《信息论基础》

第5章 纠错编码

吉小鹏

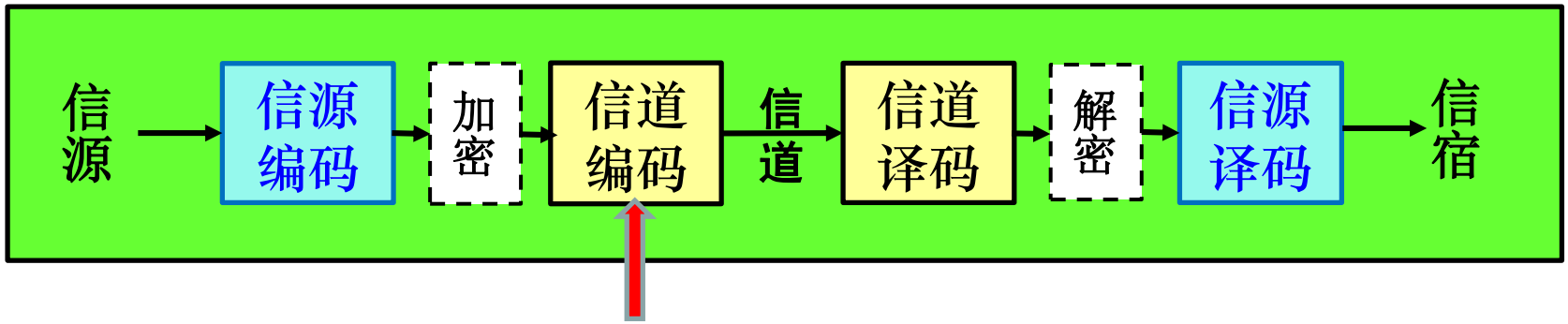
E-mail: 003163@nuist.edu.cn

南京信息工程大学 电子与信息工程学院 尚贤楼209




- 编码
  - 信源编码
    - 提高数字信号有效性
      - 将信源的模拟信号转变为数字信号
      - 降低数码率，压缩传输频带(数据压缩)
  - 信道编码
    - 提高数字通信可靠性
  - 数字信号在信道的传输过程中，由于实际信道的传输特性不理想以及存在加性噪声，在接收端往往会产生误码。

信道编码器在通信系统中的位置



- **作用** 提高信息传输时的抗干扰能力
- **目的** 增加信息传输的可靠性
- **手段** 增加信息冗余度
- **名称** **信道码、数据传输码、差错控制码**

提纲

5.1 纠错编码的基本概念

5.2 线性分组码

5.3 循环码

5.1 纠错编码的基本概念

差错率的定义

- 差错率是衡量传输质量的重要指标之一，它有几种不同的定义。
- 码元差错率/符号差错率
  - 指在传输的码元总数中发生差错的码元数所占的比例(平均值)，简称**误码率**。
  - 是指信号差错概率
- 比特差错率 /比特误码率：
  - 在传输的比特总数中发生差错的比特数所占比例。
  - 是指信息差错概率
- 对二进制传输系统，符号差错等效于比特差错；对多进制系统，一个符号差错对应多少比特差错却难以确定。

5.1 纠错编码的基本概念

差错率的定义

- 根据不同的应用场合对差错率有不同的要求:
  - 在电报传送时，允许的比特差错率约为:

    $10^{-4}\sim10^{-5}$;
  - 计算机数据传输，一般要求比特差错率小于:

    $10^{-8}\sim10^{-9}$;
  - 在遥控指令和武器系统的指令系统中，要求有更小的误比特率或码组差错率。

5.1 纠错编码的基本概念

差错图样

- 为定量地描述信号的差错，定义差错图样E
- E=C－R　(模M)
- 最常用的二进制码可当作特例来研究，其差错图样等于收码与发码的模2加，即
- E = C⊕R 或 C = R⊕E
- 设发送的码字C　1 1 1 1 1 1 1 1 1 1 1
  接收的码字R　1 0 0 1 0 0 1 1 1 1
  差错的图样E　0 1 1 0 1 1 0 0 0 0
- 差错图样中的“1”既是符号差错也是比特差错，差错的个数叫汉明距离。

0:传输中无错

1:传输中有错

5.1 纠错编码的基本概念

- **差错图样**
  - 随机差错：
    - 差错是相互独立的,不相关
    - 存在这种差错的信道是无记忆信道或随机信道
  - 突发差错：
    - 指成串出现的错误,错误与错误间有相关性,一个差错往往要影响到后面一串字
  - **E:** 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 1 1 1 0 0 0 0 0 0 0

    突发长度 = 4　　突发长度 = 6

5.1 纠错编码的基本概念

- 差错控制系统

1 前向纠错方式

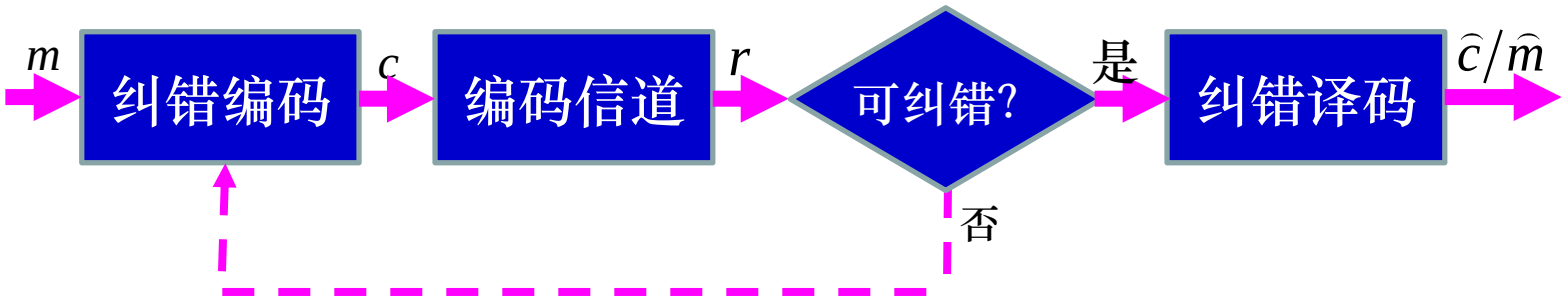
FEC方式差错控制系统模型

2 反馈重传方式

ARQ方式差错控制系统

- 差错控制系统

3 混合纠错方式



HEC方式差错控制系统

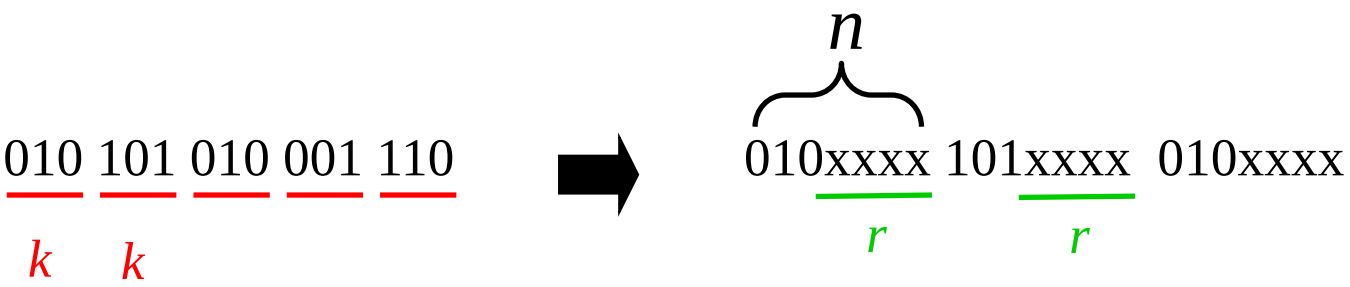
5.1 纠错编码的基本概念

- **纠错码的分类**
  - 从功能角度讲，差错码分为检错码和纠错码
    - **检错码**：用于发现差错
    - **纠错码**：能自动纠正差错
  - 纠错码与检错码在理论上没有本质区别，只是应用场合不同，而侧重的性能参数也不同。
  - 按照适用的差错类型，分成:
    - 纠随机差错码：用于随机差错信道，其纠错能力用码组内允许的独立差错的个数来衡量。
    - 纠突发差错码：针对突发差错而设计，其纠错能力主要用可纠突发差错的最大长度来衡量。

5.1 纠错编码的基本概念

纠错码的分类

- 按照对信息序列的处理方法，有分组码和卷积码
- 分组码：
  - 将*k*个信息码元分成一组，由这*k*个码元按照一定规则产生*r*个监督(校验)码元，组成长度n = k + r的码字，记为(n, k)码。



- 卷积码：
  - 先将信息序列分组(k_0)，不同的是编解码运算不仅与本组信息有关，而且还与前面若干组(L)有关，记为(n, k_0, L)码。

5.1 纠错编码的基本概念

纠错码的分类

- 按照码元与原始信息位的关系，分为
  - 线性码：所有码元(校验元)均是原始信息元的线性组合，编码器不带反馈回路。
  - 非线性码：码元(校验元)并不都是信息元的线性组合，可能还与前面已编的码元有关，编码器可能含反馈回路。
- 由于非线性码的分析比较困难，早期实用的纠错码多为线性码，但当今发现的很多好码恰恰是非线性码。

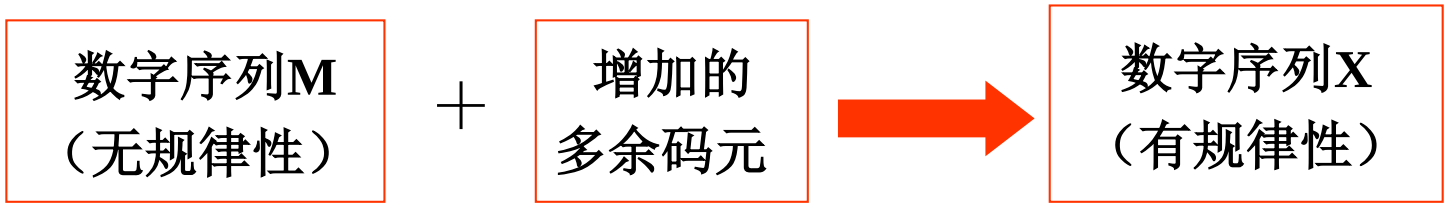
5.1 纠错编码的基本概念

纠错编译码原理

对于无噪无损信道只要对信源进行适当的编码，总能以信道容量无差错的传递信息。但是一般信道总会存在噪声和干扰，那么在有噪信道中进行无错传输可以达到的最大信息传输率是多少呢？

信道编码的对象：信源编码器输出的数字序列（信息序列） M。通常是0、1符号构成的序列。

信道编码：

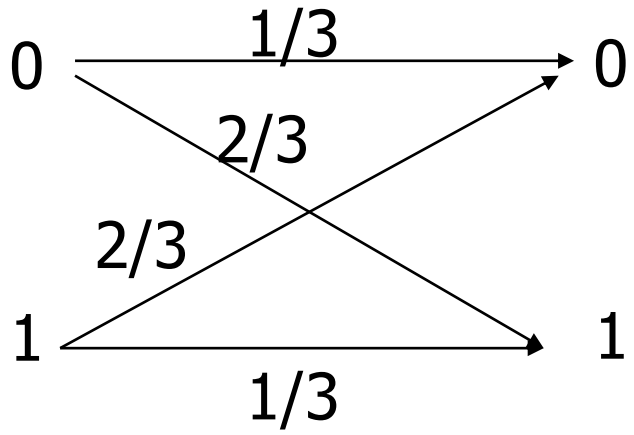


- **纠错编译码原理**

(1) 错误概率和译码规则

在有噪信道中传输消息是会发生错误的。错误概率与信道统计特性有关。信道的统计特性可由信道的转移矩阵来描述。

例：有一个BSC信道，如图所示。



译码规则：

① 若收到“0”译作“0”，收到“1”译作“1”，则平均错误概率为：

$$P_E = P(0)P_e^{(0)} + P(1)P_e^{(1)} = \frac{2}{3}$$

② 若收到“0”译作“1”，收到“1”译作“0”，则平均错误概率为1/3，可见错误概率与译码准则有关。

错误概率既与信道的统计特性有关，又与译码的规则有关。

5.1 纠错编码的基本概念

- 纠错编译码原理

（1）错误概率和译码规则

定义译码准则：

译码规则　设离散单符号信道的输入符号集为 $A=\{a_i\}$, $i=1,\ 2,\ \cdots,\ r$ 输出符号集 $B=\{b_i\}, j=1,\ 2,\ \cdots,\ s$ 。若对每一个输出符号 b_j 都有一个确定的函数 $F(b_j)$ 使 b_j 对应于惟一的一个输入符号 a_i ，则这样的函数称为译码规则，记为

$$F(b_j)=a_i \qquad \begin{matrix} i=1,\ 2,\ \cdots,\ r \\ j=1,\ 2,\ \cdots,\ s \end{matrix}$$

5.1 纠错编码的基本概念

- 纠错编译码原理

(1) 错误概率和译码规则

例

$$[P]=\begin{matrix}a_1\\a_2\\a_3\end{matrix}\begin{matrix}\begin{bmatrix}0.5 & 0.3 & 0.2\\0.2 & 0.3 & 0.5\\0.3 & 0.3 & 0.4\end{bmatrix}\end{matrix}\quad\begin{matrix}b_1 & b_2 & b_3\end{matrix}$$

可以设计译码准则：

$$A:\quad \begin{aligned}F(b_1)&=a_1\\F(b_2)&=a_2\\F(b_3)&=a_3\end{aligned}$$

$$B:\quad \begin{aligned}F(b_1)&=a_1\\F(b_2)&=a_3\\F(b_3)&=a_2\end{aligned}$$

译码规则的选择应该有一个依据，一个自然的依据就是使平均错误概率最小。

- 纠错编译码原理

(1) 错误概率和译码规则

有了译码规则以后，收到 b_j 的情况下，译码的条件正确概率为：

$$P(F(b_j)|b_j)=P(a_i|b_j)$$

而错误译码的概率为收到b_j后，推测发出除了a_i之外其它符号的概率：

$$P(e|b_j)=1-P(a_i|b_j)$$

可以得到平均错误译码概率为：

$$P_E=\sum_{j=1}^{m}p(b_j)P(e|b_j)=\sum_{j=1}^{s}p(b_j)(1-P(a_i|b_j))$$

它表示经过译码后平均接收到一个符号所产生错误的大小，也称平均错误概率。

如何设计译码规则 $F(b_j)=a_i$，使P_E 最小？

- 纠错编译码原理

（1）错误概率和译码规则

由前边的讨论可以看出，为使 $P(e|b_j)$ 最小，就应选择 $P(F(b_j)|b_j)$ 为最大，即选择译码函数 $F(b_j)=a^*$ 并使之满足条件：

$$P(a^*|b_j) \geq P(a_i|b_j) \qquad a_i \neq a^*$$

也就是说，收到一个符号以后译成具有最大后验概率的那个输入符号。这种译码准则称为“最大后验概率准则”或“最小错误概率准则”。

根据贝叶斯定律，上式也可以写成

$$\frac{P(b_j|a^*)P(a^*)}{P(b_j)} \geq \frac{P(b_j|a_i)P(a_i)}{P(b_j)}$$

即： $P(b_j|a^*)P(a^*) \geq P(b_j|a_i)P(a_i)$

- 纠错编译码原理

(1) 错误概率和译码规则

$$P(b_j \mid a^*)P(a^*) \geq P(b_j \mid a_i)P(a_i)$$

当信源等概分布时，上式为：

$$P(b_j \mid a^*) \geq P(b_j \mid a_i)$$

这称为最大似然译码准则，方法是收到一个 b_j 后，在信道矩阵的第 j 列，选择最大的值所对应的输入符号作为译码输出。

可进一步写出平均错误概率：

$$\begin{aligned} P_E &= \sum_Y P(b_j)P(e \mid b_j) = \sum_Y \{1 - P[F(b_j) \mid b_j]\}P(b_j) \\ &= 1 - \sum_Y P[F(b_j)b_j] = \sum_{X,Y} p(a_i b_j) - \sum_Y P[F(b_j)b_j] \\ &= \sum_{X,Y} p(a_i b_j) - \sum_Y P[a^* b_j] = \sum_{Y, X-a^*} P(a_i b_j) \end{aligned}$$

5.1 纠错编码的基本概念

- 纠错编译码原理

(1) 错误概率和译码规则

$$P_E = \sum_{Y,X-a^*} P(a_i b_j) \quad \text{也可写成：} \quad P_E = \sum_{Y,X-a^*} P(b_j \mid a_i) P(a_i)$$

上式也可写成对行求和：

$$P_E = \sum_X P(a_i) \sum_Y \left\{ P(b_j \mid a_i) \qquad F(b_j) \neq a^* \right\}$$

$$= \sum_X P(a_i) P_e^{(i)}$$

如果先验概率相等，则：$P_E = \frac{1}{r} \sum_X P_e^{(i)}$

5.1 纠错编码的基本概念

- **纠错编译码原理**

（1）错误概率和译码规则

$$P=\begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.2 & 0.3 & 0.5\\ 0.3 & 0.3 & 0.4\end{bmatrix}$$

例：信道矩阵*P*，输入等概率分布。

根据最大似然准则可选择译码函数为B：

$$\begin{cases}F(b_1)=a_1\\ F(b_2)=a_3\\ F(b_3)=a_2\end{cases}$$

$$P_E=\frac{1}{3}\sum_{Y,X-a^*}P(b_j\mid a_i)=\frac{1}{3}[(0.2+0.3)+(0.3+0.3)+(0.2+0.4)]=0.567$$

若采用前边讲到的译码函数A，则平均错误率为：

$$P_E^{'}=\frac{1}{3}\sum_{Y,X-a^*}P(b_j\mid a_i)=\frac{1}{3}[(0.3+0.2)+(0.3+0.3)+(0.2+0.5)]=0.6$$

若输入不等概分布，其概率分布为：$P(a_1)=\frac{1}{4},P(a_2)=\frac{1}{4},P(a_3)=\frac{1}{2}$

$$P_E^{''}=\sum_{Y,X-a^*}P(a_i)\cdot P(b_j\mid a_i)=\frac{1}{4}(0.3+0.2)+\frac{1}{4}(0.3+0.3)+\frac{1}{2}(0.2+0.5)=0.6$$

5.1 纠错编码的基本概念

- **纠错编译码原理**

（1）错误概率和译码规则

$$P=\begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.2 & 0.3 & 0.5\\ 0.3 & 0.3 & 0.4\end{bmatrix}$$

若输入不等概分布，其概率分布为：$P(a_1)=\frac{1}{4},P(a_2)=\frac{1}{4},P(a_3)=\frac{1}{2}$

若采用最小错误概率译码准则，则联合矩阵为：

$$\left[P(a_ib_j)\right]=\begin{bmatrix}0.125 & 0.075 & 0.05\\ 0.05 & 0.075 & 0.125\\ 0.15 & 0.15 & 0.2\end{bmatrix}$$

所得译码函数为C：$\begin{cases}F(b_1)=a_3\\ F(b_2)=a_3\\ F(b_3)=a_3\end{cases}$

平均错误率为：$P_E^{'''}=\sum_{Y}\sum_{X-a^*}P(a_i)P(b_j/a_i)=(0.125+0.05)+(0.075+0.075)+(0.05+0.125)=0.5$

- 纠错编译码原理

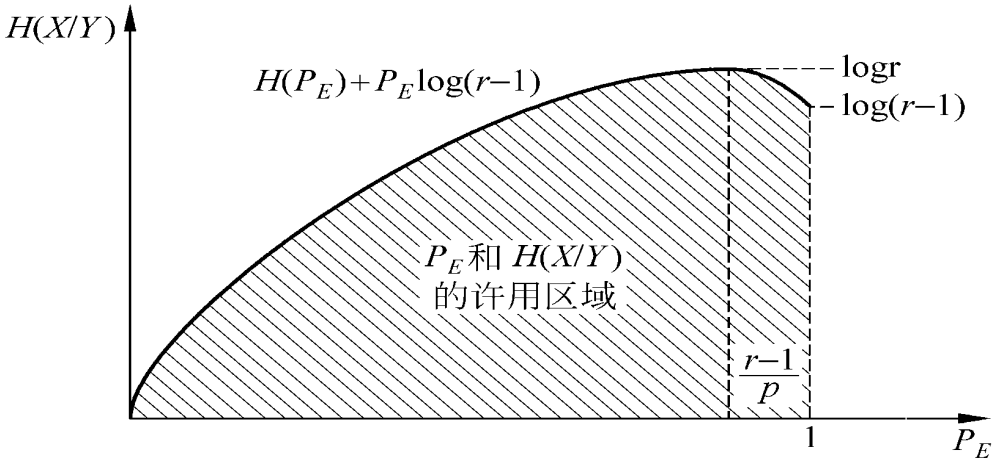
(1) 错误概率和译码规则

平均错误概率 P_E 与译码规则有关，而译码规则又由信道特性决定。于是，平均错误概率 P_E 与信道疑义度 $H(X/Y)$ 是有一定关系的，满足

费诺不等式

$$H(X|Y) \leqslant H(P_E) + P_E \log(r-1)$$

接收到 Y 后关于 X 的平均不确定性分为两部分：①是否产生 P_E 错误的不确定性 $H(P_E)$；②输入符号发送而造成的最大不确定性 $P_E \log(r-1)$。



费诺不等式的曲线图

- 纠错编译码原理

（2）错误概率和编码方法

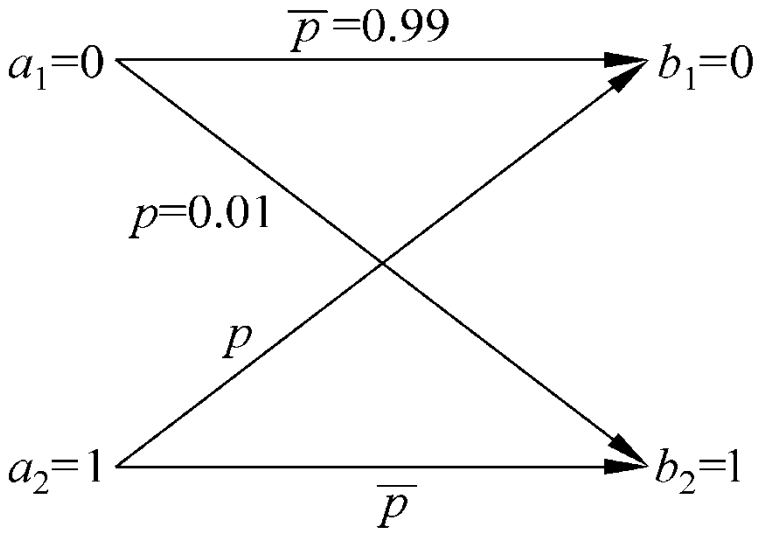
一般信道传输时都会产生错误，而选择译码准则并不会消除错误，那么如何减少错误概率呢？下边讨论通过优选信道编码方法来降低错误概率。

例：对于如下二元对称信道

$$[P]=\begin{bmatrix}0.99 & 0.01\\ 0.01 & 0.99\end{bmatrix}$$

$$P_E=0.01=10^{-2}$$

（输入等概率）



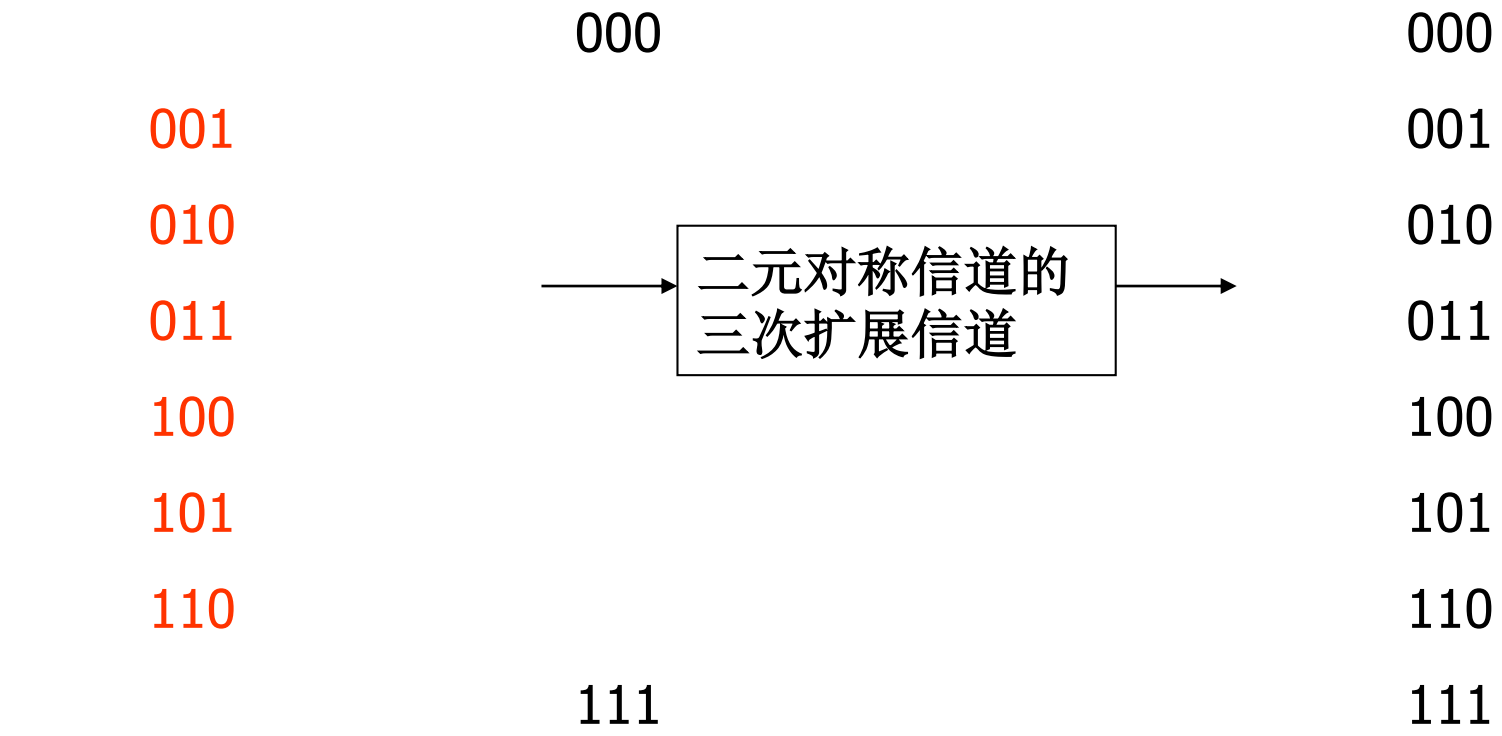
二元对称信道

5.1 纠错编码的基本概念

- 纠错编译码原理

（2）错误概率和编码方法

如何提高信道传输的正确率呢？可以尝试用下面的方法　简单的重复编码



- 纠错编译码原理

（2）错误概率和编码方法

信道矩阵 $$P=\begin{bmatrix} \bar{p}^3 & \bar{p}^2 p & \bar{p}^2 p & \bar{p}p^2 & \bar{p}^2 p & \bar{p}p^2 & \bar{p}p^2 & p^3 \\ p^3 & \bar{p}p^2 & \bar{p}p^2 & \bar{p}^2 p & \bar{p}p^2 & \bar{p}^2 p & \bar{p}^2 p & \bar{p}^3 \end{bmatrix}$$

根据最大似然译码准则，可得译码函数为：

F(000)=000　F(001)=000　F(010)=000　F(011)=111

F(100)=000　F(101)=111　F(110)=111　F(111)=111

$$P_E=\sum_{Y^3C-a^*} P(\alpha_i)P(\beta_j/\alpha_i)=3*10^{-4}$$

此时，译码可以采用“选择多数译码”，即根据接收序列中0多还是1多，0多就判作0，1多就判作1。错误概率降低了两个数量级，这种编码可以纠正码字中的一位码元出错。若重复多次可进一步降低错误率。

- 纠错编译码原理

（2）错误概率和编码方法

$$P_E = \text{错3个码元的概率} + \text{错2个码元的概率}$$

$$= C_3^3 p^3 + C_3^2 \bar{p} p^2 = p^3 + 3\bar{p}p^2 \simeq 3*10^{-4}$$

$$n=5 \quad P_E \approx 10^{-5} \qquad n=9 \quad P_E \approx 10^{-8}$$

$$n=7 \quad P_E \approx 4\times 10^{-7} \qquad n=11 \quad P_E \approx 5\times 10^{-10}$$

当n很大时，使P_E很小是可能的。但是又出现了一个新的问题，n很大时，信息传输率会降低很多。

在上例中：M=2 $R = \dfrac{\log M}{n}$

当n=1时 R=1；当n=3时 R=1/3；当n=5时 R=1/5.....

结论：利用简单重复编码来减少平均错误概率是以降低信息传输率R作为代价的，即“牺牲有效性，换取可靠性”。

5.1 纠错编码的基本概念

- 信道编码定理

若一离散平稳无记忆信道，其容量为C，输入序列长度为L，只要待传送的信息率$R<C$，总能找到一种编码。当L足够长时，译码差错概率

$$p_e < \varepsilon$$

ε 为任意正数。

反之，当$R>C$，任何编码的p_e必大于0，或无法任意小。

5.1 纠错编码的基本概念

- **码距与检错、纠错能力**

设 $\alpha_i=(a_{i_1}a_{i_2}\cdots a_{i_n}),\beta_i=(b_{j_1}b_{j_2}\cdots b_{j_n})$ 为两个长度为 n 的二元序列（码字），则 α_i 和 β_i 之间的汉明距离定义为

$$D(a_i,\beta_j)=\sum_{k=1}^{n}a_{i_k}\oplus b_{j_k}$$

如：$\alpha_i=101111$　　$\beta_j=111100$　　则

$$D(\alpha_i,\beta_j)=3$$

在某一码簿中，任意两个码字的汉明距离的最小值称为该码C的最小距离。

$$d_{\min}=\min\{D(C_i,C_j)\}\qquad C_i\neq C_j$$

我们来讨论前边的4种码的距离：

5.1 纠错编码的基本概念

- 码距与检错、纠错能力

	码A	码B	码C	码D
码字	000 111	000 011 101 110	000 001 010 100	000 001 010 011 100 101 110 111
消息数M	2	4	4	8
最小距离 dmin	3	2	1	1
信息传输率 R	1/3	2/3	2/3	1
错误概率	$3*10^{-4}$	$2*10^{-2}$	$2.28*10^{-2}$	$3*10^{-2}$

很明显，$d_{\min}$越大，P_E越小，在M相同的情况下也是一样。

5.1 纠错编码的基本概念

- **码距与检错、纠错能力**

定理：若纠错码的最小距离为d_{min}，

(1)可以检测出任意小于等于$l = d_{min} - 1$个差错

(2)可以纠正任意小于等于 $t = \left[\frac{d_{min} - 1}{2}\right]$ 个差错

(3)可以检测出任意小于等于l同时纠正小于等于t个差错，

– 其中l、t满足： $l + t \le d\min - 1$

$t < l$

5.2 线性分组码

- 线性分组码的基本概念

信道编译码方法的最初范例

基本思路： 将码字分成两段

信息位： 含有信源信息的比特位(红色部分)，长度用k表示。

校验位： 按一定规则添加的码位。用于监督码组内部码元的关联关系，可发现或纠正传输错误(蓝色部分)，亦称监督位。

码　　字： 信息位和校验位组成的相对独立码组(红色和蓝色整体)，长度用n表示。

5.2 线性分组码

- 线性分组码的基本概念

汉明重量： 码字中非零码元的个数，亦称码重。

1	0	1	1	0	1	0	1

汉明重量(码重)= 5

汉明距离： 两个码字相应码元取不同数值的码元数，也称码距。

1	0	1	1	0	1	0	1

1	1	0	1	0	0	1	1

汉明距离(码距)= 4

5.2 线性分组码

- 线性分组码的基本概念

$$d(c,c') = \sum_{i=1}^{n} c_i \oplus c'_i, \quad c_i \neq c'_i \tag{5.2.1}$$

码的最小距离： 任意两码字之间的最小汉明距离，简称最小码距。

$$d_{\min} = \min_{c \neq c'} d(c,c') \tag{5.2.2}$$

例

0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0

1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

(4,3)偶校验码的最小码距= 2

5.2 线性分组码

- 线性分组码的基本概念

分组码： 将信息序列以 k 为长度进行等长划分，然后按部门规则添加一定数目冗余位的编码方法，冗余位亦称校验位。

线性分组码： 信息位和校验位之间满足线性关系的分组码。

总结线性分组码的基本参数如下：

码　　长：n

信息位长：k

码 字 数：M

监督位长：r

最小码距：d_{min}

编码效率：$R = k/n$

5.2 线性分组码

- 线性分组码的编码

设$\boldsymbol{m}$为k维信息矢量，$\boldsymbol{c}$为n维码矢量，一般线性分组码可以(n,k)两个参数表示。线性分组码的编码规则很简单：

$$\boldsymbol{c}=\boldsymbol{m}\boldsymbol{G} \tag{5.2.3}$$

式中$\boldsymbol{G}$是k行n列$(n \geq k)$的秩为k的矩阵，称为生成矩阵。

$$\boldsymbol{G}=\begin{bmatrix} g_{0,0} & \cdots & g_{0,n-1} \\ \vdots & \ddots & \vdots \\ g_{k-1,0} & \cdots & g_{k-1,n-1} \end{bmatrix} \tag{5.2.4}$$

对于二元编码，$\boldsymbol{c}$和$\boldsymbol{m}$都是二元向量，$\boldsymbol{G}$为GF(2)上的 $k \times n$矩阵，$g_{i,j} \in \{0,1\}$，向量与矩阵之间，矩阵与矩阵之间均为模2运算。

5.2 线性分组码

- 线性分组码的编码

例5.2.1 (4,3)偶校验码是一个(4,3)线性分组码，其生成矩阵

$$\boldsymbol{G}=\begin{bmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\end{bmatrix}$$，当$\boldsymbol{m}=(m_0,m_1,m_2)=(101)$时，求其码字。

解：按照偶校验编码规则，并将$\boldsymbol{m}$代入(5.2.3)得

$$\boldsymbol{c}=\left(\boldsymbol{c}_0\boldsymbol{c}_1\boldsymbol{c}_2\boldsymbol{c}_3\right)=(101)\begin{bmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\end{bmatrix}=(1010) \qquad (5.2.5)$$

例5.2.2 3重复码是(3,1)线性分组码，生成矩阵为[111]，码字只有

0	0	0

1	1	1

5.2 线性分组码

- 线性分组码的编码

当G给定时，线性分组码有如下**性质**：

1. 零向量$\boldsymbol{\theta} = (0,0,\cdots,0)$一定是一个码字，称为零码字；
2. 任意两码字的和或差仍是一个码字；**线性性**
3. 任意码字c是G的行向量$g_0, g_1, \cdots, g_{k-1}$的线性组合；**对称性**
4. 线性分组码的最小距离等于最小非零码字重量，即

$$d_{\min} = \min_{c \neq \theta} w(c) \tag{5.2.6}$$

最小码距 ⟷ 最小码重

- 线性分组码的译码

对于GF(2)上的$k \times n$矩阵$\boldsymbol{G}$，存在$(n-k)\times n$矩阵$\boldsymbol{H}=\begin{bmatrix} h_{0,0} & \cdots & h_{0,n-1} \\ \vdots & \ddots & \vdots \\ h_{n-k-1,0} & \cdots & h_{n-k-1,n-1} \end{bmatrix}$，

使得

$$\boldsymbol{G}\boldsymbol{H}^{T}=[0]_{k\times(n-k)} \tag{5.2.8}$$

$\boldsymbol{H}$：一致校验矩阵。$\boldsymbol{H}^{T}$为$\boldsymbol{H}$的转置。由式(5.2.3)有

$$\boldsymbol{c}\boldsymbol{H}^{T}=\boldsymbol{m}\boldsymbol{G}\boldsymbol{H}^{T}=\boldsymbol{m}[0]_{\boldsymbol{k}\times(\boldsymbol{n}-\boldsymbol{k})}=\boldsymbol{\theta} \tag{5.2.9}$$

$\boldsymbol{\theta}$为$(n-k)$维零向量。

5.2 线性分组码

- 线性分组码的译码

如果生成矩阵G具有形式 $$G = G_s = \left[I_k \ Q_{k\times(n-k)} \right] \tag{5.2.10}$$

则成该码为系统码，其中 I_k 为 $k \times k$ 单位阵。

相应的一致校验矩阵H_S

$$H = H_s = \left[Q^T_{k\times(n-k)} \ I_{(n-k)} \right] \tag{5.2.11}$$

G_S与H_S仍然满足

$$G_s H_s^T = [0]_{k\times(n-k)}$$

G和H的角色可以互换，即H作为生成矩阵，G为校验矩阵。由H生成的码字称为c的对偶码。

- 线性分组码的译码

例 5.2.3 一个(5,3)线性分组码的生成矩阵$\boldsymbol{G}=\begin{bmatrix}1&0&1&1&0\\0&1&0&1&1\\1&1&0&1&0\end{bmatrix}$。求对应的$\boldsymbol{G}_S$和$\boldsymbol{H}_S$。

对$\boldsymbol{G}$进行初等变换，用R_i表示$\boldsymbol{G}$的第i行

$$R_3 \leftarrow R_2 \oplus R_3, R_1 \leftarrow R_1 \oplus R_3, R_1 \leftrightarrow R_3$$

$$\boldsymbol{G}_s=\left[\begin{array}{ccc|cc}1&0&0&0&1\\0&1&0&1&1\\0&0&1&1&1\end{array}\right] \qquad \boldsymbol{H}_s=\left[\begin{array}{ccc|cc}0&1&1&1&0\\1&1&1&0&1\end{array}\right]$$

5.2 线性分组码

- 线性分组码的译码

表5.2.1 (5,3)线性分组码码例

消息m	000	001	010	011	100	101	110	111
G生成码字	00000	11010	01011	10001	10110	01100	11101	00111
G_s生成码字	00000	00111	01011	01100	10001	10110	11010	11101
对偶码码字			00000	11101	01110	10011		

- 线性分组码的译码

码字通过信道传输可能出错，若信道无记忆，接收向量可表示成码矢量和差错的线性叠加。

接收向量：$\boldsymbol{r}=(r_0,r_1,\cdots,r_{n-1})$　　码矢量：$\boldsymbol{c}=(c_0,c_1,\cdots,c_{n-1})$

错误图样：$\boldsymbol{e}=(e_0,e_1,\cdots,e_{n-1})$　　伴随式：$\boldsymbol{s}=(s_0,s_1,\cdots,s_{n-k-1})$

$$\boldsymbol{r}=\boldsymbol{c}\oplus\boldsymbol{e}=(c_0\oplus e_0,c_1\oplus e_1,\cdots,c_{n-1}\oplus e_{n-1}) \tag{5.2.12}$$

$$\boldsymbol{s}=\boldsymbol{r}\boldsymbol{H}^T=\boldsymbol{c}\boldsymbol{H}^T\oplus\boldsymbol{e}\boldsymbol{H}^T \tag{5.2.13}$$

若$\boldsymbol{s}\neq\boldsymbol{\theta}$，则传输中一定有错误发生；若$\boldsymbol{s}=\boldsymbol{\theta}$，传输中无差错或错误图案恰好为一个码字。

5.2 线性分组码

- 线性分组码的译码

伴随式纠错译码步骤：

1. 计算伴随式，构造伴随式-差错图案表(s,e)；
2. 对接收向量计算伴随式 $\boldsymbol{s}=\boldsymbol{r}\boldsymbol{H}^{T}$
3. 查$(\boldsymbol{s},\boldsymbol{e})$表得$\boldsymbol{e}$;
4. 纠错:$\hat{\boldsymbol{c}}=\boldsymbol{r}\oplus\boldsymbol{e}$ 。

5.2 线性分组码

- 线性分组码的译码

例 5.2.4 (6,3)线性分组码，系统生成矩阵$\boldsymbol{G}_S=\begin{bmatrix}1&0&0&1&1&0\\0&1&0&0&1&1\\0&0&1&1&0&1\end{bmatrix}$，

接收矢量$\boldsymbol{r}$=(111001)，对其译码。

1 求系统校验矩阵。

$$\boldsymbol{H}_s=\begin{bmatrix}1&0&1&1&0&0\\1&1&0&0&1&0\\0&1&1&0&0&1\end{bmatrix}$$

2 由 $\boldsymbol{s}=\boldsymbol{e}\boldsymbol{H}_S^T$， 构造$(\boldsymbol{s},\boldsymbol{e})$表。

差错图案$\boldsymbol{e}$	000000	100000	010000	001000	000100	000010	000001
伴随式$\boldsymbol{s}$	000	110	011	101	100	010	001

5.2 线性分组码

- 线性分组码的译码

3 由式$\boldsymbol{s}=\boldsymbol{r}\boldsymbol{H}_s^T$计算接收向量伴随式，得$\boldsymbol{s}$=(001)。

4 查($\boldsymbol{s}$,$\boldsymbol{e}$)表可知，对应的差错图案$\boldsymbol{e}$=(000001)。

5 纠错计算得码字估值

$$\hat{\boldsymbol{c}}=\boldsymbol{r}\oplus\boldsymbol{e}=111001\oplus 000001=111000$$

6 恢复原始消息。对于系统(6,3)码，前三位即发送消息，即m=(111)，非系统码需要查消息-码字对照表恢复原始消息。

- 线性分组码的译码

定理： 若纠错码的最小距离为$d_{\min}$，如下任何一个结论独立成立。

1. 可以检测出任意小于等于$l=d_{\min}-1$个差错；
2. 可以纠正任意小于等于$t=\left[\frac{d_{\min}-1}{2}\right]$个差错；
3. 可以检测出任意小于等于l同时纠正小于等于t个差错，其中l和t满足$l+t\leq d_{min}-1,\ t<l$。

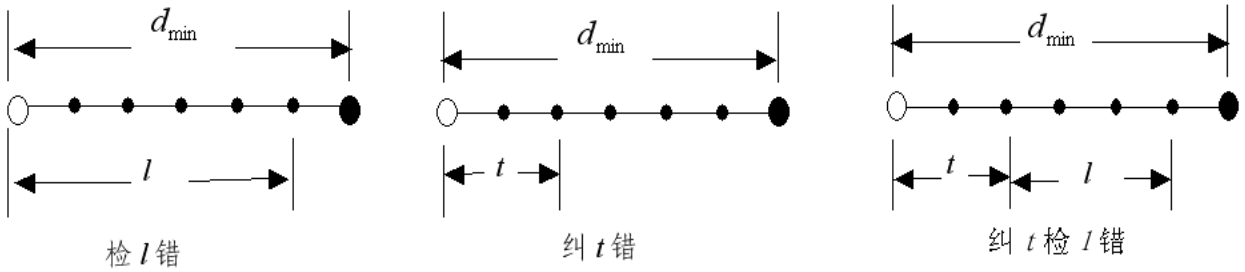


图5.2.2 最小码距与检错和纠错能力

- 典型码例

多个校验位的汉明码

每个校验位是部分或全部信息位按模二和规则确定。汉明码满足下列条件

码　　长：$n=2^r-1$

信息位长：$k=n-r=2^r-r-1$

码 字 数：$M=2^k$

监督位长：$r=n-k$

最小码距：$d_{min}=3$

纠错能力：$t=1$

- 典型码例

例5.2.5 二元(7,4)汉明码的系统码生成矩阵和校验矩阵分别为

$$\boldsymbol{G}_s=\begin{bmatrix}1&0&0&0&1&0&1\\0&1&0&0&1&1&1\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}$$

$$\boldsymbol{H}_s=\begin{bmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\1&1&0&1&0&0&1\end{bmatrix}$$

等价编码方程

$$\begin{cases}c_i=m_i, i=0,1,2,3\\ c_4=m_0\oplus m_1\oplus m_2\\ c_5=m_1\oplus m_2\oplus m_3\\ c_6=m_0\oplus m_1\oplus m_3\end{cases}$$

伴随式方程

$$\begin{cases}s_0=r_0\oplus r_1\oplus r_2\oplus r_4\\ s_1=r_1\oplus r_2\oplus r_3\oplus r_5\\ s_2=r_0\oplus r_1\oplus r_3\oplus r_6\end{cases}$$

$\boldsymbol{u}=(1101)$

$\boldsymbol{c}=\boldsymbol{u}\boldsymbol{G}_s=(1101001)$

- 典型码例

若传输无差错：

$$cH^{\mathrm{T}} = [1101001]\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 0$$

译码得：

$$\boldsymbol{u} = (1101)$$

5.2 线性分组码

对消息序列$\boldsymbol{m}=(m_0,m_1,m_2,m_3)=(1101)$编码，若传输过程中第4个比特出错，对接收矢量译码。

m_0	m_1	m_2	m_3	c_4	c_5	c_6
c_0	c_1	c_2	c_3	c_4	c_5	c_6

1. $\boldsymbol{C}=(1100001)$
2. 计算$(\boldsymbol{s},\boldsymbol{e})$表
3. 计算伴随式$\boldsymbol{s}=(011)$
4. 查$(\boldsymbol{s},\boldsymbol{e})$表，得错误图案$\boldsymbol{e}=(0001000)$
5. 码估值，$\hat{\boldsymbol{c}}=\boldsymbol{r}\oplus\boldsymbol{e}=(1101000)$
6. 恢复消息，$\boldsymbol{m}=(1101)$

表5.2.4 $(\boldsymbol{s},\boldsymbol{e})$表

e	s
0000000	000
1000000	101
0100000	111
0010000	110
0001000	011
0000100	100
0000010	010
0000001	001

- 循环码的基本概念

如果一个线性分组码的任意一个码字c(n元组)都是另外一个码字c'的循环移位，称此码为线性循环码。

将例5.2.5中的(7,4)汉明码生成矩阵做如下初等变换

$$R_1 \leftarrow R_1 \oplus R_3 \oplus R_4,\quad R_2 \leftarrow R_2 \oplus R_4$$

得到一个新的生成矩阵

$$\boldsymbol{G}' = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

矩阵的每一行都是第1行循环右移的结果。

循环码也可由多个码字分别循环构成

5.3 循环码

- 循环码的描述

1 循环码的多项式描述

记n维矢量或n元组a的一次右移循环为$a^{(1)}$,i次右移循环为$a^{(i)}$:

$$\begin{cases} a=(a_0,a_1,\cdots,a_{n-2},a_{n-1}),\ a_i\in\{0,1\} \\ a^{(1)}=(a_{n-1},a_0,a_1,\cdots,a_{n-3},a_{n-2}) \\ a^{(i)}=(a_{n-i},a_{n-i+1},\cdots,a_{n-1},a_0,a_1,\cdots,a_{n-1-i}),\ i=1,2,\cdots,n \end{cases}$$

任一矢量可用多项式$a(x)$来描述

$$a(x)=a_0+a_1x+\cdots+a_{n-1}x^{n-1} \tag{5.3.1}$$

$a_0,a_1,\ldots,a_{n-1}$为多项式系数。

5.3 循环码

多项式的次数$\partial^0 a(x)$定义为

$$\partial^0 a(x) = \max\left\{ i \middle| a_i \neq 0, i = 0,1,\cdots,n-1 \right\}$$

零次多项式 $\Longrightarrow a(x) = a_0$

零多项式 $\Longrightarrow a(x) = 0$

首一多项式：次数项系数为1的多项式。二元多项式均为首一多项式。

多项式加法$a(x) \oplus g(x)$

$$a(x) \oplus g(x) = (a_0 \oplus g_0) + (a_1 \oplus g_1)x + \cdots + (a_{n-1} \oplus g_{n-1})x^{n-1}$$

多项式乘法 $a(x) = a_0$

$$a(x)g(x) = f(x) = f_0 + f_1 x + \cdots + f_l x^l$$

$$l = \partial^\circ f(x) = \partial^\circ a(x) + \partial^\circ g(x) = n + m - 2$$

5.3 循环码

$$f_i=\begin{cases}\sum_{j=0}^{i}a_j g_{i-j}, & i=0,1,\cdots,m-1,\ m\le n\\ \sum_{j=i-m}^{n}a_j g_{i-j}, & i=m+1,\cdots m+n-2\end{cases}$$

其中$a(x)$称为$f(x)$的因式，$f(x)$称为$a(x)$的倍式。如果$f(x)\neq a(x)g(x)$，两者关系可用下式表示

$$f(x)=a(x)g(x)+r(x),\quad 0\le\partial^{\circ}r(x)<\partial^{\circ}g(x)$$

$$f(x)=r(x)\pmod{g(x)}$$

余　式　　模多项式

多项式模运算使用长除法。

5.3 循环码

采用多项式$a(x),a^{(1)}(x),a^{(i)}(x)$ 来一对一描述向量和其循环移位$a,a^{(1)},a^{(i)}$

$$\begin{cases} a(x)=a_{n-1}x^{n-1}+a_{n-2}x^{n-2}+\cdots+a_1x+a_0 & a_i\in\{0,1\} \\ a^{(1)}(x)=a_{n-2}x^{n-1}+a_{n-3}x^{n-2}+\cdots+a_1x^2+a_0x+a_n \\ a^{(i)}(x)=a_{n-1-i}x^{n-1}+a_{n-2-i}x^{n-2}+\cdots+a_1x^{i+1}+a_0x^i+a_{n-1}x^{i-1}+\cdots+a_{n-i} \end{cases} \tag{5.3.2}$$

比较$a(x)$ 和$a^{(i)}(x)$ 可以发现

$$\begin{aligned} a^{(1)}(x)&=xa(x) \pmod{(x^n+1)} \\ a^{(i)}(x)&=x^ia(x) \pmod{(x^n+1)} \end{aligned} \tag{5.3.3}$$

码多项式(码式)

$$C=\left\{c\middle|c=b^{(i)},b\in C\right\}\leftrightarrow C(x)=\left\{c(x)\middle|c(x)=b^{(i)}(x),b(x)\in C(x)\right\}$$

5.3 循环码

- 循环码的描述

定理1 (n,k)循环码$C(x)$中存在一个非零的、首一的、最低次数为$r(r<n)$的码式，满足：

1. $g(x)$是唯一的；
2. $g(x)$的零次项 $g_0 \neq 0$ ；
3. $c(x)$是码式当且仅当$c(x)$是$g(x)$的倍式；
4. $r=n-k$。

5.3 循环码

- 循环码的描述

满足定理1的码式$g(x)$称为(n,k)循环码的生成多项式。

定理2 $g(x)$是(n,k)循环码的生成多项式，当且仅当$g(x)$是x^n+1的$r=n-k$次因式。

循环码由生成多项式$g(x)$的倍式组成。

$$C(x)=\{c(x)\,|\,c(x)=a(x)g(x),\partial^{\circ}a(x)<k\}$$

5.3 循环码

- 循环码的描述

例5.3.2 x^7+1 的因式分解为

$$x^7+1=(x+1)(x^3+x^2+1)(x^3+x+1)$$

因此，码长n=7的循环码可以由上式任意因式生成。

1. (7,4)循环码A，生成多项式$g(x)=(x^3+x^2+1)$，码式

$$c(x)=(a_0+a_1x+a_2x^2+a_3x^3)(x^3+x^2+1)$$

2. (7,4)循环码B，生成多项式 $g(x)=(x^3+x+1)$，码式

$$c(x)=(a_0+a_1x+a_2x^2+a_3x^3)(x^3+x+1)$$

3. (7,3)循环码C，生成多项式$g(x)=(x^3+x+1)(x+1)$，码式

$$c(x)=(a_0+a_1x+a_2x^2+a_3x^3)(x^4+x^2+x+1)$$

5.3 循环码

- 循环码的描述

2 循环码的矩阵描述

将码式$g(x)$的系数用n维向量表示，(n,k)循环码的生成矩阵为

$$G=\begin{bmatrix} g_0 & g_1 & g_2 & \cdots & g_{r-1} & g_r & 0 & \cdots & 0 & 0 \\ 0 & g_0 & g_1 & \cdots & g_{r-2} & g_{r-1} & g_r & 0 & \cdots & 0 \\ & & & \cdots & & & & \cdots & & \\ 0 & 0 & 0 & \cdots & g_0 & g_1 & g_2 & \cdots & g_{r-1} & g_r \end{bmatrix}_{k\times n}$$

定义(n,k)循环码的一致校验多项式

$$h(x)=\frac{x^n-1}{g(x)}=h_0+h_1x+\cdots+h_{k-1}x^{k-1}+h_kx^k$$

5.3 循环码

- 循环码的描述

对应的(n,k)循环码一致校验矩阵为

$$H=\begin{bmatrix} h_k, h_{k-1}, & \cdots & h_0, 0, & 0, & \cdots 0 \\ 0, & h_k, & \cdots & h_1, h_0, 0, & \cdots 0 \\ \cdots & & \cdots & \cdots & \cdots \\ 0, \cdots 0, & & h_k, & h_{k-1}, \cdots & h_0 \end{bmatrix}_{r\times n}$$

注意生成矩阵的行向量为$g(x)$系数的升幂排列时，对应的一致校验矩阵$h(x)$系数为降幂排列；如果G为降幂排列，则H为升幂排列。

- 循环码的描述

例5.3.3

(7,4)循环码A，$h(x)=x^4+x^3+x^2+1$，$g(x)$为升幂排列时的$\boldsymbol{G}$和$\boldsymbol{H}$为

$$\boldsymbol{G}=\begin{bmatrix}1&0&1&1&0&0&0\\0&1&0&1&1&0&0\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{bmatrix}$$

$$\boldsymbol{H}=\begin{bmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\0&0&1&1&1&0&1\end{bmatrix}$$

- 循环码的描述

3 系统循环码

循环码的系统码码式为

码多项式

$$c(x) = x^r m(x) \oplus \left(x^r m(x) \bmod g(x) \right) \tag{5.3.12}$$

相应的循环码系统码的生成矩阵为

$$\boldsymbol{G}_s = \begin{bmatrix} p_{00} & p_{01} & \cdots & p_{0,r-1} & 1 & 0 & 0 \\ p_{10} & p_{11} & \cdots & p_{1,r-1} & 0 & 1 & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ p_{k-1,0} & p_{k-1,1} & & p_{k-1,r-1} & 0 & 0 & 1 \end{bmatrix} \tag{5.3.13}$$

5.3 循环码

- 循环码的描述

例5.3.4 (7,4)循环码$g(x)=1+x+x^3$

$$x^3=g(x)+(1+x), \qquad c_0(x)=1+x+x^3$$
$$x^4=xg(x)+(x+x^2), \qquad c_1(x)=x+x^2+x^4$$
$$x^5=(x^2+1)g(x)+(1+x+x^2), \qquad c_2(x)=1+x+x^2+x^5$$
$$x^6=(x^3+x^2+1)g(x)+(1+x^2), \qquad c_3(x)=1+x^2+x^6$$

$$G_s=\begin{bmatrix}1&1&0&1&0&0&0\\0&1&1&0&1&0&0\\1&1&1&0&0&1&0\\1&0&1&0&0&0&1\end{bmatrix}\sim\begin{pmatrix}g(x)\\xg(x)\\(x^2+1)g(x)\\(x^3+x^2+1)g(x)\end{pmatrix}$$

5.3 循环码

- 循环码的伴随多项式与检错

$$r(x) = c(x) \oplus e(x) \tag{5.3.14}$$

$$\begin{aligned} s(x) &= r(x)(\mathrm{mod}\, g(x)) = (c(x) \oplus e(x))(\mathrm{mod}\, g(x)) \\ &= (a(x)g(x) \oplus e(x))(\mathrm{mod}\, g(x)) = e(x)(\mathrm{mod}\, g(x)) \end{aligned} \tag{5.3.16}$$

其中 $c(x) = a(x)g(x) = 0 \ \ (\mathrm{mod}\, g(x))$

如果$s(x) \neq 0$，一定有差错产生；如果$s(x) = 0$，则无差错或者有不可检差错。循环码在自动请求重传(ARQ)方式中得到广泛应用。用于ARQ的循环码，生成多项式称为CRC多项式。

$$CRC12(x) = x^{12} + x^{11} + x^3 + x^2 + x + 1$$

5.3.4 BCH码与RS码

在一个有2^m个元素的有限域中如果选择本原元的某些连续幂次的元素为$g(x)$的全部根，则由此得到的循环码称为二元本原BCH码。

对于任意的整数m和可达的纠错数t，均可构造距离为d_0的二元本原BCH码，其码长、校验位长和最小距离满足(码例见表5.3.1)

$$n = 2^m - 1,\ r = n - k \le mt,\ d_{\min} \ge d_0 = 2t + 1 \qquad (5.3.17)$$

RS码是多元BCH码的一个子类，其码长n、符号长m、校验位长r和纠错数t的关系为

$$n = m \cdot (2^m - 1)(bit),\ \ r = n - k = m \cdot 2t(bit) \qquad (5.3.19)$$