

2008.04.

创刊 第01期



刊首寄语

当这份酝酿了 4 个月的杂志，终于出现在我的面前，我一下想到的是：布尔巴基学派的事业是由巴黎高师的一群年轻人开始做起来的以及清华园的算学系，曾经是群星璀璨。

我希望，这份立足于本科生的数学杂志不仅限于展示大家学习过程中做出的题目，更着眼于以开放独立的态度去学习数学，思考数学，讨论数学，乃至进一步创造。

数学中，有自由灵动，也有简朴凝重。风格不同，可同为精妙。做数学不同于应付考试。数学史上有领悟力创造力极高的天外来客，也有如郭靖之类的一代宗师。所以，每个热爱数学，渴望用数学这门上帝的语言来解释世界的同学，都不妨来给自己一个机会，试着与“好”数学亲密接触，看自己能走多远。我们要逐渐培养起自己的数学审美观，不做人云亦云的八哥；我们要试图建立自己的数学风格，形成自己的一技乃至多技之长。这份杂志即可成为我们的练武场。

我们已经看到，绝大多数“好”数学都是相互关联的。喜欢不同分支的同学须有整体的观念，浚泉源，固根本，共同探讨，互相启发。同时，有志于其他学科而又热爱数学的同学需知，只有领悟了数学中的招数，才可能独立而不是跟风地把数学应用到各自学科中。所以，所有挚爱数学的同学们，请来以这份杂志为媒介，以良友为伍，多多谈数学，不做好龙的叶公，而能志存高远，取法乎上，共同努力透过现代数学的技术外壳，看到其后的本质内核，乃至水到渠成，走上独立科研的道路，令清华园的夜空再一次群星璀璨。

最后，我愿这份诞生于春天的杂志，如轻舟一叶，驶过险峰万重，召集沿途同志，一路驶向大海，化成雄伟的舰队，创造属于自己的历史。

马辉
2008 年 4 月

目录

发刊词 1

名篇选读

数学的用场与发展 华罗庚 2

研究探讨

Maxwell Electromagnetic Theory from a Viewpoint of Differential Forms	杜升华 郝诚 胡悦科 惠瑜铭 石权 王力 吴玉清 9
A Brief Introduction to Yang-Mills' Theory	周杰 蔡稼途 22
Cauchy 问题 $\frac{dy}{dx} = x^2 + y^2, y(0) = 0$ 的最大存在区间	李超 31
关于李超同学作业的评注	杨利军 36
机票预售价格和策略的数学模型	吴玉清 孟昭时 郭晓江 39
有关等度连续函数序列一致有界性的问题	杜升华 57
$\sum_{n=1}^{\infty} \frac{n^k}{n!}$ 是 e 的整数倍的证明	刘立达 60

数学争鸣

三次数学危机中的问题猎手	杜升华 62
感悟科研	杜升华 67

人物与传记

天才与愚蠢 — 埃瓦里斯 · 伽罗瓦的悲剧故事 70

清华数学人

在孙念增数学分析奖学金颁奖会上的发言	吴昊 73
对卢旭光老师的访谈	吴昊 75

《荷思》发刊词

伴随着岁月的脚步，我们的数学系也走过了她的第九九八十一一个年头。由我们数学系本科生自主创办的学术刊物《荷思》终于来到了亲爱的读者们面前。

本刊主要以清华大学本科生数学研究各类成果为内容，兼顾介绍国内外研究之进展及动态，古今数学之思想与历史等等；同时也会包括数学家的趣闻轶事，数学谜语之类趣味性内容。

长时间以来，同学们在数学的学习上也许还不够积极主动，甚至数学系的不少同学都只停留在课堂上听教师讲授，课下自己读课本、做习题的阶段。客观上同学们彼此交流的机会相对则比较少，自主钻研的意识较为淡薄。很多同学，无论是否以数学为专业，在这样的学习气氛之中，往往错过了研习数学的乐趣，也没有充分地体会到数学之优雅与动人——仅仅将其当作一门课程来应付，实在可惜。另一方面，在我们钟灵毓秀的清华园中，并不乏在数学上勤奋好学、善于思考的同学，他们将自己学习数学的心得体会，或是点滴所得整理成篇。不过同样令人惋惜的是，他们写出的文稿往往仅于三五好友之间流传，甚至仅仅当作自娱自乐。而如何能使得这种情形有所改善？恐怕为大家建立一个有效的交流平台或学习园地是其中必不可少的一环。

在这种使命感的鞭策之下，本刊应运而生。本科生所写的数学文章往往难以发表在正规刊物上，但若能够将它们集中起来，作为我们日常学习的辅助材料则难能可贵。我们办刊的宗旨乃是传播数学文化知识、交换数学学习心得，启迪创造性思维。希望此刊物能够激发起大家对数学的兴趣与热情，增强同学之间的学术气氛，并且为广大清华师生，尤其是本科阶段的同学们提供一个在数学上交流、讨论、学习的平台。

热切的盼望着我们的刊物能够对大家的数学学习切实有所帮助，成为我们在数学研究领域初显身手，走出清华、迈向世界的起点，进而也为清华大学数学的复兴尽一份绵薄之力。虽然文章内容和版式设计都还稚嫩，但翻阅着第一期《荷思》，我们不禁想到清华园秀美的景物，想到课堂上老师的谆谆教诲，想到脑海中动人心魄的数学；而此时此刻，荷塘边，春风里，睿智的你，又在思考着什么？

我们深知要将这一刊物办好，单凭自身的热情与努力是不够的。更需要的是广大读者朋友们的支持。数学之美是永恒的，同时她更是属于我们每一个人。我们竭诚欢迎同学老师们不吝赐稿，或向我们提出宝贵意见及建议。

谢谢大家的支持和帮助！

《荷思》编辑部

2008-04

数学的用场与发展¹

华罗庚

编者按：华罗庚（1910—1985）先生指出：“宇宙之大，粒子之微，火箭之速，化工之巧，地球之变，生物之谜，日用之繁，无处不用数学。”他以一个大数学家对数学的深刻理解和一个爱国者对生产实践的关心，在文章中从上述诸方面举了大量生动的例子，广泛而深入地论述了数学在各行各业的巨大作用，高屋建瓴，很有大师风范。本文虽是1978年的文章，但作为应用数学方面科普文章的经典之作，对我们今天的数学系学生仍有很强的现实意义：从中我们能体会到广博的科学文化知识对于数学工作者的重要性，以及一位年近古稀的老人热爱科学、关心人民、报效祖国的情怀。

数与量

数（读作shù）起源于数（读作shǔ），如一、二、三、四、五……，一个、两个、三个……。量（读作liàng）起源于量（读作liáng）。先取一个单位作标准，然后一个单位一个单位地量。天下虽有各种不同的量（各种不同的量的单位如尺、斤、斗、秒、伏特、欧姆和卡路里等等），但都必须通过数才能确切地把实际的情况表达出来。所以“数”是各种各样不同量的共性，必须通过它才能比较量的多寡，才能说明量的变化。

“量”是贯穿到一切科学领域之内的，因此数学的用处也就渗透到一切科学领域之中。凡是研究量、量的关系、量的变化、量的关系的变化、量的变化的关系的时候，就少不了数学。不仅如此，量的变化还有变化，而这种变化一般也是用量来刻画的。例如，速度是用来描写物体的变化的动态的，而加速度则是用来刻画速度的变化。量与量之间有各种各样的关系，各种各样不同的关系之间还可能有关系。为数众多的关系还有主从之分——也就是说，可以从一些关系推导出另一些关系来。所以数学还研究变化的变化，关系的关系，共性的共性，循环往复，逐步提高，以至无穷。

数学是一切科学得力的助手和工具。它有时由于其它科学的促进而发展，有时也先走一步，领先发展，然后再获得应用。任何一门科学缺少了数学这一项工具便不能确切地刻画出客观事物变化的状态，更不能从已知数据推出未知的数据来，因而就减少了科学预见的可能性，或者减弱了科学预见的精确度。

思格斯说：“纯数学的对象是现实世界的空间形式和数量关系。”数学是从物理模型抽象出来的，它包括数与形两方面的内容。以上只提要地讲了数量关系，现在我们结合宇宙之大来说明空间形式。

宇宙之大

宇宙之大，宇宙的形态，也只有通过数学才能说得明白。天圆地方之说，就是古代人民用几何形态来描绘客观宇宙的尝试。这种“苍天如圆盖，陆地如棋局”的宇宙形态的模型，后来被航海家用事实给以否定了。但是，我国从理论上对这一模型提出的怀疑要早得多，并且也同样的有力。论点是：“混沌初开，乾坤始奠。气之轻清，上浮者为天；气之重浊，下凝者为地。”但不知轻清之外，又有何物？也就是圆盖之外，又有何物？三十三天之上又是何处？要想解决这样的问题，就必须借助于数学的空间形式的研究。

四维空间听来好像有些神秘，其实早已有之。即以“宇宙”二字来说，“往古来今谓之宙，四方上下谓之宇”（《淮南子·齐俗训》）。就是“宇”是东西、南北、上下三维扩展的空间，而

¹本文曾以《大哉数学之为用》为题于一九五九年五月二十八日发表在《人民日报》上，后以《数学的用场与发展》为题转载在《现代科学技术简介》（科学出版社，一九七八年）上。

“宙”是一维的时间。牛顿时代对宇宙的认识也是如此。宇宙是一个无边无际的三维空间，而一切的日月星辰都安排在这框架中运动。找出这些星体的运动规律是牛顿的一大发明，也是物理模型促进数学方法，而数学方法则是用来说明物理现象的一个好典范。由于物体的运动不是等加速度，要描绘不是等加速度，就不得不考虑速度时时在变化的情况，于是乎微商出现了。这是刻画加速度的好工具。由牛顿当年一身而二任焉，既创造了新工具——微积分，又发现了万有引力定律。有了这些，宇宙间一切星辰的运动初步统一地被解释了。行星凭什么以椭圆轨道绕日而行的，何时以怎样的速度达到何处等，都可以算出来了。

有人说西方文明之飞速发展是由于欧几里得几何的推理方法和进行系统实验的方法。牛顿的工作也是逻辑推理的一个典型。他用简单的几条定律推出整个的力学系统，大至解释天体的运行，小到造房、修桥、杠杆、称物都行。但是人们在认识自然界时建立的理论总是不会一劳永逸完美无缺的，牛顿力学不能解释的问题还是有的。用它解释了行星绕日公转，但行星自转又如何解释呢？地球自转一天二十四小时有昼有夜，水星自转周期和公转一样，半面永远白天，半面永远黑夜。一个有名的问题：水星进动每百年 $42''$ ，是牛顿力学无法解释的。

爱因斯坦不再把“宇”、“宙”分开来看，也就是时间也在进行着。每一瞬间三维空间中的物质在占有它一定的位置。他根据麦克斯韦-洛伦兹的光速不变假定，并继承了牛顿的相对性原理而提出了狭义相对论。狭义相对论中的洛伦兹变换把时空联系在一起，当然并不是消灭了时空特点。如向东走三里，再向西走三里，就回到原处。但时间则不然，共用了走六里的时间。时间是一去不复返地流逝着。值得指出的是有人推算出狭义相对论不但不能解释水星进动问题，而且算出的结果是“退动”。这是误解。我们能算出进动 $28''$ ，即客观数的三分之二。另外，有了深刻的分析，反而能够浅出，连微积分都不要用，并且在较少的假定下，就可以推出爱因斯坦狭义相对论的全部结果。

爱因斯坦进一步把时、空、物质联系在一起，提出了广义相对论，用它可以算出水星进动是 $43''$ 。这是支持广义相对论的一个有力证据。由于证据还不多，因此对广义相对论还有不少看法，但它的建立有赖于数学上的先行一步，如先有了黎曼几何。另一方面它也给数学提出了好些到现在还没有解决的问题。对宇宙的认识还将有多么大的进展，我不知道，但可以说，每一步都是离不开数学这个工具的。

粒子之微

佛经上有所谓“金粟世界”，也就是一粒粟米也可以看作一个世界。这当然是佛家的幻想。但是我们今天所研究的原子却远远地小于一粒粟米，而其中的复杂性却不亚于一个太阳系。

即使研究这样小的原子核的结构也还是少不了数学。描述原子核内各种基本粒子的运动更是少不了数学。能不能用处理普遍世界的方法来处理核子内部的问题呢？情况不同了。在这里，牛顿的力学，爱因斯坦的相对论都遇到了困难。在目前人们应用了另一套数学工具，如算子论、群表示论、广义函数论等。这些工具都是近代的产物。即使如此，也还是不能完整地说明它。

在物质结构上不管分子论、原子论也好，或近代的核子结构、基本粒子的互变也好，物理科学上虽然经过了多次的概念革新，但自始至终都和数学分不开。不但今天，就是将来，也有一点是可以肯定的，就是一定还要用数学。

是否有一个统一的处理方法，把宏观世界和微观世界统一在一个理论之中，把四种作用力统一在一个理论之中，这是物理学家当前的重大问题之一。不管将来他们怎样解决这个问题，但是在处理这些问题的数学方法必须统一。必须有一套既可以解释宏观世界又可以解释微观世界的数学工具。数学一定和物理学刚开始的时候一样，是物理科学的助手和工具。在这样的大问题的解决过程中，也可能如牛顿同时发展天体力学和发明微积分那样，促进数学的新分支的创造和形成。

火箭之速

在今天用“一日千里”来形容慢则可，来形容快则不可了。人类所创造的物体的速度远远地超过了“一日千里”。飞机虽快到日行万里不夜，但和宇宙速度比较，也显得缓慢得很。古代所幻想的“朝昆仑而暮苍梧”，在今天已不足为奇。

不妨回忆一下，在星际航行的开端——由诗一般的幻想进入科学现实的第一步，就是和数学分不开的。早在牛顿时代就算出了每秒钟近八公里的第一宇宙速度，这给科学技术工作者指出了奋斗目标。如果能够达到这一速度，就可以发射地球卫星。一九七〇年我国发射了第一颗人造卫星。数学工作者自始至终都参与这一工作（当然，其中不少工作者不是以数学工作者见称，而是运用数学工具者）。作为人造行星环绕太阳运行所必须具有的速度是11.2公里/秒，称为第二宇宙速度；脱离太阳系飞向恒星际空间所必须具有的速度是16.7公里/秒，称为第三宇宙速度。这样的目标，也将会逐步去实现。

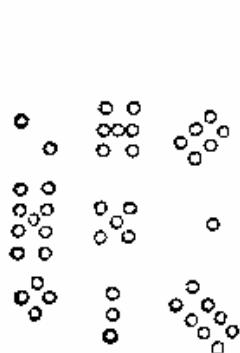


图 1

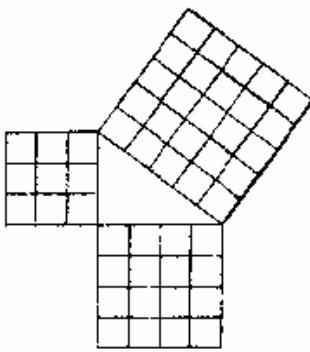


图 2

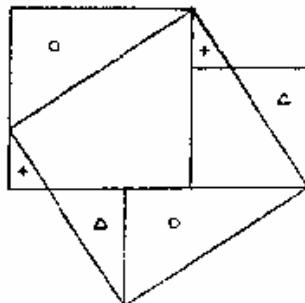


图 3

顺便提一下，如果我们宇宙航船到了一个星球上，那儿也有如我们人类一样高级的生物存在，我们用什么东西作为我们之间的媒介？带幅画去吧，那边风景殊，不了解；带一段录音去吧，也不能沟通。我看最好带两个图形去：一个“数”一个“数形关系”（勾股定理）（图1和图2）。

为了使那里较高级的生物知道我们会几何证明，还可送去上面的图形，即“青出朱入图”（图3）。这些都是我国古代数学史上的成就。

化工之巧

化学工业制造出的千千万万种新产品，使人类的物质生活更加丰富多彩，真是“巧夺天工”，“巧夺造化之工”。在制造过程中，它的化合与分解方式是用化学方程来描述的，但它是在变化的。因此，伟大革命导师恩格斯明确指出：“表示物体的分子组合的一切化学方程式，就形式来说是微分方程式。但是这些方程式实际上已经由于其中所表示的原子量而积分起来了。化学所计算的正是量的相互关系为已知的微分。”

为了形象化地说明，例如，某种物质中含有硫，用苯提取硫。苯吸取硫有一定的饱含量，在这个过程中，苯含硫越多越难再吸取硫，剩下的硫越少越难被苯吸取。这个过程时刻都在变化，吸收过程速度在不断减慢着。实验本身便是这个过程的积分过程，它的数学表达形式就是微分方程式及其求解。简单易作的过程我们可以用实验去解决，但对于复杂、难作的过程，则常常需要用数学手段来加以解决。特别是选取最优过程的工艺，数学手段更成为必不可少的手段。特别是量子化学的发展，使得化学研究提高到量子力学的阶段，数学手段——微分方程及矩阵、图论更是必需的数学工具。

应用了数学方法还可使化学理论问题得到极大的简化。例如，对于共轭分子的能级计算，

在共轭分子增大时十分困难。应用了分子轨道的图形理论，由图形来简化计算，取得了十分直观和易行的效果，便是一例，其主要根据是如果一个行列式中的元素为 0 的多，那就可以用图论来简化计算。

地球之变

我们所生活的地球处于多变的状态之中，从高层的大气，到中层的海洋，下到地壳，深入地心，都在剧烈地运动着，而这些运动规律的研究也都用到数学。

大气环流，风云雨雪，天天需要研究和预报，使得农民可以安排田间农活，空中交通运输可以安排航程。飓风等灾害性天气的预报，使得海军、渔民和沿海地区能够及早预防，减少损害。而所有这些预报都离不了数学。

“风乍起，吹皱一池春水。”风和水的关系自古便有记述，“无风不起浪”。但是风和浪的具体关系的研究，则是近代才逐步弄清的，而在风与浪的关系中用到了数学的工具，例如偏微分方程的间断解的问题。

大地每年有上百万次的地震，小的人感觉不到，大的如果发生在人烟稀少的地区，也不成大灾。但是每年也有几次在人口众多的地区的大震，形成大灾。对地壳运动的研究，对地震的预报，以及将来进一步对地震的控制都离不开数学工具。

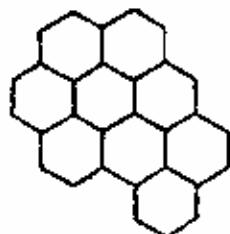


图 4

生物之谜

生物学中有许许多多的数学问题。蜜蜂的蜂房为什么要象如下的形式（图 4），一面看是正六角形，另一面也是如此。但蜂房并不是六棱柱，而它的底部是由三个菱形所拼成的。图 5 是蜂房的立体图。这个图比较清楚，更具体些，拿一支六棱柱的铅笔未削之前，铅笔一端形状是 ABCDEF 正六角形（图 6）。通过 AC，一刀切下一角，把三角形 ABC 搬置 AOC 处。过 AE，CE 也如此同样切三刀，所堆成的形

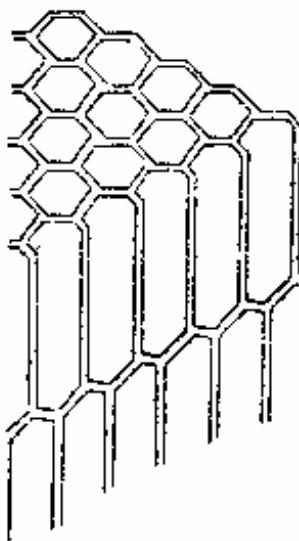


图 5

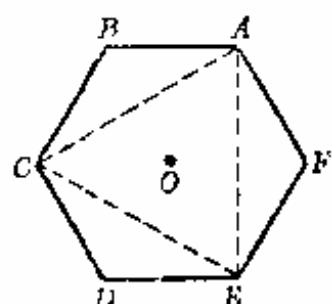


图 6

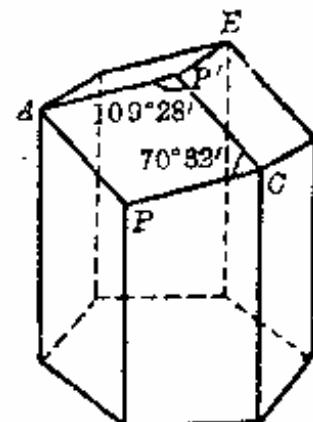


图 7

状就是图 7，而蜂巢就是两排这样的蜂房底部和底部相接而成。

关于这个问题有一段趣史：巴黎科学院院士数学家克尼格，从理论上计算，为使消耗材料最少，菱形的两个角度应该是 $109^{\circ} 26'$ 和 $70^{\circ} 34'$ 。与实际蜜蜂所做出的仅相差 2 分。后来苏格兰数学家马克劳林重新计算，发现错了的不是小小的蜜蜂，而是巴黎科学院的院士，因克尼格用的对数表上刚好错了一个字。这十八世纪的难题，一九六四年我用它来考过高中生，不

少高中生提出了各种各样的证明。

这一问题，我写得篇幅略长些，目的在于引出生物之谜中的数学，另一方面也希望生物学家给我们多提些形态的问题，蜂房与结晶学联系起来，这是“透视石”的晶体。

再回到化工之巧，有多少种晶体可以无穷无尽、无空无隙地填满空间，这又要用到数学。数学上已证明，只有 230 种。

还有如胰岛素的研究中，由于复杂的立体模型也用了复杂的数学计算。生物遗传学中的密码问题是研究遗传与变异这一根本问题的，它的最终解决必然要考虑到数学问题。生物的反应用数学加以描述成为工程控制论中“反馈”的泉源。神经作用的数学研究为控制论和信息论提供了现实的原型。

日用之繁

日用之繁，的确繁，从何谈起真为难。但也有容易处。日用之繁与亿万人民都有关，只要到群众中去，急群众之所急，急生产和国防之所急，不但可以知道哪些该搞，而且知道轻重缓急。群众是真正的英雄，遇事和群众商量，不但政治上有提高，业务上也可以学到书本上所读不到的东西。像我这样自学专攻数学的，也在各行各业师傅的教育下，学到了不少学科的知识，这是一个大学一个专业中所学不到的。

我在日用之繁中搞些工作始于一九五八年，但真正开始是一九六四年接受毛主席的亲笔指示后。并且使我永远不会忘记的是在我刚迈出一步写了《统筹方法平话》下到基层试点时，毛主席又为我指出了“不为个人，而为人民服务，十分欢迎”的奋斗目标。后来在周总理关怀下又搞了《优选法》。由于各省、市、自治区的领导的关怀，我曾有机会到过二十个省市，下过数以千计的工矿农村，拜得百万工农老师，形成了有工人、技术人员和数学工作者参加的普及、推广数学方法的小分队。通过群众性的科学实验活动证明，数学确实大有用场，数学方法用于革新挖潜，能为国家创造巨大的财富。回顾已往，真有“抱着金饭碗讨饭吃”之感。

由于我们社会主义制度的优越性，在这一方面可能有我们自己的特点，不妨结合我下去后的体会多谈一些。

统筹方法不仅可用于一台机床的维修，一所房屋的修建，一组设备的安装，一项水利工程的施工，更可用于整个企业管理和大型重点工程的施工会战。大庆新油田开发，万人千台机的统筹，黑龙江省林业战线采、运、用、育的统筹，山西省大同市口泉车站运煤统筹，太原铁路局太钢和几个工矿的联合统筹，还有一些省市公社和大队的农业生产统筹等等，都取得了良好效果。看来统筹方法宜小更宜大。大范围的过细统筹效果更好，油水更大。特别是把方法交给广大群众，结合具体实际，大家动手搞起来，由小到大、由简到繁，在普及的基础上进一步提高，收效甚大。初步设想可以概括成十二个字：大统筹，理数据，建系统，策发展，使之发展成一门学科——统筹学，以适应我国具体情况，体现我们社会主义社会特点。统筹的范围越大，得到和用到的数据也越多。我们不仅仅是消极地统计这些数据，而且还要从这些数据中取出尽可能多的信息来作为指导。因此数据处理提到了日程上来。数据纷繁就要依靠电子计算机。新系统的建立和旧系统的改建和扩充，都必须在最优状态下运行。更进一步就是策发展，根据今年的情况明年如何发展才更积极又可靠，使国民经济的发展达到最大可能的高速度。

优选法是采用尽可能少的试验次数，找到最好方案的方法。优选学作为这类方法的数学理论基础，已有初步的系统研究。实践中，优选法的基本方法，已在大范围内得到推广。目前，我国化工、电子、冶金、机械、轻工、纺织、交通、建材等等方面都有较广泛的应用。在各级党委的领导下，大搞推广应用优选法的群众活动，各行各业搞，道道工序搞，短期内就可以应用优选法开展数以万计项目的试验，使原有的工艺水平普遍提高一步，在不添人、不增设备、不加或少加投资的情况下，就可收到优质、高产、低耗的效果。例如，小型化铁炉，优选炉形尺寸和操作条件，可使焦铁比一般达 1:18。机械加工优选刀具的几何参数和切削用量，工效可成倍提高。烧油锅炉，优选喷枪参数，可以达到节油不冒黑烟。小化肥工厂搞优选，既节煤

又增产。在大型化工设备上搞优选，提高收率潜力更大。解放牌汽车优选了化油器的合理尺寸，一辆汽车一年可节油一吨左右，全国现有民用汽车都来推广，一年就可节油六十余万吨。粮米加工优选加工工艺，一般可提高出米率百分之一、二、三，提高出粉率百分之一，若按全国人数的口粮加工总数计算，一年就等于增产几亿斤粮食。

最好的生产工艺是客观存在的，优选法不过是提供了认识它的、尽量少做试验、快速达到目的的一种数学方法。

物资的合理调配，农作物的合理分布，水库的合理排灌，电网的合理安排，工业的合理布局，都要用到数学才能完满解决，求得合理的方案。总之一句话，在具有各种互相制约、互相影响的因素的统一体中，寻求一个最合理（依某一目的，如最经济，最省人力）的解答便是一个数学问题，这就是“多、快、好、省”原则的具体体现。所用到的数学方法很多，其中确属适用者我们也准备了一些，但由于林彪、“四人帮”一伙的干扰破坏，没有力量进行深入的工作。今天，在开创社会主义建设事业新局面的同时，数学研究和应用也必将出现一个崭新的局面。

数学之发展

宇宙之大，粒子之微，火箭之速，化工之巧，地球之变，生物之谜，日用之繁，无处不用数学。其它如爱因斯坦用了数学工具所获得的公式指出了寻找新能源的方向，并且还预示出原于核破裂发生的能量的大小。连较抽象的纤维丛也应用到了物理当中。在天文学上，也是先从计算上指出海王星的存在，而后发现了海王星。又如高速飞行中，由次音速到超音速时出现了突变，而数学上出现了混合型偏微分方程的研究。还有无线电电子学与计算技术同信息论的关系，自动化与控制技术同常微分方程的关系，神经系统同控制论的关系，形态发生学与结构稳定性之间的关系等等，不胜枚举。

数学是一门富有概括性的学问。抽象是它的特色。同是一个方程，弹性力学上是描写振动的，流体力学上却描写了流体动态，声学家不妨称它是声学方程，电学家也不妨称它为电报方程，而数学家所研究的对象正是这些现象的共性的一面——双曲型偏微分方程。这个偏微分方程的解答的性质就是这些不同对象的共同性质，数值的解答也将是它所联系各学科中所要求的数据。

不但如此，这样的共性，一方面可以促成不同分支产生统一理论的可能性，另一方面也可以促成不同现象间的相互模拟性。例如：声学家可以用相似的电路来研究声学现象，这大大地简化了声学实验的繁重性。这种模拟性的最普遍的应用便是模拟电子计算机的产生。根据神经细胞有兴奋与抑制两态，电学中有带电与不带电两态，数学中二进位数的0与1，逻辑中的“是”与“否”，因而有用电子数字计算机来模拟神经系统的尝试，及模拟逻辑思维的初步成果。

我们作如上的说明，并不意味着数学家可以自我陶醉于共性的研究之中。一方面我们得承认，要求数学家深入到研究对象所联系的一切方面是十分困难的，但是这并不排斥数学家应当深入到他所联系到的为数众多的科学之一或其中的一部分。这样的深入是完全必要的。这样做既对国民经济建设可以做出应有的贡献，而且就是对数学本身的发展也有莫大好处。

客观事物的出现一般讲来有两大类现象。一类是必然的现象——或称因果律。一类是大数现象——或称机遇律。表示必然现象的数学工具一般是方程式，它可以从已知数据推出未知数据来，从已知现象的性质推出未知现象的性质来。通常出现的有代数方程、微分方程、积分方程、差分方程等等（特别是微分方程）。处理大数现象的数学工具是概率论与数理统计。通过这样的分析便可以看出大势所趋，各种情况出现的比例规律。

数学的其它分支当然也可以直接与实际问题相联系。例如：数理逻辑与计算机自动化的设计，复变函数论与流体力学，泛函分析与群表示论之与量子力学，黎曼几何之与相对论等等。在计算机设计中也用到数论。一般说来，数学本身是一个互相联系的有机整体，而上面所提到的两方面是与其它科学接触最多、最广泛的。

计算数学是一门与数学的开始而俱生的学问，不过今天由于快速大型计算机的出现特别显示出它的重要性。因为对象日繁，牵涉日广（一个问题的计算工作量大到了前所未有的程度），解一个一百个未知数的联立方程是今天科学中常见的（如水坝应力，大地测量，设计吊桥，大型建筑等等），仅靠笔算就很困难。算一个天气方程，希望从今天的天气数据推出明天的天气数据，单凭笔算要花成年累月的时间。这样算法与明天的天气何干？一个讽刺而已！电子计算机的发明就满足了这种要求。高速度大存储量的计算机的发展改变了科学的研究的面貌，但是近代的电子计算机的出现丝毫没有减弱数学的重要性，相反地更发挥数学的威力，对数学的要求提得更高。繁重的计算劳动减轻了或解除了，而创造性的劳动更多了。计算数学是一个桥梁，它把数学的创造同实际结合起来。同时它本身也是一个创造性的学科。例如推动了一个新学科计算物理学的发展，

除掉上面所特别强调的分支以外，并不是说数学的其余部分就不重要了。只有这些重点部门与其它部分环环扣紧，把纯数学和应用数学都分工合作地发展起来，才能既符合我国当前的需要，又符合长远需要。

从历史上数学的发展的情况来看，社会愈进步，应用数学的范围也就会愈大，所应用的数学也就愈精密，应用数学的人也就愈多。在日出而作、日入而息的古代社会里，会数数就可以满足客观的需要了。后来由于要定四时，测田亩，于是需要窥天测地的几何学。商业发展，计算日繁，便出现了代数学。要描绘动态，研究关系的变化，变化的关系，因而出现了解析几何学、微积分等等。

数学的用处在物理科学上已经经过历史考验而证明。它在生物科学和社会科学上的作用也已经露出苗头，存在着十分宽广的前途。

最后，我得声明一句，我并不是说其它学科不重要或次重要。应当强调的是，数学之所以重要正是因为其它学科的重要而重要的，不通过其它学科，数学的力量无法显示，更无重要之可言了。

需要指出的是，“四人帮”为了复辟资本主义，疯狂地破坏生产，破坏科学技术的发展，他们既破坏理论研究工作，更疯狂地打击从事应用数学的工作者。他们的遗毒需要彻底清除，不可低估。为了实现“四个现代化”，把我国建成强大的社会主义国家这一伟大目标，发展数学的重要性是无可置辩的。

数学名言

事实上在“纯的”与“应用的”数学之间找不到严格的分界线。

——（德国-美国数学家）R·柯朗

数学的发展与完善和国家的繁荣富强紧密相关。

——（法国皇帝）拿破仑

一门科学，只有当它成功地运用数学时，才能达到真正完善的地步。

——（德国哲学家）马克思

Maxwell Electromagnetic Theory from a Viewpoint of Differential Forms*

杜升华, 郝诚, 胡悦科,
惠瑜铭, 石权, 王力, 吴玉清

January, 2008

编者按:这篇关于 *Maxwell* 电磁场理论和下一篇关于 *Yang-Mills* 理论的论文均为”微分流形”课上小组合作完成。这两篇文章都体现了数学,更确切的说是几何学,与理论物理学的密切联系。这篇文章用微分形式的语言深入解读 *Maxwell* 电磁场理论,并涉及主丛的相关理论;后者成为在下一篇文章中详细描述的 *Yang-Mills* 理论的核心基础。

§1 Background

§1.1 Maxwell Equations

As we all know, the Maxwell equations are:

$$\nabla \cdot \mathbf{D} = \rho \quad (\text{Coulomb's Law}) \quad (1.1)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (\text{Ampere's Law}) \quad (1.2)$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0 \quad (\text{Faraday's Law}) \quad (1.3)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (\text{Absence of Free Magnetic Poles}) \quad (1.4)$$

The first two equations are inhomogeneous, while the other two are homogeneous. Here \mathbf{E} is the electric field vector, \mathbf{D} is the electric displacement, \mathbf{H} is the magnetic field and \mathbf{B} is the magnetic induction. We know that in vacuum (for simplicity we assume that we are always dealing with electromagnetic fields in vacuum) they are related by:

$$\mathbf{D} = \epsilon_0 \mathbf{E} \quad (1.5)$$

$$\mathbf{H} = \frac{1}{\mu_0} \mathbf{B} \quad (1.6)$$

where ϵ_0 and μ_0 are constants satisfying $\epsilon_0 \mu_0 = 1/c^2$, and c is the speed of light in vacuum. \mathbf{J} is the current density, following the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (1.7)$$

where ρ is the charge density.

*This article is a final report of the course *Differential Manifolds*.

§1.2 Vector and scalar potentials

It is convenient to introduce the scalar potential Φ and the vector potential \mathbf{A} , obtaining a smaller number of equations. According to (1.4), we can define \mathbf{B} in terms of a vector potential:

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (1.8)$$

Then the equation (1.3) can be written as

$$\nabla \times (\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}) = 0 \quad (1.9)$$

So, in this equation, the quantity with vanishing curl can be written as the gradient of some scalar potential Φ . We have

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla \Phi \quad (1.10)$$

That is

$$\mathbf{E} = -\nabla \Phi - \frac{\partial \mathbf{A}}{\partial t} \quad (1.11)$$

From the above equations we find that \mathbf{B} and \mathbf{E} are determined by \mathbf{A} and Φ , and the dynamic behavior of \mathbf{A} and Φ are determined by the two inhomogeneous Maxwell equations, which can be written as follows:

$$\nabla^2 \Phi + \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = -\rho/\varepsilon_0 \quad (1.12)$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla(\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \Phi}{\partial t}) = -\mu_0 \mathbf{J} \quad (1.13)$$

Considering that $\nabla \times (\nabla \Lambda) = 0$ where Λ is any scalar function, \mathbf{B} will be left unchanged by the transformation

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla \Lambda \quad (1.14)$$

\mathbf{E} is unchanged as well if

$$\Phi \rightarrow \Phi' = \Phi - \frac{\partial \Lambda}{\partial t} \quad (1.15)$$

We can choose a set of potentials (\mathbf{A}, Φ) to satisfy the *Lorenz condition*

$$\nabla \cdot \mathbf{A} + \frac{1}{c^2} \frac{\partial \Phi}{\partial t} = 0 \quad (1.16)$$

Then the equations (1.12) and (1.13) have the form of wave equations, one for Φ and one for \mathbf{A} :

$$\nabla^2 \Phi - \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} = -\rho/\varepsilon_0 \quad (1.17)$$

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J} \quad (1.18)$$

These equations show us that the electromagnetic fields have similar characters as waves.

§1.3 Maxwell equations in special theory of relativity

The laws of mechanics should be the same in different coordinate systems moving uniformly relative to one another. For a long time, we believe that the laws of mechanics are invariant under Galilean transformation. However, the form of the wave equation is not invariant under classical Galilean transformations, which is in contrast with electromagnetic phenomena.

Based on the results of Michelson's and Morley's experiments, Lorentz got new transformations, under which Maxwell equations (1.17) and (1.18) were invariant. Einstein improved the theory and finally gave the well-known Lorentz transformations. The transformations from one frame moving with velocity parallel to x_1 -axis is:

$$\left. \begin{array}{l} x'_0 = \gamma(x_0 - \beta x_1) \\ x'_1 = \gamma(x_1 - \beta x_0) \\ x'_2 = x_2 \\ x'_3 = x_3 \end{array} \right\} \quad (1.19)$$

where $\beta = |\frac{\mathbf{v}}{c}|$, $\gamma = (1 - \beta^2)^{(-1/2)}$

Or we may have the matrix form of Lorentz transformations

$$\begin{pmatrix} x'_0 \\ x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{pmatrix} \cosh \zeta & -\sinh \zeta & 0 & 0 \\ -\sinh \zeta & \cosh \zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (1.20)$$

The 4-dimensional space-time replaces the traditional space. We get an invariance of

$$s^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2 \quad (1.21)$$

In four-dimensional vector space, we will employ the notations

$$\partial_\alpha \equiv (\frac{\partial}{\partial x^\alpha}, \nabla) \quad (1.22)$$

$$\square \equiv \frac{\partial^2}{\partial(x^0)^2} - \nabla^2 \quad (1.23)$$

where $x^0 = ct$. With these notations, we will see that if we put Φ and \mathbf{A} together to form a 4-vector potential $A^\alpha = (\Phi, c\mathbf{A})$, and define $J^\alpha = (c\rho, \mathbf{J})$, then the wave equations (1.17)(1.18) have the form

$$\square A^\alpha = \frac{1}{c\varepsilon_0} J^\alpha \quad (1.24)$$

and the Lorenz condition turns into

$$\partial_\alpha A^\alpha = 0 \quad (1.25)$$

If we define

$$(F^{\alpha\beta}) = \begin{pmatrix} 0 & -E_{x^1} & -E_{x^2} & -E_{x^3} \\ E_{x^1} & 0 & -cB_{x^3} & cB_{x^2} \\ E_{x^2} & cB_{x^3} & 0 & -cB_{x^1} \\ E_{x^3} & -cB_{x^2} & cB_{x^1} & 0 \end{pmatrix} \quad (1.26)$$

and

$$(\mathfrak{F}^{\alpha\beta}) = \begin{pmatrix} 0 & -cB_{x^1} & -cB_{x^2} & -cB_{x^3} \\ cB_{x^1} & 0 & E_{x^3} & -E_{x^2} \\ cB_{x^2} & -E_{x^3} & 0 & E_{x^1} \\ cB_{x^3} & E_{x^2} & -E_{x^1} & 0 \end{pmatrix} \quad (1.27)$$

then the inhomogeneous Maxwell equations (1.1) (1.2) can be written as

$$\partial_\alpha F^{\alpha\beta} = \frac{1}{c\varepsilon_0} J^\beta \quad (1.28)$$

and the homogeneous Maxwell equations (1.3) (1.4) are

$$\partial_\alpha \mathfrak{F}^{\alpha\beta} = 0 \quad (1.29)$$

§2 Maxwell equations on 3-dimensional manifolds

In this section, we shall translate the classical Maxwell equations into the language of differential forms and discuss gauge invariance.

§2.1 Translation into the language of differential forms

Let M be a 3-dimensional manifold, which is contractible or, at least, has trivial cohomology groups $H^k(M)$ for $k \geq 1$. We express the physical quantities $\mathbf{E} = \mathbf{E}(t)$, $\mathbf{B} = \mathbf{B}(t)$ and so on as differential forms depending on the parameter $t \in \mathbb{R}$ smoothly.

More precisely, taking \mathbf{E} for example, by saying that $E = E(t), t \in \mathbb{R}$ is a family of smooth differential 1-forms depending on t smoothly, we mean that E can be represented as $E(t, x^1, x^2, x^3) = E_i(t, x^1, x^2, x^3)dx^i$ in a certain local coordinate system, where E_i has continuous partial derivative with respect to t for $i = 1, 2, 3$. It is obvious that the smooth dependence on t defined above does not depend on the choice of local coordinate systems.

We consider $E = E(t), H = H(t)$ as 1-forms, $D = D(t), B = B(t), J = J(t)$ as 2-forms, and $\rho = \rho(t)$ as a 3-form. We define two operators $\frac{\partial}{\partial t} : \Omega^k(M) \times \mathbb{R} \rightarrow \Omega^k(M) \times \mathbb{R}$ and $* : \Omega^k(M) \rightarrow \Omega^{3-k}(M)$ as follows ($k = 0, 1, 2, 3$).

Suppose $\omega = \omega(t)$ is a family of k -forms depending on t smoothly. In a local coordinate system, it can be represented as $\omega(t, x^1, x^2, x^3) = \omega_{i_1, \dots, i_k}(t, x^1, x^2, x^3)dx^{i_1} \wedge \dots \wedge dx^{i_k}$. ¹ Define

$$\frac{\partial}{\partial t} \omega(t, x^1, x^2, x^3) = \frac{\partial \omega_{i_1, \dots, i_k}}{\partial t}(t, x^1, x^2, x^3)dx^{i_1} \wedge \dots \wedge dx^{i_k}$$

in the same coordinate system. Since ω_{i_1, \dots, i_k} is continuously differentiable with respect to all its parameters, it is easy to check that the operator $\frac{\partial}{\partial t}$ is well defined. For example, $\frac{\partial E}{\partial t}(t, x^1, x^2, x^3) = \frac{\partial E_i}{\partial t}(t, x^1, x^2, x^3)dx^i$ in a local coordinate system, and $\frac{\partial E}{\partial t}(t, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3) = \frac{\partial \tilde{E}_i}{\partial t}(t, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3)d\tilde{x}^i$ in another. Furthermore, with the help of local representation, it can be seen that $\frac{\partial}{\partial t}$ is commutative with the exterior differential operator d .

¹More precisely, we should have written $\varphi^* \omega(t, x^1, x^2, x^3)$, where φ is a local chart, and φ^* is the corresponding pull-back map.

We define a linear mapping $* : \Omega^k(M) \rightarrow \Omega^{3-k}(M)$ for $k = 0, 1, 2, 3$ by:

$$*1 = dx^1 \wedge dx^2 \wedge dx^3$$

$$*dx^1 = dx^2 \wedge dx^3, *dx^2 = dx^3 \wedge dx^1, *dx^3 = dx^1 \wedge dx^2$$

$$*dx^2 \wedge dx^3 = dx^1, *dx^3 \wedge dx^1 = dx^2, *dx^1 \wedge dx^2 = dx^3$$

$$*dx^1 \wedge dx^2 \wedge dx^3 = 1$$

in a local coordinate system. It is not difficult to verify that the operator $*$ is well defined and commutative with $\frac{\partial}{\partial t}$. Obviously, $*^2 = * \circ *$ is an identity map.

Now we can transform the classical Maxwell equations as follows:

$$dD = \rho \quad (2.30)$$

$$dH = J + \frac{\partial D}{\partial t} \quad (2.31)$$

$$dB = 0 \quad (2.32)$$

$$dE + \frac{\partial B}{\partial t} = 0 \quad (2.33)$$

In our case, E and H are 1-forms but D and B are 2-forms. So we should use the operator $*$ to express their relations precisely:

$$D = \epsilon_0 * E \quad (2.34)$$

$$H = \frac{1}{\mu_0} * B \quad (2.35)$$

§2.2 Potentials

In classical electrodynamics, it is convenient to introduce the scalar potential Φ and the vector potential \mathbf{A} , obtaining a smaller number of equations. Now we do the analogous things in the language of differential forms.

Since $dB = 0$ and the manifold M has trivial cohomology groups, there exists a 1-form A such that

$$B = dA \quad (2.36)$$

Then equation (2.33) can be written as

$$d \left(E + \frac{\partial A}{\partial t} \right) = 0 \quad (2.37)$$

This means that $E + \frac{\partial A}{\partial t}$ is the exterior differential of a 0-form, namely, a smooth function $-\Phi$:

$$E + \frac{\partial A}{\partial t} = -d\Phi, \quad \text{or} \quad E = -d\Phi - \frac{\partial A}{\partial t} \quad (2.38)$$

The definition of B and E in terms of the 1-form A and the 0-form Φ according to (2.36) and (2.38) satisfies identically the two homogeneous Maxwell equations (2.32) and (2.33). Now let's substitute these expressions into the other two Maxwell equations.

According to (2.34) and (2.38), $D = -\varepsilon_0 (*d\Phi + \frac{\partial}{\partial t}(*A))$, so $dD = -\varepsilon_0 (d * d\Phi + \frac{\partial}{\partial t}(d * A))$. It is not difficult to see² that $d * d\Phi = *\nabla^2\Phi$.

According to (2.35) and (2.36), $H = \frac{1}{\mu_0} * dA$, so $dH = \frac{1}{\mu_0} d * dA$. Using the above expression for D and the commutativity of $\frac{\partial}{\partial t}$ with $*$ and d , we can see that $\frac{\partial D}{\partial t} = -\varepsilon_0 * (d(\frac{\partial \Phi}{\partial t}) + \frac{\partial^2 A}{\partial t^2})$. Thus we can transform equation (2.31) into

$$d * dA + \frac{1}{c^2} * \left(d \left(\frac{\partial \Phi}{\partial t} \right) + \frac{\partial^2 A}{\partial t^2} \right) = \mu_0 J \quad (2.39)$$

It is not difficult to verify that $d * d * A$ corresponds to $\nabla(\nabla \cdot \mathbf{A})$, and $*d * dA$ to $\nabla \times (\nabla \times \mathbf{A})$.³ In fact, if we introduce an operator $\delta = *d* : \Omega^k(M) \rightarrow \Omega^{k-1}(M)$, then δA corresponds to $\nabla \cdot \mathbf{A}$. Recalling the formula $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$, we can define an operator $\nabla^2 : \Omega^k(M) \rightarrow \Omega^k(M)$ by $\nabla^2 = [d, \delta] = d\delta - \delta d = d * d * - * d * d$ for $k = 1, 2, 3$.⁴ It will turn out to be useful for simplifying the Maxwell equations.

Now we can rewrite the four equations into two:

$$*\nabla^2\Phi + \frac{\partial}{\partial t}(d * A) = -\rho/\varepsilon_0 \quad (2.40)$$

$$*(\nabla^2 A - \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2} - d(*d * A + \frac{1}{c^2} \frac{\partial \Phi}{\partial t})) = -\mu_0 J \quad (2.41)$$

§2.3 Gauge invariance

To simplify the above equations, we must exploit the arbitrariness involved in the definition of the potentials and choose what is convenient. Since B is defined through (2.36) by $B = dA$ and $d^2 = 0$, it is left unchanged by the transformation,

$$A \rightarrow A' = A + d\Lambda \quad (2.42)$$

For E defined through (2.38) to be unchanged as well, Φ must be simultaneously transformed,⁵

$$\Phi \rightarrow \Phi' = \Phi - \frac{\partial \Lambda}{\partial t} \quad (2.43)$$

The freedom implied by (2.42) and (2.43) means that we can choose a set of (A, Φ) to satisfy the *Lorenz condition* (1867),

$$*d * A + \frac{1}{c^2} \frac{\partial \Phi}{\partial t} = 0 \quad (2.44)$$

This will uncouple the equations (2.40) and (2.41), leaving two wave equations only:

$$*(\nabla^2\Phi - \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2}) = -\rho/\varepsilon_0 \quad (2.45)$$

²In a local coordinate system, $*d\Phi = \frac{\partial \Phi}{\partial x^i} dx^1 \wedge \widehat{dx^i} \wedge dx^3$, and $d * d\Phi = \nabla^2\Phi dx^1 \wedge dx^2 \wedge x^3 = *\nabla^2\Phi$.

³For example, $d * d * A = d * \left(\frac{\partial A_i}{\partial x^i} dx^1 \wedge dx^2 \wedge dx^3 \right) = \frac{\partial}{\partial x^j} \left(\frac{\partial A_i}{\partial x^i} \right) dx^j$. The other one is left to the reader.

⁴For $k = 0$, $d * d * - * d * d = -\sum_{i=1}^3 \frac{\partial^2}{\partial(x^i)^2}$, which differ from the classical Laplacian operator by a negative sign. So we use ∇^2 in the sense of the above definition only for $k \geq 1$, especially for $k = 1$, and as the classical Laplacian operator for $k = 0$.

⁵ $E = -d\Phi - \frac{\partial A}{\partial t} = -d\Phi + d(\frac{\partial \Lambda}{\partial t}) - \frac{\partial}{\partial t}(d\Lambda) - \frac{\partial A}{\partial t} = -d\Phi' - \frac{\partial A'}{\partial t}$

$$*(\nabla^2 A - \frac{1}{c^2} \frac{\partial^2 A}{\partial t^2}) = -\mu_0 J \quad (2.46)$$

Equations (2.45) and (2.46), plus (2.44), form a set of equations equivalent to the Maxwell equations in vacuum. The transformation (2.42) and (2.43) is called a *gauge transformation*, and the invariance of the fields under such transformations is called *gauge invariance*.

In fact, suppose A and Φ do not satisfy (2.44), making a gauge transformation (2.42) and (2.43), we can demand that A' , Φ' satisfy Lorenz condition:

$$*d * A' + \frac{1}{c^2} \frac{\partial \Phi'}{\partial t} = 0 = *d * A + \frac{1}{c^2} \frac{\partial \Phi}{\partial t} + \nabla^2 \Lambda - \frac{1}{c^2} \frac{\partial^2 \Lambda}{\partial t^2} \quad (2.47)$$

provided a gauge function Λ can be found to satisfy

$$\nabla^2 \Lambda - \frac{1}{c^2} \frac{\partial^2 \Lambda}{\partial t^2} = - \left(*d * A + \frac{1}{c^2} \frac{\partial \Phi}{\partial t} \right) \quad (2.48)$$

Suppose the Lorenz condition is satisfied by A, Φ , then it is preserved by the *restricted gauge transformation*,

$$\begin{aligned} A &\rightarrow A + d\Lambda \\ \Phi &\rightarrow \Phi - \frac{\partial \Lambda}{\partial t} \\ \nabla^2 \Lambda - \frac{1}{c^2} \frac{\partial^2 \Lambda}{\partial t^2} &= 0 \end{aligned} \quad (2.49)$$

All potentials satisfying the Lorenz condition are said to belong to the *Lorenz gauge*. The Lorenz gauge is a concept independent of coordinate system chosen and so fits naturally into the considerations of special relativity.

§3 Maxwell equations in special relativity

§3.1 4-dimensional Maxwell Equations in differential forms

In special relativity, we take space and time as a whole into consideration and deal with the 4-dimensional space $\mathbb{R}^{1,3}$ called space-time with a constant Lorentz metric. Now we will study the 4-dimensional case.⁶

Recall that we expressed the physical quantities $\mathbf{E} = \mathbf{E}(t)$, $\mathbf{B} = \mathbf{B}(t)$ and so on as differential forms depending on $t \in \mathbb{R}$ smoothly in 3-dimensional case. Now it is convenient to, for example, express \mathbf{E} as a single differential form $E(x^0, x^1, x^2, x^3) = E_i(x^0, x^1, x^2, x^3)dx^i$, where $\{x^0, x^1, x^2, x^3\}$ is a coordinate system and $x^0 = ct$.⁷ We use E_i, B_i, J_i and so on to denote the components of corresponding physical quantities, which are smooth functions of (x^0, x^1, x^2, x^3) .

⁶We are unable to study general 4-manifolds in special relativity because the metric matrix is required to be unchanged under coordinate transformations. But what we do here in the language of differential forms may be easily generalized to general relativity.

⁷In $\mathbb{R}^{1,3}$ it can be even chosen as a global coordinate system.

We rewrite the classical Maxwell equations as follows:(where we have used the relations (1.5), (1.6) and $x^0 = ct$)

$$\nabla \cdot \mathbf{B} = 0 \quad (3.50)$$

$$\nabla \times \mathbf{E} + c \frac{\partial \mathbf{B}}{\partial x^0} = 0 \quad (3.51)$$

$$\nabla \cdot \mathbf{E} = \frac{1}{\varepsilon_0} \rho \quad (3.52)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial x^0} \quad (3.53)$$

To simplify the homogeneous Maxwell equations (3.50) and (3.51), we define a 2-form F , called "Faraday":

$$\begin{aligned} F = & -E_1 dx^0 \wedge dx^1 - E_2 dx^0 \wedge dx^2 - E_3 dx^0 \wedge dx^3 \\ & + c(B_1 dx^2 \wedge dx^3 + B_2 dx^3 \wedge dx^1 + B_3 dx^1 \wedge dx^2) \end{aligned} \quad (3.54)$$

We compute

$$\begin{aligned} dF = & (c \nabla \cdot \mathbf{B}) dx^1 \wedge dx^2 \wedge dx^3 \\ & + (c \frac{\partial B_1}{\partial x^0} + \frac{\partial E_3}{\partial x^2} - \frac{\partial E_2}{\partial x^3}) dx^0 \wedge dx^2 \wedge dx^3 \\ & + (c \frac{\partial B_2}{\partial x^0} + \frac{\partial E_1}{\partial x^3} - \frac{\partial E_3}{\partial x^1}) dx^0 \wedge dx^3 \wedge dx^1 \\ & + (c \frac{\partial B_3}{\partial x^0} + \frac{\partial E_2}{\partial x^1} - \frac{\partial E_1}{\partial x^2}) dx^0 \wedge dx^1 \wedge dx^2 \\ = & 0 \end{aligned} \quad (3.55)$$

where we have used the first two equations. Conversely, $dF = 0$ also implies the validity of these two equations.

To deal with the last two equations, observe that by putting $c\mathbf{B} \rightarrow \mathbf{E}$ and $\mathbf{E} \rightarrow -c\mathbf{B}$ in the first two equations, we actually have something quite similar to the last two. More precisely, we may as well define a new 2-form M , called "Maxwell":

$$\begin{aligned} M = & c(B_1 dx^0 \wedge dx^1 + B_2 dx^0 \wedge dx^2 + B_3 dx^0 \wedge dx^3) \\ & + E_1 dx^2 \wedge dx^3 + E_2 dx^3 \wedge dx^1 + E_3 dx^1 \wedge dx^2 \end{aligned} \quad (3.56)$$

It is now convenient to introduce a new operator $*$ for 4-dimensional case. In general, on an n-manifold with a metric g , the *Hodge star operator* $* : \Omega^k \rightarrow \Omega^{n-k}$ is such a linear operator that

$$\alpha \wedge * \beta = g(\alpha, \beta) d\text{vol}_g \quad (3.57)$$

where $\alpha = \alpha_{i_1 \dots i_k} dx^{i_1} \wedge \dots \wedge dx^{i_k}$, $\beta = \beta_{j_1 \dots j_k} dx^{j_1} \wedge \dots \wedge dx^{j_k}$, $g(\alpha, \beta) = \alpha_{i_1 \dots i_k} \beta_{j_1 \dots j_k} g^{i_1 j_1} \dots g^{i_k j_k}$, $d\text{vol}_g = \sqrt{|\det(g_{ij})|} dx^1 \wedge \dots \wedge dx^n$ and (g^{ij}) is the inverse of the metric matrix (g_{ij}) . In special

relativity the metric is given by

$$(g_{ij}) \equiv \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix} \quad (3.58)$$

and the star is determined by the following rules:

$$\begin{aligned} *1 &= dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 \\ *dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 &= -1 \\ \\ *dx^0 &= dx^1 \wedge dx^2 \wedge dx^3 & *dx^1 \wedge dx^2 \wedge dx^3 &= dx^0 \\ *dx^1 &= dx^0 \wedge dx^2 \wedge dx^3 & *dx^0 \wedge dx^2 \wedge dx^3 &= dx^1 \\ *dx^2 &= dx^0 \wedge dx^3 \wedge dx^1 & *dx^0 \wedge dx^3 \wedge dx^1 &= dx^2 \\ *dx^3 &= dx^0 \wedge dx^1 \wedge dx^2 & *dx^0 \wedge dx^1 \wedge dx^2 &= dx^3 \\ \\ *(dx^0 \wedge dx^1) &= -dx^2 \wedge dx^3 & *(dx^0 \wedge dx^2) &= -dx^3 \wedge dx^1 & *(dx^0 \wedge dx^3) &= -dx^1 \wedge dx^2 \\ *(dx^1 \wedge dx^2) &= dx^0 \wedge dx^3 & *(dx^3 \wedge dx^1) &= dx^0 \wedge dx^2 & *(dx^2 \wedge dx^3) &= dx^0 \wedge dx^1 \end{aligned}$$

It can be concluded that $*\omega = (-1)^{k+1}\omega$ if ω is a k -form, $k = 0, \dots, 4$.

Then it is ready to see that

$$M = *F \quad (3.59)$$

and we compute (keep the last two equations in mind)⁸

$$\begin{aligned} dM &= (\nabla \cdot \mathbf{E}) dx^1 \wedge dx^2 \wedge dx^3 \\ &\quad + \left(\frac{\partial E_1}{\partial x^0} - c \left(\frac{\partial B_3}{\partial x^2} - \frac{\partial B_2}{\partial x^3} \right) \right) dx^0 \wedge dx^2 \wedge dx^3 \\ &\quad + \left(\frac{\partial E_2}{\partial x^0} - c \left(\frac{\partial B_1}{\partial x^3} - \frac{\partial B_3}{\partial x^1} \right) \right) dx^0 \wedge dx^3 \wedge dx^1 \\ &\quad + \left(\frac{\partial E_3}{\partial x^0} - c \left(\frac{\partial B_2}{\partial x^1} - \frac{\partial B_1}{\partial x^2} \right) \right) dx^0 \wedge dx^1 \wedge dx^2 \\ &= \frac{1}{\varepsilon_0} \rho dx^1 \wedge dx^2 \wedge dx^3 \\ &\quad - c\mu_0 J_1 dx_0 \wedge dx^2 \wedge dx^3 - c\mu_0 J_2 dx_0 \wedge dx^3 \wedge dx^1 - c\mu_0 J_3 dx_0 \wedge dx^1 \wedge dx^2 \end{aligned} \quad (3.60)$$

or

$$*d *F = \frac{1}{c\varepsilon_0} \tilde{J} \quad (3.61)$$

where \tilde{J} is a 1-form, called the "4-current".⁹

$$\tilde{J} = c\rho dx^0 - (J_1 dx^1 + J_2 dx^2 + J_3 dx^3) \quad (3.62)$$

⁸Here we regard ρ as a smooth function, not a 3-form.

⁹Don't forget that $c^2\mu_0\varepsilon_0 = 1$.

Introducing a new operator $\delta = *d* : \Omega^k \rightarrow \Omega^{k-1}$, we have

$$\delta F = \frac{1}{c\varepsilon_0} \tilde{J} \quad (3.63)$$

We remark that the relation above implies the last two equations.

To sum up, we have reduced Maxwell equations into two very simple identities involving the differential forms:

$$\begin{aligned} dF &= 0 \\ \delta F &= \frac{1}{c\varepsilon_0} \tilde{J} \end{aligned}$$

§3.2 Discussions

Since $dF = 0$ in \mathbb{R}^4 , we must have

$$F = dI \quad (3.64)$$

where the "potential" 1-form I is

$$I = -\Phi dx^0 + c(A_1 dx^1 + A_2 dx^2 + A_3 dx^3) \quad (3.65)$$

From the relation $F = dI$ we deduce

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (3.66)$$

$$\mathbf{E} = -c \frac{\partial \mathbf{A}}{\partial x_0} - \nabla \Phi \quad (3.67)$$

where $\mathbf{A} = (A_1, A_2, A_3)$. Our choice of I is not unique. Indeed I is determined up to an "exact" 1-form, i.e.

$$I' = I + d\Lambda \quad (3.68)$$

serves our demand as well, thus we have the following *gauge transformation* leaving \mathbf{E} and \mathbf{B} unchanged:

$$\mathbf{A}' = \mathbf{A} + \nabla \Lambda \quad (3.69)$$

$$\Phi' = \Phi - c \frac{\partial \Lambda}{\partial x_0} \quad (3.70)$$

I also plays a fundamental role in our differential form of Maxwell equations, since now we have only one equation remained:

$$\delta dI = *d * dI = \frac{1}{c\varepsilon_0} \tilde{J} \quad (3.71)$$

Curiously, we might want to know about $*d * I$:

$$\begin{aligned} *I &= -\Phi dx^1 \wedge dx^2 \wedge dx^3 + c(A_1 dx_0 \wedge dx^2 \wedge dx^3 + A_2 dx_0 \wedge dx^3 \wedge dx^1 + A_3 dx_0 \wedge dx^1 \wedge dx^2) \end{aligned} \quad (3.72)$$

$$d * I = -\left(\frac{\partial \Phi}{\partial x^0} + c\nabla \cdot \mathbf{A}\right) dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 \quad (3.73)$$

$$*d * I = \frac{\partial \Phi}{\partial x^0} + c\nabla \cdot \mathbf{A} \quad (3.74)$$

If $*d*I = 0$, we say that I satisfies the *Lorenz condition*:

$$\frac{\partial \Phi}{\partial x^0} + c\nabla \cdot \mathbf{A} = 0 \quad (3.75)$$

which is equivalent to (1.16). It can be proved that we can always choose I (more precisely, Λ) to satisfy the Lorenz condition.

From the identity

$$\delta F = \frac{1}{c\varepsilon_0} \tilde{J}$$

we deduce

$$\delta \tilde{J} = 0 \quad (3.76)$$

or more explicitly

$$c \frac{\partial \rho}{\partial x^0} + \nabla \cdot \mathbf{J} = 0 \quad (3.77)$$

i.e.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (3.78)$$

This is the same as equation (1.7) and also called *law of conservation* which states the total charge is invariant.

§4 Classical electrodynamics as the curvature of a line bundle

§4.1 About principal bundle

To interpret electromagnetism as a gauge theory, we need first to introduce some definitions.

A differentiable fiber bundle consists of following ingredients:

1. Three manifolds: the total space E , the base space X , and the fiber F .
2. A map $\pi : E \rightarrow X$, the projection.
3. A covering U of X by a family $(U_\alpha)_{\alpha \in A}$, of open sets.
4. For each U_α in U a diffeomorphism, the local trivialization, $h_\alpha : \pi^{-1}(U_\alpha) \times F$ such that for U_β in U with $U_\alpha \cap U_\beta \neq \emptyset$ the map

$$h_\alpha \circ h_\beta^{-1} : (U_\alpha \cap U_\beta) \times F$$

is given by

$$h_\alpha \circ h_\beta^{-1}(x, f) = (x, h_{\alpha\beta}(x)(f))$$

where

$$h_{\alpha\beta} : U_\alpha \cap U_\beta \rightarrow Aut(F)$$

and are called the transition maps.

Next, a principal G-bundle is a fibre bundle whose fibre is a Lie group G, and each of its transition maps $h_{\alpha\beta}$ corresponds to the left action of an element $g_{\alpha\beta}$ in G.

A differential-geometric G-connection on a principal G-bundle is a 1-form ω on E taking its values in the Lie algebra \mathfrak{g} of G, with the following two properties:

(1)the "normalization" property: restriction of the form ω to the fiber G yields the above-defined form $\omega_0 = -(dg)g^{-1}$;

(2)invariance property: under the natural left action of the group G on E we have that $g^*\omega = Ad(g)\omega = g\omega g^{-1}$.

On each local trivialization $h_U : \pi^{-1}(U) \rightarrow U \times G$, the connection defined above can be represented by a \mathfrak{g} -valued 1-form A_U on U, which we shall call the *local principal gauge potential*. Note that in doing so we have passed from an object defined on the total space P to one defined on the base space M. As in physics M is generally space-time, this is in keeping with the notion that physical field theory deals with objects defined directly on space-time.

Conversely, suppose we have a set of trivializations $h_U : \pi^{-1}(U) \rightarrow U \times G$ such that the open sets U cover X. Suppose that for each trivialization we have a \mathfrak{g} -valued 1-form A_U such that if $U \cap V \neq \emptyset$, then in $U \cap V$, the equality

$$A_V = Ad_{g_{VU}} A_U - dg_{VU} \cdot g_{VU}^{-1} \quad (*)$$

holds, in which g_{VU} is the fibre transition map. Then there is a unique invariant connection in P for which the representatives with respect to the given trivializations are the given A_U .

From gauge potential we can define a \mathfrak{g} -valued 2-form in U called the *curvature 2-form* F that satisfies $F(\eta, \xi) = dA(\eta, \xi) + [A(\eta), A(\xi)]$, or in short, $F = dA + [A, A]$.

With the definitions above, we can now further the understanding of electromagnetism. The 1-form A in electromagnetism takes its value in \mathbb{R} which is the Lie algebra of Lie group $U(1)$. By checking the equation (*) one can conclude that A is the local principal gauge potential of an invariant connection on the principal $U(1)$ -bundle. What's more, the curvature 2-form $\tilde{F} = dA + [A, A] = dA$ because of the commutativity of \mathbb{R} , thus the electromagnetic 2-form is actually the curvature 2-form of the connection that A decides.

§4.2 Developments after Maxwell's equations

In the preceding subsection, we can attach the quantities (differential forms) produced by electromagnetic fields to our 4-dimensional space-time by description to electromagnetic fields in 4-dimensional space. Thus our universe can be explained as a structure of 4-dimensional manifold, with a structure of connections deduced from electromagnetic fields attached to it.

This means that our description to the universe is deeper than 4-dimensional space.

Of course, what we do here is actually a special case of Yang-Mills theory.

But an idea to construct models for the universe is given here: 4-dimensional manifold plus structure of principal bundle.

In fact, what we added is $U(1)$, for it is a simplest Lie group. A particularity of $U(1)$ is that it is Abelian, which is not a common property of Lie groups. When we replace $U(1)$ into a more complex Lie group, Yang-Mills theory comes out.

参考文献

- [1] John David Jackson: *Classical Electrodynamics*, Third Edition
- [2] George Svetlichny: *Preparation for Gauge Theory*
- [3] Laurent Claessens: *Field theory from a bundle point of view*

数学谜语

1. What is yellow and complete?
2. What is round, purple and commutative?
3. How is a set in a topological space different from a door? (James R. Munkres: TOPOLOGY, p93)
4. (原创) 谁是抽象代数学的“外祖父”?

(答案见第 76 页)

数学家趣闻

▲ Levy, Paul (1886-1971) 这个人数学做的虽然不错,但是记忆力却很差。有一次 Errera, Alfred (1886-1960) (Landau 的一个学生) 为 Levy 举办了一场晚宴。第二天, Errera 碰见 Levy, 毕恭毕敬的说:“我很高兴昨天度过一个美好的夜晚”“恩?那么你昨晚在哪?”

▲一次拓扑课, Minkowski (闵可夫斯基) 向学生们自负的宣称:“这个定理没有证明的最主要的原因是至今只有一些三流的数学家在这上面花过时间。下面我就来证明它……”于是 Minkowski 开始拿起粉笔。这节课结束的时候,没有证完,到下一次课的时候, Minkowski 继续证明,一直几个星期过去了……一个阴霾的早上, Minkowski 跨入教室,那时候,恰好一道闪电划过长空,雷声震耳, Minkowski 很严肃的说:“上天被我的骄傲激怒了,我的证明是不完全的……”

▲由于 Fermat 大定理的名声,在 New York 的地铁车站出现了乱涂在墙上的话: $x^n + y^n = z^n$ 没有解,对此我已经发现了一种真正美妙的证明,可惜我现在没时间写出来,因为我的火车正在开来。

A Brief Introduction to Yang-Mills Theory*

Jie Zhou , Jiatu Cai[†]

January 20, 2008

Abstract

The main part of our report is our understandings on how to generalize physical theories from the basic principles and in a natural way, from the example of the Yang-Mills Theory. In this report, we will introduce the origin of gauge transformation and gauge invariance. Then we investigate the possible process of generalization, in which our principle is the invariance of Lagrangian under local gauge symmetry, with “the least” assumption on physical laws and mathematical forms. In the last section, we appeal to language of fiber bundle and its connection, during which progress we will find at last the quantities introduced in previous sections come out elegantly.

1 Introduction

1.1 Gauge Symmetry: Invariant under Transformation

In physics, gauge theories are a class of physical theories based on the idea that symmetry transformations can be performed locally as well as globally. Many powerful theories in physics are described by Lagrangians which are invariant under certain symmetry transformation groups. When they are invariant under a transformation identically performed at every space-time point they are said to have a **global symmetry**. Gauge theory extends this idea by requiring that the Lagrangians must possess **local symmetries** as well, it should be possible to perform these symmetry transformations in a particular region of space-time without affecting what happens in another region.

For example, in quantum physics, symmetry is a transformation between physical states that preserves the expectation values of all observables O (in particular the Hamiltonian)

$$\begin{aligned} S : |\varphi\rangle &\mapsto |\psi\rangle = S|\varphi\rangle \\ |\langle\psi|O|\psi\rangle|^2 &= |\langle\varphi|O|\varphi\rangle|^2 \end{aligned}$$

The usual formulation of the physics theories uses fields, which sometimes are not physical quantities. Such are the gauge fields (fiber bundle connections for the mathematicians), which provide a redundant but convenient description of the physical degrees of freedom. The gauge (local) “symmetries” are a reflection of this redundancy. The physical quantities are certain equivalence classes of gauge fields. An analogy can be made with the construction of the real numbers. We can use sequences of rational numbers that have the same limit. Of course, each real number is represented by infinitely many such sequences. We can choose a particular well defined sequence to be a representative of the real number. This corresponds to the procedure of gauge fixing in gauge theories. The fact that gauge fields are not physical degrees of freedom becomes very clear when we try to quantize them. Then we are forced to work in one way or another with the physical quantities by removing the redundancy (the gauge symmetry).

*Course Report for Differentiable Manifolds, Fall 2007.

[†]数41班

1.2 Motivation : From Global Symmetry to Local Symmetry

We use the scalar $O(n)$ gauge theory to illustrate how local gauge invariance can be “motivated” heuristically starting from global symmetry properties.

Consider a set of n non-interacting (free) scalar fields, with equal masses m . This system is described by an action which is the sum of the (usual) action for each scalar field φ_i

$$\mathcal{S} = \int d^4x \sum_{i=1}^n \left[\frac{1}{2}(\partial_\mu \varphi_i)(\partial^\mu \varphi_i) - \frac{1}{2}m^2 \varphi_i^2 \right]$$

By introducing a vector of fields $\Phi = (\varphi_1, \varphi_2, \dots, \varphi_n)^T$, the Lagrangian (density) can be compactly written as (use the Einstein notation)

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi)^T \partial_\mu \Phi - \frac{1}{2}m^2 \Phi^T \Phi$$

It is now transparent that the Lagrangian is invariant under the transformation $\Phi \mapsto g\Phi$ whenever g is a constant matrix belonging to the n -by- n orthogonal group $O(n)$. This is the global symmetry of this particular Lagrangian, and the symmetry group is often called the **gauge group**.

Now, demanding that this Lagrangian should have local $O(n)$ -invariance requires that the G matrices (which were earlier constant) should be allowed to become functions of the **space-time** coordinates x .

Unfortunately, the g matrices do not “pass through” the derivatives. When $g = g(x)$,

$$\partial_\mu(g\Phi)^T \partial^\mu(\Phi) \neq \partial_\mu \Phi^T \partial^\mu \Phi$$

This suggests defining a new Lagrangian, i.e., replace the functor ∂_μ by a new “derivative” D with the property

$$\mathcal{L}(g\Phi) = \mathcal{L}(\Phi)$$

that is,

$$\frac{1}{2}(D_\mu(g\Phi))^T D^\mu(g\Phi) - \frac{1}{2}m^2(g\Phi)^T(g\Phi) = \frac{1}{2}(D_\mu \Phi)^T D^\mu \Phi - \frac{1}{2}m^2 \Phi^T \Phi$$

Since $g(x) \in O(n)$ we have $(g\Phi)^T(g\Phi) = \Phi^T \Phi$, from the previous case it is tempting to define the new “derivative” (called covariant derivative) D as follows:

$$D_\mu = \partial_\mu + \gamma A_\mu(x)$$

such that

$$D_\mu(g\Phi) = gD_\mu \Phi$$

We now consider the form of $A_\mu(x)$ in order that the local symmetry holds. Since

$$D_\mu(g\Phi) = \partial_\mu(g\Phi) + \gamma \tilde{A}_\mu(g\Phi) = (\partial_\mu g)\Phi + g(\partial_\mu \Phi) + \gamma \tilde{A}_\mu(g\Phi)$$

it suffices to require that

$$(\partial_\mu g)\Phi + \gamma \tilde{A}_\mu g\Phi = g(\gamma A_\mu \Phi)$$

that is, the **gauge field** $A(x)$ is defined to have the transformation law

$$\tilde{A}_\mu(x) = g(x)A_\mu(x)g^{-1}(x) - \gamma^{-1}(\partial_\mu g(x))g^{-1}(x)$$

Our picture of classical gauge theory is almost complete except for the fact that to define the covariant derivatives D , a.e. $A_\mu(x)$ which satisfies the local transformation mentioned above, one needs to know the value of the gauge field $A(x)$ at all space-time points. Instead of manually specifying the values of this field, it can be given as the solution to a field equation from the principles of mechanics, besides satisfying the local invariance..

In fact, **Weil's Principle of Gauge Invariance** implies that if ψ satisfies Schrödinger's equations which involves the potential A , then $\exp(\frac{i}{e}\hbar f(x))\psi$ satisfies the Schrödinger's equations in which A has been replaced by $A + df$.

It will be indicated that the gauge field is an element of the Lie algebra, and can therefore be expanded as

$$A_\mu(x) = \sum_a A_\mu^a(x) T^a$$

There are therefore as many gauge fields as there are generators of the Lie algebra. We now have a **locally gauge invariant** Lagrangian

$$\mathcal{L}_{\text{loc}} = \frac{1}{2}(D_\mu \Phi)^T D^\mu \Phi - \frac{1}{2}m^2 \Phi^T \Phi$$

The difference between this Lagrangian and the original globally gauge-invariant Lagrangian is seen to be the interaction Lagrangian

$$\begin{aligned} \mathcal{L}_{\text{local}} &= \mathcal{L}_{\text{global}} + \mathcal{L}_{\text{int}} \\ \mathcal{L}_{\text{int}} &= \frac{g}{2} \Phi^T A_\mu^T \partial^\mu \Phi + \frac{g}{2} (\partial_\mu \Phi)^T A^\mu \Phi + \frac{g^2}{2} (A_\mu \Phi)^T A^\mu \Phi \end{aligned}$$

The term \mathcal{L}_{int} introduces interactions between the n scalar fields just as a consequence of the demand for local gauge invariance. In the quantized version of this classical field theory, the quanta of the gauge field $A(x)$ are called gauge bosons. The interpretation of the interaction Lagrangian in quantum field theory is of scalar bosons interacting by the exchange of these gauge bosons.

2 Yang-Mills's Theory

2.1 Review of Electromagnetism

We will use the Hamilton theory to drive the equations describing electromagnetic fields (i.e. Maxwell's equations) which turn out to be just the Euler-Lagrange equations corresponding to the particular action

$$S = S_f + S_m + S_{mf}$$

We begin with the term S_m . This is that part of the action due to the particles (i.e. charges) considered separately from the field through which they move, i.e. the action of the charges assuming the field absent. The appropriate action, as it is usually defined, is given by

$$S_m = - \sum_i m_i c \int_a^b dl,$$

where the sum is taken over all of the particles in the field, of masses m_i , c is the speed of light, and the integral $\int_a^b dl$ (where l denotes arc length) is taken over the arc of the world-line of the particle in $\mathbb{R}_{1,3}^4$ between the two fixed events corresponding to the positions of the particle

at an initial time t_1 and a later time t_2 . The term S_{mf} , representing that part of the action determined by the mutual interaction of the particles and the field, is usually defined by

$$S_{mf} = - \sum_j \frac{e_j}{c} \int A_k^{(j)} dx^k$$

where again the summation is over all particles indexed by j , where e_j is the charge on the j th particle, where, as for S_m , the integral is taken over an arc of the world-line of the the particle. Here (A_i) is a given 4-covector defined on $\mathbb{R}_{1,3}^4$ (the so-called “4-potential”) which characterizes the field Finally, the term S_f is that part of the action depending on the properties of the field alone, i.e. the action due to the field in the assumed absence of charges. If we are interested only in the motion of the particles in a given electromagnetic field, then the term S_f need not to be considered; on the other hand this term is crucial if our interest lies rather in finding equations characterizing the field. By way of preparing for the definition of S_f we introduce some already familiar concepts of electromagnetic-field theory, defining them in terms of the basic 4-potential (A_i) . The three spatial components A^1, A^2, A^3 of the 4-vector (A^i) obtained by raising the index of the tensor (A_i) (for this purpose resorting , of course, to the Minkowski metric), define a 3-vector \mathbf{A} called the *vector-potential* of the field. The remaining component A^0 , perhaps more familiarly denoted by φ , is called the *scalar potential* of the field. The *electric field strength* is then the 3-vector

$$\mathbf{E} = \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} - \text{grad} \varphi$$

while the *magnetic field strength* is by definition the 3-vector

$$\mathbf{H} = \text{curl} \mathbf{A}.$$

Finally the *electromagnetic field tensor* (F_{ik}) is defined by

$$F_{ik} = \frac{\partial A_k}{\partial x^i} - \frac{\partial A_i}{\partial x^k}$$

We are now ready for the definition of S_f : we set

$$S_f = a \int 2(E^2 - H^2) d^4x,$$

where $H^2 = \langle \mathbf{H}, \mathbf{H} \rangle$, $E^2 = \langle \mathbf{E}, \mathbf{E} \rangle$ are the (Euclidean) scalar squares of the 3-vectors \mathbf{H} and \mathbf{E} , where a is a constant (usually taken as $\frac{1}{16\pi c}$), and where with respect to the spatial co-ordinates x^1, x^2, x^3 , the integral is taken over the whole of 3-space , while with respect to the variable x^0 (proportional to the time) it is taken over the interval between two fixed instants. Noting that $F_{ik}^2 \equiv F_{ik} F^{ik} = 2(H^2 - E^2)$, and substituting for a its customary value, we have

$$S_f = -\frac{1}{16\pi c} \int 2(H^2 - E^2) d^4x = \frac{1}{16\pi c} \int F_{ik}^2 d^4x.$$

Putting this together with the formulae for S_m and S_{mf} we obtain the formula for the total action S of an electromagnetic field containing charged particles:

$$S = - \sum \int mcdl - \sum \int \frac{e}{c} A_k dx^k - \frac{1}{16\pi c} \int F_{ik}^2 d^4x.$$

We can sometimes regard the total charge as being distributed continuously throughout space. In this case the amount of charge contained in the 3-dimensional volume element $dV = dx^1 \wedge dx^2 \wedge dx^3$ is given by ρdV , where ρ denotes the point-density of charge (thus ρ depends on

x^1, x^2, x^3 and the time t). We may also parameterize the world -line in $\mathbb{R}_{1,3}^4$ of a (variable) point -charge by the time:

$$x^0 = ct, x^i = x^i(t), i = 1, 2, 3.$$

Then (dx^i/dt) is the 4-dimensional velocity vector of the point-charge, and it is natural to call the 4-vector (j^i) defined as follow the *current 4-vector*

$$j^i = \rho \frac{dx^i}{dt}$$

. The three spatial components of this 4-vector define the usual current 3-vector

$$\mathbf{j} = \rho v$$

where v is the charge velocity at the given point, while the component j^0 is just $c\rho$. Direct calculation shows that in terms of the current vector (j^i) the total action takes the form

$$S = - \sum \int mcdl - \frac{1}{c^2} \int A_8 j^i d^4x - \frac{1}{16\pi c} \int F_{ik}^2 d^4x.$$

Having defined the action S for an electromagnetic field we are now ready to show that it is an appropriate one, in the sense that Maxwell's equations for the field are just the Euler-Lagrange equations corresponding to S. Since we are interested only in the field , we may take the motions of the charges (i.e. the current) as predetermined, i.e. known in advance. Thus since we are, as it were, given the trajectories (i.e. world-lines) of the charges in advance, we can restrict our attention to the action $S = S_{mf} + S_f$. Our problem is therefore that of finding the conditions (in the form of the Euler-Lagrange equations) which the 4-potential (A_i) must satisfy for S to have an extreme value. Taking into account the assumption that in the term S_{mf} the current (j^i) is not to be regarded as subject to variation, we have that the corresponding Lagrangian is

$$L = L(A_i, \frac{\partial A_i}{\partial x^\alpha}) = -\frac{1}{c}(\frac{1}{c}j^i A_i + \frac{1}{16\pi}F_{ik}^2),$$

where $F_{ik} = \partial A_k / \partial x^i - \partial A_i / \partial x^k$. Looking into the Euler-Lagrange equations

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_\alpha} - \frac{\partial L}{\partial q_\alpha} = 0$$

After some calculation we get

$$\frac{\partial F^{ik}}{\partial x^k} = -\frac{4\pi}{c}j^i, i = 0, 1, 2, 3$$

2.2 From Electrodynamics to Yang-Mills Theory

In order to describe the two charge states of strong interaction holding together nucleons in atomic nuclei. Yang-Mills generalized the gauge invariance of electromagnetism to explore the possibility of requiring all interactions of the isotopic spin at all space-time points. In analogy to the case of electrodynamics, let ψ be a two-component wave function with the isotopic spin $\frac{1}{2}$. Under an isotopic gauge transformation

$$\psi = S\tilde{\psi}$$

As discussed above we know

$$D_\mu = \partial_\mu - i\epsilon B_\mu$$

with B_μ satisfies

$$\tilde{B}_\mu = S^{-1}B_\mu S + \frac{i}{\epsilon}S^{-1}\frac{\partial S}{\partial x_\mu}$$

Again in analogy to the procedure of obtaining gauge invariant field strengths in the electromagnetism case, Yang-Mills suggest that

$$F_{\mu\nu} = \frac{\partial B_\mu}{\partial x_\nu} - \frac{\partial B_\nu}{\partial x_\mu} + i\epsilon[B_\mu, B_\nu]$$

instead of

$$F_{\mu\nu} = \frac{\partial B_\mu}{\partial x_\nu} - \frac{\partial B_\nu}{\partial x_\mu}$$

in the former case, in order that $F_{\mu\nu}$ satisfies

$$\tilde{F}_{\mu\nu} = S^{-1}F_{\mu\nu}S$$

which also comes from the invariance of the Lagrangian. The form of $F_{\mu\nu}$ that Yang-Mills suggested was thought to be the most brilliant and powerful step in Yang-Mills theory.

So why $F_{\mu\nu}$ can be chosen in that way? Is it **natural**? We now consider what $F_{\mu\nu}$ should look like in order that $\tilde{F}_{\mu\nu} = S^{-1}F_{\mu\nu}S$. We suppose the form

$$F_{\mu\nu} = \frac{\partial B_\mu}{\partial x_\nu} - \frac{\partial B_\nu}{\partial x_\mu} + G_{\mu\nu} \equiv H_{\mu\nu} + G_{\mu\nu}$$

Then it requires that

$$\tilde{G}_{\mu\nu} - S^{-1}G_{\mu\nu}S = S^{-1}H_{\mu\nu}S - \tilde{H}_{\mu\nu}$$

$$\begin{aligned} \partial_\nu \tilde{B}_\mu &= \partial_\nu(S^{-1}B_\mu S + \frac{i}{\epsilon}S^{-1}\frac{\partial S}{\partial x_\mu}) \\ &= (S^{-1})_\nu B_\mu S + S^{-1}\partial_\nu B_\mu S + S^{-1}B_\mu S_\nu + \frac{i}{\epsilon}(S^{-1})_\nu S_\mu + \frac{i}{\epsilon}S^{-1}S_{\mu\nu} \end{aligned}$$

$$\begin{aligned} \partial_\mu \tilde{B}_\nu &= \partial_\mu(S^{-1}B_\nu S + \frac{i}{\epsilon}S^{-1}\frac{\partial S}{\partial x_\nu}) \\ &= (S^{-1})_\mu B_\nu S + S^{-1}\partial_\mu B_\nu S + S^{-1}B_\nu S_\mu + \frac{i}{\epsilon}(S^{-1})_\mu S_\nu + \frac{i}{\epsilon}S^{-1}S_{\nu\mu} \end{aligned}$$

$$\begin{aligned} \tilde{H}_{\mu\nu} &= (S^{-1})_\nu B_\mu S - (S^{-1})_\mu B_\nu S + S^{-1}H_{\mu\nu}S \\ &\quad + S^{-1}B_\mu S_\nu - S^{-1}B_\nu S_\mu + \frac{i}{\epsilon}(S^{-1})_\nu S_\mu - \frac{i}{\epsilon}(S^{-1})_\mu S_\nu \end{aligned}$$

$$\begin{aligned} \tilde{H}_{\mu\nu} - S^{-1}H_{\mu\nu}S &= (S^{-1})_\nu B_\mu S - (S^{-1})_\mu B_\nu S \\ &\quad + S^{-1}B_\mu S_\nu - S^{-1}B_\nu S_\mu + \frac{i}{\epsilon}(S^{-1})_\nu S_\mu - \frac{i}{\epsilon}(S^{-1})_\mu S_\nu \\ &= S^{-1}G_{\mu\nu}S - \tilde{G}_{\mu\nu} \end{aligned}$$

It can be checked that when

$$G_{\mu\nu} = i\epsilon[B_\mu, B_\nu]$$

the transformation property holds, but it is apparent **there are still others which also work**, so in this sense that we choose $F_{\mu\nu}$ to be that form seems not so natural?

It can be seen that in the case of electrodynamics, in which $[B_\mu, B_\nu]$ vanishes, that is the difficulty generalizing the gauge invariance of electromagnetism to the particle physics since it is hard to imagine that there exist this special term which vanish in the case of electromagnetism but do not vanish in the latter case. Yang wrote in a commentary on his article as follows:

... Then I tried to define the field strengths $F_{\mu\nu}$ by $F_{\mu\nu} = \frac{\partial B_\mu}{\partial x_\nu} - \frac{\partial B_\nu}{\partial x_\mu}$, which was a “natural” generalization of electromagnetism. This led to a mess, and I had to give up. But the basic motivation remained attractive, and I came back to it several times in the next few years, always getting stuck at the same point ...

2.3 Yang-Mills equation

We now consider *Lagrangians* $L = L(A_\mu)$ corresponding to a gauge field A_μ alone (and to a prescribed matrix group G). Such a lagrangian must satisfy the following two conditions:

- (i) $L(A_\mu)$ is a scalar.
- (ii) $L(A_\mu)$ is invariant under gauge transformations.

The simplest functional satisfying these requirements has the form

$$L = -\frac{1}{4}g^{\mu\lambda}g^{\nu\chi} < F_{\mu\nu}, F_{\lambda\chi} >,$$

where $F_{\mu\nu}$ is the curvature form of the connection A_μ , $g_{\mu\nu}$ is an arbitrary metric on the region of interest of the underlying space, and $<, >$ denotes the Killing form on the Lie algebra of the group G, defined by

$$< X, Y > = -\text{tr}(ad_X, ad_Y),$$

ad_X being the linear transformation of the Lie algebra defined by $ad_X(A) = [X, A]$.

Suppose that the metric $g_{\mu\nu}$ is Euclidean or pseudo-Euclidean, i.e. that (after choosing coordinates suitably) $g_{\mu\nu} = \epsilon_\mu \delta_{\mu\nu} = \pm 1$, then

$$S[A_\mu] = \int -\frac{1}{4} < F_{\mu\nu}, F_{\mu\nu} > d^n x,$$

where here the subscripts μ, ν are summed over and with the signs $\epsilon_\mu = \pm 1$ taken into account.

Theorem 1. *The extremals of the previous functional satisfy the equations*

$$\nabla_\mu F_{\mu\nu} = 0$$

where μ is summed over.

Gauge fields corresponding to the group $SU(2)$ are generally known as *Yang-Mills fields* and the previous equations as *Yang-Mills equations*.

3 Natural from Mathematical Viewpoint?

Gauge theories are usually discussed in the language of differential geometry. Mathematically, a gauge is just a choice of a (local) section of some principal bundle. A gauge transformation is just a transformation between two such sections.

Physical Language	Mathematical Language
Gauge	Section
Gauge Potential	Connection Form
Field Strength	Curvature Form

To make things clear, we express the above ideas in mathematical terminology. that is, to use concepts from the theory of fiber bundle and connections.

Let M be a manifold of dimension n , oriented. Let P the bundle of its frames and $\pi : P \rightarrow M$ is the projection. Given a section $s : M \ni U \rightarrow P$ of the bundle, to a frame field $s = \{s_i\}_{1 \leq i \leq n}$ the connection is given by an matrix of 1-forms

$$\omega = (\omega_i^j)$$

such that

$$\nabla s = sw$$

Suppose $\tilde{s} = \{\tilde{s}_i\}_{1 \leq i \leq n}$ is another frame on the bundle. Let

$$\tilde{s} = sg$$

where $g(x) \in G$ and the elements in the matrix $g(x)$ are all smooth functions on U . We can easily get

$$g\tilde{w} = dg + wg$$

then take the exterior differential,

$$d(g\tilde{w}) = gd\tilde{w} + dg \wedge \tilde{w} = d(dg) + dwg - w \wedge dg$$

which implies that

$$g(d\tilde{w} + \tilde{w} \wedge \tilde{w}) = (dw + w \wedge w)g$$

Thus it is natural to define

$$\Omega = dw + w \wedge w$$

we obtain

$$\tilde{\Omega} = g^{-1}\Omega g$$

$\Omega = (\Omega_i^j)$ is called the curvature of the connection.

Now we take a look at what the above formulae stand for in physics. If we treat the a gauge as choice of a (local) section of some principal bundle, then we have the following corresponding expressions

$$w_\mu = -i\epsilon B_\mu$$

in which $w = w_\mu dx^\mu$. Then the transformation $g\tilde{w} = dg + wg$ can be describe as

$$\tilde{B}_\mu = g^{-1}B_\mu g + \frac{i}{\epsilon}g^{-1}dg$$

which is consistent with section 2.2. Furthermore,

$$\begin{aligned} \Omega_{\mu\nu} &= \frac{\partial w_\nu}{\partial x_\mu} - \frac{\partial w_\mu}{\partial x_\nu} + [w_\mu, w_\nu] \\ &= -i\epsilon\left(\frac{\partial B_\nu}{\partial x_\mu} - \frac{\partial B_\mu}{\partial x_\nu}\right) - \epsilon^2[B_\mu, B_\nu] \\ &= i\epsilon\left(\frac{\partial B_\mu}{\partial x_\nu} - \frac{\partial B_\nu}{\partial x_\mu} + i\epsilon[B_\mu, B_\nu]\right) \end{aligned}$$

Thus the way that Yang-Mills defined the field strength by

$$F_{\mu\nu} = \left(\frac{\partial B_\mu}{\partial x_\nu} - \frac{\partial B_\nu}{\partial x_\mu}\right) + i\epsilon[B_\mu, B_\nu]$$

with the transformation law

$$\tilde{F} = g^{-1}Fg$$

now seems more natural from mathematics viewpoint.

References

- [1] Wikipedia. en.wikipedia.org.
- [2] C.N.Yang, R.L Mills. *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phy.Rev (1954).
- [3] George Svelichny. *Preparation for Gauge Theory*.
- [4] B.A.Dubrovin, Fomentko, Novikov. *Modern Geometry methods and Applications*. Springer-Verlag
- [5] Shoshichi Kobayashi, Katsumi Nomizu. *Foundation of differential geometry, volume 1* 1963, Interscience publishers.
- [6] Chris J Isham. *Modern differential geometry for physicists* World Scientific.
- [7] C.N.Yang. *Integral Formulation for gauge fields* Phys.Rev,33,7(1974).
- [8] Ryoyu Utiyama. *Invariant theoretical interpretation of interaction* Phys.Rev,101,5(1956).
- [9] Jian Zhou. *Lecture Notes on Differentiable manifolds*.
- [10] Alexandre Guay. *Geometrical aspects of local gauge symmetry*, 2004.
- [11] Hou Boyuan,Hou Boyu. *Differential Geormert of Physicsits (Send Edition)*Sicentific Press, 2004.
- [12] S.S.Chern, Chen Weihuan. *Lecture Notes on Differential Geometry(Sencond Edition)*Peking University Press, 2001.

Motto

Read Euler, read Euler. He is the master of us all

—Pierre-Simon Laplace

Pauca sed matura. (Few, but ripe.)

—Carl Friedrich Gauss

Cauchy 问题 $\frac{dy}{dx} = x^2 + y^2, y(0) = 0$ 的最大存在区间

李超*

2007 年 12 月

编者按：本文是李超同学对杨利军老师在“常微分方程”课上提出的问题的探讨。我们在此文后面一并附上杨利军老师的评注。

1 问题的具体化

我们的目的是求 Cauchy 问题

$$\frac{dy}{dx} = x^2 + y^2, y(0) = 0 \quad (1)$$

的饱和解的最大存在区间。在课上已经证明问题 (1) 的最大存在区间有限，由于解曲线是关于原点对称的，故我们不妨只考虑右行的饱和解，设其最大存在区间为 $[0, \beta]$ ，其中 $\beta < +\infty$ 。

2 Bessel 方程与 Bessel 函数

Bessel 函数是下列 Bessel 方程

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (x^2 - \nu^2)y = 0 \quad (2)$$

的解。

Bessel 方程的两个线性无关的解为 [1]:

$$J_\nu(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{\Gamma(\nu + k + 1)\Gamma(k + 1)} \left(\frac{x}{2}\right)^{2k+\nu} \quad (3)$$

与

$$Y_\nu(x) = \frac{J_\nu(x) \cos \nu \pi - J_{-\nu}(x)}{\sin \nu \pi}. \quad (4)$$

其中 $\Gamma(x)$ 为 Euler Gamma 函数。

$J_\nu(x)$ 称为 ν 阶的第一类 Bessel 函数，当 $\nu \geq 0$ 或 ν 为负整数时， J_ν 在整个 \mathbb{R} 上有定义，而当 $\nu < 0$ 且不为整数时， J_ν 在 $(0, +\infty)$ 上有定义，而在 0 处附近无界。

$Y_\nu(x)$ 称为 ν 阶的第二类 Bessel 函数（或 Neumann 函数）， $Y_\nu(x)$ 在 $(0, +\infty)$ 上有定义，而在 0 处无界。当 ν 为整数时，右端为不定式，此时定义应被理解为关于 ν 的极限。

* 基科 61

方程

$$x^2 \frac{d^2y}{dx^2} + (1 - 2a)x \frac{dy}{dx} + (b^2 c^2 x^{2c} + (a^2 - c^2 \nu^2))y = 0 \quad (5)$$

可以通过作变换 $y = x^a z$, $t = bx^c$ 转化为

$$t^2 \frac{d^2z}{dt^2} + t \frac{dz}{dt} + (t^2 - \nu^2)z = 0. \quad (6)$$

而方程 (6) 正为 Bessel 方程的形式, 故方程 (5) 的通解为

$$y(x) = x^a (c_1 J_\nu(bx^c) + c_2 Y_\nu(bx^c)), \quad (7)$$

其中 c_1, c_2 为任意常数.

3 问题的转化

我们作变换

$$u(x) = e^{-\int_0^x y(s)ds}, \quad (8)$$

将方程 (1) 化为简单形式

$$\frac{d^2u}{dx^2} + x^2 u = 0, u(0) = 1, u'(0) = 0, \quad (9)$$

两边同乘 x^2 , 得到

$$x^2 \frac{d^2u}{dx^2} + x^4 u = 0. \quad (10)$$

这恰巧是在方程 (5) 中取 $a = \frac{1}{2}$, $b = \frac{1}{2}$, $c = 2$, $\nu = \frac{1}{4}$ 得到的特殊情形. 从而由 (7) 可得方程 (10) 的通解为

$$u(x) = x^{\frac{1}{2}} \left(c_1 J_{\frac{1}{4}} \left(\frac{1}{2}x^2 \right) + c_2 Y_{\frac{1}{4}} \left(\frac{1}{2}x^2 \right) \right), \quad (11)$$

且通过计算可得

$$u'(x) = x^{\frac{3}{2}} \left(c_1 J_{-\frac{3}{4}} \left(\frac{1}{2}x^2 \right) + c_2 Y_{-\frac{3}{4}} \left(\frac{1}{2}x^2 \right) \right). \quad (12)$$

利用初值条件 $u(0) = 1, u'(0) = 0$, 得到

$$u(x) = cx^{\frac{1}{2}} \left(J_{\frac{1}{4}} \left(\frac{1}{2}x^2 \right) - Y_{\frac{1}{4}} \left(\frac{1}{2}x^2 \right) \right), \quad (13)$$

$$u'(x) = cx^{\frac{3}{2}} \left(J_{-\frac{3}{4}} \left(\frac{1}{2}x^2 \right) - Y_{-\frac{3}{4}} \left(\frac{1}{2}x^2 \right) \right). \quad (14)$$

其中常数 $c = \frac{2}{\Gamma(\frac{3}{4})}$.

再利用

$$J_{-\nu} = J_\nu(x) \cos \nu x - Y_\nu(x) \sin \nu x, \quad (15)$$

由方程 (8) 得到

$$\boxed{y(x) = -\frac{u'(x)}{u(x)} = \frac{x^{\frac{3}{2}} J_{\frac{3}{4}} \left(\frac{1}{2}x^2 \right)}{x^{\frac{1}{2}} J_{-\frac{1}{4}} \left(\frac{1}{2}x^2 \right)}}. \quad (16)$$

当然从推导过程也可以知道, 式 (16) 在 0 处的值是在极限意义下理解的.

4 求解最大存在区间

我们画出 $y(x)$ 表达式 (16) 分母的直观的图像来, 如图 1.

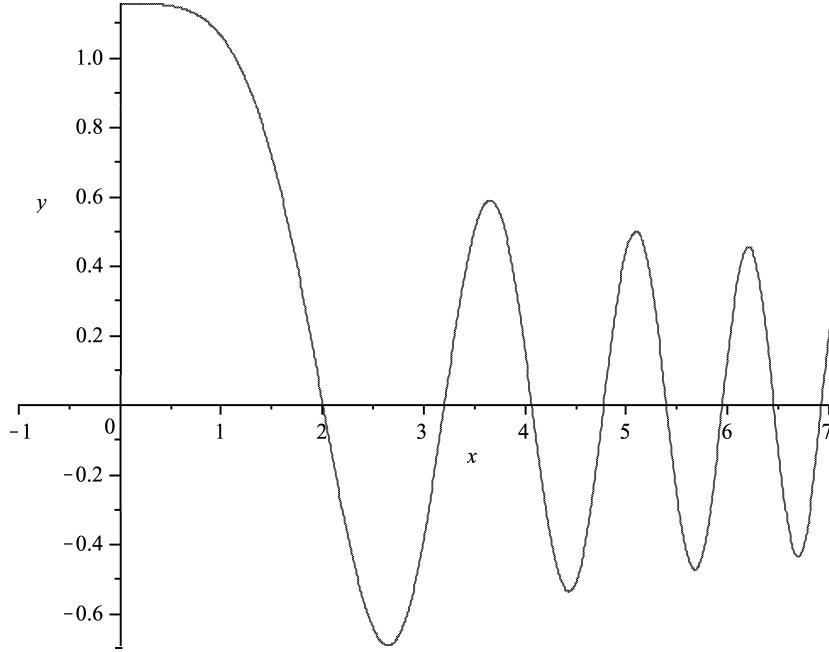


图 1 $g(x) = x^{\frac{1}{2}} J_{-\frac{1}{4}}(\frac{1}{2}x^2)$ 的图像

由式 (16), 自然的, 为求问题 (1) 的最大存在区间的右端点 β , 只需求出函数 $J_{-\frac{1}{4}}(\frac{1}{2}x^2) \mid_0$ 处向右的第一个零点即可. 更进一步的, 设 x_0 为 $-\frac{1}{4}$ 阶的第一类 Bessel 函数 $J_{-\frac{1}{4}}(x)$ 的第一个零点, 则有 $\beta = \sqrt{2x_0}$.

于是原问题最终归结为求解 Bessel 函数的零点问题了. 当然, 如果我们除了 Bessel 函数的定义式 (3) 之外对其性质一无所知的话, 这种化归是无助于事的. 幸好 Bessel 函数由于其广泛的应用而得到了前人的充分研究, 对其零点的性质也较为清楚. 例如 Watson 在 [3] 中用了将近 50 页的篇幅来讨论 Bessel 函数的零点. 我们在这里只简要地给出 Segura 在 [2] 中给出的求零点的方法. 这个方法也被软件 *Mathematica* 所采用, 参见 <http://library.wolfram.com/infocenter/MathSource/6777/>.

Segura 考虑函数

$$\mathcal{C}_\nu(x) = \cos \alpha J_\nu(x) - \sin \alpha Y_\nu(x), \quad (17)$$

$$H_\nu(x) = \frac{\mathcal{C}_\nu(x)}{\mathcal{C}_{\nu-1}(x)}, \quad (18)$$

$$f_\nu(x) = x^{2\nu-1} H_\nu(x). \quad (19)$$

Segura 证明了 $H_\nu(x)$ 和 $f_\nu(x)$ 及其导数在 $x > 0$ 时的单调性, 由此建立了基于 $f_\nu(x)$ 的 Newton 迭代法并证明了其收敛性, 从而求得对任意 ν, α 函数 $\mathcal{C}_\nu(x)$ 的第 k 个零点, 由此可求得 Bessel 的零点.

这里我们给出实际的精确到 200 位的计算结果.

```

In[1]:= N[ $\sqrt{2 \text{BesselJZero}\left[-\frac{1}{4}, 1\right]}, 200$ ]
Out[1]= 2.00314735942688470800461097905429922381014481722899...
          615650492862520734802244098959020931976087107741226...
          123800407945177342346641683567089788797604051459786...
          37607182307449868648326185010118032797752119152

```

图 2 *Mathematica* 6 求零点结果

```

Digits := 200;
fsolve(BesselJ(-1/4, 1/2*x^2), x, 0..5);
200
2.0031473594268847080046109790542992238101448172289961565...
049286252073480224409895902093197608710774122612380040...
794517734234664168356708978879760405145978637607182307...
449868648326185010118032797752119152

```

图 3 *Maple* 11 求零点结果

其中图 2 使用专门的 Bessel 零点函数, 而图 3 只用到普通的方程数值解的函数. 尽管求值的方法不同, 但我们可以看到两者给出的结果是完全一样, 相当可信.

由于我们对解曲线的未知, 通常只能给出解曲线的近似估计, 通过比较定理来得出最大存在区间的范围. 对解曲线了解的越清晰, 估计越精确, 得到的范围也就越准确. 在这个问题中, 我们完全确定了解曲线 (例如, 由式 (3) 和式 (16)), 从而完全确定了最大存在区间 $[0, \beta]$. 当然, 虽然可以得到数 β 任意精度的值, 但或许不能期望 β 有一个简明的表示 (哪怕五次的代数方程的零点也是如此), 正如常数 π 正可以理解为通过幂级数定义的函数 $\sin(x)$ 的零点一样.

在这个意义上可以说, 我们已经对问题 (1) 的解的最大存在区间 $[0, \beta]$ 有了一个较为精确的认识. 最后我们写出, 在小于 10^{-200} 的差距范围内:

$$\begin{aligned} \beta \approx & 2.00314735942688470800461097905429922381014481722899 \\ & 6156504928625207348022440989590209319760871077412261 \\ & 2380040794517734234664168356708978879760405145978637 \\ & 607182307449868648326185010118032797752119152. \end{aligned} \quad (20)$$

解曲线如图 4.

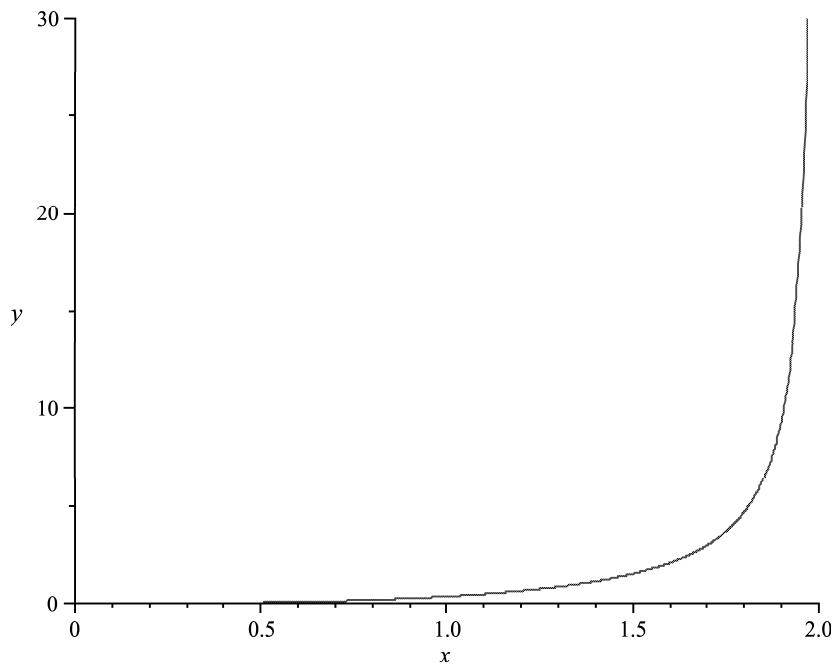


图 4 解曲线草图

参考文献

- [1] 系定平,贝塞尔函数, 北京: 高等教育出版社; 德国: 施普林格出版社, 1998.
- [2] J. Segura, *A global newton method for the zeros of cylinder functions*, Numerical Algorithms **18** (1998), 259-276.
- [3] G.N. Watson, *A treatise on the theory of bessel functions*, Cambridge University Press, 1995.

数学家趣闻

▲ 40 年代的时候, Michael Golomb 曾看见 Erdos 正和一位国际象棋高手 Nat Fine 下棋: ”Erdos 战胜对手的机会很少, 而且总是通过心理战术……我看 Nat 双手托着下颌, 仔细盘算着如何走下一步棋, 而 Erdos 却似乎在全神贯注地研究一本厚厚的医学大百科全书……我问他: ‘Paul, 你在干嘛呢? 你不是正在跟 Nat 下棋吗?’ 他回答说: ‘别打断我, 我正在证明一个定理。’”

关于李超同学作业的评注

杨利军

李超同学所提交的关于初值问题 $\dot{x} = t^2 + x^2, x(0) = 0$ 的饱和解的最大存在区间的估计, 解法新颖, 结论正确. 他所用到的一些工具, 是我在课堂上尚未介绍的. 他的解答组织得整齐清晰. 可以想见李超同学为解答这道题所下的功夫和所付出的努力. 这值得大力表彰和肯定.

布置这道习题的目的是希望同学们学会利用比较定理, 估计解的存在区间的大小. 李超解法非常规方法, 其思想和步骤大致如下: 先将方程 $\dot{x} = t^2 + x^2$ 转化成二阶线性方程 $\ddot{u} + t^2 u = 0$. 然后再将这个二阶线性方程化为标准的 Bessel 方程 $t^2 \ddot{v} + t \dot{v} + (t^2 - p^2)v = 0$, 其通解可由 Bessel 函数表示. 可以证明原始初值问题解的最大存在区间的右端点, 对应着某个 Bessel 函数的最小正零点. 最后利用数值方法, 确定这个最小正零点.

关于第二次转化, 还值得进一步讨论. 因为二阶线性方程 $\ddot{u} + t^2 u = 0$ 看起来比 Bessel 方程 $t^2 \ddot{v} + t \dot{v} + (t^2 - p^2)v = 0$ 要简单得多, 因此这样的转化有化简为烦之嫌. 虽然关于 Bessel 函数已有不少的研究成果, 但这些成果目前看来尚不能直接用来解决我们的问题. 实际上, 正如有些同学所做的, 我们可以对原初值问题, 进行数值积分从而得到右侧最大存在区间的估计, 所得到的结果与李超解答类似.

我个人认为, 数值方法对于工程应用而言很重要, 但对于常微定性分析而言, 数值方法是最后的, 不得已的方法. 例如, 我们可以利用各种数值方法求出如下无穷级数的任意精度的近似值, (以下是 40 位小数的近似值):

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots = 1.6449340668482264364724151666460251892189 \cdots$$

但这远不如 Euler 在 1734 年所得到的结果

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} + \cdots = \frac{\pi^2}{6}$$

美妙. Euler 的这个结果被认为是历史上最伟大的定理之一, 参见科普书 “天才引导的历程”, 威廉邓纳姆著, 1990 (中译本). Euler 的这个结果有三个很值得一学的, 被认为是来自圣经 (即上帝) 的证明, 参见 “Proofs from THE BOOK”, Martin Aigner and Guenter M. Ziegler, Third Edition, 2004. 对有志于数学的各位同学, 我强烈地推荐这两本书.

为了使得李超同学的解答看起来更容易些, 以下介绍一些关于 Riccati 方程与二阶线性齐次方程之关系的一些结论. 这可以看成对李超解答的补充说明.

形如

$$\dot{x} = a(t)x^2 + b(t)x + c(t) \quad (1)$$

的一阶方程称为 Riccati 方程, 其中 $a(t), b(t)$ 和 $c(t)$ 通常假设为开区间上的连续函数. 习题中的方程 $\dot{x} = t^2 + x^2$ 是典型的 Riccati 方程. 我们回忆二阶线性齐次方程是指如下形式的方程

$$\ddot{u} + p(t)\dot{u} + q(t)u = 0 \quad (2)$$

其中 $p(t), q(t)$ 为开区间上的连续函数.

定理 1: 考虑 Riccati 方程 (1). 假设 $a(t)$ 在开区间 I 连续可微且 $a(t) \neq 0, \forall t \in I$, 则方程 (1) 则在变换

$$u(t) = \exp \left\{ - \int_{t_0}^t a(\tau)x(\tau)d\tau \right\}, \quad t, t_0 \in I \quad (3)$$

下, 可化为一个相应的二阶线性齐次方程 (2), 其系数函数 $p(t), q(t)$ 如下确定

$$p(t) = -b(t) - \dot{a}(t)/a(t), \quad q(t) = a(t)c(t). \quad (4)$$

这里”化为”的意思是: (i) 若 $x(t)$ 是 Riccati 方程 (1) 的解, 则由方程 (3) 所确定的二次连续可微的函数 $u(t)$ 是二阶线性齐次方程 (2) 的正解; (ii) 若 $u(t)$ 是 (2) 的正解, 则由变换 (3) 的逆变换所确定的连续可微的函数

$$x(t) = -\frac{\dot{u}(t)}{a(t)u(t)}, \quad t \in I \quad (5)$$

是 Riccati 方程 (1) 的解.

注: 这里要求 $u(t)$ 是正解并不构成一个限制, 因为若 $u(t)$ 是 (2) 的解, 则 $-u(t)$ 也是. 实际上, 若 $u(t)$ 是 (2) 的一个解, 则由方程 (5) 确定了若干个 Riccati 方程 (1) 在不同的开区间 I_k 上的不同的解, 而这些开区间 I_k 是由解 $u(t)$ 的零点隔离开区间 I 而得到的.

定理 1 的证明: (i) 设 $x(t)$ 是 Riccati 方程 (1) 的解, 函数 $u(t)$ 是由方程 (3) 所确定, 或等价地, $x(t)$ 可由方程 (5) 来表示. 对方程 (5) 求导得

$$\dot{x} = -\frac{\ddot{u}}{au} + \frac{\dot{u}(\dot{a}u + a\dot{u})}{(au)^2}.$$

另一方面,

$$\dot{x} = ax^2 + bx + c = a\left(\frac{-\dot{u}}{au}\right)^2 - b\frac{\dot{u}}{au} + c.$$

对上述两个方程稍加整理即可得到

$$\ddot{u} + (-\dot{a}/a - b)\dot{u} + cau = 0. \quad (6)$$

因此结论 (i) 成立.

(ii) 设 $u(t)$ 是方程 (6) 的解, 函数 $x(t)$ 是由方程 (5) 所确定. 对方程 (5) 求导得

$$\begin{aligned} \dot{x} &= -\frac{\ddot{u}}{au} + \frac{\dot{u}(\dot{a}u + a\dot{u})}{(au)^2}. \\ &= \frac{1}{au} \left((-\dot{a}/a - b)\dot{u} + cau \right) + \frac{\dot{u}(a\dot{u} + \dot{a}u)}{(au)^2} \\ &= c + (\dot{a}/a + b)x + ax^2 - \dot{a}x/a = ax^2 + bx + c. \end{aligned}$$

结论 (ii) 成立. #

同理, 二阶线性齐次方程 (2) 可化为 Riccati 方程 (1).

定理 2: 二阶线性齐次方程 (2) 在变换 (5) 下, 可化为相对应 Riccati 方程 (1), 其系数函数为

$$a(t) = 1, \quad b(t) = -p(t), \quad c(t) = q(t), \quad (7)$$

这里”化为”的意义同定理 1.

证明: 证明思想同定理 1, 但更简单. 略. #

对于习题中的 Riccati 方程 $\dot{x} = t^2 + x^2$, 其对应的二阶线性齐次方程为 $\ddot{u} + t^2u = 0$. 两个方程的解有如下关系:

$$x(t) = -\frac{\dot{u}(t)}{u(t)} \quad \text{或} \quad u(t) = \exp \left\{ - \int_0^t x(\tau) d\tau \right\} \quad (8)$$

这里需要需要注意的是, 两个解的定义区间通常是不一样的. 解 $x(t)$ 的定义区间为 $u(t)$ 不为零的连通的开区间.

根据上述定理，我们不难得得到如下结论.

定理 3: 一阶 Riccati 方程初值问题

$$\dot{x} = t^2 + x^2, \quad x(0) = 0 \quad (9)$$

饱和解的右侧最大存在区间的右端点是二阶线性齐次方程初值问题

$$\ddot{u} + t^2 u = 0, \quad u(0) = 1, \dot{u}(0) = 0 \quad (10)$$

解的最小正零点.

证明: 记问题 (10) 的解为 $u(t)$. 根据定理 1 知, $x(t) = -\dot{u}(t)/u(t)$ 是 Riccati 方程 $\dot{x} = t^2 + x^2$ 的解, 其初值条件为 $x(0) = 0$, 即 $x(t) = -\dot{u}(t)/u(t)$ 是问题 (9) 的解. 由此可见, 解 $x(t)$ 右侧最大存在区间的右端点, 正是 $u(t)$ 的最小正零点. 证毕. #

注 1: 对于初值问题 (10) 解的最小零点的确定, 李超同学的方法是再做进一步的转化, 转化为 Bessel 函数的零点问题. 如前所述, 这种转化是一种可行的, 但不一定是最理想的. 关于 Bessel 方程以及 Bessel 函数, 因为涉及的内容较多, 很难用简短的篇幅说清楚, 我这里就不细谈了. 有兴趣的同学可参阅丁同仁李承治的书, §7.5, page 236.

注 2: 若将问题 (10) 的初值条件 $u(0) = 1$ 中的 1 可以换成任何非零数 $u_0 \neq 0$, 即 $u(0) = u_0$. 这是因为相应的解的零点个数及其位置不变.

注 3: 设 $u(t)$ 是问题 (10) 的解, 再记 $v(t)$ 为二阶线性齐次方程 $\ddot{v} + t^2 v = 0$ 满足初值条件 $v(0) = 0, \dot{v}(0) = 1$ 的解, 则 Riccati 方程 $\dot{x} = t^2 + x^2$ 的满足初值条件为 $x(0) = \mu$ 的解 $x(t, \mu)$ 可表为 $x(t, \mu) = -\dot{u}(t, \mu)/u(t, \mu)$, $u(t, \mu) := u(t) - \mu v(t)$.

注 4: 请有兴趣的同学进一步思考, 如何解析的估计或确定初值问题 (10) 解的最小正零点?
评注完毕.

数学名言

数统治着宇宙。

—[古希腊] 毕达哥拉斯

数学, 科学的女皇; 数论, 数学的女皇。

—[德] C · F · 高斯

你, 自然, 我的女神, 我要为你的规律而献身。

—[德] C · F · 高斯

无限! 再没有其他的问题如此深刻地打动过人类的心灵。

—[德] D · 希尔伯特

机票预售价格和策略的数学模型

吴玉清 孟昭时 郭晓江

编者按:本文是2006年清华大学数学建模竞赛校内赛的论文,在该赛事中本文获得第一名。发表到本刊时作者做了删改。

摘要

本文对机票价格的波动规律以及对不同飞机型号时预售的票数进行了建模求解并对结果作了详细的讨论。

在问题1中,我们由粗到细分别建立了独立数据拟合模型、非平稳时间序列模型,并对后者采用了确定性和随机性两种分析方法。在随机性分析中,观察到数据有很强的周期性,我们建立ARIMA(求和自回归移动平均)模型^[1]。首先经过对序列的两次差分运算,将序列化为平稳序列,然后考察该平稳序列的自相关系数,偏自相关系数并对其进行白噪声检验,再用BIC准则定阶,建立比较合适的平稳序列的ARMA(自回归滑动平均)模型,最后用SAS软件进行拟合。对机票价格序列拟合的模型为: $g_t = g_{t-1} + g_{t-7} - g_{t-8} + \varepsilon_t + 0.30956\varepsilon_{t-1}$ ($t=9, 10, 11\dots$) 用前九周的数据对第十周的数据进行检验得到误差基本小于5%,这说明了该模型很好地符合了机票价格变动规律。我们用前十周的数据预测后两周的机票价格如下表:

	周一	周二	周三	周四	周五	周六	周日
第11周	500	500	500	590	590	490	490
第12周	490	490	490	580	580	480	480

在问题2中,我们首先提出并分析了一个一般性的基本模型:已知客舱的容量 n ,求最佳预订票数 m 。通过Matlab数值计算此模型得到了很好地解决。通过计算,120经济舱座位和8头等舱座位条件下得出公司纯利润为固定损耗的67%。对于有三个客舱并优先考虑头等舱和公务舱利益的情况,提出了优先级原则,降级原则,又提出一种新的机票发售方式并证明了这种方式和经典的机票发售方式等价,从而将问题分解为3个上面基本模型的解,最终得到航空公司的纯利润值占成本的66%~68%。以波音747飞机为例,求出了预售票方式的最佳解。随后,又详尽分析了公司利润对参数 b/g 和顾客误机概率 p 的灵敏度,绘出了不同的 p 下三个客舱最佳预售票张数的曲线。通过上述分析计算,定性的给出了航空公司在不同情况下应采取的措施,提出了自己的建议。

本文在最后比较了问题1中三种模型的优缺点并在问题2中提出并解决了一个推广的模型并通过化归方法解决,这两个讨论都很有实用和普遍价值。

(一) 问题的重述

问题背景:某航空公司每天都有一班航班飞往某著名旅游城市,该航班的所有机票均通过互联网来预订,这种预订具有很大的不确定性,旅客很可能由于各种原因取消预订或误机。航空公司制定了一系列的条约。首先,航空公司采用浮动价格在网上进行机票销售,其价格随市场需求情况进行调整,一般来说周末价格会比较高些。其次,合约规定:要求旅客在网上预订机票时先用信用卡预付机票价格的20%作为定金。如果旅客在航班起飞前24小时前就取消预订,

定金将如数退还，否则（包括误机）定金将不再退还。

在旅游旺季，航空公司往往可以在网上预订出超过飞机容量的机票数，以减低旅客由于取消预订或误机时给航空公司所带来的损失。当然这样做可能会带来新的风险，因为万一届时有超出飞机容量的订票旅客出现，将有一部分旅客不能登机，这势必导致航空公司的声誉受损，并引起纠纷。因此，航空公司承诺，那些预先订购机票但不能飞走的旅客，可以不加任何费用乘坐下一班机或者退订，并付给一定的赔偿金。

问题一：试根据所给的十周内预售机票价格建立机票预售价格的数学模型，用模型说明价格变动的规律，并据此估计今后2周的经济舱机票销售的参考价格。

问题二：用小飞机载客的时候，如何确定经济舱和头等舱的预定票数量使得航空公司的总收益最大？

问题三：改用大飞机执行该航班任务时，如何制定合适的登机策略才能合理的保障头等舱和公务舱旅客的权益？在该登机策略下如何确定三种票的预售数量使得航空公司获得最大的利润？

（二）基本假设

1. 在问题一中，我们假设不会出现突发事件而导致机票价格或订票乘客的数量有突然的改变；
2. 在问题二中，用小飞机载客时，我们假设不充分保证头等舱顾客的利益，即如果头等舱已经坐满，多出的持头等舱预订票的乘客不能进入经济舱，只能等待下一班飞机并获得赔偿；
3. 在问题二中，用大飞机载客时，我们假设头等舱和公务舱的乘客的利益能够得到充分的保证，即如果头等舱或公务舱已经坐满，那么持相应的预订票的乘客可以进入下一级的机舱并保证有座位的同时获得赔偿，无论下一级的机舱是否坐满；而持经济舱预订票的顾客如果所有座位都已经坐满，则要等待下一班飞机；
4. 我们假设这个模型在旺季中建立，即所有预售票都能售出。
5. 假设公务舱票价是经济舱的1.3倍，这一点可以从网上实际数据中得到验证。

（三）本文用到的部分符号及说明

t ：时间

g ：飞机票票价

S_k ：周期指数

n ：座位个数

b ：造成超员赔偿给每位乘客的钱

s ：单次飞行公司的收益

r ：飞行中的固定损耗

p ：买票的人不能登机的概率

m ：预售的票数

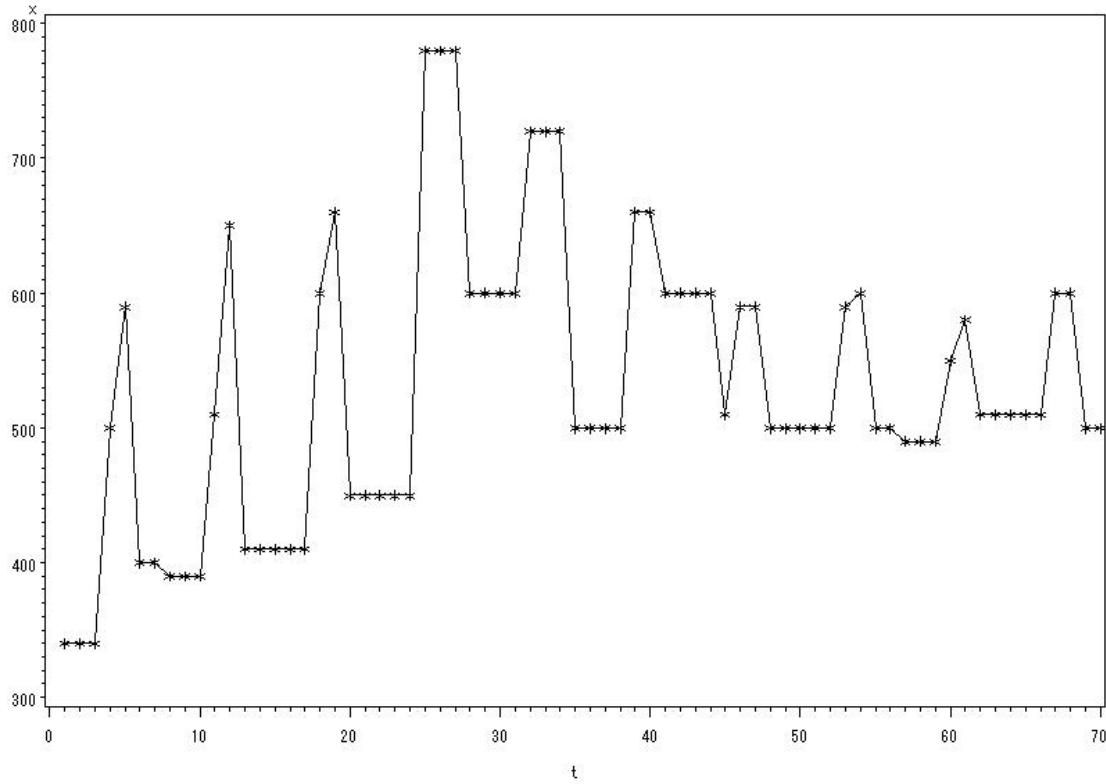
J ：单次飞行公司所得到的纯利润占固定损耗的比例

bg ：超员赔偿给顾客的钱和机票价格的比，即 $bg=b/g$

其他的符号会在所出现处作具体说明。

(四) 问题一模型的建立、求解与检验

我们作出价格的时序图如下：



图一

观察可知价格随着时间的变化以 7 天为周期有明显的周期性波动趋势，同时每周的平均价格也有一个波动趋势，对这种比较复杂的变化规律，我们尝试用三种模型来分析和预测。

模型一： (略)

模型二： 非平稳时间序列确定性分析模型(略)

模型三： 非平稳序列的随机分析模型。

1. 模型的建立：我们考虑求和自回归移动平均模型^[1]，简记为 ARIMA(p, d, q) 模型：

$$\begin{cases} \Phi(B)\nabla^d g_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(g_s \varepsilon_t) = 0, \forall s < t \end{cases} \quad (*)$$

式中：

∇ 为差分算子， B 为延迟算子： $\nabla x_t = x_t - x_{t-1}$ ， $B x_t = x_{t-1}$ ， 则 $\nabla^d = (1-B)^d$

$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ ， 为平稳可逆 ARMA(p, q) 模型的自回归系数多项式

$\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ ， 为平稳可逆 ARMA(p, q) 模型的移动平滑系数多项式

(*) 也可简记为：

$$\nabla^d g_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t, \text{ 式中, } \{\varepsilon_t\} \text{ 为零均值白噪声序列。}$$

2. 用 SAS 软件对模型进行求解：

考虑到机票的价格的波动以 7 天为周期, 所以对原序列作 7 步差分, 差分后的时序图如图 8。

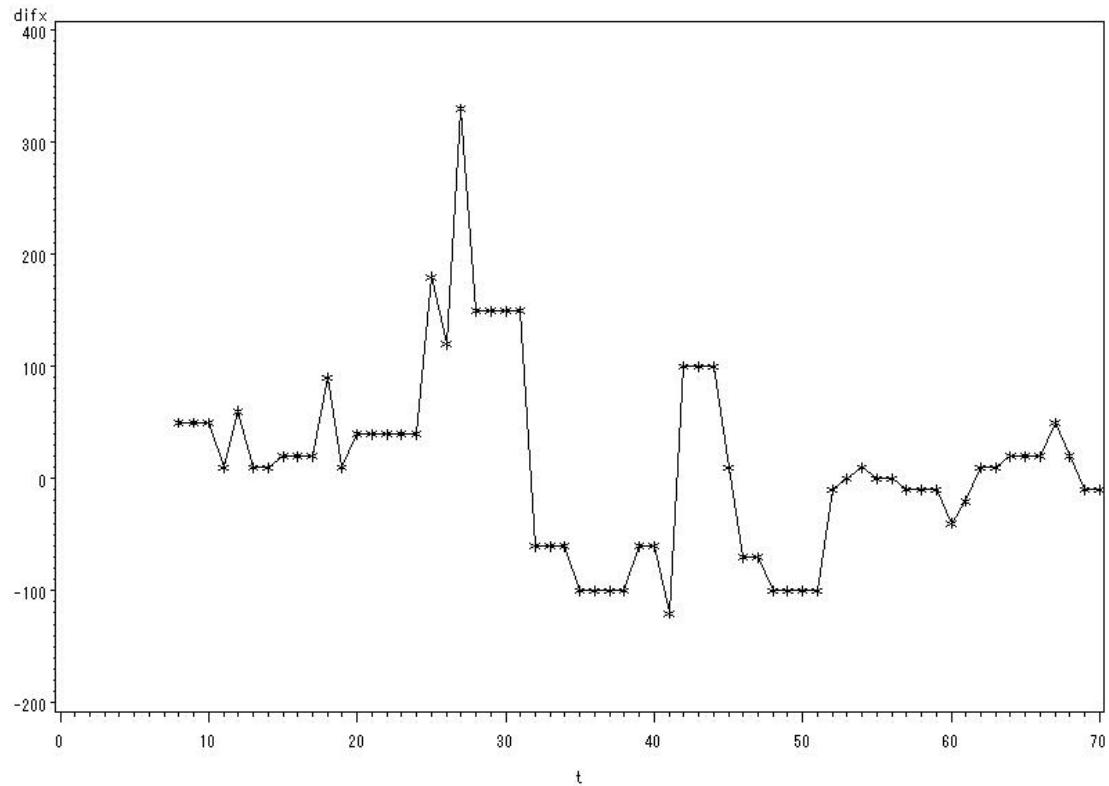


图 8

可见此时周期性已经被消除了，但是还有长期趋势，于是再做一阶差分，如图 9 所示。

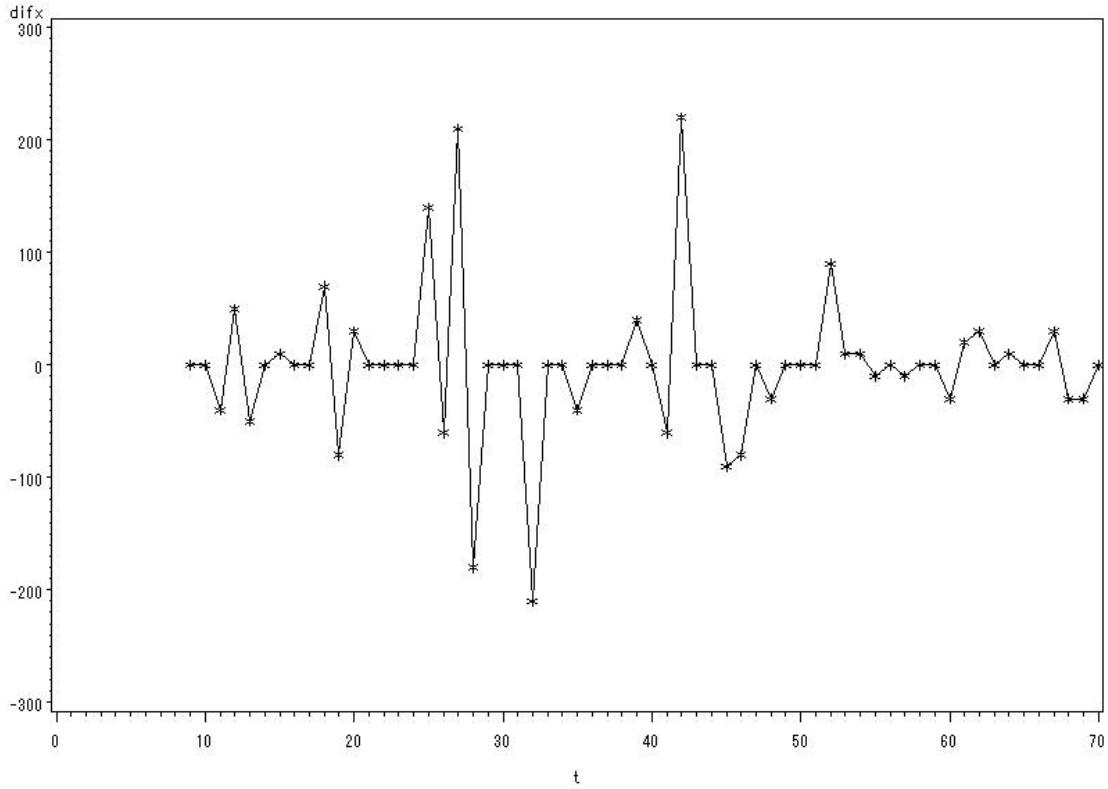


图 9

此时的时序图已无显著趋势或周期，随机波动比较平稳。可以用平稳序列 ARMA 模型进行处理。对差分后的序列计算 ACF（自相关系数），PACF（偏自相关系数）如下图所示：

Autocorrelations																								
Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	Std Error
0	3979.709	1.00000																						0
1	-1233.210	-.30987												*****	.									0.127000
2	696.276	0.17496												*	***	.								0.138660
3	-388.131	-.09753												.	**	.								0.142176
4	308.993	0.07764												.	**	.								0.143251
5	-627.339	-.15763												.	***	.								0.143928
6	-145.096	-.03646												.	*	.								0.146686
7	-640.584	-.16096												***	.									0.146832
8	-543.825	-.13665												***	.									0.149651
9	721.820	0.18138												.	****	.								0.151650
10	-586.051	-.14726												.	***	.								0.155110
11	-106.930	-.02687												.	*	.								0.157348
12	23.231849	0.00584												.										0.157422
13	874.830	0.21982												****	.									0.157426
14	-981.481	-.24662												*****	.									0.162301
15	1158.826	0.29118												.	*****	.								0.168237

"." marks two standard errors

图 10

The ARIMA Procedure																						
Partial Autocorrelations																						
Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1
1	-0.30987																					
2	0.08732																					
3	-0.02340																					
4	0.03222																					
5	-0.12927																					
6	-0.14266																					
7	-0.20100																					
8	-0.27788																					
9	0.11128																					
10	-0.06279																					
11	-0.20197																					
12	-0.12215																					
13	0.12425																					
14	-0.24942																					
15	0.06284																					

Autocorrelation Check for White Noise												
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations								
8	11.15	8	0.0839	-0.310	0.175	-0.098	0.078	-0.158	-0.036			
12	18.56	12	0.0996	-0.161	-0.137	0.181	-0.147	-0.027	0.006			

图 11

从白噪声检测结果来看，各阶延迟下 LB 统计量的 P 值都显著大于 0.05，可以认为这个拟合模型的残差序列属于白噪声序列。而从自相关系数和偏自相关系数图中可以看出，自相关系数显示出明显的一阶截尾性质，而偏自相关系数显示出明显的拖尾性，因此可以尝试使用 AR(1) 模型来拟合。为了确保模型的最优化，用 Minic 命令来计算阶数小于等于 5 的 ARMA(p, q) 模型的 BIC 信息量，并找出信息量达到最小的模型的阶数，输出结果如下：

Minimum Information Criterion						
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	8.261645	8.238426	8.266045	8.294116	8.360544	8.386607
AR 1	8.234224	8.296212	8.332421	8.35895	8.425157	8.447387
AR 2	8.268551	8.310909	8.371131	8.417919	8.484182	8.513046
AR 3	8.316378	8.355571	8.420425	8.48331	8.549783	8.570875
AR 4	8.381432	8.419988	8.485703	8.537203	8.58739	8.635322
AR 5	8.432867	8.467554	8.534009	8.571139	8.637662	8.700965

Error series model: AR(8)
Minimum Table Value: BIC(1,0) = 8.234224

图 12

从结果可以看出用模型 AR(1, 0) 拟合为最优。使用条件最小二乘估计法，得到未知参数的估计值为：

Autoregressive Factors
Factor 1: 1 + 0.30956 B***(1)

图 13

即 $g_t = (1 + 0.30956B)\varepsilon_t$

综合考虑前面的差分运算，对机票价格序列拟合的模型为：

$$(1 - B)(1 - B^7)g_t = (1 + 0.30956B)\varepsilon_t$$

即 $g_t = g_{t-1} + g_{t-7} - g_{t-8} + \varepsilon_t + 0.30956\varepsilon_{t-1} \quad t = 9, 10, \dots$

3. 模型检验

我们用前九周的价格数据作拟合，并预测第十周的价格，列表如下：

表五 模型三对第十周价格的预测结果

	周一	周二	周三	周四	周五	周六	周日
预测值	500	500	500	560	590	520	520
实际值	510	510	510	600	600	500	500
误差(%)	1.96	1.96	1.96	6.67	1.67	4.00	4.00

从上表中可以看出，用 ARIMA 模型所得到的结果和实际值误差很小，基本在 5% 以内。因此，本模型是很优秀的，用来预测也是十分可信的。

将前十周的序列拟合值和实际值联合作图如下：

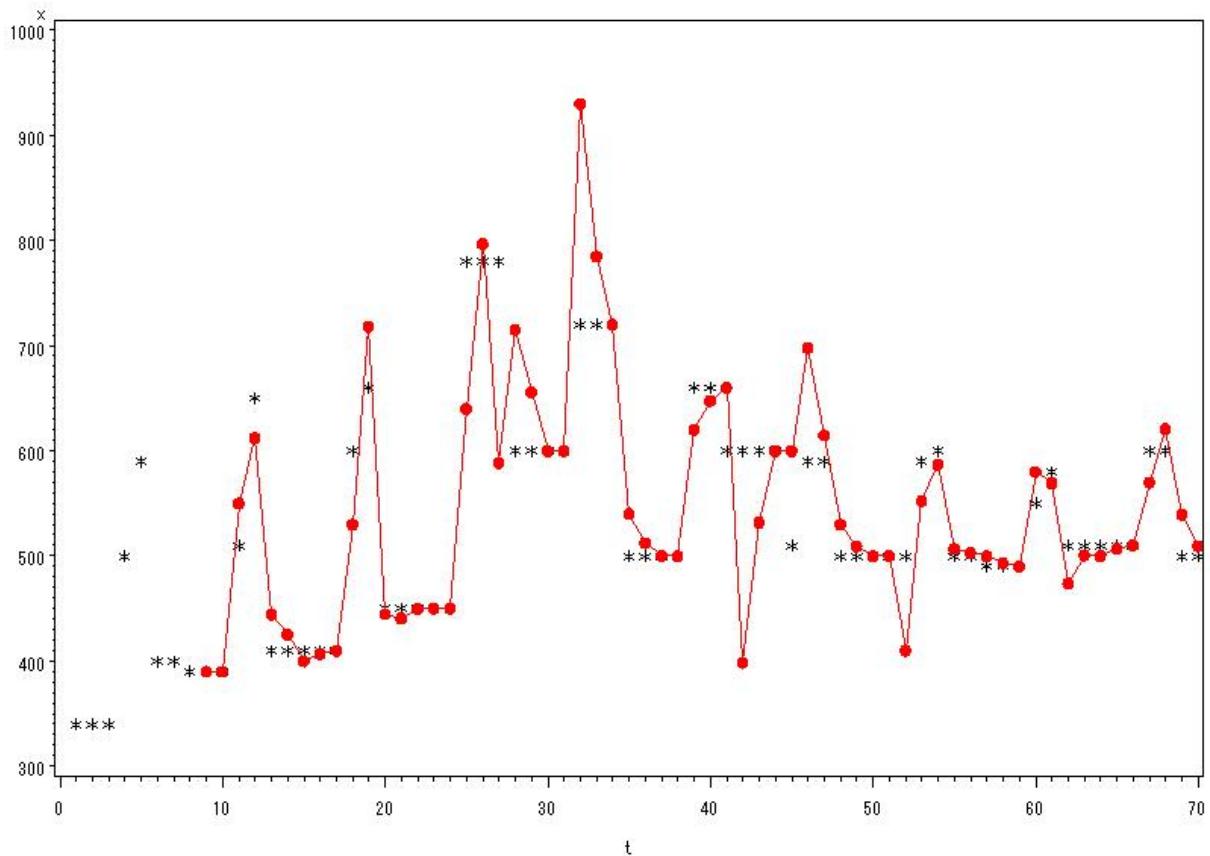


图 14
图中星号为实际值；曲线为拟合值。

从上图中可以看出，拟合曲线和每个点结合的也很不错

4. 预测

利用前十周的数据拟合后的模型对第十一，十二周进行预测，结果如下表：

表六 模型三对十一十二周价格的预测

Obs	Forecast	Std Error	95% Confidence	Limits
71	500.0000	60.4830	381.4556	618.5444
72	500.0000	73.4987	355.9452	644.0548
73	500.0000	87.5420	328.4208	671.5792
74	590.0000	98.7812	396.3925	783.6075
75	590.0000	109.1008	376.1664	803.8336
76	490.0000	118.4582	257.8263	722.1737
77	490.0000	127.1480	240.7945	739.2055
78	490.0000	165.9637	164.7170	815.2830
79	490.0000	187.8259	121.8681	858.1319
80	490.0000	209.9188	78.5666	901.4334
81	580.0000	229.1729	130.8294	1029.1706
82	580.0000	247.1375	95.6195	1064.3805
83	480.0000	263.8217	-37.0810	997.0810
84	480.0000	279.5293	-67.8674	1027.8674

把预测后的曲线和实际值画成联合曲线如下图：

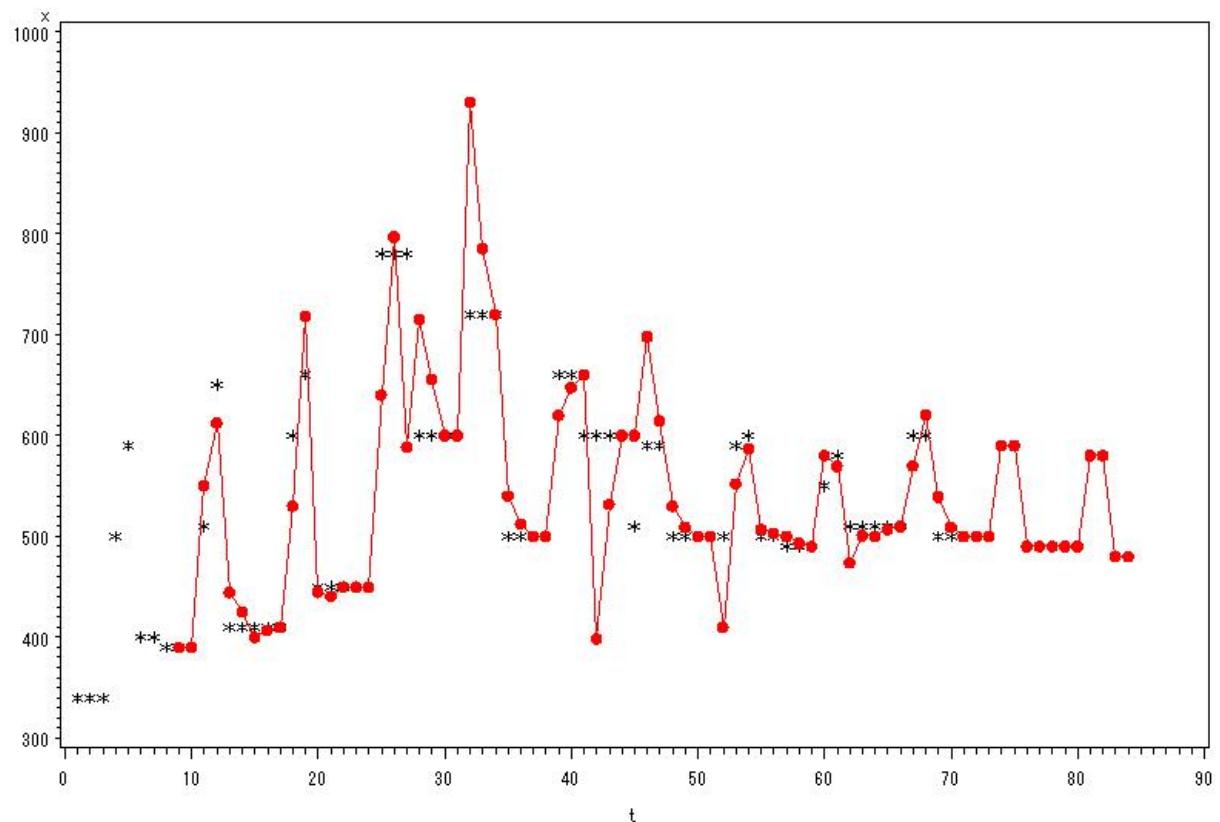


图 15

从整体的趋势来看，预测值是可信的。把预测结果整理如下：

表七

模型三对十一、十二周机票价格的预测结果

	周一	周二	周三	周四	周五	周六	周日
第 11 周	500	500	500	590	590	490	490
第 12 周	490	490	490	580	580	480	480

(五) 问题二模型的建立、求解与检验

A. 一个基本的模型的求解

本问题中可以抽象出一个基本的问题，即：给定座位个数 n ，票价均为 g ，买票的人不能登机的概率为 p ，如果造成超员赔偿给每位乘客的钱是 b ，预售出多少张票 m ，可以使利润最大？我们把这种模型称为 $P(p, bg, n)$ 模型（其中 $bg = b/g$ ，在后面的分析中我们可以看到，模型只和 b ， g 的比值有关）

问题的解：

公司的经济利益可以用总的收益 s 来衡量，设 r 为飞行中的固定损耗，当 m 位乘客中有 k 位不按时前来时：

$$s = \begin{cases} (m-k)g + \frac{1}{5}kg - r, & m-k \leq n \\ ng - r - (m-k-n)b + \frac{1}{5}kg, & m-k > n \end{cases} \quad (1)$$

不按时前来登机的乘客数 k 符合二项分布，于是概率为

$$p_k = C_m^k p^k q^{m-k}, \quad q = 1-p \quad (2)$$

收益 s 的期望为

$$s(m) = \sum_{k=0}^{m-n-1} \left(ng - r - (m-k-n)b + \frac{1}{5}kg \right) p_k + \sum_{k=m-n}^m \left((m-k)g + \frac{1}{5}kg - r \right) p_k \quad (3)$$

化简上式，并注意到 $\sum_{k=0}^m kp_k = mp$ ，可得

$$s(m) = qmg - (g+b) \sum_{k=0}^{m-n-1} (m-k-n)p_k - r \quad (4)$$

可以假设飞行中固定损耗 r 和满员的机票收入成一定的比例关系，即 $r = \lambda ng$ ，(4)式两端同时除以 r ，得到：

$$J(m) = \frac{s(m)}{r} = \frac{1}{\lambda n} \left[qm - (1 + \frac{b}{g}) \sum_{k=0}^{m-n-1} (m-k-n)p_k \right] - 1 \quad (5)$$

$J(m)$ 的经济意义是公司纯利润占固定损耗的比例。该模型没有解析解，但可以借助 Matlab 软件进行数值计算，求得最大值点。

B. 小飞机的情况

在这种情况下，根据假设不优先考虑头等舱乘客的利益（即如果头等舱已经坐满，多出的持头等舱票的乘客不能进入经济舱，只能等待下一班飞机并获得赔偿）。根据这个假设，我们便可以将经济舱和头等舱分开来考虑。利用参考文献中的数据^[2]，我们可以认为 $\lambda = 0.6$ ， $p = 0.04$ ， $b/g = 0.2$ 是相对固定的。

i. 经济舱情况。 $n_1 = 120$ ，将这组数带入基本模型中可求得当 $m = 127$ 时 $J(m)$ 有最大值为

$J_{1\max} = 0.6711$ ， m 和 $J(m)$ 关系图线如下：

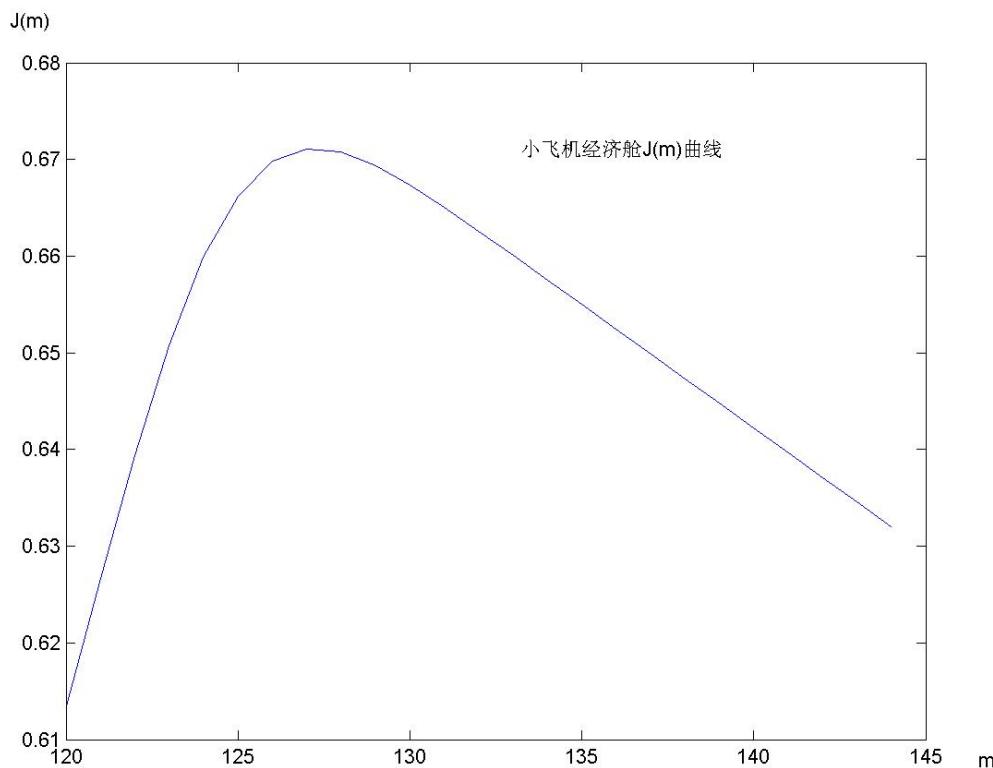


图 16

ii. 头等舱情况。和经济舱一样，取 $n = 8$ ， $b/g = 0.2/1.5 = 0.133$ 代入基本模型中可得 $m = 9$ ，

$J_{2\max} = 0.6515$

综上，小飞机总的收入最大是在发出 127 张经济舱票，9 张头等舱的票，总的收益为

$$J = \frac{(J_{1\max} \times n_1 g_1 + J_{2\max} \times n_2 g_2)}{n_1 g_1 + n_2 g_2} = 0.6693，\text{ 即纯收入为投入的 } 66.93\%。$$

C. 小飞机改成大飞机后的模型求解

我们假设经济舱、公务舱、头等舱的座位个数分别 n_1, n_2, n_3 ，并且满足关系 $n_1 \geq n_2 \geq n_3$ ，

票价分别为 g_1, g_2, g_3 ，航空公司发出的票分别为 m_1, m_2, m_3 张。为了叙述方便，我们定义 4 个优先级别：头等舱（3 级），公务舱（2 级），经济舱（1 级），被挤下飞机（0 级）。假设使顾客享受降一级的服务时赔偿为 b 。我们假设在高一级的舱位没有坐满时，低一级的乘客可以坐到高一级的舱位而不用额外付出代价（比如商务舱没有坐满而经济舱超员，当然可以让多出的经济舱的乘客坐到商务舱去，以避免付给赔偿金 b ），这符合公司追求利益最大化的原则，是合理而必要的。

1. 优先级原则。

对于大飞机模型，要求充分考虑头等舱和公务舱乘客的利益，即如果高一级的客舱已满，则多出来的持该舱票的乘客优先享受次一级的待遇并得到赔偿。例如，如果一个持公务舱票的乘客发现公务舱（2 级）已经坐满，那么他最终会乘坐经济舱（1 级），而不会被挤下飞机（0 级），当然，为了达到这一点，很可能要挤下去一个经济舱的乘客。

2. 降级原则。

下面证明，在实际情况中，在优先级原则下，任何优先级的人都不可能得到低两级的服务（比如说头等舱（3 级）的乘客不会乘坐经济舱（1 级），他至少乘坐商务舱）。

证明：如果这种情况出现，必然是因为头等舱的乘客将头等舱和公务舱占满或者是公务舱的乘客将公务舱和经济舱占满。但这就要求 $m_3 \geq n_2 + n_3$ 或者 $m_2 \geq n_1 + n_2$ ，根据实际飞机的数据， n_1 至少是 n_2 的 6 倍， n_2 至少是 n_3 的 2 倍。如果这种情况出现，说明航空公司发出了 3 倍于头等舱座位个数的头等舱票或 7 倍于商务舱座位个数的商务舱票，这样风险极大，会使很多头等舱或商务舱乘客得不到满意的服务而使公司名誉下降，经济受损。因此可以断言，在实际背景下，所有人所得到的服务最多降一级，这就是降级原则。

3. 改变机票的定义，利用等效性定理，化归为基本模型。

我们可以这样设想这样的一种预售票过程，航空公司分三次预售机票，每次售出的是“升级票”：买 1 级票可以从 0 级升到 1 级（经济舱），买 2 级票可以从 1 级升至 2 级（公务舱），依此类推。当然，买到低级的票才有资格购买高一级的票。1 级票 $m_1 + m_2 + m_3$ 张，价格为 $g_1 = g$ ；2 级票 $m_2 + m_3$ 张，价格为 $g_2 - g_1 = 0.3g$ ；3 级票 m_3 张价格为 $g_3 - g_2 = 0.2g$ 。三种票都要求误机违约金 20% 以及承诺客满降级赔偿为 b 。可以验证，在这种票的定义和原来票的定义下，乘客可以花相同的钱享受到相同级别的服务，只不过是票这种载体形式不同罢了。下面进一步证明（等效性定理），公司的收入对于这两种票的定义，也是相同的。

证明：公司的收入分为以下四部分：误机违约金，票款收入，固定损耗，降级赔偿金。对于前三项，由于乘客享受相同级别所花的钱在两种定义下是相同的，它们的期望值也不会变。对于降级赔偿金，由降级原则知若购买到了 i 级票，则至少享受到 $i-1$ 级的服务，也就是说，他买的 $1 \sim i-1$ 级票都不会赔付降级赔偿金，而有一定的概率被赔付 i 级机票的降级赔偿金 b ，而

这个概率和（原始机票定义中）买 i 等级的票被赔付 b 的概率是相同的，都是 $\sum_{k=i}^3 m_k$ 个乘客来不少

于 $\sum_{k=i}^3 n_k$ 人的概率。故降级赔偿金的期望亦相同，故公司总收入也相同。

由等效性定理可以得出该模型可以化为三个基本模型的叠加，即 $P(p, b/g, m_1 + m_2 + m_3) + P(p, b/0.3g, m_2 + m_3) + P(p, b/0.2g, m_3)$ ，可以化为基本模型的

求解。至此，本问题在理论上圆满地解决了。

下面以大飞机 ($n_1=360$, $n_2=40$, $n_3=18$)为例做定量的计算(取 $\lambda=0.6$, $b/g=0.2$, $p=0.04$): 用 Matlab 作数值计算可以得到

$$m_1 + m_2 + m_3 = 440$$

$$m_2 + m_3 = 61$$

$$m_3 = 19$$

此时公司利润取得最大值, 可以解出 $m_1 = 379$, $m_2 = 42$, $m_3 = 19$, 即头等舱多发出 1 张票, 公务舱多发出 2 张票, 经济舱多发出 19 张票, 所得的收益为 $J = 0.6693$, 收益最大期望为飞行固定费用的 66.93%

4. 灵敏度分析。

变量 λ 是长期以来航空公司的经验值, 所以可以认为它是一个常量, 下面对 b/g 和 p 两个变量做灵敏度分析。

(1)最大收益期望值对 b/g 的灵敏度分析。

由于 $bg = b/g$ 是公司制定的, 因此, 在大飞机座位个数确定的前提下, 给定 bg 便可以算出此时公司的最大收益期望值。所以, 以 $bg = 0.01 \sim 2.00$ 作为自变量, 固定 bg 下最大收益期望值为因变量绘出图形如下:

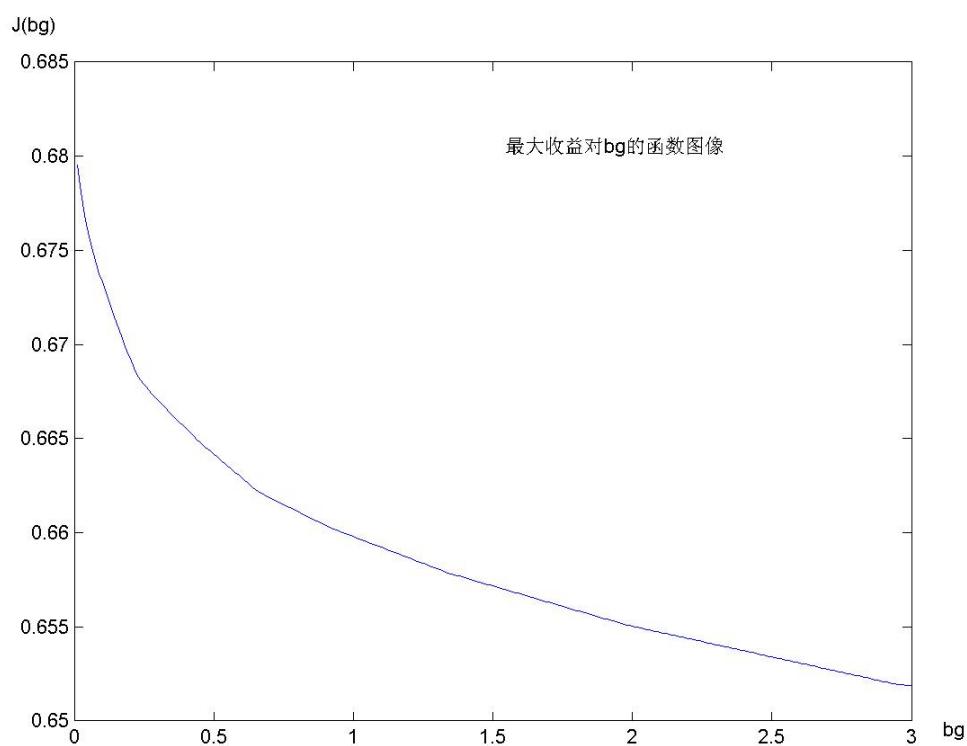


图 17

在图像上可以看出，随着 bg 的上升，刚开始的一段时间里利润下降的比较快，在 $bg = 0.7$ 以后，呈线性单调减少。这是因为如果航空公司赔付的比例越大，预售出的额外的座位也越少，机票收入损失的也越多，收益就小一些。但是当然不能定 $bg = 0$ 而使收益最大，因为公司同时得考虑其社会声誉，如果它是一毛不拔的铁公鸡，不能视顾客为上帝，就会间接的影响其利润。因此权衡利弊，将 bg 大约定在 0.2~0.4 范围内还是比较合适的。

下面给出在最优发票策略下， bg 和 m_1 , m_2 , m_3 的关系，用 Matlab 软件对 bg 从 0.01 到 2.50 进行扫描，得到如下三条曲线。

(i)头等舱 m_3 对 bg 的曲线

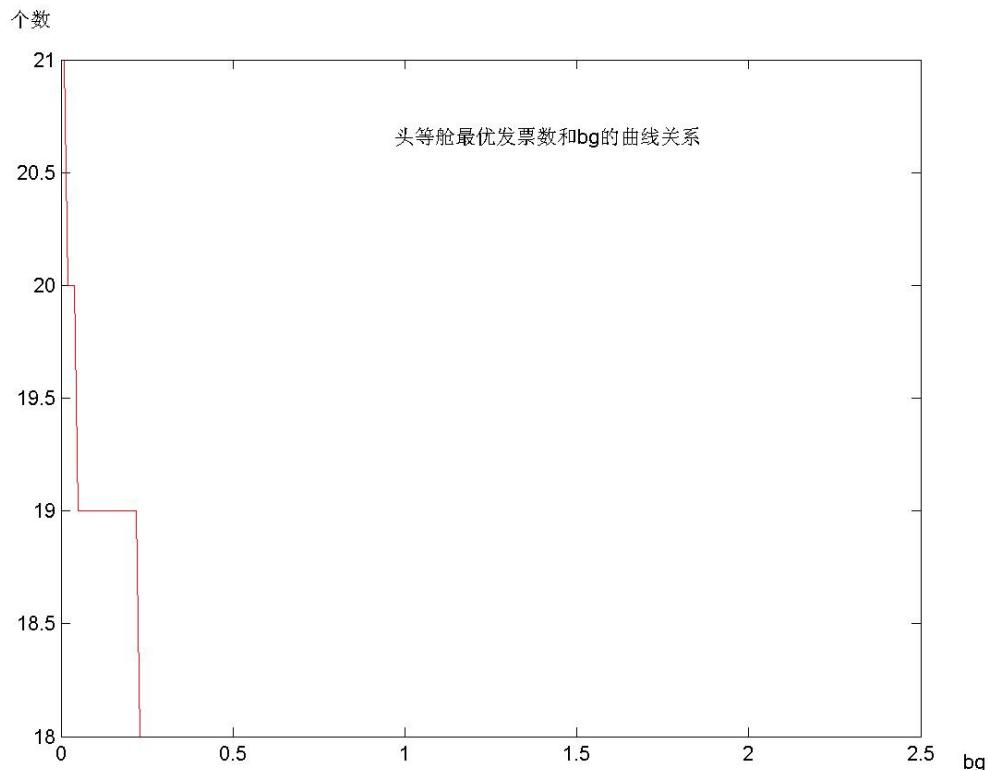


图 18

(ii)公务舱 m_2 对 bg 的曲线。

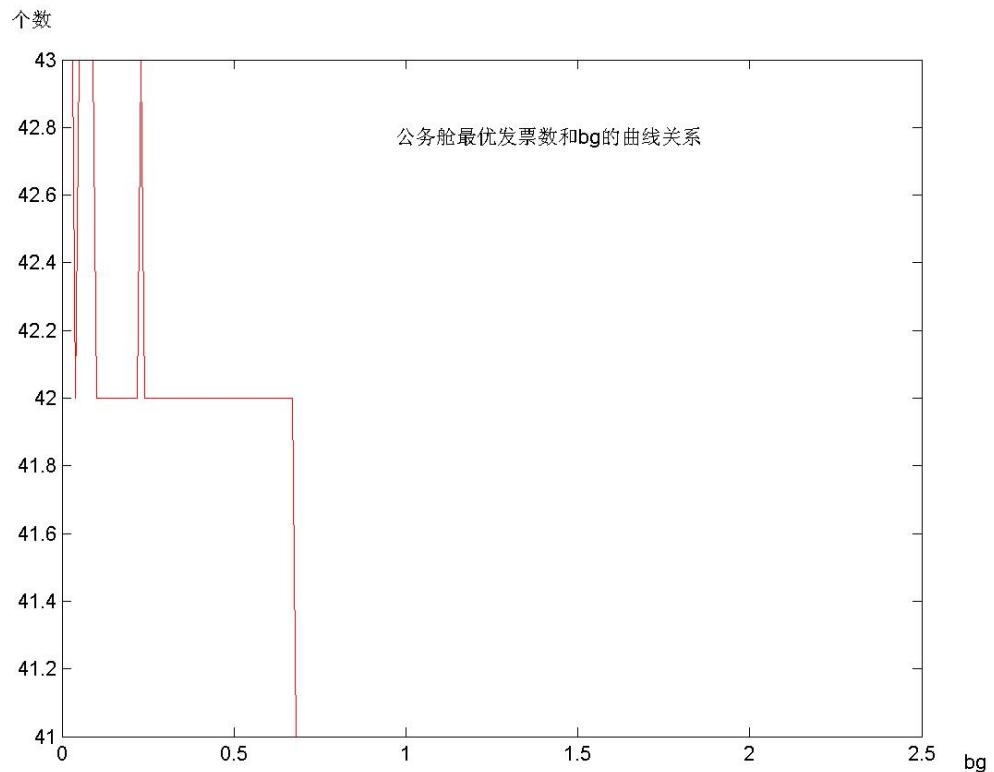


图 19

(iii) 经济舱 m_1 对 bg 的曲线。

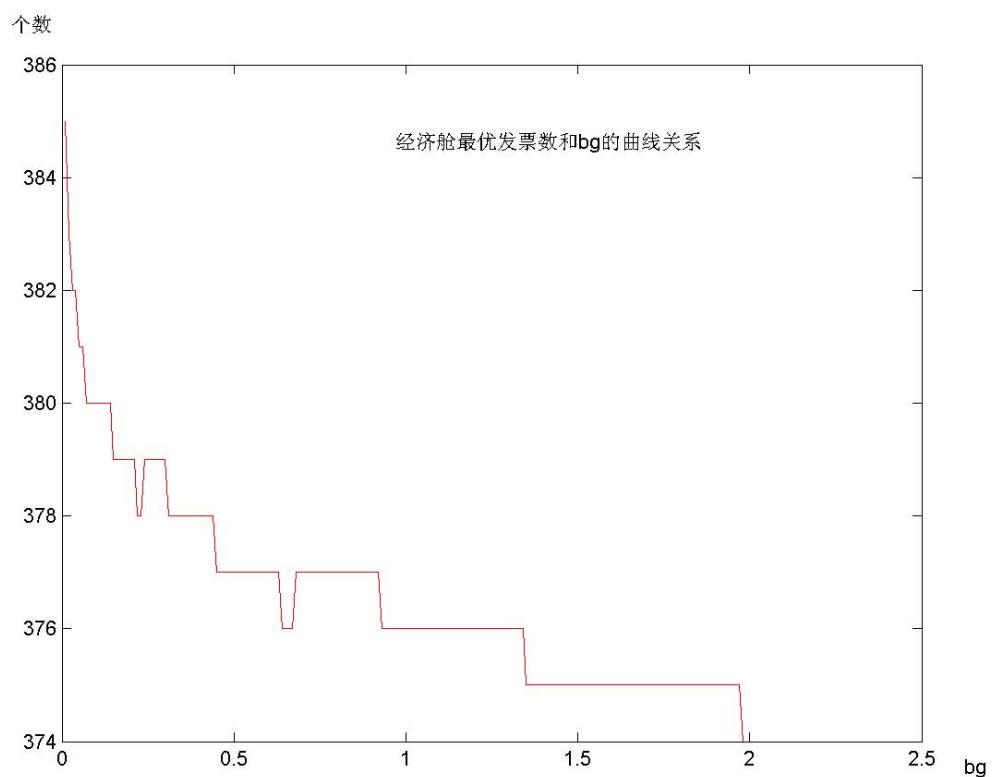


图 20

(2)最大收益期望值对顾客误机概率 p 的灵敏度分析。

由于 p 是航空公司估计的值，所以得知其灵敏度至关重要。我们取 $bg = 0.2$ ，其余数值同上，根据实际情况，知道 p 的值在没有重大事件冲击的时候应该在 0.2 以内，于是在 0~0.3 范围内做出最大收益期望值对于 p 的函数如下：

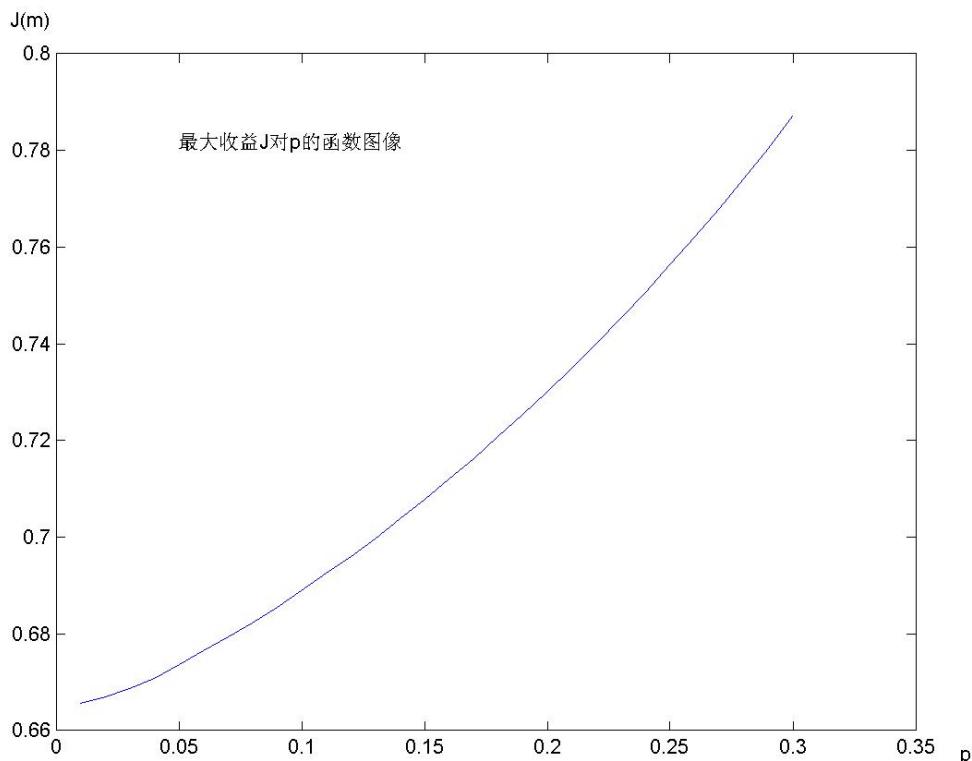


图 21

从图线上可以明显看出这是一个递增下凸的曲线，表明曲线随着 p 增长加速增长，但是比较实际的范围是 $p=0.05$ 附近，在这个区域内，最大收益的变化值相对不是特别的明显。但是值得指出的是，和 bg 的相比， J 对 p 更加灵敏一些，这说明了航空业对外界的依赖性比较大，而自身的一些决策可能对利润影响不是特别的显著。

(六) 模型的优缺点和进一步讨论

- A. 对问题一所建立的三个模型各自的优缺点总结如下：
- 1) 独立数据拟合模型操作简单，但十分粗糙，它无法提取出序列中足够的信息，从而也无法进行有效的拟合和合理的预测。但是作为粗略的估计，在有些情况下还是可以考虑的。
 - 2) 非平稳序列的确定性分析模型操作也比较简捷，它能够有效的提取出序列的总趋势和周期波动趋势，但是对随机信息的浪费十分严重，往往残差数据点还具有一定的相关性。这种模型适合拟合总波动趋势十分强劲，随机信息的影响可以忽略的序列。

3) 非平稳序列的随机性分析有很多模型。本文用的ARIMA模型是一种经过多次差分消除序列中所有的相关性信息，从而转化为对一个稳定序列的分析。利用十分成熟的ARMA模型及其相应的检验准则和参数估计方法可以得到对原序列相当精确的拟合。这种方法虽然操作比较繁琐，而且往往建立的模型需要多次优化，对操作者的经验也有较高要求，但是这种模型可以适用于很多种波动规律的序列（因为非平稳序列经过差分总可以转化为平稳序列），而且拟合精度十分令人满意。在SAS等软件中，ARIMA模型可以很方便的实现，所以这种方法在当今社会的各个领域已经得到了越来越广泛的应用。

B. 对问题二中模型的推广。

本模型在降级原则下，可以推广至 n 个有不同优先级的等级(舱)的预售票收益最大问题。设 n 个等级为 $0 \sim n-1$ 级，分别固定有 n_i 个座位，票价分别为 g_i ，降级赔偿为 b ，违约金为 $c\%$ ，求最佳预售票的策略。可以通过等效性原则将模型转化为 n 个基本模型之和，再通过数学软件对每个基本模型进行数值求解，最终得到模型的数值解，亦可以对不同的变量按照不同的需求作灵敏度分析。

C. 关于公司的声誉的讨论

在满足了以上的种种条件之后，公司应当考虑到不能让太多的旅客来了之后没有登机，当然这都是持经济舱票的旅客。这时我们应当限制总的预售票数量，也就是给出 $m_1 + m_2 + m_3$ 的上限。我们考察的是当 $m_1 + m_2 + m_3$ 取每个值的时候，至少有 j 个人来了但是无法登机的概率。这也就等价于有不多于 $m_1 + m_2 + m_3 - n_1 - n_2 - n_3 - j$ 个人没有来，这个事件的概率为 $P_j(m) = \sum_{k=0}^{m-n-j} p_k$ 。分别考察 $b/g=0.2$ 和 0.4 以及 $j=5$ 和 $j=10$ 情况，用 Matlab 计算可以得到如下表格：

表八 至少有 j 个人无法登机的概率及相应的收益

m	p=0.04			
	b/g=0.2	b/g=0.4	j=5	j=10
420.0000	0.6211	0.6211	0	0
422.0000	0.6288	0.6288	0	0
424.0000	0.6365	0.6365	0.0000	0
426.0000	0.6442	0.6442	0.0000	0
428.0000	0.6517	0.6517	0.0005	0.0000
430.0000	0.6589	0.6588	0.0042	0.0000
432.0000	0.6653	0.6649	0.0206	0.0001
434.0000	0.6703	0.6694	0.0687	0.0014
436.0000	0.6736	0.6720	0.1680	0.0087
438.0000	0.6753	0.6727	0.3215	0.0356
440.0000	0.6756	0.6718	0.5053	0.1027
442.0000	0.6751	0.6699	0.6814	0.2245
444.0000	0.6740	0.6674	0.8196	0.3933
446.0000	0.6727	0.6646	0.9101	0.5775
448.0000	0.6713	0.6617	0.9605	0.7405

450.0000	0.6698	0.6587	0.9846	0.8597
452.0000	0.6684	0.6557	0.9947	0.9331
454.0000	0.6669	0.6527	0.9983	0.9718
456.0000	0.6655	0.6497	0.9995	0.9894
458.0000	0.6640	0.6467	0.9999	0.9965
460.0000	0.6625	0.6437	1.0000	0.9989

p=0.1				
m	b/g=0.2	b/g=0.4	j=5	j=10
420.0000	0.5407	0.5407	0	0
422.0000	0.5480	0.5480	0	0
424.0000	0.5553	0.5553	0.0000	0
426.0000	0.5626	0.5626	0.0000	0
428.0000	0.5698	0.5698	0.0005	0.0000
430.0000	0.5766	0.5765	0.0042	0.0000
432.0000	0.5826	0.5822	0.0206	0.0001
434.0000	0.5872	0.5864	0.0687	0.0014
436.0000	0.5901	0.5886	0.1680	0.0087
438.0000	0.5914	0.5889	0.3215	0.0356
440.0000	0.5914	0.5876	0.5053	0.1027
442.0000	0.5905	0.5853	0.6814	0.2245
444.0000	0.5890	0.5824	0.8196	0.3933
446.0000	0.5873	0.5792	0.9101	0.5775
448.0000	0.5856	0.5759	0.9605	0.7405
450.0000	0.5837	0.5726	0.9846	0.8597
452.0000	0.5819	0.5692	0.9947	0.9331
454.0000	0.5800	0.5658	0.9983	0.9718
456.0000	0.5782	0.5624	0.9995	0.9894
458.0000	0.5763	0.5590	0.9999	0.9965
460.0000	0.5745	0.5557	1.0000	0.9989

从表中可以看出，当 $p=0.04$ ，在最大利润时 ($m=440$)，5 个乘客不能乘飞机的概率为 50.53%，而 10 个乘客不能乘飞机的概率为 10.27%，有点偏大，但是同时可以看出当赔偿金加倍时 ($b/g=0.4$)，利润并没有很大损失，所以航空公司可以考虑用增加赔偿金的方式获得好的声誉。

D. 关于 p 值突变的讨论

通过 $J(m) \sim p$ 图象可以看出，对于一个确定的误机概率 p ，航空公司应该制定合适的预订票数量 $m(p)$ ，使得总利润达到最大。但是如果 p 发生突变，航空公司来不及改变发售预订票的政策，就可能使利润大幅减少。下面以一组具体数据来说明：

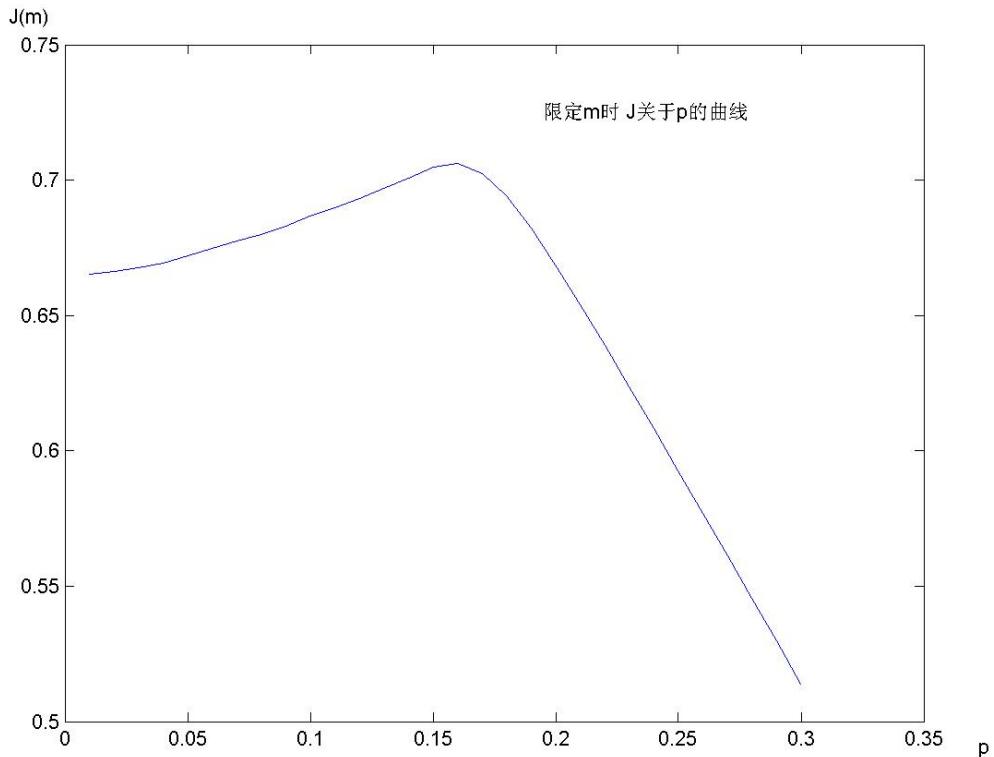


图 22 限定 m 不超过 n 的 20% 情况下 J 关于 p 的变化曲线

从图中可看出当 p 大于某个值时利润开始下降，甚至有可能出现负值。

前几年发生的“911 事件”就可以看作这个讨论的现实背景，当“911 事件”发生后，乘客的误机概率会在极短时间内大幅上升，如果航空公司不作出及时调整，将会受到严重损失。这时航空公司可以减小飞行损耗（比如把大飞机换成小飞机），或者增加预订票数量 m 和赔偿金 b 来减少损失（这是建立在旺季顾客源充足的前提下的）。但是从现实情况来看，当发生了类似于“911 事件”的突发事件后，旺季模型已经不适用，此时航空公司可以采取的更现实的措施是设法减小 p 值，例如减小误机所付定金的百分比，增加顾客意外伤害的保险金额等等来吸引顾客。但这已经超出了本模型讨论的范围。

(七) 参考文献

- [1] 王燕，应用时间序列分析，北京：中国人民大学出版社，2005。
- [2] 姜启源，数学模型（第三版），北京：高等教育出版社，2003。
- [3] 王炜（火斤），应用时间序列分析，桂林：广西师范大学出版社，1999。
- [4] 张 卓，SAS 软件的应用，统计与信息论坛，第20卷第4期：104–107页，2005。
- [5] 杨云霞，时间序列预测模型及其应用，太原师范学院学报(自然科学版)，第4卷第4期：4–7 页，2005。

(八) 附录（略）

有关等度连续函数序列一致有界性的问题¹

杜升华²

§ 1 问题

在高等分析课上，我们学习了如下形式的 Arzelà-Ascoli 定理：

设 $K \subset \mathbb{R}^d$ 为紧集， $\{f_n\} \subset C(K)$ 满足：(1)一致有界，即 $\|f_n\| := \sup_{x \in K} |f_n(x)| \leq C, \forall n \geq 1$ (C

与 n 无关)；(2)等度连续，即 $\forall \varepsilon > 0$ ， $\exists \delta > 0$ 使得当 $x, y \in K$ ， $|x - y| < \delta$ 时，

$\sup_{n \geq 1} |f_n(x) - f_n(y)| < \varepsilon$ 。则存在子列 $\{f_{n_k}\}$ 使得 $\{f_{n_k}\}_{k=1}^{\infty}$ 在 K 上一致收敛，即 $\exists f \in C(K)$ 使得 $\|f_{n_k} - f\| \rightarrow 0 (k \rightarrow \infty)$ 。

由于证明过程中只用到逐点有界的条件，但根据卓里奇的教材，一致有界是一个必要条件，所以有同学提出这样的问题：是否 $\{f_n\}$ 的一致有界性可由逐点有界和题目的其他条件推出？或者考虑一个更强些的命题：是否紧集上逐点有界的连续函数序列一定是一致有界的？

以下是我课后给出的结论和证明——

§ 2 紧集上逐点有界的等度连续函数序列必一致有界

设 X 是一个度量空间， $K \subset X$ 是一个紧集， $f_n \in C(K, \mathbb{R})$ 是 K 上的等度连续函数序列，

$\forall x \in K$ ，序列 $\{f_n(x)\}$ 有界，则函数序列 $\{f_n\}$ 在 K 上一致有界。

证明：假设不然，那么 $\sup_{x \in K} \left(\sup_{n \geq 1} |f_n(x)| \right) = +\infty$ 。对于 1，存在 $x_1 \in K$ 使得 $\sup_{n \geq 1} |f_n(x_1)| > 1$ ；

对于 2，存在 $x_2 \in K$ 使得 $\sup_{n \geq 1} |f_n(x_2)| > 2$ ；……；对于 $k \in \mathbb{N}$ ，存在 $x_k \in K$ 使得

$\sup_{n \geq 1} |f_n(x_k)| > k$ ……由此得到紧集 K 中序列 $\{x_k\}$ ，它有在 K 中收敛的子列 $\{x_{k_j}\}$ ，设 $\lim_{j \rightarrow \infty} x_{k_j} = x_0$ 。

而序列 $\{f_n(x_0)\}$ 是有界的，设 $\sup_{n \geq 1} |f_n(x_0)| = M$ 。由函数序列 $\{f_n\}$ 在 K 上等度连续知 $\exists \delta > 0$ 使

得 $\forall x, y \in K$ ， $d(x, y) < \delta$ 有 $|f_n(x) - f_n(y)| < 1$ ， $\forall n \geq 1$ 。特别地，当 $x \in B(x_0, \delta)$ 时，

$|f_n(x)| < |f_n(x_0)| + 1 \leq M + 1$ ， $\forall n \geq 1$ 。但是当 j 充分大时，有 $d(x_{k_j}, x_0) < \delta$ 且

$\sup_{n \geq 1} |f_n(x_{k_j})| \geq k_j \geq j > M + 1$ ，矛盾。

也可直接证明：由等度连续定义， $\exists \delta > 0$ 使得 $\forall x, y \in K$ ， $d(x, y) < \delta$ 有 $|f_n(x) - f_n(y)| < 1$ ，

¹ 本文根据作者在 2007 年春季学期高等分析课程学习期间推导出的结果整理而成。

² 基科 58-基数 53

$\forall n \geq 1$ 。 $\bigcup_{x \in K} B(x, \delta)$ 构成紧集 K 的开覆盖，从中可取出有限覆盖： $\bigcup_{i=1}^N B(x_i, \delta) \supset K$ 。令

$$M = \max_{1 \leq i \leq N} \left\{ \sup_{n \geq 1} |f_n(x_i)| \right\} + 1。 \quad \forall x \in K, \exists x_i \text{ 使得 } x \in B(x_i, \delta), \text{ 则 } |f_n(x) - f_n(x_i)| < 1, \forall n \geq 1。$$

于是有 $|f_n(x)| < |f_n(x_i)| + 1 \leq \sup_{n \geq 1} |f_n(x_i)| + 1 \leq M$ 。由 x 与 n 的任意性知函数序列 $\{f_n\}$ 在 K 上一致有界。

值得注意的是，条件中“等度”二字不能去掉。因为可以构造出随着 n 的增大在某一点处变化得越来越“厉害”的函数序列。下面举一个反例：

$$\text{设 } K = [0, 2] \text{ 为 } \mathbb{R} \text{ 中紧集, } f_n(x) = \begin{cases} n^2 x, & 0 \leq x \leq \frac{1}{n} \\ 2n - n^2 x, & \frac{1}{n} \leq x \leq \frac{2}{n} \\ 0, & \frac{2}{n} \leq x \leq 2 \end{cases}, \text{ 它的图像是底边在 } x \text{ 轴上、}$$

以 $\left(\frac{1}{n}, n\right)$ 为顶点的等腰三角形的两腰以及 x 轴上的一条线段。 $\{f_n\}_{n=1}^\infty$ 显然是 $[0, 2]$ 上的连续函数序列，且不难验证它们在每一点处都是有界的，但显然并不一致有界。

§ 3 相关问题

任课老师对我的上述证明表示肯定，并指出可能只要“函数列在某一点处的函数数列有界”加上“等度连续”即可推出“一致有界”。后来补充说当紧集 K 不连通时上述命题不总成立，而连通时应当成立。于是我对此作了严格证明——

§ 4 在连通紧集中某点处有界的等度连续函数序列一致有界

设 K 是度量空间 X 内的连通紧集， $\{f_n\}$ 是 K 上的等度连续函数序列， $x_0 \in K$ ，数列

$\{f_n(x_0)\}$ 有界，则 $\{f_n\}$ 在 K 上一致有界。

证明：根据已证明的“紧集上逐点有界的等度连续函数序列必一致有界”的命题，只需证 $\{f_n\}$ 在 K 上逐点有界。

由 $\{f_n\}$ 等度连续知， $\exists \delta > 0$ 使得 $\forall x, y \in K$ ，只要 $d(x, y) < \delta$ 就有 $|f_n(x) - f_n(y)| < 1$ ， $\forall n \geq 1$ 。

设 $E = \{x \in K \mid \exists M > 0, \forall n \geq 1, |f_n(x)| < M\}$ ，来证明 E 是 K 中的非空开闭集。非空性是已知的。

任取 $x_1 \in E$ ，设 $\sup_{n \geq 1} |f_n(x_1)| = M$ 。考虑 K 中开集 $B(x_1, \delta) \cap K$ ， $\forall x \in B(x_1, \delta) \cap K$ ，有

$|f_n(x)| \leq |f_n(x_1)| + |f_n(x) - f_n(x_1)| < M + 1$, $\forall n \geq 1$, 故 $B(x_1, \delta) \cap K \subset E$, E 是 K 中开集³。

任取 $x_2 \in K \setminus E$ 。考虑 K 中开集 $B(x_2, \delta) \cap K$, 假设 $\exists a \in B(x_2, \delta) \cap K$ 使得 $\{f_n(a)\}$ 有界,

设 $\sup_{n \geq 1} |f_n(a)| = M$, 则 $|f_n(x_2)| \leq |f_n(a)| + |f_n(x_2) - f_n(a)| < M + 1$, $\forall n \geq 1$ 。由此推出 $x_2 \in E$,

矛盾。故 $\forall x \in B(x_2, \delta) \cap K$, $\{f_n(x)\}$ 无界, 即 $B(x_2, \delta) \cap K \subset K \setminus E$, 所以 $K \setminus E$ 是 K 中开集, 也即 E 是 K 中闭集。

由 K 的连通性知 $E = K$, 即 $\{f_n\}$ 在 K 上逐点有界, 从而一致有界。

注: K 的紧性要在“逐点有界”推“一致有界”这一步用到。

§ 5 结论

至此, 我们可以把开头提到的 \mathbb{R}^d 中紧集上的 Arzelà-Ascoli 定理在更弱的条件下表述出来⁴:

设 (X, d) 是一个度量空间, $K \subset X$ 是一个紧集, $f_n \in C(K, \mathbb{R})$ 是 K 上的等度连续函数序列, 并满足如下两条件之一: ① $\forall x \in K$, 序列 $\{f_n(x)\}$ 有界; ② K 是连通集, 且 $\exists x_0 \in K$ 使得序列 $\{f_n(x_0)\}$ 有界。则存在子列 $\{f_{n_k}\}$ 使得 $\{f_{n_k}\}_{k=1}^\infty$ 在 K 上一致收敛。

数学名言

一个没有几分诗人气的数学家永远成不了一个完全的数学家。

——[德] K•维尔斯特拉斯

一个名副其实的科学家, 尤其是一个数学家, 他在工作中会感受到与艺术家相同的巨大愉快。

——[法] H•庞加莱

在“真正的数学”中, 存在着严肃性、寓意深远、美丽、一般性、深刻性、意外性、必然性和经济性。

——[英] G•H•哈代

³ 所谓“ E 是 K 中开集”是指相对开集(诱导拓扑意义下的开集), 即存在 X 中开集 U 使得 $E = K \cap U$ 。事实上, 这里可以取 $U = \bigcup_{x_1 \in E} B(x_1, \delta)$ 。对于下文中的 $K \setminus E$ 也是如此。

⁴前面的 \mathbb{R}^d 显然可以替换为任意一个度量空间 (X, d) 。

$\sum_{n=1}^{\infty} \frac{n^k}{n!}$ 是 e 的整数倍的证明

刘立达¹

2007 年 8 月

对 $\forall k \in \mathbb{N}, k \geq 1$, $\sum_{n=1}^{\infty} \frac{n^k}{n!}$ 是 e 的整数倍.

证明:

首先, 当 $k=1$ 时, 该式等于 e , 结论显然成立.

当 $k \geq 2$ 时, 我们来分拆 n^k , 将 n^k 分拆为阶梯状乘积的和.

令:

$p_{1,1}$ 为 $n(n-1)\dots(n-k+1)$ 的展开式中 n 的系数;

$p_{1,2}$ 为 $n(n-1)\dots(n-k+1)$ 的展开式中 n^2 的系数;

.....

$p_{1,k}$ 为 $n(n-1)\dots(n-k+1)$ 的展开式中 n^k 的系数 ($=1$);

$p_{2,1}$ 为 $n(n-1)\dots(n-k+2)$ 的展开式中 n 的系数;

.....

$p_{2,k-1}$ 为 $n(n-1)\dots(n-k+2)$ 的展开式中 n^{k-1} 的系数 ($=1$);

.....

$p_{k,1}$ 为 n 的展开式中 n 的系数, 即 1.

即下表:

n^k 的系数	n^{k-1} 的系数	...	n 的系数
$n(n-1)\dots(n-k+1)$	$p_{1,k}$	$p_{1,k-1}$	\dots
$n(n-1)\dots(n-k+2)$		$p_{2,k-1}$	\dots
\vdots		\ddots	\vdots
n			$p_{k,1}$

令 $n^k = [q_1 n(n-1)\dots(n-k+1)] + [q_2 n(n-1)\dots(n-k+2)] + \dots + q_k n$, 其中 $q_1=1$ 。由于等式两边 n, n^2, \dots, n^{k-1} 的系数都为 0, 所以有:

$$\begin{cases} q_1 p_{1,1} + q_2 p_{2,1} + \dots + q_k p_{k,1} = 0 \\ q_1 p_{1,2} + q_2 p_{2,2} + \dots + q_{k-1} p_{k-1,2} = 0 \\ \vdots \\ q_1 p_{1,k-2} + q_2 p_{2,k-2} + q_3 p_{3,k-2} = 0 \\ q_1 p_{1,k-1} + q_2 p_{2,k-1} = 0 \end{cases}$$

由此解得:

$$q_1 = 1;$$

$$q_2 = -p_{1,k-1};$$

$$q_3 = -(p_{1,k-2} + q_2 p_{2,k-2});$$

.....

$$q_k = -(p_{1,1} + q_2 p_{2,1} + \dots + q_{k-1} p_{k-1,1}).$$

那么我们得到 $n^k = [q_1 n(n-1)\dots(n-k+1)] + [q_2 n(n-1)\dots(n-k+2)] + \dots + q_k n$, 其中 q_1, \dots, q_k 的表达式如上。

至此, 我们将 n^k 拆成了 k 个阶梯状乘积的和. 从而得到下式:

¹ 基科 58-基数 53

$$\begin{aligned}
& \sum_{n=1}^{\infty} \frac{n^k}{n!} \\
&= \sum_{n=1}^{\infty} \frac{\sum_{j=1}^k q_j \prod_{i=1}^j (n-i+1)}{n!} \\
&= \sum_{j=1}^k q_j \sum_{n=1}^{\infty} \frac{\prod_{i=1}^j (n-i+1)}{n!} \\
&= \sum_{j=1}^k q_j \sum_{n=j}^{\infty} \frac{1}{(n-j)!} \\
&= \sum_{j=1}^k q_j \sum_{n=0}^{\infty} \frac{1}{n!} \\
&= e^{\sum_{j=1}^k q_j}
\end{aligned}$$

鉴于诸 q_j 都是整数，其和必为整数，因而结论得证。

数学笑话

微分

常函数和指数函数 e^x 走在街上，远远看到微分算子，常函数吓得慌忙躲藏，说：“被它微分一下，我就什么都没有啦！”指数函数不慌不忙道：“它可不能把我怎么样，我是 e^x ！”指数函数与微分算子相遇。指数函数自我介绍道：“你好，我是 e^x 。”微分算子道：“你好，我是 d/dy ！”

黑色的羊

物理学家、天文学家和数学家走在苏格兰高原上，碰巧看到一只黑色的羊。

“啊！”天文学家说道，“原来苏格兰的羊是黑色的。”

“得了吧，仅凭一次观察你可不能这么说。”物理学家道，“你只能说那只黑色的羊是在苏格兰发现的。”

“也不对，”数学家道，“由这次观察你只能说：在这一时刻，这只羊，从我们观察的角度看过去，有一侧表面上是黑色的。”

处处不可导

有一位国外的学者（搞数学研究的）到我们学校访问，住在学校外宾招待所，他要走的时候，我问他对我们学校的印象如何，他说：“你们学校的招待所太差了，以后再也不敢住了！”我急忙问其原因。教授说道：“那吃饭的碗，碗口处处不可导，这哪是给人用的！”

三次数学危机中的问题猎手 *

杜升华

2007 年 11 月

关键词： 数学危机 问题猎手

摘要

本文通过回顾历史上三次数学危机的产生和解决以及在此过程中杰出数学家发现问题、解决问题的途径，来阐释成为一个优秀的问题猎手对于数学工作者的意义。

伟大的法国数学家庞加莱 (H. Poincaré, 1854-1912) 曾说过：“如果我们希望预知数学的将来，适当的途径是研究这门学科的历史和现状。”对于数学系的学生和未来的数学工作者来说，了解数学史，尤其是历史上伟大的数学家如何发现和解决重大数学问题，是十分必要的。

数学的发展并非坦途。历史上，数学家们曾三次遇到令人困惑不解的难题，甚至引起整个数学界的震惊、混乱与反思，史称“数学危机”。在这三次数学危机中，无论难题的提出者还是解决者，都是了不起的问题猎手，都极大地推动了数学的发展。

§1 第一次数学危机 — 无理数的发现

古希腊的毕达哥拉斯 (Pythagoras, 约公元前 569–前 500) 是数学史上一位有重要影响的数学家，他把“证明”这一基本数学思想引入了数学。他在意大利南部成立了一个有宗教色彩的秘密帮会，致力于数学与哲学的研究。这一学派被称为毕达哥拉斯学派。现在一般认为该学派发现了著名的毕达哥拉斯定理（即勾股定理）：设 a 、 b 、 c 是一个直角三角形的三边长，其中 c 为斜边长，则 $a^2 + b^2 = c^2$ 。据传他们在发现了这一定理时无比欣喜，以至宰杀了一百头牛祭神庆贺。毕达哥拉斯学派笃信的基本信条是“万物皆数”，这里他们所说的“数”指的仅仅是能够写成两个正整数之比的数，即有理数。这在几何上相当于说，对于任何两条给定的线段，总能找到某第三线段，以它为单位线段能将给定的两条线段划分为整数段。希腊人称这样两条给定线段为“可公度量”，意即有公共的度量单位 [1]。

可是该学派成员希帕苏斯 (Hipasus) 恰恰由这一学派最著名的定理得出了违背他们的信条的结论。考虑直角边长为 1 的等腰直角三角形，设其斜边长为 l ，根据毕达哥拉斯定

*本文是作者在选修课 科研思维方法 中写的期中论文。

理，有 $l^2 = 1^2 + 1^2 = 2$ 。容易证明这个 l 不可能写成两个正整数之比¹。这种数的发现严重地冲击了毕达哥拉斯的信徒们的观念，因为它意味着正方形的对角线长与边长是不可公度的。是放弃“万物皆（有理）数”的基本信条，还是否认由他们最得意的定理导出的几何学事实，毕达哥拉斯学派面临两难的选择和信仰的危机，困惑、惊恐、不知所措，最终决定对外保守这个可怕的秘密。后来希帕苏斯还是把它透露出去了，结果按毕达哥拉斯学派的纪律，把希帕苏斯投入大海，葬身鱼腹 [2]！

大约一个世纪后，欧多克斯 (Eudoxus，约公元前 408—前 347) 提出新的比例理论²，在几何上回避了两个量是否可公度的问题，从而在当时的认识水平上解决了第一次数学危机。至于这一问题的根本解决，则要归功于德国数学家戴德金 (J. W. R. Dedekind, 1831—1916) 等人严格建立起来的无理数理论，那是 19 世纪后半叶的事情了。限于篇幅，这里不再介绍相关的理论。

§2 第二次数学危机 — 微积分的严密化

众所周知，牛顿 (I. Newton, 1642—1727) 和莱布尼兹 (G. W. Leibniz, 1646—1716) 创立了微积分，开启了高等数学的大门，这是数学史上一场意义深远的革命。自微积分创立以来，许多用初等方法难以计算的几何和物理问题被解决，相关的种种工程学科得到迅猛发展，社会生产力得以飞速提高。由于人类认识和改造自然的能力的增强，科学思想得以传播，旧的神学观念受到冲击，这自然引起保守势力的不满。从某种意义上说，微积分奠定了上述朝着资本主义方向迈进的新现象的基础，因而它受到某些神学家的责难就不难理解了。

英国哲学家和牧师伯克莱在 1734 年出版了一本小册子《分析学家，或致一位不信神的数学家》，点名道姓地攻击牛顿、莱布尼兹及其拥护者的微积分成果是诡辩 [3]。事实上，牛顿和莱布尼兹创立的微积分，确实存在逻辑基础和推导过程不严密的问题，尤其“流数”、“无穷小量”等概念没有明确的定义，曾引起不少科学家善意的批评和建议。但伯克莱牧师的动机另当别论。

伯克莱在这本小册子里提出了后来以他的名字命名的著名悖论：“在推导任意次幂的流数时，如果让增量消失，亦即让增量变成零，那么原来的关于增量存在的假设也就不能成立，而由这一假设引出的结果即借助于增量而得到的表达式却必须保留，这种推理是站不住脚的。因为我们如果假设增量消失了，理所应当也就必须假设它们的比、它们的表达式以及由于假设其存在而导出的一切东西都必须随之消失。”接下来伯克莱说了些很难听的话：“这些消逝的量是什么呢？难道我们不能称它们为消逝的鬼魂吗？”“分明是诡辩，是招摇撞骗，把人们引入歧途。”……

数学家们，牛顿和莱布尼兹的追随者们，自然对此展开反击。可是一方面数学家们忙于开拓微积分理论的广泛应用而无暇顾及基础理论，另一方面当时也无人能对极限、导数（流数）等基本概念作出严格的定义，在这场数学与神学、科学与宗教的辩论中伯

¹ 假设存在正整数 p, q 使得 $l = \frac{p}{q}$ ，我们可以假定 p 与 q 是互素的，否则约分化简即可。于是 $lq = p$ ，两边平方得 $2q^2 = p^2$ 。这表明 p 是偶数，可设 $p = 2k$ ，则 $2q^2 = 4k^2$ ，即 $q^2 = 2k^2$ ，从而 q 也为偶数，这同 p 与 q 互素矛盾。

² 收录在欧几里得 (Euclid，约公元前 330—前 275) 著名的几何《原本》第五卷。

克莱似乎占了上风。这就是历史上骇人听闻的第二次数学危机。到了 19 世纪，如何巩固微积分的基础，巩固近代以来飞速发展的自然科学的基础，成为一个摆在数学家面前的重大问题。

数学史家贝尔以如下一个比喻来说明把严格性引进数学分析的重要意义 [4]：“假定整个民族多少世纪以来一直是向谬误的神明顶礼膜拜，而突然间向他们揭示了他们的错误。在引进严格性之前，数学分析就是整整一座谬误之神的万神殿。在这方面，柯西与高斯和阿贝耳一起，是伟大的先驱者。”柯西 (O. L. Cauchy, 1789–1857) 以严格化为目标，对许多基本概念作了明确的定义，并在此基础上重建和拓展了微积分的重要事实与定理。例如：[1]

变量 “依次取许多互不相同的值的量叫做变量。”

函数 “当变量之间这样联系起来的时候，即给定了这些变量中的一个值，就可以决定所有其他变量的值的时候，人们通常想象这些量是用其中的一个来表达的，这时这个量就取名为自变量，而由这些自变量表示的其他量就叫做这个自变量的函数。”这个定义很接近于现代对函数的定义，突破了以往函数必须有解析表达式的限制。

极限 “当同一变量逐次所取的值无限趋向于一个固定的值，最终使它的值与该定值的差要多小就多小，那么最后这个定值就称为所有其他值的极限。”现代的中学教科书一般就是这样描述极限概念的。

导数与微分 柯西把导数明确定义为差商

$$\frac{\Delta y}{\Delta x} = \frac{f(x+h) - f(x)}{h}, \Delta x = h$$

当 h 无限地趋向于零的极限 $f'(x)$ ，函数的微分则定义为 $dy = f'(x)dx$ 。而以往的数学家常常是取通过被伯克莱所指责的不严格运算得到的微分作为基本概念，再把导数作为“微分系数”而引入。

柯西的工作向分析的全面严格化迈出了关键的一步，引起科学界很大的轰动。但他的结果还只能算是比较严格，“无限趋向”、“要多小就多小”等语言仍只是直觉的描述。真正为微积分奠定严格基础的，是 19 世纪后半叶的“分析算术化”运动。德国数学家维尔斯特拉斯 (K. Weierstrass, 1815–1897) 由于在这一运动中的伟大贡献而被视为“现代分析学之父”和德意志的民族英雄。可以毫不夸张地说，他是数学史上的英雄。

维尔斯特拉斯引入的关于极限的“ $\varepsilon - \delta$ 语言”，至今仍是大学教科书中函数极限概念的标准定义：设 x_0 是函数 $f(x)$ 定义域 E 中的一个值， A 为一个实数。如果对于任意的正数 ε ，存在正数 δ ，使得当 $0 < |x - x_0| < \delta, x \in E$ 时， $|f(x) - A| < \varepsilon$ ，则称当 x (在 E 中) 趋于 x_0 时， $f(x)$ 的极限是 A 。与此同时，维尔斯特拉斯、戴德金、康托尔 (这个人我们很快就会提到) 等人对实数集的结构给出了确切的刻画。至此，微积分的基础得到加固，第二次数学危机得到圆满的解决³。反科学的神学家们再也无法撼动建立在严

³当然，此后数学分析理论仍有很大发展。建立在集合论基础之上的实变函数论、泛函分析等学科的创立和发展被称为数学分析的第三次严格化。不过通常认为第二次严格化 (即维尔斯特拉斯等人的那一次) 已经把第二次数学危机解决了。

格化了的数学分析理论之上的自然科学成就的地位了。

§3 第三次数学危机 — 集合论与悖论

德国大数学家康托尔 (G. Cantor, 1845-1918) 在研究实数理论的过程中，抽象出“集合”的概念，成为集合论的创始人。集合论的出现，特别是无限集合的出现，是数学思想史上的一次伟大革命。因为长期以来数学家一直只承认潜在的无穷而否认实在的无穷，即只把“无穷”看作某种没有尽头的变化趋势，而不承认“无穷”的事物作为一个固定实体而存在。集合论尽管从一开始就受到某些保守的数学家的反对，但它使人们对实数集结构的认识，乃至对整个世界中一切数学对象的认识，上升到了一个新的、更抽象的高度，并逐渐成为现代数学各个分支的理论基础。19、20世纪之交最伟大的数学家之一希尔伯特 (D. Hilbert, 1862-1943) 充满自信地说：“没有人能够把我们驱逐出康托尔为我们创造的乐园。” [4]

但是康托尔创立集合论时还来不及构建严密的公理体系，不加限制地讨论“实无穷”和“所有集合的集合”迟早会出问题。1902年，英国数理逻辑学家罗素 (Russell) 提出了下述著名的悖论：作集合

$$E = \{A : A \notin A\}$$

即 E 由这样的元素 A 组成， A 是一个集合，同时它不是自身的元素。问题是： E 是否属于 E ？容易看出，无论怎样回答都会引出矛盾⁴。这一悖论引起数学界极大的震惊，它使人们看到集合论的体系是不完善的，需要利用公理对所讨论的对象和研究的手段加以限制。有人甚至认为数学的基础由此发生动摇，于是称这一事件为第三次数学危机。

不久，1908年，德国数学家策墨罗 (E. F. F. Zermelo) 和弗伦克尔 (Fraenkel) 合作提出一套集合论的公理体系 [2]，被称为 ZF 公理体系，对集合的构造作了限制。特别地，这里禁止谈“一切集合的集合”，这就避免了罗素悖论中那样的集合的出现。因篇幅所限，不再介绍该公理体系的详细内容。在经过策墨罗等人改造的公理集合论中，至今没有人发现任何悖论或矛盾出来。于是，在这个意义上，第三次数学危机基本解决。

§4 总结与思考

三次数学危机是数学史上的大事，它们促使人们创造新的观念，开拓新的理论，从而极大地促进了数学的发展。在这三场危机中，问题的提出者和解决者都是出色的问题猎手，他们发现问题或解决问题的敏锐眼光值得我们学习。

第一次数学危机的挑起者希帕苏斯毫无疑问是值得纪念的英雄。他勇于怀疑权威和教条，依靠理性思维作出新的发现，为此甚至牺牲了生命。第二次数学危机的挑起者伯克莱牧师不是数学家，而是数学家的敌人，他提出悖论的目的是驳倒牛顿、莱布尼兹创立的微积分和以此为基础的近代自然科学。但在客观上他的悖论促使数学家们寻求微积分的严格基础，从而推动了分析学的发展。至少他追求严格性的精神是值得我们学习的。

⁴假如 E 属于 E ，那么根据 E 的定义，它不是自身的元素，即 $E \notin E$ ，矛盾；假如 E 不属于 E ，同样根据 E 的定义，推出它是自身的元素，即 $E \in E$ ，又是矛盾。

罗素悖论的提出者同样有追求严格性的挑剔眼光，他能从集合论的“乐园”中看出混乱与矛盾，引发整个数学理论的大厦来自基础的震荡——而罗素的目的和最终导致的结果恰恰是使这一基础变得更加坚实。总之，一个怀疑精神，一个严谨品质，是使这三位问题猎手有机会捕捉到巨大猎物的强有力武器。

问题的解决者同样伟大。柯西、维尔斯拉斯、戴德金、康托尔、策墨罗……无一例外都熟知他所工作的这一领域面临什么样的困难，前人做过哪些工作，更重要的是，前人的工作有什么不尽人意的地方——也就是不严格的地方。法国数学家韦伊 (A. Weil, 1906–1998) 有一句名言：“严格性对于数学家，就如道德之于人。”要成为一个数学家，就必须有穷根究底、不放过任何可能的错误的精神。数学就是在这种与问题和错误做斗争的不断严格化的过程中发展、成熟的。

当然，要捕获重要问题，更离不开自由的数学思想。让我们引用康托尔的一段话作为结束 [3]：“数学在其发展过程中应该是完全自由的，对数学研究设定任何多余的限制都只会带来更大的危险。数学的本质在于它的自由！……我宣布，我们的数学科学必须摆脱形而上学的桎梏，我们需要自由发展。”

参考文献

- [1] 李文林, 数学史概论, 高等教育出版社, 2002 年 8 月第 2 版
- [2] 王树禾, 数学思想史, 国防工业出版社, 2003 年 4 月第 1 版
- [3] 王树禾, 数学演义, 科学出版社, 2004 年 10 月第 1 版
- [4] E. T. Bell, *MAN OF MATHEMATICS*, 中译本 数学精英 – 数学家的故事

数学家趣闻

▲ Rota 曾讲了一个 Lefschetz 的故事，关于他的课是如何难懂，因为他经常语无伦次。这是几何课的开场白：“一个 Riemann 曲面是一定形式的 Hausdorff 空间。你们知道 Hausdorff 空间是什么吧？它也是紧的，好了。我猜想它也是一个流形。你们当然知道流形是什么。现在让我给你们讲一个不那么平凡的定理 —Riemann-Roch 定理。”要知道第一节 Riemann 曲面的课如果这样进行的话，恐怕 Riemann 复生也未必可以听懂。： -)

▲ Hilbert (希尔伯特) 曾有一个学生，给了他一篇论文来证明 Riemann (黎曼) 猜想，尽管其中有个无法挽回的错误，Hilbert 还是被深深的吸引了。第二年，这个学生不知道怎么回事就死了，Hilbert 要求在葬礼上做一个演说。那天，风雨瑟瑟，这个学生的家属们哀不胜收。Hilbert 开始致词，首先指出，这样的天才这么早离开我们实在是痛惜呀，众人同感，哭得越来越凶。接下来，Hilbert 说，尽管这个人的证明有错，但是如果按照这条路走，应该有可能证明 Riemann 猜想，再接下来，Hilbert 继续热烈的冒雨讲道：“事实上，让我们考虑一个单变量的复函数. ”众人皆倒。

感悟科研：站在巨人的肩膀上，创造新的价值！¹

杜升华
2008年1月

作为一个数学系学生和未来的数学工作者，我想结合自己的专业，从数学（也涉及物理学）的角度谈谈关于科学的研究的感悟。首先回顾一些历史事实，来说明数学家的工作是“站在巨人的肩膀上创造新的价值”；然后由此引发关于数学的价值以及她为什么值得数学家投入一生的时间进行研究的原因的讨论。

让我们从这句话的提出者开始——据说牛顿(I. Newton, 1642-1727)曾说过(1)：“如果我比其他人都看得更远些，那是因为我站在巨人的肩上。”这句话并不仅仅是自谦之辞，因为牛顿的伟大贡献确实在前人的杰出成就基础上做出的。在这些巨人之中，最高大的有：笛卡儿、开普勒和伽利略。

从笛卡儿那里，牛顿继承了解析几何。笛卡儿(R. Descartes, 1596-1650)是法国数学家，在数学中以创立解析几何学而著称，代表作是《几何学》(2)。他将平面上的点与一种由坐标确定的数对(x,y)联系起来，将几何问题化为代数问题，引入了变量与函数的观点。从此，人类进入变量数学的时代。难以设想，如果没有坐标系，没有质点的运动方程，牛顿的运动定律和微积分基本定理将用什么语言来表述。

从开普勒那里，牛顿继承了天体运动的三个基本原理。开普勒(J. Kepler)于1609和1619年先后提出了行星运动的三条规律，即开普勒三定律(3)：①行星围绕太阳运行的轨道是椭圆，太阳位于所有这些椭圆的一个焦点上；②连结太阳与行星的线段在相等的时间内扫过相等的面积；③行星公转周期的平方与它绕太阳公转轨道的半长轴的立方成正比。在此基础上，牛顿归纳、提炼出了著名的万有引力定律：自然界中任何两个物体都是互相吸引的，引力的大小与这两个物体的质量的乘积成正比，与它们的距离的平方成反比。

从伽利略那里，牛顿得到了“成为他自己动力学奠基石的运动三定律中的头两个”(1)。牛顿的运动三定律是：①任何物体在不受外力作用的情况下都保持静止或做匀速直线运动；②物体的加速度与它所受的外力大小成正比，与它的质量成反比，且加速度与外力方向一致（即 $F=ma$ ）；③作用力与反作用力大小相等，方向相反。而伽利略所做的斜面实验与自由落体实验及相关研究分别为前两个牛顿运动定律奠定了直接基础。

由此可见，牛顿之所以做出伟大的贡献，确实是因为他“站在巨人的肩膀上”，继承和发展了前人的工作。但“砖头还不是大厦；牛顿是动力学和天体力学的建筑师”(1)。如果牛顿没有创造出新的价值的话，也就不会在科学史上占有那样崇高的地位。事实上，牛顿不仅创造了新的价值，而且创造了远远超过他的前辈的巨大的价值。

先谈谈物理方面：开普勒的三大定律是从他的老师第谷(Tycho Brahe)的观测数据中经过二十二年“非人的计算”由经验发现的，而牛顿万有引力定律揭示了自然界更为本质的客观规律；伽利略只是提出了前两个牛顿运动定律的思想，而牛顿把这三个定律明确地表述出来——尤其是，把第二定律以精确的数学语言表述出来。

对数学来说，最重要的或许是解释第二定律中的一个词——“加速度”，也即速度的改变率。归根到底，这是关于“改变率”的问题，因为速度就是位移的改变率。为了解决这个问题，牛顿发明了“流数法”，找到了“揭开改变率及其度量的全部秘密的万能钥匙”(1)——

¹ 本文是作者在选修课科研思维方法中写的期末论文。

微分学。与改变率有关的相反的问题又使牛顿掌握了积分学：如何计算一个速度每时每刻都在变化的质点在给定的时间内通过的距离呢？牛顿把这两方面的问题联系起来考虑，用他的流数理论给出了解答，从而作出了一个伟大的发现：微积分基本定理。当然，德国数学家莱布尼兹(G. W. Leibniz, 1646-1716)也几乎同时独立地作出了这一伟大发现，他们之间关于优先权的争吵是数学史上一件不幸的事。他们的发现标志着微积分学的最终创立，开启了高等数学的大门，使得许多用初等方法难以计算的几何和物理问题获得解决，相关的种种工程学科得到迅猛发展，社会生产力得以飞速提高。

18世纪，欧拉(L. Euler, 1707-1783)、拉格朗日(J. L. Lagrange, 1736-1813)、拉普拉斯(P. S. Laplace, 1749-1827)等数学大师站在牛顿和莱布尼兹的肩膀上，创立了微分方程、变分法、复变函数论等新的数学分支，解决了大量的物理学和力学问题，促成了分析学的大繁荣。

但这种繁荣背后也隐藏着危机，因为18世纪数学家们忙于发展微积分的广泛应用而无暇顾及基础理论，对极限、导数、连续性等基本概念很少过问。于是英国牧师伯克莱的悖论导致了数学史上骇人听闻的第二次数学危机。

在解决这场危机的过程中，高斯(C. F. Gauss, 1777-1855)、柯西(A. L. Cauchy, 1789-1857)和阿贝尔(N. H. Abel, 1802-1829)是把严格性引入数学分析的先驱；维尔斯特拉斯(K. T. T. W. Weierstrass, 1815-1897)等人发起“分析算术化”运动，用“ $\epsilon - \delta$ 语言”对分析学的基本概念给出了今天已被广泛采用的精确定义，将柯西等人的结果进一步严格化，从而解决了第二次数学危机；勒贝格(H. L. Lebesgue, 1875-1941)、希尔伯特(D. Hilbert, 1862-1943)等人在严格化了的分析学基础上进一步发展相关理论，创立了实变函数论、泛函分析等学科，开启了现代分析学的大门。

以上就是分析学的三次严格化运动。关于这个问题我在期中作业里有过比较详细的论述²。值得指出的是，早在18世纪，欧拉、拉格朗日和达朗贝尔(J. L. R. D'Alembert)已经对建立微积分的严格基础做出了重要尝试，他们的工作为19世纪微积分的严格表述提供了方向(2)。所以，19世纪的分析学家也都是站在巨人的肩膀上的。

而他们创造出新的价值更是巨大的。我想，无论怎样评价分析严格化运动的伟大意义都不过分，因为它不但把微积分及其辉煌成就建立在一个严格的基础上，使得任何反科学的神学家都无法撼动，而且体现了空前的创造精神与高度的严格精神——数学精神的实质。可以说，分析的严格化不仅把数学分析本身提升为与粗糙的微积分截然不同的精密的学科，而且深刻影响和改变了整个数学的面貌；并且，它所体现的严谨的美、思维的美、逻辑的美成为吸引现代数学家走上数学研究之路的强大力量。

英国数学家哈代(G. H. Hardy, 1877-1947)在《一个数学家的辩白》中写道(4)：“我的眼界最初是由乐甫教授打开的，他只教了我几个学期的课，却使我对分析的严谨概念有了最初的了解。我从乐甫教授处获益最大的是他建议我读乔丹的著名的《分析教程》(Cours d' analyse)。我永远也不会忘记这部著作所给予我的震撼，不会忘记那本书对我这一代数学家的激励。读了这部著作我才第一次懂得了数学的真谛；也是从那时起我走上了真正的数学家的道路，对数学树立了正确的目标，对数学有了真正的热情。”

现在，让我们谈一谈数学的价值以及数学值得人们进行认真研究的理由。

首先，数学的有用性似乎是不用辩白的。数学在各行各业有着日益广泛的应用，其重要性正变得越来越明显。即使“纯”的数学，最抽象、最“无用”的学科，也总有一天会被其他学科用作实际工具，从而对现实世界产生影响。历史已经证明了这一点。比如哈代深信不会有任何实际用处的数论，在科技迅猛发展的今天已被广泛地应用于通讯编码、自动控制、计算机科

² 见本刊62页《三次数学危机中的问题猎手》一文。

学等领域。因此，我想我们可以得出这样的结论：即使对一个“纯”数学家来说，他的工作也是在间接地为人类社会的发展做贡献，只不过这种贡献的实际效果一般不能马上显现出来。

以上谈的是数学家可能间接地做出的物质上的贡献，这一点或许不太明显，也可能存在争议，但数学家在精神上的贡献则是确切无疑的。“正如历史所充分证明的那样，不论数学内在的本质价值何在，其成就是一切成就中最持久的。”哈代给出了一个堪称经典的例子(4)：“希腊数学是‘不朽的’，甚至比希腊的文学还要持久。当爱斯奇里斯(Aeschylus)被遗忘时，阿基米德(Archimedes)仍将为人们铭记，因为语言文字会消亡，而数学的思想却永不会死亡。‘不朽’这个词可能不太高明，不过也许数学家与它的含义最投缘了。”这段话已成为名言，它令每一个有志于从事数学研究事业的人心潮澎湃。

我不在乎非数学家对数学如何评价，但我知道，我将要投身的事业是一门美丽的学科：她有至少两千五百年的历史，却依然焕发着活力；她贵为科学的女皇，却时常出现在我们的身边；她可以谦恭地为市场服务，却也能高达星宿——这就是数学！

没有什么比马克思在 17 岁时所作的《青年在选择职业时的思考》中的一段话更能表达我的想法了，让我们以此作为结束：“如果我们选择了最能为人类福利而劳动的职业，我们就不会为它的重负所压倒，因为这是为全人类所作的牺牲；那时我们感到的将不是一点点自私而可怜的欢乐，我们的幸福将属于千万人，我们的事业并不显赫一时，但将永远存在；而面对我们的骨灰，高尚的人们将洒下热泪。”

参考文献

- (1)E. T. Bell: *MAN OF MATHEMATICS*, 中译本《数学精英——数学家的故事》
- (2)杜瑞芝（主编）：《数学史辞典》，
http://166.111.121.20:9080/mathdl/search_service/browse.htm
- (3)郭奕玲 沈慧君：《物理学史》，清华大学出版社，1993 年 7 月第 1 版
- (4)G. H. Hardy: 《一个数学家的辩白》，《科学家的辩白》，江苏人民出版社，1999 年 9 月第 1 版

数学家趣闻

▲1983 年，Erdos 第一次到路易斯维尔拜访 Michael Jacobson。那时 Jacobson 还没结婚，于是就跟 Erdos 住在一起。他对 Erdos 的故事早有耳闻，已经作好准备与之一起拼命工作，但情况显然远超过他的预料。“头一天我们就工作到半夜 1 点，我已经精疲力竭，上楼睡觉去了。他呆在楼下的客房里。凌晨 4 点半，我听到厨房有响声，他把锅碗瓢盆弄得一片响，那是在告诉我该起床了。大约 6 点钟时，我跌跌撞撞地下了楼。你猜他说的第一句话是什么？不是‘早上好！’也不是‘睡得好吗？’而是‘设 n 是整数， k 是……’我只穿了一件浴衣，半裸着身体，眼睛还是半睁半闭，矇矇眬眬。我认输了。我告诉他，不洗一个淋浴的话，我是没法接着研究数学的。”

▲一次在 Hilbert 的讨论班上，一个年轻人报告，其中用了一个很漂亮的定理，Hilbert 说：“这真是一个妙不可言 (wunderschön) 的定理呀，是谁发现的？”那个年轻人茫然的站了很久，对 Hilbert 说：“是你……”。

天才与愚蠢——埃瓦里斯特·伽罗瓦的悲剧故事¹

我需要我的全部勇气在二十岁时死去。——Évariste Galois

如果这个世界上有天才的话，那么埃瓦里斯特·伽罗瓦（Évariste Galois, 1811.10.25—1832.5.31）无疑是最杰出的天才人物之一，可能也是其中最年轻的一位。他的过于短促的光辉的一生，或许是数学史上最震撼人心的悲剧故事。

与其说伽罗瓦死于一场决斗，倒不如说他死于周围的人们的愚蠢，不可饶恕的愚蠢。他在同一个接着一个的不可战胜的蠢才的斗争中，耗尽了他的生命。

1811年10月25日，埃瓦里斯特诞生在巴黎城外莱茵河畔的布尔-拉-林（Bourg-la-Reine）小城里。他一生中最初十一年是幸福的。

1823年，伽罗瓦十二岁时进了巴黎的路易大帝皇家学院（collège Louis-le-Grand）。那地方阴森可怕，戒备森严，看上去像一座监狱，实际上也简直就是一座监狱。

随后的一年，他对文学和古典文学的俯首贴耳的兴趣消失了，变得厌烦了；他的数学天才已经在生气勃勃地活跃起来。然而他的教师们终于如愿以偿地让这个“成绩低劣”“行为放荡”的学生降了级。他被迫吞下他的天才所拒绝的残汤剩饭。

正是在这非常厌烦的一年，他发现了勒让德的光辉的几何学。据说就连学生当中比较好的数学人才，要掌握勒让德，一般也需要两年的时间。伽罗瓦从头到尾读完了几何学，就像其他孩子读一本关于海盗的故事那样容易。一个富于创造性的数学家所创作的这件艺术品唤起了他的热情。

学校提供的代数方面的课本使他厌恶，伽罗瓦轻蔑地把它扔在一边，直接从当时最伟大的大师拉格朗日那里去学他的代数。后来他阅读了阿贝耳的著作。他在课堂上的数学成绩平平；传统的课程对于一个数学天才说来是微不足道的，也是掌握真正的数学所需要的。

他的教师们抱怨他在给他指定的工作中——当他打算在上面花费一点心思时——马虎潦草和古里古怪，故意地用无休止的“胡闹”来使人厌烦。这最后一点不是指干坏事，因为伽罗瓦并没有恶意。那仅仅是用一个强烈的措词，来描述一个第一流的数学天才，决不能把他的聪明才智浪费在迂腐的学究们详细阐述的、无用的修辞学上。

当伽罗瓦在十六岁时已经很好地开始了他那重大发现的事业时，他的数学教师仍在不断地对他唠唠叨叨，就像一只母鸡孵出了一只小鹰，而不知道怎样把这个不守规矩的家伙的双脚固定在谷仓院子里那一大堆脏土上一样。

伽罗瓦没有准备就去参加了巴黎综合工科学校的竞争性的入学考试，但结果是，“一个高等智力的报考者，败在一个智力低下的主考人手中了。*Hic ego barbarus sum quia non intelligor illis* [因为他们不了解我，我是一个野蛮人。]”²。这次失败使他的生活更加痛苦。

伽罗瓦十七岁时就在方程理论方面做着划时代的重要发现，并写成论文准备呈交科学院。柯西答应递交这篇论文，但是忘记了，最后竟不称职到把作者的摘要遗失了。这只不过是一连串类似的灾难中的第一件。

在他十八岁的时候，又有两个灾难给伽罗瓦的性格带来了最后的影响。他第二次参加综合工科学校的入学考试。一些连给他削铅笔也不配的人坐在那里评判他。结果是可想而知的。伽罗瓦失败了。这是他最后的机会；综合工科学校的大门永远对他关闭了。

那次考试已经成为一个传奇。伽罗瓦几乎完全凭脑子工作的习惯，使他在黑板面前处于非

¹本文主要依据E. T. Bell著《数学精英》第二十章节选、改编而成，也参考了其他文献。

²将近四分之一世纪以后，专门维护投考综合工科学校和师范学院的考生利益的数学杂志《新数学年鉴》的编辑泰尔奎姆这样评论伽罗瓦的失败和另一个例子中主考人的难以理解的裁决。

常不利的地位。粉笔和板刷妨碍了他——直到他对其中的一样找到合适的用途才好一些。在考试的口试部分，一个主考人冒昧地和伽罗瓦辩论一道数学难题。这个人明明错了，但又很固执。伽罗瓦眼见他作为一个数学家和综合工科学校的民主自由的职业斗士的全部希望和整个生活正在落空，他完全失去了忍耐。他知道他已经失败了。在一阵愤怒和失望中，他把板刷扔到那个折磨他的人的脸上，狠狠地击中了。

最后的打击是伽罗瓦的父亲的惨死。……从此以后，他怀疑到处都存在着他所憎恨的不公正，他在任何事情上都看不见善良。

1830年2月，伽罗瓦十九岁时，写出了包含了一点他关于代数方程论的伟大工作的论文，并满怀希望地提交给科学院，参加数学大奖赛。它远远超过了当时已经做过的任何东西，正像伽罗瓦完全公正地说的：“我作出的这些研究成果，将使许多著名学者对他们的研究踌躇不前。”

傅立叶把手稿带回家去审阅，但是还没有来得及看就去世了。人们在他死后搜寻他的文件时，这篇手稿连一点踪迹也没有找到。这就是伽罗瓦最后一次听说起它。他的憎恨加深了，他全力投入政治活动，站在当时被禁止的激进派、共和派一边。

被学校开除后，伽罗瓦在泊松的鼓励下向科学院呈交了一篇有关方程的一般求解——现在称为“伽罗瓦理论”——的论文，但是泊松提交了一份敷衍塞责的报告，说这篇论文是“不可理解的”。这是终于使人不能忍受的最后一击。伽罗瓦把他的全部力量投入了革命的政治活动。他写道：“如果需要一具尸体来唤起人民，我愿献出我的。”

1831年5月9日，大约两百个青年共和党人举行了一次宴会。这次聚会的整个气氛是革命的和挑衅的。伽罗瓦站起来，一手拿着酒杯，一手拿着打开的小刀，提议：“为国王路易·菲利普干杯！”他的同伴们误解了祝酒的意义，把他嘘了下来。然后他们看见了打开的小刀。他们把这解释为对国王的性命的威胁，狂喊着表示赞同。伽罗瓦一时间成了英雄；第二天被捕，并被投进了圣·佩拉热（Sainte-Pélagie）监狱。

伽罗瓦虽被无罪释放，但他的自由没有保持多久。1831年7月14日，共和党人准备举行一次庆祝活动之时，伽罗瓦由于在当局眼中被视为一个“危险的激进派”，无端地被拘禁了。

“怎么？你只喝水？离开共和党，回到你的数学那里去吧。”——“没有酒和女人，你永远不会成为一个人。”专心于数学的伽罗瓦被这些酒徒刺激得受不住了，他抓起一瓶白兰地，不知道它是什么，也不管它是什么，就喝了下去。一个正派的同监犯照顾他，直到他清醒过来。当他认识到他做了什么时，他的这次丢脸使他受到极大的打击。

伽罗瓦假释出狱后，偶然地经历了他的第一次、也是唯一的一次恋爱。在这件事上，也像在其他事情上一样，他是不幸的。一个不足取的姑娘（“低级客栈里的卖弄风骚的女人”）找上了他。伽罗瓦对待这件事很狂暴，他讨厌爱情，讨厌他自己，也讨厌这个姑娘。

从两封信中摘录的话，提出了通常作为事实真相被接受的情况：伽罗瓦刚刚获释就与政敌发生了争吵。这些“爱国者”只是一心想要打架，不幸的伽罗瓦命中注定在一场决斗中和他们较量。在一封日期为1832年5月29日的“致全体共和党人书”中，伽罗瓦写道：“我请求爱国者和我的朋友们不要为了我不是为祖国而死责备我。我是作为一个无耻的、卖弄风情的女子的受害者而死。我的生命是在一场可悲的争吵中熄灭的。”在另一封写给两个未指名的朋友的信中，他写道：“两个爱国者已经向我提出挑战——拒绝，对我来说是不可能的。……既然命运没有给我足够的时间，让我活到我的国家知道我的名字，请保留对我的记忆吧。我至死是你们的朋友。”

这些就是他写的最后的话。在写这些信之前，整个晚上，他把飞逝的时间用来焦躁一气写出他的科学上的最后遗言，在死亡之前（他预见到能够追上他）尽快地写，把他丰富的思想中那些伟大的东西尽量写一些出来。他不时中断，在纸边空白处写上“我没有时间，我没有时间(*Je n'ai pas le temps*)”然后又接着涂写下一个极其潦草的提纲。他在天亮之前那最后几个小时拼命写出的东西，将使世世代代的数学家们忙上几百年。他一劳永逸地给一个折磨了数学家

们达几世纪之久的谜，找出了真正的解答。³

伽罗瓦把他的遗嘱委托给他忠实的朋友奥古斯特·谢瓦利埃 (Auguste Chevalier)。他结束说：“请雅可比或高斯公开提出他们的意见，不是对这些定理的正确性，而是对它们的重要性。我希望以后会有人发现，辨读这一堆写得很潦草的东西，对他们是有益的。”⁴

1832 年 5 月 31 日⁵清晨很早的时候，伽罗瓦在“决斗场”与他的敌手相遇。决斗是在二十五步的距离用手枪对射。伽罗瓦倒下了，肠子被打穿。九点钟的时候，一个路过的农民把他送到利尚医院。他的弟弟，他的家人中唯一得到通知的一个，流着泪赶到了。伽罗瓦努力以一种坚忍精神去安慰他的弟弟：“不要哭，”他说，“我需要我的全部勇气在二十岁时死去。”

1832 年 5 月 31 日上午，伽罗瓦在未满二十一周岁之时去世了。他被埋葬在南公墓的普通壕沟里，所以今天埃瓦里斯·伽罗瓦的坟墓已无踪迹可寻。他不朽的纪念碑是他的著作，共六十页。

愚蠢的人们夺去了一位天才的生命，但抹不去真理的光辉。他为之奠定基础的整个现代数学理论的大厦见证了它的荣耀。愿埃瓦里斯·伽罗瓦在后世数学家的景仰与怀念中安息！

参考文献

- [1]E. T. Bell: *MAN OF MATHEMATICS*, 中译本数学精英-数学家的故事
- [2]J. Liouville (edit) : *OEUVRES MATHÉMATIQUES D'ÉVARISTE GALOIS*, Journal de mathématiques pures et appliqués, XI, 1846, P381-444
- [3]A. Dalmas: 伽罗瓦传, 邵循岱译, 商务印书馆, 1981 年 9 月

数学名言

我此生没有什么遗憾，死亡并不可怕，它只不过是我遇到的最后一个函数。 ——
[法] J · L · 拉格朗日

在我看来，一个人如果要在数学上有所进步，他必须向大师学习，而不是向大师的徒弟学习。

—— [挪威] N · H · 阿贝尔

如果我们想要预见数学的未来，适当的途径是研究这门科学的历史和现状。
—— [法] H · 庞加莱

³ 一个多项式的根如何用它的系数经过四则运算和开方表示出来，即所谓用根式解方程的问题，是 19 世纪以前代数的一个主要问题。人们在公元 1500 年左右获得了三、四次方程的求根公式，从此以后致力于五次以上方程的代数解法，但经过近三百年的努力一直未能成功。后来，鲁菲尼和阿贝尔先后独立地证明了五次以上一般方程的根是不能用根式解的。天才的伽罗瓦悉心研究了拉格朗日、阿贝尔和高斯的著作之后，将一个多项式 $f(x)$ 的根可用根式解的条件转化成它的群 G_f 所应具备的条件，那就是 G_f 应是可解的。参见聂灵沼、丁石孙著《代数学引论》第二版第 238、239 页。

⁴ 十四年后的 1846 年，刘维尔把伽罗瓦生前的论文整理并发表在他所创办的《纯粹与应用数学杂志》(Journal de mathématiques pures et appliqués) 上，此时伽罗瓦的理论才受到人们的重视。他在序言中评价道，这一真正值得数学家们注意的方法，足以确保我们的同胞跻身于堪称奠基人的为数不多的学者的行列。(Cette méthode, vraiment digne de l'attention des géomètres, suffirait seule pour assurer à notre compatriote un rang dans le petit nombre des savants qui ont mérité le titre d'inventeurs.)

⁵ 一说 5 月 30 日。可能《数学精英》的说法是不准确的。

在孙念增数学分析奖学金颁奖会上的发言

吴昊¹

尊敬的各位老师、同学：

大家下午好。

经过两年大学学习，我很荣幸获得“孙念增教授数学分析奖学金”。孙教授是一位把教育作为自己终身事业的好老师，而这份奖学金也正是他的学生为纪念孙教授，并继承孙教授对清华大学及中国教育事业贡献力量的遗志所设立的。今天，我获得了这份奖学金感到的是一份肯定与希望。

回想自己两年的学习生活，脑中闪现的是数学系老师的循循善诱，春风化雨，甘霖润物；是班集体中同学们的朝夕相处，情同手足，亲如姐妹；是班主任、辅导员无微不至的关心与爱护。我成绩的取得离不开周围老师、同学的帮助，今天作为对他们的回报，我想在这里谈一点学习心得体会，希望能够为新同学提供一点帮助，使你们能够更快地适应大学学习生活。

来到大学，你们面临的第一个难题恐怕就是接受这样一个事实：曾经辉煌的你们、曾经被老师同学关注的你们，现在很有可能只是清华默默无闻的一个普通学生。这个事实确实有些残忍，但是请你们先默默地接受它。在清华这样一个人才济济的地方，默默无闻是绝大部分同学的归宿，但是我认为正是这种“默默无闻”为我们提供了求知的良好环境——那就是安静、专心、和时间，这些都是在失意寂寞时才有，在得意热烈时所缺的。歌德说得好：“追求伟大事物的人必须全力以赴，巨匠在限制中才能表现自己。”失意挫折期往往是机遇期。在受挫落寞时，我们更要全力以赴，并请牢记华罗庚老师的一句话“一分辛苦一分才”。你的任何付出必将得到回报。

其次是学习方法从高中向大学的转变。在我印象里，概括最好的是张贤科老师的“动脚，动手，动脑”。

动脚——迈动你的双脚，到图书馆去查阅资料。对于一个像我一样进入大学后刚刚开始使用图书馆的学生，查阅资料的一个重要方面是熟悉各学术期刊的情况，熟悉各种文摘评论的查阅方法，熟悉图书馆的各种资料的分布、排列、版制等等。这些看起来琐碎，其实对于掌握信息是必不可少的，也只有在实践中不断积累。其次，图书馆也是一个诱人读书的极妙环境，是理科治学者陶冶心灵的最佳场所。古者孟母三迁，亚圣乃出。科学大家必在书侧。

动手，动脑——平时要勤动手写札记、眉批、摘记等，要多动脑筋，看是否有所启

¹ 基科 53-基数 51；作者在全部四门数学分析课中取得 100 分，因而获得孙念增数学分析奖学金。

发。动手，学习过程中的“百思不得其解”，一定要及时记录下来，跟踪追击，刨根问底，切不可得过且过；读书过程中的“偶有所得”，一定要及时记录下来，这是孕育创造性的温床，是科研想法的萌芽。动脑，那就是学习过程中的反思与举一反三。理科学习时要重写改述，重理思路，多多举例，动手计算，反向思考，提出问题，给出新证明，推广应用，得到新结果。华罗庚曾有著名的读书公式：薄——厚——薄。第一步，见其大概；第二步，详细研读，加眉批注记，加纸条笔记，加心得体会甚至推广；第三步，经过反复，逐渐达到融会贯通，切实掌握，书中理论变得自然了然。而这些过程中，都离不开向老师、助教、同学的虚心的请教与积极的讨论，我本人也从中受益匪浅。定义、定理的来龙去脉更为清晰了，复杂的证明过程变得信手拈来。老师对自己想法的一点肯定，使我更有动力，原本枯燥的学习过程变得其乐无穷。

最后，有一点我想说明，那就是“一本书主义”。治学之路盘曲而上，由一段一段阶梯组成。在每段阶梯，要读烂一本书。我认为没有必要拿许多属于同一阶梯水平的书反复对看。开学后，你们会慢慢发现，大学的学习时间是非常宝贵的，容不得你把同一件事做很多遍，那么选择一本合适的书，精读细研，直到切实理解掌握，应该是一种比较节省时间的方法。待到对一个环节的基本理论真正掌握后，如果有时间，再翻看其他同类的书，就会发现这些书多是大同小异，讲法、符号不同而已。当然也有部分章节内容是新的，逐渐补上就很容易了。掌握一个环节后，最好及时转入更高环节，不要在原有环节上徘徊。攀登之路尚远，前面更美的境界在等待。

以上只是我的一点学习体会，供新同学参考。你们可以对其中的想法做出自己的判断，逐渐摸索出适合自己的大学学习方法。希望清华大学四年的生活不但能够丰富你们的知识，还能够给你们更多启发。谢谢。

寄青年

华罗庚

1981年

发愤早为好，苟晚休嫌迟。

最忌不努力，一生都无知。

访谈

采访对象：卢旭光老师

记者：吴昊

时间：2008-3-18 下午 3: 00-5: 45

受访者基本介绍：

卢旭光教授，1955年11月出生于辽宁，清华数学系党委书记，教过基科54-58班三个学期的数学分析课，以及基科5字班的高等分析课，现为基科7字班数学分析教师。卢老师以过人的拼劲全身心地投入到教学工作中。他对数学推理和演算的细节一丝不苟，不仅使我们打下了数学分析的坚实基础，而且展现了19世纪以来经Cauchy、Weierstrass、Cantor、Lebesgue等伟大数学家严格化了的近现代分析学的精髓。卢老师以其特有的严谨、勤奋、热情和幽默感染着课堂上的每一个学生，引导我们走进了高等数学的大门。

卢老师：

咱们先漫谈一些吧。

谈起中国数学的发展，要从“传统”开始。欧洲一些国家早期就崇尚科学，他们一直以来对科学有着认同，并且他们重视逻辑（演绎、归纳等）。他们有着长远的关于严谨逻辑分析的传统。比如俄罗斯，即使在集权时期，他们也比较重视数学家。他们的传统文化和价值观，使那里的人民更乐于从事科学工作。而同一时期，中国还在实行“科举”，那时人们的思想理念偏重于人文方面，崇尚写意，衡量标准依赖个人的经验和感觉，难以普及。就是说在中国近代科学的发展起步较晚。

而在当今社会中，中国乃至全世界都开始重视数学。二十多年前美国总统的一位科学顾问（E. David）指出：“很少有人认识到，当前被如此广泛称颂的高科技，本质上是数学技术”。数学正逐步成为高科技的核心，数学的发展前景也更为广阔，我认为原因有三：

1. 效率。随着社会飞速发展，推新、抢先、高效是企业发展的命脉，这些离不开方案的合理设计。合理的方案以及对效益的预测都需要数学。

2. 教训。众所周知的“千禧年危机”，是个很好的例子。由于系统基础不完善造成返工，其中的代价是高额的，这严重抵消了已有利润。新一代企业家意识到了这一点，他们开始更重视“标准化”，完善系统的基础，进行合理论证。这些论证是数学上严谨的论证，只有这样才能够预见可能的损失，并把它最小化。这也是现在很多企业更愿意招收数学系学生的一个重要原因。虽然开始的时候这些学生可能不熟悉业务，但经过一段适应期，他们的强大的逻辑思维和深入的分析能力将使企业收益。

3. 兴趣。这一点跟物理学家有相似处。物理学家认为支配世界的规律应该是简单的。他们想借助数学，发现事物之间的联系，建立更高层次的统一。所以支持很多数学家、物理学家从事科学研究的主要是他们的兴趣，他们认为科学是有趣的。今天来看，这兴趣将随着社会需要的增加而更加广泛、持久。

记者：卢老师，您认为清华的学生存在什么问题？数学系的学生现在怎样？

卢老师：

调查显示，现在中国科学界名气大的人物例如院士群体中，本科清华80年代以来毕业的人很少。这与国家的投入不成比例。个人认为，原因可能是这样的，不少清华的学生从小学到中学到大学一路领先走来很辛苦，同时已经得到了足够多的荣誉，这些可能导致他们进入社会

做出一定成绩后有歇歇脚吃老本的念头，拼劲小了，倾向于安逸的生活。另一方面，能够在数学，或者说科学领域坚持的人少了。这可能与年轻一代受到的不同价值或人生观的影响有关。不过对于青年人，相比于个人的性格气质来说，价值观的转变或许更容易些。当然我们教师在这个问题上也有责任……近十几年由于清华数学系师资力量迅速增强，杰出青年基金获得者和长江教授逐渐增多，加上本科规模扩大和基科班办学，生源很好，涌现出了一些真正对数学感兴趣、有志在数学领域有所作为的优秀学生，他们的学习和学术研究进步很快，潜力很大，在国外的声誉也不错。我觉得对清华的多数学生而言，关键问题只是能否做得更长久。十年后看结果吧。

记者：那您觉得我们还应该注意哪些方面呢？

卢老师：

1. 希望能有更多的人保持对数学的兴趣。
2. 一定要坚持，要有拼劲，这当然与兴趣相辅相成。目标始终如一，踏实苦干，并且相信只要坚持下去就一定可以成功。我清楚记得，南开的张伟平教授（去年当上院士）几年前在我系作学术报告，在谈到著名人物的工作时讲道：“勤奋就是天才”。他的这句话我们都很理解！是的，有大智慧有远大抱负的人最懂得勤奋，勤奋对他们来说是常态而不是累，勤奋导致他们天才工作的诞生。
3. 在选择研究方向和研究课题时尽可能考虑对象的“生命力”和“发展潜力”。
4. 学术工作取得一定成绩或数学技术达到一定程度后，应注意思想观点上的提高：向大师学习，追求对问题的理解，探索不同事物间的联系。事实上这正是有趣之所在。
5. 随着年龄的增加，要逐渐有责任感，当发觉自己从事的学科方向有些散乱，某些发展似乎偏离主线时，应感到自己有责任去澄清或统一，至少应有这方面的意识。为科学做出了基础性贡献，为后人带来了方便，才是我们最大的欣慰。

最后对你们现在的学习说一点具体的。学习是要下功夫的，要靠平常的努力，靠考试前的突击不是学习的办法，因为忘得快，以后会觉得荒废了时光。要多跟老师交流，老师们都是很乐于与同学交流的。要充分利用大学资源，你们有了自己的想法后，通过与老师交流可以让你们提高更多更快，并且少走弯路。

记者：谢谢卢老师！

卢老师：不用谢，欢迎你们常来。我系教师有很多高手，采访他们收获会更大。

数学谜语答案

1. Bananach space.
2. Abelian grape.
3. A door must be either open or closed, and cannot be both, while a set can be open, or closed, or both, or neither!
4. 德国数学家 Max Noether，因为他的女儿 Emmy Noether 被称为“抽象代数之母”。

征稿启事

《荷思》是清华大学数学系学生自主创办的数学学术刊物，主要面向读者为各个院系中对数学感兴趣的本科生。

本刊欢迎全校师生任何与数学有关的投稿，无论是长篇的论述，还是精彩的小品，抑或学习/教学的心得、习题的妙解。在原作者的允许下，推荐他人的作品也同样欢迎。

为了编辑方便，建议投稿者能够提供电子版，并且采用 Word 或 LATEX 排版。

来稿请注明作者，联系方式。

在此我们对所有热心支持本刊的读者表示由衷的感谢！

投稿请寄： THUmath@googlegroups.com

《荷思》编辑部
2008-04

主办： 清华大学 数学科学系 《荷思》编辑部
主编： 吴玉清

编委： (依姓名笔划排序)

石 权 杜升华 吴玉清
吴 昊 陈凌骅 胡悦科

封面： 陈凌骅

排版： 陈凌骅 杜升华

联系本刊： THUmath@googlegroups.com

北京市 海淀区 清华大学 紫荆 9 号楼 205B 100084
010-515-31844 吴玉清



我们所做的事可能是渺小的，
但它具有某些永恒的性质。

——〔英〕G.H.哈代

清华大学
数学科学系

