A New Proof and an Adjoint Filter Interpretation for Linear Discrete Time Smoothing

Hagander, Per

1973

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Hagander, P. (1973). *A New Proof and an Adjoint Filter Interpretation for Linear Discrete Time Smoothing*. (Research Reports TFRT-3060). Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

TFRT - 3060

A NEW PROOF AND AN ADJOINT FILTER INTERPRETATION FOR LINEAR DISCRETE TIME SMOOTHING

PER HAGANDER

Report 7330 September 1973
Lund Institute of Technology
Division of Automatic Control

A NEW PROOF AND AN ADJOINT FILTER INTERPRETATION FOR LINEAR DISCRETE TIME SMOOTHING

Per Hagander

Abstract

Linear discrete time systems, usually formulated using difference equations, can also be described by operators, which is more general. The covariances for a stochastic system are expressed as operators, and the solution of the fixed interval smoothing problem is obtained by use of the projection theorem:

$$\hat{x} = R_{xy}R_y^{-1}y$$

The computation of $\hat{x}$ is conveniently done if R_y can be factored into two Volterra (triangular) operators. It is shown how this factorization can be carried out using the Riccati equation, so that the estimate can be expressed as two adjoint coupled filters, the Bryson-Frazier formulation.

From the operator identity used for factorization it is seen that the one step ahead predictor is fundamental. Both the forward backward difference equations and the weighting function representation are presented, and the weighting function is shown to be the error covariance of the one step ahead predictor.

1. Introduction

The two approaches to linear estimation problems, the Wiener filter using covariance functions and the Kalman filter directly using difference or differential equations can be unified by use of the Riccati equation. [7,12]. In [7] the continous time linear control and estimation problems were analysed using operators in function spaces. The same technique is applicable in the discrete time case. This is demonstrated here on the smoothing problem. The projection theorem gives an equation in covariance operators from which the difference equations are obtained by operator factorization using the discrete Riccati equation. The resolvent identity searched for by Kailath and Frost [9] is thus presented.

2. Notations

Consider a discrete time system for $t \in [t_o, t_1]$

$$x(t+1) = \phi(t+1,t)x(t) + v(t), \qquad x(t_o) = x_o \tag{2.1}$$

The state of (2.1) $x(t)$, at all times t during the discrete time interval $[t_o, t_1]$, can be formed as a long vector with $n(t_1 - t_o + 1)$ elements, but it can also be regarded as a time function on $[t_o, t_1]$. The difference equation (2.1) can be formulated using linear operators in a space X of such functions:

$$x = gx_o + Lv$$

with $L: X \to X$ and $g: R^n \to X$. The operator formulation is more general than (2.1), and (2.1) thus introduces special structure on the operators.

Using the long vector interpretation these operators are in fact large matrices. Since L is causal it corresponds to a lower block triangular matrix.

Define in X the scalar product

$$x_1 \cdot x_2 = \sum_{t=t_o}^{t_1} x_1^T(t) x_2(t)$$

giving the adjoint of L:

$$L^*: X \to X; \quad z = L^*x, \quad z(t) = \sum_{s=t+1}^{t_1} \phi^T(s,t+1)x(s), \qquad \phi(t,s) = \prod_{i=s}^{t-1} \phi(i+1,i)$$

Define also the functions from X to R^n:

$$T_o x = x(t_o), \qquad T_1 x = x(t_1)$$

3. Linear Stochastic, Time Discrete Systems

The space X can be extended to contain stochastic processes generated by linear systems driven by white noise. Such a Hilbert space is often used in the theory of stochastic processes, cf [4]. Let v and e of

$$x(t+1) = \phi x(t) + v(t), \qquad x(t_o) = x_o$$
$$y(t) = \theta x(t) + e(t) \tag{3.1}$$

be zero mean, independent white noise with covariances R_1 and R_2 $(R_2>0)$, and let x_o have zero mean value, covariance R_o and be independent of v and e. The operators L and g are directly generalized. A new scalar product

$$x_1 \cdot x_2 = E \sum_{t=t_o}^{t_1} x_1^T(t) x_2(t)$$

gives the same adjoints. Notice that the deterministic functions constitute a subspace. The covariance operator of x is easily obtained from the reformulation of (3.1).

$$x = Lv + gx_o$$

using the matrix point of view:

$$R_x = LR_1L^* + gR_og^*$$

where R_1 is now an operator in X (or a block diagonal matrix).
Moreover, $R_{xy} = R_x\theta^T$ and $R_y = \theta R_x \theta^T + R_2$. θ^T is a diagonal operator with $\theta^T(t)$ in the diagonal.

4. Smoothing Estimate

All linear estimates of x based on $\{y(t_o),\ldots,y(t_1)\}$ can be written

$$\hat{x} = Fy$$

If the operator F is such that the error variance in each component $\hat{x}_i(t)$ is minimized then $\hat{x}$ is the smoothing estimate of x.

Using the projection theorem [4], F must satisfy

$$R_{xy} = FR_y$$

In the continuous time case it has been shown that the Riccati equation decomposes operators like R_y into a product of a causal and an anti-causal part [7,8,12]. The algebra of the discrete time case is more involved and a corresponding identity has not been obtained previously, cf[9].

In its most simplified form the discrete time identity can be formulated as

Theorem 1: Let $P(t)$ be the solution of

$$P(t+1) = \phi P(t)\phi^T + I - \phi P(t)(I+P(t))^{-1}P(t)\phi^T$$
$$P(t_o) = 0$$

then

$$I + LL^* = (I+P+L\phi P)(I+P)^{-1}(I+P+P\phi^T L^*)$$

where P is a blockdiagonal operator with $P(t)$ in the diagonal.

Proof: With the forward and backward shift operators q and q^{-1} defined by $qx(t)=x(t+1)$, $qx(t_1)=0$, $q^{-1}x(t)=x(t-1)$, $q^{-1}x(t_o)=0$, it is easy to prove that

$$L(q-\phi) = I - gT_o \tag{4.1}$$
$$(q^{-1}-\phi^T)L^* = I - (gT_o)^* \tag{4.2}$$

so that the proposition

$$I + LL^* = I + P + L\phi P + P\phi^T L^* + L\phi P(P+I)^{-1}P\phi^T L^*$$

could be written

$$I+LL^* = I+L(q-\phi)P(q^{-1}-\phi^T)L^* + L\phi P(q^{-1}-\phi^T)L^* + L(q-\phi)P\phi^T L^* + L\phi P(P+I)^{-1}P\phi^T L^* +$$
$$+ gT_oP + L(q-\phi)P(gT_o)^* + L\phi P(gT_o)^* + gT_oP\phi^T L^*$$

When $T_oP = 0$ this requires

$$L\,\{I - qPq^{-1} + \phi P\phi^T - \phi P(I+P)^{-1}P\phi^T\}\,L^* = 0$$

which is true for $P(t)$ from the Riccati equation. □

The problem of decomposing R_y is solved by a generalization of Theorem 1.

Corollary 1: Let P be defined by

$$P(t+1) = \phi P(t)\phi^T + R_1 - \phi P\theta^T(\theta P\theta^T+R_2)^{-1}\theta P\phi^T$$
$$P(t_o) = R_o \tag{4.3}$$

then

$$R_x = L\phi P + P + P\phi^T L^* + L\phi P\theta^T(\theta P\theta^T+R_2)^{-1}\theta P\phi^T L^*$$
$$R_{xy} = P(I+\phi^T L^*)\theta^T + L\phi P\theta^T(\theta P\theta^T+R_2)^{-1}(\theta P\theta^T+R_2+\theta P\phi^T L^*\theta^T)$$
$$R_y = [\theta P\theta^T+R_2+\theta L\phi P\theta^T](\theta P\theta^T+R_2)^{-1}[\theta P\theta^T+R_2+\theta P\phi^T L^*\theta^T]$$

Proof: The only difficulty compared with Theorem 1 is the initial value. Just prove that

$$g^* = T_o(\phi^T L^* + I)$$

and use $T_oP = R_oT_o$. □

The smoothing formula can now be obtained using inversion of operators.

Theorem 2: The smoothing estimate for the system (3.1) is given by

$$\hat{x}(t|t_1) = \hat{x}(t|t-1) + P(t)\lambda(t-1) \qquad t_o \le t \le t_1$$

where $\hat{x}(t|t-1) = \hat{x}_p(t)$ is the one step ahead predictor.

$$\hat{x}_p(t+1) = \psi(t+1,t)\hat{x}_p(t) + K(t)y(t), \qquad \hat{x}_p(t_o) = 0 \tag{4.4}$$

$$\psi(t+1,t) = \phi(t+1,t) - K(t)\theta(t) \tag{4.5}$$

$$K = \phi P\theta^T(\theta P\theta^T+R_2)^{-1} \tag{4.6}$$

and λ the solution to an adjoint equation

$$\lambda(t-1) = \psi^T(t+1,t)\lambda(t) + \theta^T(\theta P\theta^T+R_2)^{-1}(y(t)-\theta\hat{x}_p(t)), \quad \lambda(t_1) = 0 \tag{4.7}$$

$P(t)$ is defined by (4.3).

Proof: Since the operator $[(\theta P\theta^T+R_2) + \theta L\phi P\theta^T]$ represents an invertible dynamical system [6], R_y is also invertible and using corollary 1 and (4.2):

$$\hat{x} = \{P(I+\phi^T L^*)\theta^T(\theta P\theta^T+R_2+\theta P\phi^T L^*\theta^T)^{-1}(\theta P\theta^T+R_2)+L\phi P\theta^T\}(\theta P\theta^T+R_2+\theta L\phi P\theta^T)^{-1}y$$

Introduce an operator M analogously to L:

$$(Mx)(t) = \sum_{s=t_o}^{t-1} \psi(t,s+1)x(s) \quad , \; \psi(t,s) = \prod_{i=s}^{t-1} \psi(i+1,i) \tag{4.8}$$

with ψ from (4.5). Using the same technique as in [6] it is possible to prove the following operator identities:

$$(\theta P\theta^T+R_2+\theta L\phi P\theta^T)^{-1} = (\theta P\theta^T+R_2)^{-1} - (\theta P\theta^T+R_2)^{-1}\theta MK$$

$$L\phi P\theta^T(\theta P\theta^T+R_2+\theta L\phi P\theta^T)^{-1} = MK$$

$$L^*\theta^T(\theta P\theta^T+R_2+\theta P\phi^T L^*\theta^T)^{-1} = M^*\theta^T(\theta P\theta^T+R_2)^{-1}$$

with K from (4.6), so that

$$\hat{x} = MKy + P(I-\theta^T K^T M^*+\phi^T M^*)\theta^T(\theta P\theta^T+R_2)^{-1}(I-\theta MK)y = MKy + P(I+\psi^T M^*)\theta^T(\theta P\theta^T+R_2)^{-1}(I-\theta MK)y$$

Note that $\hat{x}_p$ and λ defined by (4.4) and (4.7) can be written as

$$\hat{x}_p = MKy \tag{4.9}$$

$$\lambda = M^*\theta^T(\theta P\theta^T+R_2)^{-1}(y-\theta\hat{x}_p) \tag{4.10}$$

which proves the theorem. □

5. Adjoint Filter Interpretation

Introduce

$$\tilde{x}_p(t) = x(t) - \hat{x}_p(t)$$

$$\tilde{y}_p(t) = y(t) - \theta\hat{x}_p(t)$$

and the covariance function $P(t,s)$ of $\tilde{x}_p$, cf [2].

$$P(t,s) = \begin{cases} \psi(t,s)P(s) & t > s \\ P(t) & t = s \\ P(t)\psi^T(s,t) & t < s \end{cases} \tag{5.1}$$

and define

$$y_m = \theta^T R_2^{-1} y \tag{5.2}$$

$$\nu_m = \theta^T(\theta P \theta^T + R_2)^{-1} \tilde{y}_p \tag{5.3}$$

The adjoint filter form of the smoothing estimate was derived in [9] for the continuous time case. A corresponding formula for discrete time can be formulated:

Corollary 2: The smoothing estimate for the system (3.1) is given by

$$\hat{x}(t) = \hat{x}(t|t_1) = \sum_{s=t_o}^{t-1} P(t,s) y_m(s) + \sum_{s=t}^{t_1} P(t,s)\nu_m(s) \tag{5.4}$$

with $P(t,s)$, y_m and ν_m defined by (5.1), (5.2) and (5.3).

Proof: First notice that

$$K = \phi P \theta^T(\theta P \theta^T + R_2)^{-1} = (\phi - \phi P \theta^T(\theta P \theta^T + R_2)^{-1}\theta) P \theta^T R_2^{-1} = (\phi - K\theta) P \theta^T R_2^{-1} \tag{5.5}$$

Hence from (4.9)

$$\hat{x}_p(t) = \sum_{s=t_o}^{t-1} \psi(t,s) P(s) \theta^T R_2^{-1} y(s) = \sum_{s=t_o}^{t-1} P(t,s) y_m(s)$$

and from (4.10)

$$P(t)\lambda(t-1) = P(t) \sum_{s=t}^{t_1} \psi^T(s,t)\theta^T(\theta P \theta^T + R_2)^{-1} \tilde{y}_p(s) = \sum_{s=t}^{t_1} P(t,s)\nu_m(s)$$

which proves the Corollary. □

Remark: The fixed point smoothing problem is directly solved from eq (5.4).

$$\begin{aligned} \hat{x}(t|s+1) &= \hat{x}(t|s) + B(s+1)\nu_m(s+1) \qquad s \geq t-1 \\ B(t) &= P(t) \\ B(s+1) &= B(s)\psi^T(s+1,s) = B(s)[\phi^T(s+1,s) - \theta^T(s)\, K^T(s)] \end{aligned} \tag{5.6}$$

Eq (5.6) could also be used to evaluate the fixed interval estimate $\hat{x}(t|t_1)$. This recursion is on the stability boundary but it has computational advantages since it can be performed in parallel with the one step ahead predictor and the Riccati equation. Variants of Theorem 2 and (5.6) have been presented earlier, see [3,9]. Some other formulations contain an unstable recursion or inversions to be performed in each time step [1,5,10,11].

6. Conclusions

This note presents a new proof of the discrete time smoothing problem by means of an operator identity searched for by Kailath and Frost [9]. It also gives the adjoint formula in the sense of [9, p 656]. The main difference between earlier derivations and this one is that the estimation is done once for all directly in the function space giving an analogue of the Wiener Hopf equation, and not performed recursively in time.

The role of the Riccati equation in the factorization of the critical operator R_y is made clear, and the importance of the one step ahead predictor is more obvious.

References

[1] B.D.O. Anderson and S. Chirarattananon: New Linear Smoothing Formulas, IEEE Trans. Automat. Contr vol AC-17, pp 160-161, Feb 1972.

[2] K.J. Åström: Introduction to Stochastic Control Theory. Academic Press, New York 1970.

[3] A.E. Bryson and M. Frazier: Smoothing for Linear and Nonlinear Dynamic Systems, Proc. Optimum Sys. Synthesis Conf., U.S. Air Force Tech. Rept. ASO-TOR-63-119, Feb 1963.

[4] H. Cramér and M.R. Leadbetter: Stationary and Related Stochastic Processes, John Wiley, New York 1967.

[5] D.C. Fraser and J.E. Potter: The Optimum Linear Smoother as a Combination of Two Optimum Linear Filters, IEEE Trans. Automat Contr vol AC-14, pp 387-390, Aug 1969.

[6] P. Hagander: Inversion of a Dynamical System by an Operator Identity, Automatica 8, p 361-362, May 1972.

[7] P. Hagander: The Use of Operator Factorization for Linear Control and Estimation, Automatica 9, Sept 1973.

[8] T. Kailath: Application of a Resolvent Identity to a Linear Smoothing Problem, SIAM J Control vol 7, pp 68-74, Feb 1969.

[9] T. Kailath and P.Frost: An Innovation Approach to Least-Squares Estimation, Part II Linear Smoothing in Additive White Noise, IEEE Trans. Automat. Contr. vol AC-13, pp 655-660, Dec 1968.

[10] J.S. Meditch: Stochastic Optimal Linear Estimation and Control, McGraw Hill, New York 1969.

[11] H.E. Rauch, F. Tung, and C.T. Striebel: Maximum Likelihood Estimates of Linear Dynamic Systems, AIAA J. vol 3, pp 1445-1450, Aug 1965.

[12] A. Schumitsky: On the Equivalence between Matrix Riccati Equations and Fredholm Resolvents, Journal of Computer and System Sciences, vol 2, pp 76-87, June 1968.