Orthonormal basis selection for LPV system identification, the Fuzzy-Kolmogorov c-Max approach

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Proceedings of the 45th IEEE Conference on Decision & Control
Manchester Grand Hyatt Hotel
San Diego, CA, USA, December 13-15, 2006

ThA05.6

Orthonormal basis selection for LPV system identification, the Fuzzy-Kolmogorov c-Max approach

R. Tóth, P.S.C Heuberger, and P.M.J Van den Hof

Abstract— A fuzzy clustering approach is developed to select pole locations for Orthonormal Basis Functions (OBFs), used for identification of Linear Parameter Varying (LPV) systems. The identification approach is based on interpolation of locally identified Linear Time Invariant (LTI) models, using globally fixed OBFs. Selection of the optimal OBF structure, that guarantees the least worst-case local modelling error in an asymptotic sense, is accomplished through the fusion of the Kolmogorov n-width (KnW) theory and Fuzzy c-Means (FcM) clustering of observed sample system poles.

I. Introduction

In general, many physical systems and control problems suffer from parameter variations due to non-stationarity, non-linear behavior, or dependence on independent variables, such as space coordinates. These systems vary in size and complexity from highly advanced aircrafts [1] to induction motors [2], but they share the common need for accurate and efficient control of the relevant process variables, which has to satisfy the rapidly increasing industrial performance demands. However, accurate modelling of such systems is in general a complex and tedious task, involving the use of non-linear partial differential equations, leading to models with many parameters and high computational complexity.

For processes with mild non-linearities, the theory of LPV systems offers an attractive framework for modelling and handling non-linear or time-varying dynamics. These systems are generally described in a *state-space representation* (SSR), where the state-space matrices are usually affine functions of a time-varying parameter vector $\zeta : \mathbb{Z} \to \Gamma$. Here Γ denotes the parameter space. Furthermore, control design in the LPV framework can be carried out by using LTI control theory via *gain scheduling* [3]. Therefore, the LPV approach can offer a useful venue to meet recent industrial demands. However, existing methods for identification of such systems often produce models with high complexity or – for instance with subspace techniques – with substantial computational load. Because most control design methods require low-order models and fast iterations in the identification process, it is a challenge to develop efficient methods for LPV system identification that yield models with limited complexity and computation time. An additional point of concern is that the McMillan degree of the system may change due to variations of ζ, especially when the approach is based on interpolation of local models. One way to overcome these

problems is to use a fixed-order model structure, capable to represent the system globally on Γ. Then, identification of the LPV system is performed by estimating multiple 'local' LTI systems around well-placed parameter set-points $\{\zeta_i\} \in \Gamma$ [4], [5]. These local models are subsequently interpolated to synthesize the desired low-order LPV model. However, if the LTI model structure, used for local identification, is not linear in the parameters, the interpolation of the estimated models represents a NP-hard non-linear optimization problem with pitfalls of local minima or existence of solution. Therefore, the choice of an easily interpolatable model structure is a crucial point of this identification approach.

The OBFs-based model representation offers such a structure with a well worked-out theory in the context of LTI system approximation and identification [6]. The basis functions 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

$$F(z) = \sum_{k=1}^{\infty} f_k \phi_k(z), \tag{1}$$

where $\{f_k\}_{k=1}^{\infty}$ is the set of coefficients and $\{\phi_k\}_{k=1}^{\infty}$ represents the sequence of OBFs. 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 modelling error with a relatively small number of parameters, due to the infinite impulse response characteristics of the basis. Furthermore, interpolation of these models can easily take place through the interpolation of the coefficients if the set of OBFs is the same for each local representation. Here, an essential challenge is to derive a set of OBFs, 'sufficiently rich' to describe the varying LPV dynamics at each local parameter point ζ_i, with a predefined number of parameters.

In practice, if the physical system is stable, it is a reasonable assumption that some sampled pole locations of the ζ-dependent pole movements of the LPV system, possibly with uncertainty bounds, are available as prior information. These can for instance be the result of preliminary local identification experiments in different points. If a region Ω in the unit disc $\mathbb{D}$ is also given, where the pole variations are guaranteed to take place, then the problem of efficient OBF set selection with a pre-defined number (n) of basis functions, can be tackled through the usage of the KnW theory for OBFs, derived by Oliveira e Silva [7]. This

approach provides the selection of a set of OBFs, that ensures the least possible worst-case local modelling error for the LPV system at any point of Γ. However, determining the region Ω from the sampled pole locations in a robust sense is not a trivial task. In this paper we aim at a well applicable solution, based on the joint application of the KnW result and FcM clustering, which is capable of determining such region Ω and its associated asymptotically optimal basis in the KnW sense. This solution, which is an improved version of the algorithms presented in [8] and [9], is developed to solve the first two steps of a proposed approach for identifying LPV systems. This approach consists of the following steps:

1) Determination of pole regions by FcM clustering of sampled pole locations
2) Determination of OBF's, based on KnW optimization
3) Estimation of local models, with the optimal OBF's
4) Interpolation of the local model coefficients

In the sequel, the pole clustering approach is only presented in the context of Step 1 and 2. The paper is organized as follows: Section 2 introduces the description and properties of OBFs; Section 3 describes the n-width result for OBFs to be used later on; in Section 4 the mechanism of the KnW based FcM pole clustering is given; in Section 5 the applicability of the introduced method is shown through an example; and finally, in Section 6, the main results of the paper are discussed.

II. ORTHONORMAL BASIS FUNCTIONS

Because of space limitations we only consider the case of real rational (finite-dimensional) discrete-time, SISO transfer functions. For details see [10], [11], [6]. Let $G_0 = 1$ 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 balanced SSRs of G_i. Let $\{\xi_1, \xi_2, \ldots\}$ denote the collection of all poles of the inner functions $G_1, G_2, \cdots$. Under the (completeness) condition that $\sum_{i=1}^\infty (1-|\xi_i|) = \infty$, the scalar elements of the sequence of vector functions

$$V_k(z) = (zI - A_k)^{-1} B_k \prod_{j=0}^{k-1} G_j(z), \qquad (2)$$

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 $\mathbb{D}$, and square integrable on the unit circle $\mathbb{T}$ with norm $\|.\|_{\mathcal{H}_2}$. These functions (2) are often referred to as the *Takenaka-Malmquist functions*. The special cases when all G_i are equal, i.e. $G_i(z) = G_b(z)$ for $\forall i > 0$, where G_b has McMillan degree $n_b > 0$, are known as *Hambo functions* or *generalized orthonormal basis functions* (GOBFs) for arbitrary n_b, *2-parameter Kautz functions* for $n_b = 2$, and as *Laguerre functions* for $n_b = 1$. 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 $n_b > 0$ and input-balanced SSR $\{A_b, B_b, C_b, D_b\}$. Define $V_1(z) = (zI - A_b)^{-1} B_b$ and $\phi_j = (V_1)_j$, $j = 1, \cdots, n_b$. The Hambo basis then consists of the functions $\{\phi_j(z)G_b^i\}_{j=1,\cdots,n_b}^{i=0,\cdots,\infty}$. An

important aspect of these bases is that the inner function G_b is, modulo the sign, completely determined by its poles $\{\xi_1, \cdots, \xi_{n_b}\} = \Xi_{n_b}$:

$$G_b(z) = \pm \prod_{j=1}^{n_b} \frac{1 - z\xi_j^*}{z - \xi_j}, \qquad (3)$$

and it is immediate that the function V_1 has the same poles. Any $F \in \mathcal{H}_{2-}(\mathbb{E})$ can be written as

$$F(z) = \sum_{i=0}^\infty \sum_{j=1}^{n_b} f_{ij}\phi_j(z)G_b^i(z), \qquad (4)$$

and it can be shown that the rate of convergence of this series is bounded by $\max_k |G_b(\lambda_k^{-1})|$, where $\{\lambda_k\}$ are the poles of F. In the best case, where the poles of F are the same as the poles of G_b, only the terms with $i = 0$ in (4) are non-zero.

III. KOLMOGOROV n-WIDTH FOR OBFs

Finding appropriate model sets to perform system identification is a much studied problem with the main conclusion, that in general each particular identification problem requires a model set that is tailored to the characteristics of the system to be identified. An arbitrary model set is adequate only to approximate a certain subset of $\mathcal{H}_{2-}(\mathbb{E})$, in the sense that the model set is sufficiently rich to describe only the systems belonging to that subset, with a relatively small number of statistically meaningful parameters. One approach to find appropriate model sets is based on the n-width concept [12], which was shown to result in appropriate model sets for robust modelling of linear systems [13]. Oliveira e Silva [7], [6, Ch. 11] showed that GOBF model structures are optimal for specific subsets of systems. In the following, the basic ingredients of this approach for discrete time, stable, SISO systems are described.

Let $\mathcal{S} \subset \mathcal{H}_{2-}(\mathbb{E})$ denote a set of systems whose optimal approximation is needed. Let $\Phi_n = \{\phi_i\}_{i=1}^n$ be a sequence of n linearly independent elements of $\mathcal{H}_{2-}(\mathbb{E})$, and let $\Psi_n = \text{Span}(\Phi_n)$. The distance $d_{\mathcal{H}_{2-}}(F, \Psi_n)$ between $F \in \mathcal{H}_{2-}(\mathbb{E})$ and Ψ_n is defined as

$$d_{\mathcal{H}_{2-}}(F, \Psi_n) = \inf_{H \in \Psi_n} \|F - H\|_{\mathcal{H}_2}. \qquad (5)$$

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

$$w_n(\mathcal{S}, \mathcal{H}_{2-}(\mathbb{E})) = \inf_{\Psi_n \in M_n} \sup_{F \in \mathcal{S}} d_{\mathcal{H}_{2-}}(F, \Psi_n), \qquad (6)$$

which means the smallest possible approximation error for the worst-case F in $\mathcal{S}$. The subspace $\breve{\Psi}_n \in M_n$, for which w_n is minimal, is called the optimal subspace in the KnW sense. A well know result in this context is that the set of *pulse functions* $\{z^{-i}\}_{i=1}^n$ is optimal for the class of stable systems analytical in the region $(|z| > R)$, $R \in \mathbb{R}_0^+$. The worst-case approximation error is proportional to R^n.

Let G_b be an inner function with McMillan degree $n_b > 0$, and let $\{\phi_j\}_{j=1}^{n_b}$ be the first n_b Hambo functions as defined in the previous section. Denote by $\mathcal{S} \subset \mathcal{H}_{2-}(\mathbb{E})$ the set of functions that are analytic in the region $\{z, |G_b(z^{-1})| > \rho\}$,

and are square integrable on the boundary of that region, where $\rho > 0$ is often referred to as the *decay rate*.

Proposition 1: (Oliveira e Silva, 1996). For $\forall n_e \in \mathbb{N}$, Span $\left\{ \phi_j(z) G_b^i(z) \right\}_{j=1,\ldots,n_b}^{i=0,\ldots,n_e-1}$ is optimal in the Kolmogorov $n = n_e \times n_b$-width sense for the set $\mathcal{S}$. The worst-case approximation error is proportional to ρ^{n_e}.

This remarkable result shows that for the specified region one can not improve on the worst case error by adding new poles to the n_b basis poles.

In practical situations we are dealing with the opposite problem, referred as the *inverse Kolmogorov problem*, where we are given a region of non-analyticity $\Omega \subset \mathbb{D}$, for instance based on sampled pole locations and we want to find the inner function G_b to describe/approximate this region in the form $\Omega(\rho, \Xi_{n_b}) = \{ z, |G_b(z^{-1})| \leq \rho \}$ with ρ as small as possible. For a given number of poles n_b, this comes down to the following min-max problem:

$$\min_{\xi_1, \cdots, \xi_{n_b}} \max_{z \in \Omega} \prod_{j=1}^{n_b} \left| \frac{z - \xi_j}{1 - z\xi_j^*} \right|. \tag{7}$$

See [6, Chapters 10 and 11] for details on this non-linear optimization problem and solution methods.

IV. FUZZY-KOLMOGOROV c-MAX CLUSTERING

A. The clustering algorithm

Objective-function-based, fuzzy clustering algorithms, such as the FcM, have been used in a wide collection of applications [14], [15]. Generally, FcM partitions the data into overlapping groups, that describe an underlying structure within the data [16]. This enables the determination of the region Ω on the basis of the observed poles, not only by hard borders, but with membership based, overlapping areas, which incorporate both the local and the global data coherency. Moreover, FcM clustering does not rely on assumptions about the data common to other methods, making it useful in the situation of pole clustering where little prior knowledge exists. To exploit these fruitful properties, in the following such a *Fuzzy-Kolmogorov c-Max* (FKcM) algorithm is presented, which provides a joint solution for Step 1 and 2 of the proposed identification scheme, based on the fusion of the KnW theory and the FcM technique.

Let c be the number of clusters and $\mathcal{I}_c$ an index set defined as $\mathcal{I}_x = \{1, \ldots, x\} \subset \mathbb{N}$. A cluster $i \in \mathcal{I}_c$ is represented by its center (or *prototype*) $v_i \in \mathbb{D}$. Furthermore, membership functions $\mu_i : \mathbb{D} \to [0\ 1]$ determine the 'degree of membership' to cluster i for $z \in \mathbb{D}$. By using a *threshold value* ε, we obtain a set

$$\Omega = \{ z \in \mathbb{D} \mid \exists i \in \mathcal{I}_c \quad \mu_i(z) \geq \varepsilon \}. \tag{8}$$

We can now formalize the problem we will consider.

Problem 2: For a given c, find a region Ω, as described above, such that Ω contains all sampled pole locations (and possibly the associated uncertainty areas), and such that the GOBFs, with poles in the cluster centers $\{v_i\}_{i=1}^c$, are optimal in the KnW sense, $n = c$, with respect to Ω and with the corresponding decay rate ρ as small as possible.

The solution is based on finding clusters in concord with the KnW concept, and subsequently finding a maximal value for ε such that all sampled poles are inside Ω. The latter is equivalent to minimizing ρ. In the sequel we focuss on the first part, i.e. finding n-width-based clusters.

Let $\mathbf{Z} = [z_k]_{k=1}^N \in \mathbb{D}^N$, be the set of $N \in \mathbb{N}$ observed poles for clustering. Denote $\mathbf{V} = [v_i]_{i=1}^c$, and introduce the membership matrix $\mathbf{U} = [\mu_{ik}]_{c \times N}$, where μ_{ik} is the degree of membership of z_k to cluster i. To constrain the clustering it is required that $\mathbf{U} \in \mathcal{M}_{cN}$, where

$$\mathcal{M}_{cN} = \left\{ \mathbf{U} \in [0,1]^{c \times N} \mid \sum_{i=1}^c \mu_{ik} = 1 \text{ for } \forall k \in \mathcal{I}_N, \right. \tag{9}$$
$$\left. 0 < \sum_{k=1}^N \mu_{ik} < N \text{ for } \forall i \in \mathcal{I}_c \right\}$$

characterizes the *fuzzy constraints*.

Furthermore, distances d_{ik} are introduced between v_i and z_k to measure dissimilarity of $\mathbf{Z}$ with respect to each cluster. The dissimilarity measure needs to be closely connected with the original KnW problem to derive an algorithmic solution of Problem 2. Thereto, the following metric on $\mathbb{D}$ is chosen:

$$\kappa(x, y) = \left| \frac{x - y}{1 - x^* y} \right| : \mathbb{D} \times \mathbb{D} \to \mathbb{R}_0^+, \tag{10}$$

called the *Kolmogorov metric* (KM), and we will use $d_{ik} = \kappa(v_i, z_k)$. In the sequel, it is going to be shown that KM relates the FcM asymptotically to the KnW theory, and in this way to the solution of Problem 2.

Fuzzy clustering can be viewed as the minimization of the FcM-*functional* [14], $J_m(\mathbf{U}, \mathbf{V}) : \mathcal{M}_{cN} \times \mathbb{D} \to \mathbb{R}_0^+$, which in the FK$c$M case can be formulated as

$$J_m(\mathbf{U}, \mathbf{V}) = \max_{k \in \mathcal{I}_N} \sum_{i=1}^c \mu_{ik}^m d_{ik}. \tag{11}$$

Here the design parameter $m \in (1, \infty)$ determines the fuzziness of the resulting partition. It can be observed, that (11) corresponds to a *worst-case (max) sum of error* criterion, contrary to the *mean squared error* criterion of the original FcM, see [14]. The exact relation of (11) with the KnW optimality of $(\mathbf{U}, \mathbf{V})$ is explained later. The following theorem yields the ingredients for the approach to solve Problem 2:

Theorem 3: (Optimal Partition) For a fixed $m \in (1, \infty)$, the fuzzy partition $(\hat{\mathbf{U}}, \hat{\mathbf{V}})$ can only be a local (global) minimum of J_m if for $\forall i \in \mathcal{I}_c$ and $\forall k \in \mathcal{I}_N$:

$$\hat{\mu}_{ik} = \begin{cases} \left[\sum_{l=1}^c \left(\frac{d_{ik}}{d_{lk}} \right)^{\frac{1}{m-1}} \right]^{-1} & \text{if } \mathcal{I}_s^{(k)} = \varnothing, \\ \frac{1}{n_k} & \text{if } i \in \mathcal{I}_s^{(k)}, \\ 0 & \text{if } i \notin \mathcal{I}_s^{(k)} \neq \varnothing, \end{cases} \tag{12}$$

$$\hat{v}_i = \arg \min_{v \in \mathbb{D}} \gamma_i(v, \hat{\mathbf{U}}), \tag{13}$$

where $d_{ik} = \kappa(\hat{v}_i, z_k)$, $\mathcal{I}_s^{(k)} = \{ i \in \mathcal{I}_c \mid d_{ik} = 0 \}$ (*singularity set*) with Card $\left(\mathcal{I}_s^{(k)} \right) = n_k$ and $\gamma_i(v, \mathbf{U})$ is the minimal

value of $\gamma \in [0, 1]$ fulfilling the constraints:

$$\begin{bmatrix} v \\ 1 \end{bmatrix}^* \left(\mu_{ik}^{2m} \Upsilon_{1k} - \gamma^2 \Upsilon_{2k} \right) \begin{bmatrix} v \\ 1 \end{bmatrix} \leq 0, \ \forall k \in \mathcal{I}_N, \quad (14)$$

where

$$\Upsilon_{1k} = \begin{bmatrix} 1 & -z_k \\ -z_k^* & z_k^* z_k \end{bmatrix}, \ \Upsilon_{2k} = \begin{bmatrix} z_k^* z_k & -z_k \\ -z_k^* & 1 \end{bmatrix}.$$

The proof is omitted because of space limitations.
In the FcM case, minimization of (11) subject to (9) is usually tackled by alternating optimization [14], steering the solution towards a settling partition in the sense of Theorem 3. For the FKcM this yields the following algorithm:

Algorithm 4: (Fuzzy-Kolmogorov c-Max)

A5.1 *Initialization*
Fix c and m; and initialize $\mathbf{V}^{(0)} \in \mathbb{D}^c$, $l = 0$.

A5.2 *Membership update*
With (12), solve $\mathbf{U}^{l+1} = \arg \min_{\mathbf{U} \in \mathcal{M}_{cN}} J_m \left(\mathbf{U}, \mathbf{V}^l \right)$.

A5.3 *Cluster center update*
With (13), solve $\mathbf{V}^{l+1} = \arg \min_{\mathbf{V} \in \mathbb{D}^c} J_m \left(\mathbf{U}^{l+1}, \mathbf{V} \right)$.

A5.4 *Check of convergence*
If $J_m \left(\mathbf{U}^{l+1}, \mathbf{V}^{l+1} \right)$ has converged, then stop, else $l = l + 1$ and goto Step A5.2.

B. Properties of the FKcM

In order to explain the specific choices for the fuzzy functional (11) and the dissimilarity measure (10), we use the following theorem.

Theorem 5: (Limiting property of J_m)

Let $\{(\mathbf{U}, \mathbf{V})\}$ be a collection of optimal partitions for J_m each associated with a different m and with no singularity (i.e. $\mathcal{I}_s^{(k)} = \varnothing$ for $\forall k \in \mathcal{I}_N$ and $\forall m$), then:

a. $\lim_{m \xrightarrow{+} 1} J_m(\mathbf{U}, \mathbf{V}) = \max_{k \in \mathcal{I}_N} \min_{i \in \mathcal{I}_c} \{d_{ik}\},$
which corresponds to the hard partitioning of $\mathbf{Z}$, i.e. $\mu_{ik} \in \{0, 1\}$ for $\forall (i, k) \in \mathcal{I}_{cN} = \mathcal{I}_c \times \mathcal{I}_N$. Here, $\mathbf{V}$ corresponds to a 1-width optimal Laguerre base.

b. $J_2(\mathbf{U}, \mathbf{V}) = \max_{k \in \mathcal{I}_N} \left[\sum_{i=1}^c d_{ik} \right]^{-1},$
which is the maximum of the *harmonic means* based distance of each z_k to the resulting clusters.

c. $\lim_{m \to \infty} J_m(\mathbf{U}, \mathbf{V}) = \lim_{m \to \infty} c^{1-m} \max_{k \in \mathcal{I}_N} \left[\prod_{i=1}^c d_{ik} \right]^{1/c}$
Furthermore, $J_m(\mathbf{U}, \mathbf{V})$ decreases monotonically with m, and $J_\infty(\mathbf{U}, \mathbf{V}) = 0$.

Thus for a large m, J_m corresponds to a close approximation of (7), enabling the FKcM to solve directly Problem 2. However, if $m \to \infty$, then $\mu_{ik} \to 1/c$ for $\forall (i, k) \in \mathcal{I}_{cN}$, which can cause numerical problems. Therefore, to obtain a well approximating solution of Problem 2, a value of $m \in (2, \infty)$ should be used. Based on theoretical investigation and experience, $m \in [5, 10]$ yields satisfactory results.

The FKcM-functional (11) is a bounded $(0 \leq J_m \leq c)$ monotonically descending function both in $\{d_{ik}\}$ and $\mathbf{U}$, which allows Algorithm 4 to always converge (in practice). This convergence point can either be a local minimum or a saddlepoint of J_m, fulfilling Theorem 3. For the standard FcM, convergence to a strict local minimum can be shown, but the underlying reasoning does not hold for the FKcM case, as J_m is discontinuous on $\mathcal{M}_{cN} \times \mathbb{D}^c$. The convergence of the FKcM is still subject of research.

Generally, (13-14) is a *bilinear matrix inequalities* (BMIs) constrained minimization problem if γ_i and v_i are both unknowns. However, for a fixed $\gamma = \gamma^*$, (13-14) becomes[1]:

Problem 6: For a given γ^* does exist any $v \in \mathbb{D}$ such that

$$\begin{bmatrix} v \\ 1 \end{bmatrix}^* Q_k \begin{bmatrix} v \\ 1 \end{bmatrix} \leq 0, \ \forall k \in \mathcal{I}_N? \quad (15)$$

Here $Q_k = \left(\mu_{ik}^{2m} \Upsilon_{1k} - (\gamma^*)^2 \Upsilon_{2k} \right)$.

Problem 6 is a well defined LMI feasibility problem (both $\mathbb{D}$ and (15) are convex). Therefore, if (15) is feasible for γ^*, then it characterizes a convex set of solutions $\mathcal{V}(\gamma^*) \subset \mathbb{D}$ fulfilling the constraints of (14). Then, convexity of Problem 6 implies that $\exists \gamma^0 \in [0, 1]$ with $\text{Card} \{\mathcal{V}(\gamma^0)\} = 1$ and $\nexists \gamma^* < \gamma^0$ satisfying Problem 6. Based on this, an upper bound for γ^0 with arbitrary precision can be found by bisection based search [17].

Initialization of Algorithm 4 effects the obtained partition as it defines which local minimum of J_m will be attractive. Therefore, it is advisable to repeat the algorithm multiple times with different initial choices for $\mathbf{V}^{(0)}$ on $\mathbb{D}$ and then select the best by comparison of the achieved decay rate $\rho_m = \max_{z \in \Omega} \prod_{j=1}^c \left| \frac{z - v_j}{1 - z v_j^*} \right|$, $n_b = c$, and visual inspection of the tight fitting of the boundary region $\{z \mid G_b(z^{-1}) = \rho_m\}$ with respect to $\mathbf{Z}$, where $\Xi_c = \mathbf{V}$.

C. Cluster merging

The determination of the number of 'natural' groups in $\mathbf{Z}$, i.e. the best suitable c for clustering, is also important for the successful application of the FKcM method. Similarity based *adaptive cluster merging* (ACM) is frequently used for this purpose [15], but other strategies exist also. ACM is more 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, $\epsilon_s \in [0, 1]$.

V. RESULTS OF APPLICATION

As an example, we consider an asymptotically stable LPV system $\mathcal{S}$, also considered in [8], given in a parameter-affine controller canonical form:

$$x(k+1) = (\mathbf{A}_0 + \mathbf{A}_1 \zeta(k)) x(k) + \mathbf{B}u(k) \quad (16)$$
$$y(k) = \mathbf{C}x(k) + \mathbf{D}u(k) \quad (17)$$

where $\zeta : \mathbb{Z} \to [-1, 1]$, is the scheduling parameter and

$$\mathbf{A}_0 = \begin{bmatrix} \mathbf{0}_{1 \times 4} & \mathbf{A}_{021} \\ \mathbf{I}_{4 \times 4} & \mathbf{A}_{022} \end{bmatrix}; \mathbf{A}_1 = \frac{21}{800} \cdot \begin{bmatrix} \mathbf{0}_{1 \times 4} & \mathbf{A}_{121} \\ \mathbf{I}_{4 \times 4} & \mathbf{A}_{122} \end{bmatrix};$$

---

[1]It can easily be shown that for $\gamma = 1$ a solution always exists

$$\mathbf{B} = \begin{bmatrix} 1 & \mathbf{0}_{1 \times 4} \end{bmatrix}^T ; \mathbf{C} = \begin{bmatrix} \mathbf{I}_{1 \times 5} \end{bmatrix} ; \mathbf{D} = 0,$$

where $\mathbf{A}_{021} = 0.155$, $\mathbf{A}_{121} = 0.1$, $\mathbf{A}_{022} = [-1.155, 3.244, -4.561, 3.302]^T$, and $\mathbf{A}_{122} = [-0.1, 1, -1, 0.1]^T$.

By fixing ζ to set-points $\{\zeta_i\}_{i=1}^{n_p} = \{-1; -1 + h; \ldots; 1\}$, where $h = 0.4$, $n_p = 6$ local LTI representations of $\mathcal{S}$ are obtained, whose pole locations are samples of the ζ-dependent pole movements. In our identification approach, these LTI systems represent the results of local identification.

For the obtained $N = 6 \cdot 5$ pole locations, the FKcM algorithm was applied with different values of m and both with fixed number of clusters $c = 5$ (denoted by *m2c5* for $m = 2$) and also with the application of ACM (denoted by *ad*). In [8], 3 different metrics-based FcM solutions were given which clustered the same pole set into 5 fuzzy sets. This number agrees with the number of sets by visual inspection and the order of the local LTI systems.

The results of the algorithm are presented in Table I and Figure 1. By using the cluster centers as basis poles, $\Xi_{n_b = c} = \mathbf{V}$, the resulting Kolmogorov region $\Omega(\rho, \Xi_c)$ is also given in the figures (bold line) with the corresponding ρ in Table I. In Figure 2, the membership function of a cluster is presented for $m = 2$ and 25. Based on these and the previous results of [8], the following observations can be made:

- The values of N_{av}[1] are relatively low, but they are growing with m. Explanation lays in Theorem 5, by which $J_m \to 0$ as $m \to \infty$. This property introduces both increased computational error and flat shapes of membership surfaces for large m (compare Fig. 2.b to 2.a). Flat surfaces give smaller improvement towards the minimum of J_m in each iteration of Algorithm 4, than the peak-weighting of the $m = 2$ case. Compared to the algorithms of [8], the magnitude of N_{av} is the same, however the bisection based solution of (13) results in a significantly increased computational time.

- The FKcM with ACM ($\varepsilon_s = -15dB$) always converges to a 5-cluster-based partition for low m, but in case of higher values of m, the merging will also have different attractive solutions, like the *m8ad5* and *m8ad8* cases. Here both the 5 and the 8 cluster-based partitions are attractive, depending on the initial condition. However, *m8ad5* achieves a lower H_p[3] than *m8ad8*, suggesting that *m8ad5* corresponds better to the natural data structure. In this way, if the FKcM with ACM converges to partitions with different c, then selection of the one with the lowest H_p yields the most valid partition.

- χ[2] is small in all of the cases, showing that each partition represents well the underlying structure. However, χ is not comparable for different m. The results in [8] used $m = 2$, thus comparison only to *m2c5* is available, giving that *m2c5* performs worse. Explanation lies in the peaking value of the *m2c5* associated membership functions, like in Figure 2.a. As χ has a decreasing tendency with growing c and an increasing tendency for growing m, therefore the fact that $\chi_{25c5} < \chi_{8ad5}$ supports that *m25c5* corresponds better to the underlying data structure in the KnW sense than *m8ad5*.

TABLE I

COMPARISON OF ALGORITHMIC RESULTS IN TERMS OF N_{av}, THE AVERAGE NUMBER OF ITERATIONS[1]; c, THE NUMBER OF OBTAINED CLUSTERS; χ, THE XIE-BENI VALIDITY INDEX[2]; ρ, THE WORST-CASE KOLMOGOROV DISTANCE; AND H_p, THE NORMALIZED ENTROPY[3]

Test case	m2c5	m8ad5	m8ad8	m25c5
N_{av}	28	41	45	67
c	5	5	8	5
χ (dB)	-19.00	-11.75	-13.15	-12.82
ρ (dB)	-10.82	-14.32	-26.52	-15.67
H_p	1.38	1.93	2.79	1.93

- The resulting $\Omega(\rho, \Xi_c)$ Kolmogorov region is relatively tight in all cases except *m2c5*. ρ is also acceptable which means small modelling error if the Ξ_c defined GOBFs are used for identification. In the *m8ad8*-case, ρ is the best, which is the consequence only of the larger ($c = 8$) number of GOBFs. Among the $c = 5$ partitions, *m25c5* is the best in the KnW sense with a very tight bound, which is in agreement with Theory 5. It is also superb to the solutions of [8] where the best achieved ρ was -13.48dB. However, *m2c5* is worse compared to these results which suggests that larger values of m can ensure only the quality of the FKcM obtained solution.

In conclusion, the FKcM solutions for the considered example are converging relatively fast to optimal partitions in terms of Theorem 3. In concord with Theorem 5, as m increases, these partitions are giving better solutions of Problem 2. ACM also ensures proper selection of an efficient number of GOBFs in the KnW 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 each cases. It is important to note that by applying a gradient search method, like the mechanism described in [6, Ch. 11], in order to solve (7), similar results can be obtained, however the computational load of such approaches is much higher and convergence to a global minimum is also not guaranteed. Further, problems may rise at points where (7) is discontinuous as there the gradients may not exist. With the presented FKcM method a good trade of between computational complexity and quality of the solution can be achieved by choosing m wisely. It is true, that for even a relatively small m, the algorithm delivers almost optimal results with a much smaller computational load than the gradient search methods and no problems of local non-differentiability. Furthermore, through the specific choices of m and J_m, the cost function (11) can be shaped such that some of the local minima of (7) can be avoided. Moreover, online selection of the number of needed OBFs can be hardly implemented theoretically into a gradient search method, while in the fuzzy domain, strategies like the ACM work

---

[1]N_{av} is based on the results of 10 runs starting from random $\mathbf{V}^{(0)}$.

[2]For checking the validity of $(\mathbf{U}, \mathbf{V})$ the Xie-Beni validity index χ [18] was applied, which gives a common ground of comparison between different FcM partitions. The smaller χ is, the better the corresponding fit.

[3]In case of naturally separated data (like in the example), for checking that clusters produced by FcM algorithms are also well-separating, the Normalized Entropy [14] can give a handle. The smaller H_p is, the more valid the hypothesis is that the clusters match with the natural data groups.
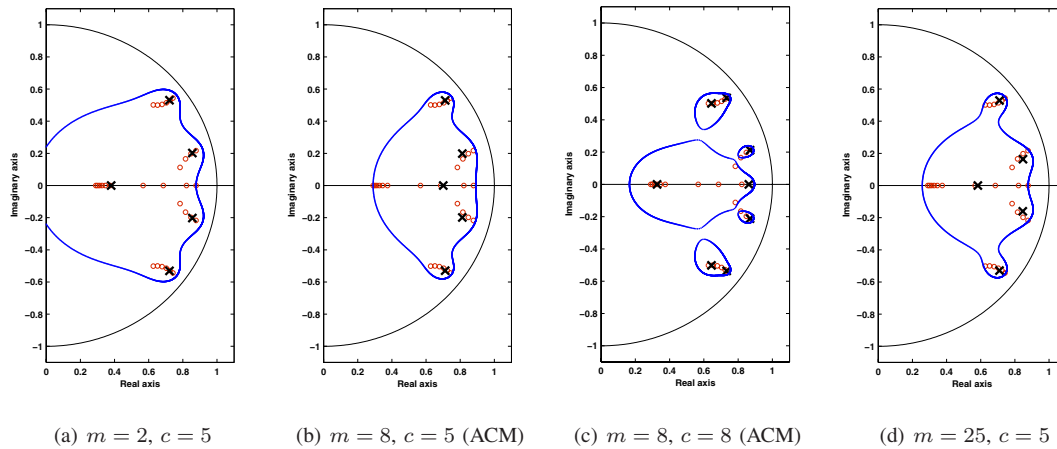
(a) $m = 2$, $c = 5$ (b) $m = 8$, $c = 5$ (ACM)   (c) $m = 8$, $c = 8$ (ACM)   (d) $m = 25$, $c = 5$

Fig. 1.   Results of FKcM clustering: sampled poles (o), resulting cluster centers ($\times$), and Kolmogorov boundaries (bold lines).

perfectly. Besides these, the FKcM also produces additional information, namely it delivers exact definitions of pole clusters. An other important remark is that in case of LTI identification, uncertainty bounds of the estimated model are available, which can be mapped into $\mathbb{D}$ in terms of pole uncertainties. By modification of the FKcM algorithm it is possible to deliver a solution for Problem 2 in a robust sense. Because of the complexity of these modifications, which lead through the utilization of robust LMIs and the hyperbolic geometry of $\mathbb{D}$ for convex uncertainty regions in the KM based topological sense, the exact details are omitted.
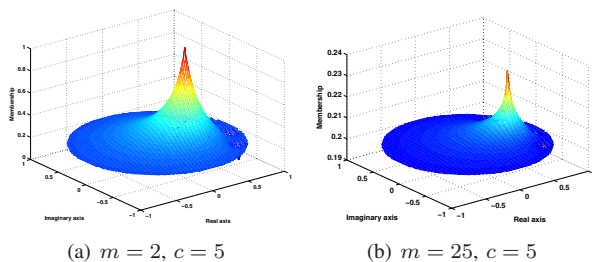

(a) $m = 2$, $c = 5$ (b) $m = 25$, $c = 5$

Fig. 2.   Membership functions of a cluster in FKcM clustering

VI. CONCLUSIONS

The FKcM-based pole clustering, presented in this paper, offers an attractive procedure to determine pole regions and the associated asymptotically KnW optimal GOBFs, based on local observations of an LPV system. The determined bases can be used for fixed-order local identification of the physical process. This contribution enables the direct use of the KnW result, improving the approach of [8] as no other optimization is needed after obtaining the pole clusters. The method, discussed in this paper is the first step of realizing the presented identification approach. A next step in our research will focuss on the interpolation of the local models.

The authors thank Carsten Scherer and Niels Vergeer for their contribution through many fruitful discussions.

REFERENCES

[1] B. L. Stevens and F. L. Lewis, *Aircraft Control and Simulation*. John Wiley and Sons Inc., 2003.

[2] R. Tóth and D. Fodor, "Speed sensorless mixed sensitivity linear parameter variant h_∞ control of induction motor," in *Proceedings of the 43rd IEEE Conf. on Decision and Control*, 2004, pp. 4435–4440.

[3] W. Rugh and J. Shamma, "Research on gain scheduling," *Automatica*, vol. 36, no. 10, pp. 1401–1425, 2000.

[4] R. Murray-Smith and T. A. Johansen, Eds., *Multiple Model Approaches to Modelling and Control*. Taylor and Francis, 1997.

[5] V. Verdult and M. Verhaegen, "Subspace identification of multivariable linear parameter-varying systems," *Automatica*, vol. 38, no. 5, pp. 805–814, 2002.

[6] P. S. C. Heuberger, P. M. J. Van den Hof, and Bo Wahlberg, Eds., *Modeling and Identification with Rational Orthonormal Basis Functions*. Springer-Verlag, 2005.

[7] T. Oliveira e Silva, "A n-width result for the generalized orthonormal basis function model," in *Preprints of the 13th IFAC World Congress*, vol. I, San Francisco, USA, 1996, pp. 375–380.

[8] R. Tóth, P. S. C. Heuberger, and P. M. J. Van den Hof, "Optimal pole selection for LPV system identification wih OBFs, a clustering approach," in *Preprints of the 14th IFAC Symp. on System Ident.*, Newcastle, Australia, 2006, pp. 356–361.

[9] N. Vergeer, "Identification of LPV systems using orhonormal basis functions," Master's thesis, Delft Technical University, 2005.

[10] P. S. C. Heuberger, P. M. J. Van den Hof, and O. H. Bosgra, "A generalized orthonormal basis for linear dynamical systems," *IEEE Tr. Aut. Contr.*, vol. 40, no. 3, pp. 451–465, 1995.

[11] B. M. Ninness and F. Gustafsson, "A unifying construction of orthonormal bases for system identification," *IEEE Tr. Aut. Contr.*, vol. 42, no. 4, pp. 515–521, 1997.

[12] A. Pinkus, *n-Widths in Approximation Theory*. New York: Springer-Verlag, 1985.

[13] P. M. Mäkilä and J. R. Partington, "Robust aproximate modelling of stable linear systems," *Int. J. of Control*, vol. 58, no. 3, pp. 665–683, 1993.

[14] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York and London: Plenum Press, 1981.

[15] U. Kaymak and M. Setnes, "Fuzzy clustering with volume prototypes and adaptive cluster merging," *IEEE Tr. Fuzzy Sys.*, vol. 10, no. 6, pp. 705–711, December 2002.

[16] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. New Jersey: Englewood Cliffs, Prentice Hall, 1988.

[17] S. P. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia: SIAM, 1994.

[18] X. L. X. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Tr. Patt. Anal. Machine Intell.*, vol. 13, no. 8, pp. 841–847, August 1991.