The Maximum-Likelihood Decoding Performance of Error-Correcting Codes

Henry D. Pfister
ECE Department
Texas A&M University

August 27th, 2007 (rev. 0)
November 21st, 2013 (rev. 1)

1 Performance of Codes

1.1 Notation

- $\mathcal{X}, \mathcal{Y}, \mathcal{S}$ Sets are denoted by calligraphic letters
- X, Y, Z Random variables are denoted by capital letters
- X_i, x_i Single elements of vectors are denoted by a subscript index
- X_i^j, x_i^j The interval subvectors (i.e., $X_i, X_{i+1}, \ldots, X_j$) of a vector
- $\underline{X}, \underline{x}$ Complete vectors are denoted by underlines

1.2 Optimal Decoding Rules

Let $\mathcal{X}$ be an arbitrary alphabet and $\mathcal{C} \subset \mathcal{X}^n$ be a length-n code. Assume a random codeword $X_1^n \in \mathcal{C}$ is chosen with probability $p\left(X_1^n\right)$ and transmitted through a DMC with transition probability $W(y|x)$. For sequences, the conditional probability of the observed sequence $Y_1^n \in \mathcal{Y}^n$ is given by

$$Pr\left(Y_1^n = y_1^n | X_1^n = x_1^n\right) \triangleq W\left(y_1^n | x_1^n\right) = \prod_{i=1}^{n} W\left(y_i | x_i\right).$$

Choosing the codeword $x_1^n \in \mathcal{C}$ which maximizes $Pr(X_1^n = x_1^n | Y_1^n = y_1^n)$ is known as *maximum a posteriori* (MAP) decoding and this minimizes the probability of block error. Using Bayes's rule, we find that

$$Pr\left(X_1^n = x_1^n | Y_1^n = y_1^n\right) = \frac{p\left(x_1^n\right) W\left(y_1^n | x_1^n\right)}{\sum_{\widetilde{x}_1^n \in \mathcal{C}} p\left(\widetilde{x}_1^n\right) W\left(y_1^n | \widetilde{x}_1^n\right)}.$$

Since the denominator is a constant for all x_1^n, we find that

$$\mathcal{D}_{MAP}\left(y_1^n\right) = \arg\max_{x_1^n \in \mathcal{C}} p\left(x_1^n\right) W\left(y_1^n | x_1^n\right).$$

The *maximum likelihood* (ML) decoding rule is defined as

$$\mathcal{D}_{ML}\left(y_1^n\right) = \arg\max_{x_1^n \in \mathcal{C}} W\left(y_1^n | x_1^n\right).$$

Notice that $\mathcal{D}_{MAP}\left(y_1^n\right) = \mathcal{D}_{MAP}\left(y_1^n\right)$ if $p\left(x_1^n\right) = \frac{1}{|\mathcal{C}|}$ is a constant for all $x_1^n \in \mathcal{C}$. Therefore, ML decoding is optimal for equiprobable transmission.

1.3 Maximum Likelihood Decoding

The ML decoding rule implicitly divides the received vectors into decoding regions known as Voronoi regions. The Voronoi region (i.e., decision region) for the codeword $x_1^n \in \mathcal{C}$ is the subset of $\mathcal{Y}^n$ defined by

$$V\left(x_1^n\right) \triangleq \left\{y_1^n \in \mathcal{Y}^n | W\left(y_1^n | x_1^n\right) > W\left(y_1^n | \widetilde{x}_1^n\right) \, \forall \, \widetilde{x}_1^n \in \mathcal{C}, \, \widetilde{x}_1^n \neq x_1^n\right\}.$$

In this case, the average probability of block error is given by

$$P_B = 1 - \sum_{x_1^n \in \mathcal{C}} \frac{1}{|\mathcal{C}|} \sum_{y_1^n \in V\left(x_1^n\right)} W\left(y_1^n | x_1^n\right).$$

It is worth noting that this formula breaks down if ties may occur. This can be rectified by directing the ML decoder to choose a codeword randomly in this case. In this case, the above expression for P_B only gives an upper bound.

1.4 Channel Symmetry and Linear Codes

In many cases, the channel satisfies a symmetry condition that allows us to simplify things. For simplicity, we will assume that $\mathcal{X}$ forms an Abelian group under "+" and that the channel symmetry is defined by $W(y|x+z) = W\left(\pi_z(y)|x\right)$ for a set of $\mathcal{Y}$-permutations π_x indexed by $x \in \mathcal{X}$. Each permutation is a one-to-one mapping $\pi_x : \mathcal{Y} \to \mathcal{Y}$ that satisfies $\pi_{x+z}(y) = \pi_x(\pi_z(y)) = \pi_z(\pi_x(y))$ and therefore the set of permutations forms a group which is isomorphic to $\mathcal{X}$. This type of channel is known as *output symmetric*. We extend this symmetry to length-n sequences by defining

$$W\left(y_1^n | x_1^n + z_1^n\right) = W\left(\pi_{z_1^n}\left(y_1^n\right) | x_1^n\right) \triangleq \prod_{i=1}^{n} W\left(\pi_{z_i}\left(y_i\right) | x_i\right)$$

with $\pi_{z_1^n}(y_1^n) \triangleq (\pi_{z_1}(y_1), \pi_{z_2}(y_2), \ldots, \pi_{z_n}(y_n))$. It is worth noting that this symmetry condition is sufficient to imply that a uniform input distribution achieves the capacity of this DMC.

Example. Consider the BSC where $\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \{0,1\}$, and

$$W(y|x) = \begin{cases} p & \text{if } x \neq y \\ 1-p & \text{if } x = y. \end{cases}$$

Then, $\pi_0(y) = y$ and $\pi_1(y) = \overline{y}$ defines the natural symmetry of the channel.

Example. Consider the binary-input AWGN channel where $\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \mathbb{R}$, and $Y \sim N\left((-1)^x, \sigma^2\right)$. Although this is not a DMC, similar results hold when sums are replaced by integrals. In this case, $\pi_0(y) = y$ and $\pi_1(y) = -y$ defines the natural symmetry of the channel.

If the code is also a group code (i.e., sum of any two codewords is a codeword), then the Voronoi region of any codeword can be written as a transformation of $V(\mathbf{0})$ with

$$\begin{aligned} V\left(\mathbf{0} + x_1^n\right) &= \left\{y_1^n \in \mathcal{Y}^n | W\left(y_1^n | 0 + x_1^n\right) > W\left(y_1^n | \widetilde{x}_1^n\right) \, \forall \, \widetilde{x}_1^n \in \mathcal{C}, \, \widetilde{x}_1^n \neq \mathbf{0} + x_1^n\right\} \\ &= \left\{y_1^n \in \mathcal{Y}^n | W\left(\pi_{x_1^n}\left(y_1^n\right) | \mathbf{0}\right) > W\left(\pi_{x_1^n}\left(y_1^n\right) | \widetilde{x}_1^n - x_1^n\right) \, \forall \, \widetilde{x}_1^n \in \mathcal{C}, \, \widetilde{x}_1^n \neq \mathbf{0} + x_1^n\right\} \\ &= \left\{y_1^n \in \mathcal{Y}^n | W\left(\pi_{x_1^n}\left(y_1^n\right) | \mathbf{0}\right) > W\left(\pi_{x_1^n}\left(y_1^n\right) | \mathbf{0}\right) \, \forall \, z_1^n \in \mathcal{C}, \, z_1^n \neq \mathbf{0}\right\}. \\ &= \pi_{x_1^n}^{-1}\left(V(\mathbf{0})\right). \end{aligned}$$

The last step follows from the fact that $y_1^n \in V\left(x_1^n\right)$ implies that $\pi_{x_1^n}\left(y_1^n\right) \in V\left(\mathbf{0}\right)$. We can also use this to simplify the probability of block error to

$$\begin{aligned} P_B &= 1 - \sum_{x_1^n \in \mathcal{C}} \frac{1}{|\mathcal{C}|} \sum_{y_1^n \in V\left(x_1^n\right)} W\left(y_1^n | x_1^n\right) \\ &= 1 - \sum_{x_1^n \in \mathcal{C}} \frac{1}{|\mathcal{C}|} \sum_{y_1^n \in \pi_{x_1^n}^{-1}(V(\mathbf{0}))} W\left(\pi_{x_1^n}\left(y_1^n\right) | \mathbf{0}\right) \\ &= 1 - \sum_{y_1^n \in V(\mathbf{0})} W\left(y_1^n | \mathbf{0}\right). \end{aligned}$$

This shows that the probability of ML decoding error for a group code over an output-symmetric channel is independent of the transmitted codeword.

1.5 The Pairwise Error Probability (PEP)

1.5.1 Discrete Memoryless Channels

Since computing the exact probability of error requires extensive knowledge of the code, it is often useful to have bounds that are easier to compute. The basis of many of these bounds is the pairwise error probability (PEP) between any two codewords. The PEP, denoted $P\left(x_1^n \to \widetilde{x}_1^n\right)$, is the probability that the ML decoder chooses $\widetilde{x}_1^n$ when x_1^n was transmitted. This probability can be written as

$$P\left(x_1^n \to \widetilde{x}_1^n\right) = \sum_{y_1^n \in \mathcal{Y}^n} W\left(y_1^n | x_1^n\right) I\left(W\left(y_1^n | x_1^n\right) \leq W\left(y_1^n | \widetilde{x}_1^n\right)\right),$$

where $I(E)$ is the indicator function for the event E (i.e., it equals 1 if the argument is true and 0 otherwise). The indicator function is upper bounded by

$$I\left(W\left(y_1^n | x_1^n\right) \leq W\left(y_1^n | \widetilde{x}_1^n\right)\right) \leq \left(\frac{W\left(y_1^n | \widetilde{x}_1^n\right)}{W\left(y_1^n | x_1^n\right)}\right)^s,$$

for any $s \in [0,1]$, because the LHS is zero if the RHS is less than one and the LHS is one when the RHS is greater than one. In general, the best bound is found by minimizing over s. For binary-input symmetric-output channels, the minimum occurs at $s = 1/2$ and the implied bound is

$$\begin{aligned} P\left(x_1^n \to \widetilde{x}_1^n\right) &= \sum_{y_1^n \in \mathcal{Y}^n} W\left(y_1^n | x_1^n\right) \left(\frac{W\left(y_1^n | \widetilde{x}_1^n\right)}{W\left(y_1^n | x_1^n\right)}\right)^{1/2} \\ &= \prod_{i=1}^{n} \sum_{y_i \in \mathcal{Y}} \sqrt{W\left(y_i | x_i\right) W\left(y_i | \widetilde{x}_i\right)} \\ &= \prod_{i=1}^{d_H\left(x_1^n, \widetilde{x}_1^n\right)} \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)\, W(y|1)}, \end{aligned}$$

because the sum is one if $x_i = \widetilde{x}_i$. This bound is known as the Bhattacharyya bound and is typically written as

$$P\left(x_1^n \to \widetilde{x}_1^n\right) \leq \gamma^{d_H\left(x_1^n, \widetilde{x}_1^n\right)},$$

where $\gamma = \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)}$ is the Bhattacharyya constant of the channel. For the BSC channel, this gives $\gamma_{BSC} = 2\sqrt{p(1-p)}$. For the binary-input AWGN (BIAWGN) channel with energy per symbol E_s and noise spectral density N_0, we have

$$W(y|x) = (\pi N_0)^{-1/2} e^{-(y - \sqrt{E_s}(-1)^x)^2 / N_0}$$

and

$$\begin{aligned} \gamma_{BIAWGN} &= \int_{-\infty}^{\infty} (\pi N_0)^{-1/2} \left[e^{-(y-\sqrt{E_s})^2/N_0} e^{-(y+\sqrt{E_s})^2/N_0}\right]^{1/2} dy \\ &= \int_{-\infty}^{\infty} (\pi N_0)^{-1/2} e^{-(y^2+E_s)/N_0} dy = e^{-E_s/N_0}. \end{aligned}$$

It is also known that γ is the best possible constant for bounds of the form γ^{d_H}.

1.5.2 The AWGN Channel

If the channel consists of a modulator $M\left(x_1^n\right) \to \mathbb{R}^n$ and zero-mean AWGN with variance σ^2 per-dimension, then the PEP can be computed exactly. In this case, the memoryless channel (it is no longer discrete) is defined by the conditional p.d.f.

$$W\left(y_1^n | x_1^n\right) = \left(2\pi\sigma^2\right)^{-n/2} e^{-\frac{1}{2\sigma^2}\left\|y_1^n - M\left(x_1^n\right)\right\|^2}.$$

Since the p.d.f. depends only on the distance between the received and transmitted vectors, we find that the ML decoder picks the codeword whose transmitted vector is closest to the received vector. To analyze this, we can project the received vector y_1^n onto the difference vector $w_1^n = M\left(\widetilde{x}_1^n\right) - M\left(x_1^n\right)$ to get the decision variable

$$Z = \frac{\sum_{i=1}^n w_i Y_i}{\sqrt{\sum_{i=1}^n w_i^2}}.$$

One can verify that Z is a zero-mean Gaussian random variable with variance σ^2. Furthermore, the decoder will make an error if an only if $Z \geq \left\|w_1^n\right\|/2$ (i.e., the received vector is closer to $\widetilde{x}_1^n$ than x_1^n). This allows us to rewrite the PEP as

$$\begin{aligned}
P\left(x_1^n \to \widetilde{x}_1^n\right) &= \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2 \cdots \int_{-\infty}^{\infty} dy_n W\left(y_1^n | x_1^n\right) \, I\left(W\left(y_1^n | x_1^n\right) \leq W\left(y_1^n | \widetilde{x}_1^n\right)\right) \\
&= \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2 \cdots \int_{-\infty}^{\infty} dy_n W\left(y_1^n | x_1^n\right) \, I\left(\left\|y_1^n - M\left(x_1^n\right)\right\| \geq \left\|y_1^n - M\left(\widetilde{x}_1^n\right)\right\|\right) \\
&= \int_{-\infty}^{\infty} f_Z(z) \, I\left(z \geq \frac{1}{2}\left\|M\left(\widetilde{x}_1^n\right) - M\left(x_1^n\right)\right\|\right) dz \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\left\|M\left(\widetilde{x}_1^n\right) - M\left(x_1^n\right)\right\|/2}^{\infty} e^{-z^2/(2\sigma^2)} dz.
\end{aligned}$$

A change of variables shows that this integral is equal to

$$P\left(x_1^n \to \widetilde{x}_1^n\right) = Q\left(\frac{1}{2\sigma}\left\|M\left(\widetilde{x}_1^n\right) - M\left(x_1^n\right)\right\|\right),$$

where $Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-z^2/2} dz$ is the tail probability of zero-mean unit-variance Gaussian.

Recall that the binary-input AWGN channel with $M(x) = \sqrt{E_s}(-1)^x$ has energy per transmitted symbol E_s and noise spectral density $N_0 = 2\sigma^2$. Therefore, the Euclidean distance is $\left\|M\left(\widetilde{x}_1^n\right) - M\left(x_1^n\right)\right\|^2 = 4E_s d_H\left(x_1^n, \widetilde{x}_1^n\right)$. Substituting these into our expression gives

$$P\left(x_1^n \to \widetilde{x}_1^n\right) = Q\left(\sqrt{2\, d_H\left(x_1^n, \widetilde{x}_1^n\right) \, E_s/N_0}\right).$$

Applying the standard bound, $Q(\alpha) \leq e^{-\alpha^2/2}$, to the Q-function gives an alternate proof of the Bhattacharyya bound for AWGN.

1.6 The Union Bound

Since every decoding error is caused a by pairwise error, we find that

$$P_B \leq \sum_{x_1^n \in \mathcal{C}} \frac{1}{|\mathcal{C}|} \sum_{\widetilde{x}_1^n \in \mathcal{C},\, \widetilde{x}_1^n \neq x_1^n} P\left(x_1^n \to \widetilde{x}_1^n\right).$$

This is only an upper bound because the received vector may be closer to two other codewords than it is to the transmitted codeword, and this causes "overcounting" of the error probability. If we assume that the code is linear and that the PEP is a function $f(h)$ of the Hamming distance h, then we get

$$\begin{aligned}
P_B &\leq \sum_{x_1^n \in \mathcal{C}} \frac{1}{|\mathcal{C}|} \sum_{\widetilde{x}_1^n \in \mathcal{C},\, \widetilde{x}_1^n \neq x_1^n} f\left(d_H\left(x_1^n, \widetilde{x}_1^n\right)\right) \\
&= \sum_{x_1^n \in \mathcal{C}} \frac{1}{|\mathcal{C}|} \sum_{\widetilde{x}_1^n \in \mathcal{C},\, \widetilde{x}_1^n \neq \mathbf{0}} f\left(d_H\left(\mathbf{0}, \widetilde{x}_1^n\right)\right) \\
&= \sum_{\widetilde{x}_1^n \in \mathcal{C},\, \widetilde{x}_1^n \neq \mathbf{0}} f\left(d_H\left(\mathbf{0}, \widetilde{x}_1^n\right)\right) \\
&= \sum_{h=1}^{n} A_h f(h),
\end{aligned}$$

where A_h is the number of codewords of weight h. For binary codes, the function f is either chosen to be $Q\left(\sqrt{2h\,E_s/N_0}\right)$ (for the AWGN channel) or γ^h (for an arbitrary DMC with γ equal to the Bhattacharyya constant). In this case, the weight enumerator (WE) is often given in the polynomial form, $A(H) = \sum_{h\geq 0} A_h H^h$, and we can use the Bhattacharyya bound (i.e., $f(h) \leq \gamma^h$) to write

$$P_B \leq A(\gamma) - 1.$$

Example. The [7,4,3] Hamming code has the WE $A(H) = 1 + 7H^3 + 7H^4 + H^7$. This implies that ML decoding of this code on the BIAWGN channel has a block error probability which satisfies

$$P_B \leq 7e^{-3E_s/N_0} + 7e^{-4E_s/N_0} + e^{-7E_s/N_0}.$$

1.7 Bit Error Probability

In many cases, we are interested not only in the probability of *block error* P_B but also in the probability of *bit error* P_b (or symbol error P_s for non-binary codes). To compute P_b we need to compute the average number of message bit (or symbol) errors that occur as the result of a block error. Let $E : \mathcal{U}^k \to \mathcal{X}^n$ be an encoder which maps any length k input sequence to a length n output sequence. Consider any two input-output pairs, $x_1^n = E\left(u_1^k\right)$ and $\widetilde{x}_1^n = E\left(\widetilde{u}_1^k\right)$, and notice that the pairwise error $x_1^n \to \widetilde{x}_1^n$ implies the message error $u_1^n \to \widetilde{u}_1^n$ and produces $d_H\left(\widetilde{u}_1^k, u_1^k\right)$ symbol errors in the decoded message. Using the union bound, we can bound P_s with

$$P_s \leq \sum_{u_1^k \in \mathcal{U}^k} \frac{1}{|\mathcal{U}|^k} \sum_{\widetilde{u}_1^k \in \mathcal{U}^k,\, \widetilde{u}_1^k \neq u_1^k} P\left(E(u_1^n) \to E(\widetilde{u}_1^n)\right) \frac{d_H\left(\widetilde{u}_1^k, u_1^k\right)}{k}.$$

While this quantity can be computed or bounded for any code, it can be simplified for codes with linear encoders. Now, we will assume that $\mathcal{U}$ has a field structure and that E is linear so that, for $\alpha, \beta \in \mathcal{U}$,

$$E\left(\alpha u_1^k + \beta \widetilde{u}_1^k\right) = \alpha\, E\left(u_1^k\right) + \beta\, E\left(\widetilde{u}_1^k\right).$$

In this case, the linearity implies that $\widetilde{x}_1^n - x_1^n = E\left(\widetilde{u}_1^k - u_1^k\right)$ and that the pairwise error $x_1^n \to \widetilde{x}_1^n$ produces $w_H\left(\widetilde{u}_1^k - u_1^k\right)$ symbol errors in the decoded message. If we also assume that the PEP is a function $f(h)$ of the Hamming distance h, then we can write

$$\begin{aligned}
P_s &\leq \sum_{u_1^k \in \mathcal{U}^k} \frac{1}{|\mathcal{U}|^k} \sum_{\widetilde{u}_1^k \in \mathcal{U}^k,\, \widetilde{u}_1^k \neq u_1^k} f\left(w_H\left(E(u_1^k) - E(\widetilde{u}_1^k)\right)\right) \frac{w_H\left(\widetilde{u}_1^k - u_1^k\right)}{k} \\
&= \sum_{\widetilde{u}_1^k \in \mathcal{U}^k,\, \widetilde{u}_1^k \neq \mathbf{0}} f\left(w_H\left(E(\widetilde{u}_1^k)\right)\right) \frac{w_H\left(\widetilde{u}_1^k\right)}{k} \\
&= \sum_{w=1}^{k} \sum_{h=1}^{n} A_{w,h} f(h) \frac{w}{k},
\end{aligned}$$

where the input-ouput weight-enumerator (IOWE), $A_{w,h}$, is the number of codewords with input weight w and output weight h. The IOWE of a linear code is often given in polynomial form as $A(W,H) = \sum_{w=1}^{k} \sum_{h=1}^{n} A_{w,h} W^w H^h$. For binary codes, we can therefore use the Bhattacharyya bound (i.e., $f(h) \leq \gamma^h$) to write

$$P_b \leq \frac{1}{k} \left[\frac{d}{dW} A(W,\gamma)\right]_{W=1}.$$

Example. One encoder for the [7,4,3] Hamming code has the IOWE $A(W,H) = 1 + (3W + 3W^2 + W^3)H^3 + (W + 3W^2 + 3W^3)H^4 + W^4 H^7$. This implies that ML decoding of this code on the BIAWGN channel has a block error probability which satisfies

$$P_b \leq \frac{12}{4} e^{-3E_s/N_0} + \frac{16}{4} e^{-4E_s/N_0} + \frac{4}{4} e^{-7E_s/N_0}.$$