is an $(n, d-1)R$ code. Similarly to Theorem 2, we prove that C is abnormal. □

There are three important consequences from Theorem 6.

Corollary 1: If Construction C is applied on the extended Hamming code of length 2^m, we obtain an abnormal $(2^m, 3)2$ code.

Corollary 2: If Construction C is applied on the punctured Preparata code of length $2^{2m}-1$, we obtain an abnormal $(2^{2m}-1, 4)3$ code.

Corollary 3: If Construction C is applied on the Preparata code of length 2^{2m}, we obtain an abnormal $(2^{2m}, 5)4$ code.

References

[1] G. D. Cohen, A. C. Lobstein, and N. J. A. Sloane, "Further results on the covering radius of codes," *IEEE Trans. Inform. Theory,* vol. IT-32, pp. 680–694, Sep. 1986.

[2] G. D. Cohen, M. G. Karpovsky, H. F. Mattson, and J. R. Schatz, "Covering radius—Survey and recent results," *IEEE Trans Inform. Theory,* vol. IT-31, pp. 328–343, May 1985.

[3] I. S. Honkala and H. O. Hämäläinen, "Bounds for abnormal binary codes with covering radius 1," *IEEE Trans. Inform. Theory,* vol. 37, pp. 372–375, Mar. 1991.

[4] X. Hou, "some results on the norm of codes," *IEEE Trans. Inform. Theory,* vol. 36, pp. 683–685, May 1990.

[5] ——, "Binary linear quasi-perfect codes are normal," *IEEE Trans. Inform. Theory,* vol. 37, pp. 378–379, Mar. 1991.

[6] K. E. Kilby and N. J. A. Sloane, "On the covering radius problem for code II. Codes of low dimension; normal and abnormal codes," *SIAM J. Algebraic Discrete Methods,* vol. 8, pp. 619–627, Oct. 1987.

[7] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error–Correcting Codes.* Amsterdam: North-Holand, 1977.

[8] G. J. M. van Wee, "More binary covering codes are normal," *IEEE Trans Inform. Theory,* vol. 36, pp. 1466–1470, Nov. 1990.

[9] S. G. Vladuţ and A. N. Skorobogatov, "Covering radius for long BCH codes," *Probl. Peredach. Inform.*, vol. 25, pp. 38–45, Jan.–Mar. 1989.

Distributed Estimation and Quantization

John A. Gubner, *Member, IEEE*

***Abstract*—An algorithm is developed for the design of a nonlinear, n-sensor, distributed estimation system subject to communication and computation constraints. The algorithm uses only bivariate probability distributions and yields locally optimal estimators that satisfy the required system constraints. It is shown that the algorithm is a generalization of the classical Lloyd–Max results.**

***Index Terms*—Nonlinear estimation, distributed estimation, sensor fusion, Lloyd–Max algorithm.**

I. Introduction

Consider the distributed estimation system shown in Fig. 1. The system consists of n sensor platforms whose respective measurements, $Y_1, \cdots, Y_n$, are related to some unobservable quantity, say X. Each sensor platform processes its respective measurement and transmits the result over a communication channel to a common fusion center. The sensors do not communicate with each other, and there is no feedback from the fusion center to the sensor platforms. The task of the fusion center is to estimate the unobserved quantity X. We denote this estimate by $\hat{X}$. Clearly, $\hat{X}$ is a function of $Y_1, \cdots, Y_n$, and we can write $\hat{X} = f(Y_1, \cdots, Y_n)$ for some function f. The problem then is to choose the function f so that $\hat{X}$ is close to X in some sense. For example, it is well known that in the appropriate probabilistic setting, the minimum-mean-square-error estimate of X given $Y_1, \cdots, Y_n$ is the conditional expectation of X given $Y_1, \cdots, Y_n$, denoted $\mathsf{E}[X \mid Y_1, \cdots, Y_n]$. However, there are many situations in which the conditional expectation does not provide a satisfactory solution to the problem of choosing f.

1) In general, the functional form of $\mathsf{E}[X \mid Y_1, \cdots, Y_n]$ as a function of $Y_1, \cdots, Y_n$ is difficult to determine, and it requires knowledge of the joint probability distribution of $X, Y_1, \cdots, Y_n$. In practice this complete joint distribution may not be available.
2) To compute $\mathsf{E}[X \mid Y_1, \cdots, Y_n]$, the fusion center must in general have access to all of the sensor measurements $Y_1, \cdots, Y_n$. Hence, even if the sensor platforms have local processing capability, it is of little use in computing $\mathsf{E}[X \mid Y_1, \cdots, Y_n]$. If the number of sensor platforms is very large, the burden of computing $\mathsf{E}[X \mid Y_1, \cdots, Y_n]$ at the fusion center, even if the formula is relatively simple, may be prohibitive. Such considerations are important if the estimate of X must be computed in real time. By using a suboptimal estimator of X for which some of the processing can be done locally at the sensor platforms, it may be possible to design an acceptable estimator that can operate in real time.
3) As indicated in Fig. 1, the sensor platforms transmit their data to the fusion center. However, using any physical communication system, it is not possible to transmit real-valued quantities without distortion. In this situation, the conditional expectation, or even the best linear estimate, is generally a physically unrealizable solution.

In this correspondence, we develop an algorithm to design solutions to the distributed estimation problem that do not suffer from these difficulties.

II. Background and Notation

Our approach is to consider quantization for distributed estimation systems. The goal of quantization in such systems is to provide a good estimate of the unobservable, X, rather than to reconstruct the sensor measurements $Y_1, \cdots, Y_n$ as in [3]. Quantization for estimation has been studied for a single sensor by Ephraim and Gray [2] and by Ayanoglu [1]. The multisensor case has been studied by Lam and Reibman [5], and we discuss their work in more detail below. Zhang and Berger [9] considers an asymptotic estimation problem in which the observations are discrete random variables taking finitely many values and the unobservable quantity is not a random variable, but a deterministic and unknown parameter in some finite-dimensional Euclidean space.

A. System Model

Let $X, Y_1, \cdots, Y_n$ be real-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Each sensor platform k processes its measurement Y_k to obtain an output Z_k. Each Z_k is then transmitted

Manuscript received August 10, 1992; revised December 15, 1992. This work was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-90-0181. This work was presented in part at the 1990 IEEE International Symposium on Information Theory, San Diego, CA, January 14–19, 1990.

The author is with the Department of Electrical and Computer Engineering, University of Wisconsin, 1415 Johnson Drive, Madison, WI 53706–1691.

IEEE Log Number 9209594.

0018–9448/93$03.00

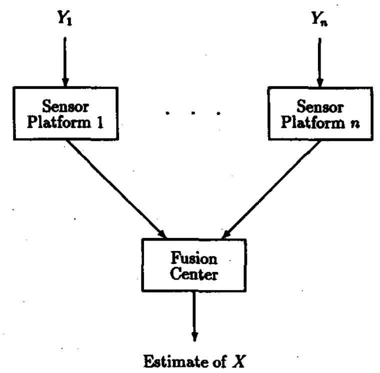

Fig. 1. A distributed estimation system.

to the fusion center. We assume that the communication channel connecting the sensors to the fusion center has a positive capacity, and that the use of error-correcting codes permits us to view the channel as noiseless. We suppose that the channel can transmit messages of $\log_2 N$ bits without error, where $N \geq 2$ is an integer. For each k, let $A_{k1}, \cdots, A_{kN}$ be a partition of the real line, $\mathbf{R}$. We require that the sensor platform output Z_k be given by

$$Z_k \triangleq \sum_{i=1}^{N} (i-1) I_{A_{ki}}(Y_k), \tag{1}$$

where $I_A(y)$ denotes the indicator function of the set $A \subset \mathbf{R}$; i.e., $I_A(y) = 1$, if $y \in A$ and $I_A(y) = 0$, otherwise.

Under the preceding constraints, the function f discussed in Section I must be of the form

$$f(Y_1, \cdots, Y_n) = h(Z_1, \cdots, Z_n),$$

where each Z_k is equal to the function of Y_k determined by (1).

B. Relation to [5]

We now briefly summarize the approach in [5]. If the sets $\{A_{ki}\}$ are fixed, one wants to find a function $h(Z_1, \cdots, Z_n)$ that minimizes the mean-square error,

$$\mathsf{E}[\,|X - h(Z_1, \cdots, Z_n)|^2\,]; \tag{2}$$

hence, the optimal h is the conditional expectation,

$$h(z_1, \cdots, z_n) = \mathsf{E}[\,X \mid Z_1 = z_1, \cdots, Z_n = z_n\,].$$

If we set $i_1 = z_1 + 1, \cdots, i_n = z_n + 1$ and let

$$B \triangleq A_{1i_1} \times \cdots \times A_{ni_n},$$

then this conditional expectation is given by

$$\frac{\int_B \mathsf{E}[\,X \mid Y_1 = y_1, \cdots, Y_n = y_n\,]\, dF_{Y_1 \cdots Y_n}(y_1, \cdots, y_n)}{\mathsf{P}(Y_1 \in A_{1i_1}, \cdots, Y_n \in A_{ni_n})}. \tag{3}$$

Clearly, in order to compute (3), we need to know $\mathsf{E}[\,X \mid Y_1, \cdots, Y_n\,]$. If the entire joint distribution $F_{XY_1 \cdots Y_n}$ is not available, computation of h will not be possible in general. Another consideration in some applications is the computation of (3) in real time. If (3) is not computable in real time, all the different possible values of the right-hand side of (3) will have to be precomputed and stored. For an n-sensor system with N-component partitions, there are N^n different numbers to compute and store. Finally, if more sensors are added at a later date, there will be no way to take advantage of the work already done to develop the n-sensor system; all of the numbers given by (3) will have to be recomputed for an even larger value of n.

The preceding paragraph assumed that the partitions were given. If h is arbitrary and given, and the partitions $\{A_{ki}\}_{i=1}^N$ are given for $k \neq l$, then the remaining partition should satisfy (in order to minimize (2) [5])

$$y \in A_{li} \iff V_{li}(y) \leq V_{lj}(y), \qquad \text{for all } j = 1, \cdots, N, \tag{4}$$

where,

$$V_{li}(y) \triangleq \mathsf{E}[\,|\mathsf{E}[\,X \mid Y_1, \cdots, Y_n\,] - h_l(i-1)|^2 \mid Y_l = y\,],$$

and $h_l(i-1) \triangleq h(Z_1, \cdots, Z_{l-1}, i-1, Z_{l+1}, \cdots, Z_n)$. The approach in [5] was to use (3) and (4) as the basis of an algorithm for computing a locally optimal quantizer for a distributed estimation system. Briefly, one starts with an arbitrary initial quantizer and computes a function $h^{(1)}$ given by (3). Using $h^{(1)}$ and the initial partition, a new partition is generated using (4). One then repeats these steps using the new partition to generate $h^{(2)}$ according to (3), and so on. The algorithm stops when the relative change in the mean-square error falls below a preset threshold.

As the preceding discussion indicates, the computational size of this problem grows exponentially with the number of sensors n. Next, we impose constraints on h so that the size of the problem of finding a locally optimal quantizer grows linearly with n.

III. Constraining the Fusion Center

Our approach [4] is to constrain the computational capabilities of the fusion center a priori as follows. We require that

$$\hat{X} = h(Z_1, \cdots, Z_n) = \sum_{k=1}^{n} g_k(Z_k), \tag{5}$$

where

$$g_k(Z_k) \triangleq \sum_{i=1}^{N} c_{ki} I_{\{i-1\}}(Z_k). \tag{6}$$

Remark 1: In spite of the sums in (5) and (6), the fusion center is performing a *nonlinear* operation on the input data $Z_1, \cdots, Z_n$. In fact, since the Z_k are discrete random variables, the set of possible inputs does not constitute a vector space over $\mathbf{R}$. Similarly, each g_k in (6) is a nonlinear function of Z_k, and in (1), Z_k is a nonlinear function of Y_k.

Combining (5) and (6), and recalling that $Z_k = i - 1$ if and only if $Y_k \in A_{ki}$, we have

$$\hat{X} = \sum_{k=1}^{n} \sum_{i=1}^{N} c_{ki} I_{A_{ki}}(Y_k). \tag{7}$$

Clearly, $\hat{X}$ is a nonlinear function of $Y_1, \cdots, Y_n$. However, if the partitions at the sensors are fixed, choosing the $\{c_{ki}\}$ that minimize $\mathsf{E}[\,|X - \hat{X}|^2\,]$ is a linear-estimation problem whose solution is given by the usual normal equations. In this case, we will have Nn equations in Nn unknowns. Hence, the number of equations will grow linearly with the number of sensors n. The moments needed to write down the normal equations are

$$\mathsf{E}[\,I_{A_{ki}}(Y_k) I_{A_{lj}}(Y_l)\,] = \begin{cases} \mathsf{P}(Y_k \in A_{ki})\delta_{ij}, & l = k, \\ \mathsf{P}(Y_k \in A_{ki}, Y_l \in A_{lj}), & l \neq k, \end{cases} \tag{8}$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise, and

$$\mathsf{E}[\,X I_{A_{ki}}(Y_k)\,] = \int_{A_{ki}} \mathsf{E}[\,X \mid Y_k = y\,]\, dF_{Y_k}(y). \tag{9}$$

Note that one needs only the joint distributions of the form $F_{Y_k Y_l}$ and F_{XY_k} and not the entire joint distribution $F_{XY_1 \cdots Y_n}$.

Definition 1: Given a partition $\{A_{ki}\}_{i=1}^N$ for each sensor k, we write

$$\boldsymbol{A} = (\{A_{1i}\}_{i=1}^N, \cdots, \{A_{ni}\}_{i=1}^N).$$

We denote the procedure of solving the previously mentioned linear estimation problem by $\mathrm{SN}(\boldsymbol{A})$. Letting C denote the $n \times N$ matrix with elements c_{ki}, we write $C = \mathrm{SN}(\boldsymbol{A})$.

We now discuss how to find a good partition. If the matrix C is fixed, it is now very natural to ask how the best partition is characterized. To obtain the answer to this question, fix any $l = 1, \cdots, n$, and write

$$\mathsf{E}[|X - \hat{X}|^2] = \mathsf{E}[|X - g_l(Z_l) - \sum_{k \neq l} g_k(Z_k)|^2]. \tag{10}$$

The right-hand side of (10) can be expanded into nine terms; however, only five terms will involve Z_l. Denoting the sum of these five terms by J_l, we have

$$J_l = \mathsf{E}[g_l(Z_l)\{g_l(Z_l) - 2(X - \sum_{k \neq l} g_k(Z_k))\}].$$

Recalling that Z_l is a function of Y_l (cf. (1)), we can use the smoothing property of conditional expectation to write

$$J_l = \mathsf{E}[g_l(Z_l)\{g_l(Z_l) - 2(\mathsf{E}[X \mid Y_l] - \sum_{k \neq l} \mathsf{E}[g_k(Z_k) \mid Y_l])\}].$$

We can then write

$$J_l = \sum_{i=1}^N \int_{A_{li}} \varphi_{li}(y)\, dF_{Y_l}(y),$$

where $\varphi_{li}(y) \triangleq c_{li}(c_{li} - 2r_l(y))$ and

$$\begin{aligned} r_l(y) &\triangleq \mathsf{E}[X \mid Y_l = y] - \sum_{k \neq l} \mathsf{E}[g_k(Z_k) \mid Y_l = y] \\ &= \mathsf{E}[X \mid Y_l = y] \\ &\quad - \sum_{k \neq l} \sum_{j=1}^N c_{kj} \mathsf{P}(Y_k \in A_{kj} \mid Y_l = y). \end{aligned} \tag{11}$$

Clearly, if the $\{c_{ki}\}$ are fixed for all k and i, and if the partitions $\{A_{ki}\}_{i=1}^N$ are fixed for $k \neq l$, then we should put

$$y \in A_{li} \iff \varphi_{li}(y) \leq \varphi_{li'}(y) \quad \text{for all } i' = 1, \cdots, N.$$

If we assume that $c_{l1} < \cdots < c_{lN}$, then this is equivalent to

$$A_{li} = \left\{ y \in \mathbf{R} : \frac{c_{l,i-1} + c_{li}}{2} < r_l(y) \leq \frac{c_{li} + c_{l,i+1}}{2} \right\}. \tag{12}$$

(The choice of $<$ and $\leq$ is arbitrary and is made so that the $\{A_{li}\}_{i=1}^N$ will be disjoint.) Observe that the function r_l depends on the $\{c_{kj}\}_{j=1}^N$ for all $k \neq l$. Also, the set A_{li} in (12) is *not* an interval, but rather the inverse image of an interval. It is also important to observe that to compute r_l for $l = 1, \cdots, n$ only requires knowing the two-dimensional joint distributions F_{XY_l} and $F_{Y_k Y_l}$ for all k and l. An important consequence of this fact is that if we decide to add another sensor to measure, say Y_{n+1}, our prior knowledge of F_{XY_l} and $F_{Y_k Y_l}$ for $k, l \leq n$ can be reused. Of course, we would still need to obtain $F_{XY_{n+1}}$ and $F_{Y_k Y_{n+1}}$ for $k = 1, \cdots, n$.

We conclude this section with a final definition and a remark.

Definition 2: We introduce the procedure $U_l(C, \boldsymbol{A})$. Recall our notation in Definition 1. Let $\hat{\boldsymbol{A}} = U_l(C, \boldsymbol{A})$ be obtained from $\boldsymbol{A}$ by replacing $\{A_{li}\}_{i=1}^N$ with $\{\hat{A}_{li}\}_{i=1}^N$, where each $\hat{A}_{li}$ is given by the right-hand side of (12).

Remark: If $n = 1$ and $X = Y_1$, then the normal equations reduce to

$$\mathsf{P}(Y_1 \in A_{1i})c_{1i} = \mathsf{E}[Y_1 I_{A_{1i}}(Y_1)],$$

or

$$c_{1i} = \frac{\int_{A_{1i}} y\, dF_{Y_1}(y)}{\mathsf{P}(Y_1 \in A_{1i})}.$$

Further, $r_1(y) = y$, and so

$$A_{1i} = \left\{ y \in \mathbf{R} : \frac{c_{1,i-1} + c_{1i}}{2} < y \leq \frac{c_{1i} + c_{1,i+1}}{2} \right\}.$$

In other words, we recover the classical Lloyd–Max conditions for locally optimal quantizers [6], [7].

IV. The Design Algorithm

Using the basic procedures SN and U_l, $l = 1, \cdots, n$, defined in the preceding section, there are two, almost identical, algorithms for generating approximately locally optimal quantizers for distributed estimation systems.

Algorithm 1:

```
        Let A = ({A_1i}_{i=1}^N, ..., {A_ni}_{i=1}^N) be given.
        C := SN(A)
loop1:  FOR l = 1 TO n
            A := U_l(C, A)
        NEXT l
        C := SN(A)
        IF stopping criterion not met, GO TO loop1
        END
```

Algorithm 2:

```
        Let A = ({A_1i}_{i=1}^N, ..., {A_ni}_{i=1}^N) be given.
        C := SN(A)
loop2:  FOR l = 1 TO n
            A := U_l(C, A)
            C := SN(A)
        NEXT l
        IF stopping criterion not met, GO TO loop2
        END
```

While the preceding algorithms appear simple enough, their implementation is nontrivial. The two main difficulties in implementing the algorithms are the computation of the function $r_l(y)$ in (11) and the characterization of the inverse images in (12). Note that even if $X, Y_1, \cdots, Y_n$ are jointly Gaussian, we cannot write (11) in closed form even if the A_{kj} are intervals. Hence, the sets A_{li} in (12) must be determined numerically and then a description of them must be stored in a suitable data structure.

A set of programs has been developed [8] to implement Algorithms 1 and 2 when provided with subroutines to compute the particular moments and probabilities for a given situation. Several examples of the form

$$Y_k = X + W_k, \quad k = 1, 2,$$

where X, W_1, and W_2 are statistically independent were considered.

Example [8, Example 8]: Let X have density

$$p(x) = \begin{cases} \frac{d}{b}[\frac{5}{4} - \cos(\frac{3\pi x}{2b})], & |x| \leq b, \\ 0, & \text{otherwise}, \end{cases} \tag{13}$$

where $d \approx 0.3419$ is a normalization constant and $b = 2$ (see Fig. 2). We let W_1 and W_2 have the same density except that $b = 1$. With

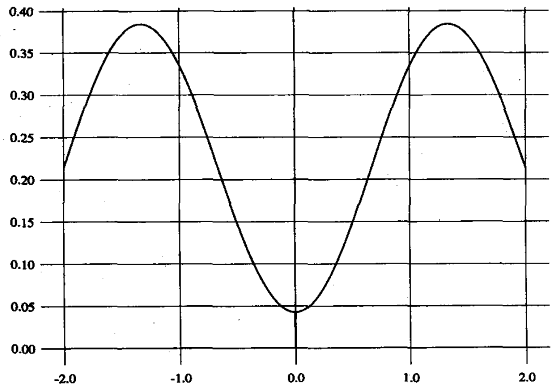

Fig. 2. Density $p(x)$ in (13) with $b = 2$.

$N = 8$ (3-bit quantizers), the Lloyd–Max partitions for the random variables Y_k, $k = 1, 2$, are identical and are given by

$$\begin{aligned} A_{k1} &= (-\infty, -1.3273] \\ A_{k2} &= (-1.3273, -1.2419] \\ A_{k3} &= (-1.2419, -1.0374] \\ A_{k4} &= (-1.0374, 0] \\ A_{k5} &= (0, 1.0374] \\ A_{k6} &= (1.0374, 1.2419] \\ A_{k7} &= (1.2419, 1.3273] \\ A_{k8} &= (1.3273, \infty). \end{aligned} \tag{14}$$

Solving the normal equations for $C = \mathrm{SN}(A)$, we have

$$\mathsf{E}[\,|X - \hat{X}|^2\,] = 0.18129.$$

Note that the minimum mean square error achievable by a linear estimator is 0.18534. Thus, just using the Lloyd–Max quantizers and doing linear estimation on Z_1, Z_2 can do better than pure linear estimation. Using Algorithm 1, we initialized A to the Lloyd–Max partition in (14). After 5 passes through the loop in Algorithm 1, the partitions were

$$\begin{aligned} A_{11} &= (-\infty, -2.1802] \\ A_{12} &= (-2.8102, -1.7950] \cup (-1.3697, -0.8177] \\ A_{13} &= (-1.7950, -1.3697] \cup (-0.8177, -0.4109] \\ A_{14} &= (-0.4109, 0.0005686] \\ A_{15} &= (0.0005686, 0.4130] \\ A_{16} &= (0.4130, 0.8200] \cup (1.3739, 1.7979] \\ A_{17} &= (0.8200, 1.3739] \cup (1.7979, 2.1834] \\ A_{18} &= (2.1834, \infty) \end{aligned}$$

and

$$\begin{aligned} A_{21} &= (-\infty, -2.1056] \\ A_{22} &= (-2.1056, -0.6504] \\ A_{23} &= (-0.6504, -0.3207] \\ A_{24} &= (-0.3207, -0.1805] \\ A_{25} &= (-0.1805, 0.2863] \\ A_{26} &= (0.2863, 0.6240] \\ A_{27} &= (0.6240, 2.0960] \\ A_{28} &= (2.0960, \infty). \end{aligned}$$

The minimum mean square error for these partitions is

$$\mathsf{E}[\,|X - \hat{X}|^2\,] = 0.12655,$$

which is more than a 30% improvement over the performance of the Lloyd–Max partition and over pure linear estimation.

Remark 2: After 5 passes through the algorithm, the mean square error was not significantly reduced.

Remark 3: The final partitions for the sensors are not the same, even though sensors 1 and 2 play interchangeable roles in this example. The reason for this is that the algorithm treats one sensor at a time.

Remark 4: As a general rule, it was found in [8] that Algorithm 2 yielded results almost identical to those of Algorithm 1.

V. Conclusion

We have developed an algorithm for the design of a distributed estimation system with n sensors and a single fusion center that is subject to communication and computation constraints. The algorithm uses only bivariate probability distributions and yields locally optimal estimators that satisfy the required system constraints.

While this work was motivated by problems in sensor fusion, the ideas can also be applied in a general nonlinear estimation context. In other words, estimators of the form (7) constitute a class of nonlinear estimators, and the algorithm presented here can be used to obtain a locally optimal nonlinear estimator from this class.

Acknowledgment

The author thanks W. Wehbe for developing the programming system used to obtain the numerical results in Section IV.

References

[1] E. Ayanoglu, "On optimal quantization of noisy sources," *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1450–1452, Nov. 1990.
[2] Y. Ephraim and R. M. Gray, "A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization," *IEEE Trans. Inform. Theory*, vol. 34, pp. 826–834, July 1988.
[3] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 773–787, Nov. 1987.
[4] J. A. Gubner, "Constrained distributed estimation and quantization for distributed estimation systems," in *Abstracts of Papers, 1990 IEEE Int. Symp. Inform. Theory*, San Diego, CA, pp. 32–33, Jan. 1990.
[5] W. M. Lam and A. R. Reibman, "Quantizer design for decentralized estimation systems with communication constraints," *Proc. Twenty-Third Ann. Conf. Inform. Sci. Syst.*, Johns Hopkins Univ., Baltimore, MD, Mar. 1989; also to appear in *IEEE Trans. Commun.*
[6] S. P. Lloyd, "Least squares quantization in PCM," unpublished Bell Lab. Memo., July 31, 1957; also *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–137, Mar. 1982.
[7] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Mar. 1960.
[8] W. H. Wehbe, "Quantization for distributed estimation systems," M.S. Rep., Dept. of Elect. Comput. Eng., Univ. of Wisconsin, Madison, 1992.
[9] Z. Zhang and T. Berger, "Estimation via compressed information," *IEEE Trans. Inform. Theory*, vol. 34, pp. 198–211, Mar. 1988.