TRABAJOS DE ESTADISTICA
Vol. 6, núm. 2, 1991, pp. 111 a 119

KALMAN FILTER WITH A NON-LINEAR NON-GAUSSIAN OBSERVATION RELATION

T. CIPRA
Dept. of Statistics
Charles University of Prague
Sokolovska 83
186 00 Prague 8

A. RUBIO
Dpto. de Matemáticas
Universidad de Extremadura
10071 Cáceres

ABSTRACT

The dynamic linear model with a non-linear non-Gaussian observation relation is considered in this paper. Masreliez's theorem (see Masreliez's (1975)) of approximate non-Gaussian filtering with linear state and observation relations is extended to the case of a non-linear observation relation that can be approximated by a second-order Taylor expansion.

Key words: Kalman filter; non-linear non-Gaussian filtering.
A.M.S. classification: 62M20, 62F35, 62F15, 60G35.

RESUMEN

El modelo lineal dinámico con observación nolineal y no-Gausiano se estudia en este artículo. Se extiende el teorema de Masreliez (ver Masreliez (1975)) como una aproximación de filtrado no-Gausiano con ecuación de estado lineal y ecuación de observaciones también lineal, al caso en que la ecuación de observaciones nolineal pueda aproximarse mediante la extesión de Taylor de segundo orden.

Palabras clave: Filtro de Kalman; filtrado nolineal no-Gausiano.
Clasificación A.M.S.: 62M20, 62F35, 62F15, 60G35.

Recibido, febrero 1991.
Revisado, Julio 1991.

1. Introduction

The discrete Kalman filter (see e.g. Jazwinski (1970)) provides the recursive estimation of the state vector in the dynamic model (or state space model)

$$x_t = F_t x_{t-1} + w_t, \tag{1.1}$$

$$y_t = H_t x_t + v_t, \tag{1.2}$$

where F_t is a known $n \times n$ transition matrix in the state relation (1.1) describing the development of the n-dimensional state vector x in time, H_t is a known $m \times n$ observation matrix in the observation relation (1.2) assigning the state x to the m-dimensional observation vector y. The $n \times 1$ and $m \times 1$ disturbance vectors w_t and v_t form zero mean, uncorrelated and mutually uncorrelated sequences with known covariance matrices W_t and V_t, respectively. Moreover, appropriate initial conditions are given.

The Kalman filter is usually used to obtain recursive formulae for the linear minimum variance estimator $\hat{x}_t^t$ of the state x_t and for its covariance matrix

$$P_t^t = E\{(x_t - \hat{x}_t^t)(x_t - \hat{x}_t^t)' | Y^t\} \tag{1.3}$$

in a current time period t using all previous information $Y^t = \{y_0, y_1, \ldots, y_t\}$:

$$\hat{x}_t^t = \hat{x}_t^{t-1} + P_t^{t-1} H_t' (H_t P_t^{t-1} H_t' + V_t)^{-1} (y_t - H \hat{x}_t^{t-1}), \tag{1.4}$$

$$P_t^t = P_t^{t-1} + P_t^{t-1} H_t' (H_t P_t^{t-1} H_t' + V_t)^{-1} H_t P_t^{t-1}, \tag{1.5}$$

where

$$\hat{x}_t^{t-1} = F_t \hat{x}_{t-1}^{t-1}, \tag{1.6}$$

$$P_t^{t-1} = F_t P_{t-1}^{t-1} F_t' + W_t \tag{1.7}$$

are predicted values constructed for time t at time $t - 1$.

The standard Kalman filter assumes normality of the disturbance vectors, i.e.

$$w_t \sim N(O, W_t) \quad , \quad v_t \sim N(O, V_t). \tag{1.8}$$

In this case it follows that

$$\hat{x}_t^t = E(x_t | Y^t) \tag{1.9}$$

so that $\hat{x}_t^t$ is the minimum variance estimator of the state.

However, in many practical situations two types of problem need to be addressed: (1) data are frequently distributed according to highly non-Gaussian densities or those that have heavier tails than the normal distribution so that robustification of the Kalman filter is necessary to protect the state estimator against outliers; (2) the relations in the basic state space model (1.1), (1.2) can be non-linear.

There have been many suggestions how to robustify the Kalman filter (see e.g. Cipra and Romera (1991), Peña and Guttman (1989) and other works). One of successful approaches to non-Gaussian filtering suggested by Masreliez (1975) for the linear model (1.1), (1.2) with non-normal observation disturbances v_t is based on the assumption that the predicted state density $p(x_t|Y^{t-1})$ (i.e. the conditional probability density of x_t under the condition Y^{t-1}) is approximately Gaussian with mean

$$\hat{x}_t^{t-1} = E(x_t | Y^{t-1}) \tag{1.10}$$

and covariance matrix

$$P_t^{t-1} = E\{(x_t - \hat{x}_t^{t-1})(x_t - \hat{x}_t^{t-1})' | Y^{t-1}\}. \tag{1.11}$$

Theorem 1 (Masreliez (1975))

Let the predicted state density $p(x_t|Y^{t-1})$ be Gaussian with mean $\hat{x}_t^{t-1}$ and covariance matrix P_t^{t-1}. Further let the predicted observation density $p(y_t|Y^{t-1})$ be twice differentiable. Then if follows that,

$$\hat{x}_t^t = \hat{x}_t^{t-1} + P_t^{t-1} H_t' g_t(y_t), \tag{1.12}$$

$$P_t^t = P_t^{t-1} - P_t^{t-1} H_t' G_t(y_t) H_t P_t^{t-1}, \tag{1.13}$$

where

$$g_t(y_t) = -[p(y_t|Y^{t-1})]^{-1} \frac{\partial p(y_t|Y^{t-1})}{\partial y_t}, \tag{1.14}$$

$$G_t(y_t) = \frac{\partial g_t(y_t)}{\partial y_t}. \tag{1.15}$$

This paper generalizes Masreliez's theorem to the case with a non-linear (and simultaneously non-Gaussian) observation relation that can be approximated sufficiently by its second-order Taylor expansion. Masreliez's theorem has proved to be useful in practice for robust time series analysis since, for example, it has motivated the so called ACM (Approximate Conditional-Mean) filters (see Martin (1979, 1981)). Therefore the generalization of the theorem due to Masreliez given in this paper may serve as one of the theoretical steps from linear to non-linear robust filtering.

2. Generalization of Masreliez's theorem

Let the observations y_t be scalar (i.e. $m = 1$) and let the observation relation have the form

$$y_t = H(x_t) + v_t, \tag{2.1}$$

where the function $H : R^n \to R^1$ can be approximated sufficiently by its second-order Taylor expansion about the point $\hat{x}_t^{t-1}$ so that

$$\begin{aligned} y_t = {} & H(\hat{x}_t^{t-1}) + H_x(\hat{x}_t^{t-1})'(x_t - \hat{x}_t^{t-1}) + \\ & + \frac{1}{2}(x_t - \hat{x}_t^{t-1})' H_{xx}(\hat{x}_t^{t-1})(x_t - \hat{x}_t^{t-1}) + v_t, \end{aligned} \tag{2.2}$$

where

$$H_x(\hat{x}_t^{t-1}) = \left(\frac{\partial H(\hat{x}_t^{t-1})}{\partial x_i}\right)_{i=1,\ldots,n} \tag{2.3}$$

is the $n \times 1$ vector of the first partial derivatives and

$$H_{xx}(\hat{x}_t^{t-1}) = \left(\frac{\partial^2 H(\hat{x}_t^{t-1})}{\partial x_i \partial x_j}\right)_{i,j=1,\ldots,n} \tag{2.4}$$

is the $n \times n$ matrix of the second partial derivatives of the function $H(x)$ at the point $x = \hat{x}_t^{t-1}$. (This possibility is also mentioned by West (1981, p. 159).)

Theorem 2

Let the observation relation (1.2) be replaced by (2.2) and let the assumptions of theorem 1 be satisfied. Then the difference $\hat{x}_t^t - \hat{x}_t^{t-1}$ can be found as the solution of the following system of differential equations

$$\frac{\partial}{\partial y_t}(\hat{x}_t^t - \hat{x}_t^{t-1}) + [B_t^{-t} - g_t(y_t)I](\hat{x}_t^t - \hat{x}_t^{t-1}) = \\ = g_t(y_t)H_{xx}(\hat{x}_t^{t-1})^{-1}H_x(\hat{x}_t^{t-1}) \tag{2.5}$$

under the initial condition

$$E_{y_t}(\hat{x}_t^t - \hat{x}_t^{t-1}|Y^{t-1}) = \int_{-\infty}^{\infty}(\hat{x}_t^t - \hat{x}_t^{t-1})p(y_t|Y^{t-1})dy_t = 0, \tag{2.6}$$

where I denotes the identity matrix and

$$B_t = P_t^{t-1}H_{xx}(\hat{x}_t^{t-1}). \tag{2.7}$$

Proof:

It follows from the Bayes' Law that,

$$\hat{x}_t^t - \hat{x}_t^{t-1} = \int_{R^n}(x_t - \hat{x}_t^{t-1})p(x_t|Y^t)dx_t =$$

$$= [p(y_t|Y^{t-1})]^{-1}P_t^{t-1}\int_{R^n}p(y_t|x_t)(P_t^{t-1})^{-1}(x_t - \hat{x}_t^{t-1})p(x_t|Y^{t-1})dx_t.$$

The density $p(x_t|Y^{t-1})$ is assumed to be Gaussian so that

$$\frac{\partial p(x_t|Y^{t-1})}{\partial x_t} = -(P_t^{t-1})^{-1}(x_t - \hat{x}_t^{t-1})p(x_t|Y^{t-1}).$$

Thus

$$\hat{x}_t^t - \hat{x}_t^{t-1} = -[p(y_t|Y^{t-1})]^{-1}P_t^{t-1}\int_{R^n}p(y_t|x_t)\frac{\partial p(x_t|Y^{t-1})}{\partial x_t}dx_t.$$

Integration by parts yields

$$\hat{x}_t^t - \hat{x}_t^{t-1} = [p(y_t|Y^{t-1})]^{-1}P_t^{t-1}\int_{R^n}\frac{\partial p(y_t|x_t)}{\partial x_t}p(x_t|Y^{t-1})\,dx_t.$$

Due to the form (2.2) of the observation relation it follows that

$$\hat{x}_t^t - \hat{x}_t^{t-1} = [p(y_t|Y^{t-1})]^{-1}P_t^{t-1}\int_{R^n}\frac{\partial p(y_t|x_t)}{\partial y_t}[-H_x(\hat{x}_t^{t-1}) -$$

$$- H_{xx}(\hat{x}_t^{t-1})(x_t - \hat{x}_t^{t-1})]p(x_t|Y^{t-1})dx_t =$$

$$= -[p(y_t|Y^{t-1})]^{-1}P_t^{t-1}\Big\{H_x(\hat{x}_t^{t-1})\frac{\partial}{\partial y_t}\int_{R^n} p(y_t|x_t)p(x_t|Y^{t-1})dx_t +$$

$$+ H_{xx}(\hat{x}_t^{t-1})\frac{\partial}{\partial y_t}\Big\{p(y_t|Y^{t-1})\int_{R^n}(x_t - \hat{x}_t^{t-1})p(y_t|x_t)p(x_t|Y^{t-1}) \times$$

$$\times [p(y_t|Y^{t-1})]^{-1}dx_t\Big\}\Big\} = -P_t^{t-1}H_x(\hat{x}_t^{t-1})[p(y_t|Y^{t-1})]^{-1}\frac{\partial p(y_t|Y^{t-1})}{\partial y_t} -$$

$$- P_t^{t-1}H_{xx}(\hat{x}_t^{t-1})[p(y_t|Y^{t-1})]^{-1}\frac{\partial}{\partial y_t}[p(y_t|Y^{t-1})(\hat{x}_t^t - \hat{x}_t^{t-1})]$$

since

$$\int_{R^n} p(y_t|x_t)p(x_t|Y^{t-1})dx_t = p(y_t|Y^{t-1}),$$

$$\int_{R^n}(x_t - \hat{x}_t^{t-1})p(x_t|Y^t)dx_t = \hat{x}_t^t - \hat{x}_t^{t-1}.$$

Using the relation (1.14), the last expression can be rewritten as

$$\hat{x}_t^t - \hat{x}_t^{t-1} = P_t^{t-1}H_x(\hat{x}_t^{t-1})g_t(y_t) + \\ + P_t^{t-1}H_{xx}(\hat{x}_t^{t-1})\left[g_t(y_t)(\hat{x}_t^t - \hat{x}_t^{t-1}) - \frac{\partial}{\partial y_t}(\hat{x}_t^t - \hat{x}_t^{t-1})\right], \tag{2.8}$$

which implies equation (2.5) holds.

The initial condition (2.6) is satisfied as

$$E_{y_t}(\hat{x}_t^t - \hat{x}_t^{t-1}|Y^{t-1}) = E_{y_t}\{E(x_t|Y^t) - E(x_t|Y^{t-1})|Y^{t-1}\} =$$
$$= E(x_t|Y^{t-1}) - E(x_t|Y^{t-1}) = 0.$$

Remark 1

If the observation relation is linear of the form (1.2) then $H_x(\hat{x}_t^{t-1}) = H_t'$ and $H_{xx}(\hat{x}_t^{t-1}) = 0$ so that (2.8) reduces to formula (1.12) from Masreliez's theorem.

Remark 2

Given the function $g_t(y_t)$, the system (2.5) can be solved by the methods recommended for linear non-homogenous systems of first order differential equations. In the special case where $n = 1$ the following explicit solution is obtained

$$\hat{x}_t^t = \begin{cases} \hat{x}_t^{t-1} + P_t^{t-1} H_x(\hat{x}_t^{t-1}) g_t(y_t) & \text{for } H_{xx}(\hat{x}_t^{t-1}) = 0, \\ \hat{x}_t^{t-1} + \dfrac{H_x(\hat{x}_t^{t-1})}{H_{xx}(\hat{x}_t^{t-1})} \left[B_t^{-1} \dfrac{\int_{-\infty}^{y_t} \exp(B_t^{-1} u) p(u|Y^{t-1}) du}{\exp(B_t^{-1} y_t) p(y_t|Y^{t-1})} - 1 \right] & \text{for } H_{xx}(\hat{x}_t^{t-1}) > 0, \\ \hat{x}_t^{t-1} + \dfrac{H_x(\hat{x}_t^{t-1})}{H_{xx}(\hat{x}_t^{t-1})} \left[-B_t^{-1} \dfrac{\int_{y_t}^{\infty} \exp(B_t^{-1} u) p(u|Y^{t-1}) du}{\exp(B_t^{-1} y_t) p(y_t|Y^{t-1})} - 1 \right] & \text{for } H_{xx}(\hat{x}_t^{t-1}) < 0, \end{cases} \tag{2.9}$$

where $p(u|Y^{t-1})$ denotes the predicted observation density in the integrand (see Cipra (1990)). Then again, theorem 2 can be reformulated for vector observations, i.e. for $m > 1$.

Remark 3

Using similar approach to that employed in the proof of theorem 2, the following linear non-homogenous system of second order differential equations for the covariance matrix P_t^t can be derived:

$$P_t^t - B_t \left\{ \frac{\partial^2}{\partial y_t^2} P_t^t - g_t(y_t) \frac{\partial}{\partial y_t} P_t^t - [G_t(y_t) - (g_t(y_t))^2] P_t^t \right\} B_t =$$

$$= P_t^{t-1} - P_t^{t-1} \Big\{ [G_t(y_t) - (g_t(y_t))^2] H_x(\hat{x}_t^{t-1}) H_x(\hat{x}_t^{t-1})' - g_t(y_t) H_{xx}(\hat{x}_t^{t-1}) -$$

$$- [p(y_t|Y^{t-1})]^{-1} \Big\{ H_{xx}(\hat{x}_t^{t-1}) \frac{\partial^2}{\partial y_t^2} [p(y_t|Y^{t-1})(\hat{x}_t^t - \hat{x}_t^{t-1})(\hat{x}_t^t - \hat{x}_t^{t-1})'] H_{xx}(\hat{x}_t^{t-1}) +$$

$$+ H_{xx}(\hat{x}_t^{t-1}) \frac{\partial^2}{\partial y_t^2} [p(y_t|Y^{t-1})(\hat{x}_t^t - \hat{x}_t^{t-1})] H_x(\hat{x}_t^{t-1})' + \tag{2.10}$$

$$+ H_x(\hat{x}_t^{t-1}) \frac{\partial^2}{\partial y_t^2} [p(y_t|Y^{t-1})(\hat{x}_t^t - \hat{x}_t^{t-1})'] H_{xx}(\hat{x}_t^{t-1}) \Big\}\Big\} P_t^{t-1} -$$

$$- (\hat{x}_t^t - \hat{x}_t^{t-1})(\hat{x}_t^t - \hat{x}_t^{t-1})'.$$

The procedure for the solution of the system (2.10) can be complicated even in simple cases. Therefore it would seem prudent to approximate its solution by the formula (1.13) from the linear case.

Although the obtained results are difficult to apply in practice, one can draw some conclusions from them in a similar way as in Masreliez (1975, pp. 108-109): (1) the formulae obtained in theorem 2 also stress the importance of the score function $g(y_t)$ for the predicted observation density $p(y_t|Y^{t-1})$ in non-linear non-Gaussian filtering (the explicit formula (2.9) with $n = 1$ uses the predicted observation density for the non-linear case in contrast to the application of its score function for the linear case) thereby showing a relation to the maximum likelihood method; (2) a qualitative role of non-Gaussian distribution is also indicated for non-linear filtering. The possible application of this theory to ACM filters is the topic of further investigations.

References

CIPRA, T. (1990): «Note on approximate non-Gaussian filtering with nonlinear observation relation», *Comment. Math., Univ. Carolinae*, 31, 601-605.

CIPRA, T., and ROMERA, R. (1991): «Robust Kalman filter and its application in time series analysis», *Kybernetika*, 27.

JAZWINSKI, A. H. (1970): *Stochastic Process and Filtering Theory*, New York, Academic Press.

MARTIN, R. D. (1979): «Robust estimation for time series autoregressions», *Robustness in Statistics* (R. L. Launer, and G. Wilkinson, eds.), New York, Academic Press, 147-176.

MARTIN, R. D. (1981): «Robust methods for time series», *Applied Time Series II* (D. F. Findley, ed.), New York, Academic Press, 683-759.

MASRELIEZ, C. J. (1975): «Approximate non-Gaussian filtering with linear state and observation relations», *IEEE Transactions on Automatic Control*, AC-20, 107-110.

PEÑA, D., and GUTTMAN, I. (1989): «Optimal collapsing of mixture distributions in robust recursive estimation», *Commun. Statist.-Theory Meth.*, 18, 817-833.

WEST, M. (1981): «Robust sequential approximate Bayesian estimation», *Journal of the Royal Statistical Society*, B43, 157-166.