EECS 121: Coding for Digital Communication and Beyond	**Fall 2013**

Lecture 7 — September 24

Lecturer: Anant Sahai *Scribe: Ankush Gupta*

7.1 Overview

This lecture introduces ***affine and linear codes***. Orthogonal signalling and random codes are reviewed first. Impracticality of random codes (exponential table size in the number of bits) motivates affine codes. Linear codes then fall through naturally from the discussion of affine codes.

7.2 Review

7.2.1 Orthogonal Signalling

Consider the problem of communicating k bits simultaneously. In *orthogonal signalling*, each of the 2^k combinations is encoded as a function $x(t)$ from the family of orthogonal cosine functions. The received signal is decoded by passing it through a bank of correlators followed by thresholding the correlation value. This process can be made more robust by adding an additional layer of coding, such as Reed-Solomon error correcting codes.

This process of communication is summarized in the following schematic:

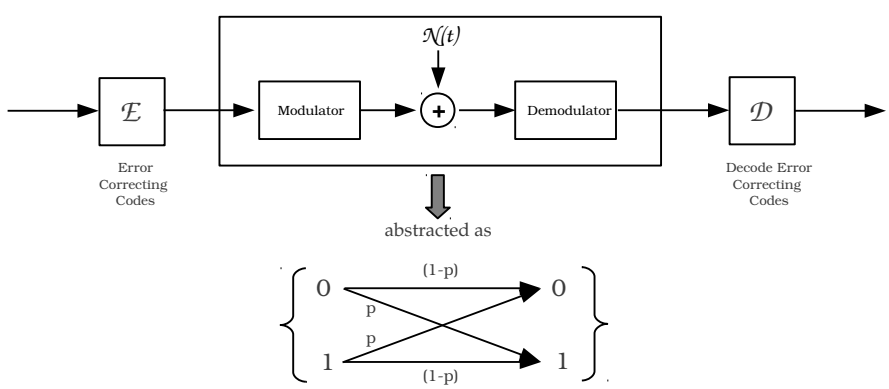

Figure 7.1: Communication through a noisy channel and its abstractions.

7.2.2 Random Codes

A limitation of orthogonal signalling is that it uses a lot of bandwidth — to transmit k bits simultaneously it requires a bandwidth of 2^k frequencies. This is due to the requirement of using orthogonal functions to represent symbols. These orthogonal functions form a sub-space; as the number of orthogonal vectors in a subspace is equal to its dimension, this means, we are embedding 2^k symbols in a 2^k dimensional space, thereby assigning one dimension to each symbol. By relaxing the requirement of perfect orthogonality and instead using approximate orthogonality, these symbols can be embedded in a much lower dimensional space. This is achieved using ***random codes***.

Encoding/ Decoding

In random codes, each of the 2^k symbols is encoded as an i.i.d. $Bernoulli(\frac{1}{2})$ codeword $\vec{X} \in \{0,1\}^n$ of length $n > k$. During transmission, a codeword is corrupted by i.i.d. $Bernoulli(p)$ noise $\vec{N} \in \{0,1\}^n$. The noise is assumed to be independent of the codebook, i.e., the table of the symbols and their corresponding codewords.[1] Then, the received signal $\vec{Y}$ can be represented as $\vec{Y} = \vec{X} + \vec{N}$. The received signal $\vec{Y}$ can be decoded to the maximum matching code-word.[2]

Probability of Error

The noise moves a codeword around in $\{0,1\}^n$. Let 'decoding box' of a codeword be the region in $\{0,1\}^n$, such that if the received signal $\vec{Y}$ lies in that region, it is decoded to that codeword. An error in decoding occurs if :

- Noise pushes $\vec{Y}$ out of the decoding box of the true codeword $\vec{X}_{true}$.
- Noise pushes $\vec{Y}$ into the decoding box of some of false codeword $\vec{X}_{false}$.

Let us formally define the decoding-box of a codeword $\vec{X}$ as the set,

$$\text{Decoding-Box}(\vec{X}) = \{\vec{y} \in \{0,1\}^n | \#\text{of 1's in } \vec{y} - \vec{X} \leq (p+\epsilon)n\}$$

where, p is the probability of noise and n is the length of the codeword. Note, all the codewords are identically distributed (as they are all $\{Bernoulli(1/2)\}^n$. Therefore, setting $\vec{X}$ to $\vec{0}$ above, we have a definition of a decoding-box, independent of the codeword:

$$\text{Decoding-Box} = \{\vec{b} \in \{0,1\}^n | \#\text{of 1's in } \vec{b} \leq (p+\epsilon)n\}$$

[1] This is a valid assumption as the noise originates in nature and should be independent of our choice of the bit-string

[2] The distance (or cost) function for this matching is like the hamming distance, $dist(\vec{Y}, \vec{X}) = \sum_{i=1}^{n} \delta(Y_i - X_i)$, where $\delta(.)$ is the kronecker delta function.

Let us now examine each of the above two possibilities of errors:

1. **Error due to $\vec{Y}$ out of the decoding box of the true codeword $\vec{X}_{true}$**

$$
\begin{aligned}
P(\vec{Y} \notin \text{Decoding-box of } \vec{X}_{true}) &= P(\# \text{ of } 1's \text{ in noise } \vec{n} \geq (p+\epsilon)n) \\
&= \sum_{i=(p+\epsilon)n}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\
&\approx Q\left(\frac{n(p+\epsilon) - np}{\sqrt{np(1-p)}}\right) && \text{central limit theorem} \\
&= Q\left(\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \leq exp\left(-n\frac{\epsilon^2}{2p(1-p)}\right)
\end{aligned}
$$

We note that this probability decays exponentially with n.

2. **Error due to $\vec{Y}$ being in the decoding box of some of false codeword $\vec{X}_{false}$.**

We first note that any two codewords, $\vec{X}_1$ and $\vec{X}_2$ are independent : this is because they are made up of i.i.d. $Bernoulli(\frac{1}{2})$ random entries. Further, the noise is i.i.d. $Bernoulli(p)$ and independent of the code-book. Hence, a received signal $\vec{Y} = \vec{X} + \vec{N}$, is independent of all the false code-words $\vec{X}_{false}$: this is because, $\vec{Y}$ is a function of two variables which are independent of $\vec{X}_{false}$. Moreover, the marginal distribution of the received signal is uniform in $\{0,1\}^n$, this is because it is a sum of noise and $\{Bernoulli(\frac{1}{2})\}^n$ codeword.[3]

Then the probability that a received signal $\vec{Y}$ lies in some false codeword's decoding box is the ratio of the number of vectors in this decoding box and the total number of vectors in $\{0,1\}^n$:

$$
P(\vec{X}_{false} \text{ claims } \vec{Y}) = \frac{|\text{Decoding Box}(\vec{X}_{false})|}{2^n} \qquad \text{by independence of } \vec{Y} \text{ and } X_{false} \text{ and, uniformity of } \vec{Y} \text{ in } \{0,1\}^n
$$

$$
\Rightarrow P(\exists \vec{X}_{false} \text{ claims } \vec{Y}) \leq 2^k \frac{|\text{Decoding Box}|}{2^n} \qquad \text{by identically-distributed and union bound}
$$

Note, above the number of false codewords is actually $2^k - 1$, but 2^k is used for simplicity.

One way of knowing the size of the decoding-box is to use Asymptotic Equipartition Principle.[4] To apply A.E.P., we need a slight modification in our definition of the

[3] Proof in the Appendix.

[4] The other way is to use Stirling's approximation for factorials and evaluate the sum of binomial distribution.

typical-set (the decoding box):

$$\begin{aligned}\text{Decoding-Box} &= \{\vec{b} \in \{0,1\}^n | \#\text{of 1's in } \vec{b} \leq (p+\epsilon)n \text{ and } \#\text{of 1's in } \vec{b} \geq (p-\epsilon)n\} \\ &= \{\vec{b} \in \{0,1\}^n | \quad |(\#\text{of 1's in } \vec{b}) - pn| \leq n\epsilon\}\end{aligned}$$

The probability of bit-string in the typical-set is basically the same and is equal to $(1-p)^{\#1's}p^{\#0's}$, and they all have approximately the same number of ones $= np$. The size of the decoding box according to A.E.P. is then,

$$|\text{Decoding-Box}| \leq 2^{n(H(p)+\epsilon')}$$

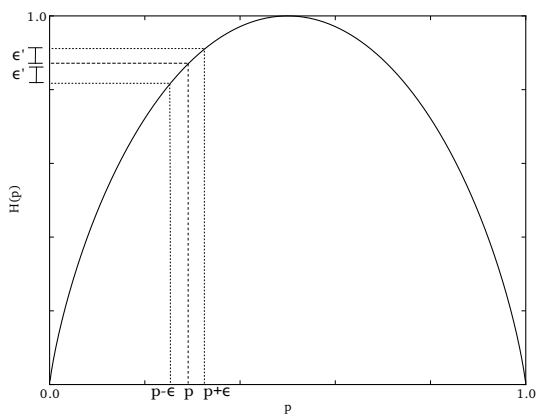

Figure 7.2: Entropy $H(p) = p\log(\frac{1}{p}) + (1-p)\log(\frac{1}{(1-p)})$

where, $H(p) = p\log_2(1/p) + (1-p)\log_2(1/(1-p))$ is the entropy of the noise and, ϵ' is such that $\epsilon' \to 0$ as $\epsilon \to 0$ (see Figure 7.2). Hence, we have,

$$\begin{aligned}P(\exists \vec{X}_{false} \text{ claims } \vec{Y}) &\leq 2^{k-n+n(H(p)+\epsilon')} \\ &= 2^{-n(1-(H(p)+\epsilon')-R)}\end{aligned}$$

if $k = Rn$, R is the channel rate. For exponential decay of error probability with codeword length, we require, $R < 1 - (H(p) + \epsilon')$. Note, the noise probability bounds the channel rate.

From the analysis of the two cases of decoding error, we conclude that the probability of error in decoding is a decaying exponential function in the length of the codeword.

Towards Affine Codes

Even though random codes remedy the problem of using too much bandwidth, yet they suffer from a practical problem — the size of the code-book. For each of the 2^k codes, there is a corresponding n length codeword, hence the size of this codebook is $n \times 2^k$; this is exponential in the number of bits (k). This is further remedied by using ***affine codes***.

7.3 Affine Codes

7.3.1 Encoding

Consider the following coding scheme : a message vector $\vec{m} \in \{0,1\}^k$, is coded as

$$\vec{X} = G\vec{m} + \vec{b} \tag{7.1}$$

where, $G \in \{0,1\}^{n\times k}, \vec{b} \in \{0,1\}^n$. The entries of the matrix G, called the *generator matrix*, and the vector $\vec{b}$ are i.i.d. $Bernoulli(\frac{1}{2})$. Note, this is a very compact representation of the codebook — it requires only $n \times k$ entries for G and n entries for $\vec{b}$, i.e., a total of $Rn^2 + n$ entries, if $k = Rn, R < 1$.

7.3.2 Probability of Error in Decoding

Let us now consider the probability of error in decoding. The error probability for random codes decayed exponentially with the length of the codeword. We would like to have similar error probability characteristics for affine codes. The error analysis for random codes rests on the following three important independence assumptions:

1. Noise is i.i.d. and independent of the codebook.
2. Codewords are marginally (individually) distributed uniformly over $\{0,1\}^n$.
3. Codewords are independent hence, the received signal $\vec{Y} = \vec{X} + \vec{N}$ is independent of all false codewords $\vec{X}_{false}$. In random coding, we have full mutual independence among all codewords, but the analysis only requires pairwise independence among the codewords.

Let us inspect if the above three assumptions hold for affine codes.

1. The first assumption holds true for affine codes as the noise originates in nature and should be independent of what coding scheme is used.
2. $\vec{X} = G\vec{m} + \vec{b}$, i.e., $\vec{X}$ is a sum where one of the terms viz., $\vec{b}$ is $\{Bernoulli(\frac{1}{2})\}^n$. But the sum of any random variable and a $Bernoulli(\frac{1}{2})$ random variable is $Bernoulli(\frac{1}{2})$.[5]

[5] See Appendix for a proof of this.

Hence, $\vec{X} \sim \{Bernoulli(\frac{1}{2})\}^n$, i.e. $\vec{X}$ is uniformly distributed in $\{0,1\}^n$.

3. $\vec{X}_1, \vec{X}_2$ are (pairwise) independent where $\vec{X}_1 = G\vec{m}_1 + \vec{b}$ and $\vec{X}_2 = G\vec{m}_2 + \vec{b}$ for some $\vec{m}_1, \vec{m}_1 \in \{0,1\}^k, \vec{m}_1 \neq \vec{m}_1$.

   Proof : Let $G = [\vec{g}_1\ \vec{g}_2\ \ldots\ \vec{g}_k]$, where $\forall i \in \{1,2,\ldots,k\}, \vec{g}_i \in \{0,1\}^n \sim \{Bernoulli(\frac{1}{2})\}^n$

   Further, let $\vec{m}_1 = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}$ and $\vec{m}_2 = \begin{bmatrix} \tilde{b}_1 \\ \vdots \\ \tilde{b}_k \end{bmatrix}$

   Define $S = \{i | i \in 1,\ldots,k, b_i = \tilde{b}_i\}$, i.e. set of indices where the bits of $\vec{m}_1$ and $\vec{m}_2$ agree and similarly, $D = \{i | i \in 1,\ldots,k, b_i \neq \tilde{b}_i\}$, i.e. set of indices where the bits are different. Then, we know:

   $\vec{X}_1 = \vec{b} + \sum_{i=1}^k b_i \vec{g}_i$, and

   $\vec{X}_2 = \vec{b} + \sum_{i=1}^k \tilde{b}_i \vec{g}_i = \vec{b} + \sum_{i \in S} b_i \vec{g}_i + \sum_{i \in D} \tilde{b}_i \vec{g}_i$

   But note that $\forall i \in D$, $\tilde{b}_i \vec{g}_i = (b_i + 1)\vec{g}_i = b_i \vec{g}_i + \vec{g}_i$, because, b_i and $\tilde{b}_i$ are complements. Therefore,

   $\vec{X}_2 = \vec{b} + \sum_{i \in S} b_i \vec{g}_i + \sum_{i \in D} b_i \vec{g}_i + \sum_{i \in D} \vec{g}_i$

   $\Rightarrow \boxed{\vec{X}_1 = \vec{X}_2 + \sum_{i \in D} \vec{g}_i}$

   Without loss of generality, there must $\exists j \in D$, s.t. $\vec{g}_j$ is not included in $\vec{X}_2$, i.e., $\tilde{b}_j = 0$ (if not, since we know $b_j = \tilde{b}_j + 1$, hence, this must be the case for $\vec{X}_1$). Therefore, $\vec{g}_j$ and $\vec{X}_2$ are independent.

   Hence, $\vec{X}_1$ is a sum of $\vec{X}_2$ and some $\vec{g}_j \in \{Bernoulli(\frac{1}{2})\}^n$ independent of $\vec{X}_2$. Therefore, $\vec{X}_1$ and $\vec{X}_2$ are independent.[6] $\square$

As all three independence assumptions hold true for affine codes, the same analysis of probability of error applies as the one used for random codes. *Therefore even for affine codes, the probability of error decays exponentially with the length of codewords.*

[6] The fact that $X = Y + B$ is independent of $Y \sim Bernoulli(p)$, if $B \sim Bernoulli(\frac{1}{2})$ and independent of Y is also used in point (2.) and proved in the Appendix.

7.3.3 Decoding and Linear Codes

Now consider the decoder. A message $\vec{m}$ was encoded as $\vec{X} = G\vec{m} + \vec{b}$. The received signal is $\vec{Y} = \vec{X} + \vec{N} = G\vec{m} + \vec{b} + \vec{N}$, where $\vec{N}$ is noise $\sim \{Bernoulli(p)\}^n$ (i.i.d.).

Hence, the decoder might first subtract $\vec{b}$ from $\vec{Y}$ and then, correlate $\vec{Y} - \vec{b} = G\vec{m} + \vec{N}$ with all $G\vec{m}_i$. But, subtraction in $\mathbb{Z}^2$ is the same as addition.[7] Hence, we have the following diagram for affine codes:

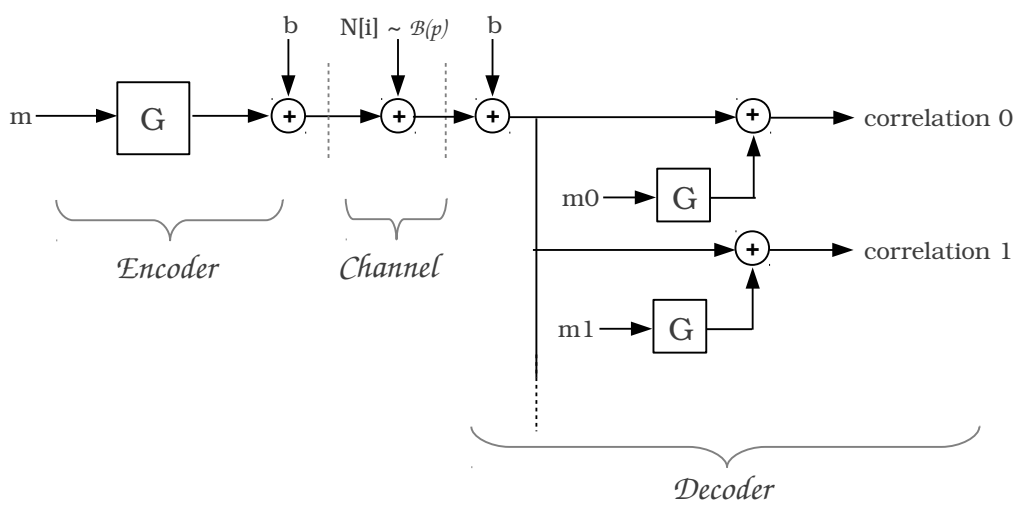

Figure 7.3: Encoding and decoding with affine codes

But note that the noise $\vec{N}$ is additive and independent of the signal $\vec{X}$. Hence, in the above encoding/ decoding scheme the addition of noise and the addition of vector $\vec{b}$ can commute. Note, had the noise not been additive or dependent on the signal then it would be incorrect to change the order of application of the noise and addition of $\vec{b}$.

So the equivalent encoding/decoding scheme now looks like the following:

Therefore, we can consider the equivalent encoding scheme to be : $\vec{m} \mapsto G\vec{m} + \vec{b} + \vec{b}$. But, in $\mathbb{Z}^2$, we have, $G\vec{m} + \vec{b} + \vec{b} = G\vec{m} + 2\vec{b} \texttt{ mod } 2 = G\vec{m}$. Therefore, the new encoding is :

$$\vec{m} \mapsto G\vec{m}, G \in \{Bernoulli(\frac{1}{2})\}^{n\times k} \tag{7.2}$$

We note that the above encoding scheme is a linear function of the message $\vec{m}$. Hence, ***linear codes*** suffice.[8] Linear codes inherit the same exponentially decaying probability

[7] $x - y \bmod 2 = x + (-1)y \bmod 2 = x + y \bmod 2$.

[8] The affine part $\vec{b}$ is used only in the proof (see independence assumption 2); it is not required in a practical system with independent additive noise.

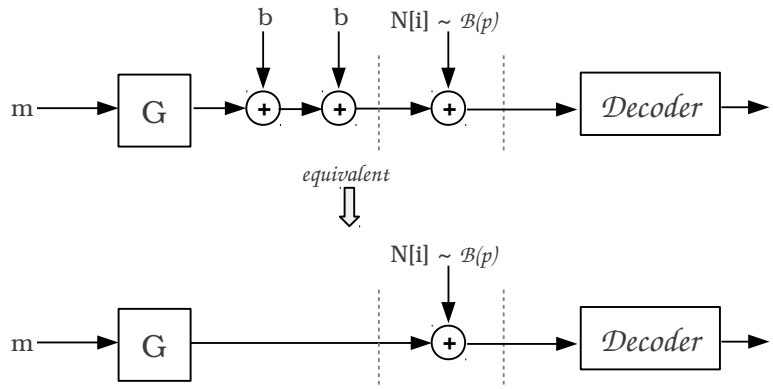

Figure 7.4: Affine Codes to Linear Codes : Equivalent system with order of $\vec{b}$ and $\vec{N}$ switched.

of error with n — the length of the codeword and have a compact representation of the codebook (Rn^2 terms)!

7.3.4 Further Explorations

We have proved that the probability of error in decoding decays exponentially with the length of the codeword. But the number of codewords also increases exponentially with the length. Hence, one might ask if there exist some messages for a given coding strategy (G and $\vec{b}$), for which the probability of error is large.

We first note that codewords form a sub-space of $\{0,1\}^n$. Sub-spaces are closed under addition, i.e., adding two vectors in a sub-space results in another which is also in the sub-space. Further, addition does the 'same thing' to all the vectors — it is not biased against any special vector of the sub-space. Now, if there existed a particular message $\vec{m}$ for which probability of error was large, that means that the noise when added to the encoded $\vec{m}$ takes it closer to another codeword or out of the sub-space. But, then the noise must do the same thing to all the codewords! Hence, no such exceptionally bad message $\vec{m}$ should exist.

It is important to note that if the sub-space had some 'boundaries', i.e., if it was not closed under addition then there could exist some messages which could have large probabilities of error.

The symmetry of codewords discussed above is used when analyzing communication systems. As no codeword is special, in our analysis of probability of erroneous decoding, we can pick a particular codeword, say the all zero codeword, and analyze the error probability of that. The results would then generalize to all the codewords.

7.4 Appendix

1. **Claim :** If $Y = X + B$, $X \sim Bernoulli(p)$, $B \sim Bernoulli(\frac{1}{2})$, B and X are independent then:

   (a) $Y \sim Bernoulli(\frac{1}{2})$

   (b) Y is independent of X

   Proof :

   (a)

$$\begin{aligned} P(Y = 1) &= P(X = 0)P(B = 1) + P(X = 1)P(B = 0) \\ &= (1 - p)\frac{1}{2} + p\frac{1}{2} \\ &= \frac{1}{2} \end{aligned}$$

   $\Rightarrow P(Y = 1) = P(Y = 0) = \frac{1}{2}$. Hence, $Y \sim Bernoulli(\frac{1}{2})$.
   Trivially, the above claim generalized to vectors with independent random-variable enteries.

   (b) $P(Y = y, X = x) = P(Y = y|X = x)P(X = x) = P(B = y - x)P(X = x) = \frac{1}{2}P(X = x) = P(Y = y)P(X = x) \Rightarrow Y$and X are independent.