OPTIMUM LINEAR CODING FOR ADDITIVE NOISE SYSTEMS USING INFORMATION FEEDBACK

Thesis by

Stanley Butman

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1967

(Submitted May 24, 1967)

ACKNOWLEDGEMENT

The author wishes to thank Professor Thomas L. Grettenberg for his advice, guidance, and many helpful discussions prior to, and during the course of this work.

The research described in this report was conducted during the tenure of a Radio Corporation of America Fellowship. The author is grateful for this assistance. A portion of this work was also supported by Naval Ordnance Test Station, Pasadena Annex Contract No. N00123-67-C-1188.

The author also wishes to acknowledge several helpful discussions with Dr. Aaron Wyner of the Bell Telephone Laboratories, Professor R. Sridhar and Mr. S. Farber. Finally, the skillful typing of Mrs. Doris Schlicht is greatly appreciated.

ABSTRACT

The feedback coding problem for Gaussian systems in which the noise is neither white nor statistically independent between channels is formulated in terms of arbitrary linear codes at the transmitter and at the receiver. This new formulation is used to determine a number of feedback communication systems. In particular, the optimum linear code that satisfies an average power constraint on the transmitted signals is derived for a system with noiseless feedback and forward noise of arbitrary covariance. The noisy feedback problem is considered and signal sets for the forward and feedback channels are obtained with an average power constraint on each. The general formulation and results are valid for non-Gaussian systems in which the second order statistics are known, the results being applicable to the determination of error bounds via the Chebychev inequality.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. INTRODUCTION

Two-way communication systems have the capability of transmitting information about the current status of a message being decoded at the receiver back to the transmitting point. The returned information can be used to simplify the coding and decoding operations in the forward channel and to provide a lower probability of error for a given coding delay than could be achieved without feedback. A potentially useful application of information feedback is in the design of efficient data retrieval systems for space vehicles, where the transmitting power is restricted to be several orders of magnitude less than the transmitting power of the ground based receiving equipment.

The analysis of a feedback communication system is similar to the one-way communication problem in that it can be separated into a decision or decoding problem and a signal selection or coding problem. It differs only in the sense that in the coding problem it is possible to optimize over both the forward and feedback signal sets. Previous authors have approached the feedback communication problem by assuming a specific functional relationship between the feedback signals and the receiver's estimate of the message, a functional form for the decision procedure, and solving the remaining signal selection problem for the forward signal set. This approach [6,14,17,19] and other methods [1-19] have succeeded in developing a number of efficient feedback communication schemes, mainly for the additive white Gaussian noise (AWGN) channel with a noiseless feedback link and the binary symmetric (BS) channel with a noiseless BS feedback path [13]. However, because of the structural assumptions in these schemes, the optimum linear feedback system still remains to be determined even for the AWGN channel with a noiseless return link. Systems in which the noise is not white have received only passing mention. Attempts to take feedback noise into account [17,18] by using a Kalman filter [20] at the transmitter have not used an optimum decoding procedure nor an optimum set of feedback signals.

In this paper, the feedback communication problem for Gaussian systems in which the noise is neither white nor statistically independent between channels is formulated in terms of arbitrary linear codes at the transmitter and at the receiver. The maximum likelihood decision rule, which is optimum for an equiprobable message source, is determined and the signal selection problem is posed for both the forward and feedback signal sets. This new formulation, is developed in Chapter II, and is used to determine a number of feedback communication systems. In Chapter III the optimum linear code is derived for a system with a noiseless feedback channel and an average power constraint on the transmitted signals. The noisy feedback problem is considered in Chapter IV where signal sets for the forward and feedback channels are obtained with an average power constraint on each.

Chapter V considers the use of Kalman filtering at the transmitter combined with an optimization at the receiver.

The present approach is valid for non-Gaussian systems provided second order statistics are available, the results being applicable for the determination of error bounds by such methods as the Chebychev inequality.

The general formulation may be classified as a fixed-time-of-decision or block-coding system in opposition to sequential-decision systems in which the time-of-decision is a random variable. This is the fundamental dichotomy which separates all of the feedback communication systems reported in the literature. In each class, information feedback, if it is sufficiently accurate, can provide improved performance. A special case of the latter category is when feedback is used to inform the transmitter only of the event that a decision has been made so that a new message may be initiated. This has been generally referred to as decision-feedback and was studied by Bloom, Chang, Harris, Hauptschein, Metzner, Morgan, Schwartz, and more recently by Viterbi [5-9,12].

They consider the transmission of binary messages using signals that are also binary (two-levels) over the AWGN channel with a noiseless feedback link. Viterbi also considers the M-ary case and uses M orthogonal binary signals. He obtains exponential bounds for the error probability and shows that the negative exponent is four times the exponent for the best available error bound on the one way channel when the rate of transmission exceeds half the channel capacity.

Sequential-decision systems using information feedback to continually inform the transmitter of the state of the receivers knowledge (or uncertainty) of the message being sent have been investigated by Horstein [13] in the case of a BS channel with a BS noiseless feedback link, and by Turin [14] and Horstein [19] for the AWGN channel with a noiseless feedback link. Ideally, in sequential decision the receiver updates the a posteriori probability over the message set as new signals are received and selects the message whose a posteriori probability, relative to the other messages, is first to exceed a threshold. The thresholds are set by the desired probability of error. Continually returned feedback information allows the transmitter to select signals that will maximize the a posteriori probability of the message being sent and informs the transmitter when a decision has been made.

The approach in [14] and [19] is based on the continuous time channel and makes the assumption of instantaneous feedback. A binary message source is used and the likelihood function (the logarithm of the ratio of the two a posteriori probabilities) is continuously computed from the received time function. The transmitted signal is a linear function of the message being sent (0 or 1) and the current value of the likelihood function available from feedback. The evolution of the likelihood function is governed by a Langevin differential equation which is driven by White Gaussian noise. Thus, the likelihood function is a continuous Markov process whose probability density satisfies a Fokker-Plank partial differential equation. A decision is made the instant that the likelihood function first crosses one of two thresholds. The time-of-decision, and hence duration of a message is a random variable. The transmitted signals are constrained in peak and average power and are chosen by Turin to minimize the average duration of a message. The result is

$$\overline{T} = \frac{N_0}{P} \ln 2 \qquad \text{seconds/bit}$$

The probability of error vanishes when the bandwidth is infinite and the peak power constraint is removed, so that a rate $\frac{1}{\overline{T}} = \frac{P}{N_0}$ nats/second which is equal to the capacity of the infinite bandwidth AWGN channel is achieved.

Horstein's earlier work [13] on the BS channel is similar in the sense that the number N of binary channel symbols per binary message takes the place of T. N is a random variable for which a bound on the mean $\overline{N}$ is found as function of the rate R, channel capacity C and probability of error Pe.

The operation of a block-system is based on the principle that the number or block of signals associated with each message is a deterministic quantity. The block length and the instants at which signals are transmitted are known to the receiver. The receiver may compute the a posteriori probability over the message set either continually as new signals arrive or after the entire block is received. However, the decision is made only after a complete block has been received so that the decision time is deterministic and decision feedback alone is of no use. On the other hand, information feedback that is continually provided to the transmitter allows the transmitter to select signals that will maximize the a posteriori probability of the message being sent. This approach has been used by Elias [10], Schalkwijk and Kailath [15,16] and Omura [17] for the AWGN channel with noiseless feedback and an equiprobable M-ary message set. Their results achieve the finite and infinite bandwidth capacity limit of Shannon [1] when an average power constraint is imposed, but are not optimum because the coding in [15,16] is not optimum while in [17] the maximum likelihood rule is not used.

Although some authors evaluate the performance of their noiseless feedback codes in the presence of feedback noise, procedures specifically designed to minimize feedback noise were not available prior to the work of Omura and Kashyap [18]. They use a Kalman filter at the transmitter to form the best estimate of the receiver's "state". However, the choice of feedback signals and decision procedure is not optimum.

II. FORMULATION OF THE FEEDBACK COMMUNICATION PROBLEM

2.1. Introduction.

In this chapter the feedback communication problem for a system in which the forward and return channels are both corrupted by additive Gaussian noise is formulated using arbitrary linear operations at the transmitter and at the receiver. This kind of an approach is applicable to situations that are more general than systems with Gaussian noise. However, the Gaussian assumption (and the assumption of linearity) allows simple closed form expressions for the optimum decision rule, and the probability of error, to be determined. In order to set up the signal selection problem, the above expressions are augmented by equations for the average energy transmitted in the forward channel and the average energy required to send feedback information.

2.2. Description of the General Linear Coding Procedure.

A linear feedback communication system using a sequence of N signals to transmit a message θ is shown below

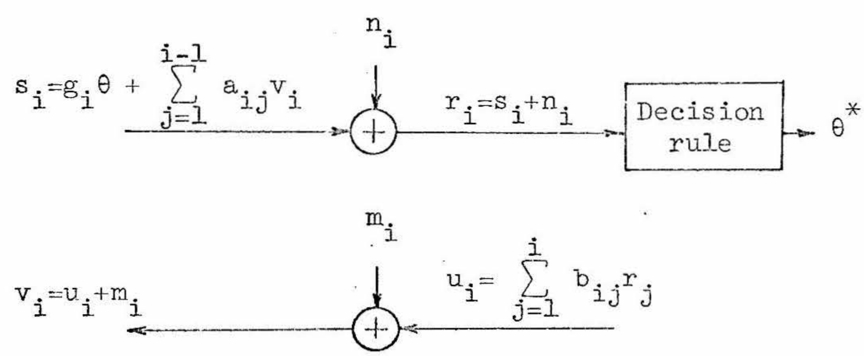

Fig. 1. Linear Feedback Communication System.

Only the discrete version of an additive noise channel is considered. The connection between the discrete and continuous formulations is well known [21-23] and will not be discussed further. The process begins with $s_1 = g_1\theta$ being transmitted, $r_1 = g_1\theta + n_1$, being received, $u_1 = b_{11}r_1$ is the first feedback signal which is observed at the receiver as $v_1 = u_1 + m_1$. The next signal to be sent can be a linear function of θ and v_1, and is thus written as $s_2 = g_2\theta + a_{21}v_1$. The general term is

$$s_i = g_1\theta + \sum_{j=1}^{i-1} a_{ij}v_j \qquad i = 1, 2, \cdots, N \qquad (2.1)$$

and

$$u_i = \sum_{j=1}^{i} b_{ij}r_j \qquad i = 1, 2, \cdots, N-1 \qquad (2.2)$$

The last, or N-th, feedback signal is not used in s_N and is therefore not fed back, nor is it generated.

It is convenient to write the above and remaining analysis using vector and matrix notation. Therefore, let $s = \mathrm{col}(s_1, s_2, \cdots, s_N)$, $r = \mathrm{col}(r_1, \cdots, r_N)$, $u = \mathrm{col}(u_1, u_2, \cdots, u_N)$, $v = \mathrm{col}(v_1, v_2, \cdots, v_N)$, $n = \mathrm{col}(n_1, n_2, \cdots, n_N)$, $m = \mathrm{col}(m_1, m_2, \cdots, m_N)$, $g = \mathrm{col}(g_1, g_2, \cdots, g_N)$ and let $(A)_{ij} = a_{ij}$, $(B)_{ij} = b_{ij}$ be $N \times N$ lower triangular matrices with the main diagonal of A and the N-th row of B identically zero. Then

$$u = Br \tag{2.3}$$

$$v = u + m \tag{2.4}$$

$$s = g\theta + Av \tag{2.5}$$

$$r = s + n \tag{2.6}$$

Note that A anihilates the N-th component of v, while the zero N-th row of B causes u_N to be zero as required. The system is now equivalent to the N-dimensional vector channel below

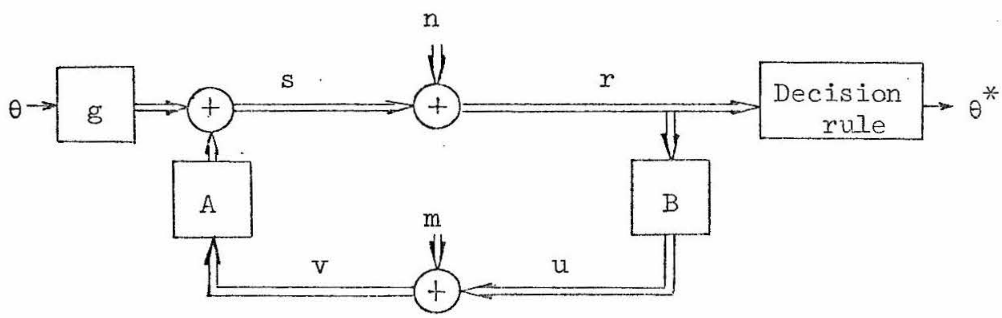

Fig. 2. Matrix Formulation of the Feedback Communication Process.

Now let m and n be jointly normal with zero-mean, covariance K_m and K_n, and cross-covariance K_{mn}, where $K_m = E[mm^T]$, $K_n = E[nn^T]$, $K_{mn} = E[mn^T]$ and $E[\cdot]$ is the expectation operator while "T" denotes transpose. The conditional probability density $p(r/\theta)$ can be found after r is written as a function of the random variables θ, m and n. This is done by substituting (2.4) into (2.5) and the result into (2.6)

$$r = (I-AB)^{-1}(g\theta + Am + n) . \tag{2.7}$$

The conditional mean $E[r/\theta]$ and covariance $E[(r-E[r/\theta])(r-E[r/\theta])^T/\theta]$ are respectively

$$\bar{r}_\theta = (I-AB)^{-1} g\theta \tag{2.8}$$

and

$$K_r = (I-AB)^{-1} K(I-AB)^{-T} \tag{2.9}$$

where,

$$K = AK_m A^T + AK_{mn} + K_{mn}^T A^T + K_n \ . \tag{2.10}$$

Then

$$p(r/\theta) = [(2\pi)^N \det K_r]^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\|r-\bar{r}_\theta\|^2_{K_r^{-1}}] \tag{2.11}$$

or since $\|r-\bar{r}_\theta\|^2_{K_r^{-1}} = \|(I-AB)r-g\theta\|^2_{K^{-1}}$ and $Br = u$,

$$p(r/\theta) = [(2\pi)^N \det K]^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\|r-Au-g\theta\|^2_{K^{-1}}] \tag{2.12}$$

where $\det K_r = \det K$ because $\det(I-AB)$ is unity, a result that follows from the fact that AB is lower triangular with zeros along the main diagonal so that $I - AB$ is lower triangular with ones down the main diagonal.

2.3. Derivation of the Maximum Likelihood Decision Rule.

In general, the optimum decision procedure for minimum error is the ideal observer rule. The procedure is for the receiver to select the message that maximizes the a-posteriori probability distribution $p(\theta/r)$. If $p(\theta)$ is the a-priori probability distribution on the message set, then Bayes' rule gives

$$p(\theta/r) = \frac{p(r/\theta)p(\theta)}{p(r)} \tag{2.13}$$

When θ is equiprobable over a finite set of M real points, $p(\theta) = \frac{1}{M}$, then

$$p(\theta/r) = \frac{p(r/\theta)}{M\,p(r)} \quad . \tag{2.14}$$

Maximization of $p(\theta/r)$ is equivalent to maximization of $p(r/\theta)$, which is the maximum likelihood rule. $p(r/\theta)$ is maximized over θ if and only if $\|r-\bar{r}_\theta\|^2_{K_r^{-1}}$ is minimized. Let $\hat{\theta}_N$ be an arbitrary scalar

$$\begin{aligned}\|r-\bar{r}_\theta\|^2_{K_r^{-1}} &= \|r-Au-g\theta\|^2_{K^{-1}} \\ &= \|r-Au-g\hat{\theta}_N + g(\hat{\theta}_N-\theta)\|^2_{K^{-1}} \\ &= \|r-Au-g\hat{\theta}_N\|^2_{K^{-1}} - 2(\hat{\theta}_N-\theta)(\hat{\theta}_N\|g\|^2_{K^{-1}} - \langle g, K^{-1}(r-Au)\rangle) \\ &\qquad + (\hat{\theta}_N-\theta)^2\|g\|^2_{K^{-1}}\end{aligned} \tag{2.15}$$

The middle term drops out if $\hat{\theta}_N$ is chosen to be

$$\hat{\theta}_N = \frac{\langle g, K^{-1}(r-Au)\rangle}{\|g\|^2_{K^{-1}}} \tag{2.16}$$

Therefore

$$\|r-\bar{r}_\theta\|^2_{K_r^{-1}} = \|r-Au-g\hat{\theta}_N\|^2_{K^{-1}} + (\hat{\theta}_N-\theta)^2\|g\|^2_{K^{-1}} \quad . \tag{2.17}$$

Since $\hat{\theta}_N$ can take on any value on the real line while θ is only one of a set of M discrete points, it is now obvious that choosing θ closest to $\hat{\theta}_N$ maximizes $p(r/\theta)$ over θ. This is the maximum likelihood estimate of θ and is denoted by θ^*, it of course runs only over the finite set of M message points. It is easy to show that $\hat{\theta}_N$ is in fact the minimum-variance unbiased linear estimate of θ. The conditional mean and variance of $\hat{\theta}_N$ are

$$E[\hat{\theta}_N/\theta] = \theta \tag{2.18}$$

$$E[(\hat{\theta}_N-\theta)^2/\theta] = \frac{1}{\|g\|^2_{K^{-1}}} \tag{2.19}$$

2.4. The Probability of Error.

An error occurs at the receiver each time θ is transmitted but $\theta^* \neq \theta$, that is, $|\hat{\theta}_N-\theta|$ is not a minimum. If the M equiprobable messages are equispaced on the interval $[-L, L]$ on the real line, the nearest neighbor distance is $L/(M-1)$ and the condition for an error is $|\hat{\theta}_N-\theta| \geq L/(M-1)$ when θ is one of the $M-2$ interior points of

$[-L, L]$. If θ is one of the end points, $\pm L$, the condition is $\hat{\theta}_N - \theta \lesseqgtr \mp L/(M-1)$ respectively. The conditional error probability for $\theta \neq \pm L$ is $P_e = \mathrm{pr}\{|\hat{\theta}_N - \theta| \geq \frac{L}{M-1}/\theta\}$. Thus,

$$P_e = 1 - \int\limits_{|\hat{\theta}_N - \theta| \leq L/(M-1)} dp(\hat{\theta}_N/\theta) \tag{2.20}$$

$$= \mathrm{erfc}\sqrt{\frac{3\sigma_\theta{}^2 \|g\|^2_{K^{-1}}}{2(M^2-1)}} \tag{2.21}$$

where

$$\mathrm{erfc}\ x = \frac{2}{\sqrt{x}} \int_x^\infty e^{-x^2}\, dx \tag{2.22}$$

$$p(\hat{\theta}_N/\theta) = \sqrt{\frac{\|g\|^2_{K^{-1}}}{2\pi}}\ e^{-\frac{1}{2}(\hat{\theta}_N - \theta)^2 \|g\|^2_{K^{-1}}} \tag{2.23}$$

and,

$$\sigma_\theta{}^2 = E[\theta^2] \qquad (E[\theta] = 0)$$

$$= \frac{L^2(M+1)}{3(M-1)} \tag{2.24}$$

When θ is one of the end points, P_e is slightly lower but negligibly so, therefore, the average error equals the conditional error above.

2.5. The Signal Selection Problem.

The signal selection problem is to choose A, B and g to minimize P_e subject to constraints on the forward and feedback signal sets. Important constraints are the average power in the forward and feedback channels, or equivalently, the average energy per transmitted message in the forward and feedback channels.

$$E_{av} = E \sum_{i=1}^{N} s_i^2 \tag{2.25}$$

$$= E[\mathrm{Tr}\ ss^T]$$

$$= \mathrm{Tr}[(I-AB)^{-1}(\sigma_\theta gg^T + AK_m A^T + 2AK_{mn}B^T A^T + ABK_n B^T A^T)(I-B^T A^T)^{-1}] \tag{2.26}$$

and, since θ is statistically independent of n and m,

$$E_{fb} = E \sum_{i=1}^{N-1} u_i^2 \quad ; \quad (u_N \equiv 0) \tag{2.27}$$

$$= E[\mathrm{Tr}\ Brr^T B^T]$$

$$= \mathrm{Tr}[B(i-AB)^{-1}(\sigma_\theta^2 gg^T + K)(I-B^T A^T)^{-1} B^T] \tag{2.28}$$

where $\mathrm{Tr}[\cdot]$ is the trace operator defined by $\mathrm{Tr}[Q] = \sum_{i=1}^{N} q_{ii}$.

Let $N = 2TW$ where T is the duration of the message, or coding delay, and W, which is defined here to be N/2T, is the "bandwidth" of the forward channel. Note that the time duration of the (N-1) feedback signals is 2W/(N-1) so that the average power in the forward channel is

$$P = \frac{2W}{N} E_{av} \tag{2.29}$$

while

$$P_{fb} = \frac{2W}{N-1} E_{fb} \tag{2.30}$$

is the feedback power.

Other constraints may be either substituted for, or added to, the above conditions. Since erfc x is a monotonically decreasing function of x, minimization of P_e is equivalent to maximization of $\sigma_\theta^2 \|g\|^2_{K^{-1}}$ and hence $1 + \sigma_\theta^2 \|g\|^2_{K^{-1}}$.

III. NOISELESS FEEDBACK

3.1. Introduction.

This chapter presents a number of new and interesting results for noiseless feedback systems. Primarily, the sequential form of the optimum operation of the receiver and transmitter is derived for a general forward covariance matrix. Secondly, Theorem 2 proves that it is possible to achieve capacity for the wideband and finite bandwidth AWGN channel in an essentially uncountable number of non-optimum ways. Thirdly, it is shown that Schalkwijk's scheme follows from the solution of the signal selection problem with additional constraints. Fourth, it becomes evident that the dynamic programming approach of Omura, for which it was necessary to assume the functional form of the receiver, uses the optimum signal set, but not the optimum decision rule. The discrepancy disappears in the limit as the block length goes to infinity because the minimum variance estimate is asymptotically unbiased. Finally, an almost optimum code for a channel with first order Markov noise is obtained. The critical rate of this code achieves the theoretical capacity when the bandwidth is large, and almost the capacity when the bandwidth is finite.

A model that assumes a noiseless feedback channel may be used to represent a system in which the noise in the forward channel predominates. Calculations based on this assumption are valid until the cumulative effect of the small feedback noise becomes comparable to the forward noise.

The absence of feedback noise is reflected in the general formulation by the vanishing of the noise vector m, the covariance K_m and the cross-covariance K_{mn}. As a result, equations (2.10), (2.26) and (2.28) simplify to

$$K = K_n \tag{3.1}$$

$$E_{av} = \mathrm{Tr}[(I-AB)^{-1}(\sigma_\theta^2 gg^T + ABKB^TA^T)(I-B^TA^T)^{-1}] \tag{3.2}$$

and

$$E_{fb} = \mathrm{Tr}[B(I-AB)^{-1}(\sigma_\theta^2 gg^T + K)(I-B^TA^T)^{-1}B^T] \tag{3.3}$$

However, E_{fb} can be arbitrarily small for any choice of E_{av} and P_e, because B may be scaled down to ϵB while A is scaled by $\frac{1}{\epsilon}$. Thus, the product AB remains unchanged, E_{av} is unaffected, $\|g\|^2_{K^{-1}}$ and hence P_e is not affected, but E_{fb} is scaled down by ϵ^2 which can be made arbitrarily small. The signal selection problem for the noiseless feedback case is therefore not constrained by feedback power. The problem of maximizing $\|g\|^2_{K^{-1}}$ for a fixed value of E_{av} is equivalent to minimizing E_{av} for fixed $\|g\|^2_{K^{-1}}$. Note that only the product matrix AB and g need be found. When there is feedback noise it will be necessary to solve for both A and B.

3.2. Selection of the Coding Matrix AB.

Let $I + C = (I-AB)^{-1}$ then $C = (I-AB)^{-1}AB = AB(I-AB)^{-1}$ is lower triangular with zeros along the main diagonal, note that $AB = C(I+C)^{-1}$.

$$E_{av} = \sigma_\theta^2\|(C+I)g\|^2 + \mathrm{Tr}[CKC^T] \tag{3.4}$$

The $N(N-1)/2$ non-zero elements of C are arbitrary because the $N(N-1)/2$ non-zero elements of AB are arbitrary. The constraint on $\|g\|^2_{K^{-1}}$ does not affect the choice of C, therefore, the minimization over the elements of C can be performed using ordinary calculus for all values of g. Subsequently, when E_{av} is found in terms of g, its optimization over g will have to include the constraint $\|g\|^2_{K^{-1}}$.

Let $c(i-1) = \text{col}(c_{i1}, c_{i2}, \cdots, c_{ii-1})$ denote the first $i-1$ elements of the i-th row of C, the remaining elements in the row are zero because of the lower triangular zero-diagonal form of C. Also, let $g(i-1) = \text{col}(g_1, g_2, \cdots, g_{i-1})$ then

$$E_{av} = \sum_{i=1}^{N} [\sigma_\theta^2(g_i + \langle c(i-1), g(i-1)\rangle)^2 + \langle c(i-1)K(i-1)c(i-1)\rangle] \tag{3.5}$$

where $K(i-1) = K_n(i-1) = E[n(i-1)n^T(i-1)]$, $n(i-1) = \text{col}(n_1, n_2, \cdots, n_{i-1})$. Note, that the individual signal energies, $e_i = E[s_i^2]$ are given by

$$e_i = \sigma_\theta^2(g_i + \langle c(i-1), g(i-1)\rangle)^2 + \langle c(i-1)K(i-1)c(i-1)\rangle. \tag{3.6}$$

Now, setting $\text{grad}_{c(i-1)} E_{av} = 0$ for $i = 2, 3, \cdots, N$ $(c(1) \equiv 0)$ gives

$$c(i-1) = -\sigma_\theta^2(g_i + \langle c(i-1), g(i-1)\rangle)K^{-1}(i-1)g(i-1) . \tag{3.7}$$

Taking the inner product of both sides with $g(i-1)$ and solving for $\langle c(i-1), g(i-1)\rangle$ in terms of g_i and $\|g(i-1)\|^2_{K^{-1}(i-1)}$ gives

$$c(i-1) = - \frac{\sigma_\theta^2 g_i}{1 + \sigma_\theta^2 \|g(i-1)\|^2_{K^{-1}(i-1)}} K^{-1}(i-1)g(i-1) \ . \qquad (3.8)$$

Therefore,

$$e_i = \frac{\sigma_\theta^2 g_i^2}{1 + \sigma_\theta^2 \|g(i-1)\|^2_{K^{-1}(i-1)}} \qquad i = 1, 2, \cdots, N \qquad (3.9)$$

and

$$E_{av} = \sum_{i=1}^{N} \frac{\sigma_\theta^2 g_i^2}{1 + \sigma_\theta^2 \|g(i-1)\|^2_{K^{-1}(i-1)}} \ . \qquad (3.10)$$

It remains to minimize E_{av} with respect to g which is constrained by $\|g\|^2_{K^{-1}}$. This is accomplished by setting $\frac{\partial}{\partial g_i}(E_{av} - \nu\|g\|^2_{K^{-1}}) = 0$ and solving the resulting N non-linear equations for the g_i's. The solution when $K = K_n$ is diagonal is straight forward and is available in closed form. When K is not diagonal the problem is more involved. However, it is not difficult to obtain a good, although not necessarily optimum, choice for g. Also, the functional form of the transmitted signals and a sequential form for the computation of $\hat{\theta}_N$ at the receiver can be derived without knowledge of g.

3.3. General Form of the transmitted Signals.

From $s = (I-AB)^{-1} g\theta + (I-AB)^{-1}ABn = (I+C)g\theta + Cn$ it follows that

$$s_{i+1} = (g_{i+1} + \langle c(i), g(i)\rangle)\theta + \langle c(i), n(i)\rangle$$

$$= \frac{g_{i+1}}{1 + \sigma_\theta^2 \|g(i)\|^2_{K^{-1}(i)}} (\theta - \sigma_\theta^2 \langle g(i), K^{-1}(i)n(i)\rangle). \quad (3.11)$$

Since $r\text{-}Au = g\theta + n$, the minimum variance unbiased estimate of θ after N observations is obtained from (2.16) as

$$\hat{\theta}_N = \theta + \frac{\langle gK^{-1}n\rangle}{\|g\|^2_{K^{-1}}}, \tag{3.12}$$

and after i observations it is

$$\hat{\theta}_i = \theta + \frac{\langle g(i), K^{-1}(i)n(i)\rangle}{\|g(i)\|^2_{K^{-1}(i)}} . \tag{3.13}$$

Therefore

$$s_{i+1} = g_{i+1}\left(\theta - \frac{\sigma_\theta^{\,2}\|g(i)\|^2_{K^{-1}(i)}}{1 + \sigma_\theta^{\,2}\|g(i)\|^2_{K^{-1}(i)}}\,\hat{\theta}_i\right). \tag{3.14}$$

Define

$$x_i = \frac{\sigma_\theta^{\,2}\|g(i)\|^2_{K^{-1}(i)}}{1 + \sigma_\theta^{\,2}\|g(i)\|^2_{K^{-1}(i)}}\,\hat{\theta}_i \tag{3.15}$$

then

$$s_{i+1} = g_{i+1}(\theta - x_i) \tag{3.16}$$

is the desired result.

There is no difficulty in showing that x_i is the <u>minimum variance</u> linear estimate of θ given the observations $r_1, r_2, \cdots, r_i$. Note that x_i is a <u>biased estimate</u> ($E[x_i] \neq \theta$) as opposed to the estimate $\hat{\theta}_i$ which is the <u>minimum variance unbiased</u> estimate.

3.4. General Sequential Form of the Receiver.

The positive definite symmetric covariance matrix K may always be factored into the product $K = QQ^T$, where Q is lower triangular and positive definite. Similarly, $K(i) = Q(i)Q^T(i)$ where $Q(i)$ is the upper left $i \times i$ corner of Q. Let $g = \frac{1}{\sigma_\theta} Qf$, then

$g(i) = \frac{1}{\sigma_\theta} Q(i)f(i)$ and $\|g(i)\|^2_{K^{-1}(i)} = \frac{1}{\sigma_\theta^2} \|f(i)\|^2$ where

$f(i) = \mathrm{col}(f_1, f_2, \cdots, f_i)$. Then

$$\|f(i)\|^2\hat{\theta}_i = \|f(i)\|^2\theta + \sigma_\theta\langle f(i), Q^{-1}(i)n(i)\rangle$$

$$= \|f(i-1)\|^2\theta + \sigma_\theta\langle f(i-1)Q^{-1}(i-1)n(i)\rangle + f_i^{\,2}\theta + \sigma_\theta f_i\langle h(i), n(i)\rangle \tag{3.17}$$

where $h(i) = \mathrm{col}(h_{i1}, h_{i2}, \cdots, h_{ii})$ denotes the i-th row of Q^{-1}, that is $h_{ij} = (Q^{-1})_{ij}$ for $j \le i = 1, 2, \cdots, N$. Thus,

$$\|f(i)\|^2\hat{\theta}_i = \|f(i-1)\|^2\hat{\theta}_{i-1} + f_i(f_i\theta + \langle h(i), n(i)\rangle) . \tag{3.18}$$

But $f = \sigma_\theta Q^{-1}g$, therefore $f_i = \sigma_\theta\langle h(i), g(i)\rangle$ and

$$\|f(i)\|^2\hat{\theta}_i = \|f(i-1)\|^2\hat{\theta}_{i-1} + \sigma_\theta f_i\langle h(i), (g(i)\theta + n(i))\rangle \tag{3.19}$$

But, recall that $\left[1 + \|f(i)\|^2\right]x_i = \|f(i)\|^2\theta_i$, therefore,

$$(1 + \|f(i)\|^2)x_i = (1 + \|f(i)\|^2)x_{i-1} - f_i^{\,2}x_{i-1} + \sigma_\theta f_i\langle h(i), (g(i)\theta + n(i))\rangle \tag{3.20}$$

$$x_i = x_{i-1} + \frac{\sigma_\theta f_i}{1 + \|f(i)\|^2} \langle h(i), (g(i)\theta + n(i) - g(i)x_{i-1})\rangle$$

$$= x_{i-1} + \frac{\sigma_\theta f_i}{1 + \|f(i)\|^2} \sum_{j=1}^{i} h_{ij}(g_j\theta + n_j - g_j x_{i-1}) \ . \tag{3.21}$$

Since $s_j = g_j(\theta - x_{j-1})$, $r_j = s_j + n_j$ becomes

$$g_j\theta + n_j = r_j + g_j x_{j-1} \ , \tag{3.22}$$

it follows immediately that

$$x_i = x_{i-1} + \frac{\sigma_\theta f_i}{1 + \|f(i)\|^2} \sum_{j=1}^{i} h_{ij}[r_j + g_j(x_{j-1} - x_{i-1})] \ . \tag{3.23}$$

The above result is valid for any positive definite covariance matrix. In particular, when the noise is white so that $K = \sigma^2 I$, $Q^{-1} = \frac{1}{\sigma} I$, $h_{ij} = \frac{1}{\sigma}\delta_{ij}$, $f = \frac{\sigma_\theta}{\sigma} g$ it reduces to

$$x_i = x_{i-1} + \frac{\sigma_\theta^2 g_i}{\sigma^2 + \sigma_\theta^2 \|g(i)\|^2} r_i \tag{3.24}$$

If the noise is generated by a first order difference equation $n_i = \alpha n_{i-1} + w_i$, where w_i is white noise with covariance $K_w = \sigma_w^2 I$, then $K_n = (I-\alpha J)^{-1} K_w (I-\alpha J^T)^{-1}$ where $J_{ij} = \delta_{ij+1}$. Thus $Q^{-1} = \frac{1}{\sigma_w}(I-\alpha J)$ so that $\sigma_w h(i) = \mathrm{col}(0, 0, \cdots, 0, -\alpha, 1)$. Then

$$x_i = x_{i-1} + \frac{\sigma_\theta^2 (g_i - \alpha g_{i-1})}{\sigma_w^2 + \|g(i) - \alpha g(i-1)\|^2} [r_i - \alpha r_{i-1} + \alpha g_{i-1}(x_{i-1} - x_{i-2})] \tag{3.25}$$

which is a linear difference equation of order two. In general, if the noise obeys a linear difference equation of order m which is driven by white noise, then the minimum variance estimate obeys a linear difference equation of order $m + 1$.

The decision rule requires $\hat{\theta}_N$. It can be computed from x_N via the relationship $\hat{\theta}_N = (1 + \|f\|^2)x_N/\|f\|^2$, or

$$\hat{\theta}_N = \frac{\sigma_\theta^{\ 2}\|g\|^2_{K^{-1}}}{1 + \sigma_\theta^{\ 2}\|g\|^2_{K^{-1}}}\, x_N \ . \tag{3.26}$$

The coefficient of x_N approaches unity as $\|g\|^2_{K^{-1}}$ goes to infinity with N, and x_N is asymptotically unbiased.

3.5. Sequential Operation of the Transmitter.

Since the transmitted signals are related to the minimum variance estimates according to $s_i = g_i(\theta - x_{i-1})$ it follows that $x_{j-1} - x_{i-1} = \frac{s_i}{g_i} - \frac{s_j}{g_j}$. Substituting this into the general formula for x_i produces

$$s_{i+1} = \frac{g_{i+1}}{g_i}\left(\frac{1 + \sigma_\theta^{\ 2}\|g(i-1)\|^2_{K^{-1}(i-1)}}{1 + \sigma_\theta^{\ 2}\|g(i)\|^2_{K^{-1}(i-1)}}\right)\left(s_i - \frac{g_i f_i \langle h(i), (r(i)-s(i))\rangle}{1 + \sigma_\theta^{\ 2}\|g(i-1)\|^2_{K^{-1}(i-1)}}\right)$$

$$= \frac{g_i e_{i+1}}{g_{i+1} e_i}\left(s_i - \frac{f_i}{g_i \sigma_\theta^{\ 2}}\, e_i \langle h(i), (r(i)-s(i))\rangle\right)$$

which reduces to

$$s_{i+1} = \sqrt{\frac{e_{i+1}}{e_i}\left(1 + \frac{e_i}{\sigma^2}\right)}\left(s_i - \frac{e_i}{e_i + \sigma^2}\, r_i\right) \tag{3.27}$$

when the noise is white.

3.6. Selection of g When the Noise is Uncorrelated.

If the noise is statistically independent between signals, the covariance matrix K_n and therefore also K is diagonal. The diagonal elements are: $k_{ij} = E[n_i n_j] = \sigma_i^2 \delta_{ij}$. Let $f_i = \frac{\sigma_\theta}{\sigma_i} g_i$, then

$$\frac{e_i}{\sigma_i^2} = \frac{f_i^2}{1 + \|f(i-1)\|^2} \tag{3.28}$$

But

$$1 + \|f(i)\|^2 \equiv \prod_{j=1}^{i} \left(\frac{1 + \|f(j)\|^2}{1 + \|f(j-1)\|^2} \right) \qquad (\|f(0)\|^2 = f_0 = 0)$$

$$= \prod_{j=1}^{i} \left(1 + \frac{f_j^2}{1 + \|f(j-1)\|^2} \right) \tag{3.29}$$

$$= \prod_{j=1}^{i} \left(1 + \frac{e_j}{\sigma_j^2} \right) \tag{3.30}$$

hence also

$$g_{i+1}^2 = \frac{e_{i+1}}{\sigma_\theta^2} \prod_{j=1}^{i} \left(1 + \frac{e_j}{\sigma_j^2} \right) \tag{3.31}$$

and

$$1 + \sigma_\theta^2 \|g\|^2_{K^{-1}} = \prod_{i=1}^{N} \left(1 + \frac{e_i}{\sigma_i^2} \right) \tag{3.32}$$

Therefore, the constraint on $||g||^2_{K^{-1}}$ is now a constraint on the signal energies. It is convenient to use $\ln(1 + \sigma_\theta^2 ||g||^2_{K^{-1}})$ and to set

$$\frac{\partial}{\partial e_i} \sum_{i=1}^{N} \left[e_i - \nu \ln\left(1 + \frac{e_i}{\sigma_i^2}\right)\right] = 0 \tag{3.33}$$

where $\sum_{i=1}^{N} e_i = E_{av}$, and ν is a Lagrange multiplier. The result is

$$e_i + \sigma_i^2 = \nu = \frac{E_{av} + \mathrm{Tr}\ K}{N} \qquad i = 1, 2, \cdots, N \tag{3.34}$$

The above condition is meaningful only if $e_i = \nu - \sigma_i^2 > 0$, a situation that does not always prevail. If $\nu - \sigma_i^2 \leq 0$ the answer is to set $e_i = 0$ (thus also $s_i = 0$), $N = N\text{-}1$ and recalculate ν. Assuming that this does not occur (E_{av} sufficiently large or $\sigma_i^2 = \sigma^2$) let $\frac{1}{N} \mathrm{Tr}\ K = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 = \overline{\sigma^2} = \frac{\overline{N_0}}{2}$. Then, since $E = PT$ and $N = 2WT$

$$1 + \sigma_\theta^2 ||g||^2_{K^{-1}} = \left(1 + \frac{P}{\overline{N_o} W}\right)^{2WT} \prod_{i=1}^{N} \left(\frac{\overline{\sigma^2}}{\sigma_i^2}\right) \tag{3.35}$$

$$\geq \left(1 + \frac{P}{\overline{N_o} W}\right)^{2WT}$$

3.7. Optimum Performance When the Noise is White.

If the noise, in addition to being uncorrelated, is stationary then $k_{ij} = k|i-j| = \sigma^2 \delta_{ij} = \frac{N_o}{2} \delta_{ij}$, where $\frac{N_o}{2}$ is the two-sided noise power spectral density of white noise. Then

$$1 + \sigma_\theta^2 ||g||^2_{K^{-1}} = \left(1 + \frac{P}{N_o W}\right)^{2WT} \tag{3.37}$$

and

$$P_e = \operatorname{erfc} \sqrt{\frac{3}{2} \frac{(e^{2CT}-1)}{(e^{2RT}-1)}} \tag{3.38}$$

where

$$C = W \ln \left(1 + \frac{P}{N_o W}\right) \text{ nats/sec} \tag{3.39}$$

is the theoretical capacity of the channel and

$$R = \frac{1}{T} \ln M \quad \text{nats/sec} \tag{3.40}$$

is the rate of the message source.

By the well known properties of erfc,

$$\lim_{T\to\infty} P_e = \begin{cases} 0 & R < C \\ \operatorname{erfc} \frac{3}{2} & R = C \\ 1 & R > C\,. \end{cases} \tag{3.41}$$

The asymptotic behavior of the error is

$$P_e = \frac{2}{3\sqrt{\pi}}\, e^{-\frac{3}{2} e^{2(C-R)T} + (C-R)T} \,. \tag{3.42}$$

The double exponential decrease of the error as a function of coding delay T, or block length N, is characteristic of codes using noiseless feedback. If thought of as an increase in the effective block length due to a larger effective input alphabet then this result was predicted by Shannon [4].

3.8. Suboptimum Codes that Achieve Channel Capacity.

There is an uncountable suboptimum choice of the signal energies $e_1, e_2 \cdots$ that allows communication at a positive rate, in fact, up to channel capacity. Define

$$R_c(N) = \frac{1}{2T} \sum_{i=1}^{N} \ln(1 + \frac{e_i}{\sigma^2})$$

$$= \frac{P}{N_o} \frac{\sum_{i=1}^{N} \ln(1 + \frac{e_i}{\sigma^2})}{\sum_{i=1}^{N} \frac{e_i}{\sigma^2}} \qquad (3.43)$$

Therefore

$$P_e = \operatorname{erfc} \sqrt{\frac{3}{2} \; \frac{e^{2R_cT} - 1}{e^{2RT} - 1}} \quad . \qquad (3.44)$$

Theorem 1

A necessary and sufficient condition for achieving zero error for $0 < R < (1-\delta)R_c$ is $\frac{1}{\sigma^2} E_{av} = \infty$. This is an obvious condition.

Proof

$$\sum_{i=1}^{N} \frac{e_i}{\sigma^2} \geq \sum_{i=1}^{N} \ln(1 + \frac{e_i}{\sigma^2}) \geq \ln(1 + \sum_{i=1}^{N} \frac{e_i}{\sigma^2})$$

thus

$$\frac{E_{av}}{\sigma^2} \geq R_c T \geq \ln(1 + \frac{E_{av}}{\sigma^2}) \tag{3.45}$$

But P_e is zero in and only if $R_c T = \infty$, thus $\frac{E_{av}}{\sigma^2} = \sum_{i=1}^{\infty} \frac{e_i}{\sigma^2} = \infty$.

Definition

Let $R_c(\infty)$ be the maximum rate at which zero error can be achieved. It is defined here by the two conditions: (a) $\sum_{i=1}^{\infty} \frac{e_i}{\sigma^2} = \infty$, and (b)

$$R_c(\infty) = \lim_{N\to\infty} \frac{P}{2\sigma^2} \frac{\sum_{i=1}^{N} \ln(1 + \frac{e_i}{\sigma^2})}{\sum_{i=1}^{N} \frac{e_i}{\sigma^2}} \quad . \tag{3.46}$$

Theorem 2

If the sequence $\{e_i\}_{i=1}^{\infty}$ converges to a limit then $R_c(\infty)$ is given by

$$R_c(\infty) = \begin{cases} \text{(a)} \quad C_W = W\ln(1 + \frac{P}{N_o W}) & \text{if } \lim_{i\to\infty} e_i = \frac{P}{N_o W} \\ \text{(b)} \quad C_\infty = \frac{P}{N_o} & \text{if } \lim_{i\to\infty} e_i = 0 \\ \text{(c)} \quad 0 & \text{if } \lim_{i\to\infty} e_i = \infty \end{cases} \tag{3.47}$$

The proof is in Appendix I.

3.9. A Class of Codes that Achieve C_∞ for the Additive White Noise Channel.

Theorem 2(b) indicates that there exists a variety of feedback codes that can achieve the infinite bandwidth capacity limit $\frac{P}{N_o}$. One method of classifying these schemes is to examine how N, or W, increases with T. For example, consider the class of codes in which the signal energies are of the form

$$e_i = \left(\frac{1}{i}\right)^{\gamma} \tag{3.48}$$

where $0 \le \gamma \le 1$. First, $\gamma = 0$ corresponds to the optimum scheme in which $P/W = 1$, $C_W = \ln 2$ so that W is finite and independent of T. The other extreme is for $\gamma = 1$ which gives

$$R_c T = \frac{1}{2} \ln(1 + N)$$

$$N = e^{2R_c T} - 1$$

therefore

$$W = (e^{2R_c T} - 1)/2T \tag{3.49}$$

When $0 < \rho < 1$ and T is sufficiently large, $R_c \approx C_\infty = P/N_o$ and

$$R_c T = \sum_{i=1}^{N} \ln(1 + i^{-\gamma})$$

$$\sim \sum_{i=1}^{N} i^{-\gamma} \qquad \text{for large } N$$

$$\sim \frac{N^{1-\gamma}}{1 - \gamma}$$

thus

$$N \sim [2C_\infty(1-\gamma)]^{\frac{1}{1-\gamma}} T^{\frac{1}{1-\gamma}}$$

and

$$W \sim \frac{1}{2} [2C_\infty(1-\gamma)]^{\frac{1}{1-\gamma}} T^{\frac{\gamma}{1-\gamma}} \qquad (3.50)$$

In general, the growth of W with T is bounded between a constant and $(e^{2C_\infty T} - 1)/T$ for all T. The asymptotic behavior of W as a function of T and γ is illustrated in Fig. 3.

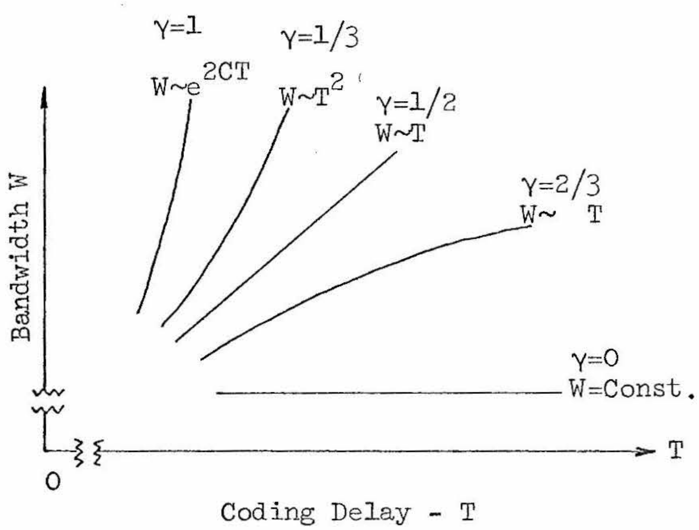

Fig. 3. Bandwidth vs Coding Delay for a Class of Codes.

3.10. Signal Selection with an Additional Constraint.

The coding scheme of Kailath and Schalkwijk can be derived by solving the signal selection problem subject to an additional constraint on AB, or $C = (I-AB)^{-1}AB$, and g. In their scheme, $E[s_1/\theta] = g_1\theta$ while $E[s_i/\theta] = 0$ for all $i > 1$. In vector form this translates into setting $E[s/\theta] = E[(I+C)g\theta + Cn/\theta] = (I+C)g\theta = \mathrm{col}(g_1\theta, 0, \cdots, 0)$. Thus there is the auxiliary constraint $(I+C)g = \mathrm{col}(g_1, 0, \cdots, 0)$ or

$$g_i + \langle c(i-1), g(i-1)\rangle = 0 \qquad i = 2, 3, \cdots, N . \tag{3.51}$$

In the white noise case considered by Schalkwijk, the present formulation gives

$$E_{av} = \|(I+C)g\|^2 \sigma_\theta^{\,2} + \sigma^2 \mathrm{Tr}\; CC^T \tag{3.52}$$

$$= \sigma_\theta^{\,2} g_1^{\,2} + \sigma^2 \sum_{i=1}^{N} \|c(i-1)\|^2 . \tag{3.53}$$

The minimization over $c(i-1)$ must now include the $N-1$ constraint equations, which are easily incorporated via Lagrange multipliers $\lambda_2, \lambda_3, \cdots \lambda_N$. Thus, let

$$F = \sigma_\theta^{\,2} g_1^{\,2} + \sigma^2 \sum_{i=2}^{N} \|c(i-1)\|^2 + \sum_{i=2}^{N} \lambda_i (g_i + \langle c(i-1), g(i-1)\rangle) \tag{3.54}$$

and set $\mathrm{grad}_{c(i-1)} F = 0$ for $i = 2, 3, \cdots, N$ in order to obtain

$$c(i-1) = -\frac{\lambda_i}{2\sigma^2} g(i-1), \tag{3.55}$$

hence,

$$\frac{\lambda_i}{2\sigma^2} = \frac{g_i}{\|g(i-1)\|^2} \qquad i = 2, 3, \cdots N \tag{3.56}$$

and,

$$e_i = \sigma^2\|c(i-1)\|^2$$

$$= \sigma^2 \frac{g_i^{\ 2}}{\|g(i-1)\|^2} \quad .$$

Next, note that

$$\sigma_\theta^{\ 2}\|g\|^2_{K^{-1}} \equiv \frac{\sigma_\theta^{\ 2}}{\sigma^2} (g_1^{\ 2} + g_2^{\ 2} + \cdots + g_N^{\ 2})$$

$$\equiv \frac{\sigma_\theta^{\ 2}}{\sigma^2} g_1^{\ 2} \prod_{i=2}^{N} \left(1 + \frac{g_i^{\ 2}}{\|g(i-1)\|^2}\right)$$

therefore,

$$\sigma_\theta^{\ 2}\|g\|^2_{K^{-1}} = \frac{e_1}{\sigma^2} \prod_{i=2}^{N} \left(1 + \frac{e_i}{\sigma^2}\right) \tag{3.57}$$

while,

$$E_{av} = \sum_{i=1}^{N} e_i$$

The optimum choice of signal energies is $e_1 - \sigma^2 = e_2 = e_3 = , \cdots e_N = \frac{E_{av}-\sigma^2}{N} - \sigma^2$. Therefore,

$$\sigma_\theta\|g\|^2_{K^{-1}} = \left(1 + \frac{E_{av}-\sigma^2}{\sigma^2 N}\right)^N \tag{3.58}$$

$$\leq \left(1 + \frac{E_{av}}{\sigma^2 N}\right)^N - 1 \qquad \text{(the optimum result)}$$

with equality when $N = 1$ and of negligible difference as N increases. However, in this case e_1 exceeds the remaining e's by σ^2. If all the signal energies are made equal, the result is

$$\sigma_\theta^2 \|g\|^2_{K^{-1}} = \frac{E_{av}}{\sigma^2 N}\left(1 + \frac{E_{av}}{N\sigma^2}\right)^{N-1} \tag{3.59}$$

which is considerably less than $\left(1 + \frac{E_{av}}{N\sigma^2}\right)^N$ when $\frac{E_{av}}{N\sigma^2} = \frac{P}{N_o W}$ is small.

The form of the transmitted signals is obtained by eliminating $c(i-1)$ from the expression for s_i as in Section 3.3. The result is

$s_i = g_i(\theta - \hat{\theta}_{i-1})$. Similarly, it is easy to show that $\hat{\theta}_i = \hat{\theta}_{i-1} + \frac{g_i \sigma^2}{\sigma_\theta^2 \|g\|^2} r_i$ is the sequential operation to use at the receiver.

3.11. Selection of g When the Noise is Correlated.

The problem of optimizing E_{av} over g when $\|g\|^2_{K^{-1}}$ is fixed, is in the case of correlated noise complicated by the non-diagonal nature of K. The problem appears to be formidable even in the case of a simple first order markov process. It is not difficult, however, to guess a good g and to compute the resulting performance. The definition for R_c is still given by

$$R_c = \frac{1}{2T} \ln(1 + \sigma_\theta \|g\|^2_{K^{-1}}) \tag{3.60}$$

Factorization of K into the product QQ^T where Q is lower triangular and positive definite, and setting $f = \sigma_\theta Q^{-1} g$, $g = \frac{1}{\sigma_\theta} Qf$, gives $1 + \sigma_\theta^2 \|g\|^2_{K^{-1}} = 1 + \|f\|^2$. The identity

$$1 + \|f\|^2 \equiv \prod_{i=1}^{N}\left(1 + \frac{f_i^{\,2}}{1 + \|f(i-1)\|^2}\right) \tag{3.61}$$

can be used to obtain

$$R_c = W \frac{1}{N} \sum_{i=1}^{N} \ln\left(1 + \frac{f_i^{\,2}}{1 + \|f(i-1)\|^2}\right). \tag{3.62}$$

However,

$$E_{av} = \sum_{i=1}^{N} \frac{\langle f(i), q(i)\rangle^2}{1 + \|f(i-1)\|^2} \tag{3.63}$$

is not simplified. Here, $f(i) = \mathrm{col}(f_1, f_2, \cdots, f_i)$, $q(i) = \mathrm{col}(q_{i1}, q_{i2}, \cdots, q_{ii})$ is the i-th row of Q, and the individual signal energies are of the form $e_i = \langle f(i), q(i)\rangle^2/(1 + \|f(i-1)\|^2)$. Thus, there does not exist a simple relationship between $\gamma_i = f_i^{\,2}/(1 + \|f(i-1)\|^2)$ and e_i as in the case of a diagonal covariance matrix. Nevertheless, at least two good guesses are available. The first is to let $\gamma_i = \gamma$ a constant independent of i. This gives

$$R_c = W \ln(1 + \gamma) \tag{3.64}$$

and

$$\pm f_i = \sqrt{\gamma(1 + \gamma)^{i-1}} \tag{3.65}$$

so that

$$E_{av} = \gamma \sum_{i=1}^{N} \left[(1+\gamma)^{1-i} \left(\sum_{j=1}^{i} \pm (1+\gamma)^{\frac{j-1}{2}} q_{ij} \right)^2 \right]$$

$$= \gamma \sum_{i=1}^{N} \left(\sum_{j=1}^{i} \pm (1+\gamma)^{\frac{j-1}{2}} q_{ij} \right)^2 \tag{3.66}$$

Note that the sign in front of each f_i affects the terms in E_{av} but not those in R_c where the f_i appear squared. This allows a partial minimization of E_{av} over sgn f_i to be carried out.

The second guess is to maintain constant average power by choosing all the e_i's equal to a constant, say e. Then

$$E_{av} = Ne$$

and

$$R_c = W \frac{1}{N} \sum_{i=1}^{N} \ln\left(1 + \frac{\langle h(i), g(i)\rangle^2}{g_i^2}\, e\right) \tag{3.67}$$

where $h(i) = \text{col}(h_{i1}, h_{i2}, \cdots, h_{ii})$ is the i-th row of Q^{-1}, $g_i = \frac{1}{\sigma_\theta} \langle q(i), f(i)\rangle$, and $f_i = \sigma_\theta \langle h(i), g(i)\rangle$. Once again the signs of the components in the inner product are available for partial optimization of R_c. An example for the case of first order Markov noise is worked out in Appendix II where it is shown that for both $\gamma_i = \gamma$, and $e_i = e$

$$\frac{P}{N_o W} = \frac{\gamma}{\left(1 + \frac{|\alpha|}{\sqrt{1+\gamma}}\right)^2} \tag{3.68}$$

where α is the parameter of the Markov noise $(n_i = \alpha n_{i-1} + w_i)$. The critical rate achievable is then

$$R_c = W \ln(1 + \gamma)$$

$$= \begin{cases} W \ln(1 + \frac{P}{N_o W}) & \text{for } \frac{P}{N_o W} \gg 1 \qquad (3.69) \\ \\ (1 + |\alpha|)^2 \frac{P}{N_o W} & \text{for } \frac{P}{N_o W} \ll 1 \qquad (3.70) \end{cases}$$

where $\frac{N_o}{2}$ is the spectral power density of the white noise which drives the difference equation to produce Markov noise, $(E[w_i w_j] = \frac{N_o}{2} \delta_{ij})$.

IV. NOISY FEEDBACK

4.1. Introduction.

The inclusion of feedback noise is essential in the representation of a physically realizable system. Otherwise, as it is shown in Section 3.1, it would be possible to use arbitrarily weak feedback signals for conveying information to the transmitter. Thus, in the case of noisy feedback, the probability of error is determined by the feedback power as well as the power used in the forward channel.

The main purpose of this chapter is to establish the relationship between the forward and feedback power and the probability of error for systems in which the noise in the forward and feedback channels is white and independent. There are many practical situations in which this assumption is valid. Thus, if σ^2 and σ_m^2 are the variances of the forward and feedback noise, then $K_n = \sigma^2 I$, $K_m = \gamma\sigma^2 I$ and $K_{mm} = 0$, where $\sigma^2 = \frac{N_o}{2}$ is the two-sided noise power spectral density in the forward channel and $\gamma = \sigma_m^2/\sigma^2$. Equation (2.10) simplifies to

$$K = \sigma^2(I + \gamma AA^T) , \qquad (4.1)$$

and, letting $I + C = (I-AB)^{-1}$, where $C = AB(I-AB)^{-1}$ is lower triangular with zeros along the main diagonal allows E_{av}, as given by (2.29), to be written

$$E_{av} = \mathrm{Tr}[(I+C)(\sigma_\theta^2 gg^T + \sigma_m^2 AA^T)(I+C)^T + \sigma^2 CC^T] . \qquad (4.2)$$

However,

$$Tr[CC^T] = Tr[(I+C)(I+C)^T - C - C^T - I] \tag{4.3}$$

and $Tr[C] = 0$ because the diagonal elements of C are identically zero. Also $Tr[I] = N$, therefore

$$Tr[CC^T] = Tr[(I+C)(I+C)^T] - N . \tag{4.4}$$

Substitution of (4.4) into (4.2) produces

$$E_{av} = Tr[(I+C)(\sigma_\theta^{\ 2} gg^T + \sigma_m^{\ 2} AA^T + \sigma^2 I)(I+C)^T] - \sigma^2 N \tag{4.5}$$

$$N_\sigma^{\ 2} + E_{av} = Tr[(I+C)(\sigma_\theta^{\ 2} gg^T + K)(I+C)^T] \quad . \tag{4.6}$$

Note that the above simplification is possible only because K_n was assumed to be diagonal. The feedback energy as given by equation (2.30) may also be written in terms of C,

$$E_{fb} = Tr[B(I+C)(\sigma_\theta^{\ 2} gg^T + K)(I+C)^T B^T] \quad . \tag{4.7}$$

Finally, the quantity that controls the probability of error is

$$E[(\hat{\theta}_N - \theta)^2/\theta] = \frac{1}{\|g\|^2_{K^{-1}}} \tag{4.8}$$

where

$$\hat{\theta}_N = \frac{\langle g, K^{-1}(r-Au)\rangle}{\|g\|^2_{K^{-1}}} \quad . \tag{4.9}$$

It is easy to see that the simple scaling argument $B \to \epsilon B$, $A \to \frac{1}{\epsilon} A$, $AB \to AB$ of Section 3.1, that caused E_{fb} to be arbitrarily small without affecting E_{av} and $\|g\|^2_{K^{-1}}$ fails; because, in the present case $K = \sigma^2(I+\gamma AA^T)$ becomes $\sigma^2(I + \frac{\gamma}{\epsilon^2} AA^T)$, thereby causing $\|g\|^2_{K^{-1}}$ to decrease while E_{av} increases as ϵ is reduced.

A more subtle but equally unrewarding pursuit is to choose g in the null space of A^T. This choice is deceptively promising but it is easy to prove that it leads to the no-feedback solution $A \equiv 0$. First, in order to show why such a choice is seemingly good, note that

$$\begin{aligned} K^{-1} &= \frac{1}{\sigma^2} (I+\gamma AA^T)^{-1} \\ &= \frac{1}{\sigma^2} [I-\gamma AA^T(I+\gamma AA^T)^{-1}] \ , \end{aligned} \tag{4.10}$$

therefore,

$$\|g\|^2_{K^{-1}} = \frac{1}{\sigma^2} \|g\|^2 - \gamma\langle g, AA^T g\rangle_{K^{-1}} \tag{4.11}$$

$$\leq \frac{1}{\sigma^2} \|g\|^2 \tag{4.12}$$

with equality if and only if $A^T g = 0$. Thus it seems that $E[(\hat{\theta}_N-\theta)^2/\theta]$ and the probability of error, are minimal and independent of the feedback noise when $A^T g = 0$. Since the rank of A is at most $N-1$, there exists a non-trivial choice of g that satisfies the above condition.

Now, in order to prove that this condition leads to the no feedback case $A \equiv 0$ as the optimum solution, consider the expression for E_{av} as given by (4.2).

$$\begin{aligned} E_{av} &= \mathrm{Tr}[(I+C)(\sigma_\theta{}^2 gg^T + \sigma_m{}^2 AA^T)(I+C)^T + \sigma^2 CC^T] \\ &= \sigma_\theta{}^2 \|(I+C)g\|^2 + \sigma^2\ \mathrm{Tr}[(I+C)\gamma AA^T(I+C)^T] + \sigma^2\ \mathrm{Tr}[CC^T] \\ &= \sigma_\theta{}^2(\|g\|^2 + 2\langle g, Cg\rangle + \|C_g\|^2) + \sigma^2(\gamma\|(I+C)A\|^2 + \|C\|^2) \\ &\geq \sigma_\theta{}^2(\|g\|^2 + 2\langle g, C_g\rangle) \end{aligned} \tag{4.13}$$

But $C = AB(I-AB)^{-1}$, therefore

$$\langle g, C_g\rangle = g^T AB(I-AB)^{-1} g \tag{4.14}$$

vanishes when $g^T A = (A^T g)^T = 0$. Thus, if $A^T g = 0$ then

$$E_{av} \geq \sigma_\theta{}^2 \|g\|^2 \tag{4.15}$$

with equality if and only if $A \equiv 0$. In fact, the only quantity that can become negative, and thereby play a major role in minimizing E_{av}, is (4.14). This implies (without proof, however) that A and B should have, as in the noiseless case, rank $N-1$, which is the maximum rank allowed in the formulation.

4.2. Solution of the Signal Selection Problem.

An important relationship between E_{av}, E_{fb} and $||g||^2_{K^{-1}}$ for systems with white and independent noise can be established by the use of several well known matrix properties. Let

$$H \triangleq (I-AB)^{-1}\left(\frac{\sigma_\theta^2}{\sigma^2} gg^T + I + \gamma AA^T\right)(I-B^TA^T)^{-1} \qquad (4.16)$$

Since AB is lower triangular with zeros down the main diagonal, $I - AB$ has ones down the main diagonal and $\det(I-AB)^{-1} = 1$. Therefore,

$$\det H = \det(I-AB)^{-2} \det\left(\frac{\sigma_\theta^2}{\sigma^2} gg^T + I + \gamma AA^T\right)$$

$$\det H = \det\left(\frac{\sigma_\theta^2}{\sigma^2} gg^T + I + \gamma AA^T\right)$$

$$= \left(1 + \sigma_\theta^2 ||g||^2_{K^{-1}}\right) \det(I+\gamma AA^T) \; , \qquad (4.17)$$

where, in general, $\det(I+xx^T) = 1 + ||x||^2$ for any vector x. The proof of this fact is simple. The vector x is itself an eigenvector of the matrix $I + xx^T$ because $(I+xx^T)x = x + x\langle x,x\rangle = (1 + ||x||^2)x$. Thus the eigenvalue associated with x is $1 + ||x||^2$. Since $I + xx^T$ is symmetric the remaining $N-1$ eigenvectors must be orthogonal to x. Thus, if y is any other eigenvector, $\langle x,y\rangle = 0$, and $(I+xx^T)y = y$, proving that the remaining eigenvalues are all equal to unity. Since the determinant is equal to the product of the eigenvalues, the proof is complete.

The result for $1 + \sigma_\theta^2 \|g\|^2_{K^{-1}}$ is obtained by letting $QQ^T = I + \gamma AA^T$ and $x = \frac{\sigma_\theta}{\sigma} Q^{-1} g$, so that $\left(\frac{\sigma_\theta^2}{\sigma^2} gg^T + I + \gamma AA^T \right) = Q(xx^T + I)Q^T$. Thus

$$1 + \sigma_\theta^2 \|g\|^2_{K^{-1}} = \frac{\det H}{\det(I+\gamma AA^T)} = \frac{\prod_{i=1}^{N} \nu_i}{\det(I+\gamma AA^T)} \tag{4.18}$$

where $\nu_1, \nu_2, \cdots, \nu_N$ are the eigenvalues of H. Also

$$N + \frac{E_{av}}{\sigma^2} = \operatorname{Tr} H = \sum_{i=1}^{N} \nu_i , \tag{4.19}$$

and

$$\frac{E_{fb}}{\sigma^2} = \operatorname{Tr}[B^T BH] \tag{4.20}$$

where $\operatorname{Tr}[B^T BH] = \operatorname{Tr}[BHB^T]$ follows from the invariance of the trace operator under cyclic permutations of the matrices.

Since $\operatorname{erfc} x$ is a monotonically decreasing function of x, the minimization of P_e given E_{av} and E_{fb} is equivalent to maximizing $1 + \sigma_\theta^2 \|g\|^2_{K^{-1}}$ subject to constraints on E_{av} and E_{fb}. But, since $\nu_i > 0$ for all i, the following product inequality always holds

$$\prod_{i=1}^{N} \nu_i \leq \left(\frac{1}{N} \sum_{i=1}^{N} \nu_i \right)^N \tag{4.21}$$

with equality if and only if $\nu_i = \frac{1}{N} \sum_{i=1}^{N} \nu_i = \nu$ for all i. Thus,

$$1 + \sigma_\theta^2 \|g\|^2_{K^{-1}} \leq \frac{\left(1 + \frac{E_{av}}{N\sigma^2}\right)^N}{\det(I+\gamma AA^T)} \tag{4.22}$$

with equality if and only if all of the eigenvalues of H are equal, that is, $\nu_i = \nu = (1 + P/N_oW)$. In order to achieve the upper bound it is necessary to hold $\det(I+\gamma AA^T)$ fixed while varying A B and g. Since every symmetric matrix whose eigenvalues are all the same is a multiple of the identity matrix, $H = \nu I$.

Note that (4.22) immediately provides the solution to the noiseless feedback problem $(\gamma=0)$ in which the forward noise is white.

The condition $H = \nu I$ may not be achievable for some choices of B or A even if of rank $(N-1)$. However, if it is achieved then the best choice of B (from the set for which $H = \nu I$) can be found as follows:

$$\begin{aligned} \frac{E_{fb}}{\sigma^2} &= \mathrm{Tr}[B^T BH] \\ &= \nu \, \mathrm{Tr}[B^T B] \\ &= \nu \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} b_{ij}^2 \end{aligned} \tag{4.23}$$

the rank constraint on B requires that

$$\sum_{i=1}^{N-1} b_{ii}^2 > 0$$

Therefore, if the last row of B is disregarded because it is identically zero, the determinant of the resulting $(N-1) \times (N-1)$ matrix must be positive. Let $\beta_1^2, \beta_2^2, \cdots, \beta_{N-1}^2, \beta_N^2$ be the eigenvalues of B^TB then β_N^2 is identically zero, and $\sum_{i=1}^{N-1} \beta_i^2 = \sum_{i=1}^{N-1} b_{ii}^2 > 0$. Since $E_{fb} = \nu \, Tr[BB^T] = \beta_i^2$, it is minimized by choosing $\beta_i^2 = \beta^2$ for all $i = 1, 2, \cdots, N-1$. This immediately indicates that $B^TB = \beta^2 I_{N-1}$ where I_{N-1} is the $(N-1) \times (N-1)$ identity matrix. Therefore $B = \beta I_{N-1}$. (Because this is the only solution when B is lower triangular.) Thus it transpires that

$$\frac{E_{fb}}{\sigma^2} = (N-1)\ \nu\beta^2$$

or

$$\frac{\nu\beta^2}{\gamma} = \frac{P_{fb}}{\gamma N_o W}$$

$$= \rho_{fb} \tag{4.24}$$

where ρ_{fb} is the feedback signal-to-noise ratio. Also, there is

$$\nu = 1 + \frac{P}{N_o W}$$

$$= 1 + \rho \tag{4.25}$$

where ρ is the signal-to-noise ratio of the forward channel. It is now necessary to solve for A and g from the condition $H = \nu I$.

4.3. Selection of A and g when $B = \beta I$.

The starting point for the determination of A and g is the matrix equation $H = \nu I$. Note that this simple matrix equation represents a total of $N(N+1)/2$ independent scalar equations. The number of unknown elements in the lower-triangular zero diagonal matrix A is $N(N-1)/2$ while the number of unknown components in the vector g is N, giving a combined total of $N(N+1)/2$ unknowns. Thus, a solution is to be expected. Note that

$$\frac{\sigma_\theta^2}{\sigma^2} gg^T + I + \gamma AA^T = (I-AB)H(I-AB)^T$$

$$= (I-\beta A)\nu I(I-\beta A)^T$$

$$= \nu I - \nu\beta(A+A^T) + \nu\beta^2 AA^T$$

Therefore

$$\frac{\sigma_\theta^2}{\sigma^2} gg^T + I = \nu I - \nu\beta(A+A^T) + (\nu\beta^2-\gamma)AA^T \,. \tag{4.26}$$

Completing the "square" on the right-hand side gives

$$\frac{\sigma_\theta^2}{\sigma^2} gg^T + \left(\frac{\nu\beta^2+\nu\gamma-\gamma}{\nu\beta^2-\gamma}\right) I = \left(\frac{\nu^2\beta^2}{\nu\beta^2-\gamma}\right)\left(I - \frac{\nu\beta^2-\gamma}{\nu\beta} A\right)\left(I - \frac{\nu\beta^2-\gamma}{\nu\beta} A\right)^T \,. \tag{4.27}$$

This reduces to

$$ff^T + I = \alpha^2(K-D)(I-D)^T \tag{4.28}$$

where

$$D = \left(\frac{\nu\beta^2-\gamma}{\nu\beta}\right)A \tag{4.29}$$

$$f = \left(\frac{\nu\beta^2-\gamma}{\nu\beta^2 + \nu\gamma-\gamma}\right)^{1/2} \frac{\sigma_\theta}{\sigma}\, g \tag{4.30}$$

and

$$\alpha^2 = \frac{\nu^2\beta^2}{\nu\beta^2 + \nu\gamma-\gamma} \quad . \tag{4.31}$$

The solution is to factor the positive definite symmetric matrix $ff^T + I$ into the product of a lower triangular matrix with its transpose, and to identify this with the right-hand side. It is also not difficult to solve the $N(N+1)/2$ equations for f and D directly since they are recursive. For example, $f_1^2 + 1 = \alpha^2$, $f_2f_1 = -\alpha^2 d_{21}$, and $f_2^2 + 1 = \alpha^2 d_{21}^2$, this gives $f_2^2 = \alpha^2(\alpha^2-1)$etc. Here, a more economical method will be demonstrated.

Let $(I+ff^T)^{-1} = (I-hh^T)$ where $h = (1 + \|f\|^2)^{-\frac{1}{2}} f$, then

$$(I-D)^T(I-hh^T)(I-D) = \frac{1}{\alpha^2} I \; . \tag{4.32}$$

Note that $I-hh^T$ is positive definite since it is the inverse of a positive definite matrix. It may be easily verified that minimization of the trace of $(I-D)^T(I-hh^T)(I-D)$ subject to a constraint on the determinant gives the condition that $(I-D)^T(I-hh^T)(I-D)$ be diagonal.* Therefore, the diagonalization can be replaced with a minimization of the trace. Note that $\det(I-D)^T(I-hh^T)(I-D) = 1 - \|h\|^2 = \alpha^{-2N}$ so that D is not constrained by the determinant.

Let $d(i) = \mathrm{col}(d_{(i+1)i}, d_{(i+2)i}, \cdots, d_{Ni})$ denote the nonvanishing portion of the i-th column of D and let $h(i) = \mathrm{col}(h_{i+1}, h_{i+2}, \cdots, h_N)$. Note that $d(N) = h(N) = 0$ while $d(N-1) = d_{NN-1}$ and $h(N) = h_N$ are scalars. The trace of $(I-D)(I-hh^T)(I-D)$ is the sum of the diagonal elements,

$$1 + \|d(i)\|^2 - (h_i - \langle h(i), d(i)\rangle)^2 = \frac{1}{\alpha^2}\,;\ i = 1, 2, \cdots, N \tag{4.33}$$

Setting $\mathrm{grad}_{d(i)} = 0$ (or taking $\mathrm{grad}_{d(i)}$ of both sides) gives

$$d(i) + (h_i - \langle h(i), d(i)\rangle)h(i) = 0\ . \tag{4.34}$$

Taking the inner product with $h(i)$ gives

$$h_i - \langle d(i), h(i)\rangle = \frac{h_i}{1 - \|h(i)\|^2} \tag{4.35}$$

so that

$$d(i) = \frac{-h_i}{1 - \|h(i)\|^2}\, h(i)\ . \tag{4.36}$$

Substituting for $d(i)$ produces

---

* This is true for any positive definite matrix, e.g. H.

$$1 - \frac{h_i^{\;2}}{1 - \|h(i)\|^2} = \frac{1}{\alpha^2} \tag{4.37}$$

therefore,

$$\begin{aligned} h_i^2 &= \frac{1}{\alpha^2}\, h_{i+1} \\ &= \alpha^{-2(N-i)}\; h_N^2 \\ &= \frac{\alpha^2-1}{\alpha^2}\;\; \alpha^{-2(N-i)} \end{aligned} \tag{4.38}$$

where

$$h_N^2 = \frac{\alpha^2-1}{\alpha^2} \quad .$$

Consequently

$$d_{ji} = \begin{cases} -\;\dfrac{\alpha^2-1}{\alpha^2}\,\alpha^{j-i} & \text{for} \quad j > i \\ \\ 0 & \text{for} \quad j \le i \quad . \end{cases} \tag{4.39}$$

Next,

$$\begin{aligned} f_i^2 &= (1 + \|f\|^2)^{\frac{1}{2}}\; h_i^2 \\ &= \alpha^{2N}\left(\frac{\alpha^2-1}{\alpha^2}\right)\alpha^{-2(N-i)} \end{aligned}$$

hence

$$f_i = \left(\frac{\alpha^2-1}{\alpha^2}\right)^{\frac{1}{2}} \alpha^i \tag{4.40}$$

$$\frac{\sigma_\theta}{\sigma} g_i = (\nu-1)^{\frac{1}{2}} \alpha^{i-1} \tag{4.41}$$

and therefore

$$a_{ji} = - \frac{\nu-1}{\nu\beta} \begin{cases} \alpha^{j-i} & j > i \\ 0 & j \le i \end{cases} . \tag{4.42}$$

4.4. Evaluation of $\det(I+\gamma AA^T)$ and Probability of Error.

It is necessary to calculate $\det(I+\gamma AA^T)$ in order to determine the probability of error of the feedback code. Define the lower triangular matrix J such that $(J)_{ij} = \delta_{ij+1}$, note that J has zeros on the diagonal. Using this notation enables A to be written as

$$A = - \alpha \frac{\nu-1}{\nu\beta} J(I-\alpha J)^{-1} , \tag{4.43}$$

and therefore

$$I + \gamma AA^T = I + \delta^2 J(I-\alpha J)^{-1}(I-\alpha J^T)^{-1} J^T \tag{4.44}$$

where,

$$\delta^2 = \frac{\gamma\alpha^2(\nu-1)^2}{\nu^2\beta^2} .$$

However, AA^T occupies only the $(N-1) \times (N-1)$ lower right corner of its $N \times N$ format. Therefore all the off-diagonal elements in the first row and column of $I + \gamma AA^T$ are zero. Since the first diagonal element is unity, it is obvious that $\det(I+\gamma AA^T) = \det(I+\gamma AA^T)_{N-1}$ where the subscript is used to indicate that the matrix is $(N-1) \times (N-1)$.

Now,

$$(I+\gamma AA^T)_{N-1} = [I + \delta(I-\alpha J)^{-1}(I-\alpha J^T)^{-1}]_{N-1}$$

$$= (I-\alpha J)^{-1}_{N-1}[(I-\alpha J)(I-\alpha J)^T + \delta^2 I]_{N-1}(I-\alpha J^T)^{-1}_{N-1} .$$

Since $\det(I-\alpha J)^{-1}_{N-1} = \det(I-\alpha J)_{N-1} = 1$, it follows that

$$\det(I+\gamma AA^T) = \det(I+\gamma AA^T)_{N-1}$$

$$= \det[(I-\alpha J)(I-\alpha J)^T + \delta^2 I]_{N-1}$$

$$= \det Q_{N-1} \qquad (4.45)$$

where Q_{N-1} is the tri-diagonal matrix $[(I-\alpha J)(I-\alpha J)^T + \delta^2 I]_{N-1}$

$$Q_{N-1} = \begin{bmatrix} 1+\delta^2 & -\alpha & 0 & 0 & \cdots & & 0 \\ -\alpha & (1+\alpha^2+\delta^2) & -\alpha & 0 & \cdots & & 0 \\ 0 & -\alpha & (1+\alpha^2+\delta^2) & -\alpha & \cdots & & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & -\alpha & (1+\alpha^2+\delta^2) \end{bmatrix} \qquad (4.46)$$

The determinant of such a matrix is well known, however, in order to obtain an expression that is suitable for the present analysis it must be rederived. Observe that Q_{N-1} is almost of the form $(aI-bJ)_{N-1}(aI-bJ)^T_{N-1}$ where

$$a^2 + b^2 = 1 + \alpha^2 + \delta^2 \tag{4.47}$$

and

$$a^2 b^2 = \alpha^2 \tag{4.48}$$

The only difference is in the first element, which instead of equaling $a^2 = 1 + \alpha^2 + \delta^2 - b^2$ is equal to $1 + \delta^2$. Thus

$$Q_{N-1} = (aI-bJ)_{N-1}(aI-bJ)^T_{N-1} + (b^2-\alpha^2)e_1 e_1^T \tag{4.49}$$

where $e_1 = \mathrm{col}(1, 0, \cdots, 0)$ and $e_1 e_1^T$ is a matrix of rank one which is empty except for the first element which is unity. Now let $e_1 = (aI-bJ)z$ where

$$z = (aI-bJ)^{-1}_{N-1}\, e_1 \tag{4.50}$$

then

$$Q_{N-1} = (aI-bJ)_{N-1}\left[I + (b^2-\alpha^2)zz^T\right]_{N-1}(aI-bJ)^T_{N-1} \tag{4.51}$$

and

$$\det(I+\gamma AA^T) = \det Q_{N-1}$$

$$= a^{2(N-1)}[1 + (b^2-\alpha^2)||z||^2] \ . \qquad (4.52)$$

However,

$$z = \text{col}\left(\frac{1}{a} , \frac{b}{a^2} , \cdots , \frac{b^{N-2}}{a^{N-1}}\right) \qquad (4.53)$$

therefore

$$||z||^2 = \frac{1 - \left(\frac{b}{a}\right)^{2(N-1)}}{a^2-b^2} \qquad (4.54)$$

giving

$$\det(I+\gamma AA^T) = \frac{b^{2N}(1-a^2) + a^{2N}(b^2-1)}{b^2 - a^2} \qquad (4.55)$$

The last step is to solve for a^2 and b^2 as functions of ν, and $\frac{\beta^2}{\gamma}$. Observe that

$$a^2 + b^2 = \alpha^2 + 1 + \delta^2$$

$$= \frac{\nu^2\beta^2 + \nu\beta^2 + \nu\gamma - \gamma + \gamma(\nu^2-2\nu+1)}{\nu\beta^2 - \gamma + \nu\gamma}$$

$$= \nu + \frac{\nu\beta^2}{\nu\beta^2 + \nu\gamma-\gamma} \qquad (4.56)$$

and

$$a^2b^2 = \nu \frac{\nu\beta^2}{\nu\beta^2 + \nu\gamma - \gamma} \tag{4.57}$$

Consequently,

$$b^2 = \nu$$

$$= 1 + \rho \tag{4.58}$$

$$a^2 = \frac{\nu\beta^2}{\nu\beta^2 + \nu\gamma-\gamma}$$

$$= \frac{\rho_{fb}}{\rho_{fb} + \rho} \quad , \tag{4.59}$$

and

$$\det(I+\gamma AA^T) = \frac{(1+\rho)^N \ (\rho+\rho_{fb})^{-1} + (1+\rho)^{-N}(1+\rho/\rho_{fb})^{-N}}{1 + (\rho+\rho_{fb})^{-1}} . \tag{4.60}$$

Therefore

$$\sigma_\theta^2 \|g\|^2_{K^{-1}} = \frac{(1+\rho)^N}{\det(I+\gamma AA^T)} - 1$$

$$= \frac{(1+\rho)^N(1+\rho/\rho_{fb})^N - 1}{(\rho+\rho_{fb})^{-1}(1+\rho)^N(1+\rho/\rho_{fb})^N + 1} \tag{4.61}$$

$$< \rho + \rho_{fb}$$

with equality when N is infinite. However, half of the upper bound is achieved with $N = N_B$, where

$$N_B = \frac{\ln(\rho+\rho_{fb})}{\ln(1+\rho)(1+\rho/\rho_{fb})} \tag{4.62}$$

Substitution of $C = W \ln(1+\rho)$, $\epsilon = W \ln(1+\rho/\rho_{fb})$ into Eq.(61) gives

$$P_e = \text{erfc} \sqrt{\frac{3}{2} \; \frac{\sigma_\theta^2 \|g\|^2_{K^{-1}}}{M^2-1}}$$

$$= \text{erfc} \sqrt{\frac{3}{2}\left(1 + (\rho+\rho_{fb})^{-1} e^{2T(C+\epsilon)} {}^{-1}\right)\left(\frac{e^{2T(C+\epsilon)}-1}{e^{2TR}-1}\right)}. \tag{4.63}$$

Thus, for $T < T_B$, where $T_B = \frac{N_B}{2W} \approx \frac{1}{2W} \frac{C_{fb}}{C}$, and $C_{fb} = W \ln(1+\rho_{fb})$ the error decreases almost as fast as in the noiseless feedback case; and approaches the constant value

$$\text{erfc} \sqrt{\frac{3}{2} \; \frac{\rho+\rho_{fb}}{M^2-1}} \quad . \tag{4.64}$$

as T increases beyond the break point T_B.

4.5. Selection of the Block Length.

Unlike noiseless feedback, noisy feedback alone cannot be used to drive the error to zero. Thus, it is necessary to employ one-way coding in addition to, or even in exclusion of, feedback coding.

Let N_t be the total block length of a code consisting of a combination feedback and one-way codes, and let N be the length of the feedback block. The length of the one-way code is then N_t/N, because each block of N signals encoded via feedback can be treated as a new element in the one-way block. The effective signal-to-noise ratio per feedback block is

$$\rho_{eff}(N) = \sigma_\theta^2 \|g(N)\|^2_{K^{-1}(N)}$$

and the error resulting from the combined code is

$$P_e(\rho_{eff}(N), R, N_t) = e^{-E_o(\rho_{eff}, R, N_t)\frac{N_t}{N}}$$

where E_o is the reliability of the one-way code and $R = \frac{2W \ln M}{N_t}$ is the rate of the message source. The length of the feedback code may be determined by selecting N to minimize $P_e(\rho_{eff}, R, N_t)$ or to maximize $\frac{1}{N} E_o(\rho_{eff}, N, R)$.

When the one-way coding procedure is limited to be a repetition of the feedback process (i.e., no one-way coding), the total number of repetitions is N_t/N and the probability of error is

$$P_e(\rho_{eff}, R, N_t) = \text{erfc} \sqrt{\frac{N_t}{N} \; \frac{3}{2} \; \frac{\rho_{eff}}{M^2-1}} \; .$$

It is evident that N must be selected to maximize $\frac{1}{N}\rho_{eff}(N)$ where $N = 1, 2, \cdots, N_t$. In most cases of interest $(\rho_{fb} \gg \rho)$ the optimum choice is $N = \min(N_t, N_B)$. If $\rho_{fb} < \infty$ and $N_t > N_B$ then

$$P_e = \text{erfc} \sqrt{\frac{3}{2} \; \frac{N_t}{N_B} \; \frac{(\rho + \rho_{fb})}{2(M^2-1)}} \; . \tag{4.65}$$

P_e approaches zero as N_t approaches infinity provided that $M^2 < N_t$.

Consequently, $R = \frac{2W}{N_t} \ln M < 2W \frac{1}{N_t} \ln N_t$ approaches zero unless W is infinite. For a more efficient one-way code, the choice of feedback block length becomes even less. The minimum choice for N is of course unity, and it corresponds to the no feedback case.

It is of some interest to observe that in this scheme the average power in both the forward and feedback channels is independent of N, and therefore also of T. That is, P_{av} and P_{fb} are constant with time.

4.6. Mechanization of the Code.

Although the solution of the system is available in matrix form, it is desirable from the standpoint of implementation to derive a more compact representation for the coding and decoding operations. The feedback signals are of course

$$u_i = \beta r_i$$

The transmitted signals are given by

$$s_{i+1} = g_1 \alpha^i \theta - \frac{\nu - 1}{\nu\beta} \sum_{j=1}^{i} \alpha^{i+1-j} v_j$$

$$= \alpha \left(s_i - \frac{\nu-1}{\nu\beta} v_i \right) .$$

Therefore

$$s_{i+1} = \sqrt{\frac{1+\rho}{1+\rho/\rho_{fb}}} \left(s_i - \frac{\rho}{\beta(1+\rho)} v_i \right) \qquad (4.66)$$

with "initial condition"

$$s_1 = \sqrt{\frac{\sigma^2 \rho}{\sigma_\theta^2}} \; \theta$$

The computation of $\hat{\theta}_N$ involves the inversion of $I + \gamma AA^T$. Since

$$I + \gamma AA^T = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & (I+\gamma AA^T)_{N-1} & \\ 0 & & & \end{array}\right],$$

it is obvious that

$$(I+\gamma AA^T)^{-1} = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & (I+\gamma AA^T)^{-1}_{N-1} & \\ 0 & & & \end{array}\right].$$

Next, let $g = g_1 \operatorname{col}(1, \alpha, \alpha^2, \cdots, \alpha^{N-1})$ be partitioned into $g = g_1 \operatorname{col}(1, \alpha y)$ where $y = \operatorname{col}(1, \alpha, \alpha^2, \cdots, \alpha^{N-2})$. Note that $(I-\alpha J)_{N-1} y = e_1$ where $e_1 = \operatorname{col}(1, 0, 0, \cdots, 0)$ and $(aI-bJ)^{-1}_{N-1} e_1 = z$ where $z = \frac{1}{a} \operatorname{col}\left(1, \frac{b}{a}, \frac{b}{a}, \cdots, \frac{b}{a}^{N-2}\right)$. Then, since $K^{-1} = \frac{1}{\sigma^2} (I+\gamma AA^T)^{-1}$,

$$K^{-1} g = \frac{1}{\sigma^2} g_1 \left[\begin{array}{c|ccccc} 1 & 0 & \cdots & 0 & \cdots & 0 \\ \hline 0 & & & & & \\ \vdots & & & (I+\gamma AA^T)^{-1}_{N-1} & & \\ 0 & & & & & \end{array}\right] \left[\begin{array}{c} 1 \\ \hline \alpha y \end{array}\right]$$

$$= \frac{g_1}{\sigma^2} \left[\begin{array}{c} 1 \\ \hline \alpha (I+\gamma AA^T)^{-1}_{N-1} y \end{array}\right]$$

However,

$$(I+\gamma AA^T)^{-1}_{N-1}y = (I-\alpha J^T)_{N-1}(aI-bJ^T)^{-1}_{N-1}\left(I + (b^2-\alpha^2)zz^T\right)^{-1}(aI-bJ)^{-1}(I-\alpha J)_{N-1}y$$

$$= (I-\alpha J)^T_{N-1}(aI-bJ^T)^{-1}_{N-1}\left(I + (\beta^2-\alpha)^2 zz^T\right)^{-1} z$$

$$= \frac{1}{1 + (b^2-\alpha^2)\|z\|^2}(I-\alpha J)^T_{N-1}(aI-bJ^T)^{-1}_{N-1}z \qquad (4.67)$$

and

$$\|g\|^2_{K^{-1}}\hat{\theta}_N = \langle K^{-1}g, (I-A\beta)r\rangle$$

$$= \frac{g_1}{\sigma^2}\begin{bmatrix} 1 & ay^T(I+\gamma AA^T)^{-1}_{N-1}\end{bmatrix}\begin{bmatrix} 1 & \\ \frac{\nu-1}{\nu}\,ay & (I-A\beta)_{N-1}\end{bmatrix}\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N\end{bmatrix}$$

Since $(I-A\beta) = I - \frac{\nu-1}{\nu}\alpha J(I-\alpha J)^{-1} = (I-\alpha J)^{-1}(I - \frac{\alpha}{\gamma}J)$, it follows that $(I-A\beta)_{N-1} = (I-\alpha J)^{-1}_{N-1}(I - \frac{a}{b}J)_{N-1}$, where $\frac{\alpha}{\gamma} = \frac{a}{b}$. Therefore, if $r' = \text{col}(r_2, r_3, \cdots, r_N)$ then

$$\|g\|^2_{K^{-1}}\hat{\theta}_N = \frac{g_1}{\sigma^2}\left[\left(1 + \frac{(\alpha^2-a^2)\|z\|^2}{1 + (b^2-\alpha^2)\|z\|^2}\right)r_1 + \frac{a\langle z, (aI-bJ)_{N-1}(I - \frac{a}{b}J)_{N-1}r'\rangle}{1 + (b^2-\alpha^2)\|z\|^2}\right]$$

$$= \frac{g_1}{b^2\sigma^2}\left(1 + \sigma^2_\theta\|g\|^2_{K^{-1}}\right)\left(r_1 + \frac{a}{b}r_2 + \cdots + \left(\frac{a}{b}\right)^{N-1}r_N\right). \qquad (4.68)$$

$\hat{\theta}_N$ is the minimum variance unbiased estimate of θ, while

$$x_N = \frac{\sigma_\theta^2 \|g\|^2_{K^{-1}}}{1 + \sigma_\theta^2 \|g\|^2_{K^{-1}}} \hat{\theta}_N \tag{4.69}$$

is the minimum variance (biased) estimate of θ. Thus

$$x_N = x_{N-1} + \frac{\sigma_\theta^2 g_1}{\sigma^2 b^2} \left(\frac{a}{b}\right)^{N-1} r_N$$

and

$$x_i = x_{i-1} + \sqrt{\frac{\sigma_\theta^2}{\sigma^2} \frac{\rho}{1+\rho} \left(1 + \frac{\rho}{\rho_{fb}}\right)} \left[(1+\rho)\left(1 + \frac{\rho}{\rho_{fb}}\right)\right]^{-\frac{i}{2}} r_i \tag{4.70}$$

with

$$x_o = 0$$

is the recursion formula to use at the receiver for calculating x_N, $\hat{\theta}_N$ and θ_N^*.

4.7. Selection of the Feedback Signals for a Specific Set of Forward Signals.

The diagonalization of H is equivalent to the minimization of the probability of error for a given amount of transmitted power in the forward channel. The resulting matrix equation $H = \nu I$ represents $N(N+1)/2$ independent scalar equations in the N^2 unknown parameters of A, B and g. Therefore, $N(N-1)/2$ parameters can be determined by other means. Thus, the condition $B = \beta I$ in the preceding sections can be viewed as a particular choice of parameters. It is of course possible to make a different assignment of the $N(N-1)/2$ free parameters in A, B, and g. For example, the matrix A can be specified instead of B. The condition $H = \nu I$ is then sufficient to provide a unique solution for B and g. The feedback energy, $E_{fb} = \nu \text{ Tr } B^T B$, is then indirectly specified by A. In fact, if A is chosen to be as in the preceeding section then B must turn out to equal βI_{N-1}, where I_{N-1} is the $(N-1) \times (N-1)$ identity matrix.

This section considers the problem of determining B and g when A is specified. Let $C = AB(I-AB)^{-1}$, then $(I-AB)^{-1} = I + C$, also let $f = \frac{\sigma_\theta}{\sigma^2} Q^{-1} g$ where $QQ^T = I + \gamma AA^T$ and Q is lower triangular. The matrix C is lower triangular with zeros along the main diagonal, and has a one to one correspondence with B when A has rank N-1. The problem now is to select B and g such that $E_{av} = \text{Tr } H$ is a minimum for a fixed value of $1 + \sigma_\theta^2 \|g\|^2_{K^{-1}}$. However,

$$\begin{aligned} \text{Tr } H &= \text{Tr } \left(I + C\right)\left(\frac{\sigma_\theta^2}{\sigma^2} gg^T + QQ^T\right)\left(I + C\right)^T \\ &= \text{TR}[(Q+CQ)(ff^T+I)(Q^T+Q^TC^T)] \\ &= \text{TR}[(\Lambda+D)(ff^T+I)(\Lambda+D^T)] \end{aligned} \tag{4.71}$$

where $\Lambda = \text{diag } Q(\lambda_{ij} = (Q)_{ij}\delta_{ij})$, and $D = CQ + Q - \Lambda$ is, again, a lower triangular matrix with zeros along the main diagonal. Note that minimization of E_{av} with respect to D is the same as with respect to C since there is a one to one correspondence between C and D. The constraint on $\|g\|^2_{K^{-1}}$ affects only the selection of g.

Let $f(i\text{-}1) = \mathrm{col}(f_1, \cdots, f_{i-1})$ and let $d(i\text{-}1) = \mathrm{col}(d_{i1}, \cdots, d_{ii-1})$ denote the nonvanishing portion of the i-th row of D. Then, since $\mathrm{diag}(\Lambda D) \equiv 0$ so that $\mathrm{TR}(\Lambda D) \equiv 0$,

$$\mathrm{Tr}\, H = \|(\Lambda+D)f\|^2 + \mathrm{Tr}(DD^T+\Lambda^2)$$

$$\frac{E_{av}}{\sigma^2} + N = \sum_{i=1}^{N} (\lambda_i f_i + \langle d(i\text{-}1), f(i\text{-}1)\rangle)^2 + \|d(i\text{-}1)\|^2 + \lambda_i^2\Big]$$

$$= \sum_{i=1}^{N}\left(\frac{e_i}{\sigma^2} + 1\right) \qquad (4.72)$$

where

$$1 + \frac{e_i}{\sigma^2} = (\lambda_i f_i + \langle d(i\text{-}1), f(i\text{-}1)\rangle)^2 + \|d(i\text{-}1)\|^2 + \lambda_i^2 \qquad (4.73)$$

Setting $\mathrm{grad}_{d(i-1)} e_i = 0$ for $i = 2, 3, \cdots, N$ obtains D as a function of f. Thus

$$d(i\text{-}1) = -\frac{\lambda_i f_i}{1 + \|f(i\text{-}1)\|^2}\, f(i\text{-}1) \qquad (4.74)$$

therefore

$$1 + \frac{e_i}{\sigma^2} = \lambda_i^2\left(1 + \frac{f_i^2}{1 + \|f(i\text{-}1)\|^2}\right). \qquad (4.75)$$

Next, since $\sigma_\theta^2\|g\|^2_{K^{-1}} = \|f\|^2$, and because of the identity

$$1 + \|f\|^2 \equiv \prod_{i=1}^{N} \left(1 + \frac{f_i^2}{1 + \|f(i-1)\|^2}\right) \tag{4.76}$$

there results

$$1 + \sigma_\theta^2 \|g\|^2_{K^{-1}} = \frac{\prod_{i=1}^{N} \left(1 + \frac{e_i}{\sigma^2}\right)}{\det(I+\gamma AA^T)} \quad , \tag{4.77}$$

where

$$\prod_{i=1}^{N} \lambda_i^2 = \det(I+\gamma AA^T) \ .$$

The optimum choice for e_i is obtained simply by observing that

$$\prod_{i=1}^{N} \left(1 + \frac{e_i}{\sigma^2}\right) \le \left(1 + \frac{E_{av}}{\sigma^2 N}\right)^N$$

with equality if and only if $e_i = \frac{E_{av}}{N}$ $i = 1, 2, \cdots, N$. Thus, since $\nu - 1 = \rho = \frac{P}{N_o W} = \frac{E_{av}}{\sigma^2 N}$, the result is

$$1 + \sigma_\theta^2 \|g\|^2_{K^{-1}} = \frac{(1+\rho)^N}{\det(I+\gamma AA^T)} \quad . \tag{4.78}$$

The components of the vector f satisfy a recursion formula

$$\frac{f_i^2}{f_{i-1}^2} = \left(\frac{\nu - \lambda_i^2}{\lambda_i^2}\right)\left(\frac{\lambda_{i-1}^2}{\nu - \lambda_{i-1}^2}\right)\left(\frac{1 + \|f(i-1)\|^2}{1 + \|f(i-2)\|^2}\right)$$

$$= \left(\frac{\nu - \lambda_i^2}{\lambda_i^2}\right)\left(\frac{\lambda_{i-1}^2}{\nu - \lambda_{i-1}^2}\right)\left(1 + \frac{f_{i-1}^2}{1 + \|f(i-2)\|^2}\right)$$

$$= \left(\frac{\nu - \lambda_i^2}{\lambda_i^2}\right)\left(\frac{\nu}{\nu - \lambda_{i-1}^2}\right) \tag{4.79}$$

Therefore

$$f_i^2 = \left(\frac{\nu - \lambda_i^2}{\lambda_i^2}\right)\prod_{j=1}^{i-1}\left(\frac{\nu}{\lambda_j^2}\right)$$

$$\left(\frac{\lambda_i^2 f_i^2}{1 + \|f(i-1)\|^2}\right)\left(\frac{f_j}{\lambda_i f_i}\right) = \left[\frac{(\nu - \lambda_i^2)(\nu - \lambda_j^2)}{\lambda_j^2}\prod_{k=j}^{i-1}\left(\frac{\lambda_k^2}{\nu}\right)\right]^{1/2} .$$

Thus,

$$d_{ij} = \begin{cases} \sqrt{(\nu - \lambda_i^2)(\nu - \lambda_j^2)}\, \dfrac{1}{\lambda_j} \displaystyle\prod_{k=j}^{i-1} \left(\dfrac{\lambda_k}{\sqrt{\nu}}\right) & ; j < i \\ 0 & ; j \geq i . \end{cases} \tag{4.80}$$

This provides the solution, because $C = (D-Q+\Lambda)Q^{-1}$ and then $AB = C(I+C)^{-1}$ can be used together with the pseudo inverse $A^{(-1)}$ $(A^{(-1)}A = J)$ to give

$$JB = A^{(-1)}AC(I+C)^{-1} . \tag{4.81}$$

J is a lowering operator that anihilates the last row of B, which is identically zero anyway.

As an example, consider the code with $A = aJ$. This is the simplest lower triangular matrix with a zero diagonal with rank $N-1$. In this case $1 + \sigma_\theta^2\|g\|^2_{K^{-1}} = (1+\rho)^N/(1+\gamma a^2)^{N-1}$ increases exponentially with N for $\gamma a^2 < \rho$. It will be shown, however, that the feedback power is also an exponentially increasing quantity.

Note that $I + \gamma AA^T = I + \gamma a^2 JJ^T = \Lambda^2$ is diagonal with $\lambda_1^2 = 1$ and $\lambda_i^2 = 1 + \gamma a^2$ for $i > 1$. Consequently,

$$d_{ij} = - \begin{cases} \varphi_1 \delta^{i-2} & \text{for } j = 1;\ i = 2, 3, \cdots, N \\ \varphi\delta^{i-j-1} & \text{for } 1 < j < i = 3, 4, \cdots, N \\ 0 & \text{for } j \ge i = 1, 2, \cdots, N \end{cases} \tag{4.82}$$

where $\varphi_1 = (\nu-\lambda^2)(\nu-1)/\nu$, $\varphi = (\nu-\lambda^2)/\nu$ and $\delta = \lambda/\nu$. Thus,

$$D = - \begin{bmatrix} 0 & 0 & & 0 \\ 1 & 0 & & \\ \delta & 1 & & \\ \delta^2 & \delta & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \delta^{N-2} & \delta^{N-3} & 1 & 0 \end{bmatrix} \begin{bmatrix} \varphi_1 & 0 & 0 & & & 0 \\ 0 & \varphi & 0 & & & \\ & & & & & \\ 0 & 0 & \varphi & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & 0 & 0 & 0 & \varphi & 0 \end{bmatrix}$$

$$= - J(I-\delta J)^{-1}\Phi \tag{4.83}$$

where $\varphi_{ij} = \varphi_i \delta_{ij}$ and $\varphi_i = \varphi$ for $i > 1$.

Next, $C = (D-Q+\Lambda)Q^{-1} = D\Lambda^{-1}$, since $Q = \Lambda$ in this case, therefore

$$C = - J(I-\delta J)^{-1}\Phi\Lambda^{-1} \tag{4.84}$$

$$(I+C) = [\Lambda\Phi^{-1}(I-\delta J) - J](I-\delta J)^{-1}\Phi\Lambda^{-1} ,$$

and

$$aJB = C(I+C)^{-1}$$

$$= - J[I - (\delta I+\Phi\Lambda^{-1})J]^{-1} \Phi\Lambda^{-1} .$$

Therefore

$$JB = \frac{1}{a} J(I-\psi J)^{-1}(\psi-\delta I) \qquad (4.85)$$

where $\psi = \delta I + \Phi\Lambda^{-1}$. Since J annihilates only the last, all zero, row

$$B = \frac{1}{a} \begin{bmatrix} 1 & 0 & 0 & & & 0 \\ \psi_1 & 1 & 0 & & & 0 \\ \psi_1\psi & \psi & 1 & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \psi_1\psi^{N-3} & \psi^{N-4} & & \psi & 1 & 0 \\ 0 & 0 & & & & 0 \end{bmatrix} \begin{bmatrix} \psi_1-\delta & 0 & 0 & & 0 & 0 \\ 0 & \psi-\delta & 0 & & & \\ 0 & 0 & \psi-\delta & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & 0 & & 0 & \psi-\delta & 0 \\ 0 & 0 & & & 0 & 0 \end{bmatrix}$$

The feedback energy may be computed from $E_{fb} = \nu \, Tr[B^TB]$ but, it is simpler to use the fact that $Tr[BHB^T] = Tr[JBHB^TJ^T]$ since all that J does is annihilate the last feedback signal, which is zero anyway. Now, however, $JB = A^{(-1)}AB = A^{(-1)}C(I+C)^{-1}$ because $A^{(-1)}A \triangleq J$, thus, $Tr[JBHB^TJ^T] = Tr[A^{(-1)}CQ(ff^T+I)Q^TC^TA^{T(-1)}]$ giving

$$\frac{E_{fb}}{\sigma^2} = \mathrm{Tr}[A^{(-T)}A^{(-1)}D(ff^T+I)D^T]$$

$$\frac{E_{fb}}{\gamma\sigma^2} = \frac{1}{\gamma a^2} \sum_{i=1}^{N-1} (\langle d(i-1), f(i-1)\rangle^2 + \|d(i-1)\|^2)$$

$$= \frac{1}{\gamma a^2} \sum_{i=1}^{N-1} \lambda_i^2 f_i^2 \quad \frac{\|f(i-1)\|^2}{1 + \|f(i-1)\|^2} \;, \qquad (4.86)$$

and

$$\frac{E_{fb}}{\gamma\sigma^2} \approx \frac{1 + \gamma a^2}{\gamma a^2} \|f(i-1)\|^2$$

$$\approx \frac{1}{\gamma a^2} \; \frac{(1+\rho)^{N-1}}{(1+\gamma a^2)^{N-2}} \qquad (4.87)$$

which grows exponentially with N. N is determined by the maximum allowed average feedback power, or signal-to-noise ratio. It can be seen that while it is desirable to make γa^2 small, it comes at the expense of increased feedback power. Also, if a is chosen to give a value for $\det(I+\gamma AA^T)$ which is equal to the same value as in section 4.5 then it is obvious from the fact that B is not diagonal that E_{fb} is greater than that in Section 4.5.

The mechanization of this scheme is simple. The transmitted signals are given by

$$s_{i+1} = g_i \theta + av_{i-1}$$

$$= \frac{g_i}{g_{i-1}} (s_{i-1} - av_{i-2}) + av_{i-1} \qquad (4.88)$$

The feedback signals are of the form

$$u_i = \sum_{j=1}^{i} b_{ij} r_j$$

$$= \psi \left(u_{i-1} + \frac{\psi - \delta}{\psi} r_i \right) \qquad (4.89)$$

The receiver computation of $\hat{\theta}$ is given by

$$\|g\|^2_{K^{-1}} \hat{\theta}_N = g_1 r_1 + \frac{1}{1 + \gamma a^2} \sum_{j=g}^{N} g_j (r_j - a u_{j-1})$$

Therefore

$$\hat{\theta}_i = \hat{\theta}_{i-1} - \frac{g_i^2}{(1+\gamma a^2) g_1^2 + g_2^2 + \cdots + g_i^2} \left[\hat{\theta}_{i-1} - \frac{1}{g_i} (r_i - a u_{i-1}) \right] \qquad (4.90)$$

4.8. Discussion.

The condition $H = \gamma I$ is not optimum for the noisy feedback problem. This may be seen by comparing the present result with a calculation for $N = 2$ made by Elias in 1956 [10]. Elias obtains

$$[\sigma_\theta^2 \|g\|^2_{K^{-1}}]_{\text{Elias}} = \rho_1 + \rho_2 + \frac{\rho_1 \rho_2 \rho_{fb}}{(1+\rho_1)(1+\rho_2) + \rho_{fb}} \tag{4.91}$$

where $\rho_1 = e_1/\sigma^2$ and $\rho_2 = e_2/\sigma^2$. It may be verified that the optimum choice for ρ_1 and ρ_2 when $\rho_1 + \rho_2 = \frac{E_{av}}{\sigma^2}$ is $\rho_1 = \rho_2 = \rho = \frac{E_{av}}{2\sigma^2}$. Thus

$$[\sigma_\theta^2 \|g\|^2_{K^{-1}}]_{\text{Elias}} = 2\rho + \frac{\rho^2 \rho_{fb}}{(1+\rho)^2 + \rho_{fb}} \; . \tag{4.92}$$

Now, when $H = (1+\rho)I$ the result is

$$[\sigma_\theta^2 \|g\|^2_{K^{-1}}]_{H=(1+\rho)I} = \frac{(1+\rho)^2(1+\rho/\rho_{fb}) - 1}{\dfrac{(1+\rho)^2(1+\rho/\rho_{fb})^2}{\rho + \rho_{fb}} + 1} \tag{4.93}$$

and it is clear that the expression in (4.93) is always less than or equal to that in (4.92), with equality occurring at $\rho_{fb} = \infty$.

An explanation as to why $H = \gamma I$ is not optimum is obtained from the requirement that $\det(I+\gamma AA^T)$, in the expression

$$1 + \sigma_\theta \|g\|^2_{K^{-1}} = \frac{\det H}{\det(I+\gamma AA^T)} \; , \tag{4.94}$$

be kept fixed. When $N = 2$ this amounts to holding the entire matrix A (since it has only one non-zero entry) fixed. Thus, in effect the optimization is only over B. Equation (4.91) results from an optimization of (4.94) over A.

It is instructive to go through the steps in greater detail. Without loss of generality, let σ_θ^2/σ^2 and $\gamma = \sigma_m^2/\sigma^2$ be unity, then

$$A = \begin{bmatrix} 0 & 0 \\ a & 0 \end{bmatrix} , \qquad B = \begin{bmatrix} b & 0 \\ 0 & 0 \end{bmatrix} ,$$

$$C = \begin{bmatrix} 0 & 0 \\ ab & 0 \end{bmatrix} , \qquad I + \gamma AA^T = \begin{bmatrix} 1 & 0 \\ 0 & 1+a^2 \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & 0 \\ ab & 1 \end{bmatrix} \begin{bmatrix} 1+g_1^2 & g_1 g_2 \\ g_1 g_2 & 1+g_2^2+a^2 \end{bmatrix} \begin{bmatrix} 1 & ab \\ 0 & 1 \end{bmatrix} .$$

From this it may be seen that ρ_1, ρ_2 and ρ_{fb} , as given by $(H)_{11} - 1$, $(H)_{22} - 1$, and $b^2(H)_{11}$, respectively, are:

$$\rho_1 = g_1^2 \tag{4.95}$$

$$\rho_2 = (g_2+abg_1)^2 + a^2b^2 + a^2 \tag{4.96}$$

$$\rho_{fb} = (1+\rho_1)\, b^2 . \tag{4.97}$$

Now

$$\sigma_\theta^2 \|g\|^2_{K^{-1}} = g_1^2 + \frac{g_2^2}{1 + a^2} \tag{4.98}$$

$$= g_1^2 + \frac{\left(abg_1 + \sqrt{\rho_2 - a^2(1+b^2)}\right)^2}{1 + a^2} \tag{4.99}$$

Maximization over a of the above expression yields equation (4.91). It may be verified that the resulting optimum value of a does not diagonalize H.

The diagonalization of H occurs if g_1, g_2 and a are held fixed, while the minimization of the forward energy $E_{av} = \sigma^2(\rho_1+\rho_2)$ is attempted by varying over b (and thus over ρ_{fb}). Since only ρ_2 depends on b, the indicated choice is $b = -g_1g_2/a(1+g_1^2)$. Equation (4.93) is obtained when $a = g_1g_2/b(1+g_1^2)$ and $b = \rho_{fb}/(1+g_1^2)$ are substituted into (4.98), followed by the setting of $\rho_1 = \rho_2 = \rho$.

The optimization of $\|g\|^2_{K^{-1}}$ over A for values of $N > 2$ has, up to the present, proven to be practically impossible.

V. RECURSIVE CODING FOR NOISY FEEDBACK

5.1. Introduction.

This chapter develops a linear coding procedure, for the AWGN channel with an AWGN feedback link, using an approach that differs from the matrix methods in the preceding chapters. The approach is suggested by the form of the noiseless feedback scheme, where the receiver employs a first order difference equation to compute the minimum variance biased and unbiased estimates of the message. Here, the feedback signals are assumed to be the successive estimates of the message. The strategy of the transmitter is to compute the optimum estimate $\overline{\theta}_i$ of the receiver's estimate $\hat{\theta}_i$, and to transmit

$$s_i = g_i(\theta - \overline{\theta}_{i-1}) \qquad i = 1, 2, \cdots, N \qquad (5.1)$$

Because of the initial structural assumptions, it is not possible to claim optimality. In fact, the recursive estimation performed by the receiver is optimum only in the sense that the receiver computes the best estimate from the data contained in the preceding estimate and the latest observation. This can, however, be generalized to a receiver that takes a linear combination of all the preceding estimates and the latest observation, which is equivalent to taking a weighted sum of all the observations and varying the weighting coefficients so as to minimize the variance.

5.2. Description of the process.

The coding procedure is diagrammed in Fig. 4. It is in principle equivalent to that of Omura [17] and Kashyap [18] except that an optimization over the receiver's parameters is also carried out. This last step is important and it leads to a significant improvement in system performance. It might be added that the work described in this chapter predates the matrix approach of the earlier chapters, and, like that of [17] and [18], was prompted by the interesting paper of Schalkwijk and Kailath [15,16] who use the Robbins-Monro method of stochastic approximation.

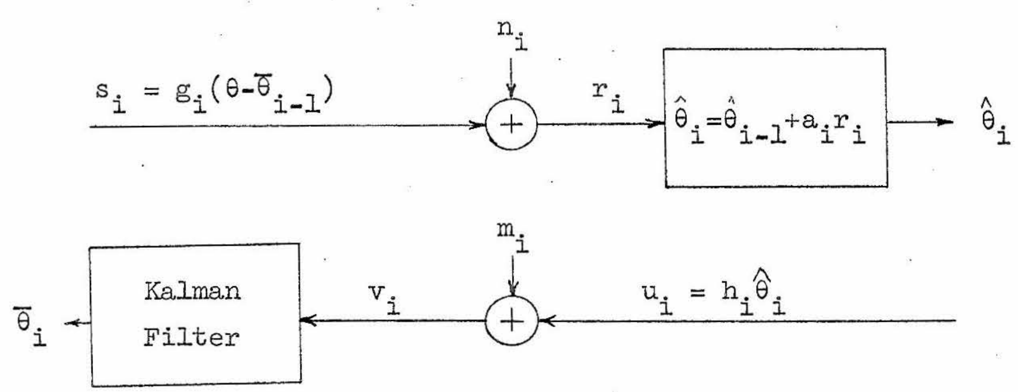

Fig. 4. Feedback Coding with Recursive Estimation at the Transmitter.

The first signal is $s_1 = g_1\theta$, the receiver observes $r_1 = s_1 + n_1$ and computes the minimum variance unbiased estimate $\hat{\theta}_1 = \frac{r_1}{g_1}$. The first feedback signal is $u_1 = h_1\hat{\theta}_1$, the transmitter observes $v_1 = u_1 + m_1$ and computes the best estimate of $\hat{\theta}_1$ from its knowledge of θ and v_1. By "best" it is meant that $E[(\bar{\theta}_1-\hat{\theta}_1)^2]$ is a minimum, where $\bar{\theta}_1$ is the transmitter's estimate of $\hat{\theta}_1$. It is given by

$$\bar{\theta}_1 = g_1\theta + \frac{p_1h_1^2}{\sigma_m^2}\left(\frac{v_1-g_1\theta}{h_1}\right) \tag{5.2}$$

where p_1 is chosen to minimize $E[(\overline{\theta}-\theta_1)^2]$. Let $\hat{\theta}_1 = \overline{\theta}_1+\widetilde{\theta}_1$ then

$$E[\tilde{\theta}_1^2] = \left(1 - \frac{p_1 h_1^2}{\sigma_m^2}\right)^2 \frac{\sigma^2}{g_1^2} + \left(\frac{p_1 h_1^2}{\sigma_m^2}\right)^2 \frac{\sigma_m^2}{h_1^2} \tag{5.3}$$

taking $\frac{\partial}{\partial p_1} E[\tilde{\theta}_1^2] = 0$ gives

$$p_1 = \frac{1}{\frac{h_1^2}{\sigma_m^2} + \frac{g_1^2}{\sigma^2}} \tag{5.4}$$

which, after substitution into (5.3) provides

$$E[\tilde{\theta}_1^2] = p_1 . \tag{5.5}$$

The form of the transmitted signals is taken to be $s_i = g_i(\theta-\overline{\theta})$, the received observations are $r_i = s_i + n_i$ and the receivers recursive determination of $\hat{\theta}_i$ is assumed to obey

$$\hat{\theta}_i = \hat{\theta}_{i-1} + a_i r_i \qquad i = 1, 2, \cdots , N \tag{5.6}$$

where a_i is to be determined. The coefficient before $\hat{\theta}_{i-1}$ is unity because it is required that $E[\hat{\theta}_i/\theta] = \theta$ for all i. Since $E[r_i/\theta] = E[g_i(\theta-\overline{\theta})+n_i/\theta]=E[g_i(\hat{\theta}-\theta_i) + g_i\widetilde{\theta}_i/\theta]=0$; it follows that $E[\hat{\theta}_i/\theta] = E[\hat{\theta}_{i-1} + a_i r_i/\theta]=E[\hat{\theta}_{i-1}/\theta] = \theta$.

5.3. Recursive Estimation at the Transmitter (Kalman Filtering).

The Kalman Bucy method for estimating the output of a linear dynamical system developed in 1960-61 [20] is directly applicable to the present situation. Therefore, let $\bar{\theta}_i$ be the best estimate of $\hat{\theta}_i$ that the transmitter can make on the basis of the feedback observations $v_1, v_2, \cdots, v_i$ and the knowledge of θ. The Kalman-Bucy filtering equation for $\bar{\theta}_i$ is then

$$\bar{\theta}_i = \bar{\theta}_{i-1} + a_i g_i(\theta-\bar{\theta}_{i-1}) + \frac{h_i^2 p_i}{\sigma_m^2}\left(\frac{v_i}{h_i} - [\bar{\theta}_{i-1} - a_i g_i(\theta-\bar{\theta}_{i-1})]\right). \quad (5.7)$$

The transmitter error, $\tilde{\theta}_i = \hat{\theta}_i - \bar{\theta}_i$, is

$$\tilde{\theta}_i = (\tilde{\theta}_{i-1} + a_i n_i)\left(1 - \frac{h_i^2 p_i}{\sigma_m^2}\right) - \frac{h_i^2 p_i}{\sigma_m^2}\frac{m_i}{h_i} \quad . \quad (5.8)$$

Since n_i and m_i are statistically independent and white,

$$E[\tilde{\theta}_i^2] = \left(E[\tilde{\theta}_{i-1}^2] + a_i^2\sigma\right)\left(1 - \frac{h_i^2 p_i}{\sigma_m^2}\right)^2 + \left(\frac{h_i^2 p_i}{\sigma_m^2}\right)^2 \frac{\sigma_m^2}{h_i^2} \quad . \quad (5.9)$$

Since $E[\tilde{\theta}_i^2]$ is a minimum, it must be a minimum over p_i. Thus, let $\frac{\partial}{\partial p_i} E[\tilde{\theta}_i^2] = 0$ and solve for p_i to find that

$$p_i = \left(1 - \frac{p_i h_i^2}{\sigma_m^2}\right)\left(E[\tilde{\theta}_{i-1}^2] + a_i\sigma^2\right) , \quad (5.10)$$

which, after substitution into (5.9) gives

$$E[\tilde{\theta}_i^2] = p_i \left(1 - \frac{h_i^2 p_i}{\sigma_m^2}\right) + \left(\frac{h_i^2 p_i}{\sigma_m^2}\right)^2 \frac{\sigma_m^2}{h_i^2} \quad (5.11)$$

$$= p_i \quad . \quad (5.12)$$

Now Equation (5.11) reads

$$\frac{1}{p_i} = \frac{h_i^2}{\sigma_m^2} + \frac{1}{p_{i-1} + a_1^2\sigma^2} \tag{5.13}$$

or setting $q_i = \frac{g_i^2 p_i}{\sigma^2}$, $\frac{1}{\gamma_i} = \frac{h_i^2\sigma^2}{g_i^2\sigma_m^2}$ and $w_i^2 = a_i^2 g_i^2$ gives

$$\frac{1}{q_i} = \frac{1}{\gamma_i} + \frac{1}{\frac{g_i^2}{g_{i-1}^2} q_i + w_i^2} \tag{5.14}$$

5.4. Recursive Estimation at the Receiver.

The receiver is supposed to compute the minimum variance estimate $\hat{\theta}_i$ of θ given only the preceding estimate $\hat{\theta}_{i-1}$ and the newest observation r_i. Thus,

$$\hat{\theta}_i = \hat{\theta}_{i-1} + a_i r_i \tag{5.15}$$

$$\hat{\theta}_i - \theta = \hat{\theta}_{i-1} - \theta + w_i(\theta - \overline{\theta}_{i-1}) + \frac{w_i}{g_i} n_i \tag{5.16}$$

$$= (\overline{\theta}_{i-1} - \theta)(1 - w_i) + \tilde{\theta}_{i-1} + \frac{w_i}{g_i} n_i \tag{5.17}$$

Now note that $E[\overline{\theta}_i \tilde{\theta}_i] = 0$. This may be verified by direct computation from equations (7),(8), and (10) for $\overline{\theta}_i$ and $\tilde{\theta}_i$ and p_i. It is simply a statement of the projection theorem, which states that the error θ_i is statistically orthogonal to the estimate $\overline{\theta}$. Since $\tilde{\theta}_i$ is independent of θ it follows that $E[\theta\tilde{\theta}_i] = 0$. Therefore,

$$E[(\hat{\theta}_i-\theta)^2] = E[(\overline{\theta}_i+\tilde{\theta}_i-\theta)^2] \tag{5.18}$$

$$= E[(\overline{\theta}_i-\theta)^2] + E[\tilde{\theta}_i^2] \ . \tag{5.19}$$

Letting $\alpha_i = \frac{g_1^2}{\sigma^2} E[(\hat{\theta}_i-\theta)^2]$, and $\beta_i = E[(\overline{\theta}_i-\theta)^2]$ gives

$$\frac{\sigma^2}{g_i^2}\alpha_i = \beta_i^2 + p_i^2 \qquad i = 1,\ 2,\ \cdots\ ,\ N \tag{5.20}$$

From (5.17) and (5.19) and (5.20) it follows that

$$E[(\hat{\theta}_i-\theta)^2] = (1-w_i)^2\ E[(\overline{\theta}_{i-1}-\theta)^2] + w_i^2\frac{\sigma^2}{g_i^2} + E[\tilde{\theta}_{i-1}^2] \tag{5.21}$$

$$\frac{\sigma^2}{g_i^2}\alpha_i = (1-w_i)^2\ \beta_{i-1}^2 + w_i^2\frac{\sigma^2}{g_1^2} + p_{i-1} \ . \tag{5.22}$$

The minimization over w_i is straight forward; $\frac{\partial\alpha_i}{\partial w_i} = 0$ gives

$$w_i = \frac{\beta_{i-1}}{\beta_{i-1} + \frac{\sigma^2}{g_1^2}}\ , \tag{5.23}$$

which, when substituted into (5.22), gives

$$\frac{\sigma^2}{g_i^2}\alpha_i = w_i\frac{\sigma^2}{g_i^2} + p_{i-1} \tag{5.24}$$

$$\alpha_i = w_i + \frac{g_i^2}{g_{i-1}^2} q_{i-1} \ . \tag{5.25}$$

The inverse of both sides of (5.23) is

$$\frac{1}{w_i} = 1 + \frac{1}{\beta_{i-1} \frac{g_i^2}{\sigma^2}} \tag{5.26}$$

substituting (5.20) obtains

$$\frac{1}{w_i} = 1 + \frac{1}{(\alpha_{i-1} - q_{i-1}) \frac{g_i^2}{g_{i-1}^2}} \quad . \tag{5.27}$$

In summary, the operation of the system is completely characterized by Equations (5.14), (5.25) and (5.27). They are restated for future convenience as the system of coupled equations (5.28abc).

$$\frac{1}{w_i} = 1 + \frac{1}{(\alpha_{i-1} - q_{i-1}) \lambda_i} \qquad \text{(a)}$$

$$\frac{1}{q_i} = \frac{1}{\gamma_i} + \frac{1}{w_i^2 + \lambda_i q_{i-1}} \qquad \text{(b)} \tag{5.28}$$

$$\alpha_i = w_i + \lambda_i q_{i-1} \qquad \text{(c)}$$

where $\lambda_i = g_i^2/g_{i-1}^2$ and $\gamma_i = \sigma_m^2 g_i^2/\sigma^2 h_i^2$. The initial conditions for starting the system of equations (5.28abc) are $w_1 = 1/g_1$ and $q_0 = 0$.

After the sequences $\{\lambda_i\}_{i=1}^N$ and $\{\gamma_i\}_{i=1}^N$ are specified, the system (528abc) can be simulated on a digital machine and the parameters of the code $a_i = w_i/g_i$, and $p_i = \sigma^2 q_i/g_i^2$ determined. These numbers can be precomputed and stored at the transmitter and receiver, or special purpose computers can be built to generate them iteratively at each station.

5.5. Equations for P, P_{fb}, P_e and R_c.

The average power transmitted in the forward channel is

$$P = \frac{1}{T} \sum_{i=1}^{N} E[g_i^2(\theta-\bar{\theta}_{i-1})^2]$$

but $E[(\theta-\bar{\theta}_{i-1})^2] = \beta_{i-1} = (\alpha_{i-1} - q_{i-1})/g_{i-1}^2$, thus,

$$P = \frac{2W\sigma^2}{N} [g_1 \frac{\sigma_\theta^2}{\sigma^2} + \sum_{i=2}^{N} \lambda_i(\alpha_{i-1}-q_{i-1})] \quad . \tag{5.29}$$

Similarly, the average feedback power is

$$P_{fb} = \frac{2W}{N-1} \sum_{i=1}^{N-1} h_i^2 \, E[\hat{\theta}_i^2]$$

but, $E[\hat{\theta}_i^2] = E[(\hat{\theta}_i-\theta+\theta)^2] = E[(\hat{\theta}_i-\theta)^2] + E[\theta^2]$, therefore,

$$P_{fb} = \frac{2W\sigma^2}{N-1} \sum_{i=1}^{N-1} \frac{h_i^2}{g_i^2} (\alpha_i + g_i^2\sigma_\theta^2) \quad . \tag{5.30}$$

The error probability after N iterations is found by the same method as in Section 2.4 with $\sigma_\theta^2\|g\|_{K^{-1}}^2$ replaced by $\sigma_\theta^2 g_N^2/\sigma^2\alpha_N$. Thus

$$P_e = \text{erfc}\sqrt{\frac{3}{2}\;\frac{\sigma_\theta^2 g_N^2}{\alpha_N(M^2-1)}}\;, \tag{5.31}$$

and the expression for the critical rate is

$$R_c(N) = \frac{1}{2T}\ \ln\left(1 + \frac{\sigma_\theta^2 g_N^2}{\sigma^2 \alpha_N}\right) \;. \tag{5.32}$$

5.6. Asymptotic Performance of the Code.

The performance of the scheme in the limit of large N can be determined after establishing the limiting form of α_N, w_N and q_N as N approaches infinity.

There are two situations of interest. One is the infinite bandwidth case, which is obtained by setting $\lambda_i = \lambda = 1$, and $\gamma_i = \gamma$. The other is the finite bandwidth case which is obtained by setting $\lambda_i = \lambda > 1$ and $\gamma_i = \gamma$. It can be seen that $g_i^2 = g_{i-1}^2\lambda = g_1^2\lambda^i$ implies that the feedback power must grow exponentially with N when $\lambda > 1$, and linearly when $\lambda = 1$. This follows from the fact that $h_i^2 = \frac{\sigma_m^2}{\gamma\sigma^2}\, g_i^2$. It is also evident that the feedback power is zero when $\sigma_m^2 = 0$.

Theorem 3.

(a) The assumption that $w_N = w$, and $q_N = q$, are constants (independent of N) for $\lambda > 1$, is consistent with the system of equations (5.28abc).

(b) The assumption that w_N, and q_N are of the limiting form $w_N = \frac{w}{N-1}$ and $q_N = \frac{q}{N}$, with $q = w(w-1)$ and $w(w-1)^2/(2w-1) = \gamma$ as

$N \to \infty$, is also consistent with (5.28abc).

Proof.

The proof is obtained by assuming the hypothesis for N and showing that it satisfies (5.28abc). First it is convenient to eliminate α_i by substituting (5.28c) into (5.28a)

$$\frac{1}{w_i} = 1 + \frac{1}{(w_{i-1} + \lambda q_{i-2} - q_{i-1})\lambda} \quad ; \tag{5.33}$$

also,

$$\frac{1}{q_i} = \frac{1}{\gamma} + \frac{1}{w_i^2 + \lambda q_{i-1}} \quad . \tag{5.28b}$$

(a) Let $w_i = w$ and $q_i = q$. Then (5.33) and (5.28) provide two simultaneous equations in w and q as functions of λ and γ,

$$\frac{1}{w} = 1 + \frac{1}{\lambda w + \lambda(\lambda-1)q} \tag{5.34}$$

$$\frac{1}{q} = \frac{1}{\gamma} + \frac{1}{w^2 + \lambda q} \quad . \tag{5.35}$$

It is easy to see that $\lambda > 1$ is necessary, for if $\lambda = 1$, there would be $\frac{1}{w} = 1 + \frac{1}{w}$ which is impossible. ($\lambda < 1$ is not considered since it implies $g_N^2 \to 0$).

(b) Let $w_{i+1} = \frac{w}{i}$, and let $q_{i+1} = \frac{q}{i+1}$ then (5.33) (for $\lambda = 1$) gives,

$$\frac{i}{w} = 1 + \frac{1}{\frac{w}{i-1} + \frac{q}{i-1} - \frac{q}{i}}$$

$$= 1 + \frac{i-1}{w} \frac{1}{1 + \frac{q}{w_i}}$$

$$= 1 + \frac{i-1}{w} \quad 1 - \frac{q}{w_i} + O(\frac{1}{i})$$

as $i \to \infty$ all the terms $O(\frac{1}{i})$ are negligible. By equating the surviving terms there results

$$q = w(w-1) \ . \qquad (5.36)$$

Similarly, from (5.28b)

$$\frac{i+1}{q} = \frac{1}{\gamma} + \frac{1}{\frac{w^2}{i^2} + \frac{q}{i}}$$

$$= \frac{1}{\gamma} + \frac{i}{q} \frac{1}{1 + \frac{w^2}{q_i}}$$

$$= \frac{1}{\gamma} + \frac{i}{q} - \frac{w^2}{q^2} + O(\frac{1}{i}) \ ,$$

thus

$$w^2 + q = \frac{q^2}{\gamma} \ . \qquad (5.37)$$

By substituting for q from (5.36) it is evident that

$$2w^2 - w = \frac{w^2(w-1)^2}{\gamma}$$

or

$$\frac{w(w-1)^2}{2w-1} = \gamma \tag{5.38}$$

thus completing the proof.

5.7. The Wideband and Narrowband Rates.

The asymptotic form for α_N is obtained from (5.28c) which in the infinite bandwidth case $(\lambda=1)$ is

$$\alpha_N = w_N + q_{N-1}$$

$$= \frac{w}{N-1} + \frac{w^2-w}{N-1}$$

$$= \frac{w^2}{N-1} \tag{5.39}$$

The average power, from (5.29) is

$$\frac{P}{2\sigma^2} = \frac{1}{2T}\left(g_1 \frac{\sigma_\theta^2}{\sigma^2} + \sum_{i=2}^{N} (\alpha_{i-1} - q_{i-1}) \right)$$

$$\approx \frac{w}{2T} \ln(N-1) \tag{5.40}$$

The critical rate is therefore

$$R_c(N) = \frac{1}{2T} \ln \left(1 + \frac{\sigma_\theta^2}{\sigma^2} \; \frac{g_1^2(N-1)}{w^2} \right)$$

$$= \frac{1}{w} \frac{P}{2\sigma^2} \; \frac{\ln\left[1 + \frac{\sigma_\theta^2 g_1^2}{\sigma^2 w^2} (N-1)\right]}{\ln(N-1)}$$

Taking the limit as $N \to \infty$ gives

$$R_c(\infty) = \frac{1}{w} C_\infty \tag{5.41}$$

where $C_\infty = \frac{P}{2\sigma^2}$ and $\frac{w(w-1)^2}{2w-1} = \gamma$. The nontrivial solution for w when $\gamma = 0$ is $w = 1$, thus achieving C_∞ when the feedback becomes noiseless.

The feedback power as given by (5.30) and (5.39) grows linearly with the bandwidth W, or N.

$$\frac{P_{fb}}{2\sigma_m^2} = \frac{W}{\gamma} \; \frac{1}{N-1} \sum_{i=1}^{N-1} (\alpha_i + g_1^2 \sigma_\theta^2)$$

$$= \frac{W}{\gamma} \; \frac{w^2 \ln(N-1)}{N-1} + g_1^2 \sigma_\theta^2$$

$$\approx \frac{W}{\gamma} g_1^2 \sigma_\theta^2 \tag{5.42}$$

Note that the forward power is increasing only logarithmically with N, thus the relative growth of P_{fb} to that of P is exponential. The same arises in the bandlimited case $(\lambda > 1)$, except that P increases linearly with N while P_{fb} grows exponentially.

That the bandwidth is finite for $\lambda > 1$ can be seen from

$$R_c(N) = \frac{1}{2T} \ln \left(1 + \frac{\sigma_\theta}{\sigma^2} \frac{g_N^2}{\alpha_N}\right)$$

$$= \frac{1}{2T} \ln \left(1 + \frac{\sigma_\theta}{\sigma^2} \frac{\lambda^N}{\alpha}\right) \quad (5.43)$$

where again from (5.28c), but with $\lambda > 1$

$$\alpha = w + \lambda q \quad (5.44)$$

and, w and q are given by the solution to (5.34) and (5.35). Thus, as $N \to \infty$ (5.43) yields

$$R_c(\infty) = W \ln \lambda \quad . \quad (5.45)$$

This time

$$\frac{P}{2\sigma^2 W} = \frac{1}{N} [g_1 \theta^2 + (N-1)(\alpha - q)\lambda]$$

or

$$= \frac{w}{1-w} \quad (5.46)$$

When the feedback noise is small, $\gamma \ll 1$, (5.35) indicates that $q \approx \gamma$, and equation (5.34) can be used to obtain

$$w \approx \frac{\lambda-1}{\lambda} + \gamma$$

and

$$\lambda \approx \frac{\left(1 + \frac{P}{2W}\right)}{1 + \gamma\left(1 + \frac{P}{2W}\right)} \quad . \tag{5.47}$$

The improvement due to the partial optimization of the receiver operation can be appreciated by comparing the present calculation for $R_c(\infty)$ in the wide band case, to that of Kashyap [18], who retains the noiseless receiver with $\hat{\theta}_i = \hat{\theta}_{i-1} + \frac{1}{i} r_i$. Kashyap obtains

$$R_c(\infty) \approx \frac{C_\infty}{1 + 4.5\gamma + \frac{1}{2}\sqrt{\gamma^2 + 4\gamma}} \tag{5.48}$$

$$\approx \begin{cases} \dfrac{C_\infty}{1 + \sqrt{\gamma}} & \text{for } \gamma < .01 \\[2ex] \dfrac{C_\infty}{1 + 5\gamma} & \text{for } \gamma \geq 1 \end{cases} \tag{5.49}$$

The result obtained here is $R_c(\infty) = C_\infty/w(\gamma)$, which gives

$$R_c(\infty) = \begin{cases} \dfrac{C_\infty}{1 + \sqrt{\gamma}} & \text{for } \gamma < .1 \\[2ex] \dfrac{C_\infty}{1 + \sqrt{2\gamma}} & \text{for } \gamma > 10 \end{cases} \tag{5.50}$$

VI. CONCLUSIONS

The work described in this report employs a new and general linear formulation of the feedback communication problem. The utility of the formulation has been demonstrated in Chapter III, by the first derivation of the best linear coding procedure for systems with arbitrary Gaussian forward noise and noisefree feedback. The formulation was used in Chapter IV to obtain a partially optimum scheme for a system with independent additive white Gaussian noise in both the forward and return channels. Chapter V describes a recursive procedure for calculating the parameters for the noisy feedback problem, with new results, concerning improved estimation at the receiver, being obtained.

An important advantage of the general approach is in the derivation of the optimum receiver, and the expression for the variance, $\sigma_\theta^2 \|g\|^2_{K^{-1}}$. This obviates the need for finding the optimum receiver configuration, or its performance. Thus, an optimization over the receiver parameters, such as is necessary in Chapter V and also in Refs.[17] and [18] is eliminated.

Of interest is a comparison of the performance of the noisy schemes in Chapters IV and V to the result of Elias, who found the optimum for $N = 2$. To be very simple, let $\rho = 1$, and $\rho_{fb} = 2$. Then, if ρ_{eff} is the resulting signal to noise ratio, Elias obtains $\rho_{eff} = 2\frac{1}{3}$, the method in Chapter IV gives $\rho_{eff} = 1\frac{2}{3}$, while the approach in Chapter V yields $\rho_{eff} = 1\frac{1}{3}$.

It is not possible to say how well the optimum system will perform when $N > 2$. The solution of this problem would constitute an important future contribution. It is conjectured here that the general behavior will still be one that requires the feedback power to be exponentially related to the forward power.

As a matter of general principle, there exist problems in learning theory, or pattern recognition, that closely parallel the feedback communication problem, and which would be amenable to the same analysis. A case in point is the problem of learning with a teacher. (The teacher, naturally, provides feedback information.)

APPENDIX I
PROOF OF THEOREM II

Proof of (a)

Let $x_i = \frac{e_i}{\sigma^2}$ and let $x = \frac{P}{N_o W}$, then

$$R_c(\infty) = \frac{P}{N_o} \lim_{N\to\infty} \frac{\sum_{i=1}^{N} \ln(1+x_i)}{\sum_{i=1}^{N} x_i}$$

By hypothesis, given any $\epsilon > 0$ there exists an integer N_ϵ such that $x - \epsilon \le x_i \le x + \epsilon$ for all $i \ge N_\epsilon$. Therefore, if $A = \sum_{i=1}^{N_\epsilon - 1} \ln(1+x_i)$ and $B = \sum_{i=1}^{N_\epsilon - 1} x_i$,

$$\sum_{i=1}^{N} \ln(1+x_i) = A + \sum_{i=N_\epsilon}^{N} \ln(1+x_i)$$

$$\le A + (N-N_\epsilon+1) \ln(1+x_i+\epsilon)$$

and

$$\sum_{i=1}^{N} x_i = B + \sum_{i=N_\epsilon}^{N} x_i$$

$$\ge B + (N-N_\epsilon+1)\ (x-\epsilon)$$

consequently,

$$R_c(\infty) \le \frac{P}{N_o} \lim_{N\to\infty} \left[\frac{\frac{A}{N-N_\epsilon+1} + \ln(1+x+\epsilon)}{\frac{B}{N-N_\epsilon+1} + x - \epsilon} \right]$$

$$= \frac{P}{N_o} \frac{\ln(1+x+\epsilon)}{x - \epsilon}$$

similarly,

$$R_c(\infty) \geq \frac{P}{N_o} \; \frac{\ln(1+x-\epsilon)}{x+\epsilon}$$

Since this is true for all $\epsilon > 0$ no matter how small, it follows that

$$R_c(\infty) = \frac{P}{N_o} \; \frac{\ln(1+x)}{x}$$

$$= C_W$$

Proof of (b)

Since $x_i \geq 0$ for all i and $\lim_{i\to\infty} x_i = 0$ it follows that for any $\epsilon > 0$ there exists an integer N_ϵ such that $x_i \leq \epsilon$ for all $i \geq N_\epsilon$. Now, $\ln(1+x_i) \geq x_i - \frac{1}{2} x_i^2$ therefore

$$R_c(\infty) \geq \frac{P}{N_o} \lim_{N\to\infty} \left[\frac{A + \sum_{i=N_\epsilon}^{N} x_i(1 - \frac{1}{2} x_i)}{B + \sum_{i=N_\epsilon}^{N} x_i} \right]$$

but $1 - \frac{x_i}{2} \geq 1 - \frac{\epsilon}{2}$ thus,

$$R_c(\infty) \geq \frac{P}{N_o} \lim_{N\to\infty} \left[\frac{A + (1 - \frac{1}{2}\epsilon) \sum_{i=N_\epsilon}^{N} x_i}{B + \sum_{i=N_\epsilon}^{N} x_i} \right]$$

$$\geq (1 - \tfrac{1}{2}\epsilon) \frac{P}{N_o}$$

The upper bound is simplest, $\ln(1+x_i) \le x_i$ for any $x_i \ge 0$ therefore

$$R_c(\infty) \le \frac{P}{N_o} \lim_{N\to\infty} \left(\frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i} \right),$$

consequently, $R_c(\infty) = \frac{P}{N_o}$.

Proof of (c)

By hypothesis, given any number $M < \infty$ there exists a number N_M such that $1 + x_i \ge e^M$ for all $i \ge N_M$. Let $y_i = \ln(1+x_i)$ then $\lim_{i\to\infty} y_i = \infty$ and $y_i \ge M$ for all $i \ge N_M$. Now,

$$0 \le R_c(\infty) = \frac{P}{N_o} \lim_{N\to\infty} \left(\frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} (e^{y_i}-1)} \right)$$

$$\le \frac{P}{N_o} \lim_{N\to\infty} \left[\frac{\sum_{i=1}^{N_M-1} y_i + \sum_{i=N_M}^{N} y_i}{\sum_{i=1}^{N_M-1} (e^{y_i}-1) + \sum_{i=N_M}^{N} (y_i + \frac{1}{2} y_i^2 + \cdots)} \right]$$

$$\le \frac{P}{N_o} \lim_{N\to\infty} \left[\frac{A + \sum_{i=N_M}^{N} y_i}{B + (1 + \frac{1}{2M}) \sum_{i=N_M}^{N} y_i} \right]$$

thus

$$0 \le R_c(\infty) \le \frac{P}{N_o} \frac{1}{1 + \frac{1}{2} M}$$

Because M is arbitrary and can be chosen as large as desired, $R_c(\infty)$ must be zero.

APPENDIX II
THE CRITICAL RATE OF A FIRST ORDER MARKOV CHANNEL WITH FEEDBACK

First order Markov noise is described by a first order linear difference equation which is driven by a white Gaussian process. Thus

$$n_i = \alpha n_{i-1} + w_i \qquad i = 1, 2, \cdots, N \tag{II.1}$$

hence

$$\begin{bmatrix} n_1 \\ n_2 \\ \cdot \\ \cdot \\ \cdot \\ n_N \end{bmatrix} = \alpha \begin{bmatrix} 0 & 0 & & & & 0 \\ 1 & 0 & & & & \\ 0 & 1 & & & & \\ \cdot & 0 & & & & \\ \cdot & \cdot & & & & \\ 0 & 0 & \cdot\cdot & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ \cdot \\ \cdot \\ \cdot \\ n_N \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ \cdot \\ w_N \end{bmatrix} + \alpha \begin{bmatrix} n_0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

thus

$$n = \alpha J n + w' \tag{II.2}$$

where $(J)_{ij} = \delta_{ij+1}$, $n = \text{col}(n_1, n_2, \cdots n_N)$, $w = \text{col}(w_1, w_2, \cdots w_N)$, $w' = \text{col}(w_1 + \alpha n_0, w_2, w_3, \cdots, w_N)$, and $K_W = E[ww^T] = \sigma_w^2 I = \frac{N_o}{2} I$.

Stationarity requires that the covariance of the noise $K_n = E[nn^T]$ be given by $(K_n)_{ij} = k_{ij} = E[n_i n_j] = k(|i-j|)$. This condition is satisfied if and only if $k_{ii} = E[n_i^2] = \sigma^2$ for all $i = 1, 2, \cdots, N$. By statistical independence, therefore

$$\sigma^2 = \alpha^2\sigma^2 + \sigma_w^2$$

giving

$$\sigma^2 = \frac{\sigma_w^{\ 2}}{1 - \alpha^2} \tag{II.4}$$

which has meaning if and only if $\alpha^2 < 1$, and it implies that $E[n_o^{\ 2}] = \sigma^2 = \sigma_w^{\ 2}/(1-\alpha^2)$. Since $n = (I-\alpha J)^{-1}\, w'$,

$$K_n = (I-\alpha J)^{-1} K_{w'} (I-\alpha J^T)^{-1} \tag{II.5}$$

where $K = E[w'w'^T]$ is diagonal, and except for the fact that $(K_{w'})_{11} = E[(w_1+\alpha n_o)^2] = E[w_1^{\ 2}] + \alpha^2\, E[n_o^{\ 2}] = \sigma_w^{\ 2} + \alpha^2\sigma_w^{\ 2}/(1-\alpha^2) = \sigma_w^{\ 2}/(1-\alpha^2)$, its elements are the same as those of K_w. From (II.5) it may be verified that

$$k_{ij} = \sigma_w^{\ 2}\ \frac{\alpha^{|i-j|}}{1 - \alpha^2} \tag{II.6}$$

$$= \sigma^2\ \alpha^{|i-j|}$$

The elements of the lower triangular matrix Q, from $QQ^T = K_n$, are therefore

$$q_{ij} = \begin{cases} \sigma\, \alpha^{i-1} & \text{for } 1 = j \leq i \\ \sigma_w \alpha^{i-j} & \text{for } 1 < j \leq i \\ 0 & \text{for } i < j \end{cases} \tag{II.7}$$

Also, it is obvious that $Q^{-1} = (I-\alpha J)K_{w'}^{\frac{1}{2}}$, so that $h_{11} = (Q^{-1})_{11} = 1 - \alpha^2/\sigma_w$, while $h(i) = \text{col}(h_{i1}, h_{i2}, \cdots h_{ii}) = \frac{1}{\sigma_{w'}}\ col$ $(0, 0, \cdots, -\alpha, 1)$ for all $i = 2, 3, \cdots, N$.

Now, in order to calculate $R_c(\infty)$ for the case of $\gamma_i = \gamma$ being constant, recall that $f_i^2 = (1+\gamma)f_{i-1}^2$, therefore

$$f_i = \pm\sqrt{1 + \gamma}\, f_{i-1} \; . \qquad \text{(II.8)}$$

Take $f_i = -\sqrt{1 + \gamma}\; f_{i-1} \operatorname{sgn} \alpha$, then

$$e_i = \gamma \left(\sum_{j=1}^{i} \pm q_{ij}\; f_j \right)^2$$

$$= \sigma_w^{\,2}\gamma \left[\frac{1}{\sqrt{1 - \alpha^2}} \left(\frac{-|\alpha|}{\sqrt{1 + \gamma}} \right)^{i-1} + \sum_{j=i}^{i} \left(\frac{-|\alpha|}{\sqrt{1 + \gamma}} \right)^{i-j} \right]^2$$

$$= \frac{N_o}{2} \frac{\gamma}{\left(1 + \frac{|\alpha|}{\sqrt{1 + \gamma}}\right)^2} \left[1 + \left(\frac{1 + \gamma + |\alpha|}{\sqrt{1 - \alpha^2}} - 1 \right) \left(\frac{-|\alpha|}{\sqrt{1 + \gamma}} \right)^{i-1} \right]^2 ,$$

Therefore

$$\frac{P}{N_o W} = \frac{1}{N} \sum_{i=1}^{N} e_i$$

$$= \frac{\gamma}{\left(1 + \frac{|\alpha|}{\sqrt{1 + \gamma}}\right)^2} + O(\tfrac{1}{N}) \qquad \text{(II.9)}$$

Since $\frac{P}{N_o W} < \gamma$, it follows that $\frac{P}{N_o W} \gg 1$ gives $\gamma \gg 1$, and $\frac{|\alpha|}{1 + \gamma} \ll 1$. This indicates that $\gamma \approx \frac{P}{N_o W}$ so that

$$R_c = W \ln(1+\gamma)$$

$$= W \ln\left(1 + \frac{P}{N_o W} + \epsilon\right) \tag{II.10}$$

where $\epsilon = \left(\gamma - \frac{P}{N_o W}\right) \approx \frac{-2|\alpha|}{\sqrt{1+\gamma}} \approx \frac{-2|\alpha|}{\sqrt{1 + \frac{P}{N_o W}}}$.

However, it also follows from (II.9) that $\gamma < \frac{4P}{N_o W}$, therefore, $\gamma \ll 1$ when $\frac{P}{N_o W} \ll 1$, which shows that $\frac{|\alpha|}{\sqrt{1+\gamma}} \approx |\alpha|$. Thus

$$R_c = W \ln(1+\gamma)$$

$$= W \ln\left[1 + (1+|\alpha|)^2 \frac{P}{N_o W} - \epsilon\right], \tag{II.11}$$

in fact,

$$\lim_{W\to\infty} R_c = (1+|\alpha|)^2 \frac{P}{N_o} \quad . \tag{II.12}$$

Next, consider the case when $e_i = e = \frac{P}{2W}$. This time

$$\gamma_1 = eh_{11}^2 = \frac{P}{N_o W}(1-\alpha^2)$$

while

$$\gamma_i = \frac{P}{N_o W}\left(1-\alpha \frac{g_{i-1}}{g_i}\right)^2$$

$$= \frac{P}{N_o W}\left(1 + \frac{|\alpha|}{\eta}\right)^2 \tag{II.13}$$

where $\eta = -\frac{g_i}{g_{i-1}} \operatorname{sgn} \alpha$. However,

$$1 + \sigma_\theta^{\,2}\|g\|^2_{K^{-1}} = 1 + \sigma_\theta^{\,2}\,\frac{2g_1^{\,2}}{N_o}\left[(1-\alpha^2)+\left(1+\frac{|\alpha|}{\eta}\right)^2 \sum_{i=2}^{N} \eta^{2(i-1)}\right]$$

$$= 1 + \frac{P}{N_oW}\left[(1-\alpha^2) + \left(1+\frac{|\alpha|}{\eta}\right)^2 \frac{(\eta^{2N}-\eta^2)}{\eta^2-1}\right]$$

or

$$\sigma_\theta^{\,2}\|g\|^2_{K^{-1}} \approx \frac{P}{N_oW}\left[\left(1+\frac{|\alpha|}{\eta}\right)^2 (\eta^{2(N-1)}-1)\,\frac{\eta^2}{\eta^2-1}\right] \qquad \text{(II.14)}$$

But this should equal $\prod_{i=1}^{N}(1+\gamma_i) \approx \left[1+\frac{P}{N_oW}\left(1+\frac{|\alpha|}{\eta}\right)^2\right]^{N-1} \times$

$\left(1+\frac{P}{N_oW}(1-\alpha^2)\right)$. Consequently

$$\eta^2 \approx 1 + \frac{P}{N_oW}\left(1+\frac{|\alpha|}{\eta}\right)^2 \qquad \text{(II.15)}$$

thus from (II.13)

$$\eta^2 - 1 \approx \gamma_i = \gamma$$

and now (II.15) implies that

$$\gamma \approx \frac{P}{N_oW}\left(1+\frac{|\alpha|}{\sqrt{1+\gamma}}\right)^2$$

which is the same as (II.9).

REFERENCES

[1] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication", The University of Illinois Press, Urbana, Ill., 1949.

[2] C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel", I.R.E. Trans. on Information Theory, Vol. IT-2, pp. 8-19, September 1956.

[3] S. S. L. Chang, "Theory of Information Feedback Systems," I.R.E. Trans. on Information Theory, Vol. IT-2, pp. 29-40, September 1956.

[4] C. E. Shannon, "Channels with Side Information at the Transmitter", I.B.M. Journal, Vol. 2, pp. 289-293, October 1958.

[5] F. J. Bloom, et.al., "Improvement of Binary Transmission by Null-Zone Reception", Proc. I.R.E. Vol. 45, pp. 963-975, July 1957.

[6] B. Harris, A. Hauptschein and L. S. Schwartz, "Optimum Decision Feedback Systems", I.R.E. Conv. Record, Pt. 2, pp. 3-10, 1957.

[7] B. Harris, A. Hauptschein, K. C. Morgan, and L. S. Schwartz, "Binary Decision Feedback Systems for Maintaining Reliability under Conditions of Varying Field Strength," Proc. Nat. Elect. Conf., Vol. 13, pp. 126-140, 1957.

[8] B. Harris and K. C. Morgan, "Binary Symmetric Decision Feedback Systems", A.I.E.E. Communication and Electronics, No. 38, pp. 436-443, September 1958.

[9] J. J. Metzner, "Binary Relay Communication with Decision Feedback Circuits", I.R.E. Conv. Record, Pt. 4, pp. 112-119, 1959.

[10] P. Elias, "Channel Capacity Without Coding", in Lectures on Communication System Theory, Baghdady, Ed., McGraw-Hill, New York, 1961, pp. 363-366.

[11] P. E. Green, "Feedback Communication Systems", ibid, pp. 345-368.

[12] A. J. Viterbi, "The Effect of Sequential Decision Feedback on Communication over the Gaussian Channel", Information and Control, Vol. 8, pp. 80-92, February 1965.

[13] M. Horstein, "Sequential Transmission Using Noiseless Feedback", I.E.E.E. Trans. on Information Theory, Vol. IT-9, pp. 136-143, July 1963.

[14] G. L. Turin, "Signal Design for Sequential Detection Systems with Feedback", I.E.E.E. Trans. on Information Theory, Vol. IT-11, pp. 401-408, July 1965.

[15] J. P. M. Schalkwijk and T. Kailath, "A Coding Scheme for Additive Noise Channels with Feedback - Part 1: No-Bandwidth Constraint", I.E.E.E. Trans. on Information Theory, Vol. IT-12, pp. 172-182, April 1966.

[16] J. P. M. Schalkwijk, "A Coding Scheme for Additive Noise Channels with Feedback - Part II: Band-Limited Signals", I.E.E.E. Trans. on Information Theory, Vol. IT-12, pp. 183-189, April 1966.

[17] J. K. Omura, "Signal Optimization for Channels with Feedback", Rept. SEL-66-068, Stanford Electronics Labs., Stanford, Calif., August 1966.

[18] R. L. Kashyap, "Sequential Coding Schemes for an Additive Noise Channel with a Noisy Feedback Link", Rept. Harvard University, Cambridge, Mass., June 1966.

[19] M. Horstein, "On the Design of Signals for Sequential and Non-sequential Detection Systems", I.E.E.E. Trans. on Information Theory, Vol. IT-12, pp. 448-455, October 1966.

[20] R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory", J. Basic Eng., Trans. ASME, pp. 95-108, March 1961.

[21] J. M. Wozencraft and I. M. Jacobs, "Principles of Communication Engineering", J. Wiley, New York, 1965.

[22] R. B. Ash, "Information Theory", J. Wiley, New York, 1965, Ch. 8.

[23] R. M. Fano, "Transmission of Information", M.I.T. Press, Cambridge, Mass., 1961.