Input design for Hybrid System Identification for Accurate Estimation of Submodel Regions

Hiroshi Suzuki and Masaki Yamakita

Abstract— This paper addresses about designing of input sequence for identification of hybrid systems as PWARX model. Identification of PWARX model requires estimation of parameter vectors and estimation of discriminant surfaces of submodels in regressor space. Since we especially focus on accuracy of discriminant surface, we propose a method to determine input which generates regressor vector data nearby the discriminant surfaces by optimizing a cost function. Since the cost function requires information about discriminant surfaces, we propose a sequential identification algorithm, which updates the model and the cost function alternatively. By this method, we can obtain a high accurate model with short input sequence. Validity of the proposed method is illustrated in numerical examples.

I. INTRODUCTION

Hybrid system is a system whose behavior is defined by both continuous and discrete dynamics, which is drawing attention from various disciplines. For example, they are systems with continuous dynamics and logic operations, biological systems and approximations of nonlinear dynamics. Analysis, control approaches and other researches for hybrid systems have done in the last decades. And also methods of hybrid system identification to obtain the models are actively studied recently.

Piece-Wise Affine (PWA) system, a standard model of hybrid system, is considered in researches of hybrid system identification. Especially, Piece-Wise affine Auto Regressive eXogenous (PWARX) model, a class of PWA system, is most popular and some approaches have been proposed [1]. Those researches are discussed on that we have already obtained rich input and output (IO) data. But method for designing identification experiments, as far as we know, have not been proposed a lot, although the importance of it has been mentioned.

So in this paper, we address about a method of designing input for hybrid system identification to obtain IO data. Since we especially focus on estimation of discrete dynamics, we consider efficient and high accurate region estimation problem for PWARX model. Since the accuracy of the submodel regions is closely related with the dynamic characteristics of PWARX model such that observability [2], it is important to focus on accuracy of the region estimation. To estimate regions with high accuracy, we have to locate the regressor vector data nearby the discriminant surfaces. So we propose a cost function which has minimal value on the discriminant surfaces, and design input by optimizing the cost function.

H.Suzuki and M.Yamakita are with Dept. of Mechanical and Control Engineering, Tokyo Institute of Technology, 2-12-1, Oh-Okayama, Meguroku, Tokyo, Japan
Contact author `yamakita@ac.ctrl.titech.ac.jp`

We also propose a sequential identification algorithm to use the cost function effectively.

In the following, fist we explain about PWARX model in section II, next we address about conventional method of hybrid system identification in section III, then we propose the method of input design and sequential identification in section IV and V. We apply the proposed method to numerical simulation in section VI.Finally some concluding remarks are given in section VII.

II. PWARX MODEL

In this research, we use one of the popular class of hybrid systems, discrete PWARX model. Assume $u_k \in \mathbb{R}$, $y_k \in \mathbb{R}$ and $\eta_k \in \mathbb{R}$ are input, output and noise at a time instance k. Then we define single input single output PWARX model as

$$\begin{cases} y_k := P_i^\top \begin{bmatrix} \boldsymbol{x}_k \\ 1 \end{bmatrix} + \eta_k \quad \text{if} \quad \boldsymbol{x}_k \in I_i, \, i = 1,...,s \\ \boldsymbol{x}_k := \begin{bmatrix} y_{k-1},...,y_{k-n_y}, u_{k-1},...,u_{k-n_u} \end{bmatrix}^\top \end{cases} \quad (1)$$

where, I_i is a region of ith submodel. We also define $\boldsymbol{r}_k := [\boldsymbol{x}_k^\top, 1]^\top$ and rewrite (1) as $y_k = P_i^\top \boldsymbol{r}_k + \eta_k$ if $\boldsymbol{x}_k \in I_i, \, i = 1,...,s$.

In this paper I_i is defined by using a discriminating function $f(\boldsymbol{x})$ and decision function $g_i(\boldsymbol{x}), \, i = 1,...,s$ as

$$I_i := \{\boldsymbol{x} : f(\boldsymbol{x}) = i\} \quad (2)$$
$$f(\boldsymbol{x}) := \arg\max_i g_i(\boldsymbol{x}), \, i \in \{1,...,s\}. \quad (3)$$

From (2) and (3), the discriminant surface of regions, I_i and I_j, is a set of $\boldsymbol{x}$ such that $g_i(\boldsymbol{x}) = g_j(\boldsymbol{x})$.

A. Decision Function

If the regions of submodels are linearly separable [5], the decision function can be written as linear functions. Assume

$$g_i(\boldsymbol{x}) := \boldsymbol{w}_i^\top \boldsymbol{x} + b_i, \quad (4)$$

then the discriminant surface of I_i and I_j is a hyperplane as

$$H_{ij} : \boldsymbol{w}_{ij}^\top \boldsymbol{x} + b_{ij} = 0, \quad (5)$$
$$\boldsymbol{w}_{ij} := \boldsymbol{w}_i - \boldsymbol{w}_j, \quad b_{ij} := b_i - b_j.$$

We will treat nonlinear decision functions in section III-B.

III. HYBRID SYSTEM IDENTIFICATION

In this section, we briefly explain about methods of hybrid system identification,when we assume that IO data are given.

Generally, identification of PWARX model requires 3 STEPS : 1. data clustering, 2. estimation of the parameter vector and 3. esitimation of the discriminant surface. In this

paper, we use Sparsification Procedure [3] for STEP 1 & 2, and Multiclass Support Vector Machine [5] for STEP 3.

A. Sparsification Procedure

Sparsification procedure can cluster the data and estimate parameter vectors at the same time with a performance criteria: 1. estimation error is lower than ϵ and 2. estimated number of submodels is minimum.

Assume $P_{(k)}$ is a parameter vector at time k for data $(y_k, \boldsymbol{r}_k)$. Then we consider the following sparsification optimizing problem.

$$
\begin{aligned}
\min_{P_{(k)}, \tilde{P}} \quad & \|\{P_{(k)} - \tilde{P}\}\|_0 \\
\text{s.t.} \quad & \left| y_k - P_{(k)}^\top \boldsymbol{r}_k \right| \le \epsilon, \quad \forall k
\end{aligned}
\tag{6}
$$

where $\|\{\boldsymbol{a}\}\|_0$ denotes l_0 quasi norm of vector sequence which is defined as the number of nonzero vector in vector sequence $\{\boldsymbol{a}\} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, ...\}$. Therefore, the optimization problem (6) becomes the problem finding a parameter vector $\tilde{P}$ which meets the condition $|y - \tilde{P}^\top \boldsymbol{r}_k| \le \epsilon$ for as many k as possible. Although (6) is a NP-hard problem, we can calculate the solution approximately by reductioning it to a linear programming problem (see Lemma 1. of [3]). When optimization is done, we obtain $\tilde{P}$ as a parameter vector and cluster, and remove all data which meet the condition $\left| y_k - \tilde{P}^\top \boldsymbol{r}_k \right| \le \epsilon$. By repeating the procedure until all data are clustered, we cluster the data and estimate the parameter vectors. We can also estimate the number of submodels as the number of parameter vectors.

B. Multiclass Support Vector Machine

After clustering the data is completed, we estimate the discriminant surfaces of submodel regions in the regressor space.

Support Vector Machine (SVM)[4] is a powerful method for binary classification problem proposed by Vapnik. In this paper, we use Multiclass SVM (MSVM) extended from SVM.

Assume the regions are linearly separable for a moment and there are decision functions like (4). We define a margin $M_{H_{pq}}$ as the distance between hyperplane H_{pq} and the data nearest from H_{pq}. We can solve the problem to maximize the margins as a convex quadratic programing problem. By solving the problem, we can obtain data which have minimum margins. We call them support vectors. The decision functions are represented with support vectors as

$$
\begin{aligned}
g_p(\boldsymbol{x}) &= \sum_{\substack{q \in S, \\ q \ne p}} \left(\sum_{i \in SV_p} \alpha_{pqi} \boldsymbol{x}_{pqi}^\top \boldsymbol{x} - \sum_{i \in SV_q} \alpha_{pqi} \boldsymbol{x}_{qpi}^\top \boldsymbol{x} \right) + b_p \\
&= \boldsymbol{w}_p^\top \boldsymbol{x} + b_p
\end{aligned}
\tag{7}
$$

where, $S = \{1, .., s\}$, SV_p is a set of class p support vectors, $\boldsymbol{x}_{pqi}$ is corresponding support vector and α_{pqi} is a weight coefficient.

(M)SVM can be used for classifying nonlinear discriminant surfaces, by replacing inner products to kernel functions as

$$
g_p(\boldsymbol{x}) = \sum_{q \in S, q \ne p} \left(\sum_{i \in SV_p} \alpha_{pqi} K(\boldsymbol{x}, \boldsymbol{x}_{pqi}) \right.
$$
$$
\left. - \sum_{i \in SV_q} \alpha_{pqi} K(\boldsymbol{x}, \boldsymbol{x}_{qpi}) \right) + b_p, \quad (8)
$$

$$
K(\boldsymbol{x}_i, \boldsymbol{x}_j) := \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma_g^2} \right). \tag{9}
$$

In this research, we use gaussian kernel (9) which is a popular kernel function and the method is called kernel trick.

IV. METHOD OF INPUT DESIGN

Our purpose is estimating the discriminant surfaces of submodel regions as well as system parameters with less input sequence in order to shorten the time for identification experiment. For this purpose, we will design inputs that intensively locate the regressor vectors nearby the discriminant surfaces. To be more precise, first we design a cost function that has the minimal value at the $\boldsymbol{x}$ data on the discriminant surfaces. Then, we design the inputs by minimize the cost function. Please note that though regressor vectors are accumulated near the discriminant surfaces, SVM method will maximize the margins and give some robustness for disturbances.

A. Designing A Cost Function

In this section, we design a cost function of regressor vector $\boldsymbol{x}$ that has the minimal value at the $\boldsymbol{x}$ data on the discriminant surfaces.

1) Designing Function by Parameter Vector: We define I_i and P_i as the region and the parameter vector of the ith submodel. We also define P_{i*} as

$$
P_{i*} \quad := \quad P_i + \Delta P_i. \tag{10}
$$

Assume that P_{i*} and ΔP_i keep the following constraint:

$$
P_{i*}^\top \begin{bmatrix} \boldsymbol{x}_{c,i} \\ 1 \end{bmatrix} = P_i^\top \begin{bmatrix} \boldsymbol{x}_{c,i} \\ 1 \end{bmatrix} \tag{11}
$$

$$
\Longleftrightarrow \quad \Delta P_i^\top \begin{bmatrix} \boldsymbol{x}_{c,i} \\ 1 \end{bmatrix} = 0 \tag{12}
$$

where, $\boldsymbol{x}_{c,i}$ is a reference point of I_i. Hereafter, we consider $\boldsymbol{x}_{c,i}$ as the center point of I_i.

Now, consider a hyperplane $y = P_i^\top \boldsymbol{r}$. Then $y = P_{i*}^\top \boldsymbol{r}$ represents a hyperplane which is tilted by ΔP_i with a constraint of the center point in region I_i.

If we define

$$
\Delta P_i := \begin{bmatrix} \boldsymbol{w}_{\Delta P_i} \\ b_{\Delta P_i} \end{bmatrix}, \qquad \boldsymbol{w}_{\Delta P_i} \in \mathbb{R}^n, \quad b_{\Delta P_i} \in \mathbb{R}, \tag{13}
$$

$\boldsymbol{w}_{\Delta P_i}$ represents the tilt direction of the hyperplane.

It should be noted that the farther form $\boldsymbol{x}_{c,i}$ in the direction of $\pm \boldsymbol{w}_{\Delta P_i}$ $\boldsymbol{x}$ locates, the larger the $\left| \Delta P_i^\top \boldsymbol{r} \right| = \left| P_{i*}^\top \boldsymbol{r} - P_i^\top \boldsymbol{r} \right|$ become.

Therefore, if we choose ΔP_i as a vector that directs to the discriminant surface, we can say that $\boldsymbol{x}$ which increases $\left| \Delta P_i^\top \boldsymbol{r} \right|$ is near the discriminant surface.

2) Parameter Design for Tilt in the Direction to the Discriminant Surfaces: Assume that discriminant surfaces are linear. From (5), hyperplane H_{ij} discriminating ith and jth regions is written as an equation $\boldsymbol{w}_{ij}^\top \boldsymbol{x} + b_{ij} = 0$. Now we design a vector ΔP_{ij} which tilts $P_i^\top \boldsymbol{r}$ in the direction to H_{ij}. Let $\boldsymbol{d}_{ij}$ be a unit direction vector to the discriminant surface as

$$\boldsymbol{d}_{ij} := \frac{\boldsymbol{w}_{ij}}{\|\boldsymbol{w}_{ij}\|}. \tag{14}$$

Since there is a constraint (12), we define

$$\Delta P_{ij} := \lambda_{ij} \begin{bmatrix} \boldsymbol{d}_{ij} \\ -\boldsymbol{d}_{ij}^\top \boldsymbol{x}_{c,i} \end{bmatrix} \tag{15}$$

where λ_{ij} is a scalar coefficient. For convenience, we set λ_{ij} that $\Delta P_{ij}^\top \boldsymbol{r} = -a_{ij}$ when $\boldsymbol{x}$ is on the hyperplane (5), and we have

$$\lambda_{ij} = a_{ij} \frac{\|\boldsymbol{w}_{ij}\|}{\left|\boldsymbol{w}_{ij}^\top \boldsymbol{x}_{c,i} + b_{ij}\right|}. \tag{16}$$

Now by using the parameter vector (15), we get a function $\Delta P_{ij}^\top [\boldsymbol{x}^\top, 1]^\top$ whose value decreases as $\boldsymbol{x}$ approaches to the discriminant surface.

3) Design the Cost Function: By using ΔP_{ij}, we define a piecewise linear cost function $J_\ell(\boldsymbol{x})$ as

$$J_\ell(\boldsymbol{x}) := \Delta P_i^\top(\boldsymbol{x}) \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} \quad \text{if } \boldsymbol{x} \in I_i, \ i = 1, ..., s \tag{17}$$

where

$$\Delta P_i(\boldsymbol{x}) := \arg\min_{\Delta P_{ij}} \left(\Delta P_{ij}^\top \begin{bmatrix} \boldsymbol{x} \\ 1 \end{bmatrix} \right), \quad j \in C_i \tag{18}$$

$$\Delta P_{ij} := \lambda_{ij} \begin{bmatrix} \boldsymbol{d}_{ij} \\ -\boldsymbol{d}_{ij}^\top \boldsymbol{x}_{c,i} \end{bmatrix} \tag{19}$$

$$\boldsymbol{d}_{ij} = \frac{\boldsymbol{w}_{ij}}{\|\boldsymbol{w}_{ij}\|}, \quad \lambda_{ij} = a_{ij} \frac{\|w_{ij}\|}{\left|\boldsymbol{w}_{ij}^\top \boldsymbol{x}_{c,i} + b_{ij}\right|}$$

$$a_{ij} = a_{ji}, \quad j \in C_i$$

$$C_i \ : \ \text{index set of the regions neighboring } I_i.$$

C_i is an index set of the regions neighboring I_i. For example, if I_i neighbors I_2 and I_3, $C_1 = \{2, 3\}$. By switching the parameter ΔP_{ij} according to (17) and (18), $J_\ell(\boldsymbol{x})$ has $-a_{ij}$ as a minimal value when $\boldsymbol{x}$ is on the discriminant surface H_{ij}.

4) Extension to Nonlinear Surface Estimation: Next we extend the function developed in section IV-A.3 for estimation of nonlinear discriminant surfaces. Referring (7) and (8), by replacing inner product to the kernel function of $\Delta P_{ij}^\top [\boldsymbol{x}^\top, 1]^\top$ and with simple calculation, we can get a nonlinear function $p_{ij}(\boldsymbol{x})$ as

$$p_{ij}(\boldsymbol{x}) = a_{ij} \left(\frac{g_i(\boldsymbol{x}) - g_j(\boldsymbol{x})}{g_i(\boldsymbol{x}_{c,i}) - g_j(\boldsymbol{x}_{c,i})} - 1 \right). \tag{20}$$

Following is a new cost function $J(\boldsymbol{x})$ which is derived from (17) for nonlinear surface estimation.

$$J(\boldsymbol{x}) := p_i(\boldsymbol{x}) \quad \text{if} \quad \boldsymbol{x} \in I_i, i = 1, ..., s \tag{21}$$

$$p_i(\boldsymbol{x}) := \min_j (p_{ij}(\boldsymbol{x})), \quad j \in C_i \tag{22}$$

$$p_{ij}(\boldsymbol{x}) := a_{ij} \left(\frac{g_i(\boldsymbol{x}) - g_j(\boldsymbol{x})}{g_i(\boldsymbol{x}_{c,i}) - g_j(\boldsymbol{x}_{c,i})} - 1 \right) \tag{23}$$

$$a_{ij} = a_{ji} > 0, \quad j \in C_i$$

$$C_i \ : \ \text{index set of the regions neighboring } I_i.$$

$J(\boldsymbol{x})$ has also minimal value when $\boldsymbol{x}$ is on the discriminant surface. Since $J(\boldsymbol{x})$ can represent the cost function for linear surfaces by expressing $g_i(\boldsymbol{x})$ as (4), it can be said as a general form.

When we estimate nonlinear discriminant surfaces by MSVM, the support vectors must spread well for good estimation. So we consider the following cost function $J_{var}(\boldsymbol{x})$.

$$J_{var}(\boldsymbol{x}) := c \sum_{j \in C_i} \sum_{p \in SV_{ij}} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}_{jp}\|^2}{2\sigma^2} \right)$$
$$\text{if} \quad \boldsymbol{x} \in I_i \tag{24}$$

where, SV_{ij} is a set of support vectors of class j and p, $\boldsymbol{x}_{jp}$ is a support vector. c and σ are constant values. By choosing $\boldsymbol{x}$ to minimize $J_{var}(\boldsymbol{x})$, we can expect that support vector data will spred well.

With (21) and (24), the cost function for nonlinear discriminant surface estimation is defined as

$$J_n(\boldsymbol{x}) := J(\boldsymbol{x}) + J_{var}(\boldsymbol{x}). \tag{25}$$

B. Designing Input

We determine a input value u_k by solving the following optimization problem:

$$\begin{aligned} \min \quad & J_\ell(\boldsymbol{x}_{k+1}) \\ \text{s.t.} \quad & u_{min} \le u_k \le u_{max} \end{aligned} \tag{26}$$

where, u_{max} and u_{min} are upper bound and lower bound. By using u_k which optimize (26), we will obtain the data $\boldsymbol{x}_{k+1}$ nearby the discriminant surface.

V. SEQUENTIAL IDENTIFICATION

As in section IV, the cost function needs information of discriminant surfaces of the PWARX model. To use the cost function effectively, we propose a sequential identification method which improve the model and cost function alternatively. In the identification process, we focus on the margin values as an index of accuracy of estimation though the other parameters are determined simultaneously.

A. Margin of SVM estimation

In (M)SVM estimation, a discriminant surface of region i and j mast be between class i data and class j data. So, if the margin value is large, the reliability of estimation is low. If the margin value is small, we can expect the reliability of estimation is high contrary.

B. Algorithm of Sequential Identification

First let us assume that submodel regions are linearly separable. In the following algorithm, the model is improved almost evely input step to the system. The algorithm ends when all margin of the discriminant surfaces is lower than threshold and we obtain the parameter vectors $\hat{P}_i$, $i = 1, ..., s$ and discriminant surface parameter $(\hat{w}_i, \hat{b}_i)$, $i = 1, ..., s$. Also in the following, subscripts i and j represent the all $i = 1, ..., s$ and $j \in C_i$.

0) **Initial parameter identification**
   a) Do hybrid system identification with short random input sequence and obtain $\left(P_i^{(0)}, (w_i^{(0)}, b_i^{(0)}), (w_{ij}^{(0)}, b_{ij}^{(0)}) \right)$.
   b) Define the set of class i support vectors as $SV_i^{(0)}$. Define the set of class i (y, x) data as $D_i^{(0)}$.
   c) Estimate $x_{c,i}$.
   d) Set $k = 0$.

1) **Checking margin values**
   a) Calculate the maximum of the margin values $M_{max}^{(k)}$
   $M_{max}^{(k)} = \max \left(M_{H_{ij}^{(k)}} \right)$ ($M_{H_{ij}^{(k)}}$: value of the margin of the discriminant surface $H_{ij}^{(k)}$).
   b) If $M_{max}^{(k)} \leq \bar{M}$, then $\left(\hat{P}_i, (\hat{w}_i, \hat{b}_i) \right) = \left(P_i^{(k)}, (w_i^{(k)}, b_i^{(k)}) \right)$, $k_f = k$ and quit the algorithm.

2) **Constructing cost function**
   a) Calculate $a_{ij}^{(k)}$ $a_{ij}^{(k)} = a_0 M_{H_{ij}}^{(k)}$.
   b) Construct the cost function $J_\ell^{(k)}(x)$ according to (17) with $\left((w_i^{(k)}, b_i^{(k)}), (w_{ij}^{(k)}, b_{ij}^{(k)}), a_{ij}^{(k)} \right)$.

3) **Obtaining data**
   Set up input u_k by optimizing $J_\ell^{(k)}(x)$ according to (26). Then obtain (y_{k+1}, x_{k+1}).

4) **Data clustering and updating the parameter vectors**
   a) Obtain $\left(P_i^{(k+1)}, D_i^{(k+1)} \right)$ by using the data (y_{k+1}, x_{k+1}) and $D_i^{(k)}$.
   b) If number of submodels is changed, then re-estimate $x_{c,i}$.

5) **Updating the discriminant surfaces**
   Estimate the discriminant surfaces $\left((w_i^{(k+1)}, b_i^{(k+1)}), (w_{ij}^{(k+1)}, b_{ij}^{(k+1)}) \right)$, by using $\forall x \in SV_i^{(k)}$ and x_{k+1}.

6) **Updating $SV_i^{(k)}$**
   $SV_i^{(k+1)} = SV_i^{(k)}$.
   If x_{k+1} is selected as Support Vector at STEP 5) ,then store x_{k+1} into $SV_i^{(k+1)}$.

7) $k = k + 1$ and return to STEP 1).

When we treat nonlinear discriminant surfaces, just replace $J_\ell^{(k)}(x)$ to $J_n^{(k)}(x)$.

C. Convergence of Algorithm

$J_\ell^{(k)}(x)$ has a minimal value on the discriminant surface estimated at kth step. Therefore, if we can solve the opti-

mization problem (26), margin value should decrease. But depending of the characteristic of the systems, sometimes input sequence can be constant value. To avoid this situation, we should use random input instead of optimizing input at a rate of 10 % in the algorithm in practice.

D. Estimation of Region Center Point $x_{c,i}$

$x_{c,i}$ is a reference point which make the cost function value 0. So $x_{c,i}$ need not to be an accurate center of the region I_i. It only need to be a point which stay well off the discriminant surfaces. Therefore, if the nonlinearity of the discriminant surface is weak, 1. define $x_{c,i}$ as arithmetic average of the data, or 2. if kernel function is gaussian kernel, define as the solution of the following optimization problem

$$\begin{array}{ll} \max_x & g_i(x) - \max_{j \in C_i} g_j(x) \\ \text{s.t.} & x \in I_i \end{array} \quad . \tag{27}$$

In the section VI, we apply former method to Example 1. and latter method to Example 2. If the nonlinearity of the discriminant surface is strong, it can be applied EM Algorithm [6] to define $x_{c,i}$ from the data distribution.

E. Weight Coefficient $a_{ij}^{(k)}$

The weight coefficient $a_{ij}^{(k)}$ of the cost function $J^{(k)}(x)$ is linear to the margin values $M_{H_{ij}^{(k)}}$ of each discriminant surface $H_{ij}^{(k)}$. Therefore, it takes heavier weight to the dicriminant surface which margin value is larger.

a_0 is a positive constant value. It should be 100 to 10000 times of $\bar{M}$ according to our simulations.
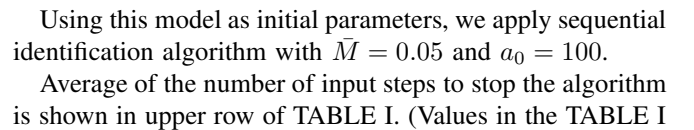
VI. NUMERICAL SIMULATION

Now we apply the proposed method to the numerical examples to verify the effectiveness. We use MATLAB for simulation.

A. Example 1. (Linear Discriminant Surface)

We apply a PWARX model whose parameters are $n_y = n_u = 1$, $x = [y_{k-1}, u_{k-1}]^\top \in \mathbb{R}^2$, $s = 3$.

$$\begin{cases} P_1 = \begin{bmatrix} 0.7, & 1.0, & -0.5 \end{bmatrix}^\top \\ P_2 = \begin{bmatrix} 0.2, & -0.5, & 2.0 \end{bmatrix}^\top \\ P_3 = \begin{bmatrix} 0.5, & -0.1, & 0 \end{bmatrix}^\top \end{cases} \tag{28}$$

We define a PWARX model **sys1** as the parameter vectors are (28) and the discriminant surfaces are shown as linear broken lines in Fig.1. Noise is set as $\eta_k = 0$ and upper and lower bound of input are $u_{max} = 2$, $u_{min} = -2$.

For initial parameter identification, with 15 steps IO data, setting $\epsilon = 0.02$ of Sparsification Procedure, then we obtained estimated discriminant surfaces shown as solid lines in Fig.1. Fig.1 also shows support vectors as '$\bigcirc$' and $x_{c,i}$ as '$+$'.

Using this model as initial parameters, we apply sequential identification algorithm with $\bar{M} = 0.05$ and $a_0 = 100$.

Average of the number of input steps to stop the algorithm is shown in upper row of TABLE I. (Values in the TABLE I
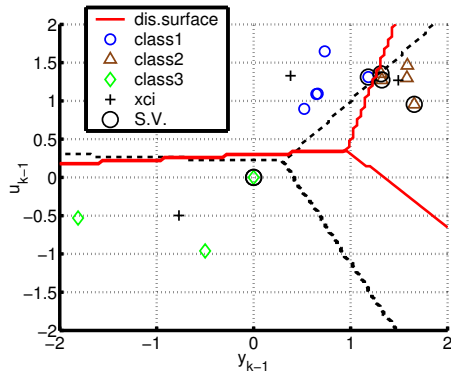
Fig. 1. Discriminant Surface Estimated by initial parameter identification (for **sys1**)

are average and sample variance of 200 times simulations.) In the case that we use random inputs instead of optimized input in STEP 3) in the sequential identification algorithm, the average of number of input steps to stop the algorithm is shown in lower row of TABLE I. The result shows that the optimized input decreased the margins effectively.

TABLE I
LENGTH OF INPUT SEQUENCE (FOR **sys1**)

Input Design	ave.	var.
Optimizing $J_\ell(\boldsymbol{x})$	34.1	220.1
Random	315.7	9611.3

Example of the initial cost function value $J_\ell^{(0)}(\boldsymbol{x})$ and final cost function value $J_\ell^{(k_f)}(\boldsymbol{x})$ are shown in Fig.2 and Fig.3. According to the contour plot in the figures, ti can be seen that finally the cost function become a function which has minimal values on the true discriminant surfaces.
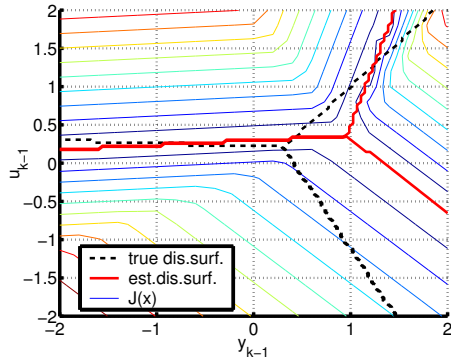

Fig. 2. Initial Cost Function $J_\ell^{(0)}(\boldsymbol{x})$ (for **sys1**)

The final estimated disiciminant surfaces and support vectors are plotted in Fig.4. It shows that we obtain a high accurate model by the proposed method.

To evaluate the identified models, we calculate RMSE of the estimated output values for 50-step test input sequence. We test **(a)** the model of initial parameter identification, **(b)** the final models of sequential identification, **(c)** the models identified by conventional identification method with IO data
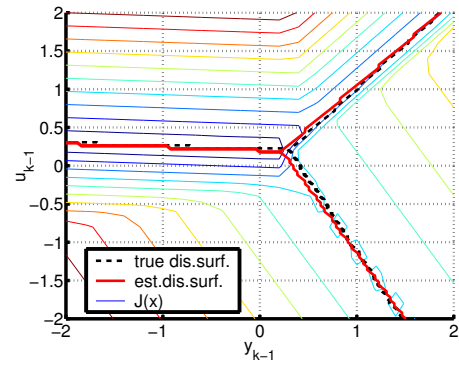

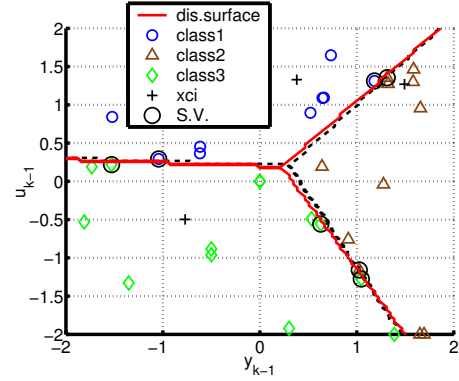Fig. 3. Final Cost Function $J_\ell^{(k_f)}(\boldsymbol{x})$(for **sys1**)

Fig. 4. Discriminant Surface Estimated by sequential Identification (for **sys1**)

of 50 random inputs. Results are shown in TABLE II. (Values in the TABLE I are average and sample variance of 200 times simulations.)

TABLE II
RMSE (FOR **sys1**)

Model	ave.	var.
(a) Initial model of proposed method	0.98	——
(b) Final model of proposed method	0.12	0.0072
(c) Model of conventional method (50 data)	0.29	0.057

Number of input steps in the initial parameter identification is 15, and the average number of input steps in the sequential identification is 34.1. So the average number of input steps that the proposed method requires is 49.1.By comparing to the results by the conventional method using 50-step input, it can be seen that accuracy of the models of proposed method are higher than that of the models of conventional identification method. It shows that the proposed method can identify high accurate model with short input sequence.

B. Example 2. (Nonlinear Discriminant Surface)

We define a PWARX model **sys2** as the parameter vectors are (28), the discriminant surface are shown as curved broken lines in Fig.5 and other parameters are the same as **sys1**.

The parameter of $J_n(\boldsymbol{x})$ are set as $c = \min\{a_{ij}^{(k)}\}$,

$i = 1, ..., s, j \in C_i, \sigma = 0.2$. The other condition are set as in Example 1. Then, the simulation result are shown in Fig.5-8 and TableIII-IV. Result shows that the proposed method work effectively to the PWARX models which has nonlinear discriminant surfaces as well.
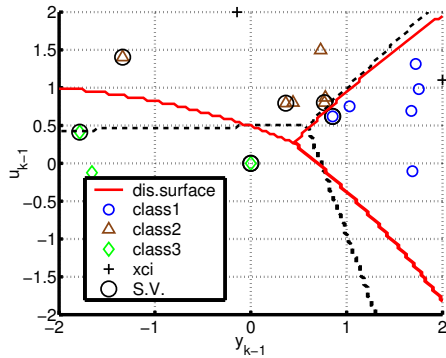

Fig. 5.   Discriminant Surface Estimated by initial parameter identification (for **sys2**)

TABLE III

LENGTH OF INPUT SEQUENCE (FOR **sys2**)

Input Design	ave.	var.
Optimizing $J_n(\boldsymbol{x})$	130.9	5430.9
Random	386.2	9320.5

TABLE IV

RMSE (FOR **sys2**)

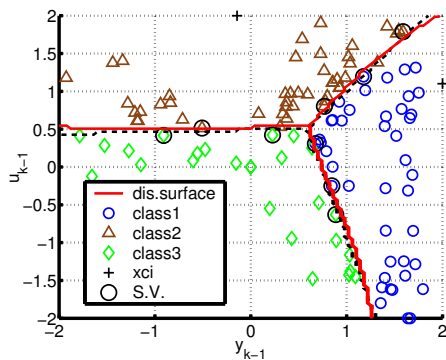| Model	ave.	var.
(a) Initial model of proposed method	0.49	—
(b) Final model of proposed method	0.11	0.014
(c) Model of conventional method (150 data)	0.29	0.048

Fig. 6.   Discriminant Surface Estimated by sequential Identification (for **sys2**)

VII. CONCLUSION

In this paper, we discussed about input design for hybrid system identification as PWARX model by focusing on accuracy of estimation of submodel regions. We have to locate the regressor vector data nearby the discriminant surfaces for high-accuracy estimation of the surface in the regressor
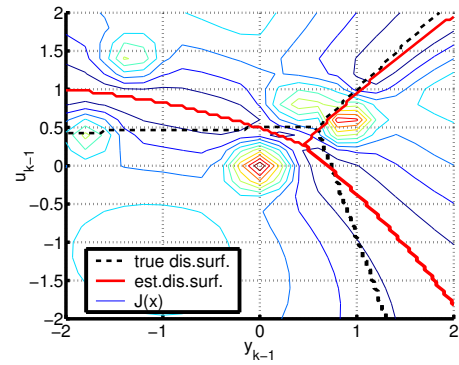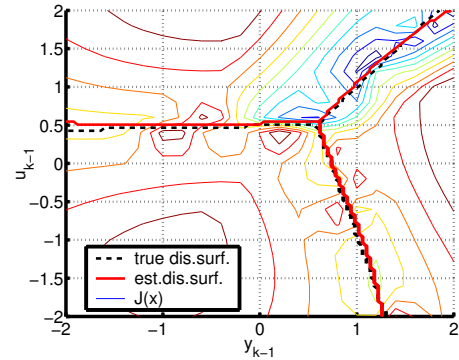

Fig. 7.   Initial Cost Function $J_n^{(0)}(\boldsymbol{x})$ (for **sys2**)

Fig. 8.   Final Cost Function $J_n^{(k_f)}(\boldsymbol{x})$(for **sys2**)

space. So we proposed a cost function which has minimal values on the discriminant surfaces, and designed input by minimizing the cost function. Since the cost function needs information about the discriminant surface, we proposed a sequential identification algorithm to use the function effectively. In the algorithm, parameter of the model and cost function are improved almost every step of input. We applied the proposed method to 1 degree PWARX models with linear and nonlinear discriminant surfaces, and it was shown that our method worked effectively.

As future study, we are considering to apply the method to high dgree PWARX models and models as approximation of nonlinear dynamics.

REFERENCES

[1] S. Paoletti, A.L. Juloski, G. Ferrari-Terecate and R. Vidal, "Indentification of Hybrid Systems A Tutorial," *European Journal of Control*, vol.13, pp.242-260, 2007

[2] M. Babbali and M. Egerstedt, "On the Observability of Piecewise Linear Systems," *43th IEEE Conference on Decision and Control*, pp.26-31, 2004

[3] N. Ozay, M. Szainer, C. Lagoa and O. Camps, "A Sparsification Approach to Set Membership Identification of a Class of Affine Hybrid Systems," *47th IEEE Conference on Decision and Control*, pp.123-130, 2008

[4] Vladimir N. Vapnik, *Statistical Learning Theory*, A Wiley-Interscience Publication, 1998

[5] K. Tatsumi, K. Hayashida, H. Higashi and T. Tanino, "Multi-objective Multiclass Support Vector Machine for Pattern Recognition," *SICE Annual Conference 2007*, pp.1095-1098, 2007

[6] J. A. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," *Technical Report TR-97-021, International Computer Science Institute and Computer Science Division, University of California, Berkeley*,1998