


E. Quality Measurement Studies: Legibility, Moire Effects, and Scan Aperture Functions


F. Related Bibliographies


On the Computation of Rate-Distortion Functions

I. CSISZÁR

Abstract—In a recent paper [1], Blahut suggested an efficient algorithm for computing rate-distortion functions. In this correspondence we show that the sequence of distributions used in that algorithm has a limit yielding a point on the R(d) curve if the reproducing alphabet is finite, and we obtain a similar but weaker result for countable reproducing alphabets.

I. INTRODUCTION

In a recent paper [1], Blahut suggested efficient algorithms for computing channel capacity and rate-distortion functions; for the former case, the same algorithm appears (though in less generality) in the work of Arimoto [2]. The convergence proofs in [1] contain a gap, claiming that a product \( b_1 \cdots b_k \) cannot have a limit (as \( k \to \infty \)) if a subsequence of the \( b_i \) has a limit \( > 1 \); this is true for a positive limit only. In [2] the proofs are complete, and, moreover, the distributions themselves used in the algorithm are shown to converge to one achieving capacity. We shall use a method similar to that of [2] to prove an analogous result for the rate-distortion function, supposing that the reproducing alphabet is finite. A similar but weaker result will also be obtained for an infinite reproducing alphabet. As reviewers pointed out, Blahut’s original convergence statement [see (12)] has also been proved in a paper of Boukris [3], as well as by Blahut himself in his Ph.D. dissertation [4].

II. PRELIMINARIES

Suppose that symbols selected from a finite or countable alphabet with distributions \( p = (p_k) \) are to be reproduced by symbols of another finite or countable alphabet. For a given loss matrix \( (p_{ik}) \), the rate-distortion function \( R(d) \) is defined as the minimum amount of information needed for reproduction with average loss \( \leq d \), i.e.,

\[
R(d) = \inf \left\{ H(Q) \mid \sum_k p_k Q_{k|j} \log \frac{Q_{k|j}}{p_k} \right\}
\]

Manuscript received March 6, 1973; revised July 30, 1973.
The author is with the Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary.
where the inf refers to \( Q \): \( D(Q) \leq d \), with
\[
D(Q) = \sum_k \sum_j p_j Q_{kj} p_{jk}.
\]
(2)
(The arguments \( p \) are omitted because \( p \) will be fixed throughout.)

Setting \( F_k = \inf Q \{ f(Q) - sD(Q) \} \), the \( R(d) \) curve is the envelope of the lines with slope \( s \leq 0 \) and \( k \)-axis intercept \( F_k \), see e.g., [5]. It suffices therefore to compute \( F_k \). Blahut's algorithm [1] is based on the observation that \( F_k = \inf Q \{ F_k(Q) \} \) where
\[
F_k(Q) = \sum_j p_j Q_{kj} \log \frac{Q_{kj}}{q_{kj}} - sD(Q)
\]
(3)
and it consists in successive minimizations with respect to \( Q \) and \( q \).

Starting from an arbitrary \( Q(1) \), set recursively
\[
Q(n) = \arg \min Q \{ F(Q) \}
\]
(4)
where \( Q_j \) stands for the distribution \( (Q_k, j) \), for fixed \( j \), and \( I(q \parallel q') \) denotes the nonnegative quantity
\[
I(q \parallel q') = \sum_k q_k \log \frac{q_k}{q'_k}.
\]
(5)

This was established in [1] by Lagrange multipliers. A more direct proof is given by the easily checked identities
\[
F_k(Q) = F_k(Q) + I(q \parallel q') + \sum_j p_j I(Q_j \parallel Q(Q))
\]
(6)
where \( Q_j \) stands for the distribution \( (Q_{kj}, j) \), for fixed \( j \), and \( I(q \parallel q') \) denotes the nonnegative quantity
\[
I(q \parallel q') = \sum_k q_k \log \frac{q_k}{q'_k}.
\]
(7)

Starting from an arbitrary \( q(1) \) with \( q(1) > 0 \), for each \( k \), set recursively \( Q^{(n)} = Q(q^{(n-1)}) \), \( q^{(n)} = Q(q^{(n)}), \) \( n = 2,3, \ldots \). Then clearly
\[
F_k(Q^{(2)}) \geq F_k(Q^{(2)}, q^{(2)}) \geq F_k(Q^{(3)}, q^{(3)}) \geq \ldots .
\]

Theorem 1: If the reproducing alphabet is finite, then \( q^{(n)} \) exists for any \( n \), and \( q^{(n)} \) converges to \( q^* \) as \( n \to \infty \), where \( q^* \) denotes the nonnegative quantity
\[
I(q \parallel q') = \sum_k q_k \log \frac{q_k}{q'_k}.
\]
(8)

This shows, in particular, that the series
\[
\sum_n \left[ F_k(Q^{(n)}, q^{(n-1)}) - F_k(Q, q) \right]
\]
converges; thus
\[
\lim_{n \to \infty} \left[ F_k(Q^{(n)}, q^{(n-1)}) - F_k(Q, q) \right] = 0
\]
(9)
if \( I(q \parallel q') < \infty \), proving
\[
\lim_{n \to \infty} F_k(Q^{(n)}, q^{(n)}) = \lim_{n \to \infty} F_k(Q^{(n)}, q^{(n-1)}) = F_k = \inf F_k(Q, q)
\]
(10)
provided that the infimum in (10) can be approached by \( q \) with \( I(q \parallel q') < \infty \). The latter condition is trivially fulfilled if the reproducing alphabet is finite; if it is countable, an easily checked sufficient condition is
\[
d^* = \inf q_k p_j < \infty.
\]
(11)

For a countable reproducing alphabet, this proof breaks down at two points: i) it may not be possible to extract a subsequence from \( q^{(n)} \) converging to a distribution \( q^* \); ii) even if such \( q^{(n)} \to q^* \) can be found, \( I(q^* \parallel q^{(n)}) \to 0 \) does not necessarily follow.

Theorem 2: Even for a countable reproducing alphabet, if \( q^* \) and \( Q^* = Q(q^*) \) achieving \( F_k \) exist and \( I(q^* \parallel q^{(n)}) \to 0 \) \( \forall n \), the backward probabilities (9) always converge to those corresponding to \( Q^* \) [see (7)], for each \( k \) with \( q_k^* > 0 \). Moreover,
\[
\sum_k q_k \exp (sp_{jk}) \to \sum_k q_k \exp (sp_{jk}), \quad \text{for all } j.
\]
(12)

Proof: (8) implies not only (10), but similarly:
\[
\sum_k q_k \log \frac{q_k}{q_k^{(n-1)}} = I(q \parallel q^{(n)}) \to I(q \parallel q^*).
\]
(13)

IV. DISCUSSION

We have shown that the sequence of distributions appearing in Blahut's [1] algorithm has a limit yielding a point on the \( R(d) \) curve, if the reproducing alphabet is finite, and established the convergence of the "backward probabilities" also for a countable reproducing alphabet. The cardinality of the original alphabet
has been irrelevant. Let us remark that in general, each limiting
distribution of a subsequence of \(q^{(b)}\) (if any) achieves \(F_b\), on
account of the lower semicontinuity of \(F_b(Q,q)\). Under a com-
 pactness condition, such as that for each \(j\) and \(K\), the set of \(K\)
with \(p_{jk} < K\) be finite, in which case \(Q^*\) and \(q^* = q(Q^*)\)
achieving \(F_b\) certainly exist, one can therefore always assert \(q^{(b)} \to q^*\) if \(q^*\) is unique. While \(q^*\) need not be unique, the sums
\(\sum q_{jk} e^x(p_{jk})\) are by Theorem 2. The case of a countable
reproducing alphabet has been included in spite of its limited
practical interest because the results obtained for this case have
straightforward extensions to abstract alphabets, replacing prob-
abilities by densities and sums by integrals.

REFERENCES
[1] R. E. Blahut, "Computation of channel capacity and rate-distortion
1972.
Blahut algorithm for computing rate-distortion functions," IEEE

A Generalization of Huffman Coding for Messages with
Relative Frequencies Given by Upper and Lower Bounds

STEPHEN A. SMITH

Abstract—A generalization of the Huffman coding procedure is
given for cases in which the source letter probabilities are known only
to fall in certain ranges.

I. INTRODUCTION TO THE PROBLEM

Given a random variable \(X\) taking on \(N\) discrete values with
\[ p_i = P(X = x_i), \quad i = 1, 2, \cdots, N \]
and an alphabet with a finite number of elements \(D\), the classical
noiseless coding problem is to construct from the alphabet a code
consisting of \(N\) distinct words corresponding to the respective
values of \(X\) so that the average or expected word length for
transmitting \(X\) is minimized. If we let
\[ l_c(e), l_1(e), \cdots, l_N(e) \]
denote the respective word lengths for a given code \(c\), the Huff-
man coding procedure [3] gives us a solution to this problem
\[
\min \sum_{c \in A} p_{i} l_c(e) \tag{1.1}
\]
where \(A\) is the set of admissible codes.

We will consider the following generalization of the problem.
Suppose that the probability distribution for \(X, p = p_1, p_2, \cdots, p_N\)
is specified only within upper and lower bounds of the form
\[
0 \leq \alpha_i \leq p_i \leq \beta_i \leq 1, \quad i = 1, 2, \cdots, N
\]
giving a set
\[
G = \{ p \mid \alpha_i \leq p_i \leq \beta_i \text{ and } \sum p_i = 1 \}
\]
of possible probability distributions. Such a situation might
occur due to lack of precise information on the probabilities or
because the message frequencies are known to fluctuate as a
function of time or other variables, e.g., the same channel and
transmission code might be used for several types of commu-
nication with each type using less than or equal to \(N\) possible
codewords.

Suppose we want to maximize the rate of transmission \(T\) that
can be guaranteed for all probability distributions \(p \in G\). That
is, we wish to find a code \(c \in A\) that minimizes the largest average
word length for probability distributions in \(G\) or, using our
previous notation, solves the problem
\[
\min \max_{c \in A} \sum_{i=1}^{N} p_i l(c). \tag{1.2}
\]
A solution \(\hat{c}\) would determine the best lower bound on trans-
mmission rate for given \(G\), namely,
\[
T = \frac{1}{\max_{c \in G} \sum_{i=1}^{N} p_i l(c)}. \tag{1.3}
\]

II. PROPOSED MINMAX SOLUTIONS

An equivalent formulation of the problem (1.2) is to determine
a pair of points \(\hat{b} \in G, \hat{a} \in \hat{A}\) satisfying
\[
\sum_{i=1}^{N} p_i l_{\hat{c}}(c) \leq \sum_{i=1}^{N} p_i l_{\hat{c}}(c), \quad \text{for all } p \in G, \tag{2.1}
\]
\[
\sum_{i=1}^{N} p_i l_{\hat{c}}(c) \geq \sum_{i=1}^{N} p_i l_{\hat{c}}(c), \quad \text{for all } p \in G. \tag{2.2}
\]
The proposed solutions are as follows. For \(\hat{b}\), take the prob-
ability distribution in \(G\), illustrated in Fig. 1, that is defined by
\[
\hat{b}_i = \begin{cases} \Delta, & \text{if } \alpha_i < \Delta < \beta_i \\ \alpha_i, & \text{if } \alpha_i \geq \Delta \\ \beta_i, & \text{if } \beta_i \leq \Delta \end{cases} \tag{2.3}
\]
where the value of \(\Delta\) is determined by the requirement that
\[
\sum_{i=1}^{N} \hat{b}_i = 1. \tag{2.4}
\]
It is clear that \(\hat{b}\) is unique since modifying it through increasing
or decreasing the value of \(\Delta\) immediately violates the require-
ment \(\hat{b}_1 + \cdots + \hat{b}_N = 1\). In addition, by using a graphical representation
like the one in Fig. 1, an appropriate value of \(\Delta\) can be
determined quickly by trial and error. For \(\hat{a}\) we then use the
Huffman Code that is optimal for \(\hat{b}\) and this code can, of course,
be determined by the Huffman coding procedure. The optimality
of these solutions is demonstrated in the Appendix.

The form of the solutions can be intuitively explained using
Shannon’s information measure [4] and the concept of maximum
entropy. Since the entropy of \(p\) for an alphabet of \(D\) characters
\[
-\sum_{i=1}^{N} p_i \log p_i
\]
is a lower bound on the actual minimum average word length
\[
\min_{c \in A} \sum_{i=1}^{N} p_i l(c)
\]

2 Note that if \(\sum_{i=1}^{N} \alpha_i < 1 < \sum_{i=1}^{N} \beta_i\), such a \(\Delta\) always exists, but is not
unique in some cases.