Optimal Linear Coding for Vector Channels

KYONG-HWA LEE, MEMBER, IEEE, AND DANIEL P. PETERSEN, SENIOR MEMBER, IEEE

Abstract—This paper is concerned with the problem of obtaining the optimal linear vector coding (transformation) method that matches an r-dimensional vector signal and a k-dimensional channel under a given channel power constraint and mean-squared-error criterion. The encoder converts the r correlated random variables into r independent random variables and selects at most k independent random variables which correspond to the k largest eigenvalues of the signal covariance matrix Q. The encoder reinserts cross correlation into the k random variables in such a way that the largest eigenvalue of Q is assigned to the smallest eigenvalue of the channel noise covariance matrix R and the second largest eigenvalue of Q to the second smallest eigenvalue of R, etc.

When only the total power for all k channels is prescribed, the optimal individual channel power assignments are obtained in terms of the total power, the eigenvalues of Q, and the eigenvalues of R.

When the individual channel power limits are constrained by $P_1, \cdots, P_k$ and R is a diagonal matrix, the necessary conditions of an inverse eigenvalue problem must be satisfied to optimize the vector signal transmission system. An iterative numerical method has been developed for the case of correlated channel noise.

I. INTRODUCTION

In many communication systems, it is required to transmit a vector signal through a noisy vector channel. A color television signal or a set of measurements in a telemetering system may be considered as a vector signal. A multichannel system, using multiconductor cables or any type of multiplexing in a single medium, is an example of a vector signal transmission system.

The problem to be considered in this paper is the optimal coding of vector signals in the presence of white Gaussian (vector) channel noise. The term "optimal" is used in the minimum-mean-squared-error sense and the term "coding" is used not in the Shannon context but in the sense of a linear transformation. We assume that the signal from the source is an r-dimensional memoryless[1] vector and the channel is a k-dimensional vector channel. The optimal encoder is a linear transformation process that matches the r-dimensional signal vector and the k-dimensional channel under the given channel power constraint.

The fundamental problem in encoding can be separated into two component problems, namely: 1) What information should be transmitted? 2) How should it be transmitted? In order to solve Problem 1, we must first extract the significant information from the r-dimensional source signal and must order the data as to their importance according to our cost function. Next we must select the data that should be transmitted.

Problem 2 is a simple problem if the channel is a single channel since our coding is a linear transformation under the channel power constraint. However, it is not simple if the channel is a set of channels whose channel noises may be correlated with each other and individual channels may have different noise variance and channel power constraints. The main questions in Problem 2 are how to assign individual channels to the selected data, and how to reinsert cross correlation into the data to provide the best signal transmission over the specified channel.

II. PROBLEM STATEMENT

We shall be concerned with the vector signal transmission system whose block diagram is given in Fig. 1. The source is assumed to be a time-discrete memoryless Gaussian vector source whose output s is an r-dimensional stationary random vector sequence. We assume that successive signal vectors are independent but the elements of the signal vector s may be correlated. The vector channel is assumed to be a memoryless time-discrete additive noise channel which has k inputs and k outputs. The channel noise ν is independent of the input signal s and is a k-dimensional white Gaussian stationary vector sequence.

The encoder transforms the r-dimensional signal vector into a k-dimensional vector by linear transformation according to the equation

$$y = Cs$$

where C is a $k \times r$ encoding matrix of real numbers. Likewise, the decoder converts the k-dimensional channel output vector into an r-dimensional vector as an estimate $\hat{s}$ according to the equation

$$\hat{s} = F(y + \nu)$$

where F is an $r \times k$ decoding matrix of real numbers.

The criterion of performance is the weighted sum of error variances which may be expressed as

$$\mathcal{E} = \operatorname{tr} E[W(s - \hat{s})(s - \hat{s})^T] \tag{1}$$

where tr is trace operator, $W = \operatorname{diag}(w_1, w_2, \cdots, w_r)$, and w_i is the ith weighting factor. In this paper we investigate the design of the combination of encoder and decoder that minimizes $\mathcal{E}$. In order to avoid a degenerate solution, it is necessary to constrain the channel input power in some way. Otherwise,

Paper approved by the Editor for Communication Theory of the IEEE Communications Society for publication without oral presentation. Manuscript received October 30, 1975; revised April 22, 1976 and August 11, 1976.

K.-H. Lee is with the Data Systems Division, Gulton Industries, Inc., Albuquerque, NM 87123.

D. P. Petersen is with the Department of Electrical Engineering and Computer Science, University of New Mexico, Albuquerque, NM 87131.

[1] The extension to dynamical signals via the Kalman formulation is straightforward, cf. [1], [3], [4].

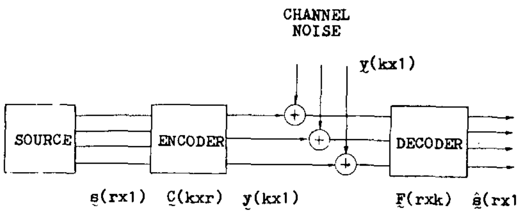

Fig. 1. Vector signal transmission system. Characters with tildes beneath are boldface in text.

making the elements of C very large results in an arbitrarily large signal-to-noise ratio at the channel output and reduces the problem to the noise-free case [3], [4]. We shall consider the following two types of constraints.

1) Total channel power constraint:

$$\sum_{i=1}^{k} E[y_i^2] = P. \tag{2}$$

2) Individual subchannel power constraints:

$$E[y_i^2] = P_i, \qquad i = 1, 2, \cdots, k. \tag{3}$$

III. NECESSARY CONDITIONS

Since $\hat{s} = F(Cs + \nu)$, the total error $\mathcal{E}$ becomes

$$\mathcal{E} = \mathrm{tr}\ [E[W(s - FCs - F\nu)(s - FCs - F\nu)^T]].$$

We assume that the means of the signal vector and the noise vector are zero, respectively, and we denote the covariance matrices of s and ν as follows:

$$E[ss^T] = Q \qquad E[\nu\nu^T] = R$$

where Q, R are positive definite matrices. Then the total error $\mathcal{E}$ is

$$\mathcal{E} = \mathrm{tr}\ [W(Q - 2FCQ + FCQC^TF^T + FRF^T)]. \tag{4}$$

If the total channel power is fixed by P, then this constraint can be written by

$$\mathrm{tr}\ [CQC^T] = P. \tag{5}$$

The unconstrained Lagrangian function can thus be written as

$$L_1 = \mathrm{tr}\ [W(Q - 2FCQ + FCQC^TF^T + FRF^T)] + \gamma(\mathrm{tr}\ [CQC^T] - P) \tag{6}$$

where γ is the Lagrange multiplier.

If each individual subchannel power is fixed by P_i, then this constraint can be written by

$$c_iQc_i^T = P_i, \qquad i = 1, 2, \cdots, k \tag{7}$$

where c_i is the ith row vector of C.

The unconstrained Lagrangian function for this case can be written as

$$L_2 = \mathrm{tr}\ [W(Q - 2FCQ + FCQC^TF^T + FRF^T)] + \mathrm{tr}\ [\Gamma(CQC^T - P)] \tag{8}$$

where $\Gamma = \mathrm{diag}\,(\gamma_1, \gamma_2, \cdots, \gamma_k)$ is the matrix Lagrange multiplier.

To optimize the decoder F, we differentiate L_1 or L_2 with respect to the components of F, keeping the coder C constant:

$$-2WQC^T + 2WFCQC^T + 2WFR = \mathbf{0}. \tag{9}$$

The optimal decoding matrix F becomes

$$F = QC^T(CQC^T + R)^{-1}. \tag{10}$$

Note that this criterion does not depend on W.[2] Likewise, differentiating L_1 with respect to the components of C and holding F constant, we get

$$(F^TWF + \gamma I)C = F^TW. \tag{11}$$

For the individual subchannel power constraint case, differentiating L_2 with respect to the components of C yields

$$(F^TWF + \Gamma)C = F^TW. \tag{12}$$

Equations (5), (10), and (11) are necessary conditions for the total power constraint problem and (7), (10), and (12) are necessary conditions for the subchannel power constraint problem. If we use the necessary condition (9) or (10), we can simplify the total error expression. By substitution of (9) into (4) and using (10) we get

$$\mathcal{E} = \mathrm{tr}\ [W(Q - QC^T(CQC^T + R)^{-1}CQ)]. \tag{13}$$

IV. SOURCE ENCODER AND SOURCE DECODER

In order to simplify our problem it is helpful to isolate the effect of the vector source in a vector signal transmission system from that of the vector channel. This can be done by breaking the encoder and decoder of Fig. 1 each into two parts as shown in Fig. 2.

The purpose of the source encoder is to remove intercomponent correlations from the r-dimensional source signal vector. The source encoder transforms a correlated source signal vector s into the uncorrelated signal vector x by a nonsingular column-weighted orthogonal transformation matrix. The channel encoder transforms the uncorrelated r-dimensional signal vector x into the k-dimensional vector y for input to the channel. If the subchannel noises are correlated, then the channel encoder reinserts cross correlations into the signal vector to provide optimal immunity to the channel noise.

For convenience, let $UW^{1/2}$ be the source encoder matrix. We choose U as an orthogonal matrix such that

$$W^{1/2}QW^{1/2} = Q' = U^T\Lambda U \tag{14}$$

where Λ is a diagonal matrix whose diagonal elements are

[2] If any $w_l = 0$, the lth row of F is arbitrary and will be assigned according to (10) for convenience.

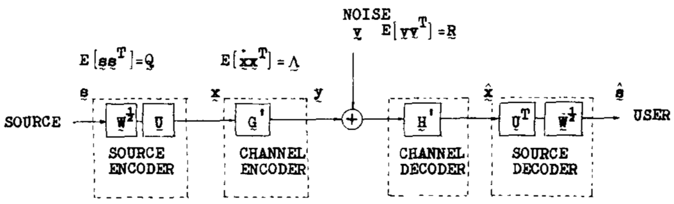

Fig. 2. Block diagram of vector signal transmission system with encoder and decoder each split into two parts. Characters with tildes beneath are boldface in text.

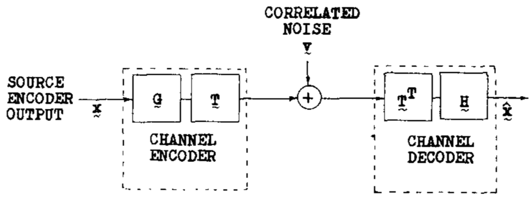

Fig. 3. Channel encoder and decoder. Characters with tildes beneath are boldface in text.

eigenvalues of Q'. We note that these eigenvalues depend both on the signal covariance matrix Q and on the performance weighting matrix W.

We shall denote the channel encoding matrix as G'. By substitution of $C = G'UW^{1/2}$ into (13), we obtain

$$\mathcal{E} = \operatorname{tr}\,[W^{1/2}QW^{1/2} - W^{1/2}QW^{1/2}U^TG'^T \cdot (G'UW^{1/2}QW^{1/2}U^TG'^T + R)^{-1}G'UW^{1/2}QW^{1/2}].$$

Since $W^{1/2}QW^{1/2} = U^T\Lambda U$, and the trace is unaffected by a similarity transformation [5], we get

$$\mathcal{E} = \operatorname{tr}\,[\Lambda - \Lambda G'^T(G'\Lambda G'^T + R)^{-1}G'\Lambda]. \tag{15}$$

The weighted error is thus expressed in terms of the diagonal matrix Λ whose elements are eigenvalues of Q'. Thus the problem is to find G' for this uncorrelated signal x.

V. OPTIMAL ENCODER FOR TOTAL POWER CONSTRAINT

Let the channel encoder G' and decoder H' be further broken into two parts as shown in Fig. 3. The channel encoder is now a cascade of transformations G and T, where T is an orthogonal matrix such that $R = TNT^T$ and $N = \operatorname{diag}(n_1, \cdots, n_k)$.

With the substitution $G' = TG$ into (15), the error $\mathcal{E}$ becomes

$$\mathcal{E} = \operatorname{tr}\,[\Lambda - \Lambda G^T(G\Lambda G^T + N)^{-1}G\Lambda]. \tag{16}$$

The error is now expressed in terms of the diagonal matrices Λ and N, and encoder G. Consider an equivalent vector signal transmission system as shown in Fig. 4. The minimum error for this system (with $W = I$) is obviously the same as (16).

The total power constraint for Fig. 3 is

$$P = \operatorname{tr}\,[G\Lambda G^T]. \tag{17}$$

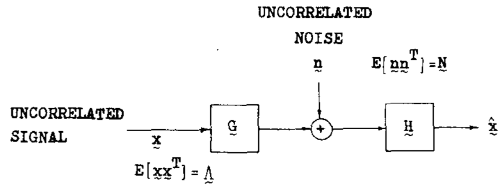

Fig. 4. Equivalent vector signal transmission system. Characters with tildes beneath are boldface in text.

It is the same as the total power constraint for Fig. 4. Thus, in this section we investigate how to design the optimal encoder G and decoder H for the vector signal transmission system given in Fig. 4. Once we find the optimal G and H, then these also provide the solution to the system given in Fig. 3.

The optimal decoder H is given in terms of G or (see (10))

$$H = \Lambda G^T(G\Lambda G^T + N)^{-1} \tag{18}$$

where N is the diagonal covariance matrix of n.

The necessary condition (11) is now given by

$$H^THG + \gamma G = H^T. \tag{19}$$

Postmultiplying by ΛG^T we have[3]

$$H^THG\Lambda G^T + \gamma G\Lambda G^T = H^T\Lambda G^T. \tag{20}$$

From (18) we know that

$$HG\Lambda G^T = \Lambda G^T - HN. \tag{21}$$

Substituting (21) into (20), we obtain

$$\gamma G\Lambda G^T = H^THN. \tag{22}$$

If we substitute (18) into (22), then the necessary condition becomes

$$\gamma G\Lambda G^T = (G\Lambda G^T + N)^{-1}G\Lambda^2G^T(G\Lambda G^T + N)^{-1}N. \tag{23}$$

We shall present some important properties of G, which are proved in Appendix A. We assume temporarily that no two elements of Λ or N are equal.

Property 1: The matrices $G\Lambda G^T$ and $G\Lambda^2G^T$ are diagonal.

Property 2: The nonzero column vectors of G are eigenvectors of a diagonal matrix D whose nonzero elements are also elements of Λ (for $k \leqslant r$) or include all elements of Λ (for $k \geqslant r$). D is defined by

$$D = G\Lambda^2G^T(G\Lambda G^T + N)^{-1} + \gamma(G\Lambda G^T + N). \tag{24}$$

Property 3: Every column vector of G has at most one nonzero element. Likewise, every row vector of G has at most one nonzero element.

These three properties serve as background for the optimal coder design problem. We may now rearrange the columns (or

[3] This does not affect the generality since the rank remains min (k,r).

rows) of G such that it consists of a partitioned matrix: one partition is diagonal of size min (r,k), and the other partition is zero. Let the elements in the diagonal partition be designated g_i, $i = 1, \cdots, \min(r,k)$, and their associated eigenvectors of the Λ matrix are λ_i. Now the total error $\mathcal{E}$ in (16) can be written as[4]

$$\mathcal{E} = \operatorname{tr}\left[\Lambda - G\Lambda^2 G^T (G\Lambda G^T + N)^{-1}\right]$$

$$= \sum_{i=1+\min(r,k)}^{r} \lambda_i + \sum_{i=1}^{\min(r,k)} \frac{\lambda_i n_i}{g_i^2 \lambda_i + n_i}. \tag{25}$$

Furthermore, in view of Property 1, the necessary condition (23) can be written as

$$\gamma G\Lambda G^T = G\Lambda^2 G^T N (G\Lambda G^T + N)^{-2} \tag{26}$$

whose ith diagonal element is

$$\gamma g_i^2 \lambda_i = \frac{g_i^2 \lambda_i^2 n_i}{(g_i^2 \lambda_i + n_i)^2}$$

so that (provided $g_i^2 \lambda_i \neq 0$)

$$\sqrt{\gamma} = \frac{\sqrt{\lambda_i n_i}}{g_i^2 \lambda_i + n_i} \tag{27}$$

or

$$\sqrt{\gamma}(g_i^2 \lambda_i + n_i) = \sqrt{\gamma_i n_i}. \tag{28}$$

Let the number of nonzero elements of G be $l \leqslant \min(r,k)$. Then summing (28) over i and employing (17) yields

$$\sqrt{\gamma} = \frac{\sum_{i=1}^{l} \sqrt{\lambda_i n_i}}{P + \sum_{i=1}^{l} n_i} \tag{29}$$

where P is the total channel power.

The error $\mathcal{E}$ in (26) can now be expressed in terms of the Lagrangian multiplier γ. From (26) and (27) we get

$$\mathcal{E} = \sum_{i=l+1}^{r} \lambda_i + \sqrt{\gamma} \sum_{i=1}^{l} \sqrt{\lambda_i n_i}. \tag{30}$$

The minimum $\mathcal{E}$ can be obtained if $\sqrt{\gamma}$ is minimized [1]. This is a nontrivial statement since γ depends on the identification of the nonzero elements of G. To find the minimum $\sqrt{\gamma}$, let the elements of Λ and N be ordered in such a way that

$$\lambda_1 > \lambda_2 > \cdots > \lambda_r$$

and

$$n_1 < n_2 < \cdots < n_k.$$

The minimum $\sqrt{\gamma}$ is given by [1]

$$\sqrt{\gamma} = \frac{\sum_{i=1}^{l} \sqrt{\lambda_i n_i}}{P + \sum_{i=1}^{l} n_i}. \tag{31}$$

Here the largest eigenvalue λ_i is assigned to the smallest noise n_i, the second largest λ_i to the second smallest noise n_i, and so on.

If we substitute (31) into (30), we get the minimum total error in terms of the total power:

$$\mathcal{E} = \frac{\left(\sum_{i=1}^{l} \sqrt{\lambda_i n_i}\right)^2}{P + \sum_{i=1}^{l} n_i} + \sum_{i=l+1}^{r} \lambda_i. \tag{32}$$

From (28) finally we get the optimum encoder G

$$\begin{aligned} g_{ii} &= \left(\sqrt{\frac{\lambda_i n_i}{\gamma}} - n_i\right)^{1/2} \Big/ \sqrt{\lambda_i}, &&\text{for } i \leqslant l \\ g_{ij} &= 0, &&\text{when } i \neq j. \end{aligned} \tag{33}$$

Here g_{ii} must be real. Therefore, from (33)

$$\sqrt{\frac{\lambda_i n_i}{\gamma}} - n_i > 0, \qquad \text{for } 1 \leqslant i \leqslant l \tag{34}$$

or

$$\frac{\lambda_i}{n_i} > \gamma, \qquad \text{for } 1 \leqslant i \leqslant l. \tag{35}$$

The optimal integer l is thus the maximum integer number which satisfies (35).

Up to this point, we assumed that no two elements of N are equal. Let us consider the case in which $N = \sigma_n^2 I$. If we add ϵ, 2ϵ, 3ϵ, $\cdots$ to the diagonal elements of N where ϵ is an arbitrary small value, then the previous results can be used,[5] and the encoder G becomes a diagonal matrix. However for $\epsilon = 0$, we find that the optimal encoder is not unique. Suppose the subchannels have equal noise powers or $N = \sigma_n^2 I$, then we may assign the subchannels arbitrarily. Furthermore, we may transform the encoded signal by an arbitrary orthogonal matrix T. Let G be an optimal encoder and $\mathcal{E}$ be given by

$$\mathcal{E} = \operatorname{tr}\left[\Lambda - \Lambda G^T (G\Lambda G^T + \sigma_n^2 I)^{-1} G\Lambda\right].$$

If we transform the encoded signal by an orthogonal matrix T, then $\mathcal{E}$ is

[4] Note that $\operatorname{tr} AB = \operatorname{tr} BA$ for any $A(k \times r)$ and $B(r \times k)$. By convention, $\Sigma_{i=j}^{k} Q_i = 0$ when $j > k$.

[5] Note that the error is continuous in the elements of N.

$$\begin{aligned}\mathcal{E} &= \operatorname{tr}\,[\Lambda - \Lambda G^T T^T (TG\Lambda G^T T^T + \sigma_n{}^2 I)^{-1} TG\Lambda] \\ &= \operatorname{tr}\,[\Lambda - \Lambda G^T T^T T (G\Lambda G^T + \sigma_n{}^2 T^T T)^{-1} T^T TG\Lambda] \\ &= \operatorname{tr}\,[\Lambda - \Lambda G^T (G\Lambda G^T + \sigma_n{}^2 I)^{-1} G\Lambda].\end{aligned}$$

We see that the error is unchanged. Hence we conclude that if $N = \sigma_n{}^2 I$ and G is the optimum diagonal encoding matrix, then TG is also an optimal encoder.

We note that this orthogonal transformation does not change the total power in the vector channel or the power assigned to each eigenvalue of Λ, but the individual subchannel powers are changed by this transformation. We return to this idea in the next section.

Fig. 5 shows the variation of the subchannel powers for various total channel power specifications when $\Lambda = \operatorname{diag}(10, 9, 8, \cdots, 2, 1)$ and $N = I$ where I is a 10×10 identity matrix. Note that two of the subchannels are unused for the case $P = 5$, and the two smallest signal eigenvalues are not transmitted.

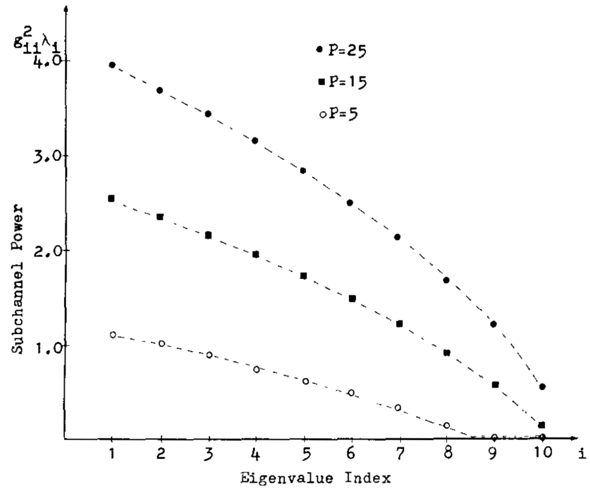

Fig. 5. Subchannel power distribution for example.

VI. OPTIMAL ENCODER FOR SUBCHANNEL POWER CONSTRAINTS

In this section we shall investigate how to design the optimal channel encoder G' for an uncorrelated signal x when the individual subchannel power is constrained by $P_1, P_2, \cdots, P_k$, or in matrix form

$$G'\Lambda G'^T = \begin{bmatrix} P_1 & * & \cdots & * \\ * & P_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ * & \cdots & * & P_k \end{bmatrix} \triangleq P.$$

A necessary condition for the minimum $\mathcal{E}$ is given by (12) with substitution $C = G'UW^{1/2}$ and $F = W^{-1/2}U^T H'$, viz.,

$$(H'^T H' + \Gamma)G' = H'^T.$$

Postmultiplying by $\Lambda G'^T$, we obtain

$$H'^T H' G' \Lambda G'^T + \Gamma G' \Lambda G'^T = H'^T \Lambda G'^T. \tag{36}$$

Since G'^T has rank min (k,r), the generality of the necessary condition is not reduced.

From (9) and (14) we know that

$$H' G' \Lambda G'^T = \Lambda G'^T - H'R.$$

By inserting this into (36), we obtain a necessary condition given by

$$\Gamma G' \Lambda G'^T R^{-1} = H'^T H'. \tag{37}$$

First we shall discuss the optimal encoder when the channel noise components are uncorrelated but unequal. Subsequently we shall present a numerical method for the general case in which the channel noise components are correlated.

A. Uncorrelated Noise Channel

When the channel noise components are uncorrelated or

$$R = N = \operatorname{diag}(n_1, n_2, \cdots, n_k)$$

then the signal-to-noise ratio at the ith subchannel is given by P_i/n_i. Consider a vector channel whose channel noise covariance matrix is the unity matrix and the subchannel power is constrained by P_i/n_i for $i = 1, 2, \cdots, k$.

We will call this channel the equivalent channel. We see that the performance of this equivalent channel and of the original channel can be the same. The mean-square-error of the original channel is, from (16),

$$\mathcal{E} = \operatorname{tr}\,[\Lambda - \Lambda G'^T (G'\Lambda G'^T + N)^{-1} G'\Lambda]. \tag{38}$$

Denoting $G = N^{-1/2}G'$, we have

$$\mathcal{E} = \operatorname{tr}\,[\Lambda - \Lambda G^T (G\Lambda G^T + I)^{-1} G\Lambda]. \tag{39}$$

This is the mean-squared error of the equivalent channel whose subchannel power is constrained by P_i/n_i, and the solution to the original coding problem is $G' = N^{1/2}G$.

Hence, from now on, we consider the equivalent channel with the modified subchannel power constraint. For the equivalent channel the necessary condition (37) becomes

$$\Gamma G \Lambda G^T = H^T H \tag{40}$$

where $H = \Lambda G^T (G\Lambda G^T + I)^{-1}$.

In the equivalent channel, the total power is $P = \sum_{i=1}^k P_i/n_i$. Suppose we found the optimal coder under this constraint alone. Then $\Gamma = \gamma I$, and (40) becomes the necessary condition of the total power constraint problem. In this case, we can use the previous results for γ and $\mathcal{E}$. In the total power constraint problem, the encoder matrix G was a diagonal matrix. However, as mentioned before, when the channel noise components are uncorrelated and equal, any orthogonal transformation of the encoded signal does not change the optimality. Let D be a diagonal encoder matrix whose diagonal elements are expressed by (see (33))

$$d_{ii} = \left(\sqrt{\frac{\lambda_i}{\gamma}} - 1\right)^{1/2} \Big/ \sqrt{\lambda_i}, \qquad \text{for } i \leqslant l. \tag{41}$$

Let the diagonal matrix M be

$$M \triangleq D\Lambda D. \tag{42}$$

Suppose we can find an orthogonal transformation matrix Z which satisfies the channel power constraint or

$$ZMZ^T = \begin{bmatrix} P_1/n_1 & * & \cdots & * \\ * & P_2/n_2 & \cdots & \vdots \\ \vdots & & \ddots & * \\ * & \cdots & * & P_k/n_k \end{bmatrix} \triangleq P'. \tag{43}$$

Then the optimum encoder is $G = ZD$. In fact, the matrix Z can be obtained by a numerical inverse Jacobi's method [1] if the following necessary condition is satisfied.

Necessary condition of an inverse eigenvalue problem: If the sum of l largest eigenvalues (or elements of M) is not less than the sum of l largest diagonal elements of P' where $1 \leqslant l \leqslant k - 1$ and $\operatorname{tr} \Lambda = \operatorname{tr} P'$, then there exists an orthogonal matrix such that $ZMZ^T = P'$. (See Appendix B.)

When the necessary condition of an inverse eigenvalue problem is violated, then at least two diagonal elements of Γ must be unequal and our assumption on Γ was wrong. Suppose the diagonal elements of Γ are composed of either γ_1 or γ_2 and $\gamma_1 \neq \gamma_2$. We may rearrange the order of the subchannels so that we can write

$$\Gamma = \begin{bmatrix} \gamma_1 I_1 & 0 \\ 0 & \gamma_2 I_2 \end{bmatrix} \tag{44}$$

where I_1, I_2 are $p \times p$ and $(k - p) \times (k - p)$ unit matrices, respectively. From the necessary condition (40), we know that $G\Lambda G^T$ must be a symmetric matrix. Noting that $\gamma_1 \neq \gamma_2$, $G\Lambda G^T$ must be given in the form of

$$G\Lambda G^T = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix} \tag{45}$$

where S_1 and S_2 are $(p \times p)$ and $(k - p) \times (k - p)$ symmetric matrices. It tells us that the p encoded signals are uncorrelated to the $(k - p)$ remaining encoded signals. If we rearrange the order of components in Λ so that the first p component signals should be transmitted over the first p subchannels, then the necessary condition can be expressed by the partitioned submatrices, i.e.,

$$\Gamma G\Lambda G^T = \begin{bmatrix} \gamma_1 I_1 & 0 \\ 0 & \gamma_2 I_2 \end{bmatrix} \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \cdot \begin{bmatrix} G_1{}^T & 0 \\ 0 & G_2{}^T \end{bmatrix} = \begin{bmatrix} H_1{}^T H_1 & 0 \\ 0 & H_2{}^T H_2 \end{bmatrix}.$$

Hence the system can be partitioned into two groups and the necessary condition for each group is the same as the total power constraint problem. Therefore, G_1 or G_2 could be solved by using the previous results and could be written by $G_1 = Z_1 D_1$ and $G_2 = Z_2 D_2$ where D_1, D_2 are diagonal matrices and Z_1, Z_2 are orthogonal matrices. The error is

$$\mathcal{E} = \operatorname{tr}\left[\Lambda - \Lambda^2 D^2 (D^2 \Lambda + I)^{-1}\right]. \tag{46}$$

Denoting m_i as the elements of a diagonal matrix M where $M = \Lambda D^2$, the error becomes

$$\mathcal{E} = \sum_{i=1}^{k} \frac{\lambda_i}{m_i + 1} + \sum_{i=k+1}^{r} \lambda_i. \tag{47}$$

Up to this point, we assumed that γ_1, γ_2 are known. But actually we do not know γ_1, γ_2 and, even if we know the values of γ_1 and γ_2, still we do not know which components of Λ belong to γ_1. However, we know that the error can be expressed by m_i, and m_i's must satisfy the inverse eigenvalue problem. Now we can present our problem in a different way.

Minimize the function

$$\mathcal{E} = \sum_{i=1}^{k} \frac{\lambda_i}{m_i + 1} + \sum_{i=k+1}^{r} \lambda_i, \qquad \lambda_1 > \lambda_2 > \cdots > \lambda_r \tag{48}$$

subject to the constraints

$$\begin{aligned} m_1 &\geqslant p_1 \\ m_1 + m_2 &\geqslant p_1 + p_2 \\ m_1 + m_2 + m_3 &\geqslant p_1 + p_2 + p_3 \\ &\vdots \\ \sum_{i=1}^{k} m_i &= \sum_{i=1}^{k} p_i . \end{aligned}$$

We have $k - 1$ inequality constraints and one equality constraint. This constrained optimization problem is easily solved by using the Lagrange-multiplier technique [2].

B. Correlated Noise Channel

In this section we present a numerical method to find the optimal channel covariance matrix P. We will also show that the minimum error and the optimal encoder can be derived from the optimal covariance matrix P.

Let P be expressed by

$$P = \begin{bmatrix} P_1 & \rho_1 P_{12} & \rho_2 P_{13} & \cdots \\ \rho_1 P_{12} & P_2 & \rho_3 P_{23} & \cdots \\ \rho_2 P_{13} & \rho_3 P_{23} & P_3 & \cdots \\ \vdots & \vdots & & \ddots & P_k \end{bmatrix} \tag{49}$$

where $P_{ij} = \sqrt{P_i P_j}$ and $|\rho_n| \leqslant 1$ for $n = 1, 2, \cdots, k(k - 1)/2$. The normalized covariance coefficients ρ_n are unknown but the diagonal elements and P_{ij}'s are known by the given power constraints.

Let $\rho(i) = [\rho_1 \rho_2 \cdots]^T$ denote a $k(k - 1)/2$-dimensional vector at the ith iteration. Let $\nabla\mathcal{E}(i)$ be the gradient of error with respect to $\rho(i)$. An unconstrained gradient search algorithm would then be described by

$$\rho(i+1) = \rho(i) - \alpha \nabla \mathcal{E}(i) \tag{50}$$

where $\alpha > 0$ is the step size.

As noted earlier, the components of $\rho(i+1)$ must be bounded at each step between -1 and $+1$. The gradient $\nabla\mathcal{E}$ can be computed by a numerical method:

$$[\nabla \mathcal{E}(i)]_j \approx [\mathcal{E}(i+1) - \mathcal{E}(i)]/\Delta_j \tag{51}$$

where Δ_j is the jth component step size of $\rho(i)$. Here the error $\mathcal{E}(i)$ is the minimum error for the given channel covariance matrix $P(i)$ at the ith iteration.

For P in the form (49), we have to find the optimum encoder G which minimizes $\mathcal{E}$. To do this we write the minimum error:

$$\mathcal{E} = \operatorname{tr}\,[\Lambda - \Lambda G'^T (G' \Lambda G'^T + R)^{-1} G' \Lambda]. \tag{52}$$

Let us make the following transformation:

$$G' = TN^{1/2}ZG \tag{53}$$

where T is an orthogonal matrix, $N^{1/2}$ is a diagonal matrix such that $R = TNT^T$, and Z is an orthogonal matrix such that

$$Z^T N^{-1/2} T^T P T N^{-1/2} Z = M = G \Lambda G^T \tag{54}$$

where M is a diagonal matrix. Using (53), P and R in (52) can be diagonalized simultaneously and the error expression becomes

$$\begin{aligned}\mathcal{E} &= \operatorname{tr}\,[\Lambda - \Lambda G^T Z^T N^{1/2} T^T (T N^{1/2} Z G \Lambda G^T Z^T N^{1/2} T^T \\ &\quad + T N^{1/2} Z Z^T N^{1/2} T^T)^{-1} T N^{1/2} Z G \Lambda] \\ &= \operatorname{tr}\,[\Lambda - \Lambda G^T (G \Lambda G^T + I)^{-1} G \Lambda] \\ &= \operatorname{tr}\,[\Lambda - \Lambda G^T (M + I)^{-1} G \Lambda]\end{aligned}$$

or

$$\mathcal{E} = \operatorname{tr}\,[\Lambda - G \Lambda^2 G^T (M + I)^{-1}]. \tag{55}$$

Consider a vector channel whose channel noise covariance matrix R is I and whose signal covariance matrix P is characterized by a set of eigenvalues which are diagonal elements of M, which at the given stage of iteration are known. The error of this special system is also expressed by (55). To minimize this error we define the unconstrained Lagrangian function as

$$\begin{aligned}L &= \operatorname{tr}\,[\Lambda - (M+I)^{-1} G \Lambda^2 G^T] + \gamma_1 (g_1 \Lambda g_1{}^T - m_1) + \cdots \\ &\quad + \gamma_k (g_k \Lambda g_k{}^T - m_k)\end{aligned} \tag{56}$$

where g_i is the ith row vector of G and $\gamma_1, \cdots, \gamma_k$ are Lagrange multipliers.

We differentiate L with respect to the components of G, obtaining

$$\frac{\partial L}{\partial G} = \mathbf{0} = -2(M+I)^{-1} G \Lambda^2 + 2 \Gamma G \Lambda \tag{57}$$

where $\Gamma = \operatorname{diag}(\gamma_1, \cdots, \gamma_k)$.

Since $(M+I)\Gamma$ is a diagonal matrix, the column vectors of G are eigenvectors of $(M+I)\Gamma$ which have at most one nonzero element. We see that

$$G \Lambda^2 G^T = (M+I) \Gamma M. \tag{58}$$

Therefore $G\Lambda^2 G^T$ is a diagonal matrix.

By the same argument given in Section V and assuming that $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k \geqslant \cdots \geqslant \lambda_r$ and $m_1 \geqslant m_2 \geqslant \cdots \geqslant m_k$, we have

$$G \Lambda^2 G^T = \Lambda_k M \tag{59}$$

where Λ_k is a $k \times k$ submatrix of Λ or $\Lambda_k = \operatorname{diag}(\lambda_1, \cdots, \lambda_k)$. We have the minimum $\mathcal{E}$ for the given $P(i)$ as

$$\begin{aligned}\mathcal{E}(i) &= \operatorname{tr}\,[\Lambda - \Lambda_k M (M+I)^{-1}] \\ &= \sum_{i=1}^{k} \frac{\lambda_i}{m_i + 1} + \sum_{i=k+1}^{r} \lambda_i.\end{aligned} \tag{60}$$

Here the error $\mathcal{E}$ is expressed by eigenvalues of $N^{-1/2}T^T \cdot PTN^{-1/2}$. Thus at each step, the eigenvalues of $N^{-1/2}T^T \cdot PTN^{-1/2}$ must be computed. If we choose the components of the encoder matrix G' as unknown variables, we have $k \times k$ unknown variables. However, using ρ as the unknown vector, we have $k(k-1)/2$ unknown variables and the complexity of the numerical calculation is greatly reduced. As an example, if $k = 2$, we have 4 unknowns in G'. But using the scalar ρ as the unknown variable, we have only one unknown.

VII. CONCLUDING REMARKS

In this paper we have investigated the optimum linear coding problem for the vector signal transmission system. The eigenvalues of the signal covariance matrix and the noise covariance matrix together with a performance weighting matrix W determine the optimum coder design. An inverse eigenvalue problem must be considered if the individual subchannel powers are constrained.

The vector coding process can also be applied to a scalar signal transmission system if we allow a certain amount of time delay. As an example, we may take r samples from the sequence of the source signal and transmit it by k encoded signals in a sequence. In this case, the source signal can be considered as an r-dimensional vector and the channel can be considered as a k-dimensional channel.

A number of possible applications have been examined in [1] and many more remain to be investigated.

APPENDIX A

We will give here the proofs of properties of G which are used in Section V.

Proof of Property 1

From (22)

$$\gamma G\Lambda G^T = H^T HN.$$

Note that the matrix $G\Lambda G^T$ is symmetric. Suppose $H^T H$ is a nondiagonal symmetric matrix. Since N is diagonal and no two elements are equal, $H^T HN$ cannot be a symmetric matrix. Therefore, $H^T H$ must be a diagonal matrix. Since $H^T HN$ is diagonal, $G\Lambda G^T$ is also a diagonal matrix.

From (23) we obtain

$$G\Lambda^2 G^T = \gamma(G\Lambda G^T + N)G\Lambda G^T N^{-1}(G\Lambda G^T + N).$$

Since $G\Lambda G^T$ is diagonal, the right side of this equation is diagonal. Thus $G\Lambda^2 G^T$ is diagonal.

Proof of Property 2

Substituting (18) into (19), we have

$$(G\Lambda G^T + N)^{-1} G\Lambda^2 G^T (G\Lambda G^T + N)^{-1} G + \gamma G$$
$$= (G\Lambda G^T + N)^{-1} G\Lambda.$$

Premultiplying by $(G\Lambda G^T + N)$ yields

$$[G\Lambda^2 G^T (G\Lambda^2 G^T + N)^{-1} + \gamma(G\Lambda G^T + N)]\, G = G\Lambda$$

or

$$DG = G\Lambda.$$

It is clear that the nonzero column vectors of G are the eigenvectors of D. Since every element of the diagonal matrix D is an eigenvalue of D, the elements of D must also appear in Λ if $k \leqslant r$ and vice versa. We note that D is a diagonal matrix because $G\Lambda G^T$ and $G\Lambda^2 G^T$ are diagonal matrices.

Proof of Property 3

We will assume that no two elements of the diagonal matrices D or Λ are the same. (If D or Λ has two identical diagonal elements, then we assume that these two elements differ by a small quantity ϵ.) The eigenvector corresponding to the ith diagonal element (eigenvalue) is parallel to the ith coordinate axis. Thus the ith component of the column vector is nonzero but other elements are all zero. Since $\Lambda G^T = G^T D$, the second statement is also true.

APPENDIX B

In this section we prove the necessary condition of an inverse eigenvalue problem which is used in Section IV.

Theorem

Suppose that a symmetric matrix P is diagonalized by the orthogonal matrix Z, so that $P = ZMZ^T$ where M is diagonal. Then the sum of the l largest diagonal elements of P cannot exceed the sum of its l largest eigenvalues (or elements of M).

Proof

From $P = ZMZ^T$, $p_{ii} = \Sigma_{j=1}^{k} z_{ij}^2 m_{jj}$ and $\Sigma_{i=1}^{l} p_{ii} = \Sigma_{j=1}^{k} (\Sigma_{i=1}^{l} z_{ij}^2) m_{jj}$. Let $m_{11} \geqslant m_{22} \geqslant \cdots \geqslant m_{kk}$ and $p_{11} \geqslant p_{22} \geqslant \cdots \geqslant p_{kk}$. Since $\Sigma_{i=1}^{l} z_{ij}^2 \leqslant 1$, the maximum possible value of $\Sigma_{i=1}^{l} p_{ii}$ is equal to $\Sigma_{j=1}^{l} m_{jj}$ if $(\Sigma_{i=1}^{l} z_{ij}^2) = 1$ for $j \leqslant l$; otherwise $\Sigma_{i=1}^{l} p_{ii} < \Sigma_{j=1}^{l} m_{jj}$ (c.f. [1, pp. 40-44]).

REFERENCES

[1] K. H. Lee, "Optimal linear coding for a multichannel system," Ph.D. dissertation, Univ. of New Mexico, Albuquerque, 1975 (University Microfilm is available).

[2] B. S. Gottfried and J. Weisman, *Introduction to Optimization Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1973, p. 49.

[3] D. P. Petersen, "Reexamination of the prefiltering problem in a state-variable formulation," *IEEE Trans. Inform. Theory,* vol. IT-16, pp. 480-484, July 1970.

[4] —, "A proof of the separability of prefiltering and coding," *IEEE Trans. Inform. Theory,* vol. IT-21, pp. 570-572, Sept. 1975.

[5] I. Satake, *Linear Algebra.* New York: Marcel Dekker, 1975, p. 37.

★

Kyong-Hwa Lee (S'72-M'75) was born in Masan, Korea, on February 7, 1937. He received the B.S. degree in electronic engineering from Seoul National University, Seoul, Korea, in 1962, and the M.S. and Ph.D. degrees in electrical engineering from the University of New Mexico, Albuquerque, in 1971 and 1975, respectively.

From 1964 to 1969 he was an Instructor at the Telecommunications Training Center, Seoul, Korea. From 1969 to 1975 he was a Graduate and Teaching Assistant in the Department of Electrical Engineering and Computer Science, University of New Mexico. Since 1975 he has been with the Data Systems Division, Gulton Industries, Albuquerque, NM, where he is involved in analog and digital signal processing.

★

Daniel P. Petersen (M'53-SM'56) was born in Buffalo, NY, on September 27, 1923. He received the B.M.E. and D.Eng.Sc. degrees from Rensselaer Polytechnic Institute (RPI), Troy, NY, in 1947 and 1963, respectively, and the M.S.M.E. degree from the Massachusetts Institute of Technology, Cambridge, MA in 1948.

He served as an Aviation Weather Officer during World War II and in various engineering capacities with Sperry Gyroscope Company, W. L. Maxson Corp., Allen D. Cardwell Company, U.S. Time Corp., and United Aircraft Corp. During the period from 1955 to 1965 he taught courses in system theory as an Adjunct Lecturer for the University of Connecticut and as Adjunct Associate Professor for the RPI Hartford Graduate Center. In 1965 he joined the University of New Mexico as an Associate Professor and is now Professor of Electrical Engineering and Computer Science. During 1972-1973 he served as Visiting Scientist at the National Center for Atmospheric Research, Boulder, CO, studying the design of observation and data assimilation schemes for the Global Atmospheric Research Program. He has served as a consultant for NSF, U.S. Air Force, U.S. Weather Bureau, NASA, and several industrial and research organizations.

Dr. Petersen is a member of the American Society for Engineering Education and the National Society of Professional Engineers. He is also a member of the Committee on Probability and Statistics of the American Meteorological Society, and a reviewer for several professional journals.