On the role of exact models in approximate modeling problems

Citation for published version (APA):

Stoorvogel, A. A., & Weiland, S. (1996). *On the role of exact models in approximate modeling problems*. (Measurement and control systems : internal report; Vol. 96I/04). Eindhoven University of Technology.

Document status and date:

Published: 01/01/1996

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Eindhoven University of Technology
Department of Electrical Engineering

On the Role of Exact Models in Approximate Modeling Problems

Siep Weiland and Anton A. Stoorvogel

August, 1996

**Measurement and Control Systems
Internal Report, 96 I/04**

Reprint of paper published in Proceedings of 35th Conference on Decision
and Control, Kobe, Japan, 1996.

Eindhoven, August 1996

On the Role of Exact Models in Approximate Modeling Problems

Siep Weiland¹

Department of Electrical Engineering
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands
E-mail: s.weiland@ele.tue.nl

Anton A. Stoorvogel

Department of Mathematics
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands
E-mail: wscoas@win.tue.nl

Abstract

The behavioral theory of dynamical system is used to address a deterministic system identification problem with a newly defined measure of misfit between data and linear time-invariant systems. An approximate model identification problem is formalized using this misfit criterium. In particular, Pareto optimal models are defined as feasible trade-offs between low complexity and low misfit models. The main result of this paper provides a complete characterization of bounded misfit and bounded complexity models. It is shown that this entire class of approximate models corresponds to the set of most powerful unfalsified models of reduced data sets. The reduced data sets are derived from Hankel norm approximations of the data. The main result therefore emphasizes the relevance of the exact modeling problem for the identification of approximate systems. The set of all Pareto optimal models is characterized as a simple consequence of this result.

Keywords

System identification, Approximate modeling, Hankel operators, Behavioral theory, Linear systems.

1 Introduction

A fundamental problem in system identification is to optimally model observed data. The most common formulations of this problem involve the parametrization of a set of models whose quality is evaluated on the basis of a criterion which expresses the discrepancy between data and model. There is a well developed theory

in which the complexity of models is expressed in terms of the dimension and structure of the parameter space and in which the accuracy of models is expressed as a measure of how well the model predicts the outputs as function of the inputs. In such a setting, optimal models correspond to optimal parameters and the system identification problem is basically a parameter estimation problem.

Since model classes allow many different parametrizations, it is natural to ask the question whether it is possible to separate the concept of optimal models from the parametrization issue. It is the aim of this paper to introduce concepts of complexity and misfit of dynamical systems which do not depend on model parametrizations.

In this paper, by a *model* we will mean a linear time-invariant dynamical system. *Optimal* will mean the best possible trade-off between the objectives of low complexity and high accuracy. We will formalize these concepts for a class of ℓ_2 discrete-time dynamical systems by introducing a notion of Pareto optimality. Roughly speaking, a Pareto optimal model is characterized by the property that every model of lower complexity is less accurate and every model of higher accuracy is more complex. It is in this sense that Pareto optimal models constitute a feasible compromise between the desirable objectives of low complexity and high accuracy.

The main result of this paper provides a characterization of Pareto optimal models. More specifically, we adopt the behavioral framework [7, 8, 9] to obtain a deterministic system identification problem in which system variables are treated in a symmetric way, without distinguishing between inputs and outputs. For a given data set, a key result of this paper (stated in Theorem 5.2 below) shows that the class of approximate models with bounded complexity and bounded misfit corresponds to the set of *exact models* for a related,

¹Part of this research has been made possible by a grant from the European Community for the Systems Identification and Modeling Network (SIMONET).

reduced data set. In particular, this result shows the relevance of the *exact modeling problem* for the identification of approximate systems. It is shown that Pareto optimal models are obtained as a simple corollary from this result (see Theorem 5.3 below).

The paper is organized as follows. Some notation is introduced in the next section. A formal problem statement is presented in Section 3. Section 4 treats representations of ℓ_2 systems. The main results are presented in Section 5 and further discussed in Section 6. In the latter section we also consider a number of generalizations of the misfit function treated in this paper. Conclusions are given in Section 7. Because of space limitations we do not include proofs of the main results in this paper. We refer to [12] for the proofs and a more extensive discussion of the results of this paper.

2 Notation

For $T \subseteq \mathbb{Z}$ we denote by $\ell_2(T, \mathbb{R}^q)$ the set of functions $w : T \rightarrow \mathbb{R}^q$ for which

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

is finite. We write ℓ_2^+ for $\ell_2(\mathbb{Z}_+, \mathbb{R}^q)$ and ℓ_2^- for $\ell_2(\mathbb{Z}_-, \mathbb{R}^q)$. $\mathcal{L}_2$ denotes the space of all complex valued functions which are square integrable on the unit disc $\mathbb{D} := \{z \in \mathbb{C} \mid |z| = 1\}$. The Hardy space $\mathcal{H}_2^+$ consists of all square integrable functions on $\mathbb{D}$ with analytic continuation outside the unit circle (including ∞). The orthogonal complement of $\mathcal{H}_2^+$ in $\mathcal{L}_2$ is denoted $\mathcal{H}_2^-$. $\mathcal{H}_\infty^+$ and $\mathcal{H}_\infty^-$ denote the Hardy spaces of complex valued functions which are bounded on the unit circle with analytic continuation in $|z| > 1$ and $|z| < 1$, respectively. The prefix $\mathcal{R}$ is used to denote rational elements of Hardy spaces. The canonical projections from $\mathcal{L}_2$ to $\mathcal{H}_2^+$ and $\mathcal{H}_2^-$ will be denoted by Π_+ and Π_- , respectively.

3 The identification problem

Consider a finite set of real valued multivariable time-series

$$w_i : \mathbb{Z}_+ \rightarrow \mathbb{R}^q, \quad i = 1, \dots, N \quad (3.1)$$

where $N > 0$ is the number of observed time-series and $q > 0$ is the dimension of the signal space. It will be assumed that $w_i \in \ell_2^+$ for $i = 1, \dots, N$. Using the Hilbert space isomorphism between ℓ_2^+ and $\mathcal{H}_2^+$ it follows that the data (3.1) can compactly be represented by the $\mathcal{H}_2^+$ function

$$W(z) := [\hat{w}_1(z) \quad \hat{w}_2(z) \quad \dots \quad \hat{w}_N(z)]$$

where $\hat{w}_i$ denotes the z -transform of w_i , i.e.,

$$\hat{w}_i(z) := \sum_{t=0}^{\infty} w_i(t) z^{-t}.$$

Using the behavioral formalism (see [7, 8, 9]) a (discrete time) dynamical system is a subset $\mathcal{B}$ of the set of all functions $w : \mathbb{Z}_+ \rightarrow \mathbb{R}^q$. Here, we will be interested in the class of ℓ_2 systems which is defined as follows.

Definition 3.1 (ℓ_2 -systems) A dynamical system $\mathcal{B}$ is an ℓ_2 -system if $\mathcal{B}$ is a closed¹ subset of ℓ_2^+ .

An ℓ_2 -system $\mathcal{B}$ is *linear* if $\mathcal{B}$ is a real linear subspace of ℓ_2^+ , it is *time-invariant* if $\sigma\mathcal{B} \subseteq \mathcal{B}$ where $[\sigma w](t) = w(t+1)$ and it is said to be *ℓ_2 -complete* if the collection of all finite truncations of system trajectories define the system, that is if

$$\{w \in \ell_2^+, w|_{[0,t]} \in \mathcal{B}|_{[0,t]}, \forall t \in \mathbb{Z}_+\} \Rightarrow \{w \in \mathcal{B}\}.$$

(cf. [9]). Introduce the model class

$$\mathbb{B}_2 := \{ \mathcal{B} \subseteq \ell_2^+ \mid \mathcal{B} \text{ is linear, time-invariant, and } \ell_2\text{-complete} \}.$$

Since ℓ_2 -completeness implies closedness of subspaces of ℓ_2 the set $\mathbb{B}_2$ defines a model class of ℓ_2 systems.

The interest in this class of models stems from the fact that elements in $\mathbb{B}_2$ admit simple representations in terms of *rational operators* (see Theorem 4.1 below) and, more importantly, elements in $\mathbb{B}_2$ can equivalently be considered in the frequency domain by defining, for each $\mathcal{B} \in \mathbb{B}_2$,

$$\hat{\mathcal{B}} := \{\hat{w} \in \mathcal{H}_2^+ \mid w \in \mathcal{B}\}$$

where $\hat{w}$ is the z -transform of w .

The complexity of a system is indicated by the number of independent time series which a system can conceive in time windows of finite length. We formalize this as follows.

Definition 3.2 (Complexity) The complexity of a system $\mathcal{B} \in \mathbb{B}_2$ is the pair of integers $c(\mathcal{B}) = (m, n)$ which satisfy

$$\dim(\mathcal{B}|_{[0,t-1]}) = mt + n \quad (3.2)$$

for all $t \geq n$.

The numbers m and n are well defined in this way ([5, 6, 8]). m corresponds to the dimension of the input-space of the system $\mathcal{B}$ and n equals the state space dimension in any minimal state space representation of $\mathcal{B}$. A *lexicographic ordering* is imposed on system complexities.

¹closed in the ℓ_2 topology.

Definition 3.3 (Misfit) The misfit between a model $\mathcal{B} \in \mathbb{B}_2$ and the data W is defined as

$$d(\mathcal{B}, W) := \sup \left\{ \frac{\langle W\lambda, v \rangle}{\|\lambda\|_2 \|v\|_2} \mid v \in \hat{\mathcal{B}}^\perp, \lambda \in \mathcal{H}_2^- \right\} \quad (3.3)$$

where the norms are the standard norms in the Hilbert spaces $\mathcal{H}_2^+$ and $\mathcal{H}_2^-$ and $\hat{\mathcal{B}}^\perp$ is the orthogonal complement of $\hat{\mathcal{B}}$ in $\mathcal{H}_2^+$.

This misfit function has the following interpretation. The set

$$\mathcal{B}_0 := \overline{\text{span}}\{\sigma^t w_i \mid i = 1, \dots, N; t \in \mathbb{Z}_+\} \quad (3.4)$$

defines a linear, time invariant ℓ_2 -system which is unfalsified by the data (3.1) in the sense that the data $w_i \in \mathcal{B}_0$ for $i = 1, \dots, N$. By associating with W a multiplicative operator $W : \mathcal{H}_2^- \rightarrow \mathcal{L}_2$, it is easily seen that $\hat{\mathcal{B}}_0$ is the closure of $\Pi_+ W \mathcal{H}_2^-$ and that the misfit $d(\mathcal{B}_0, W) = 0$. For any $\mathcal{B} \in \mathbb{B}_2$, the values of $\langle w, v \rangle$ with normalized elements $w \in \mathcal{B}_0$ and $v \in \mathcal{B}^\perp$ therefore indicate to what extent linear functionals implied by $\mathcal{B}$ fail to explain the unfalsified model $\mathcal{B}_0$ of the data. The misfit (3.3) can therefore equivalently be viewed as a distance measure between the exact model $\mathcal{B}_0$ and the ℓ_2 -system $\mathcal{B} \in \mathbb{B}_2$. To make this precise, consider $\mathcal{B}_0$ and equip $\hat{\mathcal{B}}_0$ with the *data weighted norm*

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

Then it is easy to show that

$$d(\mathcal{B}, W) = \sup \left\{ \frac{\langle w, v \rangle}{\|w\|_W \|v\|_2} \mid w \in \mathcal{B}_0, v \in \mathcal{B}^\perp \right\}. \quad (3.5)$$

The misfit therefore incorporates a weighting along the principal axes of the image of $\Pi_+ W$ under $\mathcal{H}_2^-$ and can be interpreted as a *data-weighted* distance measure between $\mathcal{B}_0$ and $\mathcal{B}$.

Remark 3.4 It is interesting to point out the relation between the expression (3.5) and the *gap* between the two closed subspaces $\mathcal{B} \subseteq \ell_2^+$ and $\mathcal{B}_0 \subseteq \ell_2^+$. The gap $[2, 3]$ is defined as

$$g(\mathcal{B}, \mathcal{B}_0) := \sup \left\{ \frac{\langle w, v \rangle}{\|w\|_2 \|v\|_2} \mid w \in \mathcal{B}_0, v \in \mathcal{B}^\perp \right\}. \quad (3.6)$$

Compared with (3.5), the gap has therefore a different normalization of elements in $\mathcal{B}_0$.

Remark 3.5 We would like to emphasize the importance of this data-weighting for applications of the misfit (3.3) in system identification. Indeed, if we consider

the example of an observed scalar valued time series $w(t) = \lambda_1^t + \epsilon \lambda_2^t$, with $t \in \mathbb{Z}_+$, $|\lambda_i| < 1$, $i = 1, 2$ and $\epsilon > 0$, then $w \in \ell_2^+$ and the linear span of the time series λ_1^t and λ_2^t defines the exact model $\mathcal{B}_0$ in (3.4). It has complexity $c(\mathcal{B}_0) = (0, 2)$. A lower order approximate model obtained by minimizing the gap (3.6) over all $\mathcal{B} \in \mathbb{B}_2$ with complexity $c(\mathcal{B}) = (0, 1)$ would therefore result in a model which does not depend on the value of ϵ . That is, for small $\epsilon > 0$, the relative weight of the component λ_1^t of the data sequence w is not discriminated by the gap criterion. However, in the misfit-criterion (3.3) it is.

Remark 3.6 Clearly, $d(\mathcal{B}, W) \geq 0$ and $d(\mathcal{B}, W) = 0$ if and only if $w_i \in \mathcal{B}$ for $i = 1, \dots, N$. Further, it is immediate from Definition 3.3 that $\mathcal{B}_1 \subseteq \mathcal{B}_2$ implies that $d(\mathcal{B}_1, W) \geq d(\mathcal{B}_2, W)$.

Definition 3.7 (Pareto optimality) A model $\mathcal{B} \in \mathbb{B}_2$ is called Pareto optimal for the data W if for all $\mathcal{B}' \in \mathbb{B}_2$ the implications

$$\begin{aligned} c(\mathcal{B}') \leq c(\mathcal{B}) &\Rightarrow d(\mathcal{B}', W) \geq d(\mathcal{B}, W) \\ d(\mathcal{B}', W) \leq d(\mathcal{B}, W) &\Rightarrow c(\mathcal{B}') \geq c(\mathcal{B}) \end{aligned}$$

hold simultaneously.

Hence, a model is Pareto optimal with respect to the data (3.1) if all models of smaller complexity have larger misfit, and all models of smaller misfit have larger complexity.

4 Kernel representations

Before presenting the main results we address in this section the problem of representing elements of the model class $\mathbb{B}_2$. It is shown in [10, 11] that $\mathbb{B}_2$ consists of those systems which can be represented as the kernel of a rational operator. We show here that the complexity and the misfit function introduced in the previous section admit simple characterizations in terms of kernel representations.

Let $\Theta \in \mathcal{RH}_\infty^-$. Then Θ defines a map $\Theta : \mathcal{H}_2^+ \rightarrow \mathcal{L}_2$ defined by the multiplication $[\Theta w](z) = \Theta(z)w(z)$. Associate with each such Θ a model

$$\begin{aligned} \hat{\mathcal{B}}_{\ker}(\Theta) &:= \{w \in \mathcal{H}_2^+ \mid \Theta w \in \mathcal{H}_2^-\} \\ &= \ker \Pi_+ \Theta. \end{aligned}$$

Let $\mathcal{B}_{\ker}(\Theta)$ denote its inverse z -transform. It is then easy to see that $\mathcal{B}_{\ker}(\Theta)$ belongs to $\mathbb{B}_2$. In fact, the following result claims that every model in $\mathbb{B}_2$ admits such a representation.

Theorem 4.1 (Kernel representation theorem) $\mathcal{B} \in \mathbb{B}_2$ if and only if there exists $\Theta \in \mathcal{RH}_\infty^-$ such

that $B = B_{\ker}(\Theta)$. Moreover, every $B \in \mathbb{B}_2$ can be represented as $B = B_{\ker}(\Theta)$ with $\Theta \in \mathcal{RH}_{\infty}^-$ co-inner, i.e. $\Theta\Theta^{\sim} = I$ where $\Theta^{\sim}(z) := \Theta^T(z^{-1})$.

Proof. See [10] or [11]. $\square$

The Hardy space $\mathcal{RH}_{\infty}^-$ therefore provides a parametrization of the model class $\mathbb{B}_2$. The next results completely characterize the complexity and misfit in terms of kernel representations.

Theorem 4.2 Let $\Theta \in \mathcal{RH}_{\infty}^-$ be a co-inner kernel representation of $B \in \mathbb{B}_2$. Then the following statements are equivalent

1. B has complexity $c(B) = (m, n)$.
2. $m = q - \text{rank}(\Theta)$ and $n = \deg(\Theta)$, where $\deg(\Theta)$ denotes the McMillan degree of Θ .

Proof. See [12]. $\square$

To characterize the misfit function, define the *Hankel norm* of $G \in \mathcal{L}_{\infty}$ as

$$\|G\|_H := \sup_{x \in \mathcal{H}_2^-} \frac{\|\Pi_+ Gx\|_2}{\|x\|_2}$$

Thus $\|G\|_H$ is the induced norm of the *Hankel operator* $\Gamma_G := \Pi_+ G \Pi_-$ acting on $\mathcal{H}_2^-$ as a multiplicative operator. Using this definition, the misfit function defined in 3.3 is characterized as follows.

Theorem 4.3 Let $\Theta \in \mathcal{RH}_{\infty}^-$ be a co-inner kernel representation of $B \in \mathbb{B}_2$. Then

$$d(B, W) = \|\Theta W\|_H$$

Proof. See [12]. $\square$

Hence, the misfit is equal to the Hankel norm of the $\mathcal{L}_{\infty}$ function ΘW .

5 Main results

The important notion of a most powerful unfalsified model was introduced by Willems (see e.g. [9]) and plays a crucial role in the sequel to characterize both exact and Pareto optimal approximate models for a given data set. An element $B \in \mathbb{B}_2$ is said to be an *unfalsified model* for $W \in \mathcal{H}_2^+$ if $d(B, W) = 0$. It is called the *most powerful unfalsified model* for W if in addition $B' \in \mathbb{B}_2$ with $d(B', W) = 0$ implies that $B \subseteq B'$. The most powerful unfalsified model is unique whenever it

exists and can be interpreted as the smallest ℓ_2 system that explains the data, i.e., the smallest covering of W by elements in $\mathbb{B}_2$. If it exists, we will write

$$B = \text{mpum}(W)$$

to denote the most powerful unfalsified model corresponding to W .

If the data W is *rational* then the most powerful unfalsified model actually exists and is easily obtained:

Theorem 5.1 (Exact models) If $W \in \mathcal{RH}_2^+$ then there exists a left coprime factorization $W = \Theta_0^{-1} \Psi_0$ over $\mathcal{RH}_{\infty}^-$. For any such factorization there holds that

1. $B_{\ker}(\Theta_0)$ is the most powerful unfalsified model for W .
2. $B_0 = B_{\ker}(\Theta_0)$.
3. $B_{\ker}(\Theta)$ is an unfalsified model for W if and only if $\Theta = \Lambda \Theta_0$ with $\Lambda \in \mathcal{RH}_{\infty}^-$.

Proof. See [12] or [1]. $\square$

Hence, all left-coprime factorizations of W define kernel representations of the most powerful unfalsified model. Moreover, all unfalsified models of W are parametrized by left multiplication of a left-coprime factor of W with rational operators in $\mathcal{RH}_{\infty}^-$. The second main result of this paper characterizes the set of all autonomous systems of given complexity and prescribed misfit level.

Theorem 5.2 (Approximate models) Let $B \in \mathbb{B}_2$ and $\sigma > 0$ be given and assume that $W \in \mathcal{RH}_2^+$. Let $\sigma_1 \geq \dots \geq \sigma_n > 0$ denote the ordered Hankel singular values of W . Then the following statements are equivalent

1. B has complexity $c(B) \leq (0, k)$ and misfit $d(B, W) \leq \sigma$.
2. $\sigma \geq \sigma_{k+1}$ and there exists $W_k \in \mathcal{RH}_2^+$ such that $\|W - W_k\|_H \leq \sigma$, $\deg(W_k) \leq k$ and $B = \text{mpum}(W_k)$.

Proof. See [12]. $\square$

Theorem 5.2 characterizes an entire set of feasible complexity and misfit combinations of dynamical systems for a given data set W . In words, it claims that the class of approximate autonomous models of the data set W is characterized as the class of *exact models* of data sets W_k which are obtained as Hankel norm approximants of W . We believe that this result has conceptually interesting consequences. Instead of finding

approximate models directly from data, this result suggests to determine a related data set, W_k , whose most powerful unfalsified model has guaranteed properties of complexity and misfit.

Pareto optimal models are easily derived using the characterization of Theorem 5.2. The following result shows that Pareto optimal models precisely correspond to the most powerful unfalsified models of optimal Hankel norm approximants of the data W .

Theorem 5.3 (Pareto optimal models) *Let $W \in \mathcal{RH}_2^+$ be given and let $\mathcal{B} \in \mathbb{B}_2$ have complexity $c(\mathcal{B}) = (0, k)$ and misfit $d(\mathcal{B}, W) = \sigma$. Assume moreover that the Hankel singular value $\sigma_k(W) > \sigma_{k+1}(W)$. Then the following statements are equivalent:*

1. $\mathcal{B}$ is Pareto optimal for the data W ,
2. $\sigma = \sigma_{k+1}$,
3. there exists an optimal Hankel norm approximant W_k of W of McMillan degree k such that $\mathcal{B} = \text{mpum}(W_k)$.

Proof. See [12]. □

6 Discussion and generalizations

Remark 6.1 The assumption on rationality of W may seem very restrictive at first sight. However, we remark that this assumption is satisfied for important signal sets like polynomial-exponential data, impulse responses of linear time-invariant lumped systems, finite support time-series or any finite set of frequency response samples. Also, the rationality of the Laplace transformed data is a common assumption in many problems in realization theory and the theory of rational interpolation.

Remark 6.2 In the presented setting no assumptions were made on noise characteristics. We emphasize that finite data sequences and finite number of frequency domain measurements always imply the rationality of W irrespective of noise models. Specific assumptions on noise are therefore not necessary.

Remark 6.3 A conceptual and constructive procedure for the computation of bounded complexity and bounded misfit models can be derived from the results of Section 5 and the results of Glover in [4] as follows. Given $W \in \mathcal{RH}_2^+$, $k > 0$ and $\sigma > 0$ one can compute the (ordered) Hankel singular values σ_i of W and verify whether $\sigma \geq \sigma_{k+1}$. If $\sigma_k > \sigma > \sigma_{k+1}$ define the augmented operator

$$W_\sigma(z) := \begin{pmatrix} W(z) & 0 \\ 0 & \sigma(z+1)/z \end{pmatrix},$$

otherwise take $W_\sigma = W$. Let W_σ^k be an optimal Hankel norm approximant of McMillan degree k to W_σ (see [4]) and let W_k be the truncation of W_σ^k to the augmented operator. W_k will then satisfy $\|W - W_k\|_H \leq \sigma$. Next, let $W_k = \Theta_k^{-1} \Psi_k$ be a left-coprime factorization of W_k over $\mathcal{RH}_\infty^-$ and put $\mathcal{B} = \mathcal{B}_{kr}(\Theta_k)$. Then $\mathcal{B}$ belongs to $\mathbb{B}_2$ and satisfies the criteria stated in Theorem 5.2.1. Moreover, if σ_k is strictly larger than σ_{k+1} then $\mathcal{B}$ is Pareto optimal if and only if $\sigma = \sigma_{k+1}$.

Remark 6.4 In Theorem 5.2 and Theorem 5.3 only autonomous models in $\mathbb{B}_2$ are considered. Similar results for non-autonomous systems have not been derived yet. However, since the definition of misfit warrants that $d(\mathcal{B}', W) \leq d(\mathcal{B}, W)$ whenever $\mathcal{B} \subseteq \mathcal{B}'$, we have that (in the notation of Theorem 5.2) any unfalsified model $\mathcal{B}'$ of W_k satisfies

$$d(\mathcal{B}', W) \leq \sigma.$$

Hence, any such model has a guaranteed upperbound on the misfit. In particular, this class of unfalsified models includes non-autonomous systems. Precisely, if $W_k = \Theta_k^{-1} \Psi_k$ is a left-coprime factorization over $\mathcal{RH}_\infty^-$ of W_k , then it follows from Theorem 5.1 and Theorem 4.2 that for any $\Lambda \in \mathcal{H}_\infty^-$ the system

$$\mathcal{B}' = \mathcal{B}_{ker}(\Lambda \Theta_k)$$

admits $m = q - \text{rank}(\Lambda \Theta_k)$ inputs while it has a guaranteed misfit level $d(\mathcal{B}', W) \leq \sigma$. Note that such a system may have a kernel representation with McMillan degree strictly smaller than k by choosing Λ appropriately.

Remark 6.5 The misfit function (3.3) is defined in terms of the standard inner product on $\mathcal{L}_2$ and involves normalization of λ and v in terms of their standard norms. As an alternative, consider the misfit criterion

$$d_\infty(\mathcal{B}, W) := \sup \left\{ \frac{\langle W\lambda, v \rangle}{\|\lambda\|_\infty \|v\|_2} \mid v \in \hat{\mathcal{B}}^\perp, \lambda \in \mathcal{H}_\infty^- \right\}. \quad (6.1)$$

Here, $\|\cdot\|_\infty$ denotes the $\mathcal{H}_\infty^-$ norm. In fact, if the data consists of one observed time series ($N = 1$), then the misfit criterion (6.1) is equal to the distance between W and $\hat{\mathcal{B}}$ measured in the $\mathcal{H}_2^+$ -norm. Precisely, if $N = 1$ then

$$d_\infty(\mathcal{B}, W) = \inf_{W' \in \hat{\mathcal{B}}} \|W - W'\|_2.$$

The latter criterion has been investigated in [5, 6] for the purpose of system identification.

Remark 6.6 As another generalization, we can replace the ℓ_2 setting of the identification problem treated in this paper by an ℓ_∞ framework. Let ℓ_∞ denote the real normed linear space of all vector valued, magnitude bounded sequences $w : \mathbb{Z}_+ \rightarrow \mathbb{R}^q$.

Let c_0 denote the subset of ℓ_∞ consisting of all sequences $w \in \ell_\infty$ which vanish in the limit, i.e. for which $\lim_{t \rightarrow \infty} w(t) = 0$. Assume that the data (3.1) belongs to c_0 and consider the model class $\mathbb{B}_\infty$ consisting of all linear time-invariant and complete subsets of c_0 . The 'orthogonal complement' $\mathcal{B}^\perp$ of $\mathcal{B} \in \mathbb{B}_\infty$ is defined as the set of all bounded continuous functionals which vanish at $\mathcal{B}$. Equipped with its induced norm, $\mathcal{B}^\perp$ becomes a subspace of the normed linear space ℓ_1 and we define the misfit function

$$d_1(\mathcal{B}, W) := \sup \left\{ \frac{\langle W * \lambda, v \rangle}{\|\lambda\|_1 \|v\|_1} \mid v \in \mathcal{B}^\perp, \lambda \in \ell_1^- \right\} \quad (6.2)$$

where $W := [w_1, \dots, w_N]$ is defined in the time-domain. We claim that the misfit (6.2) is again characterized as the induced norm of a Hankel operator. That is, if $\Theta \in \ell_1^-$ has norm $\|\Theta\|_1 = 1$, then Θ defines a normalized kernel representation of a system $\mathcal{B} \in \mathbb{B}_\infty$ by putting

$$\mathcal{B} = \mathcal{B}_{\ker}(\Theta) := \{w \in c_0 \mid [\Theta * w](t) = 0 \text{ for all } t \in \mathbb{Z}_+\}.$$

In this setting, the misfit is characterized as follows.

Theorem 6.7 For all $W \in c_0$ and $\mathcal{B} \in \mathbb{B}_\infty$

$$d_1(\mathcal{B}, W) = \|\Theta * W\|_{H_1}$$

where $\|\cdot\|_{H_1}$ denotes the ℓ_1 induced norm of the Hankel operator, i.e. if we associate with $G \in \ell_1(\mathbb{Z}, \mathbb{R}^{q \times q})$ the convolution operator $\mathcal{G} : \ell_1^- \rightarrow \ell_1^+$ defined as

$$\mathcal{G}(u) := G * u$$

then

$$\|\mathcal{G}\|_{H_1} := \sup_{u \in \ell_1^-} \frac{\|\mathcal{G}(u)\|_1}{\|u\|_1}.$$

7 Conclusions

For a model class of ℓ_2 dynamical systems we proposed definitions for misfit and system complexity which were shown to have simple characterizations in terms of rational kernel representations. These characterizations have been used to parameterize a class of optimal approximate models as exact models of related data sets. Pareto optimal models were defined and it has been shown how a set of Pareto optimal models can be obtained by determining optimal Hankel norm approximants derived from the data. In particular, these Pareto optimal models are characterized as the most powerful unfalsified models of Hankel reduced data sets. A conceptual procedure for the computation of these models has been given and various generalizations of the misfit criterion were discussed.

References

- [1] A.C. ANTOUNAS, J.C. WILLEMS, "A Behavioral Approach to Linear Exact Modeling," IEEE Transactions on Automatic Control, Vol. 38, pp. 1776-1802, 1993.
- [2] A.K. EL-SAKKARY, "The Gap Metric: Robustness of Stabilization of Feedback Systems," IEEE Trans. Automatic Control, Vol. AC-30, pp. 240-247, 1985.
- [3] T.T. GEORGIU and M.C. SMITH, "Optimal Robustness in the Gap Metric," IEEE Transactions on Automatic Control, Vol. 35, pp. 673-685, 1990.
- [4] K. GLOVER, "All Optimal Hankel Norm Approximations of Linear Multivariable Systems and their L_∞ Bound," Int. Journal of Control, Vol. 39, pp. 1115-1193, 1984.
- [5] B. ROORDA, "Global Total Least Squares: a method for the construction of open approximate models from vector time series", Ph.D. Thesis, Tinbergen Institute Research Series, Rotterdam, the Netherlands, Vol. 88, 1995.
- [6] B. ROORDA AND C. HEIJ, "Global Total Least Squares Modeling of Multivariable Time Series", IEEE Transaction on Automatic Control, Vol. 40:1, pp. 50-63, 1995.
- [7] J.C. WILLEMS, "From time Series to Linear System; Part I: Finite Dimensional Linear Time Invariant Systems"; Part II: "Exact Modeling," Automatica, Vol. 22, no. 5, pp. 561-580, resp. Vol. 22, no. 6, pp. 675-694, 1986.
- [8] J.C. WILLEMS, "From Time Series to Linear System; Part III: Approximate Modelling," Automatica, Vol. 23, no. 1, pp.87-115, 1987.
- [9] J.C. WILLEMS, "Paradigms and Puzzles in the Theory of Dynamical Systems," IEEE Trans. Automatic Control, Vol. 36, No. 3, pp. 259-294, 1991.
- [10] S. WEILAND AND A.A. STOORVOGEL, "Rational Representations of Behaviours: Well-posedness, Stability, and Stabilizability", Submitted to Math. Contr. Sign. and Syst., 1996.
- [11] S. WEILAND AND A.A. STOORVOGEL, "On Representing Behaviors in the Frequency Domain," Proceedings 13th IFAC World Congress, Vol. I, pp. 7-12, San Francisco, June 30-July 5, 1996.
- [12] S. WEILAND AND A.A. STOORVOGEL, "Optimal Hankel Norm Identification of Dynamical Systems", Submitted to Automatica, 1996.