The MMSE with the optimum linear processor is:

\[ J_p(t) = E \left\{ ( \hat{v}(t) - v(t))^2 \right\} \]

\[ = K_d(t, t) + E \int_{T_i}^{T_f} h(t, \tau_1) r(\tau_1) d\tau_1 \int_{T_i}^{T_f} h(t, \tau_2) r(\tau_2) d\tau_2 \]

\[ - z E \int_{T_i}^{T_f} h(t, \tau) r(\tau) d\tau \]

\[ = K_d(t, t) + \int_{T_i}^{T_f} \int_{T_i}^{T_f} h(t, \tau_1) h(t, \tau_2) K_p(\tau_1, \tau_2) d\tau_1 d\tau_2 \]

\[ - z \int_{T_i}^{T_f} h(t, \tau) K_d(t, \tau) d\tau \]

For optimum \( h(t) \):

\[ \int_{T_i}^{T_f} h(t, \tau_1) K_{dr}(t, \tau_1) d\tau_1 \]

\[ \int_{T_i}^{T_f} h(t, \tau_2) K_{dr}(t, \tau_2) d\tau_2 \]
$K_r(t, u)$ and $K_{dr}(t, u)$ are the only quantities needed to find the linear MMSE point estimate. Any further statistical information about the processes can not be used. All processes, Gaussian or non-Gaussian, with the same $K_r(t, u) \text{ and } K_{dr}(t, u)$ lead to the same processor and the same linear MMSE.

Recall

$$K_{dr}(t, u) = \int_{T_i}^{T_f} h_o(t, \tau) \ K_r(\tau, u) \ dr,$$

$T_i < u < T_f$
When the Gaussian assumption holds, the optimum MMSE linear processor is the cheat of any type, i.e., a nonlinear processor cannot give an estimate with a smaller MSE.

**Pf.**

\[ r(t) \xrightarrow{h(t, r)} \hat{d}(t) \]

Optimum linear processor; i.e., \( \hat{d}(t) \) is obtained when \( E[\hat{d}(t)^2] \) is minimized (MMSE) when the processor is constrained to be linear.

\[ r(t) \xrightarrow{f(\cdot)} d_*(t) \]

Nonlinear processor; i.e., \( d_*(t) = f(t : r(u), T, U \leq U \leq T_f) \).

With the above, when the Gaussian assumption holds, (1) performs better (or at least the same) than (2) in MSE sense.
\[ f^*_x(t) = E\left\{ [d^*_x(t) - d(t)]^2 \right\} \]

and recall that

\[ f^p(t) = E\left\{ [\hat{d}(t) - d(t)]^2 \right\} \]

wish to show that

\[ f^p(t) \leq f^*_x(t) \]

Consider,

\[ f^*_x(t) = E\left\{ [d^*_x(t) - d(t)]^2 \right\} = E\left\{ (d^*_x(t) - \hat{d}(t)) + (\hat{d}(t) - d(t))^2 \right\} + \frac{f^p(t)}{2} \]

\[ = E\left\{ (d^*_x(t) - \hat{d}(t))^2 \right\} + E\left\{ (\hat{d}(t) - d(t))^2 \right\} + \frac{f^p(t)}{2} + \frac{f^p(t)}{2} \]

\[ = E\left\{ (d^*_x(t) - \hat{d}(t))^2 \right\} + E\left\{ (\hat{d}(t) - d(t))^2 \right\} + \frac{f^p(t)}{2} \left( 1 + E\left\{ e^2(t) \right\} \right) \]

\[ = E\left\{ (d^*_x(t) - \hat{d}(t))^2 \right\} + f^p(t) + \frac{f^p(t)}{2} \left( 1 + E\left\{ e^2(t) \right\} \right) \]

\[ \geq E\left\{ (d^*_x(t) - \hat{d}(t))^2 \right\} + f^p(t) + \frac{f^p(t)}{2} \left( 1 + E\left\{ e^2(t) \right\} \right) \] (nonnegative)

\[ \text{if This is zero, the proof is done.} \]
Consider,

\[ E \left\{ \left[ d_\#(t) - \hat{d}(t) \right] e_0(t)^2 \right\} \]

\[ = E \left\{ \left[ f(t; r(u), T_i \leq u \leq T_f) - \int_{T_i}^{T_f} h_0(t, u) r(u) \, du \right] e_0(t)^2 \right\} \]

**Brief Aside**

If \( d(t), r(t), \) and \( a(t) \) are jointly Gaussian, the error using the optimum linear processor is statistically independent of the input \( r(u) \) at every point in the observation interval. This is due to the fact that uncorrelated Gaussian variables are s.i.i.d.

**Summary:**

Know \( E \left\{ e_0(t) r(u) \right\} = 0, \quad T_i < u < T_f \)

For Gaussian case, orthogonality principle

\[ P \left( e_0(t) r(t) \right) = P \left( e_0(t) \right) P \left( r(t) \right) \]
\[ A = D = E \left\{ f(t; r(u), T_i \leq u \leq T_f) e_0(t)^2 \right\} - \int_{T_i}^{T_f} h_o(t; u) E \left\{ r(m) e_0(t)^2 \right\} du \]

\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_{r, e_0} \, dP_{r, e_0} \]

\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_{r, e_0} \, dP_{r, e_0} \, dr \, de \]

\[ \Rightarrow \text{which proves the property, i.e.,} \]

\[ \int_{\rho}^{\infty} (t) \leq \int_{\rho}^{\infty} (t) \]

Conclusions

1. If the Gaussian assumption holds, we are studying the best possible processor.

2. Even if the Gaussian assumption does not hold (or can't be justified) we shall have found the best possible linear processor.
**Wiener Filters**

We now consider the IE

\[
K_d(t, \sigma) = \int_{T_i}^{T_f} h_0(t, \tau) K_r(\tau, \sigma) d\tau, \quad T_i < \sigma < T_f
\]

with

\[
T_f = t \\
T_i = -\infty
\]

\[r(t) = \text{stationary}\]

\[r(t), \, d(t) \text{ are jointly stationary}\]

\[
K_d(t - \sigma) = \int_{-\infty}^{t} h_0(t, u) K_r(u - \sigma) du, \quad -\infty < \sigma < t
\]

where we try

\[h_0(t, u) = h_0(t - u)\]

Let \[T = t - \sigma\]

\[V = t - u\]

\[\implies T - V = t - \sigma - (t - u) = u - \sigma\]
\[ K_d (\tau) = \int_0^\infty h_0 (\tau) K_r (\tau - \nu) \, d\nu, \quad 0 < \tau < \infty \]

This is commonly known as the Wiener-Hopf eqn.

\[ \text{Solution of WH eqn.} \]

Let us assume that \( r(t) \) is white,

\[ K_r (\tau) = \delta (\tau) \]

\[ (\ast) \]

\[ K_d (\tau) = \int_0^\infty h_0 (\tau) \delta (\tau - \nu) \, d\nu \]

\[ = \left\{ \begin{array}{ll}
h_0 (\tau), & \tau \geq 0 \\
0, & \tau < 0
\end{array} \right. \]

or

\[ h_0 (\tau) = K_d (\tau) u (\tau). \]
It is, however, unlikely that \( r(t) \) be white. Thus, we can make it white as shown:

\[
\begin{align*}
    r(t) & \xrightarrow{W_r(r)} z(t) \\
    \text{whitening filter} & \\
\end{align*}
\]

Note that

\[
\gamma_z(\tau) = |W(j\omega)|^2 \gamma_r(\omega) = 1
\]

Recall that for the filter to be invertible, we should have

\[
\begin{align*}
    W(\tau) \ast w^{-1}(\tau) &= \delta(\tau) \\
    \mathcal{F} \left[ w^{-1}(\tau) \right] &= \frac{1}{W(j\omega)} = W^{-1}(j\omega)
\end{align*}
\]
Let

\[ S_r(w) = \frac{2k}{w^2 + k^2} \]

wish to find the TF of the whitening filter so that \( \mathbf{w} \) is realizable and

\[ S_z(w) = S_r(w) \left| W(jw) \right|^2 = 1 \]

\[ \implies W(jw) W^*(jw) = \frac{1}{S_r(w)} \]

Let's consider \( S_r(w) \) as

\[ S_r(w) = \frac{2k}{w^2 + k^2} = \left( \frac{\sqrt{2k}}{jw + k} \right) \left( \frac{\sqrt{2k}}{-jw + k} \right) = \begin{bmatrix} G^*(jw) \end{bmatrix} [G^*(jw)]^* \]

\[ \begin{bmatrix} G^*(jw) \end{bmatrix} [G^*(jw)]^* \]

\( \text{we use } ^* \text{ because we are}
\]

\( \text{to find a causal (realizable)}
\]

\( \text{whitening filter; it means it is zero for } t < 0. \)

\[ W(jw) W^*(jw) = \frac{1}{\begin{bmatrix} G^*(jw) \end{bmatrix}} \cdot \frac{1}{\begin{bmatrix} [G^*(jw)]^* \end{bmatrix}} \]

\[ \begin{bmatrix} G^*(jw) \end{bmatrix} \]

\[ \begin{bmatrix} G^*(jw) \end{bmatrix}^* \]

\[ W(jw) = \frac{1}{G^*(jw)} = \frac{jw + k}{\sqrt{2k}} \]

\( \text{differentiator + gain in parallel.} \)
\[ h_0(t) = K_{de}^u(t) u(t) \]

Optimum operation on \( z(t) \)

Diagram:
- \( r(t) \) to \( \dot{W}(j\omega) \) to \( \frac{j\omega + K}{\sqrt{2K}} \) to \( z(t) \) to \( H_0(j\omega) \) to \( d(t) \)
- \( r(t) \) to \( \dot{W}(j\omega) \) to \( \dot{W}(j\omega) \) to \( H_0(j\omega) \) to \( \hat{d}(t) \)
We have now proved that we can always find a realizable, reversible whitening filter.

![Diagram](image)

With its design, this filter is such that it operates on \( z(t) \) in such a way that it produces the MMSE estimate of \( d(t) \).

\[
K(\tau) = \int_{-\infty}^{\infty} h_o(v) K_e(\tau-v) \, dv, \quad 0 < \tau < \infty
\]

\[
h_o(\tau) = K^{d^2}(\tau) u(\tau).
\]

If we knew this, our problem would be complete.

But

\[
K^{d^2}(\tau) = E \left\{ d(t) z(t-\tau)^3 \right\}
\]

\[
= E \frac{1}{2} d(t) \int_{-\infty}^{\infty} W(v) R(t-\tau-v) \, dv \, dv
\]
\[ = \int_{-\infty}^{\infty} W(v) \ k_{dr}(\tau + v) \ dv \]

Transforming

\[ S_{dz}(j\omega) = W^*(j\omega) \ S_{dr}(j\omega) = \frac{1}{[G^*(j\omega)]^*} \ S_{dr}(j\omega) \]
We can denote the transform of $K_{d_2}(\tau)$ for $\tau \geq 0$ by the symbol

$$\left[ S_{d_2}(jw) \right]_+ = \int_0^\infty K_{d_2}(\tau) e^{-jw\tau} d\tau$$

$$H_0'(jw) = \left[ S_{d_2}(jw) \right]_+ = \left[ w^*(jw) S_{dr}(jw) \right]_+$$

$$= \left[ \frac{1}{[G^*(jw)]^* S_{dr}(jw)} \right]_+$$

This is a single stage, although it appears to be two stages; one is the "whitening" and the other is contraction of $H_0(jw)$.
once again consider

\[ r(u) = c(u) \alpha(u) + n(u), \quad T_i \leq u \leq t \]

where \( \alpha(t) \) and \( n(t) \) are \( \text{m.m} \) random processes with
\( k_{\alpha}(t, u) \) and \( \frac{N_0}{2} S(t - u) \), and \( \sigma(t) = \alpha(t) \).

Recall that the optimum processor consists of a linear filter \( h_0(t, \tau) \) where

\[ k_{ar}(t, \tau) = \int_{T_i}^{t} h_0(t, \tau) k_{\alpha}(\tau, \sigma) \, d\tau, \quad T_i < \sigma < t. \]

In this chapter we use the following conjectures:

1. Instead of describing the processes of interest in terms of their covariance functions, characterize them in terms of the linear (possibly time-varying) systems that would generate them when driven with white noise.†

2. Instead of describing the linear system that generates the message in terms of a time-varying impulse response, describe it in terms of a differential equation whose solution is the message. The most convenient description will turn out to be a first-order vector differential equation.

3. Instead of specifying the optimum estimate as the output of a linear system which is specified by an integral equation, specify the optimum estimate as the solution to a differential equation whose coefficients are determined by the statistics of the processes. An obvious advantage of this method of specification is that even if we cannot solve the differential equation analytically, we can always solve it easily with an analog or digital computer.

In this section we make these observations more precise and investigate the results.

First, we discuss briefly the state-variable representation of linear, time-varying systems and the generation of random processes. Second, we derive a differential equation which is satisfied by the optimum estimate. Finally, we discuss some applications of the technique.

The original work in this area is due to Kalman and Bucy [23].