Automatica 45 (2009) 1359–1370

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Asymptotically optimal orthonormal basis functions for LPV system identification☆

Roland Tóth*, Peter S.C. Heuberger, Paul M.J. Van den Hof

Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 27 February 2008
Received in revised form
29 June 2008
Accepted 22 January 2009
Available online 10 March 2009

Keywords:
Linear parameter-varying systems
System identification
Orthonormal basis functions
Pole clustering

ABSTRACT

A global model structure is developed for parametrization and identification of a general class of Linear Parameter-Varying (LPV) systems. By using a fixed orthonormal basis function (OBF) structure, a linearly parametrized model structure follows for which the coefficients are dependent on a scheduling signal. An optimal set of OBFs for this model structure is selected on the basis of local linear dynamic properties of the LPV system (system poles) that occur for different constant scheduling signals. The selected OBF set guarantees in an asymptotic sense the least worst-case modeling error for any local model of the LPV system. Through the fusion of the Kolmogorov n-width theory and Fuzzy c-Means clustering, an approach is developed to solve the OBF-selection problem for discrete-time LPV systems, based on the clustering of observed sample system poles.

1. Introduction

In general, many physical systems and control problems exhibit parameter variations due to non-stationary or nonlinear behavior or dependence on external variables, such as space coordinates, in particular found in servo-mechanical applications. These systems vary in size and complexity, but they share the common need for accurate and efficient control of the relevant process variables. However, accurate modeling of such systems is in general a complex and tedious task, involving the use of nonlinear differential equations, leading to models with many parameters and high computational complexity. For processes with mild nonlinearities or dependence on external variables, the theory of *Linear Parameter-Varying* (LPV) systems offers an attractive modeling framework (Rugh & Shamma, 2000). Discrete-time LPV systems are generally described in either a *State-Space* (SS) or an *Input/Output* (I/O) representation (Tóth, Felici, Heuberger, & Van den Hof, 2007), where the parameters are functions of a time-varying *scheduling signal* $p(k) : \mathbb{Z} \to \mathbb{P}$, that schedules between local *Linear Time Invariant* (LTI) behaviors of the system. The compact set $\mathbb{P} \subset \mathbb{R}^{n_p}$ denotes the *scheduling space*. Practical use of the LPV framework is stimulated by the fact that control design for LPV systems is well worked out. For this class of systems, application of LTI control theory via *gain scheduling* (Rugh & Shamma, 2000) and LPV control synthesis techniques like μ*-synthesis* (Zhou & Doyle, 1998) or *Linear Matrix Inequality* (LMI)-based optimal control (Scherer, 1996) offer fast and reliable controller design, proved by a wide range of applied LPV control solutions from aerospace applications (Marcos & Balas, 2004) to CD players (Dettori & Scherer, 2001). However, it still remains a problem how to develop LPV models in a systematic fashion.

Recently several methods have been worked out, aiming at global identification of discrete-time LPV models from given measured data. This comprises methods based on multiple-model approaches (Murray-Smith & Johansen, 1997; Steinbuch, van de Molengraft, & van der Voort, 2003; Wassink, van de Wal, Scherer, & Bosgra, 2004), set-membership methods (Mazzaro, Movsichoff, & Pena, 1999; Milanese & Vicino, 1991), subspace techniques (dos Santos, Ramos, & de Carvalho, 2007; Felici, van Wingerden, & Verhaegen, 2006, 2007; Verdult & Verhaegen, 2002), basis functions (Tóth, Heuberger, & Van den Hof, 2007), LMI-based optimization (Sznaier, Mazzaro, & Inanc, 2000), simple *Least Mean Squares* (LMS) approaches (Giarré, Bauso, Falugi, & Bamieh, 2006; Wei & Del Re, 2006), and parameter estimation based gradient searches (Lee & Poolla, 1996; Verdult, Ljung, & Verhaegen, 2002). Most of these approaches build on the fact that an LPV system $\mathcal{S}$ can always be viewed as a collection of "local" behaviors and p-dependent weighting functions, i.e. *scheduling functions* that

☆ The material in this paper was partially presented at 45th IEEE Conf. on Decision and Control. 2006 Dec., San Diego, California, USA. This paper was recommended for publication in revised form by Associate Editor Brett Ninness under the direction of Editor Torsten Söderström.

* Corresponding author. Tel.: +31 (0) 631382635; fax: +31 (0)15 2786679.
E-mail addresses: tothrola@ieee.org, r.toth@tudelft.nl (R. Tóth), p.s.c.heuberger@tudelft.nl (P.S.C. Heuberger), p.m.j.vandenhof@tudelft.nl (P.M.J. Van den Hof).

0005-1098/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2009.01.010

schedule between them (Rugh & Shamma, 2000; Tóth, Heuberger et al., 2007). For any constant scheduling signal: $p(k) = \bar{\mathrm{p}}$ for all $k \in \mathbb{Z}$ where $\bar{\mathrm{p}} \in \mathbb{P}$, the LPV system $\mathcal{S}$ is identical to an LTI system $\mathcal{F}_{\bar{\mathrm{p}}}$. Thus, the set of local behaviors of $\mathcal{S}$ is given as $\mathfrak{F}_{\mathbb{P}} = \{\mathcal{F}_{\bar{\mathrm{p}}}\}_{\bar{\mathrm{p}}\in\mathbb{P}}$. The p-dependent scheduling function set, that schedules on $\mathfrak{F}_{\mathbb{P}}$, is denoted by $\mathfrak{H}_{\mathbb{P}} = \left\{h_{\bar{\mathrm{p}}}(.)\right\}_{\bar{\mathrm{p}}\in\mathbb{P}}$.

Identification of $\mathfrak{F}_{\mathbb{P}}$ is commonly accomplished in a sampled sense by LTI identification of $\mathcal{S}$ for a set of constant scheduling signals, associated with (for instance equidistant) points in the scheduling space $\mathbb{P}$. Then, assuming that the scheduling functions $\{h_{\bar{\mathrm{p}}}\}$ have a particular structure of dependence, like polynomial, an interpolation problem is formulated on $\mathbb{P}$ to give a global approximation of $\mathcal{S}$. Recently it was exposed that this approach should be handled with care for several reasons (Tóth, Felici et al., 2007; Tóth, Heuberger et al., 2007). In Tóth, Felici et al. (2007) it was shown that for general discrete-time LPV systems each $h_{\bar{\mathrm{p}}}$ is a function of time-shifted versions of p (dynamic dependence). Then, if the particular interpolation structure of $\{h_{\bar{\mathrm{p}}}\}$ is chosen to be too simple (dependence only on $p(k)$ (static dependence), linear dependence, etc.) the interpolation based on state-space or I/O model parametrization can result in significantly different models (Tóth, Felici et al., 2007). An additional concern of interpolation is that the McMillan degree of the local systems $\{\mathcal{F}_{\bar{\mathrm{p}}}\}$ may vary for different values of $\bar{\mathrm{p}} \in \mathbb{P}$. This shows that the choice of an easily interpolatable model structure which can incorporate aspects of dynamical dependence and local order changes is a crucial point of this identification approach.

The *Orthonormal Basis Function* (OBF)-based model representation offers such a structure with a well worked-out theory in the context of LTI system approximation and identification (Heuberger, Van den Hof, & Wahlberg, 2005). The basis functions, that provide bases for the system space $\mathcal{H}_2$ (Hilbert space of complex functions that are squared integrable on the unit circle), are generated by a cascaded network of stable all-pass filters, whose pole locations represent the prior knowledge about the system at hand. This approach characterizes the transfer function of a strictly proper LTI system as

$$F(z) = \sum_{i=1}^{\infty} w_i \phi_i(z), \tag{1}$$

where $\{w_i\}_{i=1}^{\infty}$ is the set of coefficients and $\Phi_\infty = \{\phi_i\}_{i=1}^{\infty}$ represents the sequence of OBFs. This implies that every $\mathcal{F}_{\bar{\mathrm{p}}} \in \mathfrak{F}_{\mathbb{P}}$ can be represented as a linear combination of a given Φ_∞, i. e. $\mathfrak{F}_{\mathbb{P}} \subset \operatorname{span}\{\Phi_\infty\}$. In practice, only a finite number of terms is used in (1), like in *Finite Impulse Response* (FIR) models. In contrast with FIR structures, the OBF parametrization can achieve almost zero modeling error with a relatively small number of parameters, due to the infinite impulse response characteristics of the basis functions. In this way, it is always possible to find a finite $\Phi_n \subset \Phi_\infty$, with a relatively small number of functions $n \in \mathbb{N}$, such that the representation error for all $\mathcal{F}_{\bar{\mathrm{p}}}$ is negligible. Using this idea in the time-domain (substitution of z with the forward time-shift operator q), it is possible to prove that LPV systems also have a series expansion representation in terms of LTI basis functions, but with coefficients $\{w_i\}_{i=1}^{\infty}$ dependent on p. Thus in terms of a finite OBF set $\Phi_n \subset \Phi_\infty$, the following approximation of the I/O map of $\mathcal{S}$ can be introduced:

$$y \approx \sum_{i=1}^{n} w_i(p)\phi_i(q)\, u, \tag{2}$$

where $\{w_i\}_{i=1}^{n}$ is a set of *coefficient functions*, with dynamic dependence on p. Note that in this structure, Φ_n gives the basis set used to approximate each element of $\mathfrak{F}_{\mathbb{P}}$ while $\{w_i\}_{i=1}^{n}$ describes the scheduling functions $\mathfrak{H}_{\mathbb{P}}$. Thus for a given $\Phi_n = \{\phi_i\}$, identification of the LPV system based on (2) simplifies to the identification of the scheduling functions. Assuming static dependence of $\{w_i\}_{i=1}^{n}$, such a task can be accomplished via two approaches:

- **Local approach:** Identify some $\mathcal{F}_{\bar{\mathrm{p}}} \in \mathfrak{F}_{\mathbb{P}}$ for constant $p(t) = \bar{\mathrm{p}}$ with the LTI OBF model structure

$$\hat{y} = \sum_{i=1}^{n} r_{\bar{\mathrm{p}},i}\phi_i(q)\, u. \tag{3}$$

  Based on a chosen functional dependence, e.g. polynomial, interpolate the resulting $\{r_{\bar{\mathrm{p}},i}\}$ for an estimate of $\{w_i\}_{i=1}^{n}$ in (2), such that $w_i(\bar{\mathrm{p}}) = r_{\bar{\mathrm{p}},i}$.
- **Global approach:** Parametrize the functional dependence of $\{w_i\}_{i=1}^{n}$ linearly (e.g. polynomial). Then for a data record with varying p, the estimation of the parameters of $\{w_i\}_{i=1}^{n}$ reduces to linear regression based on (2) in a least-squares prediction error setting.

There are many beneficial properties of the structure (2). For instance, the obtained model simplifies control design (see Section 6) and this parametrization is not affected by local order changes. The problem that remains to be solved with the proposed OBF-based identification approaches is to choose the set of OBFs Φ_n, "sufficiently rich" to describe $\mathfrak{F}_{\mathbb{P}}$ with a predefined number of functions. Seeking the solution for this problem is the purpose of the present paper.

Even in the case of LTI systems, the choice of OBFs to approximate a given system $\mathcal{F}$ in an "optimal" sense (based on some error measure) is a highly non-trivial task (Heuberger et al., 2005). For the LTI case, already quite some effort has been put into tackling the basis function selection problem resulting in methods of nonlinear optimization (Heuberger et al., 2005) and iterative search (Bodin, Villemoes, & Wahlberg, 1997). One of the concepts used for this purpose is the *Kolmogorov n-width* (K*n*W) theory for OBFs (Oliveira e Silva, 1996), which establishes optimality in the sense of the worst-case modeling error for any LTI system with pole locations in a given region of the complex plane. Denote by

$$\Omega_{\mathbb{P}} = \{\lambda \in \mathbb{C} \mid \lambda \text{ is a pole of } \mathcal{F}_{\bar{\mathrm{p}}} \in \mathfrak{F}_{\mathbb{P}} \text{ for } \bar{\mathrm{p}} \in \mathbb{P}\},$$

the collection of all pole locations belonging to the local behaviors of the LPV system $\mathcal{S}$. Then, based on $\Omega_{\mathbb{P}}$, the K*n*W theory can be evidently applied (e.g. by the approach of Heuberger et al. (2005)) to solve the optimal selection of OBFs with respect to $\mathfrak{F}_{\mathbb{P}}$. However, this approach is not applicable if $\mathfrak{F}_{\mathbb{P}}$ is unknown. This underlines the need for a mechanism that guarantees optimality of the OBF selection (selection of Φ_n) based on the available information.

In this paper, we assume as a starting point that a collection of pole locations, some samples of $\Omega_{\mathbb{P}}$, is available that are obtained from local linear behaviors of the LPV system $\mathcal{S}$. This set of pole samples $\bar{\Omega} \subset \Omega_{\mathbb{P}}$ can result – but not necessarily – from identification of the related local linear models. Based on $\bar{\Omega}$, we aim at the derivation of a basis function selection mechanism, that is capable of accomplishing the following objectives:

- Reconstruction of $\Omega_{\mathbb{P}}$ from $\bar{\Omega}$.
- Determination of the set of OBFs, which has the least possible worst-case modeling error for any LTI system with pole locations in $\Omega_{\mathbb{P}}$, therefore for all $\mathcal{F}_{\bar{\mathrm{p}}} \in \mathfrak{F}_{\mathbb{P}}$.

This choice of model structure leads to the local and global identification methods. The proposed method is the joint application of the K*n*W theory and *Fuzzy c-Means* (F*c*M) clustering (Jain & Dubes, 1988). The contribution of this method is to provide a practical model structure selection tool for the local and global LPV identification methods based on globally fixed OBFs. Earlier work along this line is proposed in Tóth, Heuberger, and Van den Hof (2006a), Tóth, Heuberger, and Van den Hof (2006b) and Vergeer (2005).

The paper is organized as follows: Section 2 introduces the description and properties of OBFs while Section 3 describes the n-width result with respect to these functions; in Section 4, the mechanism of the KnW-based FcM pole clustering is given that solves simultaneously the determination of $\Omega_{\mathbb{P}}$ from sampled poles and the selection of optimal OBFs with respect to $\Omega_{\mathbb{P}}$; in Section 5, the OBF-based LPV system identification scheme is specified to provide a brief description how the selected basis functions are used in an identification scenario; in Section 6, the applicability of the introduced method is shown through an example; and finally, in Section 7, the main results of the paper are discussed.

2. Orthonormal basis functions

We consider only the case of real rational (finite-dimensional) discrete-time, SISO transfer functions. For details see Heuberger et al. (2005), Heuberger, Van den Hof, and Bosgra (1995) and Ninness and Gustafsson (1997). Let $G_0 \equiv 1$ and $\{G_i\}_{i=1}^{\infty}$ be a sequence of inner functions (i.e. stable transfer functions with $G_i(z)G_i(\frac{1}{z}) = 1$), and let $\{A_i, B_i, C_i, D_i\}$ be minimal balanced SS representations of G_i. Let $\{\xi_1, \xi_2, \ldots\}$ denote the collection of all poles of the inner functions $G_1, G_2, \ldots$. Under the (completeness) condition that $\sum_{i=1}^{\infty}(1 - |\xi_i|) = \infty$, the scalar elements of the sequence of vector functions

$$V_n(z) = (zI - A_n)^{-1}B_n \prod_{i=0}^{n-1} G_i(z), \tag{4}$$

constitute a basis for $\mathcal{H}_{2-}(\mathbb{E})$, the *Hardy space* of functions, which are 0 for $z = \infty$, analytic on $\mathbb{E}$, the exterior of the unit disk $\mathbb{D}$, and squared integrable on the unit circle $\mathbb{T}$ with norm $\|.\|_{\mathcal{H}_2}$. In this way $\mathcal{H}_{2-}(\mathbb{E})$ is the space of all stable strictly proper transfer functions. These functions (4) are often referred to as the *Takenaka–Malmquist functions*. The special cases when all G_i are equal, i.e. $G_i(z) = G_{\mathrm{b}}(z)$, $\forall i > 0$, where G_{b} has McMillan degree $n_{\mathrm{b}} > 0$, are known as *Hambo functions* or *generalized orthonormal basis functions* (GOBFs) for arbitrary n_{b}, *2-parameter Kautz functions* for $n_{\mathrm{b}} = 2$, and as *Laguerre functions* for $n_{\mathrm{b}} = 1$. Note that for these cases the completeness condition is always fulfilled. In the remainder we will only consider the set of Hambo functions. Let G_{b} be an inner function with McMillan degree $n_{\mathrm{b}} > 0$ and minimal balanced SS representation $\{A_{\mathrm{b}}, B_{\mathrm{b}}, C_{\mathrm{b}}, D_{\mathrm{b}}\}$. Define $V_1(z) = (zI - A_{\mathrm{b}})^{-1}B_{\mathrm{b}}$ and $\phi_j = [V_1]_j$, $j \in \mathbb{I}_1^{n_{\mathrm{b}}}$, where $\mathbb{I}_{s_1}^{s_2} = \{s_1, s_1 + 1, \ldots, s_2\} \subset \mathbb{Z}$ is the index set. The Hambo basis then consists of the functions $\Phi_{n_{\mathrm{b}}}^{\infty} = \{\phi_j G_{\mathrm{b}}^i\}_{j=1,\ldots,n_{\mathrm{b}}}^{i=0,\ldots,\infty}$. An important aspect of these bases is that the inner function G_{b} is, modulo the sign, completely determined by its poles $\Xi_{n_{\mathrm{b}}} := \{\xi_1, \ldots, \xi_{n_{\mathrm{b}}}\}$:

$$G_{\mathrm{b}}(z) = \pm \prod_{j=1}^{n_{\mathrm{b}}} \frac{1 - z\xi_j^*}{z - \xi_j}, \tag{5}$$

where $*$ denotes complex conjugation, and it is immediate that the function V_1 has the same poles. Any $F \in \mathcal{H}_{2-}(\mathbb{E})$ can be decomposed as

$$F(z) = \sum_{i=0}^{\infty} \sum_{j=1}^{n_{\mathrm{b}}} w_{ij}\phi_j(z)G_{\mathrm{b}}^i(z), \tag{6}$$

and it can be shown that the rate of convergence of this series expansion is bounded by $\rho = \max_k |G_{\mathrm{b}}(\lambda_k^{-1})|$, called the *decay rate*, where $\{\lambda_k\}$ are the poles of $F(z)$. In the "best" case, where the poles of F are the same (with multiplicity) as the poles of G_{b}, only the terms with $i = 0$ in (6) are non-zero. The I/O relation of the OBF parametrization (6) is illustrated in Fig. 1.

In practice, only a finite number of terms $\Phi_{n_{\mathrm{b}}}^{n_{\mathrm{e}}} = \{\phi_j(z)G_{\mathrm{b}}^i\}_{j=1,\ldots,n_{\mathrm{b}}}^{i=0,\ldots,n_{\mathrm{e}}}$ with $n_{\mathrm{e}} \geq 0$ is used in (6), like in *Finite Impulse Response* (FIR) models. In contrast with FIR structures, which are described by (6) as a finite linear combination of *pulse basis functions* $\phi_1(z) = z^{-1}$ with $n_{\mathrm{b}} = 1$ and $G_{\mathrm{b}}^i(z) = z^{-i}$, the OBF parametrization uses a broad class of basis functions with infinite impulse responses. Therefore, OBF parametrization can achieve almost zero modeling error with a relatively small number of parameters due to the faster convergence of the series representation than in the FIR case. Moreover, the reduced number of parameters in the structure results in decreased variance of the final model estimate.

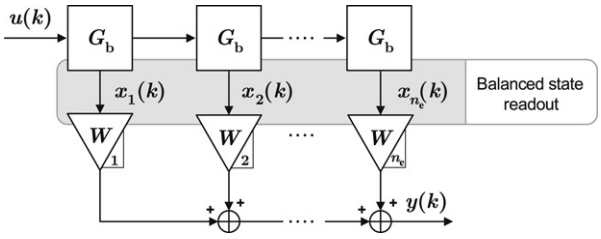


Fig. 1. I/O signal flow graph of the OBF model structure described by (6) for a finite n_{e} number of extensions of G_{b} and with $W_i = \left[w_{i1}, \ldots w_{in_{\mathrm{b}}}\right]$.

Identification of any $F \in \mathcal{H}_{2-}(\mathbb{E})$ based on a predefined set of OBFs $\Phi_{n_{\mathrm{b}}}^{n_{\mathrm{e}}}$ consisting of $n = (n_{\mathrm{e}} + 1)\, n_{\mathrm{b}}$ basis functions, is performed as a linear regression with respect to the basis coefficients $W_{n_{\mathrm{b}}}^{n_{\mathrm{e}}} = \left[w_{ij}\right]_{j=1,\ldots,n_{\mathrm{b}}}^{i=0,\ldots,n_{\mathrm{e}}}$ due to the linear parametrization of (6). The OBF-based identification has valuable properties. Non-asymptotic variance bounds of the estimates are computable through reproducing kernels and the identified models are unbiased if the input signal is uncorrelated to the noise. This is explained by the *Output Error* (OE) like structure of the OBF parametrization (Heuberger et al., 2005). However, selection of the basis function set has a major impact on the outcome of the identification process as the distance between basis poles and the original system poles determines the convergence rate of the coefficients, meaning that with a "better" basis function set a better approximation can be achieved.

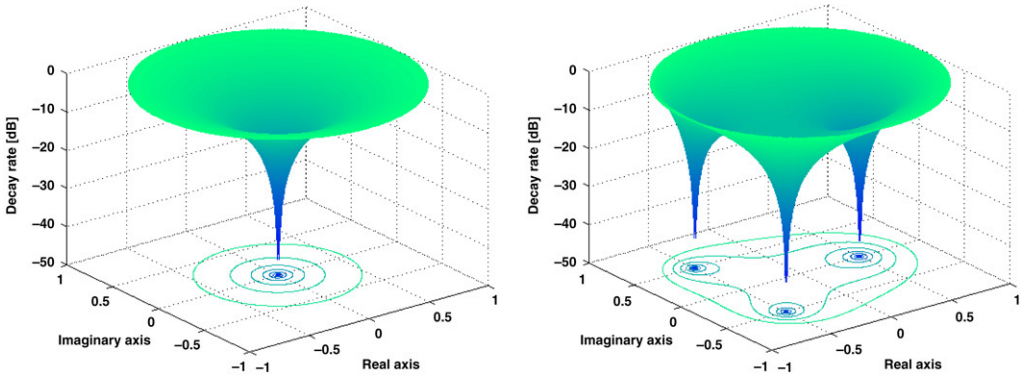
As discussed, OBF-based parametrization can be effectively used for LTI system representation and in this way to describe each $\mathcal{F}_{\bar{\mathrm{p}}} \in \mathfrak{F}_{\mathbb{P}}$ of an LPV system $\mathcal{S}$. However, if the same OBFs are used to compose each $\mathcal{F}_{\bar{\mathrm{p}}}$, then it is required that the basis function set is "well chosen" with respect to the entire $\mathfrak{F}_{\mathbb{P}}$. In the next section, the concept of optimality of an OBF set with respect to $\mathfrak{F}_{\mathbb{P}}$ is established, giving the key theorem to solve the basis function selection problem of the proposed identification scheme.

3. Kolmogorov n-width for OBFs

In the proposed LPV identification approach, it is crucial to find an appropriate model set, i.e. set of basis functions $\Phi_{n_{\mathrm{b}}}^{n_{\mathrm{e}}}$ for the local behaviors $\mathfrak{F}_{\mathbb{P}}$, in the sense that $\Phi_{n_{\mathrm{b}}}^{n_{\mathrm{e}}}$ is sufficiently rich to describe the systems belonging to $\mathfrak{F}_{\mathbb{P}}$, with a relatively small number of statistically meaningful parameters. In LTI system identification, one approach to find appropriate model sets is based on the n-width concept (Pinkus, 1985), which was shown to result in appropriate model sets for robust modeling of linear systems (Mäkilä & Partington, 1993). Using this concept, Oliveira e Silva (1996), (Heuberger et al., 2005, Ch. 11) showed that OBF model structures are optimal in the n-width sense for specific subsets of systems. In the following, the basic ingredients of this theory for discrete-time, stable, SISO systems are described.

Let $\mathfrak{F}$ denote a set of systems with transfer functions $\{F\} = \mathfrak{T} \subseteq \mathcal{H}_{2-}(\mathbb{E})$, that we want to approximate with the linear combination of n elements of $\mathcal{H}_{2-}(\mathbb{E})$. Let $\Phi_n = \{\phi_i\}_{i=1}^{n}$ be a sequence of n linearly independent elements of $\mathcal{H}_{2-}(\mathbb{E})$, and let $\Psi_n = \mathrm{span}(\Phi_n)$. The distance $d_{\mathcal{H}_{2-}}(F, \Psi_n)$ between $F \in \mathcal{H}_{2-}(\mathbb{E})$ and Ψ_n is defined as

$$d_{\mathcal{H}_{2-}}(F, \Psi_n) = \inf_{G \in \Psi_n} \|F - G\|_{\mathcal{H}_2}. \tag{7}$$



(a) $G_b(z) = z^{-1}$, pole at the origin. (b) $G_b(z)$ with poles 0.5 and -0.5 ± 0.5i.

Fig. 2. The plot of the function $\left|G_b\left(z^{-1}\right)\right|$ for different choices of the inner function G_b and the decay rate ρ (in dB). Level sets of $\left|G_b\left(z^{-1}\right)\right|$ give the boundaries of the regions $\{z \in \mathbb{C}, \left|G_b\left(z^{-1}\right)\right| \leq \rho\}$. Optimality of the G_b generated basis is ensured with a worst-case decay rate ρ^{n_e+1} for systems with pole locations inside the regions defined by the level set boundaries.

If $\mathfrak{M}_n$ is the collection of all n-dimensional subspaces of $\mathcal{H}_{2-}(\mathbb{E})$, then the Kolmogorov n-width of $\mathfrak{T}$ in $\mathcal{H}_{2-}(\mathbb{E})$ is

$$\pi_n\left(\mathfrak{T}, \mathcal{H}_{2-}(\mathbb{E})\right) = \inf_{\Psi_n \in \mathfrak{M}_n} \sup_{F \in \mathfrak{T}} d_{\mathcal{H}_{2-}}(F, \Psi_n), \tag{8}$$

which means the smallest possible approximation error for the worst-case F in $\mathfrak{T}$. The subspace $\breve{\Psi}_n \in \mathfrak{M}_n$, for which π_n is minimal, is called the optimal subspace in the KnW sense. Now we can formulate this concept for OBFs:

Proposition 1 (*Oliveira e Silva, 1996*). *Let G_b be an inner function with McMillan degree $n_b > 0$, with poles Ξ_{n_b} and $n_e \in \mathbb{N}$. Consider the subspace $\Psi_n = \operatorname{span}\{\phi_j(z) G_b^i(z)\}_{j=1,\ldots,n_b}^{i=0,\ldots,n_e}$. Then the subspace Ψ_n is optimal in the Kolmogorov n-width sense, with $n = (n_e + 1) n_b$, for the set of systems with transfer functions analytic in the complement of the region*

$$\Omega\left(\Xi_{n_b}, \rho\right) = \{z \in \mathbb{C}, \left|G_b\left(z^{-1}\right)\right| \leq \rho\}, \tag{9}$$

and squared integrable on its boundary. The worst-case approximation error is proportional to ρ^{n_e+1}.

This remarkable result shows that for the specified region (9) one cannot improve on the worst-case error by adding new poles to the n_b basis poles. It also generalizes the well-know fact that the set of *pulse functions* $\{z^{-i}\}_{i=1}^{n}$ is optimal for the class of stable systems analytical outside the circular region $\Omega(0, \rho) = \{|z| \leq \rho\}$, $\rho > 0$. The boundary of $\Omega(0, \rho)$ is given in Fig. 2(a) as a function of the decay rate ρ. For a given $\rho > 0$, the boundary of the region results as the level set of this function, like the contour lines at the bottom of the figure. The worst-case approximation error in this case is proportional to ρ^n. This implies the optimality of FIR model structures with respect to the identification of such systems. However in case of arbitrary regions, like the regions in Fig. 2(b), the level sets are commonly non-circular containing separate regions that merge for increasing values of ρ. For these regions, the optimal choice of a basis has to be found among general basis functions (OBF model structures).

In the LPV identification scenario we are dealing with the opposite problem, referred to as the *inverse Kolmogorov problem*, where we are given a region of non-analyticity $\Omega_{\mathbb{P}} \subset \mathbb{D}$ and we want to find an inner function G_b to describe/approximate this region in the form $\Omega\left(\Xi_{n_b}, \rho\right)$ with ρ as small as possible. The reason is that in terms of Proposition 1, the inner function G_b, associated with the best fitting $\Omega\left(\Xi_{n_b}, \rho\right)$, generates the n-width optimal basis functions with respect to $\Omega_{\mathbb{P}}$. For a given number of poles n_b, this comes down to the following min–max problem:

$$\min_{\xi_1,\ldots,\xi_{n_b}} \max_{z \in \Omega_{\mathbb{P}}} \prod_{j=1}^{n_b} \left|\frac{z - \xi_j}{1 - z\xi_j^*}\right|. \tag{10}$$

See Heuberger et al. (2005, Chapters 10 and 11) for details on this nonlinear optimization problem and solution methods.

The previous results show that the Kolmogorov n-width theory for OBFs provides an effective way to choose appropriate basis functions for the description[1] of $\mathfrak{F}_{\mathbb{P}}$ based on $\Omega_{\mathbb{P}}$. However, in an identification scenario we are facing the situation where $\Omega_{\mathbb{P}}$ is unknown. Thus, to enable the application of this theory, we will focus on the problem of reconstruction of $\Omega_{\mathbb{P}}$ based on some sample pole locations of this set. As we will see, the joint solution of this reconstruction problem and the optimization (10) can be found through a clustering approach.

4. Fuzzy–Kolmogorov *c*-Max clustering

4.1. The pole clustering algorithm

In the following we propose a particular data clustering algorithm, which by weighting-function-based separation, the so-called *fuzzy clustering* of sampled pole locations $\bar{\Omega}$ of the LPV system, can effectively handle the reconstruction of $\Omega_{\mathbb{P}}$ jointly with the solution of (10).

Objective-function-based fuzzy clustering algorithms, such as the *Fuzzy c-Means* (FcM), have been used in a wide collection of applications like pattern recognition, data analysis, image processing and fuzzy modeling (Bezdek, 1981; Kaymak & Setnes, 2002). Generally, FcM partitions the data into overlapping groups so called clusters, where each data element is associated with a set of membership levels with respect to these clusters. These indicate the strength of the association between that data element and a particular cluster. In this way, fuzzy clustering is a process of assigning these membership levels such that the resulting clusters describe the underlying structure within the data (Jain & Dubes, 1988). This enables the determination of the region $\Omega_{\mathbb{P}}$ on the basis of the observed poles by exploring the underlying data coherency. To exploit this fruitful property, in the following such a *Fuzzy–Kolmogorov c-Max*(FKcM) algorithm is presented, which provides an effective OBF selection approach based on the fusion of the KnW theory and the FcM technique.

Let $c > 1$ be the number of clusters or data groups and let $Z = [z_k]_{k=1}^{N} \in \mathbb{D}^N$, be the set of $N \in \mathbb{N}$ observed poles for clustering. A cluster is represented by its center (or *prototype*) $v_i \in \mathbb{D}$, $i \in \mathbb{I}_1^c$.

[1] Note that $\mathfrak{F}_{\mathbb{P}}$ is a set of LTI systems which can be represented in the frequency domain. However, it must be clear that the global LPV system $\mathcal{S}$ is not described in frequency domain terms.

Furthermore, membership functions $\mu_i : \mathbb{D} \to [0, 1]$ determine the "degree of membership" to the cluster for all $z \in \mathbb{D}$. By using a *threshold value* ε, we obtain a set

$$\Omega_\varepsilon = \{z \in \mathbb{D} \mid \exists i \in \mathbb{I}_1^c, \mu_i(z) \geq \varepsilon\}. \tag{11}$$

We can now formulate the problem we will consider.

Problem 2. For a set of sampled pole locations Z and for a given number of clusters c, find a set of cluster centers $\{v_i\}_{i=1}^c$, a set of membership functions $\{\mu_i\}_{i=1}^c$, and the maximum of ε, such that

- Ω_ε contains Z.
- With respect to Ω_ε, the OBFs, with poles Ξ_c in the cluster centers $\{v_i\}_{i=1}^c$, are optimal in the KnW sense, where $n = c$.

The solution is based on finding clusters in accordance with the KnW concept and subsequently finding a maximal value for ε, such that all sampled poles are inside Ω_ε. The latter is equivalent to minimizing ρ in the optimization problem of (10). Note that optimality of the OBFs is sought as $n_\mathrm{e} = 0$. According to the principle of KnW theory, this might result in repetitive optimal poles and therefore similar clusters. In the following we will focus on finding n-width-based clusters.

Denote $V = [v_i]_{i=1}^c$ and introduce the membership matrix $U = [\mu_{ik}]_{c\times N}$, where μ_{ik} is the degree of membership of z_k to cluster i. To constrain the clustering it is required that $U \in \mathcal{U}_c^N$, where

$$\mathcal{U}_c^N = \left\{ U \in [0, 1]^{c\times N} \,\middle|\, \sum_{i=1}^{c} \mu_{ik} = 1 \text{ for } \forall k \in \mathbb{I}_1^N, \quad 0 < \sum_{k=1}^{N} \mu_{ik} \text{ for } \forall i \in \mathbb{I}_1^c \right\} \tag{12}$$

characterizes the *fuzzy constraints*.

Furthermore, distances d_{ik} are introduced between v_i and z_k to measure dissimilarity of Z with respect to each candidate cluster. To derive an algorithmic solution of Problem 2, the Kolmogorov metric[2] (KM) of $\mathbb{D}$:

$$\kappa(x, y) := \left| \frac{x - y}{1 - xy^*} \right| : \mathbb{D} \times \mathbb{D} \to \mathbb{R}_0^+, \tag{13}$$

with $\mathbb{R}_0^+ = \{r \in \mathbb{R} \mid r \geq 0\}$ is used, which is the 1-width version of the cost function of (10). As a notation, $d_{ik} = \kappa(v_i, z_k)$ is introduced. It will be shown that KM relates the FcM asymptotically to the KnW theory and to the solution of Problem 2.

Fuzzy clustering can be defined as the minimization of the FcM-*functional* (Bezdek, 1981), $J_m(U, V) : \mathcal{U}_c^N \times \mathbb{D}^c \to \mathbb{R}_0^+$. For Problem 2, J_m is formulated as

$$J_m(U, V) = \max_{k\in\mathbb{I}_1^N} \sum_{i=1}^{c} \mu_{ik}^m d_{ik}. \tag{14}$$

Here, the design parameter $m \in (1, \infty)$ determines the fuzziness of the resulting partition. It can be observed, that (14) corresponds to a *worst-case (max) sum-of-error* criterion, contrary to the *mean-squared-error* criterion of the original FcM, see Bezdek (1981). The exact relation of (14) with the KnW optimality of the partition (U, V) is explained later. The following theorem yields the ingredients for the approach to solve Problem 2:

Theorem 3 (*Optimal Partition*). *Let $m > 1$, a data set $Z \in \mathbb{D}^N$, and a fuzzy partition $(U, V) \in \mathcal{U}_c^N \times \mathbb{D}^c$ be given. Denote $[V]_i = v_i$ and $[U]_{ij} = \mu_{ij}$. Define $\gamma_i(v, U)$ as the minimal value of $\gamma \in [0, 1]$ fulfilling the quadratic constraints:*

$$\begin{bmatrix} |1 - z_k^* v|^2 & \mu_{ik}^m (z_k - v) \\ \mu_{ik}^m (z_k - v)^* & \gamma^2 \end{bmatrix} \succeq 0, \quad \forall k \in \mathbb{I}_1^N, \tag{15}$$

where $v \in \mathbb{D}$. Additionally, let $d_{ik} = \kappa(v_i, z_k)$ be the dissimilarity measure of z_k with respect to V and $\mathcal{I}_\mathrm{s}^{(k)} = \left\{ i \in \mathbb{I}_1^c \mid d_{ik} = 0 \right\}$ be the singularity set of z_k with $n_\mathrm{s}^{(k)} = \mathrm{card}(\mathcal{I}_\mathrm{s}^{(k)})$ (number of elements). Then (U, V) is a local minimum of J_m, if for any $(i, k) \in \mathbb{I}_1^c \times \mathbb{I}_1^N$:

$$\mu_{ik} = \begin{cases} \left[\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{1}{m-1}} \right]^{-1} & \text{if } \mathcal{I}_\mathrm{s}^{(k)} = \emptyset, \\ \frac{1}{n_\mathrm{s}^{(k)}} & \text{if } i \in \mathcal{I}_\mathrm{s}^{(k)}, \\ 0 & \text{if } i \notin \mathcal{I}_\mathrm{s}^{(k)} \neq \emptyset, \end{cases} \tag{16}$$

and $v_i = \arg\min_{v\in\mathbb{D}} \quad \gamma_i(v, U).$ (17)

The proof is given in Appendix A. In the FcM case, minimization of (14) subject to (12) is usually tackled by alternating optimization (Picard iteration) (Bezdek, 1981), steering the solution towards a settling partition in the sense of Theorem 3. For the FKcM this yields Algorithm 1, formulated next, where V_l and U_l denote the actual fuzzy partition in iteration step l. In the following part, we will discuss the main properties of this algorithm and clarify each step in detail.

Algorithm 1. Fuzzy–Kolmogorov c-Max

(1) *Initialization*
Fix c and m; and initialize $V_0 \in \mathbb{D}^c$, $l = 0$.
(2) *Membership update*
With (16), solve $U_{l+1} = \arg\min_{U\in\mathcal{U}_c^N} J_m(U, V_l)$.
(3) *Cluster center update*
With (17), solve $V_{l+1} = \arg\min_{V\in\mathbb{D}^c} J_m(U_{l+1}, V)$.
(4) *Check of convergence*
If $J_m(U_{l+1}, V_{l+1})$ has converged, then stop, else $l = l + 1$ and go to Step 2.

4.2. Properties of the FKcM

In order to explain the specific choices for the fuzzy functional (14) and the dissimilarity measure (13), we use the following theorem.

Theorem 4 (*Limiting Property of J_m*). *Given a data set $Z \in \mathbb{D}^N$, $N > 0$, and a set of cluster centers $V \in \mathbb{D}^c$, $c > 0$, such that $d_{ik} = \kappa(v_i, z_k) \neq 0$ for all $(i, k) \in \mathbb{I}_1^N \times \mathbb{I}_1^c$ (no singularity). Define U_m as a membership matrix of V satisfying (16) for $m > 1$. Then*

a. $\lim_{m\to 1} J_m(U_m, V) = \max_{k\in\mathbb{I}_1^N} \min_{i\in\mathbb{I}_1^c} \{d_{ik}\}$, *which corresponds to the hard partitioning of Z, i.e. $\mu_{ik} \in \{0, 1\}$, $\forall (i, k) \in \mathbb{I}_1^c \times \mathbb{I}_1^N$. Here, the optimal partition corresponds to a collection of 1-width optimal basis functions with respect to each reconstructed pole region.*
b. $J_2(U_2, V) = \max_{k\in\mathbb{I}_1^N} \left[\sum_{i=1}^{c} d_{ik} \right]^{-1}$, *which is the maximum of the harmonic-means-based distance of each z_k with respect to the clusters.*
c. $J_m(U_m, V) = c^{1-m} \max_{k\in\mathbb{I}_1^N} \left[\prod_{i=1}^{c} d_{ik} \right]^{1/c} + \mathcal{O}(\mathrm{e}^{-m})$. *Furthermore, $J_m(U_m, V)$ decreases monotonically with m, and $J_\infty(U_\infty, V) = 0$.*

[2] Note that KM is not a distance in $\mathbb{D}$, only $\mathrm{arctanh}(\kappa(x, y))$, called the *Poincaré distance*, bears this property (Brannan, Esplen, & Gray, 1999). However in fuzzy clustering, the dissimilarity measure does not need to qualify as a distance.

The proof is presented in Appendix A. Based on Theorem 4, the minimization of J_m corresponds to a close approximation of (10) for large m, enabling the FK*c*M to solve Problem 2 directly. However, if $m \to \infty$, then in the optimal partition $\mu_{ik} \to 1/c$ for all $(i,k) \in \mathbb{I}_1^c \times \mathbb{I}_1^N$, which can cause numerical problems in the minimization of (17). Therefore, to obtain a well-approximating solution of Problem 2, an appropriately large value of $m \in (1,\infty)$ should be used. Based on experience, $m \in [5, 10]$ usually yields satisfactory results.

For $m > 1$, the FK*c*M-functional (14) is a bounded ($0 \le J_m \le 1$) monotonically descending function both in $\{d_{ik}\}$ and U, which allows Algorithm 1 to converge in practice. The convergence point, which is directly dependent on the initial V^0, can either be a local minimum or a saddle point of J_m, fulfilling Theorem 3. Therefore, it is advisable to repeat the algorithm multiple times with different initial choices for V^0 and then select the best resulting set of OBFs by comparison of the achieved decay rate

$$\bar{\rho} = \max_{z\in Z} \prod_{i=1}^{c} \left| \frac{z - v_i}{1 - z v_i^*} \right|, \tag{18}$$

and by visual inspection of the region $\Omega(\Xi_c = V, \bar{\rho})$ with respect to Z. In practice, uniformly random choices for V^0 are suggested.

4.3. Optimization and numerical conditioning

While the membership update step in Algorithm 1 can be analytically computed through (16), the cluster center update step requires the solution of (17) which is a minimization problem with a *Quadratic Constraint* (QC) where γ is the optimization variable and v is the decision variable. Based on Scherer and Hol (2006), it is possible to derive *Sum-of-Squares* (SoS) relaxations of such constraints, through which (15) is turned into LMIs. The resulting LMIs constrained convex minimization of γ is a *Linear Semi Definite Programming* (LSDP) problem that can be efficiently solved by a variety of (interior-point-based) solvers like SeDuMi (Sturm, 1999) or CSDP etc. Alternatively, bisection-based recursive search can also be utilized to obtain the minimization of γ in (17). In each step of this bisection-based minimization, the QCs with a fixed γ are rewritten as LMI constraints. Checking feasibility of the constraints indicates how to proceed with the minimization of γ.

For high values of m, the QCs (15) become numerically ill-conditioned which can be overcome by the normalization of $\{\mu_{ik}^m\}_{k=1}^N$:

$$\bar{\mu}_{ik} = \frac{\mu_{ik}^m}{\breve{\mu}_i}, \quad \text{with } \breve{\mu}_i = \sum_{k=1}^{N} \mu_{ik}^m. \tag{19}$$

4.4. Termination criterion

In Algorithm 1, the cost function J_m flattens when m increases. This yields that for high values of m, J_m drastically drops in a local minimum, while J_m is almost constant for other points. To avoid unnecessary termination, the relative evolution of J_m, in each iteration step l, has to be checked in a windowed sense:

$$1 - \frac{\max_k [J_m(U_k, V_k) - J_m(U_{k-1}, V_{k-1})]}{\max_k J_m(U_k, V_k)} < \varepsilon_t \tag{20}$$

where $k \in \mathbb{I}_{l-n_w}^{l}$, $n_w \in \mathbb{N}$ is the length of the window, and $0 \ll \varepsilon_t < 1$ is a user defined termination constant. For $m \in [5, 10]$, $\varepsilon_t = 0.99$ with $n_w = 3$ usually works well.

4.5. Cluster merging

The determination of the number of "natural" groups in Z, i.e. the best suitable c for clustering, is important for the successful application of the FK*c*M method. Similarity-based *adaptive cluster merging* (ACM) is frequently used for this purpose (Kaymak & Setnes, 2002), but other strategies exist also. ACM is suitable for problems where little is known about the statistical properties of the data, like in the pole clustering case. The basic idea is the following: a measure of similarity is introduced with respect to cluster pairs. A cluster pair is merged when its similarity does not decrease between iterations and if also this pair is the most similar of all cluster pairs. However, merging is only applied if the similarity measure exceeds a certain threshold value, $\varepsilon_a \in [0, 1]$ arbitrarily chosen by the user. In F*c*M clustering, most commonly the following similarity measure is applied:

Definition 5 (*Inclusion Similarity Measure (Kaymak & Setnes, 2002)*). The fuzzy-inclusion-similarity measure (given point-wise on Z) for two fuzzy clusters i and j is defined as

$$s_{ij} = \frac{\sum_{k=1}^{N} \min\left(\mu_{ik}, \mu_{jk}\right)}{\min\left(\sum_{k=1}^{N} \mu_{ik}, \sum_{k=1}^{N} \mu_{jk}\right)}. \tag{21}$$

This measure takes into account the contribution to similarity from all $\{z_k\}_{k=1}^N$. For the theoretical details see Kaymak and Setnes (2002). Then, in the lth iteration of Algorithm 1, the most similar cluster pair can be selected as

$$(\hat{\imath}, \hat{\jmath}) = \underset{(i,j)\in\mathbb{I}_1^c\times\mathbb{I}_1^c, i>j}{\arg\max} \{s_{ij}^{(l)}\}. \tag{22}$$

Merging is applied if $|s_{\hat{\imath}\hat{\jmath}}^{(l-1)} - s_{\hat{\imath}\hat{\jmath}}^{(l)}| < \varepsilon_s$, where $0 < \varepsilon_s \ll 1$ is a threshold value to judge the significance of decrease of cluster similarity between iterations. However as the partition converges, similarity changes a little between iterations, therefore merging is only applied if $s_{\hat{\imath}\hat{\jmath}}^{(l)} > \varepsilon_a^{(l)}$ where $\varepsilon_a^{(l)} \in [0, 1]$ is an adaptive threshold. In Kaymak and Setnes (2002), it is suggested to use $\varepsilon_a^{(l)} = (c^{(l)} - 1)^{-1}$ which is observed empirically to work well if the initial number of clusters c^0 satisfy $c^0 < \frac{1}{2}N$.

In this way, the FK*c*M algorithm with ACM provides the possibility to automatically choose the number of required OBFs for the model structure based on Z. So by starting from a large c, the algorithm converges to a partition which contains only the necessary number of clusters representing the data. However in terms of Proposition 1, the setting of Problem 2 implies that repetitive basis poles can be part of the optimal solution. With the ACM, these solutions are not accessible, since repetitive poles result in perfectly similar clusters which are immediately joined. Therefore, ACM only provides convergence to partitions with distinct cluster centers.

5. LPV system identification

In the previous section, an OBF selection algorithm has been proposed to obtain an adequate selection of the model structure (2) with respect to an unknown LPV system $\mathcal{S}$. The fact that an LPV system can be viewed as a set of local LTI behaviors $\mathfrak{F}_{\mathbb{P}}$ which are combined by a set of scheduling functions $\mathfrak{H}_{\mathbb{P}}$ is the motivation to select the optimal model structure based on $\mathfrak{F}_{\mathbb{P}}$. In the following, according to the LPV system identification approach of Section 1, it is briefly shown how these OBFs can be used for identifying a discrete-time $\mathcal{S}$ efficiently, i.e. how the scheduling functions can be

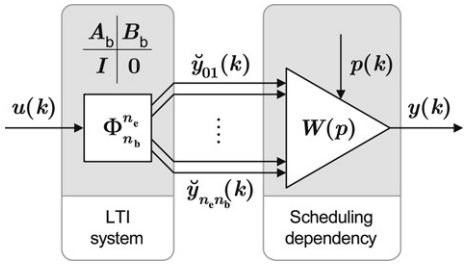


Fig. 3. I/O signal flow graph of the W-LPV OBF model structure.

estimated. Based on (2), we introduce a model structure presented in Fig. 3, where the selected OBFs are set up as a filter bank followed by a p-dependent weighting function (Tóth, Heuberger et al., 2007). Due to the similarity of this model structure to Wiener models we will call it a Wiener-LPV (W-LPV) OBF model.

Let $\Phi^{n_e}_{n_b}$ be a set of OBFs in $\mathcal{H}_{2-}(\mathbb{E})$. Denote by $\{A_b, B_b, C_b, D_b\}$ a minimal balanced SS realization of $G^{n_e}_b$, where G_b is the inner function associated with $\Phi^0_{n_b}$. Let $\mathcal{S}$ be a data generating SISO LPV system without a feedthrough term. By applying the W-LPV model of $\mathcal{S}$ in an OE setting and with static coefficient dependence, leads to the following 1-step ahead predicted output

$$\hat{y}(k) = \sum_{i=0}^{n_e} \sum_{j=1}^{n_b} w_{ij}(p(k)) \underbrace{\phi_j(q) G^i_b(q) u(k)}_{\breve{y}_{ij}(k)}, \tag{23}$$

where q denotes the forward time-shift operator and $\{w_{ij}\}$ is a set of functions. The SS equivalent representation of (23), is defined as

$$x(k+1) = A_b x(k) + B_b u(k), \tag{24a}$$

$$\hat{y}(k) = W(p(k)) x(k), \tag{24b}$$

where $x^T = [\breve{y}_{01} \ \dots \ \breve{y}_{n_e n_b}]$ and $W(p) = [w_{01}(p) \ \dots \ w_{n_e n_b}(p)]$. Note that using the model structure (23) in an OE setting has many attractive properties. For instance, it is linear in the coefficient functions $\{w_{ij}\}$, the noise model is independently parametrized from the process part, and this model structure has a direct state-space realization via (24a) and (24b) where only the output equation has dependence on p. The latter implies that LPV control design simplifies for the obtained model estimate. Furthermore, it can be shown based on series expansion representation of LPV systems, that the W-LPV OBF structure can represent the general class of LPV systems with arbitrary precision in case the weighting functions W have dynamic dependence (dependence on the past/future of p) (Tóth, Heuberger et al., 2007). In case of static dependence, i.e. dependence on $p(k)$ (the instantaneous value of p), approximation of a wide class of LPV systems is available by this model structure. This class however is hard to characterize as the approximation error results both from the finite number/quality of the chosen OBFs and the assumption of static dependence. Identification of an LPV system $\mathcal{S}$ by this model class can be accomplished through the local and global approach presented in Section 1. The schematic view of these approaches is given in Fig. 4.

5.1. Local approach

In this method, the identification is based on data records $\mathcal{D}_l = \{y(k), u(k), \bar{p}_l\}_{k=1}^{N_d}$ gathered from $\mathcal{S}$ for constant scheduling points $\mathcal{P}_n = \{\bar{p}_1, \dots, \bar{p}_n\}$ with persistently exciting (in the LTI sense) u. Then local LTI models of $\mathcal{S}$

$$\hat{y}(k) = \sum_{i=0}^{n_e} \sum_{j=1}^{n_b} r_{\bar{p}_l,i,j} \phi_j(q) G^i_b(q) u(k), \tag{25}$$

in terms of the OBFs $\Phi^{n_e}_{n_b}$ with $r_{\bar{p}_l,i,j} \in \mathbb{R}$ are estimated, e.g. by linear regression, based on $\mathcal{D}_l$. The obtained coefficients $\{\hat{r}_{\bar{p}_l,i,j}\}$ are the estimated samples of $\{w_{ij}\}$ in (23) for the applied constant scheduling functions, i.e. $w_{ij}(\bar{p}_l) \approx \hat{r}_{\bar{p}_l,i,j}$. Then by choosing a particular structure of the static functional dependence of each w_{ij}, interpolation (by arbitrary method) of $\{\hat{r}_{\bar{p}_l,i,j}\}$ gives an estimate $\{\hat{w}_{ij}\}$ of $\{w_{ij}\}$ for which $\hat{w}_{ij}(\bar{p}_l) = \hat{r}_{\bar{p}_l,i,j}$. In this way, the local approach simply provides an extension of the classical LTI OBF identification approaches with all their beneficial properties to the LPV case.

Note that an adequate choice of $\mathcal{P}_n$ is required for an efficient model estimate and adequateness of $\mathcal{P}_n$ depends on the variation

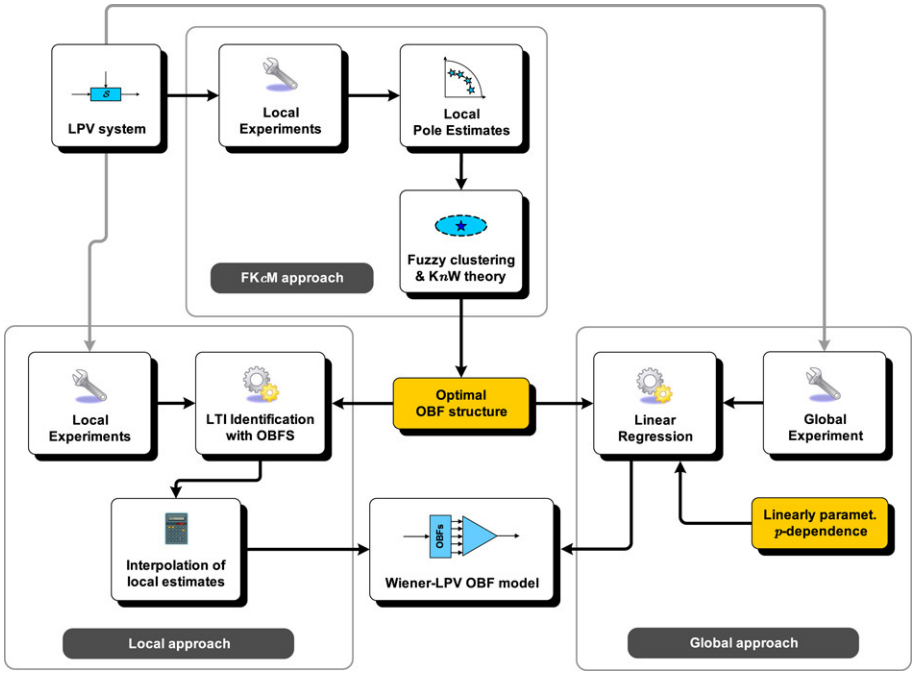


Fig. 4. Block diagram of local and global identification methods.

of the dynamical properties in $\mathcal{F}_{\mathbb{P}}$ (see Murray-Smith, Johansen, and Shorten (1999) for more on this issue). Furthermore, in case of systems where p cannot be held constant, the local approach is hardly applicable. This approach also estimates the global behavior of $\mathcal{S}$ based only on samples of $\mathcal{F}_{\mathbb{P}}$ which cannot describe the transient behavior imposed in the scheduling functions. These disadvantages motivate the following alternative:

5.2. Global approach

In the global case, identification is based on a single data record $\mathcal{D} = \{y(k), u(k), p(k)\}_{k=1}^{N_d}$ of $\mathcal{S}$ with a varying p. If each w_{ij} in (23) is parametrized linearly:

$$w_{ij}(p(k)) = \sum_{l=0}^{n_\varphi} r_{lij}\varphi_l(p(k)), \tag{26}$$

where φ_l are arbitrary chosen functions, e.g. polynomials, and $r_{lij} \in \mathbb{R}$, then (23) becomes linear in the unknown parameters $\{r_{lij}\}$. This implies that estimation of $\{w_{ij}\}$ based on (26) and $\mathcal{D}$ can be accomplished through linear regression. Consistency of the parameter estimation can be shown under minor conditions (Tóth, Heuberger et al., 2007) together with the extension of some classical results on the variance, bias, etc. of OBF model estimates.

6. Results of application

As an example, an asymptotically stable discrete-time LPV system $\mathcal{S}$ is considered, in an I/O representation:

$$\sum_{i=0}^{5} a_i\left(p\left(k\right)\right) y\left(k-i\right) = b_1\left(p\left(k\right)\right) u\left(k-1\right), \tag{27}$$

where $p : \mathbb{Z} \to \mathbb{P}$ is the discrete-time scheduling signal with $\mathbb{P} = [0.6, 0.8]$ and $a_0(p) = 0.58 - 0.1p$, $a_1(p) = -\frac{511}{860} - \frac{48}{215}p^2 + 0.3(\cos(p) - \sin(p))$, $a_2(p) = \frac{61}{110} - 0.2\sin(p)$, $a_3(p) = -\frac{23}{85} + 0.2\sin(p)$, $a_4(p) = \frac{12}{125} - 0.1\sin(p)$, $a_5(p) = -0.003$, $b_1(p) = \cos(p)$. In Fig. 5(a), the local pole set $\Omega_{\mathbb{P}}$ of $\mathcal{S}$ is presented, while in Fig. 5(b) the impulse responses of the local subsystem set $\mathfrak{F}_{\mathbb{P}}$ is given. By these pictures, it can be concluded that the dynamic changes of $\mathcal{S}$ are quite heavy between different constant scheduling points.

6.1. OBF selection by FKcM clustering

By using constant scheduling signals with values $\{0.6; 0.6 + \tau; \ldots; 0.8\}$, where $\tau = 0.02$, 11 local LTI representations of $\mathcal{S}$ are obtained, whose pole locations are samples of $\Omega_{\mathbb{P}}$ (see Fig. 5(a)). In our basis function selection approach, these LTI systems represent the results of local identification. With the obtained $N = 11 \cdot 5$ pole locations, the FK*c*M algorithm has been applied with different values of m and both with fixed number of clusters $c = 8$ (denoted by *m2c8* for $m = 2$) and also with the application of ACM starting from $c^{(0)} = N/2$ (denoted by *m8ad11* for $m = 8$ resulting in $c = 11$ clusters). The number of clusters 8 agrees with the number of sets by visual inspection (two times 3 sets for the complex and 2 sets for the real poles) and as will follow, also with the number of clusters selected by ACM. The results of the algorithm are presented in Table 1 and in Fig. 6. The comparison in Table 1 is presented in terms of N_{av}, the average number of iterations based on 10 runs of the algorithm starting from random V_0; c, the number of obtained clusters; H_p, the Normalized Entropy[3]; χ, the Xie–Beni validity index[4]; $\bar{\rho}$, the achieved decay rate; and $\varepsilon_{\max}^{n_e}$, the worst-case absolute representation error of the local impulse responses with n_e extension of the cluster centers generated OBFs. By using the cluster centers as basis poles, $\Xi_{n_b=c} = V$, the resulting Kolmogorov region $\Omega(\Xi_c, \bar{\rho})$ is also given in Fig. 6. Based on these, the following observations can be made:

Table 1
Comparison of algorithmic results.

	m2c8	*m8ad8*	*m8ad11*	*m25c8*
N_{av}	21	37	65	56
c	8	8	11	8
χ (dB)	−17.49	−12.42	−8.44	−13.20
$\bar{\rho}$ (dB)	−55.86	−58.38	−83.11	−61.36
H_p	1.79	2.41	2.94	2.43
$\varepsilon_{\max}^{n_e=1}$ (dB)	−43.73	−46.90	−77.33	−45.34
$\varepsilon_{\max}^{n_e=3}$ (dB)	−146.61	−171.41	−249.63	−168.83

- The FK*c*M with ACM ($\varepsilon_s = -15$ dB) converges to a 8-cluster-based partition for low m, but in case of higher values of m, the merging, starting from $c^0 = \frac{1}{2}N$, will have different attractive solutions, like the *m8ad8* and *m8ad11* cases. Here both the 8 and the 11 cluster-based partitions are attractive, depending on the initial position of the cluster centers. However, *m8ad8* achieves a lower entropy H_p than *m8ad11*, suggesting that *m8ad8* corresponds better to the natural data structure. As different initial conditions can drive the FK*c*M with ACM to converge to partitions with different c, it is suggested to the user to choose the one with the lowest H_p, as it most likely yields the "best" partition.
- χ is small in all cases, showing that each partition represents the underlying structure well. However, χ is not comparable for different m. χ has a decreasing tendency with growing c and an increasing tendency for growing m, therefore the fact that $\chi_{m25c8} < \chi_{m8ad8}$ supports that *m25c8* corresponds better to the underlying data structure in the K*n*W sense than *m8ad8*.
- The resulting Kolmogorov region $\Omega(\Xi_c, \bar{\rho})$ is relatively tight in all cases except for *m2c8*. $\bar{\rho}$ is also acceptable, which means small modeling error if the corresponding OBFs are used for identification. In the *m8ad11*-case, $\bar{\rho}$ is the best, which is the consequence of the larger ($c = 11$) number of OBFs only. By using extension of the derived poles associated inner functions such that the number of generated basis function is equal, comparison of the K*n*W performance of these cases becomes available. Based on such a comparison, it follows that *m25c8* is better in the K*n*W sense, which is in agreement with Theorem 4. The partition *m2c8* is the worst among these results which suggests that only larger values of m can ensure the quality of the obtained solution.
- Fig. 7 and Table 1 show the representation errors of the local impulse responses of $\mathfrak{F}_{\mathbb{P}}$ by the selected OBFs with poles in the obtained cluster centers. From these results it follows that the obtained OBFs result in negligible representation error with respect to $\mathfrak{F}_{\mathbb{P}}$, which is our main objective to achieve with the presented basis function selection approach (see Section 1). Among the solutions with 8 basis functions, surprisingly *m 8c8* has the lowest representation error instead of *m25c8*. Based on the previous results, one would expect, that the representation error drops for OBFs generated with higher m, however this is not the case here, due to the fact that $\Omega_{\mathbb{P}}$ is sampled. Even if *m25c8* delivers a better choice with respect to the sampled pole locations, it is not guaranteed that the reconstruction of $\Omega_{\mathbb{P}}$,

[3] *Normalized Entropy* (Bezdek, 1981) describes the separation of clusters. The smaller the H_p, the more valid the hypothesis that the clusters match with naturally separated data groups.

[4] The Xie–Beni validity index χ (Xie & Beni, 1991) gives a common ground of comparison between different FcM partitions. The smaller the χ, the better the corresponding fit to the data.

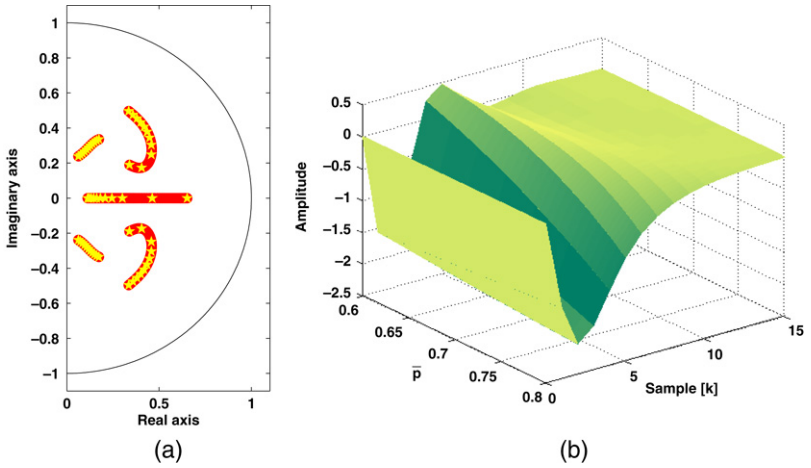
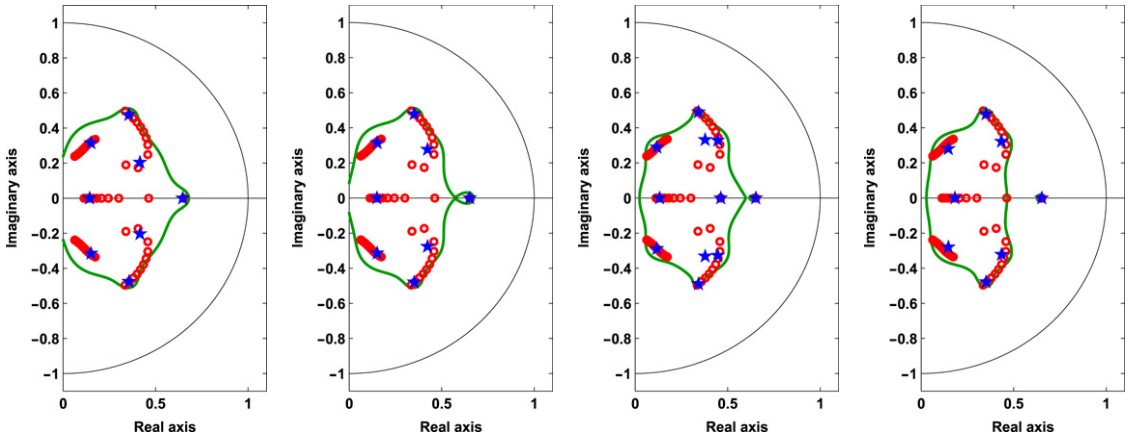


Fig. 5. (a) The local pole set $\Omega_{\mathbb{P}}$ (solid line) of the LPV system $\mathcal{S}$. Sampled pole locations are denoted by $\star$. (b) Impulse responses of the local subsystem set $\mathfrak{F}_{\mathbb{P}}$ of $\mathcal{S}$ associated with constant scheduling signals $p(k) \equiv \bar{p} \in \mathbb{P}$.



(a) *m2c8*. (b) *m8ad8*. (c) *m8ad11*. (d) *m25c8*.

Fig. 6. Results of FKcM clustering: sampled poles (o), resulting cluster centers ($\star$), and Kolmogorov boundaries (bold lines).

based on the available information, resulted in a better estimate than in the other case. By comparing the results of H_p of these cases, such a phenomena is clearly indicated. The quality of the information with respect to the pole samples is highly significant in establishing optimality between the sampled-poles-based OBFs and the original system.

In conclusion, the FK*c*M solutions for the considered example are converging relatively fast to optimal partitions in terms of Theorem 3. In accordance with Theorem 4, as m increases, these partitions give better solutions of Problem 2. ACM also ensures proper selection of an efficient number of OBFs in the K*n*W sense, if the different settling partitions are compared in terms of H_p. Furthermore, validity of the derived partitions is supported by low χ in all cases.

Comparison of results to solutions provided by the gradient search method (Heuberger et al., 2005, Ch. 11), is only possible if the number of available samples of $\Omega_{\mathbb{P}}$ is so high that there is no need for the reconstruction of $\Omega_{\mathbb{P}}$. Thus an advantage of the FK*c*M approach is that it provides a solution for the practical case when only few samples of $\Omega_{\mathbb{P}}$ are available. In the unrealistic case, when $\Omega_{\mathbb{P}}$ is known, the algorithms converge to similar solutions, but with a lower computational time in the FK*c*M case. The two algorithms also have similar properties in the sense that they only provide convergence to local minima. As online selection of the efficient number of OBFs is very difficult to implement into the gradient search method, the FK*c*M approach, with strategies like the ACM, has a second advantage over the gradient method.

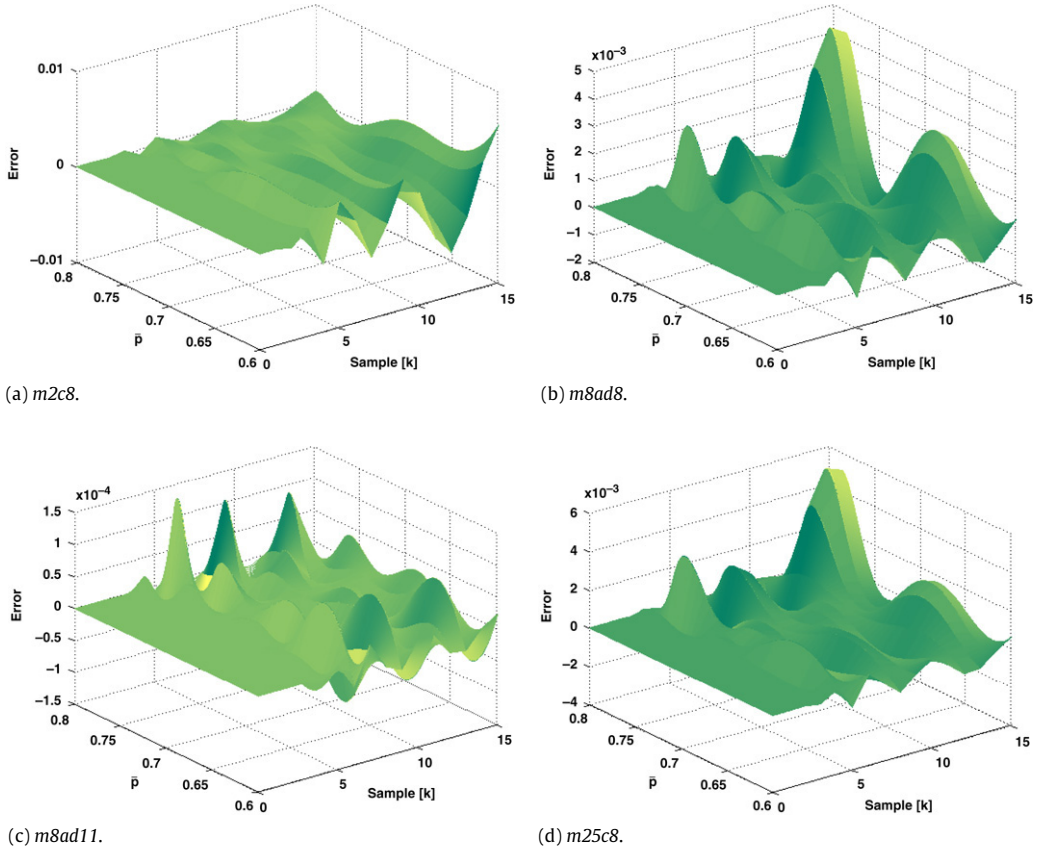
6.2. Identification by the W-LPV OBF model structure

Using the basis functions of the *m8c8* case, identification of $\mathcal{S}$ with the W-LPV OBF structure has been accomplished by the global approach with a 500 sample long data record $\mathcal{D}$. $\mathcal{D}$ was generated by uniform noise $u \in \mathcal{U}(-1, 1)$, $p \in \mathcal{U}(0.6, 0.8)$ and with additive, white output noise: $e \in \mathcal{N}(0, 0.5)$. For the estimation of $W(p)$, 2nd-order polynomial parametrization has been used. In Fig. 8, the (in)validation result of the model estimate is shown with an MSE[5] of 0.0572, BFT[6] of 83.69%, and VAF[7] equals to 97.34%. For the (in)validation of the obtained model, signals $u \in \mathcal{U}(-1, 1)$ and $p \in \mathcal{U}(0.6, 0.8)$ have been used that are different from the signals applied for model estimation. Due to the absence of dynamic dependence in the parametrization of $W(p)$, the W-LPV OBF structure could not cope fully with the variations in the $\{a_l\}_{l=0}^{5}$ parameters, but for such a heavy nonlinear system, the method provided quite acceptable result in terms of the investigated error measures.

[5] *Mean Squared Error*, the expected value of the squared estimation error (Ljung, 1999), often computed in a sampled form: $\widehat{\mathrm{MSE}} = \frac{1}{N}\sum_{k=0}^{N-1}\left(y(k) - \hat{y}(k)\right)^2$.

[6] *Best Fit* percentage, the percentage of the output variation that is explained by the model (Ljung, 2006). $\mathrm{BFT} = 100\% \cdot \max\left(1 - \frac{\|y-\hat{y}\|_2}{\|y-\bar{y}\|_2}, 0\right)$ where $\bar{y}$ is the mean of y.

[7] *Variance Accounted For* percentage is defined as $\mathrm{VAF} = 100\% \cdot \max\left(1 - \frac{\mathrm{var}(y-\hat{y})}{\mathrm{var}(y)}, 0\right)$ and computed on noise free y.



(a) *m2c8*. (b) *m8ad8*.

(c) *m8ad11*. (d) *m25c8*.

Fig. 7. Representation error of the local impulse responses with the FK*c*M clustering obtained OBFs.

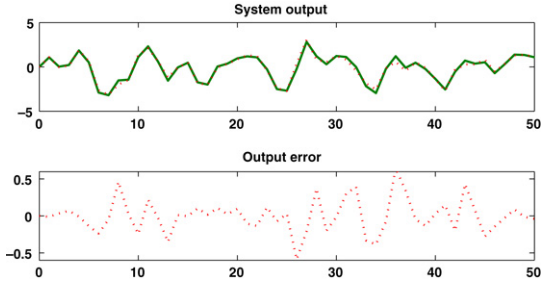


Fig. 8. Comparison of the identified (dotted line) and the true model (solid line) of $\mathcal{S}$ by their responses for $u \in \mathcal{U}(-1, 1)$ and $p \in \mathcal{U}(0.6, 0.8)$.

7. Conclusions

In this paper, an OBF-based model structure selection algorithm has been proposed for the identification of LPV systems. The fact that any LPV system can be viewed as a set of local LTI behaviors combined by a scheduling function set, motivated to formulate the OBF selection algorithm based on the local behavior set. For this set of LTI systems, an optimality condition for OBFs is expressed in the K*n*W sense, which requires the knowledge of the complex regions where the local pole locations of the system lie. In an identification scenario, such knowledge is often only available in a sampled sense, e.g. in the form of poles of some local LTI systems. To overcome this problem, a pole clustering algorithm, the FK*c*M method, is introduced which offers an attractive procedure to determine the pole regions of an unknown system and the associated asymptotically optimal OBFs in the K*n*W sense, based on the available information. This contribution enables the direct use of the K*n*W result for the OBF-based model structure selection in an LPV identification scenario. As a next step of the research, we will focus on the robust extension of the algorithm in order to attenuate the effects of uncertainty in the pole samples.

Acknowledgments

The first and second author were supported by the Netherlands Organization for Scientific Research (NWO). The authors thank Carsten Scherer and Sjoerd G. Dietz for their contribution to constraint relaxations and for many fruitful discussions. Also special thanks to Niels Vergeer for his help in the early development of the FK*c*M mechanism.

Appendix A

Proof of Theorem 3. The proof is given in an alternating minimization sense. First, fix V and define $\hat{J}_m(U) = J_m(U, V)$, for $U \in \mathcal{U}_c^N$. Since the membership values $[\mu_{ik}]_{i=1}^c$ of z_k to the fixed clusters are not depending on the memberships of other data points, the columns of U degenerate to each other (decoupled) in the minimization of $\hat{J}_m(U)$, therefore:

$$\min_{U\in\mathcal{U}_c^N} \hat{J}_m(U) = \min_{U\in\mathcal{U}_c^N} \max_{k\in\mathbb{I}_1^N} \sum_{i=1}^{c} \mu_{ik}^m d_{ik} = \max_{k\in\mathbb{I}_1^N} \min_{U\in\mathcal{U}_c^N} \sum_{i=1}^{c} \mu_{ik}^m d_{ik}.$$

Denote $\hat{J}_m^{(k)}(U) = \sum_{i=1}^c \mu_{ik}^m d_{ik}$. To introduce the constraints $\mathcal{U}_c^N$, the Lagrangian $\Lambda_k(\lambda_k, U)$ of $\hat{J}_m^{(k)}(U)$ is defined for each $k \in \mathbb{I}_1^N$ as

$$\Lambda_k(\lambda_k, U) = \sum_{i=1}^{c} \mu_{ik}^m d_{ik} - \lambda_k \left[\left(\sum_{i=1}^{c} \mu_{ik} \right) - 1 \right]. \tag{28}$$

Assume that $\mathcal{I}_s^{(k)} = \emptyset$, then (λ_k, U) is a stationary point for Λ_k, only if $\nabla_{\lambda,U}\Lambda_k(\lambda_k, U) = \left(0^N, 0^{c\times N}\right)$ for all $k \in \mathbb{I}_1^N$.Setting all of these gradients equal to zero yields that

$$\frac{\partial \Lambda_k(\lambda_k, U)}{\partial \lambda_k} = \sum_{i=1}^{c} \mu_{ik} - 1 = 0, \tag{29a}$$

$$\frac{\partial \Lambda_k(\lambda_k, U)}{\partial \mu_{ik}} = m\mu_{ik}^{m-1} d_{ik} - \lambda_k = 0, \tag{29b}$$

for every $k \in \mathbb{I}_1^N$ and $i \in \mathbb{I}_1^c$. From (29b), it follows that

$$\mu_{ik} = \left(\frac{\lambda_k}{m d_{ik}}\right)^{\frac{1}{m-1}}. \tag{30}$$

Moreover, by substitution of (30) into (29a):

$$0 = \sum_{l=1}^{c} \left(\frac{\lambda_k}{m}\right)^{\frac{1}{m-1}} \left(\frac{1}{d_{lk}}\right)^{\frac{1}{m-1}} - 1 \tag{31a}$$

$$\left(\frac{\lambda_k}{m}\right)^{\frac{1}{m-1}} = \left[\sum_{l=1}^{c} \left(\frac{1}{d_{lk}}\right)^{\frac{1}{m-1}}\right]^{-1}. \tag{31b}$$

If (31b) is substituted back into (30), it follows that

$$\mu_{ik} = \frac{\left(\frac{1}{d_{ik}}\right)^{\frac{1}{m-1}}}{\sum_{l=1}^{c} \left(\frac{1}{d_{lk}}\right)^{\frac{1}{m-1}}} = \frac{1}{\sum_{l=1}^{c} \left(\frac{d_{ik}}{d_{lk}}\right)^{\frac{1}{m-1}}}. \tag{32}$$

In this way we have proved that in a local minimum of $J_m(U, V)$, all μ_{ik} have to satisfy (16). If $\mathcal{I}_s^{(k)} \neq \emptyset$, then (32) is singular. In this situation, choosing μ_{ik} as given by (16) results in $\hat{J}_m^{(k)}(U) = 0$, because the non-zero weights are placed on zero distances, while positive distances with non-zero weights would increase $\hat{J}_m^{(k)}(U)$, contradicting minimality. As the zero-distances can have arbitrary weights, for the sake of simplicity equal weights are considered fulfilling (16). Note, that such a singularity hardly occurs in reality, since machine round-off prevents its encounter.

To establish (17), fix $U \in \mathcal{U}_c^N$ and define $\check{J}_m(V) = J_m(U, V)$. Minimization of $\check{J}_m(V)$ is unconstrained on $\mathbb{D}^c$, and it is decoupled for each v_i. Therefore

$$\min_{V\in\mathbb{D}^c} \check{J}_m(V) = \min_{V\in\mathbb{D}^c} \max_{k\in\mathbb{I}_1^N} \sum_{i=1}^{c} \mu_{ik}^m d_{ik} = \sum_{i=1}^{c} \min_{V\in\mathbb{D}^c} \check{J}_m^{(i)}(V),$$

where $\check{J}_m^{(i)}(V) = \max_{k\in\mathbb{I}_1^N} \mu_{ik}^m d_{ik}$, depending only on v_i. This means that

$$v_i = \arg\min_{V\in\mathbb{D}^c} \check{J}_m^{(i)}(V) = \arg\min_{v_i\in\mathbb{D}} \max_{k\in\mathbb{I}_1^N} \mu_{ik}^m d_{ik}. \tag{33}$$

Optimization (33) can be formulated as a matrix inequality constrained minimization problem. Denote

$$\gamma_i = \check{J}_m^{(i)}(V) = \max_{k\in\mathbb{I}_1^N} \mu_{ik}^m d_{ik}, \tag{34}$$

then the solution of (33) can be obtained by solving

$$\begin{aligned} &\text{minimize} && \gamma_i \geq 0, \\ &\text{subject to} && \mu_{ik}^m \left|\frac{z_k - v}{1 - z_k^* v}\right| \leq \gamma_i, \forall k \in \mathbb{I}_1^N, \\ & && v \in \mathbb{D}. \end{aligned}$$

The constraints of this minimization can be written for each k as

$$\mu_{ik}^m \left|\frac{z_k - v}{1 - z_k^* v}\right| \leq \gamma_i, \tag{35a}$$

$$\mu_{ik}^{2m} |z_k - v|^2 \left|1 - z_k^* v\right|^{-2} \leq \gamma_i^2. \tag{35b}$$

From the Schur-complement of (35b) it follows that (35a) holds iff

$$\begin{bmatrix} |1 - z_k^* v|^2 & \mu_{ik}^m (z_k - v) \\ \mu_{ik}^m (z_k - v)^* & \gamma_i^2 \end{bmatrix} \succeq 0, \quad \forall k \in \mathbb{I}_1^N, \tag{36}$$

where $v \in \mathbb{D}$. Then a sufficient but not necessary condition for (U, V) being a local minimum of J_m is to satisfy (32) and (33). This concludes the proof. It is important to remark that $J_m(U, V)$ has more stationary points than what can be reached through alternating minimization, however all points fulfilling Theorem 3 are stationary points of $J_m(U, V)$. □

Proof of Theorem 4. As the cluster centers of V are assumed to be "non-singular" with respect to Z, i.e. $d_{ik} > 0$ for all $(i, k) \in \mathbb{I}_1^c \times \mathbb{I}_1^N$, thus based on the optimality of U_m, substitution of (32) into (14) implies, that for $m > 1$:

$$\begin{aligned} J_m(U_m, V) &= \max_{k\in\mathbb{I}_1^N} \sum_{i=1}^{c} \mu_{ik}^m d_{ik} = \max_{k\in\mathbb{I}_1^N} \sum_{i=1}^{c} \mu_{ik} \mu_{ik}^{m-1} d_{ik} \\ &= \max_{k\in\mathbb{I}_1^N} \sum_{i=1}^{c} \mu_{ik} \frac{d_{ik}}{d_{ik}\left[\sum_{l=1}^{c} \left(\frac{1}{d_{lk}}\right)^{\frac{1}{m-1}}\right]^{m-1}} \\ &= \max_{k\in\mathbb{I}_1^N} \left[\sum_{l=1}^{c} (d_{lk})^{\frac{1}{1-m}}\right]^{1-m}, \end{aligned}$$

holds as $\sum_{i=1}^{c} \mu_{ik} = 1$. Now introduce

$$\bar{J}_t^{(k)}(V) = \left[\sum_{i=1}^{c} \frac{1}{c} (d_{ik})^t\right]^{1/t}, \tag{37}$$

with $t = \frac{1}{1-m}$. Then

$$J_m(U_m, V) = J_{\frac{t-1}{t}}(U_{\frac{t-1}{t}}, V) = c^{1/t} \max_{k\in\mathbb{I}_1^N} \bar{J}_t^{(k)}(V).$$

Eq. (37) is called the *Hölder* or *generalized mean* (Bullen, 2003) of d_{ik}. Based on the properties of the generalized mean in terms of t, the following hold:

Case $m \to 1 \Leftrightarrow t \to -\infty \Rightarrow \bar{J}_t^{(k)}(V) \to \min_{i\in\mathbb{I}_1^c}\{d_{ik}\}$ for all $k \in \mathbb{I}_1^N$. Since $c^{1-m} \to 1$, the minimum over $\mathbb{I}_1^c$ is unique for each k:

$$\lim_{m\to 1} J_m(U_m, V) = \max_{k\in\mathbb{I}_1^N} \min_{i\in\mathbb{I}_1^c} \{d_{ik}\}. \tag{38}$$

Case $m = 2 \Leftrightarrow t = -1$. Then $\bar{J}_{-1}^{(k)}(V)$ is the harmonic mean of $\{d_{ik}\}_{i=1}^c$ for each $k \in \mathbb{I}_1^N$, so

$$J_2(U_2, V) = \frac{1}{c} \max_{k\in\mathbb{I}_1^N} \frac{c}{\sum_{i=1}^{c} \frac{1}{d_{ik}}}. \tag{39}$$

Case $m \to \infty \Leftrightarrow t \to 0$. Then, the asymptotic convergence of the generalized mean to the geometric mean yields: $\bar{J}_t^{(k)}(V) = \left[\prod_{i=1}^{c} d_{ik}\right]^{1/c} + \mathcal{O}(e^{\frac{1}{t}})$, which gives

$$J_m(U_m, V) = c^{1-m} \max_{k\in\mathbb{I}_1^N} \left[\prod_{i=1}^{c} d_{ik}\right]^{\frac{1}{c}} + \mathcal{O}(e^{-m}),$$

and since $c^{1-m} \to 0$, therefore

$$\lim_{m\to\infty} J_m(U_m, V) = 0. \quad \square$$

References

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Plenum Press.

Bodin, P., Villemoes, L., & Wahlberg, B. (1997). An algorithm for selection of best orthonormal rational basis. In *Proc. of the 36th IEEE Conf. on decision and control*.

Brannan, D. A., Esplen, M. F., & Gray, J. J. (1999). *Geometry*. Cambridge University Press.

Bullen, P. S. (2003). *Handbook of means and their inequalities*. Kluwer Academic Publishers.

Dettori, M., & Scherer, C. W. (2001). LPV design for a CD player: An experimental evaluation of performance. In *Proc. of the 40th IEEE conf. on decision and control* (pp. 4711–4716).

dos Santos, P. L., Ramos, J. A., & de Carvalho, J. L. M. (2007). Identification of linear parameter varying systems using an iterative deterministic-stochastic subspace approach. In *Proc. of the European control conf.* (pp. 4867–4873).

Felici, F., van Wingerden, J. W., & Verhaegen, M. (2006). Subspace identification of MIMO LPV systems using a periodic scheduling sequence. *Automatica*, *43*(10), 1684–1697.

Felici, F., van Wingerden, J. W., & Verhaegen, M. (2007). Dedicated periodic scheduling sequences for LPV system identification. In *Proc. of the European control conf.* (pp. 4896–4902).

Giarré, L., Bauso, D., Falugi, P., & Bamieh, B. (2006). LPV model identification for gain scheduling control: An application to rotating stall and surge control problem. *Control Engineering Practice*, *14*, 351–361.

Heuberger, P. S. C., Van den Hof, P. M. J., & Wahlberg, Bo (2005). *Modeling and identification with rational orthonormal basis functions*. Springer-Verlag.

Heuberger, P. S. C., Van den Hof, P. M. J., & Bosgra, O. H. (1995). A generalized orthonormal basis for linear dynamical systems. *IEEE Transactions on Automatic Control*, *40*(3), 451–465.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice Hall.

Kaymak, U., & Setnes, M. (2002). Fuzzy clustering with volume prototypes and adaptive cluster merging. *IEEE Transactions on Fuzzy Systems*, *10*(6), 705–711.

Lee, L. H., & Poolla, K. R. (1996). Identification of linear parameter-varying systems via LFTs. In *Proc. of the 35th IEEE conf. on decision and control* (pp. 1545–1550).

Ljung, L. (1999). *System identification, theory for the user*. Prentice Hall.

Ljung, L. (2006). *System identification toolbox, for use with Matlab*. The Mathworks Inc.

Mäkilä, P. M., & Partington, J. R. (1993). Robust approximate modeling of stable linear systems. *International Journal of Control*, *58*(3), 665–683.

Marcos, A., & Balas, G. J. (2004). Development of linear-parameter-varying models for aircraft. *Journal of Guidance, Control and Dynamics*, *27*(2), 218–228.

Mazzaro, M. C., Movsichoff, E. A., & Pena, R. S. S. (1999). Robust identification of linear parameter varying systems. In *Proc. of the American control conf.* (pp. 2282–2284).

Milanese, M., & Vicino, A. (1991). Optimal estimation theory for dynamic systems with set membership uncertainty : An overview. *Automatica*, *27*(6), 997–1009.

Murray-Smith, R., & Johansen, T. A. (1997). *Multiple model approaches to modeling and control*. Taylor and Francis.

Murray-Smith, R., Johansen, T. A., & Shorten, R. (1999). On the interpretation of local models in blended multiple model structures. *International Journal of Control*, *72*(7–8), 620–628.

Ninness, B. M., & Gustafsson, F. (1997). A unifying construction of orthonormal bases for system identification. *IEEE Transactions on Automatic Control*, *42*(4), 515–521.

Oliveira e Silva, T. (1996). A n-width result for the generalized orthonormal basis function model. In *Proc. of the 13th IFAC world congress* (pp. 375–380).

Pinkus, A. (1985). *n-Widths in approximation theory*. Springer-Verlag.

Rugh, W., & Shamma, J. (2000). Research on gain scheduling. *Automatica*, *36*(10), 1401–1425.

Scherer, C. W. (1996). Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control for time-varying and linear parametrically-varying systems. *International Journal of Robust and Nonlinear Control*, *6*(9–10), 929–952.

Scherer, C. W., & Hol, C. W. J. (2006). Matrix sum-of-squares relaxations for robust semi-definite programs. *Mathematical Programming*, *107*(1), 189–211.

Steinbuch, M., van de Molengraft, R., & van der Voort, A. (2003). Experimental modeling and LPV control of a motion system. In *Proc. of the American control conf.* (pp. 1374–1379).

Sturm, J. (1999). *Optimization Methods and Software*, *11–12*, 625–653.

Sznaier, M., Mazzaro, C., & Inanc, T. (2000). An LMI approach to control oriented identification of LPV systems. In *Proc. of the American control conf.* (pp. 3682–3686).

Tóth, R., Felici, F., Heuberger, P. S. C., & Van den Hof, P. M. J. (2007). Discrete time LPV I/O and state space representations, differences of behavior and pitfalls of interpolation. In *Proc. of the European control conf.* (pp. 5418–5425).

Tóth, R., Heuberger, P. S. C., & Van den Hof, P. M. J. (2006a). Optimal pole selection for LPV system identification with OBFs, a clustering approach. In *Proc. of the 14th IFAC symposium on system identification* (pp. 356–361).

Tóth, R., Heuberger, P. S. C., & Van den Hof, P. M. J. (2006b). Orthonormal basis selection for LPV system identification, the Fuzzy-Kolmogorov c-Max approach. In *Proc. of the 45th IEEE Conf. on decision and control* (pp. 2529–2534).

Tóth, R., Heuberger, P. S. C., & Van den Hof, P. M. J. (2007). LPV system identification with globally fixed orthonormal basis functions. In *Proc. of the 46th IEEE conf. on decision and control* (pp. 3646–3653).

Verdult, V., Ljung, L., & Verhaegen, M. (2002). Identification of composite local linear state-space models using a projected gradient search. *International Journal of Control*, *75*(16–17), 1125–1153.

Verdult, V., & Verhaegen, M. (2002). Subspace identification of multivariable linear parameter-varying systems. *Automatica*, *38*(5), 805–814.

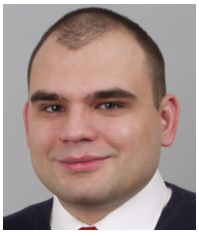
Vergeer, N. (2005). *Identification of LPV systems using orthonormal basis functions. Master's thesis*. Delft University of Technology.

Wassink, M. G., van de Wal, M., Scherer, C. W., & Bosgra, O. (2004). LPV control for a wafer stage: Beyond the theoretical solution. *Control Engineering Practice*, *13*, 231–245.

Wei, X., & Del Re, L. (2006). On persistent excitation for parameter estimation of quasi-LPV systems and its application in modeling of diesel engine torque. In *Proc. of the 14th IFAC symposium on system identification* (pp. 517–522).

Xie, X. L. X., & Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(8), 841–847.

Zhou, K., & Doyle, J. C. (1998). *Essentials of robust control*. Prentice-Hall.



Roland Tóth was born in 1979 in Miskolc, Hungary. He received the B.Sc. degree in Electrical Engineering and the M.Sc. degree in Information Technology in parallel at the University of Pannonia, Hungary in 2004. In 2008, he received his Ph.D. at the Delft Center for Systems and Control (DCSC) of the Delft University of Technology, The Netherlands. Currently he is a post doctoral researcher at DCSC.

Roland Tóth's research interests are in issues of linear parameter-varying (LPV) and nonlinear system identification and modeling, electrical motor control, and behavioral system theory. He is a member of IEEE.

Peter S.C. Heuberger was born in Maastricht, The Netherlands, in 1957. He obtained the M.Sc. degree in Mathematics from the University of Groningen in 1983, and the Ph.D. degree from the Mechanical Engineering Department of Delft University of Technology in 1991.

From 1991 to 2006 he was with the Dutch National Institute for Public Health and the Environment (RIVM) and from 2006 to 2007 with the Dutch Environmental Assessment Agency (MNP). From 1996 to 2007 he also held a position at Delft University of Technology, in the Mechanical Engineering Systems and Control Group and the Signals, Systems and Control Group of the Department of Applied Physics. Since 2007 he is an assistant professor at the Delft Center for Systems and Control. His research interests are in issues of system identification, uncertainty, model reduction, linear parameter-varying systems and in the theory and application of orthogonal basis functions.

Paul M.J. Van den Hof received the M.Sc. and Ph.D. degree from the Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands in 1982 and 1989, respectively. In 1986 he joined the Mechanical Engineering Systems and Control Group of Delft University of Technology, The Netherlands until 1999 when he was appointed as full professor in the Department of Applied Physics at the same university. Currently he is Director of the Delft Center for Systems and Control (established in 2004), which is a merger of three former systems and control groups from different departments within Delft. Since 2005 he also serves as scientific director of the Dutch Institute of Systems and Control (disc).

His research work has concentrated around system identification and identification for control, including aspects of closed-loop experiments, uncertainty modeling and parametrization issues. His research interests further cover signal processing and (robust) control design, while dealing with applications in industrial process control systems, physical measurement systems and motion control systems.

He has acted as General Chair for the 13th IFAC Symposium on System Identification, that was held in Rotterdam, the Netherlands in 2003. He has been a member of the editorial board of Automatica (1992–2005), the IFAC Council (1999–2005), the Administrative Council of the European Union Control Association EUCA (1999–2005) and the Board of Governors of IEEE's Control System Society (2003–2005). He is an IFAC Fellow and a Fellow of IEEE.