On the Blizard Decoding Algorithm

L. R. Welch
University of Southern California

This article presents an analysis and modification of the new Blizard decoding algorithm, which promises to give performance superior to any known practical decoding algorithm on the deep space channel.

I. Introduction

In Ref. 1, R. B. Blizard describes a new method for decoding binary linear error correcting codes. The algorithm was presented ad hoc and the description was sketchy. In an attempt to understand his process, I have developed a modification which has a partly logical, partly heuristic derivation as an approximation to the maximum likelihood estimator. The computational complexity of these algorithms is within the range of practicability, while for most codes it is impractical to implement the maximum likelihood estimate.

The mathematical foundation, given here, reveals the assumptions needed to derive the algorithms. It is, however, necessary that an investigation be carried out to determine the suitability of these algorithms for specific codes and channels.

II. Preliminaries

Let $m_1, \cdots, m_k$ be random variables which take on the values 0 and 1. Let G be a k by n matrix of 0's and 1's and define

$$T_j = \sum_{i=1}^{k} m_i G_{ij} \bmod 2, \qquad j = 1, 2, \cdots, n$$

The m_i's are message bits, the T_j's are transmitted bits, and G is the generating matrix of the code. Finally, let $p(Z|0)$ and $p(Z|1)$ be two probability densities and $Z_1, \cdots, Z_n$ be random variables whose joint density, given $m_1, \cdots, m_k$, is

$$P(Z_1, \cdots, Z_n | m_1, \cdots, m_k) = \prod_{j=1}^{n} p(Z_j | T_j) \qquad (1)$$

The Z_i's are the received symbols of a memoryless channel.

The decoding problem is to determine functions $\widehat{m}_i(Z_1, \cdots, Z_n)$ which satisfy some performance criterion. If all messages are equally likely and minimum probability of word error is desired, then the estimator is the maximum likelihood estimator, $(m_1^*, \cdots, m_k^*)$ which satisfies

$$P(Z_1, \cdots, Z_n | m_1^*, \cdots, m_k^*) = \max_{m_1, \cdots, m_k} P(Z_1, \cdots, Z_n | m_1, \cdots, m_k) \qquad (2)$$

In practice this estimator is not easily computed (with the exception of Viterbi's dynamic programming algorithm for convolutional codes with small memory) so that approximations are required to minimize equipment.

III. Derivation of the Algorithm

The algorithm begins with tentative probabilities assigned to the m_i's and attempts to repetitively improve these estimates by using the estimates as *a priori* probabilities and replacing them by *a posteriori* probabilities using the Z_j's. The goal is to achieve the maximum likelihood estimator.

Let $\{P_\theta : \theta = (\theta_1, \cdots, \theta_k), |\theta_i| \leqq 1\}$ be a parametric family of probability functions defined as follows:

$$P_\theta(Z_1, \cdots, Z_n, m_1, \cdots, m_k) = \prod_{j=1}^{n} P(Z_j | T_j) \prod_{i=1}^{k} \left(\frac{1 + (-1)^{m_i} \theta_i}{2} \right) \tag{3}$$

where T_j is defined as before. There are other methods for parameterizing this family but the θ's are convenient since $E_\theta[(-1)^{m_i}] = \theta_i$.

Observe that

$$\begin{aligned} P_\theta(Z_1, \cdots, Z_n) &= \sum_{m_1, \cdots, m_k} P(Z_1, \cdots, Z_n | m_1, \cdots, m_k) P_\theta(m_1, \cdots, m_k) \\ &\leqq \sum_{m_1, \cdots, m_k} P(Z_1, \cdots, Z_n | m^*, \cdots, m_k^*) P_\theta(m_1, \cdots, m_k) \\ &= P(Z_1, \cdots, Z_n | m_1^*, \cdots, m_k^*) \\ &= P_{\theta^*}(Z_1, \cdots, Z_n) \end{aligned} \tag{4}$$

where $(m_1^*, \cdots, m_k^*)$ is defined by Eq. (2) and $\theta_j^* = (-1)^{m_j^*}$. Therefore, θ^* is a solution to the problem of finding that θ which maximizes $P_\theta(Z_1, \cdots, Z_n)$. Conversely, if the maximum likelihood estimate is unique, the solution to the parametric maximization problem is unique and is at $\theta_j^* = (-1)^{m_j^*}$.

From the probability model defined by θ, the *a posteriori* probabilities $P_\theta(m_i | Z_1, \cdots, Z_n)$ and the *a posteriori* expectations

$$\theta_i'(\theta, Z_1, \cdots, Z_n) = E_\theta[(-1)^{m_i} | Z_1, \cdots, Z_n] \tag{5}$$

can be computed. The substitution of θ' for θ induces a transformation, $\sigma(\theta)$ defined by Eq. (5) on the parameter space. This transformation has been studied (Refs. 2 and 3) and is known that

$$P_{\sigma(\theta)}(Z_1, \cdots, Z_n) \geqq P_\theta(Z_1, \cdots, Z_n)$$

with equality only if $\theta = \sigma(\theta)$. Further, $\theta = \sigma(\theta)$ only at stationary points of $P_\theta(Z_1, \cdots, Z_n)$ (regarded as a function of only θ). This fact suggests the following procedure: select some θ^0 and define recursively $\theta^t = \sigma(\theta^{t-1})$ for $t = 1, 2, \cdots$. The function values $P_{\theta^t}(Z_1, \cdots, Z_n)$ increase and for almost all choices of θ^0 (i.e., except for a set of Lebesque measure 0), the sequence will converge to a local maximum (Ref. 3). Hopefully, if θ^0 is in a neutral position, the maximum point will have enough influence on the trajectory to be the point of convergence.

Next, consider the formula for $\sigma(\theta)$. Algebraic manipulation of Eqs. (3) and (5) results in the equation

$$\frac{1 + \theta_i'}{1 - \theta_i'} = \frac{P_\theta(Z_1, \cdots, Z_n | m_i = 0)}{P_\theta(Z_1, \cdots, Z_n | m_i = 1)} \frac{1 + \theta_i}{1 - \theta_i} \tag{6}$$

where

$$P_\theta(Z_1, \cdots, Z_n | m_i = a) = \sum_{\substack{m_1, \cdots, m_k = 0, 1 \\ m_i = a}} \prod_{j=1}^{n} P(Z_j | m_1, \cdots, m_k) \prod_{l \neq i} \left(\frac{1 + (-1)^{m_l} \theta_l}{2} \right) \tag{7}$$

This formula is of little practical value since the work required to evaluate it grows exponentially with k. However, if we assume that $P_\theta(Z_1, \cdots, Z_n | m_i = a)$ is well approximated by

$$\prod_{j=1}^{n} P_\theta(Z_j | m_i = a)$$

(that is, the Z's are conditionally independent given m_i), then the work is significantly reduced and Eq. (6) becomes

$$\frac{1+\theta_i'}{1-\theta_i'} = \frac{1+\theta_i}{1-\theta_i} \prod_{j=1}^{n} \frac{P_\theta(Z_j | m_i = 0)}{P_\theta(Z_j | m_i = 1)} \tag{8}$$

The jth factor can be written in terms of channel probabilities as

$$\frac{P_\theta(Z_j | m_i = 0)}{P_\theta(Z_j | m_i = 1)} = \frac{p(Z_j|0) P_\theta(T_j = 0 | m_i = 0) + p(Z_j|1) P_\theta(T_j = 1 | m_i = 0)}{p(Z_j|0) P_\theta(T_j = 0 | m_i = 1) + p(Z_j|1) P_\theta(T_j = 1 | m_i = 1)} \tag{9}$$

Now

$$T_j = \prod_{\ell=1}^{k} g_{\ell j} m_\ell$$

so that

$$E_\theta[(-1)^{T_j} | m_i = 0] = E_\theta\left[\prod_{\ell} (-1)^{m_\ell} \,\middle|\, m_i = 0\right] = \prod_{\ell \neq i} \theta_\ell \tag{10}$$

where the products are over all ℓ for which $g_{\ell j} = 1$. Labelling the last product in Eq. (10), β_{ij}, reduces Eq. (9) to

$$\frac{P_\theta(Z_j | m_i = 0)}{P_\theta(Z_j | m_i = 1)} = \frac{p(Z_j|0)(1+\beta_{ij}) + p(Z_j|1)(1-\beta_{ij})}{p(Z_j|0)(1-\beta_{ij}) + p(Z_j|1)(1+\beta_{ij})} \tag{11}$$

provided $g_{ij} = 1$. If $g_{ij} = 0$, the ratio is 1. Equation (8) can now be written

$$\frac{1+\theta_i'}{1-\theta_i'} = \frac{1+\theta_i}{1-\theta_i} \prod_{j} \frac{p(Z_j|0)(1+\beta_{ij}) + p(Z_j|1)(1-\beta_{ij})}{p(Z_j|0)(1-\beta_{ij}) + p(Z_j|1)(1+\beta_{ij})} \tag{12}$$

where the product is over all j with $g_{ij} = 1$.

With the exception of the factor $(1+\theta_i)/(1-\theta_i)$, Eq. (12) is equivalent to Blizard's transformation. This additional factor has a conservative effect. If the probability of $m_i = 0$ is close to one and the product in Eq. (12) is less than one, then the transformation with the $(1+\theta_i)/(1-\theta_i)$ factor lessens the probability a small amount. While the transformation without the factor switches the probability to less than one-half.

Blizard's initial probabilities can be obtained in a natural manner from Eq. (12) as follows. If all messages are assumed equally probable, the initial parameter $\theta^0 = (0, \cdots, 0)$. In this case $\beta_{ij} = 0$ unless the product defining β_{ij} is empty, in which case $\beta_{ij} = 1$. This corresponds to $Tj = m_i$. If there is only one value of j for which this is true, Eq. (12) reduces to

$$\frac{1+\theta_i'}{1-\theta_i'} = \frac{p(Z_j|0)}{p(Z_j|1)} \tag{13}$$

which corresponds to Blizard's initial probabilities.

The suitability of these algorithms depends upon two things. First, the assumption

$$P_\theta(Z_1, \cdots, Z_n | m_i = a) = \prod_{j} P_\theta(Z_j | m_i = a)$$

should not do too much violence to the probability model. And secondly, the product factor in Eq. (12) should have a distribution which is not clustered too near 1.

References

1. Blizard, R. B., *Study of Applications of Digital Techniques to Apollo S-Band Communications, Phase 2, Final Report*, NASA-MCR 70-419. Martin Marietta Corp., Nov. 1970.

2. Baum, L. E., and Eagon, J. A., "An Inequality with Applications to Statistical Prediction for Functions of Markov Processes," *Bull. Am. Math. Soc.*, Vol. 73, pp. 360–363, 1967.

3. Baum, L. E., and Sell, G. R., "Growth Transformations for Functions on Manifolds," *Pacific J. Math.*, Vol. 27, pp. 211–222, 1968.