Communications in Information and Systems
Vol. 6, No. 3, pp. 203-220, 2006

004

# A NEW NONLINEAR FILTER

ROBERT J. ELLIOTT* AND SIMON HAYKIN†

**Abstract.** A discrete time filter is constructed where both the observation and signal process have non-linear dynamics with additive white Gaussian noise. Using the reference probably framework a convolution Zakai equation is obtained which updates the unnormalized conditional density. Our work obtains approximate solutions of this equation in terms of Gaussian sum when second order expansions are introduced for the non-linear terms.

**Acknowledgements.** The authors thank NSERC for its continued support.

## 1. Introduction.

The most successful filter has without doubt been the Kalman filter. This considers noisy observations of a signal process where the dynamics are linear and the noise is additive and Gaussian. Extensions of the Kalman filter to cover non-linear dynamics were obtained by taking first order Taylor expansions of the non-linear terms about the current mean. The resulting filter is the so-called extended Kalman filter, or EKF.

A popular method in recent years has been the so-called particle filter approach. However, this is only based on Monte-Carlo simulation.

In this paper the reference probability framework is used to obtain a discrete time version of the Zakai equation. This looks like a convolution equation and it provides an update for the unnormalized conditional density of the state process given the observations. If first order Taylor series approximations are used for the non-linear terms in the signal and observation processes, the convolution equation can be solved explicitly and the extended Kalman filter re-derived. Taylor expansions of the non-linear terms to second order are then considered and approximate solutions in terms of Gaussian sums obtained.

Detailed proofs and numerical work will appear in [2].

## 2. Dynamics.

Consider non-linear signal and observation processes $x, y$ where the noise is additive and Gaussian. Suppose the processes are defined on $(\Omega, F, P)$ where $w = \{w_k,\, k = 0, 1, 2, \dots\}$, $v = \{v_k,\, k = 0, 1, 2, \dots\}$ are sequences of independent $N(0, I_n)$, resp. $N(0, I_m)$, random variables.

*Haskayne School of Business, University of Calgary, Calgary, Alberta, Canada, T2N 1N4, E-mail: relliott@ucalgary.ca

†Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada, E-mail: haykin@mcmaster.ca

Suppose for each $k = 0, 1, 2, \dots$

$$A_k : R^n \to R^n$$
$$C_k : R^n \to R^m.$$

are measurable functions.

We suppose the signal dynamics have the form:

$$x_{k+1} = A_k(x_k) + B_k w_{k+1}. \tag{2.1}$$

and the observation dynamics have the form

$$\text{Observation} \qquad y_k = C_k(x_k) + D_k v_k. \tag{2.2}$$

To simplify notation we suppose the coefficients are time independent.

Further we suppose that $B : R^n \to R^n$ and $D : R^m \to R^m$ are symmetric and non-singular.

**Measure Change.** We shall follow the methods of Elliott and Krishnamurthy [3] and show how the dynamics (2.1), (2.2) can be modelled starting with a reference probability $\overline{P}$.

Suppose on $(\Omega, F, \overline{P})$ we have two sequences of random variables

$$x = \{x_k,\ k = 0, 1, 2, \dots\}$$
$$y = \{y_k,\ k = 0, 1, 2, \dots\}.$$

We suppose that under $\overline{P}$ the $x_k$ are independent $n$-dimensional $N(0, I_n)$ random variables and the $y_k$ are independent $N(0, I_m)$ random variables. For $x \in R^n$, $y \in R^m$ write

$$\psi(x) = (2\pi)^{-n/2} \exp\Big(-\frac{x'x}{2}\Big)$$
$$\phi(y) = (2\pi)^{-n/2} \exp\Big(-\frac{y'y}{2}\Big).$$

Here the prime $'$ denotes transpose. Define the $\sigma$-fields

$$\mathcal{G} = \sigma\{x_0, x_1, \dots, x_k, y_0, y_1, \dots, y_k\}$$
$$\mathcal{Y} = \sigma\{y_0, y_1, \dots, y_k\}.$$

Then $\mathcal{G}_k$ represents the (possible) histories of $x$ and $y$ to time $k$ and $\mathcal{Y}_k$ represents the history of $y$ to time $k$.

For any square matrix $B$ write $|B|$ for its determinant. Write

$$\lambda_0 = \frac{\phi\big(D^{-1}(y_0 - C(x_0))\big)}{|D|\phi(y_0)}$$

and for $\ell \geq 1$

$$\lambda_\ell = \frac{\phi\big(D^{-1}(y_\ell - C(x_\ell))\big)}{|D|\phi(y_\ell)} \cdot \frac{\psi\big(B^{-1}(x_\ell - A(x_{\ell-1}))\big)}{|B|\psi(x_\ell)} .$$

For $k \geq 0$ define

$$\Lambda_k = \prod_{\ell=0}^{k} \lambda_\ell.$$

We can define a new probability measure $P$ on $|\Omega, V_k \mathcal{G}_k)$ by setting

$$\frac{dP}{d\overline{P}}\Big|_{\mathcal{G}_k} = \Lambda_k.$$

For $\ell = 0,1,2,\dots$ define $v_\ell := D^{-1}\big(y_\ell - C(x_\ell)\big)$. For $\ell = 1,2,\dots$ define $w_\ell := B^{-1}\big(x_\ell - A(x_{\ell-1})\big)$.

As in [3] we can then prove:

LEMMA 2.1. *Under the measure $P$*

$$v = \{v_\ell,\ \ell = 0,1,2,\dots\}$$
$$w = \{w_\ell,\ \ell = 1,2,\dots\}$$

*are sequences of independent $N(0, I_m)$ and $N(0, I_n)$ random variables, respectively.*

That is, under the measure $P$

$$x_\ell = A(x_{\ell-1}) + Bw_\ell$$
$$y_\ell = C(x_\ell) + Dv_\ell.$$

However, $\overline{P}$ is a nicer measure under which to work.

**3. Recursive Densities.** The filtering problem is concerned with the estimation of $x_k$, and its statistics, given the observations $y_0, y_1, \dots, y_k$, that is, given $\mathcal{Y}_k$. The problem would be completely solved if we could determine the conditional density $\gamma_k(x)$ of $x_k$ given $\mathcal{Y}_k$. That is,

$$\gamma_k(x)dx = P(x_k \in dx|\mathcal{Y}_k) = E[I(x_k \in dx)|\mathcal{Y}_k].$$

Using a form of Bayes Theorem, (see [1]), for any measurable function $g : R^n \to R$

$$E[g(x_k)|\mathcal{Y}_k] = \frac{\overline{E}[\Lambda_k g(x_k)|\mathcal{Y}_k]}{\overline{E}[\Lambda_k|\mathcal{Y}_k]}$$

where the expectations on the right are taken under the measure $\overline{P}$. The numerator $\overline{E}[\Lambda_k g(x_k)|\mathcal{Y}_k]$ is an unnormalized conditional expectation of $g(x_k)$ given $\mathcal{Y}_k$. The denominator is the special case of the numerator obtained by taking

$$g(\,\cdot\,) = 1.$$

The map $g \to \overline{E}[\Lambda_k g(x_k)|\mathcal{Y}_k]$ is a continuous linear functional and so defines a measure. We assume this measure has a density $\alpha_k(x)$ so that

$$\overline{E}[\Lambda_k g(x_k)|\mathcal{Y}_k] = \int_{R^n} g(x)\alpha_k(x)dx$$

and heuristically

$$\overline{E}[\Lambda_k I(x_k \in dx)|\mathcal{Y}_k] = \alpha_k(x)dx.$$

Then

$$E[g(x_k)|\mathcal{Y}_k] = \frac{\int_{R^n} g(x)\alpha_k(x)dx}{\int_{R^n} \alpha_k(x)dx}$$

and $\alpha_k(x)$ is an unnormalized conditional density of $x_k$ given $\mathcal{Y}_k$.

Modifying the proof in [3] we have:

THEOREM 3.1. *$\alpha_k$ satisfies the recursion:*

$$\alpha_k(x) = \frac{\phi\big(D^{-1}(y_k - C(x))\big)}{|B|\ |C|\phi(y_k)} \cdot \int_{R^n} \alpha_{k-1}(z)\psi\Big(B^{-1}\big(x - A(z)\big)\Big)dz. \tag{3.1}$$

*Proof.* See [3]. □

REMARK 3.2. This equation provides the recursion for the unnormalized conditional density of $x_k$ given $\mathcal{Y}_k$. It is a discrete time version of the Zakai equation. The problem is now to solve this equation.

As shown in the book [1] and the paper [3], when the dynamics are linear, $A(x) = Ax$ and $C(x) = Cx$, for suitable matrices $A$ and $C$, with linear Gaussian noise terms the integral in (3.1) can be evaluated and a Gaussian density obtained for $\alpha_k(x)$.

The extended Kalman filter is obtained below by linearizing $A$ and $C$ and solving (3.1). We then obtain approximate solutions when second and higher order terms are included in the approximations of $A(x)$ and $C(x)$.

**4. Notation.** Recall our model is

$$\begin{aligned} &\text{SIGNAL} && x_k = A(x_{k-1}) + Bw_k \in R^n \\ &\text{OBSERVATION} && y_k = C(x_k) + Dv_k \in R^m \end{aligned}$$

where $A$ and $C$ can be non-linear functions: $A : R^n \to R^n$, $C : R^n \to R^m$.

NOTATION 4.1. The transpose of any matrix (or vector) $M$ will be denoted by $M'$,

Write

$$\begin{aligned} A(x) &= \big(A_1(x), \dots, A_n(x)\big)' \\ C(x) &= \big(C_1(x), \dots, C_m(x)\big)'. \end{aligned}$$

If $A$ (resp. $C$) are (twice) differentiable we shall write

$$\nabla A = (\nabla A_1, \dots, \nabla A_n)' = \begin{pmatrix} \frac{\partial A_1}{\partial x_1}, & \frac{\partial A_1}{\partial x_2}, \dots, & \frac{\partial A_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial A_n}{\partial x_1}, & \dots & \frac{\partial A_n}{\partial x_n} \end{pmatrix},$$

$$\nabla C = (\nabla C_1, \dots, \nabla C_m)'.$$

$\nabla^2 A$ will denote the matrix of Hessians

$$\nabla^2 A = \begin{pmatrix} \nabla^2 A_1 \\ \nabla^2 A_2 \\ \vdots \\ \nabla^2 A_n \end{pmatrix}$$

where

$$\nabla^2 A_k = \begin{pmatrix} \frac{\partial^2 A_k}{\partial x_1^2} & \frac{\partial^2 A_k}{\partial x_1 \partial x_2} \cdots & \frac{\partial^2 A_k}{\partial x_1 \partial x_n} \\ \vdots & & \\ \frac{\partial^2 A_k}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 A_k}{\partial x_n^2} \end{pmatrix}.$$

Similarly,

$$\nabla^2 C = \begin{pmatrix} \nabla^2 C_1 \\ \vdots \\ \nabla^2 C_m \end{pmatrix}.$$

**Approximations 4.2.** For any $\mu \in R^n$ we can consider the Taylor expansions of $A$ and $C$ to second order:

$$\begin{aligned} A(z) \simeq A(\mu) &+ \nabla A(\mu) \cdot (z - \mu) \\ &+ \frac{1}{2} \nabla^2 A(\mu) \cdot (z - \mu)^2 \end{aligned} \tag{4.1}$$

$$\begin{aligned} C(x) \simeq C\big(A(\mu)\big) &+ \nabla C\big(A(\mu)\big) \cdot \big(x - A(\mu)\big) \\ &+ \frac{1}{2} \nabla^2 C\big(A(\mu)\big) \cdot \big(x - A(\mu)\big)^2. \end{aligned} \tag{4.2}$$

Here $\nabla A(\mu) \cdot (z - \mu) \in R^n$ and $\nabla^2 A(\mu) \cdot (z - \mu)^2$ denotes the vector

$$\big((z-\mu)'\nabla^2 A_1(\mu)(z-\mu), \dots, (z-\mu)'\nabla^2 A_n(\mu) \cdot (z-\mu)\big)'.$$

From the Gaussian densities involved in the recursion for $\alpha_k$ we shall be interested in scalar quantities of the form:

$$\frac{1}{2}\big(x - A(\mu) - \nabla A(\mu) \cdot (z - \mu)\big)' B^{-2} \nabla^2 A(\mu)(z - \mu)^2.$$

Write $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ for the vector $B^{-2}\big(x - A(\mu) - \nabla A(\mu)\cdot(z-\mu)\big)$. Then

$$
\begin{aligned}
\frac{1}{2}\,&\big(x - A(\mu) - \nabla A(\mu)\cdot(z-\mu)\big)' B^{-2}\nabla^2 A(\mu)\cdot(z-\mu)^2 \\
&= \frac{1}{2}\,\xi'\nabla^2 A(\mu)\cdot(z-\mu)^2 \\
&= \frac{1}{2}\,(\xi_1, \ldots, \xi_n)\begin{pmatrix}(z-\mu)'\nabla^2 A_1(\mu)(z-\mu)\\ \vdots \\ (z-\mu)'\nabla^2 A_n(\mu)(z-\mu)\end{pmatrix} \\
&= (z-\mu)' H(\mu)(z-\mu)
\end{aligned}
$$

where $H(\mu)$ is the symmetric matrix defined symbolically by

$$
\begin{aligned}
H(\mu) &= \frac{1}{2}\,\xi'\begin{pmatrix}\nabla^2 A_1(\mu)\\ \vdots \\ \nabla^2 A_n(\mu)\end{pmatrix} \\
&= \frac{1}{2}\,\big(\xi_1\nabla^2 A_1(\mu) + \xi_2\nabla^2 A_2(\mu)\cdots + \xi_n\nabla^2 A_n(\mu)\big).
\end{aligned}
\tag{4.3}
$$

Similarly, for $C$ $\nabla^2 C\big(A(\mu)\big)\cdot\big(x - A(\mu)\big)^2$ will denote the vector

$$
\Big(\big(x - A(\mu)\big)'\nabla^2 C_1\big(A(\mu)\big)\big(x - A(\mu)\big), \ldots, \big(x - A(\mu)\big)'\nabla^2 C_m\big(A(\mu)\big)\big(x - A(\mu)\big)\Big).
$$

In the exponential involving $C$ terms we shall consider scalar quantities of the form

$$
\frac{1}{2}\,\Big(y - C\big(A(\mu)\big)\Big)' D^{-2}\cdot\nabla^2 C\big(A(\mu)\big)\cdot\big(x - A(\mu)\big)^2.
$$

This can be written

$$
\big(x - A(\mu)\big)' Z^{-1}(\mu)\big(x - A(\mu)\big)
$$

where $Z^{-1}(\mu)$ is the symmetric matrix

$$
\frac{1}{2}\,\zeta'\begin{pmatrix}\nabla^2 C_1\big(A(\mu)\big)\\ \vdots \\ \nabla^2 C_m\big(A(\mu)\big)\end{pmatrix} = \frac{1}{2}\,\Big(\zeta_1\nabla^2 C_1\big(A(\mu)\big) + \zeta_2\nabla^2 C_2\big(A(\mu)\big) + \cdots + \zeta_m\nabla^2 C_m\big(A(\mu)\big)\Big)
\tag{4.4}
$$

with $\zeta = D^{-2}\big(y - C\big(A(\mu)\big)\big)$.

**5. The EKF.** We shall take the first two terms in the Taylor expansions of $A(x)$ and $C(x)$ and show how equation (3.1) can be solved to re-derive the EKF (Extended Kalman Filter).

Of course, if the dynamics are linear the calculations below show how the Kalman filter can be derived from (3.1).

**Linear Approximations.** SIGNAL The signal has dynamics

$$x_k = A(x_{k-1}) + Bw_k \in R^n.$$

If $\mu_{k-1}$ is the conditional mean determined at time $(k-1)$ we consider the first two terms in the Taylor expansion of $A(x_{k-1})$ about $\mu_{k-1}$ and write

$$x_k \simeq A(\mu_{k-1}) + \nabla A(\mu_{k-1})(x_{k-1} - \mu_{k-1}) + Bw_k. \tag{5.1}$$

Similarly for the

$$\text{OBSERVATION} \qquad y_k = C(x_k) + Dv_k \in R^m$$

we take the first two terms of the Taylor expansion about $A(\mu_{k-1})$ and write

$$y_k \simeq C\big(A(\mu_{k-1})\big) + \nabla C\big(A(\mu_{k-1})\big)\big(x_k - A(\mu_{k-1})\big) + Dv_k \in R^m. \tag{5.2}$$

We are supposing that the conditional density of $x_{k-1}$ given $\mathcal{Y}_{k-1}$ is $N\big(\mu_{k-1}, \Sigma_{k-1}\big)$, that is the (normalized) conditional density of $x_{k-1}$ given $\mathcal{Y}_{k-1}$ is

$$\big|\Sigma_{k-1}\big|^{-1/2}\psi\Big(\Sigma_{k-1}^{-1}(x - \mu_{k-1})\Big). \tag{5.3}$$

Therefore, $\alpha_{k-1}(z) \sim \psi\big(\Sigma_{k-1}^{-1}(z - \mu_{k-1})\big)$. Let us note that

$$\begin{aligned}
\widehat{x}_{k|k-1} :&= E[x_k|\mathcal{Y}_{k-1}] \\
&= E[A(x_{k-1}) + Bw_k]\mathcal{Y}_{k-1}] \\
&= E[A(x_{k-1})|\mathcal{Y}_{k-1}] \\
&\simeq E[A(\mu_{k-1}) + \nabla A(\mu_{k-1})(x_{k-1} - \mu_{k-1})|\mathcal{Y}_{k-1}] \\
&= A(\mu_{k-1})
\end{aligned} \tag{5.4}$$

as $\mu_{k-1} = E[x_{k-1}|\mathcal{Y}_{k-1}]$.

In the paper of Ito and Xiong [5] a more accurate approximation is given using quadrature for $E[A(x_{k-1})|\mathcal{Y}_{k-1}]$. This could be used rather than $A(\mu_{k-1})$. Also write

$$\begin{aligned}
\Sigma_{k|k-1} &= E[(x_k - \widehat{x}_{k|k-1})(x_k - \widehat{x}_{k|k-1})'|\mathcal{Y}_{k-1}] \\
&\simeq E\big[\big(A(x_{k-1}) + Bw_k - A(\mu_{k-1})\big)\big(A(x_{k-1} + Bw_k - A(\mu_{k-1})\big)'|\mathcal{Y}_{k-1}\big] \\
&\simeq E\big[\big(\nabla A(\mu_{k-1})(x_{k-1} - \mu_{k-1}) + Bw_k\big) \\
&\qquad \times \big(\nabla A(\mu_{k-1})(x_{k-1} - \mu_{k-1}) + Bw_k\big)'|\mathcal{Y}_{k-1}\big] \\
&= \nabla A(\mu_{k-1})\Sigma_{k-1}\nabla A(\mu_{k-1})' + B^2.
\end{aligned} \tag{5.5}$$

Recall the matrix inversion Lemma (MIL:)

LEMMA 5.1. *For suitable matrices*

$$(B + A\Sigma A')^{-1} = B^{-1} - B^{-1}A(\Sigma^{-1} + A'B^{-1}A)^{-1}A'B^{-1}.$$

COROLLARY 5.2.

$$\Sigma^{-1}_{k|k-1} = B^{-2} - B^{-2}\nabla A(\mu_{k-1})\big(\Sigma^{-1}_{k-1} + \nabla A(\mu_{k-1})'B^{-2}\nabla A(\mu_{k-1})\big)^{-1}\nabla A(\mu_{k-1})'B^{-2}.$$

Linear algebra and the MIL then enable us to obtain the following result.

THEOREM 5.3 (EKF). *Suppose*

$$\widehat{x}_{k-1|k-1} = E[x_{k-1}|\mathcal{Y}_{k-1}]$$

*is* $N(\mu_{k-1}\Sigma_{k-1})$. *Then, approximately,* $\widehat{x}_{k|k}$ *is* $N(\mu_k, \Sigma_k)$ *where*

$$\begin{aligned}\Sigma_k &= \Sigma_{k|k-1} - \Sigma_{k|k-1}\nabla C\big(A(\mu_{k-1})\big)' \\ &\quad \times \Big(D^2 + \nabla C\big(A(\mu_{k-1})\big)\Sigma_{k|k-1}\nabla C\big(A(\mu_{k-1})\big)\Big)^{-1}\nabla C\big(A(\mu_{k-1})\big)\Sigma_{k|k-1}\end{aligned} \tag{5.6}$$

*and*

$$\begin{aligned}\mu_k &= A(\mu_{k-1}) + \Sigma_{k|k-1}\nabla C\big(A(\mu_{k-1})\big)' \\ &\quad \times \Big(D^2 + \nabla C\big(A(\mu_{k-1})\big)\Sigma_{k|k-1}\nabla C\big(A(\mu_{k-1})\big)'\Big)^{-1}\Big(y_k - C\big(A(\mu_{k-1})\big)\Big).\end{aligned} \tag{5.7}$$

*Proof.* See [2]. □

**6. A New Nonlinear Filter.** We extend the EKF by including second order terms in the approximations for $A(x)$ and $C(x)$.

This has two consequences:

1) we can no longer explicitly evaluate the integral $I$. We approximate this using Gauss-Hermite quadrature rules after some algebra.
2) when the integral is multiplied by $\phi\big(D^{-1}(y_k - C(x))\big)$ we would like to obtain Gaussian densities. By modifying the algebra of Section 5 we are able to do this

**Expansions 6.1.** We recall the second order expansions (4.1) and (4.2)

$$A(z) \simeq A(\mu) + \nabla A(\mu)\cdot(z-\mu) + \frac{1}{2}\,\nabla^2 A(\mu)\cdot(z-\mu)^2 \tag{6.1}$$

$$C(x) \simeq C\big(A(\mu)\big) + \nabla C\big(A(\mu)\big)\cdot\big(x - A(\mu)\big) + \frac{1}{2}\,\nabla^2 C\big(A(\mu)\big)\cdot\big(x - A(\mu)\big)^2. \tag{6.2}$$

THEOREM 6.2. *Suppose* $\alpha_{k-1}(z)$ *is given by a weighted sum of Gaussian densities:*

$$\alpha_{k-1}(z) = \sum_{i=1}^{N(k-1)} \lambda_{k-1,i}\, \exp\Big(-\frac{1}{2}\,(x-\mu_{k-1,i})'\Sigma^{-1}_{k-1,i}(x-\mu_{k-1,i})\Big).$$

*Then*

$$\alpha_k(x) \simeq \sum_{i=1}^{N(k-1)} \sum_{j=1}^{m(k,i)} \lambda_{k,i,j} \, \exp\Big(-\frac{1}{2}\,(x-\mu_{k,i,j})\,\Sigma_{k,i}^{-1}(x-\mu_{k,i,j})\Big)$$

*where*

$$\begin{aligned}\Sigma_{k,i,j} &= \Gamma_{k,i,j} - \Gamma_{k,i,j}\nabla C\big(A(\mu_{k-1,i})\big)' \\ &\quad\times \Big(D^2 + \nabla C\big(A(\mu_{k-1})\big)\Gamma_{k,i,j}\nabla C\big(A(\mu_{k-1,i})\big)'\Big)^{-1}\nabla C\big(A(\mu_{k-1,i})\big)\Gamma_{k,i}\end{aligned}$$

*and*

$$\begin{aligned}\mu_{k,i,j} &= A(\mu_{k-1}) \\ &\quad+ \Gamma_{k,i,j}\nabla C\big(A(\mu_{k-1,i})\big)\Big(D^2 + \nabla C\big(A(\mu_{k-1,i})\big)\Gamma_{k,i,j}\nabla C\big(A(\mu_{k-1,i})\big)'\Big)^{-1} \\ &\quad\times \Big(y_k - C\big(A(\mu_{k-1,i})\big) - \Sigma_{k,i,j}\gamma_{k,i,j}\Big).\end{aligned}$$

The values of $\Gamma_{k,i,j}$, $g_{k,i,j}$ and $\lambda_{k,i,j}$ are given, respectively, by (6.19), (6.12) and (6.18).

*Proof.* Suppose that $\alpha_0(z)$ is a sum of Gaussian densities:

$$\alpha_0(z) = \sum_{i=1}^{N(0)} \lambda_{oi} \, \exp\Big(-\frac{1}{2}\,(z-\mu_{oi})'\Sigma_{oi}^{-1}(z-\mu_{o,i})\Big)$$

with

$$\lambda_{oi} = \rho_{oi}(2\pi)^{-n/2}|\Sigma_{oi}|^{-1/2} > 0.$$

We can, and shall, suppose $\alpha_0(z)$ is a normalized density, that is

$$\int_{R^n} \alpha_0(z)dz = 1,$$

$$\text{so} \qquad \sum_{i=1}^{N(0)} \rho_{oi} = 1.$$

Later (unnormalized) conditional densities are obtained from the recursion (3.1):

$$\alpha_k(x) = \frac{\phi\big(D^{-1}(y_k - C(x))\big)}{|B|\,|D|\phi(y_k)} \int_{R^n} \alpha_{k-1}(z)\psi\Big(B^{-1}\big(x - A(z)\big)\Big)dz.$$

Suppose $\alpha_{k-1}(z)$ is approximated as a sum of Gaussian densities:

$$\alpha_{k-1}(z) \simeq \sum_{i=1}^{N(k-1)} \lambda_{k-1,i} \exp\Big(-\frac{1}{2}\,(z-\mu_{k-1,i})'\Sigma_{k-1,i}^{-1}(z-\mu_{k-1,i})\Big)$$

where again $\lambda_{k-1,i} = \rho_{k-1,i}(2\pi)^{-n/2}|\Sigma_{k-1,i}|^{-1/2} > 0$ and $\sum_{i=1}^{N(k-1)} \rho_{k-1,i} = 1$. Then from the recursion (3.1)

$$\begin{aligned} \alpha_k(x) \simeq \frac{\phi\big(D^{-1}(y_k - C(x))\big)}{|B|\,|D|\phi(y_k)} \int_{R^n} \sum_{i=1}^{N(k-1)} \lambda_{k-1,i} \\ \times \exp\Big(-\frac{1}{2}\big[(z-\mu_{k-1,i})'\Sigma_{k-1,i}^{-1}(z-\mu_{k-1,i}) \\ + \big(x - A(z)\big)'B^{-2}\big(x - A(z)\big)\big]\Big)dz. \end{aligned} \tag{6.3}$$

Consider one integral term in the sum. To simplify the notation we shall drop the suffices for the moment. Using the second order Taylor expansion for $A(z)$ the integral term is approximately:

$$\begin{aligned} I_{k-1,i} = \int_{R^n} \exp\Big(-\frac{1}{2}\big[(z-\mu)'\Sigma^{-1}(z-\mu) \\ + \big(x - A - \nabla A(z-\mu) - \frac{1}{2}\nabla^2 A(z-\mu)^2\big)'B^{-2} \\ \times \big(x - A - \nabla A(z-\mu) - \frac{1}{2}\nabla^2 A(z-\mu)^2\big]\Big)dz. \end{aligned} \tag{6.4}$$

Here, as in Section 5,

$$\begin{aligned} A &= A(\mu_{k-1,i}) \in R^n \\ \nabla A &= \nabla A(\mu_{k-1,i}) \in R^{n\times n} \end{aligned}$$

and $\nabla^2 A$ is the matrix of Hessians.

Dropping the $(z-\mu)^4$ term this is approximately

$$\begin{aligned} \simeq \int_{R^n} \exp\big(-\frac{1}{2}\big[(z-\mu)'\Sigma^{-1}(z-\mu) \\ + \big(x - A - \nabla A(z-\mu)\big)'B^{-2}\big(x - A - \nabla A(z-\mu)\big)\big]G(x,z)dz \end{aligned}$$

where

$$\begin{aligned} G(x,z) &= \exp\Big(\frac{1}{2}\big(x - A - \nabla A(z-\mu)\big)'B^{-2}\nabla^2 A(z-\mu)^2\Big) \\ &= \exp\Big(\frac{1}{2}(z-\mu)'H(x,\mu)(z-\mu)\Big) \end{aligned}$$

in the notation of (4.3), where

$$H(x,\mu) = \frac{1}{2}\big(x - A - \nabla A(z-\mu)\big)'B^{-2}\begin{pmatrix}\nabla^2 A_1 \\ \vdots \\ \nabla^2 A_n\end{pmatrix}.$$

Completing the square we have

$$I_{k-1,i} \simeq K(x)\int_{R^n} \exp\Big(-\frac{1}{2}(z-\sigma_{ki}\delta_{ki})'\sigma_{ki}^{-1}(z-\sigma_{ki}\delta_{ki})\Big)G_{k,i}(x,z)dz$$

where

$$\sigma_{ki}^{-1} = \Sigma_{k-1,i}^{-1} + \nabla A(\mu_{k-1,i})' B^{-2} \nabla A(\mu_{k-1,i}) \tag{6.5}$$

$$\begin{aligned} \delta_{k,i} = {} & \Sigma_{k-1,\mu}^{-1} \mu_{k-1,i} + \nabla A(\mu_{k-1,i})' B^{-2} \big( x - A(\mu_{k-1,i}) \\ & + \nabla A(\mu_{k-1,i}) \mu_{k-1,i} \big) \end{aligned} \tag{6.6}$$

$$\begin{aligned} G_{k,i}(x,z) = {} & \exp \Big( \frac{1}{2} \big( x - A(\mu_{k-1,i}) - \nabla A(\mu_{k-1,i})(z - \mu_{k-1,i}) \big)' \\ & \times B^{-2} \nabla^2 A(\mu_{k-1,i})(z - \mu_{k-1,i})^2 \Big) \end{aligned} \tag{6.7}$$

and

$$\begin{aligned} K(x) = \exp \Big( \frac{1}{2} \big( & \delta_{ki}' \sigma_{ki} \delta_{ki} - \mu_{k-1,i}' \Sigma_{k-1,i}^{-1} \mu_{k-1,i} + \big( x - A(\mu_{k-1,i}) + \\ & \nabla A(\mu_{k-1,i}) \mu_{k-1,i} \big)' \times B^{-2} \big( x - A(\mu_{k-1,i}) + \nabla A(\mu_{k-1,i}) \mu_{k-1,i} \big) \big) \Big). \end{aligned} \tag{6.8}$$

The problem is now to evaluate the integral

$$\int_{R^n} \exp \Big( -\frac{1}{2} (z - \sigma_{ki}\delta_{ki})' \sigma_{ki}^{-1} (z - \sigma_{ki}\delta_{ki}) \Big) G_{ki}(x,z) dz.$$

As in Ito and Xiong, [5], one way is to use Gauss-Hermite quadrature (of which the Julier-Uhlmann unscented filter is a special case). That is, we assume

$$\sigma_{ki} = S_{ki}' S_{ki}$$

and change coordinates writing

$$z = S_{ki}' t + \sigma_{ki}\delta_{ki}, \quad t \in R^n.$$

The integral is then

$$|S_{ki}| \int_{R^n} \exp \Big( -\frac{t't}{2} \Big) G(x, S_{ki}' t + \sigma_{ki}\delta_{ki}) dt. \tag{6.9}$$

Using the Gauss-Hermite formula this is approximately

$$\begin{aligned} & \sum_{j=1}^{m(k,i)} \omega_{k,i,j} G(x, S_{ki}' t_j + \sigma_{ki}\delta_{ki}) \\ = & \sum_{j=1}^{m(k,i)} \omega_{k,i,j} \exp \Big( \big( x - A(\mu_{k-1,i}) - \nabla A(\mu_{k-1,i}) \cdot (S_{ki}' t_j + \sigma_{ki}\delta_{ki} - \mu_{k-1,i}) \big)' \gamma_{k,i,j} \Big) \end{aligned}$$

where

(6.10)

$$\gamma_{k,i,j} = \frac{1}{2} B^{-2} \big( S_{ki}' t_j + \sigma_{ki}\delta_{ki} - \mu_{k-1,i} \big)' \begin{pmatrix} \nabla^2 A_1(\mu_{k-1,i}) \\ \vdots \\ \nabla^2 A_n(\mu_{k-1,i}) \end{pmatrix} \big( S_{ki}' t_j + \sigma_{ki}\delta_{ki} - \mu_{k-1,i} \big).$$

The order of approximation $m(k,i)$ can be chosen. Weights $\overline{\omega}_{k,i,j}$ are determined as in Golub [4]. In fact if $J$ is the symmetric tridiagonal matrix with a zero diagonal and $J_{i,i+1} = \sqrt{i/2}$, $1 \leq i \leq m(k,i) - 1$ then the $t_j$ are the eigenvalues of $J$ and the $\overline{\omega}_{k,i,j}$ are given by $(v_i)(1)^2$, where $v_i(1)$ is the first element of the $i^{\text{th}}$ normalized eigenvector of $J$.

As our integrand in (6.9) is not a $N(0, I_n)$ density in our case

$$\omega_{k,i,j} = (2\pi)^{n/2} |S_{ki}| \overline{\omega}_{k,i,j}.$$

Write $X_i = x - A(\mu_{k-1,i})$ and recall from (6.6) that

$$\delta_{k,i} = \Sigma_{k-1,i}^{-1} \mu_{k-1,i} + \nabla A(\mu_{k-1,i})' B^{-2} \big(X_i + \nabla A(\mu_{k-1,i}) \mu_{k-1,i}\big).$$

Also write

$$M_{kij} = S_{ki}' t_j + \sigma_{ki} \big(\Sigma_{k-1,i}^{-1} \mu_{k-1,i} + \nabla A(\mu_{k-1,i})' B^{-2} \nabla A(\mu_{k-1,i}) \mu_{k-1,i}\big) - \mu_{k-1,i}$$

and

$$F_{ki} = \sigma_{ki} \nabla A(\mu_{k-1,i}) B^{-2}.$$

Then

$$\begin{aligned} I_{k-1,i} \simeq \sum_{j=1}^{m(k,i)} \omega_{k,i,j} K_{ki} \exp \Big[ \frac{1}{2} \big(X_i - \nabla A(\mu_{k-1,i})(M_{kij} + F_{ki} X_i)\big)' \\ \times (M_{kij} + F_{ki} X_i)' \nabla^2 A(\mu_{k-1,i}) \cdot (M_{kij} + F_{ki} X_i) \Big]. \end{aligned} \tag{6.11}$$

Here again $\nabla^2 A(\mu_{k-1,i})$ denotes the matrix of Hessians.

Consider one term in this sum. Retaining only terms of order $\big|\big(x - A(\mu_{k-1,i})\big)\big|^2 = |X_i|^2$ and dropping suffices this is:

$$\omega K(x) \exp \Big[ \frac{1}{2}\, X' H^{-1} X + g' X + \kappa \Big]$$

where:

$$\begin{aligned} H^{-1} = H_{kij}^{-1} &= 2(I - \nabla A F)' B^{-2} \cdot M' (\nabla^2 A) F \\ &\quad - (B^{-2} \nabla A M)' \cdot (\nabla^2 A) F \\ g = g_{kij} &= \frac{1}{2} (I - \nabla A F)' B^{-2} \big(M' (\nabla^2 A) M\big) \\ &\quad - F' (\nabla^2 A) M' B^{-2} \nabla A' M \end{aligned} \tag{6.12}$$

and

$$\kappa = \kappa_{kij} = -M' \nabla A B^{-2} \cdot \big(M' (\nabla^2 A) M\big).$$

Therefore,

$$I_{k-1,i} \simeq \sum_{j=1}^{m(k,i)} \omega_{kij} K_{ki}(x) \exp\left[\frac{1}{2} X_i' H_{kij}^{-1} X_i + g_{kij}' X_i + \kappa_{kij}\right]$$

where $X_i = x - A(\mu_{k-1,i})$.

Substituting in (6.3)

$$\alpha_k(x) \simeq \frac{\phi\big(D^{-1}(y_k - C(x))\big)}{|B|\,|D|\phi(y_k)} \sum_{i=1}^{N(k-1)} \sum_{j=1}^{m(k,i)} \lambda_{k-1,i}\omega_{kij} K_{ki}(x)$$
$$\times \exp\left[\frac{1}{2} X_i' H_{kij}^{-1} X_i + g_{kij}' X_i + \kappa_{kij}\right].$$

Write

$$\rho_{kij} = \frac{\lambda_{k-1,i}\omega_{kij} \exp(\kappa_{kij})}{|B|\,|D|\phi(y_k)} \tag{6.13}$$

so

$$\alpha_k(x) \simeq \sum_{i=1}^{N(k-1)} \sum_{j=1}^{m(k,i)} \rho_{kij}\phi\big(D^{-1}(y_k - C(x))\big) K_{ki}(x) \exp\left[\frac{1}{2} X_i' H_{kij}^{-1} X_i + g_{kij}' X_i\right].$$

Consider one term only in the sum and replace $C(x)$ by its second order Taylor expansion about $A(\mu_{k-1,i})$. Again, drop suffices to simplify the notation. Then

$$\begin{aligned}
L_{kij}(x) = L(x) :&= \rho\phi\Big(D^{-1}\big(y_k - C(x)\big)\Big) K(x) \\
&\quad \times \exp\left[\frac{1}{2}(x-A)' H^{-1}(x-A) + g'(x-A)\right] \\
&\simeq \rho K(x) \exp\left[\frac{1}{2} X_i' H^{-1} X_i + g' X_i\right] \\
&\quad \times \exp\Big[-\frac{1}{2}\big(y - C - \nabla C X_i - \frac{1}{2} X_i' \nabla^2 C X_i\big)' \\
&\quad \times D^{-2}\Big(y - C - \nabla C X_i - \frac{1}{2} X_i' \nabla^2 C X_i\Big)
\end{aligned}$$

Here again $C = C\big(A(\mu_{k-1,i})\big) \in R^m$

$$\begin{aligned}
\nabla C &= \nabla C\big(A(\mu_{k-1,i})\big) \in R^{m\times n} \\
\nabla^2 C &= \begin{pmatrix} \nabla^2 C_1\big(A(\mu_{k-1,i})\big) \\ \vdots \\ \nabla^2 C_m\big(A(\mu_{k-1,i})\big) \end{pmatrix} \\
A &= A(\mu_{k-1,i}) \in R^n
\end{aligned}$$

and

$$X_i = x - A(\mu_{k-1,i}) = x - A.$$

Keeping only terms of order $|x-A|^2 = |X_i|^2$ this gives

$$L(x) \simeq \rho K(x) \exp\Big(-\frac{1}{2}\big(y - C - \nabla C X_i)\big)' D^{-2}\big(y - C - \nabla C X_i\big)\Big)$$
$$\times \exp\Big(\frac{1}{2}(y-C)' D^{-2} \cdot X_i' \nabla 2 C X_i\Big) \exp\Big[\frac{1}{2} X_i' H^{-1} X_i + g' X_i\Big].$$

As in (4.4) we can write

$$\exp\Big(\frac{1}{2}(y-C)' D^{-2} \cdot X_i' \nabla^2 C X_i\Big)$$

as

$$\exp\Big(\frac{1}{2} X_i' R^{-1} X_i\Big)$$

where

$$R^{-1} = \Big(y_k - C\big(A(\mu_{k-1,i})\big)\Big)' D^{-2} \cdot \nabla^2 C\big(A(\mu_{k-1,i})\big).$$

Write

$$Z^{-1} := R^{-1} + H^{-1}.$$

Then dropping suffices

$$L(x) \simeq \rho K(x) \exp\Big[\frac{1}{2} X_i' Z^{-1} X_i + g' X_i\Big]$$
$$\times \exp\Big(-\frac{1}{2}\big(y - C - \nabla C X_i\big)' D^{-2}\big(y - C - \nabla C X_i\big)\Big)$$

and

$$\alpha_k(x) \simeq \sum_{i=1}^{N(k-1)} \sum_{j=1}^{m(i,j)} \rho_{k,i,j} L_{k,i,j}(x).$$

Recalling $X = X_i = x - A(\mu_{k-1,i})$ and substituting for $K(x)$ from (6.8) we have

$$L_{kij}(x) = L(x) \simeq \rho \exp\Big(\frac{1}{2}\big[\big(\Sigma^{-1}\mu + \nabla A B^{-2}(X + \nabla A\mu)\big)'$$
$$\times \sigma_k\big(\Sigma^{-1}\mu + \nabla A B^{-2}(X + \nabla A\mu)\big)$$
$$- \mu'\Sigma^{-1}\mu + (X + \nabla A\mu)' B^{-2}(X + \nabla A\mu)\big]\Big)$$
$$\times \exp\Big(\frac{1}{2} X' Z^{-1} X + g' X\Big)$$
$$\times \exp\Big(-\frac{1}{2}\big(y - C - \nabla C X\big)' D^{-2}\big(y - C - \nabla C X\big)\Big).$$

Collecting terms in $X = X_i = x - A(\mu_{k-1,i})$ this is:

$$= G \exp\Big(-\frac{1}{2}\,(X'\Sigma^{-1} X - 2\Delta' X)\Big)$$

where

$$\Sigma^{-1} = \Sigma_{kij}^{-1} = \nabla C\big(A(\mu_{k-1,i})\big)' D^{-2} \nabla C\big(A(\mu_{k-1,i})\big) \\ + B^{-2} - B^{-2} \nabla A(\mu_{k-1,i}) \cdot \sigma_{k,i} \nabla A(\mu_{k-1,i})' B^{-2} - Z_{kij}^{-1} \tag{6.14}$$

and

$$\begin{aligned} \Delta &= \Delta_{kij} \\ &= \nabla C\big(A(\mu_{k-1,i})\big) D^{-2} \Big(y_k - C\big(A(\mu_{k-1,i})\big)\Big) - B^{-2} \nabla A(\mu_{k-1,i}) \mu_{k-1,i} \\ &\quad + g_{k,i,j} + B^{-2} \nabla A(\mu_{k-1,i}) \\ &\quad \times \sigma_{k,i} \big(\Sigma_{k-1,i}^{-1} \mu_{k-1,i} + \nabla A(\mu_{k-1,i})' B^{-2} \nabla A(\mu_{k-1,i}) \mu_{k-1,i}\big). \end{aligned} \tag{6.15}$$

Further

$$G = G_{kij} = \rho \, \exp\, \Big( \frac{1}{2} \, (\Sigma^{-1}\mu + \nabla A \cdot B^{-2} \nabla A' \mu)' \sigma \big(\Sigma^{-1}\mu + \nabla A B^{-2} \nabla A' \mu\big) \\ - \mu' \Sigma^{-1} \mu - \mu' \nabla A' B^{-2} \nabla A \mu - \frac{1}{2} \, (y - C)' D^{-2} (y - C) \Big). \tag{6.16}$$

Completing the square we have

$$L_{k,i,j}(x) = L(x) \simeq \lambda_{k,i,j} \, \exp\, \Big[ - \, \frac{1}{2} \, (X_i - \Sigma\Delta)' \Sigma^{-1} (X_i - \Sigma\Delta) \Big] \tag{6.17}$$

where

$$\lambda_{k,i,j} = G_{kij} \, \exp\, \Big( \frac{1}{2} \, \Delta' \Sigma \Delta \Big). \tag{6.18}$$

Now as in (5.5)

$$\Sigma_{k|k-1,i} := \nabla A(\mu_{k-1,i}) \Sigma_{k-1,i} \nabla A(\mu_{k-1,i})' + B^2$$

is an approximate conditional variance of $x_k$, given $\mathcal{Y}_{k-1}$, if

$$\widehat{x}_{k-1|k-1} \sim N\big(\mu_{k-1,i}, \Sigma_{k-1,i}\big).$$

Then, as in Corollary 5.2

$$\Sigma_{k|k-1,i}^{-1} = B^{-2} - B^{-2} \nabla A(\mu_{k-1,i}) \sigma_{k,i} \nabla A(\mu_{k-1,i})' B^{-2}$$

so, from (6.14)

$$\Sigma^{-1} = \Sigma_{k,i,j}^{-1} = \nabla C' D^{-2} \nabla C + \Sigma_{k|k-1,i}^{-1} - Z^{-1}.$$

Write $\Gamma^{-1} = \Gamma_{k,i,j}^{-1} := \Sigma_{k|k-1,i}^{-1} - Z_{k,i,j}^{-1}$ so, using the MIL, Lemma 5.1

$$\Gamma = \Sigma_{k|k-1,i} - \Sigma_{k|k-1,i} \big(\Sigma_{k|k-1,i} - Z_{k,i,j}\big)^{-1} \Sigma_{k|k-1,i} \, . \tag{6.19}$$

Also, as

$$\Sigma^{-1} = \Sigma^{-1}_{k,i,j} = \nabla C' D^{-2} \nabla C + \Gamma^{-1}$$

again using Lemma 5.1

$$\Sigma = \Sigma_{k,i,j} = \Gamma - \Gamma \nabla C' (D^2 + \nabla C \Gamma \nabla C')^{-1} \nabla C \Gamma. \tag{6.20}$$

Finally we evaluate

$$\mu_{k,i,j} = \Sigma_{k,i,j} \Delta_{k,i,j}.$$

Dropping suffices, from (6.14) and (6.15) this is

$$\begin{aligned}
&\big[\Gamma - \Gamma \nabla C'(D^2 + \nabla C \Gamma \nabla C')^{-1} \nabla C \Gamma\big] \nabla C' D^{-2}(y - C) \\
&\quad + \Sigma\big[B^{-2} \nabla A \sigma\big(\Sigma^{-1}\mu + \nabla A' B^{-2} A\mu)\big)\big] + \Sigma g - \Sigma B^{-2} \nabla A \mu \\
&\quad = \big[\Gamma \nabla C' - \Gamma \nabla C'(D^2 + \nabla C \Gamma \nabla C')^{-1}(\nabla C \Gamma \nabla C' + D^2 - D^2)\big] D^{-2}(y - C) \\
&\qquad + \Sigma\big[B^{-2} \nabla A \cdot \sigma\big(\Sigma^{-1} + \nabla A' B^{-2} \nabla A\big)\mu\big] + \Sigma g - \Sigma B^{-2} \nabla A \mu.
\end{aligned}$$

Recalling from (6.5) that

$$\sigma = (\Sigma^{-1} + \nabla A' B^{-2} \nabla A)^{-1}$$

this is

$$= \Gamma \nabla C' (D^2 + \nabla C \Gamma \nabla C')^{-1} (y - C) + \Sigma g.$$

Therefore, in terms of the variable

$$X = X_i = x - A(\mu_{k-1,i})$$

the mean is

$$\Sigma_{k,i,j} \Delta_{k,i,j} = \Gamma \nabla C' (D^2 + \nabla C \Gamma \nabla C')^{-1} (y - C) + \Sigma g$$

so in terms of the original variable $x$ the mean is

$$= A + \Gamma \nabla C' (D^2 + \nabla C \Gamma \nabla C')^{-1} (y - C) - \Sigma g.$$

That is, now writing in suffices, we have obtained a Gaussian density in the sum which determines $\alpha_k(x)$ which has a variance

$$\begin{aligned}
\Sigma_{k,i,j} = \Gamma_{k,i,j} &- \Gamma_{k,i,j} \nabla C\big(A(\mu_{k-1,i})\big)' \Big(D^2 + \nabla C\big(A(\mu_{k-1,i})\big)\Gamma_{k,i,j} \\
&\times \nabla C\big(A(\mu_{k-1,i})\big)'\Big)^{-1} \times \nabla C\big(A(\mu_{k-1,i})\Gamma_{k,i,j},
\end{aligned}$$

where

$$\Gamma_{k,i,j}^{-1} = \Sigma_{k|k-1,i}^{-1} - Z_{k,i,j}^{-1},$$

and a mean

$$\begin{aligned}\mu_{k,i,j} &= A(\mu_{k-1,i}) + \Gamma_{k,i,j}\nabla C\big(A(\mu_{k-1,i})\big)\\ &\quad\times\Big(D^2 + \nabla C\big(A(\mu_{k-1,i})\big)\Gamma_{k,i,j}\nabla C\big(A(\mu_{k-1,i})\big)'\Big)^{-1}\\ &\quad\times\Big(y_k - C\big(A(\mu_{k-1,i})\big)\Big) - \Sigma_{k,i,j}g_{k,i,j}.\end{aligned}$$

Here $g_{k,i,j}$ is given by (6.12).

Also, the new weights $\lambda_{k,i,j}$ are given by (6.18).

Therefore,

$$\alpha_k(x) \simeq \sum_{i=1}^{N(k-1)}\sum_{j=1}^{m(k,i)} \lambda_{k,i,j}\exp\Big(-\frac{1}{2}(x-\mu_{k,i,j})'\Sigma_{k,i,j}^{-1}(x-\mu_{k,i,j})\Big).$$

Pruning can be effected by selecting only so many of the terms based on the size of the $\lambda_{k,i,j}$. □

**7. Conclusion.** A discrete time version of the Zakai equation has been obtained. Using first order linearizations the equation can be solved and the extended Kalman filter derived.

Second order Taylor expansions of the non-linear terms were then considered and approximate solutions of the Zakai equation derived in terms of Gaussian sums. Formulae for updating the means, variances and weights of the terms in the sums are given. Detailed proofs will appear in [2].

## REFERENCES

[1] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*, Springer-Verlag, Berlin, Heidelberg, New York, Second edition 2006.

[2] R. J. Elliott and S. Haykin, *An extended EKF*, University of Calgary preprint (2006), (to appear).

[3] R. J. Elliott and V. Krishnamurthy, *New finite dimensional filters for parameter estimation of discrete time linear Gaussian models*. IEEE Trans. Automatic Control, 44(1998), pp. 938-951.

[4] G. H. Golub, *Some modified matrix eigenvalue problems*, SIAM Review, 15(1973), pp. 318-334.

[5] K. Ito and K. Xiong, *Gaussian filters for non-linear filtering problems*, IEEE Trans. Automatic Control, 45(2000), pp. 910-927.

[6] H. Kushner and A. S. Budhiraja, *A nonlinear filtering algorithm based on an approximation of the conditional distribution*, IEEE Trans. Automatic Control, 45(2000), pp. 580-585.