2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

Gaussian Process Models Trained by Particle Swarm Optimization for Continuous-Time Nonlinear System Identification

Tomohiro Hachino[†], Shinichiro Yoneda[†] and Hitoshi Takata[‡]

[†]Department of Electrical and Electronics Engineering, Kagoshima University
1-21-40 Korimoto, Kagoshima, 890-0065 Japan
[‡]Professor Emeritus, Kagoshima University
Email: hachino@eee.kagoshima-u.ac.jp

Abstract—This paper proposes a nonparametric identification of continuous-time nonlinear systems by using a Gaussian process (GP) model. The GP prior model is trained by particle swarm optimization. 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 given by the predictive covariance of the GP.

1. 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. Nonlinear system identification based on the continuous-time model has been developed by using neural network [1], RBF model [2], and other methods. However, since these methods are categorized into the parametric identification, one needs many weighting parameters of any basis functions to describe the nonlinearity. Moreover, any confidence measures for the estimated nonlinear function are not given in such identification methods.

In recent years, the Gaussian process (GP) model has received much attention for nonlinear system identification [3, 4]. The GP model is a nonparametric model and fits naturally into Bayesian framework [5]. Since it has fewer parameters called *hyperparameters* than the parametric models, we can describe the nonlinearity of the objective system in a few parameters. Therefore, in this paper, we proposes a nonparametric identification of continuous-time nonlinear systems by using the GP model. The GP prior model is trained by using particle swarm optimization (PSO) [6] which is a swarm intelligence optimization technique inspired by the social behavior of a flock of birds or a shoal of fish. 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 given by the predictive covariance of the GP. To perform the nonlinear system identification in the GP framework, the GP prior model has to be trained by minimizing the negative log marginal likelihood of the identification data. The conjugate gradient algorithms have been utilized to train the GP model, but the gradient-based optimization algorithms still suffer from the local minima problem unless the initial guess is suitable. An alternative approach is use of the genetic algorithm (GA), but the GA requires complicated coding and genetic operations such as crossovers and mutations. In this paper, the hyperparameters of the covariance functions are searched by PSO, and the weighting parameters of the prior mean function and the parameters in the linear term corresponding to each candidate hyperparameter vector are estimated by the linear least-squares method. PSO is simpler than the GA, because the algorithm of PSO consists of only the basic arithmetic operations and does not require complicated coding and genetic operations. Therefore, the use of PSO increases the efficiency of nonlinear system identification.

This paper is organized as follows. In section 2 the problem is formulated. In section 3 the GP prior model for the identification is derived. In section 4 the GP prior model is trained by using PSO and the nonlinear function are estimated with the confidence measure in the GP framework. In section 5 simulation results are shown to illustrate the effectiveness of the proposed method. Finally conclusions are given in section 6.

2. Statement of the Problem

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

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

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

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

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

where $u(t)$ and $x(t)$ are the true input and output signals, respectively. $y(t)$ is the noisy output which 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 aim 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.

3. GP model for Identification

Equation (1) can be rewritten as

$$
\begin{aligned}
p^n y(t) &= f(\boldsymbol{w}(t)) - \sum_{\substack{i=1 \\ i\neq n_1,n_2,\cdots,n_\alpha}}^{n} a_i p^{n-i} y(t) + \sum_{\substack{j=0 \\ j\neq m_1,m_2,\cdots,m_\beta}}^{m} b_j p^{m-j} u(t) + \varepsilon(t) \\
\boldsymbol{w}(t) &= [p^{n-n_1}y(t), p^{n-n_2}y(t), \cdots, p^{n-n_\alpha}y(t), \\
&\quad p^{m-m_1}u(t), p^{m-m_2}u(t), \cdots, p^{m-m_\beta}u(t)]^{\mathrm{T}}
\end{aligned} \tag{2}
$$

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

Multiplying both sides of (2) by the state variable filter $F(p)$ which has a transport lag characteristic yields

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

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

$$
\boldsymbol{y} = \boldsymbol{v} + \boldsymbol{G}\boldsymbol{\theta}_l \tag{4}
$$

where

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

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:

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

where

$$
\begin{aligned}
\boldsymbol{f} &= [f(\boldsymbol{w}^f(t_1)), f(\boldsymbol{w}^f(t_2)), \cdots, f(\boldsymbol{w}^f(t_N))]^{\mathrm{T}} \\
\boldsymbol{w} &= [\boldsymbol{w}^f(t_1), \boldsymbol{w}^f(t_2), \cdots, \boldsymbol{w}^f(t_N)].
\end{aligned} \tag{7}
$$

$\boldsymbol{w}$ is the input (variable) 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 [5]. In this paper the mean function is expressed by the first order polynomial, i.e., a linear combination of the input variable:

$$
\begin{aligned}
m(\boldsymbol{w}^f(t)) &= (\boldsymbol{w}^f(t))^{\mathrm{T}}\boldsymbol{\theta}_m \\
\boldsymbol{\theta}_m &= \left[\theta_{n_1}, \theta_{n_2}, \cdots, \theta_{n_\alpha}, \theta_{m_1}, \theta_{m_2}, \cdots, \theta_{m_\beta}\right]^{\mathrm{T}}
\end{aligned} \tag{8}
$$

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:

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

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:

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

where $\|\cdot\|$ denotes the Euclidean norm. Eq. (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 $\boldsymbol{w}^f(t_q)$.

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

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

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

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

4. Identification

4.1. Training of GP prior model by PSO

At the first stage of the identification, the GP prior model is trained by optimizing the unknown parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_m^{\mathrm{T}}, \boldsymbol{\theta}_l^{\mathrm{T}}, \boldsymbol{\theta}_c^{\mathrm{T}}]^{\mathrm{T}}$. Although this is a nonlinear optimization problem, we can separate the linear optimization part and the nonlinear optimization part. Therefore, in this paper, we propose a method that combines the LS method with PSO. Only $\boldsymbol{X} = [\boldsymbol{\theta}_c^{\mathrm{T}}, \omega_c]^{\mathrm{T}}$ is searched for using PSO, where ω_c is the cutoff frequency of the state variable filter. The proposed training algorithm is as follows:

Step 1: Initialization

Generate an initial population of Q particles with random positions $\boldsymbol{X}^0_{[i]} = [\boldsymbol{\theta}^{\mathrm{T}}_{c[i]}, \omega_{c[i]}]^{\mathrm{T}}$ and velocities $\boldsymbol{V}^0_{[i]}$ $(i = 1, 2, \cdots, Q)$.

Set the iteration counter l to 0.

Step 2: Filtering of the identification data

Construct Q candidates of the state variable filter using $\omega_{c[i]}$ $(i = 1, 2, \cdots, Q)$. Calculate the filtered input $u^f_{[i]}(t)$, filtered output $y^f_{[i]}(t)$ and their higher-order derivatives, using each candidate of the state variable filter. Then construct Q candidates of $\boldsymbol{y}_{[i]}$ and $\boldsymbol{Z}_{[i]} = [\boldsymbol{w}^{\mathrm{T}}_{[i]} \vdots \boldsymbol{G}_{[i]}]$ $(i = 1, 2, \cdots, Q)$.

Step 3: Construction of covariance matrix

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

Step 4: Estimation of $\boldsymbol{\theta}_{ml}$

Estimate Q candidates of $\boldsymbol{\theta}_{ml[i]} = [\boldsymbol{\theta}^{\mathrm{T}}_{m[i]}, \boldsymbol{\theta}^{\mathrm{T}}_{l[i]}]^{\mathrm{T}}$ corresponding to $\boldsymbol{X}^l_{[i]}$ $(i = 1, 2, \cdots, Q)$:

$$\boldsymbol{\theta}_{ml[i]} = (\boldsymbol{Z}^{\mathrm{T}}_{[i]}\boldsymbol{K}^{-1}_{[i]}\boldsymbol{Z}_{[i]})^{-1}\boldsymbol{Z}^{\mathrm{T}}_{[i]}\boldsymbol{K}^{-1}_{[i]}\boldsymbol{y}_{[i]}. \tag{14}$$

Step 5: Evaluation value calculation

Calculate the evaluation values which are the values of the negative log marginal likelihood of the identification data:

$$\begin{aligned} J(\boldsymbol{X}^l_{[i]}) = &\frac{1}{2}\log|\boldsymbol{K}_{[i]}| + \frac{1}{2}(\boldsymbol{y}_{[i]} - \boldsymbol{Z}_{[i]}\boldsymbol{\theta}_{ml[i]})^{\mathrm{T}}\boldsymbol{K}^{-1}_{[i]} \\ &\times(\boldsymbol{y}_{[i]} - \boldsymbol{Z}_{[i]}\boldsymbol{\theta}_{ml[i]}) + \frac{N}{2}\log(2\pi). \end{aligned} \tag{15}$$

Step 6: Update of the best positions *pbest* and *gbest*

Update $\boldsymbol{pbest}^l_{[i]}$, which is the personal best position, and $\boldsymbol{gbest}^l$, which is the global best position among all particles as follows:

If $l = 0$ then

$$\begin{aligned} &\boldsymbol{pbest}^l_{[i]} = \boldsymbol{X}^l_{[i]} \\ &\boldsymbol{gbest}^l = \boldsymbol{X}^l_{[i_{best}]} \quad i_{best} = arg \min_i J(\boldsymbol{X}^l_{[i]}) \end{aligned} \tag{16}$$

otherwise

$$\begin{aligned} &\boldsymbol{pbest}^l_{[i]} = \begin{cases} \boldsymbol{X}^l_{[i]} & (J(\boldsymbol{X}^l_{[i]}) < J(\boldsymbol{pbest}^{l-1}_{[i]})) \\ \boldsymbol{pbest}^{l-1}_{[i]} & (\text{otherwise}) \end{cases} \\ &\boldsymbol{gbest}^l = \boldsymbol{pbest}^l_{[i_{best}]} \quad i_{best} = arg \min_i J(\boldsymbol{pbest}^l_{[i]}). \end{aligned} \tag{17}$$

Step 7: Update of positions and velocities

Update the particle positions and velocities using (18):

$$\begin{cases} \boldsymbol{V}^{l+1}_{[i]} = w^l \cdot \boldsymbol{V}^l_{[i]} + c_1 \cdot rand_1() \cdot (\boldsymbol{pbest}^l_{[i]} - \boldsymbol{X}^l_{[i]}) \\ \qquad\qquad +c_2 \cdot rand_2() \cdot (\boldsymbol{gbest}^l - \boldsymbol{X}^l_{[i]}) \\ \boldsymbol{X}^{l+1}_{[i]} = \boldsymbol{X}^l_{[i]} + \boldsymbol{V}^{l+1}_{[i]} \end{cases} \tag{18}$$

where w^l is an inertia factor, c_1 and c_2 are constants representing acceleration coefficients, and $rand_1()$ and $rand_2()$ are uniformly distributed random numbers with amplitude in the range $[0, 1]$.

Step 8: Repetition

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

Step 9: Determination of the GP model

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

$$m(\boldsymbol{w}^f(t)) = (\boldsymbol{w}^f(t))^{\mathrm{T}}\hat{\boldsymbol{\theta}}_m \tag{19}$$

$$\begin{cases} s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) = \hat{\sigma}^2_y \exp\left(-\dfrac{\|\boldsymbol{w}^f(t_p) - \boldsymbol{w}^f(t_q)\|^2}{2\hat{\ell}^2}\right) \\ k(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) = s(\boldsymbol{w}^f(t_p), \boldsymbol{w}^f(t_q)) + \hat{\sigma}^2_n\delta_{pq}, \end{cases} \tag{20}$$

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.

4.2. 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:

$$\begin{aligned} \begin{bmatrix} \boldsymbol{y} \\ f(\boldsymbol{w}^f_*(t)) \end{bmatrix} \sim \mathcal{N}\Bigg(&\begin{bmatrix} \boldsymbol{m}(\boldsymbol{w}) + \boldsymbol{G}\hat{\boldsymbol{\theta}}_l \\ m(\boldsymbol{w}^f_*(t)) \end{bmatrix}, \\ &\begin{bmatrix} \boldsymbol{K} & \boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w}^f_*(t)) \\ \boldsymbol{\Sigma}(\boldsymbol{w}^f_*(t), \boldsymbol{w}), & s(\boldsymbol{w}^f_*(t), \boldsymbol{w}^f_*(t)) \end{bmatrix}\Bigg). \end{aligned} \tag{21}$$

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

$$f(\boldsymbol{w}^f_*(t))|\boldsymbol{w}, \boldsymbol{G}, \boldsymbol{y}, \boldsymbol{w}^f_*(t) \sim \mathcal{N}(\hat{f}(\boldsymbol{w}^f_*(t)), \hat{\sigma}^2_*(t)) \tag{22}$$

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

$$\begin{aligned} \hat{f}(\boldsymbol{w}^f_*(t)) = &\, m(\boldsymbol{w}^f_*(t)) \\ &+\boldsymbol{\Sigma}(\boldsymbol{w}^f_*(t), \boldsymbol{w})\boldsymbol{K}^{-1}(\boldsymbol{y} - \boldsymbol{m}(\boldsymbol{w}) - \boldsymbol{G}\hat{\boldsymbol{\theta}}_l) \end{aligned} \tag{23}$$

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

$$\hat{\sigma}^2_*(t) = s(\boldsymbol{w}^f_*(t), \boldsymbol{w}^f_*(t)) - \boldsymbol{\Sigma}(\boldsymbol{w}^f_*(t), \boldsymbol{w})\boldsymbol{K}^{-1}\boldsymbol{\Sigma}(\boldsymbol{w}, \boldsymbol{w}^f_*(t)) \tag{24}$$

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

5. Illustrative example

Consider a system described by

$$\begin{cases} \ddot{x}(t) = f(\boldsymbol{z}(t)) + b_0u(t) \\ f(\boldsymbol{z}(t)) = -(0.8 + 2x^2(t))\dot{x}(t) + 4.0x(t) \\ b_0 = 4.0 \\ y(t) = x(t) + e(t). \end{cases} \tag{25}$$

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 is utilized as a delayed state variable filter. The design parameters for PSO are chosen as follows:

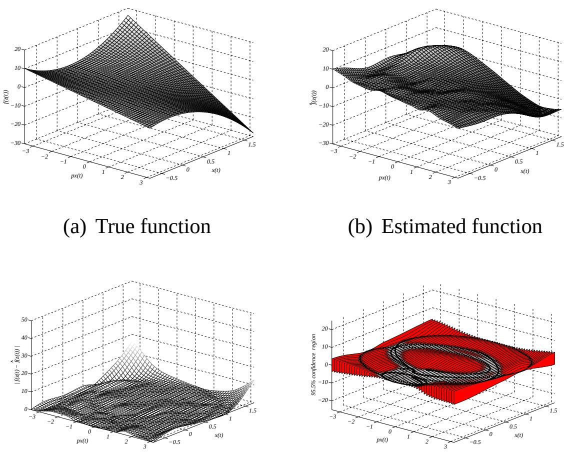
1) particle size: $Q = 30$

2) inertia factor: $w^l = w_{max} - (w_{max} - w_{min})l/l_{max}$, ($w_{max} = 0.8$, $w_{min} = 0.4$)

3) acceleration coefficients $c_1 = 0.8$, $c_2 = 0.8$

4) maximum iteration number $l_{max} = 100$

The hyperparameters of the covariance function and the cutoff frequency of the state variable filter have been determined by PSO as $\hat{\boldsymbol{X}} = [\hat{\sigma}_y, \hat{\ell}, \hat{\sigma}_n, \hat{\omega}_c]^{\mathrm{T}} = [1.846, 0.359, 0.097, 9.032]^{\mathrm{T}}$. Estimate of the parameter in the linear term is $\hat{b}_0 = 4.052$, which is very close to the true value $b_0 = 4.0$. 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 Fig.1, 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.2 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.



(a) True function

(b) Estimated function

(c) Absolute error

(d) 95.5% confidence region

Figure 1: Estimated nonlinear function

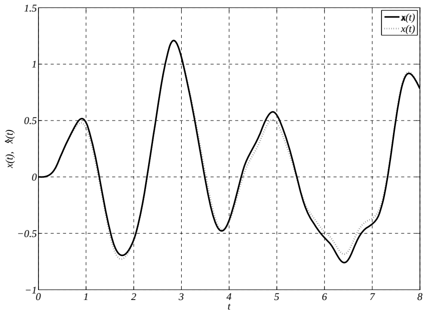


Figure 2: True output and output by the estimated model

6. Conclusions

In this paper we have proposed an identification method of continuous-time nonlinear systems using the GP model. The GP prior model is trained by the aid of PSO so that the negative log marginal likelihood of the identification data is minimized. The proposed identification method is categorized into the nonparametric identification and does not need the determination of the model structure. Simulation results show that the proposed method can accurately estimate the nonlinear function with the confidence measure.

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. on Electronics, Information and Systems*, vol.114-C, no.5, pp.595-602, 1994.

[2] G. P. Liu and V. Kadirkamanathan, "Stable Sequential Identification of Continuous Nonlinear Dynamical Systems by Growing Radial Basis Function Networks", *Int. J. Control*, vol.65, no.1, pp.53-69, 1996.

[3] 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.

[4] 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.

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

[6] J. Kennedy and R. C. Eberhart, "Particle Swarm Optimization", *Proc. of IEEE International Conference on Neural Networks*, pp.1942-1948, 1995.