

Iterative procedures for identification of nonlinear interconnected systems

Eleni Pepona Paresh Date

September 17, 2007

Abstract

This work addresses the identification problem of a discrete-time nonlinear system composed by linear and nonlinear subsystems. Systems in this class will be represented by Linear Fractional Transformations. Iterative identification procedures are examined, that alternate between the estimation of the linear and the nonlinear components. The burden of identification falls naturally on the nonlinear subsystem, as techniques for identification of linear systems have long been established. Two approaches are examined. A point-wise identification of the nonlinearity, recently proposed in the literature, is applied and its advantages and drawbacks are outlined. An alternative procedure that employs piecewise affine approximation techniques is proposed. Numerical examples demonstrate the efficiency of the proposed algorithm.

Contents

Abstract	i
Table of Contents	ii
1 Introduction	1
2 Problem Formulation	3
3 Point-wise identification	7
3.1 Kalman filtering	8
3.1.1 The Kalman Filter	9
3.1.2 Kalman smoother	12
3.2 LFT identification algorithm	15
4 PWA identification	19
Conclusions	24
Bibliography	25

Chapter 1

Introduction

Whenever a linear model fails to explain a system accurately, nonlinear models are employed. Nonlinear system identification has been an active research field in the last few decades and a number of black-box identification approaches have been proposed in the literature. For an extensive overview the interested reader is referred to the survey paper [16].

However, real world systems often consist of interconnected linear and nonlinear components. The simplest and most commonly arising examples are the Hammerstein and Wiener systems, a serial connection of a nonlinear followed by a linear part, or vice versa, respectively. More complex networked structures are also encountered. A general framework for representing such interconnected systems based on the Linear Fractional Transformations (LFTs) was introduced in [3]. Exploiting networked structures in the identification process is expected to be beneficial in terms of complexity and accuracy of the identified model. In spite of this, identification of interconnected systems has received little attention so far.

In this work we address the identification problem for an LFT interconnection composed by a linear time-invariant system and a static nonlinearity. First we study the work of [6], where the linear part is assumed to be known and the authors propose a nonparametric, point-wise, estimation algorithm for the static nonlinear map of the interconnected system. We integrate this algorithm into an iterative procedure that interchanges between the identification of the linear and the nonlinear part. Iterative schemes have been extensively used in nonlinear system identification, see for example the seminal paper [12].

Further, we develop an alternative iterative scheme, based on Piece Wise

Affine (PWA) approximation of the static nonlinearity. Piece Wise Affine system identification has recently attracted a lot of attention due to the universal approximation properties of PWA maps [7, 2]. In PWA identification one needs to estimate a suitable partition of the regressors domain and the parameters of the affine submodels in each region of the partition. Recently proposed identification techniques can be divided into two classes. The ones that fix the number of affine submodels to be estimated a priori [5, 13, 15, 11, 8] and these that estimate it from the data [10, 4, 1]. An extensive overview on PWA system identification can be found in [14]. Here, we make use of the bounded-error identification technique proposed in [1] to approximate the system nonlinearity by a PWA map. Identification of the linear part is tackled via standard techniques [9]. Numerical examples demonstrate that the proposed iterative scheme is able to exploit the knowledge of the system interconnection structure effectively.

In Chapter 1 we formulate the identification problem rigorously. In Chapter 2 we summarize the point-wise identification approach of [3] and integrate it into an iterative scheme for the identification of an LFT system structure. Motivated by the limitations of this method, in Chapter 3 we develop a new iterative algorithm that employs the PWA identification technique of [1] for the approximation of the nonlinear component. Numerical examples demonstrate the suitability of the proposed method. We conclude this report with a discussion on the identifiability of the proposed algorithm and directions for future research.

Chapter 2

Problem Formulation

The problem of identifying a discrete time networked system composed by linear and nonlinear components is addressed in this project. Such systems are represented under a general framework by Linear Fractional Transformations (LFTs). LFTs are a very flexible structure that can be used to model any dynamical system consisting of interconnections of linear and nonlinear components. Moreover, any interconnection of two or more LFTs results in a new LFT, so LFT can be regarded as a generalized form of any dynamical system consisting of linear and nonlinear parts.

A Linear Fractional Transformation is of the form of Figure 2.1, where $\mathbf{u}_k \in \mathbb{R}^n$ and $\mathbf{y}_k \in \mathbb{R}^m$ are given (measured) input and output sequences, $\mathbf{e}_k \in \mathbb{R}^l$ is a noise term sequence, possibly of a known class. In Figure 2.1, \mathcal{L}

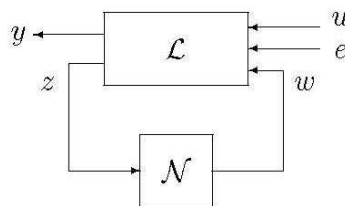


Figure 2.1: LFT form of a dynamical system

describes the overall linear dynamics of the system and can be partitioned as

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_{yu} & \mathcal{L}_{ye} & \mathcal{L}_{yw} \\ \mathcal{L}_{zu} & \mathcal{L}_{ze} & \mathcal{L}_{zw} \end{bmatrix}. \quad (2.1)$$

Moreover, \mathcal{N} is a nonlinear function, $\mathcal{N} : \mathbb{R}^q \rightarrow \mathbb{R}^p$ with $\mathcal{N}(\mathbf{z}) = \mathbf{w}$.

In [6] a method is proposed for the identification of the nonlinear map \mathcal{N} under the assumptions that the linear system is known, the input signal to the nonlinearity, z , is measurable and the nonlinear map \mathcal{N} is static. An estimate $\hat{\mathcal{N}}$ is then identified by specifying its input-output pairs (z, \hat{w}) such that $\hat{w} = \hat{\mathcal{N}}(z)$. That is, the method consists in point-wise estimation of the unknown signal w .

In this work we assume that the linear system belongs in a known class and its parameters are to be identified. Moreover, z is a regression vector which at each time point $k = 1, \dots, N$ consists of n_y past values of the output $\{y_k, k = 1, \dots, N\}$ and the current and n_u past values of the input $\{u_k, k = 1, \dots, N\}$, i.e.

$$\mathbf{z}_k = [\mathbf{y}_{k-1}^T \quad \dots \quad \mathbf{y}_{k-n_y}^T \quad \mathbf{u}_k^T \quad \dots \quad \mathbf{u}_{k-n_u}^T]^T, \quad (2.2)$$

hence, $z \in \mathbb{R}^q$ with $q = n_y m + (n_u + 1)n$. Finally, the nonlinear map is assumed to be unknown. In contrast to [6] we do not impose any restrictions on the structure of the nonlinear function \mathcal{N} . Rather we wish to approximate it by a piece wise affine (PWA) map so that we obtain an analytical form for the unknown nonlinearity. A PWA map is a function $g : \mathbb{R}^q \rightarrow \mathbb{R}^p$ of the form

$$g(\mathbf{z}) = \begin{cases} \phi^T \theta_1 & \text{if } \mathbf{z} \in \mathcal{X}_1 \\ \vdots & \vdots \\ \phi^T \theta_s & \text{if } \mathbf{z} \in \mathcal{X}_s \end{cases}$$

where $\phi = [\mathbf{z}^T \quad 1]^T$, $\theta_i \in \mathbb{R}^{q+1}$, $i = 1, \dots, s$ are the parameters of each affine submodel, $\mathcal{X} \subseteq \mathbb{R}^q$ is the set of domain of the regression vector \mathbf{z} and $\{\mathcal{X}_i, i = 1, \dots, s\}$ is a complete partition of \mathcal{X} . Each \mathcal{X}_i , $i = 1, \dots, s$ is assumed to be a convex polyhedron described by

$$\mathcal{X}_i = \{ \mathbf{z} \in \mathbb{R}^q : H_i^T \phi \preceq 0 \}, \quad (2.3)$$

where H_i , $i = 1, \dots, s$ are constant vectors in \mathbb{R}^{q+1} . In [1] an algorithm is proposed for the identification of PWA models, where the minimum number

s of submodels along with their parameters θ_i 's and the parameters of the convex polyhedra H_i 's, $i = 1, \dots, s$ are estimated. Our target is to identify the LFT structure in an iterative fashion interchanging between identification of the linear and nonlinear components, using standard techniques for the identification of the linear system, such as the prediction error method, and the PWA techniques of [1] for the identification of an approximating PWA map for the unknown nonlinearity.

For the identification process, we have at our disposal N pairs of input/output measurements $\{\mathbf{u}_k, \mathbf{y}_k\}_{k=1}^N$.

We consider the model

$$A(q)y_k = B(q)u_k + w_k + C(q)e_k \quad (2.4a)$$

$$w_k = \mathcal{N}(z_k) \quad (2.4b)$$

for $k = 1, \dots, N$, where $z_k, k = 1, \dots, N$ is given by (2.2), $p = l = m = n = 1$, $q = n_y + n_u + 1$ and $A(q)$, $B(q)$ and $C(q)$ are given by

$$A(q) = 1 - a_1q^{-1} - a_2q^{-2} - \dots - a_{n_a}q^{-n_a} \quad (2.5a)$$

$$B(q) = b_1 + b_2q^{-1} + \dots + b_{n_b}q^{-n_b} \quad (2.5b)$$

$$C(q) = 1 + c_1q^{-1} + \dots + c_{n_c}q^{-n_c} \quad (2.5c)$$

where the orders n_a , n_b and n_c are known. In this setting we then have

$$\begin{aligned} \mathcal{L}_{yu} &= \frac{B(q)}{A(q)} \\ \mathcal{L}_{yw} &= \frac{1}{A(q)} \\ \mathcal{L}_{ye} &= \frac{C(q)}{A(q)} \end{aligned}$$

This choice of \mathcal{L}_{yu} , \mathcal{L}_{yw} and \mathcal{L}_{ye} is not representative of the general case for the linear system \mathcal{L} . However this structure is preferred in terms of simplicity of the resulting model.

As discussed above, our target is to identify the system in an iterative fashion, interchanging between identification of linear and nonlinear parts. For the linear part the identification task reduces to the estimation of the linear model's parameters

$$\begin{aligned} \mathbf{a} &= [1 \quad a_1 \quad a_2 \quad \dots \quad a_{n_a}] \\ \mathbf{b} &= [b_1 \quad b_2 \quad \dots \quad b_{n_b}] \\ \mathbf{c} &= [1 \quad c_1 \quad \dots \quad c_{n_c}] \end{aligned}$$

Identification is performed using standard techniques like the prediction error method. The nonlinear function \mathcal{N} is of unknown structure and only its input sequence $\{\mathbf{z}_k\}_{k=1}^N$ is available to us. We explore two different approaches for the identification of the nonlinear component. First we implement the point-wise identification method of [6]. Then we integrate the PWA identification algorithm for nonlinear dynamical systems of [1] in the iterative scheme to obtain a PWA approximating map of the unknown nonlinearity. Finally, we conclude on the performance of these algorithms and we propose further directions for future investigation.

Chapter 3

Point-wise identification

Identification of the nonlinear component \mathcal{N} of a LFT structure is performed in [6]. There are several assumptions imposed in this work, the most important of which are that the linear system \mathcal{L} is known and the nonlinear map \mathcal{N} is static. An estimate $\hat{\mathcal{N}}$ is formed by specifying the pairs (z, \hat{w}) such that $\hat{w} = \hat{\mathcal{N}}(z)$. In this respect the identification problem reduces to an estimation problem, that is infer the internal signal w given the input-output sequences of the system (u, y) .

Remember that in general the LFT structure of Figure 2.1 is described by

$$y = \mathcal{L}_{yu}u + \mathcal{L}_{yw}w + \mathcal{L}_{ye}e \quad (3.1)$$

As in all estimation problems under this setting, given the input-output sequences (u, y) there may be many pairs of signals (\hat{e}, \hat{w}) that are consistent with (3.1). In [6], the assumption that the nonlinear operator is static is exploited to the development of a metric called the dispersion function $\mathcal{D}(z, w)$. This metric forces the estimation algorithm described below to prefer those signals (e, w) that are consistent with the static nature of the nonlinear map. The dispersion function $\mathcal{D}(z, w)$ is given by

$$\mathcal{D}(z, w) = N(\|\Delta\Pi_z z\|^2 + \|\Delta\Pi_z w\|^2), \quad (3.2)$$

where Π_z is a permutation operator that sorts z in an increasing order and Δ is a first order difference operator. In [6] it is proved that the dispersion of (z, w) is quadratic in w , i.e. there exists a positive definite matrix Q and a scalar r such that

$$\mathcal{D}(z, w) = w^T Q w + r. \quad (3.3)$$

To estimate the unknown signals (e, w) the following optimization problem is solved :

$$\begin{aligned} \min_{e,w} \quad & \|e\|^2 + \beta w^T Q w \\ \text{s.t.} \quad & (e, w) \in \mathbb{C}_{\mathcal{L}}(u, y) \end{aligned} \quad (3.4)$$

where $\mathbb{C}_{\mathcal{L}}(u, y) = \{(e, w) \in \mathbb{R}^l \times \mathbb{R}^p \mid y = \mathcal{L}_{yu}u + \mathcal{L}_{yw}w + \mathcal{L}_{ye}e\}$ and β is a weighting parameter that is used to trade-off between the error and the desired smoothness properties.

Problem (3.4) is a quadratic minimization problem with $2N$ unknowns to be estimated. In [6] a parameterization of the unknown signals (e, w) is proposed that reduces the complexity of the problem. All consistent signals $(e, w) \in \mathbb{C}_{\mathcal{L}}(u, y)$ are parameterized as

$$(e, w) = (e^0 + \mathcal{B}_e f, w^0 + \mathcal{B}_w f), \quad (3.5)$$

where f is a free signal and $[\mathcal{B}_e \ \mathcal{B}_w]$ is a null system of $[\mathcal{L}_{ye} \ \mathcal{L}_{ye}]$ such that $\mathcal{L}_{ye}\mathcal{B}_e + \mathcal{L}_{yw}\mathcal{B}_w = 0$. In (3.5), (e^0, w^0) is a particular solution of (3.1) which is computed using standard techniques such as the Kalman Filter or Kalman Smoother. Under the above parameterization the optimization problem (3.4) is reformulated as an unconstrained least squares problem :

$$\hat{f} = \arg \min_f \left\| \begin{bmatrix} e^0 \\ \sqrt{\beta} Q^{\frac{1}{2}} w^0 \end{bmatrix} + \begin{bmatrix} \mathcal{B}_e \\ \sqrt{\beta} Q^{\frac{1}{2}} \mathcal{B}_w \end{bmatrix} f \right\|^2 \quad (3.6)$$

so that $(\hat{e}, \hat{w}) = (e^0 + \mathcal{B}_e \hat{f}, w^0 + \mathcal{B}_w \hat{f})$.

In the sequel we explore the performance of the Kalman Filter and the Kalman Smoother as a mean of estimating the unknown sequence w for the model structure (2.4). Further we develop an iterative algorithm for the identification of the LFT model in the case where the linear system need also to be identified.

3.1 Kalman filtering

In this section we estimate the unknown sequence $\{w_k\}_{k=1}^N$, the output sequence of the nonlinear map \mathcal{N} of Figure 2.1. For simplicity of notation we consider the model structure (2.4) and we assume that $n_b \leq n_a$.

The input-output model system (2.4) is transformed into state-space form as follows : Model (2.4) is equivalent to

$$y_k = y_k^1 + y_k^2 \quad (3.7a)$$

$$y_k^1 = \frac{B(q)}{A(q)}u_k + \frac{1 - A(q)}{A(q)}w_k + \frac{C(q) - A(q)}{A(q)}e_k \quad (3.7b)$$

$$y_k^2 = w_k + e_k \quad (3.7c)$$

where the system (3.7b) is strictly proper and admits a canonical observation state space form as :

$$x_{k+1} = Fx_k + G \begin{bmatrix} u_k \\ e_k \\ w_k \end{bmatrix} \quad (3.8a)$$

$$y_k^1 = Hx_k \quad (3.8b)$$

where, $F = \begin{bmatrix} 0 & \dots & 0 & -a_{n_a} \\ 1 & \dots & 0 & -a_{n_a-1} \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 1 & -a_1 \end{bmatrix}$, $G = \begin{bmatrix} b_{n_a} & c_{n_a} - a_{n_a} & -a_{n_a} \\ \vdots & \vdots & \vdots \\ b_1 & c_1 - a_1 & -a_1 \end{bmatrix}$ and $H = [0 \dots 0 \ 1]$.

Thus, from (3.7) and (3.8) our system is given in state-space form by :

$$x_{k+1} = Fx_k + G \begin{bmatrix} u_k \\ e_k \\ w_k \end{bmatrix} \quad (3.9a)$$

$$y_k = Hx_k + [0 \ 1 \ 1] \begin{bmatrix} u_k \\ e_k \\ w_k \end{bmatrix} \quad (3.9b)$$

This description is used for the application of the Kalman Filter and Smoother.

3.1.1 The Kalman Filter

In order to estimate $\{w_k\}_{k=1}^N$, w_k needs to be included in the state of model (3.9). Thus, we define a new state $z_k = \begin{bmatrix} x_k \\ w_k \end{bmatrix}$ and the system is now described

by

$$z_{k+1} = \underbrace{\begin{bmatrix} F & G_3 \\ 0 & 0 \end{bmatrix}}_{\bar{A}} z_k + \underbrace{\begin{bmatrix} G_1 \\ 0 \end{bmatrix}}_{\bar{B}} u_k + \underbrace{\begin{bmatrix} G_2 & O \\ 0 & 1 \end{bmatrix}}_{\Gamma} \begin{bmatrix} e_k \\ w_{k+1} \end{bmatrix} \quad (3.10a)$$

$$y_k^1 = \underbrace{\begin{bmatrix} H & 1 \end{bmatrix}}_{\bar{C}} z_k \quad (3.10b)$$

where $G_i, i = 1, 2, 3$ is the i^{th} column of matrix G defined above. So system (3.7) has state-space representation

$$z_{k+1} = \bar{A}z_k + \bar{B}u_k + \Gamma \begin{bmatrix} e_k \\ w_{k+1} \end{bmatrix} \quad (3.11a)$$

$$y_k = \bar{C}z_k + e_k \quad (3.11b)$$

To estimate the state of system (3.11) we assume that the noise sequence is Gaussian with $e_k \sim N(0, \sigma_e^2)$ and we apply Kalman filtering with correlated noises. Let $\xi_k = \begin{bmatrix} e_k \\ w_{k+1} \end{bmatrix}$ then the covariance matrices to be used in Kalman filtering equations are :

$$Q = \mathbf{E}\xi_k\xi_l^T = \begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_w^2 \end{bmatrix} \delta_{kl} \quad (3.12a)$$

$$R = \mathbf{E}e_k e_l^T = \sigma_e^2 \delta_{kl} \quad (3.12b)$$

$$S = \mathbf{E}\xi_k e_l^T = \begin{bmatrix} \sigma_e^2 \\ 0 \end{bmatrix} \delta_{kl} \quad (3.12c)$$

Kalman filter equations

$$P_{0,0} = \text{Var}(x_0)$$

$$\hat{x}_{0,0} = \mathbf{E}x_0$$

$$K_k = \Gamma S R^{-1}$$

$$P_{k,k-1} = (\bar{A} - K_k \bar{C}) P_{k,k} (\bar{A} - K_k \bar{C})^T + \Gamma Q \Gamma^T - K_k R K_k^T$$

$$W_k = P_{k,k-1} \bar{C}^T (\bar{C} P_{k,k-1} \bar{C}^T + R)^{-1}$$

$$\hat{x}_{k,k-1} = \bar{A} \hat{x}_{k-1,k-1} + \bar{B} u_{k-1} + K_k (y_{k-1} - \bar{C} \hat{x}_{k-1,k-1})$$

$$\hat{x}_{k,k} = \hat{x}_{k,k-1} + W_k (y_k - \bar{C} \hat{x}_{k,k-1}) \text{ for } k = 1, \dots, N$$

u_k distribution	w_k distribution			
	N(0,1)	N(0,10)	N(0,100)	u_{k-1}^2
U(0,1)	22.70 %	61.58 %	86.90 %	29.37 %
U(-1,1)	21.17 %	58.24 %	86.12 %	0
N(0,10)	26.18 %	69.81 %	76.29 %	0

Table 3.1: Measure of fit between true and estimated values of w

Note that $\{\xi_k, k = 1, \dots, N\}$ is not a Gaussian process, therefore the Kalman filter is applied as the best available linear estimator.

Example 1

Consider the system

$$y_k + y_{k-1} + 0.2y_{k-2} - 0.05y_{k-3} = u_{k-1} + u_{k-2} + 0.2u_{k-3} + w_k + e_k - 0.5e_{k-1}, \quad (3.13)$$

where $e_k \sim N(0, 1)$. System (3.13) is of the form of system (2.4) with $\mathbf{a} = [1 \ 1 \ 0.2 \ -0.05]$, $\mathbf{b} = [0 \ 1 \ 1 \ 0.2]$ and $\mathbf{c} = [1 \ -0.5]$.

In order to check the performance of the Kalman filter in estimating the unknown sequence $\{w_k\}_{k=1}^N$ we experiment with different input distributions for $\{u_k\}_{k=1}^N$ and different distributions for $\{w_k\}_{k=1}^N$. As a measure of fit between the real values of $\{w_k\}_{k=1}^N$ and the estimated ones $\{\hat{w}_k\}_{k=1}^N$ we use the following

$$FIT(w, \hat{w}) = \left(1 - \frac{\|w - \hat{w}\|_2}{\|w - \bar{w}\|_2}\right)100\% \quad (3.14)$$

where $\|\cdot\|_2$ is the Euclidean norm and \bar{w} is the mean of the sequence $\{w_k\}_{k=1}^N$. The results are shown in Table 3.1.

We see that the distribution of the input signal $\{u_k\}_{k=1}^N$ does not play any significant role in the performance of the Kalman filter. In contrast the distribution of the signal to be estimated $\{w_k\}_{k=1}^N$ and the distribution of the noise sequence $\{e_k\}_{k=1}^N$ determine the quality of fit between the true and the estimated values of w_k . In cases where the variances of noise $\{e_k\}_{k=1}^N$ and input $\{u_k\}_{k=1}^N$ sequences satisfy $\frac{\sigma_w^2}{\sigma_e^2} \simeq 1$ the estimation of the input sequence is very poor and for $\frac{\sigma_w^2}{\sigma_e^2} \ll 1$ there is no fit between estimated and true input sequence. The quality of fit, however, increases for $\frac{\sigma_w^2}{\sigma_e^2} \gg 1$ as shown in Table

3.1.

Further, we compute the innovations process which gives us the estimates of the noise sequence $\{\hat{e}_k\}_{k=1}^N$. From (3.11)

$$\hat{e}_k = y_k - \bar{C}z_k. \quad (3.15)$$

So we have at our disposal estimates $\{\hat{w}_k\}_{k=1}^N$ and $\{\hat{e}_k\}_{k=1}^N$. Filtering these back through (3.13) we obtain estimates of the output sequence $\{\hat{y}_k\}_{k=1}^N$. For all of the cases in Table 3.1 the quality of fit is very satisfactory, $FIT(y, \hat{y}) \geq 95\%$.

Remember that our system is of the form (3.9b)

$$y_k = Hx_k + \underbrace{w_k + e_k}_{unknown} \quad (3.16)$$

and we use the Kalman filter to estimate the input sequence $\{w_k\}_{k=1}^N$. We conclude that when the variance of the noise and input sequences are of the same degree the filter cannot distinguish in between them, therefore resulting in bad estimates $\{\hat{w}_k\}_{k=1}^N$, whereas when the input signal is "richer", i.e. its variance is of higher degree than that of the noise sequence, the two signals are distinguishable and the filter produces better estimates $\{\hat{w}_k\}_{k=1}^N$.

◇

3.1.2 Kalman smoother

Because of the difficulties that arose in the estimation of w by use of the Kalman filter we now explore the Kalman smoother, since we have a priori knowledge of the input/output sequences.

The system is given in state-space form by (3.9)

$$\begin{aligned} x_{k+1} &= Fx_k + G_1u_k + [G_2 \ G_3] \begin{bmatrix} e_k \\ w_k \end{bmatrix} \\ y_k &= Hx_k + [1 \ 1] \begin{bmatrix} e_k \\ w_k \end{bmatrix} \end{aligned}$$

Let $\epsilon = \begin{bmatrix} e_k \\ w_k \end{bmatrix}$ and $B_u = G_1$, $B_\epsilon = [G_2 \ G_3]$, $D_\epsilon = [1 \ 1]$, so the system is given by

$$\mathcal{S} \begin{cases} x_{k+1} &= Fx_k + B_uu_k + B_\epsilon\epsilon_k \\ y_k &= Hx_k + D_\epsilon\epsilon_k \end{cases}$$

The Kalman filter equations in state / process innovation form are given by

$$\mathcal{K} \begin{cases} m_{k+1} &= (F + K_k H)m_k + B_u u_k - K_k y_k \\ \nu_k &= H m_k - y_k \end{cases}$$

where $Q = \mathbf{E} \epsilon_k \epsilon_k^T = \begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_w^2 \end{bmatrix}$ and

$$\begin{aligned} R_k &= \text{Cov}(x_{k+1}, x_{k+1}) = F P_k F^T + B_\epsilon Q B_\epsilon^T \\ S_k &= \text{Cov}(x_{k+1}, y_k) = F P_k H^T + B_\epsilon Q D_\epsilon^T \\ T_k &= \text{Cov}(y_k, y_k) = H P_k H^T + D_\epsilon Q D_\epsilon^T \end{aligned}$$

and

$$\begin{aligned} P_{k+1} &= R_k - S_k T_k^{-1} S_k^T, \text{ the Ricatti equation} \\ K_k &= -S_k T_k^{-1}, \text{ the Kalman gain} \end{aligned}$$

Then, the system $\mathcal{J} = \mathcal{KS}$ gives the Kalman filter in state-innovations / sequence-innovations form as

$$\mathcal{J} \begin{cases} (m-x)_{k+1} &= (F + K_k H)(m-x)_k - (B_\epsilon + K_k D_\epsilon) \epsilon_k \\ \nu_k &= H(m-x)_k - D_\epsilon \epsilon_k \end{cases}$$

i.e.

$$\mathcal{J} \sim \left[\begin{array}{c|c} F + K_k H & -(B_\epsilon + K_k D_\epsilon) \\ \hline H & -D_\epsilon \end{array} \right]$$

In order to estimate ϵ_k we need to find a right inverse system of \mathcal{J} , i.e. a system \mathcal{J}^* such that $\mathcal{J}\mathcal{J}^* = I$. We compute the right inverse of \mathcal{J} as in [17]. For the matrix D_ϵ there exists a right inverse $D_\epsilon^* = [d_1 \ d_2]$, such that $D_\epsilon D_\epsilon^* = I$ or $d_1 + d_2 = 1$, so that any choice of d_1 gives a right inverse of D_ϵ as

$$D_\epsilon^* = \begin{bmatrix} d_1 \\ 1 - d_1 \end{bmatrix} \quad (3.17)$$

The inverse system \mathcal{J}^* is given by

$$\mathcal{J}^* \sim \left[\begin{array}{c|c} F - B_\epsilon D_\epsilon^* H & -(B_\epsilon D_\epsilon^* + K_k) \\ \hline -D_\epsilon^* H & -D_\epsilon^* \end{array} \right]$$

	Distributions	Kalman Filter	Kalman Smoother
Case I	$e_k \sim N(0, 1)$ $w_k \sim N(0, 1)$	29.38%	10.76% $d_1 = 0.2$
			29.71% $d_1 = 0.5$
			19.00% $d_1 = 0.8$
Case II	$e_k \sim N(0, 1)$ $w_k \sim N(0, 10)$	58.25%	63.98% $d_1 = 0.1$
			57.94% $d_1 = 0.3$
			44.14% $d_1 = 0.5$
Case III	$e_k \sim N(0, 1)$ $w_k \sim N(0, 100)$	71.38%	77.78% $d_1 = 0.1$
			63.86% $d_1 = 0.3$
			46.14% $d_1 = 0.5$
Case IV	$e_k \sim N(0, 1)$ $w_k = u_{k-1}^2$	19.23%	12.01% $d_1 = 0.2$
			19.56% $d_1 = 0.5$
			11.18% $d_1 = 0.7$

Table 3.2: Kalman filter vs Kalman smoother. Measure of fit between true and estimated values of w

Kalman smoother equations

$$\xi_{k-1} = (F - B_\epsilon D_\epsilon^* H) \xi_k - (B_\epsilon D_\epsilon^* + K_k) \nu_k$$

$$\epsilon_k = -D_\epsilon^* H \xi_k - D_\epsilon^* \nu_k, \text{ for } k = N - 1, \dots, 1.$$

Example 1 (continued)

For the system (3.13) we apply the Kalman smoother to estimate the unknown input sequence $\{w_k\}_{k=1}^N$ and we compare the quality of fit attained by using the Kalman filter. Since we can choose any D_ϵ^* as in (3.17) we experiment for different D_ϵ^* 's as shown on Table 3.2.

The degree of freedom in choosing d_1 in (3.17) can be used to trigger the performance of the smoother. If the degree of $\frac{\sigma_\epsilon^2}{\sigma_w^2}$ is known and $\frac{d_1}{1-d_1}$ is chosen accordingly we always get a better quality of fit using the smoother. In cases where no such knowledge is available an immediate solution is to take $D_\epsilon = [0.5 \ 0.5]^T$. In such cases however, the quality of fit can be at best as good as the Kalman smoother's one (Case's I, IV).

In conclusion the use of the Kalman smoother is beneficial only in cases

where a priori knowledge for the distribution (variance) of the noise sequence and the unknown input sequence is available.

3.2 LFT identification algorithm

Under the assumption that the linear system \mathcal{L} of the LFT model of Figure 2.1 is known, point-wise estimation of the unknown input sequence w can be performed by solving the optimization problem (3.6). We now relax this assumption. Instead we explore a special structure of LFT given by

$$A(q)y_k = B(q)u_k + w_k + e_k \quad (3.18a)$$

$$w_k = \mathcal{N}(z_k) \quad (3.18b)$$

i.e. we consider the model structure (2.4) with $C(q) = 1$ for ease of computations. In (3.18), z_k is given by the regression vector (2.2) where we assume to know the orders n_y and n_u .

The system can be written in semi-regression form as

$$y_k = \theta^T \varphi_k + w_k + e_k, \text{ for } k = 1, \dots, N \quad (3.19)$$

where $\theta = [-a_1 \ \dots \ -a_{n_a} \ b_0 \ b_1 \ \dots \ b_{n_b}]^T$ is the unknown parameters' vector and $\varphi_k = [y_{k-1}^T \ \dots \ y_{k-n_a}^T \ u_k^T \ \dots \ u_{k-n_b}^T]^T$ is the regression vector. Let $\mathbf{y} = [y_1^T \ \dots \ y_N^T]^T$ and let \mathbf{u} , \mathbf{w} , \mathbf{e} and φ be defined accordingly. The identification algorithm works as follows :

Algorithm 3.1

Initialization : Set $j=0$, $\hat{\mathbf{w}}^0 = \mathbf{0}$.

Iteration step j : $j = 1, \dots, K$, K fixed.

Identification of the linear component

Given the estimated sequence $\hat{\mathbf{w}}^{j-1}$ estimate the linear part's parameters by solving the optimization problem

$$\hat{\theta}^j = \min_{\theta} \|\mathbf{e}\|^2 = \min_{\theta} \|\mathbf{y} - \hat{\mathbf{w}}^{j-1} - \theta^T \varphi\|^2 \quad (3.20)$$

which is a linear least squares problem.

Identification of the nonlinear component

Compute matrices F^j , B_u^j , B_e^j , H^j and D_e^j of system (\mathcal{S}).

Get Kalman Smoother estimates of $\epsilon_0^j = \begin{bmatrix} \mathbf{e}_0^j \\ \mathbf{w}_0^j \end{bmatrix}$

Solve optimization problem (3.6)

$$\hat{f} = \arg \min_f \left\| \begin{bmatrix} \mathbf{e}^0 \\ \sqrt{\beta} \mathbf{Q}^{\frac{1}{2}} \mathbf{w}^0 \end{bmatrix} + \begin{bmatrix} \mathcal{B}_e \\ \sqrt{\beta} \mathbf{Q}^{\frac{1}{2}} \mathcal{B}_w \end{bmatrix} \right\|^2$$

using parameter β as a tuning knob to trade off between identification error and the required smoothness properties.

Compute $\hat{\mathbf{w}}^j = \mathbf{w}_0^j + \mathcal{B}_w \hat{f}$.

The initialization of the algorithm with $\hat{\mathbf{w}}^0$ is not restrictive. This choice however forces the algorithm to fit at the first iteration a linear model that explains the input-output data as best as possible, i.e. as a first guess we assume the best case scenario, no nonlinearity exists.

Application of the above algorithm showed that a considerable number of iterations is needed in order for the identification of both linear and nonlinear parts to be satisfactory (see Example 2.1). This was attributed to the fact that at first iterations $\hat{\mathbf{w}}$ may not be well estimated and the linear least squares estimates in problem (3.6) take into account part of the nonlinear contribution in the system. A modification is thus applied in problem (3.6) as follows.

Algorithm 3.2

The unknown sequence $\{w_k\}_{k=1}^N$ can be decomposed as

$$w_k = \bar{w} + \tilde{w}_k, \text{ for } k = 1, \dots, N$$

where \bar{w} is the mean of $\{w_k\}_{k=1}^N$ and $\{\tilde{w}_k\}_{k=1}^N$ is a zero-mean sequence.

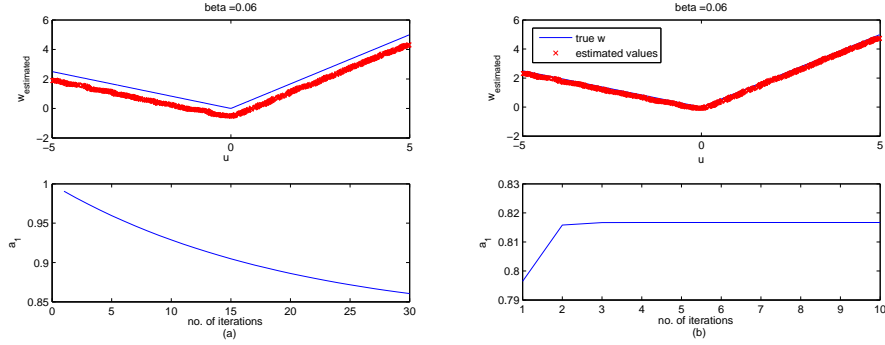


Figure 3.1: Estimated parameter of the linear model and point-wise approximated output.

Let \bar{w} be estimated along with the parameters of the linear model, so that $\phi = [\varphi^T \mathbf{1}]^T$ and $\vartheta = [\theta^T \bar{w}]^T$. Then the optimization problem for the linear part at iteration step j becomes

$$\hat{\vartheta}_j = \min_{\vartheta} \|\mathbf{e}\|^2 = \min_{\vartheta} \|\mathbf{y} - \hat{\mathbf{w}}_{j-1} - \vartheta^T \phi\|^2 \quad (3.21)$$

Under this formalization the iteration numbers required for the identification of the system is dramatically reduced. The following examples demonstrate these results.

Example 2.1

First we consider a system with piecewise linear nonlinearity as below.

$$y_k = 0.8y_{k-1} + w_k + e_k \quad (3.22a)$$

$$w_k = \begin{cases} u_{k-1} & , \text{ if } u_{k-1} \geq 0 \\ -0.5u_{k-1} & , \text{ if } u_{k-1} < 0 \end{cases} \quad (3.22b)$$

where u is uniformly distributed in $[-5, 5]$ and the noise is also uniformly distributed in $[-0.2, 0.2]$. In Figure 3.1.a we see the results of the initial algorithm. Indeed, after 30 iterations the parameter $a_1 = 0.8$ of the linear model is not well estimated and the nonlinearity although well recovered in shape is not correctly shifted. Applying the modification described above the algorithm converges in 4 iterations, as shown in Figure 3.1.b, which is a significant reduction. Although the nonlinearity is very well approximated,

the parameter of the linear component is not well identified, however we see that no improvement is possible. We note here that these are the best results obtained for several choices of the weighting parameter β .

Chapter 4

PWA identification

Having explored the performance and limitations of the point-wise identification algorithm we now propose a viable alternative for the identification of the nonlinear map \mathcal{N} . As discussed above we are interested in estimating a PWA function that approximates the nonlinear map. Recall that a PWA map is a function $g, g : \mathbb{R}^q \rightarrow \mathbb{R}^p$ of the form

$$g(\mathbf{z}) = \begin{cases} \phi^T \theta_1 & \text{if } \mathbf{z} \in \mathcal{X}_1 \\ \vdots & \vdots \\ \phi^T \theta_s & \text{if } \mathbf{z} \in \mathcal{X}_s \end{cases}$$

where $\phi = [\mathbf{z}^T \ 1]^T$, $\theta_i \in \mathbb{R}^{q+1}$, $i = 1, \dots, s$ are the parameters of each affine submodel, $\mathcal{X} \subseteq \mathbb{R}^q$ is the set of domain of the regressor vector \mathbf{z} and $\{\mathcal{X}_i, i = 1, \dots, s\}$ is a complete partition of \mathcal{X} . Each $\mathcal{X}_i, i = 1, \dots, s$ is assumed to be a convex polyhedron described by

$$\mathcal{X}_i = \{\mathbf{z} \in \mathbb{R}^q : H_i^T \phi \preceq 0\},$$

where H_i 's, $i = 1, \dots, s$ are constant vectors in \mathbb{R}^{q+1} . Again we consider the system (3.18)

$$\begin{aligned} A(q)y_k &= B(q)u_k + w_k + e_k \\ w_k &= \mathcal{N}(z_k) \end{aligned}$$

which is given in semi-regression form by (3.19)

$$y_k = \theta^T \varphi_k + w_k + e_k, \text{ for } k = 1, \dots, N.$$

The identification algorithm works as follows :

Algorithm 3.1

Initialization : Set $j=0$, $\hat{\mathbf{w}}^0 = \mathbf{0}$.

Iteration step j : $j = 1, \dots, K$, K fixed.

Identification of the linear component

Decompose $\{w_k\}_{k=1}^N$ as

$$w_k = \bar{w} + \tilde{w}_k, \text{ for } k = 1, \dots, N$$

where \bar{w} is the mean of $\{w_k\}_{k=1}^N$ and $\{\tilde{w}_k\}_{k=1}^N$ is a zero-mean sequence.

Given the estimated sequence $\hat{\mathbf{w}}^{j-1}$ solve the optimization problem

$$\hat{\vartheta}_j = \min_{\vartheta} \|\mathbf{e}\|^2 = \min_{\vartheta} \|\mathbf{y} - \hat{\mathbf{w}}_{j-1} - \vartheta^T \phi\|^2 \quad (4.1)$$

where $\phi = [\varphi^T \ \mathbf{1}]^T$ and $\vartheta = [\theta^T \ \bar{w}]^T$ as in Algorithm 2.2.

Identification of the nonlinear component

Compute $\mathbf{v}^j = \mathbf{y} - (\hat{\theta}^j)^T \varphi$. It holds

$$\mathbf{v}^j = \mathbf{w} + \epsilon^j \quad (4.2)$$

where ϵ^j describes the overall noise due to system's noise \mathbf{e} and the linear identification error term. Under the assumption that ϵ^j is bounded by some bound δ , estimate a minimum positive integer s , a set of parameter vectors $\{\theta_i\}_{i=1}^s$ and a polyhedral partition $\{\mathcal{X}_i\}_{i=1}^s$ of the regressor set \mathcal{X} such that the corresponding PWA function \hat{g} satisfies

$$|\mathbf{v}_k^j - \hat{g}(z_k)| \leq \delta, \text{ for all } k = 1, \dots, N. \quad (4.3)$$

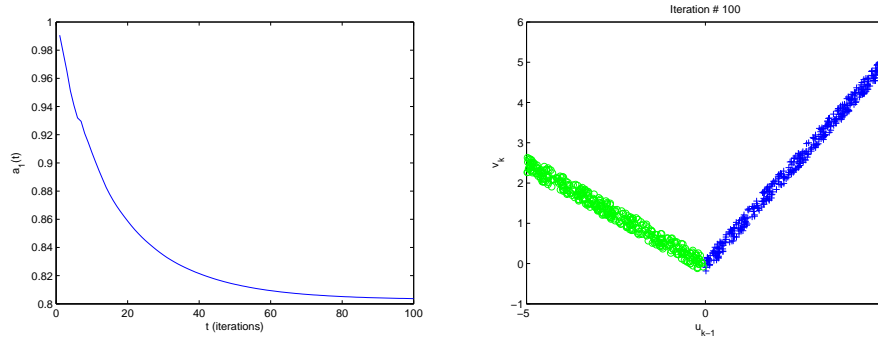


Figure 4.1: Estimated parameter of the linear model and approximated output by PWA model.

To solve this problem apply the identification algorithm of [1] to the data $(\mathbf{z}, \mathbf{v}^j)$.

Compute $\hat{\mathbf{w}}^j = \text{PWA}(\mathbf{z})$.

Example 3.1

Consider the system of Example 2.1.

$$\begin{aligned}
 y_k &= 0.8y_{k-1} + w_k + e_k \\
 w_k &= \begin{cases} u_{k-1} & , \text{ if } u_{k-1} \geq 0 \\ -0.5u_{k-1} & , \text{ if } u_{k-1} < 0 \end{cases}
 \end{aligned}$$

where u is uniformly distributed in $[-5, 5]$ and the noise is also uniformly distributed in $[-0.2, 0.2]$.

Estimating the linear component as in (3.6) we obtain the results of Figures 4.1 and 4.2. As shown, nearly 100 iterations are needed for the system to be identified satisfactorily. In Figure 4.2 the parameters γ_1 and γ_2 are the constant terms of the PWA function, which in the true system have nil value. However solving the least squares problem (4.1) instead, only 3 iterations are required which is a dramatic reduction. The respective results are shown in Figures 4.3 and 4.4.

◇

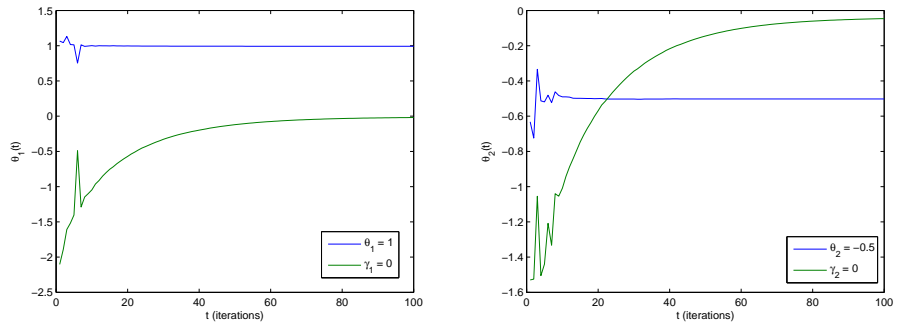


Figure 4.2: Estimated parameters for the PWA function.

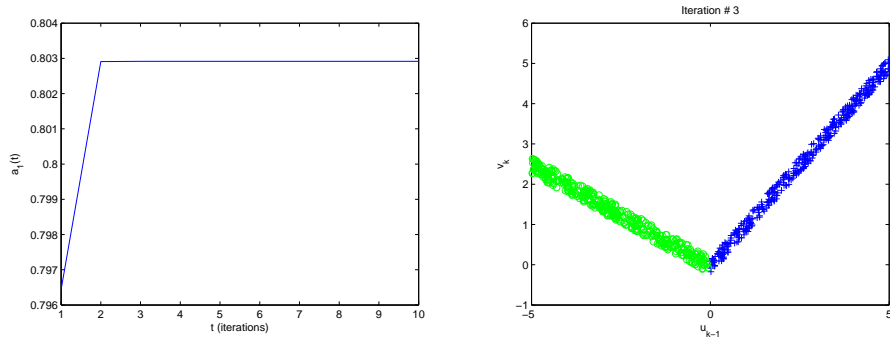


Figure 4.3: Estimated parameter of the linear model and approximated output by PWA model.

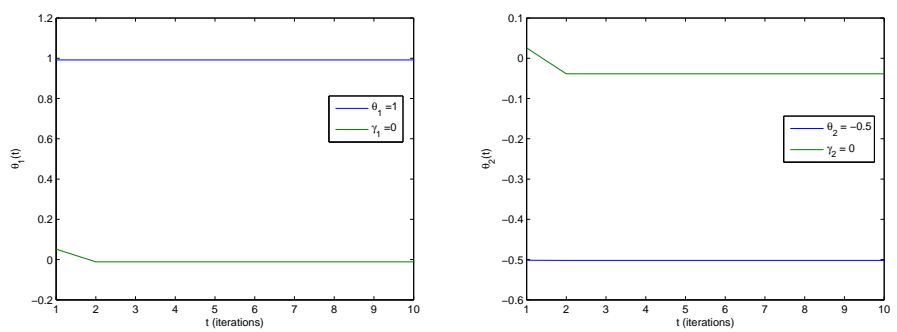


Figure 4.4: Estimated parameters for the PWA function.

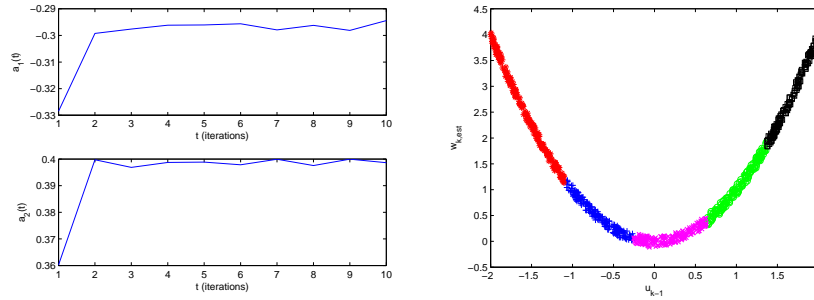


Figure 4.5: Estimated parameters of the linear model and approximated output \hat{w} by PWA function \hat{g} in Example 1.b .

Example 3.2

Now consider the Hammerstein system

$$y_k = -0.3y_{k-1} + 0.4y_{k-2} + w_k + e_k \quad (4.4a)$$

$$w_k = u_{k-1}^2 \quad (4.4b)$$

where u is uniformly distributed in $[-2, 2]$ and the noise is also uniformly distributed in $[-0.1, 0.1]$. In Figure 4.5 the estimated parameters of the linear system and the estimated sequence \hat{w} are shown. Although convergence is not as clear as in the previous example, the estimation again is satisfactory after 4 iterations. The nonlinearity in this example is a parabola in u_{k-1} and it is approximated by a PWA map that consists of 5 modes, as shown in Figure 4.5.

◇

We now continue to a more challenging situation.

Example 3.3

Let the system be given by

$$y_k = -0.49y_{k-2} + 0.5u_{k-1} - 1.5u_{k-2} + w_k + e_k \quad (4.5a)$$

$$w_k = \begin{cases} -0.4y_{k-1} + u_{k-1} + 1.5, & \text{if } 4y_{k-1} - u_{k-1} + 10 < 0 \\ 0.5y_{k-1} - u_{k-1} - 0.5, & \text{if } 5y_{k-1} + u_{k-1} - 10 \geq 0 \\ & \text{and } 5y_{k-1} - u_{k-1} - 6 \leq 0 \\ -0.3y_{k-1} + 0.5u_{k-1} - 1.7, & \text{if } 5y_{k-1} + u_{k-1} - 6 \geq 0 \end{cases} \quad (4.5b)$$

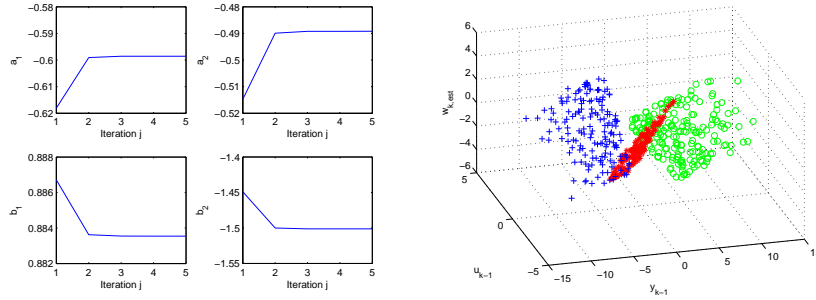


Figure 4.6: Estimated parameters of the linear model and approximated output \hat{w} by PWA function \hat{g} in Example 2 .

In this example $z_k = \begin{bmatrix} y_{k-1} \\ u_{k-1} \end{bmatrix}$. As shown in Figure 4.6 the algorithm converges in 5 iterations. However the structure turns out to be non-identifiable. Although the estimates of a_2 and b_2 , the parameters corresponding to y_{k-2} and u_{k-2} respectively, are very close to the true values, the estimated parameters corresponding to y_{k-1} and u_{k-1} , namely a_1 and b_1 , are quite different from the true ones. Since y_{k-1} and u_{k-1} take part in both linear and nonlinear maps, it is immediate that the linear model fitted to the data explains as linearly as possible the data. The nonlinearity in this example is of the form

$$w_k = \begin{cases} \alpha_1 y_{k-1} + \beta_1 u_{k-1} + \gamma_1, & H_1 \phi \leq 0 \\ \alpha_2 y_{k-1} + \beta_2 u_{k-1} + \gamma_2, & H_2 \phi \leq 0 \\ \alpha_3 y_{k-1} + \beta_3 u_{k-1} + \gamma_3, & H_3 \phi \leq 0 \end{cases}$$

which can be written as

$$w_k = \tilde{a}_1 y_{k-1} + \tilde{b}_1 u_{k-1} + \begin{cases} (\alpha_1 - \tilde{a}_1) y_{k-1} + (\beta_1 - \tilde{b}_1) u_{k-1} + \gamma_1, & H_1 \phi \leq 0 \\ (\alpha_2 - \tilde{a}_1) y_{k-1} + (\beta_2 - \tilde{b}_1) u_{k-1} + \gamma_2, & H_2 \phi \leq 0 \\ (\alpha_3 - \tilde{a}_1) y_{k-1} + (\beta_3 - \tilde{b}_1) u_{k-1} + \gamma_3, & H_3 \phi \leq 0 \end{cases}$$

in which case the model equation (4.5) can be written as

$$y_k = \underbrace{(a_1 + \tilde{a}_1)}_{\mathbf{a}_1} y_{k-1} + a_2 y_{k-2} + \underbrace{(b_1 + \tilde{b}_1)}_{\mathbf{b}_1} u_{k-1} + b_2 u_{k-2} + (w_k - \tilde{a}_1 y_{k-1} - \tilde{b}_1 u_{k-1}) + e_k \quad (4.6)$$

Indeed the iterative algorithm produces estimates for \mathbf{a}_1 and \mathbf{b}_1 .

◇

Conclusions

In this report we developed two iterative algorithms for the identification of LFT system structures. The linear part of such systems has been identified using standard techniques available in the literature, while for the identification of the nonlinear component we have employed a non-parametric identification method based on point-wise approximation and a parametric method that delivers PWA approximation. The non-parametric method has turned out to be inadequate for our purposes. The PWA one however has produced promising results.

There are many issues to be addressed. One is whether the use of the PWA iterative procedure is beneficial compared to the use of the Piecewise Autoregressive Exogenous (PWARX) model to identify the system as a whole, i.e. without distinguishing between linear and nonlinear components, as described in [1]. It is our belief that the separate estimation procedure may reduce considerably the number of PWA submodels in the system. Another point of interest is how the method behaves when the linear part is described by a state-space model instead of an input/output one. Convergence results are usually ambiguous on iterative procedures, nevertheless it is an area worth to explore. Finally, issues of identifiability of the LFT structure arise in more complex systems as demonstrated in Example 3.3. Therefore, rigorous identifiability conditions need to be established.

Bibliography

- [1] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino. A bounded-error approach to piecewise affine system identification. *IEEE Transactions on Automatic Control*, 50(10):1567–1580, 2005.
- [2] L. Breiman. Hinging hyperplanes for regression, classification, and function approximation. *IEEE Transactions on Information Theory*, 39(3):999–1013, 1993.
- [3] M. S. Claassen. *System identification for structured nonlinear systems*. PhD thesis, Department of Mechanical Engineering, University of California at Berkeley, USA, 2001.
- [4] G. Ferrari-Trecate and M. Muselli. Single-linkage clustering for optimal classification in piecewise affine regression. In *Proc. of the IFAC Conference on Analysis and Design of Hybrid Systems*, Saint-Malo, France, 2003.
- [5] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari. A clustering technique for the identification of piecewise affine systems. *Automatica*, 39(2):205–217, 2003.
- [6] K. Hsu, T. Vincent, C. Novara, M. Milanese, and K. Poola. Identification of nonlinear maps in interconnected systems. In *Proc. of the 44th IEEE Conference on Decision and Control and the European Control Conference*, pages 6430–6435, Seville, Spain, 2005.
- [7] J.-N. Lin and R. Unbehauen. Canonical piecewise-linear approximations. *IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications*, 39(8):697–699, 1992.

- [8] A. Lj. Juloski, S. Wieland, and W. P. M. H. Heemels. A Bayesian approach to identification of hybrid systems. *IEEE Transactions on Automatic Control*, 50(10):1520–1533, 2005.
- [9] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, NJ, 1999.
- [10] E. Münz and V. Krebs. Identification of hybrid systems using apriori knowledge. In *Proc. of the 15th IFAC World Congress*, Barcelona, Spain, 2002.
- [11] H. Nakada, K. Takaba, and T. Katayama. Identification of piecewise affine systems based on statistical clustering technique. *Automatica*, 41(5):905–913, 2005.
- [12] K. S. Narendra and P. G. Gallman. An iterative method for the identification of nonlinear systems using a Hammerstein model. *IEEE Transaction on Automatic Control*, 11(3):546–550, 1966.
- [13] J. Ragot, G. Mourot, and D. Maquin. Parameter estimation of switching piecewise linear systems. In *Proc. of the 42nd IEEE Conference on Decision and Control*, pages 5783–5788, Maui, HI, 2003.
- [14] J. Roll. *Local and piecewise affine approaches to system identification*. PhD thesis, Department of Electrical Engineering, Linköping University, Linköping, Sweden, 2003.
- [15] J. Roll, A. Bemporad, and L. Ljung. Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40(1):37–50, 2004.
- [16] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P-Y. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: A unified overview. *Automatica*, 31(12):1691–1724, 1995.
- [17] K. Zhou and J.C. Doyle. *Essentials of Robust Control*. Prentice Hall, Upper Saddle River, New Jersey, 1998.