Week 7

Hamming codes

Version 2023-11-04. To accessible online version of this chapter

Synopsis. *Hamming codes are essentially the first non-trivial family of codes that we shall meet. We give a construction of a q-ary Hamming code and prove that it is perfect with minimum distance 3. We show that syndrome decoding works for Hamming codes in an especially simple way.*

Finding a check matrix

Before we can construct Hamming codes, we need to discuss check matrices further and prove a result (the Distance Theorem) which will allow us to find the minimum distance of a linear code from its check matrix.

The following result allows us to find a generator matrix of $C^\perp$, assuming that C has a generator matrix in standard form.

Theorem 7.1: a check matrix construction

Assume that C has a $k \times n$ generator matrix $G = [\, I_k \,|\, A \,]$ in standard form. Then the dual code $C^\perp$ has generator matrix

$$H = [\, -A^T \,|\, I_{n-k} \,].$$

Proof. H has $n-k$ rows which are linearly independent (due to I_{n-k} present). It is enough to show that each row $\underline{r}$ of H is a codevector of $C^\perp$: indeed, we have $n-k$ linearly independent vectors in $C^\perp$, and $n-k$ is the dimension of $C^\perp$ by Theorem 5.1, so a linearly independent set of $n-k$ vectors must be a basis of $C^\perp$.

By Theorem 5.1, it is enough to show that $\underline{r}G^T = \underline{0}$. We will show this at once for all rows

of H, by proving that HG^T is the zero matrix. Indeed,

$$[\,-A^T \,|\, I_{n-k}\,] \begin{bmatrix} I_k \\ \hline A^T \end{bmatrix} = -A^T I_k + I_{n-k} A^T = -A^T + A^T = 0. \qquad \square$$

How can one find a check matrix of C if C has no generator matrix in standard form? We address this question below.

Linearly equivalent codes

Definition: linearly equivalent codes

Two linear codes $C, C' \subseteq \mathbb{F}_q^n$ are **linearly equivalent**, if C' can be obtained from C by a sequence of linear transformations of the following types:

(C1) choose indices i, j; in every codeword, swap symbols x_i and x_j;

(C2) choose index i and non-zero $\lambda \in \mathbb{F}_q$; in every codeword, multiply x_i by λ.

Exercise. Linearly equivalent codes have the same length, dimension and weight. They have the same weight enumerator. (*Reason*: (C1) and (C2) do not change the weight of any vector.)

Fact: known from linear algebra

Every generator matrix can be brought into the standard form by using row operations (R1), (R2), (R3) considered above and column operations (C1).

Reason: any matrix can be brought to reduced row echelon form, RREF, by (R1)–(R3); a generator matrix has linearly independent rows so the RREF won't have zero rows and will have a leading entry 1 in each of the k rows; the k columns which contain the leading entries are columns of the identity matrix of size k; use (C1) to move all these columns to the left.

Conclusion: we can always find a generator matrix in standard form for a linearly equivalent code.

The Distance Theorem

We already know how to read the length and the dimension of a linear code C off a check matrix H of C:

- the number of columns of H is the length of C;
- the number of columns minus the number of rows of H is the dimension of C.

The following theorem tells us how to determine the minimum distance of C using H.

Theorem 7.2: Distance Theorem for linear codes

Let $C \subseteq \mathbb{F}_q^n$ be a linear code with check matrix H. Then $d(C) = d$ if and only if every set of $d-1$ columns of H is linearly independent and some set of d columns of H is linearly dependent.

Proof. Let e be the size of a smallest linearly dependent subset of the set $\{\overline{h}_1, \ldots, \overline{h}_n\}$ of columns of H. The theorem claims that $e = d(C)$. Note that e is the minimum positive number of non-zero coefficients x_i in the linear combination

$$x_1\overline{h}_1 + x_2\overline{h}_2 + \ldots + x_n\overline{h}_n = \overline{0},$$

i.e., the minimum weight of non-zero $\underline{x} = (x_1, \ldots, x_n)$ such that $\underline{x}H^T = \underline{0}$. By Theorem 5.1, such vectors $\underline{x}$ are exactly the codevectors of C, so $e = w(C) = d(C)$ as claimed. □

Example: calculate $d(C)$ using the Distance Theorem

Use the Distance Theorem to find the minimum distance of the ternary linear code with check matrix $H = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 2 & 0 & 1 & 1 \end{bmatrix}$.

Solution. *Step 1.* H has no zero columns. Hence every set of 1 column is linearly independent (a one-element set is linearly dependent iff that element is zero). So $d \geq 2$.

Step 2. Any two columns of H are linearly independent, because no two columns are proportional to each other. So $d \geq 3$.

Step 3. There are three linearly dependent columns in H: for example, columns 1, 2 and 3 form linear combination $\begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \overline{0}$. Therefore, $d = 3$.

Hamming codes: the construction

Definition: line, representative vector, projective space

A **line** is a 1-dimensional subspace of the vector space $\mathbb{F}_q^n$.
A **representative vector** of a line is a non-zero vector $\underline{u}$ from that line. The line is then given by $\{\lambda\underline{u} \mid \lambda \in \mathbb{F}_q\}$.
The **projective space** $\mathbb{P}_{n-1}(\mathbb{F}_q)$ is the set of all lines in $\mathbb{F}_q^n$.

Remark: the terminology comes from euclidean geometry — in the euclidean plane, the set of all vectors proportional to a given non-zero vector is a straight line through the origin. Projective spaces over the field $\mathbb{R}$ of real numbers are well-studied geometric objects.

For example, $\mathbb{P}_1(\mathbb{R})$ — the set of all lines through the origin in the euclidean plane — can be thought of as the unit circle with antipodes identified. We are working over a finite field $\mathbb{F}_q$ where these notions are less intuitive.

Definition: Hamming codes

Let $r \geq 2$ be given. We let $\mathbf{Ham}(\boldsymbol{r}, \boldsymbol{q})$ denote an $\mathbb{F}_q$-linear code whose check matrix has columns which are representatives of the lines in $P_{r-1}(\mathbb{F}_q)$, exactly one representative vector from each line.

Remark: $\mathrm{Ham}(r, q)$ is not one code but a class of linearly equivalent codes

$\mathrm{Ham}(r, q)$ is defined up to a linear equivalence. Indeed, we can:

- multiply a column by non-zero λ to get another representative of the same line;
- put columns in any order.

This means that $\mathrm{Ham}(r, q)$ is not just one code but a class of linearly equivalent codes. We will therefore say "*a* $\mathrm{Ham}(r, q)$ *code*" to mean any of the linearly equivalent codes.

Let us see how the construction works in historically the first example of a Hamming code.

Example: $\mathrm{Ham}(3, 2)$

Construct a parity check matrix for a binary Hamming code $\mathrm{Ham}(3, 2)$. Then find a generator matrix in standard form for $\mathrm{Ham}(3, 2)$.

Solution: we need to take one non-zero column from each line in $\mathbb{F}_2^3$. For binary vectors, a line $\{\lambda \underline{u} \mid \lambda \in \mathbb{F}_2\}$ consists only of two points, $\underline{0}$ and $\underline{u}$. This means that a check matrix for a **binary** Hamming code consists of **all non-zero binary columns** or the required size.

Start filling in the check matrix by putting the identity columns at the end (this is convenient for finding a generator matrix). In total, there are 7 non-zero binary vectors of size 3:

$$H = \left[\begin{array}{cccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

From this H, we can reconstruct the generator matrix $G = [I_k \mid A]$ by Theorem 7.1:

$$G = \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}\right]$$

This is, up to linear equivalence, the generator matrix of the original code of R. Hamming.

Historical remark. Despite their name, the q-ary Hamming codes for $q > 2$ were not invented by Hamming. Richard Hamming told Claude Shannon (who he shared an office with at Bell Labs) about his binary $[7, 4, 3]$-code, and Shannon mentioned it in his paper of 1948. That paper was read by **Marcel J. E. Golay** (1902–1989), a Swiss-born American mathematician and electronics engineer, who then suggested the $\mathrm{Ham}(r, q)$ construction in his paper published in 1949. Golay went further and constructed two perfect codes which are not Hamming codes. He asked whether there are any more perfect codes.

We will see the Golay codes, and will learn about an answer to Golay's question about perfect codes, later in the course.

Parameters of a Hamming code

We considered an example of a $\mathrm{Ham}(3, 2)$ code, which — by looking at its generator matrix — turns out to be a $[7, 4, d]_2$ code. It is not difficult to see directly that $d = 3$. By explicitly computing the Hamming bound, one can show that all $[7, 4, 3]_2$-codes are perfect.

We will now generalise this and show that all Hamming codes are perfect.

Theorem 7.3: properties of Hamming codes

$\mathrm{Ham}(r, q)$ is a perfect $[n, k, d]_q$ code where $n = \dfrac{q^r - 1}{q - 1}$, $k = n - r$, $d = 3$.

Proof. The length n of the code is equal to the number of columns in the check matrix, which is $\#\mathbb{P}_{r-1}(\mathbb{F}_q)$, the number of lines in $\mathbb{F}_q^r$.

Observe that two lines intersect only at one point, namely $\overline{0}$. The set $\mathbb{F}_q^r \setminus \{\overline{0}\}$ is therefore a disjoint union of lines. Each line $\{\lambda\overline{u} : \lambda \in F\}$ contains $q - 1$ non-zero points.

So the number of lines in $\mathbb{F}_q^r$ can be found as $\dfrac{\#(\mathbb{F}_q^r \setminus \{\overline{0}\})}{q - 1} = \dfrac{q^r - 1}{q - 1}$.

We have $k = \dim \mathrm{Ham}(r, q) = n - r$ since, by construction, the check matrix H has r rows.

To find d, we use the Distance Theorem for linear codes. Any two columns of H are linearly independent because they are from different lines in $\mathbb{F}_q^r$. (Two vectors are linearly dependent only if they are proportional to each other, i.e., belong to the same line.) Therefore, $d \geq 3$.

On the other hand, H has columns $(a,0,0,\ldots,0)^T$, $(0,b,0,\ldots,0)^T$ and $(c,c,0,\ldots,0)^T$, from three different lines (where $a,b,c \in \mathbb{F}_q \setminus \{0\}$). These columns are linearly dependent:

$$a^{-1}\begin{bmatrix} a \\ 0 \\ \vdots \\ 0 \end{bmatrix} + b^{-1}\begin{bmatrix} 0 \\ b \\ \vdots \\ 0 \end{bmatrix} - c^{-1}\begin{bmatrix} c \\ c \\ \vdots \\ 0 \end{bmatrix} = \overline{0}.$$

So $d = 3$ by the Distance Theorem.

It remains to show that $\mathrm{Ham}(r,q)$ is perfect. We calculate $t = [(d-1)/2] = [2/2] = 1$. The Hamming bound (in logarithmic form) then says

$$k \leq n - \log_q\left(\binom{n}{0} + \binom{n}{1}(q-1)\right) = n - \log_q\left(1 + n(q-1)\right).$$

By the already proved formulae for n and k we have $n(q-1) = q^r - 1$ and $k = n - r$. Hence the bound is $n - r \leq n - \log_q(q^r) = n - r$ — attained. Thus, $\mathrm{Ham}(r,q)$ is perfect. □

Remark: $(q^r-1)/(q-1)$ is an integer

The proof shows that the fraction $\dfrac{q^r-1}{q-1}$ is an integer. In fact, this can be seen for all integers $q, r > 1$ by a formula for summing a geometric progression, $\dfrac{q^r-1}{q-1} = q^{r-1} + q^{r-2} + \cdots + q + 1$; the right-hand side is obviously an integer.

Decoding a Hamming code

Algorithm 7.4: decoding algorithm for a Hamming code

Let a Hamming code be given by its check matrix H. Suppose a vector $\underline{y}$ is received.

- Calculate $S(\underline{y}) = \underline{y}H^T$. If $S(\underline{y}) = \underline{0}$, $\texttt{DECODE}(\underline{y}) = \underline{y}$.
- Otherwise, $S(\underline{y}) = \lambda \times$ some column of H. Let this be the ith column of H.
- Subtract λ from the ith position in $\underline{y}$. The result is the codevector $\texttt{DECODE}(\underline{y})$.

Proof of validity of the algorithm. We prove that the algorithm outputs the nearest neighbour of $\underline{y}$ in the code C. This is clear if $S(\underline{y}) = \underline{y}H^T = \underline{0}$: by Proposition 5.2 $\underline{y}$ is a codevector, and so its own nearest neighbour in C. Hence it is correct to decode $\underline{y}$ to itself.

If $\underline{y}H^T \neq \underline{0}$, the line in $\mathbb{F}_q^r$ which contains $\underline{y}H^T$ has a representative column in H — say, h_i. As $\underline{y}H^T$ lies on the line spanned by h_i, we must have $\underline{y}H^T = \lambda h_i$ for some $\lambda \in \mathbb{F}_q$.

Note that λh_i equals $(\lambda \underline{e}_i)H^T$ where $\underline{e}_i$ is the unit vector with symbol 1 in position i and zeros elsewhere. It follows that

$$(\underline{y} - \lambda \underline{e}_i)H^T = \lambda h_i - \lambda h_i = \underline{0},$$

hence $\underline{y} - \lambda \underline{e}_i$ is a codevector. Finally, since $d(\underline{y}, \underline{y} - \lambda \underline{e}_i) = 1$, and no codevector can be at distance **less** than 1 from $\underline{y}$, we conclude that $\underline{y} - \lambda \underline{e}_i$ is the nearest neighbour of $\underline{y}$ in C. □

Remark: properties of a Hamming decoder

The Algorithm and the proof above imply:

- Every coset leader of $\mathrm{Ham}(r, q)$ is $\underline{0}$ or $\lambda \underline{e}_i$, i.e., a vector of weight 0 or 1.
- The decoder changes at most one symbol in the received vector.

Note that the fact that every coset leader is of weight ≤ 1 also follows, in a different way, from Exercise 4.3.

For $\mathrm{Ham}(3, 2)$, a clever ordering of columns in the parity check matrix can make the decoding algorithm especially elegant:

Example: special check matrix for $\mathrm{Ham}(3, 2)$

Construct a decoder for the $\mathrm{Ham}(3, 2)$ code with parity check matrix

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

Solution. If $\underline{y} \in \mathbb{F}_2^7$ is received, $\underline{y}H^T$ is either $\underline{0}$ or one of the columns of H. Now note that, by Algorithm 7.4,

- if $\underline{y}H^T = 001$, the decoder must subtract 1 from the first bit in $\underline{y}$, because 001 is the first column of H;
- if $\underline{y}H^T = 010$, the decoder must subtract 1 from the second bit in $\underline{y}$, because 010 is the second column of H;

and so on. Subtracting 1 from a bit in $\mathbb{F}_2$ is the same as "flipping" the bit, i.e., replacing 0 by 1 and 1 by 0.

Thus, to decode the received vector $\underline{y}$, we calculate the syndrome $\underline{y}H^T$. If this is 000, output $\underline{y}$, otherwise *read the syndrome $\underline{y}H^T$ as the binary representation of a number $i \in \{1, 2, \ldots, 7\}$ and decode by flipping the ith bit in $\underline{y}$.*