Essay on error-correcting linear codes

Jonas Dix

5th January 2024

How should we transmit a message, such that the receiver can correct errors, which happened in the channel? An answer to that question is given by error-correcting linear codes. The idea is to embellish the message with some redundancy, which helps to detect and correct errors in the received message, so that we get the original message back. Application of those codes are given everywhere, where messages are sent. For example in deep space communication, where an electro-magnetic wave from a satelite to earth is sent or, when we hear music from a compact disc, where the message is music sent to our ears.
The following graph shows a communication channel. If the message would get sent directly through the channel without modification, a single error would distort the message, so that it would not be recoverable. Thus the message gets encoded into a codeword, where redundancy is introduced. Then it gets sent through the channel, where noise in form of an error distorts it. Finally in the decoding process small errors can be identified and corrected.

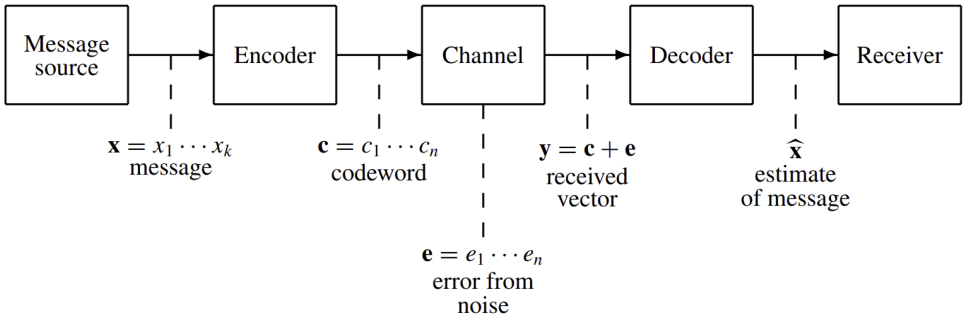

Figure 1: Communication channel

We illustrate this in the following example. We want to send a message $\mathbf{x} \in \mathbb{F}_2^4$ through a noisy channel. We agree a priori that codewords $c \in \mathbb{F}_2^7$ are given by the equations

$$\begin{cases} c_5 = c_1 + c_2 + c_3 \\ c_6 = c_2 + c_3 + c_4 \\ c_7 = c_1 + c_3 + c_4. \end{cases}$$

We thus encode the message $\mathbf{x}$ into the codeword $\mathbf{c} = (x_1, x_2, x_3, x_4, x_1 + x_2 + x_3, x_2 + x_3 + x_4, x_1 + x_3 + x_4)$. If $\mathbf{r} \in \mathbb{F}_2^7$ with possible errors is received, we calculate

$$\begin{cases} s_1 = r_5 - (r_1 + r_2 + r_3) \\ s_2 = r_6 - (r_2 + r_3 + r_4) \\ s_3 = r_7 - (r_1 + r_3 + r_4). \end{cases}$$

Now clearly $\mathbf{r}$ is a codeword if and only if $s = (s_1, s_2, s_3) = 0$. So if this is the case, it looks like $\mathbf{r} = \mathbf{x}$ so that no errors have been made. If for example $s = (1,0,0)$, it seems reasonable, that we have to correct r_5 to $r_5 + 1$ to get $(s_1, s_2, s_3) = 0$. But under which assumptions are we sure, that we correct properly by this proceeding?

We describe this procedure in a general setting and make precise statements about error correction. A linear code is a subspace $C \subset \mathbb{F}_q^n$, where q is a prime number and $\mathbb{F}_q = \mathbb{Z}_q$ is the field with q elements. The number n is the length of the code. If k is the dimension of C, we have $|C| = q^k$ codewords.

A generator matrix G of a k dimensional linear code $C \subset \mathbb{F}_q^n$ is a $k \times n$ matrix, whose rows generate C. If $G = [\mathbb{I}_k A]$, where $\mathbb{I}_k$ is the $k \times k$ identity matrix and A is a $k \times (n-k)$ matrix, then we say that G is of standard form. Having a generator matrix in standard form is always possible by going over to an equivalent code, i.e. an code C' generated by the matrix GPD, where P is a $n \times n$ permutation matrix and D is a $n \times n$ diagonal matrix.

We encode the message $\mathbf{x} \in \mathbb{F}_q^k$ to the codeword $\mathbf{c} = \mathbf{x}G \in C$. If G is of standard form, then the redundancy is in the last $n-k$ coordinates. Through a noisy channel we receive a vector $\mathbf{r} = \mathbf{c} + \mathbf{e} \in \mathbb{F}_q^n$, where $\mathbf{e} \in \mathbb{F}_q^n$ is the unkown error. Now the question appears, whether we can detect that an error happened and if so, whether we can even correct this error. For this we need to measure the error. On $\mathbb{F}_q^n$ we define the metric $d(x,y) := \#\{i \in \{1,\dots,n\} : x_i \neq y_i\}$ and the weight of $x \in \mathbb{F}_q^n$ by $w(x) := d(x,0)$. The minimal distance of C is $d(C) := \min\{w(c) : c \in C \setminus \{0\}\} = \min\{d(c,\hat{c}) : c, \hat{c} \in C, c \neq \hat{c}\}$.

We want to decode by minimal distance decoding, meaning that we want to decode the received $\mathbf{r} \in \mathbb{F}_q^n$ by a $c \in C$ that minimizes $\min\{d(\mathbf{r}, \tilde{c}) : \tilde{c} \in C\}$. If we know a priori that $w(\mathbf{e}) < d$, we immediately know if an error happend or not. We only have to check, if $\mathbf{r}$ is a codeword. If $\mathbf{r} \notin C$, of course an error happened and if $\mathbf{r} \in C$, $d(\mathbf{r}, \mathbf{c}) = w(\mathbf{e}) < d$ implies $\mathbf{r} = \mathbf{c}$. If we even know a priori, that $w(\mathbf{e}) \leq \frac{d-1}{2}$, then we can correct $\mathbf{r}$ to the unique codeword $c \in C$ closest to $\mathbf{r}$. This correction is indeed right, since for a $c \in C$ with minimal distance to $\mathbf{r}$, we have $d(\mathbf{r}, c) \leq d(\mathbf{r}, \mathbf{c}) \leq \frac{d-1}{2}$ and thus get $d(c, \mathbf{c}) \leq d(\mathbf{r}, c) + d(\mathbf{r}, \mathbf{c}) \leq d - 1$, so that $\mathbf{c} = c$, which also implies the stated uniqueness.

Finding the closest codeword can be very expensive. How can we find the error efficiently? This can be done via syndrome decoding, a decoding method that makes use of the linearity of C. If we define $H : \mathbb{F}_q^n \to \mathbb{F}_q^{n-k}$ by mapping C to 0 and $C^\perp$ bijectively to $\mathbb{F}_q^{n-k}$, we get a matrix $H \in \mathbb{F}_q^{(n-k)\times n}$, called parity-check matrix of C. If G has standard form, i.e. $G = [\mathbb{I}_k A]$, then it can be chosen as $H = [-A^T \mathbb{I}_{n-k}]$. For $x \in \mathbb{F}_q^n$ we set $\mathrm{syn}(x) = Hx$. Since $\ker(\mathrm{syn}) = C$, we have $\mathrm{syn}(x) = 0$ if and only if $x \in C$. We get a bijection between syndromes and cosets of $\mathbb{F}_q^n$ w.r.t C by $x + C \mapsto \mathrm{syn}(x)$. We say that $x \in y + C$ is (not necessarily unique) coset leader of $\mathrm{syn}(y)$ if $w(x)$ minimizes $\{w(z) : z \in y + C\} = \min\{w(z) : \mathrm{syn}(z) = \mathrm{syn}(y)\}$. Now the error is in fact given by the unique coset leader of $\mathrm{syn}(r)$, of course again under the assumption, that the weight of the error is less or equal than $\frac{d-1}{2}$. To see that let $e \in \mathbb{F}_q^n$ be coset leader of $\mathrm{syn}(\mathbf{r})$. We have $\mathrm{syn}(\mathbf{r} - e) = 0$ and get for all $c \in C$ that

$$d(\mathbf{r} - e, \mathbf{r}) = w(-e) = w(e) \leq w(\mathbf{r} - c) = d(c, \mathbf{r}),$$

since $\mathrm{syn}(\mathbf{r} - c) = \mathrm{syn}(\mathbf{r})$. Thus $\mathbf{r} - e$ is the codeword closest to $\mathbf{r}$, so indeed $\mathbf{c} = \mathbf{r} - e$, i.e. $\mathbf{e} = e$.

We thus have to find all coset leaders with weight smaller or equal than $\frac{d-1}{2}$ to be able to decode all received messages $\mathbf{r}$ with error satisfying $w(\mathbf{e}) \leq \frac{d-1}{2}$. They can be found by listing all vectors of $\mathbb{F}_q^n$ with weight less or equal $\frac{d-1}{2}$.

Now we want to apply this theory to our example. The linear code $C \subset \mathbb{F}_2^7$ is given by the

generator matrix G with parity-check matrix H, where

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}; \quad H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

We have $d(C) = 3$. It is clear that there are exactly 7 cosets with coset leader of weight 1, namely with the coset leaders $(1,0,0,0,0,0,0), \ldots, (0,0,0,0,0,0,1)$. The message $\mathbf{x} = (1,0,0,1)$ is encoded as $\mathbf{x}G = (1,0,0,1,1,1,0)$. If the vector $\mathbf{r} = (1,0,0,0,1,1,0)$ is received, we calculate $\text{syn}(\mathbf{r}) = (0,1,1)$, which has the coset leader $e = (0,0,0,1,0,0,0)$. Thus we correct $\mathbf{r}$ to $\mathbf{r} - e = (1,0,0,1,1,1,0)$, which indeed corresponds to the original message $\mathbf{x} = (1,0,0,1)$. We see $e = \mathbf{e}$, which was also guaranteed due to the fact that $w(\mathbf{e}) \leq 1 = \frac{3-1}{2}$.