# Non-recursive and recursive methods for parameter estimation in filtering problems

Michael A. Kouritzin, Hongwei Long, Xinjian Ma and Wei Sun

MITACS-PINTS
Department of Mathematical and Statistical Sciences
University of Alberta
Edmonton, Canada T6G 2G1
{mkouritz, long, victor, wsun}@math.ualberta.ca

## ABSTRACT

Nonlinear filtering is an important and effective tool for handling estimation of signals when observations are incomplete, distorted, and corrupted. Quite often in real world applications, the signals to be estimated contain unknown parameters which need to be determined. Herein, we develop and analyze non-recursive and recursive methods, which can deal with combined state and parameter estimation for nonlinear partially-observed stochastic systems. For the non-recursive method, we obtain the unknown parameters through solving a system of non-singular finite order linear equations. For the recursive method, we generalize the least squares method and develop a particle prediction error identification algorithm so that it can be applied to general nonlinear stochastic systems. We use the branching particle filter to do the signal state estimation and implement simulations for both methods.

**Keywords:** filtering, nonlinear stochastic system, parameter estimation, asymptotic method of moments, particle prediction error identification method, particle filter.

## 1. INTRODUCTION

Nonlinear filtering is an important signal processing tool that can be used to help identify or recognize a target through its movement. In this and many other applications, the signal contains unknown parameters that must be identified in parallel with the filtering process. Thus, there is some overlap between signal estimation and parameter estimation. In this paper, we develop and analyze non-recursive and recursive methods which can deal with combined state and parameter estimation for nonlinear partially-observed stochastic systems.

It is well known that parameter identification is a challenging problem, especially when the systems are nonlinear and one wishes to do both parameter and state estimation. There are a number of significant works on this subject, including the books of Ljung and Soderström [14], Goodwin and Sin [6], Benveniste, Métiver and Priouret [2], and Kushner and Yin [12], as well as the papers of Dembo and Zeitouni [4], Campillo and Le Gland [3], and Hansen and Scheinkman [7]. Common methods used in parameter identification include least squares methods, methods of moments, maximum likelihood methods (often via the EM algorithm), and filtering methods.

In this paper, we introduce a non-recursive *asymptotic method of moments* (AMM) and a recursive *particle prediction error identification method* (PPEIM) for parameter estimation in filtering problems. For the asymptotic method of moments, we find the moments of the stationary distribution for the signal state by virtue of the Birkhoff ergodic theorem and obtain the unknown parameters through solving a system of non-singular finite order linear equations. For the particle prediction error identification method, we generalize the recursive least squares method and develop a novel algorithm so that it can be applied to general nonlinear stochastic systems.

For both methods, the branching particle method of nonlinear filtering introduced by Kouritzin (see Ballantyne, Chan and Kouritzin [1]) is used to do the signal state estimation. The algorithms developed in this paper have good robustness for unmodeled dynamics through the parameter estimation, since the output error

between the true system and the model at hand is minimized in the parameters at the observation times. Hence, even if we do not know what the exact mathematical form of a true system is, we can still often obtain satisfactory modeling and prediction results.

This paper is organized as follows. In Section 2, we propose algorithms for combined state and parameter estimation. Then, in Section 3, we present some simulation results based upon our algorithms designed in Section 2, which demonstrate that our algorithms provide effective numerical solutions to parameter estimation problems for nonlinear partially-observed stochastic systems. Finally, in Section 4, we give the derivation of the algorithms.

## 2. COMBINED STATE AND PARAMETER ESTIMATION

### 2.1. Asymptotic Method of Moments

Continuous-time Markov processes can be characterized conveniently by their infinitesimal generators. In Hansen and Scheinkman [7], generalized method of moments estimators and tests have been constructed for observable stationary and ergodic Markov processes using their generators. In this note, we consider partially-observed stochastic systems. The asymptotic method of moments for parameter estimation introduced in this section may be used for more general (discrete- or continuous-time) ergodic Markov processes taking values in a compact subset of $\mathbb{R}^d$. To explain explicitly, we constrain our signal process $X$ to be a one-dimensional continuous time stochastic process living within the interval $[0,1]$ and evolving according to the following Skorohod stochastic differential equation (SDE):

$$\begin{aligned} dX_t \;=\; & -\frac{1}{2}(a_0 + 2a_1X_t + \cdots + (n+1)a_nX_t^n)dt \\ & +\sigma_1 dB_t + \sigma_1^2\chi_{\{0\}}(X_t)d\xi_t^{(0)} - \sigma_1^2\chi_{\{1\}}(X_t)d\xi_t^{(1)}, \end{aligned} \tag{1}$$

where $a_0, a_1, \ldots, a_n$ are unknown parameters, $\sigma_1 > 0$ is a fixed constant, $B_t$ is a $\mathbb{R}^1$-standard Brownian motion, $\chi$ is the indicator function, and $\xi_t^{(0)}$, $\xi_t^{(1)}$ are the local times of $X_t$ at 0, 1, respectively. Throughout this note, we suppose that $\varepsilon > 0$ is a fixed constant and set $t_m = m\varepsilon$ for $m = 1, 2\ldots$ Let $\{V_{t_m}\}$ be a sequence of i.i.d. standard normal random variables independent of $\{X_t\}$ and $h$ be a continuous function on $[0,1]$ with a continuous inverse $h^{-1}$. We define

$$Y_{t_m} = h(X_{t_m}) + \sigma_2 V_{t_m},$$

where $\sigma_2 > 0$ is a fixed constant. Then, $Y$ is a "noisy distorted observation" of $X$.

We define a probability measure $\mu$ on the Borel $\sigma$-algebra of $[0,1]$ by

$$d\mu = \exp\left\{-\frac{1}{\sigma_1^2}(O_0 + a_0x + a_1x^2 + \cdots + a_nx^{n+1})\right\}dx,$$

where

$$O_0 \stackrel{\triangle}{=} \sigma_1^2 \log\left\{\int_0^1 \exp\left\{-\frac{1}{\sigma_1^2}(a_0x + a_1x^2 + \cdots + a_nx^{n+1})\right\}dx\right\}.$$

Then, $\mu$ is the unique invariant probability measure for $X_t$, i.e., $\int_0^1 p_t f d\mu = \int_0^1 f d\mu$ for any continuous function $f$ on $[0,1]$ and any $t \geq 0$, where $(p_t)_{t\geq 0}$ is the transition semigroup of $(X_t)_{t\geq 0}$. We define $c_k \stackrel{\triangle}{=} \int_0^1 h^k d\mu$, $k \geq 1$. Then, from the argument in Section 4.1, we find the following recursive formula

$$c_k = \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} Y_{t_i}^k - \sum_{j=0}^{k-1}\binom{k}{j}c_j EV_{t_1}^{k-j},$$

where

$$EV_{t_1}^l = \begin{cases} (2p-1)!!(\sigma_2^2)^p & \text{if } l = 2p \text{ for some } p \in \mathbb{Z}_+, \\ 0 \quad \text{otherwise.} \end{cases} \tag{2}$$

By approximating $h^{-1}$ via Bernstein's polynomials, we can obtain an expression of all the moments $\{O_k\}$ of $\mu$ via the limits of the observations, where $O_k = \int_0^1 x^k \mu(dx)$.

The algorithm of combined state and parameter estimation consists of the following steps:

**Initialization:** Let $\{\xi_t^i\}_{i=1}^{N_t}$ be a set of particles, where $N_t$ is the number of particles at time $t$. In this step, all particles are initialized so that their empirical measure approximates the initial signal distribution.

**Parameter estimation:** The parameter vector is estimated by the following asymptotic method of moments. First, based upon the accumulative observations, the moments $\{O_k\}_{k=1}^{2n+2}$ of $\mu$ are approximated using the method mentioned above. Then, the parameters are approximated via the solution of the following system of linear equations:

$$\begin{cases} \sigma_1^2(2O_1 - 1) = a_0(O_2 - O_1) + 2a_1(O_3 - O_2) + \cdots + (n+1)a_n(O_{n+2} - O_{n+1}), \\ \sigma_1^2(3O_2 - 2O_1) = a_0(O_3 - O_2) + 2a_1(O_4 - O_3) + \cdots + (n+1)a_n(O_{n+3} - O_{n+2}), \\ \cdots\cdots \\ \sigma_1^2((n+2)O_{n+1} - (n+1)O_n) = a_0(O_{n+2} - O_{n+1}) + 2a_1(O_{n+3} - O_{n+2}) \\ \qquad\qquad + \cdots + (n+1)a_n(O_{2n+2} - O_{2n+1}). \end{cases} \tag{3}$$

**Evolution:** In the evolution stage, each of the particles is evolved independently for the time period between observations (e.g. $t_m \to t_{m+1}$) according to the evolution equation (1) of the signal (e.g. $\xi_{t_m}^i \to \xi_{t_{m+1}-}^i$). Here, we use the approximate parameter vector obtained in the previous step.

**Selection:** Particles are branched according to their likelihood given the current observation. We define

$$\tau_{t_{m+1}}^i \triangleq \exp\left\{ \frac{\langle h(\xi_{t_{m+1}-}^i), Y_{t_{m+1}} \rangle - \frac{1}{2}|h(\xi_{t_{m+1}-}^i)|^2}{\sigma_2^2} \right\} - 1, \quad i = 1, 2, \cdots, N_{t_m}$$

and generate a uniform-$[0,1]$ random variable $\eta_{t_{m+1}}^i$ for each particle $\xi_{t_{m+1}-}^i$. When $\tau_{t_{m+1}}^i < 0$: the particle $\xi_{t_{m+1}-}^i$ is removed if $-\tau_{t_{m+1}}^i > \eta_{t_{m+1}}^i$, otherwise it maintains its current state. When $p \le \tau_{t_{m+1}}^i < p+1$ for some $p \ge 0$: $p$ independent copies of $\xi_{t_{m+1}-}^i$ are produced, and in addition, another copy of $\xi_{t_{m+1}}^i$ is produced if $\tau_{t_{m+1}}^i - p > \eta_{t_{m+1}}^i$. We denote the resulting set of particles by $\{\xi_{t_{m+1}}^i\}_{i=1}^{N_{t_{m+1}}}$.

The parameter estimation, evolution, and selection steps are repeated at each observation time.

### 2.2. Particle Prediction Error Identification Method

For parameter estimation in stochastic systems, there are two classes of identification methods: off-line and on-line (or recursive) methods. Many problems require the availability of a model with parameters updated in real time as the dynamical system evolves. Thus, in typical applications such as adaptive control, signal processing, and target detection and tracking, one generally prefers recursive algorithms due to their relative computational simplicity. Ljung [13] introduced a recursive prediction error method (so-called Ljung's scheme) which includes classical and generalized sequential least squares method, Robbins-Monro scheme (cf. [15]), Kiefer-Wolfowitz scheme (cf. [9]), and stochastic gradient parameter estimation method (cf. Kushner and Clark [11]) as special cases. The "prediction error" termed for the discrepancy between the measured data (observations) and the predicted data (calculated by conditional expectation from previous data according to a candidate model), is minimized over the model set using a stochastic approximation approach. The convergence analysis of Ljung's scheme is closely related to the asymptotic stability of a deterministic ordinary differential equation associated to the algorithm, giving rise to the so-called ODE method of stability analysis. Also, to keep the estimator process in a compact domain, Ljung used a *projection* (or resetting) mechanism that was not adequately described and analyzed until Gerencser [5] provided a rigorous convergence analysis for Ljung's scheme via stopping arguments. Generally speaking, Ljung's scheme can only be applied to parameter

estimation of linear stochastic systems. In this note, motivated from Ljung [13], Gerenscer [5] and Kouritzin [10], we propose a recursive particle prediction error method for parameter estimation in partially-observed nonlinear stochastic systems. In this section, we discuss the algorithm in detail. The derivation of the algorithm will be given in Section 4.2. The convergence analysis of the algorithm will be discussed in a separate paper.

Let us consider a discrete-time nonlinear stochastic system defined on a probability space $(\Omega, \mathcal{F}, P)$ with state-space form:

$$\begin{cases} X^{\beta}_{t_{m+1}} = f(\beta, X^{\beta}_{t_m}) + g(\beta, X^{\beta}_{t_m}) B_{t_{m+1}} \\ Y^{\beta}_{t_{m+1}} = h(X^{\beta}_{t_{m+1}}) + V_{t_{m+1}}, \ \ m = 0, 1, 2, \dots \end{cases} \tag{4}$$

where $\{X^{\beta}_{t_m}\}$ and $\{Y^{\beta}_{t_{m+1}}\}$ are $n-$ and $k-$dimensional system state and observation sequences respectively, $\beta$ is a $d-$dimensional unknown parameter vector to be estimated, $f(.,.)$, $g(.,.)$ and $h(.,.)$ are known $n \times 1$, $n \times s$ and $k \times 1$ nonlinear matrices respectively. The state noise $\{B_{t_{m+1}}\}$ and measurement noise $\{V_{t_{m+1}}\}$ are second-order $s-$ and $k-$dimensional random vector sequences respectively, and $\{V_{t_{m+1}}\}$ is independent of $(\{B_{t_{m+1}}\}, X_0)$. We assume that the "true" value of the parameter vector is $\beta^*$ and for convenience, we use $\{Y_{t_{m+1}}\} = \{Y^{\beta^*}_{t_{m+1}}\}$ to denote our actual observation data which correspond to our real signal with the "true" parameter vector. We define a filtration $\{\mathcal{F}_{t_m}\}_{m \geq 1}$ by

$$\mathcal{F}_{t_m} = \sigma\{Y_{t_1}, \cdots, Y_{t_m}\},$$

and denote by $E$ the expectation with respect to $P$. Furthermore, we assume that

$$E(B_{t_{m+1}}|\mathcal{F}_{t_m}) = 0, \quad E(V_{t_{m+1}}|\mathcal{F}_{t_m}) = 0, \quad E(V_{t_m} V^T_{t_l}) = \begin{cases} 0 & m \neq l \\ v & m = l, \end{cases} \tag{5}$$

where $v$ is a known $k \times k$ semi-positive definite matrix. We also assume that $f(\beta, x)$, $g(\beta, x)$ are continuously differentiable with respect to $\beta$ and $x$, and $h$ is continuously differentiable in $x$.

We assume that the parameter vector $\beta$ ranges over a compact subset $D$ of $\mathbb{R}^d$. Given the observations, we will estimate the conditional distribution for the signal state and the unknown parameters. The algorithm of combined state and parameter estimation consists of the following steps:

**Initialization**: $N$ particles $\{\xi^i\}^N_{i=1}$ are independently initialized so that their empirical measure approximates the initial signal distribution, and the initial guess $\widehat{\beta}_0$ of the parameters is set.

**Evolution**: Each particle is evolved independently according to the signal model:

$$\xi^i_{t_{m+1}-} = f(\widehat{\beta}_{t_m}, \xi^i_{t_m}) + g(\widehat{\beta}_{t_m}, \xi^i_{t_m}) B^i_{t_{m+1}},$$

where $B^i_{t_{m+1}}$ are independent with the same distribution as $B_{t_{m+1}}$ for $i = 1, 2, \cdots, N_{t_m}$; $N_{t_m}$ being the number of particles at time $t_m$. Also, we have the following evolution for the gradient of each particle with respect to parameters:

$$\begin{aligned} \nabla_\beta \xi^i_{t_{m+1}-} = & [\partial_\beta f(\widehat{\beta}_{t_m}, \xi^i_{t_m}) + \partial_{\xi^i_{t_m}} f(\widehat{\beta}_{t_m}, \xi^i_{t_m}) \nabla_\beta \xi^i_{t_m}] + \\ & + [\partial_\beta g(\widehat{\beta}_{t_m}, \xi^i_{t_m}) B^i_{t_{m+1}} + \partial_{\xi^i_{t_m}} g(\widehat{\beta}_{t_m}, \xi^i_{t_m}) B^i_{t_{m+1}} \nabla_\beta \xi^i_{t_m}]. \end{aligned}$$

**Parameter estimation**: We update parameter estimators via

$$\widehat{\beta}_{t_{m+1}} = \widehat{\beta}_{t_m} + \frac{1}{(m+1)^\alpha} \varphi_{t_{m+1}} [Y_{t_{m+1}} - \widehat{Y}_{t_{m+1}}],$$

where

$$\widehat{Y}_{t_{m+1}} = E[h(X^{\beta}_{t_{m+1}})|\mathcal{F}_{t_m}]\big|_{\beta = \widehat{\beta}_{t_m}}$$

which is approximated by

$$\frac{1}{N_{t_m}}\sum_{i=1}^{N_{t_m}} h(\xi^i_{t_{m+1}-}),$$

and

$$\varphi_{t_{m+1}} = E\left[\nabla h(X^\beta_{t_{m+1}})\nabla_\beta X^\beta_{t_{m+1}}\Big|\mathcal{F}_{t_m}\right]^T\Big|_{\beta=\widehat{\beta}_{t_m}}$$

which is approximated by

$$\frac{1}{N_{t_m}}\sum_{i=1}^{N_{t_m}}\left(\nabla h(\xi^i_{t_{m+1}-})\nabla_{\widehat{\beta}_{t_m}}\xi^i_{t_{m+1}-}\right)^T$$

In order to keep the estimators within $D$, we use the following resetting mechanism. If $\widehat{\beta}_{t_{m+1}-}\in D$, we let $\widehat{\beta}_{t_{m+1}} = \widehat{\beta}_{t_{m+1}-}$, otherwise, we let $\widehat{\beta}_{t_{m+1}} = \widehat{\beta}_0$. We can use averaging technique to speed up the convergence of the parameter estimators by defining

$$\overline{\beta}_{t_{m+1}} = \frac{m}{m+1}\overline{\beta}_{t_m} + \frac{1}{m+1}\widehat{\beta}_{t_{m+1}},$$

which converges to $\beta^*$ as $m\to\infty$ at a noticeably better rate.

**Selection:** Particles are branched according to their likelihood given the current observation. This step is similar to that in Section 2.1.

The evolution, parameter estimation and selection steps are repeated at each observation time.

## 3. SIMULATION RESULTS

### 3.1. Signal Models

For both methods, we generate a one dimensional signal within the interval [0,1]. For the asymptotic method of moments, we use the polynomial model (1) as in Section 2.1. For the particle prediction error identification method, the signal model is

$$\begin{aligned} dX_t \;=\; & -\frac{1}{2}\left(a_0+\sqrt{2}\sum_{i=1}^{n}(-a_i\sin 2\pi i X_t + b_i\cos 2\pi i X_t)\right)dt \\ & +\sigma_1 dB_t + \sigma_1^2\chi_{\{0\}}(X_t)d\xi^{(0)}_t - \sigma_1^2\chi_{\{1\}}(X_t)d\xi^{(1)}_t, \end{aligned}$$

where $a_0, a_1, b_1, \ldots, a_n, b_n$ are unknown parameters, $B_t$ is a $\mathbb{R}^1$-standard Brownian motion, $\chi$ is the indicator function, and $\xi^{(0)}_t, \xi^{(1)}_t$ are local times of $X_t$ at $0, 1$, respectively. Due to reflection at boundary points 0 and 1, the signal $X_t$ stays within the interval [0,1]. Let $t_m = m\varepsilon$ ($\varepsilon > 0$ is a small constant). For our recursive particle prediction error identification method discussed in Section 2.2, we use Euler scheme to discretize the (continuous) signal model to be:

$$\begin{aligned} X_{t_{m+1}-} \;=\; & X_{t_m} - \frac{\varepsilon}{2}\left(a_0+\sqrt{2}\sum_{i=1}^{n}(-a_i\sin 2\pi i X_{t_m} + b_i\cos 2\pi i X_{t_m})\right) \\ & +\sigma_1(B_{t_{m+1}} - B_{t_m}). \end{aligned} \tag{6}$$

If $X_{t_{m+1}-}\in[0,1]$, then we set $X_{t_{m+1}} = X_{t_{m+1}-}$. Otherwise, we let $X_{t_{m+1}}$ be the value of the reflecting point of $X_{t_{m+1}-}$ within [0,1]. For instance, if $X_{t_{m+1}-} = 1.01$, then $X_{t_{m+1}} = 0.99$. So, in this discretized model, we have

$$f(\beta, X_{t_m}) = X_{t_m} - \frac{\varepsilon}{2}\left(a_0+\sqrt{2}\sum_{i=1}^{n}(-a_i\sin 2\pi i X_{t_m} + b_i\cos 2\pi i X_{t_m})\right) \tag{7}$$

and $g(\beta, X_{t_m}) = \sigma_1$ with $\beta = (a_0, a_1, b_1, \cdots, a_n, b_n)^T$. In our simulation, the particles will be evolved independently according to the law of the discrete model (6). One can easily compute the partial derivatives of $f(\beta, x)$ with respect to $\beta$ and $x$ using the expression (7). We omit the details.

### 3.2. Observations

For the asymptotic method of moments, we use the nonlinear observation model

$$Y_{t_m} = \sqrt{X_{t_m}} + V_{t_m}, \tag{8}$$

where $V_{t_m}$ are i.i.d. random variables with $N(0, \sigma_2)$-distribution. For the particle prediction error identification method, the domain is chopped into pixels and at any time $t_m$ the observation $Y_{t_m}$ is a vector which consists of the information from every pixel. More precisely, the observation has the form

$$Y_{t_m} = \langle Y^0_{t_m}, Y^1_{t_m}, \ldots, Y^{N_{obs}}_{t_m} \rangle.$$

Here $N_{obs}$ is the total number of pixels on the domain,

$$Y^i_{t_m} = \frac{1}{\sqrt{2\pi}\lambda} exp\left( -\frac{(X_{t_m} - \frac{i}{N_{obs}})^2}{2\lambda^2} \right) + V^i_{t_m}, \quad i = 0, 1, \cdots, N_{obs}, \tag{9}$$

$V^i_{t_m}$, $i \in \{0, 1, \cdots, N_{obs}\}$, are i.i.d. random variables with $N(0, \sigma_2)$-distribution and $\lambda > 0$ is a fixed constant.

Simulation results are presented for three-dimensional parameter vectors. In our simulation for filtering model (1,8), we use the true value of the parameter vector $(a_0, a_1, a_2) = (0.3, -0.5, -0.3)$. The values of signal noise and observation noise are $\sigma_1 = 0.4$, $\sigma_2 = 0.1$, respectively. In our simulation for filtering model (6,9), we use the true values of the parameters $a_0 = -0.5, a_1 = 0.3$ and $b_1 = -0.3$, and take $\sigma_1 = 0.2, \sigma_2 = 0.4$, $\lambda = 0.1$ and $N_{obs} = 10$. For both methods, the simulation time is 22.2 minutes and the time period between observations $\varepsilon = 4.44 \times 10^{-5}$ minute.

In Figures 1 and 2, the solid lines represent the true parameters whereas the dotted are the current parameter estimates for Model (1,8) and Model (6,9) based upon AMM and PPEIM, respectively. Clearly, all three parameter estimates are converging to the true parameters. Figures 3 and 4 compare the true signal states to their approximate-filter conditional mean estimates for Model (1,8) and Model (6,9), respectively. Here, the filters do not have access to the true parameters, but rather must contend with noisy estimates. As time increases these estimates improve and the filter performs markedly better.

The simulation results show that our non-recursive asymptotic method of moments and recursive particle prediction error identification method provide effective numerical solutions to the problem of parameter estimation for nonlinear partially-observed stochastic systems.

## 4. DERIVATION OF THE ALGORITHMS

### 4.1. Non-Recursive Asymptotic Method of Moments

From the argument below, one can see that the method introduced in this section can also be used to handle weakly correlated noises satisfying some mixing condition. Note that for the reflecting diffusion described by Skorohod SDE (1), the associated diffusion generator $(\mathcal{L}, D(\mathcal{L}))$ is given by

$$\begin{cases} \mathcal{L}f(x) = \frac{1}{2}\sigma_1^2 f''(x) - \frac{1}{2}(a_0 + 2a_1 x + \cdots + (n+1)a_n x^n) f'(x), \ f \in D(\mathcal{L}), \\ D(\mathcal{L}) \stackrel{\triangle}{=} \left\{ f \in L^2([0,1], dx) : f'' \in L^2([0,1], dx) \text{ and } f'(0) = f'(1) = 0 \right\}. \end{cases}$$

We denote by $P$ the law of $X_\cdot$ with a fixed initial distribution and $E$ the expectation with respect to $P$. Then, the signal process $(X_t, P)$ is ergodic, that is, for every event $G$ in the path space $C_{[0,1]}[0, \infty)$ that is invariant under time shifts, $P(X_\cdot \in G) = 0$ or 1.

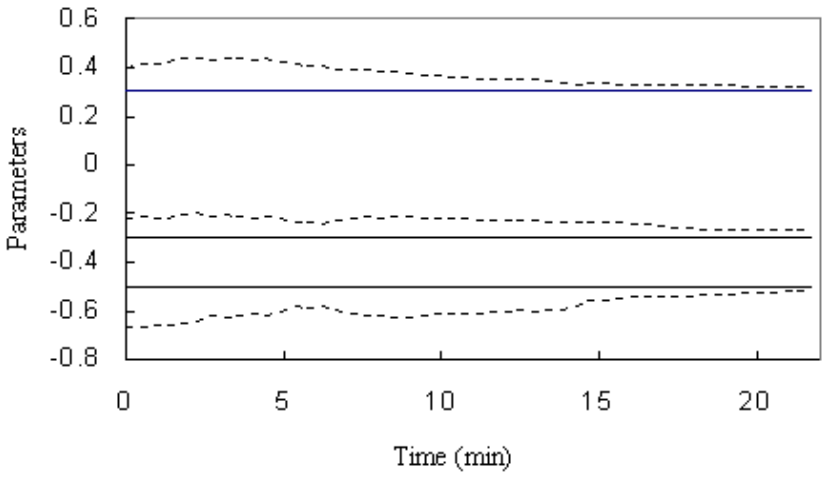


**Figure 1.** Parameter estimation via AMM for Model (1,8) with $a_0 = 0.3, a_1 = -0.5$ and $a_2 = -0.3$

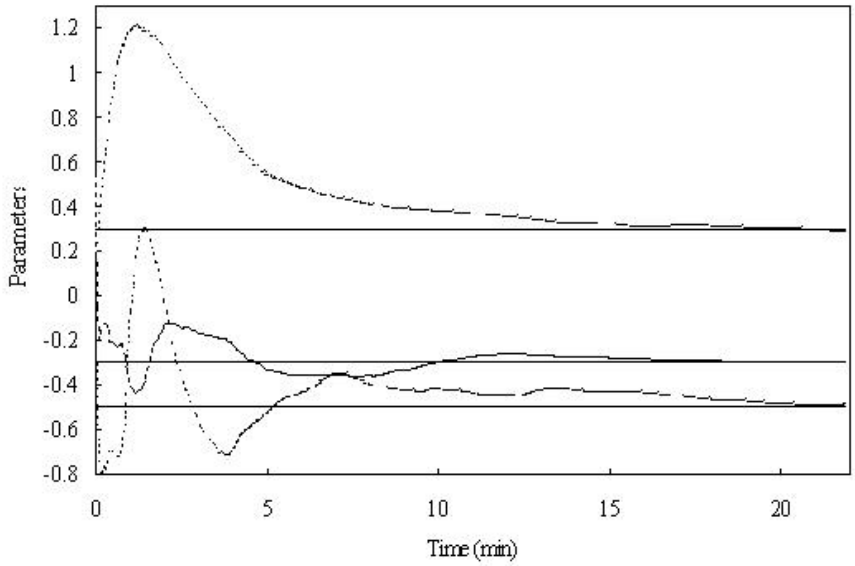


**Figure 2.** Parameter estimation via PPEIM for Model (6,9) with $a_0 = -0.5, a_1 = 0.3$ and $b_1 = -0.3$

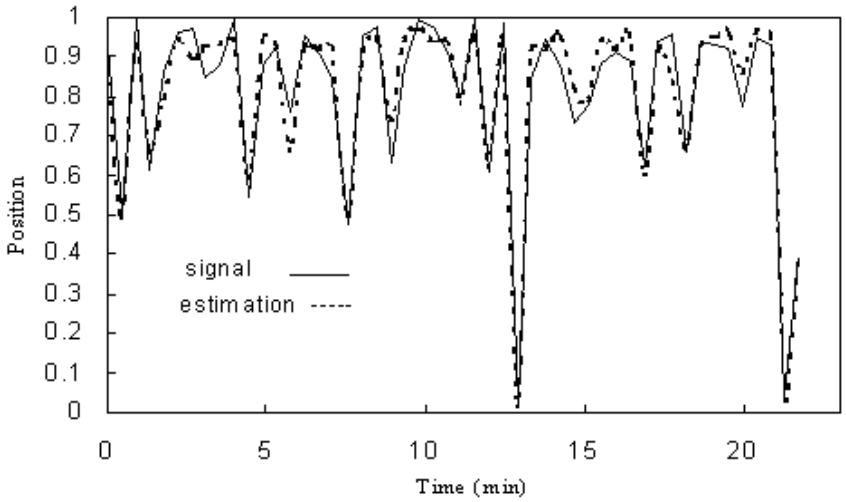


**Figure 3.** Branching particle filter for Model (1,8) using AMM parameter estimates

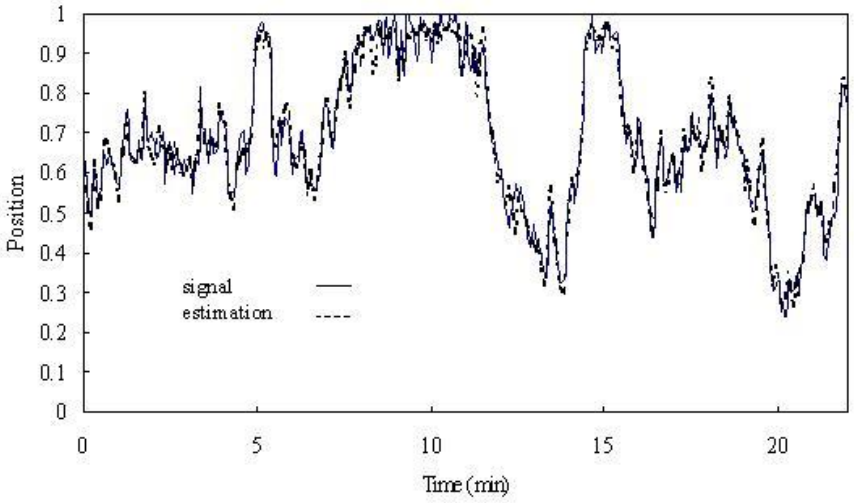


**Figure 4.** Branching particle filter for Model (6,9) using PPEIM parameter estimates

By independence, the boundedness of $h$ and the Birkhoff ergodic theorem, one has that for arbitrary $k \in \mathbb{N}$ and $0 \le j \le k$,

$$\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} h^j(X_{t_i}) V_{t_i}^{k-j} = \left( \int_0^1 h^j d\mu \right) E V_{t_1}^{k-j}, \quad P - a.s.$$

Therefore, we find that

$$\begin{aligned} \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_{t_i}^k &= \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} (h(X_{t_i}) + V_{t_i})^k \\ &= \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} \sum_{j=0}^{k} \binom{k}{j} h^j(X_{t_i}) V_{t_i}^{k-j} \\ &= \sum_{j=0}^{k} \binom{k}{j} c_j E V_{t_1}^{k-j}. \end{aligned} \tag{10}$$

Hence, by induction, (10) and (2), we find that

$$c_1 = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_{t_i}, \quad c_2 = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_{t_i}^2 - \sigma_2^2, \quad c_3 = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_{t_i}^3 - 3c_1\sigma_2^2, \quad \dots$$

By approximating $h^{-1}$ via Bernstein's polynomials, we can obtain an expression for all the moments $\{O_k\}$ of $\mu$ via the limits of the observations.

By integration by parts, we get for $k \in \mathbb{Z}_+$,

$$\begin{aligned} (k+1)O_k &= (k+1) \int_0^1 x^k \mu(dx) \\ &= (k+1) \int_0^1 x^k \exp\left\{ -\frac{1}{\sigma_1^2} (O_0 + a_0 x + a_1 x^2 + \cdots + a_n x^{n+1}) \right\} dx \\ &= \exp\left\{ -\frac{1}{\sigma_1^2} (O_0 + a_0 + a_1 + \cdots + a_n) \right\} \\ &\quad + \frac{1}{\sigma_1^2} (a_0 O_{k+1} + 2a_1 O_{k+2} + \cdots + (n+1) a_n O_{k+n+1}). \end{aligned} \tag{11}$$

Subtracting $kO_{k-1}$ from $(k+1)O_k$, we get from (11) the system of linear equations (3).

We define a new measure $\eta(dx) \stackrel{\triangle}{=} (x - x^2)\mu(dx)$. Then, the non-singularity of the system of linear equations (3) is equivalent to that of the following matrix

$$M = \begin{pmatrix} \int_0^1 \eta(dx) & \int_0^1 x\eta(dx) & \cdots & \int_0^1 x^n \eta(dx) \\ \int_0^1 x\eta(dx) & \int_0^1 x^2\eta(dx) & \cdots & \int_0^1 x^{n+1}\eta(dx) \\ \cdots\cdots & & & \\ \int_0^1 x^n\eta(dx) & \int_0^1 x^{n+1}\eta(dx) & \cdots & \int_0^1 x^{2n}\eta(dx) \end{pmatrix}.$$

However, non-singularity of this matrix follows from the fact that for any $\xi_0, \xi_1, \dots, \xi_n \in \mathbb{R}$ with $\xi_0 \cdot \xi_1 \cdots \xi_n \neq 0$, we have that

$$(\xi_0, \xi_1, \dots, \xi_n) M (\xi_0, \xi_1, \dots, \xi_n)^T = \int_0^1 (\xi_0 + \xi_1 x + \dots + \xi_n x^n)^2 \eta(dx) > 0.$$

Thus, $M$ is in fact a positive definite matrix. Therefore, the unknown parameter vector $(a_0, \dots, a_n)$ is the unique solution to the system of linear equations (3).

### 4.2. Recursive Particle Prediction Error Identification Method

For the discrete-time partially-observed nonlinear stochastic system described by (4), we will use the least squares method with discounted measurements to estimate the unknown parameters. Let $\{\gamma(i)\}_{i=1}^{\infty}$ be a sequence of positive real numbers tending to zero and satisfying $\sum_{i=1}^{\infty}\gamma(i)=\infty$. Also, we let

$$\delta(m,i)=\gamma(i)\prod_{k=i+1}^{m}[1-\gamma(k)],\ 1\le i\le m-1,\quad \delta(m,m)=\gamma(m).$$

It is easy to see that for $1\le i\le m-1$

$$\delta(m,i)=[1-\gamma(m)]\delta(m-1,i). \tag{12}$$

Take $||\cdot||$ to be Euclidean distance. Then, the "off-line" estimator for $\beta$ is defined as:

$$\tilde{\beta}_{t_{m+1}}=\arg\min_{\beta\in D}\left\{\frac{1}{m+1}\sum_{i=0}^{m}\delta(m+1,i+1)||Y_{t_{i+1}}-E[Y^{\beta}_{t_{i+1}}|\mathcal{F}_{t_i}]||^2\right\},$$

where the predictor $E[Y^{\beta}_{t_{i+1}}|\mathcal{F}_{t_i}]$ is the prediction of the $(i+1)$-th observation based upon measurements up to time $t_i$. By (4) and the second formula of (5), we find that

$$\tilde{\beta}_{t_{m+1}}=\arg\min_{\beta\in D}\left\{\frac{1}{m+1}\sum_{i=0}^{m}\delta(m+1,i+1)||Y_{t_{i+1}}-E[h(X^{\beta}_{t_{i+1}})|\mathcal{F}_{t_i}]||^2\right\}. \tag{13}$$

To get a recursive estimator close to $\tilde{\beta}_{t_{m+1}}$, we differentiate (13) and use the chain rule to yield

$$-\sum_{i=0}^{m}\delta(m+1,i+1)E\left[\nabla h(X^{\beta}_{t_{i+1}})\nabla_{\beta}X^{\beta}_{t_{i+1}}\Big|\mathcal{F}_{t_i}\right]^T\{Y_{t_{i+1}}-E[h(X^{\beta}_{t_{i+1}})|\mathcal{F}_{t_i}]\}=0, \tag{14}$$

where $\nabla_{\beta}X^{\beta}_{t_{i+1}}$ is an $n\times d$ matrix and satisfies the following equation:

$$\begin{aligned}\nabla_{\beta}X^{\beta}_{t_{i+1}} &= [\partial_{\beta}f(\beta,X^{\beta}_{t_i})+\partial_{X^{\beta}_{t_i}}f(\beta,X^{\beta}_{t_i})\nabla_{\beta}X^{\beta}_{t_i}]\\ &\quad+[\partial_{\beta}g(\beta,X^{\beta}_{t_i})B_{t_{i+1}}+\partial_{X^{\beta}_{t_i}}g(\beta,X^{\beta}_{t_i})B_{t_{i+1}}\nabla_{\beta}X^{\beta}_{t_i}].\end{aligned}$$

Here, we have used the smoothness of $f,g,h$ and the bounded convergence to interchange conditional expectation and differentiation. Using Taylor's theorem to expand $h(X^{\beta}_{t_{i+1}})$ in the unknown parameter $\beta$, we have that

$$h(X^{\beta}_{t_{i+1}})-h(X^{\beta^*}_{t_{i+1}})=\nabla h(X^{\beta}_{t_{i+1}})\nabla_{\beta}X^{\beta}_{t_{i+1}}(\beta-\beta^*)+R_{t_{i+1}}(\beta,\beta^*), \tag{15}$$

where $R_{t_{i+1}}(\beta,\beta^*)$ is the remainder. Substituting (15) into (14), we then find that

$$\begin{aligned}&\sum_{i=0}^{m}\delta(m+1,i+1)E\left[\nabla h(X^{\beta}_{t_{i+1}})\nabla_{\beta}X^{\beta}_{t_{i+1}}\Big|\mathcal{F}_{t_i}\right]^T E\left[\nabla h(X^{\beta}_{t_{i+1}})\nabla_{\beta}X^{\beta}_{t_{i+1}}\Big|\mathcal{F}_{t_i}\right](\beta-\beta^*)\\ &\quad=\sum_{i=0}^{m}\delta(m+1,i+1)E\left[\nabla h(X^{\beta}_{t_{i+1}})\nabla_{\beta}X^{\beta}_{t_{i+1}}\Big|\mathcal{F}_{t_i}\right]^T(Y_{t_{i+1}}-E[h(X^{\beta^*}_{t_{i+1}})|\mathcal{F}_{t_i}])\\ &\qquad-\sum_{i=0}^{m}\delta(m+1,i+1)E\left[\nabla h(X^{\beta}_{t_{i+1}})\nabla_{\beta}X^{\beta}_{t_{i+1}}\Big|\mathcal{F}_{t_i}\right]^T\overline{R}_{t_{i+1}}(\beta,\beta^*),\end{aligned}$$

where

$$\overline{R}_{t_{i+1}}(\beta,\beta^*)\triangleq E[R_{t_{i+1}}(\beta,\beta^*)|\mathcal{F}_{t_i}].$$

We let the recursive estimators $\{\widehat{\beta}_{t_{m+1}},\, m=0,\, 1,\, \cdots\}$ for $\beta$ be determined by:

$$\begin{aligned}\widehat{\beta}_{t_{m+1}} - \beta^* &= \left(\sum_{i=0}^{m} \delta(m+1, i+1)\varphi_{t_{i+1}}\, \varphi_{t_{i+1}}^T\right)^{-1} \\ &\quad \times \sum_{i=0}^{m} \delta(m+1, i+1)\varphi_{t_{i+1}} \left\{Y_{t_{i+1}} - E[h(X_{t_{i+1}}^{\beta^*})|\mathcal{F}_{t_i}] - \overline{R}_{t_{i+1}}(\widehat{\beta}_{t_i}, \beta^*)\right\}, \end{aligned} \tag{16}$$

where

$$\varphi_{t_{i+1}} \stackrel{\triangle}{=} E\left[\nabla h(X_{t_{i+1}}^{\beta})\nabla_\beta X_{t_{i+1}}^{\beta}\Big|\, \mathcal{F}_{t_i}\right]^T\Big|_{\beta:=\widehat{\beta}_{t_i}} \in \mathbb{R}^{d\times k}, \quad \forall i \geq 0, \tag{17}$$

and $\varphi_0$ is chosen such that $\varphi_0\varphi_0^T > 0$. We set

$$Q_{m+1} = \sum_{i=0}^{m} \delta(m+1, i+1)\varphi_{t_{i+1}}\varphi_{t_{i+1}}^T.$$

Then, one can easily find that

$$Q_{m+1} = (1-\gamma(m+1))Q_m + \gamma(m+1)\varphi_{t_{m+1}}\varphi_{t_{m+1}}^T. \tag{18}$$

It follows from (16), (18) and (12) that

$$\begin{aligned}\widehat{\beta}_{t_{m+1}} - \beta^* &= Q_{m+1}^{-1}\left\{\sum_{i=0}^{m-1} \delta(m+1, i+1)\varphi_{t_{i+1}}(Y_{t_{i+1}} - E[h(X_{t_{i+1}}^{\beta^*})|\mathcal{F}_{t_i}] - \overline{R}_{t_{i+1}}(\widehat{\beta}_{t_i}, \beta^*))\right. \\ &\quad \left. +\gamma(m+1)\, \varphi_{t_{m+1}}(Y_{t_{m+1}} - E[h(X_{t_{m+1}}^{\beta^*})|\mathcal{F}_{t_m}] - \overline{R}_{t_{m+1}}(\widehat{\beta}_{t_m}, \beta^*))\right\} \\ &= Q_{m+1}^{-1}\{(1-\gamma(m+1))Q_m(\widehat{\beta}_{t_m} - \beta^*) \\ &\quad +\gamma(m+1)\varphi_{t_{m+1}}(Y_{t_{m+1}} - E[h(X_{t_{m+1}}^{\beta^*})|\mathcal{F}_{t_m}] - \overline{R}_{t_{m+1}}(\widehat{\beta}_{t_m}, \beta^*))\} \\ &= Q_{m+1}^{-1}\{Q_{m+1}(\widehat{\beta}_{t_m} - \beta^*) + \gamma(m+1)\varphi_{t_{m+1}}[-\varphi_{t_{m+1}}^T(\widehat{\beta}_{t_m} - \beta^*) + Y_{t_{m+1}} \\ &\quad -E[h(X_{t_{m+1}}^{\beta^*})|\mathcal{F}_{t_m}] - \overline{R}_{t_{m+1}}(\widehat{\beta}_{t_m}, \beta^*)]\} \\ &= \widehat{\beta}_{t_m} - \beta^* + \gamma(m+1)Q_{m+1}^{-1}\varphi_{t_{m+1}}\left(Y_{t_{m+1}} - E[h(X_{t_{m+1}}^{\beta})|\mathcal{F}_{t_m}]|_{\beta=\widehat{\beta}_{t_m}}\right). \end{aligned}$$

Thus, we have

$$\widehat{\beta}_{t_{m+1}} \stackrel{\triangle}{=} \widehat{\beta}_{t_m} + \gamma(m+1)Q_{m+1}^{-1}\varphi_{t_{m+1}}(Y_{t_{m+1}} - \widehat{Y}_{t_{m+1}}), \tag{19}$$

where

$$\widehat{Y}_{t_{m+1}} = E[h(X_{t_{m+1}}^{\beta})|\mathcal{F}_{t_m}]\Big|_{\beta=\widehat{\beta}_{t_m}}. \tag{20}$$

The prediction error identification algorithm for the system (4) is made up of (19), (20), (17), and (18). In the above algorithm, we choose the initial value $\widehat{\beta}_0$ arbitrarily in a compact domain $D$. We remark that we can prove that $Q_{m+1}^{-1}$ converges to some positive definite matrix under certain conditions. Thus, in the practical implementation of the above algorithm, we take $\gamma(i) = 1/i^\alpha$ $(0 < \alpha < 1)$ and replace $Q_{m+1}^{-1}$ by the identity $d \times d$ matrix which does not affect the convergence of the parameter estimators. Since we are dealing with the nonlinear stochastic systems, we employ the branching particle filter method discussed in Section 2 to calculate the approximate values for $\varphi_{t_{i+1}}$ in (17) and $\widehat{Y}_{t_{m+1}}$ in (20).

## ACKNOWLEDGMENTS

The authors gratefully acknowledge support from Lockheed Martin Naval Electronics and Surveillance Systems, NSERC, Lockheed Martin Canada, APR. Inc., and PIMS through sponsorship of the MITACS-PINTS centre of excellence.

## REFERENCES

[1] D. Ballantyne, H. Chan, and M. Kouritzin, "A novel branching particle method for tracking", *Proceedings of SPIE, Signal and Data Processing of Small Targets 2000*, **4048**, pp. 277-287, 2000.

[2] A. Benveniste, M. Metivier, and P. Priouet, *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, Berlin, Heidelberg, 1990.

[3] F. Campillo and F. Le Gland, "MLE for partially observed diffusions: direct maximization vs. the EM algorithm", *Stochastic Processes and their Applications*, **33**, pp. 245-274, 1989.

[4] A. Dembo and O. Zeitouni, "Parameter estimation of partially observed continuous-time stochastic processes via EM algorithm", *Stochastic Processes and their Applications*, **23**, pp. 91-113, 1986.

[5] L. Gerencser, "Rate of convergence of recursive estimators", *SIAM J. Control and Optimization*, **30**, pp. 1200-1227, 1992.

[6] G. C. Goodwin and K. Sin, *Adaptive Filtering, Prediction, and Control*, Prentice Hall, Englewood Cliffs, New Jersey, 1984.

[7] L.P. Hansen and J.A. Scheinkman, "Back to the future: generating moment implications for continuous-time Markov processes", *Econometrica*, **63**, pp. 767-804, 1995.

[8] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[9] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function", *Ann. Math. Statist.*, **23**, pp. 462-466, 1952

[10] M.A. Kouritzin, "Inductive methods and rates of $r^{\text{th}}$-mean convergence in adaptive filtering", *Stochastic and Stochastic Reports*, **51**, pp. 241-266, 1994.

[11] H.J. Kushner and D.S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*, Springer-Verlag, Berlin and New York, 1978.

[12] H.J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York, Berlin, Heidelberg, 1997.

[13] L. Ljung, "Analysis of recursive stochastic algorithms", *IEEE Trans. Automat. Control*, **AC-22**, pp. 551-575, 1977.

[14] L. Ljung and T. Soderström, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, 1983.

[15] H. Robbins and S. Monro, "A stochastic approximation method", *Ann. Math. Statist.*, **22**, pp. 400-407, 1951.