

第一章 怎样参加大学生数学建模竞赛

叶其孝

北京理工大学 应用数学系

提 要

本章论述了在世界范围人才培养竞争中数学教育所处的重要地位、数学建模及其发展、数学建模竞赛出现的必然性、大学生数学建模竞赛培养学生什么样的能力以及和大学数学教育改革的关系。本章还就怎样参加大学生数学建模竞赛以及同学们关心的问题作了较详细、深入的论述、分析和建议。本章可以作为初次参赛的师生了解大学生数学建模竞赛的参考材料，也可以作为向广大公众作宣传的参考材料。

§ 1.1 人才培养竞争中数学教育 所处的重要地位

人类已经进入了以计算机、网络、数码、光纤和多媒体等为代表的信息时代，进入了科学和技术以惊人速度发展的知识经济时代、“全球化”经济的时代。人们越来越明白并深切地体验到 21 世纪经济竞争的关键是科学技术的竞争，科学技术竞争的关键就是人才培养的竞争，而人才培养的竞争就是教育的竞争，教育改革迫在眉睫。然而，教育改革应该怎样进行，主要应该抓什么，众说纷纭。因而许多国家，特别是西方发达国家，早在十多年前就开始进行有关教育情况、教育改革的调查研究，这些调查研究的共同发现之一就是：科学和数学的教育改革处在关键的重要地位。1989 年经合组织（OECD）在其会员国中发起了一场关

于科学、数学和技术教育正在发生怎样的变化的大规模的调查。调查在 1991~1995 年进行，经历四年之久。于 1996 年 4 月 8 日在美国华盛顿召开的一个会议上公布了由美国斯坦福大学的 J. M. Atkin 教授和英国帝国学院的 P. J. Black 教授执笔撰写的题为“改变课程：科学、数学和技术教育中的创新”的长达 230 页的调研报告。该报告指出：“全世界工业国家的决策者们愈来愈把这些（科学、数学和技术教育）领域看做是经济增长的关键。”（见〔1〕）在 1996 年 7、8 月号《未来学家》杂志公布的一份题为“学生必须掌握哪些知识和技能才能在 21 世纪立于不败之地”的调研报告指出“运用数学、逻辑和推理的技能；熟练的读写能力；以及了解统计学。正如人们一直强调的，数学是一种语言，是一种交流和认识世界的方法。应建立用以指导学校制订和评价数学课程的内容、教学以及评估的标准。掌握数学的概念、计算和问题解决的能力对一个真正有文化的人来说是至关重要的。”（见〔2〕）由美国参议员格伦（John H. Glenn，他是在 1962 年作第一次绕地球轨道飞行的宇航员）担任主席的“21 世纪数学和科学教育委员会”于 2000 年 9 月 27 日公布的题为“亡羊补牢，犹未为晚”的调研报告指出：“在新世纪和千禧年已经来临之际，本委员会确信我们国家和人们未来的安康不仅有赖于我们对我们的孩子们的一般教育有多好，而且特别有赖于我们对他们的数学和科学教育有多好。”（见〔3〕）美国教育部长赖利（Richard W. Riley）在 1998 年 1 月对美国数学研究界和数学教育界的报告中鼓励他们“让我们的国家认清数学的重要性，以使所有的教师教更好的数学以及更好地教数学”的同时，更深情地说出了“‘数学等于机会’。当我们为即将来临的世纪作准备时，我们不可能送给美国的父母和学生别的更关键的信息了”。这是分量很重的话语（见〔4〕）。所有这些充分表明了在培养富有创新能力和竞争力的人才中数学教育的关键重要性，各级各类教育中数学教育改革的紧迫性。当然，并非各级领导、企业界、学术界

和广大的公众都已对此有了共识，甚至可以说距离达成共识还相当中远。尽管世界各国都很重视从幼儿园(学前)、小学、中学、大学和研究生院的数学教育，甚至数学课是每学期都要修的必修课，然而数学教育在相当程度上却没有能跟上科技、经济和社会的迅速发展和变化。数学教学的内容、方法和手段变化甚微，不能体现数学在科技和社会以及个人工作和生活中的重要作用，甚至不能满足专业后继课程的需要。特别是，就业市场的变化以及对计算机技术和软件的迅速发展及其作用的过分的宣传也误导了人们，使人们以为只要掌握计算机和软件就能确保个人的事业前途。各科学分支学科的迅速发展，知识膨胀也要求更多的课时，导致数学课时的减少，教材陈旧，教学方法和手段的落后，数学教师的进修没有及时跟上，加上公众本来对数学的作用缺乏了解甚至有着种种误解。而且，由于各级学校的数学教育中还存在不少问题，更由于许多发达国家正在潜心研究我国数学教育成功经验的时候，在我国却出现了对数学的重要性相当缺乏了解的种种误解的言论(例如见[5]和[6])。这实在是相当令人担忧的。怎样让人们真正了解数学的重要性是数学界和数学教育界的迫切而又艰巨的任务。数学建模的迅速发展，特别是，大学生数学建模竞赛的成功举办为我们完成这个迫切而又艰巨的任务创造了很好的条件。

§ 1.2 数学建模再现雄风

1.2.1 数学建模是用数学来解决实际问题的桥梁

数学科学(Mathematical Sciences,简称数学(Mathematics))是研究现实世界中抽象出来的数量关系和空间形式的科学。数学是一切自然科学的基础，它为其他科学提供语言、观念和方法。自然科学中几乎所有的重大发现无不依赖于数学的发展与进步。数学也是一切重大技术发展的基础，计算机的发明以及当今各个高科

技领域和经济领域的发展是最好的明证。数学又是人类的一种文化，它在教育中一直占有特殊地位，在提高人的逻辑推理能力、分析判断能力、想像力和创造力上具有其他学科所不能替代的重要作用。但是要用数学去真正解决实际问题还需要在实际问题和数学理论和方法之间“搭建桥梁”，数学建模就是“桥梁”之一。实际上，数学的发展始终是和解决各种各样的实际问题密切相联的，数学建模的思想和方法自古以来一直在使用。事实上，数学模型(Mathematical Model)这一术语早就有了(见[7])，它有两个含义，一是用以表达抽象数学概念的物理实物模型，另一个就是我们现在所谓的数学建模。因而不能说数学建模是新东西，不过数学建模(Mathematical Modeling)这一术语确实在 20 世纪 60 年代之后才流行起来的。

确切地说，数学建模就是通过对实际问题的分析，通过抽象和简化，明确实际问题中最重要的变量和参数，通过某些“规律”建立变量和参数间的数学问题(我们也可以说是把实际问题“翻译”为数学问题，或称之为这一简化阶段的一个数学模型)，再用精确的或近似的数学方法求解之，然后把数学的结果“翻译”成普通人能懂的语言，并用现场实验数据或历史记录数据或其他手段来验证结果是否符合实际并用来解决实际问题。这样的过程的多次执行和完善就是数学建模的全过程。欧几里得几何和微积分的发明就是科学史上最光辉、最成功的两个数学模型，正是由于欧氏几何是建立在高度抽象的没有厚度的面、没有长度的线和没有长、宽、高的点的基础上的公理系统，才能应用于土地的丈量等工农业问题的解决；也只有把太阳、地球和月亮等天体抽象成只有质量而没有长、宽、高的点才能把牛顿的微积分用上，从而发现了万有引力。微积分在 17、18 世纪广泛应用的成果也充分显示了数学建模的威力。很有意思的是，从百科全书或字典上看 Modelling (Modeling) 的意思是“塑造艺术”(见[8])或“造型术”。[8] 中说：“……塑造和雕刻相反，它是一种添加性工艺，

它不同于雕刻之处在于塑造过程中可以修正形象”。这和“这样的过程的多次执行和完善”是一致的，因此，也有专家认为数学建模带有艺术的特点。所以，我们也可以把数学建模形象地称为“数学的塑造艺术”。

1.2.2 计算和信息技术的发展使数学建模雄风再现

但是，用数学建模来解决实际问题却强烈依赖于多种因素，不仅要对实际问题有深刻的理解，能抓住主要因素，并能从中作出正确的数学抽象，更有赖于计算技术的发展。微积分发明以来，数学建模用于解决各种各样的实际问题从未停止，但往往停留于某个数学模型（明确的数学问题）的表述而计算不出结果，只好放在一边。浏览一下 20 世纪 50 年代以前的许多工程杂志不难看到这种情况。只有到了 20 世纪后半世纪，计算机、计算技术、仿真技术及数学软件得到了迅速发展（数学理论同样为计算机的发明作好了充分的理论准备），才使数学建模焕发了新生，得到了迅猛的发展。数学建模正在成为科学的研究和工程技术的关键工具。正如〔9〕指出的“一切科学和工程技术人员的教育必须包括数学和计算科学的更多的内容。数学建模和与之相伴的计算正在成为工程设计中的关键工具。科学家正日益依赖于计算方法，而且在选择正确的数学和计算方法以及解释结果的精度和可靠性方面必须具有足够的经验。对工程师和科学家的数学教育需要变革以反映这一新的现实”。此外，历史的经验告诉我们任何一项成功的技术最终必然会进入到培养人的教育领域。事实也是如此，到了 20 世纪 60 年代，英国牛津大学开始为研究生开设数学建模课程，并进一步开创了“牛津工业研究小组（Oxford Study Group with Industry，缩写为 OSGI）”研究课题，每年有 10 天到两周的时间，牛津大学的教授、博士生和工业、企业界的代表会聚一起研讨如何用数学的方法解决工业中的各种问题，并取得了很大的成功。从此，在大学里开设数学建模课程、建立工业数

学研究课题在工业发达国家得到了迅速发展。为了更好地进行交流，多种国际性数学建模杂志也相继创建。现在我们可以毫不夸张地说数学建模的思想和方法研究渗透到科学、技术、工程、经济、管理以至人们的日常生活的一切方面。诺贝尔经济学奖所有获奖者的工作都是有关各种经济行为的成功的数学建模就足以说明这一事实。也正是在这个意义下，可以说“高科技本质上就是数学技术”。这决不是哗众取宠的夸张。

§ 1.3 大学生数学建模竞赛

1.3.1 大学生数学建模竞赛的出现

教育的任务就是要教给学生最基本的知识，特别是要教给学生能使他们在今后的学习和工作中能展现其智慧和能力的思想、方法和顽强的意志力。因此，教学的内容和教学方式就变得十分重要了。由于“数学科学对经济竞争力至关重要，数学是一种关键的、普遍适用并授予人以能力的技术”。（见[10]）“数学的思考方式有着根本的重要性。简言之，数学为组织和构造知识提供方法。一旦数学用于技术，它就能产生系统的、可再现的并能传授的知识，分析、设计、建模、模拟和应用便会变成可能的高效的富有结构的活动。”（见[10]，pp. 44~45）更由于“数学除了锻炼敏锐的理解力、发现真理以外，它还有一个训练全面考虑科学系统的头脑的开发功能”。（见[11]）以及我们前面论述过的数学的重要性充分说明了数学教育对人的教育中的重要地位。但是传统的数学教育往往不能适应科技和经济的迅速发展和变化的形势，使得学生不能充分了解数学的功底对他们今后一生的事业和生活的重要性，许多学生学习数学的积极性下降了。而且数学教育大规模的变革又不太可能，数学界和数学教育界都为此十分担忧。与此同时，一些教师注意到不少学生对用数学去解决各种实际问题非常感兴趣。他们还注意到竞赛实际上也是一种培养学

生的很好的教学活动，能否通过搞竞赛来适应这样一部分学生的要求，同时又能推动教学改革呢？到了 1983 年，在美国终于有一位大学教授提出了能否创办一个和传统的普特南（Putnam）数学竞赛不同的应用数学类型的竞赛呢？经过一年多的讨论，他的建议终于得到认可，并得到美国科学基金会的资助，于 1985 年在美国创办了一个名为“数学建模竞赛”（Mathematical Competition in Modeling 后改名为 Mathematical Contest in Modeling，缩写均为 MCM）一年一度的大学水平的竞赛（见〔12〕或〔13〕, pp. 308~314）。MCM 和有着悠久历史、培养了许多优秀数学家的面向美国大学生的“普特南（Putnam）数学竞赛”（这是一种类似于中学奥林匹克数学竞赛那样彻底闭卷的竞赛，可参阅〔14〕或〔15〕）完全不同，是一种彻底公开的竞赛。MCM 每年只有若干个来自不受限制的任何领域的实际问题，学生以三人组成一队的形式参赛，在三天（72 小时）内任选一题，完成该实际问题的数学建模的全过程，并就问题的重述、简化和假设及其合理性的论述、数学模型的建立和求解（及软件）、检验和改进、模型的优缺点及其可能的应用范围的自我评述等内容写出论文。由专家组成的评阅组进行评阅，评出优秀论文，并给予某种奖励肯定。MCM 只有惟一的一条禁律，就是在竞赛期间不得与队外任何人（包括指导教师）讨论赛题，但可以利用任何图书资料、互联网上的资料、任何类型的计算机和软件等等，这就为充分发挥参赛学生的创造性提供了广阔的空间。我国大学生从 1989 年起就组队参加 MCM，近年来我们的参赛队数已占到全部 MCM 参赛队数的近三分之一，并取得了优异的成绩（见〔13〕）。从 1992 年起我国开始创办我们自己的大学生数学建模竞赛，特别是，教育部高瞻远瞩于 1994 年把全国大学生数学建模竞赛定为仅有的少数几项大学生课外教学和竞赛活动之一，并得到了各级教学行政领导、广大师生和企业界的热烈响应和支持。全国大学生数学建模竞赛发展迅速，现在已经有 26 个省、市（自治区）建立了赛

区，全国 511 所大学的近万名大学生紧张地参加了“2000 年网易杯全国大学生数学建模竞赛”，香港特区的大学生也于今年首次参赛。从 1999 年开始还设立了大专赛题（C 和 D 题）（详细情况可参阅〔13〕）。全国组委会按照“发挥创新意识、发扬团队精神、提倡重在参与、坚持公平竞争”的原则组织竞赛，更特别强调推动数学教育改革和不断扩大受益面，在广大师生的不断实践中很好地推动了我国的大学和中学的数学教育改革，积累了丰富的培养富有创新能力和竞争力的人才的经验。

1.3.2 大学生数学建模竞赛培养学生什么样的能力

为什么大学生数学建模竞赛会受到如此多的大学生的欢迎，为什么我们又希望有更多的大学生来参加这项竞赛呢？这就要首先来看看大学生数学建模竞赛培养了学生什么样的能力。经过十多年来广大参赛同学、指导教师和有关的教育行政领导的总结，至少有以下几个方面是值得提出的。

1. 应用数学进行分析、推理、计算的能力，特别是，“双向”翻译的能力大大提高。贯穿数学建模过程的每一个阶段都不能离开灵活的数学分析、推理、计算，特别是，要把用普通语言表述的实际问题的主要方面用数学的语言表述为明确的数学问题，才能用数学的方法去解决它们。这就是我们所说的一个方向上的翻译能力；当得到数学的结果后，又要能用普通人能懂的语言“翻译”出来，才能用于实际。这就是我们所说的另一个方向上的翻译能力。正如著名的应用数学家 Richard Courant（美国名牌大学纽约大学的 Courant 数学研究所就是以他的名字命名的）指出的，“应用数学的任务是面向外部提出的问题，适合这些问题的形式，把它们翻译成数学语言，分析其模型表示的抽象问题，然后是最后的也是最主要一步，从理论分析转回现实语言并使之合于使用。”（见〔16〕，pp. 59~60，第 169 条）

2. 应用计算机、相应数学软件以及因特网（Internet）的能力大大提高。几乎所有的赛题涉及大量的计算或逻辑运算，因此不掌握计算机和相应数学软件的使用是难以取得好成绩的。又由于赛题可以来自不同的领域，事先又不知道，而现今大量的资料发布在网上，学会在网上迅速查到相关资料也非常有助于取得好成绩。值得指出的是，在常规的教学中这方面的能力是难以得到训练的。
3. 获得应变能力（独立查找文献、在短时间内阅读、消化、应用的能力）的培养。由于事先不知道赛题来自什么领域，从而任何人不可能在赛前就作好一切准备，而恰恰是要求参赛者具有能在短时间内独立查找相关文献、阅读并能部分消化、应用的能力。这样的应变能力在同学们参加工作后尤为重要。因为我们就生活在一个迅速发展变化的社会，我们每个人经常会面临自己不曾想到过的挑战。因而，能否临变而不乱，有没有应变能力对我们自己的工作和事业的成功都是至关重要的。在常规的教学活动中一般是不容易培养应变能力的。大学生数学建模竞赛却给我们提供了初步培养这种能力的机会。
4. 培养和发展同学们的创造力、想像力、联想力和洞察力。因为通过教学活动来培养学生的创新能力是眼下谈论教育改革的重要主题，我们略微多讨论一点。曾有人在谈到人类进入21世纪时，什么会是决定未来的重要元素？是电脑科技？是政治经济？还是别的什么？他的回答是：长远看来，推动人类未来的动力还是那个将人类从蛮荒带向文明的元素：人类的创造力（creativity）。那么，什么是创造力呢？〔17〕中的解释为，“对已积累的知识和经验进行科学的加工和创造，产生新概念、新知识、新思想的能力。大体上由感知力、记忆力、思考力、想像力四种能力构成。”也由于创造力是那么重要，近年来许多专家从不同的角度对创造力和怎样培养创造

力进行了广泛的研究。例如，曾任芝加哥大学心理系主任、现任教于哈佛大学的心理学教授 Csikszentmihalyi 集 30 年研究人类的创造力写成的著作^[18]中就有很详细的论述。创造力的具体实现是各种发明，而这些发明改变了人们的生活方式。创造力的发挥需要寻找或创造实现目标所需要的全部信息的能力，而且能以富有成效和有意义的方式运用这种信息。对个人而言，创造的过程充满了挑战和喜悦。一个真正重要的发明创造不是来自一时的灵感，而是经由多年的辛勤努力。一个有创造力的想法或成品，不仅仅是产生于个人的独立的钻研和努力，往往是依靠许多资源的协助和客观环境配合下众人集体努力的成果。想创造发明一件新事物，必须先学习旧的法则，或者说从模仿到创造是必由之路。创造力并不是与生俱来的，需要一点一点培养。我们的教育在某种意义下就是在潜移默化地培养学生的创造力，但这是不够的。大学生数学建模竞赛提供了一个培养创造力的极好的载体，可让同学们充分体会到创造过程的紧张、艰辛和喜悦。想像力也是非常重要的，正如伟大的物理学家 A. Einstein 所指出的“想像力比知识更重要，因为知识是有限的；而想像力却抓住了整个世界，激励着产生进化的进步”。（见 [19]，p. iii）当然，这里说的想像力不是胡思乱想，必须是合乎逻辑的、有根据的想像。联想力就是把看似不同的事物或现象的本质的、同样的（或相似的）量的关系联系起来。例如，热传导、分子扩散、生态学中的群体动力学、金融行为等现象均可用扩散反应方程来描述。当然，这种由此及彼的能力是经过长期的知识积累和实践才能获得的。但是，有意识的培养也是非常重要的。至于洞察力，就是所谓的“一眼看穿事物本质的能力”，它可以说是上述各种能力的综合表现，同样是需要长期培养的。

5. 培养了学生组织、管理、协调（合作）以及及时妥协的能力。

大学生数学建模竞赛的成功不仅仅取决于队员个人的基础和努力，更依赖于三个队员间的合作和团队精神的发挥。我们在竞赛实践中就曾发现三个优秀学生组成的队没有取得好成绩的例子，原因就在于团队精神没有发挥好。学习成绩优秀的学生往往有其负面，即过于自信、好强，不容易耐心听取别人的意见，不容易发现别人的长处。曾有一个队，三个队员学习成绩都很好，但都过于自信、好强，在讨论方案时争持不下，一天过去了还在争，没能及时妥协，等到意识过来，已经晚了，没有能取得应有的成绩。同样，我们在竞赛过程中看到有的队组织能力很强，能通过大体上的分工合作把每个队员的长处充分发挥出来。例如，数学强一点的同学多负责一些数学分析，计算机强一点的同学多负责一点编程和上机，写作能力强一点的同学从一开始就要考虑论文初稿的写作，而且在开始的方案讨论中能分页记下每一条假设，使得当发现模型有缺陷而需要改进时，可以很方便地把曾讨论过而没有被采用的假设重新提出来，从而节省了时间，提高了效率。当然，我们也看到有的队组织比较混乱，从而导致效率低下，甚至成绩较差。许多参加过竞赛的同学对此深有体会。确实如此，怎样发挥每个队员的长处，特别是在培训、竞赛的全过程中，既能充分看到自己的优点，更能看到自己的不足之处，看到别的队员的长处，对自己能扬长避短，又能充分发挥别人的长处，当自己处于少数情形时，尽管自己认为自己的设想是正确的，为了避免无休止的争论，能及时妥协，尽自己的最大努力来实现多数人的想法，尽可能使之取得最大成功。这才是团队精神，这才是真正有能力，是做大事的人应该具有的品质。但是在常规的教学活动中是很难有机会来培养这样的优秀品质的。凡是意识到这一点的同学，特别是能在竞赛过程中有意识地培养自己的同学，则体会更深，收获更大（参阅〔13〕和〔20〕）。

6. 培养了交流、表达和写作能力。能否在竞赛中取得成功，还有赖于队员之间能否互相把自己的意见、想法清楚地表述出来，这也包括能否耐心地倾听他人的有关表述，并抓住其要点。这说起来容易，做起来是很不容易的。在常规的教学中，同学间、师生间的口头交流的机会是比较少的。但是，对于参赛同学来说，必须通过在讨论班上报告优秀论文、文献或自己的心得，进行讨论，甚至争论并了解什么是大学生数学建模竞赛，怎样才能写出优秀论文，或者学习、掌握、应用新的知识和方法。逐步地把很多同学从羞于说话，不敢大胆表述自己的意见，不愿也不敢争论的初态引导到敢于并善于说话、争论的优良状态。许多同学在赛后的总结中强调了这方面的收获。当然，评奖主要是对论文进行评阅，而且论文表述的清晰与否也是重要标准，这就不但要求口头表达能力，还要求书写表达能力。写作能力的强弱对于同学们今后事业的成功也是极其重要的。有一个例子很说明问题。某赛区五位专家花了好几个小时仔细阅读一个队的论文，觉得似乎有创新之处。但由于该文的表述实在太差，无法看懂，没有评上奖。事后，队员们不服气。有关老师组织了一次讨论班让该队的队员们仔细报告，实际上是不断的提问和答辩，最后大家明白了该队的论文确有创新之处。不过，当老师们再问该队的队员该不该给奖的问题时，队员们心服口服地说确实不该得奖，因为写得实在太差了。同学们真正认识到写作能力的重要性。
7. 获得了竞争意识、坚强的意志力的培养：竞赛当然就是竞争，似乎当然就有竞争意识。事实并非如此，在竞赛中我们往往能看到有的队不是知难而进，往往原谅自己这次竞赛准备不足，下次再说，甚至放弃，看不出有什么斗志。而我们也看到许多队，他们的物质条件很差，计算机和软件都是比较落后的，但是他们不气馁，斗志昂扬、分秒必争地做到最后一刻。

分钟，有的取得了好成绩，有的尽管没有取得好成绩，但大大地前进了一步。他们的共同体会就是大学生数学建模竞赛是培养他们竞争意识和坚强的意志力的极好的载体。

8. 培养了同学们自律、“慎独”的优秀品质。〔2〕中特别强调自律的重要性，指出“‘自律’对行为负责，运用伦理准则以及制定和评估目标的能力（在各种不良影响和腐蚀下），学生要想在 21 世纪立足、生存并大展宏图，必须严以律己，律己有赖于道德准则以及制定和评估他们朝着自己的目标进展情况的能力。”大学生数学建模竞赛的惟一的一条禁律：在竞赛期间不得与队外任何人（包括指导教师）讨论赛题，恰恰是难以监督的。尽管组委会可以组织一定的巡视，采取一定的措施，但最根本的还是要靠自觉遵守。事实上，也确实有极少数同学，甚至教师，由于获奖心切等不正确思想的驱动而作出了犯规的事情。实际上，他们丢掉的却是能培养自己在未来事业中取得成功的优秀品质的大好机会。
9. 培养了正确的数学观（正确理解数学的作用、数学和外界的关系）。尽管我们在前面论述过数学的重要性，但现实情况决不是人人都已经真正认识到了。而大学生数学建模竞赛的参赛者由于亲身参加了竞赛的全过程，他们既深切体会到数学的极端重要性，又不会盲目地认为数学是万能的，可以单打独斗的。他们知道成功地把数学这一工具用于解决实际问题，一定要了解与实际问题相关的领域和学科的知识相结合，要与有关专家相结合。他们中间会有许多人成为将来的学术带头人，甚至各级政府的领导人，他们是否具有正确的数学观对于我们国家的发展是至关重要的。对于数学教师来说，也是极为重要的。因为，大部分数学教师是从数学系培养出来的，由于种种局限性，多数人对数学本身比较了解，但对数学和外界的关系往往是了解不够，甚至片面。因而很难成为一个真正优秀的数学教师。

正因为大学生数学建模竞赛的全过程在相当程度上“模拟”了同学们今后到了工作岗位后的实际情况以及为取得成绩和成功所需要的能力，受到了广大师生的热烈欢迎。他们用不同的表述方式表达了同一个体会，即“一次参赛，终身受益。”（见[13]、[20]等）正因为培养了这些能力，许多参加过大学生数学建模竞赛的毕业生在工作中、在深造学习中都作出了优秀成绩，受到用人单位的好评，有的甚至被委以重任。

1.3.3 怎样参加大学生数学建模竞赛

对于有兴趣参加大学生数学建模竞赛活动的同学，为了更好、更主动、积极地参加这项活动还应对竞赛的全过程有一个了解，因为竞赛决不是仅仅三天的事。

大学生数学建模竞赛大体上可分为三个阶段：赛前培训，竞赛三天的拼搏，赛后的继续。下面我们来对这三个阶段作较为详细的说明。

1. 赛前培训

赛前的培训活动的目的最主要的是了解什么是数学建模，什么是大学生数学建模竞赛。当然，这种了解是相对而言的。有的同学只需要自己阅读若干资料大体上了解一些情况就可以参赛，这当然是非常好的。但是，大学生数学建模竞赛作为一种希望面向更多同学的课外科技、教学活动，只有某些系或专业的一些同学参加是不够的。根据我国很多学校多年来的经验，为了更好地培训选手，也为了吸引更多的同学参赛，赛前的培训活动是必要的，它不是出于仅仅为了夺取金牌的纯锦标主义的功利目的，而是出于怎样更好地培养同学。培训活动大体上包括以下内容。

开设必修或选修的数学建模课，其主要内容是通过教师讲授与学生的实践，既初步学习到什么是数学建模，什么是大学生数学建模竞赛，也初步学习一些数学软件的使用。数

学建模课的前修课程通常是高等数学、线性代数、概率和统计初步，开设该课的时间通常是在二年级。开设数学建模课程也是在大学生中普及数学建模的思想和方法的最切实可行、最有效的方法。当然，怎样开好这门课程，使之能吸引更多的同学是教学改革的重要方面。

对于有兴趣参加大学生数学建模竞赛的同学开设数学建模讨论班，讨论班的主要内容是，让自己报告已有的大学生数学建模竞赛优秀论文，或者相关的数学内容（它们可以是由教师帮忙寻找、指定，也可以由学生自己确定）。关键是学生自己阅读、消化后在讨论班上报告，相互展开讨论，甚至争论。教师既是讨论班的组织者、领导人，更要通过质疑和答疑扮演引导者的角色，把讨论班搞得生动活泼来培养学生的表达、交流能力，也为相互了解、进一步合作打下良好的基础。只有通过同学们自己对优秀论文的阅读、消化、报告和讨论（争论），他们才能真正了解竞赛要求的论文应该是什么样的。为了在竞赛中取得好成绩，这是至关重要的。教师从讨论班上也可以得到有关数学教育改革的很多启发。

适当地扩大学生的知识面，在最优化、图论、概率统计、微分方程、数学软件等方面适当开设一些讲座或启发性的短课程。时间通常也就是几个小时，既讲了主要内容和方法，更提供相当多的参考资料，鼓励同学们自己去进一步阅读、消化，有的可以在讨论班上报告、讨论。同时也鼓励同学们自己去寻找相关的文献和资料。这样做一定会大大提高同学们的自学能力和应变能力。

学习、掌握有关数学软件的使用。众所周知，数学软件的使用入门介绍是容易的，不需要很多时间，但是要真正掌握，并能熟练、灵活地应用是要花很多时间的，而且必须亲自去实践。怎样才能用尽可能少的时间来做到这一点正是教学改革的重要任务之一。这里问题的设计，特别是结合实

际问题中的应用可能是一种好方法。

做一到二次模拟考试. 通过完全实战性的模拟考试，测试一个队三人合作能否在 72 小时内完成竞赛的各项要求，同时也进一步磨合三人的配合。指导教师可以给出试题，并对学生的论文进行评阅并讲评。同学们自己也可以进行研讨、讲评，甚至和以前的优秀论文进行比较研究，找出自己的优点和有待改进之处。由于能报名参赛的队是有限的，有的学校也利用这种模拟考试来作为挑选最后报名参赛队的参考。

当然，我们不赞成搞太多的赛前培训，因为这会影响到正常的教学秩序。但是，我们支持适度的、能调动学生学习主动性，特别是有利于大学数学教育改革的培训。有关大学生数学建模竞赛和大学数学教育改革的联系，我们将在 § 1.4 中作适当的论述。

2. 竞赛三天的拼搏

参加竞赛就是要力争领先，这是完全可以理解的，但是更应该把它看做是自己在漫漫人生道路上无穷多次挑战中的一次，对自己是否具有坚强意志力的一次考验。这样就能做到心态平静地全力以赴，最大限度地发挥自己。同时也要考虑适合于自己队的战略和策略。队员们在理解题意后一定会有许多各种各样的想法，而三天的时间是很短的，只可能完成其中的一部分想法。因此，要尽可能把三个人的想法集中到大家都同意的想法上，众志一心地去完成。这里，我们要特别强调认真审题，以及首先要明确回答试题要求参赛者回答的问题，然后再进行发挥或者对试题进行讨论甚至批判。还要注意表述一定要明确、清晰。因为，论文评阅的标准，或者说原则是“假设的合理性，建模的创造性，结果的正确性，表述的清晰性”。有的同学由于审题不仔细，结果出现某种“答非所问”；又由于赛题都是开放题，完全可以有不同的理解，甚至在赛题中也会出现某种疏漏。这本来是完全可以

理解的，但是有的同学却因此而把注意力放到对题目的批判上去，而没有首先回答赛题要求回答的问题，又形成一种“答非所问”，使得评阅组很难作出判断。当然，更重要的是要在三天的拼搏中尽量发挥出前面提到的种种能力。

3. 赛后的继续

多数同学在竞赛结束后最迫切需要的是好好睡一觉，然后就等成绩的公布。如获奖则高兴，如没有获奖，就不高兴，甚至抱怨。我们说，这是很消极的态度。事实上，真正的收获并不完全在于获不获奖，而在于竞赛是否考验、锻炼了自己的能力，是否做到了“赛而后知不足”。善于总结，才能往更高的境界前进。特别是，竞赛中的许多想法并没有实现，为什么不继续去做呢？更何况竞赛已经结束，竞赛的惟一禁律已经不起作用，老师和同学可以在一起切磋、讨论。这就是我们所说的赛后继续的阶段，它也往往是培养同学的更重要的方面，正引起越来越多的师生、教育行政领导的重视。事实也是如此，许多学校通过讨论班、交流会报告各队的解法然后进行讲评，讲述自己的体会，对教育改革提出建议等方面进行总结交流。例如，不少队在赛后与指导教师一起对赛题进行深入研究，获得了更好、更完善的结果，有的甚至被有相当水平的学术杂志录用发表，这时师生们的喜悦更大、体会更深。有的学校还成立了学生数学建模协会、学生工业与应用数学学会等科技活动组织，从而使数学建模的教学和科技活动成为经常性的活动。赛后继续的阶段正引起越来越多的重视。许多学校行政领导也进一步意识到大学生数学建模竞赛的三个阶段的全过程正是培养优秀的富有创造力和竞争力的人才的有效方法，正在进一步研究使之更加完善、更加系统。

1.3.4 大学生们经常会提出的几个问题

1. 数学上我要作多少准备才能参赛?

之所以提出这样的问题是很自然的，完全可以理解的。因为，不同水平的体育竞赛就首先要通过所谓的“及格赛”、“资格赛”。是的，大学生数学建模竞赛原则上只允许大学生参加。至于说到要有多少数学上的准备或资格才能参赛，那就不太好回答了。数学水平越高、实际知识越多当然越好，但这决不是参赛资格。为了回答这个问题，首先要弄清楚大学生数学建模竞赛是一项普及数学建模、培养创造性的课外科技、教学活动，它是由有机结合的三个阶段组成的培养人的载体。每个大学生都可以充分自由地利用这个载体，因而，可以说不存在资格的问题，只存在想不想利用它挑战自己的问题。当然，对数学也不是没有一点要求，至少要学过大学的高等数学（微积分）、线性代数和概率统计初步。然而，更重要的是能否应用数学的知识、思想和方法去解决数学建模的问题。更大的收获也许是通过竞赛逐步了解在实际应用中需要哪些数学。

2. 很想学好数学，特别是数学建模，但正课学时很紧，活动也多，想参加又怕耽误正常的功课。想学好数学一定要参加大学生数学建模竞赛吗？

当然不是，应该说学好、掌握好数学的方法和途径是多种多样的，条条大路通罗马！我们鼓励学有余力、对大学生数学建模竞赛有点心动的同学花一点时间来更多地了解这项竞赛、参加这项竞赛。我们并不鼓励学习有困难的同学来参加这项竞赛，但是我们希望这些同学也能多多了解这项竞赛。此外，同学们各自的兴趣也未必相同，可以参加电子设计竞赛、计算机（信息）竞赛、外语演讲比赛等等。大学生数学建模竞赛只是一项可供选择的课外科技、教学活动。

3. 我们是普通大学的学生，和重点大学的学生一起参赛是不平等竞争；我们是农、林、医，甚至是文科专业的学生，和理工科的学生一起参赛是不平等的竞争。

确实来自不同大学和专业的学生情况是不同的，他们的数学基础、学习条件等方面是有差距的，但是这不能说是不平等的竞争。因为，竞赛的规则是一样的，阅卷的标准是一样的，允许利用的资源是一样的，竞赛主要是考核参赛者在数学建模中的创造性，是为了通过这项活动更好地掌握数学建模的思想和方法，从而培养、发展自己的创造力、竞争力。当然，来自重点大学的同学他们确实有着主、客观的优势和有利条件，只要他们一样努力，他们作出优异的成绩，取得更多的好名次本来就是应该的、合理的。但是，这决不是说来自非重点大学或非理工科的队就一定不能取得好成绩，事实也确实不是这样。不少获得全国一、二等奖的队来自非重点大学、非理工科专业。关键在于是否真正认识到大学生数学建模竞赛是一项很好的挑战自己、锻炼自己的活动，主要是看自己是否有进步。此外，如同赛跑一样，和比自己跑得快的人一起跑往往更能提高自己的成绩。参赛者要有这样的志气：我就是要和强者一起比赛看看自己究竟有多少差距（可以通过与优秀论文相比较，可以邀请获得高等级奖的同学来作报告进行交流，从而来看自己队的优缺点），激励自己更加奋发向上。

4. 没有拿到好名次的奖是否就说明自己不行？

当然不是。常胜将军并不是从来没有打过败仗，更何况“失败是成功之母”。当然，我们不是说现在拿到好名次的同学不行，应该说他们在前进的道路上有了一个良好的开始。我们要说的是，这次没有拿到好名次应该说是一次挫折，但决不是“完蛋”，关键在于是否善于总结，是否真正找到自己的长处和短处，真正知道差距在哪里，应该往何处努力，而

且真正去努力了。真正的强者一定是善于总结、有顽强毅力、不断进取的人，他们既能做到“人贵有自知之明”，又能努力做到“不以一时一地的胜败论英雄”。这里有非常深刻的哲理！

§ 1.4 大学生数学建模竞赛与数学教育改革

任何竞赛活动总是一部分(甚至是一少部分)人的活动，但是竞赛活动的影响却常常可以遍及活动以外的很多人，更可能产生巨大的影响。大学生数学建模竞赛就是这样一项活动。大学生数学建模竞赛实际上是一项不打乱正常教学秩序、规模不小的数学教学改革试验。近年来的教改实践也充分表明了确实如此。无论是在课程教学中切入应用(数学建模)，在课程教学中切入数学实验，把数学建模的思想融合到主干课程中去等诸多方面都反映了大学生数学建模竞赛的实实在在的影响。

我们要特别提出两点。学生是整个教学工作的主体，大学数学教育改革成功的关键是激起大学生对数学重要性的切身会与真正的认识，从而激发起学生出自内心的努力学习，以及学好数学的激情和动力。大学生数学建模竞赛及相关活动的开展确实起到了这样的作用。其次，高质量的教师是动力，他(她)们通过其教学活动不断地指导与提高学生的学习质量和活力。十多年来通过指导学生参加竞赛培养了一批优秀的青年教师，他们既有扎实的数学理论基础，又了解怎样通过数学建模的方法去解决实际问题，并能在教学、科研中运用数学建模的思想、方法和技术手段，从而为培养更多的优秀学生打下了坚实的基础。

综上所述，我们完全有理由相信在教育部的正确领导和指导下，在全国组委会，各省、市、自治区教育厅，各赛区组委会，学校领导和广大师生的共同努力和通力合作下，一定会有越来越

多的大学生参加到各种形式的数学建模的教学活动中来，全国大学生数学建模竞赛一定会得到进一步健康地发展，为祖国培养出更多的富有创新能力和竞争力的人才。

参 考 文 献

- [1] A global revolution in science, mathematics and technology education, FORUM—Education Week, April 10, 1996, 1~8. 参见《参考消息》1996, 4, 30 社会·文教版(6版)题为“如何使科学、数学、技术三位一体——经合组织成员国掀起一场教育运动”的简要报道。
- [2] 1996年7, 8月号《未来学家》发表的由 Donna Uchida, Marvin J. Cetron 和 Floretta Mckenzie 执笔写的“What students must know to succeed in the 21st century (学生必须掌握哪些知识和技能才能在21世纪立于不败之地)”的特别报告；也可参考《参考消息》1996, 8, 10~18 社会·文教版何海光的译文。
- [3] Report of the Senior Assessment Panel for the International Assessment of the U. S. Mathematical Sciences, NSF, 1998, 3; 美国科学基金会发布的“资深评估小组对美国的数学科学的国际评估报告”。
- [4] Richard W. Riley, The state of mathematics education: Building a strong foundation for the 21st century, Notices of AMS, v. 45 (1998), no. 4, 487~491; 数学译林, v. 17 (1998), no. 3, 252~256, 207.
- [5] 苏文洋, 学非所用的浪费, 北京晚报, 1996年6月16日, 3版。
- [6] 王晋堂, “过剩”的话题, 光明日报, 1997年11月27日, 2版。
- [7] 《简明不列颠百科全书》, 7卷, 中国大百科全书出版社, 1986, 366~367。
- [8] 《简明不列颠百科全书》, 7卷, 中国大百科全书出版社, 1986, 547。
- [9] A. Friedman, J. Glimm, J. Lavery, The mathematical and computational sciences in emerging manufacturing technologies and management

practices (正在出现的制造技术和管理实践中的数学和计算科学) —SIAM Report on Issues in the Mathematical Sciences—, SIAM, 1992, pp. 62~63.

- (10) Mathematical sciences, technology, and economic competitiveness, Edited by James G. Glimm, National Academy Press, Washington, D. C., 1991, p. 12; 数学科学·技术·经济竞争力, 邓越凡译, 南开大学出版社, 1992, p. 3.
- (11) R. E. Moritz Ed., On mathematics and mathematicians, Dover new edition, 1958, p. 226; 数学译林, v. 7 (1988), no. 4, p. 340.
- (12) 叶其孝, 美国大学生数学模型竞赛及一些想法, 高校应用数学学报, v. 4 (1989), no. 1, 137~145.
- (13) 李大潜主编, 中国大学生数学建模竞赛, 高等教育出版社, 1998.
- (14) Leonard F. Klosinski et al., The fifty-ninth William Lowell Putnam mathematical competition, The American Mathematical Monthly, v. 106 (1999), no. 9, 842~849.
- (15) Leonard F. Klosinski et al., 第五十七届 William Lowell Putnam 数学竞赛, 数学译林, v. 18 (1999), no. 2.
- (16) J. N. Kapur, 数学家谈数学本质 (1973), 王庆人译, 北京大学出版社, 1989.
- (17) 辞海, 上海辞书出版社, 1999 年版, 善及本.
- (18) Mihaly Csikszentmihalyi, Creativity—Flow and the psychology of discovery and invention, (创造力及发现和发明的心理学) Harper Collins Publishers, 1996 (第 1 版), viii + 456pp.
- (19) M. F. Rubinstein, Tools for thinking and problem solving, Prentice-Hall, Inc., 1986.
- (20) 李尚志, 一次参赛, 终身受益——MCM 参赛者谈收获, 数学的实践与认识, 1995 年第 4 期, 35~42.

第二章 混凝土地板的温度变化

叶其孝

北京理工大学 应用数学系

提 要

本章介绍了 1994 年美国大学生数学建模竞赛 (MCM-1994) 的竞赛情况以及论文的评阅和奖励，特别是介绍了 A 题的优秀论文，实践者的评述和我们的评注。主要内容：MCM-1994 的评阅、结果和奖励；MCM-1994A 题；北卡科学与数学学校队的优秀论文；实践者的评述；我们的注记(包括可供参考的其他优秀论文)。

§ 2.1 MCM-1994 的评阅、结果和奖励

本次竞赛共有包括美国、中国、香港等 9 个国家和地区的 192 所大学的 315 个队参加，其中中国有 33 所大学的 84 个队参加。

初评(“粗评”，*triage*, n. A system designed to produce the greatest benefit from limited treatment to those who may survive and not to who have no chance of survival and those who will survive with out it.) 是在马里兰州的 *Salisbury* 大学进行的，共有 14 位评阅人。每篇论文由两个初评评阅人评阅，摘要和论文的组织是论文评定的基础。如果两个评阅人的评分不同则进行协商，如果协商后还不一致，则再由第三位评阅人来评阅。终评是在加州的哈维·马德 (Harvey Mudd) 学院进行的，A 题评阅人有 9 位，B 题评阅人有 17 位。

MCM-1994A 题是由加拿大阿尔伯达 (Alberta) 大学的 Murray Klamkin 提供的, MCM-1994B 题是由纽约市立大学 (CUNY) 约克 (York) 学院的 Joe Malkevitch 和美国军事学院 (西点军校) (US Military Academy) 的 Steve Horton 提供的。评出的最后结果是:

	O	M	H	P	合计
MCM-1994A 题获奖队数(中国队数)	1(0)	13(6)	21(4)	35(4)	70(14)
MCM-1994B 题获奖队数(中国队数)	5(0)	39(12)	59(16)	142(42)	245(70)

其中, O=Outstanding=特等奖, M=Meritorious=一等奖, H=Honorable Mention=二等奖, P=Successful Participant=成功参赛奖。

每个参赛队都将获得由竞赛主任和每题的评阅组长签名的证书。美国运筹学会 (ORSA) 给予两个获得特等奖的队全程路费资助, 参加于 1994 年 4 月在波士顿举行的该学会的年会, 请他们在特设的 MCM 分会上作报告, 并给以现金奖励, 这两个队分别是来自衣阿华州的格林内尔 (Grinnell) 学院队和加拿大的多伦多大学队(他们做的都是 B 题)。

美国工业与应用数学学会 (SIAM) 对每题指定一个特等奖队作为 SIAM 的获奖队, 每个队员获得 \$ 150 的现金奖励, 每个队得到半程路费资助, 参加于 1994 年 7 月在 San Diego 举行的 SIAM 年会, 并在特设的小型讨论会上作报告。这两个队分别来自美国的北卡科学与数学学校队(A 题)和北卡大学队(B 题)。

§ 2.2 MCM-1994A 题

美国住房与城市发展部 (HUD) 正在考虑建造从单幢住宅到公寓楼大小不同的住宅, HUD 主要关心的是使房主定期支付的费用——特别是暖气和冷气的费用——最少。建房地区位于全

年温度变化不大的温带地区。

通过特殊的建筑技术，HUD 的工程师能建造不依靠对流，即不需要依靠开门、开窗来帮助调节温度的住宅。这些住宅都是只用混凝土地板作为仅有的地基的单层住宅。你们被聘为顾问来分析混凝土地板中的温度变化，由此决定地板表面的平均温度能否全年保持在给定的温度舒适范围内。如果可能的话，什么样的尺寸和形状的混凝土地板能做到这点。

第一部分：地板温度

给定了由表 1 给出的每天的温度变化范围，试研究混凝土地板中温度的变化。假定最高温度在中午达到，最低温度在午夜达到。试决定能否在只考虑辐射的条件下设计混凝土地板使其表面的平均温度保持在给定的温度舒适范围内。一开始，热是通过暴露在外的混凝土地板的周边传入的，而混凝土地板的上、下表面是绝热的。试就这些假定是否恰当，假定的敏感性作出评述。如果你们不能找到满足表 2-1 条件的解，你们能找到满足由你们自己提出的表 2-1 的混凝土地板的设计吗？

表 2-1 以华氏表示的环境温度的日变化

	周边环境温度	舒适温度范围
最高	85°F	76°F
最低	60°F	65°F

第二部分：建筑物温度

试分析一开始所作假定的实用性，并将其推广到分析单层住宅内的温度变化。住宅内温度能否保持在舒适温度范围内。

第三部分：建筑费用

考虑到建筑物的各种限制和费用，试提出一种考虑 HUD 关于降低甚至免去暖气和冷气费用这一目标的设计。

§ 2.3 创新的供热技术——北卡科学与数学 学校队的优秀论文^[1]

2.3.1 摘要

代之以通过门窗的对流，室内的加热和致冷系统，住宅的温度变化仅仅是作为混凝土地板的传热的结果。给定了周围环境温度的一张表，什么样形状和尺寸的混凝土地板才能保持舒适的地板和室内的平均温度。

我们考虑的第一个模型认为地板的顶部和底部绝热，只是通过地板周边的简单热传导的模型。我们对任意形状的地板完成了一般分析，并找到了一个能很好近似地板上任意一点温度的正弦函数；通过地板表面的数值积分来得到平均温度。我们把这个模型用到圆形和方形地板上去，发现大小至少为 $30m \times 30m$ 的方形地板将能有效地控制住宅内的温度。

假设地板的底部绝热可能是不现实的，在我们的第二个模型中地板底部和周边一样可以有热传导。我们再次找到了一条正弦曲线，并应用于方形地板。板上任意一点的温度变化总是在地基温度的周围振动。

然而，如果我们希望通过混凝土地板的传热来加热居室的话，假设地板的顶部绝热是不得要领的。在我们的第三个模型中，我们考虑了通过地板的辐射来加热住宅中的空气。（UMAP 编者注：我们略去了这个模型的细节）

我们建议住房与城市发展部建造能与地基交换热量的相对大的方形住宅。

2.3.2 问题的重述

代之以通过门窗的对流，室内的加热和致冷系统，住宅的温度变化仅仅是作为混凝土地板的传热的结果。表 2-2 给出了用绝

对温度表示的最高和最低的周围环境温度和舒适温度.

表 2-2

用绝对温度给出的温度

	周边环境温度	舒适温度范围
最高	302.6°K	297.6°K
最低	288.7°K	291.5°K

因为我们假定周围温度按正弦曲线变化，按所述的最高和最低温度的近似正弦函数为

$$T_e(t) = 6.94 \cos\left(\frac{2\pi t}{1440}\right) + 295.7, \quad (1)$$

其中 $t=0$ 表示正午时刻， t 的单位用分来表示.

如果地板的顶部和底部都绝热的话，地板的形状和尺寸应怎样才能保持舒适温度？当把顶部和底部的绝热条件去掉时我们的解答会怎样变化以及怎么能使建筑物内达到舒适温度？

2.3.3 假设

问题陈述中隐含的假设

- 周边温度按近似的正弦曲线变化，正午温度最高，午夜温度最低。周边温度曲线可以随、也可以不随季节变化。若它随季节变化，变化也是微小的，而且最高和最低温度之差不变。
- 住宅的墙壁绝热性良好，所以墙壁外的温度变化只影响混凝土地板的温度。
- 混凝土地板的顶部和底部可以绝热，也可以不绝热。若不绝热的话：
 - 热从地基向地板底传热。
 - 地基的温度保持常温 $70^{\circ}\text{F} = 294.3^{\circ}\text{K}$ ，而且不因混凝土地板的热损失/辐射而变化。
 - 在居室中热从混凝土地板向居室空间传送。

附加的假设

- 混凝土地板是由标准的 Portland 和砂子、碎石混合组成的。

其比热为 $2\text{J/kg}^{-\circ}\text{K}$ (2 焦耳/ $\text{公斤}^{-\circ}\text{K}$ ，实际上是 $1.4 \sim 3.6$)，辐射系数为 0.9 ，密度为 3000kg/m^3 (3000 公斤/ 立方米)。

- 混凝土地板的厚度是均匀的。
- 混凝土地板周边的温度实际上就是周围的环境温度（若周边不绝热就是这种情况）。
- 当考虑两个方向的热流时，一个方向的热流与另一个方向的热流无关。类似地，当把混凝土地板看成是由无穷多根导热杆组成时，单个杆之间没有热传导。
- 在居室的空气中可以有也可以没有湍流。如果有湍流，它实际上是使空气中的温度变均匀。

表 2-3

符号索引

A	地面的面积；混凝土地板的顶表面或底面的面积
C	杆的横截面的面积
c	材料的比热；对混凝土而言为 $2\text{J/kg}^{-\circ}\text{K}$
d	圆内一点 P 到圆周上一点的距离
h	混凝土地板的厚度
J	见(8)
k	材料的热传导系数；对混凝土而言为 $90\text{J/minute}^{-\text{m}}^{-\circ}\text{K}$ (90 焦耳/ $\text{分钟}^{-\text{米}^{-\circ}\text{K}}$)
O	平面上的点
ΔQ	热
r	热流
R	从一点向外辐射的不同的杆的长度
s	圆板的半径
t	板的周边的弧长
T	时间
T_a	混凝土地板顶部一点处的温度
T_s	周围环境的温度
u	地基温度， 294.3°K
θ	正方形混凝土地板周长的一半
ρ	密度

2.3.4 混凝土地板顶部和底部都绝热的模型

一般的凸混凝土板中的热传导

我们认为混凝土板周边的温度为周围环境的温度。因为混凝土板顶部和底部都是绝热的，所以板中的传热只能以热传导的方式发生。在一杆中，从温度为 T 的一端到温度为 T_a 的另一端的稳定热流由

$$\frac{\Delta Q}{\Delta t} = kC \frac{T_a - T}{r}$$

给出，其中 k 是材料的均匀的热传导系数， C 是杆的不变的横截面面积， r 是杆的长度，我们把这个方程理想化为

$$\frac{dQ}{dt} = kC \frac{T_a - T}{r}. \quad (2)$$

热流 dQ 也与杆的质量 m ，比热 c 和温度的改变 dT 成比例，
 $dQ = mcdT.$ (3)

把 dQ 代入(2)，我们得到 (UMAP 编者注：如同在后面的实践者的评论中指出的，这里的两个 dQ 是不同的)

$$\frac{dT}{dt} = \frac{kC(T_a - T)}{mcr}. \quad (4)$$

现在我们考虑面积为 A ，密度为 ρ ，厚度为 h 的凸混凝土板。我们把要研究其温度的混凝土板看成是由板上一点向外辐射开去的角度从 0 到 2π 的许多杆组成的，如图 2-1 所示，其横截面取作杆在周边处的面积，即， $dC = hds$ ， ds 是周边处的弧长。(UMAP 编者注：本文计算经常是假设 $ds = rd\theta$ ，就结果而言，这给出了 O 位于中心处的情形的大体上正确的结果。事实上，在极坐标系中微分 ds 的表达式是 $ds^2 = r^2 d\theta^2 + dr^2$ ；例如参见 L. Ross, L. Finney, George B. Thomas, Jr., Calculus, p. 703. (Reading, MA: Addison-Wesley, 1990)).

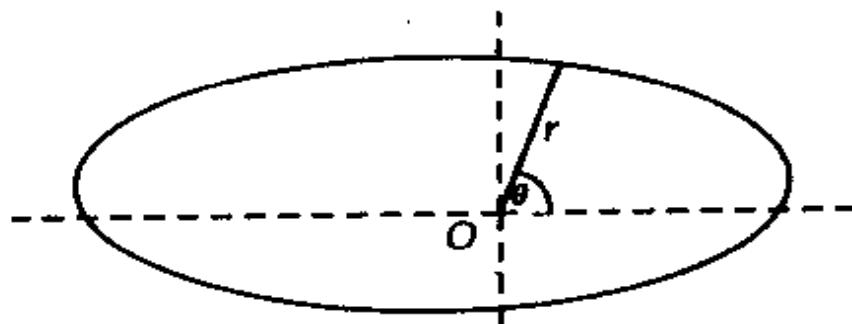


图 2-1 一般的混凝土地板

此外, $m = \rho Ah$, 故(4)成为

$$\frac{dT}{dt} = \frac{\kappa(T_a - T)}{\rho Ahc} \sum_n \frac{C_n}{r_n} = \frac{\kappa(T_a - T)}{\rho Ahc} \int \frac{dC}{r(\theta)}. \quad (5)$$

因为 $dC = hds$, 我们可以把(5)写作

$$\frac{dT}{dt} = \frac{\kappa(T_a - T)}{\rho Ahc} \int_0^{2\pi} \frac{ds}{r(\theta)}. \quad (6)$$

注意 T 与混凝土地板的厚度无关.

利用(1)中关于 T_a 的表达式并应用待定系数法, 我们求得(6)的解为:

$$T = \left[\frac{6.94 \cdot \left(\frac{2\pi}{1440} \right) I}{\left(\frac{2\pi}{1440} \right)^2 + I^2} \right] \sin\left(\frac{2\pi}{1440}t\right) + \left[\frac{6.94 I^2}{\left(\frac{2\pi}{1440} \right)^2 + I^2} \right] \cos\left(\frac{2\pi}{1440}t\right) + 295.7, \quad (7)$$

其中

$$I = \frac{\kappa}{\rho Ahc} \int_0^{2\pi} \frac{ds}{r(\theta)}, \quad (8)$$

我们通过简单地改变与不同的混凝土地板的形状相应的 $r(\theta)$ 和 A 把这个方程用于许多不同的情形.

为求平均温度, 我们在地板面上求温度的积分并除以地板的面积. 我们有

$$\bar{T} = \frac{\iint_A T dA}{A}$$

非凸的混凝土地板的模型

在计算非凸混凝土地板中的温度时碰到了极大的困难。总可以把一个非凸区域分成若干个较小的凸区域，它们每个都可以看做是杆的聚集物。于是一点处温度的基本表达式具有如同凸混凝土地板同样的一般特征。然而，凸混凝土地板问题的分析是极难进行的，因为沿着区域边界的“周围环境温度”是在变化的。此外，暴露在空气中的非凸混凝土地板的周边长度要大于具同样面积的凸混凝土地板的周边长度。因此，在非凸混凝土地板中的温度震荡要比在凸混凝土地板中的温度震荡厉害，而且其震荡很可能超出舒适区域的范围。所以，住房与城市开发部大概不应该建造地板形状为非凸区域的住宅。

圆形混凝土地板

我们首先讨论圆形混凝土地板，因为它具有最小的周长—面积比（即，周长一定面积最大）。考虑半径为 R 的圆形混凝土地板上的一点 P ， P 到中心的距离为 $(R-d)$ ，如图 2-2 所示。

利用余弦定律可得 $r(\theta)$ 的表达式：

$$r(\theta)^2 = R^2 + (R-d)^2 - 2R(R-d)\cos\theta.$$

把表达式代入(6)就给出(7)中的解，其差别只是 I 的公式中的 $r(\theta)$ 不同而已。图 2-3(a)和 2-3(b)表示了地板上一点关于时间变化的温度。

正方形混凝土地板

由于造价的原因，不会有太多人取造圆柱形的房屋。所以，我们考虑如图 2-4 所示的周长为 $2u$ 的正方形混凝土地板。

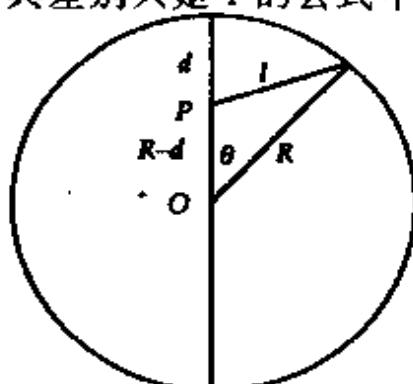


图 2-2 圆形混凝土地板

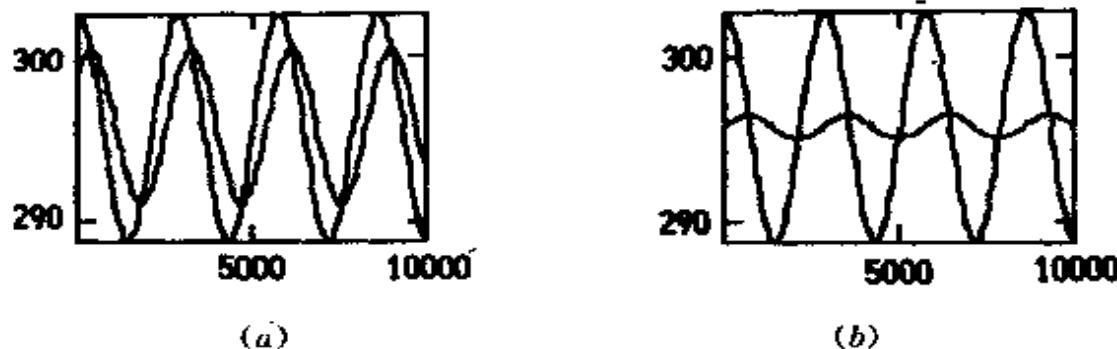


图 2-3 关于时间变化的周围环境温度(变化较大)以及圆形混凝土地板上一点的温度(变化较小)

(a) $R=10$ 而 $d=5$. (b) $R=30$ 而 $d=15$

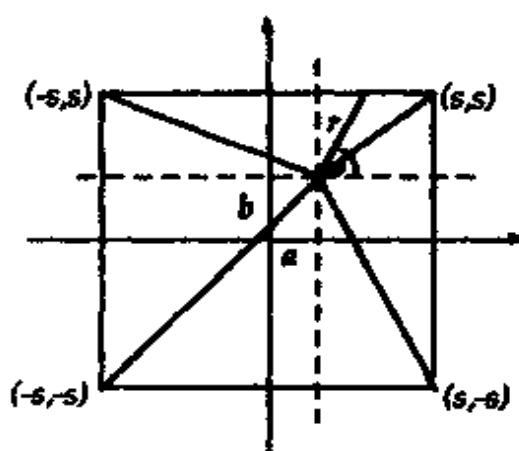


图 2-4 正方形混凝土地板

设 a 和 b 是图 2-4 所示板上一点的直角坐标. 正方形混凝土地板的 $r(\theta)$ 的分段定义如下:

$$r(\theta) = \begin{cases} \frac{s-a}{\sin\theta}, & \arctan\left(\frac{s-a}{s-b}\right) < \theta \leq \arctan\left(\frac{s+b}{s-a}\right) + \frac{\pi}{2}; \\ -\frac{s+b}{\cos\theta}, & \arctan\left(\frac{s+b}{s-a}\right) + \frac{\pi}{2} < \theta \leq \arctan\left(\frac{s+a}{s+b}\right) + \pi; \\ -\frac{s+a}{\sin\theta}, & \arctan\left(\frac{s+a}{s+b}\right) + \pi < \theta \leq \arctan\left(\frac{s+a}{s-b}\right) + \frac{3\pi}{2}; \\ \frac{s-b}{\cos\theta}, & \arctan\left(\frac{s+a}{s-b}\right) + \frac{3\pi}{2} < \theta \leq \arctan\left(\frac{s-a}{s-b}\right) + 2\pi. \end{cases}$$

相应的微分方程给出和(7)一样的解, 只是 I 稍有不同. 图 2-5 展示了两种情形.

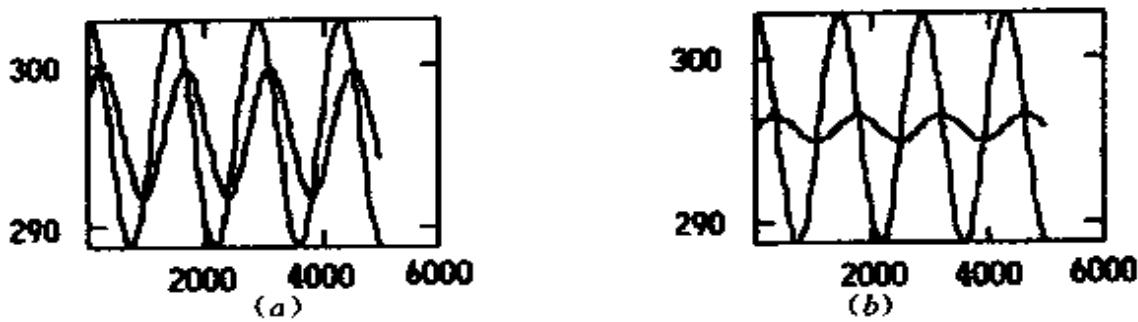


图 2-5 关于时间变化的周围环境温度(变化较大)以及圆形混凝土地板上一点的温度(变化较小)
(a) $u=10$ 而 $a=b=7$. (b) $u=20$ 而 $a=b=8$

我们的建议

具有较小的周长—面积比的混凝土地板的温度很可能保持在舒适温度的范围内。由于矩形板容易制造，又因为周长—面积比最小的矩形是正方形，因此我们建议住房与城市开发部建造正方形板基的房屋。为免去加热和致冷的费用，板基应相对大一点，至少为 $30m \times 30m$ ，这样大小的混凝土地板能使板面平均温度保持在舒适温度的范围内。因为我们用了非常粗糙的数值方法，所以我们的结果也是很粗略的。

假设的实用性和敏感性

改变表示周围温度的函数不会造成很大的误差，混凝土地板的温度仍将在最高和最低温度的范围内并围绕着平均温度震荡。因为我们假定余弦函数的振幅不变，又因为我们考虑的所有函数都只依赖于温度之差而与其绝对值无关。平均温度的最小和最大值关于时间的图形是一样的，只是关于最高和最低温度的算术平均值有一个上下的移动而已。如果混凝土板基很大，就像是很大的公寓单元房的地基板，则温和的季节温度变化(温度变化小于一个数量级)还是可以允许的。但是由于舒适温度的范围只有 10°F ，季节温度变化一定要很温和，否则仅仅由混凝土地板的作用是不能保证屋内温度保持在舒适温度范围内的。

然而，住房与城市开发部肯定不应考虑在周围环境温度随季节的变化会很大的非温和气候地区建造这样的住房。如果在一年中任何时间内平均周围温度的升和降超出舒适温度的范围，那么只靠混凝土地板来加热和致冷以使屋内平均温度保持在舒适温度范围内是不可能的。

本模型对于假定墙壁是不传热的这点来说是很敏感的。如果墙壁是导热的，那么它们的作用和混凝土地板是一样的，从而将改变混凝土和建筑物内空气的温度。然而，尽管有这种敏感性，本假设还是有效的，因为美国的建筑标准中有建筑物墙壁绝热的标准等级，这些材料几乎没有导热性。

另一方面，混凝土地板顶部绝热使得数学处理更容易，但就人们希望利用它来加热混凝土地板上的居室而言，这是很愚蠢的想法。但地下温度不在舒适温度范围内时，使混凝土地板的底部绝热是一个很好的想法，但并不总是现实的。所以，我们将考虑混凝土地板的顶部和底部不一定绝热的情形。

2.3.5 混凝土地板的顶部或底部不绝热的情形

混凝土地板的作用是把热传入室或传送出室外，而绝热会阻止这样做。

当底部不绝热时，如图 2-6 所示，对于小的 h （混凝土地板的厚度），在房屋地板附近温度震荡的平均值将趋于地下的温度。混凝土地板的温度将紧紧围绕地下温度震荡。因此，在地下温度超出舒适温度范围的地区，住房与城市发展部可能希望混凝土地板的底部绝热，既不在顶部也不在底部绝热的混凝土地板的情形没有太大的差别。

图 2-7 表示了在正方形 一半边长范围内最高和最低温度的变化。混凝土地板不必如前面建议的那么大，因为居室中的空气起到了绝热的作用。具体说，混凝土地板必须有一边长为 $20 \pm 3m$ 。（UMAP 编者注：全面的分析和计算略去了。）

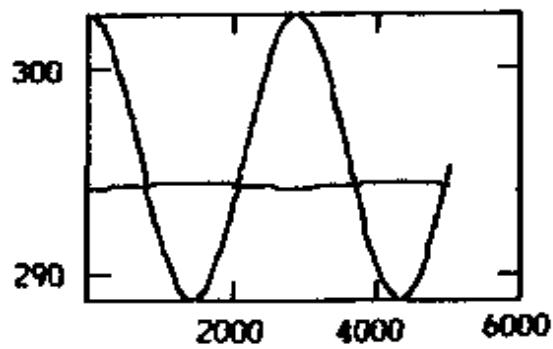


图 2-6 关于时间变化的周围环境温度(变化较大)以及正方形混凝土地板上一点的温度(变化较小), 包括地下温度的效果, $u=5$, $a=b=0$

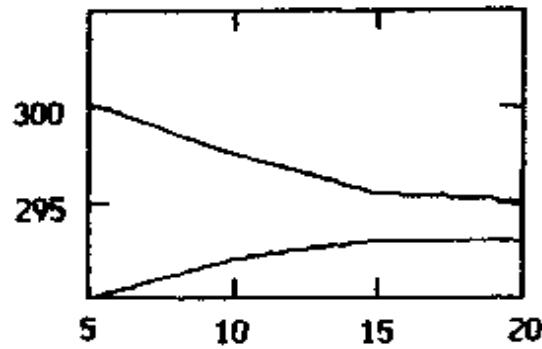


图 2-7 未经绝热的正方形混凝土地板在一半边长范围内(以米为单位)最高和最低温度

2.3.6 模型设计的分析

敏感性分析

当混凝土地板的底部没有绝热时, T_e 在决定混凝土地板和居室温度时起决定作用, 致使所有其他的因素, 甚至周围环境温度的因素, 几乎都可以忽略。而且在周围环境温度范围内, 如果 T_e 在周围环境温度范围内的话, 混凝土地板和室内温度总是绕着接近 T_e 的一个值震荡。所以, 在地下温度于舒适温度范围内的气候条件下, 则无需去做混凝土地板底部的绝热, 从而降低了成本, 同时又提供了很好的解决办法。此外, 在通常建造的地基板下面几英尺深的地方温度不随时间变化。如果季节温度变化很大而地下温度在舒适温度范围内, 这种情形将确保在加热和致冷方面只要花一点点钱或者根本不用花钱。如果地下温度不是很接近或不在舒适温度范围内, 而且周围环境温度又不是那么有运气的话, 因此就无法以这样的方式控制房屋的温度。因此, 我们就要来分析一下有绝热或没有绝热的混凝土地板的成本。

当混凝土地板的底部绝热时, 平均温度仍不特别敏感。但是, 平均温度的小的改变会在混凝土地板的特定点上得到放大。

我们建议，如果住房与城市发展部对所用的混凝土的类型要说什么的话，那就是要采用高强度和高比热的混凝土。从(7)可知，如果这些值大，那么 I 就小，从而温度的震荡就比较小。然而，相当小的能效的增加，尤其是对大建筑物来说，往往并不能说明花费的增加是值得的。

成本分析

混凝土各组成部分在 1994 年 2 月的价格是：

- Portland 水泥 \$ 63.55/吨；
- 砂子 \$ 7.11/吨；
- 碎石 \$ 6.25/吨。

因为构成房基的混凝土通常是按大致 1 : 8 : 8 的比例混合而成的，这种混凝土的实际价格为 \$ 7/吨 或 \$ 22/立方米。绝热的成本是每 16 平方英尺 \$ 2.11，或 \$ 1.42/平方米。这种混凝土一般用的厚度为 0.1~0.15 米。如果我们大量生产 0.15 米的混凝土地板的话，这种房基板(绝热加混凝土)的价格约为 \$ 2.50/平方米，但是不绝热的房基板的价格为 \$ 1/平方米。对于 500 平方米的建筑物来说，绝热要花掉 \$ 710 多。若地下的温度超出了舒适温度范围，而且是要居住多年的房屋是值得绝热的。如果地下温度不在舒适温度范围内，则混凝土地板的底部需要绝热；而且我们推荐采用面积不小于 400 平方米的正方形混凝土地板。如果地下温度在舒适温度范围内，那么我们推荐采用未经绝热过的大正方形混凝土地板。

实际上，住房的地板应铺上瓷砖使之与空气间的传热更有效，因为地毯和地板的导热性并不好。

2.3.7 讨论

模型的测试

为测试本模型，住房与城市发展部应建造一幢以混凝土地板

为基础的样板楼，并用温度传感器来验证本模型关于平均温度随时间变化的预测。

Felix Trombe 博士曾在法国南部建造了一所住房，其南墙是混凝土墙。该墙的作用类似于我们的问题中的混凝土地板的作用，但我们的问题中还讨论季节温度的变化。因为该墙是直立的，它接受的太阳辐射热流随季节和太阳辐射的角度的变化而变化以调节房屋的温度。从地基板下的地下传热的房屋也曾建造过。根据我们的模型，以及他们的观察，他们的加热和致冷系统应该是很合算的。

模型的改进

我们模型的一个缺点就是本模型严重依赖于假定混凝土地板是由无穷多根导热杆组成的，杆之间互不影响。对于精确求解这一假设可能并不成立。

本模型的另一个缺点就是除了盖一幢样板楼外无法进行测试。

2.3.8 参考文献

Giancoli, Douglas C. 1991. *Physics: Principles with Applications*. Englewood Cliffs, NJ: Prentice-Hall.

Kong, F. K., et. al. 1983. *Handbook of Structured Concrete*. New York: McGraw-Hill.

Materials prices. 1994. ENR (7 February 1994).

Neville, A. M. *Properties of Concrete*. New York: John Wiley and Sons, 1973.

Press, William H., Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. 1990. *Numerical Recipes in C: The Art of Scientific Computing*. New York: Cambridge University Press.

§ 2.4 实践者的评述

2.4.1 作者介绍

本题优秀论文评述的作者 Dave Dobson 在 Berkeley 加州大学获化学学士和物理学博士学位。在美国 Lawrence Livermore 实验室研究核武器扩散四年之后于 1968 年到 Beloit 大学任教，在 70 年代早期他对一般的能源，特别是对太阳能产生兴趣。1983 年他设计了一幢房子，利用了太阳能和超级绝热技术，并于 1984 年在他的妻子和几个朋友的帮助下开始建房。自 1987 年起这幢房子就成了他的住所和实验室。

2.4.2 评述文章的主要内容

Dobson 的评述文章包括以下内容：

引言

作为要对之进行建模的物理系统，混凝土地板问题是一个好的且有适当挑战性的问题。但是建议把该系统作为房屋设计的一种可信的方法，在我看来是完全不切实际和不能令人信服的。所以我要把我的评述分为三类：

- 对 North Carolina School of Science and Mathematics(北卡科学与数学学校)队论文的解法的简短的批评；
- 给出所讨论问题的理想化问题的一个解；
- 对任何房屋设计的与这种模型相关的问题的评述。

批评

北卡队论文的作者所用的“杆模型”的主要问题是绕一点把混凝土地板分割成的馅饼形状的楔子不是均匀的等截面杆。

另一个问题是作者在(2)中用到的 dQ 是通过一点的热量而不是(3)中用到的储存在一点的热量 dQ ——它们不是相同的概

念！

尽管有这些批评，但考虑到他们在该问题上工作的严格的时间限制，对于该队的很好的尝试还是值得祝贺的。

理想化的问题

厚度和成分都是均匀的混凝土地板的顶部和底部都绝热，而且其周边的温度总是和被规定为时间的函数的外界温度一样。一般说，这是一个二维的热的流动问题。

(作者推导出一维、二维的热传导方程，见〔2〕pp. 220～221。我们略去这部分内容，读者可参看任何一本偏微分方程的教材，也可参看〔3〕。) 一维、二维的热传导方程分别为

$$\frac{\partial^2 T}{\partial x^2} = \frac{\rho c}{\kappa} \cdot \frac{\partial T}{\partial t}, \quad \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = \frac{\rho c}{\kappa} \cdot \frac{\partial T}{\partial t},$$

其中 $T = T(x, t)$ 或 $T = T(x, y, t)$ 。

求理想化的问题的解

挑战在于要求得到在混凝土板周边上满足规定的周围环境温度的边界条件的热传导方程的解。

我们再次来考察一维的情形（我们把它称为长条形房子！）对此，我们有

$$\frac{\partial T}{\partial t} = \lambda \frac{\partial^2 T}{\partial x^2}, \quad \lambda = \frac{\kappa}{\rho c}.$$

考察试探解

$$T(x, t) = T_0 + T_1 \cos(\omega_0 t - \beta[L-x]) e^{-\gamma(L-x)}, \quad (0 \leq x \leq L),$$

其中 β 和 γ 是代入方程后再来确定的量。在一维的混凝土板的面 ($X=L$) 上满足

$$T(L, t) = T_0 + T_1 \cos \omega_0 t,$$

这正是问题所要求的！

对试探解求微商，得到

$$\begin{aligned}\frac{\partial T}{\partial t} &= -T_1 \omega_0 \sin(\omega_0 t - \beta[L-x]) e^{-\gamma(L-x)}, \\ \frac{\partial T}{\partial x} &= -T_1 \beta \sin(\omega_0 t - \beta[L-x]) e^{-\gamma(L-x)} \\ &\quad + T_1 \gamma \cos(\omega_0 t - \beta[L-x]) e^{-\gamma(L-x)}, \\ \frac{\partial^2 T}{\partial x^2} &= -T_1 \beta^2 \cos(\omega_0 t - \beta[L-x]) e^{-\gamma(L-x)} \\ &\quad - 2T_1 \beta \gamma \sin(\omega_0 t - \beta[L-x]) e^{-\gamma(L-x)} \\ &\quad + T_1 \gamma^2 \cos(\omega_0 t - \beta[L-x]) e^{-\gamma(L-x)}.\end{aligned}$$

仅当代入热传导方程后得到的正弦和余弦项前面的系数分别等于零时， $T(x, t)$ 才能满足方程，即，若 $\beta = \gamma = \sqrt{\frac{\omega_0}{2\lambda}} = \sqrt{\frac{\omega_0 \rho c}{2\kappa}}$ ，则 $T(x, t)$ 既满足方程又在 $x=L$ 处满足边界条件。

因此，一维长条形房子的混凝土地板的温度可以表示为

$$T(x, t) = T_0 + T_1 \cos\left(\omega_0 t - \left[\frac{L-|x|}{L_c}\right]\right) e^{-\frac{|x|}{L_c}},$$

其中

T_0 是周围(板的周边)环境温度的平均；

T_1 是周围环境温度变化的振幅；

ω_0 是周围环境温度变化的角频率；

L 是长条形房子(它从 $x=-L$ 延伸到 $x=L$)的半宽；

L_c 是表征温度变化随离开板边时的减弱以及温度变化随离开板边时的相移的特征距离：

$$L_c = \sqrt{\frac{2\kappa}{\omega_0 \rho c}}.$$

对于全天(24 小时)中的周围环境温度的变化以及混凝土的典型特征而言， $L \approx 7\text{m}$ 。

要求二维的混凝土地板的精确解就更困难了，但是对于任何形状的板都可以近似求解。特别是，如果比之于特征距离 L_c 要

大的板来说是可以近似求解的。靠近板的周边的地方(比如说，在20米内)可以利用一维问题的解。远离周边的地方，温度变化会相当小，也可能可以忽略。

(注：这里有误。因为 $T(x,t)$ 在 $x=0$ 处不是二次连续可微的，因而不满足方程。正确的解可参看[3]。)

房屋设计

没有一个精神正常的人会用这种方法来设计房屋！就能效设计而言混凝土板是很好的，但如果周围环境温度超出了舒适温度范围，那么总是要把板的周边绝热！(周长×板的厚度) 面积不大，绝热是不太费钱的。当周边没有绝热时，靠近板的外周边的地方的温度总是接近于周围环境温度；从计算中我们看到在离周边几米远的地方仍保持这种状况。

建筑物下的地温接近于全年的年平均周围环境温度，所以，如果这个温度在舒适温度范围内的话，那么，除在靠近周边的地方外，板的底部不应该绝热，这样，你们就能从土地作为热源或热流的有限效应中获得最大的好处。

北卡科学与数学学校队指出对混凝土地板的顶部绝热并不合理。这是正确的。我相信命题人心中想的是房屋的墙壁、屋顶的绝热能使混凝土地板与周围情况隔热。地板向住宅房间以及房间表面向地板的热辐射，将把房间内部以及混凝土地板本身一起导致靠近热平衡的状态。当然，板中的温度变化将使这种情形的分析复杂化，但是我想这可能就是命题人要建议求“板表面平均”温度的原因吧。但是，对于实际建筑物来说，这种平均并不蕴涵着外墙附近处于舒适温度的条件。

屋内空气的热容量被完全忽略了

在一幢绝热很好的房子里，对流也是不重要的。存在着某种分层(屋顶附近热一点，地板附近凉点)，但是，如果房屋是绝热良好，对空气渗漏来说是密闭良好的，而且有高效(隔热玻璃或更好的材料做成的)窗户的话，辐射热的传输效应不大。

在确定房屋内部的热容量时，应把石膏灰胶纸夹板、内墙、房屋的固定装置、家具等的热容量加进混凝土地板的热容量。作为估算，可假定所有这些东西都有和混凝土一样的热容量，于是，所有这些就是包含在房屋内要绝热的所有东西总质量。

在问题陈述中描述的非常温和的气候条件下，可能无需加热或致冷就很容易保持在舒适温度范围内。但是，把湿度控制在舒适范围内仍是一个挑战。为防止内部产生毒气、臭气、人呼出和洗澡产生的湿气等，每天要有新鲜空气进入室内是很重要的。通常在温和气候的条件下在最佳时间（白天或夜晚，有赖于季节）开开窗就够了。但是，如果外边的空气极其潮湿、冷凝、发霉的话，即使温度是合适的，也会使人感到不舒服。这时，就可能需要空调或太阳能干燥器。

每种气候都对注重能效的建筑提出了自己独特的挑战。考虑周到并创造性地应用可利用的一切手段，包括混凝土地板、绝热技术、窗户位置、建筑物朝向、吊顶等等会给出实际而且优美的解决。

让我们来建造城镇住房

一个传统的“美国梦”就是有一幢（在郊区某处的）供单个家庭住的房子。城市规划的分区制常常是从地产边缘后撤 10 英尺，还要保证住宅间 20 英尺的“浪费掉的（不能作它用的）空间”。

排屋（比较时髦的话叫做城镇住房）的建造把浪费掉的空间减少为零，致使土地的利用效率增加 30%~50%，通过公共墙壁没有热损失，从而对于这种墙壁（除隔音外）无需特别的绝热。现代的防火材料大大提高了防火安全性，而这正是所有木头房屋的主要缺点。

东西方向的排屋没有东、西朝向的窗户（使房屋过分加热的主要热源）。排屋总是要东西向，窗户则朝北或朝南，永远不是其他方向。南窗的吊顶（在热带地区南北窗户都要吊顶）把不想要的低角度的太阳光线挡在外。

§ 2.5 评注

引言

正如实践者的评论指出的，这道题出得不太好，但是作为训练数学建模的一个物理模型对学生还是有一定的挑战的。

对于学过偏微分方程（或数学物理方程）的学生可能很快会想到这至少是一个在任意有界平面区域上，给定已知的周期变化的周围环境温度，求热传导方程的周期解，并回答什么样形状的混凝土地板能使板面（或室内）的平均温度能在舒适温度范围内。这样一来，数学问题是明确的，但是真正的解答却是不容易的。即使是对实践者评述中所说的长条形房屋来说也不能说是平凡的。我们估计多数参赛学生并不具备足够的偏微分方程知识，但是他们却能从分析问题入手，抓住主要矛盾作出很有创意的工作。北卡科学与数学学校队，还有其他队的论文，就是这样的富有一定创意的论文。

北卡科学与数学学校队论文中怎样从常微分方程(6)得到解(7)

由于(8)式，记 $a = 6.94 \times I$, $b = 295.7 \times I$, $\omega = 2\pi/1440$, $IT_a = a \cos \omega t + b$, 则常微分方程(6)可写为 $\frac{dT}{dt} = a \cos \omega t + b - IT$, 两边同乘 $e^{\int dt}$, 得 $\frac{d}{dt}(Te^{\int dt}) = e^{\int dt}(a \cos \omega t + b)$, 从 0 到 t 积分，得 $T(t) = [T(0) - \frac{aI}{\omega^2} - b]e^{-at} - \frac{\omega a}{\omega^2 + I^2} \sin \omega t + \frac{Ia}{\omega^2 + I^2} \cos \omega t$, 若令初始值 $T(0) = \frac{aI}{\omega^2} + b$, 就得到(7)式（注意(7)中丢了一个“-”号）。

值得阅读的其他优秀论文——获一等奖的北京大学物理系队的论文（见[4]）

该文叙述明确、有条理。

该文的第一个模型只考虑热辐射。他们指出(或者说是假设),一般建筑地板平面尺度远大于地板厚度,“尽管混凝土地板对周围并不是绝热的,它们对于热量传输的影响也小到可以忽略的程度,所以我们认为形状不重要…”他们利用物理规律对板面的平均温度导出了一个非线性微分方程,并数值求解之,发现一些规律性的知识,得到了结论。

该文的第二个模型研究了室内温度能否在舒适温度范围内的问题。他们考虑了一个长方体房间,并认识到其困难,从而只讨论一个最简单的问题——双容薄壁系统,也发现了一些规律性的知识,得到了结论。

该文的第三个模型利用了热传导方程的一个物理上合理的解,对把住房建造在地下提出了建议。

最后,他们对三个模型作出了评价。

参 考 文 献

- (1) C. Ahn, K. Paur, E. Tytell, Innovative heating technology, UMAP, v. 15 (1993), no. 3, 207~217.
- (2) David A. Dobson, Practitioner's commentary: The outstanding concrete slab paper, UMAP, v. 15(1994), no. 3, 219~224.
- (3) 叶其孝, 差分、微分方程建模,《大学生数学建模竞赛辅导教材》(二)(叶其孝主编),湖南教育出版社,1997, 231~247.
- (4) 江洪春、庄磊、陈纲(指导教师:王兰革),无需冷气、暖气的舒适住宅——A题,数学的实践与认识,1994第4期,78~83.

第三章 通讯网络问题

韩继业

中国科学院数学与系统科学研究院 应用数学研究所

提 要

本章介绍 1994 年美国大学生数学建模竞赛 B 题的背景和有关情况，部分优秀答案，一些评注和参考文献等。

§ 3.1 问题

本章讨论的 1994 年美国大学生数学建模竞赛 B 题的内容如下：

在你们的公司里，各部门每天都要分享信息。这种信息包括前一天的销售统计和当前的生产指南。尽快传出这些信息是十分重要的。

假设一个通讯网络被用来从一台计算机向另一台计算机传输数据组（文件）。作为例子，考虑下面的图模型：

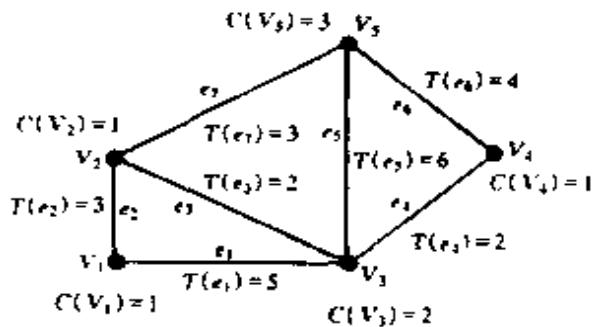


图 3-1 文件传输网络的例子

顶点 V_1, V_2, \dots, V_m 表示计算机，边 e_1, e_2, \dots, e_n 表示（由边的端点表示的计算机之间）要传输的文件。 $T(e_x)$ 表示传输文件 e_x 所需的时间， $C(V_y)$ 表示计算机 V_y 可同时传输多少个文件的容量。文件的传输必须占用两个有关计算机为传输该文件所需的全部时间。 $C(V_y)=1$ 表示计算机 V_y 一次只能传输一个文件。

我们有兴趣的是以最优的方式来安排传输，使得传输完所有的文件所用的总时间最少。这个总时间被称为完工时间（makespan）。请为你们的公司考虑以下三种情况：

情况 A

你们公司有 28 个部门，每个部门有一台计算机，每台计算机被表示为图 3-2 中的一个顶点。每天必须传输 27 个文件，在图 3-2 中用边来表示文件传输。对于这个网络，对所有的 x 和 y ，有 $T(e_x)=1, C(V_y)=1$ 。试找出该网络的一个最优时间表（进度表，schedule）和相应的完工时间。你们能向你们的主管人员证明你们对该网络求得的完工时间是最小可能（最优）的吗？叙述你们求解该问题的方法。你们的方法适用于一般情形吗，即是否适用于 $T(e_x), C(V_y)$ 以及图的结构都是任意的情形？

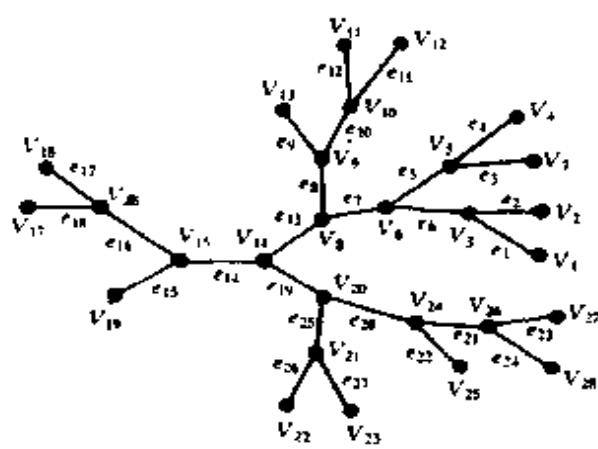


图 3-2 情况 A 和 B 的网络

情况 B

假设你们的公司改变了传输要求。现在你必须在相同的基本

网络结构（见图 3-2）上考虑不同类型和数量的文件。传输这些文件所需的时间由表 3-1 中每条边的 $T(e_x)$ 项表示。对全部 y 仍有 $C(V_y) = 1$ 。试对新网络找出一个最优时间表和其完工时间。你们能证明对新网络而言所求得的完工时间是最小可能的吗？叙述你们求解该问题的方法。你们的方法适用于一般情形吗？试对任何特异的或出乎意料的结果发表评论。

表 3-1 情况 B 的文件传输时间数据

x	1	2	3	4	5	6	7	8	9
$T(e_x)$	3.0	4.1	4.0	7.0	1.0	8.0	3.2	2.4	5.0
x	10	11	12	13	14	15	16	17	18
$T(e_x)$	8.0	1.0	4.4	9.0	3.2	2.1	8.0	3.6	4.5
x	19	20	21	22	23	24	25	26	27
$T(e_x)$	7.0	7.0	9.0	4.2	4.4	5.0	7.0	9.0	1.2

情况 C

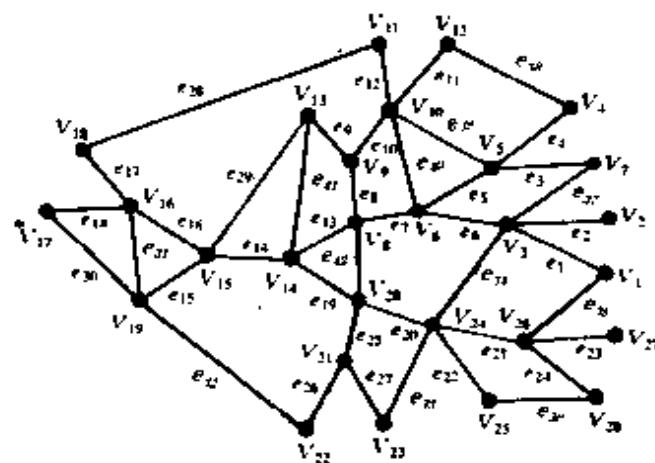


图 3-3 情况 C 的网络

你们公司正在考虑扩展业务。如果公司这样做的话，每天有一些新文件（边）要传输。这种业务扩展还包括计算机系统的升级换代。28个部门中的某些部门将配备新的计算机使之每次能传输不止一个文件。所有这些变化都在图 3-3 及表 3-2、表 3-3 中说明。你们能找到的最优时间表和完工时间是什么？你们能证明对

该网络而言你们的完工时间是最小可能的吗？叙述你们求解该问题的方法。试对任何特异的或者出乎意料的结果发表评论。

表 3-2 情况 C 的文件传输时间数据

x	1	2	3	4	5	6	7	8	9
$T(e_i)$	3.0	4.1	4.0	7.0	1.0	8.0	3.2	2.4	5.0
x	10	11	12	13	14	15	16	17	18
$T(e_i)$	8.0	1.0	4.4	9.0	3.2	2.1	8.0	3.6	4.5
x	19	20	21	22	23	24	25	26	27
$T(e_i)$	7.0	7.0	9.0	4.2	4.4	5.0	7.0	9.0	1.2
x	28	29	30	31	32	33	34	35	36
$T(e_i)$	6.0	1.1	5.2	4.1	4.0	7.0	2.4	9.0	3.7
x	37	38	39	40	41	42			
$T(e_i)$	6.3	6.6	5.1	7.1	3.0	6.1			

表 3-3 情况 C 的计算机容量数据

y	1	2	3	4	5	6	7	8	9
$C(V_y)$	2	2	1	1	1	1	1	1	2
y	10	11	12	13	14	15	16	17	18
$C(V_y)$	3	1	1	1	2	1	2	1	1
y	19	20	21	22	23	24	25	26	27
$C(V_y)$	1	1	1	2	1	1	1	2	1
y	28								
$C(V_y)$	1								

§ 3.2 问题的背景

本章的通讯网络问题是在受传输时间和计算机容量等因素的限制下，研究在一网络中如何安排一些文件的传输，使得完成全

部文件传输的工作时间为最短。这种问题是“时间表问题”(scheduling problems, 或称为“调度问题”)中的一类问题。时间表问题属于离散最优化这一学科。它研究在某些因素(如物资、人力、财力等)的一定限制下, 寻求若干工作相对于某一目标而言的最优安排。“完工时间”(makespan)即是常用的一种目标。时间表问题的研究在国际上仅有四十多年, 由于它在国民经济和军事方面具有广泛的应用, 目前它已是运筹学、管理科学、系统工程和计算机科学等领域中的重要研究课题, 有大量的文献资料。

时间表问题按照其结构的不同而分成许多不同的类型, 不同类型的问题在性质与算法上有很大的差别。从复杂性而言, 70%以上的时间表问题属于NP-困难问题。因此, 大多数时间表问题当其规模相当大时, 最优解的寻求是困难的。对于一个属于NP-困难的规模较大的时间表问题, 为了很快地计算出最优解, 文献中常使用一些多项式时间的近似方法, 且要求近似方法求得的近似解能比较好地接近问题的最优解。为了合理地衡量这种接近程度, 常用的衡量指标是“最坏情况下的性能比”。本章的通讯网络问题的三种情况(情况A, 情况B和情况C)都还是规模较小的问题。与其他的离散最优化问题一样, 规模较小的时间表问题可以使用“穷举方法”(如分支定界法)来求出其最优解。针对同一问题可以设计不同的穷举方法, 较简便的穷举法比烦琐的穷举法可显著地提高计算效率。但NP-困难问题的任何穷举方法只能是指数性质的, 它的计算时间随着问题的规模的增大而呈现指数的增长。由于实际问题的规模都比较大, 这样就大大地促进了多项式近似方法在近十年内的迅速发展。

§ 3.3 优秀竞赛论文

在参赛的315个队中有245个队选择“通讯网络”这一赛

题，结果有 5 个队获得了特等奖，39 个队获得了一等奖，59 个队获得了二等奖。获得特等奖的 5 个队是来自：美国的 Beloit 学院，Grinnell 学院，北卡罗莱纳大学，加拿大的 Calgary 大学，多伦多大学等。我国有 12 个队获得了一等奖，他们分别属于：复旦大学、哈尔滨船舶工程学院、河海大学、吉林大学、南京航空航天大学、上海师范大学、东南大学、清华大学、中国科技大学（两个队）和西安电子科技大学（两个队）。

本节介绍多伦多大学工业工程系一个参赛队的获特等奖的优秀论文，题目是：

“计算机网络中的同步文件传输”

该文利用图论的一些结果和最大匹配设计了一类方法，并且证明了对于情况 A，B 和 C，所计算出的结果都是最优的。对于任意图的结构和计算机容量，所提出的方法的结果也是最优的。

3.3.1 基本概念和基本假设

1. 图被表示为 $G=(V,E)$ ，其中 V 是“顶点”集， E 是“边”集。这里只讨论无向图，即图中任何一个边 (x,y) 都没有方向， $x, y \in V$ 。如果 E 中任何一边 (x,y) 都有 $x \neq y$ ，则图 G 被称为“简单图”。通讯网络中的图即是简单图。

2. 顶点 x 的“次”（或“度”）是指与 x 相关联的边的个数。
3. 图 G 是“连通”的，如果对每一对顶点 $x, y \in V$, $x \neq y$ ，都存在一条边 $e_1, e_2, \dots, e_n \in E$ ，使得 $e_1 = (x_0, x_1), e_2 = (x_1, x_2), \dots, e_n = (x_{n-1}, x_n)$ ，且 $x_0 = x, x_n = y$ 。

4. 如果 $V' \subseteq V, E' \subseteq E$ ，且由 $(x,y) \in E'$ 可知 $x, y \in V'$ ，则称 $G' = (V', E')$ 是 G 的一个“子图”。

5. G 的一条“路”是 G 的一个连通子图，而且子图中的两个顶点的次为 1，其他的顶点的次为 2。

6. G 的一个连通子图被称为“圈”，如果子图中的顶点的次都是 2。

7. 图 $G=(V, E)$ 被称为一个“二分图”，如果 $V=X \cup Y$, X 与 Y 是不相交的顶点集，且 E 中任何一边 (u, w) 都有 $u \in X$, $w \in Y$, 或者 $u \in Y$, $w \in X$. 二分图也表示成 $G=(X, Y, E)$.

8. 图 G 被称为“树”，如果它是连通的没有圈的简单图. 树都是二分图.

9. E 中子集 M 被称为 G 的一个“匹配”，如果 M 中的边的端点都互不相同. 显然 G 中任一顶点至多与匹配 M 中一条边关联.

10. 如果顶点 V_s 与匹配 M 中一条边相关联，则 V_s 被称为“ M 饱和顶点”；反之， V_s 被称为 M 不饱和顶点.

11. 匹配 M 被称为“最大基数匹配”，如果 G 中没有另一匹配 M' ，使得 $|M'| > |M|$.

12. M 被称为“完美匹配”，如果 G 中任一顶点都是 M 饱和顶点. 显然完美匹配必是一最大基数匹配.

13. 路 P 是匹配 M 的一条“交错路”，如果路 P 中的边交错地一条边属于 M ，下一条边不属于 M .

14. 匹配 M 的交错路 P 被称为 M 的“增广路”，如果交错路 P 的首尾两个端点都是 M 的非饱和顶点. 显然增广路 P 的边的个数是奇数，其中不属于 M 的边数比属于 M 的边数多一个.

图论中关于最大基数匹配的一个重要定理是：

匹配 M 是图 G 的最大基数匹配，当且仅当 G 中没有 M 的增广路(参见参考文献[2]，定理 4.1).

这个定理实际上给出了求最大基数匹配的算法的基本思想：如果已有一个匹配，在图 G 中寻找它的增广路，如果可以找到一增广路，则可以得到一个增广的匹配；否则匹配已是最大基数匹配. 根据这一基本思想，出现了寻求最大基数匹配的一些重要算法，与本问题有关的是下面的几个结果：

1. 二分图的最大基数匹配的算法（算法的详细叙述可参见参考文献[2] § 4.2，或参考文献[3]第 7 章 § 2）.

2. 一般图的最大基数匹配的算法（算法的详细叙述可参见参考文献[2] § 4.3）。

3. 网络上的最大权匹配。在情况 B 和 C 中每一边（文件传输）的传输时间 $T(e_x)$ 可被视为该边的“权”。对于情况 B 是寻求二分图中的最大权匹配，即寻求一匹配，它的所有边的权数的总和为最大（算法的详细叙述可参见参考文献[3] 第 7 章 § 3）。

基本假设是：

1. 通讯网络问题的图是无向的简单图，且任意两个顶点间没有重边。这意味着两台计算机之间传输的全部讯息可包括在一个文件中。

2. 所有的文件都为传输作好准备，也即文件的“准备时间”都为零，由此可知通讯网络在文件传输过程中没有空闲时间，它们可以连续地进行。

3. 每一文件的传输时间都已确定，它们是固定的数。时间的单位不妨约定是分钟。

4. 所有文件都是相互独立的，它们之间没有先后顺序的规定，不存在优先传输的文件。

5. 任一文件在它的传输时间内不能被中断。一个文件一经开始传输，就要连续地被传输完。

6. 通讯网络中计算机在文件传输过程中都不会出现故障。

3.3.2 关于建模的分析

对于情况 A，由于假定了每一文件的传输时间 $T(e_x) = 1$ 和每一计算机的容量 $C(V_y) = 1$ ，所以问题较简单。显然情况 A 中的图是二分图（见图 3-2）。我们观察情况 A 的任何一个可行的文件传输方案（不一定是最优方案），可发现在传输过程中每一分钟内有一些不同的文件在传输，它们对应的边集是情况 A 的图的一个匹配。不同的匹配不含重复的边，它们的并即构成图 3-2。所以情况 A 的最优传输时间表就是将图 3-2 分解为最少个不相重

的匹配，不相重的匹配的最少个数即是最优完工时间(makespan).

对于情况 B 和 C, 由于传输时间 $T(e_s)$ 不全是 1, 情况 C 的计算机容量也不全为 1, 使问题复杂一些. 情况 B 的图是二分图, 情况 C 的图(图 3-3)已不是二分图.

在现有的时间表理论(调度理论)中没有合适的算法可以直接用来求解文件传输问题, 也没有关于这一问题的复杂性的结果. 我们可以把通讯网络问题归结为一类非同型的非抢占的时间表问题(nonidentical nonpreemptive scheduling problem), 并且利用 LPT 算法(加工时间最长的工作首先被处理)的基本思想. 文件传输时间相当于“加工时间”, 待传输的文件相当于“工作”. 在情况 B 和 C 中文件传输时间不相同, 传输时间比较长的文件需要尽早被传输出去. 最优传输方案必然是在文件传输过程的每一时刻被传输的文件个数尽量多. 故此我们发展一种动态的最大基数匹配算法来求解通讯网络问题, 称这种方法为“局部调整方法”, 它不断地对尚未传输的文件构成的子网络使用最大基数算法, 且尽可能先挑选传输时间长的文件.

3.3.3 关于情况 A 的方法

文件的传输时间都是 1 分钟, 每一分钟内文件传输必同时开始、同时结束. 我们基于二分图的最大基数匹配算法设计出如下方法:

方法 A

第一步 令 $E(u)$ 是 G 的所有边的集合. 置

STAGE=1.

第二步 寻求 $E(u)$ 中一个最大基数匹配. 匹配中的边被称为已标号的.

第三步 从 $E(u)$ 中删除已被标号的边.

第四步 如 $E(u)$ 不为空集, 到第五步; 否则, 计算终止, STAGE

即是传输完所有文件的时间。

第五步 令 STAGE=STAGE+1，返回第二步。

将情况 A 的图表示成二分图的形式，可以得到图 3-4。在图 3-5 中用黑线代表一个匹配，由于在图 3-4 中已没有此匹配的增广路，所以它是阶段 1 (stag 1) 的最大基数匹配。图 3-6 指出了在原来的图 3-2 中这一匹配的边。

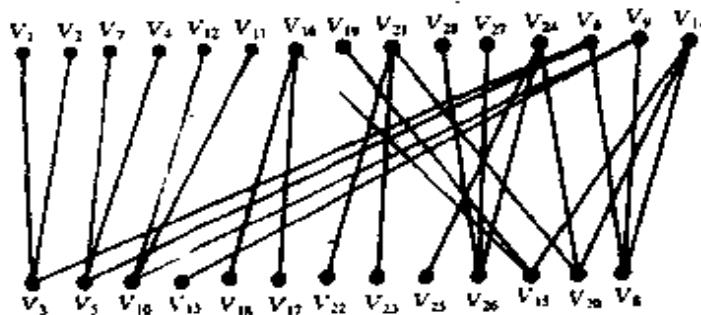


图 3-4 情况 A 的网络的二分图表示

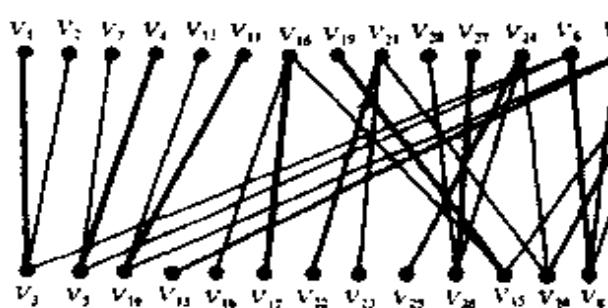


图 3-5 情况 A 的一个最大基数匹配

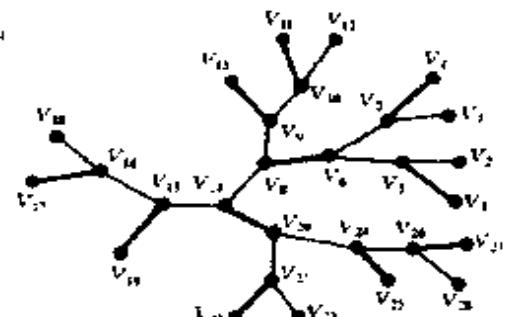


图 3-6 情况 A 的原来网络中的最大基数匹配

3.3.4 关于情况 B 的方法

情况 B 的图仍为二分图。因文件传输时间不全为 1，将文件传输时间作为对应边的“权”，我们可以利用二分图的最大权匹配算法来设计情况 B 的方法。下面给出两种方法：

方法 B1

第一步 利用二分图的最大基数匹配算法求出初始的文件传输集（边集） M ：令 $E(p)$ 表示已传输完的文件集，置 $E(p)$:

$= \varphi$.

- 第二步 在匹配中文件传输的某一时刻已有文件被传输完毕，将已传输完的文件并入 $E(p)$. 令 $M' = M \setminus E(p)$.
- 第三步 如 $E(p)$ 已包含情况 B 的图中所有边，计算终止.
- 第四步 对子网络 $(V, E \setminus E(p))$ 求最大基数匹配，并使匹配包括 M' . 返回第二步.

在方法 B1 中每一次对二分图寻求最大基数匹配时都需要利用局部调整方法的思想，尽可能使匹配包括传输时间比较长的文件（边）。匹配中的文件就是要传输的文件。

方法 B2

将方法 B1 中每次求最大基数匹配换成求最大权匹配，计算步骤不变，就构成方法 B2.

3.3.5 关于情况 C 的方法

由于情况 C 的计算机容量不全为 1，所以寻求最大基数匹配或最大权匹配对此情况已无意义，它们不能直接被利用。解决情况 C 的一种途径是引进“伪顶点”的概念。

如果一个顶点的容量为 3，则在情况 C 的网络中引进两个伪顶点，并设原顶点与两个伪顶点的容量都为 1. 这样，我们可以

在变化了的通讯网络中使用最大基数匹配算法或最大权匹配算法。与伪顶点关联的边等价于原顶点。在图 3-7 中举例说明了这一概念： $C(V_1) = 3$, $C(V_2) = C(V_3) = C(V_4) = C(V_5) = 1$ ，实线表示原图中的边，虚线表示由伪顶点 V_1^* 和 V_2^* 产生的边， (V_1, V_2) , (V_1^*, V_2) 与 (V_1^*, v_2) 被称为“等价边”，它们表示同一个文件传输。

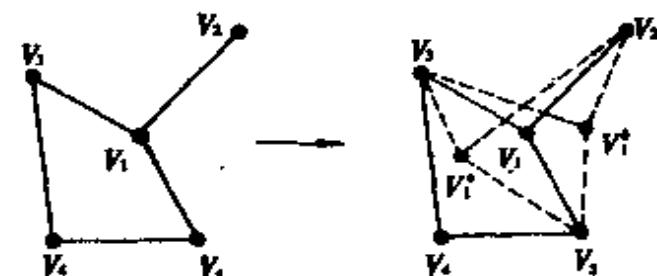


图 3-7 伪顶点变换与等价边

由于伪顶点的出现及等价边的影响，我们虽可对经过伪顶点变换所得的图 (V', E') （每个顶点的容量为1）使用最大基数匹配算法或最大权匹配算法，但在匹配中只能包括等价边中的一条边。这使得对情况C的方法做适当调整：当等价边中的一边已选入匹配中，其他的等价边随即从网络中删除。

方法 C1

- 第一步 对情况C的网络进行伪顶点变换。利用一般图的最大基数匹配算法求出初始的文件传输边 M 。令 $E(p)$ 表示已传输完的文件集。置 $E(p) := \varphi$ 。
- 第二步 从 $E' \setminus (E(p) \cup M)$ 中除去匹配内的边的等价边，并并入 $E(p)$ 。
- 第三步 在匹配中文件传输的某一时刻有文件已传输完毕，将已传输完的文件并入 $E(p)$ 。令 $M' = M \setminus E(p)$ 。
- 第四步 如 $E(p)$ 已包含图中所有边，计算终止。
- 第五步 对于图 $(V', E' \setminus E(p))$ 求最大基数匹配 M ，并使匹配包括 M' 。返回第二步。

方法 C2

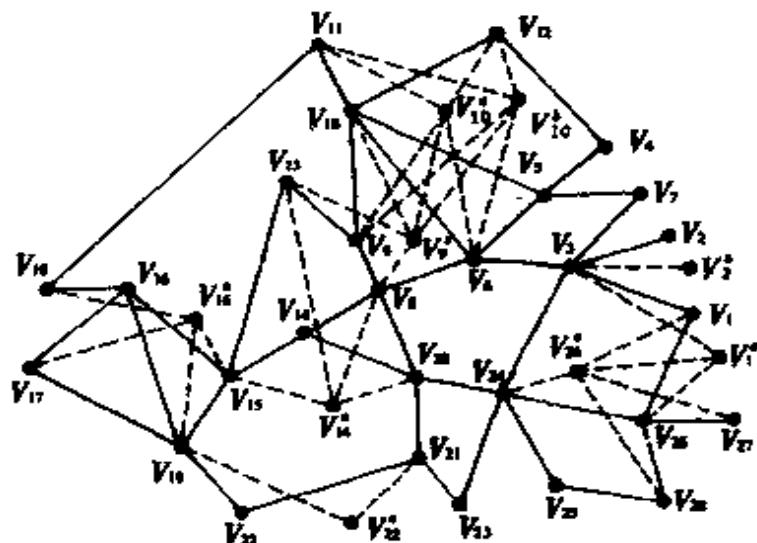


图 3-8 情况 C 的网络的伪顶点变换

将方法 C1 中每次求最大基数匹配换成求最大权匹配，计算
• 56 •

步骤不变，即构成方法 C2.

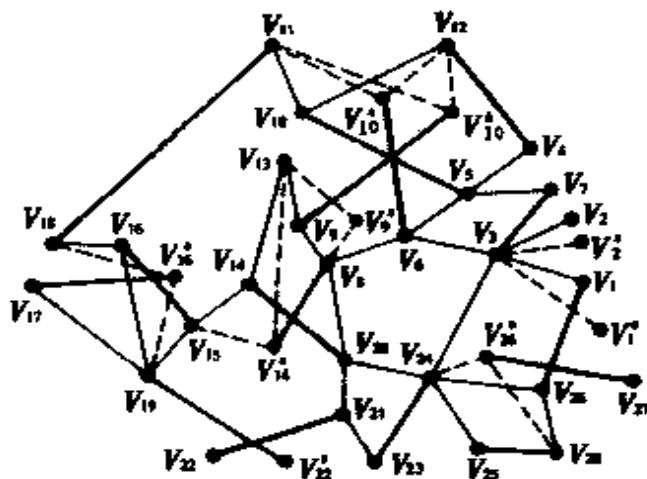


图 3-9 最大基数匹配(粗线)及等价边消去后的图

由于伪顶点及等价边的增加，使原来的网络的规模随着容量的增大而迅速增大，这会增加求解的计算量。数值计算的结果显示，当每次求出最大基数(最大权)匹配后，随着很多等价边从网络中消除，而使网络的规模显著地减小，计算效率仍比较好。图 3-8 和图 3-9 说明情况 C 的网络的伪顶点变换和等价边的消去。

3.3.6 计算结果和最优化证明

情况 A

最小完工时间是 3 分钟。在第一分钟内传输 11 个文件（见图 3-5 中最大基数匹配），第二分钟内传输 9 个文件，第三分钟内传输 7 个文件。这个答案的最优性的证明可以参考图论中关于边色数的重要结果：二分图的边色数等于顶点的最大次。显然情况 A 的图的顶点最大次是 3。我们将第一分钟传输的 11 个文件（边）染一种色，第二分钟传输的 9 个文件染一种色，第三分钟传输的 7 个文件染一种色，这构成图的一个“正常三边染色”（即相邻的边的染色不同）。这证明了计算结果的最优性。

情况 B

利用方法 B1 和方法 B2 所计算的完工时间都是 23 分钟，其中由方法 B1 得到的文件传输时间表见图 3-10。这一结果是最优的，其理由可解释如下：

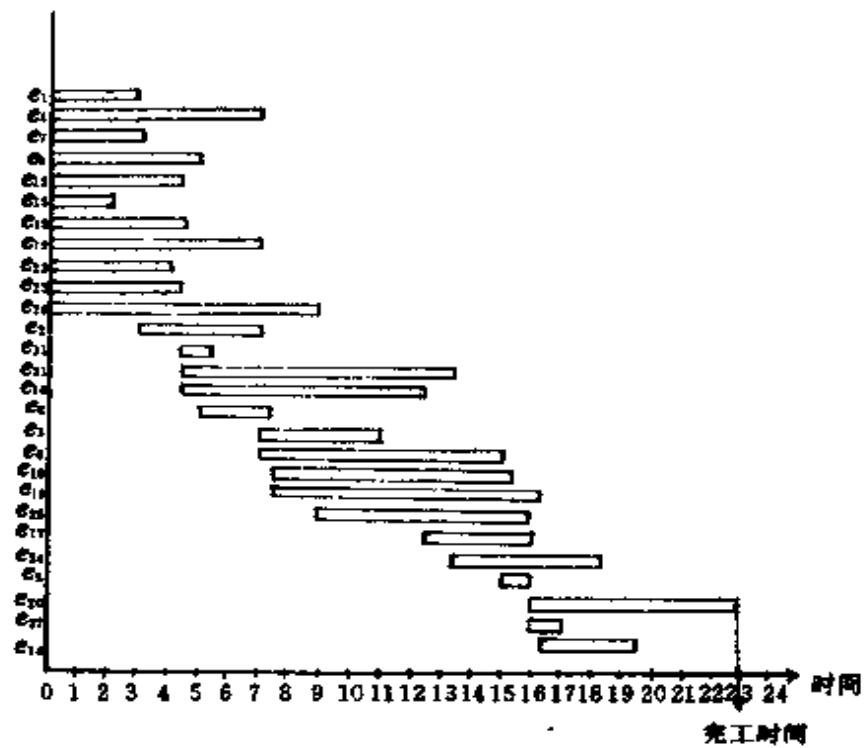


图 3-10 情况 B 的文件传输时间表

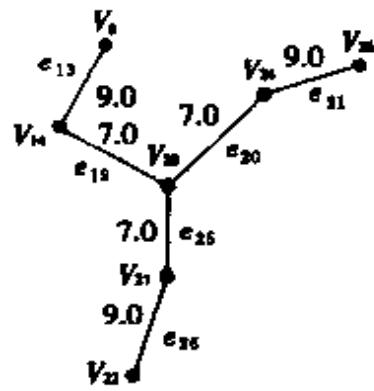


图 3-11 情况 B 的子网络

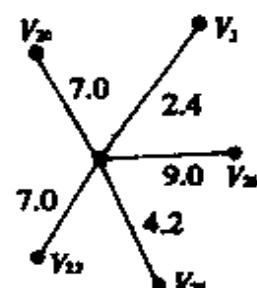


图 3-12 情况 C 的子网络

考虑情况 B 的子网络 $(V_{20}, V_{21}, V_{22}, V_{24}, V_{25}, V_{14}, V_8; e_{20}, e_{21}, e_{25}, e_{26}, e_{19}, e_{13})$ ，此子网络的结构见图 3-11。文件传输时间 $T(e_{20}) = 7.0$, $T(e_{21}) = 9.0$, $T(e_{25}) = 7.0$, $T(e_{26}) = 9.0$, $T(e_{19}) = 7.0$, $T(e_{13}) = 9.0$ 。它的文件传输时间至少是 23 分钟。因情况 B 的讯

息传输的最优完工时间不会少于任一子网络的最优传输时间，所以方法 B1 和方法 B2 求得的完工时间(23 分钟)是最优的.

情况 C

利用方法 C1 和方法 C2 计算的完工时间是 29.6 分钟. 为了证明这一结果是最优的，考虑子网络 $(V_{24}, V_{20}, V_{23}, V_{25}, V_{26}, V_3; e_{20}, e_{33}, e_{22}, e_{21}, e_{34})$ ，文件传输时间 $T(e_{20}) = 7.0$, $T(e_{33}) = 7.0$, $T(e_{22}) = 4.2$, $T(e_{21}) = 9.0$, $T(e_{34}) = 2.4$. 它的结构见图 3-12. 由于子网络的文件传输时间已是 29.6 分钟，所以方法 C1 及方法 C2 计算的结果是最优的.

3.3.7 敏感度分析

我们研究增大顶点的容量对最优完工时间的影响. 显然增大一些顶点的容量有可能缩短完工时间，增大那些传输量很大的顶点的容量，对完工时间的影响更显著. 称一个顶点为“瓶颈顶点”，如果它的容量为 1，且它的总传输时间最长. 对于情况 C，瓶颈顶点是 V_{24} ，它的总文件传输时间是 29.6. 如果令 $C(V_{24}) = 2$ ，利用方法 C1 可得最优完工时间为 27.1.

3.3.8 优点与缺点

对于二分图我们容易求解其最大基数问题. 然而对于具有任意图结构的通讯网络问题，算法必须改进. 方法 C1 与方法 C2 的主要优点是可以应用于任意结构的图. 方法依赖于根据一已有的匹配寻找一条增广路.

问题中给出的三个网络都相当小，用手算也可比较快地解决. 但是我们的方法可以编成计算机程序，用来解决更大规模的网络问题. 最大的麻烦将是方法的第一步，即寻求初始最大基数匹配（或者最大权匹配），以后的计算步骤只需随时注视那些与可行顶点相关联的边，而不必细察整个网络.

方法 C1 和 C2 有一缺点，即增加了伪顶点和等价边。但很多等价边在求出初始最大基数匹配后被删去，其余的等价边也在以后每一步被消去。

3.3.9 讨论

下面对问题在有代表性的计算条件下研究其结果。

1. 非零的准备时间

上面的讨论假设了全部文件在开始时都可被传输。但这是不实际的，因为不可能全部文件都在同一时刻被准备好传输，它们能够被传输的开始时刻是有差别的。我们只需对方法做一修改，即可适用于文件的准备时间不为零的情况：在方法的迭代计算过程中只在已完成准备的文件（边）中寻求最大基数匹配（最大权匹配）。

2. 相关性

上面的讨论假设文件的传输是相互无关的，实际上一些文件的传输有先后顺序的限制。我们对方法稍做修改使之适用于文件相关的情况：在网络中暂时除去需后传输的文件，当需先传输的文件已传输完毕，再将后传输的文件（边）加入网络。

3. 优先权

上面的讨论中未对文件传输的优先权做任何假设。这允许我们可优先挑选传输时间最长的文件。在实际的一些网络中（如 Advanced Peer to Peer Network）文件被安排了传输优先权。我们能调整文件的权，使之符合它的优先权。我们的方法也适应于优先权的情况。

4. 抢占

上面的讨论中任一文件的传输过程没有中断现象，即不允许有抢占。如果我们在方法的迭代计算中利用最大权匹配算法，没有抢占的假设将被违背。在一文件的传输中如发生中断，有两种策略被用来对待该文件的重新传输：一是只需传输尚未传输的讯息；另一是从头开始传输该文件。在实际情况下，具有抢占的文

件传输时间常介于两种策略之间。因此，在具有抢占的情况下，最优完工时间甚至比不允许抢占的最优完工时间要长。

§ 3.4 其他获奖论文的模型与方法

北卡罗莱纳大学参赛队的获奖论文对情况 A 利用最大基数匹配的概念提出了类似的模型与方法，对于情况 B 与 C，则利用一种多项式时间的“贪婪算法”的思想，提出了一种近似的求解方法，它在迭代计算的每一步尽量挑选传输时间长的文件进行传输。贪婪算法对一般结构的网络可以得到比较好的近似解。

Beloit 学院参赛队的获奖论文在设计模型与方法时也是基于“贪婪算法”的思想，它希望在迭代计算时由每一步的局部最优化而得到问题的整体最优化（一般而言，只是近似最优解）。论文对网络的顶点安排了一个优先顺序，使得在计算的每一步尽量选出优先权最大的两个顶点（计算机）来传输文件。顶点之间顺序的安排根据一种改变的“装箱算法”。对任意顶点 V_i ，设有 $C(V_i)$ 个箱子，箱子是一维的，只考虑其长度，且箱子的长度是足够大。设通过 V_i 传输的文件有 k 个，将 k 个文件重新编号为 $e_{(1)}, e_{(2)}, \dots, e_{(k)}$ ，满足 $T(e_{(1)}) \leq T(e_{(2)}) \leq \dots \leq T(e_{(k)})$ 。装箱算法如下：令 $c = C(V_i)$ ，

第一步 如 $k \leq c$ ，则将每一文件放入一箱。 $T(e_{(k)})$ 为计算机 V_i 需负担的最短工时。

第二步 如 $k > c$ ，则置 $e_{(1)}$ 入第一箱， $e_{(k-1)}$ 入第二箱， \dots ， $e_{(k-c+1)}$ 入第 c 箱。再依次将 $e_{(k-c)}, \dots, e_{(1)}$ 置入一箱，此箱子已有的文件传输时间的和为最小。最后令 c 只箱子中文件传输时间之和的最大者为 V_i 需负担的最短工时。

将每一计算机的最短工时按增加的顺序加以排列，相应的计算机的顺序作为它们的优先权的顺序。

Harvey Mudd 学院参赛队在论文的附录中证明了一个有意义的结果：在“树”结构的通讯网络中寻找最优完工时间问题是“NP-完全问题”。这一结果是基于“划分问题”的任一实例都可在多项式时间内约化为树结构的通讯网络的完工时间问题的一实例。

§ 3.5 评注

通讯网络的完工时间问题是一个很好的数模竞赛题，不但在实际中有广泛的应用背景，而且是有理论意义的问题。就题型而言，它属于离散最优化（组合最优化）中的时间表问题（scheduling problems）。时间表问题是一个极富挑战性的研究领域，其中的问题有多种类型。通讯网络问题属于“Job-shop”类型的时间表问题，文件相当于“工件”（job），传输时间相当于工件的“工时”（processing time），每个工件的“工序”（operation）都为 1。通讯网络问题的特点是每个工件都需两台机器（计算机）同时进行加工，这一特点不同于普通的一台机器加工一个工件的情形。更一般的，是多台机器同时加工一个工件的时间表问题，这是近年来国际上时间表问题新的研究课题，难度很大。“完工时间”（makespan）是一种常用的优化目标。就本竞赛题而论，它可表述为与树图（情况 A 和 B）和简单图（情况 C）上的匹配有关的“边染色”问题。特别是情况 A，由于顶点的容量与文件传输时间都是 1，最优完工时间就是树图的“边色数”，它等于树图的最大次，即是 3（见图 3-2）。对于情况 B 和 C 中的一般简单图，多数获奖的竞赛论文注意了通讯网络与匹配和边染色的联系，并利用了图论中关于边色数性质的 Vizing 定理的结果和“贪婪方法”的思想来设计模型和求解方法。上面介绍的多伦多大学参赛队的论文的方法是有代表性的，它借助于计算出一列最大基数匹配或最大权匹配来求出通讯网络的最优完工时间，其他获奖论文的方法

也多是在每一步尽量寻求局部最优，以期达到整体最优。但对于一般结构的通讯网络问题，“贪婪方法”不能保证求出精确最优解，而只是近似解。个别论文指出了这一点。竞赛题中的三个数值问题由于问题规模较小，结构较简单（即使用手算也可在不很长的时间内找出最优完工时间），用“贪婪方法”可以顺利地求出最优解，但不能保证“贪婪方法”对于大规模通讯网络问题也求出最优解。这个竞赛题的价值还表现在，它会引导一些参赛者和感兴趣的读者去思考如何设计效果更好的多项式时间近似方法。这就是要分析近似方法的结果与精确最优解之间的误差。如果对竞赛题中的网络结构与数据加以改变，就可检验所提出的方法的计算结果的误差情况。似乎所有发表的获奖论文未讨论这一点。当然，从理论上研究近似方法的误差，即研究近似方法在最坏情况下的性能比，这是关于离散最优化的深层次的问题，已不是对大学生数学建模竞赛所应要求的。

参 考 文 献

- (1) Baker, K. R., *Introduction to Sequencing and Scheduling*. New York, Wiley, 1974.
- (2) 田丰、马仲善，图与网络流理论，科学出版社，1987。
- (3) 刘家壮、徐源，网络最优化，高等教育出版社，1991。
- (4) Papadimitriou, C. H., Steiglitz, K., *组合最优化：算法和复杂性*，刘振宏，蔡茂诚译，清华大学出版社，1988。
- (5) Evans, J. R., Minieka, E., *Optimization Algorithms for Networks*. New York, Marcel Dekker, 1992.

第四章 螺旋线的交点

孙山泽

北京大学 数学科学学院

提 要

本章介绍了 1995 年美国大学生数学建模竞赛(MCM-1995)的竞赛情况、评阅和奖励。特别介绍了 A 题的背景、阅卷人对参赛论文的评论和两篇优秀论文。我们对这些内容作了适当的编辑、补充和评述。

§ 4.1 MCM-1995 的评阅、结果和奖励

本次竞赛共有包括美国、中国、香港等 9 个国家和地区的 194 所大学的 320 个队参加，其中中国有 31 所大学的 84 个队参加。

各队的论文在 COMAP 的总部进行编号使得评阅人不知道论文作者的姓名和所属的学校。

初评是在马里兰州的索尔兹伯里(Salisbury)州立大学进行的，共有 14 位评阅人。每篇论文由两个初评评阅人评阅，摘要和论文的组织是论文评定的基础。如果两个评阅人的评分不同则进行协商，如果协商后还不一致，则再由第三位评阅人来评阅。终评是在加州的 Harvey Mudd 学院进行的，A 题评阅人有 10 位，B 题评阅人有 12 位。

MCM-1995A 题是由华盛顿州东华盛顿大学的 Yves Nieuwergelt 提供的，MCM-1995B 题是由马里兰州的索尔兹伯里(Salis-

bury)州立大学的 Kathleen M. Shannon 提供的,数据取自该校的公开信息.

评出的最后结果是:

	O	M	H	P	合计
MCM-1995A 题获奖队数 (中国队数)	3(0)	18(6)	43(10)	82(20)	146(36)
MCM-1995B 题获奖队数 (中国队数)	4(0)	26(7)	41(16)	103(25)	174(48)

其中, O=Outstanding=特等奖, M=Meritorious=一等奖, H=Honorable Mention=二等奖, P=Successful Participant=成功参赛奖.

每个参赛队的指导教师和队员都将获得由竞赛主任和每题的评阅组长签名的证书. 美国运筹学和管理科学学会(ORSA)给予两个获得特等奖的队现金奖励和三年的会员资格. 这两个队分别是明尼苏达州的麦卡莱斯特(Macalester)学院队(A题)和加州的哈维·马德(Harvey Mudd)学院队(B题), 将请他们在特设的MCM分会上作报告, 并给以现金奖励.

美国工业与应用数学学会(SIAM)对每题指定一个特等奖队作为SIAM的获奖队, 每个队员获得现金奖励, 每个队得到半程路费资助参加于1995年7月在加州圣迭戈(San Diego)举行的SIAM年会, 并在特设的小型建模讨论会上作报告. 这两个队分别是衣阿华州衣阿华(Iowa)州立大学队(A题)和阿拉斯加州阿拉斯加大学的费尔班克斯(Alaska Fairbanks)分校队(B题).

§ 4.2 问题的背景

该问题是Eastern Washington University的Niebergelt教授提供的, 问题是求出空间中一般位置的一条螺旋线和一个平面的

全部交点。这来源于美国西部的一家小公司。该公司从事医疗技术的开发研制工作，他们的一个设备有一个螺旋形部件，医生和技术人员一起制造这个装置，它应该能适用于对每一个病人的特殊测量。病人的 X-射线数据载入一台数值计算机，计算机能够处理三维图像，且要有一个程序计算希望获得的交点。医生和技术人员可以迅速改变螺旋线的参数，使得螺旋形部件在空间与病人的模型重叠，然后扫描一个平面片段观察那些临界的位置。模糊的医学描述转换成数学的精确叙述，这是数学的实际应用中常遇到的典型情况，螺旋线交点问题就是要用一种计算方法去解这样一个叙述简单的问题。

§ 4.3 评判者对参赛论文的评论

螺旋线交点问题的呈述、解的技术、结果的解释、答案的检查等解题的步骤是明确有序的，事实上提交评判的论文差不多都完成了上述各步，因而评判的准则聚焦于每一步是否实现得比较好，以及论文的结构和叙述是否清晰。评判中关注了下列一些方面。

各种可能情况的彻底分析。一个平面和一个螺旋线可以没有交点、有限数的交点或无限数的交点（在无限螺旋线的情形或零螺距的退化螺旋线），一个计算程序必须能对这些情况的每一种都近似地找到全部交点。

解决数值问题的简单性。给定一个螺旋线和一个平面，直接写出螺旋线的参数方程（依赖一个变量）和平面方程，再将参数方程代入平面方程产生一个单变量方程。找出这个单变量方程的根比找一个多变量方程的根要简单得多。

问题的数值解法。如果适当的端点被给定，二分法可以保证找到一个零值，但这个方法很慢。另一种选择是非线性问题牛顿法，它寻解是相当快的，然而一些队显然不知道牛顿法并非总是收敛的（例如，邻近的多重根）。由于可以找到适当的数值界限，

因而二分法是可用的。一些队用了这一方法。评判者认为，在寻解上加上可认定的界限的二分法技术优于未提到收敛问题的牛顿法。Macalester 学院队用了牛顿迭代法和二分法(当牛顿法失败时)两种方法。

结果的检验。为解决这一问题，编制的计算程序的结果可用多种方法进行检验。一些队用图的方法，另一些队则利用了用于更一般目的的方程解的结论。对一些软件产品(如 Mathematica)结果的可靠性似乎常存在混乱。

对问题的灵敏性。原来的问题要求涉及确定交点的计算速度，评判注意到这一要求。说清这一点有许多方法，给出计算每一个交点的时间，与一个更一般的数学解法(如 Mathematica)进行比较看节省的计算时间，或者论述算法计算的复杂性。

另外，真正突出的论文不仅求解了问题，而且考虑了可能的推广和可能的限制，这样做会使问题变难还是变易，以及可否使问题用到其他领域等等。

下面介绍 Macalester 学院队和 Harvey Mudd 学院队的两篇优秀参赛论文。

§ 4.4 Macalester 学院队的论文(节选)

4.4.1 问题

要求设计、执行、检验一个数学算法。这个算法应能“实时”地给出空间中一般位置的一个平面和一条螺旋线的全部交点。必须证明设计的算法在数学上和计算上均是正确的。

问题的解包括：确定设计的要求；构成问题的数学模型；设计并执行一个算法；调整和检测这一算法；最后给出评价。

4.4.2 建模分析

对螺旋线和平面，我们严格地给出一条圆柱形螺旋线的数学定义，同时认为平面是无穷展开的。设计的算法也可适用于其他类型的螺旋线，如椭圆柱形螺旋线。

数值计算有足够的精度。

“实时”意味着解题的时间应较短。当使用者给出螺旋线和平面的定位后，计算机屏幕应立即有反应。

4.4.3 数学模型

1. 一般情形

考虑一个一般的 3×4 的旋转-平移矩阵

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix}$$

和“单位”螺旋线(向量形式)

$$(\cos(at-t_0), \sin(at-t_0), t)$$

产生一般的螺旋线

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix} \begin{pmatrix} \cos(at-t_0) \\ \sin(at-t_0) \\ t \\ 1 \end{pmatrix}$$

得到在空间中一般螺旋线的方程

$$x = a_{11} \cos(at-t_0) + a_{12} \sin(at-t_0) + a_{13}t + a_{14},$$

$$y = a_{21} \cos(at-t_0) + a_{22} \sin(at-t_0) + a_{23}t + a_{24},$$

$$z = a_{31} \cos(at-t_0) + a_{32} \sin(at-t_0) + a_{33}t + a_{34}.$$

展开正弦、余弦项即可得到在笛卡儿坐标系中空间一般螺旋线的参数方程

$$x = a_{11} \cos at + a_{12} \sin at + a_{13}t + a_{14},$$

$$y = a_{21} \cos at + a_{22} \sin at + a_{23}t + a_{24},$$

$$z = a_{31} \cos \alpha t + a_{32} \sin \alpha t + a_{33} t + a_{34}.$$

空间中平面的方程可以写成

$$ax + by + cz - d = 0.$$

将螺旋线的 x , y , z 的表示式代入平面方程左边, 如此产生一个 t 的函数 $f(t)$

$$f(t) = A \cos t + B \sin t + Ct - D,$$

其中 A , B , C , D 是适当的系数. 用 t/α 替代原来的 t , 则将原 \cos 和 \sin 项中的 α 归入到 C 中. 现在的任务是解方程 $f(t) = 0$. 这个方程只能求数值解. 给定参数 A , B , C 和 D , 用数值技术确定这个方程的全部根. 对问题给定确定界限的情况下, 许多已知文献的算法保证可收敛到根, 并且会给出结果的误差的界限. 方程 $f(t) = 0$ 的一些信息对我们采用的数值技术有很大的影响. 例如, 我们知道周期地出现函数的极端值.

第一步考察函数的一般形式, 如图 4-1. 可以看到全部根一定处于两个极端值之间. 这两个极端值在 t 轴的两侧, 仅当根同时是一个极端值或弯曲点时是一种例外, 这种情形要给予特别处理.

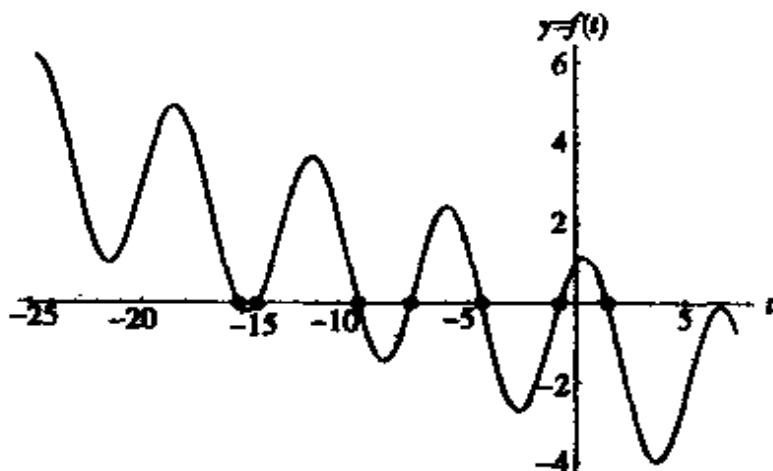


图 4-1 一般函数形式

我们从确定函数的极端小值和极端大值开始, 它们具有周期 2π . 可解出下述方程得到这些点.

$$f'(t) = -A\sin t + B\cos t + C = 0.$$

从

$$\cos t = \frac{A\sin t + C}{B}$$

和 $\cos^2 t + \sin^2 t = 1$, 得

$$\sin t = \frac{AC \pm B\sqrt{A^2 + B^2 - C^2}}{A^2 + B^2}.$$

由于开方会产生虚根, 要将这些值代回 $f'(t)$ 进行检查, 看是否为 0.

然后用直线连接两个极端值, 内插求出根, 如图 4-2, 并过滤这些内插值. 这样构成我们的求根算法.

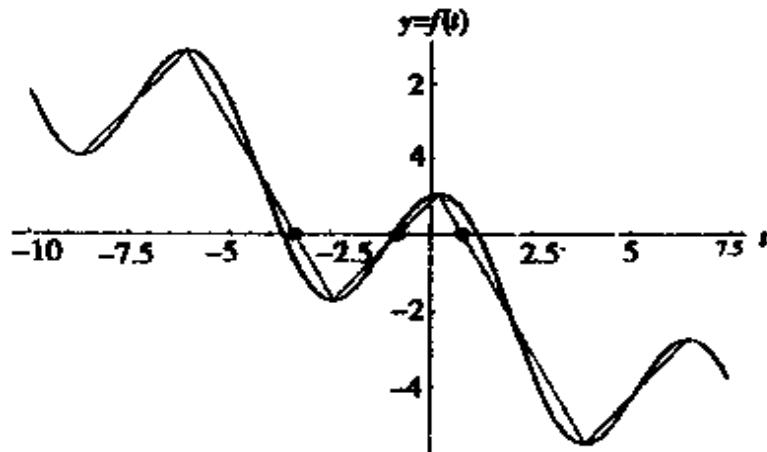


图 4-2 内插法

我们选择 $t_0 = D/C$ 作为数值算法的初始值, 围绕此值的 2π 邻近区域内会有一个根(如果根存在). 方程 $f(t) = Acost + Bsint + Ct - D$, 可看作参数方程 $x = \cos t$, $y = \sin t$, $z = t$ 确定的螺旋线和平面 $Ax + By + Cz = D$ 的交点. 这儿有一个半径为 1 的竖直螺旋线和平面. 螺旋线的中心线与平面的交点为 $t_0 = D/C$. 方程的根差不多围绕该点对称分布. 这可以从下述事实看出: 空间中一个平面和一个柱面的交点是一个椭圆, 螺旋线在柱面上, 它和平面的交点必在椭圆上. 椭圆的中心是平面和螺旋线中心轴的交点, 如果有根存在, 它们必定在绕椭圆中心的一圈完整转动之内, 因此用中心点作初始点.

2. 特殊情况

如果 C 为 0, 函数 $f(t)$ 是在一条直线附近周期振荡, 振幅不超过 $\sqrt{A^2 + B^2}$. 如果 $|D| > \sqrt{A^2 + B^2}$, 则函数与 t 轴无交点, 无根, 平面与螺旋线不相交. 如果 $|D| \leq \sqrt{A^2 + B^2}$, 则有无穷多个根.

另一个重要情况是仅有一个根. 当 $A^2 + B^2 - C^2 \leq 0$ 时, 会出现此种情况. 这时方程有一个实数解或无解. 如果有解, 它一定在弯曲点上. 我们的算法判定这一条件, 并用二分法替代 Newton 法来进行处理. 当根是弯曲点时, Newton 法收敛非常慢.

4.4.4 算法描述

算法描述有四个层次. 第一层用框图说明需解决的主要子问题(编者按: 限于篇幅, 发表时省略了). 第二是论述算法的工作细节. 第三是数学证明和详细的解释. 第四是 C++ 代码.

有下列一些约定.

输入——平面可用三种方式定义: 一般的笛卡儿方程 $ax + by + cz = d$; 用两个向量和一个点; 用三个点.

螺旋线用两种方式: 用一般的参数方程; 用三个欧拉角和映像 z 轴到螺旋线中心轴的变换向量.

输出——如果螺旋线和平面不相交, 无根被输出.

如果存在无限多个解, 则提供充足的信息, 使得使用者能产生全部交点.

其他情形给出交点的 x , y , z 坐标.

可装载性——我们的算法在 ANSI C++ 中运行, 足以保证适用于多数计算机.

4.4.5 检验和质量控制

我们花了很多的力量进行检验. 在四个不同的方面审查想法和运行的正确性.

数学模型. 我们使用的全部变换和函数形式是用 Mathematica 的标准构件和 Vector Analysis and Rotations 包产生的. 方程的全部字符解用 Mathematica 进行审查.

算法设计. 我们的求根步骤进行了仔细的选择, 必要时将引用二分法.

运行情况. 为了我们的运行, 转录表示式时, 我们将 Mathematica 表示式转换成 C 代码. 极小化了进入错误表达式的机会. 我们的求根步骤来自 Plybon [1992]. 用已知的 Mathematica 的 FindRoot 程序对它进行了检验. 我们的步骤在找出一个根上从未失败过, 也从未报告过一个不存在的根. 在几个 FindRoot 失败的情形, 我们的步骤能进行正确的处理, 找出一个根.

贮存分配. 对算法输出我们作了三种类型的审查.

1. 产生了五十多个 $f(t) = A\cos t + B\sin t + Ct - D$ 形式的函数, 检查算法是否能正确地发现他们的根. 运行结果与由一组 Mathematica 程序求得的结果是一致的. 在多数情形, 我们还直观地查看了 $f(t)$ 的图形, 以保证没有根被遗漏, 且无假的根被导入. 检查包括了无根、有一个根、有多个根和有无限多根的各种情形.

2. 我们输入五十多组螺旋线和平面, 用我们的算法找它们的交点. 然后对每一次用 Mathematica 作平面和螺旋线的三维图, 从各个视点查看, 以保证没有遗漏交点且无多余交点. 我们的算法通过了全部检查.

3. 我们设计了一种检验. 用一组已知交点的情况检查我们的算法. 在全部五十多个情形, 算法均执行正确.

4.4.6 评价

1. 正确性

检验已经说明我们的算法数学上和计算上均是正确的. 在一些情况我们还研究了数值计算复合误差的可能性.

在计算 $f(t)$ 的系数 A, B, C 和 D 时，会产生复合误差。我们不能确定这些误差的界限，但和高精度下 Mathematica 运行结果比较，显示这种误差小于 10^{-12} 。

根据使用的 Newton 法和二分法，当给出 A, B, C 和 D 的正确值时，我们能保证找出的根可以达到希望的精度。

当计算每一个根的包含区域时也会遇到不利的误差。如果数值计算的误差将包含区域放到根的一侧，根就会丢失。但我们从未遇到这种情形。当然可以采取一些措施避免这种错误，但会影响求根的速度。

另外，在求 $f(t)=0$ 的解时，如果两个根非常接近，可能被误认为一个。然而在数值误差范围内区分它们也无实际意义。

2. 稳健性

我们检查了可能是陷阱的各种特殊情形，如相切、弯曲点、双根等等。

3. 效率

由于我们对每一个交点做单独求根，算法的执行时间与交点数呈线性关系。求根的复杂性依赖于函数的形状和要求的有效数字位数。用 Newton 法时，求每一个根需 5、6 次迭代，即可达 12 位的精度。用二分法时，迭代少于 44 次。

4.4.7 改进的建议

算法的结构非常接近数学模型，无什么改进余地。算法的执行在下列几方面尚可改进。

可以修改表示式的求值以减少归整误差和复合误差。

可以对由计算系数 A, B, C, D 引起的误差找出一个界限。

对算法运行中的不可接收的计算错误加以监视，但这会影响速度。

输入可以包括其他螺旋线的定义形式。

在检验中使用的 Mathematica 程序可发展为无限精确的计

算，运行此模式的随机检查，用“实时算法”提供解的正确性。
算法可发展到处理其他形式的螺旋线。

§ 4.5 Harvey Mudd 学院队的论文(节选)

4.5.1 假定

问题中有一些潜在的未阐明的参数，我们对研究的范围作了下列假定：

非退化常规有限螺旋线。我们假定在任一固定时刻，螺旋线是圆圈的、有限高度的、有一致的倾斜和非零的直径。

有效的无限平面。我们考虑一个无限平面，因为螺旋线在一个平面片段的边缘可以跑出跑进，这会给问题增加很多麻烦。

相对坐标系。一个交点的位置由算法程序在它自己的坐标系中产生，而被表示成命名的函数。如果需要这个命名的函数可以经尺度变换，将这些交点放到适当的坐标系中，以适合于使用者的界面，也可投影这些三维点到一个二维屏幕。

4.5.2 构造模型

我们仅关心两个物体的相对旋转和位置。按照这种考虑，我们可以自由地选择一种安排，使问题在数学上比较易于处理。

指定一个坐标系。首先我们指定螺旋线的一端作为始点，另一端作为终点。 z 轴为螺旋线的轴， x 轴包含始点，因此坐标系的原点被固定于螺旋线轴的底部。为便于讨论，我们仅考虑右手螺旋线，但讨论很易推广到左手螺旋线。

为了在这个坐标系中放置平面，取 $(0, 0, z_0)$ 为此切割平面与 z 轴的交点，这个相交要保证平面不平行于 z 轴（特殊情形另行考虑）。由于设定了平面上的一个点 $(0, 0, z_0)$ ，平面可用两个角度 θ_0 和 φ_0 完全描述。 θ_0 是由 x 轴逆时针方向转动至该切割平面与平面 $z=z_0$ 的交线方向的转角， φ_0 是该平面对 z 轴的倾斜角。

可以这样设想，用 z_0 , θ_0 , φ_0 描述平面，是从一个由 x - z 轴构成的垂直平面开始，绕 z 轴逆时针旋转平面 θ_0 弧度，再从 z 轴向后转 φ_0 弧度，最后平移到 z_0 位置。

用这样的方式定义的坐标系，我们可以保存螺旋线和切割平面间的相对转动，而同时固定螺旋线在垂直位置。在这个坐标系中能够容易地用明晰的方程描述这两个物体。如果螺旋线有半径 R 、螺距 P 和高度 L ，用参数方程在直角坐标系中的表示为

$$x = R \cos 2\pi t, y = R \sin 2\pi t, z = Pt, \\ 0 \leq t \leq L/P. \quad (1)$$

类似地，平面在此直角坐标系中的描述为

$$z = -\sin \theta_0 \tan \left(\frac{\pi}{2} - \varphi_0 \right) x + \cos \theta_0 \tan \left(\frac{\pi}{2} - \varphi_0 \right) y + z_0. \quad (2)$$

现在考虑一个搁在 x - y 平面上，中心为原点，半径为 R ，高为 L 的正圆柱面，这个柱面为

$$x = R \cos 2\pi t, y = R \sin 2\pi t, z = s, 0 \leq s \leq L. \quad (3)$$

这个柱面包含螺旋线，它和切割平面的交点必定包含螺旋线与平面的全部交点。柱面与平面的交点正好是柱面在平面上的投影，由(2)和(3)可以解出这个曲线方程为

$$z = R \tan \left(\frac{\pi}{2} - \varphi_0 \right) \sin(2\pi t - \theta_0) + z_0,$$

从而得到螺旋线和平面交点的参数方程为

$$Pt = R \tan \left(\frac{\pi}{2} - \varphi_0 \right) \sin(2\pi t - \theta_0) + z_0. \quad (4)$$

这是一个超越方程，当 $\varphi_0 \neq 0$ 时，不能找出解析解。当 $\varphi_0 = 0$ ，这个方程不是螺旋线与平面交点的可用描述。这种特殊情形另行处理。

方程(4)的处理是相当困难的，为了易于分析，定义

$$\tau = 2\pi t, \beta = \frac{2\pi z_0}{P}, \sigma = \frac{2\pi R \tan \left(\frac{\pi}{2} - \varphi_0 \right)}{P}.$$

这样无量纲化该方程，我们可以考虑方程

$$\tau = \sigma \sin(\tau - \theta_0) + \beta,$$

定义函数

$$f(\tau) = \sigma \sin(\tau - \theta_0) + \beta - \tau. \quad (5)$$

解方程 $f(\tau) = 0$ ，通过此方程的根，我们找出螺旋线与平面的交点。设找出 $f(\cdot)$ 的根为 τ^* ，用 2π 除这个值并代入方程(1)，从而确定出在笛卡儿坐标系中的交点。

4.5.3 模型分析

在分析这个模型前，我们先看一下函数 $f(\cdot)$ ，注意函数 $f(\cdot)$ 有上边界限 $\sigma + \beta - \tau$ 和下边界限 $-\sigma + \beta - \tau$ 。这两条线在 $\beta \pm \sigma$ 处与 τ 轴相交，所以我们可以限定在 $(\beta - \sigma, \beta + \sigma)$ 内对 τ 找寻 f 函数的根。当然这个区间限制并不足以使解的速度达到“实时”的要求。这个区间的大小依赖于使用者控制的变量，无法保证是小的。况且由于这个函数快速振动，多数求根的标准算法，如牛顿法和二分法，对这个函数将不能很好地运用。

作为第一步近似，我设计一种寻根的技术。由 $f(\cdot)$ 的一阶导数的根确定出 $f(\cdot)$ 的极小值点和极大值点。解

$$\frac{df}{d\tau} = \sigma \cos(\tau - \theta_0) - 1 = 0,$$

得到

$$\tau = \theta_0 \pm \arccos \frac{1}{\sigma} + 2\pi n, \quad \sigma > 1,$$

($|\sigma| \leq 1$ 的情形，另行处理)，用这些点构造一个锯齿状线性近似 f 的函数 g ，即连接 f 的极大值点和极小值点。通过寻找 g 的根，可以容易而快速地粗估出 f 的根。应该注意 f 与 g 有相同数目的根，这就保证这一方法不会丢失也不会添加交点。

当 τ 递增时， $f(\cdot)$ 的根出现在极大值点和接着的一个极小值点之间或者出现在极小值点和接着的一个极大值点之间。我们

称前一种为“下降根”，后一种为“上升根”。

然而在许多情形，这种 f 的线性逼近的根可能不能满足精度要求，这时要用更精确的算法去寻找 $f(\cdot)$ 的根。

以 g 的根作为种子值，采用修改的牛顿法去寻找 f 的根。修改之处是在种子值处采用 f 的导数的一个近似值，基本上我们使用了两个近似值，对上斜采用常数近似值 $\pi/2\sigma$ ，对下斜采用常数近似值 $-\pi/2\sigma$ ，它分别用于确定“下降根”和“上升根”。这样做是基于下列两个原因。其一是每一个寻数值计算都要计算一个余弦函数，使用常数值显著地减少了计算时间；其二是对 $f(\cdot)$ 函数这样的具有周期导函数的函数，用牛顿法求根常会产生一些麻烦。

4.5.4 特殊情形的处理

1. 有垂直切割平面的情形

当 $\varphi_0 = 0$ 时，切割平面平行于 z 轴，参数 z_0 没有解释。为了描述平面，设平面到 z 轴的最短竖直距离为 r_0 ，则平面的方程为

$$r = \frac{r_0}{\sin(2\pi t - \theta_0)}.$$

注意，如果 $r_0 > R$ ，平面和螺旋线没有交点。当 $r_0 \leq R$ ，寻找垂直切割平面和螺旋线的交点等价于寻找满足下列方程的全部 t ：

$$R = \begin{cases} \frac{r_0}{\sin(2\pi t - \theta_0 - 2\pi n)}, \\ \frac{r_0}{\sin(\pi - 2\pi t + \theta_0 - 2\pi m)}. \end{cases}$$

这里 n, m 是整数。解出

$$t = \begin{cases} \frac{\theta_0 + \arcsin \frac{r_0}{R}}{2\pi} + n, \\ \frac{\theta_0 - \arcsin \frac{r_0}{R}}{2\pi} + m + \frac{1}{2}. \end{cases}$$

找到这些 t 值后，仅需将它们代回螺旋线的方程即可确定全部的交点。

2. $\sigma \leq 1$ 的情形

当 $\sigma \leq 1$ 时，导数 $df/d\tau = \sigma \cos(\tau - \theta_0) - 1$ 对一切 τ 值都是非正的，如此 f 是一个单调非增函数，与 τ 轴的交点仅有一个，因为这个根一定在 $\beta - \sigma$ 和 $\beta + \sigma$ 之间，故用简单的二分法即可，我们的程序在 β 处开始二分法，最初步长为 $\sigma/2$ 。

4.5.5 检验

程序是用 C++ 代码写成的，并做了计算速度和精度的若干试验。

运行时间。我们设定倾斜角 φ_0 为 45° ，转动角 θ_0 每 10° 确定一个切割平面。我们认为这样基本上代表了使用者的典型情况。每一个切割平面与螺旋线平均大约有 10 个交点，算出这 10 个交点大致花费 $1/25$ 秒的时间。

精度。在每一个试算中，我们以三维空间中估算的交点与实际交点的距离作为精度的描述。试算表明我们的寻根程序比通用的 Mathematica 软件的寻根程序（该程序使用牛顿法）计算的结果要精确。

另外还通过模拟产生图像进行了检验。

螺旋线的交点这一问题，其叙述是直接而清楚的。优秀的参赛论文都研究了一个圆柱形螺旋线，但其基本解法也适用于椭圆柱形螺旋线和锥形螺旋线。这些论文使用了通常求空间中曲面和曲线交点常用的方法，即隐式描述曲面和参数化曲线，然后当参数形式满足隐式方程时，产生一个交点方程。讨论这个方程的一般形式，解出交点。但各队建立方程的途径和限制假定稍有不同，主要的差异是寻找交点方程根的策略。

参 考 文 献

The UMAP Journal, vol. 16, No 3, 1995.

第五章 工资调整方案

姜启源

清华大学 数学科学系

提 要

本文取材于 1995 年美国大学生数学建模竞赛 B 题及发表在 UMAP vol. 16, No. 3 上的优秀论文^{[1][2]}. 按照题目给出的原则, 建立了基于教师职称和教龄的工资调整的理想目标函数, 分别就不考虑货币贬值和考虑货币贬值两种情况, 给出了从目前工资向理想目标过渡的方案, 并用计算机模拟对方案的可行性及对模型参数的敏感性作了检验.

§ 5.1 问题的提出

Atuacha Balacava 学院聘了一位新主任, 学院教师的工资标准是前主任制订的. 新主任把设计一个公正、合理的工资体系作为她的第一个任务, 现在她请你的队为顾问, 设计一个能反映以下情况和原则的工资体系.

情况

教师职称由低到高分 4 个等级: 讲师, 助理教授, 副教授, 教授. 获博士学位的聘为助理教授, 正在读博士的聘为讲师, 并在他完成学位时自动提升为助理教授. 在副教授职称上工作 7 年(或 7 年以上) 的可申请提升教授, 提升由主任会同学院的委员会讨论决定, 你不必考虑.

教师工资以 10 个月(每年 9 月到来年 6 月) 为一期, 在 9

月以前增加工资就有效。可用于增加工资的总额每年不同，一般在每年3月才能确定这个数额。

没有教学经验的讲师的起始工资是\$27000，助理教授是\$32000。教师可以拿到在其他学院受聘、多达7年教学经验的资格证明。

原则

- 只要有足够的钱，所有教师每年都增加工资。
- 职称提升应带来实质性的利益，若某人在最短可能的时间内提升，那么他的所得应大致等于正常情况下（不提升）7年增加的工资。
- 按时提升（在一个职称上7年或8年）并具有25年（或更多）教龄的教师的工资应大致是有博士学位新教师工资的2倍。
- 同一职称等级中经历（年限）长的应比经历短的工资高，但是这种影响应随着年限的增加而渐减，即两个同一等级教师的工资应随着年限的增加渐趋一致。

目标

先不考虑生活费用的增长设计一个新的工资体系，然后再将生活费用增长的因素加入。最后，在不减少每人现在工资的条件下，设计一个从目前的工资体系到你给出的新体系的过渡方案，并讨论能够改进你的新体系的任何想法。目前教师的工资、职称等级、教龄见表5-1（表中职称用0, 1, 2, 3分别表示讲师、助理教授、副教授、教授）。

新主任需要详细的、能够实施的工资体系计划，及一份短的执行纲要，其语言要清晰，使她能向委员会和教师宣布。纲要应给出模型及其假设、优缺点和预期的结果。

表 5-1

序号	教龄	职称	工资	序号	教龄	职称	工资	序号	教龄	职称	工资
1	4	2	54000	35	23	3	60576	69	6	1	49134
2	19	1	43508	36	20	2	48926	70	4	1	29500
3	20	1	39072	37	9	3	57956	71	4	1	30186
4	11	3	53900	38	32	2	52214	72	7	1	32400
5	15	3	44206	39	15	1	39259	73	12	2	44501
6	17	1	37538	40	22	2	43672	74	2	1	31900
7	23	3	48844	41	6	0	45500	75	1	2	62500
8	10	1	32841	42	5	2	52262	76	1	1	34500
9	7	2	49981	43	5	2	57170	77	16	2	40637
10	20	2	42549	44	16	1	36958	78	4	2	35500
11	18	2	42649	45	23	1	37538	79	21	3	50521
12	19	3	60087	46	9	3	58974	80	12	1	35158
13	15	2	38002	47	8	3	49971	81	4	0	28500
14	4	1	30000	48	23	3	62742	82	16	3	46930
15	34	3	60576	49	39	2	52058	83	24	3	55811
16	28	1	44562	50	4	0	26500	84	6	1	30128
17	9	1	30893	51	5	1	33130	85	16	3	46090
18	22	2	46351	52	46	3	59749	86	5	1	28570
19	21	2	50979	53	4	2	37954	87	19	3	44612
20	20	1	48000	54	19	3	45833	88	17	1	36313
21	4	1	32500	55	6	2	35270	89	6	1	33479
22	14	2	38642	56	6	2	43037	90	14	2	38624
23	23	3	53500	57	20	3	59755	91	5	1	32210
24	21	2	42488	58	21	3	57797	92	9	2	48500
25	20	2	43892	59	4	2	53500	93	4	1	35150
26	5	1	35330	60	6	1	32319	94	25	3	50583
27	19	2	41147	61	17	1	35668	95	23	3	60800
28	15	1	34040	62	20	3	59333	96	17	1	38464
29	18	3	48944	63	4	1	30500	97	4	1	39500
30	7	1	30128	64	16	2	41352	98	3	1	52000
31	5	1	35330	65	15	3	43264	99	24	3	56922
32	6	2	35942	66	20	3	50935	100	2	3	78500
33	8	3	57295	67	6	1	45365	101	20	3	52345
34	10	1	36991	68	6	2	35941	102	9	1	35978

序号	教龄	职称	工资	序号	教龄	职称	工资	序号	教龄	职称	工资
103	24	1	43925	137	5	1	32210	171	23	3	51571
104	6	2	35270	138	21	2	43160	172	12	3	46500
105	14	3	49472	139	2	1	32000	173	6	1	35978
106	19	2	42215	140	7	1	36300	174	7	1	42256
107	12	1	40427	141	9	2	38624	175	23	2	46351
108	10	1	37021	142	21	3	49687	176	22	3	48280
109	18	2	44166	143	22	3	49972	177	3	1	55500
110	21	2	46157	144	7	2	46155	178	15	2	39265
111	8	1	32500	145	12	1	37159	179	4	1	29500
112	19	2	40785	146	9	1	32500	180	21	2	48359
113	10	2	38698	147	3	1	31500	181	23	3	48844
114	5	1	31170	148	13	0	31276	182	1	1	31000
115	1	0	26161	149	6	1	33378	183	6	1	32923
116	22	3	47974	150	19	3	45780	184	2	0	27700
117	10	2	37793	151	4	3	70500	185	16	3	40748
118	7	1	38117	152	27	3	59327	186	24	2	44715
119	26	3	62370	153	9	2	37954	187	9	2	37389
120	20	2	51991	154	5	2	36612	188	28	3	51064
121	1	1	31500	155	2	1	29500	189	19	0	34265
122	8	2	35941	156	3	3	66500	190	22	3	49756
123	14	2	39294	157	17	1	36378	191	19	1	36958
124	23	2	51991	158	5	2	46770	192	16	1	34550
125	1	1	30000	159	22	1	42772	193	22	3	50576
126	15	1	34638	160	6	1	31160	194	5	1	32210
127	20	2	56836	161	17	1	39072	195	2	1	28500
128	6	0	35451	162	20	1	42970	196	12	2	41178
129	10	1	32756	163	2	3	85500	197	22	3	53836
130	14	1	32922	164	20	1	49302	198	19	2	43519
131	12	2	36451	165	21	2	43054	199	4	1	32000
132	1	1	30000	166	21	3	49948	200	18	2	40089
133	17	3	48134	167	5	3	50810	201	23	3	52403
134	6	1	40436	168	19	2	51378	202	21	3	59234
135	2	2	54500	169	18	2	41267	203	22	3	51898
136	4	2	55000	170	18	1	42176	204	26	2	47047

这道题目共有 4 队获优秀奖，其中 Alaska Fairbanks 大学的

一个队为 SIAM (工业与应用数学学会) 优胜者, Harvey Mudd 学院的一个队为 INFORMS (运筹与管理协会) 优胜者. 下面 § 5.2 和 § 5.3 分别以这两个队的获奖论文为基础介绍建立模型的全过程, § 5.4 介绍论文评阅者的意见.

§ 5.2 Logistic 模型与尺度法⁽¹⁾

1. 假设条件

- 1) 不论是过渡阶段还是将来, 每个人都不减工资.
- 2) 如果所有人都拿到了理想工资, 就不一定要把所得到的款额都发工资.
- 3) 工资标准不看业绩, 只根据职称和教龄.

2. 理想工资函数

如果有足够的钱, 建立一个理想工资函数, 满足以下原则:

- 1) 一次提升顶 7 年教龄所加的工资.
- 2) 新讲师工资为 \$ 27000.
- 3) 新助理教授工资为 \$ 32000.
- 4) 25 年(或更多)教龄的教授工资为 \$ 64000.
- 5) 教龄大于 25 年的教授工资比只有 25 年教龄教授的工资要多.
- 6) 教龄增加时同一职称的工资渐趋一致.

为将教龄和职称统一考虑, 按原则 1) 定义指标 $x = \text{教龄} + 7 \times \text{职称等级} (0, 1, 2, 3)$, 其中 0, 1, 2, 3 分别表示讲师、助理教授、副教授、教授. 理想工资函数记作 $I(x)$, 满足:

$$I(0) = 27000, \quad I(7) = 32000,$$

$$I(46) = 64000 \pm 5\%, \quad I(71) = 64000 \pm 10\%,$$

$I(71) > I(46)$, 这里 71 表示教龄 50 年, 是随便取的.

当 x 较大时, $\frac{d^2 I}{dx^2} > 0$ (为什么? 请读者考虑).

检查了多项式、平方根、立方根、指数函数、幂函数及 Logistic 函数，最后选定 Logistic 函数作为理想工资函数，形式为

$$I(x) = \frac{k}{1+ae^{bx}}, \quad a, k > 0, b < 0. \quad (1)$$

若 k 已知，则 a, b 可由原则 2), 3) 求出。

用试探法确定 k ：对不同的 k 计算 $I(46)$ 和 $I(71)$ ，得到 $k=83000$ ，然后由 $I(0)$ 和 $I(7)$ 可得到 $a=56/27=2.07, b=-0.0376$ ，即理想工资函数为

$$I(x) = 83000 \left(1 + \frac{56}{27} e^{-0.0376x} \right)^{-1}. \quad (1)'$$

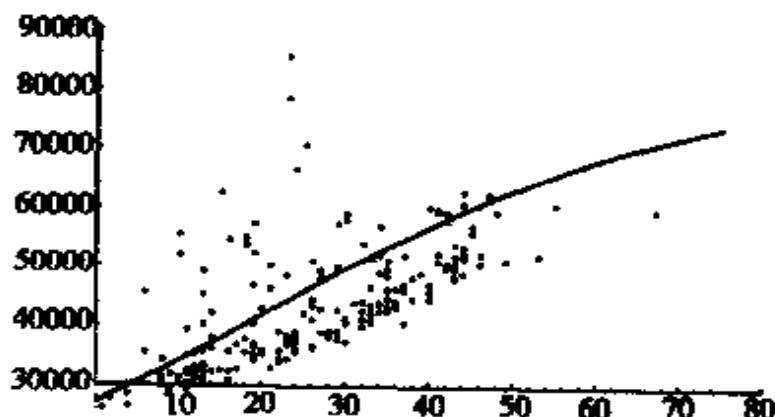


图 5-1 理想工资函数曲线和目前教师工资

图 5-1 是理想工资函数曲线(1)和目前教师的工资(用圆点表示)，可以看出，少数人的工资在曲线之上很多，多数人的工资在曲线之下一点。

3. 不考虑生活费用增长时的方案

讨论如何分配每年得到的款额给各个教师，作为工资的增加。

设每人有一固定的增加额，记作 R （其数值应由主任决定），记序号 i 教师的指标为 x_i ，目前工资为 S_i ，称

$$D_i = I(x_i) - (S_i + R) \quad (2)$$

为 i 教师的工资欠额。

设从总款项中扣除每人的固定增额 R 后剩余额为 M , 下面讨论 M 的分配方法 (若总款额不足以每人增加 R , 则平均分配之):

1) 欠额——比例法 (下记作方法 A)

M 按欠额的比例分配, i 下一年的工资为

$$S_i^+ = S_i + R + D_i M / \sum_i D_i. \quad (3)$$

若 $D_i < 0$, 按 $D_i = 0$ 处理.

2) 尺度法(下记作方法 B)

使尽可能多的人的工资按同一尺度增加, 靠近理想函数 $I(x)$, 设尺度(比例)为 c ($0 < c < 1$), 则 i 下一年的工资为

$$S_i^+ = \max\{S_i + R, cI(x_i)\} \quad (4)$$

记 $k' = ck$ (k 为(1)式中的系数); 确定 c 归结为确定 k' , 显然 k' 在 $I(0)$ 与 k 之间, 可用如下的二分试探法求出:

a) $k_{hi} := k$,

b) $k_{low} := I(0)$,

c) $k' := k = \frac{1}{2}(k_{hi} + k_{low})$,

d) 令 i 的临时工资为 $T_i^+ = \max\{S_i + R, cI(x_i)\}$,

e) 若 $\sum_i T_i^+$ 在总款额的允许误差(指四舍五入误差)之内, 则

$S_i^+ = T_i^+$,

f) 若 $\sum_i T_i^+$ 大于总款额, $k_{hi} := k$, 转 c,

g) 若 $\sum_i T_i^+$ 小于总款额, $k_{low} := k$, 转 c.

两种方法的比较:

两种方法的效果可由图 5-2、图 5-3 看出. 方法 A 使原来低于理想曲线的人的工资有了很大提高, 但对指标 x 相同的两位教师, 原来工资低的仍然低. 方法 B 则有所改进; 指标 x 相同的教师的工资将会很快地变为相同.

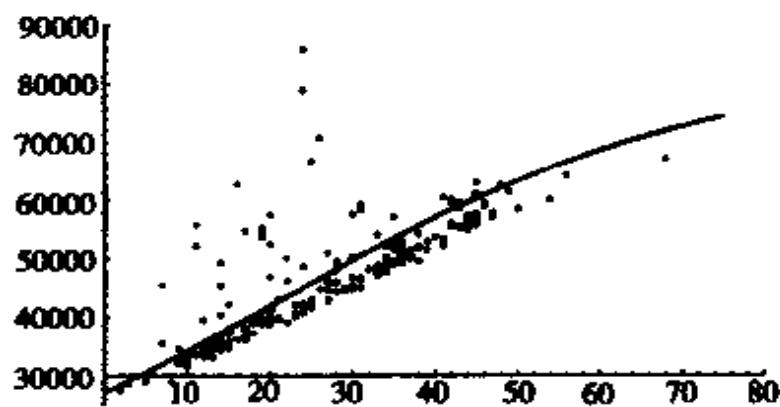


图 5-2 欠额-比例法(方法 A), 总款额 $\$ 9.5 \times 10^6$

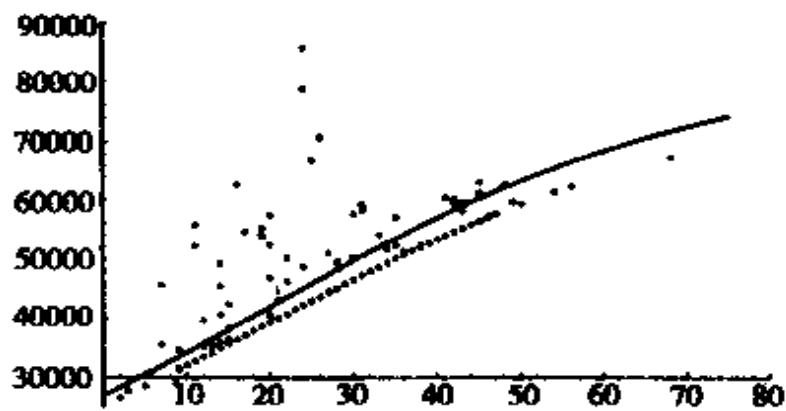


图 5-3 尺度法(方法 B), 总款额 $\$ 9.5 \times 10^6$

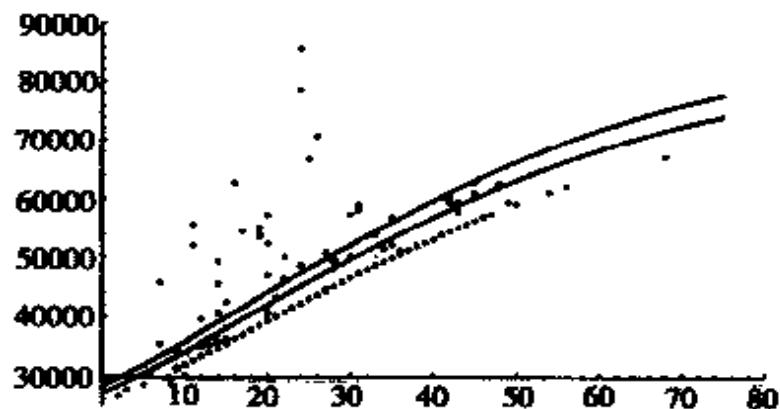


图 5-4 考虑生活费用指数 $c_1 = 5\%$ 的尺度法,
(方法 C) 总款额 $\$ 9.5 \times 10^6$

4. 考虑生活费用增长时的方案

设生活费用增长指数为 c_1 , 有两种方法考虑指数 c_1 .

1) 提升理想曲线(下记作方法 C)

将理想工资函数 $I(x)$ 中的系数 k 增加为 $k(1+c_1)$, 然后用尺度法计算, 结果如图 5-4.

2) 提升理想曲线同时增加工资 (下记作方法 D)

同上将 k 增至 $k(1+c_1)$. 同时, 若有足够的钱, 每人的工资都乘以 $(1+c_1)$; 否则, 以最大可能的比例给每人提薪, 即

$$S_i^+ = S_i + S_i M / \sum_i S_i.$$

两种方法的比较:

方法 C 不能立即提高每人的工资, 这对资金不足时是合适的, 长期来看, 它比方法 D 能使更多的人更快地提高工资, 因为后者使所有人都增加工资, 包括那些已高于理想曲线的, 这会消耗部分资金, 一旦这些人离去, 就会抵消低于理想曲线的人的利益.

5. 过渡时期

若总款额不足, 则方法 A, D 不如方法 B, C. 过渡时期的长短与高工资教师留在学院中任职的时间有关, 因为若资金充足就可以很快地使低工资的教师完成过渡, 达到理想曲线, 而不能将高工资者降下来.

过渡时期和未来阶段的实施方案是一样的, 这也是本体系的一个优点.

6. 检验与分析

为检验提出的各种方法, 作整个工资变动过程的模拟, 包括教师指标 x 随教龄增加与提升的改变, 及退职、聘用新人等.

模拟假设:

1) 各职称的人数不变, 用聘用新人的办法填补空缺, 任何职称的新人都可聘到.

2) 不论什么职称, 工作满 25 年者均有资格退职, 但学院不作强迫.

- 3) 对有资格提升的教师能否得到提升,由概率决定.
- 4) 有资格提升但未获提升的将离开学院.
- 5) 从助理教授到副教授的提升与任职的年限无关.

在上述假设下用 Mathematica 编制程序, 模拟运行 10 年、20 年甚至 100 年, 观察随着指标的变化, 各个教师的工资如何向理想曲线移动, 并用 Mathematica 作统计分析, 结果如下:

公平的工资体系应使指标相同的人有相同的工资, 因此用指标相同者工资的差距来衡量所提出的方法的优劣, 而这个差距可用相对标准差度量.

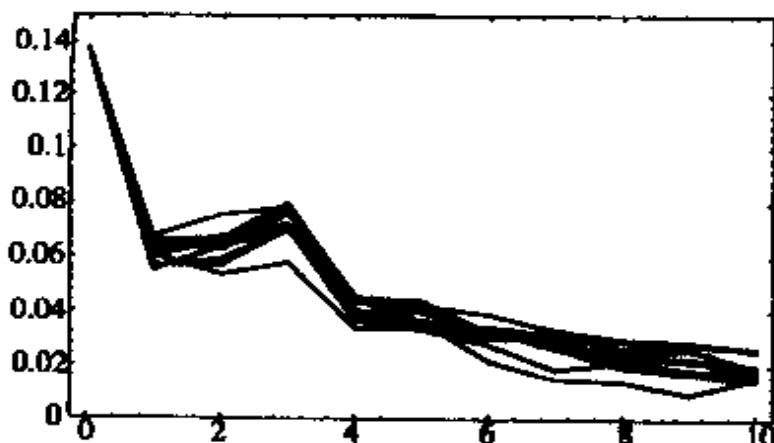


图 5-5 方法 A, 总款额年增 2%, 10 次运行
(横轴为年, 纵轴为相对标准差, 下同)

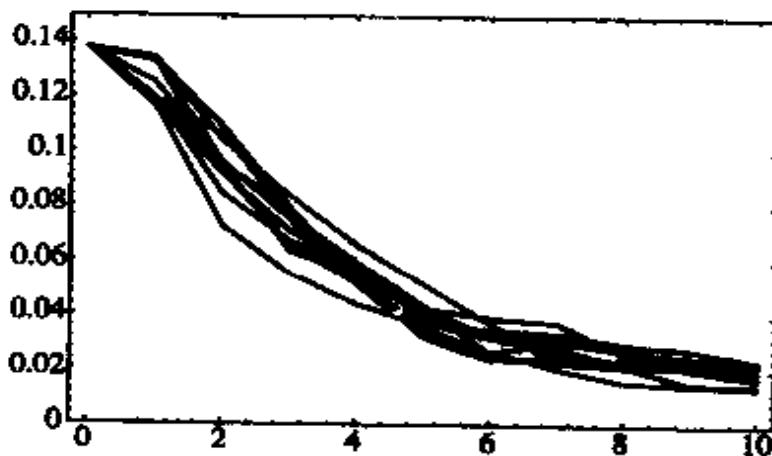


图 5-6 方法 B, 总款额年增 2%, 10 次运行

对于方法 A, B, C, D 各运行 10 年, 总数项每年增加 2%, 或 5%, 每年计算指标相同者工资的相对标准差, 然后在整个指标范围内取平均, 得到图 5-5~5-8, 每个图中有 10 条线表示模拟运行了 10 次, 以观察结果的随机性.

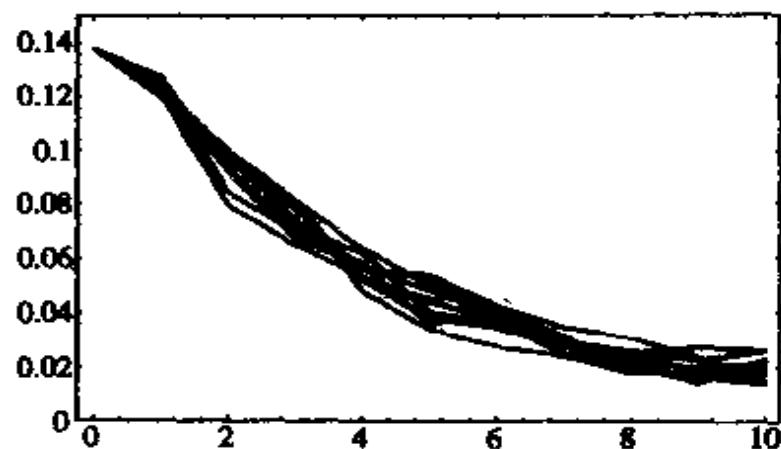


图 5-7 方法 C, 总款额年增 2%, 生活费
用指数 3%, 10 次运行

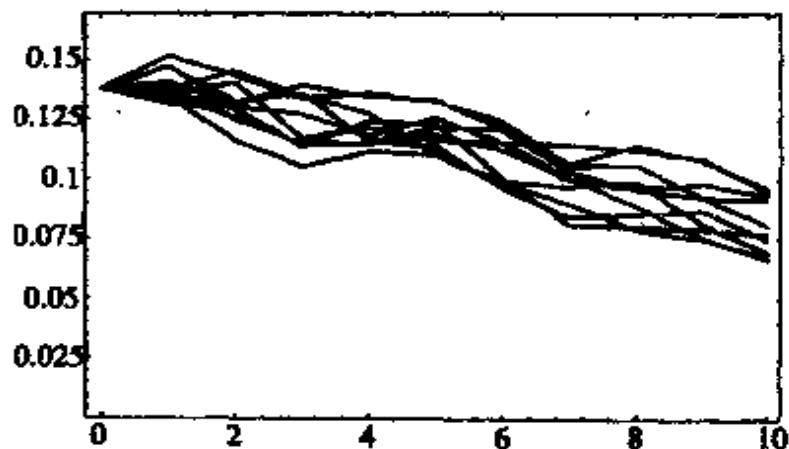


图 5-8 方法 D, 总款额年增 2%, 生活费
用指数 3%, 10 次运行

结果表明, 方法 B, C 在多次运行时较为一致, 且相对标准差较快地降到 0.02, 在总款额增加较多时方法 C 又比 B 好 (但是图中未给出总款额年增 5% 的结果); 方法 A 虽然也较快地使相对标准差降到 0.02, 但中间有波动; 方法 D 对每次运行的一

致性较差，且相对标准差降低不大。

§ 5.3 线性模型与对数模型的比较⁽²⁾

1. 假设条件

- 1) 即使资金不足以提高工资，也至少可以支付现有工资。
- 2) 25 年教龄的不一定退休，但最长工作 60 年。
- 3) 题目中所给“提升大致相当 7 年(不提升)增加的工资”和“按时提升并具有 25 年教龄者工资是新教师(博士学位)的两倍”，均指的是实际值，因为 25 年内即使加倍，也不能抵消生活费用指数上升 3% 的贬值。
- 4) 工资计算均指现值，这样可能不足以抵消生活费用指数上升的贬值。
- 5) 对提升慢的(多于 7 年)当提升时，所加工资应多于正常提升者。
- 6) 当有足够的资金时才聘新人，不从用于升工资的钱中聘人；退休者的工资不纳入用于升工资的资金，但可用于聘新人。
- 7) 由助理教授升副教授也需 7 年教龄(题目中无此规定)。

2. 问题分析与模型设计

应全面考虑一定职称、一定教龄的教师工资应是多少，以什么比例增加工资，鼓励不鼓励退休，是否聘用没有教学经验的教授等。

将设计两种方案供行政方面和教师选择。先不考虑贬值问题，然后再加以调整，并满足总资金的限制。对工资已经超过新标准的人，也有两种意见。最后将考虑超支的情形。

3. 两种方案

设教师 i 在第 t 年的工资目标值为 $T(i, t)$ ，他的起始工作年份为 t_i ，用 $i \in k=0, 1, 2, 3$ 分别表示 i 的职称是讲师、助理教授、副教授、教授，用 d_0, c_0, b_0, a_0 分别表示职称 $k=0, 1,$

2, 3 的起始工资, 按题目所给, 应有 $d_0 = 27000$, $c_0 = 32000$, $d_0 < c_0 < b_0 < a_0$, 且 $a_0 < 2c_0$ (因为教授工作 25 年后工资才大致为 $2c_0$), 由此可估计出 $b_0 = 36000$, $a_0 = 40000$ (见附录).

对目标值 $T(i, t)$ 设计如下两种方案:

对数函数

$$T(i, t) = \begin{cases} d_0 \log_{10}[d(t-t_i)+10], & i \in k=0; \\ c_0 \log_{10}[c(t-t_i)+10], & i \in k=1; \\ b_0 \log_{10}[b(t-t_i)+10], & i \in k=2; \\ a_0 \log_{10}[a(t-t_i)+10], & i \in k=3. \end{cases} \quad (5)$$

线性函数

$$T(i, t) = \begin{cases} c_0 + c(t-t_i-7), & i \in k=0; \\ c_0 + c(t-t_i), & i \in k=1; \\ b_0 + b(t-t_i), & i \in k=2; \\ a_0 + a(t-t_i), & i \in k=3. \end{cases} \quad (6)$$

选用这两种方案都可满足题目所给的条件: 对职称相同者当 t 充分大时工资差别渐趋消失. 对数函数使工资之差趋于 0; 线性函数使工资之比趋于 1.

虽然看起来这两条工资曲线没有水平渐近线, 但是我们限制最长工作年限为 60 年, 所以实际上工资是有上界的.

(5)、(6)式中的 a , b , c , d 将由假设条件确定 (见附录). (6)式中讲师的工资线由助理教授的工资线平移得到, 这是因为讲师提升的年限 (即他读博士学位的时间) 不确定, 平移的距离使讲师的起始工资刚好为 27000.

图 5-9 将这两族曲线画在一起以作比较. 4 条曲线自下而上分别为讲师、助理教授、副教授、教授在对数函数方案下的理想工资, 4 条直线则对应于线性函数方案. 可以看出, 对数方案在最初的若干年内工资上升较快, 而时间较长以后就上升平缓, 如教授工资在 25 年后增加显著变慢, 所以这一方案实际上在鼓励退休.

3. 若干实际问题的处理

1) 贬值

设生活费用指数引起的贬值率为 $r(t)$ (一般取 3%), 在第 t 年对下年的贬值率估计为 $r(t+1)$, 若新计划自 t_0 年开始执行, 则 $t+1$ 年的工资现值应为

$$N(i, t+1) = T(i, t + 1) r(t+1) \prod_{j=t_0}^t r(j). \quad (7)$$

2) 资金不足

已知 t 年的实际工资(现值) $n(i, t)$ 和 $t+1$ 年用于提高工资的资金 $M_r(t+1)$, 按(7)式计算 $t+1$ 年为达到理想工资所需资金为

$$M_n(t+1) = \sum [N(i, t+1) - n(i, t)]. \quad (8)$$

当 $M_r(t+1) < M_n(t+1)$ 时, 按差额比例分配, 即 $t+1$ 年的实际工资为

$$n(i, t+1) = n(i, t) + \frac{M_r(t+1)}{M_n(t+1)} [N(i, t+1) - n(i, t)]. \quad (9)$$

这样可使远离目标值的教师得到较大的比例, 能较快地接近理想线.

对于 $n(i, t)$ 已超过 $N(i, t+1)$ 的那些教师 i , 则只考虑贬值, 即用

$$u(i, t+1) = r(t+1) n(i, t) \quad (10)$$

代替(8)、(9)式中的 $N(i, t+1)$ 计算.

3) 资金有余

有几种处理方式, 如

- 提高每人的工资. 若资金余量足够大, 可使每人工资超

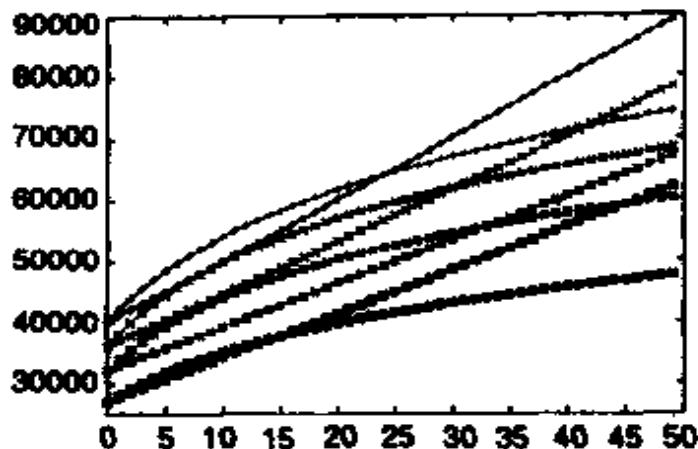


图 5-9 对数和线性两种方案的理想工资(横轴为年)

过理想值，但可能给院方未来的工资计划带来压力。

■ 给每人发奖金，这可能是比提高工资更灵活的，院方愿意采纳的方式。

■ 设立基金。

4. 模型检验

用题目所给数据按两个模型(对数、线性)各运行 50 年(自 1995 年开始)，研究贬值、有限资金、聘用新教师、职称提升、退职等因素及参数 a_0 、 b_0 变化对模型性能的影响。

1) 各因素对模型的影响

图 5-10、5-11 是在没有聘用、提职、退职且忽略贬值及

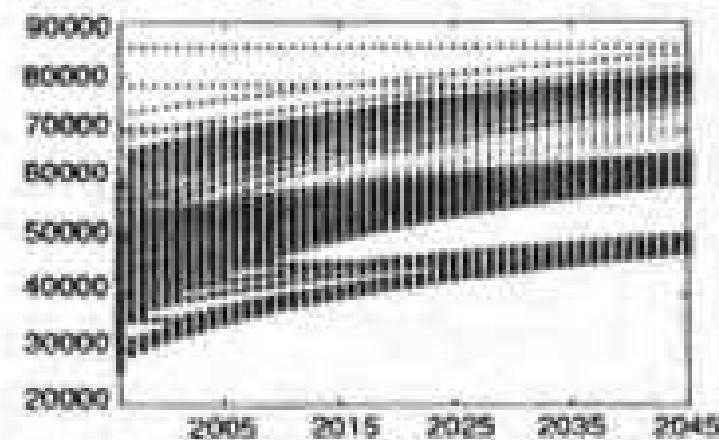


图 5-10 对数模型(无聘用、提升、退职、贬值、资金限制)

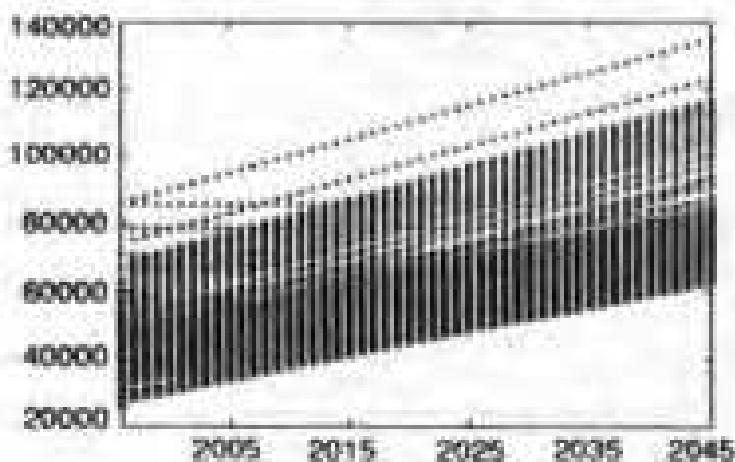


图 5-11 线性模型(无聘用、提升、退职、贬值、资金限制)

资金限制情况下的结果，原来低于目标值的随着时间增加将趋向

目标值，而高于目标值将维持不变，直到目标值赶上他们的现有工资。

图 5-12、5-13 考虑了 3% 的贬值及退休，而资金无限制，当时间较大时两个模型趋于一致。

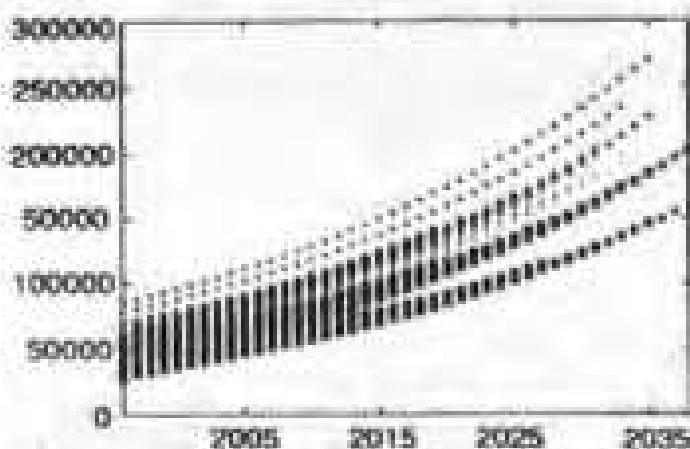


图 5-12 对数模型(有 3% 的贬值及退休)

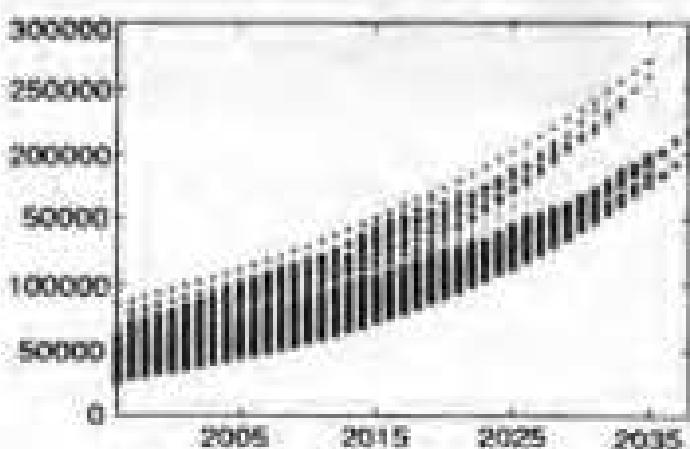


图 5-13 线性模型(有 3% 的贬值及退休)

图 5-14、5-15 考虑提升和退休，而他们原来的职位被聘用的新入代替。

图 5-16、5-17 考虑资金限制，而不考虑提升，在对数模型（图 5-16）实际工资与目标值的差随着时间增加渐小，而线性模型（图 5-17）则相反，这两个图形反映了这两个模型调节资金的能力。

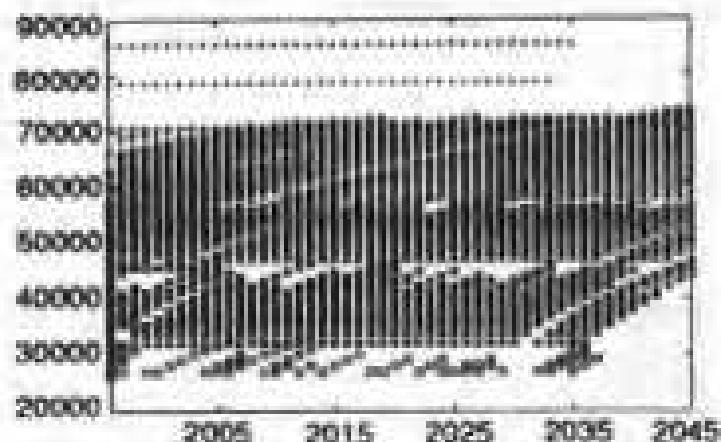


图 5-14 对数模型(考虑晋升、退休和聘用)

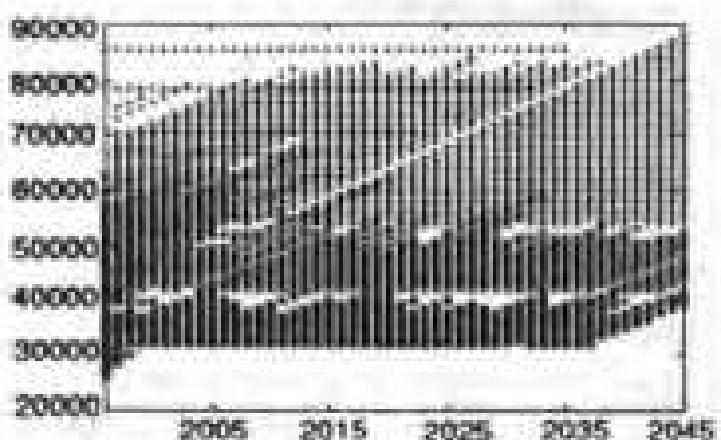


图 5-15 线性模型(考虑晋升、退休和聘用)

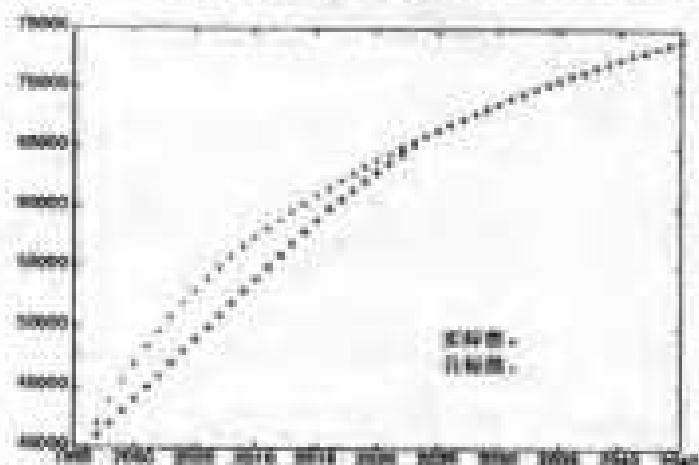


图 5-16 对数模型(有资金限制)

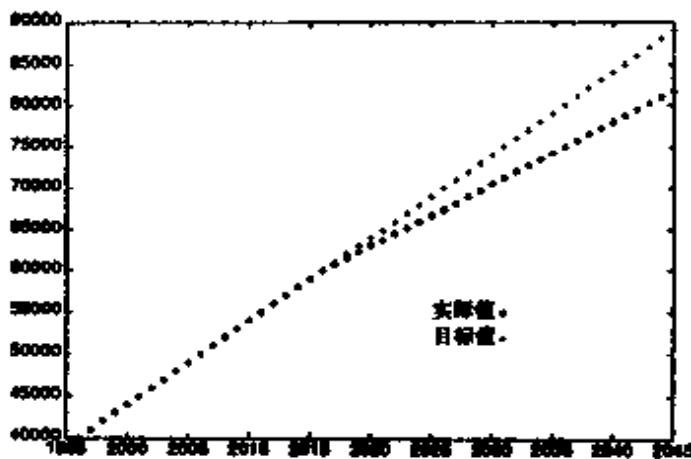


图 5-17 线性模型(有资金限制)

2) 灵敏性分析

改变 a_0 , b_0 同样地运行 50 年, 用 50 年后各职称的工资研究这两个起始工资对模型的灵敏性. (c_0 , d_0 不变, 因为它们是题目给定的), 结果列入表 5-2, 5-3.

由表 5-2 可知, 尽管 a_0 , b_0 有 10% 的改变, 50 年后各职称工资最多有 2% 的差别, 而在表 5-3 中则约有 8% 的变化, 所以我们的模型, 特别是对数模型对初始参数是相当不敏感的.

表 5-2 按对数模型, a_0 , b_0 改变对 50 年后工资的影响

a_0	b_0	教授	副教授	助理教授	讲师
40000	36000	74107	68312	60180	47556
42000	36000	73996	68545	60365	47601
40000	38000	74017	68306	60898	47744
42000	38000	73996	68548	61061	47788
38000	34000	73938	68048	59530	47391

3) 长期性能

模拟一些实际因素对模型的影响, 如需要决定每年有多少人提升、退职和聘用, 以及资金和贬值率等.

表 5-3 按线性模型, a_0 , b_0 改变对 50 年后工资的影响

a_0	b_0	教授	副教授	助理教授	讲师
40000	36000	89000	78000	67000	62000
42000	36000	86917	79944	67972	62833
40000	38000	89000	75333	71667	66000
42000	38000	86917	77278	72639	66833
38000	34000	91083	78722	61361	57166

在检查了所给数据后, 决定在模拟中采用聘用人数服从平均值为 9、标准差为 5 的离散化的正态分布; 多少有点随意地假定退职前的工作年限服从平均值为 40、标准差为 2 的离散化正态分布; 助理教授工作 7 年后有 50% 提升副教授, 8 年后再提 25%, 以后每年提升剩下的一半; 副教授提升教授也按此方案; 贬值率定为 3%. 模拟分资金无限和有限两种情况.

5. 附录

为推导(5)、(6)式中的 a , b , c , d , 当 $i \in k=0, 1, 2, 3$ 时将 $T(i, t)$ 记作 $T_k(t)$. 按题目所设, 应有

$$\begin{aligned} T_3(24) &= 2T_1(0), T_3(14) = T_2(21), T_2(7) = T_1(14), \\ T_1(1) &= T_0(8). \end{aligned}$$

代入(5)式可得

$$\begin{cases} a = (10^{2c_0/a_0} - 10)/24, \\ b = [(14a + 10)^{a_0/b_0} - 10]/21, \\ c = [(7b + 10)^{b_0/c_0} - 10]/14, \\ d = [(c + 10)^{c_0/d_0} - 10]/8. \end{cases} \quad (11)$$

代入(6)式可得

$$\begin{cases} a = (2c_0 - a_0)/24, \\ b = (14a + a_0 - b_0)/21, \\ c = (7b + b_0 - c_0)/14. \end{cases} \quad (12)$$

前面给出的 $a_0 = 40000$, $b_0 = 36000$ 是这样估计出来的; 对线性模型, 由(6)式讲师的起始工资是 $c_0 - 7c = 27000$, 而助理教授的起始工资是 $c_0 = 32000$, 于是 $c = 5/7$. 将 c_0 和 c 代入 (12) 式后有 a_0 , b_0 , a , b 共 4 个未知量. 上述 a_0 , b_0 是满足(12)式及 $a_0 < 2c_0$ 的一组合适的解.

§ 5.4 整体方案与特殊处理

在评阅者对优秀论文的评注中有以下几点值得注意⁽³⁾:

- 多数论文注意到这是一类曲线拟合问题, 其中优秀论文选取的拟合函数较好, 有的比较了多种函数才确定下来. 对 4 种职称可以用统一模型, 也可以分别处理.
- 优秀论文都对问题作了完整的、成熟的处理, 包括假设条件、精细的实施计划、灵敏性分析、生活费用指数的考虑等.
- 许多论文(包括优秀论文)没有将工资非常高、教龄又短的个别教授的工资作为“奇异点”(outlier)处理, 来自 North Carolina 数学科学学校的一个队这样做了(参阅优秀论文集).

参考文献

- [1] Liam Forbes, Marcus Martin, Michael Schmahl: How to Keep Your Job as Provost. UMAP Vol. 16, No. 3.
- [2] Jay Rosenberger, Andrew M. Ross, Dan Suyder: Paying Professors What They're Worth. UMAP Vol. 16, No. 3.
- [3] Donald E. Miller: Judge's Commentary: The Outstanding Faculty Salaries Papers. UMAP Vol. 16, No. 3.

第六章 潜艇探测问题

谭永基

复旦大学 数学系

提 要

本章介绍了 1996 年美国大学生数学建模竞赛 (MCM-1996) 的竞赛情况、评阅和奖励。特别介绍了 A 题的背景和部分优秀答案，并作了评述，对大学生数学建模竞赛的命题与阅卷发表了若干看法。

§ 6.1 MCM-1996 的评阅、结果和奖励

本次竞赛共有包括美国、中国、香港等 9 个国家和地区的 225 所大学的 393 个队参加，其中中国有 39 所大学的 115 个队参加。

各队的论文在 COMAP 的总部进行编号使得评阅人不知道论文作者的姓名和所属的学校。

初评是在蒙大拿州的卡罗尔 (Carroll) 学院进行的，共有 9 位评阅人。每篇论文由两个初评评阅人评阅，摘要和论文的组织是论文评定的基础。如果两个评阅人的评分不同则进行协商，如果协商后还不一致，则再由第三位评阅人来评阅。终评是在加州的哈维·马德 (Harvey Mudd) 学院进行的，A 题评阅人有 11 位，B 题评阅人有 16 位。

MCM-1996 A 题和 B 题都是由位于马萨诸塞州阿灵顿的 Zwillinger & Associate 的 Daniel Zwillinger 提供的。

评出的最后结果是：

	O	M	H	P	合计
MCM-1996A 题获奖队数(中国队数)	4(0)	16(9)	37(19)	70(18)	127(46)
MCM-1996B 题获奖队数(中国队数)	5(2)	38(6)	76(7)	148(54)	267(69)

其中, O=Outstanding=特等奖, M=Meritorious=一等奖, H=Honorable Mention=二等奖, P=Successful Participant=成功参赛奖.

每个参赛队的指导教师和队员都将获得由竞赛主任和每题的评阅组长签名的证书. 美国运筹学和管理科学学会(ORSA)给予两个获得特等奖的队队员现金奖励和三年的会员资格. 这两个队分别是宾夕法尼亚州的葛底斯堡(Gettysburg)学院队(B题)和密苏里州的 Washington 大学队(B题). ORSA 还给每个获一、二等奖的队一年的免费会员资格.

美国工业与应用数学学会(SIAM)对每题指定一个特等奖队作为 SIAM 的获奖队, 每个队员获得现金奖励, 这两个队分别是加州波莫纳(Pomona)学院队(A题)和纽约州的圣博纳旺蒂尔(St. Bonaventure)大学队(B题). 他们将于 1996 年 7 月在密苏里州的堪萨斯(Kansas)城举行的 SIAM 年会特设的分组会上作报告.

美国数学协会(MAA)指定一个特等奖队作为 MAA 的获奖队. 该队是加州波莫纳(Pomona)学院队(A题).

§ 6.2 问题和背景

6.2.1 问题

本章讨论 1996 年美国大学生数学建模竞赛 A 题——潜艇探测问题. 题目全文如下:

世界上的海洋都有一个环绕噪声场. 地震扰动、海面船舶航行和海洋动物都是这一噪声场的根源, 它们分别为噪声场提供了

不同频率范围的声源。我们希望考虑如何用这一环绕噪声场去探测位于海面下的大的移动的物体，比如潜艇。假设，潜艇本身不发出噪声，试研究一种仅仅用测量环绕噪声场的变化获得的信息来探测运动潜艇的出现，它的速度、它的尺寸和它的前进方向的方法。可先从一种固定频率和振幅的噪声着手。

6.2.2 问题的背景

水下声学有悠久的历史，早在古希腊，学者们就以听水下鱼声为乐；中国古代人们将一竹筒一端插入水中，在另一端听水下的声音用来发现浅水底的鱼群。这一方法到 1490 年有了进一步的发展，Leonardo 和 Vinci 用一根管子听水下的声音可以发现远处航行的船只，然而直到 1826 年，Daniel Colladon 和 Charles Sturm 测量了水下声速，水下声学才有了第一个定量研究的结果。令人惊讶的是，他们得到的水下声速和我们现在公认的水下声速的误差还不到 3%。

水下声学探测方法的显著进展发生在 20 世纪，两次世界大战中的潜艇和反潜艇战斗是水下声学探测技术发展的动力。在反潜艇战中，首先要发现潜艇。探测潜艇可以有多种方法，但是由于海洋本质上对声音信号是透明的而对其他信号有严重的阻碍作用，所以用水下声学探测的方法是最好的。

用水下声学方法探测潜艇最早采用被动法，即简单地侦听目标潜艇发出的声音。第一次世界大战期间主要就是采用这种方法来探测潜艇的。随着潜艇技术的发展，潜艇发出的噪声水平大大降低，这种被动探测的效率大大降低。

针对这种情况，在第一次世界大战末期，英国和美国都致力于研究主动探测方法。他们将声脉冲发射到水下，用测量捕捉从存在的目标反射回来的信号探测潜艇。但因为这种方法出现得太晚，来不及在第一次世界大战中发挥举足轻重的作用。主动探测技术的主要代表是声纳系统，声纳系统在二次世界大战之间得到

了充分的发展，在第二次世界大战中成为从水上船只或水下探测潜艇的主要手段。然而主动探测也有其弊病，在探测过程中因为不断主动发射信号，探测者很容易被发现并遭到攻击。

水下声学探测不仅可以用于军事目的，而且还有大量其他应用，如：海底勘查，海底石油和天然气工业，鱼群定位和监测以及水下生物的保护等。在这些用途中，有时可以用被动或主动探测技术，但是有时，这两种技术都不适用。比如，当目标发出声音很小或根本不发声音时，被动探测技术无法发挥作用；而当探测对象（如海豚和鲸鱼）会被发射的信号干扰时，主动探测技术也无法应用。

在这样的情况下，近年出现了不同于主动或被动探测的新方法。这一方法基于海洋的环绕噪声场在某种程度类似于光场，因此利用目标对噪声的反射和散射引起噪声场的扰动可以探测海洋水下目标的想法。这一想法是由加利福尼亚大学海洋学院水下物理实验室的 Buckingham 首先提出来的。他认为，既然噪声像大气中的日光一样有随机散射场和有分量向各个方向传播的特性，就有可能像利用日光照相一样，利用环绕噪声场来形成目标物的形象。这种技术称为水下成像术。最近几年这一技术的发展取得了一定的成功。利用某种“声学透镜”最近已经成功地得到了一个大小约 40 m 的目标的几何形状的分辨率为 126 个像素的像。无论在实验上还是在理论上，计算机模拟和计算机软件方面都得到了不小的进展，出现了若干重要文献（如 [1]，[2]，[3]）。Buckingham 认为目前水下声学成像技术就像电视技术处于图像模糊的早期阶段一样，虽然远谈不上清晰，图像至少已经存在。

既然水下声学成像技术是一种非常有用并且很可能成功的技术，又是一种目前尚很不完善的技术，其中还有许多问题需要解决。命题人利用 Buckingham 的研究提供的素材加工成本题，期望参赛学生发挥创造精神，在竞赛过程中获得锻炼的同时为水下声学成像技术提供新鲜的思想。

6.2.3 竞赛结果

在参赛的 393 个队中有 126 个队选择了本题，结果有 4 个队获得了特等奖 (Outstanding)，16 个队获得一等奖 (Meritorious)，37 个队获得二等奖 (Honorable Mention)。获得特等奖的 4 个队分别来自 Pomona 学院，北卡罗来纳大学，Wake Forest 大学和 Worcester 综合工学院。我国北京工学院、重庆大学、南开大学、东南大学(2 队)、清华大学、中国科技大学、湘潭大学和中山大学共 9 个队获得一等奖。

6.2.4 本章其余各节的安排

在 § 6.3 和 § 6.4 中，我们分别介绍 Worcester 综合工学院和 Wake Forest 大学二队的优秀论文；在 § 6.5 中我们对其余二篇优秀论文作一概述；在 § 6.6 中对几篇优秀论文作简单的评述，并讨论此题的命题和阅卷给我们的启示。

在介绍获奖的优秀论文时，我们略去了一些枝节的内容，并尽量使模型和方法更加清晰化。

§ 6.3 用环绕噪声使水下目标成像

本节介绍 Worcester 综合工学院优秀获奖论文。该文利用抛物面反射器接收目标对环绕噪声场的反射信号，聚焦后判别目标的存在并决定其速度和目标的尺寸。

6.3.1 基本假设

对测量仪器和外界环境文章作了以下基本假设：

(1) 假设目标是一个几乎完全反射的物体，即潜艇几乎不吸收声音。

由于目标几乎是一个完全反射体，只需从海水中可以传播的

频率范围内选定一种单一的频率进行监测就可以了。这一范围界于 5 至 50 kHz 之间^[2]。其理由是完全反射体对各种频率都产生同样的特征响应^[4]，因此采用多种频率就会产生信息冗余。所以采用能够较强反映环绕噪声场的单一频率更为合适。

（2）假设环绕噪声场有方向性。如果没有环绕噪声场在某些方向的偏差，就不能探测出背景噪声和目标反射之间的差别。实验研究已经证明这样的偏差是存在的^[5]。

（3）假设环绕噪声场在进行搜索的过程中相对保持稳定。在很短的时间区间内，噪声场频繁的变化使发现目标变得十分困难。

（4）假设下述两个条件至少有一个满足：

- a. 搜索的目标是运动的；
- b. 在目标进入搜索范围之前，在相同的海洋条件下进行过背景搜索。

（5）假设目标的最小尺寸为 10 m，这是美国潜艇的平均宽度，其他国家的潜艇也差不多。

（6）测量仪器由多个抛物反射器构成，它们通常用来将微弱的信号聚焦到焦点。在每个抛物反射器的焦点，安置一个水下听声器（hydrophone），一种测量水下声强度的装置。而且设仪器控制角度的精度可以达到 0.1°。

6.3.2 二维情形目标探测的数学模型与方法

为使问题更加直观，首先考察一种二维的特殊情形：假设用两台抛物面反射器，这两台抛物面反射器均靠近海岸且位于水下的同一深度；又设目标也位于水下同一深度，但不与两台反射器共线（参见图 6-1）。

测量时让两台抛物面反射器面向海洋由外向里相向旋转 180°，同时监测记录某一固定的频率。若对任意角度背景噪声为常量，噪声-角度曲线就是一条水平直线。但是由于环绕噪声场

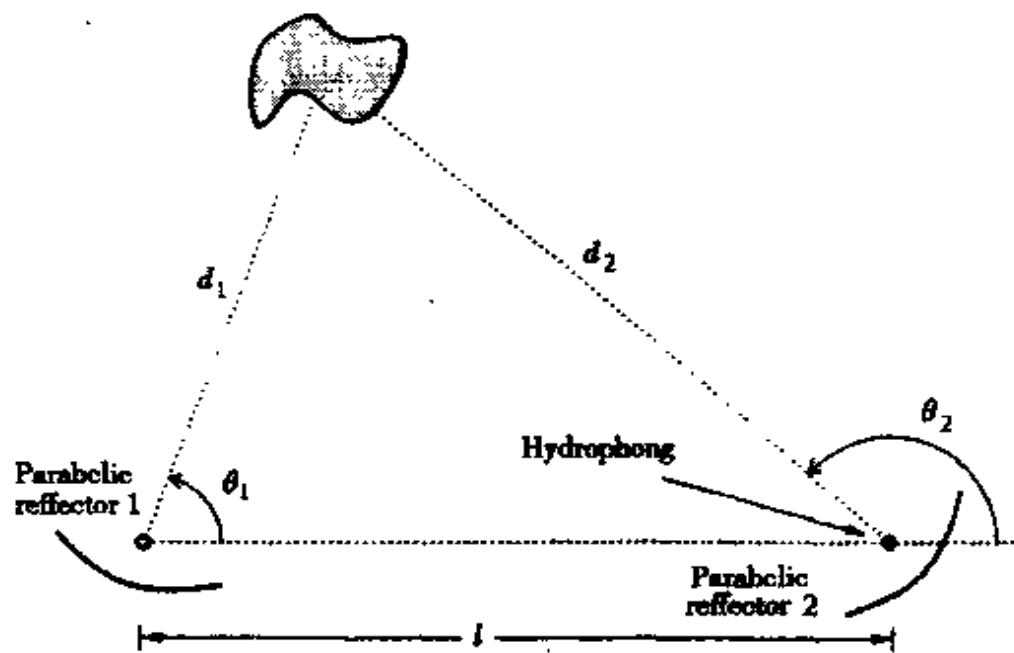


图 6-1 二维点的位置

的方向偏差，因而噪声是反射器偏转角度的函数。所以进行背景监测是有必要的，所谓背景监测就是在无目标出现时，用两台抛物面反射器旋转监测同一固定频率的噪声，用以和以后的探测进行比较。

在进行目标探测时，当抛物面反射器面对目标时就会接收到目标反射的噪声，经聚焦后这一频率的噪声强度会有明显增加。特别是将目标探测值减去背景监测值后，目标存在的特征可以看得更加明显。这一扰动的中心对应的角度可以近似为目标中心的视角。根据两个抛物面反射器决定的两个视角 θ_1 和 θ_2 以及两台反射器之间的距离 l 就可以决定目标的位置。

事实上，设 d_i 为反射器 i 至目标的距离， $i=1, 2$ ，由正弦定理

$$\frac{d_1}{\sin(\pi - \theta_2)} = \frac{d_2}{\sin \theta_1} = \frac{l}{\sin[\pi - \theta_1 - (\pi - \theta_2)]},$$

即

$$\frac{d_1}{\sin \theta_2} = \frac{d_2}{\sin \theta_1} = \frac{l}{\sin(\theta_2 - \theta_1)}.$$

解得

$$d_1 = \frac{l \sin \theta_2}{\sin(\theta_2 - \theta_1)}, d_2 = \frac{l \sin \theta_1}{\sin(\theta_2 - \theta_1)}.$$

将坐标原点取在反射器 1 的位置，取 x 轴为从反射器 1 指向反射器 2 的向量建立坐标系，那么，目标中心位置的坐标为 $(d_1 \cos \theta_1, d_1 \sin \theta_1)$.

噪声强度图上受目标扰动的模式不仅提供了目标位置视角的信息，还提供了目标大小的信息。目标越大，受扰动的范围越宽。根据目标的距离和受到扰动的角的幅度就可近似地算出目标的尺寸。

在很短的时间区间内连续两次测量目标的位置就可以计算出目标的速度。

例如在岸边水下相距 100 m 同一深度放置两台抛物反射器，在同一深度有一目标。第一次探测在噪声强度图上 $\theta_1 = 84.3^\circ$ 和 $\theta_2 = 91.9^\circ$ 处为峰值，用前述方法可计算出目标位置：

$$x = \frac{l \sin \theta_2 \cos \theta_1}{\sin(\theta_2 - \theta_1)} = \frac{100 \sin 91.9^\circ \cos 84.3^\circ}{\sin 7.6^\circ} \approx 75.1 \text{ m},$$

$$y = \frac{l \sin \theta_1 \sin \theta_2}{\sin(\theta_2 - \theta_1)} = \frac{100 \sin 84.3^\circ \sin 91.9^\circ}{\sin 7.6^\circ} \approx 752.0 \text{ m}.$$

5 秒钟后再探测一次，发现目标的视角变为 $\theta_1 = 84.5^\circ$, $\theta_2 = 92.4^\circ$ ，再次计算得目标新位置：

$$\bar{x} \approx 69.7 \text{ m}, \bar{y} \approx 723.6 \text{ m}.$$

将两位置向量相减，得运动方向

$$\text{方向} \approx (69.7 - 75.1, 723.6 - 752.0) = (-5.4, -28.4),$$

其速度为

$$v = \sqrt{(-5.4)^2 + (-28.4)^2} / 5 \approx 5.8 \text{ m/s}.$$

当目标与两个反射器共线时，必须采用第三个反射器，这第三个反射器当然不能置于前两个反射器的连线上。

在无法进行背景监测或背景噪声发生较大改变时目标探测仍

然可以进行。但这时我们记录下来的噪声强度曲线是目标噪声反射和环绕噪声场的迭加。为了断定噪声强度图上的扰动是真正由目标引起的扰动还是由噪声场的随机涨落引起的变化，可以在较短时间内探测二次，若扰动的模式相同且有微小偏移即可断定为目标反射引起的扰动。

6.3.3 目标引起的扰动的模拟

目标探测的二维模型的基础是目标对环绕噪声场的反射会引起接收噪声强度曲线的扰动。在一时无法做实验的情况下，用模拟和分析的方法也可加以验证。

设用一个抛物反射器探测一个平面线段目标面且反射器恰好位于该线段中点的垂线上。设反射器到目标中点的距离 $r = 19.7$ m，目标宽度 $x = 6.74$ m，反射器直径 $w = 5$ m（见图 6-2）。

目标的边缘至反射器中心的距离

$$d = \sqrt{\left(\frac{x}{2}\right)^2 + r^2} = 20 \text{ m.}$$

当反射器直接对准目标的左端时反射器中心与目标左端连线与水平线的夹角

$$\varphi = 90^\circ - \arcsin\left(\frac{x}{2d}\right) = 80^\circ,$$

同样，反射器对准目标的右端时，反射器中心与目标右端连线与水平线的夹角

$$\sigma = 90^\circ + \arcsin\left(\frac{x}{2d}\right) = 100^\circ.$$

为了直观地考察信号从目标的反射，引入如下环绕噪声场

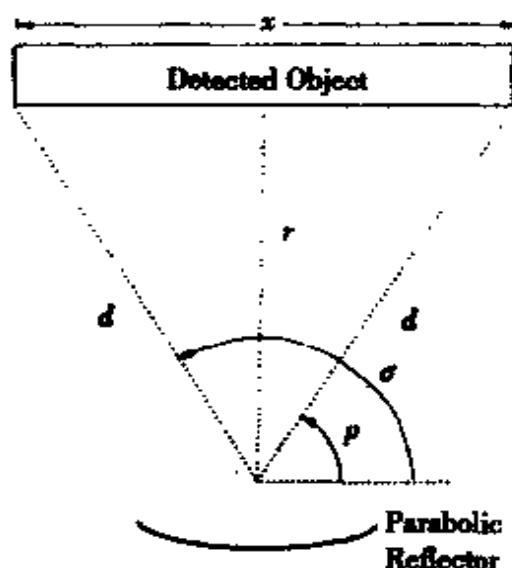


图 6-2 目标与反射器的关系

$$I(\lambda) = -2\sin\lambda + 90,$$

其中 λ 为监测角, $I(\lambda)$ 为无目标时的噪声强度。这类似于照镜子时，太阳从背后照射的效果。图 6-3 是噪声强度的函数图像。

当目标进入探测范围时，在反射器对准目标之前就接收到反射噪声。图 6-4 中最大的区域 Δ 表示有某些反射噪声被接收到的区域。在目标的两端都有一个小小的过渡区域。在这些过渡区域中，并不全部接收到目标的反射信号，其中的一部分（图中的 E）主要接收到目标的反射信号。在区域 Δ 的内部，完全接收到目标反射的噪声，而在区域 Δ 的外部，只接收到背景噪声。



图 6-3 强度 I 与扫描角 λ
(单位：度) 的关系

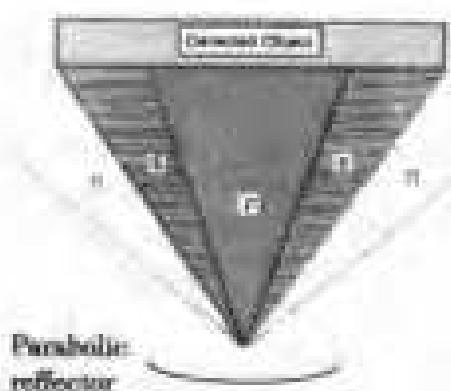


图 6-4 反射区域

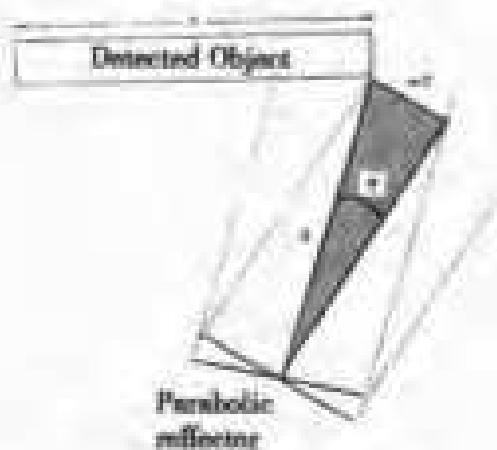


图 6-5 角 Ψ 的定义

左右两侧过渡区域可表示为 $(\sigma - \psi, \sigma + \psi)$ 和 $(\varphi - \Psi, \varphi + \Psi)$ 。参见图 6-5，其中

$$\Psi = \arcsin\left(\frac{w}{2d}\right).$$

由于假设目标是完全反射体，又由于抛物反射面的性质，它只将平行于它轴线的平行方向传播的噪声信号进行聚焦。因此在

区域 E 内噪声强度正比于 $2\sin\lambda + 90$, 从而在区域 $(\varphi + \Psi, \sigma - \Psi)$ 中, 噪声强度扰动呈正弦势态, 当 $\lambda = 90^\circ$ 时达到最大值。设在 $(\varphi + \Psi, \sigma - \Psi)$ 中接收到的噪声强度为 $A(\sin\lambda + 45)$, 注意到在 $(\varphi - \Psi, \sigma + \Psi)$ 之外, 只接收到背景噪声, 即噪声强度为 $-2\sin\lambda + 90$ 。由于噪声强度是连续变化的, 因此在过渡区域中可以利用区域 Γ 的边界和 Ω 的边界值来进行插值。例如在右边的过渡区域 $(\varphi - \Psi, \varphi + \Psi)$ 中, 噪声强度可表示为

$$I(\lambda) = A(\sin(\varphi + \Psi) + 45) \left[\frac{-2\sin(\varphi - \Psi) + 90}{A(\sin(\varphi + \Psi) + 45)} \right]^{\frac{(\lambda - \varphi - \Psi)^2}{4\Psi^2}}.$$

若忽略海水引起的衰减, 上式可以改写为

$$I(\lambda) = (2\sin(\varphi + \Psi) + 90) \left[\frac{-2\sin(\varphi - \Psi) + 90}{2\sin(\varphi + \Psi) + 90} \right]^{\frac{(\lambda - \varphi - \Psi)^2}{4\Psi^2}}.$$

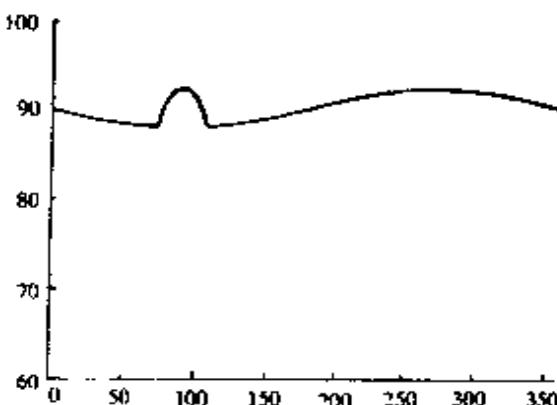


图 6-6 总响应与角 λ (单位: 度) 的关系

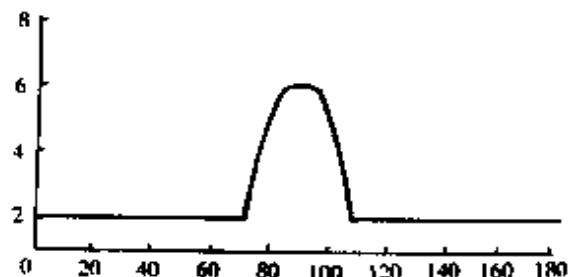


图 6-7 差图

图 6-6 是用前述具体数据画出的接收到的噪声强度图。若将此噪声强度函数减去背景噪声强度, 由目标反射引起的扰动及其模式就揭示得更加清楚了(参见图 6-7)。

6.3.4 三维情形目标的探测

一般三维情形, 探测目标比二维情形花费的时间多得多(见图 6-8), 每个反射器要作全方位的旋转, 目标的位置可以从两个已知点与它的中心的连线的角度来决定。每个反射器的方向用

两个角度通过球面坐标或用向量表示. 当目标与两个反射器共线时就使用第三个反射器.

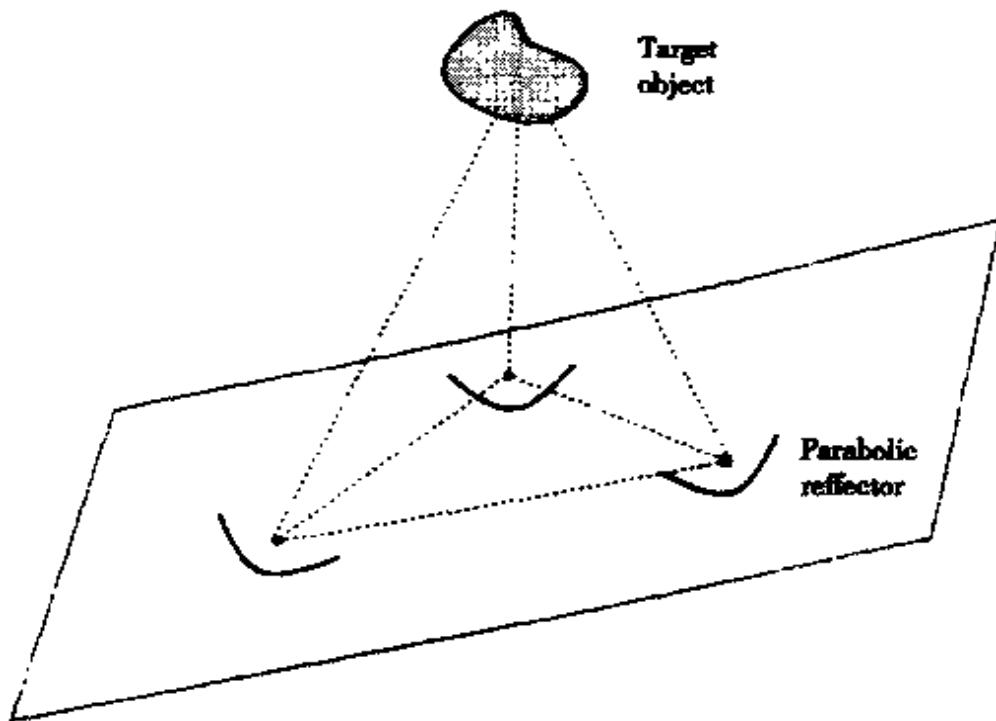


图 6-8 三维目标位置

设两个反射器已经捕捉到目标, 两个反射器中心指向目标中心的方向向量分别为 $\vec{V}_1 = (a_1, b_1, c_1)$, $\vec{V}_2 = (a_2, b_2, c_2)$. 取第一个反射器中心为原点, 它和第二反射器的连线为 x 轴建立三维直角坐标系, 并设这两个反射器的距离为 l .

目标的位置可由具有方向 \vec{V}_1 和 \vec{V}_2 的两条直线的交点给出. 这两条直线的参数方程为

$$\begin{cases} x = sa_1, \\ y = sb_1, \\ z = sc_1, \end{cases}$$

$$\begin{cases} x = l + ta_2, \\ y = tb_2, \\ z = tc_2. \end{cases}$$

它们的交点是方程组

$$\begin{cases} sa_1 = t + ta_2, \\ sb_1 = tb_2, \\ sc_1 = tc_2 \end{cases}$$

的解。由此得

$$t = \frac{b_1}{a_1 b_2 - b_1 a_2}.$$

目标中心位置的坐标为 (sa_1, sb_1, sc_1) 。

为了改进三维目标检测的效果，可以采用多个抛物面反射器构成的线性阵列（图 6-9），抛物面槽和抛物面环（图 6-10(a) 和图 6-10(b)）。

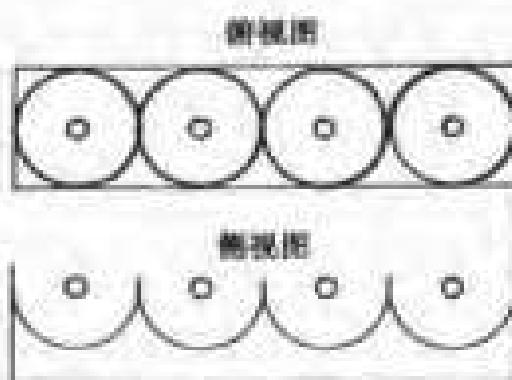


图 6-9 线性阵列结构



(a) 抛物面槽



(b) 抛物面环

图 6-10 其他建议的反射器

6.3.5 优点和局限

本模型所作的假设，有的是必需的，有的虽然可以放松或去掉，但情况会发生改变。由于设备和环境的原因，本模型与方法有其本质的局限性。

(1) 目标是几乎完全反射体的假设可以放松。具有很强声音

吸收性的目标与背景相比会产生噪声强度的负增益区，这与几乎完全反射体产生的正增益区是不同的。若允许依赖频率的反射性，我们可以通过监测一定频率范围来决定目标的声学“色彩”。

(2) 若没有环绕噪声场有方向性的假设，除非对有很强声吸收性的目标，探测变得十分困难。吸收在背景环绕噪声场中总是看起来像一个“洞”。由于具有强声吸收性的目标比较少，建议在实施时还是保留这一假设。

(3) 若对运动目标采用多次监测方法，背景监测可以不要；若可以得到可信的背景监测，可以用来验证探测是否正确。为探测静止的目标，可靠的背景监测仍然是必需的。

(4) 由于接近海面环绕噪声场是很不稳定的，因此静止的目标很难探测。但在海底，环绕噪声场主要由有方向的地震引起的，与海面情况相反它是非常稳定的，通常在一个季节中保持不变。

(5) 对建议的几种设备，不能确定在实际中哪一种是最好的，因此要进行实验。

(6) 为了精确地发现目标，对跟踪设备控制角度的精度要求是比较苛刻的。只要略微放松一点要求，在定位时就会产生较大的误差，特别是目标距离较远时，误差更大。

(7) 另一局限性是由于海水对声能的吸收引起的，能量的衰减，严重地缩短了我们的探测距离。若使用 40 kHz 频率，探测距离只能达到 1 km，这是令人失望的。

(8) 由下一节可以看到探测距离受到抛物面反射器直径的限制。目标越远要求反射器的直径越大。但在实际应用中，过大的反射器是不切实际的。

6.3.6 附注：抛物面反射器直径与探测距离

给定了反射器的直径 w ，运用 Rayleigh 准则⁽⁴⁾，若声音的波长为 λ ，反射器可以分辨的最小角度为

$$\theta = \frac{1.22\lambda}{w},$$

文献[4]指出

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

其中 μ 为声波的频率, c 是某个依赖于海水温度的常数. 例如当水温为 25°C 时, $c=1531$ m/s. 此时,

$$\theta = \frac{1.22c}{w\mu}.$$

设目标宽度为 x , 与反射器的距离为 r . 目标两端与反射器中心连线的张角为 φ , 设 φ 较小, 成立

$$\frac{\varphi}{2} \approx \tan\left(\frac{\varphi}{2}\right) = \frac{x}{2r}.$$

要分辨出目标, 必须成立 $\varphi \geq \theta$, 即

$$\frac{x}{r} \geq \frac{1.22c}{w\mu} \text{ 或 } w \geq \frac{1.22cr}{x\mu}.$$

设目标宽度为 10 m, 频率为 40 kHz, 水温为 25°C, 探测距离和至少需要的反射器直径的关系由下表所示. 不难发现若需探测 20 km 远处的一艘潜艇, 反射器本身就要像潜艇一样大, 这显然是不现实的. 附表列出了目标的距离和反射器直径的关系.

距离 r (km)	1	2	5	10	20
直径 w (m)	5	9	23	47	93

§ 6.4 由环绕噪声场的涨落探测无声潜艇

本节介绍 Wake Forest 大学队的优秀获奖论文. 该队建立的模型基于用四个监听站, 每个监听站由隔开一定距离的四个麦克风构成. 通过用 Fourier 分析计算每个麦克风接收的振幅谱, 并与先前测量的环绕噪声场基准谱进行比较, 两者谱之差是潜艇反

射的噪声。

用位于特定位置的四个麦克风测定潜艇噪声谱振幅峰值的梯度。由于振幅反变于至潜艇的距离，因此可以用每个监听站的振幅和梯度计算出潜艇的位置。假设潜艇为球状的，用距离和峰值振幅可计算它的大小即半径。利用Doppler效应可计算潜艇的速度。

6.4.1 基本假设

(1) 设在海洋中音速是常数。虽然音速是随温度变化的，但因所用设备所处的范围足够小，音速的微小变化可以忽略。

(2) 环绕噪声在处处都有同样的频率和振幅，因此不必考虑海面和海底的反射。

(3) 假设潜艇的形状均近似于球，由钢制成并将射向它的声能的百分比为 k 的能量反射出来。虽然潜艇的表面，作为三维物体的二维弯曲表面不是一个简单的谐振子(SHO)，但SHO是一种合理的类比。此时SHO既是强迫的(由于环绕噪声)又是阻尼的(由于水的作用和钢铁的柔韧性)，对强迫和阻尼SHO，稳定状态解是具有强迫频率的振动。高阻尼系数意味着潜艇的响应与频率无关。又因为考虑的距离大大超过潜艇的大小，潜艇的反射波可视作球面波。

(4) 海底和海面对潜艇反射波的再次反射强度已经可以忽略，即麦克风接收到的非环绕场噪声只有来自潜艇的直接反射。

(5) 设海洋是均匀的和一致的。因此设没有大的动物和其他目标在潜艇旁边，这些东西会严重影响声波的传播。而且还假设在任何时刻只有一艘潜艇出现。

(6) 假设潜艇不发出噪声，且它的运动不会产生影响声波传播的湍流。

(7) 设潜艇运动的速度不会超过20m/s。

(8) 设没有明显的水流流动，监听站相对于海水是静止的。

(9) 首先假设环绕噪声场只有一种频率，然后推广至有多种频率的情形。

6.4.2 模型的叙述

测量系统由四束麦克风构成。取直角坐标系，使 $x-y$ 平面平行于海平面，取 z 轴向上。将四束麦克风置于原点和点 $(d, 0, 0)$, $(0, d, 0)$ 和 $(0, 0, d)$ 处（见图 6-11）。

由于潜艇一般不潜到 1500m 以下，可将坐标系的原点置于水下 1000m 处，所以其中一束麦克风位于水下 1000dm 深处，其余三束位于水下 1000m 深处。令 $d = 500m$ ，探测麦克风束分布在潜艇可能出现的深度范围内又不至于与海面的船只或锚碰撞。

每一束麦克风（监听站）又由四个麦克风组成，其中一个恰好位于上述麦克风束的位置，其余三个离开它一小段距离 δ ，且每一个都在一个不同的坐标方向上（见图 6-12）。

首先测量环绕噪声场在不出现潜艇时的波形，这样就可以决定环绕噪声场的频率和相应的振幅（开始只有一种频率出现）。然后短时间测量每一个麦克风所在位置的声音，进行 Fourier 分析，得到波的模式来决定出现的频率和相应的振幅，用这些数据来算出潜艇的位置、速度、前进方向和大小。

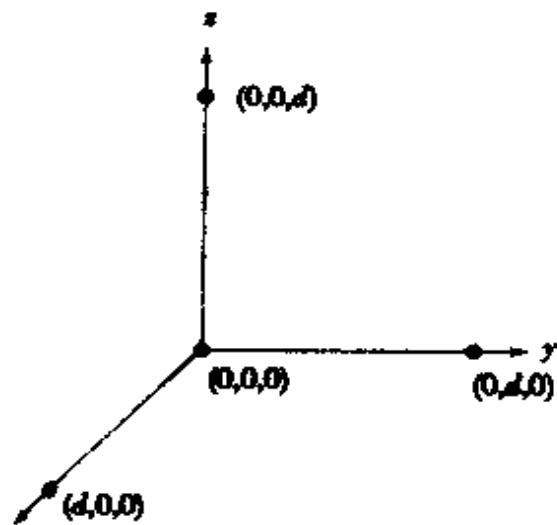


图 6-11 麦克风束阵列

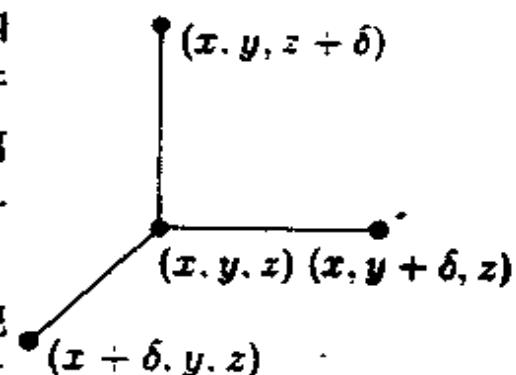


图 6-12 监听站麦克风的配置

6.4.3 计算需要的数据

由于潜艇被处理成球体，决定它的大小即为决定其半径 R ，决定潜艇的位置即决定球心的位置。因此问题完整的解应该包括半径 R ，位置坐标 (x, y, z) 和速度向量 $\vec{v} = (v_x, v_y, v_z)$ 。

可以用来计算这些量的数据由麦克风阵列接收到的频率和相应的振幅。下面列出计算需要的常数和变量：

f =环绕噪声的频率，在此选择 1000Hz 作为单频率模拟的数据，它位于实际海洋环绕噪声的频率范围内。

I_0 =环绕噪声的强度。1000Hz 频率的合理的强度是 $5.4457 \times 10^{-10} \text{ Pa}^2$ ⁽⁶⁾。

A_0 =噪声的振幅。振幅的平方正比于声强度。由于比例常数已经在计算 I_0 时加以考虑，有 $A_0 = \sqrt{I_0} = 2.3336 \times 10^{-5} \text{ Pa}$ 。

k =潜艇表面反射的声能（强度）所占的百分比，取 $k=0.86$ 。由于振幅正比于强度的平方根，从潜艇表面直接反射后的声波振幅为 $\sqrt{k} A_0 = 0.9274 A_0$ 。

$A(s)$ =从潜艇表面反射的声波到距潜艇中心 s 处的振幅。由能量守恒，注意到反射波可近似为球面波，因此有

$$4\pi R^2 k A_0^2 = 4\pi s^2 A^2(s),$$

由此得

$$A(s) = \frac{\sqrt{k} A_0 R}{s}.$$

6.4.4 探测潜艇是否存在

用每个麦克风接收到的原始数据是一小段时间的波形，图 6-13 是一个多频率噪声图的例子。

为将其转化为有用的频率振幅图，可采用快速 Fourier 变换。计算后得到对应于频率 f 的振幅值 A_{ij} ，其中 i 是监听站的

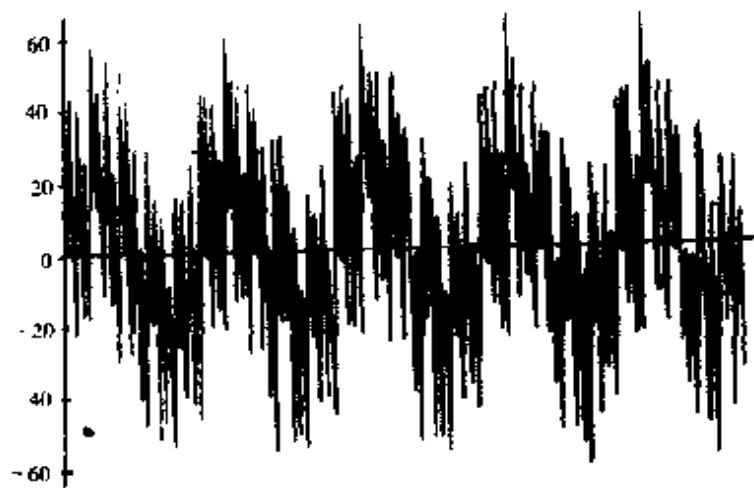


图 6-13 具有 5 种频率的环绕噪声场。
压力为时间的函数

序号， j 是在监听站内麦克风的序号。

判别是否出现潜艇的算法是这样的，测量计算好无潜艇时的振幅谱 \bar{A}_{ij} ，再将现在测量计算得到的振幅谱 A_{ij} ，减去 \bar{A}_{ij} 得

$$\tilde{A}_{ij} = A_{ij} - \bar{A}_{ij}.$$

若 $\tilde{A}_{ij} \approx 0$ ($i, j = 1, 4$)，目前测得的噪声只是均匀的环绕噪声场的噪声，即没有出现潜艇，可进行下一次探测。

若上述差中存在不接近 0 的，就要设法判别它是由环绕噪声场改变引起的，还是由潜艇出现引起的。如果谱的变化是由海洋环绕噪声场的改变引起的，所有的麦克风应记录下同样的数据。但如果谱的改变是由潜艇的出现而引起的，每个监听站乃至每个麦克风记录下来的数据均有差别。所以比较每个监听站第一个麦克风的振幅谱差 \tilde{A}_{1j} ，若无显著差别则断言改变是由环绕噪声场本身引起的，并无潜艇出现。但此时应该将这次探测计算得到的振幅谱作为新的环绕噪声场的基准谱。

如果各监听站的振幅谱差 \tilde{A}_{1j} 之间有明显的差异，这意味着有潜艇出现了。此时，找出每个麦克风具有最大振幅的频率。由于一切频率的振幅均有相同的反射比例 \sqrt{k} ，此峰值振幅和相应的频率必然体现了环绕噪声场的峰值振幅对应的频率的噪声的反

射。在以后的算法中可以考虑由多频率或单频率构成的环绕噪声场，但只考虑其中的具有峰值振幅的频率的噪声（见图 6-14）。

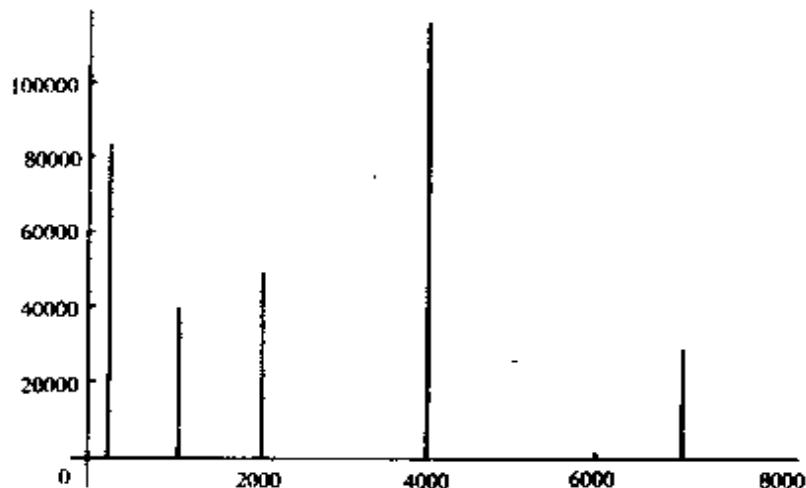


图 6-14 环境噪声的振幅谱

6.4.5 潜艇的位置和大小

利用每个监听站中四个麦克风位置的差异可近似求得反射噪声振幅的梯度，再利用反射波是近似的球面波这一性质，可以求出潜艇距该监听站的距离。

令

$$A(x, y, z) = A(s)$$

表示潜艇反射振幅函数，其中 s 表示点 (x, y, z) 到潜艇中心的距离。设第 i 个监听站的位置是 (x_0, y_0, z_0) ，则

$$\frac{\partial A}{\partial x}(x_0, y_0, z_0) \approx \frac{A(x_0 + \delta, y_0, z_0) - A(x_0, y_0, z_0)}{\delta},$$

而 $A(x_0 + \delta, y_0, z_0)$ 和 $A(x_0, y_0, z_0)$ 即为该监听站两个不同麦克风分析记录的振幅差峰值。设位于 $(x_0 + \delta, y_0, z_0)$ 的麦克风为 2 号麦克风，则

$$\frac{\partial A}{\partial x}(x_0, y_0, z_0) \approx \frac{\tilde{A}_2 - \tilde{A}_1}{\delta}.$$

类似可以求出 $\frac{\partial A}{\partial y}$ 和 $\frac{\partial A}{\partial z}$ 从而得到 $\nabla A(x_0, y_0, z_0)$ 。

设潜艇的位置为 (a, b, c) , 则 $\vec{s} = (x-a, y-b, z-c)$,

$$s = \sqrt{(x-a)^2 + (y-b)^2 + (z-c)^2}.$$

由

$$\frac{\partial A}{\partial x} = \frac{\partial A}{\partial s} \cdot \frac{\partial s}{\partial x} = \frac{\partial A}{\partial s} \cdot \frac{x-a}{s},$$

因此

$$\nabla A = \frac{\partial A}{\partial s} \cdot \frac{\vec{s}}{s}.$$

已知

$$A(s) = \frac{\sqrt{k}A_0 R}{s},$$

可得

$$\frac{\partial A}{\partial s} = -\frac{\sqrt{k}A_0 R}{s^2},$$

从而

$$|\nabla A| = \frac{\sqrt{k}A_0 R}{s^2}.$$

对比 $A(s)$ 和 $|\nabla A|$ 的表达式可知

$$s = A / |\nabla A|.$$

由 \vec{s} 方向与 ∇A 方向相反

$$\vec{s} = -s \frac{\nabla A}{|\nabla A|} = -\frac{A \nabla A}{|\nabla A|^2},$$

即

$$(a, b, c) = (x_0, y_0, z_0) + \frac{A \nabla A}{|\nabla A|}.$$

对四个监听站作同样的计算, 然后取平均值即可.

由 $A(s)$ 的表达式中解出 R 得

$$R = \frac{A \cdot s}{\sqrt{k}A_0}.$$

对四个监听站计算后取平均值可得到 R 的近似.

6.4.6 潜艇的速度

由于环绕噪声场的频率经过潜艇反射后会有改变，可以利用 Doppler 效应来计算潜艇的前进速度。一般的声 Doppler 效应为

$$f_0 = f_s \left(\frac{c - v_0}{c + v_s} \right),$$

其中 f_0 是观察者接收到的频率， f_s 是源的频率（这里是环绕噪声场的频率）， v_0 和 v_s 分别是观察者和源的速度在二者连线上的投影， c 是音速。由于监听站是静止的，因此 $v_0 = 0$ 。令潜艇远离监听站时的运动速度为正，解出

$$v_s = \left(\frac{f_s}{f_0} - 1 \right) c.$$

对第 i 个监听站，利用以上公式可求出潜艇速度在监听站和潜艇中心连线上的投影 v_i ，同时，我们已经得到了该连线的方向。设该监听站第 1 麦克风的坐标为 (x_i, y_i, z_i) ，则该方向单位向量 \vec{u}_i 为

$$\vec{u}_i = \frac{\nabla A}{|\nabla A|}(x_i, y_i, z_i).$$

记 $\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4$ 中任意三个不共面向量为 $\vec{u}_i, \vec{u}_j, \vec{u}_k$ ，在其上的速度分量为 v_i, v_j, v_k ，则潜艇的速度公式为

$$\vec{u} = v_i \vec{u}_i + v_j \vec{u}_j + v_k \vec{u}_k.$$

如果四个方向向量均不共面，应选由三个方向向量构成的矩阵有最大行列式的方向来进行计算。

6.4.7 模型的推广

虽然模拟仅对只有一种频率的环绕噪声场进行，但由于采用的算法是挑选出峰值振幅及其对应的频率来处理，从而这一模型也适用于多频率的情形。

所用的算法有根据变化调整环绕噪声场的功能。然而如果潜艇引起振幅谱的改变与环绕噪声场自身的显著改变恰好一致，探测就会出现错误。

由于计算速度很快，这一方法还可以用来在一段时间内跟踪一艘特定的潜艇。

在同一区域内出现两艘潜艇对模型和算法提出了很多复杂的问题，对环绕噪声场产生较小影响的潜艇难以被探测。但可以修改算法，计算出探测到的潜艇对环绕噪声场的影响。扣除这一影响后，再来分析另一艘潜艇是否存在。

模型假设中关于水的流动和监听站静止的假设可以放松。Doppler 效应的一般公式已经提供了观察者运动时的处理方法，因此监听站的移动是易于处理的。类似地，不变的环流可以综合到 Doppler 方程中加以考虑，只是使计算变得略微复杂一些。

6.4.8 模拟结果

为了模拟算法的实现，编制了一个 Fortran 语言程序。该程序仅仅用以下数据： k ；所有麦克风的位置；表示环绕噪声场的波形和对每一个麦克风，一个表示有潜艇出现时的波形。

要运行这一模拟程序，还得产生供麦克风接收的声数据。为此另外用一个 Fortran 程序产生一个每秒有 8000 样本的离散波形数据。先产生一组只有环绕噪声出现的数据作为参照，然后再加上由特定大小的位于特定位置和用特定速度前进的潜艇引起的附加噪声。数据生成的方式十分便于对模拟程序的精度进行校核。

先生成具有一个固定频率和振幅的环绕噪声数据，然后对潜艇半径、位置和速度的三种不同的选择生成数据文件。对这三组数据运行模拟程序，模拟结果如表 6-1 所示。

这些结果表明在确定潜艇的位置和速度是很成功的，虽然有少许误差出现。但是在计算潜艇的半径时，模拟结果并不好。但

是三个模拟结果的误差的百分比是相当一致的。这说明此误差有可能是系统的误差，是可纠正的。这一误差可能是 Fourier 分析计算过程中产生的。

表 6-1 一种频率、三种数据下潜艇大小、距离和速度的输入和输出

Simulation	R(m)	x(m)	y(m)	z(m)	v_x (m/s)	v_y (m/s)	v_z (m/s)
1	Input	13.8	-3000	-2000	300	15	10
	Output	9.73	-2765	-1843	281	14.9	10.5
2	Input	5	200	1000	-500	2	10
	Output	4.02	190	953	-464	1.3	9.5
3	Input	10	2000	2000	-500	5	5
	Output	7.95	1871	1871	-468	4.8	4.8

尽管这一模拟结果是鼓舞人心的，但必须指出，这一计算程序仅仅是这一求出潜艇位置、大小和速度的一般数学算法的一种理想化。若使用更加完全和精确的噪声数据或用更精确的快速 Fourier 变换算法，结果可能会得到改善。

对多频率的环绕噪声也进行了模拟，结果如表 6-2 所示。

表 6-2 误 差

Simulation	R	x-coordinate	y-coordinate	x^* -coordinate
1	.043	.034	.014	.023
2	.078	.009	.009	.023

6.4.9 误差和敏感性分析

表 6-2 给出了潜艇位置、大小和速度的真实数据和计算结果的差别。位置坐标的误差的主要原因来自算法对振幅和振幅导数的依赖，而它们显然不是非常精确的。一个原因是计算需要离散化数据，这人为地造成了部分的误差。第二个原因是计算 ∇A 的误差，导数用差分近似，但步长 δ 不可能取得太小，因为这需要

更加灵敏的麦克风，技术上实现有困难。由于计算位置是四次计算（对4个监听站的计算）的平均，对半径计算也是有误差的，去掉系统误差后的相对误差在表6-3中列出。正如已经指出的，该误差来自振幅计算的Fourier分析。从表6-1可以看出，当Fourier变换计算出的振幅偏小时，计算得到的半径也偏小。

引起速度计算的误差的主要因素有两个，一个是位置计算的误差，另一个是观察频率 f 的误差。当环绕噪声中低频占支配地位时，观察频率的误差就变得十分重要了。

还作了一些模拟考察系数 δ 的影响，其结果在表6-3中列出。由表中的数据可以看出，结果对参数 δ 是十分敏感的。

表6-3 δ 的影响

δ (m)	Simulation	R (m)	x (m)	y (m)	z (m)	$ \bar{v} $ (m/s)
5	1	-.195	-.041	-.046	-.068	-.058
	2	-.295	-.078	-.080	-.054	.007
10	1	-.192	-.049	-.046	-.070	-.058
	2	-.294	-.078	-.079	-.061	.007
15	1	-.197	-.057	-.047	-.072	-.057
	2	-.278	-.078	-.077	-.068	.007
20	1	-.198	-.067	-.047	-.075	-.056
	2	-.272	-.077	-.076	-.075	.007

§ 6.5 其他模型和方法

6.5.1 Pomona学院的方法

类似于Worcester综合技术学院队，Pomona学院队用水下听声器监测潜艇反射噪声对环绕噪声场的扰动，来探测潜艇的存在、位置、大小和前进的速度。但是他们用的设备是水下听声器有方向的阵列，使用另一组有方向的水下听声器阵列可以获得潜

艇的三维形象。他们的思路是最接近于环绕噪声场不仅可以用来探测而且可以用来成像的设想。

这个队的学生在竞赛中不仅进行了理论上的分析，还设计了一个实验。他们在空气中用扬声器来模拟环绕噪声场，用旋转的废罐作为目标。当废罐出现在声源和接收器之间时，得到了目标的轮廓像。论文中还提出了一些用环绕噪声场产生潜艇图像的方法。一种方法是声学照相机，另一种方法是 Schlieren 成像术的一个变种。似乎他们并不熟悉 Schlieren 成像术，但自己得到了这一方法的主要原理。

6.5.2 北卡罗来纳大学队的模型

北卡罗来纳大学队的数学模型将潜艇的形状表示为椭球，在环绕噪声场中主要对声波传播起阻碍作用，不考虑它对噪声的反射。这样潜艇使得接收器接收到的声强度明显地下降。

文章建议用布置在直交网格点的接收器接收到的信号，生成声强度等高线的方法确定潜艇的位置和速度以及潜艇的轮廓。

§ 6.6 几点启示

“潜艇探测问题”是美国大学生数学建模竞赛中最好的题目之一。首先，水下目标的探测有重大的实际意义。除了用于潜艇等水下军事目标的探测外，还可用于水下探矿、海洋动物研究和海洋生态、环境保护等方面。这样的问题很容易激发参赛学生的强烈兴趣。其次，这个问题又是一个正在开展研究取得一定进展，但离开真正解决尚有一大段距离的问题。如果用一个虽然很重要的意义，但实际上已经解决得差不多的问题作为竞赛的题目，虽然有时仍不失为一个好的题目，但多少有点练习题的味道。而本题涉及的课题是一个有待于突破的课题，提出这个问题除了竞赛本身外，很明显地希望从学生的论文答卷中获得新的思

想和见地。再次，涉及用海洋噪声探测水下目标或水下目标成像的文献十分稀少，参赛学生必须充分发挥自己的想像力和创造性才可能取得进展，在竞赛中取得好的成绩。综上所述，“潜艇探测问题”的确是一个有很大实际意义、激发学生想像力和创造精神，有很大发挥余地的一个开放性的问题。

参赛学生中的佼佼者的确像命题人希望的那样，淋漓尽致地发挥了他们的想像力和创造精神。Worcester 综合工学院队提出了一种用几个抛物面反射器捕捉信号，用三角方法定位的方法；Wake Forest 大学队用一定方式排列的水下听声器构成的一定排列的监听站，利用反射球面波的性质巧妙地决定潜艇的位置；又提出了用多普勒效应决定潜艇速度的新思想。Pomona 学院队巧妙地在陆地的空气中设计了一个模拟海洋环绕噪声场探测的实验并获得了成功；他们还独立地提出了用二组有方向的水下听声器构成的陈列装置重构目标的三维轮廓的设想，和独立地提出了一种用激光干涉技术观察声场中压力涨落从而成像的设想。这种设想与 Schlieren 成像术几乎不谋而合…正如水下噪声成像领域的最高权威 Buckingham 评论的：“这些文章的质量是很高的，提出的某些设想可能在实际中是很有效的。…或许，我们会看到学生文章中提出的某些新颖的思想会在将来的实际系统中实现。”

本题的阅卷也给我们留下了深刻的印象。从阅卷人写的评述和挑选出来的优秀论文我们可以看出，他们评判竞赛答卷的优劣主要有两方面的标准：第一，是否有新颖的思想和创造性的思维；第二，答卷是否采用了正确的数学建模的过程和符合数学建模的一般规律。阅卷人评论说，这些优秀论文之所以高出一筹的原因是“他们将新颖的思想结合到问题的分析中去，这表现在他们在设计接收器时采用十分聪明的构思，因此在问题给定的条件下是很有效的。虽然有些队在数学的分析方面有些欠缺，但只要他们作创新的尝试时有创造性思想总会引起评阅人的青睐”。从他们将北卡罗来纳大学队的论文评为优秀论文可以清楚地看出

这一点。这个队的主要假设是潜艇将噪声全部散射，因而几乎没有反射的效应。探测方法主要依据是潜艇阻碍噪声的传播。这一假设是和其他几篇优秀论文的基本假设相悖，是值得商榷的。此外，文章的表达也有欠缺。但鉴于这篇文章提出的用等高线图确定目标的方法很有创造性，仍被评为优秀论文。

阅卷人的评论认为数学建模应该包括以下几个主要步骤，第一，将一个科学问题用数学术语归结出来；第二，解此数学问题或在这一过程中发明一种新的方法；第三，用原来实际问题的眼光解释获得的数学结果。在第三步中还包括模型精度的分析和模型的评价等，还包括模型的改进和以上三步的适当重复。优秀论文在这三个步骤上保持了良好的平衡。评阅人也指出有许多答卷像信号处理课程的家庭作业，一页复一页充斥着“理论”和所提问题没有任何清楚的联系。虽然有些文章中不乏数学技巧，但几乎没有建模和模型。这种答卷是完全不符合数学建模竞赛宗旨的。

我国的大学生数学建模竞赛也已开展多年，在命题、阅卷等方面也积累了一些经验，竞赛的规模也已大大超过了美国大学生数学建模竞赛，目前正继续健康、稳健地发展。

通过对 1996-A 题即潜艇探测问题的命题、阅卷和得奖优秀论文的剖析，我们发现还是有许多值得借鉴之处。

首先，要进一步提高命题的质量，要发掘更多实际意义大，鼓励创造精神、发挥余地大的开放性问题。要做到这一点既要扩大命题人的队伍，发动第一线上的科研工作者和工业与其他产业的工程技术人员参与命题，用科技、生产中的新颖课题充实我国数学建模竞赛的题库；另一方面又要解放思想，敢于将尚未解决的、没有现成答案的或答案不惟一的有意义的问题经过加工提炼，作为赛题。

第二，在阅卷中要将鼓励学生发扬创新精神作为主要的原则，将是否具有数学建模和解决实际问题的能力和是否掌握正确的数学建模方法作为重要的考核标准。目前由于我国的大学生数

学建模竞赛规模很大，无法像美国一样采取集中阅卷的方式而采取赛区和全国二级阅卷的制度。为了保证阅卷的公平性和标准的一致性，目前采用提供参考答案的办法。但参考答案有时会产生副作用。个别阅卷人过于拘泥于参考答案，否决了有些富有创造精神，但模型和方法与参考答案不符或有某些缺陷的答卷。另外，也有少数阅卷人，在阅卷中过分偏爱数学方法和技巧，有时不自觉地将数学水平的优劣放在首位而不是将建模或建立的模型的优劣作为主要的判别标准。这就需要我们通过不断地努力，提高阅卷队伍的自身素质，逐步形成一支业务水平高，对大学生数学建模竞赛的宗旨和意义有深刻认识的阅卷队伍。

第三，要鼓励学生充分发挥自己的想像力和创造精神，在竞赛中敢于提出新思想，敢于尝试新方法。我国学生学习刻苦，基础扎实是得到公认的。但是不能不看到由于种种原因，和某些国家的青年学生相比，在发挥想像力和敢于创新方面中国学生有一定的差距。我们举办大学生数学建模竞赛的本意之一就是要提高我国大学生的这一重要素质。同时，也只有真正发挥出高度的创造精神才会在竞赛中获得优异的成绩。

参 考 文 献

- (1) Buckingham, M. J., Theory of acoustic imagine in the ocean with ambient noise, *Journal of Computational Acoustic* 1, 1993; 117~140.
- (2) Buckingham, M. J., S. A. L. Glegg, Imaging the ocean with ambient noise, *Nature* 356 (1992); 327~329.
- (3) Buckingham, M. J., John R. Potter, Chad L. Epifanio, Seeing underwater with background noise, *Scientific American* 274 (2) (1996); 86~90.
- (4) Pedrotti, F. L., L. S. Pedrotti, *Introduction to Optics*, 2nd

ed. Englewood Cliffs, Prentice Hall, 1993.

(5) Stephens, R. W. B (ed), Underwater Acoustics, Wiley-Interscience, 1970, New York.

(6) Murk, Walter H., Peter Worcester, Carl Wunsch, Ocean Acoustic Tomography, 1995, Cambridge University Press, New-York.

第七章 竞赛择优问题

王 强

北京应用物理与计算数学研究所

提 要

数学教育的目标,是培养学生运用基础知识解决科技发展和生产实践中的问题. 教育目标的实现程度,是要经过测量和评价的.

数学建模竞赛中产生的大量答卷或论文,以数学形式全面而深刻地反映出教育工作的质量. 如何评阅和择优,则是各类赛事中最难于处理的.

本章介绍 1996 年美国大学生数学建模竞赛 B 题及其相关情况,力求对三份优秀答卷加以比较. 此外,简略说明中国和美国在竞赛择优问题上的实际操作情况.

§ 7.1 问题和背景

7.1.1 问题的提法

在确定像数学建模竞赛这种形式的比赛的优胜者时,常常要评阅大量的答卷,比如说,有 $P=100$ 份答卷. 一个由 J 位评阅人组成的小组来完成评阅任务,基于竞赛资金对于能够聘请的评阅人数量和评阅时间的限制,如果 $P=100$,通常取 $J=8$. 理想的情况是每个评阅人看所有的答卷,并将它们一一排序,但这种方法工作量太大. 另一种方法是进行一系列的筛选,在一次筛选中每个评阅人只看一定数量的答卷,并给出分数. 为了减少所看答卷的数量,考虑如下的筛选模式: 如果答卷是被排序的,则在

每个评阅人给出的排序中排在最下面的 30% 答卷被筛除；如果答卷被打分（比如说从 1 分到 100 分），则某个截止分数线以下的答卷被筛除。这样，通过筛选的答卷重新放在一起返回给评阅小组，重复上述过程。人们关注的是，每个评阅人看的答卷要显著地小于 P 。评阅过程直到剩下 W 份答卷时停止，这些就是优胜者。当 $P=100$ 时通常取 $W=3$ 。

你的任务是利用排序、打分及其他方法的组合，确定一种筛选模式，按照这种模式，最后选中的 W 份答卷只能来自“最好的” $2W$ 份答卷（所谓“最好的”是指，我们假定存在着一种评阅人一致赞成的答卷的绝对排序）。例如，用你给出的方法得到的最后 3 份答卷将全部包括在“最好的”6 份答卷中。在所有满足上述要求的方法中，希望你能给出使每个评阅人所看答卷份数最少的一种方法。

注意在打分时存在系统偏差的可能。例如，对于一批答卷，一位评阅人平均给 70 分，而另一位可能给 80 分。在你给出的模式中如何调节尺度来适应竞赛参数(P , J 和 W)的变化？

7.1.2 问题的背景

这个问题，反映着数学建模竞赛中答卷排序工作的实际困难以及对解决办法的需要。为了改进工作，明白无误地在向莘莘学子求计求谋。诚如 Bell Labs 的 V. B. Mendiratta 所说，这是一个富有挑战性的现实问题。自然，它的应用也会扩展到诸多领域的决策方面。

在中国，全国数学建模竞赛历经数年，业已走上良性循环的道路。评阅答卷的工作是分为赛区和全国两级进行的，这更是加重了其复杂程度。如何确保评阅工作的公正性和科学性，一直是一个受到多方面关注和不断讨论研究的课题。在以往的竞赛中，评阅人深入细致地分析赛题并理解和把握它的各种解法，从而形成较为一致的评分标准。这是一条关键性的成功经验。其次，便

是依靠评阅人的公正与求实精神，在管理工作方面，也应当具备与此相适应的一套切实有效的办法。

评阅过程显示出，评阅人往往会对适合其业务倾向的答卷给出较高分数，反之则较低，这是一种系统误差；评阅人也会由于某种因素的影响而判分偶然偏高或偏低，这是一种偶然误差。然而，从数学上界定这两种真实存在的误差是困难而模糊的。再者，如何理解题目假定的“绝对排序”？如何在开始几轮筛选中保护 2W 个“最好的”答卷不被筛除？这一切似乎也是无章可循的。

基于上述情况，我们选择具有典型意义的三份答卷，从如下两个方面加以介绍：第一，如何完成题目所给的基本任务；第二，怎样分析评阅人的判分误差。

7.1.3 竞赛结果

在参赛的 393 个队中有 267 个队选择了本题，结果有 5 个队获得了特等奖，38 个队获得一等奖，77 个队获得二等奖。获得特等奖的 5 个队分别来自复旦大学、Gettysburg 学院、St. Bonaventure 大学、中国科技大学和 Washington 大学。我国华东科技大学、信息工程学院、国防科技大学、华南科技大学、西安电子科技大学和浙江大学共六个队获得一等奖。

7.1.4 本章其余各节的安排

在 § 7.2, § 7.3 和 § 7.4 中，我们分别介绍复旦大学、Gettysburg 学院和中国科技大学三个队的优秀论文；在 § 7.5 中我们对其余二篇优秀论文作一概述；在 § 7.6 中对两位评阅专家的意见进行归纳。

§ 7.2 基于模拟分析的论文优选方案

——复旦大学优秀获奖论文(答卷)

这份答卷共建立五个模型：理想模型，圆桌模型，传统模

型，筛除模型和高级圆桌模型。在这些模型之间，彼此略有逻辑关联。此处，我们仅引用具有代表性的高级圆桌模型。

在答卷的假设条件中，认为存在着评阅人一致同意的绝对的排序和判分，而且其判分服从正态分布 $N(70, 100)$ 。其次，认为评阅人在判分过程中会出现偶然误差和系统误差。再者，明确指出，题目所要求的“最后 W 份优胜答卷只能来自‘最好的’ $2W$ 份答卷”是一个概率事件。

这个模型使用排序和判分相结合的方法。在最后一轮筛选之前，评阅人对每组答卷排序而且组与组之间部分地交换答卷，依此反复，以求实现筛除 30% 最下面答卷的题目要求。这个筛选方案结构稳定而且易于执行。每轮中对答卷交换次数的设定使整个过程相当灵活。在最后一轮，利用答卷平均分来确定优胜者。

1. 把答卷平均分配给每位评阅人。在每一轮中筛除比例取为 30%。在 n 轮筛选之后，每位评阅人仅留下一份答卷。后面会说明，一旦给定评阅人的能力参数，我们可以确定在每轮中的交换次数 K_i 。

对于本题 $n=6$ ，每轮之后每位评阅人留有的答卷份数为 9, 6, 4, 3, 2 和 1。每轮中筛除答卷的份数为 4, 3, 2, 1, 1 和 1。

2. 评阅人围着圆桌就坐。设 $K_i = i$ （随后我们充分讨论选择 K_i 值的方法）。在第一轮中， $K_0 = 0$ ，评阅人不交换答卷只是给它们排序并筛除最差的 30%。
3. 在第二轮中， $K_1 = 1$ 。每位评阅人把最差的 30% 答卷向右传。之后，每位评阅人为得到的新答卷判分而且为所有答卷（包括未经传递者）重新排序，而后筛除最差的 30%。
4. 对于 $K_i \geq 2$ ，向右传、判分，再排序和筛出最差的 30%，这组步骤进行 K_i 次。
5. 当每位评阅人只有一份答卷时，答卷已被分发和评阅过 n 次。从 n 个判分的平均值中，我们选取最好的三份答卷作为优胜者。

为什么要交换最下面的 30% 答卷呢？题目的要求是，每位评阅人可以筛除最下面 30% 的答卷，此处给出的是一个限度。评阅人会依据自己手中答卷的优劣情况自行决定（在限度内）筛除量，这导致方案的复杂化和不稳定。我们的办法是为评阅人手中较差答卷提供流通、比较和再评阅机会。例如，考虑一份已由 J_1 传给 J_2 的最下面 30% 的答卷。如果在 J_2 重新排序后，它依然属于 J_2 手中答卷最下面的 30%，则它确应被筛除；如果 J_2 的最下面 30% 答卷中包含不是来自 J_1 的答卷（仅由 J_2 排序），则筛除它也不会破坏 30% 的限度。

每轮中传递答卷多少次？首先，我们注意 $\{K_i\}$ 的两个性质：

1. $\{K_i\}$ 是有界的，即， $0 \leq K_i < J$ 。
2. $\{K_i\}$ 是单调上升的，即， $i < j \Rightarrow K_i \leq K_j$ 。

只有 J 位评阅人，而且所有答卷分成了 J 组。当 $K_i \geq J$ 时，评阅人会发现先前传递出去的答卷又回来了。所以 $K_i < J$ 。

在最后的几轮中，优秀答卷（顶级的 6 份答卷）似乎更易于被筛除，应逐步增加每轮的传递次数，所以 $\{K_i\}$ 是单调上升的。

其次，在 $\{K_i\}$ 和花费 C 之间存在依赖关系。也就是， C 是随着所有评阅人排序过的答卷总数 K 而单调上升的， $K = 8 \sum_{i=0}^J P_i K_i + 100$ ，此处 P_i 是在第 i 轮筛除答卷的份数。在这个模型中，每位评阅人阅读答卷的数量几乎是相等的，所以 C 随着 K 单调上升。

显然，阅卷花费和筛选方案的精度是相互矛盾的：花费越少，产生不合格优胜者的概率就越大。我们也可以从其相应的 C 为最小的 $\{K_i\}$ 开始，通过按顺序逐一地试验 $\{K_i\}$ ，我们一步步增加花费，与此同时也增加了方案的精度。当精度适合要求时，由这组 $\{K_i\}$ 相应的方案就是最佳方案。我们可以利用下面要谈的模拟程序来检验一个方案的精度。

花费函数 减低评阅人评阅答卷数量是为了节省开支，不同评阅人评阅数量应尽可能相等。花费函数取决于实际情况，我们

采用如下形式：评阅 1 至 20 份答卷则每份 m 元，21 至 50 份则每份 $2m$ 元，而 51 至 100 份则每份 $4m$ 元。于是

$$C = m \sum_{i=1}^J \{a_i + (a_i - 20)u(a_i - 20) + 2(a_i - 50)u(a_i - 50)\},$$

其中 a_i 是第 i 位评阅人评阅答卷的份数且

$$u(x-a) = \begin{cases} 0, & x < a; \\ 1, & x \geq a. \end{cases}$$

花费函数的微小变化对于筛选方案仅有微小影响。此处取 $m = 10$ 。

评阅人 初步模拟显示，评阅人的能力是确定筛选方案的最重要因素。我们使用两个参数来描述评阅人的能力：

- 判分偶然误差的方差。方差越小，表明评阅人的经验越丰富而且判分越精确，反之亦然。这个方差可由该评阅人的以往工作效能得到。
- 系统偏差的量度。在现实生活中，把个人偏差量化是困难的。所以，我们把问题简化而把评阅人和答卷分为三种类型：激进的，中立的和保守的。一位激进评阅人对激进答卷给高分而对保守答卷给低分，对于中立答卷没有偏差。一位保守评阅人的情况则相反。中立评阅人则完全没有偏差。

模拟算法 下面给出对评阅人判分过程的模拟：

1. 生成 100 个服从 $N(70, 100)$ 分布的界于 1 至 100 之间的随机整数。把它们作为答卷的“真实”分数放入数组 $\text{paper-score}[1, i]$ 。
2. 取定常数 d 作为全体评阅人偶然误差的上界。生成 8 个随机整数 d_j 作为评阅人偶然误差的标准差，使之在 0 至 d 之间呈离散均匀整数分布，并把这些数放入数组 $\text{judge}[1, j]$ 中。
3. 取定常数 $e > 0$ 作为系统偏差值。设 1, 0, -1 分别代表激进，中立和保守。给每份答卷一个取自 {1, 0, -1} 的数并放入数组 $\text{paper-score}[0, i]$ 。给每位评阅人一个取自 {1, 0, -1} 的

数并放入数组 $\text{judge}[0, j]$. 我们用

$$s = e \cdot \text{paper-score}[0, i] \cdot \text{judge}[0, j]$$

来计算系统偏差 s . 例如, 一位保守评阅人遇到激进答卷, 则 $s = -e$.

4. 评阅人 j 为答卷 i 判分的模式: 设

$$u = \text{paper-score}[1, i] + s.$$

在区间 $[1, 100]$ 中按正态分布 $N(u, d^2)$ 生成随机整数并放入数组 $\text{judge-score}[i, j]$. 这样, 我们就可以生成评阅人的判分矩阵.

对参数 d 和 e 的处理方法 从概率论知识可以得到

引理 设 X_1, X_2, \dots, X_n 是分别具有方差 σ_i^2 ($i = 1, 2, \dots, n$) 的独立随机变量, 则 $\bar{X} = \sum X_i / n$ 的方差是

$$\sigma^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2.$$

于是, 有如下推论:

推论 1 $\frac{1}{\sqrt{n}} \min_{1 \leq i \leq n} \{\sigma_i\} \leq \sigma \leq \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} \{\sigma_i\}$.

由此我们知道, 在高级圈桌模型的最后一轮中, 对每份答卷采用其平均分排序确实能够提高评阅过程的精确度.

利用 Cauchy 不等式, 可以得出

$$\sigma^2 \geq \frac{1}{n^2} \left(\sum_{i=1}^n \sigma_i \right)^2,$$

从而有

推论 2 $\sigma \geq \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n \sigma_i}{n}$.

作者进一步假定 σ 在 $[0, d]$ 上呈现离散均匀分布, 从而

$$\frac{\sum_{i=1}^n \sigma_i}{n} \approx \frac{d}{2}.$$

对于 $n \leq 8$, 这就得到平均偶然误差的标准差估值:

$$\sigma \geq \frac{1}{2\sqrt{2}} \frac{\sum_{i=1}^n \sigma_i}{n} \approx \frac{\sqrt{2}d}{8}.$$

此外, 经过大量的计算机模拟, 作者发现如下的规律:

$$d < 10.$$

而且也知道, d 对结果影响巨大, 而 e 的作用则是很小的。自然, 可以把 d 和 e 取为同样的量级。计算机模拟是工作基础, 对 d 和 e 的不同取值来寻找最优方案。

模拟实践表明, 取 $e \in \{0, 5, 10\}$ 和 $d \in \{1, 3, 5, 7, 9\}$ 是适当的。

表 7-1 参数选取和失败率

偏差 e	最大方差	K_n	失败率(%)
0	5	1,1,1,1,1,2,4	1.8
	7	1,1,1,1,1,4,5	3.9
	9	1,1,1,2,2,4,5	4.8
5	5	1,1,1,2,2,2,4	0.7
	7	2,2,2,2,2,4,8	2.7
	9	2,2,2,2,2,4,8	6.7

由上述可知, 高级圆桌模型的操作步骤清楚, 易于理解和投入使用。对于给定的参数 P , J 和 W , 该模型首先就要求出最优的每轮交换次数(K_n)。无庸讳言, 优化(K_n)是消耗计算机资源的。

在这个模型中, 由于它的筛选机制和参数优化, 评阅人的答卷评阅量是很低的。这是一个相当突出的优点。

对于 P , J 和 W 的不同取值, 作者对本模型进行了稳定性检验; 并且相应于题目中所给的一组数据, 找出了最优筛选方案, 其中, 总阅读量应低于 170, 失败率为 3%。

§ 7.3 论文评阅的优化模式

——Gettysburg 学院优秀获奖论文

这份答卷的假设条件很是细致, 考虑了评阅中几种实际情况。

况。在对答卷的假设中，认为存在绝对排序；另外，答卷份数远远多于优胜者个数。在对评阅人的假设中，认为评阅人要有能力解决提出的问题；评阅人有自己对于答卷内容的偏好，评阅工作是在某种误差界限内进行；评阅人每次坐下来至多能公正地评阅 20 份答卷；要有首席评阅人，他只负责仲裁歧见和在最后一轮投票；评阅人至少要有五位，其中包括首席评阅人。

常数和术语的定义：

P ：答卷总份数。

J ：评阅人总数，不包括首席评阅人。

J_k ：代表第 k 位评阅人。

W ：优胜者总数。

read：一位评阅人一次评阅一份答卷。

round：一个筛除过程，把一组答卷剩 W 份。

R_a ：代表第 a 轮。

S_a ：在第 a 轮中堆数。一堆是一组份数少于 P 的答卷。

N ：在一堆中的答卷份数。

S_{jk} ：代表 k 轮中的 j 堆。

error：与绝对排序相矛盾的一位评阅人的排序。

评阅前准备工作 我们首先确定第一轮所需要的堆数 S_1 。为保证对称的筛选，我们需要 S_1 是 2 的方幂。由我们的假定，每位评阅人一次至多可以读 20 份答卷，所以每堆的份数不能超过 20。在每堆中的答卷份数是 $N=P/2^n$ ，此处 n 是满足

$$N=P/2^n \leqslant 20$$

的最小值。如果 2^n 不整除 P ， N 向上加一。答卷尽可能均匀地分成 S_1 堆。我们给每位评阅人发一堆。如果最后剩下答卷，某些评阅人将被要求重复第一轮的操作。

第一轮 评阅人 J_1 和 J_2 分得 S_{11} 堆和 S_{21} 堆。评阅人 J_1 从 S_{11} 堆中选择 W 份答卷，以便保证最好的 W 份答卷不会在 R_1 轮中筛除。选好后，他们再把手中的堆交换。于是，评阅人 J_1 从

S_{21} 中筛选 W 份，而 J_2 从 S_{11} 中筛选 W 份。尔后，他们共同比较两次评阅的答卷排序进而由 S_{11} 和 S_{21} 的合并中选出 W 份答卷来。如果存在争论，由首席评阅人决定提交哪份答卷。依此办法每两个堆都筛选出 W 份答卷。第一轮完成时，有 $S_2 = 2^{n-1}$ 个堆，而且每堆有 $N=W$ 份答卷。

中间轮 将有 $n-2$ 个“中间轮”。在第 r 轮 R_r 开始时，有 $S_r = 2^{n-r-1}$ 堆和 $N=W$ 份答卷。操作方法与前相似，比如，把 S_{1r} 和 S_{2r} 分给两位未曾经手过的评阅人。每位评阅人都从 S_{1r} 和 S_{2r} 的合并中选出 W 份答卷，而后再共同确定 W 份答卷向上提交。首席评阅人解决任何歧见问题。其他堆也如此处理。这个过程逐轮重复直到 $n-1$ 轮 R_{n-1} ，这轮完成时，便只剩下 $2W$ 份答卷。

决胜轮 最后一轮 R_n 是投票过程。为了保证公正性并考虑到决胜的重要性，推选包括首席评阅人在内五位评阅人来评价各答卷。这些评阅人评阅剩下的 $2W$ 份答卷并且排序。一位官方的可能是附加的评阅人记票。得票数排在前面的 W 份答卷就是优胜者。如果出现票数相同而难分胜负，首席评阅人投票解决。

人的因素 一个不能控制的变量是人的因素。作者用描述评阅人行为方式的概率分布来模拟人的因素。每位评阅人对答卷内容持有偏好。最普通的例子是，一位评阅人把形式看得比内容重要，而另一位评阅人却更侧重后者。这样，两份答卷的排序就可能颠倒。此处，用 d 表示在绝对排序尺度下两份答卷的距离。作者进而选择函数

$$E(P, d) = \frac{1.46 + \arctan(1 - 60d/P)}{2.92 + \frac{\pi}{2}}$$

作为一位评阅人对两份答卷的排序不同于绝对排序的概率，此处 P 表示竞赛中答卷的总份数。作者指出，随着两份答卷的距离增大，错排的概率迅速下降。当两份答卷的距离为 $0.01P$ 时，错

排的概率近似于 50%. 所以，对于 $P=100$ 在答卷 5 和答卷 6 之间选择完全是随机的。而当距离大于 $0.17P$ 时，错排概率为零。距离界于 $0.01P$ 和 $0.17P$ 之间的错排概率值正是真实情况的表达——两份答卷越接近，评阅人对答卷风格的个人偏好越可能影响它们的排序。类似地，两份答卷分离越远，评阅人的偏好很少会影响对它们的比较。

作者指出，对人的因素建模是困难的。上述函数是本文对人的本性的最佳估计。没有数据可供参照，以便搞清楚在这些环境下人们实际上是如何动作的。

部分结果 不包括首席评阅人的仲裁工作，经过 n 个筛选轮后，总的阅读量是

$$2P + \sum_{i=2}^{n-1} 2^{n-i-2} W + 5 \cdot (2W).$$

其中，第一项是在 R_1 中的阅读数，第二项考虑 R_2 到 R_{n-1} ，第三项用于最后一轮 R_n 。

这个模型的失败概率特别低，通常小于 0.1%。但它的总阅读量是比较大的，以题目所给数据为例，总阅读量为 254.

§ 7.4 快速选择优胜者 ——中国科技大学优秀获奖论文

在这份答卷的假设条件中，认为存在一种客观准则以判定两份答卷的优劣次序，而且这种排序具有传递性。由此可见，作者认为存在全部答卷的绝对排序。

其后，作者给出了简明的问题分析和符号系统。通常，在评阅中使用排序和判分两种方法。在判分方案中可能存在系统偏差，也就是，每位评阅人在判分中可以有主观倾向，这导致不同评阅人所判分数的不可比较性。不过有理由相信，同一位评阅人对不同答卷的判分是可比较的，即使是取自不同的筛选轮中。所以，同排序法相比，为了在早期筛选轮中记录评阅结果，判分是

更加有意义的方式。作者采用判分方案，因而同一位评阅人不必重读一份答卷。此处应注意的是，不直接比较不同评阅人的判分，而主要是利用判分来获得排序。

作者指出，在判分过程中存在偶然发生的“误判”，也存在评阅人之间系统性的主观差异。

表 7-2 符号

记号	含 义
P	答卷总份数
W	优胜者人数
J	评阅人总数
T	总评阅时间
P_i	答卷 i
$P_{(i)}$	具有绝对排序 i 的答卷
S_i	答卷 i 的绝对分数
$P_i > P_j$	答卷 i 在绝对排序下优于答卷 j
R_i	当前已知优于 P_i 的答卷份数
ORD	表达每两份答卷间已知关系的矩阵
$[x]$	不小于 x 的最小整数
$N(\mu_0, \sigma_0^2)$	具有均值 μ_0 和标准差 σ_0 的正态分布
σ_1	评阅人判分的标准差
μ_j, σ_j	评阅人 j 判分的均值和标准差
$\hat{\mu}_j, \hat{\sigma}_j$	μ_j, σ_j 的估值
Perror	发生误判的概率

评阅答卷的过程 作者简明扼要地介绍了对于评卷过程的设计：

- 把评阅过程分为若干个筛选轮，而且在每轮中遵从下述原则直到剩下 W 份答卷。
- 采用判分方案。
- 不比较相异评阅人的判分。
- 在第一轮中，把答卷平均分给所有评阅人。判分后，从每组中选出上方 $2W$ 份答卷而进入下一轮。

- 在每轮结束时，计算每份答卷上方的答卷份数（称之为该答卷的当前等级），之后筛除其当前等级多于 $W-1$ 的每份答卷。
- 在每轮开始时，尽可能把当前等级相近的答卷分给同一位评阅人。
- 每位评阅人在每轮中分得答卷的份数应尽可能相同。

在上述操作中，作者利用矩阵 ORD 来描述已知的答卷排序。其定义为

$$ORD_{ij} = \begin{cases} 1, & \text{如果 } P_i > P_j; \\ -1, & \text{如果 } P_i < P_j; \\ 0, & \text{如果 } P_i = P_j; \\ \infty, & \text{如果 } P_i \text{ 和 } P_j \text{ 没经评阅人比较.} \end{cases}$$

在每轮开始时，总是把答卷分给评阅人而由评阅人给每份答卷判分。而后，在完成的各轮中找出由同一位评阅人判分的 P_i 和 P_j 并且填充 ORD_{ii} 和 ORD_{jj} 。此外，还要用 ORD 的传递闭包来代替 ORD ，简言之，就是把所有由 ORD_{ij} 得到的非直接答卷排序加到 ORD 矩阵中去。依据这个矩阵就可以计算一份答卷的当前等级。

考虑误判 误判指的是最后 W 份答卷不是最好的。如果最后 W 份答卷不是完全属于最好的 $2W$ 份，则发生了误差。

假定，对于一份具有绝对分数 μ_1 的答卷，某位评阅人的判分是遵从正态分布 $N(\mu_1, \sigma_1^2)$ 的一个随机数。误判来源于评阅人判分对于绝对分数的偏差。

必然存在所有答卷绝对分数的一个分布，假定它是正态分布 $N(\mu_0, \sigma_0^2)$ ，从而，比值 σ_1/σ_0 就反映评阅人辨别答卷质量的能力。给定 P_i , j , W 和 σ_1/σ_0 ，就可以利用基本模型来估计误差出现的概率 P_{error} 。如果这个概率充分小，我们就可以期望该模型提供令人满意的结果。

矩阵 ORD 的精度是可以改善的。例如，要确定 P_i 和 P_j 的实际排序，只须找到同时读过这两份答卷的评阅人的判分，比较

两个判断之和就能确定新的 ORD_{ij} .

对误判概率的粗估和观察 下列两个模拟是根本性的:

- 模拟绝对分数的分布. 一般地说, 有理由采用一个正态分布. 为指定 100 份答卷的分数, 可在 0 到 100 之间生成遵从 $N(60, 30^2)$ 的 100 个随机数.
- 模拟对一份答卷的判断. 通过在该答卷的绝对分数上添加一个正态随机数来模拟一位评阅人的判断.

比值 σ_1/σ_0 应是相当小(比如, ≤ 0.1), 因为评阅人应精于评阅. 在我们的模拟中取 $\sigma_1/\sigma_0 = 1/30, 2/30$ 和 $3/30$ 三种情形.

作者也对这些情形从理论上估计了最差环境中的误判概率 Perror. 当 $P_{(7)}, P_{(8)}, \dots$ 之中的一份或多份答卷进入最后的三份优胜卷, 误差便发生. $P_{(7)}$ 进入最后三份的概率在 Perror 中所占份额最大. 设 $Perror(i, j)$ 为误判答卷 i 和 j 的概率. 作者通过 $P_{(7)}$ 进入最后三份的概率来近似表示 Perror:

$$\begin{aligned} Perror &\approx \frac{1}{8} Perror(3, 7) \\ &+ \left(\frac{1}{8}\right)^2 \sum Perror(3, i)Perror(i, 7) + \dots \\ &\approx \frac{1}{8} Perror(3, 7) \\ &+ \left(\frac{1}{8}\right)^2 \sum_{i=4}^6 Perror(3, i)Perror(i, 7). \end{aligned}$$

模拟结果与理论估计一致.

表 7-3 数值实验的结果与理论估值
 $P=100, J=8, W=3$

σ_1/σ_0	Errors	观察 Perror	Perror 粗估
1/30	0	.000	10^{-7}
2/30	0	.000	.0006
3/30	4	.004	.004

评阅人之间的系统偏差 当考虑到不同评阅人评分倾向的差别时，在第一轮中由每位评阅人选出相同数量的答卷就不是好办法，因为很可能在第一轮就把优秀的答卷筛除。

另一种代替办法是，当第一轮评阅结束时，把每组答卷的评分输入计算机，它会给出每位评阅人参数(平均分和标准差)，而且会相应于一个确定的绝对水平对每组答卷计算出用于筛除答卷的分数界限。按这个办法，优秀答卷很少可能在第一轮被筛除。对所有评阅人参数的估值使我们能够在某种程度上比较来自不同评阅人的评分。

我们利用 Bayes 估计来确定评阅人 j 的参数 (μ_j, σ_j) 的估值。假定评阅人 j 给答卷 P_1, \dots, P_n 的评分是 S_1, \dots, S_n 。我们采用最大似然法来估计 σ_j ：

$$\hat{\sigma}_j^2 = \frac{1}{n} \sum [S - E(S_j)]^2.$$

之后，再利用 Bayes 方法估计 μ_j 。事实上，我们可能有关于每位评阅人评分倾向的先验知识。即使没有，我们依然有理由假定一个评阅人评分的先验分布。如果其先验参数是 (μ_0, σ_0^2) ，那么后验参数就是

$$\hat{\mu}_j = \frac{n \cdot E(S)}{n + \left(\frac{\hat{\sigma}_j}{\sigma_0}\right)^2} + \frac{\left(\frac{\hat{\sigma}_j}{\sigma_0}\right)^2 \cdot \mu_0}{n + \left(\frac{\hat{\sigma}_j}{\sigma_0}\right)^2}.$$

这样，我们就能把分位数 $(N(\hat{\mu}_j, \hat{\sigma}_j^2), \text{LEVEL})$ 用作评分界限，此处， $1 - \text{LEVEL}$ 是应被保留的答卷的预期份额。一个适当的取值是

$$\text{LEVEL} = 1 - \frac{W \cdot J}{P}.$$

若干建议

- 适当减少在第一轮中筛除答卷的份数，会降低错判概率。
- 在每一轮中按需要改变筛除份数，有助于在竞赛中确定不同

层次的参赛者.

- 为了切合实际和高效率, 建议首先预评答卷, 也就是, 筛除明显低质量答卷.
- 在各个筛选轮中间, 评阅人进行若干讨论以使大家获得有关答卷整体水平的知识. 这样的反馈机制肯定有助于降低评阅的标准差.
- 当剩下 $2W$ 份左右答卷时, 如果时间允许的话, 所有评阅人应汇集在一起共同阅读留下来的答卷, 以便尽可能准确地选出顶级的 W 份答卷.

对模型的检验 作者通过改变 P , J , W 的值来检验模型稳定性. 从中发现, 当 W/P 过于小(比如, $<1/100$)时, 本模型不能很好工作, 不过适当减少在每轮中筛除答卷的份数, 会使误差概率 Perror 降低. 对题目所给基本模型的模拟计算是成功的. 作者还运用具有随机生成的均值和方差的正态分布来模拟不同评阅人的评分, 结果是, 随着 LEVEL 上升, 总的评阅时间下降, 但发生更多的误差.

§ 7.5 其余模型简介

7.5.1 St. Bonaventure 大学的模型

这个模型的工作基础是把问题分为四个主要方面并分别加以处理: 在评阅人中间分发答卷、打分的办法, 每轮淘汰答卷的数量以及总的评阅轮数安排.

在评阅进程中, 着重于维持公正性和所有评阅手续的灵活处理. 在每轮中筛除尽可能多的答卷, 使总的评阅轮数降到最低, 而且最重要的是, 要使上述各个要点均衡地实现.

本模型假定, 预算资金只影响评阅人的数量. 而且还假定在评阅人中间存在一个近似的“绝对排序”系统, 也就是, 如果每位评阅人对所有答卷打分或排序, 那么, 诸位评阅人的结果会基本

上一致，仅在一些地方答卷相邻顺序有颠倒。

7.5.2 Washington 大学的模型

作者指出，这个模型充分健壮，不管在评阅过程中随机误差和系统误差如何，保证筛选出真正优秀的答卷。该模型引入评阅过程中的不协调性，也就是把实际的答卷得分表示为固有分数、评阅人系统偏差和误差项之和：

$$S_{ip} = S_p + B_i + \epsilon_{ip}.$$

本模型利用计算机指引下的迭代处理来确定评阅手续。在每轮评阅之后，计算机程序利用偏差估计为每份答卷计算出其固有分数的置信区间。这些置信区间用于筛除尽可能多的答卷，与此同时，在指定的置信水准上保证顶级 W 份答卷提交到下一轮。

作者就一系列参数值进行了计算机模拟并给出结论：固有分数服从均值 50 和标准差 20 的正态分布，偏差和协调性参数则是变化的。

本模型的特色是导出了若干解析结果，并用于评阅过程中。这个成绩的基础是如下的假定：答卷的固有分数、所有评阅人的系统偏差范围以及评阅人的随机偏差均服从正态分布，而且，正态分布中的均值和标准差是可以在评阅过程中近似获得，其误差认为是可忽略的。

§ 7.6 来自评阅专家的评论

竞赛择优问题也应用于其他决策领域，例如，奖学金的发放，或者在众多申请者中为一个特定职位进行筛选。在这种情形下，决策人或评阅人就必须以某种方式甄别申请者以求确定谁是“最优”。而且，这些决定要在时间限制下作出，就使得每位决策人不可能去评价所有申请者；即便可能，各种评价也难得完全一致。正是这个特征使竞赛论文的评阅工作复杂起来。

关于绝对排序的假定 (There is an absolute rank-ordering to which all judges would agree) 使问题貌似简单，实际上是一个陷阱，对各色各样的论文产生影响。有些论文假定绝对排序而对基本问题提供了启发式解法；更有些论文明知该假定不真实却依然用它建模，因为这是题目所要求的。另外一些使用这个假定的论文明显地不相信它，论文中发展了更复杂的算法而对“绝对排序”予以摒弃。还有若干论文试图运用图论命题乃至模糊集合概念等等理论来改进这个问题。数学建模实践表明，不切实际的假定会使模型很少能有实用价值。作为建模者，应严格地审视问题的所有假定，如有必要，应改进问题使之具有真实性。

建模的问题不同，在评阅过程中判定论文优劣的标准也是各异的。就这个问题来说，较好的论文应具备如下特征：首先，解决了基本问题；更进一步，便是完成一个具有适应复杂性的改进来精确建模，而它的用法却是简单的；再者，每个模型叙述清楚，并用例子展示用法。

更为理想的论文应能显示解法最优或接近最优；应能适应不同的模型参数；再进一步，应能处理评阅人偏差，把算法的成功率表示为评阅人偏差的函数并加以测量。最后，应能考查其他算法，指明模型的优点和缺点等。

在本次评阅中，获得特等奖的论文是基于如下特征脱颖而出：第一，对评阅人偏差的有效处理；第二，有助于实现模型应用的适当文档。

参 考 文 献

- (1) 戴忠恒：《教育统计、测量与评价》，中国科学技术出版社，1991年。

- (2) 佟庆伟等：《教育科研中的量化方法》，中国科学技术出版社，1997年。
- (3) Box, G. E. P. and Tiao, G. C. 1973. Bayesian Inference in Statistical Analysis. Reading, MA: Addison-Wesley.
- (4) Wang, Yihé. 1986. Introduction to Discrete Mathematics. Harbin, China: Harbin Instituto of Technology Press.
- (5) DeGroot, Morris H. 1986. Probability and Statistics. Reading, MA: Addison-Wesley.

第八章 恐龙捕食问题

叶其孝

北京理工大学 应用数学系

提 要

本章介绍了 1997 年美国大学生数学建模竞赛(MCM-1997)的竞赛情况、评阅和奖励，特别是介绍了 A 题的优秀论文、评阅人的评述和我们的评注。主要内容：MCM-97 的评阅、结果和奖励；密歇根州的卡尔文(Calvin)学院队的优秀论文；评阅人的评述；我们的注记(包括对可供参考的其他优秀论文的简要评注)。

§ 8.1 竞赛的评阅、结果和奖励

本次竞赛共有包括美国、中国、香港等 8 个国家和地区的 226 所大学的 409 个队参加，其中中国有 38 所大学的 107 个队参加。

各队的论文在 COMAP 的总部进行编号使得评阅人不知道论文作者的姓名和所属的学校。A 题的初评是在康涅狄格州的南康涅狄格大学进行的，共有 7 位评阅人。B 题的初评是在蒙大拿州的卡罗尔(Carroll)学院进行的，共有 5 位评阅人。每篇论文由两位初评评阅人评阅，摘要和论文的组织是论文评定的基础。如果两位评阅人的评分不同则进行协商，如果协商后还不一致，则再由第三位评阅人来评阅。

A 题是由佐治亚州佐治亚(Georgia)学院数学系的 Jack Robertson 和生物和环境科学系的 William Wall 提供的。B 题是由印第安纳州圣玛丽学院(St. Mary's College)的 Don Miller 提供

的。

终评是在加州的哈维·马德(Harvey Mudd)学院进行的，A题评阅人有19位，B题评阅人有13位。评出的最后结果是：

	O	M	H	P	合计
MCM-1997A题获奖队数(中国队数)	5(0)	37(13)	58(28)	134(23)	234(64)
MCM-1997B题获奖队数(中国队数)	4(1)	25(6)	43(18)	103(18)	175(43)

其中，O=Outstanding=特等奖，M=Meritorious=一等奖，H=Honorable Mention=二等奖，P=Successful Participant=成功参赛奖。

A题共有五篇优秀论文^{(1), (6)~(8)}。

每个参赛队都将获得由竞赛主任和每题的评阅组长签名的证书。

美国运筹学和管理科学学会(ORSA)给予两个获得特等奖队的队员现金奖励和三年的会员资格。这两个队分别是来自美国密歇根州的卡尔文(Calvin)学院队(A题)和印第安纳州的罗斯-哈尔曼(Rose-Hulman)理工学院队(B题)。此外，美国运筹学和管理科学学会还给获一、二等奖的队的每个队员一年的免费会员资格。

美国工业与应用数学学会(SIAM)对每题指定一个特等奖队作为SIAM的获奖队，每个队员都有现金奖励，每个队将于1998年7月在Stanford大学举行的SIAM年会特设的小型研讨会上作报告。这两个队是密苏里州的华盛顿大学队(A题)和加拿大的多伦多大学队(B题)。

美国数学协会(MAA)对每题指定一个特等奖队作为MAA的获奖队。他们是哈佛大学队(A题)和明尼苏达州的麦卡莱斯特(Macalester)学院队(B题)。两个队将在1998年8月在佐治亚州的亚特兰大举行的MAA的数学节(Mathfest)上作报告。

§ 8.2 MCM-1997A 题

Velociraptor, Velociraptor mongoliensis, 是生活在距今约 7500 万年前晚白垩纪(译注:白垩纪为距今 1.36~0.65 亿年的地质年代,是中生代最后的纪)的一种食肉(捕食其他动物的)恐龙.

古生物学家认为这是一种非常顽强的猎食其他动物的野兽,而且可能是成对或成群地外出追猎.然而,不幸的是无法像观察现代哺乳食肉动物在野外是如何追猎其食物的行为那样,观察到 *Velociraptor* 在野外的追猎行为.一组古生物学家来到你们队请求你们在 *Velociraptor* 的追猎行为的建模方面给予帮助,他们希望把你们的结果与研究狮子、老虎及其他类似的食肉动物行为的生物学家的研究报告相比较.

成年的 *Velociraptor* 平均长 3 米, 高 0.5 米, 体重约为 45 千克. 据估计, 这种动物跑得非常快. 速度可达 60 千米/小时, 持续时间约 15 秒. 在一开始的突然加速后, 它要停下来并在其肌肉中积聚乳酸以恢复体力.

假设 *Velociraptor* 捕食一种称为 *Thescelosaurus*(太西龙属) *neglectus* 的大小与 *Velociraptor* 差不多的双足食草动物. 从 *Thescelosaurus* 化石的生物力学分析得知 *Thescelosaurus* 可以 50 千米/小时的速度长时间奔跑.

第一部分

假设 *Velociraptor* 是一只独居的猎食其他动物的野兽, 试设计单个的 *Velociraptor* 潜近猎物并追猎一只单个的 *Thescelosaurus* 策略以及被追捕物逃避追捕的策略的数学模型. 假设当 *Velociraptor* 潜近到 *Thescelosaurus* 的 15 米范围内时, *Thescelosaurus* 总能觉察到, 根据栖息地及气候条件的不同, 甚至在(多达 50 米的)更大的范围内觉察欲捕食它的动物的存在. 此外, 由于 *Velociraptor* 的身体结构及体能, 它在全速奔跑时的拐弯半径是受到

限制的。据估计，拐弯半径大约是其體高的三倍。另一方面，*Thescelosaurus* 却是极其灵活的，其拐弯半径只有 0.5 米。

第二部分

更现实地假设 *Velociraptor* 是成对外出追猎，试设计一个新的关于成对的 *Velociraptor* 潜近猎物并追猎一只单个的 *Thescelosaurus* 的策略以及被追捕物逃避追捕的策略的数学模型。利用第一部分给出的假设和限制。

§ 8.3 Calvin 学院队的优秀论文——晚白垩纪的追逃对策⁽¹⁾

8.3.1 摘要

应用微分对策理论的方法，我们用半离散的计算机算法来对 *Velociraptor*（疾走恐龙）捕食问题进行建模。

按照简单、直观的原则定义了捕食者和食饵（被捕食者）的行为，我们识别预定来对抗另一组策略的一组策略，使得没有一组纯捕食者策略或者食饵策略能确定一种最优行为模式。作为替代，理想的策略应在二个或者更多的纯策略之间，以一种本质上是不可预测的或变化多端的方式转换使用。最终的最优行为展示了 *Thescelosaurus*（太西龙属）的伪装、吓唬和真的转弯的混合，以及 *Velociraptor* 的有预兆的拦截和简单的追逐的混合。

最后，利用这些策略，我们证明了成对追猎比单独追猎有着肯定的优势。

8.3.2 引言

我们用追逃的微分对策的半离散的计算表示来描述 *Velociraptor* 的追猎策略和食饵的逃跑策略。

首先，我们评述了惯用的非微分对策理论的形成以及它的原则在微分系统分析中的推广，并对 *Velociraptor* 问题的独特的方

面作了仔细的注记。

其次，我们提出了为把解析对策化归为数值时间迭代计算机算法所需要的最少的一组假设。

第三，我们考察了单个追猎者和单个逃跑者策略优化的完全和不完全信息假设的含义，然后把结论推广到更多追逐者的情形。

最后，我们评注了本模型固有的局限性并提出了我们的评估。我们得出结论：对于不完全信息对策被捕食者不存在纯策略，但是最佳的替代是用转弯和伪装交替使用的无法预测的方式行动。然而，捕食者有一个在有预报价值的算法和简单的追逐之间的妥协的清楚的控制策略。

8.3.3 用对策论术语表示的数学陈述

捕食者和被捕食者之间的拦截和逃避的竞争属于线性微分对策的广泛而形式不同的范畴。在追捕者和逃跑者的微分对策中，两个或多个局中人在遵从对他们的运动的某些限制下试图极大化或极小化它们之间的距离。

和经典的对策论不同，微分对策必须包括微分方程方法的应用以及利用定义局中人的状态和目标的连续的波动。

在传统的或者微分的对策中，局中人试图在一组可供选择的策略中挑选策略使得称为支付函数的值极大化，其中每个局中人的支付是由所有局中人的选择的某个函数确定的。在零和对策情形，所有局中人互相都处于直接竞争的地位使得由它们的支付函数表示的目标正好符号相反。追逃对策属于这个分类；追逐者试图使自己和逃跑者之间的距离极小，而逃跑者却试图使这个距离极大。

在 *velociraptor-thescelosaurs* 这样的追逃对策的情形，支付函数是一个简单的二元函数；仅有的有关的结果就是抓住或者逃走。相对于度对策，这种对策称为类对策，度对策的支付可以在更大的可能值上取值。在有些情形，把类对策嵌入到度对策中，同时

把抓住时间(或者逃跑者的分离时间)作为支付函数可能是有用的. 尽管在纯粹确定性的试验中这不是必要的, 但是一对特定的策略对的平均抓住时间可能提供了一个策略的功效的有用的统计度量.

在微分对策中有两类变量: 状态变量, 它规定了任意给定时刻整个系统的完全的结构, 还有控制变量, 对策的参与者用它以有利于极化它们自己的支付函数的方式来改变状态函数. 在惯用的追逃对策中通常的状态变量是参与者的空间坐标; 控制变量可以包括最大旋转角或加速度向量.

所有局中人都可存取在任意给定点和时刻处的状态变量的一个完全集的对策称为完全信息对策; 若不是这样的对策称为非完全信息对策. 非完全信息对策缺少精确的解析方法. 通常, 评价这种对策的最好的方法是用离散模型, 它可以借助于简单的计算算法来实现. 不幸的是纯粹离散模型往往冒如下的风险, 即会模糊可能依赖于没有量化的值的连续性的系统的本质细节. 仍然容许进行求解的计算方法的合理的妥协就是半离散方法: 时间离散化了, 而空间坐标仍可取整个实数值(Isaacs 1967, 42). 我们把猛禽捕猎作为一种半离散计算算法来实现.

velociraptor 问题具体化了两个局中人追逃对策中最有趣的情形. 如果逃避者的最大速度大于追逐者的最大速度, 那么逃避者的最优策略就是以最大速度直接逃离追逐者, 并总能成功地逃掉. 类似地, 如果追逐者在速度和机动性方面都更强, 那么追逐者的最优策略就是直接向逃避者运动, 并保证能成功地猎获逃避者. 但在追逐者速度较高, 而逃避者更为灵活的情形, 那就没有平凡的解法了, 最优对局可能需要复杂的或非决定性的策略.

8.3.4 假设和模型的研制

- 初始结构

直到捕食者移动到一个容易辨认的追逐半径内, 被捕食动物

一般无视捕食者的存在. 当被捕食者和一个捕食者之间的距离小于这个半径值时, 就引发了追逐应答. 试题给我们的 *thescelosaurs* 的追逐半径小于 50 m 大于 15 m, 致使当 *thescelosaurs* 借察到 *velociraptor* 的移动落入这个距离时, *thescelosaurs* 将立即逃走. 类似地, 我们假设 *velociraptor* 为抓住 *thescelosaurs* 而蓄意潜近它, 因此, 捕食者将处于一组蹲伏位置, 并在见到第一个逃窜信号时立即追逐. 尽管 *thescelosaurs* 可能会大吃一惊, 但它还是启动了追逐和逃避的运动序列, 从而被捕食者和猛禽几乎同时开始运动.

我们的模型打算通过或者选择每个捕食者和被捕食者的位置和朝向, 或者执行一个用以初始分隔距离的简单的概率潜近模型来规定初始状态. 在大多数情形, 我们令捕食者和被捕食者间的起始分隔距离接近所给的最小可能值 15 m. 这是因为, 对于大多数大的分隔距离, 被捕食者的最优策略是直接以最大速度逃离捕食者, 直到捕食者又接近它为止. 因此, 直到被捕食者由于接近捕食者而被迫离开直线路为止, 策略行为的差异通常并不明显.

在多个捕食者的情形, 我们或假定它们在半径为 15 m 的圆周的相对方向潜近猎物, 或假定它们从几乎相同的位置出发.

8.3.5 感觉的灵敏度

对于最简单的近似而言, 只要作如下假设就够了, 即假设追逃对策中所有的参与者都能完全且同时达到所有时刻的状态函数, 致使它们介入的是完全信息对策. 然而, 对于大多数真实的物理系统而言这是个相对不准确的假设. 估计距离和方向的能力总是遵从随机误差, 而视觉限于角度小于 180° 视域. 追于依靠其他感觉, 通常是听觉, 去追逐一个运动的对象是不那么理想的, 特别是会导致距离估计的相当大的误差. 这些限制对捕食者特别成问题, 它们有一个狭窄的能调节眼睛焦距的前视域, 而且严重地依赖于不仅能估计被捕食者现在而且能估计其将来的位置的能力.

我们的模型除了假定完全信息外, 在感觉感受中同时引进随

机的系统误差。在转向支配控制变量的程序前，随机误差是通过对捕食者和被捕食者之间的位移向量的大小和角度各乘上在 0.95 和 1.05 之间的一个随机数来实现的。由视域的局限性造成的方程中的不确定性（系统误差），是用与视觉方向和被观察物体间的角位移成比例的线性增长的差距来估计的。距离和角度各乘以一个范围在 $1.00 \pm s\theta/\pi$ 内的不同的随机数，其中对于角位移 $s=0.05$ ，而对于线性位移则为 $s=0.25$ 。最后，为强调视觉接触的重要性，捕食者的这些因素要分别增加 50%，即分别增加到 $s=0.075$ 和 $s=0.375$ 。

相关的问题是反应时间问题。在完全信息的情形，状态向量的知识是立即传递的。更为实际的假设是仅当经过大脑的心理处理后这些数据才能使用。通常，对环境的最新的感知过程可以认为是立即的感知；但在一个以百分之一秒计的比赛中，拒绝被捕食者对自己的行动的知识优于捕食者将遮掩模型的本质因素。为防止恐龙对其环境变化的瞬时反应，我们延迟它们对其他恐龙的状态变量，对 velociraptor 是 0.05 秒，对 thescelosaurs 为 0.037 秒。

8.3.6 运动的物理约束

假定每只恐龙的质心是按牛顿力学在二维平面上运动的质点。改变质点运动的两种可供采用的选择（遵从牛顿运动学方程）是加上线性加速度或角加速度。由生物力学分析提供的数据表明 velociraptor 的最大转弯半径是 1.5 m，而 thescelosaurs 的最大转弯半径只有 0.5 m。（我们理解试题陈述中的意思是指这些值可用于最高速度，即使这会导致 velociraptor 能作 32-g（译注： $g \geq 9.8 \text{m/sec}^2$ ）的转弯。）在这种情形，这暗示在任意给定的时间区间内的最大角位移可以定义为 $d\theta = a(dt)/v$ ，其中 v 是当前的速度， dt 是时间区间的长度， a 是向心加速度，其中 $a = v^2/r$ ，而 v 和 r 分别是在最大速度处的速率和最大曲率半径。没有相应的恐龙的线性加速度的数据，有必要援引类似的论据。非洲猎豹是起着和 ve-

lociraptor 类似的生态作用的现代捕食者。猎豹还有着许多属于 *velociraptor* 的同样的策略（高速、有限持续时间以及比它们的主要食饵（被捕食者）的转弯半径要大）。人们可以合理地假定 *velociraptor* 的线性加速能力是类似的。猎豹可以在大约 2 秒钟里加速到最高速度（超过 90 千米/小时）。但是 *velociraptor* 的最高速度要低一点，而且体重比猎豹要轻一点。这提示 *velociraptor* 的加速度可能会稍大一点。在假设 *velociraptor* 与其身体尺寸相比具有同样的相对加速能力下运作，并认为肌肉产生的力与身体的线性尺寸的平方成正比，而身体的质量与其立方成正比，因子 $(2/3)/(1.25)^{2/3} \approx .57$ 并非是对较轻的 *velociraptor* 的一个不合理的修正（其中 1.25 是猎豹的质量比之 *velociraptor* 的质量的近似比值）。

8.3.7 策略

可以预期动物是按照直接的启发性的原则来行事的，我们测试过的每种策略都反映了这个原则。为了进行模拟，我们假设每只 *velociraptor* 和每只 *thescelosaurs* 在追逃对策开始时必须只采取一种策略，尽管我们在后面会考虑在追逃过程中匹配策略可能会带来的好处。

捕食者的策略

- 捕食者策略-0（默认的）：以最大可能的速度不减速地向着被捕食者现在的位置直接运动（即使会背离被捕食者，也不能减速）。
- 捕食者策略-1（预言性的）：通过假定被捕食者将沿现在的方向以现在的速度继续运动来估计被捕食者未来的位置，并测定拦截路线。
- 捕食者策略-2（半预言性的）：取由捕食者策略-0 和捕食者策略-1 所指示的角度的平均值。

被捕食者的策略

- 被捕食者策略-0 (默认的): 以最大可能的速率从捕食者现在的位置不减速地直接运动开(即使会向着捕食者, 也不能减速).
- 被捕食者策略-1 (不变的转弯): 类似于被捕食者策略-0, 但是每当捕食者进入 1.5 m 圈时, 急转 90°, 再直线运动.
- 被捕食者策略-2 (变着转弯): 类似于被捕食者策略-1, 但是代之以旋转 90°, 以与到捕食者的距离成比例地不停地旋转.
- 被捕食者策略-3 (佯攻并虚张声势): 类似于被捕食者策略-1, 但是代之以旋转 90°后直线前进, 再以最大的速率转 270° 转回原地.

8.3.8 抓住和抓不住

通常, 在追逃微分对策的解析研究中, 抓住的条件只是追逐者和逃避者之间的距离小于某个设定值, 即抓获半径. 在我们的模型中, 我们采用一种或两种机制使抓住的概念更加现实.

- *velociraptor* 和 *thescelosaurs* 相距在 1m 内, *velociraptor* 可能试图“猛冲”并发动一次决定性的攻击, 刺中 *thescelosaurs* 身体的一部分 (颚和爪子) 用以伤害或阻止 *thescelosaurs* 以突然进发出的加速度前冲. 这是由包括两个概率性抓获条件来表示的, 一个因素依赖于从 0.5 m 到 1 m 线性地变化的距离, 另一个因素依赖于从 0 弧度到 π 弧度的线性地变化的角度. 如果生成的 0 和 1 之间的随机数小于这两个因素之积, 那么抓住就发生了.
- 另一个机制是, 如果 *velociraptor* 和 *thescelosaurs* 的运动相距在 0.5 米内, 它们身体猛烈相撞, 那么对 *thescelosaurs* 来说总是发生灾难性后果. 我们决定不对这种情形下 *velociraptor* 可能受到伤害进行建模.

如果在 15 秒内抓获没有发生, 则模拟结束, 并假定 *thescelosaurs* 已经逃走, 因为 *velociraptor* 被迫要停下来休息. 即使

velociraptor 可能并非在 15 秒内都以最高速度运动，加速或减速的过程至少也要和以常速度奔跑时的能量需求一样。

8.3.9 模拟器

模拟器由两个主要部分组成：负责更新每次迭代对策态势的外循环，以及运动生成器。后者，各自的生成器分别对捕食者和被捕食者执行策略。这是由下列五步来完成的：

1. 数据获得阶段：决定并记录每个对手的方位和距离。
2. 数据操作阶段：随机变更上述每个值以模拟感觉感受的不精确性。
3. 策略化阶段：基于该数据以及所选择的策略，选定“最好情形”的移动。
4. 限制阶段：按照动物相应的能力把物理限制加在所选的移动上。
5. 运动阶段：这个最终值传回到执行外循环。

如果在任一点外循环确定抓获发生，或已超过 15 秒的时间限制，则模拟停止，并报告最终状态。

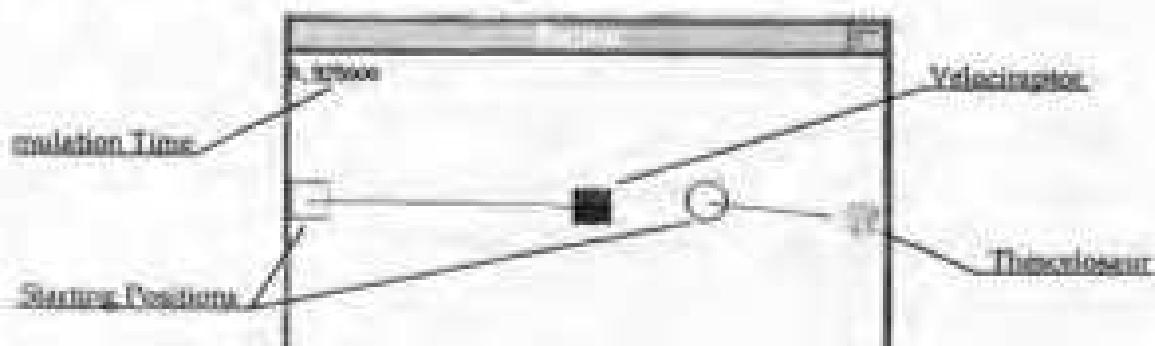


图 8-1 模拟窗口

模拟器是用 ANSI C++（以及 Hewlett-Packard 标准模块库）编写的。程序用一个 X Windows (X11R5) 显示界面以图示跟踪捕食者和被捕食者的位置。代码是用从 Gnu 软件基金会免费提供的 C++ 编译程序编译和连接的，并在 Sun Microsys-

tems SparcStation 5 工作站上执行完成的。源代码可以通过电子邮件 smenni23@calvin.edu 得到。

8.3.10 结果

一只 velociraptor

在完全信息以及前面确定的对感觉灵敏度的假设下，我们详细研究了单独狩猎的情形。在前一个假设下，每个初始位置的输出结果是确定性的并且是可重复的。因此，捕食者的每个策略要么成功，这时其结果是抓获被捕食者；要么失败，因为在指定的时间内没有抓获被捕食者。这就导致用二元矩阵的自然表述（表 8-1），其中 1 表示抓获，而 0 表示逃走。

表 8-1 完全信息对策下的抓获

	被捕食者 策略-0	被捕食者 策略-1	被捕食者 策略-2	被捕食者 策略-3
捕食者策略-0	1	0	0	1
捕食者策略-1	1	1	1	0
捕食者策略-2	1	1	1	1

捕食者策略-0 和捕食者策略-1（或被捕食者策略-0，被捕食者策略-1，被捕食者策略-2，被捕食者策略-3）相对排序非传递性值得进一步评注。关于只包含这些选择的简化对策，这对应捕食者（或者被捕食者）没有最优策略的态势。由于 thescelosaurs 是半径更小的急转弯，当 thescelosaurs 作急转弯时，velociraptor 能抓住 thescelosaurs 的仅有的方法时预先考虑 thescelosaurs 未来的位置。然而，这有导致 thescelosaurs 转向被捕食者策略-3 的潜在可能。由被捕食者所作的假的急转弯（对此，不那么复杂的捕食者策略-0 是可以不受其影响的）的后果，捕食者被迫过分介入。同样的分析反过来可用于 thescelosaurs，thescelosaurs 协调地提出的每个策略，velociraptor 可以匹配一个每次都能打

击到 *thescelosaurs* 的新的策略。因此，任一局中人坚持用单个的决定性策略来对策，那么另一个局中人就可利用这个对策的弱点。这至少可保证另一只恐龙不能先发制人地、协调地考虑到导致这个策略的战术。幸运的是，在完全信息对策中，*velociraptor* 可以由另一种选择：捕食者策略-2 以非随机的方式结合另两种策略的决定性策略，如果 *velociraptor* 经受着一个反应时间的延迟，如同下面要叙述的不完全信息的变形的情形，那么同样的这个选择就无效了。由于引进由不完全信息所产生的影响，决定性的输出结果由概率性的输出结果所替代。某些策略仍将实际上总能成功地击败其对手，但在许多情形，输出结果是充分随机的，因而只能用统计的方法来处理。为获得不完全信息对策中相对概率性的思想，我们对每一对在起始结构中带有随机变化的策略进行 10 次试验。这里报告的矩阵值就是由前面的试验决定的捕杀的概率(表 8-2)，这些值只能作为粗略的近似。

表 8-2 不完全信息对策下的捕获

	被捕食者 策略-0	被捕食者 策略-1	被捕食者 策略-2	被捕食者 策略-3
捕食者策略 0, 0	90%	50%	100%	40%
捕食者策略 1, 1	90%	100%	100%	100%
捕食者策略 2, 2	100%	100%	100%	70%
捕食者策略 0, 1	100%	100%	100%	70%

两只 *velociraptor*

在除两种情形外的所有情形中，两个 *velociraptor* 的策略优于单个 *velociraptor*，假定两个 *velociraptor* 的行为方式相同。如果一个 *velociraptor* 以预言性方式采用捕食者策略-1 行动，而另一个 *velociraptor* 以默认策略采用捕食者策略-0 行动，结果两个“半预言性”的捕食者策略-2 的 *velociraptor* 的情形极为相似，这暗示捕猎角色的专业化可能会提供很可观的优势(例如，考察一下 0,0

策略对 0,1 策略的输出结果.). 这些结果暗示, velociraptor 以群体去捕猎应该会更刺激, 特别是, 如果缺乏经验的 velociraptor(采用捕食者策略-0)能和更有经验的 velociraptor(采用捕食者策略-1)合作的话. 由两个“0”和两个“1”的组合的 velociraptors 所捕杀的总数不仅大于四个单打独斗的 velociraptors 的猎获数, 也大于没有经验的组合的成对捕猎的捕杀总数.

8.3.11 缺点

- 由于缺乏有关恐龙的生物力学性质的直接证据, 某些参数, 特别是最大线性加速度和反应时间的延迟的数据是猜测的、有争议的, 需要合理的争论. 类似于现代捕食者的论证是没有把握的, 而且如果目的在于比较现代哺乳动物捕食者的实验观测到的行为的模拟结果, 那么采用这些资料作为经验参数可能是具有肯定意义的资源.
- 由于半离散方法固有的局限性, 我们不能确保真的找到了最优解, 而只能保证给定解关于其他所考虑的解是最优的.
- 两个 velociraptors 的策略是不实际的, 他们假定了 thecelosaurs 完全忽略了两个 velociraptors 在任何时候的更远的距离.
- 引入反应时间延迟大大复杂了 velociraptors 的任务, 而且由于相当简单的假装转弯受欺骗而使之处于易受攻击的处境. 本模型缺乏“学习”即使 是重复的行为的能力, 没有融合在程序中的显式设计来反击这些行为的新策略.
- 由于忽略了诸如地形、能见度和障碍等因素, 本模型给出的是比在天然情形下 velociraptors 更多的优势. 这导致了不合理的高捕获率.

8.3.12 结论

我们的模型为成对捕猎对 velociraptors 更有优势的假设提供了有说服力的支持. 此外, 本模型证明了不存在一个最优的追策

略或逃策略，而且捕食者和被捕食者都喜欢转换技巧的灵活性。

8.3.13 附录：图示行进中的被捕食者策略的有代表性的模拟输出

虽然捕食者的策略是相当明显的，然而被捕食者的策略可能更难看出来。下面的图说明在最可能出现的情形中各个非默认的被捕食者的策略，即与相应的最弱的捕食者策略相对的策略。此外，图 8-4 展示了由不完全信息产生的复杂情况。注意，当特别与图 8-2 相比时，*thescelassurus* 的趋势时转弯转得太远，而 *velociraptors* 延迟了对转弯的反应。

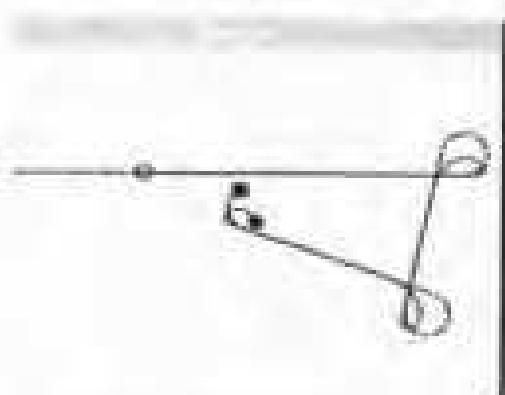


图 8-2 完全信息时被捕食者策略-1

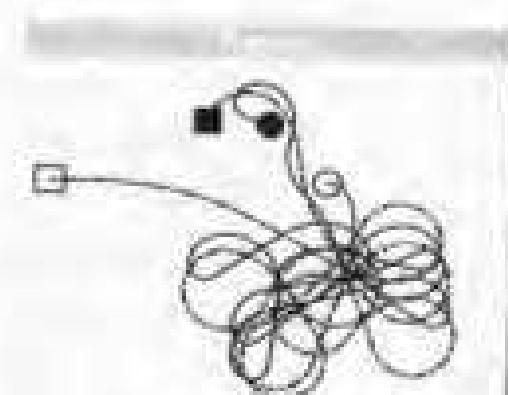


图 8-3 完全信息时被捕食者策略-2

8.3.14 参考文献

Crichton, Michael, and Steven Spielberg. 1993. Jurassic Park. Motion picture. Hollywood, CA: Universal Pictures.

Curio Eberhard. 1976. *The Ethology of Predation*. New York: Springer-Verlag.

Isaacs, Rufus. 1967. *Differential Games*. New York: John Wiley and Sons.

Miller, Geoffrey F., and Dave Cliff. 1997a. Co-evolution



图 8-4 不完全信息时被捕
食者策略-1

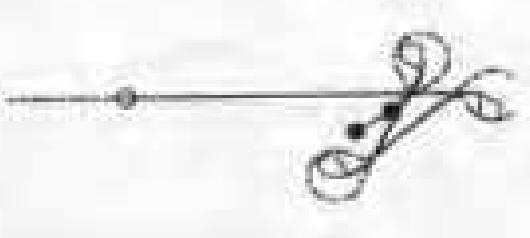


图 8-5 完全信息时被捕
食者策略-3

of pursuit and evasion. I: Biological and game-theoretic foundations. Submitted to *Adaptive Behavior*. Available via the World Wide Web at <http://www.cogs.susx.ac.uk/cgr-bin/htmlcogsreps?csrp311>.

_____. 1997b. Co-evolution of pursuit and evasion. II: Simulation methods and results. In *From Animals to Animats 4, Proceedings of the Fourth International Conference On Simulation of Adaptive Behavior*, edited by Patti Maes et al. Cambridge, MA: MIT Press Bradford Books. Available via the World Wide Web at <http://www.cogs.susx.ac.uk/users/davec/sab96.ps>. Z.

Webb, Paul G. 1986. Locomotion in predator-prey relationships. In *Predator-Prey Relationships*, edited by Martin E. Feder and George V. Lauder, 24~41. Chicago, IL: University of Chicago Press.

§ 8.4 评阅人的评述⁽²⁾

8.4.1 作者介绍

John S. Robertsson 博士是(位于佐治亚州 Milledgeville 的)佐

治亚学院和州立大学,伟大的南方作家 Flannery O'Connor 的母校的数学和计算机科学系的系主任、教授。他把自己描述为一位“指甲下的污物(没有用处的、被人瞧不起的)”应用数学家,而且既深爱数学在其他学科中的应用,也深爱数学本身。他和他的家庭幸福地安居在美国南方腹地(指美国最具南方特点和保守的一片地区,尤指南卡罗来纳、佐治亚、阿拉巴马和密西西比等州),他在那里喜欢观看美国北方冬天降雪的天气预报,一点不漏。

8.4.2 评述文章的主要内容

● 引言

生命科学为科学建模者提供了可以应用其技艺的特别肥沃的领域。物理和工程是应用数学家传统的献艺领域,而生物学及其子学科常被认为是不需要或很少需要数学的领域——如果你愿意的话,可称之为软科学。这种旧观念恰恰不是真的。(只要查阅一下任一卷 UMAP,就会了解这一点。)

对于数学建模者来说古生物学是一个特别肥沃的领域,因为根本没有可利用的观测数据。例如,不可能看到凶暴的恐龙猎获其晚餐或翼手飞龙飞翔在空中。关于古代动物我们所知道的就是以大量推理侦查性工作为基础的事情,这常常需要大量的应用数学的工作。

Michael Crichton 的小说《侏儒纪公园》和《失去的世界》的成功(以及在 Steven Spielberg 精湛导演下的电影的广泛普及)不料却起到了提高许多人,特别是年轻人对古生物学的兴趣。在这种意义上,本问题的时机掌握得不能再好了。

● 假设的重要性

建模过程取决于所作的假设。在 *velociraptor* 和 *thescelosaurus* 的情形,古生物学家基于残留下来的化石的分析提供了这些恐龙的某些物理特征。在分析这些猛禽所采用的策略时古生物学家需要帮助。

所提供的数据暗示着两种恐龙作急转弯——几乎是求生所必须的机动动作——时所具有的不切实际的高度控制力。多数队意识到了这点；但是由于他们不能从古生物学家那里获得修正过的数据，所以他们作了合理而适度的假设。在现实世界中这种困难并不陌生，是否能识别出所提供的数据中所包含的明显的缺陷，并以某种合理而适当的方式作出反应，对于各队能否获得成功是很重要的。例如，不少队利用了从文献中容易得到的在大小和行为特征方面类似于本问题中的恐龙的哺乳类动物的有关资料。这就使这些队能以一种现实的方式来调整数据，从而提高了他们随后的结果为客户所真正利用的可能性。

另一个必须对假设进行讨论的主要问题是有关潜近(猎物)、追猎和捕获的几何和机制。最好的队提供了有关他们的选择的清楚、详细的想法。好的工作不可避免地要引向各队必须面对的解释问题。

● 模型的选择

一旦各队讲清楚了他们的假设，他们就会应用令人惊讶的不同方法来完成数学建模的任务。有些队能够只用代数和几何而不用微积分来形成其模型。另一些队用到了微分方程，递交的最优秀的论文之一用到了微分几何。

在几乎所有的情形，各队转而利用计算机来完成模型的计算。评阅人看到了各种各样的方法。像 Matlab 那样的计算机代数系统是很普及的了。为此，许多队用了一种很好的有点过时的编程语言(大多数情形是 C++ 语言)。

本问题特别适合于这样那样的图形解释，大多数队提供了一个或两个捕食者追猎(猎物)过程的进行和结果的图形和图表。当和(诸如古生物学家那样的)可能尚未掌握所有的专门数学的人一起工作时，图形分析特别重要。古老的谚语——图顶千字——在这种情形确实是对的，因为各队的图解说明对于他们的分析或模型的预测来说是绝对重要的。

● 结果的分析

最优秀的论文就他们的结果和预测给出了透彻的论述。大多数情形，他们考虑了未被其结果所包含的模型的缺点。这些缺点往往要追溯到建模一开始所作的假设。对于建模者来说，这一步也是重要的。顾客并不总能，更谈不上认真地评估缺点的存在。这些缺点不一定是致命的，但是指出哪儿去找更好的数据、更精确的假设或不同的方法是值得鼓励的。这就会导致建模者和顾客之间的重要相互作用。没有这种相互作用整个数学建模过程就失去了中心。

● 结论

优秀论文展示了这些队对所用到的生物学和数学知识，以及把两者结合在数学建模过程中的洞察。评阅人享受着阅读这些优秀论文的美好时光。本试题显然在参赛者中激起了广泛的兴趣。恐龙问题无疑已经激励着许多学生仔细考虑把研究生物学作为数学家的令人激动和富有成果的研究领域。

§ 8.5 评注

8.5.1 微分对策^[3]

对策这一概念有许多推广，微分对策是其中之一，而且出现较早，发展也较成熟。

微分对策是局中人在每一时刻 t 皆要作出一个决定的连续情况，例如追逃问题。追和逃的每时每刻皆要作出某种选择。设在时刻 t ，对局的状态变量（例如，位置、方向、速度等）为

$$x(t) = (x_1(t), x_2(t), \dots, x_n(t)).$$

设在此时，局中人甲选取的策略（方向、速度等）为

$$\varphi = (\varphi_1, \varphi_2, \dots, \varphi_p),$$

其中 $\varphi_i = \varphi_i(t)$ ，一般可设 $a_i \leq \varphi_i \leq b_i$ ， a_i, b_i 是常数；局中人乙选取的策略

$$\psi = (\psi_1, \psi_2, \dots, \psi_q)$$

也满足类似的关系。

φ, ψ 称为控制变量(策略), 他们按照微分方程组

$$\frac{dx_i}{dt} = f_i(x; \varphi; \psi) \quad (i=1, 2, \dots, n)$$

来控制状态变量的运动。当状态变量达到某一给定的闭集时对局即告结束。

寻求最优的 φ, ψ 是微分对策的基本课题之一。

由于本题提供的信息无法确定连续的 $f_i(x; \varphi; \psi)$, 所以只能用离散的方法来做。

有关微分对策的一本重要著作就是[4]。本题的优秀论文中有两篇论文引用了该书。

8.5.2 其他可供参考的优秀论文^{(5)~(9)}

正如评阅人的评注指出的, 这些优秀论文都能对试题提供的信息和数据是否合理、足够提出自己的有根据的看法。例如, 哈佛大学队根据查到的文献指出, 试题中提出碰到 velociraptor 追捕 theselosaurs 的情形的可能性很小, 因为 velociraptor 的化石只在蒙古发现, 而 theselosaurs 的化石只在美国和加拿大的中西部发现。波莫纳学院队则指出 velociraptor 和 theselosaurs 以最大可能线速度转弯不合理, 提出了他们认为比较合理的数据。北京航空航天大学队指出由于缺乏太多的重要数据, 不能精确地描述概率函数的分布, 等等。各队都能尽可能根据自己能查阅到的文献资料和数据提出较为合理的假设, 有的队还考虑了随机性, 并建立相关的数学模型。

在求解数学模型方面各队也各有千秋。哈佛大学队在假设 velociraptor 和 theselosaurs 都认为转弯比减速、停下、改变方向、再加速更为有利的前提下, 定义了一种距离度量, 只用到简单的平面几何。阿拉斯加大学费尔班克斯分校队考虑了随机性(被捕食者幸存概率), 分别建立了数学模型和计算机模拟模型。

密苏里州华盛顿大学队的建模基础是〔4〕中提出的两辆汽车追逐问题的数学理论.

各队都广泛使用计算机图示来展示自己的结果.

参 考 文 献

- 〔1〕 Edward L. Hamilton, Shawn A. Menninga, David Tong (Calvin College), Pursuit-Evasion Games in the Late Cretaceous, UMAP, v. 18 (1997), no. 3, 213~224.
- 〔2〕 John S. Robertson, Judge's Commentary: The Outstanding Velociraptor Papers, UMAP, v. 18 (1998), no. 3, 293~295.
- 〔3〕 对策论, 中国大百科全书——数学卷, 中国大百科全书出版社, 北京, 1988, 138~140.
- 〔4〕 Rufus Isaacs, *Differential Games*, John Wiley & Sons, New York, 1965.
- 〔5〕 Charlene S. Ahn, Edward Boas, Benjamin Rahn (Harvard University), The geometry and the game theory of chases, UMAP, v. 18 (1998), no. 3, 225~242.
- 〔6〕 Hei (Celia) Chan, Robert A. Moody, David Young (Pomona College), Gone huntin': modeling optimal predator and prey strategies, UMAP, v. 18 (1998), no. 3, 243~254.
- 〔7〕 Gordon Bower, Orion Lawler, James Long (University of Alaska Fairbanks), Lunch on the run, UMAP, v. 18 (1998), no. 3, 255~276.
- 〔8〕 Lance Finney, Jade Vinson, Derek Zaba (Washington University), A three-phase model for predator-prey analysis, UMAP, v. 18 (1998), no. 3, 277~292.
- 〔9〕 童中华、俞慧斌、李海明, 追捕与逃跑策略, 数学的实践与认识, v. 28 (1998), no. 2, 178~184.

第九章 会议分组安排

姜启源

清华大学 数学科学系

提 要

本文取材于 1997 年美国大学生数学建模竞赛 B 题及发表在 UMAP vol. 18 No. 3 上的优秀论文^{(1)~(4)}. 按照题目所给的会议分组需使与会人员“充分混合”的原则，建立了整数规划、关联分数、不利函数等数学模型，用改进的贪婪算法、模拟退火等方法求解，讨论了模型推广等问题.

§ 9.1 问题的提出

近来流行开小组会讨论一些重要事宜，如长期计划。因为一般认为大型会议不利于充分讨论，并且容易受某些权势人物的控制和支配。公司在召开全体董事会议之前常常先开小组会。小组会也会有被权势人物支配的危险，为减少这种危险通常把会议分成若干段，在每段让不同组的与会者充分混合。

An Tostal 公司董事会共 29 位成员，其中 9 位是在职董事（即公司雇员）。一天的会议分成上午 3 段和下午 4 段，每段 45 分钟，从上午 9:00 直到下午 4:00。上午每段分 6 个组，每组由一位资深职员（非董事）主持，于是每位资深职员要主持 3 个不同的小组会。下午每段分 4 个组，没有资深职员主持。

董事长希望得到一份 7 段小组会的分组名单安排计划，这种安排应将各位董事尽可能地混合。理想的安排是使每一董事和另

一董事在同一小组中开会的次数相同，并使不同段的小组中一起开会的董事最少。

安排计划要满足以下准则：

1. 上午的 3 段不允许任一董事参加同一位资深职员主持的两次会议；

2. 在职董事均匀地分配在每段的各小组中。

给出一份 1~9 号在职董事、10~29 号董事和 1~6 号资深职员的分组名单，说明它在多大程度上满足上述准则要求。因为有可能有的董事在最后一分钟宣布不参加会议，或者不在名单上的董事表示要出席，所以秘书需要一个临时调整的办法。如果算法能够用于不同类型、不同水平的与会者的会议安排，当然更好。

这道题共有 4 队获优秀奖：中国华东理工大学的一个队（下记 A 队），Macalester 学院的一个队（B 队），Rose-Hulman 科技学院的一个队（C 队），加拿大 Toronto 大学的一个队（D 队）。本文拟综合介绍这 4 篇优秀论文。^{[1]~[4]}

本质上这是一个优化问题，确定目标函数和约束条件以构成模型是关键的第一步。下面 § 9.2~§ 9.4 是模型建立、求解及检验和推广等内容，每一部分分别介绍 4 篇优秀论文有特色的地方，§ 9.5 为评阅者的意见。

§ 9.2 模型建立

1. 混合整数规划模型（A 队）

题目所需要的分组名单安排计划属于分派问题，常用 0~1 变量表示分派的决策变量，令

$$x_{itk} = \begin{cases} 1, & \text{第 } i \text{ 董事在第 } l \text{ 段分入第 } k \text{ 组;} \\ 0, & \text{否则.} \end{cases}$$

$i=1,2,\dots,29$, 其中 $1,2,\dots,9$ 为在职董事; $t=1,2,\dots,7$, 其中 $1,2,3$ 为上午, $4,5,6,7$ 为下午; $k=1,2,\dots,6$, 且当 $t=1,2,3$ 时 $k=1,2,\dots,6$; 当 $t=4,5,6,7$ 时, $k=1,2,3,4$.

按照题目所给的准则及分派问题的当然需要, 容易写出模型的约束条件:

a) 按准则 1, 每位董事 i 上午的 3 段 ($t=1, 2, 3$) 不应参加同一资深职员 k (每一资深职员对应一组, $k=1, 2, \dots, 6$) 主持的 2 次会, 于是

$$0 \leq \sum_{t=1}^3 x_{it} \leq 1, \quad i = 1, \dots, 29, \quad k = 1, \dots, 6 \quad (1)$$

b) 按准则 2, 9 位在职董事 ($i=1, \dots, 9$) 要均匀分配在各小组中, 上午 6 个组每组只能 1 或 2 位; 下午 4 个组每组只能 2 或 3 位, 于是

$$\begin{cases} 1 \leq \sum_{i=1}^9 x_{ik} \leq 2, & t = 1, 2, 3, \quad k = 1, 2, \dots, 6; \\ 2 \leq \sum_{i=1}^9 x_{ik} \leq 3, & t = 4, \dots, 7, \quad k = 1, 2, 3, 4. \end{cases} \quad (2)$$

c) 每位董事在每一段必须且只能分入一组, 于是

$$\begin{cases} \sum_{k=1}^6 x_{ik} = 1, & i = 1, \dots, 29, \quad t = 1, 2, 3; \\ \sum_{k=1}^4 x_{ik} = 1, & i = 1, \dots, 29, \quad t = 4, 5, 6, 7. \end{cases} \quad (3)$$

d) 直观上在每一段中 29 位董事分配得越均匀, 整体考虑时才可能混合得越好 (可以从反面设想 29 人全在一组的极端情况), 即在上午每段的 6 组中应分别有 5, 5, 5, 5, 5, 4 人, 下午每段的 4 组中应分别有 7, 7, 7, 8 人, 于是

$$\begin{cases} 4 \leq \sum_{i=1}^{29} x_{ik} \leq 5, & t = 1, 2, 3, \quad k = 1, \dots, 6; \\ 7 \leq \sum_{i=1}^{29} x_{ik} \leq 8, & t = 4, \dots, 7, \quad k = 1, \dots, 4. \end{cases} \quad (4)$$

在确定目标函数之前，先计算一下每对董事的平均相遇（即在 7 段会议中参加同一组会）次数。因为按照约束条件 d （即(4)式），为一对董事相遇提供的机会数为

$$3 \times (5C_5^2 + C_4^2) + 4(3C_7^2 + C_8^2) = 532, \quad (5)$$

而组合成对的董事共有 $C_{29}^2 = 406$ 对，所以每对董事平均相遇次数为 $\bar{q} = 532/406 \approx 1.3$ 。这个结果表明，为使分组计划安排得好，即混合均匀，理想的情况应是多数董事对只相遇 1 次，少数对相遇 2 次，没有对相遇 3 次、4 次…的，也没有对相遇 0 次，即不相遇的，因为如果有不相遇的对，那么必然要增加相遇 2 次、3 次…的对。

基于以上分析，模型以不相遇的对数尽量少，及相遇次数与平均值 $\bar{q}=1.3$ 的差尽量小为目标。为了写出目标函数，记某对董事 (i, j) 相遇的次数为 q_{ij} ，显然有

$$q_{ij} = \sum_{l=1}^3 \sum_{k=1}^6 x_{ik} x_{jk} + \sum_{l=4}^7 \sum_{k=1}^4 x_{ik} x_{jk}, \quad i, j = 1, \dots, 29, i \neq j. \quad (6)$$

当 $i=j$ 时定义 $q_{ii}=0$ ，且记 $Q=\{q_{ij}\}_{29 \times 29}$ 。再令

$$p_{ij} = \begin{cases} 1, & q_{ij}=0, \\ 0, & q_{ij} \neq 0, \end{cases} \quad i, j=1, \dots, 29. \quad (7)$$

$$T=Q-\bar{q}(E-I), \quad (8)$$

其中 E 为全 1 矩阵， I 为单位矩阵， $T=\{t_{ij}\}_{29 \times 29}$ ， t_{ij} 表示董事对 (i, j) 的相遇次数与平均值 \bar{q} 之差。

两个目标函数定义为

$$f(x) = \sum_{i=1}^{29} \sum_{j=1}^{29} p_{ij}, \quad (9)$$

$$g(x) = \sum_{i=1}^{29} \sum_{j=1}^{29} t_{ij}^2. \quad (10)$$

这里 x 表示 $\{x_{ik}\}$ 。

2. 构造不利函数模型(B 队)

按照对题目所给准则和要求的损害程度构造目标函数，称不利函数 (Badness)，由以下 4 个子函数加权组合而成，求解极小化不利函数的优化问题。

a) 各位董事参加同一位资深职员主持的 2 次(及 2 次以上)小组会的总数，记作 f_1 。按准则 1，不许出现这种情况，所以希望 $f_1 = 0$ 。

b) 在职董事的不均匀数，记作 f_2 。在职董事的均匀分配结果对上午的 3 段应是 1 或 2，对下午的 4 段应是 2 或 3，不均匀数 f_2 定义为与这些数字的差(绝对值)。按准则 2，应该均匀分配，所以希望 $f_2 = 0$ 。

c) 董事对相遇的异常数，记作 f_3 。按要求，每对董事相遇(在同一小组)的次数尽量相同，推导出理想次数是 1 和 2，将相遇次数异于 1 或 2 的董事对数求和，得到 f_3 。

董事对相遇的最大数，记作 f_4 。它衡量同一对董事反复在同一小组出现的严重程度。 f_3 和 f_4 都只能希望尽量小，而不可能等于 0。

d) 两组中含有共同董事数的异常数 f_5 ，及最大共同董事数 f_6 。这两个指标与 f_3 ， f_4 相似，不过它们是从小组的角度而不是董事的角度考虑。计算方法是先算出平均数，只将大于平均数的组(对)数求和得到 f_5 ，最大数为 f_6 。

目标函数(不利函数)定义为

$$f = \sum_{i=1}^6 w_i f_i - \lambda_1 g_1 - \lambda_2 g_2, \quad (11)$$

其中 w_i 是 f_i 的权， g_1 ， g_2 定义为

$$g_1 = \begin{cases} 1, & f_1 = 0, \\ 0, & f_1 \neq 0; \end{cases} \quad g_2 = \begin{cases} 1, & f_2 = 0, \\ 0, & f_2 \neq 0. \end{cases} \quad (12)$$

λ_1 ， λ_2 是 g_1 ， g_2 的权。(12)式表明，在目标函数(11)式中加入 g_1 ， g_2 是对 $f_1 = 0$ ， $f_2 = 0$ 的额外“奖励”。通过 $-\lambda_1 g_1 - \lambda_2 g_2$ 降低满足准则的不利函数的数值。

权由试探和修正确定，下面是得到的结果： $w_1 = \lambda_1 = 1200$ ， $w_2 = \lambda_2 = 1000$ ， $w_3 = 400$ ， $w_4 = 4000$ ， $w_5 = 100$ ， $w_6 = 500$ 。以这组数值为权，得到的安排方案中， $f_1 = f_2 = 0$ （即准则 1, 2 均满足），董事对相遇最大数 f_4 不超过 3，两组中最大共同董事数 f_6 也不超过 3。

对权的几点说明：

- 所谓额外“奖励”的权 λ_1 , λ_2 没有另外调整，而是令 $\lambda_1 = w_1$, $\lambda_2 = w_2$ ，这样做效果不错。
- $w_3 : w_4 = 1 : 10$ ，意味着算法对董事对相遇异常数和最大数的重要性之比为 1 : 10。
- w_5 , w_6 与 w_3 , w_4 相比小得多，是因为发现 f_5 , f_6 与 f_3 , f_4 有很强的相关性，不用再给予很大注意。

3. 构造另一种不利函数模型(C 队)

董事对相遇 i 次的总数记为 e_i ($i=2, 3, \dots$)，在职董事分配的不均匀性用任两组中在职董事数相差超过 1 的数量 d 表示。目标函数(不利函数)定义为

$$b(s) = 1000d + \sum_{i=2}^{\infty} e_i 4^{i-2}, \quad (13)$$

这里 s 表示一种分组安排。

4. 关联分数模型(D 队)

董事 i 与 j 的相遇次数记作 a_{ij} ($i, j=1, \dots, N (=29)$, $i < j$) 称关联数。从满足题目要求看， a_{ij} 最好取 1 或 2，若用

$$r = \sum_{1 \leq i < j \leq N} a_{ij} \quad (14)$$

(r 称关联和)作为目标函数，可能会得到 a_{ij} 取 0, 3, 4, … 的安排；若用 a_{ij} 的方差或标准差作为目标函数，只能使 a_{ij} 相互靠近，不能使它们均匀地小。作为折衷，这里取

$$s = \sum_{1 \leq i < j \leq N} a_{ij}^2 \quad (15)$$

为目标函数，称关联分数。容易看出当 s 小时关联和 r 也会变

小。

下面求 s 的理想下界。

与 A 队的分析一样，得到平均相遇次数为 1.3，所以理想安排的 a_{ij} 只取 1, 2 二值，记 $d=1$ 和 $d+1=2$ 。设总共 C_N^2 个董事对中有 c_1 对的关联数 a_{ij} 取 $d=1$ ，余下 c_2 对的关联数 a_{ij} 取 $d+1=2$ 。于是可得

$$\begin{cases} c_1 + c_2 = C_N^2, \\ c_1 d + c_2 (d+1) = r, \\ c_1 d^2 + c_2 (d+1)^2 = s. \end{cases} \quad (16)$$

由此算出关联分数 s 的下界

$$s_{\min} = (2d+1)r - d(d+1)C_N^2. \quad (17)$$

在尽可能均匀分组的情况下（即上午每组 5, 5, 5, 5, 5, 4 人，下午 7, 7, 7, 8 人），可得 $r=532$ （见 A 队(5)式）。再将 $d=1$, $N=29$ 代入(17)式得 $s_{\min}=784$ 。

在用 s 为目标函数寻最优解时，可以与 s_{\min} 比较，衡量所得解的优劣。

§ 9.3 ·模型求解

1. 算法实现

1) 对于混合整数规划模型 (A 队)，采用坐标轮换与级别优先法。

由于目标函数是非线性的，该模型没有一般解法，用枚举法有 986 个变量，至少要比较 6^{986} 个解，这是不可能的。这里先寻找一个好的可行解，再迭代调整它，逼近最优解。

a. 找一初始可行解。将一个个董事依次分入一组，在每次分配之前检查他分入哪一组最好。

b. 迭代。先将第 1 个目标函数 $f(x)$ ((9)式)调整到尽可能小，再极小化第 2 个目标函数 $g(x)$ ((10)式)。

对 $l=7$, 使 $f(x)$ 尽可能小, 然后通过对换 i, j (见 c) 在不影响 $f(x)$ 条件下降低 $g(x)$; 对 $l=6, 5, \dots$ 重复之.

c. 对换. 将不在同一组的二董事 i, j 对换, 若能使 f, g 减少, 则接受; 否则, 舍弃之.

2) 对不利函数模型(B队)采用模拟退火法(Simulated Annealing).

在比较了诸如启发式方法(直观加经验)、梯度算法、整数规划、遗传算法之后, 从快捷、简单和该队的经历等方面决定选用模拟退火法, 该法的具体实施请参阅专门的书籍, 此处从略.

3) 对另一种不利函数模型(C队)采用改进的贪婪算法.

a. 按 $l=1, 2, \dots, 7$ 的顺序将在职董事进行尽可能均匀的分配;

b. 仍按上述 l 的顺序将其余董事分配入各组, 使目标函数尽可能小;

c. 对同一 l , 将董事 i 与 j 对换, 考察目标函数是否继续下降.

编者注: 作者认为步骤 a, b 是贪婪算法, c 是对它的改进. 实际上 a, b 只是在同一 l 内的枚举, 并未作全盘的枚举.

C队还对用随机分配、贪婪算法和改进的贪婪算法得到的结果作了统计分析和比较.

4) 对关联分数模型(D队)采用另一种改进的贪婪算法.

与上面的贪婪算法稍有不同, 这里是按董事的序号迭代地进行;

a. 按 $i=1, 2, \dots, 29$ 的顺序, 在满足两个准则的条件下将董事一个个地分为所有各段的组中, 使关联分数尽量小.

b. 对于每次迭代(即某个 i 的分配)进行调整, 即董事 i 与 j 的对调(i, j 同属在职董事, 或同属非在职董事), 或将 i 从一组分入另一组.

2. 基本结果

利用上面的模型和算法, 4个队都给出了最终结果, 如表

9-1, 9-2, 9-3, 9-4 所示。

表 9-1 模型 1 (A 队) 的分配名单 ($i=1, 2, \dots, 29$ 分入每一段的各组内)

	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	第 6 组
上午 第 1 段	1 6 13 19 24	2 3 14 25 26	5 8 10 15 27	7 16 20 22 28	9 11 21 23 29	4 12 17 18
上午 第 2 段	3 10 11 26 28	5 12 19 22 23	6 16 20 21 29	1 2 17 24 27	4 7 15 18 25	8 9 13 14
上午 第 3 段	5 7 14 17 29	4 13 15 16 21	1 9 19 26 28	3 11 18 23	6 8 12 25 27	2 10 20 22 24
下午 第 1 段	4 6 11 14 20 23 27	7 8 9 16 18 24 26	2 5 17 19 21 25 28	1 3 10 12 13 15 22 29		
下午 第 2 段	1 4 10 14 16 23 25	5 6 11 15 17 22 26	2 8 13 18 19 27 28 29	3 7 9 12 20 21 24		
下午 第 3 段	6 9 10 18 21 22 27	1 2 11 12 14 15 16	7 8 13 17 20 23 26 19	3 4 5 24 25 28 29		
下午 第 4 段	1 5 11 13 18 20 25	2 6 9 15 23 24 28	3 7 10 16 17 19 27	4 8 12 14 21 22 26 29		

这个结果满足模型 1 的约束条件(1)~(4)式, 且使目标函数 $f(x)=81, g(x)=19.44$. 董事对相遇 0, 1, 2, 3 次的对数分别为 26, 253, 102, 23, 没有相遇 4 次及 4 次以上的.

表 9-2 模型 2 (B 队) 的分配名单

	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	第 6 组
上午 第 1 段	1 17 20 22 27	3 4 13 21 28	8 10 11 18 29	7 14 19 23 24	5 9 12 15 16	2 6 25 26
上午 第 2 段	2 6 13 28 29	9 10 16 19 25	7 12 20 22 23	3 5 11 17 26	4 8 14 20 27	1 15 21 24
上午 第 3 段	5 8 16 19 21	1 2 18 22 23	9 14 15 17 26	4 12 25 27 29	2 6 11 13 20	3 7 10 28
下午 第 1 段	2 8 10 20 24 26 27	3 4 6 15 18 19 22	1 5 11 14 16 23 25 28	7 9 12 13 17 21 29		
下午 第 2 段	5 6 10 12 21 23 27	2 3 14 15 20 25 29	1 8 9 13 19 22 26 28	4 7 11 16 17 18 24		
下午 第 3 段	1 2 4 10 13 15 16	5 8 12 17 22 24 25	6 7 14 18 20 21 28	3 9 11 19 23 27 29		
下午 第 4 段	5 7 13 18 19 25 27	2 9 10 11 14 21 22	1 3 6 12 16 24 26 29	4 8 15 17 20 23 28		

这个结果满足题目所给的 2 个准则, 董事对相遇 0, 1, 2, 3 次的对数分别为 40, 214, 138, 14. 没有相遇 4 次及以上的, 于是相遇异常数 $f_3 = 54 (= 40 + 14)$. 两组含有共同董事数也不超过 2.

表 9-3

模型 3 (C 队) 的分配名单

	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	第 6 组
上午 第 1 段	1 3 10 12 26	2 8 17 21 25	6 13 22 28	5 11 18 23 27	4 14 15 19 24	7 9 16 20 29
上午 第 2 段	7 11 13 17 24	1 4 15 22 27	6 9 16 18 26	6 10 14 20 25	3 5 21 28 29	2 12 19 23
上午 第 3 段	4 5 25 29 26	6 7 11 19 20 23	3 15 17 21 24	9 12 13 18 22	2 10 16 18 22	1 8 14 27 28
下午 第 1 段	2 3 6 18 24 26 27 29	8 9 11 12 14 15 22 25	1 5 13 16 19 20 21	4 7 10 17 23 28		
下午 第 2 段	5 6 7 12 15 18 21	2 4 11 14 20 26 28	1 9 17 19 22 24 29	3 8 10 13 16 23 25 27		
下午 第 3 段	5 8 20 22 23 24 26	3 7 9 18 19 25 28	1 2 10 11 13 15 29	4 6 12 14 16 17 21 27		
下午 第 4 段	5 9 10 15 17 26 27	4 8 12 13 18 19 20 29	1 6 11 16 23 24 25 28	2 3 7 14 21 22		

这是从 679 次计算中选出的最好结果，目标函数 $b(s)=168$ ((13)式)。

表 9-4 模型 4 (D 队) 的分配名单

	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	第 6 组
上午 第 1 段	1 4 14 22 25	2 9 12 21 28	3 10 15 23 27	5 6 18 20 26	7 11 17 24 29	8 13 16 19
上午 第 2 段	2 7 18 19 27	1 8 13 23 25	4 9 17 20 29	3 11 16 22 28	5 10 14 21 26	6 12 15 24
上午 第 3 段	3 9 17 23 26	4 10 19 20 24	1 11 12 18	3 13 14 15 29	6 8 16 22 27	5 7 21 25 28
下午 第 1 段	1 5 9 16 23 24 27	2 6 10 13 17 22 26 28	3 7 12 14 19 20 25	4 8 11 15 18 21 29		
下午 第 2 段	1 3 6 17 19 21 27 29	2 5 11 15 20 22	8 9 10 14 18 24 25 28	4 7 12 13 16 23 26		
下午 第 3 段	1 2 10 15 16 17 25	7 8 9 20 21 26 27	3 5 12 13 18 22 24 29	4 6 11 14 19 23 28		
下午 第 4 段	1 7 15 19 22 26 28 29	2 3 4 16 18 21 24	6 9 10 11 13 25 27	5 8 12 14 17 20 23		

这个结果满足所有准则，董事对相遇 0, 1, 2, 3 次的对数分别为 33, 226, 134, 13。没有相遇 4 次及以上的。目标函数——关联分数 $s = 879$ (s 的下界 $s_{\min} = 784$, 见(15)、(17)式)。

3. 临时调整

当董事临时改变出席会议计划时的调整办法，各队相差不大，下面介绍其中两种：

1) A 队的处理办法

a. 当有人要加入时，可一组组试探，看看分配在哪一组可在满足约束条件下目标函数最小。

b. 当有人要退出时，试探每一组与退出者相同类型的董事（均为在职，或均为非在职），看看哪一组中的董事退出后使目标函数下降最多，就用该位董事取代那位退出者。

c. 既有要加入者也有要退出者。若同类型董事中加入人数为 r_1 ，退出人数为 r_2 ，当 $r_1 = r_2$ 时对换即可， $r_1 > r_2$ 时采用程序 a， $r_1 < r_2$ 时采用程序 b。

2) C 队的处理办法

先统计同一类型董事中要加入者和要退出者的人数，去掉可以对换的外，余下的如要安排加入者，则对每一段会议先排除同一资深职员主持 2 次的那些组，再排除人数较多的那些组，最后在剩下的组中寻求加入后使相遇次数最少的那个组。如要安排退出者，办法与 A 队的程序 b 相同。

§ 9.4 模型检验与推广

各队给出了检验的原则和推广的办法，而在实际计算方面作的比较多的是 B 队的文章，现简介如下：

1) 改变董事数目。将董事总数由 29 逐步增至 100，计算时间随之增加的结果如下(以本题的计算时间为 1)：

在职董事数	非在职董事数	总 数	计算时间
9	20	29	1
15	33	48	1.5
18	41	59	3.3
30	70	100	8.75

在总数 29 的情形还将在职与非在职董事数改变为(4, 25)与

(14, 15), 都可得到在职董事不均匀数 $f_2=0$, 和董事对相遇最大数, 两组含有共同董事最大数均为 3 的结果.

2) 改变段数及每段的组数. 段数改变到上午 1 段、下午 6 段, 每段的组数改变到 2 组, 基本上都得到与原来一样好的结果. 惟一的例外出现在 3 段会议且每段由 3 位资深职员主持 3 组时, 很难得到 6 组时 $f_1=0$ 的结果, 当数 $w_1=2000$ 时才有较好的结果.

3) 个别董事参加会议的特殊模式作了如下试验: 某些在职董事不参加下午的会议; 某些非在职董事不参加上午的会议; 某些新成员只参加上午的一段会议. 这些情形均得到了与原来类似的结果.

§ 9.5 “充分混合”的度量标准

作为这题的提供者和评阅人之一, Donald E. Miller 撰文发表了如下意见⁽⁵⁾.

这是一个没有标准答案的“充分混合”(good mix)问题, 对于什么是“充分”, 题目并未提出明确定义, 因此参赛者必须首先对结果的“充分性”(goodness)给出度量方法.

为建立这种度量标准, 即目标函数, 必须作出某些假设, 这些假设可以从回答以下问题得出:

- 两位董事第 3 次相遇比第 2 次坏吗?
- 两位在职董事第 2 次相遇比两位非在职董事第 2 次相遇坏吗?
- 如何评估一位在职董事和一位非在职董事的第 2 次相遇?
- 两位董事在某两段的小组会相遇, 如果将这两段相隔得远些(如分在上午和下午), 能减少不利因素吗?
- 小组内共同董事是 A—B—C 形式(有 3 位共同成员)比 A—B 和 C—D 形式(两对两位共同成员)坏吗?

参赛者作出的假设可以多种多样，评阅者会很重视对这些假设的衡量。

论文共同的缺点是不能同时提供模型与解法。有的给出了一个拼凑出来的解或模拟解而没有模型，有的给出了有新意的模型却没有表现出解决这道题目的功效。如果只对题目本身凑出一个结果，那么论文在评阅过程的初期就会被剔除。

虽然有各种解法，最常用的是贪婪算法和模拟退火，但是用贪婪算法有的队得到是局部最优——只在一段的水平上选优。

4篇优秀论文有不少相似之处，他们都希望各组的董事数尽量平均，于是得到需至少安排532次董事对相遇机会，而董事对共有406个。将4篇论文提出的结果中董事对相遇次数为0, 1, 2, 3的董事对数目列入表9-5作一比较：

表9-5 4篇论文最终结果的比较

	相遇次数				总相遇次数
	0	1	2	3	
A队	26	253	102	25	532
B队	40	214	138	14	532
C队	32	218	152	4	534
D队	33	226	134	13	533

可以看出，尽管4个队的目标函数不同，而他们的结果却类似。C队的结果中相遇3次的董事对只有4个，因为目标函数中对相遇次数加的权重是4的幂。

B队论文的优点在于它对模拟退火如何用于解决本问题的说明，以及目标函数的全面考虑。C队论文的优点在于3种解法的统计分析和比较，以及好的文字表达。D队则提供了一些有关解的界的证明。

参 考 文 献

- (1) Han Cao, Hui Yang, Zheng Shi; An Assignment Model for Fruitful Discussions. UMAP vol. 18 No. 3
- (2) David Castro, John Renze, Nicholas Weininger; Using Simulated Annealing to Solve the Discussion Groups Problem. UMAP vol. 18 No. 3
- (3) Joshua M. Horstman, Jamie Kawabata, James C. Moore, IV; Meetings, Bloody Meetings! UMAP vol. 18 No. 3
- (4) Adrian Corduneanu, Cyrus C. Hsia, Ryan O'Donnell; A Greedy Algorithm for Solving Meeting Mixing Problems; UMAP vol. 18. No. 3
- (5) Donald E. Miller; Judge's Commentary: The Outstanding Discussion Groups Papers. UMAP vol. 18 No. 3

第十章 扫描问题

唐 云

清华大学 数学科学系

提 要

本章介绍了 1998 年美国大学生数学建模竞赛 (MCM-1998) 的竞赛情况、评阅和奖励，特别介绍了 A 题及其优秀答案，并对本题的解答进行评述。

§ 10.1 MCM-1998 的评阅、结果和奖励

本次竞赛共有包括美国、中国、香港等 8 个国家和地区的 246 所大学的 472 个队参加，其中中国有 45 所大学的 138 个队参加。

各队的论文在 COMAP 的总部进行编号使得评阅人不知道论文作者的姓名和所属的学校。

A 题的初评是在康涅狄格州的南康涅狄格大学进行的，共有 5 位评阅人。B 题的初评是在蒙大拿州的卡罗尔 (Carroll) 学院进行的，共有 6 位评阅人。每篇论文由两个初评评阅人评阅，摘要和论文的组织是论文评定的基础。如果两个评阅人的评分不同则进行协商，如果协商后还不一致，则再由第三位评阅人来评阅。

终评是在加州的哈维·马德 (Harvey Mudd) 学院进行的，A 题评阅人有 12 位，B 题评阅人有 17 位。

MCM-1998A 题是由华盛顿州东华盛顿大学的 Yves Niev-

ergelt 提供的, MCM-1998B 题是由位于马萨诸塞州阿灵顿的 Zwillinger & Associate 的 Daniel Zwillinger 提供的.

评出的最后结果是:

	O	M	H	P	合计
MCM-1998A 题获奖队数(中国队数)	4(1)	31(13)	47(22)	106(28)	189(64)
MCM-1998B 题获奖队数(中国队数)	3(0)	48(8)	69(25)	163(43)	283(74)

其中, O=Outstanding=特等奖, M=Meritorious=一等奖, H=Honorable Mention=二等奖, P=Successful Participant=成功参赛奖.

每个参赛队的指导教师和队员都将获得由竞赛主任和每题的评阅组长签名的证书. 美国运筹学和管理科学学会 (ORSA) 给予两个获得特等奖的队队员现金奖励和三年的会员资格. 这两个队分别是明尼苏达州的麦卡莱斯特(Macalester)学院队(A 题)和佛罗里达州的斯特森(Stetson)学院队(B 题). 此外, 美国运筹学和管理科学学会还给获一、二等奖的队的每个队员一年的免费会员资格.

美国工业与应用数学学会 (SIAM) 对每题指定一个特等奖队作为 SIAM 的获奖队, 这两个队分别是明尼苏达州的麦卡莱斯特(Macalester)学院队(A 题)和加州的哈维·马德(Harvey Mudd)学院队(B 题). Harvey Mudd 学院队将于 1998 年 7 月在加拿大 Toronto 举行的 SIAM 年会特设的分组会上报告他们的结果, 他们队的每个队员也将获得 300 美元的现金奖励. Harvey Mudd 学院获得一个装在烫金镜框里亲笔签名的证书.

美国数学协会(MAA)对每题指定一个特等奖队作为 MAA 的获奖队. 他们是俄勒冈州的东俄勒冈大学队(A 题)和北卡罗来纳州杜克(Duke)大学队(B 题). 两个队将在 1998 年 7 月在加拿大多伦多(Toronto)举行的 MAA 的数学节(Mathfest)上报告他

们的解决。MAA 的当选理事长 Tom Banchoff 将授予每个队员证书。

§ 10.2 问题和竞赛结果

本节先介绍扫描问题的全文，然后介绍问题的背景并作初步分析，最后介绍本题的竞赛结果。

10.2.1 扫描问题

引言

被称为磁共振成像仪(MRI)的工业与医用诊断机对像脑那样的三维物体进行扫描，并将扫描的结果以三维像素阵列的形式传送。每个像素由一个标志其颜色或灰度的数构成，它对被扫描物体中像素所在的位置处的一个小区域内含水量(浓度)的度量进行编码。例如，0能以黑色描绘出高含水量(脑室、血管)，128能以灰色来描述出中等含水量(脑核和灰质)，而255能以白色来描述出低含水量(组成有髓体轴的浓脂白质)。这种磁共振成像扫描仪还包括能在屏幕上画出通过该三维像素阵列的平行或垂直片(与三个笛卡儿坐标轴平行的切片)的设备。

但是，能通过斜片画出切片的算法是专卖的。眼下的算法是：

- 在利用角度及所提供的参数选择方面受到限制，
- 大量使用专用的工作站才能执行，
- 缺乏在切片前的画面上作点的输入能力，以及
- 使原始像素之间明晰的边界变得模糊并“减弱”。

一个能在个人计算机上实现的更为可靠的、灵活的算法对于以下几个方面来说是很有用的：

- 设计尽可能少的技术处理；
- 校准磁共振成像仪；

- 研究诸如动物研究中尸体解剖组织部分那样在空间中的斜向结构；
- 使斜面能以任意角度与黑白图线组成的脑图谱相交。

为设计出这样的算法，要能存取像素的值和位置，而不是仅由扫描仪收集到的原始数据。

问题

设计并测试一种算法，它能沿空间中任意指向的平面产生三维阵列的截面，并尽可能保持原始的灰度值。

数据集

典型的数据集由数 $A(i, j, k)$ 的三维阵列 A 组成，该数 $A(i, j, k)$ 表示物体在位于 $(x, y, z)_{i,j,k}$ 处的浓度。 $A(i, j, k)$ 通常的取值范围可以从 0 到 255。在大多数的应用中，该数据集是相当大的。参赛队要设计出用于测试并论证其算法的数据集。数据集应能反映有可能具有诊断意义的状况。参赛队还应描述使其算法有效性受到限制的数据集的特性。

总结

算法必须生成由空间中一个平面与三维阵列相交的切片图像。这种平面在两空间中可以有任意的指向和位置（该平面可能会漏掉一些或全部的数据点）。算法的结果应是所扫描的物体在所选平面上的一个浓度模型。

10.2.2 问题的背景和初步分析

磁共振成像仪对身体的各个部位进行扫描，以诊断体内某些疾病和异常，这种手段已越来越被医学界所采用。其中一个最常用的途径是对脑中的异常体，如肿瘤进行探测。由于健康组织和肿瘤出现的形态不同，在取截面时，边界的清晰度以及区域的光滑程度就显得尤其重要。本问题就是针对这类课题提出来的。本问题的命题人是美国东华盛顿大学数学系 Nievergelt 教授，他就曾指出这个问题的提出与有脑疾的猴子做大脑内的药物注射试验

有关。

对本问题的解答通常是从插值法入手，但插值法一般要求数据是连续变化的，而在机体各组织间，特别是异常物的边界，其水密度的分布是有间断的，这就需要对插值法予以改进，使之适应这种情形，这也是解决本问题的关键。

10.2.3 竞赛结果

在参赛并通过的 472 个队中有 189 个队选择了做扫描问题这道题，其中获特等奖 (Outstanding) 的有 4 个队，获一等奖 (Meritorious) 的有 11 个队，获二等奖 (Honorable Mention) 的有 47 个队。获特等奖的 4 个队分别来自美国的 Eastern Oregon 大学、Harvey Mudd 学院、Macalester 学院和中国的清华大学。在获一等奖的队中，中国有 13 个队，他们是：华东理工大学、复旦大学、南开大学、(长沙) 国防科技大学 (2 个队)、华南理工大学、东南大学 (2 个队)、清华大学、中国科技大学、西安交通大学 (2 个队) 和西安电子科技大学。

10.2.4 本章其余各节的安排

在 § 10.3 到 § 10.6 中我们将依次介绍上述获特等奖的优秀论文，在 § 10.7 中对上述这些论文予以评述。在介绍优秀论文时对原文作了某些改动。

§ 10.3 三维网格数据截面的取法

本节我们介绍美国的 Eastern Oregon 大学获特等奖的优秀论文。

10.3.1 摘要

有效的三维磁共振成像 (MRI) 要求用一种精确的方式来取平

面截面。但若取定一种斜截面，该平面可能不与任何已知的数据点相交，因而需要有一种方法来对数据点之间的水密度进行插值。

插值法假定密度是连续的，但人体内不同类型组织的边界是不连续的。大多数的插值法都是想对这些清晰的边界作磨光处理，这使得数据变得模糊，并且有可能毁掉有用的信息。

为定性地抓住这个问题的关键难点，我们创造一套生物模拟数据集，如脑与手，每个都有它特别的缺陷。数据集为每边有 100 个元素，总共一百万个元素的三维阵列，在每一点用一个范围在 [0,255] 的整数来标记水密度。对每个数据集，用一些可微函数来描述互相之间不连续的几种类型组织。

为分析这些数据，我们创立了一套算法，用 C++ 执行，并比较它们在生成精确截面时的有效性。我们用的是局部插值法，因为数据从总体水平上是不连续的。最终的算法是寻找组织之间的间断处。如果在一点找到，就将最近点的水密度指定为该点的密度，以保持其清晰的边界。若不存在间断，该算法就将三维的多项式拟合到最靠近的 64 个数据点，并对水密度进行插值。

我们通过在截面的每一点寻找在插值的水密度和实际水密度之间的平均绝对差来衡量算法的精确性。最终的算法其误差低于单纯的最近点法的 16%，低于连续线性插值法的 17%，低于不探测间断的连续多项式插值法的 22%。

假设

- MRI 是一个具有等间隔的数据网格，取为 $100 \times 100 \times 100$ 阵列。
- 每个阵列元是一个表示该点水密度的整数，其范围从 0 到 255。
- 所取的截面分辨率等于数据集的分辨率。（如果阵列元具有 1 微米的间隔，则该截面应有同样的间隔。）
- 我们承认，所有的插值法都假定数据点之间是连续的，因而

假定生命组织内的水密度可以表示成在组织之间有间断的连续可微函数。

10.3.2 模拟数据集

因为只能找得几个已有的三维阵列数据集，我们来构造模拟数据集。实际的生物器官是极其复杂的，一组模拟器官应能定性地表示出这类问题，使 MRI 扫描典型地用于研究。虽然现实的 MRI 数据集有着更大的分辨率，所寻求的特征应是相同的：肿瘤、骨折或普通畸形。一般来说，比起周围的组织来，这些发生间断的区域有着不同的水密度。我们提出下面一些有缺陷的模仿器官：

1. 小体：连续的、重复的球型，在中央达到密度峰值（图 10-1）。
2. 手臂：具有两根骨的光滑组织，其中一根有一小球状孔（图 10-2）。
3. 腹类器官：充满着一些间断区域的圆柱状体（图 10-3）。
4. 脑：紧密的头盖骨、周期变化着的灰质，以及在一突起部分有着不同密度的小区域（图 10-4）。

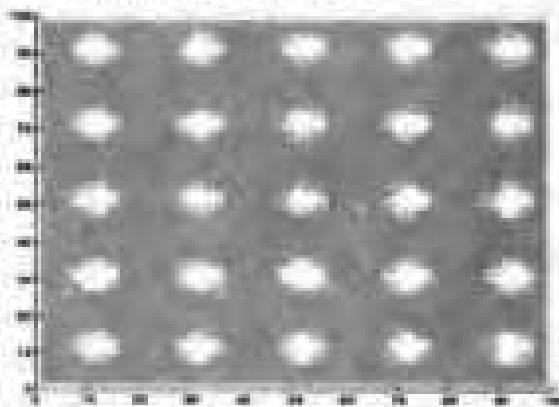


图 10-1 小体
坐标系与定义

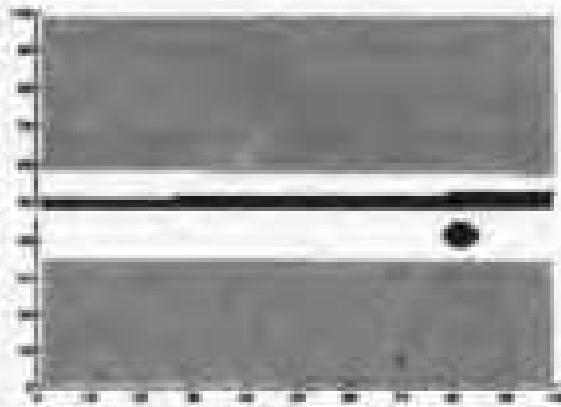


图 10-2 手臂

通过已有的数据阵列在该问题中加上笛卡儿坐标系。若每个

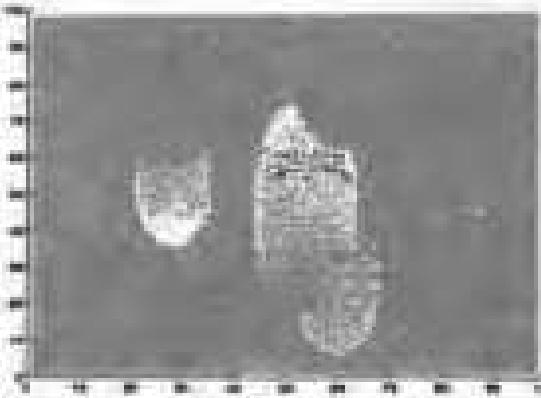


图 10-3 肝

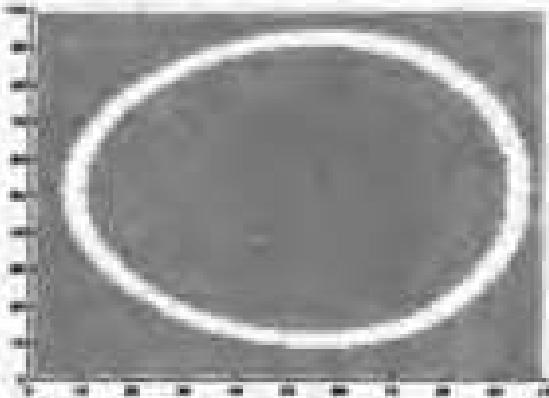


图 10-4 脑

方向有 n 个数据点，则坐标范围从 $(0,0,0)$ 到 (n,n,n) 。我们通过取该坐标系中的一个点 (x_0, y_0, z_0) 及表示平面与正 x 轴和正 y 轴夹角的两个角 (θ, φ) 来定义截面。这点成为平面的原点，而新的 x 轴为该 x 轴在这个平面上沿 z 方向的投影。故单位向量为

$$x' = x \cos \theta + z \sin \theta.$$

可以解出单位向量 y' ，如果我们要求它垂直于 x' ，与单位向量 x 的夹角为 φ ，且具有单位长度，则

$$y' = -z \sin \varphi \sin \theta + y \cos \varphi + z \sin \varphi \cos \theta.$$

这样，我们可以从 (x', y') 坐标系变回到阵列系，为

$$x = x_0 + x' \cos \theta - y' \sin \varphi \cos \theta,$$

$$y = y_0 + y' \cos \theta,$$

$$z = z_0 + x' \sin \theta + y' \sin \varphi \cos \theta.$$

我们称这已知点 (x, y, z) 为数据点，而未知点 (x', y') 为平面点。

10.3.3 插值算法

平面点通常不能与已有的数据点相交。我们知道每个平面点周围的水密度，所以必须用插值来估计平面的点密度。

有两类主要的插值法：

- 整体方法：用集中每个数据点去估计每个平面点的密度。
- 局部方法：仅用到数据点的一个小子集。

因为插值法假定是连续的，整体方法对于这个问题并不合适。我们知道，器官仅仅是分片连续与可微的，而在组织之间并不连续。因此，我们的所有算法都用局部插值法。

1. 邻近方法

这种算法对平面点指定最接近于该平面点的数据点的密度。该方法似乎简朴，但它都保留了清晰的边界而不弄模糊。对于平面的每一点 (x', y') ，计算原阵列中的 x , y 和 z ，并把它们进位到整数值 (X, Y, Z) ，这就给出最接近的数据点。

2. 密度平均值

该方法用更多的信息来估计每一点的水密度。我们可将每个平面点看成是在一个以数据点为其顶点的方体的内部。为估计内部的值，我们取周围八个点的密度的算术平均值。尽管利用了更多的信息，这种方法仍使间断的边变得模糊。

3. 三线性插值法

这种算法是用同样的八个点作为密度平均值的方法，但对密度(ρ)值作加权平均。这种方法假定斜率 $d\rho/dx$ 是数据点之间的常数。对于有低分辨数据集，该方法得出的是不精确的；但随着分辨率的增加，斜率将愈来愈保持恒定，因为在足够小的范围内检验时，任一可微函数出现的是线性的。其线性插值公式(见[9], p. 104~105)为

$$\rho(x') = \sum_{i=1,2} (1 - T_i) \rho_i,$$

其中 $T_i = |x' - x_i|$ 。

我们对于值 ρ_i 所给的权等同于到相反点的距离。这里，每对相继数据点之间具有一个单位的距离，所以相反点的距离是 $1 - T_i$ 。作为三线性插值，将这个和式扩充到所有点，故得

$$\rho(x', y', z') = \sum_{i,j,k=1,2} (1 - T_i)(1 - U_j)(1 - V_k) \rho_{ijk},$$

其中 $T_i = |x' - x_i|$ ， $U_j = |y' - y_j|$ 和 $V_k = |z' - z_k|$ 。

4. 多项式插值法

为更好地估计水密度函数，多项式插值方法要用到更多的数据。我们扩展八个点的周围方块，从每一方向加上一个方块，它在每一面有四个点，这就得到 64 个最近点。多项式可更好地拟合可微函数，因为它们有更多的导数，并且可以伴随函数更大的趋势。回忆两点确定唯一的直线，三点确定唯一的二次式，四点确定唯一的三次式。这样做，可以发展一种方法来拟合三维的函数。通过按 x , y 和 z 方向的一系列拟合，我们可以把这些综合成空间中一点的密度估计。这就把问题分解成一系列的一维插值来做。

设 (x, y, z) 为平面点，数据点为 $(x_{1\dots 4}, y_{1\dots 4}, z_{1\dots 4})$ ，我们先拟合 x_1 和 y_1 ，把四个点 $(x_1, y_1, z_{1\dots 4})$ 拟合成一个方块，并在 (x_1, y_1, z) 处对密度作插值。再对 x_1, y_2 作增量，并且同样一直做到在点 $(x_1, y_1, z), (x_1, y_2, z), (x_1, y_3, z), (x_1, y_4, z)$ 有密度。然后，沿 y 方向对这四个点用多项式拟合，并对 (x_1, y, z) 处密度作插值。我们重复这个全过程，找到 $(x_2, y, z), (x_3, y, z)$ ，和 (x_4, y, z) ，再将最后一个多项式拟合到这些点，以对 (x, y, z) 处密度进行插值。

有许多方法来进行多项式拟合。我们利用 Lagrange 公式（见 [1]），因为它的计算强度最小：

$$\begin{aligned}\rho(x') = & \frac{(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_2)(x_1-x_3)(x_1-x_4)}\rho_1 \\ & + \frac{(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_1)(x_2-x_3)(x_2-x_4)}\rho_2 \\ & + \frac{(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_1)(x_3-x_2)(x_3-x_4)}\rho_3 \\ & + \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_1)(x_4-x_2)(x_4-x_3)}\rho_4,\end{aligned}$$

其中 ρ_i 为 x_i 处的密度。

这种方法将会使边界变得模糊，但在用可微函数所描述的整个区域上会是做得很好的。

5. 混合算法

以上方法都具有优缺点。这些在可微区域内是最强的方法（三线性、多项式），在间断处却是最弱的，因为它们要对清晰的边界进行磨光处理。而最能保持间断性的方法（邻近法）在认同函数的光滑趋势方面是最弱的。

为利用两种方法的长处，我们提出混合算法。在对一组点作插值之前，用混合法来找它们中的间断之处；如果找到了，就用邻近方法，否则就用连续方法。这种混合算法通过测定平面点周围一对截然相反的点之间的密度($\Delta\rho$)差来确定间断处。如果出现间断， $\Delta\rho$ 就会变大，就用邻近方法。否则， $\Delta\rho$ 将会小，就用连续方法，即三线性或多项式法。为区分这两种情形，我们设定 $\Delta\rho_0$ 的阈值。这样，混合算法使我们用上每一种在达到最强时要用的方法。

10.3.4 验证与结果

因为已对四种模拟数据集的每一种精确定义了水密度函数，所以能将插值与实际的密度作比较，并发现其残差。为度量截面的精度，我们对整个平面上的点计算其平均绝对残差（即残差绝对值的平均）。

为比较这些算法，取 12 个截面，它们通过每个模拟数据集，并取不同的角度、点和间断的幅度。我们通常在数据中心附近选点，以生成较大的平面区域。

表 10-1 表示出每个算法在用于每个数据集时的平均绝对残差及对所有数据集的平均。我们依次对每个数据集（表 10-1 中的列）来讨论这些结果。

小体

总体来说，平均法最能精确地生成斜面，而三线性和多项式插值法都提供了具有相似精度的截面。对该数据集，混合算法得到的比纯连续方法差（图 10-5），邻近法最差，大概是因为在小球数据集中没有边界，而连续性从不破坏。三线性和多项式插值

法在每个球中心也还会带来峰值的麻烦。

表 10-1 用于插值方法的平均绝对误差

算法	小球	手臂	器官	肺	组合
邻近法	1.62	0.97	0.89	3.25	1.73
密度平均法	1.39	1.89	1.34	5.52	2.53
三线性插值法	1.54	1.27	0.70	3.49	1.75
多项式插值法	1.55	1.46	0.75	3.67	1.66
混合三线性插值法					
($\Delta\rho = 20$)	1.55	1.01	0.59	3.49	1.49
($\Delta\rho = 30$)	1.54	1.01	0.59	2.82	1.49
($\Delta\rho = 40$)	1.54	1.01	0.61	2.82	1.50
混合多项式插值法					
($\Delta\rho = 20$)	1.65	0.99	0.71	3.14	1.62
($\Delta\rho = 30$)	1.63	0.99	0.61	2.86	1.52
($\Delta\rho = 40$)	1.61	0.99	0.53	2.56	1.45

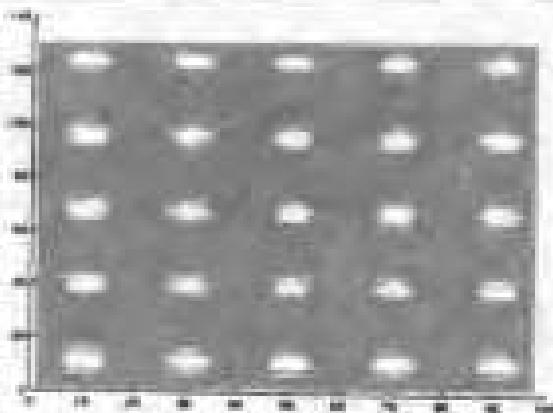


图 10-5 小体, 多项式混合方法, $\theta = 45^\circ$, $\varphi = 0^\circ$, ($45, 45, 50$)

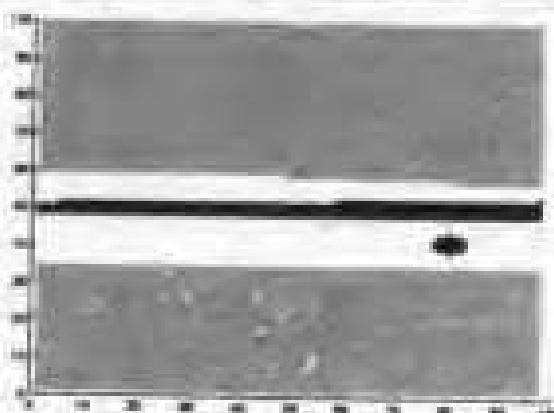


图 10-6 手臂, 多项式混合方法, $\theta = 10^\circ$, $\varphi = 0^\circ$, ($40, 80, 50$)

手臂

这里我们发现几乎全相反的情形。邻近方法与混合算法(图 10-6)执行得比纯连续方法要好得多。而平均方法做得特差。这是很合理的。因为我们的手臂从肌肉到骨骼有着许多轮廓很分明的边界。平均方法在许多间断处无效。这就是我们所看到的。间断探测似乎在进行。因为混合算法比三线性和多项式方法执行得

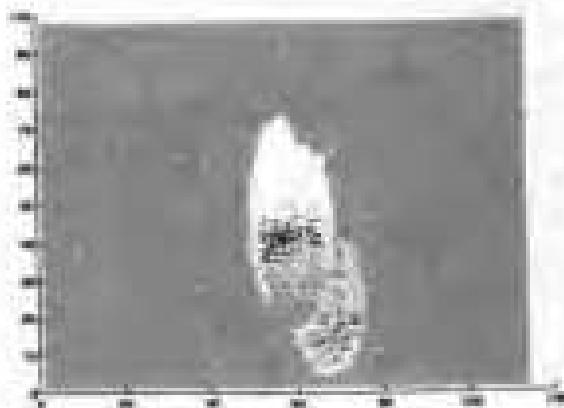


图 10-7 脑类器官:多项式混合方法, $\theta=0^\circ$, $\varphi=30^\circ$, (50, 50, 50)

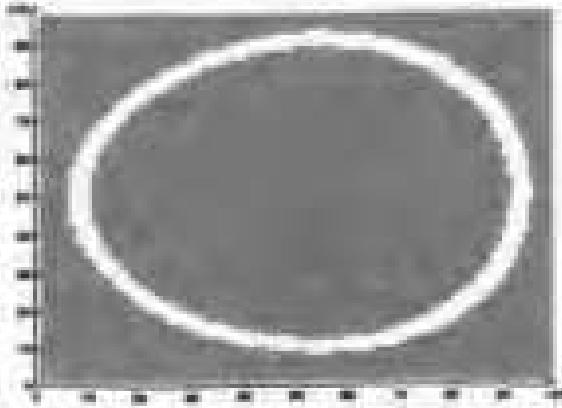


图 10-8 脑:多项式混合方法, $\theta=5^\circ$, $\varphi=0^\circ$, (50, 50, 50)

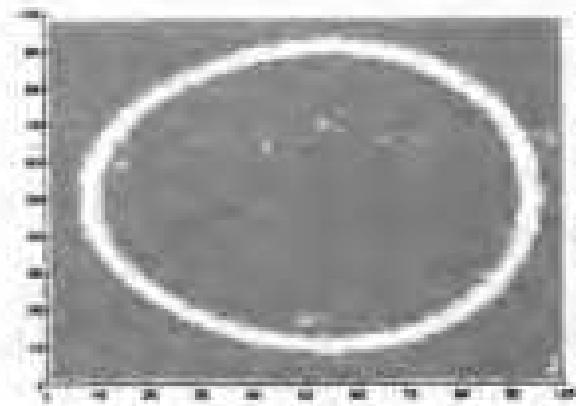


图 10-9 脑:多项式方法, $\theta=0^\circ$, $\varphi=30^\circ$, (50, 50, 50)

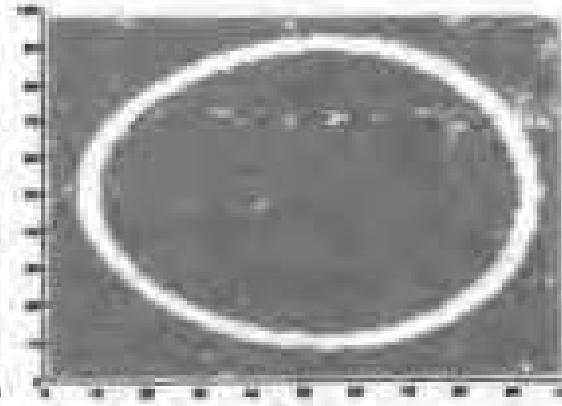


图 10-10 脑:多项式邻近方法, $\theta=5^\circ$, $\varphi=0^\circ$, (50, 50, 50)

好得多。

脑类器官

我们发现三线性与多项式法,两种混合公式都得出很好的结果(图 10-7),再次看到算术平均法执行得很差,接着是邻近法,多项式和三线性法,这些结果都是可以比较的。我们猜想这是因为对于这些数据集我们用了光滑函数来生成每个不同类型的组织。

脑

由于每个突起的普通光滑性和头盖骨的鲜明对照,具有高 $\Delta\rho_0$ 值的混合多项式方法得到了最精确的结果。逼近法得到很精

精确的结果，超过三线性和多项式方法，但不及混合方法（图 10-8~10-10）。最不精确的结果是由算术平均法得到的。图 10-11 清楚地表明由于要对头盖骨的边缘作磨光处理而如何失效的。

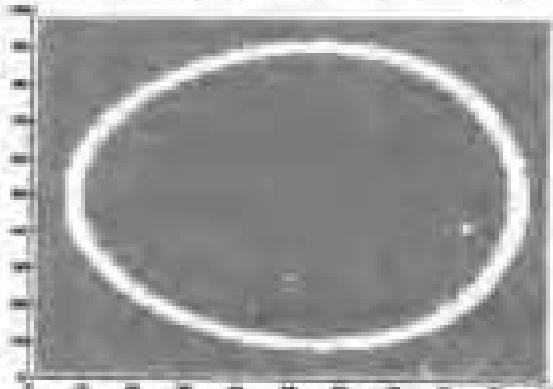


图 10-11 脑：密度平均法 $\theta = 0^\circ, \varphi = 30^\circ, (50, 50, 50)$



图 10-12 脑：邻近法残差 $\theta = 10^\circ, \varphi = 0^\circ, (50, 50, 50)$

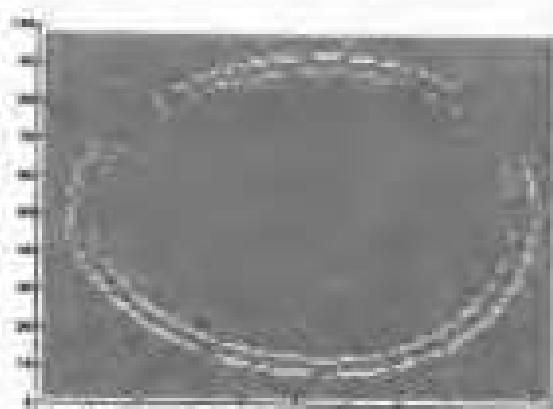


图 10-13 脑：密度平均，残差， $\theta = 0^\circ, \varphi = 30^\circ, (50, 50, 50)$

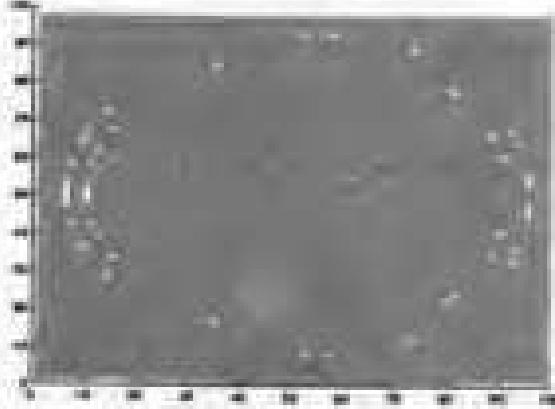


图 10-14 脑：多项式混合方法，残差， $\theta = 5^\circ, \varphi = 0^\circ, (50, 50, 50)$

残差图

检验这些算法的另一种方式是作出残差图，它使我们确切地看到方法在何处失效的。在图 10-12~10-14 中大的正或负的残差用白色或灰色标明。

邻近方法对于脑的大部分一般是精确的，除了一叶的暗区及在少许区域头盖骨的间断处也产生误差（图 10-12）。

算术平均算法在所有边界周围明显出现差错(图 10-13).

最后, 混合多项式算法在头盖骨边缘附近不精确, 但它处理暗区很好.

总体结果

- 在四个数据集上作平均(见表 10-1 最后一列), 对于通过三维数据的阵列生成斜平面来说, 最精确的算法是混合多项式算法(这里 $\Delta\rho_0 = 40$), 它产生的平均误差低于邻近法的 16%.
- 混合三线性算法(这里 $\Delta\rho_0 = 30$)执行得几乎很好, 生成的平均误差低于邻近法的 14%.
- 邻近算法合理地产生精确的平面, 其平均误差低于连续多项式算法的 7%, 比连续三线性算法强 1%左右.
- 三线性算法比多项式方法强, 但在处理间断方面所得到的结果也不如混合方法那样好.
- 算术平均算法得到特别差的结果, 因其在任何一种间断处附近都产生大的误差.

10.3.5 优缺点

混合多项式算法是灵活的, 并可容易地推广到处理任意大小的数据阵列. 可以通过任意一点以任意角度取切片. 而且, 它在通过光滑区域作插值是有效的, 但仍保留原数据中边界的轮廓. 即使当最近的点落在间断处错误的一边时, 映像从定性上仍是对的; 它显示出有轮廓的边界. 阈值常数 $\Delta\rho_0$ 由用户选取以做得何等光滑.

混合算法会错过数据中小的间断, 它并不去找不可微的点. 如果有尖点, 该算法大概注意不到它们, 而想越过这些点光滑化. 即使这样做并不合适.

10.3.6 进一步的工作

下一步是要用更大的点集对连续区域的插值实施更好的方

法。三次和四次的样条会是有效的，其他类型的多项式法，也许还有有理函数插值方法也会如此。

间断探测算法可以改进，扩展到寻求尖点和不可微点。间断处也可执行自动组织的定型；这种算法因而应能自动地输出仅显示大脑灰度的映像，或仅显示肿瘤组织。即使用流行的软件，对于 $\Delta\rho_0$ 的动力学的更细致的研究仍是很用的。

最重要的是，这种算法需要通过与实际 MRI 数据做对照来检测。

§ 10.4 任意平面成像的模型

本节我们介绍美国的 Harvey Mudd 学院获特等奖的优秀论文。

10.4.1 摘要

我们基于均匀取样的 MRI 数据直线阵列，提出一种算法来对一个三维密度函数的任意斜片作成像处理。

我们

- 发展一种线性插值格式以确定像平面上点的密度；
- 组合成一种离散的卷积滤波器 (filter) 以补偿由插值引起的不必要的模糊；并且
- 根据有限差分提供一种探测边界。

所得到的算法使用个人电脑是会足够快的，并且允许户用参数控制。

我们从典型人脑的模拟 MRI 扫描方面，以及为检验该模型的局限性而设计的人为的数据结构方面展示出对算法的检验结果。由插值造成的过滤的畸变及不精确的建模出现于某些极端场合。然而，我们发现我们的算法适用于现实的医疗成像。

10.4.2 构造模型

我们的模型主要由四个部分组成：

- 首先，我们发展一种方法来把平面置于 R^3 中的任何位置.
- 然后我们将数据从 R^3 的包含数据的区域插值到平面上.
- 接着我们用锐化的方法去除因插值引起的多余的模糊.
- 最后，我们构造出一种差分阵列，并用它来作出表现映像边界的线条图.

假设

- 源目标中的密度变化很能体现出合理性，并且是连续的. 像清晰的边界那样的间断之处将在模型中得到逼近，但是仅当他们在某些阵列元的尺度内是孤立的时候. 类似地，不规则的行为和粗糙的涨落可以精确地模型化，只在他们位于某些像素的范围内. 该模型应是对数据阵列源，而不是对阵列本身成像，但斜切片的精度取决于阵列中数据的精度.
- 数据阵列各向同性地表现出等间隔的样本. 阵列 $A(i, j, k)$ 包含着从连续的三维空间中离散化了的样本，此空间我们用坐标 (x, y, z) 表示. 分量 $A(i, j, k)$ 表示在某个点 (x_i, y_j, z_k) 处的密度 f . 我们假定该源是均匀取样的，故有

$$x_i = i\delta x, \quad y_i = j\delta y, \quad z_i = k\delta z,$$

其中 δx , δy 和 δz 为样本间的常数距离(典型地，接近于 1mm). 我们还假定这些距离都相等(否则，我们改变坐标的尺度以补偿).

- 设定的计算平台是当前典型的个人电脑. 因而，输入数据阵列不会比典型 PC 机的存储更大. 我们假定阵列是到 $256 \times 256 \times 256$ 的整数元(容量 16MB)，每个在 $[0, 255]$ 中. 我们还通过典型的 PC 机处理器速度来测定计算时间.

平面阵列的截面

为在计算机屏幕上表示物片，我们在三维目标空间和屏幕平面之间建立一个映射。用与 x - y 平面和 z 轴间的角度以及原点的位移来表示 R^3 中的任意平面。由

$$T(u, v) = R \begin{pmatrix} u \\ v \\ 0 \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$$

给出的映射 $T: R^2 \rightarrow R^3$ ，将 R^2 中的点变成平面上的一个点，这里点 (x_0, y_0, z_0) 为原点的平移，而 R 是旋转矩阵

$$R = \begin{pmatrix} \cos\varphi \cdot \cos\theta & -\sin\theta & \sin\varphi \cdot \cos\theta \\ \cos\varphi \cdot \sin\theta & \cos\theta & \sin\varphi \cdot \sin\theta \\ -\sin\varphi & 0 & \cos\varphi \end{pmatrix}.$$

角度 φ 和 θ 为垂直于该平面的向量在球坐标下的极角和方位角。

10.4.3 插值法

为了表示映像，我们寻求与计算机屏幕像素对应的离散化的点 (u_i, v_i) 的有规则间隔的阵列。因为点 $T(u_i, v_i)$ 不必与该数据的点 (x_k, y_k, z_k) 一致，需要能逼近 R^3 中任意点的密度值。因而，我们从 A 给出的点的附近来插值该数据。

有一小点概念上的滥用，设 $g(x, y, z)$ 为该像在 (x, y, z) 处的灰度值，使得

$$g(T(u, v)) = g(u, v) \text{ 和 } g(x_k, y_k, z_k) = A(i, j, k).$$

从插值的一系列技巧，我们寻求一种算法，它将光滑地逼近密度而在计算上也不难操作。

最近邻域逼近

设 (x^*, y^*, z^*) 为我们想知道其密度的点。该点含在一大小为 $\delta x \times \delta y \times \delta z$ 的方体单元内，它有阵列 A 给出的已知密度的顶点。从这八个顶点只找到最靠近 (x^*, y^*, z^*) 的点 (x_a, y_b, z_c) ，且设 $g(x^*, y^*, z^*) = A(a, b, c)$ 。

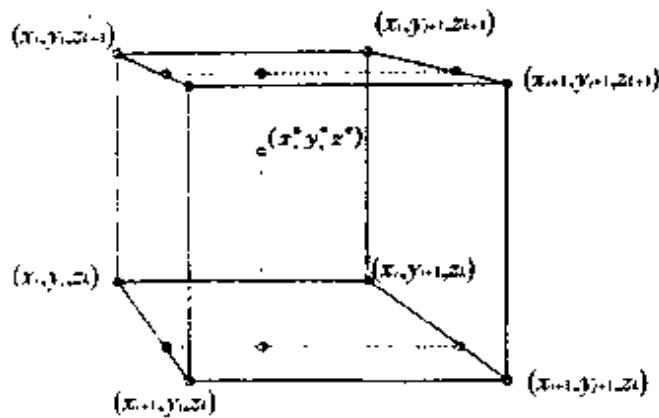


图 10-15 表示三维插值格式的立方体

3-D 线性逼近

我们也发展一种称之为 3-D 线性插值的技术。对于这种方法，希望从含 (x^*, y^*, z^*) 的方块顶点的密度值出发，找到对该方块里面数据的一种光滑延拓。该方法基于依次解一、二和三维的 Laplace 方程。

$$\Delta g = \frac{\partial^2 g}{\partial x_1^2} + \dots + \frac{\partial^2 g}{\partial x_n^2} = 0.$$

我们取该块的两个相邻顶点，并且用这些顶点的密度作为边值解一维 Laplace 方程。这给出两点间最光滑的函数——直线。然后用直线作为边界条件，在方体单元的面上解二维 Laplace 方程。最后我们用这些面上的值作为边界条件来解三维的 Laplace 方程以填满该方体。

细节并不困难。方体单元中点的表示如图 10-15 中。沿边界上的值通过简单的线性插值来构造；例如，沿图 10-15 中的方块的左下边的值由下式给出

$$\begin{aligned} g(x^*, y_j, z_k) &= g(x_i, y_j, z_k) \\ &+ \frac{g(x_{i+1}, y_j, z_k) - g(x_i, y_j, z_k)}{x_{i+1} - x_i} (x^* - x_i) \\ &= A(i, j, k) \\ &+ \frac{A(i+1, j, k) - A(i, j, k)}{x_{i+1} - x_i} (x^* - x_i). \end{aligned}$$

类似有

- 沿着 $A(i, j+1, k)$ 和 $A(i+1, j+1, k)$ 所在的右下边的值 $g(x^*, y_{j+1}, z_k)$,
- 沿着 $A(i, j, k+1)$ 和 $A(i+1, j, k+1)$ 所在的左上边的值 $g(x^*, y_j, z_{k+1})$,
- 沿着 $A(i, j+1, k+1)$ 和 $A(i+1, j+1, k+1)$ 所在的右上边的值 $g(x^*, y_{j+1}, z_{k+1})$.

我们继续用线性插值来得到左下边的值 $g(x^*, y_j, z_k)$ 和右下边的值 $g(x^*, y_{j+1}, z_k)$ 所在的底面上的值 $g(x^*, y^*, z_k)$, 以及左上边的值 $g(x^*, y_j, z_{k+1})$ 和右上边的值 $g(x^*, y_{j+1}, z_{k+1})$ 所在的顶面上的值 $g(x^*, y^*, z_{k+1})$.

作为最后一步, 再一次用线性插值法由底面上的值 $g(x^*, y^*, z_k)$ 和顶面上的值 $g(x^*, y^*, z_{k+1})$ 得到值 $g(x^*, y^*, z^*)$. 结果是由最靠近的八个顶点关于 $g(x^*, y^*, z^*)$ 的惟一的值, 它与插值的顺序无关. [原编者注: 这里省略作者关于这个事实的证明, 他们是通过 $g(x^*, y^*, z^*)$ 的直接计算并注意到 x , y 和 z 的对称性而得到的.] 而且,

$$\begin{aligned}\frac{\partial^2 g}{\partial x^2}(x^*, y^*, z^*) &= \frac{\partial^2 g}{\partial y^2}(x^*, y^*, z^*) \\ &= \frac{\partial^2 g}{\partial z^2}(x^*, y^*, z^*) = 0,\end{aligned}$$

它表明 Laplace 方程在该方块中满足. 由具有 Dirichlet 边界条件的 Laplace 方程解的惟一性定理可知, 用这方法可得到仅有的解.

其他方法

我们考虑三维样条, 这里是取三次插值使一阶导数连续. 然而, 这种技术上的复杂性和计算耗时使得它不可取. 我们也考虑过空间加权平均法. 不过在把这种方法应用到简单的二维情形时, 我们发现与真实映像严重脱节(简单的线性的斜坡被改变成

好像是一系列波状的阶梯).

最后, 我们发现

- 最近邻域方法对于粗略的成像分析最有用, 并且,
- 3-D 的线性插值对于更现实的成像是最有效的方法.

10.4.4 映像锐化

插值不可避免地使实像 f 变得模糊, 或难以通过滤波器. 所以, 我们对该算法加上一个步骤, 它锐化所记录的像 g , 或者使之易通过滤波器. 我们来考虑锐化的各种技术.

边界的复原

我们锐化的方式是探测映像中的边或边界的位置, 然后在这些位置附近回复成最邻近的像素的测定. 我们发现这种方式对于增加收益和像素有不利的一面.

点散布函数

另一种方法是安德鲁斯(Andrews)和亨特(Hunt)(见[2])的, 即设所记录的映像是将实像与 $h(x, y)$ 所表示的点散布函数 (PSF) 作卷积. 这样,

$$g(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x - \xi, y - \eta) f(\xi, \eta) d\xi d\eta. \quad (1)$$

该实像则通过离散的傅立叶变换的反演得到. 这样的 PSF 可以事先计算或事后测定.

或者, 数据离散性质和插值过程的线性性质把我们引导到改写(1)成为矩阵方程

$$g(u_p, v_q) = \sum_{m,n=1}^N a(p, m) b(q, n) f(u_m, v_n),$$

其中 $a(p, m)$ 为 $N \times N$ 矩阵, 它使数字化平面映像的列变得含糊, 而 $b(p, m)$ 为使行变得含糊的 $N \times N$ 矩阵. 这种“含糊的”矩阵可以由其他矩阵来逼近, 该类矩阵的主对角元在单位附近, 而邻近的非对角元等于某个小“混合”参数. 因而映像 f 可以通过取这类矩阵的逆而得到复原.

最后，我们认为这个方法与傅立叶 PSF 方法都太耗计算量。
卷积滤波器

我们偏好的方法是用卷积滤波器，这出自 [10]。该方法基于出现模糊是一个扩散过程的假设。如果实际映像是扩散方程

$$\kappa \nabla^2 g = \frac{\partial g}{\partial x}$$

的初始条件，则通过依时量 $g(u, v; t)$ 在小时间值 τ 处展开，我们得到

$$\begin{aligned} f(u, v) &= g(u, v; 0) = g(u, v; \tau) - \tau \frac{dg}{dt}(u, v; 0) + O(\tau^2) \\ &= g - \kappa \tau \nabla^2 g + O(\tau^2). \end{aligned}$$

这样， f 便可由 g 减去 g 的 Laplace 值而得到恢复。这种方法通常称为消除轮廓，在我们的模型中特别有吸引力；因为我们所取的插值格式使映像 g 的插值区域满足 Laplace 方程 $\nabla^2 g = 0$ 。

在应用中，Laplace 值通过有限差来逼近。命

$$\Delta_u g(u_p, v_q) = g(u_p, v_q) - g(u_{p-1}, v_q),$$

$$\Delta_v g(u_p, v_q) = g(u_p, v_q) - g(u_p, v_{q-1}).$$

高阶差分算子通过重复的一阶差分来定义，即为

$$\Delta_u^2 g(u_p, v_q) = \Delta_u g(u_{p+1}, v_q) - \Delta_u g(u_p, v_q),$$

它导出

$$\begin{aligned} \nabla^2 g &= \Delta_u^2 g(u_p, v_q) - \Delta_v^2 g(u_p, v_q) \\ &= g(u_{p+1}, v_q) + g(u_{p-1}, v_q) + g(u_p, v_{q+1}) \\ &\quad + g(u_p, v_{q-1}) - 4g(u_p, v_q). \end{aligned} \tag{2}$$

对整个矩阵应用 Laplace 算子可以看成关于卷积的离散的类似物。即我们可以在“掩模(mask)”矩阵

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

的分量附近的 3×3 邻域中通过对每个值按分量相乘，并将所得的 3×3 矩阵的所有分量相加，由此得到映像矩阵分量 $g(u_p,$

v_q) 的 Laplace 值.

按照(2), 我们希望与该掩模作卷积

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} - \alpha \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -\alpha & 0 \\ -\alpha & 1+4\alpha & -\alpha \\ 0 & -\alpha & 0 \end{bmatrix},$$

它是从该函数自身减去 Laplace 乘以一个控制参数 α (类似于 $\kappa\tau$).

这种方法的自然扩充是利用 Laplace 算子的高阶逼近(因而与一个更大的掩模作卷积), 或增加具有某种易通过滤波器的掩模. 考虑到时间及计算的复杂性, 我们不发展这种方法.

10.4.5 边界探测

为使在我们的映像中更能看到边界, 探测边界及生成相应的直线图是有用的. 为此, 我们运用已经讨论过的有限差方法一个变种. 在每一点 (x_i, y_j, z_k) , 我们计算

$$\Delta_{2x}(i, j, k) = A(i-1, j, k) - A(i+1, j, k),$$

$$\Delta_{2y}(i, j, k) = A(i, j-1, k) - A(i, j+1, k),$$

$$\Delta_{2z}(i, j, k) = A(i, j, k-1) - A(i, j, k+1).$$

这组定义将差集中在问题中的点的周围.

我们构造一个新的阵列 Γ , 这里

$$\Gamma(i, j, k) = \max\{\Delta_{2x}(i, j, k), \Delta_{2y}(i, j, k), \Delta_{2z}(i, j, k)\},$$

它度量了 A 的值是如何快地通过一个已知点而变化的.

这种方法有着许多优点:

- Γ 的值容易计算.
- 该方法并不会使所遇到的边界类型(笔直, 弯曲, 倾斜等等)发生偏离.
- Γ 的值仍留在最初的灰度范围内(见[10]).

于是, 对于通过由 Γ 给出而不是由 A 给出的数据盒的平面, 我们来运用插值法. 只要我们有了这个新的二维映像, 我们就通

过应用阈值条件把它复原成双映像：每个插值在阈值之上的点标上黑色，而每个插值在阈值之下的点标上白色。这将差分值高的区域转成黑色区域，而与白色背景相对照。因为边界显示出大的密度差，黑色区域表示了边界。

10.4.6 对模型的分析

我们在 Unix 制图工作站上执行了平面成像算法，并对几个不同的数据组分析了算法的行为。有一个数据组是对人脑的模拟 MRI 扫描，这是一个期待在现实的医疗透视环境中收到的模型数据量的例子。为了对已知的结构检验上述算法，我们提出其他一些人为的数据集，从而揭示出该模型的局限性。

计算机工具

我们运用 C++ 语言和 OpenGL 的 3-D 的图像库来把模型建成相互作用的图像应用。所设计的程序有着许多特点，这也是实际医疗场合用到的平面成像系统所具备的。取表现密度阵列 A 的任意尺寸的三维字节值(0~255)作为输入。本程序对用户提供两个显示窗口：

- 第一个窗口看到的是 (x, y, z) 坐标空间，它显示出输入数据阵列和投影平面的线框表示。
- 第二个窗口显示出由我们的平面成像算法生成的平面上的扫描映像。

用户能用键盘和鼠标把成像平面移动到 A 中不同的位置 (x_0, y_0, z_0) 和角度 (φ, θ) ，实时看着投影映像的变化。我们使用这个程序产生了本文的全部图形。

对于在 A 中作插值，以及在投影平面中作出锐化映像的编码算法，它们都是前面给出的数学陈述的直接转移。我们对于这种成像算法有着特别的理解，它可以确定源数据是位于 (u, v) 平面的哪个部分(见 10.4.9 附录)。

用户能控制模型的全部参数，包括锐化控制因子和边界探测

阈值，可以选择不同的插值法，锐化与边界探测滤波器也能很好地套上。程序执行得相当快；但是，因为我们要少许注意到建立最优算法，为了加快操作的速度（如锐化滤波器）还有许许多多的潜力。

脑 MRI 数据的结果

我们的模型的主要检测数据组是包含 181 片，每片有 181×217 个像素的模拟人脑 MRI 组。这些数据是由非常现实的 MRI 计算机仿真输出的（见[3]），因而可能是使用户有诊断兴趣的实际 MRI 扫描仪所反映的数据。

斜方向结构的探测

图 10-16~10-19 是我们关于脑 MRI 数据组的平面成像算法按垂直的和倾斜的几个不同平面方向输出的示例。在图像计算机的显示器上动态地看到这样的输出，将使医生从任何可能的位置和方向探测到脑中的结构。

精细结构

脑 MRI 数据表明了应用锐化滤波器的好处。图 10-20 显示两个脑映像：左边的一个是已锐化的，而右边的一个还没有。锐化的映像已没有不锐化映像那种多余的模糊。它也清楚地显示出脑中的精细结构，这对计划用最小创伤程序的外科医生是有兴趣的。

线条描绘

我们描述的探测边界的算法可用来生成在解剖图册中能找到的那种黑白线条图像（见图 10-21）。这样的图像对于寻找脑中结构清晰的边界是有用的。

关于人为的数据组结果

我们模型提出的一个问题是，锐化滤波器能否去除由插值算法引起的模糊，或者仅由数据阵列留下来的含糊。为了要回答这个问题，我们检验了具有完全离散边界的数组模型，这里数据组由“厚片”组成，即具有某种表而深度的一组平行的等间隔平面。它包含着最大的强像素。图 10-22 显示一成像平面通过与厚

片成 35° 角的量(任意选择)的精密观察结果。比较两个不同映像中厚片边界的清晰度表明，我们的锐化滤波器在去除插值效应方面执行得很好，只要边界以充分大的角度穿过该成像平面。当成像平面与厚片成小角度时，我们看到即使在锐化映像中也有模糊的边界，这是我们三维插值格式的必然结果。

数据组的价值可以看成是我们模型想要遇到行为的离散极端。为了验证一个相反的连续的极端，我们制作了数据组，其强度在单个的 $x-y$ 平面上不变，但随 z 的三次函数变化。我们模型中的插值算法将许多变化着的非线性数据线性化了，其结果是锐化的滤波器导致变形。

10.4.7 模型的限制

我们的模型有几条限制：

- 非均匀或各向异性的取样数据阵列不考虑。
- 其平面边平行或几乎平行于投影平面的物体成像并不精确。
- 插值和滤波器产生得到更多的时间和工作也许更高的顺序计划的使用减轻的小变形。像十进制输入那样使用补充性的发明的数据结构，可以进一步照亮为了这样的变形的事业。

10.4.8 结论

我们的计算机设备提供的图解描述了我们的模型对于描写倾斜平面上的密度变化的能力。我们的算法以允许容许平面任意定向，迅速地生成映像，以及包括锐化和边界探测滤波器，优于大部分已有的算法。我们的模型对于模拟脑的 MRI 成像结果表明了其现实医疗成像的实用性。

10.4.9 附录：成像平面的边界

为从计算上优化我们的平面成像算法，我们推导平面与数据盒相交区域的界限。这使我们的算法在保证得到所有相交数据所

需的最小的 (u, v) 区域上迭代.

$T(u, v)$ 变换给出了三个方程

$$\begin{aligned}x &= \cos\varphi \cos\theta u - \sin\theta v + x_0, \\y &= \cos\varphi \sin\theta u - \cos\theta v + y_0, \\z &= -\sin\varphi u + z_0.\end{aligned}$$

我们可以把它们改写成

$$\begin{aligned}x - x_0 &= \cos\varphi \cos\theta u - \sin\theta v, \\y - y_0 &= \cos\varphi \sin\theta u - \cos\theta v, \\z - z_0 &= -\sin\varphi u.\end{aligned}$$

我们想知道何时当该平面穿过数据盒的边界. 这个边界在三个变量 (x, y, z) 中将有两个是常量. 我们能把这两个值连接在上述适当的方程中，并且解出 u 和 v .

例如，假如我们想知道在哪一点 (u, v) 平面与平行于 z 轴的数据盒的边界相交. 此时，我们知道 x 和 y ，因为我们知道 R^3 中该盒子的大小和位置. 利用关于 $x - x_0$ 和 $y - y_0$ 的方程，我们有

$$\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} = \begin{pmatrix} \cos\varphi \cos\theta & -\sin\theta \\ \cos\varphi \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

注意，如果 $\varphi = \pi n + \frac{\pi}{2}n$ ，其中为 n 整数，这变换没有逆. 此时该平面平行于 z 轴，因而为该数据盒的任意垂直边. 如果 $\varphi \neq \pi n + \frac{\pi}{2}$ ，则我们可以求出上述变换的逆，得到

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{\cos\varphi} \begin{pmatrix} \cos\theta & \sin\theta \\ -\cos\varphi \sin\theta & \cos\varphi \cos\theta \end{pmatrix} \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix}.$$

对于 R^3 中具有 φ 和 θ 类似限制的其他两个方向，我们可以实行类似运算. 因而我们得到 $u-v$ 平面上的 12 个点，或者更少一些，如果平面平行于其中一个轴的话. 用这 12 个值，我们仅仅选取 u 和 v 的最大和最小值，就得到一个长方体，它紧挨着那个平面

和数据集的交.

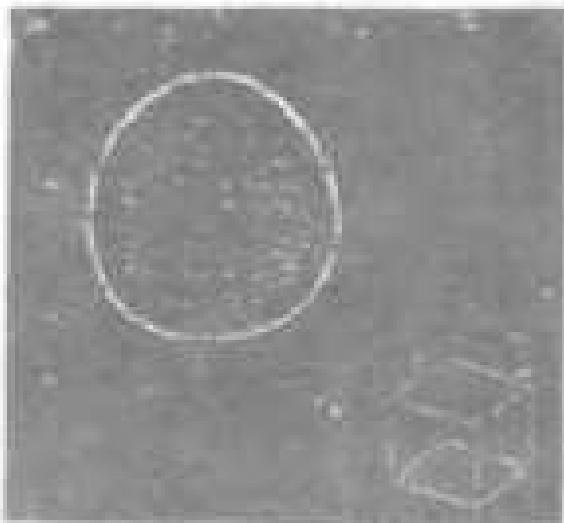


图 10-16 中心在(91, 108, 120)且 $(\varphi, \theta) = (0, 0)$ 的映像。映像平面垂直于 z 轴



图 10-17 中心在(91, 108, 120)且 $(\varphi, \theta) = (35, 75)$ 的映像。映像平面垂直于 z 轴

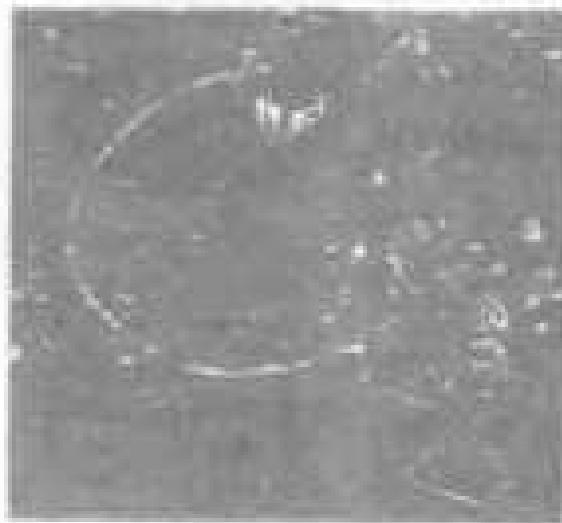


图 10-18 中心在(110, 100, 70)且 $(\varphi, \theta) = (130, -30)$ 的映像。

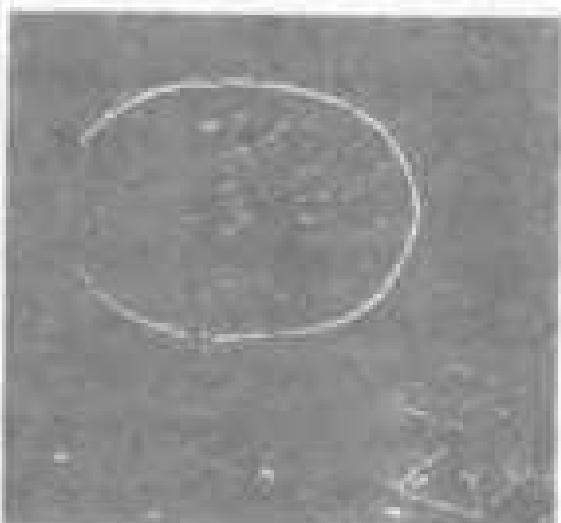


图 10-19 中心在(100, 100, 45)且 $(\varphi, \theta) = (-20, 90)$ 的映像。

§ 10.5 MRI 成像截面的三维三次插值算法

本节我们介绍美国的 Macalester 学院获特等奖的优秀论文。

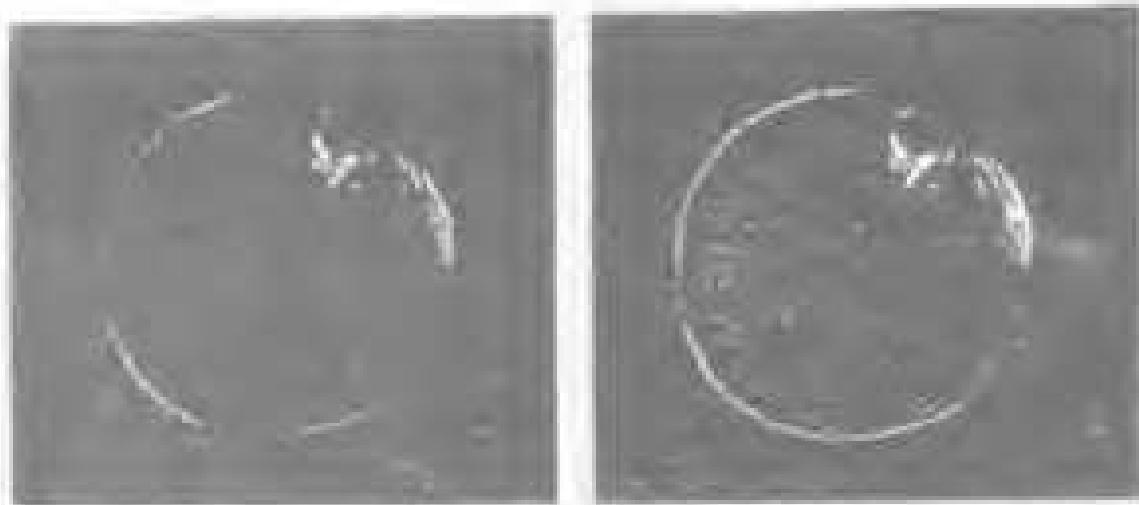


图 10-20 两个脑 MRI 数据斜平面扫描的比较。右边的映像穿过极化滤波器，而左边的还没有。映像平面以(63,116,58)为心且(φ,θ) = (0,0)

10.5.1 引言

我们设计并执行了一种程序，它能

- 接受一 Byte 灰度像素量的三维阵列，并从 MRI 仪输出。
- 通过沿任意平面阵列作斜切片，及
- 用插值法得到由平面描述的截面映像。

我们容许用户选择由平面上三点指定的，而不是由一点和两角指定的截平面。我们顾及不同维数中大小不等的像素量的可能性。但是，假定它们对每个维数的间隔都相等。然后用三维三次插值算法产生截面映像。这方法是广泛用于二维映像的二维三次插值算法的推广。选择三维三次方法是因为它在准确性和运算速度方面提供了最佳平衡。最后，容许用户给数据的重要的部分“着色”，或者涂上颜色。

我们试用简单的几何图形程序来检验其正确性。然后在两个脑的实际的 MRI 映像切片上予以验证，得到非常令人满意的结



图 10-21 穿过脑 MRI 数据的斜平面的黑白线条画。线条用边界轮廓画。映像平面与图 10-17 中的相同。

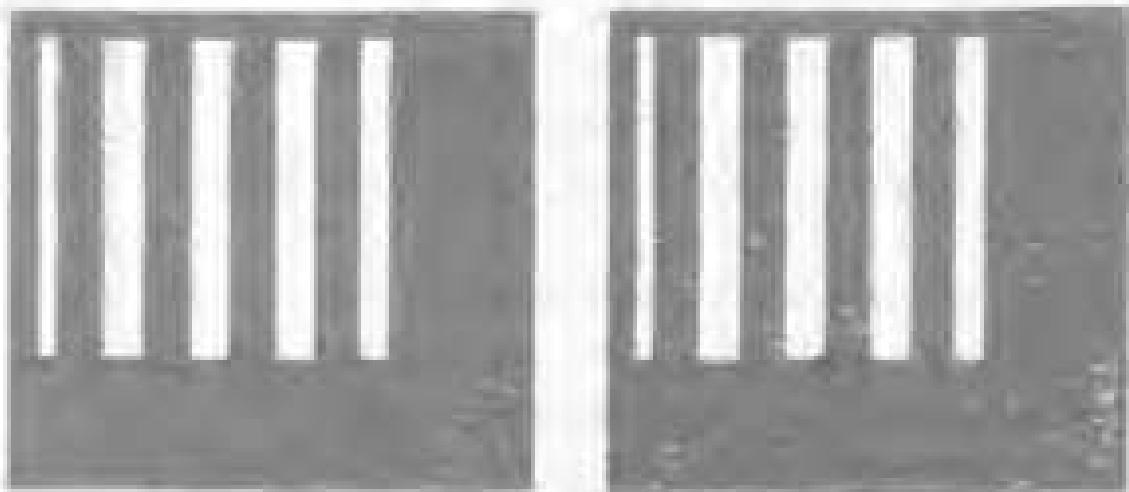


图 10-28 “厚片”数据的两个斜平面扫描比较。快模平面与可视厚片网的投射角为 35° 。右边的映像穿过锐化滤波器，而左边的还没有。

果，发现重要的映像特点得以妥善保存，并且映像的着色对直可视化很有用。插值算法以线性时间运行：它在几秒内产生一个 $256 \times 256 \times 256$ 个数据量的映像。

最后，构造一些数据集以指出我们算法和问题本身的局限性。这些局限性与我们算法在远离样本点的最不确定性区域的性质有关。

10.5.2 设计考虑

1. MRI 成像的典型运用

如同 Rodriguez [1995] 所描述的那样，身体各部位的 MRI 扫描被用来诊断大范围的紊乱。一个最常用的是对脑中的异常体，如肿瘤、囊肿和血肿的探测。由于健康的组织和肿瘤在外形上的变化，在取截面时边界的清晰和模糊，以及区域的一般形状和亮度均应予以保留。这是要緊的。

分析程序应该依据空间的直观理解，而且还为用户提供直接的方式来指定随后的截面浏览量。

2. 映像数据的特征

数据大小

一个 MRI 切片的常用数据大小为 256×256 个灰度像素。为得到覆盖整个三维目标的数据，需要多重切片。对每个切片扫描所需的时间依赖于扫描过程中称为重复次数的参数；一个典型的切片可能要花几分钟来扫描，虽然多重切片有时可以同时扫描（见[6]）。因为患者呆在仪器前面的时间有限，所能取得的切片数与切片分辨率相比是很小的。我们实际试验数据的数据库（见[7]）典型地取 25~60 个切片对整个脑作扫描。

这说明由每个像素所表示的实际空间量不像是个方体。代之以有可能是矩形的棱柱，在一维方向比其他两个要长得多；该算法需要把这事实考虑进去，以便不产生变形的输出。此外，如果一个像素沿着一根轴比沿着别的要长得多，在这一维就需要更多的插值，这样，平面所得的平行于那个轴的映像可能是很不精确的。

数据的加工品

很多不同种类的加工品也许表现在 MRI 图像数据中；有些是仪器的不正确的操作或配置的结果，而另一些则是在扫描过程中具有物理属性的产品（见[3]、[6]）。

因为大部分这种加工品反映出仪器配置中的问题，它们有可能产生错误映像，故被保存在截面里，使得 MRI 的操作员可以看到它们并适当调整仪器，这点是很重要的。

抽样特点

数据中所有这些映像的特点表现为基本上都与离散抽样的性质相关联。我们可以对这些数据进行分类，在这里离散样本点描述了连续函数（如我们数据所做的），无论它是过少抽样，过多抽样，还是临界抽样（见图 10-23），这取决于抽样的分辨率与映像中实际细节的对应程度。

- 过多抽样的数据：样本网格比映像细节更细。这样的映像趋势看上去很模糊，邻近的网格点仅倾向于轻微变化，而且包



图 10-23 映像的典型数据抽样的特点

括本质上多余的信息。细节的这种高水平使映像适合于精确的插值和强化。

- 过少抽样的数据：映像比样本网格包含着更细的枝节，并且在邻近的像素之间，特别是在映像中物体的边界处只有极少的关联。如果每个像素的实际样本区比像素所表现的样本区要小，破映像可通过齿形边界和清晰对照来描述。这样的映像使插值法和强化法成为研究和检测的一个问题。
- 临界抽样的数据在过少和过多抽样的数据之间。MRI 数据进入这一类中。像过多抽样的数据那样，边界线倾向于去除多余量(平滑)，并且映像甚至会出现轻微的模糊；不过，和不足抽样的数据一样，像素水准的细节是重要的，而且插值法可能受到限制。

10.5.3 插值算法

我们输入的数据是从离散点取得的一组映像值，但想取的截面可以不与这些点相交。因而，需要有一种方式来根据样本点的映像值估计任意点的映像值。这就是说，根据我们的样本阵列 $A_{i,j,k}$ ，想要一插值函数 $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ ，使得

$$f(i,j,k) = A_{i,j,k}$$

其中 i, j 和 k 为整数，且使得 f 对于非整数 i, j 和 k 也能取到合理值。(插值函数与样本点相比较的约定是合理的，因 MRI 映像趋于完善，并且有高的信号噪声比。)

在选插值函数时，必须在映像生成的精确性和由 MRI 数据典型的临界抽样性质所限制的运行时间之间作截断。我们取三次方法，发现它对实际的 MRI 数据是非常之快并且很精确。

1. 三维三次插值法

三次插值法是 Lagrange 插值法的特例，后者是通过 n 个数据点求出 $(n-1)$ 次惟一的多项式的一种简单方法。

我们先考虑一维情形。三次插值开始于最靠近目标点 x 的四个样本点，每一边有两个最近的邻点 $[x]$, $[x]-1$, $[x]$ 和 $[x]+1$ 。并将它们拟合成一个三次函数 $p: R \rightarrow R$; $p(x)$ 给出了 x 处的插值（见图 10-24）。注意，由 x 附近的这四个点所描述的特殊的 f 只在中间两点的区间中给出值。因而，对整个映像作插值的函数 f 是许多不同的三次函数的分段组合。

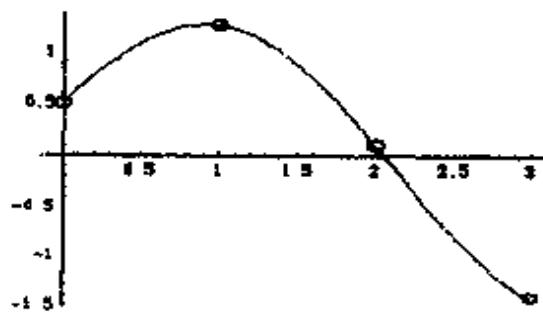


图 10-24 一维三次插值法

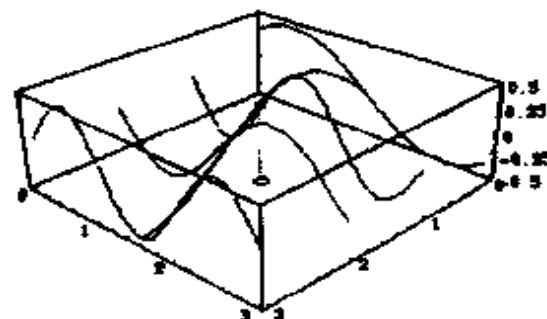


图 10-25 二维三次插值法

这个过程易推广到高维。它不要求所想像的，构造精细的多变量多项式或解大的方程组；事实上，只要对每一维相继进行插值就够了。

先对二维来看这过程也许更容易。如图 10-25 所示，我们把目标点周围的 16 个点分成四条线，每条四个点。我们沿每条线作一维三次插值，再在含目标点的垂直线上的点处计算所得的方体。此后用所得的这四个数对另一个能在目标点作计算的方体进行插值。

然后我们以显然的方式把这推广到三维：把这 64 个点分到四个平面，每个平面 16 个点。在每个这样的平面上执行这种二

擦过程，以得到过目标点的直线上的四个插值点。最后进行插值以得到我们目标点的函数值。这要求总数 21 个一维插值：每个平面五个，再加上最后一个。这过程如图 10-26 所示。其运行的像素总数为时间线性。

关键的问题是：怎样选择平面及每个平面中的线？结果是，在目标点处所得到的最终值与插值所取的维数的次序无关。

三次插值法尤适合于临界抽样的数据。正是依靠相邻点间的某些关联和连续性，所生成的光滑曲线能很好地对稍光滑的物体边界进行拟合，而不产生在原始映像中没有的人工细节。不过，它不是过分光滑化，或者乱砍越过单像素程度的细节。它引进最低限度的加工品。虽然三次插值法对于过少抽样或锯齿状的数据执行得很差，它还是适于 MRI 数据。

此外，因为它仅仅与简单的算术有关及在数据集大小中的线性，所以二维三次插值法对于产生接近于实时而不是高档工作站的典型的 MRI 映像是足够快的。二维三次法放大的结果如图 10-27a。

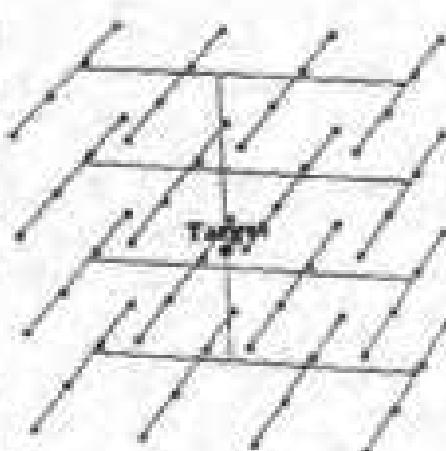


图 10-26 三维三次插值格式



a. 双三次

b. 最近邻域

c. 双线性

图 10-27 不同插值算法的放大

2. 交替的插值方法



图 10-28 过于锐化增加噪声和伪边界

我们考虑并抛弃一些交替的插值方法。

最近邻插值法

在最近邻插值法中，任意点的映像值是邻近的样本点值。这是很快的算法，它对每一维仅需要一个舍入算子。

最近邻插值法常很适合于过少的抽样数据，因为它保存这样的数据的锯齿形而不想让映像穿过清晰的边界时变弱。对黑白线条医学图表的截面来说它也许是最合适的选择。

不过，由于同样的理由，它对临界抽样的数据执行得很差。它的舍取实际上是在光滑的插值保持不变之处，局部地使映像的比例变形（图 10-27b）。

线性插值法

线性插值法取一点的映像值为所知邻近的映像值，按其远近程度作加权平均。这也是很简单且很快的算法。不过它比起最近邻法来有些模糊，并且有许多问题。特别是，它使物体边界保持参差不齐，即使它们没有假，模糊还是太多（图 10-27c）。

基于卷积的方法（略）

10.5.4 映像强化法

在插值时或插值后，我们选择强化映像，使得模糊的区域变得清晰，强化边界，或者去掉枝节。不过，我们发现三维三次方

法执行得相当好，而大部分这些方法不适当或者不必要。人的眼睛极熟练地对朦胧的细节作插值，而三维三次插值法倾向于通过在不定区域中产生模糊但有启发的输出来利用这一点。我们核查的强化算法并没有揭示出眼睛看不到的细节。考虑到在医疗透视中引出的人工细节的危险，我们决定对三维三次方法不予强化。

去伪法

对于清晰或过少抽样的数据，将强对比的边界弄模糊些是有好处的。不过，在我们的情形这是要去掉一些东西；我们的映像里枝节太多，但一般并不是要把物体的边界弄假。

锐化法

传统的锐化算法是将像素值从它邻近的平均移出，也许还要加权使得这种锐化局部化为物体的边界。

这种锐化的主要问题是，它会产生出齿状边，并增加数据的噪声效果（图 10-28）。因为三维三次插值法在很大的插值区域中会产生模糊输出，这种锐化可以用。不过我们发现，它在实际数据的截面上造成极少的视差，且并没有展现重大的新细节。

边界拟合法

许多算法通过寻找映像中的边界来强化锯齿形或模糊的边界（这里的区域其像素不同于其邻近的平均值），用曲线对其拟合并予以强化，或通过去伪，或选择强化。但这种算法通常用于美感方面的强化，而不是在需要科学精度的场合；虽然他们的输出可以很悦目。

映像染色法（略）

10.5.5 执行我们的算法

1. 指定截平面

我们容许用户通过下面两种方式之一来指定截平面：

- 在该截平面上取三个不共线的点，一个很好的方法是基于映像特性取原始的任意值；或者

- 选取要包含的任意点，并且指定两个角，称为欧拉角。第一个角是 $x-y$ 平面与截平面之间的角；第二个是 x 轴和截平面与 $x-y$ 平面的交线所夹的角。这是将映像连续运转的好方法。

为计算截面，我们需要将输入数据变换为三元组 $(\vec{p}, \bar{x}, \bar{y})$ ，这里 \vec{p} 为平面上的任意点，而 (\bar{x}, \bar{y}) 组成平面上的正交基。

2. 三点表示

为从三点表示 $(\vec{p}_1, \vec{p}_2, \vec{p}_3)$ 得到 $(\vec{p}, \bar{x}, \bar{y})$ ，我们取两个向量 $\vec{p}_2 - \vec{p}_1$ 和 $\vec{p}_3 - \vec{p}_1$ 的叉积，以得到垂直于该平面的法向量 \vec{n} 。然后我们从方程组

$$\vec{n} \cdot \bar{x} = \vec{n} \cdot \bar{y} = 0, \quad \bar{x} \cdot \bar{x} = \bar{y} \cdot \bar{y} = 1, \quad \bar{x} \cdot \bar{y} = 0, \quad \bar{x}_x = 0$$

解得 \bar{x} 和 \bar{y} 。前两个方法保证基向量在平面上；次两个保证它们有单位长度；第五个使它们垂直。这五个方程有六个未知量，得不到唯一的基，所以我们需要更多的约束。我们取 $\bar{x}_x = 0$ 作为最后的约束，因为它大大简化了所得的公式。最后我们设 $\vec{p} = \vec{p}_1$ 。

3. 点加欧拉角

对于形如 $(\vec{p}, \varphi, \theta)$ 的输出，我们把平面设想成 $x-y$ 平面关于设在 \vec{p} 的原点作旋转。我们先将平面绕 x 轴旋转 φ ，然后绕 z 轴旋转 θ 。所得的变换表成

$$\bar{x} = R_\theta R_\varphi \bar{i}, \quad \bar{y} = R_\theta R_\varphi \bar{j},$$

其中 \bar{i}, \bar{j} 是 $x-y$ 平面的标准基，而 R_θ 和 R_φ 是相应的旋转

$$R_\varphi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix},$$

$$R_\varphi = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

我们从这些方法得到的向量对我们并非总是最好的。我们想显示与用户的量概念相对应的方向，即只要可能，向上和向左都保持不变。因而，我们想排列基向量尽可能靠近 $x-y$ 平面的基。为此，我们旋转平面的基向量，使得 $\bar{x} \cdot \bar{i}$ 为最大，这使向量 \bar{x} 尽可能靠近真 x 轴。然后我们将方向翻转（有效地翻转映像），只要这种翻转增加 $\bar{y} \cdot \bar{j}$ 值。

我们再次调整基，如果平面与数据量的 12 条边中的一条相交的话，我们就予以计算。这些边的交点定义了量的截面的边界。我们定义的数据量平行于 $(0,0,0)$ 和 $(x_{\max}, y_{\max}, z_{\max})$ 处的顶点。每条边有两个固定的坐标。因而，我们可以通过解两个未知量的两个方程来计算每条边的交。

〔译者注：我们省略作者对算法的执行、检验与讨论。〕

10.5.6 结论

我们的算法用实际数据运作的好处是：

- 截面的稳定与直观的特性，
- 保留大范围的特点，
- 保留原映像中的细节，
- 保留映像与给出像素维数成比例，
- 保留诊断兴趣的特点，接受映像中三维特点的有用色彩，以及
- 为能找到数据所需要的实时操作。

当然，还有大量的地方可改进。

§ 10.6 MRI 切片成像

本节介绍清华大学获特等奖的优秀论文以及译文参考(见[12]).

10.6.1 摘要

为了从 MRI 三维采样数据生成空间中任一位置及任一方向的切片图像，我们在物体空间中及计算机屏幕上建立了两套坐标系，引入了六个参数来描述切割平面，推导了从屏幕坐标到物体空间的坐标的映射公式，设计了六种密度估计算法，即三线性插值法、最近邻法、中值法、控制力法、梯度法及 GNP 综合法，用于从所给的数据来估计空间中任意位置的密度，所有的算法都在某些情况下显现了它们的优点。

我们建立了一个由 10 个尺寸，方向，密度各不相同的椭球组成的三维头模型，通过在物体空间的均匀采样来生成数据集，使用了多组参数来检验模型和算法的成像能力。

在对算法结果进行了主、客观的比较之后，我们总结了这些算法的优、缺点，对于一般的应用，我们推荐梯度法与 GNP 综合法。在大多数情况下，这两种算法都能产生平滑且明显的边界。

算法的测试与比较使用了我们自己的一个基于 WINDOWS 95 的程序。

10.6.2 有关 MRI 的一些事实

MRI 有如下一些与本问题有关的特性：

- 高精度 MRI 的扫描精度约为 1~3mm，也就是说，MRI 可以容易地区分 1~3mm 尺寸的细微差别。一般使用的 MRI 切片尺寸不超过 25cm×25cm (见[13])。

- 高对比度 MRI 的优点之一是它所生成的图像的高对比度，这使得器官的边缘十分明显，利于诊断（见[13]和[5]）。
- 长扫描时间 MRI 的扫描时间大约为几分钟。例如，以 $T = 1.5\text{s}$ 的脉冲重复时间对物体进行扫描从而得到一典型的二维图像 ($128 \times 128 \times 256$)，需要大约 6 分钟（见[13]）。虽然在实际应用中已经采用了一些技术来加速整个过程，但这些仍然是 MRT 的主要缺点之一。因此，我们不能期望已知的数据集已足够充分来产生一个好的切片图像（那可能需要过多的时间）。切片算法也不应太复杂或者太耗时。
- 重建算法 通常被用来从 MRT 生成的原始数据中重建三维信息的两种方法是投影重建(PR)和 Fourier 变换。它们都可以在空间均匀间隔的数组的形式给出结果。因此，可以假定题目所给的数据集在空间里是均匀分布的。

10.6.3 假设、坐标和符号

1. 模型假设

从本问题的要求和 MRI 的特性出发，我们作出如下假设：

- (1) 被扫描物体的最大尺寸为 $256\text{mm} \times 256\text{mm} \times 256\text{mm}$ 。这样的尺寸在绝大多数情况下是足够的。如果扫描的是一个比这更大的物体，我们可以把扫描得到的数据分为几个 $256\text{mm} \times 256\text{mm} \times 256\text{mm}$ 的立方体。
- (2) 期望的切片成像精度是 1mm 。我们将在计算机屏幕上显示由我们的算法生成的切片，每个像素代表 $1\text{mm} \times 1\text{mm}$ 的区域。
- (3) 给定的已知数据集是一个三维数组 $A(i, j, k)$ ，这个数组是在整个物体空间里沿坐标轴以某个均匀的间隔采样得到的。这样的间隔大约是 $2 \sim 4\text{mm}$ ，对于 MRI 在一个不很长的时间里完成扫描是足够大的。稍后本文将讨论给定数据不是等间隔采样的情形。
- (4) $A(i, j, k)$ 取 0 到 255 范围内的整数值，这些值从高到低代表

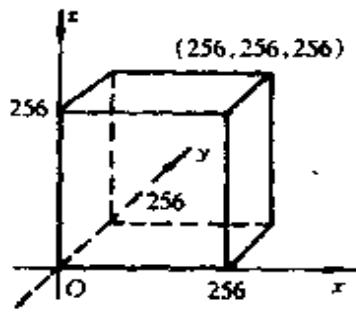


图 10-29 数据
空间坐标系

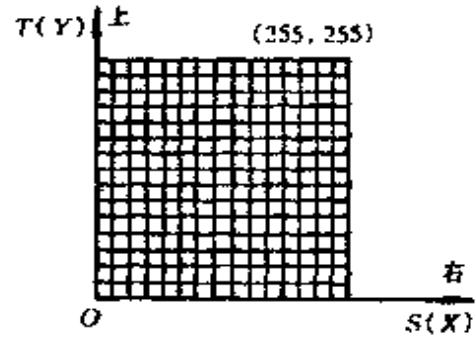


图 10-30 屏幕坐标系
(原点在屏幕的左下角)

了水分的密度。在我们的屏幕显示中，0 用黑色表示，255 用白色表示，而其他值则用不同程度的灰度表示。

- (5) 扫描的物体由不同的成分组成。假定在一种成分里密度的变化不会很大。例如，在头盖骨中水分浓度总是很低的。所以在同种成分中，颜色或灰度将几乎不变化，或仅在一个小范围内变化。
- (6) 用 MRI 检验的物体是某种动物或者人类的身体，因为身体中的器官或者组织一般较为柔软，我们可以想像在绝大多数情况下边缘是平滑而明显的。例外只会发生于某些骨头之间，诸如脊椎（它们的边缘是明显的，但不平滑），或一些病态的组织中。
- (7) 一个位置的未知密度是受所有给定数据影响的。不过，点之间的距离在本问题中起着很大的作用，远离未知点（例如 50mm）的位置被假定为没有影响或者仅有很小的影响。

2. 坐标系

根据上面的假设，我们在数据（或物体）空间中和计算机屏幕上建立了两个坐标系，分别如图 10-29 和图 10-30 所示。为了方便起见，这两个坐标系中的单位（在屏幕图像里为像素）都是

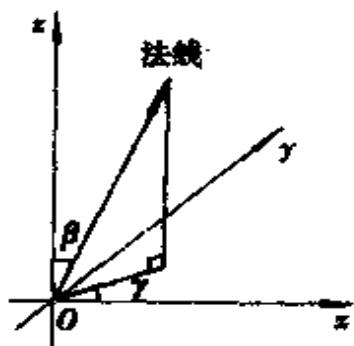


图 10-31 将 P 的
法线旋转至指定
方向

1mm，因为物体尺寸为 256mm×256mm×256mm，所以数据空间正好是 $0 \leq x, y, z \leq 256$ 。

3. 用语及符号说明

在本文中“密度”和“浓度”都指密度在屏幕上的 256 级灰度表示。短语“未知点”或者“未知位置”指的是这样的点（或位置），在该点处物体的密度并未作为已知数据给定，需要用我们的算法计算出来。另外，用于对成像质量的评价，我们用“锐度”表示物体内不同成分间边缘的显著程度，用“平滑度”表示边缘的光滑程度。

另外，本文中经常使用的符号概述如下：

$A(i, j, k)$ 给定的三维数据，表示空间某位置的密度。在一些上下文中， A 也表示那个位置。

s_x, s_y, s_z 分别是沿笛卡尔坐标轴的三个采样间隔。这样 $A(i, j, k)$ 就是位置 (is_x, js_y, ks_z) 处的密度。

$\alpha, \beta, \gamma, x_0, y_0, z_0$ 用来定义切面的六个参数，在下面“切片平面”中定义。

$D(x, y, z)$ 在位置 (x, y, z) 处物体的密度。

10.6.4 问题分析

考虑一个与被扫描的物体相交的平面，我们想知道的是物体在这个平面上各点的密度。如果我们能够将切面中点的坐标转换成物体真实的三维坐标，并且计算出相应的密度，问题就解决了。只要对空间几何有一些基本的知识，第一步就可以解决。但是，如何计算未知的密度呢？

1. 能知道未知的密度吗？

从著名 Nyquist 采样定理，我们知道要准确地重建被扫描物体的全部密度信息，采样间隔必须满足不等式：

$$\max(s_x, s_y, s_z) \leq \frac{1}{2f_m}, \quad (1)$$

其中 f_m 是物体密度在空间分布的频率的上界。在我们的问题中，如果需要准确绘出任意方向和位置的切片的话，不等式(1)也需要被满足。不过情况并不是这样的。

一方面，在实际情况下(1)不可能成立，因为一个物体的 f_m 总是很大——实际上是无穷大，没有什么采样间隔可以满足这样的不等式；另一方面，为了绘出切片我们并不需要准确地知道未知的密度，因为灰度级是从 0 到 255 的整数，小于一个灰度级的误差是容许的。事实上，某种程度的模糊总是不可避免的，在这种意义上，我们能够知道未知的密度。

2. 怎样知道未知的密度？

因为我们不能准确地知道未知的密度，我们必须尝试找到某种方法来估计未知的密度。我们要在如下方面作出选择：

- (1) 简单和复杂；(2) 局部和全局信息；(3) 静态和自适合算法。

10.6.5 模型描述

我们的模型由三部分组成：给定的数据、切片的描述以及用来估计物体在任意位置的密度的算法。切片的描述用于将点的屏幕坐标转换成空间坐标，而密度估计算法则用于由给定的数据获得那个点的密度。这样切片就很容易被显示在屏幕上。

下一节将详细讨论我们所提出的六种不同的密度估计算法。给定的数据在问题和“模型假设”中已有描述，所以本节将给出第二部分即切面的描述，同时给出从屏幕坐标到空间坐标的映射公式。

1. 切片平面

为了在空间定义一个有着任意方向和位置的切面，我们可以通过以下 4 个步骤来把 X-Y 平面转换到空间中任意平面：

- (1) 把一个平面 P 与它自身的 S-T 坐标系(与屏幕坐标系相同)，放到 X-Y 平面上并且使两个坐标系的原点和方向

相重合.

- (2) 绕平面 P 的法线(即 Z 轴)将 P 旋转 α 角, 使 $S-T$ 的方向不同于 $X-Y$.
- (3) 绕原点将 P 的法线旋转至指定的方向. 在图 10-31 中, 这个方向用角 β 和 γ 来定义.
- (4) 平移平面 P , 把 P 的原点移至空间中的给定点 (x_0, y_0, z_0) .

这样用六个参数 $(\alpha, \beta, \gamma, x_0, y_0, z_0)$, 我们可以在空间任何方向和位置, 用一个和屏幕坐标系统相同的坐标系定义一个平面.

2. 从屏幕到空间的映射

因为屏幕坐标系和切面坐标系是相同的, 我们可以把屏幕坐标改为切面, 然后使用前面所描述的变换关系, 把切面坐标转换为空间坐标.

假设屏幕中位置在 (s, t) 的像素在空间中对应于点 (x, y, z) , 根据变换我们可得:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix} \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \quad (2)$$

(2) 是从屏幕到空间的映射公式. 这样我们解决了问题的第一步.

10.6.6 密度估计算法

考虑屏幕上的某个像素及其在物体空间 (x, y, z) 的对应点 U . 密度估计算法的目的就是估计 U 处的密度 $D(x, y, z)$.

我们尝试了五种基本类型的密度估计算法, 即三线性插值法、最近邻法、中值法、控制力法和梯度法. 基于这些算法的实

验结果，我们还设计了一种整体上最有效的算法，即 GNP 综合法。下面是这些算法的描述和简评。算法的成像效果及相互比较在本章的下一节中给出。

1. 三线性插值法

通常，线性插值能给出较满意的结果。在三维空间中，我们用三线性插值，对三个坐标方向上的 8 个邻点进行插值，即

$$\begin{aligned}
 D(x, y, z) = & A(i, j, k) \cdot (1-u) \cdot (1-v) \cdot (1-w) \\
 & + A(i+1, j, k) \cdot u \cdot (1-v) \cdot (1-w) \\
 & + A(i, j+1, k) \cdot (1-u) \cdot v \cdot (1-w) \\
 & + A(i, j, k+1) \cdot (1-u) \cdot (1-v) \cdot w \\
 & + A(i+1, j+1, k) \cdot u \cdot v \cdot (1-w) \\
 & + A(i+1, j, k+1) \cdot u \cdot (1-v) \cdot w \\
 & + A(i, j+1, k+1) \cdot (1-u) \cdot v \cdot w \\
 & + A(i+1, j+1, k+1) \cdot u \cdot v \cdot w, \quad (3)
 \end{aligned}$$

其中 ($[x]$ 是不大于 x 的最大整数)：

$$\begin{aligned}
 i &= \left[\frac{x}{s_x} \right], \quad j = \left[\frac{y}{s_y} \right], \quad k = \left[\frac{z}{s_z} \right], \\
 u &= \frac{x}{s_x} - i, \quad v = \frac{y}{s_y} - j, \quad w = \frac{z}{s_z} - k.
 \end{aligned}$$

这种方法用 8 个邻点的密度来确定点 U 的密度，在几乎所有情况下，能缓解边界不连续的问题。但是，由于内在的低通滤波的性质，它常使锐度明显的边缘变模糊。

2. 最近邻法

为了保持边缘的锐度，我们尝试了最近邻法。这种方法假定点 U 和它在空间最近的已知邻点属于同一成分，因此被赋予同一密度值。

最近邻法非常简单，计算量很小。但这种方法的性能很不稳定，尽管有时它会给出较好的结果，然而，它部分保持了边的锐度，如果恰当地与别的方法结合，可能会有出色的效果。

3. 中值法

在信号处理与图像处理中，中值滤波是很有名的。它能在平滑信号的同时保持边缘的锐度不被破坏。在我们的算法中，我们把 U 的邻点的中间密度值赋给 U 。这种算法给出了明显的边缘，但同样也出现了羽状突起，使边界看起来不符合真实的情况。

4. 控制力法

既然我们认为点与点间的距离很重要，我们就可设想与 U 的距离在一合理范围内的点对 U 的密度都有一定的控制力，这种力使 U 的密度与控制点的密度相近，并且这种控制力随着距离的增加而减少。最后所估计的 U 的密度是这些点的密度依其控制力的大小的加权平均。

在我们的算法中，与 U 距离为 d 的点 $A(i, j, k)$ 的控制力定义为：

$$p = \frac{1}{1 + e^{5(d/d_0 - 1)}},$$

其中 $d = \sqrt{(x - i \cdot s_x)^2 + (y - j \cdot s_y)^2 + (z - k \cdot s_z)^2}$ ，而 d_0 是距离阈值（ $d = d_0$ 时控制力 $p = \frac{1}{2}$ ）。 U 处的密度估计为：

$$D(x, y, z) = \frac{\sum_{d_i \leq 2d_0} p_i^* \cdot A(i_i, j_i, k_i)}{\sum_{d_i \leq 2d_0} p_i^*}, \quad (4)$$

求和对所有距离在 $2d_0$ 内的已知点进行。当采样间隔为 2mm 时， d_0 取 1mm。

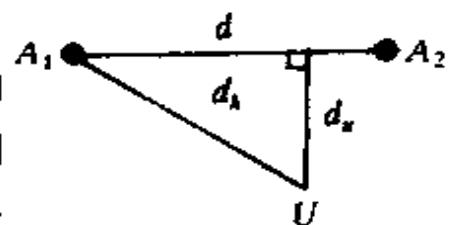
尽管公式(4)与三线性插值法的公式(3)很相似，但控制力定义中的非线性性使这种方法给出的边界比三线性插值法所给出的更平滑，但也更模糊。

5. 梯度法

上面的算法都是基于单个点对未知点的影响。当一对点对未知点的影响被考虑时，我们引入了梯度算法。

图 10-32 中 $A_1(i_1, j_1, k_1), A_2(i_2, j_2, k_2)$ 为两已知点密度值, U 为未知点。 A_1, A_2 间的距离为 d , $\overrightarrow{A_1U}$ 到 $\overrightarrow{A_1A_2}$ 的投影为 d_h (当 $\overrightarrow{A_1U}$ 与 $\overrightarrow{A_1A_2}$ 所夹为钝角时, d_h 为负), d_v 是点 U 到 $\overrightarrow{A_1A_2}$ 的距离, 如果只考虑 A_1 到 A_2 的梯度的话, U 处的密度为:

$$D(x, y, z) = A_1 + \frac{d_h}{d} (A_2 - A_1).$$



然而, 当考虑到点 U 附近的其他点对时, 密度 D 同样也应是所有梯度作用的加权平均, 其权重(类似于控制力)为:

$$p = \begin{cases} e^{-d_v}, & \text{当 } d_h \geq 0 \text{ 时;} \\ \frac{1}{4}e^{-d_v}, & \text{当 } d_h < 0 \text{ 时.} \end{cases}$$

这种算法不但利用了未知点 U 附近的密度信息, 而且利用了局部范围内密度的变化趋势。这赋予了算法一定程度的自适应能力。不仅如此, 我们还可以增加一些全局信息。例如在算法的实现中, 当 A_1 与 A_2 很接近时 ($|A_1 - A_2| < 20$), 权重 p 就被乘以 3, 因为此时 A_1 和 A_2 极可能处于同一组成成分中, 从而使 U 也处于这一成分中的可能性很高。相反, 当 $|A_1 - A_2| > 80$ 时, 权重 p 被乘以 0.7, 因为 A_1, A_2 极可能属于不同组成成分。

6. GNP 综合法

实验结果(见下一节)表明梯度法与控制力法能给出平滑但稍有模糊的边缘, 而最近邻法则给出对比强烈、边缘粗糙的图像。综合它们的优点, 我们设计了 GNP 综合法。简单地说, 便是将梯度法、最近邻法和控制力法以 3:2:1 的比例加权平均。

10.6.7 算法检验

1. 数据集-头模型

算法的检验和不同算法之间的比较都需要有合适的数据集。

似乎实际的 MRI 数据是最合适的。但除了获取数据上的不方便外，使用实际的 MRI 数据的另一个困难在于我们很难比较成像的切片和实际的切片，因为后者无法得到。

受被广泛采用的 S-L 头模型(见[13])的启发，我们扩展并设计了三维头模型，它是由十个位置、形状、方向、密度不同的椭球组成的(详细描述可见附录(略))。头模型中空的部分用环境密度填充。(注意：环境密度不同于成像切片中的背景，背景表示头模型以外的 MRI 数据部分。)

我们采用椭球，是因为它简单，并且不同类型的椭球组合起来可模拟很多真实物体，例如脑和胃等。我们设计了三种有着不同密度分布的椭球来模拟真实世界：

类型 1——密度均匀；

类型 2——从中心到边缘密度呈线性变化；

类型 3——在类型 2 的基础上加一随机噪声。在我们的实验中，这种类型未被考虑，因为噪声过滤超出了本文的范围。

当采样间隔确定后，通过判断采样点位于哪个椭球内，就可以方便地产生数据集(附录略)，这样生成的数据集很大，例如当采样间隔为 2mm 时，数据集为 $128 \times 128 \times 128$ ，即 2M 字节。

除此之外，我们还利用头模型产生实际的切片图像，其计算过程与上面产生数据集的过程相似。

2. 实验及结果

一个重要的问题是怎样评价不同算法的输出结果。一种办法是主观的视觉观察。在本文所处理的问题中，主要的两个目标(边缘的锐度和平滑度)很难定量地测出，因此我们主要用视觉观察的方法来比较不同算法的性能。同时，虽然成像切片与实际切片密度差的均方根(RMS)不能全面合理地反映成像切片的质量，它仍然有益于对成像效果的评价。所以我们把 RMS 误差最小化当作我们的第三目标。(这个目标只在我们的仿真程序中有用，因为在实际应用中，真实的切片是未知的。) 最后，因为数

据集相对来说很大，我们也考虑了算法的计算量。

我们做了大量实验来比较各种算法的效果，其中有代表性的几种情况可见图 10-33~10-36（图下附有相应切片的参数。所有图中 $x_0 = y_0 = z_0 = 2.5$ ）。以下将逐个介绍这些实验结果，并由此对算法作出评价。图 10-33 中，切片在物体中间且与 X-Z 平面平行。这种情况下，切片穿过所有 10 个椭球，这有利于比较各种算法的整体性能。

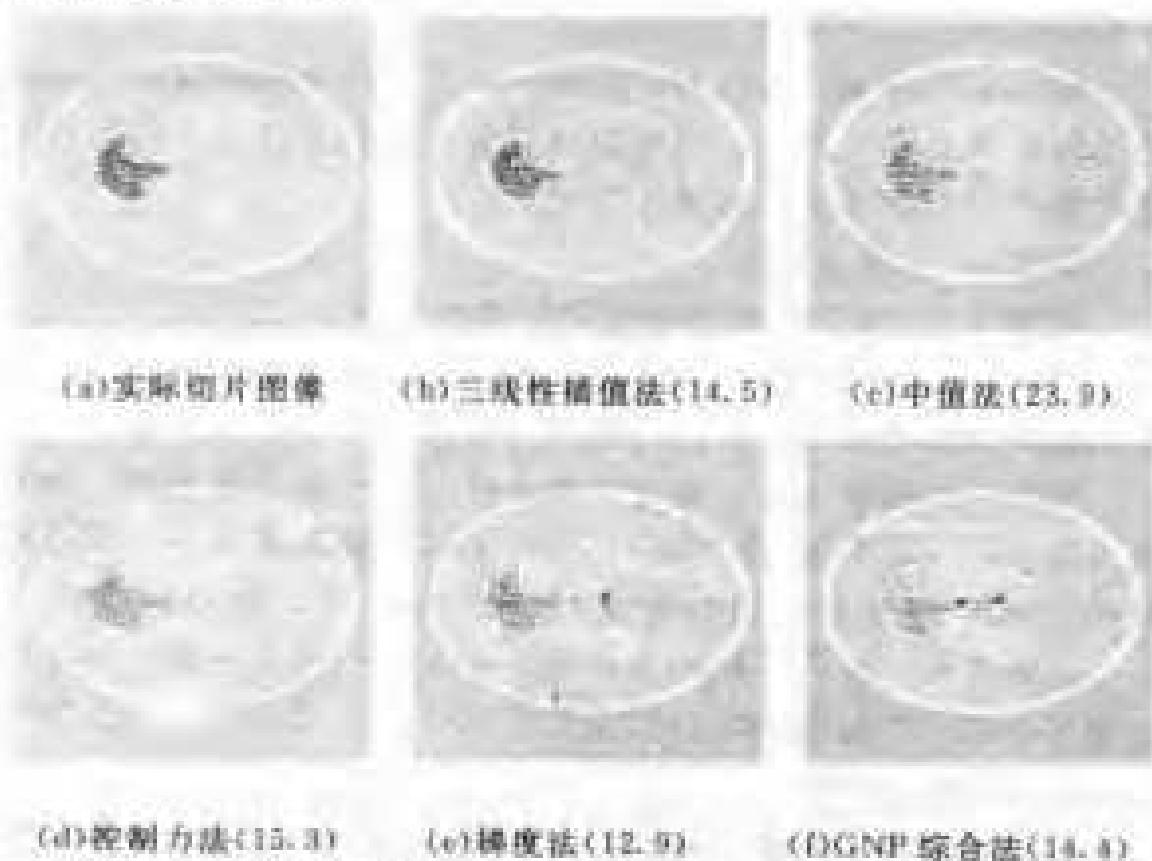


图 10-33 切面参数为 $x_0 = 0, y_0 = 128, z_0 = 0, \alpha = 0, \theta = 90^\circ, \gamma = 90^\circ$
(括号内的数给出了算法的 RMS 错差)

图 10-34 中，切面是倾斜的，所有算法都给出了较好的结果。图 10-35（略）中的切面是由图 10-34 中切面平移一小段距离得到的，但部分算法的效果差了许多。图 10-36 中，切面处在一个临界的位置，这时我们考察各种算法在这种极端情况下的性能。

3. 算法评价

一般情况下，三线性方法性能不错。这种方法的 RMS 误差较小，计算时间较短，但它通常模糊了图像，并且在极端情况下很令人失望（图 10-36 (b)）。当物体中不同组成成分的密度差别较小时，不应使用这种方法。

最近邻法和中值法能保持明显的边缘，并且用时最少。但它们给出的边界不平滑，并且不能区分小物体（图 10-36(c)）。切面的小位移会使它们给出锯齿形的边界（图 10-35(c)略）。不可避免地，它们的 RMS 误差很大。但是，当数据集很大以至计算时间很重要，或者锯齿状的边缘不影响后处理时，这两种方法仍是适用的。

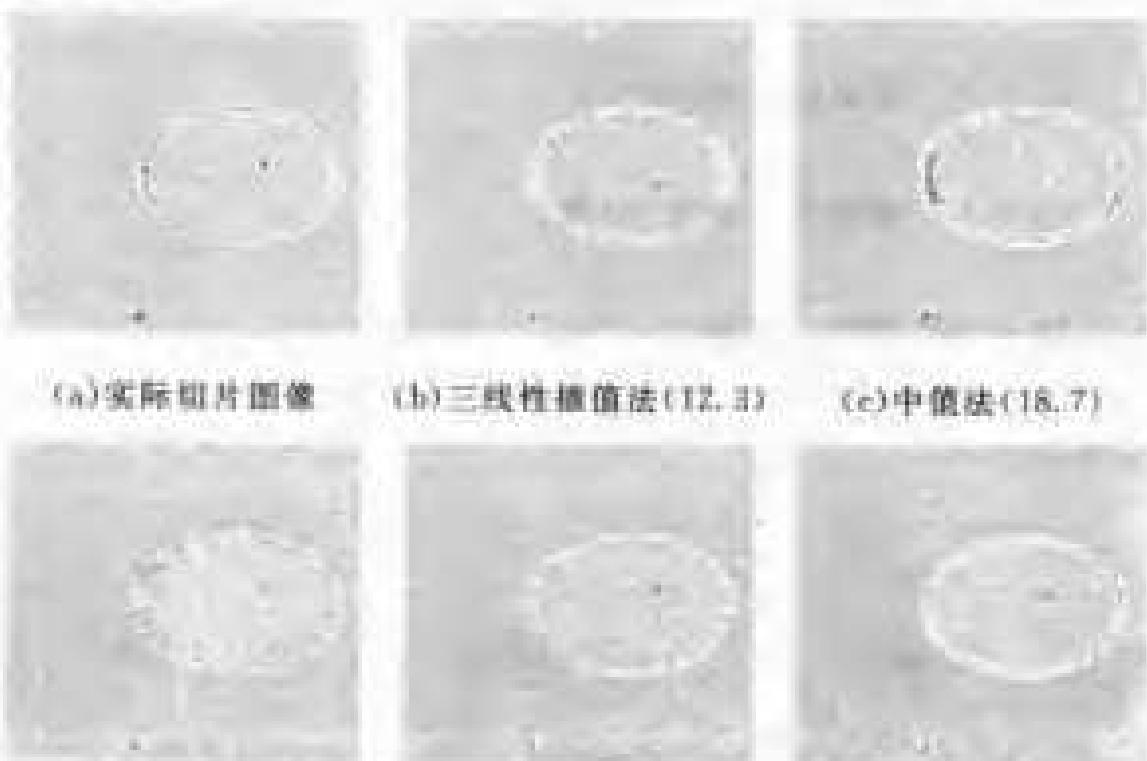


图 10-34 重建切面的参数 $x_0=0, y_0=128, z_0=0, \alpha=0, \beta=45^\circ, \gamma=90^\circ$
(切片中左边的灰色区域为扫描物体之外的空间)

梯度法在所有情况下都给出了最小的 RMS 误差。而且，成

像切片的平滑度和锐度都很令人满意。GNP 综合法的 RMS 误差比三线性插值法的稍大，但当同时考虑平滑度和锐度时，它超过了所有其他算法。因此，当计算时间不甚重要时（目前这两种方法在奔腾 166 上花时 10~14 秒，另几种算法为 2~3 秒），这两种方法最有竞争力。在切面的位置、方向处于临界、极幅情况下，它们尤其有效。

4. 采样间隔太大时的成像效果

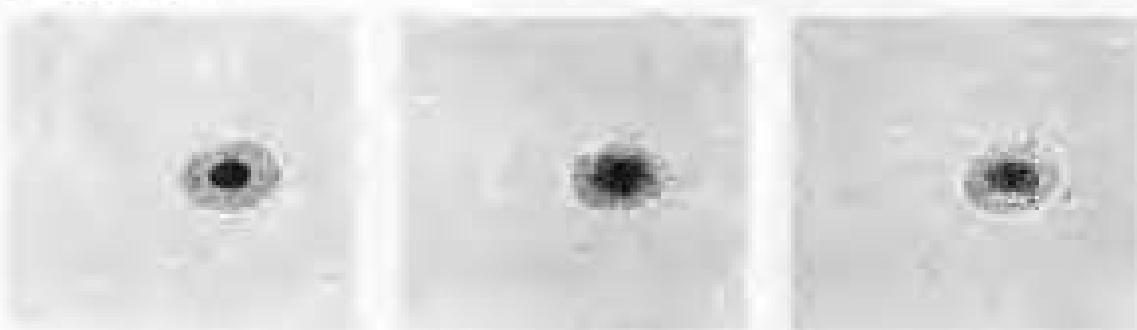
为了验证我们关于采样间隔的讨论（见“能知道未知的密度吗？”），我们测试了当 $s_x = s_y = s_z = 4$ 时的成像结果。不出所料，切片成像的质量差了很多，例如图 10-37（略）中一些相连的细边被打断了。

10.6.8 模型的优缺点（略）

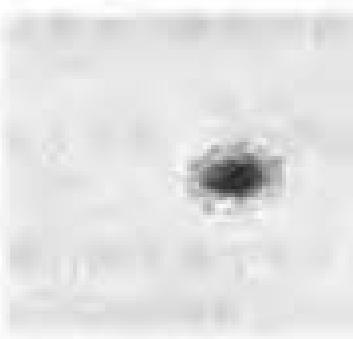
10.6.9 模型推广及建议（略）

10.6.10 附录（略）

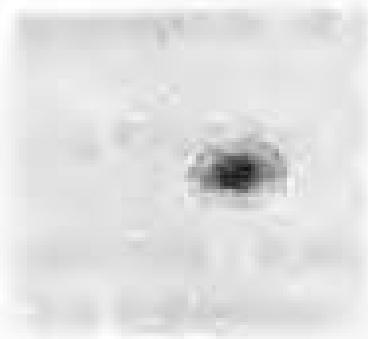
（原文中给出了仿真程序的使用说明以及头模型的具体参数及构造方法）。



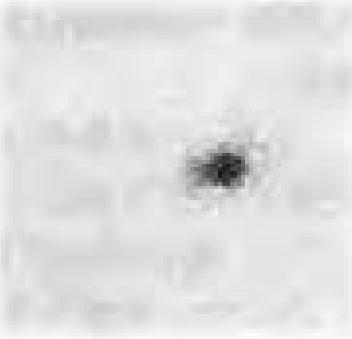
(a) 实际切片图像 (b) 三线性插值法(12,3) (c) 中值法(18,3)



(a) 控制方法(13.8)



(b) 梯度法(13.2)



(c) GNP 综合法(13.4)

图 10-36 倾斜切面, 参数为 $x_0=0, y_0=120, \alpha=0, \beta=0, \theta=70^\circ, \gamma=60^\circ$

§ 10.7 论文评述

这里我们对上述优秀论文的解答作出评述，并介绍本题阅卷者对成绩的评定以及命题人对参赛队解答的一些评价。最后附本章的参考文献。

10.7.1 对扫描问题解答的评述

按本问题所述，现在给出的 MRI 像像皆为垂直的或水平的三维扫描切片，要求参赛队设计一种算法，它能在任意斜平面上确定出尽可能保持原有灰度值的映像，并对所得的算法进行测试，这里包括确定斜平面的位置、选定适当的插值算法和提出数据实例予以检验等三方面的问题。

1. 首先是是如何确定三维空间 R^3 中的斜平面的位置，有几种解决方法：

(1) 常见的方法是提出一个平面 $Ax + By + Cz = D$ ，并用标准的矩阵变换让它旋转，这个平面通常可取为 (x, y) 平面，如 Harvey Mudd 学院和清华大学参赛队所做的那样。

(2) 选择一个点和两个角来定义其平面，如 Eastern Oregon 大学参赛队所做的那样。

(3) 取平面上不共线的三点，其中一点为原点，再其余两点

可得到一组特定的正交基，如 Macalester 学院的参赛队所做的那样。

(4) 还有其他方法，如选定三维空间中的两个点，并定义这些点之间的平面。

2. 选择并推广插值算法，以尽可能保持原有的灰度值。据阅卷者说，参赛队大多意识到关键之处在于将三维数据集映射到斜平面上的坐标系，以得到平面中元素的灰度格式(0~255)。三维数据集由三个整数来定义，而斜平面上的点为实数。必须对所选用的方法予以发展，使得能对斜平面上所有点处的灰度值作插值。在这方面，这些参赛队运用并发展了多种方法。

下面这些方法是几乎各个参赛队都用到的：

- (1) 用八个或更多的点的最近邻算法；
- (2) 加权点算法；
- (3) 样条法(由三次的线性)；
- (4) Lagrange 多项式法；
- (5) 几种方法的混合算法。

有些参赛队发展出一些特别的插值方法，如 Harvey Mudd 学院提出一种基于 Laplace 方程的三维线性逼近。为得到在边界处的较好的成像效果，一些队还提出各种强化和锐化的方法，并结合算例进行讨论。

各参赛队通常想用多于一种方法来得到灰度值。但据阅卷者说，他们对各种方法的比较论述总的较为一般，缺乏深刻的分析。这里所选的几份，有些(如 Eastern Oregon 队)也特别比较和分析了他们的方法和结果，并得到有效的结论，算是其中较好的。

3. 设计出一组或几组适当的数据，用于检验所提出的算法。几乎所有的队都选择大脑来构造其模拟数据集，有些队还选择了其他组织器官，如 Eastern Oregon 大学还选择了手臂等。本问题的叙述要求参赛队设计并测试一种算法，它能对于数据集所在

的空间中的任意一个斜平面产生一个阵列截面，使之尽量接近原来的灰度值。对灰度的利用程度是可以区分的。有些队用颜色强化他们的表示；这是可接受的，只要灰度不都删去。但有些队想去着色是因为他们的灰度分辨率不能探测出某些诊断元素，这与问题的意愿相违，因为问题的叙述要求用到灰度。

10.7.2 阅卷者对成绩的评定

要严格地区分各个参赛队的名次总是困难的，阅卷者提出他们更密切关注的是问题叙述中的“必须”和“应该”两个要求：

1. 参赛队的算法必须在空间中得到按平面的三维阵列切片图。这成为参赛队是否至少能列入顺利参加者范围之内的关键因素。阅卷者希望见到的是图像，而不是矩阵的记叙。图像被仔细地检查，以看到它们是否作为斜切片出现。

2. 参赛队应该：

- (1) 设计数据集来检验并演示他们的算法。
- (2) 生成的数据集应反映出诊断值所属的条件。
- (3) 描述限制其算法有效性的数据集。

因而，阅卷者寻找所选数据集的好的描述和诊断值元素的描述。有些队在寻找体内肿瘤或异常物时用口语来表述，这是可接受的。有些队还提出将方块内的球作为他们所表示的数据集。如果参赛队在其大范围的三维元中设置某种诊断值，他们的数据集仍可被接受。

有些队对数据集的刻画使其算法的有效性受到限制，这大大背离了“应该”的要求。在有效性方面受到限制的任意数据集，对其内容作口头上的描述也是可以被接受的，以此可区分出最优秀的论文。

阅卷者看到另一个未一致完成的要点是某种误差分析。很少的队为精度与他们三维空间数据集中相应的整数点作对照，来检验他们平面上的整数值。阅卷者赞扬完成这些工作的队，几乎每

一队都引用他们作为“输出”的图来解释或想显示出精度。有些队用“模糊对清晰”的边界来作为他们仅有的分析基础。

阅卷者认为，叙述的风格和明晰性也是被看做另一个要点。各个队的组织以及对他们方法解释的能力将参赛者区分开来。好的组织和团结的集体帮助了优秀队。

10.7.3 命题人的评价

命题人 Nievergelt 教授认为，这次关于这个纯数学建模问题的优胜，解答的一个引人注目的新事，是这些学生队掌握了几种电子工具，与数学一起熟练地使用。

1. 第一种电子工具是万维网 WWW，各队用它在不同程度找到关于磁共振成像的一般医疗信息、人脑的实际和模拟的三维数据集以及二维插值的数学算法。不过有两个队找到了他们需要的印刷物信息，然后熟练地生成他们自己的试验数据。
2. 第二种电子工具是计算机图形，所有的队有效地运用它来传播自己的结果。对本问题，有一个队注意到似乎不存在对演示所作的能替代最终可视医疗诊断的任意数字估计（如根平均平方或其他范数），因而图像可用以最好的方式将算法与现实作比较。
3. 第三种电子工具由计算机程序组成，参赛队利用它来作坐标变换，实际是对空间中的平面进行等距参数化，以及作三维插值。
4. 所有队适当地用到的第四种电子工具是准备其最终的论文，包括叙述、数学公式与图像。

所有这些工具当然帮助了问题的本质部分，即数学。参赛队以数学显示出对概念和细节的很好主导。第一个例子是将概念的关键之处放在开头，这里所有队意识到该实际问题能变成三维插值的数学问题。第二个例子是在从一或二维到三维插值法作推广。

时细节变得很重要。有一个队(清华大学)已经知道这结果，其他队(Eastern Oregon 大学和 Harvey Mudd 学院)对他们的数学推广提供了很好的解释和证明。

最后，所有队表明有效地利用了他们的时间，在用于研究的时间和用于室内产生数据和算法之类的时间方面取得平衡。在寻找和再创作之间的这种平衡行为对于按时交上计算机工作程序来说实际上是重要的。例如，没有队表明他们用过从 WWW 得到的三维插值计算机程序，也许是因为它不能明显得到。事实上，对于 <http://neilib2.cs.utk.edu> 处 Netlib 关于“三维插值法”作搜索表明，一维和二维的程序为 toms/474 (双三次插值法)，而没有揭示任何特别的三维程序。但是这种程序是存在的，而找到它们并使用它们也许比得到要费更多的时间。例如，在 <http://dtinet33~199.dt.navy.mil/dtnurbs/about.htm> 有运用非一致有理 B-样条的多维(不限维数)插值法(NURBS)程序。

参 考 文 献

- (1) Acton, F., 1990. Numerical Methods That Work. Washington DC: Mathematical Association of American.
- (2) Andrews, H. C., and B. R. Hunt. 1977. Digital Image Restoration. Englewood Cliffs, NJ: Prentice Hall.
- (3) Ballinger, R., 1997, Gainesville VAMC MRI Teaching Files.
<http://www.xray.ufl.edu/~rball/teach/mriteach.html>.
- (4) Cocosco, C., V. Kollokian, R. Kwan, and A. Evans. 1997. Brain Web: Simulated Brain Database.
- (5) Frommhold, H., and R. Ch. Otto. 1985. New Methods of Medicine Imaging and Their Application (German). (中译本：新医学成像方法及其临床价值，中国医学科技出版社，北京，1988)

- (6) Hornak, J., 1997. The Basics of MRI. <http://www.cis.rit.edu/htbooks/mri/>.
- (7) Johnson, K., and A. Becker, 1997. The Whole Brain Atlas. <http://www.med.harvard.edu/AANLIB/home.html>.
- (8) Lancaster, P. And K. Saltauskas, 1986. Curve and Surface Fitting. London: Academic Press.
- (9) Press, W. Et al. 1988. Numerical Recipes in C. New York: Cambridge University Press.
- (10) Rosenfeld, A. And A. Kak. 1982. Digital Picture Processing. 2 vols. San Diego, CA: Academic.
- (11) Press, Summit, S., C Programming FAQs. 1996. New York: Addison-Wesley.
- (12) 倪江、陈俊、李凌. 1998. MRI 切片成像. 数学的实践与认识, 第 28 卷第 3 期, 271~280.
- (13) 高上凯, 1996. 医学成像系统, 清华大学电机系.

第十一章 分数贬值问题

俞文毓

华东理工大学 应用数学研究所

提 要

本章介绍 1998 年美国大学生数学建模竞赛 B 题的提法，有关情况分析，部分优秀答卷和若干其他答卷，并作出了讨论。

§ 11.1 问题和背景

11.1.1 问题

以下先给出美国大学生数学建模竞赛 1998 B 题的翻译。

问题（1998 B 题）：分数贬值问题

背景：ABC 学院的一些管理者关注课程评分问题。平均地说，ABC 学院的教师经常给出高分（目前的平均分为 A-），从而难以区分好学生与普通学生。根据一项奖学金的规定，仅允许前 10% 的学生得到资助，所以需要作出课程排名。

学院院长的想法是，将每个学生与其他学生在各个课程中作比较，用此种信息建立排名。例如，如果一个学生在某课程得到 A，而其他学生在此课程中也得到 A，那么该学生在此课程中仅仅是“中等”。另一方面，如果一个学生在某课程中得到唯一的 A，那么该学生显然是“优于中等”。综合从多个课程得到的信息，或许可以使学院内的学生被分成“十分点排名”（Decile Ranking），指前 10%，次 10%，等等。

问题：

假如所采用的计分是 (A_+ , A , A_- , B_+ , ...), 学院院长的上述想法是否行得通?

假如所采用的计分仅是 (A , B , C , ...), 学院院长的上述想法是否行得通?

其他方式能产生所需要的排名吗?

一个担心是, 单个课程的计分可能改变许多学生的“十分点排名”. 这是否可能?

数据集:

参赛队应当设计数据集, 以检验与显示所采用的算法. 参赛队还应指明, 数据集存在什么特征时, 所采用的算法会受到局限?

11.1.2 该问题的有关情况

问题 1998 B 的上述背景说明涉及美国大学中广泛采用的评分体系, 更清楚地说, 采用的等级分(等级点, Grade Point)构成 12 级计分制如下:

A_+ , A , A_- , B_+ , B , B_- , C_+ , C , C_- , D_+ , D ,
 F

它们可以定量化为下列分数:

4.3, 4, 3.7, 3.3, 3, 2.7, 2.3, 2, 1.7, 1.3, 1, 0.

在一个学院里, 一般有上千名或数千名学生, 衡量学生成绩的常用方法是采用“等级分平均值”GPA(Grade Point Average), 从而可将学生按 GPA 排名. 一般来说, 这种排名方法的优点是: 简单, 基本公平与逻辑合理的. 但是, GPA 排名方法容易引导学生去选读较容易或者较易取得好分数的课程, 从而也间接地引导教师给学生以较好分数, 造成分数贬值及与此相关的恶性循环. 在选课自由度较大及同年级学生所读课程差异颇大的情况下, 对于这个在分数贬值情况下的学生成绩排名问题, 不得不加

以深入地分析与处理，以便更加公正地选拔应该得到奖学金的学生，例如，公正地选拔百分之十的学生。

以下是一个用以说明上述问题的小规模虚拟例子，在该例子中，4个学生分别在9门课程中选读5门课，结果如下：

	学生 1	学生 2	学生 3	学生 4	课程 GPA
课程 1	B ₊			B	3.00
课程 2	C ₊		G		2.15
课程 3			A	B ₊	3.65
课程 4	C ₋	D			1.35
课程 5		A		A ₋	3.85
课程 6	B ₊			B	3.15
课程 7		B ₊	B		3.15
课程 8	B ₊	B	B ₋	C ₊	2.83
课程 9		B	B ₋		2.85
学生 GPA	2.78	2.86	2.88	3.00	-

在上例中，学生 1 的 GPA 低于学生 4 的 GPA，但是学生 1 与学生 4 共同选读的课程中，学生 1 的分数总比学生 4 的分数要好一些。这个例子能说明，1998 B 题中 ABC 学院院长的观点是合理的，他认为，应该强调学生之间在选读同一课程时的成绩比较，以此作为学生成绩排名的依据。

该问题属于教学管理研究的范围，从 1970 年以来在美国陆续有一些研究报告与研究论文发表，这方面可参见 V. E. Johnson, An alternative to traditional GPA evaluating student performance, *Statistical Science*, 12 (4), 1997, pp. 251~278. 也可查找该论文所引用的文献。

总的来说，该问题尚处在研讨之中，还未得到公认的可代替 GPA 的成绩排名方法。这样，1998 B 题的优秀论文也为该问题的研究提供了有价值的探索。

11.1.3 竞赛结果

在 1998 年参加美国大学生数学建模竞赛的 472 个参赛队中，有 283 个队选择本题。结果有 3 个队获得特等奖 (Outstanding)，48 个队获得一等奖 (Meritorious)，69 个队获得二等奖 (Honorable Mention)。获得特等奖的 3 个队分别来自 Duke 大学 (North Carolina 州)，Harvey Mudd 学院 (California 州) 和 Stetson 大学 (Florida 州)。我国有 8 个队获得一等奖，分别来自清华大学 (2 个队)，中国科技大学，东南大学，华东理工大学，上海师范大学，西安电子科技大学，国防科技大学。

11.1.4 本章其余各节的安排

在 § 11.2 中，我们介绍对 GPA 进行修正的两个方法，取自 Duke 大学队的论文与 Stetson 大学队的论文。在 § 11.3 中，我们介绍能力分的概念与相应的最小二乘法，该方法取自 Duke 大学队的论文。在 § 11.4 中，我们再介绍其他几个有意义的处理方法。最后，我们在 § 11.5 中进行一些讨论。

§ 11.2 对 GPA 进行修正的两个方法

11.2.1 GPA 的公式

设某个学生在课程 k 的等级分 (Grade Point) 为 g_k ，而课程 k 的学分数为 c_k ，则该学生的 GPA (等级分平均值) 的公式为

$$GPA = \sum c_k g_k / \sum c_k,$$

其中二个求和均取该学生所选读的所有课程。

用 GPA 数值的大小来进行学生成绩排名至今还是最简便最常用的方法，但在分数贬值及前而 11.1.2 中所述的情况下，缺点是明显的。

11.2.2 标准化 GPA 方法

一种比较合理的替代处理是，它以 SGPA 来代替 GPA，公式是：

$$SGPA = \sum c_k \cdot (g_k - \mu) / \sigma / \sum c_k,$$

其中求和时取遍该学生所选修的课程， μ 为该课程的平均分， σ 为均方差，其表达式为

$$\mu = \sum g_k / m,$$

$$\sigma = (\sum (g_k - \mu)^2 / m)^{\frac{1}{2}},$$

这里， m 为选修该课程的学生数，求和时取遍所有该课程的学生。

该方法的优点是简单易算，每个课程可以被单个处理，缺点是未能体现课程的难易，且学生之间在一个课程的等级分差异可能被人为地拉大。

另外，也可以考虑用一个课程中所有学生的 GPA 的中位值来代替平均分，也可以考虑将等级分 g_k 的修正值 $(g_k - \mu) / \sigma$ 改变为其他形式的函数变换值。

11.2.3 调整 GPA 的方法

另一种对 GPA 进行修正的思路着眼于课程之间进行比较。假如有一个课程是特别困难的，那么选读该课程的学生所得的 GPA 会比他们选读其他课程所得的 GPA 为差，从而对此课程在处理过程中适当加分，是合理的。类似地，对于特别容易的课程，也可以适当地减分。这样的处理过程称为调整 GPA 方法。记课程 k 的 GPA 平均值为 μ_k ，调整量为 x_k ，那么，调整准则是

$$\mu_k + x_k = \sum_j AGPA_j / m_k \quad (k = 1, 2, \dots, n_c),$$

其中， m_k 是选读该课程的学生人数， n_c 为课程总数，AGPA 表示学生 j 的调整 GPA 的数值，求和是针对上述 m_k 个学生来做

的.

据 Duke 大学参赛队的论文所述, 该方法中的调整量 x_k 在 10 次迭代之后, 就能使上述等式近似地被满足.

模拟数据还表明, 该方法所得的调整 GPA (即 AGPA) 能很好地用于学生成绩排名. 同时, 该方法对于课程难易因素的考虑能为局外人所接受. 该方法的缺点是不能对各门课程分别进行计算, 而必须将所有课程放在一起进行计算, 从而这种计算不可能由每个学生自己校核. 当然, 从教学管理部门来说, 对于相当规模的问题, 这种计算也并不困难, 例如, 有 1000 个学生, 200 个课程, 每个学生 6 个课程, 也能很快完成相应的计算.

调整准则的含义是, 调整后的课程 GPA 等于调整后的选课学生 AGPA 的平均值, 注意 AGPA 应当线性地依赖于 $x = (x_1, x_2, \dots, x_n)$, 所以这个调整准则实际上是关于 x 的线性方程组. 我们认为, 这个线性代数方程组应当具有一些好的性质, 可以方便于求解. 有兴趣的读者可以对此作进一步的思考.

§ 11.3 能力分的最小二乘法

对学生的等级分进行重新处理的另一途径是引入能力分的概念, 它有二个基本假设:

(1) 设学生 i 的能力分为 x_i ($i=1, 2, \dots, n$). 设学生 i 与学生 j 选读同一课程 k , 那么他们的能力分的差异应当体现在他们取得的等级分 c_{ik} 与 c_{jk} 的差异上.

(2) 属于自然科学专业的学生在选读同类专业课程的成绩会比选读社会科学课程(专业转移)的成绩要好一些, 社会科学专业的学生亦有类似情况. 这个差异可用一个非负常数 δ 来体现, 它可称为专业转移的加分值, 其数值待定.

根据以上二个假设, 相应于每个分数对 (c_{ik}, c_{jk}) , 可得如下等式:

$$x_i - x_j = c_{ik} - c_{jk} + \lambda\delta,$$

其中 $\lambda = \pm 1$, 或 0; 当学生 i 与学生 j 选读课程 k 均是专业转移或均不是时, $\lambda = 0$; 当学生 i 选读课程 k 为专业转移而学生 j 不是时, $\lambda = 1$; 当学生 j 选读课程 k 为专业转移而学生 i 不是时, $\lambda = -1$.

设课程 k ($k=1, 2, \dots, m$) 的学生数为 m_k , 由该课程所确定的上述等式个数共有 $m_k(m_k-1)/2$. 所以所得方程的总数为

$$q = \sum_{k=1}^m m_k(m_k-1)/2.$$

看一个数值例子, 设有 2000 个学生选读 400 门课程, 每个学生选 6 门课, 平均地说, 每门课程有 30 个学生, 于是

$$n=2000, m=400, m_k=30,$$

$$q=400 \cdot (30 \cdot 29/2)=174000.$$

令 $\delta = x_{n+1}$, 上述 q 个方程可组成一个关于 $x = (x_1, x_2, \dots, x_{n+1})'$ 的方程组:

$$Ax=b,$$

其中 A 为 $q \times (n+1)$ 矩阵, b 为 q 维列向量. 由于总有 $q > n+1$, 上述方程组是超定的, 因此 x 应作为求解最小二乘问题而得到:

$$\min_x \|Ax-b\|^2 = \min_x (x' A' - b')(Ax-b),$$

它等价于求解

$$A'Ax=A'b,$$

其中 $A'A$ 为 $n+1$ 阶方阵, $A'b$ 为 $n+1$ 阶向量. 设方程组所相应的齐次方程组有一组非零解 $(1, 1, \dots, 1, 0)'$, 它表示 $x_{n+1} = \delta = 0$ 及其他所有 $x_i = 1$. 这意味着 $\det(A'A) = 0$, 也表明学生的能力分相差一个公共常数(且 δ 保持不变)时, 方程组保持被满足. 对此的简单处理是, 取定其中某个 x_i 的值, 如取 $x_1 = 3$ (相当于等级分 B).

上述方法称为能力分的最小二乘法. 根据 Duke 大学参赛队的论文的模拟计算结果, 最小二乘法能给出数千名学生成绩按能

力分的排名，相当合理。按此方法的观点，每个学生在一个课程的得分应当等于三部分之和，一是能力分，二是专业转移分，三是随机分，而最小二乘法有助于消除随机分，提取所需的能力分。该方法的缺点是计算工作量较大，据 Duke 大学队的论文，对数千名学生进行此项计算，一般需在微型计算机上计算数小时。

§ 11.4 成绩排名的其他方法

11.4.1 可否对等级分不作变换

以上两节所介绍的标准化 GPA 方法，调整 GPA 方法与能力分的最小二乘法有一个共同点是：对学生的课程等级分作一个变换，得到另一个较合理的更能体现学习成绩的“分数”。本节所介绍的方法不考虑对等级分作变换，其思想是直接利用各门课程中学生的排名次序，或者直接利用各个学生在课程学习中所得到的几个等级分，来进行学生之间的比较。

11.4.2 课程排名选拔法

举例来说，设有 600 名学生，各人选读 6 门课程，共有 100 门课程，这样，平均地说，每门课程有 36 名学生，现要求根据学生所得的课程等级分选拔 10% 的优秀学生，即 60 名学生。如果选拔的指导思想是“全面较好，大部领先”，那么按以下标准来进行选拔是适当的：

(1) 该学生在选读的每一门课程中所得的等级分应在平均分以上。

(2) 该学生在超过半数课程(如 4 门课程)中所得的等级分应位于前列，即从此课程成绩最好的学生算起，该学生的排名位数在比值 r (与全体选读此课程的学生人数之比) 之内，其中 r 可取 5%，6%，7%，…，15%，等等。

先取 $r=5\%$, 符合(1)与(2)的学生数一般会小于学生总人数的 10% (所述例子中为 60 名), 而取 $r=15\%$, 符合(1)与(2)的学生数一般会大于学生总人数的 10%. 为达到所需的被选拔学生数, 可以用对分法确定 r 的值.

上述课程排名选拔法能很好地体现合理的指导思想, 但缺点是, 未考虑课程的难度与专业转移等因素.

11.4.3 字典序排名法

该方法也不需要考虑分数的变换. 该方法的指导思想是, 在学生的成绩排名中优先考虑各门课程取得最高等级分的学生. 举例来说, 设每个学生选读 6 门课, 由这 6 门课的成绩可以得到一个 12 维的整数向量(称为该学生的等级分向量):

$$z = (z_1, z_2, \dots, z_{12})'$$

其中 z_1 表示该学生取得等级分 A₊ 的课程数, z_2 表示该学生取得等级分 A 的课程数, …, z_5 表示该学生取得等级分 B 的课程数, …, z_{12} 表示该学生取得等级分 F 的课程数, 也就是说, 所有 z_i 分别相应于所有可能的等级分:

$$A_-, A, A_+, B_-, B, B_+, C_+, C, C_-, D_+, D, F.$$

因此, 自然成立

$$z_1 + z_2 + \dots + z_{12} = 6.$$

设 2 个学生的等级分向量分别为 $z = (z_1, z_2, \dots, z_{12})'$ 与 $z' = (z'_1, z'_2, \dots, z'_{12})'$. 我们认为 $z \geq z'$ (z 优于 z'), 如果 $z = z'$, 或者存在某个 k , 使

$$z_1 = z'_1, \dots, z_{k-1} = z'_{k-1}, z_k > z'_k.$$

这样, 按上述“优先”关系可将所有学生按成绩排名, 其特点是, 取得高等级分次数越多的学生将被排在前面. 该方法的缺陷是可能会导致某些学生放弃个别课程, 因为个别课程成绩可以很差, 但不会影响该学生的排名. 弥补的方法是, 在选拔 10% 优秀学生时, 补充如同 11.4.2 中条件(1)那样的条件.

§ 11.5 若干讨论

11.5.1 不同处理方法的比较

按照 Duke 大学队的论文所述，模拟数据的计算结果表明，关于 § 11.2 与 § 11.3 中的四种对学生按成绩排名方法，从效果较好到较差，依次为：能力分的最小二乘法，调整 GPA 法，标准 GPA 法，GPA 法。实际上，这些方法连同 § 11.4 所介绍的两种方法各具特点，各自代表一定的价值取向与引导功能，也相应地具有它们各自的优点与缺点，下面对此作一简述。

能力分的最小二乘法能够突出地体现同课程不同学生之间的“比较”效果，且能鼓励学生跨专业选课。但是该方法也会有副作用，或许会促使一些学生避开一些能力强的学生较多选读的课程。

调整 GPA 法着眼于课程之间进行比较，通过加减的迭代调整，使课程 GPA 近似等于学生调整 GPA 平均值（仅对选读该课程的所有学生），从而使各课程的 GPA 具有协调性。它的缺点是，当一个课程中所给出的等级分相当密集时，该方法所采用的调整无助于改变此状况。

标准化 GPA 法有助于改变某些课程等级分密集的状况，但不能体现各门课程的 GPA 与选读课程的学生水平的协调性。

课程排名选拔法与字典序排名法的特点是，不必对等级分进行某种变换，可以说，对于所给的原始成绩单而言，该方法更加“原汁原味”，更加稳妥可靠，但如何处理“兼顾全面”与“强调按课程名列前茅”两者的关系，容易表现出人为的因素。

11.5.2 关于 ABC 学院院长的想法

该院长强调，要将各个课程中学生等级分的比较作为学生成绩排名的依据，通过上面介绍的一些处理方法及计算试验，应当说，这种想法是可行的，在 12 级计分制下，即计分制 (A_+ , A ,

A_+, A, A_-, \dots 下，要选拔10%的优秀学生，在多数情况下，是可以找到合理答案的。但在5级计分制下，即计分制(A, B, C, D, F)下，要选拔10%的优秀学生，则容易出现中间的模糊地带。大体上可以说，要选拔更多层次的优秀学生时，计分制的等级数应当更多一些。

对于“分数贬值”现象的处理方法，不仅在于作出“调整”去适应它，而且更在于运用适当的方法去引导改变它，使它失去存在的机理，特别地，要使它不会泛滥。从这个意义上说，上述多种方法在供决策者选用时，可以联系考虑这样一个问题：当前需要作什么样的引导，使得学生的课程评分能更加合理。在不同的情况下，也许适宜采取不同的方法。

顺便提及，这些方法在计算上都是稳定可靠的，不会因个别课程的等级分变化对许多学生产生影响。

11.5.3 数据集的产生

自然，进行计算试验所用的数据集要用随机的方法来产生。我们认为，按照行能力分的基本思想，将每个学生在课程所得的成绩看做为几项之和，是比较符合实际的。以 c_{ik} 表示学生*i*在课程*k*中的成绩，可以按下式来考虑：

$$c_{ik} = x_i + \lambda\delta + \theta_{ik},$$

其中 x_i 表示学生*i*的能力分， δ 表示专业转移加分（例如 $\delta=0.7$ ），当学生*i*的主修专业与课程*k*不一致时， $\lambda=-1$ ，否则， $\lambda=0$ ，此外， θ_{ik} 表示随机分。

以12等级分制(A_+, A, A_-, \dots)来考虑，还应限制 c_{ik} 的最大值为4.3，最小值为0。上述 c_{ik} 的公式中， x_i 应取为某种正态分布的随机变量，其取值可在区间[2, 4.3]之内，同时，对于每个确定的*i*， θ_{ik} 可取为某个小区间上的均匀分布随机变量。有关这些随机变量的参数，可通过合理性分析与计算试验加以确定。

第十二章 小行星撞击地球问题

叶其孝

北京理工大学 应用数学系

提 要

本章介绍了 1999 年美国大学生数学建模竞赛 (MCM-1999) 的竞赛情况、评阅和奖励，特别是介绍了 A 题的优秀论文、评阅人的评述和我们的评注。主要内容：MCM-1999 的评阅、结果和奖励；哈维·马德学院队的优秀论文；评阅人的评述；我们的注记(包括对可供参考的其他优秀论文的简要评注)。

§ 12.1 MCM-1999 的评阅、结果和奖励

本次竞赛共有包括美国、中国、香港等 9 个国家和地区的 223 所大学的 479 个队参加，其中中国有 40 所大学和 3 所中学的 155 个队参加。

各队的论文在 COMAP 的总部进行编号使得评阅人不知道论文作者的姓名和所属的学校。

A 题的初评是在康涅狄格州的南康涅狄格大学进行的，共有 7 位评阅人。B 题的初评是在蒙大拿州的卡罗尔 (Carroll) 学院进行的，共有 4 位评阅人。C 题的初评是在新罕布什尔州的新罕布什尔 (New Hampshire) 大学进行的，共有 6 位评阅人。每篇论文由两位初评评阅人评阅，摘要和论文的组织是论文评定的基础。如果两位评阅人的评分不同则进行协商，如果协商后还不一致，则再由第三位评阅人来评阅。

终评是在加州的哈维·马德(Harvey Mudd)学院进行的，A题评阅人有15位，B题评阅人有16位，C题评阅人有7位。

A题是由佐治亚(Georgia)州的佐治亚学院和佐治亚州立大学的Jack Robertson提供的，B题是由纽约城市大学约克(York)学院的Joe Malkevitch提供的，C题是由东华盛顿大学的Yves Nievergelt提供的。

评出的最后结果是：

	O	M	H	P	合计
MCM-1999A题获奖队数 (中国队数)	5(0)	34(9)	61(28)	112(16)	212(53)
MCM-1999B题获奖队数 (中国队数)	5(0)	39(8)	72(31)	91(38)	207(77)
MCM-1999C题获奖队数 (中国队数)	2(1)	9(6)	17(11)	32(7)	60(25)

其中，O=Outstanding=特等奖，M=Meritorious=一等奖，H=Honorable Mention=二等奖，P=Successful Participant=成功参赛奖。

A题共有五篇优秀论文(见[1]，[7]~[10])。

每个参赛队都将获得由竞赛主任和每题的评阅组长签名的证书。

美国运筹学和管理科学学会(ORSA)给予两个获得特等奖队的队员现金奖励和三年的会员资格。这三个队分别是华盛顿州的皮吉特海峡(Puget Sound)大学队(A题)，北卡罗来纳州的杜克(Duke)大学队(B题)和中国的浙江大学队(C题)。此外，美国运筹学和管理科学学会还给获一、二等奖的队的每个队员一年的免费会员资格。

美国工业与应用数学学会(SIAM)对每题指定一个特等奖队作为SIAM的获奖队，他们是来自加州的哈维·马德(Harvey Mudd)学院的两个队(A题和B题)，来自印第安纳州的厄勒姆

(Earlham)学院队(C题). 每个队员都将得到300美元的现金奖励, 他们的学校都将获得一个装在镀金镜框里亲笔签名的证书. 哈维·马德学院队将于2000年5月在佐治亚州的亚特兰大举行的SIAM年会特设的小型研讨会上报告他们的结果.

美国数学协会(MAA)对(B题和C题)分别指定一个特等奖队作为MAA的获奖队. 他们是来自阿拉斯加州的阿拉斯加大学费尔班克斯分校(University of Alaska Fairbanks)队(B题)和来自印第安纳州的厄勒姆(Earlham)学院队(C题). 两个队将在2000年8月在罗得岛州的普罗维登斯(Providence)举行的MAA数学节(Mathfest)的特设分组上报告他们的解法. 美国数学协会当选理事长Thomas Banchoff会授予每个队员证书.

§ 12.2 MCM-1999A题 强烈的碰撞

美国国家航空和航天局(NASA)从过去某个时间以来一直在考虑一颗大的小行星撞击地球会产生的后果.

作为这种努力的组成部分, 要求你们队来考虑这种撞击的后果, 假如该小行星撞击到了南极洲的话. 人们关心的是撞到南极洲比撞到地球的其他地方可能会有很不同的后果.

假设小行星的直径大约为1000米, 还假设它正好在南极与南极洲大陆相撞.

要求你们队对这样一颗小行星的撞击提供评估. 特别是, NASA希望有一个关于这种撞击下可能的人类人员伤亡的数量和所在地区的估计, 对南半球海洋的食物生产区域造成的破坏的估计, 以及由于南极洲极地冰岩的大量融化造成的可能的沿海岸地区的洪水的估计.

§ 12.3 哈维·马德(Harvey Mudd)学院 队的优秀论文——强烈的碰撞

12.3.1 摘要

我们考虑直径为 1000 米的小行星撞击地球。这种量级的撞击会造成严重后果，包括局部地区的地震和海啸，可能的全球气候变化以及由于射向大气的尘土造成的突变性的农业灾害。

可幸的是，在南极洲的撞击造成的后果远没有那么大的灾害性。通过对小行星可能的轨道的建模，我们确定撞击发生的角度相对说来较小，只造成较小、较浅的陨石坑。因为南极洲为很厚的冰盖所覆盖，很少量的尘土会射向大气中。撞击的热量会融化掉少量的冰。最坏的情况是，如果撞击造成的激波会引起海啸，所以我们预测了哪些将会被淹没。

12.3.2 初始假设

1. 小行星是球形的，直径为 1000 米，具有典型的组成成分和密度，并在南极和地球相撞。
2. 小行星源自我们的太阳系，在撞击之前绕太阳运行在轨道平面上，该平面和地球绕太阳运行的轨道平面是一样的 [Transcript—Plane of the Solar System 1996]。
3. 只有太阳、地球和月亮会对小行星的轨道产生重大影响。这四个行星的轨道可用 Newton 引力模型来预测。
4. 在南极附近，南极洲的冰盖的深度是均匀的，深度为 2 公里，粗略地讲冰盖的密度也是均匀的，冰盖的温度处处为 -76°C 。

12.3.3 小行星的性质

撞击的位置、角度和速度

我们考察撞击在南极或其他地方的相对概率。因此我们用

Newton 引力模型来模拟太阳、地球、月亮和小行星的运动，在 Newton 引力模型中，

$$F = \frac{Gm_1 m_2}{d^2}$$

描述了作用在质量为 m_1, m_2 相距为 d 的两个物体上的力。力的方向沿连接每两个物体质心的直线。万有引力常数 G 的值为 $6.67259 \times 10^{-20} \text{ km}^3 \text{ s}^{-2} \text{ kg}^{-1}$ ，引力按 $\vec{a} = \vec{F}/m$ 加速物体。加速度改变了物体的速度 \vec{v} ，从而影响到物体的位置 \vec{x} 。

我们采用时间离散化的数值模拟。物体 i 在时刻 $t + \Delta t$ 的位置 $\vec{x}_{i,t+\Delta t}$ 是利用了除物体 i 在时刻 t 的位置、速度和质量外的其他行星的位置和质量来算得的。特别是，我们对系统中的每个物体执行下列计算：

$$\begin{aligned}\vec{F} &= \sum_{j,j \neq i} \frac{Gm_i m_j}{|\vec{x}_{i,t} - \vec{x}_{j,t}|^2} \times \frac{\vec{x}_{i,t} - \vec{x}_{j,t}}{|\vec{x}_{i,t} - \vec{x}_{j,t}|}, \\ \vec{a}_{i,t+\Delta t} &= \frac{\vec{F}}{m_i}, \\ \vec{v}_{i,t+\Delta t} &= \vec{v}_{i,t} + \vec{a}_{i,t+\Delta t} \times \Delta t, \\ \vec{x}_{i,t+\Delta t} &= \vec{x}_{i,t} + \vec{v}_{i,t+\Delta t} \times \Delta t + \frac{1}{2} \vec{a}_{i,t+\Delta t} \times (\Delta t)^2.\end{aligned}$$

太阳、地球和月亮一开始具有表 12-1 中的特征 [Lide, 1992, 14~26, 14~27]。

表 12-1 太阳、地球和月亮的质量、半径、位置和速度

	Sun	Earth	Moon
m	$1.99 \times 10^{30} \text{ kg}$	$5.97 \times 10^{24} \text{ kg}$	$7.35 \times 10^{22} \text{ kg}$
r	$6.96 \times 10^6 \text{ km}$	$6.38 \times 10^3 \text{ km}$	$1.74 \times 10^3 \text{ km}$
\vec{x}	$(0,0,0) \text{ km}$	$\vec{x}_{\text{Sun}} + (1.50 \times 10^8, 0, 0) \text{ km}$	$\vec{x}_{\text{Earth}} + (0, 3.84 \times 10^5, 0) \text{ km}$
\vec{v}	$(0,0,0) \text{ km s}^{-1}$	$\vec{v}_{\text{Sun}} + (0, 29.8, 0) \text{ km s}^{-1}$	$\vec{v}_{\text{Earth}} + (-1.02, 0, 0) \text{ km s}^{-1}$

我们选择坐标系使得太阳位于坐标原点，而地球和月亮都在 x - y 平面上。由假设 1，小行星是直径为 1000 米的圆球。因此，

其体积为 $V_{\text{ast}} = \frac{4}{3}\pi(0.5\text{ km})^3$, 或者 0.524 km^3 . 典型的小行星的密度为 $\rho_{\text{ast}} = 2.5 \times 10^{12} \text{ kg km}^{-3}$ [Toon et al. 1997, 44]. 该小行星的质量为 $m_{\text{ast}} = 1.31 \times 10^{12} \text{ kg}$. 我们把在太阳系平面内飞向地球的小行星和从该平面外飞向地球的小行星加以区分. 从该平面外飞来的小行星会不会更可能击中南极呢? 为求得解答, 我们对这两种情形都进行了模拟.

我们把小行星放在距地球为 $1.54 \times 10^6 \text{ km}$ 任意位置处 (大约是月亮到地球的距离的 4 倍). 我们给予小行星和地球相对于太阳的速度同样的速度, 就好像小行星将会进入和地球的轨道重合似的. 我们通过对小行星加上一个速度为 10 km s^{-1} 方向指向离地心不超过 $9.57 \times 10^3 \text{ km}$ 的任意一点 (即, 小行星向以地心为中心, 以地球半径的 1.5 倍为半径的圆内的任意一点) 处, 开始来设定小行星和地球相撞的过程.

我们对 $\Delta t = 10 \text{ s}$ 进行模拟. 如果小行星和地球间的距离小于它们的半径之和, 那么碰撞就发生了. 我们从向量计算撞击处的纬度. 向量 $\vec{x}_{\text{ast}} - \vec{x}_{\text{Earth}}$ 和 $x-y$ 平面的夹角决定了撞击处的纬度.

$$\Delta \vec{x} = \vec{x}_{\text{ast}} - \vec{x}_{\text{Earth}},$$

$$\text{latitude} = \arctan\left(\frac{\Delta \vec{x} \times (0, 0, 1)}{\sqrt{(\Delta \vec{x} \times (1, 0, 0))^2 + (\Delta \vec{x} \times (0, 1, 0))^2}}\right).$$

类似地, 我们从 $\vec{x}_{\text{ast}} - \vec{x}_{\text{Earth}}$ 和 $\vec{v}_{\text{ast}} - \vec{v}_{\text{Earth}}$ 计算撞击角度和速度, 因为该速度向量和地球表面在撞击点处切平面的夹角就是撞击角, 速度向量的大小就是撞击速度.

$$\Delta \vec{x} = \vec{x}_{\text{ast}} - \vec{x}_{\text{Earth}},$$

$$\Delta \vec{v} = \vec{v}_{\text{ast}} - \vec{v}_{\text{Earth}},$$

$$\text{angle} = -\arcsin\left(\frac{\Delta \vec{x}}{|\Delta \vec{x}|} \times \frac{\Delta \vec{v}}{|\Delta \vec{v}|}\right),$$

$$\text{speed} = |\Delta \vec{v}|.$$

我们对小行星进行了 20000 次模拟, 一半从太阳系平面内飞

向地球，一半从该平面外飞向地球。在两种情形，约有略少于四分之一的次数小行星能避免和地球相撞。图 12-1 展示了击中地球的纬度分布。对于这两种飞向地球的方式而言，大约有 1% 的机会撞击在大于南纬 80° 的地方。

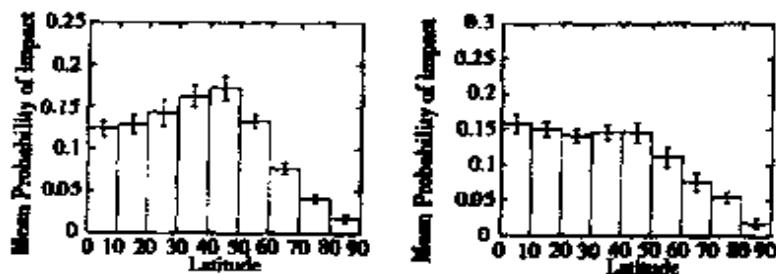


图 12-1 小行星从太阳系平面内(左图)和
外(右图)撞击地球的概率

但是在从太阳系平面内，在地球的高纬度地区撞击地球的比较可能的是较小的入射角： $18^\circ \pm 1^\circ$ ，而从太阳系平面外撞击地球的入射角为： $45^\circ \pm 5^\circ$ (见图 12-2)。

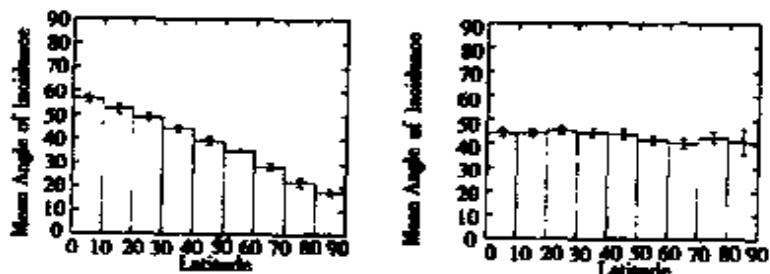


图 12-2 小行星从太阳系平面内(左图)和
外(右图)撞击地球的入射角

对于半径超过(译注：似应为“不超过”)100m 小行星，空气阻力可以忽略；这种小行星以其原来动能的大部分撞击地球 [Hills and Mader 1995]。我们的模拟表明以相对速度 15km/s 撞击地球与文献一致 [Chapman and Morrison 1994, 34]。

计算中没有考虑地球(轨道)相对于 $x-y$ 平面的倾斜(约为 22°)，因为我们没有碰撞年份的信息。因此，在南极撞击的概率不能通过简单的读一读图 12-1 的高度得到，也不能期望从图

12-2 直接读出撞击的入射角.

碰撞的动力学

利用算得的质量 $m_{\text{ast}} = 1.31 \times 10^{12} \text{ kg}$ 以及速度 15 km/s , 我们算出小行星到达地球大气时的动能为

$$E = \frac{mv^2}{2} = \frac{(1.31 \times 10^{12} \text{ kg})(15 \text{ km/s})^2}{2}$$
$$= 1.5 \times 10^{20} \text{ Joules.}$$

这相当于 3.5×10^4 百万吨的 TNT(如果是彗星而不是小行星撞击地球, 那么撞击速度为 50 km/s ; 但由于彗星的较小的密度 1 g/cm^3 (编译者注: 似应为 1 g/cm^3), 撞击能量还要稍小一点 [Toon et al. 1997, 44]).

覆盖在南极洲的冰盖使得预测陨石坑的大小成了问题. 因为, 如果是撞击在陆地上, Toon 给出了陨石坑直径(公里)的下列两个公式:

$$D = 0.64 \left(\frac{Y}{\rho_t} \right)^{1/3.4} \left(\frac{20000}{v_i} \right)^{0.1} (\cos\theta)^{0.5} \left(\frac{\rho_t}{\rho_i} \right)^{0.083},$$

$$D = 0.53 c_f \left(\frac{Y}{\rho_t} \right)^{1/3.4} (\cos\theta)^{2/3},$$

其中

Y 是以百万吨计的能量,

ρ_t 是目标物的密度,

ρ_i 是撞击物的密度,

v_i 是撞击物的速度,

c_f 是数值约为 1.37 的修正因子, 而

θ 是撞击角度 [1997, 44] (编译者注: 似应为 [Toon et al. 1997, 44]).

对于从太阳系平面内飞向地球的小行星来说, $\theta=18^\circ$ 这个值有点小了, 因为小行星可能会对该平面有某种扰动. 因此我们采用 $\theta=30^\circ$, 冰的密度为 0.9 g/cm^3 , 两个公式都给出了陨石坑的直径大约为 15 km .

因为“典型的”小行星产生的陨石坑的深度与直径之比约为 1 : 5 或者 1 : 7 [Terrestrial Impact Craters 1999]，因此直径为 15km 的陨石坑的深度约为 2.5 到 3km。但是，“典型的”小行星并不像我们模型中的小行星那样以较小的角度撞击地球，这将犁出一条大冰沟从而造成更宽但不那么深的陨石坑。如果，尽管小行星会降低它的向下的冲力，它会冲裂冰层，然而它会碰到更为密实的厚为 2 到 2.5km 的基岩。所以，我们并不认为陨石坑的深度会超过 2km。

因为冰比岩石容易融化，我们可能低估陨石坑的大小。南极周围的冰厚 2km，温度为 -76°C。直径为 15km 以及平均深度为 1km 的陨石坑只会搬走 175cm^3 的冰。然而，要融化掉这么多冰是不大可能的。碰撞也会把大量的冰水以及一些基岩抛向空中，但是比撞击在其他大陆地区所抛出的岩石要少。

12.3.4 对南极洲的影响

小行星撞击南极洲和撞击其他地方的后果很不相同。虽然南极洲远离大多数人口稠密的中心，但是冰盖融化的问题人们还是关心的。我们的计算采用了表 12-2 中的数据。

表 12-2 有关南极洲冰盖的数据

特征	大小	资料来源
体积	$30 \times 10^6 \text{ km}^3$	Virtul Antarctica [1999]
面积	$14 \times 10^6 \text{ km}^2$	World Factbook [1998]
平均厚度	2km	Computerworld Antarctica [1999]
平均温度	-76°C	Assumption 4

从理论上讲，足够大的撞击可以融化掉整个冰盖，从而把全球的海平面提高 70m [Computerworld Antarctica 1999]。暂时假定撞击的全部能量转化为热量。为计算撞击能融化的冰的体积，我们需要水的一些热力学性质(见表 12-3)。

表 12-3 水的热力学性质 [Lide 1992, 6~172, 6~174]

相态	传热性($\text{Wm}^{-1}\text{K}^{-1}$)	比热($\text{JK}^{-1}\text{kg}^{-1}$)	$k/c_p(\text{m}^2\text{s}^{-1})$
冰	1.88	2030	9.26×10^{-7}
水	0.61	4810	1.27×10^{-7}
汽	0.027	2020	1.34×10^{-8}

融化热焓 $3.33 \times 10^5 \text{ J kg}^{-1}$

汽化热焓 $2.26 \times 10^6 \text{ J kg}^{-1}$

我们假设撞击的全部能量 $1.5 \times 10^{20} \text{ J}$ 用于把质量为 M_{ice} 的冰温度提高到 0°C 的冰，并融化为 0°C 的水。于是我们有：

$$1.5 \times 10^{20} \text{ J} = (273.2 \text{ K} - 197.2 \text{ K}) \times 2030 \text{ JK}^{-1} \text{ kg}^{-1} \\ \times M_{\text{ice}} + 3.33 \times 10^5 \text{ J kg}^{-1} \times M_{\text{ice}}$$

求解后给出冰的质量为 $M_{\text{ice}} = 3.1 \times 10^{14} \text{ kg}$ 。利用冰的密度为 0.9 g/cm^3 ，那么撞击可以融化掉 1340 cm^3 的冰，稍多于南极洲冰盖总体积的 $1/100000$ ！如果所有的水都注入海洋，则将提高海平面不到 1 mm （毫米）。然而，南极离开最近的南极洲海岸的距离超过 500 km ，因此任何这样得到的水都不太可能到达海洋。对于南极洲冰盖的热量的更为精确的模型来说，我们假定全部能量 $1.5 \times 10^{20} \text{ J}$ 用来提高小行星下面的冰（我们用一个直径为 1 km 的圆柱体来近似表示）的温度。在直径为 1 km 的圆下面的冰的质量为 $1.5 \times 10^{12} \text{ kg}$ 。

- 为把冰加热 76°C （达到它的熔点）的能量应是质量乘上温度乘上冰的比热，这给出 $2.31 \times 10^{17} \text{ J}$ 。
- 冰的融化热焓为 $3.33 \times 10^5 \text{ J kg}^{-1}$ ，所以要融化这些冰的能量应是 $5.00 \times 10^{17} \text{ J}$ 。
- 要把这些冰再提高 100°C 应消耗的能量为 $7.2 \times 10^{17} \text{ J}$ 。
- 在沸点汽化水所需的能量为 $3.39 \times 10^{18} \text{ J}$ （汽化热焓是 $2.26 \times 10^6 \text{ J kg}^{-1}$ ）。

从初始能量减去这四部分能量值还剩下 $1.46 \times 10^{20} \text{ J}$ （编译

者注：似应为 1.45159×10^{20} J.) 如果我们假定所有剩余的能量都用来使水汽化，那么水温可提高到 48000°C.

全部能量不大可能都用于对冰加热，一些水汽在把它的热量传送到周围的冰之前已经逃逸到空中，所以这个模型对撞击的影响给出了过分的估计。

热传导方程

我们用热传导方程来对温度分布的传播进行建模。因为冰盖平均厚度只有 2km，但宽却有 6000km，所以，我们把冰盖建模为二维的冰片。令 $u(x, y, t)$ 表示该冰片在位置 (x, y) (以米为单位) 处和时间 t (以秒为单位) 处的温度 (°C)。设定坐标系使 $u(0, 0, 0)$ 是撞击时刻撞击中心处的温度。在热传导方程

$$\frac{\partial u}{\partial t} = \frac{k}{c\rho} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right)$$

中，常数 k 是物质的热传导系数。 c 是比热， ρ 是密度。但是当冰变成水，水变为汽(编译者注：即固、液、汽三个相态)时， k ， c ， ρ 的值会有变化，所以只有数值求解的方法。我们采用的方法是基于 [Burden and Fairs 1997] 给出的算法。令相邻迭代的时间步长为 Δt ，又令温度读出点间的距离为 Δx ， Δy 。为导出有限差分法，首先考虑 u 对 t ， x 和 y 的 Taylor 级数逼近：

$$\begin{aligned}\frac{\partial}{\partial t}(x, y, t) &= \\ \frac{u(x, y, t + \Delta t) - u(x, y, t)}{\Delta t} &- \frac{\Delta t}{2} \frac{\partial^2}{\partial t^2} u(x, y, t), \\ \frac{\partial^2}{\partial x^2} u(x, y, t) &= \\ \frac{u(x + \Delta x, y, t) - 2u(x, y, t) + u(x - \Delta x, y, t)}{(\Delta x)^2} &- \frac{(\Delta x)^2}{12} \frac{\partial^4}{\partial x^4} u(x, y, t), \\ \frac{\partial^2}{\partial y^2} u(x, y, t) &= \end{aligned}$$

$$\frac{u(x, y + \Delta y, t) - 2u(x, y, t) + u(x, y - \Delta y, t)}{(\Delta y)^2} \\ - \frac{(\Delta y)^2}{12} \frac{\partial^4}{\partial y^4} u(x, \psi, t),$$

对某些 $\tau \in (t, t + \Delta t)$, $\chi \in (x - \Delta x, x + \Delta x)$, 和 $\psi \in (y - \Delta y, y + \Delta y)$ 成立. 我们假定 $\Delta x = \Delta y$, 所以我们可以结合更多的项, 代入上述热传导方程, 并把误差项分离开来, 记为 $E(x, y, t)$, 就给出了下面的关系式:

$$\frac{u(x, y, t + \Delta t) - u(x, y, t)}{\Delta t} \\ = \frac{k}{c\rho(\Delta x)^2} \{ u(x + \Delta x, y, t) + u(x - \Delta x, y, t) + u(x, y + \Delta y, t) + u(x, y - \Delta y, t) - 4u(x, y, t) \}, \\ E(x, y, t) = \frac{\Delta t}{2} \frac{\partial^2}{\partial t^2} u(x, y, \tau) \\ - \frac{k}{c\rho} \frac{(\Delta x)^2}{12} \left(\frac{\partial^4}{\partial x^4} u(x, y, t) + \frac{\partial^4}{\partial y^4} u(x, \psi, t) \right).$$

常数可以集合为一项 K : $K = \frac{k}{c\rho} \frac{\Delta t}{(\Delta x)^2}$.

(编译者注: 这就是所谓的显式差分格式, 知道了 u 在 t 时刻各格点处的值就可算得 u 在 $t + \Delta t$ 时刻各格点处的值. 至于这种格式的收敛性、稳定性, 这里都没有涉及.) 求解 $u(x, y, t + \Delta t)$ 就给出能从给定的时刻 t 的温度分布解出 $t + \Delta t$ 时刻的温度分布:

$$u(x, y, t + \Delta t) = u(x, y, t)(1 - 4K) + K(u(x + \Delta x, y, t) \\ + u(x - \Delta x, y, t) + u(x, y + \Delta y, t) + u(x, y - \Delta y, t) - 4u(x, y, t)).$$

模拟的结果

我们用 C 语言编了一个程序来求解初始温度在直径为 1km 的圆外为 -76°C , 而在圆内为 48000°C 的该方程. 我们发现如果所有的热水汽都呆在原地(而不是我们原来预期的那样被抛到空中)的话, 那么最多能融化 $5.7 \times 10^7 \text{ m}^3$ 的冰(足以提高海平面 2

$\times 10^{-1}$ m)，而且融化这么多冰要花 10 天的时间！在达到这种状态前很久，人们就可以期望所有的东西都会冷却下来。即使能被融化掉的水也很难流入海洋，因为南极洲表面的很多地方被冰的可观的重量压到低于海平面了 [Virtual Antarctica 1999]。

如果人们考虑到冰和水的热传导性的量级差别，模拟的结果就不会太令人吃惊了。一开始的热量很快融化了大量的冰，但是当温度升高时温度的增长率很快慢下来，几乎达到小于 100°C 的平衡温度。这种过程提供了在假设时要把热传递出去的热气和周围只是很热的水和冰之间的一个隔热层，当这些冰融化时它并不传递太多的热量到下一层冰去。（编译者注：0°C 的冰要转化为 0°C 的水要吸收一定的热量，即潜热。在数学上，讨论这样的冰化水的问题就是所谓的两个相态（固相和液相）的相变问题，其精确的数学模型可归结为偏微分方程中所谓的自由边界问题。）

结论：我们要担心的最后一件事就是大量的冰融化的问题。

南极洲的生态系统

在南极洲周围的海域栖息着名为磷虾（krill）的小的甲壳类动物，它们对于该地区的食物链来说是重要的。水温的微小差别，或任何打乱自然界平衡的自然灾害都会影响到它们的群体数，这具有全局的影响。然而，几乎没有冰会融化而且也不会由于撞击而产生的其他的重大的长期影响，因而我们的最好估计是自然界会随着时间来修复自己。

12.3.5 全球规模的影响

我们需要担心的最严重的事情之一就是撞击是否会引起足够大的地震，从而可能引发海啸（潮汐波），如果引发靠近地面的地震，那么会更加严重。海啸的浪可高达 10m 到 30m，从其波前到波后可延伸达 4km 到 5km [Monastersky 1998b]，这种波通常以大约 800km/hr 的速度行进 [Monastersky 1998a]。（技术上讲，“海啸”一词指的是大的波，当它撞击海岸线前面的大陆架时

会慢下来，并增加其高度，不过我们在更广泛的意义下使用这个术语时，指的是任何具有能变成海啸的大的波。)

海啸特别难于预测，不是地震的量级而是其频率决定了波的高度；特别是，原因是推动波越来越高的关于时间的长期的振动 [Monastersky 1998a]。当波撞击到海岸线时甚至升得更高，并以一堵水墙淹没大地，造成物质的破坏。在 1992 年到 1997 年期间作为太平洋海啸的后果，有 200 人丧生；而于 1998 年 7 月发生在巴布亚新几内亚的海啸，据声称至少有 2500 多人丧生。丧生的主要原因是几乎没有要发生海啸的预警。

由于小行星撞击南极洲造成的海啸要用一个多小时才能达到南美洲的最南端，而且我们早就会事先知道这么大小的小行星的来到（在 10 年内，90% 的穿过地球轨道的这么大小的小行星应该被识别出来，而且可以画出它们的轨道 [Asteroid Comet Impact Hazards 1999]）。所以，通过疏散海岸线附近的人口，人类的伤亡率几乎可以完全避免。

海啸行向内陆的最大距离可以由当海啸击到海岸时的高度、海岸线处的水深、地形的高低不平性以及离开海岸线的海岸的斜率来决定。作为例子，相应于应该典型的发育地区的地形来说，一个 40-m 的海啸可行进到内陆 9km，而 10-m 的海啸可行进到内陆 100km [Hills and Mader 1995]。

海啸造成的全球规模的影响的精确模拟要用到复杂的流体动力学方程组，但是如果没有非常好的初始数据，这些模拟都将是没有意义的。作为替代，我们创造了一个更为简单的模型。假設由于小行星的撞击引起的激波能在南极洲大陆行进得足够快，使得它能在大约相同的时间到达海岸线的所有地方。那么，初始波前将在南极洲的形状向北行进。考虑一个表示地球表面的二维网格点。一开始时，给每个点标记上是水域还是陆地。表示水域的点给以两个变量：波的高度（标量）以及波的运动的方向（向量）。一开始，除了在南极洲边界所有点处的一个指向远离大陆

的波前外，所有标以水域的点处的高度为零，没有方向。每个时间步长，波沿运动方向传播，并与波的其他部分进行建设性或破坏性的相互干扰。除非有另一个波前作用，高出海平面的水将回流到海洋。



撞击后一小时，南极洲半岛的形状已经造成了以避免撞击到南美洲的东南海岸。



一小时后，海啸袭击南美洲。



几小时后，海啸淹没了非洲和澳大利亚的一部分。

这个模型是很有局限性的，但是在没有可能会引发破坏性的地震类型的更多的信息的情况下，人们能期望的大概也就这么多了。图 12-3[编译者注：这仅仅是计算机模拟输出的示意图]展示了



最后的图像，用黑边来展示所有的被淹没的海岸线。

图 12-3 海啸波前的计算机生成模型
了我们对海啸的计算机模拟在不同时刻的输出。

尘土和冰的荷载

对于住在西藏的人来说，海啸不会对他们造成什么影响，但是如果大量的尘土抛向大气，西藏人会发现天气会反常地变冷，西藏人赖以生存作为食物的植物将停止其光合作用从而死掉。这种全球性的影响可能会毁掉所有文明的遗迹。

我们可以从火山活动中看到尘土能做什么。1991 年 6 月 15 日菲律宾的 Pinatubo 火山爆发就大气中尘土微粒的水平而言，它是本世纪的任何事件的后果中最严重的后果 [McCormick et al., 1995, 399]。以当量为 10^{12} MT 的 TNT 的能量撞击大陆的小行星大约也会把这么多的尘土送入大气。若是当量为 10^{13} MT 的 TNT 的能量的话，产生的后果将类似于 1815 年 Tambora 火山的爆发 [Toon et al., 1992, 59]（编译者注：坦博拉火山是印度尼西亚松巴哇岛北岸休眠火山，原来海拔 13000 英尺，1815 年爆发时山顶削去很大部分，现海拔 9354 英尺。这次火山爆发使岛上居民数万人丧生）。Pinatubo 火山爆发使全球温度下降了约 0.5°C [McCormick et al., 1995] 而 Tambora 火山的爆发使全球温度下降了约 0.75°C [Tambora, Indonesia, 1815-1999]。

这给出了对于具有 3.4×10^{13} MT 的 TNT 能量的小行星预期能产生的后果一个粗略的等级估计：虽然以小角度撞击厚冰层比撞击其他大陆地区产生的尘土要少。

被抛射到空中的尘土没有足够的能量进入轨道，所以将散落在周围的大气中 [Toon et al. 1992, 57]. 我们认为由于占优势的风，造成的反方向吹的风带限制了尘土的扩散。被风吹起跨过能到达北半球的尘土将会几次大大地改变其高度。由 Pinatubo 火山爆发喷射出的尘土到达北半球和南半球要花 2 到 3 个月的时间。我们认为需要差不多的时间滞后（如果不是更长的话）才能把尘土输送到北半球。

尽管由于撞击而喷向空中的尘土会全球气候造成影响，但不大可能会严重破坏人类文明。它将对农业，特别是南半球的农业，产生温和的、暂时的影响。

除尘土外，大量的冰也将被喷射到空中。大部分空中的冰将变成雨或雪而降落回地球，还可能把空中的一些尘土一起带回地球。不断增加的水汽将降低上层大气层的温度，因为水是很强的红外线辐射体，造成更多的水气凝聚且凝成水或霜降落下来 [Toon et al. 1992, 68~69]。

以能量为 10^4 MT 的 TNT 撞击在海洋时将会喷射出两倍的水到上层大气层。这会产生较小的温室效应，由于冰雾挡住太阳多少能消除一点温室效应。这不会产生任何严重的影响，除非海洋温度改变的反应时间持续超过 10 年 [Toon et al. 1992, 69]。

12.3.6 结论

一颗直径为 1000m 的小行星会造成严重的全球性灾害。在靠近人口稠密地区的撞击会造成物质破坏和人员伤亡，而且撞击到海洋会造成极大的海啸。如果小行星撞击在大陆地区而不是撞击在南极的话，那么将会喷射出足够的尘土到空中，从而造成长期的环境危害。与任何一种这样的情景相比，撞击在南极洲就远没有那么严重的灾害。

对来自我们的太阳系的小行星而言，撞击角度大概是小的，比之于直接撞击它将造成一个更宽但较窄的陨石坑。因为南极洲

的冰盖厚为 2km，喷射到空中的大部分是冰的碎片而不是尘土。由于占优势的风流，喷出的尘土不会向北行进到人口更为稠密的地区。能进入大气的冰可能会造成温室效应，但仅当它们能留下许多年之后；可能大部分的冰将以雨的形式落回地面。

海啸的可能性是实实在在的，但无法预测。我们的模型预测了可能淹没的位置，但是这些模拟的完成需要有关更精确的关于世界的模型的更多的细节以及更复杂的海啸模型。由于预先的警报，在任何情况下沿海岸地区都可以疏散。严重的洪水可能危及几百万平方公里的食物生产区域，但这种影响只是短期的。

人们希望有足够的警告来疏散在南极洲做考察和研究工作的 4000 名科考人员。

12.3.7 参考文献

- Asteroid Comet Impact Hazards. 1999. <http://impact.arc.nasa.gov/index.html>.
- Burden, Richard L., and J. Douglas Faires. 1997. *Numerical Analysis*. New York: Brooks/Cole.
- Chapman, R. C., and D. Morrison. 1994. Impacts on the Earth by asteroids and comets: Assessing the hazard. *Nature* 367 (1994): 33~40.
- Computerworld Antarctica. 1999. <http://antarctica.computerworld.com/>.
- Hills, Jack G., and Charles L. Mader. 1995. Tsunami produced by the impacts of small asteroids. *Proceedings of the Planetary Defense Workshop*. Livermore, CA. Available at <http://www.llnl.gov/planetary>.
- Lide, David R., editor. 1992. *CRC Handbook of Chemistry and Physics*. 73rd ed. Boca Raton, FL: Chemical Rubber Company Press. Thermal properties of water: 6~172, 6~

174.

- McCormick, M. P., L. W. Thomason, and C. R. Trepte, 1995. Atmospheric effects of the Mt. Pinatubo eruption. *Nature* 373: 399~404.
- Monastersky, Richard. 1998a. How a middling quake made a giant tsunami. *Science News* 154 (1 August 1998): 69.
- , 1998b Waves of death; Why the New Guinea tsunami carries bad news for North America. *Science News* 154 (3 October 1998): 221~223.
- Montgomery, Carla W. 1995. *Environmental Geology*. 4th ed. Dubuque, IA: Wm. C. Brown Communications, Inc.
- Tambora, Indonesia, 1815. 1999. <http://volcano.und.nodak.edu/vwdocs/Gases/tambora.html>.
- Terrestrial Impact Craters.
<http://www.cpk.lv/hata/solarsys/solar/tercrate.htm>.
- Toon, O. B., R. P. Turco, and C. Covey. 1997. Environmental perturbations caused by the impacts of asteroids and comets. *Reviews of Geophysics* 35: 41~78.
- Transcript — Plane of the Solar System (October 20, 1996). 1996. <http://www.earthsky.com/1996/es961020.html>.
- Virtual Antarctica.
<http://www.terraquest.com/va/science/snow/snow.html>.
- World Factbook. 1998. <http://www.odci.gov/cia/publications/factbook/ay.html>.

§ 12.4 评阅人的评注^[2]

12.4.1 作者介绍

Patrick J. Driscoll 是美国西点军校数学科学系的教授。他也是 MCM-1999A 题的评阅组长。他在斯坦福大学获运筹学和工程经济系统硕士，在弗吉尼亚理工学院获工业与系统工程博士。目前，他是西点军校数学选修课主任。他的研究集中于线性与非线性最优化中的重建线性化技术 (reformulation—linearization techniques)。

12.4.2 评阅文章主要内容

很长时间以来，参赛者第一次读到 MCM 的问题 A 时，在感受到数学分析的挑战的同时总是呈现出神话般的联想。然而，经过队员们进一步的讨论和考察，这个问题通常总是不可抵挡地要用到相当直接的数学结合着创造性的思考。小行星撞击地球的问题继续着这种趋势，即，为参赛者提供机会来全力对付一个复杂而又具挑战性的问题。尽管参赛者可以从图书馆或互联网上得到大量的参考资料，但清楚地识别小行星撞击地球的短期和长期的影响的任务仍然留给参赛者大量的要探索的想法。

在过去几年里，专业不同的大学生参赛者给参赛队带来了处理竞赛问题的多方面的专业能力。这通常导致许多有趣的混合杂交建模的方法。但从今年的情况来看，不管各队队员们的专业结构有多少不同，解决的方法似乎有点收敛到少数几种方法。我们的推测是这种后果是由于经由互联网的网络联系的急剧增加。

正如许多队在周末三天的努力中所发现的，利用互联网作为支持他们的分析的信息源是一把双刃剑。在诸如桑迪亚 (Sandia) 实验室和喷射推进实验室的网址上就有讨论小行星撞击地球的有趣的，而且在有些情形下是精确而相关的信息。不幸的是，许多

论文显示，凡是首先没有对本问题进行思考和讨论就从这些网址上提取信息的队很快就发现他们自己陷入麻烦，被哄骗进了使他们在竞赛规定的时间内不能成功完成的数学方法之中。而且就他们缺乏直接的支持文件和推理来看，他们最终发现他们自己对清楚地解释和充分地理解这些网址所提供的基本假设和数学推理方面并未作好准备。就如在大多数建模努力中见到的，这导致了他们的论文中的致命的缺陷。还值得提出的第二个评注是有关在竞赛截止日期前完成任务应采取的策略的问题。大多数好的论文表现出了该队的如下的策略：当面临着就是否对问题的一部分做复杂的数学分析，或是试图对问题作一个完整的（数学）建模努力作出决策时，选择后者。优秀论文包含了全面回答试题中提出的问题的有见地的见解，如何平衡这两个方面，随不同的论文而不同。但是有一点是清楚的，各队要在完成建模的努力下选择一二种数学方法（例如，偏微分方程、动能建模等）。

总的说来，优秀论文提供了他们队在寻求支持信息前曾花了大量的时间来思考该问题的明确的证据。这样的选择看来能使他们在识别确切的建模参数与作出合理的假设相比，要付出的代价和会得到的好处之间作出权衡，并能用近似的、在一定范围内的参数值来进行建模。许多事实和本问题相关——诸如，小行星和南极洲的地质成分，飞向地球的小行星的典型的原始资料，撞击的角度，人类的人口分布，以及大气气流和环流——都将置于采用这种策略的考虑之下。那些未能（无论是明显地或不明显地）提供曾考虑过的重要问题的特征的论文将不予进一步的评阅。至少，应能识别和解释特殊特征（例如上部大气风流）的影响，然后出于数学易处理性的理由，明白地选择不把这些因素包括在内的话，就会更好一点。

建模的假设有两大范畴：需要通过讨论来验证的物理假设，以及可以从所列出的文献得到的数值参数的假设。两类假设的似是而非性和可应用性，都直接依赖于各队能更好地把一个特定的

假设与 MCM 所述的问题联系起来，而不是参考资料文献中所述的某个问题。不论论文的计算如何，地球海洋的 10 米瞬时提升与所提问题的结果，即使是对虔诚的科幻小说的追随者也能接受的结果来说差得太远了。

和过去的竞赛一样，不能过分强调在报告的正文需要有确切的支持文件。优秀论文都在他们的正文中传递了可验证的、可靠的信息来源的清晰的联系。差一点的论文虽展示了支持信息有赖于互联网网址，但没有包括关于为什么某些参数值是合适的以及他们的方法是建立在什么样的假设上的清楚的说明。尽管直接从互联网上的资料剪贴的诱惑是很强的，但这样做多数是关于未得到支持的“事实”的陈述的文章，而不是展示一个队对模型有清晰了解的文章。此外，花很多时间去推导已知的关系（例如，Kepler 的引力吸引定律）对论文几乎不会加分。

最后，较好的论文提供了完全的摘要，只有少数的语法错误，甚至没有语法错误，并提供设计得很好的表和图，以图示说明他们队的基本的解析推理。

§ 12.5 评注

1766 年普鲁士天文学家、物理学家和生物学家提丢斯 (Johann Daniel Titius, 1729, 1, 2 ~ 1796, 12, 11) 发现了一个表示行星到太阳距离的定则（见〔3〕），此定则于 1772 年为德国天文学家波得 (Johann Elert Bode, 1747, 1, 19 ~ 1826, 11, 23)（见〔4〕）所证实，后合称为提丢斯-波得定则。1801 年元旦，意大利天文学家皮亚齐 (Giuseppe Piazzi, 1746, 7, 16 ~ 1826, 7, 22)（见〔5〕）根据提丢斯-波得定则在距太阳的距离为 2.8 天文单位处果真发现了第一颗小行星谷神星。到 1940 年，具有永久性编号的小行星已达 1564 颗（见〔6〕）。尽管太阳系中第一颗小行星的发现只有 200 年的历史，但有关小行星，特别是某

些小行星是否会撞击地球的问题长期以来一直是人们关心、讨论和研究的问题。因此，要在图书馆或互联网上查找有关文献资料是很容易的。从这次发表的优秀论文来看也确实如此。

正如我们在第一章中论述的大学生数学建模竞赛培养学生的能力时指出过的应变能力的培养，即在看到试题并确定要做的题目后能迅速查阅到相关的资料，并能在较短的时间内适度的消化和应用。但正如评阅人的评注中指出的，首先要对题目有充分的理解和讨论，有自己的见解，才不会被文献牵着鼻子走，更不会被“引入歧途”。因为要真正研究、解决小行星撞击地球的问题决不是短短的三天能解决的。只有根据自己的洞察，紧紧围绕主要特征进行数学建模，并能应用适当的数学方法和计算机模拟技巧，加上清晰的表述，才能在竞赛中取得较好的成绩。

优秀论文〔7〕～〔10〕也是很值得一阅的。

参 考 文 献

- (1) Dominic Mazzoni, Deep Impact, UMAP, v. 21 (2000), no. 3, 211～224.
- (2) Patrick J. Driscoll, Judge's Commentary, The Outstanding Asteroid Impact Papers, UMAP, v. 21(2000), no. 3, 269～271.
- (3) 简明不列颠百科全书, 卷 7, 中国大百科全书出版社, 1986, 732.
- (4) 简明不列颠百科全书, 卷 1, 中国大百科全书出版社, 1985, 762.
- (5) 简明不列颠百科全书, 卷 6, 中国大百科全书出版社, 1986, 483.
- (6) 简明不列颠百科全书, 卷 8, 中国大百科全书出版社, 1986, 590～591.

- (7) Micheal Rust, *Asteroid Impact at the South Pole, A Model-Based Risk Assessment*, UMAP, v. 21 (2000), no. 3, 225~240.
- (8) Nicholas R. Baeth, *Antarctic Asteroid Effects*, UMAP, v. 21 (2000), no. 3, 241~252.
- (9) Mikhail Shpitser, *Not an Armageddon*, UMAP, v. 21 (2000), no. 3, 253~261.
- (10) Daniel Forrest, Garrett Aufderberg, Murray Johnson, *The Sky is Falling!*, UMAP, v. 21 (2000), no. 3, 263~268.

第十三章 公众场所的法定容量

姜启源

清华大学 数学科学系

提 要

“公众场所的法定容量”是1999年美国大学生数学建模竞赛的B题，207个队选作了此题，5个队获特等奖，39个队获一等奖，72个队获二等奖。获特等奖的队中，Duke University的一个队被评为INFORs (Institute for Operations Research and Management Science) 的优胜者；Harvey Mudd College一个队被评为SIAM (Society for Industrial and Applied Mathematics) 的优胜者；University of Alaska Fairbanks的一个队被评为MAA (Mathematical Association America) 的优胜者。以上3个美国的学会(协会)都是这项竞赛的协办者和支持者。UMAP Journal 1999年第3期刊登了5篇特等奖论文和评阅人的文章，这里对它们作较详细的介绍。

§ 13.1 是题目，§ 13.2，§ 13.3，§ 13.4 分别是上述3篇优胜者的论文，§ 13.5 是评阅人和专家对竞赛论文的评论。

§ 13.1 公众场所的法定容量

在公众聚集场所的房间，你会看到有“违法”字样的标记，指的是这个房间的人数超过了规定的限额，估计这个限额是根据在紧急情况下人们从房间的出口疏散的速度制定的。类似地，电梯和其他设施也常有“最大容量”的告示。

试为如何确定这种“法定容量”建立一个数学模型，模型中要讨论确定“违法”占用房间（或空间）人数的准则（除了在火警或其他紧急情况下所考虑的公众安全之外），还要考虑如果你的对象是带有可移动家具的房间（如有桌椅的食堂），还是体育馆、公共游泳池、有排椅和通道的演讲厅时，有什么区别。你可以比较在不同环境（如电梯、演讲厅、游泳池、食堂或体育馆）下人们可以作什么样的模型。像滚石乐队音乐会、英式足球赛这样的场合，也可以作为特殊情况。

将模型应用于你们学校（或邻近城镇）的一个或多个公众场所，把你的结果与实际标出的容量（如果有的话）相比较。如果你的模型能用，你会受到想增加容量的用户的诘问，用你的分析给地方报纸写一篇答辩性的文章。

§ 13.2 速率模型与房间分解

这是 Duke University 一个队的论文。

本文对一个门的房间建立了 3 个模型：一个假设人的流动速率为常数；另一个将速率用房间人员密度的线性函数界定；第三个则用人员密度的负二次函数界定。在每种情形人员撤离的速率都大致正比于所有门的组合流动速率，对常数和二次模型，计算机模拟给出了与校园中食堂相一致的结果。一旦知道了 N 个人疏散需多长时间，就能反过来确定在 T 时间所撤离的最大人数，而 T 可以根据实际情况确定。

两个重要的影响容量的因素是：

- 紧急问题：在人员不受伤害条件下，用极小化撤离时间确定最大容量；
- 舒适问题：在房间变得过热或 CO_2 显著超标之前，对于给定的空隙，确定房间内可以容纳多少人。

对紧急问题给出两个模型，每一个都给出一定人员撤离建筑

物所需的时间，反过来，确定在一定时间内撤离的最大人数。

对舒适问题给出在一定时间内可以舒适地占用给定空间的最大人数。

1. 一般假设

- 人员指体重在 100~300 磅的成年人；
- 撤离时没有警卫员或负责人，每个人都希望尽快离开建筑物，并用同样的方法决定最好的路线；
- 一个人从一个房间到另一个房间的时间，与所有人撤离房间的时间相比，可以忽略；
- 一个人需要 1m^2 (9ft^2) 站立或舒适地走动，特殊地方（如电梯）可以是 0.75m^2 。

2. 常数速率模型

补充如下假设：

- 人员以常数速率通过各个门；
- 一个人到门边的时间与他离开房间的时间相比可以忽略；
- 在撤离中所有门都畅通，人员总是拥向最近的门或最快撤离的门；
- 人群聚集在每个门边，在房间几乎撤空之前，有足够的站在门边。

1) 一个门的单房间

由假设，总有足够的人充分利用门，若门允许人以速率 r 通过，房门内有 n 个人，则需 $t = n/r$ 使房间变空。

2) 多个门的单房间

由假设，在房间变空前所有门都是拥挤的。设有 k 个门，通过速率为 r_1, \dots, r_k ，人数为 n_1, \dots, n_k ，则所有队列同时走出，即 $t = \frac{n_1}{r_1} = \dots = \frac{n_k}{r_k}$ ，则 $n = \sum n_i = t \sum_{i=1}^k r_i$ 。定义

$$r = \frac{n}{t} = \sum r_i, \quad (1)$$

即多个门的房间等价于有一个大门的房间，其速率 γ 为各个 γ_i 之和。

3) 子房间和通道分解

考虑家具和其他障碍物。首先设想一个有大量桌椅的餐厅（图 13-1），尽管妨碍人们走向通道，但前面的假设仍成立，于是（1）成立。

另外，障碍可将一个房间分成子房间和通道。如有若干

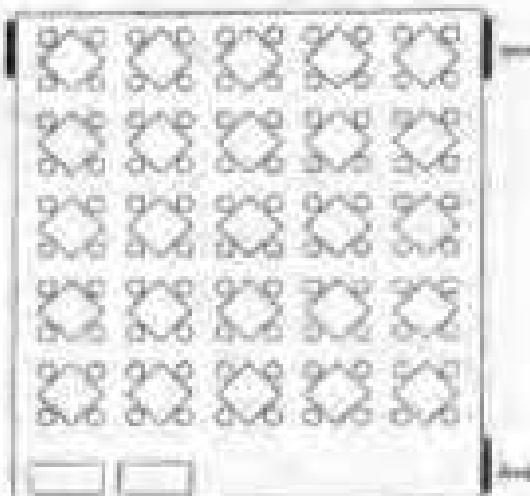


图 13-1 餐厅(俯视图)

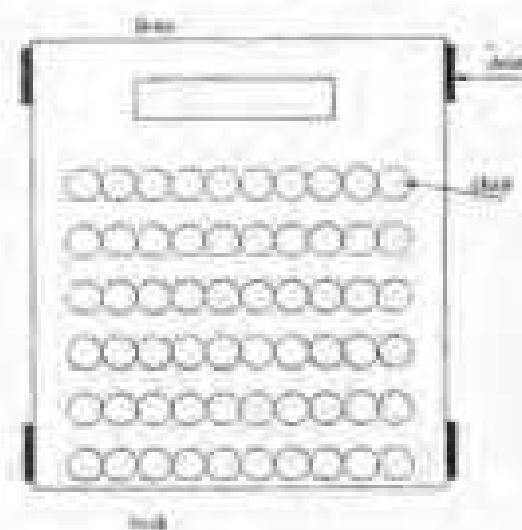


图 13-2 演讲厅(俯视图)

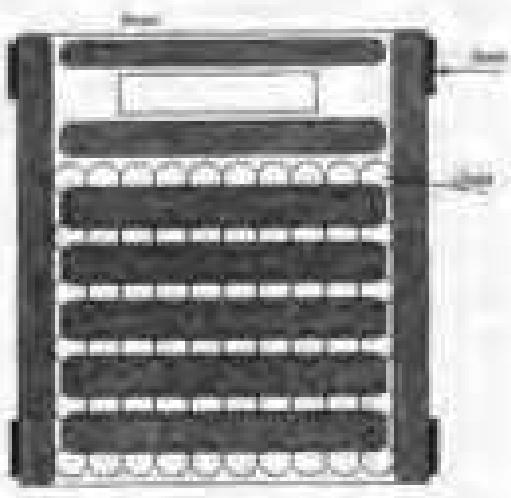


图 13-3 演讲厅的通道

排座位的小演讲厅（图 13-2），可以作图 13-3 样式的分解。这与餐厅不同，因为它严重限制了人员的流动方向，人员要从两端撤离，会堵塞。一旦分解完毕，就变成复合房间的撤离（图 13-4），出口处于最大容量下，撤离时间由组合速率确定。

更复杂的如自助食堂（图 13-5），这是一个真实的校园中的建筑，大多数房间由开着的通道（视为通过速率很大的门）相连（图 13-6），撤离时间由 4 个外门的速率确定。但如果有一个由单

个小门连到穿堂的大房门，及一个连接穿堂到外面的大门，则撤离时间要依于人们进入穿堂的速率，即小门有时是瓶颈，有时不是。

4) 极大流模型

复合房间的撤离问题可以忽略房间内的人数。假设离开与进入复合房间的速率相同，用 Ford-Fulkerson 算法求图的最大流。设一个有向图，每个弧有已知的最大流量，一个结点为源，另一个为沟，赋予每弧以实际的流量，若有从源到沟的通路，则每个弧的流量可增加，这种赋值可改进。若没有这种通路，赋值就是极大的。Ford-Fulkerson 算法寻找所有可能的通路，直到不能改进。

n 个人离开的时间由 n 除以最大流来估计，用 Ford-Fulkerson 算法需构造两个结点，源(结点)以无限容量连通到各个房间；沟(结点，指外部)与复合体的出口连通，其容量等于外门的容量。

图 13-7 是食堂的图，标记的(极大)流已不能再改进，因为所有连到沟的弧都是极大流，所以撤离速率由外门的流动速率决定，即无内部瓶颈。

3. 二次速率模型

1) 线性速率模型

人员撤离的速率 $f(t)$ 由房间内人数的线性函数界定，撤离

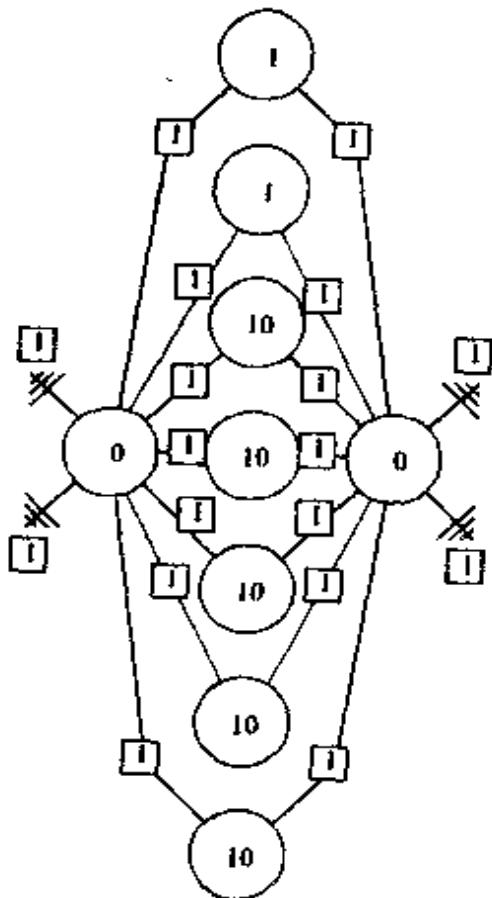


图 13-4 演讲厅的子房间和通道分解简图。圆圈表示子房间，直线表示通道，接地符号表示通向外面的门。每个子房间标注了里面的人数，每条通道标注了每秒可以通过的人数

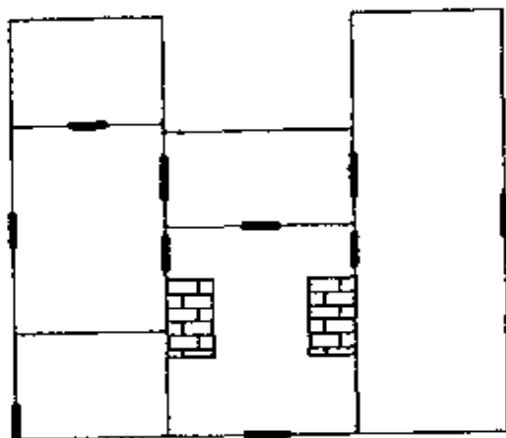


图 13-5 自助食堂(俯视图)

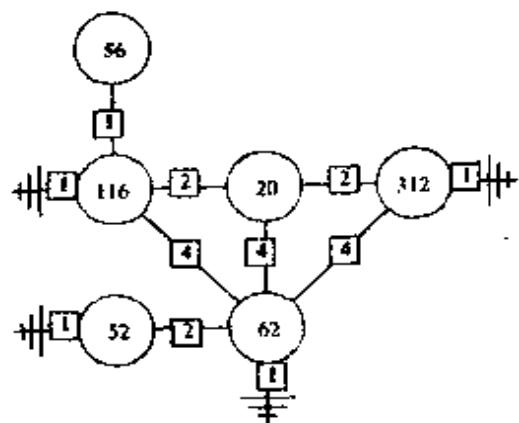


图 13-6 自助食堂分解简图

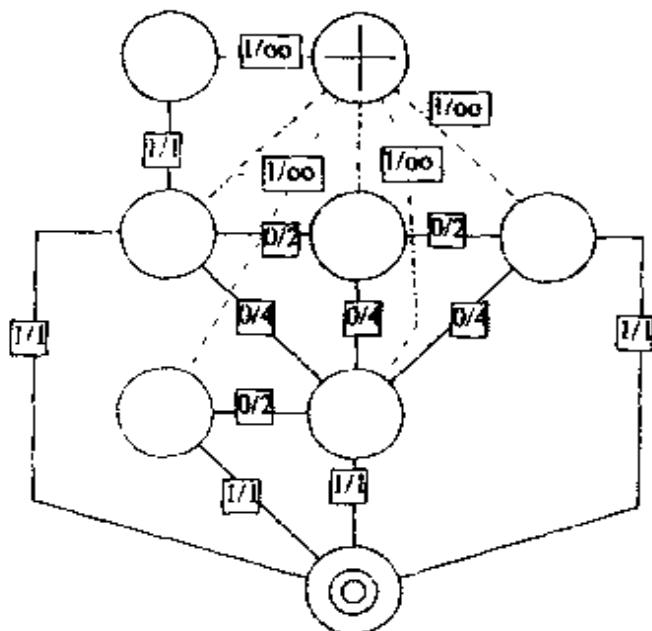


图 13-7 Ford-Fulkerson 算法用于食堂的图.
+号表示源, 牛眼表示沟.

问题表示为

$$\begin{aligned} & \max \int_0^T f(s) ds \\ \text{s. t. } & 0 \leq f(t) \leq a \int_t^T f(s) ds + b, \quad \forall 0 \leq t \leq T \end{aligned} \quad (2)$$

这里目标函数是给定时间 T 内的撤离人数, 约束条件中的积分分为 t 时刻留在房内的人数. 模型表示房内人数对通过出口速率的影响是线性的, a 反映强度, b 表示影响可忽略时的速率. 模型

未考虑当流量足够大时，表示速率上界的容量函数应减少到零。

2) 负二次模型的建立

将(2)式中人员撤离速率 $f(t)$ 的上界改为 t 时刻房内人数的二次函数，有

$$\max \int_0^T f(s) ds \quad (3)$$

$$\text{s. t. } 0 < f(t) < q - r \left(\int_t^T f(s) ds - p \right)^2, \quad \forall 0 < t < T$$

当房间内人数为最佳容量 p 时， $f(t)$ 达到最大值 q 。

负二次模型的两个假设为

- 当房中人数显著小于 p 时， $f(t)$ 的上界将减小，因为与撤离时间相比，人们走到门的时间不能再忽略；

- 当房中人数显著大于 p 时，相挤、碰撞等使 $f(t)$ 减小。

p 依于空间面积 A 和临界密度 d （当单位面积人数超过 d 时，流量减小）， $p = Ad$ 。这里设 $d = 0.75 \text{ 人}/\text{ft}^2 \doteq 7 \text{ 人}/\text{m}^2$ 。

为用(3)解撤离问题，设最大流出现，于是(3)化为

$$f(t) = q - r \left(\int_0^t f ds - \int_0^t f ds - p \right)^2, \quad \forall 0 < t < T$$

微分可得

$$\begin{cases} f'' f - f'^2 + 2rf^3 = 0, \\ f(T) = q - rp^2, f'(T) = 0. \end{cases}$$

解出

$$f(t) = \frac{c}{\cos^2(\sqrt{rc}(T-t))}, \quad c = q - rp^2 > 0. \quad (4)$$

由此，时间 T 内由房间撤离的最大人数 N 为

$$N(T) = \int_0^T f(t) dt = \sqrt{\frac{c}{r}} \operatorname{tg} \sqrt{rc} T, \quad c = q - rp^2. \quad (5)$$

反过来得

$$T(N) = \frac{1}{\sqrt{rc}} \operatorname{tg}^{-1} N \sqrt{\frac{r}{c}}. \quad (6)$$

3) 负二次模型的有关问题

紧急情况下，某些人会因拥挤、倒下造成伤害。二次模型合理地假定，当超过临界密度时，会使撤离变慢。

为解释模型的推测，考虑面积 $A = 1000\text{ft}^2$ 的房间，最大速率 $q = 90$ 人/分，最优容量为 $p = Ad = 1000 \times 0.75 = 750$ 人，设有 $T = 6$ 分钟时间撤离。取^①

$$r = \frac{a}{p^2} = \frac{0.01}{750^2} = 1.8 \times 10^{-8}.$$

由(5)得到 $N(6) = 540$ ，而用线性模型得 $N(6) = 557$ ，似乎相差不大，是由于 p 在最优值附近。

若令 $p = 10000$ ，则 $N(6) = 501$ ； $p = 100000$ ，则 $N(6) = 195$ ，即拥挤使撤离效果变差^②。

4) 负二次模型的局限

负二次模型是为从一块区域而非整个建筑物的撤离设计的，当用于校园的食堂时它与常数速率模型的结果相同。将用负二次模型、线性模型、常数模型模拟各种紧急情况，比较结果。

模拟运行时要计算一个人在给定时间内离开屋子的概率。当房内人数太少时，二次模型得到 0 或负概率，所以当人数少于 10 时要用线性模型。

我们估计了 p 和 d ，由于模型结果严重地依于这些参数，所以要精确估计它们。

4. 通风模型

对最大容量，另一个要考虑的因素是舒适水平，如

- 室温应在 $65\sim90^\circ\text{F}$ ($18\sim32^\circ\text{C}$)，通风系统应能散去人员产生的热量；
- 空气中毒素应保持在无伤害水平，如 CO_2 应在 0.1% 以下

① 作者未给出 $a=0.01$ 的理由。

② 笔者无法得到这两个结果。

(8%致死)；

- 若允许吸烟，要增加通风装置。

根据资料，人体产生热量的速率为 60W(睡眠)到 600W(剧烈运动)，适度运动为 100W。散热依于绝缘、窗、空调等。用了几小时的房间应能散热 100W/人，使室温大体不变。每人每秒钟至少要 0.21 新鲜空气，来冲淡 CO₂ 浓度(若可吸烟，需 25l)。

氧在空气的比例可以降到 13%，密闭空间中人体自然产生的 CO₂ 是关键因素。人体正常呼吸一次约 500cc，其中 4.1% 为 CO₂，持续 4 秒钟，于是人体以 5×10^{-3} mol/s 产生 CO₂。

给定房间容积 V，空气克分子量 N 满足 $PV = NRT$ 。记 r 为产生毒素的常数速率 (mol/s)，q 为空气中毒素比例，则 $qN = rt$ ，于是时间

$$t = \frac{qV}{r} \cdot \frac{P}{RT}, \quad (7)$$

在室温、1 个大气压下， $\frac{P}{RT} = 41.4 \text{ mol/m}^3$ ，q 代以致死浓度，可得到 t。

例如一电梯 3m × 3m × 3m，12 人，因拥挤而空气不足，人占了一半空间。由(7)，CO₂ 达到 8% 需 2.5 小时，因此，可以用救援队打开电梯的时间来限制容量。当然，电梯一般通风良好，CO₂ 不会积累起来。

5. 两个特例

1) 游泳池

室外池的撤离不成问题；室内池可类比开着的房门，人们可以从各边撤离。

在池中活动空间比陆地要大，每人约 3m² 空间，可在各方向 1m 范围内运动，在跳板、扶梯附近要有更大的空间，如 4m 的圆。

2) 电梯

电梯的门宽，且人少，紧急情况下的撤离时间可以忽略。

已经考虑了通风问题，更重要的因素是载重和空间，制造商提供了载重限制，而每人 0.5m^2 即能提供充分的个人空间。

6. 优缺点和附注

- 模型相当强健(Robust)，二次模型更实际、更精确地模拟了紧急情况，但对大房间，给出的结果有问题。
- 对二次模型，可以将确定 ρ 值的分析推广到外部因素如何减慢人群的撤离上。还可建立一种方法量测临界密度 d ，如观察多少人在不同时间内可以撤离，用这个数据估计最大流出现时的临界值。
- 对舒适问题，可以更好地估计多长时间使房间过热。

7. 附录——计算机模拟

检验复合房间的撤离：不同类型的门(开；时开时闭；不同速率)；不同选择通路的方式。每个门旁有一队人在等待，在每一(时间)步所有门使一定数量的人进入下一个房间，每人有机会根据他们的感受加入另一队。外部为一特殊房间，当指定人数到达时停止模拟。

§ 13.3 房间的六边形分割

这是 Harvey Mudd College 一个队的论文。

本文给出一个建筑物中人群的移动模型，模型把屋子分割成六边形，利用最近邻调和平均的等待时间函数，确定所有人撤离的时间，将它与基于建筑物大小的目标时间作比较。模型的参数可以对几种建筑物进行修正，并考虑各种特殊情况，给出最大容量。

1. 模型假设

1) 当几个人去抢占一个空出的位置时，某人占有的概率与他在当前位置等待多长时间无关。虽然等待时间长的人似乎更有

利，但这种影响可以被队伍向出口移动的趋势所补偿。

2) 要撤离的人的移动总会减小总的撤离时间，因为人们总是选择最快的撤离通道。

3) 人们可以很快地将行走速度提高到至少 6ft/s (正常行走速度是 4ft/s)。

4) 当人们撤离时挤在一起形成一个直径 1.4ft 的小区域。

5) 可移动的家具不会堵塞出口，虽然它们可在出口附近影响撤离。

建模中必须考虑的因素：

- 在拥挤过程中离开的人数，与按有序的队通过门能离开的最大人数不一定相同，为得到离开需要的时间，不能简单地将总人数除以给定时间内通过门的人数。

- 个人的移动只由人的位置和对出口的利用所决定。

2. 定义

将房间划分为标准的六边形，边长 0.75ft，这是人群挤在一起想离开时一个人所处的区域。假定这个矩形房间只在北墙有一个门。

以下的定义一般均对一个六边形(为方便起见，用 H 表示六边形)而言。

- H 的近邻～与 H 有公共边的 6 个 H (在边界上可少于 6 个)。

- 允许移动～从一个 H 向任一近邻的移动。

- H 的向径 R ～从 H 移出门外的最小允许移动次数。

- R 的等值线～ R 相同的 H 集合。

- 等同～两个 H 的 R 相同。

- H 的好近邻～ R 最小的一个近邻。

- H 好近邻数～好近邻的个数。

- H 的希望近邻～或者是好近邻；或者是 R 相同且好近邻数更多的近邻。若只允许向 R 较小的 H 移动，则妨碍了向 R 相

同但前景更好的 H 移动，给每个 H 一个 R ，及好近邻数，就使这种移动成为可能。

- H 的固有等待时间～通过这个 H 的时间，这与它的近邻是空、是满无关。

- H 的实际等待时间～给定固有等待时间、它的近邻的等待时间和竞争态势，花在这个 H 上的预期时间。

- H 的等效等待时间～实际等待时间乘以竞争这个 H 的人数。

- H 的预期离开时间～形成从这个 H 到出口最小通路的各个 H 的等待时间之和。

- (时间)步长～人员离开房间的基本时间单位，即一个人离开一个靠近门的 H 的时间，于是当门有 3 个 H 宽，且每秒能使 6 人通过时，步长为 0.5 秒。

3. 模型构造

根据预计通过每一个六边形所需的时间，赋予它一个固有等待时间。在无障碍下可设为 0.25 秒，即人以速度 6ft/s 移动 1.5ft (H 的宽度)。桌子等障碍物可视为六边形，且等待时间很大。

图 13-8 是一座假想剧院内一些六边形的固有等待时间。图 13-9 表示某些六边形的 R 等值线和好近邻数。

在给定固有等待时间之后，可以确定(时间)步长为，门的宽度(以六边形个数计)除以离开速率(出门人数/秒)。

与门相邻的六边形($R=1$)的实际等待时间为 1(步长)，是门边某人预期的实际等待时间。 $R=2$ 的六边形中，有些只与一个 $R=1$ 的相邻，有的则与两个 $R=1$ 的相邻。显然后者较优。一般规则为：

- R 不同的两个六边形， R 小者较优；
- R 相同者，好近邻多者较优。

考察所有最优的六边形，确定其实际等待时间，然后计算次

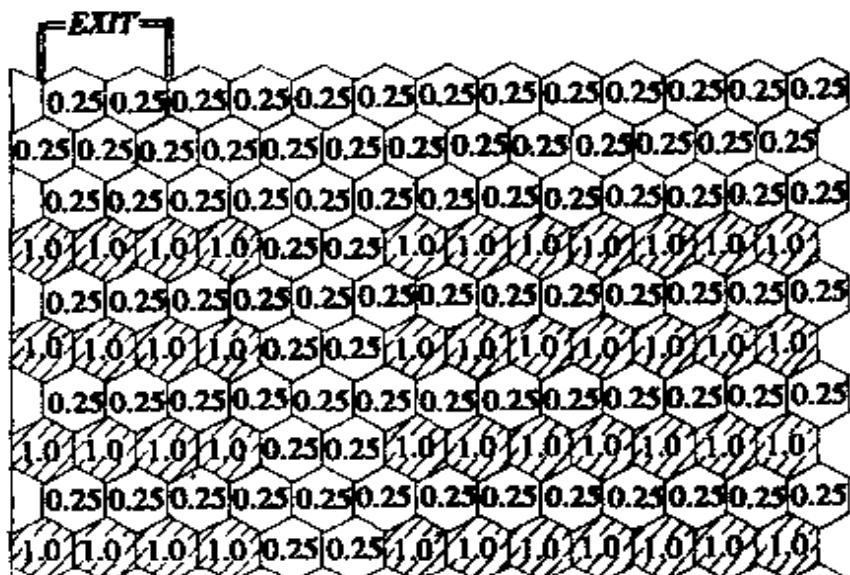


图 13-8 假想刷院内一些六边形的固有等待时间. 标注 0.25(秒)的六边形是自由空间, 标注 1.0 的六边形是固定座位

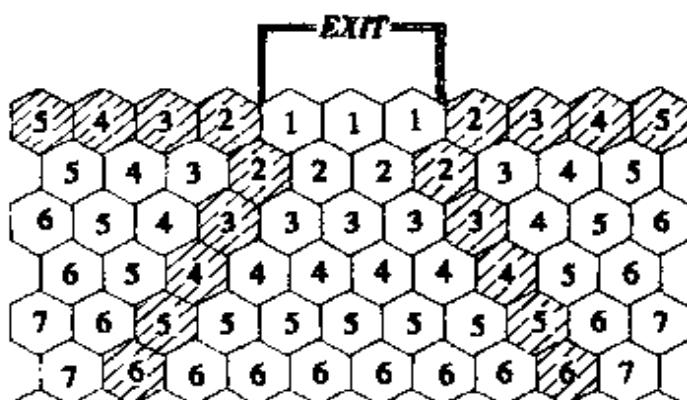


图 13-9 刷院内一些六边形的 R 值. 带阴影的只有一个好近邻, 其他的或者靠近门, 或者有两个好近邻
优的实际等待时间.

因为实际等待时间只依赖于它的希望近邻的实际等待时间、这些近邻的竞争态势和它自己的固有等待时间, 所以我们不必知道比希望近邻更差的那些六边形的实际等待时间.

4. 实际等待时间的计算

对每一希望近邻计算等效等待时间: 实际等待时间乘以与它竞争的六边形数. 若 3 个人都试图得到一个六边形, 则他们的等效等待时间是它实际等待时间的 3 倍, 即竞争会减慢人的前进

——每人都想竞争这 $1/3$ 的机会。然而很多六边形有不只一个希望近邻，有机会就要去占有某一个，我们用约化调和平均（调和平均除以希望近邻数）处理多个希望近邻的影响，即（约化调和平均记作 RHM——reduced harmonic mean）

$$RHM(A, B) = \frac{AB}{A+B} = \frac{1}{\frac{1}{A} + \frac{1}{B}},$$

$$RHM(A, B, C) = \frac{ABC}{AB+BC+CA} = \frac{1}{\frac{1}{A} + \frac{1}{B} + \frac{1}{C}}.$$

对应用调和平均可作如下解释：

- 希望近邻的实际等待时间好比电阻，有两个或两个以上希望近邻时，像电阻并联一样，与 RHM 公式相同。
- 关心的是停留在给定六边形中的时间，若一个六边形每 A 步打开一次，另一个每 B 步打开一次，那么在 AB 步中第一个打开 B 次，第二个打开 A 次，于是每步平均打开 $AB/A+B$ 次，与 RHM 公式相同。

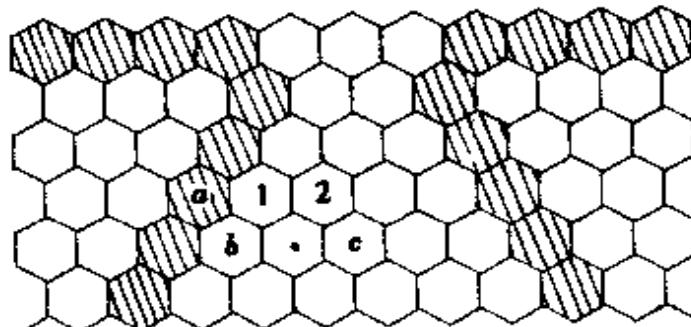


图 13-10 矩形房间，门有 3 个六边形宽，计算标有 (*) 的六边形的等待时间。先找出它的好近邻（标有 (1), (2) 的六边形），和 R 相同但好近邻多的六边形（图中没有）。对于 (*) 要竞争的两个六边形 (1), (2)，确定竞争者的总数：(1) 有 3 个 ((a), (b), (*)), (2) 有 2 个 ((*), (c))，于是 (1), (2) 的等待时间要乘以竞争者的数目，而它们的等效等待时间的调和平均要除以 (*) 竞争的六边形总数。这个约化调和平均加上固有等待时间，就是 (*) 的实际等待时间

各个希望近邻的等效等待时间的约化调和平均计算后，加上

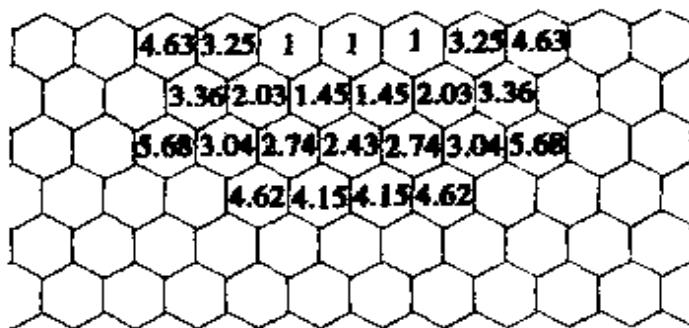


图 13-11 图 13-10 中六边形的等待时间

固有等待时间就是实际等待时间，如图 13-10，图 13-11.

5. 目标时间

确定一建筑物的最大容量，要有一个目标时间，我们决定根据实际标出的容量（对很少障碍物的简单建筑），用拟合方法得到目标时间 $T = 0.4A^{3/4}$ ，其中 A 是建筑物面积(ft^2)。

6. 模型检验

考察一个较空旷的建筑物(如体育馆)，只有一个门，若模型准确，模型给出的时间会比占有数与最大数(给定时间内离开的)之比稍高，因为人们会相互竞争位置。

有一篮球场、一排球场的体育馆 $80\text{ft} \times 110\text{ft}$ ，划分为 84×53 个六边形，在长边中央有一个出口，6 个六边形宽，每秒每个六边形出去 2 人，故最大速率是 12 人/s。

设有 875 人在馆内(最大容量 1350)，很快形成聚集，用简单的方法求撤离时间。考虑离门最远为 $\sqrt{(110/2)^2 + 80^2} = 97.1\text{ft}$ ；距门的向径为 R 的六边形的数目是 R^2 的函数，系数为 1.5 (六边形对边距)；最远的一个 R 约为 $\sqrt{p/\pi} = \sqrt{p/3}$ ， p 为初始人数，于是 $p(t) = 875 - 12t$ (12 人/s 为最大速率)。这样，六边形长度按 $(1.5 + 1.3)/2 = 1.4$ 计，最远的人走到队最后，形成排队的时间为(人的行走速度为 6ft/s)

$$T = \frac{97.1 - 1.4\sqrt{p(t)/3}}{6}, \quad p(T) = 875 - 12T,$$

代入得 $T = 10.9$ s, $p(T) = 875 - 131 = 744$, 即 10.9 秒后出口处形成 744 人的队伍.

在分割并计算各个六边形的实际等待时间后, 找出每个六边形的最短路, 计算最短离开时间, 直到第 744 个, 得到约 221 秒, 再加上上面得到的 11 秒, 共 232 秒, 这个数字约为简单地用人数除以速率(即 $875/12=73$ 秒)的 3 倍, 更合乎实际.

7. 模型结果

带有各种设备房间的固有等待时间如表 13-1. 模型的最终结果, 即各种建筑物在不同撤离时间下的最大容量如表 13-2, 表 13-2 还提供了两个备用的撤离时间和人数. 由表 13-2 可知, 模型可以比较各种设施不同排列情况下的结果.

表 13-1 各种设施的固有等待时间

设 施	时 间 (秒)
自由空间	0.25
剧院座椅	1
娱乐厅中的曲径	3
桌子	0.7
写字台	1.2
隔板	20
排水沟	20

表 13-2 模型的最终结果

房间结构	面 积 (ft^2)	时 间 (s)	最 大 容 量	时 间 人 数	时 间 人 数
剧院, 15 排座位, 门在一角	900	63	98	30 54	120 153
剧院, 15 排座位, 门在后面中间	900	63	98	30 54	120 174
歌舞厅	375	34	83	10 27	60 125
电梯	60	6	21	3 10	10 26
大教室, 4 排长条桌(从一头到另一头)	750	57	99	3 57	120 187
大教室, 7 个写字台一行, 7 行	750	57	96	30 55	120 186
小教室, 3 排长条桌(从一头到另一头)	300	25	51	15 27	45 75
小教室, 4 个写字台一行, 5 行	300	25	44	15 27	45 67
浴室, 5 块隔板, 4 条排水沟	200	21	25	10 15	30 33
娱乐厅	1200	82	32		

8. 优缺点及改进

- 用一般方式和很少参数对各种类型的建筑物建模，可讨论如门的位置，家具放置的改变引起撤离时间的变化。
- 它给出了很简单的极大安全容量，可对任何容量给出撤离时间。
- 只需用微机，运行时间对划分区域的面积是多项式时间的，我们计算运行时间不超过 1 分钟。
- 模型只限于一个门，只针对矩形建筑（这不是模型本身的限制），未考虑屋子的高度，而对某些紧急情形，房子的容积可能比面积更重要。
- 模型的最大局限也许是某些建筑用现有的规范确定目标撤离时间函数，如果最简单的建筑不能有一个精确的最大容量，那么这个目标时间函数必定要改变。
- 经进一步的计算，允许任意多个门，如游泳池的一个边是一个门，一个六边形可以有几个不同的 R （每个门一个），用最小的一个 R 决定去哪个门。
- 模型可以用于门在某时刻被关闭的情况。

§ 13.4 图-流模型和粒子模拟模型

这是 University of Alaska Fairbanks 一个队的论文。

本文给出两个模型研究最大容量和撤离时间。一个是基于图的网络流模型，将人看做可压缩的流体，流向出口，用连续过程研究人们的相互作用；另一个是离散单元（粒子）模型，将人看做圆盘，人的相互作用来自以个人为单元的假设。在比较、估计模型的输出后，分析房间的容量。

1. 图-流模型

模型用代表一区域（如房间）的图表示，图由结点集 N 和有向弧集 E 组成。每个结点赋以一小块区域的人数，弧标出从一点到另一点的方向。人离开结点的能力受该点拥挤程度和通向另

一结点的弧的带宽限制。带宽指人在结点间移动的速率，结点间没有障碍物和门时带宽较大，出口限制人流时带宽较小。

进入一点的人数受离开该点的人数和该点紧缩的趋势(流入率)限制。

因为结点间存在着相依关系，在每一(离散时间)点从出口以串级形式计算，一个结点的流出率计算出以后，就可以计算它的流入率，由此确定整个图的流率。

1) 模型假设及其缺点

- 所有人都知道紧急情况发生，并试图退出。实际上不是每人都知道并愿意退出。
- 人们只向一个出口移动。实际上人们会观察，并奔向较少拥挤的出口，这个假设排除了瓶颈的存在。
- 人们到达出口结点就认为是安全的，并退出模拟。这忽略了出口的排放能力，实际上离开出口的人数影响总的退出时间，这个假设限制模型只对单个房门。
- 聚集人群的移动速度由聚集密度决定；移动受人们要通过区域的宽度限制。这个假设来自文献，它是描述交通集散地的行人移动而非撤离房间的情形。
- 人们都将尽快移动到出口，不管人群密度的影响。这未考虑人的智力或可能出现的管理人员。
- 人群密度的增加是有限的。意思是在很短时间内结点不大可能从空置达到最大容量。
- 人视为连续流，可以有分数。这虽违反现实，但可不严格地认为，人可以部分地跨越两个结点的边界，这是模型用小的时间步长时所需要的。

2) 模型的数学结构

赋以下记号：

N_i ~ 结点(房间的一个小区域)；

$E_i \sim N_i$ 可以到达的所有结点集合；

I_i ~ 可以到达 N_j 的所有结点集合;
 P_i ~ N_i 的人数(人);
 A_i ~ N_i 的面积(ft^2);
 W_{ij} ~ 从 N_i 到 N_j 的流率(人/ ft)(两结点之间为 0.541; 到出口结点为 0.325);
 S_{\min} ~ 最大拥挤下的最小移动速度(2.5 ft/s);
 T ~ 人群的最大密度(3 人/ ft^2);
 r_s ~ 流入率常数(4.333 ft/s);
 S_e, S_g ~ 两个移动常数($S_e = 58.678, S_g = 58.669$).

流率——容量方程

记 S_i 为拥挤状态下 N_i 内的移动速度(ft/s), 有

$$S_i = \max[S_e + S_g \ln \frac{A_i}{P_i}, S_{\min}], \quad \textcircled{1}$$

FR_i 为流入率(t 时间内可以进入 N_i 的最大人数), 有

$$FR_i = r_s t \frac{A_i T - p_i}{A_i T}, \quad (A_i T \text{ 是 } N_i \text{ 的最大容量}) \quad \text{(2)}$$

OF_i 为流出量(t 时间内从 N_i 到 N_j 的人数), 有

$$OF_i = t S_i W_{ij}, \quad \text{(3)}$$

IF_i 为最大流入量(t 时间内从各方向进入 N_i 的人数), 有

$$IF_i = \sum_{j \in E_i} FFA_{ij} + FR_i, \quad \text{(4)}$$

其中 FFA_{ij} 是从 N_i 到 N_j 的实际人数, 有

$$FFA_{ij} = \begin{cases} FF_{ij}, & p_i \geq \sum_{k \in E_i} FF_{ik}; \\ p_i \frac{FF_{ij}}{\sum_{k \in E_i} FF_{ik}}, & \text{其他.} \end{cases} \quad \text{(5)}$$

式中 FF_{ij} 是在 N_j 能接受的情况下, 从 N_i 到 N_j 的人数, 有

① 作者未给出此式的来源, 笔者亦无法理解.

$$FF_{ij} = \begin{cases} OF_{ij}, & IF_j \geq \sum_{k \in I_j} OF_{kj} \text{ 或 } N_j \text{ 是出口;} \\ OF_{ij} \frac{OF_{ij}}{\sum_{k \in I_j} OF_{kj}}, & IF_j < \sum_{k \in I_j} OF_{kj}. \end{cases}$$
(6)

IF_j 是 FFA_{ij} 的函数，而 FFA_{ij} 又与从 N_i 流入的那些 N_j 有关，由于这种相依性，图必须是非循环的，如果图有回路，就无法计算 IF_j 。

(6) 式表示，当从各方向进入 N_j 的流量不超过 IF_j 时， OF_{ij} 可被 N_j 接受，否则按比例分配。

(5) 式表示，当从 N_i 流向 E_j 的人数超过 P_j 时， N_i 到 N_j 的实际人数要乘以比例因子，否则就等于 FF_{ij} 。

2. 粒子模拟模型

将人员视为离散的、相互独立的粒子，紧急情形下室内人员向一个选定的出口移动，直到全部出去。

1) 假设及其缺点(仅列出与上述模型不同处)

- 根据拥挤程度(靠近出口的人数)、距离和可见程度选择出口。实际上人们经常随大流拥向不能看到的出口。
- 步速为 4ft/s. 这未考虑小孩、老人等。
- 人可以瞬间地改变移动方向和速度。虽有违于基本物理规律，但简化了相互作用。
- 若在拟定的路线上遇到其他人，要停下来改变方向。其实人们有预定的行走计划，会避免遇到他人而停下来。
- 人不能通过墙或家具，为此将人视为硬的圆盘。
- 计划的通道会绕过家具，通向出口。这忽略了在惊慌中直接跳过家具的人。

2) 一个例子

400 人在 110ft × 120ft 的体育馆内均匀分布，接着很快在门

附近形成人群。随着门边的人向外的移动，人群逐渐撤出。

若一人预定的路线与他人相交，则停止，并试验其他（随机的）方向。为确定这种相互作用，我们规定，如碰到人，绕过他，向右。这样可能会出现循环等待情形，正如计算机中的死锁。我们利用伪随机数发生器决定移动方向。

3. 模型检验

利用粒子模拟模型和图-流模型估计从试验房间的撤离时间，房间大小为 15ft × 15ft，左面墙中央有一个 3ft 宽的出口。

两个模型重复运行，得到图 13-12，两个模型对试验房间的结果都接近线性，而直线的斜率有别。因为两个模型所用的参数有不确定性（如图-流模型的带宽，粒子模拟模型的人的活动范围），这种差别是合理的。虽然每个模型从一组独立的假设和数据导出，但模型的趋势有强相关性。

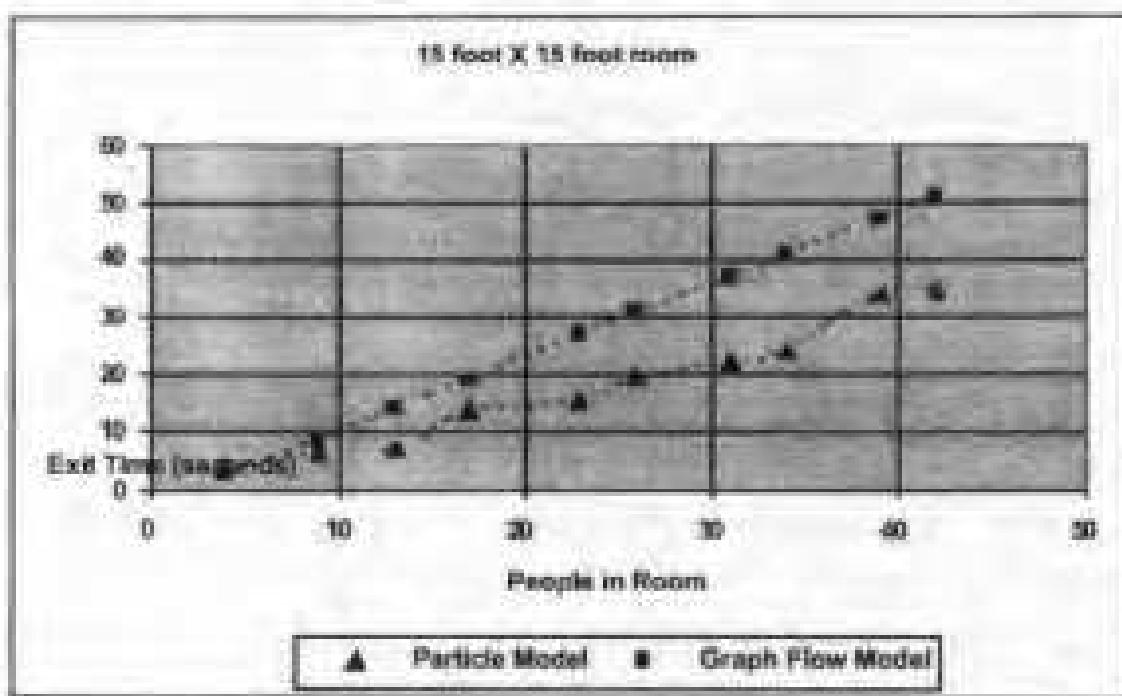


图 13-12 两个模型的模拟结果

4. 优缺点

图-流模型的缺点在于，它不是将人视为可分的，而是流体。

其结果依于(图的)带宽和源结点选择的任意性. 其优点在于, 人的行为是确定的, 而许多数学结构是由实际研究导出的.

粒子模拟模型的缺点是, 结果与人活动范围的大小有关, 决策非确定性, 对输入的微小变化敏感, 模型也受限于通道及非人为的行为. 其优点是把人视为单一的不可分的实体, 可以不依于他的近邻而移动, 对整个房间的总体流动不需作假设.

5. 结论

给出了两个模型来确定多长时间可以从一个房间撤离. 虽然它们的方法和假设不同, 两个模型都与我们的检验相吻合.

根据检验和分析, 撤离时间对房间人数近似于线性关系, 直线的斜率与房间的配置、出口的大小和数量有关.

原以为当更多的人挤向出口时, 撤离速率会降低, 但实际上对两个模型它是常数. 我们认为这是由于, 即使只有不多的人拥向出口, 出口也会很快变得拥挤, 这与人们的经验相符.

我们模拟的建筑物的最大容量是合乎要求的, 在最大容量下3分钟内撤离是可以接受的.

为确定一个房间的最大容量, 建议首先向火警站长询问可以接受的最长撤离时间, 然后用模拟器找出在此时间内可以撤出的最大人数. 而由于二者之间的关系近似于线性, 这样做是容易的.

§ 13.5 评阅人和专家的忠告

本节介绍一位评阅人, University of South Carolina 的 Jerry R. Griggs 教授的文章, 和一位专家, Richard Hewitt 博士的文章, 它们对于参加数学建模竞赛, 以及用数学工具真正解决实际问题, 都是大有帮助的.

1. 不足、成绩与希望

作为评阅人, Griggs 教授通过比较参赛队的论文, 指出以

下一些不足之处：

- 用过于简单的方法确定房间的容量，如设每个出口每秒可出去 r 人，有 n 个出口，安全撤离的时间为 s 秒，则容量为 rns 人。

- 忽略题目要求的项目，如考虑不同的房间布置及环境条件；将模型与已有的标记或规范比较；讨论安全以外的准则；为报刊写文章。

对于一些好的文章的做法，评阅人指出了以下几点：

- 模型包含了各种影响因素，如房间内（不只是出口）人员的流动，由房间形状、家具形成的瓶颈造成的拥挤，以及人群的初始分布等。

- 考虑到聚集在门旁的人群可以转向路程较远、但不太拥挤的出口。

- 参考已有的规范，对聚集的人群进行观察，收集数据。
- 给出撤离时间作为容量函数的图形，使决策者容易使用。
- 对模型运行时间复杂性进行分析，并作简化计算。
- 考虑更多因素如惊慌、通风、救援人员作用的可行性。

评阅人对参赛队提出的希望有：

- 尽力给出问题要求的所有主要项目，少作一些项目会被淘汰。

- 一个周到的、资料充分的摘要是重要的，文章虽强而摘要很弱也会在早期淘汰。摘要中不要只是重复问题，应指出如何建模，结果如何。摘要不要过分技术化。

- 建立人们能用的、容易被接受的模型。一般我们不喜欢难以捉摸的、过多的变量和方程。好例子会提高文章的可读性。给出的算法应让人充分理解，而不少文章只是给出计算机编程，没有对“为什么”这样做，及“如何”这样做给出解释。

- 对文章起支持作用的信息是重要的，如图形、表格等。还应完整列出参考文献，告诉人们，你的想法来自何方。

2. 你的回答还不是问题的解决

Richard Hewitt 以专家的身份发表了一篇题为 “The Answer Is Not the Solution” 的评注文章。他有 3 个学位：运筹学博士、数学硕士、经济学硕士，作过许多实际问题：油气开采、公共安全、通讯等。

在向所有参赛队表示祝贺、鼓励之后，他首先对一些文章作的假设提出意见，如：

- “人们在紧急状态中从最近的门撤出”。实际上，按照一位火警站长的说法，“火警中人们大多从他进来的那个门跑出，而不是最近的门，他们通常只记得来的路，按这条路走，并不能按照理智和法则行动”。这样，问题就复杂化了，需要了解人们如何进入建筑物，才能确定最大容量。

- “人们平均以每秒 x 英尺退出”。Richard Hewitt 有进出失火建筑物的亲身经历，他认为能见度是关键因素，在烟雾中出口都难以找到，何谈速度。一些人的所谓经验通常来自火警演习，在演习中人们知道门在何处，但知道与做是两回事。

- “在紧急情况下人们有理智地” 如何如何。实际上，那位火警站长举出了不少在火警中人们丧失理智的例子。

所以，你的模型和解法必须考虑现实条件，否则将影响你的结果的可靠性和可接受性。

Richard Hewitt 认为，要想使人们接受你的解，必须首先弄清谁是关键人物（如最大容量问题是建筑物的承包商等人），与他们交谈、沟通，确定解决问题的范围，他们心中的轻重缓急，需要作出若干折中（考虑政治、经济、风险等），在这个基础上，再加入你的数学的解答。他说，在学术领域之外，从未看到数学的解答统治其他决策准则，它只是补充。

Richard Hewitt 将他的建议归纳为：

- 根据你的能力去设计问题，与他人交流你的假设和解法。
- 一定要检验你的假设，保证它们合乎现实。

- 找出关键人物，引起他们的关注，证实听到的和考虑到的每个要求。
- 用例子和清晰扼要的语言进行交流，引起、强化听众的希望，着重说明他们能得到的效益，以及你的解答如何能超过他们现有的结果。

参 考 文 献

- [1] Frank Giordano, Results of the 1999 Mathematical Contest in Modeling, *The UMAP Journal* 20 (3) (1999)
- [2] Samuel W. Malone, et al, Determining the People Capacity of a Structure, *The UMAP Journal* 20 (3) (1999)
- [3] David Rudel, et al, Hexagonal Unpacking, *The UMAP Journal* 20 (3) (1999)
- [4] Gregg A. Christopher, et al, Room Capacity Analysis Using a Pair of Evacuation Models, *The UMAP Journal* 20 (3) (1999)
- [5] Jerrold R. Griggs, Judge's Commentary: The Outstanding Lawful Capacity Papers, *The UMAP Journal* 20 (3) (1999)
- [6] Richard Hewitt, Practitioner's Commentary: The Outstanding Lawful Capacity Papers; The Answer Is Not the Solution, *The UMAP Journal* 20 (3) (1999)

第十四章 大地污染问题

叶其孝

北京理工大学 应用数学系

提 要

本章介绍了 1999 年美国大学生数学建模竞赛(MCM-1999)的 C 题的优秀论文、评阅人的评述和我们的评注。主要内容：厄勒姆学院队和浙江大学队的优秀论文；评阅人和命题人的评述；我们的注记(包括对可供参考的其他优秀论文的简要评注)。

§ 14.1 MCM-1999 C 题 大地污染

[注] MCM-1999 的一个新的特点是增加了第三题(C 题)：一个与数学、化学、环境科学和环境工程有关的跨学科的实际问题。参赛队可以从网上得到一个实际的污染问题的数据。参赛 C 题的队要单独报名，一个学校可以有两个队报名参赛 C 题，这样一来每个学校至多可以报名 6 个队参加 MCM-1999。实际上，C 题的另一个名称是 ICM=Interdisciplinary Contest in Modeling。

背景

若干实践中重要但理论上困难的数学问题与污染的评估有关。这种问题之一就是根据只是在被怀疑为已污染地区的周围而不必直接在该地区中，测得的很少的测量数据来导出不易进入的地下的渗漏污染物的位置和数量，以及污染源的精确估计。

例子

数据集可通过 <http://www.comap.com/mcm/prod-a>

ta. xis 找到。

该数据集(一种超常文件 (an Excel file), 它能卸载到大多数电子数据表(spreadsheets))展示了从 1990 年到 1997 年在 10 个监测井处地下水中污染物的测量数据. 单位是微克/升($\mu\text{g/l}$). 8 个测井的位置和高度是已知的并在下表给出. 头两个数是在一张地图的直角格点上井的位置的坐标. 第三个数是井中水面高出平均海平面的高度(以英尺计).

井号	x-坐标(英尺计)	y-坐标(英尺计)	高度(英尺计)
MW-1	4187.5	6375.0	1482.23
MW-3	9062.5	4375.0	1387.92
MW-7	7625.0	5812.5	1400.19
MW-9	9125.0	4000.0	1384.53
MW-11	9062.5	5187.5	1394.26
MW-12	9062.5	4562.5	1388.94
MW-13	9062.5	5000.0	1394.25
MW-14	4750.0	2562.5	1412.00

数据集中另两个井(MW-27 和 MW-33)的位置和高度不知道. 在该数据集中你还会看到数字后面的字母 T (Top), M (Middle), 或 B (Bottom), 它们分别表示测量是在井的含水层的顶部、中部和底部进行的. 因此, MW-7B 和 MW-7M 是来自同一个井, 但分别是底部和中部的测量. 此外, 其他的测量数据表明水有流向该区域中的 MW-9 号井的趋势.

问题 1 试建立一个数学模型来决定在由该数据集表示的区域和时间里是否有任何新的污染物产生. 若有, 试识别新的污染物并估计它们的污染源的位置和时间.

问题 2 在收集任何数据之前, 会提出下列问题: 是否拟议中的数据类型和模型能给出关于污染物所在位置和数量的我

们想要的估计。液态的化学物质会从埋置在均匀的土壤中的存储设备中，许多类似的存储罐中的一个存储罐中渗漏。因为若要在许多大罐的下面去探测的费用会过分昂贵而且危险，所以只能在存储设备的边缘地区附近或在看来是更合意的地区的表面进行测量。试决定只是在整个存储罐的边界的外面或表面，进行什么类型的测量以及测量数目可以用于一个数学模型以决定渗漏是否发生，何时发生，何处（从哪个罐）发生，以及渗漏多少液体。

§ 14.2 厄勒姆学院队的优秀论文 ——污染的侦测：经由水文-化学 物质的分析对地下溢出的建模⁽¹⁾

14.2.1 摘要

来自一个被怀疑有地下污染的地区的 10 个监测井的数据，用来评估释放到大地中的污染物的源（的位置）、时间和（污染物）总量。基于对所记录的随时间变化的化学物质的浓度的分类来决定哪些是在数据收集期间积极活动的污染物，并说明由于不完全的数据集所造成的不相符之处。在这段时间内所发现的积极活动的化学物质同时改变浓度，说明了每种化学物质是与两处溢出有关的一个流体渗出中的（渗出物的）组成成分。把所选的那些积极活动的化学物质结合起来形成了一种复合指示物，它提供了在每个监测井处、每个日期该化学物质的浓度值。该复合指示物揭示有两次溢出发生，第一次在 1991 年 7 月到 1993 年 3 月间，第二次在 1995 年 1 月到 1997 年 4 月间，可能还会继续到数据收集期的最后时刻为止，渗出流体的主要化学成分得到了识别。

采用 Delaunay 的三角化方法，每个测量日期监测井之间复合指示物的浓度梯度。（我们的）结果指出在该区域内总的地下

水流动指向第 9 号井，基于浓度梯度随时间的变化，污染物的时间和位置可以近似表出。估计溢出源自点(8000, 4500)的周围地区。在初始三角化后，Voronoi 多腔形被用来构成表示溢出的总体积和位置(污染区域的体积)的一个凸包。多腔形包括较小的段，每一个都是特定的均匀浓度。程序 Geomview 是用来生成这些多腔形和凸包的。在每个浓度值处都可以计算体积，如果起初的受污染液体中复合指示物的浓度已知的话，最终可求得受污染液体的总体积。

最后，考察了各种测试和内插方法并融合到评估地下污染的方法中去。每种方法都是通过它在试题第二问的方案的应用中得到讨论，并利用数据集中给出的信息来测试方法的有效性。

14.2.2 引言

给定了 8 个地下水监测井的位置和高度(另两个井的位置不知道)，在 1990 年～1997 年期间在每个井处定期地进行了完全的化学分析，还分析了地下水流的总的指向，有可能精确地估计出源的位置，污染的来源以及渗入地下的污染物的总体积。在被怀疑为有渗漏的建筑的均匀土质上的化学药品存储设施的情形，由于费用和安全性的原因，禁止在被怀疑有溢出的地区直接收集数据。把来自布置在被怀疑有溢出的地区的外围，而不必布置在该地区的监测井的数据用于一个数学模型以决定：该地区是否有渗漏发生，若有，渗漏发生的时间和位置，在数据收集期间渗出液体的总量。

14.2.3 假设

- 所有的监测井都在地下井穿过一个含水层(或称为蓄水层，是一种能存储和输运大量水的一种地质结构)。该含水层有一个畅通无阻的不变的流动流速，流速和土介质的多孔性成反比。容许自由流体流经其测量仪器，对该地区的化学和地

质构成没有影响，并能提供邻近地区的的确切情况。这保证了监测井本身不会污染要评估的地区 [Soliman et al. 1997, 32].

- 在每个监测井中流体的体积是不变的，而且所有的监测井都有相同的流体体积。监测井中流体体积相同的假设使得要估计的污染物与每个监测井中的溶解物的浓度成正比。
- 不同的化学物质可以用不同的速度流经含水层。化学物质具有不变和特殊的能力在水溶液中移动，这种能力取决于分子的极性、亲水性以及每种复合物的初始浓度。
- 在数据集中发现存在某些化学物质是在地下水巾自然出现的，不会产生污染。在数据集收集期间在监测井中没有呈现出明显改变的化学物质都可以从数据集中去掉，不予考虑。此外，可以认为某些自然出现在地下水中的化学物质的成分可以在某个标准水平上浮动。
- 污染物的浓度在其源头最高，当时间和离开其源头的距离增加时污染物的浓度会减小。
- 所给的数据是不完全的。由于缺少可利用的数据，有些趋势可能误述或者整个被忽略掉。还有，对所给的值必须予以适当的评估，使得不至于把不能用的值当作零来处理。
- 数据的不一致可能是由于所用的仪器或在研究过程中样本和分析方法的差异造成的，数据的不一致，特别是出现于测试过的样本中的同样的数据的差异不能总是解释为环境的变化。
- 污染 (pollution) 定义为对有机体有害的污染物，而玷污 (contamination) 指的是比不一定造成伤害的自然出现的物质浓度更高的浓度 [Blatt 1997, 76]。在本问题中，我们假定这两个术语指的都是地下区域中人为造成的污染物，而不管污染物是否对有机体有什么影响。

14.2.4 处理数据

为有效地利用和解释如此大量而且各种各样的数据集，必须要用特殊的准则来组织和分类已知的信息。我们把原来的数据表（spreadsheet）形式的数据转换成一个数据库，使之可以建立查询并能有选择地进入到信息的任何部分。

数据中有些部分不是彻底的化学分析所需要的化学浓度，而只是其他的因素，例如特殊的传导性和完全软化的土，把这些数据分出来，并存储在另一个数据表中。尽管对污染问题进行建模的有些方法利用了这些测量数据，但是我们的模型却没有用到它们，因为我们不能在这些数据中侦查出有或者没有有意义的污染的模式。

利用反映所有日期所有井处给出的浓度的线图，我们识别出浓度展现出不可忽略的变化的化学物质。把这些化学物质的数据移走，并存储在一个单独的数据表中。这样做就把原有的 106 个测量类目只留下了 23 个。

14.2.5 测定污染的存在

在化学浓度关于时间的线图中显示出来的快速增加，在该地区的测试期间显然存在新的污染发生。所侦查到的作为新的污染物的这些化学物质包括：丙酮(acetone)，氨(ammonia)，砷(ar-senic)，钡(barium)，碳酸氢盐(bicarbonate)，钙(calcium)，氯化物(chloride)，铁(iron)，铅(lead)，镁(magnesium)，镍(nickel)，亚硝酸盐(nitrate/nitrite)，钾(potassium)，钠(sodium)，完全溶解了的固体(TDS= Total Dissolved Solid)，硫酸盐(sulfate=sulphate)，钒(vanadium)，锌(zinc)。

数据集中积极活动的化学物质的浓度升高了，并流在一起，这指示着每种这样的化学物质是单个的溢出中的流体的组成部分。尽管数据集中所有积极活动的化学物质的浓度都遵从明显的

趋势，但有些化学物质浓度变化的增强要远大于其他化学物质浓度的变化。我们把这些增强的化学物质作为指示器化学物质，以追踪溢出的运动。为进一步简化对溢出的侦查，我们把这些指示器化学物质（氯、硫酸盐、亚硝酸盐）亚硝酸盐的浓度合计在一起，形成一个复合指示器化学物质，其浓度指示了每个测试点处在给定的日期污染的存在。我们之所以选这些化学物质还因为它们是通常的污染物的成分，而且常被用来监测污染[B. C. Ministry of Environment, Land, and Parks 1999]。

在选择化学物质作为溢出的指示器时，重要的是要找到在整个数据收集期间在同样的日期在所有井处的测量是连贯一致的。数据集中满足这个准则的三种化学物质是：氯、硫酸盐、亚硝酸盐，因而我们把它们用于复合指示物中。因为数据集是不完全的，而且测量并不是对所有化学物质在所有井点或所有日期进行的，因此，确保这些复合指示物的浓度由于在给定井处或给定的日期缺少或没有数据不会误传溢出的运动是极为重要的。我们仔细检查了数据集，并去掉了重复记录两次数据的日期中的一次（取每个井处列出的浓度的平均值），还纠正了数据集中反常的数据，直到这三种化学物质中的每一个在每个测井处在每个所需要的日期恰好只有一个浓度值为止。这样做的例外情形包括在一开始测量是没有数据的那些井；把数据加进数据集使之成为有数据。

14.2.6 溢出的时间

可用展示给定井处各时刻复合指示物的浓度的一连串线图来估计溢出的时间。当画在一起时使得每条线表示一个检测井，这些浓度随时间变化的图展示了何时浓度首先开始增加，以及在哪些井处这种增长被首先记录下来。表示浓度首先增加的这些井处的记录也提供了溢出位置的粗略估计。〔编者注：我们在这里无法用黑白两色来有效地复制作者的图。〕

有两次溢出可能发生，第一次在 1991 年 7 月到 1993 年 3 月间。在这期间，认为是最靠近溢出的井处的浓度增加，然后退回到正常水平。第二次溢出大概发生在 1995 年 1 月并持续到 1997 年 1 月。这时，浓度开始降下来，但这也可能是因为溢出速率的下降，因而可能并不表示渗漏已经停止。

14.2.7 溢出源头的定位

从数据库查询生成的线图对测定溢出的存在、溢出发生的时间是极其有用的。但是，求溢出的源头则要用表明每个井处复合指示物的浓度关于时间变化的形象化的值来更有效地完成。这样，我们就能测定哪里浓度首先升高以及溢出移动的总的方向。知道了溢出的总的方向，我们就可以研究溢出的源头必须位于何处的范围。在三维情形，可以通过构造 Voronoi 多胞形来完成。这种插值方法把数据点和它们的邻近点组织成三角形，并把每个已知点的邻域分成多胞形，使得多胞形中的任一点到该数据点的距离比到其他数据点的距离更近。映射的三角剖分是惟一的，并有效地权重了该区域中任一点的值作为它到三维自然邻域的距离的函数。

尽管线图展示了一处溢出可能发生的日期，以及在哪些井处侦查到了浓度的变化，Voronoi 多边形用已知数据点的插值来更精确地定出溢出源头的位置。从每个井处和所选的日期处复合指示器化学物质的浓度的一系列图解来看，溢出的发展是显然的，而溢出源头的位置可以通过流动模式在第一次发生溢出的井点在地下往回追踪求得。〔编者注：我们在这里不复制作者的图了。〕

14.2.8 评估地下污染的方法

侦查地下流体的存在性是一个古老的问题，而且由于它在确定水源、油藏和矿藏的位置中的应用，因而有大量的有关的技术

和方法的信息可以利用。为侦查一个地下区域的确切性质，在一些点打一些取样井或检测井显然是必要的。然而，这种取样及随后的分析是费时、危险和费钱的，而且会有污染或破坏该地区地下水流的潜在可能。有许多地面或水面的测量有助于决定这种井的最有效的位置。此外，从现有井处得到的资料有助于决定是否需要增加井位以及在何处设井。一些重要的数据可以在打井前从地表的地球物理勘测中得到，这些数据包括重力的、电力和磁力传导力的仪器读数，这包括通过地表土的电流或磁场，测量电压降或潜在的磁场，以及给定位置处的密度。通过比较在该地区不同位置处表土的传导性，常常可侦查出地表下沙土或砂砾河床的存在 [Walton 1970, 61]。这是有用的，因为沙土或砂砾河床具有很高的多孔性，有能力包含以称为蓄水层的地下水的形式的自由流动流体。这些蓄水层是容纳和传播地下污染物的途径，所以为精确地预测污染的定位或污染源，了解蓄水层的流动和方向是至关重要的[Soliman et al. 1997, 32].

打井并进行监测

一旦确定了一开始的井的定位（地表测量应表明有蓄水层），就有几种类型的井可以利用。因为一开始的打井的钻孔是打井过程中最危险和最费钱的，因此诸如本问题中用来收集数据的永久性检测井从长远观点看是最经济的。这种井应能侦查出蓄水层中流体流动的方向和速率，以及提供要在实验室中进行化学分析的最重要的样本。这些井对于该地区的水来说必须是可渗透的，而且不会使蓄水层中的流动断流，或者由于打井过程或随时间增加而导致的腐蚀而引入新的污染物。

需要打多少口井？

为侦查地下化学物质溢出的源头、时间和体积所需要的井数随溢出的情况而变化很大。就这里所述的模型而言，量少要三口井。水系的方向和流动可以从一开始的井处和溶解在地下水中的化学物质的浓度值一起定出。另外的井应该沿水面的路径来打。

同时考虑到化学存储设施或其他被怀疑为污染源的一般位置，如果知道的话。如果在一开始的井处侦查出有污染，那么其他的井应该在溢出的“下游”开挖，而若没有侦查到污染，则应在“上游”打井，或根据该地区的地质情况沿不同的蓄水层打井。

当为侦查污染的流向和浓度而打的至少三口井都打好时，下面的模型可用来估计任何发现存在的污染物源的位置，以及溢出时间和溢出流体的总体积。在这种情形，如果利用更多井处收集的数据，那么对溢出可以作出更精确的预测。

模型 1：图方法

本模型需要至少在三个不同的位置，每个位置给出至少三个日期的化学物质的估计数。收集数据的井数愈多，模型的精确性愈高。

第一步是把浓度值置于数据库使得可以通过日期、收集数据的位置、化学物质或浓度来进入该数据库。利用该数据库，可以在每个井或数据收集点处生成按日期显示的所有化学物质的线图。可以从这些图决定污染的存在，以及所测量的化学物质的浓度的大的变化的日期。浓度的剧增指出了本模型中污染物的引入。如同本赛题那样的许多情形，在化学分析中侦查到的许多化学物质的同时升高或骤降指示着它们是一个共同的污染物的组成部分。化学物质有可能落入清晰可见的两组泄漏表明有两组流体泄漏。这时，一种化学物质，或者我们更愿意说是侦查到的一组化学物质结合在一起，形成一种指示器化学物质。这使得我们只要用一个浓度值就行了，从而简化了将来的图。

在创建统一的指示器化学物质时，数据的相容是极为重要的。同类型的数据必须在所有井点和所有日期处都有，或者调整必须融合进去，以防止会严重误传地下水中侦查到的化学物质浓度的数据集中的反常性。

有了把经由每个井位和给定时刻处有代表性的化学物质的浓度合在一起形成的合成指示器后，我们就可以生成测定溢出时间

和源头的新的图，表明在每个井处随时间变化的浓度的重量的图可以有效地用来测定溢出的时间。采集足够的数据以产生有待测量的化学物质的底线浓度是有用的。由 EPA (编译者注：美国国家环境保护局，Environment Protection Agency)，不列颠哥伦比亚环境署，Land 公司和 Parks 公司以及其他管理部门出版的表，列出了地下水中各种化学物质浓度的正常范围，它在区分水系中自然出现的化学物质和会造成污染的化学物质时也是有用的。解释这些图是相对容易的。在一张展示每个井或测试点位置的图中，识别出永远不显示指示器浓度的肯定增加的那些井(如果有这种井的话)。然后，一定要识别出首先达到高浓度值或总体展示最高浓度值的那些井。利用这些信息以及(在每个井处测得的或在问题的陈述中给出的)水系的流动方向和速率，可以把污染物追溯出一个估计是源头的位置。

对所指出的两处溢出中的每一个的这种过程的结果包含在图 14-1 中。这两处溢出都在 #9 和 #12 处达到高潮，然后在 #3, #11, 和 #7 处出现增高。这表明两处溢出都在点 (8000, 4500) 的周围区域中的某处开始。

溢出滞后到达 #7 可能是由于蓄水层中的流动模式造成的。初一看，会觉得奇怪，#7 离溢出处要远得多，为什么 #3 处的浓度会比 #7 处的浓度低。这可能是由于 #7 处的高度较低。在监测井的顶部、中部和底部的测量表明溢出是向下游(顺流)方向渗漏的，而且从未到达过井的底部。这也说明了为什么位于溢出的路径上的 #13

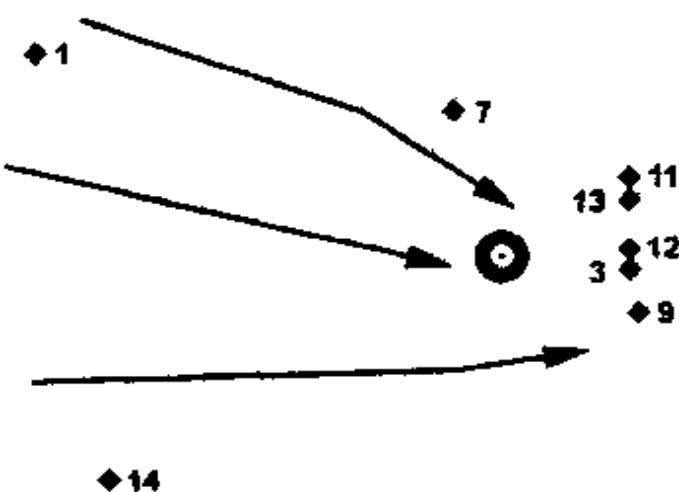


图 14-1 由牛眼指示的溢出位置以及蓄水层中流动的方向

从未显示化学物质的增长——因为其抽样只是在井的底部.

优、缺点

本方法的最大优点是其大量复杂数据集的有效性，大多数计算机能从一个数据表建立并利用这样一个数据库，而且一旦数据被组织起来，计算和解释结果所需的时间是最少的。因为它把新井的布置基于从现有井中采集到的信息之上，并提供了少至三个井时的粗略近似，从打井和良好维护来说，本模型是非常有效的。但是，本模型只是对溢出时间和源头的一个很一般的近似，而且会由于井位的不规则或地下水系流动模式的不规则而大受影响。在被怀疑为污染源头的下面没有直接可以利用的采样数据的情况下，直到渗漏早已渗入地下水供应之前侦查不出渗漏，从而妨碍了在对周围的地区造成问题之前试图制止渗漏的努力。本方法无法精确地测定溢出污染液体的体积。

模型 2：三维空间中的三角剖分插值

利用前面描述过的同样的方法，为了侦查污染的存在和有关的化学物质本模型需要创建数据库和线图。本模型还要用到监测污染物流动的化学物质的复合指示器的浓度的改变。计算几何中讲述了一种称为自然邻接点插值（natural-neighbor interpolation）的方法，用这种方法可以把高度不规则的数据组织起来，并形象地表示出来。利用 Delaunay 三角剖分法可以在任意点集中排列出惟一的一组三角形。任意一点处的值是基于其最邻近的三个已知点，即该点所属的三角形的三个顶点，完全局部地确定。

Delaunay 三角剖分和 Voronoi 多胞形对这类系统的插值之所以非常有效，有两个主要的原因：

- 它们提供了一个线性方程组，据此可以决定任意点的值，如果用这个方程组来求解，就可以准确地重新找到原来的数据点。
- 每一点的插值只受到其自然邻域的影响使得数据集中的不规

则性能反映在本模型中，但又不曲解其他点处模型的精确性 [Sambridge et al. 1995, 3].

计算机程序 Geomview 把浓度存储在数据库和监测井的位置坐标处，并生成了表示整个溢出的凸包。该凸包就是 Voronoi 多胞形最外面的表面，它是由较小的四面体组成的，每个四面体称为一个已知件，它表示了三个“自然邻接点”之间的空间。每个已知件，基于定义它的三个点处的已知浓度，有一个特定的不变的浓度 [Watson 1992, 108].

进入该程序的数据被区分开来，使得只有超过所设定的基线才会出现在表示溢出的形象化模型中。该程序根据每个特定浓度处的体积来加权，并把所有这些加了权的浓度加在一起计算溢出流体的总体积。在本模型中该凸包可以用来形象化地表示溢出，的位置和整个受污染区域的体积。我们用 Geomview 生成的图来展示由数据收集期间的 6 个数据集定义的溢出以及该时期受污染区域的体积[编者注：我们不复制作者的图了].

本模型在决定化学物质溢出的源头时也是有用的。它生成一个关于每个日期的新图解，这是基于进入到该日期从数据集中得来的。在数据收集期的开始，没有可见的污染。随着时间的推移，该图解清楚地指出哪些井经历着高于指示器化学物质的正常浓度水平。当把几个相继的凸包放在一起看时，污染物的近似源头以及在地下水水平岩石层中的流动方向是明显的。

误差分析

为检验本模型及线性模型的误差，可以对污染物源头的位置、发生渗漏的日期以及流出液体的总体积已知时间的数据集进行计算。位置、时间和体积的预测值和准确值之差除以准确值就决定了插值的百分比误差。误差量依赖于个别检测特定的诸多因素，包括源头离最近的监测井有多远、地下水系的速率、溢出的大小、监测井的数目以及许多其他的因素。

优、缺点

本模型强于图模型是因为它考虑了数据的三维形象化，还因为它应用了一种惟一的算法插值，来评估已知点间污染的存在。当用于对高度不规则的数据集时，由于自然邻接点的原则，可以从数据集中导出的 Delaunay 三角剖分和 Voronoi 多胞形以及凸包是极为精确的。不规则点或者点分布中的大的差异的出现反映在所产生的预测中，但不会导致数据的误传以及对已知点的值的不精确的歪曲 [Sambridge et al. 1995]。可利用的数据点愈多，这种插值愈精确，因为像任何插值一样，本模型在最靠近已知点的地方是非常精确的。

因为本问题给出的数据集极不规则，而且只包含很少几个采样井位，本方法的预测不可能完全精确，但本方法大大优于大多数插值方法。

至于前一个模型，直到溢出早已进入水层，该方法无法侦查出溢出。

还有，误差分析的仅有的显然的方法就是检验更多的点。给出的值仍然是近似的，而且对这么大的数据集进行分类仍然是相当冗长乏味的。如果从课题的一开始就采用这种方法，并且以一种特定而相容的方式采集数据，那么应用本模型的分析将使应用数据库和这里描述的过程变得相对简单。

本模型能决定是否有渗漏发生，溢出时间的近似值，并给出溢出源头的位置估计。本模型的最大优点就是它能测定污染区域的体积，并对其地下位置进行建模。当渗出液体中化学物质的初始浓度已知时，本模型还能测定渗出流体的体积。

本模型是非常或功有效的，基于已知信息生成溢出的近似边界，相邻的井应在此些边界处打，以确保溢出实际上是由当前的信息精确地预测的。如果新的井侦查到另外的溢出区域，那么新的边界就会生成，而且这个过程可以重复进行。在任一种情况下，一旦识别出污染区域决不会打不必要的井。

14.2.9 结果和建议

在测试期间的 1990 年到 1997 年期间该区域发生了新的污染。两个分隔开的同样的流体的溢出发生了，第一个从 1991 年 7 月开始约在 1993 年 3 月结束，而另一个从 1995 年 1 月开始大约在 1997 年 2 月逐渐减少，虽然在数据收集期间的其余时间也可能继续溢出。在图 14-1 中标出了这些溢出源自的区域。溢出是由下列化学物质组成的，它们是作为污染物释放到地下的：丙酮(acetone)，氨(ammonia)，砷(arsenic)，钡(barium)，碳酸氢盐(bicarbonate)，钙(calcium)，氯化物(chloride)，铁(iron)，铅(lead)，镁(magnesium)，镍(nickel)，亚硝酸盐(nitrate/nitrite)，钾(potassium)，钠(sodium)，完全溶解了的固体(TDS = Total Dissolved Solid)，硫酸盐(sulfate=sulphate)，钒(vanadium)，锌(zinc)。在 1997 年前污染区域的总体积是 3 千 2 百万立方英尺。(当流体中复合指示器的初始浓度可以得到的话，有可能通过对每个 Voronoi 多胞形已知件加权，并把浓度加起来乘以污染区域的总体积决定渗出流体的总体积。)

对将来的测试的建议包括：

- 识别地面的性质以更有效地设置一开始的监测井。
- 在同样的日期，在所有井处测试同样的化学物质以确保完整而准确的数据。
- 利用先前井所生成的信息预测溢出的边界，并把附加的井沿这个边界设置以极小化为精确，测定溢出大小所需的测试井位的个数。
- 确定监测井所在地的蓄水层水流的速率和方向，以精确预报溢出源头的时间和位置。

14.2.10 参考文献

Blatt, Harvey. 1997. *Our Geologic Environment*. Upper Sad-

- dle River, NJ: Prentice Hall.
- British Columbia Ministry of Environment, Land, and Parks. 1999.
<http://www.env.gov.bc.ca/wat/gws/gwbc/>.
- Flanagan, David. *Java in a Nutshell*. 1996. Sebastopol, CA: O'Reilly and Associates.
- Levy, Stewart, Tamara Munzer, and Mark Phillips. 1996. GeomView—1996.
<http://www.geom.umn.edu/software/download/geomview.html>.
- Sambridge, Malcolm, Jean Braun, and Herbert McQueen. 1995. Geophysical parameterization and interpolation of irregular data using natural neighbours. *Geophysical Journal International* 122: 837~857.
<http://rses.anu.edu.au/geodynamics/nm/SBM95/SBM.html>.
- Sedgewick, Robert. 1988. *Algorithms*. 2nd ed. Reading, MA: Addison-Wesley.
- Soliman, Mostafa M., Philip E. LaMoreaux, Bashir A. Memon, Fakhry A. Assaad, and James W. LaMoreaux. 1997. *Environmental Hydrogeology*. New York: Lewis Publishers.
- Levy, Stewart,, Tamara Munzer, and Mark Phillips. 1996. GeomView—1996.
<http://www.geom.umn.edu/software/download/geomview.html>.
- Todd, David Keith. 1959. *Groundwater Hydrology*. 2nd ed. New York: John Wiley and Sons.
- Walton, William C. 1970. *Groundwater Resource Evaluation*.

New York: McGraw Hill.

Watson, David F. 1992. *Contouring: A Guide to the Analysis and Display of Spatial Data*. New York: Pergamon Press.

§ 14.3 浙江大学队的优秀论文 ——污染源的定位^[2]

14.3.1 摘要

我们就侦查新的污染的策略研制了一个模型。三个过程支配了污染物在地下水中的运动：对流、扩散和阻滞。来自测井的信息由于：

- 决定地下水运动的速率和方向，
- 决定污染物的水平和垂直的延伸，
- 分析地下的结构和特征。

由于所给的数据的多样性和复杂性，我们采用两步数据选择以决定在数据收集期间最可能造成新的污染的污染物质。我们选择精选数据中最能代表这期间变化的那些化学物质，然后利用格点搜索算法，编写了一个模拟运动过程和识别污染源的位置和开始时间的计算机程序。程序用 C 语言编写，在 PC 上运行。确定了四个新的污染源的位置。我们的模型所给出的图和问题所给的数据相当一致。最后，我们测试了参数的敏感性。

14.3.2 假设

- 所有的土和蓄水层的性质在漫透水层和未漫透水层都是均匀的、各向同性的。
- 蓄水层由土和砂砾组成。
- 稳定、均匀的水流只发生在整个未漫透水层的垂直方向，以及在漫透水层水平(纵向)平面中，沿地下水流动方向流动。

- 物理过程起着最重要的作用，而化学过程可以忽略不计。
- 在整个监测期间描述各层特征的所有参数都是常数。
- 所有的污染物种都是点源。

14.3.3 问题 1

问题 1 要求估计污染源的位置和开始时间，所以我们考虑污染物的运动和地下的结构。

数据的分析和处理

我们假定污染物之间没有相互作用，所以我们可以分别处理各个污染物。用所给的各个井的坐标和水位的数据，利用对监测井高度的线性插值我们画出了如图 14-2 所给出的水位图。为计算简单起见，我们假定所有的地下水流沿同一方向流动。

数据的选择

因为我们由关于各种各样的污染物浓度的几千个数据点，我们必须小心地仔细地选择数据。我们用下列步骤来做这件事：

- 因为污染物强烈地受到不同渗透性的地层的影响，所以主要参数和污染物浓度的测量需要在蓄水层的一段高度区间内进行。我们需要一种在蓄水层不同高度处采样的方法。通过对数据集的分析，几乎每种污染物都只影响到井的一部分（顶部、中部或底部）。因此，对每种污染物我们只需考虑其对井的某一层的影响。而且，每个井底部的数据（如果有的话）保持或几乎保持不变，因此我们可以略去这种数据。
- 我们去掉某些污染物的数据，诸如四氯化乙烷（tetrachloroethane），丙烯醛（acrolein），苯（benzene），含溴乙烷。

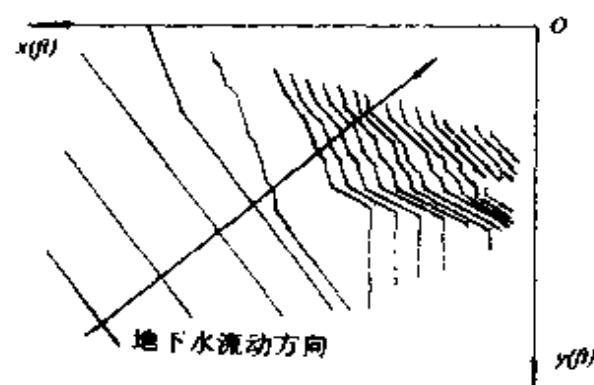


图 14-2 水位图

(bromomethane), 氯化苯 (chlorobenzene), 钴(cobalt), 等等. 因为在每个井中这些污染物的浓度几乎没有改变.

- 我们认为有些污染物, 例如锰 (manganese), 在相对稳定期间的脉冲浮动是由随机因素引起的. 因此, 把这种污染物质从数据集中去掉.
- 有一种成分, Carbon Total Organic, 其浓度显著下降, 从高于 1000 下降到低于 1.5, 因此, 我们把它去掉了.
- 现在只剩下四种污染物质, 钙 (calcium), 氯化物 (chloride), 镁 (magnesium) 和完全溶解了的固体 (TDS=Total Dissolved Solid).

数据的再选择

对于每种留下的污染物质, 为精确反映其浓度变化的趋势, 我们以如下方式再选择其数据:

- 对每个井每年我们选两个浓度值, 上半年一个, 下半年一个.
- 因为我们不知道 #MW-27 和 #MW-33 的位置, 而且在这两个井处的浓度变化小, 所以我们不考虑这两个井处的数据.
- 按照地下水流动的方向, 在 #MW-9 的平均浓度不应该高于 #MW-3 和 #MW-12 处的平均浓度, 这和所给的钙、氯等的数据矛盾. 对于钡 (barium) 也是如此. (1997 年在 #MW3-M 和 #MW-12M 处的浓度从 50 变到 85, 而在 #MW-9M 处却从 80 变到 95.) 所以, 我们认为 #MW-9 是一个泵井 (图 14-3). 所以在我们的分析中不采用 #MW-9 处的数据.

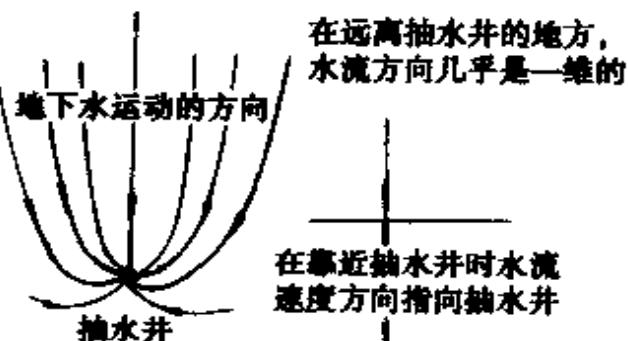


图 14-3 抽水井附近地下水的运动

最后, 我们在表 14-1 中列出了我们用以计算源的位置的钙

的数据。

表 14-1 模型中采用的污的数据

Date	MW-3M	MW-7M	MW-11M	MW-12M
12/7/1993	41	50	39	42
3/7/1994	42	50	43	47
9/19/1994	42	45	41	41
7/10/1995	36.5	54.3	44.7	59.5
10/10/1995	19.2	53	43.2	54.7
3/6/1996	62.4	65.1	50.7	82.4
10/9/1996	60.2	61.9	53.3	87.6
3/18/1997	63.8	125	53.2	87.6
12/15/1997	61.4	115	63.8	88.4

表 14-2 本文所用到的记号

a_L	水平扩散系数 (m)
a_T	垂直扩散系数 (m)
C	污染物质的浓度 (mg/liter)
C_b	背景浓度 (前面已经说过)
C_s	污染源处的浓度 (mg/liter)
D	渗透系数 (m^2/s)
H	水位 (ft)
I	水力梯度
K	水力传导性 (gal/day/ft ²)
L	水流方向的水平距离 (ft)
m	污染物质的排放率 (mg/day)
n	有效多孔性
q	污染物质的排放率 (mg/day)
R_d	阻滞因子
S	复合参数
t_0	污染开始的时间 (yr)
θ	地下水流方向和 x 轴的夹角
V_d	地下水的速度 (ft/day)
W	hantush 函数
(x_0, y_0)	污染源的坐标

- 根据数据可见，有些污染物质是在早年，例如在 1990 年值

查到的；我们称其浓度为背景浓度（background concentration）(C_b)，我们认为污染物以后的浓度是由背景浓度加上新注入的浓度构成的。

按照图 14-2, #MW-1 一处于水源水位。而且，根据数据集在这期间其底部的数据几乎不变。所以，我们采用#MW-1B 处的数据以如下方式来估计 C_b 。

C_b = 在这期间某种污染物在#MW-1B 处的浓度的算术平均值。

在表 14-2 中我们集中表示了本文所用的记号及其定义。

14.3.4 模型的设计

模型的形成

污染物质的运动包括对流、扩散和阻滞。而且，注意到区域很大，垂直运动可以忽略。因此，污染物在(浸透和未浸透)土中的运动可用下列二维方程来描述

$$R_d \frac{\partial C}{\partial t} = V_{d\alpha_L} \frac{\partial^2 C}{\partial x^2} + V_{d\alpha_I} \frac{\partial^2 C}{\partial y^2} - V_d \frac{\partial C}{\partial x}. \quad (1)$$

模型的解释

模型方程适用于稳定均匀的流动。把污染物质的连续(阶梯函数)和脉冲输入作为边界条件时，可求得该方程的解析解。阶梯函数适用于在无穷长时间里不变浓度的化学物质的输入。术语“无穷”和“有限”是相对于分析中的时间范围而言的。

我们假定和下列边界条件一起把一个阶梯函数(连续地)加在污染源上：

$$C(x, y, 0) = 0, (x, y) \neq (0, 0);$$

$$C(0, 0, t) = C_0;$$

$$C(\pm\infty, y, t) = C(x, \pm\infty, t) = 0, t \geq 0.$$

模型的解

这是一个偏微分方程。这类方程适用于包括质量迁移、流体

动力学和热转换在内的广泛的一类问题.

对于在 $t=0$ 处的瞬时点源, 有一个形为

$$C(x, y, t) = S \exp\left(\frac{x}{2\alpha_L}\right) [W(0, b) - W(t, b)], \quad (2)$$

其中

$$m = C_0 q, S = \frac{m}{4\pi V_d (\alpha_L \alpha_T)^{1/2}},$$

而 $W(u, b)$ 是 hantush 函数,

$$W(u, b) = \int_u^{\infty} \frac{\exp\left[-y - \frac{b^2}{2y}\right]}{y} dy,$$

其中

$$b = \sqrt{\frac{x^2}{4\alpha_L^2} + \frac{y^2}{4\alpha_L \alpha_T}}.$$

在计算前, 我们首先按照前面的假设对所用到的参数分分类:

- 数据处理中污染源的坐标和时间是未知的, 从而 m 的值也是未知的. 因此, x_0 , y_0 , t_0 和 S 是变量.
- 参数 α_L , α_T , θ 和 V_d 都是常数.

主要的任务是求得污染源的位置和时间. 因此, 为得到最优解我们研制了一个格点最优化程序.

- 我们用如下的步骤来估计污染源的位置和转移坐标:
 - 置污染源为新的坐标原点,
 - 置新的 x 轴和地下水水流方向平行,
 - 置新的 y 轴垂直于新的 x 轴.
- 我们构造一个方程来计算污染物质在地下的运动. 我们计算在每个井处的浓度改变, 并与数据集中的变化相比较. 我们反复地调整污染源的位置、 S 的值、 t_0 的值(细节见后)直到满意的一致为止. 收敛准则是数据和预测值间的残差的平方和. 要求极小的目标函数是

$$\sum_i [(C_i - C_b) - C'_i]^2,$$

其中 C_i 是第 i 个井处污染物质浓度的数据值, C'_i 是模型的预测值, 而 C_b 是背景浓度.

参数的估计

我们对浸透层作如下估计:

- 水力传导系数 K : 我们只在水平方向考虑水力传导性, 以每天每平方英尺加仑来度量. 根据文献 $K = 265 \text{ gpd}/\text{ft}^2$ ($1 \text{ gpd}/\text{ft}^2 = 4.72 \times 10^{-5} \text{ cm/sec}$).
- 水力梯度: 根据由插值作出的图 14-2, 我们假定地下水的流动是一维的.
- 地下水(孔隙裂缝中的水)的速度 V_d : 按达西(Darcy)定律, V_d 定义为

$$V_d = -KI/n,$$

其中 I 是水力梯度, K 是水力传导系数, n 是有效的多孔性. 我们假定浸透层的土的类型是多孔性为 20% 的沙土, 所以, 我们估计 $V_d = 1.5 \text{ ft/day}$.

- 弥散系数 α : 该系数融合了两种形式的弥散: 动态弥散和分子扩散. 根据文献, 水平弥散系数和垂直弥散系数近似相等, 其估计值为 25 ft.
- 阻滞因子: 阻滞是基于污染物的特征和蓄水层的组成. 因为它的影响不大, 我们估计 $R_d = 1$.
- 污染源的浓度: 根据文献, 当地下水位足够高致使污染物直接进入地下水时, C_0 的值就是污染物浓度的估计值.

结果

有四种新的污染物质: 钙、氯、镁和 TDS. 我们模型预测的污染源的位置和开始时间见表 14-3.

最后, 我们模拟了污染物质的运动过程, 并与所给的数据集比较(图 14-4). 从这些图我们得出结论:

- 在几乎理想的条件下, 本模型是合适的; 对常规的应用而言, 希望有一个更为强健的模型.

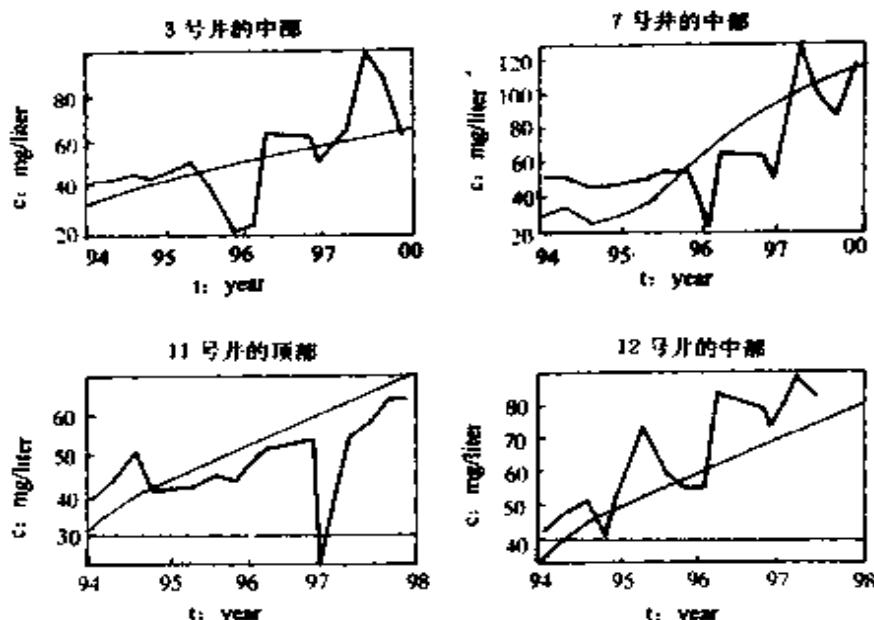


图 14-4 在四个井处钙的浓度，粗线是数据，
细线是模型的预测

表 14-3 污染物质的源头和开始时间

污染物	x-坐标(ft)	y-坐标(ft)	开始时间(月/日/年)
TDS	7077	6538	8/12/1991
镁	6423	7461	1/1/1994
氯	6931	5823	5/18/1991
钙	7750	6040	9/1/1993

- 即使数据曲线和模型预测曲线拟合得并不好，它们仍然展示了相似的趋势。

敏感度分析

我们作了敏感度分析以说明我们模型的稳定性。我们分别改变常数 α_L , α_T , θ 和 V_d 的值 10%，并计算了相应的污染源头的位置和时间(见表 14-4)。

本模型显示了很好的稳定性，但 θ 对本模型的结果有相对大的影响。因此，有理由把 θ 作为变量并在 $(\theta, x_0, y_0, t_0, S)$ 这个五维空间里重复我们的格点搜索算法。对钙而言，我们得到的

比较结果见表 14-5.

表 14-4

参数扰动的影响

参数	位置变化(ft)	时间变化(ft)
θ	70	0.2
a_L	<10	<0.1
a_T	10	<0.1
V_d	<10	<0.1

在扩展了的模型中， θ 的值增大了 7%，我们认为地下水方向有所偏转，如图 14-5 所示。

表 14-5

4-和 5-维模型的比较

维数	θ	x_0 (ft)	y_0 (ft)	t_0 (yr)	$S \times 10^6$
4	0.785	7750	6040	93.75	2.1
5	0.84	7750	6100	93.60	2.2

14.3.3 问题 2

地方性的假设

- 存储罐位于地下浸透层。
- 地下水流动方向保持不变。
- 浸透层是半无穷的。
- 渗透过程是连续的，因为钢制地下存储系统渗漏的主要原因是腐蚀。

模型设计

为快速、精确地侦查出污染物质，我们研制了一种三步骤方法。

1. 根据存储罐的形状、大小以及地下水流动的方向，我们确定

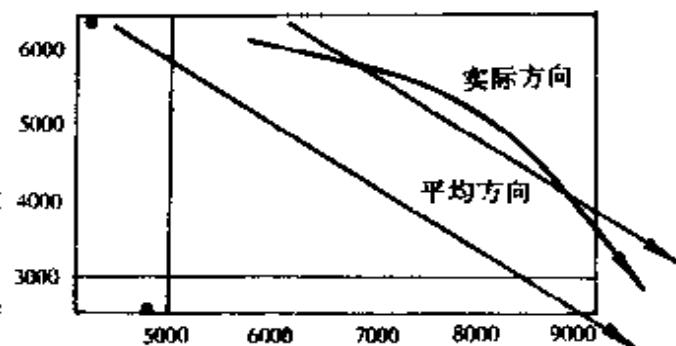


图 14-5 地下水流动的偏转

第一组井的位置和数目。如果存储罐是一边长为 S 米的正方体，则第一组井的数目为 $N=S/20$ ，即在地下水流动方向的垂直线上每隔 20 米打一口井，如图 14-6 所示。我们监测来自这些井处的数据。

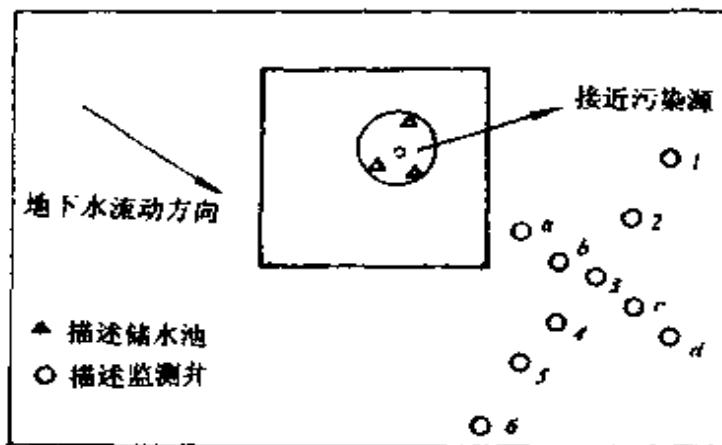


图 14-6 监测井的定位。空圆表示起初的监测井；标有 a, b, c, d 的圆是在侦查到有污染并发现大多数污染影响到 #3 后打的井

- 一旦有污染的证据，我们要决定哪个井受污染的影响最大。在该井的附近，我们沿地下水流动方向打一系列井（也许 5 个或者更多）。因此，我们可以构造一个三维的公式来计算污染物质浓度的浮动。这里存储设施占的面积可能不太大（边长小于 1000ft），所以，我们不能用(1)。我们利用三维方程

$$R_d \frac{\partial C}{\partial t} + V_d \frac{\partial C}{\partial x} = D \left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2} \right) + \frac{m}{n}.$$

因为渗透是一个连续过程，我们假定污染源的作用如同一个阶梯函数（连续地）并满足下列边界条件：

$$C(x, y, z, 0) = 0, (x, y, z) \neq (0, 0, 0),$$

$$m(x, y, z, t) = qC_0 \delta(x, y, z),$$

$$C(\pm\infty, y, z, t) = C(x, \pm\infty, z, t) = C(x, y, \pm\infty, t) = 0, \\ t \geq 0.$$

对于在 $t=0$ 时刻的瞬时点源，该方程有一个形为

$$C(x, y, z, t) = \frac{R_d q C_0}{8 \pi n D r} \exp\left(\frac{V_d}{2D}\right) \\ \times \left\{ \exp\left(\frac{V_d x}{2D}\right) \operatorname{erfc}\left[\frac{r+V_d t}{2} \left(\frac{R_d}{D}\right)^{1/2}\right] \right. \\ \left. + \exp\left(\frac{-V_d x}{2D}\right) \operatorname{erfc}\left[\frac{r-V_d t}{2} \left(\frac{R_d}{D}\right)^{1/2}\right] \right\}.$$

的解析解，其中 $r = (x^2 + y^2 + z^2)^{1/2}$.

当 $t \rightarrow \infty$ 时定常态方程给出

$$C(x, y, z, t) = \frac{R_d q C_0}{4 \pi n D r} \exp\left(\frac{V_d(r-x)}{2D}\right). \quad (3)$$

为方便起见，我们用记号 $C_m(x, y, z, t)$ 来表示(3)的右端.

对常数 V_d , R_d , n , q 和 D , 我们可以画一个如图 14-7 所示的浓度值为 $0.01C_0$ 的等浓度面.

设高度 Height 就是等浓度面的最大高度. 我们以问题 1 中的同样方法转换直角坐标.



图 14-7 水的溢出

对于在 (x, y) 的监测井以及厚为 b 的蓄水层，我们对三种情形来考虑井中的浓度：

- 若 $b \ll H$ 或 $b \ll$ 存储设施的大小，我们可以把(3)转换成如(1)那样的二维方程.
- 若 $b \geq H/2$, 有理由认为 $b = \infty$. 因此，问题可以简化. 我们假定蓄水层中的物质除原点外不能进入未漫透层. 因此， $\partial C / \partial z_{z=0} = 0$. 而且对于水面下的每一点 (x, y, z) 处其浓度为半无穷空间情形下由(3)给出的浓度的二倍

$$C(x, y, z, t) = 2C_m(x, y, z, t).$$

- 其他情形，在蓄水层的上表面和下表面处(见图 14-8)，我们有

$$\frac{\partial C}{\partial z} \Big|_{z=0} = \frac{\partial C}{\partial z} \Big|_{z=-b} = 0.$$

画污染源关于轴 A' (下表面)

○ 虚拟源 2

对称的虚拟源 1. 因此，满足条

件 $\partial C / \partial z_{z=0} = 0$, 尽管 $\partial C / \partial z_{z=b} = 0$ 不再成立，用同样的方法

我们画出与轴 A (上表面)对称的

虚拟源 2. 重复这个过程，我们

得到虚拟源 3 等等. 上表面或下表面处的浓度可以认为是所有源(包括虚拟源)的浓度值的积累的结果. 即，

$$C_i(x, y, z) = 2 \sum_0^r C_m(x, y, z + 2(-1)^{r+1} [\frac{i+1}{2}] b, t).$$

实际上，出于以下原因我们只需要考虑前三个虚拟源：

- 从这些源到 A (A') 的距离是最小的，所以它们对 C_i 最有影响. 其他的虚拟源离 A (A') 很远，它们之间的距离一般都大于值 Height. 所以，我们忽略这些虚拟源.
- 从离开蓄水层很远地方的虚拟源处释放出来的污染物要经过很长时间才能达到蓄水层.

最后，我们把(3)变换为

$$C_i(x, y, z) = 2[C_m(x, y, z, t) + C_m(x, y, z + 2b, t) \\ + C_m(x, y, z - 2b, t)].$$

因此，我们得到解析解 $C(x, y, z, t)$. 然后，我们采用在问题 1 中用过的某些以计算机为基础的方法来计算污染源的近似位置和时间.

3. 最后一步我们画一个中心在近似源点 Q 半径为 25 米(或更大)的圆在这个圆内，我们从地表取了一些土样并分析其化学成分以求得最大值. 因此，我们能精确地识别污染源的位置.

数值积分格式

为计算渗漏必须对空间变量 C 积分. 不幸的是(3)的积分没有解析表示，因而必须做数值积分. 对本模型我们用了一种三维的积分格式. 渗出液体的分子质量 M 计算为

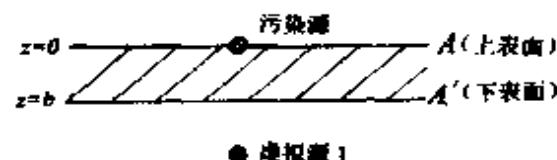


图 14-8 蓄水层中的污染源

$$M = \iiint C(x, y, z, t) dx dy dz \approx \sum_{i,j,k} C_{ijk} \delta x \delta y \delta z,$$

其中 C_{ijk} 是在“微元”(ijk)处算出的浓度。我们采用了一致的空间步长 $\delta x = \delta y = \delta z = 1$ 米。

一种更好的估计质量的方法

当用计算机处理数据时，为得到准-最优解我们极小化了变差。同时我们估计出了 m 的值，所以我们可以更方便有效地计算渗出液体的分子质量为

$$M = mt.$$

14.3.6 模型的优、缺点

优点

- 模型有很好的实践性，而所给出的算法几乎没有时间复杂性。对于所给问题的规模，我们为格点搜索算法编的 C 语言程序在 Pentium-166 计算机上运行不到 2 分钟。
- 模型给出了预测值和数据的很好的一致。它是快速、有效和稳定的。
- 至于精选数据以简化计算，准确性并没有降低。我们在表 14-6 中列出钙的数据作为说明。

表 14-6 为简化计算精选数据的效果

数据点的数目	x_0 (ft)	y_0 (ft)	t_0 (yr)(从 1900 年算起)
60(原来的)	7750	6060	93.70
36(精选后)	7750	6040	93.75

缺点

- 如果要侦查的区域不够大，就会有一些误差。当污染源和监测井间的距离增大时，计算精度增加了，测量精度降低了，应答时间增加了。
- 为降低计算复杂性，我们简化了地下水水流网，这会影响到结果的精度。

- 在模型中没有考虑统计因素，使得我们的模型不能和未经提炼的数据相拟合。

14.3.7 参考文献

Abriola, Linda M., editor. 1989. *Groundwater Contamination*. Wallingford, U. K.: International Association of Hydrological Sciences.

Barcelona, Michael, Allen Wehrhann, Joseph F. Keely, and Wayne A. Pettyjohn. 1990. *Contamination of Ground Water: Prevention, Assessment, Restoration*. Park Ridge, NJ: Noyes Data Corp.

Guswa, J. H., W. J. Lyman, A. S. Donigian, T. Y. R. Lo, and E. W. Shanahan. 1984. *Groundwater Contamination and Emergency Response Guide*. Park Ridge, NJ: Noyes.

Ward, C. H., W. Giger, and P. L. McCarty, editors. 1985. *Ground Water Quality*. New York: Wiley.

§ 14.4 评阅人的评注

——关于大地污染问题的论文⁽³⁾

14.4.1 作者介绍

David L. Elliott 是在密苏里州圣路易斯(St. Louis)的华盛顿大学的数学系统方向的退休教授，自 1992 年起一直是在马里兰州科利奇帕克 (College Park) 的马里兰大学的系统研究院的资深访问研究科学家。

他于 1953 年在加州波莫纳 (Pomona College) 学院获数学学士，于 1959 年在南加州大学 (USC) 获数学硕士学位，于 1969 年在洛杉矶加州大学 (UCLA) 获工程博士学位。在美国海军海洋中

心(U. S. Naval Ocean Center)从事控制系统和海洋声学方面的研究工作后，分别在洛杉矶加州大学、华盛顿大学教书，他也是罗得岛州布朗大学(Brown University)和洛杉矶加州大学的访问学者。1987年～1989年他是美国科学基金会系统理论的课题主任。他的研究是在非线性控制理论和应用数学(包括血凝固的运动学——他患有血友病)。

他是美国电气及电子工程师学会(IEEE)的成员(Fellow)和美国工业与应用数学学会(SIAM)、美国数学会(AMS)、美国数学协会(MAA)和西格马克茜科学研究所(Sigma xi, The Scientific Research Society)的会员。他曾是几种数学杂志的助理编辑，曾主编《控制神经系统》(*Neural Systems for Control*, Academic Press, 1997)一书。他和MCM的早先关联是作为MCM-1985和MCM-1986特等奖获得者华盛顿大学队的指导教师。

14.4.2 评论文章的主要内容

各队都倾向于花较多的时间做第一个问题，这里的评述关心的也是第一个问题。最优秀的论文两个问题都掌握得很好。我看过的论文把第一个问题分成几个子问题：为着手解决这些问题需要除了问题描述以外的假设，最好的论文清楚、明白地提出了其假设，并尽可能证明其合理性。子问题包括：

1. 列出“污染”的种类，
2. 污染物传输的数学模型，
3. 污染物溢出的时间和溢出个数，
4. 溢出源的位置(利用1～3)。

对问题的回答差异很大，甚至在最好的论文中也是这样，这有赖于假设和对数据的解释。优胜者展示了关于有关文献及其解释的小心仔细的检索；

- 子问题提得好，并找到了从中能得到有用的回答的数学模型；
- 能以清楚和有说服力的方式表述他们的结果；
- 避免了大的错误（这些错误看来常常是由于队员之间沟通交流很差造成的）。

网址上的问题的陈述可能会很好地给出有关场所（垃圾堆？仓库？）的某些描述，有关土/蓄水层类型的描述，以及由该领域的专业人员给出的其他的定性信息。

数据表的列是按化验过的化学物质的种类加以标记的，数据表还包括几个井处、几个深度处以及分开的时间序列处的化学物质的浓度。参赛者对哪些种类是“污染物质”几乎没有一致的意见：大多数种类的化学物质的浓度（例如，有机氯化物）可以忽略，其他种类随时间不增加或者可能是自然生成的（在雨中或在土中），有些列看来用了多于一种的度量单位。

优秀论文就是表明模型可以有多么不同的很好的例证。浙江大学队把时间拟合为一个简单的偏微分方程的解。注意，在该文中提到的“扩散”主要是动态的起因（渗透，虽然在我看过的任一篇文章中都没有见到这个术语）。有的参赛者似乎认为热的（布朗）扩散是重要的，但它们太小了，在大多数流体中是观察不到的。厄勒姆学院队忽略了扩散，但假定了不同种类的化学物质以不同的速率行进。该队用时间序列图来很好地选择要察看种类和估计时间。

其他的队在求得流动的方向，把扩散和对流一起考虑，或选取数据的理由等方面碰到麻烦。有的队没有找到有助于建模的有关文献。

§ 14.5 命题人的评述——关于大地污染 问题优秀论文的评述^[4]

14.5.1 作者介绍

Yves Nievergelt 于 1976 年毕业于瑞士洛桑联邦理工学院 (Ecole Polytechnique Federale de Lausanne) 数学系，主修偏微分方程的泛函和数值分析。他于 1984 年在华盛顿州的华盛顿大学获博士学位，在 James R. King 的指导下博士论文是研究多复变函数的。

他现在是 *The UMAP Journal* (大学生数学及其应用杂志) 的助理编辑。他是若干个 UMAP Modules (大学生数学及其应用教学单元)、低层次数学应用案例研究的参考文献 (*The UMAP Journal* 6 (2) (1985): 37~56)，以及《企业管理中的数学》 (*Mathematics in Business Administration* (Irwin, 1989)) 一书的作者。他最近的新书是《小波使事情变得容易》 (*Wavelet Made Easy* (Birkhauser, 1999))。

14.5.2 评述文章的主要内容

本赛题的来源

11月的一个深夜，在荒凉的西部的一个伐木采矿小镇的一家小酒店里，我碰巧坐在一位水文地质学家的旁边。除了经常能在这类小酒店里听到的典型的有关当地酿造的饮料的品味、河里的鱼少了、在你的后院里蹑手蹑脚地潜近鹿的美洲狮，以及在前院里捣毁你的垃圾桶的熊外，谈话转向监测该地理区域的私人的或政府的机构的讨论，本赛题的数据就是来自该地区。

数据是由真实的、原始的、未经改变的数据表组成，其中列出了私人或政府机构利用的在某个地区打的通过蓄水层的井中污染物的测量数据。数据不仅是真实的，也是有意义的，意即某些

人可能从这些数据的分析和解释中受损或者得益。

经过一些讨论后，这位水文地质学家同意在 MCM 中利用这些数据，条件是在问题的陈述中不可以包含可能识别出有关当事人的信息。

对优秀论文的评述

获特等奖的两个队用了两种基本上不同的方法，但每个队都显示了他们对问题情景的深刻理解以及对数学概念的很强的运用能力。两个队对问题情景的理解使他们能够

- 不为庞大的数据弄得不知所措；
- 没有因为数据集中有些地方没有数据、有些地方数据重复而停步；
- 没有因为不熟悉的化学物质的名词而受阻；
- 合乎规律地识别出潜在的化学污染物质的出现；
- 对看来几乎是不变的、随机的，或有可能展现趋势的化学物质按浓度进行分类；
- 能找到并利用来自文献或因特网(互联网)的参考资料。

除了对问题的这种理解外，两个队还采用了很不相同的数学求解方法。

浙江大学队采用了最小二乘法(最小变差)来拟合偏微分方程建模中对流、耗散和滞后的物理参数：扩散系数、地下水的速度、地层的多孔性，以及可能的污染源的时间和定位。他们的结果提出了四个污染物溢出，分别在 1991 年坐标为(7077, 6538) 和(6931, 5823) 的地方，以及 1993 年靠近年底坐标为(7750, 6040) 和(6423, 7461) 的地方。

厄勒姆学院队利用 Delaunay 三角化以及 Voronoi 多胞形来内插浓度梯度，以及最终用于探测污染物浓度的突然增加，并追溯到一个假定存在的污染源。他们的结果提出两个溢出，第一个是 1992 年，然后又于 1996 年在坐标(8000, 4500)附近。

虽然这两个结果互不相同，但他们提出了在地图右上角

$[0, 10000] \times [0, 7000]$ 的区域里有污染物的增长. 两个队的论文还包括有关假设、模型、方法和结果的仅仅是一个周末的工作的令人印象深刻的表述.

§ 14.6 评注

浙江大学队的三位队员沈权(电机系)、杨振宇(数学系)和何晓飞(计算机系)在竞赛三天的拼搏中表现确实非常突出, 他们真正发挥了团队精神, 充分发挥了各自的长处, 团结合作取得了好成绩. 面对巨量的数据, 他们能冷静对待, 不惊不乱, 沉着处理. 特别值得提出的是, 虽然他们偏微分方程的知识很有限, 但却能正确地选用对流-反应扩散方程(看来他们并不知道这个专门名词!)作为他们的数学模型的基础, 并能从文献中找到点源情形等的解析解, 结合格点搜索算法求得较好的结果. 在求解方程时, 较为精确地设定方程中的系数(参数)极为重要, 因为即使模型基本正确, 但如果模型中的参数离开实际情况很远, 那么也不会得到正确的结果. 三位队员能根据问题给出的数据, 经过精选, 并能进一步作出一些合理的假设, 很好地确定了这些参数.

此外, 地下水层中污染的侦查也是极其重要的. 所以, 本问题完全可以让有兴趣的同学进一步去思考.

参 考 文 献

(1) James R. Garlick, Savannah N. Crites, Pollution Detection: Modeling an Underground Spill through Hydrochemical Analysis. UMAP, v. 21 (2000), no. 3, 343~354.

(2) Shen Quan, Yang Zhenyu, He Xiaofei, Locate the Pollution

Source, UMAP, v. 21 (2000), no. 3, 355~368.

(3) David L. Elliott, **Judge's Commentary: The Ground Pollution Papers**, UMAP, v. 21 (2000), no. 3, 369~370.

(4) Yves Nievergelt, **Author's Commentary: The Outstanding Ground Pollution Papers**, UMAP, v. 21 (2000), no. 3, 371~372.

附录 I 1998 年~2000 年中国及美国 大学生数学建模竞赛试题

1. 1998 年中国大学生数学建模竞赛试题 (1998, 9, 22~24)

A 题 投资的收益和风险

市场上有 n 种资产(如股票、债券……) S_i ($i=1, \dots, n$) 供投资者选择, 某公司有数额为 M 的一笔相当大的资金可用作一个时期的投资。公司财务分析人员对这 n 种资产进行了评估, 估算出在这一时期内购买 S_i 的平均收益率为 r_i , 并预测出购买 S_i 的风险损失率为 q_i 。考虑到投资越分散, 总的风险越小, 公司确定, 当用这笔资金购买若干种资产时, 总体风险可用所投资的 S_i 中最大的一个风险来度量。

购买 S_i 要付交易费, 费率为 p_i , 并且当购买额不超过给定值 u_i 时, 交易费按购买 u_i 计算(不买当然无须付费)。另外, 假定期银行存款利率是 r_0 , 且既无交易费又无风险($r_0=5\%$)。

1) 已知 $n=4$ 时的相关数据如下:

S_i	$r_i(\%)$	$q_i(\%)$	$p_i(\%)$	$u_i(\text{元})$
S_1	28	2.5	1	103
S_2	21	1.5	2	198
S_3	23	5.5	4.5	52
S_4	25	2.6	6.5	40

试给该公司设计一种投资组合方案, 即用给定的资金 M , 有选择地购买若干种资产或存银行生息, 使净收益尽可能大, 而总体风险尽可能小。

2) 试就一般情况对以上问题进行讨论, 并利用以下数据进行计算。

S_i	$r_i(\%)$	$q_i(\%)$	$p_i(\%)$	$a_i(\text{元})$
S_1	9.6	42	2.1	181
S_2	18.5	54	3.2	407
S_3	49.4	60	6.0	428
S_4	23.9	42	1.5	549
S_5	8.1	1.2	7.6	270
S_6	14	39	3.4	397
S_7	40.7	68	5.6	178
S_8	31.2	33.4	3.1	220
S_9	33.6	53.3	2.7	475
S_{10}	36.8	40	2.9	248
S_{11}	11.8	31	5.1	195
S_{12}	9	5.5	5.7	320
S_{13}	35	46	2.7	267
S_{14}	9.4	5.3	4.5	328
S_{15}	15	23	7.6	131

(本题是由浙江大学陈叔平提供的)

B 题 灾情巡视路线

下图为某县的乡(镇)、村公路网示意图，公路边的数字为该路段的公里数。

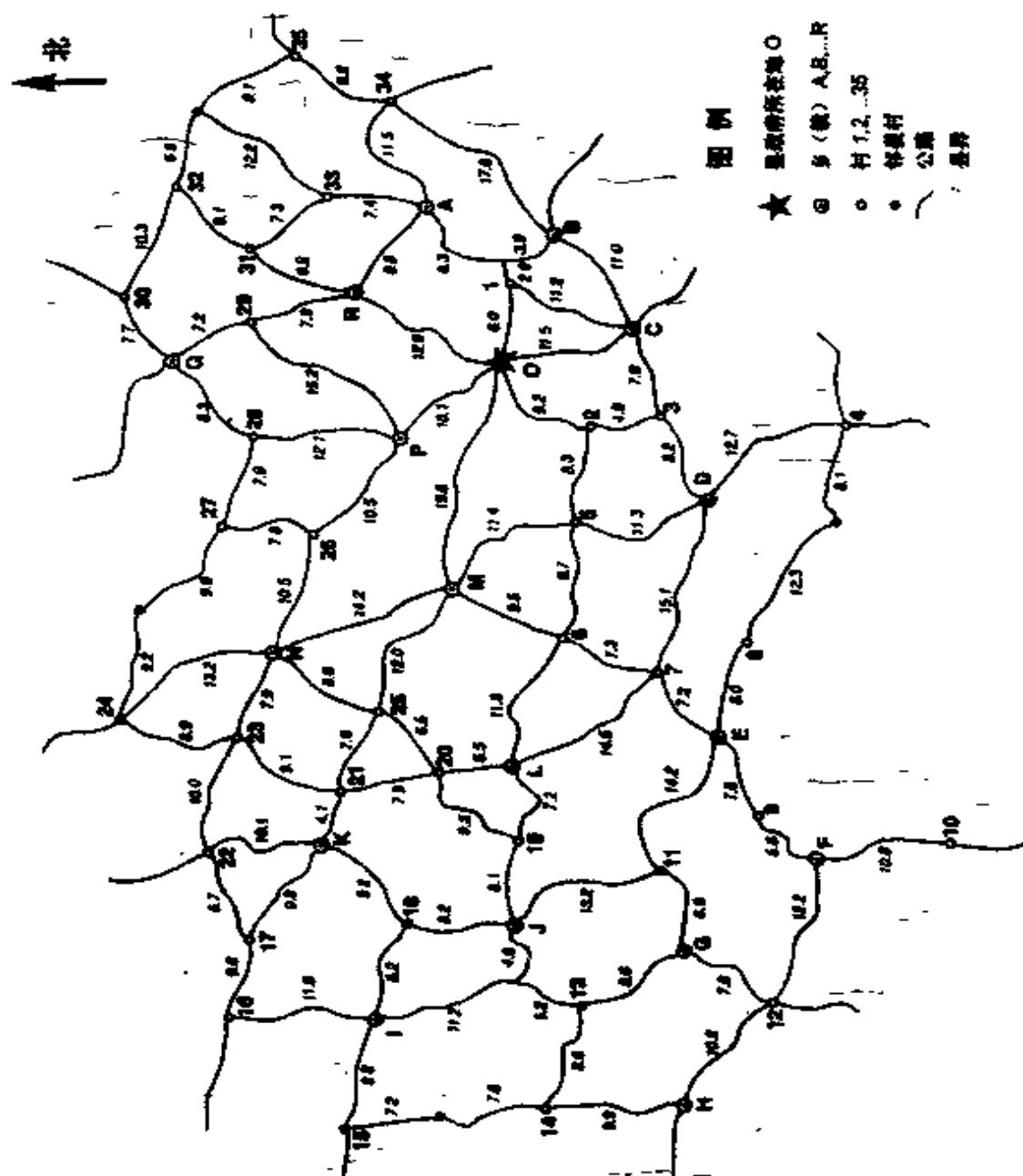
今年夏天该县遭受水灾，为考察灾情、组织自救，县领导决定，带领有关部门负责人到全县各乡(镇)、村巡视。巡视路线指从县政府所在地出发，走遍各乡(镇)、村，又回到县政府所在地的路线。

1. 若分三组(路)巡视，试设计总路程最短且各组尽可能均衡的巡视路线。

2. 假定巡视人员在各乡(镇)停留时间 $T=2$ 小时，在各村停留时间 $t=1$ 小时，汽车行驶速度 $V=35$ 公里/小时。要在 24 小时内完成巡视，至少应分几组；给出这种分组下你认为最佳的巡视路线。

3. 在上述关于 T , t 和 V 的假定下, 如果巡视人员足够多, 完成巡视的最短时间是多少; 给出在这种最短时间完成巡视的要求下, 你认为最佳的巡视路线.

4. 若巡视组数已定 (如三组), 要求尽快完成巡视, 讨论 T , t 和 V 改变对最佳巡视路线的影响.



(本题是由上海海运学院丁颂康提供的)

2. 1999 年创维杯中国大学生数学建模竞赛试题 (1999, 9, 21~23)

A 题 自动化车床管理

一道工序用自动化车床连续加工某种零件，由于刀具损坏等原因该工序会出现故障，其中刀具损坏故障占 95%，其他故障仅占 5%。工序出现故障是完全随机的，假定在生产任一零件时出现故障的机会均相同。工作人员通过检查零件来确定工序是否出现故障。

现积累有 100 次刀具故障记录，故障出现时该刀具完成的零件数如附表。现计划在刀具加工一定件数后定期更换新刀具。

已知生产工序的费用参数如下：

故障时产出的零件损失费用 $f = 200$ 元/件；

进行检查的费用 $t = 10$ 元/次；

发现故障进行调节使恢复正常后的平均费用 $d = 3000$ 元/次（包括刀具费）；

未发现故障时更换一把新刀具的费用 $k = 1000$ 元/次。

1) 假定工序故障时产出的零件均为不合格品，正常时产出的零件均为合格品，试对该工序设计效益最好的检查间隔（生产多少零件检查一次）和刀具更换策略。

2) 如果该工序正常时产出的零件不全是合格品，有 2% 为不合格品；而工序故障时产出的零件有 40% 为合格品，60% 为不合格品。工序正常而误认有故障停机产生的损失费用为 1500 元/次。对该工序设计效益最好的检查间隔和刀具更换策略。

3) 在 2) 的情况，可否改进检查方式获得更高的效益。

附：100 次刀具故障记录(完成的零件数)

459	362	624	542	509	584	433	748	815	505
612	452	434	982	640	742	565	706	593	680

926	653	164	487	734	608	428	1153	593	844
527	552	513	781	474	388	824	538	862	659
775	859	755	649	697	515	628	954	771	609
402	960	885	610	292	837	473	677	358	638
699	634	555	570	84	416	606	1062	484	120
447	654	564	339	280	246	687	539	790	581
621	724	531	512	577	496	468	499	544	645
764	558	378	765	666	763	217	715	310	851

(本题是由北京大学孙山泽提供的)

B 题 钻井布局

勘探部门在某地区找矿。初步勘探时期已零散地在若干位置上钻井，取得了地质资料。进入系统勘探时期后，要在一个区域内按纵横等距的网格点来布置井位，进行“撒网式”全面钻探。由于钻一口井的费用很高，如果新设计的井位与原有井位重合（或相当接近），便可利用旧井的地质资料，不必打这口新井。因此，应该尽量利用旧井，少打新井，以节约钻探费用。比如钻一口新井的费用为 500 万元，利用旧井资料的费用为 10 万元，则利用一口旧井就节约费用 490 万元。

设平面上有 n 个点 P_i ，其坐标为 (a_i, b_i) , $i=1, 2, \dots, n$, 表示已有的 n 个井位。新布置的井位是一个正方形网格 N 的所有结点（所谓“正方形网格”是指每个格子都是正方形的网格；结点是指纵线和横线的交叉点）。假定每个格子的边长（井位的纵横间距）都是 1 单位（比如 100 米）。整个网格是可以在平面上任意移动的。若一个已知点 P_i 与某个网格结点 X_j 的距离不超过给定误差 ϵ ($=0.05$ 单位)，则认为 P_i 处的旧井资料可以利用，不必在结点 X_j 处打新井。

为进行辅助决策，勘探部门要求我们研究如下问题：

- 1) 假定网格的横向和纵向是固定的（比如东西向和南北

向), 并规定两点间的距离为其横向距离(横坐标之差绝对值)及纵向距离(纵坐标之差绝对值)的最大值. 在平面上平行移动网格 N , 使可利用的旧井数尽可能大. 试提供数值计算方法, 并对下面的数值例子用计算机进行计算.

2) 在欧氏距离的误差意义下, 考虑网格的横向和纵向不固定(可以旋转)的情形, 给出算法及计算结果.

3) 如果有 n 口旧井, 给出判定这些井均可利用的条件和算法(你可以任意选定一种距离).

数值例子. $n=12$ 个点的坐标如下表所示:

i	1	2	3	4	5	6	7	8	9	10	11	12
a_i	0.50	1.41	3.00	3.37	3.40	4.72	4.72	5.43	7.57	8.38	8.98	9.50
b_i	2.00	3.50	1.50	3.51	5.50	2.00	6.24	4.10	2.01	4.50	3.41	0.80

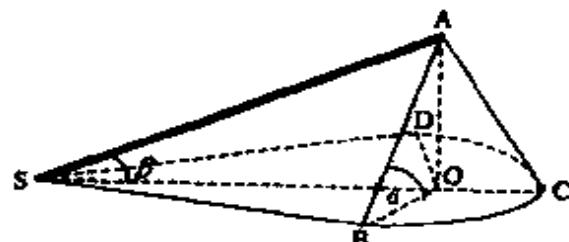
(本题是由郑州大学林治勋提供的)

中国大学生数学建模竞赛试题(大专组)

C 题 煤矸石堆积

煤矿采煤时, 会产生无用废料——煤矸石. 在平原地区, 煤矿不得不征用土地堆放矸石. 通常矸石的堆积方法是:

架设一段与地面角度约为 $\beta=25^\circ$ 的直线形上升轨道(角度过大, 运矸车无法装满), 用在轨道上行驶的运矸车将矸石运到轨道顶端后向两侧倾倒, 待矸石堆高后, 再借助矸石堆延长轨道, 这样逐渐堆起如右图所示的一座矸石山来.



现给出下列数据:

矸石自然堆放安息角(矸石自然堆积稳定后, 其坡面与地面形成的夹角) $\alpha \leqslant 55^\circ$;

研石容重(碎研石单位体积的重量)约 2 吨/米³；
运研车所需电费为 0.50 元/度(不变)；
运研车机械效率(只考虑堆积坡道上的运输)初始值(在地平面上)约 30%，坡道每延长 10 米，效率在原有基础上约下降 2%；
土地征用费现值为 8 万元/亩，预计地价年涨幅约 10%；
银行存、贷款利率均为 5%；
煤矿设计原煤产量为 300 万吨/年；
煤矿设计寿命为 20 年；
采矿出研率(研石占全部采出的百分比)一般为 7%~10%。
另外，为保护耕地，煤矿堆研土地应比实际占地多征用 10%。

现在煤矿设计中用于处理研石的经费(只计征地费及堆积时运研车用的电费)为 100 万元/年，这笔钱是否够用？试制订合理的年度征地计划，并对不同的出研率预测处理研石的最低费用。

(本题是由太原理工大学贾晓峰提供的)

D 钻井布局

勘探部门在某地区找矿，初步勘探时期已零散地在若干位置上钻井，取得了地质资料。进入系统勘探时期后，要在一个区域内按纵横等距的网格点来布置井位，进行“撒网式”全面钻探。由于钻一口井的费用很高，如果新设计的井位与原有井位重合(或相当接近)，便可利用旧井的地质资料，不必打这口新井。因此，应该尽量利用旧井，少打新井，以节约钻探费用。比如钻一口新井的费用为 500 万元，利用旧井资料的费用为 10 万元，则利用一口旧井就节约费用 490 万元。

设平面上有 n 个点 P_i ，其坐标为 (a_i, b_i) ， $i = 1, 2, \dots, n$ ，表示已有的 n 个井位。新布置的井位是一个正方形网格 N 的所有结点(所谓“正方形网格”是指每个格子都是正方形的网格)；

结点是指纵线和横线的交叉点). 假定每个格子的边长(井位的纵横间距)都是 1 单位(比如 100 米). 整个网格是可以在平面上任意移动的. 若一个已知点 P_i 与某个网格结点 X_j 的距离不超过给定误差 ϵ (=0.05 单位), 则认为 P_i 处的旧井资料可以利用, 不必在结点 X_j 处打新井:

为进行辅助决策, 勘探部门要求我们研究如下问题:

1) 假定网格的横向和纵向是固定的(比如东西向和南北向), 并假定距离误差是沿横向和纵向计算的, 即要求可利用点 P_i 与相应结点 X_j 的横坐标之差(取绝对值)及纵坐标之差(取绝对值)均不超过 ϵ . 在平面上平行移动网格 N , 使可利用的旧井数尽可能大. 试提供数值计算方法, 并对下面的数值例子用计算机进行计算.

2) 在问题 1) 的基础上, 考虑网格的横向和纵向不固定(可以旋转)的情形, 给出算法及计算结果.

数值例子. $n=12$ 个点的坐标如下表所示:

i	1	2	3	4	5	6	7	8	9	10	11	12
a_i	0.50	1.41	3.00	3.37	3.40	4.72	4.72	5.43	7.57	8.38	8.98	9.50
b_i	2.00	3.50	1.50	3.51	5.50	2.00	6.24	4.10	2.01	4.50	3.41	0.80

(本题是由郑州大学林治勤提供的)

3. 2000 年网易杯中国大学生数学建模竞赛试题

A 题 DNA 序列分类

2000 年 6 月, 人类基因组计划中 DNA 全序列草图完成, 预计 2001 年可以完成精确的全序列图, 此后人类将拥有一本记录着自身生老病死及遗传进化的全部信息的“天书”. 这本大自然写成的“天书”是由 4 个字符 A, T, C, G 按一定顺序排成的长约 30 亿的序列, 其中没有“断句”也没有标点符号, 除了这 4 个字符表示 4 种碱基以外, 人们对它包含的“内容”知之甚少,

难以读懂。破译这部世界上最巨量信息的“天书”是 21 世纪最重要的任务之一。在这个目标中，研究 DNA 全序列具有什么结构，由这 4 个字符排成的看似随机的序列中隐藏着什么规律，又是解读这部天书的基础，是生物信息学（Bioinformatics）最重要的课题之一。

虽然人类对这部“天书”知之甚少，但也发现了 DNA 序列中的一些规律性和结构。例如，在全序列中有一些是用于编码蛋白质的序列片段，即由这 4 个字符组成的 64 种不同的 3 字符串，其中大多数用于编码构成蛋白质的 20 种氨基酸。又例如，在不用于编码蛋白质的序列片段中，A 和 T 的含量特别多些，于是以某些碱基特别丰富作为特征去研究 DNA 序列的结构也取得了一些结果。此外，利用统计的方法还发现序列的某些片段之间具有相关性，等等。这些发现让人们相信，DNA 序列中存在着局部的和全局性的结构，充分发掘序列的结构对理解 DNA 全序列是十分有意义的。目前在这项研究中最普通的思想是省略序列的某些细节，突出特征，然后将其表示成适当的数学对象。这种被称为粗粒化和模型化的方法往往有助于研究规律性和结构。

作为研究 DNA 序列的结构的尝试，提出以下对序列集合进行分类的问题：

1) 下面有 20 个已知类别的人工制造的序列(见后面)，其中序列标号 1~10 为 A 类，11~20 为 B 类。请从中提取特征，构造分类方法，并用这些已知类别的序列，衡量你的方法是否足够好。然后用你认为满意的方法，对另外 20 个未标明类别的人工序列（标号 21~40）进行分类，把结果用序号(按从小到大的顺序)标明它们的类别(无法分类的不写入)：

A 类 _____； B 类 _____。

请详细描述你的方法，给出计算程序。如果你部分地使用了现成的分类方法，也要将方法名称准确注明。

这 40 个序列也放在如下地址的网页上，用数据文件 Art-

model-data 标识，供下载：

网易网址：www.163.com 教育频道 在线试题；

教育网：www.cbi.pku.edu.cn News mcm2000

教育网：www.csiam.edu.cn/mcm

2) 在同样网址的数据文件 Nat-model-data 中给出了 182 个自然 DNA 序列，它们都较长。用你的分类方法对它们进行分类，像 1)一样地给出分类结果。

提示：衡量分类方法优劣的标准是分类的正确率，构造分类方法有许多途径，例如提取序列的某些特征，给出它们的数学表示：几何空间或向量空间的元素等，然后再选择或构造适合这种数学表示的分类方法；又例如构造概率统计模型，然后用统计方法分类等。

(本题是由北方工业大学孟大志提供的)

Art-model-data

1. aggcacggaaaaacgggaataacggaggaggacttggcacggcattacacggagga
cgaggiaaggaggcttgtctacggccggaagtgaagggggatatgaccgtttgg
2. cggaggacaaacgggatggcggtattggaggtggcggactgttcgggaattattcg
gtttaaacggacaaggaaggcggctggaacaaccggacggtgccagcaaagga
3. gggacggatacggattctgccacggacggaaaggaggacacggcggacatacagg
cggcaacggacggaacggaggaaggagggcggcaatcggtacggaggcggcgg
4. atggataacggaaacaaaccagacaaaacttcggttagaaatacagaagcttagatgcata
tgtttttaaataaaatttgtatttatggtatcataaaaaaaggttgcga
5. cggctggcggacaacggactggcggttccaaaaacggaggaggcggacggaggct
acaccacccgtttcggcggaaaggcggaggctggcaggaggcttacggggag
6. atggaaaatttcggaaggcggcaggcaggaggcaaaggcggaaaggaaac
ggcggataittcggaaagtggatattggagggcggaaataaggaaacggcggcaca
7. atgggattattgaatggcggaggaagatccggaaataaaatatggcggaaagaacttgt
tttcggaaatggaaaaaggacttaggaatcggcggcaggaaggatggaggcgg
8. atggccgatcggcttaggcttggcggaaacaaataggcggaaattaaggaaaggcgttctc

gctttcgacaaggaggcggaccataggaggcggatttaggaacggttatgagg
9. atgcggaaaaaggaaatgttggcatggcgggctccggcaactggagggttcggcc
atggagggcgaaaatcggtggcggcggcagcgctggcggagttttagggagcgcg
10. tggccgcggagggcccgtggcgcggatttataaggccttgtaaggagg
tggeatccaggcgctgcacgcgtgcgcggcaggaggcacgcggaaaaaacg
11. gttagatttaacgttttatggaatttatggaattataaatttaaaaatttatattttta
ggtaagtaatccaacgttttattactttaaaatttaattttatt
12. gtttaattactttatcattaatttagtttaatttaatttagtaagatgaat
ttggtttttttaaggtagtttttaattatcgtaaggaaagttaaa
13. gtattacaggcagaccttatttaggttattattattttttttttttta
agttAACGAAATTATTTCTTAAAGACGTTACTTAATGTCAATGC
14. gttagtcttttagattaaattatttagattatgcagtttttacataagaaaaattttt
ttcgagttcatattcaatctgttttattaaatcttagagatatta
15. gtattatatttttatttttatttttagaatataattttaggtatgttttttttttttttt
tt
16. gttttttaaatttaattttaaattttaaaataaaaaattttactttctaaaattggctc
tggatcgataatgtasacttattgaatctatagaattacattttgt
17. gtatgtctatttacggaaagaatgcaccactatgtattgaaattatctatggctaaa
acccteagtaaaatcaatccctaaacccttaaaaaacggcggctatccc
18. gtttaatttatttattcccttacggcaatttaattttattacggtttttttacaatttttt
tttgcctatagagaaattacttacaaaacgttattttacataactt
19. gttacatttttatttatttattccgttacgataatttttaccccttttgcgtgagttttt
ttcttactttttctttatataaggatctcatttaataatetttaa
20. gtatttaactctttacttttttttactctctacatttcatttctaaaactgtttgatt
taaaacttttttttaaggattttttacttatectctgttat
21. tttagcttagtcagctagtagttacaatttgcacaccagtttgcaccatcttaat
ttcgatccgtacgtaatttagcttagatttggatttaaggatttagattga
22. tttagtacagtagctcagtcagaacgatgttaccgtiaacgtgacgtaccgtacgc
taccgtiaccggattccggaaagccgattaaggaccgatcgaaaggg

23. cggggggataggccgacggggaccggattcgggacccgagggaaattcccggtt
atagaaggtttagttcccccggatttagggccggatggctgggaccc
24. tttagctactttagctattttagtagctagccagccttaaggctagtttagct
agcattgttcatttgggacccaagttcgactttacgattttagtttgcacgt
25. gaccaaaagggtggcttagggacccgatgccttagtcgcagctggaccagggttccccagg
gtttaggcaaaagctgacggcaattgcaatttaggcttaggcca
26. gatttactttagcatttttagctgacgtagcaagcattagcttagccaatttcgcatt
tgccagtttcgcagctcagtttaacgcgggatcttagcttcaagcttttac
27. ggattcggatttacccggggattggcggaaacgggaccccttaggtcgggacccattagg
gatggaaatgccaaaggacgttgttttagccagtcgttaaggcttag
28. tccttagatttcagttactatatttacttacagtcttttagatttccttacgatttgac
ttaaaatttagacgttagggcttatcgttatggatataatttagcttattttcga
29. ggc当地有更多行，每行包含一个DNA序列，长度不一，最长的行有40个字符。

gaaatgcccaaaggagggccacgggttagatgccasagtgcaccgt
 38. aacttttaggcattccagtttacgggttatttcccagttaaacttgcaccatttac
 gtgttacgatttacgtataatttaccttatttggacactttagttgggttac
 39. ttagggccaagtcggaggcaaggaattctgatccaagtccaatcacgtacagtccaa
 gtcaccgttgcagctaccgttaccgtacgttcaagtcaaatccat
 40. ccatttagggttatttacctgttattttcccgagaccttaggttaccgtacttttaa
 cggttaccttgaaattttgacttagctaccctggatttaacggccagtt

B 题 钢管订购和运输

要铺设一条 $A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_{15}$ 的输送天然气的主管道，如图一所示(见后面). 经筛选后可以生产这种主管道钢管的钢厂有 S_1, S_2, \dots, S_7 . 图中粗线表示铁路，单细线表示公路，双细线表示要铺设的管道(假设沿管道或者原来有公路，或者建有施工公路)，圆圈表示火车站，每段铁路、公路和管道旁的阿拉伯数字表示里程(单位 km).

为方便计，1km 主管道钢管称为 1 单位钢管.

一个钢厂如果承担制造这种钢管，至少需要生产 500 个单位. 钢厂 S_i 在指定期限内能生产该钢管的最大数量为 S_i 个单位，钢管出厂销价 1 单位钢管为 p_i 万元，如下表：

i	1	2	3	4	5	6	7
S_i	800	800	1000	2000	2000	2000	3000
p_i	160	155	155	160	155	150	160

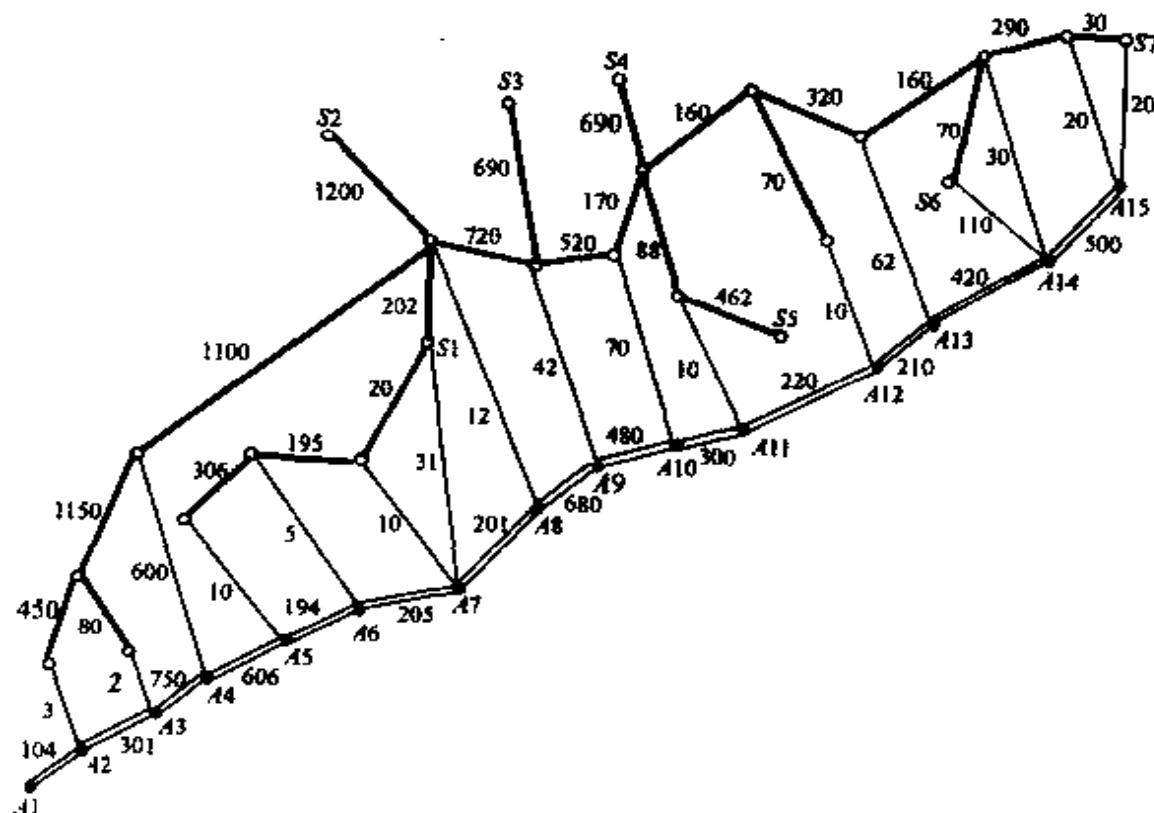
1 单位钢管的铁路运价如下表：

里程(km)	≤ 300	$301 \sim 350$	$351 \sim 400$	$401 \sim 450$	$451 \sim 500$
运价(万元)	20	23	26	29	32
里程(km)	501 ~ 600	601 ~ 700	701 ~ 800	801 ~ 900	901 ~ 1000
运价(万元)	37	44	50	55	60

1000km 以上每增加 1 至 100km 运价增加 5 万元.

公路运输费用为 1 单位钢管每公里 0.1 万元 (不足整公里部分按整公里计算).

钢管可由铁路、公路运往铺设地点 (不只是运到点 A_1 , A_2 , ..., A_{15} , 而是管道全线).

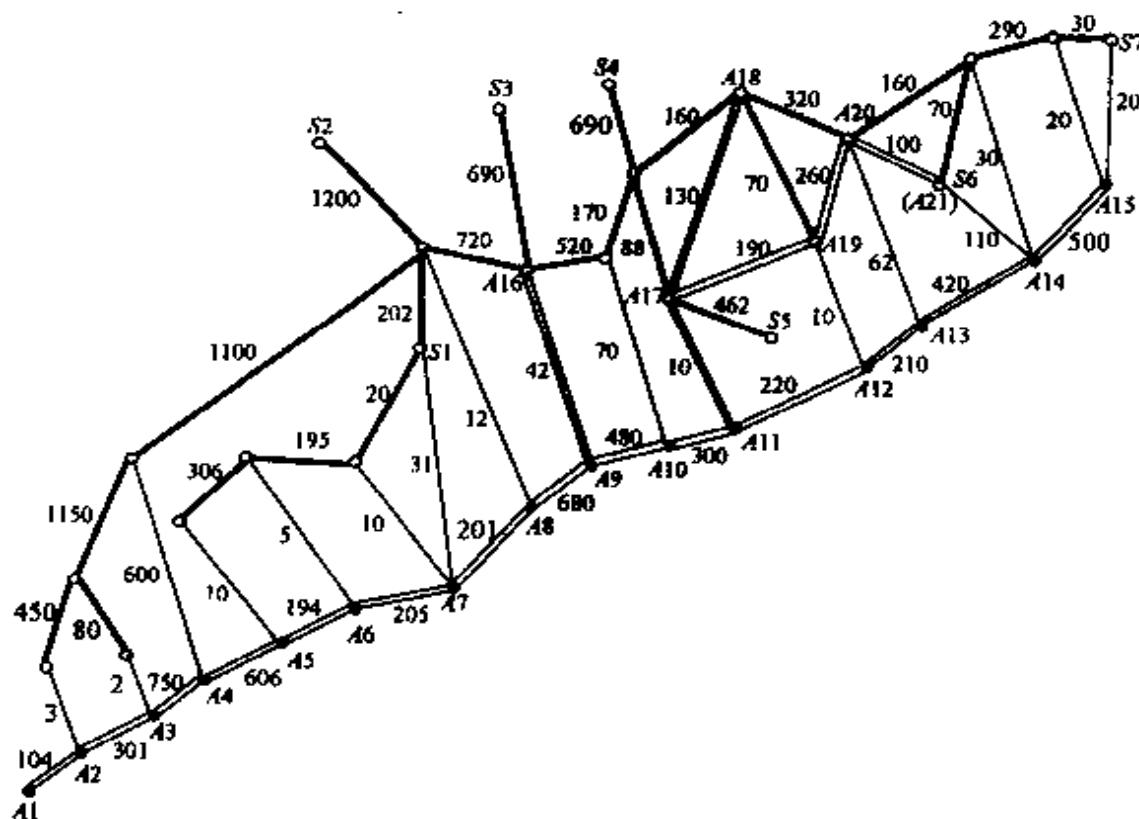


图一

(1) 请制定一个主管道钢管的订购和运输计划, 使总费用最小(给出总费用).

(2) 请就(1)的模型分析: 哪个钢厂钢管的销价的变化对购运计划和总费用影响最大, 哪个钢厂钢管的产量的上限的变化对购运计划和总费用的影响最大, 并给出相应的数字结果.

(3) 如果要铺设的管道不是一条线, 而是一个树形图, 铁路、公路和管道构成网络, 请就这种更一般的情形给出一种解决办法, 并对图二按(1)的要求给出模型和结果.



二

(本题是由武汉大学费浦生提供的)

2000 年网易杯中国大学生数学建模竞赛试题(大专组)
(由 C, D 题中任选 1 题)

C层 飞越北极

今年6月，扬子晚报发布消息：“中美航线下月可飞越北极，北京至底特律可节省4小时”，摘要如下：

7月1日起，加拿大和俄罗斯将允许民航班机飞越北极，此改变可大幅度缩短北美与亚洲间的飞行时间，旅客可直接从休斯敦、丹佛及明尼阿波利斯直飞北京等地。据加拿大空中交通管制局估计，如飞越北极，底特律至北京的飞行时间可节省4个小时。由于不需中途降落加油，实际节省的时间不止此数。

假设：飞机飞行高度约为 10 公里，飞行速度约为每小时

980 公里；从北京至底特律原来的航线飞经以下 10 处：

- A1 (北纬 31 度，东经 122 度)；
- A2 (北纬 36 度，东经 140 度)；
- A3 (北纬 53 度，西经 165 度)；
- A4 (北纬 62 度，西经 150 度)；
- A5 (北纬 59 度，西经 140 度)；
- A6 (北纬 55 度，西经 135 度)；
- A7 (北纬 50 度，西经 130 度)；
- A8 (北纬 47 度，西经 125 度)；
- A9 (北纬 47 度，西经 122 度)；
- A10 (北纬 42 度，西经 87 度)。

请对“北京至底特律的飞行时间可节省 4 小时”从数学上作出一个合理的解释，分两种情况讨论：

- (1) 设地球是半径为 6371 千米的球体；
- (2) 设地球是一旋转椭球体，赤道半径为 6378 千米，子午线短半轴为 6357 千米。

(本题是由复旦大学谭永基提供的)

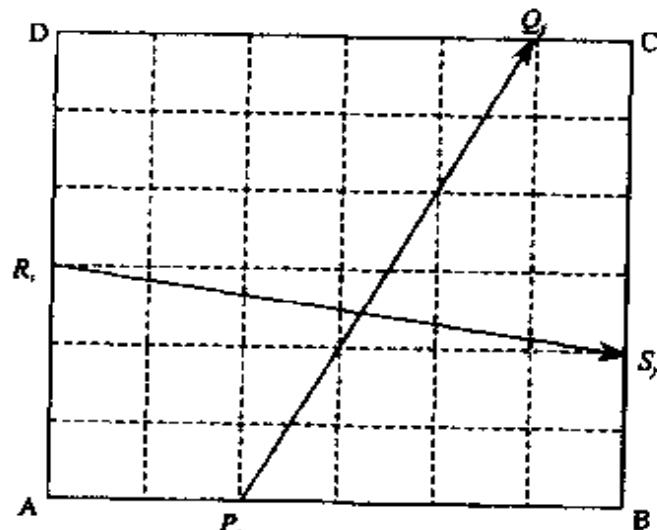
D 题 空洞探测

山体、隧道、坝体等的某些内部结构可用弹性波测量来确定。一个简化问题可描述为：一块均匀介质构成的矩形平板内有一些充满空气的空洞，在平板的两个邻边分别等地设置若干波源，在它们的对边对等地安放同样多的接收器，记录弹性波由每个波源到达对边上每个接收器的时间，根据弹性波在介质中和在空气中不同的传播速度，来确定板内空洞的位置。现考察如下的具体问题：

一块 240(米)×240(米)的平板(如图)，在 AB 边等地设置 7 个波源 P_i ($i=1, \dots, 7$)，CD 边对等地安放 7 个接收器 Q_j ($j=1, \dots, 7$)，记录由 P_i 发出的弹性波到达 Q_j 的时间 t_{ij} (秒)；

在 AD 边等距地设置 7 个波源 R_i ($i=1, \dots, 7$)， BC 边对等地安放 7 个接收器 S_j ($j=1, \dots, 7$)，记录由 R_i 发出的弹性波到达 S_j 的时间 τ_{ij} (秒)。已知弹性波在介质和空气中的传播速度分别为 2880(米/秒) 和 320(米/秒)，且弹性波沿板边缘的传播速度与在介质中的传播速度相同。

- 1) 确定该平板内空洞的位置。
- 2) 只根据由 P_i 发出的弹性波到达 Q_j 的时间 t_{ij} ($i, j=1, \dots, 7$)，能确定空洞的位置吗？讨论在同样能够确定空洞位置的前提下，减少波源和接受器的方法。



t_{ij}	Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7
P_1	0.0611	0.0895	0.1996	0.2032	0.4181	0.4923	0.5646
P_2	0.0989	0.0592	0.4413	0.4318	0.4770	0.5242	0.3805
P_3	0.3052	0.4131	0.0598	0.4153	0.4156	0.3563	0.1919
P_4	0.3221	0.4453	0.4040	0.0738	0.1789	0.0740	0.2122
P_5	0.3490	0.4529	0.2263	0.1917	0.0839	0.1768	0.1810
P_6	0.3807	0.3177	0.2364	0.3064	0.2217	0.0939	0.1031
P_7	0.4311	0.3397	0.3566	0.1954	0.0760	0.0688	0.1042

τ_{ij}	S_1	S_2	S_3	S_4	S_5	S_6	S_7
R_1	0.0645	0.0602	0.0813	0.3516	0.3867	0.4314	0.5721
R_2	0.0753	0.0700	0.2852	0.4341	0.3491	0.4800	0.4980
R_3	0.3456	0.3205	0.0974	0.4093	0.4240	0.4540	0.3112
R_4	0.3655	0.3289	0.4247	0.1007	0.3249	0.2134	0.1017
R_5	0.3165	0.2409	0.3214	0.3256	0.0904	0.1874	0.2130
R_6	0.2749	0.3891	0.5895	0.3016	0.2058	0.0841	0.0706
R_7	0.4434	0.4919	0.3904	0.0786	0.0709	0.0914	0.0583

(本题是由东北电力学院关信提供的)

4. 1999 年美国大学生数学建模竞赛试题

(MCM-1999, 1999, 2, 5~8)

MCM-1999 问题 A 强烈的碰撞

美国国家航空和航天局 (NASA) 从过去某个时间以来一直在考虑一颗大的小行星撞击地球会产生的后果.

作为这种努力的组成部分, 要求你们队来考虑这种撞击的后果, 假如该小行星撞击到了南极洲的话, 人们关心的是撞到南极洲比撞到地球的其他地方可能会有很不同的后果.

假设小行星的直径大约为 1000 米, 还假设它正好在南极与南极洲大陆相撞.

要求你们队对这样一颗小行星的撞击提供评估. 特别是, NASA 希望有一个关于这种撞击下可能的人类人员伤亡的数量和所在地区的估计, 对南半球海洋的食物生产区域造成的破坏的估计, 以及由于南极洲极地冰岩的大量融化造成可能的沿海岸地区的洪水的估计.

(本题是由 Jack Robertson(Mathematics Department, Georgia College and State University) 提供的)

MCM-1999 问题 B 非法的集会

在许多公众设施的用于公众集会的房间里都有指示牌，指明如果在本室的人员超过指定的数目，那将是非法的。这个指定的数目可能是根据一旦有紧急情况时能从房间出口撤离的速度来确定的。类似地，在电梯和其他设施中常有“最大容量”之类的张贴告示。

试研制一个数学模型：什么数目可以作为“合法的容量”张贴在指示牌上。作为你们的求解的一部分，你们要讨论与火警或其他紧急情况不同的决定房间（或空间）中的人数为“非法”的准则。还有，你们构造的模型要考虑在诸如（带有桌、椅的）自助餐厅那样带有可移动家具的房间、体育馆、游泳池，以及有成排座位和走道的报告厅之间的差别。你们可能希望对比在各种不同的环境下——电梯、报告厅、游泳池、自助餐厅或体育馆——可能得出的结论的相似和不同之处。收集诸如摇滚乐音乐会和足球比赛那些能提出特定条件的数据。把你们的模型应用于你们学院（或邻镇）的一个或多个公众设施。试把你们的结果和这些设施所指示的容量（如果有张贴的指示的话）进行比较。如果用了后，你们的模型看来会引起想提高容量的当事人的兴趣的话，试给当地的报纸写一篇捍卫你们的分析的文章。

（本题是由 Joe Malkevitch (Mathematics Department, York College, City University of New York) 提供的）

MCM-1999 问题 C 大地污染

[注] MCM-1999 的一个新的特点是增加了第三题（C 题）；一个与数学、化学、环境科学和环境工程有关的跨学科的实际问题。参赛队可以从网上得到一个实际的污染问题的数据。参赛 C 题的队要单独报名，一个学校可以有两个队报名参赛 C 题。因此，一个学校至多可以有 6 个队报名参加 MCM-1999。

背景

若干实践中重要但理论上困难的数学问题与污染的评估有关。这种问题之一，就是根据只是在被怀疑为已污染地区的周围而不必直接在该地区中测得的很少的测量数据，来导出不易进入的地下的渗漏污染物的位置和数量、以及污染源的精确估计。

例子

数据集可通过 <http://www.comap.com/mcm/prodata.xls> 找到。

该数据集（一种超常文件（an Excel file），它能卸载到大多数电子数据表（spreadsheets））展示了从 1990 年到 1997 年在 10 个监测井处地下水中污染物的测量数据。单位是微克/升 ($\mu\text{g/l}$)。8 个测井的位置和高度是已知的井在下表给出。头两个数是在一张地图的直角格点上井的位置的坐标。第三个数是井中水面高出平均海平面的高度（以英尺计）。

井号	x-坐标(英尺计)	y-坐标(英尺计)	高度(英尺计)
MW-1	4187.5	6375.0	1482.23
MW-3	9062.5	4375.0	1387.92
MW-7	7625.0	5812.5	1400.19
MW-9	9125.0	4000.0	1384.53
MW-11	9062.5	5187.5	1394.26
MW-12	9062.5	4562.5	1388.94
MW-13	9062.5	5000.0	1394.25
MW-14	4750.0	2562.5	1412.00

数据集中另两个井(MW-27 和 MW-33)的位置和高度不知道。在该数据集中你还会看到数字后面的字母 T(Top), M

(Middle), 或 B(Bottom), 它们分别表示测量是在井的含水层的顶部、中部和底部进行的。因此，MW-7B 和 MW-7M 是来自同一个井，但分别是底部和中部的测量。此外，其他的测量数据表明水有流向该区域中的 MW-9 号井的趋势。

问题 1 试建立一个数学模型来决定在由该数据集表示的区域和时间里是否有任何新的污染物产生。若有，试识别新的污染物并估计它们的污染源的位置和时间。

问题 2 在收集任何数据之前，会提出下列问题：是否拟议中的数据类型和模型能给出关于污染物所在位置和数量的我们想要的估计。液态的化学物质会从埋置在均匀的土壤中的存储设备中许多类似的存储罐中的一个存储罐中渗漏。因为若要在许多大罐的下面去探测的费用会过分昂贵而且危险，所以只能在存储设备的边缘地区附近或在看来是更合意的地区的表面进行测量。试决定只是在整个存储罐的边界的外面或表面进行什么样类型的测量，以及测量数目可以用于一个数学模型以决定渗漏是否发生，何时发生，何处(从哪个罐)发生，以及渗漏多少液体。

(本题是由 Yves Nievergelt (Mathematics Department, Eastern Washington University) 提供的)

(叶其孝译)

为使读者能更准确地了解题意，特把英文赛题刊载如下：

1999 Mathematical Contest in Modeling Problem A—Deep Impact

For some time, the National Aeronautics and Space Administration (NASA) has been considering the consequences of a large asteroid impact on the earth.

As part of this effort, your team has been asked to consider the effects of such an impact were the asteroid to land in Antarctica.

There are concerns that an impact there could have considerably different consequences than one striking elsewhere on the planet.

You are to assume that an asteroid is on the order of 1000 m in diameter, and that it strikes the Antarctic continent directly at the South Pole.

Your team has been asked to provide an assessment of the impact of such an asteroid. In particular, NASA would like an estimate of the amount and location of likely human casualties from this impact, an estimate of the damage done to the food production regions in the oceans of the southern hemisphere, and an estimate of possible coastal flooding caused by large-scale melting of the Antarctic polar ice sheet.

Problem B—Unlawful Assembly

Many public facilities have signs in rooms used for public gatherings which state that it is “unlawful” for the rooms to be occupied by more than a specified number of people. Presumably, this number is based on the speed with which people in the room could be evacuated from the room’s exits in case of an emergency. Similarly, elevators and other facilities often have “maximum capacities” posted.

Develop a mathematical model for deciding what number to post on such a sign as being the “lawful capacity”. As part of your solution discuss criteria, other than public safety in the case of a fire or other emergency, that might govern the number of people considered “unlawful” to occupy the room (or space). Also, for

the model that you construct, consider the differences between a room with movable furniture such as a cafeteria (with tables and chairs), a gymnasium, a public swimming pool, and a lecture hall with a pattern of rows and aisles. You may wish to compare and contrast what might be done for a variety of different environments: elevator, lecture hall, swimming pool, cafeteria, or gymnasium. Gatherings such as rock concerts and soccer tournaments may present special conditions.

Apply your model to one or more public facilities at your institution (or neighboring town). Compare your results with the stated capacity, if one is posted. If used, your model is likely to be challenged by parties with interests in increasing the capacity. Write an article for the local newspaper defending your analysis.

1999 Mathematical Contest in Modeling Problem C—Ground Pollution

Background

Several practically important but theoretically difficult mathematical problems pertain to the assessment of pollution. One such problem consists in deriving accurate estimates of the location and amount of pollutants seeping inaccessibly underground, and the location of their source, on the basis of very few measurements taken only around, but not necessarily directly in, the suspected polluted region.

Example

A data set is located at: <http://www.comap.com/mcm/>

proedata.xls

The data set (an Excel file which can be downloaded into most spreadsheets) shows measurements of pollutants in underground water from 10 monitoring wells (MW) from 1990 to 1997. The units are micrograms per liter ($\mu\text{g/l}$). The location and elevation for eight of the wells is known and given below. The first two numbers are the coordinates of the location of the well on a Cartesian grid on a map. The third number is the altitude in feet above Mean Sea Level of the water level in the well.

Well Number (ft)	x-Coordinate (ft)	y-Coordinate (ft)	Elevation (ft)
MW-1	4187.5	6375.0	1482.23
MW-3	9062.5	4375.0	1387.92
MW-7	7625.0	5812.5	1400.19
MW-9	9125.0	4000.0	1384.53
MW-11	9062.5	5187.5	1394.26
MW-12	9062.5	4562.5	1388.94
MW-13	9062.5	5000.0	1394.25
MW-14	4750.0	2562.5	1412.00

The locations and elevations of the other two wells in the data set (MW-27 and MW-33) are not known. In the data set you will also see the letter T, M or B after the well number, indicating the measurements were taken at the Top, Middle, or Bottom of the aquifer in the well. Thus, MW-7B and MW-7M are from the same well, but from the bottom and from the middle. Also, other measurements indicate that water tends to flow toward

well MW-9 in this area.

Problem One

Build a mathematical model to determine whether any new pollution has begun during this time period in the area represented by the data set. If so, identify the new pollutants and estimate the location and time of their source.

Problem Two

Before the collection of any data, the question arises whether the intended type of data and model can yield the desired assessment of the location and amount of pollutants. Liquid chemicals may have leaked from one of the storage tanks among many similar tanks in a storage facility built over a homogeneous soil. Because probing under the many large tanks would be prohibitively expensive and dangerous, measuring only near the periphery of the storage facility or on the surface of the terrain seems preferable. Determine what type and number of measurements, taken only outside the boundary or on the surface of the entire storage facility, can be used in a mathematical model to determine whether a leak has occurred, when it occurred, where (from which tank) it occurred, and how much liquid has leaked.

5. 2000 年美国大学生数学建模竞赛试题 (MCM-2000, 2000, 2, 4~7)

MCM-2000 A 题 空中交通管理

——纪念美国联邦航空局前首席科学家 Robert Machol 博士

为改善安全性和降低空中交通管理人员的工作负担，美国联邦航空局(Federal Aviation Agency, 缩写为 FAA)正考虑在空中交通管理系统中增加软件，该软件将能自动侦察出飞行器飞行路线可能的冲突，并提请管理人员注意。为此，FAA 的分析员提出下列问题。

要求 A：对于给定在空中飞行的两架飞机，在何种情况下空中交通管理人员应认为飞机靠得太近而且需要干预？

要求 B：一个空域分区指的就是在整个三维空域中一位空中交通管理人员所管制的区域。给定任一空域分区，我们如何度量从空中交通工作量的角度来看该空域有多复杂？

在多大程度上复杂性是由同时飞经该空域分区的飞行器的架数决定的：

- (1) 在任何时刻？
- (2) 在任意给定的时间区间内？
- (3) 在一天中的某个特定的时间内？

在这些时间里可能出现的飞行冲突的数目是怎样影响到复杂性的？

自动预测冲突的附加软件工具的出现以及提请管理人员注意会否降低或增加这种复杂性？

除了你的报告的准则外，试写一个为你的结论辩护的概述(不超过 2 页)，以使联邦航空局的分析人员可以把它提交给联邦航空局的行政管理 Jane Garvey，为你的结论进行答辩。

MCM-2000 B 题 无线电信道的分配

我们要对在一个大的平面区域上的匀称的发射台址网络分配无线电信道进行建模以避免干扰。一种基本的方法就是把区域划分成如图 1(见本书 p. 370 Figure 1)所示的(蜂窝状)规则的六边

形，发射台址位于每个六边形的中心。

一个频谱区间被分配给各发射台的频率。该区间将被划分为间距规则的信道，并用整数 1, 2, 3, … 来表示。给每个发射台分配一个正整数信道。如果能避免来自附近发射台的干扰，那么同一个信道可用于许多台址。

我们的目的是要极小化为分配能满足某些约束的信道所需的频谱区间的宽度。这个目的是用跨度的概念来实现的。跨度就是在满足约束的所有分配中任一个台址所用到的最大信道的极小值。这并不要求在达到该跨度一种分配中小于该跨度的每个信道都被使用。

令 s 是正六边形的边长。我们集中考虑有两种干扰水平的情形。

要求 A：频率分配有几种约束。首先，同一个信道不能分配给相互间距离为 $4s$ 的两个发射台。其次，由于频谱传播的原因，对于相互间距离为 $2s$ 的发射台不能给予同一个或者相邻的信道；它们的信道至少要差 2。在这些约束下，就图 1 所示的网状区域的跨度我们能说些什么？

要求 B：重复要求 A，又假定例子中的网状区域可以在所有方向向外无限扩展。

要求 C：重复要求 A 和 B，除了更一般地假定相互间距离为 $2s$ 的发射台的信道至少要差某个给定的整数 k ，还要求相互间距离为 $4s$ 的发射台的信道仍然至少要差 1。关于跨度以及关于设计频率分配作为 k 的函数的有效策略我们能说些什么？

要求 D：考虑问题的推广，例如几个干扰水平或者不规则的发射台设置。要考虑的可能是重要的其他因素是什么？

要求 E：为当地报纸写一篇说明你们研究结果的（不超过 2 页的）文章。

ICM-2000 问题：大象：什么时候是足足有余了？

〔注〕从 2000 年起 MCM-C 题改为 ICM (Interdisciplinary Contest in Modeling, 跨学科建模竞赛)

“不幸的是，如果栖息地被大象令人不愉快地改变了，那么就应该考虑它们的迁移——即使要从象群中挑出一些进行迁移。”
National Geographic (国家地理杂志) (Earth Almanac(地球年鉴))-December 1999

南非一个巨大的国家公园里大约有 11000 头大象。管理方针要求有一个有益于健康的环境使之能保持在 11000 头大象的稳定的象群。每年公园管理人员都要计算象群的数目。在过去的 20 年间，为了保持象群数目尽可能接近 11000 头，整个象群曾被迁移过。这个过程包括射杀(大部分是这种情形)以及有时候每年重新安置大约 600 到 800 头大象到异地。

最近，出现了反对射杀这些大象的公众的强烈抗议。此外，每年重新安置即使是一小群大象也已经不再可行。然而，已经研制出了避孕针刺的方法，它能在两年中阻止母象怀孕。

以下是有该公园中的大象的一些信息：

- 几乎没有大象的迁进或迁出。
- 性别比非常接近 1 : 1，而且采取控制措施以保持相同的比例。
- 新生象崽的性别比大约也是 1 : 1。在这段时间里出生的双胞胎约占 1.35%。
- 母象第一次怀胎大约在 10 到 12 岁之间，平均每 3.5 年生一头象崽，直到 60 岁左右为止。怀孕期约为 22 个月。
- 避孕针刺会引起母象每月一次的动情期（但不会怀胎）。通常每 3.5 年大象才会有一次求偶，所以月周期的动情会造成额外的压力。
- 母象可以每年做一次避孕针刺而不会产生额外的不良影响。在上一次避孕针刺后的两年里母象不会怀孕。

- 新生象崽能活到一岁的比例在 70% 到 80% 之间。此后，在直到大约 60 岁的每个年龄段的存活率是均匀的而且是很高的（超过 95%）；大象在 70 岁之前会死掉是一个很好的假设。
- 公园里不准打猎，偷猎也是微不足道的。

公园管理部门有关于过去两年中从此地迁移出去的大象的大致的年龄和性别的粗略的数据。在以下网址有该数据：www.comap/icm/icm2000data.xls。不幸的是没有被射杀或还留在公园的大象的数据。

你们的总的的任务就是研制模型来研究避孕针刺怎样可以用于象群的控制。特别是：

- 任务 1：研制并利用模型来推测大象从 2 到 60 岁的大概的存活率。也要推测当前象群的年龄结构。
- 任务 2：试估计为使象群数目保持在 11000 头的水平每年要有多少头母象要做避孕针刺。说明由你使用的数据中的不确定性会怎样影响到你的估计。就象群的年龄结构的任何变化以及这种变化会怎样影响到旅游者发表你的意见。（你或许要前瞻大约 30~60 年。）
- 任务 3：如果每年重新安置 50 到 300 头大象是可行的话，那么将会怎样影响到要进行避孕针刺的母象的数目？就针刺和重新安置之间的协调使用提出你的意见。
- 任务 4：一些避孕针刺的反对者争辩说如果发生大量大象的突然损失（由于疾病或无法控制的偷猎），即使立即停止避孕针刺象群重新增长的能力也会严重受阻。试研究并应答这种担忧。
- 任务 5：公园管理部门对建模抱怀疑态度。特别是，他们争辩说缺乏完整的数据使得利用模型来指导他们的决策的任何企图都将贻笑大方。除了你们的技术报告外，还要包括一个明确为公园管理部门精心写出的报告（最多 3 页）来回答他们的担忧并提出建议。还要提出能提高公园管

理人员对你们的模型和结论的信心的方法。

任务 6：如果你们的模型能付诸应用，那么非洲其他的大象公园会有兴趣用你们的模型。试为各种大小的公园（300~25000 头大象）准备一份避孕针刺计划，同时要考虑稍有不同的存活率和搬运的可能性。

（叶其孝译）

为使读者能更准确地了解题意，特把英文赛题刊载如下：

Problem A Air Traffic Control

Dedicated to the memory of Dr. Robert Machol, former chief scientist of the Federal Aviation Agency

To improve safety and reduce air traffic controller workload, the Federal Aviation Agency (FAA) is considering adding software to the air traffic control system that would automatically detect potential aircraft flight path conflicts and alert the controller. To that end, an analyst at the FAA has posed the following problems.

Requirement A: Given two airplanes flying in space, when should the air traffic controller consider the objects to be too close and to require intervention?

Requirement B: An airspace sector is the section of three-dimensional airspace that one air traffic controller controls. Given any airspace sector, how do we measure how complex it is from an air traffic workload perspective? To what extent is complexity determined by the number of aircraft simultaneously passing through that sector

(1) at any one instant?

- (2) during any given interval of time?
- (3) during a particular time of day?

How does the number of potential conflicts arising during those periods affect complexity? Does the presence of additional software tools to automatically predict conflicts and alert the controller reduce or add to this complexity?

In addition to the guidelines for your report, write a summary (no more than two pages) that the FAA analyst can present to Jane Garvey, the FAA Administrator, to defend your conclusions.

Problem B

Radio Channel Assignments

We seek to model the assignment of radio channels to a symmetric network of transmitter locations over a large planar area, so as to avoid interference. One basic approach is to partition the region into regular hexagons in a grid (honeycomb-style), as shown in Figure 1, where a transmitter is located at the center of each hexagon.

An interval of the frequency spectrum is to be allotted for transmitter frequencies. The interval will be divided into regularly spaced channels, which we represent by integers 1, 2, 3, Each transmitter will be assigned one positive integer channel. The same channel can be used at many locations, provided that interference from nearby transmitters is avoided.

Our goal is to minimize the width of the interval in the frequency spectrum that is needed to assign channels subject to some constraints. This is achieved with the concept of a span. The span

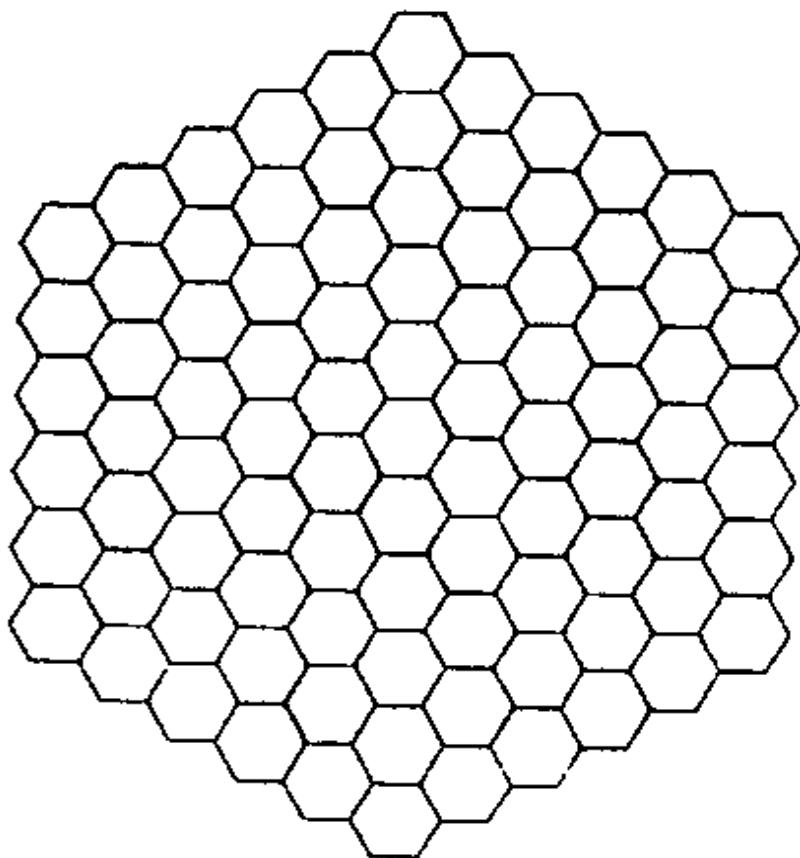


Figure 1

is the minimum, over all assignments satisfying the constraints, of the largest channel used at any location. It is not required that every channel smaller than the span be used in an assignment that attains the span.

Let's be the length of a side of one of the hexagons. We concentrate on the case that there are two levels of interference.

Requirement A: There are several constraints on frequency assignments. First, no two transmitters within distance $4s$ of each other can be given the same channel. Second, due to spectral spreading, transmitters within distance $2s$ of each other must not be given the same or adjacent channels; Their channels must differ by at least 2. Under these constraints, what can we say

about the span in Figure I?

Requirement B: Repeat Requirement A, assuming the grid in the example spreads arbitrarily far in all directions.

Requirement C: Repeat Requirements A and B, except assume now more generally that channels for transmitters within distance $2s$ differ by at least some given integer k , while those at distance at most $4s$ must still differ by at least one. What can we say about the span and about efficient strategies for designing assignments, as a function of k ?

Requirement D: Consider generalizations of the problem, such as several levels of interference or irregular transmitter placements. What other factors may be important to consider?

Requirement E: Write an article (no more than 2 pages) for the local newspaper explaining your findings.

Interdisciplinary Contest in Modeling Elephants: When is Enough, Enough?

"Ultimately, if a habitat is undesirably changed by elephants, then their removal should be considered even by culling." *National Geographic* (Earth Almanac) — December 1999

A large National Park in South Africa contains approximately 11,000 elephants. Management policy requires a healthy environment that can maintain a stable herd of 11,000 elephants. Each year park rangers count the elephant population. During the past 20 years whole herds have been removed to keep the population as close to 11,000 as possible. This process involved shooting (for the most part) and occasionally relocating approximately 600 to 800 elephants per year.

Recently, there has been a public outcry against the shooting of these elephants. In addition, it is no longer feasible to relocate even a small population of elephants each year. A contraceptive dart, however, has been developed that can prevent a mature elephant cow from conceiving for a period of two years.

Here is some information about the elephants in the Park:

- There is very little emigration or immigration of elephants.
- The gender ratio is very close to 1 : 1 and control measures have endeavored to maintain parity.
- The gender ratio of newborn calves is also about 1 : 1. Twins are born about 1.35% of the time.
- Cows first conceive between the ages of 10 and 12 and produce, on average, a calf every 3.5 years until they reach an age of about 60. Gestation is approximately 22 months.
- The contraceptive dart causes an elephant cow to come into oestrus every month (but not conceiving). Elephants usually have courtship only once in 3.5 years, so the monthly cycle can cause additional stress.
- A cow can be darted every year without additional detrimental effects. A mature elephant cow will not be able to conceive for 2 years after the last darting.
- Between 70% and 80% of newborn calves survive to age 1 year. Thereafter, the survival rate is uniform across all ages and is very high (over 95%), until about age 60; it is a good assumption that elephants die before reaching age 70.
- There is no hunting and negligible poaching in the Park.

The park management has a rough data file of the approximate ages and gender of the elephants they have transported out of the region during the past 2 years. This data is available on website: www.comap.com/icm/icm2000data.xls. Unfortunately no data is available for the elephants that have been shot or remain in the Park.

Your overall task is to develop and use models to investigate how the contraceptive dart might be used for population control. Specifically,

Task 1: Develop and use a model to speculate about the likely survival rate for elephants aged 2 to 60. Also speculate about the current age structure of the elephant population.

Task 2: Estimate how many cows would need to be darted each year to keep the population fixed at approximately 11, 000 elephants. Show how the uncertainty in the data at your disposal affects your estimate. Comment on any changes in the age structure of the population and how this might affect tourists. (You may want to look ahead about 30—60 years.)

Task 3: If it were feasible to relocate between 50 and 300 elephants per year, how would this reduce the number of elephants to be darted? Comment on the trade-off between darting and relocation.

Task 4: Some opponents of darting argue that if there were a sudden loss of a large number of elephants (due to disease or uncontrolled poaching), even if darting stopped immediately, the ability of the population to grow again would be seriously impeded. Investigate and respond to this concern.

Task 5: The management in the Park is skeptical about modeling. In particular, they argue that a lack of complete data

makes a mockery of any attempt to use models to guide their decisions. In addition to your technical report, include a carefully crafted report (3-page maximum) written explicitly for the park management that responds to their concerns and provides advice. Also suggest ways to increase the park managers confidence in your model and in your conclusions.

Task 6: If your model works, other elephant parks in Africa would be interested in using it. Prepare a darting plan for parks of various sizes (300—25,000 elephants), with slightly different survival rates and transportation possibilities.

附录Ⅱ 20世纪最好的10个算法

Barry A. Cipra

Algos 是希腊字，意思是“疼”，*Algor* 是拉丁字，意思是“冷却”。这两个字都不是 *algorithm*(算法)一词的词根，*algorithm* 一词却与 9 世纪的阿拉伯学者 al-Khwarizmi 有关，他写的书《al-jabr w'al muqabalah(代数学)》演变成为现在中学的代数教科书。[译注：al-Khwarizmi(阿尔柯瓦利兹米)于大约 825 年写的该书，有关介绍可参阅《古今数学思想》，第一册，p. 218 ~ 220，上海科学技术出版社，1979；《数学百科全书》(第 1 卷)，p. 69，科学出版社，1994.] al-Khwarizmi 强调求解问题的有条理的步骤。如果他能活到今天的话，他一定会被以他的名字而得名的方法的进展所感动。

由美国物理协会(American Institute of Physics)和美国电气和电子工程师学会(IEEE)的计算机协会联合出版的 *Computing in Science & Engineering* (CiSE)，2000 1 月/2 月这一期把注意力集中于计算机时代的一些最好的算法。客座主编田纳西大学和橡树岭国家实验室的 Jack Dongarra 和国防分析研究所计算科学中心的 Francis Sullivan 把他们称为的“本世纪最好的 10 个算法”放在一起列了一张表^①。

Dongarra 和 Sullivan 写道：“我们试图收集 20 世纪中对科学和工程的研究和实践影响最大的 10 个算法”。他们承认任何 10 个的最好的算法，入选或者不入选，一定会有争议的。当涉及到挑选算法最好时，似乎就没有最好的算法了。

① 原题：The Best of the 20th Century: Editors Name Top 10 Algorithms, 译自 SIAM NEWS, v. 33(2000), no. 4(May), pp. 1, 3.

不啰嗦了，以下就是 CiSE 列出的按年代次序排列的最好的 10 个算法。（和算法一起列出的日期和人名应看做是一阶近似。大多数算法是经过很多时间经过许多人的贡献才形成的）

1946：在洛斯阿拉莫斯科学实验室工作的 John von Neumann, Stan Ulam 和 Nick Metropolis 编制了 Metropolis 算法，也称为 Monte Carlo 方法。

Metropolis 算法旨在通过模仿随机过程来得到具有难以控制的大量的自由度的数值问题和具有阶乘规模的组合问题的近似解法。以数字计算机对确定性计算的盛名为前提，说生成随机数是该算法最早的应用之一是恰当的。

1947：兰德公司的 George Dantzig 创造了线性规划的单纯形方法。

就其广泛的应用而言，Dantzig 的算法一直是最成功的算法之一。线性规划对于那些要想在经济上站住脚而有赖于是否具有在预算和其他约束条件下最优化的能力的工业界有着决定性的影响。（当然，工业中的“实际”问题往往是非线性的；使用线性规划有时候是由估算的预算促成的。）单纯形法是一种能达到最优解的精细的方法。尽管理论上讲其效果是（译注：变量个数的增加）指数衰减的，但在实践中该算法是高度有效的——它本身说明了有关计算的本质的一些有趣的事情。

1950：来自美国国家标准局的数值分析研究所的 Magnus Hestenes, Eduard Stiefel 和 Cornelius Lanczos 开创了Krylov 子空间迭代法的研制。

这些算法处理看似简单的求解形如 $Ax = b$ 的方程的问题。当然隐藏的困难在于 A 是一个巨型的 $n \times n$ 矩阵，致使代数解 $x = b/A$ 是不容易计算的（确实，矩阵的“相除”不是一个实际上有用

的概念) 迭代法——诸如求解形为 $Kx_{i+1} = Kx_i + b - Ax_i$ 的方程, 其中 K 是一个理想地“接近” A 的较为简单的矩阵——导致了 Krylov 子空间的研究。以俄罗斯数学家 Nikolai Krylov 命名的 Krylov 子空间是由作用在初始“余量”向量 $r_0 = b - Ax_0$ 上的矩阵幂张成的。当 A 是对称矩阵时, Lanczos 找到了一种生成这种子空间的正交基的极好的方法。对于对称正定的方程组, Hestenes 和 Stiefel 提出了称为共轭梯度法的甚至更妙的方法。过去的 50 年中, 许多研究人员改进并扩展了这些算法。当前的一套方法包括非对称方程组的求解技巧, 像字首缩拼词为 GMRES 和 Bi-CGSTAB 那样的算法(GMRES 和 Bi-CGSTAB 分别首次出现于 1986 和 1992 *SIAM Journal on Scientific and Statistical Computing*(美国工业与应用数学学会的科学和统计计算杂志)上)

1951: 橡树岭国家实验室的 Alston Householder 系统阐述了矩阵计算的分解方法。

研究证明能把矩阵因子分解为三角、对角、正交和其他特殊形式的矩阵是极其有用的。这种分解方法使软件研究人员能生产出灵活有效的矩阵软件包。这也促进了数值线性代数中反复出现的大问题之一的舍人误差分析问题。(1961 年伦敦国家物理实验室的 James Wilkinson 基于把矩阵分解为下和上三角矩阵因子的积的 LU 分解, 在美国计算机协会(ACM)的杂志上发表了一篇题为“矩阵逆的直接方法的误差分析”的重要文章。)

1957: John Backus 在 IBM 领导一个小组研制 Fortran 最优编译程序。

Fortran 的创造可能是计算机编程历史上独一无二的最重要的事件: 科学家(和其他人)终于可以无需依靠像地狱那样可怕的机器代码就可告诉计算机他们想要做什么。虽然现代编译程序的

标准并不过分——Fortran I 只包含 23500 条汇编语言指令——早期的编译程序仍然能完成令人吃惊的复杂计算。就像 Backus 本人在 1998 年在 *IEEE Annals of the History of Computing* 发表的有关 Fortran I, II, III 的近代历史的文章中回忆道：编译程序“所产生的如此有效的代码使得其输出令研究它的编程人员都感到吓了一跳。”

1959~1961：伦敦 Ferranti Ltd. 的 J. G. F. Francis 找到了一种称为QR 算法的计算本征值的稳定的方法。

本征值大概是和矩阵相连在一起的最重要的数了——而且可能是最需要技巧来计算它们。把一个正方矩阵转换成一个“几乎是”上三角的矩阵——意即在紧挨着矩阵主对角线下面的——斜列上可能有非零元素——是相对容易的。但要想不产生大量的误差就把这些非零元素消去就不是平凡的事了。QR 算法正好是能达到这一目的的方法。基于 QR 分解， A 可以写成一个正交矩阵 Q 和一个上三角矩阵 R 的乘积，这种方法迭代地把 $A_i = Q_i R_i$ 变成 $A_{i+1} = R_i Q_i$ ，就加速收敛到上三角矩阵而言多少有点不能指望。20 世纪 60 年代中期 QR 算法把一度难以对付的本征值问题变成了例行程序的计算。

1962：伦敦 Elliott Brothers, Ltd. 的 Tony Hoare 提出了快速(按大小)分类法。

把 n 个事物按数或字母的次序排列起来在心智上是不会有什么触动的单调平凡的事。智力的挑战在于发明一种快速完成排序的方法。Hoare 的算法利用了古老的分割开和控制的递归策略来解决问题：挑一个元素作为“主元”，把其余的元素分成“大的”和“小的”两堆（当和主元比较时），再在每一堆中重复这一过程。尽管可能要做受到严厉责备的做完全部 $N(N-1)/2$ 次的比较（特别是，如果你把主元作为早已按大小分类好的表列的第一个元素

的话!)快速分类法运行的平均次数具有 $O(N \log N)$ 的有效性. 其优美的简洁性使之成为计算复杂性的著名的例子.

1965: IBM T. J. Watson 研究中心的 James Cooley 以及普林斯顿大学和 AT & T 贝尔实验室的 John Tukey 向公众透露了 快速 Fourier 变换(方法)(FFT).

应用数学中意义最深远的算法无疑是使信号处理发生突破性进展的 FFT. 其基本思想要追溯到 Gauss(他需要计算小行星的轨道), 但是 Cooley-Tukey 的论文弄清楚了 Fourier 变换计算起来有多容易. 就像快速分类法一样 FFT 有赖于用分割开和控制的策略把表面上令人讨厌的 $O(N^2)$ 降到令人欢乐的 $O(N \log N)$. 但是不像快速分类法, 其执行(初一看)是非直观的而且不那么直接. 其本身就给计算机科学一种推动力去研究计算问题和算法的固有复杂性.

1977: Brigham Young(杨伯翰)大学的 Helaman Ferguson 和 Rodney Forcade 提出了 整数关系侦查算法.

这是一个古老的问题: 给定一组实数, 例如说 x_1, x_2, \dots, x_n , 是否存在整数 a_1, a_2, \dots, a_n (不全为零), 使得 $a_1x_1 + a_2x_2 + \dots + a_nx_n = 0$? 对于 $n=2$, 历史悠久的欧几里得算法能做这项工作, 计算 x_1/x_2 的连分数展开中的各项. 如果 x_1/x_2 是有理数, 展开会终止, 在适当解开后就给出了“最小的”整数 a_1 和 a_2 . 如果欧几里得算法不终止——或者如果你只是简单地由于厌倦计算——那么解开的过程至少提供了最小整数关系的大小的下界. Ferguson 和 Forcade 的推广更有威力, 尽管这种推广更难于执行(和理解). 例如, 他们的侦查算法被用来求得逻辑斯谛(logistic)映射的第三和第四个分歧点, $B_3 = 3.544090$ 和 $B_4 = 3.564407$ 所满足的多项式的精确系数. (后者是 120 阶的多项式, 它的最大的系数是 257^{30} .) 已证明该算法在简化量子场论中的

Feynman 图的计算中是有用的.

1987: 耶鲁大学的 Leslie Greengard 和 Vladimir Rokhlin 发明了 快速多极算法.

该算法 N 体模拟中最头疼的困难之一：经由引力或静电力相互作用的 N 个粒子运动的精确计算（想像一下银河系中的星体，或者蛋白质中的原子）看来需要 $O(N^2)$ 的计算量——比较每一对质点需要一次计算。该算法利用多极展开（净电荷或质量、偶极矩、四矩，等等）来近似遥远的一组质点对当地一组质点的影响。空间的层次分解用来确定当距离增大时比以往任何时候都更大的质点组。快速多极算法的一个明显优点是具有严格的误差估计，这是许多算法所缺少的性质。

21 世纪将会带来什么样的新的洞察和算法？对于又一个一百年完整的回答显然是不知道的。然而，有一点似乎是肯定的，正如 Sullivan 在最好的 10 个算法一文的介绍中所写的，“新世纪对我们来说既不会是很宁静的，也不会是弱智的。”

（叶其孝译，吴庆宝校）