On worst-case approximation of feasible system sets via orthonormal basis functions

M. Casini, A. Garulli, and A. Vicino

***Abstract*—This paper deals with the approximation of sets of linear time-invariant systems via orthonormal basis functions. This problem is relevant to conditional set membership identification, where a set of feasible systems is available from observed data, and a reduced-complexity model must be estimated. The basis of the model class is made of impulse responses of linear filters. The objective of the paper is to select the basis function poles according to a worst-case optimality criterion. Suboptimal conditional identification algorithms are introduced and tight bounds are provided on the associated identification errors.**

I. Introduction

A typical objective of set membership identification is that of estimating models which are suitable for robust control purposes (see [1] and related references). In this respect, a fundamental requirement is that model uncertainties need to be quantified in the estimated model. The set membership approach considers the set of feasible systems (those compatible with available input/output data and a priori information) and picks a nominal model in a given model class, by minimizing the worst-case error with respect to the feasible set. Hence, worst-case approximations of a set of systems must be tackled.

A well-known problem of several approaches to identification for control (see papers in [2] and the survey [3]) is the high complexity of the estimated models. On the other hand, robust control design techniques require nominal models of reduced complexity (low dimensional, linearly parameterized), along with a measure of the associated nonparametric perturbation. For this reason, it is common to select as nominal model class the linear combination of impulse responses of suitable linear filters. Typical examples, widely used in system identification, are Laguerre functions [4], Kautz functions [5], generalized orthonormal basis functions [6], [7], [8]. This choice is motivated by the fact that these expansions provide linearly parameterized model classes, which assure good approximations of linear time-invariant (LTI) systems by means of few parameters. Estimation of the model that minimizes the worst-case identification error, within the chosen reduced-complexity model class, is usually addressed as *conditional* set membership identification, and has been deeply investigated in recent years (see e.g. [9], [10]). Special attention has been devoted also to the optimal choice of pole locations for the basis functions, in order to find the optimal approximating model for a given LTI system [11].

This paper studies the worst-case approximation of an entire set of feasible systems, via orthonormal basis functions. The problem is formulated in the conditional set membership setting, and optimal pole selection for the chosen basis expansion is addressed. It is shown that the computation of optimal poles with respect to a set of systems, requires the solution of complicated min-max optimization problems. Suboptimal identification algorithms are then introduced, in which pole selection is performed by optimizing with respect to a single element related to the feasible set (namely, its Chebyshev center or a generic element of the set). Tight bounds on the worst-case error of these suboptimal algorithms are provided. Finally, illustrative examples for the case of set approximation via Laguerre basis functions are presented.

Authors are with Dipartimento di Ingegneria dell'Informazione, Università di Siena, 53100 Siena, Italy (email: casini@ing.unisi.it; garulli@ing.unisi.it; vicino@ing.unisi.it).

II. Conditional set membership identification

In set membership identification, the uncertainty associated to an identified model is usually measured according to the worst-case error with respect to a set of admissible systems. This set, called *feasible system set*, accounts for two main information sources: i) a priori knowledge on the plant; ii) input/output measurements. The former may be represented by a set to which the true system is known a priori to belong; the latter is typically a finite record of noisy data, and assumptions are made on the nature of the noise.

In this paper, LTI discrete-time SISO systems are considered. The impulse response sequence of a system is denoted by $h = \{h_i\}_{i=0}^{\infty}$, and belongs to a linear normed space $\mathcal{H}$, equipped with the norm $\|\cdot\|_{\mathcal{H}}$. A priori knowledge on the system is expressed as $h \in \mathcal{S}$, where $\mathcal{S}$ is a set contained in $\mathcal{H}$. Data consists of N input/output pairs $Z = \{(u_k, y_k),\ k = 0,\dots,N-1\}$, related by

$$y = T(u)h^N + v \tag{1}$$

where $y = [y_0 \dots y_{N-1}]'$ is the output vector, $T(u)$ is the lower triangular Toeplitz matrix of inputs $u_0,\dots,u_{N-1}$, $h^N = [h_0 \dots h_{N-1}]'$ is the truncated impulse response, and $v = [v_0 \dots v_{N-1}]'$ is the vector of disturbances affecting the measurements. It is assumed that v is unknown-but-bounded, i.e.

$$\|v\|_Y \le \varepsilon \tag{2}$$

where $\|\cdot\|_Y$ denotes a suitable norm in $\mathbb{R}^N$ and ε is a known positive scalar. According to (1)-(2), the feasible system set is defined as

$$\mathcal{F} = \{h \in \mathcal{S} :\ \|y - T(u)h^N\|_Y \le \varepsilon\}. \tag{3}$$

Following the terminology of the IBC theory (see [12] for a thorough treatment), an identification algorithm ϕ is a mapping from i/o data to the space of models. When the latter coincides with the space of systems $\mathcal{H}$, it results $\phi : \mathbb{R}^N \times \mathbb{R}^N \to \mathcal{H}$. The *worst-case identification error* associated to the identification algorithm turns out to be

$$E[\phi] = \sup_{h\in\mathcal{F}} \|h - \phi(Z)\|_{\mathcal{H}}. \tag{4}$$

In identification for robust control, it is customary to identify reduced-complexity models, that are suitable for robust control design techniques. A typical choice is to select a linearly parameterized model class such as

$$\mathcal{M} = \{h :\ h = M_p\theta,\ \theta \in \mathbb{R}^n\} \tag{5}$$

where M_p is a linear operator, $M_p :\ \mathbb{R}^n \to \mathcal{H}$ and θ is the n-dimensional parameter vector to be identified, $n < N$.

Hence, a conditional identification algorithm is a mapping $\phi : \mathbb{R}^N \times \mathbb{R}^N \to \mathcal{M}$. In system identification theory, it is common to choose the basis of $\mathcal{M}$ as a collection of impulse responses of linear filters, such as Laguerre or Kautz functions. The notation M_p emphasizes the dependence of the model class on some variables $p \in \mathbb{R}^m$ that characterize the chosen basis. For example, for Laguerre filters p is the real Laguerre pole ($m = 1$), while for Kautz functions p denotes the pair of complex conjugate poles ($m = 2$).

The identification of a model within the class $\mathcal{M}$ has been addressed in the literature as *conditional* set membership identification [9], [10], [13]. In particular, the optimal model is given by the so-called *conditional central algorithm*, which minimizes the worst-case error among the elements of $\mathcal{M}$, i.e. $\phi_{cc}(Z) = M_p\theta_{cc}$ where

$$\theta_{cc} \;=\; \arg \inf_{\theta\in\mathbb{R}^n} \sup_{h\in\mathcal{F}} \|h - M_p\theta\|_{\mathcal{H}}. \tag{6}$$

A procedure for computing $\phi_{cc}(Z)$ has been given in [10] for the case $\|\cdot\|_{\mathcal{H}} = \|\cdot\|_Y = \|\cdot\|_2$, while suboptimal identification algorithms have been considered in [13]. In all these results, the model class $\mathcal{M}$ is fixed, i.e. the poles p were selected a priori on the basis of some knowledge of the true system. In the next section, optimal choice of the basis poles, based on the available data, will be addressed.

III. Optimal choice of the model class

The approximation of LTI stable systems, by means of orthogonal basis expansions, has been studied since long time, and has witnessed a renewed interest in the last decade, due to significant applications not only in system identification, but also in signal processing and other fields. Several results have been provided concerning the approximation of a given LTI system, via different basis expansions (see [8] and references therein). The optimal choice of pole locations for the basis functions, in order to find the optimal approximating model for a fixed number of terms in the expansion, is also a widely addressed problem (see [11] for a survey on this topic). For a given function $h \in \mathcal{H}$ and a model class $\mathcal{M}$, the optimal pole is provided by

$$p^*(h) \;=\; \arg \inf_{p\in\mathcal{P}} \inf_{\theta\in\mathbb{R}^n} \|h - M_p\theta\|_{\mathcal{H}} \tag{7}$$

where $\mathcal{P}$ is the set of admissible pole locations, which usually rely on a priori knowledge on the true system.

In the context of conditional set membership identification, the selection of the optimal pole must be performed with respect to all the elements in the feasible set, i.e. via the minimization of the worst-case error. This leads to the following optimization problem

$$p^*(\mathcal{F}) \;=\; \arg \inf_{p\in\mathcal{P}} \inf_{\theta\in\mathbb{R}^n} \sup_{h\in\mathcal{F}} \|h - M_p\theta\|_{\mathcal{H}}. \tag{8}$$

This is, in general, a very complicated min-max optimization problem for which the derivation of simple conditions appears to be an awkward task. For this reason, suboptimal solutions will be pursued in this paper, based on the computation of the optimal pole $p^*(h)$ as in (7), for some suitable h related to $\mathcal{F}$.

For a given system h, let us introduce the optimal (with respect to p) projection operator Π_n, such that

$$\Pi_n h = M_{p^*(h)}\theta^*(h) \tag{9}$$

where $\theta^*(h)$ is the parameter vector corresponding to the projection of h onto the linear manifold $\mathcal{M}$ with optimal poles $p^*(h)$, i.e.

$$\theta^*(h) = \arg \inf_{\theta\in\mathbb{R}^n} \|h - M_{p^*(h)}\theta\|_{\mathcal{H}} \tag{10}$$

(notice that both $p^*(h)$ and $\theta^*(h)$ may not be unique; in these cases, $\Pi_n h$ is equal to one of the possible choices of $M_{p^*(h)}\theta^*(h)$). Moreover, let us adopt the notation $\phi(Z;p)$ to stress the dependence of the identification algorithm on the choice of poles p. Then, the algorithm providing the minimum worst-case identification error is $h_{cc} = \phi_{cc}(Z;p^*(\mathcal{F}))$. In the following, the optimal error $E[\phi_{cc}(Z;p^*(\mathcal{F}))]$ will be compared to that of two classes of suboptimal conditional identification algorithms:

- *central projection algorithm*

$$\phi_{cp}(Z;p^*(h_c)) = h_{cp} \triangleq \Pi_n h_c \tag{11}$$

  where $h_c = \arg \inf_{h\in\mathcal{H}} \sup_{\tilde{h}\in\mathcal{F}} \|h - \tilde{h}\|_{\mathcal{H}}$ is the Chebyshev center of $\mathcal{F}$ in the norm $\mathcal{H}$;
- *interpolatory projection algorithm*

$$\phi_{ip}(Z;p^*(h_i)) = h_{ip} \triangleq \Pi_n h_i \tag{12}$$

  where $h_i \in \mathcal{F}$.

These algorithms have been introduced in conditional set membership identification (see e.g. [14], [9], [13]), because in several problems it is much easier to compute the center h_c or a generic element h_i of $\mathcal{F}$, and then project it onto the model class $\mathcal{M}$, rather than solving problem (6). For example, h_c is easily obtained every time the set $\mathcal{F}$ has a symmetry center (e.g., if it is an ellipsoid), while algorithms are available for computing h_c when $\mathcal{F}$ is a generic polytope [15]. In this case, a $h_i \in \mathcal{F}$ can be obtained via linear programming.

When also the model class $\mathcal{M}$ must be selected via optimization with respect to poles p, as in the computation of (11) or (12), one can first compute h_c or h_i, then exploit conditions to obtain $p^*(h_c)$ or $p^*(h_i)$, and hence h_{cp} or h_{ip} via the projection operator (9). In the next section, results will be provided on the suboptimality degree of algorithms (11)-(12), with respect to the minimum identification error, provided by $\phi_{cc}(Z;p^*(\mathcal{F}))$.

IV. Suboptimal pole choice and error bounds

The aim of this section is to derive tight bounds on the identification error provided by the projection algorithms (11) and (12), i.e. to determine the minimum $\kappa \geq 1$, such that

$$E[\phi_{cp}(Z;p^*(h_c))] \leq \kappa \cdot E[\phi_{cc}(Z;p^*(\mathcal{F}))]$$

for all possible y and $\mathcal{F}$ (and similarly for an interpolatory projection algorithm ϕ_{ip}).

Before proceeding, it is useful to recall that in the conditional set membership identification setting of Sect. II, several problems of interest can be restricted to the finite dimensional space $\mathbb{R}^N$. In fact, let T_N denote the truncation operator in $\mathcal{H}$, such that $T_N h = h^N$, and R_N be the remainder operator $R_N h = \{h\}_{i=N}^{\infty}$. Then, under the mild assumption that $R_N\mathcal{S}$ is a balanced set (i.e., if $h \in R_N\mathcal{S}$, then also $-h \in R_N\mathcal{S}$), it can be shown that for any ℓ_p norm, $1 \leq p < \infty$, one has $E[T_N\phi] \leq E[\phi]$ for any conditional algorithm ϕ, model class $\mathcal{M}$ and feasible set $\mathcal{F}$. Therefore, one can consider only truncated basis expansions such as

$$\mathcal{M} = \{h^N \in \mathbb{R}^N :\ h^N = M_p\theta,\ \theta \in \mathbb{R}^n,\ n < N\}. \quad (13)$$

Consequently, an identification algorithm turns out to be a mapping from $\mathbb{R}^N \times \mathbb{R}^N$ (the i/o data space) to an n-dimensional subspace of $\mathbb{R}^N$ (the truncated model class). Moreover, one has

$$E[\phi,\mathcal{M}] = \left(\sup_{h^N \in T_N(\mathcal{F})} \|h^N - \phi(Z)\|_p^p + \sup_{h\in\mathcal{S}} \|R_N h\|_p^p \right)^{1/p} \quad (14)$$

(in the following, dependence on $\mathcal{M}$ will be omitted to simplify notation). The rightmost term in (14) depends only on $\mathcal{S}$ and can be computed a priori. Hence, in the following, the finite-dimensional feasible set $\mathcal{F}_N = T_N\mathcal{F} \subset \mathbb{R}^N$ will be considered, when computing the estimates $\phi_{cc}(Z;p^*(\mathcal{F}_N))$, $\phi_{cp}(Z;p^*(h_c))$, $\phi_{ip}(Z;p^*(h_i))$ introduced in Sect. III.
In this paper, the ℓ_2 identification error is considered, i.e. $\|\cdot\|_{\mathcal{H}} = \|\cdot\|_2$ (to simplify notation, the ℓ_2 norm will be denoted by $\|\cdot\|$). Let us also denote the Chebyshev radius of $\mathcal{F}_N$ by $\text{rad}(\mathcal{F}_N) = \sup_{h\in\mathcal{F}_N} \|h - h_c\|$, where h_c is the Chebyshev center of $\mathcal{F}_N$ in the ℓ_2 norm. Now, the main result of the paper can be stated.

Theorem 1. *Let $r = rad(\mathcal{F}_N)$ and $d = \|h_c - h_{cp}\|$. Then,*

$$E[\phi_{cp}(Z;p^*(h_c))] \leq \sqrt{2 - \frac{(r-d)^2}{r^2+d^2}} \cdot E[\phi_{cc}(Z;p^*(\mathcal{F}_N))]. \quad (15)$$

Proof. See Appendix A. □

It is worth remarking that Theorem 1 holds for any N (dimension of the feasible set depending on the data set) and for any n (model order), $n < N$. The maximum bound with respect to all possible feasible sets (i.e. with respect to all $r \geq 0$, $d \geq 0$) is given by $E[\phi_{cp}(Z;p^*(h_c))] \leq \sqrt{2} \cdot E[\phi_{cc}(Z;p^*(\mathcal{F}_N))]$, which occurs when $r = d$. The next theorem shows that the bound on the identification error is *tight*, i.e. it is possible to find a feasible set $\mathcal{F}_N$ and a family of model classes $\mathcal{M}(p)$, depending on p, such that in (15) equality holds.

Theorem 2. *Let $\mathcal{M}(p)$ be a family of orthonormal bases as in (13), with $M_p \in \mathbb{R}^{N\times n}$. Assume that for some $\bar{h} \in \mathbb{R}^N$, the infimum in (7) is achieved for two distinct pole vectors p_1^*, p_2^*, and let θ_1^*, θ_2^* be the corresponding parameter vectors in (10), so that*

$$\|\bar{h} - M_{p_1^*}\theta_1^*\| = \|\bar{h} - M_{p_2^*}\theta_2^*\| \leq \|\bar{h} - M_p\tilde{\theta}\|\ , \quad \forall p\ ,\ \ \forall\tilde{\theta} \in \mathbb{R}^n. \quad (16)$$

Moreover, let $(\bar{h} - M_{p_1^*}\theta_1^*)'(\bar{h} - M_{p_2^*}\theta_2^*) = 0$. Then, there exists a feasible set $\mathcal{F}_N \subset \mathbb{R}^N$ such that

$$E[\phi_{cp}(Z;p^*(h_c))] = \sqrt{2} \cdot E[\phi_{cc}(Z;p^*(\mathcal{F}_N))].$$

Proof. See Appendix B. □

A bound on the identification error can be provided also for interpolatory projection algorithms. Unfortunately, this turns out to be much larger than that given by Theorem 1.

Theorem 3. *Let $r = rad(\mathcal{F}_N)$ and $d_i = \|h_i - h_{ip}\|$. Then,*

$$E[\phi_{ip}(Z;p^*(h_i))] \leq \min\left\{2 + \frac{d_i}{r}, 1 + \frac{2r}{d_i}\right\} \cdot E[\phi_{cc}(Z;p^*(\mathcal{F}_N))]. \quad (17)$$

Proof. See Appendix C. □

In order to obtain the maximum value of the bound in a worst-case setting, one has to consider the worst interpolatory estimator, i.e. the projection of the worst $h_i \in \mathcal{F}_N$, and all possible feasible sets. This corresponds to maximizing (17) with respect to all $r \geq 0$, $d_i \geq 0$, thus obtaining the bound $E[\phi_{ip}(Z;p^*(h_i))] \leq 3 \cdot E[\phi_{cc}(Z;p^*(\mathcal{F}_N))]$, which follows immediately from Theorem 3 when $r = d_i$. Also this bound turns out to be tight, as one can find a feasible set $\mathcal{F}_N$, an element $h_i \in \mathcal{F}_N$ and a family of model classes $\mathcal{M}(p)$, depending on p, such that the error of the interpolatory projection algorithm is arbitrarily close to three times the error of the optimal algorithm ϕ_{cc}.

It is interesting to compare the above results to those obtained in [13] for the case of a fixed model class $\mathcal{M}$, with poles p assigned a priori. It has been shown that $E[\phi_{cp}] \leq \sqrt{4/3}\, E[\phi_{cc}]$ and $E[\phi_{ip}] \leq 2\, E[\phi_{cc}]$; moreover, there exist feasible sets $\mathcal{F}_N$ and model classes $\mathcal{M}$ for which the equality holds. Obviously, in the context of the present paper, the ratio between the worst-case identification error of the projection algorithms and the minimum achievable error is larger, due to the fact that suboptimal algorithms select the poles p by optimizing over a single element related to $\mathcal{F}_N$ (namely h_c or h_i), while the minimum error is achieved by choosing $p^*(\mathcal{F}_N)$ as in (8), where the whole feasible set is considered. Nevertheless, the bounds provided by the results in this section are useful, as they clarify that the maximum possible gap between a “set-oriented” choice of the poles and a choice based only on a single element is not very large. This is especially true for the central projection algorithm, which is in turn much easier to compute than $\phi_{cc}(Z;p^*(\mathcal{F}_N))$, in particular when the feasible set admits a symmetry center.

A. Error bounds for ellipsoidal feasible sets

In some identification problems, the feasible set has a special structure that can be exploited in the computation of bounds on the identification error. If the noise in (2) is bounded in the ℓ_2 norm, and the a priori set $\mathcal{S}$ provides constraints only on the tail $R_N h$ of the impulse response (the so called *residual* a priori information, often adopted in the literature [9]), the feasible system set $\mathcal{F}_N$ is an N-dimensional ellipsoid. Theorems 1 and 2 guarantee that the error provided by the central projection algorithm is not greater than $\sqrt{2}$ times the minimum error. However, for ellipsoidal feasible sets this bound can be further reduced, exploiting the special structure of the set.

Theorem 4. *Let $\mathcal{F}_N \subset \mathbb{R}^N$ be an ellipsoid of center h_c, and let L_M and L_m be the lengths of its maximum and minimum semi-axis, respectively. Moreover, define $d = \|h_c - h_{cp}\|$. Then,*

$$\frac{E[\phi_{cp}(Z; p^*(h_c))]}{E[\phi_{cc}(Z; p^*(\mathcal{F}_N))]} \leq$$
$$\leq \begin{cases} \dfrac{L_M + d}{L_M \sqrt{1 + \dfrac{d^2}{L_M^2 - L_m^2}}} & \text{if } d < \dfrac{L_M^2 - L_m^2}{L_m} \text{ or } L_m = 0, \\ \dfrac{L_M + d}{L_m + d} & \text{if } d \geq \dfrac{L_M^2 - L_m^2}{L_m} \text{ and } L_m > 0. \end{cases} \tag{18}$$

Proof. See Appendix D. □

Observe that, for the ellipsoidal feasible set above, $L_M = rad(\mathcal{F}_N)$. Hence, when $L_m = 0$ one obtains the same bound as in Theorem 1. Conversely, when L_m tends to L_M, the ratio $E[\phi_{cp}]/E[\phi_{cc}]$ tends to 1, as expected, because for a spherical feasible set the conditional center with respect to any $\mathcal{M}$ coincides with the projection of the center of the sphere onto $\mathcal{M}$.

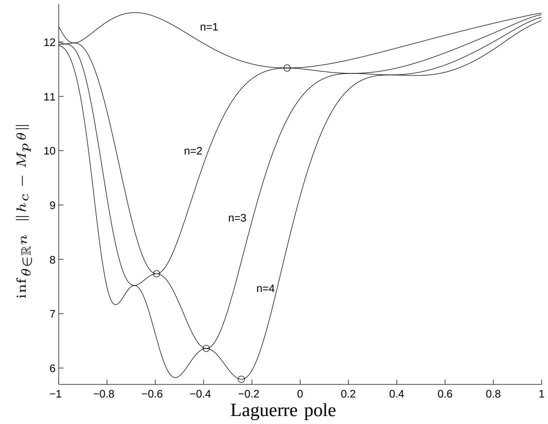
V. EXAMPLES

In the first example reported below, the suboptimal pole choice performed by the central projection algorithm is compared to the optimal one (8), in the case of Laguerre basis functions. The optimal model $h_{cc} = \phi_{cc}(Z; p^*(\mathcal{F}_N))$ is obtained by applying the procedure presented in [10] for computing the conditional central estimate of an ellipsoidal feasible set, for each model class $\mathcal{M}$ with fixed Laguerre pole p, and then minimizing with respect to p via a one-dimensional gridding on the interval $(-1, 1)$.

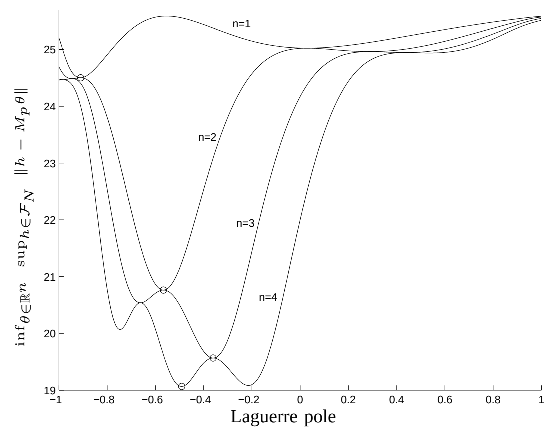
Example 1. Consider the transfer function

$$H(z) = \frac{5 + 10.7\, z^{-1} + 5.002\, z^{-2}}{1 + 2.3\, z^{-1} + 2.06\, z^{-2} + 0.72\, z^{-3}}$$

and let $N = 50$ i/o data be available, with the input u_k being a unitary step. Assume that the truncated feasible system set is given by $\mathcal{F}_N = \{h^N : \ \|y - T(u)h^N\| \leq \sqrt{N}\,\varepsilon\}$ with noise $\{v_k\}_{k=0}^{N-1}$ being a Gaussian random sequence, satisfying $(1/\sqrt{N})\|v\| \leq \varepsilon$, and $\varepsilon = 1$. Let $h_c = T^{-1}(u)\,y$ and $\mathcal{M}$ be the model class given by Laguerre filters. In Figure 1a, the error $\inf_{\theta \in \mathbb{R}^n} \|h_c - M_p\theta\|$ is plotted as a function of the Laguerre pole, for different model orders $n = 1, 2, 3, 4$. This error clearly does not depend on the feasible set, but just on the approximation of its center h_c. The global minimum of each curve corresponds to the pole $p^*(h_c)$ for each model order n, which is the one picked by the central projection algorithm ϕ_{cp}. These values are reported in Table I, together with the associated worst-case identification errors $E[h_{cp}]$; the latter are obviously larger than the errors at the minima in Fig. 1a, which are computed only with respect to the center h_c and not to the whole set $\mathcal{F}_N$.



(a)



(b)

Fig. 1. (a) Approximation error for the center h_c versus Laguerre pole, for different model orders; (b) Worst-case error for the conditional central algorithm versus Laguerre pole, for different model orders.

Conversely, Figure 1b shows the worst-case error (with respect to $\mathcal{F}_N$) of the conditional Chebyshev center versus the Laguerre pole. The values of the optimal poles and the corresponding error $E[h_{cc}]$ are also reported in Table I, together with the ratio between the errors of suboptimal and optimal algorithm and the upper bound provided by Theorem 4.

From the above example, it can be observed that the suboptimal pole $p^*(h_c)$ selected by the central projection algorithm can be quite far from the optimal one $p^*(\mathcal{F}_N)$, due to the presence of local minima whose cost is close to the global one (in Example 1, this happens for $n = 1$ and $n = 4$). Nevertheless, the worst-case identification error of the suboptimal algorithm turns out to be pretty close to the minimum achievable error $E[\phi_{cc}]$, as shown by Table I.

The next example shows that there exist conditional identification problems, with model class $\mathcal{M}$ given by Laguerre expansions, for which the conditions of Theorem 2 are satisfied, and hence the upper bound is actually achieved.

Example 2. Let $\mathcal{M}$ be the model class of discrete Laguerre

TABLE I
SELECTED POLES, CORRESPONDING WORST-CASE IDENTIFICATION ERRORS, ACTUAL ERROR RATIO AND UPPER BOUND PROVIDED BY THEOREM 4, FOR DIFFERENT MODEL ORDERS.

n	$p^*(h_c)$	$p^*(\mathcal{F}_N)$	$E[h_{cp}]$	$E[h_{cc}]$	$E[h_{cp}]/E[h_{cc}]$	Bound Thm. 4
1	-0.054	-0.910	25.0408	24.5024	1.0220	1.4069
2	-0.595	-0.567	20.8556	20.7634	1.0044	1.3573
3	-0.389	-0.361	19.6296	19.5662	1.0032	1.3223
4	-0.244	-0.491	19.1362	19.0656	1.0037	1.3045

filters, which are defined as

$$L_j(z;p) = \mathcal{Z}\{l_j(p)\} = \frac{\sqrt{1-p^2}}{1-p\,z^{-1}} \left(\frac{z^{-1}-p}{1-p\,z^{-1}}\right)^{j-1}$$

where $j = 1, 2, \ldots$, and $p \in \mathbb{R}$, $|p| < 1$ is the Laguerre pole. Let $h = l_1(-0.7746) + l_1(0.7746)$, and let us set $N = 100$ and $n = 2$. For $\bar{h} = T_N h$ there are two distinct optimal Laguerre poles, given by $p_1^* = -0.5773$ and $p_2^* = 0.5773$. The corresponding optimal parameter vectors (10) can be computed as $\theta_1^* = M^{'}_{p_1^*}\bar{h}$ and $\theta_2^* = M^{'}_{p_2^*}\bar{h}$, where $M_{p_i^*} \in \mathbb{R}^{N\times 2}$ is the truncated basis matrix of $\mathcal{M}$, with pole p_i^*. Let us choose $h_{cp} = M_{p_1^*}\theta_1^*$. Then, it can be verified that $\forall p$ and $\forall \tilde{\theta} \in \mathbb{R}^n$

$$\|\bar{h} - h_{cp}\| = \|\bar{h} - M_{p_2^*}\theta_2^*\| = 0.9129 < \|\bar{h} - M_p\tilde{\theta}\|$$

and $(\bar{h}-h_{cp})'\,(\bar{h}-M_{p_2^*}\theta_2^*) = 0$. Therefore, due to Theorem 2, there exists a feasible set for which $E[h_{cp}] = \sqrt{2}\cdot E[h_{cc}]$. Indeed, let $\mathcal{F}_N = \{x \in \mathbb{R}^N : \|x - \bar{h}\| \le 0.9129,\, h'(M_{p_2^*}\theta_2^* - \bar{h}) = 0\}$. By applying the procedure in [10] for the computation of the conditional Chebyshev center and minimizing with respect to p, one gets $h_{cc} = \phi_{cc}(Z; p^*(\mathcal{F}_N)) = M_{p_2^*}\theta_2^*$. Moreover, one has $E[h_{cp}] = 1.8258$ and $E[h_{cc}] = 1.291$.

VI. CONCLUSIONS

In this paper the problem of selection of an orthonormal function basis for the approximation of a set of feasible systems has been addressed in a worst case identification setting. Impulse responses of increasing order linear filters are considered as candidate basis functions and suboptimal conditional estimators are derived, using the information provided by specific optimal or suboptimal point estimators of the feasible system set. Tight bounds on the worst case errors of these suboptimal algorithms with respect to the optimal estimation error attainable through the optimal function basis are derived.

REFERENCES

[1] A. Garulli, A. Tesi, and A. Vicino (Eds.). *Robustness in Identification and Control*. Lecture Notes in Control and Information Sciences. Springer, London, 1999.

[2] R. L. Kosut, G. C. Goodwin, and M. P. Polis (Eds.). Special issue on system identification for robust control design. *IEEE Transactions on Automatic Control*, 37:899–1008, 1992.

[3] P. M. Mäkilä, J. R. Partington, and T. K. Gustafsson. Worst-case control-relevant identification. *Automatica*, 31(12):1799–1819, 1995.

[4] B. Wahlberg. System identification using Laguerre models. *IEEE Transactions on Automatic Control*, 36:551–562, 1991.

[5] B. Wahlberg. System identification using Kautz models. *IEEE Transactions on Automatic Control*, 39:1276–1282, 1994.

[6] P. M. J. Van den Hof, P. S. C. Heuberger, and J. Bokor. System identification with generalized orthonormal basis functions. *Automatica*, 31(12):1821–1834, 1995.

[7] B. M. Ninness, H. Hjalmarsson, and F. Gustafsson. The fundamental role of generalized orthonormal bases in system identification. *IEEE Trans. on Aut. Control*, 44:1384–1406, 1999.

[8] P. Van den Hof, B. Whalberg, P. Heuberger, B. Ninness, J. Bokor, and T. Oliveira e Silva. Modelling and identification with rational orthogonal basis functions. In *Proc. of IFAC SYSID 2000*, Santa Barbara, CA, 2000.

[9] L. Giarrè, B. Z. Kacewicz, and M. Milanese. Model quality evaluation in set membership identification. *Automatica*, 33(6):1133–1139, 1997.

[10] A. Garulli, A. Vicino, and G. Zappa. Conditional central algorithms for worst-case set membership identification and filtering. *IEEE Trans. on Aut. Control*, 45(1):14–23, 2000.

[11] T. Oliveira e Silva. Optimal pole conditions for Laguerre and two-parameter Kautz models: a survey of known results. In *Proc. of IFAC SYSID 2000*, Santa Barbara, CA, 2000.

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

[13] A. Garulli, B. Z. Kacewicz, A. Vicino, and G. Zappa. Error bounds for conditional algorithms in restricted complexity set membership identification. *IEEE Transactions on Automatic Control*, 45(1):160–164, 2000.

[14] B. Z. Kacewicz, M. Milanese, and A. Vicino. Conditionally optimal algorithms and estimation of reduced order models. *Journal of Complexity*, 4:73–85, 1988.

[15] N. D. Botkin and V. L. Turova-Botkina. An algorithm for finding the Chebyshev center of a convex polyhedron. *Applied Mathematics and Optimization*, 29:211–222, 1994.

[16] A. Garulli, B. Z. Kacewicz, A. Vicino, and G. Zappa. Reliability of projection algorithms in conditional estimation. *J. of Optimization Theory and Applications*, 101(1):1–14, 1999.

APPENDIX

A. Proof of Theorem 1

In order to prove Theorem 1, the following lemma is needed (for a proof, see [16]).

Lemma 1. *Let h_c be the Chebyshev center of $\mathcal{F}_N$ in the ℓ_2 norm and consider the closed halfspace $\mathcal{Q} = \{h \in \mathbb{R}^N :\ a_q^T h \ge b_q,\ a_q \in \mathbb{R}^N, b_q \in \mathbb{R}\}$, such that $a_q^T h_c = b_q$. Then, there exists $h_e \in \mathcal{F}_N \bigcap \mathcal{Q}$ such that $\|h_e - h_c\|_2 = rad(\mathcal{F}_N)$.*

Now, let us first consider $E[h_{cp}] = E[\phi_{cp}(Z; p^*(h_c))]$. One has

$$E[h_{cp}] = \sup_{h\in\mathcal{F}_N} \|h_{cp}-h\| \le \|h_{cp}-h_c\| + \sup_{h\in\mathcal{F}_N} \|h_c-h\| = d+r \tag{19}$$

with d and r defined as in the statement of Theorem 1.

Then, let us analyze the minimum error $E[h_{cc}] = E[\phi_{cc}(Z; p^*(\mathcal{F}_N))]$. Consider the halfspace $\mathcal{Q}_{cc} = \{h \in \mathbb{R}^N :\ (h-h_c)'(h_c-h_{cc}) \ge 0\}$. Lemma 1 guarantees that there exists $h_e \in \mathcal{F}_N \bigcap \mathcal{Q}_{cc}$ such that $\|h_e - h_c\| = r$. Therefore, one has

$$\begin{aligned} E[h_{cc}]^2 &\geq \|h_e - h_{cc}\|^2 \\ &= \|h_e - h_c\|^2 + \|h_c - h_{cc}\|^2 \\ &\quad +2\frac{(h_e-h_c)'(h_c-h_{cc})}{\|h_e-h_c\|\|h_c-h_{cc}\|} \\ &\geq \|h_e - h_c\|^2 + d^2 \ = \ r^2 + d^2 \end{aligned} \quad (20)$$

where it has been exploited the fact that $\|h_c - h_{cc}\| \geq d$, which follows from the definitions of h_{cp} and d. Then, from (19) and (20) one has $\frac{E[h_{cp}]}{E[h_{cc}]} \leq \frac{r+d}{\sqrt{r^2+d^2}} = \sqrt{2 - \frac{(r-d)^2}{r^2+d^2}}$ which proves the theorem.

B. Proof of Theorem 2

Let $\mathcal{F}_N = \{h \in \mathbb{R}^N : \, h = \bar{h} + \alpha\,(\bar{h} - h_{cp}), \ |\alpha| \leq 1\}$, where $h_{cp} = M_{p_1^*}\theta_1^*$ has been chosen as the central projection (this choice is correct, as $\bar{h} = \text{cen}(\mathcal{F}_N)$ and (16) holds). Moreover, let $d = \|\bar{h} - h_{cp}\|$. It is easy to show that

$$\begin{aligned} E[h_{cp}] &= \sup_{h\in\mathcal{F}_N} \|h_{cp} - h\| \\ &= \|h_{cp} - (\bar{h} + \bar{h} - h_{cp})\| \\ &= \|2\,(h_{cp} - \bar{h})\| = 2\,d. \end{aligned} \quad (21)$$

Now, let $\hat{h} = M_{p_2^*}\theta_2^*$. Then, from the assumptions of the theorem, one has $\|\bar{h} - \hat{h}\| = d$ and $(\bar{h} - \hat{h})'(\bar{h} - h_{cp}) = 0$. Therefore, it follows that

$$\begin{aligned} E[\hat{h}]^2 &= \sup_{h\in\mathcal{F}_N} \|\hat{h} - h\|^2 \\ &= \|(\hat{h} - \bar{h}) - (h_{cp} - \bar{h})\|^2 \\ &= \|\hat{h} - \bar{h}\|^2 + \|h_{cp} - \bar{h}\|^2 = 2\,d^2. \end{aligned} \quad (22)$$

Hence, from (21) and (22) one has $\frac{E[h_{cp}]}{E[\hat{h}]} = \frac{2\,d}{\sqrt{2}\,d} = \sqrt{2}$ which means that $h_{cc} = \hat{h}$ (otherwise Theorem 1 would be violated) and the upper bound is achieved.

C. Proof of Theorem 3

For any $h \in \mathcal{F}_N$ one has

$$\begin{aligned} \|h_{ip} - h\| &\leq \|h_{ip} - h_c\| + \|h_c - h\| \\ &= \|\Pi_n h_i - h_c\| + \|h_c - h\| \\ &\leq \|\Pi_n h_i - h_i\| + \|h_i - h_c\| + \|h_c - h\| \\ &\leq d_i + r + \|h_c - h\| \end{aligned} \quad (23)$$

where it has been exploited the fact that $h_i \in \mathcal{F}_N$. Then, maximizing both sides of (23) over $h \in \mathcal{F}_N$ one gets $E[h_{ip}] \leq d_i + 2r$. On the other hand, one has that $d_i = \|\Pi_n h_i - h_i\| = \inf_p \inf_{\theta\in\mathbb{R}^n} \|h_i - M_p\theta\| \leq \inf_p \inf_{\theta\in\mathbb{R}^n} \sup_{\tilde{h}\in\mathcal{F}_N} \|\tilde{h} - M_p\theta\| = E[h_{cc}]$. Since, by definition, $r = \text{rad}(\mathcal{F}_N) \leq \sup_{\tilde{h}\in\mathcal{F}_N} \|h_{cc} - \tilde{h}\| = E[h_{cc}]$, one has $E[h_{cc}] \geq \max\{d_i, r\}$ and the result follows immediately.

D. Proof of Theorem 4

In order to prove Theorem 4, the following lemmas are needed.

Lemma 2. *Let $\mathcal{E} = \{x \in \mathbb{R}^N : \ x'Q\,x \leq 1\}$ be a non-degenerate axes-oriented ellipsoid, such that $Q = diag\{q_i\}_{i=1}^N$, with $0 < q_1 \leq q_2 \leq \ldots \leq q_N$. Moreover let $\mathcal{B} = \{z \in \mathbb{R}^N : \ z'z \geq d \ , \ d > 0\}$. Define*

$$z^* = \arg \inf_{z\in\mathcal{B}} \sup_{x\in\mathcal{E}} \|z - x\|_2^2. \quad (24)$$

Then, $z_1^* = 0$.

Proof. W.l.o.g., it can be assumed $d = 1$. Moreover, it is straightforward to show that the minimum in (24) is reached on the boundary of $\mathcal{B}$, i.e. $\|z^*\| = 1$. From Theorem 2 in [10], one has that for any z such that $z_1 \neq 0$

$$\max_{x\in\mathcal{E}} \|z - x\| = \|(I_N - \lambda^* Q)^{-1}\,\lambda^* Q\,z\|^2, \quad (25)$$

where λ^* is the largest real solution of the equation

$$z'(I_N - \lambda Q)^{-2} Q\,z - 1 = 0. \quad (26)$$

Moreover, it is known that $\lambda^* > \dfrac{1}{q_1}$. Hence, using (25) one has that z^* is the solution of

$$\inf_{z:\ z'z=1} \quad \sum_{i=1}^N \left(\frac{\lambda^* q_i z_i}{1 - \lambda^* q_i}\right)^2. \quad (27)$$

Exploiting (26) and substituting $z_1^2 = 1 - \sum_{i=2}^N z_i^2$ into (27), one has

$$\begin{aligned} \sum_{i=1}^N \left(\frac{\lambda^* q_i z_i}{1 - \lambda^* q_i}\right)^2 &= \left(\frac{\lambda^* q_1}{1 - \lambda^* q_1}\right)^2 \\ &\quad + \sum_{i=2}^N \left\{\left(\frac{\lambda^* q_i}{1 - \lambda^* q_i}\right)^2 - \left(\frac{\lambda^* q_1}{1 - \lambda^* q_1}\right)^2\right\} z_i^2. \end{aligned}$$

Since $q_1 \leq q_i$, for $i \geq 2$, it is easy to show that $\gamma_i \triangleq \left(\frac{\lambda^* q_i}{1-\lambda^* q_i}\right)^2 - \left(\frac{\lambda^* q_1}{1-\lambda^* q_1}\right)^2 < 0$, $i = 2, \ldots, N$ and hence the minimum in (27) is achieved for some $z_2^*, \ldots, z_N^*$ such that $(z_2^*)^2 + \ldots + (z_N^*)^2 = 1$. This implies $z_1^* = 0$. $\square$

A straightforward extension of Lemma 2 is obtained by translating the center of the ellipsoid $\mathcal{E}$ and by rotating the ellipsoid semi-axes onto a new reference system spanned by the orthonormal basis $\{v_1, \ldots, v_N\}$. Hence, the next result holds.

Lemma 3. *Let $\mathcal{E} = \{x \in \mathbb{R}^N : \ x = h_c + \sum_{i=1}^N \alpha_i\,v_i, \ \frac{\alpha_1^2}{L_M^2} + \frac{1}{L_m^2}\sum_{i=2}^N \alpha_i^2 \leq 1; \ h_c \in \mathbb{R}^N; \ \|v_i\| = 1, \ i = 1, \ldots, N; \ v_i'v_j = 0, \ i \neq j; \ L_M > L_m > 0\}$ and $\mathcal{B} = \{z : \ \|z - h_c\| \geq d \ , \ d > 0\}$. Define*

$$z^* = \arg \inf_{z\in\mathcal{B}} \sup_{x\in\mathcal{E}} \|z - x\|_2^2.$$

Then, $(z^* - h_c) \in span\{v_2, \ldots, v_N\}$ and $\|z^* - h_c\| = d$.

Now Theorem 4 can be proven. Let us first consider $E[h_{cp}] = E[\phi_{cp}(Z;p^*(h_c))]$. One has

$$\begin{aligned} E[h_{cp}] &= \sup_{h\in\mathcal{F}_N} \|h_{cp}-h\| \\ &\leq \|h_{cp}-h_c\| + \sup_{h\in\mathcal{F}_N} \|h_c-h\| \; = \; d+L_M. \end{aligned} \quad (28)$$

Then, let us analyze the minimum error $E[h_{cc}] = E[\phi_{cc}(Z;p^*(\mathcal{F}_N))]$. Let $\overline{\mathcal{F}}_N$ be an ellipsoid with the same center and axes orientation as $\mathcal{F}_N$, maximum semi-axis of length L_M and all other semi-axes of length L_m. By construction, $\overline{\mathcal{F}}_N \subseteq \mathcal{F}_N$. Moreover, $\overline{\mathcal{F}}_N$ coincides with $\mathcal{E}$ in Lemma 3 (with $v_1,\ldots,v_N$ being the directions of the semi-axes of $\mathcal{F}_N$). By definition of h_{cp} in (11), one has that the conditional center h_{cc} satisfies $\|h_{cc}-h_c\| \geq \|h_{cp}-h_c\| = d$. Hence $E[h_{cc}] = \sup_{h\in\mathcal{F}_N}\|h_{cc}-h\| \geq \sup_{\overline{h}\in\overline{\mathcal{F}}_N}\|h_{cc}-\overline{h}\| \geq \inf_{z:\ \|z-h_c\|\geq d}\ \sup_{\overline{h}\in\overline{\mathcal{F}}_N}\|z-\overline{h}\|$. Then, from Lemma 3 one gets

$$E[h_{cc}] \geq \sup_{\overline{h}\in\overline{\mathcal{F}}_N} \|z^*-\overline{h}\| \quad (29)$$

for some z^* such that $(z^*-h_c) \in span\{v_2,\ldots,v_N\}$ and $\|z^*-h_c\| = d$. W.l.o.g., assume $(z^*-h_c)\in span\{v_N\}$ (a rotation of the axes $v_2,\ldots,v_N$ can be applied without affecting $\overline{\mathcal{F}}_N$) and set $z^* = h_c - d\,v_N$ (the sign of v_N can be chosen arbitrarily). First, observe that when $L_m = 0$ the ellipsoid $\overline{\mathcal{F}}_N$ collapses onto a segment with extremal points $\overline{h} = h_c \pm L_M\,v_1$. Then, one has $\|z^*-\overline{h}\|^2 = L_M^2+d^2$, and hence from (28) and (29) $\frac{E[h_{cp}]}{E[h_{cc}]} \geq \frac{L_M+d}{\sqrt{L_M^2+d^2}}$ as stated in the upper part of (18).
Now, let us examine the case $L_m > 0$. A generic point on the boundary of $\overline{\mathcal{F}}_N$ can be written as $\overline{h} = h_c + \alpha_1\,v_1 + \ldots + \alpha_N\,v_N$, where

$$\frac{\alpha_1^2}{L_M^2} + \frac{1}{L_m^2}\sum_{i=2}^{N}\alpha_i^2 = 1. \quad (30)$$

Then,

$$\begin{aligned} \|z^*-\overline{h}\|^2 &= \|h_c - d\,v_N - h_c - \alpha_1\,v_1 - \ldots - \alpha_N\,v_N\|^2 \\ &= \alpha_1^2+\alpha_2^2+\ldots+(\alpha_N+d)^2 \\ &= \alpha_1^2\left[\frac{L_M^2-L_m^2}{L_M^2}\right] + 2\,\alpha_N\,d + L_m^2 + d^2 \end{aligned}$$

where the last equality has been obtained by using (30). Exploiting the above expression, the maximization of $\|z^*-\overline{h}\|$ with respect to $\overline{h}\in\overline{\mathcal{F}}_N$ is a straightforward exercise that leads to

$$\sup_{\overline{h}\in\overline{\mathcal{F}}_N}\|z^*-\overline{h}\| = \begin{cases} L_M\sqrt{1+\dfrac{d^2}{L_M^2-L_m^2}} & \text{if } d < \frac{L_M^2-L_m^2}{L_m}, \\ L_m+d & \text{if } d \geq \frac{L_M^2-L_m^2}{L_m}. \end{cases} \quad (31)$$

Then, (18) is an immediate consequence of (28), (29) and (31).