A Qualitative Introduction to Linear Least square Estimation

*Dr. S. Chirarattananon**

1. Introduction

The concept of electrical signal filtering, whereby an incoming signal is "filtered" by passing the signal through an electrical device to eliminate the undesirable components of the signal, is a familiar one for many readers. In radio communications, particularly in the reception of commercial broadcasting, the receiver is tuned to a particular station by adjusting the input filter to accept only signal of required frequencies.

In satellite tracking the observation data of the positions of the satellite from the tracking stations are "filtered" or processed and used to estimate the position of the satellite. In communications via satellite signals are processed to obtain the desired messages.

The concept of estimation, of which filtering is only a part, may be defined to mean the processing of a set of observations to obtain a desired set of information. The set of observations may come in the form of electromagnetic signals e.g. induced voltages, or geometrical measurements of length, or in fact any analog or digital forms. The observations are in all cases not perfect in that they are contaminated by undesired disturbances and noises. In filtering the observations signal passes through a prearranged (which may be adaptive) set of equipments or processors the outcome of which is the filtered signal which in some cases include the estimate we desire. The concepts of filtering and estimation are not totally distinct. In many instances the two are interchangeable.

*Lecturer, Faculty of Engineering, Khon Kaen University.

Other examples exemplifying the application of estimation theory include the filtering of an output signal from a servomechanism to obtain estimates of the various quantities such as the angular position or the velocity of the motor shaft. Figure 1 depicts a situation in which the output y of the system or plant as observed or measured is contaminated by noise (w). The filter processes the noisy measurement to obtain the estimate $\hat{y}$ of y. This estimate is then used by the controller to compute an appropriate feedback signal for the purpose of controlling the system to attain efficient operation. This situation may be generalized to include an economic system. Here the controller will be the government whose task is to find proper fiscal or monetary policies and to implement these policies for the economic well being of the population.

More recent applications may be found in radar and sonar tracking of objects. The return signals are usually highly contaminated by noise such that a statistical method called correlation must be used to obtain the buried signal. This method is time consuming and expensive. Recent results in filtering theory has been applied successfully to these cases. The new procedure gives the estimate of the required signal "on line", ie. an estimate of the signal is given almost instantaneously as the measurements are received.

It is interesting to note that estimation theory has its analytical foundation in Gauss' least-squares method. Gauss developed a technique in 1795 to establish and predict the orbit of Ceres. Prior to this, geometric ideas had been utilized extensively in orbit determination, generally with great frustration. Following Gauss' work very little appeared until the 1930's. The new era started with the works of Kolmogorov (1) and Wiener (2). The most recent interest is stimulated in large measure by the works of Kalman (3) and Kalman and Bucy (4). Although it is true that the present interest is now focussed on nonlinear estimation, it is felt that linear estimation and estimation theory in general is not yet widely known, and that opportunity for research in this field is not exhausted. In particular, only relatively few application notes on linear estimation theory have appeared. In the sequel we shall consider the Wiener filter and the Kalman-Bucy filter.

2. Problem definition.

We consider a continuous-time measurement process model of the form

$$z(t) = y(t) + w(t) \qquad \text{.........................} (1)$$

where $y(t)$ is the signal process and $w(t)$ the noise process. Our aim here is to obtain an estimate of $y(t)$ or some other processes related to $y(t)$ by linear operation. In all cases the estimate we seek satisfies certain criteria. First, the estimate is a best estimate in that the error in the estimation is smallest in the meansquare sense (this will be discussed further on). Second, the operation on the measurement process $z(t)$ which we undertake to obtain the estimate must be linear-hence the concept "linear least-squares estimation". Even when the linearity constraint is enforced general solution for our problem is not directly obtainable in closed form. Further fundamental assumptions are required. The process $y(t)$ and $w(t)$ are guassian or normal processes. Technically, this means that the joint probability density of $y(t_1)$, $y(t_2)$, ...$y(t_n)$ must be gaussian for any set of time variables t_i, and any integer n (see (5) and (6) for precise definitions). Many commonly occurring processes meet this assumption, though often only approximately. Examples are :

(1) Thermodynamic noise, such as occurs in resistors.

(2) Shot noises, such as occurs in active devices.

(3) Noise arising from the superposition of huge number of tiny independent random disturbances (here the central limit theorem applies, see 5)

(4) Interference on earth – satellite communication channels.

(5) Noise appearing at the output of a linear circuit of system when the input is gaussian noise.

It is true that many commonly occurring noises are not gaussian. However, the filters we shall discuss offer robust performance in that the degradation in performance is minor. Albeit, we can consider the filters to be the best linear filters for non-gaussian noises (Kalman (3)) will be made when we discuss the Wiener or the Kalman filter.

Let us consider a process $s(t)$, related to $y(t)$ by a linear operation. Suppose we want to obtain an estimate of $s(t)$ from the observations $z(t)$. By definition, the whole trajectory of $z(t)$ is utilized in obtaining the estimate. We may classify the estimate $\hat{s}$ of s into three types.

If $\hat{s}(t/b)$ is the estimate of $s(t)$, and $\hat{s}(t/b)$ is obtained from the trajectory $z(\tau)$, $t_o < \tau < b$,

where $b < t$ and t_0 is some initial time we have a predicted estimate of s. Here we utilize the information of z from t_0 to b, where $b < t$. If $b = t$, we have $\hat{s}(t/t)$ as the filtered estimate of $s(t)$, where the information from $z(\tau)$ is utilized from $\tau = t_o$ to $\tau = t$.

When $b > t$, we have $\hat{s}(t/b)$ as a smoothed estimate of $s(t)$.

We may classify the smoothed estimate further into three types. When $t = t_1$, a constant and b is running variable $(b > t_1)$ we have a fixed-point smoothed estimate ; when b = a constant and t is a running variable, we have a fixed-interval smoothed estimate and when $b = t + \triangle$, where $\triangle$

is a fixed time lag, we have a fixed-lag smoothed estimate.

3. The Wiener Filter

The situation in which Wiener filtering applied is when $s(t)$ is scalar and $s(t) = y(t)$, the signal process. Further assumptions on the processes $y(t)$ and $w(t)$ for the solution to the estimation problem are

i) The statistical means, the autocorrelation and the cross-correlation of the process $y(t)$ and $w(t)$ are known.

ii) The processes $y(t)$ and $w(t)$ are stationary. Mathematically this implies the following.*

$$E[\,y(t)\,] = E[\,y(t+T)\,] = m_s$$
$$E[\,w(t)\,] = E[\,w(t+T)\,] = m_w$$
$$Ryy(t, \tau) = E[\,y(t+T)\, y(\tau+T)\,] = E[\,y(t)\, y(\tau)\,],$$
$$Ryw(t, \tau) = E[\,y(t+T)\, w(\tau+T)\,] = E[\,y(t)\, w(\tau)\,],$$
$$Rww(t, \tau) = E[\,w(t+T)\, w(\tau+T)\,] = E[\,w(t)\, w(\tau)\,],$$

for all t, τ and T; where Ryy and Rww are called the autocorrelation and Ryw is the cross-correlation.

Physically, the last assumption means that the mechanism generating the random processes $y(t)$ and $w(t)$ is not time-varying. It also means that the power spectra can be associated with $y(t)$ and $w(t)$, these being frequency domain functions measuring the intensity of the component of the processes present at a certain frequency.

Most of the frequently encountered random processes are stationary if certain condition holds. The noise generated by passing a current through a resistor is stationary if the temperature of the resistor is kept constant. In almost all cases temperature variation is the main cause of time variation in the statistics of the random processes. In most cases, temperature variation is small for a relatively short time-interval such that the random

---

*E denotes ensemble averaging or expectation.

processes can be considered stationary, eg. the ambient temperature in the surroundings of a device changes relatively slowly in the course of, say, an hour.

To enable one to obtain neat solution and to physically implement the solution in a straight-forward manner certain assumptions are further made on the processes. These are that the noise $w(t)$ is white and that the power spectrum of the process $s(t)$ falls off at high frequency. Physically, white noise is an impossibility. Whiteness implies a flat power spectrum for all frequencies. Fortunately, noise with power specturm flat to very high frequency is common. Also, since Wiener filter offers robust performance the falling off of power spectrum does not affect its performance greatly.

In order to obtain the Wiener filter, we first write down the integral equation derived from the technique of calculus of variation under the condition that the meansquare error $E\left\{ y(t) - \hat{y}(t/b) \right\}^2$ is minimized by the use of the Wiener filter. This is the well - known Wiener - Hopf equation. By applying a transformation technique we obtain the frequency domain version of this equation. Spectral factorization is then used to solve for this equation (2), (8), which under normal circumstances is achieved by factorizing polynomial in a special way.

Several comments on the Wiener filter is relevant at this stage.

i) The Wiener filter may contain differentiator. As we know caution must usually be exercised in differentiating a random process. When it occurs in the Wiener filter it simply is a necessary procedure for obtaining an optimum filter.

ii) The Wiener filter obtained as a solution from spectral factorization is usually a band-limited system. This means that the transfer function $H(s)$ of the Wiener filter goes to zero as s goes to infinity.

iii) The Wiener filter is normally stable. This means that its transfer function has no right half plane pole and its construction and use are thus practical propositions.

Before proceeding we must make a certain reservation regarding this last comment. One can imagine the filter as a device which produces an estimate $\hat{y}(t/t)$ of $y(t)$ with $z(t)$ as its input at that instant. Thus we obtain an "on-line" estimate of $y(t)$. One can also obtain a fixed-lag smoothed estimate $\hat{y}(t/t+\triangle)$ of $y(t)$ from the fixed-lag smoother obtained by the same mechanism which produces the Wiener filter. The estimate $\hat{y}(t/t+\triangle)$ is also an "on-line" estimate in the sense that the device produces $\hat{y}(t/t+\triangle)$ as an estimate of $y(t)$ with $z(t+\triangle)$ as the input at that instant. As it turns out the fixed-lag smoother contains terms like $[\exp(-s\triangle) - \exp(-\alpha\triangle)](s-\alpha)^{-1}$, where α is a positive constant. The problem with this terms is that it is difficult (not to say impossible) to realize a stable system with this type of transfer function. This point will be taken up later.

iv) The performances of the Wiener filter and the associated smoother or predictor, measured as the meansquare error $E\{y(t) - \hat{y}(t/t)\}^2$ etc., are readily computable from Ryy, Ryw and Rww. This means that from the a priori information we possess, we can also predict the performance of the filter, etc., in advance. We can choose which type of estimator to use to give us the desired accuracy. Furthermore, as the results of research in this direction indicate use of smoother can be quite advantageous. First, the improvement in performance resulting from the use of fixed-lag smoothing as compared to filtering can be very significant. Second, the value of the delay $\triangle$ needs not necessarily be large to enable one to obtain all or almost all of the improvement (9). This situation applies also to the Kalman-Bucy filter to be discussed.

4. The Kalman–Bucy Filter

In a broad sense, the situation in which the Kalman–Bucy filter applies is similar to that for the Wiener filter. The original paper of Kalman and Bucy, however, requires an additional knowledge on the detailed structure of the mechanism generating the signal process, while it also enlarges the number of applicable cases by allowing time–varying signal processes. Recent discovery has shown that the requirement of the detailed knowledge on the structure of the signal process is not entirely necessary. The knowledge of the type of structure or model for the signal process is sufficient to obtain a Kalman–Bucy filter.

To proceed let us consider the original Kalman–Bucy filter by assuming that there is known the state–space equations of a linear system with a deterministic input and a white noise input;

$$\begin{aligned} \dot{x} &= F(t)\,x + G(t)\,\left[u(t) + v(t)\right], \quad E\left[x(t_o)\right] = x_o, \\ z(t) &= y(t) + w(t) \\ y(t) &= H'(t)\,x(t) \end{aligned} \qquad (2)$$

Here **x** is the state of the system, u is a known deterministic input, $v(t)$ is a zero mean, white gaussian process (noise), with known covariance matrix

$$E\left[v(t)\,v'(\tau)\right] = Q(t)\,\delta(t-\tau)$$

Where δ is the Dirac delta function. The superscript prime denotes matrix transposition. The case of a vector process is easily treated with the Kalman–Bucy Theory, in contrast to the Wiener Theory.

The initial time for (2) is t_o, with $t_o - \infty$ allowed; $x(t_o)$ is a guassian random variable of known mean and known covariance matrix, and is independent of v (t) and w (t) for all $t > t_0$. Figure 2 illustrates the arrangement.

The above assumption prompts the following notes

i) The Kalman-Bucy filter assumes a detailed knowledge of the mechanism generating the signal process. The signal y (t) is the output of a band - limiting system describable by ordinary differential equations. If F, G and H are constant, there exists a transfer function (matrix), for the signal process, relating the Laplace transform of u to the Laplace transform of y, with this transfer function (matrix) rational and tending to zero as s tends to ∞. The Wiener theory requires only that Ryy (t, τ) be known, in retrospect. The difference lies in the fact that many different systems will lead to the same Ryy (t, τ).

ii) The Wiener theory demands stationarity, the Kalman - Bucy Theory does not; the stationary case in the Kalman-Bucy theory is a special case when F,G and H are constant, the various processes are stationary and when $t_0 = -\infty$.

iii) The Kalman-Bucy theory primarily demands that the measurement noise w be a white gaussian process, although this assumption can be removed to a certain extent.

It has recently been found (10) that less restrictive assumptions are actually required to compute a Kalman-Bucy filter design than noted above. These are

i) The signal y is the output of some linear finite-dimensional system excited by white noise.

ii) The measurement noise w is a zero mean, white gaussian process with nonsingular covariance.

iii) The covariance functions Ryy and Ryw are known. These changes on the basic assumptions bring the Kalman–Bucy Theory into much closer parallel with the Wiener Theory. Moreover, the change has practical significance, since there exist nonstationary filtering problems where detailed knowledge of the mechanism generating y is not present but the knowledge of Ryy and Ryw is present. (see (11)).

The Kalman–Bucy Theory primarily concerns with the optimum estimate $\hat{x}(t/t)$ of the state $x(t)$, given the trajectory $z(\tau)$, $t_o < \tau < t$. Again, the estimate is optimum in the sense that the meansquare error $E\,[x(t) - \hat{x}(t/t)]^2$ is minimized. The optimum signal estimate $\hat{y}(t/t)$ is obtained from the state estimate simply as

$$\hat{y}(t/t) = H'(t)\,\hat{x}(t/t).$$

As is known the state description of a system gives much more detailed knowledge on the system. In addition the design of feedback controllers which position the closed–loop transfer function matrix poles of a system such as described by

$$\dot{x} = Fx + Gu$$

$$y = H'x$$

to any arbitrary values can be regarded as a problem of selecting a state feedback law

$$u = K'x + u_{ext}$$

and implementing that law. If the system states are not available, then it necessary to compute a state estimate $\hat{x}$ from the measurement z, to implement

$$u = K\hat{x} + u_{ext}.$$

This is one situation where the Kalman-Bucy filter is applicable.

The optimum Kalman-Bucy filter is a finite-dimensional linear system. In addition, it inherently contains a model of the system generating y. Figure 3 depicts a Kalman-Bucy filter. The linear gain $K(t)$ is obtained via a sequence of calculations. The filter input comprises u, the deterministic input to the signal generating process model, the noisy measurement z and the output signal y which goes through a feedback path.

Similar to the Wiener filter, the Kalman-Bucy filter is normally stable. Further, the same procedure used to obtain the Kalman-Bucy filter can be applied to obtain a fixed-lag smoother. Here again, a problem of stability is encountered (12). A close examination will reveal that this problem with the fixed-lag smoother obtained through the use of the Kalman-Bucy Theory is identical to the problem arises from the use of the Wiener Theory. A route around this problem exists, however. It is possible to obtain stable fixed-lag smoothers which are optimum for the case of discrete-time smoothing (13) and suboptimal in the case of continuous-time smoothing (14).

Summary

The filtering theories we discussed are usually classified as linear recursive estimation theory. Non-recursive estimation theory is in fact the forerunner of the estimation theory; dating back to the time of Gauss

Kepler, etc. Present research however, focusses on the development of recursive estimation, as this is the case where physical implementation is an attractive practical proposition.

Our discussion has been concentrated upon the continuous-time filtering. In modern communication, however, most systems are more appropriately described as discrete-time model. The actual transmitted signals are often continuous-time. Sampling takes place at the receiving end and the signals are processed using both digital and analog equipments. The estimate obtained from the discrete-time measurement sequence is given both as discrete-time and continuous-time. This estimation procedure is known as discrete-continuous filtering. This concept can also be applied to fixed-lag smoothing (15). The smoother obtained normally possesses the same stability property as that possessed by the Kalman-Bucy filter.

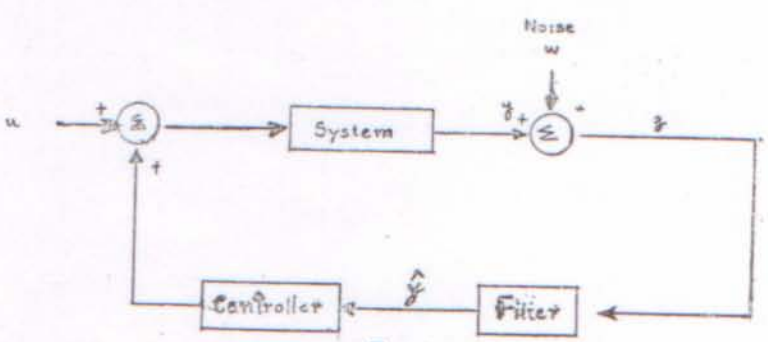

Fig 1

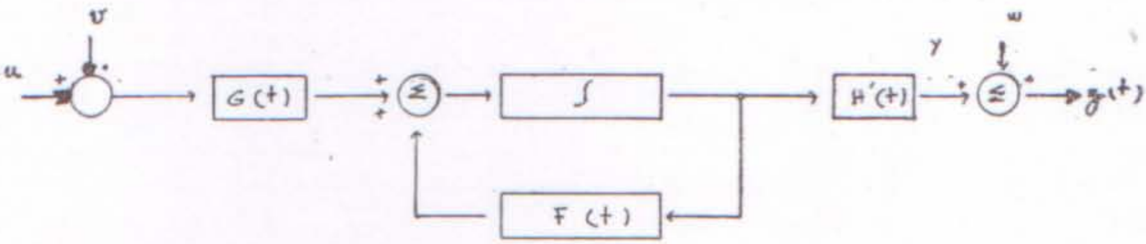

Fig 2

Generation of Signal and measurement processes

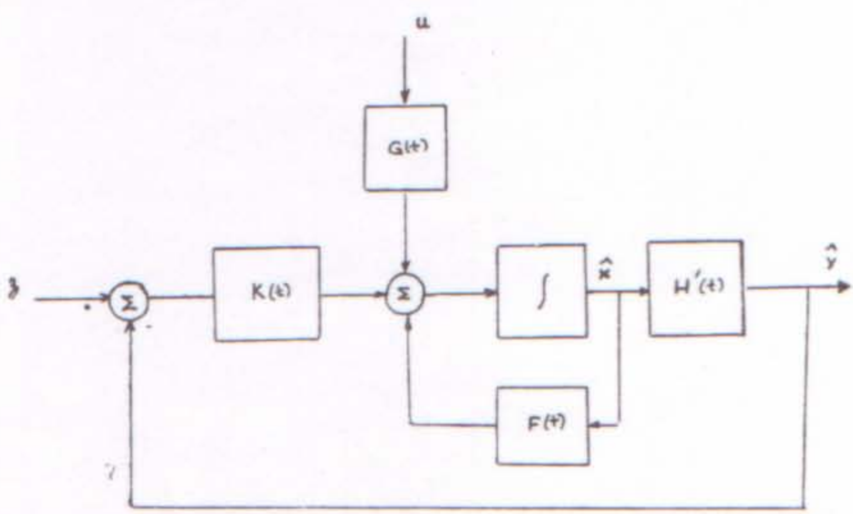

Fig. 3 The Kalman-Bucy Filter.

REFERENCES

(1) AN. Kolmogorov, "Interpolation and Extrapolation von stationaren Zufalligen Folgen", Bull, Acad. Sci., USSR, Ser. Math., Vol. 5, 1941.

(2) N. Wiener, "Extrapolation, Interpolation, and Smoothing of Stationary Time Series", John Wiley and Sons, New York, 1949,

(3) R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", Trans ASME, J. Basic Engineering, 82, 1960.

(4) R.E. Kalman and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory", Trans. ASME, J. Basic Engineering, Ser.D, 83, 1961.

(5) J.L. Doob, "Stochastic Processes", Wiley, New York, 1953.

(6) A.H. Jazwinski, "Stochastic Processes and Filtering Theory", Academic Press, London, 1970.

(7) B.D.O. Anderson and S. Vongpanitlerd, "Network Synthesis, A State Space Approach", Prentice Hall, 1973.

(8) H.L. Van Trees, "Detection, Estimation and Modulation Theory: Volume I", John Wiley & Sons, New York, 1968.

(9) B.D.O. Anderson and S. Chirarattananon, "Smoothing as An Improvement On Filtering : A Universal Bound", Electronics Letters Vol. 7, 1973.

(10) B.D.O. Anderson and T. Kailath, "The Choice of Signal Process Models in Kalman-Bucy Filtering", J. Mathematical Analysis and Applications, Vol. 35, 1971.

(11) R.S. Bucy, "Linear and Nonlinear Filtering", Proc. IEEE , Vol. 58, 1970.

(12) C.N. Kelly and B.D.O. Anderson, "On the Stability of Fixed - lag Smoothing Algorithms", J. Franklin Inst., Vol. 291, 1971.

(13) S. Chirarattananon and B.D.O. Anderson, "The Fixed-lag Smoother as a Finite Dimensional Linear System", Automatica, Vol. 7, 1973.

(14) S. Chirarattananon and B.D.O. Anderson, "Stable Fixed-lag Smoothing of Continuous-time Processes", IEEE - Transactions on Information Theory, Vol. 18, 1974.

(15) S. Chirarattananon, "Optimal and Suboptimal Linear Fixed - lag Smoothing", a PhD. dissertation, 1973.