应用数学丛书

编码理论

肖国镇　卿斯汉　编著

.4

国防工业出版社

（京）新登字106号

内 容 简 介

本书较全面、系统地介绍了编码理论的有关知识和数学基础，取材广泛，并有重点地介绍了典型码类。通过大量的例题与应用背景的阐述，使读者对有关内容获得清晰的了解。本书还对近年来兴起的热门课题——代数几何编码理论进行了简要的介绍。

本书可供从事通信、计算机、应用数学等方面研究的广大工程技术人员参考，亦可作为大专院校有关专业师生的教学参考书。

应用数学丛书

编码理论

肖国镇 梁传甲 王育民 编著

国防工业出版社出版发行

（北京市海淀区紫竹院南路23号）

（邮政编码 100044）

新华书店经售

北京昌平长城印刷厂印装

850×1168毫米 32开本 印张12 1/8 316千字

1993年10月第一版 1993年10月第一次印刷 印数：0001—2 000册

ISBN 7-118-01082-0/O·87 定价：11.10元

应用数学丛书目录

*1.	Z变换与拉普拉斯变换	关肇直	王恩平	编著
*2.	常微分方程及其应用	秦化淑	林正国	编著
*3.	实变函数论基础		胡钦训	编著
*4.	正交函数及其应用		柳重堪	编著
*5.	沃尔什函数与沃尔什变换	关肇直	陈文德	编著
*6.	圆柱函数		刘　颖	编著
*7.	集合论		程极泰	编著
*8.	图论		王朝瑞	编著
*9.	概率论		狄昂照	编著
*10.	矩阵理论	王耕禄	史荣昌	编著
*11.	复变函数论		杨维奇	编著
*12.	逼近论 徐利治	周蕴时	孙玉柏	编著
*13.	矢量与张量分析 冯潮清	赵愉深	何浩法	编著
*14.	模糊数学	李洪兴	汪培庄	编著
*15.	编码理论	肖国镇	卿斯汉	编著
*16.	应用泛函分析		柳重堪	编著
17.	偏微分方程		丁夏畦	编著
18.	球函数及其应用		楼仁海	编著
*19.	椭圆函数及其应用		高本庆	编著
*20.	应用离散数学		陈文德	编著
*21.	拓扑理论及其应用	王则柯	凌志英	编著
*22.	网络理论		张正宾	编著

* 为已经出版或即将出版的书目。

*23.	广义函数及其解析表示	李邦河	李雅卿	编著
*24.	群论		刘木兰	编著
*25.	数理逻辑		沈百英	编著
*26.	线性系统与多变量控制		叶庆凯	编著
27.	最优化计算方法 马仲蕃	应玫茜	陈光亚	编著
28.	应用数理统计	李国英	吴启光	编著
*29.	多项式与多项式矩阵	王恩平	王朝珠	编著
30.	索伯列夫空间		丁夏畦	编著
31.	旋转群与四元素方法		毕大川	编著
*32.	信息论与最优编码		章照止	编著
33.	场的数学理论及物理应用		杜瑰	编著
34.	系统的动态辨识		张永光	编著
*35.	非线性系统分析与应用		司徒荣	编著
36.	数学物理数值方法	应隆安 滕震环	韩厚德 黄禄平	编著
*37.	误差理论与数据处理		贾沛璋	编著
38.	可计算性与计算复杂性		李未	编著
*39.	随机过程理论及应用		熊大国	编著
*40.	估计理论与随机控制		卢伯英	编著
*41.	应用组合数学		刘振宏	编著
*42.	渐进分析方法及应用	徐利治	陈文忠	编著
43.	有限元方法		应隆安	编著
44.	经济数学	苑凤歧	林寅	编著
*45.	预测的数学方法		张有为	编著
46.	粘性流体理论	吴望一	韩厚德	编著
47.	塑性理论		黄筑平	编著
48.	变分法及其应用	叶庆凯	郑应平	编著

出 版 说 明

近二十年来电子工程、控制工程、系统工程及其它领域都获得巨大发展。众所周知，这些科学技术研究的发展与现代逐渐形成的应用数学学科紧密相联，相辅相成。尤其是近年发展起来的边缘学科，更与数学紧密结合。但一般数学专著比较偏重于论证严谨，全面系统，篇幅较大，理论较深。广大科技工作者学习此类著作往往需时较多，与工作结合不紧，收效不大。本丛书将为目前在电子工程、控制工程、系统工程等领域工作的同志在数学基础的提高上，提供适合其工作特点的数学参考书。

本丛书是一种介于现代应用数学专著与工程专业理论书籍之间的桥梁参考著作，更着重于科技工作中应用较多的数学概念、分析和解题的基本技巧，也包括一部分适合于实际工作者为学习更高深的现代应用数学专著所需之基础知识。

本丛书选材包括三个方面：基础数学；应用数学有关领域的基础介绍；应用于科技中的典型基础专业理论。出版采用分册形式。各册内容独立，自成系统，但仍有少量交叉，分期分批出版。

丛书可供大专院校有关专业研究生、教师、从事科研生产的工程师参考。

此为试读，需要完整PDF请访问：www.ertongbook.com

前　言

编码理论自 40 年代末由仙农 (Shannon)、汉明 (Hamming) 等人创立以来，已经有 40 余年的历史。这期间，由于工程技术的实际需要，编码理论获得了不断的发展。特别是近年来，它在卫星通信、计算机技术、保密技术以及磁盘与光盘技术方面具有许多重要应用，愈来愈受到人们的重视。

在我国，编码理论的研究大约始于 50 年代末。中国电子学会信息论委员会的领导人蔡长年教授、周炯槃教授、胡征教授及陈太一教授等，对于促进这一学科的发展，把这一学科应用于通信工程及计算机技术领域起了重要作用。对于这一学科理论基础研究及其深入发展，则应提到万哲先教授、曾肯成教授等一批数学家的参与及贡献。在今天，我国在这一学科领域中已经拥有一批中青年科学工作者，他们在老科学家的带领下从事卓有成效的研究工作，并且在国际上具有一定的水平。

本书的初稿是作者之一于 70 年代初为在西安电子科技大学由梁传甲、王育民及王新梅等教授共同参加的编码讨论班所编写的讲义。这本讲义还曾作为作者之一在西安电子科技大学及在有关的研究所举办的编码讲习班上使用过，也曾多次作为研究生的参考教材。现在奉献给读者的这本书是由另一作者在原来讲义的基础上经过精心加工、修改及补充而成的。由于本书是应用数学丛书的一本，因此在取材上力求选取编码理论中的基础概念及基本理论，未涉及有关的工程技术。同时，由于篇幅所限，对于近年来发展起来的多用户编码理论也未写进本书，希望以后有机会加以补救。香港大学肖文强教授在我国讲学期间谈到的关于代数编码的经验，在本书中也有所体现。

对于前面提到的学者，我们表示衷心的感谢。本书作者之一

还对中国科学院计算中心的有关领导与教授在本书写作中所给予的支持和鼓励表示感谢。还应特别感谢国防工业出版社汪栋臣，侯迁，李端，王华诸同志的督促与帮助。

由于作者水平所限，书中谬误与不妥之处一定不少，恳切希望读者的批评与指正。

作　者

目　录

第一章 引 论

§1.1 历史的与现实的背景

编码理论导源于现代通信技术与电子计算机技术中差错控制研究的实际需要。美国数学家仙农（C.E.Shannon）在1948年发表的著名论文“通信的数学理论”，开创了一门在现代科学技术中具有重大意义的崭新的学科，即信息论。编码理论是信息论的一个专门分支。

汉明（R.W.Hamming）在1950年发表的论文“检错码与纠错码”是开拓编码理论研究的第一篇论文。这篇论文主要考虑在大型计算机中如何纠正所出现的单个错误。从这种能够纠正单个错误的汉明码过渡到能够纠正多个错误的所谓BCH码，整整经历了10年的时间。因此，可以说60年代是代数编码理论发展的鼎盛时期。70年代出现了高帕码（Goppa *Codes*），从而又把编码理论推向了一个新的高峰。到了80年代，茨伐斯曼(Tsfasman)等人运用代数几何的方法推广了高帕码的思想，指出存在$GF(q)$（$q=p^r$）上的一列码，它超过所谓基尔伯特-瓦尔莎莫夫限(Gilbert-Varshamov *bound*)。这一令人吃惊的结果给编码理论的进一步发展带来了新的希望。除此之外，基于组合理论及有限几何所建立的各种码类，以及在工程技术中具有重大实用价值的卷积码类的研究都大大地丰富了编码理论的研究。

如所周知，一个通信系统可以概括为图1-1所示的模型。来自信源的消息经过信源编码器被变换成能够表达这些消息的符号（为了提高有效性，在对消息进行编码时应当尽量减少多余度）。再经过信道编码器对信源编码器的输出符号进行变换，使变换后的符号具有抗击信道中噪声干扰的能力。最后，经过调制器将信

此为试读，需要完整PDF请访问：www.ertongbook.com

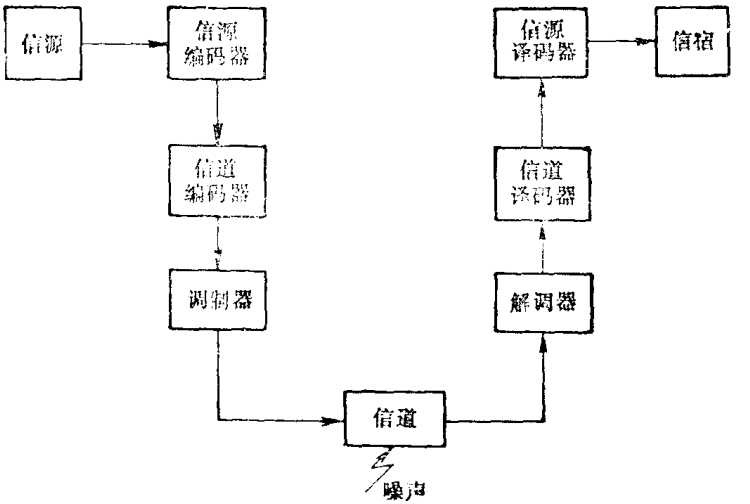

图 1-1

道编码器的输出调制成在带宽、功率及波段等适合于信道传输要求的信号。在收端，经过解调器、信道译码器及信源译码器来恢复发端所发送的消息。本书所要涉及的是信道编码理论，或者说是纠错编码理论。这一理论主要研究具有抗干扰能力的码类的构造及相应的编码及译码方法。

为了说明编码理论在当代空间技术中的应用，最好是看一看美国国家航空和航天管理局（NASA）所发射的探测外行星的“旅行者1号”及“旅行者2号”飞船在1979年及1986年发回地球的彩色照片。在这里即采用了编码技术把许多格点的光暗程度以二元数字传送回地球，经过解调、信道译码及信源译码还原后，便可看到有关木星及天王星的清晰的彩色图象。

从以上的简短回顾，我们可以看出编码理论是当代高科技与基础理论和谐统一、相互促进的一个典范。

§1.2 编码理论的基本思想与仙农信道编码定理

大家知道，在数字电路中，总假定来自信源的信息是二进制数字序列，即 0 —1 序列。对于这种二进制序列中的符号 0 与 1 可定义加法与乘法如下：

$$\left.\begin{array}{l} 0\oplus 0=0,\ 0\oplus 1=1,\ 1\oplus 0=1,\ 1\oplus 1=0 \\ 0\cdot 0=0,\ 0\cdot 1=0,\ 1\cdot 0=0,\ 1\cdot 1=1 \end{array}\right\} \tag{1-1}$$

如此定义的加法与乘法通常称为**模 2 加法**与**模 2 乘法**，其相应的逆运算——减法和除法，可定义如下：

$$\left.\begin{array}{l} 0\ominus 0=0,\ 1\ominus 1=0,\ 1\ominus 0=1,\ 0\ominus 1=1 \\ \dfrac{0}{1}=0,\qquad \dfrac{1}{1}=1 \end{array}\right\} \tag{1-2}$$

由式（1-2）可知，$-1=1$。因此，在模 2 算术中，加法与减法是一样的。并且，平常算术中的运算规律，诸如结合律，交换律及分配律等，在这里依然有效。

两个符号 0，1 所构成的集合，如果与它所定义的算术运算式（1-1）与式（1-2）联系起来，便称为**二元域**（或**二进制域**），记为 $GF(2)$。

对于来自信源的信息序列（即二进制数字序列），首先将其分成消息组，每个消息组由 k 位接续的信息数字组成，从而总共有 2^k 种不同的消息。其次，编码器按照一定的规则把每个消息变换成较长的 n 位（$n>k$）二进制数字组，称其为**码字**。由这 2^k 个消息所获得的 2^k 个码字的全体，便称为**码组长为 n，信息位为 k 的分组码**。每个消息所增加的 $n-k$ 个数字称为**多余数字**，它们不含有任何新的消息，其作用仅在于使码字在有干扰的信道中传输时能够修正传输中产生的错误。称比值

$$R=\frac{k}{n}$$

为该分组码的**信息率**。或者，更为一般地，假定有 s 个不同的消息，每个消息由 k 位二进数字组成，将每个消息增加 $n-k$ 个多余

数字，从而得到含 s 个码字的分组码 C。称比值

$$R=\frac{\log_2|C|}{n}$$

为该分组码 C 的**信息率**，其中 $|C|$ 代表该分组码中码字的总数，即为 s。当 $s=2^k$ 时，便有 $R=k/n$。

为什么将消息数字适当增加些多余数字，就会提高消息在传输过程中的抗干扰能力呢？这是常识范围内所能想像到的。比如你见到一个英语字母组合 *Infomation*，马上会想到它是英语单词 *Information* 之误。但是，如果你见到 *tull* 这个字母组合，便无法猜出它是 *tall* 之误，还是 *tell* 或 *till* 之误。道理很简单，长的字即使其中有一两个字母错了，它还是像原来那个字多于像其它的字。短的字母则不然，印错的字可能与许多另外的字都相象。因此，编码的基本思想就是将原来要传送的消息数字适当加长，以便使所有传送的消息在传输过程中所产生的错误容易辨认及纠正。

把消息数字适当加长使其变换为码字的过程，称为**编码**。把接收的数字组按照一定的准则恢复成码字的过程，称为**译码**。

一个合理的译码准则是**所谓极大后验概率译码准则**。当接收端收到长为 n 的数字组 $\boldsymbol{r}$（称为接收字）时，对于所有可能的码字 $\boldsymbol{c}_i$，计算全部条件概率 $p(\boldsymbol{c}_i|\boldsymbol{r})$，$i=1,2,\cdots,s$，若某一码字 $\boldsymbol{c}_i$ 使 $p(\boldsymbol{c}_i|\boldsymbol{r})$ 最大，便把 $\boldsymbol{r}$ 译作 $\boldsymbol{c}_i$。这就是所谓最大后验概率译码准则。但条件概率 $p(\boldsymbol{c}_i|\boldsymbol{r})$ 实际上不易计算。为此，依概率计算法则可将 $p(\boldsymbol{c}_i|\boldsymbol{r})$ 写成

$$p(\boldsymbol{c}_i|\boldsymbol{r})=\frac{p(\boldsymbol{c}_i)\,p(\boldsymbol{r}|\boldsymbol{c}_i)}{p(\boldsymbol{r})}$$

如果进一步假设全部码字在信道上传送是等概率的：

$$p(\boldsymbol{c}_i)=\frac{1}{s},\quad i=1,2,\cdots,s$$

便有

$$p(\boldsymbol{c}_i|\boldsymbol{r})=\frac{p(\boldsymbol{r}|\boldsymbol{c}_i)}{sp(\boldsymbol{r})}$$

由此可见，当且仅当 $p(\boldsymbol{r}|\boldsymbol{c}_i)$ 最大时，$p(\boldsymbol{c}_i|\boldsymbol{r})$ 为最大。按

$p(\boldsymbol{r}|\boldsymbol{c}_i)$最大的译码方案称为**极大似然译码准则**。从信息论的观点看，当一个信道给定时，意味着转移概率$p(\boldsymbol{r}|\boldsymbol{c}_i)$是完全确定的。不过，按极大似然译码准则进行译码，这种算法的复杂度仍然很大。有人估计，对于$n=10$，$R=1/2$的二进制码，按极大似然译码准则的计算量竟达至10^{15}次之多！

为进一步简化计算，考虑图 1-2 所示的二元对称信道。在这种信道中，发 0 收 0 或发 1 收 1（传输正确）的概率均为q，而发 0 收 1 或发 1 收 0（传输错误）的概率均为$p=1-q$，并且假设$p<\frac{1}{2}$。这种信道的转移概率矩阵是

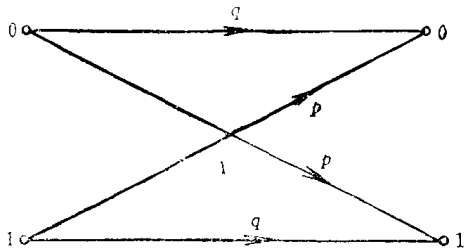

图 1-2

$$T=\begin{bmatrix} p(0|0) & p(1|0) \\ p(0|1) & p(1|1) \end{bmatrix}=\begin{bmatrix} q & p \\ p & q \end{bmatrix}$$

称这种信道为**二元对称信道**。进一步还假定：在传输中每个符号传输正确与否与别的符号传输正确与否是两个独立的事件。

让我们再回到转移概率$p(\boldsymbol{r}|\boldsymbol{c}_i)$的计算上来。设$d_i$代表$\boldsymbol{r}$与$\boldsymbol{c}_i$对应位分量不相同的数目，称之为$\boldsymbol{r}$与$\boldsymbol{c}_i$之间的**汉明距离**。例如$\boldsymbol{r}=(01001)$与$\boldsymbol{c}_i=(10111)$之间的汉明距离为 4。于是，在上述假定下便有

$$p(\boldsymbol{r}|\boldsymbol{c}_i)=p^{d_i}q^{n-d_i}$$

根据微积分中极值理论不难看出，当且仅当d_i最小时，$p(\boldsymbol{r}|\boldsymbol{c}_i)$最大。由此，我们又得出一个译码方案：收到$\boldsymbol{r}$后，在全部码字$\boldsymbol{c}_i(i=1,2,\cdots,s)$中寻找与$\boldsymbol{r}$的汉明距离最近的$\boldsymbol{c}_i$判决为原来发送的码字（这就是我们前面提到过的印错的字更像谁的问题）。称这种译码方案为**最小距离译码准则**。

总之，在二元对称信道相应的假定下，极大后验概率译码准

则、极大似然译码准则与最小距离译码准则是一回事情。

紧接着的一个问题便是按照一定的译码方案来计算错误译码概率的问题。一个合理的译码方案均应是使错误译码概率为最小的方案。

在二元对称信道中，长为 n 的码字在传输中全然无误的概率应当是 q^n。如果给定的码能纠正至多 1 个错误，这种模式共有 $\binom{n}{1}=n$ 种，每一种产生的概率均为 pq^{n-1}。因而对于那种能纠正至多 1 个错误的信道及相应的码，接收端正确译码的概率应当是 q^n+npq^{n-1}。一般，对于那种能纠正至多 t 个错误的码，接收端正确译码的概率应当是

$$p_c=\sum_{i=0}^{t}\binom{n}{i}p^iq^{n-i}$$

而错误译码的概率应当是

$$p_e=1-p_c=\sum_{i=t+1}^{n}\binom{n}{i}p^iq^{n-i} \tag{1-3}$$

这一错误译码概率简称为**误码率**。

例如，考虑所谓**重复码**，把待发送的信息数字 0 与 1 分别重复$2N+1$次：

$$0\mapsto\underbrace{0\ 0\cdots0}_{2N+1};\quad 1\mapsto\underbrace{1\ 1\cdots1}_{2N+1}$$

译码按照所谓**大数逻辑准则**，即当接收到 $r=a_0a_1\cdots a_{2N}$ 时，若其中 1 的个数多于N个，则将r译作 $11\cdots1$，否则便将r 译作 $00\cdots0$。这种译码方案显然是最小距离译码方案。按这种方案错误译码的概率应当是（注意 $p<1/2$）

$$\begin{aligned}p_e(N)&=\sum_{k=0}^{N}\binom{2N+1}{k}q^kp^{2N+1-k}\leqslant(pq)^N\sum_{k=0}^{N}\binom{2N+1}{k}\\&=(pq)^N2^{2N}=(4pq)^N\end{aligned}$$

注意当 $p=1/2$ 时，$4pq=4p(1-p)$ 取得最大值 1，故当

$p<\frac{1}{2}$时，$4pq<1$。因而当$N\to\infty$时，$p_e(N)\to 0$。这表明，只要增大信息的重复次数便可使误码率变得任意小。但是，这样做便使该码的信息率$R=1/(2N+1)$变得任意小，在实践上是很不经济的。

令人振奋的是仙农在40年代末确立了下述的深刻定理。

仙农信道编码定理 每一个信道都有一个确定的信道容量C（它标志该信道最大传输信息的能力。例如对于二元对称信道，$C=1+p\log_2 p+q\log_2 q$）。对于任意的$\varepsilon>0$及给定的R，$0<R<C$，必存在一个分组长度n足够大而信息率为R的码，当采用最大似然译码准则时可使错误译码概率$p_e<\varepsilon$。

这一定理的证明，读者可在信息论的专著中找到。为了说明这一定理的结果是可信的，我们举一个例子说明在信息率R不变的情况下，可以通过增大码长n使误码率降低。假定$R=1/2$，$p=0.01$，$q=0.99$，考虑重复码：$0\mapsto 00$；$1\mapsto 11$。按照大数逻辑译码准则，当且仅当所收到的$r=a_0a_1$中两位均无错时，才能正确译码。因而正确译码的概率$p_c=q^2=0.9801$，错误译码的概率$p_e=1-p_c=0.0199$。假定我们等到传2个信息元之后才一并传4个元（后面2个是多余码元）。按下述规则编码：

$$00\longmapsto 0000,$$
$$01\longmapsto 0111,$$
$$10\longmapsto 1001,$$
$$11\longmapsto 1110。$$

亦即假定码字$c=a_0a_1a_2a_3$之后两位多余码元按下述规则添加：

$$\begin{aligned} a_2&=a_1 \\ a_3&=a_0+a_1 \end{aligned} \qquad (1\text{-}4)$$

由此不难看出，如果采用译码算法：当式(1-4)成立时，认为收到的$r=a_0a_1a_2a_3$无错；当式(1-4)不成立时，认为a_3无错而a_0，a_1，a_2中有一位错；则这种码可纠正前3位中不多于1个的错误。事实上，如果式(1-4)中$a_2\neq a_1$，$a_3=a_0+a_1$，则可断定a_2有错。同理，当$a_2=a_1$而$a_3\neq a_0+a_1$时，a_0有错。最后，当

此为试读，需要完整PDF请访问：www.ertongbook.com