Faculteit der Economische Wetenschappen en Econometrie

Serie Research Memoranda

A Synopsis of the Smoothing Formulae Associated with the Kalman Filter

H.R. Merkus
D.S.G. Pollock
A.F. de Vos

Research Memorandum 1991-79
November 1991

vrije Universiteit *amsterdam*

A SYNOPSIS OF THE SMOOTHING FORMULAE ASSOCIATED WITH THE KALMAN FILTER

by H.R. Merkus, D.S.G. Pollock and A.F. de Vos
Vrije Universiteit, Amsterdam,
The Netherlands.

This paper provides straightforward derivations of a wide variety of smoothing formulae which are associated with the Kalman filter. The smoothing operations are of perennial interest in the fields of communications engineering and signal processing. Recently they have begun to interest statisticians. It is often asserted that it is tedious and difficult to derive the formulae. We show that this need not be so.

1. Introduction

The object of this paper is to provide a synopsis of the various algorithms which can be used for the retrospective enhancement of the state-vector estimates generated by the Kalman filter.

In its normal mode of operation, the Kalman filter generates an estimate of the current state of a system using information from the past and the present. Often an estimate can be improved greatly in the light of subsequent observations. In many real-time signal-processing applications, there is scope for a brief delay between the reception of a signal and the provision of the state estimate; and this delay can be used for gathering and processing additional observations. The classical fixed-lag smoothing algorithm is then the appropriate device for improving the estimate.

In recent years, statisticians have begun to use the Kalman filter in contexts where there is virtually no real-time constraint; and their attention has been concentrated upon the algorithms of fixed-interval smoothing which bring all of the information in a fixed sample to bear upon the estimation of a sequence of state vectors. The consequence of this renewed interest has been the discovery of several new algorithms as well as the rediscovery of older, partly-forgotten, algorithms.

Diverse approaches have been taken in the derivation of the various algorithms, and a welter of alternative notation has arisen. We fear that, nowadays, only the few veritable cognoscenti feel at ease in this specialised but highly profitable area of statistical theory; and we believe that the time is ripe for a synopsis of the results which aims to be both brief and accessible.

In pursuance of this aim, we feel bound to begin with a complete and self-contained derivation of the Kalman filter. With the help of the calculus of conditional expectations, this can be accomplished within a page. The same calculus is the ideal method for deriving the majority of the smoothing algorithms. The exceptions are the forward–backward algorithms, presented in the final section, for which a Bayesian approach is more appropriate.

2. Equations of the Kalman Filter

We shall present the basic equations of the Kalman filter in the briefest possible manner. The state-space model, which underlies the Kalman filter, consists of two equations

$$y_t = H_t\xi_t + \eta_t, \qquad \textit{Observation Equation} \tag{1}$$

$$\xi_t = \Phi_t\xi_{t-1} + \nu_t, \qquad \textit{Transition Equation} \tag{2}$$

where y_t is the observation on the system and ξ_t is the state vector. The observation error η_t and the state disturbance ν_t are mutually uncorrelated random vectors of zero mean with dispersion matrices

$$D(\eta_t) = \Omega_t \qquad \text{and} \qquad D(\nu_t) = \Psi_t. \tag{3}$$

It is assumed that the matrices H_t, Φ_t, Ω_t and Ψ_t are known for all $t = 1, \ldots, n$ and that an initial estimate x_0 is available for the state vector ξ_0 at time $t = 0$ together with a dispersion matrix $D(\xi_0) = P_0$. The empirical information available at time t is the set of observations $\mathcal{I}_t = \{y_1, \ldots, y_t\}$.

The Kalman-filter equations determine the state-vector estimates $x_{t|t-1} = E(\xi_t|\mathcal{I}_{t-1})$ and $x_t = E(\xi_t|\mathcal{I}_t)$ and their associated dispersion matrices $P_{t|t-1}$ and P_t. From $x_{t|t-1}$, the prediction $\hat{y}_{t|t-1} = H_t x_{t|t-1}$ is formed which has a dispersion matrix F_t. A summary of these equations is as follows:

$$x_{t|t-1} = \Phi_t x_{t-1}, \qquad \textit{State Prediction} \tag{4}$$

$$P_{t|t-1} = \Phi_t P_{t-1}\Phi_t' + \Psi_t, \qquad \textit{Prediction Dispersion} \tag{5}$$

$$e_t = y_t - H_t x_{t|t-1}, \qquad \textit{Prediction Error} \tag{6}$$

$$F_t = H_t P_{t|t-1} H_t' + \Omega_t, \qquad \textit{Error Dispersion} \tag{7}$$

$$K_t = P_{t|t-1} H_t' F_t^{-1}, \qquad \textit{Kalman Gain} \tag{8}$$

$$x_t = x_{t|t-1} + K_t e_t, \qquad \textit{State Estimate} \tag{9}$$

$$P_t = (I - K_t H_t) P_{t|t-1}. \qquad \textit{Estimate Dispersion} \tag{10}$$

We shall also define

$$M_t = \Phi_t K_{t-1} \quad \text{and} \tag{11}$$

$$\Lambda_t = \Phi_t(I - K_{t-1}H_{t-1}). \tag{12}$$

Alternative expressions are available for P_t and K_t:

$$P_t = (P_{t|t-1}^{-1} + H_t'\Omega_t^{-1}H_t)^{-1}, \tag{13}$$

$$K_t = P_t H_t'\Omega_t^{-1}. \tag{14}$$

By applying the well-known matrix inversion lemma to the expression on the RHS of (13), we obtain the original expression for P_t given under (10). To verify the identity $P_{t|t-1}H_t'F_t^{-1} = P_t H_t'\Omega_t^{-1}$ which equates (8) and (14), we write it as $P_t^{-1}P_{t|t-1}H_t' = H_t'\Omega_t^{-1}F_t$. The latter is readily confirmed using the expression for P_t from (13) and the expression for F_t from (7).

A variant of the Kalman filter known as the information filter is available which replaces the variables $x_{t|t-1}$ and x_t of (4) and (9) respectively by the variables $a_{t|t-1} = P_{t|t-1}^{-1}x_{t|t-1}$ and $a_t = P_t^{-1}x_t$, thereby transforming the equations into

$$a_{t|t-1} = P_{t|t-1}^{-1}\Phi_t P_{t-1}a_{t-1}, \qquad x_{t|t-1} = P_{t|t-1}a_{t|t-1}; \tag{15}$$

$$a_t = a_{t|t-1} + H_t'\Omega_t^{-1}y_t, \qquad x_t = P_t a_t. \tag{16}$$

The first of these comes immediately from (4). The second is established by writing the combination of equations (9) and (6) as

$$x_t = (I - K_t H_t)x_{t|t-1} + K_t y_t \tag{17}$$

or, equivalently,

$$a_t = P_t^{-1}(I - K_t H_t)P_{t|t-1}a_{t|t-1} + P_t^{-1}K_t y_t, \tag{18}$$

whence the result is obtained with the use of the equations (10) and (14). The inverse matrices $P_{t|t-1}^{-1}$ and P_t^{-1} are obtained with reference to (5) and (13).

Derivation of the Kalman Filter. The equations of the Kalman filter may be derived using the ordinary algebra of conditional expectations which indicates that, if x, y are jointly distributed variables which bear the linear relationship $E(y|x) = \alpha + B\{x - E(x)\}$, then

$$E(y|x) = E(y) + C(y,x)D^{-1}(x)\{x - E(x)\}, \tag{19}$$

$$D(y|x) = D(y) - C(y,x)D^{-1}(x)C(x,y), \tag{20}$$

$$E\{E(y|x)\} = E(y), \tag{21}$$

$$D\{E(y|x)\} = C(y,x)D^{-1}(x)C(x,y), \tag{22}$$

$$D(y) = D(y|x) + D\{E(y|x)\}, \tag{23}$$

$$C\{y - E(y|x), x\} = 0. \tag{24}$$

Of the equations listed under (4)—(10), those under (6) and (8) are merely definitions.

To demonstrate equation (4), we use (21) to show that

$$\begin{aligned} E(\xi_t|\mathcal{I}_{t-1}) &= E\{E(\xi_t|\xi_{t-1})|\mathcal{I}_{t-1}\} \\ &= E\{\Phi_t\xi_{t-1}|\mathcal{I}_{t-1}\} \\ &= \Phi_t x_{t-1}. \end{aligned} \tag{25}$$

We use (23) to demonstrate equation (5):

$$\begin{aligned} D(\xi_t|\mathcal{I}_{t-1}) &= D(\xi_t|\xi_{t-1}) + D\{E(\xi_t|\xi_{t-1})|\mathcal{I}_{t-1}\} \\ &= \Psi_t + D\{\Phi_t\xi_{t-1}|\mathcal{I}_{t-1}\} \\ &= \Psi_t + \Phi_t P_{t-1}\Phi_t'. \end{aligned} \tag{26}$$

To obtain equation (7), we substitute (1) into (6) to give $e_t = H_t(\xi_t - x_{t|t-1}) + \eta_t$. Then, in view of the statistical independence of the terms on the RHS, we have

$$\begin{aligned} D(e_t) &= D\{H_t(\xi_t - x_{t|t-1})\} + D(\eta_t) \\ &= H_t P_{t|t-1} H_t' + \Omega_t = D(y_t|\mathcal{I}_{t-1}). \end{aligned} \tag{27}$$

To demonstrate the updating equation (9), we begin by noting that

$$\begin{aligned} C(\xi_t, y_t|\mathcal{I}_{t-1}) &= E\{(\xi_t - x_{t|t-1})y_t'\} \\ &= E\{(\xi_t - x_{t|t-1})(H_t\xi_t + \eta_t)'\} \\ &= P_{t|t-1}H_t'. \end{aligned} \tag{28}$$

It follows from (19) that

$$\begin{aligned} E(\xi_t|\mathcal{I}_t) &= E(\xi_t|\mathcal{I}_{t-1}) + C(\xi_t, y_t|\mathcal{I}_{t-1})D^{-1}(y_t|\mathcal{I}_{t-1})\{y_t - E(y_t|\mathcal{I}_{t-1})\} \\ &= x_{t|t-1} + P_{t|t-1}H_t'F_t^{-1}e_t. \end{aligned} \tag{29}$$

The dispersion matrix under (10) for the updated estimate is obtained via equation (20):

$$\begin{aligned} D(\xi_t|\mathcal{I}_t) &= D(\xi_t|\mathcal{I}_{t-1}) - C(\xi_t, y_t|\mathcal{I}_{t-1})D^{-1}(y_t|\mathcal{I}_{t-1})C(y_t, \xi_t|\mathcal{I}_{t-1}) \\ &= P_{t|t-1} - P_{t|t-1}H_t'F_t^{-1}H_tP_{t|t-1}. \end{aligned} \tag{30}$$

Innovations and the Information Set. The remaining task of this section is to establish that the information of $\mathcal{I}_t = \{y_1, \ldots, y_t\}$ is also conveyed by the prediction errors or innovations $\{e_1, \ldots, e_t\}$ and that the latter are mutually uncorrelated random variables.

First we demonstrate that each error e_t is a linear function of $y_1, \ldots, y_t$. From equations (9), (6) and (4), or, equally, from equations (17) and (4), we obtain the equation $x_{t|t-1} = \Lambda_t x_{t-1|t-2} + M_t y_{t-1}$. Repeated backsubstitution gives

$$x_{t|t-1} = \sum_{j=1}^{t-1} \Lambda_{t,j+2} M_{j+1} y_j + \Lambda_{t,2} x_{1|0}, \tag{31}$$

where $\Lambda_{t,j+2} = \Lambda_t \cdots \Lambda_{j+2}$ is a product of matrices which specialises to $\Lambda_{t,t} = \Lambda_t$ and to $\Lambda_{t,t+1} = I$. It follows that

$$\begin{aligned} e_t &= y_t - H_t x_{t|t-1} \\ &= y_t - H_t \sum_{j=1}^{t-1} \Lambda_{t,j+2} M_{j+1} y_j - H_t \Lambda_{t,2} x_{1|0}. \end{aligned} \tag{32}$$

Next, we demonstrate that each y_t is a linear function of $e_1, \ldots, e_t$. By backsubstitution in the equation $x_{t|t-1} = \Phi_t x_{t-1|t-2} + M_t e_{t-1}$ obtained from (4) and (9), we get

$$x_{t|t-1} = \sum_{j=1}^{t-1} \Phi_{t,j+2} M_{j+1} e_j + \Phi_{t,2} x_{1|0}, \tag{33}$$

wherein $\Phi_{t,j+2} = \Phi_t \cdots \Phi_{j+2}$ is a product of matrices which specialises to $\Phi_{t,t} = \Phi_t$ and to $\Phi_{t,t+1} = I$. It follows that

$$\begin{aligned} y_t &= e_t + H_t x_{t|t-1} \\ &= e_t + H_t \sum_{j=1}^{t-1} \Phi_{t,j+2} M_{j+1} e_j + H_t \Phi_{t,2} x_{1|0}. \end{aligned} \tag{34}$$

Given that there is a one-to-one linear relationship between the observations and the prediction errors, it follows that we can represent the information set in terms of either. Thus we have $\mathcal{I}_{t-1} = \{e_1, \ldots, e_{t-1}\}$; and, given that $e_t = y_t - E(y_t|\mathcal{I}_{t-1})$, it follows from (24) that e_t is uncorrelated with the preceding errors $e_1, \ldots, e_{t-1}$. The result indicates that the prediction errors are mutually uncorrelated.

3. The Smoothing Operations

The object of smoothing is to improve our estimate x_t of the state vector ξ_t using information which has arisen subsequently. For the succeeding observations $\{y_{t+1}, y_{t+2}, \ldots\}$ are bound to convey information about the state of the system which can supplement the information $\mathcal{I}_t = \{y_1, \ldots, y_t\}$ which was available at time t.

There are several ways in which we might effect a process of smoothing. In the first place, there is fixed-point smoothing. This is used whenever the object is to enhance the estimate of a single state variable ξ_t repeatedly, using successive observations. The resulting sequence of estimates is described by

$$\{x_{t|n} = E(\xi_t|\mathcal{I}_n); n = t+1, t+2, \ldots\}. \qquad \textit{Fixed-Point Smoothing} \qquad (35)$$

The second mode of smoothing is fixed-lag smoothing. In this case, enhanced estimates of successive state vectors are generated with a fixed lag of, say, t periods:

$$\{x_{n-t|n} = E(\xi_{n-t}|\mathcal{I}_n); n = t+1, t+2, \ldots\}. \qquad \textit{Fixed-Lag Smoothing} \qquad (36)$$

Finally, there is fixed-interval smoothing. This is a matter of revising each of the state estimates for a period running from $t = 1$ to $t = n$ once the full set of observation in $\mathcal{I}_n = \{y_1, \ldots, y_n\}$ has become available. The sequence of revised estimates is

$$\{x_{n-t|n} = E(\xi_{n-t}|\mathcal{I}_n); t = 1, 2, \ldots, n\}. \qquad \textit{Fixed-Interval Smoothing} \qquad (37)$$

Here, instead of $x_{t|n}$, we have taken $x_{n-t|n}$ as the generic element, which gives the sequence in reverse order. This is to reflect the fact that, with most algorithms, the smoothed estimates are generated by running backwards through the initial set of estimates.

There is also a variant of fixed-interval smoothing which we shall describe as *Intermittent Smoothing*. For, it transpires that, if the fixed-interval smoothing operation is repeated periodically to take account of new data, then some use can be made of the products of the previous smoothing operation.

For each mode of smoothing, there is an appropriate recursive formula. We shall derive these formulae, in the first instance, from a general expression for the expectation of the state vector ξ_t conditional upon the information contained in the set of innovations $\{e_1, \ldots, e_n\}$ which we have shown to be identical to the information contained in the observations $\{y_1, \ldots, y_n\}$.

4. Conditional Expectations and Dispersions of the State Vector

Given that the sequence $e_1, \ldots, e_n$ of Kalman-filter innovations are mutually independent vectors with zero expectations, it follows from (19) that

$$E(\xi_t|\mathcal{I}_n) = E(\xi_t) + \sum_{j=1}^{n} C(\xi_t, e_j)D^{-1}(e_j)e_j. \qquad (38)$$

However, the sum is recursive in the sense that

$$E(\xi_t|\mathcal{I}_j) = E(\xi_t|\mathcal{I}_{j-1}) + C(\xi_t, e_j)D^{-1}(e_j)e_j; \qquad (39)$$

and so we have

$$E(\xi_t|\mathcal{I}_n) = E(\xi_t|\mathcal{I}_m) + \sum_{j=m+1}^{n} C(\xi_t, e_j)D^{-1}(e_j)e_j. \tag{40}$$

In a similar way, we see from equation (20) that the dispersion matrix satisfies

$$D(\xi_t|\mathcal{I}_n) = D(\xi_t|\mathcal{I}_m) - \sum_{j=m+1}^{n} C(\xi_t, e_j)D^{-1}(e_j)C(e_j, \xi_t). \tag{41}$$

The task of evaluating the expressions under (40) and (41) is to find the generic covariance $C(\xi_t, e_k)$. For this purpose, we must develop a recursive formula which represents e_k in terms of $\xi_t - E(\xi_t|\mathcal{I}_{t-1})$ and in terms of the state disturbances and observation errors which occur from time t.

Consider the expression for the innovation

$$\begin{aligned} e_k &= y_k - H_k x_{k|k-1} \\ &= H_k(\xi_k - x_{k|k-1}) + \eta_k. \end{aligned} \tag{42}$$

Here the term $\xi_k - x_{k|k-1}$ follows a recursion which is indicated by the equation

$$\xi_k - x_{k|k-1} = \Lambda_k(\xi_{k-1} - x_{k-1|k-2}) + (\nu_k - M_k\eta_{k-1}). \tag{43}$$

The latter comes from subtracting from equation (2) the equation $x_{t|t-1} = \Lambda_t x_{t-1|t-2} + M_t(H_{t-1}\xi_{t-1} + \eta_{t-1})$, obtained by substituting (1) into (17) and putting the result, lagged one period, into (4). By running the recursion from time k back to time t, we may deduce that

$$\xi_k - x_{k|k-1} = \Lambda_{k,t+1}(\xi_t - x_{t|t-1}) + \sum_{j=t}^{k-1} \Lambda_{k,j+2}(\nu_{j+1} - M_{j+1}\eta_j), \tag{44}$$

wherein $\Lambda_{k,k+1} = I$ and $\Lambda_{k,k} = \Lambda_k$. It follows from (42) and (44) that, when $k \geq t$,

$$\begin{aligned} C(\xi_t, e_k) &= E\{\xi_t(\xi_t - x_{t|t-1})\Lambda'_{k,t+1}H'_k\} \\ &= P_{t|t-1}\Lambda'_{k,t+1}H'_k. \end{aligned} \tag{45}$$

Using the identity $\Phi_{t+1}P_t = \Lambda_{t+1}P_{t|t-1}$ which comes via (10), we get for $k > t$

$$C(\xi_t, e_k) = P_t\Phi'_{t+1}\Lambda'_{k,t+2}H'_k. \tag{46}$$

Next we note that

$$C(\xi_{t+1}, e_k) = P_{t+1|t}\Lambda'_{k,t+2}H'_k. \tag{47}$$

It follows, from comparing (46) and (47), that

$$C(\xi_t, e_k) = P_t \Phi'_{t+1} P^{-1}_{t+1|t} C(\xi_{t+1}, e_k). \tag{48}$$

If we substitute the expression under (45) into the formula of (40) where $m \geq t-1$, and if we set $D^{-1}(e_j) = F_j^{-1}$, then we get

$$\begin{aligned} E(\xi_t|\mathcal{I}_n) &= E(\xi_t|\mathcal{I}_m) + \sum_{j=m+1}^{n} C(\xi_t, e_j) D^{-1}(e_j) e_j \\ &= E(\xi_t|\mathcal{I}_m) + \sum_{j=m+1}^{n} P_{t|t-1} \Lambda'_{j,t+1} H'_j F_j^{-1} e_j \\ &= E(\xi_t|\mathcal{I}_m) + P_{t|t-1} \Lambda'_{m+1,t+1} \sum_{j=m+1}^{n} \Lambda'_{j,m+2} H'_j F_j^{-1} e_j. \end{aligned} \tag{49}$$

An expression for the dispersion matrix is found in a similar way:

$$\begin{aligned} D(\xi_t|\mathcal{I}_n) = D(\xi_t|\mathcal{I}_m) \\ - P_{t|t-1} \Lambda'_{m+1,t+1} \Big\{ \sum_{j=m+1}^{n} \Lambda'_{j,m+2} H'_j F_j^{-1} H_j \Lambda_{j,m+2} \Big\} \Lambda_{m+1,t+1} P_{t|t-1}. \end{aligned} \tag{50}$$

Notice that the sums in the two final expressions may be accumulated using recursions running backwards in time of the form

$$\begin{aligned} q_t &= \sum_{j=t}^{n} \Lambda'_{j,t+1} H'_j F_j^{-1} e_j \\ &= H'_t F_t^{-1} e_t + \Lambda'_{t+1} q_{t+1} \end{aligned} \tag{51}$$

and

$$\begin{aligned} Q_t &= \sum_{j=t}^{n} \Lambda'_{j,t+1} H'_j F_j^{-1} H_j \Lambda_{j,t+1} \\ &= H'_t F_t^{-1} H_t + \Lambda'_{t+1} Q_{t+1} \Lambda_{t+1}. \end{aligned} \tag{52}$$

These recursions are initiated with $q_n = H'_n F_n^{-1} e_n$ and $Q_n = H'_n F_n^{-1} H_n$.

5. The Classical Smoothing Algorithms

An account of the classical smoothing algorithms is to be found in the book by Anderson and Moore [1] which has become a standard reference for the Kalman filter.

Anderson and Moore have adopted a method for deriving the filtering equations which depends upon an augmented state-transition equation wherein the enlarged state vector contains a sequence of the state vectors from the original transition equation. This approach is common to several authors including Willman [13] who deals with fixed-point smoothing, Premier and Vacroux [11] who treat fixed-lag smoothing and Farooq and Mahalanabis [8] who treat fixed-interval smoothing. We believe that an approach via the calculus of conditional expectations is more direct.

The Fixed-Point Smoother. Of the classical smoothing algorithms, the fixed-point smoothing equations are the easiest to derive. The task is as follows: given $x_{t|n} = E(\xi_t|e_1,\ldots,e_n)$, we must find an expression for $x_{t|n+1} = E(\xi_t|e_1,\ldots,e_{n+1})$ with $n \geq t$. That is to say, we must enhance the estimate of ξ_t by incorporating the extra information which is afforded by the new innovation e_{n+1}. The formula is simply

$$E(\xi_t|\mathcal{I}_{n+1}) = E(\xi_t|\mathcal{I}_n) + C(\xi_t, e_{n+1})D^{-1}(e_{n+1})e_{n+1}. \tag{53}$$

Now, (45) gives

$$\begin{aligned} C(\xi_t, e_n) &= P_{t|t-1}\Lambda'_{n,t+1}H'_n \\ &= L_n H'_n \end{aligned} \tag{54}$$

and

$$\begin{aligned} C(\xi_t, e_{n+1}) &= P_{t|t-1}\Lambda'_{n+1,t+1}H'_{n+1} \\ &= L_n\Lambda'_{n+1}H'_{n+1}. \end{aligned} \tag{55}$$

Therefore we may write the fixed-point algorithm as

$$\begin{aligned} &E(\xi_t|\mathcal{I}_{n+1}) = E(\xi_t|\mathcal{I}_n) + L_{n+1}H'_{n+1}F^{-1}_{n+1}e_{n+1} \\ &\text{where} \qquad L_{n+1} = L_n\Lambda'_{n+1} \qquad \text{and} \qquad L_t = P_{t|t-1}. \end{aligned} \tag{56}$$

The accompanying dispersion matrix can be calculated from

$$D(\xi_t|\mathcal{I}_{n+1}) = D(\xi_t|\mathcal{I}_n) - L_{n+1}H'_{n+1}F^{-1}_{n+1}H_{n+1}L'_{n+1}. \tag{57}$$

The fixed-point smoother is initiated with values for $E(\xi_t|\mathcal{I}_t)$, $D(\xi_t|\mathcal{I}_t)$ and $L_t = P_{t|t-1}$, which are provided by the Kalman filter. From these initial quantities, a sequence of enhanced estimates of ξ_t is calculated recursively using subsequent observations. The values of e_{n+1}, F_{n+1} and K_n, needed in computing (56) and (57), are also provided by the Kalman filter, which runs concurrently with the smoother.

The Fixed-Interval Smoother. The next version of the smoothing equation to be derived is the fixed-interval form. Consider using the identity of (48) to rewrite equation (40), with m set to t, as

$$E(\xi_t|\mathcal{I}_n) = E(\xi_t|\mathcal{I}_t) + P_t\Phi'_{t+1}P^{-1}_{t+1|t}\sum_{j=t+1}^{n} C(\xi_{t+1}, e_j)D^{-1}(e_j)e_j. \tag{58}$$

Now

$$E(\xi_{t+1}|\mathcal{I}_n) = E(\xi_{t+1}|\mathcal{I}_t) + \sum_{j=t+1}^{n} C(\xi_{t+1}, e_j)D^{-1}(e_j)e_j; \tag{59}$$

so it follows that equation (58) can be rewritten in turn as

$$E(\xi_t|\mathcal{I}_n) = E(\xi_t|\mathcal{I}_t) + P_t\Phi'_{t+1}P^{-1}_{t+1|t}\Big\{E(\xi_{t+1}|\mathcal{I}_n) - E(\xi_{t+1}|\mathcal{I}_t)\Big\}. \tag{60}$$

This is the formula for the fixed-interval smoother.

A similar strategy is adopted in the derivation of the dispersion of the smoothed estimate. According to (41), we have

$$D(\xi_t|\mathcal{I}_n) = D(\xi_t|\mathcal{I}_t) - \sum_{j=t+1}^{n} C(\xi_t, e_j)D^{-1}(e_j)C(e_j, \xi_t) \tag{61}$$

and

$$D(\xi_{t+1}|\mathcal{I}_n) = D(\xi_{t+1}|\mathcal{I}_t) - \sum_{j=t+1}^{n} C(\xi_{t+1}, e_j)D^{-1}(e_j)C(e_j, \xi_{t+1}). \tag{62}$$

Using the identity of (48) in (61) and taking the result from (62) enables us to write

$$P_{t|n} = P_t - P_t\Phi'_{t+1}P^{-1}_{t+1|t}\Big\{P_{t+1|t} - P_{t+1|n}\Big\}P^{-1}_{t+1|t}\Phi_{t+1}P_t. \tag{63}$$

An Interpretation. Consider $E(\xi_t|\mathcal{I}_n)$, and let us represent the information set, at first, by

$$\mathcal{I}_n = \Big\{\mathcal{I}_t, h_{t+1}, e_{t+2}, \ldots, e_n\Big\} \quad \text{where} \quad h_{t+1} = \xi_{t+1} - E(\xi_{t+1}|\mathcal{I}_t). \tag{64}$$

We may begin by finding

$$E(\xi_t|\mathcal{I}_t, h_{t+1}) = E(\xi_t|\mathcal{I}_t) + C(\xi_t, h_{t+1}|\mathcal{I}_t)D^{-1}(h_{t+1}|\mathcal{I}_t)h_{t+1}. \tag{65}$$

Here we have

$$\begin{aligned} &C(\xi_t, h_{t+1}|\mathcal{I}_t) = E\Big\{\xi_t(\xi_t - x_t)'\Phi'_{t+1} + \xi_t\nu'_t|\mathcal{I}_t\Big\} = P_t\Phi'_{t+1} \quad \text{and} \\ &D(h_{t+1}|\mathcal{I}_t) = P_{t+1|t}. \end{aligned} \tag{66}$$

It follows that

$$E(\xi_t|\mathcal{I}_t, h_{t+1}) = E(\xi_t|\mathcal{I}_t) + P_t\Phi'_{t+1}P^{-1}_{t+1|t}\Big\{\xi_{t+1} - E(\xi_{t+1}|\mathcal{I}_t)\Big\}. \tag{67}$$

Of course, the value of ξ_{t+1} in the RHS of this equation is not observable. However, if we take the expectation of the equation conditional upon all of the information in the set $\mathcal{I}_n = \{e_1, \ldots, e_n\}$, then ξ_{t+1} is replaced by $E(\xi_{t+1}|\mathcal{I}_n)$ and we get the formula under (60). This interpretation was published by Ansley and Kohn [2]. It highlights the notion that the information which is used in enhancing the estimate of ξ_t is contained entirely within the smoothed estimate of ξ_{t+1}.

The Intermittent Smoother. Consider the case where smoothing is intermittent with m sample points accumulating between successive smoothing operations. Then it is possible to use the estimates arising from the previous smoothing operation.

Imagine that the operation is performed when $n = jm$ points are available. Then, for $t > (j-1)m$, the smoothed estimate of the state vector ξ_t is given by the ordinary fixed-interval smoothing formula found under (60). For $t \leq (j-1)m$, the appropriate formula is

$$E(\xi_t|\mathcal{I}_n) = E(\xi_t|\mathcal{I}_{(j-1)m}) + P_t\Phi'_{t+1}P^{-1}_{t+1|t}\Big\{E(\xi_{t+1}|\mathcal{I}_n) - E(\xi_{t+1}|\mathcal{I}_{(j-1)m})\Big\}. \tag{68}$$

Here $E(\xi_t|\mathcal{I}_{(j-1)m})$ is being used in place of $E(\xi_t|\mathcal{I}_t)$. The advantage of the algorithm is that it does not require the values of unsmoothed estimates to be held in memory when smoothed estimates are available.

A limiting case of the intermittent smoothing algorithm arises when the smoothing operation is performed each time a new observation is registered. Then the formula becomes

$$E(\xi_t|\mathcal{I}_n) = E(\xi_t|\mathcal{I}_{n-1}) + P_t\Phi'_{t+1}P^{-1}_{t+1|t}\Big\{E(\xi_{t+1}|\mathcal{I}_n) - E(\xi_{t+1}|\mathcal{I}_{n-1})\Big\}. \tag{69}$$

The formula is attributable to Chow [4] who provided a somewhat lengthy derivation. Chow proposed this algorithm for the purpose of ordinary fixed-interval smoothing, for which it is clearly inefficient.

The Fixed-Lag Smoother. The task is to move from the smoothed estimate of ξ_{n-t} made at time n to the estimate of ξ_{n+1-t} once the new information in the prediction error e_{n+1} has become available. Equation (39) indicates that

$$E(\xi_{n+1-t}|\mathcal{I}_{n+1}) = E(\xi_{n+1-t}|\mathcal{I}_n) + C(\xi_{n+1-t}, e_{n+1})D^{-1}(e_{n+1})e_{n+1}, \tag{70}$$

which is the formula for the smoothed estimate, whilst the corresponding formula for the dispersion matrix is

$$D(\xi_{n+1-t}|\mathcal{I}_{n+1}) = D(\xi_{n+1-t}|\mathcal{I}_n) - C(\xi_{n+1-t}, e_{n+1})D^{-1}(e_{n+1})C(e_{n+1}, \xi_{n+1-t}). \tag{71}$$

To evaluate (70), we must first find the value of $E(\xi_{n+1-t}|\mathcal{I}_n)$ from the value of $E(\xi_{n-t}|\mathcal{I}_n)$. On setting $t = k$ in the fixed-interval formula under (60), and rearranging the result, we get

$$E(\xi_{k+1}|\mathcal{I}_n) = E(\xi_{k+1}|\mathcal{I}_k) + P_{k+1|k}\Phi'^{-1}_{k+1}P_k^{-1}\Big\{E(\xi_k|\mathcal{I}_n) - E(\xi_k|\mathcal{I}_k)\Big\}. \quad (72)$$

To obtain the desired result, we simply set $k = n - t$, which gives

$$\begin{aligned} E(\xi_{n+1-t}|\mathcal{I}_n) &= E(\xi_{n+1-t}|\mathcal{I}_{n-t}) \\ &+ P_{n+1-t|n-t}\Phi'^{-1}_{n+1-t}P^{-1}_{n-t}\Big\{E(\xi_{n-t}|\mathcal{I}_n) - E(\xi_{n-t}|\mathcal{I}_{n-t})\Big\}. \end{aligned} \quad (73)$$

The formula for the smoothed estimate also comprises

$$C(\xi_{n+1-t}, e_{n+1}) = P_{n+1-t|n-t}\Lambda'_{n+1,n+2-t}H'_{n+1}. \quad (74)$$

If Λ_{n+1-t} is nonsingular, then $\Lambda_{n+1,n+2-t} = \Lambda_{n+1}\{\Lambda_{n,n+1-t}\}\Lambda^{-1}_{n+1-t}$; and thus we may profit from the calculations entailed in finding the previous smoothed estimate which will have generated the matrix product in the parentheses.

In evaluating the formula (71) for the dispersion of the smoothed estimates, we may use the following expression for $D(\xi_{n+1-t}|\mathcal{I}_n) = P_{n+1-t|n}$:

$$\begin{aligned} P_{n+1-t|n} &= P_{n+1-t|n-t} \\ &- P_{n+1-t|n-t}\Phi'^{-1}_{n+1-t}P^{-1}_{n-t}(P_{n-t} - P_{n-t|n})P^{-1}_{n-t}\Phi^{-1}_{n+1-t}P_{n+1-t|n-t}. \end{aligned} \quad (75)$$

This is demonstrated is the same manner as equation (73).

A process of fixed-lag smoothing, with a lag length of t, is initiated with a value for $E(\xi_1|\mathcal{I}_{t+1})$. The latter is provided by running the fixed-point smoothing algorithm for t periods. After time $t+1$, when the $(n+1)$th observation becomes available, $E(\xi_{n+1-t}|\mathcal{I}_n)$ is calculated from $E(\xi_{n-t}|\mathcal{I}_n)$ via equation (73). For this purpose the values of $x_{n+1-t|n-t}$, x_{n-t}, $P_{n+1-t|n-t}$ and P_{n-t} must be available. These are generated by the Kalman filter in the process of calculating e_{n-t}, and they are held in memory for t periods. The next smoothed estimate $E(\xi_{n+1-t}|\mathcal{I}_{n+1})$ is calculated from equation (70), for which the values of e_{n+1}, F_{n+1} and K_n are required. These are also provided by the Kalman filter which runs concurrently.

6. Variants of the Classical Algorithms

The attention which statisticians have paid to the smoothing problem recently has been focussed upon fixed-interval smoothing. This mode of smoothing is, perhaps, of less interest to communications engineers than the other modes; which may account for the fact that the statisticians have found scope for improving the algorithms.

Avoiding an Inversion. There are some modified versions of the classical fixed-interval smoothing algorithm which avoid the inversion of the matrix $P_{t|t-1}$. In fact, the basis for these has been provided already in section 4. Thus, by replacing the sums in equations (49) and (50) by q_{m+1} and Q_{m+1}, which are the products of the recursions under (51) and (52), we get

$$E(\xi_t|\mathcal{I}_n) = E(\xi_t|\mathcal{I}_m) + P_{t|t-1}\Lambda'_{m+1,t+1}q_{m+1}, \tag{76}$$

$$D(\xi_t|\mathcal{I}_n) = D(\xi_t|\mathcal{I}_m) - P_{t|t-1}\Lambda'_{m+1,t+1}Q_{m+1}\Lambda_{m+1,t+1}P_{t|t-1}. \tag{77}$$

These expressions are valid for $m \geq t-1$.

Setting $m = t-1$ in (76) and (77) gives a useful alternative to the classical algorithm for fixed-interval smoothing:

$$x_{t|n} = x_{t|t-1} + P_{t|t-1}q_t, \tag{78}$$

$$P_{t|n} = P_{t|t-1} - P_{t|t-1}Q_tP_{t|t-1}. \tag{79}$$

We can see that, in moving from q_{t+1} to q_t via equation (51), which is the first step towards finding the next smoothed estimate $x_{t-1|n}$, there is no inversion of $P_{t|t-1}$. The equations (78) and (79) have been derived by De Jong [6].

The connection with the classical smoothing algorithm is easily established. From (78), we get $q_{t+1} = P^{-1}_{t+1|t}(x_{t+1|n} - x_{t+1|t})$. By setting $m = t$ in (76) and substituting for q_{t+1} we get

$$\begin{aligned} x_{t|n} &= x_t + P_{t|t-1}\Lambda'_{t+1}P^{-1}_{t+1|t}(x_{t+1|n} - x_{t+1|t}) \\ &= x_t + P_t\Phi'_{t+1}P^{-1}_{t+1|t}(x_{t+1|n} - x_{t+1|t}), \end{aligned} \tag{80}$$

where the final equality follows from the identity $\Phi_{t+1}P_t = \Lambda_{t+1}P_{t|t-1}$ already used in (46). Equation (80) is a repetition of equation (60) which belongs to the classical algorithm.

Equation (63), which also belongs to the classical algorithm, is obtained by performing similar manipulations with equations (77) and (79).

Smoothing via State Disturbances. Given an initial value for the state vector, a knowledge of the sequence of the state-transition matrices and of the state disturbances in subsequent periods will enable one to infer the values of subsequent state vectors. Therefore the estimation of a sequence of state vectors may be construed as a matter of estimating the state disturbances. The information which is relevant to the estimation of the disturbance ν_t is contained in the prediction errors from time t onwards. Thus

$$E(\nu_t|\mathcal{I}_n) = \sum_{j=t}^{n} C(\nu_t, e_j)D^{-1}(e_j)e_j. \tag{81}$$

Here, for $j \geq t$, the generic covariance is given by

$$\begin{aligned} C(\nu_t, e_j) &= E\{\nu_t \nu_t' \Lambda_{j,t+1}' H_j'\} \\ &= \Psi_t \Lambda_{j,t+1}' H_j', \end{aligned} \tag{82}$$

which follows from the expression for e_t which results from substituting (44) in (42). Putting (82) into (81) and setting $D^{-1}(e_j) = F_j^{-1}$ gives

$$\begin{aligned} E(\nu_t|\mathcal{I}_n) &= \Psi_t \sum_{j=t}^{n} \Lambda_{j,t+1}' H_j' F_j^{-1} e_j \\ &= \Psi_t q_t, \end{aligned} \tag{83}$$

where q_t is a sum which may be accumulated using the recursion under (51).

By taking the expectation of the transition equation conditional upon all of the information in the fixed sample, we obtain the recursive equation which generates the smoothed estimates of the state vectors:

$$\begin{aligned} x_{t|n} &= \Phi_t x_{t-1|n} + E(\nu_t|\mathcal{I}_n) \\ &= \Phi_t x_{t-1|n} + \Psi_t q_t. \end{aligned} \tag{84}$$

The initial value is $x_{0|n} = x_0 + P_0 \Phi_1' q_1$. This is obtained by setting $t = 0$ in the equation $x_{t|n} = x_t + P_t \Phi_{t+1}' q_{t+1}$ which comes from (80).

Equation (84) has been presented recently in a paper by Koopman [9]. A similar approach has been pursued by Mayne [10].

With some effort, a connection can be found between equation (84) and equation (78) which is its counterpart in the previous algorithm. From (4) and (9), we get $x_{t|t-1} = \Phi_t(x_{t-1|t-2} + K_{t-1}e_{t-1})$. From (5) and (10), we get $P_{t|t-1} = \Phi_t P_{t-1|t-2}(I - K_{t-1}H_{t-1})'\Phi_t' + \Psi_t$. Putting these into (78) gives

$$x_{t|n} = \Phi_t x_{t-1|t-2} + \Psi_t q_t + \Phi_t(K_{t-1}e_{t-1} + P_{t-1|t-2}\Lambda_t' q_t). \tag{85}$$

Equation (78) lagged one period also gives an expression for $x_{t-1|t-2}$ in terms of $x_{t-1|n}$:

$$x_{t-1|t-2} = x_{t-1|n} - P_{t-1|t-2} q_{t-1}. \tag{86}$$

Using the identity $q_{t-1} = H_{t-1}' F_{t-1}^{-1} e_{t-1} + \Lambda_t' q_t$ and the latter equation, we can rewrite (85) as

$$\begin{aligned} x_{t|n} &= \Phi_t x_{t-1|n} + \Psi_t q_t - \Phi_t P_{t-1|t-2}(H_{t-1}' F_{t-1}^{-1} e_{t-1} + \Lambda_t' q_t) \\ &\quad + \Phi_t(K_{t-1}e_{t-1} + P_{t-1|t-2}\Lambda_t' q_t) \\ &= \Phi_t x_{t-1|n} + \Psi_t q_t, \end{aligned} \tag{87}$$

where the final equality follows from equation (8). This is (84) again.

An alternative algorithm exists which also uses estimates of the state disturbances. In contrast to the previous algorithm, it runs backwards in time rather than forwards. The basic equation is

$$x_{t-1|n} = \Phi_t^{-1} x_{t|n} - \Phi_t^{-1} \Psi_t q_t, \tag{88}$$

which comes directly from (84). The value of q_t is obtained via equation (51). However, because we have a backward recursion in (88), an alternative recursion for q_t is available, which reduces the number of elements which must be held in memory. A reformulation of equation (51) gives

$$\begin{aligned} q_t &= H_t' F_t^{-1} e_t + \Lambda_{t+1}' q_{t+1} \\ &= H_t' F_t^{-1} e_t + (I - K_t H_t)' \Phi_{t+1}' q_{t+1} \\ &= H_t' s_t + \Phi_{t+1}' q_{t+1}, \end{aligned} \tag{89}$$

where s_t is defined as

$$s_t = F_t^{-1} e_t - K_t' \Phi_{t+1}' q_{t+1}. \tag{90}$$

Now, consider the smoothed estimates of the observation errors. Because η_t is independent of $y_1, \ldots, y_{t-1}$, these are given by

$$E(\eta_t | \mathcal{I}_n) = \sum_{j=t}^{n} C(\eta_t, e_j) D^{-1}(e_j) e_j. \tag{91}$$

The covariances follow once more from equations (42) and (44). For $j > t$, we get

$$C(\eta_t, e_j) = -\Omega_t M_{t+1}' \Lambda_{j,t+2}' H_j', \tag{92}$$

whereas, for $j = t$, we have $C(\eta_t, e_t) = \Omega_t$. Substituting these in (91) gives

$$\begin{aligned} E(\eta_t | \mathcal{I}_n) &= \Omega_t \Big\{ F_t^{-1} e_t - M_{t+1}' \sum_{j=t+1}^{n} \Lambda_{j,t+2}' H_j' F_j^{-1} e_j \Big\} \\ &= \Omega_t \{ F_t^{-1} e_t - K_t' \Phi_{t+1}' q_{t+1} \} \\ &= \Omega_t s_t; \end{aligned} \tag{93}$$

from which

$$s_t = \Omega_t^{-1} E(\eta_t | \mathcal{I}_n) = \Omega_t^{-1} \{ y_t - H_t x_{t|n} \}, \tag{94}$$

where the final equality is justified by the observation equation (1). Notice that, in order to calculate s_t from this expression, we need $x_{t|n}$, which is available only because we are using a backward smoothing algorithm. Thus s_t is calculated from (94) using the previous smoothed estimate. Then it is substituted in (89) to obtain q_t. Finally, the smoothed estimate of the state vector is obtained from equation (88). Whittle [12] has derived this algorithm by maximising a log-likelihood function.

Comparing the Fixed-Interval Smoothers. By its avoidance of a matrix inversion, the algorithm of equations (78) and (79), which we may call De Jong's [6] algorithm, is more efficient than the classical fixed-interval smoother; and we can advise that it should be used in preference. Our attention must be focussed, therefore, on a comparison of the latter algorithm with the two algorithms which are based upon estimates of the state disturbances.

De Jong's algorithm requires the values of $x_{t|t-1}$, $P_{t|t-1}$, e_t, F_t^{-1} and K_t to be computed in a forward pass of the Kalman filter. The backward recursions for q_t and Q_t, which employ equations (51) and (52), then can be executed; and in each step the used values of e_t, F_t^{-1} and K_t can be deleted from memory. In combining the results by means of equations (76) and (77), we are able to generate both the smoothed estimate and its dispersion matrix.

The first of the state-disturbances algorithms, which is Koopman's [9] algorithm, uses successively a forward, a backward and a forward run to obtain the smoothed estimates. First, e_t, F_t^{-1} and K_t are calculated for all t via the Kalman filter. Then a backward recursion is used to generate the values of q_t which are committed to memory. Finally, the smoothed estimates of the state vector are calculated using the forward recursion of (84).

Since Koopman's algorithm and De Jong's algorithm both entail the calculation of q_t, their comparison amounts to the comparison of the equations $x_{t|n} = \Phi_t x_{t-1|n} + \Psi_t q_t$ and $x_{t|n} = x_{t|t-1} + P_{t|t-1}q_t$ of (84) and (78) respectively. The latter equation—De Jong's—is favoured by the fact that it uses one less vector–matrix multiplication in each step. However, Koopman claims that the former equation leads to a more efficient algorithm when the structure of the matrices Φ_t and Ψ_t is taken into account. A clear advantage of this algorithm over De Jong's is its limited use of memory, since there is no need to retain the values of $P_{t|t-1}$ and $x_{t|t-1}$. However, this advantage does not extend to the calculation of the dispersion matrix of the smoothed estimates.

The second of the state-disturbance algorithms, which is Whittle's [12] algorithm, consists of a forward and a backward pass. Of the products of the forward pass, which involves the Kalman filter, only the value of $x_{n|n}$ is used in further calculations. The backward pass is initialised with $x_{n|n}$ and $q_{n+1} = 0$; and, in each step, values of s_t and q_t are calculated, via (94) and (89) respectively. The smoothed estimate follows from equation (88). This algorithm is efficient in both time and memory; and virtually no storage is required. Its disadvantage is that it is prone to numerical instability; which limits the size of the sample to which it can be safely applied. The primary source of this instability is in the calculation of s_t via equation (94) wherein the elements of Ω^{-1} and of $y_t - H_t x_{t|n}$ are liable to have disparate magnitudes.

7. The Forward–Backward Algorithm

The approach pursued in this final section differs from those found elsewhere in the paper. Instead of conditional expectations, Bayesian analysis is used in deriving a smoothing algorithm; and it is assumed that all random variables are normally distributed.

The forward–backward algorithm which is presented here has been derived by Mayne [10] via the principle of least-squares. It has been rediscovered recently De Vos and Merkus [7], who have used the principle of combining informationto develop a variety of algorithms.

Combining Information. Imagine that the sample is split into two sets $\mathcal{I}_1 = \{y_1, \ldots, y_{t-1}\}$ and $\mathcal{I}_2 = \{y_t, \ldots, y_n\}$. Then, by applying Bayes' rule twice, we get

$$\begin{aligned} N(\xi_t|\mathcal{I}_1, \mathcal{I}_2) &\propto N(\mathcal{I}_2|\xi_t, \mathcal{I}_1)N(\xi_t|\mathcal{I}_1) \\ &= N(\mathcal{I}_2|\xi_t)N(\xi_t|\mathcal{I}_1) \\ &\propto \frac{N(\xi_t|\mathcal{I}_1)N(\xi_t|\mathcal{I}_2)}{N(\xi_t)}, \end{aligned} \tag{95}$$

where the symbol of proportionality indicates that a factor has been omitted. The omitted factors make no reference to the state vector ξ_t.

Within the final expression, the factor $N(\xi_t|\mathcal{I}_1)$ stands for the density function associated with an estimate of ξ_t based upon the information of $\mathcal{I}_1$ and upon prior information. The factor $N(\xi_t|\mathcal{I}_2)$ relates to the density function of an estimate based upon prior information and upon the information $\mathcal{I}_2$ from the second half of the sample. The factor $N(\xi_t)$ stands for a Bayesian prior relative to the state vector.

The formula under (95) indicates how the three factors may be combined to form a smoothed estimate of ξ_t based upon all of the information. The presence of the prior in the denominator of the final expression indicates that the prior information must be subtracted somehow from one or other of the numerator factors to avoid its being used twice in forming $N(\xi_t|\mathcal{I}_1, \mathcal{I}_2)$.

The decision to subtract the prior information from $N(\xi_t|\mathcal{I}_2)$ leads to a so-called inverse model which satisfies

$$N(\xi_t|\mathcal{I}_2; \text{Inv}) \propto \frac{N(\xi_t|\mathcal{I}_2)}{N(\xi_t)}. \tag{96}$$

Substituting (96) into (95) gives

$$N(\xi_t|\mathcal{I}_1, \mathcal{I}_2) \propto N(\xi_t|\mathcal{I}_1)N(\xi_t|\mathcal{I}_2; \text{Inv}). \tag{97}$$

Let us denote the expectation and the dispersion of the inverse model by

$$\tilde{x}_t = E(\xi_t|y_t, \ldots, y_n; \text{Inv}), \tag{98}$$

$$\tilde{P}_t = D(\xi_t|y_t, \ldots, y_n; \text{Inv}). \tag{99}$$

Then, after taking logarithms in equation (97), we can manipulate the exponents of the normal density functions to show that

$$\begin{aligned}
&-2\ln\{N(\xi_t|y_1,\ldots,y_n)\}\\
&\quad\propto -2\ln\{N(\xi_t|y_1,\ldots,y_{t-1})\} - 2\ln\{N(\xi_t|y_t,\ldots,y_n;\mathrm{Inv})\}\\
&\quad\propto (\xi_t - x_{t|t-1})'P_{t|t-1}^{-1}(\xi_t - x_{t|t-1}) + (\xi_t - \tilde{x}_t)'\tilde{P}_t^{-1}(\xi_t - \tilde{x}_t) \qquad (100)\\
&\quad\propto \xi_t'(P_{t|t-1}^{-1} + \tilde{P}_t^{-1})\xi_t - 2\xi_t'(P_{t|t-1}^{-1}x_{t|t-1} + \tilde{P}_t^{-1}\tilde{x}_t)\\
&\quad\propto (\xi_t - x_{t|n})'P_{t|n}^{-1}(\xi_t - x_{t|n}).
\end{aligned}$$

It follows, from comparing the final and the penultimate expressions, that

$$x_{t|n} = \left(P_{t|t-1}^{-1} + \tilde{P}_t^{-1}\right)^{-1}\left(P_{t|t-1}^{-1}x_{t|t-1} + \tilde{P}_t^{-1}\tilde{x}_t\right), \qquad (101)$$

$$P_{t|n} = \left(P_{t|t-1}^{-1} + \tilde{P}_t^{-1}\right)^{-1}. \qquad (102)$$

Equations (101) and (102) show that the estimate of ξ_t which uses all of the data is a weighted average of the estimate using data 'from the left' and the estimate using data 'from the right'.

Although $x_{t|t-1}$ and $P_{t|t-1}$ may be calculated via the Kalman filter, it is more efficient to employ the information filter which is defined under (15) and (16), since only $a_{t|t-1} = P_{t|t-1}^{-1}x_{t|t-1}$ and $P_{t|t-1}^{-1}$ are needed in (101) and (102). To obtain $\tilde{x}_t$ and $\tilde{P}_t$, we may apply filtering techniques to the inverse model; and, for this, we need to derive recursive equations which run in reversed time.

Derivation of the Inverse Model. First, consider the update step of the inverse model. On the one hand is the equation

$$\begin{aligned}
N(\xi_t|y_t,\ldots,y_n;\mathrm{Inv}) &\propto \frac{N(\xi_t|y_t,\ldots,y_n)}{N(\xi_t)}\\
&\propto \frac{N(y_t|\xi_t,y_{t+1},\ldots,y_n)N(\xi_t|y_{t+1},\ldots,y_n)}{N(\xi_t)} \qquad (103)\\
&\propto N(y_t|\xi_t)N(\xi_t|y_{t+1},\ldots,y_n;\mathrm{Inv}).
\end{aligned}$$

which shows how additional information is assimilated to the inverse model. The first and the final proportionalities follow from the definition of the inverse model, whilst the second one comes from applying Bayes' rule. Notice also that, in writing $N(y_t|\xi_t)$ in the final expression, we omit to make y_t conditional on the observations $y_{t+1},\ldots,y_n$ since these are redundant in predicting y_t if ξ_t is known. On the other hand is the equation

$$N(\xi_t|y_t,\ldots,y_n;\mathrm{Inv}) \propto N(y_t|\xi_t,y_{t+1},\ldots,y_n;\mathrm{Inv})N(\xi_t|y_{t+1},\ldots,y_n;\mathrm{Inv}). \qquad (104)$$

By comparing equations (103) and (104), we see that

$$N(y_t|\xi_t, y_{t+1}, \ldots, y_n; \text{Inv}) \propto N(y_t|\xi_t), \tag{105}$$

which indicates that the inverse model has the same measurement equation as the ordinary (forward) state-space model.

In the prediction step, we have, on the one hand, the equation

$$\begin{aligned} N(\xi_{t-1}|y_t, \ldots, y_n; \text{Inv}) &\propto \frac{N(\xi_{t-1}|y_t, \ldots, y_n)}{N(\xi_{t-1})} \\ &= \int \frac{N(\xi_{t-1}|\xi_t, y_t, \ldots, y_n)N(\xi_t|y_t, \ldots, y_n)}{N(\xi_{t-1})} d\xi_t \\ &\propto \int \frac{N(\xi_{t-1}|\xi_t)N(\xi_t)}{N(\xi_{t-1})} N(\xi_t|y_t, \ldots, y_n; \text{Inv}) d\xi_t \\ &= \int N(\xi_t|\xi_{t-1})N(\xi_t|y_t, \ldots, y_n; \text{Inv}) d\xi_t, \end{aligned} \tag{106}$$

where we have used the equality $N(\xi_{t-1}|\xi_t, y_t, \ldots, y_n) = N(\xi_{t-1}|\xi_t)$, which holds due to the fact that the information contained in $y_t, \ldots, y_n$ which is not in ξ_t, relates solely to observation errors. On the other hand, we have the equation

$$\begin{aligned} &N(\xi_{t-1}|y_t, \ldots, y_n; \text{Inv}) = \\ &\qquad \int N(\xi_{t-1}|\xi_t, y_t, \ldots, y_n; \text{Inv})N(\xi_t|y_t, \ldots, y_n; \text{Inv}) d\xi_t. \end{aligned} \tag{107}$$

The comparison of (106) and (107) shows that

$$N(\xi_{t-1}|\xi_t, y_t, \ldots, y_n; \text{Inv}) \propto N(\xi_t|\xi_{t-1}). \tag{108}$$

The latter implies that, in the inverse model, $y_t, \ldots, y_n$ are redundant for predicting ξ_{t-1} if ξ_t is known, and that the transition equation of the inverse model can be calculated from

$$\begin{aligned} &-2\ln\left\{N(\xi_{t-1}|\xi_t; \text{Inv})\right\} \\ &\qquad \propto -2\ln\left\{N(\xi_t|\xi_{t-1})\right\} \\ &\qquad \propto (\xi_t - \Phi_t\xi_{t-1})'\Psi_t^{-1}(\xi_t - \Phi_t\xi_{t-1}) \\ &\qquad \propto \xi_{t-1}'\Phi_t'\Psi_t^{-1}\Phi_t\xi_{t-1} - 2\xi_{t-1}'\Phi_t'\Psi_t^{-1}\xi_t \\ &\qquad \propto (\xi_{t-1} - \Phi_t^{-1}\xi_t)'(\Phi_t^{-1}\Psi_t\Phi_t^{-1\prime})^{-1}(\xi_{t-1} - \Phi_t^{-1}\xi_t). \end{aligned} \tag{109}$$

Here we have assumed that Ψ_t is invertible. However, the result also holds if this matrix is singular. The initialisation of the inverse model follows from the prediction step, for $t-1=n$. Since $N(\xi_n|\text{Inv}) \propto N(\xi_n)/N(\xi_n) = 1$, the inverse model has a non-informative prior distribution.

Equations (105) and (108) thus lead to the following equations for the inverse model:

$$y_t = H_t\xi_t + \tilde{\eta}_t, \tag{110}$$

$$\xi_{t-1} = \Phi_t^{-1}\xi_t + \tilde{\nu}_t, \tag{111}$$

where

$$\tilde{\eta}_t \sim N(0, \Omega_t), \quad \tilde{\eta}_t = \eta_t, \tag{112}$$

$$\tilde{\nu}_t \sim N(0, \Phi_t^{-1}\Psi_t\Phi_t^{-1\prime}) \tag{113}$$

are mutually independent random vectors. Application of the information filter to the inverse model shows that the backward recursions which we are seeking are

$$\tilde{a}_t = H_t'\Omega_t^{-1}y_t + \Phi_{t+1}'(\tilde{P}_{t+1} + \Psi_{t+1})^{-1}\tilde{P}_{t+1}\tilde{a}_{t+1}, \tag{114}$$

$$\tilde{P}_t^{-1} = H_t'\Omega_t^{-1}H_t + \Phi_{t+1}'(\tilde{P}_{t+1} + \Psi_{t+1})^{-1}\Phi_{t+1}, \tag{115}$$

where $\tilde{a}_t = \tilde{P}_t^{-1}\tilde{x}_t$.

To summarise, the forward-backward algorithm consists of three steps:

1. calculate $P_{t|t-1}^{-1}$ and $a_{t|t-1}$, using the information filter or the Kalman filter.
2. calculate $\tilde{P}_t^{-1}$ and $\tilde{a}_t$.
3. combine both estimates using the smoothing formulae (101) and (102) to get the smoothed estimates.

A relation with the algorithm avoiding an inversion is found by applying the matrix inversion lemma to (102); this results in

$$P_{t|n} = P_{t|t-1} - P_{t|t-1}(P_{t|t-1} + \tilde{P}_t)^{-1}P_{t|t-1}. \tag{116}$$

As is easily verified, (101) can now be rewritten as

$$x_{t|n} = x_{t|t-1} + P_{t|t-1}(P_{t|t-1} + \tilde{P}_t)^{-1}(\tilde{x}_t - x_{t|t-1}). \tag{117}$$

The comparison of (117) and (78) indicates that

$$q_t = (P_{t|t-1} + \tilde{P}_t)^{-1}(\tilde{x}_t - x_{t|t-1}); \tag{118}$$

equations (116) and (79) together show that

$$Q_t = (P_{t|t-1} + \tilde{P}_t)^{-1}. \tag{119}$$

These identities suggest that the forward–backward algorithm is less efficient than the algorithms of De Jong [6] and Koopman [9].

In concluding this section, we should mention that the forward–backward smoothing algorithm is particularly useful in computing cross-validation errors for a state-space model. The cross-validation error associated with a given sample element is the error in predicting that element using the information from the rest of the sample. The estimate of the state vector upon which the prediction is based can be calculated most efficiently by combining the products of a forward and a backward filter proceeding from either end of the sample. These filters stop short of including information from the sample element whose value is to be predicted. Alternative algorithms which serve the same purpose has been provided by De Jong [5] and by Ansley and Kohn [3].

References

[1] Anderson, B.D.O. and J.B. Moore, (1979), *Optimal Filtering*, Prentice–Hall, Englewood Cliffs, New Jersey.

[2] Ansley, C.F. and R.Kohn, (1982), A Geometrical Derivation of the Fixed Interval Smoothing Equations, *Biometrika*, 69, 486–7.

[3] Ansley, C.F. and R. Kohn, (1989), A Fast Algorithm for Signal Extraction, Influence and Cross-Validation in State Space Models, *Biometrika*, 76, 65–79.

[4] Chow, G.C., (1983), *Econometrics*, McGraw–Hill, New York.

[5] De Jong, P., (1988), A Cross-Validation Filter for Time Series Models, *Biometrika*, 75, 594–600.

[6] De Jong, P., (1989), Smoothing and Interpolation with the State-Space Model, *Journal of the American Statistical Association*, 84, 1085–1088.

[7] De Vos, A.F., and H.R. Merkus, (1991), The Prior, the Past, the Present and the Future: Smoothing Algorithms in the Kalman Filter as a Combination of Information, Discussion paper of the Department of Econometrics, The Free University of Amsterdam.

[8] Farooq, M. and A.K. Mahalanabis, (1971), A Note on the Maximum Likelihood State Estimation of Linear Discrete Systems with Multiple Time Delays, *IEEE Transactions on Automatic Control*, AC-16, 105–106.

[9] Koopman, S.J., (1990), Efficient Smoothing Algorithms for Time Series Models, discussion paper of the Department of Statistics, The London School of Economics.

[10] Mayne, D.Q., (1966), A Solution of the Smoothing Problem for Linear Dynamic Systems *Automatica*, 4, 73-92.

[11] Premier, R. and A.G. Vacroux, (1971), On Smoothing in Linear Discrete Systems with Time Delays, *International Journal of Control*, 13, 299–303.

[12] Whittle, P., (1991), Likelihood and Cost as Path Integrals, *Journal of the Royal Statistical Society, Series B*, 53, 505–538.

[13] Willman, W.W., (1969), On the Linear Smoothing Problem, *IEEE Transaction on Automatic Control*, AC-14, 116–117.

1991-1	N.M. van Dijk	On the Effect of Small Loss Probabilities in Input/Output Transmission Delay Systems
1991-2	N.M. van Dijk	Letters to the Editor: On a Simple Proof of Uniformization for Continious and Discrete-State Continious-Time Markov Chains
1991-3	N.M. van Dijk P.G. Taylor	An Error Bound for Approximating Discrete Time Servicing by a Processor Sharing Modification
1991-4	W. Henderson C.E.M. Pearce P.G. Taylor N.M. van Dijk	Insensitivity in Discrete Time Generalized Semi-Markov Processes
1991-5	N.M. van Dijk	On Error Bound Analysis for Transient Continuous-Time Markov Reward Structures
1991-6	N.M. van Dijk	On Uniformization for Nonhomogeneous Markov Chains
1991-7	N.M. van Dijk	Product Forms for Metropolitan Area Networks
1991-8	N.M. van Dijk	A Product Form Extension for Discrete-Time Communication Protocols
1991-9	N.M. van Dijk	A Note on Monotonicity in Multicasting
1991-10	N.M. van Dijk	An Exact Solution for a Finite Slotted Server Model
1991-11	N.M. van Dijk	On Product Form Approximations for Communication Networks with Losses: Error Bounds
1991-12	N.M. van Dijk	Simple Performability Bounds for Communication Networks
1991-13	N.M. van Dijk	Product Forms for Queueing Networks with Limited Clusters
1991-14	F.A.G. den Butter	Technische Ontwikkeling, Groei en Arbeidsproduktiviteit
1991-15	J.C.J.M. van den Bergh, P. Nijkamp	Operationalizing Sustainable Development: Dynamic Economic-Ecological Models
1991-16	J.C.J.M. van den Bergh	Sustainable Economic Development: An Overview
1991-17	J. Barendregt	Het mededingingsbeleid in Nederland: Konjunktuurgevoeligheid en effektiviteit
1991-18	B. Hanzon	On the Closure of Several Sets of ARMA and Linear State Space Models with a given Structure
1991-19	S. Eijffinger A. van Rixtel	The Japanese Financial System and Monetary Policy: a Descriptive Review
1991-20	L.J.G. van Wissen F. Bonnerman	A Dynamic Model of Simultaneous Migration and Labour Market Behaviour