December, 1991

NOTES ON NONLINEAR AND NONNORMAL FILTER

Hisashi Tanizaki

University of Pennsylvania

1. Introduction

The state-space model consists of two equations; measurement and transition equations. In the case of linear measurement and transition equations, we can obtain the linear recursive Kalman filter algorithm explicitly under the normality assumption for the error terms. In the case of nonlinear measurement and transition equations, however, it is difficult to get the algorithm in closed form. Therefore some approximations have to be made for estimation.

There are two approaches to nonlinear filtering algorithms. One is approximating the nonlinear measurement and transition equations by the Taylor series expansion. The filtering algorithm is obtained by applying the linearized nonlinear functions directly to the conventional linear recursive algorithm. Another approach is approximating the underlying density functions of the state vector. The recursive algorithm on the densities is derived by Bayes' formula.

For one of the approaches, the most heuristic and easiest approximation is to use the Taylor series expansion and apply the linear algorithm directly to the expanded nonlinear measurement and transition functions. The traditional filters are the extended Kalman filter, the second-order nonlinear filter and the

single-stage iteration filter. The nonlinear filters based on the Taylor series expansion are described in Wishner, Tabaczynski and Athans (1969) and Gelb (1974).

When we apply the Taylor series expansion to the nonlinear functions, we encounter some problems with the approximated error terms; (i) the expectation of the error terms is not necessarily zero, (ii) the state vector is correlated with the error terms, (iii) the error term in the measurement equation is correlated with the error term in the transition equation, and (iv) the error terms are not normal.

Thus, the approximated error terms are not well-behaved. If we approximate the error terms to be well-behaved and apply the linearized functions to the linear recursive algorithm, we have the algorithms based on the Taylor series expansion. Clearly, however, ignoring these problems of the error terms implies that the filtering estimates of the state vector are biased, because applying the linearized nonlinear functions to the conventional Kalman filter algorithm implies that non-normal error terms are approximated as normal ones. As Meinhold and Singpurwalla (1989) pointed out, Kalman filter models based on the normality assumption are known to be non-robust, which implies that when there is a large difference between the prior density and the observed data, the posterior density becomes unrealistic. Therefore, the approximation of the densities, rather than the approximation of the nonlinear functions, are essential to the nonlinear filter.

It is known that the recursive algorithm of the Kalman filter can be obtained by the conditional density function of the state variable from Bayes' formula (see, for example, Harvey (1989)). The nonlinear filters based on the density functions were developed by Sorenson and Alspach (1971), Alspach and Sorenson (1972), Kitagawa (1987), Kramer and Sorenson (1988), and Carlin, Polson and Stoffer (1990). Alspach and Sorenson (1972) and Sorenson and Alspach (1971) approximated the densities by a sum of Gaussian distributions. Carlin, Polson and Stoffer (1990) suggested applying an adaptive Monte-Carlo integration technique known as the Gibbs sampler to the density approximation. Kitagawa (1987) and Kramer and Sorenson (1988) proposed an approach of representing the densities numerically by a piecewise linear function. There, Kitagawa represented each density as the number of segments, location of nodes and the value at each node, and evaluated it through numerical integration. According to his approach, however, computational burden increases more than proportionally as the dimension of the state vector increases. Programming is also extremely tedious if the dimension is high.

Improving the problems, Tanizaki (1991) suggested the nonlinear filter by a simulation-based density approximation through random numbers, in which a recursive algorithm of the weighting functions, which are represented by the ratio of two densities, is derived through Monte-Carlo simulation. Geweke (1989) and Shao (1989) developed an approximation of prior density in Bayesian framework, so-called importance sampling theory. Here,

this approach is applied to the nonlinear filter. From the point of programming and computational time, the nonlinear filter based on the importance sampler can be easily extended to higher dimensional cases in practice.

In addition, the algorithm based on the density functions is applicable in all the following cases:

(i) nonlinear measurement and transition equations and nonnormal errors,

(ii) linear measurement and transition equations and nonnormal error,

(iii) nonlinear measurement and transition equations and normal errors,

(iv) linear measurement and transition equations and normal errors.

Thus, we can handle all the cases with the density approximation approach, because there is no distinction between non-normality and nonlinearity in the case of the nonlinear filter based on the density functions. There, rather than the nonlinear functions, the densities are approximated.

2. Overview of Nonlinear and Nonnormal Filter

The nonlinear state-space is represented by the following two equations:

$$y_t = h_t(\alpha_t, \epsilon_t), \qquad (Measurement\ equation)$$

$$\alpha_t = g_t(\alpha_{t-1}, \eta_t), \qquad (Transition\ equation)$$

$$for\ t=1,\ldots T,$$

where the error terms ϵ_t and η_t are mutually independently distributed, which are not necessarily normal. The initial state variable α_0 is assumed to be uncorrelated with ϵ_t and η_t for any time t. Define $a_{t|s}$ and $\Sigma_{t|s}$ as the conditional expectation and variance of α_t given the information up to s, i.e., $a_{t|s}=E(\alpha_t|I_s)$ and $\Sigma_{t|s}=Cov(\alpha_t|I_s)$, where $I_s=\{y_{s-1}, y_{s-2},\ldots,y_1\}$ which is the information available up to s. The evaluation of $\alpha_{t|s}$ is known as filtering if $t=s$, smoothing if $t<s$, and prediction if $t>s$, respectively. Here we focus on the filtering estimation.

In the above nonlinear measurement and transition equations, the functions g_t and h_t may depend on the other exogenous variables, say x_t. Also, in the case of filtering, g_t and h_t may include the information up to time $t-1$, i.e., I_{t-1}. Based on the above two equations, a filtering algorithm is given by the

following two equations:

$$P(\alpha_t|I_{t-1}) = \int P(\alpha_t|\alpha_{t-1}, I_{t-1})\, P(\alpha_{t-1}|I_{t-1})\, d\alpha_{t-1}, \quad \textit{(Prediction equation)}$$

$$P(\alpha_t|I_t) = \frac{P(y_t|\alpha_t, I_{t-1})\, P(\alpha_t|I_{t-1})}{\int P(y_t|\alpha_t, I_{t-1})\, P(\alpha_t|I_{t-1})\, d\alpha_t}, \quad \textit{(Updating equation)}$$

$$\textit{for } t=1,\ldots,T,$$

which are easily obtained from Bayes' formula. Unless g_t and h_t depend on the past information I_{t-1}, the densities $P(\alpha_t/\alpha_{t-1}, I_{t-1})$ and $P(y_t/\alpha_t, I_{t-1})$ are replaced by $P(\alpha_t/\alpha_{t-1})$ and $P(y_t/\alpha_t)$. Note that, so far, we have not assumed any functional forms of the nonlinear measurement and transition equations and the distribution functions of ϵ_t and η_t, i.e., we have treated a nonlinear filtering problem with a general form. In practice, however, we have to assume the nonlinear functions g_t and h_t, the distribution functions of ϵ_t and η_t and the density function of the initial state variable α_0. In the above density algorithm, the two densities $P(\alpha_t/\alpha_{t-1}, I_{t-1})$ and $P(y_t/\alpha_t, I_{t-1})$ are explicitly computed by assuming the distributions for ϵ_t and η_t and transforming the variables from ϵ_t and η_t to y_t and α_t based on the measurement and transition equations. Under the condition that these functions are known, $P(\alpha_t/I_{t-1})$ and $P(\alpha_t/I_t)$ are computed in the following way: (1) given the initial density $P(\alpha_0/I_0)$, $P(\alpha_1/I_0)$ can be obtained through the integration in the prediction equation, where $P(\alpha_1/\alpha_0)$ is derived from the

distribution of η_t, and (2) based on the density $P(\alpha_1/I_0)$ computed in the step (1), $P(\alpha_1/I_1)$ is computed from the above updating equation, where $P(y_1/\alpha_1)$ is derived from the distribution of ϵ_t. Thus, we can compute $P(\alpha_t/I_{t-1})$ and $P(\alpha_t/I_t)$ for $t=1,...T$, recursively, given $P(\alpha_t/\alpha_{t-1})$, $P(y_t/\alpha_t)$, $t=1,...,T$, and $P(\alpha_0/I_0)$. In the next section, we discuss a linear filter, which is a special case of nonlinear filter.

3. Example: Linear and Normal Case

The standard Kalman filtering algorithm developed by Kalman (1960) and Kalman and Bucy (1961) is derived in the case where g_t and h_t are linear in both state variables and error terms, and the error terms ϵ_t and η_t are mutually independently and normally distributed. Also, we need to assume that the initial state variable α_0 is normal. These additional assumptions are given by:

$$y_t = Z_t \alpha_t + S_t \varepsilon_t, \qquad (\textit{Measurement equation})$$

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t, \qquad (\textit{Transition equation})$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix}\right),$$

$$\alpha_0 \sim N(a_{0|0}, \Sigma_{0|0}),$$

where Z_t, S_t, H_t, T_t, R_t and Q_t may depend on the information available at time $t-1$, i.e., I_{t-1}.

Under these assumptions, we can have the following conventional Kalman filtering algorithm (see, for example, Anderson and Moore (1979)):

$$a_{t|t-1}=T_t a_{t-1|t-1},$$

$$\Sigma_{t|t-1}=T_t \Sigma_{t-1|t-1} T_t'+R_t Q_t R_t',$$

$$k_t=\Sigma_{t|t-1} Z_t'(Z_t \Sigma_{t|t-1} Z_t'+S_t H_t S_t')^{-1},$$

$$\Sigma_{t|t}=\Sigma_{t|t-1}-k_t(Z_t \Sigma_{t|t-1} Z_t'+S_t H_t S_t')\, k_t',$$

$$a_{t|t}=a_{t|t-1}+k_t(y_t-Z_t a_{t|t-1}),$$

$$for\ t=1,\ldots T,$$

where Z_t, S_t, T_t, R_t, Q_t, H_t, $a_{0|0}$ and $\Sigma_{0|0}$ are assumed to be known. In the case of linear measurement and transition equations, it is clear from the property of the normal distribution that α_t is normal if η_t is normally distributed, and that y_t is also normal if α_t and ϵ_t are normal. Therefore, the above algorithm is obtained from the first- and the second-moments of the normal distribution.

The first two equations in the above algorithm are called the prediction equations, which essentially predict the parameter α_t using the information up to time $t-1$. The last two equations are known as the updating equations, because they play the role of combining the new observation obtained at time t (i.e., y_t) and the past data up to time $t-1$ (i.e., I_{t-1}). Moreover, k_t is called the Kalman gain, which is chosen such that $a_{t|t}$ has minimum variance.

4. Nonlinear and Nonnormal Filtering Algorithms

As it is well-known, traditional nonlinear filters include the extended Kalman filter, the second-order nonlinear filter, the single-stage iteration filter and the Gaussian sum filter.

According to the first two filters, the nonlinear functions are approximated by the first- or the second-order Taylor series expansion, and the approximated nonlinear functions are applied to the standard linear recursive Kalman filter algorithm.

The single-stage iteration filter is based on the conventional nonlinear least squares estimation, which is equivalent to the mixed estimation method (i.e., Goldberger-Theil estimation), where a sum of the weighted residuals is minimized to derive the filtering estimates. There, however, the transition equation has to be approximated by the Taylor series expansion in obtaining the prediction equations.

Moreover, the Gaussian sum filter is a technique of approximating each density as a sum of normal distributions, but in each density the nonlinear measurement and transition equations are expanded by the first-order Taylor series approximation. Therefore, as a result, the Gaussian sum filter is known to be the weighted average of extended Kalman filters.

Thus, the traditional nonlinear filters utilize the nonlinear functions approximated by the Taylor series expansion. The problem of use of the Taylor series expansion is that the approximated error terms are not normal even if the original error terms, i.e., ϵ_t and η_t, are normally distributed. Applying the approximated nonlinear functions to the linear recursive algorithm leads to the biased filtering estimates, because this approximation implies approximating the nonnormal errors as the normal ones. Therefore, we need to consider the underlying density functions without approximating the nonlinear measurement and transition equations.

Recently, an attempt has been made to approximate the underlying density functions, not the nonlinear measurement and transition equations. Kitagawa (1987) approximated the densities by numerical integration. Tanizaki (1991) applied the importance sampling theory (see, for example, Geweke (1989)) to the density approximation in the context of the filtering theory. We expect better filtering estimates by their approaches. In this section, we discuss these two filters.

The underlying basic idea of these filters is as follows. Let x be a random variable from a density function $P(x)$. Consider evaluating the expectation of a function of x, say $g(x)$. According to the Kitagawa approach, the expectation of $g(x)$ is computed as:

$$E(g(x)) = \int g(x) P(x) dx$$

$$= \sum_{i=1}^{n} g(x_i) P(x_i) (x_i - x_{i-1}),$$

(More precisely, take below.)

$$= \sum_{i=1}^{n} \frac{1}{2} (g(x_i) P(x_i) + g(x_{i-1}) P(x_{i-1})) (x_i - x_{i-1}),$$

where n is referred to as the number of nodes and x_i, $i=0,\ldots,n$, are regarded as the nodes, which are assumed to be sorted in order of size. The integration is represented as a sum of rectangles or trapezoids.

For another approximation, taking an appropriate density function of x, say $P_x(x)$, Tanizaki (1991) utilized the following evaluation of the expectation:

$$E(g(x)) = \int g(x) P(x) dx$$

$$= \int g(x) \frac{P(x)}{P_x(x)} P_x(x) dx$$

$$= \int g(x) \omega(x) P_x(x) dx$$

$$= \frac{1}{n} \sum_{i=1}^{n} g(x_i) \omega(x_i),$$

where $P_x(x)$ is appropriately assumed by a researcher and not too different from the density $P(x)$. $\omega(x)$ is referred to as the

weighting function.

Both approaches are useful when we know the functional from of the density $P(x)$ but it is difficult to evaluate the expectation of $g(x)$ in closed form. The former evaluates the expectation by numerical integration, and the latter by the Monte-Carlo simulation, where the random numbers of x are generated based on the appropriately chosen density function $P_x(x)$. Clearly, both approaches accomplish an asymptotically unbiased estimate of $E(g(x))$.

Applying numerical integration to the density evaluation, the nonlinear filtering algorithm proposed by Kitagawa (1987) is represented by:

$$P(\alpha_{i,t}|I_{t-1}) = \sum_{j=1}^{n} P(\alpha_{i,t}|\alpha_{j,t-1}, I_{t-1})\, P(\alpha_{j,t-1}|I_{t-1})\, (\alpha_{j,t-1}-\alpha_{j-1,t-1}),$$

(*Prediction equation*)

$$P(\alpha_{i,t}|I_t) = \frac{P(y_t|\alpha_{i,t}, I_{t-1})\, P(\alpha_{i,t}|I_{t-1})}{\sum_{j=1}^{n} P(y_t|\alpha_{j,t}, I_{t-1})\, P(\alpha_{j,t}|I_{t-1})\, (\alpha_{j,t}-\alpha_{j-1,t})},$$

(*Updating equation*)

for $t=1,\ldots,T$,

where $\alpha_{i,t}$, $i=1,\ldots,n$, are regarded as the nodes which are sorted in order of size. Thus, the recursive algorithm of the densities $P(\alpha_t/I_{t-1})$ and $P(\alpha_t/I_t)$ can be derived, corresponding to each node. Note that it is possible to evaluate the integration above by a sum of the trapezoids. To make things easier, here we take a sum of rectangles, rather than trapezoids. Based on the densities evaluated at $\alpha_{i,t}$, we can easily compute the filtering estimates, i.e., $a_{t/s}$ and $\Sigma_{t/s}$, as follows:

$$a_{t|s}=\sum_{i=1}^{n}\alpha_{i,t}P(\alpha_{i,t}|I_s)\,(\alpha_{i,t}-\alpha_{i-1,t}),$$

$$\Sigma_{t|s}=\sum_{i=1}^{n}(\alpha_{i,t|s}-a_{t|s})(\alpha_{i,t|s}-a_{t|s})'P(\alpha_{i,t}|I_s)\,(\alpha_{i,t}-\alpha_{i-1,t}),$$

where $a_{t/s}=E(\alpha_t/I_s)$, $\Sigma_{t/s}=Cov(\alpha_t/I_s)$ and $s=t,t-1$.

On the other hand, according to the approach suggested by Tanizaki (1991), the following recursive algorithm of the weighting functions is derived by applying the importance sampling theory to the density approximation.

$$\omega_{i,t|t-1}=\frac{1}{n}\sum_{j=1}^{n}\frac{P(\alpha_{i,t}|\alpha_{j,t-1},I_{t-1})}{P_{\alpha}(\alpha_{j,t-1}|I_{t-1})}\omega_{j,t-1|t-1}, \quad \textit{(Prediction equation)}$$

$$\omega_{i,t|t}=\frac{P(y_t|\alpha_{i,t},I_{t-1})\,\omega_{i,t|t-1}}{\frac{1}{n}\sum_{j=1}^{n}P(y_t|\alpha_{j,t},I_{t-1})\,\omega_{j,t|t-1}}, \quad \textit{(Updating equation)}$$

$$\textit{for } t=1,\ldots,T,$$

where the weighting function $\omega_{t|s}$ is defined as the ratio of two densities, i.e., $\omega_{t|s}=P(\alpha_t|I_s)/P_{\alpha}(\alpha_t)$. $\omega_{i,t|s}$ is the weighting function evaluated at $\alpha_{i,t}$, which is a random draw generated from the appropriately chosen $P_{\alpha}(\alpha_t)$. Note that the initial weighting function $\omega_{0|0}$ has to be assumed for the algorithm. Given the weighting functions, the filtering estimates are given as follows:

$$a_{t|s}=\frac{1}{n}\sum_{i=1}^{n}\alpha_{i,t|s}\omega_{i,t|s},$$

$$\Sigma_{t|s}=\frac{1}{n}\sum_{i=1}^{n}(\alpha_{i,t|s}-a_{t|s})(\alpha_{i,t|s}-a_{t|s})'\omega_{i,t|s},$$

where $s=t,t-1$. Furthermore, it is known that the obtained estimates approach the true values as n goes to infinity (see Geweke (1989) for consistency). Each density is evaluated as $P(\alpha_{i,t}|I_s)=\omega_{i,t|s}P_{\alpha}(\alpha_{i,t})$ for each node $\alpha_{i,t}$.

Taking a logistic type of nonlinear measurement and transition equations and performing Monte-Carlo experiments, Tanizaki (1991) obtained the following results; (i) both of the filtering estimates go to the exactly same value as n approaches infinity (n=100), and (ii) the approach suggested by Tanizaki showed a good performance even if n is small (n=5), compared with the Kitagawa approach.

As a final comment, we should keep in mind the following. According to Kitagawa's approach, computational burden increases more than proportionally as the dimension of the state vector increases. Programming is also extremely tedious if the dimension is high. From the point of programming and computational time, the nonlinear filter based on the importance sampler can be easily extended to higher dimensional cases in practice.

Thus, the Tanizaki approach has many attractive features over the Kitagawa approach, and the former is more practical and applicable than the latter.

REFERENCES

Alspach, D.L. and H.W. Sorenson, 1972, "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, Vol. AC-17, No.4, pp.439-448.

Anderson, B.D.O. and J.B. Moore, 1979, *Optimal Filtering*, Prentice-Hall, New York.

Carlin, B.P., N.G. Polson and D.S. Stoffer, 1990, "A Monte Carlo Approach to Nonnormal and Nonlinear State Space Modeling," Technical Report 486, Carnegie Mellon University.

Gelb, A. (Ed.), 1974, *Applied Optimal Estimation*, MIT Press.

Geweke, J., 1989, "Modeling with Normal Polynomial Expansions," in *Economic Complexity: Chaos, Sunspots, Bubbles, and Nonlinearity*, edited by Barnett et al. Cambridge University Press.

Harvey, A.C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Kalman, R.E., 1960, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering, Transactions ASME*, Series D, 82, pp.35-45.

Kalman, R.E. and R.S. Bucy, 1961, "New Results in Linear Filtering and Prediction Theory," *Journal of Basic Engineering, Transactions ASME*, Series D, 83, pp.95-108.

Kramer, S.C. and Sorenson H.W., 1988, "Recursive Bayesian Estimation Using Piece-wise Constant Approximations," *Automatica*, Vol.24, No.6, pp.789-801.

Kitagawa, G., 1987, "Non-Gaussian State-Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, Vol.82, pp.1032-1063 (with discussion).

Meinhold, R.J. and N.D. Singpurwalla, 1989, "Robustification of Kalman Filter Models," *Journal of the American Statistical Association*, Vol.84, pp.479-486.

Shao, J., 1989, "Monte Carlo Approximations in Bayesian Decision Theory," *Journal of the American Statistical Association*, Vol.84, No.407, pp.727-732.

Sorenson, H.W. and D.L. Alspach, 1971, "Recursive Bayesian Estimation Using Gaussian Sums," *Automatica*, Vol.7, pp.465-479.

Tanizaki, H., 1991, "Nonlinear Filters - Estimation and Applications -," Dissertation, University of Pennsylvania.

Wishner, R.P., J.A. Tabaczynski and M. Athans, 1969, "A Comparison of Three Non-Linear Filters," *Automatica*, Vol.5, pp.487-496.