On continuity and consistency of l-infinite optimal models

Weiland, S.; Scherrer, W.; Deistler, M.

Published: 01/01/1999

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 author's 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.

Citation for published version (APA):
On Continuity and Consistency of $\ell_\infty$ Optimal Models

Siep Weiland, Wolfgang Scherrer and Manfred Deistler

February 1999

Measurement and Control Systems
Internal Report, 99 I/01

Paper submitted to Systems and Control Letters, 1999

Eindhoven, February, 1999
On Continuity and Consistency of $\ell_\infty$ Optimal Models*

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

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

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

Submitted February 24, 1999

Abstract

This paper is concerned with 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 symmetrically in the system variables and measured in the $\ell_\infty$ or amplitude norm. It is shown that the misfit function continuously depends on both the data and the identified model. The consequences of this result for consistency of optimal and suboptimal models are discussed.

Keywords
System identification, Linear systems, Continuity, Consistency, Behavioral theory

*Research supported by the European Research Network on System Identification (ERNSI)
†Corresponding author
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 study of system identification procedures is the continuity of the function mapping observed data to identified models. 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. Further, as will be shown in this paper, consistency properties of models identified on the basis of finite length observations crucially rely on continuity properties.

In contrast to the usual approaches, we analyze a system identification problem in a deterministic setting in which neither the observed data nor the class of candidate models are required to have an input-output structure. This means that all system variables are treated in a symmetric way, without making an a priori distinction between input and output variables. We define the misfit between model and observed data by the $\ell_\infty$ (or amplitude) norm of the "noise", i.e. the part of the data which is not explained by the model. The main result of the paper shows that this 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.

The issues of continuity and consistency have been a main topic of research in the area of system identification. See e.g. [4]. Deterministic approaches to system identification and/or symmetric modeling of system variables have been proposed by various authors, see, e.g., [5,6,10]. The results presented in this paper closely resemble recent work in [7] 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 $\ell_\infty$ models are discussed.

The paper is organized as follows. In section 2 we formalize the $\ell_\infty$ 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 deferred 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

- $\ell_\infty$ the real normed linear space of all vector valued, magnitude bounded sequences $w : \mathbb{Z}_+ \to \mathbb{R}^q$ with norm
  $\|w\|_\infty := \max_{i \in \{1, \ldots, q\}} \sup_{t \in \mathbb{Z}_+} |w_i(t)|$.

- $c_0^\infty$ the subspace of $\ell_\infty$ consisting of all sequences $w \in \ell_\infty$ which vanish in the limit, i.e.,
  $c_0^\infty := \left\{ w \in \ell_\infty : \lim_{t \to \infty} w(t) = 0 \right\}$.

- $\ell_1$ 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
  $\|w\|_1 := \sum_{i=0}^{q} \sum_{t=0}^{\infty} |w_i(t)|$. 

We will drop the integer $q$ when dimensions are clear from the context. It is well known that $\ell_\infty$, $c_0$ and $\ell_1$ are Banach spaces and that $c_0$ is a closed linear subspace of $\ell_\infty$. The prefix $U$ will be used to indicate the closed unit sphere of normed linear spaces, e.g. $U\ell_\infty := \{ w \in \ell_\infty \mid \|w\|_\infty \leq 1 \}$.

## 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 $w \in c^q_0$. In factor analysis or principal component analysis the observed time series $w$ is decomposed as

$$w = \tilde{w} + \hat{w}, \quad (2.1)$$

where $\tilde{w}$ is the latent process [5] which satisfies linear restrictions and where the time series $\hat{w}$ is the error resulting from the approximation of $w$ by the latent process $\tilde{w}$. More precisely, a model (or a system) is a set $\mathcal{B}$ of time series mapping $\mathbb{Z}_+$ to $\mathbb{R}^q$. A model 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 $\sigma$ is the left-shift defined for given $v : \mathbb{Z}_+ \to \mathbb{R}^q$ as $(\sigma v)(t) := v(t + 1)$. See [10, 11] for more details on the behavioral framework. We will be particularly interested in $\ell_\infty$-systems which are defined as follows.

**Definition 2.1** An $\ell_\infty$-system is a linear, time-invariant and closed subset $\mathcal{B}$ of $c^q_0$.

The latent process $\tilde{w}$ in the decomposition (2.1) is supposed to satisfy $\tilde{w} \in \mathcal{B}$ where $\mathcal{B}$ is an $\ell_\infty$-system. The latent process $\tilde{w} \in \mathcal{B}$ is an optimal approximant of $w$ if the error $\hat{w} := w - \tilde{w}$ is minimal in some norm. In this paper we focus on the amplitude norm of the noise $\hat{w}$.

**Definition 2.2** (Misfit) The misfit between a time series $w \in c^q_0$ and the $\ell_\infty$-system $\mathcal{B}$ is

$$\mu(w, \mathcal{B}) := \inf \| w - \tilde{w} \|_\infty \quad (2.2)$$

The misfit is therefore expressed in terms of the distance between the data point $w$ and the element $\tilde{w} \in \mathcal{B}$ (if it exists) which is closest in the $\ell_\infty$ sense to $w$. Note that, because $\mathcal{B}$ is a closed subset of $c^q_0$, the misfit satisfies $\mu(w, \mathcal{B}) = 0$ 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 $\hat{w}$ is equal to zero for some decomposition of the form (2.1).

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

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

Note that $\mathcal{B}_{\text{opt}}(w)$ may be empty, as the minimum in (2.3) need not exist. A suboptimal version of this problem amounts to characterizing all models $\mathcal{B} \in \mathcal{B}$ which have a guaranteed misfit level. More precisely, for $\epsilon \geq 0$ we define the level set

$$\mathcal{B}(\epsilon, w) := \{ \mathcal{B} \in \mathcal{B} \mid \mu(w, \mathcal{B}) \leq \epsilon \}. \quad (2.4)$$
It is clear that $B(e, w)$ is empty if $e < e^{opt}(w)$ where
\[
e^{opt}(w) := \inf_{B \in B} \mu(w, B)
\]
is the optimal misfit level for the data $w$. In what follows we will be interested in the continuity properties of the misfit map $\mu : C_0^q \times B \to \mathbb{R}$ and discuss its consequences for the consistency of identified models.

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 equipped with the usual definitions of addition and scalar multiplication of linear functionals. The value of a bounded linear functional $x^* \in \mathcal{X}^*$ at $x \in \mathcal{X}$ is denoted $\langle x, x^* \rangle$. Its norm is defined as
\[
\| x^* \| := \sup_{x \in U \mathcal{X}} | \langle x, x^* \rangle |.
\]

It is well known [1,2] that for any $q > 0$ the normed space $\ell_1^q$ is the dual of $C_0^q$ and $\ell_{\infty}^q$ is the dual of $\ell_1^q$, i.e. $\ell_1^q = (C_0^q)^*$ and $\ell_{\infty}^q = (\ell_1^q)^*$. 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 $\ell_1^q$.

**Definition 3.1 (Orthogonal complements)** Suppose $B$ is a subspace of $C_0^q$, and $L$ a subspace of $\ell_1^q$. Their orthogonal complements $B^\perp$ and $L^\perp$ are the sets
\[
B^\perp := \{ r \in \ell_1^q \mid \langle w, r \rangle = 0 \text{ for all } w \in B \},
\]
\[
L^\perp := \{ w \in C_0^q \mid \langle w, r \rangle = 0 \text{ for all } r \in L \}.
\]

**Remark 3.2** A subspace $L \subseteq \ell_1^q$ defines a second orthogonal complement $L^\perp$ in $\ell_{\infty}^q$ by putting
\[
L^\perp := \{ v \in \ell_{\infty}^q \mid \langle v, r \rangle = 0 \text{ for all } r \in L \}.
\]

We will however not use this subspace.

It is clear that $B^\perp$ and $L^\perp$ define vector spaces. The laws of a model $B \subseteq C_0^q$ are the elements of the orthogonal complement $L = B^\perp$. Hence, every model $B \subseteq C_0^q$ uniquely defines a set of laws and, conversely, a set of laws $L \subseteq \ell_1^q$ defines a model $B := L^\perp$. Since $L^\perp (B^\perp)$ may be a proper superset of $B$, the laws $L := B^\perp$ need not uniquely define $B$. For closed subsets of $C_0^q$ this is however the case (see [1,3]) and thus we have

**Proposition 3.3** For an $\ell_{\infty}$-system $B$, the laws $L := B^\perp$ uniquely define $B$ in the sense that $B = L^\perp$.

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

\(1\) Also called annihilators.
Theorem 3.4 The misfit between \( w \in c_0^2 \) and an \( \ell_\infty \)-system \( B \) satisfies

\[
\mu(w, B) := \inf_{\tilde{w} \in B} \| w - \tilde{w} \|_\infty = \max_{r \in U B^\perp} \langle w, r \rangle
\]

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

4 Main results

In this section we provide the main technical result of this paper which states that the misfit function (2.2) is jointly continuous in the observed time series and the model. Convergence of time series is defined in the strong \( \ell_\infty \) sense as follows. A sequence \( (w_n | n \in \mathbb{Z}_+) \) of elements \( w_n \in c_0^2 \) is said to converge to an element \( w \in c_0^2 \) if \( \| w - w_n \|_\infty \to 0 \) as \( n \to \infty \). The definition of convergence of \( \ell_\infty \)-systems is more involved. A sequence of \( \ell_\infty \)-systems \( (B_n | n \in \mathbb{Z}_+) \) is said to converge to an \( \ell_\infty \)-system \( B \) if the following two conditions hold:

1. For any sequence of laws \( r_n \in U B_n^\perp \) there exists a sequence of laws \( r_n^0 \in U B^\perp \) such that for all \( w \in c_0^2 \) the functional

\[
\langle w, r_n - r_n^0 \rangle \to 0
\]

as \( n \to \infty \).

2. For all laws \( r^0 \in U B^\perp \) there exists a sequence of laws \( r_n \in U B_n^\perp \) such that for all \( w \in c_0^2 \)

\[
\langle w, r^0 - r_n \rangle \to 0
\]

as \( n \to \infty \).

In words, the first requirement says that for all sequences of (normalized) laws \( r_n \in B_n^\perp \) of \( B_n \) there exists a sequence of laws \( r_n^0 \) of the limiting model \( B \) 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 \( B_n \) constitute a subset of the laws of the limiting model \( B \). Similarly, the second requirement states that all (normalized) laws of the limit model \( B \) can be obtained as a weak-star limit of laws of \( B_n \). Model convergence is therefore expressed in terms of weak-star convergence of normalized laws.

Definition 4.1 (Sequential continuity and sequential compactness) Let \( \mathcal{X} \) be a set and \( f : \mathcal{X} \to \mathbb{R} \). Then \( f \) is called sequentially continuous if for all \( x \in \mathcal{X} \) and all sequences \( (x_n | n \in \mathbb{Z}_+) \) with \( x_n \in \mathcal{X} \) and \( x_n \to x \) there holds \( f(x_n) \to f(x) \). \( \mathcal{X} \) is called sequentially compact if every sequence \( (x_n | n \in \mathbb{Z}_+) \) with \( x_n \in \mathcal{X} \) has a convergent subsequence with limiting element in \( \mathcal{X} \).

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 \( \ell_\infty \)-systems. For this reason continuity of maps and compactness are defined via convergent sequences.

The main result of this section is as follows.

Theorem 4.3 The misfit function \( \mu : c_0^2 \times B \to \mathbb{R} \) is sequentially continuous in its arguments \( (w, B) \).
Proof. Let \( w \in \mathcal{C}_0^\alpha \), \( \mathcal{B} \in \mathcal{B} \) be an \( \ell_\infty \)-system and let \( (w_n | n \in \mathbb{Z}_+) \) and \( (\mathcal{B}_n | n \in \mathbb{Z}_+) \) be sequences of time series in \( \mathcal{C}_0^\alpha \) and \( \ell_\infty \)-systems respectively such that \( (w_n, \mathcal{B}_n) \rightarrow (w, \mathcal{B}) \). Then, using theorem 3.4, we find that

\[
\mu(w_n, \mathcal{B}_n) - \mu(w, \mathcal{B}) = \max_{r_n \in U\mathcal{B}_n^\perp} (w_n, r_n) - \max_{r \in U\mathcal{B}^\perp} (w, r) = \langle w_n, r_n^{\text{opt}} \rangle - \max_{r \in U\mathcal{B}^\perp} (w, r) \leq \langle w_n - w, r_n^{\text{opt}} \rangle - \langle w, r_n^0 \rangle = \langle w_n - w, r_n^{\text{opt}} \rangle + \langle w, r_n^{\text{opt}} - r_n^0 \rangle.
\]

Here, \( r_n^{\text{opt}} \in U\mathcal{B}_n^\perp \) is such that \( \langle w_n, r_n^{\text{opt}} \rangle = \max_{r \in U\mathcal{B}_n^\perp} (w_n, r_n) \) and \( r_n^0 \in U\mathcal{B}_n^\perp \) is the corresponding sequence of laws satisfying requirement 1 for convergence of models. Further, we used that \( \max_{r \in U\mathcal{B}^\perp} (w, r) \geq \langle w, r_n^0 \rangle \). 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 l.h.s. of the above formula converges to 0 as \( n \to \infty \).

Similarly, we obtain that

\[
\mu(w, \mathcal{B}) - \mu(w_n, \mathcal{B}_n) = \max_{r \in U\mathcal{B}^\perp} (w, r) - \max_{r_n \in U\mathcal{B}_n^\perp} (w_n, r_n) = \langle w, r^{\text{opt}} \rangle - \max_{r_n \in U\mathcal{B}_n^\perp} (w_n, r_n) \leq \langle w, r^{\text{opt}} \rangle - \langle w_n, r_n^0 \rangle = \langle w, r^{\text{opt}} - r_n^0 \rangle + \langle w_n - w, r_n^0 \rangle.
\]

Here, \( r^{\text{opt}} \in U\mathcal{B}^\perp \) is such that \( \langle w, r^{\text{opt}} \rangle = \max_{r \in U\mathcal{B}^\perp} (w, r) \) 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 \( \mu \).

\( \square \)

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 \( w \in \mathcal{C}_0^\alpha \), a truncated observation

\[
w_n(t) := \begin{cases} w(t) & t \leq n \\ 0 & t > n \end{cases}
\]

clearly satisfies \( w_n \to w \).

Next, we discuss under what conditions optimal models will be consistent. In most frameworks, consistency amounts to considering the question whether a model \( \mathcal{B} \in \mathcal{B} \) can be identified from finite length observations \( w_n \) of \( w \in \mathcal{B} \) by minimizing the misfit function \( \mu(w_n, \cdot) \) over elements in \( \mathcal{B} \). We will formalize this notion in a deterministic way for models in the level sets \( \mathcal{B}(\varepsilon, w) \), defined in (2.4).

**Definition 4.4 (Consistency)** Let \( \mathcal{B} \) be a model class of \( \ell_\infty \)-systems and \( \mathcal{B}(\varepsilon, w) \subseteq \mathcal{B} \) be a non-empty level set. \( \mathcal{B}(\varepsilon, w) \) is called
1. weakly consistent if

\[
(\varepsilon_n, w_n, \mathcal{B}_n) \rightarrow (\varepsilon, w, \mathcal{B}), \quad \mathcal{B}_n \in \mathcal{B}(\varepsilon_n, w_n) \implies \mathcal{B} \in \mathcal{B}(\varepsilon, w)
\]

2. consistent if

\[
(\varepsilon_n, w_n) \rightarrow (\varepsilon, w), \quad \mathcal{B}_n \in \mathcal{B}(\varepsilon_n, w_n) \implies \exists \text{ a subsequence } \mathcal{B}_{n_m} \text{ of } \mathcal{B}_n \text{ with } \mathcal{B}_{n_m} \rightarrow \mathcal{B} \text{ and } \mathcal{B} \in \mathcal{B}(\varepsilon, w)
\]

In words, weak consistency guarantees that if elements of level sets \( \mathcal{B}(\varepsilon_n, w_n) \) converge, then they converge to elements in the limiting level set \( \mathcal{B}(\varepsilon, w) \). Consistency means that any sequence of models in \( \mathcal{B}(\varepsilon_n, w_n) \) has at least one subsequence that converges to an element in the limiting level set. Note that a consistent level set is also weakly consistent.

**Theorem 4.5** Let \( \mathcal{B} \) be a model class of \( \ell_\infty \)-systems. Then

1. \( \mathcal{B}(\varepsilon, w) \) is weakly consistent for all \( w \in c_0^\mathbb{N} \) and \( \varepsilon > \varepsilon_{\text{opt}}(w) \).
2. \( \mathcal{B}(\varepsilon, w) \) is consistent for all \( w \in c_0^\mathbb{N} \) and \( \varepsilon > \varepsilon_{\text{opt}}(w) \) whenever \( \mathcal{B} \) is sequentially compact.

**Proof.** Note that, \( \varepsilon_n \rightarrow \varepsilon \) with \( \varepsilon > \varepsilon_{\text{opt}}(w) \) implies that \( \mathcal{B}(\varepsilon_n, w_n) \) is non-empty for all but finitely many \( n \in \mathbb{N} \).

It thus makes sense to consider sequences \( \mathcal{B}_n \mid n \in \mathbb{N} \) with \( \mathcal{B}_n \in \mathcal{B}(\varepsilon_n, w_n) \).

1. Let \( (\varepsilon_n, w_n, \mathcal{B}_n) \rightarrow (\varepsilon, w, \mathcal{B}) \) and let \( \mathcal{B}_n \in \mathcal{B}(\varepsilon_n, w_n) \). By continuity of the misfit we have that \( \mu(w_n, \mathcal{B}_n) \rightarrow \mu(w, \mathcal{B}) \). As \( \mu(w_n, \mathcal{B}_n) \leq \varepsilon_n \) and \( \varepsilon_n \rightarrow \varepsilon \) it follows that \( \mu(w, \mathcal{B}) \leq \varepsilon \), i.e. \( \mathcal{B} \in \mathcal{B}(\varepsilon, w) \).

2. Let \( (\varepsilon_n, w_n) \rightarrow (\varepsilon, w) \) and \( \mathcal{B}_n \in \mathcal{B}(\varepsilon_n, w_n) \). Since \( \mathcal{B} \) is sequentially compact, there exists a subsequence \( (\mathcal{B}_{n_m} \mid n_m \in \mathbb{N}) \) of \( (\mathcal{B}_n \mid n \in \mathbb{N}) \) that converges to, say, \( \mathcal{B} \in \mathcal{B} \). Hence, \( (\varepsilon_{n_m}, w_{n_m}, \mathcal{B}_{n_m}) \) converges to \( (\varepsilon, w, \mathcal{B}) \) with \( \mathcal{B} \in \mathcal{B} \). Statement 1 now implies that \( \mu(w, \mathcal{B}) \leq \varepsilon \) so that \( \mathcal{B} \in \mathcal{B}(\varepsilon, w) \). □

The following theorem is an immediate consequence of Theorem 4.5 and shows that optimal \( \ell_\infty \)-models exist in sequentially compact model sets.

**Theorem 4.6** Let \( \mathcal{B} \) be a model class of \( \ell_\infty \)-systems and suppose that \( \mathcal{B} \) is sequentially compact. Then

1. \( \mathcal{B}_{\text{opt}}(w) \) is non-empty for all \( w \in c_0^\mathbb{N} \).
2. \( \varepsilon_{\text{opt}} : c_0^\mathbb{N} \rightarrow \mathbb{R}_+ \) is sequentially continuous.

**Proof.** (1) Let \( w \in c_0^\mathbb{N}, \varepsilon_n \downarrow \varepsilon_{\text{opt}}(w) \) and \( w_n = w \). With the notation of the proof of Statement 2 of Theorem 4.5, \( \mathcal{B} := \lim_{n \rightarrow \infty} \mathcal{B}_{n_m} \) belongs to \( \mathcal{B} \) and \( \mu(w, \mathcal{B}) \leq \lim_{n \rightarrow \infty} \mu(w, \mathcal{B}_{n_m}) = \varepsilon_{\text{opt}}(w) \). Hence, \( \mathcal{B} \in \mathcal{B}_{\text{opt}}(w) \), which proves that \( \mathcal{B}_{\text{opt}}(w) \) is non-empty.

(2) Let \( w_n \rightarrow w \). By the first claim, there exists \( \mathcal{B}_n, \mathcal{B} \in \mathcal{B} \) such that \( \varepsilon_{\text{opt}}(w_n) = \mu(w_n, \mathcal{B}_n) \) and \( \varepsilon_{\text{opt}}(w) = \mu(w, \mathcal{B}) \). Then \( \varepsilon_{\text{opt}}(w_n) = \mu(w_n, \mathcal{B}_n) \leq \mu(w_n, \mathcal{B}) \), which, taking limits on both sides and using continuity of the misfit, yields

\[
\lim_{n \rightarrow \infty} \sup \varepsilon_{\text{opt}}(w_n) \leq \varepsilon_{\text{opt}}(w).
\]
Suppose that the latter inequality is strict. By compactness of $B$, there exists a converging subsequence $(B_{n_m} | n_m \in \mathbb{Z}_+)$ of $\{B_n | n \in \mathbb{Z}_+\}$ with limit, say, $B_\infty \in B$. But then,

$$
\mu(w, B_\infty) = \lim_{m \to \infty} \mu(w_{n_m}, B_{n_m}) = \lim_{m \to \infty} \varepsilon_{\text{opt}}(w_{n_m}) \leq \limsup_{n \to \infty} \varepsilon_{\text{opt}}(w_n) < \varepsilon_{\text{opt}}(w)
$$

which contradicts the definition of $\varepsilon_{\text{opt}}(w)$.

\[ \Box \]

**Remark 4.7** It follows that sequentially compact model sets $B$ are relevant to guarantee consistency of level sets. However, it is not obvious to find such model sets. One of the most common model sets consists of systems $B \subseteq (\mathbb{R}^q)^{\mathbb{Z}_+}$ that are linear, time-invariant and closed (in the topology of pointwise convergence). See e.g. [10, 11]. In the context of $\ell_\infty$-systems, this model set is defined as follows. Associate with a finite set of laws $R = \{r_1, \ldots, r_g\}$ with $r_i \in \ell_1^q$ ($i = 1, \ldots, g$), the system

$$
B(R) := \{w \in \ell_0^q \mid (a^T w, r) = 0 \text{ for all } t \in \mathbb{Z}_+ \text{ and } r \in R\}.
$$

Then $B(R)$ is an $\ell_\infty$-system in the sense of definition 2.1. For $r \in \ell_1^q$, let

$$
\hat{r}(z) := \sum_{t \in \mathbb{Z}_+} r(t)z^t, \quad z \in \mathcal{E}
$$

de note its $z$ transform, where the existence region $\mathcal{E}$ is the set of all complex points $z$ for which the infinite sum converges. It then follows that

$$
B(R) = \{w \in \ell_0^q \mid \hat{r}^T(\sigma)w = 0 \text{ for all } r \in R\}. \quad (4.2)
$$

That is, the system $B(R)$ can be represented by a finite set of autoregressive equations. Next, suppose that each of the elements $r \in R$ has finite support, i.e., $\hat{r}$ is a polynomial with coefficients in $\mathbb{R}^q$ for all $r \in R$. In that case

$$
\hat{R}(z) := \begin{pmatrix}
\hat{r}_1(z) \\
\vdots \\
\hat{r}_g(z)
\end{pmatrix}
$$

is a polynomial matrix. Suppose that $R$ has rank $q - m$ and McMillan degree $n$ and define $B(m, n)$ as the model set of all $\ell_\infty$-systems of the form (4.2) where $R$ has rank $q - m$ and McMillan degree at most $n$. When expressed in terms of input-state-output systems, $m$ is the number of inputs, $q - m$ the number of outputs and $n$ is the (minimal) number of states of systems in $B(m, n)$.

In general, this model set is not sequentially compact, as can be seen from the following example with $q = 1, m = 0$ and $n = 1$. It is easy to see, that $B(0, 1)$ is the set of all $B = B(r)$ for $r = (1, 0, \ldots)$ or $r = (\alpha, 1, 0, \ldots) \in \ell_1$ with $|\alpha| < 1$. Let $B(r_k), r_k = ((1 - 1/k), 1, 0, \ldots), k \in \mathbb{Z}_+$ be a sequence of systems in $B(0, 1)$. Now suppose that $B(r_k) \to B$ with $B \in B(0, 1)$. Then $B = B(r_0)$ for some $r_0 = (\alpha_0, 1, 0, \ldots), |\alpha_0| < 1$. Now the requirements for convergence of models give the desired contradiction.

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 $\ell_\infty$ (or amplitude) 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. An advantage of this approach is that it allows for a symmetric treatment of variables and that no assumptions on inputs and outputs are necessary. 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. Consequences of this result for consistency of sub-optimal models have been discussed and it has been shown that sequentially compact model sets are of interest for the study of consistent models. The general class of linear time-invariant $\ell_\infty$-systems that admit autoregressive representations is not sequentially compact. In this paper only qualitative properties of $\ell_\infty$ optimal models have been investigated. The development of identification algorithms and a qualification of sequentially compact model sets are among the subjects of further investigation.

References