Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

An Algebraic Approach to Efficient Identification of a Class of Wiener Systems ⋆

B. Ozbay* **M. Sznaier*** **O. Camps ***

* *Electrical and Computer Engineering,*
Northeastern University, Boston, MA 02115.
emails ozbay.b@husky.neu.edu, {msznaier,camps}@coe.neu.edu

Abstract: This paper considers the problem of identifying the linear portion of a Wiener system, for the case of a known, but non-invertible non-linearity. It is well known that this scenario, common in many practical applications, leads to NP-hard problems in the number of experiments. Thus, existing techniques scale poorly and are typically limited to relatively few points. We show that this difficulty can be circumvented by considering an algebraic motivated approach. Specifically, we show that it is equivalent to identification of a switched linear system generated from the data. In turn, we can solve this problem by recasting it as the problem of finding the vanishing ideal of an arrangement of subspaces, a task that reduces to finding the null space of an embedded data matrix constructed from observed data.

Keywords: Wiener Systems Identification, Nonlinear Systems Identification, Factorization Based Identification.

1. INTRODUCTION

Wiener systems, consisting of the cascade of a linear time invariant (LTI) system and a static output nonlinearity are ubiquitous in many application domains ranging from biomedical (Da Silva et al., 2012; Song et al., 2015), chemical (Norquay et al., 1998), mechanical (Khan and Vyas, 2001) and control systems (Wigren, 1994), to communications and computer vision (Ayazoglu et al., 2010). In addition, they can provide (local) approximations to more general nonlinear dynamics (Boyd and Chua, 1985; Giri and Bai, 2010).

Given their importance, a large research effort has been devoted during the past decade to identification of Wiener systems. Frequency domain based methods have been proposed in Giri et al. (2009); Bai (2003); Tan (2013); Giri et al. (2013). However, these methods rely on a finite impulse response (FIR) approximation of the linear dynamics, or on a specially designed input excitation. A potential difficulty here is that FIR models are inefficient at representing inherently dynamic behavior such as resonances. Alternative approaches include methods exploiting specially designed inputs (Tiels and Schoukens, 2014; Zhang et al., 2015; Ren and Li, 2015; De Angelis et al., 2017; Bottegal et al., 2018). While these methods work well, it may not be feasible to implement the required input signals in actual experimental settings. Finally, set membership (or control oriented identification) of Wiener systems has been addressed in Sznaier et al. (2009) pursuing a risk-adjusted approach and in Ayazoglu et al. (2010), recasting the problem into a polynomial optimization form. However, the entailed computational complexity is non-trivial. Yılmaz and Sznaier (2015) improves the time complexity by considering a set inversion of the output data and exploiting the properties of interval matrices to prune the set of candidate points. While this approach results in substantial computational time reduction, its complexity grows exponentially with the identification horizon, limiting its applicability to relatively short experimental records.

Motivated by the difficulties noted above, in this paper we seek to develop a computationally efficient, scalable algorithm for identifying the linear portion of a Wiener system, in scenarios where the non-linearity is known, but non-invertible. Note that in many applications (e.g. biology) the nonlinearity is dictated by the underlying physics and thus known. Additional situations where this scenario arises include, among others, cases where the data is collected using sensors insensitive to sign such as power sensors, anemometers or pulse counters, and reconstruction of the 3D geometry from 2D images (Ayazoglu et al., 2010). It is worth noting that, as shown in Sznaier (2009), identifying the linear portion of the Wiener system in cases involving *known* even nonlinearities (and thus non-invertible), is generically NP hard, so the scenario considered here is both practically relevant and far from trivial.

The key observation motivating this paper is the fact that, for the case considered here, the one-to-many set inversion of the nonlinearity leads to a collection of points that belong to an arrangement of subspaces, obtained by considering all possible sign combinations of the parameters of the ARX model. This observation allows for recasting the Wiener identification problem as the identification of the parameters of a switched ARX system (where each of the subsystems corresponds to one sign combination). In turn, exploiting the algebraic approach proposed in Vidal et al. (2005) allows for finding the absolute value of these parameters from the coefficients of the so called *hybrid decoupling polynomial*, associated with the null space of an embedded data matrix constructed directly from the observed data. Once these values are available, the correct signs for the internal variables can be determined by minimizing the residual error over all sign combinations. Finally, a refined estimate of the ARX model parameters can be obtained from the null space of the Hankel matrix constructed from the estimated internal

⋆ This work was partially supported by NSF grants CNS–1646121, CMMI–1638234, IIS–1814631, ECCS1808381; AFOSR grant FA9550-19-1-0005 and the Alert DHS Center of Excellence under Award 2013-ST-061-ED0001.

variables. An important point worth noting is that we explicitly exploit the structure of the non-linearity and its inverse mapping to mitigate the growth of the embedded data matrix (and hence computational complexity) with the order of the system.

This paper is organized as follows. Section 2 introduces notation, background results on the Generalized Principal Component Analysis (GPCA) method, and formally states the addressed Wiener system identification problem. Section 3 shows that it can be recast as a switched systems identification problem and solved via algebraic methods. The specific algorithm is discussed in Section 4. Section 5 illustrates the effectiveness of the proposed method as a stand-alone method and when combined with existing methods to improve robustness to noise. Section 6 provides concluding remarks and future directions.

2. PRELIMINARIES

2.1 Notation

$\mathbf{A}_i$	i^{th} column of $\mathbf{A}$.
$\mathscr{P}^n_{d,h}$	subspace of n^{th} degree homogeneous multivariate polynomials in d variables.
$s_{n,d} \doteq \binom{n+d-1}{d}$	number of monomials of degree n in d variables.
$\mathbf{v}_n(\mathbf{x})$	Veronese map of degree n: $\mathbf{v}_n(x_1 \ \dots \ x_d) \doteq \left[x_1^n \ x_1^{n-1}x_2 \ \dots \ x_d^n\right]^T$
$\mathscr{E}_\mu(x)$	Expected value of x with respect to the probability density function μ.
$\mathbf{H}^{r,c}_x$	Hankel matrix with c columns and r rows associated with a vector sequence $\mathbf{x}$, with elements $(\mathbf{H}_x)_{i,j} = \mathbf{x}_{i+j-1}$
$\circ$	Hadamard product of matrices where $\mathbf{M} = \mathbf{X} \circ \mathbf{Y}$ has entries $m_{ij} = x_{ij}y_{ij}$

2.2 Subspace Arrangements

In this section we briefly recall background results on algebraic approaches to subspace clustering and switched systems identification (Ma et al., 2008; Ozay et al., 2010; Vidal et al., 2005).

Definition 1. The arrangement $\mathscr{A}(S)$ of a set of subspaces $S = \{S_i\}_{i=1}^n \subseteq R^n$ is defined as:

$$\mathscr{A}(S) \doteq S_1 \cup S_2 \cup \ldots \cup S_n \tag{1}$$

Definition 2. The vanishing ideal $I(\mathscr{A})$ of a subspace arrangement $\mathscr{A} \subseteq R^d$ is the set of all multivariate polynomials in d variables that vanish on all points in $\mathscr{A}$, that is:

$$I(\mathscr{A}) \doteq \left\{P \in \mathscr{P}^d \colon P(\mathbf{x}) = 0 \ \forall \mathbf{x} \in \mathscr{A}\right\} \tag{2}$$

The subset $I_n(\mathscr{A}) \subseteq I(\mathscr{A})$ formed by homogeneous polynomials is known as n^{th} degree homogeneous component of $I(\mathscr{A})$.

Definition 3. Given a set I of polynomials, $\mathscr{Z}(I)$, the zero set of I is the set of all common roots, that is

$$\mathscr{Z}(I) \doteq \{\mathbf{x} \in \mathbb{R}^d \colon P(\mathbf{x}) = 0 \text{ for all } P \in I\} \tag{3}$$

The following result shows that the arrangement $\mathscr{A}$ is completely characterized by its associated homogeneous ideal:

Lemma 1. (Lemma 2.8, Ma et al. (2008)). The subspace arrangement $\mathscr{A}$ is the zero set of $I_n(\mathscr{A})$, e.g. $\mathscr{A} = \mathscr{Z}\left[I_n(\mathscr{A})\right]$.

2.3 Algebraic Approach to Subspace Clustering

The results from the previous section form the basis of algebraic based approaches for finding the normals, $\mathbf{b}_i$, $i = 1, \ldots, n$, to the subspaces, S_i, in an arrangement $\mathscr{A}$. Rather than directly estimating $\mathbf{b}_i$, these approaches seek to estimate first I_n from data. Under mild conditions, I_n has dimension 1 (e.g. it is a principal ideal) and the parameters of each subspace can then be estimated from its generator, for instance via polynomial differentiation. Since each point in $\mathscr{A}$ satisfies $\mathbf{b}_i^T\mathbf{x} = 0$ for some i, it follows (Vidal et al., 2005) that all points in the arrangement $\mathscr{A}$ satisfy the *hybrid decoupling constraint*:

$$P_{dc}(\mathbf{x}) = \prod_{i=1}^{n}(\mathbf{b_i}^T\mathbf{x}) \doteq \mathbf{c_n}^T\nu_n(\mathbf{x}_t) = 0 \tag{4}$$

where $P_{dc}(\mathbf{x})$ is an n^{th} order homogeneous polynomial in d variables with coefficient vector $\mathbf{c}_n$, $\nu_n(.)$ is the Veronese map of degree $n \doteq s_{n,d}$. Collecting all data into a matrix leads to:

$$\mathbf{V}_n\mathbf{c}_n \doteq \left[\ \mathbf{v}_n^T(\mathbf{x}_1), \cdots, \mathbf{v}_n^T(\mathbf{x}_{N_p})\ \right]^T\mathbf{c}_n = \mathbf{0} \tag{5}$$

Thus, $\mathbf{c}_n$, the coefficients of the generator of the ideal $I_n(\mathscr{A})$, can be computed by simply finding a vector in the right nullspace of $\mathbf{V}_n$. The individual normals $\mathbf{b}_i$ can be recovered from $\mathbf{c}_n$ either by factoring the associated polynomial into a product of linear forms or through polynomial differentiation (see Vidal et al. (2005) for details).

2.4 Problem Statement

The goal of this paper is to identify the parameters of the linear time invariant (LTI) portion of a Wiener system of the form shown in Fig. 1, from noisy experimental data, for the case where the non-linearity is known. Specifically, we are interested in solving the following problem:

Problem 1. Consider the SISO Wiener structure shown in Fig. 1 described by a model of the form:

$$\begin{aligned} x_t &= \sum_{i=1}^{n_a} a_i x_{t-i} + \sum_{i=0}^{n_b} b_i u_{t-i} \\ \hat{y}_t &= F(x_t)\,, \quad y_t = \hat{y}_t + \eta_t \end{aligned} \tag{6}$$

where u_t and y_t denote the system's input and measured output, corrupted by additive noise η, and where $F(.)$ is a known static nonlinearity. Given upper bounds on n_a and n_b, and N input output pairs (u_t, y_t), the goal is to estimate the parameters $\mathbf{r} \doteq \left[a_1, \ldots, a_{n_a}, b_0, \ldots b_{n_b}\right]$ of the linear portion of the system.

In the sequel, we will make the following assumptions:

- A.1 $F : \mathscr{D} \to \mathscr{R}$ is a known, non-invertible even nonlinearity, that is $F(x) = F(-x)$.
- A.2 Consider now the one–to–many inverse mapping of F defined as
$$\mathscr{F}(y) = \{x \mid F(x) = y\} \tag{7}$$
and assume that, for each y, it has finitely many elements n_y. Let $n_{max} \doteq \sup_{y\in\mathscr{F}} n(y)$ [1]. In the sequel, we will assume that the input to the system operates in a region where $n_{max} = 2$.

Assumption A.1 (known, even non-linearity), restricts classes of problems that can be addressed by our framework. However, as noted in the introduction, it holds in many scenarios of practical interest ranging from biology to computer vision. Further, the class of problems considered here are amongst the most difficult Wiener systems identification problems: recall that the results in Sznaier (2009) show that these problems are NP hard in the number of experimental data points.

[1] n_{max} can be considered a measure of the "severity" of non-invertibility for the nonlinear block.

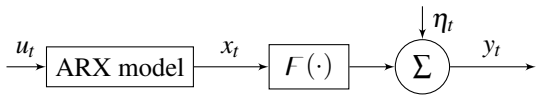

Fig. 1. General Wiener Identification Setup

Assumption A.2 leads to a simplification of the proposed approach. If it fails, some of the proposed ideas can still be used, but at the price of a far more involved algorithm. Note that, in most cases, the magnitude of the system input can be chosen so that it operates in a region where the assumption holds.

3. REDUCING THE PROBLEM TO IDENTIFICATION OF A SWITCHED SYSTEM

This section shows that Problem 1 can be reduced to identifying a switched ARX model. While in principle this is also a generically NP hard problem, efficient relaxations whose complexity is independent of the number of data points, can be obtained by exploiting the approach discussed in §2.3. The key observation motivating our approach is the fact that the elements of the set $\mathscr{F}(y)$ belong to a subspace arrangement (defined by hyperplanes whose coefficients are sign combinations of a_i). Thus, the absolute value of these coefficients can be obtained from vanishing ideal of this arrangement, and used to determine the correct signs of internal variables by minimizing the residual error over all sign combinations. Once these internal variables are found, finding the ARX model coefficients reduces to a simple LTI model identification. Before formalizing these ideas, below is a simple example to illustrate the supporting intuition:

3.1 An Illustrative Example

Consider the following Wiener system composed of the cascade of a first order LTI system and a square nonlinearity:

$$x_t = ax_{t-1} + b_0u_t + b_1u_{t-1} = ax_{t-1} + \mathbf{b}^T\mathbf{u}_t, \quad y_t = x_t^2 \tag{8}$$

Since for a (noiseless) first order system it is enough to consider two consecutive points to identify its dynamics, we consider the following (block Hankel) data matrix:

$$\mathbf{X} = \begin{bmatrix} \mathbf{H}_{|X|} \\ \mathbf{H}_{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} |x_1| & \cdots & |x_{n-1}| \\ |x_2| & \cdots & |x_n| \\ u_2 & \cdots & u_n \\ u_1 & \cdots & u_{n-1} \end{bmatrix} = \begin{bmatrix} \sqrt{y_1} & \cdots & \sqrt{y_{n-1}} \\ \sqrt{y_2} & \cdots & \sqrt{y_n} \\ u_2 & \cdots & u_n \\ u_1 & \cdots & u_{n-1} \end{bmatrix} \tag{9}$$

Next, form an extended data matrix $\mathbf{X}_{ext}$ by considering all possible 2^2 different sign combinations [2] of each column of $\mathbf{X}$:

$$\mathbf{X}_{ext} = \begin{bmatrix} \cdots & \sqrt{y_{t-1}} & -\sqrt{y_{t-1}} & \sqrt{y_{t-1}} & -\sqrt{y_{t-1}} & \cdots \\ \cdots & \sqrt{y_t} & \sqrt{y_t} & -\sqrt{y_t} & -\sqrt{y_t} & \cdots \\ \cdots & \mathbf{u}_t & \mathbf{u}_t & \mathbf{u}_t & \mathbf{u}_t & \cdots \end{bmatrix} \tag{10}$$

where $\mathbf{u}_t = [u_t, \quad u_{t-1}]^T$. Note that each sign combination of the actual intermediate signal indeed satisfies the same ARX model with different coefficients. Hence, the columns of $\mathbf{X}_{ext} \in \mathbb{R}^4$ belong to the following arrangement of four subspaces

$$\begin{aligned} S_1 \cup S_2 \cup S_3 \cup S_4 = \{\mathbf{x} : &(x_t - ax_{t-1} - \mathbf{b}^T\mathbf{u}_t = 0) \\ &\vee (x_t + ax_{t-1} - \mathbf{b}^T\mathbf{u}_t = 0) \\ &\vee (x_t - ax_{t-1} + \mathbf{b}^T\mathbf{u}_t = 0) \\ &\vee (x_t + ax_{t-1} + \mathbf{b}^T\mathbf{u}_t = 0)\} \end{aligned} \tag{11}$$

It can be easily seen that the following homogeneous quadratic polynomial generates the vanishing ideal of the arrangement:

$$\begin{aligned} &p(x_t, x_{t-1}, \mathbf{u}_t) \doteq \\ &(x_t^2 - (ax_{t-1})^2)^2 - 2(\mathbf{b}^T\mathbf{u})^2(x_t^2 + (ax_{t-1})^2) + (\mathbf{b}^T\mathbf{u})^4 \end{aligned} \tag{12}$$

and that $\mathbf{c}$, the coefficient vector of this polynomial, satisfies $\mathbf{c}^T v_4(\mathbf{d}) = 0$, where $v_4(\mathbf{d})$ denotes the Veronese map of degree four in variables $\mathbf{d} \doteq [x_t, \ x_{t-1}, \ u_t, \ u_{t-1}]$. Thus, as discussed in §2.3, $\mathbf{c}$ can be obtained by creating an embedded data matrix $\mathbf{V}$, by stacking Veronese maps corresponding to each time instant, and finding a vector in its null space. The ARX model paramaters can then be obtained, up to a sign, from elements of $\mathbf{c}$ corresponding to the highest order monomials. Further, note that (since the nonlinearity is even, with $n_{max} = 2$), $p(.)$ contains only even powers of the variables $|x_{t-1}|$ and $|x_t|$. Thus, only the 14 monomials indicated in Table 1, out of $s_{n,\mathbf{d}} \doteq \binom{n+(d-1)}{n} = \binom{7}{4} = 35$, have non-zero coefficients.

Table 1. Non zero monomials and corresponding exponents for the simple example

Term#	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\lvert x_t\rvert$	4	2	2	2	2	0	0	0	0	0	0	0	0	0
$\lvert x_{t-1}\rvert$	0	2	0	0	0	4	2	2	2	0	0	0	0	0
u_{t-1}	0	0	2	1	0	0	2	1	0	4	3	2	1	0
u_t	0	0	0	1	2	0	0	1	2	0	1	2	3	4

It follows that a reduced embedded data matrix $\mathbf{V}_{red}(\mathbf{d})$ can be constructed directly from the data matrix $\mathbf{X}$, by considering only the monomials listed in Table 1, resulting in a substantial complexity reduction. As before, the ARX model parameters (up to a sign) can be obtained from coefficients of a suitable normalized vector in the null space of $\mathbf{V}_{red}$.

Remark 1. In the case of impulse response experiments, the variable $\mathbf{b}^T\mathbf{u}$ in (12) reduces to a single coefficient b_0. Equivalently, (12) can be replaced by a non-homogeneous 2^{nd} order polynomial in the variables $|x_t|^2 = y_t$ and $|x_{t-1}|^2 = y_{t-1}$. Thus, the coefficients of the ARX model can be identified from a reduced data matrix created directly from the measured outputs.

3.2 Exploiting the Subspace Structure of Inverse Mappings of an Even Nonlinear Function

Consider a trajectory of (6) corresponding to a given input and define the regressor and model parameter vectors:

$$\begin{aligned} \mathbf{r}_{t,LTI} &\doteq [-x_t, ..., x_{t-n_a}, u_t, ..., u_{t-n_b}]^T \\ \mathbf{r}_{t,NL} &\doteq [-y_t, ..., y_{t-n_a}, u_t, ..., u_{t-n_b}]^T \\ &= [F(-x_t), ..., F(x_{t-n_a}), u_t, ..., u_{t-n_b}]^T \\ \mathbf{p}_{AR} &\doteq [1, a_1, ..., a_{n_a}, b_0, ..., b_{n_b}]^T \end{aligned} \tag{13}$$

Since $\mathbf{p}_{AR}^T\mathbf{r}_{t,LTI} = 0$ holds for all time instants, the corresponding regressor $\mathbf{r}_{t,LTI}$ lives in a subspace normal to $\mathbf{p}_{AR}$. Note that we cannot construct $\mathbf{r}_{t,LTI}$ directly from the observed data, since the internal variables $x_t, ..., x_{t-n_a}$ are not measurable. However, each measured output y_t provides a set of candidates $x_{t,i} \doteq \mathscr{F}_i(y_t), \ i = 1, \ldots, n_{max}$ defined by the inverse mapping set $\mathscr{F}(y_i)$ as in (7). The number of all possible combinations of the elements $\mathscr{F}(y_t), \cdots, \mathscr{F}(y_{t-n_a})$ is $n_{map} \doteq n_{max}^{n_a+1} = 2^{n_a+1}$. Writing these n_{map} combinations in a matrix form leads to:

$$\mathbf{R}_t = \begin{bmatrix} \mathscr{F}_1(y_t) & \cdots & \mathscr{F}_{i_0}(y_t) & \cdots & \mathscr{F}_{n_{max}}(y_t) \\ \vdots & & \vdots & & \vdots \\ \mathscr{F}_1(y_{t-n_a}) & \cdots & \mathscr{F}_{i_{n_a}}(y_{t-n_a}) & \cdots & \mathscr{F}_{n_{max}}(y_{t-n_a}) \\ \mathbf{u}_t & \cdots & \mathbf{u}_t & \cdots & \mathbf{u}_t \end{bmatrix} \tag{14}$$

[2] Entries corresponding to $\mathbf{u}_t$ are not "unfolded" since these values are known.

where $\mathbf{u}_t \doteq [u_t, \cdots, u_{t-n_b}]^T$ and $\mathbf{R}_t \in \mathbb{R}^{(n_a+n_b+2)\times n_{map}}$. In the noiseless case, the true combination matching $\mathbf{r}_{t,LTI}$ corresponds to one of the columns in $\mathbf{R}_t$ and hence lives in a subspace defined by the normal vector $\mathbf{p}_{AR}$. Further, under Assumptions A.1 and A.2 ($F(.)$ even and $n_{max} = 2$) the remaining columns of $\mathbf{R}_t$ also live in subspaces whose normals are defined by suitable permutations of the signs of $\mathbf{P}_{AR}$. Thus, at each time instant t, the columns of the corresponding matrix $\mathbf{R}_t$ belong to the union of n_{map} subspaces. In principle, the GPCA approach discussed in §2.3 could be used to find the vanishing ideal of the arrangement, that is polynomials of the form:

$$p(\mathbf{z}) = \mathbf{c}_n^T \mathbf{v}_n(\mathbf{z}) \tag{15}$$

and such that

$$\mathbf{c}_n^T [\mathbf{v}_n(\mathbf{z}_{1,1}) \ \ \cdots \mathbf{v}_n(\mathbf{z}_{t,j}) \ \ \cdots \mathbf{v}_n(\mathbf{z}_{n_{max},N})] \doteq \mathbf{c}_n^T \mathbf{V}_n = 0 \tag{16}$$

Here $\mathbf{z}_{t,j}$, $\mathbf{v}_n(.)$ and $\mathbf{V}_n$ denote the j^{th} column of $\mathbf{R}_t$, the corresponding Veronese map of degree n in $n_a + n_b + 2$ variables, and the embedded data matrix, respectively.

In this context, finding the coefficients $\mathbf{c}_n$ simply reduces to finding vectors in the null space of $\mathbf{V}_n$. A potential difficulty here stems from the size of $\mathbf{V}_n$ in (16), with $s \doteq \binom{n_{map}+n_a+n_b+1}{n_{map}}$ rows and $N \times n_{map}$ columns. As shown next, this difficulty can be avoided by exploiting the symmetry of the problem. Note that since the non-linearity is even, the polynomial (15) contains only even powers of the variables $\mathscr{F}_i(y_t)$. Hence the elements of $\mathbf{c}_n$ corresponding to monomials with odd powers of variables are zero and corresponding columns can be eliminated from $\mathbf{V}_n$. Further, once these columns are eliminated, all rows corresponding to a given matrix $\mathbf{R}_t$ become identical. It follows that the non-zero elements of $\mathbf{c}_n$ can be obtained by considering a reduced embedded data matrix $\mathbf{V}_{red}$ containing only the columns corresponding to even powers of the variables $\mathscr{F}_i(y_t)$ and N rows, each obtained from Veronese map of the first column of the corresponding $\mathbf{R}_t$. As shown in §3.1, this leads to a substantial reduction in the number of columns/rows of $\mathbf{V}_n$.

3.3 Handling noise

So far, we have considered the ideal case of noiseless data. In order to show that the same approach extends to the case where the data is corrupted by noise we introduce the following result:

Proposition 1. Given a vector random variable $\mathbf{z}$ with probabilty distribution μ, consider the problem of finding a (normalized) minimum variance polynomial of degree n, that is:

$$\min_{\mathbf{c}_n} \mathscr{E}_\mu [\mathbf{c}_n^T \mathbf{v}_n(\mathbf{z})]^2 \text{ subject to } \|\mathbf{c}_n\|_2 = 1$$

Then, the solution is given by:

$$\mathbf{c}_n^* = \text{minimum singular vector of } \mathbf{M}_n$$

where $\mathbf{M}_n \doteq \mathscr{E}_\mu [\mathbf{v}_n(\mathbf{z}) \mathbf{v}_n(\mathbf{z})^T]$.

Proof: Follows by noting that

$$\begin{aligned} \mathscr{E}_\mu [\mathbf{c}_n^T \mathbf{v}_n(\mathbf{z})]^2 &= \mathscr{E}_\mu [\mathbf{c}_n^T \mathbf{v}_n(\mathbf{z}) \mathbf{v}_n^T(\mathbf{z}) \mathbf{c}_n] \\ &= \mathbf{c}_n^T \mathscr{E}_\mu [\mathbf{v}_n(\mathbf{z}) \mathbf{v}_n^T(\mathbf{z})] \mathbf{c}_n = \mathbf{c}_n^T \mathbf{M}_n \mathbf{c}_n \end{aligned}$$

Hence

$$\mathbf{c}_n^* = \underset{\|\mathbf{c}_n\|=1}{argmin} \mathbf{c}_n^T \mathbf{M}_n \mathbf{c}_n = \text{minimum singular vector of } \mathbf{M}_n. \quad \square$$

Remark 2. Note that when the random variable $\mathbf{z}$ belongs to an arrangement of n subspaces, each defined by its normal $\mathbf{b}_i$, then

$$\begin{aligned} p_n^*(\mathbf{z}) &= \prod_{i=1}^{n} (\mathbf{b}_i^T \mathbf{z}) = \mathbf{c}_n^T \mathbf{v}_n(\mathbf{z}) = 0 \\ &\Rightarrow \mathscr{E}_\mu [\mathbf{c}_n^T \mathbf{v}_n(\mathbf{z})]^2 = \mathbf{c}_n^T \mathbf{M}_n \mathbf{c}_n = 0 \end{aligned}$$

It follows that $\mathbf{M}_n$ is singular, with its null space spanned by $\mathbf{c}_n$. The corresponding polynomial p^* is precisely a generator of the vanishing ideal of the subspace arrangement.

In the case of interest to this paper, when y_t is corrupted by noise, the true probability distribution of the pre-images $\mathscr{F}_i(y_t)$ is, in general, difficult to compute. Thus, we will replace μ and its corresponding moments matrix by the empirical distribution:

$$\mathbf{M}_{emp} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{v}_n(\mathbf{z}_i) \mathbf{v}_n^T(\mathbf{z}_i) = \frac{1}{N} \mathbf{V}_n \mathbf{V}_n^T$$

where $\mathbf{V}_n$ is precisely the embedded data matrix in (16), and find a polynomial that has minimum variance with respect to this empirical distribution. When applied to noisy data, this approach can be thought of as finding the coefficients of the polynomial that minimizes the expected value of the quadratic fitting error to the associated subspaces arrangement.

4. PROPOSED ALGORITHM

The proposed algorithm, based on the ideas discussed in the previous sections, has three main components:

(i) Forming a data matrix $\mathbf{X}$ for the internal variables $\mathbf{x}_t$
(ii) Forming a reduced embedded data matrix $\mathbf{V}_{red}$ from $\mathbf{X}$
(iii) Estimating coefficients of the LTI block and its output

Algorithm 1 Wiener system identification

Require: Hankelized inverse mapping of the output signal $\mathbf{H}_{\mathscr{F}_1(\mathbf{y})}$, Hankelized input signal $\mathbf{H}_\mathbf{u}$, output order n_a, input order n_b, horizon length N

Forming a data matrix X for the internal variables $\mathbf{x}_t$
$n_{map} \leftarrow 2^{n_a+1}$ ▷ number of possible inverse mappings
$\mathbf{X} \leftarrow \begin{bmatrix} \mathbf{H}_{\mathscr{F}_1(\mathbf{y})} & \mathbf{H}_\mathbf{u} \end{bmatrix}^T$ ▷ Data matrix
Forming a reduced embedded data matrix $\mathbf{V}_{red}$ from X
$n_y \leftarrow n_{map}/2, \quad d_y \leftarrow (n_a+1))$
$s_{n_y,d_y+1} \leftarrow \binom{n_y+(d_y+1)-1}{n_y}$ ▷ # of the polynomial p elements
Compute $\mathbf{V}_{red,lumped} \in \mathbb{R}^{s_{n_y,d_y+1} \times N-n_a-1}$ ▷ Reduced Veronese mapping in n_a+2 variables as represented by $\tilde{\mathbf{d}}$
$\mathbf{V}_{red} \leftarrow$ Extended $\mathbf{V}_{red,lumped}$ ▷ unfolds $\tilde{u}$ in $\tilde{\mathbf{d}}$
$[\mathbf{U}, \mathbf{S}, \mathbf{V}] \leftarrow \mathbf{svd}(\mathbf{V}_{red}^T \mathbf{V}_{red})$
$\mathbf{c_n} \leftarrow \mathbf{V}(:, end)$ ▷ Coefficients of polynomial p
Estimating coefficients of the LTI block and its output
$(\hat{c}_{abs}) \leftarrow (c_{n,i})^{\frac{1}{n_{map}}}$ ▷ $\hat{c}_{abs} \in \mathbb{R}^{n_a+n_b+1}$ estimation of $\mathbf{c}_{abs}$
$\mathbf{S}_k \leftarrow$ all possible sign combinations for $\mathbf{x}$
$k_i = argmin_k |(\mathbf{S}(\mathbf{c}_{abs} \circ \mathbf{X}_i)_k|$
$\mathbf{x}_{sign,i} \leftarrow \mathbf{S}_{k_i}(n_a+1)$ where $i = 1, \cdots, N_{p,y}$
$\hat{\mathbf{x}} \leftarrow \mathbf{X}(n_a+1)\mathbf{x}_{sign,i}$ ▷ estimate of the internal variables
$\mathbf{H}_{\hat{\mathbf{x}}} \leftarrow \mathbf{H}_{\hat{\mathbf{x}}}^{n_a+1,N-2\times n_a-1}$ ▷ Hankelize $\hat{\mathbf{x}}$
$\mathbf{X}_{LTI} \leftarrow [\mathbf{H}_{\hat{\mathbf{x}}} \quad \mathbf{H}_\mathbf{u}]^T$ ▷ Data matrix with estimated variables
$[\mathbf{U}, \mathbf{S}, \mathbf{V}] \leftarrow \mathbf{svd}(\mathbf{X}_{LTI} \mathbf{X}_{LTI}^T)$
$\hat{\mathbf{c}}_{\mathbf{ARX}} \leftarrow \mathbf{V}(:, end)/\mathbf{V}(n_a+1, end)$ ▷ estimated ARX model

Forming the data matrix. The data matrix is formed as:

$$\mathbf{X} = \begin{bmatrix} \mathbf{H}_{\mathscr{F}_1(\mathbf{y})}, \mathbf{H}_\mathbf{u} \end{bmatrix}^T = \begin{bmatrix} \mathbf{H}_{\mathscr{F}_1(\mathbf{y})}^{n_a+1,N-n_a-1}, \mathbf{H}_\mathbf{u}^{n_b+1,N-n_a-1} \end{bmatrix}^T \tag{17}$$

Note that in here we consider, for each y_t, one element of the set $\mathscr{F}$. As discussed in Section 3, the reduced embedded data

matrix can be constructed by considering just the first column of the matrices $\mathbf{R}_t$. Thus, without loss of generality, we can take this element to be the one corresponding to $i = 1$.

Forming the reduced embedded data matrix. As shown earlier in §3, it suffices to consider a reduced embedded data matrix containing only the columns corresponding to even powers of variables $\mathscr{F}_1(\mathbf{y}_t)$. This reduced matrix can be constructed by generating monomials of a Veronese mapping of order $n_{map}/2$ in n_a+2 variables $\tilde{\mathbf{d}} \doteq [\mathscr{F}_1^2(y_t),\ \cdots,\ \mathscr{F}_1^2(y_{t-n_a}),\ \tilde{u}]$, where the effects of inputs are lumped in a single variable $\tilde{u} \doteq (\mathbf{b}^T\mathbf{u}_t)^2$. The reduced embedded data matrix $\mathbf{V}_{red}$ can then be obtained by "unfolding" the elements of $\tilde{u}$, leading to a matrix whose columns correspond to all monomials of degree n in the variables $[\mathscr{F}_1^2(y_t),\ \cdots,\ \mathscr{F}_1^2(y_{t-n_a}), u_t,\ \cdots, u_{t-n_b}]$

Estimating the coefficients of the LTI block and its output. Once the matrix $\mathbf{V}_{red}$ is available, the next step is to compute the coefficients $\mathbf{c}_n$ of the minimum variance polynomial p by finding the singular vector of $\mathbf{V}_{red}$ associated with its minimum singular value. The absolute values of ARX model parameters can be estimated from $|a_i| = (c_{n,i})^{\frac{1}{n_{map}}}$, where $c_{n,i}$ denotes the element of $\mathbf{c}_n$ associated with the monomial $\mathscr{F}_1^{n_{map}}(y_{t-i})$, with a similar expression for $|b_i|$. To obtain the correct signs for these coefficients, begin by writing the model (6) as:

$$-sign(x_t)|x_t| + \sum_{i=1}^{n_a} sign(a_i x_{t-i})|a_i x_{t-i}| + \sum_{j=0}^{n_b} sign(b_j)|b_j|u_{t-j} = 0 \tag{18}$$

Define $\mathbf{c}_{abs} \doteq [|a_{n_a}|\ \cdots\ |a_1|\ -1\ |b_{n_b}|\ \cdots\ |b_0|]^T$ and the matrix

$$\mathbf{S} \doteq \begin{bmatrix}\mathbf{sg}_1\ \cdots\ \mathbf{sg}_{n_{map}}\end{bmatrix}^T$$

where $n_{map} = n_{max}^{n_a+1}$, the first n_a+1 elements of the vectors $\mathbf{sg}_i$ contain all possible sign combinations of ± 1 and the last n_b elements are 1. Note that as discussed in §3.1, when the input signal is an impulse, there is no need to generate the sign combinations for entries corresponding to $\mathbf{u}_t$.

Let $\mathbf{X}_i$ denote the i^{th} column of the data matrix $\mathbf{X}$ in (17) and consider the vector $\mathbf{r} \doteq \mathbf{S}(\mathbf{c}_{abs} \circ \mathbf{X}_i)$. In the noiseless case, since at least one of the elements of the matrix $\mathbf{S}$ contains the correct sign combination, from (18) it follows that there exist at least one k such that r_k, the corresponding entry of $\mathbf{r}$, must be zero. When k is unique, the corresponding row of $\mathbf{S}$, $\mathbf{sg}_k$, indeed contains the correct signs for elements b_i and products $(a_i x_{t-i})$. Particularly, since $a_0 = -1$, the sign of x_t, for $t \geq n_a+1$ can be read directly from this vector. In degenerate cases where more than one component of $\mathbf{r}$ is zero, the correct signs for x_t can be recovered by imposing consistency of signs in additional columns of $\mathbf{X}$. Once x_t estimates are available, from the null space of its Hankel matrix coefficients a_i can be recovered . In the noisy case where, generically, all elements $r_k \neq 0$, we propose to select the smallest (in absolute value) element, that is:

$$k_i = \underset{k}{argmin}\,|[\mathbf{S}(\mathbf{c}_{abs} \circ \mathbf{X}_i)]_k|$$

Optional: refining the model. As we will illustrate in §5 the proposed algorithm successfully identifies the correct model in scenarios where the noise level is low. However, as shown there, performance tends to degrade quickly as the noise increases, due to the lack of robustness of factorization based algebraic methods. This difficulty can be circumvented by combining the proposed method with Wiener identification methods that rely on local optimization, such as the one implemented by MATLAB nlhw(.) command. As shown below, using the proposed method to provide an initial estimate to nlhw(.) results in both improved performance and reduced computational time.

5. NUMERICAL EXPERIMENTS AND COMPARISONS

In order to illustrate the advantages of the proposed approach we benchmarked our algorithm against Matlab's nlhw(.) command and the approach recently introduced by Risuleo et al. (2018) [3], on the following system, taken from Hagenblad et al. (2008); Tiels and Schoukens (2014):

$$\begin{aligned} x_t &= -0.3x_{t-1} + 0.3x_{t-2} + u_t - 0.3u_{t-1} + 0.3u_{t-2} \\ \hat{y}_t &= x_t^2 \quad , \quad y_t = \hat{y}_t + \eta_t \end{aligned} \tag{19}$$

The above system was excited with a Gaussian random input signal. The identification performance was tested under 40dB, 35dB, 30dB, 25dB and 20dB signal to noise ratio (SNR) values and horizon length $N = 600$. The results from 100 random runs are summarized in Table 2. Performance of the algorithms was evaluated both in terms of Fitting Accuracy defined as [4] $FA \doteq (1 - \frac{\|\hat{\mathbf{y}}-\tilde{\mathbf{y}}\|_2}{\|\hat{\mathbf{y}}-\text{mean}(\hat{\mathbf{y}})\|_2})$, where $\tilde{\mathbf{y}}$ is the estimated output of the Wiener system, and parameter estimation error, defined as $\text{PE} \doteq \max\|\mathbf{c}_{AR} - \tilde{\mathbf{c}}_{AR}\|_2$. Note that since Risuleo et al. (2018) is a semi-parametric method, it does not estimate the ARX model parameters and thus the PE criteria is not applicable.

Table 2. Comparisons on Numerical Examples

SNR	Method	FA	PE	Time (secs)
Noiseless	W-SA	0.82	N/A	0.5660
	nlhw only	0.60	0.5758	0.5801
	proposed only	1	1.20e-14	0.0233
	proposed&nlhw	0.98	0.0028	0.1894
40dB	W-SA	0.82	N/A	0.6325
	nlhw only	0.59	0.6113	0.4981
	proposed only	0.35	0.7571	0.0226
	proposed&nlhw	0.98	0.0051	0.2000
35dB	W-SA	0.82	N/A	0.5417
	nlhw only	0.57	0.6077	0.5734
	proposed only	0.33	0.7765	0.0229
	proposed&nlhw	0.97	0.0079	0.2225
30dB	W-SA	0.81	N/A	0.6219
	nlhw only	0.55	0.5659	0.4761
	proposed only	0.29	0.8011	0.0224
	proposed&nlhw	0.96	0.0133	0.2001
25dB	W-SA	0.81	N/A	0.5332
	nlhw only	0.60	0.5068	0.5331
	proposed only	0.24	0.8384	0.0230
	proposed&nlhw	0.93	0.0235	0.2262
20dB	W-SA	0.78	N/A	0.6271
	nlhw only	0.41	0.6138	0.4788
	proposed only	0.17	0.7103	0.0226
	proposed&nlhw	0.88	0.0419	0.2043

As shown in Table 2, the proposed algorithm outperforms the alternatives, both in accuracy and timing, in noiseless scenarios. However, its performance quickly degrades as the noise level increases. On the other hand, using the proposed method to generate an initial estimate to nlhw, provided by MATLAB System Identification Toolbox to identify Hammerstein-Wiener systems, leads to both, substantial accuracy improvement and reduction in the computational time, even for large noise values.

[3] In order to have a fair comparison, explicit information about the nonlinearity was used in all of the methods.

[4] From this definition it follows that $FA = 1$ corresponds to a perfect fit.

6. CONCLUSION

Wiener system identification problems arise in many domains from control to biomedical applications. When the output nonlinearity is non-invertible, even if it is known, the problem is computationally challenging (Sznaier, 2009). To circumvent this difficulty, we propose a computationally tractable method, based on an algebraic approach. The main idea is based on the observation that the one-to-many inverse set mappings of the output belong to an arrangement of subspaces, whose normals are sign combinations of true model coefficients. This observation allows for recasting the problem into a switched systems identification (or subspace clustering) form, where a polynomial in the vanishing ideal of the arrangement is first found from the null space of an embedded data matrix, and the model coefficients are then recovered from this polynomial. While in principle the size of the embedded data matrix scales combinatorially with the system order, as shown in the paper, this growth can be mitigated by exploiting the structure of the nonlinearity inverse mapping. The proposed approach effectiveness was illustrated on an example taken from the literature. As shown in the paper, combining the proposed method with existing tools leads to a substantial improvement in accuracy and computational time. Note that this combination was needed to robustify the method, given the fragility of factorization based methods to noise. Ongoing research seeks a more computationally efficient approach to achieve robustness against noise by considering the approach proposed in Sznaier and Camps (2018).

REFERENCES

Ayazoglu, M., Sznaier, M., Lagoa, C., and Camps, O. (2010). A moments-based approach to estimation and data interpolation for a class of Wiener systems. In *49th IEEE Conference on Decision and Control (CDC)*, 5674–5680.

Ayazoglu, M., Sznaier, M., and Camps, O. (2010). Euclidean structure recovery from motion in perspective image sequences via hankel rank minimization. In *European Conference on Computer Vision*, 71–84. Springer.

Bai, E.W. (2003). Frequency domain identification of Wiener models. *Automatica*, 39(9), 1521–1530.

Bottegal, G., Castro-Garcia, R., and Suykens, J.A. (2018). A two-experiment approach to Wiener system identification. *Automatica*, 93, 282–289.

Boyd, S. and Chua, L. (1985). Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Transactions on circuits and systems*, 32(11), 1150–1161.

Da Silva, M.M., Wigren, T., and Mendonça, T. (2012). Nonlinear identification of a minimal neuromuscular blockade model in anesthesia. *IEEE Transactions on Control Systems Technology*, 20(1), 181–188.

De Angelis, A., Schoukens, J., Godfrey, K., and Carbone, P. (2017). Measuring the best linear approximation of Wiener systems using multilevel sequences. In *Instrumentation and Measurement Technology Conference (I2MTC), 2017 IEEE International*, 1–5. IEEE.

Giri, F. and Bai, E.W. (2010). *Block-oriented nonlinear system identification*, volume 1. Springer.

Giri, F., Rochdi, Y., and Chaoui, F.Z. (2009). An analytic geometry approach to Wiener system frequency identification. *IEEE Transactions on Automatic Control*, 54(4), 683–696.

Giri, F., Rochdi, Y., Radouane, A., Brouri, A., and Chaoui, F. (2013). Frequency identification of nonparametric Wiener systems containing backlash nonlinearities. *Automatica*, 49(1), 124–137.

Hagenblad, A., Ljung, L., and Wills, A. (2008). Maximum likelihood identification of Wiener models. *Automatica*, 44(11), 2697–2705.

Khan, A. and Vyas, N. (2001). Nonlinear bearing stiffness parameter estimation in flexible rotor–bearing systems using Volterra and Wiener approach. *Probabilistic Engineering Mechanics*, 16(2), 137–157.

Ma, Y., Yang, A.Y., Derksen, H., and Fossum, R. (2008). Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM review*, 50(3), 413–458.

Norquay, S.J., Palazoglu, A., and Romagnoli, J. (1998). Model predictive control based on Wiener models. *Chemical Engineering Science*, 53(1), 75–84.

Ozay, N., Sznaier, M., Lagoa, C., and Camps, O. (2010). GPCA with denoising: A moments-based convex approach. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3209–3216. IEEE.

Ren, J. and Li, G. (2015). Identification of Wiener systems with non-invertible nonlinearity. In *Image and Signal Processing (CISP), 2015 8th International Congress on*, 214–218. IEEE.

Risuleo, R.S., Lindsten, F., and Hjalmarsson, H. (2018). Semiparametric kernel-based identification of Wiener systems. In *2018 IEEE Conference on Decision and Control (CDC)*, 3874–3879. IEEE.

Song, J.H., Kim, C., and Yoo, Y. (2015). Vein visualization using a smart phone with multispectral Wiener estimation for point-of-care applications. *IEEE journal of biomedical and health informatics*, 19(2), 773–778.

Sznaier, M. and Camps, O. (2018). Sos-rsc: A sum-of-squares polynomial approach to robustifying subspace clustering algorithms. In *IEEE CVPR*, 8033–8041.

Sznaier, M., Ma, W., Camps, O.I., and Lim, H. (2009). Risk adjusted set membership identification of Wiener systems. *IEEE Transactions on Automatic Control*, 54(5), 1147–1152.

Sznaier, M. (2009). Computational complexity analysis of set membership identification of Hammerstein and Wiener systems. *Automatica*, 45(3), 701–705.

Tan, A.H. (2013). Direct synthesis of pseudo-random ternary perturbation signals with harmonic multiples of two and three suppressed. *Automatica*, 49(10), 2975–2981.

Tiels, K. and Schoukens, J. (2014). Wiener system identification with generalized orthonormal basis functions. *Automatica*, 50(12), 3147–3154.

Vidal, R., Ma, Y., and Sastry, S. (2005). Generalized principal component analysis (GPCA). *IEEE Trans. PAMI*, 27(12), 1945–1959.

Wigren, T. (1994). Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model. *IEEE Transactions on Automatic Control*, 39(11), 2191–2206.

Yılmaz, B. and Sznaier, M. (2015). Efficient identification of Wiener systems using a combination of atomic norm minimization and interval matrix properties. In *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*, 109–114. IEEE.

Zhang, Q., Laurain, V., and Wang, J. (2015). Weighted principal component analysis for Wiener system identification-regularization and non-Gaussian excitations. *IFAC-PapersOnLine*, 48(28), 602–607.