

本书第一版荣获

■ 国家图书馆第八届文津图书奖

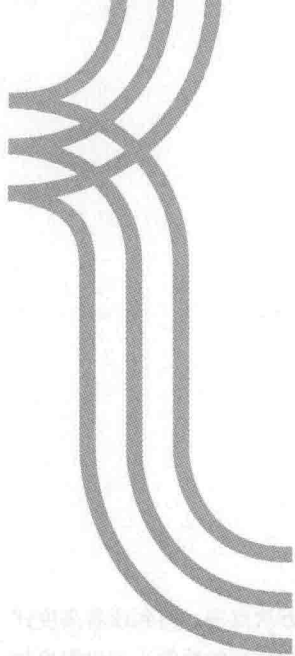
# 数学之美

吴军 著 第二版

JUST { PUB

1

人民邮电出版社  
POSTS & TELECOM PRESS



JUST PUB

# 数学之美

第二版

Beauty of Mathematics

吴军 著

人民邮电出版社  
北京

## 图书在版编目(CIP)数据

数学之美 / 吴军著. -- 2版. -- 北京: 人民邮电出版社, 2014. 11  
ISBN 978-7-115-37355-7

I. ①数… II. ①吴… III. ①电子计算机—数学基础  
IV. ①TP301.6

中国版本图书馆CIP数据核字(2014)第235987号

## 内 容 提 要

几年前,“数学之美”系列文章原刊载于谷歌黑板报,获得上百万次点击,得到读者高度评价。正式出版前,吴军博士几乎把所有文章都重写了一遍,为的是把高深的数学原理讲得更加通俗易懂,让非专业读者也能领略数学的魅力。

本书第一版上市后深受广大读者欢迎,并荣获国家图书馆第八届文津图书奖。读者说,读了《数学之美》,才发现大学时学的数学知识,比如马尔可夫链、矩阵计算,甚至余弦函数原来都如此亲切,并且栩栩如生,才发现自然语言和信息处理这么有趣。

而今,数学在信息产业中的应用越来越广泛,因此,作者在第二版中增加了一些内容,尤其是针对大数据和机器学习的内容,以便满足人们对当下技术的学习需求。

- 
- ◆ 著 吴 军
  - 责任编辑 俞 彬
  - 审稿编辑 李琳骁
  - 策划编辑 周 筠
  - 责任印制 焦志炜
  
  - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
  - 邮编 100164 电子邮件 315@ptpress.com.cn
  - 网址 <http://www.ptpress.com.cn>
  - 北京铭成印刷有限公司印刷
  
  - ◆ 开本: 720×960 1/16
  - 印张: 21.5
  - 字数: 284千字
  - 印数: 166 001 - 181 000册
  - 2014年11月第2版
  - 2014年11月北京第1次印刷
- 

定价: 49.00元

读者服务热线: (010)81055410 印装质量热线: (010)81055316  
反盗版热线: (010)81055315

本书谨献给我的家人。

愿科学之精神在国民中得到普及，愿中国年轻的一代涌现更多的杰出专业人才。

## 第一版读者赞誉

其实这本书越早读越好，最好在中学阶段，中国教育最失败的就是学生从上课的第一天到考试结束，都不知道他学的东西能做什么。《数学之美》正好能告诉学生（包括大学生），从小到大学的那些数学知识可以如何改造世界。

—— gj83 / 亚马逊读者

很棒，抽象思维的极致。探讨如何将现实问题转化为合适的数学模型，从不同思路剖析问题的本质。是很好的思维训练。

—— matthewluzju / 亚马逊读者

很多知识讲的都浅显易懂，用作者本人的话讲就是：一个技术或者问题“术”和“道”之分！这本书讲的比较偏向“道”！非常好，这种书越早看越好！

—— 羽翼 / 亚马逊读者

我是学计算机的。大学里学了很多数学课，当时并不知道这些数学知识有什么用。读研时，因为接触到数据挖掘，这才发现，很多看起来很智能的技术，其实是统计学的应用。例如语音识别、图像识别、OCR、输入法中的智能联想、自然语言处理中的词性标注。为此，回头重读了概率论、随机过程、离散数学、计算方法，等等。如果在大学时能看到这本《数学之美》的话，我相信数学会学得更好的。好可惜。现在的年轻人有福，可以直接从吴军博士的这本书里启蒙。

—— 东风“专注成就专业” / 亚马逊读者

我们本科阶段在学习线代、统计、图论、通信原理时常常会怀疑这些理论到底有什么用呢？读了这本书算是长了见识，原来这些理论还可以这么玩。比如计算机自然语言处理可以抽象成非常简单的通信模型和统计学模型，然后一个简单条件概率公式加上一个马尔可夫假设就可以做到机器翻译和语音识别……读这本书的过程也是对数学建模思维的训练，让自己从更高的数学模型层面去抽象问题，去寻找一个正确的解决方案。

——花生 / 豆瓣读者

看完这本书后，我发现我还真是低估了数学的作用，很多复杂的问题，竟然用那么简单的数学模型就解决了，这对我的冲击很大。另外，在这书里我特别喜欢阿米特·辛格博士的那一章。吴军博士就用寥寥几页讲解了辛格博士处理事情的方法和原则，先帮用户解决主要的问题，再决定要不要纠结在次要的部分上；要知道修改代码的所作所为，知其所以然；能用简单方法解决就用简单的，可读性很重要。

——好了个yin / 豆瓣读者

读了这本书，才第一次惊讶的发现，数学原来并不只是枯燥的考试题，而是人类理解世界最重要的一种工具。数学不仅真的很美，而且真的很有用。这本书最大的价值就在于，它没有直接告诉你答案，而是让你带着新的启示、新的方法以及新的眼光、新的境界来重新理解这个世界。世界太复杂了，必须分步、分类从而化繁为简，数学就是一件非常有力的工具。从此，“数学无用论”、“学数学只是锻炼思维”的观点，在我这里没有市场了。从此，再也不要将数学高手看成不名一文的书呆子了，没准他们里头就有下一个维特比。

——野笛无腔 / 豆瓣读者

推荐有一定数据分析基础的同学读吴军博士的《数学之美》，特别是有文本处理基础的人读起来更加受益。但即使没有基础的不妨读一读，相

信会宏观上认识到——数学不仅有用而且还有无穷魅力。

——池明旻 / 新浪微博读者

在备考之余品味《数学之美》带给我的感悟。让人对科学充满敬仰和激情。对大师的渴望。这本书带给我的不仅仅是书中的那些东西，似乎某些东西被它点亮了。

——刺客 -110 / 新浪微博读者

《数学之美》展示的是一个我们陌生而又熟悉的世界，熟悉是因为它展示的东西我们天天使用，百度搜索、密码加密、语音识别，等等；陌生是因为它剥去山竹顽固的外壳，露出藏在里边白嫩的果实一般诱人的数学算法。在不禁惊叹数学奥妙之时，也在感慨自己为何当初不学好数学，就这样错过一个进入新世界发现新知识的接口。希望小学弟小学妹们能看看这本书，喜欢上数学，给自己一个机会去接触去挖掘一个近在咫尺蕴含无穷乐趣的新世界。

——蜀黍怪怪 / 京东读者

个人的角度这本书更适合给初中或者高中的学生看。这本书中尽管掺杂了大量看似很复杂的数学公式和数学原理，但并非抽象的数学自然哲理，而是简单实用的实例展示。简单的实例举证，让读者能理解在 Google 搜索以及在计算机语言识别中所使用的数学模型等。如果能将这种对数学的运用以及思考理念运用到中学的教育当中，让大量参加数学奥赛的学生能真正领略到数学的魅力而不是单一的解题，从而引导学生的数学以及各方面的思考能力，每年数以百万计的奥赛大军当中难道就不能诞生出一个新的 Google “拉里佩奇”！

——戡乱 K1 / 当当读者



## 第二版出版说明

“数学之美”最初是从2006年起在Google中国的官方博客——谷歌黑板报上连载的系列文章。当时应原黑板报版主吴丹丹女士之邀，写一些博客介绍Google的技术，于是我根据自己当时在Google和之前在约翰·霍普金斯大学的研究经历，写了一系列短文，介绍Google产品背后的基本技术，尤其是数学原理。当初我并没有一个完整的写作计划，只打算有空就抽时间写一点，写到哪儿算哪儿。不曾想刊登了几篇之后，受到IT行业广大从业人员和大学生们的关注和喜爱，在互联网上被转载了上万次，读者有上百万之众。于是，在大家的鼓励下，我便陆陆续续写了20多篇。后来因为工作不断变动，写作一度中断过一段时间，但令我感动的是，在这期间始终有读者持续关注这个系列，时不时来问我能否将这个系列写完，有无可能出书。因此，从2010年起，我陆续完成了这个系列最后的几篇，并且开始把它修订成书。

出书比写博客要求高很多。一本好书需要结构系统而文字严谨。为了达到出书的要求，从2010年到2012年，我几乎重写了所有的内容。因此，本书各个章节的标题和主题虽与原来的博客大体相同，但是内容和文字都是新的。在系统性方面，为了便于非IT读者的阅读，我对每个专题都做了背景介绍；同时，为了起到给从事相关工作的工程师做参考的目的，在一些专题的最后，我都给出了一节“延伸阅读”。非IT读者完全可以跳过这些延伸阅读部分，这样并不会影响阅读其他内容。本书在系统性



方面的另一个改进就是调整了章节的顺序，以帮助读者阅读。在严谨性方面，我在腾讯工程师王益等人的帮助下，更正了原来博客中的一些错误，并尽可能补充完善诸多公式推导的过程。

本书的素材来源于我本人的工作。语言信息的处理、互联网技术、数据挖掘和机器学习等都是博大精深而又快速发展的领域，我所做的研究工作也只涵盖了其中很小的一部分。对于我没有涉足过的领域，我没有信心也没有资格写。因此，这本书在内容上并没有全面覆盖上述领域，比如对当今数据挖掘领域的算法、互联网上各种推荐系统的数学模型都鲜有提及。对这些内容有兴趣的读者可以查阅相应的文章和书籍，也希望今后有这方面的专家能够将自己工作的心得写出来，供大家学习参考。因此，从另一个角度讲，这本书对专家学者们来说也算是抛砖引玉；而对于大众读者，则希望这本书能帮助大家领悟数学之道，以便今后解决实际问题时能够举一反三。

2012年，在很多朋友的关心和帮助下，《数学之美》终于以纸本书的形式正式出版了，并且荣获国家图书馆第八届文津图书奖。尤其可喜的是，很多年轻读者，包括中学生，在阅读完这本书后对数学产生了浓厚的兴趣，并尝试将学到的数学知识自觉地应用到日常学习或工作中。从这本书第一版出版至今，已经过去了两年半的时间，虽然数学本身的发展和变化并不大，但是它在信息产业中的应用越来越广泛，因此在第二版中我增加了一些内容，尤其是针对大数据和机器学习的内容，以便满足人们对当下技术的学习需求；同时，根据专家和读者的反馈做了勘误，并更新了部分内容。

在《数学之美》第二版即将出版之际，我要感谢很多人给予我的教育、帮助和鼓励。首先，我要感谢那些把我带到数学王国和信息处理领域的人，包括在我幼年时培养我对数学和自然科学兴趣的父亲，后来把我带进语音和语言处理王国的三位导师：王作英教授、库坦普教授和贾里尼克教授，以及在 Google 不断提携我的诺威格博士和辛格院士。其次，我要感谢在

数学上和我不断交流，并且对我的写作尤其是这本书的创作长期给予支持和帮助的同事和朋友，包括李开复博士、清华大学的李星教授和马少平教授、斯坦福大学的张首晟教授、华中科技大学的周笠教授、亚马逊的郭进博士以及 Google 过去与现在的很多员工。此外，我还要感谢我的夫人张彦帮我校对了全书，我的两个女儿吴梦华和吴梦馨绘制了本书的许多插图。

《数学之美》能够从系列博客成为一本获大奖的畅销书，JUSTPUB 出版团队和人民邮电出版社功不可没。其中 JUSTPUB 的负责人周筠女士主持了本书的出版工作，审稿编辑李琳骁对书稿进行了多次精心的审读和校对，设计师胡文佳对这本书进行了细致认真的排版，人民邮电出版社的俞彬和刘涛等多位朋友为这本书的印制、发行尽心尽力。是整个团队的精诚合作和持续不断的服务，使得《数学之美》受到了这么多读者的欢迎。在此，我向他们表示最诚挚的感谢。

最后要感谢所有热心的读者，尤其是那些帮助本书更正错误、完善内容的朋友。当然，也要感谢在互联网上积极传播这个系列博客以及推荐这本书的媒体、网站和个人。我也希望大家能继续支持《数学之美》。

由于本人水平有限，书中难免存有疏漏和错误，希望读者朋友继续不吝赐教，共同将这本书打造得更完美。

吴军

2014 年 10 月于硅谷

# 第一版序言

《数学之美》是一本非常值得读的书。这本书展现了吴军博士在他多年的科研经历中对科学问题的深入思考。

我于1991年从美国回到清华大学电子工程系工作，与吴军博士是同事，对他在汉语语音识别方面的深入研究印象非常深刻。后来他到美国工作，出版了一本介绍硅谷的书《浪潮之巅》，使我对他的写作激情和水平有了新的认识。

这些年来我在清华大学教书，一直思考着如何让学生能真正欣赏和热爱科学研究，这将有助于他们深入理解自己所从事的研究的价值，进而能逐渐成长为所在领域的大师和领军人物。在这一过程中，恰好发现了吴军博士在 Google 中国的官方博客——谷歌黑板报上连载的“数学之美”系列文章，我非常欣赏这些文章。因此，在很多场合都建议学生跟踪阅读这个系列的博客文章。今天本书出版，与原先的博客文章相比，其内容的系统性和深度又上升到了一个新的境界。

我读《数学之美》有下面几点体会，与大家分享。

## 1. 追根溯源

《数学之美》用了大量篇幅介绍各个领域的典故，读来令人兴趣盎然。典故里最核心的是相关历史事件中的人物。我们必须问：提出巧妙数学思想的人是谁？为什么是“他/她”提出了这个思想？其思维方法有何特点？成为一个领域的大师有其偶然性，但更有其必然性。其必然性就是大师们的思维方法。

## 2. 体会方法

从事科学研究，最重要的是掌握思维方法。在这里，我举两个例子。

牛顿是伟大的物理学家和数学家，他在《自然哲学的数学原理》中叙述了四条法则。其中有“法则 1：除那些真实而已足够说明其现象者外，不必去寻找自然界事物的其他原因”。这条法则后来被人们称作“简单性原则”。正如爱因斯坦所说：“从希腊哲学到现代物理学的整个科学史中，不断有人力图把表面上极为复杂的自然现象归结为几个简单的基本概念和关系。这就是整个自然哲学的基本原理。”这个原理也贯穿了《数学之美》本身。

WWW 的发明人蒂姆·伯纳斯·李谈到设计原理时说过：“简单性和模块化是软件工程的基石；分布式和容错性是互联网的生命。”虽然在软件工程和互联网领域的从业人员数量极其庞大，但能够真正体会到这些核心思想的人能有多少呢？

我给学生出过这样的考题：把过去十年来重要 IT 杂志的封面上重点推荐的技术专题找来看看，瞧一瞧哪些技术成功了，哪些技术是昙花一现，分析一下原因？其答案很有意思：“有正确设计思想方法的技术”未必能够成功，因为还有非技术的因素；但“没有正确设计思想方法的技术”一定失败，无一例外。因此，我也建议本书的读者结合阅读，体会凝练创造《数学之美》的方法论。

### 3. 超越欣赏

数学既是对自然界事实的总结和归纳，如英国的哲学家培根所说“一切多依赖于我们把眼睛紧盯在自然界的事实之上”；又是抽象思考的结果，如法国哲学家笛卡尔所说“我思故我在”。这两个方法成就了目前绚丽多彩、魅力非凡的数学，非常值得欣赏。《数学之美》把数学在 IT 领域，特别是语音识别和搜索引擎方面的美丽之处予以了精彩表达。但在这里我想说的是：欣赏美不是终极目的，更值得追求的是创造美的境界。希望本书的读者，特别是年轻读者能够欣赏数学在 IT 技术中的美，学习大师们的思想方法，使自己成为大师，创造新的数学之美。

李星

2012年4月于北京

## 第二版序言

几年前我曾经给吴军的《浪潮之巅》和《数学之美》的第一版写序，很高兴《数学之美》后来获得了文津奖，并且第二版也即将出版！

《数学之美》最初是作为 Google 资深研究员的吴军应邀在谷歌黑板报上撰写的一系列文章。刚开始，黑板报的版主还有点担心这个系列会不会让读者觉得太理论太枯燥，但很快这个顾虑就被打消了。《数学之美》用生动形象的语言，结合数学发展的历史和实际的案例，谈古论今，系统地阐述了与现代科技领域相关的重要的数学理论的起源、发展及其作用，深入浅出，受到广大读者尤其是科技界人士的喜爱。

我在《浪潮之巅》的序言中曾经说过，在我认识的顶尖研究员和工程师里，吴军是极少数具有强大叙事能力和对科技、信息领域的发展变化有很深的纵向洞察力，并能进行有效归纳总结的人之一。在《数学之美》里，吴军再次展示了这一特点。与《浪潮之巅》不同的是，这次吴军集中阐述了他对数学和信息处理这些专业学科的理解，尤其是他在语音识别、自然语言处理和信息搜索领域多年来积累的认识。从数字和信息的由来，到搜索引擎对信息进行处理背后的数学原理，到与搜索相关的众多领域中奇妙的数学应用，吴军都娓娓道来。他把数学后面的本质思维写得透彻、生动。不得不说，他的文字，引人入胜，也确实让我们体会到数学的美。在他的笔下，数学不是我们一般联想到的枯燥深奥的符号，而是实实在

在源于生活的有趣的现象和延伸。数学，其实无处不在，而且有一种让人惊叹的韵律和美！

伽利略曾经说过，“数学是上帝描写自然的语言”；爱因斯坦也曾说过，“纯数学使我们能够发现概念和联系这些概念的规律，这些概念和规律给了我们理解自然现象的钥匙。”我多年来一直也对信息处理、语音识别领域有着一定的研究，深深体会到数学在所有科学领域起到的基础和根本的作用。“哪里有数，哪里就有美。”在这里，我把《数学之美》真诚推荐给每一位对自然、科学、生活有兴趣有热情的朋友，不管你是搞理科还是搞文科的，读一读数学的东西，会让你受益良多，同时能感受到宇宙和世界的美好与奇妙。这本书尤其可贵之处在于，作者所介绍的内容不仅是他熟知的，而且是他在工作中长期使用的。作者不仅告诉大家为什么可以用那些形式上简单的数学模型解决非常复杂的工程问题，而且还清楚地讲述了他（和同事们）的思考过程，这一点没有实际经验的学者是不可能办到的。

2012年，吴军把之前谷歌黑板报上的“数学之美”系列文章编辑成《数学之美》第一版，花费了大量的心血和时间。他本着十分严谨的态度，在繁忙的工作之余，补充完善了之前的系列，并几乎重写了所有的文章，既照顾了普通读者的兴趣，又兼顾了专业读者对深度的要求，很让人钦佩。此后，吴军把他这两年在 Google 工作的体会总结成新增的两章，通过本书的第二版介绍给大家，相信读者们能因此而进一步理解数学之美。

有时我在想，现在的社会多了一点压力和浮躁，少了一点踏实和对自然科学本质的好奇求知。吴军的这本《数学之美》真的非常好。非常希望吴军今后能写出更多这样深入浅出的好书，它们会是给这个社会 and 年轻人最好的礼物。

李开复

2014年10月



## 第二版前言

数学一词在西方源于古希腊语 μάθημα，意思是通过学习获得的知识。从这个角度来说，早期的数学涵盖的范围比我们今天讲的数学要广得多，和人类的生活也更接近一些。

早期的数学远不如今天神秘，它是非常真实的。与任何事物一样，数学也在不断地演化，而这个发展过程使得数学变得高深起来。数学演化的过程，实际上是将我们生活中遇到的具体事物及其运动的规律不断抽象化的过程。经过几千年的抽象化，大家头脑里能想象的数学只剩下数字、符号、公式和定理了。这些东西和我们的生活似乎渐行渐远，甚至在表面上毫不相关了。今天，除了初等数学，大家一般对数学，尤其是纯粹数学（Pure Mathematics）的用途甚至产生了怀疑。很多大学生毕业后，在大学所学的数学可能一辈子都没有机会应用，几年后就忘得差不多了。这样，很多人也产生了为什么要学习数学的疑问。更加不幸的是，数学专业的毕业生连就业也颇为困难，在中国和美国都是如此。在很多国人眼里，数学家都像陈景润那样戴着厚厚的镜片、言行举止多少有些木讷。因此，在一般人看来，无论是这些抽象的数字、符号、公式和定理，还是研究它们的数学家，似乎都和美没有什么联系。

然而，事实上数学的用途远远超乎人们的想象，甚至可以说在我们的生活中无处不在。且不说那些和我们生活联系相对较少的领域，比如原子

能和航天，都需要用到大量的数学知识。就说我们天天用的产品和技术，背后都有支持它们的数学基础。作为一名工作了 20 多年的科学工作者，我在工作中经常惊叹数学语言应用于实际问题时的魔力。因此，我也很希望把这种神奇讲给大家听。

在古代，最重要的知识，除了对世界的认识 and 了解，就是人与人之间的互通和交流了，我们把它称为广义上的通信。本书的内容也将从这里开始。为了展示数学的美妙之处，我之所以选择了以通信这个领域为切入点，一来是因为数学在通信中应用非常普遍，二来通信和我们的生活息息相关。从工业社会起，通信就占据了人们生活的大量时间。当人类进入电的时代后，通信的扩展不仅拉近了人与人的距离，而且成为带动世界经济增长的火车头。如今，通信及其相关产业可能占到世界 GDP 很大一部分。今天城市里的人们花时间最多的，无非是在电视机前、互联网上、电话上（不论是固定电话还是手机），这些都是这样或那样的通信方式。甚至原本必须人到现场的很多活动，比如购物，也被建立在现代通信基础之上的电子商务逐渐取代。而现代通信，追溯到 100 多年前的莫尔斯电报码和贝尔的电话，再回到今天的电视、手机和互联网，都遵循着信息论的规律，而整个信息论的基础就是数学。如果往更远处看，我们人类的自然语言和文字的起源背后都受着数学规律的支配。

“信”字作为“通信”一词的 50%，表明了信息的存储、传输、处理和理解的重要性。今天每个人都要使用的搜索，以及我们都觉得很神奇的语音识别、机器翻译和自然语言处理也被包括在其中。也许大家想不到，解决这些问题的最好工具就是数学。人们不仅能够非常清晰地用一些通用的数学模型来描述这些领域里看似不同的实际问题，而且能给出非常漂亮的解决办法。每当人们应用数学工具解决了一个个和信息处理相关的问题时，总会感叹数学之美。虽然人类的语言有成百上千种，但处理它们的数学模型却是相同或者相似的，这种一致性也是数学之美的表现。在这本书中，我们将介绍一些数学工具，看看人们是如何利用这些数学工具来处理信息，开发出生活中每天都会用到的产品。