Nonparametric filtering for stochastic nonlinear oscillations

Isao Shoji
Tokyo University of Science, Fujimi, Chiyoda, Tokyo 102-0071, Japan

(Received 6 July 2020; accepted 25 October 2020; published 30 November 2020)

This paper proposes a method of nonparametric filtering for stochastic nonlinear oscillations with particular interest in their derivative estimation. Based on a second-order ordinary differential equation, a stochastic oscillation is modeled by a two-variate stochastic differential equation without specifying the function form of the drift function, where the first variable is assumed to be observable but not the other. Given the discrete time series with observation error, the proposed method enables us to estimate the values of the drift function and its derivatives including those of the unobservable variable. According to the results of numerical experiments to compare estimation accuracy with a parametric method, the proposed method shows better performance in the estimation of nonlinear models.

DOI: 10.1103/PhysRevE.102.052221

I. INTRODUCTION

Nonlinear oscillations can be seen in so many fields of sciences that various methods for their nonlinear analysis are reported from natural science to social science; for example, in neuroscience neural oscillations are modeled by nonlinear oscillations which are estimated from EEG time series [1–5], in geoscience oscillations of climate are modeled by nonlinear stochastic differential equations and applied to proxy data for temperature [6,7], in astronomy nonlinear oscillations are used for modeling time evolution of the light emitted from the center of galaxies [8,9], in finance those models are used for describing time series of asset prices [10,11], and so on. Mathematically their dynamics could be modeled by the following second-order ordinary differential equation:

$$\ddot{x} = f(x, \dot{x}).$$

To study the dynamic behavior of the solution to the differential equation from a geometric viewpoint, analysis of the trajectory on the phase space can be performed. For this purpose, the following equivalent first-order differential equation is often used as well:

$$dx_1 = x_2 dt$$
$$dx_2 = f(x_1, x_2)dt,$$

where $x_1 = x$ and $x_2 = \dot{x}$.

Then, investigating a trajectory implied by the differential equation on the two-dimensional phase space, we can grasp geometrically the dynamic behavior of the original x. Various trajectories $(x_1(t), x_2(t))$ can be produced depending on the function form of f; circular trajectories can be observed in some cases where f is linear and more complex trajectories can be seen in nonlinear cases, Van der Pol oscillation and Duffing oscillation, for example. In addition, the derivatives of f are necessary for investigating the behavior of a trajectory because the trajectory can be characterized by its vector field $V = (\dot{x}_1, \dot{x}_2) = (x_2, f)$ and the acceleration $dV/dt = (\dot{x}_2, \dot{f}) = [f, (\partial f/\partial x_1)x_2 + (\partial f/\partial x_2)f]$. Hence, the values of f and its derivatives are essential quantities for analysis of the trajectory.

Turing our attention to real applications, internal or external disturbances cause unexpected fluctuations so that stochastic differential equations (SDEs) may be more useful for describing the dynamics than the ordinary differential equation. Taking the above differential equation, its stochastic version can be expressed, for example, as follows:

$$dX_{1,t} = X_{2,t}dt,$$
$$dX_{2,t} = f(X_{1,t}, X_{2,t})dt + \sigma dB_t,$$

where X_1 and X_2 correspond to x_1 and x_2, respectively, and B_t is a standard Brownian motion. So, fitting (X_1, X_2) into realized trajectories, we can characterize the dynamic behavior of our interest more clearly by the estimated model.

However, the modeling is not easy. In the first place, we have almost no information on the function form of f before analysis, especially for highly complex real systems, although what function is used for f plays a crucial role in describing the dynamics. In addition, we cannot always obtain full information, e.g., information of both X_1 and X_2, necessary for estimation. Suppose modeling the dynamics of heart beat or a stock price index for example. We can easily observe electrical impulses of heart beat or its price but not observe their time rate of change, e.g., their velocity, much less their acceleration. Therefore the estimation must be conducted only by partially observed data.

Taking these restriction into account, we have to estimate an SDE. When estimating models without specifying the function form, we could use such a nonparametric method as the kernel regression which is a method of fitting a polynomial function into data locally [12–14]. However, the regression method is not useful for our analysis because it needs data of the objective and explanatory variables and thus data of both X_1 and X_2 would be required for estimation. To estimate

a model from partially observed data with nonlinear and/or non-Gaussian systems, unscented Kalman filter [2,6,15–19], particle filters [20,21] or sequential Monte Carlo methods [22,23] could be used. However, because our main interest is to estimate f and particularly its derivatives, direct application of these methods does not fit our purpose.

The aim of this paper is to propose a method of estimating stochastic oscillations under the restrictions above. To this end, it may be useful to use a nonparametric method proposed in Ref. [24], which was developed to estimate $X_{1,t}$ and $X_{2,t}$ from partially observed data without specifying f. As a by-product, the method could produce estimates of the values of f and its derivatives. Though the paper investigated the estimation accuracy for $X_{1,t}$ and $X_{2,t}$, the paper gave no quantitative evidence of the estimation accuracy for f nor its derivatives. Moreover, the paper assumed no observation errors, which is not general enough to handle filtering problems, although it may be desirable for practical use to include the observation errors. Due to these drawbacks, it is questionable whether the method can practically work as an estimation method of f and its derivatives. In this paper, we conduct a numerical study of investigating the estimation accuracy of f and its derivatives together with observation errors. We compare the performance of the proposed method with that of the parametric method where f is modeled by a polynomial function which contains the exact function form of f as a subset. Through the comparison, we will show that the proposed method can be used to estimate nonparametrically f and its derivatives from partially observed data.

The paper is organized as follows. In Sec. II we propose a nonparameteric filtering method to estimate f and its derivatives from discrete observations. In Sec. III we conduct numerical experiments to compare estimation accuracy of the proposed method with that of a parametric method. We conclude in the final section.

II. NONPARAMETRIC FILTERING: NPF METHOD

We consider a stochastic process which satisfies the following two-variate stochastic differential equation:

$$dX_{1,t} = X_{2,t}dt, \tag{1}$$

$$dX_{2,t} = f(X_{1,t}, X_{2,t})dt + \sigma dB_t. \tag{2}$$

Here we assume $X_{1,t}$ to be observable but $X_{2,t}$ to be unobservable. In addition, given observed discrete time series data $\{Z_{t_k}\}_{1\leqslant k\leqslant n}$, because data always contain observation errors, the process Z_{t_k} is assumed to be given by

$$Z_{t_k} = X_{1,t_k} + \varepsilon_k, \tag{3}$$

where ε_k is independently identically normally distributed with mean zero and variance σ_ε^2. Hence the statistical model for estimation comes to a state space model, whose system equation is characterized by SDE (1) and (2) and whose observation equation is expressed as (3).

Here note the qualitative difference between the system equation and the observation equation; the former is formulated in continuous time but the latter in discrete time. To estimate from discrete observations, the SDE needs to be discretized with attention to no assumption on the function form of f. To this end, we use a nonparametric method proposed in Ref. [24]. On the basis of mth-order expansion, the approximate discretized version of the state space model is given as follows. Let the state vector ξ_k at discrete time t_k be a vector in $R^M[M = 2 + m(m+1)/2)]$, which is given by

$$\xi_k = \big(X_{1,t_k}, X_{2,t_k}, Y_{t_k}^{(0,0)}, Y_{t_k}^{(1,0)}, Y_{t_k}^{(0,1)}, \ldots, Y_{t_k}^{(m-1,0)}, \ldots, Y_{t_k}^{(0,m-1)}\big)^T,$$

where

$$\begin{aligned} Y_t^{(0,0)} &= f(X_t), \\ Y_t^{(1,0)} &= \frac{\partial f}{\partial x_1}(X_t), \; Y_t^{(0,1)} = \frac{\partial f}{\partial x_2}(X_t) \\ &\vdots \\ Y_t^{(m-1,0)} &= \frac{\partial^{m-1} f}{\partial x_1^{m-1}}(X_t), \ldots, \; Y_t^{(0,m-1)} = \frac{\partial^{m-1} f}{\partial x_2^{m-1}}(X_t) \end{aligned}$$

with

$$X_t = (X_{1,t}, X_{2,t}).$$

Since we make no assumption about the function form of f, we estimate the values of the function and its partial derivatives from discrete observations, $\{Z_{t_k}\}_{1\leqslant k\leqslant n}$. That is why additional $(M-2)$ states are considered as unobservable states.

The state vector ξ_k satisfies the following linear system and observation equations. See Appendix A 1 for outline of the derivation.

$$\xi_{k+1} = F_k\xi_k + c_k + e_{k+1}, \tag{4}$$

$$Z_{t_k} = H\xi_k + \varepsilon_{t_k}, \tag{5}$$

where

$$F_k = I_M + G_kA_k, \tag{6}$$

$$c_k = G_kb_k, \tag{7}$$

$$e_{k+1} = G_k\epsilon_{k+1}, \tag{8}$$

with the $h \times h$ identity matrix I_h and

$$A_k = \begin{pmatrix} 0 & A_{1,2} & A_{1,3} & 0 & \cdots & 0 \\ 0 & A_{2,2} & A_{2,3} & 0 & \cdots & 0 \\ 0 & & \cdots & & & 0 \\ \vdots & & & & & \vdots \\ 0 & & \cdots & & & 0 \end{pmatrix},$$

$$\begin{pmatrix} A_{1,2} & A_{1,3} \\ A_{2,2} & A_{2,3} \end{pmatrix} = \big(J_{t_k}\big)^{-1}\{\exp(J_{t_k}\Delta t) - I_2\},$$

$$J_{t_k} = \begin{bmatrix} 0 & 1 \\ Y_{t_k|t_k}^{(1,0)} & Y_{t_k|t_k}^{(0,1)} \end{bmatrix},$$

$$G_k = \begin{bmatrix} 1 & 0 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & \cdots & & 0 \\ Y_{t_k|t_k}^{(1,0)} & Y_{t_k|t_k}^{(0,1)} & 1 & \cdots & & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \theta_{m-2} & \theta_{m-1} & 0 & & \ddots & 0 \\ \theta_{m-1} & \theta_m & 0 & & \cdots & 1 \end{bmatrix},$$

$$b_k = \begin{bmatrix} \left(J_{t_k}^{-1}\right)^2\{\exp(J_{t_k}\Delta t) - I_2 - J_{t_k}\Delta t\}M_{t_k} \\ \frac{\sigma^2}{2}Y^{(0,2)}_{t_k|t_k}\Delta t \\ \frac{\sigma^2}{2}Y^{(1,2)}_{t_k|t_k}\Delta t \\ \vdots \\ 0 \end{bmatrix},$$

$$\epsilon_{k+1} = \begin{bmatrix} \int_{t_k}^{t_{k+1}} \exp\{J_{t_k}(t_{k+1}-u)\}\sigma dB_u \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$H = (1 \quad 0 \quad \cdots \quad 0),$$

$$M_{t_k} = \begin{bmatrix} 0 \\ \frac{\sigma^2}{2}Y^{(0,2)}_{t_k|t_k} \end{bmatrix},\ S = \begin{pmatrix} 0 \\ \sigma \end{pmatrix}.$$

Since the above state space model is linear with respect to ξ_k, we can apply the Kalman updating formula like the extended Kalman filtering, the filter and prediction of ξ_k, which are given as $\xi_{k|k} = E[\xi_k|\{Z_{t_j}\}_{1\leqslant j\leqslant k}]$ and $\xi_{k+1|k} = E[\xi_{k+1}|\{Z_{t_j}\}_{1\leqslant j\leqslant k}]$, can be computed as follows:

$$\xi_{k+1|k} = F_k\xi_{k|k} + c_k, \tag{9}$$

$$\Sigma_{k+1|k} = F_k\Sigma_{k|k}F_k^T + Q_k, \tag{10}$$

$$K_k = \Sigma_{k|k-1}H^T\left(H\Sigma_{k|k-1}H^T + \sigma_\varepsilon^2\right)^{-1}, \tag{11}$$

$$\xi_{k|k} = \xi_{k|k-1} + K_k\left(Z_{t_k} - H\xi_{k|k-1}\right), \tag{12}$$

$$\Sigma_{k|k} = (I - K_kH)\Sigma_{k|k-1}, \tag{13}$$

where $\Sigma_{k+1|k}$ and $\Sigma_{k|k}$ are the covariance matrix of prediction and filtering, respectively, and Q_k is the covariance matrix of e_{k+1}. Then the estimates of the unobservable states of our interest can be found in the components of $\xi_{k|k}$ or $\xi_{k+1|k}$. For example, the estimate of the unobservable state X_{2,t_k} is given by the second component of $\xi_{k|k}$, and the predicted values of $X_{1,t_{k+1}}$ and $X_{2,t_{k+1}}$ are given by the first and second components of $\xi_{k+1|k}$, and so on.

Here note that the system has several parameters to be estimated; σ and σ_ε and the nuisance parameters $\theta_0, \theta_1, \ldots, \theta_m$, which correspond to $Y_t^{(m,0)}, Y_t^{(m-1,1)}, \ldots, Y_t^{(0,m)}$. These parameters can be estimated by the quasi-maximum likelihood estimation from $\{Z_{t_k}\}_{1\leqslant k\leqslant n}$ by using the following likelihood function. For the parameter vector θ,

$$L(\theta) = \Pi_{k=1}^{n-1}p(Z_{t_{k+1}}|Z_{t_k};\theta), \tag{14}$$

$$p(Z_{t_{k+1}}|Z_{t_k};\theta) = \left(2\pi\left(H\Sigma_{k+1|k}H^T + \sigma_\varepsilon^2\right)\right)^{-1/2} \times \exp\left\{-\frac{(Z_{t_{k+1}} - H\xi_{k+1|k})^2}{2\left(H\Sigma_{k+1|k}H^T + \sigma_\varepsilon^2\right)}\right\}. \tag{15}$$

Then maximizing $L(\theta)$ with respect to θ, the maximum likelihood estimate $\hat{\theta}$ is obtained.

III. NUMERICAL EXPERIMENT

We conduct numerical experiments to compare estimation accuracy of the NPF method with that of such a parametric method of filtering as discussed in Refs. [25,26], called the PF method here, in which a cubic polynomial function is used as f in (2) but its values of parameters unknown, which are estimated from discrete time series. Once the polynomial function is estimated, its derivatives can be easily computed as well. The formula of the PF method are given in Appendix A 2.

Here consider the following examples of f.

Model	$f(x_1, x_2)$
Linear	$-x_1 - 0.001x_2$
Van der Pol	$-x_1 + 0.75\left(1 - x_1^2\right)x_2$
Cubic	$-0.2\left(x_1^3 + x_2^3\right)$
Unforced Duffing	$x_1 - x_1^3 - 0.5x_2$

In the experiments, we are particularly interested in the estimation of the partial derivatives of f. Because the examples of f considered here are at most cubic polynomial, the polynomial function used for the PF method certainly includes the exact function forms of the models as subsets and thus it may be reasonable that the performance of the PF method is taken as a benchmark for the comparison.

TABLE I. RMSE of each state by each method is presented; the RMSE is computed from one sample path of size 4000.

Method	x_1	x_2	f	$\partial_1 f$	$\partial_2 f$	$\partial_1^2 f$	$\partial_1\partial_2 f$	$\partial_2^2 f$
Linear								
PF	0.0007	0.0399	1.5150	1.9987	1.0074	0.0027	0.0030	0.0205
NPF	0.0008	0.0594	0.7752	1.0541	1.4821	1.9977	1.7913	0.0643
Van der Pol								
PF	0.0008	0.0541	2.3210	2.1741	1.8426	2.4001	2.0215	0.8502
NPF	0.0006	0.0157	0.1732	0.6300	0.6472	0.6917	0.7691	1.0707
Cubic								
PF	0.0006	0.0231	0.5591	0.7635	0.9270	1.6736	0.1587	0.3975
NPF	0.0003	0.0034	0.0187	0.0295	0.0453	0.0722	0.0315	0.1279
Duffing								
PF	0.0004	0.0062	0.0581	0.3139	0.2089	1.2063	1.2110	0.8397
NPF	0.0002	0.0019	0.0176	0.0148	0.0127	0.1321	0.0167	0.0540

Throughout the experiments, Δt, σ and σ_ε are fixed with $\Delta t = 1/100$, $\sigma = \sigma_\varepsilon = 1/1000$ for all the examples. Estimation accuracy is measured by root mean squared errors (RMSE) in estimating the states of X_1, X_2, $f(X_1, X_2)$ and its partial derivatives up to second order. In simulation the discretized process of an SDE is required for generating its discrete observations $\{Z_{t_k}\}_{1\leqslant k\leqslant n}$. Except for the linear model, however, because it is difficult to derive the exact discretized process from a nonlinear SDE, an approximate discretized process is needed. In our experiments the same discretization method as used in the PF method, called the local linearization (LL) method discussed in Ref. [25], is applied for the derivation. Of course there may be an influence of the approximation on evaluation of the estimation accuracy. However, if any, it would favor the performance of the PF method.

The first experiment is the comparison on the basis of one sample path of size 4000. As for the NPF method, its order of expansion must be determined before the comparison. Here the expansion of up to sixth order is considered and then we select the optimal order on the basis of the maximized log-likelihood, showing that the sixth order is optimal for all examples. The result of RMSEs are presented in Table I and the true curves (in red line) and their estimated (in blue line) are displayed in Figs. 1–8. Looking at the result of the linear model, Table I shows that the PF method produces slightly more accurate estimates than the NPF method. In the case of the linear model the second-order derivatives are all zero. However, the curves estimated by the both methods in Figs. 1 and 2 display periodic waves. Anyway, we can see that the two methods show comparable performance. On the other hand, the results of the nonlinear models in Table I show that the estimation by the NPF method is much more accurate than by the PF method. This is also confirmed graphically. Looking at the difference between Figs. 3 and 4, the estimates by the NPF method look more accurate than those estimates by the PF method, especially for the second-order derivatives. This can be seen in the estimation of the Cubic model and the Duffing model as well.

These results are obtained certainly from the experiment of only one sample path. Hence it may be questionable if these results might be a special outcome from a special sample path. To exclude the possibility of particular choice of sample paths, the same experiment is carried out 100 times and then the average of RMSEs are computed, which are given by Table II. In repeated experiments, the values of parameters and the order of expansion of the NPF method are fixed at the same as the first experiment. The results show that the averaged RMSEs of the both methods are almost the same as those of one sample path, suggesting that the estimation accuracy is not so affected by choice of sample paths.

Finally, we conduct an experiment to see the influence of sample size on the performance of estimation referring to Ref. [27]. Here consider four different sample sizes $N = 500$, 1000, 2000, and 4000. For each sample size N, we proceed the same experiment as the above of computing the averaged RMSEs. For simple evaluation, the averaged norm of the RMSE vectors $\{\boldsymbol{R}_i\}_{1\leqslant i\leqslant 100}$ is used, each of which is defined by $\boldsymbol{R}_i = (\mathrm{RMSE}(x_1), \ldots, \mathrm{RMSE}(\partial_2^2 f))$ for each ith experiment. Then, the averaged norm is defined by the average of

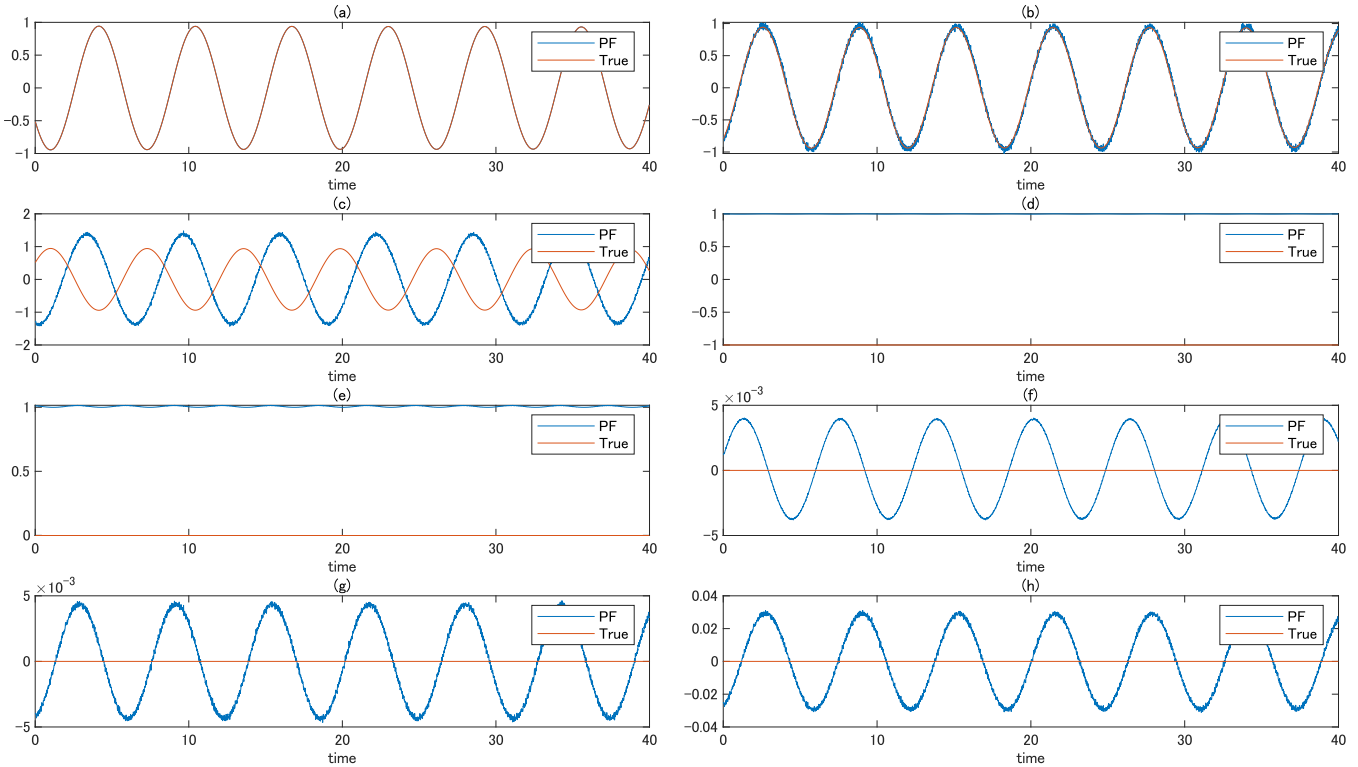

FIG. 1. State estimation of the Linear model by the PF method: Panels (a)–(h) show the estimated states of $x_1, x_2, f, \partial_1 f, \partial_2 f, \partial_1^2 f, \partial_1\partial_2 f, \partial_2^2 f$ in this order. The curves in red display the true and the curves in blue the estimates.

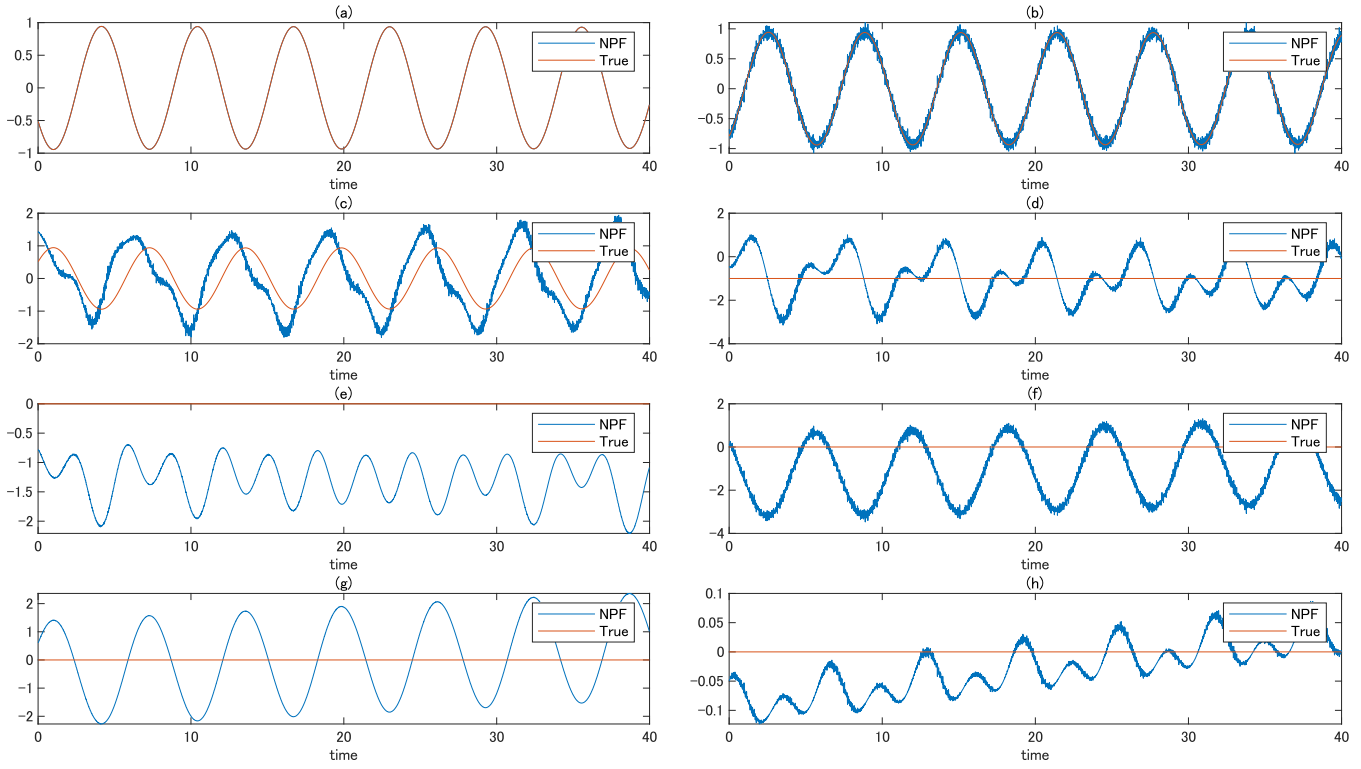

FIG. 2. State estimation of the Linear model by the NPF method: Panels (a)–(h) show the estimated states of $x_1, x_2, f, \partial_1 f, \partial_2 f, \partial_1^2 f, \partial_1 \partial_2 f, \partial_2^2 f$ in this order. The curves in red display the true and the curves in blue the estimates.

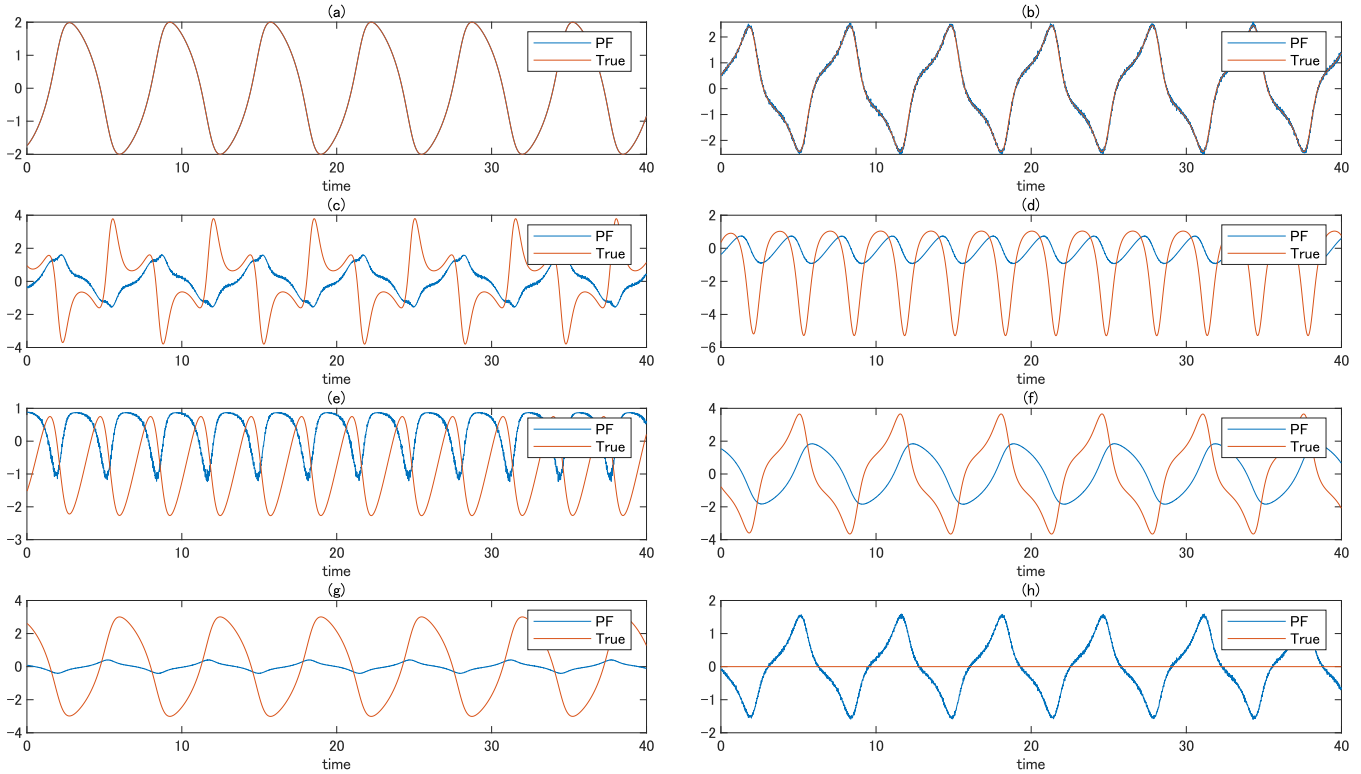

FIG. 3. State estimation of Van der Pol model by the PF method: Panels (a) to (h) show the estimated states of $x_1, x_2, f, \partial_1 f, \partial_2 f, \partial_1^2 f, \partial_1 \partial_2 f, \partial_2^2 f$ in this order. The curves in red display the true and the curves in blue the estimates.

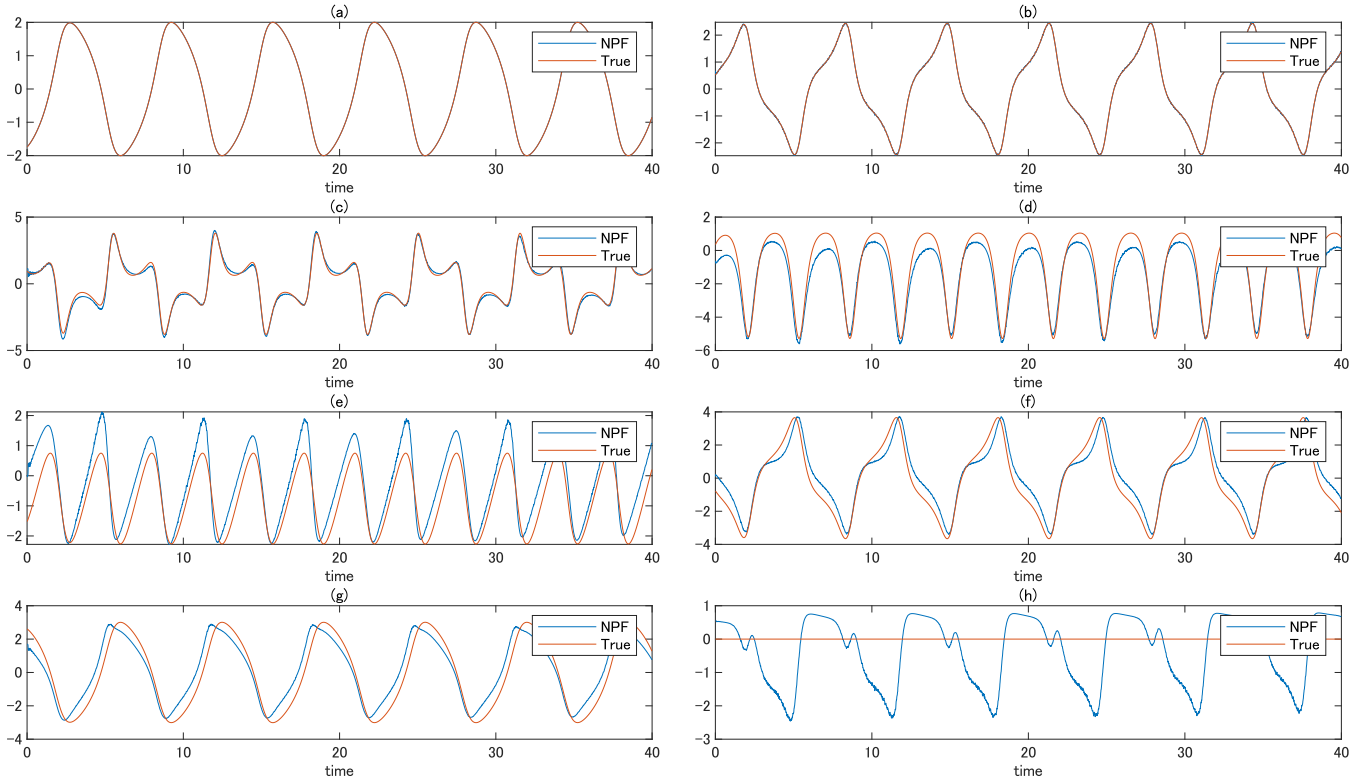

FIG. 4. State estimation of Van der Pol model by the NPF method: Panels (a)–(h) show the estimated states of $x_1, x_2, f, \partial_1 f, \partial_2 f, \partial_1^2 f, \partial_1\partial_2 f, \partial_2^2 f$ in this order. The curves in red display the true and the curves in blue the estimates.

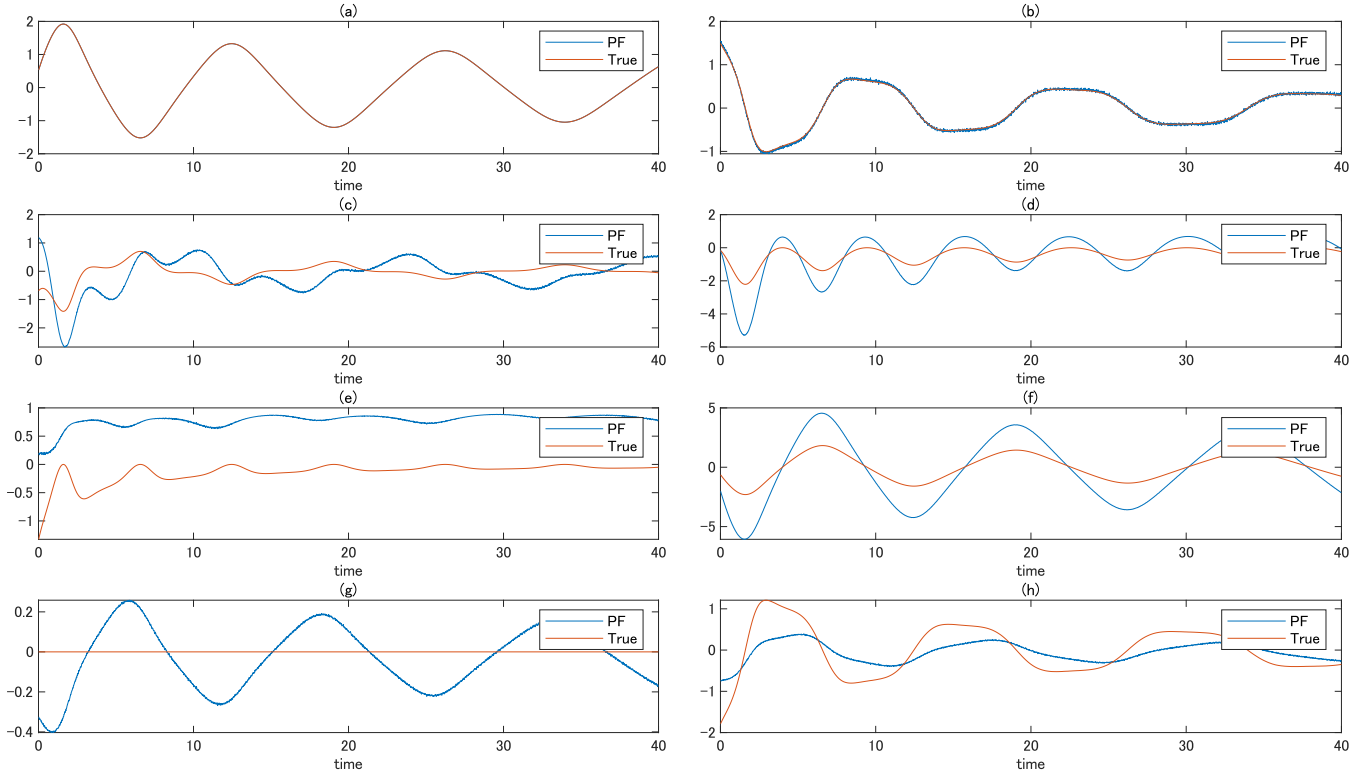

FIG. 5. State estimation of the cubic model by the PF method: Panels (a)–(h) show the estimated states of $x_1, x_2, f, \partial_1 f, \partial_2 f, \partial_1^2 f, \partial_1\partial_2 f, \partial_2^2 f$ in this order. The curves in red display the true and the curves in blue the estimates.

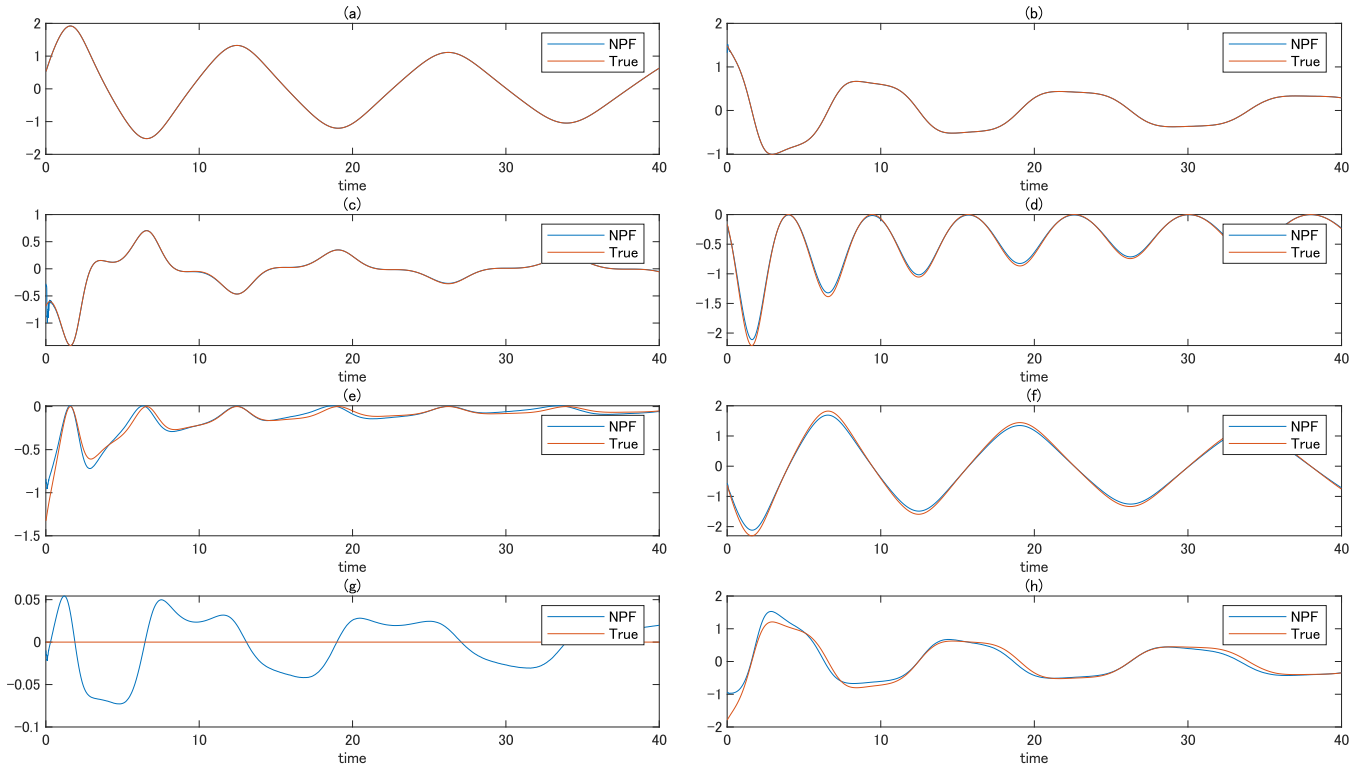

FIG. 6. State estimation of the cubic model by the NPF method: Panels (a)–(h) show the estimated states of $x_1, x_2, f, \partial_1 f, \partial_2 f, \partial_1^2 f, \partial_1 \partial_2 f, \partial_2^2 f$ in this order. The curves in red display the true and the curves in blue the estimates.

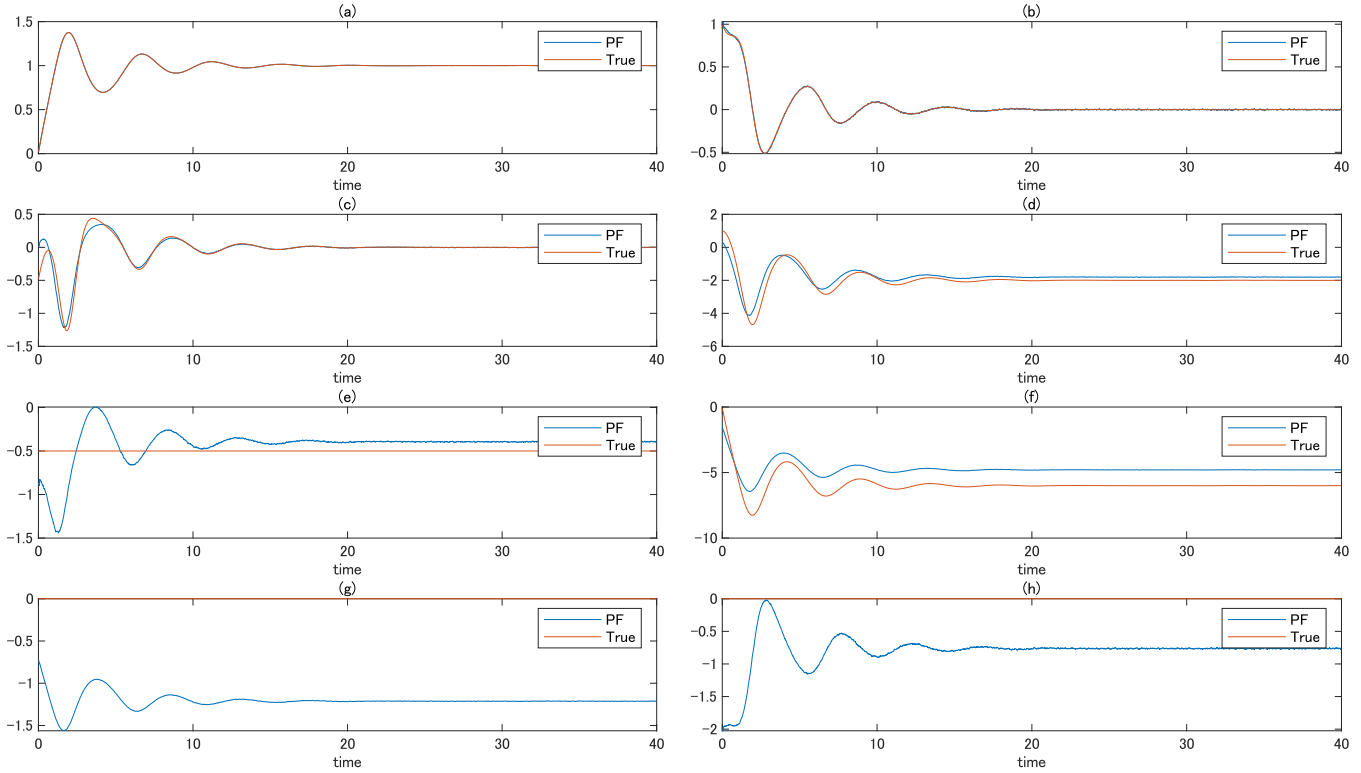

FIG. 7. State estimation of unforced Duffing model by the PF method: Panels (a)–(h) show the estimated states of $x_1, x_2, f, \partial_1 f, \partial_2 f, \partial_1^2 f, \partial_1 \partial_2 f, \partial_2^2 f$ in this order. The curves in red display the true and the curves in blue the estimates.

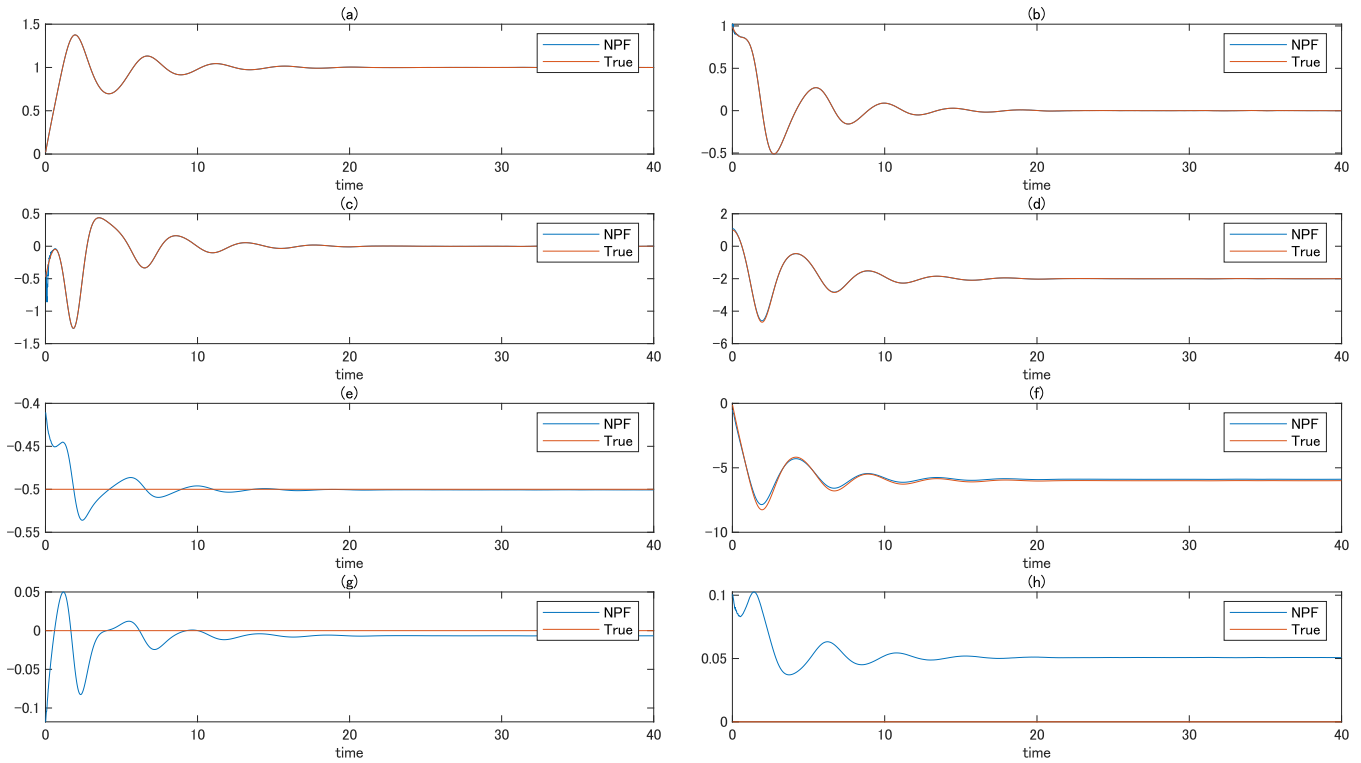

FIG. 8. State estimation of unforced Duffing model by the NPF method: Panels (a)–(h) show the estimated states of $x_1, x_2, f, \partial_1 f, \partial_2 f, \partial_1^2 f, \partial_1\partial_2 f, \partial_2^2 f$ in this order. The curves in red display the true and the curves in blue the estimates.

$\{|\boldsymbol{R}_i|\}_{1\leqslant i\leqslant 100}$, or $\sum_{i=1}^{100}|\boldsymbol{R}_i|/100$. The results are presented in Table III. Similarly to the previous experiment, the estimation performance of the NPF method is comparable to that of the PF method for the linear model, but the NPF method shows considerably better performance than the PF method for the nonlinear models. However, the difference of the performance between the two methods is somewhat affected by the sample size.

Hence, while showing a comparable performance with the PF method for the linear model, the NPF method can produce considerably more accurate estimates than the PF method, especially in the estimation of derivatives of the nonlinear models.

IV. CONCLUSION

This paper proposes a nonparametric method of filtering for stochastic oscillations modeled by two-variate stochastic differential equation, where one variable is observable but not the other. Under this setting, we have particular interest in the estimation of derivatives of the drift function of the stochastic differential equation without specifying its function form.

In addition to the two variables of the stochastic differential equation, making the values of the drift function and its derivatives the additional components of a state vector, we construct a state space model with the updating formula of the Kalman filtering, and thereby the unobservable values of

TABLE II. The average of the RMSE of each state by each method is presented. The average is computed from 100 experiments.

Method	x_1	x_2	f	$\partial_1 f$	$\partial_2 f$	$\partial_1^2 f$	$\partial_1\partial_2 f$	$\partial_2^2 f$
Linear								
PF	0.0007	0.0402	1.5109	1.9987	1.0074	0.0027	0.0030	0.0204
NPF	0.0008	0.0606	0.7719	1.0662	1.4754	2.0101	1.7766	0.0645
Van der Pol								
PF	0.0008	0.0549	2.3221	2.1790	1.7767	2.3046	1.9628	0.8359
NPF	0.0007	0.0260	0.6426	0.6138	0.8409	0.7200	0.7018	0.9909
Cubic								
PF	0.0008	0.0496	0.6728	0.7679	0.9400	1.6781	0.1608	0.4046
NPF	0.0006	0.0081	0.0252	0.0414	0.0670	0.0720	0.0325	0.1402
Duffing								
PF	0.0004	0.0068	0.0585	0.3140	0.2091	1.2063	1.2110	0.8396
NPF	0.0002	0.0027	0.0146	0.0193	0.0127	0.1326	0.0169	0.0546

TABLE III. The average norm of the RMSE vector is computed from 100 experiments, where the RMSE vector $\boldsymbol{R}_i$ of ith experiment is defined by $\boldsymbol{R}_i = (\mathrm{RMSE}(x_1), \ldots, \mathrm{RMSE}(\partial_2^2 f))$. The averaged norm is defined by the average of $\{|\boldsymbol{R}_i|\}_{1 \leqslant i \leqslant 100}$.

	Linear		Van der Pol		Cubic		Duffing	
N	NPF	PF	NPF	PF	NPF	PF	NPF	PF
500	5.1575	4.2999	3.4682	4.5616	2.0373	2.3177	0.9829	1.4559
1000	1.8701	2.7247	2.1036	4.8127	0.5179	2.2321	0.2175	1.2794
2000	1.8555	2.7214	1.8807	5.8269	0.2296	1.9858	0.2332	0.9245
4000	3.3338	2.7008	1.8932	6.1614	0.1880	1.6567	0.1471	0.7726

the function and its derivatives can be estimated from discrete time series of the observable variable.

We conducted numerical experiments to evaluate estimation accuracy of the proposed method by comparing with a parametric method in which the drift function is expressed as a cubic polynomial function containing the true function form as a subset. Four examples of stochastic oscillations were considered; one is a linear model and the others are nonlinear ones. Then, we tried to estimate the values of the derivatives up to second order including the values of the drift function and the unobservable variable.

The results of the comparison on the basis of one sample path showed that the proposed method considerably outperforms the parametric one in the estimation of the nonlinear models, particularly in the estimation of the second-order partial derivatives, while estimation accuracy of the proposed method is comparable with that of the parametric one in the case of the linear model. The additional experiments based on one hundred sample paths revealed almost the same results as on one sample path, suggesting that the estimation by the proposed method is more accurate, particularly for nonlinear models, regardless of sample paths. Moreover, due to the experiment to see the influence of sample size on the estimation accuracy, while the difference is somewhat affected by the sample size, the dominance of the proposed method over the parametric one in the estimation of the nonlinear models does not change.

ACKNOWLEDGMENTS

We thank anonymous referees for their helpful comments and suggestions.

APPENDIX

1. Outline of derivation for NPF method

This Appendix presents an outline of derivation for the NPF method. More detail discussion can be seen in Ref. [24].

First suppose the truncated Taylor's expansion of f around $\boldsymbol{x}_0 = (x_0, y_0)$ up to mth order. Then we get

$$\begin{aligned} f(\boldsymbol{x}) &\approx \tilde{f}(\boldsymbol{x}), \\ &= f(\boldsymbol{x}_0) + \frac{\partial f}{\partial x}(\boldsymbol{x}_0)(x - x_0) + \frac{\partial f}{\partial y}(\boldsymbol{x}_0)(y - y_0) \\ &\quad \vdots \\ &\quad + \frac{1}{m!}\frac{\partial^m f}{\partial x^m}(\boldsymbol{x}_0)(x - x_0)^m + \cdots + \frac{1}{m!}\frac{\partial^m f}{\partial y^m}(\boldsymbol{x}_0)(y - y_0)^m. \end{aligned}$$

Let $X_t = (X_{1,t}, X_{2,t})^T$ and define $Y_t = \tilde{f}(X_t)$ by putting X_s into $\boldsymbol{x}_0$ where $s < t$. Because $\tilde{f}$ is a polynomial function of degree m, repeated application of the Ito's formula over time interval $u \in [s, t)$ gives

$$\begin{aligned} Y_t^{(0,0)} - Y_s^{(0,0)} &= \int_s^t Y_u^{(1,0)} dX_{1,u} + \int_s^t Y_u^{(0,1)} dX_{2,u} + \frac{\sigma^2}{2}\int_s^t Y_u^{(0,2)} du, \\ Y_t^{(1,0)} - Y_s^{(1,0)} &= \int_s^t Y_u^{(2,0)} dX_{1,u} + \int_s^t Y_u^{(1,1)} dX_{2,u} + \frac{\sigma^2}{2}\int_s^t Y_u^{(1,2)} du, \\ Y_t^{(0,1)} - Y_s^{(0,1)} &= \int_s^t Y_u^{(1,1)} dX_{1,u} + \int_s^t Y_u^{(0,2)} dX_{2,u} + \frac{\sigma^2}{2}\int_s^t Y_u^{(0,3)} du, \\ &\vdots \\ Y_t^{(0,m)} - Y_s^{(0,m)} &= 0. \end{aligned}$$

where

$$Y_t^{(i,j)} = \frac{\partial^{i+j}\tilde{f}}{\partial x^i \partial y^j}(X_t).$$

To discretize the process at discrete times $\{t_k\}_{1\leqslant k\leqslant n}$ $(\Delta t = t_k - t_{k-1})$, we use a usual assumption that each integrand is a constant over the time interval $[t_k, t_{k+1})$. However, because of no knowledge about a function form of f, we must estimate $Y_t^{(i,j)}$ from discrete observation $\{Z_{t_k}\}_{1\leqslant k\leqslant n}$. To this end we replace $Y_{t_k}^{(i,j)}$ with $Y_{t_k|t_k}^{(i,j)} = E[Y_{t_k}^{(i,j)}|\{Z_{t_l}\}_{1\leqslant l\leqslant k}]$. This replacement is used for the formulation of the extended Kalman filter algorithm; see Refs. [28,29]. Furthermore the last $(m+1)$ equalities imply that $Y_{t_k}^{(i,j)}(i+j=m)$ are constant. Denoting them by θ_j $(0\leqslant j\leqslant m)$, we estimate θ_j as nuisance parameters from data. Under the above setting, the following linear discrete time model is obtained:

$$\begin{aligned}
Y_{t_{k+1}}^{(0,0)} - Y_{t_k}^{(0,0)} &= Y_{t_k|t_k}^{(1,0)}\big(X_{1,t_{k+1}} - X_{1,t_k}\big) + Y_{t_k|t_k}^{(0,1)}\big(X_{2,t_{k+1}} - X_{2,t_k}\big) + \frac{\sigma^2}{2}Y_{t_k|t_k}^{(0,2)}\Delta t,\\
Y_{t_{k+1}}^{(1,0)} - Y_{t_k}^{(1,0)} &= Y_{t_k|t_k}^{(2,0)}\big(X_{1,t_{k+1}} - X_{1,t_k}\big) + Y_{t_k|t_k}^{(1,1)}\big(X_{2,t_{k+1}} - X_{2,t_k}\big) + \frac{\sigma^2}{2}Y_{t_k|t_k}^{(1,2)}\Delta t,\\
Y_{t_{k+1}}^{(0,1)} - Y_{t_k}^{(0,1)} &= Y_{t_k|t_k}^{(1,1)}\big(X_{1,t_{k+1}} - X_{1,t_k}\big) + Y_{t_k|t_k}^{(0,2)}\big(X_{2,t_{k+1}} - X_{2,t_k}\big) + \frac{\sigma^2}{2}Y_{t_k|t_k}^{(0,3)}\Delta t,\\
&\vdots\\
Y_t^{(0,m-1)} - Y_s^{(0,m-1)} &= \theta_{m-1}(X_{1,t_{k+1}} - X_{1,t_k}) + \theta_m(X_{2,t_{k+1}} - X_{2,t_k}).
\end{aligned}$$

Meanwhile, let the SDE be expressed as

$$dX_t = \mu(X_t)dt + SdB_t,$$

where $\mu(x) = (x_2, f(x))^T$ with $x = (x_1, x_2)$. Here consider the local linearization (LL) method in which the drift function of the SDE is approximated by the second-order Taylor's expansion around each filtered state; its detail explanation can be seen in Refs. [25,26,30,31]. Applying the LL method,

$$\begin{aligned}
X_{t_{k+1}} = X_{t_k} &+ D\mu\big(X_{t_k}\big)^{-1}\big[\exp(D\mu(X_{t_k})\Delta t) - I_2\big]\mu\big(X_{t_k}\big) + \big(D\mu(X_{t_k})^{-1}\big)^2\big\{\exp\big(J_{t_k}\Delta t\big) - I_2 - J_{t_k}\Delta t\big\}M_{t_k}^o\\
&+ \int_{t_k}^{t_{k+1}} \exp\big\{D\mu\big(X_{t_k}\big)(t_{k+1} - u)\big\}SdB_u,
\end{aligned}$$

where $D\mu$ is a Jacobi matrix of μ. Approximate $D\mu(X_{t_k}) \approx D\mu(X_{t_k|t_k})$ and $f \approx \tilde{f}$, then

$$D\mu(X_{t_k}) \approx \begin{bmatrix} 0 & 1\\ Y_{t_k|t_k}^{(1,0)} & Y_{t_k|t_k}^{(1,0)}\end{bmatrix} = J_{t_k}$$

Similarly,

$$\begin{aligned}
\mu(X_{t_k}) &= \big(X_{2,t_k}, f(X_{t_k})\big)^T\\
&\approx \big(X_{2,t_k}, \tilde{f}(X_{t_k})\big)^T = \big(X_{2,t_k}, Y_{t_k}^{(0,0)}\big),^T
\end{aligned}$$

and

$$\begin{aligned}
M_{t_k}^o &= \big(0, \big(\sigma^2/2\big)\partial_2^2 f\big(X_{t_k}\big)\big)^T\\
&\approx \big(0, \big(\sigma^2/2\big)Y_{t_k|t_k}^{(0,2)}\big)\big)^T = M_{t_k}.
\end{aligned}$$

We get

$$X_{t_{k+1}} = X_{t_k} + J_{t_k}^{-1}\big\{\exp\big(J_{t_k}\Delta t\big) - I_2\big\}\big(X_{2,t_k}, Y_{t_k}^{(0,0)}\big)^T + \big(J_{t_k}^{-1}\big)^2\big\{\exp\big(J_{t_k}\Delta t\big) - I_2 - J_{t_k}\Delta t\big\}M_{t_k} + \int_{t_k}^{t_{k+1}} \exp\big\{J_{t_k}(t_{k+1} - u)\big\}SdB_u.$$

In this equation and the equations of $Y_{t_{k+1}}^{(i,j)} - Y_{t_k}^{(i,j)}$, putting the processes at time t_{k+1} in the left-hand side and those at time t_k in the right-hand side, we get

$$G_k^{-1}\xi_{k+1} = \big(G_k^{-1} + A_k\big)\xi_k + b_k + \epsilon_{k+1}.$$

Multiplying both hand sides by G_k, the formula of the NPF method is obtained.

2. PF method

The PF method is a continuous-discrete filtering method derived from a discretization of an SDE by applying the LL method. The extended Kalman filter (EKF) is a special version of the PF method because the first-order Taylor's expansion is used in EKF. For the sampling time interval Δt, denote the discrete state by $\xi_k = X_{t_k} = (X_{1,t_k}, X_{2,t_k})^T$ at discrete time t_k. After applying the LL method to the SDE where its drift function f is a two-variate cubic polynomial one, the discrete state space model is

given by

$$\xi_{k+1} = F_k \xi_k + c_k + e_{k+1},$$
$$Z_{t_k} = H\xi_k + \varepsilon_{t_k},$$

where

$$F_k = \exp(J_{t_k}\Delta t),$$
$$c_k = \left[I_2 - \exp\left(J_{t_k}\Delta t\right)\right]\xi_{k|k} + J_{t_k}^{-1}\left[\exp\left(J_{t_k}\Delta t\right) - I_2\right]\mu_{t_k} + \left(J_{t_k}^{-1}\right)^2\left[\exp\left(J_{t_k}\Delta t\right) - I_2 - J_{t_k}\Delta t\right]M_{t_k},$$
$$e_{k+1} = \int_{t_k}^{t_{k+1}} \exp\left[J_{t_k}(t_{k+1} - u)\right]SdB_u,$$
$$H = (1\ 0),$$

with

$$\mu(x) = (\mu_1(x), \mu_2(x))^T,$$
$$\mu_1(x) = x_2,$$
$$\mu_2(x) = a_0 + a_{1,0}x_1 + a_{0,1}x_2 + \cdots + a_{3,0}x_1^3 + \cdots + a_{0,3}x_2^3,$$
$$x = (x_1, x_2)^T,$$
$$J_t = D\mu(X_{t|t}),\ \ \mu_t = \mu(X_{t|t}),$$
$$M_t = \left(\frac{1}{2}\mathrm{tr}(SS^T H_{1,t}), \frac{1}{2}\mathrm{tr}(SS^T H_{2,t})\right)^T,$$
$$H_{i,t} = \left(\frac{\partial^2 \mu_i}{\partial x_j \partial x_k}\right)_{1\leqslant j,k\leqslant 2}(X_{t|t}),$$
$$S = (0\ \sigma)^T.$$

In addition, the covariance matrix Q_k of e_k is given by

$$Q_k = \int_0^{\Delta t} \exp\left(J_{t_k}u\right)SS^T \exp\left(J_{t_k}u\right)^T du.$$

The EKF is given by simply putting $M_t = 0$.

[1] M. Havlicek, K. J. Friston, J. Jan, M.Brazdil, and V. D. Calhoun, Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering, NeuroImage **56**, 2109 (2011).

[2] V. N. Livina, F. Kwasniok, G. Lohmann, J. W. Kantelhardt, and T. M. Lenton, Changing climate states and stability: From Pliocene to present, Clim. Dynam. **37**, 2437 (2011).

[3] J. J. Riera, J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, T. Ozaki, and R. Kawashima, A state-space model of the hemodynamic approach: Nonlinear filtering of BOLD signals, NeuroImage **21**, 547 (2004).

[4] K. E. Stephan, L. Kasper, L. M.Harrison, J. Daunizeau, H. E. M. den Ouden, M. Breakspear, and K. J. Friston, Nonlinear dynamic causal models for fMRI, NeuroImage **42**, 649 (2008).

[5] P. A. Valdes-Sosa, J. M. Sanchez-Bornot, R. C. Sotero, Y. Iturria-Medina, Y. Aleman-Gomez, J. Bosch-Bayard, F. Carbonell, and T. Ozaki, Model driven EEG/fMRI fusion of brain oscillations, Hum. Brain Map. **30**, 2701 (2009).

[6] J. H. Gove and D. Y. Hollinger, Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange, J. Geophys. Res. **111**, D08S07 (2006).

[7] F. Kwasniok and G. Lohmann, Deriving dynamical models from paleoclimatic records: Application to glacial millennial-scale climate variability, Phys. Rev. E **80**, 066104 (2009).

[8] R. Misra and A. A. Zdziarski, Damped harmonic oscillator interpretation of the soft-state power spectra of Cyg X-1, Mon. Not. Roy. Astron. Soc. **387**, 915 (2008).

[9] R. A. Phillipson, P. T. Boyd, and A. P. Smale, The chaotic long-term X-ray variability of 4U 1705-44, Mon. Not. Roy. Astron. Soc. **477**, 5220 (2018).

[10] C. Chiarella, H. Hung, and T. D. To, The volatility structure of the fixed income market under the HJM framework: A nonlinear filtering approach, Comput. Stat. Data Anal. **53** (2009).

[11] P. Date and K. Ponomareva, Linear and non-linear filtering in mathematical finance: A review, IMA J. Manage. Math. **22**, 195 (2011).

[12] J. Fan and I. Gijbels, *Local Polynomial Modeling and Its Applications* (Chapman & Hall, London, 1996).

[13] J. Fan and Q. W. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods* (Springer, Berlin, 2005).

[14] J. Fan and C. Zhang, A reexamination of diffusion estimators with applications to financial model validation, J. Am. Stat. Assoc. **98**, 118 (2003).

[15] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Trans. Sign. Process. **50**, 174 (2002).

[16] Y. Che, B. Liu, H. Li, M. Lu, J. Wang, and X. Wei, Robust stabilization control of bifurcations in Hodgkin-Huxley model with aid of unscented Kalman filter, Chaos, Solitons Fract. **101**, 92 (2017).

[17] S. Mariani and A. Ghisi, Unscented Kalman filtering for nonlinear structural dynamics, Nonlin. Dynam. **49**, 131 (2007).

[18] M. Quach, N. Brunel, and F. d'Alche-Buc, Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference, Bioinformatics **23**, 3209 (2007).

[19] A. Sitz, U. Schwarz, J. Kurths, and H. U. Voss, Estimation of parameters and unobserved components for nonlinear systems from noisy time series, Phys. Rev. E **66**, 016210 (2002).

[20] Z. Chen, Bayesian filtering: From Kalman filters to particle filters, and beyond, Statistics **182**, 1 (2003).

[21] G. Kitagawa, Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, J. Comput. Graph. Stat. **5**, 1 (1996).

[22] A. Doucet, N. D. Freitas, and N. Gordon, *An Introduction to Sequential Monte Carlo Methods* (Springer, New York, 2001).

[23] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proc. F Radar Sign. Process. **140**, 107 (1993).

[24] I. Shoji, A nonparametric method of estimating nonlinear dynamical system models, Phys. Lett. A **277**, 159 (2000).

[25] T. Ozaki, *Time Series Modeling of Neuroscience Data* (CRC Press, Boca Raton, FL, 2012).

[26] I. Shoji, A comparative study of maximum likelihood estimators for nonlinear dynamical system models, Int. J. Contr. **71**, 391 (1998).

[27] J. Miskiewicz, Improving quality of sample entropy estimation for continuous distribution probability functions, Physica A **450**, 473 (2016).

[28] B. D. O. Anderson and J. B. Moor, *Optimal Filtering* (Prentice-Hall, Upper Saddle River, NJ, 1979).

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

[30] J. C. Jimenez, I. Shoji, and T. Ozaki, Simulation of stochastic differential equations through the local linearization method: A comparative study, J. Stat. Phys. **94**, 587 (1999).

[31] E. Platen and N. Bruti-Liberati, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance* (Springer, Berlin, 2010).