Separable Least-squares Approach for Gaussian Process Model Identification Using Firefly Algorithm

Tomohiro Hachino, Hitoshi Takata, Shigeru Nakayama, Seiji Fukushima, and Yasutaka Igarashi

Abstract**—**This paper presents a separable least-squares (LS) approach combining the linear LS method with firefly algorithm (FA) to train a Gaussian process (GP) prior model for nonlinear system identification. The hyperparameter vector of the GP prior covariance is searched for by FA, while the system parameter vector of the objective system and the weighting parameter vector of the GP prior mean are estimated by the linear LS method. The nonlinear function of the objective system is estimated as the predictive mean function of the GP, and the confidence measure of the estimated nonlinear function is evaluated by the predictive covariance of the GP. The proposed identification method is applied to modeling of a simplified electric power system on numerical simulation.

Keywords**—**Firefly algorithm, Gaussian process model, identification, nonlinear system, separable least-squares approach.

I. Introduction

SINCE most practical systems are continuous-time nonlinear systems, the development of accurate identification algorithm of such systems is a key problem for precise analysis or control design. For this problem, many parametric identification methods have been exploited using neural network model [1], orthogonal least-squares (LS) estimator [2], radial basis function model [3], [4], and so forth. However, these models need many weighting parameters to describe the nonlinearity of the objective systems. Moreover, any confidence measures of the estimated nonlinear functions are not obtained in such approaches.

In recent years, the Gaussian process (GP) model has received much attention for nonlinear system identification [5]–[7]. The GP model was originally utilized for the regression problem by O' Hagan [8] and has been utilized for regression or classification problem [9]–[11]. The GP model is a nonparametric model and fits naturally into Bayesian framework. Since it has fewer parameters than parametric models such as the neural network model, we can describe the nonlinearity of the objective system in a few parameters. Moreover, the GP gives us not only the mean function but also the covariance function. In this paper, we deal with a nonparametric identification of continuous-time nonlinear systems using the GP model.

The GP prior model derived by applying the delayed state variable filter has to be appropriately trained based on the identification data. Generally this training becomes nonlinear optimization problem. In this paper, the separable LS approach combining the linear LS method with firefly algorithm (FA) is proposed to train the GP prior model. The hyperparameter vector of the GP prior covariance is searched for by FA, while the system parameter vector of the objective system and the weighting parameter vector of the GP prior mean are estimated by the linear LS method. FA is an optimization technique inspired by an intelligent behavior of firefly swarms [12]. In FA, for any two flashing fireflies, the less brighter firefly moves toward the brighter one according to the attractiveness. The attractiveness is proportional to the light intensity observed by the partner and monotonically decreases as the distance between two fireflies increases, owing to the inverse square law and the absorption property of light. FA consists of only the basic arithmetic operations and does not require complicated coding and genetic operations such as crossovers and mutations of genetic algorithm (GA). Moreover, the performance and computational cost of FA are shown to be better than those of other population-based algorithms such as GA and particle swarm optimization (PSO) [12], [13]. These advantages suggest that the use of FA increases efficiency when the GP prior model for identification is trained.

This paper is organized as follows. In section II, the problem is formulated. In section III, the GP prior model for the identification is derived. In section IV, the separable LS approach using FA is presented for training the GP prior model. In section V, the nonlinear function with the confidence measure is estimated from the GP posterior distribution. In section VI, numerical simulation for a simplified electric power system is carried out to illustrate the effectiveness of the proposed identification method. Finally some conclusions are given in

Tomohiro Hachino is with the Department of Electrical and Electronics Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (phone and fax: +81-99-285-8392; e-mail: hachino@eee.kagoshima-u.ac.jp).

Hitoshi Takata is with the Department of Electrical and Electronics Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (e-mail: takata@eee.kagoshima-u.ac.jp).

Shigeru Nakayama is with the Department of Information Science and Biomedical Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (e-mail: shignaka@ibe.kagoshima-u.ac.jp).

Seiji Fukushima is with the Department of Electrical and Electronics Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (e-mail: fukushima@eee.kagoshima-u.ac.jp).

Yasutaka Igarashi is with the Department of Electrical and Electronics Engineering, Kagoshima University, Kagoshima, 890-0065 Japan (e-mail: igarashi@eee.kagoshima-u.ac.jp).

section VII.

II. Statement of the Problem

Consider a single-input, single-output, continuous-time nonlinear system described by

$$\sum_{\substack{i=0\\ i\neq n_1,n_2,\cdots,n_\alpha}}^{n} a_i p^{n-i}x(t) = f(\boldsymbol{z}(t)) + \sum_{\substack{j=0\\ j\neq m_1,m_2,\cdots,m_\beta}}^{m} b_j p^{m-j}u(t) \quad (1)$$

$$(a_0 = 1, \quad n \geq m)$$

$$\boldsymbol{z}(t) = \big[p^{n-n_1}x(t), p^{n-n_2}x(t), \cdots, p^{n-n_\alpha}x(t),\\ p^{m-m_1}u(t), p^{m-m_2}u(t), \cdots, p^{m-m_\beta}u(t)\big]^{\mathrm{T}}$$

$$y(t) = x(t) + e(t)$$

where $u(t)$ and $x(t)$ are the true input and output signals, respectively. $y(t)$ is the noisy output that is corrupted by the measurement noise $e(t)$. $f(\cdot)$ is an unknown nonlinear function, which is assumed to be stationary and smooth. p denotes the differential operator. n, n_i $(i = 1, 2, \cdots, \alpha)$, m and m_j $(j = 1, 2, \cdots, \beta)$ are assumed to be known. The purpose of this paper is to identify the parameters $\{a_i\}$ and $\{b_j\}$ of the linear terms and the nonlinear function $f(\cdot)$ with the confidence measure, from the true input and noisy output data in the GP framework.

III. GP Prior model for Identification

Equation (1) can be rewritten as

$$p^n y(t) = f(\boldsymbol{w}(t)) - \sum_{\substack{i=1\\ i\neq n_1,n_2,\cdots,n_\alpha}}^{n} a_i p^{n-i}y(t) + \sum_{\substack{j=0\\ j\neq m_1,m_2,\cdots,m_\beta}}^{m} b_j p^{m-j}u(t) + \varepsilon(t)$$

$$\boldsymbol{w}(t) = [p^{n-n_1}y(t), p^{n-n_2}y(t), \cdots, p^{n-n_\alpha}y(t),\\ p^{m-m_1}u(t), p^{m-m_2}u(t), \cdots, p^{m-m_\beta}u(t)]^{\mathrm{T}} \quad (2)$$

where $\varepsilon(t)$ is an error caused by the measurement noise $e(t)$.

Multiplying both sides of (2) by the delayed state variable filter $F(p)$ [2] yields

$$p^n y^f(t) = f(\boldsymbol{w}^f(t)) - \sum_{\substack{i=1\\ i\neq n_1,n_2,\cdots,n_\alpha}}^{n} a_i p^{n-i}y^f(t) + \sum_{\substack{j=0\\ j\neq m_1,m_2,\cdots,m_\beta}}^{m} b_j p^{m-j}u^f(t) + \varepsilon^f(t) \quad (3)$$

where $u^f(t) = F(p)u(t)$, $y^f(t) = F(p)y(t)$, and $\boldsymbol{w}^f(t) = F(p)\boldsymbol{w}(t)$ are the filtered signals, and $\varepsilon^f(t)$ is assumed to be zero mean Gaussian noise with variance σ_n^2.

Putting $t = t_1, t_2, \cdots, t_N$ into (3) yields

$$\boldsymbol{y} = \boldsymbol{v} + \boldsymbol{G}\boldsymbol{\theta}_l \quad (4)$$

where

$$\boldsymbol{y} = [p^n y^f(t_1), p^n y^f(t_2), \cdots, p^n y^f(t_N)]^{\mathrm{T}}$$
$$\boldsymbol{v} = [f(\boldsymbol{w}^f(t_1)) + \varepsilon^f(t_1), f(\boldsymbol{w}^f(t_2)) + \varepsilon^f(t_2),\\ \cdots, f(\boldsymbol{w}^f(t_N)) + \varepsilon^f(t_N)]^{\mathrm{T}}$$
$$\boldsymbol{\theta}_l = [a_1, \cdots, a_i, \cdots, a_n, b_0, \cdots, b_j, \cdots, b_m]^{\mathrm{T}} \quad (5)$$
$$\boldsymbol{G} = [\boldsymbol{g}(t_1), \boldsymbol{g}(t_2), \cdots, \boldsymbol{g}(t_N)]^{\mathrm{T}}$$
$$\boldsymbol{g}(t) = [-p^{n-1}y^f(t), \cdots, -p^{n-i}y^f(t), \cdots, -y^f(t),\\ p^m u^f(t), \cdots, p^{m-j}u^f(t), \cdots, u^f(t)]^{\mathrm{T}}$$

A GP is a Gaussian random function and is completely described by its mean function and covariance function. We can regard it as a collection of random variables with a joint multivariable Gaussian distribution. Therefore, the function values $\boldsymbol{f}$ can be represented by the GP:

$$\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{w}), \boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w})) \quad (6)$$

where

$$\boldsymbol{f} = [f(\boldsymbol{w}^f(t_1)), f(\boldsymbol{w}^f(t_2)), \cdots, f(\boldsymbol{w}^f(t_N))]^{\mathrm{T}}$$
$$\boldsymbol{w} = [\boldsymbol{w}^f(t_1), \boldsymbol{w}^f(t_2), \cdots, \boldsymbol{w}^f(t_N)] \quad (7)$$

$\boldsymbol{w}$ is the input of the function $\boldsymbol{f}$, $\boldsymbol{m}(\boldsymbol{w})$ is the mean function vector, and $\boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w})$ is the covariance matrix. The mean function is often represented by a polynomial regression [11]. In this paper, the mean function is expressed by the first order polynomial, i.e., a linear combination of the input variable:

$$m(\boldsymbol{w}^f(t)) = (\boldsymbol{w}^f(t))^{\mathrm{T}}\boldsymbol{\theta}_m$$
$$\boldsymbol{\theta}_m = \big[\theta_{n_1}, \theta_{n_2}, \cdots, \theta_{n_\alpha}, \theta_{m_1}, \theta_{m_2}, \cdots, \theta_{m_\beta}\big]^{\mathrm{T}} \quad (8)$$

where $\boldsymbol{\theta}_m$ is the unknown parameter vector for the mean function. Thus, the mean function vector $\boldsymbol{m}(\boldsymbol{w})$ is described as follows:

$$\boldsymbol{m}(\boldsymbol{w}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\theta}_m \quad (9)$$

The covariance $\Sigma_{pq} = s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q))$ is an element of the covariance matrix $\boldsymbol{\Sigma}$, which is a function of $\boldsymbol{w}^f(t_p)$ and $\boldsymbol{w}^f(t_q)$. Under the assumption that the nonlinear function is stationary and smooth, the following Gaussian kernel is utilized in this paper:

$$\Sigma_{pq} = s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) = \sigma_y^2 \exp\left(-\frac{||\boldsymbol{w}^f(t_p) - \boldsymbol{w}^f(t_q)||^2}{2\ell^2}\right) \quad (10)$$

where $||\cdot||$ denotes the Euclidean norm. Equation (10) means that the covariance of the outputs of the nonlinear function depends only on the distance between the inputs $\boldsymbol{w}^f(t_p)$ and $\boldsymbol{w}^f(t_q)$. A high correlation between the outputs of the nonlinear function occurs for the inputs that are close to each other. The overall variance of the random function can be controlled by σ_y, and the characteristic length scale of the process can be changed by ℓ.

From (6), the vector $\boldsymbol{v}$ of the noisy function values in (4) can be written as

$$\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{w}), \boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w})) \tag{11}$$

where

$$\begin{aligned} \boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w}) &= \boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w}) + \sigma_n^2 \boldsymbol{I}_N \\ \boldsymbol{I}_N &: N \times N \quad \text{identity} \quad \text{matrix} \end{aligned} \tag{12}$$

and $\boldsymbol{\theta}_c = [\sigma_y, \ell, \sigma_n]^{\mathrm{T}}$ is called the *hyperparameter* vector. From (4) and (11), the GP model for the identification is derived as

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{w}) + \boldsymbol{G}\boldsymbol{\theta}_l, \boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w})) \tag{13}$$

In the following, $\boldsymbol{K}(\boldsymbol{w}, \boldsymbol{w})$ is written as $\boldsymbol{K}$ for simplicity.

IV. Separable LS approach by FA

At the first stage of the identification, the GP prior model is trained by optimizing the unknown parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_m^{\mathrm{T}}, \boldsymbol{\theta}_l^{\mathrm{T}}, \boldsymbol{\theta}_c^{\mathrm{T}}]^{\mathrm{T}}$. This training is carried out by maximizing the log marginal likelihood of the identification data:

$$\begin{aligned} J &= \log p(\boldsymbol{y}|\boldsymbol{w}, \boldsymbol{G}, \boldsymbol{\theta}) \\ &= -\frac{1}{2}\log|\boldsymbol{K}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}_{ml})^{\mathrm{T}}\boldsymbol{K}^{-1}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}_{ml}) \\ &\quad -\frac{N}{2}\log(2\pi) \end{aligned} \tag{14}$$

where

$$\begin{aligned} \boldsymbol{Z} &= [\ \boldsymbol{w}^{\mathrm{T}} \ \vdots \ \boldsymbol{G}\ ] \\ \boldsymbol{\theta}_{ml} &= [\boldsymbol{\theta}_m^{\mathrm{T}}, \boldsymbol{\theta}_l^{\mathrm{T}}]^{\mathrm{T}} \end{aligned} \tag{15}$$

Although this problem is a nonlinear optimization one, we can separate the linear optimization part and the nonlinear optimization part. The partial derivative of (14) with respect to the parameter vector $\boldsymbol{\theta}_{ml}$ is as follows:

$$\frac{\partial J}{\partial \boldsymbol{\theta}_{ml}} = \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{y} - \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{Z}\boldsymbol{\theta}_{ml} \tag{16}$$

Note that if the candidates of the hyperparameter vector $\boldsymbol{\theta}_c$ of the covariance function are given, the candidates of the covariance matrix $\boldsymbol{K}$ can be constructed. Therefore, the parameter vector $\boldsymbol{\theta}_{ml}$ can be estimated by the linear LS method from (16):

$$\boldsymbol{\theta}_{ml} = (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{y} \tag{17}$$

However, even if the parameter vector $\boldsymbol{\theta}_{ml}$ is known, the optimization with respect to $\boldsymbol{\theta}_c$ is a complicated nonlinear problem and might suffer from the local optima problem. Therefore, in this paper, we propose the separable LS approach combining the linear LS method with FA. Only $\boldsymbol{X} = \boldsymbol{\theta}_c = [\sigma_y, \ell, \sigma_n]^{\mathrm{T}}$ is represented with the positions of fireflies in the search space and is searched for by FA. The detailed training algorithm is as follows:

step 1: Initialization

Generate an initial population of Q fireflies with random positions $\boldsymbol{X}_{[i]}$ $(i = 1, 2, \cdots, Q)$.

Set the iteration counter l to 0.

step 2: Construction of the covariance matrix

Construct Q candidates of the covariance matrix $\boldsymbol{K}_{[i]}$ using $\boldsymbol{X}_{[i]}$ $(i = 1, 2, \cdots, Q)$.

step 3: Estimation of $\boldsymbol{\theta}_{ml}$

Estimate Q candidates of $\boldsymbol{\theta}_{ml[i]}$ $(i = 1, 2, \cdots, Q)$:

$$\boldsymbol{\theta}_{ml[i]} = (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}_{[i]}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{K}_{[i]}^{-1}\boldsymbol{y} \tag{18}$$

step 4: Light intensity calculation

Calculate the light intensity I_i of each firefly from the log marginal likelihood of the identification data:

$$\begin{aligned} I_i(\boldsymbol{X}_{[i]}) &= -\frac{1}{2}\log|\boldsymbol{K}_{[i]}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}_{ml[i]})^{\mathrm{T}}\boldsymbol{K}_{[i]}^{-1} \\ &\quad \times(\boldsymbol{y} - \boldsymbol{Z}\boldsymbol{\theta}_{ml[i]}) - \frac{N}{2}\log(2\pi) \end{aligned} \tag{19}$$

step 5: Sorting of the fireflies

Sort the fireflies in ascending order of their light intensities and find the current best position:

$$\boldsymbol{X}_{best}^{l} = \boldsymbol{X}_{[Q]} \tag{20}$$

step 6: Movement of the fireflies

If $I_i(\boldsymbol{X}_{[i]}) < I_j(\boldsymbol{X}_{[j]})$, move a firefly i at position $\boldsymbol{X}_{[i]}$ toward a brighter firefly j at position $\boldsymbol{X}_{[j]}$ by

$$\begin{aligned} \boldsymbol{X}_{[i]} = \boldsymbol{X}_{[i]} &+ \beta_0 \exp(-\gamma r_{ij}^2)(\boldsymbol{X}_{[j]} - \boldsymbol{X}_{[i]}) \\ &+\alpha_l \cdot rand() \end{aligned} \tag{21}$$

where r_{ij} is the Euclidean distance between $\boldsymbol{X}_{[i]}$ and $\boldsymbol{X}_{[j]}$, β_0 is the attractiveness at $r_{ij} = 0$, γ is the media absorption coefficient, α_l is the randomization parameter, and $rand()$ is uniformly distributed random number with amplitude in the range $[-0.5, 0.5]$. $\beta = \beta_0 \exp(-\gamma r_{ij}^2)$ is the attractiveness between the fireflies i and j.

step 7: Repetition

Set the iteration counter to $l = l + 1$ and go to *step 2* until the prespecified iteration number l_{max}.

step 8: Determination of the GP prior model

Determine the vector $\hat{\boldsymbol{X}} = \hat{\boldsymbol{\theta}}_c = [\hat{\sigma}_y, \hat{\ell}, \hat{\sigma}_n]^{\mathrm{T}}$ and the corresponding parameter vector $\hat{\boldsymbol{\theta}}_{ml} = [\hat{\boldsymbol{\theta}}_m^{\mathrm{T}}, \hat{\boldsymbol{\theta}}_l^{\mathrm{T}}]^{\mathrm{T}}$ using the best position $\boldsymbol{X}_{best}^{l_{max}}$ of firefly. Construct the suboptimal prior mean function and prior covariance function:

$$m(\boldsymbol{w}^f(t)) = (\boldsymbol{w}^f(t))^{\mathrm{T}}\hat{\boldsymbol{\theta}}_m \tag{22}$$

$$\begin{cases} s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) = \hat{\sigma}_y^2 \exp\left(-\dfrac{||\boldsymbol{w}^f(t_p) - \boldsymbol{w}^f(t_q)||^2}{2\hat{\ell}^2}\right) \\ k(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) = s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) + \hat{\sigma}_n^2\delta_{pq}, \end{cases} \tag{23}$$

where $s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q))$ is an element of covariance matrix $\boldsymbol{\Sigma}$, $k(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q))$ is an element of covariance matrix $\boldsymbol{K}$, and δ_{pq} is the Kronecker delta, which is 1 if $p = q$ and 0 otherwise.

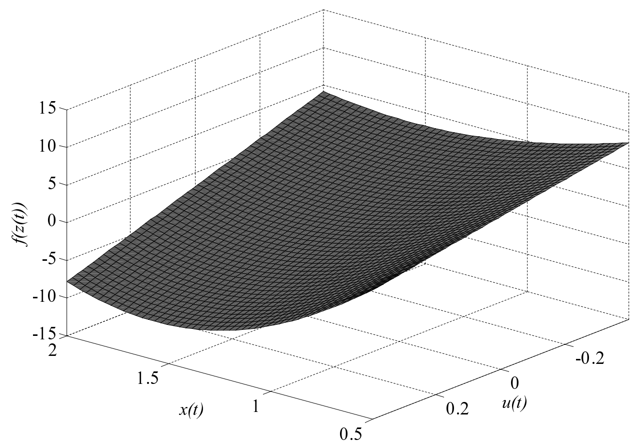

Fig. 1 True nonlinear function

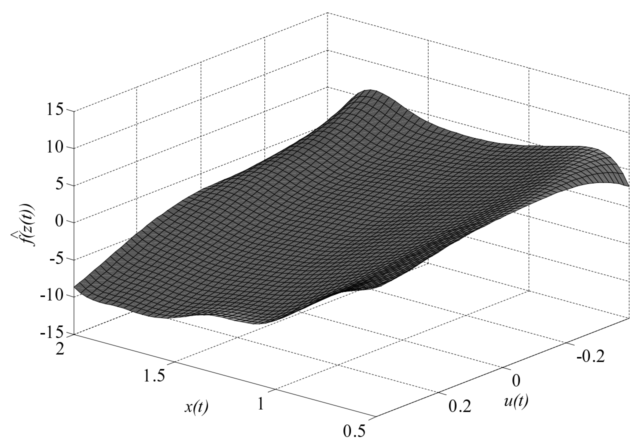

Fig. 2 Estimated nonlinear function

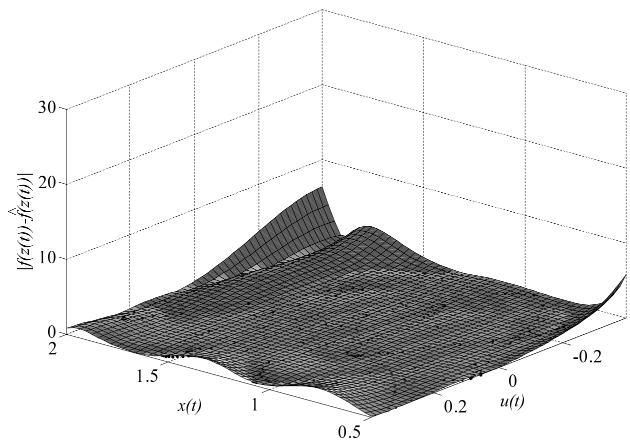

Fig. 3 Absolute error between true and estimated nonlinear functions

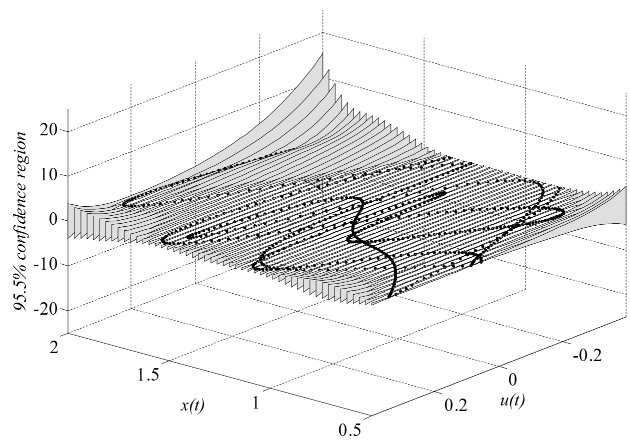

Fig. 4 95.5% confidence region

V. Estimation of the Nonlinear Function

For a new input $\boldsymbol{w}_*^f(t)$ and the corresponding function $f(\boldsymbol{w}_*^f(t))$, we have the following joint Gaussian distribution:

$$\begin{bmatrix} \boldsymbol{y} \\ f(\boldsymbol{w}_*^f(t)) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{m}(\boldsymbol{w}) + \boldsymbol{G}\hat{\boldsymbol{\theta}}_l \\ m(\boldsymbol{w}_*^f(t)) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w}_*^f(t)) \\ \boldsymbol{\Sigma}(\boldsymbol{w}_*^f(t), \boldsymbol{w}), & s(\boldsymbol{w}_*^f(t), \boldsymbol{w}_*^f(t)) \end{bmatrix} \right) \tag{24}$$

From the formula for conditioning a joint Gaussian distribution [14], the posterior distribution for specific test data is

$$f(\boldsymbol{w}_*^f(t))|\boldsymbol{w}, \boldsymbol{G}, \boldsymbol{y}, \boldsymbol{w}_*^f(t) \sim \mathcal{N}(\hat{f}(\boldsymbol{w}_*^f(t)), \hat{\sigma}_*^2(t)) \tag{25}$$

where the mean function $\hat{f}$ is given as

$$\hat{f}(\boldsymbol{w}_*^f(t)) = m(\boldsymbol{w}_*^f(t)) + \boldsymbol{\Sigma}(\boldsymbol{w}_*^f(t), \boldsymbol{w})\boldsymbol{K}^{-1}(\boldsymbol{y} - \boldsymbol{m}(\boldsymbol{w}) - \boldsymbol{G}\hat{\boldsymbol{\theta}}_l) \tag{26}$$

which is used as the estimated nonlinear function of the objective system. And its covariance $\hat{\sigma}_*^2$ is evaluated as

$$\hat{\sigma}_*^2(t) = s(\boldsymbol{w}_*^f(t), \boldsymbol{w}_*^f(t)) - \boldsymbol{\Sigma}(\boldsymbol{w}_*^f(t), \boldsymbol{w})\boldsymbol{K}^{-1}\boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w}_*^f(t)) \tag{27}$$

which is used for the confidence measure of the estimated nonlinear function.

VI. Illustrative example

Consider an electric power system [15] described by

$$\begin{cases} \ddot{x}(t) + a_1\dot{x}(t) = f(\boldsymbol{z}(t)) \\ f(\boldsymbol{z}(t)) = -\dfrac{P_e}{\tilde{M}} + \dfrac{P_{in}}{\tilde{M}} \\ \qquad = -\dfrac{P_{em}}{\tilde{M}}(1 + u(t))\sin x(t) + \dfrac{P_{in}}{\tilde{M}} \\ \boldsymbol{z}(t) = [x(t), u(t)]^{\mathrm{T}} \\ y(t) = x(t) + e(t) \end{cases} \tag{28}$$

where $x(t) = \delta(t)$: phase angle, $u(t) = \Delta E_{fd}(t)$: increment of excitation voltage, $\tilde{M}$: inertia coefficient, $\tilde{D}$: damping coefficient, P_e : generator output power, P_{in} : turbine output

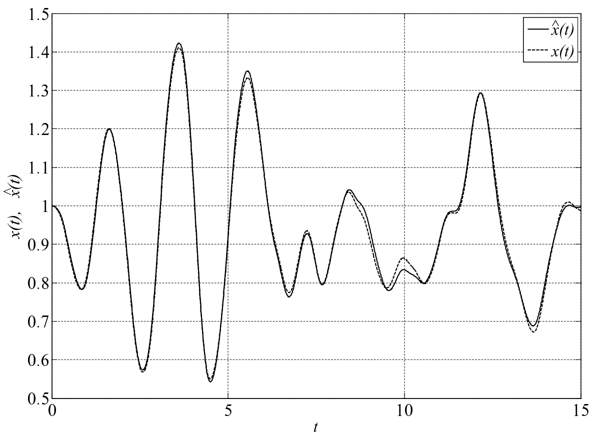

Fig. 5 True output and output by the estimated model

power. In numerical example, $\tilde{M} = \tilde{D} = 0.06$, $P_{em} = 1.0$, $P_{in} = 0.8$ and $a_1 = \tilde{D}/\tilde{M} = 1.0$ are set. The measurement noise $e(t)$ is white Gaussian noise, where noise-to-signal ratio is about 1.5%. The number of input and output data for identification is taken to be $N = 800$. The third-order Butterworth filter with the cutoff frequency $\omega_c = 10$ (rad/s) is utilized as a delayed state variable filter. The setting parameters of FA are chosen as follows:

(i) firefly size: $Q = 50$
(ii) attractiveness at $r_{ij} = 0$: $\beta_0 = 1.0$
(iii) media absorption coefficient: $\gamma = 1.0$
(iv) randomization parameter: $\alpha_l = 1.0 \times (0.97)^l$
(v) maximum iteration number: $l_{max} = 100$

The hyperparameter vector of the covariance function has been determined by FA as $\hat{\boldsymbol{\theta}}_c = [\hat{\sigma}_y, \hat{\ell}, \hat{\sigma}_n]^{\mathrm{T}} = [45.875, 0.343, 0.181]^{\mathrm{T}}$. Estimate of the parameter in the linear term is $\hat{a}_1 = 0.962$, which is very close to the true value $a_1 = 1.0$. The true nonlinear function $f(\boldsymbol{z}(t))$, the estimated nonlinear function $\hat{f}(\boldsymbol{z}(t))$, the absolute error between $f(\boldsymbol{z}(t))$ and $\hat{f}(\boldsymbol{z}(t))$, and the double standard deviation confidence interval (95.5% confidence region) around the estimated nonlinear function are shown in Figs. 1–4, respectively, where the thick curves depict the trajectories of the identification data. Clearly the estimated nonlinear function $\hat{f}(\boldsymbol{z}(t))$ is shown to be very close to the true nonlinear function $f(\boldsymbol{z}(t))$ on the data region. The confidence region of the estimated nonlinear function grows as $\boldsymbol{z}(t)$ goes away from the data region. On the other hand, the confidence region of the estimated nonlinear function is very small on the data region. Fig. 5 shows the true output $x(t)$ and the output $\hat{x}(t)$ by the estimated model, where the outputs were generated by the inputs for validation. This figure indicates that $\hat{x}(t)$ matches $x(t)$ considerably.

For comparison, the PSO-based GP (PSOGP) method [7] has been carried out for this identification problem. The mean squares error of the output $\sum_{k=1}^{N}(x(t_k) - \hat{x}(t_k))^2/N$ is 2.651×10^{-4} for the proposed method and 2.757×10^{-4} for PSOGP method. The computational time of the training is 242.7 (s) for the proposed method and 303.4 (s) for PSOGP method (CPU: Intel(R) Core(TM) i7-2640M 2.80GHz). We can confirm that the proposed method gives more accurate model of the objective electric power system with smaller computational burden.

VII. Conclusions

In this paper, we have proposed a separable LS approach combining the linear LS method with FA to train a GP prior model for nonlinear system identification. The GP prior model is trained by maximizing the log marginal likelihood of the identification data. The proposed identification method is categorized into the nonparametric identification and does not need the determination of the model structure. Since FA is simple and has a high potential for global optimization, the proposed training algorithm makes system identification be more efficient. Simulation results show that the proposed method can be successfully applied to modeling of the electric power system.

References

[1] C. Z. Jin, K. Wada, K. Hirasawa, and J. Murata, Identification of nonlinear continuous systems by using neural network compensator (in Japanese), *IEEJ Trans. C*, Vol. 114, No. 5, pp. 595–602, 1994.

[2] K. M. Tsang and S. A. Billings, Identification of continuous time nonlinear systems using delayed state variable filters, *International Journal of Control*, Vol. 60, No. 2, pp. 159–180, 1994.

[3] G. P. Liu and V. Kadirkamanathan, Stable sequential identification of continuous nonlinear dynamical systems by growing radial basis function networks, *International Journal of Control*, Vol. 65, No. 1, pp. 53–69, 1996.

[4] T. Hachino, I. Karube, Y. Minari, and H. Takata, Continuous-time identification of nonlinear systems using radial basis function network model and genetic algorithm, *Proc. of the 12th IFAC Symposium on System Identification*, Vol. 2, pp. 787–792, 2001.

[5] J. Kocijan, A. Girard, B. Banko, and R. Murray-Smith, Dynamic systems identification with Gaussian processes, *Mathematical and Computer Modelling of Dynamical Systems*, Vol. 11, No. 4, pp. 411–424, 2005.

[6] T. Hachino and H. Takata, Identification of continuous-time nonlinear systems by using a Gaussian process model, *IEEJ Trans. on Electrical and Electronic Engineering*, Vol. 3, No. 6, pp. 620–628, 2008.

[7] T. Hachino, S. Yoneda, and H. Takata, Gaussian process models trained by particle swarm optimization for continuous-time nonlinear system identification, *Proc. of 2011 International Symposium on Nonlinear Theory and Its Applications*, pp. 276–281, 2011.

[8] A. O'Hagan, Curve fitting and optimal design for prediction (with discussion), *Journal of the Royal Statistical Society B*, Vol. 40, pp. 1–42, 1978.

[9] C. K. I. Williams, Prediction with Gaussian processes: from Linear regression to linear prediction and beyond, in *Learning and Inference in Graphical Models*, Kluwer Academic Press, pp. 599–621, 1998.

[10] M. Seeger, Gaussian processes for machine learning, *International Journal of Neural Systems*, Vol. 14, No. 2, pp. 1–38, 2004.

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

[12] X. S. Yang, Firefly algorithms for multimodal optimization, in *Stochastic Algorithms: Foundations and Applications*, *SAGA 2009, Lecture Notes in Computer Sciences*, Vol. 5792, pp. 169–178, 2009.

[13] S. K. Pal, C. S. Rai, and A. P. Singh, Comparative study of firefly algorithm and particle swarm optimization for noisy non-linear optimization problems, *International Journal of Intelligent Systems and Applications*, Vol. 4, No. 10, pp. 50–57, 2012.

[14] R. von Mises, *Mathematical Theory of Probability and Statistics*, Academic Press, 1964.

[15] H. Takata, An automatic choosing control for nonlinear systems, *Proc. of the 35th IEEE CDC*, pp. 3453–3458, 1996.