Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

A Hybrid Modeling Method for Precise Positioning Systems

Mitsuo Hirata [*] **Sakae Noguchi** [*] **Shuichi Adachi** [**]

[*] *Utsunomiya University, Tochigi, Japan (e-mail: hirata@cc.utsunomiya-u.ac.jp)*
[**] *Keio University, Yokohama, Japan (e-mail: adachi@appi.keio.ac.jp)*

Abstract: In this paper, a system identification method for hybrid systems switched by the magnitude of velocity is proposed. First, it is shown that the regression vector space of a mechanical system switched by the magnitude of velocity cannot be separated by a hyperplane. Then a method based on support vector machines with a polynomial kernel is proposed. The effectiveness of the proposed method is shown by simulations.

Keywords: System identification, hybrid systems, support vector machine, polynomial kernel

1. INTRODUCTION

In positioning control, the characteristics of the plant might change for long span motion and short span motion. It is considered that the phenomenon is caused by nonlinear friction and/or hysteresis (Futami et al., 1990). However, in general, it is difficult to model these characteristics by a physical interpretation. Therefore, it will be useful, if the nonlinear model can be identified from input-output data.

Recently, modeling and design of hybrid systems have received more attention both academia and industry. It provides a unified framework to the mixed system with continuous dynamics and discrete events. Ferrari-Trecate et al. (2003) proposed an identification method of hybrid systems. The method can identify a hybrid system described by a piecewise affine (PWA) system from input-output data and estimate the switching rules at the same time.

A class of system with nonlinear friction can be approximated by a hybrid system (Noguchi et al., 2005). For example, a system in which the friction force is changed for higher speed and lower speed can be modeled as a hybrid system with two sub-models that are switched by the magnitude of velocity. However, in the case that the switching condition depends on the magnitude of velocity, it is difficult to apply the methods proposed by Ferrari-Trecate et al. (2003).

In this paper, a new identification method for a mechanical hybrid system switched by the magnitude of velocity is proposed. First, it is shown that the regression vector space of the hybrid system cannot be separated by a hyperplane. Then a method based on support vector machines with a polynomial kernel is proposed. The effectiveness of the proposed method is shown by simulations.

2. IDENTIFICATION OF HYBRID SYSTEMS

In this section, the identification method proposed by Ferrari-Trecate et al. (2003) is summarized. The plant is assumed to be described by a piecewise ARX (PWARX) model with s sub-modes:

$$y[k] = \begin{cases} \boldsymbol{\theta}_1^T \boldsymbol{\varphi}[k] + w[k], \text{ if } \boldsymbol{\varphi}[k] \in C_1 \\ \vdots \\ \boldsymbol{\theta}_s^T \boldsymbol{\varphi}[k] + w[k], \text{ if } \boldsymbol{\varphi}[k] \in C_s \end{cases} \quad (1)$$

$$\boldsymbol{\theta}_i := [a_{i1}, \ldots, a_{in}, b_{i1}, \ldots, b_{in}]^T$$

$$\boldsymbol{\varphi}[k] := [\,-y[k-1], \ldots, -y[k-n], u[k-1], \ldots, u[k-n]\,]^T$$

where $w[k]$ is white noise, $\boldsymbol{\theta}_i$ $(i = 1, \ldots, s)$ are parameter vectors, $\boldsymbol{\varphi}[k]$ is a regression vector. n is an order of the PWARX model, and it is assumed that the order for each sub-model is the same. $u[k]$ and $y[k]$ are input and output respectively. $\{C\}_{i=1}^s$ are polytopic and they satisfy well-posedness condition

$$\cup_{i=1}^s C_i = C, \quad C_i \cap C_j = \emptyset, \quad \forall i \neq j$$

2.1 Clustering of data points

A given input-output data $(u[k], y[k])_{k=1}^N$ are clustered according to the following procedure.

Step 1 Data points $(\boldsymbol{\varphi}[k], y[k])_{i=1}^N$ are classified into clusters $\mathcal{C}_j$ including N_c data points. $\mathcal{C}_j$ is built by collecting $N_c - 1$ neighboring data points of $(\boldsymbol{\varphi}[j], y[j])$. Euclidean norm is used to evaluate the distance between data points as

$$\|\boldsymbol{\varphi}[j] - \boldsymbol{\varphi}[i]\|.$$

Step 2 For each cluster $\mathcal{C}_j$, a parameter vector $\boldsymbol{\theta}^{LS,j}$ is obtained using the least squares method:

$$\boldsymbol{\theta}^{LS,j} = (\boldsymbol{\Phi}_j^T \boldsymbol{\Phi}_j)^{-1} \boldsymbol{\Phi}_j^T \boldsymbol{y}_{\mathcal{C}_j}, \quad j = 1, \cdots, N$$

$$\boldsymbol{\Phi}_j = [\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_{N_c}]^T,$$

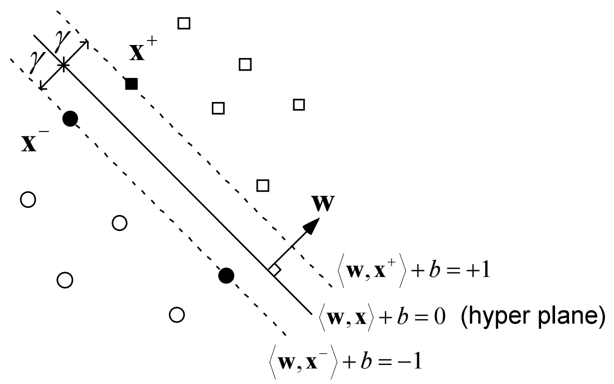


Fig. 1. Training data and hyperplane

$$\boldsymbol{y}_{\mathcal{C}_j} = [y_1, \cdots, y_{N_c}]^T$$

where $(\boldsymbol{\varphi}_i, y_i)_{i=1}^{N_c}$ are data points that belong to the cluster $\mathcal{C}_j$.

Step 3 Cluster the parameter vectors $\boldsymbol{\theta}^{LS,j}$ into s disjoint subsets $\mathcal{D}_i$ using K-means algorithm which is well-known clustering algorithm, and it updates the parameter vector in the cluster $\mathcal{D}_i$, $(i = 1, \ldots, s)$ and the centers $\boldsymbol{\mu}_i$ of clusters in such a way to minimize the clustering function:

$$J = \sum_{i=1}^{s} \sum_{\boldsymbol{\theta}^{LS,j} \in \mathcal{D}_i} \|\boldsymbol{\theta}^{LS,j} - \boldsymbol{\mu}_i\|^2$$

Step 4 Classify the original data points $(\boldsymbol{\varphi}[k], y[k])_{i=1}^{N}$ into s classes by using the bijective maps between parameter vectors $\boldsymbol{\theta}^{LS,j}$ and clusters $\mathcal{C}_j$ and between clusters $\mathcal{C}_j$ and data points $(\boldsymbol{\varphi}[k], y[k])_{i=1}^{N}$.

Since the original data points are now classified, the final s ARX sub-models can be identified by applying the least squares method to the data points in each cluster $\mathcal{D}_j$.

2.2 Estimation of a hyperplane that separates sub-models using SVM

A hyperplane that separates sub-models on the regression space is estimated by using Support Vector Machine (SVM) that is one of powerful tools for classification (Cristianini and Shawe-Taylor, 2000; Adachi, 2004). The modes are estimated by a discriminant function with respect to the regression vector $\boldsymbol{\varphi}[k]$.

For given training data points $(\boldsymbol{x}_i, y_i)$ composed of input data $\boldsymbol{x}_i \in \mathcal{R}^n$ and class label $y_i \in \{-1, 1\}$, the problem is to find a linear discriminant function that separates the data points $\boldsymbol{x}_i$ into two classes:

$$f(\boldsymbol{x}_i) = \langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b \tag{2}$$

where $\langle \cdot, \cdot \rangle$ is an inner-product, $\boldsymbol{w}$ is a normal vector of the hyperplane, and b is a bias term. As shown in Fig. 1 SVM separates data points by the hyperplane so that the distance from the hyperplane to the nearest data point is maximized. The data points nearest to the hyperplane are called the support vectors.

The problem that maximizes the margin between the hyperplane and the nearest data point can be formulated

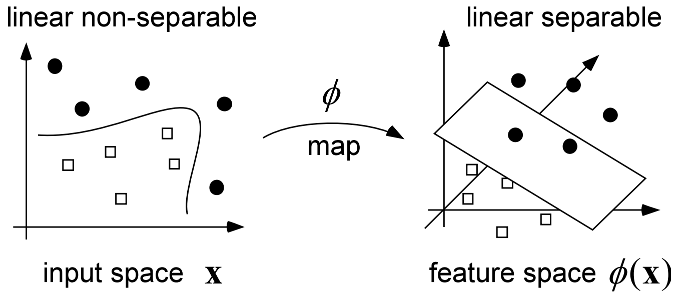


Fig. 2. Mapping to higher dimensional feature space

as a quadratic programming optimization problem (primal problem):

$$\min_{\boldsymbol{w}, b} \mathcal{J}(\boldsymbol{w}) = \frac{1}{2}\langle \boldsymbol{w}, \boldsymbol{w} \rangle \tag{3}$$

$$\text{subject to} \quad (\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) y_i \geq 1, \quad i = 1, \cdots, N \tag{4}$$

where the maximum margin is $\gamma = 1/\|\boldsymbol{w}\|$. The dual problem is given as follows:

$$\begin{aligned} &\max_{\alpha} \mathcal{Q}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \\ &\text{subject to} \quad \sum_{i=1}^{N} y_i \alpha_i = 0 \\ &\qquad\qquad\quad \alpha_i \geq 0, \ i = 1, \cdots, N \end{aligned} \tag{5}$$

When the data points are not linearly separable, they might be linearly separable by mapping them into higher-dimensional space using a nonlinear function $\phi(\cdot)$ as shown in Fig. 2. It is equivalent to use a nonlinear discriminant function

$$f(\boldsymbol{x}_i) = \langle \boldsymbol{w}, \boldsymbol{\phi}(\boldsymbol{x}_i) \rangle + b \tag{6}$$

instead of Eq.(2). In general, the dimension of $\phi(\boldsymbol{x}_i)$ tends to be very high, and the optimization might be difficult.

However, in the dual problem Eq.(5), higher dimensional space does not appear explicitly as:

$$\max_{\alpha} \mathcal{Q}(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \alpha_i \alpha_j \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle.$$

It just requires to calculate the inner product of $\phi(\cdot)$, and

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle$$

is called kernel function (Cristianini and Shawe-Taylor, 2000). This technique is well known as the kernel trick. A polynomial kernel, a Gaussian kernel, and a Sigmoid kernel are commonly used. For example, the polynomial kernel is defined as

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle^d \tag{7}$$

or

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle + 1)^d. \tag{8}$$

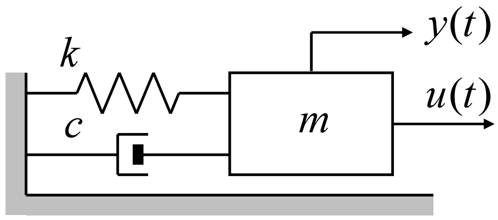


Fig. 3. Spring-mass-damper system

3. IDENTIFICATION OF A MECHANICAL HYBRID SYSTEM SWITCHED BY THE MAGNITUDE OF VELOCITY

3.1 A system switched by the magnitude of velocity

Let us consider a spring-mass-damper system as shown in Fig. 3, in which m [kg] is a mass, k [N/m] is a spring constant, and c [Ns/m] is a damping coefficient. The output $y(t)$ is a displacement [m] from the equilibrium position, and the input $u(t)$ is a force [N] applied to the mass. The equation of motion is described by

$$m\ddot{y}(t) + c\dot{y}(t) + ky(t) = u(t). \quad (9)$$

Now, it is assumed that the damping coefficient c is switched between c_1 and c_2, and the discrete-time model of Eq.(9) with a sampling period T_s is obtained:

$$\begin{aligned} y[k] &= \boldsymbol{\theta}_i^T \boldsymbol{\varphi}[k], \quad i = 1, 2 \\ \boldsymbol{\varphi}[k] &:= [-y[k-1], -y[k-2], u[k-1], u[k-2]]^T \end{aligned} \quad (10)$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are parameter vectors for $c = c_1$ and c_2 respectively, $\boldsymbol{\varphi}[k]$ is a regression vector, $y[k] := y(T_s k)$, and $u[k] := u(T_s k)$.

The velocity is defined by the backward difference of $y[k]$ as

$$v[k-1] := \frac{y[k-1] - y[k-2]}{T_s}, \quad (11)$$

and consider a hybrid system switched by the magnitude of velocity as:

$$y[k] = \begin{cases} \boldsymbol{\theta}_1^T \boldsymbol{\varphi}[k] & \text{if } |v[k-1]| < V_{sw} \\ \boldsymbol{\theta}_2^T \boldsymbol{\varphi}[k] & \text{if } |v[k-1]| \geq V_{sw} \end{cases}. \quad (12)$$

This system has two modes: mode 1 for $|v[k-1]| < V_{sw}$ and mode 2 for $|v[k-1]| \geq V_{sw}$

For simplicity, $u[k] = 0, \forall k$ is assumed, and let us consider the relation between the trajectory of $y[k]$ and the corresponding modes on $y[k-2]$—$y[k-1]$ plane, i.e., in regression space. Fig. 4 is obtained from Eqs.(11) and (12), and it shows that the trajectory of $y[k]$ is on C_b in mode 1 and on C_a or C_c in mode 2. Therefore, it is impossible to separate these modes by one hyperplane.

Let us divide mode 2 into two modes; mode 2a and mode2c that are corresponding to C_a and C_c respectively, then between C_a and C_b and between C_b and C_c are separated by a hyperplance. However, the method proposed by Ferrari-Trecate et al. (2003) can not apply to this case because the both of mode 2a and mode 2c have same parameter vector $\boldsymbol{\theta}_2$ and it is impossible to distinguish mode 2a and mode 2c in the parameter space.

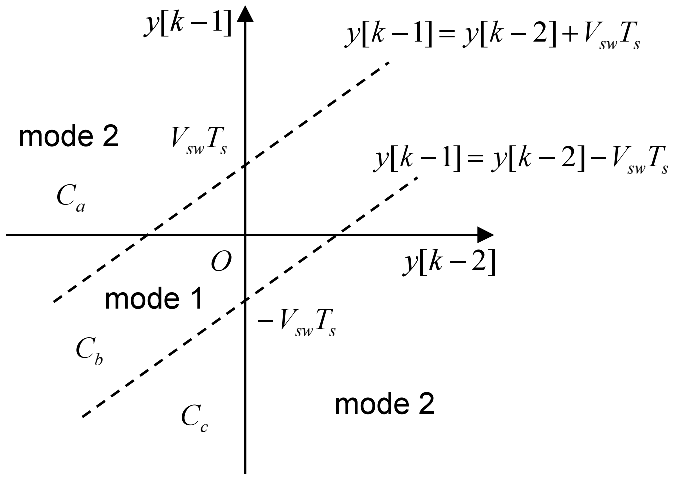


Fig. 4. Regression vector space

3.2 Identification of hybrid system using polynomial kernel

As explained in section 3.1, the regression space cannot be separated into mode 1 and mode 2 if the system is switched by the magnitude of velocity. Thus we consider to separate the regression space by a nonlinear discriminant function. From Fig. 4, it is easy to confirm that C_b and $C_a \cup C_c$ can be represented by:

$$C_b = \left\{ \boldsymbol{y} \mid \boldsymbol{y}^T \boldsymbol{M} \boldsymbol{y} < (V_{sw} T_s)^2 \right\}, \quad (13)$$

$$C_a \cup C_c = \left\{ \boldsymbol{y} \mid \boldsymbol{y}^T \boldsymbol{M} \boldsymbol{y} \geq (V_{sw} T_s)^2 \right\} \quad (14)$$

where

$$\boldsymbol{y} := \begin{bmatrix} y[k-1] \\ y[k-2] \end{bmatrix}, \quad \boldsymbol{M} := \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \quad (15)$$

This means that it is possible to decide the mode to which the trajectory of $y[k]$ belongs by the sign of

$$\begin{aligned} &\boldsymbol{y}^T \boldsymbol{M} \boldsymbol{y} - (V_{sw} T_s)^2 \\ &= y^2[k-1] - 2y[k-1]y[k-2] \\ &\quad + y^2[k-2] - (V_{sw} T_s)^2. \end{aligned} \quad (16)$$

Therefore, by a quadratic function with respect to the regression vector as shown below, the regression space can be separated into mode 1 and mode 2:

$$f(\boldsymbol{\varphi}[k]) = \boldsymbol{\varphi}^T[k] \boldsymbol{M} \boldsymbol{\varphi}[k] + b, \quad (17)$$

where $\boldsymbol{\varphi}[k]$ is a regression vector, and $\boldsymbol{M} = [m_{ij}]$ is a symmetric matrix.

Eq.(17) can be formulated by a polynomial kernel. As an example, $\boldsymbol{\varphi} = [\varphi_1, \varphi_2]^T$ is assumed, and a nonlinear function to the higher dimensional feature space is defined as:

$$\phi(\boldsymbol{\varphi}) = \left[\varphi_1^2 \ \sqrt{2}\varphi_1\varphi_2 \ \varphi_2^2 \right]^T.$$

It is easy to show that Eq.(17) is equivalent to Eq.(6) by defining

$$\boldsymbol{w} = [m_{11}, \sqrt{2} m_{12}, m_{22}]$$

The inner product of $\phi(\cdot)$ is

Table 1. Plant parameters

parameter	value	unit
m	2.5	[kg]
k	1	[N/m]
c_1	1.2	[Ns/m]
c_2	1.2×10^{-3}	[Ns/m]

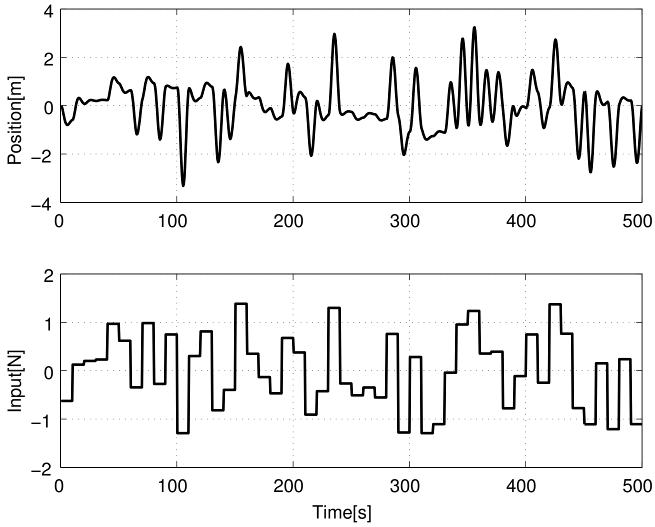


Fig. 5. Input-output data (upper: output, lower: input)

$$\begin{aligned}\langle \phi(\boldsymbol{z}), \phi(\boldsymbol{x}) \rangle &= (z_1 x_1)^2 + 2 z_1 z_2 x_1 x_2 + (z_2 x_2)^2 \\ &= \langle \boldsymbol{z}, \boldsymbol{x} \rangle^2, \end{aligned} \quad (18)$$

and it corresponds to Eq.(7) for $d = 2$. Therefore, the nonlinear discriminant function of Eq.(17) is achieved by SVM using a 2nd order polynomial kernel.

4. SIMULATIONS

4.1 A simulation model and a input-output data

In order to show the effectiveness of the proposed method, simulation results are shown for a simple spring-mass-damper system as shown in Fig. 3. The parameters are shown in Table 1. An input-output data was collected by the discrete-time model of Fig. 3 with a sampling period $T_s = 1$ [s], and it is shown in Fig. 5. The number of data is $N = 1000$ and the first half of the data is used for the estimation of model, and the second half is used for model validation. A 2nd order ARX model is used for an identification model:

$$y[k] = \begin{cases} \boldsymbol{\theta}_1^T \boldsymbol{\varphi}[k] + w[k] \ \text{(mode 1)} \\ \boldsymbol{\theta}_2^T \boldsymbol{\varphi}[k] + w[k] \ \text{(mode 2)} \end{cases} .$$

where $w[k]$ is white noise, $\boldsymbol{\varphi}[k]$ is the regression vector defined by Eq.(10), and $\boldsymbol{\theta}_i \in \mathcal{R}^4$, $i = 1, 2$ are parameter vectors of mode 1 and mode 2.

4.2 Data clustering and identification of sub-models

The data points $(\boldsymbol{\varphi}[k], y[k])$ are classified into two classes corresponding to mode 1 and mode 2 by using the method described in section 2.1. The result is shown in the upper figure of Fig. 6. The true clustering result is also shown below in Fig. 6 for comparison, and the estimated result has a good agreement with the true result.

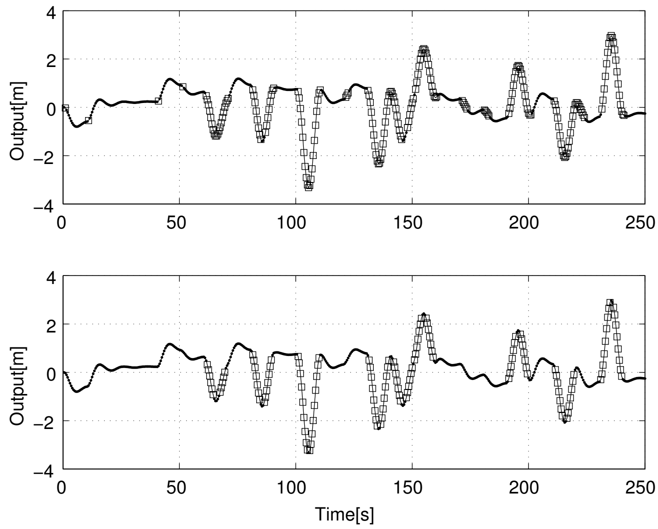


Fig. 6. Clustering result (upper: clustered output, lower: true output; dots: mode 1, squares: mode 2)

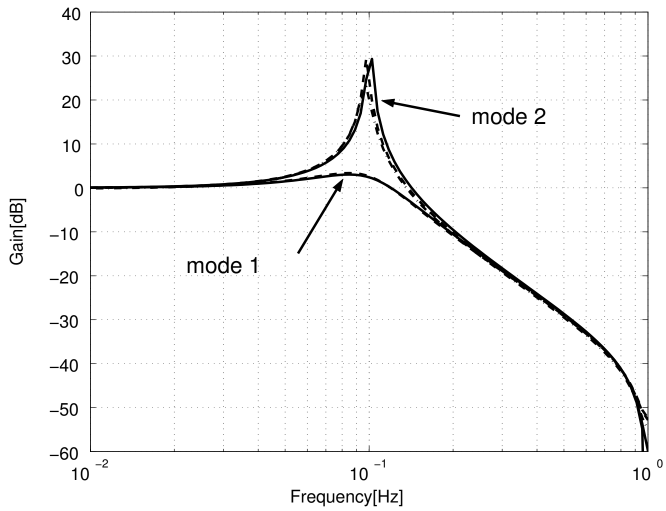
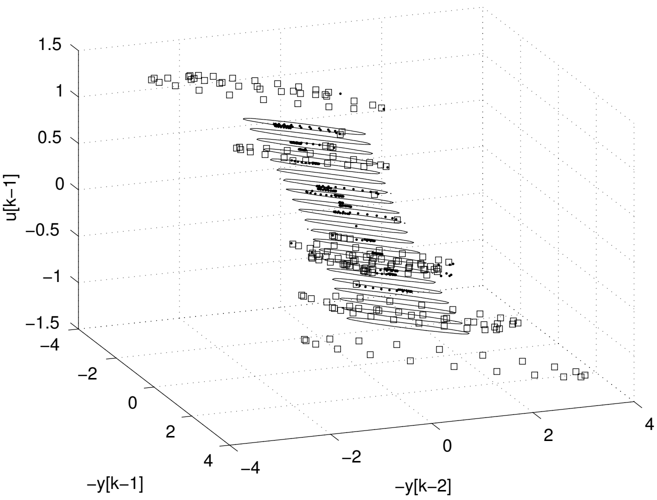


Fig. 7. Frequency responses of identified sub-models (solid line: true models of mode 1 and mode 2, dotted line: mode 1, dashed line: mode 2, dashdot line: single ARX model)

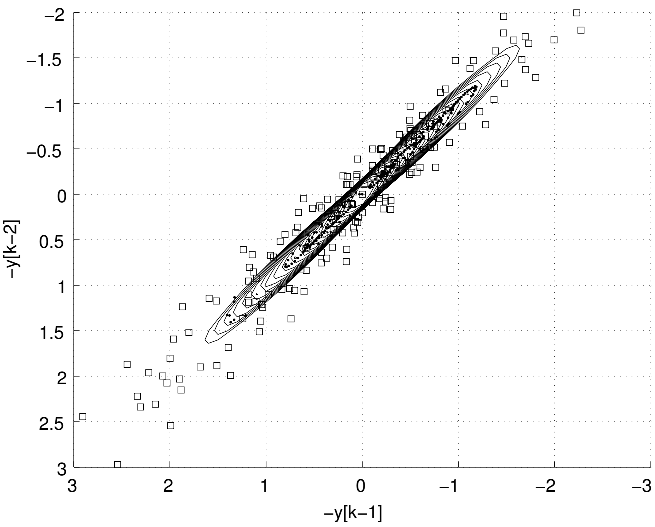
The parameter vectors $\boldsymbol{\theta}_i$ for mode 1 and mode 2 are estimated by using the least squares method. The frequency responses of the estimated ARX models are shown in Fig. 7. In this figure, the frequency responses of mode 1 and mode 2 are shown by dotted line and dashed line respectively. For comparison, the frequency responses of the true modes are also shown by solid line. The identified sub-models of mode 1 and mode 2 have a good agreements with the true sub-models. The identification result using a single ARX model is also shown by dashdot line, and it identifies mode 2 only.

4.3 Estimation of a nonlinear discriminant function

The regression space is separated by using a 2nd order polynomial kernel. In order to allow a small amount of



(a)



(b)

Fig. 8. Regression space (dots: mode 1, squares: mode 2, solid line: estimated boundary between mode 1 and mode 2)

mislabeled data points, a soft margin SVM is used. As the solver for soft margin SVM with polynomial kernel, "SVM and Kernel Methods Matlab Toolbox" is used (Canu et al., 2005).

The result is shown in Fig. 8 in which the trajectory of

$$[-y[k-1], -y[k-2], u[k-1]]^T$$

is shown from different view points because the regression vector is four dimensions and it is difficult to visualize directly. From this figure, it is confirmed that a nonlinear boundary that separates mode 1 and mode 2 is estimated correctly.

Fig. 9 shows a time response of $y[k]$ when the second half of the input data shown in Fig. 5 is applied to the identified model. The time response of the single ARX model and the hybrid model are shown by dashed line and the dashdot line respectively. The time response of the hybrid model has a good agreement with that of the true model. The fit ratio of the identified hybrid model was 89%, while that of the single ARX model was 68.2%. From these results, the effectiveness of the proposed method based on SVM with polynomial kernel is confirmed.

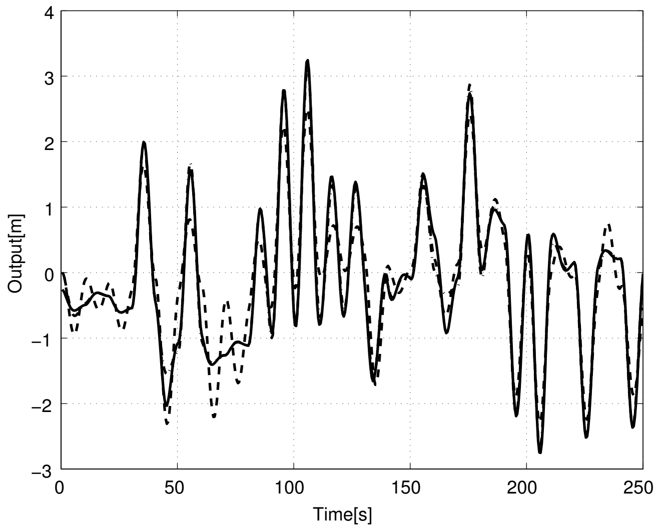


Fig. 9. Time responses of identified models (solid line: true model, dashed line: single ARX model, dashdot line: hybrid model)

5. CONCLUSION

In this paper, a system identification problem for the hybrid system switched by the magnitude of velocity has been considered. First, it was shown that the conventional method cannot be applied directly to the system. Then the identification method based on SVM with polynomial kernel has been proposed. The proposed method separates the regression space by a nonlinear discriminant function.

From the simulation results, it was shown that the obtained parameters of sub-models have good agreement with that of the true systems. The time responses were also compared with that of the true system, and it was confirmed that the fit ratio of the proposed method is improved compared with that of the conventional method based on the single ARX model.

Future investigations will focus on the implementation of the proposed method to the actual systems such as a head positioning control system of hard disk drives (HDDs). The head actuator of HDDs has a nonlinearity such as a pivot friction, and the identification of the nonlinearity is required in order to improve the positioning performance.

REFERENCES

Adachi, S. *Advanced System Identification for Control Using MATLAB.* Tokyo Denki Univ. Press, 2004. (in Japanese).

Canu, S., Y. Grandvalet, V. Guigue, and A. Rakotomamonjy. *SVM and Kernel Methods Matlab Toolbox.* `http://asi.insa-rouen.fr/~arakotom/toolbox/`, 2005

Cristianini, N., and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press, 2000.

Ferrari-Trecate, G., M. Muselli, D. Liberati, and M. Morari. A Clustering Technique for the Identification of Piecewise Affine Systems. *Automatica*, **39**–2, pp.205–217, 2003

Futami, S., A. Furutani, and S. Yoshida. Nanometer Positioning and Its Micro Dynamics. *Nanotechnology*, **1**–1, pp.31–37, 1990.

Noguchi, S., M. Hirata, and S. Adachi. Hybrid Modeling of Hard Disk Drives Based on Input-Output Data. In *Proc. of SICE Annual Conference 2005*, pp.1132–1135, 2005.