ON PREDICTION AND FILTERING OF TIME SERIES

by

Sureshchandra L. Vasa

B. E., 1963
University of Bombay, India

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements of the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1965

Approved by:

Sachy S. Haye
Major Professor

TABLE OF CONTENTS

INTRODUCTION 1

TOPICS ON GENERAL RELATIONS 2

DIFFERENT KINDS OF ERROR CRITERIA 6

STATIONARY TIME SERIES 8

Prediction and Filtering 8

Wiener's Method 8

Mean Square Error 11

Approximation of Wiener's Method 12

Bode-Shannon Method 17

Mean Weighted Square Error Criterion 20

NON-STATIONARY TIME SERIES 22

An Approach by Ragazzini and Zadeh 22

A General Problem of Non-Stationary Time Series 25

CONCLUSION 28

ACKNOWLEDGMENT 29

REFERENCES 30

INTRODUCTION

In many fields of engineering, various types of messages need to be transferred from place to place. This is done by different types of transmission. The messages may be dots or dashes as in Morse code or sound wave pattern as in radio transmission or waves containing images as in telephoto or television. The message to be transmitted is developed into a time series by means of different operations such as coding, scanning, etc. In communication engineering the messages are converted into electrical voltages or currents. The time series is transmitted from one end to other end by means of a transmission device. The time series passes through a number of stages in transmission. At every stage an operation is performed on the time series transforming it to another time series. Also the time series is distorted due to thermal noise, tracking errors and poor characteristics of transmitting or receiving equipments.

It is important that the information received at the receiving end of the transmission line should yield the same information as was transmitted. The detection problem is to obtain the original time series from the received time series. In some cases a future estimate of the time series is required, from the knowledge of the past values of the time series. The two problems may be combined, as to separate the original time series and then to estimate a future value of the time series.

TOPICS ON GENERAL RELATIONS

Time Series. A time series is a sequence of quantitative data assigned to specific points in time domain. A time series may represent the sequences of electrical voltages or currents, or it may be some data such as stock prices, weather reports, etc. Usually the sequences are assigned over the certain interval of time and the time series can be assumed to be specified over time varying from $-\infty$ to $+\infty$.

The sequences may be discrete or continuous. If the quantitative data is assigned over discrete values of time, the time series is known as a Discrete Time Series. Similarly if the quantitative data is assigned over every value of time, the time series is known as a Continuous Time Series.

Prediction. Prediction is also known as extrapolation. In general, prediction can be said to be an estimation of the future value of a time series when its past values or characteristics are known. An operator is a device which operates on the time series and converts it into another time series. First an operator is assumed to give a time series; then the error, which is the difference between the desired output and the obtained output, is minimized using a particular error criterion and a suitable operator is found. If $s(t)$ is a time series, then the prediction is an estimate of the most probable value of $s(t+T)$, where T 0, obtained from the past characteristics of $s(t)$.

Filtering. The received time series has been altered from the transmitted time series due to the different stages of the transmission of the time series and the noise in the transmitting and receiving equipment. The received time series can be represented as a sum of $s(t)$ and $n(t)$, where $s(t)$ is transmitted time series and $n(t)$ is another time series resulting due to noise and the different operations during transmission. The separation of $s(t)$ from $\left[s(t)+n(t)\right]$ is known as filtering.

Many times the problems of filtering and prediction are combined. Sometimes it is necessary to obtain $s(t+T)$ from $[s(t)+n(t)]$. If T is positive, the problem is a combined problem of filtering and predicting and the operator is known as a filter with lead characteristics. If T is negative, then the operator is known as a filter with lag characteristics.

Autocorrelation Function. The autocorrelation function for a discrete time series

$x_{-2}, x_{-1}, x_0, x_1, x_2, x_3,$ .... is given by

$$\phi(t_1, t_2) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{+N} x_{t+t_2} \cdot x^*_{t+t_1} \tag{1}$$

where x^* is conjugate of x.

For a continuous time series s(t), the autocorrelation function is given by

$$\phi(t_1,t_2) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} s(t+t_2) \quad s^*(t+t_1)\, dt \tag{2}$$

The mean value M of a discrete time series is given by

$$M_1 = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{+N} x_t$$

whereas that of a continuous time series is

$$M_1 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} s(t)\, dt$$

Stationarity. If, for a particular time series, the autocorrelation function $\phi(t_1,t_2)$ is a function of the difference between t_2 and t_1 and the mean value M_1 is a constant, then the time series is known as 'stationary in wide sense' (Yaglom, 1962). In this report such a time series shall be referred to simply as a 'stationary time series'. The autocorrelation of a stationary time series can be written as

$$\phi(t_1,t_2) = \phi(t_2-t_1) = \phi(\tau) \qquad \text{where} \quad \tau = t_2-t_1,$$

or

$$\phi(\tau) = \lim_{T\to\infty} \frac{1}{2T+1} \sum_{t=-T}^{+T} x_{t+\tau} \cdot x^*_t \quad \text{for discrete time series} \tag{3}$$

and

$$\phi(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} s(t+\tau) \quad s^{*}(t) \quad dt$$

for continuous time series (4)

If the above requirements are not satisfied by a time series, then the time series is known as a non-stationary time series.

Cross-correlation Function. Let $x(t)$ and $y(t)$ be any two time series. Then the cross-correlation function is defined as

$$\phi_{xy}(t,T) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{+N} x_{t+T} \; y^{*}(t) \text{ for discrete time series} \quad (5)$$

and

$$\phi_{xy}(t,T) = \lim_{N\to\infty} \frac{1}{2N} \int_{-N}^{+N} x(t+T)\; y^{*}(t)\; dt \text{ for continuous time}$$

series (6)

If the cross-correlation between the two series is zero then the two time series are known as statistically independent of each other.

Power Spectral Density for the Stationary Time Series. For a stationary time series the auto correlation function and the power spectral density are the Fourier transformations of each other. Let $\phi(T)$ be the autocorrelation function of the stationary time series and $G(w)$ be its power spectral density function; then

$$G(w) = \int_{-\infty}^{+\infty} \phi(T)\; e^{-jwT} \; dT \quad (7)$$

and by inverse Fourier transformation

$$\phi(T) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(w) \quad e^{jwT} \; dw \quad (8)$$

General Discussion. Physical prediction or filtering depends basically on the past values and is supposed to represent most probable value in the future. Although this argument cannot be supported by rigorous mathematics, yet like the laws of physics, the prediction and filtering, although on much weaker basis, are based on statistical results.

White Noise. When the autocorrelation function of a time series is an impulse function, and therefore its power spectral density is a constant, the time series is known as a white noise. In practice, it is not possible to generate such a stationary time series, as this will require infinite power. Certain time series can be approximated by time series having white noise over the range of frequencies of interest.

Weighting Function. In frequency domain an operator is represented by the transfer function of the operator. Similarly in the time domain the operator is represented by its weighting function. A weighting function $w(t,T)$ is defined as the output at time T after a unit impulse function is applied to the operator at time t. The system is assumed at rest prior to the time when the impulse is applied. For the linear operator $W(t,T)$ the weighting function is the function of $(t-T)$, and can be written as

$$W(t,T) = W(t-T)$$

If the input to the operator having weighting function $W(t-T)$ is $x(t)$, then the output $y(t)$ is given by

$$y(t) = \int_{-\infty}^{t} W(t-T)x(T)dT \tag{9}$$

or

$$y(t) = \int_{0}^{\infty} W(T)x(t-T)dT \tag{10}$$

Shaping Filter. Let $s(t)$ be a stationary time series with power spectral density $G(w)$, then the output of the shaping filter for the time series $s(t)$ is the itself when its input is white noise.

If $G_{xx}(w)$ is the power spectral density of the input time series to an operator with transfer function $Y(jw)$ and the output is $y(t)$ with power spectral density as $G_{yy}(w)$, then

$$G_{yy}(w) = |Y\ (jw)|^2\ G_{xx}(w) \tag{11}$$

The spectral density $G_{xx}(w)$, for a shaping filter, is unity, and therefore the above equation becomes

$$|Y(jw)|^2 = G_{yy}(w) \tag{12}$$

If $G_{yy}(w)$ satisfies the expression

$$\int_{-\infty}^{+\infty} \frac{\left|\log |G_{yy}(w)|\right|}{1+w^2}\, dw < \infty \qquad (13)$$

then $G_{yy}(w)$ can be written as $G_{yy}(w) = |H_1(w)|^2$, where $H_1(w)$ has no singularities in the lower half plane or below the real axis of w-plane, and hence

$Y(jw) = H_1(w)$

If $Y(jw)$ is the shaping filter for the time series $s(t)$, the inverse of shaping filter is given as $Y^{-1}(jw)$. The output of an inverse shaping filter is white noise when its input is the time series $s(t)$.

DIFFERENT KINDS OF THE ERROR CRITERIA

In order to obtain satisfactory performance of a system, it is conventional to minimize the system output error. For a time series, it is physically unrealizable to minimize the error corresponding to all the values of time. An error weighting function is defined in order to measure the system performance with reference to its optimum performance. The average error function is minimized, then the optimal system is defined.

The average error may seem to be a useful criterion, but it has a serious disadvantage that the positive and negative values cancels each other in the averaging process. Hence another appropriate error function of the magnitude of the error is required to be minimized in order to obtain optimum performance of the system. Some of those error criteria are discussed below.

Laning and Battin (1956) show that s simple error function may be defined as

$$F(e) = \begin{cases} 0 & \text{if } e < a \\ 1 & \text{if } e \geq a \end{cases}$$

where an error below a particular value a is neglected while errors greater thana a are treated equal while the average is taken.

Wiener (1949) used the mean square error criterion using the error function

$F(e) = e^2(t)$

If $C_d(t)$ is the desired output and $C_a(t)$ is the actual output, then the error is $e(t) = C_d(t)-C_a(t)$, and the mean value of $F(e)$ is given as

$$E = \overline{F(e)} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left| C_d(t)-C_a(t)\right|^2 dt \tag{14}$$

Wiener's work has been followed by investigation of a number of different error functions, some of which are discussed below. Some criteria are primarily valid for deterministic inputs, or the value of the error function is zero for negative values of t. These can be applied to the time series with proper modifications. For instance Graham and Lanthrop (1953) proposed the use of the "time multiplied absolute error" function as given below

$$F(e) = t|e(t)|$$

The average error E is then

$$E = \int_0^\infty t|e(t)| \quad dt \tag{15}$$

Nims (1951) used the time multiplied error to obtain the mean error

$$E = \int_0^\infty T \quad e(T) \quad dT \tag{16}$$

Schultz and Rideout (1957) discussed a general case

$$F(e) = W(t-T) \quad A(e(t,T))$$

where $A(e(t,T))$ is an arbitrary function of $e(t,T)$, $W(t)$ is taken as a suitable function of t, and $e(t)$ is made function of t and T by using input function as the function of t and T, resulting in

$$E = \int_0^\infty W(t-T) \quad A(e(t,T)) \; dT \tag{17}$$

Spooner and Rideout (1956) considered a special case of above by taking $W(t)$ as 1, and $A(e(t,T))$ as $(e(t,T))^2$. Murphy and Bold (1960) analyzed a case proposed by Schultz and Rideout by taking T=0 and $A(e(t,T))$ as $e^2(t)$. This yields the mean error as

$$E = \int_0^\infty W(t) \; e^2(t) \; dt \tag{18}$$

and for estimation of E from a truncated (-T,T) portion of the time series it is modified to

$$E = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} W(t)\, e^2(t)\, dt \tag{19}$$

This is known as 'Integral of the Weighted Square Error' (IWSE Criterion).

STATIONARY TIME SERIES

Prediction and Filtering for Stationary Time Series

For a stationary time series the statistical characteristics are invariant for all the sections of the time series. This characteristic of a stationary time series is used to obtain an operator to improve accuracy. The autocorrelation function $\phi(t_1,t_2)$ of a stationary time series is a function of t_2-t_1 only, and it is sufficient to characterise the time series.

Wiener's Method

The following assumptions are used by Wiener in his work:

(1) Both signal and noise time series are assumed to be stationary time series.

(2) The operation being performed on the time series is assumed to be a linear operation on the past history of the given signal and noise series.

(3) The linear operator is physically realizable by employing mechanical or electrical circuits.

(4) The mean-square-error criterion is used.

The first assumption about stationarity of the time series is usually valid in practice and hence the knowledge of the autocorrelation function is sufficient.

Bode and Shannon (1950) show that a linear operator is usually encountered although a non-linear operator may be preferable to a linear one, but the advantages of simplicity and realizability favor the latter.

Physical realizability is one of the necessary conditions to be met by an operator and the mean square error is preferable as it is an even function of the error, and is proportional to the power, if the time series represents voltage.

Prediction

For a given stationary time series s(t), free from noise, prediction implies finding an operator having a weighting function W(t) such that the resulting operating on s(t), yields a prediction after a positive time α. The error is

$$e(t) = s(t+\alpha) - \int_0^\infty s(t-\tau)\ W(\tau)\ d\tau \tag{20}$$

Taking mean-square value of the error yields

$$E = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left| s(t+\alpha) - \int_0^\infty s(t-\tau)\cdot W(\tau)\,d\tau \right|^2 dt \tag{21}$$

After expansion and by the definition of autocorrelation function, the equation (21) may be written as

$$E = \phi(0) - 2\mathrm{Re}\left(\int_0^\infty \phi(\alpha+\tau) W^*(\tau)\ d\tau\right) + \int_0^\infty W(\tau) d\tau \int_0^\infty W(\sigma)\phi(\tau-\sigma) d\sigma \tag{22}$$

where

$$\phi(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} s(t+\tau)\ s^*(t)\ dt \tag{23}$$

and Re represents the real part of the expression following it. In order to minimize E, $W(\tau)$ is changed to $[W(\tau) + \epsilon(\delta W(\tau))]$, and differentiating it with respect to ϵ and equating this derivative to zero the following integral equation is obtained

$$\phi(\alpha+\tau) - \int_0^\infty \phi(\tau-\sigma)\ W(\sigma)\ d\sigma = 0 \tag{24}$$

This is the Wiener-Hopf equation of the first kind. Wiener (1949) shows that this represents a true minimum value of the error. Wiener's solution of this integral equation is discussed below.

Let G(w) be the power spectral density function of the time series s(t) and if it satisfies the condition

$$\int_{-\infty}^{+\infty} \frac{|\log|G(w)||}{1+w^2} \quad dw \quad < \infty \tag{25}$$

then $G(w)$ can be expressed as the product of the factors as $G(w) = H_1(w) \cdot H_2(w)$, where $H_1(w)$ is free from singularities in lower half-plane of w-plane, and $H_2(w)$ is free from singularities in upper half-plane and further more $G(w) = |H_1(w)|^2$.

If $Y(jw)$ is the Fourier transform of the weighting function $W(t)$, the solution of the integral equation (24) is given as

$$Y(jw) = \frac{\int_0^{\infty} \Psi(\alpha+t) \; e^{-jwt} \quad dt}{H_1 \; (w)} \tag{26}$$

where $\Psi(t)$ is the inverse Fourier transform of $H_1(w)$. Therefore the transfer function $Y(jw)$ of the operator is given as

$$Y(jw) = \frac{1}{2\pi H_1 \; (w)} \int_0^{\infty} e^{-jwt} \; dt \int_{-\infty}^{+\infty} H_1(w) \; e^{jw(t+\alpha)} \; dw \tag{27}$$

If $H_1(w)$ is written as

$$H_1(w) = \sum_{m,n} \frac{A_{m,n}}{(w-w_n)^m} \tag{28}$$

then equation (27) becomes

$$Y(jw) = \frac{\sum_{m,n} A_{m,n} \; e^{jw_n\alpha} \sum_{k=0}^{M-1} \frac{(j\alpha)^{m-1-k}}{(m-1-k)! \; (w-w_n)^{k+1}}}{\sum_{m,n} \frac{A_{m,n}}{(w-w_n)^m}} \tag{29}$$

Since the poles of $Y(jw)$ lie above the real axis in w-plane, the predictor is said to be physically realisable.

Filtering

Let the time series be a mixture of signal s(t) and noise n(t), where both of these time series are stationary. Let W(t) be the operator which, when operating on these two series, gives the approximation to s(t-). Then the mean square error is given as

$$E = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T}\left|s(t+\alpha)-\int_0^\infty\left[s(t-\tau) + n(t-\tau)\right] W(\tau)\, d\tau\right|^2 dt \tag{30}$$

This is expanded to give

$$E = \phi_{11}(0) - 2\mathrm{Re}\left[\int_0^\infty\left|\phi_{11}(\alpha+\tau) + \phi_{12}(\alpha+\tau)\right| W^*(\tau)\, d\tau\right] + \int_0^\infty W(\sigma)\, d\sigma \cdot$$
$$\int_0^\infty W^*(\tau)\, d\tau\, \left(\phi_{11}(\tau-\sigma) + \phi_{12}(\tau-\sigma) + \phi_{22}(\tau-\sigma) + \phi_{12}^*(\sigma-\tau)\right) \tag{31}$$

where $\phi_{11}(\tau)$ and $\phi_{22}(\tau)$ are autocorrelation functions of s(t) and n(t) respectively, while $\phi_{12}(\tau)$ is the cross-correlation function between s(t) and n(t). This expression is minimized by Wiener to obtain

$$Y(jw) = \frac{1}{2\pi H_1(w)} \int_0^\infty e^{-jwt} dt \int_{-\infty}^{+\infty} H_3(u)\, e^{ju(t+\alpha)} du. \tag{32}$$

and $G_1(w) = G_{ss}(w) + G_{ns}(w) + G_{sn}(w) + G_{nn}(w)$

where $G_{ss}(w)$ and $G_{nn}(w)$ are power spectral density of s(t) and n(t) respectively, while $G_{sn}(w)$ and $G_{ns}(w)$ are cross-spectral densities of s(t) and n(t). $H_1(w)$ is free from singularities in lower half plane of w-plane and

$$H_3(w) = \frac{G_2(w)}{H_1(w)} \tag{33}$$

Mean Square Error when the Operator is Known

If the input to an operator having transfer function Y(jw), has a spectral density as $G_{xx}(w)$ and if the output has a spectral density as $G_{yy}(w)$, then

$$G_{yy}(w) = |Y(jw)|^2 \; G_{xx}(w)$$

Laning and Battin (1956) show that if the input to the operator with transfer function $Y(jw)$, contains signal $s(t)$ and noise $n(t)$, which have power spectral densities $G_{ss}(w)$ and $G_{nn}(w)$ respectively, the spectral density of the error $G_{ee}(w)$ is given as

$$G_{ee}(w) = |Y_1(jw)|^2 \; G_{ss}(w) \; + Y_1^*(jw) \; Y(jw) \; G_{sn}(w) + Y_1(jw) \; Y^*(jw) \; G_{ns}(w) + |Y(jw)|^2 \; G_{nn}(w) \tag{34}$$

where $Y_i(jw)$ is the ideal operator which when operating on only $s(t)$ gives the desired output. A table of such operators is given below. $Y_1(jw)$ is the difference between the transfer function of the operator and the ideal operator $Y_i(jw)$. The total mean square error is obtained by integrating of the spectral density over the positive values of w, as

$$E = \overline{e^2(t)} = \int_0^\infty G_{ee}(w) \; dw = \frac{1}{2}\int_{-\infty}^{+\infty} G_{ee}(w) \; dw \tag{35}$$

Table 1. Table for ideal operator $Y_i(jw)$. (Laning and Battin, 1956).

Estimation of the Time Series	$Y_i(jw)$	Corresponding Weighting Function
Present value of $s(t)$	1	$\delta(t)$
Future value of $s(t) = s(t+T)$	e^{jwT}	$\delta(t+T)$
Past value of $s(t) = s(t-T)$	e^{-jwT}	$\delta(t-T)$
Derivative of $s(t) = s'(t)$	jw	$\delta'(t)$

Approximation of Wiener's Method

Levinson (1947) gave an approximate procedure for solving the same filter problem considered by Wiener. This method approximates the solution of Wiener's integral equation by replacing integral by a finite sum and minimizing the resulting error. A new measure similar to the weighting function is introduced. This new measure $A(t)$ is the output at time t corresponding to an input of a unit step function $u(t)$ defined as

$$u(t) = 1 \text{ for } t > 0$$
$$= 0 \text{ for } t < 0 \qquad (36)$$

For such an operator operating on an input function $x(t)$ yields the output $y(t)$ as

$$y(t) = \int_0^{\infty} A'(\tau)\, x(t-\tau)d\tau + A(0)\, x(t) \qquad (37)$$

where $A(0)$ is the limit of $A(t)$ as t approaches zero from the positive values, and $A'(t)$ is derivative of $A(t)$ with respect to t.

The integral on the right hand in equation (37) can be approximated by summing the areas over small interval of h, such that the area of the nth interval is given as $hA'(nh)\, x(t-nh)$. Summing over all values of n from 0 to ∞, $y(t)$ can be written as

$$y(t) = h \sum_{n=0}^{\infty} A'(nh)x(t-nh) + A(0)\, x(t) \qquad (38)$$

Writing $A(0)$ as A_0, and $hA'(nh)$ as A_n in the equation (38) yields

$$y(t) = \sum_{n=0}^{\infty} A_n\, x(t-nh) \qquad (39)$$

The right hand side of the equation represents an infinite summation. In such an approximation this sum may be terminated at n=M if the values of $A_n\, x(t-nh)$ does not increase significantly in the summation, the determination of a suitable M is considered later in the text. Such an approximation is valid in case of the time series, and equation (39) can be written as

$$y(t) = \sum_{n=0}^{M} A_n \, x(t-nh) \tag{40}$$

Then the value of $y(t)$ at $t=kh$ is given as

$$y(kh) = \sum_{n=0}^{M} A_n \, x((k-n)h) \tag{41}$$

and on writing y_k for $y(kh)$ and x_{k-n} for $x((k-n)h)$, it becomes

$$y_k = \sum_{n=0}^{M} A_n x_{k-n} \tag{42}$$

Calculation of a filtering operator

Let $A(t)$ be the response of the operator at time t to a unit step function, $f(t)$ be a time series containing both the signal $s(t)$ and noise $n(t)$ and the problem is to separate the signal from signal and noise. If a_k and b_k represent the values of $s(t)$ and $f(t)$ respectively at time $t=kh$, and when the output $h(t)$ when input is $f(t)$ is given as c_k, then

$$c_k = \sum_{n=0}^{M} A_n \, b_{k-n} \tag{43}$$

where A_n is defined previously, the error at time $t=kh$ is therefore

$$e_k = a_k - c_k = a_k - \sum_{n=0}^{M} A_n \, b_{k-n} \tag{44}$$

The mean square error is obtained by averaging the square of the error over all the values of k from $-\infty$ to $+\infty$. Hence, the mean square error E is

$$E = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=-N}^{+N} \left(a_k - \sum_{n=0}^{M} A_n b_{k-n}\right)^2 \tag{45}$$

Let the autocorrelation function of $f(t)$ be

$$\phi_{bb}(n) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{+N} b_t b_{t-n} \tag{46}$$

and the autocorrelation function of $s(t)$ be

$$\phi_{aa}(n) = \lim_{N \to \infty} \frac{1}{2N+1} \sum_{t=-N}^{N} a_t a_{t-n} \tag{47}$$

and $\phi_{ba}(n)$ be the cross-correlation function between f(t) and s(t). After expanding and substituting the values of correlation functions equation (45) becomes (Levinson, 1947)

$$E = \phi_{aa}(0) - 2 \sum_{n=0}^{M} A_n \phi_{ba}(n) + \sum_{n=0}^{M} \sum_{n=0}^{M} A_n A_m \phi_{bb}(m-n) \tag{48}$$

In order to minimize the total mean square error, the equation (48) is partially differentiated with respect to A_k, k=0,1,2,......M, and equated to zero as

$$\frac{\partial E}{\partial A_k} = -2 \phi_{ba}(k) + 2 \sum_{n=0}^{M} A_n \phi_{bb}(k-n) = 0 \tag{49}$$

Hence, the condition for maximum or minimum of E becomes (Levinson, 1947)

$$\sum_{n=0}^{M} A_n \phi_{bb}(k-n) = \phi_{ba}(k) \qquad k = 0,1,.......M \tag{50}$$

The condition yields M+1 equations for M+1 unknowns, A_n. Substitution of this condition in equation (48) results in the minimum E_m as

$$E_m = \phi_{aa}(0) - \sum_{n=0}^{M} A_n \phi_{ba}(n) \tag{51}$$

Dividing equation (51) by $\phi_{aa}(0)$, which is non-zero, writing E_m as V and the ratio of $\phi_{ba}(n)$ to $\phi_{aa}(n)$ as r_n, equation (51) becomes

$$V = 1 - \sum_{n=0}^{M} A_n r_n = 1 - H_m \tag{52}$$

where

$$H_m = \sum_{n=0}^{M} A_n r_n \tag{53}$$

The value of E_m lies between zero and $\phi_{aa}(0)$.

The value of H_m depends on the value of M. For large value of M, H_m increases and approaches unity, and consequently the value of E_m decreases. At a particular value of M, any further increase in M does not significantly effect the value of H_m, and as $M \to \infty$ H_m will tend to a limit H.

The equation (50) can be written as

$$\sum_{n=0}^{M} A_n r_{k-n} = \mu_k \tag{54}$$

where

$$\mu_k = \left[\frac{\phi_{ba}(k)}{\phi_{aa}(o)}\right] \tag{55}$$

for $k = 0,1,2,\ldots\ldots M$

Levinson (1947) developed an iterative process to solve these equations for A_n. For each index M, an equation is obtained, giving a value of A_n. Hence, $A_k^{(M)}$, for $k=0, 1\ldots\ldots M$ represents the values of A_k's at a particular M. Its values are given in the following equations.

$$A_o^{(o)} = \frac{\mu_o}{r_o} \tag{56}$$

$$A_{M+1}^{(M+1)}\left[r_o - \sum_{k=0}^{M} C_k^{(M)} r_{M+1-k}\right] = \mu_{M+1} - \sum_{k=0}^{M} A_k^{(M)} r_{M+1-k} \tag{57}$$

and

$$A_k^{(M+1)} = A_k^{(M)} - C_k^{(M)} A_{M+1}^{(M+1)} \quad \text{for } k = 0,1,\ldots\ldots M \tag{58}$$

The values of H_M's are given as

$$H_o = \frac{\mu_o^2}{r_o} \tag{59}$$

$$H_{M+1} = H_M + A_{M+1}^{(M+1)}\left[\mu_{M+1} + \sum_{k=0}^{M} C_k^{(M)} \mu_k\right] \tag{60}$$

where $C_k^{(M)}$ $k = 0,1,2,\ldots\ldots M$ are given as

$$C_o^{(o)} = \frac{r_1}{r_o} \tag{61}$$

$$C_o^{(M)}\left[r_o - \sum_{k=o}^{M-1} C_k^{(M-1)} r_{M-k}\right] = \mu_{M+1} - \sum_{k=o}^{M} C_{k-1}^{(M-1)} r_k \qquad (62)$$

$$C_k^{(M)} = C_{k-1}^{(M-1)} - C_o^{(M)} C_{M-k}^{(M-1)} \quad \text{for } k = 0,1,2,\ldots\ldots M \qquad (63)$$

As the value of M increases, the operator becomes more complicated and difficult to realize in practice, although better accuracy is obtained as M increases. Therefore a compromise is necessary for selecting the values of M. In practice different values of M are chosen and H_m's and A_k's as given by (57, 59, 60) are calculated. After comparing successive results similarly, a suitable operator is selected.

Levinson also considered the case of filters with lead or lag characteristics. In these cases, the error expression becomes

$$e_k = a_{k+s} - \sum_{n=o}^{M} A_n b_{k-n} \qquad (64)$$

where a_{k-s} is the value of signal s(t) at time t = (k+s)h. Hence, if s is positive, it is a filter with lead characteristics and if s is negative, it is a filter with lag characteristics. A solution similar is obtained by Levinson for this case and as follows. The equations (56, 58, 59, 61, 62, 63) remain unchanged and equations (57, 61) are modified to

$$A_{M+1}^{(M+1)}\left[r_o - \sum_{n=o}^{M} C_n^{(M)} r_{M+1-n}\right] = \mu_{M+1-s} - \sum_{n=o}^{M} A_n^{(M)} r_{M+1-n} \qquad (65)$$

and

$$H_{M+1} = H_M + A_{M+1}^{(M+1)}\left[\mu_{M+1-s} - \sum_{n=o}^{M} C_n^{(M)} \mu_{n+s}\right] \qquad (66)$$

Bode-Shannon Method

In Wiener's work the condition of physical realizability is a basic condition for obtaining prediction and filtering operators. Bode and Shannon (1950) solved the same problem, but did not introduce the condition of physical realizability at the start. First an operator which minimized the mean of the error function, is obtained and then realizability condition is imposed on it, as discussed.

Let $Y(jw)$ be transfer function of a filter which can also be represented as $Y(jw) = A(w) \exp\left[jB(w)\right]$, where $A(w)$ and $B(w)$ are the real valued polynominals in w. After substituting $Y_1(jw)$ as e^{jwT}, where T may be positive, negative or zero depending on the output, in equation (35) one obtains

$$e^2 = \int_0^{\infty}\Big[|Y(jw) - e^{jwT}|^2 G_{ss}(w) + [Y(-jw) - e^{-jwT}]Y(jw)\, G_{sn}(w) + [Y(jw) - e^{jwT}]Y(-jw)\, G_{ns}(w) + |Y(jw)|^2 G_{nn}(w)\Big]dw \tag{67}$$

where $G_{ss}(w)$ and $G_{nn}(w)$ are power spectral densities of signal $s(t)$ and noise $n(t)$ respectively, while $G_{sn}(w)$ and $G_{ns}(w)$ are cross-power spectral densities. Substituting $Y(jw) = A(w)\, e^{jB(w)}$, equation (69) becomes

$$e^2 = \int_0^{\infty}\Big\{A^2(w)\left[G_{ss}(w) + G_{ns}(w) + G_{sn}(w) + G_{nn}(w)\right] + G_{ss}(w) - A(w)\left[G_{ss}(w) - G_{ns}(w)\right] e^{j(wT-B(w))} - A(w)\left[G_{ss}(w) - G_{sn}(w)\right] e^{-j(wT-B(w))}\Big\}dw \tag{68}$$

Bode and Shannon (1950) show that minimum of equation (68) is obtained when $Y(jw)$ satisfies

$$Y(jw) = \frac{G_{ss}(w) + G_{ns}(w)}{G_{ss}(w) + G_{ns}(w) + G_{sn}(w) + G_{nn}(w)}\, e^{jwT} \tag{69}$$

If $Y(jw)$ is physically realizable, it will represent a filter for $[s(t) + n(t)]$ with lead or lag characteristics. A realizable operator with weighting function $W(t)$ must satisfy.

$$\begin{aligned} W(t) &= 0 \text{ for } t < 0 \\ &= \text{bounded for } t > 0 \end{aligned} \tag{70}$$

If $Y(jw)$ in equation (65) is physically unrealizable a modified procedure must be used. Let $Y_1(jw)$ be the shaping filter for $[s(t) + n(t)]$, and if $G_{xx}(w)$

$$+G_{yy}(w) = |H_1(w)|^2, \tag{71}$$

where $H_1(w)$ is free from singularities in the lower half w plane then

$$Y_1(jw) = H_1(w). \tag{72}$$

$Y_2(jw)$ is obtained from $Y_2(jw) = Y_1(jw)\, Y(jw)$. (73)

The weighting function $W_2(t)$ corresponding to $Y_2(jw)$ is obtained by taking inverse Fourier transformation as

$$W_2(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y_2(jw) \quad e^{jwt}\, dw \tag{74}$$

Then $W_3(t)$ is found such that

$$\begin{aligned} W_3(t) &= W_2(t) && \text{for } t \geqslant 0 \\ &= 0 && \text{for } t < 0 \end{aligned} \tag{75}$$

and the corresponding transfer function $Y_3(jw)$ is obtained by taking Fourier transform of $W_3(t)$. Finally the transfer function of the required operator is given as

$$Y_4(jw) = Y_3(jw)\, Y_1^{-1}(jw) \tag{76}$$

This is physically realizable as it satisfies the condition given in equation (70). The final results obtained by Wiener's method, and by this procedure yields the same result.

Prediction by Bode-Shannon Method

The Bode-Shannon procedure discussed above can also be used in prediction. A special case of $n(t) = 0$ is discussed here. The shaping filter $Y_1(jw)$ is given by $Y_1(jw) = H_1(w)$, where $G(w) = |H_1(w)|^2$ and $H_1(w)$ is free from singularities in lower half w-plane. The corresponding weighting function $W_2(t)$ is obtained by inverse Fourier transformation. Then $W_3(t)$ is constructed such that

$$\begin{aligned} W_3(t) &= W_2(t+T) && \text{for } t \geqslant 0 \\ &= 0 && \text{for } t < 0 \end{aligned} \tag{77}$$

where T is time after which prediction is required. Then the transfer function

$Y_3(jw)$ is obtained by Fourier transformation of $W_3(t)$, and the final predictor $Y_4(jw)$ is obtained as

$$Y_4(jw) = Y_3(jw)\, Y_1^{-1}(jw) \tag{78}$$

Mean Weighted Square Error Criterion

Murphy and Bold (1960) used the following method to minimize the weighted square error expression as

$$E = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} W(t)\, e^2(t)\, dt \tag{79}$$

where $W(t)$ is function of t alone and is selected according to the desired system performance. It is for this reason that this procedure is known as Mean Weighted Square error criterion.

Let the time series consist of stationary time series having signal $s(t)$ and stationary noise $n(t)$. The desired output $y(t)$ may be written as an ideal operator $g_1(t)$ operating on the signal $s(t)$. Let $W_1(t)$ be the weighting function of an operator which operates on $r(t)$, where $r(t) = [s(t) + n(t)]$, the output $h(t)$ is given as

$$h(t) = \int_{-\infty}^{+\infty} W_1(\tau)\, r(t-\tau)\, d\tau \tag{80}$$

The difference between desired output and actual output is given as

$$e(t) = y(t) - h(t) \tag{81}$$

The mean weighted square error is given by

$$E = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} W(t) \Big[y^2(t) - 2y(t) \int_{-\infty}^{+\infty} W_1(\tau)\, r(t-\tau)\, d\tau + \int_{-\infty}^{+\infty} W_1(\tau)\, r(t-\tau) \int_{-\infty}^{+\infty} W_1(\mu)\, r(t-\mu)\, d\mu\, d\tau \Big] dt \tag{82}$$

A new type of weighted correlation functions are defined as

$$\phi_{wyy}(\tau,\mu) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} W(t)\, y(t+\tau)\, y(t+\mu)\, dt \tag{83}$$

$$\phi_{wyr}(\tau,\mu) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} W(t)\, y(t+\tau)\, r(t+\mu)\, dt \tag{84}$$

and

$$\phi_{wrr}(\tau,\mu) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} W(t)\, r(t+\tau)\, r(t+\mu)\, dt \tag{85}$$

These are referred by Murphy and Bold as correlation functions in two-space, and if $W(t)$ is unity, these correlation functions are reduced to auto- and cross-correlation functions. A substitution of these functions in equation (82) results in

$$E = \Big[\phi_{wyy}(\tau,\mu) - \int_{-\infty}^{\infty} W_1(x)\, \phi_{wyr}(\tau,\mu-x)dx - \int_{-\infty}^{\infty} W(x)\, \phi_{wyr}(\mu,\tau-x)dx + \int_{-\infty}^{\infty} W(x) \int_{-\infty}^{\infty} W(z)\, \phi_{wrr}(\tau-x,\mu-z)\; dz\;\; dx \Big] \quad \text{for } \tau=\mu=0 \tag{86}$$

Murphy and Bold minimized E by calculus of variations methods. The condition for minimization

$$\int_{-\infty}^{\infty} W(z) \left[\phi_{wrr}(\mu-x,\tau-z) + \phi_{wrr}(\tau-x,\mu-z) \right] dz - \phi_{wyr}(\tau,\mu-x) - \phi_{wyr}(\mu,\tau-x) = 0 \qquad \text{for } x>0 \text{ and } \tau=\mu=0 \tag{87}$$

Two dimensional Fourier transformations are made assuming that the integral is absolutely convergent, the result is

$$\left[G(jw')\Psi_{wrr}(jw,jw') - \Psi_{wyr}(jw',jw) \right] e^{-jwx} + \left[G(jw)\Psi_{wrr}(jw',jw) - \Psi_{wyr}(jw,jw') \right] e^{-jw'x} = 0 \qquad \text{for } \tau=\mu=0. \tag{88}$$

where $G(jw)$ is the Fourier transformat of $W_1(t)$ and hence, it is the transfer function of the operator. The function $\Psi_{wrr}(jw,jw')$ is double Fourier transformation of $\phi_{wrr}(\tau,\mu)$, and similarly other such functions in equation (88) are obtained.

A solution of equation (88) yields the operator transfer function as

$$G(jw) = \frac{\Psi_{wyr}(jw, jw')}{\Psi_{wrr}(jw', jw)} \tag{89}$$

An ideal operator $g_i(\tau)$ operating on signal only gives this output

$$y(t) = \int_{-\infty}^{\infty} g_i(\tau)\, s(t-\tau)\, d\tau \tag{90}$$

If the input consists of uncorrelated signal $s(t)$ and noise $n(t)$,

$$\Psi_{wyr}(jw, jw') = G_i(jw)\Psi_{wss}(jw, jw') \tag{91}$$

where $G_i(jw)$ is Fourier transformation of $g_i(\tau)$ and $\Psi_{wss}(jw,jw')$ is double Fourier transformation of $\phi_{wss}(\tau,\mu)$, and equation (89) becomes

$$G(jw) = \frac{G_i(jw)\ \Psi_{wss}(jw, jw')}{\Psi_{wss}(jw, jw') + \Psi_{wnn}(jw, jw')} \tag{92}$$

This gives an operator which minimizes the weighted mean square error E. If this operator is not physically realizable, then a procedure similar to that of Bode and Shannon may be used to obtain physical realizability.

NON-STATIONARY TIME SERIES

If the autocorrelation $\phi(t_2, t_1)$ of a time series is not a function of (t_2, t_1), then the time series is non-stationary. The problem of filtering and prediction for such a time series is solved in a modified way depending on the nature of time series and the type of performance required. Numerous approaches have been reported in the literature and only two of those approaches will be discussed here.

An Approach by Ragazzini and Zadeh

A specific problem of a non-stationary time series is considered by Zadeh and Ragazzini (1950). The problem is to separate a non-stationary time series from the same time series with additive stationary noise. It is further assumed that the signal time series $s(t)$ can be divided into a summation of two series, one of which is a stationary time series $M(t)$ and the other a time series $P(t)$ which is a polynominal in t of degree n.

The desired output can be obtained by an ideal operator, with weighting function $W(t)$ operating on the input signal time series.

Let $W_1(t)$ be the weighting function of an operator through which the total series $f(t)$ is passed, and its output $y(t)$ is given as

$$y(t) = \int_0^{\infty} W_1(\tau)\, f(t-\tau)\, d\tau \tag{93}$$

In practice it is usually found necessary to restrict the duration of sampling of the input time series to a finite duration T, which implies that $W_1(t)$ must be zero outside the interval $0 \leq t \leq T$. Therefore equation (93) can be written as

$$y(t) = \int_0^T W_1(\tau)\, f(t-\tau)\, d \qquad 0 \leqslant t \leqslant T \tag{94}$$

The polynominal $P(t- \ )$ can be expanded into Taylor's series such as

$$P(t- \ ) = P(t) - \tau P^{(1)}(t) + \frac{\tau^2}{2!} P^{(2)}(t) - \cdots + (-1)^n \frac{\tau^n}{n!} P^{(n)}(t) \tag{95}$$

where $P^{(k)}(t)$ is the k th derivative of $P(t)$ with respect to t. Its derivative $P^{(m)}(t)$ of order higher than n are zero because $P(t)$ is polynominal of n th order. Now $f(t) = [M(t) + P(t) + N(t)]$, where $N(t)$ is additive noise. Substituting in equation (94) gives

$$y(t) = \int_0^T \left[M(t-\tau) + P(t-\tau) + N(t-\tau)\right] W_1(\tau) d\tau \tag{96}$$

A substitution of $P(t- \ )$ in equation (95) in equation (96)

$$y(t) = \int_0^T W_1(\tau) \left[M(t-\tau) + N(t-\tau)\right] d\tau - \int_0^T \left[P(t) - \tau P^{(1)}(t) \ldots\ldots\right] W_1(\tau)\, d\tau \tag{97}$$

$$= \int_0^T W(\tau)\left[M(t-\tau) + N(t-\tau)\right] d\tau + \mu_0 P(t) - \mu_1 P^{(1)}(t) \ldots\ldots$$

$$+ (-1)^n \frac{\mu_n}{n!} P^{(n)}(t) \tag{98}$$

where μ_k represents k th moment of $W(t)$

In order to find values of μ_k it is assumed that $M(t)$ and $N(t)$ are normalized to have zero mean value, hence, the average value of $y(t)$ is given by

$$y(t)_{av} = \mu_0 P(t) - \mu_1 P^{(1)}(t) + \ldots\ldots\ldots\ldots\ldots\ldots + (-1)^n \frac{\mu_n}{n!} P^{(n)}(t) \tag{99}$$

If $y_1(t)$ is the desired output, the average value of $y_1(t)$ is given as

$$y_1(t)_{av} = \left[\int_0^\infty W_1(t)\left[M(t-\tau) + P(t-\tau)\right] d\tau\right]_{av}$$

$$= \left[\int_0^\infty W_1(t)\, M(t-\tau)\, d\tau\right]_{av} + \left[\int_0^\infty W_1(t)\, P(t-\tau)\, d\tau\right]_{av} \tag{100}$$

Since $M(t)$ is stationary, it follows that the ensemble means of $y_1(t)$ depend only on the non-random component of the signal, hence equation (100) becomes

$$y_1(t)_{av} = \left[\int_0^\infty W_1(\tau) P(t-\tau) d\tau\right]_{av} \tag{101}$$

Equating equation (99) and (101) yields

$$\left[\int_0^\infty W_1(\tau) P(t-\tau)\, d\tau\right]_{av} = \mu_0 P(t) - \mu_1 P^{(1)}(t) + \ldots + (-1)^n \frac{\mu_n}{n!} P^{(n)}(t) \tag{102}$$

The integral on the left hand side can be expanded into Taylor's expansion of $P(t)$. Equating the co-efficients of $P(t)$ on both the sides the values of μ_k's are obtained.

Now the mean square error is given by

$$E = \int_0^\infty W_1(\tau_1) W_1(\tau_2)\ \phi_{mm}(\tau_1 - \tau_2)\, d\tau_1\, d\tau_2 - 2\int_0^\infty\int_0^\infty W_1(\tau_1)\ W(\tau_2)$$

$$\phi_{mm}(\tau_1 - \tau_2)\, d\tau_1\, d\tau_2 + \int_0^T\int_0^T W(\tau_1) W(\tau_2)\left[\phi_{mm}(\tau_1 - \tau_2)\right.$$

$$+\phi_{nn}(\tau_1-\tau_2)\Big]\,d\tau_1 d\tau_2 \tag{103}$$

where $\phi_{mm}(t)$ and $\phi_{nn}(t)$ are the autocorrelation functions of $M(t)$ and $N(t)$.

The problem is now reduced to minimizing the equation (103), with the constraints of the μ_k's. As the first term of equation (103) is independent of $W(\tau_2)$, and hence it is not taken into account in finding a minimum. And the expression to be minimized is given as

$$I = \int_0^T W(\tau_1)\,d\tau_1\Bigg[\int_0^T W(\tau_2)\Big[\phi_{mm}(\tau_1-\tau_2)+\phi_{nn}(\tau_1-\tau_2)\Big]\,d\tau_2$$
$$-2\int_0^\infty W_1(\tau_2)\,\phi_{mm}(\tau_1-\tau_2)\,d\tau_2 - 2\lambda_0 - 2\lambda_1\tau_1 - 2\lambda_2\tau_2 \cdots\cdots$$
$$-2\lambda_n\tau_n\Bigg] \tag{104}$$

where $\lambda_1, \lambda_2 \ldots\ldots \lambda_n$ are Lagragian's multipliers. The equation (104) is minimized when $W_1(t)$ satisfies

$$\int_0^T W(\tau)\Big[\phi_{mm}(t-\tau)+\phi_{nn}(t-\tau)\Big]\,d\tau = \lambda_0 + \lambda_1 t + \ldots\ldots\ldots + \lambda_n t^n +$$
$$\int_{-\infty}^{\infty} W_1(t)\,\phi_{mm}(t-\tau)\,d\tau \qquad 0 \leqslant t \leqslant T \tag{105}$$

This integral can be solved by various methods including numerical methods. One such method is given by Ragazzini and Zadeh (1950).

A General Problem of Non-Stationary Time Series

R. C. Booton, Jr (1952) assumed a time-variant system having weighting function $W(t,\tau)$. If $x(t)$ is the input to the system having a weighting function $h(t,\tau)$, the output $y(t)$ is given by

$$y(t) = \int_0^\infty h(t,\tau)\,x(t-\tau)\,d\tau \tag{106}$$

Let the time series $f(t)$ which is to be filtered, be a summation of non-stationary time series with s(t) and n(t). Let y(t) be the desired output which is obtained with an ideal operator with weighting function $g_1(t,\tau)$, operating on the signal s(t) alone, hence

$$y(t) = \int_{-\infty}^{\infty} g_1(t,\tau)\, s(t-\tau)\, d\tau \tag{107}$$

For the ideal operator $g_1(t,\tau)$, the lower limit of the integration is taken as $-\infty$.

The autocorrelation function is given as

$$\phi_{ss}(t_1,t_2) = \langle s(t_1)\, s(t_2)\rangle \tag{108}$$

where $\langle\cdots\rangle$ represents the time average.

The autocorrelation and cross-correlation are known. The autocorrelation of f(t) is given by

$$\phi_{ff}(t_1,\ t_2) = \langle (s(t_1) + n(t_1))\ (s(t_2) + n(t_2))\rangle \tag{109}$$

$$= \phi_{ss}(t_1,t_2) + \phi_{sn}(t_1,t_2) + \phi_{ns}(t_1,t_2) + \phi_{nn}(t_1,\ t_2) \tag{110}$$

If $W(t,\tau)$ is the weighting function of an operator, its output h(t) is given as

$$h(t) = \int_0^{\infty} W(t,\tau)\, f(t-\tau)\, d \tag{111}$$

and the error is

$$e(t) = y(t) - \int_0^{\infty} W(t,\tau)\, f(t-\tau)\, d\tau \tag{112}$$

The mean square error is then obtained as

$$E = \langle e^2(t)\rangle = \langle\left[(y(t) - \int_0^{\infty} W(t,\tau)\, f(t-\tau)\, d\tau)^2\right]\rangle$$

$$= \phi_{yy}(t,t) - 2 \int_0^\infty W(t,\tau)\ \phi_{yf}(t-\tau,t)d\tau$$

$$+ \int_0^\infty \int_0^\infty W(t,\tau_1)\ W(t,\tau_2)\ \phi_{ff}(t-\tau_1,t-\tau_2)\ d\tau_1\ d\tau_2 \qquad (113)$$

To minimizing E for $W(t,\tau)$, let $W(t,\tau)$ be increased by $B(t,\tau)$, the new error is given by

$$E_1 = \phi_{yy}(t,t) - 2 \int_0^\infty W(t,\tau)\ \phi_{yf}(t-\tau,t)d\tau + \int_0^\infty \int_0^\infty W(t,\tau_1)\ W(t,\tau_2)$$

$$\phi_{ff}(t-\tau_1,t-\tau_2)d\tau_1 d\tau_2$$

$$+ \int_0^\infty \int_0^\infty W(t,\tau_1)\ B(t,\tau_2)\ \phi_{yf}(t-\tau_1,t-\tau_2)\ d\tau_1 d\tau_2\ -2\int_0^\infty B(t,\tau)$$

$$\phi_{yf}(t-\tau,t)\ d\tau$$

$$+ \int_0^\infty \int_0^\infty W(t,\tau_2)\ B(t,\tau_1)\ \phi_{yf}(t-\tau_1,t-\tau_2)\ d\tau_1 d\tau_2$$

$$+ \int_0^\infty \int_0^\infty B(t,\tau_1)\ B(t,\tau_2)\ \phi_{yf}(t-\tau_1,t-\tau_2)\ d\tau_1 d\tau_2 \qquad (114)$$

The last term in equation (114) is non-negative. Now for minimum E, $E_1 \geqslant E$

or

$$[E_1 - E] \geq 2 \int_0^\infty B(t,\tau_1)\ \left[\int_0^\infty W(t,\tau_2)\ \phi_{ff}(t-\tau_1,\ t-\tau_2)\ d\tau_2 - \phi_{yf}\right.$$

$$\left.(t-\tau_1,t)\right]\ d\tau_1 \qquad (115)$$

Equation (115) should be zero for minimum for all values of $B(t,\tau_1)$ or

$$\phi_{yf}(t-\tau_1,t) = \int_0^\infty W(t,\tau_2)\ \phi_{ff}(t-\tau_1,\ t-\tau_2)\ d\tau_2 \qquad (116)$$

Which is the condition for minimum E. The solution of equation (116) yields the value of $W(t,\tau)$.

There can be more than one solution which satisfies the equation (116). All such solutions will give the same mean square error. Booton (1952) stated that most of the physical system have unique solution, although he did not derive such a general solution. It was also stated that a number of approximations and numerical methods can be applied to solve this equation for $W(t,\tau)$. If the operator obtained from the integral equation (116) is not physically realizable, then a procedure similar to that of Bode-Shannon can be used to assure physical realizability.

The operators obtained by the different methods are physically realizable by employing a finite number of lumped resistances, capacitances and inductances. Levinson (1947) offered a method for obtaining a filter approximating to the values of A_n's obtained in equations (56, 57, 58, 65).

CONCLUSION

Various error criteria for time series are discussed in this paper. Such errors in control systems are minimized and an optimum performance is obtained.

For statistical inputs, such techniques which minimize the weighted mean square error and thus obtaining optimum filtering and prediction are indispensable. Noise jitter observed in automatic radar and wind gust disturbances in aircraft are a few examples of such systems. A prediction problem giving the future position of a moving target is another important problem.

Most of the work in the literature stresses least-square method. Further investigation is necessary to minimize different functions of error employing different weighting function similar to work done by Murphy and Bold.

ACKNOWLEDGMENT

The author wishes to express his deep gratitude to Dr. H. S. Hayre, for the help and guidance given by him throughout this work.

REFERENCES

1. Bode, H. W. and C. E. Shannon.
   A simplified derivation of linear least square smoothing and prediction theory. Proc. I.R.E., Vol. 38. April 1950, pp 417-425.

2. Booton, R. C., Jr.
   An optimization theory for time-varying linear systems with non-stationary statistical inputs. Proc. I.R.E., Vol. 40 August 1952, pp 977-981.

3. Graham, D. and R. C. Lathrop.
   The synthesis of 'optimum' transient response: criteria and standard forms. Trans. A.I.E.E., Vol. 72, pt. 2. November 1953. pp 223-286.

4. Laning, J. H., Jr., and R. H. Battin.
   Random processes in automatic control. New York: McGraw Hill Co., Inc., 1956.

5. Levinson, N.
   The Wiener RMS (root mean square) error criterion in filter design and prediction. Journal of Mathematics and Physics, Vol. XXV, No. 4. January 1947, pp 261-278.

6. Murphy, G. J. and N. T. Bold.
   Optimization based on a square error criterion with an arbitrary weighting function. I.R.E. Trans. on Automatic Control, Vol. AC-5. January 1960, pp 24-30.

7. Nims, P. T.
   Some design criteria for automatic controls. Trans. A.I.E.E., Vol. 70, pt 1. 1951, pp 606-611.

8. Schultz, W. C. and V. C. Rideout.
   A general criterion for servo performance. Proc. Nat'l Electronics Conf., Vol. 13 1957, pp 549-560.

9. Spooner, M. G. and V. C. Rideout.
   Correlation studies of linear and non-linear systems. Proc. Nat'l Electronics Conf., Vol. 12. 1956, pp 321-325.

10. Wiener, N.
    The extrapolation, interpolation, and smoothing of stationary time series. New York: John Wiley and Sons, Inc. 1949.

11. Yaglom, A. M.
    An introduction to the theory of stationary random functions. Eanglewood Cliffs, N.S., Prentice-Hall, Inc. 1962.

12. Zadeh, L. A. and J. R. Ragazzini.
    An extension of Wiener's theory of prediction. J. Appl. Phys., Vol. 21. July 1950, pp 645-655.

ON PREDICTION AND FILTERING OF TIME SERIES

by

Sureshchandra L. Vasa

B. E., 1963
University of Bombay, India

AN ABSTRACT OF
A MASTER'S REPORT

submitted in partial fulfillment of the

requirements of the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1965

Theories of filtering and prediction of stationary and non-stationary time series are reviewed. A time series is a sequence of quantitative data assigned to specific points in time domain. A time series may be stationary or non-stationary depending upon the statistical characteristics of the autocorrelation function.

A distorted time series is composed of two time series, namely the original time series and the noise time series. Filtering is used to separate the original time series from the mixture of two series, whereas prediction is employed to determine a future value of the original time series.

The error of a system is the difference between desired output and the actual response of the system. The error function F(e) is the weighted error e(t). The mean weighted error is minimized to obtain a suitable operator. The weighting function or transfer function of the operator is modified to make it physically realizable.

For the stationary time series, the original work of least-mean-square error by Wiener is discussed briefly. An integral equation is solved to obtain the optimum result. The Levinson's method of using a finite sum in place of integral equation to linear equations and Bode-Shannon derivation of the above results with a concept of shaping filter are discussed. Murphy and Bold's method of minimising mean weighted square error is also included.

For non-stationary time series, the following two cases are discussed. Ragazzini and Zadeh considered a special case when the signal can be separated into a stationary time series and polynominal in t, whereas the noise time series is stationary. Booton derived an integral equation for a general case of non-stationary time series.