On continuity of l-infinite optimal models

Eindhoven University of Technology
Department of Electrical Engineering

On Continuity of ℓ_∞ Optimal Models

Siep Weiland, Wolfgang Scherrer
and Manfred Deistler

October 1996

Measurement and Control Systems
Internal Report, 96 I/08

Paper submitted to International Journal on Mathematical Systems, Estimation and Control

Eindhoven, October 1996

On Continuity of ℓ_∞ Optimal Models*

Siep Weiland[†] **W. Scherrer**[‡] **M. Deistler**[‡]

Paper submitted to the International Journal on Mathematical Systems, Estimation and Control
October 1996

Abstract

This paper concerns the question of continuity of the mapping from observed time series to models. The behavioral framework is adopted to formalize a model identification problem in which the observed time series is decomposed into a part explained by a model and a remaining part which is ascribed to noise. Neither the observed time series nor the set of candidate models are assumed to have an input-output structure. The misfit between data and model is defined symmetric in the system variables and measured in the ℓ_∞ or amplitude norm. It is shown that the misfit function continuously depends on the data and the identified model. The consequences of the continuity of the misfit function for optimal and suboptimal models are discussed.

Keywords
System identification, Linear systems, Continuity, Behaviors

1 Introduction

The central question in system identification amounts to finding models which give a "good" description of the observed data. An important issue in the qualitative study of system identification procedures is the continuity of identified models as function of the observed data. This question is particularly relevant for the issue of well-posedness and robustness of identification algorithms with respect to variations in the observed data. Also, many model validation techniques assess the quality of identified models on the basis of large numbers of data sets and therefore implicitly test the continuity of identified models as function of the data.

In this paper we analyze a system identification problem in which neither the observed data nor the class of candidate models are required to have an input-output structure. One of the advantages of such a formalism is that system variables are treated in a symmetric way, without making a distinction between input and output variables. We introduce a notion of misfit between model and data in which the noise part of the observed data is measured in the ℓ_∞ (or amplitude) norm. The main result of the paper shows that the misfit function is continuous in both the data as well as the model provided suitable topologies are defined on the set of data sequences and the class of candidate models.

*Research supported by the European Research Network on System Identification (ERNSI)

†Corresponding author; Department of Electrical Engineering; Eindhoven University of Technology; P.O. Box 513; 5600 MB Eindhoven; The Netherlands.

‡Institut für Ökonometrie, Operations Research und Systemtheorie; Technische Universität Wien; Argentinierstraße 8; A-1040 Vienna; Austria.

The issue of continuity and consistency of models has been a main topic of research in the area of system identification. See e.g. [4]. The results presented in this paper closely resemble recent work in [5] where a stochastic approach is taken to investigate continuity and consistency properties of dynamic factor models. In this paper we investigate amplitude norms for the purpose of system identification and provide weaker topological conditions for the continuity of misfit functions. The consequences of the continuity results for optimal ℓ_∞ models are discussed.

The paper is organized as follows. In section 2 we formalize the ℓ_∞ optimal identification problem. Section 3 considers the duality between models and their laws by using orthogonal complements of subspaces of Banach spaces. The main results are collected in Section 4 and conclusions are defered to section 5.

Notation

Let $\mathbb{Z}_+$ denote the set of non-negative integers and suppose that q is a positive integer. Then we define

- ℓ_∞^q the real normed linear space of all vector valued, magnitude bounded sequences $w : \mathbb{Z}_+ \to \mathbb{R}^q$ with norm
$$\| w \|_\infty := \max_{i\in\{1,\dots,q\}} \sup_{t\in\mathbb{Z}_+} |w_i(t)|.$$
- c_0^q the subspace of ℓ_∞^q consisting of all sequences $w \in \ell_\infty^q$ which vanish in the limit, i.e.,
$$c_0^q := \left\{ w \in \ell_\infty^q \mid \lim_{t\to\infty} w(t) = 0 \right\}.$$
- ℓ_1^q the real normed linear space of all vector valued sequences $w : \mathbb{Z}_+ \to \mathbb{R}^q$ whose entries are absolutely summable functions. The norm is defined
$$\| w \|_1 := \sum_i^q \sum_{t=0}^\infty |w_i(t)|.$$

We will drop the integer q when dimensions are clear from the context. It is well known that ℓ_∞, c_0 and ℓ_1 are Banach spaces and that c_0 is a closed linear subspace of ℓ_∞. The prefix U will be used to indicate the closed unit sphere of normed linear spaces, e.g. $U\ell_\infty := \{w \in \ell_\infty \mid \|w\|_\infty \leqslant 1\}$.

2 The identification problem

Let a multivariable time series
$$w : \mathbb{Z}_+ \to \mathbb{R}^q$$
be observed. Throughout, it will be assumed that $w(t)$ converges to 0 as $t \to \infty$, i.e., we assume that $w \in c_0^q$. The identification problem amounts to finding linear models which explain the data w up to some level of accuracy. Here, by model (or *system*) we mean any collection $\mathcal{B}$ of time series mapping the (discrete) time set $\mathbb{Z}_+$ to the real valued signal space $\mathbb{R}^q$. A model $\mathcal{B}$ is called *linear* if $\mathcal{B}$ is a real linear subspace of $(\mathbb{R}^q)^{\mathbb{Z}_+}$. It is called *time-invariant* if $\sigma\mathcal{B} \subseteq \mathcal{B}$ where σ is the left-shift defined for given $v : \mathbb{Z}_+ \to \mathbb{R}^q$ as $(\sigma v)(t) := v(t+1)$. See [6, 7] for more details on the behavioral framwork. We will be particularly interested in ℓ_∞-systems which are defined as follows.

Definition 2.1 An ℓ_∞-system is a linear, time-invariant and closed subset $\mathcal{B}$ of c_0^q.

For a given ℓ_∞-system $\mathcal{B}$ the observed time series w admits a decomposition

$$w = \hat{w} + \tilde{w}, \text{ where } \hat{w} \in \mathcal{B}. \tag{2.1}$$

Here, $\hat{w}$ represents a component of the observed time series which can be explained by the ℓ_∞-system $\mathcal{B}$ in the sense that $\hat{w} \in \mathcal{B}$. The time series $\tilde{w}$ is an unexplained part of the observed data w and represents the noise. The trajectory $\tilde{w}$ will therefore be interpreted as a component of the observed time series which results from the error $w - \hat{w}$ due to the approximation of w by a trajectory $\hat{w} \in \mathcal{B}$. Obviously, the decomposition (2.1) is non-unique. For a given ℓ_∞-system $\mathcal{B}$, we will call $\hat{w} \in \mathcal{B}$ an *optimal approximant of* w if the noise or error signal $\tilde{w} := w - \hat{w}$ is minimal in some norm. In this paper we focus on the amplitude norm of the noise or error signal $\tilde{w}$.

Definition 2.2 (Misfit) The *misfit* between a time series $w \in c_0^q$ and the ℓ_∞-system $\mathcal{B}$ is

$$\mu(w, \mathcal{B}) := \inf_{\hat{w} \in \mathcal{B}} \| w - \hat{w} \|_\infty \tag{2.2}$$

The misfit is therefore expressed in terms of the distance between the data point w and the element $\hat{w} \in \mathcal{B}$ (if it exists) which is closest in the ℓ_∞ sense to w. Note that, because $\mathcal{B}$ is a closed subset of c_0^q the misfit $\mu(w, \mathcal{B}) = 0$ if and only if $w \in \mathcal{B}$. In that case $\mathcal{B}$ is said to be an *unfalsified model* for the data w and the noise component $\tilde{w} = 0$ for some decomposition of the form (2.1).

Let $\mathbb{B}$ denote a class of ℓ_∞-systems. Given the data w, the identification problem amounts to finding those models $\mathcal{B} \in \mathbb{B}$ which minimize the misfit $\mu(w, \mathcal{B})$, i.e. we wish to find optimal models

$$\mathbb{B}^{\text{opt}}(w) := \arg\min_{\mathcal{B} \in \mathbb{B}} \mu(w, \mathcal{B}). \tag{2.3}$$

Note that $\mathbb{B}^{\text{opt}}(w)$ may be empty, as the minimum in (2.3) need not exist. A suboptimal version of this problem amounts to characterizing all models $\mathcal{B} \in \mathbb{B}$ which have a guaranteed misfit level. For $\varepsilon \geqslant 0$ we define the *level set*

$$\mathbb{B}(\varepsilon, w) := \{ \mathcal{B} \in \mathbb{B} \mid \mu(w, \mathcal{B}) \leqslant \varepsilon \}. \tag{2.4}$$

It is clear that $\mathbb{B}(\varepsilon, w)$ is empty if $\varepsilon < \varepsilon^{\text{opt}}(w)$ where

$$\varepsilon^{\text{opt}}(w) := \inf_{\mathcal{B} \in \mathbb{B}} \mu(w, \mathcal{B})$$

is the optimal misfit level for the data w. In what follows we will be interested in the continuity properties of the misfit map $\mu : c_0^q \times \mathbb{B} \to \mathbb{R}$.

3 Duality

For the analysis of consistency and continuity we exploit the interrelations between a normed linear space and its corresponding dual. For a normed linear space $\mathcal{X}$, its dual will be denoted $\mathcal{X}^*$ and consists of all bounded linear functionals on $\mathcal{X}$. $\mathcal{X}^*$ is a complete normed space when equiped with the usual definitions of addition and scalar multiplication of linear functionals. The value of a bounded linear functional x^* at $x \in \mathcal{X}$ is denoted $\langle x, x^* \rangle$. Its norm is defined as

$$\| x^* \| := \sup_{\|x\| \leqslant 1} | \langle x, x^* \rangle |.$$

It is well known [1, 2] that for any $q > 0$ the normed space ℓ_1^q is the dual of c_0^q and ℓ_∞^q is the dual of ℓ_1^q, i.e. $\ell_1^q = (c_0^q)^*$ and $\ell_\infty = (\ell_1^q)^*$. Borrowing terminology from the theory of Hilbert spaces, we call $w \in c_0^q$ and $r \in \ell_1^q$ *orthogonal* if $\langle w, r \rangle = 0$. This induces the following notions of orthogonal complements of subsets in c_0^q and ℓ_1^q.

Definition 3.1 The *orthogonal complement* of a subspace $\mathcal{B} \subseteq c_0^q$ is the set

$$\mathcal{B}^\perp := \left\{ r \in \ell_1^q \mid \langle w, r \rangle = 0 \text{ for all } w \in \mathcal{B} \right\}. \tag{3.1}$$

The *orthogonal complement* of a subspace $\mathcal{L}$ of ℓ_1^q is the set

$${}^\perp\mathcal{L} := \left\{ w \in c_0^q \mid \langle w, r \rangle = 0 \text{ for all } r \in \mathcal{L} \right\}. \tag{3.2}$$

Remark 3.2 A subspace $\mathcal{L} \subset \ell_1^q$ defines a second orthogonal complement $\mathcal{L}^\perp$ in ℓ_∞^q by putting

$$\mathcal{L}^\perp := \left\{ v \in \ell_\infty^q \mid \langle v, r \rangle = 0 \right\}.$$

We will however not use this subspace.

Definition 3.3 The *laws* of a model $\mathcal{B} \subseteq c_0^q$ are the elements of the orthogonal complement $\mathcal{L} := \mathcal{B}^\perp$.

Hence, every model $\mathcal{B} \subseteq c_0^q$ uniquely defines a set of laws. Conversely, a set of laws $\mathcal{L} \subseteq \ell_1^q$ defines a model $\mathcal{B} := {}^\perp\mathcal{L}$. Since ${}^\perp(\mathcal{B}^\perp)$ may be a proper superset of $\mathcal{B}$, the laws $\mathcal{L} := \mathcal{B}^\perp$ need not uniquely define $\mathcal{B}$. For *closed* subsets of c_0^q this is however the case (see [3, 1]) and thus we have

Proposition 3.4 *For an ℓ_∞-system $\mathcal{B}$, the laws $\mathcal{L} := \mathcal{B}^\perp$ uniquely define $\mathcal{B}$ in the sense that $\mathcal{B} = {}^\perp\mathcal{L}$.*

The following result is standard [3, 2] and characterizes the misfit in terms of dual spaces.

Theorem 3.5 *The misfit between $w \in c_0^q$ and an ℓ_∞-system $\mathcal{B}$ satisfies*

$$\mu(w, \mathcal{B}) := \inf_{\hat{w} \in \mathcal{B}} \| w - \hat{w} \|_\infty = \max_{r \in U\mathcal{B}^\perp} \langle w, r \rangle \tag{3.3}$$

where the maximum is achieved for some $r^{opt} \in U\mathcal{B}^\perp$.

4 Main results

In this section we provide the main technical result of this paper which states that the misfit function is jointly continuous in the observed time series and the model. Convergence of time series is defined in the strong ℓ_∞ sense as follows. A sequence $\{w_n\}_{n=0}^\infty$ of elements $w_n \in c_0^q$ is said to converge to an element $w \in c_0^q$ if $\| w - w_n \|_\infty \to 0$ as $n \to \infty$. A sequence of ℓ_∞-systems $\{\mathcal{B}_n\}_{n=0}^\infty$ is said to converge to an ℓ_∞-system $\mathcal{B}$ if the following two conditions hold

1. For any sequence of laws $r_n \in U\mathcal{B}_n^\perp$ there exists a sequence of laws $r_n^0 \in U\mathcal{B}^\perp$ such that for all $w \in c_0^q$ the functional

   $$\langle w, r_n - r_n^0 \rangle \longrightarrow 0$$

   as $n \to \infty$.
2. For all laws $r^0 \in U\mathcal{B}^\perp$ there exists a sequence of laws $r_n \in U\mathcal{B}_n^\perp$ such that for all $w \in c_0^q$

   $$\langle w, r^0 - r_n \rangle \longrightarrow 0$$

   as $n \to \infty$.

In words, the first requirement says that for all sequences of (normalized) laws $r_n \in \mathcal{B}_n^\perp$ of $\mathcal{B}_n$ there exists a sequence of laws r_n^0 of the limiting model $\mathcal{B}$ such that $r_n - r_n^0$ *converges in the weak-star sense* to $0 \in \ell_1$ as $n \to \infty$. This means that the weak-star limit of all laws of $\mathcal{B}_n$ constitute a subset of the laws of the limiting model $\mathcal{B}$. It is easily seen that the first requirement is equivalent to

1.' For all sequences $\{r_n\}_{n\in\mathbb{Z}_+}$, $r_n \in \mathcal{B}_n^\perp$ for which there exists an $r \in \ell_1^q$ such that for all $w \in c_0^q$

$$\langle w, r_n - r\rangle \longrightarrow 0$$

as $n \to 0$, there holds $r \in \mathcal{B}^\perp$.

Similarly, the second requirement states that for all laws in the limit model $\mathcal{B}$ there exist a sequence of laws of $\mathcal{B}_n$ that converges to it in the weak-star sense. Model convergence is therefore expressed in terms of weak-star convergence of laws.

Definition 4.1 (Sequential continuity) A map $f : \mathcal{X} \to \mathbb{R}$ is called *sequential continuous* if for all $x \in \mathcal{X}$ and all sequences $\{x_n\}_{n\in\mathbb{Z}_+}$ with $x_n \to x$ there holds $f(x_n) \to f(x)$.

Remark 4.2 It is emphasized that we only introduced a notion of convergence for systems, but not a topology on the set of all ℓ_∞-systems. For this reason continuity of maps is defined via convergent sequences.

The main result of this section is as follows.

Theorem 4.3 *The misfit function $\mu(w, \mathcal{B})$ is (sequential) continuous in its arguments $(w, \mathcal{B})$.*

Proof. Let $w \in c_0^q$, $\mathcal{B}$ be an ℓ_∞-system and let $\{w_n\}_{n\in\mathbb{Z}_+}$ and $\{\mathcal{B}_n\}_{n\in\mathbb{Z}_+}$ be sequences of time series in c_0^q and ℓ_∞-systems such that $(w_n, \mathcal{B}_n) \to (w, \mathcal{B})$. Then, using theorem 3.5, we find that

$$\begin{aligned}\mu(w_n, \mathcal{B}_n) - \mu(w, \mathcal{B}) &= \max_{r_n\in U\mathcal{B}_n^\perp}\langle w_n, r_n\rangle - \max_{r\in U\mathcal{B}^\perp}\langle w, r\rangle\\ &= \langle w_n, r_n^{\text{opt}}\rangle - \max_{r\in U\mathcal{B}^\perp}\langle w, r\rangle\\ &\leqslant \langle w_n - w + w, r_n^{\text{opt}}\rangle - \langle w, r_n^0\rangle\\ &= \langle w_n - w, r_n^{\text{opt}}\rangle + \langle w, r_n^{\text{opt}} - r_n^0\rangle.\end{aligned}$$

Here, $r_n^{\text{opt}} \in U\mathcal{B}_n^\perp$ is such that $\langle w_n, r_n^{\text{opt}}\rangle = \max_{r_n\in U\mathcal{B}_n^\perp}\langle w_n, r_n\rangle$ and $r_n^0 \in U\mathcal{B}^\perp$ is the corresponding sequence of laws satisfying requirement 1 for convergence of models. Further, we used that $\max_{r\in U\mathcal{B}^\perp}\langle w, r\rangle \geqslant \langle w, r_n^0\rangle$. Using boundedness of $\| r_n^{\text{opt}} \|_1$ and the fact that $\|w_n - w\|_\infty$ converges to 0 as $n \to \infty$ it follows that the latter expression converges to 0 as $n \to \infty$.

Similarly, we obtain that

$$\begin{aligned}\mu(w, \mathcal{B}) - \mu(w_n, \mathcal{B}_n) &= \max_{r\in U\mathcal{B}^\perp}\langle w, r\rangle - \max_{r_n\in U\mathcal{B}_n^\perp}\langle w_n, r_n\rangle\\ &= \langle w, r^{\text{opt}}\rangle - \max_{r_n\in U\mathcal{B}_n^\perp}\langle w_n, r_n\rangle\\ &\leqslant \langle w, r^{\text{opt}}\rangle - \langle w_n, r_n^0\rangle\\ &= \langle w, r^{\text{opt}} - r_n^0\rangle + \langle w - w_n, r_n^0\rangle.\end{aligned}$$

Here, $r^{\text{opt}} \in U\mathcal{B}^\perp$ is such that $\langle w, r^{\text{opt}}\rangle = \max_{r\in U\mathcal{B}^\perp}\langle w, r\rangle$ and $r_n^0 \in U\mathcal{B}_n^\perp$ is the corresponding sequence of laws satisfying requirement 2 for convergence of models. Since r_n^0 has bounded norm, and $\langle w, r^{\text{opt}} - r_n^0\rangle$ converges to 0 as $n \to \infty$ we obtain that the latter expression vanishes in the limit. Consequently,

$$\lim_{n\to\infty} |\mu(w, \mathcal{B}) - \mu(w_n, \mathcal{B}_n)| = 0$$

which proves the sequential continuity of μ. $\square$

Theorem 4.3 implies that small amplitude variations in the observed data imply small perturbations of the misfit function. Finite length observations of the data w are particularly relevant in this context. With $w \in c_0^q$, a truncated observation

$$w_n(t) := \begin{cases} w(t) & t \leqslant n \\ 0 & t > n \end{cases} \tag{4.1}$$

clearly satisfies $w_n \to w$. The following results are an immediate consequence of the continuity of $\mu(w, \mathcal{B})$. Let $\varepsilon \geqslant 0$ and consider the level set $\mathbb{B}(\varepsilon, w)$ as defined in (2.4).

Definition 4.4 (Upper semicontinuity) Let $\mathbb{B}$ be a model class of ℓ_∞-systems. The level set $\mathbb{B}(\varepsilon, w) \subseteq \mathbb{B}$ is called *upper semicontinuous* if

$$\{ (\varepsilon_n, w_n, \mathcal{B}_n) \to (\varepsilon, w, \mathcal{B}),\ \mathcal{B}_n \in \mathbb{B}(\varepsilon_n, w_n) \} \Longrightarrow \{ \mathcal{B} \in \mathbb{B}(\varepsilon, w) \}$$

In words, this property guarantees that if elements of level sets $\mathbb{B}(\varepsilon_n, w_n)$ converge, then they converge to elements of the limiting level set $\mathbb{B}(\varepsilon, w)$. In particular, if $w \in c_0^q$ and $\{\varepsilon_n\}_{n=0}^\infty$ is a sequence of non-negative numbers converging to $\varepsilon = \varepsilon^{\text{opt}}(w)$, then upper semicontinuity of $\mathbb{B}(\varepsilon, w)$ implies that any convergent sequence of ℓ_∞-systems $\mathcal{B}_n \in \mathbb{B}(\varepsilon_n, w_n)$ converges to an optimal model $\mathcal{B} \in \mathbb{B}^{\text{opt}}(w)$.

Proposition 4.5 *The level set $\mathbb{B}(\varepsilon, w)$ is upper semicontinuous.*

Proof. Let $(\varepsilon_n, w_n, \mathcal{B}_n) \to (\varepsilon, w, \mathcal{B})$ and let $\mathcal{B}_n \in \mathbb{B}(\varepsilon_n, w_n)$. By continuity of the misfit we have that $\mu(w_n, \mathcal{B}_n) \to \mu(w, \mathcal{B})$. As $\mu(w_n, \mathcal{B}_n) \leqslant \varepsilon_n \to \varepsilon$ it follows that $\mu(w, \mathcal{B}) \leqslant \varepsilon$. $\square$

The following result shows that the optimal misfit level $\varepsilon^{\text{opt}}(w)$ is a continuous function of the data. The result is a consequence of Theorem 4.3 and Proposition 4.5

Proposition 4.6 *Let the model class $\mathbb{B}$ be sequentially compact (in the sense that every sequence $\{\mathcal{B}_n\}_{n=0}^\infty$ with $\mathcal{B}_n \in \mathbb{B}$ has a convergent subsequence with limiting element in $\mathbb{B}$). Then $\varepsilon^{\text{opt}}(w)$ is a sequentially continuous function of w.*

Remark 4.7 Suppose that the data w is compatible with a model $\mathcal{B}_0 \in \mathbb{B}$ in the sense that $\mu(w, \mathcal{B}_0) = 0$ and further suppose that $\mathcal{B}_0$ is the unique element in $\mathbb{B}$ with this property. A natural question is whether the model $\mathcal{B}_0$ can be identified in a consistent way from finite length observations w_n, as defined in (4.1), by minimizing the misfit function $\mu(w_n, \mathcal{B})$ over elements $\mathcal{B} \in \mathbb{B}$. If the model set $\mathbb{B}$ is sequentially compact then the set of optimal models

$$\mathbb{B}^{\text{opt}}(w_n) := \arg\min_{\mathcal{B}\in\mathbb{B}} \mu(w_n, \mathcal{B})$$

is non-empty. Furthermore, using the upper semicontinuity of the level sets (Proposition 4.5) and the continuity of the optimal misfit levels (Proposition 4.6) it follows that every convergent sequence of optimal models $\mathcal{B}_n \in \mathbb{B}^{\text{opt}}(w_n)$ converges to $\mathcal{B}_0$. This property of convergence of optimal models can be viewed as a deterministic notion of consistency.

Remark 4.8 It follows that sequential compactness of model sets $\mathbb{B}$ is a relevant property to guarantee consistency and continuity of optimal misfit levels. One of the most common model sets is defined as follows, see e.g. [6, 7]. Associate with a law $r \in \ell_1^q$, the set

$$\mathcal{B}(r) := \{w \in c_0^q \mid \langle \sigma^t w, r\rangle = 0 \text{ for all } t \geqslant 0\}.$$

Clearly, $\mathcal{B}(r)$ is an ℓ_∞-system in the sense of definition 2.1. Let $r_1, \ldots, r_g \in \ell_1^q$ be a set of laws which are assumed to be independent and of finite support i.e. $\exists N > 0$ such that $r_i(t) = 0$ for $t \geqslant N$, $i = 1, \ldots, g$. We associate with these laws the system

$$\mathcal{B} := \bigcap_{i=1}^{g} \mathcal{B}(r_i). \tag{4.2}$$

Since r_i has finite support we can introduce the polynomial $p_i(z) := \sum_{t=0}^{N} r_i(t) z^t$. Then $p_i(\sigma)$ denotes a polynomial difference operator and we have

$$\mathcal{B}(r_i) = \{w \in c_0^q \mid p_i(\sigma)w = 0\} = \ker p_i(\sigma)$$

Introduce the matrix polynomial

$$P(z) = \begin{pmatrix} p_1(z) \\ \vdots \\ p_g(z) \end{pmatrix}.$$

Then the system (4.2) is represented by the autoregressive equation $P(\sigma)w = 0$, i.e.

$$\mathcal{B} = \mathcal{B}(P) := \{w \in c_0^q \mid P(\sigma)w = 0\}$$

For given integers g and n, the model set $\mathbb{B}(g, n)$ is defined as the set of all models $\mathcal{B}(P)$ where the matrix polynomial P has rank g and McMillan degree $\leqslant n$. It is an open question whether this model set is sequentially compact with the notion of convergence as defined in Section 4.

5 Conclusions

A system identification problem has been addressed in which optimal models are defined as those elements in a model class of linear systems which minimize the ℓ_∞ distance to the observed data. Models are defined in terms of families of system trajectories and are not required to have an input-output structure. A misfit function is proposed which measures the amplitude of the approximation error of the model with respect to the data and it is shown that the misfit continuously depends on the data and the model. These results are imperative to prove that ℓ_∞ optimal models also continuously depend on the observed data.

References

[1] K. Jänich, "Topologie", Springer Verlag, New York, 1980.

[2] G. Köthe, "Topological Vector Spaces I", Springer Verlag, Heidelberg, 1969.

[3] D.G. Luenberger, "Optimization by Vector Space Methods", John Wiley & Sons, New York, 1969.

[4] E.J. Hannan and M. Deistler, "The Statistical Theory of Linear Systems", John Wiley & Sons, New York, 1988.

[5] C. Heij, W. Scherrer, M. Deistler, "System Indentification by Dynamic Factor Models," Internal report, Institut für Ökonometrie, Operations Research und Systemtheorie, Technische Universität Wien, 1995. To appear in SIAM Journal on Control and Optimization.

[6] Willems J.C., "From time Series to Linear System, Part I: Finite Dimensional Linear Time Invariant Systems", Part II: "Exact Modeling," Automatica, Vol. 22, (1986), no. 5, p. 561–580, resp. Vol. 22, (1986), no. 6, p. 675–694.

[7] Willems J.C., "Paradigms and Puzzles in the Theory of Dynamical Systems," IEEE Trans. Automatic Control, Vol. 36, (1991), no. 3, p. 259–294.