2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

# Identification of PWA models via data compression based on $\ell_1$ optimization

Ichiro Maruta and Toshiharu Sugie

***Abstract*— In this paper, a new identification method for piecewise affine (PWA) models is introduced. The method is based on the data-based representation of PWA maps and the data compression with $\ell_1$ optimization technique, which enable the method to deal with large data sets. This method can be applied to a wide range of modeling problems, and an example with a DC motor system is shown in this paper to show the usability of the method.**

## I. INTRODUCTION

Modeling of the target system is an important stage in designing control systems, and a lot of work has been done on this research topic [1], [2]. Especially, LTI (linear time-invariant) models are commonly used since they are easy for analysis and useful in designing controllers. However, many realistic systems have non-linearity, such as friction [3] for mechanical systems and non-linearity caused by changing operating points in plant systems. Hence, more powerful modeling schemes are required.

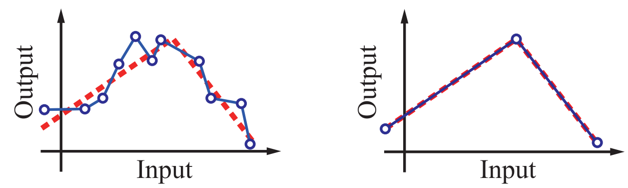
One attractive approach for this problem is to use PWA (Piece-wise affine) model, which has PWA map as substitute for standard linear map. The domain of PWA map is partitioned into several modes, and complex maps can be represented by switching among affine maps according to the mode. Since PWA map has universal approximation properties [4], [5], it is expected to have broad utility for modeling nonlinear systems, and there are many papers related to this approach (see [6] and references therein). In obtaining PWA models from I/O data, a common problem is that it is hard to estimate mode transitions in the I/O data and subsystems simultaneously. Thus, many approaches are based on the prior knowledge about the mode transition rule and the methods which do not require such knowledge tend to be computationally expensive (e.g., in [7], the estimation problem is reduced to mixed-integer programming, and numerical examples with a few data points are shown).

On the other hand, the $\ell_1$ optimization technique has attracted much attention due to its capability to provide good approximation for $\ell_0$ optimization problems, which are computationally intractable, and new techniques related to $\ell_1$ optimization, such as compressed sensing [8], are

I. Maruta is with Department of Systems Science, Graduate School of Informatics, Kyoto University, Uji, Kyoto 611-0011, Japan maruta@robot.kuass.kyoto-u.ac.jp

I. Maruta is a Research Fellow of the Japan Society for the Promotion of Science

T. Sugie is with Department of Systems Science, Graduate School of Informatics, Kyoto University, Uji, Kyoto 611-0011, Japan sugie@i.kyoto-u.ac.jp




(a) Data-based PWA model directly built from measured data, which fluctuated by stochastic noise.

(b) Ideal data-based PWA model for the example system.

Fig. 1: Two types of data-based PWA models for an example system ($x$: Input; $y$: Output). Dashed and solid lines illustrates the system and the model, respectively.

recently developed. As for the studies utilizing $\ell_1$ optimization method in the area of PWA model identification, segmentation of I/O data according to the system mode transition is studied by Ohlsson and others [9]. Also, the authors have shown the effectiveness of the $\ell_1$ based method for the modeling of a DC motor system [10]. However, these studies utilize only the sparseness of the mode transition in the I/O data sequence and does not concern with the mode partition in the domain of the map.

Thus, the objective of this paper is to propose a new computationally efficient method for constructing PWA models, which is able to capture the mode partition in the domain of the map without the prior information. To propose the new method, a data-based PWA model is introduced at first. The introduced model determines its output by interpolating some given data set, and by using the data set consists of I/O data of the target system, we can compose the PWA model of the system. Fig. 1a shows an example of such PWA model. In this figure, the dashed line shows the relationship between the input and the output of the target system; the dots shows the data set which is measured from the target system with stochastic noise; and the solid line shows the I/O relationship of the PWA model based on the data set. On the other hand, the best model for this system, which is shown in Fig. 1b, can be composed with the data set which has only three data points, thus the model shown in Fig. 1a seems to have excessive data points and complexity. From this aspect, the procedure for obtaining a model with appropriate complexity from a given measured data set, i.e., identification, can be regarded as the procedure for obtaining a small number of essential data (like Fig. 1b) from a large number of measured data (like Fig. 1a), and this procedure can be regarded as data compression. Based on this idea, we introduce a new measure

978-1-61284-799-3/11/$26.00 ©2011 IEEE

of complexity and propose the method for compressing data set according to the measure. We then can compose PWA model with appropriate complexity based on the compressed data set obtained from measured I/O data. The proposed compression procedure is reduced to an $\ell_1$ optimization problem, which can be solved efficiently, and the method can be applied to systems with multiple dimensional inputs. Thus, the proposed method can be an effective approach for problems with large data set and complicated systems, which can not be handled by the existing methods.

This paper is organized as follows: At first, we introduce the data-based PWA map in Section II, this presentation of PWA map is essential to introduce the new identification scheme. Then, the identification problem is defined in detail in Section III, and the proposed method is described in Section IV. Section V contains an experiment with a DC motor system, which illustrates the effectiveness of the proposed method, and finally Section VI concludes the paper.

For conciseness, we denote the set $\{x_1, x_2, \cdots, x_N\}$ by $\{x_k\}_{k=1}^N$, and let $P$ be a set of points in Euclidean space, we define: $\mathrm{Co}(P)$ as the convex hull of $P$; $\mathcal{DT}(P)$ as Delaunay triangulation of $P$; $\mathcal{D}_{\mathrm{nbr}}(P, S)$ as the set of the neighborhoods of Delaunay simplex $S$ in Delaunay triangulation of $P$. Here,the neighborhoods are defined as the points shared by $S$ and adjoining simplexes.

## II. DATA-BASED PWA MAP

In this section, we introduce the data-based PWA map before describing the main problem. The output of this map is determined by interpolating some data set, and its definition is straightforward as shown in Fig. 1 if its input is a scalar. However, the definition of data-based PWA map with multiple dimensional input is not straightforward and described in detail.

At first, the domain of the map considered here is $d$-dimensional Euclidean space, and we denote the vector in this space by $\boldsymbol{x}$ with suffixes. Also, the range of the map is $\mathbb{R}$ and the corresponding scalars are denoted by $y$ with suffixes. Next, consider the data set

$$D \triangleq \left\{ \begin{pmatrix} \boldsymbol{x}_D^1 \\ y_D^1 \end{pmatrix}, \begin{pmatrix} \boldsymbol{x}_D^2 \\ y_D^2 \end{pmatrix}, \ldots, \begin{pmatrix} \boldsymbol{x}_D^{N_D} \\ y_D^{N_D} \end{pmatrix} \right\} \tag{1}$$
$$\left( \boldsymbol{x}_D^1, \boldsymbol{x}_D^2, \ldots, \boldsymbol{x}_D^{N_D} \in \mathbb{R}^d, \quad y_D^1, y_D^2, \ldots, y_D^{N_D} \in \mathbb{R} \right),$$

which is composed of the $N_D$ pair of the I/O data; and denote the map based on the set $D$ by

$$\Pi_D : \mathrm{Co}\left( \left\{ \boldsymbol{x}_D^k \right\}_{k=1}^{N_D} \right) \subset \mathbb{R}^d \mapsto \mathbb{R}. \tag{2}$$

Also, we define the linear interpolation of $y$ for $\boldsymbol{x}^0$ from the data set $\left\{ \begin{pmatrix} \boldsymbol{x}^k \\ y^k \end{pmatrix} \right\}_{k=1}^{d+1}$ as

$$\mathrm{Lerp}\left( \boldsymbol{x}^0, \left\{ \begin{pmatrix} \boldsymbol{x}^k \\ y^k \end{pmatrix} \right\}_{k=1}^{d+1} \right) \triangleq \begin{bmatrix} \boldsymbol{x}^0 \\ 1 \end{bmatrix}^T \cdot \begin{bmatrix} \left(\boldsymbol{x}^1\right)^T & 1 \\ \left(\boldsymbol{x}^2\right)^T & 1 \\ \vdots & \vdots \\ \left(\boldsymbol{x}^{d+1}\right)^T & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^{d+1} \end{bmatrix} \tag{3}$$

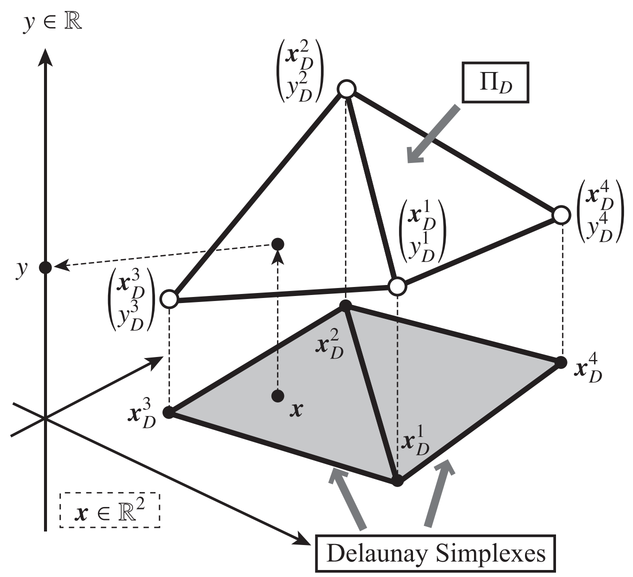



Fig. 2: Illustration of data-based PWA map $\Pi_D$ ($d = 2$, $N_D = 4$)

for conciseness. Then, the definition of the data-based PWA map is as follows:

*Definition 1 (Data-based PWA map):* For given data set $D$, we define the PWA map based on this set

$$\Pi_D : \mathrm{Co}\left( \left\{ \boldsymbol{x}_D^k \right\}_{k=1}^{N_D} \right) \subset \mathbb{R}^d \mapsto \mathbb{R} \tag{4}$$

as the map whose value $\Pi_D(\boldsymbol{x})$ is calculated by the following procedure [P1]–[P2]:

[P1] Choose the vertexes $\left\{ \boldsymbol{x}_D^{v_k} \right\}_{k=1}^{d+1}$ of the Delaunay simplex which includes $\boldsymbol{x}$ from Delaunay triangulation $\mathcal{DT}\left( \left\{ \boldsymbol{x}_D^k \right\}_{k=1}^{N_D} \right)$. (For Delaunay triangulation, see Remark 1).

[P2] Determine $\Pi_D(\boldsymbol{x})$ by interpolating the data chosen in [P1]

$$\mathrm{Lerp}\left( \boldsymbol{x}, \left\{ \begin{pmatrix} \boldsymbol{x}_D^{v_k} \\ y_D^{v_k} \end{pmatrix} \right\}_{k=1}^{d+1} \right). \tag{5}$$

■

By this definition, the value of $\Pi_D$ is determined by linearly interpolating the data set, and it becomes PWA map. Fig. 2 illustrates the relationship between the data set and the map for $d = 2$.

*Remark 1:* Delaunay triangulation is a triangulation, where the convex hull of a set of points is partitioned into triangles (simplexes). The interior of the circumcircle of any triangle in Delaunay triangulation contains no points of the set (see Fig. 3), and Delaunay triangulation also can be extended to higher dimensions. For more information about Delaunay triangulation, see, e.g., [11].

## III. PROBLEM DESCRIPTION

Next, the problem considered in this paper is described in detail. Here, the target system is assumed to be described by

$$y = f(\boldsymbol{x}) + \eta, \tag{6}$$

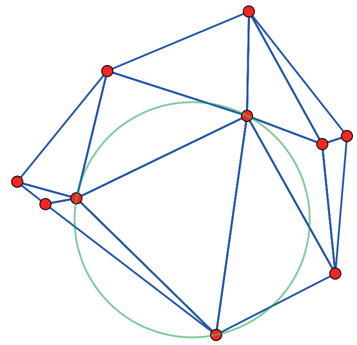


Fig. 3: An example of Delaunay triangulation for set $P$ of points in the plane. No point in $P$ is inside the circumcircle of any triangle.

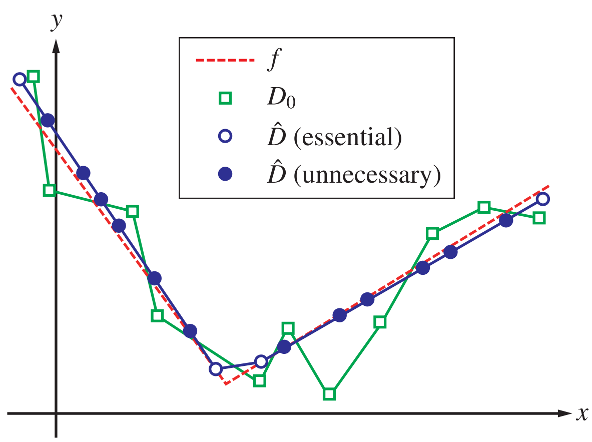



Fig. 4: Illustration of the compressed data set $\hat{D}$. If $\Pi_{\hat{D}}$ is simple PWA map, most of the data in $\hat{D}$ are unnecessary.

where $\boldsymbol{x} \in \mathbb{R}^d$ is the known input of the system; $y(t) \in \mathbb{R}$ is the measurable output of the system; and $f : \mathbb{R}^d \mapsto \mathbb{R}$ is an unknown map. Although $f$ is unknown, it is assumed that $f$ is a PWA map with a finite mode, and $\eta \in \mathbb{R}$, which is defined as the difference between $f(x)$ and $y$, can be regarded as stochastic noise.

The objective here is to obtain a PWA map which models $f$ from the given set of $N_{D_0}$ data

$$D_0 = \left\{ \begin{pmatrix} \boldsymbol{x}_{D_0}^1 \\ y_{D_0}^1 \end{pmatrix}, \begin{pmatrix} \boldsymbol{x}_{D_0}^2 \\ y_{D_0}^2 \end{pmatrix}, \dots, \begin{pmatrix} \boldsymbol{x}_{D_0}^{N_{D_0}} \\ y_{D_0}^{N_{D_0}} \end{pmatrix} \right\} \tag{7}$$

whose data satisfy the relationship (6).

More concretely, the problem here is to determine $\left\{ y_{\hat{D}}^k \right\}_{k=1}^{N_{\hat{D}}}$ to let $\Pi_{\hat{D}}$ be a good model of $f$ while assuming that $D_0$ and $\left\{ \boldsymbol{x}_{\hat{D}}^k \right\}_{k=1}^{N_{\hat{D}}}$ are given. Here, we also assume that the given set $\left\{ \boldsymbol{x}_{\hat{D}}^k \right\}_{k=1}^{N_{\hat{D}}}$ is large compared to the complexity of $f$, i.e., most of the Delaunay simplexes in $\mathcal{DT}\left( \left\{ \boldsymbol{x}_{\hat{D}}^k \right\}_{k=1}^{N_{\hat{D}}} \right)$ do not include the mode boundary of $f$. Under this assumption, most elements in $\hat{D}$ are unnecessary when $\Pi_{\hat{D}}$ is close to $f$, and $\hat{D}$ is essentially compressed for such case. Fig. 4 illustrates this aspect with an example where $d = 1$, and note that $N_{\hat{D}}$ does not need to be fewer than $N_{D_0}$.

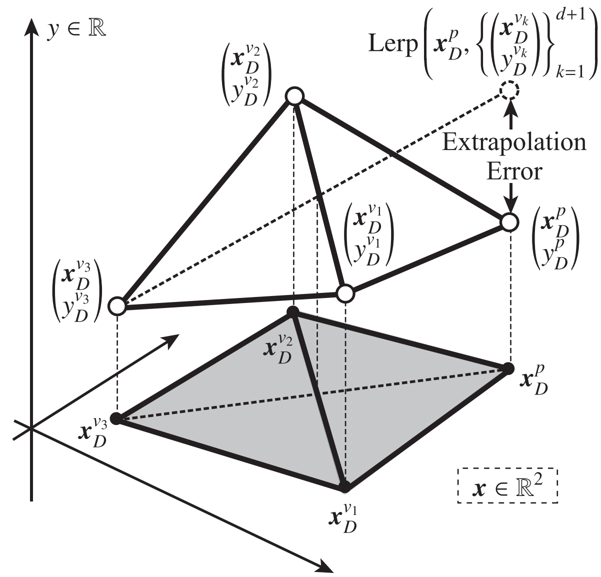



Fig. 5: Illustration of extrapolation error ($d = 2$)

## IV. IDENTIFICATION BASED ON DATA COMPRESSION

In this section, we introduce a new method which solves the problem described in Section III. The method is based on the simplicity of $f$ and is able to obtain simple and useful PWA model.

### A. Measure of Model Complexity

First, a new measure of model complexity, which is essential to propose the compression-based identification method, is introduced. To make the following compression process computationally tractable, the measure is desired to be a convex function of $y_{\hat{D}}^1, y_{\hat{D}}^2, \dots, y_{\hat{D}}^{N_{\hat{D}}}$ and has to be naturally defined for any $d$.

To introduce the complexity measure, we focus on the extrapolation error which is the difference between the expectation from the data set on a Delaunay simplex $\left\{ \boldsymbol{x}_D^{v_k} \right\}_{k=1}^{d+1} \in \mathcal{DT}\left( \left\{ \boldsymbol{x}_D^k \right\}_{k=1}^{N_D} \right)$ and the measured data at a neighborhood point $\boldsymbol{x}_D^p \in \mathcal{D}_{\mathrm{nbr}}\left( \left\{ \boldsymbol{x}_D^{v_k} \right\}_{k=1}^{d+1}, \left\{ \boldsymbol{x}_D^k \right\}_{k=1}^{N_D} \right)$, that is,

$$\left| y_D^p - \mathrm{Lerp}\left( \boldsymbol{x}_D^p, \left\{ \begin{pmatrix} \boldsymbol{x}_D^{v_k} \\ y_D^{v_k} \end{pmatrix} \right\}_{k=1}^{d+1} \right) \right|. \tag{8}$$

Fig. 5 illustrates this extrapolation error for $d = 2$. If the error is equal to zero, all data points exist on a common plane and all data are explained by one affine map. This extrapolation error can be a measure for map consistency also for $d > 2$, and is convex with respect to $y_{\hat{D}}^1, y_{\hat{D}}^2, \dots, y_{\hat{D}}^{N_{\hat{D}}}$.

Then, we consider the extrapolation error (8) for every combination of Delaunay simplex in $\mathcal{DT}\left( \left\{ \boldsymbol{x}_D^k \right\}_{k=1}^{N_D} \right)$ and its neighborhood point. If most of these extrapolation errors are zero, $\Pi_D$ is simple PWA map, most part of which is flat. Thus number of the combinations which produce non-zero extrapolation error is an appropriate measure of the map complexity; however, minimization of this measure is computationally intractable, and thus we use $\ell_1$ reduction of this measure

$$J_{\text{complexity}}(\Pi_D) \triangleq \sum_{\left\{\begin{pmatrix} \boldsymbol{x}_D^{v_k} \\ y_D^{v_k} \end{pmatrix}\right\}_{k=1}^{d+1} \in \mathcal{DT}\left(\{\boldsymbol{x}_D^k\}_{k=1}^{N_D}\right)} \sum_{\boldsymbol{x}_D^p \in \mathcal{D}_{\text{nbr}}\left(\{\boldsymbol{x}_D^{v_k}\}_{k=1}^{d+1}, \{\boldsymbol{x}_D^k\}_{k=1}^{N_D}\right)} \left| y_D^p - \text{Lerp}\left(\boldsymbol{x}_D^p, \left\{\begin{pmatrix} \boldsymbol{x}_D^{v_k} \\ y_D^{v_k} \end{pmatrix}\right\}_{k=1}^{d+1}\right) \right| \quad (9)$$

for the measure of the map complexity. This measure can be calculated efficiently in optimization problems.

### B. Reduction to Optimization Problem

From the assumption that $f$ is simple PWA map, it seems reasonable to construct $\Pi_{\hat{D}}$ by minimizing $J_{\text{complexity}}(\Pi_{\hat{D}})$ while fitting $\Pi_{\hat{D}}$ to $D_0$. Thus, we propose to reduce the identification problem to the following optimization problem

$$\text{Given} \quad D_0 = \left\{\begin{pmatrix} \boldsymbol{x}_{D_0}^k \\ y_{D_0}^k \end{pmatrix}\right\}_{k=1}^{k=N_{D_0}}, \left\{\boldsymbol{x}_{\hat{D}}^k\right\}_{k=1}^{N_{\hat{D}}} \quad (10)$$

$$\underset{\left\{y_{\hat{D}}^k\right\}_{k=1}^{N_{\hat{D}}}}{\text{minimize}} \quad J_{\text{error}}\left(D_0, \Pi_{\hat{D}}\right) + w \cdot J_{\text{complexity}}\left(\Pi_{\hat{D}}\right), \quad (11)$$

where

$$J_{\text{error}}\left(D_0, \Pi_{\hat{D}}\right) \triangleq \sqrt{\sum_{(\boldsymbol{x},y)\in D_0} \left(y - \Pi_{\hat{D}}(\boldsymbol{x})\right)^2} \quad (12)$$

indicates the disagreement between $\Pi_{\hat{D}}$ and $D_0$; and $w \in \mathbb{R}$ is a user-defined positive constant, which determines the balance between the model complexity and the disagreement. If $w$ is large, obtained $\Pi_{\hat{D}}$ is flat on most Delaunay sides and most of its data points can be removed without changing the map, i.e., $\Pi_{\hat{D}}$ is strongly compressed. And, if $w$ is small, obtained $\Pi_{\hat{D}}$ fits to the measured data set $D_0$. Thus, by choosing appropriate $w$, we can obtain simple and precise model with high capability of generalization.

## V. EXPERIMENT WITH A DC MOTOR SYSTEM

Here, the effectiveness of the proposed scheme is illustrated through an experiment with a DC motor system shown in Fig. 6. This experiment system consists of the DC motor with current input ($\triangleq i(t)$), the harmonic drive system, the inertial load, and the encoder which outputs the rotation angle ($\triangleq \theta(t)$). Here, the harmonic drive gear is a torque transmission system, which has widespread industrial applications and complex dynamic behavior [12]. It is supposed that we know $i(t)$ and can calculate the angular velocity $\omega(t)$ and its derivative $\dot{\omega}(t)$ from $\theta(t)$ (see Fig. 6).

For this system, we construct the model with a PWA map

$$\Pi_{\hat{D}} : \boldsymbol{x} \mapsto y \quad \left(\begin{bmatrix} i \\ \omega \end{bmatrix} \mapsto \dot{\omega}\right) \quad (13)$$

shown in Fig. 7. To obtain the I/O data for the identification, the input current $i(t)$ of the system is changed randomly in every 0.1 [sec], and $(i(t), \omega(t), \dot{\omega}(t))$ are measured in every

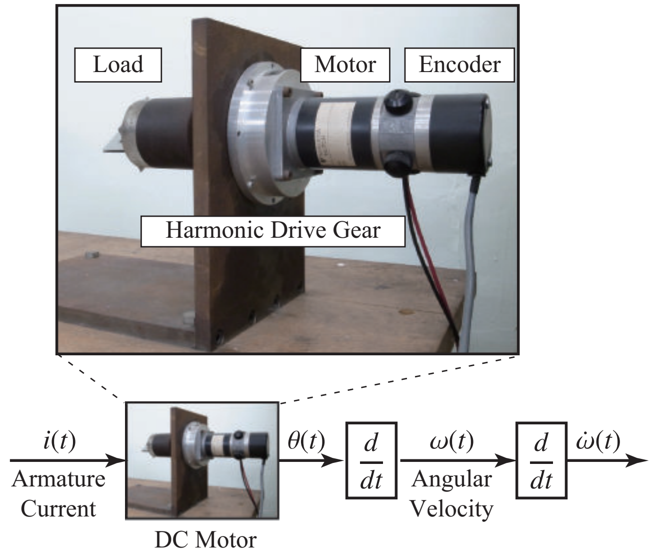



Fig. 6: DC motor system

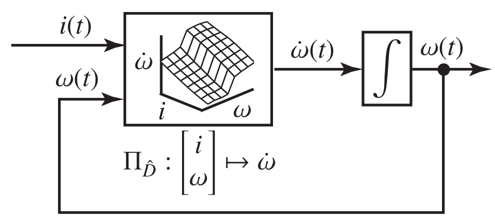



Fig. 7: Data-based PWA model for DC motor system

0.01 [sec] over 5119.95 [sec]. Thus, the problem here is to obtain the PWA map from the data set

$$D_0 = \left\{\begin{pmatrix} \boldsymbol{x}_{D_0}^k \\ y_{D_0}^k \end{pmatrix}\right\}_{k=1}^{511996} \quad (14)$$

$$\left(\begin{matrix} \boldsymbol{x}_{D_0}^k = \begin{bmatrix} i(t_k) \\ \omega(t_k) \end{bmatrix}, \quad y_{D_0}^k = \dot{\omega}(t_k), \\ t_k \triangleq 0.01 \cdot (k-1)\,[\text{sec}] \end{matrix}\right). \quad (15)$$

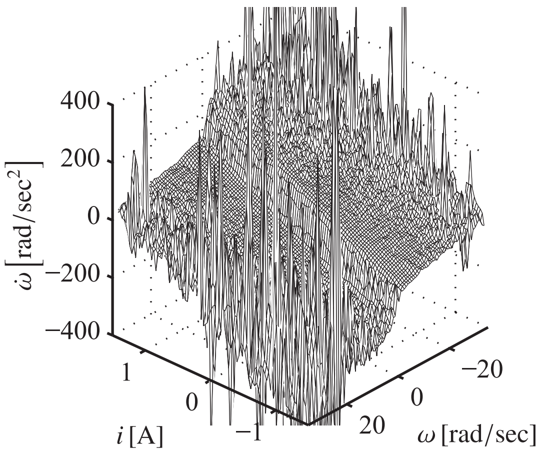
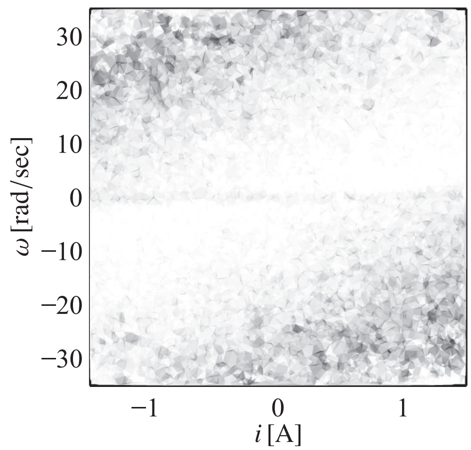
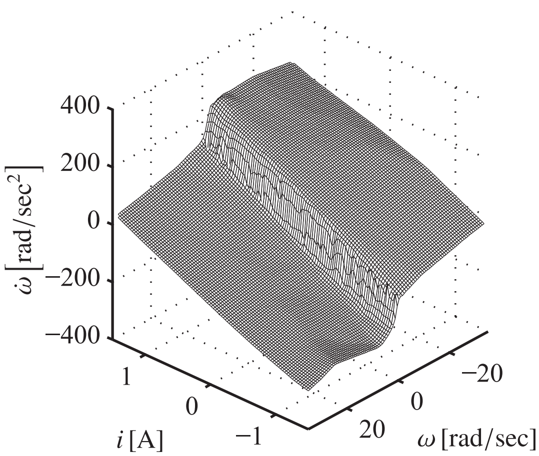
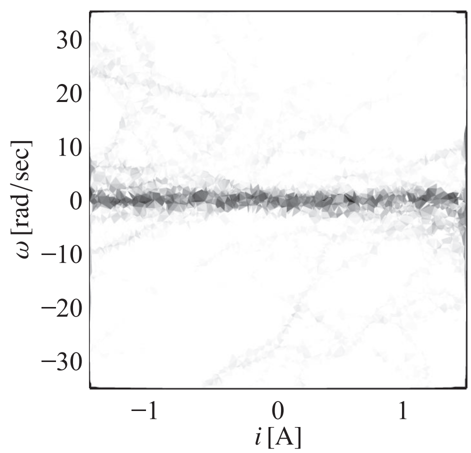
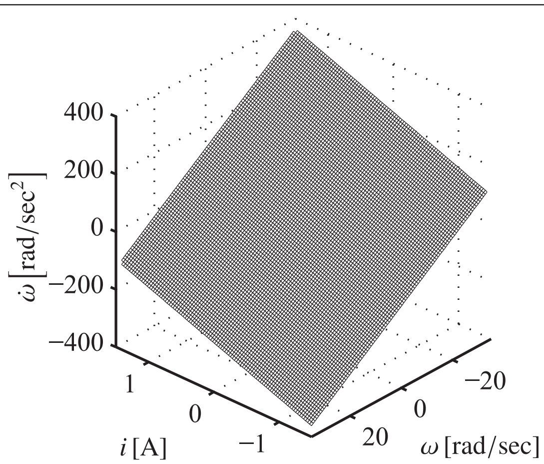
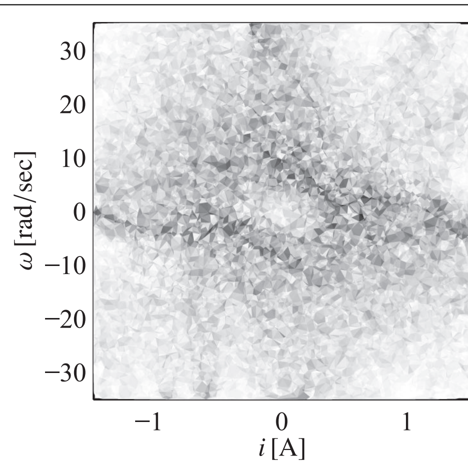
Fig. 8 shows the first 5 [sec] of the data measured from the target system. Also, the size of the data set of the constructed data-based PWA model is set to $N_{\hat{D}} = 10000$, and $\left\{\boldsymbol{x}_{\hat{D}}^k\right\}_{k=1}^{N_{\hat{D}}}$ are randomly chosen from

$$\left\{\begin{bmatrix} i \\ \omega \end{bmatrix} : \begin{matrix} -1.5\,[\text{A}] \le i \le 1.5\,[\text{A}], \\ -35\,[\text{rad/sec}] \le \omega \le 35\,[\text{rad/sec}] \end{matrix}\right\}, \quad (16)$$

which covers the region reached by the real system in normal operation. Since the elements of $\boldsymbol{x}\left([i,\omega]^T\right)$ represent different type of physical quantities, these quantities are normalized by using 1.5 [A] and 35 [rad/sec] as a unit in considering Delaunay triangulation.

With these settings, we construct $\Pi_{\hat{D}}$ by the proposed method for various $w$, and obtained $\Pi_{\hat{D}}$ are shown in Table I with the extrapolation error for the map. In the plot of extrapolation error, the contribution of each Delaunay simplex

TABLE I: Illustration of the obtained PWA map for the DC motor experiment system. The effect of compression is clearly seen by comparing three results.

| Compression Level | PWA Map | Extrapolation Error |
|---|---|---|
| Weak $\left(w = 5 \times 10^{-6}\right)$ |  |  |
| Moderate $\left(w = 5 \times 10^{-3}\right)$ |  |  |
| Strong $\left(w = 5 \times 10^{0}\right)$ |  |  |

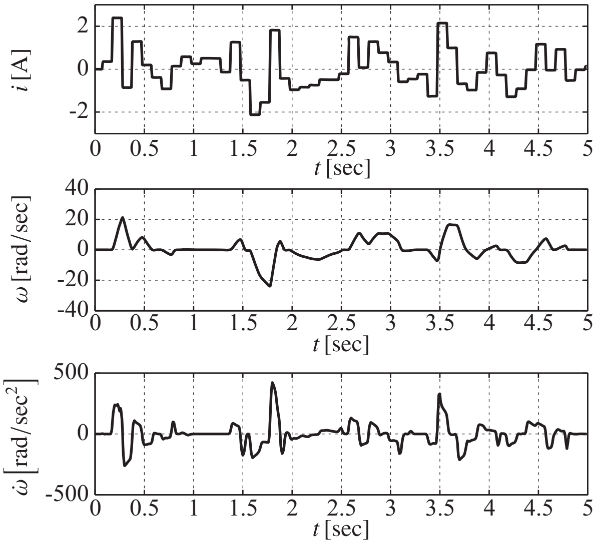



Fig. 8: I/O data for first 5 seconds

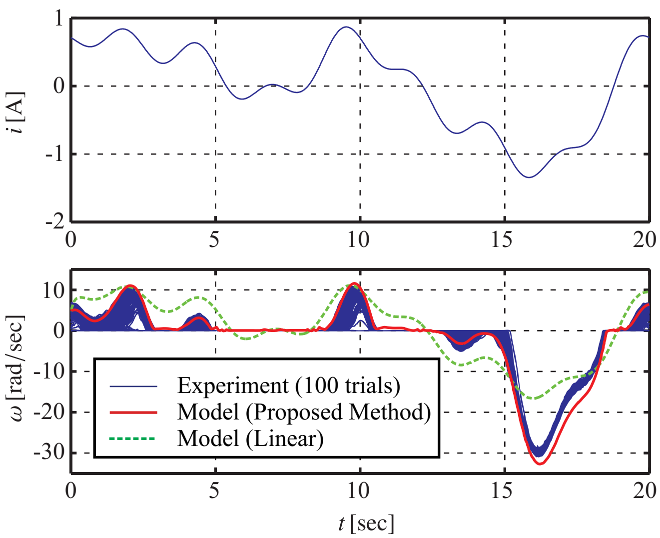



Fig. 9: Validation input and output of the experiment system and the model obtained with moderate compression level ($w = 5 \times 10^{-3}$)

to $J_{\text{complexity}}(\Pi_{\hat{D}})$ is shown, where a simplex $S$ is filled with more darker color if its contribution

$$\sum_{x_D^p \in \mathcal{D}_{\text{nbr}}\left(\{x_D^k\}_{k=1}^{N_D}, S\right)} \left| y_D^p - \text{Lerp}\left(x_D^p, S\right) \right| \tag{17}$$

is relatively larger. As seen in Table I, the result of weak compression seems to be affected by the measurement noise, while the excessively strong compression produces the standard linear map. Also, it is confirmed that the complexity of the PWA map is well controlled by the design parameter $w$, and the sparsity of the extrapolation error is a good measure of the complexity.

Next, we compare the response of the obtained model for a validation input with the real system to validate the obtained model. The periodic input for the validation and the results are shown in Fig. 9, where the solid blue thin lines show the output of the experiment system for 100 cycles and the thick red solid line shows the output of the model with the PWA map for $w = 5 \times 10^{-3}$ (see Table I). Also, the output of the standard linear model is shown as the green dashed line for comparison purpose. These results shows that an appropriate PWA model is obtained by the proposed method for the practical system, which has strong non-linearity.

## VI. Conclusion

In this paper, a new identification method for piecewise affine (PWA) models is introduced. The method is based on the data-based representation of PWA maps and the data compression with $\ell_1$ optimization technique. As illustrated through the example, the proposed method can construct good PWA models without prior knowledge about mode transition. The notable feature of the proposed method is that it can handle systems with high-dimensional input and large data sets because the proposed compression procedure results in a convex optimization problem which can be solved efficiently. Furthermore, we can easily balance the model complexity and preciseness in this scheme by adjusting one parameter, and these facts contribute high usability of the proposed method.

## References


[1] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice-Hall, 1999.

[2] H. Garnier and L. Wang, *Identification of Continuous-Time Models from Sampled Data*, ser. Advances in Industrial Control. Springer Verlag, 2008.

[3] B. Armstrong-Héouvry, P. Dupont, and C. C. D. Wit, "A survey of models, analysis tools and compensation methods for the control of machines with friction," *Automatica*, vol. 30, no. 7, pp. 1083 – 1138, 1994.

[4] L. Breiman, "Hinging hyperplanes for regression, classification, and function approximation," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 999 –1013, may 1993.

[5] J.-N. Lin and R. Unbehauen, "Canonical piecewise-linear approximations," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 39, no. 8, pp. 697 –699, aug 1992.

[6] S. Paoletti, A. Juloski, G. Ferrari-Trecate, and R. Vidal, "Identification of hybrid systems: a tutorial," *European Journal of Control*, vol. 13, no. 2-3, pp. 242–260, 2007.

[7] J. Roll, A. Bemporad, and L. Ljung, "Identification of piecewise affine systems via mixed-integer programming," *Automatica*, vol. 40, no. 1, pp. 37 – 50, 2004.

[8] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.

[9] H. Ohlsson, L. Ljung, and S. Boyd, "Segmentation of arx-models using sum-of-norms regularization," *Automatica*, vol. 46, no. 6, pp. 1107 – 1111, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/B6V21-4YX0BV8-2/2/a1cc6678aa3592a5debdafccb82792af

[10] I. Maruta and T. Sugie, "A new approach for modeling hybrid systems based on the minimization of parameter transition in linear time-varying models," in *Proc. of 49th IEEE Conference on Decision and Control*, December 2010, pp. 1177–1182.

[11] M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*, 3rd ed. Springer-Verlag, 2008.

[12] H. D. Taghirad and P. R. Bélanger, "Modeling and parameter identification of harmonic drive systems," *Journal of Dynamic Systems, Measurement, and Control*, vol. 120, no. 4, pp. 439–444, 1998. [Online]. Available: http://link.aip.org/link/?JDS/120/439/1