THE LOCAL LINEARIZATION FILTER WITH APPLICATION TO NONLINEAR SYSTEM IDENTIFICATIONS

Tohru OZAKI
Institute of Statistical Mathematics
4-6-7, Minami Azabu, Minato-ku,
Tokyo, Japan-106

Some key words: Nonlinear filter, system identification, maximum likelihood method, nonlinear dynamical systems, local linearization, Markov diffusion process, Markov chain, nonlinear random vibrations, van der Pol oscillations, innovations, nonlinear state-space representation.

1. Introduction.

Many dynamic phenomena in scientific fields are modeled by a continuous time stochastic dynamical system,

\[ \dot{z} = f(z, t; \theta) + w(t), \]

where \( w(t) \) is a Gaussian white noise. The process \( z(t) \) defined by the model is a Markov diffusion process. When we use such mathematical models in physics or engineering sciences we often have a rough idea, from some physical understanding of the process, of the size of the parameters \( \theta \) and the variance-covariance matrix \( \Sigma \) of the Gaussian white noise \( w(t) \) in the model. However when prediction or control of the process is concerned we sometime need more accurate estimates of the parameters. In some cases, we don't have much physical information helping us to guess the size of the parameters in the model. In such situations we often try to estimate parameters from observation data of the phenomenon as accurately as possible.

We consider, in this paper, the estimation problem for continuous time stochastic dynamical system models with discrete time observation data using a sufficiently small sampling time interval. The dimension \( r \) of the observation data \( x_t \) is sometimes less than the dimension \( k \) of the state \( z(t) \). Also we sometimes cannot ignore the observation errors. Thus the estimation problem concerned is that of the following continuous time state-space model with discrete time observation model,

\[ \dot{z} = f(z, t; \theta) + w(t) \]
\[ x_t = C z(t) + \nu_t \tag{1.1} \]

where \( C \) is a \( r \times k \) observation matrix, e.g.
\[ C = \begin{pmatrix} 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix} \]

If the \( k \)-dimensional stochastic dynamical system is linear, i.e.

\[ \dot{z} = L(\theta) z + w(t), \tag{1.2} \]

with some \( z \)-free matrix \( L(\theta) \), we know that the following discrete time stochastic dynamical system model,

\[ Z_{t+\Delta t} = A Z_t + B W_{t+\Delta t} \]

is a natural and good approximation to (1.2), where

\[ A = e^{L(\theta)\Delta t} = \sum_{k=0}^{\infty} \frac{(\Delta t)^k}{k!} L(\theta)^k \]

and \( B \) is a matrix whose elements are functions of eigenvalues of the matrix \( A \) (see Ozaki(1986)). The process \( W_{t+\Delta t} \) is a discrete time Gaussian white noise of variance \( \Delta t \Sigma \).

Thus in linear cases we have the following discrete time state-space model,

\[ Z_{t+\Delta t} = A Z_t + B W_{t+\Delta t} \]

\[ X_t = C Z_t + \eta_t \tag{1.3} \]

When \( \eta_t \) is zero, it is known that \( X_t \) has ARMA(k,k-1) representation(Akaike(1974)),

\[ X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_k X_{t-k} = C B W_t + C(A - \phi_1 I) B W_{t-1} + \ldots + C(A^{k-1} - \phi_1 A^{k-2} - \ldots - \phi_{k-1} I) B W_{t-k+1} \tag{1.4} \]

where \( \phi_i \)'s are the characteristic coefficients of the matrix \( A \). Therefore we can estimate the parameters of the model for \( z(t) \) by the maximum likelihood method for the linear ARMA time series models without an argument of state space representation models. We note that in (1.4) the dimension of the variable \( X_t(t=1,2,\ldots N) \) can be less than that of \( z(t) \). This is one of the reason why ARMA time series models with high orders are so powerful in many application fields where only scalar time series observation is available and the dimension of its state dynamics could be much higher. When the observation noise \( \eta_t \) is present we have ARMA(k,k) model instead of ARMA(k,k-1) model because \( X_t \) \( (i=t, t-1, \ldots, t-k) \) in (1.4) is replaced by \( X_t - \eta_t \). Thus we can still estimate the parameters by the maximum likelihood method for time series models although the parametric expression of the time series model becomes more complicated.

It may be natural to expect to have a nonlinear ARMA representation for the nonlinear stochastic dynamical system

\[ \dot{z} = f(z|\theta) + w(t) \]

so that we can estimate the model parameters through the nonlinear ARMA time series model. However this is not possible(see Ozaki(1986)), because the introduction of the ARMA representation is based on the assumption of the linearity of the system which guarantees the Cayley-Hamilton theorem, which says

\[ A^k - \phi_1 A^{k-1} - \phi_2 A^{k-2} - \ldots - \phi_k I = 0 \]

for the characteristic coefficients \( \phi_1, \phi_2, \ldots, \phi_k \) of the matrix \( A \). Here \( I \) is the unit matrix. If \( A \) is not constant and depends on \( z(t) \), the Cayley-Hamilton theorem does not hold.
On the other hand for the linear discrete time state space model (1.3), a parameter estimation method directly applied to the model without going through its ARMA time series model representation is also available. The direct estimation method is based on the Schweppe(1965)’s maximum likelihood method for discrete time linear state space models, where the Gaussianity of the innovations obtained by the Kalman filter (Kalman(1960), Kailath(1968)) is assumed. It will be natural to consider the extension of this direct approach from linear to nonlinear cases. To do that we need a nonlinear filtering method for the discrete time observation data which come from a continuous time dynamical system model.

The nonlinear filtering problem, although theoretically solved by Kushner(1962), has been a well-known difficult problem. This is because the exact solution of the nonlinear filtering problem leads to the infinite dimensional systems which are impossible to implement in practice. It has been said(Marcus, 1984), however, that "non-existence of finite dimensional filters does not address the issue of non exact but high-performance suboptimal filters, which is much more important from a practical point of view." Thus the problem of obtaining a good approximation to the exact solutions of nonlinear filtering problems has been studied by various approaches(see Hazewinkel and Willems(1981)). However as is mentioned in Frost and Kailath(1971) or Liang(1983), many of these results of nonlinear filtering provide a "representation" or "philosophy of approach" rather than a "solution" or a "procedure" leading to the derivation of practical estimators.

Liang(1983) and Schwartz and Stear(1968) discussed the difficulties of conventional nonlinear filters for systems with nonnegligible nonlinearities. A common difficulty seen in these empirical studies is the computational instability of nonlinear filtering. Even though the nonlinear filtering is properly defined and is supposed to have a finite valued solution, we often see a numerical explosion in mechanizing the prediction or filtering of the state estimate in discrete time steps. To overcome the computational instability problem we try to get an approximate nonlinear filter, not by approximate continuous time filter equations, but by discrete time filtering equations for an approximate discrete time model for the original continuous time dynamical system. The efficiency of this approach depends totally on how good the locally linearized discrete time model approximation is.

Before introducing our discretization scheme, we first explain, in section 2.1, why the conventional discretization schemes are unsatisfactory. In section 2.2 we show how the local linearization method preserves important characteristics such as stationarity of the original continuous time dynamical system model in the discrete time model. In section 2.3 the filtering scheme for the discretized model is presented. In section 2.4 the effectiveness of the filtering scheme was checked by applying the filter to the data generated from the van der Pol equation. In section 3.1 we give a maximum likelihood method for the discretized model using the local linearization filter. In section 3.2 we will show several applications of the present method for the identification of nonlinear oscillation models with numerical results. Asymptotic properties of the maximum likelihood estimates and the least squares estimates are briefly discussed with numerical results in section 3.3. Implication of the present method and related topics are discussed and conclusions are given in section 4.

2. The Local Linearization Filter.

For the purpose mentioned in the previous section, we try to introduce a nonlinear filter based on a "good" discrete time approximation to the continuous time stochastic dynamical system model for a nonlinear and non-Gaussian multivariate process \( z(t) \). Seeing first why conventional discretization schemes are not satisfactory helps for us to understand our new discretization scheme.
2.1. WHY CONVENTIONAL DISCRETIZATION METHODS ARE NOT SATISFACTORY.

For simplicity of explanation we first confine ourselves to the time discretization of the scalar stochastic dynamical system model,

\[ z = f(z | \theta) + w(t), \]  

(2.1)

where \( w(t) \) is a Gaussian white noise and \( f(z|\theta) \) is a nonlinear function of \( z(t) \). The simplest and most common discrete time model for the scalar stochastic dynamical system (2.1) is

\[ z_{t+\Delta t} = z_t + \Delta t f(z_t) + \sqrt{\Delta t} w_{t+\Delta t} \]  

(2.2)

The discretization method is a stochastic analog of Euler’s method for the discretization of deterministic dynamical system,

\[ \dot{z} = f(z | \theta) \]

The time discretization method (2.2) is consistent, i.e.

\[ \frac{z_{t+\Delta t} - z_t}{\Delta t} \to f(z_t | \theta) + w(t) \quad \text{for} \quad \Delta t \to 0. \]

Actually the Itô stochastic differential equation (and the Stratonovich stochastic differential equation as well, since the variance of \( w(t) \) is independent of \( z(t) \)) is defined as a limit of this discrete time model (see Mortensen (1979)). Therefore the distribution of \( z_t \) defined by the approximate discrete model is known to converge to the distribution of \( z(t) \) in probability. The only disadvantage of this model is that the discrete time process, i.e. the Markov chain in continuous state space defined by (2.2) is not stationary for most nonlinear function \( f(z|\theta) \) if the time interval \( \Delta t \) is fixed, no matter how small it is (Jones (1978)). We may have a non-explosive sample path if we simulate the model with a small \( \Delta t \) for a certain finite time interval. However the probability of computational explosion in a finite time is always 1, i.e. if we carry on the simulation it always ends with computational explosion after a finite number of steps. For example, for the process defined by (2.1) with \[ f(z | \theta) = -z^3 \]

the model (2.2) almost certainly explodes if \( z_t \) starts from an initial value greater than \[ \sqrt{2/\Delta t} \], even though the original continuous time process is stationary with the stationary distribution,

\[ p(x) = C \exp\left(-\frac{x^4}{2\sigma_w^2}\right) \]

and never explodes no matter how large its initial value. Here \( C \) is a normalizing constant and \( \sigma_w^2 \) is the variance of \( w(t) \).

A computationally stable time discretization scheme of deterministic dynamical system,

\[ \dot{z} = f(z | \theta) \]

(2.3)

has been used in the engineering field (p264, Sorenson (1966)). The discrete model derived from the scheme applied to the deterministic dynamical system (2.3) is

\[ z_{t+\Delta t} = e^{J_t \Delta t} z_t \]

(2.4)

where \( J_t \) is the Jacobian of \( f(z|\theta) \) at time point \( t \). The model (2.4) is known to be computationally stable whenever the original nonlinear deterministic dynamical system is stable. When \( f(z|\theta) \) is linear the trajectory of (2.4) coincides with the exact trajectory of the
original system (2.3) on the discrete time points $t=\Delta t, 2\Delta t, 3\Delta t, \ldots$. If we replace the deterministic part of the model (2.2) by (2.4) we have the following discrete time stochastic dynamical system model,

$$z_{t+\Delta t} = e^{J\Delta t}z_t + \sqrt{\Delta t} \ w_{t+\Delta t}$$  \hspace{1cm} (2.5)

The discrete time stochastic process $z_t$ (2.5), which is a Markov chain with a continuous state space, is known to be computationally stable and ergodic. Unfortunately the model is not consistent unless $f(z|\theta)$ is linear. If we let $\Delta t \to 0$, the process defined by the model (2.5) converges to

$$\dot{z}(s) = Jz(s) + w(s)$$  \hspace{1cm} (2.6)

Incidentally we note that if we assume model (2.6) on the interval $[t,t+\Delta t)$, the autocovariance function $\gamma(s)$ of the process $z(s)$ of (2.6) is

$$\gamma_t(s) = \left(\frac{\sigma_w^2}{2J_t}\right)e^{Jts}$$

Since

$$\frac{\gamma_t(\Delta t)}{\gamma_t(0)} = e^{Jt\Delta t}$$

it is reasonable to have $e^{Jt\Delta t}$ as a coefficient in the discretized model (2.5). This idea is used later when we introduce our local linearization scheme, which is consistent and computationally stable.

Some more sophisticated time discretization methods for the deterministic dynamical system (2.3) are known in numerical analysis, such as the Heun and Runge-Kutta methods. It may be natural to substitute the deterministic part of (2.2) by the deterministic discrete time models obtained by these more sophisticated methods. We call these discrete time stochastic dynamical system models the Heun and Runge-Kutta scheme models respectively. Actually these models are used for simulation of stochastic differential equations(Arnold(1974), McShane(1974)). Since the difference between these models and model (2.1) is of order $(\Delta t)^r$ with $r>1$, these models preserve the consistency property of the Euler scheme model (2.2).

Unfortunately, however, the disadvantage of model (2.2), i.e. non-stationarity, is not remedied by these more sophisticated discrete time models. For example the Runge-Kutta scheme, applied to

$$\dot{z} = -z^3 + w(t),$$

gives us a stochastic discrete time dynamic model,

$$z_{t+\Delta t} = p_{81}(z_t) + \sqrt{\Delta t} \ w_{t+\Delta t}$$  \hspace{1cm} (2.7)

where $p_{81}(z_t)$ is a 81-st order polynomial of $z_t$. Therefore the Markov chain defined by (2.7) is, like the one by (2.2) introduced by the Euler scheme, computationally explosive and is not stationary. We can often avoid the computational explosion within a given finite time steps, in the simulations, by taking $\Delta t$ very small. However, in the inferential use of the discretized model, we cannot choose the scale of the white noise variance and the time interval $\Delta t$, unlike the simulation studies, since they are already fixed when the data were sampled.
In numerical analysis it is known that the error of the Euler method applied to the
deterministic dynamical system (2.3) is of order o(Δt), of the Heun method order o(Δt)^2,
and of the Runge-Kutta method order o(Δt)^4. In the stochastic situation the goodness of the
approximations may be evaluated by the mean square one step error,

\[ E_\mathbb{Q}_t \left[ (z(t+Δt) - z_{t+Δt})^2 \right] \]

where \( E_\mathbb{Q}_t \) denotes the conditional expectation with respect to \( z_t \). The Euler scheme model
(2.2) is known to have the expected mean square error of order \( o(Δt)^2 \)(Milstein(1974); note
that our model has, unlike those of Milstein(1974) and Rumelin(1982), a special form,
where the noise variance is constant). A higher order expected mean square error are
expected from the Heun scheme and Runge-Kutta scheme models. However it is shown that
the order of the expected mean square error of these models is also \( o(Δt)^2 \). Moreover it has
been proved that the maximum possible speed of convergence attained by a discrete time
approximate model is \( o(Δt)^2 \)(Milstein(1974), Rumelin(1982)). This result is, in a sense,
natural because even though we approximate the deterministic part of the model (2.1) more
accurately by the Runge-Kutta method etc., the approximation of the white noise part in
these models stays as poor as the Euler scheme model, dominating the performance on the
whole, and is impossible to improve in general. Thus the other discrete time models with
more sophisticated time discretization schemes are no better than the Euler method (2.2).

From the above discussion we now know that we should not expect to have a new time
discretization scheme with a higher order convergence speed of the approximation. What we
need is a new discretization method which, like the Euler scheme method, is consistent and
at the same time gives a computationally stable and stationary Markov chain like (2.5) for a
fixed \( Δt \) whenever the original continuous time Markov diffusion process is stationary.

2.2. THE LOCAL LINEARIZATION SCHEME.

A key for the introduction of a new time discretization scheme lies in the scheme (2.5)
which is the only one giving a computationally stable scheme among the conventional
methods. This scheme is also the only one whose deterministic part gives a trajectory
which exactly coincides with the true trajectory of the continuous deterministic system (2.3)
at the discrete time points when \( f(z(t)) \) is linear. The scheme (2.5) can be introduced by
approximating the original process by the linear process (2.6) on each short interval \([t,t+Δt)\),
where we assume that the coefficient of the linear function of the dynamical system on the
interval is given by the Jacobian \( J_z \) of \( f(z(t)) \) as \( J_z \) at point \( t \). A disadvantage of this
scheme is that it is not consistent. This is because \( J_z \) is constant on the interval \([t,t+Δt)\) and
the function \( J_z \) does not converge to \( f(z(t)) \) for \( Δt \to 0 \). This consideration suggests that
we may be able to get a computationally stable and consistent scheme if we use some
different function \( L_z \) to approximate the original continuous model on each short time
interval \([t,t+Δt)\) with, instead of (2.6), a linear stochastic dynamical system,

\[ \dot{z}(s) = L_t z(s) + \mathbb{W}(s), \]

where \( L_t \) is some function of \( z_t \) and

\[ L_t z_t \to f(z(t)) \quad \text{for} \quad Δt \to 0. \]

A scheme which satisfies this requirement is obtained by a simple assumption, i.e. "The
Jacobian of the linear dynamical system,

\[ \dot{z}(s) = L_t z(s) \]

is equivalent to the Jacobian,
of the original dynamical system on each short time interval \([t,t+\Delta t]\). We call it the Local Linearization (L.L.) assumption. From this assumption we have

\[
\dot{z}(s) = J_t z(s)
\]
on \([t,t+\Delta t]\). If we integrate this on the interval \([t, t+\tau]\), where \(0 \leq \tau < \Delta t\), we have

\[
\dot{z}(t+\tau) = e^{J_t \tau} z(t) = e^{J_t \tau} f(z(t)).
\]

By integrating this again with respect to \(\tau\) on the interval \([0, \Delta t]\) we have,

\[
z(t + \Delta t) = z(t) + J_t^{-1} \left( e^{J_t \Delta t} - 1 \right) f(z(t)) \quad (2.10)
\]

Since \(J_t\) is given as a function of \(z(t)\), \(z(t+\Delta t)\) is explicitly given as a function of \(z(t)\) at time point \(t\). Thus the trajectory of the system (2.9) takes the value (2.10) at the discrete time point, \(t+\Delta t\). On the other hand the solution \(z(t+\Delta t)\) of the linear system (2.9) is given explicitly as a function of \( L_t \) and \( z(t) \) by integrating \( \dot{z}(s) \) of (2.9) from \( t \) to \( t+\Delta t \) to give,

\[
z(t + \Delta t) = e^{L_t \Delta t} z(t) \quad (2.11)
\]

From (2.10) and (2.11) we can write down \( L_t \) explicitly as a function of \( z(t) \),

\[
L_t = \frac{1}{\Delta t} \log \left( 1 + J_t^{-1} \left( e^{J_t \Delta t} - 1 \right) F_t \right) \quad (2.12)
\]

where

\[
F_t = \frac{f(z(t))}{z(t)}.
\]

\( L_t \) of (2.12) is defined on the region of \( z(t) \) where \( J_t z_t \neq 0 \). When \( J_t z_t = 0 \), it holds, for sufficiently small \( \Delta t \), that

\[
1 + J_t^{-1} \left( e^{J_t \Delta t} - 1 \right) F_t > 0
\]

When \( z(t) \) is on the region where \( J_t z_t = 0 \) (which is of measure zero), the \( z(t+\Delta t) \) is defined separately in a manner consistent with the rest of the region (see Ozaki(1992)).

Next we try to use \( L_t \) of (2.12) to introduce a stochastic version of the discrete time approximation to the continuous time system. One simple way is, as in those conventional discretization schemes for stochastic dynamical systems, to add \( \sqrt{\Delta t} w_{t+\Delta t} \) to the deterministic part, thus

\[
z_{t+\Delta t} = e^{L_t \Delta t} z_t + \sqrt{\Delta t} w_{t+\Delta t}. \quad (2.13)
\]

However if we stick to the the L.L. assumption (2.9) and extend it to stochastic dynamical systems, we have a more consistent model than (2.13). In the stochastic situation it will be natural to start from the assumption, "The stochastic dynamical system is locally linear and so it is Gaussian on each short time interval \([t,t+\Delta t]\), i.e. we employ the model (2.8) on the interval to approximate the original process, where \( L_t \) is given by (2.12) ". Then \( \dot{z}(s) \) of (2.8) can be integrated from \( t \) to \( t+\Delta t \) on the interval giving

\[
z_{t+\Delta t} = e^{L_t \Delta t} z_t + a_{t+\Delta t}
\]

where
\[ a_{t+\Delta t} = \int_{t}^{t+\Delta t} e^{L_t (t+\Delta t - s)} w(s) ds \]

The variance of \( a_{t+\Delta t} \) is
\[
\sigma_a^2 = \int_{t}^{t+\Delta t} e^{2L_t (t+\Delta t - s)} \sigma_w^2 \, ds \\
= \frac{(e^{2L_t \Delta t} - 1)}{2L_t} \sigma_w^2.
\]

where \( \sigma_w^2 \) is the variance of the continuous time white noise \( w(t) \). Therefore it will be natural and more consistent if we use, instead of the model (2.13), the following discrete time model,
\[
z_{t+\Delta t} = e^{L_t \Delta t} z_t + \sqrt{\frac{(e^{2L_t \Delta t} - 1)}{2L_t}} w_{t+\Delta t}
\]

(2.14)

where \( w_{t+\Delta t} \) is a discrete time Gaussian white noise of variance \( \sigma_w^2 \). Since \( L_t \) of (2.12) satisfies
\[ L_t z(t) \to f(z(t)) \text{ for } \Delta t \to 0, \]
the model (2.14) is consistent. The deterministic part of the model (2.14) inherits the zero points and Jacobian of the original continuous time deterministic dynamical system (2.3). Therefore the stability property of the original continuous time stochastic dynamical system is also preserved in scheme (2.14). The ergodicity of the discrete time model is proved by direct application of Tweedie(1975)'s theorem when a proper \( f(z | \Theta) \) of the original continuous time system is given(Ozaki(1985)). Thus the Markov chain process defined by (2.14) brings us a stationary Markov chain if the original continuous time diffusion process is stationary.

The local linearization can be applied to the multi-dimensional stochastic dynamical system,
\[
\dot{z} = f(z | \Theta) + w(t)
\]

(2.15)

without any essential change. Since the integration of the linear multi-dimensional differential equation,
\[
\dot{z} = L z(t)
\]
is
\[ z(t) = \text{Exp}(L_t) z(0), \]
the discrete time model for the multi-dimensional stochastic dynamical system (2.15) is given by
\[
z_{t+\Delta t} = A(z_t) z_t + B(z_t) \ w_{t+\Delta t}
\]

(2.16)

\[ A(z_t) = \text{Exp}(L(z_t) \Delta t) \]

\[ L(z_t) = \frac{1}{\Delta t} \text{Log} \{ 1 + J_t^{-1} (e^{J_t \Delta t} - 1)F_t \} \]

\[ J_t = \left( \frac{\partial f(z)}{\partial z} \right) z = z_t \]

\[ F_t \text{ is such that } F_t z_t = f(z_t). \]

The matrix functions \( \text{Exp}(.) \) and \( \text{Log}(.) \) are defined by
\[
\text{Exp}(L) = \sum_{k=0}^{\infty} \frac{1}{k!} L^k
\]

and

\[
\text{Log}(L) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (L - 1)^k
\]

respectively. The elements of the matrix B_t are explicitly given as a function of eigenvalues of the matrix \(L_t\) (Ozaki, 1986, 1989).

2.3. THE LOCAL LINEARIZATION FILTER.

It is well known that nonlinear filtering is a difficult problem even in the scalar case. Its main difficulty is the computational instability which comes from the nonlinearity of the state dynamics. Our idea of obtaining a computationally stable filtering scheme for,

\[
\begin{align*}
\dot{z} &= f(z_t, \theta) + w(t) \\
\dot{x}_t &= Cz_t + y_t \\
C &= \begin{pmatrix} 0 & \ldots & 1 & \ldots & 0 \\
0 & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 0 & \ldots & 1 \end{pmatrix}
\end{align*}
\]

is to introduce a stable approximate discrete time L.L. model from (2.17), instead of introducing an approximate continuous time filtering scheme, and to give a discrete time filtering scheme for the L.L. model. In (2.17), \(w(t)\) is a continuous time \(k\)-dimensional Gaussian white noise with variance-covariance matrix \(\Sigma_w\) and \(y_t\) is a discrete time \(r\)-dimensional Gaussian white noise with variance-covariance matrix \(\Sigma_y\) and \(z_t\) and \(y_t\) are independent. Thus from (2.17), using the L.L. method in the previous section, we have,

\[
\begin{align*}
\dot{z}_{t+\Delta t} &= A(z_t)z_t + B(z_t)w_{t+\Delta t} \\
\dot{x}_t &= Cz_t + y_t
\end{align*}
\]

where \(A(z_t)\) and \(B(z_t)\) are given in (2.16). \(w_{t+\Delta t}\) is a \(k\)-dimensional discrete time Gaussian white noise with variance-covariance matrix \(\Sigma_w\). Also between the i-th component \(w^{(i)}\) of \(w_{t+\Delta t}\) and j-th component \(v^{(j)}\) of \(y_t\) it holds that \(E[w^{(i)v^{(j)}}] = 0\) for \(1 \leq i \leq k\) and \(1 \leq j \leq r\). Based on the minimum variance principle (Sorenson (1966)), which is to minimize

\[
E[(\dot{z}(t) - z_{t+\Delta t})'(\dot{z}(t) - z_{t+\Delta t})]
\]

the filtering equation for (2.18) is obtained as follows,

\[
\begin{align*}
\dot{z}_{t+\Delta t} &= A(z_t)z_t + K_t y_t \\
\dot{x}_t &= Cz_t + K_t y_t \\
K_t &= P_t C'(C P_t C' + \Sigma_y)^{-1} \\
P_t &= E[(\dot{z}(t) - z_{t+\Delta t})(\dot{z}(t) - z_{t+\Delta t})'] \\
V_t &= P_t - K_t C P_t \\
&= P_t - P_t C' (C P_t C' + \Sigma_y)^{-1} C P_t
\end{align*}
\]
where \( v_t \) is the innovation of the filtering model, \( K_t \) is the filter gain and \( P_t \) is the variance-covariance matrix of the one step ahead prediction error of the state \( z_t \). The evolution of \( P_t \) is obtained from

\[
P_{t+\Delta t} = E[(z_{t+\Delta t} - z_t)(z_{t+\Delta t} - z_t)']
= E[(A(z_t)z_t - z_t) + B(z_t)w_{t+\Delta t})(A(z_t)z_t - z_t) + B(z_t)w_{t+\Delta t})']
\]

Since we are assuming \( A(z_t) \) and \( B(z_t) \) are constant on \([t, t+\Delta t)\), we have

\[
P_{t+\Delta t} = A(z_t)P_{t+\Delta t}A(z_t)' + B(z_t)\Sigma_w B(z_t)'
\]

(2.20)

For small \( \Delta t \), the function \( B(\cdot) \) is almost constant, whereas \( A(\cdot) \) is not. Since we assume that the system (2.18) is linear on \([t, t+\Delta t)\) and the system transition is characterized by \( A(z_t) \) as in (2.19), it will be reasonable to replace \( A(z_t) \) and \( B(z_t) \) in (2.19) and (4.4) by \( A(z_{it}) \) and \( B(z_{it}) \) respectively. Then we have

\[
z_{t+\Delta t | t} = A(z_t)z_t | t + w_{t+\Delta t} | t
\]

(2.21)

and

\[
P_{t+\Delta t} = A(z_t)P_{t+\Delta t}A(z_t)' + B(z_t)\Sigma_w B(z_t)'
\]

(2.22)

The above filtering equations were first introduced in Ozaki(1990a), where its effectiveness was confirmed by comparing the results by the conventional Extended Kalman filtering method using simulation data.

2.4. APPLICATION TO THE VAN DER POL OSCILLATORS.

We applied the local linearization filter to the data (see Fig.1) generated from

\[
\ddot{x}(t) - 3(1 - x(t)^2)\dot{x}(t) + x(t) = 0
\]

(2.23)

which is called the van der Pol equation. Liang(1983) says about the example, "This example was selected not only because of its third order nonlinearities but also because Kushner(1967) had previously shown that for this state-estimation problem, even with a linear measurement model the linearized filter was extremely unstable and was completely useless within a fraction of a time unit, and even the implementation of a Gaussian second order filter proved to be unstable." In Liang(1983)'s simulation studies the observation noise \( \sigma^2 \) is chosen to be \( \sigma^2 = 0.1 \).

From (2.23) we have the local linearization filter (2.19) whose \( A(\cdot) \) and \( B(\cdot) \) matrices are given by the following \( f(z_t), J_t \) and \( F_t \),

\[
f(z_t) = [3(1 - x_t^2)\dot{x}_t - x_t, \dot{x}_t]' \]

\[
J_t = \begin{pmatrix}
3(1 - x_t^2) & -6x_t\dot{x}_t - 1 \\
1 & 0
\end{pmatrix}
\]

\[
F_t = \begin{pmatrix}
3(1 - x_t^2) & -1 \\
1 & 0
\end{pmatrix}
\]
Figure 1: van der Pol oscillation data generated from (2.23) with $\sigma^2=0.1$.

Figure 2.a: Estimated states and true states.
Fig. 1 shows the simulation data generated from the model (2.23). In the paper the initial estimate of the state and the variance-covariance matrix of the initial estimate were taken as follows,

\[ z_{0.10} = (2, 0) \]
\[ V_0 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \]

Liang applied, for the data, two filtering methods; one is the stochastic linearization filter method and the other is the minimum variance filtering method. The conclusion of his analysis is that the minimum variance filtering method works better than the stochastic linearization filtering method. However, if we look at the errors of the minimum variance filter in Fig. 13 of Liang (1983), still the result is obviously far from satisfactory. We applied our locally linearized filter for the same data with the same initial condition. Obtained results, estimated state and true state are plotted in Fig. 2a. Errors of the state estimates and error bounds, are in Fig. 2b, where we took \( \Delta t = 0.001 \) (Liang (1983) does not mention the step size of his numerical integration). The figure shows a drastic improvement of the filtering performance compared with the minimum variance method of Liang (1983). Obviously such large errors by the conventional filtering schemes in Liang (1983) will cause a great difficulty when we use the filtering scheme for the estimation of parameters of nonlinear systems.

With (2.16) we can transform the time series data $z_1, z_2, ..., z_N$ into Gaussian white noise $w_1, w_2, ..., w_N$, and we can estimate the model parameters by maximizing the likelihood in terms of the parameter $\Theta$. This is the way how parameters are estimated in ordinary linear or nonlinear time series models. However when we observe an $r$-dimensional ($r < k$) vector data $x_i$, instead of the $k$-dimensional state vector $z_i$, and when observation errors are present, we cannot obtain the residuals directly from the model structure as for (2.16). We need a nonlinear filtering method to whiten the $r$-dimensional data, $x_1, x_2, ..., x_N$, into $r$-dimensional innovations, $v_1, v_2, ..., v_N$.

3.1. THE MAXIMUM LIKELIHOOD METHOD

Before we go into the maximum likelihood method for the nonlinear state space models, let us see how the linear state space model parameters are estimated by the maximum likelihood method. The parameters of a discrete time linear state space model,

\[
\begin{align*}
\dot{z} &= L(\Theta) z + w(t) \\
x_i &= C z(t) + y_i,
\end{align*}
\]

\[C = \begin{pmatrix}
0 & 0 & 1 & \ldots & 0 \\
0 & \ddots & \ddots & \ddots & 0 \\
\end{pmatrix}
\]

(3.1)

can be estimated by maximizing the likelihood,

\[p(x_1, x_2, \ldots, x_N | \Theta)\]

of the model for the given data $x_1, x_2, \ldots, x_N$. If we use the Kalman filter, the data $x_1, x_2, \ldots, x_N$ are transformed to innovations $v_1, v_2, \ldots, v_N$ as

\[v_i = x_i - \hat{x}_{i|i-1}\]

(3.2)

where $\hat{x}_{i|i-1}$ is the one-step ahead prediction value of $x_i$ at time point $i-1$ and is a function of $z_{i-1}$, $v_{i-1}$ and $\Theta$. If the initial values $z_0$ and $V_0$ are known, then the Jacobian of the transformation from $x_1, x_2, \ldots, x_N$ to the innovations $v_1, v_2, \ldots, v_N$ is 1. Usually they are unknown and may be specified by a probability density function $p(z_0, V_0 | \Theta)$. Then the likelihood is written down in terms of the innovations, obtained by applying the Kalman filter, as

\[p(x_1, x_2, \ldots, x_N | \Theta) = \int \int p(x_1, x_2, \ldots, x_N | z_0, V_0, \Theta) p(z_0, V_0 | \Theta) dz_0 dV_0\]

\[= \int \int p(v_1, v_2, \ldots, v_N | z_0, V_0, \Theta) p(z_0, V_0 | \Theta) dz_0 dV_0\]

There are many ways of specifying $p(z_0, V_0 | \Theta)$. One simple and practical way of specification is to use the delta function. If we assume a delta function for $z_0$ and $V_0$, the likelihood becomes

\[p(x_1, x_2, \ldots, x_N | z_0, V_0, \Theta) = p(v_1, v_2, \ldots, v_N | z_0, V_0, \Theta)\]
where we have two more parameter sets $z_0$ and $V_0$. Since the innovations are Gaussian white noise, we have

$$p(y_1, y_2, \ldots, y_N | z_0, V_0, \Theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi |\Sigma_i|}} \exp \left( -\frac{1}{2} y_i^{\top} \Sigma_i^{-1} y_i \right)$$

Then we have

$$(-2)\log p(x_1, x_2, \ldots, x_N | z_0, V_0, \Theta) = \sum_{i=1}^{N} (\log |\Sigma_i| + y_i^{\top} \Sigma_i^{-1} y_i) + N \log 2\pi$$

(3.4)

Then the maximum likelihood estimates of the parameters $z_0$, $V_0$ and $\Theta$ are obtained by minimizing $(-2)\log$-likelihood (3.4). The maximum likelihood method based on this innovation likelihood has been used for the estimation of parameters of linear state space models (Schweppes (1965), Harvey (1984) and Kitagawa (1981)).

In the nonlinear case,

$$\tilde{z} = f(z | \Theta) + w(t) \quad (3.5)$$
$$x_t = C \tilde{z}(t) + y_t$$

we expect to use the L.L. filter to transform the data $x_1$, $x_2$, ..., $x_N$ into Gaussian white noise innovations $y_1$, $y_2$, ..., $y_N$ as we did in the linear case. When $f(z | \Theta)$ is nonlinear the state $\tilde{z}(t)$ is not Gaussian. However if the sampling interval $\Delta t$ is sufficiently small, we can assume that the data $x_1$, $x_2$, ..., $x_N$ are generated from the discrete time state space model,

$$z_{t+\Delta t} = A(z_t) z_t + B(z_t) w_{t+\Delta t}$$
$$x_t = C z(t) + y_t \quad (3.6)$$

whose state space process $z_t$ is driven by $w_t$. Then we can get discrete time Gaussian white noise innovations, from observations $x_t$ by applying the L.L. filter, since $w_{t+\Delta t}$ is a discrete time Gaussian white noise. It is also proved by Frost and Kailath (1971) that the continuous time innovation process for a non-Gaussian diffusion process is a Gaussian white noise. Since the continuous time innovations are defined as a limit of discrete time innovations as

$$y(t) = \lim_{\Delta t \to 0} y_t$$
$$\quad = \lim_{\Delta t \to 0} (x_t - C z_{t-\Delta t})$$

their result implies that the innovation $y_t$ obtained by the L.L. filter is very close to a Gaussian white noise. This consideration justifies and guarantees the use of the following representation for the approximation of the $(2)\log$-likelihood,

$$(-2)\log p(x_1, x_2, \ldots, x_N | z_0, V_0, \Theta) = \sum_{i=1}^{N} (\log |\Sigma_i| + y_i^{\top} \Sigma_i^{-1} y_i) + N \log 2\pi$$

(3.7)

where $\Sigma_i$ is the variance-covariance matrix of the innovation $v_i$ and is given by the L.L. filter (4.3) as

$$\Sigma_i = CP_i C^{\top} + \Sigma_y$$

(3.8)

Thus the maximum likelihood estimates of the parameters $z_0$, $V_0$ and $\Theta$ of (3.1) are obtained by minimizing (3.7).
3.2. PARAMETER ESTIMATION IN PRACTICE.

It is a common view of applied time series analysts that most nonlinear models carry a lot of theoretical material without much useful applications. Unfortunately this view cannot be denied when we look through the articles and books on system identifications where almost all numerical results treated are about linear models. This is mainly because, for most typical nonlinear models such as van der Pol oscillators, the numerical difficulties of nonlinear models' computational instabilities have been unavoidable. As we saw in the previous sections, the L.L. filter provide us with a computationally stable scheme. With this L.L. filter combined with the maximum likelihood method, we try to see if the identification of nonlinear random vibration systems is possible and usable in practice. For this purpose we use the following two examples of typical nonlinear random vibrations.

Example 1. \[ \ddot{x}(t) - 3 \{1 - x(t)^2 \} \dot{x}(t) + x(t) = n(t) \] (3.9)

Example 2. \[ \ddot{x}(t) + 0.25 \dot{x}(t) + \{14.8 + 4 x(t)^2 \} x(t) = n(t) \] (3.10)

Figure 3.: Simulated data of model (3.9) with \( \sigma_n^2 = 1 \) and \( \sigma \epsilon^2 = 0.01 \). (N=1000, \( \Delta t = 0.1 \))
Fig. 3 shows simulation data for the example I, where $x(t)$ was sampled with $\Delta t=0.1$ and the observation error variance $\sigma_n^2=0.01$. The number of data points is 1000. In the Example 2, the nonlinearity is in the third term of the equation which describes the restoring force of the vibrations. When $n(t)$ is replaced by a sinusoidal wave the equation is called a Duffing equation. This type of nonlinear vibration is sometimes called Duffing type random vibration. Fig. 4 shows the simulation data of the model of Example 2 with $\Delta t=0.1$, observation noise variance $\sigma_v^2=0.001$ and $N=1000$. In Example 2, because of the nonlinear term, the restoring force increases faster than for the linear case when the amplitude increases, which leads to the frequency shift from lower frequency to higher frequency. This frequency shift phenomenon is seen in Fig. 4, where $x_i$ around $t=900$ shows a higher frequency character than $x_i$ around $t=700$.

Both examples are special cases of the following general nonlinear random vibration model,

$$\ddot{x}(t) + a(x) \dot{x}(t) + b(x) = n(t),$$  \hspace{1cm} (3.11)

where the damping coefficient $a(x)$ is $x$-dependent and the restoring function $b(x)$ is nonlinear. From (3.11) we have the following stochastic dynamical system representation for the general nonlinear random vibration,

$$\dot{z} = f(z) + \mathbf{w}(t)$$  \hspace{1cm} (3.12)

$$z(t) = [\dot{x}(t), x(t)]'$$
\[ w(t) = (n(t), 0)' \]
\[ f(z) = (a(x) \dot{x} + b(x), \dot{x})' \]

Its locally linearized model is
\[ z_{t+\Delta t} = A(z_t) z_t + w_{t+\Delta t}, \]

where
\[ A(z_t) = \text{Exp}(L_t \Delta t) \]
\[ w_{t+\Delta t} = \int_t^{t+\Delta t} \exp\{L_t(t + \Delta t - u)\} w(u) du \]
\[ L_t = \frac{1}{\Delta t} \log\{1 + J_t^{-1} (\exp^{J_t \Delta t} - I) F_t\} \]
\[ J_t = \begin{pmatrix} a(x) & a'(x) \dot{x} + b'(x) \\ 1 & 0 \end{pmatrix}, \]
\[ F_t = \begin{pmatrix} a(x) & b(x)/\dot{x} \\ 1 & 0 \end{pmatrix} \]

We note that the variance-covariance matrix \( S_t \) of \( n_{t+\Delta t} \) has full rank 2, even though the variance-covariance matrix of the continuous time white noise \( w(t) = (n(t), 0)' \) has rank one. This is because of the integration of \( w(t) \) over \([t, t+\Delta t]\). \( S_t \) is decomposed as follows,
\[ S_t = \sigma^2 U_t \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} U_t' \]

using the unitary matrix \( U_t \). The elements of \( U_t \) and eigenvalues \( \lambda_1 \) and \( \lambda_2 \) are given as functions of eigenvalues of \( L_t \) and are explicitly given as functions of \( \dot{x}_t \) and \( x_t \) (see Ozaki(1986)). Therefore we have, from (3.12), the following discrete time state space representation,
\[ z_{t+\Delta t} = A(z_t) z_t + B(z_t) w_{t+\Delta t} \]
\[ x_t = C z_t + v_t \]
\[ B(z_t) = U_t \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} \]
\[ C = (0, 1) \]
\[ z_t = (\dot{x}_t, x_t) \]

where \( w_{t+\Delta t} \) is a discrete time Gaussian white noise of variance-covariance matrix \( \sigma^2 I \) and \( A(z_t) \) is given in (3.13). Using the L.L. filter equations (2.19) and (2.22) for (3.14) we can write down the likelihood of the model for the data \( \zeta_1, \zeta_2, ..., \zeta_N \).

We applied the maximum likelihood method based on the L.L. filter and obtained the results in Table 1. In the computation of the maximum likelihood estimation we fix the initial values \( z_0 \) and \( v_0 \) as \( z_0 = (0.1, 0.01)' \) and
\[ v_0 = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.01 \end{pmatrix} \]

The estimated parameters are quite satisfactory and close to the true values in both cases.
Table 1 Maximum likelihood estimates of the model (3.9) and the model (3.10).

<table>
<thead>
<tr>
<th></th>
<th>Model (3.9) true</th>
<th>estimated</th>
<th></th>
<th>Model (3.10) true</th>
<th>estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>-3</td>
<td>-2.9178</td>
<td>$a$</td>
<td>0.25</td>
<td>0.1709</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0</td>
<td>-0.1599x10^{-1}</td>
<td>$b_1$</td>
<td>14.8</td>
<td>15.5977</td>
</tr>
<tr>
<td>$a_3$</td>
<td>3</td>
<td>2.9436</td>
<td>$b_2$</td>
<td>0</td>
<td>0.4649</td>
</tr>
<tr>
<td>$b$</td>
<td>14.8</td>
<td>14.7528</td>
<td>$b_3$</td>
<td>4</td>
<td>3.8065</td>
</tr>
<tr>
<td>$\sigma_n^2$</td>
<td>1</td>
<td>1.7432</td>
<td>$\sigma_n^2$</td>
<td>4</td>
<td>3.5093</td>
</tr>
<tr>
<td>$\sigma_v^2$</td>
<td>0.01</td>
<td>0.9132x10^{-2}</td>
<td>$\sigma_v^2$</td>
<td>0.001</td>
<td>0.9990x10^{-3}</td>
</tr>
</tbody>
</table>

3.3. MAXIMUM LIKELIHOOD ESTIMATES AND LEAST SQUARES ESTIMATES.

For the estimation of parameters of state-space representation models, apart from the maximum likelihood method, there are some other methods such as the least squares estimation method, which suggests to choose the parameter minimizing the sum of squares of the prediction errors,

$$v_t(\theta) = x_t - x_{lt-1}(\theta),$$

thus

$$\hat{\theta}_N = \arg \min_{\theta} \left\{ \frac{1}{N} \sum_{t=1}^{N} \frac{1}{2} v_t(\theta) \right\}$$

Ljung(1976), extending the least squares method, introduced the prediction error estimation method which suggests to choose the parameter,

$$\hat{\theta}_N = \arg \min_{\theta} \left\{ \frac{1}{N} \sum_{t=1}^{N} l(v_t(\theta)) \right\}$$

where $l(.)$ is a scaler-valued (typically positive) function. Asymptotic properties, such as consistency and normality, were proved, and conditions for the convergence have been refined(Caines(1976), Ljung(1978), Caines(1978) and Ljung and Caines(1979)) for the prediction error estimation method. Incidentally, we note that the importance of the subject of the present paper was pointed out, in relation to the prediction error estimation method, by Caines(1978), saying "As a result, we have 'reduced' the identification of (stationary) nonlinear systems to the set of difficult open problems known as nonlinear filtering, and nonlinear system identification constitutes, in the author's view, an exciting area for future research."

Since our maximum likelihood method is a special case of the prediction error estimation method($l(.) = - \log p(.)$), the maximum likelihood estimates we have seen in the previous section share the same asymptotic convergence properties as the prediction error estimation method. However what interests applied scientists more is not whether the consistency is weak or strong, but whether the estimates are computationally obtainable in practice and whether the obtained estimates are reliable for ordinary data analysis where number of data points are limited. If the estimates are obtainable we can easily check the asymptotic
convergence properties of the estimation method by repeatedly applying the method to
simulation data using a modern fast computer.

Using our nonlinear filtering method, i.e. the L.L. filtering scheme, we applied the
maximum likelihood method and the least squares method to the twenty sets of data
generated from the nonlinear model (3.10). The number of data points of the first each ten
data sets is 100 and the number of data points for the latter each ten data sets is 1000. In
both methods we used the same numerical optimization procedure (Ishiguro and
Akaike(1989)). The estimation results are in Table 2 and Table 3 respectively.

<table>
<thead>
<tr>
<th>Table 2. MLE and LSE (N=100) for (3.10).</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE a 0.25</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>MLE1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>LSE1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>
Table 3. MLE and LSE (N=1000) for (3.10).

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b₁</th>
<th>b₂</th>
<th>b₃</th>
<th>σₓ²</th>
<th>σᵧ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>0.25</td>
<td>14.8</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0.001</td>
</tr>
<tr>
<td>MLE1</td>
<td>0.17</td>
<td>15.60</td>
<td>0.46</td>
<td>3.81</td>
<td>3.51</td>
<td>0.00100</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
<td>14.49</td>
<td>-0.23</td>
<td>4.25</td>
<td>3.86</td>
<td>0.00099</td>
</tr>
<tr>
<td></td>
<td>0.33</td>
<td>14.39</td>
<td>-0.35</td>
<td>3.01</td>
<td>3.78</td>
<td>0.00110</td>
</tr>
<tr>
<td></td>
<td>0.41</td>
<td>14.70</td>
<td>0.05</td>
<td>4.29</td>
<td>3.86</td>
<td>0.00116</td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>14.13</td>
<td>-0.36</td>
<td>4.76</td>
<td>3.31</td>
<td>0.00090</td>
</tr>
<tr>
<td></td>
<td>0.23</td>
<td>14.91</td>
<td>-0.15</td>
<td>4.11</td>
<td>4.40</td>
<td>0.00105</td>
</tr>
<tr>
<td></td>
<td>0.32</td>
<td>13.89</td>
<td>-0.03</td>
<td>4.80</td>
<td>4.32</td>
<td>0.00099</td>
</tr>
<tr>
<td></td>
<td>0.29</td>
<td>14.09</td>
<td>0.29</td>
<td>4.38</td>
<td>4.00</td>
<td>0.00109</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
<td>14.53</td>
<td>-0.04</td>
<td>4.18</td>
<td>3.62</td>
<td>0.00108</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
<td>14.82</td>
<td>-0.40</td>
<td>3.74</td>
<td>4.78</td>
<td>0.00107</td>
</tr>
<tr>
<td>LSE1</td>
<td>0.18</td>
<td>15.62</td>
<td>0.48</td>
<td>3.79</td>
<td>5.42</td>
<td>0.00153</td>
</tr>
<tr>
<td></td>
<td>0.21</td>
<td>14.49</td>
<td>-0.23</td>
<td>4.24</td>
<td>5.03</td>
<td>0.00129</td>
</tr>
<tr>
<td></td>
<td>0.32</td>
<td>14.38</td>
<td>-0.38</td>
<td>4.69</td>
<td>5.7×10⁴</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>0.41</td>
<td>14.73</td>
<td>0.02</td>
<td>4.24</td>
<td>4.12</td>
<td>0.00124</td>
</tr>
<tr>
<td></td>
<td>0.37</td>
<td>14.25</td>
<td>-0.33</td>
<td>4.61</td>
<td>1.7×10⁴</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>0.23</td>
<td>14.78</td>
<td>-0.05</td>
<td>4.14</td>
<td>0.36</td>
<td>0.00009</td>
</tr>
<tr>
<td></td>
<td>0.27</td>
<td>13.78</td>
<td>-0.11</td>
<td>4.77</td>
<td>0.49</td>
<td>0.00012</td>
</tr>
<tr>
<td></td>
<td>0.29</td>
<td>14.02</td>
<td>0.33</td>
<td>4.45</td>
<td>1.7×10³</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
<td>14.55</td>
<td>-0.03</td>
<td>4.17</td>
<td>9.9×10⁴</td>
<td>29.54</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
<td>14.82</td>
<td>-0.39</td>
<td>3.74</td>
<td>3.2×10³</td>
<td>0.94</td>
</tr>
</tbody>
</table>

From the tables we can see that the maximum likelihood method works reasonably well even for fairly short data. The least squares method, on the other hand, sometimes gives very unreliable estimates even for long data. The extraordinary large values of the least squares estimates for σₓ² and σᵧ² are not caused by the employed nonlinear optimization method. The gradient of these optimized points were all reasonably small (less than 10⁻³). The original data of these abnormal estimates do not show any special different feature (see Fig.5 and Fig.6). The present numerical results imply that the maximum likelihood method is much more reliable than the least squares method, although both methods share the same asymptotic convergence properties in theory.
Figure 5. Ten sets of simulated data of model (3.10) with $\sigma_n^2=4$ and $\sigma_v^2=0.001$.
(N=100, $\Delta t=0.1$)
Figure 6. Ten sets of simulated data of model (3.10) with $\sigma_n^2=4$ and $\sigma_v^2=0.001$

(N=1000, $\Delta t=0.1$)

Many interesting models, such as deterministic nonlinear dynamical systems, stochastic nonlinear dynamical systems and diffusion processes, used in sciences, were unfortunately not estimable from observation data.

We have seen that the local linearization scheme provides us with a method to treat these continuous time models, by automatically converting them into discrete time models, on the same basis as nonlinear time series models. Further we have seen that the local linearization filter, applied to the state space representation models of these discrete time models, made it possible to estimate the parameters of these models from time series data.

Acknowledgement

The present work owes a great deal to Akaike's introduction of Boltzman entropy in statistics, which directed the interest of many statisticians away from proving asymptotic theory of unobtainable estimates to devising concrete methods of estimating useful models in natural and social sciences. On the occasion of Prof. Akaike's 65th birthday, I would like to express hearty thanks to him. Sharing an office with him when I was a junior young researcher is one of my most valuable experiences.

References


