Local linearization filters for non-linear continuous-discrete state space models with multiplicative noise

J. C. JIMENEZ†* and T. OZAKI‡

In this paper, the local linearization method for the approximate computation of the prediction and filtering estimates of continuous-discrete state space models is extended to the general case of non-linear non-autonomous models with multiplicative noise. The approximate prediction and filter estimates are obtained by applying the optimal linear filter to the piecewise linear state space model that emerges from a local linearization of both the non-linear state equation and the non-linear measurement equation. In addition, the solutions of the differential equations that describe the evolution of the first two conditional moments between observations are obtained, and an algorithm for their numerical computation is also given. The performance of the LL filters is illustrated by mean of numerical experiments.

1. Introduction

The estimation of the state of a continuous stochastic dynamical system from noisy discrete observations taken on the state is of central importance to solve diverse scientific and technological problems. The major contribution to the solution of this estimation problem is due to Kalman (1960, 1963) and Kalman and Bucy (1961), who provided a sequential and computational efficient solution to the linear filtering and prediction problems. However, the optimal non-linear estimation is still an active area of research. In particular, the basic need of obtaining a good approximation to the exact solution of the non-linear filtering and prediction problems remains and the subject continues to receive considerable attention.

In the last 40 years, a large variety of approximate non-linear filters have been considered (see Schwartz and Stear 1968, Lainiotis 1974, Liang 1983, Daum 1986, and references therein), e.g. the extended Kalman, the iterated extended Kalman, the Gaussian, the modified Gaussian, etc. It is well known that, even though these classical recursive filters are properly defined, they are computational unstable in the case of diverse types of non-linear problems. This unstability often becomes visible as a numerical explosion in the computation of the prediction or filtering estimates at an instant of time (Ozaki 1993). Recently, the novel projection and the particle filter methods have been extended from continuous to continuous-discrete state space models. The first one was only briefly introduced in Brigo et al. (1999) and further study is still required for its efficient implementation in practice. The second one, introduced in Del Moral et al. (2001), has strong theoretical support, but the filter is computationally expensive in practice because of the intrinsic complexity of the Monte Carlo method. Moreover, the filter could also be computational unstable in dependence of the numerical method used to simulate the state of the model.

To overcome the computational unstability of the classical recursive filters, in Ozaki (1993) an alternative approximate non-linear filter was introduced, not by approximating continuous-time filter equations as those cited in Schwartz and Stear (1968), Lainiotis (1974) and Liang (1983), but by applying the discrete-time filtering equations to an approximate discrete-time model for the original continuous-time dynamical system. The approximate discrete time model is obtained from a piece-wise linear discretization of the non-linear state equation. The approximate non-linear filter obtained in this way is called a local linearization (LL) filter. Following this approach, but using a higher order piece-wise linear discretization, Shoji (1998) introduced a LL filter that includes the original one as a particular case. Both filters are restricted to autonomous models with linear observation equations.

For the class of non-autonomous and non-linear continuous-discrete state space models a LL filter was proposed by Jimenez (1996). In this case, the LL filter is obtained approximating the expressions that define the extended Kalman filter for this type of model, being the LL discretization of the prediction equation the essential approximation. This LL filter reduces to the original one for autonomous models with linear observation equations. The high stability of LL filters is the key of their success in the solution of non-linear filtering problems for which other filtering algorithms fail (Ozaki 1993,
2. Optimal linear filter for linear state space models

Let the linear state space model defined by the continuous state equation

\[ \mathbf{d}x(t) = (A(t)x(t) + a(t))dt + \sum_{i=1}^{m} \mathbf{B}_i(t)x(t)dt + \mathbf{b}_i(t)d\mathbf{w}^i(t), \]

and the discrete observation equation

\[ \mathbf{z}_k = \mathbf{C}(t_k)x(t_k) + \sum_{i=1}^{n} \mathbf{D}_i(t_k)x(t_k)\xi^i_k + \mathbf{\eta}_k, \]

for \( k = 0, 1, \ldots, N \) (2)

where \( x(t) \in \mathbb{R}^d \) is the state vector at the instant of time \( t \), \( \mathbf{z}_k \in \mathbb{R}^r \) is the observation vector at the instant of time \( t_k \), \( \mathbf{w} \) is an \( m \)-dimensional Wiener process, and \( \{\xi^i_k : \xi^i_k \sim \mathcal{N}(0, A), A = \text{diag}(\lambda_1, \ldots, \lambda_m), k = 0, \ldots, N\} \) and \( \{\mathbf{\eta}_k : \mathbf{\eta}_k \sim \mathcal{N}(0, \mathbf{I}), k = 0, \ldots, N\} \) are sequences of random vectors i.i.d.

Let \( \mathbf{\hat{x}}_{i/\rho} = \mathcal{E}(x(t)/Z_{\rho}) \) and \( \mathbf{P}_{i/\rho} = \mathcal{E}((x(t) - \mathbf{\hat{x}}_{i/\rho})(x(t) - \mathbf{\hat{x}}_{i/\rho})^T)/Z_{\rho} \) for all \( \rho \leq t \), where \( \mathcal{E}(\cdot) \) denotes mathematical expectation and \( Z_{\rho} = \{\mathbf{z}_k : t_k \leq \rho\} \).

Suppose that \( \mathcal{E}(\mathbf{w}(t)\mathbf{w}^T(t)) = \mathbf{I}, \mathcal{E}(\mathbf{\xi}_k^i, \mathbf{\eta}_k) = \mathbf{0}^i(t_k), \mathbf{x}_{0/0} = \mathcal{E}(x(t_0))/Z_{t_0} < \infty \) and \( \mathbf{P}_{0/0} = \mathcal{E}((x(t_0) - \mathbf{x}_{0/0})(x(t_0) - \mathbf{x}_{0/0})^T)/Z_{t_0} < \infty \).

Theorem 1: (Jimenez and Ozaki, 2002 a): The optimal (minimum variance) linear filter for the linear model (1)–(2) consists of equations of evolution for the conditional mean \( \mathbf{\hat{x}}_{i/\rho} \) and covariance matrix \( \mathbf{P}_{i/\rho} \). Between observations, these satisfy the ordinary differential equation

\[ \mathbf{d}\hat{x}_{i/\rho} = (\mathbf{A}(t)\mathbf{\hat{x}}_{i/\rho} + \mathbf{a}(t))dt \]

\[ \mathbf{dP}_{i/\rho} = (\mathbf{A}(t)\mathbf{P}_{i/\rho} + \mathbf{P}_{i/\rho}\mathbf{A}^T(t))dt \]

\[ + \sum_{i=1}^{m} \mathbf{B}_i(t)(\mathbf{P}_{i/\rho} + \mathbf{\hat{x}}_{i/\rho}\mathbf{\hat{x}}_{i/\rho}^T)\mathbf{B}_i^T(t)dt \]

\[ + \sum_{i=1}^{m} (\mathbf{B}_i(t)^T\mathbf{\hat{b}}_{i/\rho}^T + \mathbf{b}_i(t)\mathbf{\hat{x}}_{i/\rho}^T)dt \]

\[ + \sum_{i=1}^{n} \mathbf{D}_i(t_k)\mathbf{\hat{x}}_{i+k/\rho} \mathbf{\hat{D}}_i^T(t_k) \]

for all \( t \in [t_k, t_{k+1}) \). At an observation at \( t_k \), they satisfy the difference equation

\[ \mathbf{\hat{x}}_{i+k/\rho} = \mathbf{x}_{i+k/\rho} + \mathbf{K}_{i+k/\rho}(z_{i+k} - \mathbf{C}(t_{k+1})\mathbf{\hat{x}}_{i+k/\rho}) \]

\[ \mathbf{P}_{i+k/\rho} = \mathbf{P}_{i+k/\rho} - \mathbf{K}_{i+k/\rho}\mathbf{C}(t_{k+1})\mathbf{P}_{i+k/\rho} \]

where

\[ \mathbf{K}_{i+k/\rho} = \mathbf{P}_{i+k/\rho} \mathbf{C}^T(t_{k+1})[\mathbf{C}(t_{k+1})\mathbf{P}_{i+k/\rho} \mathbf{C}^T(t_{k+1}) + \mathbf{I}]^{-1} \]

\[ + \sum_{i=1}^{m} \lambda_i \mathbf{D}_i(t_{k+1}) \mathbf{P}_{i+k/\rho} \mathbf{D}_i^T(t_{k+1})d\mathbf{w}^i(t_{k+1}) \]

\[ + \sum_{i=1}^{n} \mathbf{D}_i(t_{k+1}) \mathbf{\hat{x}}_{i+k/\rho} \mathbf{\hat{D}}_i^T(t_{k+1}) \]

\[ + \sum_{i=1}^{n} \mathbf{\hat{D}}_i(t_{k+1}) \mathbf{x}_{i+k/\rho} \mathbf{D}_i^T(t_{k+1}) \]
is the filter gain. The predictions \( \hat{x}_{i,p} \) and \( P_{i,p} \) are accomplished, respectively, via expressions (3) and (4) with initial conditions \( \hat{x}_{i,0} \) and \( P_{i,0} \), for \( \rho < t \).

Note that, for linear models (1)–(2) with \( a(t) \equiv 0 \) and \( B_i(t) \equiv D_i(t) \equiv 0 \) for all \( i = 1, \ldots , m \) and \( j = 1, \ldots , n \), the filter (3)–(6) reduces to the Kalman–Bucy filter for continuous-discrete models with additive noise.

3. The LL filters for non-linear state space models

Let the state space model defined by the continuous state equation

\[
\frac{\text{d}x(t)}{\text{d}t} = f(t, x(t)) + \sum_{i=1}^{m} g_i(t, x(t)) \text{d}w^i(t), \quad \text{for } t \geq t_0
\]

and the discrete observation equation

\[
z_k = h(t_k, x(t_k)) + \sum_{i=1}^{n} p_i(t_k, x(t_k)) g^i_{\alpha_k} + e_{\alpha_k},
\]

for \( k = 0, 1, \ldots , N \)

where \( f, g, h \) and \( p \) are non-linear functions, \( \{ \xi_{\alpha_k}, \zeta_{\alpha_k} \sim \mathcal{N}(0, \Sigma), \quad k = 0, \ldots , N \} \) and \( \{ e_i, \omega_i \sim \mathcal{N}(0, \Sigma), \quad k = 0, \ldots , N \} \) are sequences of random vectors i.i.d. and \( g^i_{\alpha_k} \) and \( e_{\alpha_k} \) are uncorrelated for all \( i \) and \( k \).

The functions \( f \) and \( g_i \) in (8) can be linearly approximated in two ways: by means of a truncated Taylor expansion or by a truncated Ito–Taylor expansion. In the first way, for example

\[
f(t, x(t)) \approx f(s, u) + \frac{\partial f(s, u)}{\partial s}(t-s) + J_f(s, u)(x(t) - u)
\]

while in the second one

\[
f(t, x(t)) \approx f(s, u) + \left( \frac{\partial f(s, u)}{\partial s} + \frac{1}{2} \sum_{k,j=1}^{d} [G(s, u)G^T(s, u)]_{k,j} \frac{\partial^2 f(s, u)}{\partial u^k \partial u^j} \right)(t-s)
\]

\[
+ J_f(s, u)(x(t) - u)
\]

where \( (s, u) \in \mathbb{R} \times \mathbb{R}^d \), \( J_f(s, u) \) is the Jacobian of \( f \) evaluated at the point \( (s, u) \), and \( G(s, u) \) is the \( d \times m \) matrix with entries \( [G(s, u)]_{k,j} = \frac{\partial g^i_{\alpha_k}}{\partial u^j} \). See Jimenez et al. (1999) for details.

Using these approximations for \( f \) and \( g_i \), the solution of the non-linear SDE (8) can be approximated by the solution of the piecewise linear SDE

\[
\text{d}y^\gamma(t) = (A(t_k, \hat{y}_{tk_{k+1}}^\gamma) + a^\gamma(t; t_k, \hat{y}_{tk_{k+1}}^\gamma)) \text{d}t
\]

\[
+ \sum_{i=1}^{m} (B_i(t_k, \hat{y}_{tk_{k+1}}^\gamma) + b^\gamma_i(t; t_k, \hat{y}_{tk_{k+1}}^\gamma)) \text{d}w^i(t)
\]

for all \( t \in \left[t_k, t_{k+1}\right) \), starting at \( y^\gamma(t_0) = \hat{y}_{0/0}^\gamma = \hat{x}_{0/0} \). Here, the symbol \( \gamma \) indicates the type of approximation used: \( \gamma = 1 \) for the Taylor approximation and \( \gamma = 1.5 \) for the Ito–Taylor. The remaining notations are \( \bar{y}_{i/\rho}^\gamma = \mathcal{E}(y^\gamma(t)/Z_{\rho}) \)

\[
A(s, u) = J_f(s, u)
\]

\[
B_p(s, u) = J_p(s, u)
\]

and \( a^\gamma \), \( b^\gamma \) are defined, respectively, by

\[
a^\gamma(t; s, u) = f(s, u) - J_f(s, u)u + \frac{\partial f(s, u)}{\partial s}(t-s)
\]

\[
b^\gamma_i(t; s, u) = g_i(s, u) - J_{g_i}(s, u)u + \frac{\partial g_i(s, u)}{\partial s}(t-s)
\]

for \( \gamma = 1 \), and by

\[
a^\gamma_{1.5}(t; s, u) = a^\gamma(t; s, u)
\]

\[
+ \frac{1}{2} \sum_{i,j=1}^{d} [G(s, u)G^T(s, u)]_{k,j} \frac{\partial^2 f(s, u)}{\partial u^k \partial u^j}(t-s)
\]

\[
b^\gamma_{1.5}(t; s, u) = b^\gamma_i(t; s, u)
\]

\[
+ \frac{1}{2} \sum_{i,j=1}^{d} [G(s, u)G^T(s, u)]_{k,j} \frac{\partial^2 g_i(s, u)}{\partial u^k \partial u^j}(t-s)
\]

for \( \gamma = 1.5 \).

Likewise, the functions \( h \) and \( p_i \) can be also linearly approximated. Specifically, let us consider

\[
h(s, v) \approx h(s, u) + J_h(s, u)(v - u)
\]

\[
p_i(s, v) \approx p_i(s, u) + J_{p_i}(s, u)(v - u)
\]

where \( (s, u) \in \mathbb{R} \times \mathbb{R}^d \) and \( J_h(s, u) \) and \( J_{p_i}(s, u) \) are, respectively, the derivative of \( h \) and \( p_i \) evaluated at the point \( (s, u) \). Using this approximation, the non-linear observation equation (9) could be rewritten as

\[
z_{k+1} = h(t_{k+1}, \hat{y}_{tk_{k+1}}^\gamma) + J_h(t_{k+1}, \hat{y}_{tk_{k+1}}^\gamma)(\hat{y}_{tk_{k+1}}^\gamma - \hat{y}_{tk_{k+1}}^\gamma)
\]

\[
+ \sum_{i=1}^{n} (p_i(t_{k+1}, \hat{y}_{tk_{k+1}}^\gamma) + \frac{\partial \hat{y}_{tk_{k+1}}^\gamma}{\partial x^i}(t_{k+1}, \hat{y}_{tk_{k+1}}^\gamma)) e_{tk_{k+1}}
\]

\[
+ r(t_{k+1}, x_{tk_{k+1}}^\gamma, \hat{y}_{tk_{k+1}}^\gamma, \hat{y}_{tk_{k+1}}^\gamma)
\]

for all \( k \), where \( \hat{y}_{tk_{k+1}}^\gamma \equiv \hat{y}^\gamma(t_{k+1}) \) and \( x_{tk_{k+1}} \equiv x(t_{k+1}) \) and
\[
\mathbf{r}(t_{k+1}, \mathbf{x}_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k}, \mathbf{z}^\gamma_{t_{k+1}/t_k}) = \mathbf{h}(t_{k+1}, \mathbf{x}_{k+1})
\]

Therefore, we have the following definition.

**Definition 1:** The local linearization filter for the non-linear state space model (8)–(9) is defined, between observations, by

\[
d\mathbf{y}^\gamma_{t_{j}/t_{j-1}} = (A(t_k, \mathbf{y}^\gamma_{t_{j}/t_k})\mathbf{y}^\gamma_{t_{j}/t_{j-1}} + a^*(t; t_k, \mathbf{y}^\gamma_{t_{j}/t_k})) dt
\]

\[
d\mathbf{P}^\gamma_{t_{j}/t_{j-1}} = \{A(t_k, \mathbf{y}^\gamma_{t_{j}/t_k})\mathbf{P}^\gamma_{t_{j}/t_{j-1}} + \mathbf{P}^\gamma_{t_{j}/t_{j-1}}A^T(t_k, \mathbf{y}^\gamma_{t_{j}/t_k})
\]

\[
+ \sum_{l=1}^{m} \mathbf{B}_i(t_k, \mathbf{y}^\gamma_{t_{j}/t_k})(\mathbf{P}^\gamma_{t_{j}/t_{j-1}} + \mathbf{y}^\gamma_{t_{j}/t_{j-1}}(\mathbf{y}^\gamma_{t_{j}/t_{j-1}})^T)\mathbf{B}_i^T(t_k, \mathbf{y}^\gamma_{t_{j}/t_k})
\]

\[
+ \sum_{l=1}^{m} \mathbf{b}_i(t_k, \mathbf{y}^\gamma_{t_{j}/t_k})(\mathbf{y}^\gamma_{t_{j}/t_{j-1}}(\mathbf{y}^\gamma_{t_{j}/t_{j-1}})^T)B_i^T(t_k, \mathbf{y}^\gamma_{t_{j}/t_k})
\]

\[
+ \sum_{l=1}^{m} b_i^T(t_k, \mathbf{y}^\gamma_{t_{j}/t_k})(\mathbf{y}^\gamma_{t_{j}/t_{j-1}}(\mathbf{y}^\gamma_{t_{j}/t_{j-1}})^T)B_i(t_k, \mathbf{y}^\gamma_{t_{j}/t_k})\}
\]

for all \( t \in [t_k, t_{k+1}) \), and by

\[
\mathbf{y}^\gamma_{t_{k+1}/t_{k+1}} = \mathbf{y}^\gamma_{t_{k+1}/t_k} + \mathbf{K}^\gamma_{t_{k+1}/t_k} \mathbf{z}_{t_{k+1}} - \mathbf{h}(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k})
\]

\[
\mathbf{P}^\gamma_{t_{k+1}/t_{k+1}} = \mathbf{P}^\gamma_{t_{k+1}/t_k} - \mathbf{K}^\gamma_{t_{k+1}/t_k} \mathbf{J}_h(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k})\mathbf{P}^\gamma_{t_{k+1}/t_k} / \text{for all } \mathbf{y}_{t_{j}/t_{j-1}}^\gamma \text{ and } \mathbf{P}_{t_{j}/t_{j-1}}^\gamma \text{ are accomplished, respectively, via expressions (16) and (17) with initial conditions } \mathbf{y}_{t/0}^\gamma \text{ and } \mathbf{P}_{t/0}^\gamma \text{ for } \rho < t. \]

**K**_{t_{k+1}/t_{k+1}} = \mathbf{P}^\gamma_{t_{k+1}/t_{k+1}} / \mathbf{J}_h(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k}) + \mathbf{P}^\gamma_{t_{k+1}/t_{k+1}}

\[
\times (\mathbf{J}_h(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k})\mathbf{J}_h^T(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k}) + \mathbf{P}^\gamma_{t_{k+1}/t_{k+1}})
\]

\[
+ \sum_{i=1}^{n} \lambda_i \mathbf{J}_p(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k})
\]

\[
\times (\mathbf{P}^\gamma_{t_{k+1}/t_k} + \mathbf{y}^\gamma_{t_{k+1}/t_k}(\mathbf{y}^\gamma_{t_{k+1}/t_k})^T)\mathbf{J}_p^T(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k})
\]

\[
+ \sum_{i=1}^{n} J_p(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k})\mathbf{y}^\gamma_{t_{k+1}/t_k}(\mathbf{y}^\gamma_{t_{k+1}/t_k})^T\mathbf{J}_p^T(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k})
\]

\[
+ \sum_{i=1}^{n} \theta_i(t_{k+1})(\mathbf{y}^\gamma_{t_{k+1}/t_k})^T \mathbf{J}_p(t_{k+1}, \mathbf{y}^\gamma_{t_{k+1}/t_k})^{-1}
\]

and \( \mathbf{P}_{t_{j}/t_{j-1}}^\gamma = \mathcal{E}((\mathbf{y}^\gamma(t) - \tilde{\mathbf{y}}_{t_{j}/t_{j-1}})(\mathbf{y}^\gamma(t) - \tilde{\mathbf{y}}_{t_{j}/t_{j-1}})^T/Z_{t_{j}}) \) for \( \rho \leq t \).

For linear state space models with multiplicative noise, the LL filter (16)–(19) reduces to the optimal linear filter (3)–(6). While, for linear models with additive noise and \( a(t) \equiv 0 \) and \( \mathbf{B}_i(t) \equiv \mathbf{D}_i(t) \equiv 0 \) (for all \( i = 1, \ldots, m \) and \( j = 1, \ldots, n \)), the LL filter (16)–(19) reduces to the Kalman–Bucy filter.

Besides, for non-linear state space models with additive noise this filter with \( \gamma = 1 \) and \( \gamma = 1.5 \) reduces, respectively, to the LL filter introduced by Ozaki (1993) and by Shoji (1998) for autonomous models with linear observation equations and functions \( \mathbf{f} \) with non-singular Jacobian. With \( \gamma = 1 \), the LL filter (16)–
(19) reduces to the LL filter introduced by Jimenez (1996) for non-autonomous and non-linear models.

4. Solution of the equations for the first two conditional moments between observations

This section deals with the problem of solving the differential equations (16) and (17) that describe the evolution of the predictions \( \hat{x}_{i/t_k} \) and \( P_{i/t_k} \), respectively. Here, the symbols \( \text{vec} \), \( \otimes \) and \( \circ \) will denote the vectorization operator, the Kronecker sum and product, respectively (see Magnus and Neudecker 1988, for definitions).

**Theorem 2:** Let \( A \) and \( B \) be constant matrices. Let \( a(t) = a_0 + a_1 t \) and \( b(t) = b_{0,i} + b_{1,i} t \), where \( a_0, a_1, b_{0,i} \) and \( b_{1,i} \) are constant vectors. The general solution of the differential equations

\[
\begin{align*}
\text{d} \hat{x}_{i/t_k} &= (A \hat{x}_{i/t_k} + a(t)) \text{d}t \\
\text{d} P_{i/t_k} &= \bigg\{ AP_{i/t_k} + P_{i/t_k} A^\top + \sum_{i=1}^{m} b_i(t) b_i^\top (t) \\
&+ \sum_{i=1}^{m} B_i (P_{i/t_k} + \hat{x}_{i/t_k} \hat{x}_{i/t_k}^\top) B_i^\top \\
&+ \sum_{i=1}^{m} (B_i \hat{x}_{i/t_k} b_i^\top (t) + b_i(t) \hat{x}_{i/t_k} \hat{x}_{i/t_k}^\top B_i^\top) \bigg\} \text{d}t
\end{align*}
\]

is given by

\[
\begin{align*}
\hat{x}_{i/t_k} &= \hat{x}_{i/t_k} + \int_{t_k}^{t} \exp(A(t-t_k-s))(A \hat{x}_{i/t_k} + a(s+t_k)) \text{d}s \\
\text{vec}(P_{i/t_k}) &= \exp(A(t-t_k)) \\
&\left( \text{vec}(P_{i/t_k}) + \sum_{j=1}^{8} \int_{t_k}^{t} \exp(-As) B_j \exp(A_j s) C_j e_j(s) \text{d}s \right)
\end{align*}
\]

for all \( t \geq t_k \), where \( A = A \otimes I + \sum_{i=1}^{m} B_i \otimes B_i \), and \( A_j, B_j, C_j \) and \( e_j \) are the matrices defined in table 1.

The following lemma will be useful to demonstrate the above theorem.

**Lemma 1** (Van Loan 1978): Let \( A_1, A_2, A_3 \) and \( A_4 \) be square matrices, \( n_1, n_2, n_3 \) and \( n_4 \) be positive integers, and set \( m \) to be their sum. If the \( m \times m \) block triangular matrix \( C \) is defined by

\[
C = \begin{bmatrix} A_1 & B_1 & C_1 & D_1 \\ 0 & A_2 & B_2 & C_2 \\ 0 & 0 & A_3 & B_3 \\ 0 & 0 & 0 & A_4 \end{bmatrix}
\]

then for \( s \geq 0 \)

\[
\begin{bmatrix} F_1(s) & G_1(s) & H_1(s) & K_1(s) \\ 0 & F_2(s) & G_2(s) & H_2(s) \\ 0 & 0 & F_3(s) & G_3(s) \\ 0 & 0 & 0 & F_4(s) \end{bmatrix} = \exp(sC)
\]

where

\[
\begin{align*}
F_j(s) &= \exp(A_j s), \quad j = 1, 2, 3, 4 \\
G_j(s) &= \int_{0}^{\infty} \exp(A_j(s - u)) B_j \exp(A_{j+1} u) \text{d}u, \quad j = 1, 2, 3 \\
H_j(s) &= \int_{0}^{\infty} \exp(A_j(s - u)) C_j \exp(A_{j+1} u) \text{d}u \\
K_j(s) &= \int_{0}^{\infty} \exp(A_1(s - u)) D_1 \exp(A_4 u) \text{d}u \\
&+ \int_{0}^{\infty} \exp(A_1(s - u)) [C_1 \exp(A_3(u - r)) B_3 \\
&+ B_1 \exp(A_2(u - r)) C_2] \exp(A_4 r) \text{d}r \text{d}u \\
&+ \int_{0}^{\infty} \int_{0}^{\infty} \exp(A_1(s - u)) B_1 \exp(A_2(u - r)) B_2 \\
&\times \exp(A_3(r - w)) B_3 \exp(A_4 w) \text{d}w \text{d}r \text{d}u
\end{align*}
\]

**Proof of Theorem 2:** The general solution of the linear equation (21) is very well known (Bellman 1970). It is

\[
\begin{align*}
\hat{x}_{i/t_k} &= \exp(A(t-t_k)) \left( \hat{x}_{i/t_k} + \int_{t_k}^{t} \exp(-A(s-t_k)) a(s) \text{d}s \right) \\
&= \exp(A(t-t_k)) \left( \hat{x}_{i/t_k} + \int_{0}^{t-t_k} \exp(-As) a(s+t_k) \text{d}s \right)
\end{align*}
\]

The expression (23) is obtained by replacing \( a(s+t_k) \) in the above expression by \( a(s+t_k) + A \hat{x}_{i/t_k} - A \hat{x}_{i/t_k} \), and by using the identity

\[
\int_{0}^{t-t_k} \exp(-As) \text{d}A = -(\exp(-A(t-t_k)) - 1),
\]

where \( I \) is the identity matrix.
Applying the operator \( \text{vec} \) to the equation (22) it is obtained the linear equation
\[
d\text{vec}(\mathbf{P}_{t_{i/k}}) = (\mathbf{B}(t) + \mathbf{A} \text{vec}(\mathbf{P}_{t_{i/k}}))dt
\]
whose solution is given by the expression
\[
\text{vec}(\mathbf{P}_{t_{i/k}}) = \exp(\mathbf{A}(t_{i/k})) \left( \text{vec}(\mathbf{P}_{t_{j/k}}) + \int_{0}^{t_{i/k}} \exp(-\mathbf{A}s) \mathbf{B}(s+t_{i/k})ds \right)
\]
(27)

where
\[
\mathbf{A} = \mathbf{A} \oplus 1 + \sum_{i=1}^{m} \mathbf{B}_{j} \otimes \mathbf{B}_{j}
\]
\[
\mathbf{B}(t) = \sum_{i=1}^{m} (\mathbf{B}_{j} \otimes \mathbf{B}_{j}) \text{vec}(\mathbf{x}_{t_{i/k}}^\top \mathbf{x}_{t_{i/k}}) + \sum_{i=1}^{m} \text{vec}(\mathbf{b}_{j}(t)) \mathbf{b}_{j}(t)
\]

Table 1. Values of \( \mathcal{A}_{j}, \mathcal{B}_{j}, \mathcal{C}_{j} \) and \( e_{j}(s) \) of the expression (24). Here

\[
\begin{align*}
\mathbf{C} &= \begin{bmatrix} \mathbf{A} & \mathbf{a}_{1} & \mathbf{A} \mathbf{x}_{t_{i/k}} + \mathbf{a}_{i}(t_{i/k}) \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{(d+2) \times (d+2)} \\
\mathbf{L}^\top &= [\mathbf{I}_{d} \hspace{1em} 0_{d \times 2}] \in \mathbb{R}^{d \times (d+2)}, \quad \mathbf{R}^\top = [0_{1 \times (d+1)} \hspace{1em} 1] \in \mathbb{R}^{1 \times (d+2)} \text{ and } \mathbf{I}_{m} \text{ is the } m \times m \text{ identity matrix.}
\end{align*}
\]

Rewriting the expression (23) as
\[
\dot{\mathbf{x}}_{t_{i/k}} = \mathbf{x}_{t_{i/k}} + \int_{0}^{t_{i-k}} \exp(\mathbf{A}(t - t_{k} - s))ds \left( \mathbf{A} \mathbf{x}_{t_{i/k}} + \mathbf{a}(t_{i/k}) \right)
\]
\[
\times (\mathbf{A} \dot{\mathbf{x}}_{t_{i/k}} + \mathbf{a}(t_{i/k})) + \int_{0}^{t_{i-k}} \int_{0}^{t_{i-k}} \exp(\mathbf{A}(t - t_{k} - s))du \mathbf{a}_{1} d\mathbf{x}_{t_{i/k}}
\]

and using Lemma 1 it is obtained that
\[
\dot{\mathbf{x}}_{t_{i/k}} = \mathbf{x}_{t_{i/k}} + \mathbf{L}^\top \exp((t - t_{k}) \mathbf{C}) \mathbf{R}
\]
(28)

where
\[
\begin{align*}
\mathbf{C} &= \begin{bmatrix} \mathbf{A} & \mathbf{a}_{1} & \mathbf{A} \mathbf{x}_{t_{i/k}} + \mathbf{a}(t_{i/k}) \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{(d+2) \times (d+2)} \\
\mathbf{L}^\top &= [\mathbf{I}_{d} \hspace{1em} 0_{d \times 2}] \in \mathbb{R}^{d \times (d+2)} \text{ and } \mathbf{R}^\top = [0_{1 \times (d+1)} \hspace{1em} 1] \in \mathbb{R}^{1 \times (d+2)}.
\end{align*}
\]
Substituting (28) in (27), and after some algebraic manipulations the expression (24) is obtained.

5. Numerical computation of the prediction estimates

In the previous section, expressions for the predictions were obtained. However, in general, these expressions are difficult to implement in practice by means of a computer program. In this section, simple algebraic expressions are introduced as a computational feasible alternative to (23) and (24).

**Theorem 3:** The expression (23) and (24) are equivalent to

\[
\hat{X}_{t/t_k} = \hat{X}_{t_{k-1}} + \Psi(t)
\]

and

\[
\text{vec}(P_{t/t_k}) = F_1(t) \text{vec}(P_{t_{k-1}}) + \sum_{j=1}^{4} G_j(t) C_j + K_5(t)
\]

respectively. Here, \( \Psi \) is defined by the matrix identity

\[
\begin{bmatrix}
F(t) & g_1(t) & h(t) \\
0 & 1 & g_2(t) \\
0 & 0 & 1
\end{bmatrix}
= \exp((t-t_k)C)
\]

with

\[
C = \begin{bmatrix}
A & a_1 & A\hat{x}_{t_{k-1}} + a(t_k) \\
0 & 0 & 1 \\
0 & 0 & 0
\end{bmatrix}
\in \mathbb{R}^{(d+2) \times (d+2)}
\]

\( F_1, G_j \) by

\[
F_j(t) = \begin{bmatrix} F_j(t) & G_j(t) \\
0 & F_j^t(t) \end{bmatrix} = \exp((t-t_k)J_j),
\]

with \( J_j = \begin{bmatrix} A & B_j \\
0 & A_j \end{bmatrix} \) for \( j = 1, \ldots, 4 \)

\( H_5 \) by

\[
F_5(t) = \begin{bmatrix} F_5(t) & G_5(t) & H_5(t) \\
0 & F_5^t(t) & G_5^t(t) \\
0 & 0 & F_5^{tt}(t) \end{bmatrix}
= \exp((t-t_k)J_5),
\]

with \( J_5 = \begin{bmatrix} A & B_5 & 0 \\
0 & B_5 & I \\
0 & 0 & B_5 \end{bmatrix} \)

and \( K_5 \) by

\[
F_5(t) = \begin{bmatrix} F_5(t) & G_5(t) & H_5(t) & K_5(t) \\
0 & F_5^t(t) & G_5^t(t) & H_5^t(t) \\
0 & 0 & F_5^{tt}(t) & G_5^{tt}(t) \\
0 & 0 & 0 & F_5^{ttt}(t) \end{bmatrix}
= \exp((t-t_k)J_5), \text{ with } J_5 = \begin{bmatrix} A & B_5 & B_7 & B_9 \\
0 & 0 & 2 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 \end{bmatrix}
\]

The matrices \( A, A_j, B_j \) and \( C_j \) are defined as in Theorem 2.

**Proof:** The expression (29) is straightly derived by re-writing (23) as

\[
\hat{X}_{t_{k-1}} = \hat{X}_{t_{k-1}} + \int_{t_{k-1}}^{t_k} \exp(A(t-t_s)) \Psi(t_s) \Psi^t(t_s) \Psi(t_s) ds
\]

with \( \Psi \) defined as above.

By setting \( t_s = t_{k-1}, A_1 = A, B_1 = a_1, B_2 = 1, C_1 = A\hat{x}_{t_{k-1}} + a(t_k) \) and \( A_2 = A_1 \) in Lemma 1 it is obtained that

\[
\int_{t_{k-1}}^{t_k} \exp(A(t-t_s)/C_0) ds = G_j(t) C_i
\]

for all \( i = 1, \ldots, 4 \). Obviously, \( \exp((t-t_k)A) = F_1(t) \).

Likewise, using the matrix identities

\[
\int_{t_{k-1}}^{t_k} \exp(A(t-t_s)/C_0) B_1 exp(A_s/s) \exp(A_s/s) ds = G_j(t) C_i
\]

and Lemma 1, the last two terms of expression (30) are obtained.

In this way, the numerical computation of the conditional moments \( \hat{X}_{t_{k-1}} \) and \( P_{t_{k-1}} \) is reduced to use a con-
convenient algorithm to compute matrix exponentials, e.g. those based on rational Padé approximations (Golub and Van Loan 1996), the Schur decomposition (Golub and Van Loan 1996) or Krylov subspace methods (Hochbruck and Lubich 1997). For a recent review see Sidje (1998). The selection of one of them will mainly depend of the size and structure of the matrices $C$ and $J_i$. In many cases, it is enough to use the algorithm developed by Van Loan (1978), which takes advantage of the special structure of these matrices. However, for state space models that involve more that four state equation the matrices $J_i$ become large. In that case, the Krylov subspace methods provide the more efficient and accurate algorithms to compute matrix exponentials. See Jimenez (2002) for more details.

Moreover, in the case of autonomous state space models simpler expressions than (29)–(30) can be used to evaluate the solution of (16)–(17). That is:

**Theorem 4** (Jimenez and Ozaki, 2002a): For equations (21)–(22) with $a(t) = a$ and $b_i(t) = b_i$, the expressions (23) and (24) are equivalent to

$$x_{t/t} = x_{t/0} + g(t)$$

and

$$vec(P_{t/t}) = F(t) vec(P_{t/0}) + \sum_{j=1}^{5} G_j(t) C_j$$

respectively. Here, the vector $g$ is defined by the matrix identity

$$\begin{bmatrix} F(t) & g(t) \\ 0 & 1 \end{bmatrix} = \exp((t - t_k)C)$$

with

$$C = \begin{bmatrix} A & A x_{t/0} + a \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{(d+1) \times (d+1)}$$

and the matrices $F(t)$, $G_j$ by

$$\begin{bmatrix} F_j(t) & G_j(t) \\ 0 & F_j'(t) \end{bmatrix} = \exp((t - t_k)J_i), \text{ with } J_i = \begin{bmatrix} A & B_j \\ 0 & A_j \end{bmatrix}$$

and $A = A \oplus 1 + \sum_{i=1}^{m} B_i \otimes B_i$. The matrices $A_j$, $B_j$, and $C_j$ are defined in Table 2.†

### 6. Numerical experiments

In this section, the performance of the LL filters introduced above is illustrated by means of some numerical experiments. Specifically, the LL filter estimates $x_{t/0}$ and $P_{t/0}$ are compared with the exact first two conditional moments of the solution of two non-linear state space models with multiplicative noise.

For $\gamma = 1$, $\dot{x}_{t/0}$ and $P_{t/0}$ are computed by the equations (18) and (19); where the predictions $\dot{x}_{t/0}$ and $P_{t/0}$, defined by (16) and (17), are computed by means of the expressions (31) and (32).

**Example 1:** Let the non-linear state space model with multiplicative noise

$$dx_1 = (\alpha + \theta x_1)dt + \rho \sqrt{x_1}d\omega_1$$

$$dx_2 = \beta x_2^2dt + \gamma x_2^2d\omega_2$$

$$z_k = x_2(t_k) - 0.001x_2^3(t_k) + (x_2(t_k))$$

$$- 0.01x_2^2(t_k)\xi_{t_k} + e_{t_k}$$

where $\alpha = 0.0025$, $\theta = -0.0175$, $\rho = 0.1$, $\beta = -0.01$, $\gamma = 0.05$, $\{\xi_{t_k}\}$ and $\{e_{t_k}\}$ are sequences of random vectors i.i.d; and $E(\xi_{t_k}) = 0$ for all $k$.

Equation (33) has been well studied in the framework of financial modelling by Cox et al. (1985). The exact first two conditional moments of the solution of this equation are also very well known (Bibby and Sørensen 1995).

Figure 1 shows a realization of the solution of the state equations (33)–(34) computed by the explicit Euler–Maruyama scheme (Kloeden and Platen 1995) at the instant of times $t_j = j \Delta$, with $\Delta = 0.001$ and $j = 0, \ldots, 4000$.

Figure 2 shows a realization of the observed data (35) with $\lambda^2 = 0$ and $\sigma^2 = 1$. Figure 3 (top) and (bottom) present, respectively, the first and second exact conditional moments of the variable $x_1$ together with the estimates $\dot{x}_{t/0}$ and $P_{t/0}$ for that variable. $\dot{x}_{t/0}$ and $P_{t/0}$ were computed from the set of observations $\{z_{t_k}\}_{k=0}^{N}$, where $t_k = kh$, $h = 0.01$ and $N = 400$.

Figure 4 shows a realization of the observed data (35) with $\lambda^2 = 0.001$ and $\sigma^2 = 0$ is displayed. Figure 5 (top) and (bottom) shows, respectively, the first and second exact conditional moments of the variable $x_1$ together with the estimates $\dot{x}_{t/0}$ and $P_{t/0}$ for that variable.

Figure 6 presents a realization of the observed data (35) with $\lambda^2 = 0.001$ and $\sigma^2 = 1$. Figure 7 (top) and (bottom) shows, respectively, the first and second exact conditional moments of the variable $x_1$ together with the estimated $\dot{x}_{t/0}$ and $P_{t/0}$ for that variable.

Note that, in all cases, there is no significant difference between the exact and the estimated values of the first two conditional moments.

**Example 2:** Consider the non-linear state space model with multiplicative noise

†The matrices $A$, $A_1$, $B_1$ and $B_2$ differ from those that appear in Jimenez and Ozaki (2002a) because of some print mistakes in that paper.
\[
\begin{array}{cccc}
 j & A_j & B_j & C_j \\
 1 & C \otimes I_{d+1} & \sum_{i=1}^{m} (B_i \otimes B_i)(L_i^T \otimes L_i^T) & \text{vec}(R) \\
 2 & I_j \otimes C & \sum_{i=1}^{m} (B_i \otimes B_i)(I_i \otimes L_i) & \text{vec}(R x_i^T) \\
 3 & C \otimes I_j & \sum_{i=1}^{m} (B_i \otimes B_i)(L_i^T \otimes I_i) & \text{vec}(\mathbf{x}_{i/\tau}) \\
 4 & C \otimes I_{d+1} & \sum_{i=1}^{m} (B_i \otimes B_i + B_j \otimes B_j)(L_i^T \otimes R) & \text{vec}(I_{d+1}) \\
 5 & 0 & \sum_{i=1}^{m} (B_i \otimes B_i) \text{vec}(\mathbf{x}_{i/\tau}^T) + \sum_{j=1}^{m} (B_j \otimes B_j + B_j \otimes B_j) \text{vec}(\mathbf{x}_{i/\tau}) + \text{vec}(b_{j}^T b_{j}) & 1 \\
\end{array}
\]

Table 2. Values of $A_j$, $B_j$, $C_j$ and of the expression (32). Here

\[
C = \begin{bmatrix} A & \mathbf{A} \mathbf{x}_{i/\tau} + \mathbf{a}(t_k) \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{(d+1) \times (d+1)}
\]

\[
L_i = \begin{bmatrix} I_i & 0 \end{bmatrix}_{d \times 1} \in \mathbb{R}^{d \times (d+1)}; \quad R_i = \begin{bmatrix} 0_{1 \times d} & 1 \end{bmatrix} \in \mathbb{R}^{(d+1) \times (d+1)} \quad \text{and} \quad I_m \text{ is the } m \times m \text{ identity matrix.}
\]

Figure 1. Realization of the solution of the state equations (33)-(34) computed by the explicit Euler–Maruyama scheme. Top: $x_1$; Bottom: $x_2$.

Figure 2. Realization of the observed equation (35) with $\lambda^2 = 0$ and $\sigma^2 = 1$.

\[
\begin{align*}
\text{dx}_1 &= \mu dt + \exp(x_2/2)dw_1 \\
\text{dx}_2 &= (\alpha + \beta x_2)dt + \gamma dw_2 \\
\text{z}_{t_k} &= \text{x}_1(t_k)
\end{align*}
\]

(36) (37) (38)

Figure 3. First (top) and second (bottom) exact conditional moments of the variable $x_1$ together their estimated $x_{i/\tau}^T$ (top) and $P_{i/\tau}$ (bottom), respectively, given the observation of figure 2.

Figure 4. Realization of the observed equation (35) with $\lambda^2 = 0.001$ and $\sigma^2 = 0$. 

where $\mu = 0.0117$, $\alpha = -0.0175$, $\beta = 0.0211$ and $\gamma = 0.0110$.

The model (36)–(37) has also been well studied in the framework of the option pricing modelling (Scott 1987, Wiggins 1987, Chasney and Scott 1989, Ozaki and Iino 2000). In this model, the exact first two conditional moments of the solution (37) are very easy to compute.

Figure 8 shows a realization of the solution of the state equations (36)–(37) computed by the explicit Euler–Maruyama scheme. Top: $x_1$; Bottom: $x_2$.

Figure 9 presents the first and second exact conditional moments of the variable $x_2$ together with the estimated $\hat{x}_{n/\tau}$ (top) and $P_{n/\tau}$ (bottom) for that variable, given the observation of the variable $x_1$.
Note that this example simulates the estimation of the non-observed component $x_2$ of the volatility model (36)–(37) given 10 daily observations of a currency exchange rate $x_1$ during 4 years.

7. Conclusions

In this paper the local linearization method for the approximate computation of the prediction and filtering estimates of continuous-discrete state space models was extended to the general case of non-linear non-autonomous models with multiplicative noise. An effective algorithm for numerical implementation of the prediction and filter estimates was also introduced and the performance of the method was illustrated by means of simulations.

At this point, it is opportune to point out that the approximate filtering method introduced here constitutes the main component of the approximate innovation method that has been recently proposed for the parameter estimation of discretely observed SDEs with multiplicative noise (Jimenez and Ozaki 2002 b). That inference method has allowed an adequate estimation of the non-observed states and the unknown parameters of continuous-time stochastic volatility models given a time series of financial data (Ozaki and Jimenez 2002) and, in that way, a satisfactory monitoring and control of time series of financial data (Ozaki and Jimenez 2002).

Acknowledgements

The authors are very grateful to the Ministry of Education, Science and Culture of Japan and, also, to the Mizuho Trust and Banking Co. Ltd for the partial support of this paper.

References


Ozaki, T., and Inou, M., 2000, An innovation approach to non-Gaussian time series analysis. In G. Kitagawa and T.


