时代教育·国外高校优秀教材精选

PEARSON
Prentice Hall

（英文版）

编码理论中的数学

The Mathematics of Coding Theory

（美）Paul Garrett 著

机械工业出版社
CHINA MACHINE PRESS

时代教育·国外高校优秀教材精选

编码理论中的数学

（英文版）

The Mathematics of Coding Theory

（美）Paul Garrett 著

机械工业出版社

ISBN 0-13-101967-8

北京市版权局著作权合同登记号：图字：01-2004-6881 号

图书在版编目（CIP）数据

编码理论中的数学/（美）加勒特（Garrett,P.）著.－北京：机械工业出版社，2005.1

（时代教育·国外高校优秀教材精选）

ISBN 7-111-15863-6

Ⅰ.编… Ⅱ.加… Ⅲ.编码理论－高等学校－教材－英文

Ⅳ.0157.4

中国版本图书馆 CIP 数据核字（2004）第 135411 号

机械工业出版社（北京市百万庄大街 22 号　邮政编码 100037）

责任编辑：郑　玫

封面设计：饶　薇　　责任印制：石　冉

北京中兴印刷有限公司印刷·新华书店北京发行所发行

2005 年 1 月第 1 版第 1 次印刷

1000mm×1 400mm B5·12.875 印张·501 千字

定价：38.00 元

国外高校优秀教材审定委员会

主任委员：

杨叔子

委员（按姓氏笔画为序）：

丁丽娟　王先逵　王大康　白峰杉　石德珂

史荣昌　孙洪祥　朱孝禄　陆启韶　张润琦

张　策　张三慧　张福润　张延华　吴宗泽

吴　麒　宋心琦　李俊峰　佘远斌　陈文楷

陈立周　单辉祖　俞正光　赵汝嘉　郭可谦

翁海珊　龚光鲁　章栋恩　黄永畅　谭泽光

郭鸿志

出版说明

随着我国加入 WTO，国际间的竞争越来越激烈，而国际间的竞争实际上也就是人才的竞争、教育的竞争。为了加快培养具有国际竞争力的高水平技术人才，加快我国教育改革的步伐，国家教育部近来出台了一系列倡导高校开展双语教学、引进原版教材的政策。以此为契机，机械工业出版社陆续推出了一系列国外影印版教材，其内容涉及高等学校公共基础课，以及机、电、信息领域的专业基础课和专业课。

引进国外优秀原版教材，在有条件的学校推动开展英语授课或双语教学，自然也引进了先进的教学思想和教学方法，这对提高我国自编教材的水平，加强学生的英语实际应用能力，使我国的高等教育尽快与国际接轨，必将起到积极的推动作用。

为了做好教材的引进工作，机械工业出版社特别成立了由著名专家组成的国外高校优秀教材审定委员会。这些专家对实施双语教学做了深入细致的调查研究，对引进原版教材提出了许多建设性意见，并慎重地对每一本将要引进的原版教材一审再审，精选再精选，确认教材本身的质量水平，以及权威性和先进性，以期所引进的原版教材能适应我国学生的外语水平和学习特点。在引进工作中，审定委员会还结合我国高校教学课程体系的设置和要求，对原版教材的教学思想和方法的先进性、科学性严格把关。同时尽量考虑原版教材的系统性和经济性。

这套教材出版后，我们将根据各高校的双语教学计划，举办原版教材的教师培训，及时地将其推荐给各高校选用。希望高校师生在使用教材后及时反馈意见和建议，使我们更好地为教学改革服务。

机械工业出版社

序

数字通信的飞速发展使数论和代数成为信息技术和理论的重要数学工具。自上世纪 50 年代以来，为解决通信可靠性而产生的纠错码理论得到了很大的进步。利用数论、线性代数、近世代数和有限域上代数曲线理论构作出性能良好的纠错码得到了广泛的实际应用。

本书首先用信息论和概率论的观点介绍通信理论的最主要和基本的结果：商农（Shannon）的编码理论。然后用主要篇幅讲述构作纠错码的一些重要方法，在每种方法之前讲述所需要的数论和近世代数工具，最后一章还通俗地介绍了代数几何码，内容先进。本书主要有以下特点：

首先，本书系统地介绍了数学，尤其是有限域和数论在编码理论中的应用。国内专门介绍编码方面的数学基础的教材不多，在这方面可以说是填补了一个空白。书中对编码中用到的数学知识均有介绍（如域、数论、概率论、线性代数、射影几何），使得本书自成体系，在知识的衔接方面做得很好。

其次，对于编码的内容，本书不仅系统地阐述了码论的内容，还涵盖了信息论的基础知识，如信息量、熵和信道容量。

本书的另一大特点就是由浅入深、循序渐进。此外书中配有大量的例题和习题，内容的安排适于讲授和阅读。

本书的作者是美国明尼苏达大学的著名数论教授。他兴趣广泛，写过 2 本数论专著，很受欢迎。

读完这本教材，不仅掌握了数学知识，更看到了它广阔的应用前景，同时还掌握了编码的知识。本书适合作为数学系、信息与计算科学系的代数编码理论课教材。

总之，不论是从选材还是从内容的编排上，这都是一本难得的好书。

清华大学数学系

冯克勤

Preface

This book is intended to be accessible to undergraduate students with two years of typical mathematics experience, most likely meaning calculus with a little linear algebra and differential equations. Thus, specifically, there is *no* assumption of a background in abstract algebra or number theory, nor of probability, nor of linear algebra. All these things are introduced and developed to a degree sufficient to address the issues at hand.

We will address the fundamental problem of **transmitting information effectively and accurately**. The specific mode of transmission does not really play a role in our discussion. On the other hand, we should mention that the importance of the issues of efficiency and accuracy has increased largely due to the advent of the internet and, even more so, due to the rapid development of wireless communications. For this reason it makes sense to think of networked computers or wireless devices as archetypical fundamental practical examples.

The underlying concepts of **information** and **information content** of data make sense independently of computers, and are relevant in looking at the operation of **natural languages** such as English, and of other modes of operation by which people acquire and process data.

The issue of **efficiency** is the obvious one: transmitting information costs time, money, and bandwidth. It is important to use as little as possible of each of these resources. **Data compression** is one way to pursue this efficiency. Some well known examples of compression schemes are commonly used for graphics: GIFs, JPEGs, and more recently PNGs. These clever file format schemes are enormously more efficient in terms of filesize than straightforward bitmap descriptions of graphics files. There are also general-purpose compression schemes, such as **gzip**, **bzip2**, **ZIP**, etc.

The issue of **accuracy** is addressed by **detection** and **correction** of errors that occur during transmission or storage of data. The single most important practical example is the TCP/IP protocol, widely used on the internet: one basic aspect of this is that if any of the *packets* composing a message is discovered to be mangled or lost, the packet is simply **retransmitted**. The detection of *lost* packets is based on numbering the collection making up a given message. The detection of *mangled* packets is by use of 16-bit **checksums** in the *headers* of IP and TCP packets. We will not worry about the technical details of TCP/IP here, but only note that **email** and many other types of internet traffic depend upon this protocol, which makes essential use of rudimentary error-detection devices.

And it is a fact of life that dust settles on CD-ROMs, static permeates network lines, etc. That is, there is **noise** in all communication systems. Human natural languages have evolved to include sufficient **redundancy** so that usually much less than 100% of a message need be received to be properly understood. Such

redundancy must be *designed* into CD-ROM and other data storage protocols to achieve similar robustness.

There are other uses for detection of *changes in data*: if the data in question is the operating system of your computer, a change not initiated by you is probably a sign of something bad, either failure in hardware or software, or intrusion by hostile agents (whether software or wetware). Therefore, an important component of **systems security** is implementation of a suitable procedure to detect alterations in critical files.

In pre-internet times, various schemes were used to reduce the bulk of communication without losing the content: this influenced the design of the telegraphic alphabet, traffic lights, shorthand, etc. With the advent of the telephone and radio, these matters became even more significant. Communication with exploratory spacecraft having very limited resources available in deep space is a dramatic example of how the need for efficient and accurate transmission of information has increased in our recent history.

In this course we will begin with the model of communication and information made explicit by Claude Shannon in the 1940's, after some preliminary forays by Hartley and others in the preceding decades.

Many things are omitted due to lack of space and time. In spite of their tremendous importance, we do not mention **convolutional** codes at all. This is partly because there is less known about them mathematically. Concatenated codes are mentioned only briefly. Finally, we also omit any discussion of the so-called turbo codes. Turbo codes have been recently developed experimentally. Their remarkably good behavior, seemingly approaching the Shannon bound, has led to the conjecture that they are explicit solutions to the fifty-year old existence results of Shannon. However, at this time there is insufficient understanding of the reasons for their good behavior, and for this reason we will not attempt to study them here. We *do* give a very brief introduction to **geometric Goppa codes**, attached to *algebraic curves*, which are a natural generalization of Reed-Solomon codes (which we discuss), and which exceed the Gilbert-Varshamov lower bound for performance.

The exercises at the ends of the chapters are mostly routine, with a few more difficult exercises indicated by single or double asterisks. Short answers are given at the end of the book for a good fraction of the exercises, indicated by '*(ans.)*' following the exercise.

I offer my sincere thanks to the reviewers of the notes that became this volume. They found many unfortunate errors, and offered many good ideas about improvements to the text. While I did not choose to take absolutely all the advice given, I greatly appreciate the thought and energy these people put into their reviews: John Bowman, University of Alberta; Sergio Lopez, Ohio University; Navin Kashyap, University of California, San Diego; James Osterburg, University of Cincinnati; LeRoy Bearnson, Brigham Young University; David Grant, University of Colorado at Boulder; Jose Voloch, University of Texas.

Paul Garrett
garrett@math.umn.edu
http://www.math.umn.edu/~garrett/

Contents

1

Probability

1.1 Sets and functions

Here we review some relatively elementary but very important terminology and concepts about *sets*, in a slightly abstract setting.

Naively, a **set** is supposed to be a collection of 'things' (?) described by 'listing' them or prescribing them by a 'rule'. Please note that this is *not* a precise description, but will be adequate for most of our purposes. We can also say that a **set** is an *unordered list* of *different* things.

There are standard symbols for some often-used sets:

$$\begin{array}{rcl} \phi & = & \{\} = \text{empty set} = \text{set with no elements} \\ \mathbf{Z} & = & \text{the integers} \\ \mathbf{Q} & = & \text{the rational numbers} \\ \mathbf{R} & = & \text{the real numbers} \\ \mathbf{C} & = & \text{the complex numbers} \end{array}$$

A set described by a *list* is something like

$$S = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

which is the set of integers greater than 0 and less than 9. This set can also be described by a *rule* like

$$S = \{1,2,3,4,5,6,7,8\} = \{x : x \text{ is an integer and } 1 \leq x \leq 8\}$$

This follows the general format and notation

$$\{x : \ x \text{ has some property}\}$$

If x is in a set S, then write $x \in S$ or $S \ni x$, and say that x is an *element* of S. Thus, a set is the collection of all its elements (although this remark only explains the *language*). It is worth noting that the *ordering* of a listing has no effect on a set, and if in the listing of elements of a set an element is *repeated*, this has no effect. For example,

$$\{1,2,3\} = \{1,1,2,3\} = \{3,2,1\} = \{1,3,2,1\}$$

A **subset** T of a set S is a set all of whose elements are elements of S. This is written $T \subset S$ or $S \supset T$. So always $S \subset S$ and $\phi \subset S$. If $T \subset S$ and $T \neq \phi$ and $T \neq S$, then T is a **proper** subset of S. Note that the empty set is a subset of *every* set. For a subset T of a set S, the **complement** of T (inside S) is

$$T^c = S - T = \{s \in S : s \notin T\}$$

Sets can also be elements of other sets. For example, $\{\mathbf{Q}, \mathbf{Z}, \mathbf{R}, \mathbf{C}\}$ is the set with 4 elements, each of which is a familiar set of numbers. Or, one can check that

$$\{\{1,2\},\{1,3\},\{2,3\}\}$$

is the set of two-element subsets of $\{1,2,3\}$.

The **intersection** of two sets A, B is the collection of all elements which lie in *both* sets, and is denoted $A \cap B$. Two sets are **disjoint** if their intersection is ϕ. If the intersection is *not* empty, then we may say that the two sets **meet**. The **union** of two sets A, B is the collection of all elements which lie in *one or the other* of the two sets, and is denoted $A \cup B$.

Note that, for example, $1 \neq \{1\}$, and $\{\{1\}\} \neq \{1\}$. That is, the *set* $\{a\}$ with sole element a is *not* the same thing as the item a itself.

An **ordered pair** (x, y) is just that, a list of two things in which there is a *first* thing, here x, and a *second* thing, here y. Two ordered pairs (x, y) and (x', y') are **equal** if and only if $x = x'$ and $y = y'$.

The **(cartesian) product** of two sets A, B is the set of **ordered pairs** (a, b) where $a \in A$ and $b \in B$. It is denoted $A \times B$. Thus, while $\{a, b\} = \{b, a\}$ might be thought of as an *unordered* pair, for *ordered* pairs $(a, b) \neq (b, a)$ unless by chance $a = b$.

In case $A = B$, the cartesian power $A \times B$ is often denoted A^2. More generally, for a fixed positive integer n, the n^{th} **cartesian power** A^n of a set is the set of ordered n-tuples $(a_1, a_2, \ldots, a_n)$ of elements a_i of A.

Some very important examples of cartesian powers are those of $\mathbf{R}$ or $\mathbf{Q}$ or $\mathbf{C}$, which arise in other contexts as well: for example, $\mathbf{R}^2$ is the collection of ordered

pairs of real numbers, which we use to describe points in the plane. And $\mathbf{R}^3$ is the collection of ordered triples of real numbers, which we use to describe points in three-space.

The **power set** of a set S is the *set of subsets* of S. This is sometimes denoted by $\mathcal{P}S$. Thus,

$$\mathcal{P}\phi = \{\phi\}$$

$$\mathcal{P}\{1,2\} = \{\phi, \{1\}, \{2\}, \{1,2\}\}$$

Intuitively, a **function** f from one set A to another set B is supposed to be a 'rule' which assigns to each element $a \in A$ an element $b = f(a) \in B$. This is written as

$$f : A \to B$$

although the latter notation gives no information about the nature of f in any detail.

More rigorously, but less intuitively, we can define a *function* by really telling its *graph*: the formal definition is that a function $f : A \to B$ is a *subset* of the product $A \times B$ with the property that for every $a \in A$ there is a unique $b \in B$ so that $(a, b) \in f$. Then we would write $f(a) = b$.

This formal definition is worth noting at least because it should make clear that there is absolutely no requirement that a function be described by any recognizable or simple 'formula'.

Map and **mapping** are common synonyms for *function*.

As a silly example of the formal definition of function, let $f : \{1,3\} \to \{2,6\}$ be the function 'multiply-by-two', so that $f(1) = 2$ and $f(3) = 6$. Then the 'official' definition would say that really f is the subset of the product set $\{1,3\} \times \{2,6\}$ consisting of the ordered pairs $(1,2), (3,6)$. That is, formally the function f is the *set*

$$f = \{(1,2), (3,6)\}$$

Of course, no one usually operates this way, but it is important to have a precise meaning underlying more intuitive usage.

A function $f : A \to B$ is **surjective** (or **onto**) if for every $b \in B$ there is $a \in A$ so that $f(a) = b$. A function $f : A \to B$ is **injective** (or **one-to-one**) if $f(a) = f(a')$ implies $a = a'$. That is, f is *injective* if for every $b \in B$ there is *at most one* $a \in A$ so that $f(a) = b$. A map is a **bijection** if it is both injective and surjective.

The number of elements in a set is its **cardinality**. Two sets are said to **have the same cardinality** if there is a *bijection* between them. Thus, this is a trick so that we don't have to actually *count* two sets to see whether they have the same number of elements. Rather, we can just pair them up by a *bijection* to achieve this purpose.

Since we *can* count the elements in a *finite* set in a traditional way, it is clear that *a finite set has no bijection to a proper subset of itself.* After all, a proper subset has *fewer elements.*

By contrast, for *infinite* sets it is easily possible that *proper* subsets have bijections to the whole set. For example, the set A of *all* natural numbers and the set E of *even* natural numbers have a bijection between them given by

$$n \to 2n$$

But certainly E is a *proper* subset of A! Even more striking examples can be arranged. In the end, we take as the *definition* that a set is **infinite** if it has a bijection to a proper subset of itself.

Let $f : A \to B$ be a function from a set A to a set B, and let $g : B \to C$ be a function from the set B to a set C. The **composite function** $g \circ f$ is defined to be

$$(g \circ f)(a) = g(f(a))$$

for $a \in A$.

The **identity function** on a non-empty set S is the function $f : S \to S$ so that $f(a) = a$ for all $a \in A$. Often the identity function on a set S is denoted by id_S.

Let $f : A \to B$ be a function from a set A to a set B. An **inverse function** $g : B \to A$ for f (if such g exists at all) is a function so that $(f \circ g)(b) = b$ for all $b \in B$, and also $(g \circ f)(a) = a$ for all $a \in A$. That is, the inverse function (if it exists) has the two properties

$$f \circ g = \mathrm{id}_B \qquad g \circ f = \mathrm{id}_A$$

An inverse function to f, if it exists at all, is usually denoted f^{-1}. (This is *not* at all the same as $1/f$!)

Proposition: A function $f : A \to B$ from a set A to a set B has an inverse if and only if f is a bijection. In that case, the inverse is unique (that is, there is only *one* inverse function).

Proof: Suppose that $f : A \to B$ is a bijection. We define a function $g : B \to A$ as follows. Given $b \in B$, let $a \in A$ be an element so that $f(a) = b$. Then define $g(b) = a$. Do this for each $b \in B$ to define g. Note that we use the *surjectivity* to know that there *exists* an a for each b and we use the *injectivity* to be sure of its *uniqueness*.

To check that $g \circ f = \mathrm{id}_A$, compute: first, for any $a \in A$, $f(a) \in B$. Then $g(f(a))$ is, by definition, an element $a' \in A$ so that $f(a') = f(a)$. Since f is injective, it must be that $a' = a$. To check that $f \circ g = \mathrm{id}_B$, take $b \in B$ and compute: by definition of g, $g(b)$ is an element of A so that $f(g(b)) = b$. But that is (after all) just what we want.

On the other hand, suppose that for $f : A \to B$ there is $g : B \to A$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$, and show that f is bijective. Indeed, if $f(a_1) = f(a_2)$, then apply g to both sides of this equality to obtain

$$a_1 = \mathrm{id}_A(a_1) = g(f(a_1)) = g(f(a_2)) = a_2$$

This proves injectivity of f. For surjectivity, given $b \in B$,

$$f(g(b)) = \mathrm{id}_B(b) = b$$

This completes the proof that if f has an inverse then it is a bijection. ///

1.2 Counting

Here we go through various standard elementary-but-important examples of **counting** as preparation for finite probability computations. Of course, by 'counting' we mean *structured* counting.

Example: Suppose we have n different things, for example the integers from 1 to n inclusive. The question is *how many different* **orderings** *or* **ordered listings**

$$i_1, i_2, i_3, \ldots, i_{n-1}, i_n$$

of these numbers are there? Rather than just tell the formula, let's quickly derive it. The answer is obtained by noting that there are n choices for the first thing i_1, then $n-1$ remaining choices for the second thing i_2 (since we can't reuse whatever i_1 was), $n-2$ remaining choices for i_3 (since we can't reuse i_1 nor i_2, whatever they were!), and so on down to 2 remaining choices for i_{n-1} and then just one choice for i_n. *Thus, there are*

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$$

possible **orderings** *of* n *distinct things.* This kind of product arises often, and there is a notation and name for it: n-**factorial**, denoted $n!$, is the product

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$$

It is an important and useful convention that

$$0! = 1$$

The factorial $n!$ is defined only for non-negative integers.

Example: How many **ordered** k**-tuples** of elements can be chosen (allowing repetition) from a set of n things? There are n possibilities for the first choice. For each choice of the first there are n choices for the second. For each choice of the first and second there are n for the third, and so on down to n choices for the k^{th} for each choice of the first through $(k-1)^{\text{th}}$. That is, altogether there are

$$\underbrace{n \times n \times \ldots \times n}_{k} = n^k$$

ordered k-tuples that can be chosen from a set with n elements.

Example: How many **ordered** k**-tuples** of **distinct** elements can be chosen from a set of n things? (In a mathematical context *distinct* means *all different from each other.)* There are n possibilities for the first choice. For each choice of the first