Control Matrices and Decoding

We want to find a way to detect and correct errors in a received message that works for linear codes. To do that we need the concept of dual code:

**Def** The dual code $C^\perp$ of a linear code $C$ in $F^n$ is the linear code $C^\perp = \{ y \in F^n \mid \langle x, y \rangle = 0 \text{ for all } x \in C \}$

(Here $\langle \cdot, \cdot \rangle$ denotes the scalar product $\langle x, y \rangle = x_1y_1 + x_2y_2 + \ldots + x_ny_n$)

Looking at a generator matrix $G$ for $C$, $C^\perp$ is the nullspace and $C$ the rowspace.

So by the dimension theorem $\dim C^\perp = n - \dim C = n - m$

**Note**: In $\mathbb{R}^n$ any vector can be uniquely decomposed into $u = u_1 + u_2$ $u_1 \in U$ $u_2 \in U^\perp$ for any subspace $U$. This is true for general $F^n$. (The problem is to construct an orthonormal basis for $U^\perp$ for $U = \langle (1, 1, 1) \rangle$ in $\mathbb{Z}_3^3$.)
Sometimes we even have $C = C^\perp$ a so-called self-dual code. One example is

$$G = \begin{pmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{pmatrix} \in \mathbb{Z}_5^4$$

The dual code is 2-dimensional and contains the rows in $G$ as they are perpendicular to each other and to themselves.

**Def:** A generator matrix for the dual code $C^\perp$ is called a control matrix for $C$.

We can check if $x \in C$ by checking if $x H^\perp = 0$ where $H$ is a control matrix. ($C^\perp)^\perp = C$)

**How to find a control matrix?**

**Thm 3.6:** If a linear $[n, m]$ code $C$ has the generator matrix $G = [I_m | A]$ then $H = [-A^\perp | I_{n-m}]$ is a control matrix.

**Proof:** Since $\dim C^\perp = n - \dim C = n - m$ $H$ should be a $(n-m) \times n$ matrix. A matrix of this size with $\dim$ independent rows satisfying $G \cdot \tilde{H}^\perp = 0$ has $n-m \dim$ ind. rows in $C^\perp$ and hence must be a control matrix.
Thus we just need to check that 
\[ H = [A^t \mid \mathbf{I}_{n-m}] \] is \((n-m)\times n\) and that \(GH^t = 0\).

\[
\begin{align*}
\text{size: } G &= [\mathbf{I}_m \mid A]^m \\
&\Rightarrow H = \underbrace{[A^t \mid \mathbf{I}_{n-m}]}_{n-m}^m \quad \text{OK}
\end{align*}
\]

and 
\[
\begin{align*}
GH^t &= [\mathbf{I}_m \mid A] \underbrace{[-A \mid \mathbf{I}_{n-m}]}_{n-m}^m m = -\mathbf{I}_m A + A \mathbf{I}_{n-m} = \\
&= -A + A = 0
\end{align*}
\]

Proof complete!

\[ G = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix} \] in \(\mathbb{Z}_3^4\) (the perfect code from last lecture) has control matrix 
\[ H = \begin{pmatrix} 2 & 2 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{pmatrix} \]

Detection and correction of errors

We detect errors by computing \(xH^t\). If \(xH^t = 0\), \(x \in C\), otherwise \(x \notin C\).

The vector \(xH^t\) of length \(n-m\) is called a syndrome. To correct \(x\) we look at all \(y \in \mathbb{F}_n\) with the same syndrome as \(x\). The reason is that if \(xH^t = yH^t\) then 
\[ (x-y)H^t = 0 \] that is \(x-y \in C\). As we want to alter \(x\) as little as possible we search the \(y\) of lowest weight satisfying...
$yH^t = xH^t$. Such a $y$ is called a coset leader (and the whole set $\{y \in F^n \mid yH^t = xH^t\}$ is called a coset). We replace $x$ by the code word $x-y$.

In the above example we have:

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>00</th>
<th>01</th>
<th>02</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>20</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coset</td>
<td>0000</td>
<td>0001</td>
<td>0002</td>
<td>0010</td>
<td>2000</td>
<td>0200</td>
<td>0020</td>
<td>0100</td>
<td>1000</td>
</tr>
</tbody>
</table>

To find the coset leaders we only need to look at $H$ as all syndromes here are multiples of columns. (That is because every word is within distance one from a code word in this example.)

Now assume that we received the words $x_1 = 0110$, $x_2 = 0221$, $x_3 = 1111$. We compute their syndromes using $H$ and get 01100 and 21. This shows that $0221$ is a code word, but not the other two. To correct $x_1$, subtract the corresponding coset leader 0001 so you get $0110 - 0001 = 0112$.

In the case of $x_2$, the coset leader is 0100 and the corrected word is $1111 - 0100 = 1011$. 
It is not always the case that every coset has a unique coset leader. Look at \( G = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} \) in \( \mathbb{Z}_2^5 \) and the related control matrix \( H = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \)

we have \( 2^3 = 8 \) syndromes

000 001 010 011 100 101 110 111

000 corresponds to code words and five other syndromes correspond to columns of \( H \) and therefore have a unique coset leader (since all columns of \( H \) are different in this example). However, the syndromes 110 and 111 do not occur as columns so we check if they are a linear combination of two columns. They both are, but not in a unique way. For ex 110 is \( \text{col} 3 + \text{col} 4 \) but also \( \text{col} 1 + \text{col} 2 \). Therefore either (00110) or (11000) could be chosen as coset leader. This means we do not really know how to correct words with these syndromes.
Finding the separation from the control matrix

Thm 3.10 A linear code $C$ with control matrix $H$ has separation $\sigma$ if and only if there are $\sigma$ linearly dependent columns in $H$ but any $\sigma-1$ columns are independent.

Proof: Every code word $x$ satisfies $x^t \cdot H = 0$

$\Rightarrow H^t \cdot x = 0$ and hence corresponds to the coefficients of a linear dependence of the columns of $H$. If $x$ is a code word of minimal weight $\sigma$ then exactly $\sigma$ of the coefficients in the dependence are non-zero. If we had a dependence between $\sigma-1$ columns this would correspond to a code word of weight $\leq \sigma-1$, a contradiction.

\[ H = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \]

Columns 2, 4 and 5 are dependent $c_2 - c_4 - c_5 = 0$

(This corresponds to $x = (0, 1, 0, 1, 1, 1) \in C$)

but all pairs of columns are independent.

This shows that the separation is 3 in this case.