云南大学数学与统计学院
上机实践报告

课程名称：近代密码学实验	**年级**：2015 级	**上机实践成绩**：
指导教师：陆正福	**姓名**：刘鹏	
上机实践名称：纠错密码学实验	**学号**：20151910042	**上机实践日期**：2018-06-10
上机实践编号：No.05	**组号**：	**上机实践时间**：18:45

一、 实验目的

1. 熟悉纠错密码学的基本概念；
2. 掌握纠错密码学编程的基本方法；
3. 理解纠错码用于密码体制设计的基本思路；
4. 查阅资料，尽可能将原理级的验证型实验逐渐演化为实现级的设计型实验。

二、实验内容

1. 复习“信息论基础实验”课程中 Hamming 码实验，做简单测试与分析；
2. 以 Hamming 码为基础，设计 McEliece 密码体制：密钥对生成算法，加密算法，解密算法；
3. 分析以 Hamming 码为基础的纠错密码体制的局限性；
4. 通过实验说明，纠错密码学中的基础纠错码应该具备哪些要点才能保证密码体制的安全性。
5. 分析 Gopal 码用于纠错密码体制有何优势？

说明：与前四个实验有所变化的是：前四个实验与大整数有关，本实验的必做部分则与有限域、有限域上的矩阵有关。部分基础纠错码与有限域上的多项式（或有理分式）有关。如果要做安全性更强的实验，则需要选择合适的配套代码库，或做与数据结构与算法设计有关的基础编程。

三、实验平台

Windows 10 Pro Workstation 1803；
SageMath version 8.2, Release Date: 2018-05-05；

四、实验记录与实验结果分析

1 题

复习“信息论基础实验”课程中 Hamming 码实验，做简单测试与分析。

Solution:

背景材料

信道容量：

如果在信道传输过程中存在误差，那么如何纠正所有误差？任何纠错过程本身也要受到误差的影响，这样的话纠正过程将会无穷无尽地进行下去。

为了证明只要码率小于信道容量，信息就可以通过该信道可靠地传输，香农使用了许多新的思想，这些思想包括：

- 允许任意小的非0误差概率存在；
- 连续使用信道许多次，以保证可以使用大数定律；
- 在随机选择的码簿上计算平均误差概率，这样可以使概率对称，而且可以用来证明至少存在一个好的编码。

信道编码定理（信道容量的可达性）：对于离散无记忆信道，小于信道容量C的所有码率都是可达的。具体说来，对任意码率$R < C$，存在一个$(2^{nR},\, n)$码序列，它的最大误差概率$\lambda^{(n)} \to 0$。反之，任何满足$\lambda^{(n)} \to 0$的$(2^{nR},\, n)$码序列必定有$\mathrm{R} \leq \mathrm{C}$。

信道编码定理使用分组码的方案。如果分组长度足够大的话，当码率小于信道容量时，可以用分组码以任意低的误差概率传输信息。通过重复添加冗余的方式，可以用降低码率的代价获得一定的纠错能力。寻找替代这种简单的重复比特的方法的一种思路是“通过一种巧妙的方式将比特联合起来，使得每一个额外的比特都可以用来检验某个信息比特子集中是否发生错误”。奇偶校验码就是这种思路的一个方案，不过在传输随机比特流的时候，奇偶校验码只有侦错能力，没有纠错能力。不过通过推广奇偶校验的思想，允许存在多个奇偶校验位，也可以允许奇偶校验依赖于各种各样的信息比特子集。这直接导致了汉明码的发明。

下面考虑分组长度为7的二元码。下述所有的运算都在模2意义下进行。考虑所有长度为3的非0二元向量的集合，以它们为列向量构成一个矩阵：

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

考虑$\boldsymbol{H}$的模 2 意义下的零空间，根据线性代数中的线性不定方程组理论：因为$\mathrm{rank}(\boldsymbol{H}) = 3$，所以$\boldsymbol{H}$的零空间的维数为$7 - 3 = 4$。可以通过类似实数域上的一般高斯消元法的方式，获得解空间中所有的元素（让通解中的四个自由变量取遍$\{0,\, 1\}$中的所有值之后得到2^4个变量：

0000 000	0100 101	1000 011	1100 110
0001 111	0101 010	1001 100	1101 001
0010 110	0110 011	1010 101	1110 000
0011 001	0111 100	1011 010	1111 111

因为上面这个码字集合是一个线性空间，所以具有线性空间的八条性质：加法封闭、可交换、存在加法单位元；乘法满足存在数乘单位元、数量乘法结合律、数量乘法的数量分配律、向量分配律。根据加法封闭性，可以知道以上集合里任意两个元素之和必然也在以上集合里。通过观察可以发现，除了全为零的码字之外的所有码字中，1的最小数目为3。

由此可以看出，上面这个七位的码的最小重量为3。下面断言，在这个编码中，任意两个码字之间至少在3个位置上有所不同。

Claim: 在上述的七位码中，任意两个码字之间至少在3个位置上有所不同。

Proof: 命这个码字集合为$\boldsymbol{C}$。假设$\boldsymbol{C}$中有一个重量小于3的码$\boldsymbol{c}$存在，不妨设其重量为2。根据定义，有$\boldsymbol{Hc} = \boldsymbol{0}$，设$\boldsymbol{c} = 0110\,000$（其他的也同样讨论）。在这种情况下，要使得$\boldsymbol{Hc} = \boldsymbol{0}$成立，则要求$\boldsymbol{H}$的第2，3列完全相同，但是这与$\boldsymbol{H}$的性质相悖，所以不可能存在重量小于3的码字，即上述码字集合的最小重量至少为3。然后，该码字集合构成线性空间，所以任意两个码字之和（模2意义下）也在这个集合里，即$\forall \boldsymbol{a},\, \boldsymbol{b} \in \boldsymbol{C},\, \boldsymbol{a} + \boldsymbol{b} = \boldsymbol{c} \in \boldsymbol{C}$，再定义加法逆运算（自然地，如$11 = 00 - 11$，$10 = 11 - 01$，可以知道，当且仅当两个不同的比特做差才有可能产生1），可知$\boldsymbol{a} = (\boldsymbol{c} - \boldsymbol{b}) \in \boldsymbol{C}$，$\boldsymbol{a}$的重量为3，所以可以知道$\boldsymbol{c}$与$\boldsymbol{b}$的不同至少有三处，由任意性知道，$\boldsymbol{C}$中任意两个元素至少有三个位是不同的。

根据这个性质可以知道，如果在传输过程中发生了两个及以下数量的错误，是可以纠正的，而且这个并不需要降低码率。这个并不需要做实验，因为实验仅仅只是比对两个码字表而已。关键之处是在码字表很大的情况下，有没有办法不通过穷举就找到最相近的码字？可以利用$\boldsymbol{H}$的性质进行。矩阵$\boldsymbol{H}$称为奇偶校验矩阵（parity check matrix），对于任意一个码字$\boldsymbol{c}$，都有$\boldsymbol{Hc} = \boldsymbol{0}$。如果发送一个码字$\boldsymbol{c}$，接收到的向量为$\boldsymbol{r}$，且在第$i$位发生了错误，那么记$\boldsymbol{r} = \boldsymbol{c} + \boldsymbol{e}_i$，有如下等式：

$$\boldsymbol{Hr} = \boldsymbol{H}(\boldsymbol{c} + \boldsymbol{e}_i) = \boldsymbol{Hc} + \boldsymbol{He}_i = \boldsymbol{He}_i$$

$\boldsymbol{He}_i$恰好是$\boldsymbol{H}$第i列的矩阵。比对一下列，然后还原（取反）即可。虽然可以快速找错，但是这样就使得这个码只能纠正一个错误了。这就是汉明码的原理。

2 题

以 Hamming 码为基础，设计 McEliece 密码体制：密钥对生成算法，加密算法，解密算法。

背景材料

六、实验体会

七、参考文献

[1] STALLINGS W. 密码编码学与网络安全：原理与实践 [M]. 6th ed. 北京: 机械工业出版社, 2015.
[2] https://doc.sagemath.org/html/en/reference/curves/sage/schemes/elliptic_curves/ell_point.html