1970)提出了一个令人难以解释的"罗素悖论(Russell Paradox)":

设 z 为一切不含自身为元素所组成的新集合,那么 z 是包含在自身为元素的集合中呢?还是不包含于自身为元素的集合中呢?无论包含与否都会导致矛盾。

为使这一表述通俗化,罗素进而将它改编为"理发师悖论":

> 在塞尔维亚有一位理发师,他宣称:他只给所有不给自己刮胡子的人刮胡子,不给那些给自己刮胡子的人刮胡子。可是当他自己要刮胡子时,却陷入了尴尬境地。若他不给自己刮,根据他前面的条件,应该给自己刮;若他给自己刮胡子,又由于他后面的声明,他不该给自己刮胡子。总之,无论刮与不刮,都违背了自己的诺言,这位可怜的理发师陷入为难之中。[2-34]

类似的悖论还有"我在说谎"的命题。如果此命题正确,那么"我没有说谎",与原命题矛盾;如果原命题错误,那么也是"我没有说谎",但是正说明"我在说谎",原命题又正确了。

形式逻辑的"排中律"要求对一个命题非真即假,这些又真又假的悖论带来

了数学的危机。1931 年奥地利数理逻辑学家哥德尔(Kurt Godel，1906-1978)发表了两个后人以他命名的不完备性定理(Godel Incomplete Theorem)：

第一不完备性定理　设 S 是包含算术系统在内的任意形式系统，则存在命题 F 使得 F 和它的否命题 $\bar{F}$ 都在 S 中不可证。

第二不完备性定理　在上述形式系统 S 中不能证明它本身的协调性。

哥德尔　1906-1978

哥德尔不完备性定理表明：任何形式化数学公理规则体系都是“不完备”的，因为它总是存在着不能证明，也不能证伪的命题。[2-35]

从哲学上看，哥德尔不完备性定理再次表明科学没有“终极”，任何方法，包括公理化方法，都有其局限性。哥德尔的影响是深远的，最新的一个例子足以说明。2002 年夏天北京国际数学家大会，英国物理大师霍金(Stephen William Hawking，1942-　)的报告就是《哥德尔与物理的终结》。在当今国际物理研究领域，很多科学家提出有可能存在一个能描述一切物理现象的理论，并把这一理论称为超弦理论。霍金认为，建立一个单一的描述宇宙的大统一理论是不太可能的。霍金说他的这一推测正是基于数学领域的哥德尔不完全性定理。[2-36]

霍金　1942-

不仅公理化方法有局限性，整个形式逻辑方法也有局限性。

首先，逻辑不能提出数学命题。庞加莱曾发表专文，探讨数学中直觉和逻辑的关系。他深刻地指出：“逻辑和直觉各有其必要的作用。二者缺一不可。唯有逻辑能给我们以可靠性，它是证明的工具；而直觉却是发明的工具。”[2-37]其次，逻辑为数学及其他学科提供了正确思维的框架，但不能代替数学证明。

对于非数学专业人士，学习数学最大的用处是能够定量解决本领域的实际问题。为此目的，学会建立切合实际的数学模型，并能找到适当的数学方法求解，或许比单纯逻辑证明更重要。

2.5　思考与练习

1. 数学有哪些常规证明方法？你能举例说明吗？

2. 选择费马大定理或哥德巴赫猜想或四色问题，通过查阅参考资料或网站，了解从提出至今的发展历程。

3. 验证

$$\left(\frac{1+\sqrt{5}}{2}\right)^2=\frac{3+\sqrt{5}}{2};\ \left(\frac{1-\sqrt{5}}{2}\right)^2=\frac{3-\sqrt{5}}{2}$$

并用数学法证明斐波那契数列的通项是：

$$F_n=\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n-\frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n$$

【回忆：$F_1=F_2=1$；$F_{n+1}=F_n+F_{n-1}(n\geqslant 3)$】

4. 你觉得日常生活中是否有不证自明的公理？能否举例？

5. 有人说，“只有经过数学证明的结论才是真理”，你是否同意？为什么？

6. 试从网上寻找相关资料，了解阿罗(Arrow)如何从公理出发，证明了“不可能定理”(参阅绪言)

参考文献或网站

[2-1]　亚里士多德：《亚里士多德全集》(苗力田编译)，中国人民大学出版社(1990 - 1997).

[2-2]　《墨子直解》：浙江文艺出版社(2004).

[2-3]　庞朴：《公孙龙子研究》，中华书局(1979).

[2-4]　《荀子》：广州出版社(2004).

[2-5]　金岳霖等：《形式逻辑》，人民出版社(1979).

[2-6]　《辞源》(修订本)，商务印书馆(1979).

[2-7]　向隆万：《东京审判・中国检察官向哲浚》，上海交通大学出版社(2010).

[2-8]　中央电视台《探索 发现》栏目：《丧钟为谁而鸣》，安徽教育出版社(2004).

[2-9]　倪征燠：《淡泊从容莅海牙》，法律出版社(1999).

[2-10]　《远东国际军事法庭判决书(IMTFE Judgement)》(张一波，石恒利译)，群众出版社(1986).

[2-11] 向隆万：勾股定理是谁发明的？《中国文化之谜》(第一辑)，学林出版社(1985)

[2-12] 徐迟：《歌德巴赫猜想》，人民文学出版社(2005).

[2-13] 张景星等：《宋诗别裁集》，中华书局(1975).

[2-14] 彭漪涟：《古诗词中的逻辑》，北京大学出版社(2005)

[2-15] 柯南道尔(Arthur Conan Doyle)：《福尔摩斯探案全集(Sherlock Holmes：The Complete Novels and Stories)》(俞步凡译)，译林出版社(2005).

[2-16] 《白居易诗集校注》(谢思炜校注)，中华书局(2006).

[2-17] 梅汝璈：《远东国际军事法庭》，法律出版社，人民法院出版社(2005).

[2-18] 梁昌洪：《复变函数札记》，科学出版社(2011).

[2-19] 陈恭亮：《信息安全数学基础》，清华大学出版社(2004).

[2-20] 徐俊杰：《数学难题探索——费尔马大定理和四色问题证明》，西北工业大学出版社(2007).

[2-21] 吴文俊：《几何定理机器证明的基本原理(初等几何部分)》，科学出版社(1984).

[2-22] 吴文俊：《吴文俊论数学机械化》，山东教育出版社(1996).

[2-23] 查继佐：《明史·徐光启传》，中华书局(1974).

[2-24] 徐光启：《徐光启集》，上海古籍出版社(1984).

[2-25] 陈卫平，李春勇：《徐光启评传》，南京大学出版社(2006).

[2-26] http：//en. wikipedia. org/wiki/Giuseppe_Peqno.

[2-27] 应坚刚，何萍：《概率论》，复旦大学出版社(2005).

[2-28] 斯宾诺莎：《伦理学》(中译本)，商务印书馆(1998).

[2-29] 中国经济学教育科研网 www. cenet. org. cn.

[2-30] 《康有为全集》，中国人民大学出版社(2007).

[2-31] 费保俊：《相对论与非欧几何》，科学出版社(2005).

[2-32] 希尔伯特(David Hilbert)：《几何基础(The Foundations of Geometry)》(江泽涵，朱鼎勋译)，北京大学出版社(2009).

[2-33] http：//www-groups. dcs. st-and. ac. uk/～history/Mathematicians/Poincare. html.

[2-34] http：//www. jimloy. com/logic/russell. htm.

[2-35] 王浩：《哥德尔传(Reflections on Kurt Godel)》(中译本)，上海译文出版社(2002).

[2-36] 霍金(Stephen W. Hawking)2003 年演讲：《哥德尔和物理的终结(Godel and the End of Physics)》.

[2-37] 庞加莱(Jules Henri Poincaré)：《科学基础(The Foundation of Science)》(中译文)，《世界科学》(No. 4，1988).

第三章

无限与极限

3.1 无 限 集 合

3.1.1 古代的无限问题

能够思考，特别能进行抽象思维，是人类的特征。人们接触的是有限事物，却能遐想联翩，思考着“无限”与“无穷”。

中国春秋战国是“百家争鸣”的时期，思想特别活跃。墨子(468—376 B.C.)提出过不少有深刻思想的命题，其中就有：“莫不容尺，无穷也。”[3-1] 就是说，用尺永远量不尽的量叫做“无穷”。

庄子 369—286 B.C.

芝诺 490—430 B.C.?

庄子(约 369—286 B.C.)是道家代表人物，他的思想包含着朴素辩证法。《庄子》中多次提到与无穷有关的思想。例如：“至大无外，谓之大一”(意思是至大到没有边界，称为大一，也就是“无穷大”)；“吾生也有涯，而知也无涯”(意思是

我的生命有限，而知识是无限的）；他更借惠施之口，提出“一尺之棰，日取其半，万世不竭。”[3-2]一尺之棰是一有限的物体，但它却可以无限地分割下去。这就涉及了无限。用数学表示，每天的棰长也是一个无穷数列：

$$\left\{1,\ \frac{1}{2},\ \frac{1}{2^2},\ \cdots,\ \frac{1}{2^n},\ \cdots\right\} \tag{3.1}$$

古代希腊也是学术思想特别活跃的时期。诡辩派代表人物芝诺（Zeno，490-430 B.C.?）提出过许多似是而非的命题，成为“悖论”，其中多涉及无限。例如有一个著名的悖论是：

“阿基里斯永远追不上乌龟”。

阿基里斯是古希腊奥运会长跑冠军，怎么会追不上乌龟呢？岂非荒谬！但芝诺这样辩解：“当阿基里斯到达乌龟的起跑点时，乌龟已经走在前面一小段路了，阿基里斯又必须赶过这一小段路，而乌龟又向前走了。这样，阿基里斯可无限接近它，但不能追到它。”[3-3]

芝诺因其悖论而著名，并因此在数学和哲学两方面享有不朽的声誉。他最早以非数学的语言，把无限和有限，动和静，连续和离散的关系提了出来。芝诺在哲学上被亚里士多德誉为辩证法的发明人。黑格尔更称芝诺是“辩证法的创始人”[3-4]。

上述芝诺悖论当然可以用数学表述，设初始时刻，阿基里斯落后于乌龟的距离是 S 长度单位，两者速度分别是 V_A 和 V_T，且 $V_A \gg V_T$。①

记 $\lambda = \dfrac{V_T}{V_A} \ll 1$，则经过 $t_1 = \dfrac{S}{V_A}$ 后，阿基里斯前进 S 长度单位，即乌龟的出发点，这时，乌龟前进了 $V_T t_1 = S\lambda$，乌龟领先 $S\lambda$；再过 $t_2 = \dfrac{S\lambda}{V_A}$ 后，阿基里斯又前进了 $S\lambda$ 长度单位，乌龟也前进了 $V_T t_2 = S\lambda^2$，仍领先 $S\lambda^2$ 长度单位。也就是说阿基里斯与乌龟的距离是个无穷数列：

---

① “$\gg$（$\ll$）”的符号表示左边的量大大地大于（小于）右边的量。例如记地球质量为 M，一辆汽车的质量为 m，则 $M \gg m$；又如记地球半径为 R，地球与太阳的距离为 L，则 $R \ll L$。

$$\{S, S\lambda, S\lambda^2, \cdots, S\lambda^n, \cdots\} \tag{3.2}$$

阿基里斯追赶乌龟所花的时间则是个无穷级数：

$$t_1 + t_2 + \cdots + t_n + \cdots = \frac{S}{V_A}(1 + \lambda + \lambda^2 + \cdots + \lambda^n + \cdots) \tag{3.3}$$

当 n 趋向于无穷时，3.3 式会趋向于无穷吗？有了极限概念后，就可以回答这个问题(参见 3.4.1 节)。

3.1.2 无穷集合与有限集合的差别

第一章曾提到许多很大的数，但还是有限的；只要有足够的时间，还是可以从头到尾写出来。但是无穷数列却写不完，有无穷多项。最简单的无穷数列就是自然数列

$$N = \{1, 2, \cdots, n, \cdots\} \tag{3.4}$$

其元素中存在着比我们所能写出的无论多大的数还要大的数。上面提到的式 3.1 和式 3.2 也都是无穷数列。正因为阿基里斯在乌龟之后的距离构成一个无穷数列，“无穷”，就是永远数不完，所以芝诺说阿基里斯永远赶不上乌龟。我们还可以想出一些无穷集合，如“直线上的所有点数”，“平面中的所有点数”等等。

有限集合的元素数称为这个集合的“基数”，任意两个有限集合通过比较基数，就知道哪个集合点元素多。那么，一个无穷集合有基数吗？或者说，两个无穷集合可以比哪个集合的元素多吗？第二章中提到，康托尔对集合论有奠基性的贡献。1874 年他提出用“一一对应”的方法比较两个无穷集合的元素多少。也就是将两个无穷集合的元素一一配对，如果最后这两组都一个不剩，说明这两组无穷集合的数量相等；如果有一组还有元素没有配出去，说明这组无穷集合的元素更多一些。

“一一对应”其实是一个古老的方法。古时一个囚徒被关在不见天日的黑牢中。他只知每天狱卒要倒一次马桶，于是每当狱卒倒一次马桶，他就在墙上刻一道线痕。由于“倒马桶”既和“坐牢天数”一一对应，又和“刻线痕数”一一对应，因此，“坐牢天数”就和“刻线痕数”一一对应了。

但是，当实施此方法时，却会出现有限集合中难以想象的结果。比如，第二章中提到的欧几里得的公理 5，即“整体大于部分”不再永远正确。比如，“偶数

和自然数一样多”。偶数明明只有自然数的一半，是自然数的真子集，怎么会一样多呢？从“一一对应”的角度看，记偶数集合为

$$E = \{2, 4, \cdots, 2n, \cdots\} \tag{3.5}$$

就按照式 3.4 和式 3.5 元素的顺序对应，显然，E 的每一个元素 $2n$ 都有 N 中唯一元素 n 与之对应；反之，N 中任一元素 n 在 E 中也只有唯一元素 $2n$ 与之对应。如表 3.1 所示：

表 3.1

E	2	4	6	…	$2n$	…
N	1	2	3	…	n	…

再考虑所有有理数的集合

$$Q = \left\{\frac{n}{m}\right\}，其中\ n\ 和\ m\ 是无公因数的自然数，且\ m \neq 0。 \tag{3.6}$$

在每两个相邻的自然数之间，有无限个有理数，似乎 Q 的元素比 N 的元素多得多。然而，按照下面的一一对应原则，Q 的元素居然和 N 的元素一样多！

先写下分子分母之和为 2 的分数，只有一个，即 $\frac{1}{1} = 1$；再写下分子分母之和为 3 的分数，共有 2 个，即 $\frac{2}{1} = 2$ 和 $\frac{1}{2}$；然后是分子分母之和为 4 的分数，即 $\frac{3}{1}$，$\frac{2}{2} = 1$（前面已出现，剔除）和 $\frac{1}{3}$。这样做下去，对每一个和数 K，得 $K-1$ 项，分子从 $K-1$ 开始，逐项减小到 1，分母则从 1 开始，逐项增加到 $K-1$；分子分母有公因数时，约成既约分数；与前面重复的数则剔除。这样就得到一个无穷数列，包含了所有有理数。如果在这个数列旁边写上自然数列 N，就得到 Q 和 N 的一一对应关系。如表 3.2 所示：

表 3.2

Q	1	2	1/2	3	1/3	4	3/2	2/3	1/4	5	1/5	…
N	1	2	3	4	5	6	7	8	9	10	11	…

那么，是不是所有无穷大的数都相等呢？并非如此！

考察数轴上从 0 到 1 这段区间中的点数，用

$$R_{(0,1]} = \{x \mid 0 < x \leqslant 1\} \tag{3.7}$$

表示。中学数学告诉我们：数轴上每点一一对应着一个实数。而实数包含着有理数和无理数。按式 3.6 定义，$R_{(0,1]}$中的有理数总可以表示为无限循环小数，如：

$$1/2=0.5=0.50000\cdots,2/3=0.66666\cdots,\ 9/11=0.818181\cdots,\text{等等；}$$

$R_{(0,1]}$中的无理数则可表示为无限不循环小数，如：

$$1/\sqrt{2}=0.70710678\cdots,\ 1/\pi=0.31830988\cdots,\ \text{等等。}$$

康托尔用反证法巧妙证明 0 到 1 之间的点不可能和自然数列建立一一对应的关系。

假定 0 和 1 之间的所有实数可以和自然数建立一一对应，即可以建立类似表 3.1、表 3.2 那样的表，如表 3.3 所示。

表 3.3

$\boldsymbol{R_{(0,1]}}$	0.3287…	0.4956…	0.5700…	0.9043…	…
$\boldsymbol{N}$	1	2	3	4	…

我们构造一个 $R_{(0,1]}$中的实数 k：小数点后第一位数字与和 N 中的 1 对应的数字的小数点后第一位数字不同(按表 3.3，不能是 3)；小数点后第二位数字与上表第二个数字的小数点后第二位不同(按上表，不能是 9)；以此类推，这样构造的 k 一定在上述对应表之外。因为假定 k 在表 3.3 第一行中某一列中出现，例如是第 53809 个数字 m。但不可能，因为 k 和 m 小数点后第 53809 位数字是不同的。也就是说，$R_{(0,1]}$中的无穷大实数要比无穷大自然数更多。那么，$R_{(0,2]} = \{x \mid 0 < x \leqslant 2\}$ 中的实数是否比 $R_{(0,1]}$中的实数多？结论是一样多！正如偶数集合和自然数集合的关系类似，图 3.1 中长度为 1 的线段 OA 与长度为 2 的线段 OB 相交。连 AB，则过 OA 上任一点 P 作与 AB 平行的直线，与 OB 有唯一交点 Q，反之亦然。

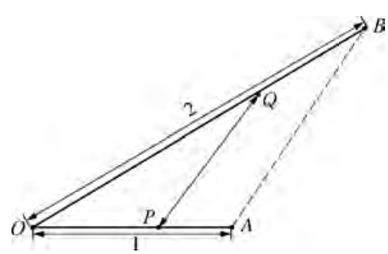


图 3.1

更不可思议的是，平面上的点与直线上的点一样多！图 3.2 上正方形

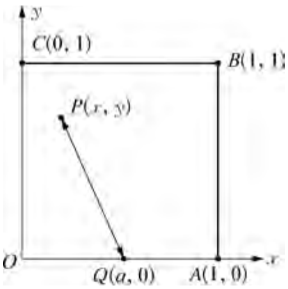


图 3.2

$OABC$中任一点 P 的坐标为(x, y)，则在线段 OA 上取这样的点Q，其对应的实数 a 这样选取：小数点后奇数位依次取 x 的小数，其小数点后偶数位依次取 y 的小数；反之亦然。例如：

$$P(0.347809\cdots, 0.984022\cdots) \longleftrightarrow Q(0.394874800292\cdots)$$

读者可以证明，立方体内所有点和线段上所有点也相同。

康托尔还证明了，所有曲线数量，或所有函数集合也是无穷集合，它的基数比实数集合的基数还要大。康托尔是犹太人后裔，他选用希伯来文的第一个字母$\aleph$(读为“阿莱夫”)表示无穷大的基数。他把全体整数序列相当于无穷大的第一级，其基数用符号$\aleph_0$(读为“阿莱夫零”)表示，这一级的无穷大数称为“可数无穷大”或“可列无穷大”(countable infinity)；有限集合和可数无穷集合合称为“离散集合(discrete set)”。他把所有和实数集合(即几何点)相同的无穷大作为第二级，其基数用$\aleph_1$(读为“阿莱夫一”)表示，称为“不可数无穷大(uncountable infinity)”；而“各种曲线的数目”是比几何点数目更大的无穷大，用$\aleph_2$(读为“阿莱夫二”)表示，当然更不可数。

这里留下了两个疑问。

第一，在$\aleph_0$和$\aleph_1$之间是否有其他基数？康托尔认为不存在。这就是著名的“连续统假设(the Continuum Hypothesis，简记为 CH)”。

第二，是否存在比$\aleph_2$更大的无穷大数$\aleph_3$？

关于“连续统假设”，曾引起数学家的争论。希尔伯特 1900 年在他提出的跨世纪 23 个数学未解问题中名列第一。20 世纪以来，许多数学家从事这项研究，取得重大进展。不过至今似乎还未彻底解决。

至于第二个问题，一百多年过去，仍踏步不前，即人们想象不出$\aleph_3$的存在。

康托尔的理论，特别是一一对应的方法造成的无穷中的悖论，与传统观念格格不入，难怪一开始他遭到那些坚持传统观念人士的强烈反对，甚至被人骂成

“疯子”。当时德国数学权威、他的老师克罗内克(Leopold Kronecker，1823－1891)的攻击尤为激烈。他说：“康托尔走进了超穷数的地狱。”他有一句名言：“上帝创造了正整数，其余的是人的工作。”就是说，人只能在正整数的有穷范围内研究，至于无穷的世界则完全超乎人的能力之外。甚至不承认康托尔为他的学生。在这种情况下，康托尔长期受到压抑和排挤，得不到柏林大学的教授职位。他郁郁不得志，一度精神崩溃，放弃数学研究，最后在一家精神病院去世。

然而，康托尔集合论的创立是人类思维发展史上的一座里程碑，它标志着人类经过几千年的努力，终于基本弄清了无穷的性质。因此越来越多的人开始承认它，并成功地把它应用到许多别的数学领域中去。关于无穷大的认识，美国科普作家盖莫夫(George Gamow)曾有通俗的阐述。[3-5]

3.2　无穷大与无穷小

3.2.1　无穷大

前面不止一次提到无穷大。什么是无穷大？无穷大是个很大的数吗？朴素的回答是：“无穷大不是一个很大的数，它是个变量，取值是要多大就多大。”下面以数列为例。

最简单的变量是数列 $\{a_n\}$，随着项数 n 的增大，a_n 的值起变化。

考察自然数列 3.4，随着项数 n 的增大，通项 $a_n=n$ 可以超过任意给定的大正数，正是“满园春色关不住，一枝红杏出墙来”。例如给出一个大正数 10000，那么从第 10001 项起的自然数都比 10000 大；即使给出的数是 1 谷歌 $=10^{100}$，那么从 $10^{100}+1$ 项起的自然数都比 10^{100} 大。自然数数列就是个无穷大数列。

再考察数列

$$\{1,\ -4,\ 9,\ -16,\ \cdots,\ (-1)^{n+1}n^2,\ \cdots\} \tag{3.8}$$

虽然各项正负相间，但绝对值越来越大，也是要多大就多大。仍给大正数 10000，要使

$$|a_n|=|(-1)^{n+1}n^2|=n^2>10000$$

只要 $n>100$，数列 3.8 的每一项都大于 10000；同理，从 $10^{50}+1$ 项起，数列 3.8

的每一项的绝对值都大于 1 谷歌。

德国数学家魏尔斯特拉斯(Karl Weierstrass，1813-1896)给出了一个逻辑严密的定义：

定义 3.1 (M-N 定义) 任意给定一个大正数 M，一定存在正整数 N，当项数 $n>N$，恒有 $|a_n|>M$，则称数列 $\{a_n\}$ 是无穷大数列。记为当 $n\to\infty$ 时，$a_n\to\infty$。

这里用符号 ∞ 表示“无穷大”，为英国数学家沃利斯(J. Wallis，1616-1703)首先使用。

中学学过函数概念，数列 a_n 可以看作定义在自然数上的函数 $f(n)$，上述定义也可表示为

定义 3.1′ (M-N 定义) 任意给定一个大正数 M，一定存在正整数 N，当项数 $n>N$，恒有 $|f(n)|>M$，则称函数 $f(n)$ 是无穷大量。记为当 $n\to\infty$ 时，$f(n)\to\infty$。

以 3.8 式的数列为例，若给出 $M=10$，则取 $N=3$(或更大的任一个数都可以)，当 $n>3$，就有 $|(-1)^{n+1}n^2|>10$，图 3.3 是其几何表述：

魏尔斯特拉斯 1813-1896

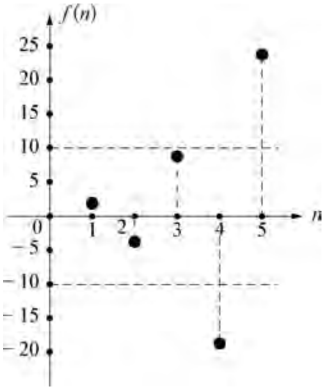


图 3.3

这里有三点要注意：

（1）无穷大量不仅越来越大，而且必须要多大有多大。例如数列 {10，11，11.1，11.11，…} 的通项是

$$a_n = \begin{cases} 10, & n = 1 \\ a_{n-1} + \dfrac{1}{10^{n-2}}, & n = 2, 3, \cdots \end{cases} \tag{3.9}$$

显然越来越大，但是不会大于 12，所以不是无穷大数列。

（2）无穷大量是指变量的绝对值，可正可负，如式 3.8 所示的数列；如果一个无穷大数列从某项之后保持为正数，我们也可称为“正无穷大”，用符号 $+\infty$ 表示；同理，如果一个无穷大数列从某项之后保持为负数，我们也可称为“负无穷大”，用符号 $-\infty$ 表示。例如 {3，2，1，0，−1，−2，…，$-n$，…} 就是个负无穷大数列，尽管前三项大于零和第四项等于零。

（3）无穷大量的绝对值必须越来越大。比如数列 {1，0，2，0，3，0，…，n，0，…} 的通项是

$$a_n = \begin{cases} \dfrac{n+1}{2}, & n = 2k-1 \\ 0, & n = 2k \end{cases} \quad k \in n \tag{3.10}$$

虽然其奇数项越来越大，要多大就多大，这是个“无界数列”；但其偶数项始终为 0，它就不是无穷大数列。

定义 3.1′可以推广到定义在实数域上的函数 $f(x)$。

定义 3.2 （M-X 定义）设 $f(x)$ 在实数域有定义。任意给定一个大正数 M，一定存在正数 X，当 $|x| > X$，恒有 $|f(x)| > M$，则称函数 $f(x)$ 是当 $x \to \infty$ 时的无穷大量，记为当 $x \to \infty$ 时，$f(x) \to \infty$。

其几何意义如图 3.4 所示。

例如 $f(x) = x$，$f(x) = x^2$，$f(x) = 2^x$，当 $x \to \infty$ 时，都有 $f(x) \to \infty$；但是对于正弦函数和余弦函数 $f(x) = \sin x$，$f(x) = \cos x$，当 $x \to \infty$ 时，函数值始终在 −1 到 1 之间摆动，就不是无穷大量。

对于定义在实数域中的函数，情况要比定义在整数域上的函数要复杂。除了当自变量趋于无穷大时，函数值可能趋于无穷大，也可能当自变量趋于某一个

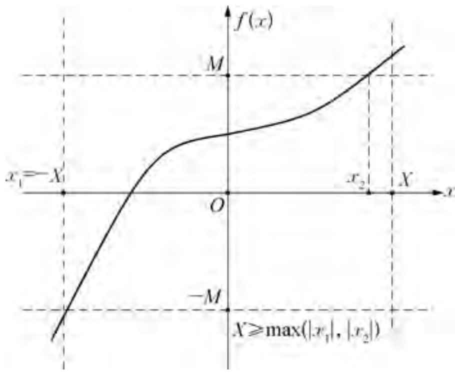


图 3.4

有限值时,函数值趋于无穷大。例如 $f(x)=\dfrac{1}{1-x}$,当 $x\to 1$ 时,$f(x)\to\infty$;又如正切函数 $f(x)=\tan x$,当 $x\to\dfrac{\pi}{2}$ 时,$f(x)\to\infty$。对于这种情况,魏尔斯特拉斯也给出了逻辑严密的定义:

定义 3.3 (M-δ 定义)设函数 $f(x)$ 在一点 x_0 的邻域有定义。任意给定一个大正数 M,一定存在正数 $\delta>0$,当 $0<|x-x_0|<\delta$,恒有 $|f(x)|>M$,则称函数 $f(x)$ 是当 $x\to x_0$ 时的无穷大量,记为当 $x\to x_0$ 时,$f(x)\to\infty$。

其几何意义如图 3.5 所示。

3.2.2 无穷小

与无穷大相反,无穷小是绝对值越来越小,且要多小有多小的变量。首先,还是以数列为研究对象。回过来考察数列 3.1 和 3.2。

数列 3.1 为:{ 1,0.5,0.25,0.125,0.0625,0.03125,0.015625,0.0078125,… }。

假设阿基里斯的速度是乌龟的 50 倍,则 $\lambda=0.02$,数列 3.2 为:

$$\{1,0.02,0.0004,0.000008,1.6\times10^{-7},3.2\times10^{-9},\cdots\}$$

可见这两个数列各项的数值越来越小,且越来越接近于零。

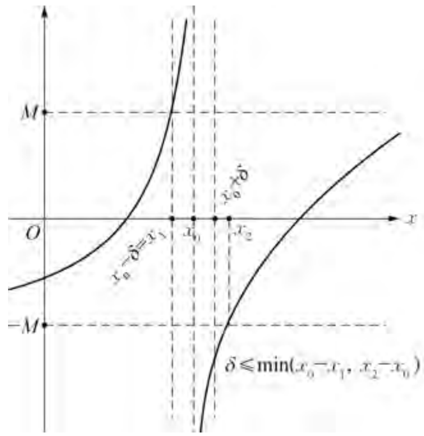


图 3.5

又如钟摆如果失去动力，它会逐渐减少摆幅，直至停下。如果记初始时刻 t_0 钟摆和垂直平衡位置的角度为 θ_0，θ_k 为 t_k 时刻的摆角，如果 $\theta_0>0$，则 $\theta_1<0$，且 $|\theta_1|<|\theta_0|$，于是我们得到一个符号相间，但绝对值逐项减小，越来越接近于零的数列：

$\{\theta_0, \theta_1, \cdots, \theta_n, \cdots\}$，且 $\theta_n\theta_{n+1}<0$，$|\theta_{n+1}|<|\theta_n|$，$\theta_n$ 越来越接近于零。

数学上怎样描述“越来越接近于零”？考察 3.1，其通项 $a_n=\dfrac{1}{2^{n-1}}$，

当 $n=1, 2, 3, \cdots, 10, \cdots, 100$，$a_n=1, 0.5, 0.25, \cdots, 0.001953, \cdots, 7.89\times10^{-31}$，越来越接近于零，而且要多小有多小。比如给出一个很小的正数 0.00001，要使

$$a_n=\frac{1}{2^{n-1}}<0.00001$$

因为 $a_{17}=0.000015258$，$a_{18}=0.0000076$，所以只要让 $n>17$，即从第 18 项起，3.1 的所有项都小于 0.00001。

用魏尔斯特拉斯的逻辑语言，无穷小数列的定义如下：

定义 3.4 （ε-N 定义）任给小正数 $\varepsilon>0$，存在自然数 N，当 $n>N$ 时，恒有 $|a_n|<\varepsilon$，则称 $\{a_n\}$ 是无穷小数列。

或

定义 3.4′ （ε-N 定义）任给小正数 $\varepsilon>0$，存在自然数 N，当 $n>N$ 时，恒有 $|f(n)|<\varepsilon$，则称 $\{f(n)\}$ 是当 $n\to\infty$ 时的无穷小量。

其几何意义如图 3.6 所示。

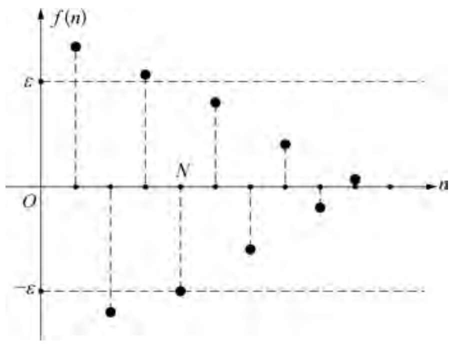


图 3.6

类似于无穷大量，也有三点注意：

(1) 无穷小量不仅越来越小，而且必须要多小有多小。例如数列 $\{10，9，8.9，8.89，\cdots\}$ 的通项是

$$a_n=\begin{cases}10, & n=1\\ a_{n-1}-\dfrac{1}{10^{n-2}}, & n=2,3,\cdots\end{cases}\tag{3.11}$$

显然越来越小，但是不会小于 8，所以不是无穷小数列。

(2) 无穷小量是指变量的绝对值，可正可负，不要把负无穷大和无穷小混淆。

(3) 无穷小量的绝对值必须越来越小。比如数列 $\{1，1/2，1，1/3，1，\cdots\}$ 的通项是

$$a_n=\begin{cases}1, & n=2k-1\\ \dfrac{1}{k+1}, & n=2k\end{cases}\quad k\in N\tag{3.12}$$

虽然其偶数项越来越小，要多小就多小；但其奇数项始终为 1，它就不是无穷小数列。

虽然无穷小是变量，但是有一个例外，常数 0 可以看作无穷小，这是和无穷大不同之处。

对于定义在实数域的函数 $f(x)$，当自变量趋于无穷大或自变量趋于有限值，都有可能是无穷小量。例如

设 $f(x) = 1/x^2$，当 $x \to \infty$ 时，$f(x) \to 0$，是无穷小；

设 $g(x) = x^2$，当 $x \to 0$ 时，$g(x) \to 0$，是无穷小；

注意，对于定义在实数域的函数 $f(x)$ 而言，是否无穷大或无穷小，必须根据自变量 x 的趋势而定。例如上述 $f(x) = 1/x^2$，当 $x \to \infty$ 时，$1/x^2 \to 0$，是无穷小；当 $x \to x_0 \neq 0$ 时，$1/x^2 \to 1/x_0$，是一个常数；当 $x \to 0$ 时，$1/x^2 \to \infty$，是无穷大。读者可以对 $g(x) = x^2$ 进行类似的判断。以下是魏尔斯特拉斯对两种趋势下无穷小量的逻辑定义：

定义 3.5　（ε-X 定义）设 $f(x)$ 在实数域有定义。任给小正数 $\varepsilon>0$，存在正数 X，当 $|x|>X$ 时，恒有 $|f(x)|<\varepsilon$，则称 $f(x)$ 是当 $x \to \infty$ 时的无穷小量。

定义 3.6　（ε-δ 定义）设 $f(x)$ 在一点 x_0 的邻域有定义。任给小正数 $\varepsilon>0$，一定存在正数 $\delta>0$，当 $0<|x-x_0|>\delta$ 时，恒有 $|f(x)|<\varepsilon$，则称 $f(x)$ 是当 $x \to x_0$ 时的无穷小量，记为 $f(x) \to 0$。

其几何意义由图 3.7 所示。

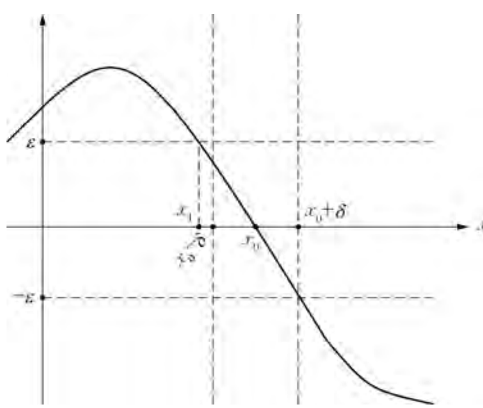


图 3.7

3.2.3 无穷大量与无穷小量的算术运算

无穷大变量的倒数形成的变量一定是无穷小量；无穷小量只要不为零，其倒数形成的变量一定是无穷大量。

例如：当 $n\to\infty$ 时，$A=\{n^2\}$ 是无穷大，则 $B=\{1/n^2\}$ 是无穷小；而当 $x\to 0$ 时 $f(x)=\sin x$ 是无穷小，则 $g(x)=1/\sin x$ 是无穷大。

无穷大和无穷小的算术运算如下，它们和有限数的运算不完全相同。

(1) 两个无穷小量之和是无穷小量；一个无穷小量与一个无穷大量之和却是无穷大量；但两个无穷大量之和不一定是无穷大量，必须根据具体对象来确定。例如：

设 $A=\{n\}$，$B=\{n^2\}$，$C=\{-n\}$，$D=\{1-n\}$，则

$A+B=\{n+n^2\}$ 仍是无穷大量；

$A+C=\{n+(-n)\}=\{0\}$ 是每一项都是 0 的常数数列，是无穷小量；

$A+D=\{n+(1-n)\}=\{1\}$ 是每一项都是 1 的常数数列，既不是无穷大量，也不是无穷小量。

(2) 两个无穷小量之积还是无穷小量；两个无穷大量之积是无穷大量；但一个无穷小量与一个无穷大量之积却无定论，也只能根据具体对象而定。例如：

设 $E=\{1/n^2\}$ 是无穷小数列，$A=\{n\}$，$B=\{n^2\}$，$C=\{n^3\}$ 都是无穷大数列，则

$EA=(1/n^2\cdot n)=\{1/n\}$ 是无穷小数列；

$EB=(1/n^2\cdot n^2)=\{1\}$ 是常数数列；

$EC=(1/n^2\cdot n^3)=\{n\}$ 是无穷大数列。

(3) 两个无穷大之商及两个无穷小之商都不确定，根据具体对象不同，可能是无穷大，也可能是无穷小，也可能两者都不是。

例如设 $f(x)=x$，$g(x)=x^2$，$h(x)=x+x^2$，显然当 $x\to\infty$ 时，三个函数都是无穷大量。这时，

$$\frac{f(x)}{g(x)}=\frac{1}{x}\to 0;\ \frac{g(x)}{f(x)}=x\to\infty;\ \frac{h(x)}{g(x)}=\frac{1}{x}+1\to 1$$

当 $x\to 0$ 时，上述三个函数都是无穷小，不难验证，这时

$$\frac{f(x)}{g(x)}=\frac{1}{x}\to\infty;\frac{g(x)}{f(x)}=x\to 0;\frac{h(x)}{g(x)}=1+x\to 1$$

上面提到的“无穷大＋无穷大”、“无穷小×无穷大”、“无穷大/无穷大”和“无穷小/无穷小”称为不定式。以后我们还会介绍其他一些不定式。

3.2.4 无穷大与无穷小的阶

在现实生活和科学技术中，无穷大和无穷小往往是相对的概念。比如地球平均半径约为 6371 千米；在机械工程中，度量尺度通常在数米之间，比如一辆汽车，尺度不过 2-3 米，一台大型机床，尺度也在 10-30 米左右。在运算中如遇到地球半径 6371 千米，后者可以看作是无穷大；但是在天体运动的研究中，一个“天文单位”等于地球到太阳的平均距离，约为 1.5 亿千米，大约等于 2.34 万个地球半径；而一“光年”指光在一年时间中行走的距离，即约九万四千六百亿千米，大约等于 14 亿 8 千 5 百万个地球半径。这时地球半径可以看作是零，也就是无穷小了。对于汽车尺度而言，1 光年和地球半径都看作“无穷大”，但是 1 光年是地球半径的“高阶无穷大”；同理，汽车尺度和地球半径对光年而言，都看作“无穷小”，而汽车尺度是地球半径的“高阶无穷小”。

数学上对无穷大和无穷小的“阶”有精确的定义。

定义 3.7 如果 $\alpha(x)\to\infty,\beta(x)\to\infty$

$$\text{且 }\frac{\alpha(x)}{\beta(x)}\to\begin{cases}\infty, & \text{高阶无穷大}\\ \text{常数 } C\text{，则称 }\alpha(x)\text{ 是 }\beta(x)\text{ 的} & \text{同阶无穷大}\\ 0, & \text{低阶无穷大}\end{cases}\tag{3.13}$$

定义 3.8 如果 $\alpha(x)\to 0,\beta(x)\to 0$

$$\text{且 }\frac{\alpha(x)}{\beta(x)}\to\begin{cases}\infty, & \text{低阶无穷小}\\ \text{常数 } C\text{，则称 }\alpha(x)\text{ 是 }\beta(x)\text{ 的} & \text{同阶无穷小}\\ 0, & \text{高阶无穷小}\end{cases}\tag{3.14}$$

特别，当 $C=1$，称 $\alpha(x)$ 是 $\beta(x)$ 的等价无穷小。

【例 3-1】 当 $x\to\infty$，x^3 是 x^2 的高阶无穷大，$2x$ 是 x^2 的低阶无穷大，$3x^2+1$ 是 x^2 的同阶无穷大；$1/x^3$ 是 $1/x^2$ 的高阶无穷小，$1/x$ 是 $1/x^2$ 的低阶无穷小，$1/(3x^2)$ 是 $1/x^2$ 的同阶无穷小，而 $3/(3x^2+1)$ 是 $1/x^2$ 的等价无穷小。（为什么?）

【例 3-2】 当 $x\to 0$，$1/x^3$ 是 $1/x^2$ 的高阶无穷大，$1/x$ 是 $1/x^2$ 的低阶无穷大，$1/(3x^2)$ 是 $1/x^2$ 的同阶无穷大；x^3 是 x^2 的高阶无穷小，$2x$ 是 x^2 的低阶无穷小，$3x^2+1$ 是 x^2 的同阶无穷小，而 x^2+1 是 x^2 的等价无穷小。（为什么?）

3.3 “无穷”的应用——求圆周率

古人对无穷的探索并非仅仅是“浮想联翩”，也用于解决实际问题。求圆面积及计算圆周率就是一个重要应用，这个过程长达数千年，至今仍在继续。

3.3.1 刘徽的“割圆术”

古代祖先在有文字之前已经认识了圆，并应用了圆。从中外考古发现中，都有器皿、车轮、装饰品等圆形物品。例如近年在四川成都发现的金沙遗址，精美的太阳鸟金箔令人叹为观止。三千多年前蜀人已经能精确将圆八等分和十二等分，如图 3.8 所示。[3-6]

图 3.8

在圆形物品的设计制造中，人们当然会遇到这样的问题：比如对于给定直径长度（例如 1 尺）的轮子，要用多少木头？即圆的面积怎么算？如果轮子周边用铁皮包箍，铁皮长度是多少？即圆的周长是多少？或者说，圆周长与直径长之比，即“圆周率”是多少？第二章曾提到过刘徽的“割圆术”，就是从 6 边形出发，用有限边长的内接多边形近似代替圆，然后多边形边数逐次加倍，无限逼近圆。《九章算术》这样描述：“割之弥细，所失弥少。割之又割，则与圆合体，而无所失矣。”[3-7] 下面作稍微详细的阐述。

由图 3.9 所示，设半径 $OA=r$。为讨论方便，根据对称性，我们讨论六分之一的圆，即扇形 OAB。

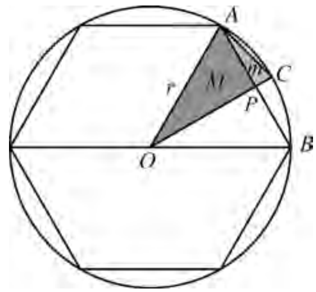


图 3.9

正三角形 OAB 的边长 $S_6 = r$，$\triangle OAB$ 的面积$= \dfrac{3\sqrt{3}}{12}r^2$；于是得到圆面积的第一个近似值 $A_6 = 6 \times \dfrac{3\sqrt{3}}{12}r^2 = 2.598076r^2$。这时，圆周长≈$6r$，圆周直径之比≈3，这正是刘徽之前人们认识到的“周三径一”。一般情况下计算 n 边形和 $2n$ 边形的程序如下：

仍用图 3.9。设 $AB = S_n$ 为 n 边形的边长，$AC = S_{2n}$ 为 $2n$ 边形的边长。

记 OP 为 H_n，由勾股定理，有

$$H_n = \sqrt{r^2 - \left(\frac{S_n}{2}\right)^2} \tag{3.15}$$

n 边形的面积为

$$A_n = \frac{nH_n S_n}{2} \tag{3.16}$$

n 边形的周长为

$$C_n = nS_n \tag{3.17}$$

圆周率的近似值

$$\pi \approx \frac{C_n}{2r} \tag{3.18}$$

因 $AP = S_n/2$，记 $PC = P_n$，则

$$P_n = r - H_n \tag{3.19}$$

再由勾股定理，$AC^2 = AP^2 + PC^2$，得

$$S_{2n} = \sqrt{\left(\frac{S_n}{2}\right)^2 + P_n^2} \tag{3.20}$$

以此类推，逐次边长加倍，直到

$$|A_{2n} - A_n| < \varepsilon\text{（给定的误差界）} \tag{3.21}$$

就得到所期望精度的圆面积、圆周长和圆周率的近似值。

刘徽觉察到，圆内接多边形面积总是小于圆面积，他进而考虑“双向夹逼”，给出圆面积或圆周率近似值误差的上界和下界。图 3.10 以内接 6 边形为例，AB 是与 CD 平行的切线。记 A_6 为内接 6 边形面积，B_6 为 A_6 加上 6 个和 $ABCD$ 面积相等的长方形面积，显然，圆面积夹在 A_6 和 B_6 之间。

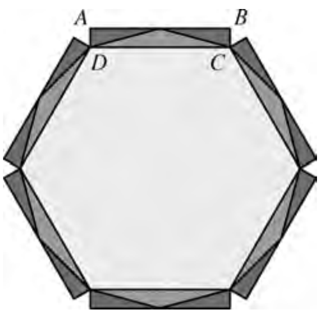


图 3.10

一般情况，在图 3.9 中假想一个长方形，以 $AB=S_n$ 为长，以 $PC=P_n$ 为宽，这 n 个长方形的面积和为

$$D_n = nS_n P_n \tag{3.22}$$

记

$$B_n = A_n + D_n \tag{3.23}$$

显然

$$A_n < A < B_n \tag{3.24}$$

其中 A 为圆面积。

不失一般性，取 $r=1$，表 3.4 为按上述程序前 5 组的计算结果。

表 3.4

边数 n	边长 S_n	内接多边形周长 C_n	内接多边形面积 A_n	内接多边形加 n 个小矩形面积 B_n	圆周率的近似值	$\Delta_n = A_n - A_{\frac{n}{2}}$
6	1	6	2.598076	3.401924	3	
12	0.517638	6.211655	3	3.408888	3.105828	0.401924
24	0.261052	6.265256	3.105824	3.211657	3.132628	0.105824
48	0.130806	6.278691	3.132624	3.235390	3.139345	0.026800
96	0.065263	6.282051	3.139342	3.190798	3.141026	0.006718
192	0.032723	6.282816	3.141083	3.166811	3.141408	0.001741

刘徽方法仅用到勾股定理，计算也只用到加减乘除和开方。在当时只能靠算筹的条件下，刘徽算到了 96 边形，圆周率的精度达到小数后 2 位，后人就把 3.14 称为“徽率”。

刘徽还进一步发明了“外推捷算法”，可以大大减少计算量。记

$$\Delta_n = A_n - A_{\frac{n}{2}} \tag{3.25}$$

刘徽从几何图形的观察和数值计算的结果中发现，

$$\Delta_n \approx \frac{1}{4}\Delta_{\frac{n}{2}} \tag{3.26}$$

（读者不妨从表 3.4 中 Δ_n 那一列数字予以验证，实际上，当多边形边长与半径之比 S_n/r 越小，式 3.26 越精确），那么，

$$\Delta_n \approx \frac{1}{4}\Delta_{\frac{n}{2}} = \left(\frac{1}{4}\right)^2 \Delta_{\frac{n}{4}} \tag{3.27}$$

这样，

$$A_n = A_{\frac{n}{2}} + \Delta_n \approx A_{\frac{n}{2}} + \frac{1}{4}\Delta_{\frac{n}{2}} \tag{3.28}$$

用上面的公式，可以算得

$$A_{192} \approx A_{96} + \frac{1}{4}\Delta_{96} = 3.139342 + \frac{0.006718}{4} = 3.141021$$

与表 3.4 中的计算结果相比，少了两次开方运算，数值却非常接近，绝对误差为

$$|3.141083 - 3.141021| = 0.000062，相对误差仅为 \frac{0.000062}{3.141083} \approx 0.002\%$$

不仅如此，式 3.28 还可以继续外推，

$$\begin{aligned}
A_{2n} &= A_n + \Delta_{2n} \approx A_n + \frac{1}{4}\Delta_n \\
A_{4n} &= A_{2n} + \Delta_{4n} \approx \left(A_n + \frac{1}{4}\Delta_n\right) + \frac{1}{4}\Delta_{2n} = A_n + \left(\frac{1}{4} + \frac{1}{4^2}\right)\Delta_n \\
A_{8n} &= A_{4n} + \Delta_{8n} \approx \left(A_n + \left(\frac{1}{4} + \frac{1}{4^2}\right)\Delta_n\right) + \left(\frac{1}{4}\right)^3 \Delta_n \\
&= A_n + \left(\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3}\right)\Delta_n \\
&\cdots\cdots \\
A_{2^k n} &\approx A_n + \left(\frac{1}{4} + \frac{1}{4^2} + \cdots + \frac{1}{4^k}\right)\Delta_n = A_n + \frac{1}{3}\left(1 - \frac{1}{4^k}\right)\Delta_n
\end{aligned} \tag{3.29}$$

最后一式的推导中用到了等比级数的求和公式：

$$a+aq+aq^2+\cdots+aq^k=\frac{a(1-q^{k+1})}{1-q}\quad\left(\text{其中 } a=q=\frac{1}{4}\right)$$

利用上式，我们不妨以96边形为基础，计算单位圆内接3072边形面积。因为 $3072=2^5\times 96$，

$$\begin{aligned}A_{3072}&\approx A_{96}+\left(\frac{1}{4}+\frac{1}{4^2}+\frac{1}{4^3}+\frac{1}{4^4}+\frac{1}{4^5}\right)\Delta_5\\&=3.139342+\frac{1}{3}\left(1-\frac{1}{4^5}\right)\Delta_{96}\\&=3.141579\approx 3.1416\end{aligned}$$

已经得到小数点后4位小数的精确度。

正如吴文俊院士所说，刘徽的割圆术是中国古代数学中“一个十分精彩的算法”[3-8]。

在刘徽之前，阿基米德也考虑过“双向夹逼”，但他用的是内接多边形和外切多边形，计算要复杂得多。他算到96边形，得到的估计不等式是3.140845＜圆周率＜3.142857，不如刘徽的估计精确。

祖冲之正是按刘徽的方法，一直算到12288边形，得到圆周率≈355/113≈3.1415929，称为“祖率”，并按刘徽外推捷算法，由不等式3.24，进而得到估计式3.14159261864＜圆周率＜3.141592706934，成为此后千年世界上最准确的圆周率。

3.3.2 人类对计算圆周率的不懈追求

为了找出π更多的准确小数位数，人类继续进行探索。

1424年波斯数学家贾姆希德(Jamshid al-Kashi，1380–1429)算出π的16位准确小数；几乎同时，印度数学家马德哈瓦(Madhawa of Sangamagrama，1350–1425)发现π可以看作无穷项相加(也就是“无穷级数”)。两百年后，苏格兰数学家格利高里(James Gregory，1638–1675)和莱布尼兹也发现了同样的公式，欧洲人曾称此公式为格利高里-莱布尼兹公式，现在被称为马德哈瓦-格利高里-莱布尼兹公式：

$$\pi = 4\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots + (-1)^{n+1}\frac{1}{2n-1} + \cdots\right) \tag{3.30}$$

这个算法简便易行，但是效率太低，前 10 个近似值是：4，2.6666，3.4666，2.8952，3.3397，2.9760，3.2837，3.0171，3.2524，3.0418，算到 4000 项才达到阿基米德的精度。

不过，人们发现，这十个近似值在 π 的上下震荡，如果取前后两个值的平均，应当更好。我们得到 9 个 1 级平均值：3.3333，3.0666，3.1809，3.1174，3.1578，3.1298，3.1504，3.1347，3.1471，比最初得到的近似值要准确。这些值仍然在 π 的上下震荡，再取 2 级平均，得到 8 个近似值：3.1999，3.1275，3.1491，3.1376，3.1438，3.1401，3.1425，3.1409，已经准确到两位小数。读者可以自行计算，3 级平均直到 9 级平均的各近似值，最终可以得到 π 的近似值为 3.1416。

1655 年英国数学家沃里斯（John Wallis，1616-1703）发现可以将 π 看作无穷项相乘（即“无穷乘积”）：

$$\pi = 2\left(\frac{2}{1} \times \frac{2}{3} \times \frac{4}{3} \times \frac{4}{5} \times \frac{6}{5} \times \frac{6}{7} \times \frac{8}{7} \times \cdots\right) \tag{3.31}$$

这个算法只用乘除，更加方便，但是计算效率也不高，前 10 个近似值是：4，2.6667，3.5556，2.8444，3.4133，2.9257，3.3437，2.9721，3.3024，3.0022，精度比 3.30 式更低一些。当然也可以逐级平均，得到更好的近似值，读者不妨自行练习。

随着微积分的发展，人们对无穷级数认识的深入，效率更高的算法不断涌现。

1706 年英国天文学家梅钦（John Machin，1686-1751）提出如下算式：

$$\pi = 4\left(4\tan^{-1}\frac{1}{5} - \tan^{-1}\frac{1}{239}\right) \tag{3.32}$$

其中 $$\tan^{-1}x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots + (-1)^{n+1}\frac{x^{2n-1}}{2n-1} + \cdots$$

他得到了 π 的 100 位准确小数；1789 年斯洛文尼亚数学家、物理学家维加（Jurij Bartolomej Vega，1754-1802）用 3.31 得出 π 的小数点后 140 位数字，虽然只有 137 位正确，但是作为世界纪录，维持了 50 年。

在这漫长的过程中，人们始终有个疑问，即 π 会不会是有限位小数？会不会是无限循环小数？1761 年法国数学家兰伯特(Johann Heinrich Lambert，1728-1777)证明了圆周率是无理数；1882 年，德国数学家林德曼(Ferdinand von Lindemann，1852-1939)进而证明了圆周率是超越数。当一个数可以被写成含有理系数的多项式方程的根的形式时，不管这个数是实数还是复数，则这个数都可以被定义为代数数。否则，就是超越数。比如无理数 $\sqrt{2}$ 是方程 $x^2-2=0$ 的一个根，就是代数数。π 的超越性证明彻底地解决了古希腊三大作图问题中的化圆为方问题，即化圆为方是不可能的。同时，π 是无限不循环小数的结论也鼓舞了人们继续求出更多小数位的热情。

1914 年，印度天才数学家拉马努金(Srinivasa Ramanujan，1929-1993)在他的论文里发表了一系列共 14 条圆周率的计算公式。其中之一是

$$\pi=2+\frac{1}{3}\left(2+\frac{2}{5}\left(2+\frac{3}{7}\left(2+\frac{4}{9}\left(2+\frac{5}{11}(2+\cdots)\right)\right)\right)\right) \tag{3.33}$$

前 10 个近似值是 2，2.6667，2.9333，3.0476，3.0730，3.1321，3.1371，3.1395，3.1405，3.1411，显然精确度比 3.30 和 3.31 式高。

拉马努金 1929-1993

拉马努金还有一个公式几乎是匪夷所思：

$$\pi=\frac{9801}{2\sqrt{2}\sum\limits_{n=0}^{\infty}\frac{(4n)!}{4^{4n}(n!)^4}\frac{1103+26390n}{99^{4n}}} \tag{3.34}$$

其中 $n!=n(n-1)(n-2)\cdots3\cdot2\cdot1$；$0!=1$。仅取第 1 项，就得到

$$\pi\approx\frac{9801}{2\sqrt{2}\frac{0!}{4^0(0!)^4}\frac{1103}{99^0}}=\frac{9801}{2\sqrt{2}\times1\times1103}=3.14159273$$

已经得到 6 位准确小数。而且每计算一项，可以得到 8 位准确小数。

电脑的问世，大大加速了求圆周率更多位数的进程。匈裔美籍数学家冯·诺依曼(Von Neumann，1903-1957)是 20 世纪少见的数学科学通才，在许多领域都有重要的基本贡献。包括数理逻辑、测度论、实分析、算子代数、对策论等领域有开拓性贡献。他对量子力学、流体力学也有很大贡献。由于流体力学中大

量非线性偏微分方程只能用数值计算方法，促使他成为今日电脑的奠基者。

冯·诺依曼和世界第一台电子计算机　图 3.11

1946 年，世界第一台电子计算机 ENIAC 制造成功，标志着人类历史迈入了电脑时代。电脑的出现导致了计算方面的根本革命。1949 年，冯·诺依曼和他的团队在 ENIAC 上根据梅钦公式 3.32 计算到 π 的 2037 位小数，包括准备和整理时间在内仅用了 70 小时。五年后，NORC(海军兵器研究计算机)只用了 13 分钟，就算出 π 的 3089 个小数位。

在 20 世纪 60 年代至 70 年代，随着美、英、法各国的电脑科学家不断地进行电脑上的竞争，π 的值也越来越精确。1973 年法国人古伊洛德(Jean Guilloud)和博伊尔(Martin Bouyer)在 CDC 7600 机器上花了 23.3 个小时发现了 π 的第一百万个小数位。1985 年美国数学家高斯帕(Ralph Willam Gosper，1943-　)用公式 3.33 计算到了 π 小数点后的 17500000 位。

1989 年，乌克兰裔美国数学家大卫·丘德诺夫斯基和格雷高里·丘德诺夫斯基兄弟(David Chudnovsky，1947-　，Gregory Chudnovsky，1952-　)将 3.33 做了改进，每算一项，可得 15 位准确小数。1994 年两兄弟利用改进公式计算到了 4044000000 位。

2002 年日本科学家金田康正的团队在一台日立超级计算机上，以每秒 200 亿次运算速度，耗时 600 小时，按照改进的梅钦公式算出 π 的 1241100000000 位小数，创造了新的世界纪录；2009 年 8 月 17 日，日本筑波大学计算科学研究中心宣布，该中心使用超级计算机计算出圆周率小数点后 25769.8037 亿位数，创下新的世界纪录。该数据由研究中心副教授高桥大介计算出。此次使用的是美国制造、连接 640 台电脑、最快计算速度可达每秒 95 万亿次的超级计算机“T2K 筑波系统”，于 2009 年 4 月分两次进行主要计算和验证计算，耗时 73 小时 36 分。随着电脑的普及，一些普通人也加入了计算圆周率的竞赛。据中国之声《全球华语广播网》报道，2010 年 10 月，日本长野县的一位公司职员、56 岁的近藤茂在家中普通电脑中加装了 48TB 硬盘驱动器，开始计算圆周率，经过整整一年，计算到小数点后 10 万亿位，创造了新的世界纪录！

从应用角度上看,算出亿万位小数似乎并没有什么实际意义。那么为什么数学家们还像登山运动员那样,奋力向上攀登,继续对 π 的探索呢?除了科学探索的好奇心外,首先,可以测试或检验超级计算机的各项性能,特别是运算速度与计算过程的稳定性。几年前,当 Intel 公司推出奔腾(Pentium)时,正是通过运行 π 的计算而发现了一个问题。其次,计算的方法和思路可以引发新的概念和思想。虽然计算机的计算速度超出任何人的想象,但毕竟还需要由数学家去编制程序,指导计算机正确运算。好的方法加上好的计算工具,才能更好解决问题。

关于圆周率的历史和最新进展可以参阅有关网站[3-9][3-10][3-11]。

3.4 变量的极限

3.4.1 数列的极限

从刘徽的割圆术,对单位圆从内接正 6 边形开始,我们得到两个越来越逼近于圆周率的数列$\{A_n\}$和$\{B_n\}$,满足不等式:

$$\begin{aligned} &A_6 < A_{12} < A_{24} < \cdots < A_{6\times 2^k} < \cdots < \pi < \cdots \\ &< B_{6\times 2^k} < \cdots < B_{24} < B_{12} < B_6 \end{aligned} \tag{3.35}$$

其中 $k=0, 1, \cdots$。当 $k \to \infty$,即 $n \to \infty$,$A_n \to \pi$,$B_n \to \pi$。我们就说,当圆内接多边形无限细分时,其面积的极限就是圆面积,对单位圆而言,极限就是圆周率 π。我们注意到,当 $k \to \infty$,即 $n \to \infty$,$A_n - \pi \to 0$ 和 $B_n - \pi \to 0$,也就是说,$A_n - \pi$ 和 $B_n - \pi$ 都是无穷小量。由此可以得到数列极限的定义:

定义 3.9 如果一个数列$\{a_n\}$和一个常数 A 之差是个无穷小量,即

$$a_n - A = \alpha_n (\alpha_n \text{ 是当 } n \to \infty \text{ 时的无穷小}) \tag{3.36}$$

则称 A 是 a_n 的极限,记为

$$\lim_{n\to\infty} a_n = A \tag{3.37}$$

用极限符号,刘徽割圆术所得的两个数列的极限可以表达为

$$\lim_{n\to\infty}A_n = \pi,\ \lim_{n\to\infty}B_n = \pi \tag{3.38}$$

可以把无穷小数列 a_n 用极限符号表达为

$$\lim_{n\to\infty}a_n = 0 \tag{3.39}$$

定义 3.10 如果一个数列存在极限，则称之为收敛数列，否则称之为发散数列。

例如$\{n\}$、$\{n^2\}$、$\{2^n\}$等都趋向于无穷大；而$\{1, 0, 1, 0, \cdots\}$始终在 0 和 1 之间摆动，$\{1, 1/2, 2, 1/3, 3, 1/4, 4, \cdots\}$的偶数项虽然趋向于 0，但奇数项却趋向无穷大；这些都是发散数列。

现在可以回答前述的芝诺悖论了。回到 3.3 式。利用等比级数的求和公式，

$T_n = \dfrac{S}{V_A}(1+\lambda+\lambda^2+\cdots+\lambda^n) = \dfrac{S}{V_A}\cdot\dfrac{1-\lambda^{n+1}}{1-\lambda}$。而 $\lambda = \dfrac{V_T}{V_A} \ll 1$，$\lim\limits_{n\to\infty}\lambda^{n+1} = 0$，所以 T_n 是收敛数列，即阿基里斯追赶乌龟的时间 T 是有限的。

$$T = \lim_{n\to\infty}T_n = \frac{S}{V_A}\lim_{n\to\infty}\frac{1-\lambda^{n+1}}{1-\lambda} = \frac{S}{V_A}\cdot\frac{1}{1-\dfrac{V_T}{V_A}} = \frac{S}{V_A-V_T}.$$

3.4.2 连续变量函数的极限

当变量 x 连续变化时，不论 $x\to\infty$ 还是 $x\to x_0$，如果函数 $f(x)$无限逼近于一个常数 A，那么我们说 A 是 $f(x)$的极限。

定义 3.11 如果一个函数 $f(x)$和一个常数 A 之差是个无穷小量，即

$$f(x)-A=\alpha(x)\ (\alpha(x)\text{ 是当 } x\to\infty \text{ 或 } x\to x_0 \text{ 时的无穷小}) \tag{3.40}$$

则称 A 是 $f(x)$的极限，记为

$$\lim_{x\to\infty}f(x) = A \quad \text{或} \quad \lim_{x\to x_0}f(x) = A \tag{3.41}$$

虽然无穷大并不是极限，但人们也常用极限符号表达。例如

$$\lim_{n\to\infty} a_n = \infty, \lim_{x\to\infty} f(x) = \infty, \lim_{x\to x_0} f(x) = \infty \tag{3.42}$$

3.4.3 极限的算术运算法则

(1) 和(差)的极限等于极限的和(差)

$$\lim f(x) = A, \lim g(x) = B \Rightarrow \lim(f(x) \pm g(x)) = A \pm B \tag{3.43}$$

这里 A 和 B 是有限数;

(2) 积的极限等于极限的积

$$\lim f(x) = A, \lim g(x) = B \Rightarrow \lim(f(x)g(x)) = AB \tag{3.44}$$

这里 A 和 B 是有限数;

(3) 在分母不为零的前提下,商的极限等于极限的商

$$\lim f(x) = A, \lim g(x) = B \Rightarrow \lim \frac{f(x)}{g(x)} = \frac{A}{B} \tag{3.45}$$

这里 A 和 B 是有限数,且 $B \neq 0$。

注意 A 和 B 都是有限数。正如前面讲述无穷大与无穷大的运算时提到的,∞、$+\infty$、$0 \cdot \infty$、$\frac{0}{0}$ 都是未定型,不能简单套用极限运算法则。

3.4.4 极限的存在准则

给定一个数列,能不能判断它有极限?如果存在极限,又如何求?

判定一个数列极限存在与否并无一定之法,有两个极限存在准则。先介绍两个概念。

定义 3.12 对于数列$\{a_n\}$,如果存在一个正数 M,满足 $|a_n| < M$,则称此数列有界。如果存在常数 A,$a_n < A$,则称此数列上有界,A 记为上界(注意,A 不一定是正数);如果存在常数 B,$a_n > B$,则称此数列下有界,B 记为下界(注意,B 不一定是负数)。

读者可以证明,有界数列一定上下有界,反之也成立。读者还可以思考,无

穷数列和无界数列的关系,收敛数列和有界数列的关系。

定义 3.13　对于数列$\{a_n\}$,如果$a_n \leqslant a_{n+1}$,则称此数列单调增,或递增;如果$a_n \geqslant a_{n+1}$,则称此数列单调减,或递减。如果将定义中的"$\leqslant$"或"$\geqslant$"改为"$<$"或"$>$",那么称此数列是严格单调增或严格单调减。

极限存在准则 1　如果数列$\{a_n\}$单调且有界(即单调增且上有界或单调减且下有界),则一定存在极限。

这条准则与实数的连续性有关,我们不严格证明了。读者不难理解,单调性排除了摆动的可能;有界性则排除了趋向无穷大的可能。

刘徽"割圆术"形成的两个数列,由 3.24 式,作为圆的内接多边形,随着边数的增加,其面积数列$\{A_n\}$是单调增的,但它有上界,$A_n < B_6$,所以$\{A_n\}$存在极限;同理,数列$\{B_n\}$是单调减的,但它有下界,$B_n > A_6$,所以$\{B_n\}$也存在极限。

博尔特创造新世界纪录　图 3.12

2008 年 8 月北京奥运会期间,牙买加运动员博尔特(Usain Bolt, 1986-)在"鸟巢"创造了男子 100 米短跑的新纪录 9.69 秒,被称为"奇迹";一年后他在柏林世界田径锦标赛上,又提高到 9.58 秒(如图 3.12 所示)。再度引起人们对人类体能极限的兴趣。男子 100 米的第一个正式纪录是 1912 年,美国人利平科特(Donald Lippincott, 1893-1963)在瑞典斯德哥尔摩奥运会所创,成绩是 10.6 秒。表 3.5 记录了此后打破世界纪录的历程,以及当时专家对人类跑 100 米极限的预测。[3-12]

表 3.5

年 份	1936	1968	1991	1999	2008	2009	?
纪录/秒	10.2	9.95	9.86	9.79	9.69	9.58	?
运动员	欧文斯 Jesee Owens	吉姆·海因 Jim Hines	刘易斯 Carl Lewis	格林 Maurice Green	博尔特 Usain Bolt	博尔特 Usain Bolt	?
国籍	美国	美国	美国	美国	牙买加	牙买加	?
预测极限	10.0	9.90	9.80	9.726	9.60	9.51	?

到底 100 米纪录有极限吗？按照上述极限存在准则，当然有极限。由表 3.5 第 2 行可见，100 米世界纪录是个单调减数列，显然，它也是下有界的，肯定大于 0。同理，所有径赛项目的速度、跳高项目的高度、跳远项目的距离、投掷项目的距离、举重项目的重量都有极限。至于极限是什么，就难以预测了，从表 3.5 最后一行可见，专家预测经常因运动员的精彩表演化为笑柄。一位荷兰数学家对 100 米的最新预测是 9.51 秒，然而博尔特在柏林比赛后放言自己可跑 9.40 秒！[3-13]看来，极限只能由运动员、教练员以及体育科技工作者的努力来逼近。

数列的极限存在准则 1 对于函数极限也是成立的。

定义 3.14 如果函数 $f(x)$ 在定义域中满足 $|f(x)| \leqslant M$，则称此函数有界。若 $f(x) \leqslant A$，称此函数上有界；$f(x) \geqslant B$，称此函数下有界。

定义 3.15 如果函数 $f(x)$ 在定义域中任两点 $x_1 < x_2$，总有 $f(x_1) \leqslant f(x_2)$，则称此函数在定义域中单调增；如果总有 $f(x_1) \geqslant f(x_2)$，则称此函数单调减。如果将定义中的"$\leqslant$"或"$\geqslant$"改为"$<$"或"$>$"，那么称此函数是严格单调增或严格单调减。

极限存在准则 1′ 如果函数 $f(x)$ 在 $a < x < +\infty$ 有定义，单调增且上有界(或单调减下有界)，则当 $x \to +\infty$ 时，函数 $f(x)$ 存在极限。

注意，上述准则只能判断极限的存在性，并不能求出极限值。

极限存在准则 2 如果一个数列 $\{a_n\}$ 的通项值始终介乎另两个数列 $\{b_n\}$ 和 $\{c_n\}$ 相应通项值之间，即 $b_n \leqslant a_n \leqslant c_n$，而 $\{b_n\}$ 和 $\{c_n\}$ 趋于同一极限，即 $\lim\limits_{n\to\infty} b_n = \lim\limits_{n\to\infty} c_n = A$，则数列 $\{a_n\}$ 存在极限，极限值与 $\{b_n\}$ 和 $\{c_n\}$ 的极限相同，即 $\lim\limits_{n\to\infty} a_n = A$。

历史上曾有人从内接 8 边形出发，再边数加倍，逐步逼近圆周率。由于内接 8 边形的面积总是大于内接 6 边形面积，又总是小于内接 12 边形面积。由于数列 $\{A_6, A_{12}, A_{24}, \cdots\}$ 和数列 $\{A_{12}, A_{24}, A_{48}, \cdots\}$ 都收敛于 $\pi = 3.14159$，所以根据极限存在准则 2，数列 $\{A_8, A_{16}, A_{32}, \cdots\}$ 一定收敛于相同的 $\pi = 3.14159$。

这个准则同样可以推广到函数极限。

极限存在准则 2′ 如果存在一个正数 M，当 $|x| > M$，函数 $f(x)$，$g(x)$，$h(x)$ 始终满足 $g(x) \leqslant f(x) \leqslant h(x)$，且 $\lim\limits_{x\to\infty} g(x) = \lim\limits_{x\to\infty} h(x) = A$，则当 $x \to \infty$ 时 $f(x)$ 存在极限，且 $\lim\limits_{x\to\infty} f(x) = A$。

极限存在准则 2″ 如果函数 $f(x)$，$g(x)$，$h(x)$ 在 x_0 的领域 $x_0 - c < x < x_0 + c$ 始终满足 $g(x) \leqslant f(x) \leqslant h(x)$，且 $\lim\limits_{x\to x_0} g(x) = \lim\limits_{x\to x_0} h(x) = A$，则当 $x \to x_0$ 时 $f(x)$ 存在极限，且 $\lim\limits_{x\to x_0} f(x) = A$。

3.4.5 两个重要极限

在微积分中会用到两个重要极限。

$$(1)\ \lim_{x\to 0} \frac{\sin x}{x} = 1 \tag{3.46}$$

由于 $\lim\limits_{x\to\infty} \sin x = 0$，$\lim\limits_{x\to 0} x = 0$，$\lim\limits_{x\to 0} \dfrac{\sin x}{x}$ 是 $\dfrac{0}{0}$ 型的未定式，不能套用极限运算法则。不妨先通过数据计算进行实验，见表 3.6。

表 3.6

x	1	0.5	0.1	0.01	0.001
$\sin x$	0.84147	0.47942	0.09983	0.00999	0.00099
$\dfrac{\sin x}{x}$	0.84147	0.95885	0.99833	0.99998	1.00000

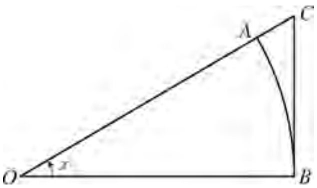


图 3.13

可见随着 $x\to 0$，$\dfrac{\sin x}{x}$ 的确 $\to 1$。但是这不能算数学证明。我们可以应用前述极限存在准则 $2'$ 来证明。图 3.13 所示扇形 OAB 中，设其半径为 1，圆心角为 x 弧度。$\angle OBC$ 为直角，C 点在 OA 的延长线上。

由于$\triangle OAB$ 的面积<扇形 OAB 的面积<$\triangle OBC$ 的面积，即

$$\frac{\sin x}{2}<\frac{x}{2}<\frac{\tan x}{2},\ 1<\frac{x}{\sin x}<\frac{1}{\cos x}\Rightarrow \cos x<\frac{\sin x}{x}<1$$

当 $x\to 0$，$\lim\limits_{x\to 0}\cos x=\cos 0=1$，$\lim\limits_{x\to 0}1=1$，所以 $\lim\limits_{x\to 0}\dfrac{\sin x}{x}=1$。

上面从几何图形出发，假定了圆心角 $x>0$，如果 $x<0$ 又如何？当 $x<0$ 时，必有 $\sin x<0$，这时

$$\frac{\sin x}{x}=\frac{\sin(-x)}{-x}\Rightarrow \lim_{\substack{x\to 0\\ x<0}}\frac{\sin x}{x}=\lim_{\substack{-x\to 0\\ -x>0}}\frac{\sin(-x)}{-x}=1$$

这就证明了 3.46 式。3.46 式又表明，当 $x\to 0$，$\sin x$ 和 x 是“等价无穷小”。

(2) $\lim\limits_{x\to\infty}\left(1+\dfrac{1}{n}\right)^n=e=2.718283\cdots$ (3.47)

我们还是先试算，见表 3.7。

表 3.7

n	1	5	10	100	1000	10000	100000
$1+\frac{1}{n}$	2	1.2	1.1	1.01	1.001	1.0001	1.00001
$\left(1+\frac{1}{n}\right)^n$	2	2.48832	2.59374	2.70481	2.71692	2.71814	2.71827

如果记 $e_n=\left(1+\dfrac{1}{n}\right)^n$，可以验证 $e_n<e_{n+1}$，以及 $e_n<3$（有兴趣的读者不妨自己尝试一下），由极限存在准则 1，可以证明数列 $\{e_n\}$ 存在极限，且在 2 与 3 之间。

e 和 π 一样，是个无理数，而且是个超越数。在理论和应用中极其重要。

例如银行存款采用复利率公式

$$A = P(1+r)^m \tag{3.48}$$

其中 P 是本金，A 是收益，r 是年利率，m 是存款年数。

实际上银行使用的是改进的公式

$$A = P\left(1+\frac{r}{k}\right)^{km} \tag{3.49}$$

这里 k 是一年分的期数。如果 $k=4$，表示按季度计算；$k=12$，表示按月计算；$k=365$，表示按日计算。那么不同的分期计算，收益相不相同呢？不妨设本金 $P=10000$ 元，年利率 $r=6\%$，$m=5$，$k=1$，4，12，365，结果见表 3.8。

表 3.8

k\m	**0**	**1**	**2**	**3**	**4**	**5**
1	10000	10600	11236	11910	12625	13382
4	10000	10614	11265	11956	12690	13469
12	10000	10617	11272	11968	12706	13490
365	10000	10618	11275	11972	12712	13498

可以发现，分期越多，收益越大，但增大的幅度越来越小。如果分期按小时，甚至按分按秒计算，收益会无限增大吗？对 3.49 式作一些数学变换

$$A = P\left(1+\frac{r}{k}\right)^{km} = P\left(1+\frac{1}{\frac{k}{r}}\right)^{\frac{k}{r}rm}$$

设 $n=\frac{k}{r}$，则

$$A = P\left(1+\frac{1}{n}\right)^{nrm} = P\left(\left(1+\frac{1}{n}\right)^{n}\right)^{rm}$$

由 3.43 式

$$\lim_{n\to\infty} P\left(\left(1+\frac{1}{n}\right)^{n}\right)^{rm} = Pe^{rm} \tag{3.50}$$

可见，上面例子中，5 年的收益不会超过 $A_{\max} = 10000e^{0.06\times 5} \approx 13499$ 元。

欧拉曾经得出一个以他命名的重要公式

$$e^{i\pi} + 1 = 0 \tag{3.51}$$

这里 1 是自然数的起点，又是实数的单位；0 既是“无”，又表征一个数的位数；$i = \sqrt{-1}$ 是虚数单位；e 和 π 则是两个最重要的无理数和超越数。这些常数居然能“安居”在一个形式特别简单的公式中，不能不赞叹数学的奇妙！欧拉是数学的多面手，在分析、微分方程、数论、代数、几何、拓扑等许多领域都有重大建树。他 64 岁时几乎失明，靠惊人的记忆，口述他的研究成果。[3-14]欧拉曾有“数学的莎士比亚”之称。比较 3.51 式和莎士比亚在《Hamlet（哈姆雷特）》中的名句：“生存还是毁灭，这是一个值得考虑的问题（to be or not to be, that is a question）”[3-15]，不是同样深刻和简洁吗！

莎士比亚　1564-1616

欧拉　1707-1783

3.5 连续与离散

3.5.1 连续函数的概念

中学学习过许多函数，例如幂函数 $y = x^a$；三角函数 $y = \sin x$，$y = \cos x$，$y = \tan x$，$y = \cot x$；反三角函数 $y = \sin^{-1} x$，$y = \cos^{-1} x$，$y = \tan^{-1} x$，$y = \cot^{-1} x$；指数函数 $y = a^x (a > 0, a \neq 1)$；对数函数 $y = \log_a x\ (a > 0, a \neq 1)$ 等等。在微积分中，这些函数统称为“基本初等函数”，从图形上看，这些函数大多为连续的曲线，部分图形如图 3.14 所示。

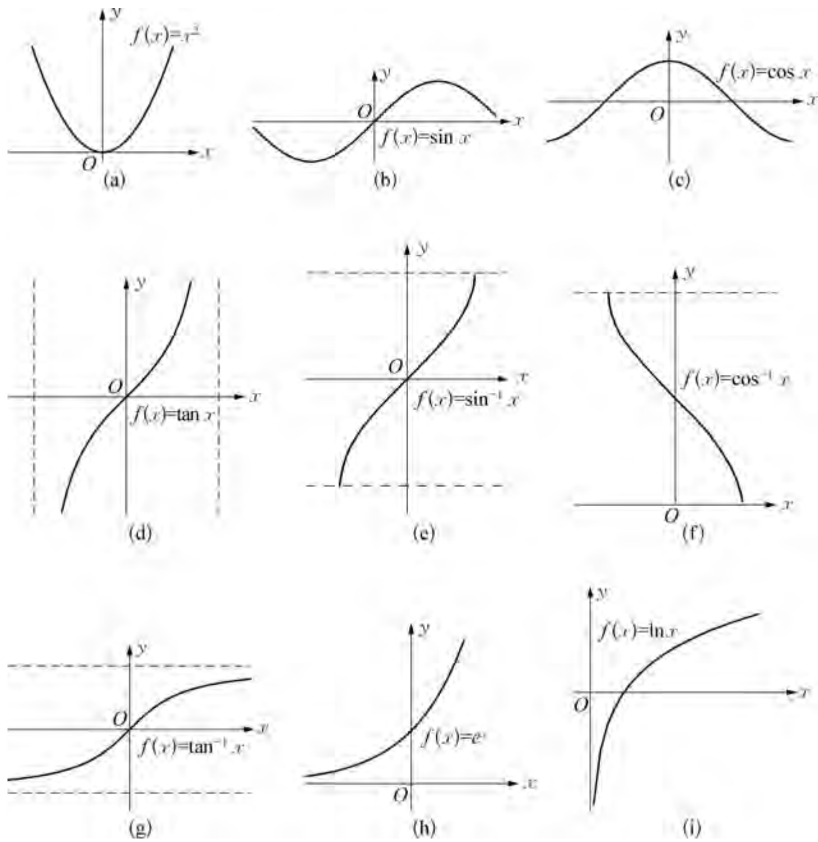


图 3.14

由于时间和空间的连续性，实际问题中研究的函数自变量和因变量多为连续变量，因此连续函数往往是主要的研究对象。但有时也会遇到不连续的自变量或因变量。例如数列就可以看作是自变量为正整数的函数，例如 $a_n = f(n) = 1 + 1/n$，其图形由图 3.15 所示，函数图形是离散的点集合。

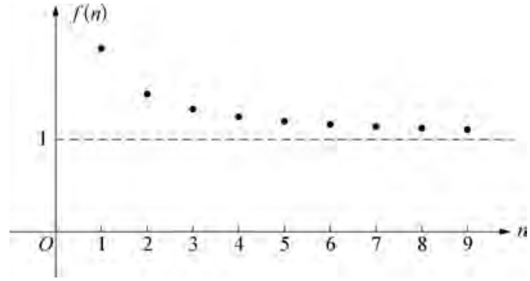


图 3.15

即使自变量是连续变量，函数曲线也可能是间断的。实际问题中不乏这样的例子。

【例 3-3】 中国邮政特快专递的资费是邮件重量的函数，单件不超过 40 千克，规定如表 3.9 所示[3-16]：

表 3.9

起重资费	续重资费				
起重 500 克及以内	续重每 500 克及其零数				
20 元	一区	二区	三区	四区	五区
	4 元	6 元	9 元	10 元	17 元

以一区收费为例，设邮件重量为 x 克，邮资为 y 元，则函数关系为

$$y = 20 + 4(k-1)，当 (k-1)500 < x \leqslant 500k，k = 1，2，\cdots，80 \tag{3.52}$$

其图形如图 3.16 所示。当 x=500，1000，1500，…，3500，函数都间断。

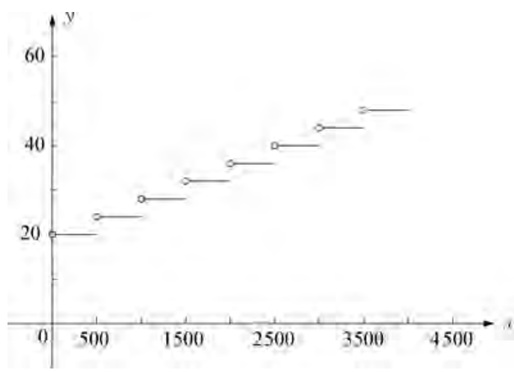


图 3.16

数学上还有处处有定义，但处处不连续的函数。

【例 3-4】 Dirichlet 函数定义如下：

$$y = \begin{cases} 1， & 当 x 是有理数； \\ 0， & 当 x 是无理数 \end{cases} \tag{3.53}$$

这样的函数我们甚至无法画出来！

数学上怎样定义函数在一点连续还是间断？这就要用到极限作为工具。

定义 3.16 设函数 $y=f(x)$ 在实数域 X 有定义，x_0 为 X 中一点。设 x 在 x_0 处有一个改变量 Δx，y 的相应改变量为

$$\Delta y = f(x_0+\Delta x)-f(x_0) \tag{3.54}$$

如果

$$\lim_{\Delta x\to 0}\Delta y = 0 \tag{3.55}$$

则称函数 $f(x)$ 在 x_0 点连续，否则称函数 $f(x)$ 在 x_0 点间断。

【例 3-3】中当 $x=500$，1000，1500，…，3500，函数都间断。以 $x=500$ 为例，当 $\Delta x<0$，即 x 从左面趋近于 500 时，$\lim\limits_{\Delta x\to 0}\Delta y=0$，但当 x 从右面趋近于 500 时，却有 $\lim\limits_{\Delta x\to 0}\Delta y=4\neq 0$。这样的间断点成为“跳跃间断点”。3.55 式也常常等价地表示为

$$\lim_{x\to x_0}f(x)=f(x_0) \tag{3.56}$$

例如 $\lim\limits_{x\to 3}x^2=3^2=9$；$\lim\limits_{x\to 0}\cos\left(x+\dfrac{\pi}{3}\right)=\cos\dfrac{\pi}{3}=\dfrac{1}{2}$；$\lim\limits_{x\to 1}\ln x=\ln 1=0$；等等。前面证明 $\lim\limits_{x\to 0}\dfrac{\sin x}{x}=1$ 过程中曾用到 $\lim\limits_{x\to 0}\cos x=\cos 0=1$，这就是余弦函数的连续性质。

定义 3.17 如果函数 $f(x)$ 在 x_0 点附近有定义，在 x_0 点可以有定义，也可以无定义，但

$$x\to x_0 \quad 时 \quad f(x)\to\infty \tag{3.57}$$

也称函数 $f(x)$ 在 x_0 点间断，这时 $x=x_0$ 称为“无穷间断点”。

图 3.13 中 $x=0$ 就是 $y=\dfrac{1}{x}$ 的无穷间断点；$x=\dfrac{\pi}{2}$ 就是 $y=\tan x$ 的无穷间断点。

定义 3.18 如果函数 $f(x)$ 在 x_0 点附近有定义，在 x_0 点可以有定义，也可以无定义，但 $x \to x_0$ 时 $f(x)$ 在不同数值间振荡而无极限，也称函数 $f(x)$ 在 x_0 点间断，这时 $x = x_0$ 称为“振荡间断点”。

例如 $y = \sin\dfrac{1}{x}$，当 $x \to 0$，$\sin\dfrac{1}{x}$ 在 -1 和 1 之间振荡，$x = 0$ 就是“振荡间断点”。如图 3.17 所示。

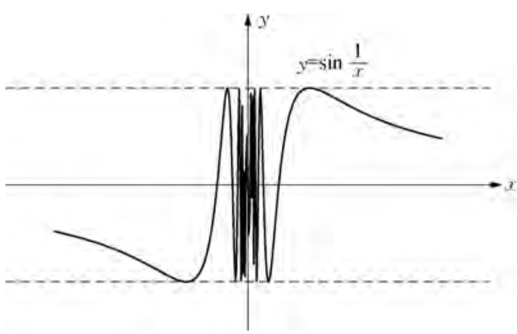


图 3.17

3.5.2 闭区间上连续函数的性质

连续函数有两个非常重要的性质，在微积分和其他学科中常会用到。先引进两个数学表达记号。

(1) 实数的开区间与闭区间。如果实数 x 的变化范围是 $a < x < b$，记作 $x \in (a, b)$，(a, b) 称为“**开区间**”；如果实数 x 的变化范围是 $a \leqslant x \leqslant b$，记作 $x \in [a, b]$，$[a, b]$ 称为“**闭区间**”；如果实数 x 的变化范围是 $a < x \leqslant b$，或 $a \leqslant x < b$，记作 $x \in (a, b]$ 或 $x \in [a, b)$，$(a, b]$ 和 $[a, b)$ 称为“半开半闭区间”。

(2) 如果函数 $f(x)$ 在开区间 (a, b) 连续，记作 $f(x) \in C(a, b)$ 或 $f \in C(a, b)$；如果函数 $f(x)$ 在闭区间 $[a, b]$ 连续，记作 $f(x) \in C[a, b]$ 或 $f \in C[a, b]$。C 是“Continuous(连续的)”一词的第一个字母。

数学表达式往往比较抽象，却非常简洁和严密。

我们不加证明，叙述两个闭区间上连续函数的性质。（证明可参阅[3-21]）

定理 3.1 （最值定理）设 $f \in C[a, b]$，则 $f(x)$ 在 $[a, b]$ 上一定能取得它的最大值和最小值。

注意，最大值和最小值既可能在区间内点达到，如图 3.18(a)所示，也可能在区间端点达到，如图 3.18(b)(c)(d)所示；达到最值的点可能不止一个，如图 3.18(b)所示

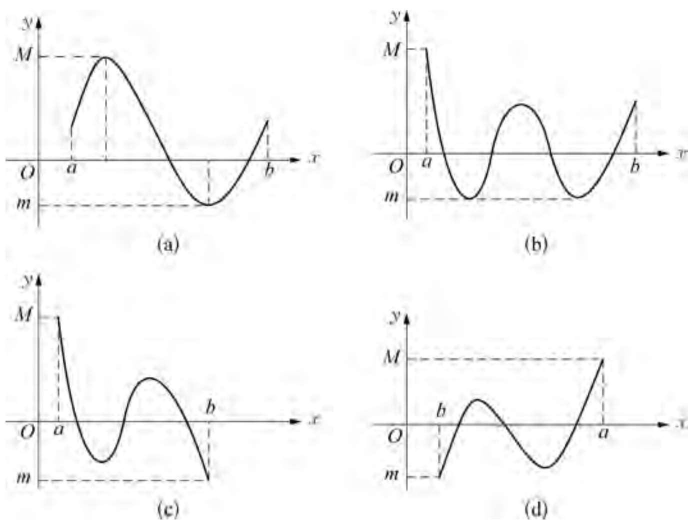


图 3.18

还要注意，如果函数有不连续点，或者函数仅在开区间连续，上述定理的结论不一定成立。例如取 $a=0$，$b=2$，$y=\dfrac{1}{x-1}$，$x=1$ 是一个无穷间断点，$\lim\limits_{x \to 1} \dfrac{1}{x-1}=\infty$，函数在$[0, 2]$中不存在最值，如图 3.19(a)所示；而 $\lim\limits_{x \to 0} \dfrac{1}{x}=\infty$，函数 $y=\dfrac{1}{x}$ 在左端点趋向于无穷，所以虽然达到最小值 $m=y(2)=\dfrac{1}{2}$，却不存在最大值，如图 3.19(b)所示。

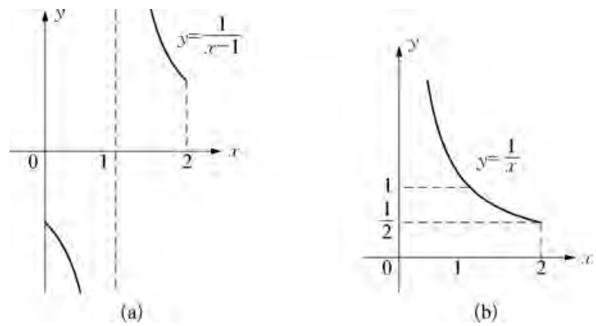


图 3.19

定理 3.2 （介值定理）设 $f \in C[a, b]$，$f(a) \neq f(b)$，μ 为介于 $f(a)$ 与 $f(b)$ 之间的任一值，则至少存在一点 $\xi \in (a, b)$，使得 $f(\xi) = \mu$，如图 3.20 所示。

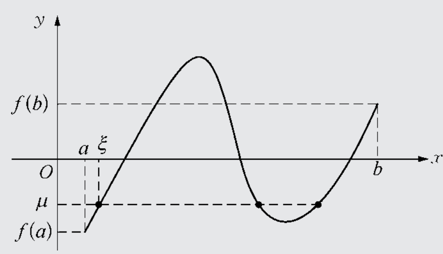


图 3.20

介值定理为非线性方程求解提供了一个有效的方法：“**对分法**”。在求 $f(x) = 0$ 的解即“零点”的问题中，如果在定义域中找到两点 a 和 b，$f(a) < 0$，$f(b) > 0$，或 $f(a) > 0$，$f(b) < 0$，由介值定理，在 (a, b) 之内至少有一个零点。然后计算 $f\left(\frac{a+b}{2}\right)$，如果 $f\left(\frac{a+b}{2}\right)$ 和 $f(a)$ 同号，则在 $\left(\frac{a+b}{2}, b\right)$ 中再找中点；如果 $f\left(\frac{a+b}{2}\right)$ 和 $f(b)$ 同号，则在 $\left(a, \frac{a+b}{2}\right)$ 中再找中点。这样，直到搜索区间达到计算精度为止。

【例 3-5】 求 $f(x) = x - 2\sin x = 0$ 在$[\pi/2, \pi]$中的零点，精度为 0.1。

解： 第 1 步：$f\left(\frac{\pi}{2}\right) = \frac{\pi}{2} - 2 = -0.929 < 0$，$f(\pi) \approx 3.1416 > 0$，知

$\left[\frac{\pi}{2}, \pi\right]$ 中至少有一零点，取中点 $x_1 = \frac{1}{2}\left(\frac{\pi}{2}+\pi\right)=\frac{3\pi}{4}$；

第 2 步：$f\left(\frac{3\pi}{4}\right)=\frac{3\pi}{4}-2\sin\frac{3\pi}{4}=0.9419>0$，知 $\left[\frac{\pi}{2}, \frac{3\pi}{4}\right]$ 中至少有一零点，取中点 $x_2 = \frac{1}{2}\left(\frac{\pi}{2}+\frac{3\pi}{4}\right)=\frac{5\pi}{8}$；

第 3 步：$f\left(\frac{5\pi}{8}\right)=\frac{5\pi}{8}-2\sin\frac{5\pi}{8}=0.1158>0$，知 $\left[\frac{\pi}{2}, \frac{5\pi}{8}\right]$ 中至少有一零点，取中点 $x_3 = \frac{1}{2}\left(\frac{\pi}{2}+\frac{5\pi}{8}\right)=\frac{9\pi}{16}\approx 1.767$；

第 4 步：$f\left(\frac{9\pi}{16}\right)=\frac{9\pi}{16}-2\sin\frac{9\pi}{16}=-0.1944<0$，知 $\left[\frac{9\pi}{16}, \frac{5\pi}{8}\right]$ 中至少有一零点，取中点 $x_4 = \frac{1}{2}\left(\frac{9\pi}{16}+\frac{5\pi}{8}\right)=\frac{19\pi}{32}\approx 1.865$；

因为 $|x_4 - x_3| = 0.098 < 0.1$，已达精度。可得零点的近似值为 $\xi \approx 1.8$。

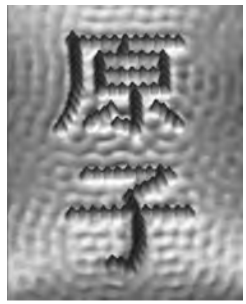
3.5.3 连续与离散的辩证关系

与“连续变量”相对，人们也常常接触“离散变量”，即一个变量的取值为有限个或可数无限个。连续变量和离散变量似乎是对立的，但是人们在研究问题中，往往会互相转化。例如前面提到的连续函数作图，我们不可能找出无限个点，而是先找出有限个点(当然是离散的)，然后再连线，画出图形；又比如气温随时间、随地理位置连续变化，而气象台只选择有代表性的时间和地点向公众报告。反过来，例如高速公路上行驶的汽车当然是离散的，可是人们也常将“汽车流”看作“连续的流体”，这样就可以用流体力学的成熟方法来研究和处理“交通流”。[3-17]

进一步看，连续和离散都是数学抽象。对于现实世界的问题，根据不同的目的和研究的条件，可以选择不同的数学模型。力学领域有一门学科叫做“连续介质力学”[3-18]，把液体、气体看作连续的介质来进行研究。可是从近代物理的角度看，液体、气体都是由分子组成，分子又由原子构成。在微观层面，分子和原子又是离散的。美国物理学家、诺贝尔物理奖获得者理查德·费曼(Rinchard. Feynman，1918-1988)早在 1959 年在美国物理学会召开的年会上，曾大胆预言：

“当我们深入并游荡在原子的周围，我们是在按不同的定律活动，我们会遇到许许多多新奇的事情，能以全新的方式生产，完成异乎寻常的工作。如果有一天可以按人的意志安排一个个原子，将会产生什么样的奇迹？！”[3-19] 30 年之后，1989 年，IBM 公司实验室科学家首先用一台扫描隧道显微镜分别搬移了 35 个氙原子，拼装成了“IBM”三个字母的标识，后来又用 48 个铁原子排列组成了汉字“原子”两个字。[3-20]

费曼 1918-1988



3.6 思考与练习

1. 试举出一两个无穷集合的例子。

2. 在芝诺悖论中，设阿基里斯落后于乌龟的距离是 $S=100$ 米，阿基里斯的速度 $V_A=8$ 米/秒，乌龟速度 $V_T=0.01$ 米/秒，试按 3.3 式，估计阿基里斯追上乌龟的时间(精确到 0.01 秒)。

3. 证明立方体内所有点和线段上所有点相同。

4. 举出两个无穷大数列的例子和两个无穷大连续变量的例子；举出两个无穷小数列的例子和两个无穷小连续变量的例子。

5. 将定义 3.1 改为：“任意给定一个大正数 M，一定存在正整数 N，$|a_N|>M$，则称数列$\{a_n\}$是无穷大数列”。行不行？

6. 举出“无穷大＋无穷大”、“无穷小×无穷大”、“无穷大/无穷大”和“无穷小/无穷小”等不定式的例子。

7. 证明【例 3-1】和【例 3-2】的结论。

8. 利用 $\sqrt{1-\varepsilon}\approx 1-\frac{1}{2}\varepsilon$, $|\varepsilon|\ll 1$, 在 3.15 式中作如下运算：

$$H_n=\sqrt{r^2-\left(\frac{S_n}{2}\right)^2}=\sqrt{1-\left(\frac{S_n}{2r}\right)}\approx r\left(1-\frac{1}{2}\left(\frac{S_n}{2r}\right)^2\right)$$

然后估计 3.26 式。

9. 如果从内接正方形开始，数次让边数加倍，仿照刘徽方法，求出 π 的近似值。

10. 求下列极限：

(1) $\lim\limits_{n\to\infty}(3n^2-2n+7)/(2n^2+5)=$　　(2) $\lim\limits_{n\to\infty}(3n^2-2n+7)/(2n+5)=$

(3) $\lim\limits_{n\to\infty}(3n^2-2n+7)/(2n^3+5)=$　　(4) $\lim\limits_{x\to 0}\frac{\tan x}{x}=$

11. 向银行贷款 50000 元，年还贷率为 5%，分别按季度和按月度计算，3 年后本利共还多少？如按 3.46 式，3 年后应还多少？

12. 判断 $x^5+x^4-3x^3-4x^2+2x+4=0$ 在[0, 2]中是否有根？如有，请找出来，精度为 0.05。

13. 连续变量与离散变量有什么差别？你如何体会两者在一定条件下的转化？

参考文献或网站

[3-1]　雷一东：《墨经校解》，齐鲁书社(2006).

[3-2]　《庄子全译》，贵州人民出版社(2009).

[3-3]　百度百科：芝诺 http://baike.baidu.com.

[3-4]　黑格尔：《哲学史讲演录》(贺麟 王太庆译)，商务印书馆(1978).

[3-5]　盖莫夫：《从一到无穷大(One, Two, Three, ..., Infinity)》(暴永宁译)，科学出版社(1978).

[3-6]　新华网：金沙遗址博物馆 http://www.sc.xinhuanet.com.

[3-7]　(汉)张苍：《九章算术》，重庆大学出版社(2006).

[3-8]　吴文俊主编：《中国数学史大系》第三卷，北京师范大学出版社(1998).

[3-9]　百度百科：圆周率 http://baike.baidu.com.

[3-10]　维基百科：Pi http://en.wikipedia.org/wiki/pi.

[3-11] 新世界教育北京网站：日本公布计算出圆周率小数点后 25769 亿位数 http://bj.neworldedu.org.

[3-12] 维基百科：男子 100 米短跑世界纪录 http://en.wikipedia.org/wiki.

[3-13] 新华社：人类速度极限到底在哪里?《新民晚报》(2009.8.16).

[3-14] 卡茨(Victor J. Katz)：《数学史通论(A History of Mathematics, An Introduction)》(李文林,邹建成,胥鸣伟等译),高等教育出版社(2004).

[3-15] 莎士比亚:《莎士比亚全集》(朱生豪译),大众文艺出版社(1999).

[3-16] 中国邮政网(www.chinapost.com.cn).

[3-17] 谭永基 俞文 :《数学模型》,复旦大学出版社(1997).

[3-18] 谢多夫:《连续介质力学(第 2 版)》(李植译),高等教育出版社(2007).

[3-19] http://wikiquote.org/wiki/Richard_Feynman.

[3-20] 斯特里克兰(Jonathan Strickland):《纳米机器可以操纵原子吗(Can nanoscopic machines manipulate atoms?)》http://science.howstuffworks.com/nanoscopic-machines.htm.

[3-21] 菲赫金哥尔茨:《微积分教程第一卷(第 8 版)》(杨弢亮,叶彦谦译),高等教育出版社(2006).

第四章

导数与微分

微积分是 17 世纪下半叶科学中最伟大的发现,它的产生开创了数学发展史的新纪元,许多科学家、思想家都给予高度评价。恩格斯曾经说过:“在一切理论成就中,未必再有什么像十七世纪下半叶微积分的发明那样看作人类精神的最高胜利了。如果在某个地方我们看到人类精神的纯粹的和唯一的功绩,那就正在这里。”[4-1]冯・诺伊曼则认为:“微积分是近代数学中最伟大的成就,对它的重要性无论做怎样的估计都不会过分。”[4-2]爱因斯坦曾不无自豪地说:“我数学一直很好,15 岁之前就掌握了微积分。……我越来越确信,大自然可以作为一种相对简单的数学结构而得到理解。……数学也许是发现——而不仅仅是描述——自然定律的一种工具。数学是大自然的剧本。”[4-3]马克思曾花许多精力研究微积分,他的《数学手稿》[4-4]包含着许多对微积分和辩证法的精彩论述。

微积分能解决什么问题? 主要思想方法是什么? 本章介绍微积分的第一个基本问题:怎样求曲线的切线,最后归结为求函数的变化率,即导数。

4.1　曲线的切线问题

怎样求出曲线的切线,是一个古老的问题。对于圆的切线,人们早就知道它是和圆只有一个交点的直线,并有如下性质:圆上任一点的切线与过这点的半径垂直。于是只要先连上圆周上某点的半径,再做此半径的垂直线,就是过圆上这点的切线。对于其他曲线,微积分诞生之前,人们研究得并不多,没有找到系统的求切线方法。

4.1.1 古希腊人的几何方法

阿波罗尼奥斯(262-190 B.C.)是欧几里得的学生的学生。第一章曾经提到,他的名著《Conic Section(圆锥曲线)》可以说是继《几何原本》之后古典希腊几何最登峰造极之作。[0-1]其中就有求圆锥曲线的切线的理论和方法。比如用他的方法,可以这样求抛物线的切线(图 4.1):

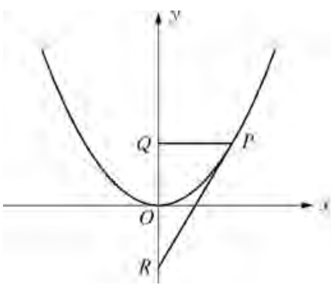


图 4.1

过抛物线上任一点 P,作与对称轴垂直的直线,交于 Q。在对称轴的轴延长线上找抛物线顶点 O 的对称点 R,即 $OR = OQ$。连 RP,由抛物线性质,此直线即为抛物线过 P 点的切线。阿氏方法基于对圆锥曲线性质的深入研究,非常巧妙,在当时,非大学者难以求解;而对于一般的曲线,这样的纯几何方法显然无能为力。

4.1.2 笛卡儿的解析几何方法

图 4.2 显示了用解析几何方法求抛物线切线的过程。

设抛物线方程为 $y = x^2$,$P(a, b)$ 为此抛物线上任一点。希望找到一个圆心在 y 轴上的圆,并与此抛物线相切,即只有一个交点,那么圆在 P 点的切线也就是抛物线在 P 点的切线。

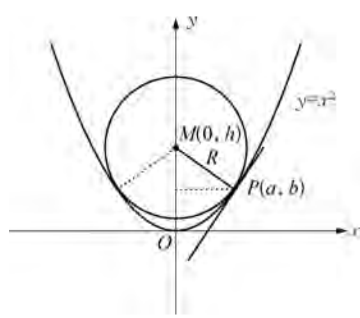


图 4.2

设圆心为 $M(0, h)$,半径为 R,由 1.18 式,则此圆方程为

$$x^2 + (y-h)^2 = R^2$$

与 $y = x^2$ 联立,解出

$$y = \frac{1}{2}(2h-1) \pm \sqrt{(1-2h)^2 - 4(h^2 - R^2)}$$

由于只有一个交点,故 $\sqrt{(1-2h)^2 - 4(h^2-R^2)} = 0$,解出 $R^2 = h - \dfrac{1}{4}$。

由勾股定理，$R^2=(h-b)^2+a^2=(h-b)^2+b$，于是，$h-\frac{1}{4}=(h-b)^2+b$，即$\left(h-b-\frac{1}{2}\right)^2=0$，故又有 $h=b+\frac{1}{2}$。

连 MP 直线的斜率为

$$k_{MP}=\frac{h-b}{0-a}=-\frac{h-b}{a}$$

解析几何知识告诉我们，两条垂直直线的斜率互为负倒数。又因抛物线过 P 点的切线与 MP 垂直，故其斜率为

$$k=-\frac{1}{k_{MP}}=\frac{a}{h-b}$$

由直线的点斜式 1.17，则切线方程为

$$y-b=k(x-a)=\frac{a}{h-b}(x-a)\quad（其中\ b=a^2）$$

笛卡儿的方法用不着用直尺和曲线板画图，却可精确计算切线的斜率和方程，显示了解析几何的威力。

【例 4-1】 用上述方法求在抛物线 $y=x^2$ 上求过 $P(2, 4)$的切线。

由前面的推导，$h=4+\frac{1}{2}=4.5$，故 M 点坐标为$(0, 4.5)$。连 MP 直线的斜率为

$$k_{MP}=\frac{4.5-4}{0-2}=-0.25$$

抛物线过$(2, 4)$的切线斜率为

$$k=-\frac{1}{k_{MP}}=4$$

于是，过 $P(2, 4)$的切线方程为 $y-4=4(x-2)$，即 $4x-y-4=0$。

用阿波罗尼奥斯方法验证一下，R 点的坐标为$(0, -4)$，按两点式，PR 直线方程为 $\frac{y+4}{8}=\frac{x}{2}$，即 $4x-y-4=0$，与前述结果相同。

笛卡儿方法虽然摆脱了几何作图，但是对更复杂的曲线还是难以应用。

从15世纪初文艺复兴时期起，欧洲的工业、农业、航海事业与商贸等都得到大规模的发展，生产实践的需要对自然科学提出了新的课题：迫切要求力学、天文学等基础科学的发展，而这些学科都是深深依赖于数学的，因而也推动了数学的发展。就以求曲线的切线而言，在几何学和运动学两方面提出了迫切的要求。

4.1.3 几何学的要求

望远镜和显微镜的发明，使得求出任意曲线的切线，成为一个迫切的实用问题。

玻璃是公元一世纪罗马人发明的。人们从生产实际中早就发现，把玻璃磨成凸透镜或凹透镜，通过光的折射，会出现将物体放大缩小的现象。望远镜和显微镜正是将凸透镜和凹透镜进行组合而成，可是它们的诞生却推迟了一千多年。

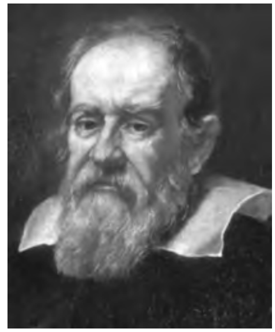
利伯希 1570-1619

17世纪初的一天，荷兰小镇的一家眼镜店的主人利伯希(Hans Lippershey，1570-1619)，为检查磨制出来的透镜质量，把一块凸透镜和一块凹镜排成一条线，通过透镜看过去，发现远处的教堂塔尖好像变大拉近了，于是在无意中发现了望远镜的秘密。1608年他为自己制作的望远镜申请专利，制造了一个双筒望远镜。

望远镜发明的消息很快在欧洲各国流传开，意大利科学家伽利略(Galileo Galilei，1564-1642)得知这个消息之后，就自制了一个。第一架望远镜只能把物体放大3倍。一个月之后，他制作的第二架望远镜可以放大8倍，第三架望远镜可以放大到20倍。1609年10月他作出了能放大30倍的望远镜。伽利略用自制的望远镜观察夜空，第一次发现了月球表面高低不平，覆盖着山脉并有火山口的裂痕。此后又发现了木星的4个卫星、太阳的黑子运动，并作出了太阳在转动的结论。所以人们把1609年作为望远镜发明年，并认为伽利略是发明者，2009年世界科学界曾隆重纪念望远镜发明400周年。不过，伽利略本人一直坚持是荷兰人发明的。[4-5]

如果说，望远镜使人看到远处的事物，那么显微镜就使人看到近处微小的事物。显微镜也是荷兰人发明的。1590年，在荷兰的米德尔堡，有一个名叫詹森

(Zacharias Janssen，1580-1632)的少年。他在父亲的眼镜制造工场玩耍时,无意间将两片凸透镜重叠放在一起,发现镜片下的小蚂蚁大了许多倍。少年被这奇怪的现象吸引住了。于是,他用薄铁片卷了两个铁筒,让小铁筒在大铁筒里滑动。他把两个凸透镜分别装在大小铁筒上,利用铁筒的滑动,调整透镜的距离,这样就可以得到较清晰的成像。这就是显微镜的雏形。詹森虽然是发明显微镜的第一人,却并没有发现显微镜的真正价值。



伽利略 1564-1642

詹森 1580-1632

列文虎克 1632-1723

詹森的发现引起科学界的兴趣。伽利略也制作过显微镜,但是贡献最大的是荷兰人列文虎克(Anthony Leeuwenhoek，1632-1723)。他一生磨制了 400 多个透镜。他的遗物中有一架简单的透镜,其放大率竟达 270 倍。1674 年他开始观察细菌和原生动物,1677 年他首次描述了昆虫、狗和人的精子。1684 年他准确地描述了红细胞,证明马尔皮基推测的毛细血管是真实存在的。1702 年他指出在所有露天积水中都可以找到微生物。列文虎克是第一个用放大透镜看到细菌和原生动物的人,对 18 世纪和 19 世纪初期细菌学和原生动物学研究的发展,起了奠基作用。他也被尊称为“显微镜之父”[4-6]。

早期望远镜和显微镜的制作都是凭工匠的经验。但要进一步提高放大倍数,就必须研究曲线的切线了。

人们早就发现将一根棍棒斜插水中会出现弯折现象,也就是说,光线穿过不

斯涅耳 1591-1626

同介质会产生折射。荷兰数学家和物理学家威里布里德·斯涅耳(Willebrord Snell Van Roijen, 1591-1626),最早发现了光的折射定律,从而使几何光学的精确计算成为可能[4-7]。

光的折射定律(斯涅耳定律):光入射到不同介质的界面上会发生折射和反射。其中入射光、反射光和折射光位于同一个平面上,并且与界面法线的夹角满足如下关系:

$$\theta_1 = \alpha_1, \frac{\sin\theta_1}{\sin\theta_2} = \frac{\gamma_2}{\gamma_1} \tag{4.1}$$

其中,γ_1, γ_2 分别是两个介质的折射率,θ_1, θ_2, α_1 分别是入射光、折射光、反射光与界面法线的夹角,称为入射角、折射角和反射角。如图 4.3 所示。

人们进一步发现,一种介质的折射率,与光在这种介质中的传播速度相关,即

$$\gamma = \frac{c}{v} \tag{4.2}$$

其中 $c=30$ 万千米/秒,为光在真空中的速度;v 为光在这种介质中的速度。由4.1式可知,光传播速度大的介质,折射率小。

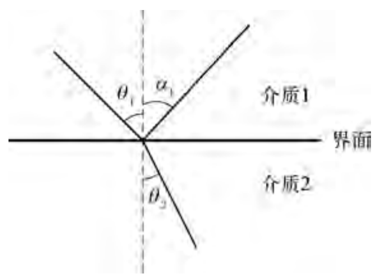


图 4.3

如果光线穿过一块凸透镜或凹透镜,计算折射角就必须求出曲线一点的切线斜率,从而得到这一点法线的斜率。以图 4.4 为例,设一玻璃凸透镜的曲面投影为一抛物线

$$y = x^2 \tag{4.3}$$

有一束水平光从空气射入 $P(1, 1)$点,问入射角是多少?若空气折射率看作 1,此玻璃折射率为 1.6,折射角是多少?

用【例 4-1】的方法,不难算出抛物线 4.3 在 P 点的切线斜率

$$k_T = 2 \tag{4.4}$$

由解析几何知识,过 P 点的法线斜率

$$k_N = -\frac{1}{k_T} = -0.5$$

于是得出入射角 $\theta_1 = \tan^{-1}0.5$

$= 0.4636$ 弧度 $= 26.56°$。

由 4.1 式，$\sin\theta_2 = \dfrac{1}{1.6}\sin\theta_1 = 0.2795$，

从而得到折射角 $\theta_2 = 0.2833$ 弧度 $=16.23°$。

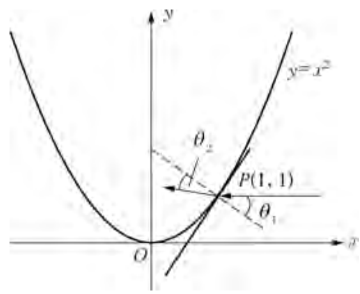


图 4.4

4.1.4 运动学的要求

运动学有一个重要课题是确定运动物体的瞬时方向。如果物体作直线运动，问题很简单，任何瞬间的运动方向都和直线方向重合；如果物体作曲线运动呢？人们很早就知道，任何瞬间的运动方向就是曲线在这点的切线方向。但是古代学者除圆周运动外，很少涉及其他类型的曲线运动。十五世纪之后，人们发现了许多圆以外曲线运动，如地球在以太阳为焦点的椭圆轨道上运行，炮弹在空气中的轨迹近似于抛物线，等等。微积分就可以完满解决这样的问题。

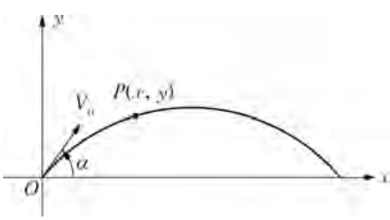


图 4.5

例如由图 4.5 所示，炮弹在坐标原点 $O(0, 0)$ 以与水平线夹角 α 发射，设初速为 V_0，如不考虑空气阻力，那么由微积分知识(参阅 5.6.2 节)，可以很容易求出任意时刻 t 炮弹位置 $P(x, y)$ 的坐标与时间的函数关系是

$$x = (V_0\cos\alpha)t,\ y = -\frac{1}{2}gt^2 + (V_0\sin\alpha)t \tag{4.5}$$

由上式消去 t，不难得出，炮弹轨迹是一条抛物线

$$y = -\frac{g}{2V_0^2\cos^2\alpha}x^2 + (\tan\alpha)x \tag{4.6}$$

由微积分知识(见 4.4.2 节)，还可方便求出任意时刻 t 炮弹瞬时速度的倾角 θ 和大小 $|V(t)|$：

$$\tan\theta = \frac{-gt + V_0\sin\alpha}{V_0\cos\alpha},\ |V(t)| = \sqrt{V_0^2 + g^2t^2 - 2gtV_0\sin\alpha} \tag{4.7}$$

人类现在发射卫星，可以精确计算出任意时刻卫星的位置、速度和方向，这些都离不开微积分的贡献。

4.1.5 曲线上任意点切线的求法

首先要弄清楚，什么是曲线的切线？对于任意曲线，用“只有一个交点的直线”来定义切线显然不可取，例如图 4.6 所示的曲线 C，在 P 点的切线是 L，但 L 与 C 还有另一个交点 Q。

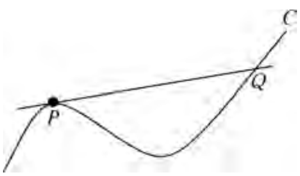


图 4.6

其实人们处理曲线问题还是有朴素的办法。比如北京天坛的回音壁是墙厚 3.72 米的圆形建筑。其直径为 61.5 米。虽然砌墙的砖块是直的，但是每块砖的长度只有几十厘米，与回音壁周长 193.2 米相比非常小，不影响整体是圆形。如图 4.7 所示。

图 4.7

我们看到古代乃至今天的许多桥涵、烟囱、水井，虽然是圆形，几乎无一例外由方砖砌成。其实，第五讲介绍过的刘徽“割圆术”，就是基于“局部用割线（即“弦”）代替圆弧”的想法。这个思路完全可以用来研究一般曲线的切线。首先局部用小的割线段近似代替曲线的切线，然后使割线段的尺寸越来越小，如果存在一个极限位置，那么这个极限位置就是切线。

下面仍以抛物线 $y=x^2$ 为例，求此抛物线过 $P(1,\ 1)$点的切线。

先在抛物线上 P 点附近找一点 $Q_1(1.1,\ 1.21)$，则割线 PQ_1 的斜率为

$$\tan\beta_1=\frac{1.21-1}{1.1-1}=2.1$$

再找更近的点 $Q_2(1.01, 1.0201)$，割线 PQ_2 的斜率为

$$\tan\beta_2 = \frac{1.0201-1}{1.01-1} = 2.01$$

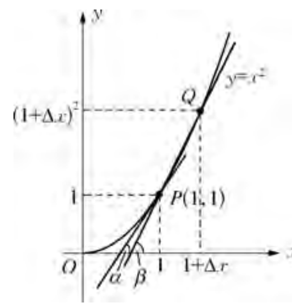


图 4.8

设抛物线上 $P(1, 1)$ 点附近任意点的坐标为 $Q(1+\Delta x, (1+\Delta x)^2)$，(见图 4.8)，则割线 PQ 的斜率为

$$\tan\beta = \frac{(1+\Delta x)^2-1}{1+\Delta x-1} = \frac{2\Delta x+(\Delta x)^2}{\Delta x} = 2+\Delta x$$

注意，和第三章对连续函数的讨论一样，这里的 Δx 表示 x 的“增量”，不是 Δ 与 x 的乘积。

当 Q 点沿抛物线向 P 点逼近，割线 PQ 有一个极限位置，就是抛物线过 P 点的切线，其斜率为

$$\tan\alpha = \lim_{\Delta x\to 0}(2+\Delta x) = 2$$

这个方法可以推广到抛物线上任一点 $P(a, a^2)$。同样，割线 PQ 斜率为

$$\tan\beta = \frac{(a+\Delta x)^2-a^2}{a+\Delta x-a} = \frac{2a\Delta x+(\Delta x)^2}{\Delta x} = 2a+\Delta x$$

抛物线过 P 点切线的斜率为

$$\tan\alpha = \lim_{\Delta x\to 0}(2a+\Delta x) = 2a$$

例如过(2, 4)点的切线斜率为 $\tan\alpha = 2\times 2 = 4$，与【例 4-1】的结果相同。

前面提到，阿波罗尼奥斯在求抛物线曲线时用到一个抛物线的性质，现在可以方便地验证。

回到图 4.1，设抛物线方程为 $y = x^2$，在抛物线上任一点 $P(a, a^2)$，在这点的切线斜率为 $y'(a) = 2a$，切线方程为 $y-a^2 = 2a(x-a)$，令 $x=0$，得 $y = -a^2$，所以 R 点的坐标为 $(0, -a^2)$，而 Q 点坐标为 $(0, a^2)$，可见 R，Q 关于 O 点对称。

【例 4-2】 设“立方抛物线”为 $y = x^3$，求过(1, 1)，(2, 8)点切线的斜率。

仿照前述求抛物线切线斜率的方法，曲线上任取一点 $P(a, a^3)$，以及相邻一点 $Q(a+\Delta x, (a+\Delta x)^3)$，割线 PQ 的斜率为

$$\tan\beta = \frac{(a+\Delta x)^3 - a^3}{a+\Delta x - a} = \frac{3a^2\Delta x + 3a(\Delta x)^2 + (\Delta x)^3}{\Delta x}$$
$$= 3a^2 + 3a\Delta x + (\Delta x)^2$$

令 $\Delta x \to 0$，则切线斜率为

$$\tan\alpha = \lim_{\Delta x \to 0}(3a^2 + 3a\Delta x + (\Delta x)^2) = 3a^2$$

当 $a = 1$，切线斜率为 3；当 $a = 2$，切线斜率为 12。

请读者证明，过 $y = x^3$ 上任一点 $P(a, a^3)$，过这点的切线截距(即与 y 轴交点的纵坐标)为 $-2a^3$。

4.2 函数的导数

4.2.1 导数的概念

回顾上面求抛物线切线的过程，我们发现，$y = x^2$ 在 $x = a$ 处的斜率为 $2a$。仍用 x 表示 a，那么斜率就是 $2x$；同理，$y = x^3$ 在 x 处的斜率是 $3x^2$。这样，$y = x^2$ 导出了一个新的函数 $y = 2x$；同样，$y = x^3$ 也导出了一个新的函数 $y = 3x^2$，都表示原来函数曲线的切线斜率。数学上就抽象出一个新的概念：导数，这也是微积分中第一个重要概念。

定义 4.1 设函数 $y = f(x)$ 在点 $x = a$ 的邻域有定义，当自变量在 $x = a$ 处有增量 Δx，函数有增量 $\Delta y = f(a+\Delta x) - f(a)$。当 $\Delta x \to 0$，若增量的比值有极限，即

$$\lim_{\Delta x \to 0}\frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0}\frac{f(a+\Delta x) - f(a)}{\Delta x} \tag{4.8}$$

存在，则称此极限值为函数 $y=f(x)$ 在 $x=a$ 处的**“导数”**，记为 $f'(a)$。这时称函数 $f(x)$ 在 a 点可导。若上述极限不存在，则称函数在 a 点不可导。

仔细分析，导数的定义有“三步曲”：“微”→“商”→“精”。

第一步 “微”：自变量取微小增量 Δx（注意 Δx 可正可负，不要从“增量”望文生义而认为 Δx 一定为正），从而得到函数的增量 Δy（同样可正可负可为零）；

第二步 “商”：作这两个增量之商 $\dfrac{\Delta y}{\Delta x}$；

第三步 “精”：求极限 $\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}$。因而导数 $f'(a)$ 也记为 $\left.\dfrac{\mathrm{d}y}{\mathrm{d}x}\right|_{x=a}$，或称为**“微商”**。

由于 a 的任意性，就用 x 表示，那么 $f'(a)$ 就成为 $f'(x)$，相当于由 $f(x)$ 导出了一个新的函数，我们把这个新的函数称为**“原函数”**的**“导函数”**，有时也简称为“导数”。例如：$y=2x$ 是 $y=x^2$ 的导函数；$y=3x^2$ 是 $y=x^3$ 的导函数；等等。

从导数定义的三步曲来看，第一、二步是“近似”，即局部以直代曲，以不变代变；第三步是“精确”，通过求极限达到。

注意，如果一个函数在一点可导，那么在这一点一定连续，因为极限 $\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}$ 存在，意味着当 $\Delta x\to 0$ 时一定有 $\Delta y\to 0$（请读者思考：为什么？）。反过来，如果函数在一点连续，并不能保证在这点可导。比如一段折线的尖点，折线在这点连续，却不存在切线，即不可导。

数学家还发现了不少处处连续却处处不可导的曲线。例如第二章中介绍过的“柯克雪花曲线”，就处处连续却处处不可导，这里就不深入探讨了。


柯西 1789—1857

导数的数学符号是“$'$”，由法国数学家柯西（Augustin Louis Cauchy，1789-1857）提出。柯西在数学领域有很高的建树和造诣。很多数学的定理和公式以他命名。他对于微积分的发展有重大贡献。

4.2.2 常数、线性函数和幂函数的导数

(1) 常数的导数为零,即

$$(C)'=0 \tag{4.9}$$

因为 $y=C$ 的图形是水平直线,每一点的切线和这条水平线都重合,自然其斜率为零。如果按照导数的定义,第一步:任取 x 和 Δx, $\Delta y\equiv 0$;第二步:$\dfrac{\Delta y}{\Delta x}\equiv 0$;第三步:$\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}\equiv 0$,结论相同。

(2) 线性函数的导数为常数,即

$$(kx+b)'=k \tag{4.10}$$

线性函数 $y=kx+b$ 表示一条斜率为 k、截距为 b 的直线,上面任一点的切线都与此直线重合,因而其斜率都是 k。按导数定义,第二步中,$\dfrac{\Delta y}{\Delta x}\equiv k$,第三步取极限,极限值仍是 k,结论当然相同。

(3) 幂函数的导数是降一次幂的幂函数,即

$$(x^m)'=mx^{m-1} \tag{4.11}$$

当 $m=0$, $y=x^0=1$,是常数,根据 4.9 式,其导数为零,满足 4.11 式;

当 $m=1$, $y=x$,是线性函数,根据 4.10 式,$(x)'=1$,满足 4.11 式;

当 $m=2$, $y=x^2$,图形是抛物线,由前面的讨论,$(x^2)'=2x$,满足 4.11 式;

当 $m=3$, $y=x^3$,由【例 4-2】的讨论,$(x^3)'=3x^2$,也满足 4.11 式。

当 m 为任意正整数,4.11 式都成立,其证明并不难,要用到代数中的二项式展开公式。

当 m 为任意负整数、任意有理数,以至任意实数,4.11 式都成立。下面按导数定义,推两个特例。

(i) 反比函数的导数

当 $m=-1$, $y=x^{-1}=\dfrac{1}{x}$

$$\frac{\Delta y}{\Delta x}=\frac{\dfrac{1}{x+\Delta x}-\dfrac{1}{x}}{\Delta x}=\frac{x-(x+\Delta x)}{\Delta x\cdot x(x+\Delta x)}=\frac{-1}{x(x+\Delta x)}$$

$\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}=\lim\limits_{\Delta x\to 0}\left(-\dfrac{1}{x(x+\Delta x)}\right)=-\dfrac{1}{x^2}=-x^2$，满足 4.11 式，或另写为

$$\left(\frac{1}{x}\right)'=-\frac{1}{x^2} \tag{4.12}$$

(ii) 平方根函数的导数

当 $m=\dfrac{1}{2}$，$y=x^{\frac{1}{2}}=\sqrt{x}$

$$\frac{\Delta y}{\Delta x}=\frac{\sqrt{x+\Delta x}-\sqrt{x}}{\Delta x}=\frac{(x+\Delta x)-x}{\Delta x(\sqrt{x+\Delta x}+\sqrt{x})}=\frac{1}{\sqrt{x+\Delta x}+\sqrt{x}}$$

$\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}=\lim\limits_{\Delta x\to 0}\dfrac{1}{\sqrt{x+\Delta x}+\sqrt{x}}=\dfrac{1}{2\sqrt{x}}=\dfrac{1}{2}x^{-\frac{1}{2}}$，满足 4.11 式，或另写为

$$(\sqrt{x})'=\frac{1}{2\sqrt{x}} \tag{4.13}$$

4.2.3 正弦、余弦函数的导数

(1) 正弦函数的导数是余弦函数，即

$$(\sin x)'=\cos x \tag{4.14}$$

按导数定义，

$$\frac{\Delta y}{\Delta x}=\frac{\sin(x+\Delta x)-\sin x}{\Delta x}=\frac{2\sin\dfrac{\Delta x}{2}\cos\dfrac{2x+\Delta x}{2}}{\Delta x}=\frac{\sin\dfrac{\Delta x}{2}}{\dfrac{\Delta x}{2}}\cdot\cos\frac{2x+\Delta x}{2}$$

上面的推导中，用到了高中三角中的“和差化积”公式：

$$\sin A-\sin B=2\sin\frac{A-B}{2}\cos\frac{A+B}{2}$$

利用第三章介绍的极限公式 3.46：$\lim\limits_{x\to 0}\dfrac{\sin x}{x}=1$，就得到

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}}\lim_{\Delta x\to 0}\cos\left(x+\frac{\Delta x}{2}\right)=1\cdot\cos x=\cos x$$

注意，上面推导过程中，用到了数学中最常用的方法之一——“换元法”。为了应用 3.46 式，引进新的变量 u（即“元”）：设 $u=\frac{\Delta x}{2}$，

$$x\to 0\Rightarrow u\to 0,\ \lim_{\Delta x\to 0}\frac{\sin\frac{\Delta x}{2}}{\frac{\Delta x}{2}}=\lim_{u\to 0}\frac{\sin u}{u}=1$$

同样方法，可以得到

（2）余弦函数的导数是负的正弦函数，即

$$(\cos x)'=-\sin x \tag{4.15}$$

读者可自行证明。注意，4.15 式等号右端是负号；而 4.14 式等号右端是正号。

4.2.4 指数函数与对数函数的导数

（1）以 e 为底的指数函数，其导数就是本身，即

$$(e^x)'=e^x \tag{4.16}$$

（2）自然对数函数的导数是倒数函数，即

$$(\ln x)'=\frac{1}{x} \tag{4.17}$$

按导数定义证明上面两式，难在求极限这一步，要用到第三章介绍的一个极限公式 3.47：$\lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^n=e$，经过“换元”：$u=\frac{1}{n}$，得到和上式等价的另一形式

$$\lim_{u\to 0}(1+u)^{\frac{1}{u}}=e \tag{4.18}$$

下面先证明相对容易一点的 4.17 式。

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{\ln(x+\Delta x)-\ln x}{\Delta x}=\lim_{\Delta x\to 0}\frac{1}{\Delta x}\ln\frac{x+\Delta x}{x}=\lim_{\Delta x\to 0}\ln\left(1+\frac{\Delta x}{x}\right)^{\frac{1}{\Delta x}}$$

为了应用 4.18 式，再次利用“换元”，设 $u=\frac{\Delta x}{x}$，上式变为

$$\lim_{u\to 0}\ln(1+u)^{\frac{1}{ux}}=(\lim_{u\to 0}\ln(1+u)^{\frac{1}{u}})^{\frac{1}{x}}=\frac{1}{x}\ln(\lim_{u\to 0}(1+u)^{\frac{1}{u}})=\frac{1}{x}\ln e=\frac{1}{x}$$

为了证明 4.16 式，关键是求极限 $\lim\limits_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim\limits_{\Delta x\to 0}\frac{e^{x+\Delta x}-e^{x}}{\Delta x}$。也可以用换元法，再应用 4.18 式而求出极限(有兴趣的读者不妨一试)。下面我们介绍反函数求导法之后，可以通过 4.17 式，更方便地得到 4.16 式。

对于任意正实数 a 为底 $(a\neq 1)$ 的指数函数及对数函数的导数分别为

$$(a^{x})'=a^{x}\ln a \tag{4.19}$$

$$(\log_{a}x)'=\frac{1}{x\ln a} \tag{4.20}$$

从上面的公式推导可见，求导数的难点在于第三步求极限。下面介绍求导的运算法则，可以大大简化对许多函数的求导。

4.3 导数的运算法则

4.3.1 和差的导数

$$(f(x)\pm g(x))'=f'(x)\pm g'(x) \tag{4.21}$$

按导数定义的“三步曲”，读者不妨自行证明。

4.3.2 乘积的导数

$$(f(x)\cdot g(x))'=f'(x)\cdot g(x)+f(x)\cdot g'(x) \tag{4.22}$$

注意，与“和差的导数是导数的和差”不同，“乘积的导数”并不等于“导数的乘积”！数学的结果不能凭猜测，只能遵循严格的逻辑推理，也就是必须通过数学证明。

按导数定义，设 $y=f(x)\cdot g(x)$

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)g(x+\Delta x)-f(x)g(x)}{\Delta x}$$

下面采用数学中常用的“加减相同项”的办法，计算上面的极限。

$$\begin{aligned}\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}&=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)g(x+\Delta x)-f(x)g(x+\Delta x)+f(x)g(x+\Delta x)-f(x)g(x)}{\Delta x}\\&=\lim_{\Delta x\to 0}\left(\frac{f(x+\Delta x)-f(x)}{\Delta x}g(x+\Delta x)+f(x)\frac{g(x+\Delta x)-g(x)}{\Delta x}\right)\\&=f'(x)g(x)+f(x)g'(x)\end{aligned}$$

4.22 式有个特例，当 $g(x)=k$（常数），则

$$(kf(x))'=kf'(x) \tag{4.23}$$

4.3.3 商的导数

$$\left(\frac{f(x)}{g(x)}\right)'=\frac{f'(x)g(x)-f(x)g'(x)}{(g(x))^2} \tag{4.24}$$

证明方法类似，有兴趣的读者可自行完成。4.24 式也有一个特例，当分子 $f(x)=1$，则

$$\left(\frac{1}{g(x)}\right)'=-\frac{g'(x)}{(g(x))^2} \tag{4.25}$$

利用上述运算法则，可以求出更多函数的导数。

【例 4-3】 求下列函数的导数

（1）$y=a_0+a_1x+a_2x^2+\cdots+a_nx^n$（$n$ 次多项式，其中系数 a_0，a_1，…，a_n 都是常数）

（2）$y=\tan x$

（3）$y=e^x\sin x$

解：（1）利用 4.21，4.23 和 4.9 及 4.11，

$$\begin{aligned}y'&=(a_0)'+(a_1x)'+(a_2x^2)'+\cdots+(a_nx^n)'\\&=0+a_1(x)'+a_2(x^2)'+\cdots+a_n(x^n)'\\&=a_1+2a_2x+\cdots+na_nx^{n-1}\end{aligned}$$

可见多项式的导数还是多项式，但是降低一次。

（2）$y=\tan x=\dfrac{\sin x}{\cos x}$，利用 4.24 和 4.14、4.15，

$$y'=\left(\frac{\sin x}{\cos x}\right)'=\frac{(\sin x)'\cos x-\sin x(\cos x)'}{\cos^2 x}$$

$$=\frac{\cos^2 x+\sin^2 x}{\cos^2 x}=\frac{1}{\cos^2 x}=\sec^2 x$$

即

$$(\tan x)'=\sec^2 x \tag{4.26}$$

类似可得

$$(\cot x)'=-\csc^2 x \tag{4.27}$$

（3）$y'=(e^x)'\sin x+e^x(\sin x)'$

$=e^x\sin x+e^x\cos x=e^x(\sin x+\cos x)$

4.3.4　反函数的导数

我们知道，从指数函数 $y=e^x$ 可以解出其反函数为对数函数 $x=\ln y$，由于指数函数的连续性，当 $\Delta x\to 0$ 时 $\Delta y\to 0$，且 $\Delta y\neq 0$，于是

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\frac{1}{\lim\limits_{\Delta y\to 0}\dfrac{\Delta y}{\Delta x}}=\frac{1}{(\ln y)'}=\frac{1}{\dfrac{1}{y}}=y=e^x$$

这就是 4.16 式，比直接按定义求导数要简单得多。

由此可推广到一般的反函数求导法则。

如果从 $y=f(x)$ 可以解出 $x=g(y)$，后者称为 $y=f(x)$ 的“反函数”。

由于 $\dfrac{\Delta y}{\Delta x}\cdot\dfrac{\Delta x}{\Delta y}=1$，而 $\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}=f'(x)$，$\lim\limits_{\Delta x\to 0}\dfrac{\Delta x}{\Delta y}=g'(y)$，所以

$$f'(x)\cdot g'(y)=1 \tag{4.28}$$

【例 4-4】　证明下列反三角函数的导数

（1）
$$(\sin^{-1}x)'=\frac{1}{\sqrt{1-x^2}} \tag{4.29}$$

$$(\cos^{-1}x)'=-\frac{1}{\sqrt{1-x^2}} \tag{4.30}$$

(2)
$$(\tan^{-1}x)' = \frac{1}{1+x^2} \tag{4.31}$$

$$(\cot^{-1}x)' = -\frac{1}{1+x^2} \tag{4.32}$$

证明:(1) 分别利用 4.14、4.15 和 4.28

$$(\sin^{-1}x)' = \frac{1}{(\sin y)'} = \frac{1}{\cos y} = \frac{1}{\cos(\sin^{-1}x)} = \frac{1}{\sqrt{1-x^2}}$$

$$(\cos^{-1}x)' = \frac{1}{(\cos y)'} = -\frac{1}{\sin y} = -\frac{1}{\sin(\cos^{-1}x)} = -\frac{1}{\sqrt{1-x^2}}$$

(2) 同理,分别利用 4.26、4.27 和 4.28

$$(\tan^{-1}x)' = \frac{1}{(\tan y)'} = \frac{1}{\sec^2 y} = \frac{1}{\sec^2(\tan^{-1}x)} = \frac{1}{1+x^2}$$

$$(\cot^{-1}x)' = \frac{1}{(\cot y)'} = -\frac{1}{\csc^2 y} = -\frac{1}{\csc^2(\cot^{-1}x)} = -\frac{1}{1+x^2}$$

可见,反三角函数的导数既不再是反三角函数,也不是三角函数,而是分式有理函数或分式无理函数。

4.3.5 复合函数的导数—链导法

如果要求 $y=\sin 2x$ 的导数,是否能根据上面的公式?

同学甲:根据 4.14 式,$(\sin x)'=\cos x$,自然有 $(\sin 2x)'=\cos 2x$;

同学乙:由初等三角公式,$\sin 2x = 2\sin x\cos x$,右端是两个函数的乘积,按 4.22 式,$(\sin 2x)'=(2\sin x\cos x)'=2((\sin x)'\cos x+\sin x(\cos x)')$,再根据 4.14 和 4.15,就得到 $(\sin 2x)'=2(\cos^2 x-\sin^2 x)=2\cos 2x$!

后一个结果是前一个结果的两倍!问题出在何处?首先我们要复习一个“基本初等函数”的概念:

(A) 常数函数 $y=C$

(B) 幂函数 $y=x^a$

(C) 指数函数 $y=a^x$

(D) 对数函数 $y = \log_a x$

(E) 三角函数 $y = \sin x$，$y = \cos x$，$y = \tan x$，$y = \cot x$

(F) 反三角函数 $y = \sin^{-1} x$，$y = \cos^{-1} x$，$y = \tan^{-1} x$，$y = \cot^{-1} x$

这五类函数称为“基本初等函数”[4-8]。

由于 $y = \sin 2x$ 不是一个基本初等函数，而是由 $y = \sin u$ 和 $u = 2x$ 的复合函数，同学甲套用 4.14，就出了问题。复合函数的导数如何求？

设 $y = f(\varphi(x))$ 是由函数 $y = f(u)$ 和 $u = \varphi(x)$ 复合而成的函数，并设 $u = \varphi(x)$ 在点 x 处可导，$y = f(u)$ 在对应点 $u = \varphi(x)$ 处也可导，则 $y = f(\varphi(x))$ 对 x 可导，且

$$y' = (f(\varphi(x)))' = f'(u)\varphi'(x) \tag{4.33}$$

再回到上面的例子，设 $u = 2x$，则 $(\sin 2x)' = (\sin u)'_u (2x)'_x = \cos u \cdot 2 = 2\cos 2x$，可见同学乙的结果是正确的。

4.33 的证明并不很难。这里我们大体说一下思路，

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u} \cdot \frac{\Delta u}{\Delta x} \Rightarrow \lim_{\Delta u \to 0} \frac{\Delta y}{\Delta u} \cdot \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = y'_u(u) \cdot \varphi'(x)$$

这并不是严格证明，比如当 $\Delta x \to 0$ 时万一 $\Delta u = 0$ 怎么办？介绍了微分概念后我们将给出一个严格的证明。

一个函数可能由 3 个或更多基本初等函数复合而成，例如 $y = f(g(h(x)))$ 可以分解为 $y = f(u)$，$u = g(v)$，$v = h(x)$，$y' = (f(g(h(x))))' = f'_u(u) g'_v(v) h'(x)$。

复合函数求导相当于把各个基本初等函数的导数像链条一样接起来，所以称为“链导法”。

【例 4-5】 求下列函数导数

(1) $y = (2x + 1)^{100}$

(2) $y = \sqrt{\tan 3x}$

(3) $y = \ln(x + e^{2x})$

解：(1) $y = u^{100}$，$u = 2x + 1$，故

$$y' = (u^{100})'_u (2x + 1)' = 100u^{99} \cdot 2 = 200(2x + 1)^{99}$$

（2）$y=\sqrt{u}$，$u=\tan v$，$v=3x$，故

$$y'=(\sqrt{u})'_u(\tan v)'_v(3x)'=\frac{1}{2\sqrt{u}}\cdot\sec^2 v\cdot 3=\frac{3}{2\sqrt{\tan 3x}}\sec^2 3x$$

（3）$y'=\dfrac{1}{x+e^{2x}}(1+2e^{2x})$（为什么？）

能把一个函数准确分解成若干基本初等函数的复合，这点特别重要。在初学阶段，尽可能把中间变量写出来，不易遗漏；熟练之后，可以把中间变量略去。上例的第(3)小题的函数由复合与四则运算混合构成，特别要仔细。

由于初等函数由基本初等函数经过有限次四则运算和复合而成[4-8]，根据上面的讨论，我们得到一个重要的结论：**任意初等函数在定义域内都可导**。

4.4 变速运动的速度和加速度

4.4.1 历史起源

研究变速运动的速度以及加速度，也是微积分的起源之一。

人们对等速直线运动，也称匀速直线运动早有研究：

$$V=\frac{S}{t}\text{（其中 }V,\ S,\ t\text{ 分别为速度、路程和时间）}\tag{4.34}$$

但是直至16世纪才对变速运动引起重视。

相传1590年，出生在比萨城的意大利物理学家伽利略，曾在比萨斜塔上做自由落体实验，将两个重量不同的球体从相同的高度同时扔下，结果两个铅球几乎同时落地，由此发现了自由落体定律：物体下落的路程与时间的平方成正比，且与其质量无关。即：

$$S=kt^2\tag{4.35}$$

这就推翻了亚里士多德认为较重物体先到达地面，落体速度同它的质量成正比的观点。伽利略是划时代的科学巨匠，也是实验科学的重要奠基人。[4-9]

如果初始时刻，再加一个向下的初速度 V_0，由运动的叠加，下落路程为

$$S = V_0 t + kt^2 \tag{4.36}$$

对于自由落体运动，瞬时速度为多少？

随着工业革命的发展，生产实践中也提出了变速直线运动问题。蒸汽机的发明与使用，具有划时代的意义。1688 年，法国物理学家德尼斯·帕潘（Danis Papin，1647-1712）用圆筒和活塞制造出第一台简单蒸汽机。1782 年，英国的仪器修理工瓦特（James Watt，1734-1819）发明了联协式蒸汽机，使得蒸汽机大规模用于工业。

瓦特从小喜欢独自沉思默想。曾有过一个广为人知的传说：有一天，小瓦特在家里看见一壶水开了，蒸汽把壶盖冲得噗噗地跳。这种常人司空见惯的现象却引起了他极浓厚的兴趣。他目不转睛地凝视那跳动的壶盖和冒出的蒸汽，苦思冥想其中的奥秘。正是这种好奇心和寻根问底的精神，后来引导他去努力探索世界的种种奥秘，攀登科学的高峰。

瓦特和他发明的蒸汽机

瓦特的名言是："青春的光辉，理想的钥匙，生命的意义，乃至人类的生存、发展……全包含在这两个字之中……奋斗！只有奋斗，才能治愈过去的创伤；只有奋斗，才是我们民族的希望和光明所在。""与其用华丽的外衣装饰自己，不如用知识武装自己。"后人为了纪念他，将功率的计量单位称为"瓦特"，用符号"W"表示。[4-10]

蒸汽机的核心是一个滑块连杆机构，如图 4.9 所示。

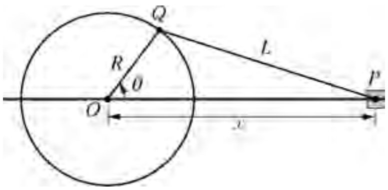


图 4.9

当水蒸汽推动活塞带动滑块 P，活塞水平运动通过连杆 PQ 带动飞轮 O 转动。设连杆长度为 L，飞轮半径为 R，OP 的距离为 x，飞轮曲柄 OQ 与 OP 的夹角为 θ，则

$$x = R\cos\theta + \sqrt{L^2 - R^2\sin^2\theta} \tag{4.37}$$

开普勒 1571-1630

滑块连杆机构也可以反过来，由曲柄 OQ 的转动转化为滑块 P 的水平往复运动。如果曲柄等速转动，其角速度为常数 ω，则 $\theta = \omega t$，代入 4.37 式，就有

$$x = R\cos\omega t + \sqrt{L^2 - R^2\sin^2\omega t} \tag{4.38}$$

现在要问，滑块 P 的水平速度是多少？

此外，人们还研究各种曲线运动。比如德国天文学家开普勒(Johannes Kepler，1571-1630)发现地球在以太阳为焦点的椭圆轨道上运动，如图 4.10 所示；又如炮弹在抛物线轨道上运动，如图 4.5 所示等等。人们要问，作曲线运动的物体(如行星、炮弹)的瞬时速度为多少？

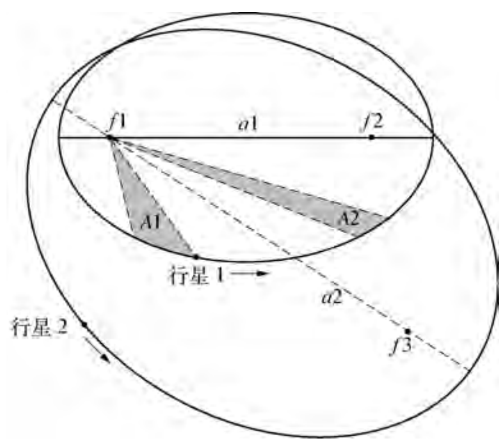


图 4.10

4.4.2 变速运动的瞬时速度

对于等速运动而言，如果时间从 t 到 $t+\Delta t$，路程从 s 到 $s+\Delta s$，不论 Δt 多

大，平均速度都是常数，即不变的速度

$$\bar{v}=\frac{\Delta s}{\Delta t}=v \tag{4.39}$$

但是对于变速运动来说，情况就不同。

【例 4-6】 求 $s=3t^2$ 分别在 $t=1$ 到 $t=2$ 和 $t=2$ 到 $t=3$ 的平均速度。

解： 当 $t=1$，$\Delta t=2-1=1$，$\Delta s=3\cdot 2^2-3\cdot 1^2=9$，$\bar{v}=\frac{\Delta s}{\Delta t}=9$；而当 $t=2$，$\Delta t=3-2=1$，$\Delta s=3\cdot 3^2-3\cdot 2^2=15$，$\bar{v}=\frac{\Delta s}{\Delta t}=15$。两者不等。

仍然考虑 $t=1$，但减小时间间隔，分别取 $\Delta t=0.5$，0.2，0.1，0.01，相应的平均速度为

$$\Delta t=0.5,\ \Delta s=3\cdot(1.5)^2-3\cdot 1^2=3.75,\ \bar{v}=\frac{3.75}{0.5}=7.5;$$

$$\Delta t=0.2,\ \Delta s=3\cdot(1.2)^2-3\cdot 1^2=1.32,\bar{v}=\frac{1.32}{0.2}=6.6;$$

$$\Delta t=0.1,\ \Delta s=3\cdot(1.1)^2-3\cdot 1^2=0.63,\bar{v}=\frac{0.63}{0.1}=6.3;$$

$$\Delta t=0.01,\ \Delta s=3\cdot(1.01)^2-3\cdot 1^2=0.0603,\bar{v}=\frac{0.0603}{0.01}=6.03$$

可见随着时间增量的减小，在起点为 $t=1$ 处的平均速度也越来越接近于 6。令 $\Delta t\to 0$，

$$v=\lim_{\Delta t\to 0}\frac{\Delta s}{\Delta t}=\lim_{\Delta t\to 0}\frac{3\,(1+\Delta t)^2-3}{\Delta t}=\lim_{\Delta t\to 0}\frac{6\Delta t+3\,(\Delta t)^2}{\Delta t}=6+3\lim_{\Delta t\to 0}\Delta t=6$$

这就是 $t=1$ 时的瞬时速度。

从任意时刻 t 出发，读者可自行推出瞬时速度为 $v=6t$。

对比导数的定义，变速直线运动的速度就是路程的导数！

这里，我们介绍牛顿对导数的另一表达符号，即用函数字母上方加一点表示这个函数的导数：

$$\dot{x}(t) \Leftrightarrow x'(t) \tag{4.40}$$

至今在物理学中,当自变量为时间 t,人们仍多用字母上加"·"的记号表示导数。

对于自由落体运动,由 4.36 式,下落物体的瞬时速度为

$$V = \dot{S}(t) = (V_0 t + kt^2)' = V_0 + 2kt \tag{4.41}$$

对于滑块连杆机构,由 4.38 式,滑块的瞬时速度为

$$v = \dot{x}(t) = (R\cos\omega t + \sqrt{L^2 - R^2\sin^2\omega t})' = -R\omega\sin\omega t - \frac{R^2\omega\sin\omega t\cos\omega t}{\sqrt{L^2 - R^2\sin^2\omega t}} \tag{4.42}$$

对于炮弹,由 4.5 式,水平方向和竖直方向的瞬时分速度 V_x 和 V_y 分别为

$$\begin{aligned} V_x &= \dot{x} = ((V_0\cos\alpha)t)' = V_0\cos\alpha \\ V_y &= \dot{y} = \left(-\frac{1}{2}gt^2 + (V_0\sin\alpha)t\right)' = V_0\sin\alpha - gt \end{aligned} \tag{4.43}$$

炮弹的瞬时速度为两者的合成。设其瞬时运动方向与 x 轴夹角为 θ,则其大小和方向分别为

$$|V(t)| = \sqrt{V_x^2 + V_y^2} = \sqrt{V_0^2 + g^2t^2 - 2gtV_0\sin\alpha}$$

$$\tan\theta(t) = \frac{V_y}{V_x} = \tan\alpha - \frac{g}{V_0\cos\alpha}t$$

这就是 4.7 式。

4.4.3 加速度与二阶导数

速度的变化率是加速度。也就是说,加速度是速度的导数,即

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} = v'(t) = \dot{v}(t) \tag{4.44}$$

由于速度是路程的导数,所以数学上称加速度是路程的二阶导数。二阶导数也同样有三种表达符号,即

$$a = \frac{\mathrm{d}^2S}{\mathrm{d}t^2} = S''(t) = \ddot{S}(t) \tag{4.45}$$

回到自由落体运动,由 4.41,

$$a = \ddot{S}(t) = \dot{v}(t) = (V_0 + 2kt)' = 2k \tag{4.46}$$

可见自由落体的加速度为常数，物理中称为“重力加速度(gravity acceleration)”，记为 g，则 $k = \frac{g}{2}$。4.36、4.41 和 4.46 分别写为

$$S = V_0 t + \frac{1}{2}gt^2,\ V = V_0 + gt \text{ 和 } a = g \tag{4.47}$$

这正是中学物理中熟知的运动学公式。

牛顿第二定律用导数的形式可以表示为

$$F = \frac{\mathrm{d}(mV)}{\mathrm{d}t} = \frac{\mathrm{d}m}{\mathrm{d}t}V + m\frac{\mathrm{d}V}{\mathrm{d}t} \tag{4.48}$$

这里 F 是外力，m 是质量，V 是速度，t 是时间；mV 是动量；如果质量是常数，则

$$F = m\frac{\mathrm{d}V}{\mathrm{d}t} = m\frac{\mathrm{d}^2 S}{\mathrm{d}t^2} = m\ddot{S} \tag{4.49}$$

中学物理介绍的牛顿第二定律 $F = ma$ 是质量和加速度都为常数时的特例。例如质量为 m 的自由落体在真空中所受的重力为 mg。如果运动物体在流体中下落，那么除重力外，还受到流体阻力，这个阻力为

$$F_d = \frac{1}{2}C_d A \rho V^2 \tag{4.50}$$

其中 F_d 为摩擦阻力(frictional drag)，C_d 为阻尼系数(drag coefficient)，A 为有效截面积，ρ 是流体密度，V 是物体相对于流体的速度。[4-11] 这时物体所受的外力为重力与流体阻力的合力，由牛顿第二定律，得到关系式

$$mg - \frac{1}{2}C_d A \rho V^2 = ma$$

写成导数形式，

$$mg - \frac{1}{2}C_d A \rho (\dot{S})^2 = m\ddot{S} \tag{4.51}$$

或

$$mg - \frac{1}{2}C_d A \rho V^2 = m\dot{V} \tag{4.52}$$

不难想象，物体下落速度越来越小。如果摩擦阻力足够大，流体足够深，到一定时刻，重力与阻力大小相等，即 $mg = \frac{1}{2}C_d A\rho V^2$，这时的速度称为"终极速度(terminal velocity)"，记为 V_T

$$V_T = \sqrt{\frac{2mg}{C_d A\rho}} \tag{4.53}$$

当物体达到终极速度后，将保持此速度等速下降。

4.4.4 高阶导数

定义 4.2 如果 $y = f(x)$ 可导，则 $y' = f'(x)$ 称为 $f(x)$ 的"一阶导数"；如果 $f'(x)$ 可导，则 $(f'(x))'$ 称为 $f(x)$ 的"二阶导数"，记为 $y'' = f''(x)$。类似可定义三阶导数 $y''' = (f''(x))' = f'''(x)$，…，$n$ 阶导数 $y^{(n)} = (f^{(n-1)}(x))' = f^{(n)}(x)$。

4.4.5 导数概念在其他学科中的应用

在物理和其他许多学科中，表征一点或一个瞬时的变化快慢，往往可以用导数描述。除前述速度和加速度外，力 $F(t)$ 是动量的瞬时变化率，即动量对时间的导数 $F = (m(t)v(t))'$；功率是功对时间的导数 $P = \dot{W}(t)$；热容是热量对温度的导数 $C = \dot{Q}(T)$；电路中的电流强度是电量对时间的导数 $I = \dot{Q}(t)$；不均匀棒的线密度是质量分布函数的导数 $\rho = m'(x)$；等等[4-11]。导数在经济领域中也有广泛应用，我们将在【选修材料之二】中专门介绍。

4.5 微　　分

4.5.1 微分概念

"微分"是微积分的一个重要概念，和导数密切相关，从历史上看，起源似乎更早。古希腊唯物主义哲学家德谟克利特(Democritus，460-370 B.C.)最早提

出物质由原子构成；他同时精通数学，也提出“数学原子论”的观点，他认为正像物质的原子不可再分一样，“数学的原子”也不可再分。他将他的方法用于几何，首次求出了锥体的体积。他的“数学原子”就有“微分”的雏形。

德氏著作多已失传，但从后人的著作中可以看到他的影响，欧几里得几何中构成线段的“点”、构成平面的“线”，以及构成立体的“面”都有“数学原子”的痕迹。毕达哥拉斯(Pythagoras)学派虽认为物质无限可分，但也提出过“无限小单子论”[0-5]。由于历史的局限，他们对于“数学原子”或“无限小单子”的概念是模糊的。比如“原子”或“单子”有没有长度？古希腊诡辩家芝诺(Zeno)曾质问，如果有长度，无限多个“单子”将构成一条无限长的直线；如果没有长度，那么无限个“单子”也没有长度！这样的质疑，直到微积分诞生之后的一百多年，仍时有可闻。

德谟克利特 460-370 B.C.

什么是微分？在物理和化学学科中，分子是能单独存在、并保持纯物质的化学性质的最小粒子。在数学中，一个变量，一个函数的“最小粒子”是什么？应该说是它的增量。看几个实际例子。

一根细棒，截面积为 A，长度为 x，密度为 ρ，则其质量为 $m(x)=\rho A x$，如果加热后，伸长了 Δx，如图 4.11(a)所示，则质量增加

$$\Delta m=\rho A(x+\Delta x)-\rho A x=\rho A \Delta x=m'(x)\Delta x \tag{4.54}$$

质量增量是长度增量的线性函数，其系数就是质量对长度的导数。

现在考虑一块正方形薄板，厚度为 h，边长为 x，密度为 ρ，则其质量为 $m(x)=\rho h x^2$，同样，均匀加热后，长宽都延伸了 Δx，如图 4.11(b)所示，则质量增加

$$\Delta m=\rho h(x+\Delta x)^2-\rho h x^2=2\rho h x \Delta x+\rho h(\Delta x)^2=m'(x)\Delta x+\rho h(\Delta x)^2 \tag{4.55}$$

对于薄板，质量增量不是长度增量线性函数，但其增量也可分离出 Δx 的线性项，其系数恰是质量函数的导数，余下的项，则是 Δx 的高阶项。当 Δx 很小时可以忽略不计。由线性主要部分项，就抽象出“微分”的概念。

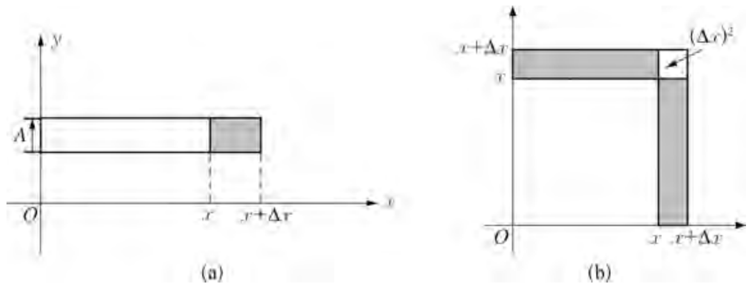


图 4.11

定义 4.3　对于可导函数 $y=f(x)$，若自变量有增量 Δx，则定义 x, y 的微分分别为

$$\mathrm{d}x=\Delta x,\ \mathrm{d}y=f'(x)\Delta x=f'(x)\mathrm{d}x \tag{4.56}$$

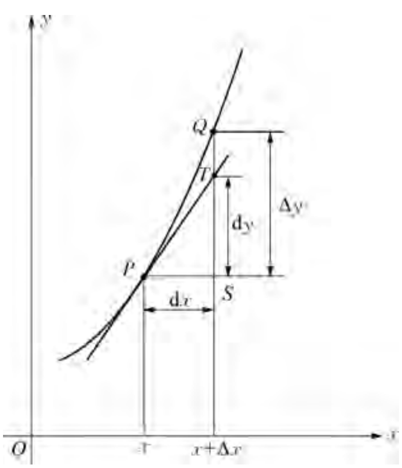


图 4.12

由导数定义，$\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}=f'(x)$，再由极限定义 3.40 式，$\dfrac{\Delta y}{\Delta x}-f'(x)=\alpha$，其中 α 是 Δx 的高阶无穷小。于是 $\Delta y-f'(x)\Delta x=\alpha\Delta x$，代入 4.39，就有

$$\Delta y=\mathrm{d}y+\alpha\Delta x \tag{4.57}$$

所以，微分 $\mathrm{d}y$ 是增量 Δy 的线性主要部分，也就是说，Δy 与 $\mathrm{d}y$ 之差是 Δx 的高阶无穷小。

图 4.12 表示了微分的几何意义。设 P，Q 是函数 $y=f(x)$ 上相邻两点。过 P 点作水平线及过 Q 点做垂直线，交于 S 点。过 P 点作曲线的切线，交 QS 于 T 点。则 $PS=\Delta x=\mathrm{d}x$；$QS=\Delta y$，而 $TS=\mathrm{d}y$。

根据微分的定义，可见一个函数可导必可微，可微必可导。

利用微分的定义，可以证明前面介绍的链导法 4.33：设 $y=f(u)$，$u=\varphi(x)$ 都可导，则

$$\Delta y=f'(u)\Delta u+\varepsilon_1\Delta u=f'(u)(\varphi(x)\Delta x+\varepsilon_2\Delta x)+\varepsilon_1(\varphi'(x)\Delta x+\varepsilon_2\Delta x)$$

$$\frac{\Delta y}{\Delta x}=f'(u)\varphi'(x)+\varepsilon_2 f'(u)+\varepsilon_1\varphi'(x)+\varepsilon_1\varepsilon_2$$

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}(f'(u)\varphi'(x)+\varepsilon_2 f'(u)+\varepsilon_1\varphi'(x)+\varepsilon_1\varepsilon_2)$$

由于当 $\Delta x\to 0$ 时，$\varepsilon_2\to 0$，根据 $f(u)$ 的连续性，$\Delta u\to 0$，从而 $\varepsilon_1\to 0$，所以 $y'(x)=f'(u)\varphi'(x)$，这就是 4.33 式。

根据微分定义 4.59,导数可以表示为微分的商

$$y'(x)=\frac{\mathrm{d}y}{\mathrm{d}x} \tag{4.58}$$

所以导数又称为“微商”。微商的符号首先由莱布尼兹采用。莱布尼兹是位哲学家,着眼于物质的最终微粒,他称这些微粒为“单子(monad)”[4-5],他把由 $\mathrm{d}x$ 和 $\mathrm{d}y$ 为直角边的三角形(图 4.12 中的三角形 PST)称为“微分三角形”[4-11]。

4.5.2 微分在近似计算中的应用

微分本身在增量估计、误差估计中也有应用。

【例 4-7】 链条因自重而下垂,形成的曲线称为“悬链线(Catenary)”,如图 4.13(a)所示。

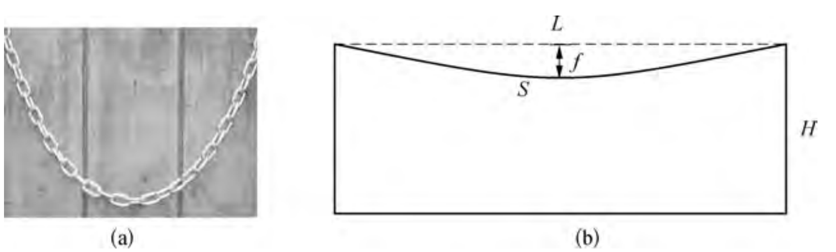


图 4.13

悬索桥、架空电缆的设计过程中都要用到悬链线原理。设一段架空电缆的跨度为 L，电缆长为 S，垂度为 f,电缆塔高 H,如图 4.13(b)所示,我们有经验公式[4-12]

$$S\approx L+\frac{8}{3}\frac{f^2}{L} \tag{4.59}$$

在设计电缆塔时，如何确定其高度 H，如果太高，浪费材料，增加成本；如果太低，影响电缆下的房舍树木。因此必须考虑电缆最低点与地面的距离 $H-f$；同时要考虑当地的冬夏温差。如果冬天的电缆长为 S，垂度为 f，夏天比冬天电缆伸长 ΔS，引起的垂度增量 Δf，那么 $H-(f+\Delta f)$ 不能低于电缆下的物体。考虑如下实际问题：在河宽 $L=200$ m 两岸架设电缆，要求不影响河道中间高为 16 m的轮船通行。设冬天电缆长为 $S=201$ m，夏天电缆伸长 0.5 m，那么电缆塔高 H 至少要多少？

解：先考虑冬天的情况。从 4.62 解出

$$f=\sqrt{\frac{3}{8}L(S-L)} \tag{4.60}$$

解出

$$f=\sqrt{\frac{3}{8}L(S-L)}=\sqrt{\frac{3}{8}\times 200\times(201-200))}=\sqrt{75}\approx 8.66\ \text{m}$$

所以塔高 $H>16+8.66\approx 24.66$ m。

再考虑夏天情况，因为 ΔS 和 S 相比要小得多，所以可以用 $\mathrm{d}S$ 和 $\mathrm{d}f$ 代替 ΔS 和 Δf。由 4.58，$\mathrm{d}S=\frac{16}{3}\frac{f}{L}\mathrm{d}f$，于是

$$\Delta f\approx \mathrm{d}f=\frac{3}{16}\frac{L}{f}\mathrm{d}S=\frac{3L}{16f}\Delta S \tag{4.61}$$

因为 $\Delta S=0.5$ m，由 4.61 式，

$$\Delta f\approx \mathrm{d}f=\frac{3}{16}\times\frac{200}{8.66}\times 0.5=2.16\ \text{m}$$

所以塔高 $H>24.66+2.16\approx 26.82$ m。在河宽 200 m 上的电缆长度仅超过河宽1.5 m，其垂度竟然达到 $8.66+2.16=10.82\approx 11$ m，有两三层楼房的高度，恐怕超过了人们的直觉。这就说明，单凭直觉设计，可能会出事故！这也是数学的魅力所在。

注意，这里我们用到悬链线的近似公式 4.58。悬链线的准确函数关系要用到曲线弧长的积分公式、双曲函数的概念以及微分方程的方法，本书不深入探

讨,有兴趣的读者可参阅[4-8]。

4.6 思考与练习

1. 阅读《古今数学思想》或《微积分概念发展史》或其他文献,了解微积分的源头和发展,思考为什么在 17 世纪产生?

2. 导数的定义是什么?描述了什么现象?

3. 证明:过 $y=x^3$ 上任一点 $P(a, a^3)$,过这点的切线截距(即与 y 轴交点的纵坐标)为 $-2a^3$。

4. 能不能找一个物理量通过导数来定义?

5. 求下列函数的导数

(1) $f(x)=100$　　(2) $f(x)=1-4x^2$

(3) $f(x)=\sqrt{1+3x}$　　(4) $g(x)=\dfrac{1}{x^2}$

(5) $g(x)=\dfrac{1+x+x^2}{8-3x}$　　(6) $g(x)=\sqrt{x}-\dfrac{1}{\sqrt{x}}$

(7) $u(x)=\sin 2x+\cos 3x$　　(8) $u(x)=x^3e^{2x}-\ln(1-x)$

(9) $v(x)=\tan 4x-\cot 4x$　　(10) $v(x)=\sin\sqrt{3\ln(2+x)}$

6. 圆面积为 $A=\pi r^2$,球体积为 $V=\dfrac{4}{3}\pi r^3$,求微分 $\mathrm{d}A$ 和 $\mathrm{d}V$,并解释其几何意义。

7. 在【例 4-10】的数字例子中,用 4.59 式直接计算夏天的垂度,从而计算与冬天垂度之差 Δf。其值与按 4.60 式的计算结果有多少误差?

附录　基本导数公式

1. 四则运算的求导公式

$$(u\pm v)'=u'\pm v'$$

$$(uv)'=u'v+uv',\qquad \text{特别},(Cu)'=Cu'$$

$$\left(\frac{u}{v}\right)'=\frac{u'v-uv'}{v^2},\ (v\neq 0),\qquad \text{特别},\left(\frac{1}{v}\right)'=-\frac{v'}{v^2}$$

2. 反函数的求导公式

若函数 $y=f(x)$ 在点 x 有不等于 0 的导数，且反函数 $x=f^{-1}(y)$ 在点 y 连续，则反函数可导，且

$$\frac{\mathrm{d}x}{\mathrm{d}y}=\frac{1}{\frac{\mathrm{d}y}{\mathrm{d}x}}=\frac{1}{f'(x)}$$

3. 复合函数的求导公式

若 $y=f(u)$，$u=\varphi(x)$ 都有导数，则

$$\frac{\mathrm{d}y}{\mathrm{d}x}=f'(u)\varphi'(x)$$

4. 基本函数的导数公式

$(C)'=0$

$(x^a)'=ax^{a-1}$　　特别，$(\sqrt{x})'=\frac{1}{2\sqrt{x}}$，$\left(\frac{1}{x}\right)'=-\frac{1}{x^2}$

$(e^x)'=e^x$　　$(a^x)'=a^x\ln x\ (a>0,\ a\neq 1)$

$(\ln x)'=\frac{1}{x}$　　$(\log_a x)'=\frac{1}{x\ln a}\ (a>0,\ a\neq 1)$

$(\sin x)'=\cos x$　　$(\cos x)'=-\sin x$

$(\tan x)'=\sec^2 x$　　$(\cot x)'=-\csc^2 x$

$(\sec x)'=\tan x\sec x$　　$(\csc x)'=-\cot x\csc x$

$(\sin^{-1}x)'=\frac{1}{\sqrt{1-x^2}}$　　$(\cos^{-1}x)'=-\frac{1}{\sqrt{1-x^2}}$

$(\tan^{-1}x)'=\frac{1}{1+x^2}$　　$(\cot^{-1}x)'=-\frac{1}{1+x^2}$

参考文献或网站

[4-1] 《马克思恩格斯全集》第 20 卷，人民出版社(1971).

[4-2] 冯·诺依曼(John Von Neumann)：《数学家 (The Mathematician)》，《Works of Mind》，Vol. 1，No. 1，University of Chicago Press(1947).

[4-3] 艾萨克森(Walltter Issacson)：《爱因斯坦：生活和宇宙(Einstein：His life and Universe)》(张卜天译)，湖南科技出版社(2009).

［4-4］ 马克思：《数学手稿》(邓东皋等译)，人民出版社(1975).

［4-5］ 阿西摩夫(Isaac Asimov)：《洞察宇宙的眼睛——望远镜的历史(Eyes on the Universe: A History of the Telesco)》(黄群等译)，科学出版社(1982).

［4-6］ 中央电化教育馆资源中心：《透镜和显微镜》(2003).

［4-7］ 玻恩(M. Born)，沃尔夫(E. Wolf)：《光学原理(Principles of Optics)》(上下册，杨葭荪等译)，电子工业出版社(2005，2006).

［4-8］ 上海交通大学数学系：《微积分(上册)》，上海交通大学出版社(2002).

［4-9］ 库兹涅佐夫：《伽利略传》(陈太先、马世元译)，商务印书馆(2001).

［4-10］ 徐榕 金稼仿 张蔚昕：《瓦特》，华东师范大学出版社(2006).

［4-11］ 卢德馨：《大学物理学》，高等教育出版社(1998).

［4-12］ 华罗庚：《高等数学引论》第一卷第一分册，科学出版社(1964).

第五章

积　分

上一章介绍了微积分的一个基本问题：怎样求曲线的切线，最后归结为求函数的导数。本章介绍微积分的另一个基本问题：怎样求曲边形的面积，并最终归结为积分。

5.1　抛物线弓形面积的计算

人们早就会求由直线构成的图形面积。比如：正方形面积＝边长的平方；长方形面积＝长×宽；三角形面积＝$\frac{1}{2}$×底×高。而多边形可以分割为若干个三角形，也不难求其面积。但是对于曲边形，就很难直接求出面积。

最简单也最常见的曲边形是圆。按照第三章介绍的“割圆术”，古代学者将圆用若干内接三角形来近似，当时没有极限概念，也没有无穷个“三角形”求和的方法。他们只能用尽可能多的有限个三角形来近似圆。

5.1.1　阿基米德的方法

除圆之外，古代学者研究得最多的是抛物线弓形面积的计算。贡献最大的则是古希腊学者阿基米德(Archemides，287-212 B.C.)。

阿基米德是伟大的力学家、数学家。他首先揭示了浮力原理和杠杆原理；他发展了欧几里得几何学，推出关于球与圆柱面积体积等50多个命题；他研究了一类螺线(后人命名为阿基米德螺线)的性质；他也通过圆的内接三角形和外切

三角形的面积逼近圆面积和得到圆周率的近似值；等等。为了纪念他的不朽贡献，今天数学最高荣誉的菲尔茨奖的奖章上就凸印着阿基米德的头像。

阿基米德 287-212 B.C.

菲尔茨奖章

下面概要介绍阿基米德如何巧妙地求抛物线弓形的面积。

图 5.1 为示意图。设 D 为抛物线上任意弦 AC 的中点。过 C 点作抛物线的切线，过 D 点作与抛物线轴平行的直线，与抛物线交于 B 点，与切线交于 E 点；抛物线上的弧段 ABC 和弦 AC 就构成抛物线弓形 ABC。再过 A 点作 ED 的平行线，与切线交于 F 点。作为力学大师，杠杆原理的创始人，阿基米德首先居然是“称出”弓形 ABC 的面积的。

阿氏把弓形看作无数条平行于轴的细条组成。任取一条 OP，延长后与 CF 交于 M；延长 CB，分别与 OM，AF 交于 N 和 K，再在此延长线上取 H，使得 $CK = KH$。

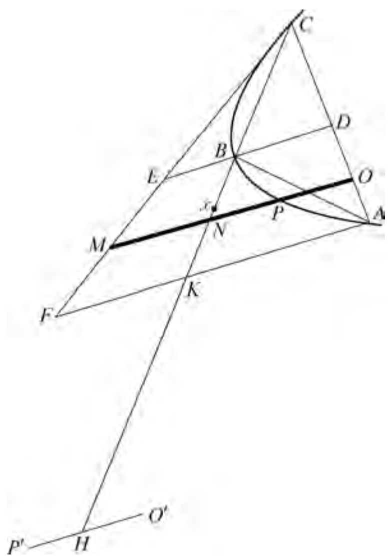


图 5.1

阿氏利用圆锥曲线的性质，$EB = BD$，以及由相似三角形原理，得 $MN = NO$。然后证明了 $\dfrac{HK}{KN} = \dfrac{MO}{OP}$，于是就有 $HK \times OP = KN \times MO$。从物理上看，这个式子意味着把 K 看作支点，把 HK 和 KN 作为杠杆的两条臂，把 OP 条放在 H 处，MO 条放在 N 处，支点两端保持平衡。如果把所有 OP 这样的细条都

放在 H 处，相当于把整个弓形 ABC 放在 H 处；而把所有 MO 这样的细条放在 $\triangle CFA$ 的重心 X 处，支点两端仍然平衡。根据三角形重心的位置，

$$KX=\frac{1}{3}KC=\frac{1}{3}HK$$

因此 $HK\times$ 弓形 $ABC=KX\times\triangle CFA=\frac{1}{3}HK\times\triangle CFA$，于是弓形 $ABC=\frac{1}{3}\triangle CFA$。另一方面，$\triangle ABC=\frac{1}{2}\triangle CKA=\frac{1}{4}\triangle CFA$，所以，

$$\text{弓形 } ABC=\frac{4}{3}\triangle ABC \tag{5.1}$$

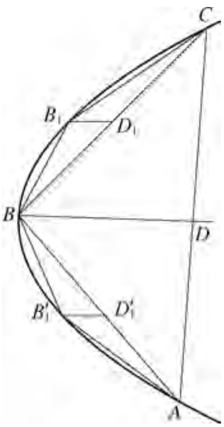


图 5.2

阿基米德没有意识到，他把弓形看作无限细条的累积，正是近两千年后积分的基本思想。阿基米德是欧几里得的学生，他知道物理类比不能算数学证明。他又通过“穷竭法”和“反证法”严格证明了 5.1 式。

在图 5.2 中 $\triangle ABC$ 称为“阿基米德三角形”，即 BD 是弦 AC 的中线，且平行于抛物线轴。下面以 BC 和 BA 为弦，构造“第 2 层”阿基米德三角形，共有两个：$\triangle BB_1C$ 和 $\triangle AB_1'B$。不难证明

$$\triangle BB_1C=\triangle AB'_1B=\frac{1}{8}\triangle ABC$$

于是，

$$\triangle ABC+\triangle BB_1C+\triangle AB'_1B=\left(1+\frac{2}{8}\right)\triangle ABC=\left(1+\frac{1}{4}\right)\triangle ABC$$

继续以 B_1C、BB_1、BB'_1 和 AB'_1 为弦，构造“第 3 层”阿基米德三角形，共有 4 个，每个面积为 $\triangle BB_1C$ 的 $\frac{1}{8}$，这 4 个小三角形面积之和为

$$4\times\frac{1}{8}\triangle BB'_1C=4\times\frac{1}{8}\times\frac{1}{8}\triangle ABC=\frac{1}{16}\triangle ABC$$

于是这 3 层阿基米德三角形的面积和为

$$\left(1+\frac{1}{4}+\frac{1}{4^2}\right)\triangle ABC$$

继续做下去，无数层“阿基米德三角形”的面积和为

$$\left(1+\frac{1}{4}+\frac{1}{4^2}+\cdots+\frac{1}{4^n}+\cdots\right)\triangle ABC$$

这些三角形将“填满”弓形 ABC。利用无穷等比级数求和公式

$$a(1+q+q^2+\cdots+q^n+\cdots)=a\frac{1}{1-q},\ (|q|<1) \tag{5.2}$$

$$\text{弓形 } ABC=\left(1+\frac{1}{4}+\frac{1}{4^2}+\cdots+\frac{1}{4^n}+\cdots\right)\triangle ABC=\frac{1}{1-\frac{1}{4}}\triangle ABC=\frac{4}{3}\triangle ABC$$

这正是 5.1 式。

不过，在阿基米德时代，人们还没有发现无穷等比级数的求和公式 5.2；也没有极限概念，无穷个“阿基米德三角形”是否恰好“填满弓形”，也存疑问。阿氏进一步用反证法，假设弓形 ABC 面积大于或小于三角形 ABC 面积的 $\frac{4}{3}$ 倍，都发生矛盾，于是 5.1 式严格成立。详细过程可参阅《古今数学思想》[0-1]。

5.1.2 和式极限方法

阿基米德的方法，充分显示了他的超人智慧和天才。在赞叹之余，一般人对这类面积问题只能望而生畏，敬而远之。16 世纪之后，生产的进步和社会的发展，提出了许多变化量积累的计算问题，除面积问题外，还有体积、曲线长度、天体引力等问题。人们迫切希望找到一种普遍的方法。

开普勒做了承上启下的工作。1615 年他用拉丁文出版了《Opera omnia（全集）》。[0-5]其中包括阿基米德测体积法，并用来测啤酒桶的容积，还推广到更一般的方法。开普勒有不少创造性思维。他把圆看成由无限个内接小三角形组成。因此圆面积$=\frac{1}{2}\times$半径$\times$圆周长，不仅如此，他进一步把球看成由无限个内接小圆锥组成，因此球体积$=\frac{1}{3}\times$半径$\times$球面面积。

伽利略的学生卡瓦列里（Bonaventura Cavalieri，1598-1647）是另一位作出重要贡献的数学家。他的著作《Geometria indivisibilibus（不可分量几何学）》蕴含着积分的基本思想。他设想一个面由等距离平行直线段组成，一个立体由等

卡瓦列里　1598-1647

距离的平行平面构成；这些元素被称为面积和体积的“不可分量”。他打了一个巧妙的比方，把一个曲面看作一块布，由许多线织成，而一个立体则看作一本书，由许多书页叠成。[0-5]

正是在许多人的“接力”下，形成了用和式极限的方法求曲边形面积的一般方法，即积分。

【例 5-1】 求以抛物线 $y=x^2$ 和 x 轴以及直线 $y=1$ 所夹的“曲边三角形”OAB 面积。如图 5.3 所示。

类似于刘徽的“割圆法”和阿基米德的“杠杆法”，我们把 OAB 分割成 n 个竖直细条，宽度为 $\frac{1}{n}$。每个细条的左端点的横坐标分别为 0，$\frac{1}{n}$，$\frac{2}{n}$，$\frac{3}{n}$，…，$\frac{n-1}{n}$。

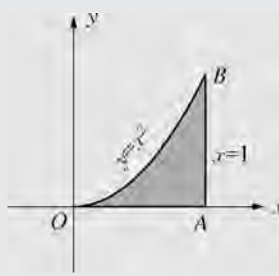


图 5.3

每个细条仍然是曲边形。我们沿用求曲线切线的基本思路“局部以直代曲”，将每个细条看作矩形，其宽仍为 $\frac{1}{n}$，其高不妨就用左端点的高，如图 5.4(a) 所示，即 0，$\left(\frac{1}{n}\right)^2$，$\left(\frac{2}{n}\right)^2$，$\left(\frac{3}{n}\right)^2$，…，$\left(\frac{n-1}{n}\right)^2$。这 n 个细矩形面积之和为

$$\begin{aligned} S_n &= \frac{1}{n}\left(0+\left(\frac{1}{n}\right)^2+\left(\frac{2}{n}\right)^2+\left(\frac{3}{n}\right)^2+\cdots+\left(\frac{n-1}{n}\right)^2\right) \\ &= \frac{1}{n^3}(0+1^2+2^2+3^2+\cdots+(n-1)^2) \end{aligned}$$

根据 2.24 式，$1^2+2^2+3^2+\cdots+n^2=\frac{1}{6}n(n+1)(2n+1)$，那么

$$S_n=\frac{1}{6n^3}(n-1)n(2n-1)=\frac{1}{6n^2}(n-1)(2n-1)=\frac{1}{6}\left(1-\frac{1}{n}\right)\left(2-\frac{1}{n}\right) \tag{5.3}$$

表 5.1 列出了随着 n 增大时 S_n 的数值。

表 5.1

n	4	8	16	50	100	1000
S_n	0.21875	0.27344	0.30273	0.3234	0.32835	0.33283

当 n 无限增大，

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{1}{6}\left(1-\frac{1}{n}\right)\left(2-\frac{1}{n}\right) = \frac{1}{3}$$

于是曲边三角形 OAB 的面积为 $S=\dfrac{1}{3}$。

上面每个细条矩形用左端点的函数值作为高，如图 5.4(a)所示，如果用右端点的函数值作为高，如图 5.4(b)所示，或用中点的函数值作为高，如图 5.4(c)所示，相应的计算式分别为

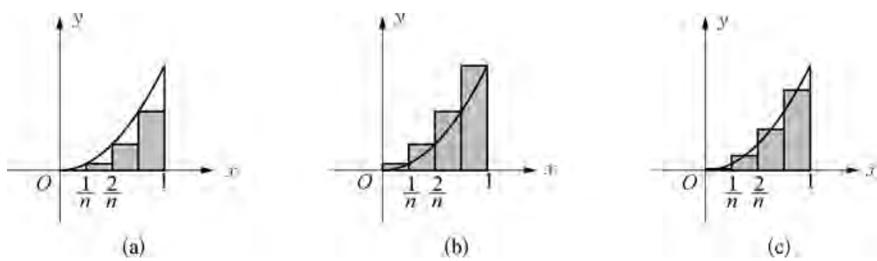


图 5.4

$$\begin{aligned} S_{Rn} &= \frac{1}{n}\left(\left(\frac{1}{n}\right)^2+\left(\frac{2}{n}\right)^2+\left(\frac{3}{n}\right)^2+\cdots+\left(\frac{n-1}{n}\right)^2+n^2\right) \\ &= \frac{1}{n^3}(1^2+2^2+3^2+\cdots+n^2) = \frac{1}{6n^3}n(n+1)(2n+1) \\ &= \frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right) \end{aligned} \tag{5.4}$$

$$S_{Mn} = \frac{1}{n}\left(\left(\frac{1}{2n}\right)^2+\left(\frac{3}{2n}\right)^2+\left(\frac{5}{2n}\right)^2+\cdots+\left(\frac{2n-1}{2n}\right)^2\right)$$

$$= \frac{1}{4n^3}(1^2+3^2+5^2+\cdots+(2n-1)^2) = \frac{1}{4n^3}\frac{1}{3}n(4n^2-1)$$

$$= \frac{1}{12}\left(4-\frac{1}{n^2}\right) \tag{5.5}$$

这里用到了等式 $1^2+3^2+5^2+\cdots+(2n-1)^2=\frac{1}{3}n(4n^2-1)$，请读者验证。

同样，

$$\lim_{n\to\infty}S_{Rn} = \lim_{n\to\infty}\frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right)=\frac{1}{3}$$

$$\lim_{n\to\infty}S_{Mn} = \lim_{n\to\infty}\frac{1}{12}\left(4-\frac{1}{n^2}\right)=\frac{1}{3}$$

从图上直观看，在有限分割的情况下，取细矩形的中点为高比较准确；表 5.1 和表 5.2 的数据也验证了这一点。但从极限的角度，无论取哪一点的高，取极限后，殊途同归，都是 $\frac{1}{3}$，结果不变。由此我们得到“定积分”的概念。

表 5.2

n	4	8	16	50	100	1000
S_{Rn}	0.4687	0.39844	0.36523	0.3434	0.33835	0.33383
S_{Mn}	0.32812	0.33203	0.33301	0.3333	0.33333	0.33333

5.2 定积分

5.2.1 定积分的定义

定义 5.1 设函数 $f(x)$ 在 $a \leqslant x \leqslant b$ 有定义。

第一步：分。在 $[a, b]$ 内任意插入 $n-1$ 个分点：

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b \tag{5.6}$$

记

$$\Delta x_1 = x_1 - x_0,\ \Delta x_2 = x_2 - x_1,\ \cdots,\ \Delta x_n = x_n - x_{n-1}$$

或者简记为

$$\Delta x_i = x_i - x_{i-1} > 0\ (i = 1,\ 2,\ \cdots,\ n) \tag{5.7}$$

这 $n-1$ 个分点不一定是等分,即各子区间 Δx_i 不一定相等。

第二步:粗。在子区间 Δx_i 中任取一点 ξ_i,即 $x_{i-1} \leqslant \xi_i \leqslant x_i$,作乘积 $f(\xi_i)\Delta x_i(i = 1,\ 2,\ \cdots,\ n)$;

第三步:合。求和 $f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \cdots + f(\xi_n)\Delta x_n$,简记为 $\sum\limits_{i=1}^{n} f(\xi_i)\Delta x_i$;

第四步:精。记最大子区间的长度为 D,即 $D = \max\Delta x_i$ $(i = 1,\ 2,\ \cdots,\ n)$。

取极限 $\lim\limits_{D\to 0}\sum\limits_{i=1}^{n} f(\xi_i)\Delta x_i$,如果极限存在,而且无论子区间如何划分,$\xi_i$ 点如何选取,极限值都相同,则称函数 $f(x)$ 在区间$[a,\ b]$上可积,称此极限值为 $f(x)$ 在$[a,\ b]$上的**定积分**(Definite Integral),记作 $\int_a^b f(x)\mathrm{d}x$,即

$$\int_a^b f(x)\mathrm{d}x = \lim_{D\to 0}\sum_{k=1}^{n} f(\xi_k)\Delta x_k \tag{5.8}$$

式中 x 为**积分变量**,$f(x)$ 为**被积函数**,$f(x)\mathrm{d}x$ 为**被积式**,$[a,\ b]$为**积分区间**,a 为**积分下限**,b 为**积分上限**。

这个定义相当长。不妨对照求曲边梯形面积的几何意义来理解,如图 5.5 所示。$\int_a^b f(x)\mathrm{d}x$ 相当于由曲线 $y = f(x)$ 和 x 轴以及二平行直线 $y = a$,$y = b$ 所夹曲边梯形的面积;而 $f(\xi_i)\Delta x_i$ 就是第 i 个细矩形的面积,用它来近似第 i 个小曲边梯形的面积。

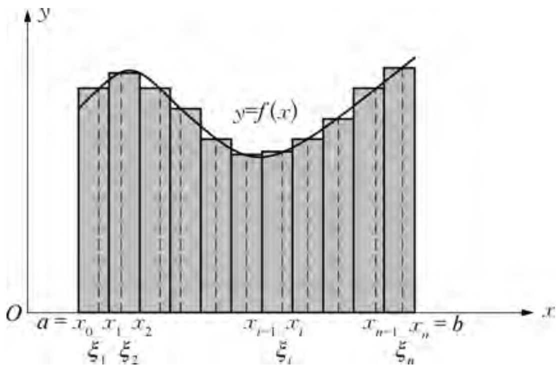


图 5.5

【例 5-1】中曲边三角形 OAB 现在可以用定积分表示为 $\int_0^1 x^2\mathrm{d}x=\frac{1}{3}$。抛物半弓形 OBD 的面积为 $1-\frac{1}{3}=\frac{2}{3}$，而三角形 OBD 的面积为 $\frac{1}{2}$，正好是半个阿基米德三角形。因此这里的抛物线弓形面积为 $\frac{4}{3}$，正好是阿基米德三角形面积的 $\frac{4}{3}$ 倍。

若 $f(x)\geqslant 0$ 曲线在 x 轴上方，$\int_a^b f(x)\mathrm{d}x\geqslant 0$，表示曲边梯形的面积；如果 $f(x)\leqslant 0$，曲线在 x 轴下方，那么 $\int_a^b f(x)\mathrm{d}x\leqslant 0$，相当于曲边梯形面积的负值。如果 $f(x)$ 在积分区间 $[a,\ b]$ 中有正有负，那么 $\int_a^b f(x)\mathrm{d}x$ 在几何上表示曲线 $y=f(x)$ 和 x 轴以及二平行直线 $y=a$，$y=b$ 所夹曲边梯形的面积的代数和。如图 5.6 所示。

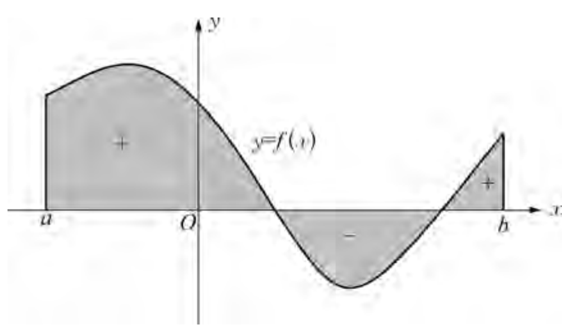


图 5.6

显然，$\int_{-\pi}^{\pi}\sin x\,\mathrm{d}x=0$（为什么?）

如果从 b 向 a 划分子区间，令 $\Delta x_1=x_{n-1}-x_n=b,\cdots,\Delta x_n=x_0-x=a$，显然，

$$\Delta x_i=x_{i-1}-x_i<0\ (i=1,2,\cdots,n) \tag{5.9}$$

按 5.8 式求和式极限，绝对值相同，符号相反，即

$$\int_a^b f(x)\,\mathrm{d}x=-\int_b^a f(x)\,\mathrm{d}x \tag{5.10}$$

注意，并不是任何函数都可积；但是只要 $f(x)$ 在$[a,b]$连续，5.8 式一定存在。

5.2.2　定积分的性质

（1）线性性质　设 $f(x)$ 和 $g(x)$ 在$[a,b]$可积，α 和 β 是给定常数，则

$$\int_a^b(\alpha f(x)+\beta g(x))\,\mathrm{d}x=\alpha\int_a^b f(x)\,\mathrm{d}x+\beta\int_a^b g(x)\,\mathrm{d}x \tag{5.11}$$

（2）单调性　设在$[a,b]$成立 $f(x)\leqslant g(x)$，则

$$\int_a^b f(x)\,\mathrm{d}x\leqslant\int_a^b g(x)\,\mathrm{d}x \tag{5.12}$$

由此得到一个推论：

$$若\ f(x)\geqslant 0，则\int_a^b f(x)\,\mathrm{d}x\geqslant 0 \tag{5.13}$$

（3）区间可加性　设 $y=f(x)$ 在区间 I 上可积。a,b,c 是 I 上任意三点，则

$$\int_a^b f(x)\,\mathrm{d}x=\int_a^c f(x)\,\mathrm{d}x+\int_c^b f(x)\,\mathrm{d}x \tag{5.14}$$

请读者从几何上说明 5.14 式。

（4）积分中值定理　设 $y=f(x)$ 在闭区间$[a,b]$上连续，则至少有一点 $\xi\in[a,b]$ 使得

$$f(\xi)=\frac{1}{b-a}\int_a^b f(x)\,\mathrm{d}x \tag{5.15}$$

我们从几何上说明，如图 5.7 所示，5.15 式可以改写为

$$f(\xi)(b-a)=\int_a^b f(x)\mathrm{d}x\ (a\leqslant\xi\leqslant b) \tag{5.16}$$

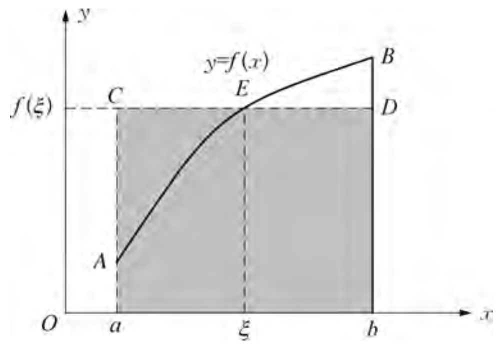


图 5.7

5.16 等式右边的几何意义是曲边梯形 $abBA$，等式左边的几何意义是矩形 $abDC$。只要 $f(x)$ 连续，总可以找到一点 $\xi\in[a, b]$ 使得曲边形 DBE 的面积等于曲边形 ACE 的面积。就好像平整土地，把“坡面”AEB 改成“水平面”CED。

5.3　牛顿-莱布尼兹公式

和式极限方法计算定积分值过于复杂，实际问题中很难算出精确极限值。试看下面的例子。

【例 5-2】 求 $\int_0^1 \sqrt{x}\mathrm{d}x$ 的值。从几何意义看，相当于求如图 5.8 所示中曲边形 OBA 的面积。

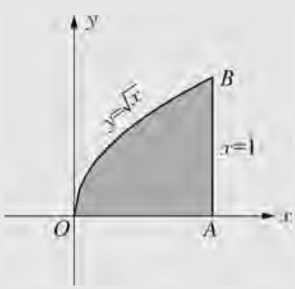


图 5.8

仿照【例 5-1】，将$[0, 1]n$ 等分，采用左矩形公式求和

$$S_n=\frac{1}{n}\left(0+\sqrt{\frac{1}{n}}+\sqrt{\frac{2}{n}}+\sqrt{\frac{3}{n}}+\cdots+\sqrt{\frac{n-1}{n}}\right)$$
$$=\frac{1}{n\sqrt{n}}(1+\sqrt{2}+\sqrt{3}+\cdots+\sqrt{n-1}) \tag{5.17}$$

表 5.3 是部分结果：

表 5.3

n	4	8	16	32
S_n	0.518	0.596	0.632	0.650

实际上，

$$\begin{aligned}\text{曲边形 } OBC \text{ 的面积} &= \text{正方形 } OABC \text{ 的面积} - \text{曲边形 } OAB \text{ 的面积} \\ &= 1 - \frac{1}{3} = \frac{2}{3} = 0.66666\cdots\end{aligned}$$

但是通过 5.17 很难得到准确的结果。如果被积函数更复杂，按 5.8 式求定积分的准确值几乎不可能。是否有更好的方法求定积分呢？牛顿和莱布尼兹独立给出了绝妙的解答。

5.3.1 牛顿-莱布尼兹公式

定义 5.2 如果在区间 I 上，$F'(x) = f(x)$，则称 $F(x)$ 是 $f(x)$ 的一个"**原函数**(original function)"，也称为"**反导数**(anti-derivative)"。

【例 5-3】 求 1、x^2、$\sqrt{x}$、$\sin x$ 和 e^x 的原函数。

解：因 $(x)' = 1$，所以 1 是 x 的一个原函数；

因 $\left(\dfrac{x^3}{3}\right)' = x^2$，所以 $\dfrac{x^3}{3}$ 是 x^2 的一个原函数；

因 $\left(\dfrac{2x\sqrt{x}}{3}\right)' = \left(\dfrac{2}{3}x^{\frac{3}{2}}\right)' = x^{\frac{1}{2}} = \sqrt{x}$，所以 $\dfrac{2x\sqrt{x}}{3}$ 是 $\sqrt{x}$ 的一个原函数；

因 $(\cos x)' = -\sin x$，即 $(-\cos x)' = \sin x$，所以 $-\cos x$ 是 $\sin x$ 的一个原函数；

因 $(e^x)' = e^x$，所以 e^x 是 e^x 的一个原函数；等等。

定理 5.1 （牛顿-莱布尼兹公式）如果 $f(x)$ 在 $[a, b]$ 可积，$F(x)$ 是 $f(x)$ 在 $[a, b]$ 上的一个原函数，则

$$\int_a^b f(x)\mathrm{d}x = F(b) - F(a) = F(x)\mid_a^b \tag{5.18}$$

回到【例 5-1】，

$$\int_0^1 x^2\,\mathrm{d}x = \frac{x^3}{3}\bigg|_0^1 = \frac{1^3}{3} - \frac{0^3}{3} = \frac{1}{3}$$

结果相同，无须求和式极限。

回到【例 5-2】，

$$\int_0^1 \sqrt{x}\mathrm{d}x = \frac{2}{3}x\sqrt{x}\mid_0^1 = \frac{2}{3} - 0 = \frac{2}{3}$$

也无须求和式极限。又如 $\int_0^{\pi} \sin x\,\mathrm{d}x = -\cos x\mid_0^{\pi} = (-\cos\pi) - (-\cos 0) = 2$。

牛顿-莱布尼兹公式是怎样得到的？下面先回到路程速度问题。

5.3.2 通过速度图求路程

对于等速运动，已知路程求速度，是做除法：

$$V = \frac{S}{t} \tag{5.19}$$

而已知速度求路程，则是做乘法：

$$S = Vt \tag{5.20}$$

“速度图”是已知速度求路程的几何方法。当速度为 V 时，物体从 0 时刻到 t 时刻的路程，$S = Vt$ 可以形象地用矩形的面积表示（如图 5.9(a)所示）。

现在进一步研究等加速运动。设加速度为 a，我们知道速度和路程公式分别为

$$V = at,\ S = \frac{1}{2}at^2 \tag{5.21}$$

能不能通过速度图求路程？考察图 5.9(b)：

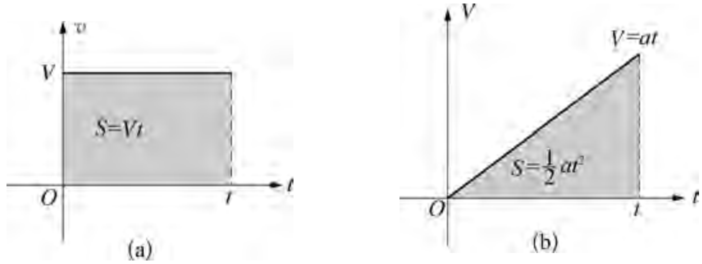


图 5.9

速度 $V=at$ 是斜率为 a 且过原点的直线;物体从 0 时刻到 t 时刻的路程 $S=\dfrac{1}{2}at^2$ 也可以形象地用三角形的面积表示。

如果速度是更一般的函数 $V=V(t)$，从 0 时刻物体从原点出发，到 t 时刻的路程 $S=S(t)$ 是否也可以用速度图中速度曲线以下、横轴之上以及时间为 t 之间的面积表示(图 5.10)? 答案是肯定的。和前面求曲边形面积的推理完全类似，把时间分成许多小区间 Δt，在每一个小区间中近似看成等速运动，在这个小区间中走过的路程是 $\Delta S=V(t)\Delta t$，

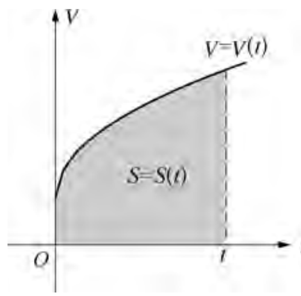


图 5.10

求和得到 $S(t)\approx\sum V(t)\Delta t$，令 $\Delta t\to 0$，总路程 $S(t)$ 的精确值就是图中有阴影的曲边形的面积。用定积分表示，

$$S(t)=\int_0^t V(t)\mathrm{d}t \tag{5.22}$$

注意，5.22 式中积分上限不是常数，而是变数 t，所以这样的积分称为“变上限积分”，其结果是上限 t 的函数 $S(t)$。

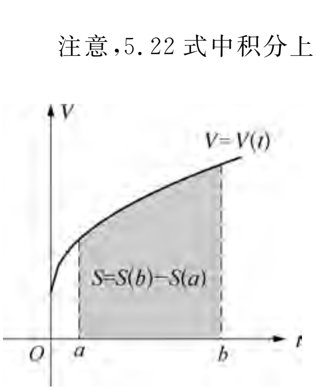


图 5.11

图 5.11 显示，当时间从 a 到 b，路程为 $S=S(b)-S(a)=\int_0^b V(t)\mathrm{d}t-\int_0^a V(t)\mathrm{d}t$，利用定积分区间可加性的性质 5.14，就有

$$S=\int_a^b V(t)\mathrm{d}t=S(b)-S(a) \tag{5.23}$$

5.23 式实际上就是牛顿-莱布尼兹公式 5.18。

在 5.22 式中，$S'(t)=V(t)$，也就是说，变上限积分所表示的函数 $S(t)$ 恰好是被积函数 $V(t)$ 的原函数。对于任意一个函数 $f(x)$ 的变上限积分所表示的函数 $F(x)=\int_a^x f(x)\mathrm{d}x$，（积分下限 a 是常数）是不是被积函数的原函数？也即是否成立 $F'(x)=f(x)$？回答也是肯定的。

5.3.3 牛顿莱布尼兹公式的证明

定理 5.2 （微积分第一基本定理）若 $f(x)$ 在 $[a,b]$ 连续，定义

$$F(x)=\int_a^x f(x)\mathrm{d}x \tag{5.24}$$

则

$$F'(x)=f(x) \tag{5.25}$$

或

$$\mathrm{d}F=f(x)\mathrm{d}x \tag{5.26}$$

证明：根据导数定义 4.1，

第一步，“微”：$\Delta F=F(x+\Delta x)-F(x)=\int_x^{x+\Delta x} f(x)\mathrm{d}x$，如图 5.12 所示。

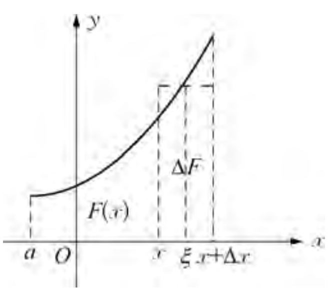


图 5.12

利用积分中值定理 5.16，$\int_x^{x+\Delta x} f(x)\mathrm{d}x=f(\xi)\Delta x$，$(x\leqslant\xi\leqslant x+\Delta x)$

第二步，“商”：$\dfrac{\Delta F}{\Delta x}=f(\xi)$

第三步，“精”：$F'(x)=\lim\limits_{\Delta x\to 0}\dfrac{\Delta F}{\Delta x}=\lim\limits_{\Delta x\to 0}f(\xi)=f(x)$，这就是 5.25 式。说明 $F(x)$ 确实是 $f(x)$ 的原函数。根据微分的定义，$\mathrm{d}F=F'(x)\mathrm{d}x=f(x)\mathrm{d}x$，这就是 5.26 式。

这也证明了牛顿莱布尼兹公式，即 $\int_a^b f(x)\mathrm{d}x=F(b)-F(a)$。

5.4 不定积分与积分法

5.4.1 原函数的一般形式

从前面的讨论可知，对于等速运动，由路程求速度用除法；由速度求路程用乘法。乘法和除法是一对逆运算。对于变速运动，由路程求速度用导数运算；由速度求路程则用积分运算。求积和求导也互为逆运算。按照牛顿-莱布尼兹公式，求定积分不必用和式极限的复杂方法，只要求出被积函数的原函数，然后把下限和上限代入相减就行。问题转化为给定一个被积函数，如何求它的原函数。

给定一个可导函数，它的导函数是唯一的。但是，一个函数如果存在原函数，其原函数却不唯一。例如 $\cos x$ 的原函数是 $\sin x$，由于 $(\sin x+1)'=\cos x$，$(\sin x+2)'=\cos x$，所以 $\sin x+1$，$\sin x+2$ 也是 $\cos x$ 的原函数。事实上，对于任意常数 C，$\sin x+C$ 都是 $\cos x$ 的原函数。可见如果一个函数存在原函数，那么原函数必有无穷多个。幸运的是，这些原函数只差一个任意常数。这就是下面的定理：

定理 5.3 （微积分第二基本定理）如果函数 $f(x)$ 在某区间 I 上有原函数 $F(x)$，则 $F(x)+C$ 也是 $f(x)$ 在此区间 I 上的原函数；而且 $f(x)$ 的任一个原函数都可表示为 $F(x)+C$ 的形式，其中 C 是任意常数。

5.4.2 不定积分

定义 5.3 函数 $f(x)$ 的全体原函数称为 $f(x)$ 的"不定积分"(Infinite Integral)，记为

$$\int f(x)\mathrm{d}x \tag{5.27}$$

和定积分一样，这里 $\int$ 称为"积分号"，x 称为"积分变量"，$f(x)$ 称为"被积函数"，$f(x)\mathrm{d}x$ 称为"被积式"。只要 $f(x)$ 连续，5.27 就有意义。

如果 $F(x)$ 是 $f(x)$ 的一个原函数，由定理 5.3，

$$\int f(x)\mathrm{d}x = F(x) + C \tag{5.28}$$

其中 C 为任意常数，也称为“积分常数”。

不难验证，不定积分有如下性质：

1）线性性质

$$\int(\alpha f(x) + \beta g(x))\mathrm{d}x = \alpha\int f(x)\mathrm{d}x + \beta\int g(x)\mathrm{d}x \tag{5.29}$$

2）微分积分互逆性质

$$\begin{aligned} &\left(\int f(x)\mathrm{d}x\right)' = f(x); \qquad \mathrm{d}\left(\int f(x)\mathrm{d}x\right) = f(x)\mathrm{d}x \\ &\int f'(x)\mathrm{d}x = f(x) + C; \qquad \int \mathrm{d}f(x) = f(x) + C \end{aligned} \tag{5.30}$$

可见，求导数或求微分和求不定积分确实是互逆运算。对一个函数先求积分再求导数等于还原；如果先求导数再求积分，结果也是还原，不过多了一个任意常数。如果不考虑任意常数，则积分号 $\int$ 和求导记号$'$可以互相抵消。

定积分和不定积分都归结为求被积函数的原函数，且都具有线性性质 5.9 和 5.29；但必须注意两者的差别。

定积分 $\int_a^b f(x)\mathrm{d}x$ 是一个数值，所以积分变量 x 换为什么字母都行，例如 $\int_a^b f(t)\mathrm{d}t$，$\int_a^b f(u)\mathrm{d}u$ 都是一样的结果，这样的积分变量称为“哑变量”；但不定积分 $\int f(x)\mathrm{d}x$ 是一个函数加上一个任意常数 $F(x)+C$，积分变量 x 就不能随意更换了。另外，与积分限有关的性质（如 5.14 和 5.15）当然只能适用于定积分。

由于积分（定积分和不定积分）归结为求被积函数的原函数，我们把找出原函数的方法称为“积分法”。常用的积分法有三种：① 查基本不定积分表；② 换元积分法；③ 分部积分法。下面作简单介绍。

5.4.3 基本不定积分表

根据第四章介绍的基本导数公式，立即可以得到下面的基本不定积分

表 5.4：

表 5.4

$\int kf(x)\mathrm{d}x = k\int f(x)\mathrm{d}x$
$\int (f(x) \pm g(x))\mathrm{d}x = \int f(x)\mathrm{d}x \pm \int g(x)\mathrm{d}x$
$\int x^{\alpha}\,\mathrm{d}x = \dfrac{x^{\alpha+1}}{\alpha+1} + C$，$(\alpha \neq -1)$
$\int \dfrac{1}{x}\mathrm{d}x = \ln\mid x \mid + C$，$(x \neq 0)$
$\int e^{x}\,\mathrm{d}x = e^{x} + C$
$\int a^{x}\,\mathrm{d}x = \dfrac{a^{x}}{\ln a} + C$，$(a > 0，a \neq 1)$
$\int \sin\mathrm{d}x = -\cos x + C$
$\int \cos\mathrm{d}x = \sin x + C$
$\int \sec^{2}\mathrm{d}x = \tan x + C$
$\int \csc^{2}\mathrm{d}x = -\cot x + C$
$\int \dfrac{\mathrm{d}x}{\sqrt{1-x^{2}}} = \sin^{-1}x + C$
$\int \dfrac{\mathrm{d}x}{1+x^{2}} = \tan^{-1}x + C$

【例 5-4】 求下列函数的不定积分：

（1）$y = 2\sin x + \dfrac{3}{x}$　　（2）$y = \dfrac{5}{\sqrt{1-x^{2}}} + 4e^{x}$

（3）$y = \dfrac{x^{2}}{1+x^{2}}$　　（4）$y = (1 - 2\sqrt{x})^{2}$

解：（1）$\int\left(2\sin x+\frac{3}{x}\right)\mathrm{d}x=2\int\sin x\,\mathrm{d}x+3\int\frac{\mathrm{d}x}{x}=-2\cos x+\ln|x|+C$

（2）$\int\left(\frac{5}{\sqrt{1-x^2}}+4e^x\right)\mathrm{d}x=5\int\frac{\mathrm{d}x}{\sqrt{1-x^2}}+4\int e^x\,\mathrm{d}x=5\sin^{-1}x+4e^x+C$

（3）$\int\frac{x^2}{1+x^2}\mathrm{d}x=\int\frac{x^2+1-1}{1+x^2}\mathrm{d}x=\int\mathrm{d}x-\int\frac{\mathrm{d}x}{1+x^2}=x-\tan^{-1}x+C$

（4）$\int(1-2\sqrt{x})^2\,\mathrm{d}x=\int\mathrm{d}x-4\int x^{\frac{1}{2}}\,\mathrm{d}x+4\int x\,\mathrm{d}x=x-\frac{4}{x^{\frac{3}{2}}}+2x^2+C$

$$=x-\frac{4}{x\sqrt{x}}+2x^2+C$$

验证不定积分是否正确，一是看任意常数 C 是否遗漏；二是对求出的结果求导数，看是否还原。

5.4.4 换元积分法

先看一个具体例子。由基本积分表，$\int\sin x\mathrm{d}x=-\cos x+C$，那么 $\int\sin 3x\,\mathrm{d}x=-\cos 3x+C$ 吗？可以通过求导验证：$(-\cos 3x)'=3\sin 3x\neq\sin 3x$！比被积函数多了一个因子3！原因是 $y=\sin 3x$ 是一个复合函数。正像复合函数求导要引入中间变量，从而用链导法一样，复合函数的积分也要“换元”。仍考虑上述例子。

令 $u=3x$，则 $\mathrm{d}u=\mathrm{d}(3x)=3\mathrm{d}x$，于是 $\mathrm{d}x=\frac{\mathrm{d}u}{3}$

$$\int\sin 3x\,\mathrm{d}x=\int\sin u\,\frac{\mathrm{d}u}{3}=\frac{1}{3}\int\sin u\,\mathrm{d}u=-\frac{1}{3}\cos u+C=-\frac{1}{3}\cos 3x+C$$

再次求导验证，$\left(-\frac{1}{3}\cos 3x+C\right)'=\frac{1}{3}\sin 3x\cdot 3=\sin 3x$，正是给定的被积函数。

这就是“换元积分法”：通过换元，对新变量可以套用基本积分表来求出原函数。实际上，“换元积分法”正是“链导法”的逆运算。写成一般公式，若经过换

元 $u=g(x)$，被积函数 $f(x)$ 可以分解为 $G'(g(x))g'(x)\mathrm{d}x$，则

$$\int f(x)\mathrm{d}x=\int G'(g(x))g'(x)\mathrm{d}x=\int G'(u)\mathrm{d}u=G(u)+C=G(g(x))+C \tag{5.31}$$

【例 5-5】 用“换元积分法”求下列不定积分：

(1) $\int\cos(4x+3)\mathrm{d}x$　　(2) $\int(a+bx)^{100}\mathrm{d}x$

(3) $\int xe^{x^2}\mathrm{d}x$　　(4) $\int\sin^3x\mathrm{d}x$

解：(1) 令 $u=4x+3$，则 $\mathrm{d}u=4\mathrm{d}x$，$\mathrm{d}x=\frac{1}{4}\mathrm{d}u$

$$\int\cos(4x+3)\mathrm{d}x=\int\cos u\cdot\frac{\mathrm{d}u}{4}=\frac{1}{4}\sin u+C=\frac{1}{4}\sin(4x+3)+C$$

(2) 令 $u=ax+b$，则 $\mathrm{d}u=a\mathrm{d}x$，$\mathrm{d}x=\frac{\mathrm{d}u}{a}$

$$\int(a+bx)^{100}\mathrm{d}x=\int u^{100}\frac{\mathrm{d}u}{a}=\frac{u^{101}}{101a}+C=\frac{(ax+b)^{101}}{101a}+C$$

(3) 令 $u=x^2$，$\mathrm{d}u=2x\mathrm{d}x$，$x\mathrm{d}x=\frac{\mathrm{d}u}{2}$

$$\int xe^{x^2}\mathrm{d}x=\int e^u\frac{\mathrm{d}u}{2}=\frac{1}{2}e^u+C=\frac{1}{2}e^{x^2}+C$$

(4) $\sin^3x=\sin x\sin^2x=\sin x(1-\cos^2x)$，令 $u=\cos x$，$\mathrm{d}u=-\sin x\mathrm{d}x$

$$\int\sin^3x\mathrm{d}x=-\int(1-u^2)\mathrm{d}u=-\int\mathrm{d}u+\int u^2\mathrm{d}u=-u+\frac{u^3}{3}+C$$
$$=-\cos x+\frac{\cos^3x}{3}+C$$

5.4.5 分部积分法

与微分学中乘积求导相对应的逆运算就是“分部积分法”。由 $(uv)'=u'v+uv'$，得到 $u'v=(uv)'-uv'$

$$\int u'v\,\mathrm{d}x = uv - \int uv'\,\mathrm{d}x \tag{5.32}$$

这就是说，如果一个被积函数 $f(x)$ 可以分解为 $u'(x)v(x)$，不容易积分，但是 $u(x)v'(x)$ 容易积分，那么就可以用上面的公式。还是通过例子说明。

【例 5-6】 求下列不定积分：

(1) $\int x\cos x\,\mathrm{d}x$　　　　(2) $\int \ln x\,\mathrm{d}x$

解：(1) 令 $u = x$，$\cos x\,\mathrm{d}x = \mathrm{d}v$，于是 $v = \sin x$，$\mathrm{d}u = \mathrm{d}x$

$$\int x\cos x\,\mathrm{d}x = \int u\,\mathrm{d}v = x\sin x - \int \sin x\,\mathrm{d}x = x\sin x + \cos x + C$$

如果令 $u = \cos x$ 行不行？这时 $\mathrm{d}v = x\mathrm{d}x = \frac{1}{2}\mathrm{d}(x^2)$，$v = \frac{x^2}{2}$，代入 5.32 式，

$$\int x\cos x\,\mathrm{d}x = \frac{x^2}{2}\cos x - \int \frac{x^2}{2}(-\sin x\,\mathrm{d}x) = \frac{x^2}{2}\cos x + \frac{1}{2}\int x^2\sin x\,\mathrm{d}x$$

得到的积分 $\int x^2\sin x\,\mathrm{d}x$ 甚至比原来的积分 $\int x\cos x\,\mathrm{d}x$ 更复杂，幂函数的次数更高了！

(2) 令 $u = \ln x$，$v = x$，则 $\int \ln x\,\mathrm{d}x = x\ln x - \int x\frac{1}{x}\mathrm{d}x = x\ln x - \int \mathrm{d}x = x\ln x - x + C$

读者不妨回忆，从算术开始，我们已经遇到过许多成对的正运算和逆运算，如加和减、乘和除、乘方和开方、指数和对数、三角函数和反三角函数等。逆运算一般比正运算难。积分运算也比导数运算难得多。甚至有的被积函数看起来简单，其原函数却无法用初等函数表达。如

$$\int \frac{\sin x}{x}\mathrm{d}x,\ \int e^{x^2}\,\mathrm{d}x,\ \int \frac{1}{\sqrt{1+x^4}}$$

等。本书对积分方法不进行更深入的讨论。有兴趣的读者可参阅其他数学教科书，例如[3－21]，[5－1]，[5-2]。

5.5 定积分的简单应用

一个图形的面积由许多小面积累积而成。既然面积可以用定积分表示，那么许多其他累积量也可以用定积分表示。例如体积、功、水坝所受的压力等等。根据 5.26 式，在用定积分表示累积量时，不必从建立和式出发，再求极限；而是找到累积量的“微元”，再求积分。下面举几个实例。

5.5.1 旋转体的体积

一个曲边形绕对称轴旋转一周得到的物体称为“旋转体”。如图 5.13 所示，

曲线 $y=f(x)$ 在 $[a, b]$ 上的曲边梯形绕 x 轴旋转一周，就得到一个旋转体。其体积微元是一个“扁平”的圆柱体，半径为 $f(x)$，高为 $\mathrm{d}x$，$\mathrm{d}V=\pi(f(x))^2\mathrm{d}x$，由此，这个旋转体体积的定积分表达式为

$$V=\pi\int_a^b (f(x))^2\mathrm{d}x \qquad (5.33)$$

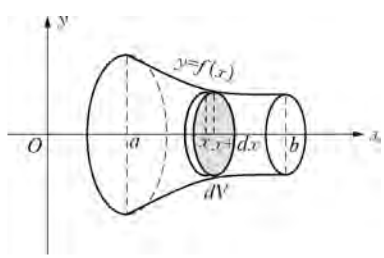


图 5.13

【例 5-7】 对于底面半径为 R，高为 H 的圆锥体，可以看作是直线 $y=\dfrac{R}{H}x$ 绕 x 轴旋转而成(图 5.14(a))，由5.33式，$V=\pi\int_0^H \dfrac{R^2}{H^2}x^2\mathrm{d}x=\dfrac{\pi R^2}{H^2}\dfrac{x^3}{3}\Big|_0^H=\dfrac{1}{3}\pi R^2 H$，这正是熟知的圆锥体积公式。

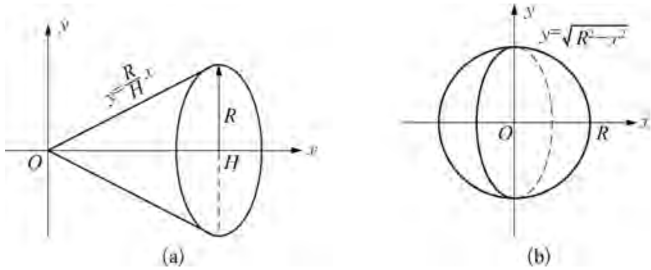


图 5.14

对于半径为 R 的球体，可以看作 $f(x)=\sqrt{R^2-x^2}$，$a=-R$，$b=R$（图

5.14(b)),代入 5.33,

$$V=\pi\int_{-R}^{R}(\sqrt{R^2-x^2})^2\,\mathrm{d}x=\pi\int_{-R}^{R}(R^2-x^2)\,\mathrm{d}x$$

$$=\pi\left(R^2x-\frac{x^3}{3}\right)\Big|_{-R}^{R}=\frac{4}{3}\pi R^3$$

这就是熟知的球体积公式。

5.5.2 水的静压力

【例 5-8】 一块宽为 L、长为 H 的矩形平板垂直插在静水中,设上底与水面齐(图 5.15(a)),侧面受到水的静压力。水的压强与水的深度成正比,如何计算平板侧面的总压力?由于同一深度的压强相同,总压力可以逐层累加。取一深度为 y 的水平细条,面积为 $L\,\mathrm{d}y$,所受压力微元为 $\mathrm{d}F=\rho g y L\mathrm{d}y$,得到总压力

$$F=\rho g L\int_0^H y\mathrm{d}y=\rho g L\,\frac{y^2}{2}\Big|_0^H=\frac{\rho g L H^2}{2}$$

其中 ρ 为液体密度,g 为重力加速度。

如果将上面的矩形板换为半径为 R 的半圆形板(图 5.15(b)),这时压力微元为 $\mathrm{d}F=\rho g y\sqrt{R^2-y^2}\mathrm{d}y$,总压力为 $F=\rho g\int_0^R y\sqrt{R^2-y^2}\mathrm{d}y$ 。求 $y\sqrt{R^2-y^2}$ 的原函数要用到"换元法"。令 $u=R^2-y^2$,
则 $\mathrm{d}u=-2y\mathrm{d}y$,于是

$$y\sqrt{R^2-y^2}\mathrm{d}y=-\frac{1}{2}\sqrt{u}\mathrm{d}u$$

$$-\frac{1}{2}\int\sqrt{u}\mathrm{d}u=-\frac{1}{2}\,\frac{u^{\frac{1}{2}+1}}{\frac{1}{2}+1}+C=-\frac{1}{3}u^{\frac{3}{2}}+C=-\frac{1}{3}(R^2-y^2)^{\frac{3}{2}}+C$$

再代入牛顿莱布尼兹公式,就得到总压力

$$F=-\rho g\,\frac{1}{3}(R^2-y^2)^{\frac{3}{2}}\Big|_0^R=\rho g\,\frac{R^3}{3}$$

定积分在经济领域也有不少应用,我们将在【选修材料之二】中作介绍。

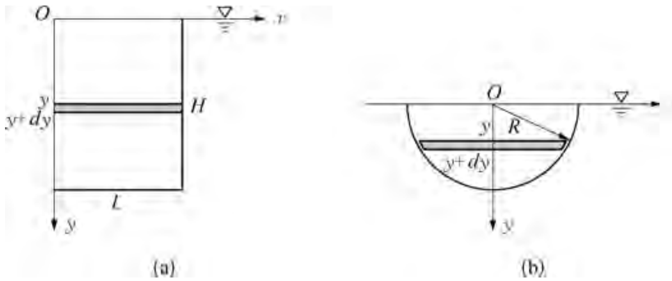


图 5.15

5.6 简单微分方程

5.6.1 什么是微分方程？

我们在小学和中学的数学课程中，接触过两类包含字母的等式："恒等式"和"方程"。

如果字母取任何值等式总是成立，这样的等式就是恒等式，恒等式的等号也常用≡表示。例如：$(a+b)^2 \equiv a^2+2ab+b^2$，$a^2-b^2 \equiv (a+b)(a-b)$，$\sin^2 x+\cos^2 x \equiv 1$，等等。

如果字母只能取特定值等式才成立，这样的等式就是方程。例如：$3x-6=0$，$x^2-3x+2=0$，$3\sin x+4\cos x=5$，$x+2y=1$，$x+2y^2=0$，$x^2/4-y^2/9=1$，等等。

方程待确定的量称为"未知量"，如果未知量取到特定值使得方程成为恒等式，这样的特定值称为方程的"解"（对多项式方程而言，"解"往往称为"根"）。"解"可以是有限个数，也可以是一个无限集合。

例如 $x=2$ 是方程 $3x-6=0$ 的解（或根）；$x=1$，$x=2$ 是方程 $x^2-3x+2=0$ 的解（或根）；而 $x=\dfrac{2k+1}{2}\pi-\tan^{-1}\dfrac{4}{3}$，$k=0, 1, 2, \cdots$ 是方程 $3\sin x+4\cos x=5$ 的解。

又如过(0，0.5)和(1，0)两点连成的直线上所有点的(x, y)集合都是方程 $x+2y=1$ 的解；而方程 $x+2y^2=0$ 和 $x^2/4-y^2/9=1$ 的解则分别是相关抛

物线和双曲线上的点的坐标集合。

如果一个方程的未知量是函数,方程中还包含未知函数的导数或微分,这样的方程称为“微分方程”。下面给出一般的定义。

定义 5.4 含有未知函数导数或微分的方程称为“微分方程”;微分方程中所含未知函数导数的最高阶数称为该方程的“阶”。如果将一个函数代入方程中的未知函数,使得方程成为恒等式,这个函数称为该方程的“解”。

【例 5-9】 设 $y = y(x)$ 是未知函数,那么

$$y' = xy \tag{5.34}$$

就是一阶微分方程。可以验证,$y = e^{\frac{x^2}{2}}$ 是解,$y = 2e^{\frac{x^2}{2}}$,$y = 3e^{\frac{x^2}{2}}$ 也是解,设 C 是任意常数,

$$y = Ce^{\frac{x^2}{2}} \tag{5.35}$$

都是解。

【例 5-10】 设 $x = x(t)$ 是未知函数,那么

$$\ddot{x} + \omega^2 x = 0 \tag{5.36}$$

是一个二阶微分方程。同样可以验证,

$$x = C_1 \cos \omega t + C_2 \sin \omega t \tag{5.37}$$

是它的解,这里 C_1,C_2 是两个任意常数。

定义 5.5 对于一个 n 阶微分方程,把包含 n 个互相独立任意常数的解称为“通解”。

根据上述定义,5.35 是一阶微分方程 5.34 的通解;5.37 是二阶微分方程 5.36 的通解。

5.6.2 几个应用问题

下面通过三个应用问题介绍微分方程的建立和求解方法。

【例 5-11】 无阻力的自由落体运动。设一质量为 m 的物体自由下落，初始速度为 v_0，求物体下落后任一时刻的瞬时速度和位置。

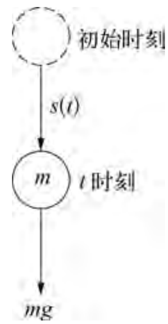


图 5.16

解：设坐标系如图 5.16 所示，S 轴方向和落体运动方向一致，指向地面；原点就设在起始点。设 t 时刻的速度和离地高度分别为 $V(t)$ 和 $S(t)$，外力只有地球的引力 mg，由牛顿第二定律，我们得到微分方程

$$m\ddot{S} = mg\text{，即 }\ddot{S} = g \tag{5.38}$$

或者与之等价的一组微分方程

$$\dot{V} = g \tag{5.39}$$

和

$$\dot{S} = V \tag{5.40}$$

5.38 是一个二阶微分方程；5.39 和 5.40 则是两个一阶微分方程。重力加速度 g 前的负号，是因为物体下落的方向和竖直坐标轴的方向相反。5.39 和 5.40 都可以直接积分。

$$V = \int g\,\mathrm{d}t = gt + C_1 \tag{5.41}$$

以及

$$S = \int (gt + C_1)\mathrm{d}t = \frac{1}{2}gt^2 + C_1 t + C_2 \tag{5.42}$$

其中 C_1，C_2 是两个任意常数。由定义 5.4，5.41 包含 1 个任意常数，是一阶方程 5.39 的通解；而 5.42 包含两个任意常数，是二阶方程 5.38 的通解。

我们把

$$S(0) = 0\text{，}\dot{S}(0) = v_0 \tag{5.43}$$

称为二阶方程 5.38 的“**初始条件**”；

$$V(0) = V_0\text{，}S(0) = 0 \tag{5.44}$$

则分别是一阶方程 5.39 和 5.40 的“初始条件”。将初始条件代入通解，可以确

定任意常数,所以初始条件又称为“**定解条件**”。将上述初始条件代入通解,求出 $C_1 = v_0$, $C_2 = 0$,于是我们得到“特解”:

$$V = gt + v_0,\ S = \frac{1}{2}gt^2 + v_0 t \tag{5.45}$$

这就是我们在中学物理中熟知的公式。

第四章提到炮弹运动轨迹,其实就是解两个微分方程。

水平方向是等速运动:$\dot{x} = V_0\cos\alpha$, $x(0) = 0$,解出 $x(t) = (V_0\cos\alpha)t$;

竖直方向看作是无阻力的自由落体运动:

$m\ddot{y} = -mg$, $\dot{y}(0) = V_0\sin\alpha$, $y(0) = 0$,解出 $y(t) = -\dfrac{1}{2}gt^2 + (V_0\sin\alpha)t$;这就是 4.5 式。

【例 5-12】 有阻力的自由落体运动。如果下落物体受到与速度大小成正比的空气阻力,初始条件和上例相同,求物体下落后任一时刻的瞬时速度和位置。

解:同样由牛顿第二定律,我们得到二阶微分方程

$$m\frac{d^2S}{dt^2} = mg - k\frac{dS}{dt} \tag{5.46}$$

或者与之等价的一组微分方程

$$m\frac{dV}{dt} = mg - kV \tag{5.47}$$

初始条件仍是 5.43 和 5.44。

对于方程 5.47,等式右边包含未知函数,不能像上例那样直接积分。但是可以把方程变形为

$$\frac{m\,dV}{mg - kV} = dt$$

对等式两边积分:

$$\int \frac{dV}{g - \dfrac{k}{m}V} = \int dt$$

右边的积分很容易,$\int dt = t + C_1$,左边的积分却要用“换元法”,令 $u = g - \dfrac{k}{m}V$,

$\mathrm{d}u = -\dfrac{k}{m}\mathrm{d}V$，$\mathrm{d}V = -\dfrac{m}{k}\mathrm{d}u$，代入上式左端

$$\int \frac{\mathrm{d}V}{g - \frac{k}{m}V} = -\frac{m}{k}\int \frac{\mathrm{d}u}{u},\ -\frac{m}{k}\ln u = t + C_1,\ 即\ \frac{m}{k}\ln\left(g - \frac{k}{m}V\right) = -t + C_1$$

得

$$\ln\left(g - \frac{k}{m}V\right) = -\frac{k}{m}t + C_1,\ g - \frac{k}{m}V = C_1 e^{-\frac{k}{m}t}$$

得到瞬时速度的通解

$$V = \frac{m}{k}(g - C_1 e^{-\frac{k}{m}t}) \tag{5.48}$$

代入初始条件 5.44，解出 $C_1 = g - \dfrac{k}{m}v_0$，得到瞬时速度的特解

$$V = \frac{mg}{k} + \left(v_0 - \frac{mg}{k}\right)e^{-\frac{k}{m}t} \tag{5.49}$$

积分一次

$$S = \int\left(\frac{mg}{k} + \left(v_0 - \frac{mg}{k}\right)e^{-\frac{k}{m}t}\right)\mathrm{d}t = \frac{mgt}{k} - \frac{m}{k}\left(v_0 - \frac{mg}{k}\right)e^{-\frac{k}{m}t} + C_2$$

由初始条件 5.44，$C_2 = \dfrac{m}{k}\left(v_0 - \dfrac{mg}{k}\right)$，得到瞬时位置的特解

$$S = \frac{mgt}{k} + \frac{m}{k}\left(v_0 - \frac{mg}{k}\right)(1 - e^{-\frac{k}{m}t}) \tag{5.50}$$

比较 5.45 和 5.49，5.50，可见，在无阻力情况下，速度越来越快；而在有与速度成正比的阻力情况下，无论是速度还是位置，$e^{-\frac{k}{m}t}$ 项快速衰减，$\lim\limits_{t\to\infty} V = \dfrac{mg}{k}$。所以当 t 很大时，接近于等速运动。

回顾第四章 4.4.3 节，讨论流体中的自由落体运动。4.51 式、4.52 式和 5.46 式、5.47 式很相似，但那里流体阻力与速度平方成正比，请读者试行求解。

【例 5-13】 人口问题。如果在 t_0 时刻人口总数为 P_0，当 $t > t_0$，能否预测未来人口总数 $P(t)$？


马尔萨斯 1766-1834

最早研究这个问题的是英国经济学家马尔萨斯(Thomas Robert Malthus，1766-1834)。马尔萨斯出身于一个富有的家庭。他的父亲丹尼尔是著名哲学家休谟(David Hume，1711-1776)和法国思想家卢梭(Jean-Jaeques Rousseau，1712-1788)的朋友。他本人与另一位经济学大师李嘉图(David Ricardo，1772-1823)也是挚友。1805年他成为世界上第一位政治经济学教授。[5-3]1798年，马尔萨斯针对18世纪末英国工业革命所造成的大批工人事业失业、贫困问题突出等社会问题以匿名的方式发表了《An Essay on the Principle of Population(人口原理)》，他提出的数学模型基于这样的假定：人口的变化率与当前人口的总数成正比，即

$$\frac{\mathrm{d}P}{\mathrm{d}t} = aP \tag{5.51}$$

初始条件为

$$P(0) = P_0 \tag{5.52}$$

5.51是一阶微分方程。虽不能直接积分，可以类似于【例5-12】那样变形为

$$\frac{\mathrm{d}P}{P} = a\,\mathrm{d}t$$

两边分别积分，$\int \frac{\mathrm{d}P}{P} = \int a\,\mathrm{d}t$，$\ln P = at + C$，于是通解为

$$P(t) = Ce^{at} \tag{5.53}$$

代入初始条件5.52，得到所需特解

$$P(t) = P_0 e^{at} \tag{5.54}$$

上式表明，人口呈指数增长。记 $e^a = k$，设时间 t 以年为单位，那么

$$\begin{aligned} &P(1) = P_0 e^a = kP_0, \\ &P(2) = P_0 e^{2a} = P_0 e^a e^a = kP_1, \cdots, P(n+1) = kP(n) \end{aligned} \tag{5.55}$$

于是，人口以几何级数(即公比为 k 的等比级数)增长。马尔萨斯认为，食物只能以算术级数(即等差级数)增长，久而久之，食物将不能供应人口的需要。只有自

然原因(事故和衰老)、灾难(战争、瘟疫及各类饥荒)、道德限制和罪恶能够限制人口的过度增长,否则将产生灾难。

虽然一开始马氏理论就遭质疑,但其影响却非常大。从 1798 年到 1826 年,三十年中马尔萨斯不断修改,他的著作出了 6 版。对马尔萨斯理论的关注也帮助了英国人口普查的实施。1801 年,英国举行了第一次现代意义上的人口普查。马尔萨斯理论对现代进化论创始人达尔文(Charles Robert Darwin,1809-1882)和华莱士(Alfred Russel Wallace,1823-1913)也产生了重大的影响。

有学者用美国从 1790 年以来的人口数据,检验马氏理论的准确性。[5-5] 他们以10 年为单位,根据 $N(1790)=3929000$,$N(1800)=5308000$,得到 $k=\dfrac{5308000}{3929000}\approx 1.351$。表 5.5 列出了实际人口数据和按式 5.55 的预测值,以及两者的误差。

表 5.5

年	实际数值/千人	预测数值/千人	绝对误差/千人	相对误差/%
1790	3929	/	/	/
1800	5308	/	/	/
1810	7240	7171	−69	0.95
1820	9638	9688	50	0.52
1830	12866	13089	217	1.68
1840	17069	17683	614	3.60
1850	23192	23890	698	3.01
1860	31443	32275	832	2.65
1870	38558	43603	5045	13.08
1880	50156	58908	8752	17.45
1890	62948	79585	16637	26.43
1900	75995	107519	31524	41.48
1910	91972	145258	53286	57.93
1920	105711	196243	90532	85.64
1930	122775	265125	142350	115.94
1940	131669	358184	226515	172.03
1950	150697	483906	333209	221.11

由表上数据可见,直到 1860 年,误差还不大;但以后误差越来越大,到 1950 年,预测值已是实际值的两倍以上,当然不可接受。说明数学模型 5.51 对于长

维尔霍斯特 1804-1849

期预测是不可行的。

1838年比利时生物数学家维尔霍斯特(Pierre-François Verhulst，1804－1849)提出一个改进的模型[5-4]：他假定，某个地区的资源最多能支持的人口数为 M，其中 M 是“承载能力”，即环境能支持最多的人口数。当 $P \to M$，$\mathrm{d}P/\mathrm{d}t \to 0$，说明当人口数达到最大值时，不再增长。于是他建立了如下微分方程：

$$\frac{\mathrm{d}P}{\mathrm{d}t} = aP\left(1 - \frac{P}{M}\right) \tag{5.56}$$

或者等价表示为

$$\frac{\mathrm{d}P}{\mathrm{d}t} = P(a - bP) = aP - bP^2, \left(\text{其中 } b = \frac{a}{M}\right) \tag{5.57}$$

将5.57和5.51相比，右端多了一项 $-bP^2$。将5.57变形，

$$\frac{\mathrm{d}P}{P(a - bP)} = a\,\mathrm{d}t$$

通过换元积分法，可以求出通解(过程略去，读者可以验证，也可自行推导，或查阅参考书[5-5])

$$P(t) = \frac{aC}{bC + a^{-at}} \tag{5.58}$$

代入初始条件5.52，解出 $C = \dfrac{P_0}{a - bP_0}$，从而得到所需特解

$$P(t) = \frac{aP_0}{bP_0 + (a - bP_0)e^{-at}} \tag{5.59}$$

可见，

$$\lim_{t \to \infty} P(t) = \frac{a}{b} = M$$

仍用美国人口数据来检验。我们用1790、1800和1810三个年份的实测数据为基础，解出参数

$$a = 0.03134,\ b = M = 197273000$$

表 5.6 是相应数据。与表 5.5 相比，维尔霍斯特模型要准确得多。5.56 或 5.57 式通常被称为“逻辑斯蒂方程”。

表 5.6

年	实际数值/千人	预测数值/千人	绝对误差/千人	相对误差/%
1790	3929	/	/	/
1800	5308	/	/	/
1810	7240	/	/	/
1820	9638	9757	119	1.2
1830	12866	13109	243	1.9
1840	17069	17506	437	2.6
1850	23192	23192	0	0.0
1860	31443	30412	−1031	−3.3
1870	38558	39372	814	2.1
1880	50156	50177	21	0.0
1890	62948	62769	−179	−0.3
1900	75995	76870	875	1.2
1910	91972	91972	0	0.0
1920	105711	107559	1848	1.7
1930	122775	123124	349	0.3
1940	131669	136653	4984	3.8
1950	150697	149053	−1644	−1.1

5.6.3 一阶微分方程

上面 3 个应用问题都归结为一阶微分方程。其一般形式为

$$\frac{\mathrm{d}y}{\mathrm{d}x} = f(x, y) \tag{5.60}$$

(1) 当 $f(x, y) = f(x)$，方程直接可积，式 5.40、5.41 就属于此类型。

(2) 当 $f(x, y) = f(x)g(y)$，称为“可分离变量方程”，当 $g(y) \neq 0$，将方程变形为

$$\frac{dy}{g(y)} = f(x)dx$$

两边分别对 y 和 x 积分，就可得到通解。式 5.47、5.51 和 5.56 就属于此类型。

(3) 当 $f(x, y) = -P(x)y + Q(x)$，方程改写为

$$y' + P(x)y = Q(x) \tag{5.61}$$

称为一阶线性微分方程。其通解为

$$y = e^{-\int P(x)dx}\left(C + \int Q(x)e^{\int P(x)dx}dx\right) \tag{5.62}$$

这里略去推导，读者可参阅[5-1][5-5]。5.47 就是一阶线性方程，其中 $P(t) = \frac{k}{m}$，代入 5.62，得到 $V = \frac{m}{k}(Ce^{\frac{k}{m}t} + g)$，与 5.48 式一致。

正像初等数学的各个分支一样，同一个微分方程可以描述不同的实际问题。例如 5.51 不仅曾用于人口预测，也可用于其他生物种群的繁衍，还可用于放射物衰减的研究。假设 $P(t)$ 为一放射物在 t 时刻钚的含量，而 15 年后测出 0.043% 的钚已衰减，由 5.54 式，$(1-0.00043)P_0 = P_0e^{15a}$，解出衰减系数

$$a = \frac{1}{15}\ln 0.00057 = -0.00002867$$

物理上有一个概念叫“半衰期”，即放射性物质衰减一半所需的时间。以钚为例，

$$\frac{P_0}{2} = P_0e^{-0.00002867T}$$

解出

$$T = \frac{\ln 2}{0.00002867} \approx 24180\ 年$$

研究放射性物质衰减有重大的现实意义。1986 年 4 月 26 日，前苏联切尔诺贝利核电站发生爆炸，8 吨多强辐射物质喷涌而出。核泄漏过程持续了 10 天，大量锶、铯、钚等放射性物质散到乌克兰、白俄罗斯以及其他欧洲国家。据不完全统计，切尔诺贝利核事故的受害者总计达 900 万人。2011 年 3 月日本大地震引发海啸，造成福岛核电站放射性物质外泄，造成环境污染并使公众受到辐射危害。

考古学家判定化石或其他古文物的年代，常用碳14方法，也是方程5.51的应用。碳14是碳的一种具放射性的同位素，于1940年首被发现。美国物理化学家利比（Willard Frank Libby，1908-1980）在1949年求出碳14的半衰期约为5730年。生物在生存的时候，由于需要呼吸，其体内的碳14含量大致不变，生物死去后会停止呼吸，此时体内的碳14开始减少。人们可通过测一件古物的碳14含量，来估计它的大概年龄，这种方法称为碳14年代测定法。[5-6]

利比 1908-1980

由于对考古学带来革命性变革，利比在1960年荣获诺贝尔化学奖。

5.6.4 二阶微分方程

许多物理问题归结为二阶微分方程。式5.38和5.46就是两个二阶方程。但是多数二阶方程并不能像这两个方程那样，立即转化为一阶方程。下面讨论一个弹簧振动问题(图5.17)[5-7]。

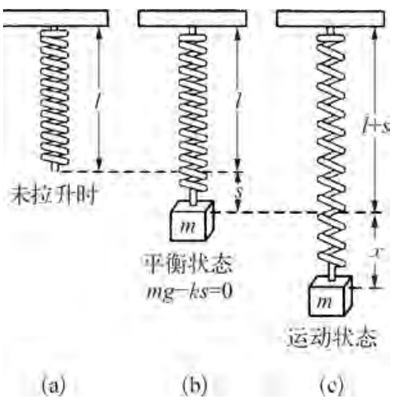


图5.17

【例5-14】 考虑一个长度为 l 的弹簧，在一端吊上质量为 m 的重物，弹簧伸长 s 后平衡。如果给重物一个扰动，物体上下振动。设物体离开平衡位置的位移为 $x(t)$，物体受到重力和弹簧力作用，由牛顿第二定律和虎克（Hook）定律，平衡方程为 $m\ddot{x}=mg-k(s+x)$，由于 $mg=ks$，我们有 $m\ddot{x}=-kx$，即

$$m\ddot{x}+kx=0 \tag{5.63}$$

令 $k/m=\omega^2$，得到 $\ddot{x}+\omega^2x=0$，这就是5.36式，其通解如5.37所示。本书不深入讨论二阶方程的求解方法，有兴趣的读者可参阅[5-1][5-5][5-7]。

微分方程对符合因果律的确定现象可以精确表达。中国的“神舟”飞船和“天宫”航天器在浩瀚宇宙中对接，时间误差不超过1秒，空间误差不超过1 cm，

这里就有微分方程的威力。微分方程在科学、工程和经济现象中有许多应用。仅从本章所列参考书目中提到的应用项目就有“伪造名画案件”、“技术革新的推广”、“放射性废物的处理”、“肿瘤生长动力学”、“大桥的坠毁”、“诊断糖尿病的数学模型”[5-5]、“可再生能源”、“溶液中浓度变化”、“圆球的温度分布”[5-7]、“导弹跟踪问题”、“行星的轨道和位置”、“生物电分析的小波方法”和“股票期权定价问题”[5-8]等等。

5.7 思考与练习

1. 定积分的定义是什么？定积分描述什么对象？能不能找一个物理量用定积分来表示？

2. 什么是一个函数的原函数？不定积分的定义是什么？

3. 不定积分和定积分有什么不同？

4. 怎样理解导数(或微分)和积分是互逆运算？

5. 求下列不定积分或定积分：

(1) $\int(1+2x-3x^2)\,\mathrm{d}x$　　(2) $\int\left(1-\frac{3}{x}+e^x\right)\mathrm{d}x$

(3) $\int\sin(2x+1)\,\mathrm{d}x$　　(4) $\int(e^x+xe^x+x^2e^x)\,\mathrm{d}x$

(5) $\int_0^2 x^4\,\mathrm{d}x$　　(6) $\int_0^2\frac{x}{1+x}\mathrm{d}x$

(7) $\int_{-1}^1 x\cos x\,\mathrm{d}x$　　(8) $\int_{-1}^1\frac{x^2}{1+x^2}\mathrm{d}x$

6. 求由直线 $y=-x+2$ 与抛物线 $y=x^2$ 所围成的面积。

7. 已知物体在 t 秒时刻的运动速度 $v(t)=2+t+3t^2$，求物体从开始到 2 秒所走过的路程。

8. 求解下列微分方程

(1) $\frac{\mathrm{d}y}{\mathrm{d}x}=-y^2$，$y(0)=1$　　(2) $y'+y=e^{-x}$，$y(0)=3$

9. 求碳 14 的半衰期。

10. 假定一个有 1000 名学生的校园,在某一天有一名学生感染了传染病毒。已知这种病毒的传播速度不仅与感染者数量 x 成正比,也与未感染者的数量 $1000-x$ 成正比。经统计,在第 4 天已有 50 名学生受到感染。如无有效措施控制,试预测第 6 天有多少学生被感染? 多少天后全部学生会被感染?

11. 试用可分离变量方法解 4.4.3 节建立的微分方程 4.52,然后解 4.51。

参考文献或网站

[5-1] 王绵森,马知恩:《工科数学分析基础(上册)》,高等教育出版社(1998).

[5-2] 同济大学数学系:《微积分(上册)》(第三版),高等教育出版社(2010).

[5-3] 百度百科:托马斯・马尔萨斯,baike. baidu. com/view/15445. htm.

[5-4] http://www-history. mcs. st-andrews. ac. uk/Mathematicians/Verhulst. html.

[5-5] M. Braun:《微分方程及其应用(Differential Equations and Their Applications)(第 4 版)》,Springer-Verlag,世界图书出版公司(1998).

[5-6] 互动百科:碳-14 测定年代 http://www. hudong. com/wiki.

[5-7] D. G. Zull, M. R. Cullen:《微分方程与边界问题(Differential Equations with Boundary-Value Problems)(第 5 版)》机械工业出版社(2003).

[5-8] 乐经良,向隆万,李世栋:《数学实验(第 2 版)》,高等教育出版社(2011).

第六章

优化问题

人们在日常生活和生产活动中，常会遇到优化问题。比如出外旅游，选择怎样的路线费用最少？或者时间最短？又如种庄稼要施肥，太少和太多都不行，施多少产量最高？一个企业，一个地区，甚至一个国家在决策过程中，为了又好又快地发展，都离不开优化问题。

6.1　用微分学方法求函数极值

6.1.1　两个实际问题

最常见的优化问题是求一个函数的最大值或最小值。先看两个实际问题。

【例 6-1】 一个边长为 1 m 的正方形纸板，在四个顶点各剪去一个小正方形后做成一无盖纸盒(图 6.1)。问如何剪，使得纸盒的容积最大？

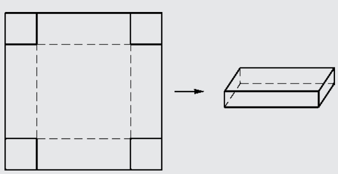


图 6.1

设剪去的小正方形边长为 x，则纸盒的容积为

$$V = x(1-2x)^2 \tag{6.1}$$

问题归结为：x 取多大时，V 达到最大值？

根据实际意义，$V>0$，所以 $0<x<\frac{1}{2}$。不妨试算一些数据，见表6.1(1)。

表 6.1(1)

X/m	0.1	0.2	0.3	0.4
V/m^3	0.064	0.072	0.048	0.016

可见 x 在(0.1, 0.2)之间,V 可能取得最大值。继续细分,得到如下数据

表 6.1(2)

X/m	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19
V/m^3	0.061	0.069	0.071	0.0726	0.0735	0.0740	0.0741	0.0737	0.0730

可见 x 在 0.17 m 左右取值,V 最大。还可继续细分,能否有更便捷的方法?

【例 6-2】 已知轮船的燃料费与速度的立方成正比,当速度为 10 km/h 时,每小时燃料费为 80 元。每小时其他费用需 480 元。设总航程为 200 km,问速度多大时,总费用最省?

设每小时燃料费为 f 元,记船速为 v,则 $f = kv^3$,由 $1000k = 80$,解出 $k = 0.08$。当船速为 v 时,航行 200 km 所需时间为 $t = 200/v$ 小时,总费用为

$$F = vt = (480 + 0.08v^3)\frac{200}{v} = \frac{96000}{v} + 16v^2 \tag{6.2}$$

问题归结为:v 取多大时,F 达到最小值?

类似上例,不妨列表试算,结果如表 6.2 所示。

表 6.2

V/(km/h)	5	10	15	20	25
V/元	19600	11200	10000	11200	13840

可见当船速为 15 km/h 左右,总费用最省。如果要更精确,需进一步细分试算。显然,这也不是好方法。

以上两个问题都属于**优化问题**。6.1 和 6.2 分别是两个问题的**目标函数**，其自变量“小正方形的边长 x”和“船速 v”分别是两个问题的**优化变量**。微分学为许多最大值问题和最小值问题提供了便捷而精确的方法。求函数最大值或最小值也是导致微积分产生的问题之一[0-1]，其基础是导数和函数性态的密切关系。

6.1.2　用导数研究函数的性态

1）微分学中值定理

先介绍一个以法国数学家拉格朗日（Joseph-Louis Lagrange，1735-1813）命名的定理。

拉格朗日　1735-1813

拉格朗日在数学、力学和天文学三个学科领域中都有历史性的贡献，在科学史中以他名字表示的概念、方法不胜枚举。例如我国“嫦娥号”到达地球和太阳之间的“拉格朗日点”，就是指“嫦娥号”在太阳和地球引力作用下，处于稳定状态的点。这类点的存在就是拉格朗日于 1772 年推导证明的。[6-1]

定理 6.1　拉格朗日中值定理　*若函数 $f(x)$ 满足：(1) 在 $[a, b]$ 连续，(2) 在 (a, b) 可导，则在 (a, b) 中至少存在一点 ξ，使得*

$$f(b)-f(a)=f'(\xi)(b-a) \tag{6.3}$$

定理的证明略去，有兴趣的读者可以阅读参考书[5-1][5-2]。我们用图形来说明。将 6.3 变形为

$$\frac{f(b)-f(a)}{b-a}=f'(\xi) \tag{6.4}$$

考察图 6.2，式 6.4 左端表示线段 AB 的斜率，将 AB 平行移动，只要函数连续可导，至少和函数曲线在一点 $P(\xi, f(\xi))$ 相切。过这点切线的斜率正是 $f'(\xi)$。注意，这样的点不一定唯一。图中 Q 点也满足条件。

2）导数的正负决定函数的增减

第三章曾经给出函数单调增减的定义。如果一个函数可导，那么由导数的正负可以方便地判断函数的增减。由图 6.3(a)可见，对于一个单调增函数 $y=$

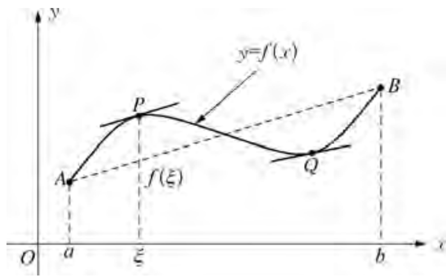


图 6.2

$f(x)$，其图形是沿 x 轴正向上升的曲线，当 $\Delta x = x_2 - x_1 > 0$，$\Delta y = f(x_2) - f(x_1) \geqslant 0$，切线与 x 轴的夹角为锐角（也有可能出现水平切线，如图中的 P 点），即 $f'(x_1) \geqslant 0$。

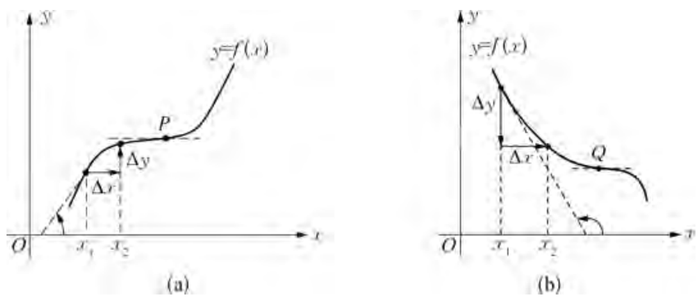


图 6.3

而对于一个单调减函数，其图形则是沿 x 轴正向下降的曲线（图 6.3(b)），当 $\Delta x = x_2 - x_1 > 0$，$\Delta y = f(x_2) - f(x_1) \leqslant 0$，切线与 x 轴的夹角为钝角（也有可能出现水平切线，如图中的 Q 点），即 $f'(x_1) \leqslant 0$。

反过来，如果切线斜率为正，函数单调增；切线斜率为负，则函数单调减。这就要用拉格朗日定理来证明了。如果在 (a, b) 中 $f'(\xi) > 0$，在 $[a, b]$ 中任取两点 $x_1 < x_2$，由 6.3 式，$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1)$，其中 ξ 点在 (x_1, x_2) 之中。由于 $f'(\xi) > 0$ 和 $x_2 - x_1 > 0$，所以 $f(x_2) - f(x_1) > 0$。

由此，我们得到一个数学结论：

定理 6.2 设函数 $y = f(x)$ 在区间 I 中可导，则此函数在 I 中单调增（减）的充要条件是 $f'(x) \geqslant 0$ $(\leqslant 0)$。

【例 6-3】 研究函数 $y = (x-2)^2 - 1$ 的单调性。

解：$((x-2)^2-1)'=2(x-2)(-1)=-2x+4$，当$-2x+4>0$，即$x>2$，函数单调增；当$x<2$，函数单调减。（图6.4）

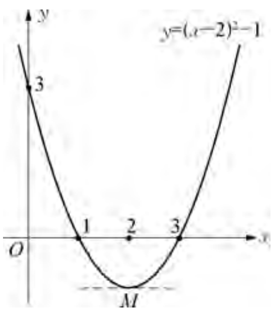


图6.4

3）驻点与函数的极值

在【例6-3】中，当$x=2$时，$y'=0$。从而可见，曲线在这点（图6.4中的M点）有水平切线，函数达到极小值，也就是$f(-2)=-1<f(x)$，x是除-2以外的所有实数。但是曲线有水平切线的地方不一定达到极大值或极小值，如图6.3的P点和Q点。对极值问题的数学表述如下：

定义6.1 如果函数$f(x)$在定义域中某点$x=a$，$f(a)\geqslant f(x)$，对x在$x=a$的邻域中所有点都成立，则称$x=a$是**极大值点**，$f(a)$为**极大值**；如果$f(a)\leqslant f(x)$，则称$x=a$是**极小值点**，$f(a)$为**极小值**。极大值点和极小值点统称为极值点；极大值和极小值则统称为极值。

定理6.3 对于可导函数，如果存在极值点$x=a$，则$f'(a)=0$。

定义6.2 可导函数$f(x)$满足$f'(a)=0$，则称$x=a$为“**驻点**”。

定理6.4 如果$x=a$是可导函数$f(x)$的驻点，

（1）当$x<a$，$f'(x)>0$，而$x>a$，$f'(x)<0$，则$x=a$是函数的极大值点；

（2）当$x<a$，$f'(x)<0$，而$x>a$，$f'(x)>0$，则$x=a$是函数的极小值点；

（3）当$x<a$和$x>a$，$f'(x)$不改变符号，则$x=a$不是函数的极值点。

图6.5描述了这三种情况。

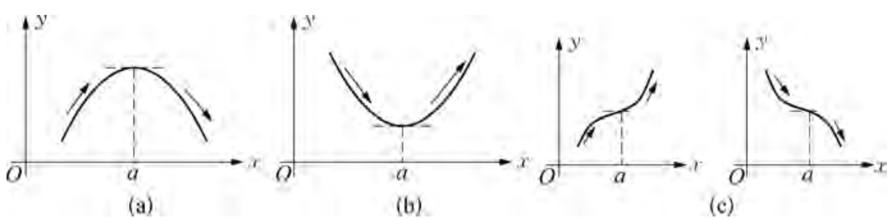


图6.5

现在我们回到前面两个实际问题。

对【例 6-1】的目标函数 6.1 求导，$\frac{\mathrm{d}V}{\mathrm{d}x}=(1-2x)^2+x\cdot 2(1-2x)(-2)=0$，即 $(1-2x)(1-6x)=0$，解出两个根：$x_1=\frac{1}{2}$，$x_2=\frac{1}{6}$。$V\left(\frac{1}{2}\right)=0$，显然不符合实际要求，实际上 $x=\frac{1}{2}$ 是函数 6.1 的极小值；当函数通过 $x=\frac{1}{6}$ 这点时，$\frac{\mathrm{d}V}{\mathrm{d}x}$ 由正变负，根据定理 6.4，$x=\frac{1}{6}=0.166\cdots$ 是极大值点，容积 $V\left(\frac{1}{6}\right)=\frac{2}{27}=0.074074(\mathrm{m}^3)$，是极大值。

对【例 6-2】的目标函数 6.2 求导，$\frac{\mathrm{d}F}{\mathrm{d}v}=-\frac{96000}{v^2}+32v=0$，解出 $v=10\sqrt[3]{3}=14.42\cdots$，当 v 通过这点时，$\frac{\mathrm{d}F}{\mathrm{d}v}$ 由负变正，根据定理 6.4，这点为极小值点，$F(10\sqrt[3]{3})=9984.4$ 元，是极小值。

从这两个实例，可以看出微分学方法的优越性。

6.1.3 函数的最大值与最小值

读者可能发现，定义 6.1 中并未要求函数可导，而定理 6.3、6.4 都针对可导函数，函数有定义但不可导的点，是否可能是极值点？前面用过“最大值、最小值”和“极大值、极小值”两种提法。两者是同义词，还是有所区别？我们作如下讨论。

(1) 在函数有定义但不可导的点上，可能取得极值。

(2) 按照定义 6.1，“极值点”和“极值”指函数的局部性质，即一点和它的邻域各点的函数值相比。而“最值”(“最大值”和“最小值”的总称)指函数的整体性质，即一点和定义域所有点的函数值相比。一个函数可能在定义域中有多个极大值或极小值，但只能有一个最大值或最小值。

(3) 根据第三章介绍的闭区间连续函数的最值定理，端点上也可能达到极值。

图 6.6 中的函数在$[a, b]$有定义，从图上可见，在(a, b)中 $x=c$，$x=e$，$x=g$ 是极小值点，在这 3 点函数均可导；而 $x=d$，$x=f$ 是极大值点，其中

$x=d$ 是可导点，$x=f$ 是不可导点；$x=a$，$x=b$ 是边界点。从图上还可见，函数在 $[a, b]$ 上的最小值点是 $x=e$，最大值点是 $x=b$。

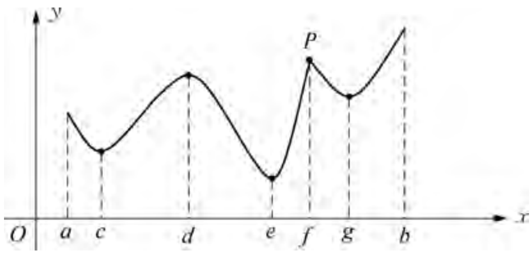


图 6.6

因此，在求最值的优化问题中，步骤如下：

第 1 步，选择优化变量，建立目标函数，并确定定义域；

第 2 步，找到全部驻点和不可导点；

第 3 步，确定这些点是否极值点；

第 4 步，比较上述极值点及目标函数定义域两端点的函数值，确定最值点，并求出最值。

如果目标函数的定义域是开区间，当然在第 4 步中不必也不能比较端点的函数值。

【例 6-4】 求函数 $f(x)=\dfrac{x^5}{5}+\dfrac{x^4}{4}-\dfrac{x^3}{3}-\dfrac{x^2}{2}+1$ 在 $[-2, 2]$ 的极值与最值。

解： $f'(x)=x^4+x^3-x^2-x=x(x+1)^2(x-1)=0$，有 3 个驻点 $x=-1, 0, 1$，没有不可导点。根据定理 6.4，$x=-1$ 不是极值点；$x=0$ 是极大值点，$x=1$ 是极小值点。和两个端点的函数值比较，

$$f(0)=1,\ f(1)=\frac{37}{60}=0.61\dot{6},\ f(-2)=-\frac{11}{15}=-0.7\dot{3},$$

$$f(2)=\frac{101}{15}=6.7\dot{3}$$

所以最大值是 $f(2)=\dfrac{101}{15}=6.7\dot{3}$，最小值是 $f(-2)=-\dfrac{11}{15}=-0.7\dot{3}$。

再次回到两个实际问题。【例 6-1】的目标函数 6.1 只有两个驻点 $x=\frac{1}{2}$ 和 $x=\frac{1}{6}$，没有不可导点；其中 $x=\frac{1}{2}$ 又是边界点，另一个边界点为 $x=0$。

因 $V(0)=V\left(\frac{1}{2}\right)=0$，$V\left(\frac{1}{6}\right)=\frac{2}{27}$，所以最大值点是 $x=\frac{1}{6}$，最大值是 $\frac{2}{27}$，最小值点是 $x=0$ 和 $x=\frac{1}{2}$，最小值是 0。对这个实际问题来说求最小值是没有意义的。

实际问题中如果目标函数可导，且只有一个驻点，一般这个驻点就是所求的极值点，无须经过定理 6.4 的判断。【例 6-2】只有唯一驻点 $v=10\sqrt[3]{3}$，这就是所求的极小值点，即最小值点。注意：在这个实际问题中求最大值是没有意义的。

下面我们再讨论一个环境污染问题。

【例 6-5】 烟囱向其周围散落烟尘而污染环境。已知落在某处的烟尘浓度与该处至烟囱距离的平方成反比，而与该烟囱喷出的烟尘量成正比。现有两座烟囱相距 20 km，其中一座烟囱喷出的烟尘量是另一座的 8 倍，试求出两座烟囱连线上的一点，使该点的烟尘浓度最小。

解： 设这点距小排量烟囱距离 x km，此点接受到的烟尘为

$$Q(x)=\frac{1}{x^2}+\frac{8}{(20-x)^2},\ x\in(0,20) \tag{6.5}$$

$Q'(x)=\frac{-2}{x^3}+\frac{(-2)\times 8(-1)}{(20-x)^3}=0$，$x=\frac{20}{3}$ km，唯一驻点，就是极小值点，也是最小值点。所以距小排量烟囱距离 6.67 km 处污染最小。目标函数 6.5 的定义域是开区间，不必也不能讨论端点。

6.1.4 二元函数的极值

以上讨论的问题中，只有一个优化变量，目标函数都是一元函数。但是许多问题有多个优化变量，目标函数就是多元函数。下面讨论二元函数的极值，对于

三元以上的函数,可以类推。先看一个实际问题。

【例 6-6】 一块铁片宽为 3 m,要把它折成一个梯形断面的水槽(图6.7),怎样才能使它的过水断面的面积最大?

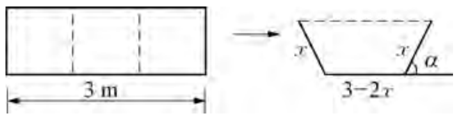


图 6.7

设水槽侧壁长为 x m,与水平面夹角为 α,水槽断面是梯形,其面积为

$$A(x,\ \alpha)=\frac{1}{2}(3-2x+3-2x+2x\cos\alpha)x\sin\alpha=(3-2x+x\cos\alpha)x\sin\alpha$$

利用三角公式 $\sin 2\alpha=2\sin\alpha\cos\alpha$,得到

$$A(x,\ \alpha)=3x\sin\alpha-2x^2\sin\alpha+\frac{1}{2}x^2\sin 2\alpha \tag{6.6}$$

这是 x 和 α 的二元函数。

对于二元函数,如果要逐点试算,工作量就更大了。下面介绍怎样用微积分方法求解。

首先引入“**偏导数**”的概念。

定义 6.3 对于二元函数 $z=f(x,\ y)$,

$$\lim_{\Delta x\to 0}\frac{f(x+\Delta x,\ y)-f(x,\ y)}{\Delta x}=\frac{\partial f}{\partial x} \tag{6.7}$$

$$\lim_{\Delta y\to 0}\frac{f(x,\ y+\Delta y)-f(x,\ y)}{\Delta y}=\frac{\partial f}{\partial y} \tag{6.8}$$

分别称为函数 $f(x,\ y)$ 对 x 和 y 的偏导数,也可记为 $\frac{\partial z}{\partial x}$, $\frac{\partial z}{\partial y}$,或 f_x, f_y,或 z_x, z_y。

【例 6-7】 求下列函数的偏导数:

(1) $z=x^2\sin y$

(2) $z=x^2+y^2-3xy^3$,求在点(1, 2)处的偏导数 $\frac{\partial z}{\partial x}$, $\frac{\partial z}{\partial y}$

解：(1) 把 y 看作常数，$\frac{\partial z}{\partial x}=2x\sin y$；把 x 看作常数，$\frac{\partial z}{\partial y}=x^2\cos y$。

(2) 把 y 看作常数，$\frac{\partial z}{\partial x}=2x-3y^3$，$\left.\frac{\partial z}{\partial x}\right|_{(1,2)}=2-3\times 2^3=-22$；

把x看作常数，$\frac{\partial z}{\partial y}=2y-9xy^2$，$\left.\frac{\partial z}{\partial y}\right|_{(1,2)}=2\times 2-9\times 1\times 2^2=-32$。

下面给出二元函数极值的定义以及叙述极值点求法(不加证明)。

定义 6.4 如果函数 $f(x, y)$ 在定义域中某点 (a, b)，$f(a, b)\geqslant f(x, y)$，对 (x, y) 在 (a, b) 的邻域中所有点都成立，则称 (a, b) 是**极大值点**，$f(a, b)$ 为**极大值**；如果 $f(a, b)\leqslant f(x, y)$，则称 (a, b) 是**极小值点**，$f(a, b)$ 为**极小值**。极大值点和极小值点统称为极值点；极大值和极小值则统称为极值。

定理 6.5 对于可导函数 $f(x, y)$，如果存在极值点 (a, b)，则

$$\frac{\partial f}{\partial x}(a, b)=0,\ \frac{\partial f}{\partial y}(a, b)=0 \tag{6.9}$$

定义 6.5 满足 6.9 的点(a, b)称为函数 $f(x, y)$的“驻点”。

和一元函数一样，一个二元函数的驻点可能是极值点，也可能不是极值点。例如 $z=xy$ 和 $z=x^2+y^2$，两者的驻点都是$(0, 0)$，函数值都为 $z(0, 0)=0$。对于 $z=xy$，在许多点函数值大于 0，例如 $z(1, 1)>0$，也有许多点上函数值小于 0，例如 $z(-1, 1)<0$。由此可见驻点不是极值点；对于 $z=x^2+y^2$，显然驻点是极小值点，也是最小值点。判断二元函数的驻点是否是极值点，比一元函数更麻烦一些，有兴趣的读者可参阅[6-2][6-3]。在实际问题中如果只有唯一的驻点，只要问题合理，驻点通常就是所需的极值点。回到【例 6-6】，

$$\frac{\partial A}{\partial x}=3\sin\alpha-4x\sin\alpha+x\sin 2\alpha=\sin\alpha(3-4x+2x\cos\alpha)=0$$

解出 $\cos\alpha=2-\frac{3}{2x}$，代入

$$\frac{\partial A}{\partial \alpha}=3x\cos\alpha-2x^2\cos\alpha+x^2\cos 2\alpha=0$$

利用三角公式 $\cos 2\alpha=2\cos^2\alpha-1$，得到

$$2x\left(2-\frac{3}{2x}\right)^2+(3-2x)\left(2-\frac{3}{2x}\right)-x=0$$

解出 $x=1$，$\cos\alpha=\frac{1}{2}$，$\alpha=60^\circ$，只有一个驻点，就是所求的极大值点。此时截面积的极大值也即最大值为 $A_{\max}=\frac{3}{4}\sqrt{3}\approx 1.299\ \mathrm{m}^2$。

下面来解决“绪论”中的一个优化问题。

【例 0-1】 某企业亏损 1050000 元。经产业结构调整，确定只生产两种支柱产品，设其产量分别为 x 和 y 件，成本函数包括固定成本 250000 元以及可变成本 $x^2+2xy+y^2$ 元。设两种产品的销售价格为 P_1 元和 P_2 元，根据市场调查，两种产品的需求函数分别是

$$x=2650-P_1,\ y=1040-0.25P_2$$

如何确定两种产品的生产水平，使得总利润最大？如果生产销售周期为 1 年，试问一年后能扭亏为盈否？

解：不难建立利润 L 对两种产品数量的依赖关系：

$$\text{总利润 } L=\text{总销售收入 } R-\text{总成本 } C$$

$$R=xP_1+yP_2=x(2650-x)+y(4160-4y)$$

$$C=x^2+2xy+y^2+250000$$

所以

$$\begin{aligned}L&=x(2650-x)+y(4160-4y)-(x^2+2xy+y^2+250000)\\&=-2x^2-2xy-5y^2+2650x+4160y-250000\end{aligned}$$

从给定条件看，有如下约束：$0\leqslant x\leqslant 2650$，$0\leqslant y\leqslant 1040$，而且 x 和 y 是整数。

$$\frac{\partial L}{\partial x}=-4x-2y+2650=0,\ \frac{\partial L}{\partial y}=-2x-10y+4160=0$$

不难解出，当 $x=505$ 件，$y=315$ 件时，总利润 L 最大，达到 1074325 元，可以扭亏为盈。

6.1.5 最优拟合直线——最小二乘法

求出实验点的最优拟合直线，是上述二元函数极值方法的重要应用。先考察一个经济学中的“恩格尔系数”问题。[6-4]

1857 年德国统计学家和经济学家恩格尔（Ernst Engel，1821-1896）对比利时不同收入的家庭消费情况进行了调查，研究了收入增加对消费需求支出构成的影响，提出了带有规律性的原理，由此被命名为“恩格尔定律”，其主要内容是指一个家庭收入越少，用于购买生存性的食物的支出在家庭收入中所占的比重就越大。对一个国家而言，一个国家越贫穷，每个国民的平均支出中，用来购买食物的费用所占比例就越大。

定义 6.6 恩格尔系数的定义为

$$E=\frac{F}{Q}\times 100\% \tag{6.10}$$

其中 F 为食物支出金额，Q 为总收入金额。

恩格尔系数的另一种定义是 6.10 式中的 Q 为总支出金额。因为总收入和总支出密切相关，用这两种定义来研究社会，基本一致。

恩格尔系数是国际上通用的衡量居民生活水平高低的一项重要指标，一般随居民家庭收入和生活水平的提高而下降。根据联合国粮农组织的标准：$E\geqslant 60\%$～贫困；$E\in[50\%, 59\%]$～温饱；$E\in[40\%, 49\%]$～小康；$E\in[30\%, 39\%]$～富裕；$E<30\%$～最富裕。从 1978 年到 2009 年，中国城镇和农村居民家庭恩格尔系数已由 57.5%和 67.7%分别下降到 36.5%和 41.0%，可见中国改革开放的巨大成就。

恩格尔 1821-1896

人们更感兴趣的是，能否根据恩格尔系数变化的历史数据，寻找出恩格尔系数和总收入的关系，以便预测未来的消费结构。

【例 6-8】 为便于说明，我们用一组实验性数据，即抽样调查某地区 8 户居民家庭的人均月收入和食品消费，见表 6.3。

表 6.3

编　号	1	2	3	4	5	6	7	8
Q/元	3060	2880	3060	2730	4140	1620	2430	1920
F/元	810	780	840	810	980	570	810	600
E/%	26.5	27.1	27.4	29.7	23.7	35.2	33.3	31.3

能否通过这些数据，得到 F 和 Q 的关系？比如当总收入达到 6000 元、8000 元或 10000 元时，预测食品支出将是多少？届时恩格尔系数又是多少？将各家的(F, Q)值描点，由图 6.8 所示。

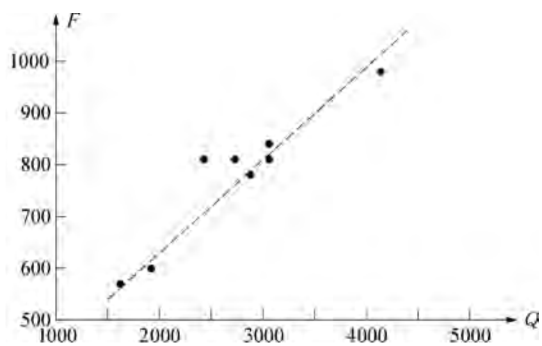


图 6.8

从图上可见，F 和 Q 的关系分布在一条直线(用虚线表示)上下。如果能求出这条直线，就可以进行预测了。

我们从一般的数学框架出发。设有两个变量 x 和 y，给定 n 组数据 (x_i, y_i) $(i = 1, 2, \cdots, n)$。希望找到一条最佳拟合直线

$$\hat{y} = kx + b \tag{6.11}$$

将 x_i 代入上式，得 $\hat{y}_i = kx_i + b$，$\hat{y}_i$ 自然不一定和 y_i 重合。记 n 个点误差平方和为

$$W = \sum_{i=1}^{n} (y_i - kx_i - b)^2 \tag{6.12}$$

显然,W 是 k 和 b 的二元函数。根据定理 6.5,要使 W 最小,必须

$$\frac{\partial W}{\partial k} = 0,\ 即 \sum_{i=1}^{n} 2(y_i - kx_i - b)(-x_i) = 0$$

$$\frac{\partial W}{\partial b} = 0,\ 即 \sum_{i=1}^{n} 2(y_i - kx_i - b)(-1) = 0$$

经整理,

$$k\sum x_i^2 + b\sum x_i = \sum x_i y_i$$

$$k\sum x_i + nb = \sum y_i$$

解此关于 k 和 b 的二元二次代数方程组,解得

$$k = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2},\ b = \frac{1}{n}\left(\sum y_i - k\sum x_i\right)$$

在实际计算中,引入算术平均值

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{6.13}$$

则

$$k = \frac{\sum x_i y_i - n\overline{x}\,\overline{y}}{\sum x_i^2 - n\overline{x}^2},\ b = \bar{y} - k\bar{x} \tag{6.14}$$

下面按 6.13 和 6.14 式求出【例 6-8】的最佳拟合直线。这里我们把 Q 和 F 分别看作公式中的 x 和 y。计算过程见表 6.4。

表 6.4

编号 i	月收入 Q_i	食品支出 F_i	$\sum Q_iF_i$	$\sum Q_i^2$	$\sum F_i^2$
1	3060	810	2478600	9363600	656100
2	2880	780	2246400	8294400	608400
3	3060	840	2570400	9363600	409600
4	2730	810	2211300	7452900	656100

续 表

编号 i	月收入 Q_i	食品支出 F_i	$\sum Q_iF_i$	$\sum Q_i^2$	$\sum F_i^2$
5	4140	980	4057200	17139600	960400
6	1620	570	923400	2624400	324900
7	2430	810	1968300	5904900	656100
8	1920	600	1152000	3686400	360000
求和	21840	6200	17607600	63829800	4631600
算术平均	2730	775			

代入 6.14，

$$k=\frac{17607600-8\times2730\times775}{63829800-8\times2730^2}=0.16203109$$

$$b=775-0.16203109\times2730=332.7$$

得到最佳拟合直线

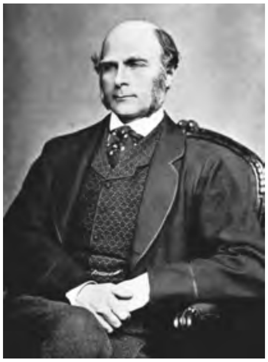
$$\hat{F}=0.162Q+332.7$$

当月收入达到 $Q=6000$ 元、8000 元、10000 元时，预测食品支出分别为 1305 元、1628.7 元、1952.7 元，恩格尔系数则分别降为 21.75%、20.36%、16.29%。

如果原始数据非常散乱，并没有直线趋势，按 6.13 式和 6.14 式也可以构造出形如 6.11 式的直线，这种拟合当然没有意义。进一步讨论需要统计学的知识，可以参阅[6-5]。在统计学中，上述拟合方法称为“回归分析”。“回归（regression）”一词最先由英国人加尔顿（Francis Galton，1882-1911）引入[6-6]。

他研究父母身高与子女身高的关系。一方面，父母身高，子女也高。另一方面，子女辈的平均身高却“回归”于全体人口的平均身高，他的原话是：“Regression to mediocrity（回归到平凡）。”加氏本人并没有用数据证实他的理论，他的朋友统计学家皮尔逊（Karl Pearson，1857-1936）收集了一千多名男子的身高记录，发现一个父辈身矮的群体，子辈平均身高高于父辈；但一个父辈身高的群体，子辈平均身高却低于父辈。子辈总体仍“回归”到平均身高。这就证实了加氏理论。[6-7]皮氏是一位兴趣广泛、涉猎很广的学者。他的时空

相对思想对爱因斯坦创立相对论有很大启发，他还研究过反物质、四维空间、时间反演等问题。



加尔顿　1882-1911

皮尔逊　1857-1936

回归分析在自然科学、工程技术、管理科学和社会科学等领域有非常广泛的应用[6-8]，也是计量经济学的重要工具[6-9]。不仅有直线拟合，还有曲线拟合；不仅有一元回归，还有多元回归。为了方便使用，许多应用软件都设置了回归分析的程序包，简单的如 excel，较专业的有 SPSS、SAS、EVIEWS，等等。使用者只要输入数据，软件自动输出回归直线或曲线的系数，以及拟合的精度[6-10]。

6.2　不用微积分的直接方法

任何方法都有局限性，用微分学方法并不能解决所有的优化问题。有的问题即使可以解决，也极其繁琐；有时初等数学方法反而更容易。

【例 6-9】　甲用户和乙用户（图 6.9 中 M 点和 N 点）共用一台变压器，A，B 分别为 M，N 在输电线上的垂足。已知 $MA = 1\,\text{km}$，$NB = 1.5\,\text{km}$，$AB = 3\,\text{km}$，问变压器设在输电干线何处，所需输电线最短？

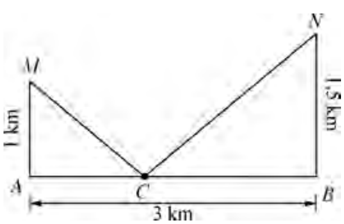


图 6.9

解：设变压器设在 C 点，优化变量为 $x=AC$，不难得到目标函数(即总输电线长)为

$$S=\sqrt{1^2+x^2}+\sqrt{1.5^2+(3-x)^2},\ (0\leqslant x\leqslant 3) \tag{6.15}$$

按照前述微分学方法，先求驻点

$$\frac{\mathrm{d}S}{\mathrm{d}x}=\frac{2x}{2\sqrt{1+x^2}}+\frac{2(3-x)(-1)}{2\sqrt{2.25+(3-x)^2}}=0$$

$$x\sqrt{2.25+(3-x)^2}=(3-x)\sqrt{1^2+x^2},$$

$$x^2(2.25+(3-x)^2)=(3-x)^2(1+x^2)$$

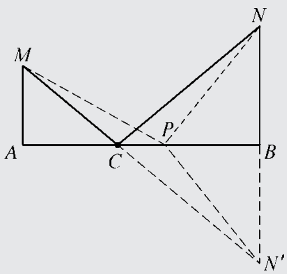


图 6.10

化简，得到一个一元二次代数方程：$1.25x^2+6x-9=0$，解出 $x=1.2$，唯一驻点就是极小值点，这时总输电线长为 $S_{\min}=S(1.2)\approx 3.905\ \mathrm{km}$。因此变压器设在输电干线离 A 点 1.2 km 处，总输电线最短。

但是这个题目用平面几何的方法更为简单。从图 6.10 可见，在输电线另侧找到 N 用户位置的对称点 N'，线段 MN' 和 AB 的交点 C 就是最佳选择。假设在干线上另取任意点 P，根据对称性和"两点之间直线最短"的性质，

$$MC+CN=MC+CN'=MN'<MP+PN'=MP+PN$$

可见变压器设置 C 处，总输电线最短。由于 $\Delta MAC\backsim\Delta N'BC$，利用"相似三角形对应边成比例"的性质，

$$\frac{CA}{CB}=\frac{MA}{N'B},\ 即\ \frac{x}{3-x}=\frac{1}{1.5}$$

容易解出 $x=1.2$，与前面的结果相同，显然要简便得多。

6.2.1 优选法

微分学方法求极值还有更大的局限，那就是目标函数必须用解析形式表达，否则无法求导数。许多实际问题的函数关系非常复杂，给出优化变量的一个值之后，必须通过试验，才得到对应的函数值。比如种粮食，如果把化肥施量作为

优化变量,粮食产量作为目标函数的因变量,很难写出精确度函数表达式,只能按实际收成确定函数值。对于这类优化问题,必须用其他方法。

下面着重介绍避开微积分的“优选法”,它是一种合理安排试验,以尽可能少的试验次数尽快找到生产和科学实验中最优方案的科学方法。

1953年美国数学家基弗(Jack C. Kiefer,1924-1981)提出单因素的“黄金分割法”[6-11],1967年美国运筹学家华尔德(Douglass J. Wilde)和皮特勒(Charles. S. Beightler)阐述了数学理论基础[6-12]。20世纪六七十年代,我国数学大师华罗庚(1910-1985)为普及推广“黄金分割法”等优化方法作出了巨大贡献。[6-13]他深入厂矿农村,在新产品、新工艺研究、仪表、设备调试等方面采用优选法,能以较少的实验次数迅速找到较优方案,在不增加设备、物资、人力和原材料的条件下,缩短工期、提高产量和质量,降低成本。“优选法”的名称就是他首先提出的。华罗庚本人还撰写了科普读物《优选法平话及其补充》和科学专著《优选学》[6-14]。

华罗庚自学成才,先后在清华大学、剑桥大学学习进修,曾任西南联大、美国普林斯顿大学和伊利诺大学教授。他是中国解析数论、多复变函数、典型群等许多领域的创始人或开拓者,有许多以他命名的数学理论和方法的成果。


华罗庚在车间、田野推广“优选法”

华罗庚　1910-1985

丘成桐　1949-

1985 年 6 月 12 日，华罗庚应邀到日本东京大学作学术报告。原定 45 分钟的报告在经久不息的掌声中被延长到一个多小时。当他结束讲话时，突然心脏病发作倒在讲台上。他用行动实践了自己的诺言："最大的希望就是工作到生命的最后一刻。"2010 年许多中外数学家撰文纪念华先生百年冥寿[6-15]。菲尔茨奖获得者丘成桐还赋词一首[6-16]：

永遇乐·华罗庚教授百岁冥辰纪念

家国飘零，关山难越，剑桥归处。翠老春湖，滇池絮落，豪杰知几许？
克难时节，干云意气，任他暴风横雨。照灵光，飞扬怒马，文章独擅俦侣。
神州再造，飞廻头雁，子弟得教七五。复变多元，堆垒难绝，矻矻求新路。
东游憔悴，高谈未尽，忍乘黄鹤归去。而今算，星沉素数，难忘隽语。

1）单因素黄金分割法(0.618 法)

当目标是一元函数 $y = f(x)$ 时，称为"单因素优选法"。仍从一个实际问题出发。

【例 6-10】 某企业生产绝缘材料，通过加催化剂以提高耐热温度。经验表明，催化剂过多过少效果都不佳。现将催化剂的用量作为优化变量，绝缘材料的耐热温度作为目标函数的因变量，如何找出催化剂的最优用量？经验告知，催化剂的用量在 1 g 到 100 g 之间，希望精确到 1 g。

显然，这个目标函数是写不出解析表达式的。假设做一次试验，成本是 1000 元，需时 5 小时。最简单的方法是"穷举法"或"均分法"，即做 100 次试验，每次催化剂的用量从 1 g 到 100 g，这样，所花成本为 $1000 \times 100 = 100000$ 元，所需时间为 $5 \times 100 = 500$ 小时。如果一天工作 8 小时，那么为了得到最优值，需花十万元和两个多月才能完成试验；即使三班倒，24 小时连续试验，也要 21 天。能不能减少试验次数，既降低成本，又节省时间？

下面介绍"黄金分割法"的步骤，可以大大减少试验次数。"绪论"中提到过"黄金分割数"：

$$\alpha = \frac{\sqrt{5}-1}{2} = 0.618033988\cdots \approx 0.618 \tag{6.16}$$

设优化变量 x 的范围是 $(a,\ b)$，目标函数为 $f(x)$。

“黄金分割法”的方法如下：

第1个试验点：

$$x_1 = a + \alpha(b-a) \approx a + 0.618(b-a) \tag{6.17}$$

第2个试验点：

$$x_2 = a + b - x_1 \approx a + 0.382(b-a) \tag{6.18}$$

用 $f(x_1)$ 和 $f(x_2)$ 分别表示目标函数在 x_1 和 x_2 的试验结果。如果 $f(x_1)$ 比 $f(x_2)$ 好，那么把试验范围 (a, x_2) 划去，剩下 (x_2, b)，这时 x_2 为“小点”，b 为“大点”，x_1 为“内点”；如果 $f(x_2)$ 比 $f(x_1)$ 好，那么把试验范围 (x_1, b) 划去，剩下 (a, x_1)，这时 a 是“小点”，x_1 是“大点”，x_2 为“内点”。

从第3个试验点起，新的点可用下面的公式计算：

$$\text{新点} = \text{小点} + \text{大点} - \text{内点} \tag{6.19}$$

以此类推，每做一次试验，可以淘汰试验范围的 $1-0.618=0.382$ 倍，直到达到所需精度为止。我们回到【例6-10】。那里 $a=0$，$b=100$，则 $x_1=61.8\approx 62$ g，$x_2=38.2\approx 38$ g。这样就剩下(0，62)或(32，100)的试验范围。以后每做一次试验，淘汰0.382倍试验范围。

做2次试验后，剩下 $100\times 0.618\approx 61.8\approx 62$；

做3次试验后，剩下 $100\times 0.618\times 0.618\approx 38.2\approx 38$；

做4次试验后，剩下 $38.2\times 0.618\approx 23.6\approx 24$；

做5次试验后，剩下 $23.6\times 0.618\approx 14.6\approx 15$；

做6次试验后，剩下 $14.6\times 0.618\approx 9.01\approx 9$；

做7次试验后，剩下 $9.01\times 0.618\approx 5.6\approx 6$；

做8次试验后，剩下 $5.6\times 0.618\approx 3.4\approx 3$；

做9次试验后，剩下 $3.4\times 0.618\approx 2.1\approx 2$；

做10次试验后，剩下 $2.1\times 0.618\approx 1.3\approx 1$

可见做10次试验后，精度已到1 g。最多做11次试验足够了。成本和时间都只有“穷举法”的十分之一！

一般说，如果优化变量取值范围是 (a, b)，区间长度为 $I_0=b-a$，那么做 k 次试验后剩下的试验范围的区间长度为

$$I_k = 0.618^{k-1} I_0 \tag{6.20}$$

上式可以决定需要做几次试验。如果试验精度为 ε,则由 $I_k = 0.618^{k-1} I_0 \leqslant \varepsilon$,解出

$$k \geqslant 1 + \frac{\ln \frac{\varepsilon}{I_0}}{\ln 0.618} \tag{6.21}$$

在【例 6-10】中，$I_0 = 100 - 0 = 100$，$\varepsilon = 1$，$k \geqslant 1 + \frac{-4.6052}{-0.4813} = 10.57 \approx 11$，与前面结果相同。

我们再以【例 6-9】为例，$a = 0$,$b = 3$，假定精度为 0.1。按均分法，从 $x = 0.1$ 开始，以 0.1 为间隔，到 $x = 2.9$ 为止，需计算目标函数 6.15 式 29 次。现用黄金分割法优选。按 6.21 式，做 8 次计算即可。

$$x_1 = 0.618 \times 3 = 1.854 \approx 1.9,\ S(1.9) = 4.007$$

$$x_2 = 0.382 \times 3 = 1.146 \approx 1.1,\ S(1.1) = 3.907$$

因 $S(1.1) < S(1.9)$，舍去(1.9，3)，在(0，1.9)继续试验。按 6.19 式，得到新的试验点

$$x_3 = 0 + 1.9 - 1.1 = 0.8,\ S(0.8) = 3.943 > S(1.1)$$

舍去(0，0.8)，在(0.8，1.9)继续试验

$$x_4 = 0.8 + 1.9 - 1.1 = 1.6,\ S(1.6) = 3.937 > S(1.1)$$

舍去(1.5，1.9)，在(0.8，1.5)继续试验

$$x_5 = 0.8 + 1.5 - 1.1 = 1.2,\ S(1.2) = 3.905 < S(1.1)$$

舍去(0.8，1.1)，在(1.1，1.5)继续试验

$$x_6 = 1.1 + 1.5 - 1.2 = 1.4,\ S(1.4) = 3.913 > S(1.1)$$

舍去(1.4，1.5)，在(1.1，1.4)继续试验

$$x_7 = 1.1 + 1.4 - 1.2 = 1.3,\ S(1.3) = 3.907 > S(1.1)$$

舍去(1.3，1.4)，留下试验区间是(1.1，1.3)。这时不必做第 8 次试验了(为什么?)，在精度为 0.1 的标准下，可以看作 $x_5 = 1.2$ 为最小值点，$S(1.2) = 3.905$

为最小值,得到的结果与前面用微分方法和几何方法完全一致。

华罗庚先生曾形象地用折纸方法推广黄金分割法,为了让大家更好记,他称之为“0.618 法”。按试验范围剪下一张纸条,画好刻度(【例 6-9】中就是[0, 3])。第一个试验点 x_1(【例 6-9】中 $x_1=1.9$)也画上去。将纸条对折,x_1 的对称点就是第二个试验点 x_2。以后,每次试验将区间对折,内点的对称点就是新的试验点。直到纸条剩下的试验区间不超过精度范围,留下的就是最优点。图 6.11 显示了用折纸方法解【例 6-9】的过程。

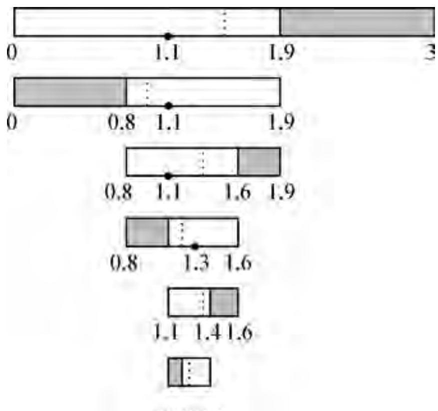


图 6.11

注意,以上讨论都蕴含着一个假定,目标函数在所求范围内只有唯一的最值点,或者说是“单峰函数”。对于“多峰函数”,有兴趣的读者可以参考[6-13][6-15],这里不进一步讨论了。

“黄金分割法”的理论基础并不高深。假定试验范围为(0, 1),第 1 个试验点为 $x_1=\alpha$,第 2 个试验点为其对称点 $x_2=1-\alpha$(图 6.12)。

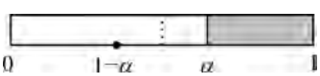


图 6.12

裁去 $1-\alpha$ 的长度之后,(例如裁去 $(\alpha, 1)$),在 $(0, \alpha)$ 中,$1-\alpha$ 的位置相当于(0, 1)中 α 的位置,成立如下比例式 $\frac{1-\alpha}{\alpha}=\frac{\alpha}{1}$,得到一个一元二次代数方程

$$\alpha^2+\alpha-1=0 \tag{6.22}$$

其正根如 6.16 式所示。

2) 单因素分数法

我们在绪论中介绍过斐波那契数列:

$$f_1=1,\ f_2=1,\ \cdots,\ f_n=f_{n-2}+f_{n-1}(n\geqslant 3)$$

华罗庚先生在《优选学》[6-14]中严格证明了它的奇妙性质:

$$\lim_{n\to\infty}\frac{f_{n-1}}{f_n}=\frac{\sqrt{5}-1}{2} \tag{6.23}$$

$\frac{f_{n-1}}{f_n}$ 的前几项是 $\frac{1}{1}$，$\frac{1}{2}$，$\frac{2}{3}$，$\frac{3}{5}$，$\frac{5}{8}$，$\frac{8}{13}$，$\frac{13}{21}$，$\frac{21}{34}$，…。对于优化变量不是连续变化的情况下，可以利用上述性质，而采用“分数法”进行优选。比如一台机床只有 12 挡速度，工人想测试哪挡速度综合效果最好。华先生在车床旁用“火柴”或零件向工人做了讲述[6-14]：因为 12 接近 $f_7=13$，建议在第 8 挡做第一次试验，再在其对称点第 5 挡做第 2 次试验。如果第 8 挡好，甩掉第 1-5 挡，在第 6-12 挡中，找到第 8 挡的对称点第 10 挡。如果还是第 8 挡好，甩掉第 10-12 挡，留下第 6-9 挡。第 8 挡的对称点是第 7 挡，如果第 7 挡好，留下第6-7 挡。最后试验第 6 挡，6、7 两挡的优者就是最好的一挡。共做了 4 次测试，而不是 12 次测试。上述过程由图 6.13 所示。

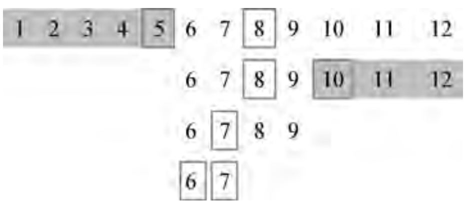


图 6.13

3）双因素黄金分割法

当目标是二元函数 $z=f(x, y)$ 时，称为“双因素优选法”。我们仍假定目标函数是“单峰”的，即在试验范围内只有唯一的最值点。单因素的黄金分割法可以推广到双因素。方法有许多种，下面介绍两种比较容易操作的方法。

（1）对开法　假设目标函数为 $z=f(x, y)$，优选范围是 $a<x<b$，$c<y<d$，求最大值。选一张长方形纸，长宽分别代表 x，y，其面积为

$$A_0=(b-a)(d-c)。$$

第 1 步，左右对折，在中线 $x=\frac{a+b}{2}$ 上，按单因素法在 P_1 点找到最大值；上下对折，在中线 $y=\frac{c+d}{2}$ 上，按单因素法在 Q_1 点找到最大值。比较目标函数在 P_1 点和 Q_1 点的函数值。如果 P_1 点函数值大于 Q_1 点函数值，裁掉下半张纸；如果 P_1 点函数值小于 Q_1 点函数值，就裁去左半张纸。不妨假定 P_1 点函数值大于 Q_1 点函数值，在 $a<x<b$，$\frac{c+d}{2}<y<d$ 继续试验，这时面积为

$$A_1=\frac{A_0}{2}；$$

第 2 步，留下的半张纸，再上下对折，在中线 $y=\frac{c+3d}{4}$ 上按单因素法找到最大值点 Q_2，比较目标函数在 P_1 点和 Q_2 点的函数值。如果 Q_2 点的函数值大于 P_1 点函数值，就裁去右半张，在 $a<x<\frac{a+b}{2}$，$\frac{c+d}{2}<y<d$ 继续试验，这时面积为 $A_2=\frac{A_0}{4}$。按上述步骤做下去，直到纸张剩下的面积小于或等于指定精度时为止。（图 6.14）

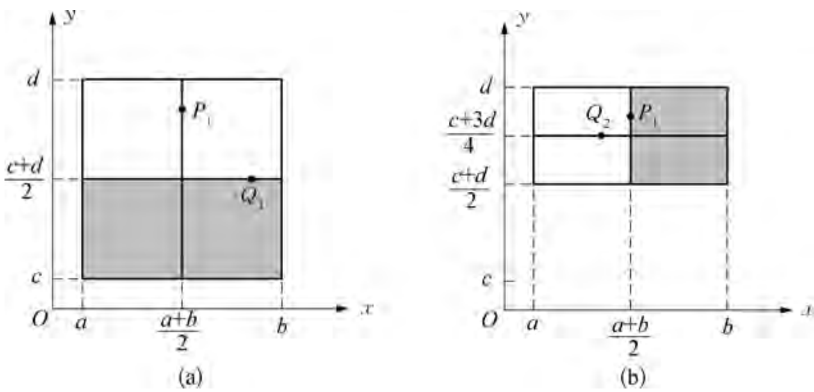


图 6.14

（2）双向 0.618 法　第 1 步，在上述长方形纸上，找到 4 个对称点：$P_1(x_1,y_1)$，$P_2(x_1,y_2)$，$P_3(x_2,y_1)$，$P_4(x_2,y_2)$，其中

$$x_1=a+0.618(b-a),\ y_1=c+0.618(d-c)$$

$$x_2=a+0.382(b-a),\ y_2=c+0.382(d-c) \tag{6.24}$$

第 2 步，比较目标函数在上述 4 点的数值。如果 P_1 点函数值最大，则保留 $x_2<x<b$，$y_2<y<d$（如果 P_2 点函数值最大，则保留 $x_2<x<b$，$c<y<y_1$；如果 P_3 点函数值最大，则保留 $a<x<x_1$，$y_2<y<d$；如果 P_4 点函数值最大，则保留 $a<x<x_1$，$c<y<y_1$）。保留长方形的面积是 $A_1=0.618^2A_0=0.382A_0$。

在留下的长方形中找到 P_1 的 3 个新的对称点 $P_5(x_1,y_3)$，$P_6(x_3,y_1)$ 和 $P_7(x_3,y_3)$，其中

$$x_3=x_2+b-x_1,\ y_3=y_2+d-y_1$$

如果 P_6 点函数值最大，则保留左上角长方形：$x_2 < x < x_1$，$y_3 < y < d$，这时长方形面积 $A_2 = 0.382A_1 = 0.382^2 A_0 \approx 0.146A_0$。

类似前述步骤，第 k 步留下的小长方形面积为 $A_{k-1} = 0.382^k A_0$，直到小于或等于指定精度时为止。(图 6.15)

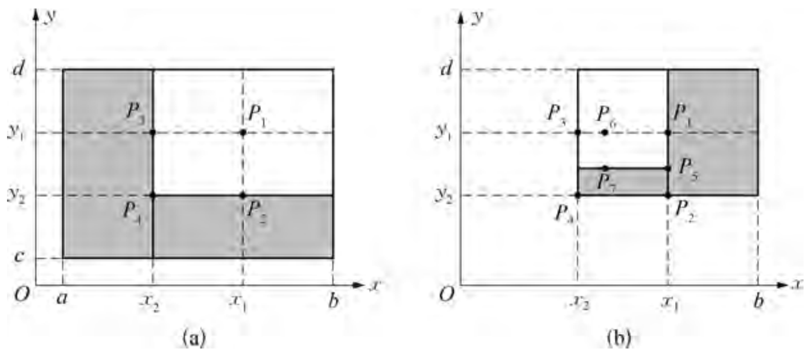


图 6.15

无论是单因素还是多因素优化，还有许多其他方法。每种方法都有优点，也都有局限。有兴趣的读者进一步探讨可参阅[6-13][6-14]。

6.2.2 统筹方法

统筹方法是通过重组、打乱等手段改变原本的固有办事格式，优化办事效率的一种办事方法。在企业管理和基本建设中，以及关系复杂的科研项目或文化艺术项目的组织与管理中，都可以应用。华罗庚先生在推广优选法的同时，还大力推广另一种数学方法："统筹法"。他也写了著名的科普著作《统筹方法平话及补充》[6-14]。下面作一简单介绍。先引用华先生的入门例子："想泡壶茶喝。当时的情况是：开水没有；开水壶要洗，茶壶茶杯要洗；火已生了，茶叶也有了。怎么办?"

办法甲：洗好开水壶，灌上凉水，放在火上；在等待水开的时候，洗茶壶、洗茶杯、拿茶叶；等水开了，泡茶喝。

办法乙：先做好一些准备工作，洗开水壶，洗茶壶茶杯，拿茶叶；一切就绪，灌水烧水；坐待水开了泡茶喝。

办法丙：洗净开水壶，灌上凉水，放在火上，坐待水开；水开了之后，急急忙忙找茶叶，洗茶壶茶杯，泡茶喝。

哪一种办法省时间？我们能一眼看出第一种办法好，后两种办法都窝了工。

这是小事，但这是引子，可以引出生产管理等方面的有用的方法来。

在这个过程中，一共有 5 项“活动(activity)”或“事件(event)”：① 洗水壶，灌凉水(1 分钟)，② 洗茶壶茶杯(2 分钟)，③ 拿茶叶(2 分钟)，④ 烧开水(15 分钟)，⑤ 泡茶。

有的活动有先后关系，成为“串行”关系，如“烧开水”前必须“灌凉水”，“灌凉水”前必须“洗水壶”；所以①→②→③就是串行关系；同样，“拿茶叶”前必须“洗茶壶茶杯”，所以④→⑤也是串行关系。但是上面“两串”活动却没有先后关系，这就是“并行”关系。可是“泡茶”却必须在这两串活动之后。我们把所有活动用箭头画出“流程图”，箭头上的数字表示完成前项活动的时间。

上面 3 种办法的流程图见图 6.16。

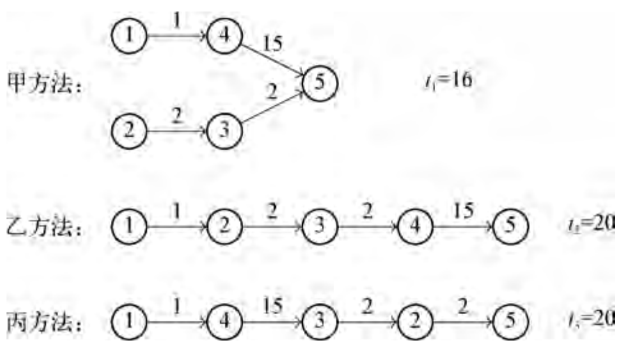


图 6.16

从这个图上可以一眼看出，办法甲总共要 16 分钟，而办法乙、丙都需要 20 分钟。如果要缩短工时、提高工作效率，应当主要抓烧开水这个环节，而不是抓拿茶叶等环节。同时，洗茶壶茶杯、拿茶叶总共不过 4 分钟，大可利用烧水的时间来做。

对于活动不多的简单情况，似乎常人都能处理；但是，在近代工业错综复杂的工艺过程中，在一个重大项目的执行和管理过程中，就不像泡茶喝这么简单了。这个方法最早是美国海军于 1957 年进行“北极星”核潜艇建造时开发的，称为“计划评审技术(Program Evaluation and Review Technique)”，缩写为 PERT[6-17]。该项目共有 3000 多项任务，参与厂商达 11000 多家。这个方法的主要着眼点是找到完成复杂项目的最短时间。在许多民用项目中也成功地得到应用，如 1968 年在法国格勒诺布尔(Grenoble)举行的冬季奥运会开幕式就成功应用了 PERT[6-18]。也在 1957 年，美国杜邦(DuPont)公司为了协调公司内部不同业务部门

的工作，研制出思路相近的“关键线路法（Critical Path Method）”，缩写为 CPM[6-19]。他们首次将此方法用于一家化工厂的筹建，缩短停工时间 47 小时。

统筹法的关键就是从流程图中找到“关键线路”，华先生称之为“主要矛盾线”，取自毛泽东《矛盾论》中的论断：“研究任何过程，如果是存在两个以上矛盾的复杂过程的话，就要用全力去找出它的主要矛盾。捉住了这个主要矛盾，一切问题就迎刃而解了。”[6-20]图 6.16 中的甲方法的主要矛盾线是①→④→⑤，共需 16 分钟。它有两个特点，第一，主要矛盾线所需时间就是整个项目所需时间。如果要缩短时间，关键要缩短主要矛盾线上各环节的时间。上面例子中如能改进加热方式，缩短把水烧开的时间，就能提早喝茶。第二，主要矛盾线上任一环节延误，就会推迟整个活动；但非主要矛盾线上的环节即使有所延误，不一定推迟整个活动。上一例子中，即使洗茶壶茶杯和拿茶叶从 4 分钟延长到 16 分钟，仍不影响喝茶时间。

在复杂的项目中，一方面，要找出主要矛盾线，并全力保证主要矛盾线上各环节如期或提前完成；另一方面，对每一项活动，给出“最早开工时间”（即不可能提早的时间）和“最迟开工时间”（即不能延后的时间），使各参与者心中有数，既不误时误事，又不浪费时间。

在主要矛盾线上每项活动的“最早开工时间”和“最迟开工时间”是相同的。在喝茶例子中，“洗水壶，灌凉水”的两个时间都是 0，“烧开水”的两个时间都是 1，“泡茶”的两个时间都是 16。

在非主要矛盾线上的活动，“最早开工时间”和“最迟开工时间”可以不同，两者之差称为“松弛”。在喝茶例子中，“拿茶叶”的“最早开工时间”是 2，“最迟开工时间”是 14，“洗茶壶茶杯”的“最早开工时间”是 0，“最迟开工时间”是 12；这两项活动都有 12 分钟的“松弛”。正由于存在“松弛”，在非主要矛盾线上就有潜力可挖。

下面我们举一个稍微复杂的例子进一步介绍。

【例 6-11】已知某机床检修的工序明细表如表 6.5 所示。找出主要矛盾线，并计算工程完成的时间。

表 6.5

工　序	A（拆卸机床）	B（检查更换零部件）	C（检查机床主体）	D（更换主轴）	E（安装零部件）	F（总装调试）	G（验收）
工时	5	4	3	5	6	7	1
紧前工序	/	/	A	C	B，C	D	E，F

流程图如图 6.17 所示。

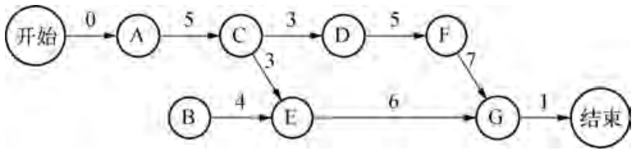


图 6.17

不难看出，主要矛盾线是：开始→A→C→D→F→G→结束，共需 21 工时。由于工序 B 和 E 不在主要矛盾线上，可以有所延误。工序 E 最早开工时间为 8（＝5＋3），最迟可以从工序 A 开工后 14（＝21－6－1）小时后开始，有 8 小时松弛；工序 B 最早开工时间为 0，最迟可以从工序 A 开工后 10（＝14－4）小时后开始，有 10 小时松弛。表 6.6 列出了各工序的“最早开工时间”、“最迟开工时间”和“松弛”。

表 6.6

工　序	*A*	*B*	*C*	*D*	*E*	*F*	*G*
最早开工	0	0	5	8	8	13	20
最迟开工	0	8	5	8	14	13	20
松　弛	0	8	0	0	6	0	0

如果要提高工效，必须缩短“拆卸机床”、“检查机床主体”、“更换主轴”和“总装调试”这 4 个工序的时间。例如把“检查更换零部件”和“安装零部件”的工人抽调一部分参加主要矛盾线某几个工序。这时要重画流程图。也要防止矫枉过正，如果抽调过多，使原来非主要矛盾线上的工序费时太长，以致增加总工时，这就得不偿失了。

以上我们讨论的例子，都是“肯定型”问题，即每项活动所需时间是肯定的。现实问题中，常有随机因素，使得各项活动的完成时间难以预料，这就是“非肯定型”问题。如何确定每项活动的完成时间？华罗庚先生有如下建议：

$$\text{预测时间}=(\text{最乐观时间}+4\times\text{最可能时间}+\text{最保守时间}) \tag{6.25}$$

优化问题和方法还有许多，有兴趣的读者可参阅有关文献[6-22][6-23][6-24]

6.2.3 线性规划简介

我们从“绪论”中的一个优化问题出发。

【例 0-2】 A，B 两个产地分别生产同一批规格产品 12 千吨和 8 千吨。而甲、乙、丙三地分别需要该产品 8 千吨、6 千吨和 6 千吨。各产销地之间每千吨的运价表如下，怎样确定运输方案，使得总运费最少？

万　元	到　甲	到　乙	到　丙
从 A	4	5	6
从 B	5	2	4

解：设从 A 运到甲地 x 千吨，运到乙地 y 千吨，运到丙地 $12-x-y$ 千吨，从 B 运到甲地 $8-x$ 千吨，运到乙地 $6-y$ 千吨，运到丙地 $8-(8-x)-(6-y)=x+y-6$ 千吨，则总运费

$$\begin{aligned}F&=4x+5y+6(12-x-y)+5(8-x)+2(6-y)+4(x+y-6)\\&=-3x+y+100\end{aligned}$$

约束条件为：$x\geqslant 0, y\geqslant 0$；$12-x-y\geqslant 0$，$8-x\geqslant 0$，$6-y\geqslant 0$，$x+y-6\geqslant 0$。

把约束条件画在坐标平面上（图 6.18）。

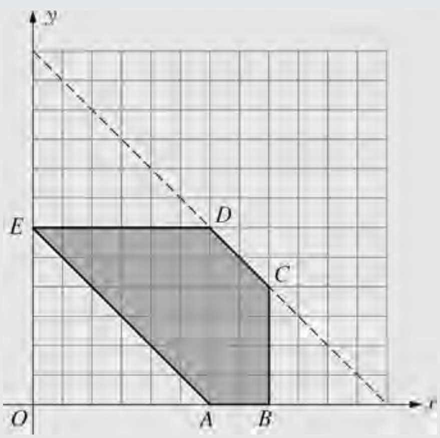


图 6.18

(x, y) 只能在多边形 $ABCDE$ 内取值。因为目标函数是线性函数，$\dfrac{\partial F}{\partial x}=-3\neq 0$，$\dfrac{\partial F}{\partial y}=1\neq 0$，根据微分法求极值原理，区域内部无极值点，最小值只能在顶点 $A(6, 0)$，$B(8, 0)$，$C(8, 4)$，$D(6, 6)$，$E(0, 6)$ 上达到。

因 $F(6, 0)=82$，$F(8, 0)=76$，$F(8, 4)=80$，$F(6, 6)=88$，$F(0, 6)=106$，所以取 $x=8$ 千吨，$y=0$ 千吨，运费为 76 万元，是最小值。

线性规划问题的一般提法是：

$$\text{目标函数(最小或最大)：} \min/\max f(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} c_i x_i \tag{6.26}$$

约束条件：

$$\begin{aligned} & x_i \geqslant 0, (i = 1, 2, \cdots, n) \\ & \sum_{j=1}^{n} a_{1j} x_j \leqslant b_1, \\ & \sum_{j=1}^{n} a_{2j} x_j \leqslant b_2, \\ & \cdots \\ & \sum_{j=1}^{n} a_{mj} x_j \leqslant b_m \end{aligned} \tag{6.27}$$

当优化变量只有两个时，可以按【例 0-2】的方法图解，如果优化变量大于等于三个时，就要用其他方法，其数学基础是线性代数[6-21]-[6-24]，本书不予细述。

6.3 思考与练习

1. 研究优化问题有什么意义？能否举一个优化的实际问题？

2. 微分学方法求最值的步骤是什么？

3. 求函数 $y = x^5 - 5x^3 + 10x - 3$ 的极值。

4. 电池的输出功率与电压 V、电阻 R、电流 I 的关系为 $P = VI - RI^2$。如果 V 和 R 为常数，求使输出功率为最大的电流值。

5. 已知某轮船的燃料费与船速的立方成正比，当船速为 10 km/h，每小时燃料费为 80 元，其他费用每小时 480 元，问船速多大时，使 20 km 航程的总费用最少？并求出在此条件下每小时的总费用。

6. 某化工厂计划制造大量长方体形状的容器，要求容积为 10 m^3。设底面材料的成本为顶面材料的 4 倍，侧面材料的成本为顶面材料的 2 倍。如何设计尺寸，使得所用材料费用最低？

7. 炼钢过程中要加化学元素来加强强度。已知炼 1 吨钢要加的某种化学元素在 50 g 和 200 g 之间，希望精确到 1 g，问用黄金分割法要做几次试验？

8. 能否用优选法验证第 6 题的结果？

9. 设某工程的工序明细表如表 6.7 所示：

表 6.7

工序代号	紧前工序	工序时间
A	/	3
B	A	2
C	A	5
D	B	1
E	B	5
F	E	1
G	A	6
H	D，C	4
I	E	4
J	F，G，H	3
K	I，J	3

画出流程图，找出主要矛盾线，计算总时间以及各工序的“最早开工时间”、“最迟开工时间”和“松弛”。

参考文献或网站

[6-1]　百度百科：拉格朗日 http：//baike.baidu.com/view/19783.htm .

[6-2]　王绵森，马知恩：《工科数学分析基础(下册)》，高等教育出版社(1998).

[6-3]　同济大学数学系：《微积分(下册)》(第三版)，高等教育出版社(2010).

[6-4]　百度百科：恩格尔系数 http：// baike.baidu.com/view/28093.htm.

[6-5]　黄良文　曾五一：《统计学原理》，中国统计出版社(2000).

[6-6]　Francis Galton：“Family Likeness in Stature(家庭身材相似性)”，Proceedings of Royal Society，Vol. 40(1886).

[6-7]　K. Pearson，A. Lee：“On the laws of inheritance(论遗传定律)”，Biometrika，Vol. 2(1903).

[6-8]　何晓群　刘文卿：《应用回归分析》(第二版)，中国人民大学出版社(2007).

[6-9]　古扎拉蒂(D. N. Gujarati)：《计量经济学(Basic Econometrics)》(第三版)林少宫译，中国人民大学出版社(1999).

[6-10] 冯力:《回归分析方法原理及 SPSS 实际操作》,中国金融出版社(2004).

[6-11] 基弗(J. Kiefer):《最大值点序列最大最小搜索(Sequential minimax search for a maximum)》, Proc. Amer. Math . Soc. No. 4(1953).

[6-12] D. J. 华尔德,C. S. 皮特勒:《优选法基础(Foundation of optimization)》(中译本),科学出版社(1978).

[6-13] 中国科学院数学研究所运筹室优选法小组:《优选法》,科学出版社(1975).

[6-14] 杨德庄:《华罗庚文集 应用数学卷 II》,科学出版社(2010).

[6-15] 丘成桐,杨乐,季理真:《传奇数学家华罗庚——纪念华罗庚诞辰 100 周年》,高等教育出版社(2010).

[6-16] 《丘成桐诗文集》,高等教育出版社(2008).

[6-17] 法查(Fazar, W):《项目评估与检查技术(Program Evaluation and Review Technique)》, *The American Statistician*, Vol. 13, No. 2, (1959).

[6-18] 1968年冬季奥林匹克官方报告)(1968 Winter Olympics official report). (English) & (French).

[6-19] 凯雷(J. Kelley), 沃克(M. Walker):《关键线路的设计与安排.(Critical-Path Planning and Scheduling)》, Proceedings of the Eastern Joint Computer Conference(1959).

[6-20] 《毛泽东选集》,人民出版社(1991).

[6-21] 马良,王波,王龙德:《基础运筹学教程》,高等教育出版社(2006).

[6-22] 陈宝林:《最优化理论与算法(第 2 版)》清华大学出版社(2005).

[6-23] 李董辉,童小娇,万中:《数值最优化算法与理论》,科学出版社(2010).

[6-24] 斯特兰(G. Strang):《线性代数导引(Introduction to Linear Algebra)》, Wellesley-Cambridge Press(1993).

选修材料之一

唐诗格律的形式体系

中国是诗歌之国。诗经、楚辞、汉赋、乐府、唐诗、宋词、元曲，直到今天的白话诗，花团锦簇，各领风骚。唐诗可以说是最光辉灿烂，影响最久远的诗体。不仅流传了大量脍炙人口的佳作，而且是今天许多人抒发感情的载体。研究任何时期的诗歌，最重要的是时代背景、地理环境、作者风格等等，研究唐诗亦然。不过，本文并不涉及这些方面，仅仅介绍唐近体诗格律的形式体系。

所谓“格律”，是指某种诗歌在格式、音律上的规定。大约在七世纪晚期，在五言古风和乐府的基础上，形成了唐近体诗，包括绝句、律诗和排律，以和古体诗相区别。后人公认是宋之问(656-712)和沈佺期(656-714)制定近体诗格律的。不过在他们之前。已有相当长的演化过程。

宋之问　656-712

沈佺期　656-714

唐诗三百首

由于近体诗短小精悍、抑扬顿挫，便于吟诵流传，在唐代就已出现大量绝句律诗，许多精品至今为人们反复咏唱。清乾隆年间孙洙(别号“蘅塘退士”)选编了一部影响很大的《唐诗三百首》[S-1]，共收入 310 首，其中五言绝句 29 首，七言

绝句51首，五言律诗80首，七言律诗53首，占总数的68.7%。其实不少乐府的格律完全和近体诗相同，加上之后，比例将更大。为行文简单，下面提到的“唐诗格律”，就是指近体诗的格律。

大家知道，数学是数与形的科学。

“自然数→小数、分数→负数→零→有理数→无理数→实数→虚数→复数”，这是“数”的逻辑发展体系；“点→线→面→体”，则是“形”的逻辑发展体系。

唐诗格律，也有一个“字(声母、韵母、声调)→词→节拍→句→诗”的逻辑发展体系。正如第二章所阐述的，数学的研究方法建立在为数不多的公理基础上，通过定义一些概念，再经过形式逻辑的推理，得到新的结果。

我认为唐诗格律也有两条“公理”：

公理1:“节拍平仄相间韵律美”;

公理2:“上下句式对仗形式美”。

在此基础上，定义“对”、“粘”、“词的门类”等，即构成唐诗体系。如果说，古代青年只能靠大量背诵来学习格律，所谓“熟读唐诗三百首，不会做诗也会吟”；那么，当代青年完全可以参照这样的逻辑发展体系来学习，当有事半功倍之效。

S.1　汉字的音韵

S.1.1　声韵与押韵

汉字的特点是单音节。例如“唐”字的拼音是 t-ang，其中 t 是声母，ang 是韵母。一个字可以没有声母，但一定有韵母。一个字只有一个音节，一个音节也只构成一个字，这是汉字和拼音文字的基本区别之一。

诗歌押韵，是指一首诗某几句的最后一个字必须有相同的韵母。下面看两首唐诗：

【例 S-1】　**登鹳雀楼**

王之涣(688-742)

白日依山尽，黄河入海流。

欲穷千里目，更上一层楼。

【例 S-2】 **下江陵**

李白(701-762)

朝辞白帝彩云间,千里江陵一日还。

两岸猿声啼不住,轻舟已过万重山。

【例 S-1】第二句和第四句的最后字分别为“流”和“楼”,韵母分别是 iu 和 ou;【例 S-2】的第一、二、四句的最后字分别为“间”、“还”和“山”,韵母都是 an 或 ian。

隋代陆法言把相同韵母的字归为同一个“韵部”,并编辑成册,称为《切韵》,以便学习、记忆和查核。小孩子启蒙时往往要求背出来。金代王文郁加以变通,形成具有 106 个韵的“平水韵”,一直沿用至明清。如果一个文人所写的诗中被发现串了韵,就是诗家大忌,会引起别人的讥笑。

王之涣　688-742

李白　701-762

李商隐　812-858

由于千百年来汉字读音的演变,有的字在唐代属于同一韵部,现在却全然不同。

【例 S-3】 **登乐游原**

李商隐(812-858)

向晚意不适,驱车登古原。

夕阳无限好,只是近黄昏。

第二句的末字为“原”,第四句的末字为“昏”,它们的韵母分别是 an 和 un。用普通话念,差别很大。可是在唐代,这两个字可以押韵。由于粤语还保留了不少唐音,如果用粤语念这两个字,还可依稀感觉,其韵母都接近于 un。相反,也有不少字,在现代普通话中有相同韵母,在唐代韵书中却分属不同韵部。例如,

“平水韵”把“东”和“冬”、“鱼”和“虞”划为不同的诗韵部的部首字，今人就很难理解了。

《红楼梦》第四十八回描写香菱学诗，她用“十四寒”韵部写了一首描写月亮的七律，薛宝钗提了意见，她便认真思索。这时，探春隔窗笑道：“菱姑娘，你闲闲吧。”香菱怔怔答道：“闲字是十五删的，错了韵了。”其实，“十五删”和“十四寒”在现代汉语中的韵母都是 an（“十五删”韵部中有部分字的韵母是 ian，如“闲”、“艰”等），可见直到曹雪芹生活的清代，文人还是恪守唐韵的。

S.1.2　声调，平与仄

汉字的另一特点是“声调”，即同一个音节，有不同的声调。做格律诗的基本功之一，就是搞清楚每个汉字属于“平声”还是“仄声”。

必须指出，唐代和现代的声调划分也有区别。“平水韵”把汉字分为“平、上、去、入”四种声调，“上、去、入”都是仄声，其中“平声”又分为“上平声”和“下平声”。影响较大的《佩文韵府》[S-2]收入 10703 个汉字，各韵目的声调及相应字数见表 S.1。

表 S.1　《佩文韵府》诗韵分布一览

声调名	上平声	下平声	上　声	去　声	入　声	总计
韵目名	东冬江支微鱼虞齐佳灰真文元寒删	先萧肴豪歌麻阳庚青蒸尤侵覃盐咸	董肿讲纸尾语麌荠蟹贿轸吻阮旱潸铣筱巧皓哿马养梗迥有寝感琰豏	送宋绛寘未御遇霁泰卦队震问愿翰谏霰啸效号箇祃漾敬径宥沁勘艳陷	屋沃觉质物月曷黠屑药陌锡职缉合叶洽	
字数	2224	2521	1855	2302	1801	10703
百分比	20.8%	23.6%	17.3%	21.5%	16.8%	100%

其中平声字占 44.4%，仄声字占 55.6%。

现代汉语则把汉字分为“第一声（阴平）、第二声（阳平）、第三声（上声）、第四声（去声）”。

唐韵的“平声”基本上属于今韵的“第一声”和“第二声”，但“上平声”和“阴平”，或“下平声”和“阳平”却并无对应关系。比如在《佩文诗韵》中，部首字“微”、

“鱼”、“虞”、“齐”、“文”、“元”、“寒”属于“上平声”，但它们都是“阳平”字；而作为“下平声”部首的“先”、“萧”、“歌”、“庚”、“青”、“蒸”、“侵”等字却是“阴平”字。好在唐诗中把汉字又分为“平、仄”两大类。押韵时只要是“平”韵，“上平”和“下平”均可。

唐韵中的“上声”及“去声”属于“仄声”，千百年来，也有演变，但总体看，和现代汉语变化不大。

问题最大的是唐韵中的“入声”。“入声”也属于“仄声”，在“仄声”压韵的诗中，往往“去声”和“上声”可以通用，而“入声”却不可与“上”、“去”通融。

【例 S-4】 **江　雪**

柳宗元(773-819)

千山鸟飞绝，万径人踪灭。

孤舟蓑笠翁，独钓寒江雪。

柳宗元　773-819

岳飞　1103-1142

这首诗就是入声韵，其中的“绝”、“灭”和“雪”属于同一入声部的韵脚字，但按现代汉语普通话的分类，则分别属于第二声、第四声和第三声。不过在不少方言中，还保留了入声。例如用吴语方言、湘语方言或粤语方言来念这首诗，就会感到强烈的诗韵。唐诗中，押入声韵的绝句或律诗并不多，但宋词中却有许多词牌，约定押入声韵。如岳飞(1103-1142)的《满江红(怒发冲冠，…)》中的“歇、烈、月、切、雪、灭、缺、血、阙”，苏轼的《念奴娇(大江东去，…)》中的“物、壁、雪、杰、发、灭、发、月”都是入声。现代人的不少作品中，仍遵循这种约定。例如毛泽东的词《满江红·和郭沫若》、《念奴娇·昆仑》等。

近人所编的《诗韵新编》[S-3]共收入 7595 个汉字(包括简体字),并按普通话将这些字分成 18 个韵目,其分布一览见表 S.2。

表 S.2　《佩文韵符》诗韵分布一览

声调 / 韵目	阴　平	阳　平	上　声	去　声	总数	入声	入声百分比
一麻	88+32	46+58	40+10	65+49	388	149	38.4%
二波	42+22	51+71	34+5	35+111	371	209	56.3%
三歌	28+14	17+59	14+5	19+91	247	169	68.4%
四皆	15+24	18+85	10+5	32+97	286	211	73.8%
五支	77+7	27+23	43+1	87+29	294	60	20.4%
六儿	0+0	4+0	7+0	2+0	13	0	0
七齐	108+37	145+62	82+9	169+129	741	237	32.0%
八微	85+0	78+0	64+0	161+0	388	0	0
九开	40+0	44+0	43+0	106+0	233	0	0
十模	87+13	104+82	107+15	124+96	628	206	32.8%
十一鱼	46+6	50+12	49+1	57+45	266	64	10.3%
十二侯	92+0	89+0	69+0	116+0	366	0	0
十三豪	152+0	150+0	107+0	110+0	523	0	0
十四寒	263+0	240+0	190+0	331+0	1024	0	0
十五痕	170+0	155+0	103+0	173+0	601	0	0
十六唐	145+0	132+0	92+0	111+0	480	0	0
十七庚	157+0	163+0	71+0	100+0	491	0	0
十八东	85+0	86+0	47+0	37+0	255	0	0
总数	1680+155 =1835	1603+452 =2055	1172+51 =1223	1835+647 =2482	7595	1305	17.2%
百分比	22.1+2.0 =24.1%	21.1+6.0 =27.1%	15.4+.7 =16.1%	24.2+8.5 =32.7%	100%	17.2%	

表中第 2、3、4、5 列中加号后面的数字为唐韵中归入入声的字数。

比较表 S.1 和表 S.2,可以看到音韵的变化趋势。唐韵的平声(上平声+下平声)占总字数的 44.4%;今韵的平声(阴平+阳平)占总字数的 51.2%,不算入声则占 43.2%,百分比变化不算大。上声的百分比由 17.3%降为 16.1%;去声的百分比则由 21.5%升至 32.7%,这是因为几乎一半的入声字(占总字数的 8.5%)归入去声,而归入上声字的只有 3.9%(仅占总字数的 0.7%)。

另一个有趣现象是，不论唐韵还是今韵，不同韵部的字数相差很大。从表S.2可见，“儿”部只有13个字，而“寒”部竟有1024个字。附录1列出了《佩文韵府》各韵部的字数，同样可以看到这一现象。古人把字数很少的韵部称为“险韵”，字数较多的韵部称为“宽韵”。有的诗人故意用险韵吟诗，以表现自己的功力。

每个汉字非平即仄，下面我们常用0表示平声，用1表示仄声。

S.2 词与节拍

S.2.1 二字词

五言或七言诗句的基本元素除了字以外就是两个字组成的词。如前例中的“白日”、“黄河”、“彩云”、“古原”、“夕阳”、“孤舟”等等。

从格律的角度看，有三种二字词值得注意。试看下面三首诗：

【例S-5】 **落 花**

李商隐

高阁客竟去，小园花乱飞。
参差连曲陌，迢递送斜晖。
肠断未忍扫，眼穿仍欲归。
芳心向春尽，所得是沾衣。

【例S-6】 **登 高**

杜甫(712-770)

风急天高猿啸哀，渚清沙白鸟飞回。
无边落木萧萧下，不尽长江滚滚来。
万里悲秋常作客，百年多病独登台。
艰难苦恨繁霜鬓，潦倒新停浊酒杯。

【例S-7】 **咏怀古迹(之二)**

杜 甫

摇落深知宋玉悲，风流儒雅亦吾师。
怅望千秋一洒泪，萧条异代不同时。

江山故宅空文藻，云雨荒台岂梦思。

最是楚宫俱泯灭，舟人指点到今疑。

（1）双声词：两个相同声母的字组成的词。如【例 S-5】中的“高（gao）阁（ge）”和“参（cen）差（ci）”；【例 S-7】的“云（yun）雨（yu）”。

（2）叠韵词：两个相同韵母的字组成的词。如【例 S-7】中的“怅（chang）望（wang）”和“萧（xiao）条（tiao）”。

（3）重叠词：两个相同的字组成的词。如【例 S-6】中的“萧萧”和“滚滚”。

杜甫　712-770

S.2.2　节拍

五言诗每句由 5 个字组成，分成 3 个节拍（或音节），即 | ** | ** | * |；

七言诗每句由 7 个字组成，分成 4 个节拍（或音节），即 | ** | ** | ** | * |。

节拍也有平仄，即最后一个字的平仄。例如“|白日|依山|尽|”有 3 个节拍，其平仄分别是：“仄平仄”（1-0-1），“|不尽|长江|滚滚|来|”有 4 个节拍，其平仄分别是：“仄平仄平”（1-0-1-0）。

S.3　绝　　句

S.3.1　五绝

五言绝句简称“**五绝**”，由四句组成。每句共 5 字，全诗共 20 字。【例 S-1，3，5】都是五绝。在音律上有如下规则：

1）押韵

第二句和第四句的最末字必须有相同韵母，称为“押韵”或“压韵”。【例 S-1，3】是二、四句押韵；【例 S-4】除二、四句外，首句也押韵。《唐诗三百首》收录 29 首五绝，首句不入韵的有 23 首，占 79.3%，可以说，多数五绝首句不入韵。

五绝以平声韵为主。【例 S-1，3】都是平声韵；而【例 S-4】是仄声韵。《唐诗

三百首》收录的 29 首五绝中，20 首为平韵，占 69%。

2）句内自对

前面已指出，五言诗每句被分为 3 个节拍，一个节拍最后字的平仄称为节拍的平仄。

为了使诗句抑扬顿挫，朗朗上口，规定每句诗的第 1 和第 2 节拍的平仄必须相反，这就是“句内自对”。以【例 S-1】为例：

“|白日|依山|尽|”：|仄仄|平平|仄|（日山尽，1-0-1）；

“|黄河|入海|流|”：|平平|仄仄|平‖（河海流，0-1-0）；

“|欲穷|千里|目|”：|仄平|平仄|仄|（穷里目，0-1-1）；

“|更上|一层|楼|”：|仄仄|仄平|平‖（上层楼，1-0-0）。

其中下划线的位置表示节拍的平仄，不能出错；无下划线框的位置表示平仄可以互换。“|”是节拍分割记号；“‖”表示韵脚。

第一和第二句的三个节拍都“对”；而第三和第四句则只有前两个节拍相“对”，即 0 对 1，1 对 0。

3）一二句和三四句相对

第一句和第二句相同位置节拍的平仄必须相反，即 1 对 0，0 对 1；第三句和第四句也有同样规定。这就是“邻句相对”，简称为“对”。【例 S-1】清楚地显示出这一规律。

4）二三句相粘

第二句和第三句前两个相同位置节拍的平仄必须相同，这就是“邻句相粘”，简称为“粘”。【例 S-1】也清楚地显示出这一规律，即 0 对 0，1 对 1。

S.3.2 七绝

七言绝句简称为“七绝”。和五绝一样，也由四句组成；所不同处，仅仅是每句为 7 个字，全诗总共 28 字。

七绝也要求第二、第四句必须押韵。第一句虽未规定必须压韵，但因 7 个字比较长，所以与五绝相比，首句押韵的七绝较多。《唐诗三百首》收录 51 首七绝，首句不入韵的有 6 首，仅占 11.8%，可以说，多数七绝首句是入韵的。

《唐诗三百首》收录的 51 首七绝全部押平韵，可见与五绝相比，七绝押仄韵

的更少。

七绝的粘对规定和五绝完全相似。每句分 4 个节拍：2-2-2-1；同句内前 3 个节拍平仄必须相间，这就是“句内自对”；一、二句和三、四句的前三个节拍必须“邻句相对”；二、三句必须“邻句相粘”。

解剖【例 S-2】，上述规则均很清晰。

|朝辞|白帝|彩云|间|，|平平|仄仄|仄平|平‖(0-1-0)

|千里|江陵|一日|还|；|平仄|平平|仄仄|平‖(1-0-1)　一二句“对”

|两岸|猿声|啼不|住|，|仄仄|平平|平仄|仄|(1-0-1)　二三句“粘”

|轻舟|已过| 万重|山|。|平平|仄仄|仄平|平‖(0-1-0)　三四句“对”

S.3.3　“一三五不论”与“孤平”

对于二字节拍而言，如前所述，其平仄与第一个字的平仄无关，故历来有“一三五不论，二四六分明”之说。也就是说，对于一句七言诗，第 1、3、5 个字的平仄没有严格规定，但第 2、4、6 个字决定节拍的平仄，不能随意。同理，对于一句五言诗，当有“一三不论，二四分明”。

不少专家认为，七言诗句的第 5 字或五言诗句的第 3 字还是要“论”的。比如【例 S-2】的第 1、4 句的第 5 字换为平声字，最后三个字变成|平平|平|；第 3 句的第 5 字换成仄声字，最后三个字变成|仄仄|仄|，都失去抑扬顿挫之美。

近体诗之大忌是“孤平”，即一句诗中只有一个平声字。【例 S-2】的第 3 句如果“一三五不论”，把第 3、5 字都换成仄声字，就出现孤平，这是绝对要避免的。所以我认为应当用下面的口诀，或许更全面一点：

“一三五不论，二四六分明；避免三同尾，大忌是孤平。”

S.3.4　首句格律决定全诗格律

按照上述规则，合乎格律的五言绝句的首句只有 4 种形式：

0-1-0 型：平平|仄仄|平(平首平尾)；

0-1-1 型：平平|平仄|仄(平首仄尾)；

1-0-0 型：仄仄|仄平|平(仄首平尾)；

1-0-1 型：仄仄|平平|仄(仄首仄尾)。

对于绝句，只要确定第 1 句的平仄，后 3 句的平仄就确定了。

如果首句是 0-1-0 型，按一二句相对以及押韵的要求，第二句一定是 1-0-0；由二三句相粘，第三句一定是 1-0-1；再按三四句相对，第四句一定是 0-1-0。全诗格律如下：

平平｜仄仄｜平｜，仄仄｜仄平｜平‖；仄仄｜平平｜仄｜，平平｜仄仄｜平‖。
（0 0　1 1　0｜1 1　1 0　0‖　1 1　0 0｜1｜　0 0　1 1　0‖）

如果首句是其他三种类型，也可以确定全诗的格律。由于七绝只是在五绝的每句前加一节拍，合乎格律的七言诗句也只有 4 种形式：

0-1-0-0 型：平平｜仄仄｜仄平｜平（平首平尾）；
0-1-1-1 型：平平｜仄仄｜平平｜仄（平首仄尾）；
1-0-1-0 型：仄仄｜平平｜仄仄｜平（仄首平尾）；
1-0-1-1 型：仄仄｜平平｜平仄｜仄（仄首仄尾）。

所以首句也只有上述四种形式，一旦确定，全诗的格律也确定了。例如取首句为 1-0-1-0 型，则二三四句的格律一定是：0-1-0-0，0-1-0-1，1-0-1-0，即

仄仄｜平平｜仄仄｜平‖，平平｜仄仄｜仄平｜平‖；
平平｜仄仄｜平平｜仄｜，仄仄｜平平｜仄仄｜平‖。

实际上，不仅首句平仄决定全诗格律，任何一句的平仄都可推出全诗的格律。比如确定五绝的第四句为 1-0-0 型，那么第三句当为 0-1-1，第二句应是 0-1-0，首句一定是 1-0-1 或 1-0-0。即

仄仄｜仄平｜平‖（或仄仄｜仄平｜仄｜），平平｜仄仄｜平‖；
平平｜平仄｜仄｜，仄仄｜仄平｜平‖。

杜甫说过："为人性僻耽佳句，语不惊人死不休。"（《江上值水如海势聊短述》）诗人有时会来灵感，得到一佳句，至于放在全诗的哪一句，则要推敲。一旦确定在第几句，则其他各句的格律都按"粘对"规则确定。

S.4 律 诗

律诗分为“五律”和“七律”，相当于把两首五绝或七绝合在一起。或者说，绝句是把一首律诗一截为二，所以有人又把绝句称为“截句”。

S.4.1 押韵

无论五律还是七律，第 2、4、6、8 句都必须押韵。和绝句相似，五律的首句以不入韵为正例，七律则以首句入韵为正例。《唐诗三百首》收录了 80 首五律和 53 首七律，其中 65 首五律首句不入韵，占 81.25%；40 首七律首句入韵，占 70.5%。

王勃 649-676

还要指出，律诗必须一韵到底。王勃(649-676)的名篇《滕王阁序》以一首七言诗收尾：

> 滕王高阁临江渚，佩玉鸣鸾罢歌舞。
> 画栋朝飞南浦云，珠帘暮卷西山雨。
> 闲云潭影日悠悠，物换星移几度秋。
> 阁中帝子今何在？槛外长江空自流。

这首诗形式上很近于七律，却不是七律。因为其 1、2、4 句的韵脚是“u”，且是仄韵；而 5、6、8 句的韵脚转为平韵的“ou”。一首诗押不同的韵，称为“转韵”。唐近体诗之前的各种诗歌都容许转韵，但律诗却不容许。

如果说，五绝还有不少仄韵，那么仄韵的五律却几乎没有。七绝已经极少仄韵，仄韵七律就更罕见了。《唐诗三百首》收录的 80 首五律和 53 首七律就全部是平韵。

S.4.2 粘对

五律和七律的句内自对和邻句粘对的规则完全与五绝和七绝相同。要注意的是：第 4 句和第 5 句必须相粘。【例 S-6】是一首七律。

第 4 句："|不**尽**|长**江**|滚**滚**|来|"的节拍平仄为**仄**—**平**—**仄**—平(1-0-1-0)；

第 5 句："|万**里**|悲**秋**|常**作**|客|"的节拍平仄为**仄**—**平**—**仄**—仄(1-0-1-1)，两句的前三节拍的平仄完全相同。

和绝句一样，律诗首句的平仄确定后，全诗的平仄也基本确定。

S.4.3 对仗

律诗与绝句的最大区别是"对仗"。

对仗是中国古典文学中常见的修辞形式，在唐近体诗形成之前早已出现。例如：

"参差荇菜，左右采之；
窈窕淑女，琴瑟友之。" （《诗经·关雎》）

"朝饮木兰之坠露兮，
夕饮秋菊之落英。" （屈原(前 340-278)《离骚》）

"惭幽闺之琴瑟，晦高台之流黄。
春宫閟此青苔色，秋帐含兹明月光。
夏簟清兮昼不暮，冬釭凝兮夜何长！
织锦曲兮泣已尽，回文诗兮影独伤" （江淹(444-505)《别赋》）

"老骥伏枥，志在千里；烈士暮年，壮心未已。
盈缩之期，不但在天；养怡之福，可得永年。"
（曹操(155-220)《步出夏门行》）

"落霞与孤鹜齐飞，
秋水共长天一色。" （王勃《滕王阁序》）

"敬业皇唐旧臣，公侯冢子，奉先君之成业，荷本朝之厚恩。宋微子之兴悲，良有以也；袁君山之流涕，岂徒然哉！是用气愤风云，志安社稷。因天下之失望，顺宇内之推心，爰举义旗，以清妖孽。南连百越，北尽山河。铁骑成群，玉轴相接。海陵红粟，仓储之积靡穷；江浦黄旗，匡复之功何远！班声动而北风起，剑气冲而南斗平。喑呜则山岳崩颓，叱咤则风云变色。以此制敌，何敌不摧；以此图功，何功不克？" （骆宾王(640-684)《讨武曌檄》）

诗经

屈原　340-278 BC

曹操　155-220

江淹　444-505

骆宾王　640-684

对仗也是中国人喜闻乐见的一种文字形式，老百姓家家户户贴的门联也是对仗，至今在许多地方仍保留着过年写春联的习俗，有雅有俗。例如：

“爆竹一声除旧岁，桃符万象更新春”；

“生意兴隆通四海，财源茂盛达三江”；

等等。

至于亭台楼阁的楹联，更是精彩的对仗。其字数少的只有五六字，多的可达数十字。例如滇池之滨的大观楼楹柱上，悬挂着一副长联：

“五百里滇池，奔来眼底，披襟岸帻，喜茫茫空阔无边。看：东骧神骏，西翥灵仪，北走蜿蜒，南翔缟素。高人韵士，何妨选胜登临。趁蟹屿螺洲，梳裹

就风鬟雾鬓；更苹天苇地，点缀些翠羽丹霞。莫辜负：四围香稻，万顷晴沙，九夏芙蓉，三春杨柳。

数千年往事，注入心头，把酒凌虚，叹滚滚英雄谁在？想：汉习楼船，唐标铁柱，宋挥玉斧，元跨革囊。伟烈丰功，费尽移山心力。尽珠帘画栋，卷不及暮雨朝云；便断碣残碑，都付与苍烟落照。只赢得：几杵疏钟，半江渔火，两行秋雁，一枕清霜。”


孙髯　1711-1773

此联为清人孙髯(1711-1773)所作。全联对仗工整，字句洗练。上联写景，气势恢弘；下联写史，意境深远，有“天下第一长联”的美誉，郭沫若(1892-1978)曾撰一联赞誉：“长联犹在壁，巨笔信如椽。”

从上面的例子可以看到，对仗或对联必须由上下两句组成。上下句不仅字数相等，节拍也相同。同句中相邻节拍以及对句中相应节拍也有某些粘对关系。尤其是上下句相同位置的词类必须一致。有的诗赋还接连使用对仗，如前面摘引的《别赋》、《步出夏门行》和《讨武曌檄》等。

对于律诗而言，人们把五律或七律分为4联。第1、2句构成“首联”；第3、4句构成“颔联”；第5、6句构成“颈联”；第7、8句构成“尾联”。每联分为上下句，规定颔联和颈联的上下句必须对仗。

试看一首五律和一首七律。

【例S-8】　**渡荆门送别**

李　白

渡远荆门外，来从楚国游。(首联)
山随平野尽，江入大荒流。(颔联)
月下飞天镜，云生接海楼。(颈联)
仍怜故乡水，万里送行舟。(尾联)

我们来分析这首五律的颔联：

上句的“山”和下句的“江”都是名词，而且都属于“地理”类；上句的“随”和下

句的“入”都是动词；上句的“平野”和下句的“大荒”都是形容词加名词组成的二字词；上句的“尽”和下句的“流”都是动词。

再分析其颈联：

“月”对“云”，名词对名词，且都属于“天文”门类；“下”对“生”，动词对动词；“飞”对“接”，动词对动词；“天”对“海”，名词（“天文”）对名词（“地理”）；“镜”对“楼”，名词（“器物”）对名词（“宫室”）。

七律则回看【例 S-6】。杜甫的这首律诗不仅颔联与颈联的上下句对仗，连首联和尾联也符合对仗要求。

首联：风—渚（名），急—清（动），天—沙（名），高—白（形），猿—鸟（名），啸—飞（动），哀—回（副）；

颔联：无边—不尽（形），落木—长江（形＋名），萧萧—滚滚（副，皆重叠词），下—来（动）；

颈联：万—百（数），里—年（名），悲秋—多病（形＋名），常—独（副），作—登（动），客—台（名）；

尾联：艰难—潦倒（副—副，皆叠韵词），苦—新（副—副），恨—停（动），繁—浊（形），霜鬓—酒杯（名）。

杜甫在二字词的对仗上非常讲究。不仅词类对仗严格，他还常选择双声词、叠韵词及重叠词进行对仗。【例 S-6】中“落木”对“长江”和“艰难”对“潦倒”，都是叠韵词相对；“萧萧”对“滚滚”，则是重叠词相对。又如他的“咏怀古迹之一”的首联也对仗：“支离东北风尘际，飘泊西南天地间”。其中“支离”对“飘泊”，是叠韵对双声；【例 S-7】的颔联是：“怅望千秋一洒泪，萧条异代不同时”。“怅望”对“萧条”，是叠韵对叠韵；《江畔独步寻花》：“留连戏蝶时时舞，自在娇莺恰恰啼”一联中，“留连”对“自在”，是双声对双声，而“时时”对“恰恰”，又是重叠词相对了。杜甫常常吟出令人一唱三叹的绝佳联句。“语不惊人死不休”，正是杜甫的态度。

王力　1900-1986

从词类来说，名词可以分很多类。对仗的范畴越小，就越工整。王力先生（1900-1986）在其名著《汉语诗律学》[S-4]中，依传统说法，对名词分成若干类，每个类再分为二至四个门。如第一类分“天文门”和“时令门”，第二

类分“地理门”和“宫室门”等(参阅附录2)。如果一联中相应的对仗词属于同一门类,则称为“工对”;如果属于邻门,则称为“邻对”;如果仅仅是名词对名词、动词对动词,则称为“宽对”。

使用对仗必须注意:一是避免出句与对句中出现相同的字;二是上下句的意思不能雷同(诗家称为“合掌”)。宋人蔡启在《蔡宽夫诗话》[S-5]中举了一个例子,“‘蝉噪林愈静,鸟鸣山更幽’之类,非不工矣,终不免此病”。也就是说,这一对仗的平仄和词类都很工整,但两对动词“噪—鸣”、“静—幽”和副词“愈—更”都是一个意思,这就是合掌之病。最佳的对仗是流畅自然,毫不雕琢,读者甚至感觉不到是对仗。

【例 S-9】 **闻官军收河南河北**

杜 甫

剑外忽传收蓟北,初闻涕泪满衣裳。

却看妻子愁何在?漫卷诗书喜欲狂。

白日放歌须纵酒,青春作伴好还乡。

即从巴峡穿巫峡,便下襄阳向洛阳。

这首诗的颔联、颈联和尾联都是对仗。作者一气呵成,淋漓尽致地抒发狂喜的爱国之情。《唐诗三百首》的评语是:“一气旋折,八句如一句,而开合动荡,元气浑然,自是神来之作。”[S-1]确实达到了对仗的最佳境界。

杜甫有时也有过于求工的地方。例如他有一联“香稻啄余鹦鹉粒,碧梧栖老凤凰枝”(《秋兴八首》),非常费解,引起千年之争。蔡启感慨地说:“诗语大忌用工太过,盖炼句胜则意必不足。故知造语用事虽同出一人之手,而优劣自异,信乎诗之难也。”[S-6]宋朝有个李廷彦曾写一首百韵长诗,请上司指教。其中有一联:“舍弟江南殁,家兄塞北亡。”上司读到此,赶紧停下来表示慰问。不料他竟回答说:“弟弟是死了,其实哥哥还健在,为了对得工整些,说得沉痛些才这样写的。”[S-7]真称得上因工害意的极致,成为千古笑柄。

S.4.4 排律

排律就是十句以上的律诗。

【例 S-10】 **湘灵鼓瑟**

钱起(751 前后)

善鼓云和瑟,常闻帝子灵。冯夸空自舞,楚客不堪听。

苦调凄金石，清音入杳冥。苍梧来怨慕，白芷动芳馨。

流水传湘浦，悲风过洞庭。曲终人不见，江上数峰青。

【例 S-11】　**和乐天重题别东楼**

元稹(779-831)

山容水态使君知，楼上从容万状移。

日映文章霞细丽，风驱鳞甲浪参差。

鼓催潮户凌晨击，笛赛婆官彻夜吹。

唤客潜挥远红袖，卖炉高挂小青旗。

剩铺床席春眠处，高卷帘帷月上时。

光景无因将得去，为郎抄在和郎诗。

这两首排律均为12句。除首联和尾联外，中间4联均对仗。排律的首联多押韵，许多排律在标题上指明多少韵，就是表示有多少句押韵。排律的长短不一，有的排律可以超过百句。如杜甫的“夔府书怀四十韵”有80句，40句押韵；刘禹锡的“武陵书怀五十韵”有100句，50句押韵；元稹的“春六十韵”有120句，60句押韵；白居易的“代书诗一百韵寄微之”更长达200句之多，其中100句押韵。

应当指出，排律远不如绝句和律诗流传得广。可能由于过长，难以记忆，至多有个别诗句为人称道，如【例 S-10】中的“曲终人不见，江上数峰青”。另外，排律的作者，要接连写出许多对联，难免人工雕琢过重，反而影响了诗人的灵气。

元稹　779-831

刘长卿　709-780

S.4.5 六言律诗

律诗以五言和七言为正宗。但诗人偶尔也作六言律诗。

【例 S-12】 **苕溪酬梁耿别后见寄**

刘长卿(709-780)

清川永路何极,落日孤舟解携。
鸟向平芜远近,人随流水东西。
白云千里万里,明月前溪后溪。
惆怅长沙谪去,江潭芳草萋萋。

闻一多 1899-1946

六言律诗也由八句四联构成,粘对和对仗的要求与五律七律相同。不同处为每句是六个字,分三个节拍:|**|**|**|,韵律为“平—仄—平”(如【例S-12】的1、4、5、8 句)或“仄—平—仄”(如【例 S-12】的 2、3、6、7 句)。

总之,了解以至掌握唐诗格律,不仅有利于更好地欣赏唐诗,也对诗歌写作大有益处。即使写白话新诗,也有助于词句精练和朗朗上口。20 世纪中国新诗大家郭沫若、闻一多(1899-1946)、徐志摩(1896-1931)、臧克家(1905-2004)、郭小川(1919-1976)等都有扎实的古文和诗词格律的基础。

徐志摩 1896-1931

臧克家 1905-2004

郭小川 1919-1976

当然,事物都有两重性,刻意在格律和对仗上求工,固然可显功力,但如果过于注重形式,也可能因辞害意。诗的优劣,第一位还是思想意境。千百年来,遵从格律的绝句律诗不计其数,真正流传的仅占很小的比例;也有许多好的唐诗,

格律却并不严格。试看诗仙李白最脍炙人口的一首五绝：

【例 S-13】　静夜思

李　白

床前明月光，疑是地上霜；

举头望明月，低头思故乡。

第二句的前两个节拍都是仄(是—上)，既句内失对，又与第一句邻句失对；第三句则前两个节拍都是平(头—明)，句内失对，且与第二句失粘，与第四句失对。但是这首诗充分表达了游子对故乡的思念，成为千古绝唱。

【例 S-14】　春　晓

孟浩然(689-740)

春眠不觉晓，处处闻啼鸟。

夜来风雨声，花落知多少。

这首诗和【例 S-4】一样，虽是仄韵，且第 2 句和第 3 句失粘，但是意境绝佳，且如同白话，至今牙牙学语的幼儿都会背诵。《唐诗三百首》中仅收入 29 首“五言绝句”，孟浩然、李白和柳宗元的这三首都在内。

本文介绍唐诗格律最基本的知识，仅供入门。进一步研究，还有不少方面会涉及。例如，有时不得不在某句平仄不依常格，该平而仄，或该仄而平，称为“拗句”，那就要在本句或下句的相应位置易仄为平，或易平为仄，就是“拗救”。【例 S-13】的第二句“疑是地上霜”的“上”该平而仄，“拗”了；于是第三句“举头望明月”的第四字“明”字是易仄为平，这就是“救”[S-4][S-6]。

关于现代人写诗是否应遵循格律，历来见仁见智。特别是 20 世纪 20 年代白话诗兴起，大量外国诗歌翻译介绍给国人，更引起激烈争论。其实，应当“不薄新诗爱旧诗”，各种体裁的诗歌都有用武之地。这方面毛泽东的看法很有见地。

毛泽东　1893-1976

1957 年《诗刊》创刊。毛泽东给《诗刊》的首任主编臧克家写了《关于诗的一封信》。其中除了表示同意《诗刊》发表他多年创作的旧体诗词 18 首外，还写了如下一段话：

“诗刊出版，很好，祝它成长发展。诗当然应以新诗为主体，旧诗可以写一些，但是不宜在青年中提倡，因这种体裁束缚思想，又不易学，这些话仅供你们参考。”[S-8]

毛泽东对臧克家还说过另一段话：

“旧体诗词源远流长，不仅像我这样的老年人喜欢，而且……中年人也喜欢。我冒叫一声，旧体诗词要发展要改革，一万年也打不倒。因为这种东西，最能反映中华民族和中国人民的特性和风尚，可以兴观群怨嘛，怨而不伤，温柔敦厚嘛……”[S-9]

我认为现代人写诗，首先当然应有感而发，而不是无病呻吟。所谓“诗言志”是也。体裁则可以自由选择。但是不论何种体裁，只要是诗，总应当语言精练、大体押韵。如果冠上“绝句”、“律诗”，那就必须遵循唐诗格律的基本准则，否则将不伦不类，贻笑大方。

20 世纪 80 年代初，我到美国访问进修应用数学。当时年轻气盛，曾到哈佛大学“燕京社”和普林斯顿大学东亚文学系讲过唐诗格律的形式体系，效果居然不错[S-10]；2008、2009、2010 年连续三年在上海交大新生研讨课中选讲，学生表现出浓厚的兴趣，不少学生的绝句律诗习作达到相当的水准[S-11]，可见当代青年，还是喜欢格律，也能够初步掌握格律的。

S.5　思考与练习

1. 诗词格律对诗人或学写诗者毕竟是一种束缚，为什么一千多年以来，绝句律诗始终绵续不绝？

2. 若一首七绝的第三句格律为 011 型(即平平仄仄平平仄)，试写出全诗的格律。

3. 给出上联为“神州燃圣火奥运精神融入五洲四海”，请对下联。

4. 尝试写一首绝句，写一首律诗。

参考文献或网站

[S-1]　(清)蘅塘退士:《唐诗三百首》,中华书局(1959).

[S-2]　中华书局上海编辑所:《诗韵新编》,中华书局(1965).

[S-3]　(清)《钦定四库全书荟要御定佩文韵府》,吉林出版集团(2005).

[S-4]　王力:《汉语诗律学》,上海教育出版社(1964).

[S-5]　中华书局:《宋诗话辑佚》.

[S-6]　涂宗涛:《诗词曲格律纲要》,天津人民出版社(2000).

[S-7]　巫唐:尽信诗不如无诗,《学习时报》(2005.3.21).

[S-8]　毛泽东:给《诗刊》编辑的信,《诗刊》第一期(1957).

[S-9]　王国钦:作为旧体诗词大家的毛泽东,《光明网》(2003.12.24).

[S-10]　向隆万:在哈佛和普林斯顿讲唐诗格律,《上海市欧美同学会会刊》No.5 (1988).

[S-11]　向隆万:雏凤新啼清丽声,《上海交大报》No.1260(2009).

【附录 S.1】

《佩文诗韵》分部表

上平声

[一东] 东、同、铜、桐、筒、童、瞳、中、衷、忠等 174 字；

[二冬] 冬、农、宗、钟、龙、松、冲、容、蓉、庸等 120 字；

[三江] 江、杠、缸、扛、窗、邦、双、腔、桩、憧等 51 字；

[四支] 支、枝、移、为、垂、吹、奇、宜、皮、离等 464 字；

[五微] 微、薇、晖、辉、挥、围、飞、非、机、希等 172 字；

[六鱼] 鱼、渔、初、书、舒、居、车、渠、余、锄等 123 字；

[七虞] 虞、愚、娱、隅、趋、无、巫、须、珠、姑等 305 字；

[八齐] 齐、黎、妻、低、题、提、鸡、西、梯、迷等 133 字；

[九佳] 佳、街、鞋、柴、钗、差、阶、偕、排、怀等 55 字；

[十灰] 灰、魁、回、槐、枚、梅、煤、雷、哀、来等 111 字；

[十一真] 真、因、辛、新、晨、臣、人、神、亲、申等 171 字；

[十二文] 文、闻、云、分、焚、坟、员、欣、芹、殷等 97 字；

[十三元] 元、原、源、烦、繁、樊、翻、喧、孙、门等 161 字；

[十四寒] 寒、韩、丹、单、安、难、餐、滩、坛、残等 139 字；

[十五删] 删、潸、关、弯、还、班、颁、般、蛮、颜等 64 字。

下平声

[一先] 先、前、千、天、坚、肩、贤、弦、烟、田等 235 字；

[二萧] 萧、箫、挑、刁、雕、迢、条、辽、消、朝等 183 字；

[三肴] 肴、巢、交、茅、嘲、钞、包、胶、敲、抛等 107 字；

［四豪］豪、毫、操、刀、萄、桃、糟、涛、号、陶等 110 字；

［五歌］歌、多、罗、河、戈、和、波、科、娥、荷等 115 字；

［六麻］麻、花、霞、家、华、沙、车、牙、蛇、瓜等 167 字；

［七阳］阳、杨、扬、香、乡、光、昌、堂、章、张等 270 字；

［八庚］庚、更、羹、盲、横、彭、棚、亨、英、平等 140 字；

［九青］青、经、形、亭、庭、停、丁、馨、星、听等 90 字；

［十蒸］蒸、承、陵、冰、鹰、应、绳、升、兴、登等 114 字；

［十一尤］尤、邮、优、流、留、刘、由、游、牛、修等 247 字；

［十二侵］侵、寻、林、临、箴、沈、深、心、琴、今等 70 字；

［十三覃］覃、潭、谭、参、南、男、谙、庵、含、贪等 96 字；

［十四盐］盐、廉、嫌、严、占、髯、谦、瞻、兼、尖等 86 字；

［十五咸］咸、函、谗、衔、岩、帆、衫、监、凡、嵌等 41 字。

上声

［一董］董、动、孔、总、笼、汞、桶、空、洞、懂等 36 字；

［二肿］肿、种、宠、垄、冗、茸、重、奉、勇、恐等 46 字；

［三讲］讲、港、棒、蚌、项等 11 字；

［四纸］纸、只、是、彼、委、傀、此、美、比、水等 212 字；

［五尾］尾、鬼、苇、卉、几、伟、斐、岂、匪、蜚等 37 字；

［六语］语、吕、旅、与、煮、汝、暑、鼠、处、女等 93 字；

［七麌］麌、雨、羽、宇、舞、父、府、鼓、虎、古等 145 字；

［八荠］荠、礼、体、米、启、陛、洗、底、弟、涕等 41 字；

［九蟹］蟹、解、骇、买、洒、楷、锴、獬、罢、矮等 21 字；

［十贿］贿、悔、改、采、彩、海、在、宰、载、待等 60 字；

［十一轸］轸、敏、允、引、尹、尽、忍、准、笋、闵等 57 字；

［十二吻］吻、粉、蕴、愤、隐、谨、近、忿、槿、刎等 22 字；

［十三阮］阮、远、本、晚、苑、反、饭、稳、很、垦等 62 字；

［十四旱］旱、管、满、短、馆、缓、款、懒、卵、伴等 48 字；

［十五潸］潸、眼、简、版、产、限、撰、串、柬、莞等 33 字；

［十六铣］铣、善、遣、浅、典、转、犬、选、冕、免等 132 字；

[十七筱] 筱、小、表、鸟、了、晓、少、扰、秒、杳等 80 字；

[十八巧] 巧、饱、卯、狡、爪、鲍、搅、绞、拗、佼等 23 字；

[十九皓] 皓、宝、藻、早、枣、老、好、道、脑、岛等 76 字；

[二十哿] 哿、火、我、娜、荷、可、左、果、朵、坐等 70 字；

[二十一马] 马、下、者、野、雅、瓦、寡、写、夏、也等 49 字；

[二十二养] 养、痒、鞅、像、象、仰、朗、奖、敞、枉等 108 字；

[二十三梗] 梗、影、景、井、岭、领、警、请、饼、省等 64 字；

[二十四迥] 迥、炯、茗、挺、艇、醒、溟、冼、等、肯等 40 字；

[二十五有] 有、酒、首、手、口、母、后、友、斗、久等 122 字；

[二十六寝] 寝、饮、锦、品、枕、审、甚、噤、朕、荏等 35 字；

[二十七感] 感、览、胆、坎、惨、敢、颔、撼、菡、喊等 60 字；

[二十八琰] 琰、焰、敛、俭、险、脸、点、贬、冉、陕等 42 字；

[二十九豏] 豏、槛、范、减、舰、犯、湛、斩、掺、砍等 30 字。

去声

[一送] 送、梦、凤、洞、众、瓮、弄、贡、冻、痛等 43 字；

[二宋] 宋、重、用、颂、诵、统、纵、讼、种、共等 23 字；

[三绛] 绛、降、巷、撞、虹、憧、幢、讧、淙等 19 字；

[四寘] 寘、置、事、意、志、治、思、泪、义、利等 266 字；

[五未] 未、味、气、贵、费、沸、尉、畏、汇、既等 52 字；

[六御] 御、处、去、虑、誉、署、据、曙、助、豫等 69 字；

[七遇] 遇、路、露、树、度、赋、布、步、固、素等 142 字；

[八霁] 霁、制、计、势、世、丽、岁、卫、济、第等 202 字；

[九泰] 泰、会、带、外、盖、大、赖、蔡、害、最等 61 字；

[十卦] 卦、挂、懈、隘、卖、画、派、债、怪、坏等 68 字；

[十一队] 队、内、塞、爱、辈、佩、代、退、载、碎等 112 字；

[十二震] 震、信、印、进、润、阵、镇、刃、顺、慎等 93 字；

[十三问] 问、闻、运、晕、韵、训、粪、奋、郡、近等 34 字；

[十四愿] 愿、论、怨、恨、万、饭、献、建、寸、顿等 40 字；

[十五翰] 翰、岸、汉、断、乱、叹、散、畔、旦、算等 100 字；

［十六谏］谏、雁、患、涧、宦、晏、慢、办、盼、豢等 48 字；

［十七霰］霰、殿、面、县、变、箭、战、扇、见、院等 133 字；

［十八啸］啸、笑、照、庙、窍、妙、召、邵、要、耀等 39 字；

［十九效］效、教、貌、校、孝、闹、豹、爆、罩、拗等 39 字；

［二十号］号、帽、报、导、盗、操、躁、灶、告、暴等 60 字；

［二十一箇］箇、个、贺、佐、作、轲、驮、大、饿、过等 42 字；

［二十二祃］祃、驾、夜、下、谢、罢、夏、霸、嫁、赦等 58 字；

［二十三漾］漾、上、望、相、将、状、帐、浪、唱、让等 84 字；

［二十四敬］敬、命、正、令、政、性、镜、盛、圣、性等 55 字；

［二十五径］径、定、听、胜、罄、应、乘、胜、赠、邓等 49 字；

［二十六宥］宥、候、就、授、售、寿、秀、绣、奏、兽等 164 字；

［二十七沁］沁、饮、禁、任、荫、浸、赁、渗、妊、甚等 34 字；

［二十八勘］勘、暗、滥、担、撼、缆、瞰、暂、淡、憨等 29 字；

［二十九艳］艳、剑、念、验、店、敛、厌、垫、欠、殓等 71 字；

［三十陷］陷、鉴、监、泛、梵、忏、赚、蘸、谗、欠等 20 字。

入声

［一屋］屋、木、竹、目、福、禄、熟、谷、肉、族等 180 字；

［二沃］沃、俗、玉、足、曲、粟、烛、属、绿、毒等 63 字；

［三觉］觉、角、较、榷、乐、捉、朔、卓、剥、浊等 92 字；

［四质］质、日、笔、出、室、实、疾、术、一、乙等 132 字；

［五物］物、佛、拂、屈、郁、乞、掘、吃、勿、不等 48 字；

［六月］月、骨、越、没、伐、罚、卒、忽、勃、粤等 110 字；

［七曷］曷、达、末、阔、活、钵、脱、夺、割、拔等 89 字；

［八黠］黠、扎、猾、八、察、杀、刹、轧、茁、滑等 64 字；

［九屑］屑、节、雪、绝、列、裂、结、穴、说、血等 161 字；

［十药］药、薄、恶、略、作、乐、罗、阁、鹤、爵等 197 字；

［十一陌］陌、石、客、白、泽、伯、迹、宅、席、策等 168 字；

［十二锡］锡、壁、历、击、绩、笛、敌、激、析、觅等 96 字；

［十三职］职、国、德、食、色、力、翼、墨、极、息等 140 字；

［十四缉］缉、辑、立、集、邑、急、入、泣、湿、习等59字；

［十五合］合、塔、答、纳、榻、阁、杂、腊、匝、衲等51字；

［十六叶］叶、帖、贴、接、猎、妾、蝶、叠、涉、捷等96字；

［十七洽］洽、狭、峡、法、甲、业、匣、压、鸭、乏等59字。

注：本附录摘自参考文献[S-2]，其中繁体字多已改用相对应的简体字。

【附录 S. 2】

名词门类表

第一类

（甲）天文门（天、空、日、月、风、雨、霜、雪、雷、电、虹、云、…）

（乙）时令门（年、岁、月、日、时、刻、春、夏、秋、冬、晨、夕、…）

第二类

（甲）地理门（地、土、山、水、江、河、湖、海、乡、村、道、路、…）

（乙）宫室门（房、宅、楼、台、门、塔、梁、柱、库、殿、寺、亭、…）

第三类

（甲）器物门（舟、船、钟、鼓、床、枕、弓、箭、刀、尺、盆、缸、…）

（乙）衣饰门（衣、裳、襟、裙、帽、带、巾、靴、裘、扇、盔、甲、…）

（丙）饮食门（酒、茶、糕、饼、药、丹、餐、饭、蔬、盐、酱、汤、…）

第四类

（甲）文具门（笔、墨、砚、纸、印、书、剑、琴、弦、箫、笛、简、…）

（乙）文学门（诗、书、赋、章、句、篇、碑、词、辞、图、画、歌、…）

第五类

（甲）草木花果门（树、木、花、草、柳、菊、桃、梅、竹、根、麦、…）

（乙）鸟兽虫鱼门（马、牛、鸡、犬、龙、凤、虎、豹、蚕、蝉、鼠、…）

第六类

（甲）形体门（身、心、肌、肤、骨、肉、首、眼、耳、鼻、手、足、…）

（乙）人事门（功、名、恩、怨、愁、才、情、歌、舞、笑、醉、梦、…）

第七类

（甲）人伦门（兄、弟、父、母、君、臣、夫、妻、儿、女、友、仙、…）

（乙）代名对（吾、我、余、汝、尔、君、他、谁、孰、自、己、人、…）

第八类

（甲）方位对（东、南、西、北、中、外、里、边、前、后、左、上、…）

（乙）数目对（一、二、三、四、十、百、千、万、两、双、独、众、…）

（丙）颜色对（红、黄、白、黑、青、绿、赤、紫、苍、粉、彩、素、…）

（丁）干支对（甲、乙、丙、丁、戊、己、庚、辛、壬、癸、子、丑、…）

第九类

（甲）人名对

（乙）地名对

第十类

（甲）同义连用词（如秦群玉　杜门："世路变**陵谷**，人情验**友朋**。"）

（乙）反义连用词（如司马礼　登河中鹳雀楼："**兴亡**留日月，**今古**共红尘。"）

（丙）连绵字（如韦庄　思归："外地见花终**寂寞**，异乡闻乐更**凄凉**。"）

（丁）重叠字（如刘兼　春日醉眠："**处处**落花春**寂寂**，**时时**中酒病**恹恹**。"）

第十一类

（甲）副词（如：忽、渐、乍、已、将、欲、即、皆、俱等）

（乙）连介词（如：与、和、共、同、并、且、还、而、则、为等）

（丙）助词（如：也、矣、焉、哉、乎、耶、尔、然、止、之等）

【注】本附录摘自参考文献[S-4]。虽然多数是名词，但其中第七类和第八

类已包含代词、副词、数词和形容词;第十类也可以包括非名词;第十一类则都是非名词。应当注意,以上分类并没有明确界限;同一个字也可能出现在不同的门类。

选修材料之二

微积分在经济问题中的若干应用

“绪论”中曾提到经济现象中蕴含着丰富的数量关系，充满着数学问题。在“优化问题”一章中，已经有涉及经济管理的例题。下面介绍微积分在经济问题中的若干应用，仅为沧海一粟，希望给读者一些启示。

J.1 供应与需求

“微观经济学”研究市场机制，其基本原理是从两条曲线，也就是两个函数开始的[J-1]。

J.1.1 供应(supply)曲线(函数)

通常取纵轴为价格 P，横轴为供应量 Q，将每一价格下刺激厂商进行生产的商品数，描点连成的曲线称为“供应曲线”；函数

$$Q = Q_S(P) \tag{J.1}$$

称为“供应函数”(图 J.1)。站在供应方的角度，如果价格上升，必然鼓励厂商生产更多，使得供应上升。所以供应函数应当是单调增的，即导数$\geqslant 0$。

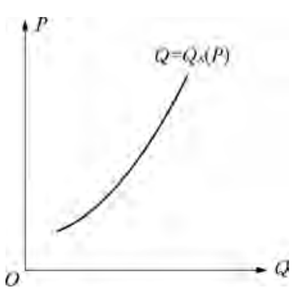


图 J.1 供应曲线

J.1.2 需求(demand)曲线(函数)

在消费者的收入、偏好及其他商品的价格不变

情况下，商品需求量 Q 与价格 P 之间的数量关系描点连成的曲线称为“需求曲线”；函数

$$Q_D = Q_D(P) \tag{J.2}$$

称为“供应函数”（图 J.2）。站在需求方的角度，如果价格上涨，必然限制消费者购买更多的商品的欲望。所以供应函数应当是单调减的，即导数$\leqslant 0$。

从上面两幅图中可见，经济学家在这里把自变量（价格 P）作为纵坐标，与数学家的习惯不相同。由于供应函数和需求函数都是单调函数，它们的反函数都存在。也可以把商品数量作为自变量，把价格作为因变量，供应函数和需求函数的反函数分别是

$$P = S(Q) \tag{J.3}$$

和

$$P = D(Q) \tag{J.4}$$

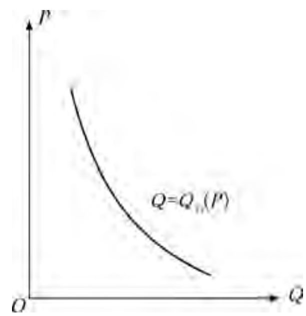


图 J.2　需求曲线

其图形不变，这时自变量（商品数量 Q）就是横坐标了。

J.1.3　均衡价格(equilibrium price)

均衡价格 P^* 是商品的供给曲线与需求曲线相交时的价格。也就是商品的供给量与需求量相等，商品的供给价格与需求价格相等时的价格（图 J.3），即

$$Q_S(P^*) = Q_D(P^*) = Q^* \tag{J.5}$$

【例 J-1】　设供应函数 $Q_S = \sqrt{5000P - 10000}$ 和需求函数 $Q_D = 400 - 20P$，P 的单位是元。求其反函数和均衡价格。

解： 供应函数的反函数是 $P = 2 + \dfrac{Q^2}{5000}$，需求函数的反函数是 $P = 20 - \dfrac{Q}{20}$。均衡价格是 $P^* = 10$ 元，对应的商品数量 $Q^* = 200$。

J.1.4　过剩(surplus)与短缺(shortage)

设 P^* 是均衡价格，当 $P > P^*$（例如图 J.3 中的 P_1），$Q_S(P_1) > Q_D(P_1)$，这时

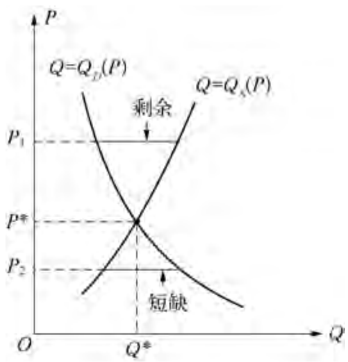


图 J.3　平衡、剩余和短缺

$$Q_S(P) - Q_D(P) > 0 \tag{J.6}$$

我们把 $Q_S(P) - Q_D(P)$ 称为“剩余”，也就是供应太多，超出购买力；

当 $P < P^*$（例如图 J.3 中的 P_2），$Q_S(P_2) > Q_D(P_2)$，这时

$$Q_S(P) - Q_D(P) < 0 \tag{J.7}$$

我们把 $Q_D(P) - Q_S(P)$ 称为“短缺”，也就是供应太少，不能满足消费者之需。

在【例 J-1】中，例如当价格为 $P_1 = 15$ 元时，商品剩余数量为

$$\sqrt{5000 \times 15 - 10000} - (400 - 20 \times 15) = 154.9 \approx 155$$

当价格为 $P_2 = 5$ 时，商品短缺数量为

$$(400 - 20 \times 5) - \sqrt{5000 \times 5 - 10000} = 177.5 \approx 178$$

J.1.5　消费者剩余(consumer's surplus)和生产者剩余(producer's surplus)

假定电影院原定放 4 场电影，可接待 800 名观众，票价为每张 30 元，学生甲准备买票；后来电影院加映两场，票价减为每张 25 元，学生甲实际上只付 25 元，比原计划省下了 5 元。这 5 元就是“**消费者剩余**”。

一般而言，需求曲线 $P = D(Q)$ 表示在价格 Q 的水平上，消费者愿为每一个单元商品 ΔQ 支付的最高价格。假定当前商品数量和价格为 $(\bar{Q}, \bar{P})$。当出售商品数 Q 低于 $\bar{Q}$ 时，消费者仍以价格 $\bar{P}$ 买进，比预期的价格 $P = D(Q)$ 低。那么，消费者感到少付了 $(D(Q) - P)\Delta Q$。累积起来，消费者感到一共少付了

$$CS = \int_0^{\bar{Q}} D(Q)\mathrm{d}Q - \bar{P}\bar{Q} \tag{J.8}$$

按上式定义的 CS 称为“消费者剩余”(图 J.4(a))。

从另一角度，假定电影院原定每张票售 20 元即可保本，现在卖 30 元，那么其差值 10 元就是“**生产者剩余**”。一般而言，供应曲线 $P = S(Q)$ 表示在价格 Q 的水平上，生产者为出售每一单元商品 ΔQ 所定的最低价格。假定当前商品数

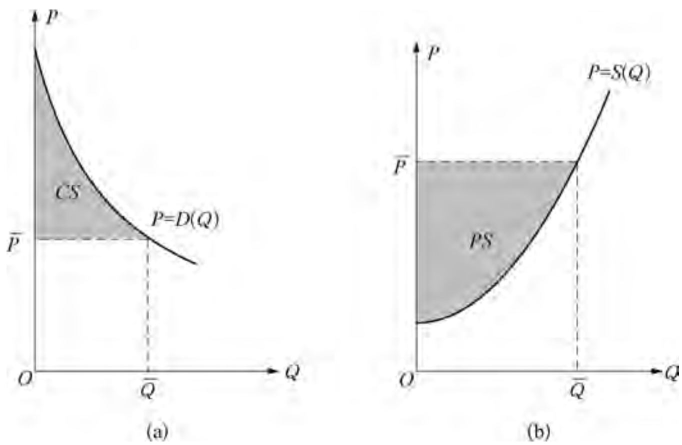


图 J.4

量和价格为 $(\bar{Q}, \bar{P})$。当出售商品数 Q 低于 $\bar{Q}$ 时，生产者仍以价格 $\bar{P}$ 出售，比预期的价格 $P=S(Q)$ 高。那么，生产者感到多赚了 $(\bar{P}-S(Q))\Delta Q$。累积起来，供应方感到一共多赚了

$$PS=\bar{P}\bar{Q}-\int_0^{\bar{Q}} S(Q)\mathrm{d}Q \tag{J.9}$$

按上式定义的 PS 称为"生产者剩余"(图 J.4(b))

回到【例 J-1】，考虑平衡点价格 $P^*=10$ 作为当前价格 $\bar{P}$，则消费者剩余为

$$CS=\int_0^{200}\left(20-\frac{Q}{20}\right)\mathrm{d}Q-10\times 200=\left(20Q-\frac{Q^2}{2\times 20}\right)\Bigg|_0^{200}-2000$$

$$=4000-\frac{40000}{40}-2000=1000\text{ 元}$$

而生产者剩余为

$$PS=10\times 200-\int_0^{200}\left(2+\frac{Q^2}{5000}\right)\mathrm{d}Q=2000-\left(2Q+\frac{Q^3}{3\times 5000}\right)\Bigg|_0^{200}$$

$$=2000-\left(2\times 200+\frac{1}{15000}\times 200^3\right)=2000-400-\frac{8000000}{15000}\approx 1067\text{ 元}$$

【例 J-2】　设需求函数为 $Q=40-P^2$(图 J.5)，当价格从每件 5 万元降到 3 万元时，消费者剩余增加了多少？

解：$Q(5)=40-5^2=15$，$Q(3)=40-3^2=31$

方法 1：将需求函数变形为 $P=\sqrt{40-Q}$，由 8.8 式，利用换元积分法，令 $u=40-Q$，$\mathrm{d}u=-\mathrm{d}Q$，当 $Q=31$，$u=9$；$Q=15$，$u=25$。

$$\begin{aligned}CS(31)-CS(15)&=\left(\int_0^{31}\sqrt{40-Q}\,\mathrm{d}Q-3\times 31\right)-\left(\int_0^{15}\sqrt{40-Q}\,\mathrm{d}Q-5\times 15\right)\\&=\int_{15}^{31}\sqrt{40-Q}\,\mathrm{d}Q-93+75\\&=-\int_{25}^{9}\sqrt{u}\,\mathrm{d}u-18=\int_9^{25}\sqrt{u}\,\mathrm{d}u-18\\&=\left.\frac{u^{\frac{1}{2}+1}}{\frac{1}{2}+1}\right|_9^{25}-18=\frac{2}{3}(5^3-3^3)-18=\frac{2}{3}\times 98-18\\&=47\frac{1}{3}\approx 47.3333\text{ 万元}\end{aligned}$$

方法 2：由定积分的几何性质，把 P 看作自变量，消费者剩余的增加量（图 J.5 的阴影部分）也可以这样计算：

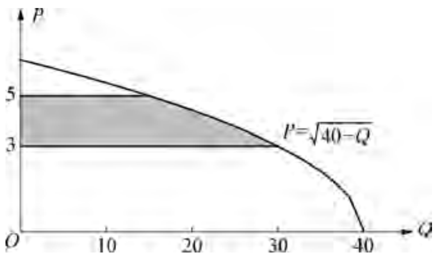


图 J.5 平衡、剩余和短缺

$$\begin{aligned}&\int_3^5(40-P^2)\,\mathrm{d}P\\&=\left.\left(40P-\frac{P^3}{3}\right)\right|_3^5\\&=40\times 5-\frac{5^3}{3}-40\times 3+\frac{3^3}{3}\\&=47\frac{1}{3}\end{aligned}$$

结果完全一致，计算简单得多。

J.1.6 局部市场均衡的稳定性分析

前面介绍过，在均衡价格下，市场处于平衡状态。如果某种干扰破坏了这种平衡状态，市场会不会自动回到平衡状态？下面我们考虑供应函数和需求函数都是线性变化的简单特例。

【例 J-3】 设某商品的供应函数和需求函数分别是

$$Q_S(P)=\alpha+\beta P\ (\beta>0) \tag{J.10}$$

$$Q_D(P)=\mu+\nu P\ (\nu<0) \tag{J.11}$$

假定价格随时间的变化与商品短缺成正比，即成立如下一阶微分方程：

$$\frac{\mathrm{d}P}{\mathrm{d}t}=\lambda(Q_D(P)-Q_S(P)) \tag{J.12}$$

这就是一个价格的动态系统。根据均衡价格的定义 J. 5 式，$Q_S(P^*)=Q_D(P^*)$，由 J. 10 式和 J. 11 式，求得均衡价格

$$P^*=\frac{\mu-\alpha}{\beta-\nu} \tag{J.13}$$

以及均衡销售数量

$$Q^*=\frac{\beta\mu-\alpha\nu}{\beta-\nu} \tag{J.14}$$

现在假定实际价格与均衡价格有偏差 $\bar{P}=P-P^*$，则 $P=\bar{P}+P^*$，代入 J. 12 式，

$$\begin{aligned}\frac{\mathrm{d}\bar{P}}{\mathrm{d}t}=\frac{\mathrm{d}P}{\mathrm{d}t}&=\lambda((\mu+\nu P)-(\alpha+\beta P))\\&=\lambda((\mu+\nu(\bar{P}+P^*)-\alpha-\beta(\bar{P}+P^*))\end{aligned}$$

即

$$\frac{\mathrm{d}\bar{P}}{\mathrm{d}t}=\lambda(\nu-\beta)\bar{P}-\lambda(\alpha-\mu+(\beta-\nu)P^*)$$

由于 J. 13 式，上式右端第 2 项为 0，故得到关于 $\bar{P}$ 的一阶微分方程

$$\frac{\mathrm{d}\bar{P}}{\mathrm{d}t}=\lambda(\nu-\beta)\bar{P} \tag{J.15}$$

这是可分离变量的形式。假定初始价格为 $P(0)=P_0$，则初始偏差为 $\bar{P}(0)=\bar{P}_0=P_0-P^*$，不难解出

$$\bar{P}(t)=\bar{P}_0e^{-\lambda(\beta-\nu)t} \tag{J.16}$$

如果外界条件不变，那么当 $t\to\infty$，$e^{-\lambda(\beta-\nu)t}\to 0$，$\bar{P}(t)\to 0$，即 $P\to P^*$。说明无论开始价格偏离多远，最终价格会趋近于均衡价格，这个系统是稳定的。

进一步讨论可参阅[J-2]；其中需要微分方程定性理论的基本知识[J-3]。由以上讨论可见，基于供应曲线（函数）和需求曲线（函数）可以研究许多经济现象。

J.2 边际分析和弹性分析

J.2.1 成本(cost)、平均成本(average cost)和边际成本(marginal cost)

1) 成本函数

生产 Q 件产品，一定有成本。一部分是“固定成本(fixed cost)”，记为 C_0，与生产数量关系不大，例如场地、固定设备、基本人员工资等，不论产量大小，看作常数；另一部分是“可变成本(variable cost)”，与生产数量密切相关，例如消耗的材质、能源、计件工资等，记为 $C_1(Q)$，于是成本函数可表示为

$$C(Q) = C_0 + C_1(Q) \tag{J.17}$$

【例 J-4】 某企业生产一种产品，固定成本为 900 元，每做 Q 件产品的可变成本为 $400Q - Q^2$ 元，求成本函数，并分别计算 $Q = 10, 20, 30$ 时的成本。

解： 成本函数为 $C(Q) = 900 - Q^2 + 400Q$

$C(10) = 4800$ 元；$C(20) = 8500$ 元；$C(30) = 12000$ 元。

平均成本函数

设生产 Q 件产品的成本函数为 $C(Q)$，定义平均成本函数为

$$\bar{C}(Q) = \frac{C(Q)}{Q} \tag{J.18}$$

即平均生产每一件产品的成本，简记为 $\bar{C}$。

在【例 J-4】中，平均成本函数为

$$\bar{C} = \frac{-Q^2 + 400Q + 900}{Q} = -Q + 400 + \frac{900}{Q}$$

$\bar{C}(10) = \dfrac{4800}{10} = 480$ 元 / 件，$\bar{C}(20) = \dfrac{8500}{20} = 425$ 元 / 件，$\bar{C}(30) = \dfrac{12000}{30} = 400$ 元 / 件。

可见当生产不同数量产品时，平均成本并不相同。

2) 边际成本

在上面例子中，生产 10 件产品时的成本是 4800 元，平均成本是 480 元。如

果再多生产 1 件，即 $\Delta Q=1$，总成本是多少？很容易计算，$C(11)=5179$ 元。这时设总成本增加量

$$\Delta C=C(11)-C(10)=5179-4800=379\text{ 元，或者 }\frac{\Delta C}{\Delta Q}=379\text{ 元/件}$$

这个成本增量称为“**边际成本**”，它既不等于 $\bar{C}(10)=480$ 元 / 件，又不等于 $\bar{C}(11)=\dfrac{5179}{11}=470.8$ 元/件。因为 $\dfrac{\Delta C}{\Delta Q}=\dfrac{C(Q+\Delta Q)-C(Q)}{\Delta Q}$，对比第四章的导数定义，当 ΔQ 与 Q 相比很小时，

$$\frac{\Delta C}{\Delta Q}\approx C'(Q)=(C_0+C_1(Q))'=C_1'(Q) \tag{J.19}$$

由此可见，边际成本与固定成本无关，它等于边际可变成本。在上面的例子中，$C'(Q)=-2Q+400$。$C'(10)=-2\times 10+400=380$ 元 / 件，和前面的结果很接近。这又是一个离散转化为连续的例子。

在实际经济量化分析问题中，经常将产量为 Q 时的边际成本 $C'(Q)$ 和此时已花费的平均成本 $C(Q)/Q$ 做比较，由两者的意义知道，如果边际成本小于平均成本，则可以再增加产量以降低平均成本，上面例子中，$C'(10)=380<\bar{C}(10)=480$，说明可增加产量。反之，如果边际成本大于平均成本，可以考虑削减产量以降低平均成本；当边际成本等于平均成本时可使产品的平均成本最低。

J.2.2　“边际”概念

1) 经济量“边际”的定义

如果一个经济指标 y 是另一个经济指标 x 的函数 $y=f(x)$，那么当自变量有改变量 Δx 时，对应有函数的改变量 Δy。在经济学中，当自变量在 x 处有一个单位改变量时，所对应的函数改变量为该函数所表示的经济指标在 x 处的边际量。

设 x 的改变量为 Δx 时，经济变量 y 的改变量为 $\Delta y=f(x+\Delta x)-f(x)$，则 y 的平均变化率是

$$\frac{\Delta y}{\Delta x}=\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

由边际的概念，在上式中取 $\Delta x=1$ 或 $\Delta x=-1$ 就可得到边际量的表达式。

但边际概念的定义和计算使我们想到能否用函数 $y = f(x)$ 的导数作为 y 的边际量呢？如果按纯粹的数学概念来讲，似乎行不通，因为导数定义要求自变量增量必须趋向于零，而实际问题中自变量 x 的经济意义通常是按计件的产量或销量作为单位的，改变量为小数且趋于零不合乎实际。但我们可以这样考虑，对于现代企业来讲，其产销量的数额和一个单位产品相比是一个很大数目，1 个单位常常是其中微不足道的量，可以认为改变一个单位的这种增量是趋近于零的。正是这个缘故，在经济理论研究中，总是用导数

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

表示经济变量 y 的边际量，即认为 $f'(x)$ 的经济意义是自变量在 x 处有单位改变量时所引起函数 y 的改变数量。

2) **边际收入**(marginal revenue)**和边际利润**(marginal profit)

在经济学中，在价格 P 水平上销售 Q 件商品所得款项称为"收入"，即

$$R = PQ = D(Q)Q \tag{J.20}$$

其中 $D(Q)$ 为需求函数 J.4 式；类似定义边际收入为 $R'(Q)$。

【例 J-5】 某企业生产家用电器，设成本函数为 $C = 7000 + 2Q$，需求函数为 $P = 10 - Q/1000$，求收入函数和边际收入，并分别计算 $Q = 1000$，4000，7000 件时的成本、平均成本、收入和边际收入。

解：

$$R = PQ = \left(10 - \frac{Q}{1000}\right)Q = 10Q - \frac{1}{1000}Q^2$$

$$R'(Q) = 10 - \frac{Q}{500}$$

不难算出，$C(1000) = 9000$ 元，$C(4000) = 15000$ 元，$C(7000) = 21000$ 元；

$\bar{C}(1000) = 9$ 元 / 件，$\bar{C}(4000) = 3.75$ 元 / 件，$\bar{C}(7000) = 3$ 元 / 件；

$R(1000) = 9000$ 元，$R(4000) = 24000$ 元，$R(7000) = 21000$ 元；

$R'(1000) = 8$ 元 / 件，$R'(4000) = 2$ 元 / 件，$R'(7000) = -4$ 元 / 件。

设利润函数为 $L = L(Q)$，它是收入函数与成本函数之差，即

$$L(Q) = R(Q) - C(Q) \tag{J.21}$$

类似可定义平均利润为

$$\overline{L}(Q) = \frac{L(Q)}{Q} \tag{J.22}$$

边际利润为

$$L'(Q) = R'(Q) - C'(Q) \tag{J.23}$$

它近似等于销量为 Q 时再销售一个单位产品所增加(或减少)的利润。

再看【例 J-5】, $L(Q) = 10Q - \frac{Q^2}{1000} - (7000 + 2Q) = -\frac{Q^2}{1000} + 8Q - 7000$

$$\overline{L}(Q) = -\frac{Q}{1000} + 8 - \frac{7000}{Q},\ L'(Q) = -\frac{Q}{500} + 8$$

$$L(1000) = 0,\ L(4000) = 9000\text{ 元},\ L(7000) = 0$$

$$\overline{L}(1000) = \overline{L}(7000) = 0,\ \overline{L}(4000) = 2.25\text{ 元/件}$$

$$L'(1000) = 6,\ L'(4000) = 0,\ L'(7000) = -6$$

由第六章,知 $Q = 4000$ 是唯一驻点, $L(4000) = 9000$ 元是最大利润。

注意,最大利润和最大平均利润不是一回事。本例中,令 $\overline{L}'(Q) = -1/1000 + 7000/Q^2 = 0$,可求出平均利润的最大值点是 $Q = \sqrt{7} \times 10^3 \approx 2646$,平均利润的最大值为每件产品获利 2.7 元;但此时总利润仅为 7168 元<9000 元。而 $Q = 4000$ 件时,平均利润为每件 9000/4000=2.25 元/件<2.7 元/件。

在经济学中还经常用到边际效用、边际产量、边际劳动生产率等概念,它和边际成本、边际收入、边际利润的经济解释方法大同小异,在此不再赘述。

3) 由边际函数求总函数

给定一个总函数,求导数就得到它的边际函数;反过来,给定边际函数,求不定积分可得到总函数。这里的“总函数”,就是第五章中的“原函数”。

【例 J-6】 已知厂商的边际成本为 $10 - 4Q + 3Q^2$,且固定成本 $C_0 = 5000$ 元,求总成本函数。

解: $C(Q) = \int (10 - 4Q + 3Q^2)\mathrm{d}Q = 10Q - 2Q^2 + Q^3 + C$

因 $C(0) = C_0 = 5000$,所以 $C = 5000$,总成本函数为 $C(Q) = 10Q - 2Q^2 + Q^3 + 5000$。

J.2.3 弹性分析

1) 弹性(elasticity)引例

函数的“边际”表达了函数 $f(x)$ 在 x 处的变化率，粗略来说，就是当自变量的值每改变一个单位所引起函数变化的数量。这里有一个很大的缺憾，就是边际的数量依赖于自变量与因变量的量纲。先看一个实例。

【例 J-7】 某果品公司出口苹果，成本为 $C=5+6x-0.1x^2$，这里 x 的单位是吨，C 的单位是万元人民币。容易计算，当 $x=4$ 吨时，边际成本为

$$C'(4)=6-2\times0.1\times4=5.2\text{（万元人民币/吨）}$$

如果 x 的单位换为 kg，成本函数应换为 $C=5+0.006x-0.0000001x^2$，这时边际成本为 $C'(4000)=0.006-2\times10^{-7}\times4\times10^3=0.0052$（万元人民币/kg）；

如果 C 的单位是千美元，x 的单位仍为吨，按 1∶6.4 元的汇率，上述成本函数换为

$$C=(5+6x-0.1x^2)\times6.4/10=3.2+3.84x-0.064x^2$$

则 $C'(4)=3.84-2\times0.064\times4=3.328$（千美元/吨）。

由上例可见，因量纲单位不同，同一个边际成本可以用三个不同数字描述。为了避免混淆，人们考虑用相对变化率之比来研究，即“弹性”。函数的弹性表达了函数 $f(x)$ 在 x 处的相对变化率，粗略来说，就是当自变量的值每改变百分之一所引起函数变化的百分数。

我们看一个简单数字例子。设 $y=x^2$，当 x 由 10 变到 11 时，y 由 100 变到 121。显然，自变量和函数的绝对改变量分别是 $\Delta x=1$，$\Delta y=21$，它们之比 $\Delta y/\Delta x=21/1=21$，这是我们熟知的当 $x=10$ 时的“平均变化率”，也是经济学中 $y=x^2$ 当 $x=10$ 时的“边际函数”值。

我们进一步考察 x 和 y 的相对改变量 $\dfrac{\Delta x}{x}$ 和 $\dfrac{\Delta y}{y}$。它们分别为

$$\frac{\Delta x}{x}=\frac{1}{10}=10\%,\ \frac{\Delta y}{y}=\frac{21}{100}=21\%$$

这表明，当自变量 x 由 10 变到 11 的相对变动为 10%时，函数 y 的相对变动为 21%，这时两个相对改变量的比为

$$E=\frac{\frac{\Delta y}{y}}{\frac{\Delta x}{x}}=\frac{21\%}{10\%}=2.1$$

E 的意义表明，$x=10$ 时，当 x 改变 1%，y 平均改变 2.1%，我们称为从$x=10$ 到 $x=11$ 时函数 $y=x^2$ 的平均相对变化率，也称为平均意义下函数 $y=x^2$ 的弹性。

2）经济量“弹性”的定义

设函数 $f(x)$ 定量描述一个经济变量对另一个经济变量的确定关系，x 的变化区间为 I。当 $x_0\in I$，如果极限

$$\lim_{\Delta x\to 0}\frac{\frac{\Delta y}{f(x_0)}}{\frac{\Delta x}{x_0}}=\lim_{\Delta x\to 0}\frac{\frac{f(x_0+\Delta x)-f(x_0)}{f(x_0)}}{\frac{\Delta x}{x_0}}$$

存在，则称此极限值为函数 $y=f(x)$ 在点 x_0处的点弹性，记为

$$\left.\frac{Ey}{Ex}\right|_{x=x_0}=\lim_{\Delta x\to 0}\frac{x_0}{f(x_0)}\cdot\frac{\Delta y}{\Delta x}=\frac{x_0}{f(x_0)}f'(x_0)\tag{J.24}$$

称 $\frac{Ey}{Ex}=\frac{x}{f(x)}f'(x)=\frac{\frac{\mathrm{d}y}{\mathrm{d}x}}{\frac{y}{x}}$ 为函数 $y=f(x)$ 在区间 I 的“点(point)弹性”函数，简称弹性函数，相当于“边际变化”除以“平均变化”。

而称取极限前段表达式

$$\frac{\frac{\Delta y}{f(x_0)}}{\frac{\Delta x}{x_0}}=\frac{\frac{f(x_0+\Delta x)-f(x_0)}{f(x_0)}}{\frac{\Delta x}{x_0}}\tag{J.25}$$

为函数 $y=f(x)$ 在以 x_0 与 $x_0+\Delta x$ 为端点的区间上的“弧(arc)弹性”函数。

弧弹性表达了函数 $f(x)$ 当自变量 x 从 x_0变到 $x_0+\Delta x$ 时函数的平均相对

变化率，而点弹性正是函数 $f(x)$ 在点 x_0 处的相对变化率。

【例 J-8】 求函数 $y=1-x^2$ 的弹性函数，并求出 $x=2$ 处的点弹性和 x 从 2 到 3 的弧弹性数值。

解：因为 $y'=-2x$，所以

$$\frac{Ey}{Ex}=\frac{x}{f(x)}f'(x)=\frac{x(-2x)}{1-x^2}=-\frac{2x^2}{1-x^2},\ \frac{Ey}{Ex}\bigg|_{x=2}$$

$$=-\frac{2x^2}{1-x^2}\bigg|_{x=2}=\frac{8}{3}\text{（点弹性）}$$

因 $\Delta x=3-2=1$，$\Delta y=(1-9)-(1-4)=-5$，

$$\left.\frac{\dfrac{\mathrm{d}y}{\mathrm{d}x}}{\dfrac{y}{x}}\right|_{x=2}=\frac{\dfrac{1}{1}}{\dfrac{1-4}{2}}=-\frac{2}{3}\text{（弧弹性）}$$

回到【例 J-7】，不论 x 和 C 取什么单位，成本的弹性都是

$$E=\frac{5.2}{27.4}\times 4=\frac{0.052}{27.4}\times 4000=\frac{3.328}{17.536}\times 4=0.759$$

3）需求弹性

需求弹性就是在需求分析中经常用来测定需求对价格反应程度的一个经济指标。

设某商品的需求函数 $Q=Q_D(P)$ 可导，按 J.24 式，称

$$\frac{EQ}{EP}=\frac{P}{Q_D(P)}Q'_D(P)=\frac{P}{Q_D}Q'_D \tag{J.26}$$

为该商品的需求价格弹性，简称为需求弹性，记为 ε_P。需求弹性可以这样解释 ε_P 的经济意义：当商品的价格为 P 时，价格改变 1% 时需求量变化的百分数。

由经济理论知道，一般商品的需求函数为价格的减函数，从而 $Q'(P)<0$，这说明需求价格弹性 ε_P 一般是负的。由此，当商品的价格上涨（或下跌）1% 时，需求量将下跌（或上涨）约 $|\varepsilon_P|$%，因此在经济学中，比较商品需求弹性的大小时，多研究弹性的绝对值 $|\varepsilon_P|$。当 $|\varepsilon_P|=1$ 时，称为单位弹性，此时商品需求量变动的百分比与价格变动的百分比相等；当 $|\varepsilon_P|>1$ 时，称为高弹性，此时商品需求量变动的百分比高于价格变动的百分比，价格的变动对需求量的影响比较

大；当 $|\varepsilon_P|<1$ 时，称为低弹性，此时商品需求量变动的百分比低于价格变动的百分比，价格的变动对需求量影响不大。

在商品经济中，商品经营者关心的是提价 $(\Delta P>0)$ 或降价 $(\Delta P<0)$ 对总收入的影响，利用需求弹性的概念，可以对此进行分析。

设收入函数为 R，则 $R=PQ_D$，此时边际收入为

$$R'(P)=Q_D+PQ'_D=Q_D\left(1+\frac{P}{Q_D}Q'_D\right)=Q_D(1+\varepsilon_P) \tag{J.27}$$

当 $|\Delta P|$ 很小时，有

$$\Delta R\approx \mathrm{d}R=R'(P)\Delta P=Q_D(1+\varepsilon_P)\Delta P=(1-|\varepsilon_P|)Q_D\Delta P \tag{J.28}$$

由此可知，当 $|\varepsilon_P|>1$（高弹性）时，商品降价时($\Delta P<0$)，$\Delta R>0$，即降价可使收入增加，商品提价时($\Delta P>0$)，$\Delta R>0$，即提价将使总收入减少。

当 $|\varepsilon_P|<1$(低弹性）时，降价使总收入减少，提价使总收入增加。

当 $|\varepsilon_P|=1$(单位弹性）时，$\Delta R=0$，提价或降价对总收入无影响。

上述分析使我们看到，根据商品需求弹性的不同，应制定不同的价格政策，以使收入快速增长。

【例 J-9】 设某种产品的需求量 Q 与价格 P 的关系为

$$Q_D(P)=1600\left(\frac{1}{4}\right)^P$$

(1) 求需求弹性；

(2) 当产品的价格 $p=10$ 时再增加 1%，求该产品需求量变化情况。

解：(1) 由需求弹性公式

$$\varepsilon_P=\frac{P}{Q_D}Q'_D=\frac{P}{1600\left(\frac{1}{4}\right)^P}\cdot\left(1600\left(\frac{1}{4}\right)^P\right)'=P\ln\frac{1}{4}\approx-1.39P$$

需求弹性为 $-1.39P$，说明产品价格 P 增加 1%时，需求量 Q 将减少 1.39%。

(2) 当产品价格 $P=10$ 时，有

$$\varepsilon_P=-1.39\times 10=-13.9$$

这表示价格 $P=10$ 时，价格增加 1%，产品需求量将减少 13.9%；如果价格降低 1%，产品的需求量将增加 13.9%。这也表明此商品的需求弹性是高弹性

的，适当降价会使销量大增。

【例 J-10】 已知某企业的产品需求弹性为 2.1，如果该企业准备明年降价 10%，问这种商品的销量预期会增加多少？总收益预期会增加多少？

解： 题中价格的改变量是相对量，所以所求的销量和总收益的改变也采用相对改变量。由需求函数弹性定义知，当 ΔP 较小时

$$\varepsilon_P = \frac{P}{Q}\cdot\frac{\mathrm{d}Q}{\mathrm{d}P} \approx \frac{P}{Q}\cdot\frac{\Delta Q}{\Delta P}$$

即

$$\frac{\Delta Q}{Q} \approx \varepsilon_P\,\frac{\Delta P}{P}$$

故当 $|\varepsilon_P| = 2.1$，$\frac{\Delta P}{P} = -0.1$ 时，有

$$\frac{\Delta Q}{Q} \approx -2.1\times(-0.1) = 21\%$$

因为 $R=PQ$，由 J.28 式有

$$\frac{\Delta P}{R} \approx (1-|\varepsilon_P|)\frac{Q}{P\cdot Q}\Delta P = (1-|\varepsilon_P|)\frac{\Delta P}{P}$$

当 $\varepsilon_P = 2.1$ 时，有

$$\frac{\Delta P}{R} \approx (1-2.1)\times(-0.1) = 11\%$$

可见，明年企业若降价 10%，企业销量将增加 21%，收入将增加 11%。

J.3 库 存 策 略

许多商家都面临库存(inventory)问题。库存商品太多，则增加仓储成本，且可能积压商品；库存商品太少，又可能供不应求，丧失商机。如何选择最佳的库存策略？我们从一个实例出发。

J.3.1 一个实例

【例 J-11】 某家电商店根据历史资料统计，发现某品牌家电以每台 2000

元价格出售时，平均日销 5 台，而进货价每台仅 1200 元，因此在下月计划进货 150 台。已知每次订货进货费为 1500 元，每天每台库存费用为进货价的 1%，即 20 元。现在的问题是分几次订货进货，成本最低，从而实现最大利润？

解：方案 I. 一次进货 150 台。这时订货进货成本为 1500 元；第 1 天库存费为 $20 \times 150 = 3000$ 元，第 2 天库存费为 $20 \times (150-5) = 2900$ 元，第 3 天库存费为 $20 \times (145-5) = 2800$ 元，以此类推。所以一个月的总库存费为(图 J. 6(a))：

$$20 \times (150 + 145 + 140 + \cdots + 5) = 20 \times 5 \times (30 + 29 + \cdots + 1) = 46500 \text{ 元}$$

相当于库存单价每天每台 20 元与阶梯形面积的乘积。加上采购费为 $1200 \times 150 = 180000$ 元，一个月的总成本为 $C = 1500 + 46500 + 1200 \times 150 = 228000$ 元，而收入 $R = 2000 \times 150 = 300000$ 元，故利润 $L = R - P = 300000 - 228000 = 78000$ 元。

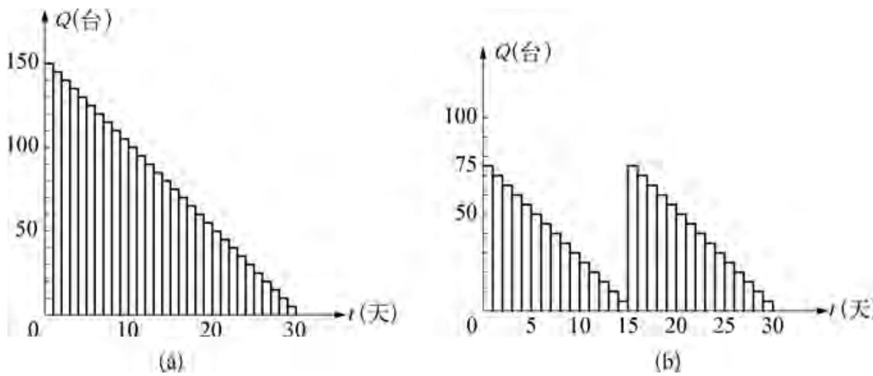


图 J.6

方案 II. 月初月中分两次进货。这时订货进货成本为 $1500 \times 2 = 3000$ 元；第 1 天库存费为 $20 \times 75 = 1500$ 元，第 2 天库存费为 $20 \times (75-5) = 1400$ 元，以此类推，前半个月的库存费为 $20 \times (75 + 70 + \cdots + 5) = 20 \times 5 \times (15 + 14 + \cdots + 1) = 12000$ 元，后半个月也是 12000 元(图 J. 6(b))，总库存费相当于单价 20 元乘上两个阶梯形的面积和，即 24000 元。采购费不变，仍为 180000 元，一个月的总成本为 $C = 3000 + 24000 + 180000 = 207000$ 元。收入同方案 I，故利润 $L = R - P = 300000 - 207000 = 93000$ 元。利润超过方案 I。

是不是为了减少库存费而尽可能增加订货进货的次数呢？极端情况是“零库存”，每天订货，直接进货到门店，这时有如下方案：

方案 III. 订货费为 $1500 \times 30 = 45000$ 元，库存费为 0。这时利润 $L = R - P = 300000 - 45000 - 180000 = 75000$ 元，低于上述两个方案。

那么究竟分几次订货最佳？设订货次数为 n，因为采购费 182000 元和销售收入 300000 元是不变的，只需考虑订货费和总库存费，二者之和最小为优。又因为每天销售 5 台，所以每次订货量必须是 5 的倍数，只有 8 种可能。前 3 个方案已经完成了 3 种可能性。所有结果见表 J.1(各种可能的库存费请读者自行验证)：

表 J.1

可能性 k	订货次数 n	每次订货数 Q	订货费 C_{0n}	库存费 C_{1n}	$C_{0n}+C_{1n}$
1	1	150	1500	46500	48000
2	2	75	3000	24000	27000
3	3	50	4500	16500	21000
4	5	30	7500	10500	18000
5	6	25	9000	9000	18000
6	10	15	15000	6000	21000
7	15	10	22500	4500	27000
8	30	5	75000	0	75000

可见分 5 次订货或 6 次订货成本最低，从而利润最大。

以上用了“穷竭法”，把 8 种可能都计算了。如果用第六章介绍的“黄金分割-分数法”，第一步，计算 $k = 5$(即 $n = 6$) 和 $k = 3$(即 $n = 3$)，由于 $k = 5$ 的效果好，舍弃 $k = 1, 2$(即 $n = 1, 2$)；第二步，计算 $k = 6$(即 $n = 10$)，还是 $k = 5$ 的效果好，舍弃 $k = 7, 8$(即 $n = 15, 30$)；第三步，计算 $k = 4$(即 $n = 5$)，效果和 $k = 5$ 相同，只须计算 4 次。

J.3.2 最优库存策略

下面用微积分方法求最优库存策略。

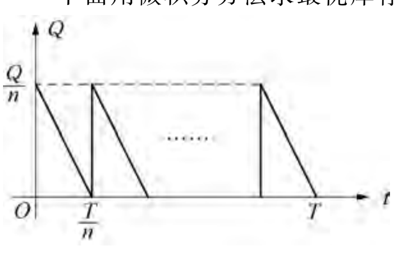


图 J.7

有几个假定：① 订货进货不会滞后；② 均匀销售(即每天销售量相同)；③ 库存曲线近似看作直线段，而不是图J.6 那样的阶梯形者线段。

正如【例 J-11】所表明，关键是求出最优订货次数 n，使得订货费 C_{0n} 和库存费

C_{1n}之和最小。设商品总数量为 Q，总销售时间为 T，每次订货费为 λ，每件商品每单位时间的库存费为 a（通常与库存商品的价格成正比，也可以是约定的单价）。

容易得出 $C_{0n}=\lambda n$，由图 J.7 可见，库存费等于库存单价 a 与 n 个三角形面积的乘积：

$$C_{1n}=an\left(\frac{1}{2}\times\frac{T}{n}\,\frac{Q}{n}\right)=\frac{aTQ}{2n}$$

$$C_{0n}+C_{1n}=\lambda n+\frac{aTQ}{2n}\text{，}\frac{\mathrm{d}}{\mathrm{d}n}\left(\lambda n+\frac{aTQ}{2n}\right)=\lambda-\frac{aTQ}{2n^2}=0$$

求出极值点

$$n=\sqrt{\frac{aTQ}{2\lambda}} \tag{J.29}$$

在【例 J-11】中，$T=30$ 天，$Q=150$ 台，$\lambda=1500$ 元，$a=20$ 元/台/天，由 J.29 式，

$$n=\sqrt{\frac{20\times30\times150}{2\times1500}}=5.477\approx5$$

要比上面的方法简捷得多。

在真实库存问题中，情况要复杂得多。进货往往和订货不一致，会产生滞后；单位时间的销售量通常是随机的。又要避免库存过多，造成商品积压；又要防止库存短缺，供不应求。这方面的深入讨论可参阅[J-4][5-8]。除库存问题外，经济领域的优化问题非常多，数学大有用武之地[J-5]。

《绪论》中提到诺贝尔经济奖与数学密切相关；实际上，许多获奖者有精深的数学造诣，其中不少在本科或研究生阶段主修数学。附录 J.1 为从 1969 年首届诺贝尔经济奖颁发以来直至 2011 年的全部获奖名单；附录 J.2 为 2000 年以来直至 2011 年有主修数学经历的获奖者名单。

J.4　思考与练习

1. 某酸乳酪商行发现酸乳酪的收入函数和成本函数分别为

$$C(Q)=4+3\sqrt{3},\ R(Q)=12\sqrt{Q}-\sqrt{Q^3},\ (0\leqslant Q\leqslant 3)$$

其中 Q 的单位为千升，$C(Q)$、$R(Q)$ 的单位为千元，求边际成本、边际收入和边际利润。

2. 某商店某商品售价为 5 元/件时每天售出 1000 件，当售价为 4.8 元/件时每天多售出 200 件，且多售出的件数与售价的降低数的平方成正比，求需求函数、需求弹性。

3. 求解以下连续供求系统的均衡点，并讨论系统稳定性

$$Q_S(P)=1+5P,\ Q_D(P)=4-2P,\ \frac{\mathrm{d}P}{\mathrm{d}t}=2(Q_D-Q_S)$$

4. 某食品商店根据历史数据和市场预测，确定下一年度某种食品的销售量为 $Q=285$ 吨。其中春、夏、秋、东四季的销售量分别是 50 吨、70 吨、125 吨、40 吨。进货单价为 1 万元/吨，销售单价为 2 万元/吨，每吨每季度的库存费为 0.04 万元，每次订货费为 0.5 万元。试确定库存订货策略。

5. 自行寻找一两个实际经济问题试图用微积分或其他数学方法求解。

参考文献或网站

[J-1] 平狄克(Robert S. Pindyck)，鲁宾费尔德(Daniel L, Rubinfeld)：《微观经济学(Microeconomics)》(第四版，张军，罗汉，尹翔硕，谢识予译)，中国人民大学出版社(2000).

[J-2] 赵胜民，周爱民：《经济数学》，科学出版社(2005).

[J-3] 王高雄，朱思铭：《常微分方程(第 3 版)》，高等教育出版社(2006).

[J-4] A. M. Law, W. D. Kelton：《仿真建模与分析(Simulation Modeling & Analysis (2nd Ed.))》，McGraw-Hill, Inc. (1991).

[J-5] 阿维纳什 K. 迪克西特(Avinash K. Dixit)：《经济理论中的最优化方法(Optimization in Economic Theory)》(冯曲，吴桂英译)，上海人民出版社(2006).

【附录 J.1】

1969–2011 年诺贝尔经济学奖一览

2011：托马斯·萨金特(Thomas J. Sargent)(美)，克里斯托弗·西姆斯(Christopher Sims)(美)(政策因素对经济的影响)

2010：彼特-戴蒙德(Peter Diamond)(美)，克里斯托弗-皮萨里迪斯(Christopher A. Pissarides)(希)，戴尔-莫滕森(Dale. T Mortensen)(英)(极大改进相关的市场理论)

2009：奥利弗·威廉森(Oliver Williamson)(美)(新制度经济学)，艾利诺-奥斯特罗姆(Elinor Ostrom，女)(美)(公共资源管理分析)

2008：保罗-克鲁格曼(Paul R. Krugman)(美)(国际贸易模式和经济活动的地域分析作出贡献)

2007：赫维奇(Leonid Hurwicz)(俄)，马斯金(Eric S. Maskin)(美)，迈尔森(Roger B. Myerson)(美)(为机制设计理论奠定基础)

2006：埃德蒙·费尔普斯(Edmund S. Phelps)(美)(通货膨胀和失业预期关系的分析)

2005：罗伯特·奥曼(Robert J. Aumann)(以/美)，托马斯·谢林(Thomas C. Schelling)(美)(通过博弈理论分析增加了世人对合作与冲突的理解)

2004：芬恩·基德兰德(Finn E. Kydland)(挪)，爱德华·普雷斯科特(Edward C. Prescott)(美)，(有关宏观经济政策的“时间一致性难题”和商业周期的影响因素)

2003：罗伯特·恩格尔(Robert F. Engle(美))，克莱夫·格兰杰(Clive W. J. Granger)(英)(分别用“随着时间变化的易变性”和“共同趋势”两种新方法分析经济时间数列)

2002：丹尼尔·卡尼曼(Daniel Kahneman)(美/以)，1954年希伯来大学心理学和数学双学士，弗农·史密斯(Vernin L. Smith)(美)(为通过实验进行可靠的经济学研究确定了标准)

2001：阿克尔洛夫(George A. Akerlet(美))，斯彭斯(Michael Spance)(美)，斯蒂格利茨(Joseph E. Stiglitz)(美)(在“对充满不对称信息市场进行分析”领域作出贡献)

2000：麦克法登(Daniel L. McFadden)(美)，赫克曼(James J. Heckman)(美)(在微观计量经济学领域作出贡献)

1999：罗伯特·蒙代尔(Robert A. Mundell)(加)(对不同汇率制度下的货币与财政政策以及最优货币区域做出了影响深远的分析)

1998：阿马蒂亚·森(Amartya Sen)(印)(对福利经济学几个重大问题理论上作出贡献)

1997：迈伦·斯科尔斯(Myron S. Scholes)(美)(提出了著名的布莱克·斯科尔斯期权定价公式)和罗伯特·默顿(Robert C. Merton)(美)(对布莱克·斯科尔斯公式所依赖的假设条件做了改进和推广)

1996：詹姆斯·米尔利斯(James A. Mirrlees)(英)(建立不对称信息条件下的经济激励理论)、威廉·维克瑞(在信息经济学、激励理论、博弈论等方面作出贡献)

1995：罗伯特·卢卡斯(Robert Lucas)(美)(倡导发展了理性预期与宏观经济学研究的运用理论)

1994：约翰·海萨尼(John C. Harasanyi)(美)、约翰·纳什(John F. Nash)(美)和莱因哈德·泽尔腾(Reinhard Selten)(德)(在非合作博弈的均衡分析理论方面作出了开创性贡献)

1993：罗伯特·福格尔(Robert W. Fogel)(美)(用经济史的新理论及数理工具重新诠释了过去的经济发展过程)和道格拉斯·诺斯(Douglass C. North)(美)(建立了包括产权理论、国家理论和意识形态理论在内的“制度变迁理论”)

1992：加里·贝克(Gary S. Becker)(美)(将微观经济理论扩展到对人类相互行为的分析)

1991：罗纳德·科斯(Ronald Harryooase)(英)(发现并阐明了交换成本和产权在经济组织和制度结构中的重要性及其在经济活动中的作用)

1990：威廉·夏普（William Sharpe）（美）、默顿·米勒（Merton H. Miller）（美）和哈里·马科维茨（Harry Markowitz）（美）（在金融和证券投资方面做出开创性工作）

1989：特里夫·哈维默（Trygve Haavelmo）（挪）（建立了现代经济计量学的基础性指导原则）

1988：莫里斯·阿莱斯（Maurice Allais）（法）（在市场理论及资源有效利用方面作出了开创性贡献；对一般均衡理论重新做了系统阐述）

1987：罗伯特·索洛（Robert M. Solow）（美）（理论上提出长期的经济增长主要依靠技术进步，而不是依靠资本和劳动力的投入）

1986：詹姆斯·布坎南（James M. Buchanan, Jr.）（美）（将经济分析扩大和应用到社会、政治法规的选择）

1985：弗兰科·莫迪利安尼（Franco Modigtiani）（意）（提出储蓄的生命周期假设）

1984：理查德·约翰·斯通（Richard Stone）（英）（国民经济统计之父）

1983：罗拉尔·德布鲁（Gerard Debreu）（美）（概括了帕累托最优理论，创立了相关商品的经济与社会均衡的存在定理）

1982：乔治·斯蒂格勒（George J. Stigler）（美）（研究工业结构、市场的作用和公共经济法规的作用与影响）

1981：詹姆士·托宾（James Tobin）（美）（阐述和发展了凯恩斯的系列理论及财政与货币政策的宏观模型）

1980：克莱因（Lawrence R. Klein）（美）（建立经济体制的数学模型）

1979：威廉·阿瑟·刘易斯（William Arthur Lewis）（英/美）和西奥多·舒尔茨（Theodore W. Schultz）（美）（研究了发展中国家在发展经济中应特别考虑的问题）

1978：赫伯特·亚·西蒙（Herbert A. Simon）（美）（建立了经济组织内的决策程序基本理论）

1977：詹姆斯·米德（James Meade）（英）与戈特哈德·俄林（Bertil Ohlin）（瑞典）（对国际贸易理论和国际资本流动作了开创性研究）

1976：米尔顿·弗里德曼（Milton Friedman）（美）（创立货币主义理论，提出了永久性收入假说）

1975：列奥尼德·康托罗维奇(Leonid Vitaliyevich Kantorovich)(苏)(创立线性规划)和佳林·库普曼斯(Tjalling C. Koopmans)(美)(将数理统计学成功运用于经济计量学)

1974：弗·冯·哈耶克(Friedrich August von Hayek)(奥)和纲纳·缪达尔(Gunnar Myrdal)(瑞典)(研究了货币理论和经济波动，并深入分析了经济、社会和制度现象的互相依赖)Paul A. Samuelson

1973：华西里·列昂惕夫(Wassily Leontief)(苏)(发展了投入产出方法)

1972：肯尼斯·约瑟夫·阿罗(Kenneth J. Arrow)(美)和约翰·希克斯(John R. Hicks)(英)(研究了经济均衡理论和福利理论)

1971：西蒙·库兹列茨(Simon Kuznets)(乌克兰)(研究了人口发展趋势及人口结构对经济增长和收入分配的关系)

1970：保罗·安·萨默尔森(Paul A. Samuelson)(美)(发展了数理和动态经济理论)

1969：简·丁伯根(Jan Tinbergen)(荷)(经济计量学模式建造者之父)和拉格纳·弗里希(Ragnar Frisch)(挪)(经济计量学的奠基人)

【附录 J.2】

2000–2011 年诺贝尔经济学奖主修数学获奖者一览

2011：克里斯托弗·西姆斯(Christopher Sims)(美)，哈佛大学数学学士(1963)

2010：彼特戴蒙德(Peter Diamond)(美)，耶鲁大学数学学士(1960)

2007：马斯金(Eric S. Maskin)(美)，哈佛大学数学学士(1970)、数学硕士(1972)、数学博士(1976)

2005：罗伯特·奥曼(Robert J. Aumann)(以/美)，纽约州立大学数学学士(1950)、MIT 数学硕士(1952)、数学博士(1955)

2004：爱德华·普雷斯科特(Edward C. Prescott)(美)，Swarthmore College 数学学士(1962)、Case Western Reserve 大学运筹学硕士(1963)

2003：克莱夫·格兰杰(Clive W. J. Granger)(英)，诺丁汉大学数学学士(1955)、统计博士(1959)

2002：丹尼尔·卡尼曼(Daniel Kahneman)(以/美)，希伯来大学心理学和数学双学士(1954)

2001：斯彭斯(Michael Spance)(美)，牛津大学数学学士(1966)

2000：赫克曼(James J. Heckman)(美)，科罗拉多学院数学学士(1965)

【编者注】从 2000—2011 年共有 30 位诺贝尔经济奖得主，其中 9 位有主修数学学历，占 30%。