Conditions for Information Capacity of the Discrete-Time Gaussian Channel to be Increased by Feedback

LlSS 22
December, 1987

Charles R. Baker
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599

*Research Supported by ONR Contract #N00014-86-K-0039.
Sufficient conditions are given for optimal causal feedback to increase information capacity for the discrete-time additive Gaussian channel. The conditions are obtained by assuming linear feedback and reformulating the problem into an equivalent no-feedback problem.

TITLE CONT.: Discrete-Time Gaussian Channel to be Increased by Feedback
Abstract

Sufficient conditions are given for optimum causal feedback to increase information capacity for the discrete-time additive Gaussian channel. The conditions are obtained by assuming linear feedback and reformulating the problem into an equivalent no-feedback problem.
Introduction

Information capacity of the discrete-time additive Gaussian channel with feedback is an open problem. It has long been speculated that causal feedback can increase capacity. We give here sufficient conditions for optimum causal linear feedback to increase information capacity for any fixed value of the constraint, for all values of the constraint, and for all sufficiently large values of the constraint.

A special case of these results is for the finite-dimensional channel with a pure power constraint. The method developed here gives the solution for that case in a particularly easy fashion; see [1].

Recent work on the capacity of feedback channels has been done by Ihara [2] (for the finite-dimensional channel) and by Cover and Pombra [3].

Problem Statement

The capacity problem will be considered for both the infinite-dimensional and the finite-dimensional discrete-time additive Gaussian channel. However, the setup will be given only for $\ell_2$; it will be seen (by substituting $\mathbb{R}^K$ for $\ell_2$) that the procedure also applies without change to $\mathbb{R}^K$, although of course the finite-dimensional channel is much simpler and does not require the full development given here.

All stochastic processes are defined on a probability space $(\Omega, \mathcal{F}, \mu)$; $E(\cdot)$ will denote expectation with respect to $\mu$. $\|x\|$ will denote the $\ell_2$ norm of the vector $x$: $\|x\|_2^2 = \sum_{n \geq 1} [x(n)]^2$.

The channel output is $Y = X - BY + N$, where $N$ is additive zero-mean Gaussian noise with strictly-positive trace-class covariance matrix $R_N$. $X$ is a
message process independent of \( N \), and \( B \) is a Hilbert-Schmidt strictly-lower-triangular (HSSLT) matrix: \( \sum_{i,j \geq 1} b_{ij}^2 < \infty \) and \( b_{ij} = 0 \) for \( j \geq i \), all \( i \geq 1 \). The mutual information of interest is that between \( X \) and \( Y \), denoted \( I(X, Y) \). The constraint will be given in terms of a trace-class covariance matrix \( R_W \). Any constraint must imply a constraint of this form if the capacity is to be finite \([4]\). The class of admissible message processes \( X \) and HSSLT matrices \( B \) consist of all \( X \) such that almost all sample paths of \( X \) belong to \( \text{range}(R_W) \), \( \text{range}(B) \) is contained in \( \text{range}(R_W^{1/4}) \), and \( E\|X-BY\|_W^2 \leq P \), where \( \|\cdot\|_W \). The capacity is then the supremum of the mutual information \( I(X, Y) \) over all such admissible pairs \((X, B)\).

The feedback capacity will be denoted by \( C_{WF}(P) \). The capacity of this channel without feedback is for the case \( B = 0 \), so that the constraint is \( E\|X\|_W^2 \leq P \). This capacity will be denoted by \( C_0(P) \).

The assumptions that \( R_N \) and \( R_W \) are strictly-positive can be dropped. Attention can be restricted to \( \text{range}(R_N) \); in order to have finite capacity, one must then have that \( R_W \) is strictly positive as an operator in \( \text{range}(R_N) \); see \([4]\) for details. However, without loss of generality, it is assumed here that both \( R_W \) and \( R_N \) are strictly positive.

Preliminaries

This section contains several mathematical definitions and small results that will be needed to prove the main result. It will be seen that much of this is obvious when one treats \( \mathbb{R}^K \).

\( \ell_2 \otimes \ell_2 \) will denote the set of all Hilbert-Schmidt operators mapping \( \ell_2 \) into \( \ell_2 \). A is in \( \ell_2 \otimes \ell_2 \) if and only if \( A \) has a matrix representation such that
\[ \sum_{i,j} [A(ij)]^2 < \infty. \]

For \( A_1 \) and \( A_2 \) in \( \ell_2 \otimes \ell_2 \), define
\[
\langle A_1, A_2 \rangle = \text{Trace } A_1 A_2^* = \sum_{i,j} A_1(ij) A_2^*(ji).
\]

\( \langle \cdot, \cdot \rangle \) defines an inner product on \( \ell_2 \otimes \ell_2 \), and it is known that \( \ell_2 \otimes \ell_2 \) is a Hilbert space under this inner product [5]. Moreover, convergence of a sequence \( (A_n) \) in \( \ell_2 \otimes \ell_2 \) to an element \( A \) in \( \ell_2 \otimes \ell_2 \) implies that \( ||A_n - A|| \to 0 \) and thus \( A_n x \to Ax \) for all \( x \) in \( \ell_2 \).

\( (\delta_n) \) will denote the natural basis vectors in \( \ell_2 \): \( \delta_n(i) = 0 \) for \( i \neq n \), \( \delta_n(n) = 1 \). Let \( H_n = \text{span}\{\delta_i, i \leq n\} \) and denote by \( P_n \) the projection operator with range space \( H_n \). \( P_n \) is a diagonal matrix with \( P_n(i,i) = 1 \) for \( i \leq n \):
\[
P_n(i,i) = 0 \text{ for } i > n.
\]

**Lemma 1:** Let \( R^n_W \) be the matrix \( P R^n W P_n \):
\[
R^n_W(ij) = R_W(ij) \text{ for } i \leq n, j \leq n
\]
\[
= 0 \text{ otherwise. Then:}
\]

1. \( R^n_W = V_n V_n^* \) for a lower triangular matrix \( V_n \) with \( V_n(ii) = c_i \) for \( i \leq n \), where \( \Pi c_i^2 = \text{determinant } R^n_W \).
2. \( V_n x = 0 \) for all \( x \) in \( H_n^1 \);
3. If \( m > n \), \( V_n = P_m V_m \);
4. \( (V_n) \) is a Cauchy sequence in \( \ell_2 \otimes \ell_2 \);
5. \( R^n_W = V W^* \), where \( V \) is a unique lower-triangular Hilbert-Schmidt matrix such that \( V(i,i) = c_i \) for all \( i \geq 1 \), and \( V = \lim_n V_n \) in the topology of \( \ell_2 \otimes \ell_2 \). Moreover, \( V_n = P_n V \) for \( n \geq 1 \).

**Proof:** Since \( H_n \) can be identified with \( \mathbb{R}^n \), and \( R^n_W \) with a covariance matrix in \( \mathbb{R}^n \), (1) is obvious, as is (2). To see (3), if \( i,j \leq n \) and \( m > n \), then
\[ R_W^{m}(ij) = R_W^{m}(ij) = \sum_{k\leq n} V_m^{(ik)}V_m^{(jk)} \]

Since \( P_n V_m \) is lower triangular, \( P_n V_m^{(ii)} = V_m^{(ii)} = V_n^{(ii)} \) for all \( i \leq n \), and the factorization \( R_W^n = V n n \) is unique when the diagonal elements of \( V_n \) are fixed [7], it follows that \( V_n = P_n V m \).

For (4), note that \( R_W^n = P R W P_n \). If \( m > n \), then

\[
\text{Trace} (V_n - V_m)(V_n - V_m)^* = \sum_{j=1}^{m} \| (V_n - V_m)^* \delta_j \| ^2 = \sum_{j=m+1}^{m} \| (P_n - I) \delta_j \| ^2 = \sum_{j=m+1}^{m} \| V_n^* \delta_j \| ^2 = \sum_{j=m+1}^{m} \langle P R W P_n \delta_j, \delta_j \rangle. \]

Since \( R_W^n \) has finite trace, this sum converges to zero as \( m, n \to \infty \), showing that \( (V_n) \) is Cauchy in \( \ell_2 \subset \ell_2 \).

To obtain (5), we first recall that convergence in \( \ell_2 \subset \ell_2 \) implies norm convergence to the same limit [5], so there exists by (4) a Hilbert-Schmidt operator \( V \) such that \( V_n \to V \) in both \( \ell_2 \subset \ell_2 \) and operator norm. \( V V_n^* \) must then converge to \( V V_n^* \) in the operator norm topology. However, \( V V_n^* = R_W^n = P R W P_n \), so \( V V_n^* \) converges to \( R_W^n \) in operator norm. Since the set of bounded linear operators on \( \ell_2 \) is a Banach space under the operator norm [6], a Cauchy sequence has a unique limit, and this gives \( R_W = V V_n^* \). \( V \) is necessarily Hilbert-Schmidt, since \( R_W \) is trace-class. To see that \( V \) must be lower-triangular, note that \( \text{Tr} (V_n - V)(V_n - V)^* \to 0 \), and \( \text{Tr} (V_n - V)(V_n - V)^* = \sum (V_n^{(ij)} - V^{(ij)})^2 \). Since \( V_n^{(ij)} = 0 \) for \( j > i \) and all \( n \geq 1 \), it follows that \( V^{(ij)} = 0 \) for \( j > i \geq 1 \). The same relations show that \( V^{(ii)} = c_i \) for all \( i \geq 1 \). Since \( P_n V V_n^* = V V_n^* n n \), we proceed as before to obtain \( V_n = P_n V \).

\[ \square \]

If \( u \) and \( v \) are two vectors in \( \ell_2 \), then \( u \otimes v \) is defined to be the element of \( \ell_2 \otimes \ell_2 \) defined by \( (u \otimes v)x = \langle v, x \rangle u \).

12/3/87 - 4
In order that the capacity without feedback be finite, it is necessary
and sufficient that \( R_N^2 = R_W^2 (I + S) R_W^2 \) where \( S \) is a self-adjoint operator in \( \ell_2 \)
such that \((I + S)^{-1}\) exists and is bounded [4]. The limit points of the spectrum
of \( S \) consist of all real numbers \( \lambda \) such that \( \lambda \) is either an eigenvalue of \( S \) of
infinite multiplicity, or the limit of a sequence of distinct eigenvalues, or
a point of the continuous spectrum (i.e., \((S - \lambda I)^{-1}\) exists and is densely-
defined but not bounded). The set of limit points of \( S \) is not empty. For
discussion of these and related facts, see [6]. \( \theta \) will be used to denote the
smallest limit point of the spectrum of \( S \). As in [4], \( (\lambda_n) \) will always be used
to denote the sequence of eigenvalues of \( S \) that are strictly less than \( \theta \); they
are ordered by \( \lambda_1 \leq \lambda_2 \leq \ldots \leq \theta \), and repeated in the sequence according to
their multiplicity. Of course, there may not be any eigenvalues strictly less
than \( \theta \). If \( \{\lambda_n, n \geq 1\} \) is not empty, then \( \{e_n, n \geq 1\} \) will denote orthonormal
eigenvectors of \( S \) corresponding to the eigenvalues \( (\lambda_n) \): \( Se_n = \lambda_n e_n, n \geq 1 \).

With \( R_W = VW^*, R_W^2 = VL \) for \( L \) a unitary operator [7]. Since \( I + S =
R_W^{-\frac{1}{2}}R_NR_W^{-\frac{1}{2}} \) (on the range of \( R_W^2 \)), \( L^* (I + S) L = I + L^* SL = V^{-1}R_NV^W_1 \). As \( L \) is
unitary, \( L^* SL \) has the same spectrum as \( S \), and so \( V^{-1}R_NV^W_1 \) has the same
spectrum as \( I + S \). Thus, \( I + \theta \) is the smallest limit point of the spectrum of
\( V^{-1}R_NV^W_1 \) and \( \{1 + \lambda_k, k \geq 1\} \) are the eigenvalues of \( V^{-1}R_NV^W_1 \) that are strictly
less than \( 1 + \theta \).

**Main Result**

**Theorem:** Let \( V \) be a lower-triangular matrix such that \( R_W = VW^* \). Fix \( P > 0 \).

(1) For the \( K \)-dimensional channel, let \( \beta_1 \leq \beta_2 \leq \ldots \leq \beta_K \) be the eigenvalues
    of \( S \), with \( J \) the largest integer \( \leq K \) such that \( J \beta_j \leq P + \sum_{i=1}^{J} \beta_i \).
$C_{WF}(P) > C_{W}(P)$ if the set of eigenvectors of $V^{-1}R_NV^{*-1}$ corresponding to the sequence of eigenvalues $(1 + \beta_i, i \leq J)$ does not contain $J$ natural basis vectors.

(2) For the infinite-dimensional channel, $C_{WF}(P) > C_{W}(P)$ if the following conditions are satisfied:

(a) \{\lambda_k, k \geq 1\} is not empty;

(b) If there exists a largest integer $J$ such that $J \lambda_j \leq P + \sum_{i=1}^{J} \lambda_i$, then

or (b') If $J \lambda_j \leq P + \sum_{i=1}^{J} \lambda_i$ for all $\lambda_j$, then $C_{WF}(P) > C_{W}(P)$ if the subspace spanned by the eigenvectors of $V^{-1}R_NV^{*-1}$ corresponding to the sequence of eigenvalues $(1 + \lambda_k, k \leq J)$ does not contain $J$ natural basis vectors.

Corollary:

(1) For the finite-dimensional discrete-time channel, causal linear feedback can increase information capacity for all $P > 0$ if the subspace spanned by the eigenvectors of $V^{-1}R_NV^{*-1}$ corresponding to its smallest eigenvalue does not contain a basis consisting entirely of natural basis vectors which are eigenvectors of $V^{-1}R_NV^{*-1}$. Causal linear feedback can increase capacity for all sufficiently large $P$ if and only if $V^{-1}R_NV^{*-1}$ is not diagonal.

(2) For the infinite-dimensional discrete-time channel, causal linear feedback can increase capacity for all $P > 0$ if \{\lambda_k, k \geq 1\} is not empty.
and the subspace spanned by the eigenvectors of $V^{-1}R_NV^{k-1}$ corresponding to its smallest eigenvalue does not contain a basis consisting entirely of natural basis vectors which are eigenvectors of $V^{-1}R_NV^{k-1}$.

$C_{WF}(P) > C_{W}(P)$ for all sufficiently large $P$ if $\{\lambda_k, k \geq 1\}$ is not empty and the subspace spanned by the eigenvectors corresponding to the eigenvalues $\{1 + \lambda_k, k \geq 1\}$ of $V^{-1}R_NV^{k-1}$ does not contain a basis for the subspace consisting entirely of natural basis vectors which are eigenvectors of $V^{-1}R_NV^{k-1}$.

Remark: The sufficient condition in (2) giving $C_{WF}(P) > C_{W}(P)$ for all sufficiently large $P$ is equivalent to the following statement: $\{\lambda_k, k \geq 1\}$ is not empty, and the restriction of $V^{-1}R_NV^{k-1}$ to the subspace spanned by the eigenvectors of $V^{-1}R_NV^{k-1}$ corresponding to the eigenvalues $\{1 + \lambda_k, k \geq 1\}$ is not a diagonal matrix.

The results stated in (1) of the Corollary were proved in [1], where the development is much streamlined because of the simpler nature of the finite-dimensional problem. That work used $R_w = I$. For the same finite-dimensional channel and constraint, Ihara has obtained the result that capacity is increased for all sufficiently large $P$ if $R_N$ is not diagonal [2], although his result is stated in a different form; his methods are quite different from those used here. He also gives as a sufficient condition for $C_{WF}(P) > C_{W}(P)$ for all $P > 0$ the condition that (in the terminology used here) $R_N$ has no natural basis vectors as eigenvectors. The corresponding sufficient condition given in (1) of the Corollary is much weaker.
Reformulation of the Problem

In this section, the original linear feedback problem is converted into an equivalent no-feedback problem. Originally, \( Y = X - BY + N \), where the matrix \( B \) is HSSLT. \((I+B)^{-1}\) exists, since \( B \) can have no non-zero eigenvalues; \((I+B)^{-1}\) is bounded, since \( B \) is compact (and thus has only zero as a limit point of the spectrum). Thus, \( Y = (I+B)^{-1}X + (I+B)^{-1}N \). Since \((I+B)^{-1}\) is 1:1,

\[
I(X, X+N) = I(X, (I+B)^{-1}(X+N)) = I(X, Y).
\]

Of course, the constraint \( \|X-BY\|_{w}^2 \leq P \) is the same as \( \|X-(I+B)^{-1}(X+N)\|_{w}^2 \leq P \).

Using \( R_{w} = VV^{*} \) with \( V \) Hilbert-Schmidt and lower triangular, the constraint can be written

\[
P \geq E\|V^{-1}X - V^{-1}B(I+B)^{-1}(X+N)\|_{w}^2
\]

or,

\[
P \geq E\|Z - D(V^{-1}Z+N)\|_{w}^2,
\]

where \( D = V^{-1}B(I+B)^{-1} \) and \( Z = V^{-1}X \). \( D \) is well-defined and bounded, since \( \text{range}(B) \subseteq \text{range}(V) \) and \((I+B)^{-1}\) is bounded. Moreover, since \( B \) is HSSLT and both \( V^{-1} \) and \((I+B)^{-1}\) are lower triangular, \( D \) must be strictly lower triangular and bounded (BSLT).

The feedback capacity problem under our initial assumptions thus becomes

maximize \( I(X, X+N) \)

subject to \( P \geq E\|V^{-1}X - D(X+N)\|_{w}^2 \)

where \( D \) is permitted to be any bounded SLT matrix in \( \ell_{2} \).

This is actually the problem that will be considered below in obtaining the sufficient conditions of the Theorem and the Corollary.

12/3/87 - 8
Analysis

Let \( H(\ell^2_2, \mu) \) be the set of all real random vectors \( u \) on \((\Omega, \beta)\) such that \( u(\omega) \in \ell^2_2 \) a.e. \( dP(\omega) \) and \( E \sum \left[ u(n, \omega) \right]^2 < \infty \). \( H(\ell^2_2, \mu) \) is a Hilbert space under the inner product \( \langle f, g \rangle_\mu = E \sum f(n, \omega)g(n, \omega) \). Let \( Y_1 \) and \( Y_2 \) be two mutually independent zero-mean Gaussian random vectors in \( \ell^2_2 \): \( E Y_1(n, \omega) Y_2(m, \omega) = 0 \) for all \( n, m \geq 1 \). Suppose that \( R_{Y_1} + R_{Y_2} \) is strictly positive. Define \( H_-(Y_1 + Y_2) \) as the set of all elements \( f \) in \( H(\ell^2_2, \mu) \) such that \( f = B(Y_1 + Y_2) \) for some bounded SLT matrix operator \( B \). \( H_-(Y_1 + Y_2) \) is clearly a linear manifold in \( H(\ell^2_2, \mu) \). To see that this linear manifold is closed, one notes that if \( (B^n) \) is a sequence of bounded SLT operators.

\[
\|B^n(Y_1 + Y_2) - B^m(Y_1 + Y_2)\|_\mu^2 = \|B^n - B^m\|^2 \|Y_1 + Y_2\|^2 = \text{Trace} (B^n - B^m)^*(R_{Y_1} + R_{Y_2})(B^n - B^m) \geq \|B^n - B^m\|^2 \tau_0
\]

where \( \tau_0 \) is the smallest eigenvalue of \( R_{Y_1} + R_{Y_2} \). Thus, \( (B^n(Y_1 + Y_2)) \) Cauchy in \( H(\ell^2_2, \mu) \) implies that \( (B^n) \) is Cauchy in operator norm, so converges to a bounded linear operator \( B \). To see that \( B \) is SLT, one notes that

\[
R_{Y_1} + R_{Y_2} = Q^*Q
\]

for some lower-triangular \( Q \), and so \( (R_{Y_1} + R_{Y_2})^{\frac{1}{2}} = QT \) for \( T \) unitary [8]. This gives

\[
\|B^n(Y_1 + Y_2) - B^m(Y_1 + Y_2)\|_\mu^2 = \|B^n - B^m\|^2 \|R_{Y_1} + R_{Y_2}\|^2 \|Q\|^2.
\]

Thus, \( (B^nQ) \) is Cauchy in \( \ell^2_2 \Theta \ell^2_2 \), so that \( BQ \) must be strictly lower triangular.

12/3/87 - 9
since $Q^{-1}$ exists and is lower triangular, this shows that $B$ is a bounded SLT matrix operator, so that $H_{-(Y_1+Y_2)}$ is closed.

Now consider our feedback problem: We wish to maximize $I(X, X+N)$ subject to $P \geq E\|V^{-1}X - D(X+N)\|^2 = E\|V^{-1}X - D(X+N)\|^2_{\mu}$, where $D$ is permitted to be any bounded SLT matrix. Given any choice of $D$ that satisfies this constraint, we know that $E\|V^{-1}X - D(X+N)\|^2_{\mu} \geq E\|V^{-1}X - P_-(V^{-1}X)\|^2_{\mu}$, where $P_-(V^{-1}X)$ is the projection of $V^{-1}X$ onto $H_{-(X+N)}$. Thus, we can assume WLOG that $D$ is the optimum bounded SLT matrix for minimizing the distance in $H(\ell_2, \mu)$ norm between $V^{-1}X$ and $H_{-(X+N)}$: $D(X+N)$ is the projection of $V^{-1}X$ onto $H_{-(X+N)}$.

Now, let $X$ be the optimum no-feedback message for the case when capacity is attained (assuming here that $C_{\ell_e}(P)$ can be attained). As the message for the feedback problem, use $aX$. Then $I(aX, aX+N) > C_{\ell_e}(P)$ if $a > 1$. Choose $a$ to satisfy the constraint:

$$P = E\|V^{-1}aX - D(aX+N)\|^2_{\mu}$$

$$= a^2 \text{Tr} V^{-1}R_XV^{-1} - \Delta,$$

where $\Delta = \Delta(a)$ is the $H(\ell_2, \mu)$ norm of the projection of $aV^{-1}X$ on $H_{-(aX+N)}$. Since $\text{Tr} V^{-1}R_XV^{-1} = \text{Tr} R_W^{-1}R_XR_W^{-1} = E\|X\|^2_{\mu} = P$, we have $P = a^2 P - \Delta$, so that $a > 1$ if $aV^{-1}X$ is not orthogonal to $H_{-(aX+N)}$.

Prop.: $aV^{-1}X$ is orthogonal to $H_{-(aX+N)}$ if and only if $V^{-1}R_XV^{*-1}$ is diagonal.

Proof: Since $X$ is independent of $N$, $(aV^{-1}X, D(aX+N))_{\mu} = a^2 \text{Tr} D_XV^{*-1} = a^2 \text{Tr} DV[V^{-1}R_XV^{*-1}].$ If $V^{-1}R_XV^{*-1}$ is diagonal, then (as $DV$ is SLT)

$$\text{Tr} D_XV^{*-1} = 0$$

for every bounded SLT matrix $D$.

If $\text{Tr} D_XV^{*-1} = 0$ for every bounded SLT matrix $D$, take $i, j$ with $i > j$. 

12/3/87 - 10
Let $D(k\ell) = 0$ unless $k = i, \ell = j$:

$$D(ij) = 1.$$  

Then, $	ext{Tr} \, D(V^{-1}R_XV^{*-1})(ji) = 0$. This shows that $R_XV^{*-1}$ must be lower triangular, so that $V^{-1}R_XV^{*-1}$ must also be lower triangular. Since $V^{-1}R_XV^{*-1}$ is symmetric, $V^{-1}R_XV^{*-1}$ must be diagonal.

In the above development, we have implicitly assumed that $\alpha$ always exists to solve the equation $\alpha^2P = P + \Delta(\alpha)$. This is not obvious, as the subspace $H_\alpha(\alpha X+N)$ changes with $\alpha$. Here we will show that a lower bound $\alpha_1$ exists; i.e., $\alpha_1 > 1$ and a bounded SLT $D$ such that $\|a_1V^{-1}X - D(a_1X+N)\|^2_\mu = P$.

Let $D$ be the optimum SLT matrix to minimize $\|V^{-1}X - D(X+N)\|^2_\mu$. Then

$$\text{Tr} \, DR_XV^{*-1} = \text{Tr} \, DR_XD^* + \text{Tr} \, DR_ED^* = \Delta(1).$$

Take $\alpha \neq 0$. Then $\|aV^{-1}X - D(\alpha X+N)\|^2_\mu = a^2(P - 2\text{Tr} \, DR_XV^{*-1} + \text{Tr} \, DR_XD^*) + \text{Tr} \, DR_ED^*$. If $P - 2\text{Tr} \, DR_XV^{*-1} + \text{Tr} \, DR_XD^* \neq 0$, then one can set $P = \|aV^{-1}X - D(\alpha X+N)\|^2_\mu$ and solve for $\alpha^2$, obtaining

$$\alpha^2 = \frac{P - \text{Tr} \, DR_ED^*}{P - \text{Tr} \, DR_ED^* - \Delta(1)}, \text{ giving } \alpha > 1.$$  

To see that $P - 2\text{Tr} \, DR_XV^{*-1} + \text{Tr} \, DR_ED^* \neq 0$ when $V^{-1}R_XV^{*-1}$ is not diagonal, we note that if inequality does not hold, then $\|V^{-1}X - D(X+N)\|^2_\mu = P - \Delta(1) = \text{Tr} \, DR_ED^*$. Similarly, for any $\alpha \neq 0$, $\|aV^{-1}X - D(\alpha X+N)\|^2_\mu = a^2(P - 2\text{Tr} \, DR_XV^{*-1} + \text{Tr} \, DR_XD^*) + \text{Tr} \, DR_ED^* = \text{Tr} \, DR_ED^*$. Thus, $P - \Delta(\alpha) \leq \text{Tr} \, DR_ED^* = P - \Delta(1)$, or $\Delta(\alpha) \geq \Delta(1)$, all $\alpha \neq 0$. This cannot hold if $V^{-1}R_XV^{*-1}$ is not diagonal, since $\Delta(\alpha) \leq \alpha^2P$, and $\Delta(1) \neq 0$ when $V^{-1}R_XV^{*-1}$ is not diagonal.

We have now shown that causal linear feedback can increase capacity provided that $V^{-1}R_XV^{*-1}$ is not diagonal, where $R_X$ is the optimum message covariance matrix in the no-feedback problem (whenever capacity can be attained in the no-feedback case). These conditions need to be converted into conditions on the noise covariance matrix $R_N$ and the constraint matrix $R_w$.

12/3/87 - 11
This will be done in the next two sections, treating the $K^*$ and $\ell_2$ channels.

Finite-Dimensional Channel ($K^*$)

From Theorem 1 of [4], the optimum no-feedback message has covariance matrix given by

$$R_X = \frac{1}{J} \left[ \sum_{i=1}^{J} \beta_i^* + P \right] \sum_{n=1}^{J} R_{w_n} \otimes \frac{1}{2} R_{u_n} - \sum_{m=1}^{J} \beta_m R_{w_n} \otimes \frac{1}{2} R_{u_n},$$

where $\{u_n, \ n \leq K\}$ are o.n. eigenvectors of $S$ corresponding to the increasing sequence of eigenvalues $\{\beta_n\}$, and $J$ is the largest integer $\leq K$ such that

$$P + \sum_{i=1}^{J} \beta_i^* \geq J \beta J.$$ 

Let $L$ be the unitary operator in $\ell_2$ such that $R_{w_n} = V L_{n}^* U_n$, then

$$V^{-1} R_N V_{**} = \frac{1}{J} \sum_{n=1}^{J} \left[ \sum_{i=1}^{J} \beta_i + P - \beta_n \right] (L_{n}^* U_n) \otimes (L_{n}^* U_n).$$

Now, $V^{-1} R_N V_{**} = L^* (I+S)L$, and $S_{n} = \beta_n U_n$, so

$$L^* (I+S)L U_n = L^* (I+S) U_n = (1+\beta_n) L_{n}^* U_n,$$

i.e., $\{L_{n}^* U_n, \ n \leq K\}$ are c.o.n. eigenvectors of $V^{-1} R_N V_{**}$ corresponding to the sequence of eigenvalues $\{1+\beta_n\}$. $V^{-1} R_X V_{**}$ is then diagonal if and only if $\{L_{n}^* U_n, \ n \leq J\}$ can be taken as natural basis vectors, proving (1) of the Theorem. For all sufficiently small $P > 0$,

$$V^{-1} R_X V_{**} = \frac{P}{M} \sum_{n=1}^{M} L_{n}^* U_n \otimes L_{n}^* U_n,$$

where $M$ is the multiplicity of the eigenvalue $\beta_1$ of $S$, and of the eigenvalue $1 + \beta_1$ of $V^{-1} R_N V_{**}$. Thus, $V^{-1} R_X V_{**}$ cannot be diagonal if $\{L_{n}^* U_n, \ n \leq M\}$ cannot be taken as natural basis vectors. For larger values of $P$, when $R_X$ has the representation given above for $J > M$, then the eigenvectors of $V^{-1} R_X V_{**}$ must include $\{L_{n}^* U_n, \ n \leq M\}$. Now, if $V^{-1} R_X V_{**}$ is diagonal, then it must have a
c.o.n. set of eigenvectors consisting of natural basis vectors. However, 
\( \text{span}\{L_u, n \leq M\} \) cannot be spanned by \( M \) natural basis vectors, so that 
\( V^{-1}R_{X,N}^{M-1} \) cannot have a c.o.n. set of eigenvectors consisting entirely of 
natural basis vectors. This shows that \( C_{WF}(P) > C_W(P) \) for every \( P > 0 \) if the 
\( M \)-dimensional eigenmanifold of \( \beta_1 \) is not spanned by \( M \) natural basis vectors.

By letting \( P \) become sufficiently large, \( J = K \), and then the above 
expressions show that \( V^{-1}R_{X,N}^{M-1} \) will be diagonal if and only if \( V^{-1}R_{N}^{M-1} \) is 
diagonal: \( V^{-1}R_{X,N}^{M-1} = \frac{1}{K} \left( \sum_{i=1}^{K} \beta_i + P + K \right) I - V^{-1}R_{N}^{M-1} \). This proves the 
sufficient conditions of the Corollary.

To see that capacity cannot be increased by causal feedback if \( V^{-1}R_{N}^{M-1} \) 
is diagonal, one notes that the feedback capacity problem is that of 
maximizing \( I(X, X+N) \) subject to the constraint 
\( \text{EI}_{I}V^{-1}D(X,N)_{2} \leq P \), where 
\( \|x\|_n^2 = \sum_{i=1}^{n} x_i^2 \) and \( D \) is a possibly non-linear operator depending only on the 
past of the second coordinate (causal): 
\( [D(x,y)]_n = D_{n}^1(x, [y_1, y_2, \ldots, y_{n-1}]) \), 
where \( D_{n}^1 \) maps \( \mathbb{R}^n \times \mathbb{R}^{n-1} \) into \( \mathbb{R} \). Write the constraint as 
\( \text{EI}_{I}V^{-1}F(T,Z)_{2} \leq P \), where 
\( T = V^{-1}X, Z = V^{-1}N, \) and \( F(x,y) = D[Vx, Vy] \). Since \( V \) is lower-triangular and \( D \) 
is causal in the second coordinate, \( F \) is also causal in the second coordinate. 
\( I(X,Y) = I(V^{-1}X, V^{-1}Y) = I[V^{-1}X, V^{-1}D(X,N) + V^{-1}N] = I[T, V^{-1}F(T,Z) + Z] \). The 
constraint is \( \text{EI}_{I}V^{-1}F(T,Z)_{2} \leq P \), and \( V^{-1}F(x,y) \) is a causal function of \( y \).
However, \( Z \) has covariance matrix \( V^{-1}R_{N}^{M-1} \). Thus, if \( V^{-1}R_{N}^{M-1} \) is diagonal, 
the original problem is equivalent to the capacity problem with causal 
feedback when the channel is without memory. It is well-known that capacity 
cannot be increased in this case.

This completes the proof of the theorem and corollary for the finite-
dimensional channel.
Infinite-Dimensional Channel ($\ell_2$)

First, assume that \(\{\lambda_n, n \geq 1\}\) is not empty. Several cases need to be considered. The various expressions for the optimum \(R_X\) (when it exists) and the value of \(C_w(P)\) are taken from [4].

1. \(\Sigma (\theta - \lambda_n) < \infty.\)

If \(P < \Sigma (\theta - \lambda_n)\), then there exists finite \(J\) such that the optimum no-feedback covariance is given by [4, Theorem 3].

\[
R_X = \frac{1}{J} \left[ \Sigma_{i=1}^{J} \lambda_i + P \right] \Sigma_{i=1}^{J} \lambda_i e_n \otimes e_n - \frac{1}{J} \left[ \Sigma_{i=1}^{J} \lambda_i e_n \otimes e_n \right] \frac{1}{J} \left[ \Sigma_{i=1}^{J} \lambda_i e_n \otimes e_n \right]
\]

(\(\ast\))

As in the finite-dimensional channel, this shows that \(VR_XV^{-1}\) will not be diagonal if the subspace spanned by the eigenvectors of \(V^{-1}R_NV^{-1}\) corresponding to the sequence of eigenvalues \((1 + \lambda_k, k \leq J)\) of \(V^{-1}R_NV^{-1}\) does not contain a basis consisting entirely of natural basis vectors which are eigenvectors of \(V^{-1}R_NV^{-1}\).

If \(P = \Sigma (\theta - \lambda_n)\), then \(J \lambda_j \leq P + \Sigma_{i=1}^{J} \lambda_i\) for every \(\lambda_j\) [4], and an optimum message covariance exists and is given by

\[
R_X = \Sigma_{n=1}^{\infty} (\theta - \lambda_n) e_n \otimes e_n.
\]

This gives

\[
V^{-1}R_XV^{-1} = \Sigma_{n=1}^{\infty} (\theta - \lambda_n) (L^* e_n) \otimes (L^* e_n),
\]

which is clearly diagonal if and only if \(\{L^* e_n, n \geq 1\}\) can be taken as natural basis vectors.

If \(\{\lambda_n\}\) is an infinite sequence and \(P > \Sigma_n (\theta - \lambda_n)\), then capacity cannot be attained in the no-feedback case. However, the capacity is given by \(\lim_{K \to \infty} I_{W}^K(P)\).

12/3/87 - 14
where $I^K_w(P)$ is the value of $I(X^K, X^{K+N})$ when
\[
R^K_X = \frac{1}{K} \left( \sum_{i=1}^{K} \lambda_i + P \right) \sum_{n=1}^{K} \frac{1}{w_n} \Theta^2 e_n - \frac{1}{w_n} \left( \sum_{i=1}^{K} \lambda_i \Theta e_n \right) R^2_w e_n
\]

Let $\Delta^K_K(1)$ be the squared $H(\ell_2, \mu)$ norm of the projection of $V^{-1}X^K$ onto $H_0(X^{K+N})$. If $\limsup \Delta^K_K(1) > 0$, then as before the capacity can be increased. That is, we choose $K$ sufficiently large so that $I(\alpha X^K, \alpha X^{K+N}) > C_w(P)$, where $\alpha > 1$, $\alpha \geq \alpha^K(1)$, with $\alpha^K(1)$ satisfying
\[
\alpha^K(1) = \frac{P - \text{Tr} D_K R_N D_K^T}{P - \text{Tr} D_K R_N D_K^T - \Delta^K_K(1)}
\]
the bounded SLT matrix that minimizes $E\|V^{-1}X^K - D_K(X^{K+N})\|^2$. The problem is now reduced to showing that $\Delta^K_K(1) \to 0$ cannot hold if there exists some $J$ such that \{${\mathbf l}^*_e n, n \leq J$\} cannot be taken to be natural basis vectors. Suppose such $J$ exists and take $K > J$.

Write $X^K = X^K_{\mathbf l} + X^K_{\mathbf o}$, where $X^K_{\mathbf o}$ is the zero-mean Gaussian process with covariance matrix $R^K_{X^K}$.

\[
R^K_{X^K} = \frac{1}{K^2} \sum_{n=1}^{J} \left( \sum_{i=1}^{K} \lambda_i + P \right) \frac{1}{w_n} \Theta^2 e_n - \frac{1}{w_n} \left( \sum_{i=1}^{K} \lambda_i \Theta e_n \right) R^2_w e_n.
\]

and $X^K_{\mathbf o}$ is independent of $X^K_{\mathbf l}$. As $K \to \infty$, $R^K_{X^K}$ converges in the operator norm topology to $\sum_{n=1}^{J} (\theta^* - \lambda_n) R^2_w e_n \Theta^2 e_n$, using the fact that $\sum_{n=1}^{J} (\theta^* - \lambda_n) < P$. Now suppose $\sum_{n \geq 1} \lambda_n - P > 0$ that $\Delta^K_K(1) \to 0$. This requires that $E[V^{-1}X_{\mathbf o}^J + V^{-1}X_{\mathbf o}^J, B(X_{\mathbf o}^J + X_{\mathbf o}^J + N)] \to 0$ for every fixed bounded SLT matrix $B$. Since $X_{\mathbf o}^J$ and $X_{\mathbf o}^J$ are independent, this implies that $E[V^{-1}X_{\mathbf o}^J, BX_{\mathbf o}^J] \to 0$ for every bounded SLT $B$, or

{\text{Trace $B R^K_{X^K} V^{-1}$}} \to 0. Since $R^K_{X^K} \to J \sum_{n=1}^{J} (\theta^* - \lambda_n) R^2_w e_n \Theta^2 e_n$ as $K \to \infty$, this implies

\[
\sum_{n=1}^{J} (\theta^* - \lambda_n) L^*_e e_n \Theta^*_e e_n
\]

(as in the proof of the Proposition) that $\sum_{n=1}^{J} (\theta^* - \lambda_n) L^*_e e_n \Theta^*_e e_n$ is diagonal. This cannot be, since by assumption \{${\mathbf l}^*_e n, n \leq J$\} cannot be taken as natural basis
vectors. This shows that optimum feedback will increase capacity when

\[ P > \sum_{n=1}^{\infty} (\theta - \lambda_n) \text{ and } (\lambda_n) \text{ is an infinite sequence.} \]

Finally, suppose that \((\lambda_n)\) is a finite sequence, \(\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_K\), and

\[ P > \sum_{n=1}^{K} (\theta - \lambda_n). \] Then there exists an infinite o.n. set \(\{u_n, n \geq 1\}\) such that

\[ \|(S-\Theta)u_n\| \to 0 \text{ and } u_n \perp \text{span}\{e_1, \ldots, e_K\} \text{ for all } n \geq 1 [6]. \]

Fix \(M < \infty\) and take \(\epsilon > 0\) such that \(\theta - \lambda_K > \epsilon\).

Let \(X_1^M\) be the zero-mean Gaussian process with covariance matrix \(R_1^M\) given by

\[
R_1^M = \sum_{n=1}^{K} \frac{\sum_{i=1}^{n} \lambda_i + p^M_1 - K\lambda_n}{nK(1+\lambda_n)} \left[ \begin{array}{c} \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \Phi \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \epsilon_i \end{array} \right].
\]

where \(P^M_1 < P\).

Choose o.n. vectors \(u_1^M, \ldots, u_2^M\) from the set \(\{u_n, n \geq 1\}\) such that

\[ |\langle (S-\Theta)u_i^M, u_1^M \rangle| \leq \epsilon \text{ for } i \leq M. \]

Let \(X_2^M\) be the zero-mean Gaussian process with covariance matrix \(R_2^M\) given by

\[
R_2^M = \frac{P - P^M_1}{M(1+\theta+\epsilon)} \sum_{i=1}^{M} R_1^M u_i^M \Phi R_1^M u_i^M.
\]

Now, let \(X^M, e\) be the zero-mean Gaussian process with covariance matrix \(R_X^M\), where \(R_X^M = R_1^M + R_2^M e\). Since \(u_1^M\) is orthogonal to \(\text{span}\{e_1, \ldots, e_K\}\) for \(i \leq M\),

\[
I(X^M, e, X^M, e+N) = I(X_1^M, X_1^M+N) + I(X_2^M, X_2^M, e+N)
\]

\[
= \frac{1}{2} \sum_{n=1}^{K} \log \left[ 1 + \frac{\sum_{i=1}^{n} \lambda_i + p^M_1 - K\lambda_n}{nK(1+\lambda_n)} \right] + \frac{1}{2} M \log \left[ 1 + \frac{P - P^M_1}{M(1+\theta+\epsilon)} \right].
\]

\(X^M, e\) satisfies the constraint for any \(P^M_1 < P\) since
\[ E[W^M, \epsilon_2] = \text{Trace} R_{W^M}^\frac{1}{2} R_{W^M} \epsilon_2 R_{W^M}^\frac{1}{2} = \text{Trace} R_{W}^\frac{1}{2} (R_{W^1}^M R_{W^2}) R_{W}^\frac{1}{2} \]

\[ \lesssim P_1^M + \frac{P - P_1^M}{M(1 + \theta + \epsilon)} M(1 + \theta + \epsilon) = P. \]

Now, define \( P_1^M \) by

\[ P_1^M = M^{-1} \left[ KP - (M-K) \sum_{i=1}^{K} \lambda_i \right]. \]

Then

\[ I(X_1^M, X_1^{M+N}) = \frac{1}{2} \sum_{n=1}^{K} \log \left[ \frac{MK + K \sum_{i=1}^{K} \lambda_i + KP + (M-K)K\theta}{MK(1+\lambda_n)} \right]. \]

As \( M \to \infty \),

\[ I(X_1^M, X_1^{M+N}) \to \frac{1}{2} \sum_{n=1}^{K} \log \left[ \frac{1 + \theta}{1 + \lambda_n} \right]. \]

Similarly,

\[ I(X_2^M, X_2^{M,N}) = \frac{1}{2} M \log \left[ \frac{M^2(1+\theta+\epsilon) + (M-K)[P + K \sum_{i=1}^{K} \lambda_i - K\theta]}{M^2(1+\theta+\epsilon)} \right]. \]

As \( M \to \infty \),

\[ I(X_2^M, X_2^{M,N}) \to \frac{P + \sum_{i=1}^{K} (\lambda_i - \theta)}{2(1+\theta+\epsilon)} \]

Thus, \( I(X_1^M, X_1^{M,N}) \) converges, as \( M \to \infty \), to

\[ \frac{1}{2} \sum_{n=1}^{K} \log \left[ \frac{1 + \theta}{1 + \lambda_n} \right] + \frac{1}{2} \frac{P + \sum_{i=1}^{K} (\lambda_i - \theta)}{1 + \theta + \epsilon}. \]

From Theorem 3 of [4], the capacity \( C_w(P) \) is equal to

\[ \frac{1}{2} \sum_{n=1}^{K} \log \left[ \frac{1 + \theta}{1 + \lambda_n} \right] + \frac{1}{2} \frac{P + \sum_{i=1}^{K} (\lambda_i - \theta)}{1 + \theta}. \]

12/3/87 - 17
so that by choosing \( \varepsilon \) sufficiently small and \( M \) sufficiently large, one can obtain \( I(X^M, \varepsilon, X^{M, \varepsilon} + N) \) arbitrarily near \( C_W(P) \).

As \( \varepsilon \to 0 \) and \( M \to \infty \), \( V^{-1}R_1^{M, \varepsilon}V^{M-1} \) converges to a diagonal matrix, since

\[
|\langle (S-I)u_i^{M, \varepsilon}, u_j^{M, \varepsilon} \rangle| \leq \varepsilon \text{ for } i \leq M. \text{ However,}
\]

\[
V^{-1}R_1^{M, \varepsilon}V^{M-1} = \sum_{n=1}^{K} \left[ (\theta - \lambda_n) + \left( P + \sum_{j=1}^{K} \lambda_j - K\theta \right) \right] e_n \otimes e_n.
\]

This matrix will not be diagonal if \( \{L^* e_n, n \leq K\} \) cannot be taken to consist entirely of natural basis vectors. This is equivalent to not having \( K \) natural basis vectors as eigenvectors of \( V^{-1}R_1^{M, \varepsilon}V^{M-1} \) corresponding to the sequence of eigenvalues \( (1 + \lambda_k, k \leq K) \). Inserting the above definition of \( P_1^M \),

\[
V^{-1}R_1^{M, \varepsilon}V^{M-1} = \sum_{n=1}^{K} \left[ (\theta - \lambda_n) + \left( P + \sum_{j=1}^{K} \lambda_j - K\theta \right) \right] e_n \otimes e_n.
\]

This matrix is independent of \( \varepsilon \); as \( M \to \infty \), it converges in the operator norm topology to

\[
V^{-1}R_1^{M, \varepsilon}V^{M-1} = \sum_{n=1}^{K} (\theta - \lambda_n) e_n \otimes e_n.
\]

Similar to the preceding part of the proof, we now consider \( A_{M, \varepsilon}(1) \), the squared \( H(\theta_2, \mu) \) norm of the projection of \( V^{-1}X^M, \varepsilon \)
onto \( H_-(X^M, \varepsilon + N) \). \( A_{M, \varepsilon}(1) \to 0 \) as \( M \to \infty \) implies, as in the preceding part of the proof, that \( V^{-1}R_1^{M, \varepsilon}V^{M-1} \) is diagonal. This is a contradiction. In fact, \( A_{M, \varepsilon}(1) \)

is bounded away from zero. Define

\[
\alpha_{M, \varepsilon}^2 = \frac{P - \text{Tr} D_{M, \varepsilon} R_1^{M, \varepsilon} D_{M, \varepsilon} - A_{M, \varepsilon}(1)}{P - \text{Tr} D_{M, \varepsilon} R_1^{M, \varepsilon} - D_{M, \varepsilon}(1)}
\]

with \( D_{M, \varepsilon}(X^M, \varepsilon + N) \) the projection of \( V^{-1}X^M, \varepsilon \) onto \( H_-(X^M, \varepsilon + N) \). Since

\[
I(X^M, \varepsilon, X^M, \varepsilon + N) \to C_W(P) \text{ as } \varepsilon \to 0, \ M \to \infty, \ \text{and } A_{M, \varepsilon}(1) \text{ is bounded away from zero, we obtain } I(\alpha_{M, \varepsilon} X^M, \varepsilon, \alpha_{M, \varepsilon}, X^M, \varepsilon + N) \to C_W(P). \text{ This completes the proof of}
\]
sufficiency in part (2) of the Theorem when \( \Sigma_{n=1}^{K}(\theta - \lambda_n) < \infty \).

(2) \( \Sigma_{n=1}^{K}(\theta - \lambda_n) = \infty \).

In this case, \( P < \Sigma_{n=1}^{K}(\theta - \lambda_n) \) for all \( P > 0 \), capacity is attained in the no-feedback case for every \( P > 0 \), and for each \( P > 0 \) there exists \( J < \infty \) (the
value of $J$ depending on $P$) such that the optimum message covariance matrix $R_X$ is given as in ($\ast$). As in case (1), it is clear that feedback can increase capacity if the set of eigenvectors corresponding to the sequence of eigenvalues $(1 + \lambda_k, k < J)$ of $V^{-1}_N V^{\ast -1}$ does not contain $J$ natural basis vectors.

This completes the proof of (2) of the Theorem. The proof of (2) of the Corollary follows from (2) of the Theorem, in the same way that (1) of the Corollary was obtained.

Verification of the Sufficient Conditions

Verification of the sufficient conditions given in the Theorem is equivalent to determining the value of $C_w(P)$, as can be seen from the expressions for $C_w(P)$ [4]. The difficulty of verifying the sufficient conditions of the Corollary is considerably less than for the Theorem. We now summarize how one can verify that $C_{WF}(P) > C_w(P)$ for all $P > 0$. This will be done by giving conditions that are equivalent to the conditions given in (1) and (2) of the Corollary for $C_{WF}(P) > C_w(P)$ for all $P > 0$.

Suppose that $V^{-1}_N V^{\ast -1}$ is nondiagonal. Write

$$V^{-1}_N V^{\ast -1} = A - D$$

where $D$ is a diagonal matrix whose non-zero elements $D(i,i)$ are the diagonal elements $\gamma_{ii}$ of $V^{-1}_N V^{\ast -1}$ such that $(V^{-1}_N V^{\ast -1})(ij) = (V^{-1}_N V^{\ast -1})(ji) = 0$ for all $j \neq i$. $C_{WF}(P) > C_w(P)$ for all $P > 0$ if the following conditions are satisfied.

1. Finite-Dimensional Channel.

$$\inf \langle Ax, x \rangle \leq \inf \{D(i,i) : D(i,i) > 0\};$$

$\|x\| = 1$
2. Infinite-Dimensional Channel.

(a) \( \inf \langle Ax, x \rangle \leq \inf \{D(i,i) : D(i,i) > 0\} \) \( \|x\| = 1 \)

(b) \( \inf \langle Ax, x \rangle \) is an eigenvalue of \( V^{-1}R_N V^{*-1} \) of finite multiplicity; \( \|x\| = 1 \)

and

(c) if \( H_0 \) is the subspace spanned by the eigenvectors of \( V^{-1}R_N V^{*-1} \) corresponding to the eigenvalue \( \inf \langle Ax, x \rangle \), then \( \inf \langle Ax, x \rangle \geq \inf \langle Ax, x \rangle \).

\( \|x\| = 1 \) \( \|x\| = 1 \)

\( x \) in \( H_0 \)

To see that (2) implies the corresponding sufficient condition in (2) of the Corollary, one can verify that (2a) and (2b) imply that the smallest eigenvalue of \( V^{-1}R_N V^{*-1} \) exists and does not have eigenspace containing a set of natural basis vectors complete for the subspace; (2b) shows that the multiplicity of this subspace is finite; and (2b) plus (2c) show that this eigenvalue is not the limit of a sequence of distinct eigenvalues.

These conditions are not complex. Consider the finite-dimensional channel. First, one inspects the matrix \( V^{-1}R_N V^{*-1} \) and locates the diagonal elements \( \gamma_{ii} \) such that the ith row and ith column are all zero except for the ii element. Denote these elements as \( \gamma_i \). This is the set of eigenvalues of \( V^{-1}R_N V^{*-1} \) corresponding to natural basis vectors as eigenvectors. If the smallest such \( \gamma_i \) is strictly greater than \( \inf \langle V^{-1}R_N V^{*-1} x, x \rangle \), then the smallest eigenvalue of \( V^{-1}R_N V^{*-1} \) has no eigenvectors that are natural basis vectors, and so \( C_{WF}(P) > C_{WF}(P) \) for all \( P > 0 \). If the smallest \( \gamma_i \) is equal to \( \inf \langle V^{-1}R_N V^{*-1} x, x \rangle \), then one must determine the multiplicity of \( \inf \gamma_i = \gamma_0 \)
as an eigenvalue of $V^{-1}R_NV^{m-1}$. If this multiplicity is strictly greater than the number of times $\tau_0$ appears among the $\{\tau_i, i \geq 1\}$, then again $C_{WF}(P) > C(P)$ for all $P > 0$.

**Necessary Conditions**

The Corollary shows that the sufficient condition for feedback to increase capacity for all sufficiently large $P$ is also necessary, in the case of the finite-dimensional channel. Although the emphasis here has been on sufficient conditions, it is our conjecture that each of the four sufficient conditions given in the Corollary is also a necessary condition for the same result.

**Concluding Remarks**

It can be seen that the capacity problem with feedback for small $P$ reduces to consideration of the eigenmanifold for the smallest eigenvalue of $V^{-1}R_NV^{m-1}$, for the finite-dimensional channel. If this eigenvalue has multiplicity one, then feedback can increase capacity for every value of $P$ if the corresponding eigenvector is not a natural basis vector.

In the case of the infinite-dimensional channel, the same situation holds, except that the additional requirement is imposed of having the smallest eigenvalue be strictly less than the smallest limit point of $V^{-1}R_NV^{m-1}$.

For the case of sufficiently large $P$, the problem can be couched in terms of the reproducing kernel Hilbert space of $R_w$, say $H_w$. If the Gaussian cylinder set measure $\mu$ on $H_w$ defined by $\mu_N = \mu_j^{-1}$, $j$ the natural injection of $H_w$ into $\ell_2$ (i.e., $jx$ is just $x$ viewed as an element of $\ell_2$ rather than as an element of $H_w$), has diagonal covariance operator, then $C_{WF}(P) = C(P)$; otherwise,
for all sufficiently large P. In essence, this states that capacity can be increased by feedback for all sufficiently large P if the noise is correlated when it is viewed as belonging to $H_\mathcal{W}$, rather than to $\ell_2$.

The setup given here is rather general, and an obvious extension is to apply the same approach to the time-continuous channel. However, the structure of (Hilbert-Schmidt) Volterra operators is more complicated in $L_2[0,T]$ than in $\ell_2$, and an arbitrary covariance operator in $L_2[0,T]$ may not have a causal decomposition of the form $R_\mathcal{W} = VV^\dagger$, $V$ Volterra. Thus, a complete extension of these results in the form stated here does not seem possible.

References


END
DATE
FILM ED
4-88
DTIC