The Effect of Non-Typical Codewords on Maximum-Likelihood Decoding

Matthew Beveridge*
mattbev@mit.edu

October, 2019

Abstract

Analysis of the effect of introducing non-typical codewords into a codebook in respect to maximum-likelihood decoding of a channel. Specific application to the *Z-channel*.

1 Introduction

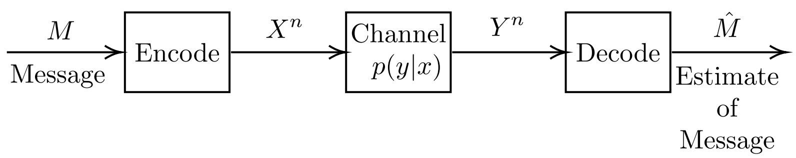

Figure 1: Generic Channel Example

Communication between a sender and a receiver can be modeled as above in Figure 1. It goes as follows:

1) Sender has some message M that they would like to send.

2) M is encoded by the encoder, outputs encoded message X^n.

3) X^n is then passed through the channel which outputs Y^n. The channel is the way the message is actually transmitted. Within it, there is a conditional distribution $p(y|x)$ which indicates the probability of outputting some sequence y given input x. This is because within the channel, there can be perturbations that cause errors to arise within the sequence during the process of transmission.

4) Y^n, the outputted sequence of the channel, is then decoded by the decoder into the outputted message $\hat{M}$.

*Massachusetts Institute of Technology

5) $\hat{M}$ is the best estimate for the inputted message M.

In step 4), there are several choices for decoding method. Going forward, we will assume use of maximum-likelihood decoding.

2 Background

Below follow several definitions that will be useful throughout the paper.

Definition 1 *Entropy of a distribution is defined as:*

$$H(X) = -\sum_{x \in X} p_x \log_2 p_x \tag{1}$$

Definition 2 *Maximum-likelihood (ML) decoding is choosing the closest codeword to the received word [1]. Formally, this is:*

$$D(y) = arg \min \Delta(y, x) \tag{2}$$

Where y is the received message and x is the inputted (encoded) message. This operation entails an exhaustive search over all possible codewords within the codebook in order to determine the minimum distance. ML decoding is the optimal decoding technique.

Definition 3 *A codeword X^n belongs to the typical set (and is thus typical) under the following condition:*

$$A_\epsilon^n := \{X^n : \left| -\frac{\log(P(X^n))}{n} - H(X) \right| < \epsilon\} \qquad \epsilon > 0 \tag{3}$$

Where n is the length of the codeword and X is the distribution from which the codewords are generated. If a codeword is not in this set, then it is non-typical.

In Shannon's proof of Channel Coding Theorem, he proves that for codebooks of typical codewords it is possible to communicate with arbitrarily small probability of error as long as the rate R is less than the channel capacity C. This was proved using jointly-typical decoding (which requires typical codewords) because the proof is simpler. In this paper, we will explore the effect of introducing non-typical codewords in to the codebook with respect to decoding error of maximum-likelihood decoding (which does not require typical codewords).

3 Optimality of Maximum-Likelihood Decoding

As previously stated, we will specifically looking at the *Z-channel.* Below, we look at efficacy of ML decoding in the *Z-channel.*

Lemma 1 *Maximum-likelihood decoding always decodes the output y to the correct input x so that $y = x$ for the* Z-channel. *This holds regardless of the distribution of inputs (i.e. whether the probability of $x = \{0, 1\}$ is uniform or not).*

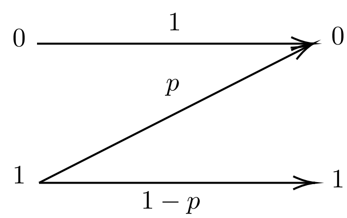

Figure 2: *Z-channel*

PROOF:

Above in Figure 2 is an example channel, the *Z-channel*. It is a binary asymmetric channel, where inputs of 1 have some probability $0 \leq p \leq 1$ of outputting incorrectly to a 0. Thus, this channel has the following probabilities:

$$\begin{aligned}
P(y = 0|x = 0) =& 1 \\
P(y = 1|x = 0) =& 0 \\
P(y = 0|x = 1) =& p \\
P(y = 1|x = 1) =& 1 - p
\end{aligned}$$

Where x is the inputted character and y is the outputted one. It follows that:

$$[P(y = 0|x = 0) = 1] \geq [P(y = 0|x = 1) = p] \tag{4}$$

$$[P(y = 1|x = 0) = 0] \leq [P(y = 1|x = 1) = 1 - p] \tag{5}$$

And ML decoding yields the correct decoded message regardless of the distribution of input 1's and 0's.

QED

4 Effect of Non-Typical Codewords

Considering a codebook of all possible codewords instead of just typical codewords yields the following:

Theorem 1 *The addition of non-typical codewords to the codebook does not affect the efficacy of maximum-likelihood decoding.*

PROOF:

As stated in Definition 3 (Equation 3), a particular codeword is typical if:

$$\begin{aligned}
\left| -\frac{\log(P(X^n))}{n} - H(X) \right| <& \epsilon \\
\left| -\frac{\log(P(X^n = 0))}{n} - H(X) \right| <& \epsilon
\end{aligned}$$

Let $z = P(X^n = 0) = [\textit{probability that input is a zero}]$ be the probability distribution over the inputs to the channel. The inputs are length $n = 1$. Therefore,

$$\begin{aligned}\left|-\frac{\log(z)}{n} - H(X)\right| &< \epsilon \\ \left|-\log(z) - H(X)\right| &< \epsilon\end{aligned}$$

Entropy as defined in Definition 1 reduces the above equation to:

$$\begin{aligned}\left|-\log(z) + [z\log(z) + (1-z)\log(1-z)]\right| &< \epsilon \\ \left|(z-1)\log(z) + (1-z)\log(1-z)\right| &< \epsilon\end{aligned}$$

For the above equation, the LHS is equal to zero if $z = \frac{1}{2}$, which will always make $LHS < \epsilon$ because $\epsilon > 0$. Thus, for $z = \frac{1}{2}$ the channel only has typical codewords as inputs and for any other value of z there may be non-typical codewords in the codebook.

In Lemma 1, we derived that for any distribution of inputs (i.e. any z value) ML decoding is always effective in regards to the *Z-channel*. Therefore, whether $z = \frac{1}{2}$ and the codebook only has typical codewords or $z \neq \frac{1}{2}$ and there could be non-typical codewords, ML decoding will perform with the same efficacy.

QED

5 Notes

The above conclusions in Sections 3 and 4 depend on the assumption that the size of the codebook is $\leq 2^{nR}$ codewords to ensure low probability of decoding error.

References

[1] Guruswami, V, "CSE 533: Error-Correcting Codes - Lecture 4: Proof of Shannon's Theorem and an Explicit Code", University of Washington, 2006.