Iterative Error Correction

Channel, Codes, and Capacity

Isha Chaudhary

July 2020

Outline

1. Channel
   - Terminology
   - Types
   - Log-Likelihood Ratio
2. Entropy, mutual information and Capacity
   - Entropy
   - Mutual Information
   - Capacity
   - Shannon's Coding Theorem
3. Codes and Decoding
   - ML Decoding
   - MAP Decoding
   - Iterative decoding
   - Performance Measures

Important Terms

- **Symmetric:** If the 0 and 1 bits get equally corrupted by the channel (with the same probability).
- **Memoryless:** If the channel output at any instant of time depends on the input at that time.

Common types of channels

All the channels considered here are Symmetric Channels. All have binary input bits (only these are considered and required for the types of codes to be considered.)

1. **Binary Symmetric Channel:** This channel has the property to flip bits, to introduce error in the transmitted codeword. Due to symmetricity, the probability of flipping the bits either way is same, p, and generally for good channels, p ¡ 0.5.
2. **Binary Erasure Channel:** This channel introduces erasures in the codeword transmitted, with same probability of converting either the 0 bit or the 1 bit to an erasure, for an equi-probable transmission.
3. **Binary Input - Additive White Gaussian Noise Channel:** This channel adds White gaussian noise to each bit of the input to give a received vector which has real numbered bits.

Log-Likelihood Ratios (LLR)

$$L(x) = log \frac{p(x=0)}{p(x=1)}$$

1. *Sign of L(x)* provides a hard decision on x.
2. *Magnitude of L(x)* is the reliability of the hard decision on x.

Received LLRs

The Log-likelihood ratio of the transmitted codeword bits, given the received codeword bits is the Received LLR, R_i. y is the received vector and c is the codeword to which y corresponds. x is the BPSK modulated vector, sent out into the channel.

$$\begin{aligned} R_i = L(x_i|y_i) &= log\frac{p(c_i=0|y_i)}{p(c_i=1|y_i)} \\ &= log\frac{p(y_i|x_i=1)p(x_i=1)}{p(y_i|x_i=-1)p(x_i=-1)} \end{aligned} \tag{1}$$

R_i, which is the LLR value for the bit c_i, provide for soft decision for c_i. When $R_i > 0$, $x_i = 1 \iff c_i = 0$ and when $R_i < 0$, $x_i = -1 \iff c_i = 1$, is a hard decision for c_i.

LLR for Binary Symmetric channel

Given the probabilities $p(x|y)$ for the BSC:

$$\begin{cases} p(x_i = 1|y_i) = 1 - \epsilon \ and \ p(x_i = 0|y_i) = \epsilon & if \ \ y_i = 1 \\ p(x_i = 1|y_i) = \epsilon \ and \ p(x_i = 0|y_i) = 1 - \epsilon & if \ \ y_i = 0 \end{cases} \tag{2}$$

the received LLRs for the i^{th} transmitted bit are:

$$R_i = L(x_i|y_i) = log \frac{p(x_i = 0|y_i)}{p(x_i = 1|y_i)} = \begin{cases} log \epsilon/(1 - \epsilon) & y_i = 1 \\ log(1 - \epsilon)/\epsilon & y_i = 0 \end{cases} \tag{3}$$

LLR for Binary Erasure channel

For an equiprobable channel, where the probability of the transmission of the

$$
\begin{gathered}
p(x = 0 \mid y = 0) = 1 \\
\mathrm{p}(\mathrm{x} = 0| \; y = 1) = 0 \\
\mathrm{p}(\mathrm{x} = 1| \; y = 0) = 0 \\
\mathrm{p}(\mathrm{x} = 1| \; y = 1) = 1 \\
\mathrm{p}(\mathrm{x} \mid y = e) = 0.5
\end{gathered}
$$

So the received LLRs are:

$$
R_i = L(x_i|y_i) = log\frac{p(x_i = 0|y_i)}{p(x_i = 1|y_i)} = \begin{cases} log\frac{0}{1} = -\inf & y_i = 1 \\ log\frac{1}{0} = \inf & y_i = 0 \\ log\frac{0.5}{0.5} = 0 & y_i = e \end{cases} \tag{4}
$$

LLR for Binary Input - AWGN channel

Here $y_i = \mu x_i + z_i$, where z_i is the Gaussian noise added in the input bit and x_i is the i^{th} transmitted symbol belonging to [-1, +1].

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma^2}$$

$$R_i = \frac{2\mu}{\sigma^2} y_i$$

Considering the relative noise level of a BI-AWGN channel, it is convenient to assume that $\mu = 1$ and adjust σ to reflect the noise quality of channel.

$$\implies R_i = \frac{2}{\sigma^2} y_i$$

LLR for BI-AWGN using SNR

For a codeword with rate r, which is getting transmitted through a BI-AWGN channel, the noise level can be expressed as:

$$\frac{1}{r}\frac{\mu^2}{2\sigma^2}(noise\ \ level)\ = \frac{1}{r}\frac{E_s}{N_0} = \frac{E_b}{N_0}(SNR)$$

So the received LLR can be given by:

$$R_i = L(x_i|y_i) = 4\frac{\sqrt{E_s}}{N_0}y_i = 4\frac{\sqrt{rE_b}}{N_0}y_i$$

When μ is taken to be 1,

$$R_i = \frac{4}{N_0}y_i$$

- Signal-to-Noise (SNR) ratio can be expressed in dB as

$$\frac{E_b}{N_0}(dB) = 10log_{10}\frac{E_b}{N_0} = 10log_{10}\frac{\mu^2}{2r\sigma^2}$$

Entropy I

- A measure for the information produced when a symbol with probability of occurrence p is received is given by:

$$I(p) = A.log_2 \frac{1}{p}$$

  where $A > 0$ and A is chosen so as to equate one binary unit to the information received from one symbol of a binary source, when both symbols are equiprobable.

$$I(0.5) = A \implies I(0.5) = 1 \; when \; A = 1$$

Entropy II

- The information content of a random variable, which is the average information over all its symbols, is its entropy, H(x). For discrete random variable,

$$H(x) = E[I(p(x))] = \sum_{j=1}^{q} p_j I(p_j) = \sum_{j=1}^{q} p_j log_2 \frac{1}{p_j} = -\sum_{j=0}^{q} p_j log_2 p_j$$

  This is when we assume that the emission of symbols is independent of time.

- For continuous random variable, we define the differential entropy as

$$H(x) = E[I(p(x))] = -\int p(x) log_2 p(x) dx$$

Entropy III

- Joint entropy of two random discrete variables, X and Y:

$$H(X, Y) = \sum p(x, y) log_2 \frac{1}{p(x, y)}$$

- Conditional entropy is the amount of uncertainty about X remaining after Y is known.

$$H(X|Y) = \sum p(x|y) log_2 \frac{1}{p(x|y)}$$

Mutual Information

Mutual Information, I(X;Y) quantifies the information common to both the random variables, X and Y.

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

It measures the amount of uncertainty about X that is removed by knowing Y, or the amount of information about X that can be transmitted through the channel.

Capacity of a channel

$$Capacity = \max_{p(x)} I(X;Y)$$

It defines the maximum amount of information that can be conveyed through the channel per transmitted symbol, over all possible input distributions.

Capacity of different channels

1. **BSC:** For the probability of flipping being ϵ,

$$c(\epsilon) = 1 - H_2(\epsilon)$$

   where $H_2(\epsilon) = H(Y|X=0) = H(Y|X=1)$

2. **BEC:** For the probability of the occurrence of erasure being ϵ,

$$c(\epsilon) = 1 - \epsilon$$

3. **BI-AWGN:** For σ being the standard deviation of the noise in the channel,

$$\begin{aligned} c(\sigma) &= H(Y) - H(Z) \\ &= -\int_{-\infty}^{\infty} \phi_\sigma(y).log_2\phi_\sigma(y)dy - \frac{1}{2}.log_2 2\pi e\sigma^2 \end{aligned} \tag{5}$$

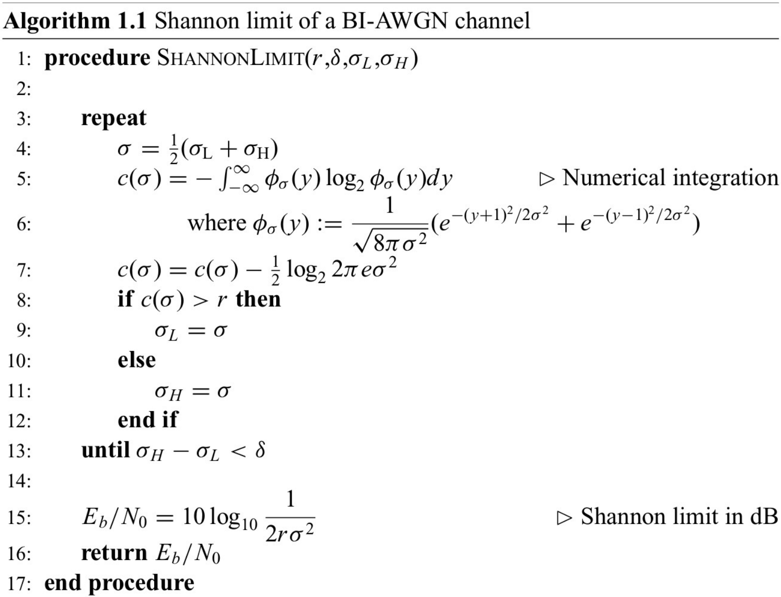
   where Z is the random variable denoting the added noise and

$$\phi_\sigma(y) = \frac{1}{\sqrt{8\pi\sigma^2}}\left(exp\left(-\frac{(y+1)^2}{2\sigma^2}\right) + exp\left(-\frac{(y-1)^2}{2\sigma^2}\right)\right)$$

Shannon's Coding Theorem

- Provided the code rate of transmission, r is less than the channel's capacity, there exits an error correction code that will achieve an arbitrarily low probability of error, despite the noise added by the channel.
- For a channel with noise level parameter x and an error correction code r, the noise level x_{Sh}, such that $r = c(x_{Sh})$ is a threshold for error correction codes with that rate. Shannon's noisy channel coding theorem says that for any noise level x below x_{Sh} there exists a rate-r code that can achieve an arbitrarily low probability of error, whereas for any noise level above x_{Sh}, no rate-r code can achieve an arbitrarily low probability of error.
- Here x_{Sh} is the *Shannon's Limit*.

Algorithm for Shannon Limit



Algorithm 1.1 Shannon limit of a BI-AWGN channel

```
procedure ShannonLimit($r$,$\delta$,$\sigma_L$,$\sigma_H$)

    repeat
        $\sigma = \frac{1}{2}(\sigma_{\mathrm{L}} + \sigma_{\mathrm{H}})$
        $c(\sigma) = -\int_{-\infty}^{\infty} \phi_\sigma(y) \log_2 \phi_\sigma(y) dy$          ▷ Numerical integration
                where $\phi_\sigma(y) := \frac{1}{\sqrt{8\pi\sigma^2}}(e^{-(y+1)^2/2\sigma^2} + e^{-(y-1)^2/2\sigma^2})$
        $c(\sigma) = c(\sigma) - \frac{1}{2}\log_2 2\pi e\sigma^2$
        if $c(\sigma) > r$ then
            $\sigma_L = \sigma$
        else
            $\sigma_H = \sigma$
        end if
    until $\sigma_H - \sigma_L < \delta$

    $E_b/N_0 = 10\log_{10}\frac{1}{2r\sigma^2}$          ▷ Shannon limit in dB
    return $E_b/N_0$
end procedure
```

Algorithm for Shannon Limit

- We require to find x_{Sh}, such that $c(x_{Sh}) = r$.
- The above pseudo-code is for the BI-AWGN channel.
- We are searching for σ_{Sh} between σ_L and σ_H, and thus finding the value of the noise level or the SNR.

$$\left(\frac{E_b}{N_0}\right)_{Sh} = 10.log_{10}\left(\frac{1}{2r\sigma_{Sh}^2}\right)$$

Maximum Likelihood Decoding

- This decoding scheme is used when no prior knowledge of the transmitted codeword is given.

$$\hat{\mathbf{c}} = argmax_{\mathbf{c} \in C}\, p(\mathbf{y}|\mathbf{c})$$

  where, assuming a memoryless channel, $p(\mathbf{y}|\mathbf{c}) = \prod_{i=1}^{N} p(y_i|c_i)$

- In general, for a code with d_{min} as the minimum distance of the code, t bit flips can be corrected by choosing the closest codeword, whenever $t \leq \lfloor (d_{min} - 1)/2 \rfloor$.
- We decode to the codeword which is closest to the received vector, either in terms of the hamming distance (for binary output symbols) or euclidean distance (for real output symbols).

Maximum a posteriori decoding

- $\hat{\mathbf{c}} = argmax_{\mathbf{c} \in C} p(\mathbf{c}|\mathbf{y})$
- If the decoder has a priori information about **c**, then MAP decoder will choose the most probable codeword after taking the extra information into account.
- If each codeword is equally likely to be sent, then ML and MAP decoding will return an identical result.
- Symbol-wise MAP decoding:

$$\hat{c}_i = argmax_{c_i \in 0,1} p(c_i|\mathbf{y})$$

The BCJR algorithm is used for the symbol-wise MAP decoding.

Iterative decoding

- In iterative error correction codes, the decoding proceeds in an iterative manner to produce accurate estimates of $p(c_i|y_i)$ using repeated low-complexity processes.
- These algorithms are generally not optimal, but can come really close to the ML or MAP performance.
- The code properties which are beneficial for ML or MAP decoding, also apply for iterative decoding.

Performance Measures

1. **Word Error Rate** or Block Error Rate: The number of times the decoder chooses the wrong codeword as a fraction of the total number of codewords decoded.
2. **Bit Error Rate:**
   - *Codeword:* The number of incorrect codeword bits as a fraction of the total number of codeword bits in all the codewords decoded.
   - *Message:* The number of incorrect message bits as a fraction of the total number of message bits in all the codewords decoded.

   The message BER is usually a more relevant quantity.

To obtain confidence in the calculated BER or WER, around 500 WER should be observed, as a rule of thumb.

Union Bounds

- It states that the total probability of error $P_e(\mathbf{c}_1)$ for c_1 can be upper bounded by the sum of probabilities that it will be decoded incorrectly to any other codeword in the code C:

$$P_e(\mathbf{c}_1) \leq \sum_{\mathbf{c} \in C, \mathbf{c} \neq \mathbf{c}_1} P_e(\mathbf{c}_1, \mathbf{c})$$

- For a binary input memoryless symmetric channel, the probability that c_1 will be selected instead of c_2, depends only on the number of bit locations, d that differ between c_1 and c_2 and not on the location of those differing entries.

Doubt

1. How to relate between ▸ these two interpretations of the Shannon's Noisy Channel Coding Theorem? Are they different, unrelated parts of the original theorem?

References

1. Iterative Error Correction, book by Sarah J. Johnson