DESIGN PROCEDURE FOR STABLE SUBOPTIMAL FIXED LAG SMOOTHERS

BRIAN D.O. ANDERSON and HOANG B. DOAN

Department of Electrical Engineering
University of Newcastle, New South Wales, 2308, Australia

Abstract

A procedure is presented for designing stable suboptimal fixed-lag smoothers for the continuous-time linear gaussian problem. Apart from incorporating an ideal delay, the smoothers are finite-dimensional. The design procedure involves the approximation by a very straightforward procedure of a matrix impulse response of compact support by a matrix impulse response with a rational and stable Laplace transform. The approximation procedure involves studying the pole-zero pattern of transfer functions derivable from the transfer function matrix being approximated.

I. Introduction

For the linear-gaussian problem, it is well known that fixed-lag smoothing offers the possibility of achieving significantly lower error covariances than normal filtering, which is equivalent to smoothing with zero lag, [1, 2]. It is perhaps less well known that difficulties often arise in implementing continuous-time fixed-lag smoothers, since optimal fixed-lag smoothers are infinite-dimensional, and likely to be unstable, [3].

Several techniques for eliminating these difficulties have been proposed, see e.g. [4-6]. In this paper, we explore in more detail one of the schemes suggested in [6], for obtaining a suboptimal, stable fixed lag smoother which is finite-dimensional, save for the inclusion of a delay. The scheme is applicable only when all processes are stationary.

II. The Smoother Equation

Suppose that the signal model is of the usual type

$$\dot{x} = Fx + Gw \qquad z = H'x + v \tag{1}$$

with $\mathrm{Re}\lambda_i(F) < 0$, with the signal model operating from $-\infty$, and with $[w' \ v']'$ a zero mean gaussian process with

$$E\left\{\begin{bmatrix} w(t) \\ v(t) \end{bmatrix} [w'(s) \quad v'(s)]\right\} = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \delta(t-s) \tag{2}$$

where R is nonsingular, (The case of correlated $w(\cdot)$ and $v(\cdot)$ can also be handled).

Let Σ be the unique nonnegative definite solution of

$$\Sigma F' + F\Sigma - \Sigma HR^{-1}H'\Sigma + GQG' = 0 \tag{3}$$

Let $\hat{x}(t|t)$ be the state estimate of $x(t)$ provided at time t by the Kalman filter, let $\nu(t) = z(t) - H'\hat{x}(t|t)$ denote the filter innovations process, and let $\bar{F} = F - KH'$ where K is the Kalman filter gain. Then the fixed lag smoothed estimate $\hat{x}(t|t+\Delta) = E[x(t)|z(s), s < t+\Delta]$ is given by [7]

$$\hat{x}(t|t+\Delta) = \hat{x}(t|t) + \Sigma\int_0^{\Delta}[\exp \bar{F}'(\Delta-\tau)]HR^{-1}\nu(t+\Delta-\tau)d\tau \tag{4}$$

A delay line can be used to store $\hat{x}(t|t)$ for a time Δ. The second contribution to $\hat{x}(t|t+\Delta)$ can be obtained by passing $HR^{-1}\nu(\cdot)$ through a linear system with impulse response

$$M(t) = \Sigma \exp \bar{F}'(\Delta - t) \qquad 0 \le t \le \Delta \tag{5}$$
$$= 0 \qquad \text{elsewhere}$$

for which the associated Laplace transform is

$$\mathcal{M}(s) = \Sigma[sI + \bar{F}']^{-1}[e^{\bar{F}'\Delta} - e^{-s\Delta}I] \tag{6}$$

III Suboptimal Smoother

It is suggested in [6] that a suboptimal smoother can be obtained by approximating (6) by a rational transfer function matrix with all poles in $\mathrm{Re}[s] < 0$. Our purpose here is to explain how such an approximation may be undertaken.

To keep matters simple, assume that $\bar{F}$ is diagonalizable. Let

$$\bar{F} = T\Lambda T^{-1}$$

with

$$\Lambda = \mathrm{diag}\left\{\lambda_1, \ldots, \lambda_k, \begin{bmatrix} \alpha_1 & -\beta_1 \\ \beta_1 & \alpha_1 \end{bmatrix} \cdots \begin{bmatrix} \alpha_j & -\beta_j \\ \beta_j & \alpha_j \end{bmatrix}\right\}$$

One can check that

$$T'\Sigma^{-1}M(s)(T')^{-1}$$

$$= \mathrm{diag}\{\ell_1(s), \ldots, \ell_k(s), \begin{bmatrix} c_1(s) & -d_1(s) \\ d_1(s) & c_1(s) \end{bmatrix}, \ldots; \begin{bmatrix} c_j(s) & -d_j(s) \\ d_j(s) & c_j(s) \end{bmatrix} \} \quad (7)$$

where

$$\ell_i(s) = \frac{e^{\lambda_i \Delta} - e^{-s\Delta}}{s + \lambda_i} \quad (8)$$

$$\begin{bmatrix} c_k(s) & -d_k(s) \\ d_k(s) & c_k(s) \end{bmatrix} = \begin{bmatrix} s+\alpha_k & \beta_k \\ -\beta_k & s+\alpha_k \end{bmatrix}^{-1} \begin{bmatrix} e^{\alpha_k \Delta}\cos\beta_k\Delta - e^{-s\Delta} & e^{\alpha_k\Delta}\sin\beta_k\Delta \\ -e^{\alpha_k\Delta}\sin\beta_k\Delta & e^{\alpha_k\Delta}\cos\beta_k\Delta - e^{-s\Delta} \end{bmatrix} \quad (9)$$

Evidently, the task of approximating $M(s)$ is equivalent to the task of approximating $\ell_i(s)$, $c_k(s)$ and $d_k(s)$. Observe that $\ell_i(s)$ is effectively defined by one parameter; then one has a one-parameter family of approximation problems to solve. On the other hand, $c_k(s)$ and $d_k(s)$ are governed effectively by two parameters.

We consider now the approximation of $\ell_i(s)$, $c_k(s)$ and $d_k(s)$.

IV. Approximation of $\ell_i(s)$

It is easily checked that $\ell_i(s)$ has no poles, and an infinite number of zeros. Further, save for a missing zero on the real axis, these zeros are uniformly spaced on a line parallel to the imaginary axis, see Figure 1.

If one plots the amplitude response $|\ell_i(j\omega)|$ as a function of ω, the response is roughly flat up to a certain frequency, and then falls off at 6db/octave. This suggests that one should approximate $\ell_i(s)$ by a rational transfer function $\bar{\ell}_i(s)$ determined in the following way:

(i) the zeros of $\bar{\ell}_i(s)$ coincide with those zeros of $\ell_i(s)$ affecting the passband response, while those zeros remote from values of $j\omega$ in the passband are discarded; thus the zero set of $\ell_i(s)$ is truncated to yield the zeros of $\bar{\ell}_i(s)$. (The possibility of truncating the zero pattern of $\ell_i(s)$ was suggested by Professor Ian Rhodes).

(ii) the poles of $\bar{\ell}_i(s)$ are chosen in the left half plane (for stability), to be one greater in number than the number of zeros [to ensure the same high-frequency roll-off of $\ell_i(s)$ and $\bar{\ell}_i(s)$] and in a Butterworth [8] configuration on a radius significantly larger than the 3db frequency of $\ell_i(j\omega)$ [thereby ensuring that their presence has little or no affect on the behaviour of $\bar{\ell}_i(s)$ in the passband].

(iii) the DC gain of $\bar{\ell}_i(s)$ is chosen to coincide with that of $\ell_i(s)$.

A typical pole-zero plot for $\bar{\ell}_i(s)$ is shown in Figure 2.

V. Approximation of $c_j(s)$ and $d_j(s)$

It turns out that one cannot vary the approach for approximating $\ell_i(s)$ to cope with the separate approximation of $c_k(s)$ and $d_k(s)$. One can however seek to approximate $c_k(s)$ and $d_k(s)$ simultaneously. A little manipulation using (9) reveals that

$$c_k(s) \pm jd_k(s) = \frac{1}{(s+\alpha_k) \pm j\beta_k}[e^{(\alpha_k \pm j\beta_k)\Delta} - e^{-s\Delta}] \quad (10)$$

There are no poles; the zero patterns are shown in Figure 3, and are like that of $\ell_i(s)$, save that they are not symmetrically located with respect to the real axis.

It is obvious that we can approximate $c_k(s) + jd_k(s)$ and $c_k(s) - jd_k(s)$ in exactly the same way as $\ell_i(s)$ provided that

(i) we allow for the zero off-set distance $\pm\beta$ from the real axis by shifting all poles and zeros of the approximant up or down by the same offset.

(ii) we allow the scaling constant used to make the true function and the approximant agree at $\omega = 0$ to be a complex constant.

Denote the approximants by $\gamma_+(s)$ and $\gamma_-(s)$; they are rational functions with complex coefficients. Clearly, it is natural, having found $\gamma_+(s)$, to arrange for $\gamma_-(s)$ to have as its poles and zeros the reflections in the real axis of the poles and zeros of $\gamma_+(s)$, while also by construction, $\gamma_+(0) = \gamma_-^*(0)$. Thus if $\gamma_+(s) = p(s)/q(s)$ for polynomial $p(\cdot)$ and $q(\cdot)$, then $\gamma_-(s) = p^*(s)/q^*(s)$, where the superscript asterisk denotes conjugation of coefficients. Then, since $2c_k(s) = [c_k(s) + jd_k(s)] + [c_k(s) - jd_k(s)]$ it is logical to take as the approximation $\bar{c}_k(s)$ of $c_k(s)$ the expression

$$\bar{c}_k(s) = \tfrac{1}{2}[\gamma_+(s) + \gamma_-(s)] \quad (11)$$

Similarly, the approximation $\bar{d}_k(s)$ of $d_k(s)$ is taken as

$$\bar{d}_k(s) = \frac{1}{2j}[\gamma_+(s) - \gamma_-(s)] \quad (12)$$

One can readily check that $\bar{c}_k(s)$ and $\bar{d}_k(s)$ are both ratios of real polynomials, the denominators being Hurwitz.

VI. Remarks

The above approximation procedure leaves to the designer the major question of how many zeros should be retained in the approximating transfer function, and the minor question of how large a radius should be taken for the Butterworth circle. (This is minor since a suitably large radius is known to be satisfactory). Rough guides in making these choices have been given above. But it is important to note that trial and error designs are perfectly feasible. This is because, using methods outlined in [6], one can always evaluate the performance of a suboptimal smoother. Consequently, if the performance is deemed unsatisfactorily far from optimum, one can increase the number of zeros.

The authors have checked that the procedure works in the sense of not requiring inordinately many zeros on a number of examples, including the following.

VII. Example

The original model is defined by

$$F = \begin{bmatrix} 0 & 1 \\ -2 & -2 \end{bmatrix} \quad G = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad H' = [1 \quad 0] \quad Q = 1000 \quad R = 1$$

and one obtains

$$\Sigma = \begin{bmatrix} 5.961 & 17.76 \\ 17.76 & 153.3 \end{bmatrix} \quad \bar{F} = \begin{bmatrix} -5.961 & 1 \\ -19.76 & -2 \end{bmatrix}$$

Eigenvalues of $\bar{F} = -\alpha \pm j\beta = -3.98 \pm j3.98$

Optimal fixed-lag error covariance with infinite

$$\Delta = \begin{bmatrix} 1.9827 & 0.00043 \\ 0.00043 & 62.812 \end{bmatrix}$$

Optimal fixed-lag error covariance with

$$\Delta \text{ of } 0.5 = \begin{bmatrix} 2.014 & 0.0007 \\ 0.0007 & 63.18 \end{bmatrix}$$

(Note that Δ is approximately twice the dominant time constant of the filter and almost all the improvement possible using smoothing is achieved with this Δ).

For the approximation, use two zeros for $\gamma_+(s)$ and for $\gamma_-(s)$ and a radius for the Butterworth circle of 2α. (This is the smallest number of zeros it is possible to use). One can then calculate

Suboptimal fixed-lag error covariance with

$$\Delta \text{ of } 0.5 = \begin{bmatrix} 2.915 & 3.392 \\ 3.392 & 82.727 \end{bmatrix}$$

This should be compared with the optimal value and the filter error covariance; it is obvious that the error covariance is significantly reduced. Better and worse results are obtained for this problem by different suboptimal smoothers described in [6].

VIII. Conclusions and Possible Future Research

Despite the difficulty of building optimal continuous-time fixed-lag smoothers, one can obtain good suboptimal smoothers in the stationary case. The design of these smoothers rests on the the approximation of a matrix impulse response of finite support. The key contribution of the paper is to explain a straightforward method for achieving the approximation. In practice, the quality of the approximation always needs to be checked, though there are heuristic guides as to the choice of the free parameters in the approximation procedure. Several questions suggest themselves.

(i) In case the signal process is narrow band and band-pass, centered around f_c with bandwidth $2f_0 \ll f_c$, the time lag to achieve significant improvement due to smoothing is intuitively several times f_0^{-1}, while the smoother bandwidth is approximately f_c. The delay-bandwidth product is accordingly a very large number, and such a smoother is extremely difficult to implement. An alternative approach is needed - perhaps involving band-pass to low-pass transformation.

(ii) There are some situations in which optimal nonlinear filters can sometimes be effectively replaced by suboptimal linear filters, e.g. in high signal-to-noise ratio FM demodulation. The question then needs to be answered as to whether linear smoother design ideas such as those of this paper could be effectively applied.

(iii) There is an alternative to the smoothing formula (4) of the form

$$\hat{x}(t|t+\Delta) = A\hat{x}(t+\Delta|t+\Delta) + \int_t^{t+\Delta} B(t+\Delta-\tau)\hat{x}(\tau|\tau)d\tau \qquad (11)$$

This formula is derived in [9]. One could seek to approximate the impulse response $B(\cdot)$ (zero outside $[0, \Delta]$) and compare the effectiveness of smoother approximations obtained this way. It is clearly advantageous that no delay here is required. It may also turn out that the approximants can be kept less complex, particularly since they are driven by a band-limited process $\hat{x}(\cdot|\cdot)$ [in contrast to the approximants derived earlier, which are driven by a white process, $\nu(\cdot)$]. The fact that the input is band-limited means that there are less stringent requirements on the accuracy of the approximation at higher frequencies.

We are aware of the existence of unpublished work of P. Hedelin on even more straightforward approaches to the design of suboptimal smoothers. The smoothers obtained for many examples with scalar measurement processes offer strikingly good performance; there may be some difficulty in obtaining such excellent results for vector measurement processes.

References

[1] J.S. Meditch, Stochastic Optimal Linear Estimation and Control, McGraw Hill Book Co., New York, 1969.

[2] B.D.O. Anderson and S. Chirarattananon, "Smoothing as an Improvement on Filtering:

A Universal Bound", *Electronics Letters*, Vol. 7, No. 18, September 1971, p. 524.

[3] C.N. Kelly and B.D.O. Anderson, "On the stability of fixed-lag smoothing algorithms", *Journal of the Franklin Institute*, Vol. 291, No. 4, April 1971, pp. 271-281.

[4] P. Tam and J.B. Moore, "Stable realizations of fixed-lag smoothing equations for continuous-time signals", *IEEE Trans. on Auto. Control*, Vol. AC-19, No. 1, February 1974, pp. 84-87.

[5] S. Chirarattananon and B.D.O. Anderson, "Outline design for stable, continuous-time fixed lag smoothers", *Electronics Letters*, Vol. 8, No. 10, May 1972, pp. 263-264.

[6] S. Chirarattananon and B.D.O. Anderson, "Stable fixed-lag smoothing of continuous time processes", *IEEE Transactions on Information Theory*, Vol. IT-20, No. 1, January 1974, pp. 25-36.

[7] T. Kailath and P. Frost, "An innovations approach to least-squares estimation: Part II Linear smoothing in additive white noise", *IEEE Trans. on Auto. Control*, Vol. AC-13, No. 6, December 1968, pp. 655-660.

[8] J.L. Stewart, Circuit Theory and Design, John Wiley, New York, 1956.

[9] B.D.O. Anderson and S. Chirarattananon, "New linear smoothing formulas", *IEEE Trans. on Auto. Control*, Vol. AC-17, No. 1, February 1972, pp. 160-161.

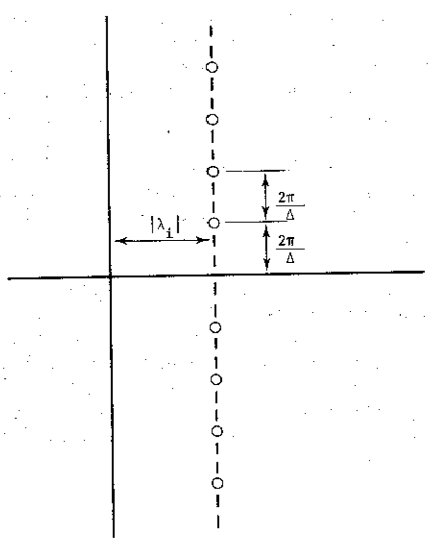

Figure 1 : Zero plot for $\ell_i(s)$

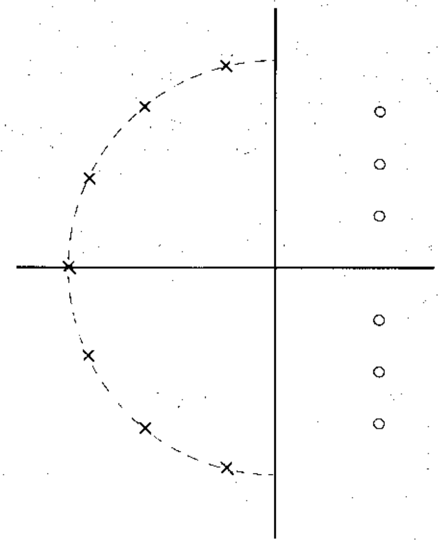

Figure 2 : Typical pole-zero plot for $\bar{\ell}_i(s)$

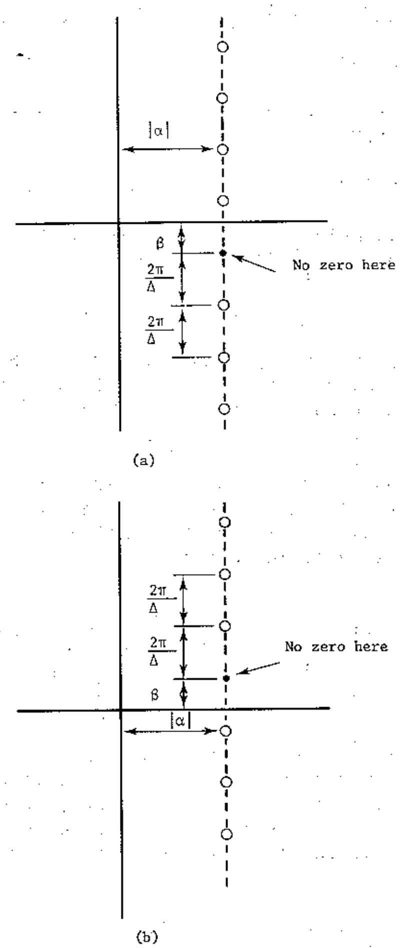

Figure 3 : The zero pattern of

(a) $c_k(s) + jd_k(s)$

and (b) $c_k(s) - jd_k(s)$