NEW LIMITATION CHANGE

TO
Approved for public release, distribution unlimited

FROM
Distribution authorized to U.S. Gov’t. agencies and their contractors; Administrative/Operational Use; JAN 1965. Other requests shall be referred to Office of Naval Research, One Liberty Center, 875 North Randolph Street, Arlington, VA 22203-1995.

AUTHORITY
onr ltr, 28 jul 1977

THIS PAGE IS UNCLASSIFIED
NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.
AN ESTIMATION THEORY APPROACH TO
PULSE AMPLITUDE MODULATION RECEIVERS

By

Toby Berger

January 5, 1965

Technical Report No. 438

Cruft Laboratory
Division of Engineering and Applied Physics
Harvard University - Cambridge, Massachusetts
AN ESTIMATION THEORY APPROACH
TO PULSE AMPLITUDE MODULATION RECEIVERS

by

Toby Berger

January 5, 1965

The research reported in this document was made possible through
support extended to Cruft Laboratory, Harvard University, by the
U. S. Army Research Office, the U. S. Air Force Office of Scientific
Research, and the U. S. Office of Naval Research under the Joint
Services Electronics Program by Contract Nonr-1866(16). Reproduc-
dition in whole or in part is permitted for any purpose of the United
States Government.

Technical Report No. 438

Cruft Laboratory
Division of Engineering and Applied Physics
Harvard University
Cambridge, Massachusetts
AN ESTIMATION THEORY APPROACH TO PULSE AMPLITUDE MODULATION RECEIVERS

by

Toby Berger

Division of Engineering and Applied Physics

Harvard University, Cambridge, Massachusetts

ABSTRACT

By generalizing a trivial problem in estimation theory, some interesting results have been obtained for a class of pulse amplitude modulation (PAM) communication channels. In particular, we consider a linear channel with memory perturbed by additive noise which is uncorrelated with the signal statistics. It is shown that, for specified second-order statistics and a mean square error criterion, the Bayes estimation rule for the specific case of zero-mean Gaussian signal and noise statistics is also both the best linear estimation rule and the minimax estimation rule. This fact is used in the theoretical design of sampled-data and continuous time receivers for such PAM systems. In the continuous time case with an infinite observation interval and stationary noise statistics, the receiver thus designed agrees with the optimum linear receiver previously derived for this case by Tufts [2] using a very different approach; the present analysis shows that Tufts's linear receiver is, therefore, also the best nonlinear receiver in the minimax sense.
In this report we consider the consequences of the following trivial estimation problem.

Problem 1:

The real random variable \( a \) has the Gaussian probability density

\[
p_1(a) = \frac{1}{\sqrt{2\pi} b} e^{-\frac{a^2}{2b^2}} ; \quad -\infty < a < \infty .
\]

(1)

The real random variable \( y \) is statistically independent of \( a \) and has the Gaussian probability density

\[
p_2(y) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{y^2}{2\sigma^2}} , \quad -\infty < y < \infty .
\]

(2)

The sum random variable \( x = a + y \) is observed. It is desired to find that estimation rule \( \hat{a}(x) \) which minimizes the mean square error \( D \) between \( \hat{a}(x) \) and \( a \). This mean square error is defined by:

\[
D = (\hat{a}(x) - a)^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\hat{a}(a+y)-a]^2 p_1(a) p_2(y) \, da \, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\hat{a}(x)-a]^2 p_1(a) p_2(x-a) .
\]

(3)

Solution:

It is well known ([1], p. 190) that the optimum estimate \( \hat{a}_1(x) \) is the mean of \( a \), conditional upon having observed \( x \), namely

\[
\hat{a}_1(x) = \int_{-\infty}^{\infty} da \, p(a|x) = \frac{\int_{-\infty}^{\infty} a p_2(x-a) p_1(a) \, da}{\int_{-\infty}^{\infty} p_2(x-a) p_1(a) \, da} .
\]

(4)
Substituting (1) and (2) into (4) and performing the indicated computations
(See Appendix I) yields the result:

\[ \hat{a}_1(x) = \left( \frac{b^2}{b^2 + \sigma^2} \right) x. \]  

(5)

The resultant minimum value of \( D \) found upon substitution of \( \hat{a}_1(x) \) from
(5) into (3) is (See Appendix I):

\[ D_1 = \frac{b^2 \sigma^2}{b^2 + \sigma^2}. \]  

(6)

We now introduce some terminology and notation which will be
employed below in discussing other estimation problems related to Problem 1.

The expected value of a function \( f(a, y) \) of two real random variables
with joint density \( p(a, y) \) will be denoted by

\[ E_p[f(a, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(a, y) p(a, y) \, da \, dy. \]  

(7)

We shall denote by \( W \) the class of all joint probability densities
\( p(a, y) \) which satisfy the following requirements

\[ E_p(a^2) = b^2 \]  

(8a)

\[ E_p(ay) = 0 \]  

(8b)

\[ E_p(y^2) = \sigma^2. \]  

(8c)
Problem 2:

It is known that the joint density of $a$ and $y$ is an element of $W$. The sum random variable $x = a + y$ is observed, and it is required that the estimation rule $\hat{a}(x)$ be linear, i.e., the only permissible estimation rules are of the form $\hat{a}(x) = kx$. We seek that estimation rule $\hat{a}_2(x)$ of this form which minimizes the mean square error $D$ defined by

$$D(\hat{a}, p) = E_p \left\{ (\hat{a}(x) - a)^2 \right\} .$$

Solution:

First note that if $\hat{a}(x)$ is linear, then $D(\hat{a}, p)$ is independent of which $p \epsilon W$ is used to compute it, because

$$D(kx, p) = E_p \left\{ (k(a+y) - a)^2 \right\} = E_p \left\{ (k-1)^2 a^2 \right\} + E_p \left\{ 2k(k-1)ay \right\} + E_p \left\{ k^2 y^2 \right\} ,$$

from which by use of (8) we obtain

$$D(kx, p) = (k-1)^2 b^2 + k^2 \sigma^2 = f(k) .$$

Setting $f'(k) = 0$ yields $k = b^2/2b^2 + \sigma^2$, while $f''(k) = 2\sigma^2 > 0$ insures a minimum. Accordingly, $\hat{a}_2(x) = \hat{a}_1(x)$.

Alternatively, one could argue as follows. Since the value of $D(\hat{a}, p)$ is independent of $p \epsilon W$ when $\hat{a}$ is linear, we can compute it using any $p \epsilon W$ whatsoever. In particular, suppose we were to use

$$\tilde{p}(a, y) = p_1(a)p_2(y)$$

where $p_1(a)$ and $p_2(y)$ are given by Eqs. (1) and (2), respectively; one easily verifies that $\tilde{p}(a, y) \epsilon W$. Now Problem 1 tells us that for any $\hat{a}(x)$ and, hence, a fortiori, for any linear $\hat{a}(x)$, we have
Since $\hat{a}_1$ is itself linear, we may conclude that $\hat{a}_2(x) = \hat{a}_1(x)$.

The alternative method of solution of Problem 2 proves useful in solving the following minimax estimation problem.

**Problem 3:**

It is desired to find the estimation rule $\hat{a}(x)$ which minimizes the maximum mean square error that occurs when $p(a, y)$ is varied over $W$, i.e., our optimum estimate $\hat{a}_3(x)$ is defined by the requirement

$$\left(\forall \hat{a}(x)\right)\left(\sup_{p \in W} D(\hat{a}, p) \geq \sup_{p \in W} D(\hat{a}_3, p)\right).$$

**Solution:**

The following argument establishes that $\hat{a}_3(x) = \hat{a}_1(x)$. Since $\hat{a}_1(x)$ is linear, we know that $D(\hat{a}_1, p) = \text{const.} = D_1$ for all $p \in W$ and, therefore,

$$\sup_{p \in W} D(\hat{a}_1, p) = D_1.$$  \hspace{1cm} (14)

On the other hand, we know from Problem 1 that $\hat{a}_1(x)$ yields the least value of $D$ when $p = \tilde{p}$. Accordingly, for all $\hat{a}(x)$ we have

$$\sup_{p \in W} D(\hat{a}, p) = D(\hat{a}, \tilde{p}) \geq D(\hat{a}_1, \tilde{p}) = D_1 = \sup_{p \in W} D(\hat{a}_1, p).$$

Comparing (15) with (13) shows that $\hat{a}_3(x) = \hat{a}_1(x)$.

In summary then, we see that in the space $W$ of joint probability densities as defined by (8), the least favorable distribution with respect
to a mean square error criterion is $\hat{p}(a, y)$ of equation (11) in which $a$ and $y$ are statistically independent, zero mean and Gaussian. Since $\hat{a}_1(x)$, the optimum estimation rule for $\hat{p}(a, y)$ and a mean square error criterion, is linear in $x$, we have seen that it is in fact the solution to all three of the estimation problems we have considered.*

We now employ the knowledge gained from the above problems in order to solve some estimation problems regarding the pulse amplitude modulation system depicted in Fig. 1.

*We could have defined $W$ somewhat more broadly and still had $\hat{a}_1(x)$ as the minimax solution. In particular, we could have replaced (8a) by $E_p(a^2) \leq b$ and (8c) by $E_p(y^2) \leq \sigma^2$, and we could have permitted probability distributions $P(a, y)$ which, because of discrete concentrations of probability, possess no density $p(a, y)$. 

---

Figure 1: Block Diagram of PAM Communication System
The transmitter forms the sum \( \sum_{k=1}^{N} a_k s(t-\tau_k) \) from the random vector \( a \), the known vector \( \tau \) and the known function \( s(t) \). This transmitter output is then sent over a linear, dispersive channel characterized by an impulse response \( h(t) \). The receiver input \( x(t) \) is an additive combination of the channel response \( u(t) \) to the transmitter output and a sample function \( y(t) \) of a random process. The receiver operates on the noisy waveform \( x(t) \) to produce the vector \( \hat{a} \) of estimates of the respective components of \( a \).

The design of the receiver may, therefore, be viewed as a problem in statistical estimation theory. We start with a relatively trivial problem analogous to Problem 1.

**Problem 4:**

Suppose that the message vector \( a \) has a zero-mean Gaussian probability density of the form

\[
p(a) = \frac{1}{(2\pi)^{N/2} |\Phi|^{1/2}} e^{-\frac{1}{2} a^T \Phi^{-1} a} = \frac{1}{(2\pi)^{N/2} |\Phi|^{1/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \phi_{ij}^{-1} a_i a_j
\]

In Eq. (16), \( \phi_{ij}^{-1} \) represents the \( i\)-\( j\)th element of the inverse of the autocorrelation matrix \( \Phi \). Further suppose that \( y(t) \) is a sample function of a zero mean, possibly nonstationary Gaussian random process statistically independent of \( a \) and having the autocorrelation function \( \gamma(t, s) = \overline{y(t)y(s)} \).

The receiver input \( x(t) \) is observed at \( M \) distinct time instants \( t_1, t_2, \ldots, t_M \); the observed random variables, therefore, constitute
an $M$-dimensional column vector $\mathbf{x}$ with components $x_m = x(t_m)$; 
$m = 1, 2, \ldots, M$. We seek that estimation rule $\hat{\mathbf{a}}(\mathbf{x})$ which minimizes the
mean square error between $\hat{\mathbf{a}}$ and $\mathbf{a}$ defined as

$$D = \left[ \hat{\mathbf{a}}(\mathbf{x}) - \mathbf{a} \right]^2 = \sum_{i=1}^{N} \int \int d\mathbf{a} d\mathbf{x} \left[ \hat{a}_i(\mathbf{x}) - a_i \right]^2 p(\mathbf{a}) q(\mathbf{x} | \mathbf{a}) . \tag{17}$$

The integration in Eq. (17) extends over the entire $N+M$ dimensional space 
of which $d\mathbf{a} d\mathbf{x} = dx_1 \ldots dx_N dx_{M+1} \ldots dx_M$ is a differential element of 
volume, and $q(\mathbf{x} | \mathbf{a})$ is the $M$-dimensional probability density of $\mathbf{x}$
for a given message vector $\mathbf{a}$.

Solution:

The signal component $u(t)$ of the receiver input waveform $x(t)$
is given by

$$u(t) = \sum_{j=1}^{N} a_j r(t - \tau_j), \text{ where } r(t) = \int_{-\infty}^{\infty} s(t - \xi) h(\xi) d\xi . \tag{18}$$

In order to write down $q(\mathbf{x} | \mathbf{a})$ explicitly, we introduce the following notation.
Let $\Gamma$ be the $M \times M$ autocorrelation matrix whose $m-n$th element is
$\gamma(t_m, t_n) = \gamma_{mn}$, and let $\gamma_{mn}^{-1}$ be the $m-n$th element of $\Gamma^{-1}$. Let

$$u(\mathbf{a}) \text{ be the column vector with components } u(t_m) = \sum_{j=1}^{N} a_j r(t_m - \tau_j) =$$

$$\sum_{j=1}^{N} c_{mj} a_j = u_m(\mathbf{a}); m = 1, 2, \ldots, M. \text{ We then may write}$$
\[
q(x|a) = \frac{e^{-\frac{1}{2} (x^T - u^T(a))\Sigma^{-1}(x - u(a))}}{(2\pi)^{N/2} |\Sigma|^{1/2}} = \frac{e^{-\frac{1}{2} \sum_{m=1}^{M} \sum_{n=1}^{M} \gamma_{mn}^{-1}(x_m - u_m(a))(x_n - u_n(a))}}{(2\pi)^{M/2} |\Gamma|^{1/2}}
\]

It is a standard result of estimation theory ([1], pp. 188-190) that the optimum estimation rule is the mean of \( a \) conditional upon having observed \( x \), the \( i \)th component of which is

\[
\hat{a}_i(x) = \frac{\int a_i p(a) q(x|a) \, da}{\int q(x|a) p(a) \, da}.
\]

The integrations indicated in Eq. (20) are performed in Appendix II and yield the result

\[
\hat{a}_i(x) = \sum_{m=1}^{M} \theta_{im} x_m; \quad i = 1, 2, \ldots, N.
\]

The constants \( \theta_{im} \) appearing in (21) are defined in Appendix II.

The most important property of the estimation rule specified by (21) is that, like the optimum estimate \( \hat{a}_1(x) \) of Problems 1, 2, and 3, it is linear in the observed random variables \( x_m \).

We now introduce some terminology and notation which will be employed below in discussing other estimation problems related to Problem 4.
The expected value of a function \( f(a, \chi) \) of two real random vectors with joint probability density \( p(a, \chi) \) will be denoted by

\[
E_p[f(a, \chi)] = \int \int f(a, \chi) p(a, \chi) \, da \, d\chi. \tag{22}
\]

We shall denote by \( V \) the class of \((M+N)\) dimensional joint probability densities \( p(a, \chi) \) which satisfy the following requirements

\[
\begin{align*}
E_p(a_j a_k) &= \phi_{jk} \quad \text{(22a)} \\
E_p(a_j \chi_m) &= 0 \quad j, k = 1, 2, \ldots, N; \ m, n = 1, 2, \ldots, M. \quad \text{(22b)} \\
E_p(\chi_m \chi_n) &= \gamma_{mn} \quad \text{(22c)}
\end{align*}
\]

**Problem 5:**

Let \( a \) be the \( N \)-component random message vector and let \( \chi \) be \( M \)-component random vector whose elements are the noise samples \( \chi(t_m) = \chi_m \). It is known that \( p(a, \chi) \in V \). The \( M \)-component random vector \( \chi \) whose elements are the receiver input samples \( x(t_m) = x_m \) is observed, and it is required that the estimation rule \( \hat{\chi}(\chi) \) be linear, i.e., the only permissible estimation rules are of the form \( \hat{\chi}(\chi) = K \chi \), where \( K \) is an \( N \times M \) rectangular matrix. We seek that estimation rule of this form which minimizes the mean square error \( D \) defined by

\[
D(\hat{\chi}, p) = E_p\left\{[\hat{\chi}(\chi) - a]^2\right\} = \sum_{i=1}^{N} E_p\left\{[\hat{\chi}_i(\chi) - a_i]^2\right\} \tag{23}
\]
Solution:

For linear $\hat{a}(x)$ we find that $D(\hat{a}, p)$ is independent of which $p \in V$ is used to compute it, since we have

$$D(Kx, p) = \sum_{i=1}^{N} E_{p} \left\{ \sum_{m=1}^{M} k_{im} x_{m} - a_{i} \right\}^{2} = \sum_{i=1}^{N} E_{p} \left\{ \sum_{m=1}^{M} k_{im} [y_{m} + u_{m}(a)] - a_{i} \right\}^{2}$$

$$D(Kx, p) = \sum_{i=1}^{N} E_{p} \left\{ \sum_{m=1}^{M} k_{im} y_{m} + \sum_{j=1}^{N} c_{mj} a_{j} - a_{i} \right\}^{2}$$

$$D(Kx, p) = \sum_{i=1}^{N} E_{p} \left\{ \sum_{m=1}^{M} k_{im} y_{m} + \sum_{j=1}^{N} \left( \sum_{m=1}^{M} k_{im} c_{mj} \right) a_{j} - a_{i} \right\}^{2} \quad (24)$$

When the squaring operation indicated in (24) is carried out and the expectation operation is commuted with the summations, it becomes clear from (22) that $D(Kx, p)$ is independent of $p \in V$. It then follows by reasoning directly analogous to the alternative method of solution of Problem 2 that the desired estimation rule is that already specified by Eq. (21) as the solution to Problem 4, i.e., the optimum estimation rule has $K = \Theta = (\theta_{im})$; the distribution analogous to $\tilde{p}(a, y)$ of Eq. (11) is

$$\tilde{p}(a, y) = \frac{e^{- \frac{1}{2} [a \Theta^{-1} a + y \Gamma^{-1} y]}}{(2\pi)^{\frac{1}{2}} |\Theta|^{\frac{1}{2}} |\Gamma|^{\frac{1}{2}}} \quad (25)$$
Problem 6:

It is desired to find that estimation rule $\hat{a}(x)$ which minimizes the maximum assumed by $D(\hat{a}, p)$ as $p$ is varied over $V$.

Solution:

In complete analogy to Problem 3, we see that Eq. (21) specifies the desired estimation rule.

In summary, the common solution of Problems 4, 5, and 6 is the optimum sampled-data mean square error receiver for the PAM system of Fig. 1, where the term "optimum" is, of course, to be construed with regard to the statistical knowledge assumed and the performance criteria adopted in these problems.

Of particular interest is the limit of infinitely large $M$ which corresponds to the case of continuous observation of the receiver input. Although it is possible to perform the limiting operations directly upon the form of the $\theta_{km}$ as derived in Appendix II, it proves somewhat more convenient to revise our method of attack slightly. (For background material regarding the analysis that follows, the reader is referred to [1], pp. 98-106.)

Suppose the receiver input $x(t)$ is observed over the interval $T_1 < t < T_2$. The revision in our method consists of replacing the vector $x$ of receiver input samples with a vector of expansion coefficients of $x(t)$ in terms of a set of functions $\{f_m(t)\}$ orthonormal over $(T_1, T_2)$.

In particular, we choose for our functions $f_m(t)$ the orthonormal solutions of the Karhunen-Loéve integral equation
Thus, the elements of $x$ are

$$x_m = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} f_m(t) x(t) \, dt .$$

Under the assumptions of Problem 4, the expansion coefficients $x_m$ for a fixed message vector $\tilde{a}$ would be statistically independent Gaussian random variables with variances $\lambda_m$ and means $\mu_m(\tilde{a})$ defined by

$$\mu_m(\tilde{a}) = \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} f_m(t) u(t) \, dt = \sum_{j=1}^{N} \left( \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} f_m(t) r(t-\tau_j) \, dt \right) a_j = \sum_{j=1}^{N} \xi_{mj} a_j .$$

Accordingly, in place of Eq. (19) we obtain

$$q(x|\tilde{a}) = \frac{\operatorname{e}^{-\frac{1}{2} \sum_{m=1}^{M} \frac{[x_m - \mu_m(\tilde{a})]^2}{\lambda_m}}}{(2\pi \lambda_m)^{1/2} \operatorname{e}^{-\frac{1}{2} \sum_{m=1}^{M} \frac{\delta_{mn}^{-1}}{\lambda_m}}} .$$

Therefore, the computations done in Appendix II remain applicable if we replace $c_{mj}$ by $\xi_{mj}$ and $\gamma_{mn}^{-1}$ by $\frac{\delta_{mn}}{\lambda_m}$. In the limit of infinite $M$, the vector $x$ of expansion coefficients provides a representation of the observed receiver input voltage $x(t)$, $T_1 < t < T_2$, which is sufficiently accurate for most practical purposes. The components $\hat{a}_i(x)$ of our optimum estimate approach linear functionals of $x(t)$, $T_1 < t < T_2$, which
will be denoted by \( \hat{a}_i[x(t), T_1 < t < T_2] = g_i \); \( i = 1, 2, \ldots, N \). We now characterize the \( g_i \) more completely by examining the limiting behavior of Eq. (21).

Let \( \theta_i(t) \) be the time function whose \( m^{th} \) Karhunen-Loève expansion coefficient is \( \theta_{im} \), i.e., let

\[
\theta_i(t) = \sum_{m=1}^{\infty} \theta_{im} \lambda_m f_m(t); \quad \theta_{im} = \int_{T_1}^{T_2} f_m(t) \theta_i(t) dt .
\]

Then, let \( z_i(t) \) be the solution of the integral equation

\[
\theta_i(t) = \int_{T_1}^{T_2} \gamma(t, s) z_i(s) ds ; \quad T_1 < t < T_2 .
\]

It follows that the \( i^{th} \) component of our optimum estimate may be expressed in the form

\[
g_i = \sum_{m=1}^{\infty} \theta_{im} x_m = \sum_{m=1}^{\infty} \int_{T_1}^{T_2} \frac{x_m}{\lambda_m} f_m(t) \theta_i(t) dt = \sum_{m=1}^{\infty} \int_{T_1}^{T_2} \frac{x_m}{\lambda_m} \int_{T_1}^{T_2} f_m(t) \gamma(t, s) z_i(s) ds dt .
\]

\[
g_i = \sum_{m=1}^{\infty} \frac{x_m}{\lambda_m} \int_{T_1}^{T_2} ds z_i(s) \int_{T_1}^{T_2} dt \gamma(t, s) f_m(t) = \sum_{m=1}^{\infty} \frac{x_m}{\lambda_m} \int_{T_1}^{T_2} ds z_i(s) \int_{T_1}^{T_2} \gamma(s, t) f_m(t) dt .
\]

\[
g_i = \int_{T_1}^{T_2} ds z_i(s) \int_{T_1}^{T_2} \frac{x_m}{\lambda_m} f_m(t) = \int_{T_1}^{T_2} z_i(s) x(s) ds .
\]
In deriving Eq. (32), use has been made of the fact that \( \gamma(t, s) \) is a symmetric kernel. The equation indicates that \( g_i \) may be formed by passing \( x(t) \) through a linear, time-invariant filter with impulse response \( w_1(t) \) given by

\[
 w_1(t) = \begin{cases} 
 z_1(T-t) ; & -T_2 + T < t < -T_1 + T \\
 0 ; & \text{elsewhere} 
\end{cases}
\]  

(33)

The filter output at time \( T \) then will be

\[
 \int_{-\infty}^{\infty} x(T-\tau)w_1(\tau) d\tau = \int_{-T_2+T}^{T_2} x(T-\tau)z_1(T-\tau) d\tau = \int_{-\infty}^{\infty} z_1(s)x(s) ds = g_i.
\]  

(34)

It appears difficult to extend the general analysis beyond this point.

In the specific case in which \( T_1 \to -\infty \) and \( T_2 \to \infty \), an explicit solution for \( w_1(t) \) can be obtained if the noise process is wide sense stationary. The solution for this case was found by Tufts [2] using a very different approach. Tufts derived the mean square error optimum linear receiver under the assumption that the only statistical knowledge available to the receiver designer was the message vector autocorrelation matrix \( \Theta \) and the noise autocorrelation function \( \gamma(t, s) = \psi(t-s) \); signal and noise were assumed additive and independent. Calculus of variations yielded an integral equation the solution of which was the transfer function \( W_1(f) \) of the optimum receiver filter, namely
\[ W_i(f) = \frac{R^*(f)}{\psi(f)} \sum_{n=1}^{N} d(i, k) e^{j2\pi f(\tau_k - T)} , \quad (35) \]

where \( W_i(f) = \int_{-\infty}^{\infty} w_i(t) e^{-j2\pi ft} dt \), \( R(f) = \int_{-\infty}^{\infty} r(t) e^{-j2\pi ft} dt \),

and \( \psi(f) = \int_{-\infty}^{\infty} \psi(\tau) e^{-j2\pi f\tau} d\tau \). Equation (35) represents a "matched filter" \( \frac{R^*(f)}{\psi(f)} \) independent of \( i \) followed by a delay line with \( N \) taps spaced like the transmission instants \( \tau_k \) and weighted by gain coefficients which vary with \( i \). The optimum estimate \( g_i \) is the sum of the weighted tap outputs of time \( T \) (see Fig. 2).

![Diagram](image)

**Fig.2-Optimum Receiver for Stationary Noise and an Infinite Observation Interval**
In Appendix III, Eq. (30) is solved for $T_1 \rightarrow -\infty$, $T_2 \rightarrow +\infty$ and stationary noise, and the solution is shown to be equivalent to Tufts' solution as given by (35) and Fig. 2. The present analysis shows that the resultant receiver is not only the best linear one, but also the best nonlinear receiver in the minimax sense. For either of these interpretations of the receiver, the joint distribution $p[a, y(t)]$ of signal and noise has been assumed to belong to the continuous analog of $V$, i.e., it is known that for all $j, k, = 1, 2, \ldots, N$ and for all $t_1, t_2 \in (-\infty, \infty)$

\begin{align}
E_p(a_j a_k) &= \phi_{jk} \quad (36a) \\
E_p[a_j y(t_1)] &= 0 \quad (36b) \\
E_p[y(t_1) y(t_2)] &= \gamma(t_1, t_2) \quad (36c)
\end{align}

**Conclusion:**

It has been shown that in mean square error estimation problems with additive, uncorrelated noise and known second-order statistics, the Bayes estimation rule for the specific case of statistically independent, zero-mean Gaussian signal, and noise statistics is also both the best linear estimation rule and the best minimax estimation rule. This fact has been used to design optimum sampled-data and continuous time receivers for a practical PAM communication system. In the continuous time case with an infinite observation interval and stationary noise statistics, the receiver thus designed agrees with the optimum linear receiver previously derived for this case.
by Tufts [2] using a very different approach; the present analysis shows that Tufts's linear receiver is, therefore, also the best nonlinear minimax receiver.

The above theory could be employed as the mathematical basis for an investigation of the extent of degradation in system performance suffered by adopting a sampled data as opposed to a continuous time receiver. It is also hoped that the above constitutes a useful contribution to the complex field of the interrelations between different system performance criteria.
\[ \int_{-\infty}^{\infty} a p_2(x-a)p_1(a)da + \int_{-\infty}^{\infty} \frac{1}{2} \left[ \frac{(x-a)^2}{\sigma^2} + \frac{a^2}{b^2} \right] da + \int_{-\infty}^{\infty} \frac{1}{2} \left[ \frac{a^2 - 2ax}{\sigma^2} + \frac{a^2}{b^2} \right] da \]

\[ a_1(x) = \frac{-\int_{-\infty}^{\infty} p_2(x-a)p_1(a)da}{\int_{-\infty}^{\infty} p_2(x-a)p_1(a)da} = \frac{-\int_{-\infty}^{\infty} \frac{1}{2} \left[ \frac{(x-a)^2}{\sigma^2} + \frac{a^2}{b^2} \right] da}{\int_{-\infty}^{\infty} \frac{1}{2} \left[ \frac{a^2 - 2ax}{\sigma^2} + \frac{a^2}{b^2} \right] da} \]

\[ a_1(x) = \sigma^2 \frac{d}{dx} \ln \left[ \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left[ \frac{(x-a)^2}{\sigma^2} + \frac{a^2}{b^2} \right]} da \right] = \sigma^2 \frac{d}{dx} \ln \left[ \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \sqrt{2\pi (b^2 + \sigma^2)} e^{-\frac{x^2}{2(b^2 + \sigma^2)}} \right] \]

\[ a_1(x) = \sigma^2 \frac{d}{dx} \left[ \frac{x^2}{2\sigma^2} - \frac{x^2}{2(b^2 + \sigma^2)} \right] = \left( \frac{b^2}{b^2 + \sigma^2} \right) x . \]

\[ D = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{b^2}{b^2 + \sigma^2} \right)^2 p_1(a)p_2(y)da dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{\sigma^2}{b^2 + \sigma^2} \right) a^2 + \left( \frac{b^2}{b^2 + \sigma^2} \right)^2 y^2 - 2 \left( \frac{b\sigma}{b^2 + \sigma^2} \right) a y \]

\[ -\frac{1}{2} \left( \frac{a^2}{b^2} + \frac{y^2}{\sigma^2} \right) da dy \]

\[ D = \left( \frac{\sigma^2}{b^2 + \sigma^2} \right)^2 b^2 + \left( \frac{b^2}{b^2 + \sigma^2} \right)^2 \sigma^2 + 0 = \frac{b^2 \sigma^2}{b^2 + \sigma^2} . \]
In what follows, omitted exponents in the denominator equal those in the numerator.

\[ \hat{a}_k(x) = \frac{\int a_k q(x|a)p(a)da}{\int q(x|a)p(a)da} \exp \left\{ -\frac{1}{2} \left[ \sum_{j=1}^{N} \sum_{k=1}^{M} \phi_{jk}^{-1} a_j a_k + \sum_{m=1}^{N} \sum_{n=1}^{N} \gamma_{mn}^{-1} (x_m - \sum_{j=1}^{N} c_{mj} a_j)(x_n - \sum_{k=1}^{N} c_{nk} x_k) \right] \right\} \]

\[ \hat{a}_k(x) = \frac{\int da a_k \exp \left\{ \frac{1}{2} \left[ 2 \sum_{m=1}^{M} x_m \gamma_{mn}^{-1} c_{nk} a_k - \sum_{j=1}^{N} \sum_{k=1}^{M} \phi_{jk}^{-1} a_j a_k - \sum_{m=1}^{M} \sum_{n=1}^{N} \gamma_{mn}^{-1} c_{mn} a_k \right] \right\}}{\int da \exp \left\{ \right\} \]

Let \( C \) be the \( M \times N \) rectangular matrix whose \( n-k \)th element is \( c_{nk} \), and consider the matrix \( \Phi^{-1} + C^T \Gamma^{-1} C = \Lambda \). Since both \( \Phi \) and \( \Gamma \) are symmetric and strictly positive definite, one easily verifies that \( \Lambda \) is also. We perform a transformation to the principal axes of the quadratic form in the exponent by defining the new coordinates \( \hat{a} \) according to the matrix relation \( a = \Lambda \hat{a} \), where \( \Lambda = (a_{jk}) \) is the matrix of normalized eigenvectors of \( \Lambda \). Denoting the eigenvalues of \( \Lambda \) by \( \nu_\ell \); \( \ell = 1, 2, \ldots, N \), we obtain
\[
\hat{a}_i(x) = \frac{\int d\beta \left( \sum_{l=1}^{N} a_{il} \beta_l \right) \exp \left\{ \frac{-1}{2} x^T \Gamma^{-1} CA - \frac{1}{2} \sum_{l=1}^{N} \nu_l \beta_l^2 \right\}}{\int d\beta \exp \left\{ \frac{-1}{2} x^T \Gamma^{-1} CA \right\}}
\]

Defining \( \xi^T = x^T \Gamma^{-1} CA \), we can express \( \hat{a}_i(x) \) in the form

\[
\hat{a}_i(x) = \sum_{l=1}^{N} a_{il} \left( \frac{\int \beta_l e^{\xi_l \beta_l - \nu_l \beta_l^2} d\beta_l}{\int e^{\xi_l \beta_l - \nu_l \beta_l^2} d\beta_l} \right) = \sum_{l=1}^{N} \frac{a_{il} \xi_l}{\nu_l}.
\]

Since \( \xi_l = \sum_{m=1}^{M} x_m \sum_{n=1}^{M} \gamma_{mn}^{-1} \sum_{k=1}^{N} c_{nk} a_{kl} \), this becomes

\[
\hat{a}_i(x) = \sum_{m=1}^{M} \theta_{im} x_m, \quad \text{where } \theta_{im} = \sum_{n=1}^{M} \gamma_{mn}^{-1} \sum_{k=1}^{N} c_{nk} \sum_{l=1}^{N} \frac{a_{il} a_{kl}}{\nu_l}.
\]
If we make the replacements \( c_{mj} = \xi_{mj} \) and \( \gamma_{mn}^{-1} = \frac{\delta_{mn}}{\lambda_m} \) as discussed after Eq. (29), the expression for \( \theta_{im} \) as given at the end of Appendix II becomes

\[
\theta_{im} = \sum_{n=1}^{M} \frac{\delta_{mn}}{\lambda_m} \sum_{k=1}^{N} \xi_{nk} \sum_{l=1}^{N} \frac{a_{il}a_{kl}}{\nu_l} = \frac{1}{\lambda_m} \sum_{k=1}^{N} \xi_{mk} \sum_{l=1}^{N} \frac{a_{il}a_{kl}}{\nu_l} .
\]

Combining this with Eq. (30) and the definition of \( \xi_{mk} \) from Eq. (28), we obtain

\[
\lambda_m \theta_{im} = \int_{T_1}^{T_2} f_m(t) \theta_1(t) \, dt = \sum_{k=1}^{N} \xi_{mk} \sum_{l=1}^{N} \frac{a_{il}a_{kl}}{\nu_l} = \sum_{k=1}^{N} \int_{T_1}^{T_2} f_m(t) r(t-\tau_k) \, dt \sum_{k=1}^{N} \frac{a_{il}a_{kl}}{\nu_l} .
\]

This may be rearranged in the form

\[
\int_{T_1}^{T_2} f_m(t) \left[ \theta_1(t) - \sum_{k=1}^{N} \sum_{l=1}^{N} \frac{a_{il}a_{kl}}{\nu_l} r(t-\tau_k) \right] \, dt = 0 .
\]

From the completeness of the \( f_m(t) \) on \((T_1, T_2)\) we may conclude that

\[
\theta_1(t) = \sum_{k=1}^{N} \sum_{l=1}^{N} \frac{a_{il}a_{kl}}{\nu_l} r(t-\tau_k) ; \quad T_1 < t < T_2 .
\]

Writing \( \gamma(t, s) = \psi(t-s) \) because the noise is stationary, and letting

\( T_1 \to -\infty \) and \( T_2 \to \infty \), we may write the integral equation (31) in the form

\[
\sum_{k=1}^{N} \sum_{l=1}^{N} \frac{a_{il}a_{kl}}{\nu_l} r(t-\tau_k) = \int_{-\infty}^{\infty} \psi(t-s) z_1(s) \, ds ; \quad t \in (-\infty, \infty) .
\]
Fourier transforming then yields
\[ \sum_{k=1}^{N} \sum_{l=1}^{N} \frac{a_1 a_{kl}}{\nu_l} R(f) e^{-j2\pi f T_k} \phi(f) = \psi(f) Z_1(f), \]

Since \( W_i(f) = Z_1^*(f) e^{-j2\pi f T} \) from (33), and since \( \psi(f) = \psi^*(f) \), the solution for the transfer function of the filter to be sampled at \( t = T \) in order to obtain \( g_i \) is
\[ W_i(f) = \frac{R^*(f)}{\psi(f)} \sum_{k=1}^{N} \left( \sum_{l=1}^{N} \frac{a_1 a_{kl}}{\nu_l} \right) e^{j2\pi f (T_k - T)}. \]

Accordingly, the optimum receiver is a matched filter followed by a tapped delay line, agreeing with Tufts's result as given in Eq. (35). The tap gain coefficients are given by \( \sum_{l=1}^{N} \frac{a_1 a_{kl}}{\nu_l} \), and it remains to establish that these are equivalent to the \( d(i,k) \) given by Tufts. Toward this end, let us recall from Appendix II that \( \Delta \) is the matrix of normalized eigenvectors of \( \Lambda \), so that we have \( \Delta^{-1} = \Delta^T \), and we may write
\[ \Delta = \Delta \begin{pmatrix} \nu_1 & 0 \\ \nu_2 & \vdots \\ 0 & \nu_N \end{pmatrix} \Delta^T; \quad \Delta^{-1} = \Delta \begin{pmatrix} 1/\nu_1 & 0 \\ 1/\nu_2 & \vdots \\ 0 & 1/\nu_N \end{pmatrix} \Delta^T. \]
Accordingly, the $i$-$k$th element of $\Lambda^{-1}$ is
\[ \sum_{l=1}^{N} \frac{a_{i,l} a_{k,l}}{v_{l}}, \]
which shows that the tap gains for the optimum receiver prescribed by the present theory are the elements of the inverse of $\Lambda = \Phi^{-1} + C^T \Gamma^{-1} C$. On the other hand, temporarily adopting Tufts's notation, we obtain by combining his Eqs. (7) and (15) the following representation for the matrix $D$ of his tap gain coefficients $d(i,k)$
\[
D = M - C = M - MQ (M^{-1} + Q)^{-1} = M - [M - (M^{-1} + Q)^{-1}] = (M^{-1} + Q)^{-1}.
\]
Accordingly, Tufts's tap gains are the elements of the inverse of $M^{-1} + Q$.

Since his $M$ is our $\Phi$, it only remains to show that his matrix $Q$ is the same as our matrix $C^T \Gamma^{-1} C$. Now the $i$-$k$th element of $Q$ is defined, in our notation, as
\[
q_{ik} = q(\tau_i - \tau_k) = \int_{-\infty}^{\infty} \frac{R(f) R^*(f)}{\Psi(f)} e^{j2\pi f(\tau_i - \tau_k)} df.
\]

We now study the behavior of $C^T \Gamma^{-1} C$ in the limit $T_1 \to -\infty$, $T_2 \to \infty$, and $M \to \infty$. First, we have $c_{mj} = \xi_{mj}$, and $\gamma_m^{-1} = \frac{\delta_m}{\lambda_m}$ because we are using Karhunen-Loève expansion coefficients, so that the $i$-$k$th element of $C^T \Gamma^{-1} C$ is
\[
\sum_{m=1}^{M} \sum_{n=1}^{M} \xi_{mi} \frac{\delta_m}{\lambda_m} \xi_{nk} = \sum_{m=1}^{M} \xi_{mi} \xi_{mk} \frac{\delta_m}{\lambda_m} = \sum_{m=1}^{M} \frac{\left( \int_{T_1}^{T_2} f_m(t) r(t-\tau_i) dt \right) \left( \int_{T_1}^{T_2} f_m(t) r(t-\tau_k) dt \right)}{\lambda_m}.
\]
It is well known ([3], p. 93 or [4], p. 167) that the Fourier coefficients of a stationary, Gaussian process become statistically independent in the limit of an infinitely long expansion interval. Accordingly, in this limit we may use as our orthonormal functions \( f_m(t) \) the trigonometric functions

\[
f_m(t) = \frac{e^{-j2\pi mt}}{(T_2-T_1)^{1/2}} ; \quad m = 0, \pm 1, \pm 2, \ldots
\]

Because these functions are complex, some minor complications arise. In particular, defining \( b_{mi} = \frac{\xi_{mi}}{\lambda m} \), we had in the real case that

\[
C^T \Gamma^{-1} C = B^T B ;
\]

in the complex case, this becomes \( C^T \Gamma^{-1} C = B^\dagger B \), where the "\( \dagger \)" indicates conjugate transpose. Moreover, summations over the index \( m \) now run from \(-M\) to \( M\) rather than from 1 to \( M\). Our expression for the \( i-k^{th} \) element of \( C^T \Gamma^{-1} C \), therefore, becomes

\[
\sum_{m=-M}^{M} \left( \int_{T_1}^{T_2} f_m^*(t) r(t-\tau_i) \, dt \right) \left( \int_{T_1}^{T_2} f_m(t) r(t-\tau_k) \, dt \right) \lambda_k
\]

Substituting the trigonometric functions, this assumes the form
We now take our limit in such a way that as \((T_2 - T_1)\) and \(m\) approach infinity, their ratio \(f\), representing frequency in cps, remains finite. In this limit, \(\frac{1}{T_2 - T_1}\) becomes the frequency differential \(df\) and \(\lambda_m\) becomes \(\Psi(f)\) ([4], p. 168). The above sum, therefore, tends to the integral

\[
\sum_{m=-M}^{M} \frac{1}{(T_2 - T_1)} \left( \int_{T_1}^{T_2} r(t) e^{j2\pi ft} dt \right) \left( \int_{T_1}^{T_2} r(t) e^{-j2\pi ft} dt \right) \lambda_m \]

This completes the demonstration of the fact that the present theory is in accord with Tufts's earlier work.
REFERENCES


By generalizing a trivial problem in estimation theory, some interesting results have been obtained for a class of pulse amplitude modulation (PAM) communication channels. In particular, we consider a linear channel with memory perturbed by additive noise which is uncorrelated with the signal statistics. It is shown that, for specified second-order statistics and a mean square error criterion, the Bayes estimation rule for the specific case of zero-mean Gaussian signal and noise statistics is also both the best linear estimation rule and the minimax estimation rule. This fact is used in the theoretical design of sampled-data and continuous time receivers for such PAM systems. In the continuous time case with an infinite observation interval and stationary noise statistics, the receiver thus designed agrees with the optimum linear receiver previously derived for this case by Tufts using a very different approach; the present analysis shows that Tufts's linear receiver is, therefore, also the best nonlinear receiver in the minimax sense.
Pulse amplitude modulation
Mean square error
Continuous-time optimum receiver
Sampled-data optimum receiver
Optimum linear estimation rule
Bayes estimation rule
Minimax estimation rule

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICER: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TB), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical content. The assignment of limits, roles, and weights is optional.