This dissertation has been microfilmed exactly as received

MARTINO, Joseph Paul, 1931-
MINIMUM-EXPECTED-COST FILTERS FOR ARBITRARY COST FUNCTIONS.

The Ohio State University, Ph.D., 1961
Mathematics

University Microfilms, Inc., Ann Arbor, Michigan
MINIMUM - EXPECTED - COST FILTERS
FOR ARBITRARY COST FUNCTIONS

DISSertation

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate School
of The Ohio State University

By

JOSEPH PAUL MARTINO, B.A., M.S.

The Ohio State University

1961

Approved by

[Signature]

Adviser
Department of Mathematics
ACKNOWLEDGMENTS

My thanks go to a large number of people who made this dissertation possible. First of all my thanks go to Prof. D. R. Whitney, who was willing to put up with a long-distance candidate during my period of off-campus research. I also wish to thank my colleagues in the Directorate of Mathematical Sciences, Air Force Office of Scientific Research, for the consideration given me in arranging time for study and research. I am grateful to the staff of Air Force Library 2819, who never ceased to amaze me with their ability to find obscure references. I also acknowledge my debt to Prof. H. B. Mann, whose influence is obvious on nearly every page. And finally I acknowledge the contribution of my wife, without whose support and encouragement I could never have finished.
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>The Problem</td>
<td>5</td>
</tr>
<tr>
<td>Necessary Theorems</td>
<td>6</td>
</tr>
<tr>
<td>Cost Function Approximation</td>
<td>10</td>
</tr>
<tr>
<td>Conclusion</td>
<td>12</td>
</tr>
<tr>
<td>Appendix</td>
<td>13</td>
</tr>
<tr>
<td>Bibliography</td>
<td>22</td>
</tr>
<tr>
<td>Autobiography</td>
<td>23</td>
</tr>
</tbody>
</table>
FIGURES

Table 1 .... 20
Graph I .... 21
I. Introduction

A problem common to a number of fields can be described as follows: A message is sent from a transmitter, over a noisy channel, to a receiver where the signal and noise are received together. The problem then is to determine what signal was sent, based on the signal plus noise received.

Presumably the signals are drawn from a population containing at least two members, for otherwise there would be no problem in determining what message was sent. The signal population may, of course, be infinite.

It must also be assumed that there is some cost associated with making an error in the determination of the transmitted signal. If an error does not cost anything, there is no point in receiving the message. The simplest way to handle the problem is to turn the receiver off. Thus the problem is to make the determination of the original signal in a way which, in some sense, minimizes the cost.

One plausible approach is to minimize the expected cost of error. However, it is clear that this is not the only approach. Under certain circumstances, it might be better to minimize the maximum possible cost, the expected cost subject to the constraint that the maximum cost never exceeds a certain limit. Finally, to consider an extreme case, assume that an Early Warning station can send either of two messages: "We are being attacked" or "We are not being attacked." It is quite difficult to assign a cost to the error of mistaking either of these signals for the other, and it is doubtful that the minimum-expected-cost criterion is applicable to this case. However, it is applicable to a wide variety of interesting cases; hence, despite its limitations it deserves consideration.

To talk about expected costs, it must be assumed that probability distributions exist over the signal and noise populations (this need not always be true, as for instance in the example of the previous paragraph). Hence the signal and noise can be considered as stochastic processes. In particular, they are processes dependent
on a continuous parameter -- time. It is assumed that the probability distributions over the signal and noise populations are not completely known, but that the first few joint moments are known, either experimentally or on some a-priori basis.

Before investigating the minimum-expected-cost approach, however, a brief survey of some techniques in current use is desirable.

The most common approach views the problem in terms of separating signal from noise. This usually involves designing filters to pass the bands of frequencies where the signal is strong, and to block the bands where the signal is weak. Most of the design effort goes into such problems as sharp cutoff at the band edges, flat response within the passband, etc. Passband edges are placed mostly by cleverness. This method is most effective against unintentional man-made interference, such as eliminating power-line hum from telephone lines. It has little effectiveness against naturally occurring noise within the signal band, and is completely ineffective against deliberate jamming.

A more sophisticated approach has come into use in the recent past. This begins by defining the error as

\[ e(t) = \int_{-\infty}^{\infty} \phi(T) x_{t-T} dT - \int_{-\infty}^{\infty} f(T) (x_{t-T} + y_{t-T}) dT \]

where \( x_t \) is the signal, \( y_t \) the noise, \( \phi(t) \) a filter representing some desired processing of the signal, and \( f(t) \) is the actual filter.

Next it is asserted that some "norm" or measure of the error is required. Plausible requirements for this norm are that it be positive and that it increase both with increasing \( e \) for \( e > 0 \) and with decreasing \( e \) for \( e < 0 \). It is observed that \( e^2 \) satisfies these requirements. Then the conclusion is that \( f(t) \) should be chosen to minimize \( \bar{e}^2 \), the mean-square error. The expression for \( e \) is squared, and its average taken:

\[ \bar{e}^2 = \lim_{L \to \infty} \frac{1}{L} \int_{-L}^{L} e^2 \, de \]
After some manipulation, including the use of Parseval's Relation, this expression is obtained:

\[ F_{\text{opt}}(j\omega) = \frac{i_{ss}(j\omega)}{i_{ss}(j\omega) + i_{nn}(j\omega)} \Phi(j\omega) \]

where \( i_{ss}(j\omega) \) is the Fourier Transform of the signal covariance; \( i_{nn}(j\omega) \) is the Fourier Transform of the noise covariance; \( \Phi(j\omega) \) is the Fourier Transform of the desired filter function; and \( F_{\text{opt}}(j\omega) \) is the Fourier Transform of the actual filter function. It should be noted that at one step in the development of (2), it is required that both signal and noise be stationary.

The filter obtained as above is commonly known as a Wiener Optimum Filter, or simply an Optimum Filter. It was originally devised by N. L. Wiener, and its theory presented in (12). The development given above is a highly condensed version of that given in (12).

The development of Wiener's filtering theory was a tremendous step forward in communication theory. Its consideration of signal and noise as random processes was a sharp break with previous practice. It had so much influence that practically every author of a modern electrical engineering text on the subject of filter theory finds it necessary to include a chapter or two on probability theory (see (6), (2), (10) for instance).

This development had great influence in another direction as well. Almost all effort in problems of statistical design of filters is directed to minimum mean-square-error filters. In (6), for instance, Y. W. Lee devotes only 8 pages out of 501 to consideration of a filter criterion other than minimum mean-square-error, and this is a special one, still requiring only the second joint moments of signal and noise. The remaining pages are devoted to Optimum Filters and related subjects (probability theory, orthonormal functions, the Wiener-Hopf integral equation, etc.). The same holds true for Davenport and Root (2). Even Doob (3), who devotes an entire chapter to linear prediction of stochastic processes
(i.e. the filter output is a prediction of what the signal will be some fixed \( T_0 \) hence), considers only linear least squares prediction, and in the Preface he acknowledges the personal influence of Wiener on his work. One of the most extensive presentations of other than minimum mean-square-error criteria is given by Laning and Battin in (5), who devote an entire chapter to this subject.

One of the chief reasons given for this avoidance of non-minimum-mean-square-error criteria is the difficulty of using them (6), p 448, for instance). The validity of this criticism will be seen below. On the other hand, as shown by eqn. 2, the minimum mean-square-error criterion is comparatively easy to use. It may be asked, then, can an Optimum Filter be used in all cases?

Truxal (10), after pointing out the drawbacks to the Optimum Filter, points out one important advantage:

"It fills an important gap in the general theory of feedback control systems by providing a straightforward design procedure which considers the actual input functions"

(p 415)

Furthermore, as shown by Pugachev (9), if both signal and noise are normal, and the cost function \( C(e) \) is convex, on a steady-state basis \( F_{\text{opt}}(j\omega) \) is the true minimum-expected-cost filter, regardless of the exact shape of \( C(e) \).

However, it was assumed above that the minimum-expected-cost filter was the one genuinely desired. Thus the possibility of non-normal signal and noise, and the possibility of non-convex cost functions, must be considered. For instance, \( C(e) \) may reach some maximum value for finite \(|e|\), and remain constant for \(|e|\) larger than this. Despite the plausible arguments about finding a "norm" for the error, what the Optimum Filter does is approximate the true \( C(e) \) by the parabola \( C^*(e) = e^2 \). In particular cases, this approximation may be a poor one. Thus \( F_{\text{opt}}(j\omega) \) may be an easy-to-compute wrong answer, if one is really interested in minimum expected cost.
II. The Problem

Let the signal be a stochastic process $x_t$, and the noise a stochastic process $y_t$. Both are assumed to be continuous in probability and bounded (i.e., there is a $k > 0$ such that $P(|x_t| > k) = P(|y_t| > k) = 0$). Furthermore, both are assumed to start at $t = 0$.

Instead of merely recovering the signal at the receiver, the possibility of operating on it is included. This desired operation will be represented by the filter function $g(t)$. The actual operation on the signal (and noise) will be represented by the filter function $f(t)$. The error is defined to be

$$e = \int_0^t g(t - T)x_T dT - \int_0^t f(t - T)x_T dT - \int_0^t f(t - T)y_T dT.$$ 

Here the first term is the desired output from the receiver, and the sum of the last two is the actual output.

Let $F(e)$ be the distribution function for the error $e$, and $C(e)$ the cost function giving cost of error for the specific application for which the filter $f(t)$ is intended. The object then is to choose $f(t)$ so as to minimize

$$E[C(e)] = \int_{-\infty}^{\infty} C(e) dF(e).$$

Since by assumption $F(e)$ is not known completely, the best that can be done is some approximation to the minimizing $f(t)$. There are basically two alternatives. One is to use the partial information about the distributions of $x_t$ and $y_t$ to approximate $F(e)$. The other is to approximate $C(e)$ by a function $C^*(e)$ such that $E[C(e)]$ can be calculated by using the partial information available. The first alternative is not only difficult, but in some cases it is impossible. The second alternative will be shown to be possible, although tedious from a computational standpoint.

5
III. Necessary Theorems

Theorem I. Let $x_t$ be a process with a covariance function $\sigma_{t_1t_2}$. The process is integrable l.i.m. if, for any $t$ in $[a, b]$, 
\[ \int_a^b \sigma_{t_1t_2} \, dt_1 dt_2 \] 
exists. The proof is given in (8) as Theorem 1.8.

Theorem II. If $w(x, y, \ldots, z)$ is a continuous function of the set of variables $x, y, \ldots, z$ and if $x_n, y_n, \ldots, z_n$ are random variables converging in probability to $x_0, y_0, \ldots, z_0$ respectively, $w(x_n, y_n, \ldots, z_n)$ converges in probability to $w(x_0, y_0, \ldots, z_0)$. The proof is given in (4), p. 191.

Theorem III. Let $X_n$ be a sequence of random variables converging in probability to $X$. Let there be a $K > 0$ such that $P(|X| > K) = P(|X_n| > K) = 0$. Let $g(x)$ be a continuous function. Then $E[g(X_n)]$ converges to $E[g(X)]$. This theorem is proven in (4), p. 193.

Corollary. Let $X_n$ be a uniformly bounded sequence of random variables converging in probability to $X$. Then $E(X_n)$ converges to $E(X)$. This follows from Theorem III since $g(x) = x$ is continuous.

Theorem IV. Let $x_t$ have a continuous covariance function $\sigma_{tt}$ on $[a, b]$. Let $g(t)$ be a continuous function on $[a, b]$. The $\int_0^t g(T)x_T \, dT$ exists l.i.m. for $t$ in $[a, b]$.

Proof. Let $z(t) = g(t)x_t$. Consider 
\[ E(z_Tz_{T'}) = E[g(t)x_tg(t')x_{t'}] = g(t)g(t') E(x_tx_{t'}) = g(t)g(t') \sigma_{tt} \] 
This is a continuous function; hence $\int \int g(T)g(T')\sigma_{TT'} \, dTdT'$ exists, and by Theorem I above, $\int_0^t g(T)x_T \, dT$ exists l.i.m.

Theorem V. Let $x_t$ be a bounded stochastic process which is continuous in probability. The $r$th joint moment $E(x_{t_1}x_{t_2}\ldots x_{t_r})$ is continuous.
Proof. First, the rth joint moment exists by the boundedness of $x_t$.
Consider $r$ points $t_1, t_2, \ldots, t_r$. For any sequences $t^k_i$ converging to $t_i$, by Theorem II above, $x_{t_1}^{k_1} \ldots x_{t_r}^{k_r}$ converges in probability to $x_{t_1} \ldots x_{t_r}$. Then by Theorem III above, $E(x_{t_1}^{k_1} \ldots x_{t_r}^{k_r})$ converges to $E(x_{t_1} \ldots x_{t_r})$.

**Theorem VI.** Let $x_1^t, x_2^t, \ldots, x_n^t$ be $n$ bounded stochastic processes which are integrable in probability in $0 \leq t < b$, and such that the joint moment $E(x_1^1 x_2^2 \ldots x_n^n)$ exists and is integrable. Then
\[
E\left(\int_0^t x_1^1 dT \left(\int_0^t x_2^2 dT\right) \ldots \left(\int_0^t x_n^n dT\right)\right) = \int_0^t E(x_1^1 x_2^2 \ldots x_n^n) dT_1 \ldots dT_n.
\]

Proof. Choose a subdivision $S_k$ of $[0, t]$:

$0 = t_1 < t_2 < \ldots < t_k = t$. Choose $t_i^*$ such that $t_{i-1} < t_i^* < t_i$, $i > 1$. Form $X_k^1 = \sum_{i=2}^k x_{t_i}^{t_{i-1}}$. By hypothesis $X_k^1$ converges in probability to $X^1 = \int_0^t x_1^1 dT$ as the modulus of $S_k$ converges to 0.

Then by Theorem II, $X_k^1 x_k^2 \ldots x_k^n$ converges in probability to
\[
\left(\int_0^t x_1^1 dT\right) \left(\int_0^t x_2^2 dT\right) \ldots \left(\int_0^t x_n^n dT\right).
\]
Now consider $E(X_k^1 x_k^2 \ldots x_k^n)$. By the hypothesis of integrability of $E(x_1^1 x_2^2 \ldots x_n^n)$, this converges to
\[
\int_0^t E(x_1^1 \ldots x_n^n) dT_1 \ldots dT_n.
\]
But by the corollary to Theorem III, $E(X_k^1 \ldots x_k^n)$ converges to $E(x_1^1 \ldots x_n^n)$. This completes the proof. It should be noted that this Theorem is a weaker form of the main Theorem of (1).

**Theorem VII.** Under the same hypotheses as Theorem VI,
\[
E\left(\int_0^t x_1^1 x_2^2 \ldots x_r^r \left(\int_0^{T_r} x_{T_r}^{r+1} dT\right) \ldots \left(\int_0^t x_n^n dT\right)\right) = \int_0^t E(x_1^1 x_2^2 \ldots x_r^{r+1} \ldots x_n^n) dT_1 \ldots dT_n,
\]
for $0 \leq t_1 \leq b$, $1 \leq i \leq r$. 
Proof. Form $X_k^j$, as above, for $r+1 < j < n$, and let $X_k^j$ equal $x^j_t$ for $1 < j < r$. Then the conclusion follows as in Theorem VI.

Theorem VIII. Let $x_t$ and $y_t$ be stochastic processes, bounded and continuous in probability. Designate them as signal and noise respectively. Let $\phi(t)$ be the desired signal filter and $f(t)$ the actual filter, both functions being continuous for $t > 0$. If the error is defined as in (3) above, then the $j$th joint moment of $e$ at time $t > 0$ is given by

$$
\sum_{r+s+u=j} \frac{(-1)^{s+u}}{r! s! u!} \int_0^t \phi(t-T_i) \int_{i=r+1}^{r+s+u} f(t-T) E(z_{T_1} \ldots z_{T_j}) \, dT_i
$$

where $z_{T_1} = x_{T_1}$ for $0 < i \leq r + s$

$z_{T_1} = y_{T_1}$ for $r + s + 1 < i < j$.

Proof. Since $x_t$ and $y_t$ are continuous in probability, by Theorem VI their covariances are continuous. Hence by Theorem V, $\int_0^t \phi(t-T)x_{T}dT$, $\int_0^t f(t-T)x_{T}dT$, and $\int_0^t f(t-T)y_{T}dT$ exist l.m. Now expanding $e^j$ by the binomial formula,

$$
e^j = \sum_{r+s+u=j} \frac{(-1)^{s+u}}{r! s! u!} \int_0^t \phi(t-T)x_{T}dT \int_{i=r+1}^{r+s+u} f(t-T) x_{T}dT \int_{j=r+s+1}^{r+s+u} f(t-T) y_{T}dT$$

Thus the hypotheses of Theorem VI are satisfied, and the conclusion follows.

Theorem IX. If $x_t$ is to be recovered only, not processed in any fashion (i.e., $\phi(t)$ is the Dirac delta function), then the $j$th moment of $e$ is

$$
\sum_{r+s+u=j} \frac{(-1)^{s+u}}{r! s! u!} \int_0^t \int_{i=r+1}^{r+s+u} f(t-T) E(z_{T_1} \ldots z_{T_j}) dT_1 \ldots dT_j
$$

where $z_{T_1} = x_t$, $0 \leq i < r$

$z_{T_1} = x_{T_1}$, $r \leq i < r+s$

$z_{T_1} = y_{T_1}$, $r+s \leq i < j$. 
Proof. This follows from Theorem VII instead of Theorem VI, in the same manner as Theorem VIII.
IV. Cost Function Approximation

By the use of Theorems VIII and IX it is possible to express the moments of \( e \) in terms of joint moments of \( x_t \) and \( y_t \), and in terms of \( \phi(t) \) and \( f(t) \). Thus if the first \( n \) joint moments of \( x_t \) and \( y_t \) are known, it is possible to find the expected value of a polynomial in \( e \), of degree \( m \leq n \). If this polynomial represents an approximation \( C^*(e) \) to \( C(e) \), the cost of error, then the \( f(t) \) minimizing \( E[C^*(e)] \) would be an approximation to the \( f(t) \) minimizing \( E[C(e)] \). In fact, if \( C(e) \) were a polynomial of degree \( m \leq n \), the \( f(t) \) minimizing \( E[C(e)] \) could be found exactly.

There is a further consideration, however. As can be seen from the expressions for \( E(e^j) \) in Theorems VIII and IX, an attempt to find an \( f(t) \) minimizing \( E[C^*(e)] \), for example by using variational techniques, would lead to an extremely complex integral equation. This would be very difficult to solve if it could be solved at all. Furthermore, should an \( f(t) \) be obtained in this fashion, it may be too complicated to synthesize in actual hardware.

However, since the \( f(t) \) to be obtained is an approximation to the desired (but unobtainable) \( f(t) \) which minimizes \( E[C(e)] \), further approximation is not fatal. \( f(t) \) will be the sum of a finite number of exponentials and sinusoids in \( t \) multiplied by polynomials in \( t \). Thus a form for \( f(t) \) can be assumed, such as

\[
f(t) = \sum_{i=1}^{r} a_i \exp(b_i t) \sin(c_i t + d_i) + \sum_{i=1}^{s} a_i \exp(b_i t)
\]

where the \( a_i, b_i \) etc., are parameters to be adjusted to minimize \( E[C^*(e)] \).

The next question is that of finding a polynomial approximation to \( C(e) \). If \( F(e) \) is the (unknown) distribution function for the error, then the best \( C^*(e) \) is the one which minimizes

\[
\int_{-\infty}^{\infty} |C(e) - C^*(e)| dF(e)
\]

Since \( F(e) \) is unknown, this best approximation cannot be obtained. This leaves no choice but to look for a "good" approximation.
There are many methods for approximating a function by a polynomial, such as approximation in the mean of order $p$ or the use of orthogonal polynomials. It must be emphasized that adopting any of these as "the" method for finding $C^*(e)$ amounts to assuming a form for $P(e)$, and it is one of the purposes of this procedure to avoid such assumptions. The real question is how to check the goodness of fit of the chosen $C^*(e)$ to $C(e)$. The general idea is that $|C^*(e) - C(e)|$ should be small when the probability density of $e$ is large. Fortunately, there are some concrete guides available.

First, it may be possible to place a bound on the error. If so, then the error of approximation for values of $e$ in excess of the bound are irrelevant. Secondly, Tchebycheff's formula

$$P(|e| \geq K) \leq E(e^2)/K^2,$$

can be used to determine the relative importance of different intervals of values of $e$, if all moments through the second are available. If moments through the fourth are available, a refinement of the Tchebycheff formula may be used (F, p. 199):

$$P(|e| \geq K) \leq E(e^2)/K^2 \text{ for } E(e^2) \leq K^2 \leq E(e^4)/E(e^2)$$

$$P(|e| \geq K) \leq \frac{E(e^4) - E(e^2)}{E(e^4) + K^4 - 2K^2E(e^2)} \text{ for } K^2 < \frac{E(e^4)}{E(e^2)}.$$

Additional refinements of the Tchebycheff formula can be made if moments of order $2n$ are known.

The use of these formulas, and of bounds on $|e|$, permits use of most, if not all, of the information available on the moments of $e$ to determine the goodness of fit of $C^*(e)$ to $C(e)$.

Thus an iterative process can be used to obtain the minimizing $f(t)$. For a given $C^*(e)$, the minimizing $f(t)$ is found; then the resulting moments of $e$ are used to check the goodness of fit of $C^*(e)$. If the $C^*(e)$ turns out to have too much error in the wrong places, it can be adjusted and a new $f(t)$ determined. This process is then repeated until a satisfactory $C^*(e)$ is obtained and the minimizing $f(t)$ for it accepted.
V. Conclusion

The procedure developed above gives a method for minimizing $E[C^*(e)]$, where $C^*(e)$ is an approximation to $C(e)$. The major advantages it has over alternative procedures are that it permits full use of all information available on joint moments of signal and noise and it contains no hidden assumptions. Its major defect is its complexity.
APPENDIX

Let the following problem be given:
The signal, $x_t$, is a random process. Its values are drawn from $[-1,1]$ with uniform density. The signal remains constant for some period, then jumps to some other value. The value after a jump is independent of the value prior to the jump. The probability of no jump between $t_1$ and $t_2$ is $\exp(-|t_1 - t_2|)$. Then the first four joint moments are:

\[
E(x_t) = 0
\]
\[
E(x_{t_1}x_{t_2}) = \frac{1}{3} \exp(t_1 - t_2), \ t_2 \geq t_1
\]
\[
E(x_{t_1}x_{t_2}x_{t_3}) = \frac{1}{3} \exp(t_1 - t_3), \ t_3 \geq t_2 \geq t_1
\]
\[
E(x_{t_1}x_{t_2}x_{t_3}x_{t_4}) = \frac{1}{9} \exp(t_1 - t_2 + t_3 - t_4) + \frac{4}{45} \exp(t_1 - t_4)
\]

The noise, $y_t$, is a random process. Its values are drawn from $[-1,1]$ with uniform density. The value of $y_t$ remains constant for some period, then jumps to some other value. The value after the jump is independent of the value prior to the jump. The probability of no jump between $t_1$ and $t_2$ is $\exp(-2|t_1 - t_2|)$. Then the first four joint moments are:

\[
E(y_t) = 0
\]
\[
E(y_{t_1}y_{t_2}) = \frac{1}{3} \exp[-2(t_2 - t_1)], \ t_2 \geq t_1
\]
\[
E(y_{t_1}y_{t_2}y_{t_3}) = \frac{1}{3} \exp[-2(t_3 - t_1)], \ t_3 \geq t_2 \geq t_1
\]
\[
E(y_{t_1}y_{t_2}y_{t_3}y_{t_4}) = \frac{1}{9} \exp[-2(t_4 - t_3 + t_2 - t_1)] +
\]
\[
+ \frac{4}{45} \exp[-2(t_4 - t_1)]
\]

The signal and noise are independent processes.
It is desired only to recover the signal, not to process it. Furthermore, the use to which the signal is put assures that it will never exceed 0.6 seconds. (An instance of this might be a fire-control system which, because of target speed, will never be able to track a target for longer than 0.6 seconds.)

The cost function \( C(e) = \begin{cases} 0 & |e| < 1 \\ 1 & |e| \geq 1. \end{cases} \)

The filter function is to be of the form \( \exp(bt) \), where \( b \) is to be selected to minimize \( E[C(e)] \). By consideration of the nature of the problem, it is clear that the solution cannot have \( b \geq 0 \), since this corresponds to a growing output, and the output should remain bounded. Thus \( b < 0 \).

Next a \( C^*(e) \) is needed. Since \( b < 0 \) in \( f(t) \), \( |f(t)| \leq 1 \). Furthermore, \( |x_t| \) and \( |y_t| \leq 1 \). Making these substitutions in the expression for error, and restricting \( 0 \leq t \leq 0.6 \), it is found that \( |e| \leq 2.2 \). Hence the behavior of \( C^*(e) \) for \( |e| > 2.2 \) is irrelevant.

As a start, suppose \( C^*(e) \) is to have values as follows:

<table>
<thead>
<tr>
<th>( e )</th>
<th>0</th>
<th>.75</th>
<th>1.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C^*(e) )</td>
<td>0.1</td>
<td>0.1</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Furthermore, \( C^*(e) \) will contain only even powers through the fourth. Solving for the coefficients,

\[
C^*(e) = 0.1 - 0.32e^2 + 0.58e^4.
\]

The minimum value of this is 0.057, occurring at \( e = 0.522 \).

Although this is a fairly good fit in the region \( 0 \leq e \leq 1 \), it can be improved by reducing the constant term to 0.5. Thus the \( C^*(e) \) becomes

\[
C^*(e) = 0.05 - 0.32e^2 + 0.58e^4,
\]
which takes on the following values:

<table>
<thead>
<tr>
<th>e</th>
<th>0</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.25</th>
<th>1.5</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C^*(e)$</td>
<td>0.05</td>
<td>0.006</td>
<td>0.05</td>
<td>0.31</td>
<td>0.95</td>
<td>2.27</td>
<td>7.07</td>
<td>44.17</td>
</tr>
<tr>
<td>$C(e)$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Then by Theorem IX,

$$E[C^*(e)] = 0.05$$

$$-.32 \int_0^1 T_2 \exp[b(2t-T_2-T_1)] \exp(T_1-T_2) dT_1 dT_2$$

$$+ \frac{2}{3} \int_0^1 T_2 \exp[b(2t-T_2-T_1)] \exp(2T_1-2T_2) dT_1 dT_2$$

$$- \frac{2}{3} \int_0^1 \exp[b(t-T)] \exp(T-t) dT$$

$$+.58 \int_0^1 T_4 \int_0^1 T_3 \int_0^1 T_2 \exp[b(4t-T_4-T_3-T_2-T_1)]$$

$$\times \exp(-T_4+T_3-T_2+T_1) dT_1 \ldots dT_4$$

$$+ \frac{96}{45} \int_0^1 T_4 \int_0^1 T_3 \int_0^1 T_2 \exp[b(4t-T_4-T_3-T_2-T_1)] \exp(-T_4+T_1) dT_1 \ldots dT_4$$

$$+ \frac{24}{9} \int_0^1 T_4 \int_0^1 T_3 \int_0^1 T_2 \exp[b(4t-T_4-T_3-T_2-T_1)]$$

$$\times \exp(-2(T_4+T_3-T_2-T_1)) dT_1 \ldots dT_4$$

$$+ \frac{96}{45} \int_0^1 T_4 \int_0^1 T_3 \int_0^1 T_2 \exp[b(4t-T_4-T_3-T_2-T_1)]$$

$$\times \exp(-2(T_4-T_1)) dT_1 \ldots dT_4$$

$$+ \frac{4}{3} \int_0^1 T_2 \exp[b(2t-T_2-T_1)] \exp(T_1-T_2) dT_1 dT_2$$

$$+ \frac{16}{15} \int_0^1 T_2 \exp[b(2t-T_2-T_1)] \exp(-t+T_1) dT_1 dT_2$$

$$+ \frac{4}{3} \int_0^1 T_2 \exp[b(2t-T_2-T_1)] \exp(-2T_2+2T_1) dT_1 dT_2$$

$$- \frac{24}{9} \int_0^1 T_3 \int_0^1 T_2 \exp[b(3t-T_1-T_2-T_3)] \exp(-t+T_3-T_2+T_1) dT_1 dT_2 dT_3$$

$$- \frac{96}{45} \int_0^1 T_3 \int_0^1 T_2 \exp[b(3t-T_1-T_2-T_3)] \exp(-t+T_1) dT_1 dT_2 dT_3$$
\[- \frac{8}{3} \int_{0}^{t} \exp[b(t-T_1)]\exp(-t+T_1)dT_1 \cdot \int_{0}^{t} \int_{0}^{T_3} \exp[b(2t-T_3-T_2)]\exp(-2T_3+2T_2)dT_2dT_3 \]
\[+ \frac{8}{3} \int_{0}^{t} \int_{0}^{T_2} \exp[b(2t-T_2-T_1)]\exp(-T_2+T_1)dT_1dT_2 \cdot \int_{0}^{t} \int_{0}^{T_4} \exp[b(2t-T_4-T_3)]\exp(-2T_4+2T_3)dT_3dT_4 \]
\[- \frac{4}{5} \int_{0}^{t} \exp[b(t-T)]\exp(T-t)dT \].

This is clearly still a function of t. The next step is to determine E[C*(e)] for various values of b, and plot the results as functions of t. For selected negative values of b, the results are as follows:

b = -0.5

E[C*(e)] = .05

\[-.32[.599 - 1.777\exp(-t) + 1.332\exp(-1.5t) + .177\exp(-2.5t)]
+.58[1.848 - 9.717\exp(-t) + 5.15\exp(-1.5t) + (1.066 + 4.266t)\exp(-2t) + 4.147\exp(-2.5t)
-2.844\exp(-3t) - 1.829\exp(-3.5t) + 2.370\exp(-4t)] \]

b = -1

E[C*(e)] = .05

\[-.32[.227 - (.166 + .333t)\exp(-2t) + .222\exp(-3t)]
+.58[.326 - (.699 + .889t)\exp(-2t) + .666\exp(-3t) -(.736 - .767t)\exp(-4t) + (.648 - .444t)\exp(-5t)] \]
\[ b = -1.3 \]

\[ E[C^*(e)] = .05 \]

\[ -.32[.232 - .677\exp(-2.3t) + .488\exp(-2.6t) \]
\[ + .288\exp(-3.3t)] \]

\[ +.58[.172 - 1.102\exp(-2.3t) + .683\exp(-2.6t) \]
\[ + .402\exp(-3.3t) + .096\exp(-3.6t) + .457\exp(-.6t) \]
\[ + .972\exp(-4.9t) - 2.018\exp(-5.2t) - 1.171\exp(-5.6t) \]
\[ + 1.706\exp(-5.9t)] \]

\[ b = -2 \]

\[ E[C^*(e)] = .05 \]

\[ -.32[.208 + .167\exp(-2t) - (.042 + .167t)\exp(-4t)] \]

\[ +.58[.115 + .042\exp(-2t) - .107\exp(-3t) \]
\[ + (.116 - .205t)\exp(-4t) + .035\exp(-5t) \]
\[ + (.002 - .167t)\exp(-6t) + .004\exp(-7t) \]
\[ - (.005 + .027t)\exp(-8t)] \]

\[ b = -4 \]

\[ E[C^*(e)] = .05 \]

\[ -.32[.229 + .034\exp(-5t) + .068\exp(-8t)] \]

\[ +.58[.098 + .081\exp(-5t) - .075\exp(-6t) + .071\exp(-8t) \]
\[ + .015\exp(-9t) + .002\exp(-10t) - .030\exp(-11t) \]
\[ + .013\exp(-13t) - .001\exp(-14t) + .009\exp(-16t)] . \]

These are shown as functions of \( t \) in Table 1, and plotted in Graph I.

Note that in Table 1, the values for \( t = 0 \) should all be \[ .05 - .32(.333) + .58(.200) = .05944. \] Deviations from this value are due largely to round-off error in the computations.
Now for any b, $E[C^*(e)]$ consists of a constant plus a set of exponentials in $t$. For large $t$, $E[C^*(e)]$ is equal to the constant. In all cases but one shown in Graph I, $E[C^*(e)]$ passes through a minimum and then climbs to its final value. In the case of $b = -4$, however, it does not have a minimum, but simply drops to its final value.

From the statement of the problem, it appears that $b = -1$ is the correct choice, since this choice minimizes $E[C^*(e)]$ in the range $0 \leq t \leq 0.6$.

Had a different problem been given, for instance one requiring correct reception of the signal at all instants in the interval $0 \leq t \leq 4$, a different choice would be made for $b$, despite the fact that the signal and noise remain the same. In this case, interest would center on the final value of $E[C^*(e)]$. Table 1 shows that this is largest for $b = -0.5$, and decreases with decreasing $b$, becoming smallest for $b = -4$. A computation of the final value for $b = -10$ shows that it is 0.046. For even larger values of $b$, the final value will increase, approaching 0.05944 as $b$ approaches minus infinity (i.e., a filter with no "memory" at all). Thus in this case the choice would be $b = -4$.

As a final step, the choice of $C^*(e)$ should be checked by Tchebycheff's formula. In this case

$$P(|e| \geq K) \leq E(e^2)/K^2.$$  

The maximum value of $E(e^2)$ in $0 \leq t \leq 0.6$ is 0.333 when $b = -1$. Then $P(|e| \geq K) \leq 1/3K^2$. Thus

$$P(|e| \geq 1) \leq 0.333$$  

$$P(|e| \geq 1.5) \leq 0.147$$  

$$P(|e| \geq 2) \leq 0.083.$$  

Hence the chosen $C^*(e)$ is a satisfactory approximation to $C(e)$. 
This example illustrates the technique developed above and emphasizes the importance of designing a filter for the specific problem at hand rather than using some general rule, such as minimization of the mean square error. It further illustrates the value of knowledge of higher moments of signal and noise, since these in general permit the use of better approximations to $C(e)$.

Finally, this example illustrates how hard it is to construct a good example. As can be seen from the statement of the problem, $E[C*\{e\}] = 0$ if the receiver is shut off and $x_t$ is assumed to be zero for all $t$. 
<table>
<thead>
<tr>
<th>t</th>
<th>-0.5</th>
<th>-1</th>
<th>-1.3</th>
<th>-2</th>
<th>-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.058</td>
<td>0.062</td>
<td>0.058</td>
<td>0.061</td>
<td>0.050</td>
</tr>
<tr>
<td>0.1</td>
<td>0.042</td>
<td>0.022</td>
<td>0.043</td>
<td>0.037</td>
<td>0.040</td>
</tr>
<tr>
<td>0.5</td>
<td>0.054</td>
<td>0.011</td>
<td>0.043</td>
<td>0.031</td>
<td>0.035</td>
</tr>
<tr>
<td>0.7</td>
<td>0.091</td>
<td>0.035</td>
<td>0.044</td>
<td>0.037</td>
<td>0.034</td>
</tr>
<tr>
<td>1.0</td>
<td>0.159</td>
<td>0.064</td>
<td>0.058</td>
<td>0.043</td>
<td>0.034</td>
</tr>
<tr>
<td>1.5</td>
<td>0.290</td>
<td>0.106</td>
<td>0.070</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>0.488</td>
<td>0.130</td>
<td>0.073</td>
<td>0.050</td>
<td>0.034</td>
</tr>
<tr>
<td>3.0</td>
<td>0.751</td>
<td>0.147</td>
<td>0.075</td>
<td>0.050</td>
<td>0.034</td>
</tr>
<tr>
<td>4.0</td>
<td>0.847</td>
<td>0.150</td>
<td>0.076</td>
<td>0.050</td>
<td>0.034</td>
</tr>
</tbody>
</table>


I, Joseph Paul Martino, was born in Warren, Ohio, 16 July 1931. I received my secondary education at Howland High School, Warren. My undergraduate training was received at Miami University, Oxford, Ohio, which granted me the Bachelor of Arts degree in 1953. From Purdue University, I received the Master of Science degree in 1955. In September of 1958 I entered the Ohio State University, completing residence requirements for the degree Doctor of Philosophy in 1960.