Zonotope parameter identification for piecewise affine systems

WANG Jianhong[*]

School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey 64849, Mexico

Abstract: The problem of how to identify the piecewise affine system is studied in this paper, where this considered piecewise affine system is a special nonlinear system. The reason why it is not easy to identify this piecewise affine system is that each separated region and each unknown parameter vector are all needed to be determined simultaneously. Then, firstly, in order to achieve the identification goal, a multi-class classification process is proposed to determine each separated region. As the proposed multi-class classification process is the same with the classical data clustering strategy, the multi-class classification process can combine the first order algorithm of convex optimization, while achieving the goal of the classification process. Secondly, a zonotope parameter identification algorithm is used to construct a set, which contains the unknown parameter vector. In this zonotope parameter identification algorithm, the strict probabilistic description about the external noise is relaxed, and each unknown parameter vector is also identified. Furthermore, this constructed set is consistent with the measured output and the given bound corresponding to the noise. Thirdly, a sufficient condition about guaranteeing our derived zonotope not growing unbounded with iterations is formulated as an explicit linear matrix inequality. Finally, the effectiveness of this zonotope parameter identification algorithm is proven through a simulation example.

Keywords: piecewise affine system, zonotope parameter identification, linear matrix inequality.

1. Introduction

In this paper, our considered piecewise affine system can be regarded as a special hybrid dynamical system, as the piecewise affine system denotes the switching principle by using a collection of linear differential or difference equations, whose state space is partitioned by a finite number of linear hyperplanes. Generally hybrid dynamical systems belong to a class of complex dynamic systems, which include interacting discrete events and continuous variable dynamical systems. All above mentioned hybrid dynamical systems are important in lots of application fields, such as embedded systems, cyber physical systems, robotics,

manufacturing systems and biomolecular networks, and they have recently been at the center of intense research activities in the control theory and artificial intelligence communities. However, in order to control a dynamical system, generally after modeling the considered plant by using the system identification theory, the process of system identification is finished and the process of control design is started. Moreover, the control performance depends on the mathematical model for the considered plant closely, i.e., the accuracy of the mathematical model will affect the latter control performance. It means the goal of system identification is to provide a mathematical basis for the next controller design, and this is the reason of the name for "identification for control". Because the considered plant and controller exist in the original closed loop system simultaneously, before identifying that unknown plant, the controller needs to be considered whether it is known or not in priori.

According to [1], the process of system identification consists of designing and conducting the identification experiment in order to collect the measurement data, selecting the structure of the model, specifying the parameters to be identified, and eventually fitting the model parameters to the obtained data. A large number of nonlinear model structures have been constructed to investigate their properties [2], where some real time fast convex algorithms are proposed to identify model parameters. Identification of the hybrid systems is an area that is related to many other research fields within nonlinear system identification, as such hybrid systems are sufficiently expressive to model many physical processes [3]. The identification of the piecewise affine system is a challenging problem, as it involves the estimation of both the parameters of the affine sub-models, and the coefficients of the hyper-planes [4]. Lauer et al. proposed to exploit the combined use of clustering, linear identification, and pattern recognition techniques [5]. In [6], the sub-model parameters were described through probability density functions. The sum of the norms regularization strategy in [7] can be computationally heavy in case of appropriate step

size. The piecewise affine system identification problem amounts to learning from a set of training data [8]. This piecewise affine system identification problem is a non-deterministic polynomial (NP) hard problem in general [9]. For the sake of simplicity, the sparse property is imposed in piecewise affine systems [10]. The strengths of the piecewise affine system identification problem of [11] are the computational efficiency and the ability to be run both in a batch and in a recursive way. A three-stage procedure of a bounded error approach for parametric identification of piecewise affine autoregressive exogenous models was proposed in [12]. The conversion of piecewise affine models from the state space input-output form was addressed by deriving necessary and sufficient conditions [13]. A convex relaxation based on L1 regulation was proposed in [14] to approximate the underlying combinatorial problem. The statistical clustering technique in [15] computed the parameters of the affine local models, then partitioned the regressor space.

It is well known that in the piecewise affine system, the space is partitioned into many separated regions and a local linear form is used for each separated region. Thus the first step in identifying the piecewise affine system is to determine these separated regions. After the separate regions are given, the second identification problem is reduced to identify the linear submodels for each region. To deal with the above mentioned steps, we reformulate the problem of determining the separated regions as a multi-class classification problem, which can be solved by the classical first order algorithm from the convex optimization theory. A multi-class classification problem coincides with a data clustering process into the separated regions. When we identify the unknown parameter in each separated region, many classical identification algorithms can be used directly. However, all the classical identification algorithms hold in case that the considered noise may be a zero mean random signal. To relax this strict probabilistic description on noise, a zonotope parameter identification algorithm is investigated in the presence of bounded noise. Generally the zonotope parameter identification algorithm computes a set that contains the parameters, and this set is consistent with the measured output and the given bound of the disturbance. To keep our derived zonotope not growing unbounded with the iteration steps, some contracting properties must be imposed.

The rest of this paper is organized as follows. In Section 2, the problem setting and the piecewise affine system are presented. In Section 3, a multi-class classification problem based on the first order algorithm from the convex optimization theory is introduced to determine the separate regions. In Section 4, a zonotope parameter identification algorithm is proposed to identify the unknown parameters in the presence of bounded noise for each separated region. In Section 5, a very simple numerical example is used to illustrate the proposed algorithm. Finally, conclusions and comments about future research are presented in Section 6.

2. Piecewise affine system

Consider the following affine model as

$$y(t) = -\sum_{i=1}^{n_a} a_i y(t-i) + \sum_{j=1}^{n_b} b_j u(t-i) + e(t) \qquad (1)$$

where $u(t)$ and $y(t)$ are input and output signals respectively, a_i $(i = 1, \ldots, n_a)$ and b_j $(j = 1, \ldots, n_b)$ are the unknown models or system parameters, and $e(t)$ is an external noise or disturbance. Two numbers n_a and n_b are priori known.

Rewrite (1) as a linear regression form, and define a regression vector $\phi(t)$ as

$$\begin{cases} \phi(t) = [-y(t-1) \quad -y(t-2) \quad \cdots \quad -y(t-n_a) \quad u(t-1) \quad u(t-2) \quad \cdots \quad u(t-n_b)]^{\mathrm{T}} \\ y(t) = \phi^{\mathrm{T}}(t)\theta + e(t) \end{cases} \qquad (2)$$

where the unknown parameter vector is stacked as

$$\theta = [a_1 \quad a_2 \quad \cdots \quad a_{n_a} \quad b_1 \quad b_2 \quad \cdots \quad b_{n_b}]^{\mathrm{T}}. \qquad (3)$$

For large enough orders n_a and n_b, the affine model can be used to approximate any linear system, but it cannot capture any nonlinear properties, so to extend the affine model into the nonlinear system, our considered piecewise affine system is obtained. At this time, the unknown parameter vector θ is dependent of the region in the regression space, where the regression vector $\phi(t)$ lies. Thus the regression space can be divided into n separated regions $R_1 \cdots R_n$. Let us define the piecewise affine system as

$$y(t) = \phi^{\mathrm{T}}(t)\theta_i + e(t), \quad \phi(t) \in R_i \qquad (4)$$

where the parameter vector θ_i depends on its corresponding separated region R_i. The identification problem for the piecewise affine system is reformulated as that, after the output and input observed data point $\{u(t), y(t)\}$ is collected by using some sensors, how to identify those unknown parameter vectors θ_i $(i = 1, \ldots, n)$? Due to the fact that the regression vector $\phi(t)$ is constituted by the output and input observed data point $\{u(t), y(t)\}$, the first step is to determine which region the regression vector belongs to.

3. Multi-class classification process

Because n separated regions $R_1 \cdots R_n$ exist, the determi-

nation about which region the regression vector lies is in conjunction with a multi-class classification process.

Observing N input-output data points $\boldsymbol{z}(t)$ ($t = 1, \ldots, N$) as follows:

$$\boldsymbol{z}(t) = [y(t) \quad \boldsymbol{\phi}(t)]^{\mathrm{T}} =$$

$$[y(t) \quad - y(t-1) \quad - y(t-2) \quad \cdots \quad - y(t-n_a)$$
$$u(t-1) \quad u(t-2) \quad \cdots \quad u(t-n_b)]^{\mathrm{T}} \quad (5)$$

where N denotes the number of observed input-output data points, and each data point $\boldsymbol{z}(t)$ is included in one of n non-overlapping classes, along with labels $\lambda_t \in \boldsymbol{R}^n$, where $\boldsymbol{R}^n$ is the real number set with dimension n, and basic quadrants in $\boldsymbol{R}^n$, and the index of the only nonzero entry in λ_t is the number of class to which $\boldsymbol{z}(t)$ belongs.

A multi-class analogy of the standard linear classifier is built as follows: a multi-class classifier is specified by a matrix $\boldsymbol{A}$ and a vector $\boldsymbol{a} \in \boldsymbol{R}^n$. Given the input-output observed data point $\boldsymbol{z}(t)$, compute the n dimensional vector $\boldsymbol{A}\boldsymbol{z}(t) + \boldsymbol{a}$, identify its maximal component, and treat the index of this component as our guess for the serial number of the class to which $\boldsymbol{z}(t)$ belongs.

Set $\overline{\lambda}_t = 1 - \lambda_t$ as the component of λ_t. Given a data point $\boldsymbol{z}$ and the corresponding label λ, let us set

$$h = h(\boldsymbol{A}, \boldsymbol{a}, \boldsymbol{z}, \lambda) = [\boldsymbol{A}\boldsymbol{z} + \boldsymbol{a}] - [\lambda(\boldsymbol{A}\boldsymbol{z} + \boldsymbol{a})] + \overline{\lambda}. \quad (6)$$

If i_* is the index of the only nonzero entry in λ, then the i_*th entry in h is zero. h is nonpositive if and only if the classifier, given by $\boldsymbol{A}, \boldsymbol{a}$ and evaluated at $\boldsymbol{z}$, recovers the class i_* of $\boldsymbol{z}$ with a margin 1, i.e., we have

$$[\boldsymbol{A}\boldsymbol{z} + \boldsymbol{a}]_j \leqslant [\boldsymbol{A}\boldsymbol{z} + \boldsymbol{a}]_{i_*} - 1, \quad j \neq i_*. \quad (7)$$

On the other hand, if the classifier fails to classify $\boldsymbol{z}$ correctly, that is,

$$[\boldsymbol{A}\boldsymbol{z} + \boldsymbol{a}]_j \geqslant [\boldsymbol{A}\boldsymbol{z} + \boldsymbol{a}]_{i_*} \quad (8)$$

for some $j \neq i_*$, then the maximal entry in h is equal to or larger than 1, so we introduce the following equation:

$$\eta(\boldsymbol{A}, \boldsymbol{a}, \boldsymbol{z}, \lambda) = \max_{1 \leqslant j \leqslant n} [h(\boldsymbol{A}, \boldsymbol{a}, \boldsymbol{z}, \lambda)]_j. \quad (9)$$

A nonnegative function is obtained, and it always vanishes for the pairs $(\boldsymbol{z}, \lambda)$. The pairs are quite reliably (with margin $\geqslant 1$) classified by $(\boldsymbol{A}, \boldsymbol{a})$, and equal to or larger than 1 for the pairs $(\boldsymbol{z}, \lambda)$ with $\boldsymbol{z}$ not classified correctly. Thus, the following function is simplified as

$$F(\boldsymbol{A}, \boldsymbol{a}) = \mathrm{E}\{\eta(\boldsymbol{A}, \boldsymbol{a}, \boldsymbol{z}, \lambda)\}. \quad (10)$$

This expectation being taken over the whole distribution of the pairs $(\boldsymbol{z}, \lambda)$ is an upper bound on the probability for the classifier $(\boldsymbol{A}, \boldsymbol{a})$ to misclassify a data point. What we would do is to minimize $F(\boldsymbol{A}, \boldsymbol{a})$ over $\boldsymbol{A}$ and $\boldsymbol{a}$. To achieve this mission, since $F(\boldsymbol{A}, \boldsymbol{a})$ is not observable, we replace the expectation by its empirical counterpart

$$F_N(\boldsymbol{A}, \boldsymbol{a}) = \frac{1}{N} \sum_{t=1}^{N} \eta(\boldsymbol{A}, \boldsymbol{a}, \boldsymbol{z}(t), \lambda_t). \quad (11)$$

For the sake of simplicity, imposing an upper bound on some norm $\|\boldsymbol{A}\|$ of $\boldsymbol{A}$, an optimization problem is obtained.

$$\min_{\boldsymbol{A}, \boldsymbol{a}} \frac{1}{N} \sum_{t=1}^{N} \max_{i \leqslant n} [\boldsymbol{A}\boldsymbol{z}(t) + \boldsymbol{a} - \lambda_t(\boldsymbol{A}\boldsymbol{z}(t) + \boldsymbol{a}) + \overline{\lambda}_t]_i$$

$$\text{s.t.} \ \|\boldsymbol{A}\| \leqslant 1 \quad (12)$$

A natural choice of the norm $\|\boldsymbol{A}\|$ is the maximum of the $\|\boldsymbol{A}\|_2$ norm. Once optimization variables $\boldsymbol{A}$ and $\boldsymbol{a}$ are obtained, then the linear classifier $\boldsymbol{A}\boldsymbol{z}(t) + \boldsymbol{a}$ is get. From the above multi-class classification process, after a data point is collected, we cluster it with a linear classifier. Thus based on the above linear classifier, all data points can be clustered together, then those data points clustering together as a class can be used in the second identification problem for an unknown parameter.

4. Zonotope parameter identification algorithm

After all collected input-output data points are clustered as these n classes, then these data points belonging to the same class can be used to identify an unknown parameter. Here we only rewrite the following piecewise affine system in the ith separate region.

$$y(t) = \boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}_i + e(t), \quad \boldsymbol{\phi}(t) \in \boldsymbol{R}_i \quad (13)$$

The goal of this section is to identify the unknown parameter vector $\boldsymbol{\theta}_i$ in case of the unknown but bounded noise. Based on this unknown but bounded noise, one of the new identification approaches, the zonotope parameter identification algorithm, is chosen to identify the unknown parameter vector $\boldsymbol{\theta}_i$. This identification algorithm obtains a set iteratively that includes the parameters consistent with the measured output signal and the given bound of the disturbance or noise. The zonotope is used to represent this obtained set. The reason of using the zonotope is that a zonotope is an affine map of a unitary hypercube.

Observing (13) again, as $e(t)$ represents the considered disturbance or external noise, and assume this external noise belongs to a bounded set, i.e.,

$$e(t) \in \{e \in \boldsymbol{R} : |e| \leqslant \sigma\} \quad (14)$$

where $\sigma \in \mathbf{R}$ is an upper bound and external noise $e(t)$ is unknown, but it has a known bound in priori.

From the set membership identification theory, given a set of measured outputs, the feasible solution set (FSS) for the unknown parameter is defined as the set of parameters that are consistent with measured outputs and the given bounds of the considered external noise. More precisely, the following definitions are given through this whole zonotope parameter identification algorithm.

Definition 1 FSS

Suppose the observed input-output pairs $\{y(t), \boldsymbol{\phi}(t)\}$ ($t = 1, 2, \ldots, N$) are given. The unknown parameter vector $\boldsymbol{\theta}_i$ is regarded to belong to the FSS if there exists $\boldsymbol{\theta}_i$, such that

$$|y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}_i| \leqslant \sigma, \quad t = 1, 2, \ldots, N. \tag{15}$$

Definition 2 Information set

Given the observed input-output pairs $\{y(t), \boldsymbol{\phi}(t)\}$ ($t = 1, 2, \ldots, N$) at time instant t, the information set I_t is deemed as a set of all feasible parameters, which are consistent with the linear regression model (13), the measured output $y(t)$ and the known bound at time instant t, namely,

$$I_t = \{\boldsymbol{\theta}_i \in \boldsymbol{R}^{n_a + n_b} : -\sigma \leqslant y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}_i \leqslant \sigma\}. \tag{16}$$

Geometrically, I_t represents a strip, which is consistent with the observed input-output pairs $\{y(t), \boldsymbol{\phi}(t)\}$ ($t = 1, 2, \ldots, N$).

The FSS at the next time instant $t + 1$, denoted as FSS_{t+1}, can be computed exactly from the one corresponding to time instant t by the following iterative recursion:

$$\mathrm{FSS}_{t+1} = \mathrm{FSS}_t \cap I_t. \tag{17}$$

Because it is difficult to compute the FSS, an outer bound of the FSS can be defined again.

Definition 3 Approximated FSS (AFSS)

An AFSS is a set that satisfies FSS. The intersection $\mathrm{FSS}_t \cap I_t$ is approximated by means of the intersection between a zonotope and a strip at time instant t.

Definition 4 Zonotope of order m

Given a vector $\boldsymbol{p} \in \boldsymbol{R}^{n_a + n_b}$ and a matrix $\boldsymbol{H} \in \boldsymbol{R}^{(n_a + n_b) \times m}$, a zonotope of order m is a set with $n_1 = n_a + n_b$ dimensional vectors.

$$Z = \{\boldsymbol{\theta}_i \in \boldsymbol{R}^{n_1} : \boldsymbol{\theta}_i \in \boldsymbol{p} \oplus \boldsymbol{H}\boldsymbol{B}^m\} \tag{18}$$

where $\boldsymbol{H}\boldsymbol{B}^m$ is a linear projection of $\boldsymbol{B}^m$ into $n_1 = n_a + n_b$ dimensional parameter space, $\boldsymbol{B}^m$ is a unit hypercube of the order m, and $\oplus$ denotes the Miniowski sum.

Using the above definitions and the AFSS on the intersection (17), then we have

$$\mathrm{FSS}_{t+1} = \mathrm{FSS}_t \cap I_t \subseteq \mathrm{AFSS}_{t+1}. \tag{19}$$

If in (19), FSS_t is denoted by a defined zonotope, and the information set I_t is a strip, then a family of zonotopes, bounding the intersection between a zonotope and a strip, are derived as the following obtained Theorem 1.

Theorem 1 Suppose a zonotope is used to denote an FSS at time instant t.

$$\mathrm{FSS}_t = \widehat{\boldsymbol{p}}_t \oplus \widehat{\boldsymbol{H}}_t \boldsymbol{B}^r \subset \boldsymbol{R}^{n_1} \tag{20}$$

and the information set or a strip is given as

$$I_t = \{\boldsymbol{\theta}_i \in \boldsymbol{R}^{n_a + n_b} : -\sigma \leqslant y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}_i \leqslant \sigma\} \tag{21}$$

and a scalar γ defines the following variables as

$$\begin{cases} \widehat{\boldsymbol{p}}_t(\gamma) = \widehat{\boldsymbol{p}}_t + \gamma(y(t) - \boldsymbol{\phi}^{\mathrm{T}}(t)\widehat{\boldsymbol{p}}_t) \\ \widehat{\boldsymbol{H}}_t(\gamma) = [(\boldsymbol{I} - \gamma\boldsymbol{\phi}(t))\widehat{\boldsymbol{H}}_t \quad \sigma\gamma] \end{cases} \tag{22}$$

where $\boldsymbol{I}$ is an identity matrix. Thus we have

$$\mathrm{FSS}_{t+1} = \mathrm{FSS}_t \cap I_t \subseteq \mathrm{AFSS}_{t+1} =$$
$$\widehat{\boldsymbol{p}}_t(\gamma) \oplus \widehat{\boldsymbol{H}}_t(\gamma)\boldsymbol{B}^{r+1}. \tag{23}$$

A scalar $\gamma \in \mathbf{R}$ is chosen by an optimization based method, through minimizing the volume of the obtained zonotope. Now the minimization of the P-radius of a zonotope is applied, as the P-radius criterion allows to guarantee the non-increasing property of the guaranteed zonotope at each time instant. It tells us that to guarantee the AFSS not growing unbounded with iteration steps, the following inequality relation between two neighboring zonotopes is imposed to guarantee that property.

$$l_t \leqslant \beta l_{t-1} + \sigma^2 \tag{24}$$

where $\beta \in (0, 1]$ is a contraction rate, and l_t is a chosen parameter or the P-radius of the zonotope parameter estimation set at time instant t, which is defined by

$$l_t = \max_{\boldsymbol{\theta}_i \in \mathrm{FSS}_t} (\|\boldsymbol{\theta}_i - \widehat{\boldsymbol{p}}_t\|_{\boldsymbol{P}}^2) \tag{25}$$

where $\boldsymbol{P}$ is an n_1-dimensional positive definite matrix.

After substituting (25) into (24), we have

$$\max_{\widehat{\boldsymbol{z}} \in \boldsymbol{B}^{r+1}} (\|\widehat{\boldsymbol{H}}_t(\gamma)\widehat{\boldsymbol{z}}\|_{\boldsymbol{P}}^2) \leqslant$$
$$\max_{\boldsymbol{z} \in \boldsymbol{B}^r} \beta(\|\widehat{\boldsymbol{H}}_{t-1}\boldsymbol{z}\|_{\boldsymbol{P}}^2) + \max_{\eta \in \boldsymbol{B}^1} \|\sigma\eta\|_2^2. \tag{26}$$

Expanding (26) to obtain the following inequity:

$$\widehat{\boldsymbol{z}}^{\mathrm{T}}\widehat{\boldsymbol{H}}_t(\gamma)\boldsymbol{P}\widehat{\boldsymbol{H}}_t(\gamma)\widehat{\boldsymbol{z}} - \beta\boldsymbol{z}^{\mathrm{T}}\widehat{\boldsymbol{H}}_{t-1}\boldsymbol{P}\widehat{\boldsymbol{H}}_{t-1}\boldsymbol{z} - \eta^2\sigma^2 \leqslant 0. \tag{27}$$

Due to the recursion property of $\widehat{\boldsymbol{H}}_t(\gamma)$ in (22), we continue to compute

$$\widehat{\boldsymbol{H}}_t(\gamma)\widehat{\boldsymbol{z}} = (\boldsymbol{I} - \gamma\boldsymbol{\phi}(t))\widehat{\boldsymbol{H}}_{t-1}\boldsymbol{z} + \sigma\gamma\eta =$$
$$(\boldsymbol{I} - \gamma\boldsymbol{\phi}(t))\overline{\boldsymbol{z}} + \sigma\gamma\eta \tag{28}$$

where we set two new defined variables

$$\begin{cases} \overline{z} = \widehat{H}_{t-1}z \\ \widehat{z} = [z \quad \eta]^{\mathrm{T}} \end{cases}.$$ (29)

Applying (28) in (27), we get

$$[\overline{z}^{\mathrm{T}}(I - \gamma\phi(t))^{\mathrm{T}} + \sigma\gamma\eta]P[(I - \gamma\phi(t))\overline{z} + \sigma\gamma\eta] - \beta\overline{z}^{\mathrm{T}}P\overline{z} - \eta^2\sigma^2 \leqslant 0.$$ (30)

Formulating the above inequality to simplify it as

$$\overline{z}^{\mathrm{T}}(I - \gamma\phi(t))^{\mathrm{T}}P(I - \gamma\phi(t))\overline{z} + \overline{z}^{\mathrm{T}}(I - \gamma\phi(t))^{\mathrm{T}}P\sigma\gamma\eta +$$

$$\sigma\gamma\eta P(I - \gamma\phi(t))\overline{z} + \sigma^2\gamma^2\eta^2 - \beta\overline{z}^{\mathrm{T}}P\overline{z} - \eta^2\sigma^2 \leqslant 0.$$ (31)

A sufficient condition for (31) to hold can be rewritten as the following linear matrix inequality:

$$\begin{bmatrix} \overline{z} \\ \eta \end{bmatrix}^{\mathrm{T}} \times \begin{bmatrix} (I - \gamma\phi(t))^{\mathrm{T}}P(I - \gamma\phi(t)) - \beta P & (I - \gamma\phi(t))^{\mathrm{T}}P\sigma\gamma \\ P\sigma\gamma(I - \gamma\phi(t)) & \sigma^2(\gamma^2 - 1) \end{bmatrix} \begin{bmatrix} \overline{z} \\ \eta \end{bmatrix} \leqslant 0,$$

$$\forall \overline{z}, \eta.$$ (32)

Using the definition and the property of the positive definite matrix, we rewrite it as

$$\begin{bmatrix} (I - \gamma\phi(t))^{\mathrm{T}}P(I - \gamma\phi(t)) - \beta P & (I - \gamma\phi(t))^{\mathrm{T}}P\sigma\gamma \\ P\sigma\gamma(I - \gamma\phi(t)) & \sigma^2(\gamma^2 - 1) \end{bmatrix} \leqslant 0,$$

$$\forall \begin{bmatrix} \overline{z} \\ \eta \end{bmatrix} \neq 0.$$ (33)

The linear matrix inequality in (33) defines the feasible solution for scalar γ, i.e., γ can be computed by solving the following eigenvalue problem:

$$\max_{\tau, \gamma} \tau$$

$$\text{s.t.} \quad \frac{(1 - \beta)P}{\max\limits_{\eta \in B} \|\gamma\eta\|_2^2} \geqslant \tau I, \quad \tau > 0$$

$$\begin{bmatrix} (I - \gamma\phi(t))^{\mathrm{T}}P(I - \gamma\phi(t)) - \beta P & (I - \gamma\phi(t))^{\mathrm{T}}P\sigma\gamma \\ P\sigma\gamma(I - \gamma\phi(t)) & \sigma^2(\gamma^2 - 1) \end{bmatrix} \leqslant 0.$$ (34)

After solving the above convex optimization algorithm, then based on this optimal scalar $\gamma \in \mathbf{R}$, a zonotopic outer approximation of the intersection between a zonotope and a strip is obtained by using the matrix inequality optimization strategy.

Finally, our zonotope parameter identification algorithm is formulated as follows.

Algorithm 1 Zonotope parameter identification algorithm

Step 1 Obtain measured input-output data points and construct the regressor vector $\phi(t)$ according to (5).

Step 2 Build a strip that bounds the consistent parameters, i.e., it is similar to the following information set:

$$I_t = \{\theta_i \in \mathbf{R}^{n_a + n_b} : -\sigma \leqslant y(t) - \phi^{\mathrm{T}}(t)\theta_i \leqslant \sigma\}.$$

Step 3 Construct a zonotope

$$\text{FSS}_t = \widehat{p}_t \oplus \widehat{H}_t B^r \subset \mathbf{R}^{n_1}$$

to denote the FSS at time instant t.

Step 4 Compute the intersection between a zonotope and a strip at time instant t and obtain a new zonotope at the next time instant $t + 1$:

$$\text{FSS}_{t+1} = \text{FSS}_t \cap I_t \subseteq \text{AFSS}_{t+1} = \widehat{p}_t(\gamma) \oplus \widehat{H}_t(\gamma)B^{r+1}$$

to denote the AFSS at time instant $t + 1$.

Step 5 Choose an optimal scalar γ through solving a matrix inequality optimization strategy according to (34).

Step 6 Repeat the above steps and terminate the recursive algorithm when the P-radius l_t is zero, then denote $\widehat{p}^*$ as the vector in the last zonotope, so the unknown parameter vector θ_i is given by

$$\widehat{\theta}_i = \widehat{p}^*.$$ (35)

It is similar to applying the above six steps to identify another unknown parameter vector.

Based on the linear matrix inequality (33), it corresponds to the contracting properties between two neighboring zonotopes, then this linear matrix inequality can be regarded as a sufficient condition to guarantee that the volume of the obtained zonotope will be decreased as the iteration steps increase. It means that after the above six steps are stopped, the volume of the final zonotope will be sufficiently small, then the center of the final zonotope can be chosen as the parameter estimation. Therefore, the convergence consistency of the proposed zonotope parameter identification algorithm can be guaranteed by the added contracting properties.

5. Numerical example

In this numerical example section, a special piecewise affine system is used to prove our ideas, such as the two class classification process and the zonotope parameter

identification algorithm. This simple piecewise affine system is given as follows:

$$y(t) = \begin{cases} \boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}_1 + e(t), & \boldsymbol{\phi}(t) > 0 \\ \boldsymbol{\phi}^{\mathrm{T}}(t)\boldsymbol{\theta}_2 + e(t), & \boldsymbol{\phi}(t) \leqslant 0 \end{cases} \tag{36}$$

where the condition that the regression vector $\boldsymbol{\phi}(t)$ satisfies $\boldsymbol{\phi}(t) > 0$ means all the elements in the regression vector $\boldsymbol{\phi}(t)$ are positive. Furthermore, similarly the condition that the regression vector $\boldsymbol{\phi}(t)$ satisfies $\boldsymbol{\phi}(t) \leqslant 0$ means all the elements in the regression vector $\boldsymbol{\phi}(t)$ are negative or zero.

In (36), the regression vector $\boldsymbol{\phi}(t)$ and two unknown parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are described as

$$\begin{cases} \boldsymbol{\phi}(t) = [-y(t) \quad -u(t)]^{\mathrm{T}} \\ \boldsymbol{\theta}_1 = [7 \quad 2]^{\mathrm{T}} \\ \boldsymbol{\theta}_2 = [2 \quad 0.5]^{\mathrm{T}} \end{cases}. \tag{37}$$

The input signal $u(t)$ is used to excite the piecewise affine system (36), then the actual input signal is plotted in Fig. 1(a). We use its approximated input signal to replace the actual input signal in our simulation, because it is not useful in practice, where the approximated input signal is similar to the sinusoidal signal in Fig. 1(b). Then we collect the output signal $y(t)$ by using some sensors, the observed output signal is plotted in Fig. 2.
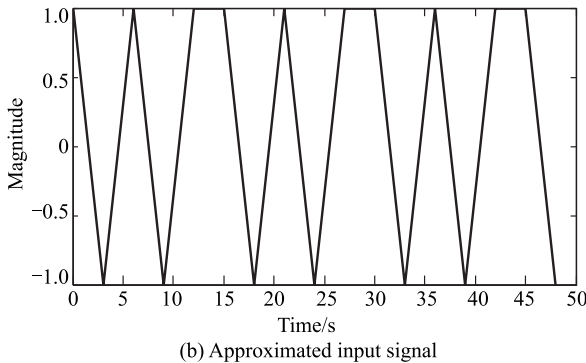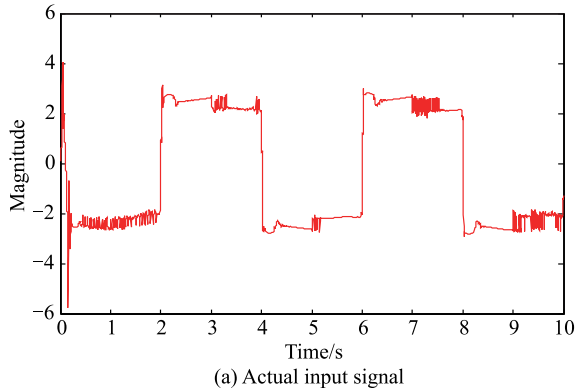

(a) Actual input signal

(b) Approximated input signal
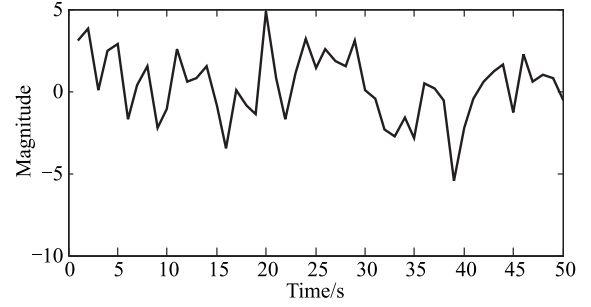
Fig. 1　Applied input signal

Fig. 2　Observed output signal

Firstly, our mentioned multi-class classification process is reduced to two class classification problems. After collecting one input-output data point $(y(t), \boldsymbol{\phi}(t))$, the region, which this data input-output point belongs to, is determined. In the whole simulation process, the number of given input-output data points is set to be $N = 500$, i.e., these 500 data points belong to one of the two classes. The detailed clustering process can be seen in Fig. 3, where the data points are clustered around two ellipsoids. As from three points deviate away these two ellipsoids, then they are regarded as outliers and they are deleted in our simulation. From Fig. 3, all data points are classified correctly, except those three data points.
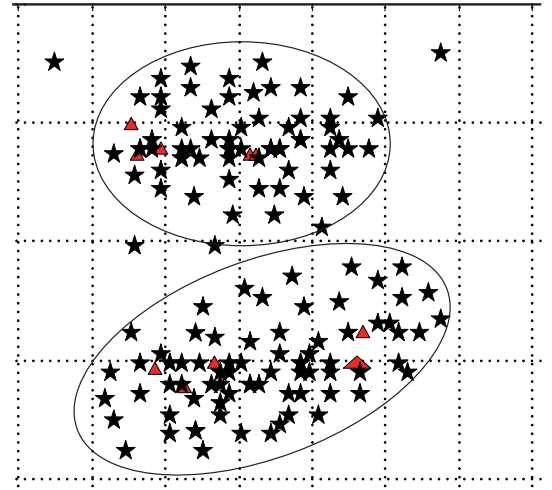

Fig. 3　Whole clustering process for estimated collected data

Secondly, in the presence of unknown but bounded external noise, choose its upper bound as

$$|e(t)| \leqslant \sigma = 0.5$$

and all initial parameter values are chosen as

$$\widehat{\boldsymbol{\theta}}_0 = \frac{1}{\boldsymbol{p}_0}\boldsymbol{I}.$$

The introduced zonotope parameter identification algorithm is used to identify those two unknown parameter vectors. Using the above six steps to construct a sequence of

candidate zonotopes iteratively, we see that after 20 iterations, these candidate zonotopes are given in Fig. 4 and Fig. 5.
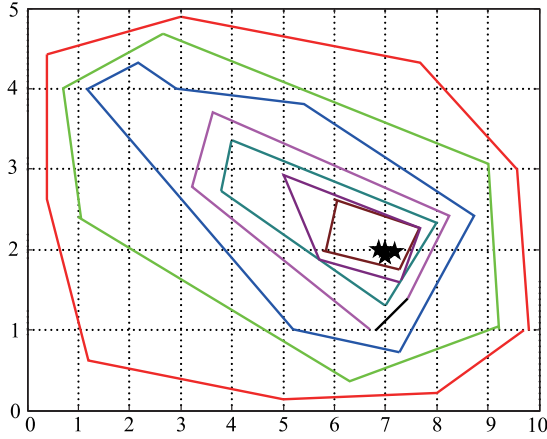

Fig. 4 Candidate zonotopes iteratively for the first optimal parameter vector
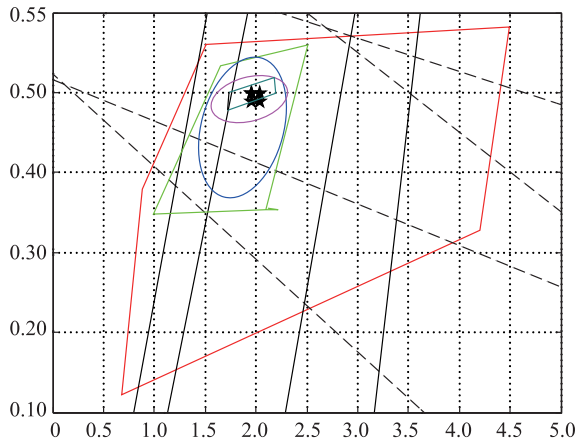

Fig. 5 Candidate zonotopes iteratively for the second optimal parameter vector

In Fig. 4, the black star denotes the optimal parameter vector to be $\boldsymbol{\theta}_1 = \begin{bmatrix} 7 & 2 \end{bmatrix}^{\mathrm{T}}$, and a sequence of candidate zonotopes are generated by the zonotope parameter identification algorithm, which include $\boldsymbol{\theta}_1 = \begin{bmatrix} 7 & 2 \end{bmatrix}^{\mathrm{T}}$ as their interior points. The volumes of these candidate zonotopes decrease with iterations, i.e., certain contracting properties are guaranteed. Generally, the other unknown parameter vector corresponding to $\boldsymbol{\theta}_1 = \begin{bmatrix} 7 & 2 \end{bmatrix}^{\mathrm{T}}$ can be chosen as the center of the smallest zonotope. Furthermore, the black star is the optimal parameter vector as $\boldsymbol{\theta}_2 = \begin{bmatrix} 2 & 0.5 \end{bmatrix}^{\mathrm{T}}$ in Fig. 5, and the results are similar to those in Fig. 4.

From Fig. 3, all data points are classified correctly, so Fig. 3 can be used to measure the performance of the multi-class classification process.

Fig. 4 and Fig. 5 are used to measure the performance of the zonotope parameter identification algorithm, and

the black star is the optimal parameter vector. After running the zonotope parameter identification algorithm, a sequence of zonotopes are obtained. Moreover, the optimal parameter vector is an interior point or a center of the smallest zonotope.

6. Conclusions

In this paper, we study the problem of identifying a special nonlinear system, the piecewise affine system, which combines linear and nonlinear properties. As this piecewise affine system is piecewise affine in the regression space, the parameter vector depends on the region in the whole considered regression space. The separated regions are determined as a multi-class classification problem, which can be solved by the classical first order algorithm from the convex optimization theory. In the presence of unknown but bounded external noise, the zonotope parameter identification algorithm is introduced to identify the unknown parameter vector in each separated region. Generally, the finite sample property of the zonotope parameter identification algorithm is our ongoing work.

References

[1] ROLL J, BEMPORAD A, LJUNG L. Identification of piecewise affine systems via mixed integer programming. Automatica, 2004, 40(1): 37 – 50.

[2] FERRARI-TRECATE G, MUSELLI M, LIBERATI D, et al. A clustering technique for the identification of piecewise affine systems. Automatica, 2003, 39(2): 205 – 217.

[3] JOULOSKI A L, WEILAND S, HEEMELS W P M H. A Bayesian approach to identification of hybrid systems. IEEE Trans. on Automatic Control, 2013, 50(10): 1520 – 1532.

[4] OHLSSON H, LJUNG L. Identification of switched linear regression models using sum of norms regularization. Automatica, 2013, 49(4): 1045 – 1050.

[5] LAUER F, BLOCH G, VIDAL R. A continuous optimization framework for hybrid system identification. Automatica, 2011, 47(3): 608 – 613.

[6] LAUER F. On the complexity of switching linear regression. Automatica, 2016, 74: 80 – 83.

[7] LAUER F. On the complexity of piecewise affine system identification. Automatica, 2015, 62: 148 – 153.

[8] BAKO L. Identification of switched linear systems via sparse optimization. Automatica, 2011, 47(4): 668 – 677.

[9] BRESCHI V, PIGA D, BEMPORAD A. Piecewise affine regression via recursive multiple least squares and multi-category discrimination. Automatica, 2016, 73: 155 – 162.

[10] BEMPORAD A, GARULLI A, PAOLETTI S, et al. A bounded error approach to piecewise affine system identification. IEEE Trans. on Automatic Control, 2005, 50(10): 1567 – 1580.

[11] PAOLETTI S, ROLL J, GARULLI A, et al. On the input-output representation of piecewise affine state space models. IEEE Trans. on Automatic Control, 2010, 55(1): 60 – 72.

[12] BAKO L, BOUKHAROUBA K, DUVIELLA E, et al. A recursive identification algorithm for switched linear/affine models. Nonlinear Analysis-Hybrid Systems, 2011, 5(2): 242 – 253.

[13] NAKADA H, TAKABA K, KATAYAMA T. Identification of

piecewise affine systems based on statistical clustering technique. Automatica, 2005, 41(5): 905 – 913.

[14] PAOLETTI S, JOULOSKI A L, FERRARI-TRECATE G, et al. Identification of hybrid systems: a tutorial. European Journal of Control, 2008, 13(2): 242 – 260.

[15] BEMPORAD A, FERRARI-TRECATE G, MORARI M. Observability and controllability of piecewise affine and hybrid systems. IEEE Trans. on Automatic Control, 2000, 45(10): 1864 – 1876.

Biographies

WANG Jianhong was born in 1980. He is a Ph.D. and a professor with the School of Engineering and Sciences, Tecnologico de Monterrey, Mexico. His current research interests are system identification, nonlinear system and control, and model predictive control.
E-mail: 9120180002@jxust.edu.cn