WEIGHTED DECODING AS A MEANS FOR REESTIMATING
A PROBABILITY DISTRIBUTION (ABSTRACT)

Gérard Battail,
Ecole Nationale Supérieure des Télécommunications,
46, rue Barrault, F - 75634 PARIS CEDEX 13, France

1 - A REDEFINITION OF DECODING

We propose to redefine weighted decoding of a redundant code as consisting of reestimating a given
prior probability distribution, available for each received symbol as the demodulator output, in order to
take into account the code constraints. The result from such redefined decoding will be referred to as
posterior probability distribution.

The usual maximum-likelihood decoding rule consists in principle of: (i) computing the probability of
the hypothesis that a given word has been transmitted, for each codeword; and (ii) choosing the
codeword which corresponds to the hypothesis whose probability is the largest. We just propose not to
perform step (ii) which makes the result simpler but irreversibly destroys part of the information
obtained from step (i).

The practical interest of the proposed redefinition becomes clear when one looks at a coding system
where two codes are concatenated i.e., where the result from encoding the data by one of the codes
(referred to as outer) is in turn encoded according to the second one (referred to as inner). If decoding
the inner code results in a probability distribution instead of a hard decision, it can be used in order to
weight the outer decoding, therefore improving it.

As redefined, decoding may be considered as a particular case of the problem of determining a
posterior distribution i.e., compatible with a given set of constraints, which is the best one with respect
to a given prior probability distribution. According to Shore and Johnson [1], the unique optimum
solution to this problem results from applying Kullback’s minimum cross-entropy principle [2].

Let \{p(x_i)\} be the prior probability distribution associated with a finite or countable set of elementary
events \{x_i\}, to be denoted by \(p\); we assume that none of the probabilities \(p(x_i)\) is zero. Kullback
principle consists of choosing, among the distributions \{q(x_i)\} which are compatible with the
constraints (to be denoted by \(q\)), the one which minimizes cross-entropy (denoted by \(q^*\)). Cross-
entropy of \(q\) and \(p\) is defined as :
\[ H(q, p) = \sum_i q(x_i) \log \frac{q(x_i)}{p(x_i)}. \]  

(1)

\( H(q, p) \) is not symmetric in \( p \) and \( q \); it is positive and in a certain sense measures the vicinity of \( q \) and \( p \).

If a block code of length \( n \) is to be decoded, the weighting information results from the demodulator evaluating the probability of the hypotheses concerning each of the transmitted symbols. Provided we assume them to be independently affected by the channel errors, a prior probability can be assigned to any \( n \)-uple by the product of the probability of its symbols, whether this \( n \)-uple belongs or not to the code. As redefined, decoding consists of recomputing the probability distribution on the \( n \)-uples in order to take into account the code constraints. The posterior distribution \( q^* \) must therefore involve probability zero for all \( n \)-uples which do not belong to the code. Clearly, the result from interverting \( p \) and \( q \) in (1) is meaningless, which expresses the decoding irreversibility.

2 - REDEFINING MAXIMUM LIKELIHOOD DECODING

Assuming that the exact prior distribution is known, we easily show that Kullback principle results in the set of probabilities which are to be compared in step (ii) of conventional maximum likelihood decoding.

In the binary case (as an example), we consider decoding a particular codeword and assume that the exact prior probability of the symbols is available as the demodulator output. The posterior probability of a given word \( c_j \) of code \( C \) which results from applying Kullback principle is precisely equal to the probability that the transmitted \( n \)-uple is \( c_j \), conditioned on the fact it belongs to code \( C \). Therefore applying Kullback principle is equivalent to maximum likelihood decoding in this case.

Kullback principle may also be applied if one assumes that a particular codeword is transmitted (e.g., the all-zero word if the code is assumed to be linear) and considers the probability distribution on the received symbols as resulting from the channel transition probabilities (hence continuous for additive continuous noise if no quantization occurs).

In both cases, the posterior distribution is degenerated in the sense that the contraints result in restricting its support. Hence \( q^* \) is proportional to \( p \) wherever it is nonzero; the proportionality constant simply renormalizes it and its logarithm is equal to \( H(q^*, p) \). It is furthermore possible to restrict the distribution support step by step. We shall make use of this remark in Paragraph 4.