

Kalman Filtering: Whence, What and Whither?

B. D. O. Anderson and J. B. Moore

Department of Systems Engineering, Research School of Physical Sciences and Engineering,
The Australian National University, GPO Box 4, Canberra ACT 2601, Australia

1 Introduction

Undoubtedly, one of the major contributions of R.E. Kalman has been the Kalman filter, [1,2], the magnitude of the contribution being specifically recognized in the award of the Kyoto Prize.

In this contribution, we shall try to put the Kalman filter in historical perspective, by recalling the state of filtering theory before its appearance, as well as some of the major developments which it spawned. It is impossible to be comprehensive in the allotted space, especially so in making a selection from major developments.

2 The Wiener Filter

Wiener filtering [3, 4] probably represented the first domain of endeavour where two important ideas were merged: dynamic systems, and optimal estimation in the presence of noise. The quantity being estimated is a “moving target”, being associated with the evolution of a dynamic system. More precisely, consider the arrangement of Fig. 2.1, which depicts a signal $y(\cdot)$, contaminating noise $v(\cdot)$ and a measurement $z(\cdot)$. Continuous and discrete-time problems can be formulated, and vector $y(\cdot), v(\cdot)$ and $z(\cdot)$ are permitted. Nevertheless, for convenience, we shall consider scalar continuous time signals only. All signals are defined on the time interval $(-\infty, \infty)$.

It is assumed that $y(\cdot)$ and $v(\cdot)$ are sample functions of stationary random processes. Normally, they are zero mean, and independent. Further, they are assumed to have known spectra, $\phi_{yy}(j\omega)$ and $\phi_{vv}(j\omega), \omega \in R$. (Certain well-behavedness properties, whose particular form does not concern us here, must be fulfilled by the spectra also.)

The task of Wiener filtering is to use the measurements $z(\cdot)$ to estimate $y(\cdot)$. More precisely, the estimation is required to be causal, on-line and optimal. Causal means that $y(t)$ is to be estimated using $z(s)$ for $s < t$; on-line means that at time t , the estimate of $y(t)$ should be available, and this should hold for all

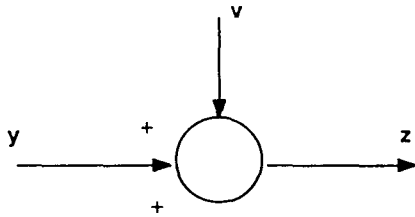


Fig. 2.1. Signal model for Wiener filter

t , i.e. as t evolves; optimal means that the estimate, call it $\hat{y}(t)$, should offer minimum mean square error, i.e. $E\{[y(t) - \hat{y}(t)]^2\}$ should be minimized. In case $y(\cdot)$ and $v(\cdot)$ are jointly gaussian, this means that $\hat{y}(t)$ is a conditional mean estimate, viz $E[y(t)|z(s), s < t]$.

The solution to this problem is depicted in Fig. 2.2. The block labelled "Wiener filter" is a linear, time-invariant, causal, stable system, describable by an impulse response $h(\cdot)$ or transfer function, thus

$$\hat{y}(t) = \int_{-\infty}^t h(t-s)z(s)ds \tag{2.1}$$

The signal $y(\cdot)$ and noise $v(\cdot)$ are often represented as the output of stable linear systems excited by white noise, see Fig. 2.3. If $\varepsilon_y(\cdot)$, $\varepsilon_v(\cdot)$ denote zero mean, unit variance, white noise, i.e.

$$E[\varepsilon_y(t)\varepsilon_y(s)] = E[\varepsilon_v(t)\varepsilon_v(s)] = \delta(t-s) \tag{2.2}$$

Then

$$\phi_{yy}(j\omega) = |W_y(j\omega)|^2, \quad \phi_{vv}(j\omega) = |W_v(j\omega)|^2 \tag{2.3}$$

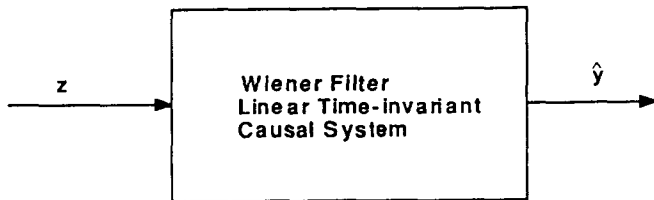


Fig. 2.2. Structure of Wiener filter

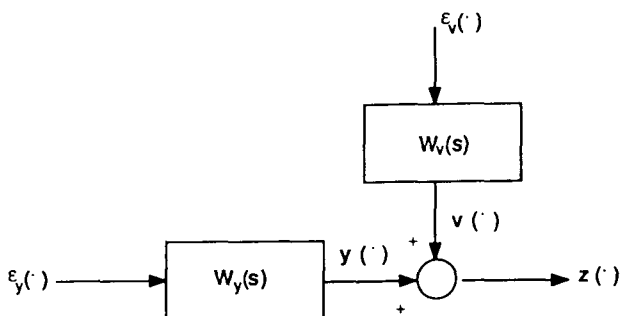


Fig. 2.3. Alternative signal model for Wiener filter

It is clear that the key problem in Wiener filtering is to describe the procedure which leads from the pair $\phi_{yy}(j\omega), \phi_{vv}(j\omega)$ (or $W_y(j\omega)$ and $W_v(j\omega)$) to the impulse response $h(t)$ or its transfer function $H(j\omega)$. The most crucial step is that of *spectral factorization*. The spectrum of $z(\cdot)$, at least when $y(\cdot)$ and $v(\cdot)$ are independent, is given by

$$\phi_{zz}(j\omega) = \phi_{yy}(j\omega) + \phi_{vv}(j\omega) \quad (2.4)$$

and the spectral factorization operation requires the determination of a transfer function $W_z(j\omega)$ such that $W_z(s)$ and $W_z^{-1}(s)$ are analytic in $\text{Re } s \geq 0$ [i.e. $W_z(\cdot)$ is causal and causally invertible] such that

$$\phi_{zz}(j\omega) = |W_z(j\omega)|^2 \quad (2.5)$$

In [3], this spectral factorization operation is presented as a crucial step in the calculation of $H(\cdot)$, while in [4], it is the key to a well-motivated derivation of the optimum filter. The point behind this derivation is the following. Define a signal $\varepsilon_z(\cdot)$ as the output of a linear system of transfer function $W_z^{-1}(j\omega)$ driven by $z(\cdot)$. Because of the causality and causal invertibility of $W_z(\cdot)$, then $\varepsilon_z(\cdot)$ is informationally equivalent to $z(\cdot)$, i.e. estimation of $y(t)$ using $\varepsilon_z(s)$ for $s < t$ must give the same result as estimation of $y(t)$ using $z(s)$ for $s < t$; but also, $\varepsilon_z(\cdot)$ is white. This is a very significant simplification, and is the key to the derivation of the optimum filter in [4].

We noted above that construction of $W_z(\cdot)$ satisfying stability constraints and (2.5) is a crucial step in the construction of $H(\cdot)$. The question arises as to how this can be done. If $\phi_{zz}(\cdot)$ is rational, polynomial factorization is the key. Otherwise, one uses

$$W_z(j\omega_0) = \min_{\varepsilon \rightarrow 0} \exp \left\{ \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\log \phi_{zz}(j\omega)}{-j(\omega - \omega_0) - \varepsilon} d\omega \right\} \quad (2.6)$$

The case of vector $z(\cdot)$ and matrix $\phi_{zz}(\cdot)$ is much more complicated.

The following table sums up some important characteristics of the Wiener filtering problem and its solution:

Table 1. Assumptions for and properties of the Wiener filter

Initial time t_0	$t_0 = -\infty$
Random Processes	Stationarity required
Signal $y(\cdot)$	(i) Spectrum $\phi_{yy}(\cdot)$ enough, but $W_y(j\omega)$ is acceptable. (ii) $y(\cdot)$ is stationary, $W_y(j\omega)$ must be stable. (iii) $\phi_{yy}(\cdot)$ and $W_y(\cdot)$ are not necessarily rational.
Measurement Noise $v(\cdot)$	(i) Usually independent of $y(\cdot)$ (ii) Stationary (iii) Not necessarily white (iv) $\phi_{vv}(\cdot)$ is not necessarily rational.
Wiener filter	Time invariant and stable but not necessarily with rational transfer function.
Main Calculation Burden	Spectral factorization
Quantity estimated	$y(t)$

3 The Kalman Filter (Continuous Time), [2]

We shall first recall the standard Kalman filter set-up, and then compare it with the Wiener filter. The signal model is (see Fig. 3.1)

$$\frac{dx(t)}{dt} = F(t)x(t) + G(t)w(t) \quad (3.1a)$$

$$z(t) = H'(t)x(t) + v(t) \quad (3.1b)$$

in which F, G, H are $n \times n, n \times m$, and $n \times p$ respectively. The processes $w(\cdot)$ and $v(\cdot)$ define zero mean white noise, such that

$$E \begin{bmatrix} w(t) \\ v(t) \end{bmatrix} \begin{bmatrix} w'(s) & v'(s) \end{bmatrix} = \begin{bmatrix} Q(t) & S(t) \\ S'(t) & R(t) \end{bmatrix} \delta(t-s) \quad (3.2)$$

with $R(t) = R'(t) > 0$ for all t . Very frequently, $S(t) \equiv 0$, i.e. $w(\cdot)$ and $v(\cdot)$ are independent. We shall make this assumption. Then $Q(t) = Q'(t) \geq 0$.

In the first instance, we shall assume a finite initial time t_0 . Further, $x(t_0)$ will be assumed to be a gaussian random variable, of mean x_0 and covariance P_0 . Equation (3.1a) defines a Gauss-Markov model—since $x(\cdot)$ is a Markov process which is gaussian. [In fact, $[x'(\cdot) \ z'(\cdot)]'$ is also gaussian and Markov.]

The estimation task is to use measurements of $z(s)$ for $s < t$ to estimate $x(t)$; the estimate, call it $\hat{x}(t)$, is to be on line and to minimize $E[\|x(t) - \hat{x}(t)\|^2]$. (This means that $\hat{x}(t)$ is necessarily a conditional-mean estimate.) The now well-known solution is obtained as follows. Define $P(t) = P'(t) \geq 0$ as the solution of

$$\dot{P} = PF' + FP - PHR^{-1}H'P + GQG' \quad P(t_0) = P_0 \quad (3.3)$$

and $\hat{x}(\cdot)$ by (see Figure 3.2)

$$\frac{d\hat{x}}{dt} = F(t)\hat{x}(t) + P(t)H(t)R^{-1}(t)[z(t) - H'(t)\hat{x}(t)] \quad (3.4)$$

(Note that Fig. 3.2 depicts a standard observer, in terms of structure, with a special gain $P(t)H(t)R^{-1}(t)$, sometimes referred to as the Kalman gain.) Further, there holds

$$E\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]'\} = P(t) \quad (3.5)$$

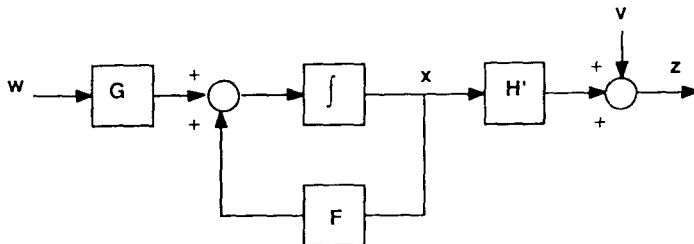


Fig. 3.1. Signal model for Wiener filter

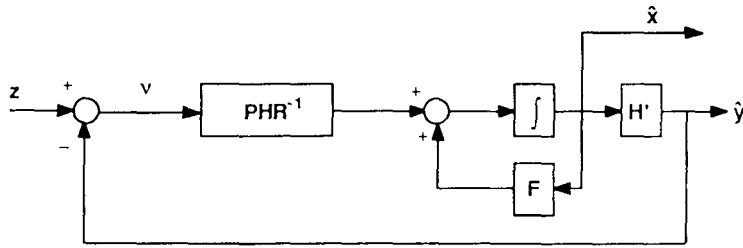


Fig. 3.2. Kalman filter

i.e. the performance of the optimal estimator, as measured by the error covariance, is given through the solution of (3.3). (The existence of a solution to (3.3) in (t_0, ∞) has to be proved, but this is not difficult.)

Let us now study some differences with the Wiener filter, considering the points listed in Table 1.

Initial time. In the above formulation, we have taken $t_0 > -\infty$. This can be adjusted, see the next section. So the Kalman filter allows either $t_0 > -\infty$ or $t_0 = -\infty$, whereas the Wiener filter allows only $t_0 = -\infty$.

Random processes. The exciting random processes $v(\cdot)$ and $w(\cdot)$ in the above signal model are not necessarily stationary, since $Q(\cdot)$ and $R(\cdot)$ in (3.2) may be time-varying. With $t_0 > -\infty$, we cannot have stationarity. Finally, time-variable $F(t)$, $G(t)$, $H(t)$ mean that, even if $Q(t)$ and $R(t)$ were constant and t_0 were $-\infty$, $z(\cdot)$ would not have to be stationary. Thus, in contrast to the Wiener filter, stationarity is *normally* not guaranteed or expected. Under special circumstances, it can be present however. We discuss this point further in the next section.

Signal $y(\cdot)$. It is clearly appropriate to regard $H'x$ in the signal model (3.1) as corresponding to the signal $y(\cdot)$ in the previous section. For the Wiener filter, it was enough to know the spectrum of $y(\cdot)$, or equivalently, its covariance. There is such a quantity as a nonstationary covariance, and so in principle one could contemplate constructing the Kalman filter knowing just the covariance of $y(\cdot)$. Whether or not this is possible, (it is, see [5]) this is certainly not the way the Kalman filter was presented—rather, a detailed signal model of $y(\cdot)$ is required, rather like requiring $W_y(j\omega)$ in the Wiener filter problem as opposed just to $\phi_{yy}(j\omega)$. Next, we note that for the Kalman filter problem, $y(\cdot)$ does not have to be stationary, and the linear system linking $w(\cdot)$ to $y(\cdot)$ does not have to be time-invariant or stable. This linear system however does have to be finite dimensional. In the time-invariant case, this would correspond to $W_y(\cdot)$ being rational, which is not a requirement in Wiener filtering. In summary, the cost of dropping the stationarity assumption of the Wiener filtering problem is the requirement for a model of $y(\cdot)$, not just of its covariance, and a finite-dimensional model into the bargain.

Measurement noise $v(\cdot)$. In both the Wiener and Kalman formulations, the noise is usually independent of $y(\cdot)$. In the Wiener formulation, it is stationary, but this is not required for the Kalman formulation. The major difference is that the noise is required to be white in the Kalman formulation. This is not

Table 2. Wiener filter and Kalman filter key differences

Wiener	Kalman
$t_0 = -\infty$	$t_0 \geq -\infty$
Stationarity	Nonstationarity acceptable
Infinite dimensional OK	Finite dimensional
Measurement noise not necessarily white	Measurement noise white
Spectral factorization	Riccati equation solution
Signal estimation	State estimation

required in the Wiener formulation (though it turns out that whiteness does carry with it significant simplifications).

Filter. The Kalman filter is in general time varying, stability is not guaranteed (and of course, over a finite interval it is of limited relevance). It is finite dimensional. The Wiener filter may not be, since its transfer function is not necessarily finite dimensional.

Main calculation Burden. Spectral factorization and Riccati matrix solution, the two key computational tasks, could not appear more dissimilar.

Quantity estimated. The Wiener filter estimates $y(t)$, the Kalman filter $x(t)$. In the Kalman filter formulation, $y(t) = H'(t)x(t)$, and an estimate of $y(t)$ follows as

$$\hat{y}(t) = H'(t)\hat{x}(t) \quad (3.6)$$

Table 2 summarizes the key differences.

The problem of prediction is almost as easily solved as that of filtering. This is the task of estimating $x(t + \Delta)$ for some positive Δ , given $z(s)$ for $s < t$. There holds

$$\hat{x}(t + \Delta) = \Phi(t + \Delta, t)\hat{x}(t) \quad (3.7)$$

where $\Phi(\cdot, \cdot)$ is the transition matrix of $F(\cdot)$. The smoothing problem (Δ is negative instead of positive) took some time to resolve; it is discussed later.

4 Kalman Filter Over an Infinite Time Interval

The deepest contribution of [2] lies in the treatment of infinite time problems. Here, concepts of controllability and observability were cleverly used to tackle issues such as time invariance and stability. We shall review the main ideas here.

Suppose that F, G, H, Q and R are all time invariant, with t_0 remaining finite.

Result 1. Given observability of $[F, H]$, $P(t)$ is bounded, and

$$\bar{P} = \min_{t \rightarrow \infty} P(t) \quad (4.1)$$

exists and satisfies the steady state Riccati equation

$$\bar{P}F' + F\bar{P} - \bar{P}HR^{-1}H'\bar{P} + GQG' = 0 \quad (4.2)$$

The boundedness of $P(t)$ is intuitively reasonable, because under observability, it is not surprising that the error in estimating $x(t)$, viz $P(t)$, should remain bounded. There are a number of other consequences of this result:

- (i) the Kalman filter (3.4) is asymptotically time invariant.
- (ii) if $t_0 \rightarrow -\infty$, rather than $t \rightarrow \infty$, the fact that the right side of the differential equation (3.3) has no explicit time dependence yields

$$\bar{P} = \min_{t_0 \rightarrow -\infty} P(t)$$

for all t .

- (iii) if $t_0 \rightarrow -\infty$, the signal model (3.1) with constant F, G may produce unbounded $E[x(t) x'(t)]$ unless $\text{Re } \lambda_i(F) < 0$, i.e. unless it is stable. And if t_0 remains finite and $t \rightarrow \infty$, the same is true.

Result 1 says nothing about stability of the Kalman filter, nor of the dependence of \bar{P} on P_0 . The circle is closed with Result 2. Again, we suppose that F, G, H, Q and R are time invariant.

Result 2. Suppose $[F, H]$ is observable and $[F, GQ^{1/2}]$ is controllable. Then \bar{P} as defined in Result 1 is independent of P_0 , and

$$\text{Re } \lambda_i(F - \bar{P}HR^{-1}H') < 0$$

Notice that this stability property is just what is required to ensure that the Kalman filter (3.4) is asymptotically stable.

Summarizing, if $\text{Re } \lambda_i(F) < 0$, with constant F, G, H, Q and R and with $t_0 \rightarrow -\infty$, the signals $x, y = H'x, v$ and z are all stationary, the Kalman filter is time invariant and asymptotically stable provided observability and controllability conditions are fulfilled. (Even if $\text{Re } \lambda_i(F) < 0$ fails, the latter statements concerning the Kalman filter remain true.)

The parallel with the Wiener filter becomes more apparent in this result.

Let us note that observability and controllability are a little stronger than needed; in fact, it is not hard to relax these requirements to detectability and stabilizability, see e.g. [5, 6].

Even in the nonstationary case, it still makes sense to contemplate the possibility of $t \rightarrow +\infty$, and to ask about the stability of the Kalman filter and the forgetting of initial conditions. A resolution of these questions was really suggested by the observation of [1, 2] that the Kalman filter problem in many ways is a dual of the linear-quadratic regulator problem, where infinite-time behavior and stability are key issues, even for time-varying systems. The fundamental paper [7] had dealt with these issues, and duality pointed the way to the corresponding filtering results:

Result 1 (TV). Suppose $[F(t), H(t)]$ is uniformly completely observable, and $F(t), G(t), H(t), Q(t)$ and $R(t)$ are bounded. Then $P(t)$ is bounded for all $t \in [t_0, \infty]$.

Moreover, if $t_0 \rightarrow -\infty$,

$$\lim_{t_0 \rightarrow -\infty} P(t) = \bar{P}(t)$$

exists and satisfies the differential equation (3.3).

Result 2 (TV). In addition to the hypothesis of Result 1 (TV) suppose that $[F(t), G(t)Q^{1/2}(t)]$ is uniformly completely controllable. Then $\bar{P}(t)$ is independent of P_0 , and the system $\dot{p} = (F(t) - \bar{P}(t)H(t)R^{-1}(t)H'(t))p$ is exponentially stable.

Relaxation of the observability/controllability to detectability/stabilizability is surprisingly difficult, and took some time, [8].

5 Kalman Filter (Discrete Time), [1]

Practically all that has been stated for the continuous-time Kalman filter carries over to the discrete-time filter. There is however one idea in the discrete-time theory that is made more transparent than in the continuous-time theory, and because of its applicability to more general problems it is worth recording. The signal model is

$$x_{k+1} = F_k x_k + G_k w_k \quad (5.1a)$$

$$z_k = H_k' x_k + v_k \quad (5.1b)$$

with

$$E \begin{bmatrix} w_k \\ v_k \end{bmatrix} \begin{bmatrix} w_l' & v_l' \end{bmatrix} = \begin{bmatrix} Q_k & 0 \\ 0 & R_k \end{bmatrix} \delta_{kl} \quad (5.2)$$

and $\{w_k\}, \{v_k\}$ are zero mean sequences. For convenience, let the initial time be $k=0$. Then the data include the mean \bar{x}_0 and variance P_0 of x_0 , which is independent of $\{w_k\}, \{v_k\}$. All variables are gaussian.

The key idea is to *distinguish* the effect of dynamics and measurements in the filter. More precisely, let $\hat{x}_{k/k}$ be the optimal estimate, again a conditional mean estimate, of x_k given $z_l, l \leq k$, and let $\hat{x}_{k+1/k}$ be $E[x_{k+1} | z_l, l \leq k]$, the one-step prediction estimate. Since w_k is independent of z_l for $l \leq k$, (5.1a) yields

$$\hat{x}_{k+1/k} = F_k \hat{x}_{k/k} \quad (5.3)$$

This shows how to update an estimate as a result of the system dynamics, when no extra measurements appear. Along with (5.3), there holds

$$\Sigma_{k+1/k} = F_k \Sigma_{k/k} F_k' + G_k Q_k G_k' \quad (5.4)$$

Hence $\Sigma_{k/k}$ and $\Sigma_{k+1/k}$ are the error covariances associated with $\hat{x}_{k/k}$ and $\hat{x}_{k+1/k}$.

The measurement update equations indicate how to pass from $\hat{x}_{k+1/k}$ and $\Sigma_{k+1/k}$ to $\hat{x}_{k+1/k+1}$ and $\Sigma_{k+1/k+1}$. They are

$$\hat{x}_{k+1/k+1} = \hat{x}_{k+1/k} + \Sigma_{k+1/k} H_{k+1} [H'_{k+1} \Sigma_{k+1/k} H_{k+1} + R_{k+1}]^{-1} \cdot [z_{k+1} - H'_{k+1} \hat{x}_{k+1/k}] \quad (5.5a)$$

$$\Sigma_{k+1/k+1} = \Sigma_{k+1/k} - \Sigma_{k+1/k} H_{k+1} [H'_{k+1} \Sigma_{k+1/k} H_{k+1} + R_{k+1}]^{-1} \cdot H'_{k+1} \Sigma_{k+1/k} \quad (5.5b)$$

Observe that F_k, G_k and Q_k enter only in the time or dynamical update equation, while H_k and R_k enter only in the measurement update equation. This separate accounting for dynamics and measurements, necessarily blurred in the continuous-time filter equations, is optimal in the discrete-time equations.

Some of these ideas are also to be found in [9].

6 Development 1: Spectral Factorization and Innovation

The key computational tool in Wiener filtering is spectral factorization, and in Kalman filtering, it is the Riccati equation.

Since there are some problems which can be treated either by Wiener or Kalman filtering methods (those where $t_0 = -\infty$, $W_y(j\omega) = H'(j\omega I - F)^{-1} G Q^{1/2}$, and $v(\cdot)$ is white), there should be a connection between the computational tools for each approach, and indeed, this is now reasonably explored even in books, see, e.g. [5, 10, 11]. We shall touch lightly on the connection here.

Take $t_0 = -\infty, F, G, H, Q$ and R constant, with F stable and the observability/controllability condition fulfilled. From the steady-state Riccati equation (4.2), it is possible to prove

$$\begin{aligned} (I + H'(sI - F)^{-1} K) R (I + K'(-sI - F')^{-1} H) \\ = R + H'(sI - F)^{-1} G Q G'(-sI - F')^{-1} H \end{aligned} \quad (6.1)$$

where

$$K = \bar{P} H R^{-1} \quad (6.2)$$

The quantity on the right side of (6.1) is the spectrum of the measurement process $z(\cdot)$. The left side defines a spectral factorization. Notice that

$$[I + H'(sI - F)^{-1} K]^{-1} = I - H'(sI - F + H K')^{-1} K$$

and $F - K H' = F - \bar{P} H R^{-1} H'$, which is guaranteed stable. So the spectral factorization produces a transfer function (matrix) which is stable, together with its inverse. Evidently, the Riccati equation has led to a spectral factorization. Conversely, if one has the spectral factorization (obtained say by Wiener filtering methods), and is able to represent the spectral factor in the form

$(I + H'(sI - F)^{-1}K)R^{1/2}$ then (4.2) and (6.2) imply

$$\bar{P}(F' - HK') + (F - KH')\bar{P} = -GQG' - KRK' \quad (6.3)$$

and this shows that \bar{P} can be defined as the solution of a linear Lyapunov matrix equation. So the two apparently distinct filter calculations, spectral factorization and (steady state) Riccati equation solution, are effectively equivalent in this case.

A related result concerns the so-called innovations process. Consider Fig. 3.2, and suppose that $P = \bar{P}$, with F, H and R all time-invariant. Then the transfer function matrix from z to $v = z - H'\hat{x}$ can be computed to be

$$\begin{aligned} I - H'(sI - F + PHR^{-1}H')^{-1}PHR^{-1} &= I - H'(sI - F + KH')^{-1}K \\ &= [I + H'(sI - F)^{-1}K]^{-1} \end{aligned} \quad (6.4)$$

It follows that the spectral matrix of v , which is termed the innovations process, is

$$\begin{aligned} \Phi_{vv}(j\omega) &= [I + H'(j\omega I - F)^{-1}K]^{-1} \Phi_{zz}(j\omega) [I + K'(-j\omega I - F')^{-1}H]^{-1} \\ &= R \end{aligned} \quad (6.5)$$

i.e. $v(\cdot)$ is a white noise process. This remarkable result continues to hold even in the nonstationary case, see [12] for a further discussion, though the proof is obviously very different. The observation has proved useful in developing various extensions of the basic theory, motivating conjectures, etc. It suggests that the time-varying Riccati equation is achieving a form of time-varying spectral factorization. (Indeed, this is true, see [13].) It also makes nice contact with the highly motivated derivation of the Wiener filter of [4].

7 Development 2: Smoothing

Filtering is concerned with obtaining $E[x(t)|z(s), s \in [t_0, t]]$ on line. Smoothing is concerned with obtaining one or more of the following:

$$\begin{aligned} E[x(t_0)|z(s), s \in [t_0, t]] & \quad (\text{fixed } t_0, \text{ varying } t) \\ E[x(t)|z(s), s \in [t_0, t + \Delta]] & \quad (\text{fixed } t_0, \Delta, \text{ and varying } t) \\ E[x(t)|z(s), s \in [t_0, T]] & \quad (\text{fixed } t_0, T, \text{ varying } t) \end{aligned}$$

These tasks are known as fixed point, fixed lag and fixed interval smoothing. We shall now outline how these problems can be solved.

Fixed point smoothing. Consider the standard signal model, augmented by a second set of integrators which are undriven and unmeasured. The initial condition on these integrators is identical with that on the main signal model. The full model is then

$$\begin{bmatrix} \dot{x} \\ \dot{x}_a \end{bmatrix} = \begin{bmatrix} F & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ x_a \end{bmatrix} + \begin{bmatrix} G \\ 0 \end{bmatrix} w \quad (7.1a)$$

$$z = [H' \quad 0] \begin{bmatrix} x \\ x_a \end{bmatrix} + v \quad (7.1b)$$

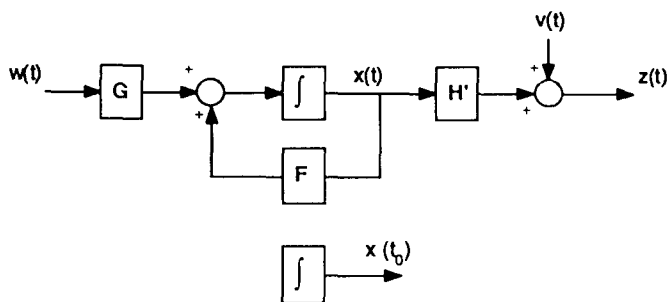


Fig. 7.1. Augmented signal model for fixed point smoothing

with

$$E \begin{bmatrix} x \\ x_a \end{bmatrix} = \begin{bmatrix} x_0 \\ x_0 \end{bmatrix}$$

$$E \left\{ \begin{bmatrix} x \\ x_a \end{bmatrix} - \begin{bmatrix} x_0 \\ x_0 \end{bmatrix} \right\} \left\{ \begin{bmatrix} x \\ x_a \end{bmatrix} - \begin{bmatrix} x_0 \\ x_0 \end{bmatrix} \right\}' = \begin{bmatrix} P_0 & P_0 \\ P_0 & P_0 \end{bmatrix} \quad (7.2)$$

The set-up is a *standard one* in terms of the Kalman filter. The best estimate of x_a , viz $\hat{x}_a(t)$, is $E[x_a(t)|z(s), s < t]$. However, $x_a(\cdot)$ is so constructed that $x_a(t) = x_a(t_0) = x(t_0)$ for all t . So $\hat{x}_a(t) = E[x(t_0)|z(s), s < t]$.

The Riccati equation for the augmented system decomposes into the Riccati equation for the original model plus linear equations, and the construction of the smoothed estimate is not at all hard.

This approach was suggested by [14, 15]. It could not really have come out of a theory requiring stationarity.

Fixed lag smoothing. This problem was examined in the Wiener filtering context, see, e.g. [3]. The optimum smoother is normally infinite dimensional (unless $\Delta = \infty$), and this does not augur well for a Kalman filtering approach. However, by switching to discrete time we avoid this problem. Consider the standard discrete-time signal model, augmented as depicted in Fig. 7.2.

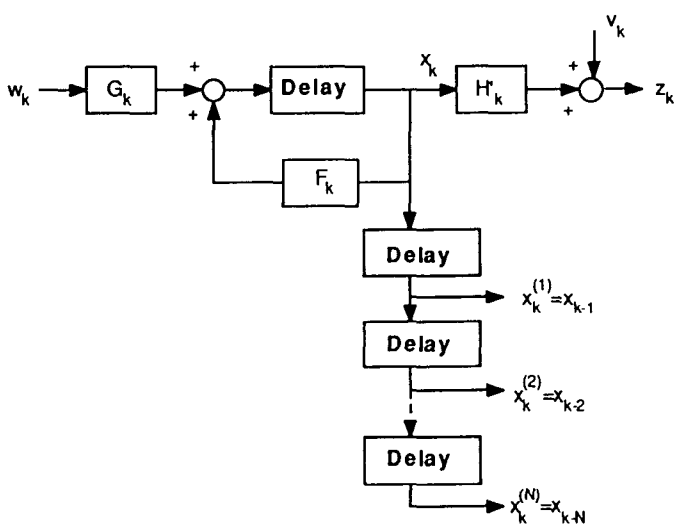


Fig. 7.2. Augmented signal model for fixed lag-smoothing

This augmented model, with state

$$[x_k' x_k^{(1)'} \dots x_k^{(N)'}]'$$

is of a standard type, and therefore a Kalman filter can be found for it. The state estimate of the Kalman filter will include as a subvector for each $j \in [1, N]$

$$E[x_k^{(j)} | z_l, l \leq k] = E[x_{k-j} | z_l, l \leq k] \quad (7.3)$$

This is nothing but a fixed-lag estimate with lag j . Again, it is easy to compute the filter for the augmented system, and therefore the fixed-lag smoother. It turns out that if not all lagged estimates between 1 and N are required, considerable simplification is possible, see [5]. This approach originally appeared in [16]. Approaches to the continuous-time problem can be found in [17].

Fixed-lag estimates will always offer a lower error covariance than filtered estimates of the same quantity, since more data is being used to generate the estimate. When the Kalman filter is exponentially stable, it turns out that all the improvement can in practice be obtained by taking the lag equal to 4 to 5 times the dominant time constraint of the Kalman filter.

Fixed interval smoothing. One way to solve the fixed interval smoothing problem becomes available when the Kalman filter is exponentially stable. Let Δ correspond to 4 to 5 times the dominant time constant t of the Kalman filter. Then

$$E[x(t) | z(s), s \in [t_0, T]] \approx E[x(t) | z(s), s \in [t_0, t + \Delta]] \quad (7.4)$$

so long as $t + \Delta \leq T$. If we define $H(t) = 0$ for $t > T$, so that measurements of $z(t)$ for $t > T$ contains no information about $x(t)$, then (7.4) holds for all $t \in [t_0, T]$.

Exact approaches are described in [18–20]. They generally involve running a Kalman filter forward in time and storing either the filtered estimates or the measurements. Then those stored estimates are run backwards to obtain the fixed-interval smoothed estimates. In discrete time, there holds

$$\hat{x}_{j-1/N} = \hat{x}_{j-1/j-1} + \Sigma_{j-1/j-1} F'_{j-1} \Sigma_{jj-1}^{-1} (\hat{x}_{j/N} - \hat{x}_{jj-1}) \quad (7.5)$$

where the interval in question is $[0, N]$. Using stored values of $\hat{x}_{jj}, \hat{x}_{jj-1}$, equation (7.5) is implemented backwards in time, thus $j = N, N-1, N-2, \dots$. Similarly,

$$\begin{aligned} \Sigma_{j-1/N} &= \Sigma_{j-1/j-1} + \Sigma_{j-1/j-1} F'_{j-1} \Sigma_{jj-1}^{-1} [\Sigma_{j/N} - \Sigma_{jj-1}] \\ &\quad \cdot \Sigma_{jj-1}^{-1} F_{j-1} \Sigma_{j-1/j-1} \end{aligned} \quad (7.6)$$

8 Miscellaneous Developments

Nonlinear Kalman filter. Following the success of the (linear) Kalman filter, it became natural to try to extend the ideas to nonlinear systems. In one thrust, see e.g. [21–22], the aim was to provide equations for the evolution of the

conditional probabilities $p(x(t)|z(s), s < t)$. Of course, describing the evolution of a function is much more complicated than describing the evolution of a finite dimensional quantity, as constituted for example by a mean and variance. Most practical applications of the nonlinear Kalman filter have therefore sought to follow the approach of describing the evolution of the conditional mean and error covariance. The catch is that approximations are inevitably involved, as linearization underpins virtually all the suggested schemes. In reference [23], the phase-locked-loop for demodulation of angle modulated signals is derived using such ideas: this is a real demonstration of their efficacy.

Computational issues. Attempts to implement the Kalman filter on practical problems soon showed that, where there was a choice of analytically equivalent procedures, numerical issues could play a major role in determining the utility of algorithms, especially for discrete-time problems, see e.g. [5] for a summary. Key ideas advanced included the following

- (a) The information filter; one works by updating the inverses of the covariance matrices, and modifies the filter equation.
- (b) Sequential processing; one processes a vector of output measurements one at a time, or in blocks.
- (c) Square root filtering; one works with the square root of the covariance matrices, rather than the matrices themselves.
- (d) Prevention of filter divergence; to avoid gain matrices becoming too small, artificial increase of the covariance of the process noise, $\{w_k\}$ or $w(\cdot)$, or exponential forgetting of old data, is introduced.
- (e) Chandrasekhar type algorithms; computational simplifications arise when the defining matrices of the signal model are time-invariant.

Controller design. For high order, linear multivariable systems, the design of controllers is a major task. One of the main tools is the linear-quadratic-Gaussian (LQG) design method, and a key component of such a design is the inclusion of a Kalman filter. The filter produces estimates $\hat{x}(t)$ of the plant state $x(t)$, and they are fed back to the plant input via an optimal linear state feedback law. For a recent treatment of LQG design, see [24].

Adaptive filtering. A critical assumption in applying Kalman filtering techniques is that the signal model is known. Since inaccurate model knowledge (which can certainly be the case in practice) may lead to poor estimation or prediction, there is motivation to “tune” or “select” filter parameters on line. The key idea of adaptive Kalman filtering is to monitor the variance of the innovations and to tune or select the Kalman gain, and perhaps other parameters, to reduce this. When the innovations covariance is white, and consequently of minimum variance, there is guaranteed optimality. A second approach is to work with signal models such that the original unknown model parameters are states, and apply Kalman filtering to estimate the parameters. These parameter estimates can then be used to tune the original Kalman filter. This coupled Kalman filter arrangement is discussed in [5].

9 Conclusions

Though the preceding sections have solely discussed theoretical issues, we should note the great practical importance of the Kalman filter. Applications in tracking and guidance abound, and as noted in the preceding section, the Kalman filter is a major constituent of many controller designs. In truth, it represents one of the major post-war advances of engineering science.

References

- [1] R.E. Kalman, "A new approach to linear filtering and prediction problems", *J Basic Eng, Trans ASME, Series D*, Vol 82, March 1960, pp 35–45
- [2] R.E. Kalman and R.S. Bucy, "New results in linear filtering and prediction theory", *J Basic Eng, Trans ASME, Series D*, Vol 83, March 1961, pp 95–108
- [3] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, MIT Press, Cambridge, Mass, 1949
- [4] H.W. Bode and C.E. Shannon, "A simplified derivation of linear least square smoothing and prediction theory", *Proc IRE*, Vol 38, April 1950, pp 417–425
- [5] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prentice Hall, Inc, Englewood Cliffs, NJ, 1979
- [6] V. Kucera, "The discrete Riccati equation of optimal control", *Kybernetika*, Vol 8, 1972, pp 430–447
- [7] R.E. Kalman, "Contributions to the theory of optimal control", *Bol Soc Matem Mex*, 1960, pp 102–119
- [8] B.D.O. Anderson and J.B. Moore, "Detectability and stabilizability of discrete-time linear systems", *SIAM J on Control & Optimization*, Vol 19, 1981, pp 20–32
- [9] P. Swerling, "A proposed stagewise differential correction procedure for satellite tracking and prediction", *J Astronaut. Sci.*, Vol 6, 1959, pp 46–59
- [10] P. Faurre, M. Clerget, and F. Germain, "*Opérateurs rationnels positifs*", Dunod, Paris, 1979
- [11] B.D.O. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis*, Prentice Hall, Inc, Englewood Cliffs, N.J., 1973
- [12] T. Kailath, "A view of three decades of linear filtering theory", *IEEE Trans Info Theory*, Vol IT-20, March 1974, pp 146–181
- [13] B.D.O. Anderson, J.B. Moore, and S.G. Loo, "Spectral factorization of time-varying covariance functions", *IEEE Trans Info Theory*, Vol IT-15, September 1969, pp 550–557
- [14] L.E. Zachrisson, "On optimal smoothing of continuous-time Kalman processes", *Information Sciences*, Vol 1, 1969, pp 143–172
- [15] W.W. Willman, "On the linear smoothing problem", *IEEE Trans Auto Control*, Vol AC-14, February 1969, pp 116–117
- [16] J.B. Moore, "Discrete-time fixed-lag smoothing algorithms", *Automatica*, Vol 19, March 1973, pp 163–174
- [17] S. Chirarattanon and B.D.O. Anderson, "Stable fixed-lag smoothing of continuous-time processes", *IEEE Trans Info Theory*, Vol IT-20, January 1974, pp 25–36
- [18] D.C. Fraser and J.E. Potter, "The optimum linear smoother as a combination of two optimum linear filters", *IEEE Trans Auto Control*, Vol AC-14, August 1969, pp 387–390
- [19] J.E. Wall, A.S. Willsky, and N.R. Sandell, "The fixed-interval smoother for continuous-time processes", *Proc 19th IEEE Conference on Decision and Control*, 1980, pp 385–389
- [20] H.E. Rauch, "Solutions to the linear smoothing problem", *IEEE Trans Auto Control*, Vol AC-8, October 1963, pp 371–372
- [21] E. Wong, *Stochastic Processes in Information and Dynamical Systems*, McGraw Hill Book Co., New York, 1971
- [22] R.J. Elliott, *Stochastic Calculus and Applications*, Springer Verlag, New York, 1982
- [23] D.L. Snyder, *The State-Variable Approach to Continuous Estimation*, MIT Press, Cambridge, Mass., 1969
- [24] B.D.O. Anderson and J.B. Moore, *Optimal Control: Linear-Quadratic Methods*, Prentice Hall, Englewood Cliffs, NJ, 1989