# On optimal deterministic identification with uncertain data treated as exact data

569350

Eindhoven University of Technology
Faculty of Electrical Engineering

# On Optimal Deterministic Identification with Uncertain Data Treated as Exact Data

L.J.J.M. Ariaans, S. Weiland, A.A.H. Damen

Internal report, 94I/04
Group Measurement and Control

Eindhoven, September 26, 1994

# On Optimal Deterministic Identification with Uncertain Data Treated as Exact Data

**L.J.J.M. Ariaans, S. Weiland, A.A.H. Damen**
Department of Electrical Engineering,
Eindhoven University of Technology,
P.O.Box 513,
5600 MB Eindhoven,
The Netherlands
Tel: +31 40 47 3795
Fax: +31 40 434 582
e-Mail: L.J.J.M.Ariaans@ele.tue.nl

September 26, 1994

**Abstract**

Experimental data obtained from a real plant is always contaminated by various disturbance effects. For the model identification problem this invariably leads to uncertainty in the identified model with respect to the true process data. In this paper, we assume deterministic bounds on the uncertainty of the data which are expressed as norm bounds. Three classes of identification problems are considered. One in which data is assumed to be exact, not corrupted by noise, and two identification problems in which data uncertainty is explicitly taken into account. For very general model sets and for any norm bound of the data uncertainty, we show that an optimal solution of the identification problem without data uncertainty also solves the more complex problems with data uncertainty.

**Keywords:** Identification, Uncertain Data, Deterministic Bounds, Norm Bounds, Optimization

## 1 Introduction

One of the difficulties inherent to process identification is that the data from which we try to identify the process is corrupted by measurement noise or process disturbances. The relationships among the process variables that we wish to uncover are further blurred by the influence of unmodelled process inputs (i.e. disturbances) or by relationships whose complexity is not contained in the model set. All these effects make that the measured process signals do not correspond to what is normally called the "true" signals. One can attribute statistical properties to these uncertainties which gives rise to maximum likelihood estimators, prediction error methods etc. Ljung (1987), Söderström and

Stoica (1989). Another approach is to bound these uncertainties in a deterministic way and reflect these uncertainties in conditions on model parameters. This approach is taken in set estimation techniques, Norton (1987), Walter and Piet-Lahanier (1990), Milanese (1989). In this paper we use deterministic bounds, but not in a set estimation context. Instead, we propose two ways of incorporating deterministic uncertainty in the identification criterion. It is shown that under very general conditions those models which are optimal for data which is not corrupted by uncertainty, are also optimal in the uncertain data case. We will show this for very general classes of model sets including the class of finite dimensional linear models.

## 2 Three classes of identification problems

We consider multivariate processes and adhere to the modelling framework of Willems (1986). Let $w : \mathbb{T} \mapsto W$ denote a possible evolution of the process variables. Here $\mathbb{T} \subseteq \mathbb{R}$ is a time set and $W$, the signal space, is the space in which $w(t)$ take their values for each $t \in \mathbb{T}$. The set of all possible evolutions is obviously equal to $W^{\mathbb{T}} =: \mathbb{W}$. We assume $\mathbb{W}$ to be a normed space and denote the norm by $\|.\|$.

A process or model is a subset of $\mathbb{W}$. Such a subset contains those trajectories in $\mathbb{W}$ which are consistent with the process or model. Let $\mathbb{B}$ denote the set of all subsets of $\mathbb{W}$, i.e. $\mathbb{B} := 2^{\mathbb{W}}$. Then $\mathbb{B}$ is the set of all possible models.

Let the measurements be given by a single trajectory $w_m \in \mathbb{W}$. Because of the uncertainty in the data, the true process behaviour $w_t$ may not have been identical to $w_m$. The uncertainty in $w_t$ is assumed to be bounded: $\|w_t - w_m\| \leq d_w$ where $d_w \geq 0$ is an a priori known constant. Given $w_m$, the "true signal" $w_t$ can therefore be any element of

$$\mathcal{W} := \{w \in \mathbb{W} \mid \|w - w_m\| \leq d_w\} \tag{1}$$

We will refer to $\mathcal{W}$ as the *uncertain data set* with center $w_m \in \mathbb{W}$. Note, that $\{w_m\} = \mathcal{W}$ if and only if $d_w = 0$.

**Definition 1** *A model $\mathcal{B} \in \mathbb{B}$ is called convex if*

$$w_1, w_2 \in \mathcal{B}, \lambda \in [0,1] \implies \lambda w_1 + (1-\lambda) w_2 \in \mathcal{B}$$

We denote by $\mathbb{B}_{cv}$ the subset of $\mathbb{B}$ consisting of all convex models.

**Definition 2** *A model $\mathcal{B} \in \mathbb{B}$ is called closed if $\mathcal{B}$ is a closed subset of $(\mathbb{W}, \|.\|)$.*

Denote by $\mathbb{B}_{cl}$ the class of all closed models in $\mathbb{B}$. Let $\mathbb{B}_{cc} := \mathbb{B}_{cv} \cap \mathbb{B}_{cl}$ be the set of all convex and closed models and denote by $\mathbb{B}_{lin}$ the subset of $\mathbb{B}$ consisting of all linear subspaces of $\mathbb{W}$.

Next suppose that $\mathbb{B}_{id} \subset \mathbb{B}_{cl}$ is a class of candidate models which we wish to consider for the purpose of identification. An identification procedure is a map $P$ which maps the uncertain data set $\mathcal{W}$ to models $\mathcal{B} \in \mathbb{B}_{id}$. We will assume, that the trivial model $\mathbb{W}$ is not an element of $\mathbb{B}_{id}$. The quality of a model $\mathcal{B} \in \mathbb{B}_{id}$ is assessed by a measure of how well $\mathcal{B}$ fits the measured data. We formalise the misfit between data and model as follows.

**Definition 3** *The nominal distance between $\mathcal{B} \in \mathbb{B}_{id}$ and the data $w_m$ is given by*

$$d_1(\mathcal{B}, \{w_m\}) := \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_m\|$$

**Definition 4** *The best case distance between $\mathcal{B} \in \mathbb{B}_{id}$ and $\mathcal{W}$ is*

$$d_2(\mathcal{B}, \mathcal{W}) := \min_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\|$$

**Definition 5** *The worst case distance between $\mathcal{B} \in \mathbb{B}_{id}$ and $\mathcal{W}$ is*

$$d_3(\mathcal{B}, \mathcal{W}) := \max_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\|$$

Note that $d_2(\mathcal{B}, \mathcal{W}) = \min_{w \in \mathcal{W}} d_1(\mathcal{B}, \{w\})$ and $d_3(\mathcal{B}, \mathcal{W}) = \max_{w \in \mathcal{W}} d_1(\mathcal{B}, \{w\})$. Each of the distance measures defines an identification problem. Given $w_m$, let $P_1$ denote the problem to find all models $\mathcal{B} \in \mathbb{B}_{id}$ such that $d_1(\mathcal{B}, \{w_m\})$ is minimised. Given $\mathcal{W}$, let $P_i$, $i = 2, 3$ denote the problem to find all models $\mathcal{B} \in \mathbb{B}_{id}$ such that $d_i(\mathcal{B}, \mathcal{W})$ is minimised. That is, $P_i$ amounts to finding

$$\arg \min_{\mathcal{B} \in \mathbb{B}_{id}} \{d_i(\mathcal{B}, \mathcal{W}_i)\}$$
$$\text{where } \mathcal{W}_1 = \{w_m\}, \mathcal{W}_2 = \mathcal{W}_3 = \mathcal{W}$$

The identification problems $P_1$, $P_2$, $P_3$ are in fact classes of identification problems, as the norm $\|.\|$ is not yet specified. Also the model set $\mathbb{B}_{id}$ can still be chosen.

Problem $P_1$ does not take the signal uncertainties into account, at least not explicitly. The bounds on the data uncertainty are not used at all.

The way in which we should take data uncertainty into account in an identification criterion is directly related to our faith (or lack of faith) in the fact that the process is in the model set $\mathbb{B}_{id}$. This general idea is reflected in the difference between $P_2$ and $P_3$. If we know that the true process is in $\mathbb{B}_{id}$, any inconsistencies between $\mathcal{B} \in \mathbb{B}_{id}$ and the measured data must come from data contamination. Under this assumption we reduce conservatism by searching in $\mathcal{W}$ for data which is closest to a $B \in \mathbb{B}_{id}$. This is formalised in problem $P_2$. The opposite situation is that we do not know whether the true process is in the model set. In this case, the distance between the true data and a model can be guaranteed to be no larger than the maximal distance between any data element in $\mathcal{W}$ and that model. This is reflected in problem $P_3$.

**Remark 1** *One can proof from the continuity of the norm and the fact that $\mathbb{B}_{id} \subseteq \mathbb{B}_{cl}$, that the minima over $\tilde{w} \in \mathcal{B}$ exist in definition 3, 4 and 5. Proof of the existence of $\min_{w \in \mathcal{W}}$ in definition 4 and $\max_{w \in \mathcal{W}}$ in 5 follows in the same way from the continuity of the norm and the compactness of $\mathcal{W}$. We assume, that the minima over $\mathcal{B} \in \mathbb{B}_{id}$ involved in $P_1$, $P_2$, $P_3$ actually exist. The identification problems $P_1$, $P_2$, $P_3$ are therefore well-posed problems.*

We do not assume, that $P_1$, $P_2$ or $P_3$ have unique solutions. If more models $\mathcal{B} \in \mathbb{B}_{id}$ belong to the argument of the minimal distances then we do not discriminate between these models.

# 3 Main result

Let $\mathbb{B}_i^{opt} = \arg\min_{\mathcal{B}\in\mathbb{B}_{id}} d_i(\mathcal{B}, \mathcal{W}_i), \quad i = 1,2,3$ denote the optimal models which solve problem $P_i$. The main results of this paper can then be formulated as follows.

**Theorem 1** *If* $\mathbb{B}_{id} \subseteq \mathbb{B}_{cc}$ *then* $\mathbb{B}_1^{opt} \subset \mathbb{B}_2^{opt}$. □

For a similar result concerning $P_1$ and $P_3$, we need an extra condition:

**Theorem 2** *Consider two conditions*

$$\mathbb{B}_{id} \subseteq \mathbb{B}_{cc} \tag{C1}$$

$$\min_{\mathcal{B}\in\mathbb{B}_{id}} \min_{\tilde{w}\in\mathcal{B}} \|\tilde{w} - w_m\| > 0 \vee \mathbb{B}_{id} \subseteq \mathbb{B}_{lin} \tag{C2}$$

*then*

$$\text{(C1)} \wedge \text{(C2)} \implies \mathbb{B}_1^{opt} \subset \mathbb{B}_3^{opt}$$ □

Hence, for the class of convex and closed model sets any optimal model that solves $P_1$ (the exact data case) is also optimal for $P_2$ and $P_3$. Stated otherwise, the nominal data point $w_m \in \mathcal{W}$ contains all necessary information to determine optimal models for the problem $P_2$ and $P_3$. Moreover, this holds for any uncertainty bound $d_w \geq 0$ and any norm on $\mathbb{W}$.

# 4 Discussion

We claim, that the results obtained here are very general. Firstly, because they hold for *any* norm. Many norms can be formulated to bound the data uncertainty. If we have a norm bound on component $w_i$, $i = 1, \ldots, q$ of $w$, we can combine these into a single norm bound on $w$. If we have for example $\|w_i\|_{(i)} \leq \alpha_i$, $i = 1, \ldots, q$ then this is equivalent to requiring $\max_i \|w_i\|_{(i)}/\alpha_i \leq 1$, where the expression to the left of the inequality sign is a norm on $w$. This does not require the norms $\|.\|_{(i)}$ to be identical. Also, if $w$ is known to be bounded in various norms, this can be combined into a single norm. If $\|w\|_{(k)} \leq 1$, $k = 1, \ldots, N$ for $N$ different norms $\|.\|_{(k)}$, then this can be expressed as $\max_k \|w\|_{(k)} \leq 1$, where the left hand side expression defines again a norm. This means that for example

$$\max_i \max_{t\in\mathbb{T}} |w_i(t)|/\alpha_i$$

defines a norm on $\mathbb{W}$ for positive constants $\alpha_i$. This expression is — from a practical point of view — very acceptable way to bound data uncertainty.

Secondly, the restrictions we put on the allowable model set are very weak. The very common practical case of linear time invariant models with a bound on their McMillan degree satisfies these restrictions. In fact, the models need not have a bounded McMillan degree or be time invariant. Also classes of nonlinear model sets satisfy the conditions of closedness and convexity.

The main restrictive element in these results is that the same norm must be used to express the data uncertainty and the optimality criterion. This precludes to some extent the use of norms which are tuned specifically towards bounding the data uncertainty as much as possible, as the optimality criterion resulting from such a norm can be hard to interpret.

As an example in which theorem 1 and 2 can be used we mention total least squares identification with a bound on the total energy in the disturbances. Also an $\ell_\infty$ bound on the disturbances combined with an $\ell_\infty$ optimality criterion for the identification satisfies the requirements for these theorems.

# 5 Conclusions

We have defined two notions of optimality for estimated models taking data uncertainty into account. One is applicable in case we know the true process is in the model set. The other takes a worst case approach and is applicable in a situation where we can not assume that the true process is in the model set. If data uncertainty can be expressed as a norm bound, then models which are optimal in the exact, noise free case, are also optimal in the case where data uncertainty is to be taken into account. This means we can solve a simpler identification problem and get the optimality with respect to the data uncertainty for free.

# A Proof of theorem 1

In this appendix we will prove theorem 1. For the proof of theorem 2 we refer to appendix B.

First, introduce the following auxiliary sets and functions

$$\widetilde{\mathcal{W}}_- : \mathbb{B}_{id} \rightarrow 2^{\mathcal{W}}, \quad \mathcal{B} \mapsto \{\tilde{w}_- \in \mathcal{B} \mid \min_{w \in \mathcal{W}} \|\tilde{w}_- - w\| = \min_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\|\}$$
$$\mathcal{W}_- : \mathbb{B}_{id} \rightarrow 2^{\mathcal{W}}, \quad \mathcal{B} \mapsto \{w_- \in \mathcal{W} \mid \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_-\| = \min_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\|\}$$

and define the subset $\mathbb{B}'_{id}$ of $\mathbb{B}_{id}$ by

$$\mathbb{B}'_{id} := \{B \in \mathbb{B}_{id} \mid \min_{w \in \mathcal{W}} \min_{\tilde{w} \in B} \|\tilde{w} - w\| > 0\}$$

**Lemma 3**

$$\forall \mathcal{B} \in \mathbb{B}'_{id}, w \in \mathcal{W}_-(\mathcal{B}) \quad \|w - w_m\| = d_w$$

□

**Proof:** Obviously, $\|w - w_m\| \leq d_w$. Now assume that $\exists w_1 \in \mathcal{W}_-$ such that $\|w_m - w_1\| < d_w$. Let $\tilde{w}_2 \in \mathcal{B}$ such that

$$\|\tilde{w}_2 - w_1\| = \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_1\|$$

Since $\mathcal{B} \in \mathbb{B}'_{id}$ we have $\tilde{w}_2 \neq w_1$ and there exists $\epsilon \in (0,1)$ such that $w_3 := w_1 + \epsilon(\tilde{w}_2 - w_1) \in \mathcal{W}$. We find

$$\|\tilde{w}_2 - w_3\| = (1 - \epsilon)\|\tilde{w}_2 - w_1\| < \|\tilde{w}_2 - w_1\|$$

This is in contradiction with $w_1 \in \mathcal{W}_-(\mathcal{B})$ so it follows $\|w_m - w_1\| \not< d_w$. This gives the stated result. □

Next introduce the function

$$w_- : \widetilde{\mathcal{W}}_-(\mathbb{B}'_{id}) \to \mathcal{W}, \quad \tilde{w} \mapsto w_m - d_w \frac{w_m - \tilde{w}}{\|w_m - \tilde{w}\|}$$

This function is well defined, because certainly $w_m \neq \tilde{w}$ for $\mathcal{B} \in \mathbb{B}'_{id}$.

**Lemma 4**

$$\forall \mathcal{B} \in \mathbb{B}'_{id}, \tilde{w}_1 \in \widetilde{\mathcal{W}}_-(\mathcal{B}) \quad \|\tilde{w}_1 - w_-(\tilde{w}_1)\| = \min_{w \in \mathcal{W}} \|\tilde{w}_1 - w\|$$

□

**Proof:** Let $\tilde{w}_1 \in \widetilde{\mathcal{W}}_-(\mathcal{B})$ and take $w_2 \in \mathcal{W}$ such that

$$\|\tilde{w}_1 - w_2\| = \min_{w \in \mathcal{W}} \|\tilde{w}_1 - w\| \tag{3}$$

Then

$$\|\tilde{w}_1 - w_m\| = \|\tilde{w}_1 - w_-(\tilde{w}_1)\| + \|w_-(\tilde{w}_1) - w_m\| = \|\tilde{w}_1 - w_-(\tilde{w}_1)\| + d_w$$
$$\|\tilde{w}_1 - w_m\| \leq \|\tilde{w}_1 - w_2\| + \|w_2 - w_m\| = \|\tilde{w}_1 - w_2\| + d_w$$

where the last equality follows from lemma 3. This leads to

$$\|\tilde{w}_1 - w_2\| \geq \|\tilde{w}_1 - w_-(\tilde{w}_1)\|$$

Together with (3) this gives

$$\|\tilde{w}_1 - w_-(\tilde{w}_1)\| = \|\tilde{w}_1 - w_2\| = \min_{w \in \mathcal{W}} \|\tilde{w}_1 - w\|$$

which completes the proof. □

**Proof of theorem 1:** First consider the case that $\mathbb{B}_{id} \setminus \mathbb{B}'_{id} \neq \emptyset$. It is easy to see that in this case $\mathbb{B}_2^{opt} = \mathbb{B}_{id} \setminus \mathbb{B}'_{id}$. The following equivalence holds

$$\hat{\mathcal{B}} \in \mathbb{B}_2^{opt} \iff \hat{\mathcal{B}} \cap \mathcal{W} \neq \emptyset \iff \min_{\tilde{w} \in \hat{\mathcal{B}}} \|\tilde{w} - w_m\| \leq d_w$$

This implies $\mathbb{B}_1^{opt} \subset \mathbb{B}_2^{opt}$; the result is shown.

Next consider the case that $\mathbb{B}_{id} \setminus \mathbb{B}'_{id} = \emptyset$. Let $\hat{\mathcal{B}}$ be any element of $\mathbb{B}_1^{opt}$ and take $\tilde{w}_1 \in \hat{\mathcal{B}}$, such that $\|\tilde{w}_1 - w_m\| = \min_{\tilde{w} \in \hat{\mathcal{B}}} \|\tilde{w} - w_m\|$. To prove theorem 1 it is sufficient to prove

$$\min_{w \in \mathcal{W}} \|\tilde{w}_1 - w\| = \min_{\mathcal{B} \in \mathbb{B}_{id}} \min_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\|$$

Because of lemma 4 the minimum on the left hand side is achieved for $w = w_-(\tilde{w}_1)$ so we need to prove

$$\|\tilde{w}_1 - w_-(\tilde{w}_1)\| = \min_{\mathcal{B} \in \mathbb{B}_{id}} \min_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\|$$

Assume to the contrary, that

$$\exists \mathcal{B}_0 \in \mathbb{B}_{id}, w_3 \in \mathcal{W}, \tilde{w}_2 \in \mathcal{B}_0 \quad \|\tilde{w}_2 - w_3\| < \|\tilde{w}_1 - w_-(\tilde{w}_1)\| \tag{4}$$

Applying lemma 4 once more we have

$$\|\tilde{w}_2 - w_-(\tilde{w}_2)\| \leq \|\tilde{w}_2 - w_3\| \quad (5)$$

Combining (4) and (5)

$$\|\tilde{w}_2 - w_m\| - d_w = \|\tilde{w}_2 - w_-(\tilde{w}_2)\| < \|\tilde{w}_1 - w_-(\tilde{w} - 1)\| = \|\tilde{w}_1 - w_m\| - d_w$$

so $\|\tilde{w}_2 - w_m\| < \|\tilde{w}_1 - w_m\|$. This contradicts $\tilde{w}_1 \in \hat{\mathcal{B}} \in \mathbb{B}_1^{opt}$. We conclude that (4) does not hold, which yields the required result. □

# B Proof of theorem 2

In this appendix, theorem 2 is shown. The line of reasoning is similar to that of appendix A, although subtle differences occur here and there.

We start with the introduction of two auxiliary functions

$$\widetilde{\mathcal{W}}_+ : \mathbb{B}_{id} \to 2^{\mathcal{W}}, \quad \mathcal{B} \mapsto \{\tilde{w}_+ \in \mathcal{B} \mid \max_{w \in \mathcal{W}} \|\tilde{w}_+ - w\| = \max_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\|\}$$

$$\mathcal{W}_+ : \mathbb{B}_{id} \to 2^{\mathcal{W}}, \quad \mathcal{B} \mapsto \{w_+ \in \mathcal{W} \mid \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_+\| = \max_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\|\}$$

and define the subset $\mathbb{B}''_{id}$ of $\mathbb{B}_{id}$ by

$$\mathbb{B}''_{id} := \{B \in \mathbb{B}_{id} \mid \min_{\tilde{w} \in B} \|\tilde{w} - w_m\| > 0\}$$

**Lemma 5**

$$\forall \mathcal{B} \in \mathbb{B}''_{id}, w \in \mathcal{W}_+(\mathcal{B}) \quad \|w - w_m\| = d_w$$

□

**Proof:** Clearly, $\|w - w_m\| \leq d_w$. Therefore, assume that $\exists w_1 \in \mathcal{W}_+(\mathcal{B})$ such that $\|w_m - w_1\| < d_w$. Let $\tilde{w}_2 \in \mathcal{B}$ such that

$$\|\tilde{w}_2 - w_1\| = \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_1\| \quad (6)$$

Since $\mathcal{B} \in \mathbb{B}''_{id}$ we have $\tilde{w}_2 \neq w_1$ and there exists $\epsilon > 0$ such that $w_3 := w_1 - \epsilon(\tilde{w}_2 - w_1) \in \mathcal{W}$. Then observe that

$$\|\tilde{w}_2 - w_3\| = (1 + \epsilon)\|\tilde{w}_2 - w_1\| > \|\tilde{w}_2 - w_1\|$$

This is only in agreement with $w_1 \in \mathcal{W}_+(\mathcal{B})$ if

$$\min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_3\| \leq \|\tilde{w}_2 - w_1\|$$

Therefore assume

$$\min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_3\| \leq \|\tilde{w}_2 - w_1\| < \|\tilde{w}_2 - w_3\|$$

and let the minimum be achieved for $\tilde{w} = \tilde{w}_4$. Consider

$$\tilde{w}_5 := \tilde{w}_2 + \frac{\tilde{w}_4 - \tilde{w}_2}{1 + \epsilon}$$

From the convexity of $\mathcal{B}$ we conclude $\tilde{w}_5 \in \mathcal{B}$. Write

$$w_1 = \tilde{w}_2 + (w_1 - \tilde{w}_2) = \tilde{w}_2 + \frac{w_3 - \tilde{w}_2}{1 + \epsilon}$$

then

$$\tilde{w}_5 - w_1 = \frac{\tilde{w}_4 - w_3}{1+\epsilon}$$
$$\|\tilde{w}_5 - w_1\| < \|\tilde{w}_4 - w_3\| \leq \|\tilde{w}_2 - w_1\|$$

Which is in conflict with (6). Conclude, that $\|w_m - w_1\| \not< d_w$, which yields the result. □

Now introduce the function

$$w_+ : \widetilde{\mathcal{W}}_+(\mathbb{B}''_{id}) \to \mathcal{W}, \quad \tilde{w} \mapsto w_m + d_w \frac{w_m - \tilde{w}}{\|w_m - \tilde{w}\|} \tag{7}$$

This function is well defined because $w_m \neq \tilde{w}$ for $\mathcal{B} \in \mathbb{B}''_{id}$.

**Lemma 6**

$$\forall \mathcal{B} \in \mathbb{B}''_{id}, \tilde{w}_1 \in \widetilde{\mathcal{W}}_+(\mathcal{B}) \quad \|\tilde{w}_1 - w_+(\tilde{w}_1)\| = \max_{w \in \mathcal{W}} \|\tilde{w}_1 - w\|$$

□

**Proof:** Let $\tilde{w}_1 \in \mathcal{W}_+(\mathcal{B})$ and take $w_2 \in \mathcal{W}$ such that

$$\|\tilde{w}_1 - w_2\| = \max_{w \in \mathcal{W}} \|\tilde{w}_1 - w\| \tag{8}$$

Then

$$\|\tilde{w}_1 - w_+(\tilde{w}_1)\| = \|\tilde{w}_1 - w_m\| + \|w_m - w_+(\tilde{w}_1)\| = \|\tilde{w}_1 - w_m\| + d_w \tag{9}$$
$$\|\tilde{w}_1 - w_2\| \leq \|\tilde{w}_1 - w_m\| + \|w_m - w_2\| = \|\tilde{w}_1 - w_m\| + d_w \tag{10}$$

where the last equality follows from lemma 5. Subtracting (9) and (10) yields This gives

$$\|\tilde{w}_1 - w_2\| \leq \|\tilde{w}_1 - w_+(\tilde{w}_1)\|$$

Combining with (8) we have

$$\|\tilde{w}_1 - w_+(\tilde{w}_1)\| = \|\tilde{w}_1 - w_2\| = \max_{w \in \mathcal{W}} \|\tilde{w}_1 - w\|$$

which proofs the result. □

**Proof of theorem 2 in case (C1):** (C1) implies that $\mathbb{B}''_{id} = \mathbb{B}_{id}$. Let $\hat{\mathcal{B}}$ be any element of $\mathbb{B}_1^{opt}$. Take $\tilde{w}_1 \in \hat{\mathcal{B}}$, such that

$$\|\tilde{w}_1 - w_m\| = \min_{\mathcal{B} \in \mathbb{B}_{id}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_m\|$$

and take $w_2 = w_+(\tilde{w}_1; \mathcal{B})$. It suffices to prove, that from

$$\forall \tilde{w} \in \hat{\mathcal{B}} \quad \|\tilde{w}_1 - w_m\| \leq \|\tilde{w} - w_m\| \tag{11}$$
$$\forall \mathcal{B} \in \mathbb{B}_{id} \quad \|\tilde{w}_1 - w_m\| \leq \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_m\| \tag{12}$$

it follows

$$\forall \tilde{w} \in \hat{\mathcal{B}} \quad \|\tilde{w}_1 - w_2\| \leq \|\tilde{w} - w_2\| \tag{13}$$
$$\forall w \in \mathcal{W} \quad \|\tilde{w}_1 - w_2\| \geq \|\tilde{w}_1 - w\| \tag{14}$$
$$\forall \mathcal{B} \in \mathbb{B}_{id} \quad \|\tilde{w}_1 - w_2\| \leq \max_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\| \tag{15}$$

Note, that (14) implies

$$\forall w \in \mathcal{W} \quad \|\tilde{w}_1 - w_2\| \geq \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\|$$

First we prove (13). Assume,

$$\exists \tilde{w}_3 \in \hat{\mathcal{B}} \quad \|\tilde{w}_1 - w_2\| > \|\tilde{w}_3 - w_2\|$$

Let $\epsilon \in (0,1)$ be such that $w_m = \tilde{w}_1 + \epsilon(w_2 - \tilde{w}_1)$. Because of the convexity of $\hat{\mathcal{B}}$

$$\tilde{w}_4 := \tilde{w}_1 + \epsilon(\tilde{w}_3 - \tilde{w}_1) \in \hat{\mathcal{B}}$$

Further

$$\tilde{w}_4 - w_m = \epsilon(\tilde{w}_3 - \tilde{w}_1) + \tilde{w}_1 - \tilde{w}_1 - \epsilon(w_2 - \tilde{w}_1) = \epsilon(\tilde{w}_3 - w_2)$$
$$\|\tilde{w}_4 - w_m\| = \epsilon\|\tilde{w}_3 - w_2\| < \epsilon\|\tilde{w}_1 - w_2\| = \|\tilde{w}_1 - w_m\|$$

This is in contradiction with (11), so that (13) is proven. To prove (14), assume

$$\exists w \in \mathcal{W} \quad \|\tilde{w}_1 - w_2\| < \|\tilde{w}_1 - w\|$$

then

$$\|\tilde{w}_1 - w\| \leq \|\tilde{w}_1 - w_m\| + \|w_m - w\|$$
$$\|w_m - w\| \geq \|\tilde{w}_1 - w\| - \|\tilde{w}_1 - w_m\| > \|\tilde{w}_1 - w_2\| - \|\tilde{w}_1 - w_m\| = d_w$$

Apparently $w \notin \mathcal{W}$, which proves (14). Finally, (15) can be shown as follows: assume to the contrary

$$\exists \mathcal{B} \in \mathbb{B}_{id}, \tilde{w}_3 \in \widetilde{\mathcal{W}}_+(\mathcal{B}) \quad \max_{w \in \mathcal{W}} \|\tilde{w}_3 - w\| < \|\tilde{w}_1 - w_2\| = \|\tilde{w}_1 - w_m\| + d_w$$

Because (lemma 6)

$$\max_{w \in \mathcal{W}} \|\tilde{w}_3 - w\| = \|\tilde{w}_3 - w_+(\tilde{w}_3)\| = \|\tilde{w}_3 - w_m\| + d_w$$

we conclude

$$\|\tilde{w}_3 - w_m\| < \|\tilde{w}_1 - w_m\|$$

This conflicts with (12). □

**Lemma 7**

$$\forall \mathcal{B} \in \mathbb{B}''_{id} \quad \max_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w_m\| > d_w$$

□

**Proof:** This is an immediate consequence of (7) and lemma 6 □

**Lemma 8**

$$\forall \mathcal{B} \in \mathbb{B}_{id} \setminus \mathbb{B}''_{id}, \mathcal{B} \textit{ linear} \quad \max_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\| = d_w$$

□

**Proof:** Consider first

$$\exists \mathcal{B}' \in \mathbb{B}_{id} \setminus \mathbb{B}''_{id} \quad \max_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}'} \|\tilde{w} - w\| = 0$$

This means $\mathcal{B}'$ can explain all data in $\mathcal{W}$. From the linearity of $\mathcal{B}'$ we can then conclude, that $\mathcal{B}' = \mathbb{W}$. We assumed such "senseless" models were excluded from $\mathbb{B}_{id}$. So now consider

$$\max_{w \in \mathcal{W}} \min_{\tilde{w} \in \mathcal{B}} \|\tilde{w} - w\| = \epsilon > 0$$

Because $w_m \in \mathcal{B}$ it holds $\epsilon \leq d_w$. Let $\min_{\tilde{w}} \|\tilde{w} - w\|$ be maximal for $w = w_1$ and $\|\tilde{w} - w_1\|$ be minimal for $\tilde{w} = \tilde{w}_2$ and take

$$w_3 = w_m + d_w \frac{w_1 - \tilde{w}_2}{\|\tilde{w}_2 - w_1\|} \in \mathcal{W}$$

It holds

$$\nexists \tilde{w}_4 \in \mathcal{B} \quad \|\tilde{w}_4 - w_3\| < d_w$$

To see this, assume such a $\tilde{w}_4$ does exist. Because of the linearity of $\mathcal{B}$

$$\tilde{w}_5 := \tilde{w}_2 + \frac{\|\tilde{w}_2 - w_1\|}{d_w}(\tilde{w}_4 - w_m) \in \mathcal{B}$$

$$\tilde{w}_5 - w_1 = \tilde{w}_2 - w_1 + \frac{\|\tilde{w}_2 - w_1\|}{d_w}(\tilde{w}_4 - w_m) = \frac{\|\tilde{w}_2 - w_1\|}{d_w}(\tilde{w}_4 - w_3)$$

$$\|\tilde{w}_5 - w_1\| = \|\tilde{w}_2 - w_1\| \frac{\|\tilde{w}_4 - w_3\|}{d_w} < \|\tilde{w}_2 - w_1\|$$

which gives a contradiction. So we have $\epsilon \geq d_w$. We already had $\epsilon \leq d_w$, which completes the proof. □

**Proof of theorem 2 in case (C2):** We need only consider the case in which (C1) does not hold, as we may otherwise use the proof given earlier. If (C1) does not hold, the set $\mathbb{B}_{id} \setminus \mathbb{B}''_{id} =: \mathbb{B}^-_{id}$ is not empty. Obviously, $\mathbb{B}_1^{opt} = \mathbb{B}^-_{id}$ then. From lemma 7 and 8 it now follows, that a model is an element of $\mathbb{B}_3^{opt}$ if and only if it belongs to $\mathbb{B}^-_{id}$. □

# References

[Ljung (1987)] Ljung, L. *System Identification: Theory for the User.* Prentice-Hall, Englewood Cliffs, New Jersey, 1987.

[Milanese (1989)] Milanese, M. *Robustness in Identification and Control*, chap. Estimation and Prediction in the Presence of Unknown but Bounded Uncertainty: A Survey. Plenum Press, 1989.

[Norton (1987)] Norton, J. Identification of Parameter Bounds for ARMAX Models from Records with Bounded Noise. *International Journal of Control*, **45**(2):375–390, 1987.

[Söderström and Stoica (1989)] Söderström, T. and P. Stoica. *System Identification.* Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

[Walter and Piet-Lahanier (1990)] Walter, E. and H. Piet-Lahanier. Estimation of parameter bounds from bounded-error data: a survey. *Mathematics and Computers in Simulation*, **32**(5 & 6):449–468, 1990.

[Willems (1986)] Willems, J. C. From Time Series to Linear System — Part I. Finite Dimensional Linear Time Invariant Systems. *Automatica*, **22**(5):561–580, 1986.