Theory and Applications of Nonlinear Control Systems
C.I. Byrnes and A. Lindquist (editors)

THE ASYMPTOTIC APPROXIMATION OF NONLINEAR FILTERS BY LINEAR FILTERS

Arthur J. Krener

Department of Mathematics
University of California
Davis, California 95616
U.S.A.

Abstract

We develop the asymptotics of linear filtering when some of the observation and driving noises are small. We show, using a Girsanov transformation, that linear filters are asymptotically optimal for certain nonlinear filtering problems.

1. INTRODUCTION

At present the most widely used algorithm for nonlinear filtering is the Extended Kalman Filter. This is a heuristically derived method, whose "...performance must be verified by monte carlo simulation. There is no guarantee that the actual estimate obtained will be close to the truly optimal estimate. Fortunately, the extended Kalman filter has been found to yield accurate estimates in a number of important practical applications. Because of this experience and its similarity to the conventional Kalman filter, it is usually one of the first methods to be tried for any nonlinear filtering problem." Gelb et. al. [3, p. 189]

In this paper we would like to briefly describe a new nonlinear filter which in implimentation resembles the Kalman and Extended Kalman Filters. It is not applicable to every nonlinear filtering problem. Loosely speaking, it is appropriate for those problems in which the nonlinearities depend only on state variables which can be estimated quickly and accurately.

If some of the observation and driving noises are small then there are such state variables, but it is not always obvious whether the nonlinearities

*Research suggested in part by NASA under NAG 2-268

depend only on these variables. This is because linearity is in the eyes of the beholder, or more precisely, in the coordinate system of the beholder. Nonlinear changes of coordinates can transform a linear filtering problem into a nonlinear one and vice versa. A key component of our approach is to choose coordinates so that the problem looks as linear as possible. In this paper we shall not describe how this is done but instead refer the reader to [9].

These changes of coordinates are computed off-line when the filtering algorithm is being developed. This increased computational burden and the restriction to problems with small noises and complimentary nonlinearities are the principle disadvantages of our method as compared with the Extended Kalman Filter. The principle advantages are that two-fold. The first is that our filter can be shown to be asymptotically optimal. As the small noises and/or the nonlinearities go to zero our filter performs asymptotically as well as the optimal nonlinear filter. The additional error incurred by using our filter instead of the optimal nonlinear filters is asymptotically smaller than the error of the optimal nonlinear filter.

The second advantage of our filter over the Extended Kalman Filter is that the filter gains can be computed off-line before filtering has begun as in the standard Kalman filter. This greatly reduces the real time computational burden and hence the filter can be implimented by a much slower processor for a given dimensional problem. Moreover, our filter can be taken to its long time limit where the filter gains are constant as in the stationary Kalman and Wiener Filters.

The derivation of our filter is rather complicated as it involves geometric, stochastic and asymptotic techniques. In this paper we give an overview of the asymptotic and stochastic aspects of the approach and only indicate the methods of proof. The full details will appear in [9].

2. ASYMPTOTICS OF LINEAR FILTERING

In this section we discuss the asymptotics of the Kalman-Bucy filter for a linear system with some small observation and driving noises. We describe how some of the states can be estimated quickly and accurately, while others cannot. Haddad [5] has considered the filtering of linear systems with two time scales. Hijab [6,7] applied WKB and large deviations techniques to nonlinear filtering problem where all the noises are small. Our approach is closer to that of Katzur, Bobrovsky and Schuss [8] and Picard [10] who considered the filtering of a one-dimensional nonlinear state process with small observation noise. See also the review article of Blankership [1] and its references.

Our approach is to expand the linear filtering equation in terms of a small parameter which measures the size of the small noises and to solve these equations for their lowest order nonzero terms. We do not give proofs but our approach is a standard one, which can be found in the usual references such as [2].

In this section we are concerned with the linear filtering model

(2.1a) $$dx = Axdt + Bdw$$

(2.2b) $$dy = Cxdt + Ddv$$

(2.2c) $$x(0) = x^0$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^p$, $w(t)$ and $v(t)$ are independent Wiener processes with covariances $Q(t\wedge s)$ and $R(t\wedge s)$ and x^o is an independent Gaussian random vector of mean $\hat{x}^0$ and covariance $P(0)$. The matrices A, B, C, D, Q and R are assumed to be autonomous and (C,A) is assumed to be an observable pair.

We also assume that certain of the observation and driving noises are "small". To make this precise it is convenient to make a change state coordinates so that the system is in observable form. We elaborate on this point. Let (C,A) have observability indices $\ell_1,\_\_\_,\ell_p > 0$. This means that we can choose coordinates

$$(2.3) \qquad x = (x_{11},\_\_\_,x_{1\ell_1},\_\_\_,x_{p1},\_\_\_,x_{p\ell_p})^*$$

so that

$$(2.4) \qquad C_i A^{j-1} x = x_{ij} \; .$$

for $i = 1,\_\_\_,p$ and $1 \le j \le \ell_i$.

Relative to these coordinates we have the observable form

$$(2.5a) \qquad A = \left[\begin{array}{ccc|ccc|ccc} 0 & 1 & & & & & & & \\ & \ddots & & & 0 & & & 0 & \\ & & 1 & & & & & & \\ * & \cdots & * & * & \cdots & * & * & \cdots & * \\ \hline & & & & \ddots & & & & \\ \hline & & & & & & 0 & 1 & \\ & 0 & & & 0 & & & \ddots & \\ & & & & & & & & 1 \\ * & * & * & * & \cdots & * & * & \cdots & * \end{array}\right]$$

$$(2.5c) \qquad C = \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 0 & & 0 & 0 & & 0 \\ \hline & 0 & & & \ddots & & & 0 & \\ \hline 0 & & 0 & 0 & & 0 & 1 & 0 & 0 \end{array}\right]$$

The p diagonal blocks of A and C are $\ell_i \times \ell_i$ and $1 \times \ell_i$ respectively.

We have chosen the coordinates (2.4) because they are convenient for the describing which noises are small. The <u>small noise indices</u> $k_1 \ge \ldots \ge k_p, \ge 0$ are integers satisfying $0 \le k_i \le \ell_i$. If $k_i > 0$ we assume that there is small observation noise in the differential equation for y_i and small driving noise in the differential equations for $x_{ij}, j=1,\_\_\_,k_i-1$. Just how small is "small" depends on k_i and j. In particular we assume that

(2.5b) $$B = \left[\begin{array}{c|c|c} \begin{matrix} \epsilon^{k_i-1} & & & \\ & \ddots & & \\ & & \epsilon^{0} & \\ & & & 1 \\ & & & & 1 \end{matrix} & 0 & 0 \\ \hline 0 & \ddots & 0 \\ \hline 0 & 0 & \begin{matrix} \epsilon^{k_p-1} & & & \\ & \ddots & & \\ & & \epsilon^{0} & \\ & & & 1 \\ & & & & 1 \end{matrix} \end{array}\right]$$

(2.5d) $$D = \begin{bmatrix} \varepsilon^{k_1} & & 0 \\ & \ddots & \\ 0 & & \varepsilon^{k_p} \end{bmatrix}$$

We subdivide the x and y vectors into <u>fast</u> and <u>slow</u> <u>components</u>,

(2.6a) $$x_f = (x_{11},\ldots,x_{1k_1},\ldots,x_{p1},\ldots,x_{pk_p})^*$$

(2.6b) $$x_s = (x_{1k_1+1},\ldots,x_{1\ell_1},\ldots,x_{p,k_p+1},\ldots,x_{p,\ell_p})^*.$$

The dimension of x_f is $n_f = k_1 + \_\_\_ + k_p$ and the dimension of x_s is $n_s = n-n_f$.

Suppose $k_i > 0$ if $i \leq q$ and $k_i = 0$ if $i > q$ then

(2.6c) $$y_f = (y_1,\ldots,y_q)$$

(2.6d) $$y_s = (y_{q+1},\ldots,y_p)$$

We make similar decompositions of the noises, $w = (w_f,w_s)$ and $v = (v_f,v_s)$.

Instead of fast and slow, one could call these the noiseless and noisy components of x and y. The ε scalings on the noises (2.5b,d) have been chosen so that dynamics of the errors in estimating x_f have a time constant of order ε. The errors in estimating x_s have a time constant of order 1. This explains the fast/slow terminologv.

Some of the driving noises may actually be smaller than indicated (e.g., Q and/or R might depend on ε). The sizes of the small noises are fixed by the small noise indices. Unless these can be increased, there is no change in the order of magnitude of the error or the speed of the filter. Of course one could consider problems involving more than one small parameter but to keep things reasonably simple we shall not do so.

It is convenient to put the model (2.1) in block form with respect to the fast and slow variables

$$\begin{bmatrix} dx_f \\ dx_s \end{bmatrix} = \begin{bmatrix} A_{ff} & A_{fs} \\ A_{sf} & A_{ss} \end{bmatrix} \begin{bmatrix} x_f \\ x_s \end{bmatrix} dt + \begin{bmatrix} B_{ff} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dw_f \\ dw_s \end{bmatrix} \tag{2.7a}$$

$$\begin{bmatrix} dy_f \\ dy_s \end{bmatrix} = \begin{bmatrix} C_{ff} & 0 \\ 0 & C_{ss} \end{bmatrix} \begin{bmatrix} x_f \\ x_s \end{bmatrix} dt + \begin{bmatrix} D_{ff} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dv_f \\ dv_s \end{bmatrix} \tag{2.7b}$$

$$\begin{bmatrix} x_f(0) \\ x_s(0) \end{bmatrix} \sim N\left(0, \begin{bmatrix} P_{ff}(0) & P_{fs}(0) \\ P_{sf}(0) & P_{ss}(0) \end{bmatrix}\right) \tag{2.7c}$$

Because of our earlier assumptions, (C_{ff}, A_{ff}) is an observable pair in dual Brunovsky form with observability indices $k_1, \_\_\_, k_q$. The matrices C_{fs} and C_{sf} are zero and B_{ff} and D_{ff} depend on ε.

To analyze the asymptotic behaviour as ε goes to zero of the filtering equations for (2.7), we rescale variables

(2.8a) $$\psi_i = \begin{cases} y_i/\varepsilon^{k_i+1/2} & 1 \le i \le q \\ y_i & q < i \le p \end{cases}$$

(2.8b) $$\xi_{ij} = \begin{cases} x_{ij}/\varepsilon^{k_i-j+1/2} & 1 \le j \le k_i \\ x_{ij} & k_i < j \le \ell_i \end{cases}$$

We define the fast and slow parts of ψ and ξ as in (2.6), then (2.7) becomes

(2.10a) $$\begin{bmatrix} d\xi_f \\ d\xi_s \end{bmatrix} = \begin{bmatrix} \varepsilon^{-1}A_{ff} & \varepsilon^{-1/2}A_{fs} \\ \mathcal{O}(\varepsilon^{1/2}) & A_{ss} \end{bmatrix} \begin{bmatrix} \xi_f \\ \xi_s \end{bmatrix} dt + \begin{bmatrix} \varepsilon^{-1/2}I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dw_f \\ dw_s \end{bmatrix}$$

(2.10b) $$\begin{bmatrix} d\psi_f \\ d\psi_s \end{bmatrix} = \begin{bmatrix} \varepsilon^{-1}C_{ff} & 0 \\ 0 & C_{ss} \end{bmatrix} \begin{bmatrix} \xi_f \\ \xi_s \end{bmatrix} dt + \begin{bmatrix} \varepsilon^{-1/2}I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dv_f \\ dv_s \end{bmatrix}$$

The Kalman-Bucy filter [3] for (2.10) is given by

(2.11a) $$\begin{bmatrix} d\hat{\xi}_f \\ d\hat{\xi}_s \end{bmatrix} = \begin{bmatrix} \varepsilon^{-1}A_{ff} & \varepsilon^{-1/2}A_{fs} \\ \mathcal{O}(\varepsilon^{1/2}) & A_{ss} \end{bmatrix} \begin{bmatrix} \hat{\xi}_f \\ \hat{\xi}_s \end{bmatrix} dt + \begin{bmatrix} K_{ff} & K_{fs} \\ K_{sf} & K_{ss} \end{bmatrix} \begin{bmatrix} d\psi_f - \varepsilon^{-1}C_{ff}\,\hat{\xi}_f\,dt \\ d\psi_f - C_{ss}\,\hat{\xi}_s\,dt \end{bmatrix}$$

where the filter gain K is given by

$$(2.11b)\quad \begin{bmatrix} K_{ff} & K_{fs} \\ K_{sf} & K_{ss} \end{bmatrix} = \begin{bmatrix} \Pi_{ff} & \Pi_{fs} \\ \Pi_{sf} & \Pi_{ss} \end{bmatrix} \begin{bmatrix} \varepsilon^{-1} C_{ff}^{*} & 0 \\ 0 & C_{ss}^{*} \end{bmatrix} \begin{bmatrix} \varepsilon^{-1} R_{ff} & \varepsilon^{-1/2} R_{fs} \\ \varepsilon^{-1/2} R_{sf} & R_{ss} \end{bmatrix}^{-1}$$

and the covariance matrix Π satisfies the Riccati differential equation

$$(2.11c)\quad \begin{bmatrix} d\Pi_{ff} & d\Pi_{fs} \\ d\Pi_{sf} & d\Pi_{ss} \end{bmatrix} = \left\{ \begin{bmatrix} \varepsilon^{-1} A_{ff} & \varepsilon^{-1/2} A_{fs} \\ \mathcal{O}(\varepsilon^{1/2}) & A_{ss} \end{bmatrix} \begin{bmatrix} \Pi_{ff} & \Pi_{fs} \\ \Pi_{sf} & \Pi_{ss} \end{bmatrix} + \right.$$

$$\begin{bmatrix} \Pi_{ff} & \Pi_{fs} \\ \Pi_{sf} & \Pi_{ss} \end{bmatrix} \begin{bmatrix} \varepsilon^{-1} A_{ff}^{*} & \mathcal{O}(\varepsilon^{1/2}) \\ \varepsilon^{-1/2} A_{fs}^{*} & A_{ss}^{*} \end{bmatrix} + \begin{bmatrix} \varepsilon^{-1} Q_{ff} & \varepsilon^{-1/2} Q_{fs} \\ \varepsilon^{-1/2} Q_{sf} & Q_{ss} \end{bmatrix}$$

$$\left. - \begin{bmatrix} \Pi_{ff} & \Pi_{fs} \\ \Pi_{sf} & \Pi_{ss} \end{bmatrix} \begin{bmatrix} \varepsilon^{-1} C_{ff}^{*} & 0 \\ 0 & C_{ss}^{*} \end{bmatrix} \begin{bmatrix} \varepsilon^{-1} R_{ff} & \varepsilon^{-1/2} R_{fs} \\ \varepsilon^{-1/2} R_{sf} & R_{ss} \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon^{-1} C_{ff} & 0 \\ 0 & C_{ss} \end{bmatrix} \begin{bmatrix} \Pi_{ff} & \Pi_{fs} \\ \Pi_{sf} & \Pi_{ss} \end{bmatrix} \right\} dt$$

We apologize to our gentle readers for displaying such an equation in polite company. However the asymptotic solution of (2.13) is relatively straightforward.

We assume the Π has a series expansion in half-integer powers of ε

$$(2.12)\qquad \Pi = \Pi(\varepsilon) = \Pi^0 + \varepsilon^{1/2}\Pi^1 + \varepsilon\Pi^2 + \mathcal{O}(\varepsilon^{3/2})$$

If we plug this into (2.13) and collect the terms associated to like powers of ε, we obtain a series of simpler equations. For example collecting terms in ε^{-1}, we obtain in the $f-f$ block

(2.13a) $$0 = \varepsilon^{-1} \{A_{ff} \Pi^0_{ff} + \Pi^0_{ff} A^*_{ff} + Q_{ff}$$

$$\Pi^0_{ff} C^*_{ff} \overline{R}_{ff}^{-1} C_{ff} \Pi^0_{ff}\} dt$$

and in the $f - s$ block

(2.13b) $$0 = \varepsilon^{-1} \tilde{A}^0_{ff} \Pi^0_{fs} dt$$

where

(2.13c) $$\overline{R}_{ff} = R_{ff} - R_{fs} R^{-1}_{ss} R_{sf}$$

(2.14) $$\tilde{A}^0_{ff} = A^0_{ff} - \Pi^0_{ff} C^*_{ff} \overline{R}_{ff}^{-1} C_{ff}$$

The first equation (2.13a) is an algebraic Riccati equation for Π^0_{ff}. This is the ε^{-1} part of the Riccati equation for a reduced order filtering problem

(2.14a) $$d\xi_f = \varepsilon^{-1} A_{ff} \xi_f dt + \varepsilon^{-1/2} \overline{B}_f dw_f$$

(2.14b) $$d\psi_f = \varepsilon^{-1} C_{ff} \xi_f dt + \varepsilon^{-1/2} d\overline{v}_f$$

where w_f is as before and $\overline{v}_f$ is a q dimensional Wiener process with covariance $\overline{R}_{ff}(t\wedge s)$.

If we use Π^0_{ff} to define the filter gain, then the resulting filter is asymptotically optimal for the reduced problem (2.14a,b) and is given by

(2.14c) $$d\hat{\xi}_f = \varepsilon^{-1} \tilde{A}^0_{ff} \hat{\xi}_f dt + \Pi^0_{ff} C^*_{ff} \overline{R}^{-1}_{ff} d\psi$$

This filter is stable since (C_{ff}, A_{ff}) is an observable pair. In particular this implies that $\tilde{A}^0_{ff}$ is invertible. From (2.13b) we see that

(2.15) $$\Pi^0_{fs} = \Pi^{0*}_{sf} = 0$$

Now we look at some equations obtained from (2.11c) by collecting terms in $\varepsilon^{-1/2}$. The f-f block yields

(2.16a) $$0 = \varepsilon^{-1/2} \{\tilde{A}^0_{ff} \Pi^1_{ff} + \Pi^1_{ff} \tilde{A}^{0*}_{ff}\}$$

and the f-s block yields

(2.16b) $$0 = \varepsilon^{-1/2} \{\tilde{A}^0_{ff} \Pi^1_{fs} + Q_{fs} + A_{fs} \Pi^0_{ss}\}$$

The invertibility of $\tilde{A}^0_{ff}$ and (2.16) imply that

(2.17a) $$\Pi^1_{ff} = 0$$

(2.17b) $$\Pi^1_{fs} = \Pi^{1*}_{sf} = -(\tilde{A}^0_{ff})^{-1}\{Q_{fs} + A_{fs} \Pi^0_{ss}\}$$

Next we look at an equation obtained from (2.11c) by collecting terms in ε^0. The s-s block yields

(2.18a) $$\begin{aligned} d\Pi^0_{ss} = \{ & A_{ss} \Pi^0_{ss} + \Pi^0_{ss} A^*_{ss} + Q_{ss} \\ & - \Pi^1_{sf} C^*_{ff} \overline{R}_{ff}^{-1} C_{ff} \Pi^1_{fs} - \Pi^0_{ss} C^*_{ss} \overline{R}_{ss}^{-1} C_{ss} \Pi^0_{ss} \\ & - \Pi^1_{sf} C^*_{ff} M C_{ss} \Pi^0_{ss} - \Pi^0_{ss} C^*_{ss} M^* C_{ff} \Pi^1_{fs}\}\, dt \end{aligned}$$

where

(2.18b) $$\overline{R}_{ss} = R_{ss} - R_{sf} R_{ff}^{-1} R_{fs}$$

and

(2.18c) $$M = -R_{ff}^{-1} R_{fs} \overline{R}_{ss}^{-1}$$

These quantities and $\overline{R}_{ff}$ arise because

(2.18d)
$$\begin{bmatrix} \varepsilon^{-1}R_{ff} & \varepsilon^{-1/2}R_{fs} \\ \varepsilon^{-1/2}R_{sf} & R_{ss} \end{bmatrix}^{-1} = \begin{bmatrix} \varepsilon\overline{R}_{ff}^{-1} & \varepsilon^{1/2}M \\ \varepsilon^{1/2}M^{*} & \overline{R}_{ss}^{-1} \end{bmatrix}$$

By substituting (2.17b) into (2.18a) we obtain a Riccati differential equation for $\Pi^{0}_{ss}(t)$. The initial condition is $\Pi^{0}_{ss}(0) = P_{ss}(0)$ of (2.7c). We solve this for $\Pi^{0}_{ss}(t)$ and then find $\Pi^{1}_{fs}(t)$ from (2.17b). These and Π^{0}_{ff} yield the lowest order terms in the asymptotic expansion of the filter gain K (2.11b),

(2.19)
$$\begin{bmatrix} K_{ff} & K_{fs} \\ K_{sf} & K_{ss} \end{bmatrix} =$$

$$\begin{bmatrix} \Pi^{0}_{ff}C^{*}_{ff}\overline{R}_{ff}^{-1}+\mathcal{O}(\varepsilon) & \varepsilon^{-1/2}\Pi_{ff}C^{*}_{ff}M+\varepsilon^{1/2}\Pi^{1}_{fs}C^{*}_{ss}\overline{R}_{ss}^{-1}+\mathcal{O}(\varepsilon) \\ \varepsilon^{1/2}(\Pi^{1}_{sf}C^{*}_{ff}\overline{R}_{ff}^{-1}+\Pi^{0}_{ss}C^{*}_{ss}M^{*})+\mathcal{O}(\varepsilon) & \Pi^{1}_{sf}C^{*}_{ff}M+\Pi^{0}_{ss}C^{*}_{ss}\overline{R}_{ss}^{-1}+\mathcal{O}(\varepsilon^{1/2}) \end{bmatrix}$$

Now suppose we filter (2.10) but instead of using the optimal filter gain K in (2.11a) we use $\overline{K}$, the lowest order nonzero terms in the asymptotic expansion (2.19)

(2.20a)
$$\begin{bmatrix} \overline{K}_{ff} & \overline{K}_{fs} \\ \overline{K}_{sf} & \overline{K}_{ss} \end{bmatrix} =$$

$$\begin{bmatrix} \Pi^{0}_{ff}\ C^{*}_{ff}\ \overline{R}_{ff}^{-1} & \varepsilon^{-1/2}\Pi_{ff}\ C^{*}_{ff}\ M \\ \varepsilon^{1/2}(\Pi^{1}_{sf}\ C^{*}_{ff}\ \overline{R}_{ff} + \Pi^{0}_{ss}\ C^{*}_{ss}\ M^{*}) & \Pi^{1}_{sf}\ C^{*}_{ff}\ M + \Pi^{0}_{ss}\ C^{*}_{ss}\ \overline{R}_{ss}^{-1} \end{bmatrix}$$

The resulting filter is not optimal but it is asymptotically optimal. This follows because the solution of the Lyapunov equation for the error covariance for the filter with $\overline{K}$ gain has the same lowest order nonzero terms as the covariance of the filter with gain K, i.e., both are of the form

$$(2.20b)\quad \begin{bmatrix} \Pi^0_{ff} + \mathcal{O}(\varepsilon) & \varepsilon^{1/2}\,\Pi^1_{sf}(t) + \mathcal{O}(\varepsilon) \\ \varepsilon^{1/2}\,\Pi^1_{sf}(t) + \mathcal{O}(\varepsilon) & \Pi^0_{ss}(t) + \mathcal{O}(\varepsilon^{1/2}) \end{bmatrix}$$

Note that the f-f block of (2.20b) is up to $O(\varepsilon)$ the same as the error covariance of the reduced order filter (2.14c). Hence this reduced order filter is asymptotically optimal not only for the reduced order problem (2.14) but also for the fast states of the full order problem (2.10). In other words the optimal filter for the fast states asymptotically decouples from the slow state estimates. Moreover this reduced order filter is autonomous to order ε and so is the f-f error covariance.

The full order filter can be made autonomous in the standard fashion. As $t \to \infty$, $\Pi^1_{fs}(t)$ and $\Pi^0_{ss}(t)$ converges to the constant solutions $\Pi^1_{fs}(\infty)$ and $\Pi^0_{ss}(\infty)$ of (2.17a) and (2.18a). In particular $\Pi^0_{ss}(\infty)$ is the unique positive definite solution obtained by substituting (2.17a) into the left side of (2.18a) and equating it to zero. From $\Pi^0_{ss}(\infty)$ and (2.17a) we obtain $\Pi^1_{fs}(\infty)$.

Next we look at the asymptotic behaviour of the eigenvalues of the filter dynamics(2.20). It is not hard to see that as $\varepsilon \to 0$, k of these eigenvalues go off to infiity like ε^{-1}. They are asymptotic to the eigenvalues of

$$(2.21b)\qquad \varepsilon^{-1}\,\tilde{A}^0_{ff}$$

The corresponding left eigenvectors converge asymptotically to the fast state coordinate functions. For the autonomous filter the remaining n-k eigenvalues remain finite and converge to the eigenvalues of

$$(2.22b)\qquad \tilde{A}_{ss}(\infty) - \Pi^1_{sf}(\infty)C^*_{ff}(\tilde{A}^0_{ff})^{-1}A_{fs}$$

where

$$(2.22c)\qquad \tilde{A}^0_{ss}(t) = A_{ss} - \Pi^0_{ss}(t)\,C^*_{ss}\,C_{ss}\ .$$

We return to the x-y coordinates of (2.3), (2.4) and (2.5).

If $1 \le i \le q$ and $1 \le j \le k_i$ then

$$(2.23a)\qquad d\hat{x}_{ij} = \hat{x}_{ij+1}\,dt + \sum_{1\le r\le q} \varepsilon^{k_i - k_r - j}\, \pi^0_{ij,ri}\,(dy_r - \hat{x}_{r1}\,dt) + \varepsilon^{k_i - j + 1} \sum_{q<r\le p} \pi^1_{ij,r1}\,(dy_r - \hat{x}_{r1}dt)$$

The symbol $\pi^0_{ij,r1}$ denotes the corresponding scalar entry of Π^0 in distinction to Π^0_{ff} which is a submatrix of Π^0. If $1 \le i \le p$ and $k_i < j \le \ell_i$ then

$$(2.23b)\qquad d\hat{x}_{ij} = \sum_{r=1}^{p} \sum_{\rho=1}^{\ell_r} a_{ij,r\rho}\hat{x}_{r\rho}\,dt + \sum_{1\le r\le q} \varepsilon^{-k_r}\, \pi^1_{ij,1r}\,(d\psi_r - \hat{x}_{r1}\,dt) + \sum_{q<r\le p} \pi^0_{ij,ir}\,(d\psi_r - \hat{x}_{r1}\,dt)$$

In (2.23b) we have included those terms of order $\mathcal{O}(\varepsilon^{1/2})$ that we neglected in the development subsequent to (2.10b).

This filter is asymptotically optimal with an asymptotic error covariance P obtained by scaling Π in accordance with (2.8). In other words

$$(2.24a)\qquad p_{ij,r\rho} = \varepsilon^{k_i + k_r - j - \rho + 1}(\pi^0_{ij,r\rho} + \mathcal{O}(\varepsilon)) \text{ if } 1 \le j \le k_i \text{ and } 1 \le \rho \le k_r.$$

$$(2.24b)\qquad p_{ij,r\rho} = \varepsilon^{k_i - j + 1}(\pi^1_{ij,r\rho} + \mathcal{O}(\varepsilon^{1/2})) \text{ if } 1 \le j \le k_i \text{ and } k_r < \rho \le \ell_r.$$

$$(2.24c)\qquad p_{ij,r\rho} = \pi^o_{ij,r\rho} + \mathcal{O}(\varepsilon^{1/2}) \text{ if } k_i < j \le \ell_i \text{ and } k_r < \rho \le \ell_r.$$

Notice that error covariance (2.24a) of the fast coordinates goes to zero $\mathcal{O}(\varepsilon)$ or faster while that (2.24c) of the slow coordinates is $\mathcal{O}(1)$. The cross variance (2.24b) between the fast and slow variables goes to zero $\varepsilon^{1/2}$ faster than would be expected from (2.24a,c). This is because $\Pi^0_{fs} = 0$. In other words asymptotically speaking the fast and slow errors are orthogonal.

3. NONLINEAR FILTERING

Suppose we consider a nonlinear perturbation of (2.7) i.e.

$$\text{(3.1a)}\quad \begin{bmatrix} dx_f \\ dx_s \end{bmatrix} = \begin{bmatrix} A_{ff} & A_{fs} \\ A_{sf} & A_{ss} \end{bmatrix} \begin{bmatrix} x_f \\ x_s \end{bmatrix} dt + \begin{bmatrix} 0 \\ \alpha_s(x_f) \end{bmatrix} dt + \begin{bmatrix} B_{ff} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dw_f \\ dw_s \end{bmatrix}$$

$$\text{(3.1b)}\quad \begin{bmatrix} dy_f \\ dy_s \end{bmatrix} = \begin{bmatrix} C_{ff} & 0 \\ 0 & C_{ss} \end{bmatrix} \begin{bmatrix} x_f \\ x_s \end{bmatrix} dt + \begin{bmatrix} D_{ff} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dv_f \\ dv_s \end{bmatrix}$$

$$\text{(3.1c)}\quad \begin{bmatrix} x_f(0) \\ x_s(0) \end{bmatrix} \sim N(0, \begin{bmatrix} P_{ff}(0) & P_{fs}(0) \\ P_{sf}(0) & P_{ss}(0) \end{bmatrix})$$

Of course this (3.1) is a very specific form of the nonlinear filtering model. In a future paper [9] we discuss how we might be able to transform a general problem into this form.

A closely related filtering model is

(3.2a) $$\begin{bmatrix} dx_f \\ dx_s \end{bmatrix} = \begin{bmatrix} A_{ff} & A_{fs} \\ A_{sf} & A_{ss} \end{bmatrix} \begin{bmatrix} x_f \\ x_s \end{bmatrix} dt + \begin{bmatrix} 0 \\ \alpha_s(\bar{x}_f) \end{bmatrix} dt$$

$$\begin{bmatrix} B_{ff} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dw_f^0 \\ dw_s^0 \end{bmatrix}$$

(3.2b) $$\begin{bmatrix} dy_f \\ dy_s \end{bmatrix} = \begin{bmatrix} C_{ff} & 0 \\ 0 & C_{ss} \end{bmatrix} \begin{bmatrix} x_f \\ x_s \end{bmatrix} dt + \begin{bmatrix} D_{ff} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} dv_f^0 \\ dv_s^0 \end{bmatrix}$$

(3.2c) $$\begin{bmatrix} x_f \\ x_s \end{bmatrix} \sim N\left(0, \begin{bmatrix} P_{ff}(0) & P_{fs}(0) \\ P_{sf}0 & P_{ss}(0) \end{bmatrix}\right)$$

where $\bar{x}_f$, $\bar{x}_s$ are defined by

(3.3a) $$\begin{bmatrix} d\bar{x}_f \\ d\bar{x}_s \end{bmatrix} = \begin{bmatrix} A_{ff} & A_{fs} \\ A_{sf} & A_{ss} \end{bmatrix} \begin{bmatrix} \bar{x}_f \\ \bar{x}_s \end{bmatrix} dt + \begin{bmatrix} 0 \\ \alpha_s(\bar{x}_f) \end{bmatrix} dt$$

$$+ \begin{bmatrix} L_{ff} & L_{fs} \\ L_{sf} & L_{ss} \end{bmatrix} \begin{bmatrix} dy_f - C_{ff}\hat{x}_f\, dt \\ dy_s - C_{ss}\hat{x}_s\, dt \end{bmatrix}$$

(3.3b) $$\begin{bmatrix} \bar{x}_f(0) \\ \bar{x}_s(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The matrix L is the gain of the Kalman-Bucy filter for (3.2a,b,c) when $\alpha_s = 0$.

Even when $\alpha_s \neq 0$, (3.2a,b,c) is a linear filtering problem because the nonlinearity $\alpha_s(\bar{x}_f)$ is a functional of the observations given by (3.2b) and

$(\bar{x}_f, \bar{x}_s)$ is the optimal estimate of the state of (3.2). The error covariance is not affected by $\alpha_s(\bar{x}_f)$.

We would like to study the relationship between (3.1) and (3.2, 3.3). Notice that the driving and observation noises are different in (3.1) and (3.2), although we assume that w and w^0 have the same distribution as does v and v^0.

Let P be the measure on the spaces of paths $\{w^0(t), v^0(t)\}$ under which they are independent Wiener processes with the desired covariances. We define $w(t)$ and $v(t)$ as functionals of $w^0(t)$ and $v^0(t)$ by the stochastic differential equations

(3.4a) $$dw_f = dw_f^0 \quad , \quad w_f(0) = w_f^0(0) = 0$$

(3.4b) $$dw_s = dw_s^0 + \left(\alpha_s(\bar{x}_f(t)-\alpha_s(x_f(t))\right)dt \quad , \quad w_s(0) = w_s^0(0) = 0$$

(3.4c) $$dv = dv^0 \quad , \quad v(0) = v^0(0) = 0.$$

where $x_f(t)$ and $\bar{x}_f(t)$ are the functionals of w^0, v^0 defined by (3.2) and (3.3).

The process $(w(t), v(t))$ defined by (3.4) is not Wiener under the measure P. But using the theorem of Girsanov [4], one can define a new probability measure P which is a Wiener measure on the space of paths $(w(t), v(t))$ and which is absolutely continuous with respect to P^0. This measure is defined by its Radon-Nikodym derivative with respect to P^0

(3.5) $$\Lambda = \frac{d\mathcal{P}}{d\mathcal{P}^0}$$

If $\mathcal{F}_t$ is the σ-algebra generated by the $\{x(s),y(s):\ 0 \le s \le t\}$ of (3.2) and (3.3) then $\Lambda(t)$ defined by

(3.6) $$\Lambda(t) = E^0(\Lambda|\mathcal{F}_t)$$

is an $\mathcal{F}_t$ - martingale. (E^0 denotes expectation with respect to $\mathcal{P}^0$ and E denotes expectation with respect to $\mathcal{P}$). Moreover $\Lambda(t)$ satisfies the stochastic differential equation.

(3.7a) $$d\Lambda(t) = \Lambda(t)\beta^*(t)dw_s^0(t) \quad , \quad \Lambda(0) = 1$$

where

(3.7b) $$\beta(t) = \alpha_s(x(t)) - \alpha_s(\bar{x}(t)).$$

Hence by using the linear model and filter (3.2) and (3.3) and Λ_t, we can construct the nonlinear model (3.1). If $\phi(x)$ is any smooth function of x, we denote by $\hat{\phi}(t)$ the conditional mean of $\phi(x(t))$ with respect to $\mathcal{P}$ given the past observations of the nonlinear model (3.1).

We denote by $\bar{\phi}(t)$ the conditional mean of $\phi(x(t)$ with respect to $\mathcal{P}^0$ given the past observations of the linear model (3.2) (3.3). Then

(3.8) $$\phi(t) + E(\phi(x_t)|\mathcal{Y}_t) = E^0(\phi(x_t)\Lambda(t)|\mathcal{Y}_t)|E^0(\Lambda(t)|\mathcal{Y}_t)$$

$$= \overline{\phi(x_t)\Lambda(t)}/\overline{\Lambda(t)} \ .$$

$\mathcal{F}_t$ is the σ-algebra generated by the past observation $\{y(s):\ 0 \le s \le t\}$ of either model.

Lemma $\quad E^0(\Lambda_t| \quad _t) = \overline{\Lambda}_t = 1 \quad$ a.s.

Proof We use a technique of E. Wong [11, p. 269]. $\overline{\Lambda}_t$ is a $\mathcal{Y}_t$ local martingale so by the martingale representative theorem,

(3.9) $\quad d\overline{\Lambda}(t) = \eta(t)dy(t)$

where $\eta(t)$ is a $\mathcal{Y}_t$ adapted process of dimension 1xp. But

$$E^0(d(\Lambda(t)y_i(t)|\mathcal{Y}_t) = E^0(d(\overline{\Lambda}(t)y_i(t)|\mathcal{Y}_t)$$

Using the martingale differential rule and (3.7) we obtain

(3.10)
$$E^0(d\Lambda(t)y_i(t)|\mathcal{Y}_t) = E^0(\Lambda(t)dy_i(t)$$
$$+ \quad y_i(t)\Lambda(t)\,\beta^*(t)dw_s^0(t) + \Lambda_t\,\beta(t)^*\, d\langle w_s^0\,(t),y_i(t)\rangle|\mathcal{Y}_t)$$
$$= \quad \overline{\Lambda}(t)C_i\overline{x}(t)dt$$

where $\langle w_s^0(t),y_i(t)\rangle$ denotes the quadratic variation and C_i is the i^{th} row in (2.5c). In a similar fashion from (3.9) we obtain

(3.11)E
$$E^0(d(\overline{\Lambda}(t)y_i(t)|\mathcal{Y}_t) = E^0(\overline{\Lambda}(t)dy_i(t)$$
$$+ y_i(t)\eta(t)dy(t) + \eta(t)d\langle y(t);y(t)\rangle|\mathcal{Y}_t)$$
$$= \overline{\Lambda}(t)\; C_i\overline{x}(t)dt + y_i(t)\eta(t)C\overline{x}(t)$$
$$+ \eta(t)DRD_i^*$$

where D and its $i^{\underline{th}}$ row D_i are from (2.5d) and $R(t\wedge s)$ is the covariance of the observation noise. Equating (3.10) and (3.11) we have

$$\eta(t)\ (C\overline{x}(t)\ y_i(t) + DRD_i^*) = 0$$

for $i = 1,\_\_\_,p$. But clearly this implies that $\eta(t) = 0$ a.s. QED.

In light of the above lemma we have a simplification of (3.8)

$$(3.12)\qquad \hat{\phi}(t) = E(\phi(x_t)|\mathcal{Y}_t) = E^0(\phi(x_t)\Lambda(t)|\mathcal{Y}_t) = \overline{\phi(t)\Lambda(t)}\ .$$

The differential equation (3.7) has the solution

$$(3.13)\qquad \Lambda(t) = \exp(\int_0^t \beta^*(\tau)dw_s^0(\tau) - \frac{1}{2}\beta^*(\tau)\beta(\tau)d_\tau)$$

where $\beta(t)$ is given by (3.7b). There are several ways to analyze (3.13) as $\varepsilon \to 0$ including the method of large deviations [2]. However for simplicity we shall use only the most straightforward approach. Recall that $x_f(t)$ and $\overline{x}_f(t)$ of (3.7b) are from the linear model and filter (3.2) and (3.3). Hence the conditional (and unconditional) distribution of $x_f(t) - \overline{x}_f(t)$ is a zero mean Gaussian with variance P_{ff} that goes to zero with ε according to the asymptotic expansion (2.24a). If $\alpha_s(x_f)$ is a smooth function of x_f satisfying a Lipschitz condition like

$$|\alpha_s(x_f^1) - \alpha_s(x_f^2)\ | \ \leq L\ |\ x_f^1 - x_f^2\ |$$

then $\beta(t)$ is approximately a zero mean Gaussian with variance $\mathcal{O}(\varepsilon L^2)$. The size of second derivative of α and ε determine how much $\beta(t)$ differs from a Gaussian.

In any case by a standard result [2, p.44] we see that

$$\mathcal{P}\{\lim_{\varepsilon \to 0} \max_{0 \le t \le T} |\Lambda_t - 1| = 0\} = 1 \tag{3.14}$$

This justifies using the linear filter (3.3) to filter the nonlinear model (3.1).

REFERENCES

[3] Gelb, A. (ed.), Applied Optimal Estimation (MIT Press, Cambridge, 1974).

[5] Haddad, A. H., Linear filtering of singularly perturbed systems, IEEE Trans. Auto. Contr. AC-31 (1976) 515-519.

[8] Katzur, R. Bobrovsky, B. Z. and Schuss, Z., Asymptotic analysis of the optimal filtering problem for one-dimensional diffussions measured in low noise channels, SIAM J. App. Math. 44 (1984) Part I:591-604, Part II:1176-1191.

[6] Hijab, O., Minimum energy estimation, Ph.D. Thesis, Department of Mathematics, University of California, Berkeley (1980).

[7] Hijab, O., Asymptotic Bayesian estimation of a first order equation with small diffusion, Ann. Prob. 12(1984) 890-902.

[1] Blankenship, G. L., Asymptotic methods in nonlinear estimation theory, Electrical Engineering Department, University of Maryland, (1983).

[10] Picard, J., Nonlinear filtering of one-dimensional diffusion in the case of a high signal-to-noise ratio, INRIA Centre Sophia Antipolis, (1985).

[2] Freidlin, M. I. and Wentzell, A. D., Random Perturbations of Dynamical Systems (Springer-Verlag, New York, 1984).

[4] Girsanov, I. V., On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Th. of Prob. and Appl. 5(1960) 285-301.

[4] Wong, E., Recent progress in stochastic processes - a survey, IEEE Tran. Inf. Th. IT-19(1973) 262-275.

[9] Krener, A. J., The geometric-asymptotic approach to nonlinear filtering, in preparation.