Convex Relaxations for Robust Identification of Wiener Systems and Applications

Burak Yilmaz   Mustafa Ayazoglu   Mario Sznaier   Constantino Lagoa

Abstract— **This paper considers the identification of Wiener systems in a worst case framework. Given some *a priori* information about the admissible set of plants, nonlinearities and measurement noise, and *a posteriori* experimental data, our goal is twofold: (i) establish whether the *a priori* and *a posteriori* information are consistent, and (ii) in that case find a model that interpolates the available experimental information within the noise level. As recently shown, this problem is generically NP hard both in the number of data points and the number of inputs to the non-linearity. Our main result shows that a computationally attractive relaxation can be obtained by recasting the problem as a rank-constrained semi-definite optimization and using existing tools specifically tailored to this type of problems. These results are illustrated with a practical application drawn from computer vision**

I. INTRODUCTION AND MOTIVATION

Many processes of practical interest can be modeled as the output of a Wiener system, consisting of the cascade of a memoryless (possibly time varying) nonlinearity and a Linear Time Invariant (LTI) plant.

Roughly speaking, existing Wiener identification techniques can be classified into two broad categories: statistical (see for instance [25], [8], [10], [3], [4], [19] and references therein), and set membership approaches [6]. The latter are attractive since they furnish hard bounds on the values of the unknown parameters of the plant, in a form that can be directly used for instance by robust control synthesis techniques. However, as recently shown in [22], set membership identification of Wiener systems is generically NP hard. Hence, obtaining scalable, computationally tractable solutions requires the use of relaxations. In this paper, we propose a new relaxation based on the use of a combination of recently developed tools for rank minimization to reduce the problem to a convex semi-definite optimization.

Specifically, we consider several scenarios with varying degrees of complexity depending on the information available on the non-linearity (ranging from just a bound on its gain to a family of rational functions known to span it) and the linear part of the system (bounds on its degree, ℓ_2 induced norm). As we show in the paper, all of these cases can be reduced to the problem of minimizing the rank of a matrix whose entries depend polynomially on the measured data.

While in principle this is a very challenging problem, we show that tractable relaxations can be obtained by appealing to a combination of arguments from the classical theory of moments and convex analysis. A salient feature of this approach is that it can seamlessly handle missing data, due for instance to outages in a communication channel, sensor failure or occlusion. In the second part of the paper, we illustrate these results with some non-trivial examples arising in computer vision: extracting geometric 3D structure from a sequence of 2D images (structure from motion). As we will illustrate there, recasting the problem as a Wiener systems identification, leads to an algorithm that outperforms the current state–of–the–art in the field.

II. PRELIMINARIES

The notation used in the paper is summarized below:

$\overline{\sigma}(\mathbf{A})$	maximum singular value of matrix $\mathbf{A}$.
$\sigma_i \mathbf{A}$	i^{th} singular value of matrix $\mathbf{A}$.
$\mathbf{A} \geq 0$	$\mathbf{A}$ is positive semidefinite.
$(\mathcal{X}, m)$	metric space of elements in $\mathcal{X}$ equipped with the metric $m(x_1, x_2)$.
$d(\mathcal{A})$	diameter of $\mathcal{A} \subseteq \mathcal{X}$: $d(\mathcal{A}) \doteq \sup_{x,a \in \mathcal{A}} m(x, a)$.
$\overline{\mathcal{BX}}(\gamma)$	closed γ-ball in a normed space $\{\mathcal{X}, \|.\|\}$: $\overline{\mathcal{BX}}(\gamma) = \{x \in \mathcal{X} : \|x\|_{\mathcal{X}} \leq \gamma\}$.
ℓ^∞	space of vector valued sequences equipped with the norm: $\|x\|_\infty \doteq \sup_i \|x_i\|_\infty$.
$\mathcal{H}_{\infty,\rho}$	space of transfer functions analytic in $|z| \leq \rho$, equipped with the norm $\|G\|_{\infty,\rho} \doteq ess \sup_{|z|<\rho} \overline{\sigma}(G(z))$. The case $\rho = 1$ will be simply denoted $\mathcal{H}_\infty$.
$\mathcal{P}_N$	Projection operator: $\mathcal{P}_N[H(z)] = \sum_{i=0}^{N-1} \mathbf{h}_i z^i$
$\mathbf{H}_x^{m,n}$	Hankel matrix associated with a vector sequence $\mathbf{x}$:

$$\mathbf{H}_x^{m,n} \doteq \begin{bmatrix} \mathbf{x}_0 & \mathbf{x}_1 \cdots & \mathbf{x}_m \\ \mathbf{x}_1 & \mathbf{x}_2 \cdots & \mathbf{x}_{m+1} \\ \vdots & \vdots \ddots & \vdots \\ \mathbf{x}_n & \mathbf{x}_{n+1} \cdots & \mathbf{x}_{m+n} \end{bmatrix}$$

$\mathbf{T}_x$	lower triangular block Toeplitz matrix associates with any finite sequence $\{x_k, k = 0, 1, \cdots, n-1\}$

Given a Linear Time Invariant (LTI) system G, we will denote by $\mathbf{g}$ its impulse response sequence (Markov param-

eters) and by $\mathbf{T_g}$ the corresponding Toeplitz matrix. Finally, when dealing with finite sequences of length N, we will use $\mathbf{g}_N$ and $\mathbf{T_g}^N$ to denote the truncated sequence and the corresponding $N \times N$ upper left sub-matrix of $\mathbf{T_g}$.
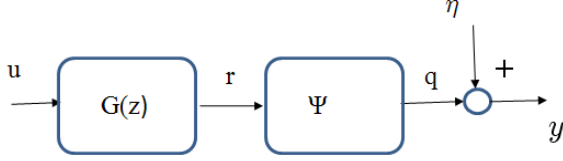
III. PROBLEM STATEMENT

Fig. 1. Wiener System Structure

Consider the Wiener system shown in Figure 1 consisting of the interconnection of a LTI system $G(z)$ and a memoryless nonlinearity $\Psi(.)$. The corresponding equations are:

$$\begin{aligned} y_k &= \Psi(r_k) + \eta_k \\ r_k &= (g * u)_k \end{aligned} \quad (1)$$

where $*$ denotes convolution and the signals $u \in R^{n_u}$ and $y \in R^{n_y}$ represent the experimental data: a known input and its corresponding output, corrupted by unknown but norm–bounded measurement noise η. Our goal is to, given experimental data consisting of N measurements of the input/output sequences $\{u_k, y_k\}_{k=0}^{N-1}$ and some *a priori* information about the plant, establish whether they are consistent, and if so, find a model that interpolates the experimental data within the measurement error level.

In the sequel, we will make the following standard assumptions about the *a priori* information:

A1.- A set description of $G(z)$ is available, e.g. $G(z) \in \mathcal{S}$, for some $\mathcal{S} \subseteq \mathcal{H}_\infty$, compact, convex.

A2.- A set description of the measurement noise is available: $\eta \in \mathcal{N}$, compact, convex.

A3.- $\Psi \in \mathcal{F}$, a family of equi-bounded, uniformly equi-continuous functions.

Remark 1: Assumptions [A1.-] and [A2.-] are standard [7]. Assumption [A3.-] is required in order to guarantee convergence of *any* interpolatory algorithm as the information is completed [24]. It is automatically satisfied in cases where the nonlinearity can be expressed as a bounded combination of a finite set of known, continuous basis functions:[a]

$$\begin{aligned} \mathcal{F} \doteq \{\Psi(.): \Psi(.) = \mathbf{B}\Phi, \ \Phi = [\phi_1(.), \ldots, \phi_{n_f}(.)]^T, \\ \phi_i(.) \text{ known}, \ \mathbf{B} \in \mathcal{B} \subset R^{n_y \times n_f} \} \end{aligned}$$

In principle the continuity assumption seems too strong, ruling out, among others, relay nonlinearities. However, note that these nonlinearities render all plants whose response to the given input has the same zero-crossings indistinguishable. Further, since the domain of the nonlinearity is restricted to

a compact subset of R^{n_r}, it can be approximated arbitrarily close by a continuously differentiable function. Indeed, it can be argued that such smooth models provide a better representation of physically realizable nonlinearities.

Under assumptions [A1–A3], the problem under consideration can be precisely stated as:

Problem 1: Given the *a priori* information $\mathcal{S}, \mathcal{N}, \mathcal{F}$ and the experimental data $\{y_k, u_k\}_{k=0}^{N-1}$:

1) determine whether the information is consistent, i.e., the consistency set $\mathcal{T}(y, N, \mathcal{N}) \doteq \{G \in \mathcal{S}: \ y_k = \Psi[(g*u)_k] + \eta_k, \ k = 0, \ldots, N-1 \text{ for some } \Psi \in \mathcal{F} \text{ and some sequence } \eta_k \in \mathcal{N}\}$ is nonempty.
2) If $\mathcal{T}(y, N, \mathcal{N}) \neq \emptyset$, find a nominal model $\{G, \Psi(.)\}$ that interpolates the data.

IV. IDENTIFICATION ALGORITHM

In this section we present the main result of the paper, a rank-constrained semi-definite optimization based identification algorithm. We will begin by considering a setup similar to the one standard in set-membership identification approaches and then indicate how to incorporate additional constraints to regularize the problem.

A. The Basic Setup

Consider first a setup similar to the one used in [24], with the following *a priori* information:

1.- The set $\mathcal{S}$ that characterizes the linear portion of the system is given by:
$\mathcal{S} \doteq \{G(z): G(z) \in \overline{\mathcal{BH}}_{\infty,\rho}(K)\}, \ \rho > 1, \ K$ given.
2.- The measurement noise satisfies: $\eta \in \mathcal{N} \doteq \{\eta: \|\eta_k\|_\infty \leq \epsilon\}$.
3.- The nonlinearity Ψ is a linear combination of a finite set of known rational functions, e.g. $\phi_i(r) = \frac{P_i(r)}{Q_i(r)}$, where $P_i(.), Q_i(.)$ are known multivariate polynomials.

Next, we recall a result from [24] that allows for recasting Problem 1 into a constrained optimization problem.

Lemma 1 ([24]): Given $K > 0$, $\rho > 1$, a nonlinear matrix function $\Psi(\cdot)$, and two vector sequences of experimental data $\mathbf{y} = [y_0, y_1, \cdots, y_{N-1}]^T$ and $\mathbf{u} = [u_0, u_1, \cdots, u_{N-1}]^T$, there exist a linear operator $\mathbf{G}(z) \in \overline{\mathcal{BH}}_{\infty,\rho}(K)$ and a nonlinear mapping $f(\cdot) = \mathbf{B}\Psi(\cdot)$ such that the consistency set $\mathcal{T}(\mathbf{y})$ is nonempty, if and only if there exist a vector $\mathbf{g}$ and a matrix $\mathbf{B}$ satisfying:

$$\mathbf{L}(\mathbf{g}) \doteq \begin{bmatrix} K\mathbf{R}^{-2} & (\mathbf{T}_g^N)^T \\ \mathbf{T}_g^N & K\mathbf{R}^2 \end{bmatrix} \geq 0 \quad (2)$$

$$\mathbf{r} = \mathbf{T}_g^N \mathbf{u} \quad (3)$$

$$[\mathbf{y} - \mathbf{B} \cdot \Psi(\mathbf{r})]_k \in \mathcal{N} \ \forall \ k \quad (4)$$

where $\mathbf{T}_g^N$ is the lower Toeplitz matrix associated with the sequence $\mathbf{g}$, $\mathbf{R} = diag[1, \rho, \rho^2, \cdots, \rho^{N-1}]$, and where, for a given vector sequence $\mathbf{v}$, $[\mathbf{v}]_k$ denotes its kth element.

---

[a]Since G has a finite ℓ^2 induced norm, for a given input, the signal r, and hence the domain of ϕ_i belong to a compact subset of R^{n_r}. Uniform equicontinuity and equiboundedness follow from continuity and finiteness of the family $\{\phi_i\}$.

Once feasibility of (2)–(4) is established, then a suitable $G(z)$ can be obtained proceeding as in [13][b].

B. A Rank Constrained SDP Relaxation

From Lemma 1 it follows that establishing consistency of the *a priori* information and the *a posteriori* experimental data reduces to checking feasibility of the set (2)–(4). Unfortunately, as recently shown in [22], this problem is NP hard both in the number of inputs to the nonlinearity Ψ and the number of experimental data pairs (u, y), even in cases where the nonlinearity $\Psi(.)$ is known. To circumvent this difficulty, in the sequel we develop a rank constrained semidefinite optimization based relaxation. The starting point is the fact that in the case where the nonlinearity is spanned by a rational basis $\Phi_i(r) = \frac{P_i(r)}{Q_i(r)}$, the feasibility set of (4) is semi-algebraic (e.g. defined by a finite set of polynomial equalities and inequalities). Indeed, simple algebra shows that the set of inequalities (4) is equivalent to:

$$\begin{aligned} q_{ik}Q_i(r_k) - b_i P_i(r_k) &= 0 \\ \|y_k - \sum_i q_{ik}\|_{\ell^\infty} &\leq \epsilon, \ \forall i, k \end{aligned} \quad (5)$$

which defines a semi–algebraic set for the unknowns q_{ik}, b_i, r_k. Next, note that the feasibility set of (2) is also semi-algebraic, since positive definiteness of a matrix can be checked by checking positivity of its leading principal minors. Clearly this is equivalent to a polynomial constraint in $\mathbf{g}$. It follows then that feasibility of (2)–(4) can be reduced to checking whether or not a semi-algebraic set is empty, a problem that can be solved using either results from semi-algebraic geometry [18], or from polynomial optimization (via the use of moments) [14]. A potential difficulty in pursuing these approaches arises from the high computational complexity entailed. For instance, when using moments, note that all elements of the sequence $\mathbf{g}$ are coupled through the inequality (2). Hence, using moments will require dealing with very large matrices (see [14] for details). Since the computational complexity of general SDP solvers is at least $\mathcal{O}(n^8)$ [20], where n is the number of decision variables, this approach becomes impractical even for moderately sized MIMO problems. Thus, in the sequel, we will pursue a different approach (although similar in spirit), where the number of variables is traded-off against rank constraints. To this effect, consider (5) for a fixed time instant k, and fixed basis function $\Phi_l(r)$. (For notational simplicity, we assume here scalar nonlinearities, but the approach extends trivially to the vector case). Assume that $deg(P_l) = t_l$ and $deg(Q_l) = s_l$ and, for each k, form the following matrices

$$\mathbf{M}_k \in R^{n_v \times n_v} \doteq \left(m_{ij}^{(k)}\right), \ m_{ij}^{(k)} = r_k^{(i+j-2)} \quad (6)$$

$$\mathbf{P}_{i,j}^{(k)} = \begin{bmatrix} 1 & m_{1,j}^{(k)} \\ b_i & \alpha_{i,j} \end{bmatrix} \text{ and } \mathbf{Q}_{i,j}^{(k)} = \begin{bmatrix} 1 & m_{1,j}^{(k)} \\ q_{i,j} & \beta_{i,j} \end{bmatrix} \quad (7)$$

with $d + 1 \geq 2n_v \geq d \doteq max(t_l, s_l)$. Here $\mathbf{M}_k, \mathbf{P}_{i,j}^{(k)}, \mathbf{Q}_{i,j}^{(k)}$ correspond to the truncated moments matrices of the joint

[b]Recall that all elements in the consistency set $\mathcal{T}(y, N, \mathcal{N})$ can be parameterized as a Linear Fractional Transformation of a free parameter $Q \in \mathcal{BH}_\infty$. The formulas in [13] provide a state–space realization for the central interpolant, e.g. the case $Q = 0$.

probability distribution of $\mathbf{r}_k, b, \mathbf{q}_k$. From the results in [14] it follows that one can obtain a hierarchy of converging relaxations by embedding these matrices into increasingly larger size ones. Further, due to the flat extension property, if (2)–(4) are feasible, then this expansion can be stopped at some finite horizon, with the resulting matrices having rank 1. However, typically pursuing this approach requires considering large horizons, and hence large optimization problems, even in cases where there are relatively few measurements. To avoid this difficulty, we will impose directly rank constraints, leading to the following result:

Theorem 1: There exists a feasible solution $\mathbf{r}, \mathbf{q}, \mathbf{g}$ to (2)–(4) if and only if there exist a feasible solution to the following rank-constrained problem:

$$\mathbf{L}(\mathbf{g}) \doteq \begin{bmatrix} K\mathbf{R}^{-2} & (\mathbf{T}_g^N)^T \\ \mathbf{T}_g^N & K\mathbf{R}^2 \end{bmatrix} \geq 0 \quad (8)$$

$$\mathbf{m}_{12}^{(k)} = [\mathbf{T}_g^N \mathbf{u}]_k \quad (9)$$

$$\text{rank}\,[\mathbf{M}_k)] \leq 1, \text{rank}\left[\mathbf{P}_{i,j}^{(k)}\right] \leq 1, \text{rank}\left[\mathbf{Q}_{i,j}^{(k)}\right] \leq 1 \quad (10)$$

$$\sum_{j=0}^{t_l} a_{i,j}\alpha_{i,j}^{(k)} - \sum_{j=0}^{s_l} c_{i,j}\beta_{i,j}^{(k)} = 0 \quad (11)$$

$$\|y_k - \sum_i q_{ik}\|_{\ell^\infty} \leq \epsilon, \quad (12)$$

Proof: (Sufficiency). Note that (10) is equivalent to $\alpha_{i,j}^{(k)} = b_i r_k^j$, $\beta_{i,j} = q_{ik} r_k^j$ and $m_{12}^{(k)} = r_k$. Hence (9)-(11) are equivalent to (3) and (5). Necessity follows immediately from the fact that if (2)–(4) admit a feasible solutions then the matrices $\mathbf{M}_k = \begin{bmatrix} 1 & r_k \ldots r_k^n \end{bmatrix}^T \begin{bmatrix} 1 & r_k \ldots r_k^n \end{bmatrix}$, $\mathbf{P}_{i,j}^{(k)} = \begin{bmatrix} 1 & r_k^j \\ b_i & \alpha_{i,j} \end{bmatrix}$ and $\mathbf{Q}_{i,j}^{(k)} = \begin{bmatrix} 1 & r_k^j \\ q_{i,j} & \beta_{i,j} \end{bmatrix}$ satisfy (10) ∎

From the result above it follows that the Wiener identification problem can be solved by solving the rank–constrained SDP defined by (8)–(12), which can be accomplished for instance by using the convex relaxation proposed in [17]. It is worth noting that missing data can be handled by simply not enforcing (4) for those time instants where data is unavailable.

C. Finding Minimum Rank Interpolants

Recall (see Theorem 1 in [24]) that in the case of Wiener systems, interpolatory algorithms[c] are no longer guaranteed to converge to the true plant as the information is completed, that is $\epsilon \to 0$ and $N \to \infty$, in the sense that $\mathcal{T}(y, N, \mathcal{N}) \to \{g^o\}$ is no longer guaranteed (except in special cases, e.g. invertible nonlinearities). Rather, any interpolatory algorithm $\mathcal{A}^I$ will converge to the set

$$\mathcal{T}^*(y) \doteq \quad \{G \in \mathcal{S} : y_k = \Psi(g * u)_k, \ k = 0, 1, \ldots, \\ \text{for some } \Psi \in \mathcal{F}\}$$

e.g. the consistency set in case of complete and uncorrupted experimental information. The diameter of this set $e^*(y) \doteq d\{\mathcal{T}(y)\}$ defines an *intrinsic local* worst-case error for identification of Wiener systems, in the sense that this is the best

[c]In the context of this paper, interpolatory algorithms are those such that the linear portion of the *true* plant $g^o \in \mathcal{T}(y, N, \mathcal{N})$.

that can be achieved by any interpolatory algorithm. Under these circumstances, it is reasonable to exploit additional *a priori* information to remove the ambiguities in the problem formulation. In particular, in many cases of practical interest, it is known that the lowest order model that explains the observed information is the correct one (an example of this situation arises in the problem of recovering 3D geometric information from 2D video data [23]). In principle, model order constraints can be handled by simply placing a constraint on the rank of the Hankel matrix associated with the (truncated) impulse response $\mathbf{g}_{i=1}^N$ of the plant $G(z)$. Note however that this does not guarantee that when the complete interpolant is formed, its rank will equal that of the truncated Hankel matrix. As we show next, this difficulty can be circumvented by considering an augmented optimization problem. To this effect, we introduce next the following result relating the singular values of an infinite Hankel matrix to those of its finite $n \times n$ upper left submatrix.

Theorem 2: Consider a transfer matrix $\mathbf{G}(z) \in \overline{\mathcal{BH}}_{\infty,\rho}(K)$ with $\rho < 1$, and let $\mathbf{H}_g$ and $\mathbf{H}_g^{n,n}$ denote the associated infinite Hankel matrix and its $(n+1) \times (n+1)$ upper left submatrix, respectively, e.g.

$$
\mathbf{H}_g \doteq \begin{bmatrix} \mathbf{g}_0 & \mathbf{g}_1 \cdots & \mathbf{g}_n & \mathbf{g}_{n+1} & \cdots \\ \mathbf{g}_1 & \mathbf{g}_2 \cdots & \mathbf{g}_{n+1} & \mathbf{g}_{n+2} & \cdots \\ \vdots & \vdots \ddots & \vdots & \vdots & \vdots \\ \mathbf{g}_n & \mathbf{g}_{n+1} \cdots & \mathbf{g}_{2n} & \mathbf{g}_{2n+1} & \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}
$$

$$
= \left[\begin{array}{c|c} \mathbf{H}_g^{n,n} & \\ \hline \mathbf{M_g} & \mathbf{H}_{\mathbf{g}_{tail}} \end{array} \right]
$$

(13)

where

$$
\mathbf{M}_g \doteq \begin{bmatrix} \mathbf{g}_{n+1} & \mathbf{g}_{n+2} \cdots & \mathbf{g}_{2n+1} \\ \mathbf{g}_{n+2} & \mathbf{g}_{n+3} \cdots & \mathbf{g}_{2n+2} \\ \vdots & \vdots \cdots & \vdots \end{bmatrix}
$$

and where $\mathbf{H}_{\mathbf{g}_{tail}}$ is the Hankel matrix associated with the transfer matrix $\mathbf{G}_{tail}(z) = z^{(n+1)}\mathbf{g}_{n+1} + z^{(n+2)}\mathbf{g}_{n+2} + \cdots$. Then, given any $\epsilon > 0$, there exist $N(\epsilon)$ such that for all $n \geq N(\epsilon)$, $\sigma_i(\mathbf{H}) \leq \sigma_i(\mathbf{H}_\mathbf{g}^{n,n}) + \epsilon$, i=1,..n.

Proof: Since

$$
\mathbf{H}_g = \left[\begin{array}{c|c} \mathbf{H}_g^{n,n} & 0 \\ \hline 0 & \end{array} \right] + \left[\begin{array}{c|c} 0 & \\ \hline 0 & \mathbf{H}_{\mathbf{g}_{tail}} \end{array} \right]
$$

$$
+ \left[\begin{array}{c|c} 0 & 0 \\ \hline \mathbf{M_g} & \end{array} \right]
$$

From Weyl's inequality it follows that

$$
\begin{aligned} \sigma_i(\mathbf{H_g}) &\leq \sigma_i(\mathbf{H_g}^{n,n}) + \sigma_1(\mathbf{M_g}) + \sigma_1(\mathbf{H}_{\mathbf{g}_{tail}}) \\ &\leq \sigma_i(\mathbf{H_g}^{n,n}) + 2\sigma_1(\mathbf{H}_{\mathbf{g}_{tail}}) \end{aligned}
$$

where we used the fact that $\mathbf{M_g}$ is a submatrix of $\mathbf{H}_{\mathbf{g}_{tail}}$. Next, recall from Theorem 2 in [5] that, for any transfer function $G(z)$, $\sigma_1(\mathbf{H_g}) \leq \|\mathbf{g}\|_1$. Applying this result to G_{tail} and using the fact that $G \in \overline{\mathcal{BH}}_{\infty,\rho}(K) \Rightarrow |g_i| \leq K\rho^{-i}$ yields

$$
\sigma_i(\mathbf{H_g}) \leq \sigma_i(\mathbf{H_g}^{n,n}) + 2K\frac{1}{\rho^{n+1}(\rho-1)}
$$

Since $\rho > 1$, the second term can be made arbitrarily small by selecting n large enough. ∎

Corollary 1: Given $\epsilon > 0$, select n so that

$$
2K\frac{1}{\rho^{n+1}(\rho-1)} < \epsilon
$$

Define $\mathbf{g}_{aug}^T \doteq \begin{bmatrix} g_0 & \cdots & g_N & g_{N+1}\cdots & g_n \end{bmatrix}^T$ and consider the following feasibility problem:

$$
\mathbf{L}(g_{aug}) \doteq \begin{bmatrix} K\mathbf{R}^{-2} & \left(\mathbf{T}_{g_{aug}}^n\right)^T \\ \mathbf{T}_{g_{aug}}^n & K\mathbf{R}^2 \end{bmatrix} \geq 0 \quad (14)
$$

$$
\left. \begin{aligned} [r - \mathbf{T}_g^N\mathbf{u}]_k &= 0 \\ [\mathbf{y} - \mathbf{B}\cdot\mathbf{\Psi}(r)]_k &\in \mathcal{N} \end{aligned} \right\} \forall\, k = 1, 2, \ldots N \quad (15)
$$

$$
\text{rank}(H_{g_{aug}}^{n,n}) \leq n_{red} \quad (16)
$$

If this problem admits a feasible solution $\mathbf{g}_{aug}$, then the Hankel singular values of the associated interpolant $G_{aug}(z)$ satisfy $\sigma_i^H(G_{aug}) \leq \epsilon$ for all $i \geq n$.

Combining the results of the corollary above with standard model reduction arguments (see for instance Chapter 9 in [21]), it follows that the linear portion of the plant identified by solving the feasibility problem above can be approximated by a transfer function with McMillan degree n_{red}.

D. Further extensions

In this section we briefly indicate how to extend the proposed identification algorithm to cases where the nonlinearity is time varying. In particular, we will consider two cases: (1) arbitrarily fast time-varying sector bound nonlinearities; and (2) switching nonlinearities, where the nonlinearity switches (at unknown times) between the elements of a given set.

Arbitrarily fast time-varying, sector bounded nonlinearity. In this case, the only information available about the non-linearity is a sector bound of the form $(q_k - \gamma_1 r_k)(q_k - \gamma_2 r_k) \leq 0$, where, as before, r_k and q_k denote the input and output to the nonlinearity, respectively. Since the constraint above is polynomial in r_k, q_k it can be directly incorporated to the set (2)–(4), with this last inequality replaced by

$$
y_k - r_k \in \mathcal{N}\ \forall\, k
$$

Switching nonlinearities Next, we consider the case of a switching Wiener system, where the nonlinearity switches randomly, at unknown times, amongst those on a given, known set. As we show next, this case can be also handled by our framework, with minimal modifications. For notational simplicity, we will assume that the system switches between two nonlinearities $\Psi_1(\xi) = \frac{P_1(\xi)}{Q_1(\xi)}$ and $\Psi_2(\xi) = \frac{P_2(\xi)}{Q_2(\xi)}$, but the argument is general. Since at any given time instant either Ψ_1 or Ψ_2 is active, then the following constraint must hold at all times:

$$
\begin{aligned} 0 &= (q_k - \Psi_1(r_k))(q_k - \Psi_2(r_k)) \iff \\ 0 &= q_k^2 Q_1(r_k)Q_2(r_k) - q_k(P_1(r_k)Q_2(r_k) \\ &+ Q_1(r_k)P_2(r_k)) + P_1(r_k)P_2(r_k) \end{aligned} \quad (17)
$$

Clearly, the constraint above can be incorporated to the feasibility problem (2)–(4) as an affine constraint by considering the appropriate elements of the matrices $\mathbf{M}_k$, $\mathbf{P}_{i,j}^{(k)}$ and $\mathbf{Q}_{i,j}^{(k)}$.

V. COMPUTING BOUNDS OF THE IDENTIFICATION ERROR

In this section we briefly address the problem of computing worst case bounds on the identification error. For a given set of experimental measurements $\mathbf{y}$, define the worst case identification error as:

$$e(\mathbf{y}, N, \mathcal{N}) \doteq \sup_{G_1, G_2 \in \mathcal{T}(\mathbf{y}, N, \mathcal{N})} \|G_1 - G_2\|_* \qquad (18)$$

where $\|.\|_*$ denotes a suitable norm. In the case of *linear* plants the quantity above can be further related to the diameter of information and, depending on the characterization of the set $\mathcal{T}$ and the choice of norm, computed by solving a linear programming problem. On the other hand, due to the presence of the non-linearity, this is no longer possible in the case of Wiener systems. Indeed, in this case, computing even an upper bound of the identification error is in principle a very challenging problem. Nevertheless, as we show in the sequel, the same ideas used to obtain tractable relaxations of the Wiener identification problem, can be exploited to compute upper bounds on $e(\mathbf{y}, N, \mathcal{N})$ by solving a semi-definite optimization problem. Start by noting that for all $G_1, G_2 \in \mathcal{T}(\mathbf{y})$ the following holds:

$$
\begin{aligned}
\|G_1 - G_2\|_\infty &\leq \|\mathcal{P}_N[G_1 - G_2]\|_\infty + \\
&+ \|(I - \mathcal{P}_N)[G_1 - G_2]\|_\infty \\
&\leq \sum_{i=0}^{N} |g_{1i} - g_{2i}| + \frac{2K}{\rho^{N+1}(\rho - 1)}
\end{aligned}
\qquad (19)
$$

Thus, $e(\mathbf{y}, N, \mathcal{N}) \leq K_1 + \frac{2K}{\rho^{N+1}(\rho-1)}$, where

$$K_1 \doteq \sup_{G_1, G_2 \in \mathcal{T}(\mathbf{y}, N, \mathcal{N})} \sum_{i=0}^{N} |g_{1i} - g_{2i}|$$

Next, note that K_1 can be computed from:

$$K_1 = \max_{\mathbf{g}_1, \mathbf{g}_2, \mathbf{x}^+, \mathbf{x}^-} \sum_{i=1}^{N} x_i^+ + x_i^- \qquad (20)$$

$$\text{subject to (2)–(4)} \quad x_i^+ \geq 0, x_i^- \geq 0$$

$$g_{1i} - g_{2i} = x_i^+ - x_i^- \qquad (21)$$

$$x_i^+ x_i^- = 0 \qquad (22)$$

This amounts to solving an identification problem of the form discussed earlier in the paper, with additional constraints (one linear, one quadratic) on the nonlinearity. Thus, the same tools applied earlier can also be used here.

VI. ILLUSTRATIVE EXAMPLES

In this section we illustrate the proposed algorithm using both academic examples and a non-trivial, computer vision motivated application.

A. Minimum Rank Interpolants

This example illustrates the regularizing effect of restricting the order of the linear portion of the plant. The experimental data consists of $n = 10$ samples of the impulse response of the system

$$G(z) = \frac{0.2}{z^2 - 1.5z + 0.765}, \quad \Phi(r) = r^2 \qquad (23)$$

with tight *a priori* constraints; $K = 1.6540$, $\rho = 1.1433$. Figure 2 compares the identification results obtained with and without enforcing a rank constraint (rank$[\mathbf{H_g}] \leq 2$) on the interpolant. Note that, due to the nonlinearity $q = r^2$, without additional priors, if a plant $G_i(z)$ explains the data, then so does the plant $-G_i(z)$, from where it follows that the radius of information (and hence the minimally achievable identification error) is at least $\|G\|$. In addition, in this case, it is not possible to shape an input that would drive the output to an invertible interval of the nonlinearity, since there is no such interval for $\Phi(r) = r^2$ (See [6] for input shaping on the problem of polynomial nonlinearity). On the other hand, imposing the rank constraint leads to the correct description.
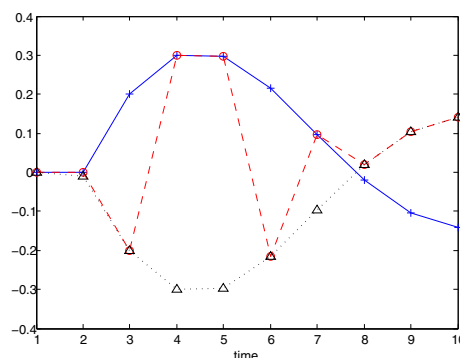

Fig. 2. (+) True impulse response, ($\triangle$) Estimated with Hankel Rank Constraint, (o) Estimated without Hankel Rank Constraint

B. Identification with missing data

This example illustrates the ability of the method to handle missing data and still obtain plants with good predictive power. The experimental data consists of the first eight measurements of the sequence obtained by filtering the impulse response of the linear plant used in the previous example through the following rational nonlinearity

$$r_k = (g * u)_k, \quad k = 0, 1, \dots$$

$$y_k = b_1 \frac{6r_k}{r_k^2 + 1} + b_2 \frac{r_k + 4}{r_k^2 + 3} + \eta_k \qquad (24)$$

where η, with $|\eta_k| \leq 0.10 \max_k |y_k| = 0.44$ denotes measurement noise. We further assumed that the measurements at $n = 6$ and $n = 7$ were missing (e.g. a gap of 25% in the data). The available *a priori* information is $K = 1.6540$, $\rho = 1.1433$, $b_1 \in [0.8, 1.2]$ and $b_2 \in [1.6, 2.4]$. The set membership description of the coefficients b_i amounts to 20 % uncertainty around their nominal values. As shown in

Figure 3, the proposed algorithm is capable of reconstructing the missing data and the resulting plant is capable of interpolating future data within the noise level.
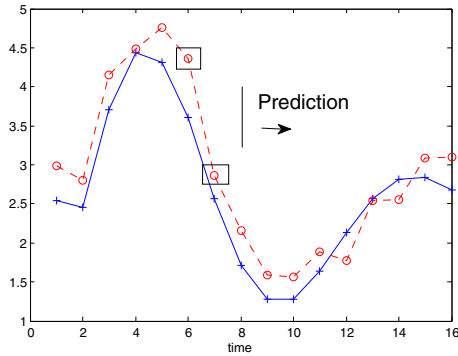

Fig. 3. Identification with missing data, (o) true signal, (+) estimated signal. Two outputs shown in rectangles missing

C. Switching Nonlinearities

This example illustrates the ability of our framework to handle switching nonlinearities. The data consists of $n = 10$ elements of the impulse response of the linear plant used in the previous examples, filtered through a nonlinearity that switches randomly between $\Psi_1(r) = \frac{6r}{r^2+1}$ and $\Psi_2(r) = \frac{r+4}{r^2+3}$ and corrupted by 10% measurement noise. As shown in Fig 4, our algorithm is capable of correctly identifying both a system that interpolates the experimental data and the nonlinearity that is active at each time instant.
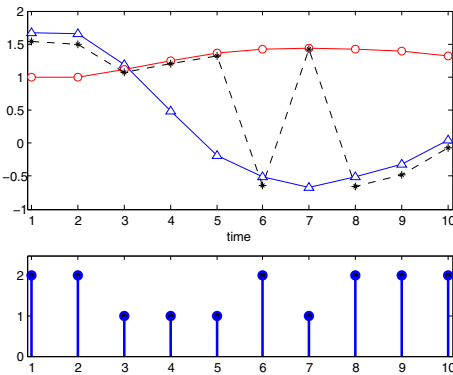

Fig. 4. Identification with a switching nonlinearity: Top ($\triangle$) Output of Ψ_1, (o) Output of Ψ_2, (+) Measured Output. Bottom: Active nonlinearity, as identified by the algorithm.

VII. APPLICATION: 3D-STRUCTURE FROM MOTION USING UNCALIBRATED CAMERAS

In this section we illustrate the potential of the proposed approach to handle realistic problems by applying it to a nontrivial computer vision problem, 3D structure from motion, where the goal is to recover the 3D geometry of a scene, using 2 dimensional data generated by a moving perspective camera. A model of the system generating the data is given by (see for instance [16]):

$$\zeta_{k+1} \doteq \mathbf{A}\zeta + \mathbf{B}e_k \qquad (25)$$

$$\mathbf{P}_{kj} \doteq \begin{bmatrix} X_{kj} \\ Y_{kj} \\ Z_{kj} \end{bmatrix} = \mathbf{C}_j \zeta_k \qquad (26)$$

$$u_{kj} = f\frac{X_{kj}}{Z_{kj}} + c_u, \ v_{kj} = f\frac{Y_{kj}}{Z_{kj}} + c_v \qquad (27)$$

where f and (c_u, c_v) denote the focal length f and principal points of the camera, respectively, and where the pair $(\mathbf{A}, \mathbf{B})$ and the associated state vector ζ_k model the 3D motion of the camera in response to the input e^{d}, $\mathbf{P}_{kj}$, $j = 1, \ldots, N$, denotes the 3D coordinates of point P_j at time k, $i = 1, \ldots, F$, and (u_{kj}, v_{kj}) denotes the 2D coordinates of the image of $\mathbf{P}_{kj}$. Clearly, the model above is a special case of a Wiener system, where the nonlinearity is rational. In this context, the problem of reconstructing the 3D geometry of the scene can be formalized as a nonlinear filtering one: estimating $\mathbf{P}_{kj}$ from (u_{kj}, v_{kj}), potentially corrupted by measurement noise. In principle, this nonlinear filtering problem can be solved using the techniques proposed in [16], [11], [1], [9]. However, proceeding in this fashion requires first identifying the Wiener system. Alternatively, several approaches have been developed in the computer vision community, based on a combination of matrix factorizations and non-linear optimization ([15], [12] and references therein). However, these approaches cannot guarantee convergence to the correct 3D geometry. In the sequel, we show that these difficulties can be circumvented by recasting the problem into a Wiener Systems identification form. Indeed, direct application of Theorem 1 in [2] shows that, in the case of *uncalibrated* cameras, the correct 3D geometry can be found by simultaneously identifying the *lowest* order dynamical system of the form (25)-(26) and a corresponding non-linearity of the form (27) such that the resulting system interpolates the 2D measurements, subject to an additional "rigidity"–type constraint on the trajectories. Specifically, given 4 points from the rigid, their 3D distances must remain constant along trajectories, that is:

$$(X_{n_1,i} - X_{n_1,j})^2 + (Y_{n_1,i} - Y_{n_1,j})^2 + \cdots$$
$$(Z_{n_1,i} - Z_{n_1,j})^2 - \cdots$$
$$-(X_{n_2,i} - X_{n_2,j})^2 - (Y_{n_2,j} - Y_{n_2,j})^2 - \cdots$$
$$(Z_{n_2,i} - Z_{n_2,i})^2 = 0$$

$$\forall n_1 \neq n_2 = 1, 2 \ldots N_f \text{ and for all } 1 \leq i \leq j \leq 4. \qquad (28)$$

Clearly, since the constraints above are polynomial, and (27) defines a family of known rational nonlinearities with unknown parameters (f, c_u, c_v), this problem fits the formalism developed in section IV-C (minimum rank interpolants). Figure 5 shows the results of applying these ideas to an artificially generated motion sequence of the Utah Teapot.

---

[d]Typically ζ_k contains the past values of the camera position, velocity, etc, with respect to a fixed frame
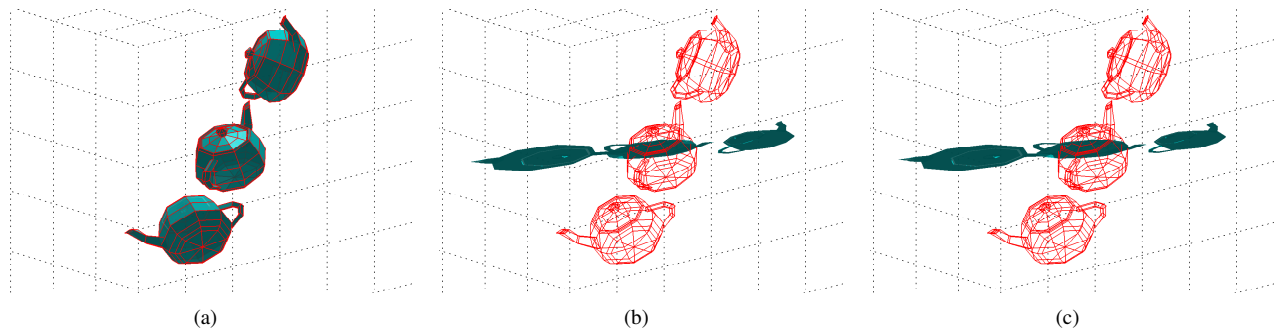
Fig. 5. Teapot 3D ground truth (red mesh) and reconstruction from 2D images (solid surface). (a) Wiener ID based. (b) Hung and Tang method [12] (c) Mahamud and Hebert method [15].

The actual camera parameters used in generating these frames were $f = 1$, $c_x = 0.5$ and $c_y = 0.5$, but during the identification we assumed that these values were only known to within 50% uncertainty level, that is the *a priori* information was $0.5 \leq f \leq 1.5$ and $0.25 \leq c_x, c_y \leq 0.75$. This *a priori* information is consistent with the fact that with current EXIF tag technology, some knowledge about the focal length of the camera is available, and it is known that the cameras' center of projection are usually close to the image center. As illustrated in Figure 5, the proposed approach is able to perfectly recover the 3D geometry of the scene, while current state–of–the–art methods (even assuming calibrated cameras) fail to do so.

VIII. CONCLUSIONS

In this paper we propose an algorithm for set membership identification of Wiener systems using time–domain data. As shown in the paper, this problem is equivalent to establishing feasibility of a semi-algebraic set. In turn, this can be reduced to a semi-definite optimization subject to a rank constraint, a problem that can be solved using recently introduced rank relaxations such as [17]. The proposed framework can handle missing data, arbitrarily fast time–varying nonlinearities and cases where the nonlinearity switches (at unknown times) amongst the elements of a given set. These results were illustrated with both academic examples and a non-trivial application arising in the context of computer vision.

REFERENCES

[1] R. Abdursul, H. Inaba, and B. K. Ghosh. Nonlinear observers for perspective time-invariant linear systems. *Automatica*, 40(3):481 – 490, 2004.
[2] M. Ayazoglu and M. Sznaier O. Camps. Euclidean structure recovery from motion in perspective image sequences via hankel rank minimization. In *Proc 2010 ECCV*, pp. 71–84, 2010.
[3] E. W. Bai. A blind approach to the hammerstein-wiener model identification. *Automatica*, 38:967–979, 2002.
[4] E. W. Bai. Frequency domain identification of wiener models. *Automatica*, 39:1521–1530, 2003.
[5] S. Boyd and J. Doyle. Comparison of peak and rms gains for discrete-time systems. *Systems and Control Letters*, (9):1–6, 1987.
[6] V. Cerone and D. Regruto. Parameter bounds evaluation of wiener models with noninvertible polynomial nonlinearities. *Automatica*, 42:1775–1781, 2006.
[7] J. Chen and G. Gu. *Control Oriented System Identification, An $\mathcal{H}_\infty$ Approach*. John Wiley, New York, 2000.

[8] C.T. Chou, B.R.J. Haverkamp, and M. Verhaegen. Linear and nonlinear system identification using separable least squares. *Eur. J. Control*, pages 116–128, 1999.
[9] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. Monoslam: Real time single camera slam. *IEEE Trans. on PAMI*, 29(6):1052–1067, 2007.
[10] W. Greblicki. Nonparametric approach to wiener system identification. *IEEE Trans. CAS*, 44:538–545, 1997.
[11] J. Hespanha. State estimation and control for systems with perspective outputs. In *Proc. of the 41st CDC.*, Dec. 2002.
[12] Y.S. Hung and W.K. Tang. Projective reconstruction from multiple views with minimization of 2d reprojection error. *IJCV*, 66(3):305–317, 2006.
[13] T. Inanc, M. Sznaier, P. A. Parrilo, and R. S. Sanchez Pena. Robust identification with mixed time/frequency domain experiments and parametric/nonparametric models: Theory and an application. *IEEE Trans. Contr. Syst. Tech*, 9(4):608–617, 2001.
[14] J. B. Lasserre. *Moments, positive polynomials and their applications*, volume 1. Imperial College Press, 2010.
[15] S. Mahamud and M. Hebert. Iterative projective reconstruction from multiple views. In *IEEE CVPR*, vol. 2, pp. 430–437, 2000.
[16] A. Matveev, X. Hu, R. Frezza, and H. Rehbinder. Observers for systems with implicit output. *IEEE Trans. Aut. Contr.*, 45:168–173, 2000.
[17] R. Orsi. LMIRank: software for rank constrained lmi problems. http://rsise.anu.edu.au/robert/lmirank/, 2005.
[18] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming Ser. B*, 96(2):293–320, 2003.
[19] R. Raich, T. Zhou, and M. Viberg. Subspace based approaches for wiener system identification. *IEEE Trans. Aut. Contr.*, 50:1629–1634, 2005.
[20] T. Roh, B. Dumitrescu, and L. Vandenberghe. Interior-point algorithms for sum-of-squares optimization of multidimensional trigonometric polynomials. In *Proc ICASSP 2007*, pages III–905–III–908.
[21] R. Sánchez Peña and M. Sznaier. *Robust Systems Theory and Applications*. Wiley & Sons, Inc., 1998.
[22] M. Sznaier. Computational complexity of set membership Hammerstein and Wiener systems identification. *Automatica*, 45(3):701–705, 2009.
[23] M. Sznaier, M. Ayazoglu, and O. Camps. Using dynamics to recover 3dimensional euclidian structure from 2dimensional perspective projections. In *Proc 2009 IEEE CDC*, pages 2414–2419, Dec. 2009.
[24] M. Sznaier, W. Ma, O. Camps, and H. Lim. Risk adjusted set membership identification of Wiener systems. *IEEE Trans. Aut. Contr.*, 54(5):1147 – 1152, 2009.
[25] D. Westwick and M. Verhaegen. Identifying mimo wiener systems using subspace model identification methods. *Signal Processing*, 52:235–258, 1996.