Optimal approximate model identification in the Hankel norm

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 94-09

Optimal approximate model identification in the Hankel norm

S. Weiland
A.A. Stoorvogel

Eindhoven, March 1994
The Netherlands

OPTIMAL APPROXIMATE MODEL IDENTIFICATION IN THE HANKEL NORM

S. WEILAND * and A.A. STOORVOGEL **†

* *Dept. of Electr. Eng., Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*

** *Dept. of Mathematics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*

Abstract. A method of optimal approximate system identification is proposed using a distance measure between an exact model for the observed data sequences and a reduced order (approximate) model for the same data but of lower complexity. A key property of this measure is that the distance is independent of specific parameterizations of the model. This distance measure can be computed in terms of induced norms of Hankel operators which are associated with the data. Using these ideas and a behavioral framework of describing dynamical systems, we put forward a new algorithm for optimal approximate identification of time series.

Key Words. System identification, approximate modeling, Hankel operators, behavioral theory

1. INTRODUCTION

In this paper we consider the problem of identifying linear time-invariant models from a given set of time series. We depart from the standard setting of this problem by refraining from a-priori assumptions on the input-output structure of the to be identified system. It is first shown that for an important class of time series optimal linear time-invariant models which explain the data are characterized by autonomous systems. Optimal (i.e. least complexity) input-output models are derived from such an autonomous model in a straightforward way using polynomial representations.

The question of approximate modeling amounts to minimizing the misfit between model and data subject to a complexity constraint on the class of candidate models. Motivated by the observation that exact identification of time series results in autonomous models, the question of approximate modeling of time series is addressed in the context of autonomous systems. We define a natural distance measure between data and model which is independent of representations of models. It is shown that this non-parametric misfit measure corresponds to the Hankel norm of an operator which is associated with the data and which maps models to model errors. Optimal approximate models are then obtained by performing optimal Hankel norm reductions. We believe that this method has important implications not only for applications in system identification, but also for problems related to data reduction, model approximation, the analysis of model uncertainty and robust control.

We conclude this section with the introduction of some notation. Let $T \subset \mathbb{Z}$ and denote by $l_2(T, \mathbb{R}^q)$ the set of functions $w : T \to \mathbb{R}^q$ for which

$$\| w \|_2^2 := \sum_{t \in T} | w(t) |^2 < \infty.$$

Here, $| \cdot |$ denotes the standard Euclidean norm in $\mathbb{R}^q$. If the dimensions of these objects will be understood from the context and if $T = \mathbb{Z}_+ := \mathbb{Z} \cap [1, \infty)$ [or $T = \mathbb{Z}_-$] then we write l_2^+ [l_2^-] for $l_2(T, \mathbb{R}^q)$. Let $\mathcal{L}_2$ denote the space of all power series $\sum_{t \in \mathbb{Z}} w(t) z^{-t}$, $z \in \mathbb{C}$, with $w \in l_2(\mathbb{Z}, \mathbb{R}^q)$. Let $\mathcal{H}_2^+$ and $\mathcal{H}_2^-$ denote the subspaces of $\mathcal{L}_2$ for which $w(t) = 0$, $t \leq 0$ and $w(t) = 0$, $t > 0$, respectively. Then $\mathcal{H}_2^+$ and $\mathcal{H}_2^-$ are Hilbert spaces with their natural norms and inner products and,

† The research of dr. A.A. Stoorvogel has been made possible by a fellowship of the Royal Netherlands Academy of Sciences and Arts.

clearly, $\mathcal{L}_2 = \mathcal{H}_2^- \oplus \mathcal{H}_2^+$. We will write Π_+ and Π_- for the canonical projections of $\mathcal{L}_2$ onto $\mathcal{H}_2^+$ and $\mathcal{H}_2^-$, respectively.

2. DATA AND UNFALSIFIED MODELS

Consider a finite set of q dimensional time-series

$$\tilde{w}_i \,:\, \mathbb{Z}_+ \longrightarrow W, \quad i = 1, \ldots, n \tag{1}$$

where $n > 0$ is the number of observed time-series and $W := \mathbb{R}^q$ is the *signal space* in which the observed variables take their values. A fundamental problem is to find an optimal model which explains this data. Following the framework of Willems (Willems, 1986; Willems, 1991), a model, or a *system*, is any subset $\mathcal{B}$ of the set of q–variate time series $w : T \to W$. That is, a system is viewed as a collection of trajectories $\mathcal{B} \subseteq W^T$, with T the *time axis* which is assumed to be $\mathbb{Z}_+$ for our application. The set $\mathcal{B}$ has the interpretation of specifying those time series w which are compatible with the laws of the system.

A few basic concepts of the behavioral framework are reviewed. A system $\mathcal{B}$ is called *linear* if $\mathcal{B}$ is a linear subspace of W^T, it is said to be *shift invariant* if $\sigma\mathcal{B} \subseteq \mathcal{B}$. Here, σ denotes the usual shift $(\sigma w)(t) = w(t+1)$. A system $\mathcal{B}$ is said to be *autonomous* if there exists a $k > 0$ such that the mapping $\pi_k : \mathcal{B} \to \mathcal{B}\mid_{[1,k]}$, defined by the restriction $\pi_k(w) := w\mid_{[1,k]}$, is injective. A system $\mathcal{B}$ is *unfalsified* by the data (1) if $\tilde{w}_i \in \mathcal{B}$ for $i = 1, \ldots, n$. A system $\mathcal{B}_1$ is said to be *more powerful* than $\mathcal{B}_2$ if $\mathcal{B}_1 \subseteq \mathcal{B}_2$. Hence, the more powerful a system is, the more trajectories it refutes.

Let $\mathbf{B}$ denote the class of all linear, time-invariant and closed (in the topology of pointwise convergence) subsets of $W^{\mathbb{Z}_+}$. The *most powerful unfalsified model* in $\mathbf{B}$, based on the data (1), is a model $\mathcal{B}_{mpum} \in \mathbf{B}$ which is unfalsified by (1) and which is more powerful than any other unfalsified model $\mathcal{B} \in \mathbf{B}$. It has been shown (see (Willems, 1986; Willems, 1991)) that for any data set (1) the most powerful unfalsified model $\mathcal{B}_{mpum} \in \mathbf{B}$ indeed exists and is unique.

Consider the power series

$$\Theta(z) := \Theta_0 + \Theta_1 z + \cdots + \Theta_k z^k + \cdots$$

where $z \in \mathbb{C}$ and Θ_i are constant real matrices, i.e., $\Theta_i \in \mathbb{R}^{g\times q}$ for some $g > 0$. It is assumed that $\Theta \in R\mathcal{H}_\infty^-$, i.e., Θ is rational and analytic for all points z inside the unit circle. Let $\Theta(\sigma)$ be an operator acting on signals $w \in l_2^+$ and consider the equation

$$\Theta(\sigma)w = 0 \tag{2}$$

This defines a linear time-invariant system $\mathcal{B} \in \mathbf{B}$ by putting

$$\mathcal{B} = \mathcal{B}(\Theta) := \{w \in l_2^+ \mid \Theta(\sigma)w = 0\}. \tag{3}$$

Here, $\Theta(\sigma)$ is said to be a *kernel representation* of $\mathcal{B}$. It is important to observe that with systems defined by (3), we restrict attention to linear, shift-invariant and closed subsets of $l_2(\mathbb{Z}_+, \mathbb{R}^q)$. Thus, $\mathcal{B} \in \mathbf{B} \cap l_2^+$. We therefore refer to these systems as l_2^+ systems in $\mathbf{B}$.

Let $\hat{\mathcal{B}}$ be defined as the image of $\mathcal{B}$ under the Laplace transform and let Π_+ denote the canonical projection $\Pi_+ : \mathcal{L}_2 \to \mathcal{H}_2^+$. Then, in the frequency domain, (3) is equivalently described by

$$\hat{\mathcal{B}} = \{\hat{w} \in \mathcal{H}_2^+ \mid \Pi_+\Theta\hat{w} = 0\} = \mathrm{Ker}\ \Pi_+\Theta.$$

Let W denote the Laplace transform of the data matrix $[\tilde{w}_1, \cdots, \tilde{w}_n]$, i.e.,

$$W(z) := \tilde{W}(1)z^{-1} + \tilde{W}(2)z^{-2} + \ldots, \tag{4}$$

where $\tilde{W}(t) = [\tilde{w}_1(t), \cdots \tilde{w}_n(t)]$, $t \in \mathbb{Z}_+$. We will make the following assumption on the data

Assumption 2.1 W is rational and belongs to $\mathcal{H}_2^+$.

Then, clearly, the Laplace transform defines a bijection between the data (1) and W. Typical examples of these data sets include finite length measurements, finite sets of frequency response measurements and polynomial-exponential time-series. A study of data sets of this type recently appeared in (Antoulas and Willems, 1993).

We will be interested in characterizing the set of all kernel representations of unfalsified models of the data (1). For this purpose, let $\mathcal{M}$, the *family of kernel representations* of (1), be defined by

$$\mathcal{M} := \{\Theta \in R\mathcal{H}_\infty^- \mid \Pi_+\Theta W = 0\}.$$

Furthermore, introduce a Hankel operator $\Gamma_W : (R\mathcal{H}_\infty^-)^{\cdot\times q} \to \mathcal{H}_2^+$ defined by the composition $\Gamma_W(\Theta) := \Pi_+\Theta W$. The model set $\mathcal{M}$ is then characterized as follows.

Theorem 2.2 Let $W = \Theta_{mpum}^{-1}\Psi$ be a left coprime factorization over $R\mathcal{H}_{\infty}^{-}$ of W. Then the following statements are equivalent

1. $\Theta \in \mathcal{M}$
2. $\Theta \in Ker\,\Gamma_W$
3. $\Theta = \Lambda\Theta_{mpum}$ for some $\Lambda \in R\mathcal{H}_{\infty}^{-}$.

Moreover, the most powerful unfalsified model $\mathcal{B}_{mpum} = \mathcal{B}(\Theta_{mpum})$ is an autonomous system.

A proof can be found in (Antoulas and Willems, 1991; Antoulas, 1993). Theorem 2.2 has the interpretation that all models (3) for the data (1) are generated from $\mathcal{B}_{mpum}$ by left-multiplication of Θ_{mpum} by elements $\Lambda \in R\mathcal{H}_{\infty}^{-}$. Note that Θ_{mpum} is a square $q \times q$ rational matrix which is uniquely determined up to pre-multiplication by units in $\mathcal{H}_{\infty}^{-}$. Observe also that Ker $\Pi_{+}\Theta_{mpum}$ is a finite dimensional subspace of $\mathcal{H}_2^{+}$.

3. THE APPROXIMATE MODELING PROBLEM

In order to formalize the approximate modeling problem we specify the *complexity* of a system $\mathcal{B} \in \mathbf{B}$ together with a criterion of *misfit* which expresses to what extent a model fails to explain a given data set (1). The approximate modeling problem then consists of finding a low complexity model which minimizes the misfit between model and data. Let $\mathcal{B} = \mathcal{B}(\Theta) \in \mathbf{B}$ with $\Theta \in R\mathcal{H}_{\infty}^{-}$.

Definition 3.1 The *complexity* $c(\mathcal{B})$ of $\mathcal{B} = \mathcal{B}(\Theta)$ is a pair $c(\mathcal{B}) := (m(\mathcal{B}), n(\mathcal{B}))$ where $m(\mathcal{B}) = q -$ rank (Θ) and $n(\mathcal{B})$ is the McMillan degree of Θ.

In this definition, $q - m(\mathcal{B})$ indicates the minimal number of independent laws which define $\mathcal{B}$, while $n(\mathcal{B})$ denotes the dimension of the space of initial conditions, i.e., the state space in any minimal state space representation of $\mathcal{B}$. We remark that $m(\mathcal{B})$ corresponds to the number of inputs (free variables) in any input-output representation of $\mathcal{B}$. See (Willems, 1991) for more details.

We will impose a lexicographic ordering on complexities. That is, $c(\mathcal{B}_1) \leq c(\mathcal{B}_2)$ if either $c(\mathcal{B}_1) = c(\mathcal{B}_2)$ or $m(\mathcal{B}_1) < m(\mathcal{B}_2)$ or, whenever $m(\mathcal{B}_1) = m(\mathcal{B}_2)$, $n(\mathcal{B}_1) < n(\mathcal{B}_2)$. In particular, for autonomous systems $\mathcal{B}$ this implies that $c(\mathcal{B}) = (0, N)$ where N equals the dimension of $\mathcal{B}$ as a subset of l_2^{+}.

When we look for models of low complexity then we have to accept a discrepancy between model and data. We formalize this by introducing a misfit function between the data and model. Recall from Theorem 2.2 that Θ is an exact model if and only if $\Gamma_W(\Theta) = \Pi_{+}\Theta W = 0$. Motivated by the definition of the Hankel operator Γ_W, a feasible choice (Antoulas, 1993) for a misfit function would be $\| \Gamma_W(\Theta) \|$, where $\| \cdot \|$ denotes the standard norm in $\mathcal{H}_2^{+}$. However, this misfit function does not take the non-uniqueness of kernel representations (2) of systems $\mathcal{B} \in \mathbf{B}$ into account. We therefore define a misfit function which is independent of specific parametrizations of the model.

Let $\mathcal{B} \in \mathbf{B} \cap l_2^{+}$ be an l_2^{+} system and denote by $\hat{\mathcal{B}}^{\perp}$ the orthogonal complement in $\mathcal{H}_2^{+}$ of the Laplace transform $\hat{\mathcal{B}}$ of $\mathcal{B}$.

Definition 3.2 The *misfit* between $\mathcal{B}$ and W is defined as

$$d(\mathcal{B}, W) := \sup \left\{ \frac{< Wx, v >}{\|v\|\,\|x\|} \mid v \in \hat{\mathcal{B}}^{\perp}, x \in \mathcal{H}_2^{-} \right\} \quad (5)$$

where the inner product and the norms are the usual ones in the Hilbert space $\mathcal{H}_2^{+}$.

This misfit function is motivated as follows. Note that in (5), the data matrix W is viewed as an *operator* mapping elements of $\mathcal{H}_2^{-}$ into elements of $\mathcal{L}_2$. Writing $W = \Pi_{-}W + \Pi_{+}W$, then the image of W is naturally decomposed in the orthogonal sets $\mathcal{W}_{-} := \Pi_{-}W\mathcal{H}_2^{-}$ and $\mathcal{W}_{+} := \Pi_{+}W\mathcal{H}_2^{-}$. Since $W \in \mathcal{H}_2^{+}$, it follows that $\mathcal{W}_{+}$ is finite dimensional and, in fact, $\mathcal{W}_{+} = \hat{\mathcal{B}}_{mpum}$. We can therefore rewrite (5) as

$$d(\mathcal{B}, W) = \sup \left\{ \frac{< w, v >}{\|v\|\,\|w\|_W} \mid v \in \hat{\mathcal{B}}^{\perp}, w \in \hat{\mathcal{B}}_{mpum} \right\}$$

where the inner product and the norm $\| \cdot \|$ are as before and

$$\| w \|_W := \inf \; \{ \; \| x \| \mid x \in \mathcal{H}_2^{-} \; w = \Pi_{+}Wx \; \}$$

is a 'data weighted' norm on elements of $\hat{\mathcal{B}}_{mpum}$.

We can interpret the misfit (5) as a measure of how far the principal axes of the data matrix W violate the laws $< w, v > = 0$ where $v \in \hat{\mathcal{B}}^{\perp}$ specify the laws which are implied by the model $\mathcal{B}$.

Note that $d(\mathcal{B}, W) \geq 0$ and $d(\mathcal{B}, W) = 0$ if and only if $\mathcal{B}$ is unfalsified by the data. Moreover, the misfit function of Definition 3.2 is independent of representations of systems $\mathcal{B} \in \mathbf{B} \cap l_2^+$.

As an example, let $q = 1$ and let the data be given by $\tilde{w}_1(t) = \lambda_1^t$, $\tilde{w}_2(t) = \epsilon\lambda_2^t$, where $|\lambda_1| < 1$, $|\lambda_2| < 1$, $t \in \mathbb{Z}_+$ and $\epsilon > 0$. Then $W(z) = [(z-\lambda_1)^{-1} \;\; \epsilon(z-\lambda_2)^{-1}]$ and the corresponding most powerful unfalsified model is obviously given by $\mathcal{B}_{mpum} = \text{span}\,\{\lambda_1^k, \lambda_2^k\}$. Clearly, for $\epsilon > 0$ small, we can no longer discriminate between the basis components $\tilde{w}_1 = \lambda_1^t$ and $\tilde{w}_2 = \epsilon\lambda_2^t$ which span $\mathcal{B}_{mpum}$. In fact, any first order approximate model $\mathcal{B} = \text{span}\,\{\lambda^t\}$ of $\mathcal{B}_{mpum}$ does not take the value $\epsilon > 0$ into account, whereas it seems logical to consider this value for defining a data weighted approximate model as in Definition 3.2.

The following result relates the misfit (5) to the Hankel norm of a specific operator.

Theorem 3.3 Let Θ be a *co-inner* kernel representation of a system $\mathcal{B} \in \mathbf{B} \cap l_2^+$. Then

$$d(\mathcal{B}, W) = \| \Pi_+ \Theta W \|_H$$

where $\| \cdot \|_H$ denotes the induced operator norm of the composite function $\Pi_+\Theta W$ viewed as a mapping from $\mathcal{H}_2^-$ to $\mathcal{H}_2^+$.

Consider dual spaces to obtain the following alternative expressions for the misfit.

Theorem 3.4 Let $\Theta \in \mathcal{H}_\infty^-$ be a co-inner kernel representation of $\mathcal{B}$. Then there holds

$$d(\mathcal{B}, W) = \|\Pi_- W^\sim \Theta^\sim\|_H = \|\Pi_- W^\sim\,|_{\mathcal{B}^\perp}\,\|_H$$

The second expression is to be interpreted as the induced operator norm of a mapping from $\mathcal{H}_2^+$ to $\mathcal{H}_2^-$. Using the fact that $\Theta^\sim$ is inner we obtain the right hand side expression where the operator norm is viewed as the induced norm of a mapping from $\hat{\mathcal{B}}^\perp$ to $\mathcal{H}_2^-$. Theorem 3.4 nicely shows that the misfit (5) is completely independent of specific parametrizations of $\mathcal{B}$.

The approximate modeling problem we will resolve in this paper is defined as follows. Given a data sequence represented by W, find a behavior $\mathcal{B} \in \mathbf{B}$ which minimizes the misfit $d(\mathcal{B}, W)$ over all behaviors of complexity less than or equal to $k > 0$

4. MAIN RESULTS

Our main result is as follows:

Theorem 4.1 Let $W \in \mathcal{H}_2^+$ be given by (4) and suppose that Assumption 2.1 holds. Let $\mathcal{B}_{mpum}$ be the most powerful unfalsified model for W. Suppose that $c(\mathcal{B}_{mpum}) = (0, n)$ and let $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_k \geq \ldots \geq \sigma_n > 0$ be the singular values of W. Let $W_k \in \mathcal{H}_2^+$ be an optimal Hankel norm approximation of W with McMillan degree k and define $\mathcal{B}_k$ to be the most powerful unfalsified model associated with the data W_k. Then

1. $\mathcal{B}_k$ is autonomous
2. $d(\mathcal{B}_k, W) = \sigma_{k+1}$
3. $\{c(\mathcal{B}_k') \leq (0,k)\} \Rightarrow \{d(\mathcal{B}_k', W) \geq d(\mathcal{B}_k, W)\}$.

Conclude from Theorem 4.1 that $\mathcal{B}_k$ is the argument of:

$$\min\{\, d(\mathcal{B}, W) \;\mid\; \mathcal{B} \in \mathbf{B} \cap l_2^+,\; c(\mathcal{B}) \leq (0,k) \,\}.$$

In other words, Theorem 4.1 gives a constructive method for the computation of an optimal approximate model of complexity $(0, k)$ when we use the distance function as defined by (5). For given data (1) and $k > 0$, this leads to the following conceptual procedure for obtaining optimal approximate models.

Algorithm:

1. Define the rational function $W \in \mathcal{H}_2^+$ according to (4).
2. Compute an optimal Hankel norm approximant W_k of W with McMillan degree k. See (Glover, 1984).
3. Let $W_k = \Theta_{mpum,k}^{-1}\Psi_k$ be a left coprime factorization over H_∞^- of W_k.
4. Put $\mathcal{B}_k = \mathcal{B}(\Theta_{mpum,k})$.

Proof of Theorem 4.1 Part 1 is one of the basic properties of the most powerful unfalsified model. Next, we focus on part 2 and 3. Let Θ_k be an inner kernel representation of $\mathcal{B}_k$. We obtain:

$$\begin{aligned} \|\Pi_+\Theta_k W\|_H &= \|\Pi_+\Theta_k(W - W_k)\|_H \\ &= \|\Pi_-(W^\sim - W_k^\sim)\Theta_k^\sim\|_H \\ &= \|\Pi_-(W^\sim - W_k^\sim)\,|_{\mathrm{Im}\,\Theta_k^\sim}\,\|_H \\ &\leq \|\Pi_-(W^\sim - W_k^\sim)\|_H \\ &= \sigma_{k+1} \end{aligned} \qquad (6)$$

Thus by Theorem 3.3 we have $d(\mathcal{B}_k, W) \leq \sigma_{k+1}$. Let $\hat{\Theta} \in \mathcal{H}_\infty^-$ be a co-inner kernel representation of some arbitrary autonomous behavior $\hat{\mathcal{B}}$ of complexity $c(\hat{\mathcal{B}}) \leq k$. Define $\Pi_{\hat{\mathcal{B}}}$ as the orthogonal projection onto $\hat{\mathcal{B}}$. Clearly, $\Pi_{\hat{\mathcal{B}}}$ has rank less than or equal to k. Furthermore,

$$\|\Pi_- W^\sim |_{\mathrm{Im}\ \hat{\Theta}^\sim} \|_H = \|\Pi_- W^\sim (I - \Pi_{\hat{\mathcal{B}}})\|_H$$

since $I - \Pi_{\hat{\mathcal{B}}}$ is an orthogonal projection onto $\mathrm{Im}\ \hat{\Theta}^\sim$. We get:

$$\begin{aligned} \|\Pi_+ \hat{\Theta} W\|_H &= \|\Pi_- W^\sim \hat{\Theta}^\sim\|_H \\ &= \|\Pi_- W^\sim (I - \Pi_{\hat{\mathcal{B}}})\|_H \\ &= \|\Pi_- W^\sim - \Pi_- W^\sim \Pi_{\hat{\mathcal{B}}}\|_H \\ &\geq \sigma_{k+1}. \end{aligned} \tag{7}$$

The last inequality follows from the fact that, since $\Pi_{\hat{\mathcal{B}}}$ has rank k, $\Pi_- W^\sim \Pi_{\hat{\mathcal{B}}}$ is, as a rank k approximant of $\Pi_- W^\sim$, never closer (in Hankel norm) than the optimal Hankel norm approximant. It is well known that the optimal Hankel norm approximant results in an error equal to σ_{k+1} and the result follows. (7) proves part 3 while (6) in conjunction with (7) for $\hat{\Theta} = \Theta_k$ proves part 2. ∎

5. CONCLUSIONS

In this paper we developed a method for optimal approximate modeling of time series. The misfit between model and data is defined in a representation independent way and we showed that the misfit criterion can be expressed in terms of the Hankel norm of an operator which is associated with the data.

We emphasize that the approximate modeling procedure discussed here is based on the assumption that the Laplace transform of the data is rational and belongs to $\mathcal{H}_2^+$. In the time domain, this includes polynomial-exponential data, impulse responses of linear time-invariant lumped systems and many other signals of practical interest. In particular, any finite length time-series or any finite set of frequency responses can be treated in the presented framework.

The optimal approximate modeling problem considered here involves the minimization of a misfit function subject to a complexity constraint on the models. As a second methodology, we may wish to specify a tolerated misfit and minimize the complexity of candidate models in $\mathbf{B}$. It is interesting to notice that this problem can entirely be solved in the context of autonomous systems.

We finally remark that the matrix W_k defined in Theorem 4.1 has the interpretation of an approximate data set. The method proposed here has therefore implications for problems related to data reduction.

6. REFERENCES

Antoulas, A.C. (1993). A new approach to modeling for control. Technical report. Rice University. To appear Lin, Alg. & Appl., Special Issue on Systems and Control, 1994.

Antoulas, A.C. and J.C. Willems (1991). A basis free approach to linear exact modeling. Technical Report 91-05. Rice University.

Antoulas, A.C. and J.C. Willems (1993). A behavioral approach to linear exact modeling. *IEEE Trans. Aut. Contr.* **38**(12), 1776–1802.

Glover, K. (1984). All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds. *Int. J. Contr.* **39**, 1115–1193.

Willems, J.C. (1986). From time series to linear systems, part I: Finite dimensional linear time invariant systems; part II: Exact modeling. *Automatica* **22**, 561–580, 675–694.

Willems, J.C. (1991). Paradigms and puzzles in the theory of dynamical systems. *IEEE Trans. Aut. Contr.* **36**(3), 259–294.