A Parametric Extension of Mixed Time/Frequency Robust Identification

Pablo A. Parrilo, Ricardo S. Sánchez Peña, and Mario Sznaier

***Abstract*—A parametric extension to the time/frequency robust identification framework is presented. The results can be applied to stable linear time-invariant systems on which time and/or frequency experiments have been performed. The parametric portion of the model should be affine in the unknown parameters, which includes practical applications such as flexible structures. The consistency problem is cast as a constrained finite-dimensional convex optimization problem that can be formulated as a linear matrix inequality. The proposed procedure provides an interpolatory identification algorithm, convergent and optimal up to a factor of two (with respect to central algorithms).**

***Index Terms*—Analytic interpolation, interpolatory algorithms, linear matrix inequalities, parametric identification, robust identification.**

I. Introduction

During the past few years a large research effort has been devoted to the problem of developing deterministic identification procedures that, starting from experimental data and an *a priori* class of models, generate a nominal model and bounds on the identification errors. These models and bounds can then be combined with standard robust control synthesis methods (such as $\mathcal{H}_\infty$, ℓ_1, or μ-synthesis) to obtain robust closed-loop systems. This problem, termed the robust identification problem, was originally posed in [11] and has since attracted considerable attention.

The case where the available experimental data is generated by frequency-domain experiments leads to $\mathcal{H}_\infty$-based identification procedures [3], [7], [11], [19]. When the experimental data available originates from time-domain experiments, ℓ_1 identification is used (see [14] and references therein). These two types of experimental data represent the response of the same system to different kinds of inputs. In the perfect information case (infinite number of samples, no noise) they are clearly equivalent. However, in practical cases (partial information contaminated by noise) using both kinds of data usually provides extra information.

Recent papers [5], [25] proposed interpolatory algorithms that use data obtained from time domain experiments to generate a nominal model together with an $\mathcal{H}_\infty$ bound on the identification error. Furthermore, a new robust identification framework that takes into account *both* time and frequency domain experiments, has been introduced in [21] and generalized in [18].

All of the above results address nonparametric identification of models with a worst case global bound. In many cases, part of the model has a clear parametric structure, and disregarding this information may lead to conservative results. This is usually the case of mechanical flexible structures, which have a well-defined parametric model for the lower frequency modes and an unknown higher frequency behavior which naturally leads to a nonparametric identification [8].

This new result extends the time/frequency identification procedure in [18], [21] and uses interpolation algorithms in [3], to the case of parametric/nonparametric model structures, based on a similar framework. The parametric portion of the model should be *affine* in the unknown parameters. This includes cases of practical interest as mechanical flexible structures, as well as models which can be concisely described in terms of a set of Laguerre or Kautz functions or, more generally, any other basis of $\mathcal{H}_2$ [12]. In the case of flexible structures, the lower frequency modes can be computed from the peaks measured from the approximate frequency response.

In general, the parametric information does not appear explicitly in the *a priori* knowledge (K, ρ) usually considered. Therefore, it is important to include it, if available, so that less conservative *a priori* estimates of (K, ρ) for the nonparametric portion can be derived. This is the case of large peaks in the frequency response of a flexible structure which should be "covered" by large values of K [8]. This is a consequence of the fact that the usual *a priori* information characterizes only the smoothness and magnitude of the whole class of models but cannot distinguish among other properties (e.g., low-frequency model structure). Instead, the *a priori* parametric knowledge provides more "structured" information.

The fact that from our procedure we obtain a consistency set allows different descriptions of the uncertainty. For instance, we can compute a global bound which covers the whole set and includes both the parametric and nonparametric uncertainty. This can also be interpreted as having "exact" nominal parameters and a global bound due to the uncertainty of the nonparametric portion. Unless we have extra information, both interpretations are indistinguishable. The approach of computing a global uncertainty bound is practical when considering robust controller design methods ($\mathcal{H}_\infty$, ℓ_1, μ-synthesis). Nevertheless, different bounds on the parameters and nonparametric portion can be computed as well, although there are no specific optimal synthesis algorithms in this case.

The paper is organized as follows. In Section II we introduce a robust identification framework using both time and frequency experiments and some background material. Section III contains the main theoretical results. We show that the problems of establishing consistency of the experimental data and the *a priori* information and of determining a nominal model can be recast into a finite-dimensional linear matrix inequality (LMI) optimization form, which generates a model that interpolates the frequency domain data points and is consistent with the time-domain experiments. Finally, in Section IV we illustrate the advantages of these results with a simple example.

II. Robust Identification Framework

The class of systems considered are discrete-time, causal, linear, and stable ones. We denote them as $H(z) = H_d(\frac{1}{z})$, with $z \in C$, and $H_d(z)$ being the usual z-transform. Therefore, causal stable systems $H(z)$ will be analytic inside the unit circle, with time and frequency representations related by

$$H(z) = \sum_{k=0}^{\infty} h(k) z^k. \tag{1}$$

For simplicity, we consider single-input/single-output (SISO) models, although all results can be applied to multi-input/multi-output (MIMO) systems, following [4].

Manuscript received August 6, 1996. Recommended by Associate Editor, S. Weiland. This work was supported by UBACyT projects IN018 and IN049, the University of Buenos Aires, and by the NSF under Grant ECS-9211169.

P. A. Parrilo is with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125 USA.

R. S. Sánchez Peña is with the Departmento de Electrónica, Univ. de Buenos Aires, Buenos Aires, Argentina and the National Commission of Space Activities (CONAE), Argentina.

M. Sznaier is with the Department of Electrical Engineering, Pennsylvania State University, University Park, PA USA.

Publisher Item Identifier S 0018-9286(99)01311-2.

The *a posteriori* experimental data can be obtained from two different sources: frequency response experiments and time-domain data. For the first one, we will consider the set of N_f samples of the frequency response of the system measured with additive bounded noise $y_k^f = \hat{h}_k + \eta_k^f$, $k = 0, \cdots, N_f - 1$, collected for notational simplicity in a vector $\mathbf{y}^f \in C^{N_f}$. Here, $\hat{\mathbf{h}}$ is a vector with its components $\hat{h}_k$ equal to $H(e^{j\Omega_k})$, $k = 0, \cdots, N_f - 1$. The additive measurement noise η^f is assumed to belong to a certain noise set, described below.

The time-domain data considered is the set of the first N_t time response samples to a known but otherwise arbitrary input, also corrupted by additive noise: $y_n^t = (U\mathbf{h})_n + \eta_n^t$, $n = 0, \cdots, N_t - 1$. We define

$$U \triangleq \begin{bmatrix} u(0) & 0 & \cdots & 0 \\ u(1) & u(0) & \cdots & 0 \\ \vdots & \ddots & \ddots & 0 \\ u(N_t-1) & \cdots & u(1) & u(0) \end{bmatrix}$$

as the Toeplitz matrix corresponding to the input sequence. Again, for notational simplicity, the measurements will be collected in a vector $\mathbf{y}^t \in R^{N_t}$.

To define the *a priori* noise classes, it is useful to introduce the concept of *LMI regions* [6]. These are regions of the complex plane, which are the feasibility domain of a given LMI. Such a class of regions is fairly general, as it can be shown [6] that its closure is the set of all convex regions of the complex plane.

In this case, the LMI regions are useful to model noise sets because they provide a simple and exact way to take into account multiple point frequency response experiments, i.e., the experiments which produce $\mathbf{y}^f$. In this situation, the "confidence region" for the true system response is the intersection of the individual confidence regions [11]. Therefore, if the latter can be defined by means of an LMI, the same is also true for the region corresponding to all the experiments. Furthermore, this intersection region can be described by the combination of all individual LMI's, i.e., the intersection of LMI regions is also an LMI region.

In the time domain case, the noise is real-valued, and therefore convex sets are in fact just intervals. However, adopting an LMI-defined set can prove advantageous if the independency assumption on noise samples turns out to be conservative. For example, it is possible to impose convex correlation constraints between samples. In this case, the noise set is no longer the Cartesian product of the individual noise intervals.

In view of the preceding remarks, we define our *a priori* noise sets $\mathcal{N}_f$, $\mathcal{N}_t$ as the feasibility regions of a set of LMI's

$$\mathcal{N}_f \triangleq \{\eta^f \in C^{N_f}, L^f(\eta^f) > 0\}$$
$$\mathcal{N}_t \triangleq \{\eta^t \in R^{N_t}, L^t(\eta^t) > 0\}$$

where $L^f(\eta^f)$ and $L^t(\eta^t)$ are hermitian and symmetric matrices, respectively, and depend affinely on their arguments. In order to compute error bounds, we also assume these *a priori* noise classes $\mathcal{N}_f, \mathcal{N}_t$ to be bounded. Note that the $\ell_\infty(\epsilon)$ noise sets usually considered [3], [18] are just special cases of the above.

To introduce the *a priori* system class, first define the following sets:

$$\mathcal{H}_\infty(\rho, K) \triangleq \left\{ H \in \mathcal{H}_{\infty,\rho} \;\middle|\; \sup_{z \in \mathcal{D}_\rho} |H(z)| < K \right\}$$

with

$$\mathcal{H}_{\infty,\rho} \triangleq \left\{ H(z) \text{ analytic in } \mathcal{D}_\rho \;\middle|\; \sup_{z \in \mathcal{D}_\rho} |H(z)| < \infty \right\}$$

where $\mathcal{D}_\rho$ is the open disk $\{z, |z| < \rho\}$ and $\overline{\mathcal{D}}_\rho$ is its corresponding closure. Also define $\mathcal{H}_\infty$ as $\mathcal{H}_{\infty,1}$, $\mathcal{D}$ as $\mathcal{D}_1$ and $\overline{\mathcal{B}}\mathcal{H}_\infty \triangleq \{F \in \mathcal{H}_\infty, \|F\|_\infty \leq 1\}$.

The class of models usually considered $\mathcal{H}_\infty(\rho, K)$ corresponds to exponentially stable systems (finite or infinite-dimensional). From a practical viewpoint, these systems have a stability margin of $(\rho - 1)$, a peak response to complex exponential inputs of K, and satisfy the time domain bound

$$|h(k)| \leq K\rho^{-k}. \tag{2}$$

The above-defined sets will be used to characterize the nonparametric part of the model $H_{\text{np}}(z)$.

We seek the worst case identification of the combined parametric and nonparametric models using both the time and frequency experiments. The models in our *a priori* class will have a mixed parametric/non parametric structure, i.e., $H = H_p + H_{\text{np}}$. To this end we define the sets of systems

$$\mathcal{S} \triangleq \{\mathcal{H}_\infty(\rho, K), \rho > 1, K < \infty\}$$

for the nonparametric part and $\mathcal{P}$ for the parametric component $H_p(z)$. The latter is the class of affine models defined as

$$\mathcal{P} \triangleq \{\mathbf{p}^T \mathbf{G}(z), \mathbf{p} \in R^{N_p}, p_i \in [a_i, b_i]\} \tag{3}$$

where the components $g_i(z)$ of the vector $\mathbf{G}(z)$ are known linearly independent functions. In Section III, in order to compute parametric error bounds, we will also require these functions to satisfy the separation condition

$$\text{span}\{g_i(z)\} \cap \mathcal{S} = \{0\}. \tag{4}$$

This condition is in fact an "identifiability" requirement and guarantees a unique parametric/nonparametric decomposition for each element $H_p(z) + H_{\text{np}}(z)$ of the *a priori* set. To see this, note that if $H(z) = A_1(z) + B_1(z) = A_2(z) + B_2(z)$, $A_i \in \mathcal{S}, B_i \in \mathcal{P}$, then $(A_1 - A_2)/2$ belongs both to $\text{span}\{g_i(z)\}$ and $\mathcal{S}$, in contradiction with (4). If this decomposition is not unique, then there are multiple parameter choices for each model in the *a priori* set, and as a consequence no "true" parameter values. Therefore, the parametric and nonparametric components will not converge separately, although the full model might converge to the real plant.

Another motivation for the above separation condition between the parametric and nonparametric components is as follows. There is a clear advantage to having a parametric component when it cannot be tightly described with a nonparametric model. In that case, a global nonparametric description would be unnecessarily conservative [8]. On the other hand, in the case where the parametric portion can be included in the nonparametric one, there is no special reason to describe it parametrically. For all the previous arguments, we consider the case where the parametric and nonparametric components are clearly separated. This type of model description includes many practical situations, e.g., flexible structures. In the latter case, the parametric/nonparametric separation is in terms of the system natural frequencies. The lower frequencies have a parametric description, while the higher frequencies can be modeled nonparametrically.

Therefore, the *a priori* information and the *a posteriori* experimental input data are

$$\begin{aligned} \mathcal{T} &= \{H(z) = H_p(z) + H_{\text{np}}(z) \mid H_p \in \mathcal{P}, H_{\text{np}} \in \mathcal{S}\} \\ \mathcal{N}_f &= \{\eta^f \in C^{N_f}, L^f(\eta^f) > 0\} \\ \mathcal{N}_t &= \{\eta^t \in R^{N_t}, L^t(\eta^t) > 0\} \\ \mathbf{y}^f &= \{\hat{\mathbf{h}} + \eta^f \in C^{N_f}\} \\ \mathbf{y}^t &= \{U\mathbf{h} + \eta^t \in R^{N_t}\}. \end{aligned}$$

Both the consistency and identification procedures are considered; therefore, the problem to be solved is as follows.

Problem 1: Given the experiments $(\mathbf{y}^f, \mathbf{y}^t)$ and the *a priori* sets $(\mathcal{T}, \mathcal{N}_f, \mathcal{N}_t)$, determine:

(1) if the *a priori* and *a posteriori* information are consistent, i.e., the consistency set

$$\mathcal{T}(\mathbf{y}^f, \mathbf{y}^t) \triangleq \left\{ H \in \mathcal{T} \;\middle|\; \begin{matrix} (\mathbf{y}^f - \hat{\mathbf{h}}) \in \mathcal{N}_f \\ (\mathbf{y}^t - U\mathbf{h}) \in \mathcal{N}_t \end{matrix} \right\} \tag{5}$$

is nonempty;

(2) a nominal model which belongs to the consistency set $\mathcal{T}(\mathbf{y}^f, \mathbf{y}^t)$.

The above *a priori* and *a posteriori* information allows a less conservative identification than in the cases where only time or frequency experiments are provided or when no knowledge of the parametric part of the model is used. Another interpretation can be made in terms of the "smaller" size of the consistency set, due to the fact that there are more experimental data and structured *a priori* information.

III. Main Results

Nevanlinna–Pick-based identification algorithms address the case where the experimental data available is purely frequency domain [3], while Carathéodory–Fejér-based identification deals only with time-domain data [5]. In this section we introduce a generalized interpolation framework and use it to solve Problem 1, obtaining a robust identification algorithm that combines both sources of data. To this effect, we divided Problem 1 into two subproblems: 1) consistency and 2) identification. The first should determine the existence of a candidate model $H \in \mathcal{T}$ which may have produced both the time- and frequency-domain experimental data. Clearly, this is a prerequisite to the second stage, the identification step, consisting of the computation of the nominal model itself and a bound on the identification error.

A. Generalized Interpolation Framework

In this section we briefly present a generalized interpolation result developed in [1] and applied to $\mathcal{H}_\infty$ control in [20]. This lemma will be used later to solve the consistency problem.

Lemma 1: There exists a transfer function $F(z) \in \mathcal{BH}_\infty$ $(\overline{\mathcal{B}}\mathcal{H}_\infty)$ such that

$$\sum_{z_o \in \mathcal{D}} \mathrm{Res}_{z=z_o} F(z) C_-(zI - A)^{-1} = C_+ \tag{6}$$

if and only if there is a unique positive (semi)definite solution of the discrete-time Lyapunov equation

$$M = A^* M A + C_-^* C_- - C_+^* C_+ \tag{7}$$

where A, C_- and C_+ are constant complex matrices of appropriate dimensions. If $M > 0$ then the solution $F(z)$ is nonunique and the set of solutions can be parameterized in terms of $Q(z)$, an arbitrary element of $\overline{\mathcal{B}}\mathcal{H}_\infty$, as follows:

$$F(z) = \frac{T_{11}(z)Q(z) + T_{12}(z)}{T_{21}(z)Q(z) + T_{22}(z)} \tag{8}$$

where $T_{ij}(z)$ are computed from A, C_+, C_-, and M.

Proof: See [1] and [20]. □

Note that the matrices A and C_- provide the structure of the interpolation problem while C_+ provides the interpolation values. It has been shown in [1] that both the Nevanlinna–Pick and the Carathéodory–Fejér problems are special cases of this theorem, corresponding to an appropriate choice of the matrices A and C_-.

B. Consistency

We will see now that the problem of determining consistency of the *a posteriori* and *a priori* information reduces to establishing whether or not there exists a model $H \in \mathcal{T}$ that interpolates the frequency experimental data within the noise bounds

$$\hat{\mathbf{h}} = \mathbf{y}^f - \eta^f, \qquad \eta^f \in \mathcal{N}_f \tag{9}$$

and has an impulse response that satisfies the following constraints:

$$U\mathbf{h} = \mathbf{y}^t - \eta^t, \qquad \eta^t \in \mathcal{N}_t \tag{10}$$

where the noiseless output $U\mathbf{h}$ is the convolution of the input vector $\mathbf{u} = [u(0)\ \ u(1)\ \ \cdots\ \ u(N_t - 1)]^T$ and the system $H(z)$.

The main result of this section shows that consistency can be established by solving a finite-dimensional convex optimization problem. To establish this result we formulate the equivalent condition for consistency. This condition, based upon the relationship between both admissible experimental noises $\eta_f \in \mathcal{N}_f$ and $\eta_t \in \mathcal{N}_t$, has the form of a linearly constrained generalized interpolation problem. In Theorems 2 and 3 we show that this generalized problem can be recast in terms of an LMI optimization.

From (5), it is clear that the *a priori* and *a posteriori* information are consistent if and only if there exists a function $H \in \mathcal{T}$ such that

$$\hat{\mathbf{h}} = \mathbf{y}^f - \eta^f, \qquad \eta^f \in \mathcal{N}_f \tag{11}$$

$$U\mathbf{h} = \mathbf{y}^t - \eta^t, \qquad \eta^t \in \mathcal{N}_t. \tag{12}$$

The next theorem provides necessary and sufficient conditions for the existence of a function $H \in \mathcal{H}_\infty(\rho, \mathcal{K})$ which interpolates *fixed* frequency domain experimental data while, at the same time, satisfies a time-domain constraint.

Theorem 1: Given N_f frequency-domain data points, $H(z_i) = w_i,\ i = 0, \cdots, N_f - 1$, and N_t time-domain data points $h(k)$, $k = 0, \cdots, N_t - 1$, there exists $H \in \mathcal{H}_\infty(\rho, \mathcal{K})$ that interpolates the frequency domain data and such that $H(z) = h(0) + h(1)z + h(2)z^2 + \cdots + h(N_t - 1)z^{N_t - 1} + \cdots$ if and only if

$$M_R(\mathbf{w}, \mathbf{h}) \triangleq \begin{bmatrix} Q - \frac{1}{K^2}\mathcal{W}_f^* Q \mathcal{W}_f & M_X \\ M_X^* & R^{-2} - \frac{1}{K^2}\mathcal{F}_t^* R^{-2} \mathcal{F}_t \end{bmatrix} > 0 \tag{13}$$

where

$$M_X = S_0 R^{-2} - \frac{1}{K^2}\mathcal{W}_f^* S_0 R^{-2} \mathcal{F}_t \tag{14}$$

$$R = \mathrm{diag}[1 \quad \rho \quad \rho^2 \quad \cdots \quad \rho^{N_t - 1}] \tag{15}$$

$$Q = \left[\frac{\rho^2}{\rho^2 - z_{i-1}^* z_{j-1}} \right]_{ij}, \quad i, j = 1, \cdots, N_f \tag{16}$$

$$S_0 = \left[\left(z_{i-1}^{j-1} \right)^* \right]_{ij}, \quad i = 1, \cdots, N_f, \quad j = 1, \cdots, N_t \tag{17}$$

$$\mathcal{W}_f = \mathrm{diag}\begin{bmatrix} w_0 & \cdots & w_{N_f - 1} \end{bmatrix} \tag{18}$$

$$\mathcal{F}_t = \begin{bmatrix} h(0) & h(1) & \cdots & h(N_t - 1) \\ 0 & h(0) & \cdots & h(N_t - 2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h(0) \end{bmatrix}. \tag{19}$$

Proof: See [18]. □

Combining the previous result (that considers only noiseless data points) with (11), (12), and the assumed parametric structure yields the following necessary and sufficient condition for consistency.

Theorem 2: Define

$$P_t = \begin{bmatrix} g_1(0) & g_2(0) & \cdots & g_{N_p}(0) \\ g_1(1) & g_2(1) & \cdots & g_{N_p}(1) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(N_t - 1) & g_2(N_t - 1) & \cdots & g_{N_p}(N_t - 1) \end{bmatrix} \tag{20}$$

and

$$P_f = \begin{bmatrix} G_1(z_0) & G_2(z_0) & \cdots & G_{N_p}(z_0) \\ G_1(z_1) & G_2(z_1) & \cdots & G_{N_p}(z_1) \\ \vdots & \vdots & \ddots & \vdots \\ G_1(z_{N_f-1}) & G_2(z_{N_f-1}) & \cdots & G_{N_p}(z_{N_f-1}) \end{bmatrix}. \quad (21)$$

Then, the *a priori* and *a posteriori* information are consistent if and only if there exist three vectors:

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{N_p} \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{N_f-1} \end{bmatrix}, \qquad \mathbf{h} = \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{N_t-1} \end{bmatrix} \quad (22)$$

such that

$$M_R(\mathbf{w}, \mathbf{h}) > 0 \quad (23)$$

$$(\mathbf{y}^f - P_f\mathbf{p} - \mathbf{w}) \in \mathcal{N}_f \quad (24)$$

$$(\mathbf{y}^t - UP_t\mathbf{p} - U\mathbf{h}) \in \mathcal{N}_t. \quad (25)$$

Note that the components of $\mathbf{w}$ and $\mathbf{h}$ are elements of the matrices $\mathcal{W}_f$ and $\mathcal{F}_t$, respectively, and $g_i(k)$ the impulse response of $G_i(z)$.

Proof: Since $H = H_p + H_{\rm np}$, we have that $\hat{\mathbf{h}} = P_f\mathbf{p} + \mathbf{w}$ and $\mathbf{h} = P_t\mathbf{p} + \mathbf{h}$, where P_f, P_t are defined above. Now, by the previous theorem, $H_{\rm np}$ is in $\mathcal{S}$ if and only if (23) holds. Finally, we can substitute $\hat{\mathbf{h}}, \mathbf{h}$ in (11) and (12) to obtain (24) and (25). □

From Theorem 2 it follows that the consistency problem can be reduced to solving a feasibility problem in terms of the parameter vector $\mathbf{p}$ and the time and frequency domain vectors $\mathbf{h}, \mathbf{w}$. In the next theorem we show that this feasibility problem is convex and can be recast in terms of LMI's. Therefore, it can be efficiently solved, using for instance interior-point methods [16], [2].

Theorem 3: The parametric/nonparametric consistency problem with mixed time/frequency-domain data is equivalent to a LMI feasibility problem.

Proof: The matrix M_R can be written as follows:

$$M_R = M_0 - \frac{1}{K^2}X^*M_0X \quad (26)$$

$$M_0 = \begin{bmatrix} Q & S_0R^{-2} \\ R^{-2}S_0^* & R^{-2} \end{bmatrix} \quad (27)$$

$$X = \begin{bmatrix} \mathcal{W}_f & 0 \\ 0 & \mathcal{F}_t \end{bmatrix}. \quad (28)$$

Positiveness of the matrix M_0 is equivalent to consistency in the case where both frequency and time domain data are zero, i.e., $w_i = 0$, $i = 0, \cdots, N_f - 1$ and $h_j = 0$, $j = 0, \cdots, N_t - 1$. Clearly, in this case the solution is not unique (we have as solutions the trivial $H(z) = 0$ and the Blaschke product). From Lemma 1 it follows that $M_0 > 0$, and thus, using Schur complements, we have

$$M_R > 0 \iff Z \triangleq \begin{bmatrix} M_0^{-1} & \frac{1}{K}X \\ \frac{1}{K}X^* & M_0 \end{bmatrix} > 0. \quad (29)$$

Clearly this is an LMI in X.

Since the admissible noise classes $(\mathcal{N}_f, \mathcal{N}_t)$ are defined by means of LMI's, the consistency constraints (24) and (25) can be rewritten as the following LMI's in the variables $\mathbf{p}, \mathbf{w}$, and $\mathbf{h}$:

$$L^f(\mathbf{y}^f - P_f\mathbf{p} - \mathbf{w}) > 0 \quad (30)$$

$$L^t(\mathbf{y}^t - UP_t\mathbf{p} - U\mathbf{h}) > 0. \quad (31)$$

Thus, the consistency problem is equivalent to finding a feasible solution to the set of LMI's (29), (30) and (31). □

An interesting question that appears at this point is related with the particular choice of a feasible solution for the LMI equations (29)–(31). For example, one could minimize the "size" of the noise set, subjected to the consistency restrictions, or maximize the smallest eigenvalue of the generalized Pick matrix, over the feasibility set, to avoid the ill-conditioning of the singular case. Alternatively, a minimization over K (a bound on the norm of the nonparametric component) could be performed. From a control perspective, the latter option seems specially useful.

C. Identification

Once consistency is established, the second step toward solving the problem stated in Section II consists of generating a nominal model in the consistency set $\mathcal{T}(\mathbf{y}^f, \mathbf{y}^t)$. The identification algorithm that we propose is based on the parameterization of all solutions of the generalized Nevanlinna–Pick interpolation problem [1] presented in Lemma 1. For simplicity we consider the case where the matrix M_R is strictly positive definite and therefore the solution is nonunique. Details for the degenerate case where there exists a unique solution can be found in [1]. The algorithm can be summarized as follows.

1) Find feasible data vectors $\mathbf{p}, \mathbf{w}, \mathbf{h}$ for the consistency problem (23)–(25) by solving the LMI feasibility problem given by (29)–(31). Note that there is no need of any kind of optimality in the search for the feasible vectors. Instead, any triple $\mathbf{p}, \mathbf{w}, \mathbf{h}$ in the admissible set will suffice.
2) Compute the generalized Pick matrix M_R in (13) (which should be positive definite), corresponding to the vectors computed in Step 1.
3) Use Lemma 1 to compute a model from the consistency set $\mathcal{T}$. Recall that all the models in $\mathcal{S}$ (i.e., all the solutions to the generalized interpolation problem) can be parameterized as a linear fractional transformation (LFT) of a free parameter $Q(z) \in \bar{\mathcal{B}}\mathcal{H}_\infty$ as follows:

$$H(z) = F_\ell[L(z), Q(z)] \quad (32)$$

$$L(z) = \begin{bmatrix} T_{12}T_{22}^{-1} & T_{11} - T_{12}T_{22}^{-1}T_{21} \\ T_{22}^{-1} & -T_{22}^{-1}T_{21} \end{bmatrix}. \quad (33)$$

In particular, if the free parameter $Q(z)$ is chosen as a constant, then the nonparametric model order is less than or equal to $N_f + N_t$.

Remark 1: Note that $T(z)$ depends on the choice of vectors $\mathbf{w}, \mathbf{h}$. Thus, there are additional degrees of freedom available in the problem [choices of $\mathbf{w}, \mathbf{h}$ and $Q(z)$] that could be used to optimize additional performance criteria (e.g., model order).

D. Analysis of the Identification Error

In this section we derive an upper bound for the worst case identification error. Since this bound is given in terms of the *radius* and *diameter* of information [11], [3] they are valid for *all* interpolatory algorithms taking as inputs the available *a priori* and *a posteriori* information (see [18] for a lower bound).

An important point that needs to be emphasized at this stage lies in the fact that the identification error bounds are dependent on the assumed *a priori* information. That is, there will be different error bounds according to *what* is considered to be *a priori*. For example, by considering a fixed parametric component it is possible to change the *a priori* assumptions without necessarily changing the identified model, and keeping consistency in the process. In this case, the problem setup is nearly the same as the nonparametric case, and therefore the error bounds can be computed, for example, using the results in [18]. On the other hand, if we keep the *a priori* information structure of the parametric-nonparametric approach, the following weaker bounds can be proved.

Since the identified model is in set $\mathcal{T}(\mathbf{y}^f, \mathbf{y}^t)$, its distance to the Chebyshev center of this set is within the diameter of information.

As a consequence the algorithm is optimal up to a factor of two as compared with central strongly optimal procedures. For the same reasons, it is also convergent, and therefore the modeling error tends to zero as the information is completed. Next we determine an upper bound on the radius of information and prove it in the spirit of [3].

Lemma 2: Assume the *a priori* noise classes $\mathcal{N}_f, \mathcal{N}_t$ to be $\ell_\infty(\epsilon_f), \ell_\infty(\epsilon_t)$. Also assume that the separation condition (4) holds. Then the radius of information $\mathcal{R}_\mathcal{I}$ can be bounded above by

$$\mathcal{R}_\mathcal{I} \leq \sum_{i=1}^{N_p} \mu_i \|g_i(z)\|_\infty + \sum_{i=0}^{M} \nu_i + \frac{K}{\rho^{M+1}} \tag{34}$$

where $M = N_t + N_f - N_p - 1$ and μ_i, ν_i are a function of the *a priori* information only.

Proof: Consider any $T \in \mathcal{T}(0,0)$ and partition it as follows:

$$T = P(z) + H(z) \tag{35}$$

$$T = \overbrace{\sum_{k=1}^{N_p} p_k g_k(z)}^{P(z)} + \underbrace{\sum_{k=0}^{M} h_1(k) z^k}_{H_1(z)} + H(z) - H_1(z) \tag{36}$$

where $P(z)$ and $H(z)$ are the parametric and nonparametric components of T, and $H_1(z)$ is the globally optimal approximation of $H(z)$ [9], which satisfies $\|H(z) - H_1(z)\|_\infty \leq K\rho^{-(M+1)}$. Consider now the first two terms of the expansion. $\mathbf{p}$ and h_1 satisfy

$$V \begin{bmatrix} \mathbf{p} \\ h_1 \end{bmatrix} = \begin{bmatrix} \eta^t \\ \eta^f \end{bmatrix}$$

$$V = \begin{bmatrix} UP_t & & U & & 0 & \\ & 1 & z_0 & z_0^2 & \cdots & z_0^M \\ P_f & 1 & z_1 & z_1^2 & \cdots & z_1^M \\ & \vdots & \vdots & \vdots & \vdots & \vdots \\ & 1 & z_{N_f-1} & z_{N_f-1}^2 & \cdots & z_{N_f-1}^M \end{bmatrix} \tag{37}$$

where $\eta^t \in \mathcal{N}_t$, $\eta^f \in \ell_\infty(\bar{\epsilon})$, and where $\bar{\epsilon} \triangleq \epsilon_f + K\rho^{-(M+1)}$. The idea of the proof is to derive bounds for $\mathbf{p}, h_1$ from those of η_f, η_t, assuming the nonsingularity[1] of V. To do this, partition V^{-1} conformally with (37) as

$$V^{-1} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{38}$$

Therefore

$$\begin{aligned} \mathbf{p} &= V_{11}\eta^t + V_{12}\eta^f \\ h_1 &= V_{21}\eta^t + V_{22}\eta^f \end{aligned} \tag{39}$$

where $\eta^t \in \ell_\infty(\epsilon_t)$, $\eta^f \in \ell_\infty(\bar{\epsilon})$. This description can be employed together with (2) to provide tighter bounds for the parameter vector and the first coefficients, using for instance linear programming. It is also possible to derive easier, but more conservative bounds. Let $(A)_i$ be the ith row of matrix A. Then, defining

$$\mu_i \triangleq \|(V_{11})_i\|_1 \epsilon^t + \|(V_{12})_i\|_1 \bar{\epsilon} \tag{40}$$

$$\nu_i \triangleq \min\left[K\rho^{-i}\left(1 - \rho^{2(i-M-1)}\right), \left\|(V_{21})_{(i+1)}\right\|_1 \epsilon^t + \left\|(V_{22})_{(i+1)}\right\|_1 \bar{\epsilon}\right] \tag{41}$$

[1]A necessary condition for V to be nonsingular is that the $g_i(z)$, $i = 1, \cdots, N_p$ be linearly independent. Even in this case, and when (4) holds, it is possible to have a singular V. This is related with the fact that the *experiment* does not provide enough information to compute an error bound, for the assumed parametric structure. That is, although (4) holds, certain finite-dimensional restriction of it does not. As an example of this, consider the case of taking only time samples up to time k and where the parametric functions have nonzero values only for times greater than $k+1$. The problem is solved simply by taking more (time or frequency) samples.

we have, from (39) and the global optimality property of $H_1(z)$, that $|\mathbf{p}(i)| \leq \mu_i, |h_1(i)| \leq \nu_i$.

The bounds on $\mathbf{p}$ and h_1 can therefore be used to give an estimate of the radius of information

$$\mathcal{R}_\mathcal{I} \leq \|T(z)\|_\infty \tag{42}$$

$$\leq \sum_{i=1}^{N_p} \mu_i \|g_i(z)\|_\infty + \sum_{i=0}^{M} \nu_i + \frac{K}{\rho^{M+1}}. \tag{43}$$

□

It is necessary to emphasize that the presented bounds (as most of the ones appearing in the literature) are *a priori* bounds, that is, they do not depend on the experimental data. The computational requirements for obtaining them are minimal, and its main application is in proving algorithm convergence. Since these bounds can be conservative, the practical application of the identification procedures usually requires alternative methods for computing uncertainty bounds [10]. In particular, it is possible to obtain better (for example, frequency dependent) bounds by solving convex optimization problems and making use of the experimental data.

To conclude this section, we briefly present a few comments on the computational complexity of the proposed procedure. The resulting convex feasibility problem (29)–(31) has a number of decision variables equal to the number of time and frequency samples (plus the number of unknown parameters). In other words, the "size" of the LMI's to be solved grows linearly with the amount of data. Furthermore, our particular problem has a lot of additional structure, that can be potentially exploited for greater efficiency. However, the available solvers do not allow (yet) the use of this extra information. Additional computational considerations have been presented in [18].

IV. Example

In this section we present an application of the proposed identification procedure. The "experimental" data proceeds from the stable component (no rigid-body modes) of the Euler–Bernoulli model of a flexible beam with viscous damping. For simplicity, we considered only frequency data, bilinearly transformed from the original continuous-time samples. Only 13 frequency points were evaluated, for equispaced frequencies between 0 and π. Fictitious complex noise, bounded in magnitude by 0.03, was added to the samples.

We assumed a parametric component having the following affine structure:

$$G(z) = \frac{p_1 z + p_2}{z^2 + 0.04z + 1.05} \tag{44}$$

where p_1 and p_2 are the uncertain parameters. The resonant poles of $G(z)$ were chosen according to the information available on the critical frequencies and damping factors of the plant. The value of ρ used was $\rho = 1.25$. The optimization criterion chosen was to minimize the value of K, subject to the consistency restrictions. This can also be formulated as a convex program and can be interpreted as an attempt to obtain the smallest possible uncertainty bound.

In Fig. 1 an identified model for the purely nonparametric identification procedure in [18] is shown. The optimal value of K was $K = 113.6$. As noted above, in this approach the highly resonant poles force very small values of ρ and/or large values of K, in order to satisfy the consistency constraints. This, in turn, causes larger error bounds, as well as less smooth interpolation functions. This issue is important in order to apply a subsequent model reduction stage. Note, however, that the oscillating characteristics of the identified model are also a consequence of the choice of the free parameter $Q(z) = 0$. It is theoretically possible to find a $Q(z)$ (dynamic) such

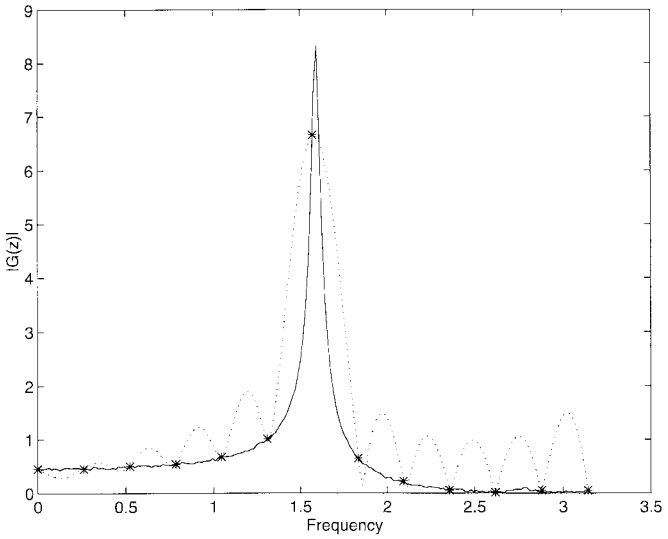

Fig. 1. Nonparametric identification results. The solid line is the physical system (with noise), and the dotted line the nonparametric identified model.

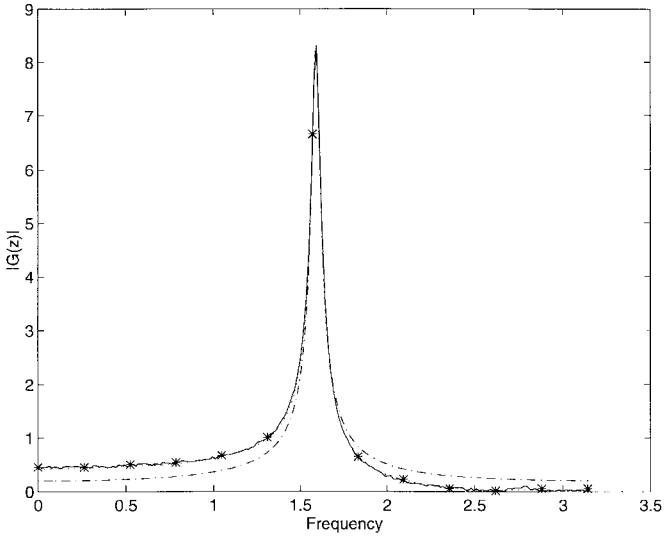

Fig. 2. Mixed parametric/nonparametric identification results. The solid line is the physical system (with noise), the dash-dotted line is the parametric model, and the dotted line is the full identified model.

that a smoother model results. However, there seems to be no easy procedure to achieve this. Further, this procedure would not change the error bound.

Instead, when the approach developed in this paper is employed, the complex dynamics of the parametric component can be "decoupled" from the nonparametric one. Fig. 2 shows the identified parametric component, as well as the full model. The optimal value of K (only for the nonparametric part) is $K = 0.35$. Note that the parametric component is *not* able to explain, by itself, the whole dynamic behavior. Its objective is to model just one particular mode, rather than the entire transfer function. Note also that the identified nominal model is almost coincident with the experiment, providing a much better fit.

ACKNOWLEDGMENT

The authors gratefully acknowledge the reviewers' comments.

REFERENCES

[1] J. A. Ball, I. Gohberg, and L. Rodman, *Operator Theory: Advances and Applications*. Boston, MA: Birkhäuser, vol. 45, 1990.
[2] S. Boyd, L. El. Ghaoui, E. Feron, and V. Balakrishnan, *SIAM Studies in Applied Mathematics*, vol. 15. Philadelphia, PA: SIAM, 1994.
[3] J. Chen, C. N. Nett, and M. K. H. Fan, "Worst-case system identification in $\mathcal{H}_\infty$: Validation of *a Priori* information, essentially optimal algorithms, and error bounds," *IEEE Trans. Automat. Contr.*, vol. 40, July 1995.
[4] J. Chen, J. A. Farrell, C. N. Nett, and K. Zhou, "$\mathcal{H}_\infty$ identification of multivariable systems by tangential interpolation methods," in *Proc. Conf. Decision and Control*, 1994.
[5] J. Chen and C. N. Nett, "The Carathéodory-Fejér problem and the $\mathcal{H}_\infty/\ell_1$ identification: A time domain approach," *IEEE Trans. Automat. Contr.*, vol. 40, Apr. 1995.
[6] P. Gahinet, A. Nemirovski, A. Laub, and M. Chilali, *LMI Control Toolbox*. Natick, MA, The MathWorks, 1995.
[7] G. Gu and P. P. Khargonekar, "A class of algorithms for identification in $\mathcal{H}_\infty$," *Automatica*, vol. 28, no. 2, 1992.
[8] J. Friedman and P. P. Khargonekar, "Application of identification in $\mathcal{H}_\infty$ to lightly damped systems: Two case studies," *IEEE Trans. Contr. Syst. Technol.*, vol. 3, Mar. 1995.
[9] G. Gu, "Suboptimal algorithms for worst case identification in $\mathcal{H}_\infty$ and model validation," *IEEE Trans. Automat. Contr.*, vol. 39, Aug. 1994.
[10] R. Hakvoort, "System identification for robust process control," Ph.D. dissertation, Delft Univ., 1994.
[11] A. J. Helmicki, C. A. Jacobson, and C. N. Nett, "Control oriented system identification: A worst case/deterministic approach in $\mathcal{H}_\infty$," *IEEE Trans. Automat. Contr.*, vol. 36, Oct. 1991.
[12] P. S. C. Heuberger, P. M. J. Van den Hof, and O. H. Bosgra, "A generalized orthonormal basis for linear dynamical systems," *IEEE Trans. Automat. Contr.*, vol. 40, Mar. 1995.
[13] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[14] M. Milanese, "Worst case ℓ_1 identification," *Bounding Approaches to System Identification*, Milanese, Norton, Piet-Lahanier, and Walter, Eds. New York: Plenum, 1994.
[15] Y. Nesterov and A. Nemirovskii, *Optimization over Positive Semidefinite Matrices: Mathematical Background and User's Manual*, USSR Acad. Science Central Econ. Mathematical Inst., 1990.
[16] ——, *SIAM Studies in Applied Mathematics*. Philadelphia, PA: SIAM, vol. 13, 1994.
[17] M. Overton, "On minimizing the maximum eigenvalue of a symmetric matrix," *SIAM J. Matrix Analysis and Appl.*, vol. 9, no. 2, 1988.
[18] P. A. Parrilo, M. Sznaier, R. S. Sánchez Peña, and T. Inanc, "Mixed time/frequency robust domain based identification," *Automatica*, vol. 34, no. 11, pp. 1375–1389, 1998.
[19] J. R. Partington, "Robust identification in $\mathcal{H}_\infty$," *J. Math. Analysis and Appl.*, vol. 166, pp. 428–441, 1992.
[20] H. Rotstein, "A Nevanlinna-Pick approach to time-domain constrained $\mathcal{H}_\infty$ control," *SIAM J. Contr. Optim.*, vol. 34, no. 4, pp. 1329–1341, 1996.
[21] R. S. Sánchez Peña and M. Sznaier, "Robust identification with mixed time/frequency experiments: Consistency and interpolation algorithms," in *Proc. Conf. Decision and Contr.*, 1995.
[22] R. Smith and J. C. Doyle, "Model validation: A connection between robust control and identification," *IEEE Trans. Automat. Contr.*, vol. 37, July 1992.
[23] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, pp. 49–95, 1996.
[24] ——, *SP—Software for Semidefinite Programming*, Nov. 1994.
[25] T. Zhou and H. Kimura, "Time domain identification for robust control," *Syst. Contr. Lett.*, vol. 20, pp. 167–178, 1993.