A GENERALIZATION OF WIENER OPTIMUM FILTERING AND PREDICTION

Thesis by

Frederick Joseph Beutler

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1957

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor C. R. DePrima who has gone far beyond his role as thesis advisor to aid the author. His inspirational teaching, constant encouragement, and invaluable suggestions have enabled the author to complete this work successfully.

Thanks are also due the many other members of the Institute staff with whom the author has been in contact. Both formal instruction and informal discussions have contributed greatly in the preparation of the thesis.

The understanding, patience, and self-denial of the author's wife have made it possible for him to initiate and continue his graduate studies.

Finally, the author is grateful for the financial assistance provided by the International Business Machines Corporation in the form of a fellowship grant.

ABSTRACT

This work generalizes the Wiener-Kolmogorov theory of optimum linear filtering and prediction of stationary random inputs. It is assumed that signal and noise have passed through a random device before being available for filtering and prediction. A random device is a unit whose behavior depends on an unknown parameter for which an a priori probability distribution is given.

Use of representation theorems and a Hilbert space structure make it possible to present the mathematical theory without the ambiguities encountered in engineering derivations. This approach also leads to a proof of the essential identity between the operator solution and a realizable lumped parameter filter.

A number of engineering applications are cited. A few of these are worked out in some detail to illustrate the optimization procedure.

TABLE OF CONTENTS

INTRODUCTION

During the past three decades the science of control and communication systems has evolved from a poorly understood art to an important branch of engineering practice. One reason for this upsurge is the development of a substantial body of theory, enabling the engineer to perform effective analysis and synthesis.

One extensively explored aspect of control and communication engineering is the behavior of constant coefficient systems, i.e., systems which relate an input $x(t)$ to an output $y(t)$ through the relation

$$\sum_{o}^{n} a_j \frac{d^j}{dt^j}\Big[y(t)\Big] = \sum_{o}^{m} b_j \frac{d^j}{dt^j}\Big[x(t)\Big] \qquad (0.1)$$

where the a_j and b_j are real constants, and $n > m$. A large number of physical systems can be assumed to act according to 0.1 over their normal operating range.

The design of constant coefficient systems to perform control or communication tasks is largely done by one of two methods. The transient response approach places requirements on the system output in response to a step (or ramp) input. For example, the overshoot may be limited to a given figure, or some part of the transition from one steady state to a new steady state output accomplished within a specified time. Frequency response design emphasizes the magnitude and phase relationship of output and input sinusoids. Upper and lower bounds are placed on both magnitude and phase; the design is considered successful if the response remains within these bounds.

Both of the above methods require that the design meet criteria which are established intuitively, and which base the performance of the system on the response to a single type of input. It is therefore no surprise that the classical design techniques are often inadequate. Many systems must deal with a wide variety of inputs, so that design according to a specific type of input may result in poor performance with other inputs*. Furthermore, the input may be corrupted by random noise, so that the system must perform some indicated operation on the signal while at the same time rejecting the noise. Lastly, the system may be required to predict future values of a random signal, when noise may again be present.

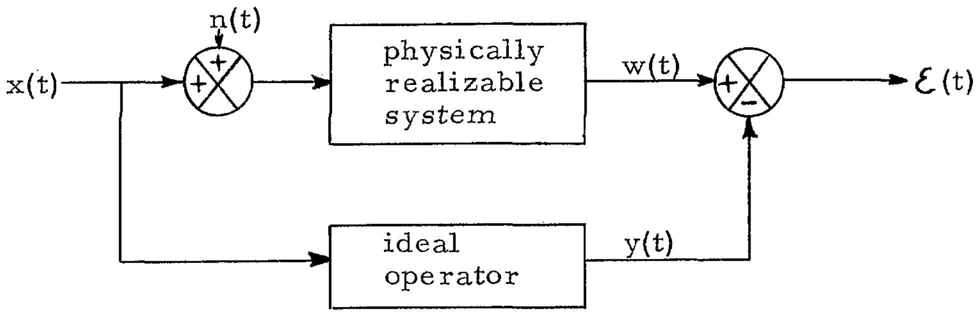
Certain design problems involving random inputs were attacked independently (and simultaneously) by Kolmogorov[1] and Wiener[2]. It is desirable to discuss their objectives as a preliminary to an explanation of the aims of this paper.

In order to give the design problem its analytical setting it is necessary to define a quantitative measure of performance. Consider a signal $x(t)$ plus noise $n(t)$ as the input of a physically realizable system, and call the system output $w(t)$. Define an ideal operator which acts on $x(t)$ to give $y(t)$. This operator is determined by the desired relationship of $y(t)$ to $x(t)$. An illustration is the prediction operator which has output $y(t + h)$ for input $x(t)$ where $h > 0$ is specified. The error, $\mathcal{E}(t)$, is

---

*This and the preceding comments are not valid for all control systems. For instance, the rocket fuel shutoff mechanism of a missile is expected to respond to one step input during its life. Therefore, the transient response to a step input is the only meaningful performance criterion.

taken to be the difference between the actual and desired outputs, that is, $\mathcal{E}(t) = w(t) - y(t)$. The error can also be described in terms of a flow chart, viz:



(0.2)

The engineer wishes to choose that realizable system which minimizes the error $\mathcal{E}(t)$. This type of optimization is feasible only if noise and signal are explicitly known beforehand, for $\mathcal{E}(t)$ (and thus the optimum system) depends on the input.

In most instances the exact system input is unknown at the time the system is being designed. However, the input may be expected to belong to a class of possible inputs. Suppose that the relative frequency with which the inputs in this class will be realized in practice can be estimated or measured. The optimum realizable system might then represent a compromise design which is most heavily influenced by those possible inputs most frequently encountered. In that case, the design objective is to minimize the error averaged over the various inputs which the system may meet in actual use.

The Wiener-Kolmogorov theory assumes that the average of $x(t)x(t + h)$ over all possible inputs is known. Furthermore, this average does not depend on t, a property referred to as stationarity*. A similar requirement is placed on $n(t)$. Then the average of $\mathcal{E}(t)\,\mathcal{E}(t+h)$ is likewise independent of t, and (with $h = 0$) the expectation (or average) of $\left|\mathcal{E}(t)\right|^2$ does not depend on t. The expectation of $\left|\mathcal{E}(t)\right|^2$ is generally called the mean square error.

The Kolmogorov-Wiener problem may now be stated as follows: find that physically realizable system which minimizes the mean square error. Of course, the minimization might have been performed with respect to some other criterion. However, there are two reasons for the choice of the mean square error as the standard. In the first place, the mean square error possesses physical significance as a measure of error power or energy. Secondly, this criterion has the distinct mathematical advantage that the theory does not involve the probability distributions of $x(t)$ or $x(t)x(t+h)$; only the second moments (covariances) enter into the problem**.

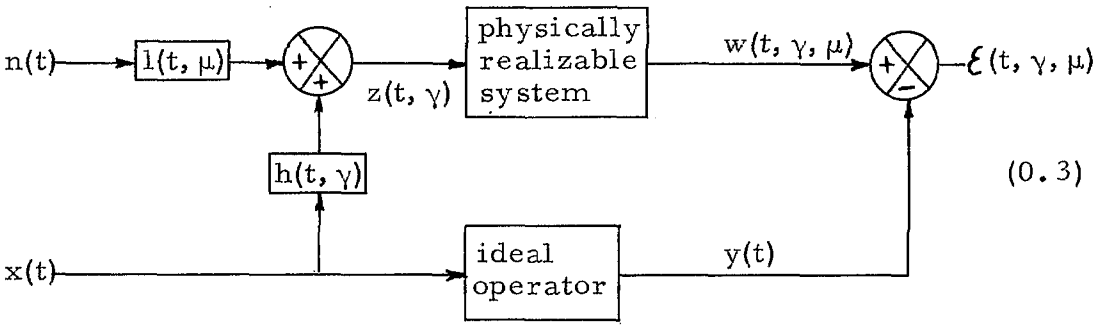
A principal objective of this thesis is to present a generalization of Wiener optimum filtering and prediction. Instead of the flow chart 0.2 applicable to the Wiener problem, the following

---

*To avoid a (trivial) deterministic version of this problem it is required that for at least one possible $x(t) + n(t)$ the future cannot be determined precisely from the past. In addition, Wiener demands that $x(t)$ be ergodic. This means that for each possible input the average over t of $x(t)x(t+h)$ must be equal to the ensemble average discussed in the text. This latter assumption is not needed if the derivations of this paper are used in place of those employed by Wiener.

**While probability distributions do not play a role here, it is interesting to observe that if $x(t)$ is a Gaussian process the linear filter which is obtained is the best of all possible filters, linear or nonlinear. Otherwise, a nonlinear filter may give even better performance relative to the mean square error criterion than the optimum linear filter. See Reference (3), pp. 74-78.

more general diagram is considered:



(0.3)

The difference between this diagram and 0.2 evidently lies in the addition of $h(t, \gamma)$ and $l(t, \mu)$. Here $h(t, \gamma)$ is a weighting function so that

$$z(t, \gamma) = \int_0^\infty x(t - \tau)\, h(\tau, \gamma)\, d\tau \qquad (0.4)$$

The parameter γ is taken to be a random variable whose probability distribution is known. The physical interpretation of $h(t, \gamma)$ is that the random process $x(t)$ is acted upon by a device which is itself random*. The output of this device constitutes the input to the system which is to be optimized with respect to the mean square error. The mean square error is now obtained by averaging $\left| \mathcal{E}(t, \gamma, \mu) \right|^2$ over γ and μ as well as all possible inputs $x(t)$ and $n(t)$. The above discussion applies to $l(t, \mu)$ in analogous fashion, except that $l(t, \mu)$ acts on the noise $n(t)$ instead of the signal $x(t)$.

Several engineering examples will serve to clarify the idea of a random weighting function, and to indicate some of the possible applications of the optimization problem depicted by 0.3.

---

* $h(t, \gamma)$ may also be a delta function representing a random gain or scale factor device.

Suppose that it is desired to design a phonograph. Ideally, this instrument would be equipped with compensation for each of the recording characteristics employed by the various record manufacturers. But commercial policies often permit only a single fixed compensation network to be used. This should be designed to represent an optimum compromise among the different recording characteristics. If the optimum is that network which minimizes the average error power the work of this thesis is applicable. Each recording characteristic is an $h(t, \gamma_i)$. The probability that $\gamma = \gamma_i$ can be taken as the number of records sold having characteristic i divided by the total number of records sold. Furthermore, estimates are available for the frequency spectrum of music. These estimates describe the statistical properties of $x(t)$.

The phonograph design example illustrates a larger class of applications. It is typical of a control or data transmission systems that corrective feedback cannot be applied to input and output devices. The compensation then takes the form of a network in series with the input and/or output device. The design of such a network is doubly difficult because the characteristics of the input and output elements in question will vary from unit to unit, and may also change with age or use. The statistical parameters of the input and output devices can be determined from a large number of laboratory tests. Then $h(t, \gamma)$ and the distribution of γ will be known. The ideal operator is determined from the over-all purpose of the control or data transmission system. If a mean square error criterion is adopted, and the system input is stationary, the methods of this thesis are applicable.

The new techniques of this thesis can also be applied to communication systems in which the error is considered to be the difference between the input to the transmitter and the output of the receiver. It is conventional practice to attach a network filter to the receiver output to reduce the average error power to the lowest value possible. In ground-to-ground microwave systems multipath problems are created by ground reflections. More simply, the transmitted signal arrives not only by a direct line of sight path, but also via multiple ground reflection paths of varying lengths. The received signal is shifted in phase, and there is cancellation of reinforcement which affects the signal amplitude. In the case of an amplitude modulated communication system* the effect is as if the signal had been passed through a filter $h(t, \gamma)$ whose Fourier transform is $1 + F(\omega, \gamma)$. Here the first term is due to the direct transmission path, and the second expresses the aggregate of the reflections in terms of the parameter γ. γ depends on the terrain over the transmission path, the transmission distance, and the height of the transmitter and receiver above ground. These factors are not known in a mobile communication complex, nor is it easy to calculate $F(\omega, \gamma)$ once they are measured. However, controlled experiments can deter-

---

*Amplitude modulation is specified because it is a linear process with sidebands whose amplitude and phase vary directly with those of the input signal. These requirements must be met if linear analysis is to be used.

mine $F(\omega, \gamma)$ under various sets of conditions. The probability that the mobile system achieves a particular $F(\omega, \gamma)$ on a given occasion is assigned in advance an a priori grounds from a knowledge of the tactical mission of the system. The problem is then of the type which has been solved in this thesis, assuming again that signal and noise are stationary with known statistical parameters.

Of course, the illustrations provided in this introduction encompass but a small part of the engineering problems which can be handled through use of the methods of the thesis. For instance, further examples are offered (and worked out in some detail) in sections five to seven of this paper.

It should not be inferred that the present work follows Wiener's derivations in arriving at the generalized results. Instead, representation theorems and a Hilbert space structure are employed to resolve some of the ambiguities in Wiener's work*.

The operator solution which appears in the frequency domain often lacks an interpretation in the time domain, and certainly cannot be synthesed by known network techniques. The writer shows that a filter can be constructed so that the mean square error is as close as desired to that of the operator solution. In fact, it is possible to find a lumped parameter filter having this property.

---

*The Hilbert space approach has been exploited by Doob (Reference (3)) to obtain a rigorous treatment of Wiener-Kolmogorov prediction and filtering.

A second point, neglected by engineering texts dealing with Wiener filtering, is that the factorization of the frequency spectrum (see section 5) is not unique. The uniqueness condition is stated here, and a time domain equivalent given.

Still other matters of interest are the limiting cases of long time prediction and lag, as discussed in section 6. Lastly, section 7 is concerned with the generalized prediction and filtering theory as applied to processes which are not themselves stationary, but which have stationary increments.

SECTION I

MATHEMATICAL FOUNDATIONS

The optimum filter problem we shall treat was discussed and motivated in the introductory section of this paper. In order to proceed to the solution of the problem we must express it in more precise terms. In that way, we are able to explore the widest conditions of validity under which the solution holds, and to show that this solution is unique and does yield the minimum mean square error.

Certain basic definitions and concepts will be used without further explanation. For example, such notions as probability space, expectation, and stationary (wide sense) process are discussed elsewhere in great detail. Any exposition of these in this paper would be redundant and lengthy. The reader is advised to consult Reference (3) for the necessary background information.

Let $x(t)$ be a process which is wide sense stationary and *continuous in the mean.* Such a process has a spectral representation of the form

$$x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} dX(\omega) \qquad (1.1)$$

The integral is defined in the sense of a "limit in mean", usually *written l.i.m. (see Reference (3),* Chap. IX and pp. 527-8). Here $X(\omega)$ is a process with orthogonal increments such that

$$E \left| X(\lambda_2 -) - X(\lambda_1 -) \right|^2 = 2\pi \int_{\lambda_1}^{\lambda_2} d\,\Lambda\,(\omega) \qquad (1.2)$$

We shall restrict ourselves to absolutely continuous spectral distribution functions. That is, there exists a spectral density $\Phi(\omega)$ such that

$$\int_{-\infty}^{\lambda} \Phi(\omega)d\omega = \Lambda(\lambda) < \infty \tag{1.3}$$

Evidently, $\Phi(\omega) \epsilon L_1$. We shall also assume that

$$\int_{-\infty}^{\infty} \frac{\log \Phi(\omega)}{1+\omega^2} d\omega > -\infty \tag{1.4}$$

It may be shown (Reference (3), p. 584) that if (and only if) 1.4 does not hold, the entire future and past of $x(t)$ may be exactly determined from the behavior of $x(t)$ over any t interval of some specified length. In other words, 1.4 is satisfied unless $x(t)$ is a trivial process from the viewpoint of prediction and filtering theory.

The output which we desire to obtain from our filter will be denoted by $y(t)$. $y(t)$ is related to $x(t)$ by an operator $u(\omega)$. We define

$$y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(\omega) e^{i\omega t} dX(\omega) \tag{1.5}$$

where the integral is understood to be l.i.m. The existence of the integral 1.5 is equivalent to the finiteness of the covariance of $y(t)$, or

$$E\left| y(t) \right|^2 = \int_{-\infty}^{\infty} \Phi(\omega) \left| u(\omega) \right|^2 d\omega \tag{1.6}$$

as shown in Reference (3), p. 534.

Since the notion of an operator is basic to this paper, a few clarifying examples will be offered at this point. Suppose that $u(\omega) = e^{i\omega a}$. Then

$$y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega(t+a)} dX(\omega) = x(t+a) \tag{1.7}$$

so that $e^{i\omega a}$ is regarded as a prediction or lag operator according to whether a is positive or negative. Consider next $u(\omega) = i\omega$. In that case we have $y(t) = \underset{h \to 0}{\mathrm{l.i.m.}} \frac{x(t+h) - x(t)}{h} \left(= \frac{dx}{dt}\right)$. As a third example, assume that $u(\omega)$ is the Fourier transform of an L_1 function $U(t)$. There follows

$$y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(\omega) e^{i\omega t} dX(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\omega t} dX(\omega) \int_{-\infty}^{\infty} e^{-i\omega\tau} U(\tau) d\tau \tag{1.8}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U(\tau) d\tau \int_{-\infty}^{\infty} e^{i\omega(t-\tau)} dX(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t-\tau) U(\tau) d\tau$$

the interchange of integrals being valid according to Reference (3), p. 431, i.e., by a direct application of Fubini's theorem.

As is seen from the above examples, the operation $u(\omega)$ is more general than that of prediction, differentiation, or application of an ordinary filter to $x(t)$. In fact, any $u(\omega)$ satisfying 1.6 is admissible, whether or not it admits of an interpretation in the time domain.

The process available to our optimum filter is yet a third process, called $z(t, \gamma)$. The z process is created by subjecting

$x(t)$ to an operator depending on a random variable γ, γ being defined on a probability field independent of that of $x(t)$. It will be assumed that this operator has meaning in the time domain. In fact, we begin by specifying that $h(t, \gamma)$ be jointly measurable on the space of t and γ, and that $h(t, \gamma) = 0$ on the product set of $t < 0$ and γ with the possible exception of a set of product measure zero. We demand, furthermore, that

$$E \left| \int_0^\infty \left| h(t, \gamma) \right| dt \right|^2 < \infty \tag{1.9}$$

and

$$E \int_0^\infty \left| h(t, \gamma) \right|^2 dt < \infty \tag{1.10}$$

It follows from 1.9 and the Schwarz inequality that

$$E \int_0^\infty \left| h(t, \gamma) \right| dt \leq \left(E \left| \int_0^\infty \left| h(t, \gamma) \right|^2 dt \right| \right)^{1/2} < \infty \tag{1.11}$$

Therefore, $h(t, \gamma) \epsilon L_1$ with probability one, so that the L_1 Fourier transform of $h(t, \gamma)$ exists with probability one; this transform is written $H(\omega, \gamma)$.

We shall have need to speak of $E\left[H(\omega, \gamma) \right]$ and $E \left| H(\omega, \gamma \right|^2$ in the future. It is convenient, therefore, to define $H(\omega)$ and $G(\omega)$ by $H(\omega) = E\left[H(\omega, \gamma) \right]$ and $G(\omega) = E \left| H(\omega, \gamma) \right|^2$. These expectations always exist, and indeed Parseval's theorem together with 1.10 gives

$$\int_{-\infty}^\infty G(\omega) \, d\omega < \infty \tag{1.12}$$

In consequence of 1.29 $G(\omega)$ is everywhere bounded so that

$$\int_{-\infty}^{\infty} \Phi(\omega)\, G(\omega)\, d\omega < \infty \tag{1.13}$$

Furthermore, 1.8 may be applied to $h(t,\gamma)$ to yield

$$z(t,\gamma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H(\omega,\gamma) e^{i\omega t}\, dX(\omega) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} x(t-\tau) h(\tau,\gamma)\, d\tau \tag{1.14}$$

This means that $\frac{1}{\sqrt{2\pi}} h(t,\gamma)$ is the weighting function of a realizable, stable filter (with probability one), in accordance with the criteria to be presented near the end of this section.

Let $r(\omega)$ be an operator applied to $z(t,\gamma)$. If

$$\int_{-\infty}^{\infty} \Phi(\omega)\, G(\omega) \left| r(\omega) \right|^2 d\omega < \infty \tag{1.15}$$

we have a new process, $w(t,\gamma;r)$ which is represented by

$$w(t,\gamma;r) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H(\omega,\gamma)\, r(\omega) e^{i\omega t}\, dX(\omega) \tag{1.16}$$

the integral again being defined in the l.i.m. sense.

In the terminology of the engineer, $r(\omega)$ describes the operation of a device with input $z(t,\gamma)$ and output $w(t,\gamma;r)$. Thus, since the desired output is $y(t)$ (see 1.5), it makes sense to speak of an error

$$\mathcal{E}(t,\gamma;r) = w(t,\gamma) - y(t) \tag{1.17}$$

The error is again a stochastic process which is wide sense stationary and continuous in the mean (Reference (3), p.534). Then we may denote the mean square value of $\mathcal{E}(t, \gamma; r)$ by $\mathcal{E}^2(r)$; i.e.,

$$\mathcal{E}^2(r) = E\left|\mathcal{E}(t, \gamma; r)\right|^2 \tag{1.18}$$

We find that

$$\mathcal{E}^2(r) = \underset{\gamma}{E}\left[\int_{-\infty}^{\infty} \Phi(\omega)\left|H(\omega, \gamma) r(\omega) - u(\omega)\right|^2 d\omega\right] \tag{1.19}$$

where $\underset{\gamma}{E}$ is the expectation on γ. 1.19 may be rewritten

$$\mathcal{E}^2(r) = \int_{-\infty}^{\infty} \Phi(\omega)\left[G(\omega)\left|r(\omega)\right|^2 - H(\omega) r(\omega)\overline{u(\omega)} - \overline{H(\omega) r(\omega)} u(\omega) + \left|u(\omega)\right|^2\right] d\omega \tag{1.20}$$

The latter form is justified by Fubini's theorem. The application of the theorem to the first and third term of 1.20 is clear. For the middle terms the Schwarz inequality gives $\left|H(\omega)\right|^2 \leq G(\omega)$ so that a second application of the inequality yields

$$\int_{-\infty}^{\infty}\left|\Phi(\omega) H(\omega) r(\omega) u(\omega)\right| d\omega \leq \left(\int_{-\infty}^{\infty} \Phi(\omega) G(\omega)\left|r(\omega)\right|^2 d\omega\right)^{1/2} \times \left(\int_{-\infty}^{\infty} \Phi(\omega) G(\omega)\left|u(\omega)\right|^2 d\omega\right)^{1/2} \tag{1.21}$$

It is convenient to denote the class of all $r(\omega)$ satisfying 1.15 by $\mathcal{H}$. As will be seen below, we may find a unique $r \in \mathcal{H}$ which minimizes the mean square error defined by 1.18. However, the engineer thinks in terms of another class, so that ultimately we shall minimize the mean square error over a class $\mathcal{M}$, the class of stable, realizable filters. In what follows, we shall give $\mathcal{M}$ a mathematically precise meaning.

A filter is defined by a weighting function, say K(t), which relates a filter input x(t) to a filter output y(t) by

$$y(t) = \int_{-\infty}^{\infty} x(t - \tau)\, K(\tau) d\tau \tag{1.22}$$

if the integral has meaning. A filter is said to be realizable if that filter is capable of acting only on signals received in the past, and not on signals which are yet to arrive in the future. Thus, a filter is defined to be realizable if and only if its weighting function $K(t) = 0$ for a.e. $t < 0$. Lastly, a filter is stable if every essentially bounded (measurable) input to the filter results in an essentially bounded (measurable) output. We have

<u>Theorem 1:</u> A filter is stable if and only if its weighting function is in L_1, that is

$$\int_{-\infty}^{\infty} \left| K(t) \right| dt < \infty \tag{1.23}$$

Proof: Using the notation of 1.22, let $x(t) \leq N$ a.e. Then $\left| y(t) \right| \leq N \int_{-\infty}^{\infty} \left| K(t) \right| dt$ which is bounded if $K(t) \in L_1$. Conversely, suppose that $K(t) \notin L_1$. If $x(t) = N$, we have

$$\int_{-\infty}^{\infty} x(t-\tau)K(\tau)d\tau = N \int_{A} K(\tau)d\tau - N \int_{X-A} \left| K(\tau) \right| d\tau \tag{1.24}$$

where $A = \left\{ t \mid K(t) \geq 0 \right\}$. At least one of the two terms on the right hand side of 1.24 is infinite because $K(t) \notin L_1$. Thus $K(t)$ cannot represent a stable filter.

We shall study the relationship between $\mathcal{M}$ and $\mathcal{H}$ further. For this purpose it is useful to utilize well-known properties of Hilbert spaces. We may define an inner product

$$(r, s) = E\left[w(t, \gamma; r)\overline{w(t, \gamma; s)} \right] = \int_{-\infty}^{\infty} \Phi(\omega) G(\omega) r(\omega) \overline{s(\omega)} d\omega \qquad (1.25)$$

where r and s are any two elements of $\mathcal{H}$. The inner product is seen to be a symmetric bilinear functional on $\mathcal{H}$. Also, $\mathcal{H}$ is complete by the Riesz-Fisher theorem. In short, the space $\mathcal{H}$ with inner product defined by 1.25 is a Hilbert space.

The null element of the Hilbert space $\mathcal{H}$ deserves mention. If $\| r \| = 0$, we have

$$\int_{-\infty}^{\infty} \Phi(\omega) G(\omega) \left| r(\omega) \right|^2 d\omega = 0 \qquad (1.26)$$

$\Phi(\omega) > 0$ a.e. is assured by 1.4. $G(\omega)$ is non-negative, since by definition $G(\omega) = E\left| H(\omega, \gamma) \right|^2$. In section three, we shall see that $\int_{-\infty}^{\infty} \frac{|\log G(\omega)|}{1 + \omega^2} d\omega < \infty$ so that $G(\omega) > 0$ a.e. It follows that 1.26 holds only if $r(\omega) = 0$ a.e.; hence, two operators in $\mathcal{H}$ are equivalent if and only if they are equal a.e.

The mean square error may be rewritten as

$$\mathcal{E}^2(r) = \left\| r - \frac{u\overline{H}}{G} \right\|^2 + \left\| \frac{u}{\sqrt{G}} \right\|^2 - \left\| \frac{u\overline{H}}{G} \right\|^2 \qquad (1.27)$$

from 1.20. We shall not develop 1.27 further at present, but only note that $\mathcal{E}^2(r)$ is minimized if

$$r(\omega) = \frac{u(\omega)\overline{H(\omega)}}{G(\omega)} \tag{1.28}$$

Let r' be any other element of $\mathcal{H}$. Then

$$\mathcal{E}^2(r') - \mathcal{E}^2(r) = \| r' - r \|^2 \tag{1.29}$$

Thus, if the choice of operators is to be restricted by some subspace $\mathcal{S} \subset \mathcal{H}$, the mean square error is minimized by the element in $\mathcal{S}$ which is the projection of $r \epsilon \mathcal{H}$ on $\mathcal{S}$.

The above discussion reveals how the optimization is to be performed with respect to $\overline{\mathcal{M}}$. Indeed, we need only show that $\mathcal{M} \subset \mathcal{H}$. From 1.8 and the properties of L_1 Fourier transforms, an element of $\mathcal{M}$ has an operator representation which is bounded and uniformly continuous. Now we have

Lemma 1: Every essentially bounded measurable operator belongs to $\mathcal{H}$.

Proof: According to 1.3, $\Phi(\omega) \epsilon L_1$. Then 1.15 is satisfied for an essentially bounded operator if $G(\omega)$ is shown to be bounded. We write for the absolute value of $G(\omega)$

$$2\pi \left| G(\omega) \right| = E \left| \int_0^\infty h(t,\gamma) e^{-i\omega t} dt \right|^2 \leq E \left| \int_0^\infty \left| h(t,\gamma) \right| dt \right|^2 < \infty \text{ by } 1.9 \tag{1.30}$$

which proves the lemma.

The projection operator of $\mathcal{H}$ on $\overline{\mathcal{M}}$ is difficult to find by direct means. Instead, we shall define a subspace $\hat{\mathcal{M}}$, project $\mathcal{H}$

on $\hat{\mathfrak{M}}$ in an obvious fashion, and then show that $\hat{\mathfrak{M}} = \overline{\mathfrak{M}}$.

Quite often, the engineer wishes to synthesize an optimum filter which is not only in $\mathfrak{M}$, but which can be constructed of lumped parameter elements (resistors, capacitors, inductors). A filter of this type is characterized by a rational Fourier transform. The class of such filters will be called $\mathfrak{M}'$. We have already demanded that $\mathfrak{M}' \subset \mathfrak{M}$. Now if, in addition, $\overline{\mathfrak{M}'} = \overline{\mathfrak{M}}$, the engineer knows that he will be able to synthesize a lumped parameter filter whose performance (as measured by the corresponding mean square error) will approach that element of $\overline{\mathfrak{M}}$ which we shall determine to be optimum. Accordingly, we shall prove

<u>Theorem 2:</u> Let $\Phi(\omega)G(\omega)$ be essentially bounded. Then $\overline{\mathfrak{M}} = \overline{\mathfrak{M}'}$.

Proof: Let $\overline{\mathfrak{M}''}$ be the subspace generated by the family of operators

$$\left\{ \frac{1}{\sqrt{\pi}} \frac{(1 - i\omega)^n}{(1 + i\omega)^{n+1}} \right\}^{*}.$$

Let $\mathcal{S}$ be the family of step functions in $(0, \infty)$; we shall think in terms of Fourier transforms of the step functions. Then it will be shown that $\mathfrak{M}'' \subset \mathfrak{M}' \subset \mathfrak{M} \subset \overline{\mathcal{S}} \subset \overline{\mathfrak{M}''}$, so that also

$$\overline{\mathfrak{M}} = \overline{\mathfrak{M}'} = \overline{\mathcal{S}} = \overline{\mathfrak{M}''} \tag{1.31}$$

It is evident that $\mathfrak{M}'' \subset \mathfrak{M}'$, since $\frac{(1 - i\omega)^n}{(1 + i\omega)^{n+1}}$ is the rational Fourier transform of an L_1 function which is zero for a.e. negative argument. By definition, we also have $\mathfrak{M}' \subset \mathfrak{M}$.

---

* These are the well-known Laguerre polynomials.

Given $\epsilon > 0$, and a $K(t) \in \mathcal{M}$, there exists an $S(t) \in \mathcal{L}$ such that

$$\left| k(\omega) - s(\omega) \right| \leq \int_{-\infty}^{\infty} \left| K(t) - S(t) \right| dt < \mathcal{E} \tag{1.32}$$

where $k(\omega)$ and $s(\omega)$ are Fourier transforms of $K(t)$ and $S(t)$, respectively. If $G(\omega) \leq M$ (see 1.30) there follows

$$\| k - s \|^2 < M \mathcal{E}^2 \int_{-\infty}^{\infty} \Phi(\omega) d\omega \tag{1.33}$$

Lastly, we prove that $\mathcal{J} \subset \overline{\mathcal{M}''}$. The set

$$\left\{ \frac{1}{\sqrt{\pi}} \; \frac{(1 - i\omega)^n}{(1 + i\omega)^{n+1}} \right\}$$

is a complete orthonormal set in L_2 norm. Therefore, (since $\mathcal{L} \subset L_2$) $\mathcal{M}''$ is dense in $\mathcal{L}$ with respect to L_2 norm. Because $\Phi(\omega)G(\omega)$ is essentially bounded the density argument holds in $\mathcal{H}$ norm also; hence $\mathcal{J} \subset \overline{\mathcal{M}''}$.

SECTION II

GENERALIZED OPTIMUM FILTERS AND THE WIENER PROBLEM

In this section we shall study some relationships between the filter problem of Wiener, and the problem presented in the preceding portions of this paper. It should be apparent that our problem reduces to the Wiener problem if $h(t, \gamma)$ is a known (rather than a random) function, i.e., a function not dependent on γ. On the other hand, if $h(t, \gamma)$ is not exactly known, it might be expected that the mean square error becomes larger, and the optimum filter solution differs from the Wiener solution. If the mean square error 1.27 is rewritten in the form

$$\varepsilon^2(r) = \left\| r - \frac{u\overline{H}}{G} \right\|^2 + \varepsilon^2_{glb} \tag{2.1}$$

in which ε^2_{glb} (the greatest lower bound mean square error) is

$$\varepsilon^2_{glb} = \left\| \frac{u}{\sqrt{G}} \right\|^2 - \left\| \frac{u\overline{H}}{G} \right\|^2 = \int_{-\infty}^{\infty} \Phi(\omega) \left| u(\omega) \right|^2 \left[1 - \frac{\left| H(\omega) \right|^2}{G(\omega)} \right] d\omega \geqslant 0 \tag{2.2}$$

we observe that a new error term foreign to Wiener filtering arises, and that this term depends only on the nature of $h(t, \gamma)$. It will be demonstrated that $\varepsilon^2_{glb} = 0$ if and only if $h(t, \gamma)$ is a sure function as defined below. In that case, the optimum element of $\mathcal{H}$ becomes*

$$r(\omega) = \frac{u(\omega)}{H(\omega)} \tag{2.3}$$

---

*If $h(t, \gamma)$ is not a sure function, we always have $\left| r(\omega) \right| < \left| \frac{u(\omega)}{H(\omega)} \right|$ a.e. except on sets where $u(\omega) = 0$.

We shall now demonstrate the validity of the above statements. A sure function is defined by the existence of a $\tilde{h}(t)$ such that

$$h(t, \gamma) = \tilde{h}(t) \tag{2.4}$$

except on a set whose product measure in t and γ is zero. An equivalent statement is given by

Lemma 2: Let $h(t, \gamma)$ be a sure function according to 2.4. If $H(\omega)$ is the Fourier transform of $\tilde{h}(t)$, we have

$$H(\omega, \gamma) = \tilde{H}(\omega) \tag{2.5}$$

except on a set whose product measure in ω and γ is zero. Conversely, if 2.5 holds as stated, $h(t, \gamma)$ is a sure function with $\tilde{h}(t)$ having $\tilde{H}(\omega)$ as its Fourier transform.

Proof: For every ω

$$H(\omega, \gamma) = \frac{1}{\sqrt{2\pi}} \int_0^\infty h(t, \gamma) e^{-i\omega t} dt = \frac{1}{\sqrt{2\pi}} \int_0^\infty \tilde{h}(t) e^{-i\omega t} dt = \tilde{H}(\omega) \tag{2.6}$$

where the middle integrals are equal except for a γ set of measure zero. Note that this γ set does not depend on the choice of ω, for with probability one

$$\left| \int_0^\infty \left[h(t, \gamma) - \tilde{h}(t) \right] e^{-i\omega t} dt \right| \leq \int_0^\infty \left| h(t, \gamma) - \tilde{h}(t) \right| dt = 0 \tag{2.7}$$

For the converse, 1.10 and Parseval's relation yields

$$E \int_0^\infty \left| h(t, \gamma) - \tilde{h}(t) \right|^2 dt = E \int_{-\infty}^\infty \left| H(\omega, \gamma) - \tilde{H}(\omega) \right|^2 d\omega = 0 \tag{2.8}$$

which can be true only if

$$h(t, \gamma) = \tilde{h}(t) \tag{2.9}$$

except on a set of product measure zero. An application of the first half of the lemma shows that $\tilde{h}(t)$ indeed has $\tilde{H}(\omega)$ for its Fourier transform.

We can now show how our optimum filtering reduces to the Wiener case when $h(t, \gamma)$ is a sure function. This is accomplished by

Theorem 3: $\left| H(\omega) \right|^2 = G(\omega)$ a.e. if and only if $h(t, \gamma)$ is a sure function.

Proof: If $h(t, \gamma)$ is a sure function we use the preceding lemma to write

$$G(\omega) = E\left| H(\omega, \gamma) \right|^2 = E\left| \tilde{H}(\omega) \right|^2 = \left| E\left[\tilde{H}(\omega) \right] \right|^2 = \left| E\left[H(\omega, \gamma) \right] \right|^2$$

$$= \left| H(\omega) \right|^2 \tag{2.10}$$

which holds for a.e. ω.

Conversely, if $\left| H(\omega) \right|^2 = G(\omega)$ a.e.

$$\int_{-\infty}^{\infty} E\left| H(\omega, \gamma) - E\left[H(\omega, \gamma) \right] \right|^2 d\omega = 0 \tag{2.11}$$

so that

$$H(\omega, \gamma) = E\left[H(\omega, \gamma) \right] = \tilde{H}(\omega) \tag{2.12}$$

except on a product set of measure zero.

Corollary: A sufficient condition that $\varepsilon^2_{glb} = 0$ is that $h(t, \gamma)$ be a sure function. The condition is also necessary if $|u(\omega)| > 0$ a.e. The proof of the corollary is obvious. It need only be added that $G(\omega) > 0$ a.e., a statement that will be proved in section three.

SECTION III

DERIVATION OF THE STABLE, REALIZABLE FILTER SOLUTION

It was stated in section one that the optimization with respect to $\mathfrak{M}$ (the subspace in $\mathcal{H}$ generated by the class of stable, realizable filters) would be accomplished by projecting the optimum element in $\mathcal{H}$ on $\mathfrak{M}$. In this section we shall determine a projection operator appropriate to this purpose, so that the engineer can find an explicit solution in terms of $\mathfrak{M}$.

The following theorem plays a central role in the work of this section:

Theorem 4: Given a function $F(\omega)$ such that

1. $F(\omega) \geqslant 0$ a.e. (3.1)

2. $F(\omega) \epsilon L_1$ (3.2)

3. $$\int_{-\infty}^{\infty} \frac{\log F(\omega)}{1+\omega^2} d\omega > -\infty \qquad (3.3)$$

there exists an $F_1(\omega)$ such that

1. $\left| F_1(\omega) \right| = \sqrt{F(\omega)}$ a.e. (3.4)

2. $F_1(\omega)$ is the L_2 Fourier transform of a function $f(t)$ with $f(t) = 0$ for a.e. $t < 0$, and (3.5) $\int_0^{\infty} f(t) e^{-izt} dt \neq 0$ for all $\sigma < 0$.

3. $$\log \left[\frac{1}{\sqrt{2\pi}} \int_0^{\infty} f(t) e^{-t} dt \right] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\log F(\omega)}{1+\omega^2} d\omega \qquad (3.6)$$

4. If $f(t) = 0$ for a.e. t in $[0, \delta]$, then $\delta = 0$. (3.7)

Furthermore, the $F_1(\omega)$ specified by 3.4, 3.5, and 3.6, is a.e. unique up to a complex constant of modulus unity.

Proof: Let $R(\omega) = \sqrt{F(\omega)} \geq 0$. Then $R(\omega) \epsilon L_2$, and satisfies 3.3 if $F(\omega)$ does so. 3.3 implies the existence of

$$\lambda(z) = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sigma \log R(x)}{\sigma^2 + (x-\omega)^2}\,dx \qquad \sigma < 0 \qquad (3.8)$$

with $z = \omega + i\sigma$. Since 3.8 is the Poisson integral applied to the lower half plane, $\lambda(z)$ is harmonic in that half plane. The function $F_1(z)$ may be defined by

$$F_1(z) = \exp\left[\lambda(z) + i\mu(z)\right] \qquad \sigma < 0 \qquad (3.9)$$

where $\mu(z)$ is the conjugate of $\lambda(z)$.

Paley and Wiener have shown* that for all $\sigma < 0$

$$\int_{-\infty}^{\infty} \left|F_1(\omega - i\sigma)\right|^2 d\omega \leq \int_{-\infty}^{\infty} F(\omega)\,d\omega < \infty \qquad (3.10)$$

The fact that $F_1(z)$ is regular in the lower half plane, together with 3.10 means that there exists an $f(t) \epsilon L_2$, $f(t) = 0$ for a.e. $t < 0$, such that**

$$F_1(\omega + i\sigma) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} f(t)e^{-it(\omega+i\sigma)}\,dt \qquad (3.11)$$

and indeed

---

* Reference (4), p. 18, (eq. 8.06)
** Reference (4), pp. 8 - 9

$$F_1(\omega) = \underset{A \to \infty}{\text{l.i.m.}} \int_0^A f(t) e^{-i\omega t} dt \qquad (3.12)$$

where 3.10 enables us to define

$$F_1(\omega) = \underset{\sigma \to 0}{\text{l.i.m.}} \; F_1(\omega + i\sigma) \qquad (3.13)$$

Thus $F_1(\omega)$ satisfies 3.5.

The theorem of Fatou* applied to the half plane yields

$$\lim_{\sigma \to 0} \lambda(\omega + i\sigma) = \log R(\omega) \qquad \text{a.e.} \qquad (3.14)$$

so that

$$\lim_{\sigma \to 0} \left| F_1(\omega + i\sigma) \right| = \sqrt{F(\omega)} \qquad \text{a.e.} \qquad (3.15)$$

If 3.13 and 3.15 are compared, we obtain

$$\left| F_1(\omega) \right| = \sqrt{F(\omega)} \qquad (3.16)$$

which verifies 3.4.

To prove 3.6, we note that the construction 3.8 and 3.9 defines $\mu(z)$ up to an arbitrary constant. This constant is now chosen in such a manner that $\mu(-i) = 0$. If we set $\omega = 0$, $\sigma = -1$, 3.9 becomes

$$F_1(-i) = \exp\left[\frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\log F(\omega)}{1 + \omega^2} d\omega \right] \qquad (3.17)$$

which, in combination with 3.11 gives 3.6 as the result.

---

* For the unit circle, this theorem may be found in Reference (5), Vol. II, pp. 147 ff.

To prove uniqueness, suppose $\tilde{F}_1(\omega)$ also satisfies 3.4, 3.5, and 3.6. Then 3.11 defines a $\tilde{F}_1(z)$ regular and non-zero (see 3.5) in the lower half plane, and such that

$$\int_{-\infty}^{\infty} \left|\tilde{F}_1(\omega + i\sigma)\right|^2 d\omega \leqslant \int_{-\infty}^{\infty} \left|\tilde{F}_1(\omega)\right|^2 d\omega < \infty \tag{3.18}$$

$$\tilde{F}_1(\omega) = \underset{\sigma \to 0}{\text{l.i.m.}}\ \tilde{F}_1(\omega + i\sigma) \tag{3.19}$$

The last two equations follow from the work of Paley and Wiener*.

According to a theorem of Kryloff[6], a $\tilde{F}_1(z)$ possessing the above properties has a representation

$$\tilde{F}_1(z) = e^{ic + iaz}\, D(z)G(z) \tag{3.20}$$

where $a \leqslant 0$, c is real, and $D(z)$ and $G(z)$ are given by

$$D(z) = \exp \frac{i}{\pi} \int_{-\infty}^{\infty} \frac{1 + xz}{x - z}\, \frac{\log \tilde{F}_1(x)}{1 + x^2}\, dx \tag{3.21}$$

and

$$G(z) = \exp \frac{1}{i\pi} \int_{-\infty}^{\infty} \frac{1 + xz}{x - z}\, dS(x) \tag{3.22}$$

respectively. Here $S(x)$ is monotone increasing and bounded. The corresponding representation for $\log\left|\tilde{F}_1(z)\right|$ is then

$$\log\left|\tilde{F}_1(z)\right| = -a\sigma + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{-\sigma \log\left|\tilde{F}_1(x)\right|}{\sigma^2 + (x - \omega)^2}\, dx + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sigma(1 + x^2)}{\sigma^2 + (x - \sigma)^2}\, dS(x) \tag{3.23}$$

---

* Reference (4), theorem V.

The second term on the right hand side of 3.23 is equal to $\log|F_1(z)|$ by 3.4 and 3.8, while the first and third terms are negative or zero. Hence

$$\log|\tilde{F}_1(z)| \leqslant \log|F_1(z)| \tag{3.24}$$

But 3.6 and 3.11 imply that $F_1(-i) = \tilde{F}_1(-i)$. Therefore, the maximum modulus theorem applied to $\log|F_1(z)| - \log|\tilde{F}_1(z)|$ gives

$$|F_1(z)| = |\tilde{F}_1(z)| \tag{3.25}$$

from which

$$F_1(z) = e^{ic}\tilde{F}_1(z) \tag{3.26}$$

Then also

$$F_1(\omega) = e^{ic}\tilde{F}_1(\omega) \tag{3.27}$$

by virtue of 3.13.

Only 3.7 remains to be demonstrated. Let $f(t) = 0$ for a.e. t in $[0, \delta]$. Then $\hat{f}(t) = f(t+\delta)$ satisfies 3.4 and 3.5, and we have

$$\hat{F}_1(z) = e^{i\delta z}F_1(z) \tag{3.28}$$

by 3.11. This gives

$$\log|\hat{F}_1(z)| = -\delta y + \log|F_1(z)| \tag{3.29}$$

with $y < 0$. If the Kryloff representation[6] is applied to $\hat{F}_1(z)$

$$\log|\hat{F}_1(z)| \leqslant \log|F_1(z)| \tag{3.30}$$

the proof being identical with that leading to 3.24. It is evident that

3.29 and 3.30 are contradictory unless $\delta = 0$. This completes the proof of the theorem.

If a function meets the conditions placed on $F(\omega)$ by theorem 4, that function will be called factorable. Its factor is the (unique) function satisfying 3.4, 3.5, and 3.6.

Theorem 4 is readily applied to $\Phi(\omega)$. That $\Phi(\omega)$ is a.e. non-negative follows from the nondecreasing nature of $\Lambda(\omega)$ as required by 1.2. The assumption that $x(t)$ is continuous in the mean leads to $\Phi(\omega) \epsilon L_1$. Lastly, 1.4 is exactly condition 3 of the theorem. Thus, $\Phi(\omega)$ is factorable; we shall denote its factor by $\Psi_1(\omega)$.

It is necessary that not only $\Phi(\omega)$ but also $G(\omega)$ be factorable. We note that $G(\omega) = E\left|H(\omega, \gamma)\right|^2 \geqslant 0$, and that $G(\omega) \epsilon L_1$ by 1.12. Then $G(\omega)$ is factorable if

$$\int_{-\infty}^{\infty} \frac{\log G(\omega)}{1 + \omega^2} d\omega > -\infty \tag{3.31}$$

We shall assume 3.31 to be true, deferring its proof for the moment. The factor of $G(\omega)$ will be called $G_1(\omega)$.

Consider now the product $\Phi(\omega)G(\omega)$. Since $G(\omega)$ is bounded according to 1.30, this product is in L_1 . Obviously, $\Phi(\omega)G(\omega)$ is a.e. non-negative. Furthermore,

$$\int_{-\infty}^{\infty} \frac{\log \Phi(\omega)G(\omega)}{1 + \omega^2} d\omega = \int_{-\infty}^{\infty} \frac{\log \Phi(\omega)}{1 + \omega^2} d\omega + \int_{-\infty}^{\infty} \frac{\log G(\omega)}{1 + \omega^2} d\omega > -\infty \tag{3.32}$$

because of 1.4 and 3.31 . Then $\Phi(\omega)G(\omega)$ is factorable. A straightforward computation shows that $\Psi_1(\omega)G_1(\omega)$ satisfies 3.4, 3.5 and

3.6, where $\Psi_1(\omega)$ and $G_1(\omega)$ are as defined above. Therefore, $\Psi_1(\omega)G_1(\omega)$ is the (unique) factor of $\Phi(\omega)G(\omega)$.

We now return to the proof of 3.31. We have

Theorem 5: Let $h(t, \gamma)$ be defined as in section one. Let 1.9 and 1.10 hold. Then 3.31 is satisfied.

Proof: By 1.9, Fubini's theorem applies to $h(t, \gamma)$ in the sense that

$$H(\omega) = \frac{1}{\sqrt{2\pi}} E \int_{-\infty}^{\infty} h(t, \gamma) e^{-i\omega t} dt = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} \left\{ E\left[h(t, \gamma) \right] \right\} e^{-i\omega t} dt \tag{3.33}$$

Thus, $H(\omega)$ is the Fourier transform of a function zero for a.e. $t < 0$. Also, 3.33 and the Schwarz inequality yield the result

$$\left| H(\omega) \right|^2 = \left| E\left[H(\omega, \gamma) \right] \right|^2 \leq E\left[\left| H(\omega, \gamma) \right|^2 \right] = G(\omega) \tag{3.34}$$

Then $G(\omega) \epsilon L_1$ (by 1.12) means that $H(\omega) \epsilon L_2$. In short, $H(\omega)$ satisfies the conditions of a theorem of Paley and Wiener* which is now used to assert that

$$\int_{-\infty}^{\infty} \frac{\log |H(\omega)|}{1 + \omega^2} d\omega > -\infty \tag{3.35}$$

Finally, we use 3.34 to write

$$\int_{-\infty}^{\infty} \frac{\log G(\omega)}{1 + \omega^2} d\omega \geq 2 \int_{-\infty}^{\infty} \frac{\log |H(\omega)|}{1 + \omega^2} d\omega > -\infty \tag{3.36}$$

which completes the proof of the theorem.

---

* Reference (4), pp. 16-20

The factorization which we have constructed is used to define a new subspace $\hat{\mathcal{M}}$. We shall define $\hat{\mathcal{M}}$ as the subspace (in $\mathcal{H}$) of all operators of the form

$$p(\omega) = \frac{\frac{1}{\sqrt{2\pi}} \underset{A \to \infty}{\text{l.i.m.}} \int_0^A P(t) e^{-i\omega t} dt}{\Psi_1(\omega) G_1(\omega)} \qquad P(t) \in L_2 \tag{3.37}$$

We see that an element of the above type is in $\mathcal{H}$. Furthermore, the fact that $\hat{\mathcal{M}}$ is closed in $\mathcal{H}$ norm is equivalent to the fact that the class of L_2 functions vanishing for a.e. negative argument is closed in $L_2(-\infty, +\infty)$ norm. Therefore, $\hat{\mathcal{M}}$ is indeed a subspace in $\mathcal{H}$.

The subspace $\hat{\mathcal{M}}$ is of great importance, for it is the subspace of all operators involving only the past of the input process. More practically, we have

<u>Theorem 6:</u> Let $\breve{\mathcal{M}}$ be the subspace (in $\mathcal{H}$) generated by operators of the type $\left\{e^{-i\omega h},\ h > 0\right\}$. Then

$$\hat{\mathcal{M}} = \breve{\mathcal{M}} \tag{3.38}$$

Proof: The proof of this theorem is omitted, since our theorem forms a part of a theorem of Doob. The reader is referred to Reference (3), pp. 586-587 (theorem 5.2).

We now assert that 3.37 is a representation of operators in $\bar{\mathcal{M}}$. The validity of this statement is a result of

<u>Theorem 7:</u> $\hat{\mathcal{M}} = \bar{\mathcal{M}}$.

Proof: It suffices to prove that $\breve{\mathcal{M}} = \bar{\mathcal{M}}$ (see theorem 6). If $\mathcal{S}$ is the family of step functions (in the time domain) defined in theorem 1, we have $\mathcal{M} \subset \bar{\mathcal{S}}$ by that theorem. But also, any step function is in L_1, so that $\mathcal{S} \subset \mathcal{M}$. Therefore* $\bar{\mathcal{M}} = \bar{\mathcal{S}}$.

To show that $\tilde{\mathcal{M}} \subset \bar{\mathcal{M}}$, we consider $e^{-i\omega h}$, $h > 0$. There exists a step function $S_a(t)$ with the Fourier transform $e^{-i\omega h}\frac{\sin \omega a}{\omega a}$. Now

$$\left\| e^{-i\omega h} - e^{-i\omega h}\frac{\sin \omega a}{\omega a} \right\|^2 \leq \int_{-A}^{A} \Phi(\omega)G(\omega) \left| 1 - \frac{\sin \omega a}{\omega a} \right|^2 d\omega$$

$$+ 2\left(\int_{\infty}^{-A} + \int_{A}^{\infty} \right) \Phi(\omega)G(\omega)\,d\omega \tag{3.39}$$

which can be made as small as desired by choosing A sufficiently large, and then a sufficiently small. Hence, $\tilde{\mathcal{M}} \subset \bar{\mathcal{S}} = \bar{\mathcal{M}}$.

Conversely, $\mathcal{M} \subset \tilde{\mathcal{M}}$. For we observe that $k \in \mathcal{M}$ has the form

$$k(\omega) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} K(t)e^{-i\omega t}\,dt \tag{3.40}$$

so that $k(\omega)$ is the limit of linear combinations of operators $e^{-i\omega h}$, $h > 0$. Given any finite ω interval and any $\varepsilon > 0$, we can find a linear combination $\sum_1^n c_j e^{-i\omega t}$, $0 < t_1 < \cdots < t_j < \cdots \leq t_n < \infty$, such that

$$\left| k(\omega) - \sum_1^n c_j e^{-i\omega t} \right| < \varepsilon \tag{3.41}$$

over the chosen finite interval. At the same time, the c_j are

---

* $\bar{\mathcal{M}} = \bar{\mathcal{S}}$ has already been proved by theorem 1; however, that proof required $\Phi(\omega)$ to be essentially bounded.

such that for every ω

$$\left| k(\omega) - \sum_{1}^{n} c_j e^{-i\omega t_j} \right| \leq \left| k(\omega) \right| + \sum_{1}^{n} \left| c_j \right| \leq 2 \int_{0}^{\infty} \left| K(t) \right| dt \qquad (3.42)$$

An argument similar to 3.39 completes the proof that $\mathcal{M} \subset \tilde{\mathcal{M}}$.

The representation 3.37 of operators in $\bar{\mathcal{M}}$ makes it easy to project $\mathcal{H}$ on $\bar{\mathcal{M}}$. In fact,

Theorem 8: The projection of $r \in \mathcal{H}$ on $\bar{\mathcal{M}}$ is given by

$$P(r) = p(\omega) = \frac{1}{2\pi \, \Psi_1(\omega) G_1(\omega)} \times$$

$$\underset{A \to \infty}{\text{l.i.m.}} \int_{0}^{A} e^{-i\omega t} \left[\underset{B \to \infty}{\text{l.i.m.}} \int_{-B}^{B} \Psi_1(\rho) G_1(\rho) r(\rho) e^{it\rho} \, d\rho \right] dt \qquad (3.43)$$

Proof: It is convenient to call $R(t)$ the inverse L_2 Fourier transform of $\Psi_1(\omega)G_1(\omega)r(\omega)$. Then 3.43 becomes

$$p(\omega) = \frac{1}{\sqrt{2\pi} \; \Psi_1(\omega) G_1(\omega)} \; \underset{A \to \infty}{\text{l.i.m.}} \int_{0}^{A} R(t) e^{-i\omega t} \, dt \qquad (3.44)$$

We also define a $q(\omega)$ such that $r(\omega) = p(\omega) + q(\omega)$. This operator is given by

$$q(\omega) = \frac{1}{\sqrt{2\pi} \; \Psi_1(\omega) G_1(\omega)} \; \underset{A \to \infty}{\text{l.i.m.}} \int_{-A}^{0} R(t) e^{-i\omega t} \, dt \qquad (3.45)$$

To show that $p(\omega)$ is the desired projection, we prove that p and q is the orthogonal decomposition of r with $p \in \hat{\mathcal{M}}$ and $q \in \hat{\mathcal{M}}^{\perp}$. That $p \in \hat{\mathcal{M}}$ is obvious from the definition 3.37. For

q , we shall demonstrate that q is orthogonal to every element of $\hat{\mathfrak{M}}$.

Given any $k \in \hat{\mathfrak{M}}$ there exists a $K(t) \in L_2$, $K(t) = 0$ for $t < 0$, such that

$$k(\omega) = \frac{1}{\sqrt{2\pi}\ \Psi_1(\omega) G_1(\omega)} \ \underset{A\to\infty}{\text{l.i.m.}} \int_{-A}^{A} K(t) e^{-i\omega t}\, dt \tag{3.46}$$

Also, we write $q(\omega)$ as

$$q(\omega) = \frac{1}{\sqrt{2\pi}\ \Psi_1(\omega) G_1(\omega)} \ \underset{A\to\infty}{\text{l.i.m.}} \int_{-A}^{A} Q(t) e^{-i\omega t}\, dt \tag{3.47}$$

where now

$$Q(t) = \begin{cases} 0 & t \geq 0 \\ R(t) & t < 0 \end{cases} \tag{3.48}$$

An application of Parseval's relation may be used to show that

$$(k, q) = \int_{-\infty}^{\infty} K(t) \overline{Q(t)}\, dt \tag{3.49}$$

But for a.e. t at least one of $K(t)$ and $Q(t)$ are zero. It follows that $(k, q) = 0$, as was to be shown.

If we project the optimum operator in $\mathcal{K}$ (as given by 1.28 i.e. $r(\omega) = \frac{\overline{H(\omega)} u(\omega)}{G(\omega)}$) on $\mathfrak{M}$ the optimum operator in $\overline{\mathfrak{M}}$ is found to be*

$$p(\omega) = \frac{1}{2\pi\ \Psi_1(\omega) G_1(\omega)} \ \underset{A\to\infty}{\text{l.i.m.}} \int_{0}^{A} e^{-i\omega t}$$

$$\left[\underset{B\to\infty}{\text{l.i.m.}} \int_{-B}^{B} \frac{\Psi_1(\rho) H(\rho) u(\rho)}{\overline{G_1(\rho)}} e^{it\rho}\, d\rho \right] dt \tag{3.50}$$

---

* That the optimum operator in $\overline{\mathfrak{M}}$ is unique, a.e. is a consequence of the unique nature of an orthogonal decomposition. See Reference (7).

in accordance with the preceding theorem. If this operator is used, the corresponding mean square error is computed to be

$$\mathcal{E}^2(p) - \mathcal{E}^2_{glb} = \| p - r \|^2 = \| (r-q) - r \|^2 = \| q \|^2 \quad (3.51)$$

by 1.29 and theorem 8 . We substitute for $\mathcal{E}^2_{glb}$ from 2.2 and evaluate $\| q \|$ to obtain

$$\mathcal{E}^2_{min} = \int_{-\infty}^{\infty} \Phi(\omega) \left| u(\omega) \right|^2 \left[1 - \frac{|H(\omega)|^2}{G(\omega)} \right] + \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \underset{A \to \infty}{\text{l.i.m.}} \int_{-A}^{A} \frac{\Psi_1(\rho)\overline{H(\rho)}u(\rho)}{\overline{G_1(\rho)}} e^{it\rho} \, d\rho \right|^2 dt \quad (3.52)$$

in which we have adopted the notation $\mathcal{E}^2_{min} = \inf_{k \in \mathfrak{M}} \left[\mathcal{E}^2(k) \right]$.

This completes the solution of the filter problem in the absence of noise inputs. The more general case of an input consisting of both signal and noise is treated in the next section of this paper.

SECTION IV

GENERAL PREDICTION AND FILTERING

We shall now consider filtering and prediction of inputs consisting of both signal and noise when each of these is subjected to a random operator. The random operators applied to the signal and noise may be the same or different, correlated or uncorrelated.

Let random operator $H(\omega, \gamma)$ be applied to the signal, and $M(\omega, \mu)$ be applied to the noise. Then the input to our filter is

$$z(t, \gamma, \mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H(\omega, \gamma) dX_s(\omega) + M(\omega, \mu) dX_n(\omega) \tag{4.1}$$

where the subscript s refers to signal and n refers to noise. The integrals, of course, are l.i.m. as before.

Let the signal and noise inputs $x_s(t)$ and $x_n(t)$ be uncorrelated, and ascribe to each of them the properties of $x(t)$ in section one. The total spectral density of the input to the filter is then

$$\Phi(\omega) = G(\omega)\,\Phi_s(\omega) + L(\omega)\Phi_n(\omega) \tag{4.2}$$

in which $L(\omega) = E\left|M(\omega, \mu)\right|^2$.

We wish $\Phi(\omega)$ to be factorable, $\chi_1(\omega)$ being its factor. $\Phi(\omega)$ meets conditions one and two of theorem 4. To satisfy condition three, it is necessary and sufficient that at least one of $\Phi_s(\omega)$ and $\Phi_n(\omega)$ satisfy this condition. Physically, this means that either signal or noise must be a non-deterministic process if signal plus noise is to be non-deterministic.

As before, we define a Hilbert space $\mathcal{H}$. The inner product is now

$$(r, s) = \int_{-\infty}^{\infty} \Phi(\omega) r(\omega) \overline{s(\omega)} d\omega \tag{4.3}$$

The work of sections two and three now follows with little, if any, modification. The optimum operator in $\mathfrak{M}$ becomes

$$p(\omega) = \frac{1}{2\pi \Psi_1(\omega)} \underset{A\to\infty}{\text{l.i.m.}} \int_0^A e^{-i\omega t} \left[\underset{B\to\infty}{\text{l.i.m.}} \int_{-B}^{B} \frac{\Phi_s(\rho)H(\rho)u(\rho)}{\overline{\chi_1(\rho)}} e^{it\rho} d\rho \right] dt \tag{4.4}$$

with a corresponding mean square error

$$\varepsilon^2_{min} = \int_{-\infty}^{\infty} \Phi_s(\omega) \left| u(\omega) \right|^2 \left[1 - \frac{\Phi_s(\omega) \left| H(\omega) \right|^2}{\Phi(\omega)} \right] d\omega + \frac{1}{2\pi} \int_{-\infty}^{0} \left| \underset{B\to\infty}{\text{l.i.m.}} \int_{-B}^{B} \frac{\Phi_s(\omega)\overline{H(\omega)}u(\omega)}{\overline{\chi_1(\omega)}} e^{it\omega} d\omega \right|^2 dt \tag{4.5}$$

We remark that $\bar{\mathfrak{M}} = \hat{\mathfrak{M}} = \check{\mathfrak{M}}$ remains unchanged; the manifolds of operators do not depend on the input process.

Of course, appropriate simplifications of the above equations yield the conventional Wiener filter formulae, or the filter derived in the previous section.

SECTION V

THE RANDOM GAIN AMPLIFIER

The operator $H(\omega, \gamma)$ has been made subject to certain assumptions (in section one) which exclude an important class of applications. We refer here to an $H(\omega, \gamma)$ of the form

$$H(\omega, \gamma) = \gamma \tag{5.1}$$

which has the physical interpretation that the signal (or input) is passed through an amplifier with random gain, or that the signal is derived from a transducer with a random scale factor.

Actually, operators such as 5.1 cause little theoretical difficulty. The restrictions of section one had the purpose of insuring that $E\left|H(\omega, \gamma)\right|^2$ be factorable, and that the process into the filter be of finite mean square. These two requirements are satisfied if $E\left|\gamma\right|^2 < \infty$. For convenience, we eliminate the case $\gamma = 0$ with probability one as trivial. Now let

$$E\gamma = \overline{\gamma} \tag{5.2}$$

and

$$E\left|\gamma\right|^2 = \overline{\gamma^2} \tag{5.3}$$

The general case of section four is not much simplified. Consider, however, the somewhat simpler assumption that both signal and noise are passed through the same random gain amplifier. Then in M_ω

$$k(\omega) = \frac{\bar{\gamma}}{\overline{\gamma^2}} k_W(\omega) \tag{5.4}$$

where $k_W(\omega)$ is the ordinary Wiener filter, viz.

$$k_W(\omega) = \frac{1}{2\pi \chi_1(\omega)} \int_0^\infty e^{-i\omega t} dt \int_{-\infty}^\infty \frac{\Phi_s(\rho)u(\rho)}{\overline{\chi_1(\rho)}} e^{it\rho} d\rho \tag{5.5}$$

the integrals being l.i.m. where the factor $\chi_1(\omega)$ constitutes the factorization of $\Phi_s(\omega) + \Phi_n(\omega)$.

The mean square error corresponding to this optimum filter is

$$\begin{aligned} \varepsilon^2_{min} = &\int_{-\infty}^\infty \Phi_s(\omega) \left| u(\omega) \right|^2 \left[1 - \frac{|\bar{\gamma}|^2}{\overline{\gamma^2}} \frac{\Phi_s(\omega)}{\Phi_s(\omega) + \Phi_n(\omega)} \right] d\omega \\ &+ \frac{|\bar{\gamma}|^2}{2\pi \overline{\gamma^2}} \int_{-\infty}^0 \left| \int_{-\infty}^\infty \frac{\Phi_s(\omega)u(\omega)}{\overline{\chi_1(\omega)}} e^{it\omega} d\omega \right|^2 dt \end{aligned} \tag{5.6}$$

We note that if $\bar{\gamma} = 0$ the optimum operation becomes $k(\omega) = 0$, and the mean square error is $\int_{-\infty}^\infty \Phi_s(\omega)|u(\omega)|^2 d\omega$. This is in accordance with the intuitive idea that if a number is equally likely to be positive or negative we guess zero if the penalty for an incorrect guess is the square of the difference between the actual number and the guess.

SECTION VI

PREDICTING AND LAGGING FILTERS

One of the most important applications of optimum filter theory lies in predicting the value of a random input process $x(t)$ (perhaps corrupted by noise) over some prediction time we shall call a . This type of prediction has been used where some component or natural phenomena introduces a lag between perception of $x(t)$ and decisions made or action applied as a result. Typical examples include anti-aircraft gun computers which must account for the motion of the target while the shell is on its way, and an automatic landing technique to be used in connection with an aircraft carrier in a stormy sea.

At the other extreme, lagging filters are useful also. Such a filter has the advantage that as the lag between output and input is increased the mean square error is lessened. For many purposes, a lag is quite acceptable and the delay in obtaining the output is gladly exchanged for enhanced accuracy. We shall give two examples where lagging filters are appropriate. On missile test flights performance data is telemetered to a ground station where it is reduced at some later date from a tape recording made during the flight. If the playback process involves a (reasonable) delay time no harm is done. The same argument may be applied to the playback of recorded music. In fact, a delay or lagging filter is applicable wherever the data in question need not be available instantaneously.

It was stated in 1.7 and the text of section one that

$$u(\omega) = e^{i\omega a} \tag{6.1}$$

constitutes a lead or lag operator. If a is positive, $u(\omega)$ corresponds to lead or prediction. Analogously, $u(\omega)$ is a delay or lag operator whenever $a < 0$*. The physical interpretation of $u(\omega)$ as in 6.1 is that the filter output $y(t)$ should lead or lag the random process $x(t)$ by a time interval $|a|$, the choice of lead or lag being determined by the sign of a.

If $u(\omega)$ is given by 6.1, the minimum mean square error in $\overline{\mathcal{M}}$ is called $\varepsilon^2_{min}(a)$, and the corresponding operation in $\overline{\mathcal{M}}$ will be denoted by $k_a(\omega)$. These are given from 4.5 and 4.4 by

$$\varepsilon^2_{min}(a) = \int_{-\infty}^{\infty} \Phi_s(\omega)\left[1 - \frac{\Phi_s(\omega)\,|H(\omega)|^2}{\Phi_s(\omega)G(\omega) + \Phi_n(\omega)L(\omega)}\right] d\omega + \int_{-\infty}^{0} \left| \int_{-\infty}^{\infty} \frac{\Phi_s(\omega)\overline{H(\omega)}}{\sqrt{2\pi}\,\overline{\chi_1(\omega)}} e^{i(t+a)\omega}\, d\omega \right|^2 dt \tag{6.2}$$

and

$$k_a(\omega) = \frac{1}{2\pi\,\chi_1(\omega)} \int_0^{\infty} e^{-i\omega t}\, dt \int_{-\infty}^{\infty} \frac{\Phi_s(\rho)\overline{H(\rho)}}{\overline{\chi_1(\rho)}} e^{i(t+a)\rho}\, d\rho \tag{6.3}$$

We may now assert

<u>Theorem 11:</u> $\varepsilon^2_{min}(a)$ is a monotone nondecreasing function of a.

---

* Our concept of a lag or delay may differ from that sometimes found in electrical engineering usage. We mean a <u>time</u> delay, i.e., a delay independent of frequency. The corresponding phase delay is then proportional to frequency.

Also

$$\inf_a \left[\mathcal{E}^2_{min}(a) \right] = \int_{-\infty}^{\infty} \Phi_s(\omega) \left[1 - \frac{\Phi_s(\omega) \left| H(\omega) \right|^2}{\Phi_s(\omega) G(\omega) + \Phi_n(\omega) L(\omega)} \right] d\omega \qquad (6.4)$$

and

$$\sup_a \left[\mathcal{E}^2_{min}(a) \right] = \int_{-\infty}^{\infty} \Phi_s(\omega) d\omega \qquad (6.5)$$

Proof: From the work of sections one and three, the mean square error in question may be expressed as

$$\mathcal{E}^2_{min}(a) = \mathcal{E}^2_{glb} + \left\| k_a - r \right\|^2 \qquad (6.6)$$

where $r(\omega) \epsilon H_\omega$ is

$$r(\omega) = \frac{\Phi_s(\omega) \overline{H(\omega)} e^{i\omega a}}{\Phi_s(\omega) G(\omega) + \Phi_n(\omega) L(\omega)} \qquad (6.7)$$

and

$$\mathcal{E}^2_{glb} = \int_{-\infty}^{\infty} \Phi_s(\omega) \left[1 - \frac{\Phi_s(\omega) \left| H(\omega) \right|^2}{\Phi_s(\omega) G(\omega) + \Phi_n(\omega) L(\omega)} \right] d\omega \qquad (6.8)$$

The theorem is proved if we show that $\left\| r - k_a \right\|^2$ is monotone and

$$\lim_{a \to -\infty} \left\| r - k_a \right\|^2 = 0 \qquad (6.9)$$

$$\lim_{a \to \infty} \left\| r - k_a \right\|^2 = \left\| r \right\|^2 \qquad (6.10)$$

It is convenient to rewrite 6.3 as

$$k_a(\omega) = \frac{1}{2\pi \mathcal{X}_1(\omega)} \int_a^{\infty} e^{-i\omega t} dt \int_{-\infty}^{\infty} \frac{\Phi_s(\rho) \overline{H(\rho)}}{\mathcal{X}_1(\rho)} e^{-t\rho} d\rho \qquad (6.11)$$

by a change of variable $t + a = r$. The integrals in 6.11 are to be taken l.i.m. In the same sense, the substitution of 6.7 into 6.11 yields

$$k_a(\omega) = r(\omega) - \frac{1}{2\pi \mathcal{X}_1(\omega)} \int_{-\infty}^{a} e^{-i\omega t} dt \int_{-\infty}^{\infty} \frac{\Phi_s(\rho) H(\rho)}{\overline{\mathcal{X}_1(\rho)}} e^{it\rho} d\rho \qquad (6.12)$$

from which

$$\| r - k_a \|^2 = \int_{-\infty}^{a} | R(t) |^2 \, dt \qquad (6.13)$$

$R(t)$ being the inverse Fourier transform in L_2 of $\mathcal{X}_1(\omega) r(\omega)$.

It is clear that $\| r - k_a \|$ is monotone. In addition, it follows from 6.13 that

$$\lim_{a \to -\infty} \| r - k_a \|^2 = \lim_{a \to -\infty} \int_{-\infty}^{a} | R(t) |^2 \, dt = 0 \qquad (6.14)$$

and

$$\lim_{a \to \infty} \| r - k_a \|^2 = \int_{-\infty}^{\infty} | R(t) |^2 dt = \int_{-\infty}^{\infty} | \mathcal{X}_1(\omega) r(\omega) |^2 d\omega = \| r \|^2 \qquad (6.15)$$

thereby completing the proof of the theorem.

A supplementary result of 6.15 is that

$$\lim_{a \to \infty} \| k_a \| = 0 \qquad (6.16)$$

or

$$\underset{a \to \infty}{\text{l.i.m.}} \left[k_a(\omega) \right] = 0 \qquad (6.17)$$

This result, of course, is also true in the case of the conventional Wiener filter.

SECTION VII

AN EXAMPLE: RANDOM DELAY TIME

Every communication system involves delay between the transmission and reception of messages. Inevitably, the information is propagated through some medium so that there is a lag directly proportional to the distance between transmitter and receiver, and inversely proportional to the propagation velocity.

Two persons conversing together experience such a delay without inconvenience. Over larger distances, and in more critical applications, the propagation delay can greatly reduce the value of the data received. In that case, Wiener filter theory yields the result that the best estimate (in the mean square sense) of the present data is achieved by passing the delayed data through a Wiener optimum prediction filter whose prediction time is exactly equal to the propagation delay.

The transmission range and therefore the delay time is likely to vary where mobile equipment is used, or where a target is tracked. A variable prediction time filter is then called for. However, range measurements may be difficult, or a fixed prediction filter may be preferable for reasons of simplicity or economy. The optimum fixed prediction filter must then represent a compromise over the transmission ranges which will be encountered in practice.

The optimum fixed prediction filter referred to above is easily derived by the methods of this paper. Let γ be the propagation time (delay), and let $F(\gamma)$ be its probability distribution function. Then

$$H(\omega, \gamma) = e^{-i\omega\gamma} \qquad \gamma \geqslant 0 \tag{7.1}$$

from which

$$H(\omega) = \int_0^\infty e^{-i\omega\gamma}\, dF(\gamma) = \overline{\varphi(\omega)} \tag{7.2}$$

where $\varphi(\omega)$ is the characteristic function of $F(\gamma)$, as generally defined. We also have, by an obvious calculation

$$G(\omega) = 1 \tag{7.3}$$

If there is no noise present the optimum operator in $\mathfrak{M}$ is

$$k(\omega) = \frac{1}{2\pi\,\Psi_1(\omega)} \int_0^\infty e^{-i\omega t} dt \int_{-\infty}^\infty \Psi_1(\rho)\varphi(\rho)e^{it\rho}\, d\rho \tag{7.4}$$

with the integrals being defined l.i.m. The mean square error becomes

$$\begin{aligned} \mathcal{E}^2_{\min} &= \int_{-\infty}^\infty \Phi(\omega)\left[1 - |\varphi(\omega)|^2\right] d\omega \\ &+ \frac{1}{2\pi} \int_{-\infty}^0 \left| \int_{-\infty}^\infty \Psi_1(\rho)\varphi(\rho)e^{it\rho} d\rho \right|^2 dt \end{aligned} \tag{7.5}$$

If the operation 7.4 were not applied to the signal the error due to the delay is calculated to be

$$\mathcal{E}^2_{(1)} = \int_{-\infty}^\infty \Phi(\omega)\left[1 - |\varphi(\omega)|^2\right] d\omega + \int_{-\infty}^\infty \Phi(\omega)\, |1 - \varphi(\omega)|^2 d\omega \tag{7.6}$$

which must be larger than 7.5; this difference in mean square error is

$$\mathcal{E}^2(1) - \mathcal{E}^2_{\min} = \int_{-\infty}^{\infty} \Phi(\omega) \left| 1 - k(\omega) \right|^2 d\omega \qquad (7.7)$$

where $k(\omega)$ is given by 7.4. It should be noted that $k(\omega) = 1$ a.e. if and only if $\varphi(\omega) \equiv 1$, that is, the delay is zero with probability one.

A classical example of Wiener optimum prediction may be described as follows: an anti-aircraft battery is equipped with a radar tracking system which determines the position and velocity of enemy aircraft. In one dimension, the aircraft velocity is stationary with spectral density

$$\Phi(\omega) = \frac{1}{1 + \omega^2} \; ^{*}.$$

The time required for the shell to attain the altitude of the aircraft is γ (γ is fixed for the present). If the anti-aircraft shell is fired at time t we must be able to predict the aircraft position at time $t + \gamma$. In other words, we require a prediction filter to compensate as nearly as possible for the delay γ inherent in the system.

Unfortunately, the aircraft position $z(t)$ is not a stationary process, so that the techniques of this paper are not immediately applicable. However, we shall show that $z(t + \gamma) - z(t)$ is stationary. It suffices to work with $z(t + \gamma) - z(t)$ since the radar measures $z(t)$ directly and exactly. Then our best estimate of $z(t + \gamma)$ is obviously attained by adding the best estimate of

---

* The appropriateness of this form of spectral density is justified in Reference (8), pp. 300 - 304.

$z(t+\gamma) - z(t)$ to the known value of $z(t)$.

In what follows, our choice of notation will be consistent (insofar as possible) with the notation used in previous sections. Accordingly, let $x(t)$ be the stationary process of aircraft velocity.

To investigate the properties of $z(t+\gamma) - z(t)$, let

$$h(t) = \begin{cases} -\sqrt{2\pi} & -\gamma \leqslant t < 0 \\ 0 & t \geqslant 0 \end{cases} \tag{7.8}$$

It is clear that $h(t) \in L_1$. Indeed, $h(t)$ has the L_1 Fourier transform

$$H(\omega) = \frac{1 - e^{i\gamma\omega}}{i\omega} \tag{7.9}$$

so that some stationary process, say $s(t)$, has the representation

$$s(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H(\omega)\, dX(\omega) \tag{7.10}$$

the integral being defined l.i.m. We also have from 1.8

$$\begin{aligned} s(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t-\tau) h(\tau)\, d\tau \\ &= -\int_{-a}^{0} x(t-\tau)\, d\tau = z(t+\gamma) - z(t) \end{aligned} \tag{7.11}$$

In this application, the input available to the optimum filter is the velocity, $x(t)$. The filter output will be denoted by $w(t)$, as in the first section. The optimum operator in $\mathcal{H}$ is designated by $r^*(\omega)$, and the optimum operator in $\mathfrak{M}$ is called $k^*(\omega)$. Then

$$\mathcal{E}(t; k^*) = w(t) - \Big[z(t+\gamma) - z(t) \Big] \tag{7.12}$$

w(t) having the representation

$$w(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} k^*(\omega)\, dX(\omega) \tag{7.13}$$

The mean square error becomes

$$\mathcal{E}^{*2}(k^*) = \int_{-\infty}^{\infty} \Phi(\omega) \left| H(\omega) - k^*(\omega) \right|^2 d\omega \tag{7.14}$$

and this error is minimized by

$$k^*(\omega) = \frac{1}{2\pi \Psi_1(\omega)} \int_0^{\infty} e^{-i\omega t} dt \int_{-\infty}^{\infty} \Psi_1(\rho) \frac{1 - e^{i\gamma\rho}}{i\rho} e^{it\rho} d\rho \tag{7.15}$$

The formula 7.15 is identical with that of Wiener.

If the spectral density $\Phi(\omega)$ is rational the inner integral in 7.15 may be evaluated by a contour integration in the upper half plane. The result is the sum of terms of the form $t^k e^{-a_n t}$, so that the second integration can be done by parts or through use of readily available integral tables. In our case, $\Phi(\omega) = \frac{1}{1 + \omega^2}$ so that the above procedure yields

$$k^*(\omega) = 1 - e^{-\gamma} \tag{7.16}$$

We note particularly that for small γ the use of the approximation $1 - e^{-\gamma} \simeq \gamma$ implies that when the prediction interval is short the best estimate of $z(t + \gamma) - z(t)$ is obtained by assuming that the aircraft continues at its present velocity $x(t)$ throughout the prediction interval γ.

The filter 7.15 is indeed optimum if it can be assumed that all enemy aircraft will fly at the altitude corresponding to shell flight time γ. Since the enemy is not expected to be so obliging, we must take into account the possibility that aircraft will come in at various altitudes. This situation can be dealt with by providing an adjustable filter which varies in accordance with the flight time of the shell as calculated from the aircraft altitude measured by the radar set. However, a fixed filter may be preferable for reasons indicated at the beginning of this section. Such a filter attempts to predict $z(t + \gamma) - z(t)$ where the aircraft altitude and thus γ are random variables. In analogy to 7.9

$$H(\omega, \gamma) = \frac{1 - e^{i\gamma\omega}}{i\omega} \tag{7.17}$$

and as in 7.14

$$\mathcal{E}^2(k) = \mathop{E}_{\gamma} \int_{-\infty}^{\infty} \Phi(\omega) \left| H(\omega, \gamma) - k(\omega) \right|^2 d\omega \tag{7.18}$$

where $\mathop{E}_{\gamma}$ denotes expectation with respect to γ. If γ assumes some given value with probability one, 7.18 reduces to 7.14 and the optimum filter is given by 7.15 as has also been shown by Wiener. In the more general case (where the aircraft may fly at more than one altitude) we suppose that γ is associated with the probability distribution function $F(\gamma)$. The only restriction placed on γ and $F(\gamma)$ are that $\gamma \geqslant 0$ with probability one, and that $F(\gamma)$ is such that

$$\int_0^{\infty} \gamma^{1 + \delta} \, dF(\gamma) < \infty \tag{7.19}$$

for some $\delta > 0$.

It is convenient to rewrite 7.18 as

$$\mathcal{E}^2(k) = \int_{-\infty}^{\infty} \Phi(\omega) \left| k(\omega) - \frac{1 - \varphi(\omega)}{i\omega} \right|^2 d\omega + \mathcal{E}^2_{glb} \qquad (7.20)$$

in which

$$\mathcal{E}^2_{glb} = \int_{-\infty}^{\infty} \frac{\Phi(\omega)}{\omega^2} \left[1 - \left| \varphi(\omega) \right|^2 \right] d\omega \qquad (7.21)$$

Here $\varphi(\omega)$ is again the characteristic function belonging to $F(\gamma)$. The computation of 7.20 and 7.21 is quite similar to the work of section three. But in addition it is also necessary to demonstrate the integrability of the integrands of 7.20 and 7.21. Since $|\varphi(\omega)| \leq 1$ the only question remaining is that of the behavior of $\frac{1 - \varphi(\omega)}{\omega}$ and $\frac{1 - |\varphi(\omega)|^2}{\omega^2}$ near $\omega = 0$. Let

$$\int_{0}^{\infty} \gamma dF(\gamma) = m \qquad (7.22)$$

Then $\left| \frac{1 - \varphi(\omega)}{\omega} \right|^2$ and $\frac{1 - |\varphi(\omega)|^2}{\omega^2}$ are integrable, for in the vicinity of $\omega = 0$, the expansion

$$\varphi(\omega) = 1 + i\omega m + O(\omega^{1 + \delta}) \qquad (7.23)$$

is valid*. It should be noted that the condition 7.19 is a sufficient but not necessary one for 7.20 and 7.21 to exist. For example, the Couchy distribution has no finite first moment; yet the characteristic function is such that $\frac{1 - \varphi(\omega)}{\omega}$ is bounded near $\omega = 0$. Of course,

---

*This is a well known theorem in probability theory. See Reference (9), p. 199.

any distribution of γ is acceptable if, for any $A > 0$,

$$\int_0^A \frac{\Phi(\omega)}{\omega^2}\, d\omega < \infty \quad .$$

Proceeding as in section three, we see that the element in $\overline{\mathfrak{M}}$ which minimizes the mean square error 7.20 is given by

$$k(\omega) = \frac{1}{2\pi \Psi_1(\omega)} \int_0^\infty e^{-i\omega t} dt \int_{-\infty}^\infty \Psi_1(\rho) \frac{1 - \varphi(\rho)}{i\rho} e^{it\rho}\, d\rho \qquad (7.24)$$

the integrals being taken as l.i.m. If the aircraft velocity spectrum $\Phi(\omega) = \frac{1}{1+\omega^2}$ as was assumed in the calculation of the conventional Wiener filter 7.16 the optimum operator 7.24 specializes to

$$k(\omega) = 1 - \varphi(i) = 1 - \int_0^\infty e^{-\gamma} dF(\gamma) \qquad (7.25)$$

This result is easily established by contour integration of the first integral in 7.24. The upper half plane is used since the integral along a sufficiently large semi-circle converges to zero. In this regard it should be noted that

$$\varphi(z) = \int_0^\infty e^{iz\gamma} dF(\gamma) \qquad (7.26)$$

is regular and bounded in the upper half plane*.

It is instructive to note the increase in the mean square error when the Wiener filter $k^*(\omega)$ is used in place of the $k(\omega)$ of 7.24. For the latter we obtain

---

* This fact may be established by a modification of a proof found in Reference (10), p. 8.

$$\mathcal{E}^2_{min} = \int_{-\infty}^{\infty} \frac{\Phi(\omega)}{\omega^2}\left[1 - |\varphi(\omega)|^2\right] d\omega$$

$$+ \frac{1}{2\pi}\int_{-\infty}^{0} \left| \int_{-\infty}^{\infty} \frac{\Psi_1(\omega)\left[1 - \varphi(\omega)\right]}{i\omega} e^{it\omega} d\omega \right|^2 dt \qquad (7.27)$$

which is analogous to 3.47 and is derived in an identical manner. Further calculations then lead to an equation like 1.28, or

$$\mathcal{E}^2(k^*) - \mathcal{E}^2_{min} = \int_{-\infty}^{\infty} \Phi(\omega) \left| k^*(\omega) - k(\omega) \right|^2 d\omega$$

$$= \frac{1}{2\pi}\int_{0}^{\infty} \left| \int_{-\infty}^{\infty} \frac{\Psi_1(\omega)\left[\varphi(\omega) - e^{i\gamma\omega}\right]}{i\omega} e^{it\omega} \right|^2 dt \qquad (7.28)$$

In the particular case $\Phi(\omega) = \frac{1}{1+\omega^2}$ we have $k^*(\omega)$ and $k(\omega)$ from 7.16 and 7.24, respectively. It is seen that 7.28 becomes

$$\mathcal{E}^2(k^*) - \mathcal{E}^2_{min} = \left| e^{-a} - \int_{0}^{\infty} e^{-\gamma} dF(\gamma) \right|^2 \times \int_{-\infty}^{\infty} \Phi(\omega) d\omega \qquad (7.29)$$

where the Wiener filter $k^*(\omega)$ has been designed on the assumption that $\gamma = a$. If the mean square error is defined as in this paper the Wiener filter is seen to be optimum if and only if $\gamma = a$ with probability one. Then also $k(\omega) = k^*(\omega)$ a.e.

REFERENCES

1. Kolmogorov, Bull. Acad. Sci. U.S.S.R. Ser. Math. (1941), 5, pp. 5-14
2. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series", Wiley and Sons (1949)
3. Doob, "Stochastic Processes", Wiley and Sons, Inc. (1953)
4. Paley and Wiener, "Fourier Transforms in the Complex Domain", AMS Colloquium Publication Volume XIX, American Mathematical Society (1934)
5. Bieberbach, "Lehrbuch der Funktionentheorie" (2 volumes), Chelsea Publishing Co. (1945)
6. Kryloff, Rec. Math. (Mat. Sbornik) N. S. (1939), 6, p. 106
7. Halmos, "Introduction to Hilbert Space and the Theory of Spectral Multiplicity", Chelsea Publishing Co. (1951)
8. James, Nichols, and Phillips, "Theory of Servomechanisms", M.I.T. Radiation Laboratory Series No. 25, McGraw-Hill Book Co., Inc. (1947)
9. Loeve, "Probability Theory", Van Nostrand Co., Inc. (1955)
10. Bochner and Chandrasekharan, "Fourier Transforms", Annals of Mathematics Studies No. 19, Princeton University Press (1949