A Qualitative Introduction to Wiener and Kalman-Bucy Filters

B. D. O. ANDERSON*†
ASSOCIATE MEMBER, I.R.E.E.

Summary

The aim of this paper is to introduce the reader to the problems which optimal linear filtering theory attempts to solve, to state in qualitative terms what optimal linear filters look like, to give sufficient guide to the literature on linear filtering to enable the reader to undertake further study and to note one or two research problems of current interest.

To try to ensure that the paper is available to as wide an audience as possible, the mathematics is very restricted. No derivations are included and the quantitative statement of results is generally omitted.

1. Introduction

The concept of tachometer feedback will be familiar to many readers. In an angle positioning servomechanism, stability of the system (or a desirable transient response) may be obtained by feeding back the position of the output shaft (proportional feedback) and the angular velocity of the output shaft (derivative feedback). The angular velocity may be determined with the aid of a tachometer. To a first approximation, the tachometer may be regarded as having a transfer function, Ks, that is, if $\theta(t)$ is the output shaft position at time t and $\bar{\theta}(s)$ is its Laplace transform, the Laplace transform of the tachometer output is $Ks\bar{\theta}(s)$. A more detailed analysis of the tachometer would probably show that the transfer function was more like $K\alpha s/(s + \alpha)$ where α is a very large positive constant; in other words, at low frequencies the tachometer would behave like a differentiator but at very high frequencies it would behave like a constant gain element.

It might be thought that a good tachometer design would be one where the constant α was very large, so that differentiation occurred over a wide frequency range, and in one sense this is quite true. From another point of view, however, it is not. The effect of approximate differentiation on noisy signals qualitatively is to cause magnification of the noise. The wider is the bandwidth of the noise, the greater is the magnification and the nearer to ideal is the approximate differentiation, the greater is the magnification. So in practice, an ideal differentiator would give terrible performance, because of the inevitable presence of a certain amount of very wide bandwidth noise (the presence of such noise in any physical system is predicted by statistical mechanics and practice bears out the theoretical prediction). Far from magnifying noise beyond bound, a tachometer actually filters the noise in the sense that it discriminates against high frequency noise through having lower gain at high frequencies.

In actual practice therefore, there are conflicting requirements on the choice of α and, if the requirements can be quantitatively stated, in a way which permitted quantitative comparison of the requirements, there would probably be an optimum value of α. Whatever the value of α however, the tachometer can still be regarded as a combination of a filter and a controlling element; quantitatively, the tachometer transfer function is the product of $\alpha/(s + \alpha)$, this term representing the filtering action, and Ks, a term representing the controlling action.

From the notions of tachometers, approximate differentiation and noise, it is easy to pass to the following sort of question. How should control system design be effected when all signals that can be measured are in fact noisy? The answer, at least in the case of linear design, is conceptually quite simple. Design can take place in two steps:

(1) A filter is determined which will process the noisy measurements, to eliminate as much of the noise as possible.

(2) A linear controller is designed whose input is assumed to be the original measurements uncontaminated with noise; in actual use, its input consists of the filter output.

*Department of Electrical Engineering, University of Newcastle, N.S.W.

†Work supported by the Australian Research Grants Committee.
Manuscript received by The Institution October 12, 1970.
Revised manuscript received by The Institution December 3, 1970.
U.D.C. number 621.372.54.

Fig. 1 illustrates the general idea for the case of a simple feedback system. Fig. 1(a) shows the arrangement that would be used if the measurements of y were noiseless; the controller in this case could conceivably contain differentiators. Fig. 1(b) shows the arrangement that could be used if the measurements, z, of the linear system output, y, are noisy; the filter recovers an estimate $\hat{y}$ of y which is fed into the linear controller. Note that the linear controllers of figs. 1(a) and 1(b) are the same.

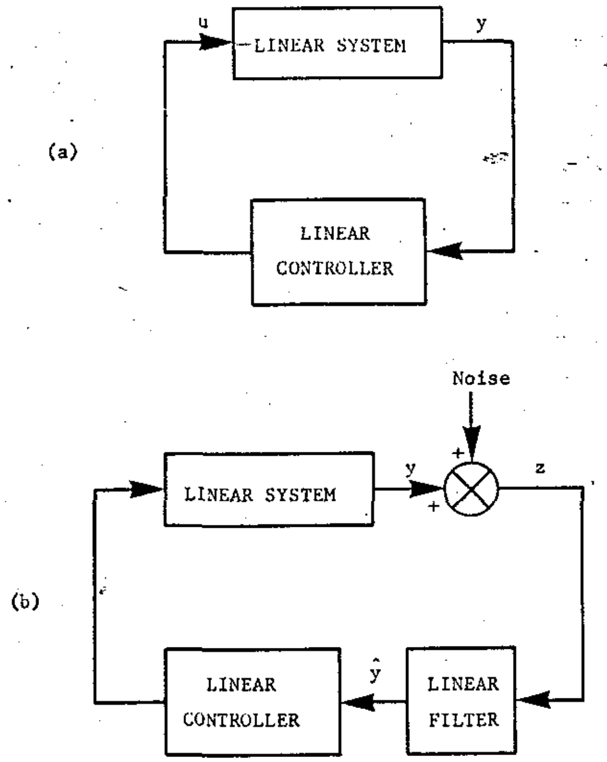

Figure 1.—*Feedback controller with*
(*a*) *Noiseless measurements*
(*b*) *Noisy measurements.*

The effect of the linear filter, besides filtering out some of the noise in z, is generally to band-limit the remaining noise, so that $\hat{y}$ consists of y contaminated by band-limited noise. This band-limited noise will certainly be magnified by any differentiating elements which might be in the linear controller but will not be magnified in an unbounded fashion, precisely because it is band-limited. Note also that the tachometer discussed earlier exemplifies this filter plus ideal controller idea.

The main purpose of this paper is to discuss features, particularly qualitative ones, of linear filters. To any control system designer faced with optimisation problems, a knowledge of filter design, particularly optimal filter design, is indispensable. Optimal filtering theory has great application to the design of communication systems too.

The bulk of the paper is devoted to a discussion of the two main types of optimal filter, the Wiener filter[1] and the Kalman-Bucy filter.[2–5] Introductions to the Wiener filter can be found in references 6 and 7 and to the Kalman-Bucy filter in references 8 and 9. Reference 6 does not require knowledge of state-space ideas but references 7 to 9 do require this knowledge; it is indispensable in the study of the Kalman-Bucy filter.

2. The Wiener Filter

The situation to which Wiener filtering is applied is depicted in fig. 2. In this figure s(·) is a signal process, n(·) a noise process and z(·) a measurement process, related to the signal and noise by

$$z(t) = s(t) + n(t) \qquad (1)$$

The Wiener filter is a device which generates from the noisy measurement, z(·), an on-line estimate of the signal, s(·).

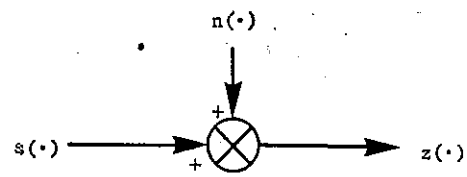

Figure 2.—Measurement process derived from signal plus noise.

2.1 *Fundamental Assumptions*

Several basic assumptions underpin the Wiener theory. First, *s*(·) *and n*(·) *are gaussian or normal processes* (in fact, jointly gaussian). Technically, this means that the joint probability density of $s(t_1), \ldots s(t_m), n(t_{m+1}), \ldots n(t_p)$ must be gaussian for any set of times, t_i, and any integers, m and p.[10] Practically, this means that many commonly occurring processes meet the assumption, though often only approximately. Examples of gaussian or approximately gaussian noise are:

(1) Thermodynamic noise, such as occurs in resistors.
(2) Shot noise, such as occurs in active devices.
(3) Noise arising from the superposition of huge numbers of tiny independent random disturbances (here, the central limit theorem[10] is at work).

1. Wiener, N., "Extrapolation, Interpolation and Smoothing of Stationary Time Series", M.I.T. Press, Cambridge, Mass. (1949).
2. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems", *J. Basic Engng., Trans. A.S.M.E.*, Series D, Vol. 82, No. 1, January 1960, p. 35.
3. Kalman, R. E. and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory", *J. Basic Engng., Trans. A.S.M.E.*, Series D, Vol. 83, No. 3, March 1961, p. 95.
4. Kalman, R. E., "New Methods and Results in Linear Filtering and Prediction Theory", Proc. Symp. Engng. Applications of Probability and Random Functions, John Wiley (1961).
5. Bucy, R. S. and Joseph, P. D., "Filtering for Stochastic Processes with Applications to Guidance", Interscience, New York (1968).
6. Franks, L. E., "Signal Theory", Prentice-Hall, N.J. (1969).
7. Schwarz, R. J. and Friedland, B., "Linear Systems", McGraw-Hill, N.Y. (1965).
8. Schultz, D. G. and Melsa, J. L., "State Functions and Linear Control Systems", McGraw-Hill, N.Y. (1967).
9. Sage, A. P., "Optimum Systems Control", Prentice-Hall, N.J. (1968).
10. Papoulis, A., "Probability, Random Variables and Stochastic Processes", McGraw-Hill, N.Y. (1965).

(4) Interference on earth-satellite communications channels.

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

Some commonly occurring noises are, however, not gaussian, including most channel noise found in point to point communications on earth.

Fortunately, Wiener filters offer " robust " performance, in that minor variations from normality of the probability densities produce only minor degradations in performance. Furthermore, one point of view which can be adopted in the absence of normality is that the Wiener filter is the best possible filter over a restricted class. We shall return to this point subsequently.

The means at time t of s(t) and n(t) are generally denoted by E[s(t)] and E[n(t)]. The auto-correlation of s(·) is denoted by

$$R_{ss}(t, \tau) = E[s(t)s(\tau)] \tag{2}$$

and of n(·) by

$$R_{nn}(t, \tau) = E[n(t)n(\tau)] \tag{3}$$

while the cross-correlation of s(·) and n(·) is denoted by

$$R_{sn}(t, \tau) = E[s(t)n(\tau)] \tag{4}$$

The second fundamental assumption of the Wiener Theory is that the means of s(·) and n(·), their auto-correlations and their cross-correlations are known.

The third fundamental assumption is that the processes are stationary,[10] namely,

$$E[s(t)] = E[s(t+T)] = m_s \quad \text{, say,} \tag{5}$$

$$E[n(t)] = E[n(t+T)] = m_n \quad \text{, say,} \tag{6}$$

where m_s and m_n are constants and the equations hold for all t and T ; also

$$R_{ss}(t, \tau) = R_{ss}(t+T, \tau+T) \tag{7}$$

$$R_{nn}(t, \tau) = R_{nn}(t+T, \tau+T) \tag{8}$$

$$R_{sn}(t, \tau) = R_{sn}(t+T, \tau+T) \tag{9}$$

for all t, τ and T. Physically, the last assumption means that the mechanism generating the random processes, s(·) and n(·), is not time-varying. It also means that *power spectra* can be associated with s(·) and n(·), these being frequency domain functions measuring the intensity of the component of the processes present at a certain frequency. In fact, it is measurement of such power spectra that offers one way for determining (experimentally) what the auto-correlations and cross-correlation are. The necessary theory for this exercise is contained in the Wiener-Khintchine theorem,[10] relating power spectra and auto-correlation functions.

Thermodynamic noise arising from a resistor is stationary, so long as the resistor is kept at a constant temperature [if the temperature varies, the intensity of the noise varies and the noise is then nonstationary]. The output of a stable linear system excited by stationary noise is also stationary.

Because the d.c. content of s(·) and n(·), namely, m_s and m_n , is frequently zero and because if it is nonzero it is very easy to subtract off prior to doing calculations, it is commonly assumed from the start that s(·) and n(·) are zero mean processes, that is, $m_s = m_n = 0$. In this case, the assumptions regarding s(·) and n(·) can be summed up as follows :

The processes s(·) and n(·) are jointly gaussian, with zero mean and known autocorrelations and cross-correlation.

The reader has probably assumed up to this point that s(·), n(·) and z(·) are scalar processes ; in fact they can be vector processes, that is, we may have noisy measurements, $z_1(\cdot)$, $z_2(\cdot)$, . . . $z_m(\cdot)$, of signal processes, $s_1(\cdot)$, $s_2(\cdot)$, . . . $s_m(\cdot)$, contaminated by noise processes, $n_1(\cdot)$, $n_2(\cdot)$, . . . $n_m(\cdot)$. Equation 1 then becomes a vector representation of the situation and all the preceding and succeeding remarks can be generalised to cope with this situation.

2.2 *Primary Problem*

In the primary problem we now describe, a further assumption on the noise processes is made, to the effect that n(·) *is a white noise process.* This means technically that its power spectral density is constant for all frequencies and that its auto-correlation function is of the form

$$R_{nn}(t, \tau) = N_0\,\delta(t - \tau) \tag{10}$$

where $\delta(\cdot)$ is the Dirac delta function ; because $\delta(t - \tau)$ is zero for $t \neq \tau$, equation 10 implies that the noise at successive instants of time is totally uncorrelated. Physically, white noise is an impossibility, though noise with a flat power spectrum up to very high frequencies is not. Thermodynamic resistor noise falls into this category, possessing a flat frequency spectrum essentially to optical frequencies. Wiener filters offer robust performance in the sense that the falling-off of the power spectrum of the noise at very high frequencies will not mar their performance greatly, even if the filters are designed on the basis that the power spectrum is constant.

An assumption is also made on the signal process s(·) ; this is that *its power spectrum falls off at high frequencies.* A signal process would have this property if it were derived at the output of a linear system or circuit driven by white noise, with the transfer function $T(j\omega)$ of the system going to zero as $\omega \to \infty$.

An optimum (Wiener) filter is a device whose input is the measurement z(·) and which at time t produces an estimate $\hat{s}(t)$ of s(t), based on the measurements $z(\tau)$, $\tau \leqslant t$; further, the estimate has the property that it is a minimum variance of error estimate, namely, the average value of $[s(t) - \hat{s}(t)]^2$ is less for the Wiener filter than for any other filter. The Wiener filter in other words computes that estimate which minimises the average mean square error between the estimate and the actual signal. (Note that the Wiener filter can also be regarded, for those familiar with the concept of conditional expectation,[10] as a device for computing $\hat{s}(t) = E[s(t)|z(\tau), \tau < t]$.)

Notice that the Wiener filter is some sort of causal system ; it is a system because it processes an input, namely, z(·), to yield an output, namely, $\hat{s}(\cdot)$. It is causal because the output at time t, namely, $\hat{s}(t)$, depends only on inputs up till time t and not on inputs after time t (if it were not causal, it would be impossible to build and operate the filter in real time).

The filter and its performance are computable from the data given; we shall comment at greater length on these points subsequently.

Other situations where Wiener filters become applicable are discussed in appendix 1. These include smoothing, predicting, derivative estimation, filtering with coloured additive noise and filtering with measurements available at discrete time instants.

2.3 *General Form of the Optimum Wiener Filter*

The basic feature of the optimum filter is that it is a *time-invariant linear system*; in other words, it is describable by a frequency domain transfer function. Of course this is of great practical significance, since linear systems can be built much more easily than nonlinear systems. A further feature is the robustness of the filter to variations from optimum; if the actual device used as a filter has a transfer function which is close to, but not the same as, the ideal, then performance will only deteriorate slightly.

As noted earlier, one of the fundamental assumptions made in order to apply the Wiener theory is that the underlying processes are normal. One might then well ask whether a filter derived on the assumption of normality of the processes is in any sense optimum if the processes are in actual fact not normal. The answer is yes. Among the class of linear filters, that computed via the Wiener theory is the best. Thus, if the processes are normal, the Wiener filter is truly optimum and, by good fortune, is linear, while if the processes are not normal and we seek the best filter that is linear, the Wiener filter does the job.*

Several additional comments regarding the optimum filter are relevant at this stage. First, the optimum filter may contain differentiators but usually does not; actually, its transfer function, $T(j\omega)$, usually goes to zero as $\omega \to \infty$, so that the Wiener filter usually is a band-limiting system. Certainly the Wiener filter associated with the primary problem defined earlier has this property.

Second, the Wiener filter is normally stable. This means that its transfer function has no right half plane poles and its construction and use are thus practical propositions.†

The third point is that a common form for realising the Wiener filter is that shown in fig. 3, that is, as a unity feedback system. Instead of building directly a system with transfer function T(s), the desired filter transfer function, one builds a filter of transfer function H(s), chosen so that

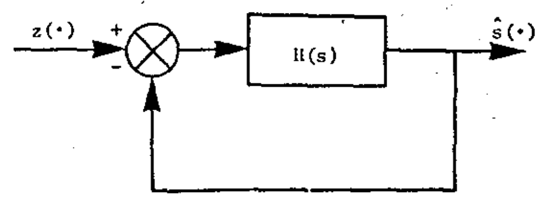

Figure 3.—Wiener filter realised as a unity feedback system.

$$\frac{H(s)}{1 + H(s)} = T(s) \qquad (11)$$

or so that the feedback system of fig. 3 has overall transfer function T(s). There can be computational advantage in seeking to find H(s) before T(s) and practical advantage in the implementation of fig. 3. This viewpoint is espoused in Van Trees.[11]

2.4 *Computation of the Wiener Filter*

The full derivation of the Wiener filter is quite extensive. From the point of view of a designer, the key points are that a formula is available for the transfer function of the optimum filter in terms of the power spectra of the signal and noise. Among the computations involved in this formula is spectral factorisation, which under normal circumstances is achieved by factoring polynomials in a special way. Spectral factorisation is thus potentially a burdensome computational task. In one case, it is not required, assuming an approximate filter will suffice, namely in a *high noise* (low signal-to-noise ratio) environment.

The concept of s.n.r. is a helpful one for describing qualitatively the shape of the amplitude response of the optimum filter. Roughly, at those frequencies where the s.n.r. is high, the amplitude response is high, while at those frequencies where the s.n.r. is low, the amplitude response is low. In other words, the amplitude response in some sense matches the variation of s.n.r. with frequency (for further remarks on computation, see appendix 2).

2.5 *Performance of the Optimum Wiener Filter*

The performance of the filter, as measured by the average mean square error $E[s(t) - \hat{s}(t)]^2$, is readily computable from R_{ss}, R_{sn}, R_{nn} and the optimum transfer function.‡ Only integration is required. The performance of a non-optimal filter may also be calculated in a reasonably straightforward way from the auto-correlations, cross-correlation and transfer function of the filter.

---

*Historically, the Wiener filter was first conceived as the best linear filter.

†In the case of fixed lag smoothing (see appendix 1), difficulty is encountered. The problem is that the transfer function of the filter tends to contain terms like $[\exp(-s\Delta) - \exp(-\alpha\Delta)](s-\alpha)^{-1}$ where α is a positive constant. Certainly the unstable pole $s = \alpha$ cancels with a zero at $s = \alpha$ but, after cancellation, the analytic form of the transfer function is extremely hard to define and it is not clear how to go about constructing a system with the desired optimum transfer function. This point, discussed in reference 12 is a subject of current research.

‡It is interesting to note[1] that in the case of the primary problem noted earlier, the filter transfer function depends only on the quantities $(R_{ss} + R_{sn})$ and R_{nn} rather than the quantities R_{ss}, R_{sn} and R_{nn} separately. Performance of the filter depends on particular R_{ss}, R_{sn} and R_{nn} and not just on $(R_{ss} + R_{sn})$ and R_{nn}. A similar phenomenon has recently been noticed in Kalman-Bucy filtering.[13]

11. Van Trees, H. L., "Detection, Estimation and Modulation Theory, Part I", John Wiley, N.Y. (1968).

12. Kelly, C. N. and Anderson, B. D. O., "On the Stability of Fixed-Lag Smoothing Algorithms", *J. Franklin Inst.* (to be published).

13. Anderson, B. D. O. and Moore, J. B., "The Kalman-Bucy Filter as a True Time-Varying Wiener Filter", *Trans. I.E.E.E.* (*Syst. Sci. and Cybernetics*), (to be published).

2.6 *Current Developments*

Since the advent of the Kalman-Bucy filter (which in many ways includes the Wiener filter), further research work on Wiener filtering theory has not been great. Among problems of current interest, we could note efforts in the area of quantitatively predicting the effect of various modelling errors, attempts at exploiting the full potential of the fixed lag smoother (see appendix 1) and attempts to apply the theory to sonar and array problems, involving multiple processes, noise with directional characteristics and so on.

3. The Kalman-Bucy Filter

In broad terms, the situation to which Kalman-Bucy filtering is applied is still that of fig. 2. In the course of this section however, we shall note structural requirements on the mechanisms for generating the signal process, $s(\cdot)$.

3.1 *The Conventional Assumptions*

As to the Wiener theory, it is assumed that $s(\cdot)$, $n(\cdot)$ and $z(\cdot)$ are gaussian random processes. There is however no assumption that the processes are stationary, which constitutes an extension of the Wiener theory.

The major difference with the Wiener theory arises through assumptions concerning the mechanism generating the signal process, $s(t)$. (Note that the case of vector $s(t)$ is very easily treated with the Kalman-Bucy theory, in some contrast to the Wiener theory and accordingly we shall not assume that $s(\cdot)$ is a scalar.) It is assumed that there is known the state-space equations of a linear system with a deterministic input and a white noise output and an output $s(t)$;

$$\dot{x} = F(t)x + G(t)u + v \quad (12)$$

$$s = H'(t)x \qquad z(t) = s(t) + n(t) \quad (13)$$

Here, x is the state of the system, u is a known deterministic input and v is a zero mean, gaussian white noise process, with known covariance matrix

$$E[v(t)v'(\tau)] = Q(t)\delta(t-\tau) \quad (14)$$

(the superscript prime denotes matrix transposition). In addition, it is necessary to state some initial condition for equation 12. The precise statement is as follows: the initial time for equation 12 is t_0, with $t_0 = -\infty$ allowed; $x(t_0)$ is a gaussian random variable of known mean, m, and known covariance matrix

$$E\{[x(t_0) - m][x(t_0) - m]'\} = P_0 \quad (15)$$

Further, $x(t_0)$ is independent of $v(\tau)$ for all $\tau \geq t_0$. Fig. 4 illustrates the arrangement.

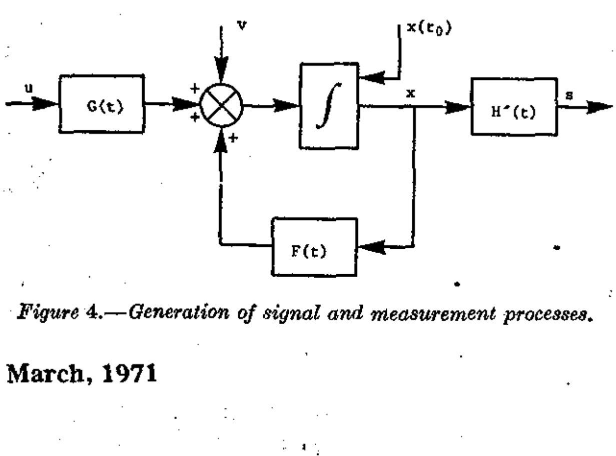

Figure 4.—Generation of signal and measurement processes.

The above assumption concerning the mechanism generating $s(\cdot)$ is enormously more detailed than in the Wiener theory. Leaving aside the fact that $s(t)$ can be a nonstationary process, being the output of a time-varying linear system, the assumption requires:

(1) $s(\cdot)$ to be the output of a band-limiting system describable by ordinary differential equations (that is, if F, G and H were constant, one could say that there existed a transfer function matrix relating the Laplace transform of u to the Laplace transform of s, with this transfer function matrix rational and tending to zero as $s \to \infty$).

(2) A full state-space description of the system generating $s(\cdot)$ to be known. Such a description is not implicit in $R_{ss}(t, \tau)$, in the sense that many different systems will lead to the same $R_{ss}(t, \tau)$.

In addition to the assumption on $s(\cdot)$, for the primary problem considered in the Kalman-Bucy theory, it is required that $n(\cdot)$ be gaussian white noise of zero mean, that is,

$$E[n(t)n'(\tau)] = R(t)\delta(t-\tau) \quad (16)$$

with, actually, $R(t)$ nonsingular for all t. The process $n(\cdot)$ is assumed independent of $x(t_0)$ and, for convenience, independent of $v(\cdot)$, although this latter assumption can generally be removed.[4]

In view of the assumption that $n(\cdot)$ is white and the fact that this assumption is removable only with some work,[14] it makes sense to compare the assumptions behind the primary Wiener and Kalman-Bucy problems. In a nutshell, the differences are as follows.

(1) The Wiener theory demands stationarity, the Kalman-Bucy theory does not; this permits initial times, t_0, which are not $-\infty$, a fact of great practical significance in many situations.

(2) The Kalman-Bucy theory demands that the process, $s(t)$, be generated as the output of a *finite-dimensional* system driven by white noise but the Wiener theory does not.

(3) The Kalman-Bucy theory demands detailed knowledge of the signal generating process, whereas the Wiener theory requires simply knowledge of the auto-correlation and cross-correlation functions R_{ss}, R_{sn} and R_{nn}.

The first difference is a plus for the Kalman-Bucy theory, while the second is not. Alternative procedures based on integral equation solution are available now which permit removal of the finite-dimensionality constraint.[15, 16] The third difference will be discussed in the next subsection.

---

14. Bryson, A. E. and Johansen, D. E., " Linear Filtering for Time-Varying Systems Using Measurements Containing Coloured Noise ", *Trans. I.E.E.E.*, Vol. AC-10, No. 1, January 1965, p. 4.
15. Gohberg, I. C. and Krein, M. G., " On the Factorization of Operators in Hilbert Spaces ", *Am. Math. Soc. Trans.*, Vol. 51, 1966, p. 155.
16. Kailath, T., " Fredholm Resolvents, Wiener-Hopf Equations, and Riccati Differential Equations ", *Trans. I.E.E.E.*, Vol. IT-15, No. 6, November 1969, p. 665.

An excellent historical perspective of the Kalman-Bucy filter is provided by Sorenson.[17] Essentially, Kalman-Bucy filtering is an extension of least squares estimation procedures used by Gauss, to estimate the planetary orbits. The extension is of course substantial. Extensions to the Gauss procedure that went halfway towards the Kalman-Bucy filter were obtained by various workers in the late 1950's as reported by Sorenson, while one worker, Swerling, published a paper in 1959,[18] prior to the papers of Kalman and Bucy, which in all but a quite insignificant respect was identical with the later work of Kalman and Bucy on discrete time filtering. In the author's opinion, it would be proper for Swerling to receive the credit for deriving the design of the (discrete time) Kalman-Bucy filter but undoubtedly the credit for the detailed performance analysis of the filter must go to Kalman and Bucy. It is interesting to note that in a recent paper of Stratonovich,[19] there occurs a claim that the results of Kalman and Bucy on continuous time filtering were published earlier by Stratonovich. Even if Stratonovich is assigned priority for deriving the filter design, again the credit for performance analysis must go to Kalman and Bucy.

3.2 *Alternative Assumptions*

It has recently been found,[13, 20, 21] that less restrictive assumptions are actually required to compute a Kalman-Bucy filter design than noted in the previous subsection. It is still the case:

(1) $s(\cdot)$ is the output of some linear *finite-dimensional* system excited by white noise.

(2) $n(\cdot)$ consists of zero mean, gaussian, white noise with covariance as in equation 15 and $R(t)$ nonsingular.

Instead, however, of requiring detailed knowledge of the system generating $s(\cdot)$, it is simply required that:

(3) the functions R_{ss} and R_{sn} are prescribed.

This change of basic assumptions brings the Kalman-Bucy theory of course into much closer parallel with the Wiener theory; moreover, the change has practical significance, since there exist nonstationary filtering problems (see, for example, reference 22) where detailed knowledge of the mechanisms generating $s(\cdot)$ is not present but knowledge of R_{ss} and R_{sn} is present.

3.3 *The Primary Problem*

In contrast to the primary problem of Wiener filtering theory, there are two primary problems of the Kalman-Bucy filtering theory. The most fundamental is *to construct a device whose input is the measurement $z(\cdot) = s(\cdot) + n(\cdot)$ and which at time t produces an estimate, $\hat{x}(t)$, of $x(t)$, $x(t)$ being the state of the system generating $s(t)$, based on the measurements $z(\tau)$, $\tau < t$; further, the estimate has the property that each entry of $\hat{x}(t)$ is a minimum variance of error estimate of the corresponding entry of $x(t)$, (or equivalently $E[\{a'[x(t) - x(t)]\}^2]$ is smaller for all t and all constant vectors, a, than the corresponding quantity computed for any other filter).*

The following variations with the Wiener filtering problem should be noted. First, as stated, the aim is to estimate $x(t)$ rather than $s(t)$; second, the concept of a minimum variance of error estimate has to be modified, to cope with the vector situation. Otherwise, the problem is virtually the same. Again for example, those familiar with the concept of conditional expectation will recognise that $\hat{x}(t) = E[x(t)|z(\tau), \tau < t]$.

As noted above there is more than one primary problem associated with the Kalman-Bucy theory and, as the reader may have surmised, the second problem is to generate an estimate, $\hat{s}(t)$, of $s(t)$. However, once the first problem is solved, the second is also; the estimate $\hat{s}(t)$ of $s(t)$ which is optimum in the minimum variance of error sense is obtained simply by setting

$$\hat{s}(t) = H'\hat{x}(t) \tag{17}$$

As for the Wiener filter, the Kalman-Bucy filter is some sort of causal system. Presently, we shall discuss its structure in more detail. Meanwhile, we wish to note a situation where the problem of estimating $\hat{x}(t)$, rather than simply $\hat{s}(t)$, is highly relevant.

Linear, time-invariant, finite-dimensional systems (that is, those with rational transfer functions) have been studied for many years; more recently, the advantages of using state-variable descriptions, particularly for multiple-input or multiple-output systems, have become evident (see, for example, references 7 and 8). Such descriptions look like

$$\dot{x} = Fx + Gu \tag{18}$$

$$y = H'x \tag{19}$$

where u, x and y are, respectively, the system input, state and output. The quantities F, G and H are real constant matrices. The design of controllers which position the closed-loop transfer function matrix poles at desired values can be regarded as a problem of selecting a state feedback law

$$u = K'x + u_{ext} \tag{20}$$

and implementing that law.[8, 23] If the system states are available, the implementation is straightforward. If not, then it is necessary to compute a state estimate, $\hat{x}$, from the output, y, to implement

$$u = K'\hat{x} + u_{ext} \tag{21}$$

17. Sorenson, H. W., "Least-squares Estimation: from Gauss to Kalman", *I.E.E.E. Spectrum*, Vol. 7, No. 7, July 1970, p. 63.
18. Swerling, P., "A Proposed Stage-wise Differential Correction Procedure for Satellite Tracking and Prediction", Report P1292, Rand Corp., Santa Monica, Calif., U.S.A., January 1958, also published in *J. Astronaut Sci.*, Vol. 6, 1959.
19. Stratonovich, R. L., "Detection and Estimation of Signals in Noise when One or Both are Non-gaussian", *Proc. I.E.E.E.*, Vol. 58, No. 5, May 1970, p. 670.
20. Anderson, B. D. O. and Moore, J. B., "State Estimation Via the Whitening Filter", Proc. Joint Automatic Control Conf., Ann Arbor, Michigan, June 1968, p. 123.
21. Anderson, B. D. O. and Moore, J. B., "Solution of a Time-Varying Wiener Filtering Problem", *Electronics Letters*, Vol. 3, No. 12, December 1967, p. 562.
22. Bucy, R. S., "Linear and Nonlinear Filtering", *Proc. I.E.E.E.*, Vol. 58, No. 6, June 1970, p. 854.

This is where the Kalman-Bucy filter is applicable; it provides a way of estimating the state of equations 18 and 19, assuming the statistical properties of certain measurement noises are known, and at the same time achieving the estimate with minimum mean square error.

Additional examples of the applicability of Kalman-Bucy filters are noted in appendix 3.

3.4 *General Form of the Kalman-Bucy Filter*

The optimum Kalman-Bucy filter, like the Wiener filter, is a causal linear system. Unlike the Wiener filter however, it has significantly more structure. To start with, it is a finite-dimensional system. In addition though, it inherently contains a model of the system generating $s(t)$. Fig. 5 depicts a Kalman filter; the linear gain, $K(t)$, is determined via a sequence of calculations, while other gain blocks are all derived directly from the signal generating model. Notice that the filter has inputs comprising $u(\cdot)$, the deterministic input to the signal generating model, and the noisy measurements $z(\cdot)$. Notice too that $K(t)$ multiplies the error between $z(\cdot)$ and the estimate $\hat{s}(t)$ of $s(t)$, which would be zero if there were no noise and perfect tracking of $s(t)$ by $\hat{s}(t)$. Inspection of fig. 5 shows where the estimates $\hat{x}(t)$ and

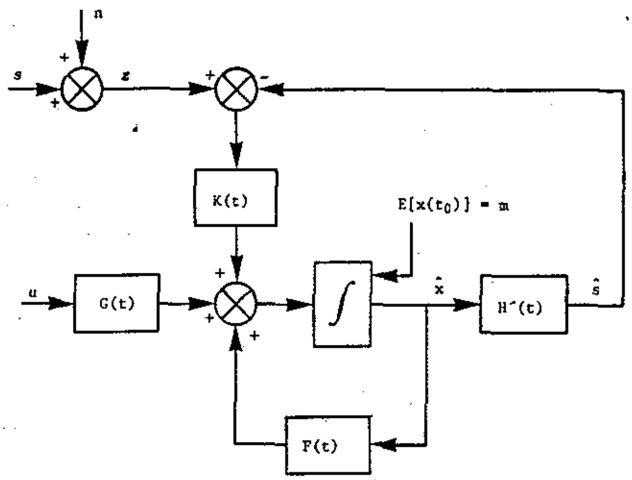

Figure 5.—Structure of the Kalman-Bucy filter.

$\hat{s}(t)$ are picked off and illustrates the relation (equation 17) between these quantities.

There are a number of parallels with the Wiener filter that we now indicate briefly.

(1) If the underlying noise processes, $v(\cdot)$ and $n(\cdot)$, are not gaussian, then the Kalman-Bucy filter is only the optimum filter among the set of linear filters, rather than the optimum of all possible filters.

(2) The arrangement of fig. 5 applies to the primary filtering problem. If, for example, $n(\cdot)$ is not white, the optimum filter will look quite different and perhaps contain differentiators.[14]

(3) It is generally the case that the optimum filter is stable. An important set of conditions guaranteeing stability (which involve highly technical concepts, admittedly becoming simple if F, G, H, Q and R are constant) is contained in reference 3. A further set of conditions has been recently discovered.[24]

(4) If the signal process, $s(\cdot)$, is stationary and the noise process, $n(\cdot)$, is stationary,* then the Kalman-Bucy filter will be time-invariant, that is, $K(t)$ will be constant. Roughly speaking, a stationary filtering problem gives rise to a time-invariant filter.

In measuring the performance of the Kalman-Bucy filter, it proves convenient to use the error covariance matrix

$$P(t) = E\{[x(t) - \hat{x}(t)]\,[x(t) - \hat{x}(t)]'\} \tag{22}$$

Notice that for a constant vector, a,

$$a'P(t)a = E\{a'[x(t) - \hat{x}(t)]^2\} \tag{23}$$

so that, by definition, the Kalman-Bucy filter minimises $a'P(t)a$ for all a and for all t.

It turns out that there is a differential equation for obtaining $P(t)$ and the filter gain, $K(t)$, is straightforwardly determined as $K = PHR^{-1}$, so the bulk of the computational effort† lies in determining P.

Reference and background on the derivation of these results are noted in appendix 4.

3.6 *Rapprochement with the Wiener Theory*

Suppose that, in a problem to which the Wiener theory applied, we had knowledge of the system generating the signal process, $s(t)$. Suppose further that this system was finite dimensional. Then it would be possible to build a Kalman-Bucy filter to provide a minimum variance of error estimate of $s(t)$, which, since the goals of the Wiener and Kalman-Bucy theories are the same, must be the same as the associated Wiener filter. This abstract argument can be verified by calculations (see, for example reference 26).

---

*The usual situation giving rise to the stationarity of s(.) and n(.) is, in technical terms, one where F, G, H, Q and R are constant, $t_0 = -\infty$ and F has left half plane eigenvalues.

†The differential equation is a nonlinear matrix Riccati equation. The boundary condition on the differential equation is a one-point one, namely, $P(t_0) = P_0$ and, despite the nonlinearity of the equation, there always exists a solution. The solution can be computed by straightforwardly integrating the equation forward in time on a digital or analog computer, since, fortunately, it is usually computationally stable. Alternatively, it is possible to set up a related linear differential equation and obtain from this the matrix $P(t)$ by some manipulations.[3] Finally, in case F, G, H, Q and R are all constant, there are many available numerical procedures for solving the equation[23] and, if $t_0 = -\infty$, these procedures are even simpler, leading to a constant value of $P(t)$.

Another case where the calculations drastically simplify arises when the signal to noise ratio is very low. This high noise case is discussed in reference 25; the conclusion is that, at worst, linear differential equations need to be solved to determine the optimal filter, at least to good approximation.

23. Anderson, B. D. O. and Moore, J. B., " Linear Optimal Control ", Prentice-Hall, N.J. (1971).

24. Anderson, B. D. O., " Stability Properties of Kalman-Bucy Filters ", *J. Franklin Inst.* (to be published).

25. Anderson, B. D. O. and Moore, J. B., " Optimal State Estimation in High Noise ", *Information and Control*, Vol. 13, No. 4, 1968, p. 286.

26. Melsa, J. L., " Frequency-Domain Derivation of the Stationary Kalman Filter ", Proc. First Ann. Houston Conf. on Circuits, Systems and Computers, May 1969.

3.7 *Further Miscellaneous Points*

In this subsection, we shall offer brief remarks on the following points : error analysis, the divergence problem, adaptive filtering, the innovations approach to filtering, nonlinear filtering and system identification from covariances.

3.7.1 *Error Analysis*

A question of practical interest is to indicate how the actual performance of a Kalman-Bucy filter will differ from the predicted performance in the face of modelling errors. For example, if it is assumed that the input noise covariance matrix is $Q_1\delta(t - \tau)$ for the purposes of filter design, when it is actually $Q_2\delta(t - \tau)$, one wants to know the resultant effects. For a survey of results in error analysis, see reference 27. In general, the results fall into two classes :

(1) Results giving quantitative variation in performance, based on quantitative information concerning modelling errors ;

(2) Qualitative results, indicating, for example, that if all noises are assumed greater in intensity than they really are, the actual filter performance will be better than that predicted.

3.7.2 *Divergence Problem*

A problem common to linear and nonlinear filtering and observed in orbit estimation work of several years ago, is that of some small modelling error having an ever increasing deleterious effect on the estimation. Thus analysis might indicate that the error variance in estimating s(t) falls off as $1/t$ as $t \to \infty$, while an experiment might reveal that the error variance rises as fast as t when $t \to \infty$. It has now been established that there are certain subtle kinds of modelling error which, no matter how small, can cause this drastic difference between theory and practice. More importantly, techniques for coping with the problem are now understood (see references 27 and 28). Broadly speaking, the techniques rely on either artificially increasing the noise intensity or rejecting information derived from old observations.

3.7.3 *Adaptive Filtering*

Because the design of a Kalman-Bucy filter does not demand stationary noise processes, one can conceive of the filter adjusting its design as the noise intensities change. Some steps in the direction of an adaptive filter design have been made[29] but far more remains to be done in this difficult problem area.

3.7.4 *Innovations Approach to Filtering*

It has been known for some time (see, for example, reference 30) that the " error signal " or " innovations process " $z(t) - \hat{s}(t)$, (see fig. 5) is white noise. This can be exploited in deriving new insights into the Kalman-Bucy filter (see references 20 and 31).

3.7.5 *Nonlinear Filtering*

Virtually all excursions into problems of nonlinear filtering which have led to practical results have in some way been based on the Kalman-Bucy filter. Two books containing material on nonlinear filtering are references 5 and 27 ; spectacular applications, some justifying empirical approaches of some years standing, may be found in references 32 and 33. The phase-locked-loop for the demodulation of frequency modulated waveforms may, for example, be viewed as an approximately optimal nonlinear filter.

3.7.6 *System Identification from Covariances*

A problem of both theoretical interest and practical importance is that of identifying from a prescribed correlation function, $R_{ss}(t, \tau)$, a system which, when driven by white noise, would possess an output with the prescribed correlation function. The problem can be solved via integral equation theory[15, 16] if there is no constraint that the generating system be finite-dimensional. Much simpler procedures where the finite-dimensionality constraint is in force are available for single-output systems in references 34 and 35, with some interesting sidelights to be found in reference 36. The multiple output problem is currently under investigation.

4. Future Directions

It would be inappropriate in a paper of this character to offer any general conclusions ; rather, we feel it important to indicate directions in which progress in filtering theory and application might take place.

In the linear filtering area, there do remain several outstanding problems. These include the obtaining of satisfactory solutions to the fixed-lag smoothing problem (see appendices 1 and 3) and the adaptive filtering problem and the generation of a sound and broadly applicable body of knowledge concerning suboptimal filtering. Other minor problems also remain, some of which we have noted in the text ; for example, there are some areas of error analysis which are still unexplored.

27. Jazwinski, A. H., " Stochastic Processes and Filtering Theory ", Academic Press, N.Y. (1970).

28. Anderson, B. D. O., " Exponential Data Weighting in the Kalman-Bucy Filter ", *(submitted for publication)*.

29. Mehra, R. K., " On the Identification of Variances and Adaptive Kalman Filtering ", *Trans. I.E.E.E.*, Vol. AC-15, No. 2, April 1970, p. 175.

30. Collins, L. D., " Realizable Whitening Filters and State-Variable Realizations ", *Proc. I.E.E.E.* (Correspondence). Vol. 56, No. 1, January 1968, p. 100.

31. Kailath, T., " An Innovations Approach to Least-Squares Estimation : Part I—Linear Filtering in Additive White Noise ", *Trans. I.E.E.E.*, Vol. AC-13, No. 6, December 1968, p. 665.

32. Snyder, D., " The State-Variable Approach to Continuous Estimation, with Applications to Analog Communication Theory ", M.I.T. Press, Cambridge, Mass. (1969).

33. Van Trees, H. L., " Detection, Estimation and Modulation Theory—Part II : Nonlinear Modulation Theory ", John Wiley, N.Y. (1970).

34. Anderson, B. D. O., Moore, J. B. and Loo, S. G., " Spectral Factorization of Time-Varying Covariance Functions ", *Trans. I.E.E.E.*, Vol. IT-15, No. 5, September 1969, p. 550.

35. Moore, J. B. and Anderson, B. D O., " Spectral Factorization of Time-Varying Covariance Functions : The Singular Case ", *Math. Syst. Theory*, Vol. 4, No. 1, 1970, p. 10.

36. Anderson, B. D. O., and Kailath, T., " The Choice of Signal-Process Models in Kalman-Bucy Filtering ", *J. of Math. Analysis and Applications*, (to be published).

The really fertile area for further development must surely be the nonlinear filtering area. Generally, there is still a search for practically workable filtering algorithms, accompanied by some general statements as to their applicability. If progress in nonlinear optimal control is any guide, progress could be a long and painful process.

Appendix 1

Additional Examples of Applicability of Wiener Filters

The theory applicable to computation of the Wiener filter is extendable to a number of other situations which we note briefly here. For further details, reference 1 should be consulted.

(1) *Smoothing Filters*

The arrangement of fig. 2 still applies but now the idea is to produce an estimate at time t of $s(t - \Delta)$, for some constant $\Delta > 0$. (This problem has been termed "smoothing with fixed lag", the term smoothing being used since measurements occurring before *and after* the time argument of the estimated quantity are used in estimating the quantity and the term fixed lag being used since the time argument of the estimated quantity is Δ seconds prior to the running time, that is, at the present time, the filter produces an estimate of what the signal was like Δ seconds ago.) Notice that a smoothed estimate of $s(t - \Delta)$ will inevitably be better than an estimate of $s(t - \Delta)$ produced according to our basic filtering scheme, since more data is used in producing the smoothed estimate; the data used in produced a filtered estimate of $s(t - \Delta)$ using the basic scheme are the measurements up till time $t - \Delta$, while to this data are added the measurements from $t - \Delta$ to t in producing a fixed lag smoothed estimate. Many expositions of the Wiener theory are pre-occupied with the case $\Delta = \infty$, which appears to have virtually no practical significance. The resulting Wiener filter has been termed the "optimum unrealisable filter", which seems an accurate enough description.[7, 11] It is possible that the interest arising in this case stems from the comparative ease of computability of the transfer function.

(2) *Predicting Filters*

The arrangement of fig. 2 still applies but now the idea is to predict some future value of $s(\cdot)$, that is, to produce an estimate at time t of $s(t + \Delta)$ for some $\Delta > 0$.

(3) *Derivative Filters*

The arrangement of fig. 2 still applies but now the aim is to estimate $\dot{s}(t)$ rather than s(t) (for example, s(t) may represent an angular position). In fact, the theory is extendable to the problem of estimating the signal resulting from any linear operation on s(t), for example, to the problem of estimating the output of a system with input s(t) and with transfer function

$$s + \frac{s + 1}{(s + 2)(s^2 + 3s + 1)}\,.$$

(4) *Coloured Noise Filtering*

The arrangement of fig. 2 still applies but now $n(\cdot)$ need not be white (if $n(\cdot)$ is not white, it is termed coloured). In general, the power spectrum of $n(\cdot)$ has to fall off slower than that of $s(\cdot)$ for the use of a Wiener filter to be a practical proposition.

(5) *Discrete Time Filtering*

Instead of the arrangement in fig. 2, wherein measurements are received in continuous time, the measurements are received at discrete instants of time, perhaps as a result of sampling of a continuous time process (such will usually be the case if signals or measurements are derived after transmission through a digital communication channel). The first worker to solve the problem of optimally filtering discrete time random processes was Kolmogorov,[37] rather than Wiener, the latter being initially concerned with continuous time processes. The results for discrete and continuous time processes are closely parallel and thus it would not be unfair to ascribe the development of optimal filtering jointly to Kolmogorov and Wiener.

(6) *Combinations*

It is of course possible to combine the above possible extensions and also possible simultaneously to have $s(\cdot)$, $n(\cdot)$ and $z(\cdot)$ vector processes.

Appendix 2

Wiener Filter Computation

The justification of the formula for the Wiener filter transfer function and its computation falls naturally into several steps, as follows.

(1) Using the calculus of variations in a very simple fashion, the impulse response of the optimum filter is shown to satisfy an integral equation, called the Wiener-Hopf equation.

(2) Using the device of spectral factorisation, the integral equation (in which the running variable is time) is transformed into an equation in the frequency domain using the Fourier transform.

(3) The frequency domain equation is solved, to yield the transfer function of the optimum filter.

We comment briefly on various aspects of this procedure. The basic reason for converting the time-domain integral equation into the frequency domain is not that the transfer function of the optimum filter is defined in the frequency domain as an end result of the process, but rather that use of frequency domain manipulations, with advanced complex variable theory, yields a solution to the integral equation. In fact, it is only recently that direct time-domain procedures have become available for solving the integral equation (see reference 15 for the basic results and reference 16 for their application to solution of the Wiener-Hopf equation).

The frequency domain equation resulting in step 2 can also be derived by a procedure due to Bode and

37. Kolmogorov, A. N., "Interpolation and extrapolation", *Bulletin de l'academie des sciences de U.S.S.R.*, Ser. Math., Vol. 5, 1941, p. 3.

Shannon.[38] Their treatment of Wiener filtering provides remarkable intuitive insights and is strongly recommended to those unfamiliar with the field.

Spectral factorisation is a mathematical operation about which enormous amounts have been written and techniques for carrying out spectral factorisation are still being researched (for example, references 1 and 39 to 44).

The frequency domain function to which the spectral factorisation procedure must be applied is the power spectrum of z, or the Fourier transform of $R_{zz}(t, 0)$. (In a few cases, spectral factorisation is not possible but, in this case, the initial filtering problem is ill-posed. A test, called the Paley-Wiener test,[1] can be applied to check the question of whether spectral factorisation is possible). If the power spectrum of z is rational, (commonly the case in practice) spectral factorisation requires factoring of the numerator and denominator polynomials of the power spectrum to isolate the factors with left half plane zeros. If, however, the power spectrum of z is not rational, a much more sophisticated technique based on use of the Hilbert transform is required to achieve the spectral factorisation. Note that, as remarked in the text, in a low s.n.r. problem, spectral factorisation is not required in computing an approximate filter. If this approximate design procedure were extendable to nonlinear filtering problems, as it might be, this would constitute a significant advance, since most nonlinear filters are poor approximations to the optimum at low s.n.r.'s. For example, the phase-locked-loop for the reception of angle modulated signals is an approximation to the optimum filter, the approximation being better at high s.n.r.'s.[32] Performance at low s.n.r.'s is marred by cycle skipping. So there would appear scope for improvement of filters for angle modulated signals at low s.n.r.

Appendix 3

Additional Examples of Applicability of Kalman-Bucy Filters

It is possible to remove the assumption that the additive noise n(·) be white.[14] The removal does not permit n(·) to be arbitrary, but requires that n(·) be the output of a linear, *finite-dimensional* system excited by zero mean, gaussian, white noise.

The removal of the assumption of finite-dimensionality requires a new approach to the nonstationary problem. One such approach is based on the use of an eigenvalue-eigenfunction of $R_{zz}(t, s)$, called a Karhunen-Loève expansion ; in fact, this has been a traditional approach for a number of years (for a recent discussion, see reference 11). A second approach involves advances in integral equation techniques, obtained in reference 15. This is reported in reference 16. As might be expected, the two techniques, Karhunen-Loève expansion and integral equation, can always be applied to situations where the finite-dimensionality assumption is in force. Which is the better technique from the computational point of view has yet to be resolved. A complete rapprochement between the Karhunen-Loève approach and the Kalman-Bucy approach has yet to be achieved ; preliminary results may be found in references 45 and 46. Rapprochement between the integral equation approach and the Kalman-Bucy approach is better understood and is surveyed in reference 16.

The *prediction problem* is one that can be very easily tackled, using the most minor of extensions of the Kalman-Bucy theory. The aim is to generate, at time t, using measurements, $z(\tau)$, $\tau < t$, an estimate of $x(t - \Delta)$, or $s(t - \Delta)$, for some $\Delta > 0$, with, of course, minimum variance of the estimation error (see reference 3 for a treatment of this problem).

On the other hand, the smoothing problem has caused great difficulty. Here, the problem is to obtain at time t, using measurements, $z(\tau)$, $\tau < t$, an estimate of $x(t - \Delta)$ or $s(t - \Delta)$ for some $\Delta > 0$, with, again, minimum variance of the estimation error. This problem has only been solved very recently ; reference 47 contains an extensive account, while reference 48 contains an instructive alternative derivation of the form of the optimal smoother. It has recently been discovered[12] that virtually all smoothers are unstable, implying physically that smoothers will not work, at least for any but small intervals of time. Research is currently in progress aimed at overcoming this problem.

Finally, we note that although the filtering problem we have posed is a continuous time problem, there exists a discrete time problem, the formulation and solution of which are very similar to the continuous time problem.[2]

38. Bode, H. W. and Shannon, C. E., " A Simplified Derivation of Linear Least Square Smoothing and Prediction Theory ", *Proc. I.R.E.*, Vol. 38, No. 4, April 1950, p. 417.

39. Wiener, N. and Masani, L., " The Prediction Theory of Multivariate Stochastic Processes ", Pts 1 and 2, *Acta Mathematica*, Vol. 98, June 1958.

40. Youla, D. C., " On the Factorization of Rational Matrices ", *Trans. I.R.E.*, Vol. IT-7, No. 3, July 1961, p. 172.

41. Wong, E. and Thomas, J. B., " On the Multidimensional Prediction and Filtering Problem and the Factorization of Spectral Matrices ", *J. Franklin Inst.*, Vol. 272, No. 2, August 1961, p. 87.

42. Loo, S. G., " Spectral Factorization by Means of Augmented Factors ", *Electronics Letters*, Vol. 3, No. 6, June 1967, p. 238.

43. Anderson, B. D. O., " An Algebraic Solution to the Spectral Factorization Problem ", *Trans. I.E.E.E.*, Vol. AC-12, No. 4, August 1967, p. 410.

44. Anderson, B. D. O., " The Inverse Problem of Stationary Covariance Generation ", *J. Statist. Phys.*, Vol. 1, No. 1, 1969, p. 133.

45. Baggeroer, A. B., " A State-Variable Approach to the Solution of Fredholm Integral Equations ", *Trans. I.E.E.E.*, Vol. IT-15, No. 4, August 1969, p. 557.

46. Anderson, B. D. O., " Karhunen-Loève Expansion for a Class of Covariances ", Proc. Int. Conf. on Syst. Sci., Hawaii, 1969, p. 779.

47. Meditch, J. S., " Stochastic Optimal Linear Estimation and Control ", McGraw-Hill, N.Y. (1969).

48. Kailath, T. and Frost, P., " An Innovations Approach to Least-Squares Estimation : Part II—Linear Smoothing in Additive White Noise ", *Trans. I.E.E.E.*, Vol. AC-13, No. 6, December 1968, p. 646.

Appendix 4

Derivation of the Kalman-Bucy Filter

One of the earliest procedures for the derivation involved taking the discrete time results (which are more easily derived) and letting the sampling interval tend to zero.[27] Other approaches rely on posing the filtering problem as a least squares estimation problem,[27] on regarding the problem as a particular case of a more general nonlinear filtering problem,[27] on solving a maximum a posteriori estimation problem[27] and on setting up an equivalent linear optimal control problem.[23] Kalman and Bucy[3] proceeded by deriving an integral equation (the Wiener-Hopf equation), necessarily satisfied by the impulse response of the optimum filter, and showing that solution of the differential equation for P(t) effectively allowed solution of the integral equation. An approach for deriving the filter design and performance from the restricted set of data noted in subsection 3.2 may be found in references 13 and 20; this approach is closest in spirit to the Wiener theory and the derivation of the Wiener filter by Bode and Shannon.[38]

The Control Function in Management

R. LAYTON*, I. D. MAIN† and J. S. RATCLIFFE‡
FELLOW, I.R.E.E.

Summary

People segregate naturally into work groups. Such work groups are made the basis of a model of the firm. The individual group is first analysed and shown to be adequately described by a first order differential equation.

The firm is then described in terms of a network of groups, and the overall response of various sequences of groups is discussed.

Attention is then turned to management and its role in setting and maintaining response parameters is described.

While the overall response function of the firm to its commercial environment is probably too complex to be of immediate use, the model is equally valid for any coherent portion of the firm.

While it is not applicable to all firms, or to all aspects of management, it is intended to give the engineer some means by which his engineering skills can be applied to some of the problems of management.

1. Introduction

When an engineer is promoted to a managerial position, he often feels that the hard-earned skills which served him so well as an engineer have lost their value. In many respects this is true but some engineering topics do have some application in the field of management.

In this paper we offer a model of the factory based on control theory which may assist the engineer in gaining a better understanding of some of the problems that confront him.

*Professor of Marketing, University of New South Wales, Kensington, N.S.W.

†Chief Chemist, Snowy Mountains Hydro-Electric Authority and M.Sc. Candidate, University of New South Wales, Kensington, N.S.W.

‡Professor of Chemical Engineering, University of New South Wales, Kensington, N.S.W.

Manuscript received by The Institution October 30, 1970.
U.D.C. number 62.007 : 658.

2. The Social Structure of Industry

When numbers of people work together, they tend to segregate into groups. The size of group is variable, rarely more than twelve and usually between four and ten. Members of the group usually share a common goal, possibly a common work place. This group develops a great deal of social cohesion and usually accepts one of its members as a leader and spokesman.

This phenomenon is analysed by a large number of management texts[1] and its broader social and industrial consequences are described by Brown.[2] It is also the basis of the model described in this paper.

3. Scope

It is intended that the model we discuss should apply to the day-to-day operation of the factory. It is intended to describe the fluctuations in production of a factory in response to the random individual fluctuations imposed on it by the normal incidents of commerce.

1. *Examples are*:
   Likert, R., "New Patterns of Management", McGraw-Hill (1961).
   Likert, R., "The Human Organisation, its Management and Value", McGraw-Hill (1967).
   McGregor, D., "The Human Side of Enterprise", McGraw-Hill (1960).
   Argyris, C., "Integrating the Individual and the Organisation", Wiley (1964).
2. Brown, J. A. C., "The Social Psychology of Industry", Penguin, Middlesex (1954).