The Strong Converse for the Time-Continuous
White Gaussian Channel with Feedback
and other Gaussian Channels

by

Ian McKeague

FSU Statistics Report No. M621

April, 1982

The Florida State University
Department of Statistics
Tallahassee, Florida 32306
1. Introduction

Wolfowitz (1957) showed for discrete memoryless channels that the channel capacity is an upper bound on the efficient rates of transmission in a stronger sense than provided by the converse to the coding theorem. This result, known as the strong converse, states that if the transmission rate is held above the channel capacity then the probability of error must approach 1. Subsequent work has extended Wolfowitz' strong converse to other channel models: the time-discrete Gaussian channel, the time-continuous stationary Gaussian channel (Yoshihara (1964)), the time-discrete Gaussian channel with feedback (Wolfowitz (1968)) and general time-discrete memoryless channels with feedback (Kemperman (1973)).

This paper provides a strong converse for the time-continuous white Gaussian channel with feedback, extending the result of Wolfowitz (1968) to the time-continuous case. Our method involves a reduction of time-continuous code words to time-discrete code words in a way that preserves the causal property of the code words. A result of Yoshihara (1964) is then applied to the time-discrete code words to yield the strong converse. We also give the strong converse for a general Gaussian channel (section 3) and the time-continuous channel with arbitrary Gaussian noise without feedback (section 4). The channel capacity of the time-continuous Gaussian channel with feedback has been investigated extensively, see Kadota, Zakai and Ziv (1971), Hitsuda and Ihara (1975) and Liptser and Shiryaev (1977). It is given by $P_0/2$, where $P_0$ is the limitation on the average power.

2. The White Gaussian Channel with Feedback

For $0 \leq t \leq T$, let $\mathcal{B}_t$ denote the $\sigma$-field on $\mathbb{C}[0, T]$ generated by the evaluation functionals $\delta_s$, $s \leq t$. Let $(\mathcal{W}_t)$, $0 \leq t \leq T$ denote a standard Wiener process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{F}_t = \sigma(\mathcal{W}_s, s \leq t)$. 

1
Allowable code words for transmission during \([0, T]\) are to be given by measurable functions \(f: [0, T] \times C[0, T] \to \mathbb{R}\) which satisfy the following conditions.

a) Causality. For each \(t \in [0, T]\), the function \(f(t, \cdot)\) is \(\mathcal{B}_t\)-measurable.

b) Uniqueness of the output. The stochastic differential equation

\[
dY_t = f(t, Y)dt + dW_t, \quad 0 \leq t \leq T
\]

has a unique \(\mathcal{F}_t\)-adapted solution.

c) Average power limitation.

\[
\int_0^T f^2(t, Y)dt \leq P_0T, \text{ a.s.}
\]

This channel model is called the continuous-time white Gaussian channel with feedback (CWGF). For a discussion of the discrete-time analog (the DWGF) see Wolfowitz (1968). A code consisting of \(k\) allowable code words and corresponding Borel measurable decoding sets in \(C[0, T]\) with maximum probability of error \(\epsilon > 0\) is called a \((k, T, \epsilon)\)-code.

**Theorem 2.1.** Let \(\delta > 0, 0 < \epsilon < 1\). For the CWGF channel there exists a \(T_0\) such that all \((k, T, \epsilon)\)-codes with \(T \geq T_0\) satisfy

\[
k < e^{-T(C_0 + \delta)},
\]

where \(C_0 = P_0/2\) is the capacity of the channel.

Our proof of Theorem 2.1 relies on the following proposition.

**Proposition 2.2.** Corresponding to each \((k, T, \epsilon)\)-code, \(0 < \epsilon < 1/4\), for the CWGF channel and \(\delta > 0\) there exists a code for the DWGF channel with code
words $\phi_i \in \mathbb{R}^k$, output $X_i \in \mathbb{R}^k$ and decoding sets $B_i \subset \mathbb{R}^k$ for $i = 1, \ldots, k$ and $\varepsilon$ sufficiently large, such that

$$P\{X_i \in B_i\} \geq 1 - 4\varepsilon$$

$$P\{\sum_{j=1}^{k} \phi_{ij}^2 \leq (P_0 + 2\delta)T\} \geq 1 - 2\delta.$$  

Proof. Let the code words and corresponding output signals of a $(k, T, \varepsilon)$-code for the CWGF channel be denoted $f_i$ and $Y_i$ respectively where $0 < \varepsilon < 1/4$, $\delta > 0$. By a standard argument there exist disjoint cylinder sets $D_i$, $i = 1, \ldots, k$ consisting of finite unions of semiclosed intervals based on a finite set $\tau \subset (0, T]$ such that

$$P\{Y_i \in D_i\} \geq 1 - 2\varepsilon, i = 1, \ldots, k.$$  

Define

$$g_i(t, \omega) = \begin{cases} f_i(t, Y_i(\omega)) & \text{if } Y_i(\omega) \in C[0, T] \\ 0 & \text{otherwise.} \end{cases}$$

$g_i$ is measurable, $F_t^n$-adapted and $E \int_0^T g_i^2 dt < \infty$. There exists a sequence \{${g_i^n}$, $n \geq 1$\} of simple processes such that

$$E \int_0^T [g_i(t) - g_i^n(t)]^2 dt \to 0 \text{ as } n \to \infty.$$  

Let

$$Y_i^n(t, \omega) = \int_0^t g_i^n(s, \omega)ds + W_t.$$  

$(Y_i^n(t))_{t \in \tau}$ converges to $(Y_i(t))_{t \in \tau}$ in distribution as $n \to \infty$. By Liptser
and Shiryayev (1977, p. 242) \( \mu_\gamma \ll \nu_W \), so that \( \mu_\gamma^{\mathbb{D}_i} (\mathbb{D}_i) = 0 \), where \( \mathbb{D}_i \) is the boundary of \( \mathbb{D}_i \) and \( \mu_\gamma^{\mathbb{R}^2} \) is the projection of \( \mu_\gamma \) onto \( \mathbb{R}^2 \). Thus

\[
\lim_{n \to \infty} \mathbb{P}\{ \omega: Y_i^n(\omega) \in \mathbb{D}_i \} = \mathbb{P}\{ \omega: Y_i(\omega) \in \mathbb{D}_i \}.
\]

Choose \( n_0 \) such that

\[
\mathbb{P}\{ \omega: Y_i^{n_0}(\omega) \in \mathbb{D}_i \} \geq 1 - 3\varepsilon,
\]

\[
\mathbb{P}\{ \omega: \int_0^T [g_i^n(s, \omega)]^2 ds \leq (P_0 + \delta)T \} \geq 1 - \delta, \ i = 1, \ldots, k.
\]

Write the simple processes \( g_i^{n_0} \), \( i = 1, \ldots, k \) in the form

\[
g_i^{n_0}(s, \omega) = \sum_{j=0}^{k_1-1} g_{ij}(\omega) I_{[t_j, t_{j+1})}(s),
\]

where \( 0 = t_0 < \ldots < t_k = T \) do not depend on \( i \), \( g_{ij} \) is \( F_{t_j}^W \) measurable and \( \tau \subset \{ t_1, \ldots, t_k \} \). For each \( g_{ij} \) there is a sequence \( \{ f_{ij}^m, m \geq 1 \} \) of random variables each of which is \( \sigma(W_{s_1}, \ldots, W_{s_r}) \) measurable for some \( s_1, \ldots, s_r \leq t_j \), and \( f_{ij}^m(\omega) \to g_{ij}(\omega) \) as \( m \to \infty \), a.s.. Define

\[
h_i^m(s, \omega) = \sum_{j=0}^{k_1-1} f_{ij}^m(\omega) I_{[t_j, t_{j+1})}(s), \ i = 1, \ldots, k,
\]

\[
z_i^m(t, \omega) = \int_0^t h_i^m(s, \omega) ds + W_t.
\]

Then \( z_i^m(t) \to Y_i^{n_0}(t) \) as \( m \to \infty \), a.s., for each \( t \in [0, T] \). Also,

\[
\int_0^T [h_i^m(s, \omega)]^2 ds \to \int_0^T [g_i^{n_0}(s, \omega)]^2 ds \text{ as } m \to \infty, \ a.s..\]
Choose \( m_0 \) such that for \( i = 1, \ldots, k \),

\[
P(\omega: Z_i^{m_0} \in D_i) \geq 1 - 4\varepsilon
\]

\[
P(\omega: \int_0^T [h_i^0(s, \omega)]^2 ds \leq (P_0 + 2\varepsilon)T) \geq 1 - 2\varepsilon.
\]

It can be assumed, without loss of generality, that \( f_{i0}^{m_0} \) is a constant and \( f_{ij}^{m_0} \) is \( \sigma(W_{t_1}, \ldots, W_{t_j}) \) measurable for \( j = 1, \ldots, \ell - 1 \). Define for \( j = 1, \ldots, \ell, i = 1, \ldots, k \),

\[
\phi_{ij}^{m_0} = f_{i,j-1}^{m_0}(t_j - t_{j-1})^\frac{1}{2}
\]

\[
\gamma_j = (W_{t_j} - W_{t_{j-1}})(t_j - t_{j-1})^{-\frac{1}{2}}
\]

\[
X_{ij} = \phi_{ij} + \gamma_j.
\]

\( \gamma_j, j = 1, \ldots, \ell \) are iid \( N(0, 1) \) random variables and for \( j \geq 2 \), \( \phi_{ij} \) is \( \sigma(X_{il}, \ldots, X_{i,j-1}) \) measurable, \( \phi_{i1} \) is a constant. The random vectors \( \phi_i, i = 1, \ldots, k \) are code words for the DWGF channel. Let \( E_i = \pi(D_i) \), where \( \pi \) denotes the projection of \( C[0, T] \) onto \( \mathbb{R}^k \), determined by \( t_1, \ldots, t_\ell \).

Define the invertible linear map \( A: \mathbb{R}^k \to \mathbb{R}^k \) by the matrix

\[
A_{ij} = \begin{cases} 
(t_j - t_{j-1})^\frac{1}{2} & \text{if } i \geq j \\
0 & \text{otherwise}.
\end{cases}
\]

Let \( B_i = A^{-1}(E_i) \). Then, treating \( B_i \) as a decoding set for \( \phi_i \),
\[ P(X_i \in E_i) = P(A(X_i) \in E_i) \]

\[ = P\left( \sum_{j=1}^P (\phi_{ij} + \gamma_j)(t_j - t_{j-1}) \right)_{\text{p}=1}^P \epsilon E_i \]

\[ = P\left( \int_0^{m_0} h_i(s, \omega) ds + \sum_{p=1}^P \omega_p \right) \epsilon E_i \]

\[ = P(Z_i \in D_i) \geq 1 - 4\epsilon. \]

Similarly,

\[ P\left( \sum_{j=1}^P \phi_{ij}^2 \leq (P_0 + 2\delta) T \right) \geq 1 - 2\delta. \]

This completes the proof of the proposition. \( \square \)

Proof of Theorem 2.1. Yoshihara (1964, pp. 220-221) gives a method which derives this result from a code for the discrete-time Gaussian channel without feedback. It is easily checked that the method carries through to the feedback case and can be applied to the DWF code given in Proposition 2.1. \( \square \)

3. The General Gaussian Channel

In this section we give the strong converse for the general Gaussian channel discussed in McKeague (1981). Let \( E \) denote a quasi-complete locally convex space with topological dual \( E^* \). Let \( \sigma(E^*) \) denote the \( \sigma \)-algebra generated by \( E^* \). Let \( \mu_N \) be a zero mean Gaussian measure on \( \sigma(E^*) \) with covariance operator \( R_N : E^* \to E \) and reproducing kernel Hilbert space (RKHS) denoted \( H_N \). When a message \( x \in E \) is transmitted the distribution of the received signal is given by \( \mu_N \circ f_x^{-1} \) where \( f_x : E \to E \) is defined by
\( f_{\varepsilon}(y) = x + y \). A \((k, n, \varepsilon)\)-code consists of \( k \) code words \( x_1, \ldots, x_k \in H_N \) having maximum probability of error \( \varepsilon > 0 \) and such that

\[
\dim \text{sp}\{x_1, \ldots, x_k\} \leq n, \quad \|x_i\|_{H_N} \leq nP_0, \quad \text{where } P_0 > 0 \text{ is the average power limitation. The coding capacity of this channel is given by } \frac{1}{2}\log(1+P_0), (\text{see McKeague (1981)}). \text{ The following result is a strong converse for the general Gaussian channel.}

**Theorem 3.1.** Let \( \delta > 0, 0 < \varepsilon < 1 \). For the general Gaussian channel, with \( H_N \) separable, there exists an \( n_0 \) such that all \((k, n, \varepsilon)\)-codes with \( n \geq n_0 \) satisfy

\[
k < \frac{n(C_0 + \delta)}{2n},
\]

where \( C_0 = \frac{1}{2}\log(1+P_0) \) is the capacity of the channel.

Our proof of Theorem 3.1 follows immediately from the next proposition and the strong converse for the discrete-time white Gaussian channel.

**Proposition 3.2.** Corresponding to each \((k, n, \varepsilon)\)-code, \( 0 < \varepsilon < 1/3 \), for the general Gaussian channel there exists a \((k, n, 3\varepsilon)\)-code for the discrete-time white Gaussian channel.

**Proof.** Let \( 0 < \varepsilon < 1/3 \). Given a \((k, n, \varepsilon)\)-code for the general Gaussian channel with code words \( x_i, i = 1, \ldots, k \), let \( D_i, i = 1, \ldots, k \) be disjoint cylinder sets whose base sets are finite unions of semiclosed intervals such that

\[
\mu_N \circ f_{x_i}^{-1}(D_i) \geq 1 - 2\varepsilon.
\]

Let \( g_1, \ldots, g_\varepsilon \in E^* \) be such that each \( D_i \) is \( \sigma(g_1, \ldots, g_\varepsilon) \) measurable. Let \( j_N \) denote the injection of \( H_N \) into \( E \). Let \( \{u_i, i \geq 1\} \subseteq \text{range}(j_N^*) \)
be a CONS for $H_N$ and let $\pi_X$ denote the Hilbert space projection

$\pi_X: H_N + \text{sp}\{u_1, \ldots, u_r\}$, $r \geq 1$. By the dominated convergence theorem

it is possible to choose $r$ sufficiently large such that

$$\mu_N \circ f^{-1}_{\pi_X(x_i)}(D_i) \geq 1 - 3\varepsilon, \ i = 1, \ldots, k.$$ 

Put $y_i = \pi_X(x_i)$. Then $\text{dim} \ sp(y_1, \ldots, y_k) \leq n$. Choose an orthonormal

sequence $e_1, \ldots, e_m$ in $H_N$ such that

$$\text{sp}(e_1, \ldots, e_n) \supset \text{sp}(y_1, \ldots, y_k)$$

and

$$\text{sp}(e_1, \ldots, e_m) = \text{sp}(y_1, \ldots, y_k, j^*_N(g_1), \ldots, j^*_N(g_N)),$$

where $m \geq n$. Choose $\phi_i \in E^*$ such that $e_i = j^*_N(\phi_i), \ i = 1, \ldots, m$. Then

define a linear map $J: E \rightarrow R^m$ by

$$J(x) = (\phi_1(x), \ldots, \phi_m(x)).$$

The code $(y_i, D_i)$ for the general Gaussian channel is now mapped by $J$ to

a code in $R^m$ with code words $u_i = J(y_i)$ and disjoint Borel measurable

decoding sets $F_i \subset R^m, \ i = 1, \ldots, k$ where $D_i = J^{-1}(F_i)$. The noise distribution $\mu_N \circ J^{-1}$ on $R^m$ is zero mean Gaussian with covariance matrix equal
to the identity. Also

$$\mu_N \circ J^{-1}(F_i - u_i) = \mu_N \circ J^{-1}(J(D_i - y_i)) \geq \mu_N(D_i - y_i) \geq 1 - 3\varepsilon,$$

$$\sum_{j=1}^{n} u_{ij}^2 = \sum_{j=1}^{n} \langle y_i, e_j \rangle_{H_N}^2 = \|\pi_X(x_i)\|_{H_N}^2 \leq \|x_i\|_{H_N}^2 \leq n\varepsilon.$$ 

Since $u_{ij} = 0$, for $j = n+1, \ldots, m$, it is easily seen that the decoding

sets $F_i \subset R^m$ can be modified to give new (disjoint) decoding sets contained
in \( \mathbb{R}^n \), without increase in error probability. The code words \( u_i \) restricted to \( \mathbb{R}^n \) and the new decoding sets form a \((k, n, 3\epsilon)\)-code for the discrete-time Gaussian channel. □

4. The Time-Continuous Channel with General Gaussian Noise

Let \( N = \{N_t, t \geq 0\} \) be a Gaussian noise process with RKHS \( H_N \). Allowable code words for transmission on \([0, T]\) are functions \( S: [0, \infty) \to \mathbb{R} \) which vanish outside \([0, T]\), whose restrictions to \([0, T]\) belong to \( H^T_N \) (the RKHS of \( N \) restricted to \([0, T]\)) and which satisfy \( \|S\|_{H^T_N}^2 \leq P_0 T \), where \( P_0 > 0 \) is the average power limitation. This channel is called the time-continuous Gaussian channel (without feedback) and it has coding capacity \( C_0 = \frac{P_0}{2} \) (see McKeague (1981)). The following strong converse for this channel is an extension of a result of Yoshihara (1964) who restricted attention to stationary Gaussian noise.

**Theorem 4.1.** Let \( \delta > 0, 0 < \epsilon < 1 \). For the time-continuous Gaussian channel there exists a \( T_0 \) such that all \((k, T, \epsilon)\)-codes with \( T \geq T_0 \) and \( H^T_N \) separable satisfy

\[
k < e^{T(C_0 + \delta)}
\]

where \( C_0 = \frac{P_0}{2} \) is the capacity of the channel.

**Proof.** A code for the time-continuous Gaussian channel may be reduced to a code for the time-discrete Gaussian channel as in the proof of Proposition 3.2. The method of Yoshihara (1964, pp. 220-221) then gives the result. □
REFERENCES


