[4] M. R. Gevers and T. Kailath, "An innovations approach to least-squares estimation—part VI: discrete-time innovations representations and recursive estimation," *I.E.E.E. Trans. Automat. Contr.*, vol. AC-18, pp. 588–600, 1973.
[5] P. Whittle, *Prediction and Regulation.* New Jersey: Van Nostrand, 1963, pp. 24–28.
[6] E. A. Robinson, *Random Wavelets and Cybernetic Systems.* London: Griffin, 1962, p. 89.
[7] M. Schoenberger, "Resolution comparison of minimum-phase and zero-phase signals," *Geophysics*, vol. 39, pp. 828–833, 1974.
[8] E. A. Robinson, *Multi-Channel Time Series Analysis with Digital Computer Programs.* San Francisco: Holden-Day, 1967, p. 72.
[9] H. Akaike, "Power spectrum estimation through autoregressive model fitting," *Ann. Inst. Statist. Math.*, vol. 21, pp. 407–419, 1969.
[10] J. Burg, "A new analysis technique for time series data," paper presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, 1971.
[11] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Statist. Math.*, vol. 21, pp. 243–247, 1969.
[12] ——, "Statistical predictor identification," *Ann. Inst. Statist. Math.*, vol. 22, pp. 203–217, 1970.
[13] J. R. Rice, *The approximation of functions*, vol. 1. Reading, Mass.: Addison-Wesley, 1964, pp. 94–118.
[14] I. Barrodale and F. D. K. Roberts, "Solution of an overdetermined system of equations in the L_1 norm," Algorithm 478, *Communications of the Assoc. for Computing Machinery*, vol. 17, pp. 319–320, 1974.
[15] J. F. Claerbout and F. Muir, "Robust modeling with erratic data," *Geophysics*, vol. 38, pp. 826–844, 1973.

Asymptotically Optimal Quantizers

ROBERT M. GRAY, MEMBER, IEEE, AND A. H. GRAY, JR., MEMBER, IEEE

***Abstract*—Asymptotically accurate approximations for quantizer performance were developed by Gish and Pierce. Variational techniques were used to obtain asymptotically optimal performance. In this correspondence, optimal performance is demonstrated simply without variational techniques using Holder's and Jensen's inequalities.**

I. INTRODUCTION

Optimal quantizers minimize the transmission rate required to meet a specified average distortion (or conversely) where transmission rate is measured either by the logarithm of the number of quantizer levels or the entropy of the quantizer output. Gish and Pierce [1] developed asymptotically accurate approximations to the average distortion and output entropy of a quantizer with a large number of levels. Variational techniques were then applied to these approximations to obtain the optimal performance in the sense of minimizing the rate subject to an average distortion constraint. This work generalized approximations and optimal performance developed for special cases (usually squared-error distortion measures) by Panter and Dite [2], Lloyd [3], Roe [5], Algazi [6], and Wood [7]. Similar asymptotic results were obtained by Elias [8] using different distortion measures (not depending on the actual value of the source, but only on which quantization interval it appears in) and by Schutzenberger [9] and Zador [10] who developed upper and lower bounds to asymptotic performance. In [1]–[7], variational techniques were used to obtain the asymptotically optimal quantizers and performance. In [9] and [10], standard inequalities (e.g., Holder's inequality [11]) were used in lengthy arguments to obtain upper and lower bounds to the asymptotically optimal performance, but not asymptotically accurate approximations (except in the known squared-error case).

In this correspondence, we show that in fact the optimality results of Gish and Pierce can be obtained quite simply without variational techniques by using Holder's and Jensen's inequalities [11]. This greatly simplifies the proof and makes global optimality obvious.

The assumptions on the distortion measure made here differ only slightly from those of Gish and Pierce [1]; ours are somewhat stronger in the case of minimum entropy, and somewhat weaker in the case of minimum alphabet.

Manuscript received March 8, 1976; revised May 28, 1976. This research was partially supported by AFOSR Contract F 44620-73-C-0065.
R. M. Gray is with the Electrical Engineering Department, Stanford University, Stanford, CA 94305.
A. H. Gray is with the Electrical Engineering and Computer Science Depts., University of California, Santa Barbara, CA 93106, and Speech Communications Research Laboratory, Santa Barbara, CA 93109.

II. ASSUMPTIONS AND RESULTS

Our approach follows Berger [12, ch. 5] rather than Gish and Pierce [1], as Berger's approach leads to answers of a simpler and more intuitive form. Let X be a random variable described by a probability density function $p(x)$ that is nonzero on the interval (a, b). An N-level quantizer q is described by quantizer thresholds (or boundaries) $a = T_0 < T_1 < \cdots < T_{n-1} < T_n = b$; the quantizer levels (or points) $\hat{x}_1 < \hat{x}_2 < \cdots < \hat{x}_n$; and the relation

$$q(x) = \hat{x}_k, \qquad \text{for } T_{k-1} < x \le T_k.$$

The distortion resulting from reproducing a letter x as another letter y is assumed to be of the form $L(x - y)$, where

1) $L(0) = 0$
2) $L(\alpha)$ is an increasing function of $|\alpha|$,
3) the function $M(v)$, defined by

$$M(v) = \frac{1}{v}\int_{-v/2}^{v/2} L(\alpha)\, d\alpha = \int_{-1/2}^{1/2} L(\alpha v)\, d\alpha,$$

is convex $\cup$.

Assumption 3) is somewhat stronger than the initial Gish–Pierce [1] assumption that $vM'(v)$ is monotonic. Since 3) implies that $vM'(v)$ is monotonic, convexity of M implies the Gish–Pierce condition and hence the applicability of their asymptotic approximations. Note that both sets of assumptions include νth law distortion measures.

When N is large, let $N(x)dx$ denote the number of quantization levels between x and $x + dx$, and let $N \to \infty$ in such a way that

$$\lim_{N\to\infty} \frac{N(x)}{N} = \lambda(x)$$

exists for all x, i.e., $\lambda(x)$ is the asymptotic level density. Note that since $\lambda(x)$ is a density function,

$$N \int_a^\alpha \lambda(x)\, dx = \text{number of quantization levels in } [a,\alpha],$$

and

$$\int_a^b \lambda(x)\, dx = 1. \tag{1}$$

By changing variables in the Gish–Pierce result, or by a straightforward generalization of Berger [12, p. 173], we have that for large N

$$E(L(x - q(x))) \cong \int_a^b dx p(x) M(1/(N\lambda(x))) = E\{M(1/(N\lambda(X)))\} \tag{2a}$$

and

$$H = -\sum_{k=1}^{N} \Pr(\hat{x}_k) \log \Pr(\hat{x}_k)$$
$$\cong h(p) + \log N + \int dx p(x) \log \lambda(x)$$
$$= h(p) - E \log (1/(N\lambda(X))) \quad (2b)$$

where $h(p)$ is the differential entropy

$$h(p) = -\int dx p(x) \log p(x). \quad (3)$$

For the special case of $L(\alpha) = \alpha^2$, similar approximations are obtained in [2]–[7].

Gish and Pierce [1] find the optimum quantizer minimizing the rate subject to a constraint on $E(L(X - q(X)))$. A complete proof involves solving the Euler–Lagrange equations for a stationary point and then demonstrating that it is unique and corresponds to a minimum. Alternatively, if we require

$$E(L(X - q(X))) = \Delta,$$

then Jensen's inequality [11] yields

$$H \cong h(p) - E \log (1/(N\lambda(X))) \geq h(p) - \log E(1/N\lambda(X)) \quad (4)$$

and

$$\Delta \cong E\{M(1/(N\lambda(X)))\} \geq M(E(1/(N\lambda(X))) \quad (5)$$

or

$$M^{-1}(\Delta) \geq E(1/(N\lambda(X))) \quad (6)$$

whence

$$H \geq h(p) - \log M^{-1}(\Delta) \quad (7)$$

with equality if and only if $\lambda(x)$ is constant. Hence, (7) gives a lower bound on H which is achievable by $\lambda(x) = (b - a)^{-1}$, $x \in [a, b]$, i.e., by a uniform quantizer. This is a simple proof of (21) of Gish and Pierce for a slightly less general case.

Alternatively, if transmission rate is measured by $\log N$ and $L(\alpha) = |\alpha|^\nu$, then (1) and application of Holder's inequality [11] to (2a) or Berger's (P5.8.1) [12, p. 173] yield

$$E(|X - q(X)|^\nu) \cong \frac{1}{(\nu + 1)(2N)^\nu} \int_a^b p(x)\, \lambda^{-\nu}(x)\, dx$$
$$= \frac{1}{(\nu + 1)(2N)^\nu} \left(\int_a^b p(x)\, \lambda^{-\nu}(x)\, dx\right)\left(\int_a^b \lambda(x)\, dx\right)^\nu$$
$$\geqq \frac{1}{(\nu + 1)(2N)^\nu} \left(\int_a^b dx (p(x)\, \lambda(x)^{-\nu})^{1/1+\nu}\, \lambda(x)^{\nu/1+\nu}\right)^{1+\nu}$$
$$= \frac{1}{(\nu + 1)(2N)^\nu} \left(\int_a^b dx\, p(x)^{1/1+\nu}\right)^{1+\nu} \quad (8)$$

with equality if and only if $\lambda(x)$ is proportional to $p(x)^{1/1+\nu}$. Note that for the case $\nu = 1$, (8) is the Cauchy–Schwartz inequality. Thus, if we require $E(|X - q(X)|^\nu) = \Delta$, then

$$N \geq \frac{1}{2}\left(\frac{1}{(1 + \nu)\Delta}\left(\int_a^b dx\, p(x)^{1/1+\nu}\right)^{1+\nu}\right)^{1/\nu} \quad (9)$$

with equality if and only if

$$\lambda(x) = p(x)^{1/1+\nu} \Big/ \int_a^b d\alpha p(\alpha)^{1/1+\nu}. \quad (10)$$

This result is a generalization of the minimum alphabet result for $\nu = 2$ in Max [4] and Gish and Pierce [1]. This generalization is reported without proof by Berger [12, p. 173].

The methods used here to demonstrate optimality can also be applied to study quantization for a different class of distortion measures called integrated sensitivity measures which arise in speech compression [13].

References

[1] H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 676–683, Sept. 1968.
[2] P. F. Panter and W. Dite, "Quantization in pulse-count modulation with nonuniform spacing of levels," *Proc. IRE*, vol. 39, pp. 44–48, 1951.
[3] S. P. Lloyd, unpublished research.
[4] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Jan. 1960.
[5] G. M. Roe, "Quantizing for minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-10, pp. 384–385, Oct. 1964.
[6] V. R. Algazi, "Useful approximations to optimum quantization," *IEEE Trans. Commun. Tech*, vol. COM-14, pp. 297–301, 1966.
[7] R. C. Wood, "Optimum quantizing in hybrid computation," Ph.D. dissertation, Dept. of Eng., University of California, Los Angeles, August, 1966 (later published as "On optimum quantization," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 248–252, Mar. 1969).
[8] P. Elias, "Bounds on performance of optimum quantizers," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 172–184, Mar. 1970.
[9] M. P. Schutzenberger, "On the quantization of finite dimensional messages," *Inform. and Control*, vol. 1, pp. 153–158, 1958.
[10] P. Zador, "Topics in the asymptotic quantization of continuous random vectors," Bell Telephone Laboratories Technical Memorandum, February 1966.
[11] Hardy, Littlewood, and Polya, *Inequalities*. London: Cambridge University, 1951.
[12] T. Berger, *Rate-Distortion Theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.
[13] A. H. Gray, Jr., R. M. Gray, and J. Markel, "Comparison of optimal quantizations of speech reflection coefficients," To appear, *IEEE Trans. Acoust., Speech and Signal Processing*, 1977.

Walsh Analysis of Power-Law Systems

MOHAMMAD MAQUSI, MEMBER, IEEE

***Abstract*—Walsh-type signals exciting memoryless power-law systems are considered. The input is assumed to consist of a sum of weighted Walsh functions. A method is given for finding the weights of the Walsh functions in the output when the input is a finite Walsh series. It is shown that the fast Walsh transform can be used to facilitate the necessary computation.**

I. Introduction

In this paper, Walsh analysis of power-law systems (devices) is considered. The class of inputs is restricted to those signals that can be described as sums of weighted Walsh functions [1] of indicated sequencies; i.e. Walsh-type signals. The output is in turn described by another sum of Walsh functions. In general, the output contains sequencies that may be different from those appearing at the input. It may, however, contain some or all of the input sequencies. However, under certain conditions the system output sequencies are the same as the input sequencies, with different weights. This occurs when a set of sequencies is suitably chosen as $S_n = \{n: 0 \leqslant n \leqslant N - 1; N = 2^k, k = \text{positive integer}\}$. This latter condition may be imposed when the input process, over a finite (time) interval, can be described by a truncated Walsh series. Consequently, it becomes of interest to determine the output (sequency) coefficients in terms of the

Manuscript received March 22, 1976; revised June 15, 1976.
The author is with the College of Engineering, University of Jordan, Amman, Jordan.