

Identification of Model Uncertainty for Control Design

Douwe K. de Vries

September 1994

Identification of Model Uncertainty for Control Design

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus
Prof. ir. K.F. Wakker,
in het openbaar te verdedigen
ten overstaan van een commissie,
door het College van Dekanen aangewezen,
op maandag 26 september 1994 te 16.00 uur

door

Douwe Klaas DE VRIES

Werktuigbouwkundig Ingenieur

geboren te Geldrop

Dit proefschrift is goedgekeurd door de promotor
Prof. ir. O.H. Bosgra
en de toegevoegd promotor Dr. ir. P.M.J. Van den Hof.

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Vries, Douwe Klaas de

Identification of Model Uncertainty for Control Design /
Douwe Klaas de Vries. - Delft : Delft University of
Technology, Faculty of Mechanical Engineering and
Marine Technology. - Ill.

Thesis Technische Universiteit Delft. - With ref.

ISBN 90-370-0110-6

Subject headings: system identification / model
uncertainty / robust control.

Abstract

This thesis deals with the problem of model uncertainty quantification from experimental data. Given a set of measured input and output data points, the aim is to establish an upper bound on a relevant norm on the difference between the true system and a suitable model of that system. More specifically, the problem addressed in this thesis is how one should represent and quantify model errors, in such a way that the identification result is suitable to serve as a basis for high performance robust control design.

A critical evaluation of the related literature is made, showing that a major part of the problem is shifted to the prior information, while no adequate procedures exist to obtain this information.

A first approach towards solving the problem makes use of a hard upper bound on the noise in the frequency domain as prior information, and yields a hard (worst case) error bound. First an error bound on a finite number of frequency points is established, and subsequently this error bound is interpolated. The merit of this procedure is that the model with respect to which the error bound is calculated can be specified by the user. A major disadvantage is however that the necessary prior information is hard to obtain.

Based on an extensive evaluation of options and consequences, a mixed averaging – worst case embedding of the modelling errors now is proposed, to effectively account for accumulations of both random as well as structural errors in a situation where only the dominant part of the system is linear and time invariant. That is, we argue that the noise should be considered as stochastic, whereas undermodelling should be regarded as unknown but bounded.

Given this choice for the embedding, a second procedure is presented to estimate a model together with a bound on the model uncertainty from experimental data. A periodic input signal and data segmentation are employed to effectively distinguish between the random and structural errors. A model is estimated over each period of the input signal, which results in a set of models. The average over this set of models yields the final estimate, while the mutual differences between the models in the set provide information about the uncertainty in this final estimate. Specializing to linear time invariant systems with stochastic noise, a closed form confidence interval for the error due to the noise now can be obtained asymptotically in the number of

data points, and an adequate estimate of the error due to undermodelling can be provided. Essentially, the only prior information that is needed with this approach is the qualitative information that the noise forgets its past, and that the system is linear time invariant and exponentially stable. Both the error due to the noise as well as the error due to undermodelling are estimated from the experimental data. Furthermore, a highly flexible model structure is provided by the use of system based orthonormal basis functions, allowing for high quality parametric estimates, and the procedure also can be applied in closed loop situations. On the basis of simulation examples and applications the performance of this model error estimation procedure is studied, also under deviations from the linear time invariant case with stochastic noise. Tight and reliable error bounds are obtained, and we are able to design high performance robust controllers. We conclude that the combination of the above properties indeed leads to models and model error bounds which are suitable to serve as a basis for high performance robust control design.

Within the above framework clear indications for model order selection and input design also can be provided, and a powerful method for model validation follows. Additionally, an asymptotic analysis is provided for frequency domain transfer function estimation using system based orthonormal basis functions, as applied in the model error estimation procedure above.

A topic for further research is to adapt the procedure to allow for quasi-stationary input signals, which certainly appears to be possible.

Acknowledgements

For being able to accomplish this thesis I am indebted to many people.

First of all, I would like to thank Professor Okko Bosgra for encouraging me to become a Ph.D. student, while giving me the opportunity to join the model research environment of the Measurement and Control group. Also, I would like to thank him for the discussions in the final stage of my work, again forcing me to look beyond accepted boundaries, and so providing a better perspective for this thesis.

I would like to express my profound gratitude to Paul Van den Hof for his excellence in advising and guiding me. Working with him was both very stimulating and relaxed, and his critical, fundamental and inspiring comments have added much to this thesis. I much appreciate the freedom he allowed to pursue my own ideas, while always keeping an eye on my work and taking the time to discuss results and problems.

Special thanks go to my room-mate Peter Bongers for the pleasant years, the fruitful discussions, and the help he was always willing to provide. At times he must have wondered what was so interesting outside our window that I kept staring at it.

To everyone in the Measurement and Control group: thanks for your help and the ever present friendly and enthusiastic atmosphere, I very much enjoyed my time as a Ph.D. student in Delft.

Finally, I would like to thank my parents for their love and support which they gave abundantly again and again over the years. They have my deepest respect and love.

Contents

Abstract	v
Acknowledgements	vii
1 Introduction	1
1.1 Motivation and background	1
1.2 A description of the problem field	4
1.3 Reader's guide and thesis outline	5
2 State of the art: a critical evaluation	7
2.1 Embedding	7
2.2 Parametric prediction error identification with filtered independent identically distributed noise	8
2.3 Spectral analysis	11
2.4 Parametric identification with stochastic embedding of noise and undermodelling	13
2.5 Parameter set identification with bounded time domain errors	16
2.6 Transfer function set identification with bounded frequency domain errors	20
2.7 Model validation	26
2.8 Identification of nominal models for control design	28
3 Problem formulation and thesis contribution	30
3.1 Problem formulation	30
3.2 The basic difficulty	32
3.3 Main criticism on existing techniques	33
3.4 Deterministic versus stochastic noise	34
3.5 Deterministic versus stochastic embedding of undermodelling	38
3.6 A broader perspective: averaging and structural errors	39
3.7 Intended use of the model and the error bounds	40
3.8 Model structure, identification criterion and undermodelling	41
3.9 Parametric techniques or spectral analysis	42
3.10 Model validation	42

3.11	Prior information and extrapolation	43
3.12	Requirements	44
3.13	Overview of thesis contributions	45
3.13.1	Error bounds from bounded frequency domain noise	45
3.13.2	A mixed averaging – worst case approach	47
4	Nonparametric frequency domain hard error bound	50
4.1	Introduction	50
4.2	Assumptions and prior information	54
4.3	Discrete error bound	55
4.3.1	Motivation	55
4.3.2	SISO systems	55
4.3.3	MIMO systems	57
4.3.4	Remarks	59
4.4	Continuous error bound	59
4.4.1	Motivation	59
4.4.2	Bounds on derivatives	60
4.4.3	Interpolation	61
4.4.4	Remarks	63
4.5	Validation of prior information	65
4.6	Relation with control design specifications	65
4.7	Example	67
4.8	Conclusions	69
5	Mixed averaging – worst case model error bounds	72
5.1	Introduction	73
5.2	Assumptions and prior information	78
5.3	Error bound for ETFE	79
5.3.1	Introduction	79
5.3.2	Nonparametric transfer function estimate	80
5.3.3	Input design	84
5.4	Error bound for generalized FIR estimate	85
5.4.1	Introduction	85
5.4.2	Parameter estimate	86
5.4.3	Parametric transfer function estimate	91
5.4.4	Mixed parametric-nonparametric uncertainty	96
5.4.5	Undisturbed output signal	98
5.4.6	Disturbed output signal	101
5.5	Remarks	106
5.6	Recapitulation of results	107
5.7	Example	107
5.8	Monte Carlo analysis	110

5.8.1	Introduction	110
5.8.2	Description	114
5.8.3	Results	115
5.8.4	Discussion	118
5.9	Application to a nonlinear simulation system	118
5.9.1	Introduction	118
5.9.2	The nonlinear simulation system	118
5.9.3	Results	119
5.9.4	Discussion	122
5.10	Application to time varying simulation systems	126
5.10.1	Simulation system 1	126
5.10.2	Simulation system 2	131
5.11	Conclusions	135
6	Quantification of uncertainty in closed loop	137
6.1	Framework	137
6.2	Coprime factor representation	138
6.3	Error quantification	140
7	Asymptotic bias and variance expressions	143
7.1	Introduction	143
7.2	Assumptions	144
7.3	Asymptotic analysis of parameter estimates	146
7.4	Asymptotic analysis of transfer function estimates	153
7.5	Conclusions	156
8	Applications	158
8.1	Application to a wind turbine system	159
8.1.1	Introduction	159
8.1.2	The wind turbine system	159
8.1.3	Model error estimation	162
8.1.4	Robust stability	170
8.1.5	Controller implementation	172
8.1.6	Application to nonlinear simulation model	173
8.1.7	Conclusions	176
8.2	Application to pick-up mechanism of CD player	178
8.2.1	Introduction	178
8.2.2	The compact disc mechanism	179
8.2.3	Model error estimation	181
8.2.4	Conclusions	197

9	Conclusions and recommendations	199
9.1	Conclusions	199
9.2	Recommendations for further research	201
A	Proofs for Chapter 4	205
A.1	Proof of Theorem 4.3.2	205
	A.1.1 Properties of the N point DFT	205
	A.1.2 Proof	206
A.2	Proof of Theorem 4.3.5	208
A.3	Proof of Proposition 4.4.1	209
A.4	Proof of Proposition 4.4.2	209
B	Proofs for Chapter 5	211
B.1	Proof of Lemma 5.3.1	211
B.2	Proof of Lemma 5.3.3	212
B.3	Proof of Lemma 5.3.4	212
B.4	Proof of Theorem 5.3.5	214
B.5	Proof of new elements in Theorem 5.4.2	215
B.6	Proof of new elements in Theorem 5.4.4	216
B.7	Proof of new elements in Theorem 5.4.5	216
B.8	Proof of new elements in Theorem 5.4.8	218
C	Proofs for Chapter 6	221
C.1	Proof of Theorem 6.3.3	221
D	Proofs for Chapter 7	227
D.1	Proof of Lemma 7.3.2	227
D.2	Proof of Theorem 7.3.3	227
D.3	Proof of Lemma 7.3.4	228
D.4	Proof of Theorem 7.3.5	230
D.5	Proof of Theorem 7.3.6	231
D.6	Proof of Theorem 7.3.7	235
D.7	Proof of Theorem 7.4.1	237
D.8	Proof of Corollary 7.4.2	238
D.9	Proof of Theorem 7.4.3	240
E	System based orthonormal basis functions	242
E.1	Properties	242
E.2	Examples	246
E.3	Merits	247
	Bibliography	248

Glossary of symbols

264

Samenvatting

270

Curriculum Vitae

272

Chapter 1

Introduction

1.1 Motivation and background

Modelling the behaviour of a process plays a central role in science and technology. These models are used to gain a better understanding of the process, and of the interaction of the process with its environment. This knowledge in turn can be used to predict and anticipate future outputs of the process, and to influence the process in such a way that it exhibits a more desirable behaviour. These three major reasons for modelling the behaviour of a process are usually called: analysis, prediction, and control.

Clearly, models can be used to obtain faster, more reliable manufacturing processes, with a higher or more constant quality. Also the energy and materials consumption, the environmental burden and safety can be improved. Of course all this also can have quite some impact on the economic profit. Examples of the use of models for analysis (simulation), prediction and control are abundant in aerospace technology, chemical industry, consumer electronics, robotics, economics, and so on.

Basically there are two ways to obtain a model of a system: white box and black box modelling. White box modelling is modelling from first principles: the model is a combination of the physical laws governing the system. In black box modelling the input and output signals of the system are measured and a mathematical model is constructed to describe the measured behaviour, without the intention to describe the internal behaviour of the system or the physical processes that are responsible for the observed behaviour. Black box modelling is also known as experimental modelling or system identification, and is the subject of this thesis. It can of course be argued that white box modelling is, at least partly, just combining generally accepted black box models of subsystems of the system. In practice both methods of modelling are often combined, to obtain both the physical insight resulting from white box modelling, as well as a model that is in accordance with the observed behaviour of the system as obtained from black box modelling. This is called grey box modelling for reasons

that should be obvious. Also, black box modelling can be useful for verification of a theoretical (white box) model.

Two important principles from the theory of black box modelling or system identification are that the model should not add to the observed data, and that the model in general only will be an approximation of the true system.

The fact that a model should not add to the observed data is obvious but not trivial in its implications and often neglected. Occam [135, 127] stated that knowledge should only be obtained from observation, logical necessity or divine revelation, and that things not known to exist should not be postulated as existing, unless absolutely necessary¹. This is also known as the principle of parsimony: the simplest model or hypothesis that is in accordance with observations should be preferred. Using the terminology of Willems [185] this can be made more precise as follows. A model that is able to reproduce the measured data is called an unfalsified model, and the behaviour of a system is defined as the class of all possible (input and output) signal trajectories. Among the class of unfalsified models, the model whose behaviour is a subset of the behaviour of another model is said to be more powerful. In other words, a model which is not powerful does state considerably more than the measured data, and thus is based only partly on the observations. If a model adds too much to the measured data it even becomes useless; the behaviour of the model can be almost anything, which is not very informative. Clearly, all unfalsified models could be true, but it must be considered a sound scientific principle to favor the more powerful ones. In system identification this principle is often reflected by the performance of the model on new data or by the tradeoff between bias and variance errors.

The fact that the model is in general only an approximation of the true system is a consequence of the fact that the true system usually is far too complex to be modelled exactly. In addition, even if the system could be modelled exactly, estimation errors will be present due to the finite length of the data set. These errors are introduced by the noise acting on the system, and unknown initial conditions of the system. Moreover, from a certain point, increasing the complexity of the model set will only increase the uncertainty in the estimated model because only a restricted amount of measured data is available: there is an increasing amount of freedom in the model set to approximately reproduce the measured data, but there is a lack of information how to choose between the different possibilities. This clearly is related to the discussion above. Furthermore, we only want to model that part of the system in which we are interested. This means that we only want to model that part of the system that is relevant for the intended use of the model. The modelling of additional effects is only disadvantageous, the model becomes more complex but it does not become better. Finally, accurate and complex models are expensive. Hence, we even do not want to model the relevant part of the system exactly, but only to some desired level of accuracy. Note that the above implies that it can be advantageous to identify different models for analysis, prediction

¹"pluralitas non est ponenda sine necessitate" to be more precise.

and control design, because the matching relevant part of the behaviour of the true system can differ significantly. In conclusion, exact modelling usually is impossible or too expensive, and unnecessary.

Given that a model is only an approximation of the system, the problem that is addressed in this thesis is the quantification of model uncertainty from experimental data, in such a way that the resulting models and model uncertainty descriptions are suitable for their intended use. By model uncertainty quantification we mean the provision of an upper bound for some norm on the difference between the model and the system (the model error). The motivation to look at this problem is primarily that a model is virtually useless without some indication of its quality. A reliable quantification of the model uncertainty is imperative when using the model, whether for analysis, prediction or control design. Without such a quantification the user of the model may be led complete astray. In order to be able to quantify the accuracy of the model, or to be able to guarantee a certain level of accuracy in view of the intended use of the model, an upper bound on the model error must be available.

An uncertainty description is also closely related with the principle of parsimony. A model having a relatively large uncertainty is an indication that the model either is not capable of approximately reproducing the data (not unfalsified), or that it adds too much to the data (not powerful). In other words, the model is too simple or too complex. Thus, the model with the smallest uncertainty could be regarded as the most powerful one. As a consequence, an uncertainty description provides an excellent tool for model order selection; the model order (model complexity) should be such that the lowest uncertainty is obtained.

This thesis is motivated by and directed towards situations where the intended use of the model is to serve as a basis for high performance robust control design, although other applications can be addressed equally well within our framework. The basic problem in identification for control design is that a small difference between system and model in open loop can result in a large difference in closed loop. The situation that the model is stabilized and the system is not, is not at all imaginary. Thus, an accurate quantification of the model uncertainty is of major importance in this context. The recent vigorous search for model uncertainty descriptions was triggered by the developments of robust control during the last decade. Given a model and an upper bound on the model error, methods have become available to design controllers that can provide a certain level of guaranteed performance when applied to the system, see e.g. [39]. Clearly, these robust control design methods only amplify the importance of an accurate quantification of the model uncertainty. The search for model uncertainty descriptions in system identification is however much older, see e.g. [86, 159, 108].

1.2 A description of the problem field

In this section we will provide a first introduction to the field of model uncertainty estimation, and the problems involved.

System identification addresses the problem of finding, from a set of measured data points, a model that is in some sense close to the system that generated the data. Model uncertainty estimation should provide a statement of how close the model and the system actually are.

To make these problems amenable for analysis and synthesis, first of all a description of the data generating process is needed. An assumption that is extensively (in fact, almost solely) used is that the system is linear, time invariant, and stable. This enormously facilitates the analysis and synthesis problem, to a level where it can be handled, although the problem still remains to be difficult enough. The above assumption is of course only an approximation of reality. Even so, the approximations involved are often justified and lead to good results in many cases, see [106]. To account for these approximations, for the fact that the system also reacts to stimuli that are not measured and therefore unknown, and for the finite precision of the measurements, noise is added to the description of the data generating process. In other words, the noise is needed to close the inevitable gap between the linear system description and the measured input and output signals. Note that the above implies that the noise can be assumed to be uncorrelated (linearly independent) with the input signal (in open loop).

Hence, it is assumed that the system, and the measurement data that is obtained from this system, allow a description

$$y(t) = G_o(q)u(t) + v(t) \quad (1.1)$$

with $y(t)$ the output signal, $u(t)$ the input signal, $v(t)$ an additive noise, q the shift operator ($q x(t) = x(t+1)$), and G_o the causal linear time invariant transfer function of the system. For the moment we will also assume that the system is stable and single-input single-output (SISO). The transfer function can be written as

$$G_o(z) = \sum_{k=0}^{\infty} g_o(k)z^{-k} \quad (1.2)$$

with $g_o(k)$ the impulse response of the plant.

A preliminary description of the problem area addressed in this thesis now reads as follows. *Given a set of N data points $y(t), u(t), t = 0, \dots, N-1$, the problem is to find the set of all models which are not unlikely to have generated the observed data.* Moreover, this thesis is directed to situations where the intended use of such a set of models is to serve as a basis for high performance robust control design. Thus, when considering the problem, we will focus on the control design point of view.

The above formulation is not yet complete. A set of N data points contains no information at all about the tail of the impulse response $g_o(k)$, $k = N, \dots, \infty$. This implies that, even if we could accurately estimate the first part of the impulse response, we still could not state anything about a set containing the system $G_o(q)$. Furthermore, if the unobserved past of the input is completely arbitrary, then the past input can arbitrarily influence the measured data. Again this implies that we can induce nothing about the system. Similarly, if nothing is known about the noise, then we can be arbitrarily fooled by the noise. Hence, the problem as formulated above does not make sense, unless further restrictions are being imposed on

- the system $G_o(q)$
- the past inputs $u(t)$, $t < 0$
- the noise $v(t)$, $t = 0, \dots, N - 1$

by means of additional assumptions or prior information, so as to restrict the solution set to a meaningful one. Thus we are actually faced with two problems

1. find a set of realistic additional assumptions and a priori information
2. using the assumptions and the a priori information, construct from the measured data a set of models, not larger than necessary, that is likely to contain the data generating system $G_o(q)$

It should be stressed that the first problem is of decisive importance to obtain model uncertainty estimation procedures that are valuable in practice. We will call this first problem the problem of *embedding*.

1.3 Reader's guide and thesis outline

Chapter 3 and Chapter 5 contain the main contributions and results of this thesis. Chapter 7 is of independent interest. One should read Chapter 3 before reading Chapter 4 or Chapter 5, and one should be familiar with the results of Chapter 5 before reading the remaining chapters of this thesis.

In Chapter 2 an overview and a critical evaluation of existing methods to solve this problem is presented.

In Chapter 3 the problem that is addressed in this thesis is formulated. The basic properties and implications of different approaches towards solving this problem are discussed, leading to specific choices with regard to the embedding (prior information and assumptions) of the problem, and a set of essential conditions and technical requirements for an adequate solution. An overview of the contributions of this thesis towards solving this problem is presented.

In Chapter 4 a frequency domain approach towards solving the problem is presented which yields a hard (worst case) error bound. First an error bound on a finite number of frequency points is established, and subsequently this error bound is interpolated. The merit of this procedure is that the model with respect to which the error bound is calculated can be specified by the user. The disadvantage is that the necessary prior information is hard to obtain.

In Chapter 5 a mixed averaging – worst case embedding of the modelling errors is employed, based on the discussions of Chapter 3. That is, a stochastic setting is used for the noise (averaging errors), whereas a deterministic framework is used for the error due to undermodelling (structural errors). Using a periodic input signal we can effectively distinguish between noise and undermodelling errors. A closed form confidence interval for the error due to the noise is obtained asymptotically in the number of data points, and an adequate estimate of the error due to undermodelling is provided. Thus, both the error due to the noise and the error due to undermodelling are estimated from the data. Essentially, the only prior information that is needed with this approach is the qualitative information that the noise satisfies a mixing condition, and that the system is linear time invariant and exponentially stable. Furthermore, a highly flexible model structure is provided by the use of system based orthonormal basis functions, and tight and reliable error bounds are obtained. The robustness of this model error estimation procedure to deviations from the linear time invariant case with mixing noise is demonstrated on the basis of simulation examples. The combination of the above properties leads to models and model error bounds which are suitable to serve as a basis for high performance robust control design.

In Chapter 4 the results are extended to multi-input multi-output (MIMO) systems.

In Chapter 6 the results of Chapter 4 and Chapter 5 are extended to closed loop situations and unstable systems.

In Chapter 7 an asymptotic analysis is provided for frequency domain transfer function estimation using system based orthonormal basis functions.

In Chapter 8 the model error estimation procedure of Chapter 5 is applied to a wind turbine system and a compact disc player, in order to validate the practical applicability of this procedure and to verify the properties claimed above.

Finally, Chapter 9 contains the conclusions and recommendations for further research.

Chapter 2

State of the art: a critical evaluation

Summary

An overview and evaluation is presented of the existing literature on model uncertainty quantification from experimental data.

2.1 Embedding

In the following sections we will discuss and critically evaluate existing methods for model uncertainty quantification. However, first we will clarify the process of embedding (see Section 1.2) and its significance in this section, because of the prominent role it plays in the different methods of model uncertainty estimation. We certainly should be aware of this process. As stated by Hjalmarsson [83], the priors used in existing model uncertainty estimation methods are governed more by a strive to get simple methods than by realism.

In system identification the set of observed data points always is embedded in a larger set to account for data sequences that also could have occurred while using the same input signal and initial conditions of the system. This means that it is assumed that the system is also affected by other signals (called noise or disturbances) that were not measured, and a characterization of these signals is given. Such a characterization is e.g. that the noise belongs to a bounded set, or that the noise is a realization of a stochastic process. As was discussed in Section 1.2, we also have to embed the system and the past input signals to be able to obtain a meaningful description of the uncertainty in the estimated model. For example, we could assume that an upper bound on the margin of relative stability of the system is known, or that an upper bound on the decay rate of the impulse response is known, and so on. For the input we could e.g. assume that an upper bound on its absolute value is known, or that it is a realization of stochastic process. Clearly, there is a lot of freedom in the actual choice for the embedding of the set of observed data points, but the embedding is required to

be such that the class of possible data sequences indeed is restricted to a meaningful one, and preferably should be such that this class is not larger than strictly necessary to account for the observed data.

As was already noted, the specific way of embedding which is used has a decisive influence on the resulting identification methods (e.g. least squares parameter estimate or solving for the feasible parameter set), on the properties of these methods (e.g. consistency, optimality), on the size of the error bounds (tight or not), and on the nature of the error bounds obtained (hard or soft, i.e. deterministic or probabilistic). This latter point deserves some clarification in view of the scope of this thesis. Hard a priori information (bounded noise, and so on) will yield hard a posteriori information, i.e. hard error bounds. Similarly, soft priors will result in soft a posteriori information. Indeed, the specific way of embedding (in the form of assumptions and available prior information) actually is one of the major topics in model uncertainty estimation.

Distinguishing between different methods of embedding of the noise and the system, the current state of model uncertainty quantification in system identification will be discussed very concisely in the following sections. We will describe the different problem formulations, the available results, and discuss specifically where these techniques fall short with respect to model uncertainty quantification. A general in-depth discussion of merits and demerits of different problem formulations, and required properties of solutions, will be given in Chapter 3, including an extensive discussion on the problem of embedding.

2.2 Parametric prediction error identification with filtered independent identically distributed noise

Comprehensible accounts of the general field of prediction error methods are given in [43, 58, 110, 106, 166]. The key results with respect to uncertainty descriptions are given in [108, 113, 105, 104, 181, 192]. Much of the credit for these results belongs to Ljung, who has been a driving force in this field. The latest contributions can be found in [180, 111, 85, 84, 83]. Based on the discussion in Section 1.1, we will only address results that allow for undermodelling, that is, where the system is not assumed to be an element of the model set.

It is assumed that the system, and the measurement data that is obtained from this system, allow a description [106]

$$y(t) = G_o(q)u(t) + H_o(q)e(t)$$

where q is the shift operator, G_o the system, which is required to be stable, H_o the noise filter, which is required to be stable and stably invertible, and $e(t)$ a sequence of independent identically distributed random variables, with zero mean, variance σ_e^2 , bounded fourth moment, and independent of the input signal $u(t)$. Let $\Phi_u(\omega)$ denote

the spectrum of $u(t)$, and let $\Phi_v(\omega) = \sigma_e^2 |H_o(e^{j\omega})|^2$ denote the spectrum of the noise $v(t) = H_o(q)e(t)$.

The model is parameterized as

$$y(t) = G(q, \theta)u(t) + H(q, \theta)e(t) \quad (2.1)$$

and the prediction error $\varepsilon(t, \theta)$ is defined as

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t|\theta) = H^{-1}(q, \theta)y(t) - H^{-1}(q, \theta)G(q, \theta)u(t)$$

Under some quite general regularity conditions there holds that, using a quadratic criterion

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{N} \sum_{t=0}^{N-1} \varepsilon^2(t, \theta)$$

the estimated parameters $\hat{\theta}$ are asymptotically (in the number of data points N) normally distributed, see [108, 106],

$$\sqrt{N}(\hat{\theta} - \theta^*) \in As\mathcal{N}(0, P) \quad (2.2)$$

where P is the asymptotic covariance matrix, and θ^* is the limiting estimate

$$\theta^* = \arg \min_{\theta} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=0}^{N-1} \mathbb{E}[\varepsilon^2(t, \theta)] \quad (2.3)$$

Furthermore

$$\hat{\theta} \rightarrow \theta^* \quad \text{as} \quad N \rightarrow \infty \quad (2.4)$$

If the limit model $G(q, \theta^*)$, $H(q, \theta^*)$ gives white prediction errors then a consistent estimate of the asymptotic (in N) covariance matrix P can be obtained easily. However, white prediction errors in general imply that there is no undermodelling and no bias in the estimate, implying that the model set must at least contain the true system. Hjalmarsson [85, 83] shows that the variance expression for white prediction errors indeed can be quite erroneous when the prediction errors in fact are strongly correlated. An expression for the asymptotic (in N) covariance matrix P of the estimated parameters in case of colored residuals has been known for long, but only recently a consistent estimate has been found, see Hjalmarsson [85, 83].

Under certain regularity conditions the estimated transfer function is also normally distributed asymptotically in N , see [104]. That is, for general black box model structures there holds for a fixed model order n and asymptotically in N

$$\sqrt{N} \begin{bmatrix} G(e^{j\omega}, \hat{\theta}) - G(e^{j\omega}, \theta^*) \\ H(e^{j\omega}, \hat{\theta}) - H(e^{j\omega}, \theta^*) \end{bmatrix} \in As\mathcal{N}(0, Q(\omega, n))$$

where

$$\lim_{n \rightarrow \infty} \frac{1}{n} Q(\omega, n) = \Phi_v(\omega) \begin{pmatrix} \Phi_u(\omega) & 0 \\ 0 & \sigma_\varepsilon^2 \end{pmatrix}^{-1} \quad (2.5)$$

under the conditions that

$$\begin{aligned} G(e^{j\omega}, \theta^*) &\rightarrow G_o(e^{j\omega}) \\ H(e^{j\omega}, \theta^*) &\rightarrow H_o(e^{j\omega}) \end{aligned}$$

as $n \rightarrow \infty$.

A more precise statement is given in [113] for finite impulse response (FIR) models. When allowing the model order n to increase with the number of data points N then, in the limit as $N \rightarrow \infty$, $n \rightarrow \infty$, and $n^2/N \rightarrow 0$ (note the doubly asymptotic nature of these conditions ¹), there holds

$$\sqrt{\frac{N}{n}} \left(G(e^{j\omega}, \hat{\theta}) - G(e^{j\omega}, \theta^*) \right) \in As\mathcal{N} \left(0, \frac{\Phi_v(\omega)}{\Phi_u(\omega)} \right) \quad (2.6)$$

and

$$|G_o(e^{j\omega}) - G(e^{j\omega}, \hat{\theta})| \rightarrow 0 \quad \text{w.p. 1} \quad \text{as} \quad N \rightarrow \infty \quad (2.7)$$

In (2.5) and (2.6) $\Phi_u(\omega)$ is a design variable, and $\Phi_v(\omega)$ can be estimated consistently using spectral analysis techniques, see e.g. [20].

An approximate result for the variance of the estimated transfer function that is not asymptotic in the model order n can be obtained from the consistent estimate of the asymptotic (in N) covariance matrix P of the estimated parameters as given by [85, 83], using a first order expansion (see e.g. [104, 83])

$$G(e^{j\omega}, \theta^*) - G(e^{j\omega}, \hat{\theta}) \approx \frac{d}{d\theta} G(e^{j\omega}, \hat{\theta})(\theta^* - \hat{\theta})$$

The corresponding approximate variance expression is

$$\mathbb{E} \left[|G(e^{j\omega}, \theta^*) - G(e^{j\omega}, \hat{\theta})|^2 \right] \approx \left(\frac{d}{d\theta} G(e^{j\omega}, \hat{\theta}) \right)^* P \frac{d}{d\theta} G(e^{j\omega}, \hat{\theta})$$

The expression is exact if the model is linear in the parameters.

An implicit asymptotic (in N) expression for the bias in the frequency domain is available as

$$\theta^* = \arg \min_{\theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} [|G_o(e^{j\omega}) - G(e^{j\omega}, \theta)|^2 \Phi_u(\omega) + \Phi_v(\omega)] \frac{1}{|H(e^{j\omega}, \theta)|^2} d\omega \quad (2.8)$$

However, while (2.8) certainly does allow for an interesting qualitative discussion of the factors affecting bias, it does not provide an explicit expression for the bias or a bound on it.

¹Some results are more asymptotic than others.

The error bounds given by Zhu [192, 193, 194] are based on (2.6) or on the analogous result for spectral analysis methods (2.11). The error bounds are obtained by neglecting the bias contribution and the fact that the variance is estimated (both the bias and the variance errors). As a result, a precise statement about the probability of the error bounds cannot be provided. In [193] it is argued that neglecting the bias contribution and the fact that the variance is estimated is reasonable, since asymptotically in the model order the estimate of the transfer function is unbiased, and the estimate of the variance is consistent. In practical applications, where the model order and the number of data points are finite, this may however be rather too optimistic a point of view. Moreover, there are basic objections against a neglectable bias error, see Section 3.8.

Clearly, while the results stated above certainly allow for an interesting analysis, they are not yet fit to actually bound the identification error. There is no explicit expression for the bias (undermodelling error) available, and the accuracy (bias and/or variance) of estimates of the variance (using spectral estimates of the noise spectrum, or estimates of the covariance matrix of the estimated parameters) is unknown. Only for finite impulse response type of models, no undermodelling, white noise and asymptotically in N , a closed form (contains no unknowns) confidence interval for the estimated parameters is established, see Ljung [106, Appendix II].

For a more general discussion on the merits of bias errors, and the demerits of the use of prior information to bound these errors, we refer to Section 3.8 and Section 3.11 respectively.

2.3 Spectral analysis

Extensive accounts of the field of spectral analysis are given in [86, 19, 20, 147, 137]. An excellent comparison between the parametric prediction error methods and spectral analysis is given by Ljung [105].

The spectral estimate $\hat{G}(e^{j\omega})$ of a transfer function $G_o(e^{j\omega})$ is defined as

$$\hat{G}(e^{j\omega}) = \frac{\hat{\Phi}_{yu}(e^{j\omega})}{\hat{\Phi}_u(\omega)} \quad (2.9)$$

where $\hat{\Phi}_{yu}(e^{j\omega})$ is an estimate of the cross spectrum between the output $y(t)$ and the input $u(t)$, and where $\hat{\Phi}_u(\omega)$ is an estimate of the input spectrum. These estimates are obtained as

$$\begin{aligned} \hat{\Phi}_{yu}(e^{j\omega}) &= \frac{1}{N} \sum_{k=1}^{N-1} W_\gamma(\omega - \zeta_k) Y(e^{j\zeta_k}) U(e^{-j\zeta_k}) \\ \hat{\Phi}_u(\omega) &= \frac{1}{N} \sum_{k=1}^{N-1} W_\gamma(\omega - \zeta_k) |U(e^{j\zeta_k})|^2 \end{aligned}$$

where $W_\gamma(\omega)$ is a real valued window function, and where $U(e^{j\zeta_k})$ and $Y(e^{j\zeta_k})$ are the discrete Fourier transforms of the input and output signal respectively. For a bounded signal $u(t)$, $t = 0, 1, \dots, N-1$, the discrete Fourier transform is defined as

$$U(e^{j\omega_k}) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} u(t)e^{-j\omega_k t} \quad \omega_k = \frac{2\pi k}{N} \quad k = 0, 1, \dots, N-1$$

To analyse the properties of the spectral estimate $\hat{G}(e^{j\omega})$, let the window $W_\gamma(\omega)$ be written as

$$W_\gamma(\omega) = \gamma W(\gamma\omega)$$

where $W(\zeta)$ is the normalized window, defining the basic shape of the window (see [105]), and where $\gamma \in \mathbb{R}$ is a scaling parameter, describing the window width. A large γ yields a narrow window. Define

$$C := \int_{-\infty}^{\infty} |\zeta|^3 W(\zeta) d\zeta \quad \tilde{W} := 2\pi \int_{-\infty}^{\infty} W^2(\zeta) d\zeta$$

Under some regularity conditions there holds in the limit as $N \rightarrow \infty$, $\gamma \rightarrow \infty$, and $\gamma/N \rightarrow 0$ (note the doubly asymptotic nature of these conditions !), that the estimated transfer function is normally distributed, see [20]. The bias and variance are asymptotically in both N and γ given as

$$G_o(e^{j\omega}) - \mathbb{E}[\hat{G}(e^{j\omega})] = \frac{1}{\gamma^2} \left[G_o''(e^{j\omega}) + G_o'(e^{j\omega}) \frac{\Phi_u'(\omega)}{\Phi_u(\omega)} \right] + \mathcal{O}\left(\frac{C}{\gamma^3}\right) + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \quad (2.10)$$

$$\mathbb{E} \left[|\hat{G}(e^{j\omega}) - \mathbb{E}[\hat{G}(e^{j\omega})]|^2 \right] = \frac{\gamma}{N} \tilde{W} \frac{\Phi_v(\omega)}{\Phi_u(\omega)} + \mathcal{o}\left(\frac{\gamma}{N}\right) \quad (2.11)$$

$\begin{matrix} \gamma, N \rightarrow \infty \\ \gamma/N \rightarrow 0 \end{matrix}$

where prime denotes differentiation with respect to frequency, $\mathcal{O}(x)$ is a term that tends to zero at the same rate as x , and $\mathcal{o}(x)$ is a term that tends to zero faster than x . Under the same asymptotic conditions the estimated noise spectrum

$$\hat{\Phi}_v(\omega) = \hat{\Phi}_y(\omega) - \hat{\Phi}_{yu}(e^{j\omega}) \hat{\Phi}_u(\omega) \hat{\Phi}_{uy}(e^{j\omega})$$

also tends to a normal distribution, and the estimated transfer function and estimated noise spectrum become independent. Unfortunately, the unknown true spectrum of the noise $\Phi_v(\omega)$ cannot be removed by combining both distributions. However, for a rectangular window the estimated noise spectrum tends to a χ^2 distribution², and a confidence interval for $\hat{G}(e^{j\omega})$ can be obtained that does not contain the true spectrum of the noise, see [20, Chapter 6].

²The connection with the above mentioned normal distribution follows from the fact that a χ^2 distribution converges to a normal one as the number of degrees of freedom tends to infinity

In Bayard [8] a confidence interval for a spectral estimate is obtained by averaging over a number of empirical transfer function estimates. A periodic input signal is applied to the system, and an empirical transfer function estimate (ETFE, a special case of (2.9), see [106]) is made over each period of the input signal. The average over these ETFE's provides the final estimate, and a confidence interval for this estimate is specified. However, to obtain this confidence interval, it is assumed that the noise is normally distributed, that the noise filter is known, and that the steady-state situation is reached before experimental data is taken. An error bound on the whole unit circle is obtained by interpolating the estimate, and using prior information on the smoothness of the transfer function of the system. The results of [8] are however nonasymptotic.

For the error bounds obtained by Zhu [192, 193] using spectral analysis techniques, we refer to the discussion of his results in Section 2.2.

In conclusion, no adequate expressions for the bias are available, and the unknown true spectrum of the noise only can be removed from the asymptotic distribution of the transfer function estimate when a rectangular window is used, while other windows will perform better in general. Priestley [147, page 535] states that *confidence intervals constructed without due consideration of the bias are likely to be quite unreliable in regions where the true spectrum has peaks or troughs*. The bias terms can be considerable even for large values of N . The bias certainly can be reduced by tapering and prewhitening, where parametric identification techniques could in turn be used to obtain an estimate of the prewhitening filter, see Brillinger [20, Chapter 5]. However, a quantification of the bias error only becomes more involved, and it becomes difficult to distinguish between noise and undermodelling.

For a more general discussion on the merits of bias errors (bias error not neglectable with respect to the variance error), and the demerits of the use of prior information to bound these errors, we refer again to Section 3.8 and Section 3.11 respectively.

2.4 Parametric identification with stochastic embedding of noise and undermodelling

The approach to use stochastic embedding also for the error due to undermodelling has been introduced and elaborated by Goodwin and co-workers, see [59, 60, 55, 57, 56, 54], and is discussed in detail in the thesis of Ninness [130]. The argumentation given in [54, 130] to use stochastic embedding for the undermodelling error is discussed in Section 3.4. The idea is to represent the class of functions from which the undermodelling error is likely to come, by a parameterized probability density function, so as to capture the on-average behaviour of the usually very complex unmodelled dynamics.

Both the noise as well as the undermodelling error are assumed to be realizations of zero mean stochastic processes. To be more precise, the noise is represented by a sequence of filtered independent identically distributed random variables with zero

mean, and the system is divided into a part that will be identified, and a part that is not modelled

$$G_o(e^{j\omega}) = G(e^{j\omega}, \theta_o) + G_\Delta(e^{j\omega})$$

where $G_\Delta(e^{j\omega})$ is represented by a stochastic process $g_\Delta(k)$

$$G_\Delta(z) = \sum_{k=0}^{\infty} g_\Delta(k)z^{-k}$$

with decaying variance and zero mean

$$\mathbb{E}[G_\Delta(e^{j\omega})] = 0$$

The distributions of the noise and undermodelling error are assumed to be known, up to a small number of free parameters which characterize the specific shape of the distribution. For normal distributions these parameters specify the covariance matrix of the noise sequence, and the covariance matrix of the impulse response of the undermodelling term, respectively. For example, the stochastic process characterizing the undermodelling is a sequence of independent normally distributed zero mean random variables with variance

$$\mathbb{E}[g_\Delta^2(k)] = \alpha\lambda^k \quad \lambda < 1 \quad (2.12)$$

This should be interpreted as follows. Since it is impossible to know exactly how the impulse response decays, it is assumed to decay "on the average" as

$$\frac{1}{2m+1} \sum_{t=k-m}^{k+m} g_\Delta^2(t) \approx \alpha\lambda^k \quad (2.13)$$

Furthermore it is assumed that the undermodelling can be approximated sufficiently close by a finite impulse response (FIR) model of a certain order $L \leq N$ (the choice of L is studied in [130]). Also the model structure $G(e^{j\omega}, \theta)$ must have a fixed denominator, i.e. only the numerator is parameterized. Examples of such model structures are FIR models, Laguerre models (see e.g. [180]), and wavelets (see e.g. [118, 25, 130]). Approximate expressions are available for ARX models.

A least squares estimate of the model parameters is made. Subsequently, the free parameters of the distribution of the noise and undermodelling error are simultaneously estimated from the prediction error sequence, using a maximum likelihood estimate. This results in a probabilistic description of the error in the least squares estimate. This confidence interval is a sum of the confidence interval for the error due to the noise, and the confidence interval for the error due to undermodelling. Hence, separate and explicit expressions are obtained for both error terms.

In [54] it is mentioned that simulations show that the precise form of the structural assumption on the undermodelling does not appear to be essential. Also it is reported

that simulation studies show the resulting confidence regions to be highly informative and discriminatory.

The approach of stochastic embedding of the unmodelled dynamics is still quite new, and we have some questions and remarks regarding this approach.

1. The simulation examples concern highly idealized situations. The simulation examples only consider normally distributed white noise, low order systems (not higher than second order with time delay), and smooth undermodelling errors.
2. Similarly, a lot of the analysis concerns highly idealized situations.
3. The actual way of stochastic embedding which is to be used for the unmodelled dynamics appears to be a difficult point. In [54] it is stated that simulations show that the choice does not seem to be essential. This statement appears to be based on the idealized simulation examples as described above. It remains questionable whether reliable results will be obtained when the undermodelling is less well-behaved.

The suggested procedure for stochastic embedding of the unmodelled dynamics (2.12) yields an error due to undermodelling $G_{\Delta}(e^{j\omega})$ which has a constant variance (and thus a constant upper bound) over frequency. This certainly will not be the case in general, and is due to the white noise embedding of the undermodelling impulse response. In [130] a choice for embedding is given that adds a frequency dependence to the undermodelling term, but this choice still seems to be far too simple to capture realistic undermodelling errors.

4. The uncertainty due to using the estimated values of the free parameters characterizing the distributions of the noise and undermodelling is not taken into account. That is, a certainty equivalence principle is used in specifying the confidence intervals: the estimated values of the parameters characterizing the distributions are used as the true values.
5. The form of the distributions of the noise and unmodelled dynamics have to be chosen a priori by the user.
6. Severe difficulties appear to arise with regard to the actual distribution of the transfer function estimate, and with regard to the maximum likelihood estimate of the parameters characterizing the distributions, when the noise or undermodelling is not normally distributed. It should be possible to regain normality asymptotically in the number of data points for the noise contribution, but this appears not to be possible for the undermodelling contribution.
7. What happens if the noise is colored and/or the undermodelling error is not smooth: is it still possible to use a small number of parameters to characterize the distributions or is more extensive modelling required after all, and is it still

possible for the maximum likelihood estimator to effectively distinguish between noise and undermodelling ?

As mentioned before, the approach of stochastic embedding of the unmodelled dynamics is still quite new, and should be given time for development. Finally, although stochastic embedding of the undermodelling error appears to be a valid and useful approach, some basic objections can be made, see the discussion of Section 3.4.

2.5 Parameter set identification with bounded time domain errors

The literature on parameter set identification is quite extensive by now. The work dates back to [186, 160, 159], and basic contributions can be found in [9, 168, 132, 122, 45]. For overviews on this topic we refer to [37, 184, 125, 4]. We will only give a very short discussion of this field, to sketch the main ideas and mechanisms.

First of all, parameter set identification techniques yield a parameter set which is guaranteed to contain the true parameters provided that the prior information is correct. The system $G_o(q) = G(q, \theta_o) + G_\Delta(q)$ is represented as

$$\begin{aligned} y(t) &= G(q, \theta_o)u(t) + w(t) \\ w(t) &= G_\Delta(q)u(t) + v(t) \end{aligned}$$

where $G(q, \theta_o)$ represents the part of the system that will be identified, $G_\Delta(q)$ represents the unmodelled dynamics, and $v(t)$ the noise. We will call $w(t)$ the error sequence or error process. It should be emphasized that the noise $v(t)$ need not be stochastic or uncorrelated with the input $u(t)$. The approach also can be applied to nonlinear systems, see e.g. [183, 89, 123, 28], and time varying systems, see e.g. [134, 144]. For clarity of presentation we will only discuss the basic approach for the linear time invariant case, but the reader should be aware of the fact that the field is much larger.

The most common embedding now is that the error sequence $w(t) = G_\Delta(q)u(t) + v(t)$ is bounded in ℓ_∞

$$|G_\Delta(q)u(t) + v(t)| \leq B \quad \forall \quad t = 0, 1, \dots, N-1 \quad (2.14)$$

The basic motivations for using this assumption is that for a small number of data points N stochastic assumptions on the noise may not be justifiable, that the undermodelling error can be accounted for explicitly, and that nonasymptotic hard error bounds can be obtained.

Assumption (2.14) may be split up in bounds on the noise $v(t)$, the input signal $u(t)$ and the undermodelling $G_\Delta(q)$. Several possibilities to obtain a hard bound on $G_\Delta(q)u(t)$ from bounds on $u(t)$ and $G_\Delta(q)$ are e.g. discussed in [182, 187, 94, 92]. Although $u(t)$ is known for $t = 0, 1, \dots, N-1$ these bounds are bound to be rather conservative.

Subsequently, upper bounds are constructed on the set of parameter vectors θ that is consistent with the information (2.14), i.e. the set of all parameters such that

$$|y(t) - G(q, \theta)u(t)| \leq B \quad \forall \quad t = 0, 1, \dots, N-1 \quad (2.15)$$

Clearly, the approach is conceptually simple and attractive, needs only minimal assumptions on the error process $w(t)$, and leads to nonasymptotic hard error bounds.

If $G(q, \theta)$ is linear in θ then the feasible set of parameters (2.15) is a convex polytope where the number of vertices increases with N . For ARMAX models the set of feasible parameters possibly is non-convex, see [133], and for more general models the set of feasible parameters even cannot be guaranteed to be connected. When the feasible set of parameters is a convex polytope, exact bounds can be calculated using linear programming or recursively, see e.g. [126, 143], but this is computationally quite demanding and leads to very complex feasible parameter sets. To solve the latter problem algorithms have been developed that outer-bound the feasible set of parameters. Given the exact feasible parameter set, optimal outer-bounding algorithms give the smallest possible outer-bound in some norm (discussed below), see e.g. [21, 124, 123], but these algorithms again are computationally very costly. Thus, one often has to resort to simpler methods that do not first compute the exact bounds, but provide some suboptimal outer-bound. A very popular example of such an algorithm is Fogel and Huang [45]. However, these algorithms potentially have poor performance since the bounds may become loose through having to be approximated. For the algorithm of [45] this is illustrated in [130].

Several norms are used to outer-bound the parameter set. Different norms yield different geometrical shapes. Ellipsoidal bounds result when using the weighted ℓ_2 norm [45, 29, 133, 124, 24, 182], and boxes are obtained with the ℓ_∞ norm [122, 123]. The ℓ_1 norm is also frequently used [116, 28, 67, 21].

In [95, 96] identification schemes are presented that ensure a monotone reduction of the parameter errors in the number of data points, in spite of the presence of a bounded nonparametric frequency domain uncertainty $G_\Delta(e^{j\omega})$, by discriminating between desirable and undesirable excitation situations. However, no parameter set estimate is provided, and only the noise-free case is addressed.

Procedures to transform a parameter set to the frequency domain are given in [187] (potentially very conservative) and [182]. In [68] a procedure is presented that yields a set of transfer functions that is guaranteed to contain the true transfer function $G_o(e^{j\omega}) = G(e^{j\omega}, \theta_o) + G_\Delta(e^{j\omega})$ directly (not via an intermediate set of parameters) from (2.14). Frequency domain uncertainty descriptions are required by H_∞ robust control design methods, see e.g. [39]. In [99, 100] the problem of robust control design directly for ellipsoidal parameter uncertainty descriptions is addressed.

As is pointed out in [45], parameter set identification may still be applied if the magnitude of the error sequence $w(t)$ is not nicely bounded. Consider for example the situation where the error process $w(t)$ has a normal distribution (noise and input

are normally distributed). Then the error sequence is not bounded, but it is possible to specify bounds for the error sequence that will hold with any desired probability. The parameter set which is obtained will then hold with this level of probability. Note however that this probability (the probability of the event that all elements of the error sequence $w(t)$ obey the specified bound simultaneously) will be very small, unless the specified bound is very large. Moreover, the distribution of the error process need be known to be able to specify this probability. Note that the error process $w(t)$ consists of both a noise contribution, and a contribution due to the response of the unmodelled dynamics to the input signal applied.

In [45, 177] it is shown that under certain conditions the outer bounding parameter set converges to θ_o if the number of data points N goes to infinity, provided that the error sequence $w(t)$ is arbitrarily close to the specified bound B at a sufficient number of time instants, without exceeding this bound. However, in many practical applications it will be quite impossible to meet this requirement, so that the parameter set will not converge to θ_o regardless of the amount of data that is used. In [117] it is noted that the data even may be of little value if the prior information B is too large. Compare this to the results (2.2) (2.4) for the stochastic noise setting ! As noted in [70], this shows the demerit of the assumption (2.14). On the one hand, the parameter set will be unnecessary large when the bounds are chosen conservatively. On the other hand, when the bounds on the error sequence are chosen too optimistic, the resulting parameter set easily can become invalid.

An approach to attenuate the problem of outliers has been proposed by [183]. A further analysis is given in [148]. The idea is based on the fact that the parameter set may (not necessary !) become empty when the bound on the error sequence is too small. Hence outliers can be removed by discarding the smallest number of measurements that leaves the parameter set nonempty. However, it can by no means be guaranteed that all outliers are removed, so that the true parameter vector certainly may not be in the resulting parameter set. Thus, the vulnerability to outliers, due to the "all or nothing" weighting of observations, remains to be a severe problem.

Other assumptions than (2.14) have been proposed. In Fogel [44] the energy of the noise is assumed to be bounded. In Hakvoort *et al.* [70] the situation is considered where $G_{\Delta}(q) = 0$ and where the sample correlation of the noise $v(t)$ with the input $u(t)$ is bounded. It is shown that the parameter set converges to the true parameter θ_o , also when the specified bound is conservative. However, when undermodelling is present the convergence of the parameter set to θ_o again is lost. In the survey paper of Combettes [26] the use of fuzzy sets to represent the prior information is discussed. A stochastic noise setting is discussed by Kosut *et al.* [94]. The noise is assumed to be filtered white noise, with a known bound on its spectrum, and a known bound on $|G_{\Delta}(e^{j\omega})|$. It is shown that the parameter set estimate contains the true parameter vector θ_o with probability one in the limit as the number of data points N goes to infinity. Again, one should compare this result to the results (2.2) (2.4).

Clearly, in practice the specified bounds on the error sequence will be a balance between two conflicting objectives. Conservative bounds are required to account for possible occasional outliers, in order to guarantee that the resulting parameter set contains the true parameter vector θ_o . On the other hand, tight bounds are necessary to obtain a reasonably sized parameter set. Therefore, in practical applications the specified bounds on the error sequence will be a compromise between the (typically unknown) probability that they are valid, and the desire to obtain a reasonably sized parameter set.

In Ninness [130] it is shown that parameter set estimation techniques using the unknown but bounded noise framework (2.14) result in error bounds which are sensitive to the prior information on the true system and on the bounds on the noise, regardless of how much data is observed. In this context it should be noted that the purpose of obtaining an error bound often is to design robust control systems which by definition should be insensitive to the accuracy of the prior knowledge. This contradiction manifests itself as overly conservative bounds, since the a priori information must be chosen conservative in order to guarantee its correctness.

To conclude, the resulting error bounds are sensitive to the prior information, which conflicts with the aims of robust control design. Moreover, when the prior bounds on the error process are too large, the feasible set of parameters will not converge to θ_o , regardless of the amount of data that is used. Thus, small error bounds cannot be obtained with cautious prior information, while such bounds are required for high performance robust control design. The experimental data even may hardly improve upon the prior information. On the other hand, when the error bounds are too small, the resulting feasible set of parameters can easily be invalid. Finally, and this is indeed of considerable importance in view of the above, how should we obtain the prior information on $v(t)$, $G_\Delta(q)$? As mentioned in [182] this is a nontrivial problem, but no systematic procedures exist to obtain this prior information sufficiently accurate and reliable. Even the probability that the prior information is correct is typically not known, apart maybe from some vague intuitive notion.

All this results in unnecessary large and possibly incorrect error bounds, without an adequate notion of the probability of correctness. Additionally, the prior bound on $G_\Delta(q)$ is not improved from the data, and will directly return in the transfer function error bound, since $G_o(q) = G(q, \theta_o) + G_\Delta(q)$.

Using a stochastic noise description to obtain (asymptotic) probability levels on the prior information is not the correct way to solve the above problems. When accepting a stochastic description of the noise, it is more logical and probably better (i.e. leading to tighter and more reliable error bounds) to maintain this description till confidence intervals on the estimated model itself are obtained. For instance, a Bayesian framework can deal with probabilistic a priori information, and an a posteriori probability is established, see e.g. [59, 60]. In [130] it is even shown that, in the case of linear regression form model structures, parameter set estimation (2.15) is just a

special case of a Bayesian framework.

Finally, as mentioned before, a basic motivation for using assumption (2.14) is that for a small number of data points N stochastic assumptions on the noise may not be justifiable. Recently the motivation often has been solely that hard error bounds are obtained, regardless of the question whether stochastic assumptions on the noise are perhaps feasible. If larger data sets are available, one certainly should be able to do better, as then stochastic noise assumptions often can be justified.

For a more general discussion on the suitability of the unknown but bounded approach (2.14) we refer to Section 3.4 and Section 3.11. We also refer to the discussion on the intended use of the model in Section 3.7.

2.6 Transfer function set identification with bounded frequency domain errors

In this line of research an upper bound on the difference between the model and the system in the frequency domain is obtained, using the following three steps

1. Find an estimate $\hat{G}(e^{j\omega_k})$ and an upper bound $\epsilon(\omega_k)$ on $|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})|$ for a finite set of frequency points $\omega_k \in \Omega \subseteq \{2\pi k/N, k = 0, 1, \dots, N-1\}$.
2. Construct an estimate which is defined for all $\omega \in [0, 2\pi)$ from the estimate $\hat{G}(e^{j\omega_k})$, possibly while using $\epsilon(\omega_k)$ as a quality measure relative to the estimate \hat{G} at ω_k . This is achieved by interpolating the estimate $\hat{G}(e^{j\omega_k})$ c.q. by fitting a parametric model to this estimate.
3. Establish an error bound for all $\omega \in [0, 2\pi)$ for the estimate arising from the second step, by using $\epsilon(\omega_k)$ and some smoothness assumption on the transfer function of the system.

All three steps are addressed in [97, 98], but the focus is on the first two. In [172, 169, 170] the first two steps are considered only. In these papers an upper bound on the discrete Fourier transform of the noise is used as a priori information in the first step. However, the literature regarding this field of research heavily focuses on the latter two steps, thus taking the bound $\epsilon(\omega_k)$ as a priori information. The latter two steps are investigated in [139, 74, 75, 76, 78, 141, 142, 115] and [63, 62, 65, 2, 22, 79, 130].

Since these methods provide H_∞ error bounds, or hard (worst case) magnitude bounds on the frequency domain identification error, this approach has come to be known as "identification in H_∞ ".

In LaMaire *et al.* [97, 98] the empirical transfer function estimate (ETFE, see [106]) is used to obtain an estimate $\hat{G}(e^{j\omega_k})$ of the system $G_o(e^{j\omega_k})$ for a finite set a frequency

points $\omega_k \in \Omega$

$$\hat{G}(e^{j\omega_k}) = \frac{Y(e^{j\omega_k})}{U(e^{j\omega_k})} \quad \omega_k \in \Omega$$

where $U(e^{j\omega_k})$ and $Y(e^{j\omega_k})$ denote the N point discrete Fourier transform (DFT) of $u(t)$ and $y(t)$ respectively, and where

$$\Omega = \left\{ \frac{2\pi k}{N}, k = 0, 1, \dots, N-1 \mid U(e^{j\omega_k}) \neq 0 \right\}$$

Subsequently, it is shown that (see also [106])

$$Y(e^{j\omega_k}) = G_o(e^{j\omega_k})U(e^{j\omega_k}) + R(e^{j\omega_k}) + V(e^{j\omega_k})$$

where $R(e^{j\omega_k})$ is an error term due to the unknown past of the input signal, and where $V(e^{j\omega_k})$ denotes the DFT of the noise $v(t)$. An upper bound on the error term $R(e^{j\omega_k})$ is derived using an upper bound on the impulse response of the system. Together with an upper bound on the DFT of the noise, $|V(e^{j\omega_k})| \leq \bar{V}(\omega_k)$, a bound $\epsilon(\omega_k)$ such that $|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \leq \epsilon(\omega_k)$ now is obtained for the finite set of frequency points $\omega_k \in \Omega$ (say N_f points). Next a transfer function model $A(e^{j\omega}, \theta)/B(e^{j\omega}, \theta)$ is fitted to the ETFE by minimizing the weighted equation error criterion

$$\sum_{k=1}^{N_f-1} |\hat{G}(e^{j\omega_k})A(e^{j\omega_k}, \theta) - B(e^{j\omega_k}, \theta)|^2 |W(\omega_k)|^2$$

Finally, a bound on $|G_o(e^{j\omega}) - A(e^{j\omega}, \hat{\theta})/B(e^{j\omega}, \hat{\theta})|$ for all $\omega \in [0, 2\pi)$ is obtained by using a bound on the first derivative with respect to frequency of the relative error. This latter bound is obtained from a set of priors which will be rather intractable in practice (see [97] for the details). The error bound obtained in this third step therefore will be of little practical significance.

In Van den Boom [172, 169, 170] only the first two steps are addressed. The procedure employed to accomplish the first step is similar to the one proposed in [97, 98], as discussed above. In the second step a parametric model is fitted to the estimate $\hat{G}(e^{j\omega_k})$ in such a way that the maximum error over the finite set of frequency points $\omega_k \in \Omega$ is minimized. An error bound for all $\omega \in [0, 2\pi)$ is not pursued.

As mentioned before, the literature regarding the "identification in H_∞ " problem as formulated at the beginning of this section heavily focuses on the latter two steps. This appears to be due to the fact that no good procedures exist to perform the first step. The use of an upper bound on the DFT of the noise as in [97, 98, 172, 169, 170] provides no real solution, since this just shifts the problem to obtaining such an upper bound.

In the remainder of this section we will address these latter two steps. That is, we will consider the problem of obtaining an estimate with an error bound for all

$\omega \in [0, 2\pi)$ from an estimate with an error bound which is defined for a finite number of frequency points ω_k only. More specifically, we will discuss the problem formulation as conceived by Parker and Bitmead [139]. This problem has been stated in canonical form by Helmicki *et al.* [74], as follows. Given

- 1 an estimate $\hat{G}(e^{j\omega_k})$ of the system $G_o(e^{j\omega_k})$ at a finite number of equidistantly spaced frequency points $\omega_k = 2\pi k/N$, $k = 0, 1, \dots, N-1$,
- 2 an upper bound ϵ such that $|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \leq \epsilon$,
- 3 values \tilde{M} and $\tilde{\rho}$ such that $1 < \tilde{\rho} < \varrho$ and $\sup_{|z|=\tilde{\rho}} |G_o(z)| < \tilde{M}$, where ϱ is such that $G_o(z)$ is analytic for $|z| < \varrho$ (i.e. ϱ is the absolute value of the pole of the system which is nearest to the unit circle),

the aim is to find a model in H_∞ and an error bound for all frequencies $\omega \in [0, 2\pi)$. All papers after [74] which consider the latter two steps of the problem as formulated in the beginning of this section, accept this problem formulation and work with these assumptions.

In Parker and Bitmead [139] harmonic estimation by means of a Kalman filter (contains the DFT as a special case) is used to obtain an estimate $\hat{G}(e^{j\omega_k})$ of $G_o(e^{j\omega_k})$ at a finite number of frequency points. A finite impulse response (FIR) model $\hat{G}_n(e^{j\omega})$ is used to interpolate between these frequency points. An error bound for all $\omega \in [0, 2\pi)$ is established as (see [130, page 74])

$$\|G_o(e^{j\omega}) - \hat{G}_n(e^{j\omega})\|_\infty \leq \frac{2\tilde{M}\tilde{\rho}}{\tilde{\rho}-1}\tilde{\rho}^{-n} + (1 + \frac{2}{\pi}\ln(n))\epsilon$$

where n is the model order. The first right hand side term of the error bound is the worst case error due to interpolation (corresponding to the noise free case), and the second term is the worst case error due to the noise (corresponding to the case where $G_o = 0$). Note that the error bound is divergent with the model order n .

Another linear algorithm is proposed by Gu and Khargonekar [63]. The N point inverse DFT is used to obtain an estimate of the impulse response of the system

$$\hat{g}(k) = \frac{1}{N} \sum_{\ell=0}^{N-1} \hat{G}(e^{j\omega_k}) e^{-j\omega_k \ell} \quad k = 0, \dots, N-1$$

Subsequently a transfer function estimate is defined by weighting the sequence $\hat{g}(k)$

$$\hat{G}_n(z) = \sum_{k=0}^{n-1} W(k)\hat{g}(k)z^k$$

where $W(k)$ is a weighting sequence. Using a triangular window $W(k) = 1 - |k|/n$ this results in the bound

$$\|G_o(z) - \hat{G}_n(z)\|_\infty \leq \frac{2\tilde{M}\tilde{\rho}^{-N+1}}{\tilde{\rho}-1} + \frac{\tilde{M}\tilde{\rho}(1-\tilde{\rho}^{-n})}{n(\tilde{\rho}-1)^2} + \left(\frac{1}{\pi}\ln(n) + C\right)\epsilon$$

where C is a constant. The constant C is known to exist, but is otherwise unknown. Based on a large number of computer calculations the value of C is claimed to be $C \approx 0.7526$ for $5n \leq N \leq 5000$. Note that this linear algorithm again yields an error bound which slowly diverges with the model order n . Ninness [130] shows that this divergence actually can occur for the true error $\|G_o(z) - \hat{G}_n(z)\|_\infty$. Partington [142] even shows that there does not exist a robustly convergent linear algorithm for the specific identification problem formulated above (robustly convergent means that the error bound converges to zero as n goes to infinity and ϵ goes to zero). However, there do exist linear algorithms which do not exhibit a divergence of the error with the model order n provided the model order is sensibly regularized, see [130, Theorem 3.1] for an example.

The divergence of the error bound with the model order n with a linear algorithm observed by [139, 142] motivated the design of various nonlinear algorithms that avoid this divergence. A two-stage method was introduced by Helmicki *et al.* [74, 75, 78], and elaborated by Gu *et al.* [63, 62, 65]. We will follow the unified description of these methods as presented in [62, 65]. The methods presented in [141, 142] also fit into this description. In these two-stage methods it is tried to keep the (weighted) H_∞ error small by using an intermediate high order L_∞ model and Nehari approximation, obtaining a finite impulse response (FIR) model. The first stage of the two-stage method is to smooth the estimates $\hat{G}(e^{j\omega_k})$. This is done using the N point inverse DFT

$$\hat{g}(k) = \frac{1}{N} \sum_{\ell=0}^{N-1} \hat{G}(e^{j\omega_k}) e^{-j\omega_k \ell} \quad k = 0, \dots, N-1$$

and subsequently defining the transfer function estimate

$$\tilde{G}_n(z) = \sum_{k=-n+1}^{n-1} W(k) \tilde{g}(k) z^k$$

where $\tilde{g}(k)$ denotes the periodic extension of $\hat{g}(k)$, and where $W(k)$ is a weighting sequence. The smoothed estimate $\tilde{G}_n(z)$ is however not analytic (not causal), so that a second stage is needed. The second stage can be seen as an H_∞ approximation of an L_∞ function

$$\hat{G}_n(z) = \arg \min_{G_n(z) \in H_\infty} \|\tilde{G}_n(z) - G_n(z)\|_\infty$$

which is known as the Nehari problem (see [1, 188]). This is a classical problem and good algorithms are available to compute a solution.

The performance of the algorithm largely depends on $W(k)$. In [62] general conditions on the window functions are given which ensure that the resulting algorithm is robustly convergent, and several windows are proposed. For the window which gave

the best results, the worst case identification error can be bounded as

$$\|G_o(z) - \hat{G}_n(z)\|_\infty \leq \frac{2\tilde{M}}{\tilde{\rho} - 1} (\tilde{\rho}^{-N+1} + \tilde{\rho}^{-N+n-m+1} + 2\tilde{\rho}^{-2m}) + \frac{2(n+m)}{n-m}\epsilon \quad (2.16)$$

for $m < n < N$, where m is a design parameter in the specification of the window. The first right hand side term of the error bound is the worst case error due to approximation (the error corresponding to the noise free case), and the second term is the worst case error due to the noise (the error corresponding to the case where $G_o = 0$). By choosing the design parameter m a trade-off between these two terms can be made. Hence, the nonlinear identification algorithm results in a worst case noise error which does not diverge with n , and in a worst case approximation error which decays exponentially with n . However, the worst case noise error does not converge when the number of frequency domain data points N increases, and is bounded from below by 2ϵ . Compare this to the results (2.6) (2.7) for the stochastic noise setting ! As a result, the error bound (2.16) is sensitive to the prior information ϵ . That is, a cautious choice for ϵ inevitably leads to a large error bound, whereas conservative choices for \tilde{M} and $\tilde{\rho}$ can be attenuated by choosing a high order model. Additionally, only prior information ϵ which is constant over frequency is considered, and that the error bound also is not frequency specific. By reformulating the problem to address a weighted version of the system this drawback can be attenuated, but the prior information \tilde{M} and $\tilde{\rho}$ then also should be reformulated to address the weighted system. Finally, note that the error bound is completely determined by the prior information. That is, the experimental data does not influence the error bound.

A method where non-uniformly spaced frequency response estimates can be used is given by Akçay *et al.* [2].

A novel approach is presented by Chen *et al.* [22]. First it is validated whether the a priori information (given by \tilde{M} , $\tilde{\rho}$ and ϵ) and the frequency response data $\hat{G}(e^{j\omega_k})$ are consistent. Using Nevanlinna-Pick interpolation (see [46]) this problem can be posed as a constrained non-differentiable convex optimization problem. If consistency is established, Nevanlinna-Pick interpolation is used to derive a magnitude bound on the identification error. It is claimed that a frequency specific error bound can be obtained.

A perspective on identification in H_∞ which differs from the unified description of [62] is given in [115].

A key problem in all the above algorithms is the assumption that a bound on $|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})|$ or $|V(e^{j\omega_k})|$ is known, for a finite set of frequency points ω_k . In [77] sinewave experiments (actually measuring the frequency response $\hat{G}(e^{j\omega_k})$ in a finite number of frequency points ω_k) are proposed to obtain a bound on $|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})|$ for a finite set of frequency points ω_k , using a time domain bound on the amplitude of the noise. However, this just shifts the problem to the time domain bound on the amplitude of

the noise, and we refer to Section 2.5 for the drawbacks on such an approach. Moreover, when using an unknown but bounded noise description, ϵ appears to be bounded from below, regardless of the amount of data that is used. That is, ϵ is bounded from below by the difference between the prior time domain bound on the noise and the actual maximum absolute value attained by the noise, or by the prior frequency domain bound on the noise, in combination with the maximum allowable amplitude of the input signal. When using a stochastic description for the noise, a confidence interval on the DFT of the noise or on the ETFE can be established asymptotically in the number of data points. Moreover, a confidence interval on the ETFE which converges to zero can be constructed (see Chapter 5). However, when accepting a stochastic noise description, far better results can be obtained by maintaining this description and derive soft error bounds (see Chapter 5), which is also more logical than to drop the stochastic noise description halfway.

Similarly, how to obtain the prior information \tilde{M} and $\tilde{\rho}$? Correct and not overly conservative choices clearly are of considerable importance. However, this is largely an open issue. Helmicki *et al.* [79] state that in practical applications such prior information typically is obtained through some "engineering leap of faith". Obviously, this is not really the way we want to proceed when providing (hard) error bounds.

In Ninness [130] a detailed analysis is made of the problem of identification from frequency response data with deterministic noise, as formulated by [139, 74] (see page 22) and the existing algorithms to solve the problem are evaluated. The conclusion of [130] is that the existing algorithms are of very limited and questionable utility. This point of view is based on the following two points of criticism.

- The assumed format of a priori information on the true system (\tilde{M} and $\tilde{\rho}$) leads to extremely conservative bounds on the undermodelling (c.q. interpolation, approximation) error. It appears to be possible to attenuate this problem by choosing a format M and ρ to bound the impulse response of the system directly, i.e. such that $|g_o(k)| \leq M\rho^{-k}$.
- The "identification in H_∞ " algorithms are compared to conventional system identification techniques by means of a simulation example with uniformly distributed white noise. Using only one twentieth of the data used by the H_∞ techniques, the conventional identification scheme arrives at a much more accurate model.

A further point of criticism raised in [130] is that the "identification in H_∞ " algorithms are focussed on a pathological disturbance sequence which is responsible for the divergence of the error bound with the model order n in the linear algorithms. It is argued that the set of disturbances for which this divergence actually occurs is very small. For further details we refer to the illuminating analysis given in [130].

The observation of [130] that the "identification in H_∞ " algorithms lead to inaccurate models can be explained as follows. In (almost) all of the methods discussed

above there is a strong connection between the identification of a model and the quantification of its uncertainty. That is, only identification methods and model structures are selected for which hard error bounds can be derived. This seems to exclude many methods and model structures that could be useful but are rather intractable when it comes to deriving error bounds. In the methods discussed above, this specifically manifests itself in the use of the FIR model structure (which in general does not lead to the most accurate models) and the fact that no identification criterion is used at all in establishing the models (the models do not arise as a result of the minimization of a prespecified criterion function).

Summarizing, the existing "identification in H_∞ " algorithms lead to awkward and inefficient identification schemes, which result in bad nominal models. The resulting error bounds are in general not frequency specific, and are sensitive to the prior information ϵ , while no adequate procedures exist in the unknown but bounded noise framework to obtain this information sufficiently accurate. There is no convergence of the worst case noise error with the number of frequency domain data points N for the existing methods (bounded from below by 2ϵ). Thus, small error bounds cannot be obtained with cautious prior information, while such bounds are required for high performance robust control design. Additionally, the acquisition of the prior information \tilde{M} and $\tilde{\rho}$ is left to the "engineering leap of faith".

Hence, both the resulting error bounds as well as the estimated models appear to be deficient, and are not suitable for robust control design. In spite of the large body of literature addressing the "identification in H_∞ " problem, no successful application of the results has been reported as yet. Finally, when the noise is indeed noisy (stochastic) it appears to be possible to obtain more accurate models and smaller error bounds

For a more general discussion on the suitability of the unknown but bounded noise approach we refer to Section 3.4 and Section 3.11. We also refer to the discussion on the intended use of the model in Section 3.7.

2.7 Model validation

Often it will be costly or dangerous to just test a model with respect to its intended use. Instead, one has to develop confidence in the model in other ways, prior to using it. This is the task of model validation.

Basically, a model is validated by answering two questions: is the model credible, and does it agree with the observed data? A model is credible when it is in agreement with experience and physical knowledge. Since this depends on the application at hand, model validation theory tends to address only the last question. Hence, as stated in Hjalmarsson [83], the role of model validation is to examine prior information, assumptions and proposed candidate models and determine whether these priors and models are inconsistent with available experimental data. If so, the priors or the model

are falsified. If a set of priors or a model is not falsified, this should be interpreted as that there is nothing in the observed data that precludes the possibility that this could be a set of priors or model that is consistent with the true system. This does not mean that they are consistent; additional data may disqualify the set of priors or the model. In other words, a set of priors or a model is validated when there is nothing in the available experimental data that is in (apparent) conflict with the set of priors or the model. As noted in Wahlberg and Ljung [182], *it are in fact the priors that must be scrutinized*, since the posterior model is a direct mathematical consequence of the assumptions, the prior information and the data. This includes the answering of questions like: does the noise indeed appear to be worst case c.q. stochastic and uncorrelated with the input signal, does the system really appear to be linear and time invariant, and so on.

The model validation phase is indeed a very important one in the process of identification, and plays a prominent role when considering the application of uncertainty descriptions in a robust control design scheme. In Wahlberg and Ljung [182] this is illustrated by a possible conversation between a model builder and a control design engineer.

Control design engineer: "I see here that you have assumed that the disturbances are random. I don't believe in that nonsense. Disturbances can be very deterministic. Then the model discrepancy could be a lot worse."

Model builder: "Well, all I can say is that there is no evidence in the particular data set we worked with that we need to be so pessimistic."

It should be noted that model validation uses no assumptions at all, it just checks whether a set of assumptions can stand up against the available data.

For the classical approach to this subject we refer to the accounts given in [43, 58, 20, 147, 131, 106, 87, 166]. The main validation tools are to test how well the model is able to reproduce new data, and residual analysis. A recent extension to this approach is given in [83], where correlation tests are given for the practical situation where the limit model ($G(q, \theta^*)$, see (2.3) and (2.1)) is not available. Recent discussions on the subject can be found in [107, 182, 112].

A new approach to model validation was introduced by Smith and Doyle [162], and elaborated in [163, 129, 161, 164]. The problem addressed is the following. Given a multivariable stable linear time invariant model G and an experimental datum (u, y) , does there exist a noise datum v and a linear time invariant error model Δ with $\|v\|_2 \leq \alpha$ and $\|\Delta\|_\infty \leq \beta$, such that

$$y = LFT(G, \Delta) \begin{bmatrix} v \\ u \end{bmatrix}$$

where LFT stands for linear fractional transformation. Algorithmic issues have not yet been clarified to the point where the techniques are readily available. However,

when Δ is a full complex block necessary conditions can be calculated. Solutions to more general problems are subject of ongoing research.

Alternatively, Poolla *et al.* [146] provide conditions under which the data can have been generated by

$$y(t) = G(y, u, \theta) + \Delta(y, u) + v(t)$$

for some linear, norm-bounded but possibly time varying operator Δ and some norm-bounded sequence $v(t)$.

One should note the relation of these model validation problems with parameter set estimation techniques (see Section 2.5). Using parameter set estimation techniques the model would be validated when the feasible parameter set is not empty. The model validation problems formulated above however (potentially) do allow for more general uncertainty structures (structured Δ).

Although these model validation formulations are conceptually attractive, it should be noted that solving these problems appears to be computationally extremely demanding.

In [163] it is argued that these model validation schemes may be developed into methods for model uncertainty quantification. However, as was also noted in [163], as yet the problem formulation is highly nonunique; the residuals can be attributed either entirely to noise or entirely to model errors. The use of structured uncertainty descriptions (structured Δ) will only aggravate this identifiability problem.

2.8 Identification of nominal models for control design

When the intended use of the identified model is control design, then it is beneficial to consider the problems of identification and control design simultaneously. This actually is the domain of adaptive control (see e.g. [6]). However, this problem is rather involved, and the interaction between the control design part and the identification part often is not well understood. Therefore, iterative schemes have emerged recently, proposing alternate turns of identification and control design, see e.g. [49, 11, 12, 101, 149, 69, 157, 155, 191] and [5, 102, 103, 140, 48, 61, 175]. The idea is to first identify a model of the system which is suitable for high performance control design, and subsequently to use this model to design the controller. However, it turns out that the high performance controller must be known to be able to identify a model which is optimally suitable for high performance control design, so that an iterative scheme arises. To fix ideas, we follow the chain of arguments used by Schrama [157, 155, 156]. Suppose that we would like to find a controller which minimizes

$$\|T(G_o, C)\|$$

where G_o is the system, C the controller, and $T(G_o, C)$ some performance objective (on the closed loop consisting of G_o and C), and where $\|\cdot\|$ denotes a (semi) norm.

The triangle inequality shows that

$$\|T(G_o, C)\| \leq \|T(\hat{G}, C)\| + \|T(G_o, C) - T(\hat{G}, C)\|$$

where \hat{G} is the identified model. For a fixed controller C , the second term on the right hand side can be made small by identifying a model \hat{G} such that the achieved performance $T(G_o, C)$ and the designed performance $T(\hat{G}, C)$ are close. This implies that the identification should be performed in closed loop, since the achieved performance is measured under closed loop. For a fixed model \hat{G} , the first term on the right hand side can be decreased by control design, solving $\min_C \|T(\hat{G}, C)\|$. This will of course change the second term, so that a new model should be identified. This obviously leads to an iterative procedure. Clearly, both terms are minimized separately, and no convergence to the global optimum can be guaranteed.

Using such an iterative scheme one can aim at high performance directly, or one can gradually increase the performance when the model becomes more accurate, to find the highest level of performance that can be reached. In the literature on this topic, it is stressed that the criterion used for identification should be the same as the criterion used for control design. In order to (approximately) achieve this, closed loop experiments and prefiltering of the experimental data are employed. Furthermore, cautious (slow) controller enhancement should be used to secure robustness of the controller against model errors. Clearly, the availability of a bound on the model error would be advantageous in this context, to determine the allowable step size in the performance enhancement.

Chapter 3

Problem formulation and thesis contribution

Summary

The problem that is addressed in this thesis is formulated. The properties, implications and suitability of different approaches towards solving this problem are discussed, leading to specific choices with regard to the embedding (prior information and assumptions) of the problem, and a set of basic conditions and technical requirements for a high quality solution. We will touch upon the basics of system identification, since they are disregarded in most current model error estimation techniques. Finally, an overview of the contributions of this thesis towards solving the problem is given.

3.1 Problem formulation

Referring to Section 1.2, we will assume that the system, and the measurement data that is obtained from this system, (approximately) allow a description

$$y(t) = G_o(q)u(t) + v(t) \quad (3.1)$$

with $y(t)$ the output signal, $u(t) \in \ell_\infty$ a bounded deterministic input signal, $v(t)$ an additive noise, being a bounded deterministic signal or a realization of a stochastic process, and $G_o \in \mathbb{RH}_\infty$ the transfer function of the system, being linear, time invariant, finite dimensional, causal, and exponentially stable.

The problem addressed in this thesis is to find the set of all systems which are not highly unlikely to have generated a given set of experimental data points. To be more precise, the problem considered in this thesis can be formulated as follows.

Find the set of all causal linear time invariant finite dimensional exponentially stable systems for which there is indeed evidence in the experimental data itself that these systems could have generated these measured sequences of input and output data $y(t), u(t), t = 0, \dots, N - 1$.

It should be noted that this set is smaller than the set of all systems which are not in apparent conflict with the data. That is, we do not want to find the set of all unfalsified (not in apparent conflict with the data) models, but only the subset of all powerful (plausible) unfalsified models. One should note that this requirement should be regarded as a basic principle of scientific modelling, as was discussed in Section 1.1.

Basically, a set of models (as formulated above) is pursued. However, (almost) all control design schemes are model based, i.e. the design proceeds from one specific model of the system (the nominal model). Additionally, (almost) all robust control design schemes can only deal with a quantification of the model uncertainty as a specific type of uncertainty, in a specific structure, using specific norms. There is no robust control design scheme which operates directly on a (arbitrary) set of models, or provides the controller which achieves the best (in some sense) performance over such a set. Thus, since this thesis is directed towards the case where the intended use of the set of models as formulated above is to serve as a basis for robust control design, we will aim at describing such a set as centered around some model of the system. Furthermore, we are especially interested in the quantification of the size of the set as a weighted ball in H_∞ . Thus, we can phrase the problem more specifically as follows.

Given a set of N measured pairs of input and output data $y(t), u(t), t = 0, \dots, N - 1$, find a model $G(e^{j\omega}) \in \mathbb{R}H_\infty$ and a bound $\gamma(\omega) \in \mathbb{R}$ such that there is evidence in the measurement data that $|G_o(e^{j\omega}) - G(e^{j\omega})| \leq \gamma(\omega)$ for all $\omega \in [0, 2\pi)$, where $\gamma(\omega)$ should not be larger than strictly necessary to reflect the uncertainty about G_o as present in the measurement data.

In a probabilistic framework "evidence" could be made more precise by providing the probability that $|G_o(e^{j\omega}) - G(e^{j\omega})| \leq \gamma(\omega)$.

Note that the formulation as given above is quite general in that we do not restrict the model G to be an element of a certain subset of $\mathbb{R}H_\infty$ a priori, nor require G_o to be an element of a certain subset of $\mathbb{R}H_\infty$. The issue of prior information and additional assumptions is left open. All this should follow as a consequence of the emphasized key element of the formulation above.

Besides the magnitude bound as given above, more general uncertainty representations will also be pursued (frequency domain ellipsoidal error bounds, structured descriptions in the form of mixed parametric-nonparametric uncertainty, coprime factor uncertainty descriptions, and error bounds on the time domain output of the system in response to an arbitrary input signal).

The solution to the problem as formulated above is by no means unique. Different models may concentrate on different aspects of the system, while a description of all

aspects only will lead to an inaccurate model, as was discussed in Section 1.1 (bias-variance tradeoff). Therefore, in addition to the problem as formulated above, we require that *it should be possible to tune both the model as well as the error bound to robust control design specifications*, and the user should be provided with tools to achieve this (e.g. by input design, model order selection, and the use of weighting functions).

Furthermore, although the system is assumed to be linear and time invariant, one should realize that this is only an approximation of reality. Therefore we also have to consider the effects of small deviations from this assumption. That is, when addressing the problem as formulated above, *we also have to consider the situation where only the dominant part of the system is linear and time invariant*. The assumption that the behaviour of the system approximately can be described as being linear and time invariant often is indeed a valid assumption, and a restriction to this situation is also necessary to successfully apply the H_∞ robust control design techniques of [39] and related methods. In general, this situation will e.g. arise when the system is controlled to remain in the neighbourhood of a certain operating condition, which is a very common setting.

3.2 The basic difficulty

As was discussed in Section 1.2, further restrictions on the system, the input signal and the noise are necessary to be able to properly address the problem as formulated in the previous section. These restrictions, in the form of additional assumptions and prior information, have a decisive influence on the resulting uncertainty quantification, as was discussed in Section 2.1. Thus, the actual choice of the format of these restrictions determines whether the problem as formulated in the previous section is indeed solved, considering the requirement that the error bound should not be larger than strictly necessary to reflect the uncertainty about G_o as present in the data.

We are of opinion that the basic difficulty with model uncertainty quantification concerns this choice of additional assumptions and prior information, and is not in the first place a mathematical one (as yet). It certainly may be tempting to just abstract the problem to a clean mathematical one, but this should not be done without having taken due account of the consequences. To be more specific, the basic difficulty to establish meaningful error bounds is to answer the following four questions.

1. What kind of assumptions and prior information do we need to establish error bounds ?
2. What kind of assumptions about the true system and the noise (i.e. about nature) appear to be appropriate (realistic) in a given situation ?
3. Which a priori information is likely to be available, or can be readily obtained with sufficient accuracy ?

4. What are the consequences of different sets of assumptions and prior information ?

Obviously, the answers to the first three questions will be conflicting: virtually any useful assumption will be incorrect or at least questionable, or the necessary matching prior information will not be available (intractable). However, we do need some assumptions to be able to arrive at a quantification of model uncertainty (see Section 1.2). In other words, *we will have to use assumptions and prior information which are known to be approximations of reality*. Therefore we have to find assumptions and priors that will lead to adequate models and error bounds, even if they are mildly violated, which leads to the last question. Additionally, the fourth question concerns the following: will the resulting models and error bounds be useful for the intended purpose ? Just finding an error bound is not enough. The error bound may be an overbound that is far too large except for a number of pathological cases, or it may be nearly impossible to efficiently improve the error bound when it is too large for the intended use. Hence the resulting error bound should be reasonably tight and flexible, and the user must be provided with clear insight and adequate tools to reduce and shape the error bound.

Clearly, there is no general correct answer to the questions mentioned above. However, in a given situation, one set of assumptions and prior information certainly can lead to error bounds which are more likely to give a realistic indication of the true error than other sets. In the following we will spend a lot of effort to shed some light on these questions, because the choices made are of decisive importance to obtain accurate models and error bounds.

3.3 Main criticism on existing techniques

From the survey of existing techniques for model error estimation in Chapter 2, we conclude that these techniques (except Zhu [192, 193], Goodwin *et al.* [54] and Ninness [130]) suffer from one severe basic deficiency: *too much of the problem is shifted to the a priori information*. This is not solving but hiding the problem. No systematic procedures exist to obtain this a priori information sufficiently accurate and reliable, while it is of major importance, see Chapter 2. Thus, we are back at the experienced guess and ad hoc procedures. Subsequently we close our eyes, hope for the best, and bravely take the "engineering leap of faith". Certainly this may work, and also cannot be avoided completely, but definitely misses the issue of model error estimation and robust control design, i.e. to guarantee stability and performance. Moreover, we even have to regard the use of such prior information as conflicting with the principle of parsimony (see Section 1.1); a principle which is fundamental to scientific modelling. In contrast, *the bounds on noise and undermodelling contributions should arise from the experimental data in a systematic way, not from a priori information, and should not be larger than strictly necessary to account for the observed data*. Clearly, we cannot do without any prior information at all, see Section 1.2. Then again, there is nothing

against the use of prior information as such, as long as it is realistic and is allowed to be approximative.

Additionally, after the quite athletic "leap of faith" with regard to the validity of the prior information, highly complex and computationally demanding techniques are applied to ensure correct (hard) error bounds, given that the prior information is correct. This is not a well-balanced approach.

Furthermore, we have the following points of criticism.

- Conventional system identification techniques [20, 106] are not able to provide explicit error bounds under realistic assumptions, that is, when the undermodelling error (bias) is not neglectable (with respect to the variance error).
- There is no adequate quantification of the undermodelling error (except maybe [54, 130]). Apart from [54, 130], the undermodelling error either just follows directly from the a priori information and depends not at all or only hardly on the measured data, or is neglected, or is unknown.
- In the stochastic framework the inaccuracy of estimated variances usually is not taken into account.
In the hard error bounding literature the noise levels are just assumed to be known a priori.
- The unknown but bounded noise (worst case) framework leads to error bounds which are sensitive to the a priori information (the bound on the noise). Cautious prior information results in error bounds which do not converge to zero and are not tight, regardless of the amount of data that is used. Still, one outlier may ruin the correctness of the error bound.
- The unknown but bounded (worst case) noise framework leads to error bounds which are hard (if the prior information is correct), but unnecessary large when the noise is not worst case.

The last two points indicate that the unknown but bounded (worst case) noise model is too crude a model of reality (allows for too much). Reliable and tight error bounds are not obtained. Thus, the error bounds are not suitable for high performance robust control design.

3.4 Deterministic versus stochastic noise

One of the main topics in uncertainty modelling is the question whether the noise should be modelled as a realization of a stochastic process, or as an unknown but bounded sequence ($|v(t)| \leq C$ for $t = 0, \dots, N - 1$ but otherwise unknown). First of all, it should be stressed that the answer to this question depends on the situation at hand. Also there is no such thing as a correct answer to this question. *Both*

settings are incomplete models of nature, having their specific merits and drawbacks. In the following we want to discuss these specific properties, and clarify the underlying assumptions about the nature of the noise. At the same time we will discuss whether these assumptions appear to be appropriate or not.

To start with, referring to [130], a general statement can be made as follows. If qualitative assumptions about the on-average properties of a disturbance realization can be made, then a stochastic framework seems appropriate, and quantitative properties can be estimated from the data. If such assumptions cannot be made then an unknown but bounded error approach may be appropriate since only a quantitative assumption is made.

The ability to make inferences about the on-average properties of a disturbance is closely related to the length of the available data records. For very short data records the unknown but bounded noise setting certainly can be the more appropriate one, see e.g. [88]. An adequate description of the noise as a realization of a stochastic process may exist, but will be hard to obtain. Qualitative stochastic assumptions on the noise may not be justifiable (assuming a normal distribution of the estimate on the basis of the central limit theorem certainly may be too optimistic), and it is almost impossible to estimate the qualitative properties of the noise from the data (no reliable estimate of the variance of the noise can be obtained). In contrast, for longer data records stochastic assumptions on the noise often are justifiable.

In this context it should be noted that this thesis is directed towards the situation where quite an accurate model of the system is required in order to enable the design of high performance robust controllers. Clearly, very short records of experimental data will in general not result in such models. Thus, we will not consider the case of very short data sequences, so that stochastic assumptions on the noise certainly may be justifiable.

When addressing the problem of providing a description of the modelling error from experimental data, the unknown but bounded noise setting allows the noise to be precisely worst case without stating the probability that this will occur. This has two important consequences. Firstly, the error bounds are determined by the worst case noise. Secondly, *the worst case noise is perfectly correlated with the input signal*, as was noted by Hjalmarsson [83]. This implies that it is assumed that nature is able to play exactly against the experimenter, no matter the input signal that is applied.¹ In a stochastic setting the probability of such an event would of course be zero. With regard to this situation Kosut [93] takes the point of view that nature is not malicious (conspiratorial), but indifferent (neutral). Hjalmarsson [83] argues that in the stochastic as well as the unknown but bounded setting a disturbance is considered to be something that is not related with the input signal. Hence, a disturbance which is correlated with the input is not, by definition, a disturbance but a part of the system

¹Even Murphy would be amazed.

dynamics. In this context, it should be noted that in the unknown but bounded noise setting the 'noise' (the error sequence $w(t)$ of Section 2.5) indeed contains a component which is used to capture the unmodelled dynamics. However, this error sequence is also used to capture signals which are not related to the input signal (compare (2.14) in Section 2.5). Thus the worst case analysis considers (at least partly) cases that by definition cannot occur, which explains the counter intuitive results that exist in this field [83], as e.g. error bounds which do not converge to zero regardless of the amount of data that is used, see Section 2.5 and Section 2.6. Note that following this line of reasoning it still can happen that the disturbance is almost worst case over a finite time interval, without being correlated with the input, but the probability of this event usually will be very small, except for very short data records.

Hjalmarsson [83] also shows that a consistent estimate of an exponentially stable linear time invariant infinite dimensional system *can* be obtained for unknown but bounded deterministic disturbances, provided that they are not correlated with the input signal. The key point is that mixing can be obtained by choosing an exponentially forgetting input signal. The estimate now behaves basically in the same way as in the stochastic disturbance case.

For some time the unknown but bounded noise setting has been considered (especially by the control theory community) as the more reliable setting, where the necessary prior information is more accessible in practice, because only an upper bound on the noise is assumed. This point of view has been seriously undermined in Goodwin *et al.* [54], and we will borrow from their argumentation. The unknown but bounded approach to noise modelling, and hence (see Section 2.1) the hard bounding approach to model error quantification, suffers from two key limitations. On the one hand, the unknown but bounded noise model is a very coarse model of physical reality, since within a compact domain the noise is allowed to be precisely worst case without stating whether this is likely or not. This is rather a pessimistic characterization of the noise, since it does not capture the fact that often the noise is on-average zero and uncorrelated with the input signal. On the other hand, the unknown but bounded noise model is far too precise a model of reality, since it is assumed that outliers will not occur. This usually is rather an optimistic point of view, and implies that the prior bounds will of necessity be very large to account for occasional but unlikely large values in the noise. In turn, this implies that the resultant hard error bounds will be unnecessary large. Still, one outlier may ruin the correctness of the error bounds. In contrast, a distribution with noncompact support is a model of reality in which the noise values are assumed to be on-average centered around some mean value without precluding the possibility of an occasional large deviation from this mean value. Of course, the resulting error bounds then become confidence regions rather than hard bounds. However it is argued that this is appropriate since prior assumptions can never be specified with absolute certainty. Indeed it is suggested that real world control problems are nearly always solved by aiming for high performance in the belief

that the set of pathological conditions associated with extreme bounds will rarely, if ever, occur. Therefore, control engineers always work with a tradeoff of uncertainty versus performance. Using a stochastic description of the disturbances precisely will enable such a tradeoff between achievable performance and the probability of failure in an accurate and well-defined way. Hard error bounds clearly fall short in this respect. That is, *confidence intervals with a high level of probability are precisely what is desired for high performance robust control design*. We want to discard the highly unlikely, in order to gain performance, but we want to guarantee robustness to everything that is not highly unlikely.

An argument which is often raised in favour of a deterministic description of the noise as unknown but bounded, is firstly that stochastic descriptions are too restrictive or require too much prior information to be specified, and secondly that the convergence in the central limit theorem to the asymptotic normal distribution, which is almost always assumed, is inaccurate for short data lengths. These arguments are heavily attacked in Ninness [130], showing that essentially only a qualitative assumption on the memory of the noise process (mixing condition) is required to obtain asymptotic normality, and that the convergence to normality can be quite fast. Furthermore, the resulting confidence intervals tend to be insensitive to precise quantitative knowledge on the noise process, and the sensitivity decreases linearly with the amount of observed data. In contrast, the existing bounded error estimation techniques deliver error bounds which are sensitive to the prior bound on the noise, see Section 2.5, Section 2.6 and [130].

We would like to add to the above argumentation that the stochastic paradigm for the noise has a long history of successful applications in a large variety of cases, and should not be abandoned as soon as hard bounds are desired or undermodelling errors are introduced, as is also noted in [130].

To conclude, undermodelling errors and noise errors certainly should be considered separately, since they differ in nature. The unknown but bounded (worst case) noise model is too crude and too hard a model of reality ("all or nothing" character). The resulting error bounds will be neither tight nor reliable, unless accurate quantitative information is available a priori, which in general will not at all be the case.

In contrast, the stochastic noise model is a very flexible model of reality. Essentially it is required only that the noise forgets its past (mixing condition). Moreover, given only the qualitative knowledge that the noise satisfies a mixing condition, reliable and tight error bounds can be obtained from the data. Additionally, the stochastic noise model precisely yields the required format of a posteriori information for high performance robust control design. That is, the resulting confidence intervals explicitly allow for a tradeoff between performance and failure probability.

3.5 Deterministic versus stochastic embedding of undermodelling

Goodwin, Gevers and Ninness [54, 130] also employ stochastic embedding for the error due to undermodelling, using the following argumentation.

”We contend that the stochastic embedding approach is a very appropriate one to choose because of the nature of undermodelling. That is, undermodelling typically arises because of physical manifestations that are too complicated to exactly describe. The best that can be hoped for is to capture the on-average properties of the undermodelling so that its most likely manifestations can be predicted. In this case, a probability density function is an appropriate choice for describing the undermodelling. Indeed we would argue that the common assumption of measurement noise existing and being modelled by a stochastic process is an equivalent injection of a probabilistic framework on an essentially deterministic underlying problem.”

Additionally, in Ninness [130, page 129] it is argued that, if it is possible to tell that undermodelling is not produced by a zero mean random variable then the model order is too low. Finally, [130] points towards the link with current thought in the more abstract area of complex nonlinear dynamical systems (chaos theory and fractals). In this theory, a trend has developed in which attempts at predicting orbits are abandoned in favour of measuring their average behaviours, see e.g. [119], since any tiny error in the estimate gives a radically large error in predicted orbits. For a further discussion on this interesting point we refer to [130].

Nevertheless, we are of opinion that stochastic embedding is less fit for the error due to undermodelling, based on the following line of reasoning. To fix ideas, we will concentrate on the use of finite impulse response (FIR) models and exponentially stable linear time invariant systems. The undermodelling error now is given by the tail of the impulse response of the system. This tail typically will be dominated by a relatively small (with respect to the significant length of the impulse response) number of exponentially weighted sinewaves. Indeed, the very tail of the impulse response often is dominated completely by only one exponentially decaying sinewave (representing the slowest dynamics), which is precisely the worst case when addressing the magnitude of the resulting transfer function error. Thus, the undermodelling error indeed can come close to having a worst case character. Moreover, given that the system is exponentially stable, a suitable hard upper bound on the tail of the impulse exists, and no outliers can occur. However, one should note that the undermodelling error is not really worst case as long as the impulse response contains a number of exponentially decaying sinewaves, instead of only one.

The situation appears to be similar to stochastic embedding of the noise. That is, if the noise is expected to contain a number of sinewaves then a stochastic description is not appropriate. In other words, noise essentially is a sequence that satisfies a mixing condition (decaying dependence with the distance between two samples), but the impulse response of a system clearly does not satisfy such a mixing condition. However, this does not exclude that a stochastic model could be adequate.

Stochastic embedding of the error due to undermodelling only provides an on-average (two norm) characterization of the undermodelling error, see (2.13), instead of a worst case (infinity norm) characterization. When the transfer function of the undermodelling error does not exhibit large peaks when evaluated over the unit circle (no poles close to the unit circle), the two norm and infinity norm are not much different. This may explain the good results obtained in the simulation examples given in [54, 130] since the transfer functions of the undermodelling errors in these examples are in fact smooth functions of frequency. However, the undermodelling transfer function typically is far from smooth !

Finally, note that traditionally, see e.g. [106, 20], the error due to undermodelling has been treated as a deterministic (unknown but bounded) error.

To conclude, stochastic embedding of the undermodelling error may be a valid approach, but some objections appear to be in place when the transfer function of the undermodelling error may contain sharp peaks. In this thesis we will employ a deterministic description. Practical applications, future developments and the possibilities to extend the different frameworks to more complex situations should decide. Also, stochastic and deterministic undermodelling descriptions both may have their own field of applicability.

3.6 A broader perspective: averaging and structural errors

As mentioned in Section 3.1, we want to consider the realistic situation where only the dominant part of the system is linear and time invariant. Thus, when the system is assumed to be linear and time invariant, we also have to consider the effects of mild violations of these assumptions. Furthermore, we should be able to capture deviations from the stochastic noise model that we will use.

In this context the choices in Section 3.4 and Section 3.5 certainly have a broader significance. These choices allow for an *appropriate distinction between averaging and structural errors*. Such a distinction should be made, since these errors differ in nature. The stochastic noise model should be interpreted as a model that is able to capture the averaging errors, whereas the deterministic framework for the undermodelling error should be interpreted as a means to account for structural errors, thus constructing a framework that should be able to account for many important deviations from a linear

time invariant system with stochastic noise.

3.7 Intended use of the model and the error bounds

When addressing the intended use of the model the central issue is again that the model will only be an approximation of the true system. This of course immediately raises the questions: which aspects of the system should be modelled, and to which level of accuracy ?

The answer to these questions of course depends on the intended use of the model. In Ljung [106, Chapter 12] some results are presented, distinguishing between simulation, prediction and control. However, identification for control design has attracted by far the most attention, see e.g. [49, 11, 12, 101, 149, 69, 157, 155, 191] and [5, 102, 103, 140, 48, 61, 175]. In this thesis we also focus on this application, but similar observations are expected to hold when the intended use of the model is analysis (simulation) or prediction.

The significance of the intended use of the model for identification has been generally recognized. Surprisingly, the implications of the intended use of the model and the uncertainty description are generally neglected in model uncertainty estimation procedures.

Identification for robust control is twofold (see [32, 34, 156]):

1. the identification of a model that is suited for control design, and
2. a relevant quantification of the model uncertainty.

A clear discussion of this topic is given by Schrama [156], and in the following we will quote his main points. The two prerequisites mentioned are complementary. On the one hand, the controller must anticipate the imperfections of the model, i.e. the controller must be robust. On the other hand, a bad model will not give rise to a good controller. Because the requirements of high performance and robustness are conflicting, the upper bound on the model error should be as tight as possible. A tight upper bound is however not sufficient to obtain a good controller: the error bound cannot be smaller than the true error (as given by the system and the model), and the latter thus limits the achievable performance. Hence, the requirement of high performance imposes limitations on the allowed shape and extent of the mismatch between the system and the model. Note that this is a conversion of the common starting point in robust control which states that the modelling error limits the achievable performance.

Now, high performance control usually implies that a small difference $|G_o - G|$ between the system G_o and the model G in a certain frequency region can result in a large difference between the controlled system and the controlled model, and vice versa. Thus, usually very small modelling errors are required in some frequency regions (typically around the desired closed loop bandwidth), whereas in other frequency

regions large errors are allowed. Hence, we must be able to accurately tune both the model as well as the shape of the upper bound on the model uncertainty to the robust performance requirements, in order to be able to guarantee a high level of performance.

In conclusion, important and generally neglected implications of the intended use of model uncertainty estimates are the following.

1. *A tight uncertainty description for an inaccurate model is rather useless in practice.*
2. *It should be possible to effectively tune both the model and the model uncertainty to the intended use of the model. In other words, it should be possible to accurately shape and reduce the error bound, and the user should be provided with tools to achieve this.*

Point 1 implies that sensible identification procedures should be used. As stated by [130], developing a theory that provides error bounds at all costs while neglecting the quality of the estimate is to "throw the baby out with the bath water". In view of point 2 above, an H_∞ error bound by no means is informative enough.

3.8 Model structure, identification criterion and under-modelling

Firstly, as also was noted in Section 3.7, the identification procedure should be of high quality. That is, the model structure should be as parsimonious as possible and yet be flexible enough to capture the dominant behaviour of the system as reflected in the data, and a relevant (with respect to the intended use of the model) criterion should be used and minimized to fit the model to the data. Clearly, this in general excludes the FIR model structure. Regarding the criterion, we prefer a least squares criterion. In [150, page 351] it is stated that it is the opinion of many mathematicians that *convergence relative to the norm in an L_2 space is particularly appropriate in the theory of approximation* because L_2 -convergence implies an overall, average measure of the remainder.

Secondly, apart from the model structure and the criterion, we have to select a model order. The selection of the model order is directly related to the concept of the most powerful unfalsified model (see Section 1.1). A lower order model is more powerful, and we should therefore select the lowest order model which is not in apparent conflict with the data. In other words, given the experimental data, we should only model what we can model reliably, and thus we should allow for undermodelling. In system identification this is specifically reflected by the tradeoff between bias and variance errors, by means of model order selection. If the bias is neglectable with respect to the variance this means that the model order is too high, and that the principle of parsimony is violated, as was discussed in Section 1.1. The model error

will be larger than necessary; a lower order model will be more accurate. Thus, the bias error should not be neglectable with respect to the variance error. This has since long been recognized in system identification; the best models are obtained when the bias and variance errors are approximately of the same size, see e.g. [106]. Therefore we state that *when undermodelling is allowed for, significantly better results can be obtained for both the model as well as the model error bounds*. Moreover, as has been argued in Section 1.1, the bias usually even cannot be neglected because the system will in general be far too complex to model exactly. Finally, note that both the bias and variance contributions should be known approximately to be able to explicitly perform this tradeoff.

3.9 Parametric techniques or spectral analysis

As a continuation of Section 3.7 and Section 3.8, emphasizing the necessity of a high quality estimate, in this section we will address the choice between spectral analysis and parametric identification methods. Ljung [105] has shown that, when compared to spectral analysis, the *convergence to zero of the mean square identification error for parametric methods can be faster*. Thus, parametric methods allow (asymptotically) for a smaller mean square error. This is due to the fact that specifically tailored model sets can be chosen when using parametric identification techniques. For a well-chosen model set the difference in mean square error certainly can be considerable. Therefore we will focus on parametric methods.

The above is in line with Priestley [147, pages 535,545-546], where it is noted that to obtain satisfactory (sufficiently small, with an adequate level of probability) confidence intervals using spectral analysis, typically requires large amounts of data, especially when the true spectrum contains sharp peaks and troughs.

3.10 Model validation

In the context of model validation the availability of reliable error bounds is desirable, since they directly attach a quality tag to the model. However, also the error bounds itself should be validated. Ljung [107] states that the prime validation tool is to test how well the model is able to reproduce new data. We will extend this, and state that *the prime validation tool is to test whether the combination of the model and the error bound indeed specifies an interval which contains the new data*. To check this, we should of course be able to specify such an interval, i.e. a bound on the difference between the disturbed output of the system and the output of the model.

To emphasize the value of such a bound, note the relation with the concept of powerful unfalsified models (see Section 1.1). That is, if the combination of a model and an error bound is unfalsified on a given set of data points but is not powerful, it is likely to be falsified over new data. Thus, we should have an adequate means

to detect whether or not the combination of a model and a model error bound is falsified (not able to capture the data), on the current as well as on a new set of data points. Especially when the assumptions of a linear time invariant system and stochastic noise are slightly violated, as will be the case in any practical application, it is of considerable importance to have a powerful means to check whether the error bounds are valid (unfalsified) in spite of these violations.

We conclude that *a reliable error bound for the disturbed output would be highly desirable*. This would provide a particularly strong validation test for both the model and the error bound, as well as the noise assumptions and the other priors used to build the model. That is, the combination of model, error bound and noise description should be able to capture new data.

3.11 Prior information and extrapolation

The idea of model error quantification is of course that the model can be safely used when due account is given to the possible errors. However, when quantifying model uncertainty from experimental data, the error bounds can account only for the behaviour of the system as significantly reflected in the observed data. Moreover, in this section we will argue that this is also *precisely what we should aim at*. That is, what is not in the experimental data should not be required to come out in the error bounds. As a consequence, *the error bounds cannot be expected to be valid for excitation or operating conditions of the system that differ considerably from the ones under which the experimental data was gathered*.

Incorporating knowledge about considerably different possible behaviours of the system via prior knowledge is certainly not the proper solution to this extrapolation problem. Firstly, how to obtain such prior knowledge sufficiently accurate? This is just the model error estimation problem that should be solved. Secondly, such prior information will almost completely determine the error bounds. The experimental data should only have a minor influence, exactly because the error bound should not be tuned to a specific experimental condition. This is in line with the observation of [117] for parameter set estimation, where it is noted that the data may be of little value if the prior bound on the noise is too large. Finally, when accounting for all conceivable behaviours of the system, the uncertainty may be that large that no robust controller can be designed which achieves a satisfactory level of robust performance.

In our opinion only the specific behaviour of the system under operating conditions should be addressed when estimating error bounds. Additionally, quantification of model uncertainty should be based on the observed data, not on prior information, as was also discussed in Section 3.3. The necessary prior information and assumptions should be minor, realistic, and allowed to be approximative. While this forbids extrapolation, it *does allow for high performance robust control*.

Again, this approach does not allow for extrapolation of the error bounds to excita-

tion or operating conditions of the system that differ considerably from the ones under which the experimental data was gathered, since the error bounds only reflect the uncertainty in the experimental data. This even holds for linear time invariant systems, since not all dynamics may have been sufficiently excited, but it holds in particular for nonlinear or time varying systems. Minor deficiencies in the model error bound, can lead to large deviations from expected behaviour when the behaviour is extrapolated from e.g. an open loop situation to a high performance closed loop situation. An example of such deviations is given in Section 5.10.1 for a slowly and periodically time varying system.

As a result, error bounds should be based on data that closely resembles the data arising in the intended use of the model. Thus, closed loop experiments and cautious controller enhancement (see [155] and Section 2.8) certainly appear to be called for. In this context, the error bounds should serve as a powerful means to indicate the amount of performance enhancement that can be safely realized, but should not be abused and used as the truth to conclude that a large performance enhancement can be achieved in one step. That is, the robustness properties of the new controller should not differ radically from the robustness properties of the old controller.

To conclude, the error bounds should account only for the behaviour of the system as significantly reflected in the experimental data. Accounting for other excitation or operating conditions via the prior information, yields error bounds that are not based on the data, that are not tight for the data at hand, and comes down to shifting the basic problem of model error estimation to establishing the prior information. Moreover, since the error bounds are not tight, the performance will be low, whereas high performance could be obtained when only the specific uncertainty in the data under operating conditions would be considered.

3.12 Requirements

From the previous sections it follows that the solution to the problem as formulated in Section 3.1 is required to have the following properties.

- Both the model and the error bounds should be suitable for their intended use.
- The procedure used to identify the model should be of high quality. That is, the model structure should be parsimonious but flexible, and we should allow for undermodelling.
- It should be possible to accurately shape and reduce the error bounds, and the user should be provided with tools to achieve this.
- An appropriate distinction should be made between averaging (noise) and structural (undermodelling) errors, since they differ in nature.

- The necessary prior information and assumptions should be minor, realistic and allowed to be approximative.
- The error bounds should arise from the experimental data in a systematic way, not from prior information, and they only should account for the behaviour of the system as reflected in the experimental data
- The system can be assumed to have a dominant linear time invariant behaviour, but not to be exactly linear and time invariant. Thus, when developing a model error estimation procedure for linear time invariant systems, it should be checked afterwards whether the procedure is robust to mild violations of this assumption. Similarly, the error bounds must be robust to mild violations of the noise assumptions.
- The error bounds should be reliable and tight. Thus, reliable and tight bounds should be established for both the error due to the noise (variance error) as well as the error due to undermodelling (bias error).

Based on these requirements, we came as yet to the following decisions.

- Use a stochastic noise model to account for averaging errors. That is, assume that the noise forgets its past (mixing condition). Both the qualitative (distribution) and quantitative (variance) information about the error due to the noise should arise from the data.
- Use a deterministic framework for the error due to undermodelling to account for structural errors. A bound on this error again should be inferred from the data.
- Let the model be specified by the user, or use a high quality identification procedure which can be tuned to the needs of a robust control design scheme. We pointed to the merits of parametric identification techniques (as opposed to spectral analysis) and a least squares criterion to identify the model.
- An error bound for the disturbed output of the system should be available for validation purposes.
- In view of the principle of parsimony, results which are not asymptotic in the model order should be pursued.

3.13 Overview of thesis contributions

3.13.1 Error bounds from bounded frequency domain noise

In this section we will provide an overview of Chapter 4 of this thesis, where a procedure for "identification in H_∞ " is presented. This chapter reports on a first approach to

solving the problem as formulated in Section 3.1, and therefore does not satisfy some of the major requirements of Section 3.12. Nevertheless, the procedure proposed possesses certain important merits when compared to existing methods for "identification in H_∞ " (see Section 2.6). An overview of the latest results is given in Section 3.13.2.

In Chapter 4 an upper bound on the discrete Fourier transform (DFT) of the noise is assumed to be known, together with an upper bound on the impulse response of the system. The upper bound on the DFT of the noise is not required to be constant over frequency. A procedure is presented to quantify the model uncertainty for *any prespecified nominal model*, directly from a sequence of measurement data from the system. This procedure consists of two steps.

In the first step, the empirical transfer function estimate (ETFE) is used to construct a nonparametric estimate of the transfer function in a finite number of frequency points, together with an upper bound on the error. This error bound is frequency dependent, which makes it more informative than a simple H_∞ bound. In order to obtain a tight error bound, a special input signal is proposed (partly periodic) which has advantages over classical sinewave experiments. Also, conditions are given to validate whether the prior information is consistent with the measurement data.

In the second step, the error bound for the ETFE is transformed to a bound which is available for all frequencies in the interval $[0, 2\pi)$. This is done by using a nominal model, which can be specified by the user. The error bound for the ETFE directly provides an error bound for this nominal model in a finite number of frequency points. The *latter error bound* now is interpolated to provide an error bound over a continuous frequency interval. The interpolation procedure does neither require the frequency points of the ETFE to be equidistantially distributed over the unit circle, nor the error bound on these frequency points to be constant over frequency. This paves the way for designing specific input signals in order to improve the estimate, and to tighten the bound.

A *frequency dependent* hard error bound results. This error bound can be split into three parts: one part due to the inherent uncertainty in the data (noise and unknown initial conditions), a second part due interpolation, and a third part due to imperfections in the prespecified nominal model (the difference between the model and the ETFE). These three components can be tuned almost independently, by appropriate experiment design or by choosing an alternative nominal model. Thus, when the error bound is too conservative in view of the robust control design specifications, we can decide whether new experiments should be performed, or the nominal model should be improved. Moreover, we can point out specifically which experiments should be performed, or how the nominal model should be improved, in order to satisfy the design requirements. Finally, since the error bound is frequency specific, it can be effectively reduced in specific frequency ranges.

To conclude, the error bound can be tailored effectively to the needs of a robust control design scheme. The procedure can be applied to multi-input multi-output

(MIMO) systems, and is extended to closed loop situations and unstable systems in Chapter 6. A simulation example is provided to illustrate the merits of the procedure.

3.13.2 A mixed averaging – worst case approach

Chapter 5 contains the latest results, and provides a well-balanced environment and a fairly complete set of tools for model uncertainty estimation from experimental data.

The main ideas behind the procedure are the following.

A probabilistic setting for the noise is chosen to account for averaging errors, whereas the effects of undermodelling are considered as being deterministic to account for structural errors. Both the error due to the noise, as well as the error due to undermodelling are estimated from the data. This constitutes the main deviation from existing literature. The framework of stochastic (averaging) noise in combination with unknown but bounded (worst case) undermodelling is exceptional in model uncertainty estimation, although this setting is quite common in the classical identification literature, see e.g. [106, 20]. To some extent this setting is also employed by Kosut *et al.* [94], Bayard [8] and Hakvoort *et al.* [70]. However, none of these authors actually estimate the undermodelling error from the data, and the error due to the noise is not [94] or only partially [8, 70] estimated from the data. The only contribution in the literature where both the noise and undermodelling errors are indeed inferred from the experimental data is the stochastic embedding approach of Goodwin *et al.* [54], which is elaborated in Ninness [130].

A parametric model and a weighted least squares criterion are employed to identify the system. System based orthonormal basis functions (see Appendix E) are used to construct the model set. These basis functions provide a powerful tool to model a system, and can be used to ensure that the model is suitable to serve as a basis for high performance robust control design.

In addition, a periodic input signal is employed to effectively distinguish between averaging (noise) and structural (undermodelling) errors. That is, repeating an experiment will give the same error due to undermodelling, whereas the realization of the noise error will be different.

Essentially, the only prior information that we need is an upper bound on the past values of the input signal. This bound need not at all be tight, and usually can be obtained from the actuator constraints. The main assumptions are that the system is linear time invariant, and that the noise satisfies a mixing condition (forgets its past).

The procedure now can be summarized as follows. An empirical transfer function estimate (ETF) is made over each period of the input signal. Subsequently, a parametric model is fitted to each of these nonparametric estimates, resulting in a set of estimated parameter vectors. Finally, this set of estimated parameter vectors directly provides a set of transfer function estimates. The errors due to the noise in

these estimates are asymptotically in the number of data points identically normally distributed and independent. The set of parametric estimates is used to provide a final estimate as the sample mean, together with an estimate of its variance as the sample variance. A closed form confidence interval for the error due to the noise now can be established. The error due to undermodelling is estimated as follows. First, using the same procedure, a confidence interval for the error due to the noise on the estimated parameters is established. Next, this confidence interval is used to estimate a lower bound on the decay rate of the parameters, resulting in an upper bound for the undermodelling error. In this way, a frequency specific error bound for the transfer function estimate is obtained. This error bound is valid asymptotically in the number of data points, and can be calculated for each frequency in the interval $[0, 2\pi)$.

Separate error bounds for the different sources of uncertainty (noise, undermodelling, unknown initial condition) are obtained. This enables an explicit bias-variance tradeoff by model order selection, and is also important when the error bound is too large in view of the robust control design specifications. The error sources that provide a major contribution to the error bound can be isolated, so that it is known how to effectively improve the error bound, by input design or by choosing an alternative basis generating system.

Using similar procedures, error bounds are established for the undisturbed output signal and on the DFT of the disturbed output signal. The former error bound is useful to assess the simulation accuracy of the model, and may be useful for robust predictive control or robust feedforward control. The error bound on the DFT of disturbed output is useful for validation of the error bounds, the model, the assumptions and the prior information, and may be useful for fault detection. Furthermore, a mixed parametric-nonparametric uncertainty description can be specified. The parametric and nonparametric uncertainty are both estimated from the data, in such a way that, when combined, they specify an error bound for the model. This appears to be new, and can be used in μ analysis or synthesis in the form of a structured real valued uncertainty, in combination with an unstructured complex valued uncertainty.

We have performed fairly extensive simulation and application studies to verify whether the mixed averaging-structural description of the error is indeed justifiable and leads to tight and reliable error bounds, also in cases where the assumptions of a linear time invariant system and stochastic noise are violated.

To assess the performance of the procedure in nonasymptotic situations, a Monte Carlo study has been carried out, showing that reliable and tight error bounds are indeed obtained.

To verify whether the procedure is robust to deviations from the assumptions (linear time invariant system, stochastic noise), several simulation studies are carried out. For the two nonlinear systems correct and tight error bounds are obtained. For a system with a slow periodic time variation, the error bounds are again quite adequate, as long as the estimates are reasonably uncorrelated, which can be easily checked. For a system

with fast random time variations, the error bound are no longer correct. Furthermore, the procedure has been applied successfully to a wind turbine system and a compact disc player in Chapter 8. Quite satisfactory error bounds are obtained, in spite of the presence of mild nonlinearities and periodic disturbances.

In Chapter 4 the model error estimation procedure is extended to multi-input multi-output (MIMO) systems.

In Chapter 6 the model error estimation procedure is extended to include closed loop situations and unstable systems.

Finally, in Chapter 7 an asymptotic analysis is provided for the frequency domain transfer function estimation procedure using system based orthonormal basis functions. That is, the statistical properties of the estimated parameters and the estimated transfer function are analyzed, asymptotically in the number of data points. When no weighting is used in the least squares parameter estimate, the estimates are shown to be consistent. That is, consistent estimates of the parameterized parts of the system are obtained. Explicit and transparent bias and variance expressions are established, which are not asymptotic in the model order. These expressions provide insight in the properties of the transfer function estimation method that we employ, and can be used to make effective choices for the basis functions and the input signal.

Chapter 4

Nonparametric frequency domain hard error bound

Summary

Identification of linear models in view of robust control design requires the identification of a control-relevant nominal model, and a quantification of model uncertainty. A procedure is presented to quantify the model uncertainty of any prespecified nominal model, from a sequence of measurement data of input and output signals from a plant. By employing a nonparametric empirical transfer function estimate (ETF_E), we are able to split the model uncertainty into three parts: the inherent uncertainty in the data due to data-imperfections, the unmodelled dynamics in the nominal model, and the uncertainty due to interpolation. A frequency-dependent hard error bound is constructed, and results are given for tightening the bound through appropriate input design. Additionally, conditions are given to validate whether the prior information is consistent with the measurement data. When the upper bound on the model uncertainty is too conservative, in view of the control design specifications, information is provided as to which additional experiments have to be performed in order to effectively improve the bound.

4.1 Introduction

Recently several approaches to the problem of model uncertainty quantification from experimental data have been presented, considering identification for robust control design. By far the most attention is paid to the construction of so-called hard error bounds. These can be divided into time and frequency domain techniques. The former are known as parameter set estimation techniques, see for example [45, 29, 132, 124,

182, 94, 187, 68], and the discussion of these techniques in Section 2.5. The frequency domain techniques often are referred to as "identification in H_∞ ", see for example [139, 75, 98, 141, 115, 62, 2, 22, 171, 79, 130], and the discussion of these techniques in Section 2.6. In [192, 194, 8, 54, 130] identification procedures are presented that provide probabilistic (soft) error bounds. A discussion of these techniques is given in Section 2.2, Section 2.3 and Section 2.4. Chapter 5 of this thesis also presents a procedure that provides soft error bounds.

In the references mentioned, there often is a strong connection between the identification of nominal models and the quantification of model uncertainty. This can be a serious drawback. Only identification methods for nominal models are selected for which (hard H_∞) error bounds can be derived. This seems to exclude many methods and model structures that could be useful but are rather intractable when it comes to deriving error bounds. When discussing the suitability of models as a basis for control system design, the availability of reliable error bounds certainly is important in order to obtain robust stability, and possibly also robust performance. However, it is the nominal model that is used as a basis for the design, and a bad nominal model will not give rise to a good controller. Thus, it is indeed quite a disadvantage when the procedure which is employed to identify the nominal model is not of high quality in view of control design. Moreover, an identification procedure which is not of high quality will give rise to unnecessary large error bounds, and seriously impairs the practical value of the error bounds obtained. Further discussions on this topic can be found in Section 3.7 and Section 3.8. As a result, the identification of nominal models, apart from the quantification of model uncertainty, is an important issue in identification for control design, see for example [11, 12, 69, 155, 157, 175, 5, 61, 140, 103, 48], and the concise overview given in Section 2.8.

We will address the problem of model uncertainty quantification from experimental data. In answer to the above observations, we will separate the quantification of model uncertainty from the identification of a nominal model. That is, in this chapter we will deal with the following problem. Given a *prespecified* nominal model G_{nom} for an unknown linear plant G_o , can we construct an error bound for

$$|G_o(e^{j\omega}) - G_{nom}(e^{j\omega})| \quad (4.1)$$

based on noise corrupted measurements from input and output samples of the plant? Note that the nominal model may be available from any (control-relevant) identification procedure. We will present a method which provides hard error bounds, and which can be considered to belong to the "identification in H_∞ " class of methods.

The problem is going to be tackled, through the construction of an intermediate data representation in the frequency domain, leading to the inequality

$$|G_o(e^{j\omega_k}) - G_{nom}(e^{j\omega_k})| \leq |G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| + |\hat{G}(e^{j\omega_k}) - G_{nom}(e^{j\omega_k})| \quad (4.2)$$

where $\hat{G}(e^{j\omega_k})$ is an -intermediate- representation of the measurement data in the frequency domain. This means that $\hat{G}(e^{j\omega_k})$ basically consists of a finite number of complex points on the unit circle, obtained from the discrete Fourier transformation (DFT) of the time domain data. The first term on the right hand side of (4.2) can be considered to reflect inherent uncertainty in the data, whereas the second term is related to the quality of the nominal model, e.g. determined by unmodelled dynamics. Having constructed a data representation $\hat{G}(e^{j\omega_k})$, the second term can be calculated exactly. Hence, to give an upper bound on the model error $|G_o(e^{j\omega_k}) - G_{nom}(e^{j\omega_k})|$, the problem is to construct an upper bound for the error $|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})|$. Note however that inequality (4.2) is only defined at the finite number of frequency points ω_k , while our aim is to bound the model error for all $\omega \in [0, 2\pi)$. The fact that the data does not contain information for frequencies $\omega \neq \omega_k$ gives rise to the uncertainty due to interpolation. The second problem therefore is to bound the model error for all $\omega \in [0, 2\pi)$ using the error bounds at ω_k . These two problems will be the main topics of this chapter.

We will call a bound on $|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})|$ or $|G_o(e^{j\omega_k}) - G_{nom}(e^{j\omega_k})|$ in a finite number of frequency points ω_k a discrete error bound, whereas a bound on $|G_o(e^{j\omega}) - G_{nom}(e^{j\omega})|$ for all $\omega \in [0, 2\pi)$ will be addressed as a continuous error bound. Similarly, we will call an estimate $\hat{G}(e^{j\omega_k})$ in a finite number of frequency points ω_k a discrete estimate.

An extensive survey of existing methods for "identification in H_∞ " can be found in Section 2.6. Summarizing, we find that almost all existing methods for "identification in H_∞ " (e.g. [75, 76, 78, 141, 142, 115, 63, 62]) only provide error bounds which are constant over frequency, and require the frequency points of the discrete estimate to be equidistantially spaced. Additionally, the required format of the prior information on the system (\tilde{M} and $\tilde{\rho}$, see Section 2.6) appears to lead to extremely conservative error bounds, and the prior information on the noise (ϵ , see Section 2.6) is required to be constant over frequency. Moreover, the resulting nominal models are not of fit for control design. These points constitute serious drawbacks in view of robust control design, as was discussed in Section 3.7. The nominal model and the error bounds cannot be tuned adequately to robust control design specifications, and the error bounds appear to be that conservative that high performance robust control is inhibited.

In the present contribution the drawbacks mentioned above are addressed and removed. To be more specific, the advantages of the procedure which will be presented in this chapter over the existing techniques for "identification in H_∞ " are the following. Firstly, the nominal model G_{nom} with respect to which the error bound is calculated can be specified directly by the user. That is, the quantification of model uncertainty is separated from the the identification of a nominal model, so that we are not restricted to (awkward) identification schemes for which error bounds can be derived. This is due to the fact that we will interpolate a bound on $|G_o(e^{j\omega_k}) - G_{nom}(e^{j\omega_k})|$ in

a finite number of frequency points ω_k in order to obtain a continuous bound, rather than trying to bound the error which could arise while using a certain technique to fit a model to the discrete estimate $\hat{G}(e^{j\omega_k})$, which is the usual procedure. Secondly, a frequency specific continuous error bound is obtained, which is more informative than a simple H_∞ bound. The discrete error bound is allowed to be frequency specific, and the frequency points ω_k are not required to be equidistantially distributed over the unit circle. This paves the way for designing specific input signals in order to improve the estimates, and tightening the bound in specific frequency ranges. As a result, the continuous error bound can be accurately tuned to robust control design specifications by input design. Finally, the prior information on the system (M and ρ , see Section 4.2) directly addresses the impulse response, and not the commonly used radius of a disc in the complex plane in which the system is analytic, and a bound on the magnitude of the system over this disc (\tilde{M} and $\tilde{\rho}$, see Section 2.6). The latter information appears to be very hard to obtain in practice, and is also shown to lead to extremely conservative error bounds in [130].

Thus, some of the major drawbacks of existing "identification in H_∞ " techniques as observed in Section 2.6 are removed. However, the essential drawbacks, as mentioned in Section 2.6, Section 3.4 and Section 3.3, attached to the hard prior information on the noise which is used, remain. In this chapter, this hard prior information takes the form of a hard prior bound on the DFT of the noise.

Related work has been published in [77, 98] where error bounds for $|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})|$ have been obtained at a finite number of frequency points. In [77] this has been done through sinewave excitation and actually measuring the frequency response in a finite number of points, and in [98] by employing the empirical transfer function estimate (ETFE, see [106]).

The vast majority of the existing methods for "identification in H_∞ " assume that a discrete estimate and error bound are given a priori, and only address the problem of finding a model in H_∞ together with a continuous error bound. In this area, the mainstream proceeds as follows, see e.g. [75, 141, 142, 65, 63, 62]. The model is obtained by applying the inverse DFT to the discrete estimate. It is tried to keep the H_∞ error small by using an intermediate high order L_∞ model and Nehari approximation, resulting in a finite impulse response (FIR) model. The continuous error bound which is obtained is constant over frequency, and the frequency points of the discrete estimate are required to be equidistantially spaced.

In [2] a method is presented which allows for non-uniformly spaced frequency response estimates. In [22] an approach is presented which uses Nevanlinna-Pick interpolation of the discrete estimate, and it is claimed that a frequency specific error bound can be obtained.

A more extensive discussion of existing methods for "identification in H_∞ " can be found in Section 2.6. For a further discussion on model error estimation techniques,

nominal models and error bounds we refer to Chapter 2 and Chapter 3.

The remainder of this chapter is organized as follows. In Section 4.3 the ETFE is used to obtain a nonparametric frequency domain estimate $\hat{G}(e^{j\omega_k})$, and a frequency specific error bound in a finite number of frequency points ω_k . In Section 4.4 a frequency dependent error bound which is valid for all $\omega \in [0, 2\pi)$ is constructed by interpolation of the discrete bound, using smoothness properties of the system. In Section 4.5 conditions are given to validate whether the prior information is consistent with the measurement data. In Section 4.6 it is shown how robust control design specifications can advocate new experiments in order to reduce the model uncertainty in specific frequency ranges. Finally, in Section 4.7 a simulation example is given to illustrate the merits of the procedure proposed.

The contents of this chapter (except Section 4.3.3 and Section 4.5) have been published as [35].

4.2 Assumptions and prior information

It is assumed that the plant, and the measurement data that is obtained from this plant, allow a description

$$y(t) = G_o(q)u(t) + v(t) \quad (4.3)$$

with $y(t)$ the output signal, $u(t)$ the input signal, $v(t)$ an additive output noise, q^{-1} the delay operator, and G_o a proper transfer function that is time invariant and exponentially stable. The transfer function can be written in its Laurent expansion around $z = \infty$, as

$$G_o(z) = \sum_{k=0}^{\infty} g_o(k)z^{-k} \quad (4.4)$$

with $g_o(k)$ the impulse response of the plant. Throughout the paper we will consider discrete time intervals for input and output signals denoted by $T^N := \mathbf{Z} \cap [0, N - 1]$, $T_{N_s}^N := \mathbf{Z} \cap [N_s, N + N_s - 1]$ with N and N_s appropriate integers. We will denote

$$\bar{u} := \sup_{t \in T^N + N_s} |u(t)| \quad (4.5)$$

For a signal $x(t)$, defined on T^N , we will denote the N point Discrete Fourier Transform (DFT) and its inverse by

$$X\left(\frac{2\pi k}{N}\right) := \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x(t)e^{-j\frac{2\pi k}{N}t} \quad \text{for } k \in T^N \quad (4.6)$$

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X\left(\frac{2\pi k}{N}\right)e^{j\frac{2\pi k}{N}t} \quad \text{for } t \in T^N \quad (4.7)$$

When a signal $x(t)$ is defined on the interval $T_{N_s}^N$, $N_s > 0$, then we will denote the N point DFT of a shifted version of the signal x , shifted over N_s time instants, by

$$X^s\left(\frac{2\pi k}{N}\right) := \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x(t + N_s) e^{-j\frac{2\pi k}{N}t} \quad \text{for } k \in T^N \quad (4.8)$$

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X^s\left(\frac{2\pi k}{N}\right) e^{j\frac{2\pi k}{N}(t - N_s)} \quad \text{for } t \in T_{N_s}^N \quad (4.9)$$

Note that this reflects the N point DFT of a signal, of which the first N_s time instants are discarded. Throughout this paper we will adopt a number of additional assumptions on the system and the generated data.

Assumption 4.2.1 *There exists a finite*

- i. \bar{u}^p , such that $|u(t)| \leq \bar{u}^p$ for $t < 0$;
- ii. pair of reals $\{M, \rho\} \in \mathbb{R}$, $\rho > 1$, such that $|g_o(k)| \leq M\rho^{-k}$ for $k \in \mathbb{N}$;
- iii. $\bar{V}^s\left(\frac{2\pi k}{N}\right)$, such that $|V^s\left(\frac{2\pi k}{N}\right)| \leq \bar{V}^s\left(\frac{2\pi k}{N}\right)$ for $k \in T^N$.

4.3 Discrete error bound

4.3.1 Motivation

The motivation to consider the ETFE is that we want $\hat{G}(e^{j\omega_k})$ to be an intermediate data representation in the frequency domain. The ETFE is the quotient of the DFT of the output signal and the DFT of the input signal. In discrete Fourier transforming a signal no information is lost or added, the mapping from time to frequency domain is one to one. Thus, the ETFE can indeed be regarded as a representation of the data in the frequency domain.

The motivation to look at input design is that the ETFE for an arbitrary input signal is in general not satisfactory. We will try to improve the quality of the frequency domain data by input design.

4.3.2 SISO systems

A nonparametric frequency domain discrete upper bound on the additive error for the ETFE of a SISO (single input, single output) system will be presented in this section. Errors due to unknown past inputs (unknown initial conditions of the system) and additive noise on the output are taken into account. We will use a partly periodic input signal for excitation, and we will discard the first part of the signals in the estimation.

Definition 4.3.1 A partly periodic signal x is a signal having the first part equal to the last part: $x = [x_1 \ x_2 \ x_1]$.

The length of x_1 will be denoted by N_s . Only the part $[x_2 \ x_1]$ will be used in the identification. The length of this part will be denoted by N . The total length of the signal x now is $N_s + N$. We will show that the value of N_s influences the error due to unknown past inputs in the estimate. Note that the largest possible value of N_s is N .

Theorem 4.3.2 Consider a SISO system, satisfying the assumptions stated in Section 4.2. Using a partly periodic input signal, $N_s \in T^{N+1}$, and the estimate

$$\hat{G}^s(\frac{2\pi\ell}{N}) := \frac{Y^s(\frac{2\pi\ell}{N})}{U^s(\frac{2\pi\ell}{N})} \quad \text{for } \ell \in \{q \in T^N \mid U^s(\frac{2\pi q}{N}) \neq 0\} \quad (4.10)$$

the following error bound is satisfied

$$|G_o(\frac{2\pi\ell}{N}) - \hat{G}^s(\frac{2\pi\ell}{N})| \leq \alpha(\frac{2\pi\ell}{N})$$

with

$$\alpha(\frac{2\pi\ell}{N}) = \frac{1}{\sqrt{N}} \frac{\bar{u}^p + \bar{u}}{|U^s(\frac{2\pi\ell}{N})|} \frac{M\rho(1 - \rho^{-N})}{(\rho - 1)^2} \rho^{-N_s} + \frac{\bar{V}^s(\frac{2\pi\ell}{N})}{|U^s(\frac{2\pi\ell}{N})|}$$

Proof: See Appendix A.1. □

The first term on the right hand side of the error bound given in the theorem is the error due to the effects of past inputs to the system, i.e. the effects of the unknown signals preceding the measurement interval. This error converges exponentially with N_s (convergence as ρ^{-N_s}). The properties of $|U^s(\frac{2\pi\ell}{N})|$ of course depend on the specific choice of the input signal $u(t)$ for $t \in T_{N_s}^N$. For a random signal the expectation of the magnitude of the N point DFT, as defined in (4.6) and (4.8), is asymptotically independent of N , see [106, lemma 6.2]. Hence, if the input is random for $t \in T_{N_s}^N$, the error due to the effects of unknown past inputs converges approximately as ρ^{-N_s}/\sqrt{N} . The second term on the right hand side is the error due to the additive noise on the output. This error does not converge at all, it is just the noise to signal ratio in the frequency domain. By designing an appropriate input signal, one can of course shape the error due to the noise. An input signal having a DFT with desired magnitude can be designed easily, see [151, 152].

We will now focus on the error due to the noise, the second term on the right hand side of the error bound given in Theorem 4.3.2. It is possible to obtain convergence for this error by choosing the input signal to be periodic. The highest rate of convergence is obtained by an input signal having an integer number of periods in the interval $T_{N_s}^N$. Let N_o denote the length of one period of the input signal and let the interval $T_{N_s}^N$ contain exactly r periods, so that $N = rN_o$. In this case $U^s(\frac{2\pi k}{N}) = 0$ if k/r is not an integer, only $U^s(\frac{2\pi k}{N_o})$ is not identically equal to zero, see [106, example 2.2]. It is now

straightforward to show that the DFT over r periods of a periodic signal is exactly \sqrt{r} times as large as the DFT over one period (see (B.1)). In conclusion, $|U^s(\frac{2\pi k}{N_o})|$ is exactly proportional to \sqrt{N} if $N = rN_o$ with $r \in \mathbb{N}$.

Corollary 4.3.3 *Consider a SISO system, satisfying the assumptions stated in Section 4.2. Using a partly periodic input signal having an integer number of periods in the interval $T_{N_s}^N$, $N_s \in T^{N+1}$, and the estimate*

$$\hat{G}^s(\frac{2\pi\ell}{N_o}) = \frac{Y^s(\frac{2\pi\ell}{N_o})}{U^s(\frac{2\pi\ell}{N_o})} \quad \text{for } \ell \in \{q \in T^{N_o} | U^s(\frac{2\pi q}{N_o}) \neq 0\}$$

the following error bound is satisfied

$$|G_o(\frac{2\pi\ell}{N_o}) - \hat{G}^s(\frac{2\pi\ell}{N_o})| \leq \alpha(\frac{2\pi\ell}{N_o})$$

with

$$\alpha(\frac{2\pi\ell}{N_o}) = \frac{1}{\sqrt{N}} \frac{\bar{u}^p + \bar{u}}{|U^s(\frac{2\pi\ell}{N_o})|} \frac{M\rho(1-\rho^{-N})}{(\rho-1)^2} \rho^{-N_s} + \frac{\bar{V}^s(\frac{2\pi\ell}{N_o})}{|U^s(\frac{2\pi\ell}{N_o})|}$$

The error bound given in the corollary goes to zero if N_s and N are going to infinity, N_o is constant, and the noise $v(t)$ does not contain a periodic component. The error due to the effects of unknown past inputs converges as ρ^{-N_s}/N . The error due to the additive noise on the output converges approximately as $1/\sqrt{N}$ if $v(t)$ is a random signal, because the expectation of the magnitude of the N point DFT of a random signal is asymptotically independent of N , see [106, lemma 6.2], while the magnitude of the DFT of the periodic input is exactly proportional to N . The price for this convergence is that less points of the transfer function are estimated (N_o instead of $N = rN_o$).

4.3.3 MIMO systems

The estimate and error bound of Theorem 4.3.2 and of Corollary 4.3.3 can be generalized easily to MIMO (multi input, multi output) systems by choosing a suitable vector input signal. That is, an error bound can be obtained for each element of the transfer function matrix.

First, we need some additional notation. Let $U_{[j]}^s$ denote the shifted DFT (4.8) of the input $u_{[j]}(t)$ to the j -th input channel, let $Y_{[i]}^s$ denote the shifted DFT (4.8) of the output $y_{[i]}(t)$ of the i -th output channel, and let $\hat{G}_{[ij]}^s$ denote the matching estimate (4.10) of the transfer function from input channel j to output channel i . Additionally, let $M_{[ij]}$ and $\rho_{[ik]}$ be such that $g_{o[ij]}(k) \leq M_{[ij]}\rho_{[ik]}^{-k}$, where $g_{o[ij]}(k)$ denotes the impulse response of the $G_{o[ij]}$, the element of the transfer function matrix G_o at row i and column j . Finally, let $V_{[i]}^s \leq \bar{V}_{[i]}^s$ where $V_{[i]}^s$ is the shifted DFT (4.8) of the additive noise $v_{[i]}(t)$ on the i -th output channel.

For a MIMO system having m inputs and p outputs the vector input signal now must be such that

$$U_{[i]}^s(\frac{2\pi\ell}{N}) U_{[j]}^s(\frac{2\pi\ell}{N}) = 0 \quad \text{for } i \neq j \quad (4.11)$$

for all $i, j = 1, 2, \dots, m$ and $\ell \in T^N$. That is, the inputs signals to the different input channels should be mutually orthogonal in the frequency domain. A time domain set of input signals which is in accordance with (4.11) can easily be realized, see [158, 151, 152].

For MIMO systems, we also have to generalize the concept of a partly periodic signal.

Definition 4.3.4 *A vector partly periodic signal x is a vector signal where each element of the vector has the first N_s points equal to the last N_s points: $x(t + N) = x(t)$ for $t \in T^{N_s}$.*

Theorem 4.3.5 *Consider a MIMO system G_o having m inputs and p outputs. Let the elements $G_{o[ij]}$, $i = 1, \dots, m$, $j = 1, \dots, p$, of the transfer function matrix G_o satisfy the assumptions stated in Section 4.2. Using a vector partly periodic input signal that is in accordance with (4.11), $N_s \in T^{N+1}$, and the estimate*

$$\hat{G}_{[ij]}^s(\frac{2\pi\ell}{N}) = \frac{Y_{[i]}^s(\frac{2\pi\ell}{N})}{U_{[j]}^s(\frac{2\pi\ell}{N})} \quad \text{for } \ell \in \{q \in T^N \mid U_{[j]}^s(\frac{2\pi q}{N}) \neq 0\}$$

the following error bound is satisfied for $i = 1, \dots, m$, $j = 1, \dots, p$

$$|G_{o[ij]}(\frac{2\pi\ell}{N}) - \hat{G}_{[ij]}^s(\frac{2\pi\ell}{N})| \leq \alpha_{[ij]}(\frac{2\pi\ell}{N})$$

with

$$\alpha_{[ij]}(\frac{2\pi\ell}{N}) = \frac{1}{\sqrt{N}} \frac{1}{|U_{[j]}^s(\frac{2\pi\ell}{N})|} \sum_{k=1}^m \left((\bar{u}_{[k]}^P + \bar{u}_{[k]}) \frac{M_{[ik]}\rho_{[ik]}}{(\rho_{[ik]} - 1)^2} \rho_{[ik]}^{-N_s} (1 - \rho_{[ik]}^{-N}) \right) + \frac{\bar{V}_{[i]}^s(\frac{2\pi\ell}{N})}{|U_{[j]}^s(\frac{2\pi\ell}{N})|}$$

Proof: See Appendix A.2. □

In comparison with the SISO case, the error in $\hat{G}_{[ij]}^s$ due to the effects of unknown past inputs becomes m (the number of inputs) times as large (if $M_{[ij]} = M$, $\rho_{[ij]} = \rho$, $\bar{u}_{[j]}^P = \bar{u}^P$ and $\bar{u}_{[j]} = \bar{u}$). This is not as bad as it may seem because the error due to the effects of unknown past inputs approaches zero exponentially fast with N_s . Hence N_s need by far not be m times as large to get an error that equals the one for the single input case. Hence, concerning experimentation time, it is advantageous to perform MIMO experiments, and not SIMO (single input, multi output) experiments, to identify the MIMO system.

The error bounds of Theorem 4.3.5 which address the elements of the transfer function matrix, can be used in a μ analysis or synthesis approach to robust control, see [38, 40, 7]. The uncertainty then can be specified as structured uncertainty in the form of a diagonal matrix with complex valued diagonal elements bounded by Theorem 4.3.5.

4.3.4 Remarks

A partly periodic signal can be seen as a generalization of a sinewave input. This generalization is useful since sinewave testing (sinewave excitation and actually measuring the frequency response in a finite number of frequency points) is time consuming. For each new sinewave input one must wait until the system has reached its steady state response. A partly periodic signal can consist of N sinewaves, but one has to wait only once for the effects of unknown past inputs to vanish.

For $N_s = 0$ the ETFE as defined in [106] arises. In this case the error due to unknown past inputs converges as $1/\sqrt{N}$ if $u(t)$ is a random signal for $t \in T^N$, as was also shown in [106]. The error due to unknown past inputs now will in general dominate the error bound. This is not just a result of a poor upper bound on the error, the actual error indeed can be large. However, for $N_s = 0$ the input signal is completely free. The choice for $N_s > 0$ hence is a choice to restrict the input signal in order to be able to obtain a tight error bound for the nominal model.

4.4 Continuous error bound

4.4.1 Motivation

We now have an upper bound $\alpha(\omega_k)$ on the error $|G_o(e^{j\omega_k}) - \hat{G}^s(e^{j\omega_k})|$. This error bound is only defined in the finite number of frequency points $\omega_k \in \Omega_N^u$, with $\Omega_N^u := \{2\pi k/N, k \in T^N \mid |U^s(e^{j(2\pi k/N)})| > 0\}$. This is due to the fact that $\hat{G}^s(e^{j\omega_k})$ is only defined at a finite number of frequency points when N , the number of data points used in the estimate, is finite. The aim is to find an upper bound $\delta(\omega)$ such that

$$|G_o(e^{j\omega}) - G_{nom}(e^{j\omega})| \leq \delta(\omega)$$

for all frequencies in the interval $[0, 2\pi)$. It is straightforward to give a discrete upper bound $\delta(\omega_k)$. First note that $\beta(\omega_k) = |\hat{G}^s(e^{j\omega_k}) - G_{nom}(e^{j\omega_k})|$ can be calculated exactly because G_{nom} is assumed to be known. From the inequality

$$|G_o(e^{j\omega_k}) - G_{nom}(e^{j\omega_k})| \leq |G_o(e^{j\omega_k}) - \hat{G}^s(e^{j\omega_k})| + |\hat{G}^s(e^{j\omega_k}) - G_{nom}(e^{j\omega_k})| \quad (4.12)$$

it now follows that a possible choice for $\delta(\omega_k)$ is $\delta(\omega_k) = \alpha(\omega_k) + \beta(\omega_k)$. Hence the problem is to find the behaviour of $\delta(\omega)$ between the estimated frequency points for the prespecified nominal model. As argued in Section 4.3.1, the data does essentially not contain more information about the transfer function of the system than is captured by the discrete estimate $\hat{G}^s(e^{j\omega_k})$. Therefore, assumptions about the system must be used to be able to bound the error at frequencies $\omega \neq \omega_k$. We will use smoothness assumptions on the system, and we will interpolate the discrete error bound $\delta(\omega_k)$ using these smoothness properties.

In order to establish a continuous error bound we will not interpolate or fit a model to the discrete estimate $\hat{G}^s(e^{j\omega_k})$, and subsequently bound the model error that could arise with the interpolation or fitting technique employed, as is done in the existing "identification in H_∞ " literature, see e.g. [139, 75, 141, 142, 62, 2, 22, 79]. As argued before, in this approach identification techniques have to be used for which error bounds can be established, resulting in procedures that sacrifice the quality of the model in order to be able to come up with hard error bounds. Moreover, the resulting model usually is of high order, so that model reduction is necessary to arrive at a nominal model which has a suitable order in view of control design, thus introducing still an additional term due to model reduction to the error bound. In contrast, we will use a prespecified nominal model, and interpolate the error bound $\delta(\omega_k)$.

4.4.2 Bounds on derivatives

Smoothness properties of the system in the form of upper bounds on the derivatives of $G_o(e^{j\omega})$ with respect to frequency, can be obtained from the assumed upper bound on the impulse response.

Proposition 4.4.1 *For a SISO system with $|g_o(m)| \leq M\rho^{-m}$ there holds*

$$\begin{aligned} \left| \frac{d G_o(e^{j\omega})}{d\omega} \right| &\leq \frac{M\rho}{(\rho-1)^2} \\ \left| \frac{d^2 G_o(e^{j\omega})}{d\omega^2} \right| &\leq \frac{M\rho(\rho+1)}{(\rho-1)^3} \end{aligned}$$

Proof: See Appendix A.3. □

The SISO system is allowed to be an element of a MIMO system. In that case G_o must be replaced by $G_{o[ij]}$, g_o by $g_{o[ij]}$, M by $M_{[ij]}$ and ρ by $\rho_{[ij]}$.

Note that the upper bounds on the derivatives are not obtained in the same way as in [75]; the M and ρ used in this paper are different from the ones used there. It is the author's opinion that it is easier in practice to obtain a good upper bound on the impulse response, than to obtain a margin of relative stability of the system together with the infinity norm of the system over the circle in the complex plane with radius equal to this margin as used in [75]. Moreover, in [130] it is shown that the prior information M and ρ used in [75] yields extremely conservative bounds on the impulse response. Hence, the prior information M and ρ should bound the impulse response directly.

To be able to bound the derivatives of the magnitude of the error system $|G_o(e^{j\omega}) - G_{nom}(e^{j\omega})|$ we need the following proposition.

Proposition 4.4.2 *For a SISO system there holds*

$$\left| \frac{d^k}{d\omega^k} |G_o(e^{j\omega}) - G_{nom}(e^{j\omega})| \right| \leq \left| \frac{d^k}{d\omega^k} (G_o(e^{j\omega}) - G_{nom}(e^{j\omega})) \right| \quad (4.13)$$

$$\leq \left| \frac{d^k G_o(e^{j\omega})}{d\omega^k} \right| + \left| \frac{d^k G_{nom}(e^{j\omega})}{d\omega^k} \right| \quad (4.14)$$

for $k = 1$ and $k = 2$.

Proof: See Appendix A.4. □

Again, the SISO system is allowed to be an element of a MIMO system.

An upper bound for (4.14) can be calculated using Proposition 4.4.1 and the knowledge of $G_{nom}(e^{j\omega})$. If an upper bound on $|g_o(m) - g_{nom}(m)|$ is known, we are able to calculate an upper bound for (4.13) directly from Proposition 4.4.1.

4.4.3 Interpolation

In this section we will address the problem of calculating an upper bound on the error $|G_o(e^{j\omega}) - G_{nom}(e^{j\omega})|$ between the frequency points ω_k where an upper bound $\delta(\omega_k)$ is known. Hence, we have to find the highest possible value $\delta(\omega)$ of this error for each frequency ω between two given points, say $\delta(\omega_k)$ and $\delta(\omega_{k+1})$. We are able to bound this error by taking into account the bounds on the first and second derivatives of $|G_o(e^{j\omega}) - G_{nom}(e^{j\omega})|$ that were derived in Section 4.4.1, say γ_1 and γ_2 respectively. The maximum value of the error $\delta(\omega)$ now arises by interpolating the discrete error bound $\delta(\omega_k)$ using the function $f(x)$ depicted in Fig. 4.1.

To explain the construction of this function $f(x)$, assume that there is a maximum between the two frequency points. Starting at the maximum ($x = 0$, $f(x) = 0$ and $df(x)/dx = 0$) we want $f(x)$, in a smooth way, to decrease as fast as possible: the faster $f(x)$ decreases, the higher the maximum lies above the two given points $\delta(\omega_k)$, $\delta(\omega_{k+1})$. Hence we use a function having a constant second derivative equal to the bound γ_2 on this derivative. In this way parts II and III of the error bound are constructed. The absolute value of the first derivative of this function will clearly increase with the distance $|x|$ to the maximum. At $|x| = \gamma_1/\gamma_2$ the first derivative becomes equal to the bound γ_1 on this derivative. Hence, for $|x| > \gamma_1/\gamma_2$ we use a function having a constant first derivative equal to the bound γ_1 . In this way part I and IV of the error bound are constructed. The function constructed in this way is unique and given by

$$\begin{aligned} f(x) &= -\frac{\gamma_2}{2} x^2 & \text{for } |x| \leq \frac{\gamma_1}{\gamma_2} \\ &= -\gamma_1 |x| + \frac{\gamma_1^2}{2\gamma_2} & \text{for } |x| > \frac{\gamma_1}{\gamma_2} \end{aligned} \quad (4.15)$$

The function $f(x)$ given in (4.15) directly gives the value of $\delta(\omega)$

$$\delta(\omega) = \delta(\omega_k) - f(\Delta x_1) + f(x) \quad \text{for } \omega \in [\omega_k, \omega_{k+1}] \quad (4.16)$$

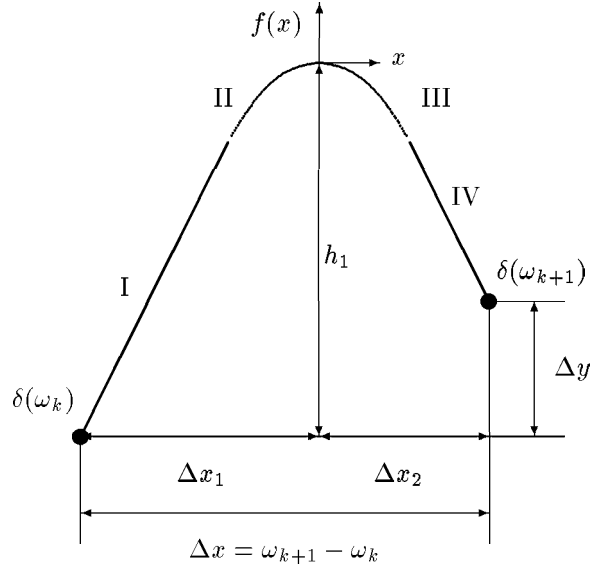


Fig. 4.1: The interpolating function $f(x)$ for the discrete error bound.

However, in (4.16) the values of Δx_1 and x are still unknown, because the location of the maximum is as yet unknown. Analytic expressions for the location of the maximum can be given, by specifying Δx_1 or Δx_2 as a function of $\delta(\omega_k)$, $\delta(\omega_{k+1})$, γ_1 and γ_2 .

To be able to provide these analytic expressions for the location of the maximum one has to distinguish between several cases, depending on which part of the interpolating function $f(x)$ as given by (4.15) actually is used. It is e.g. possible that γ_1 , γ_2 , $\delta(\omega_k)$ and $\delta(\omega_{k+1})$ are such that the interpolating function $f(x)$, see Fig. 4.1, reduces to part I. In all, there are ten possibilities: only part I, only part II, part I and II, etc.

Algorithm 4.4.3 All possibilities of the function $f(x)$ given in (4.15) to interpolate two points are given below, as a function of Δx , Δy , γ_1 and γ_2 .

A maximum occurs if

$$|\Delta y| < \gamma_1 \Delta x - \frac{\gamma_1^2}{2\gamma_2} \quad \text{and} \quad \Delta x \geq \frac{\gamma_1}{\gamma_2}$$

or if

$$|\Delta y| < \frac{\gamma_2}{2} \Delta x^2 \quad \text{and} \quad \Delta x \leq \frac{\gamma_1}{\gamma_2}$$

If a maximum occurs we can distinguish the following four cases.

1. If $\Delta x_1 \geq \gamma_1/\gamma_2$ and $\Delta x_2 \geq \gamma_1/\gamma_2$ then $\Delta x_1 = \frac{\Delta y + \gamma_1 \Delta x}{2\gamma_1}$.
All four parts of $f(x)$, as depicted in Fig. 4.1, are used.

2. If $\Delta x_1 \geq \gamma_1/\gamma_2$ and $\Delta x_2 < \gamma_1/\gamma_2$ then $\Delta x_1 = \frac{\gamma_1}{\gamma_2} + \Delta x - \sqrt{\frac{2}{\gamma_2}(\gamma_1 \Delta x - \Delta y)}$.
Parts I, II and III of $f(x)$ are used.
3. If $\Delta x_1 < \gamma_1/\gamma_2$ and $\Delta x_2 \geq \gamma_1/\gamma_2$ then $\Delta x_1 = \sqrt{\frac{2}{\gamma_2}(\gamma_1 \Delta x + \Delta y)} - \frac{\gamma_1}{\gamma_2}$.
Parts II, III and IV of $f(x)$ are used.
4. If $\Delta x_1 < \gamma_1/\gamma_2$ and $\Delta x_2 < \gamma_1/\gamma_2$ then $\Delta x_1 = \frac{\Delta y}{\gamma_2 \Delta x} + \frac{\Delta x}{2}$.
Parts II and III of $f(x)$ are used.

The maximum height h_1 above $\delta(\omega_k)$ is given by $h_1 = -f(\Delta x_1)$, where $f(x)$ is given in (4.15).

If no maximum occurs we can distinguish the following six cases.

1. If $\gamma_1 \Delta x - \frac{\gamma_2}{2} \Delta x^2 \leq \Delta y < \gamma_1 \Delta x$ then $\Delta x_1 = \frac{\gamma_1}{\gamma_2} + \Delta x - \sqrt{\frac{2}{\gamma_2}(\gamma_1 \Delta x - \Delta y)}$. Note that $\Delta x_1 \geq \Delta x$. Parts I and II of $f(x)$ are used.
2. If $\frac{\gamma_2}{2} \Delta x^2 \leq \Delta y < \gamma_1 \Delta x - \frac{\gamma_2}{2} \Delta x^2$ then $\Delta x_1 = \frac{\Delta x}{2} + \frac{\Delta y}{\gamma_2 \Delta x}$. Note that $\Delta x_1 \geq \Delta x$. Only part II of $f(x)$ is used.
3. If $\gamma_1 \Delta x - \frac{\gamma_2}{2} \Delta x^2 \leq -\Delta y < \gamma_1 \Delta x$ then $\Delta x_2 = \frac{\gamma_1}{\gamma_2} + \Delta x - \sqrt{\frac{2}{\gamma_2}(\gamma_1 \Delta x + \Delta y)}$. Note that $\Delta x_2 \geq \Delta x$. Parts III and IV of $f(x)$ are used.
4. If $\frac{\gamma_2}{2} \Delta x^2 \leq -\Delta y < \gamma_1 \Delta x - \frac{\gamma_2}{2} \Delta x^2$ then $\Delta x_2 = \frac{\Delta x}{2} - \frac{\Delta y}{\gamma_2 \Delta x}$. Note that $\Delta x_2 \geq \Delta x$. Only part III of $f(x)$ is used.
5. If $\Delta y = \gamma_1 \Delta x$ then $\Delta x_1 = \frac{\gamma_1}{\gamma_2} + \Delta x$. Only part I of $f(x)$ is used.
6. If $\Delta y = -\gamma_1 \Delta x$ then $\Delta x_2 = \frac{\gamma_1}{\gamma_2} + \Delta x$. Only part IV of $f(x)$ is used.

Using this algorithm and (4.15),(4.16) we are able to calculate an upper bound for the difference between the system and the nominal model for all $\omega \in [0, 2\pi)$.

If $|\Delta y| > \gamma_1 \Delta x$ the estimated point of the discrete estimate with the highest error bound must not be used. Interpolation from neighbouring points, although over a greater distance, gives a lower error bound. This situation can also arise when $|\Delta y| \leq \gamma_1 \Delta x$, see Fig. 4.2.

Note that, as opposed to existing algorithms for "identification in H_∞ ", e.g. [75, 76, 78, 141, 142, 115, 63, 62, 79], 2Tipf allows for a discrete error bound $\alpha(\omega_k)$ that is frequency dependent, and it yields a continuous error bound $\delta(\omega)$ that is frequency dependent. Moreover the discrete frequency points ω_k are not required to be equidistant.

4.4.4 Remarks

Taking a closer look at the results of this section and the previous one, we can summarize in the following way. In Section 4.3 a bound $\alpha(\omega_k)$ has been derived

$$|G_o(e^{j\omega_k}) - \hat{G}^s(e^{j\omega_k})| \leq \alpha(\omega_k)$$

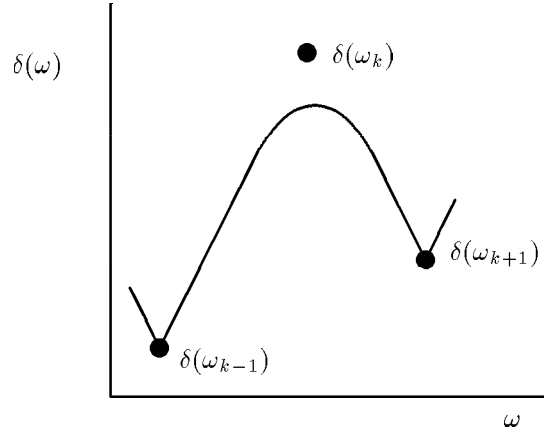


Fig. 4.2: A situation in which the point $\delta(\omega_k)$ must not be used.

for all ω_k in a set $\Omega_N^u \subset \mathbb{R} \cap [0, 2\pi)$ containing a finite number ($\leq N$) of elements. Since the nominal model is known, the error

$$\beta(\omega_k) := |\hat{G}^s(e^{j\omega_k}) - G_{nom}(e^{j\omega_k})|$$

can be calculated exactly for all $\omega_k \in \Omega_N^u$. In this Section 4.4, a continuous bound $\delta(\omega)$ is derived, such that

$$|G_o(e^{j\omega}) - G_{nom}(e^{j\omega})| \leq \delta(\omega)$$

with

$$\delta(\omega_k) = \alpha(\omega_k) + \beta(\omega_k) \quad \text{for } \omega_k \in \Omega_N^u$$

This continuous error bound $\delta(\omega)$ is robustly convergent. That is, the error bound converges to the true error when N_s and N are approaching infinity and $\bar{V}^s(\frac{2\pi\ell}{N})$ approaches zero.

In the nonparametric discrete estimate, cf. (4.4.4), no error due to undermodelling is present, i.e. no error due to approximation is made, because complete freedom exists for each frequency point to fit $G_o(e^{j\omega_k})$. The approximation error therefore is completely due to the nominal model, cf. (4.4.4).

In the procedure presented, the determination of the nominal model and the determination of the error bound clearly are completely separated. We addressed the problem of determining the error bound. Methods for tuning the nominal model to nominal control design specifications are discussed in [11, 12, 69, 155, 157, 175, 5, 61, 140, 103, 48].

4.5 Validation of prior information

Using the discrete estimate $\hat{G}^s(e^{j\omega_k})$ it is also possible to check whether the prior information on the system M and ρ and the prior information on the noise $\bar{V}(\frac{2\pi k}{N})$ are consistent. That is, we are able to detect serious conflicts between the prior information on the system and the prior information on the noise. The detection is based on the fact that a serious inconsistency will cause the discrete estimate, the discrete error bound and the bounds on the derivatives to be conflicting, as formulated in the following theorem.

Theorem 4.5.1 *The prior information on the system M and ρ and the prior information on the noise $\bar{V}(\frac{2\pi k}{N})$ are conflicting if for any $\omega_k, \omega_{k-1} \in \Omega_N^u$*

$$|\hat{G}^s(e^{j\omega_k}) - \hat{G}^s(e^{j\omega_{k-1}})| > \frac{M\rho}{(\rho-1)^2} |\omega_k - \omega_{k-1}| + \alpha(\omega_k) + \alpha(\omega_{k-1})$$

where $\alpha(\omega_k), \omega_k \in \Omega_N^u$, is given in Theorem 4.3.2.

Proof: Follows directly from Proposition 4.4.1 and Proposition 4.4.2. \square

Clearly, if a conflict arises then not all assumptions of Section 4.2 can be correct, and one should carefully check the assumed values of M , ρ and $\bar{V}(\frac{2\pi k}{N})$.

Chen *et al.* [22] also provide a method to validate the prior information. This method may be more perceptive to conflicts than the simple test of Theorem 4.5.1, but is also computationally much more involved.

4.6 Relation with control design specifications

To show the applicability of the approach presented in this paper to robust control design, we will consider the following situation. In order to verify desired robustness properties of a designed controller for the system, an allowable error bound is specified for the difference between G_o and G_{nom}

$$|G_o(e^{j\omega}) - G_{nom}(e^{j\omega})| \leq \delta_a(\omega)$$

The allowable error $\delta_a(\omega)$ is a function of the nominal model, the designed controller and the robust control design specifications. Given measurement data from the system, it now has to be verified whether a specific nominal model satisfies the specified error bound. If not, it should be determined which action should be taken in order to solve the problem: either constructing a new nominal model, or performing additional experiments to reduce the uncertainty.

The actual error bound $\delta(\omega)$ for the nominal model clearly is a function of the nominal model itself and of the discrete estimate \hat{G} . Therefore both should be tuned to

the robust control design specifications. This can be done by comparing the allowable error $\delta_a(\omega)$ with the actual error bound $\delta(\omega)$. For those values of ω where $\delta(\omega) > \delta_a(\omega)$ we can analyse $\delta(\omega)$ and evaluate its different components.

At the finite number of frequency points $\omega_k \in \Omega_N^u$, we have $\delta(\omega_k) = \alpha(\omega_k) + \beta(\omega_k)$. Therefore we know that

1. when $\alpha(\omega_k) \gg \beta(\omega_k)$, the uncertainty is mainly due to the inherent uncertainty in the data $\alpha(\omega_k)$, i.e. effects of unknown past inputs, low signal-to-noise ratio and/or restricted length of the data set. Actions to be taken to improve the bound include: increasing N_s , increasing the power of the input signal, and increasing N . In the case of periodic input signals, the signal-to-noise ratio in the frequency domain is proportional to $\sqrt{N/N_o}$. Consequently the error bound can also be improved by decreasing N_o .
2. when $\alpha(\omega_k) \ll \beta(\omega_k)$, the uncertainty is mainly due to a bad nominal model. A straightforward action is then to choose a new nominal model, that is better able to represent the system dynamics in the specific frequency range.

In between the finite number of frequency points $\omega_k \in \Omega_N^u$, say for $\omega_k < \omega < \omega_{k+1}$, the error bound $\delta(\omega)$ is determined through interpolation between the adjacent points $\delta(\omega_k)$, $\delta(\omega_{k+1})$. Therefore

3. when $\delta(\omega) \gg \max(\delta(\omega_k), \delta(\omega_{k+1}))$, the uncertainty is mainly due to the interpolation step. Note that the uncertainty due to interpolation is strongly determined by the distance between two adjacent discrete frequency points. Consequently new experiments should be performed with a smaller distance between the discrete frequency points in the specific frequency region.

Note that it is possible to determine whether the main source of the actual error is the inherent uncertainty in the data, the nominal model, or the interpolation step caused by the absence of data due to the specific excitation of the system. Also it is possible to decrease the contribution of these different error sources almost independently. Now it is possible to iteratively decrease the error bound, until the level of the allowable error is reached, successively by input design and additional experiments, and by tuning the nominal model. Using this procedure we can determine whether or not specific robust control design specifications can be met (for a fixed controller).

Both the inherent uncertainty in the data $\alpha(\omega_k)$, and the error due to interpolation can be made arbitrarily small, c.q. the error bound can be made arbitrarily tight, in a certain frequency region by improving the discrete estimate. Note that the error bound $\alpha(\omega_k)$ is essentially frequency dependent and that the frequency points $\omega_k \in \Omega_N^u$ need not be positioned equidistantially over the frequency axis. In comparison with existing techniques for "identification in H_∞ " this creates a lot of freedom to shape the error bound into an accepted (allowable) form, which -from a control point of view- definitely should be frequency dependent.

4.7 Example

To illustrate the results of this chapter a simulation was made of a fifth order system

$$G_o(z) = \frac{0.82 - 1.04z^{-1} + 0.28z^{-2} + 0.61z^{-3} - 1.05z^{-4} + 0.47z^{-5}}{1 - 2.47z^{-1} + 2.88z^{-2} - 1.97z^{-3} + 0.81z^{-4} - 0.17z^{-5}}$$

whose impulse response satisfies a bound given by $M_o = 3$ and $\rho_o = 1.256$. There was 10 percent (in amplitude) colored noise (high pass filtered white noise) on the output. The nominal model is given as

$$G_{nom}(z) = \frac{0.79 + 0.09z^{-1} - 0.24z^{-2} + 0.63z^{-3}}{1 - 1.25z^{-1} + 0.75z^{-2} + 0.05z^{-3}}$$

Based on this nominal model a controller was designed for the system

$$C_o(z) = \frac{1.60z^{-2} - 1.18z^{-3} + 0.81z^{-4}}{1 - 2.85z^{-1} + 3.45z^{-2} - 2.28z^{-3} + 0.69z^{-4}}$$

In this example we will focus on the following two questions

- When the controller is applied to the true system $G_o(z)$, will the resulting closed loop configuration be stable ?
- What is the quality of the nominal model ?

To answer the first question we will investigate whether the loop gain $G_o(\omega)C_o(\omega)$ could encircle the point -1 in the Nyquist diagram. In order to build in some safety c.q. some robust performance, we require in addition that the loop gain does not enter a circle with radius 0.3 around the point -1 . The resulting allowable uncertainty is

$$\delta_a(\omega) = \frac{|G_{nom}(\omega)C_o(\omega) + 1| - 0.3}{|C_o(\omega)|}$$

As a priori information on the impulse response we choose $M = 3$ and $\rho = 1.2$. The upper bound $\bar{V}^s(\omega_k)$ was set to $\bar{V}^s(\omega_k) = 3\sqrt{\Phi_v(\omega_k)}$, where $\Phi_v(\omega_k)$ denotes the spectrum of the noise. For normally distributed noise this upper bound has a probability of 99.99 % of being valid, and indeed was satisfied for all experiments.

The input signal was chosen to obey $\bar{u}^p = 2$ and $\bar{u} = 1$. For the first experiment we used 178 points with $N = 128$, $N_o = 128$ and $N_s = 50$. The magnitude of the DFT of the input signal in the interval $T_{N_s}^N$, $|U^s(\omega_k)|$, is given in Fig. 4.3. In Fig. 4.4 the allowable error $\delta_a(\omega)$, the error bound $\delta(\omega)$ and the error due to approximation $\beta(\omega_k)$ are given. The inherent uncertainty in the data $\alpha(\omega_k)$ equals $\delta(\omega_k) - \beta(\omega_k)$. The error due to interpolation is indicated by the curves between the points $\delta(\omega_k)$. Clearly, in the frequency interval $\omega = [0.8, 1.3]$ rad/s it is possible that the actual error is larger than the allowable one, so that no stability guarantee can be given. In order to be able to improve the error bound we will take a closer look at the error generating

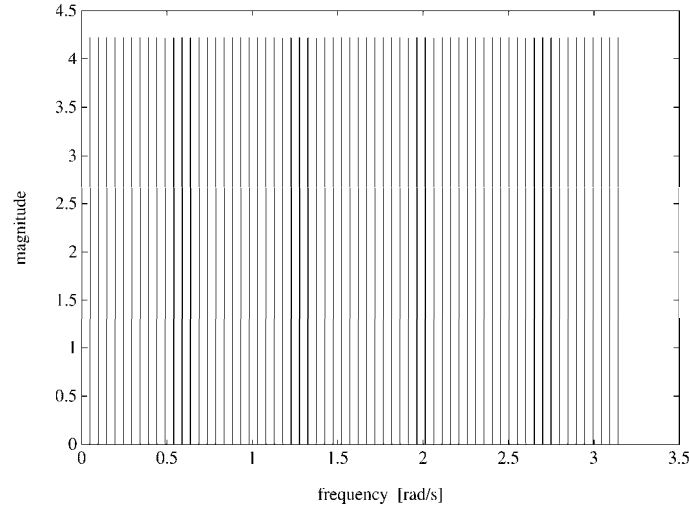


Fig. 4.3: $|U^s(\omega_k)|$, the magnitude of the DFT of the input signal in the interval $T_{N_s}^N$.

processes. The true error due to approximation is fixed, because the nominal model is fixed. Using the prior information M , ρ , \bar{u}^p and \bar{u} , it follows from Corollary 4.3.3 that the error due to unknown past inputs is less than 0.01, and therefore is neglectable. Hence we can only improve upon the errors due to noise and interpolation. From Fig. 4.4 it follows that in the frequency interval $\omega = [0.8, 1.1]$ rad/s the noise as well as the interpolation error need to be improved. In the frequency interval $\omega = [1.1, 1.3]$ rad/s it probably will suffice to decrease the error due to the noise only.

The specific improvements mentioned above can be achieved accurately by input design. The magnitude of the DFT of the new input signal is given in Fig. 4.5. We choose $N = 512$, $N_o = 256$ and $N_s = 50$. We used an input signal containing two periods, in order to obtain a higher signal to noise ratio in the frequency domain $|U^s(\omega_k)|/|V^s(\omega_k)|$ compared to the previous experiment, while maintaining the constraint $\bar{u} = 1$. Also, the period length N_o was increased to obtain a denser DFT frequency grid. This denser grid is needed in the frequency interval $\omega = [0.8, 1.1]$ rad/s to be able to reduce the distance between two frequency points ω_k when compared to Fig. 4.3, in order to decrease the error due to interpolation in this frequency region. The magnitude of the DFT of the input signal was shaped as follows. Because the error due to unknown past inputs is neglectable, we have $\alpha(\omega_k) \approx \bar{V}^s(\omega_k)/|U^s(\omega_k)|$. Also $\delta(\omega_k) = \alpha(\omega_k) + \beta(\omega_k) \leq 2\alpha(\omega_k) + |G_o(\omega_k) - G_{nom}(\omega_k)|$. Choosing $|U^s(\omega_k)| = 2(\delta_a(\omega_k) - |G_o(\omega_k) - G_{nom}(\omega_k)|)^{-1}\bar{V}^s(\omega_k)$ therefore would result in $\delta(\omega_k) \leq \delta_a(\omega_k)$. However, because G_o is unknown, we choose $|U^s(\omega_k)|$ to be proportional to $(\delta_a(\omega_k) - \beta(\omega_k))^{-1}\bar{V}^s(\omega_k)$.

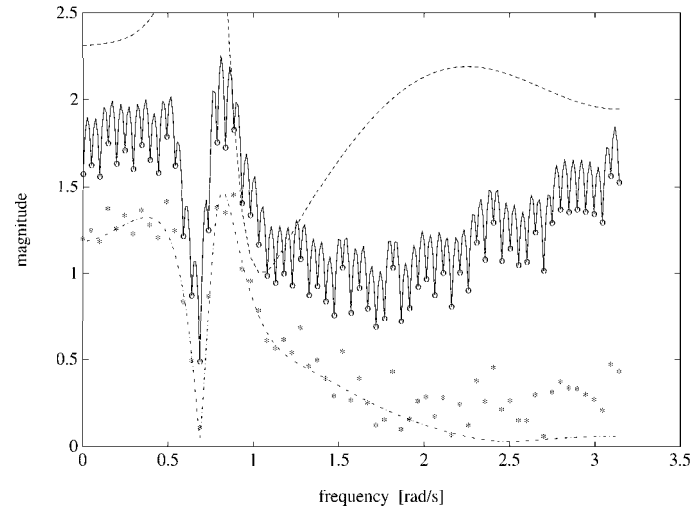


Fig. 4.4: The error bounds and the true error: $\delta_a(\omega)$ (---), $\delta(\omega)$ (—), $\delta(\omega_k)$ (o), $\beta(\omega_k)$ (*), $|G_o(\omega) - G_{nom}(\omega)|$ (-·).

In Fig. 4.6 the resulting error bounds are given, together with the allowable error and the true error. Note that $\beta(\omega_k)$ provides a good indication of the true error, and that the error bound $\delta(\omega)$ can be made almost equal to the true error by input design. Combining the two error bounds, which is possible because the nominal model is fixed, proves that the actual error is indeed lower than the allowable error, meaning that the true closed loop system will be stable. The controller therefore can be implemented safely. However, the nominal model is not correct. A better nominal model certainly is desirable. If the control requirements would have been slightly more severe, a better (higher order) nominal model even would have been necessary to be able to prove stability of the closed loop. Clearly, the modelling error in the frequency interval $\omega = [0.8, 1.1]$ rad/s is crucial. A shorter experimentation time would have been sufficient if the nominal model would have been more accurate in this frequency interval.

The above demonstrates the interplay between the accuracy of the error bound (experimentation time), modelling accuracy (modelling effort, model order), and control specifications (robustness, performance).

4.8 Conclusions

In this chapter a procedure has been presented to quantify the model uncertainty of any prespecified nominal model, given a sequence of measurement data from a plant. In the procedure presented the empirical transfer function estimate (ETF) is used

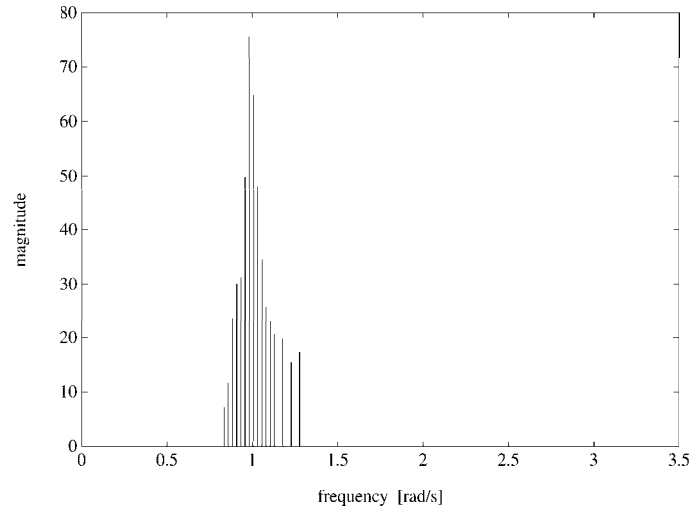


Fig. 4.5: $|U^s(\omega_k)|$, the magnitude of the DFT of the input signal in the interval $T_{N_s}^N$.

to construct a -nonparametric- estimate of the transfer function in a finite number of frequency points, together with an upper bound on the error. Through interpolation, this error bound is transformed to a bound which is available on a continuous frequency interval. A frequency dependent upper bound is obtained, which is much more tailored to the needs of a robust control design scheme, than an H_∞ bound. In order to obtain a tight error bound, a special input signal is proposed (partly periodic) which has advantages over -classical- sinewave experiments.

The estimated upper bound for the model error of a prespecified nominal model can be split into three parts: one part due to the inherent uncertainty in the data, a second part due interpolation, and a third part due to imperfections of the nominal model. These three components can be tuned almost independently, by appropriate experiment design and by choosing an appropriate nominal model. When the error bound is too conservative in view of control design specifications, information is provided as to which action should be taken (new experiments or alternative nominal model) in order to satisfy the design requirements.

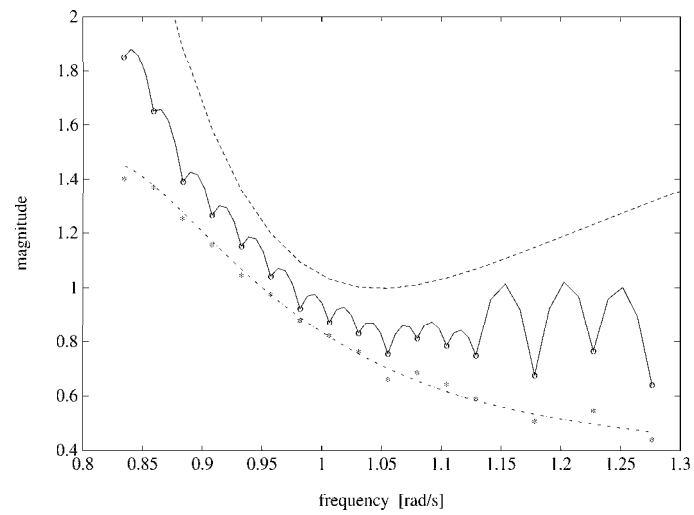


Fig. 4.6: The error bounds and the true error: $\delta_a(\omega)$ (---), $\delta(\omega)$ (—), $\delta(\omega_k)$ (o), $\beta(\omega_k)$ (*), $|G_o(\omega) - G_{nom}(\omega)|$ (-·).

Chapter 5

Mixed averaging – worst case model error bounds

Summary

We consider the problem of identification for robust control design, while focusing on the quantification of model uncertainty. A procedure is presented to identify, on the basis of time domain measurement data, a reduced order parametric model together with a bound on the model error, while using only minor a priori information. The measurement data is assumed to be contaminated with a stochastic noise disturbance with unknown spectral properties. For the parametric estimate we employ system based orthonormal basis functions, which can be generated by any (control relevant) nominal model specified by the user. By applying data segmentation, estimating a parametric model on each data segment, and averaging over the resulting set of parametric estimates, in conjunction with a periodic input signal, the statistics of the model error can be obtained asymptotically from the data. The model error consists of two parts: an averaging (probabilistic) part, due to the stochastic noise disturbance on the data, and a structural (worst case) part, due to the unmodelled dynamics and unknown past inputs. The latter is explicitly bounded with a hard error bound, while for the former a confidence interval can be specified asymptotically. For this analysis only minor assumptions are made on the distribution of the noise. Proceeding in this way, an error bound which is defined for each frequency $\omega \in [0, 2\pi)$ is established for the parametric transfer function estimate. The bound on the parametric transfer function estimate can be split up into a parametric part and a nonparametric part. Additionally, error bounds are obtained for the estimated parameters, the estimated undisturbed output, and the discrete Fourier transform (DFT) of the estimated disturbed output. These error bounds are important for validation of the prior information, the as-

sumptions, and the error bound obtained for the parametric transfer function estimate. Furthermore, they may be valuable for analysis, prediction, fault detection, feedforward control and predictive control. The resulting error bounds appear to be highly realistic and, as a consequence, suitable for high performance robust control design purposes. Moreover, separate bounds are obtained for all error sources. This enables an explicit tradeoff between bias (undermodelling) and variance (noise) terms by choosing the model order, and provides an answer to the question how the error bounds can be improved effectively when the bound is too large in view of the robust control design requirements.

5.1 Introduction

The aim of this chapter is to find, from experimental data, a reliable and tight quantification of the region containing (with a certain probability) the transfer function of the system that generated the data. That is, we will address the problem of identifying an accurate transfer function model, together with a quantification of the uncertainty that is present in this estimate.

The identification of linear models together with error bounds is an important issue that recently has attracted a lot of attention. In general terms, the transfer function is one of the main tools used in the analysis and manipulation of the properties of a linear system, and an error bound on an estimated model is indispensable for assessing the quality of any model describing the dynamical system that underlies the measured data. However, recent interest in identifying models specifically directed towards the consecutive design of robust control systems, in particular has been a stimulus for developing identification methods that provide model error bounds.

Early references addressing the problem of describing the uncertainty in a transfer function estimate include e.g. Jenkins and Watts [86], Schweppe [159] and Ljung and Caines [108], whereas comprehensive accounts of either spectral analysis or parametric identification methods are given in Brillinger [20] and Ljung [106] (for a concise description we refer to Section 2.3 and Section 2.2).

The recent approaches to this estimation problem can be divided into two areas, characterized by the type of error bounds that arise: soft (probabilistic) or hard (worst case) error bounds. The two different approaches are simply shown to be the result of different priors on the data generating process, see e.g. Ljung *et al.* [112] and Hjalmarsson [83]. Thus, the type of uncertainty quantification that is obtained originates from the way in which the experimental data is assumed to be embedded in the data generating process, as was also discussed in Section 2.1.

In view of robust control design, one would like to have available a model together with a hard bound on the model error, measured in e.g. the H_∞ norm of the additive model error. Employing system identification methods, the provision of such an error

bound is achieved by using the prior that the available experimental data is contaminated by 'hard bounded' disturbance signals, as e.g. an ℓ_∞ bounded time or frequency domain signal. This type of prior implies that the identification procedure has to allow a worst case disturbance on the data at any time instant. As a result the error bounds that are obtained will be highly conservative if this worst case disturbance on the data is not actually present. Examples of approaches to this problem set-up which use the prior that the noise is bounded in the frequency domain can be found e.g. in Helmicki *et al.* [75], LaMaire *et al.* [98], and Gu and Khargonekar [62]. For an extensive discussion on these approaches we refer to Section 2.6. Examples of approaches to which use the prior that the noise is bounded in the time domain are e.g. Wahlberg and Ljung [182], Hakvoort [68], and many works in the area of bounded error and set membership identification, as e.g. Fogel and Huang [45], Norton [132], Milanese and Vicino [125]. A further discussion of these approaches is given in Section 2.5. The demerits of using hard prior bounds on the noise are discussed Section 3.4 and Section 3.3.

Experiences with real-life measurements show that quite often it can be motivated very well to describe disturbance signals on experimental data as realizations of stochastic processes, see also Goodwin *et al.* [54] and Ninness [130]. One of the typical advantages of this description is that it allows the disturbance signal to be characterized as a random (indifferent) signal, which is contaminating the data but which is not worst case at all time instants, i.e. it is not 'playing against' the experimenter, see Hjalmarsson [83] and Kosut [93]. For an extensive discussion on these issues we refer to Section 3.4. The assumption of stochastic noise disturbances on the data leads to identification procedures that provide soft (probabilistic) bounds on the model error, see e.g. Zhu [192], Zhu and Backx [194], Goodwin *et al.* [54], and Bayard [8]. This line of research is surveyed in Section 2.2, Section 2.3 and Section 2.4.

Clearly, the presence of noise disturbance on the experimental data is not the only cause of model uncertainty, as uncertainty is also implied by undermodelling and the effect of unknown past inputs (initial conditions) on the data. This has led to a further diversification of the field of model uncertainty estimation, see Section 3.5.

Also, when discussing the suitability of models as a basis for control system design, the availability of reliable error bounds certainly is important in order to obtain robust stability, and possibly also robust performance. However, it is the model that is used as a basis for the design, and a bad model will not give rise to a good controller. As a result, the identification of suitable models, apart from the quantification of their uncertainty, is an important issue in identification for control design, see e.g. [11, 155, 175, 103, 48] and the concise overview given in Section 2.8. A further discussion of the requirements imposed on model uncertainty estimation schemes when the intended use is high performance robust control design is given in Section 3.7, emphasizing that both the model as well as the model error bound should be suitable for robust control design, and that the user should be provided with tools to shape the error bound to be in accordance with robust control design specifications.

Our problem formulation will be the following. Let the data generating system be linear and time invariant, generating data sequences according to

$$y(t) = G_o(q)u(t) + v(t)$$

with $v(t)$ a realization of a stationary zero mean stochastic process with spectral density $\Phi_v(\omega)$, G_o an exponentially stable proper transfer function, and q the forward shift operator. Identify, on the basis of an experimental data sequence of input $u(t)$ and output $y(t)$, a finite dimensional linear time invariant model, together with a bound on the additive model error, where the model is suitable for control design, and where the error bound is frequency specific and distinguishes between the three sources of uncertainty: (1) undermodelling, (2) noise disturbance and (3) unknown past inputs affecting the data.

The distinction between the different sources of uncertainty is necessary in order to obtain information a posteriori about how to improve the identification result and the error bound in order to meet required performance specifications. When evaluating the estimated error bounds, different decisions now may be made so as to effectively improve the bounds, e.g. adjusting experimental conditions, adjusting model orders, and adjusting the basis generating system (see below).

Our approach to provide a solution to the problem mentioned above contains three characteristic features. The first one is that we will consider the model errors due to unmodelled dynamics and unknown past input signals as deterministic unknown-but-bounded (worst case) quantities, whereas the noise disturbance is considered to be stochastic (averaging). The rationale behind these choices is discussed extensively in Section 3.4, Section 3.5, Section 3.6 and Section 3.3. As a result, the error bounds that we will derive will have both soft and hard components. That is, the error bounds will hold with a probability level that is bounded from below. The second feature is that we will employ a periodic input signal and data segmentation, which allows us to clearly distinguish between disturbance effects which are supposed to average out, and structural effects as unmodelled dynamics which remain unchanged when experiments are repeated. The third feature is that the error bound can be specified almost directly with respect to a low order control relevant model specified by the user. That is, system based orthonormal basis functions, introduced by Heuberger [80] and discussed in Appendix E, are used for the parametric estimate. A control relevant model now can be used to generate the basis functions. When this basis generating model indeed provides good description of the measurement data, the model resulting from the model error estimation procedure will be close to the basis generating model specified by the user. Additionally, the basis generator can be used to incorporate prior knowledge about the system, which may be uncertain or approximative. For the importance of a model which is suitable for control design we refer to Section 3.7.

The combination of hard and soft components in the model error bounds constitutes the main deviation from existing methods, where either a worst case approach is taken,

leading to the before mentioned conservativeness, or a stochastic approach is taken, which makes it complex to handle bias due to undermodelling. The deterministic embedding of undermodelling errors and the stochastic embedding of noise errors is however well-established in the classical identification literature, see e.g. [105, 106, 20].

The use of system based orthonormal basis functions should ensure that the identification procedure yields an accurate model which is suitable for its intended use. In the literature regarding model error estimation, see Chapter 2, there is without exception a strong connection between the identification of the model and the quantification of its uncertainty. Thus, only identification methods can be selected for which error bounds indeed can be derived, which often means that error bounds are established at the expense of the accuracy of the estimate.

The approach that we take consists of the following three steps. First we will apply Bartlett's procedure of periodogram averaging (see e.g. [136]). Employing a periodic input signal and data segmentation, we will estimate an empirical transfer function estimate (ETFE) on each data segment. This set of estimates is used to obtain an averaged estimate, which is only defined in a finite number of frequency points, together with an estimate of its variance. An asymptotic error bound, which reflects the effects of unknown initial conditions and noise disturbances, while accounting for the fact that the variance is estimated, is established for the averaged estimate. In the second step of the procedure we will fit a parametric model to the frequency domain estimates obtained in the first step. System based orthonormal basis functions (see Appendix E) are used to construct the model set. In this second step a bias contribution due to undermodelling is introduced, but we can reduce the variance contribution (due to the noise) by a reduction in the number of parameters that has to be estimated from the data. For the merits of such a procedure to tradeoff bias and variance errors we refer to Section 3.8. In the third step, the set of parameter estimates is used to specify a set of transfer function estimates. An frequency specific asymptotic bound on the error in the averaged transfer function estimate will be established. This error bound is defined for each frequency $\omega \in [0, 2\pi)$, and can be separated into components which account for the different sources of uncertainty. Alternatively, the third step can be used to transform the set of parameter estimates to a set of undisturbed output signals which arise in response to an arbitrary input signal specified by the user, and an asymptotic bound on the estimation error again can be established. Similarly, an error bound on the discrete Fourier transform (DFT) of the disturbed output signal is obtained. The construction of these error bound requires only minor a priori information concerning the data generating process. An upper bound on past values of the input signal, on the system's impulse response, and on the generalized expansion coefficients is required a priori. However, procedures will be provided to accurately estimate the latter two bounds from the experimental data, whereas the first bound is given by the actuator constraints. The noise essentially only has to satisfy a mixing condition.

Comparing the above results to closely related previous work, we have the following

remarks. The classical results, see e.g. Ljung [105, 104, 106], Ljung and Yuan [113] and Brillinger [20], specify the asymptotic distribution of the error for parametric identification methods and spectral analysis. Although these results are excellent for analysis, they are less fit to actually bound the error. It is difficult to provide a good quantification of the bias due to undermodelling c.q. windowing, while the bias can be considerable. Only an implicit expression for the bias are available c.q. the expression for the bias contains the unknown first and second derivative of the system with respect to frequency, and the variance contribution contains the unknown spectrum of the noise. For a further discussion we refer to Section 2.2 and Section 2.3.

The results given in Zhu [192] and Zhu and Backx [194] are in line with the classical results as stated above. The error bounds are obtained by neglecting the bias contribution and the fact that the variance is estimated.

In Bayard [8] the noise is assumed to be normally distributed, the noise filter is assumed to be known, and the deviation from a steady-state situation is not taken into account. However, the results of [8] are nonasymptotic.

In Goodwin *et al.* [54] stochastic embedding is used for both the influence of the noise as well as for the error due to undermodelling. Both the distribution of the noise and the distribution of the undermodelling error are assumed to be known, up to a number of free parameters which are estimated from the data. The fact that these parameters are estimates of the true values is neglected. This leads to nonasymptotic error bounds. For a further discussion we refer to Section 2.4 and Section 3.5.

The work on hard error bounds is reviewed in Section 2.6 and Section 2.5. Hard prior bounds on the noise in the time or frequency domain are required as prior information, and the resulting models and/or model error bounds are not suitable for high performance robust control design.

For a further and more extensive discussion on the above topics we refer to Chapter 2 and Chapter 3.

The remainder of this chapter is organized as follows. Section 5.2 contains the assumptions and a priori information. In Section 5.3 an error bound for the nonparametric empirical transfer function estimate (ETFE) is formulated. Subsequently, in Section 5.4, a reduced order parametric model is fitted to the ETFE using system based orthonormal basis functions, and an error bound is established for the transfer function of this model using the results of Section 5.3. This error bound constitutes the main result of this chapter. In Section 5.4.5 and Section 5.4.6 error bounds for the response of the system to an arbitrary input signal are derived. A simulation example is presented in Section 5.7. A Monte Carlo analysis of the results in nonasymptotic situations is performed in Section 5.8. The behaviour of the error bounds under mild violations of the assumptions is studied on a simulation level in Section 5.9 and Section 5.10. Finally, Section 5.11 contains the conclusions.

A simplified version (Section 5.3 and Section 5.4 specialized to FIR models) of the

procedure presented in this chapter has been accepted for publication as [36].

5.2 Assumptions and prior information

It is assumed that the plant, and the measurement data that is obtained from this plant, allow a description

$$y(t) = G_o(q)u(t) + v(t) \quad (5.1)$$

with $y(t)$ the output signal, $u(t)$ a bounded deterministic input signal, $v(t)$ an additive noise, q the shift operator, and G_o the proper transfer function of the system, being finite dimensional, time invariant and exponentially stable. The transfer function can be written as

$$G_o(z) = \sum_{k=0}^{\infty} g_o(k)z^{-k} \quad (5.2)$$

$$= D + \sum_{k=0}^{\infty} L_k \mathcal{V}_k(z) \quad (5.3)$$

with $g_o(k)$ the impulse response of the plant, and L_k the generalized expansion coefficients, see (E.6). We will consider scalar (single input, single output) systems. The output disturbance $v(t)$ is represented as

$$v(t) = H_o(q)e(t) \quad (5.4)$$

where $e(t)$ is a sequence of independent identically distributed random variables, with zero mean values, variance σ_e^2 , and bounded moments of all orders, and where H_o is a stable proper transfer function.

We will denote the discrete time intervals for input and output signals by the integer intervals $T^N = \mathbb{Z} \cap [0, N - 1]$, $T_{N_s}^N = \mathbb{Z} \cap [N_s, N + N_s - 1]$ with N and N_s appropriate integers. We will frequently partition the time interval $T_{N_s}^N$ with $N = rN_o$ in r time intervals of length N_o , denoting $T_i = T_{(i-1)N_o + N_s}^{N_o}$, $i = 1, \dots, r$. The subscript i will indicate a variable that originates from the i -th time interval T_i , e.g.

$$x_i(t) := x(t + (i - 1)N_o + N_s) \quad \text{where} \quad t \in T^{N_o} \quad (5.5)$$

For a signal $x_i(t)$, defined on T_i , we will denote the discrete Fourier transform (DFT) by

$$X_i(e^{j\frac{2\pi k}{N_o}}) := \frac{1}{\sqrt{N_o}} \sum_{t=0}^{N_o-1} x_i(t) e^{-j\frac{2\pi k}{N_o}t} \quad \text{for} \quad k \in T^{N_o} \quad (5.6)$$

For a signal $x(t)$ being defined on $T_{N_s}^N$, we will denote

$$X^s(e^{j\frac{2\pi k}{N}}) := \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x(t + N_s) e^{-j\frac{2\pi k}{N}t} \quad \text{for} \quad k \in T^N \quad (5.7)$$

Some specific sets of frequencies that arise in the DFT are denoted as

$$\Omega_{N_o} := \left\{ \frac{2\pi k}{N_o}, k = 0, 1, \dots, N_o - 1 \right\} \quad (5.8)$$

$$\Omega_{N_o}^{u_i} := \{ \omega_k \in \Omega_{N_o} \mid |U_i(e^{j\omega_k})| > 0 \} \quad (5.9)$$

Finally we will denote

$$\bar{u} := \max_{t \in T^{N+N_s}} |u(t)| \quad (5.10)$$

Throughout this paper we will assume to have available the following a priori information on the past input signal and the system. Note that $L_k \in \mathbb{R}^{1 \times n_b}$ for all $k \in \mathbb{N}$, where n_b is the McMillan degree of the basis generating system, see Appendix E.1.

Assumption 5.2.1 *We have as a priori information that*

- i. *there exists a finite and known $\bar{u}^p \in \mathbb{R}$, such that $|u(t)| \leq \bar{u}^p$ for $t < 0$*
- ii. *there exist finite and known $M \in \mathbb{R}$ and $\rho \in \mathbb{R}$, with $\rho > 1$, such that $|g_o(k)| \leq M\rho^{-k}$ for $k \in \mathbb{N}$.*
- iii. *there exist finite and known $\mathcal{K} \in \mathbb{R}^{1 \times n_b}$ and $\eta \in \mathbb{R}$, with $\eta < 1$, such that $|L_k| \leq \mathcal{K}\eta^k$ for $k \in \mathbb{N}$.*

The a priori information on M , ρ , \mathcal{K} and η need not be tight in first instance. These priors can be estimated or improved using the measurement data, as will be discussed in Section 5.4.2 and Section 5.4.5. The a priori information on \bar{u}^p is given by the actuator constraints.

5.3 Error bound for ETFE

5.3.1 Introduction

In this section we will construct a nonparametric estimate \hat{G} of the system's transfer function G_o , by averaging over a set of empirical transfer function estimates (ETFEs). For this estimate, which is only defined at a finite number of frequency points, we will establish a confidence interval $\alpha(\omega_k)$ such that $|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \leq \alpha(\omega_k)$ with a prespecified probability. In order to achieve this, we will pursue the following strategy. Experimental data is available over a time set of length $N_s + N$. This time set is composed of a first interval of length N_s , not used for identification, and consecutively r intervals of length N_o . We consider an input signal that is periodic with period N_o , so that in each of the r intervals the same input signal is applied. Subsequently, an ETFE is made over each interval, and \hat{G} is obtained by averaging over these ETFEs. This procedure is similar to Bartlett's procedure of periodogram averaging, see [136, 20], and is also proposed by [105]. The choice to use a periodic input signal and averaging, and the significance of N_s , will be clarified at the end of Section 5.3.2.

Because of the periodicity of the input signal $u(t)$ it follows that

$$\Omega_{N_o}^{u_i} = \Omega_{N_o}^u \quad \text{for all } i = 1, \dots, r \quad (5.11)$$

5.3.2 Nonparametric transfer function estimate

Define the following estimates

$$\hat{G}_i(e^{j\omega_k}) := \frac{Y_i(e^{j\omega_k})}{U_i(e^{j\omega_k})} \quad \text{for } i = 1, 2, \dots, r \quad \omega_k \in \Omega_{N_o}^u \quad (5.12)$$

$$\hat{G}(e^{j\omega_k}) := \frac{1}{r} \sum_{i=1}^r \hat{G}_i(e^{j\omega_k}) \quad (5.13)$$

Employing the system's equations, similar as [106], results in

$$\hat{G}_i(e^{j\omega_k}) = G_o(e^{j\omega_k}) + S_i(e^{j\omega_k}) + \frac{V_i(e^{j\omega_k})}{U_i(e^{j\omega_k})} \quad (5.14)$$

where

$$S_i(e^{j\omega_k}) = \frac{R_i(e^{j\omega_k})}{U_i(e^{j\omega_k})} \quad (5.15)$$

with $R_i(e^{j\omega_k})$ a component which is due to unknown past inputs, i.e. input samples preceding the time interval under consideration. We will split the analysis into two parts, separating bias and variance issues, by defining the auxiliary variables $\tilde{G}_i(e^{j\omega_k})$ and their average $\tilde{G}(e^{j\omega_k})$ as

$$\tilde{G}_i(e^{j\omega_k}) := \hat{G}_i(e^{j\omega_k}) - S_i(e^{j\omega_k}) = G_o(e^{j\omega_k}) + \frac{V_i(e^{j\omega_k})}{U_i(e^{j\omega_k})} \quad (5.16)$$

$$\tilde{G}(e^{j\omega_k}) := \frac{1}{r} \sum_{i=1}^r \tilde{G}_i(e^{j\omega_k}) \quad \text{for } \omega_k \in \Omega_{N_o}^u \quad (5.17)$$

Using the triangle inequality we find

$$|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \leq |G_o(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})| + |S(e^{j\omega_k})| \quad (5.18)$$

with

$$S(e^{j\omega_k}) := \hat{G}(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k}) = \frac{1}{r} \sum_{i=1}^r S_i(e^{j\omega_k}) \quad (5.19)$$

Considering the inequality (5.18), the first term on the right hand side is the variance contribution to the error, which is due to the noise disturbance on the data. The second term on the right hand side of (5.18) is a bias contribution, which is due to unknown past inputs. A confidence interval for the first (stochastic) term has to be determined on the basis of its distribution, whereas a hard bound will be derived for the second (deterministic) term.

For the bias contributions $S(e^{j\omega_k})$ and $S_i(e^{j\omega_k})$, which will be of interest in the sequel of the paper, we have the following results.

Lemma 5.3.1 Consider the variables $S_i(e^{j\omega_k})$, $S(e^{j\omega_k})$ as defined in (5.15), (5.19). Let the input signal be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$, $r \geq 1$. Then, for all $\omega_k \in \Omega_{N_o}^u$ and $i = 1, \dots, r$

$$|S_i(e^{j\omega_k})| \leq \frac{1}{\sqrt{N_o}} \frac{\bar{u}^P + \bar{u}}{|U_i(e^{j\omega_k})|} \frac{M\rho(1 - \rho^{-N_o})}{(\rho - 1)^2} \rho^{-(i-1)N_o - N_s} \quad (5.20)$$

$$|S(e^{j\omega_k})| \leq \frac{1}{r\sqrt{N_o}} \frac{\bar{u}^P + \bar{u}}{|U_i(e^{j\omega_k})|} \frac{M\rho(1 - \rho^{-N})}{(\rho - 1)^2} \rho^{-N_s} \quad (5.21)$$

Proof: See Appendix B.1. \square

Upper bounds on the bias contribution $S(e^{j\omega_k})$ in situations of general nonperiodic input signals can be found in Ljung [106] and LaMaire *et al.* [98].

Establishing the distribution of the variance contribution $|G_o(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|$ is more involved, and only results are available which are asymptotic in N_o . Because the difference $G_o(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})$ is entirely due to the noise, see (5.17) and (5.16), a first essential step is to find the asymptotic distribution of the DFT of the noise. This has been achieved by Brillinger [20].

Theorem 5.3.2 Brillinger [20]

Consider $v_i(t)$ as defined in (5.4), (5.5), and let $V_i(e^{j\omega_k})$ be the DFT of $v_i(t)$ as defined in (5.6) with $\omega_k \in \Omega_{N_o}$. Let

$$\check{V}_{N_o} = \begin{bmatrix} \operatorname{Re}\{V_i(e^{j\omega_k})\} \\ \operatorname{Im}\{V_i(e^{j\omega_k})\} \\ \operatorname{Re}\{V_i(e^{j\omega_\ell})\} \\ \operatorname{Im}\{V_i(e^{j\omega_\ell})\} \\ \operatorname{Re}\{V_m(e^{j\omega_\ell})\} \\ \operatorname{Im}\{V_m(e^{j\omega_\ell})\} \end{bmatrix}$$

where $\omega_k, \omega_\ell \in \Omega_{N_o}$, $\omega_k \neq \omega_\ell$, and $i, m = 1, \dots, r$, $i \neq m$. Then, as $N_o \rightarrow \infty$

$$\check{V}_{N_o} \xrightarrow{d} \mathcal{N}(0, \Lambda)$$

meaning that \check{V}_{N_o} converges in distribution to the normal distribution with zero mean and covariance matrix Λ , where Λ is a diagonal matrix with diagonal elements given by

$$\begin{aligned} \operatorname{var}[\operatorname{Re}\{V_q(e^{j\omega_p})\}] &= \operatorname{var}[\operatorname{Im}\{V_q(e^{j\omega_p})\}] = \frac{1}{2}\sigma_e^2 |H_o(e^{j\omega_p})|^2 & \text{for } \omega_p \neq 0, \pi \\ \operatorname{var}[\operatorname{Re}\{V_q(e^{j\omega_p})\}] &= \sigma_e^2 |H_o(e^{j\omega_p})|^2 & \text{for } \omega_p = 0, \pi \\ \operatorname{var}[\operatorname{Im}\{V_q(e^{j\omega_p})\}] &= 0 & \text{for } \omega_p = 0, \pi \end{aligned}$$

for $\omega_p \in \Omega_{N_o}$ and $q = 1, \dots, r$.

Theorem 5.3.2 states that the elements of the vector \check{V}_{N_o} are asymptotically uncorrelated and jointly normally distributed, which implies that they are asymptotically independent, see e.g. [147]. Because U_i is independent of i due to the periodic input, it follows from Theorem 5.3.2 that asymptotically the auxiliary variables $\tilde{G}_i(e^{j\omega_k})$ are independent and identically normally distributed. Considering (5.16) it now follows that $\tilde{G}(e^{j\omega_k})$ is asymptotically normally distributed with expectation $G_o(e^{j\omega_k})$. However, the variance of $\tilde{G}(e^{j\omega_k})$ is still unknown. We will quantify this variance on the basis of measurement data. To this end we define the estimate

$$\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k})) = \frac{1}{r(r-1)} \sum_{i=1}^r |\tilde{G}(e^{j\omega_k}) - \tilde{G}_i(e^{j\omega_k})|^2 \quad (5.22)$$

Although this estimate is not available from the data, as the auxiliary variables $\tilde{G}(e^{j\omega_k})$ and $\tilde{G}_i(e^{j\omega_k})$ are unknown, we can bound its difference with the estimate

$$\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k})) = \frac{1}{r(r-1)} \sum_{i=1}^r |\hat{G}(e^{j\omega_k}) - \hat{G}_i(e^{j\omega_k})|^2 \quad (5.23)$$

as formulated in the following lemma.

Lemma 5.3.3 *Consider the estimates $\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k}))$, $\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))$ as defined in (5.23), (5.22). Let $u(t)$ be a periodic input signal with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, and let $\omega_k \in \Omega_{N_o}^u$. Then*

$$|\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k})) - \hat{\sigma}_r^2(\hat{G}(e^{j\omega_k}))| \leq \mathcal{S}(\omega_k)$$

with

$$\mathcal{S}(\omega_k) = \frac{1}{r(r-1)} \sum_{i=1}^r (2|A_i(e^{j\omega_k})||B_i(e^{j\omega_k})| + |B_i(e^{j\omega_k})|^2) \quad (5.24)$$

and

$$\begin{aligned} |A_i(e^{j\omega_k})| &= |\hat{G}(e^{j\omega_k}) - \hat{G}_i(e^{j\omega_k})| \\ |B_i(e^{j\omega_k})| &= \frac{r-2}{r} |S_i(e^{j\omega_k})| + \frac{1}{r} \sum_{m=1}^r |S_m(e^{j\omega_k})| \end{aligned}$$

where $S_i(e^{j\omega_k})$, $S_m(e^{j\omega_k})$ are bounded by (5.20).

Proof: See Appendix B.2. □

Note that the difference between the random variables (5.23) and (5.22) contains known realizations of random variables only, i.e realizations of \tilde{G}_i , \hat{G} , and therefore can be bounded with a hard error bound.

Using Theorem 5.3.2 and the auxiliary estimate $\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))$, the distribution of the difference between the auxiliary transfer function $\tilde{G}(e^{j\omega_k})$ and the system's transfer function $G_o(e^{j\omega_k})$ now can be completely specified asymptotically in N_o .

Lemma 5.3.4 Consider the auxiliary transfer function $\tilde{G}(e^{j\omega_k})$, (5.17), (5.16), and the estimate of its variance (5.22). Let the input signal be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, and let $\omega_k \in \Omega_{N_o}^u$. Then, as $N_o \rightarrow \infty$

$$\frac{|G_o(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))} \xrightarrow{d} \begin{cases} F(2, 2(r-1)) & \omega_k \neq 0, \pi \\ F(1, r-1) & \omega_k = 0, \pi \end{cases}$$

where $F(n, d)$ denotes the F distribution with n degrees of freedom in the numerator and d degrees of freedom in the denominator.

Proof: See Appendix B.3. □

Combining lemma's 5.3.1, 5.3.3 and 5.3.4 and (5.18) leads to an error bound in terms of a confidence interval, that can be calculated on the basis of measurement data. In formulating this confidence interval, we will adopt the following notation

$$F_\alpha(m, n) = \{P[x \leq \alpha], x \in F(m, n)\} \quad (5.25)$$

meaning that a random variable x which is distributed as $F(m, n)$ is smaller than α with probability (w.p.) given by $F_\alpha(m, n)$.

Theorem 5.3.5 Consider the transfer function $\hat{G}(e^{j\omega_k})$, (5.13), (5.12), and the estimate of its variance (5.23). Let the input signal be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, and let $\omega_k \in \Omega_{N_o}^u$. Then, asymptotically in N_o

$$|G_o(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \leq |S(e^{j\omega_k})| + \gamma_\alpha(\omega_k) \quad w.p. \geq \begin{cases} F_\alpha(2, 2(r-1)) & \omega_k \neq 0, \pi \\ F_\alpha(1, r-1) & \omega_k = 0, \pi \end{cases}$$

where $S(e^{j\omega_k})$ is bounded by (5.21),

$$\gamma_\alpha(\omega_k) = \sqrt{\alpha} \left(\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k})) + \mathcal{S}(\omega_k) \right)^{\frac{1}{2}} \quad (5.26)$$

and $\mathcal{S}(\omega_k)$ is given by (5.24).

Proof: See Appendix B.4. □

Further analysis of the properties of the estimates being used gives rise to the following remarks.

- The estimate $\hat{G}(e^{j\omega_k})$ is a consistent estimate of $G_o(e^{j\omega_k})$; its bias decays as ρ^{-N_s}/\sqrt{rN} , see (5.21), and its variance decays as $1/r$, see (5.28). Asymptotically in N_o the estimate $\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k}))$ is a consistent estimate of $\text{var}[\hat{G}(e^{j\omega_k})]$; asymptotically in N_o the estimate is unbiased and its variance

decays as $1/(r-1)$, see (B.6) and (B.5) in Appendix B.3.

Thus, for the components of the error bound given in Theorem 5.3.5 there holds

$$\begin{aligned} |S(e^{j\omega_k})| &\rightarrow 0 \quad \text{as } \rho^{-N_s}/\sqrt{rN} \\ \mathcal{S}(\omega_k) &\rightarrow 0 \quad \text{as } \rho^{-N_s}/(r\sqrt{rN}) \\ \hat{\sigma}_r^2(\hat{G}(e^{j\omega_k})) &\rightarrow \text{var}[\hat{G}(e^{j\omega_k})] \quad \text{as } 1/(r-1) \text{ asymptotically in } N_o \\ \text{var}[\hat{G}(e^{j\omega_k})] &\rightarrow 0 \quad \text{as } 1/r \end{aligned} \quad (5.27)$$

- The variance of $\hat{G}(e^{j\omega_k})$ is given as the noise to signal ratio in the frequency domain

$$\text{var}[\hat{G}(e^{j\omega_k})] = \frac{\text{var}[V^s(e^{j\omega_k})]}{|U^s(e^{j\omega_k})|^2} = \frac{1}{r} \frac{\text{var}[V^s(e^{j\omega_k})]}{|U_i(e^{j\omega_k})|^2} \quad (5.28)$$

see (5.13), (5.12), (5.7), (B.1), where

$$\text{var}[V^s(e^{j\omega_k})] \rightarrow \Phi_v(\omega_k) = \sigma_e^2 |H_o(e^{j\omega_k})|^2 \quad \text{as } 1/N$$

see [105].

- The estimate $\hat{G}(e^{j\omega_k})$ is only defined at the finite number of frequency points in $\Omega_{N_o}^u$.

Summarizing, the periodic input yields an estimate having a variance that quickly decays with the number of averages, and enables a considerable reduction of the bias due to unknown past input signals by using N_s . Furthermore, the periodic input offers the opportunity to completely characterize the asymptotic distribution of the estimate, i.e. including the fact that an estimate of the variance is used, without heavily relying on a priori knowledge of the noise, see Lemma 5.3.4.

The advantage of the averaging (5.13), in conjunction with a periodic input signal, over windowing techniques which average over neighbouring frequency points, is that an estimate of the true transfer function with decaying variance is obtained without introducing bias due to windowing. However, averaging reduces the frequency resolution: an estimate is obtained at only N_o instead of N frequency points.

5.3.3 Input design

In view of experiment design, the error bound of Theorem 5.3.5 has the following implications. The bias contributions $|S(e^{j\omega_k})|$ and $\mathcal{S}(\omega_k)$ can be reduced most effectively by increasing N_s , see (5.21), (5.24), (5.20). Considering the estimated variance $\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k}))$ it follows from (5.27) and (5.28) that $|U_i(e^{j\omega_k})|$ should not be too small at any frequency $\omega_k \in \Omega_{N_o}^u$. To arrive at an error bound which is less than some prespecified frequency domain bound $K(\omega_k)$, resulting e.g. from the robustness properties of a controller that has been designed for the system, the input $|U_i(e^{j\omega_k})|$ should be chosen to be proportional to $\sqrt{\text{var}[V^s(e^{j\omega_k})]}/K(\omega_k)$ for all $\omega_k \in \Omega_{N_o}^u$. Such an input, having

a small time domain magnitude \bar{u} in addition, can be designed using the results of e.g. [158, 151, 152]. Filtered white noise or pseudo random binary sequences (PRBS) are not adequate input signals: $|U_i(e^{j\omega_k})|^2$ will be an erratic function of frequency, see [106, lemma 6.2], resulting in an erratic estimate and error bound.

5.4 Error bound for generalized FIR estimate

5.4.1 Introduction

In the previous section we have obtained a sequence of nonparametric estimates $\hat{G}_i(e^{j\omega_k})$, $\omega_k \in \Omega_{N_o}^u$, $i = 1, \dots, r$, together with a bound on the error in the averaged estimate $\hat{G}(e^{j\omega_k})$. This implies that we only have information in a finite number of points on the frequency axis. In this section we will use the frequency domain data, as obtained in the previous section, to provide a reduced order parametric model together with a model error bound that is defined on the whole frequency interval $\omega \in [0, 2\pi)$. In view of robust control design, based on an estimated model of the system, a model error bound that is defined for all $\omega \in [0, 2\pi)$ is essential in order to be able to reliably predict stability and performance properties of the closed loop consisting of the true system and the designed controller.

The parametric identification introduces bias due to undermodelling. However, the variance can be reduced by restricting the number of estimated parameters. Hence, by choosing the model order, bias can be traded against variance, allowing for a lower error bound. We will derive explicit bounds for the bias and variance errors, enabling a clear tradeoff between these terms.

We focus on parametric methods instead of spectral analysis techniques (windowing) to improve our results since Ljung [105] has shown that parametric methods allow for a faster convergence to zero of the mean square model error, due to the fact that specifically tailored model sets can be chosen when using parametric identification techniques.

For the parametric identification we will use system based orthonormal basis functions, which were introduced by [80]. That is, we will employ a linear combination of a finite number of these generalized orthonormal basis functions to model the system. We will refer to such a model as a generalized FIR model. These generalized basis functions include the well-known pulse functions (resulting in a FIR model), Laguerre functions and Kautz functions (see [180]). System based orthonormal basis functions have a number of very valuable properties. In Appendix E the theory and the merits of system based orthonormal basis functions are discussed, and two examples are given. The main advantage is that the basis generator can be used to incorporate prior information about the system. This information may be uncertain or approximative. However, the more accurate the basis generating system, the more simple the system representation will be. To be more precise, a basis generator whose poles are closer to

those of the system yields a higher rate of convergence of the parameters L_k to zero, i.e. a higher η in $|L_k| \leq \mathcal{K}\eta^k$. The basis generator can e.g. be a nominal model of the system, obtained by any identification technique or by physical modelling. A control relevant nominal model therefore can be used as the basis generating system. Control relevant identification procedures are discussed in [11, 155, 175, 103, 48], see also the concise overview given in Section 2.8.

First, in Section 5.4.2, we will estimate the parameters of a generalized FIR model, by using the set of nonparametric frequency domain estimates $\hat{G}_i(e^{j\omega_k})$, and an error bound for the estimated parameters will be derived. Based on these results, in Section 5.4.3 an error bound will be established for the transfer function corresponding to the estimated generalized FIR model. Additionally, in Section 5.4.5 and Section 5.4.6 the estimated generalized FIR model will be used to obtain an estimate, together with an error bound, of the undisturbed and disturbed output of the system in response to an arbitrary input signal, respectively.

The presentation will be short; the methods used are similar to those of Section 5.3.

5.4.2 Parameter estimate

We will consider a generalized FIR model structure, of order $n_p - 1$, determined by

$$\begin{aligned} G(e^{j\omega}, \theta) &= D + \sum_{k=0}^{\bar{n}_p-1} L_k \mathcal{V}_k(e^{j\omega}) \\ &= \vartheta(e^{j\omega})\theta \end{aligned} \quad (5.29)$$

with

$$\begin{aligned} \bar{n}_p &= \frac{n_p - 1}{n_b} \\ \theta &= [D \ L_0 \ \dots \ L_{\bar{n}_p-1}]^T \end{aligned} \quad (5.30)$$

$$\vartheta(e^{j\omega}) = [1 \ \mathcal{V}_0^T(e^{j\omega}) \ \dots \ \mathcal{V}_{\bar{n}_p-1}^T(e^{j\omega})] \quad (5.31)$$

where the superscript T denotes the transpose, and n_b is the McMillan degree of the basis generating system, see Appendix E.1. Define

$$\begin{aligned} \mathcal{U} &= [\vartheta^T(e^{j\zeta_1}) \ \vartheta^T(e^{j\zeta_2}) \ \dots \ \vartheta^T(e^{j\zeta_{N_p}})]^T \\ \hat{\mathcal{G}}_i &= [\hat{G}_i(e^{j\zeta_1}) \ \hat{G}_i(e^{j\zeta_2}) \ \dots \ \hat{G}_i(e^{j\zeta_{N_p}})]^T \end{aligned} \quad (5.32)$$

where $\hat{G}_i(e^{j\zeta_k})$ is defined in (5.12), and where the frequency points ζ_k , $k = 1, \dots, N_p$, constitute a set $\Omega^p \subseteq \Omega_{N_o}^u$, satisfying that the $e^{j\zeta_k}$ come in complex conjugate pairs.

Now for each time interval T_i , $i = 1, \dots, r$, an estimate of the parameters of the generalized FIR model (5.29) is obtained by the following weighted least squares criterion

$$\hat{\theta}_i := \arg \min_{\theta} \|W(\hat{\mathcal{G}}_i - \mathcal{U}\theta)\|_2^2 \quad (5.33)$$

where θ ranges over an appropriate parameter space $\Theta \subset \mathbb{R}^{n_p}$, and $W \in \mathbb{C}^{N_p \times N_p}$ is a weighting matrix. This weighting matrix, which will be discussed in more detail at the end of this section (page 89), can be used to minimize the variance of the estimated parameters, or to affect the bias distribution over frequency.

Similar to the previous section, the final estimate is obtained by averaging over the different sections of the data

$$\hat{\theta} := \frac{1}{r} \sum_{i=1}^r \hat{\theta}_i \quad (5.34)$$

We will now set out to find an expression for the difference $\hat{\theta} - \theta_o$. As an alternative expression for (5.33) we can write

$$\hat{\theta}_i = \Psi \hat{G}_i \quad (5.35)$$

with

$$\Psi = (\mathcal{U}^* W^* W \mathcal{U})^{-1} \mathcal{U}^* W^* W \quad (5.36)$$

where the superscript * denotes the complex conjugate transpose. From (5.14) it follows that

$$\hat{G}_i(e^{j\omega_k}) = \mathcal{V}(e^{j\omega_k})\theta_o + Z(e^{j\omega_k}) + S_i(e^{j\omega_k}) + F_i(e^{j\omega_k}) \quad (5.37)$$

for $\omega_k \in \Omega_{N_o}^u$, with

$$\theta_o = [D \ L_0 \ \cdots \ L_{\bar{n}_p-1}]^T \quad (5.38)$$

$$Z(e^{j\omega_k}) = \sum_{\ell=\bar{n}_p}^{\infty} L_\ell \mathcal{V}_\ell(e^{j\omega_k}) \quad (5.39)$$

$$F_i(e^{j\omega_k}) = \frac{V_i(e^{j\omega_k})}{U_i(e^{j\omega_k})} \quad (5.40)$$

where $Z(e^{j\omega_k})$ is the error due to undermodelling, $S_i(e^{j\omega_k})$ is the error due to the unknown past inputs (5.14), and $F_i(e^{j\omega_k})$ is the error due to the noise. Using (5.37) and (5.35) we now find that

$$\hat{\theta}_i - \theta_o = \Psi Z + \Psi S_i + \Psi F_i \quad (5.41)$$

with

$$Z = [Z(e^{j\zeta_1}) \ Z(e^{j\zeta_2}) \ \cdots \ Z(e^{j\zeta_{N_p}})]^T \quad (5.42)$$

$$S_i = [S_i(e^{j\zeta_1}) \ S_i(e^{j\zeta_2}) \ \cdots \ S_i(e^{j\zeta_{N_p}})]^T \quad (5.43)$$

$$F_i = [F_i(e^{j\zeta_1}) \ F_i(e^{j\zeta_2}) \ \cdots \ F_i(e^{j\zeta_{N_p}})]^T \quad (5.44)$$

Hence, using (5.34) and (5.19), it follows that

$$\begin{aligned} \hat{\theta} - \theta_o &= \Psi Z + \Psi S + \frac{1}{r} \sum_{i=1}^r \Psi F_i \\ &= z^p + s^p + \frac{1}{r} \sum_{i=1}^r \Psi F_i \end{aligned} \quad (5.45)$$

with

$$\begin{aligned} S &= [S(e^{j\zeta_1}) \ S(e^{j\zeta_2}) \ \dots \ S(e^{j\zeta_{N_p}})]^T \\ z^p &= \Psi Z \\ s^p &= \Psi S \end{aligned} \quad (5.46)$$

To separate bias and variance issues, similar to the technique presented in the previous section, see (5.16), we introduce the auxiliary variables $\tilde{\theta}_i$ and $\tilde{\theta}$

$$\begin{aligned} \tilde{\theta}_i &:= \theta_o + \Psi F_i \\ \tilde{\theta} &:= \frac{1}{r} \sum_{i=1}^r \tilde{\theta}_i \end{aligned}$$

Furthermore, we will use an estimate of the variance of the estimated parameters, analogous to (5.23). To be able to formulate this estimate, we have to introduce some additional notation.

Let $x(m)$ denote the m -th element of a vector x and $X(m, i)$ the m, i -th element of a matrix X .

The estimated variance of $\hat{\theta}\langle k \rangle$ now is defined for $k \in \mathbb{Z} \cap [1, n_p]$ as

$$\hat{\sigma}_r^2(\hat{\theta}\langle k \rangle) := \frac{1}{r(r-1)} \sum_{i=1}^r |\hat{\theta}\langle k \rangle - \hat{\theta}_i\langle k \rangle|^2 \quad (5.47)$$

In line with Lemma 5.3.3, we can bound the difference between the estimated variances of the parameter estimate $\hat{\theta}$ and the unknown auxiliary variable $\tilde{\theta}$, as formulated in the following lemma.

Lemma 5.4.1 *Consider the estimates $\hat{\sigma}_r^2(\hat{\theta}\langle k \rangle)$, $\hat{\sigma}_r^2(\tilde{\theta}\langle k \rangle)$ as defined in (5.47). Let the input signal be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, and let $k \in \mathbb{Z} \cap [1, n_p]$. Then*

$$|\hat{\sigma}_r^2(\tilde{\theta}\langle k \rangle) - \hat{\sigma}_r^2(\hat{\theta}\langle k \rangle)| \leq \mathcal{S}^p\langle k \rangle$$

with

$$\mathcal{S}^p\langle k \rangle = \frac{1}{r(r-1)} \sum_{i=1}^r (2 |A_i^p\langle k \rangle| |B_i^p\langle k \rangle| + |B_i^p\langle k \rangle|^2) \quad (5.48)$$

and

$$\begin{aligned} |A_i^p\langle k \rangle| &= |\hat{\theta}\langle k \rangle - \hat{\theta}_i\langle k \rangle| \\ |B_i^p\langle k \rangle| &\leq \frac{1}{r} \sum_{m=1}^r |s_m^p\langle k \rangle| + \frac{r-2}{r} |s_i^p\langle k \rangle| \\ |s_i^p\langle k \rangle| &\leq \sum_{m=1}^{N_p} |\Psi\langle k, m \rangle| |S_i(e^{j\zeta_m})| \end{aligned}$$

where $S_i(e^{j\zeta_m})$ is bounded by (5.20).

Proof: The proof follows along exactly the same lines as the proof of Lemma 5.3.3. \square

We are now ready to formulate an error bound on the estimated parameters.

Theorem 5.4.2 *Consider the estimated parameters $\hat{\theta}$, (5.34), (5.35), and the estimated variance (5.47). Let the input signal be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, let $n_p \leq N_p \leq N_o$, and let $m \in \mathbb{Z} \cap [1, n_p]$. Then, asymptotically in N_o*

$$|\theta_o \langle m \rangle - \hat{\theta} \langle m \rangle| \leq |s^p \langle m \rangle| + |z^p \langle m \rangle| + \gamma_\alpha^p \langle m \rangle \quad w.p. \geq F_\alpha(1, r-1)$$

with

$$\begin{aligned} |s^p \langle m \rangle| &\leq \sum_{\ell=1}^{N_p} |\Psi \langle m, \ell \rangle| |S(e^{j\zeta_\ell})| \\ |z^p \langle m \rangle| &\leq \mathcal{K} \sum_{k=\bar{n}_p}^{\bar{n}_h-1} \eta^k \left| \sum_{\ell=1}^{N_p} \Psi \langle k, \ell \rangle \mathcal{V}_k(e^{j\zeta_\ell}) \right| + \mathcal{K} \frac{\eta^{\bar{n}_h}}{1-\eta} \sum_{\ell=1}^{N_p} |\Psi \langle k, \ell \rangle| |\mathcal{V}_0(e^{j\zeta_\ell})| \\ \gamma_\alpha^p \langle m \rangle &\leq \sqrt{\alpha} \left(\hat{\sigma}_r^2(\hat{\theta} \langle m \rangle) + \mathcal{S}^p \langle m \rangle \right)^{\frac{1}{2}} \end{aligned}$$

for any $\bar{n}_h \in \mathbb{N}$, $\bar{n}_h \geq \bar{n}_p$, while $\mathcal{S}^p \langle m \rangle$ is given by (5.48), and $S(e^{j\zeta_\ell})$ is bounded by (5.21).

Proof: See Appendix B.5. \square

The bound on the parameter error formulated in Theorem 5.4.2 consists of three separate terms. The first term s^p reflects the error due to unknown past inputs, the second term z^p refers to the effect of unmodelled dynamics, whereas the third (probabilistic) term γ_α^p represents the fact that the measurements are contaminated by noise.

A standard result from statistics is that the variance of the estimated parameters $\hat{\theta}$ is minimized by choosing the matrix W^*W to be proportional to the inverse covariance matrix of the estimate $\hat{G}(e^{j\omega_k})$. Asymptotically in N this covariance matrix is proportional to a diagonal matrix with diagonal elements $\text{var}[V_i(e^{j\omega_k})]/|U_i(e^{j\omega_k})|^2$, $\omega_k = \zeta_1, \dots, \zeta_{N_p}$, see (5.28) and Theorem 5.3.2. A good choice for W therefore is $W = \text{diag}(1/\hat{\sigma}_r(\hat{G}(e^{j\zeta_1})) \dots 1/\hat{\sigma}_r(\hat{G}(e^{j\zeta_{N_p}})))$. The weighting matrix W of course also can be used to affect the bias distribution over frequency of the estimated generalized FIR model, see [105].

Comparing the above result with the results for parametric identification as obtained by [105] and [106, appendix II.2] we have the following four remarks. Firstly, we have established an expression for the distribution of the error that can be calculated from the data, also if the noise is colored. Secondly, the bias due to undermodelling has

been explicitly bounded. Thirdly, we have obtained an estimate of the variance of the estimated parameters (5.47) that is consistent in the presence of undermodelling. That is, the estimated variance converges to the true variance as N_o and r go to infinity. Recently, consistent estimates for the variance of a parameter estimate have also been established by Hjalmarsson [85, 83], considering time domain prediction error methods. Finally, a consistent estimate of the covariance matrix of the nonparametric estimate $\hat{G}(e^{j\omega_k})$ is available, and this estimate is independent of the parametric model. A good approximation of the best linear unbiased estimate (BLUE) therefore can be obtained by choosing W as mentioned above.

Because the a priori information \mathcal{K} and η on the generalized expansion coefficients of the system largely determines the bound on the error due to undermodelling $z^p(m)$, a correct and not overly conservative choice for \mathcal{K} and η is of major importance. This also holds for the frequency domain error bound on the transfer function of the estimated generalized FIR model as given in Theorem 5.4.4 and Theorem 5.4.5 in the next section. However, obtaining the a priori information \mathcal{K} and η is largely an open issue in model uncertainty estimation. Helmicki *et al.* [79] state that in practical applications such a priori information typically is obtained through some 'engineering leap of faith'. An important contribution of Theorem 5.4.2 is that it provides a simple and effective way to validate the prior information \mathcal{K} and η : the prior information should be consistent with the resulting error bound on the estimated coefficients (the estimated parameters). This implies that the prior information \mathcal{K} and η can be improved iteratively by adapting \mathcal{K} and η to match the resulting error bounds on the estimated coefficients. Clearly, such a procedure gives no guarantee that the prior information given by \mathcal{K} and η is correct, but it significantly reduces the 'engineering leap of faith' involved.

When evaluating the prior information, one should consider the following. The true parameters $\theta_o(k)$ (generalized expansion coefficients) will converge to zero with k when the system and basis generator are exponentially stable. However, the estimated parameters will only converge to a zero mean stochastic process, due to noise disturbances. Thus, the size of the confidence interval will not converge to zero, but the confidence interval will become centered around zero. Therefore, the prior information should be required to encompass the confidence interval for $k < k_c$ only, while allowing it to cross the confidence interval for $k \geq k_c$, where k_c is such that zero is contained in the confidence interval for $k \geq k_c$. Also, we would like to stress that it is important that the tails of the confidence intervals indeed contain zero. When this is not the case, it is not possible to estimate the prior information reliably, since it is not known where the prior information is allowed to cross the confidence interval.

The above observations emphasize the fact that a confidence interval is indispensable to reliably estimate values for \mathcal{K} and η . When a confidence interval is not available, it is not known whether the deviations of the estimated parameters from zero are due to undermodelling or are likely to be due to the noise.

A procedure similar to the procedure proposed above to estimate \mathcal{K} and η , can be applied to estimate M and ρ , using the results of Section 5.4.5.

The only remaining prior information now is the upper bound on the past inputs \bar{u}^p (see Assumption 5.2.1.i). This bound can be obtained from the actuator constraints. If the error due to past inputs is indeed very small with respect to the other error components (as should be the case), the error bounds are not at all sensitive to the actual value of \bar{u}^p (or the values of M and ρ). Thus, an incorrect or very conservative choice for \bar{u}^p will hardly affect the error bounds in this case.

5.4.3 Parametric transfer function estimate

Using the estimated parameters of the generalized FIR model an estimate of the transfer function with an error bound can be obtained for any frequency $\omega \in [0, 2\pi)$. The estimate of the transfer function at ω is, for any $\omega \in [0, 2\pi)$, defined as

$$\hat{G}_i^f(e^{j\omega}) := P(e^{j\omega}) [\mathcal{W} \ 0] \hat{\theta}_i \quad (5.49)$$

where

$$P(e^{j\omega}) = [1 \ \mathcal{V}_0^T(e^{j\omega}) \ \dots \ \mathcal{V}_{\bar{n}_f-1}^T(e^{j\omega})]$$

and $\mathcal{W} \in \mathbb{C}^{n_f \times n_f}$ is a weighting matrix, with $n_f = \bar{n}_f n_b + 1$, where n_b is the McMillan degree of the basis generating system. This weighting matrix, which will be discussed in more detail later at the end of this section (page 95), can be seen as a regularization of the estimated parameters towards agreement with a prior knowledge such as e.g. an exponentially decaying sequence.

Similar to the previous sections, the final estimate is obtained by averaging over the different sections of the data.

$$\hat{G}^f(e^{j\omega}) := \frac{1}{r} \sum_{i=1}^r \hat{G}_i^f(e^{j\omega}) \quad (5.50)$$

We will now set out to find an expression for the difference $G_o - \hat{G}^f$. From (5.41) we find that

$$G_o - \hat{G}_i^f = \Gamma + \Lambda - Z^f - S_i^f - \Upsilon F_i \quad (5.51)$$

with

$$\begin{aligned} \Gamma &= P(I - \mathcal{W}) \theta_o \\ \Lambda &= \sum_{k=\bar{n}_f}^{\infty} L_k \mathcal{V}_k \\ Z^f &= \Upsilon Z \\ S_i^f &= \Upsilon S_i \\ \Upsilon(e^{j\omega}) &= P(e^{j\omega}) [\mathcal{W} \ 0] \Psi \end{aligned} \quad (5.52)$$

We will denote

$$\mathcal{P} = [\mathcal{P}_D \ \mathcal{P}_0 \ \cdots \ \mathcal{P}_{\bar{n}_f-1}] := P(I - W) \quad (5.53)$$

with the appropriate dimensions for \mathcal{P}_D and \mathcal{P}_k , i.e. $\mathcal{P}_D \in \mathbb{C}$ and $\mathcal{P}_k \in \mathbb{C}^{1 \times n_b}$ for $k = 0, \dots, \bar{n}_f - 1$.

From (5.51), (5.50) and (5.19) it now follows that

$$G_o - \hat{G}^f = \Gamma + \Lambda - Z^f - S^f - \Upsilon \frac{1}{r} \sum_{i=1}^r \mathbf{F}_i$$

with

$$S^f = \Upsilon \frac{1}{r} \sum_{i=1}^r \mathbf{S}_i = \Upsilon \mathbf{S}$$

To separate bias and variance issues, similar to the technique presented in the previous sections, we introduce the auxiliary variables

$$\tilde{G}_i^f(e^{j\omega}) := G_o(e^{j\omega}) + \Upsilon \mathbf{F}_i \quad (5.54)$$

$$\tilde{G}^f(e^{j\omega}) := \frac{1}{r} \sum_{i=1}^r \tilde{G}_i^f(e^{j\omega}) \quad (5.55)$$

Furthermore, we will use an estimate of the variance of the transfer function estimate $\hat{G}^f(e^{j\omega})$, defined as

$$\hat{\sigma}_r^2(\hat{G}^f(e^{j\omega})) := \frac{1}{r(r-1)} \sum_{i=1}^r |\hat{G}^f(e^{j\omega}) - \tilde{G}_i^f(e^{j\omega})|^2 \quad (5.56)$$

Because the real and imaginary part of the averaged auxiliary variable $\tilde{G}^f(e^{j\omega})$ are not independent (as was the case for the nonparametric auxiliary $\tilde{G}(e^{j\omega_k})$ in Section 5.3.2), we will give separate confidence intervals for the real and imaginary part of the error. Subsequently, we will show that these separate confidence intervals can be combined to specify a rectangular confidence interval in the complex plane.

Lemma 5.4.3 *Consider the estimates $\hat{\sigma}_r^2(\hat{G}^f(e^{j\omega}))$, $\hat{\sigma}_r^2(\tilde{G}^f(e^{j\omega}))$ as defined in (5.56). Let the input signal be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$. Then, for all $\omega \in [0, 2\pi)$*

$$|\hat{\sigma}_r^2(\operatorname{Re}\{\tilde{G}^f(e^{j\omega})\}) - \hat{\sigma}_r^2(\operatorname{Re}\{\hat{G}^f(e^{j\omega})\})| \leq \mathcal{S}^f(\omega)$$

with

$$\mathcal{S}^f(\omega) = \frac{1}{r(r-1)} \sum_{i=1}^r (2 |A_i^f(e^{j\omega})| |B_i^f(e^{j\omega})| + |B_i^f(e^{j\omega})|^2) \quad (5.57)$$

and

$$\begin{aligned} |A_i^f(e^{j\omega})| &= |\hat{G}^f(e^{j\omega}) - \hat{G}_i^f(e^{j\omega})| \\ |B_i^f(e^{j\omega})| &\leq \frac{1}{r} \sum_{m=1}^r |S_m^f(e^{j\omega})| + \frac{r-2}{r} |S_i^f(e^{j\omega})| \\ |S_i^f(e^{j\omega})| &\leq \sum_{m=1}^{N_p} |\Upsilon(m)| |S_i(e^{j\zeta_m})| \end{aligned}$$

where $S_i(e^{j\zeta_m})$ is bounded by (5.20).

Proof: Similar to the proof of Lemma 5.3.3. \square

Theorem 5.4.4 Consider the estimated transfer function $\hat{G}^f(e^{j\omega})$, (5.50), (5.49), and the estimated variance $\hat{\sigma}_r^2(\hat{G}^f(e^{j\omega}))$, (5.56). Let the input signal be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, and let $n_f \leq n_p \leq N_p \leq N_o$. Then, asymptotically in N_o , for all $\omega \in [0, 2\pi)$

$$\begin{aligned} |\operatorname{Re}\{G_o(e^{j\omega})\} - \operatorname{Re}\{\hat{G}^f(e^{j\omega})\}| &\leq |S^f(e^{j\omega})| + |\operatorname{Re}\{\Lambda(e^{j\omega}) - Z^f(e^{j\omega})\}| \\ &\quad + |\operatorname{Re}\{\Gamma(e^{j\omega})\}| + \gamma_\alpha^f(\omega) \quad w.p. \geq F_\alpha(1, r-1) \end{aligned}$$

with

$$\begin{aligned} |S^f(e^{j\omega})| &\leq \sum_{\ell=1}^{N_p} |\Upsilon(\ell)| |S(e^{j\zeta_\ell})| \\ |\operatorname{Re}\{\Lambda(e^{j\omega}) - Z^f(e^{j\omega})\}| &\leq \mathcal{K} \sum_{k=\bar{n}_f}^{\bar{n}_p-1} \eta^k |\operatorname{Re}\{\mathcal{V}_k(e^{j\omega})\}| \\ &\quad + \mathcal{K} \sum_{k=\bar{n}_p}^{\bar{n}_h-1} \eta^k \left| \operatorname{Re}\{\mathcal{V}_k(e^{j\omega}) - \sum_{\ell=1}^{N_p} \Upsilon(\ell) \mathcal{V}_k(e^{j\zeta_\ell})\} \right| \\ &\quad + \mathcal{K} \frac{\eta^{\bar{n}_h}}{1-\eta} \left(|\mathcal{V}_0(e^{j\omega})| + \sum_{\ell=1}^{N_p} |\Upsilon(\ell)| |\mathcal{V}_0(e^{j\zeta_\ell})| \right) \\ |\operatorname{Re}\{\Gamma(e^{j\omega})\}| &\leq M |\operatorname{Re}\{\mathcal{P}_D\}| + \mathcal{K} \sum_{m=0}^{\bar{n}_f-1} |\operatorname{Re}\{\mathcal{P}_m\}|^T \eta^m \\ \gamma_\alpha^f(\omega) &\leq \sqrt{\alpha} \left(\hat{\sigma}_r^2(\operatorname{Re}\{\hat{G}^f(e^{j\omega})\}) + \mathcal{S}^f(\omega) \right)^{\frac{1}{2}} \end{aligned}$$

for any $\bar{n}_h \in \mathbb{N}$, $\bar{n}_h \geq \bar{n}_p$, while $\mathcal{S}^f(\omega)$ is given by (5.57), and $S(e^{j\zeta_\ell})$ is bounded by (5.21).

Proof: See Appendix B.6. □

Replacing $\text{Re}\{\cdot\}$ in Lemma 5.4.3 and Theorem 5.4.4 with $\text{Im}\{\cdot\}$ gives a confidence interval for the imaginary part of the error.

The result given in Theorem 5.4.4 does not directly provide a specification of a region in the complex plane that, with a certain probability, contains the true transfer function. A straightforward solution to this problem of specifying the simultaneous probability of two dependent random variables follows from the Bonferroni inequality which reduces a simultaneous probability statement to individual probability statements

$$P[x_i < a_i, i = 1, \dots, m] \geq 1 - \sum_{i=1}^m (1 - P[x_i < a_i]) \quad (5.58)$$

see e.g. [87]. Thus we can specify a lower bound on the probability that the true transfer function lies in a rectangle in the complex plane by using Theorem 5.4.4. That is, the rectangle in the complex plane, specified by using Theorem 5.4.4 for the real and imaginary part of the error, has at least probability

$$1 - (1 - F_\alpha(1, r - 1)) - (1 - F_\beta(1, r - 1)) = F_\alpha(1, r - 1) + F_\beta(1, r - 1) - 1$$

to contain $G_o(e^{j\omega})$, where $F_\alpha(1, r - 1)$ and $F_\beta(1, r - 1)$ denote the probability level for the real part and the imaginary part, respectively. Note that if the probability levels are high, then the Bonferroni inequality (5.58) yields values that are very close to the probability level that would arise if the real and imaginary part were independent. For individual probability levels of 99.5 % the Bonferroni inequality gives a simultaneous probability of 99 %, whereas the simultaneous probability would be 99.0025 % if the real and imaginary part were independent.

Alternatively, a joint confidence interval for the real and imaginary part of the error can be derived, using Hotellings T^2 statistic (see e.g. [87, page 171]). Define the estimated covariance matrix $\hat{\Sigma}_r^2(\hat{G}^f(e^{j\omega}))$ as

$$\hat{\Sigma}_r^2(\hat{G}^f(e^{j\omega})) := \frac{1}{r(r-1)} \sum_{i=1}^r \begin{bmatrix} \text{Re}\{\hat{G}_i^f(e^{j\omega}) - \hat{G}^f(e^{j\omega})\} \\ \text{Im}\{\hat{G}_i^f(e^{j\omega}) - \hat{G}^f(e^{j\omega})\} \end{bmatrix} \begin{bmatrix} \text{Re}\{\hat{G}_i^f(e^{j\omega}) - \hat{G}^f(e^{j\omega})\} \\ \text{Im}\{\hat{G}_i^f(e^{j\omega}) - \hat{G}^f(e^{j\omega})\} \end{bmatrix}^T \quad (5.59)$$

The confidence region now becomes the sum of a rectangular region due to the deterministic error terms (bias), and an ellipsoidal region due to the noise (variance), as formulated in the following theorem.

Theorem 5.4.5 *Consider the estimated transfer function $\hat{G}^f(e^{j\omega})$, (5.50), (5.49), and the estimated variance $\hat{\Sigma}_r^2(\hat{G}^f(e^{j\omega}))$, (5.59). Let the input signal be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{IN}$ and $r > 1$, and let $n_f \leq n_p \leq N_p \leq N_o$.*

Then, asymptotically in N_o , for all $\omega \in [0, 2\pi)$

$$G_o(e^{j\omega}) - \hat{G}^f(e^{j\omega}) \in \mathcal{A}(e^{j\omega}) \quad w.p. F_\alpha(2, r - 2)$$

where the set $\mathcal{A}(e^{j\omega})$ is given by all $\Delta \in \mathbb{C}$, $\Delta(e^{j\omega}) = \Delta_1(e^{j\omega}) + \Delta_2(e^{j\omega})$, $\Delta_1 \in \mathbb{C}$, $\Delta_2 \in \mathbb{C}$, satisfying

$$\begin{bmatrix} \text{Re}\{\Delta_1(e^{j\omega})\} \\ \text{Im}\{\Delta_1(e^{j\omega})\} \end{bmatrix}^T \left[\hat{\Sigma}_r^2(\hat{G}^f(e^{j\omega})) + 2\mathcal{S}^f(\omega)I \right]^{-1} \begin{bmatrix} \text{Re}\{\Delta_1(e^{j\omega})\} \\ \text{Im}\{\Delta_1(e^{j\omega})\} \end{bmatrix} \leq \frac{2(r-1)}{r-2}\alpha \quad (5.60)$$

$$|\text{Re}\{\Delta_2(e^{j\omega})\}| \leq |S^f(e^{j\omega})| + |\text{Re}\{\Lambda(e^{j\omega}) - Z^f(e^{j\omega})\}| + |\text{Re}\{\Gamma(e^{j\omega})\}| \quad (5.61)$$

$$|\text{Im}\{\Delta_2(e^{j\omega})\}| \leq |S^f(e^{j\omega})| + |\text{Im}\{\Lambda(e^{j\omega}) - Z^f(e^{j\omega})\}| + |\text{Im}\{\Gamma(e^{j\omega})\}| \quad (5.62)$$

where \mathcal{S}^f is given in (5.57), while bounds on S^f and the real and imaginary parts of $\Lambda - Z^f$ and Γ are given in Theorem 5.4.4.

Proof: See Appendix B.7. □

In Theorem 5.4.5 the inequality (5.60) specifies an ellipsoidal region in the complex plane, while the combination of (5.61) and (5.62) specifies a rectangular region in the complex plane. The size of the confidence regions specified in Theorem 5.4.5 does not differ much from the size of the rectangular confidence intervals which result from the Bonferroni inequality (5.58) and Theorem 5.4.4, but the ellipsoidal regions provide more information as to the direction of the error.

In Theorem 5.4.4 and Theorem 5.4.5 S^f and \mathcal{S}^f are the errors due to the unknown past inputs, Z^f and Λ are the errors due to undermodelling, and Γ is the error due to weighting with \mathcal{W} . For these error components hard bounds are provided. A probabilistic bound is given for the error due to the noise γ_{α}^f or Δ_1 .

Since $\mathcal{P} = P(I - \mathcal{W})$, see (5.53), it follows that $\Gamma(e^{j\omega})$ and the upper bound on $\Gamma(e^{j\omega})$ as given in Theorem 5.4.4 are zero when $\mathcal{W} = I$. A good choice for the weighting matrix \mathcal{W} is a diagonal matrix having its first diagonal elements equal to one, and decreasing thereafter (using e.g. a cosine taper or an exponential with base η). By choosing the point at which the elements start to decay appropriately, the variance can be reduced significantly, while relatively little bias is introduced. Usually, a good choice for this point is the point at which the confidence interval for the estimated generalized expansion coefficients (Theorem 5.4.2) becomes larger than the interval given by the prior information \mathcal{K} and η . For such a choice, the influence of \mathcal{W} can be seen as a regularization of the estimated parameters towards zero, thus using the prior knowledge that the generalized expansion coefficients converge to zero, which reduces the sensitivity to the actual choice for the order n_f , see [109].

Choosing \mathcal{W} as a diagonal matrix having its first (say n_w) diagonal elements equal to one, and exponentially decreasing thereafter, all deterministic error components in

Theorem 5.4.4 ($S^f, \mathcal{S}^f, Z^f, \Lambda, \Gamma$) have exponential convergence (in N_s, N_s, n_p, n_f , and n_w , respectively). Since the relative size of the bias (deterministic error components) and variance (probabilistic error component) contributions to the error in general will change with frequency, the optimal choice (yielding the smallest error bound) for the model order n_f in general also will change with frequency.

Note that we have obtained explicit and separate bounds for the errors due to noise, undermodelling, weighting and unknown past inputs, providing valuable insight in the effective ways to reduce and shape the error bound by input design and by choosing the orders, the weighting functions, and N_s . Therefore, together with the freedom to choose the basis generating system, these design choices can be used to effectively tune both the model and the error bound to robust control design specifications. As mentioned before, this is of decisive importance to indeed achieve models and error bounds which are suitable for high performance robust control design. Further insight into these choices will be provided in Chapter 7, where transparent asymptotic bias and variance expressions are established for frequency domain least squares identification using system based orthonormal basis functions, as employed in this section. Finally, considering input design, note that the procedure allows for nonuniformly spaced frequency points $\zeta_k \in \Omega^p$.

Comparing the result of Theorem 5.4.4 to the frequency domain results given in Ljung [105, 104, 106] for the asymptotic bias and variance of parametric transfer function estimates, we have the following three remarks. Firstly, in contrast to the results of [105, 104, 106], Theorem 5.4.4 holds for finite model orders n_p and n_f , and we have explicit expressions for the various bias contributions. Secondly, we use an estimate of the variance, rather than the unknown true spectrum of the noise. Thirdly, we have obtained an estimate of the variance of the estimated transfer function (5.56) that is consistent in the presence of undermodelling. That is, the estimated variance converges to the true variance as N_o and r go to infinity. Recently, consistent estimates for the variance of a transfer function estimate have also been established by Hjalmarsson [85, 83], considering time domain prediction error methods, when the model is linear in the parameters.

Additionally, we have derived explicit bounds for the bias and variance errors, enabling a clear tradeoff between these terms by e.g. selecting the model order. If bias errors are not taken into account, as in Zhu [192, 194], or if no explicit error bounds are available, as in Ljung [105, 106] and Brillinger [20], such a tradeoff cannot be made.

5.4.4 Mixed parametric-nonparametric uncertainty

In the previous sections we have specified the uncertainty either as parametric (structured) or nonparametric (unstructured) uncertainty. That is, Theorem 5.4.2 specifies the uncertainty as structured uncertainty, whereas Theorem 5.4.4 or Theorem 5.4.5 specify the uncertainty as unstructured uncertainty. Now, on the one hand, the uncer-

tainty specified by Theorem 5.4.2 does not capture the complete uncertainty about the system G_o , since it only addresses the uncertainty on the finite number of estimated parameters, while the series expansion of G_o contains an infinite number of parameters (compare (5.29) with (5.3)). On the other hand, the error bounds provided by Theorem 5.4.4 or Theorem 5.4.5 do not reflect all available information about the system. That is, these error bounds do not account for that fact that the transfer function of the system is known to be a smooth function of frequency, as reflected by the prior information \mathcal{K} and η or M and ρ (see e.g. Proposition 4.4.1), thus allowing the true transfer function to be a very very erratic function of frequency. Therefore, we will now try to combine both uncertainty descriptions.

First, note that the results of Theorem 5.4.4 or Theorem 5.4.5 can easily be extended to an estimate of a transfer function G_{of} given by the tail of the generalized expansion coefficients of G_o

$$G_{of}(z) = \sum_{k=\bar{n}_s}^{\infty} L_k \mathcal{V}_k(z) \quad (5.63)$$

with $\bar{n}_s > 0$. Now, using (5.3), G_o can be written as

$$G_o(z) = D + \sum_{k=0}^{\bar{n}_s-1} L_k \mathcal{V}_k(z) + G_{of}(z)$$

Thus, considering (5.29), Theorem 5.4.2 can be used to specify the uncertainty in the estimated parameters $D, L_0, \dots, \bar{n}_s - 1$, whereas Theorem 5.4.4 or Theorem 5.4.5 can be used to specify the uncertainty in G_{of} .

In Proposition E.1.4 a procedure is given to transform the uncertainty in the estimated parameters D, L_k , to uncertainty in the output and feedthrough matrices of a state-space description of the estimated model. Thus, the uncertainty can be divided into parametric uncertainty in a low order state-space description, due to the first $n_s = \bar{n}_s n_b + 1$ estimated parameters, and unstructured uncertainty caused by the remaining parameters. Note that choosing \bar{n}_s to be small, does not imply that \bar{n}_p in (5.29) should be chosen to be small, when determining the uncertainty in the estimated parameters, as this may induce too much bias, see Theorem 5.4.2. Especially, taking $\bar{n}_s = 1$ results in a state-space realization having the same system and input matrix as the input balanced realization of the basis generating system, see Proposition E.1.4. If the basis generating system is given by a (control-relevant) nominal model with input balanced state-space realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, then a mixed parametric-nonparametric uncertainty description is obtained for the nominal model as follows. Let $\Delta \hat{L}_0, \Delta \hat{D}$, denote the uncertainty in the estimated parameters \hat{L}_0, \hat{D} , respectively. Then the uncertainty with respect to the prespecified nominal model can be specified as parametric uncertainty $|\mathbf{C} - \hat{L}_0| + \Delta \hat{L}_0$ and $|\mathbf{D} - \hat{D}| + \Delta \hat{D}$ in, respectively, the output and feedthrough matrices \mathbf{C} and \mathbf{D} of the nominal model, and an unstructured uncertainty description representing $G_{of}(e^{j\omega})$ of (5.63) with $\bar{n}_s = 1$. Note that if the nominal

model is redefined to be $(\mathbf{A}, \mathbf{B}, \hat{L}_0, \hat{D})$, then the parametric uncertainty just becomes the uncertainty in \hat{L}_0 and \hat{D} .

The structured error bounds arising from the procedure described above can be used in a μ analysis or synthesis approach to robust control, see [38, 40, 7, 91, 138, 27]. The uncertainty then can be specified as structured uncertainty in the form of a diagonal matrix with n_s real valued diagonal elements representing the parametric uncertainty, and one complex valued diagonal element representing the nonparametric uncertainty.

5.4.5 Undisturbed output signal

In many cases it is important to have a clear idea of the true (noise free) response of the system to a certain input signal $u_a(\tau)$, and to what extent an estimate of this response can be trusted. In other words, it is often important to be able to assess the prediction accuracy of a model. Of course $u_a(\tau)$ can be chosen as the input $u(t)$ used during identification, but any other input is allowed also. Let the known input $u_a(\tau)$ be applied to the system in a time interval $\tau \in T^{N_a}$. Because the new origin of the time axis ($\tau = 0$) can be different from the one in the previous sections, we need additional prior information on the past values of the input to the system.

Assumption 5.4.6 *We have as a priori information that there exists a finite and known $\bar{u}_a^p \in \mathbb{R}$, such that $|u_a(\tau)| \leq \bar{u}_a^p$ for $\tau < 0$.*

In the following we will provide expressions for the undisturbed output signal, and the estimate of this signal that we will use. Using (E.8) and (E.9) the undisturbed output can be written as

$$y_o(\tau) = [u_a(\tau) \ u_a(\tau-1) \ \cdots] \begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & \phi_0^T(0) & \phi_1^T(0) & \cdots \\ 0 & \phi_0^T(1) & \phi_1^T(1) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} D \\ L_0^T \\ L_1^T \\ \vdots \end{bmatrix} \quad (5.64)$$

For $q \leq \tau$ the estimated output now can be calculated as

$$\hat{y}(\tau) = [u_a(\tau) \ \cdots \ u_a(\tau-q)] \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \phi_0^T(0) & \cdots & \phi_{\bar{n}_y-1}^T(0) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \phi_0^T(q-1) & \cdots & \phi_{\bar{n}_y-1}^T(q-1) \end{bmatrix} \mathcal{W}_y \begin{bmatrix} \hat{D} \\ \hat{L}_0^T \\ \vdots \\ \hat{L}_{\bar{n}_y-1}^T \end{bmatrix} \quad (5.65)$$

$$= Q(\tau) \mathcal{W}_y \hat{\theta} \quad (5.66)$$

where $\mathcal{W}_y \in \mathbb{R}^{n_y \times n_p}$ is a weighting matrix, $n_y = \bar{n}_y n_b + 1$. We can write

$$\hat{y}_i(\tau) = Q(\tau) \mathcal{W}_y \hat{\theta}_i \quad (5.67)$$

$$\hat{y}(\tau) = \frac{1}{r} \sum_{i=1}^r \hat{y}_i(\tau) \quad (5.68)$$

From (5.64) and (5.65) it follows that the true undisturbed output and its estimate can be rewritten as

$$\begin{aligned} y_o(\tau) &= Du_a(\tau) + \sum_{m=0}^{\infty} u_a(\tau-1-m) \sum_{k=0}^{\infty} L_k \phi_k(m) \\ \hat{y}(\tau) &= \check{D}u_a(\tau) + \sum_{m=0}^{q-1} u_a(\tau-1-m) \sum_{k=0}^{\tilde{n}_y-1} \check{L}_k \phi_k(m) \end{aligned} \quad (5.69)$$

where

$$[\check{D} \ \check{L}_0 \ \cdots \ \check{L}_{\tilde{n}_y-1}]^T = \mathcal{W}_y [\hat{D} \ \hat{L}_0 \ \cdots \ \hat{L}_{\tilde{n}_y-1}]^T$$

There holds

$$\hat{y}_i(\tau) = \Upsilon^y \hat{G}_i$$

with

$$\Upsilon^y = Q(\tau) \mathcal{W}_y \Psi$$

where \hat{G}_i is given in (5.32), and where Ψ is given in (5.36). To separate bias and variance issues, we introduce the auxiliary variables $\tilde{y}_i(\tau)$ and $\tilde{y}(\tau)$

$$\begin{aligned} \tilde{y}_i(\tau) &:= y_o(\tau) + \Upsilon^y F_i \\ \tilde{y}(\tau) &:= \frac{1}{r} \sum_{i=1}^r \tilde{y}_i(\tau) \end{aligned} \quad (5.70)$$

Finally, we will denote

$$\mathcal{P}^y = [\mathcal{P}_D^y \ \mathcal{P}_0^y \ \cdots \ \mathcal{P}_{\tilde{n}_y-1}^y] := Q(\tau) (\mathcal{W}_y - I)$$

with the appropriate dimensions for \mathcal{P}_D^y and \mathcal{P}_k^y , i.e. $\mathcal{P}_D^y \in \mathbb{C}$ and $\mathcal{P}_k^y \in \mathbb{C}^{1 \times n_b}$ for $k = 0, \dots, \tilde{n}_y - 1$.

With the usual expressions for the estimated variances, see (5.47), we now have the following results.

Lemma 5.4.7 *Consider the estimates $\hat{\sigma}_r^2(\hat{y}(\tau))$, $\hat{\sigma}_r^2(\tilde{y}(\tau))$, as defined in (5.47). Let the input signal $u(t)$ used for identification of the model (5.34) be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, and let $\tau \in T^{N_s}$. Then*

$$|\hat{\sigma}_r^2(\tilde{y}(\tau)) - \hat{\sigma}_r^2(\hat{y}(\tau))| \leq \mathcal{S}^y(\tau)$$

with

$$\mathcal{S}^y(\tau) = \frac{1}{r(r-1)} \sum_{i=1}^r (2 |A_i^y(\tau)| |B_i^y(\tau)| + |B_i^y(\tau)|^2) \quad (5.71)$$

and

$$\begin{aligned} |A_i^y(\tau)| &= |\hat{y}(\tau) - \hat{y}_i(\tau)| \\ |B_i^y(\tau)| &\leq \frac{1}{r} \sum_{m=1}^r |s_m^y(\tau)| + \frac{r-2}{r} |s_i^y(\tau)| \\ |s_i^y(\tau)| &\leq \sum_{m=1}^{N_p} |\Upsilon^y\langle m \rangle| |S_i(e^{j\zeta_m})| \end{aligned}$$

where $S_i(e^{j\zeta_m})$ is bounded by (5.20).

Proof: Similar to the proof of Lemma 5.3.3. \square

Theorem 5.4.8 Consider the estimated undisturbed output $\hat{y}(\tau)$, (5.68), (5.67), and the estimated variance $\hat{\sigma}_r^2(\hat{y}(\tau))$, (5.47). Let the input signal $u(t)$ used for identification of the model (5.34) be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, and let $n_y \leq n_p \leq N_p \leq N_o$. Let the input signal $u_a(\tau)$ be known, let $0 \leq q \leq \tau$, and let $\tau \in T^{N_a}$. Then, asymptotically in N_o

$$\begin{aligned} |y_o(\tau) - \hat{y}(\tau)| &\leq |s^y(\tau)| + |z^y(\tau)| + |s_a^y(\tau)| + |z_a^y(\tau)| + |\Gamma^y(\tau)| + \gamma_\alpha^y(\tau) \\ &w.p. \geq F_\alpha(1, r-1) \end{aligned}$$

with

$$\begin{aligned} |s^y(\tau)| &\leq \sum_{\ell=1}^{N_p} |\Upsilon^y\langle \ell \rangle| |S(e^{j\zeta_\ell})| \\ |z^y(\tau)| &\leq \mathcal{K} \sum_{k=\bar{n}_p}^{\bar{n}_h-1} \eta^k \left| \sum_{\ell=1}^{N_p} \Upsilon^y\langle \ell \rangle \mathcal{V}_k(e^{j\zeta_\ell}) \right| + \mathcal{K} \frac{\eta^{\bar{n}_h}}{1-\eta} \sum_{\ell=1}^{N_p} |\Upsilon^y\langle \ell \rangle| |\mathcal{V}_0(e^{j\zeta_\ell})| \\ |s_a^y(\tau)| &\leq M \sum_{m=q+1}^{\tau} |u_a(\tau-m)| \rho^{-m} + \bar{u}_a^p M \frac{\rho^{-\tau}}{\rho-1} \\ |z_a^y(\tau)| &\leq \mathcal{K} \sum_{k=\bar{n}_y}^{\bar{n}_x-1} \eta^k \left| \sum_{m=1}^q u_a(\tau-m) \phi_k(m-1) \right| \\ &\quad + \|\mathcal{K}\|_2 \eta^{\bar{n}_x} \sqrt{\frac{1}{1-\eta^2}} \sum_{m=1}^q |u_a(\tau-m)| \sqrt{1 - \sum_{k=0}^{\bar{n}_x-1} \|\phi_k(m-1)\|_2^2} \\ |\Gamma^y(\tau)| &\leq M |\mathcal{P}_D^y| + \mathcal{K} \sum_{m=0}^{\bar{n}_y-1} |\mathcal{P}_m^y|^T \eta^m \\ \gamma_\alpha^y(\tau) &\leq \sqrt{\alpha} (\hat{\sigma}_r^2(\hat{y}(\tau)) + \mathcal{S}^y(\tau))^{\frac{1}{2}} \end{aligned}$$

for any $\bar{n}_h, \bar{n}_x \in \mathbb{N}$, $\bar{n}_h, \bar{n}_x \geq \bar{n}_y$, while $\mathcal{S}^y(\tau)$ is given by (5.71), and $S(e^{j\zeta_\ell})$ is bounded by (5.21).

Proof: See Appendix B.8. □

In Theorem 5.4.8 $s^y(\tau)$ is the error due to unknown past inputs $u(t)$, $z^y(\tau)$ is the error due to undermodelling during the estimation of θ_o , $s_a^y(\tau)$ is the error due to the unknown past of $u_a(\tau)$, $z_a^y(\tau)$ is the error due to undermodelling during the estimation of the output $y_o(\tau)$, $\Gamma^y(\tau)$ is the error due to weighting with \mathcal{W}_y , and $\gamma_\alpha^y(\tau)$ is the error due to the noise. Hard error bounds are given for $s^y(\tau)$, $z^y(\tau)$, $s_a^y(\tau)$, $z_a^y(\tau)$, and $\Gamma^y(\tau)$. A confidence interval is given for the error due to the noise.

Theorem 5.4.8 provides a direct means to evaluate the prediction accuracy of the model (5.34). Note that the capability of a model to reproduce the measured output c.q. to predict future outputs is an important tool to judge the quality of a model, see e.g. Ljung [106, 107]. A good prediction accuracy often even is the prime goal of a model. Additionally, the result of Theorem 5.4.8 appears to be valuable for robust predictive control and robust feedforward control purposes. The control action should yield both a satisfactory predicted output trajectory as well as sufficiently small possible deviations between the predicted and the true output, i.e. a small confidence interval.

Finally, we have the following remarks.

- Theorem 5.4.8 provides a confidence interval for the impulse response $g_o(\tau)$ of the system when $u_a(\tau)$ is chosen as the impulse: $u_a(\tau) = \delta(\tau)$. Hence, the prior information on M and ρ can be improved iteratively by adapting M and ρ to match the resulting error bounds on the impulse response, similar to choosing \mathcal{K} and η as discussed at the end of Section 5.4.2.
- When possible, q must be chosen large, so that the error due to the fact that q is less than infinity is relatively small. Note that q can not be larger than τ , so that large errors can occur at the beginning of the response (if \bar{u}_a^p is not very small). If q is large (and N_s is large) the prior information on M , ρ and \bar{u}_a^p need not be tight to obtain satisfactory error bounds.
- If \mathcal{W}_y is block-diagonal with first diagonal block equal to the identity matrix with dimension $n_w = \bar{n}_w n_b + 1$, then

$$|\Gamma^y(\tau)| \leq \mathcal{K} \sum_{m=\bar{n}_w}^{\bar{n}_y-1} |\mathcal{P}_m^y|^T \eta^m$$

5.4.6 Disturbed output signal

In the context of model validation, the availability of reliable error bounds is desirable, since they directly attach a quality tag to the model. However, also the error bound itself should be validated. This can be done if a bound on the disturbed output can be specified. This error bound then can be compared directly to a new set of experimental

data. Thus, the availability of a confidence interval for the disturbed output is of major importance for model validation purposes, see Section 3.10 and Ljung [106, 107]. When comparing the measured output with such a confidence interval, one can check immediately whether they are compatible, and thus whether the model (including the prior information, the assumptions and the error bounds) could be a valid one. To be able to specify a confidence interval for the disturbed output in the time domain

$$y(\tau) = Du_a(\tau) + \sum_{m=0}^{\infty} u_a(\tau - 1 - m) \sum_{k=0}^{\infty} L_k \phi_k(m) + v(\tau) \quad (5.72)$$

we need the additional assumption that the noise $v(\tau)$ is normally distributed. The plausibility of this assumption can be checked using the techniques described in e.g. [87, pages 13-25 and 146-162], and a confidence interval for $y(\tau)$ can be derived. However, the assumption that the noise $v(\tau)$ is normally distributed clearly will make this confidence interval questionable in many practical applications. Therefore we will use a frequency domain formulation, so that we can use the result of Theorem 5.3.2 which states that the DFT of the noise is asymptotically normally distributed, and derive a confidence interval for the DFT of the disturbed output $y(\tau)$. This also provides a far better indication as to the frequency ranges in which the model is correct or at fault. First we will derive confidence intervals for the DFT of the noise and the DFT of the undisturbed output. Next these intervals will be combined into a confidence interval for the DFT of the disturbed output.

We will split up the time interval T^{N_a} into the two intervals T^{N_q} and $T_{N_q}^{N_b}$, where $N_b = N_a - N_q$ and $0 \leq N_q < N_a$. The first interval T^{N_q} will serve to diminish the effects of the unknown past input signals $u_a(\tau)$, $\tau < 0$, on the DFT of the predicted undisturbed output over the second interval $T_{N_q}^{N_b}$. Denote

$$V^q(e^{j\zeta\ell}) = \frac{1}{\sqrt{N_b}} \sum_{\tau=0}^{N_b-1} v(\tau + N_q) e^{-j\zeta\ell\tau} \quad \zeta\ell \in \Omega_{N_b} \quad (5.73)$$

$$\hat{Y}(e^{j\omega_k}) = \frac{1}{r} \sum_{i=1}^r Y_i(e^{j\omega_k}) \quad \omega_k \in \Omega_{N_o}^u \quad (5.74)$$

$$\hat{\sigma}^2(Y_i(e^{j\omega_k})) = \frac{1}{r-1} \sum_{i=1}^r |\hat{Y}(e^{j\omega_k}) - Y_i(e^{j\omega_k})|^2 \quad (5.75)$$

where $Y_i(e^{j\omega_k})$ is defined in (5.6). Note that the estimated variance (5.75) is independent of i . A confidence interval for the DFT of the noise is given in the following theorem.

Theorem 5.4.9 *Consider the DFT of the noise $V^q(e^{j\zeta\ell})$, (5.73), (5.72), and the estimated variance $\hat{\sigma}^2(Y_i(e^{j\omega_k}))$, (5.75). Let the input signal $u(t)$ used for identification of the model (5.34) be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$*

and $r > 1$, and let $\zeta_\ell \in \Omega_{N_o}^u$. Then, asymptotically in N_o and N_b

$$|\operatorname{Re}\{V^q(e^{j\zeta_\ell})\}| \leq \gamma_\alpha^v(\zeta_\ell) \quad \text{w.p.} \geq F_\alpha(1, r-1)$$

with

$$\begin{aligned} \gamma_\alpha^v(\zeta_\ell) &\leq \sqrt{\alpha} (\hat{\sigma}^2(\operatorname{Re}\{Y_i(e^{j\zeta_\ell})\}) + \mathcal{S}^v(\zeta_\ell))^{\frac{1}{2}} \\ \mathcal{S}^v(\zeta_\ell) &= r |U_i(e^{j\zeta_\ell})|^2 \mathcal{S}(\zeta_\ell) \end{aligned}$$

where $\mathcal{S}(\zeta_\ell)$ is given by (5.24).

Proof: Follows directly from Theorem 5.3.5. \square

Alternatively, we could formulate a confidence interval for $|V^q(e^{j\zeta_\ell})|$ similar to Theorem 5.3.5.

A confidence interval for the real and imaginary part of the DFT of the undisturbed output can be obtained from the results of Section 5.4.5 as follows. Denote for $\zeta_\ell \in \Omega_{N_b}$

$$\hat{Y}_i^q(e^{j\zeta_\ell}) = \frac{1}{\sqrt{N_b}} \sum_{\tau=0}^{N_b-1} \hat{y}_i(\tau + N_q) e^{-j\zeta_\ell \tau} \quad (5.76)$$

$$\hat{Y}^q(e^{j\zeta_\ell}) = \frac{1}{r} \sum_{i=1}^r \hat{Y}_i^q(e^{j\zeta_\ell}) \quad (5.77)$$

$$Y_o^q(e^{j\zeta_\ell}) = \frac{1}{\sqrt{N_b}} \sum_{\tau=0}^{N_b-1} y_o(\tau + N_q) e^{-j\zeta_\ell \tau} \quad (5.78)$$

$$Y^q(e^{j\zeta_\ell}) = \frac{1}{\sqrt{N_b}} \sum_{\tau=0}^{N_b-1} y(\tau + N_q) e^{-j\zeta_\ell \tau} \quad (5.79)$$

where $\hat{y}_i(\tau)$ is given by (5.67), $y_o(\tau)$ by (5.69), and $y(\tau)$ by (5.72). There holds

$$\hat{Y}_i^q(e^{j\zeta_\ell}) = \Phi^q \hat{G}_i$$

with

$$\Phi^q = P^q \mathcal{W}_y \Psi \quad (5.80)$$

$$P^q = \frac{1}{\sqrt{N_b}} [1 \ e^{-j\zeta_\ell} \ \dots \ e^{-j\zeta_\ell(N_b-1)}] [Q^T(N_q) \ Q^T(N_q+1) \ \dots \ Q^T(N_b-1)]^T$$

where \hat{G}_i is given by (5.32), $Q(\tau)$ by (5.65), (5.66), and Ψ by (5.36). To separate bias and variance issues, we introduce the auxiliary variables $\tilde{Y}_i^q(e^{j\zeta_\ell})$ and $\tilde{Y}^q(e^{j\zeta_\ell})$

$$\tilde{Y}_i^q(e^{j\zeta_\ell}) = \frac{1}{\sqrt{N_b}} \sum_{\tau=0}^{N_b-1} \tilde{y}_i(\tau + N_q) e^{-j\zeta_\ell \tau}$$

$$\tilde{Y}^q(e^{j\zeta_\ell}) = \frac{1}{r} \sum_{i=1}^r \tilde{Y}_i^q(e^{j\zeta_\ell})$$

where $\tilde{y}_i(\tau)$ is given in (5.70).

A reduction of the time needed to compute the error due to the unknown past inputs $u_a(\tau)$ can be obtained by using

$$\bar{u}_a := \max_{\tau \in T^{N_a}} |u_a(\tau)|$$

instead of the exact values $|u_a(\tau)|$ for $\tau \in T^{N_a}$, see Theorem 5.4.11. If q is chosen large enough in Theorem 5.4.11 this approximation will not significantly increase the error bound. Finally, we will denote

$$\mathcal{P}^q = [\mathcal{P}_D^q \ \mathcal{P}_0^q \ \cdots \ \mathcal{P}_{\bar{n}_y-1}^q] := P^q (W_y - I)$$

with the appropriate dimensions for \mathcal{P}_D^q and \mathcal{P}_k^q , i.e. $\mathcal{P}_D^q \in \mathbb{C}$ and $\mathcal{P}_k^q \in \mathbb{C}^{1 \times n_b}$ for $k = 0, \dots, \bar{n}_y - 1$.

Lemma 5.4.10 *Consider the estimates $\hat{\sigma}_r^2(\hat{Y}^q(e^{j\zeta_\ell}))$, $\hat{\sigma}_r^2(\tilde{Y}^q(e^{j\zeta_\ell}))$, as defined in (5.47). Let the input signal $u(t)$ used for identification of the model (5.34) be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, and let $\zeta_\ell \in \Omega_{N_b}$. Then*

$$|\hat{\sigma}_r^2(\tilde{Y}^q(e^{j\zeta_\ell})) - \hat{\sigma}_r^2(\hat{Y}^q(e^{j\zeta_\ell}))| \leq \mathcal{S}^q(\zeta_\ell)$$

with

$$\mathcal{S}^q(\zeta_\ell) = \frac{1}{r(r-1)} \sum_{i=1}^r (2 |A_i^q(e^{j\zeta_\ell})| |B_i^q(e^{j\zeta_\ell})| + |B_i^q(e^{j\zeta_\ell})|^2) \quad (5.81)$$

and

$$\begin{aligned} |A_i^q(e^{j\zeta_\ell})| &= |\hat{Y}^q(e^{j\zeta_\ell}) - \tilde{Y}_i^q(e^{j\zeta_\ell})| \\ |B_i^q(e^{j\zeta_\ell})| &\leq \frac{1}{r} \sum_{m=1}^r |S_m^q(e^{j\zeta_\ell})| + \frac{r-2}{r} |S_i^q(e^{j\zeta_\ell})| \\ |S_i^q(e^{j\zeta_\ell})| &\leq \sum_{m=1}^{N_p} |\Phi^q(m)| |S_i(e^{j\zeta_m})| \end{aligned}$$

where $S_i(e^{j\zeta_m})$ is bounded by (5.20).

Proof: Similar to the proof of Lemma 5.3.3. □

Theorem 5.4.11 *Consider the estimate \hat{Y}^q , (5.77), (5.76), and the estimated variance $\hat{\sigma}_r^2(\hat{Y}^q(e^{j\zeta_\ell}))$, (5.47). Let the input signal $u(t)$ used for identification of the model (5.34) be periodic with period N_o for $t \in T^{N+N_s}$, with $N = rN_o$, $r \in \mathbb{N}$ and $r > 1$, and let $n_y \leq n_p \leq N_p \leq N_o$. Let the input signal $u_a(\tau)$ be known for $\tau \in T^{N_a}$, let $0 \leq N_q < N_a$, and let $\zeta_\ell \in \Omega_{N_b}$. Then, asymptotically in N_o*

$$\begin{aligned} |\operatorname{Re}\{Y_o^q(e^{j\zeta_\ell}) - \hat{Y}^q(e^{j\zeta_\ell})\}| &\leq |S^q(e^{j\zeta_\ell})| + |\operatorname{Re}\{Z^q(e^{j\zeta_\ell})\}| + |\operatorname{Re}\{S_a^q(e^{j\zeta_\ell})\}| \\ &\quad + |\operatorname{Re}\{Z_a^q(e^{j\zeta_\ell})\}| + |\operatorname{Re}\{\Gamma^q(e^{j\zeta_\ell})\}| + \gamma_\alpha^q(\zeta_\ell) \\ w.p. &\geq F_\alpha(1, r-1) \end{aligned}$$

with

$$\begin{aligned}
|S^q(e^{j\zeta_\ell})| &\leq \sum_{\ell=1}^{N_p} |\Phi^q(\ell)| |S(e^{j\zeta_\ell})| \\
|\operatorname{Re}\{Z^q(e^{j\zeta_\ell})\}| &\leq \mathcal{K} \sum_{k=\bar{n}_p}^{\bar{n}_h-1} \eta^k \left| \operatorname{Re}\left\{ \sum_{\ell=1}^{N_p} \Phi^q(\ell) \mathcal{V}_k(e^{j\zeta_\ell}) \right\} \right| + \mathcal{K} \frac{\eta^{\bar{n}_h}}{1-\eta} \sum_{\ell=1}^{N_p} |\Phi^q(\ell)| |\mathcal{V}_0(e^{j\zeta_\ell})| \\
\operatorname{Re}\{S_a^q(e^{j\zeta_\ell})\} &\leq \frac{1}{\sqrt{N_b}} M \sum_{\tau=1}^{N_b-1} |\operatorname{Re}\{e^{-j\zeta_\ell \tau}\}| \sum_{m=N_q+1}^{\tau+N_q} |u_a(\tau+N_q-m)| \rho^{-m} \\
&\quad + \frac{1}{\sqrt{N_b}} \frac{\bar{u}_a^p M}{\rho-1} \frac{\rho}{\rho-1} (1-\rho^{-N_b}) \rho^{-N_q} \\
&\leq \frac{1}{\sqrt{N_b}} \frac{M}{\rho-1} \left(\bar{u}_a N_b - \frac{\bar{u}_a}{\rho-1} (1-\rho^{-N_b+1}) + \frac{\bar{u}_a^p \rho}{\rho-1} (1-\rho^{-N_b}) \right) \rho^{-N_q} \\
|\operatorname{Re}\{Z_a^q(e^{j\zeta_\ell})\}| &\leq \frac{\mathcal{K}}{\sqrt{N_b}} \sum_{k=\bar{n}_y}^{\bar{n}_x-1} \eta^k \left| \operatorname{Re}\left\{ \sum_{m=1}^{N_q} \phi_k(m-1) \sum_{\tau=0}^{N_b-1} u_a(\tau+N_q-m) e^{-j\zeta_\ell \tau} \right\} \right| \\
&\quad + \frac{\|\mathcal{K}\|_2}{\sqrt{N_b}} \frac{\eta^{\bar{n}_x}}{\sqrt{1-\eta^2}} \sum_{m=1}^{N_q} \left| \sum_{\tau=0}^{N_b-1} u_a(\tau+N_q-m) e^{-j\zeta_\ell \tau} \right| \sqrt{1 - \sum_{k=0}^{\bar{n}_x-1} \|\phi_k(m-1)\|_2^2} \\
|\operatorname{Re}\{\Gamma^q(e^{j\zeta_\ell})\}| &\leq M |\operatorname{Re}\{\mathcal{P}_D^q\}| + \mathcal{K} \sum_{m=0}^{\bar{n}_y-1} |\operatorname{Re}\{\mathcal{P}_m^q\}|^T \eta^m \\
\gamma_\alpha^q(\zeta_\ell) &\leq \sqrt{\alpha} \left(\hat{\sigma}_r^2(\operatorname{Re}\{\hat{Y}^q(e^{j\zeta_\ell})\}) + \mathcal{S}^q(\zeta_\ell) \right)^{\frac{1}{2}}
\end{aligned}$$

for any $\bar{n}_h, \bar{n}_x \in \mathbb{N}$, $\bar{n}_h, \bar{n}_x \geq \bar{n}_y$, while $\mathcal{S}^q(\zeta_\ell)$ is given by (5.81), and $S(e^{j\zeta_\ell})$ is bounded by (5.21).

Proof: Follows directly from Theorem 5.4.8. \square

Similar as in Section 5.4.3 we can now specify a rectangular confidence region in the complex plane for $Y_o^q(e^{j\zeta_\ell})$ using the Bonferroni inequality. Alternatively, we can specify an ellipsoidal confidence region using Hotellings T^2 -statistic.

Combining the confidence interval of Theorem 5.4.9 with the confidence interval of Theorem 5.4.11 provides a confidence interval for the DFT of the disturbed output $Y^q(e^{j\zeta_\ell})$, by noting that $V^q(e^{j\zeta_\ell}) = Y^q(e^{j\zeta_\ell}) - Y_o^q(e^{j\zeta_\ell})$, see (5.69) and (5.72). To obtain the probability level of this combined confidence interval we have to use the Bonferroni inequality (5.58) because the two confidence statements Theorem 5.4.9 and Theorem 5.4.11 are dependent.

Corollary 5.4.12 *Consider the situation of Theorem 5.4.11 and Theorem 5.4.9, but with $\zeta_\ell \in \Omega_{N_o}^u \cap \Omega_{N_b}$. Then*

$$\begin{aligned} |\operatorname{Re}\{Y^q(e^{j\zeta_\ell}) - \hat{Y}^q(e^{j\zeta_\ell})\}| &\leq |S^q(e^{j\zeta_\ell})| + |\operatorname{Re}\{Z^q(e^{j\zeta_\ell})\}| + |\operatorname{Re}\{S_a^q(e^{j\zeta_\ell})\}| \\ &\quad + |\operatorname{Re}\{Z_a^q(e^{j\zeta_\ell})\}| + |\operatorname{Re}\{\Gamma^q(e^{j\zeta_\ell})\}| + \gamma_\alpha^q(\zeta_\ell) + \gamma_\alpha^v(\zeta_\ell) \\ \text{w.p. } &\geq 1 - 2(1 - F_\alpha(1, r - 1)) \end{aligned}$$

where $S^q(e^{j\zeta_\ell})$, $\operatorname{Re}\{Z^q(e^{j\zeta_\ell})\}$, $\operatorname{Re}\{S_a^q(e^{j\zeta_\ell})\}$, $\operatorname{Re}\{Z_a^q(e^{j\zeta_\ell})\}$, $\operatorname{Re}\{\Gamma^q(e^{j\zeta_\ell})\}$, and $\gamma_\alpha^q(\zeta_\ell)$ are bounded in Theorem 5.4.11, and where $\gamma_\alpha^v(\zeta_\ell)$ is bounded in Theorem 5.4.9.

Replacing $\operatorname{Re}\{\cdot\}$ in Theorem 5.4.9, Theorem 5.4.11, and Corollary 5.4.12 with $\operatorname{Im}\{\cdot\}$ gives a confidence interval for the imaginary part of the error.

A comparison of the confidence interval for the DFT of the disturbed output, as given by Corollary 5.4.12, with the DFT of the measured output can very well be used for validation of the model, the assumptions, the prior information, and the error bounds. That is, the combination of the model, the error bound and the noise description should be able to capture new data. The ability of a model to reproduce measured data is stressed by Ljung [107] who states that the prime validation tool is to test how well the model is able to reproduce new data. The need to test the prior information and the assumptions is emphasized by Wahlberg and Ljung [182], stating that it is in fact the prior that must be scrutinized, since the posterior model is a direct mathematical consequence of the prior information, the assumptions and the data. Alternatively, the confidence interval for the DFT of the disturbed output can be used for fault detection purposes with a prespecified probability level for a false alarm.

5.5 Remarks

With respect to the results described in the previous sections we have a number of general remarks.

- As regards the sensitivity of the confidence intervals to the asymptotic normality of the individual estimates, Johnson and Wichern [87, page 146] state that “in situations where the sample size is large and the techniques depend solely on the behaviour of the estimated mean \bar{x} , or distances involving \bar{x} of the form $r(\bar{x} - \mu)^T \hat{S}^{-1}(\bar{x} - \mu)$ the assumption of normality for the individual observations is less crucial”, where μ is the true mean and \hat{S} is the estimated covariance matrix. This situation is the situation that is addressed in the current contribution. A similar remark is made by Kendall and Stuart [90, Chapter 31].
- Because the bias terms have exponential convergence, whereas the variance has linear convergence, the optimal model order (with respect to the size of the confidence regions) yields a considerable lower bias than variance contribution, but

the bias contribution should certainly not be neglectable with respect to the variance contribution. This is only to some extent in line with the classical results, which state that the bias and variance contributions should be approximately of the same size. Note that the variance contribution, and hence the optimal model order, changes with the probability level. Similarly, considering the transfer function estimate, the optimal model order changes with frequency, since the bias and variance contributions change with frequency.

- The confidence intervals derived in this chapter can be easily extended to multivariable systems. That is, confidence intervals for each element of a transfer function matrix can be provided. For estimates of an element of a transfer function matrix only the error due to unknown past inputs changes, since the pasts of the other inputs also influence the output. An upper bound for the error due to past inputs for multivariable systems has already been derived in Section 4.3.3 when the DFT's over one period of the input signals are orthogonal, see Theorem 4.3.5.
- The extension to situations where the system operates in closed loop, so that only the external reference signal can be chosen to be periodic, is given in Chapter 6.

5.6 Recapitulation of results

At this point, the theoretical part of this chapter (finally) comes to an end. This section provides an overview of the main results, their mutual relations and their application areas, in the form of Table 5.1.

This chapter will continue by providing several simulation studies, followed by the conclusions.

5.7 Example

To illustrate the results of this chapter a simulation is made of a fifth order system

$$G_o(z) = \frac{0.7027 - 0.8926z^{-1} + 0.24z^{-2} + 0.5243z^{-3} - 0.9023z^{-4} + 0.4009z^{-5}}{1 - 2.4741z^{-1} + 2.8913z^{-2} - 1.9813z^{-3} + 0.8337z^{-4} - 0.1813z^{-5}} \quad (5.82)$$

whose impulse response $g_o(k)$ satisfies a bound given by $M_o = 2$ and $\rho_o = 1.23$. The input signal obeys $\bar{u}^p = 2$ and is required to satisfy $\bar{u} \leq 1$. There is 10 percent (in standard deviation) colored noise $v(t) = H_o(q)e(t)$ on the output, where the noise filter is a third order highpass filter

$$H_o(z) = \frac{0.7184 - 0.2206z^{-1} + 0.2390z^{-2} - 0.0060z^{-3}}{1 + 0.1177z^{-1} + 0.3208z^{-2} - 0.0182z^{-3}} \quad (5.83)$$

	Error bound for	Basis for	Application
theorem 5.3.5	nonparametric transfer function estimate $\hat{G}(e^{j\omega_k})$, only defined for $\omega_k \in \Omega_{N_o}^u$	Theorem 5.4.2 Corollary 5.4.12	
theorem 5.4.2	estimated parameters $\hat{\theta}$ (generalized expansion coefficients)	Theorem 5.4.5 Theorem 5.4.8	estimate the a priori information \mathcal{K} and η , specifying the undermodelling error in theorems 5.4.2, 5.4.5, 5.4.8 and Corollary 5.4.12
theorem 5.4.5	parametric transfer function estimate $\hat{G}^f(e^{j\omega})$, defined for all $\omega \in [0, 2\pi)$		robust control design
theorem 5.4.8	estimated undisturbed output $\hat{y}(\tau)$	Corollary 5.4.12	prediction of future outputs, simulation, also can be used to estimate the a priori information M and ρ , specifying the error due to unknown past inputs in theorems 5.4.2, 5.4.5, 5.4.8 and Corollary 5.4.12
corollary 5.4.12	estimated DFT of the disturbed output $\hat{Y}^q(e^{j\zeta\epsilon})$		model validation

Table 5.1: Overview of the main theoretical results of this chapter.

and $e(t)$ is a *uniformly* distributed random sequence. The probability level, as given by F_α , is specified to be 99 % for all confidence regions that will be shown.

The input signal is chosen as a random-phased multi-sine, with phases uniformly distributed in the interval $[0, 2\pi)$, $\bar{u} = 1$, $N_o = 64$, $N_s = 76$ and $N = 1024$. Thus, 1100 data points are used. The magnitude of the DFT of the input signal over one period is given in Fig. 5.1. Note that the frequency points where $|U_i(e^{j\omega_k})| > 0$ are not equidistant. The magnitude and frequency grid of $U_i(e^{j\omega_k})$ express that we are particularly interested in the behaviour of the system around 0.9 rad/s , and that we do not expect the behaviour of the system to change rapidly with frequency for the higher frequencies.

First a fourth order model is identified using an output error (OE) model structure, and a least squares criterion. This model is used to generate the orthonormal basis functions. With respect to this basis generator $\mathcal{K}_o = [1.0 \ 0.8 \ 0.3 \ 0.9]$ and $\eta_o = 0.38$.

The values for M , ρ , \mathcal{K} , and η are estimated from the data iteratively. Starting from zero, we iterate till the estimated prior information is in accordance with the matching confidence intervals, i.e. till the bounds given by the estimates of the prior information cross the resulting confidence intervals only where zero is contained in the confidence interval. This results in $\hat{M} = 2.5$, $\hat{\rho} = 1.16$, $\hat{\mathcal{K}} = [1.1 \ 0.8 \ 0.7 \ 1.5]$, and $\hat{\eta} = 0.2$. These estimated values now are taken as the prior information of Assumption 5.2.1.ii and Assumption 5.2.1.iii.

Figure 5.2 depicts the true generalized expansion coefficients, the prior information given by $\hat{\mathcal{K}}$ and $\hat{\eta}$, and the confidence interval of Theorem 5.4.2, using $n_p = 21$ and $W = \text{diag}(1/\hat{\sigma}_r(\hat{G}(e^{j\omega_k})))$. Since $L_k \in \mathbb{R}^4$ the expansion coefficients are depicted in four plots, each referring to a specific component of L_k , $k = 0, \dots, (n_p - 1)/n_b - 1$. Figure 5.2 shows that the prior information given by $\hat{\mathcal{K}}$ and $\hat{\eta}$ is consistent with the resulting confidence intervals for the estimated generalized expansion coefficients. The weighting matrix \mathcal{W} , that will be used for the parametric estimate $\hat{G}^f(e^{j\omega})$, is chosen as a diagonal matrix with diagonal elements as given in Fig. 5.3. Comparing the prior bound and the confidence interval for the generalized expansion coefficients, as given in Fig. 5.2, it can be seen that \mathcal{W} is chosen as indicated on page 95. Figure 5.4 depicts the confidence interval for the nonparametric estimate $\hat{G}(e^{j\omega_k})$ of Theorem 5.3.5, which is only defined at $\omega_k \in \Omega_{N_o}^u$, and the confidence interval for a parametric estimate $\hat{G}^f(e^{j\omega})$ of Theorem 5.4.5, using $n_f = 21$ and \mathcal{W} as described above, together with the true errors. Note that very good estimates are obtained for those frequencies where $|U_i(e^{j\omega_k})|$ was chosen to be large. Comparing the parametric estimate $\hat{G}^f(e^{j\omega})$ with the nonparametric estimate $\hat{G}(e^{j\omega_k})$, it follows that the parametric estimate achieves a large reduction of the variance where the signal to noise ratio in the frequency domain is relatively small, while only little bias is introduced, so that the confidence interval becomes only slightly larger for those frequencies where the signal to noise ratio is large. Figure 5.5 gives the Nyquist plot of the true system, together with the confidence region according to the estimate $\hat{G}^f(e^{j\omega})$. Figure 5.6 depicts the true impulse response, the

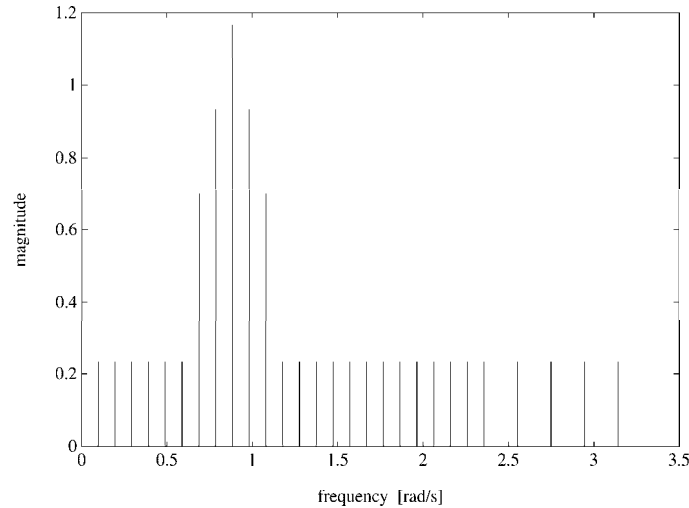


Fig. 5.1: Magnitude of the DFT over one period of the input signal, $|U_i(e^{j\omega_k})|$.

prior information given by \hat{M} and $\hat{\rho}$, and the confidence interval of Theorem 5.4.8, using $n_y = 17$ and a weighting matrix \mathcal{W}_y having a similar form as the one used for $\hat{G}^f(e^{j\omega})$ (i.e. \mathcal{W} , see Fig. 5.3). Figure 5.6 shows that the prior information given by \hat{M} and $\hat{\rho}$ is consistent with the resulting confidence interval for the estimated impulse response. Finally, Fig. 5.7 depicts the undisturbed output that arises in response to a random input signal, and the confidence interval of Theorem 5.4.8, using the same n_y and \mathcal{W}_y as for the impulse response estimate, and $q = 60$.

Note that in all cases the error bounds are tight, i.e. in each case the actual error is at some points close to the upper bound.

5.8 Monte Carlo analysis

5.8.1 Introduction

In this section we will give the results of a Monte Carlo analysis of the model error estimation procedure which has been presented in this chapter. That is, given a desired error probability (specified by the confidence level F_α), we calculated the actual error percentages, being the number of trespasses of an error bound divided by the total number of trials. We performed this analysis for the nonparametric transfer function estimate of Theorem 5.3.5, the estimated generalized expansion coefficients of Theorem 5.4.2, the estimated impulse response of Theorem 5.4.8, and the parametric transfer function estimate of Theorem 5.4.4.

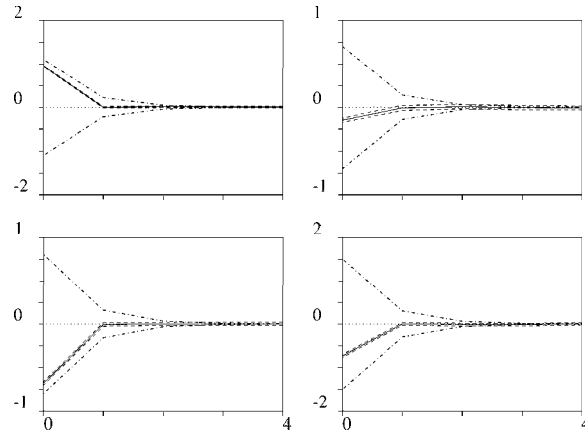


Fig. 5.2: True parameters $\theta_o\langle k \rangle$ (—), confidence interval of $\hat{\theta}\langle k \rangle$ (- -), and estimated prior information $\hat{\mathcal{K}}\hat{\eta}^k$ (- ·).

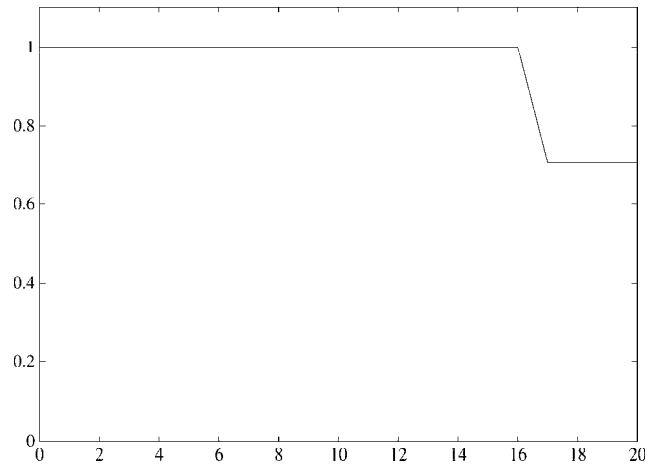


Fig. 5.3: Diagonal elements of the diagonal weighting function \mathcal{W} used for the parametric transfer function estimate $\hat{G}^f(e^{j\omega})$.

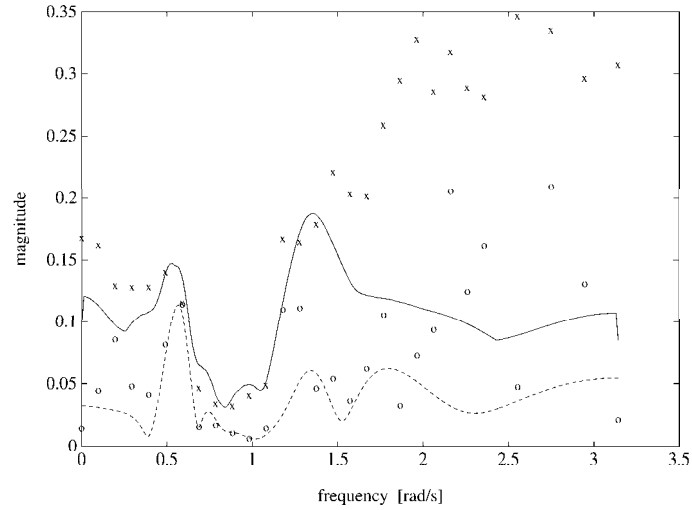


Fig. 5.4: Confidence interval of the nonparametric estimate $\hat{G}(e^{j\omega_k})$ (\times), defined at $\omega_k \in \Omega_{N_o}^u$, true error $|\hat{G}(e^{j\omega_k}) - G_o(e^{j\omega_k})|$ (\circ), confidence interval of the parametric estimate $\hat{G}^f(e^{j\omega})$ (—), defined for all $\omega \in [0, 2\pi)$, and true error $|\hat{G}^f(e^{j\omega}) - G_o(e^{j\omega})|$ (- -).

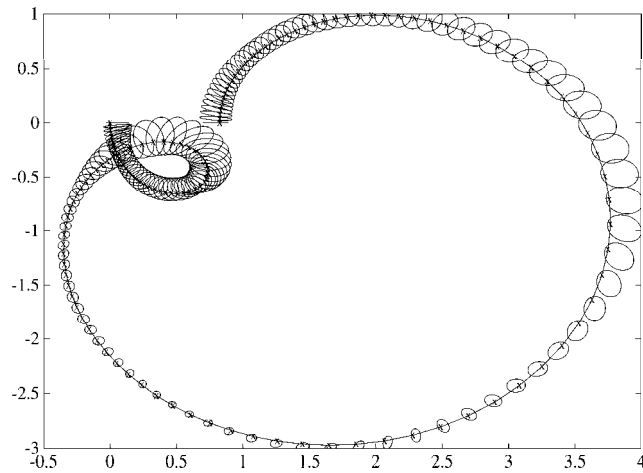


Fig. 5.5: Nyquist plot of true system and confidence region of the parametric transfer function estimate $\hat{G}^f(e^{j\omega})$.

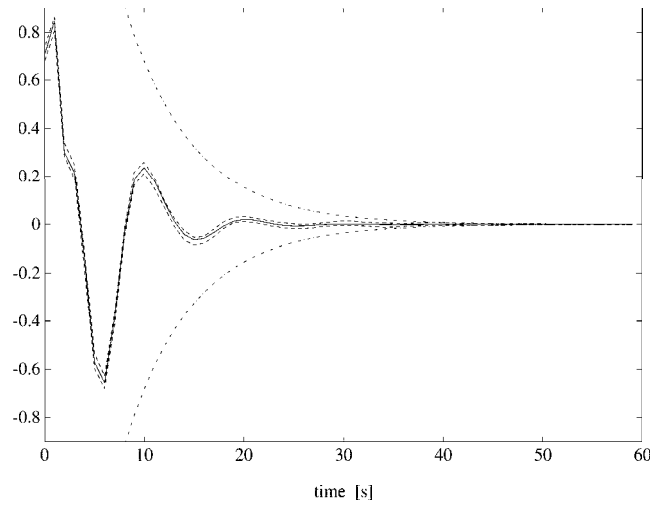


Fig. 5.6: True impulse response $g_o(k)$ (—), confidence interval of $\hat{y}(\tau)$ (- -), and estimated prior information $\hat{M}\hat{\rho}^{-k}$ (- ·).

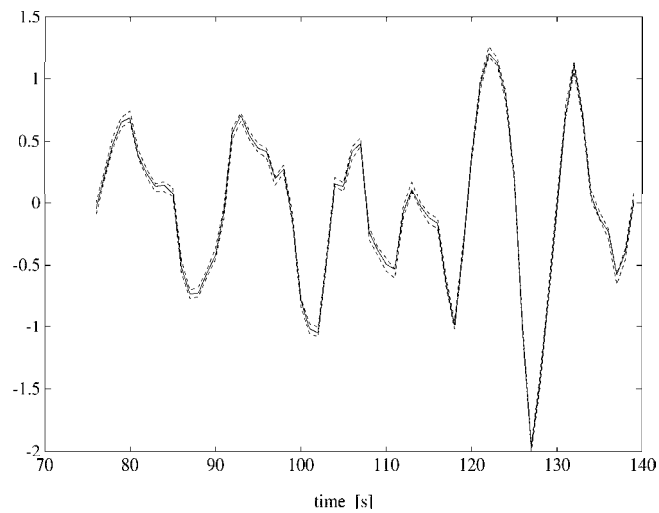


Fig. 5.7: True undisturbed response $y_o(\tau)$ (—), and confidence interval of $\hat{y}(\tau)$ (- -)

5.8.2 Description

The system and (if the noise is colored) the noise filter that we used are the same as the ones that were used in the example of Section 5.7, see (5.82) and (5.83). The basis generating system and the values for M , ρ , \mathcal{K} and η were estimated from a data set with filtered uniformly distributed noise on the output, resulting in

$$G_b(z) = \frac{0.7 - 0.36z^{-1} - 0.04z^{-2} + 0.49z^{-3} - 0.53z^{-4}}{1 - 1.7z^{-1} + 1.6z^{-2} - 0.8z^{-3} + 0.25z^{-4}}$$

$\hat{M} = 3$, $\hat{\rho} = 1.2$, $\hat{\mathcal{K}} = [1.5 \ 0.8 \ 0.7 \ 1]$ and $\hat{\eta} = 0.25$. These estimated values for M , ρ , \mathcal{K} and η now are taken as the prior information of Assumption 5.2.1.ii and Assumption 5.2.1.iii. The values for \mathcal{K} and η differ from the ones of Section 5.7 because the basis generating system is different.

We investigated the behaviour of our model error estimation procedures for five noise cases:

1. Colored normally distributed noise (noise filter H_o of (5.83)) with zero mean.
2. Noise that has been constructed by taking the sign of a zero mean normally distributed random sequence (not filtered), which results in a random binary sequence (RBS).
3. Noise that has been constructed by dividing two zero mean normally distributed random variables (not filtered) with variance one, and bounding the amplitude at 10, which results in near Cauchy noise.
4. Colored uniformly distributed noise (noise filter H_o of (5.83)) with zero mean.
5. First, zero mean uniformly distributed noise (not filtered) has been added to the output. Subsequently the disturbed output signal has been rounded to one decimal, which introduces discretization "noise".

We did perform 1000 trials for each estimate and noise distribution, resulting in a total of 5000 estimates. In each trial the same input signal was used, but a different realization of the noise was added to the undisturbed output. The noise level was 10 percent (in standard deviation) in each trial. The input signal was a random-phased multi-sine with the same amplitude for each $\omega_k \in \Omega_{N_o}$. We had $n_p = 21$ in Theorem 5.4.2, $n_f = 21$ in Theorem 5.4.4, and $n_y = 21$ in Theorem 5.4.8. The weighting matrix W for the generalized expansion coefficients of Theorem 5.4.2 was chosen to be $W = \text{diag}(1/\hat{\sigma}_r(\hat{G}(e^{j\omega_k})))$. The weighting matrix \mathcal{W} for the transfer function estimate of Theorem 5.4.4, and the weighting matrix \mathcal{W}_y for the impulse response estimate of Theorem 5.4.8, were chosen as diagonal matrices with diagonal elements as given in Fig. 5.3. For the parametric transfer function estimate we used rectangular confidence regions instead of ellipsoidal confidence regions, because for a

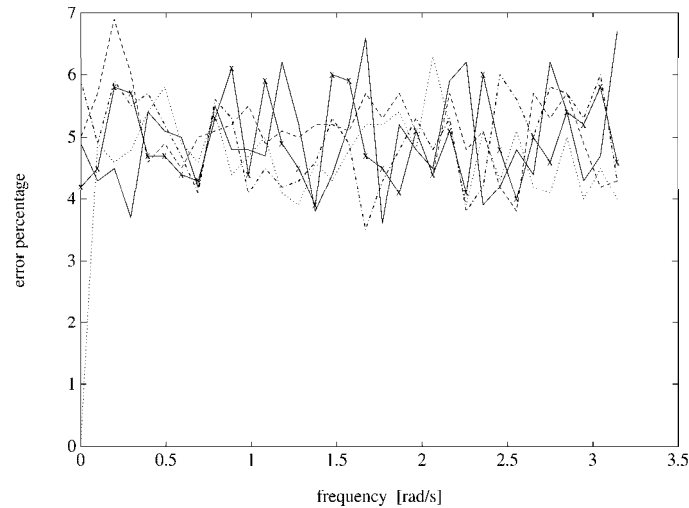


Fig. 5.8: Actual error percentages for the nonparametric transfer function estimate $\hat{G}(e^{j\omega_k})$, (—) colored normal noise, (– –) RBS noise, ($\cdot \cdot \cdot$) near Cauchy noise, (– \cdot –) colored uniform noise, (– \times –) uniform & discretization noise.

rectangular region it is much easier to detect automatically whether a point is inside or outside the region. For each trial 1100 data points were used with $N = 1024$, $N_o = 64$ and $N_s = 76$. The probability level for all confidence regions was specified to be 95 %, i.e. the specified error percentage is 5 %.

5.8.3 Results

Figure 5.8 gives the actual error percentages for the nonparametric transfer function estimate $\hat{G}(e^{j\omega_k})$ of Theorem 5.3.5, Fig. 5.9 gives the actual error percentages for the estimated generalized expansion coefficients $\hat{\theta}(k)$ of Theorem 5.4.2, Fig. 5.10 gives the actual error percentages for the estimated impulse response $\hat{g}(k)$ of Theorem 5.4.8, and Fig. 5.11 gives the actual error percentages for the parametric transfer function estimate $\hat{G}^f(e^{j\omega})$ of Theorem 5.4.4. The somewhat peculiar frequency grid of Fig. 5.11 arises because the frequency grid is chosen such that an equidistant gridding of the Nyquist plot of the system results. The mean error percentages for each estimate and noise distribution are given in Table 5.2.

The fact that the error percentages are sometimes significantly lower than the specified ones, is due to the influence of the undermodelling error. For the undermodelling error a hard error bound is used, which certainly can be too large in a number of points, so that the resulting error bound actually has a higher probability level in these points. This is also expressed by the statement "with probability larger than or

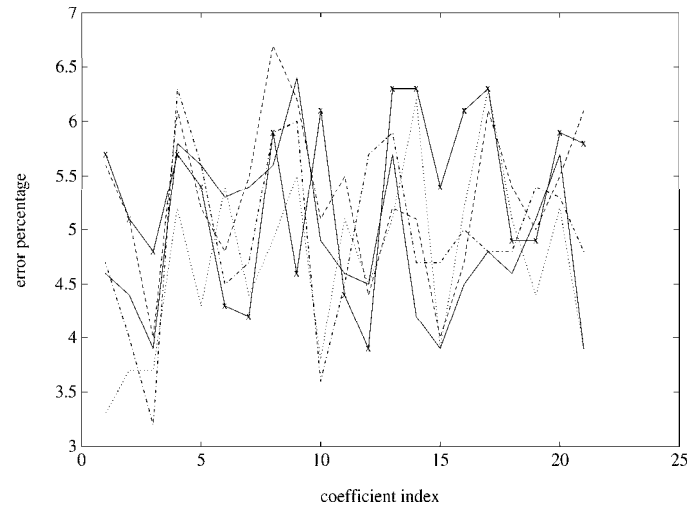


Fig. 5.9: Actual error percentages for the estimated generalized expansion coefficients $\hat{\theta}(k)$, (—) colored normal noise, (– –) RBS noise, ($\cdot \cdot \cdot$) near Cauchy noise, (– \cdot) colored uniform noise, (– \times) uniform & discretization noise.

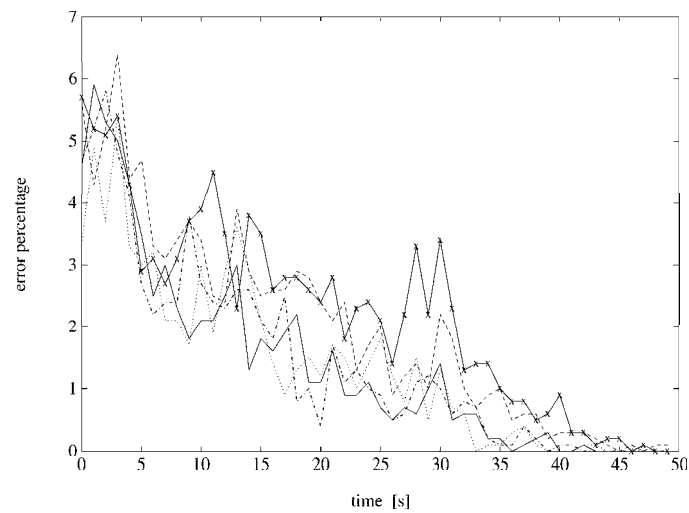


Fig. 5.10: Actual error percentages for the estimated impulse response $\hat{g}(k)$, (—) colored normal noise, (– –) RBS noise, ($\cdot \cdot \cdot$) near Cauchy noise, (– \cdot) colored uniform noise, (– \times) uniform & discretization noise.

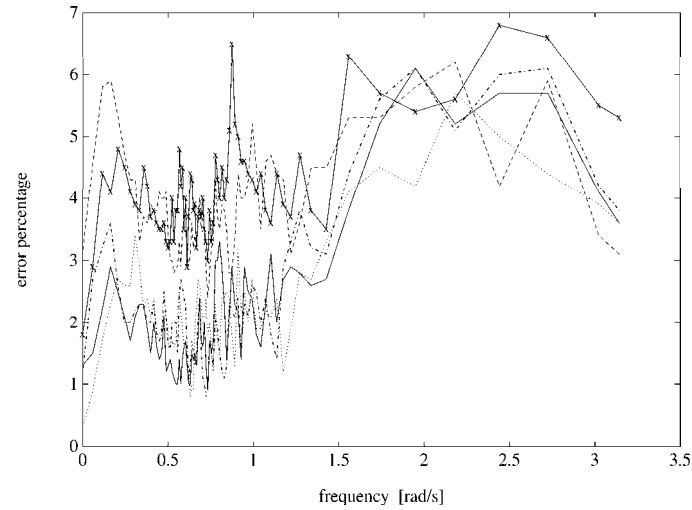


Fig. 5.11: Actual error percentages for the parametric transfer function estimate $\hat{G}^f(e^{j\omega})$, (—) colored normal noise, (- -) RBS noise, (· · ·) near Cauchy noise, (- ·) colored uniform noise, (-x-) uniform & discretization noise.

	\hat{G}	$\hat{\theta}$	\hat{g}	\hat{G}^f
colored normal noise	4.942	4.924	1.422	2.209
RBS noise	5.115	5.300	2.000	3.863
near Cauchy noise	4.621	4.719	1.412	2.273
colored uniform noise	4.982	4.957	1.460	2.333
uniform & discretization noise	4.927	5.333	2.240	4.165

Table 5.2: Actual averaged error percentages for the different estimates and noise distributions. In each case the specified error percentage was 5 %.

equal to” in the theorems concerned (Theorem 5.3.5, Theorem 5.4.2, Theorem 5.4.4, Theorem 5.4.8).

5.8.4 Discussion

The results of the Monte Carlo analysis can be summarized as follows. When the deterministic error terms (errors due to undermodelling, weighting, and unknown past inputs) are small in comparison to the probabilistic error term (error due to the noise), then the actual error percentages are very close to the specified ones. When the deterministic error terms are not small relative to the probabilistic error term, then the actual error percentages can be considerably smaller than the specified ones. However, this should be the case, because the deterministic error terms are bounded with hard (100 % guaranteed) error bounds, so that the probability level of the error bounds is bounded from below.

It should be noted that the Monte Carlo analysis was performed in a nonasymptotic situation, and that the a priori information M , ρ , \mathcal{K} and η was estimated from the data.

5.9 Application to a nonlinear simulation system

5.9.1 Introduction

In Chapter 8 the model error estimation procedure introduced in this chapter is applied to a laboratory set-up of a wind turbine system (and its nonlinear simulation model), and to the pick-up mechanism of a compact disk player. In anticipation of these application to real life systems, we will in this section and the following two sections (Section 5.10.1 and Section 5.10.2), investigate the behaviour of the model error estimation techniques under mild deviations from the assumptions that the system is linear and time invariant c.q. that the noise is stochastic and independent of the input signal, on the basis of a number of simulation studies. That is, what are the simulation results of the algorithms if the system is slightly nonlinear or time varying? An indication of the behaviour under these conditions is of course very important; if the behaviour severely deteriorates than the procedure cannot be expected to give realistic error bounds in practice.

5.9.2 The nonlinear simulation system

The structure of the nonlinear simulation system is given in Fig. 5.12. The first block in Fig. 5.12 can be seen as a finite word length digital-analog converter (DAC), the second block as a limiter, the fourth block as a threshold and the last block as a analog-digital converter (ADC) with again a finite word length. There is no noise disturbance acting on the system. The MATLABTM code of the nonlinear system is given by

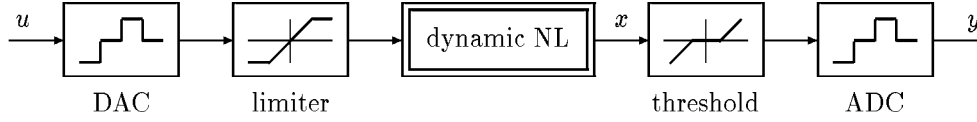


Fig. 5.12: The nonlinear dynamical simulation system.

```

u(t) = round(u(t)*256/umax)/256;           % DAC
u(t) = sign(u(t))*min(abs(u(t)),umax*0.9); % limiter

x(t) = 0.7*u(t-1) - 0.36*u(t-2) - 0.04*u(t-3) + 0.49*u(t-4)
      - 0.53*u(t-5)
      + 0.01*u(t-1)^2 - 0.01*sign(u(t-1)-u(t-2))*umax/20
      + 0.01*cos(u(t-2)) - 0.01 + 0.01*u(t-3)^3
      + 1.72*x(t-1) - 1.59*x(t-2) + 0.79*x(t-3) - 0.25*x(t-4)
      + 0.02*x(t-1)*cos(50*x(t-4))
      - 0.01*sin(100*x(t-2)*x(t-5));

y(t) = sign(x(t))*max(abs(x(t))-umax/10,0); % threshold
y(t) = round(y(t)*256/umax)/256;           % ADC

```

where $umax = \bar{u}$. The term $\text{sign}(u(t-1)-u(t-2))*umax/20$ in the code symbolizes dry friction. Terms similar to the sine and cosine terms in the code can appear in rotating mechanical systems (robots, electrical drives, wind turbines, et cetera).

5.9.3 Results

For the basis generating system we used the linear part of the nonlinear simulation system

$$G_b(z) = \frac{0.7z^{-1} - 0.36z^{-2} - 0.04z^{-3} + 0.49z^{-4} - 0.53z^{-5}}{1 - 1.72z^{-1} + 1.59z^{-2} - 0.79z^{-3} + 0.25z^{-4}} \quad (5.84)$$

We used 400 data points with $N = 256$, $N_o = 64$ and $N_s = 80$. The input signal was chosen to be a random-phased multi-sine with the same amplitude for each $\omega_k \in \Omega_{N_o}$, and with $\bar{u} = \bar{u}^p = 1$. The probability level for all confidence regions was specified to be 99 %, i.e. the specified error percentage is 1 %.

The values for M , ρ , \mathcal{K} , and η are obtained from the data iteratively. Starting from zero, we iterate till the estimated prior information is in accordance with the matching confidence intervals, i.e. till the bounds given by the estimates of the prior information cross the matching confidence intervals only where zero is contained in the confidence interval. This results in $\hat{M} = 1.5$, $\hat{\rho} = 1.1$, $\hat{\mathcal{K}} = [2 \ 1.5 \ 0.3 \ 0.8 \ 1]$, and $\hat{\eta} = 0.3$. These estimated values now are taken as the prior information of Assumption 5.2.1.ii and Assumption 5.2.1.iii.

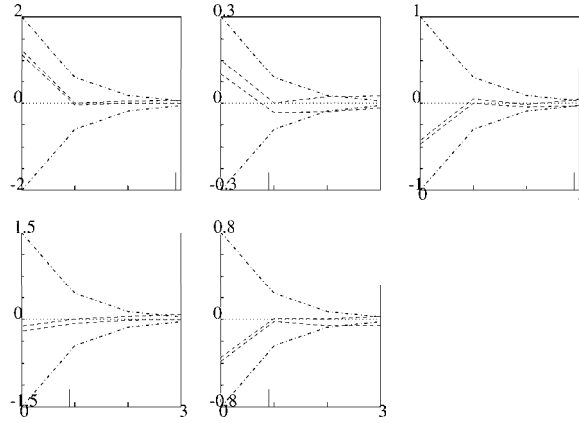


Fig. 5.13: Confidence interval of $\hat{\theta}\langle k \rangle$ and estimated prior information $\hat{\mathcal{K}}\hat{\eta}^k$, for $N = 256$.

We had $n_p = 21$ in Theorem 5.4.2, $n_f = 21$ in Theorem 5.4.5, and $n_y = 21$ in Theorem 5.4.8. The weighting matrix W for the generalized expansion coefficients of Theorem 5.4.2 was chosen to be $W = \text{diag}(1/\hat{\sigma}_r(\hat{G}(e^{j\omega_k})))$. The weighting matrix \mathcal{W} for the transfer function estimate of Theorem 5.4.5, and the weighting matrix \mathcal{W}_y for the undisturbed time domain (impulse) response estimate of Theorem 5.4.8, were chosen as diagonal matrices with diagonal elements as given in Fig. 5.3.

Figure 5.13 gives the confidence interval of the estimated generalized expansion coefficients $\hat{\theta}\langle k \rangle$ of Theorem 5.4.2, together with the estimated prior information given by $\hat{\mathcal{K}}$ and $\hat{\eta}$. The parametric transfer function estimate $\hat{G}^f(e^{j\omega})$ and the matching confidence interval of Theorem 5.4.5 are given in Fig. 5.14, together with the transfer function of the linear part of the nonlinear system as given in (5.84). Note that this linear part is contained in the confidence interval. The fact that the estimated transfer function $\hat{G}^f(e^{j\omega})$ tends to have a lower gain is explained by the influence of the limiter and the threshold in the nonlinear system.

To further test the validity of the above results, we will use the test of Theorem 5.4.8 over new data (data which has not been used to estimate the confidence intervals). Figure 5.15 gives the output of the nonlinear system, together with the confidence interval of $\hat{y}(t)$. Although there are a number of outliers, the confidence interval accounts for the observed data fairly reasonable.

An important question now is: will the error bounds still be "correct" if we increase the number of averages c.q. the period length of the input signal, or will they become

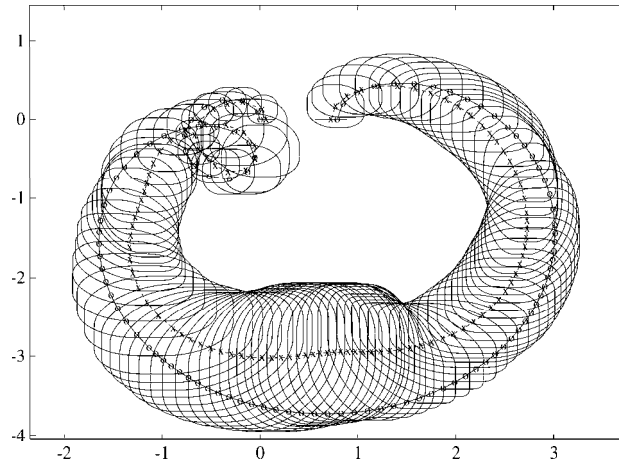


Fig. 5.14: $\hat{G}^f(e^{j\omega})$ with confidence interval, and transfer function of the linear part (5.84) of the nonlinear system, for $N = 256$.

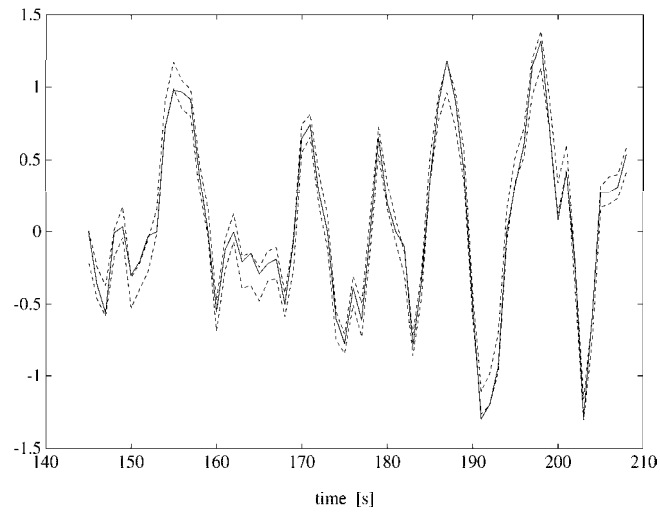


Fig. 5.15: Confidence interval of $\hat{y}(t)$ and output of the nonlinear system, for $N = 256$.

much too small ?

We will study the results of doubling both the number of averages r , and the period length of the input signal N_o . That is, we will use 1104 data points with $N = 1024$, $N_o = 128$ and $N_s = 80$. Estimating the prior information from the data, we again find $\hat{M} = 1.5$, $\hat{\rho} = 1.1$, but $\hat{\mathcal{K}} = [1.5 \ 0.8 \ 0.7 \ 1 \ 0.7]$, and $\hat{\eta} = 0.6$. It should be noted that we had some difficulty in finding priors $\hat{\mathcal{K}}$ and $\hat{\eta}$ which are consistent with the resulting error bounds. At first sight the values obtained for $\hat{\mathcal{K}}$ and $\hat{\eta}$ appear to be rather conservative. However, a slight decrease in $\hat{\mathcal{K}}$ and $\hat{\eta}$ results in smaller error bounds so that zero is no longer contained in the confidence intervals for the tails of the estimated parameter sequences. Also, since the variance error is relatively small when compared to the undermodelling error, one would expect that estimating a higher order model will reduce the error bounds. This is however not the case. Using the same prior information $\hat{\mathcal{K}}$ and $\hat{\eta}$ smaller confidence intervals are indeed obtained when the model order is increased since the undermodelling error decreases, but the confidence intervals on the estimated generalized expansion coefficients also show that the prior information $\hat{\mathcal{K}}$ and $\hat{\eta}$ is no longer consistent with these confidence intervals, and should be larger. Larger values for $\hat{\mathcal{K}}$ and $\hat{\eta}$ again will result in an increase in the undermodelling error.

As usual, the estimated values $\hat{\mathcal{K}}$ and $\hat{\eta}$ now are taken as the prior information of Assumption 5.2.1.ii and Assumption 5.2.1.iii. We now use $n_p = n_f = n_y = 36$, and the resulting confidence intervals for $\hat{\theta}(k)$ and $\hat{G}^f(e^{j\omega})$ are depicted in Fig. 5.16 and Fig. 5.17. Comparing Fig. 5.14 and Fig. 5.17, it follows that increasing the number of data points with a factor four, does not significantly reduce the error bound for parametric transfer function estimate $\hat{G}^f(e^{j\omega})$. The confidence interval still contains the transfer function of the linear part of the nonlinear system. However, the variance error has reduced considerably, whereas the undermodelling error has increased considerably. This can be seen by noting that a variance error gives rise to an ellipsoidal confidence interval, whereas undermodelling and weighting result in rectangular error bounds, see Theorem 5.4.5.

To test the validity of the above results, we will again use the test of Theorem 5.4.8 over new data. The result is given in Fig. 5.18, from which it follows that the estimated error bound indeed captures the output of the nonlinear system (there are no outliers at all).

5.9.4 Discussion

The above observed behaviour of our model error estimation procedure can be explained as follows.

Firstly, the nonlinear system, when excited with a periodic input signal, gives rise to an output signal of which only a part tends to become periodic. That is, the nonlinear system can – conceptually – be divided into a part that has a vanishing memory, which will give rise to an output that tends to become periodic in response to a periodic input,

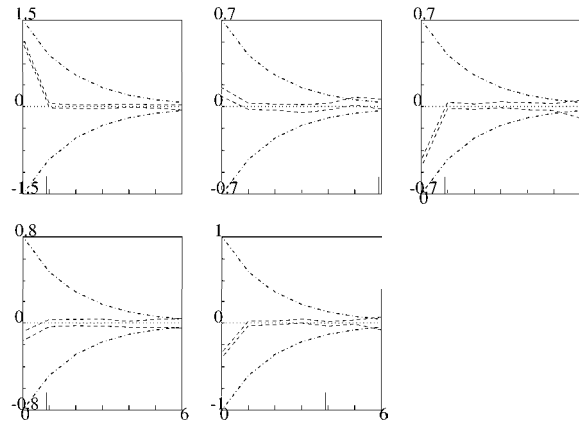


Fig. 5.16: Confidence interval of $\hat{\theta}(k)$ and estimated prior information, for $N = 1024$.

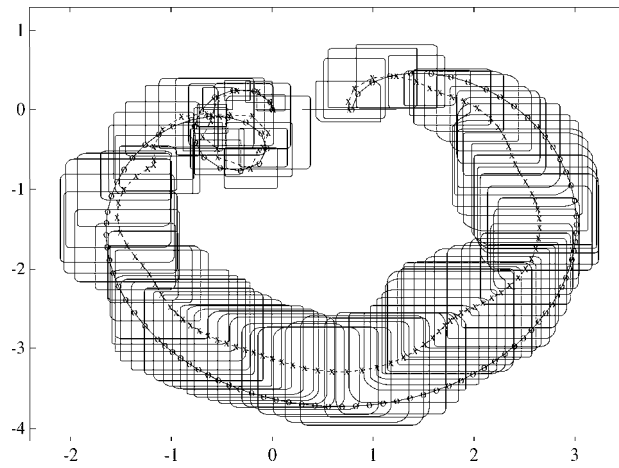


Fig. 5.17: $\hat{G}^f(e^{j\omega})$ with confidence interval, and transfer function of the linear part (5.84) of the nonlinear system, for $N = 1024$.

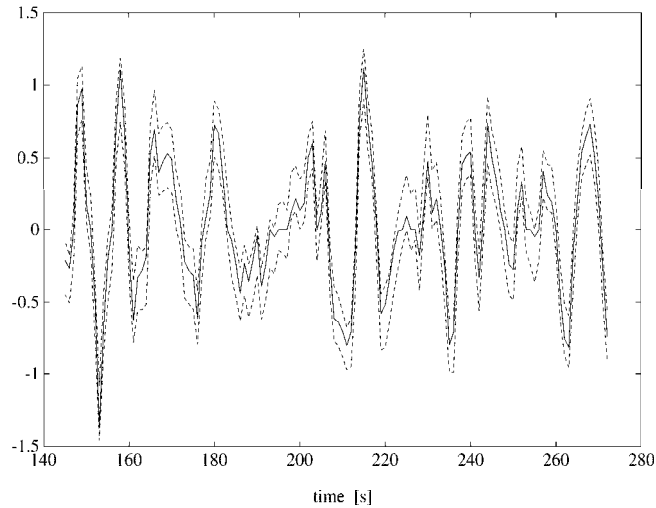


Fig. 5.18: Confidence interval of $\hat{y}(t)$ and output of the nonlinear system, for $N = 1024$.

and a part with nonvanishing memory, which will result in a chaotic, "noisy", output, even when excited with a periodic input. Hence, using the model error estimation techniques proposed, the part with nonvanishing memory will give rise to a variance (probabilistic) error, whereas the part with vanishing memory should result in a bias (deterministic) error.

Secondly, since the system is nonlinear, we cannot find a linear model which describes the observed data, even when the data is not noise disturbed (which is the situation which we are currently considering). Certainly this implies that the sequence of generalized expansion coefficients actually will not converge to zero.

It now follows that increasing the number of data points N will reduce the variance of the estimates, and hence the confidence intervals. However, this will also reveal that the undermodelling error is larger than we at first assumed: the confidence intervals for the estimated generalized expansion coefficients do no longer contain the value zero. This in turn implies that we should estimate a higher order model, which will result in an increase in the variance that will partly cancel the decrease in the variance due to the larger number of data points. Now, if the basis generating system is chosen properly (provides a good approximation of the linear part of the behaviour of the system), then the higher order generalized expansion coefficients will become small, so that zero again will lie inside the confidence interval for the tail of the higher order model. However, there is no longer a clear convergence to zero of the generalized expansion coefficients. This is due to the nonlinear part of the system that is structurally present

in the data (the part with vanishing memory), which is no longer obscured by the variance due to the nonstructural nonlinear effects in the data (due to the part of the nonlinear system with nonvanishing memory). The result is that the data indicates, by means of the fact that zero is not contained in the tails of the confidence intervals on the estimated generalized expansion coefficients, that there are structural effects left in the data (not captured by the model). These effects however cannot be adequately captured by increasing the order of the model since they are essentially nonlinear, but as a consequence will also yield only small values for the estimated generalized expansion coefficients. We now just increase the order of the model or the values of $\hat{\mathcal{K}}$ and $\hat{\eta}$ till the matching increase in the variance or bias of the estimated coefficients is such that zero lies again in the tails of the confidence intervals. It should be noted that when the model order now is again increased a smaller confidence interval may be obtained (when the decrease in the undermodelling error is larger than the increase in the variance error). This confidence interval may again invalidate the prior information \mathcal{K} and η (zero no longer lies in the confidence interval where formerly it did, or the confidence interval for the estimated additional coefficients does not contain zero). This may of course also happen for a linear time invariant system when the prior information is not correct, but for a nonlinear system this seems to be intrinsic since the sequence of generalized expansion coefficients will not converge to zero. Thus, the 'correct' values of $\hat{\mathcal{K}}$ and $\hat{\eta}$ also depend on the model order used.

To summarize, increasing the number of data points will reduce the variance error, but will lead to an increase in the error due to undermodelling, when the prior information \mathcal{K} , η is estimated from the data.

The mechanism described above is rather complex, but important when applying the model error estimation procedure in practice, see Section 8.2.3. A similar mechanism occurs when a periodic disturbance which has approximately the same period length as the input signal is acting on the system, or when the system is periodically time varying with a period length which is approximately equal to the period length of the input signal. In general, the mechanism described above will arise when some structural effect is present in the data which cannot be captured by a linear time invariant model with a stochastic noise disturbance.

We would like to stress that estimating the prior information from the data is in this case decisive to obtain good results, and that the 'correct' prior information changes with the number of data points and the model order used to estimate the error bounds.

Finally, it should be noted that no error due to undermodelling is present in the nonparametric estimate $\hat{G}(e^{j\omega_k})$, so that the above described mechanism does not arise. Indeed, when the number of averages r is large, the resulting error bounds for the nonparametric estimate are far too small. A similar situation arises when periodic disturbances, which have approximately the same period length as the input signal, act on the system. This also implies that one should not rely too much on the results obtained with a spectrum analyzer, although they often seem to be very accurate !

5.10 Application to time varying simulation systems

5.10.1 Simulation system 1

In this example we will study the behaviour of our model error estimation techniques for a time varying system. We will follow the simulation example that was presented in [83, Example 8.4], where a variance estimator for time varying systems is studied. The time varying system is given as

$$G_o(z, t) = \frac{b_1(t)z^{-1}}{1 + a_1(t)z^{-1} + a_2z^{-2}}$$

where $a_1(t)$ and $a_2(t)$ vary so that the poles move linearly forth and back on the arc $0.9e^{j\pm\alpha}$, $\alpha \in [\pi/4, 1.035]$, with a period length of 2000 points. The parameter $b_1(t)$ is such that the static gain is always unity. There is no noise acting on the system.

We will estimate a confidence interval for the transfer function of this system, using data over one complete period of the time varying dynamics. That is, we will use $N = 2048$, $N_o = 256$ and $N_s = 200$. The input signal was chosen to be random-phased multi-sine with the same amplitude for each $\omega_k \in \Omega_{N_o}$, and with $\bar{u} = \bar{u}^p = 1$. The probability level for all confidence regions is specified to be 99 %.

The basis generating system was estimated from the data, resulting in

$$G_b(z) = \frac{0.722z^{-1}}{1 - 1.1z^{-1} + 0.75z^{-2}} \quad (5.85)$$

The values for M , ρ , \mathcal{K} , and η are as usual obtained from the data iteratively, which resulted in $\hat{M} = 2$, $\hat{\rho} = 1.07$, $\hat{\mathcal{K}} = [3 \ 2]$, and $\hat{\eta} = 0.3$. These estimated values now are taken as the prior information of Assumption 5.2.1.ii and Assumption 5.2.1.iii.

We had $n_p = 9$ in Theorem 5.4.2, $n_f = 9$ in Theorem 5.4.5, and $n_y = 9$ in Theorem 5.4.8. The weighting matrix W for the generalized expansion coefficients of Theorem 5.4.2 was chosen to be $W = \text{diag}(1/\hat{\sigma}_r(\hat{G}(e^{j\omega_k})))$. The weighting matrix W for the transfer function estimate of Theorem 5.4.5, and weighting matrix W_y for the undisturbed time domain (impulse) response estimate of Theorem 5.4.8, were chosen as diagonal matrices where the diagonal has a similar form as the one that was depicted in Fig. 5.3.

Figure 5.19 shows the parametric transfer function estimate $\hat{G}^f(e^{j\omega})$ and the matching confidence interval of Theorem 5.4.5, together with a set of transfer functions which are samples of the time varying system at different time instants. These transfer functions are contained within the confidence interval. At first sight, the confidence intervals may seem to be far too large. This is however not the case, since the frequency axis changes radically over the different Nyquist plots. Figure 5.20 gives the error bound of the estimated impulse response according to Theorem 5.4.8, together with a set of impulse responses which are samples of the time varying system at different time instants. It follows that the error bound is fairly accurate, and certainly not far too large.

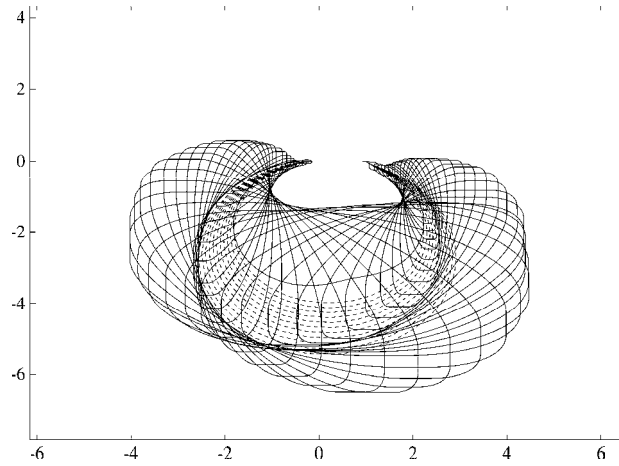


Fig. 5.19: $\hat{G}^f(e^{j\omega})$ with confidence interval (—), and Nyquist plots of the time varying system at different time instants (— —).

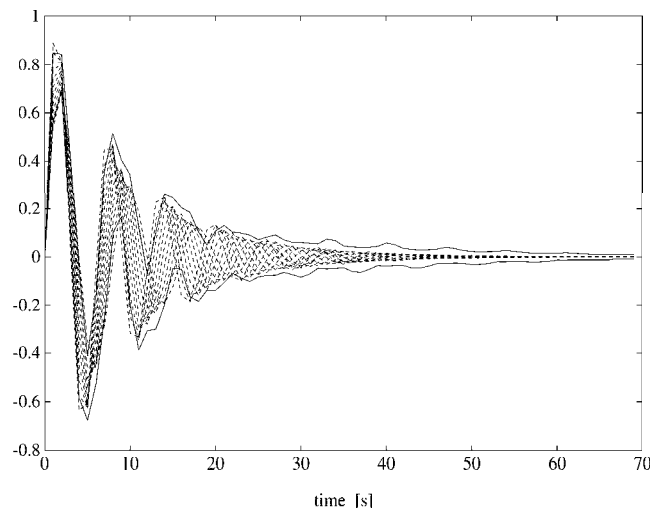


Fig. 5.20: Confidence interval of estimated impulse response (—), and impulse responses of the time varying system at different time instants (— —).

We will further test whether the error bound of Fig. 5.19 is realistic by designing two controllers for the system: a first controller that does not robustly stabilize the nominal model given by (5.85), and a second controller that does robustly stabilize the nominal model. We will check whether the first controller indeed does destabilize the true system, and whether the second controller does not. We design the two controllers using pole-placement, i.e. the controllers have the structure

$$C_R(q)u(t) = C_T(q)r(t) - C_S(q)y(t)$$

where $r(t)$ is the reference signal, and where $C_R(q)$, $C_T(q)$, $C_S(q)$ are polynomials. We also include an integrator to achieve tracking of a step in the reference signal. The first controller C_1 is given by

$$C_R = 1 - q^{-1} \quad C_T = 1.1219 \quad C_S = 2.6316 - 2.5485q^{-1} + 1.0388q^{-2}$$

and is designed to achieve a high closed loop bandwidth on the nominal model (both closed loop poles in 0.1). The second controller C_2 is given by

$$C_R = (1 - q^{-1})^2 \quad C_T = 0.0528 \quad C_S = 0.9723 - 2.3935q^{-1} + 2.5092q^{-2} - 1.0388q^{-3}$$

where we included a second integrator to achieve sufficient robustness.

For a linear time invariant system $G_o(z)$, the Nyquist criterion states the following. Given that the nominal closed loop (consisting of the nominal model $G_b(z)$ and the controller $C(z)$) is stable, then a sufficient condition for the actual closed loop (consisting of the system $G_o(z)$ and the controller $C(z)$) to be stable is

$$|G_o(e^{j\omega}) - G_b(e^{j\omega})| \leq \left| \frac{G_b(e^{j\omega})C(e^{j\omega}) + 1}{C(e^{j\omega})} \right|$$

Hence, we should have

$$\Delta(e^{j\omega}) \left| \frac{C(e^{j\omega})}{G_b(e^{j\omega})C(e^{j\omega}) + 1} \right| \leq 1 \quad (5.86)$$

where $\Delta(e^{j\omega})$ must be such that

$$|G_o(e^{j\omega}) - G_b(e^{j\omega})| \leq \Delta(e^{j\omega})$$

Although this criterion is true only for time invariant systems, it should be approximately valid in this example since the time variations are very slow. Therefore we will use this criterion to obtain an indication of the robustness properties of the designed controllers. For $\Delta(e^{j\omega})$ we will use the magnitude of the confidence intervals of $\hat{G}^f(e^{j\omega})$, plus $|G_b(e^{j\omega}) - \hat{G}^f(e^{j\omega})|$, which is known and relatively rather small.

To study the actual robustness properties of the time varying system $G_o(z, t)$ with respect to the nominal model $G_b(z)$ and the controller $C(z)$, we will use the criterion

$$|G_o(e^{j\omega}, t) - G_b(e^{j\omega})| \left| \frac{C(e^{j\omega})}{G_b(e^{j\omega})C(e^{j\omega}) + 1} \right| \leq 1$$

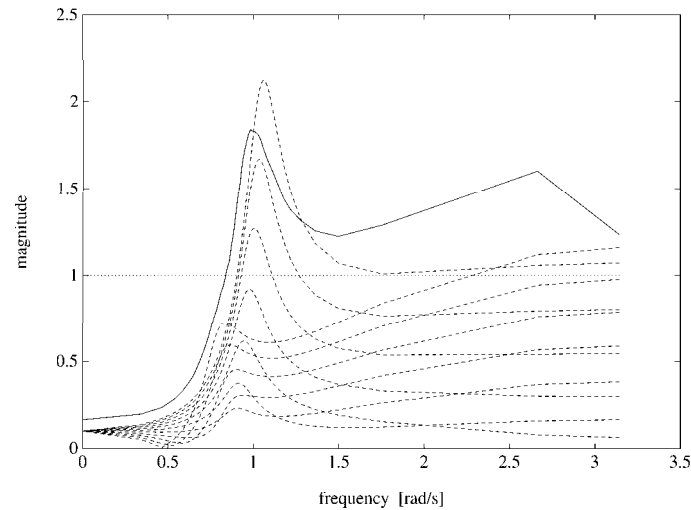


Fig. 5.21: Robustness analyses for C_1 using the estimated confidence interval (—) , and robustness analyses for C_1 using the time varying system at different time instants (- -) .

at the time instants $t = [0 \ 100 \ \dots \ 1000]$.

Figure 5.21 now depicts the behaviour of the left hand side of (5.86) for the controller C_1 , together with the robustness properties of the time varying system. Similarly, Fig. 5.22 depicts the results for the controller C_2 . Using only the estimated error bound, Fig. 5.21 indicates that the controller C_1 may violate the Nyquist criterion, so that we may expect this controller to have a bad performance on the true system, if the error bounds are indeed tight. Using only the estimated error bound, Fig. 5.22 indicates that the controller C_2 should stabilize the true system. Using the "transfer functions" of the time varying system, Fig. 5.21 and Fig. 5.22 also show that our error bounds are fairly realistic. The actual robustness properties may be a bit worse than indicated by the error bounds, but this is only the case over a short time interval. This time interval is only of minor importance in the data, so that we cannot expect it to be completely reflected in the error bound. For completeness, we note that the indication of Fig. 5.22 that the actual closed loop may become unstable turns out to be too pessimistic; for all time samples of the time varying system the loop is stable.

The responses of the actual closed loop systems to a square wave input are given in Fig. 5.23 for C_1 , and in Fig. 5.24 for C_2 . The performance achieved with C_1 indeed is very bad, whereas C_2 indeed is robust.

In conclusion, the confidence intervals which we obtained do provide realistic information as to the allowable uncertainty for control design in this case. If the number of

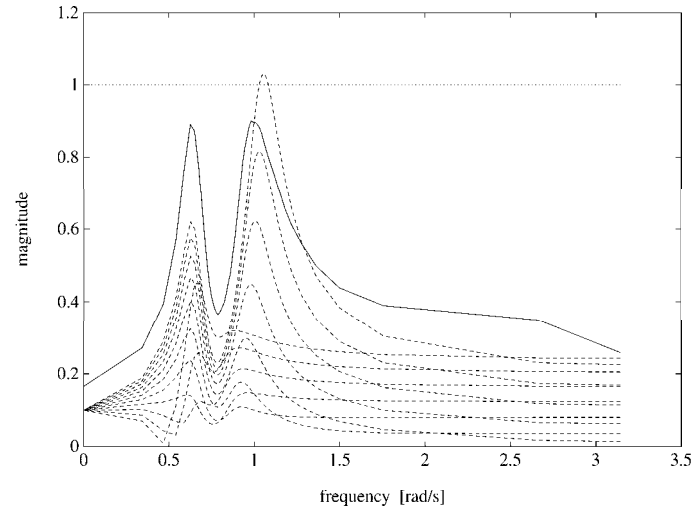


Fig. 5.22: Robustness analyses for C_2 using the estimated confidence interval (—) , and robustness analyses for C_2 using the time varying system at different time instants (- -) .

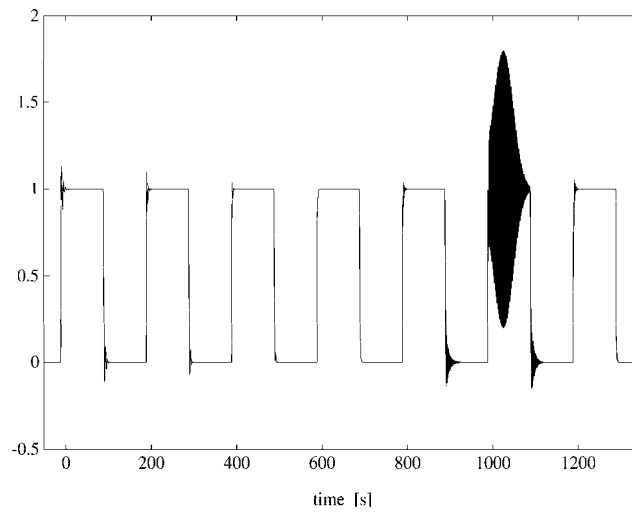


Fig. 5.23: Closed loop response to a square wave input under C_1 .

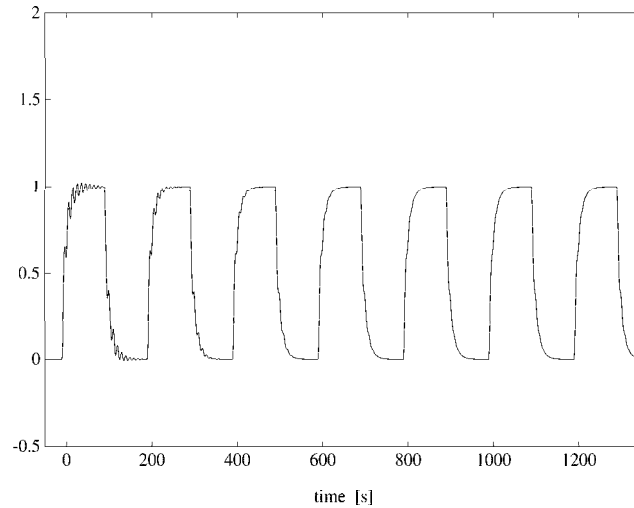


Fig. 5.24: Closed loop response to a square wave input under C_2 .

data points used for identification would be increased, then the estimates \hat{G}_i^f clearly will become highly correlated, since the time variations in the system are periodic. In this case the assumption that the "noise" on the different estimates is independent will be violated, which can be detected easily from the high "correlation" between deviations of the Nyquist plots of the different estimates from the average.

Finally it should be noted that it is dangerous to extrapolate the error bounds to conditions that differ considerably from the conditions under which the data was obtained. Figure 5.22 shows that the actual closed loop system could have been unstable, but only over a short time interval. In open loop this time interval clearly has no particularly strong influence on data, whereas in closed loop this can be radically different, if the closed loop becomes (almost) unstable over this time interval. The proper way to handle such a situation is to use cautious controller enhancement, so that specific features of the system slowly become more prominent in the data.

5.10.2 Simulation system 2

After the success story of the previous two sections (Section 5.10.1 and Section 5.9) regarding the performance of the proposed model error estimation method under deviations from the assumed linear time invariant case, we will now give a simulation example in which our method fails to give an accurate description of the behaviour of the system, although we will also argue that in view of (robust) control design this may be rather harmless. We will follow the simulation example that was presented in [83, Example 8.3], where a variance estimator for time varying systems is studied. The

data is generated as

$$y(t) = G(q, \theta_o)u(t) + \check{\theta}(t)\check{G}(q, \theta_o)u(t) + H_o(q)e(t)$$

with

$$\begin{aligned} G(q, \theta_o) &= \frac{0.3q^{-1}}{1 - 0.7q^{-1}} \\ \check{G}(q, \theta_o) &= \left. \frac{d}{d\theta} G(q, \theta) \right|_{\theta=\theta_o} = \frac{-0.3q^{-2}}{1 - 1.4q^{-1} + 0.49q^{-2}} \\ H_o(q) &= \frac{1}{1 - 0.7q^{-1}} \\ \check{\theta}(t+1) &= \begin{bmatrix} 0.9 & 0 \\ 0 & 0.9 \end{bmatrix} \check{\theta}(t) + w(t) \\ \mathbb{E}[w(t)w^T(t+\tau)] &= 10^{-4} \begin{bmatrix} 0.6 & 0.2 \\ 0.2 & 1 \end{bmatrix} \delta(\tau) \end{aligned}$$

where $\delta(\tau)$ is the Kronecker delta function. The noise sequence $e(t)$ is normally distributed white noise with zero mean and variance 0.002. The random vector sequence $w(t)$ also is normally distributed white noise with zero mean. The variance of the time varying term is approximately one fourth of the variance of the noise. To provide some more feeling for the nature of the time variations, the behaviour of $\theta_o + \check{\theta}(t)$ is depicted in Fig. 5.25.

The input signal is chosen to be periodic, by repeating a normally distributed white noise sequence with zero mean and unit variance. We choose $N = 512$, $N_o = 64$ and $N_s = 61$. The basis generating system was estimated from the data, resulting in

$$G_b(z) = \frac{0.009 + 0.308z^{-1}}{1 - 0.716z^{-1}}$$

The values for M , ρ , \mathcal{K} , and η are as usual obtained from the data iteratively, which resulted in $\hat{M} = 0.5$, $\hat{\rho} = 1.25$, $\hat{\mathcal{K}} = 0.5$, and $\hat{\eta} = 0.1$. These estimated values now are taken as the prior information of Assumption 5.2.1.ii and Assumption 5.2.1.iii.

We had $n_p = 4$ in Theorem 5.4.2, $n_f = 4$ in Theorem 5.4.5 and $n_y = 4$ in Theorem 5.4.8. The weighting matrix W for the generalized expansion coefficients of Theorem 5.4.2 was chosen to be $W = \text{diag}(1/\hat{\sigma}_r(\hat{G}(e^{j\omega_k})))$. The weighting matrix \mathcal{W} for the transfer function estimate of Theorem 5.4.5, and the weighting matrix \mathcal{W}_y for impulse response estimate of Theorem 5.4.8 was chosen as a diagonal matrix where the diagonal has a similar form as the one that was depicted in Fig. 5.3. The probability level for the confidence regions is specified to be 99 %.

Figure 5.26 shows the resulting confidence interval of $\hat{G}^f(e^{j\omega})$, together with the Nyquist plot of the time invariant term $G(q, \theta_o)$ of the system. Included also are the Nyquist plots of the time varying system at a number of time instants, and the

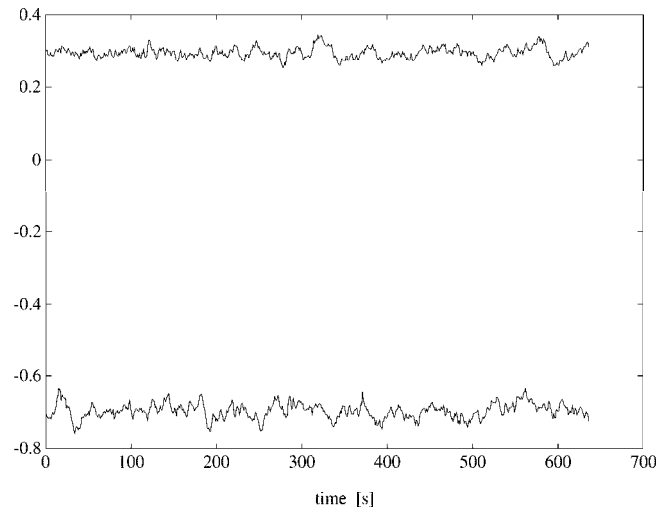


Fig. 5.25: Behaviour of $\theta_o + \check{\theta}(t)$ as a function of time.

confidence intervals for the true *frozen frequency function*¹. It follows that the estimated confidence intervals provide a reasonable indication of the variations in the time varying system, although they are too small for the higher frequencies.

When the number of data points used for identification is increased to $N = 2048$, the result becomes less acceptable. Figure 5.27 gives the resulting confidence interval of $\hat{G}^f(e^{j\omega})$, together with the Nyquist plot of the time invariant term $G(q, \theta_o)$ of the system, and the confidence intervals for the true frozen frequency function. Clearly, the estimated confidence intervals have deteriorated considerably, and no longer capture the behaviour of the true system. When the number of data points N goes to infinity, the estimated confidence intervals even will converge to zero. In contrast, the model variance estimator for time varying systems that is proposed in [83] does not exhibit this undesirable behaviour, and provides an accurate estimate of the variability in the true system, see [83, Example 8.3].

In conclusion, for long data sequences, our model error estimation procedure will no longer capture the random variations in the dynamics of the system. However, for control design this may be rather harmless. The output of the time varying term is only marginally correlated with the input signal, so that the random variations in the system are, for control design, more an issue of noise attenuation than robustness. Situations where the system differs in more structural ways from a linear time invariant system have been discussed in Section 5.10.1 and Section 5.9.

¹The transfer function obtained when keeping the time variable fixed

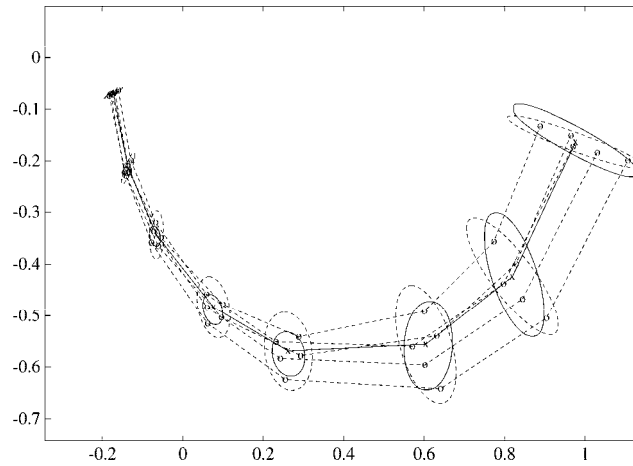


Fig. 5.26: Confidence interval of $\hat{G}^f(e^{j\omega})$ (—) for $N = 512$, Nyquist plots of the time varying system at $t = 100, 200, \dots, 500$ (- -), true 99 % confidence intervals (- -), and Nyquist plot of time invariant term (—) .

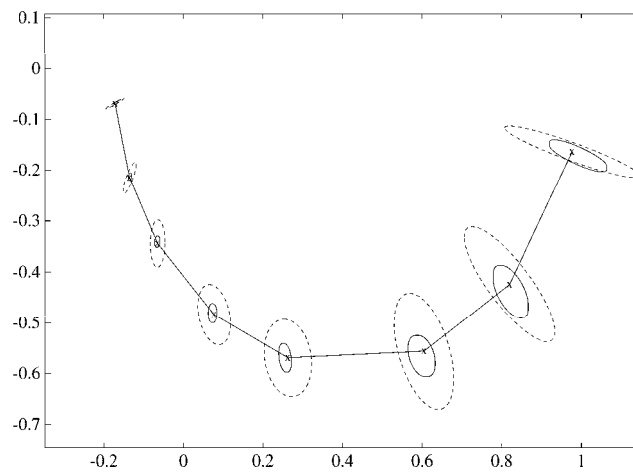


Fig. 5.27: Confidence interval of $\hat{G}^f(e^{j\omega})$ (—) for $N = 2048$, true 99 % confidence intervals (- -), and Nyquist plot of time invariant term (—) .

5.11 Conclusions

By applying a procedure similar to Bartlett's procedure of periodogram averaging in conjunction with a periodic input signal, a reliable and tight error bound has been established for a parametric transfer function estimate of a system for any frequency $\omega \in [0, 2\pi)$, while using only minor a priori information (an upper bound on the generalized expansion coefficients, on the impulse response, and on the past values of the input signal). The system is assumed to be linear and time invariant, and the additive noise on the output is assumed to be stochastic. System based orthonormal basis functions are employed to construct the parametric models. By means of these basis functions, prior information which is allowed to be inaccurate or approximative can be incorporated in the estimation procedures, so that an accurate low order description of the system, which is tuned to robust control design requirements, can be obtained. Additionally, we have obtained an error bound on the generalized expansion coefficients of the transfer function estimate, and on its response to an arbitrary input signal. These error bounds can be used to validate and improve the prior information on the generalized expansion coefficients and the impulse response, thus reducing the 'engineering leap of faith' involved in choosing this prior information. Moreover, the error bound on the response of the system to an arbitrary input signal can also be used to validate this a posteriori information, by comparing the predicted output and its error bounds to the measured output signal.

The results are based on the asymptotic normality of the DFT of a filtered sequence of independent identically distributed random variables, the use of a periodic input signal, and data segmentation. The error bounds are formulated in terms of confidence intervals which are valid asymptotically in the period length of the input signal. Simulations however display an excellent nonasymptotic performance (as demonstrated among others by a Monte Carlo study). By means of a number of simulation examples the procedure also is shown to be robust to mild violations of the assumptions on the system (linear and time invariant) and the noise (stochastic). The results of the model error estimation procedure proposed in this chapter additionally will be verified in practice in Chapter 8, corroborating the robustness and practical value of the method.

In the derivation an averaging (stochastic) setting for the noise was chosen, whereas the input signal outside the measurement interval and the effects of undermodelling were considered as being unknown-but-bounded (worst case). The fact that the variance of the noise is estimated from the data is accounted for in the asymptotic error distribution, and the estimate of the variance is consistent in the face of undermodelling.

In all cases, explicit and separate bounds are obtained for the different sources of uncertainty. This allows for an explicit tradeoff between the bias and variance contributions to the error by selecting the model order, and provides valuable insight in the effective ways to improve the error bound, when the error bound is too large in view of the robust control design requirements.

The main contributions of this chapter are that the error bounds obtained indeed appear to be reliable and tight, that the model to which the error bounds refer can be chosen to be control-relevant by choosing a suitable basis generating system to construct the orthonormal basis functions, and that the error bounds can be tuned accurately to robust control design specifications by input design. As a consequence, the models and model error bounds are suitable for high performance robust control design purposes. Hence, we conclude that the mixed averaging – structural embedding of the modelling errors is successful, and that data segmentation provides a powerful means to estimate the error components.

Chapter 6

Quantification of uncertainty in closed loop

Summary

Using a coprime factor identification framework, approximative identification and model uncertainty estimation can be performed in closed loop and for unstable systems. An additional error due to unknown past inputs arises, and an upper bound on this error is established.

6.1 Framework

In this chapter we focus on the problem of approximative identification and model uncertainty estimation in closed loop. The motivation to address this problem is twofold. Firstly, we note that in practice many systems must operate in closed loop during the measurements, for economic or safety reasons, or because the open loop system is unstable, or because the open loop system is nonlinear and must be kept in a working point. Secondly, in [11, 12, 3, 155, 156, 69, 101, 5, 48] it is motivated that identification for control design requires identification in closed loop. To solve the problem of model uncertainty estimation in closed loop we will use a coprime factor description of the system.

For the classical theory (identification techniques and identifiability conditions) concerning identification of systems in closed loop we refer to [165, 66, 106, 166, 173]. For the theory of coprime factor descriptions, and their use in control design, see e.g. [178, 179, 120, 121, 52, 47, 17, 13, 16]. The use of coprime factor representations of the system for identification was introduced by [72, 71, 73], and elaborated extensively by Schrama, see [154, 153, 157, 155, 156]. This framework is also used for system identification by [169]. The specific advantages of identifying *normalized* coprime factors of the system are exploited in [175]. The reasons for choosing this framework are the following.

- The closed loop identification problem can be recasted into two open loop identification problems, see [154]. For these two open loop identification problems the theory of uncertainty quantification developed in Chapter 4 and Chapter 5 can be used directly.
- The resulting coprime factor uncertainty description for the system can be used in robust control design, and is fairly general and nonconservative (encompasses additive and multiplicative uncertainty), see [179, 52, 120, 47, 17, 13, 16]. The more usual "standard plant" approach to robust control design, see [39, 38, 40, 138, 27], can of course also be used.

Further useful properties of the coprime factor framework are the following.

- An iterative scheme for control-relevant identification of coprime plant factors has been elaborated extensively in [157, 155, 156, 175].
- The coprime factor framework results in an errors-in-variables framework (noise on input and output), and has a clear relation with spectral analysis (numerator and denominator are estimated separately) and the instrumental variable approach (a filtered version of the reference signals is used as instrumental variable). The coprime factor framework also provides a generalization of the indirect identification approach where the transfer functions from the reference signal to the output and to the input signal of the system are estimated separately first, and next the quotient is taken to obtain an estimate of the system.
- The identification and model uncertainty estimation of unstable systems does not pose any problem when using the coprime factor framework.
- The coprime factor framework can handle unstable perturbations of the system in robustness analysis and robust control design.
- Convergence of the estimated coprime factors to the true factors (an open loop property) implies convergence with respect to closed loop properties (convergence in the graph metric, see [178, 179]).

The coprime factor framework can also be used for identification in open loop (just substitute $C = 0$ for the controller C).

6.2 Coprime factor representation

We consider the feedback configuration of Fig. 6.1. The *possibly unstable* system P_o is assumed to be LTIFD (linear time invariant and finite dimensional). The noise filter H_o is assumed to be stable and LTI. The noise $e(t)$ is assumed to be a sequence of independent identically distributed random variables that is uncorrelated with the

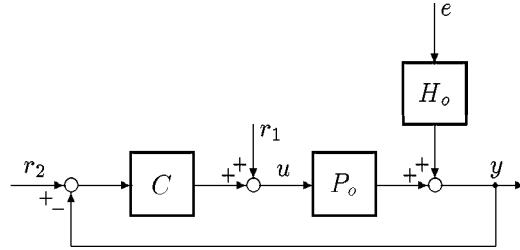


Fig. 6.1: Feedback configuration.

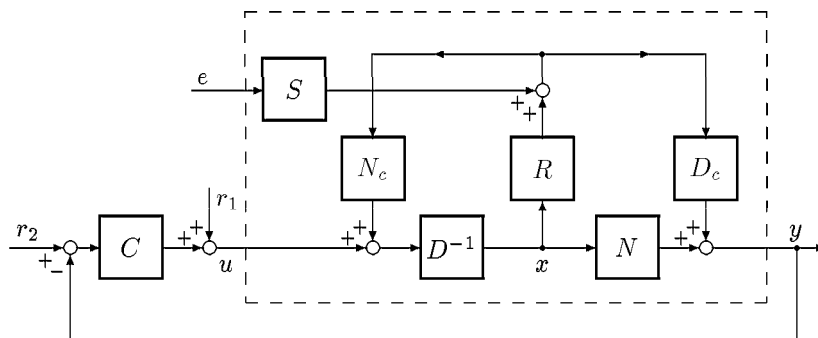


Fig. 6.2: Extended dual Youla parameterization.

external signals $r_1(t)$ and $r_2(t)$. At least one of the external signals $r_1(t)$ and $r_2(t)$ is required to be an excitation signal that can be specified by the user, the other one (if applicable) can be a disturbance signal or zero. The (possibly unstable) controller C should be LTIFD. The feedback system is required to be stable, and the controller should be known.

The feedback system is said to be stable if its transfer function mapping $[e \ r_2 \ r_1]^T$ into $[y \ u]^T$ belongs to \mathbb{RH}_∞ .

Let P be a LTIFD nominal model that is stabilized by the controller C . The system, the nominal model and the controller admit right coprime factorizations over the ring \mathbb{RH}_∞ , see [179],

$$\begin{aligned} P_o &= N_o D_o^{-1} \\ P &= N D^{-1} \\ C &= N_c D_c^{-1} \end{aligned}$$

Using the extended Youla parameterization of all systems that are stabilized by the controller C , the feedback system of Fig. 6.1 can be redrawn to Fig. 6.2, as was shown by [154, 156]. There holds, see [154, 156],

$$N_c^{-1}(Dx - u) = Rx + Se \quad (6.1)$$

$$D_c^{-1}(y - Nx) = Rx + Se \quad (6.2)$$

where $R, S \in \text{IRH}_\infty$, and where $x(t)$ is a known intermediate signal

$$\begin{aligned} x &= (D + CN)^{-1}(u + Cy) \\ &= (D + CN)^{-1}(r_1 + Cr_2) \end{aligned} \quad (6.3)$$

Note that $x(t)$ is not affected by the noise $e(t)$, while $u(t)$ and $y(t)$ are. Equivalently

$$u = D_o x - N_c S e \quad (6.4)$$

$$y = N_o x + D_c S e \quad (6.5)$$

where

$$D_o = D - N_c R$$

$$N_o = N + D_c R$$

Because $x(t)$ is uncorrelated with the noise $e(t)$ it is now possible to estimate R or D_o and N_o in an open loop fashion from $x(t)$, $u(t)$ and $y(t)$, using the known controller C .

6.3 Error quantification

The aim of this section is to find a complete error quantification for the closed loop approximate identification problem. Using the coprime factor description introduced above, this turns out not to be too difficult. It can be seen easily that all components of the error bounds of Chapter 4 and Chapter 5 remain unchanged, but that additional error terms arise because the input $x(t)$ is no longer exactly periodic for $t \geq 0$ when the excitation signals are periodic only for $t \geq 0$.

Note that the unknown past of the external excitation signals gives rise to two kinds of errors: the usual errors due to the unknown past ($t < 0$) of the input $x(t)$ to the system, which is already accounted for in the error bounds of Chapter 4 and Chapter 5, and errors due to the fact that the input $x(t)$ is no longer exactly periodic for $t \geq 0$ when the excitation signals are periodic only for $t \geq 0$. In this section we will derive an upper bound on the error due to the unknown past of the external excitation signal, which should replace the error terms due to the unknown past of the input signal in Chapter 4 and Chapter 5.

We will consider only scalar (single input, single output), linear time invariant, discrete time systems. The problem that we need to solve now can be formulated as follows. The signals are generated as, see (6.3), (6.1), (6.2), (6.4), (6.5),

$$x(t) = F(q)r(t) \quad (6.6)$$

$$z(t) = G_o(q)x(t) \quad (6.7)$$

where $F(q)$ is known, $r(t)$ is periodic for $t > 0$, and $G_o(q)$ unknown. The signal $r(t)$ can be $r_1(t)$ or $r_2(t)$. If both $r_1(t)$ and $r_2(t)$ are used as excitation signals, upper bounds for the errors due to $r_1(t)$ and $r_2(t)$ should be calculated separately, and combined. Define

$$\begin{aligned} R^s(e^{j\omega_k}) &= \sum_{t=0}^{N-1} r(t + N_s) e^{-j\omega_k t} \quad \omega_k \in \Omega_{N_o} \\ X_p^s(e^{j\omega_k}) &= F(e^{j\omega_k}) R^s(e^{j\omega_k}) \quad \omega_k \in \Omega_{N_o} \\ \Omega_{N_o} &= \left\{ \omega_k = \frac{2\pi k}{N_o}, k = 0, 1, \dots, N_o - 1 \right\} \end{aligned}$$

Note that $X_p^s(e^{j\omega_k}) \neq X^s(e^{j\omega_k})$. We want to use $X_p^s(e^{j\omega_k})$ to estimate the system $G_o(q)$, and not $X^s(e^{j\omega_k})$, in order to be able to employ the results of Chapter 5, which assumes the input signal to be periodic.

We now want to find an upper bound on $|E^s(e^{j\omega_k})|$, where $E^s(e^{j\omega_k})$ is the error term in the expression

$$Z^s(e^{j\omega_k}) = G_o(e^{j\omega_k}) X_p^s(e^{j\omega_k}) + E^s(e^{j\omega_k}) \quad (6.8)$$

given that $z(t) = G_o(q)x(t)$.

Assumption 6.3.1 *We have as a priori information that*

1. *there exists a finite and known $\bar{r}^p \in \mathbb{R}$ such that $|r(t)| \leq \bar{r}^p$ for all $t < 0$.*
2. *there exist finite and known $M, \rho \in \mathbb{R}$, with $\rho > 1$, such that $|g_o(k)| \leq M\rho^{-k}$ for $k \in \mathbb{N}$.*

Define

$$\bar{r} = \max_{t \in T^N} |r(t)|$$

and let $f(k)$ denote the impulse response of the filter $F(q)$.

Assumption 6.3.2 *The following conditions are assumed to be satisfied*

1. *$|f(k)| \leq B\mu^{-k}$ with $B, \mu \in \mathbb{R}$, B finite and $\mu > 1$.*
2. *$r(t)$ is periodic for $t \in T^{N+N_s}$, with period N_o .*
3. *$N = kN_o$, $k \in \mathbb{N}$.*

When $F(q) \in \text{IRH}_\infty$, which implies that $F(q)$ is exponentially stable, a B and μ exists and can be found easily from the impulse response $|f(k)|$ and the spectral radius of the system matrix of the known filter $F(q)$.

Theorem 6.3.3 *Let Assumption 6.3.1 and Assumption 6.3.2 hold, and let $\omega_k \in \Omega_{N_o}$. Then the error term $E^s(e^{j\omega_k})$ in (6.8) can be bounded as*

$$|E^s(e^{j\omega_k})| \leq (\bar{r}^p + \bar{r}) \frac{MB\mu}{\mu-1} \left[\frac{\mathcal{U}^s}{\mu-1} + (1-\rho^{-N}) \frac{\rho^{-N_s}}{\rho-1} \left(\frac{1}{\mu-1} + \frac{\rho}{\rho-1} \right) \right]$$

where

$$\mathcal{U}^s = \begin{cases} (N_s + 1)\mu^{-N_s} - (N_s + N + 1)\mu^{-N_s - N} & \text{for } \rho = \mu \\ \frac{\rho\mu}{|\rho-\mu|} |(1-\mu^{-N})\mu^{-N_s-1} - (1-\rho^{-N})\rho^{-N_s-1}| & \text{for } \rho \neq \mu \end{cases}$$

Proof: See Appendix C.1. □

Note that the upper bound on $|E^s(e^{j\omega_k})|$ given in Theorem 6.3.3 converges to zero exponentially in N_s .

We can retrieve the open loop situation by substituting $N_c(q) \equiv 0$ and $D_c(q) \equiv 1$, yielding $C(q) \equiv 0$. Note that in this situation we have $u(t) = D_o(q)x(t)$, with $D_o(q) = D(q)$.

The open loop situation with $u(t) = x(t) = r_1(t)$ can be retrieved by substituting $N_c(q) \equiv 0$, $D_c(q) \equiv 1$ and $D_o(q) \equiv 1$, yielding $C(q) \equiv 0$, $P_o(q) \equiv N_o(q)$, $F(q) \equiv 1$ and $u(t) = x(t) = r_1(t)$. Hence the open loop situation with $u(t) = x(t) = r_1(t)$ is obtained by substituting the discrete impulse for $f(m)$, i.e. $B = 1$ and taking the limit $\mu \rightarrow \infty$, which results in

$$|E^s(e^{j\omega_k})| \leq (\bar{u}^p + \bar{u})M\rho \frac{1-\rho^{-N}}{(\rho-1)^2} \rho^{-N_s} \quad (6.9)$$

compare (A.11) in Appendix A.1.

Chapter 7

Asymptotic bias and variance expressions

Summary

An asymptotic analysis is provided for least squares frequency domain transfer function estimation using system based orthonormal basis functions, as employed in the model error estimation procedure of Chapter 5. That is, the statistical properties of the estimated parameters and the estimated transfer function are analyzed, asymptotically in the number of data points. The estimates are shown to be consistent when no weighting is used in the least squares parameter estimate. That is, consistent estimates of the parameterized parts of the system are obtained. Explicit and transparent bias and variance expressions are established, which are not asymptotic in the model order. These expressions provide insight in the properties of the transfer function estimation method employed in Chapter 5, and can be used to make effective choices for the basis functions and the input signal.

7.1 Introduction

The aim of this chapter is to establish asymptotic (in the number of data points) bias and variance expressions for frequency domain least squares identification using system based orthonormal basis functions. Such asymptotic expressions are valuable for input design, model order selection, and the selection of the basis generating system, in the model error estimation procedure presented in Chapter 5.

To be more specific, from Chapter 5 we have available a nonparametric frequency domain estimate \hat{G} of the system G_o , defined as an average over a number (r) of empirical transfer function estimates (ETFE's) G_i over subsequent data segments (of

length N_o)

$$\begin{aligned}\hat{G}_i(e^{j\omega_k}) &:= \frac{Y_i(e^{j\omega_k})}{U_i(e^{j\omega_k})} \quad \omega_k \in \Omega_{N_o}^u \\ \hat{G}(e^{j\omega_k}) &:= \frac{1}{r} \sum_{i=1}^r \hat{G}_i(e^{j\omega_k}) \quad \omega_k \in \Omega_{N_o}^u\end{aligned}\quad (7.1)$$

see (5.13), (5.12). Using the nonparametric frequency domain estimate $\hat{G}(e^{j\omega_k})$, a least squares parameter estimate is obtained in Chapter 5 as

$$\hat{\theta} := \arg \min_{\theta} \frac{1}{p} \sum_{k=1}^p \left| \hat{G}(e^{j\omega_k}) - \phi(e^{j\omega_k})\theta \right|^2 \quad (7.2)$$

where

$$\phi(e^{j\omega_k}) = [1 \quad \mathcal{V}_0^T(e^{j\omega_k}) \quad \dots \quad \mathcal{V}_{n-1}^T(e^{j\omega_k})]$$

is a vector of frequency domain basis functions, see (7.8). Subsequently, using the estimated parameters $\hat{\theta}$, a transfer function estimate is obtained as

$$\hat{G}^f(e^{j\omega}) := \phi(e^{j\omega})\hat{\theta} \quad (7.3)$$

The parameter estimate (7.2) and transfer function estimate (7.3) are used in the model uncertainty estimation procedures of Chapter 5. In this chapter we will investigate the asymptotic behaviour of these estimates. That is, we will establish asymptotic bias and variance expressions for these estimates.

We will show that, even in the presence of undermodelling, the estimated parameters are asymptotically unbiased. That is, the estimate of that part of the system which is parameterized is asymptotically unbiased. A simple expression for the asymptotic bias in the transfer function estimate is given, which shows that the bias converges exponentially to zero. Requiring the input signal $u(t)$ to be periodic for $t \in T^{N+N_s}$, we obtain tractable and transparent expressions for the asymptotic variance of the estimated parameters and the estimated transfer function, for finite model orders.

The remainder of this chapter is organized as follows. First, we will list the assumptions. Subsequently we will address the asymptotic properties of the parameter estimate, and finally we will discuss the asymptotic behaviour of the transfer function estimate.

7.2 Assumptions

It is assumed that the system, and the measurement data that is obtained from this system, allow a description

$$y(t) = G_o(q)u(t) + v(t) \quad (7.4)$$

$$G_o(z) = \sum_{k=0}^{\infty} g_o(k)z^{-k} \quad (7.5)$$

$$v(t) = H_o(q)e(t) \quad (7.6)$$

$$\Phi_v(\omega) = \sigma_e^2 |H_o(e^{j\omega})|^2 \quad (7.7)$$

where G_o is the transfer function of the system, being proper, exponentially stable, and finite dimensional, where $e(t)$ is a sequence of independent identically distributed random variables, having zero mean, variance σ_e^2 , and bounded moments of all orders, where H_o is a stable proper transfer function, and where Φ_v denotes the spectrum of the noise $v(t)$.

Using the system based orthonormal basis functions introduced by [80, 81], see Appendix E, we can also write

$$y(t) = Du(t) + \sum_{k=0}^{\infty} L_k \mathcal{V}_k(q)u(t) + v(t) = Du(t) + \sum_{k=0}^{\infty} L_k \mathcal{X}_k(t) + v(t)$$

$$G_o(z) = D + \sum_{k=0}^{\infty} L_k \mathcal{V}_k(z) \quad (7.8)$$

$$\mathcal{V}_k(z) = \mathcal{V}_0(z) \mathbf{G}_b^k(z) = \sum_{t=0}^{\infty} b_k(t)z^{-t} \quad (7.9)$$

$$\mathbf{B} = \begin{bmatrix} b_0(0) & b_0(1) & \cdots \\ b_1(0) & b_1(1) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \quad (7.10)$$

where $\mathbf{G}_b \in \mathbb{RH}_2$ is a scalar inner function with McMillan degree n_b , $\mathcal{V}_k \in H_2^{n_b \times 1}$, $L_k \in \mathbb{R}^{1 \times n_b}$, and $b_k \in \ell_2^{n_b \times 1}[0, \infty)$ is a basis function. The basis generating system G_b is required to be finite dimensional and exponentially stable, implying that $\mathcal{V}_k(z)$ is continuous when k is finite, and is assumed to be given a priori, implying that n_b is fixed and finite.

Let T^N denote the integer interval $\mathbb{Z} \cap [0, N-1]$. For a signal $x(t)$ being defined on $T_{N_s}^N$, we will denote

$$X^s(e^{j\omega_k}) := \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x(t + N_s) e^{-j\omega_k t} \quad \omega_k \in \Omega_N \quad (7.11)$$

$$X_o(e^{j\omega_k}) := \frac{1}{\sqrt{N_o}} \sum_{t=0}^{N_o-1} x(t + N_s) e^{-j\omega_k t} \quad \omega_k \in \Omega_{N_o} \quad (7.12)$$

$$X_i(e^{j\omega_k}) := \frac{1}{\sqrt{N_o}} \sum_{t=0}^{N_o-1} x(t + (i-1)N_o + N_s) e^{-j\omega_k t} \quad \omega_k \in \Omega_{N_o} \quad (7.13)$$

where $i = 1, 2, \dots, r$. Specific sets of frequency points that arise are denoted as

$$\begin{aligned}\Omega_N &:= \left\{ \frac{2\pi k}{N}, k = 0, 1, \dots, N-1 \right\} \\ \Omega_N^u &:= \{ \omega_k \in \Omega_N \mid |U^s(e^{j\omega_k})| > 0 \}\end{aligned}\quad (7.14)$$

The input signal $u(t)$ is assumed to be periodic for $t \in T^{N+N_s}$ with period length N_o , where $N = rN_o$, $r \in \mathbb{N}$, $r \geq 1$. Note that $N_o = N$ and $N_s = 0$ are allowed. The data set used for identification has length N , being the last N points of the time interval T^{N+N_s} . The first N_s points of this time interval serve to let decay out the effects of the unknown past of the input signal (input for $t < 0$).

Additionally, to ensure that our limit expressions are well-defined, we will assume that the input signal is such that

$$\lim_{N_o \rightarrow \infty} |U_o(e^{j\omega_k})|^2 < \infty \quad \forall \quad \omega_k \in \Omega_{N_o}$$

Finally, we need the following assumption on the input signal, to be able to arrive at more elegant expressions (i.e. to be able to use integral expressions in the limit as $N_o \rightarrow \infty$ for sum expressions over k that contain $|U_o(e^{j\omega_k})|^{-2}$, see e.g. Theorem 7.3.5).

Assumption 7.2.1 *It is assumed that the input signal within one period is such that there exists a function $\Phi_u(\omega)$ such that*

$$\lim_{N_o \rightarrow \infty} |U_o(e^{j\omega_k})|^2 = \Phi_u(\omega_k) \quad \forall \quad \omega_k \in \Omega_{N_o}^u \quad (7.15)$$

where $\Phi_u^{-1}(\omega)$, $\omega \in [0, 2\pi)$, is continuous, except for at most a finite number of points, and bounded.

This situation arises e.g. when the excitation signal $u(t)$ is designed as a filtered sum of N_o sinewaves with the same amplitude, given that the filter is inversely stable.

7.3 Asymptotic analysis of parameter estimates

This section is organized as follows. We will first establish expressions for the bias and covariance of the estimated parameters, starting from the nonparametric frequency domain estimate \hat{G} . Subsequently, we will study the asymptotic behaviour of these expressions, where we will first address the covariance, and next the bias.

Because the input signal is assumed to be periodic, it follows that

$$U_i(e^{j\omega_k}) = U_o(e^{j\omega_k}) \quad \forall \quad i = 1, 2, \dots, r$$

compare (7.13) and (7.12). Straightforward calculation shows that

$$U^s(e^{j\omega_k}) = \begin{cases} \sqrt{r}U_o(e^{j\omega_k}) & \text{for } \omega_k \in \Omega_{N_o} \\ 0 & \text{for } \omega_k \in \Omega_N \setminus \Omega_{N_o} \end{cases} \quad (7.16)$$

so that Ω_N^u (7.14) equals

$$\Omega_N^u = \Omega_{N_o}^u = \{\omega_k \in \Omega_{N_o} \mid |U_o(e^{j\omega_k})| > 0\} \quad (7.17)$$

Hence, since the input signal is assumed to be periodic, the nonparametric frequency domain estimate \hat{G} (7.1) equals

$$\hat{G}(e^{j\omega_k}) := \frac{1}{r} \sum_{i=1}^r \frac{Y_i(e^{j\omega_k})}{U_i(e^{j\omega_k})} = \frac{Y^s(e^{j\omega_k})}{U^s(e^{j\omega_k})} \quad \omega_k \in \Omega_{N_o}^u \quad (7.18)$$

see (B.1), (7.16) and (7.17). Similar to (5.14) of Chapter 5 we now can express \hat{G} as

$$\hat{G}(e^{j\omega_k}) = G_o(e^{j\omega_k}) + S(e^{j\omega_k}) + F(e^{j\omega_k}) \quad (7.19)$$

with

$$F(e^{j\omega_k}) = \frac{V^s(e^{j\omega_k})}{U^s(e^{j\omega_k})} \quad (7.20)$$

and $S(e^{j\omega_k})$ a term due to the past of the input signal (see (5.14)). The true transfer function G_o can be written as

$$\begin{aligned} G_o(e^{j\omega_k}) &= D + \sum_{k=0}^{\infty} L_k \mathcal{V}_k(e^{j\omega_k}) \\ &= D + \sum_{k=0}^{n-1} L_k \mathcal{V}_k(e^{j\omega_k}) + \sum_{k=n}^{\infty} L_k \mathcal{V}_k(e^{j\omega_k}) \\ &= \phi(e^{j\omega_k})\theta_o + Z(e^{j\omega_k}) \end{aligned} \quad (7.21)$$

with

$$\phi(e^{j\omega_k}) = [1 \quad \mathcal{V}_0^T(e^{j\omega_k}) \quad \dots \quad \mathcal{V}_{n-1}^T(e^{j\omega_k})] \quad (7.22)$$

$$\theta_o = [D \quad L_0 \quad \dots \quad L_{n-1}]^T \quad (7.23)$$

$$Z(e^{j\omega_k}) = \sum_{k=n}^{\infty} L_k \mathcal{V}_k(e^{j\omega_k}) \quad (7.24)$$

Let the set $\Omega_{N_o}^u$, as given by (7.17), consist of the following p elements

$$\Omega_{N_o}^u = \{\omega_1, \omega_2, \dots, \omega_p\}$$

Combining (7.19) and (7.21), we now can write in vector-matrix notation

$$\begin{aligned} \hat{G} &= G_o + S + F \\ &= \Phi\theta_o + Z + S + F \end{aligned} \quad (7.25)$$

with

$$\begin{aligned}
\hat{G} &= [\hat{G}(e^{-j\omega_1}) \dots \hat{G}(e^{-j\omega_p})]^T \\
G_o &= [G_o(e^{-j\omega_1}) \dots G_o(e^{-j\omega_p})]^T \\
S &= [S(e^{-j\omega_1}) \dots S(e^{-j\omega_p})]^T \\
F &= [F(e^{-j\omega_1}) \dots F(e^{-j\omega_p})]^T \\
Z &= [Z(e^{-j\omega_1}) \dots Z(e^{-j\omega_p})]^T \\
\Phi &= \begin{bmatrix} 1 & \mathcal{V}_0^T(e^{j\omega_1}) & \dots & \mathcal{V}_{n-1}^T(e^{j\omega_1}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathcal{V}_0^T(e^{j\omega_p}) & \dots & \mathcal{V}_{n-1}^T(e^{j\omega_p}) \end{bmatrix} \tag{7.26}
\end{aligned}$$

where Z reflects the error due to undermodelling, S represents the error due to unknown past inputs, and F reflects the error due to the noise. The least squares parameter estimate now can be formulated as follows.

Fact 7.3.1 *The least squares parameter estimate is defined as*

$$\hat{\theta} := \arg \min_{\theta} \frac{1}{p} \sum_{k=1}^p \left| \hat{G}(e^{j\omega_k}) - \phi(e^{j\omega_k})\theta \right|^2 \tag{7.27}$$

where θ ranges over an appropriate parameter space $\Theta \subset \mathbb{R}^{n_p}$,

$$\hat{\theta} = [\hat{D} \ \hat{L}_0 \ \dots \ \hat{L}_{n-1}]^T \tag{7.28}$$

$$n_p := \dim(\hat{\theta}) = n_b n + 1$$

and $n < n_p \leq p \leq N_o$. Alternatively, this estimate can be written as

$$\hat{\theta} = \Psi \hat{G} \tag{7.29}$$

where

$$\Psi = (\Phi^* \Phi)^{-1} \Phi^* \tag{7.30}$$

Remarks

1. The least squares criterion (7.27) implies that more attention is paid to those frequency regions which have a relatively denser frequency grid $\omega_k \in \Omega_{N_o}^u$.
2. The estimate $\hat{\theta}$ exists and is unique if Φ is 'persistently exciting of order $\dim(\hat{\theta})$ ', i.e. when $\Phi^* \Phi$ is nonsingular. Note that for frequency domain parameter estimation the notion of 'persistence of excitation' is a property of the set of basis functions that is used, and of $\Omega_{N_o}^u$ (the frequency contents of the input signal). Clearly, $n_p \leq p$ is necessary for Φ to be 'persistently exciting of order n_p '.

From (7.29), (7.25) and (7.30) it follows that the least squares parameter estimate (7.27) can be expressed as

$$\hat{\theta} = \theta_o + \Psi Z + \Psi S + \Psi F$$

Since we assumed that $\mathbb{E}[e(t)] = 0$ it follows from (7.20), (7.11) and (7.6) that

$$\begin{aligned} \mathbb{E}[F(e^{j\omega_k})] &= \frac{\mathbb{E}[V^s(e^{j\omega_k})]}{U^s(e^{j\omega_k})} \\ &= \frac{1}{U^s(e^{j\omega_k})} \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \mathbb{E}[v(t + N_s)] e^{-j\omega_k t} \\ &= \frac{1}{U^s(e^{j\omega_k})} \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \mathbb{E}[H_o(q)e(t + N_s)] e^{-j\omega_k t} \\ &= 0 \quad \forall \quad \omega_k \in \Omega_{N_s}^u \end{aligned}$$

Hence, for the bias and covariance of the estimated parameters (7.27) we find

$$\mathbb{E}[\hat{\theta}] = \theta_o + \Psi Z + \Psi S \quad (7.31)$$

$$\text{cov}[\hat{\theta}] := \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\hat{\theta} - \mathbb{E}[\hat{\theta}])^T] = \Psi \mathbb{E}[\text{FF}^*] \Psi^* \quad (7.32)$$

We now will first study the asymptotic behaviour of the covariance matrix (7.32), and next the asymptotic behaviour of the bias (7.31).

Lemma 7.3.2 *For the covariance matrix of $F(e^{j\omega_k})$ (7.20) there holds that*

$$\frac{N}{N_o} \mathbb{E}[\text{FF}^*] \rightarrow Q = \begin{bmatrix} \frac{\Phi_v(\omega_1)}{|U_o(e^{j\omega_1})|^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\Phi_v(\omega_p)}{|U_o(e^{j\omega_p})|^2} \end{bmatrix} \quad \text{as } N \rightarrow \infty \quad (7.33)$$

where the rate of convergence is $1/N$ element wise.

Additionally, there exists a constant $c_1 \in \mathbb{R}$ such that

$$\left\| \frac{N}{N_o} \mathbb{E}[\text{FF}^*] - Q \right\|_2 \leq c_1 \frac{p}{N}$$

Proof: See Appendix D.1. □

Using Lemma 7.3.2 we can state the following preliminary result for the covariance of the estimated parameters, where we will address the i -th element of the vector of estimated parameters $\hat{\theta}$ (7.28) as $\hat{\theta}_i$

$$\hat{\theta}_i = \hat{\theta}\langle i \rangle$$

Theorem 7.3.3 For the covariance matrix of $\hat{\theta}$ (7.27) there holds that

$$\frac{N}{N_o} \text{cov}[\hat{\theta}_i, \hat{\theta}_k] \rightarrow (\Psi Q \Psi^*)_{ik} = \sum_{\ell=1}^p \Psi_{i\ell} \Psi_{k\ell}^* \frac{\Phi_v(\omega_\ell)}{|U_o(e^{j\omega_\ell})|^2} \quad \text{as} \quad \frac{p}{N} \rightarrow 0 \quad (7.34)$$

if $\|\Psi\|_2$ is uniformly bounded.

Proof: See Appendix D.2. □

Theorem 7.3.3 provides an asymptotic expression for the covariance of the estimated parameters which can be calculated. However, (7.34) does not provide much insight in the underlying mechanisms, and we do not know the conditions under which $\|\Psi\|_2$ is uniformly bounded (i.e. bounded independent of N_o and n_p).

To improve upon the result of Theorem 7.3.3 we clearly have to find an expression for the behaviour of $\Psi = (\Phi^* \Phi)^{-1} \Phi^*$ as N goes to infinity. To find such an expression we will first focus on the behaviour of $\Phi^* \Phi$.

Lemma 7.3.4 For Φ (7.26) there holds for $\Omega_{N_o}^u = \Omega_{N_o}$ that

$$\frac{1}{N_o} \Phi^* \Phi \rightarrow I \quad \text{as} \quad \frac{n}{N_o} \rightarrow 0$$

where the rate of convergence is n/N_o element wise.

Additionally, for $\Omega_{N_o}^u = \Omega_{N_o}$ there exists a constant $c_4 \in \mathbb{R}$ such that

$$\left\| \frac{1}{N_o} \Phi^* \Phi - I \right\|_2 \leq c_4 \frac{n_p n}{N_o}$$

Proof: See Appendix D.3. □

We now come to our final theorem regarding the covariance of the estimated parameters, where we will denote the grouped elements (block elements) of the vector of estimated parameters (7.28) as

$$\begin{aligned} \hat{\theta}_0 &:= \hat{D} \\ \hat{\theta}_i &:= \hat{L}_{i-1}^T \quad i = 1, \dots, n \end{aligned}$$

Theorem 7.3.5 For the estimated parameters (7.27) there holds for $\Omega_{N_o}^u = \Omega_{N_o}$ and $\frac{n_p n}{N_o} \rightarrow 0$ that for $i, k \geq 1$

$$\begin{aligned} \lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} N \text{cov}[\hat{\theta}_i, \hat{\theta}_k] &= \lim_{N_o \rightarrow \infty} \frac{1}{N_o} \sum_{\ell=1}^{N_o} \mathcal{V}_0(e^{j\omega_\ell}) \mathcal{V}_0^*(e^{j\omega_\ell}) \mathbf{G}_b^{i-k}(e^{j\omega_\ell}) \frac{\Phi_v(\omega_\ell)}{|U_o(e^{j\omega_\ell})|^2} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{V}_0(e^{j\omega}) \mathcal{V}_0^*(e^{j\omega}) \mathbf{G}_b^{i-k}(e^{j\omega}) \frac{\Phi_v(\omega)}{\Phi_u(\omega)} d\omega \quad (7.35) \end{aligned}$$

that for $i \geq 1$ and $k = 0$

$$\lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} N \operatorname{cov}[\hat{\boldsymbol{\theta}}_{\mathbf{i}}, \hat{\boldsymbol{\theta}}_{\mathbf{k}}] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{V}_0(e^{j\omega}) \mathbf{G}_b^{i-1}(e^{j\omega}) \frac{\Phi_v(\omega)}{\Phi_u(\omega)} d\omega \quad (7.36)$$

and that for $i = 0$ and $k = 0$

$$\lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} N \operatorname{cov}[\hat{\boldsymbol{\theta}}_{\mathbf{i}}, \hat{\boldsymbol{\theta}}_{\mathbf{k}}] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_v(\omega)}{\Phi_u(\omega)} d\omega \quad (7.37)$$

Proof: See Appendix D.4. □

Regarding (7.35) one should note that $\mathbf{G}_b^{-1} = \mathbf{G}_b^*$, since \mathbf{G}_b is a scalar inner function, see (E.4).

Remarks

1. We explicitly used the fact that the input signal is periodic to obtain Theorem 7.3.5.
2. The variance of an estimated parameter depends on the matching element of the vector $|\mathcal{V}_0(e^{j\omega_k})|^2$.
3. If the weighted (weighting function $\Phi_v(\omega)/\Phi_u(\omega)$) integral over frequency of an element of $|\mathcal{V}_0(e^{j\omega})|^2$ is small then the variance of the matching estimated parameter is small. Hence, considering the variance of the estimated parameters, $\mathcal{V}_0(e^{j\omega})$ should be small where the noise to signal ratio in the frequency domain is large. Note that, see (D.5),

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{V}_0(e^{j\omega}) \mathcal{V}_0^*(e^{j\omega}) d\omega = I$$

4. The asymptotic variance of the estimated parameters does not depend on the number of estimated parameters $n_p = n_b n + 1$.
5. If $\Phi_v(\omega)/\Phi_u(\omega)$ is constant over frequency then the estimated parameters are uncorrelated, see (D.5).
6. For the pulse basis (FIR model) it follows that $|\mathcal{V}_0(e^{j\omega})| = 1$ so that all parameters have the same variance.
7. From (7.35), (7.36) and (7.37) it follows that the asymptotic covariance of the estimated parameters does not depend on $r = N/N_o$.
8. The parameter estimate is numerically very well conditioned if $\Omega_{N_o}^u = \Omega_{N_o}$: asymptotically the condition number (quotient of maximum and minimum singular value) of $\Phi^* \Phi$ becomes one, which is optimal.

The following theorem addresses the asymptotic bias of the estimated parameters.

Theorem 7.3.6 *For the estimated parameters (7.27) there holds for $\Omega_{N_o}^u = \Omega_{N_o}$ and $\frac{r_p r_u}{N_o} \rightarrow 0$ that*

$$\lim_{N_o \rightarrow \infty} \mathbb{E}[\hat{\theta}] = \theta_o$$

Proof: See Appendix D.5. □

Remarks

1. Theorem 7.3.6 requires that $\Omega_{N_o}^u = \Omega_{N_o}$ and that $N_o \rightarrow \infty$. Theorem 7.3.6 therefore addresses the situation that the input is not periodic and the number of data points goes to infinity.
2. The estimated parameters are asymptotically in N_o unbiased. Also the variance converges to zero as N_o and r go to infinity, and the estimated parameters are asymptotically normally distributed. Hence, the parameter estimate is consistent, also in the case of undermodelling. For time domain parameter estimation using a FIR model structure an uncorrelated (white noise) input signal is necessary to obtain this result, see [174].
3. $S(e^{j\omega_\ell})$, the error due to past inputs, can be made relatively very small using input design, see (D.26).

Finally, we will take a closer look at the parameter estimation criterion, including a continuous weighting function $W(e^{j\omega_k})$. Using (7.19) and (7.20) we find

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta} \frac{1}{p} \sum_{k=1}^p |\hat{G}(e^{j\omega_k}) - \phi(e^{j\omega_k})\theta|^2 |W(e^{j\omega_k})|^2 \\ &= \arg \min_{\theta} \frac{1}{p} \sum_{k=1}^p \left| G_o(e^{j\omega_k}) - \phi(e^{j\omega_k})\theta + \frac{V^s(e^{j\omega_k})}{U^s(e^{j\omega_k})} + S(e^{j\omega_k}) \right|^2 |W(e^{j\omega_k})|^2 \end{aligned}$$

To be able to make some more explicit statements we will focus on the parameter values θ^* that minimize the expectation of the criterion in the limit as $p \rightarrow \infty$.

$$\theta^* = \arg \min_{\theta} \lim_{p \rightarrow \infty} \frac{1}{p} \sum_{k=1}^p \mathbb{E} \left[|\hat{G}(e^{j\omega_k}) - \phi(e^{j\omega_k})\theta|^2 \right] |W(e^{j\omega_k})|^2 \quad (7.38)$$

We have the following theorem.

Theorem 7.3.7 *For the estimated parameters (7.27) there holds that*

$$\begin{aligned} \theta^* &= \arg \min_{\theta} \lim_{p \rightarrow \infty} \frac{1}{p} \sum_{k=1}^p |G_o(e^{j\omega_k}) - \phi(e^{j\omega_k})\theta|^2 |W(e^{j\omega_k})|^2 \\ &= \arg \min_{\theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} |G_o(e^{j\omega}) - \phi(e^{j\omega})\theta|^2 |W(e^{j\omega})|^2 d\omega \quad \text{for } \Omega_{N_o}^u = \Omega_{N_o} \end{aligned}$$

When $W(e^{j\omega_k}) = 1$ there holds for $\Omega_{N_o}^u = \Omega_{N_o}$ that

$$\begin{aligned}\theta^* &= \arg \min_{\theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi(\theta_o - \theta)|^2 d\omega \\ &= \theta_o\end{aligned}$$

Proof: See Appendix D.6. □

Remarks

1. Considering (7.27) it follows that the parameter estimate is a direct, nonasymptotic weighted fit in the frequency domain. Considering (D.29) and (D.30) we find that asymptotically the parameter estimate is a weighted fit on the true transfer function, where the weighting function is directly chosen by the user. The direct choice of the weighting function is an advantage when considering the problem of identification for control design.
2. The result given in (D.30) is the frequency domain identification analog of the time domain prediction error identification result given by [106, Equation (8.66)].

7.4 Asymptotic analysis of transfer function estimates

The nonweighted ($W = 1$) estimated transfer function is obtained as

$$\hat{G}^f(e^{j\omega}) := P(e^{j\omega})\hat{\theta} = P(e^{j\omega})(\Phi^* \Phi)^{-1} \Phi^* \hat{G} \quad (7.39)$$

$$P(e^{j\omega}) = [1 \quad \mathcal{V}_0^T(e^{j\omega}) \quad \dots \quad \mathcal{V}_{n-1}^T(e^{j\omega})] \quad (7.40)$$

Again, we will first address the asymptotic behaviour of the variance, and next the asymptotic behaviour of the bias of this transfer function estimate.

Theorem 7.4.1 *For the estimated transfer function (7.39) there holds for $\Omega_{N_o}^u = \Omega_{N_o}$, $\omega \in [0, 2\pi)$, and $\frac{n_p n}{N_o} \rightarrow 0$ that*

$$\lim_{\substack{N_o \rightarrow \infty \\ n_p \rightarrow \infty}} \frac{N}{n_p} \text{var}[\hat{G}^f(e^{j\omega})] = \lim_{N_o \rightarrow \infty} \frac{1}{N_o} \sum_{\ell=1}^{N_o} \mathcal{P}_n(\omega, \omega_\ell) \frac{\Phi_v(\omega_\ell)}{|U_o(e^{j\omega_\ell})|^2} \quad (7.41)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{P}_n(\omega, \zeta) \frac{\Phi_v(\zeta)}{\Phi_u(\zeta)} d\zeta \quad (7.42)$$

where

$$\mathcal{P}_n(\omega, \zeta) = \frac{1}{n_p} \left| 1 + \mathcal{V}_0^T(e^{j\omega}) \mathcal{V}_0(e^{-j\zeta}) \frac{\left(\frac{\mathbf{G}_b(e^{j\omega})}{\mathbf{G}_b(e^{j\zeta})} \right)^n - 1}{\frac{\mathbf{G}_b(e^{j\omega})}{\mathbf{G}_b(e^{j\zeta})} - 1} \right|^2 \quad (7.43)$$

If in addition $\Phi_v(\zeta)/\Phi_u(\zeta) = 1$ for all $\zeta \in [0, 2\pi)$ then

$$\lim_{\substack{N \rightarrow \infty \\ r \rightarrow \infty}} \frac{N}{n_p} \text{var}[\hat{G}^f(e^{j\omega})] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{P}_n(\omega, \zeta) d\zeta = \frac{1 + n \|\mathcal{V}_0(e^{j\omega})\|_2^2}{n_p}$$

Proof: See Appendix D.7. □

Note that for any $x \in \mathbb{C}$ there holds that

$$\lim_{x \rightarrow 1} \frac{x^n - 1}{x - 1} = n \quad (7.44)$$

$$\left| \frac{x^n - 1}{x - 1} \right| = \left| \sum_{k=0}^{n-1} x^k \right| \leq n \quad \text{if } |x| = 1 \quad (7.45)$$

$$\left| \frac{x^n - 1}{x - 1} \right| \leq \frac{|x|^n + 1}{|x - 1|} = \frac{2}{|x - 1|} \quad \text{if } |x| = 1 \quad (7.46)$$

$$\frac{x^{n+1} - 1}{x - 1} = x^n + \frac{x^n - 1}{x - 1} \quad (7.47)$$

which, together with (7.4.1), provides a reasonable indication of the behaviour of $\mathcal{P}_n(\omega, \zeta)$ in (7.43).

Remarks

1. We explicitly used the fact that the input signal is periodic to obtain Theorem 7.4.1.
2. Note that (7.41) and (7.43) provide an accessible expression for the variance for finite n , as opposed to the expressions established for time domain prediction error methods, see e.g. [189, 105, 104, 113, 106, 192, 180]. At first sight (7.41) or (7.42) and (7.43) may seem a bit complicated, but when given a basis generating system and a model order, then $\mathcal{P}_n(\omega, \zeta)$ can be calculated directly, and the effect of adding (or subtracting) a model order is given by (7.47).
3. The variance will be relatively large where $\|\mathcal{V}_0(e^{j\omega})\|_2^2$ is large (near the poles of the basis generating system), provided that $\Phi_v(\omega)/|U_o(e^{j\omega})|^2$ is not relatively small for those frequencies. This implies that a large amount of noise for the higher frequencies (as is rather common in practical applications) usually is relatively harmless, because the true system, and hence a properly chosen basis generating system, usually will have a lowpass character.
4. The frequency region (with regard to ζ and a fixed ω) where $\mathcal{P}_n(\omega, \zeta)$ is relatively large becomes more concentrated around $\zeta = \omega$ when n increases. Considering the variance error it is therefore advantageous to have a high model order n where $\Phi_v(\omega)/|U_o(e^{j\omega})|^2$, the noise on \hat{G} , is small, and vice versa. Hence, a high

order model can have a lower variance in certain frequency regions (the frequency regions where $\Phi_v(\omega)/|U_o(e^{j\omega})|^2$ is small) than a low order model. It follows that, up to a certain model order, variance considerations do not conflict with bias considerations for those frequency regions where $\Phi_v(\omega)/|U_o(e^{j\omega})|^2$ is relatively small.

5. Considering (D.9), it follows that the conditions stated in Theorem 7.4.1, namely $\Omega_{N_o}^u = \Omega_{N_o}$ and $\frac{n_p n}{N_o} \rightarrow 0$, are not only theoretical: simulations show that if the gap between two adjacent frequency points is large, relative to the smoothness of the basis functions that are used and the number of basis functions that is used, then the bias and variance can increase considerably between these frequency points.
6. If an estimate of $\Phi_v(\omega_k)$ is known, then the asymptotic variance expression (7.41) can be estimated easily, which is important for model order selection. An estimate of $\Phi_v(\omega_k)$ is given by $\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k}))|U^s(e^{j\omega_k})|^2$, see (5.23).
7. From (7.42) it follows that the asymptotic variance of the estimated transfer function does not depend on $r = N/N_o$.

The model order n will, for a basis generating system that is chosen properly, typically lie between 3 and 6. Asymptotic results in the model order are therefore of little practical significance, and we will not pursue them. We will only give the asymptotic result in the model order n for the pulse basis $\mathcal{V}_k(e^{j\omega}) = e^{-j\omega(k+1)}$, to be able to compare our results with the time domain results of [113], and because in this case asymptotic results in the model order may be useful.

Corollary 7.4.2 *Let \mathcal{V}_k be the pulse basis $\mathcal{V}_k(e^{j\omega}) = e^{-j\omega(k+1)}$. Then, for the estimated transfer function (7.39) there holds for $\Omega_{N_o}^u = \Omega_{N_o}$, $\omega \in [0, 2\pi)$ and $\frac{n_p^2}{N_o} \rightarrow 0$ that*

$$\begin{aligned} \lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} \frac{N}{n_p} \text{var}[\hat{G}^f(e^{j\omega})] &= \lim_{N_o \rightarrow \infty} \frac{1}{N_o} \sum_{\ell=1}^{N_o} \frac{1}{n_p} \left| 1 + e^{j\frac{\omega_\ell - \omega}{2}(n+1)} \frac{\sin(n\frac{\omega_\ell - \omega}{2})}{\sin(\frac{\omega_\ell - \omega}{2})} \right|^2 \frac{\Phi_v(\omega_\ell)}{|U_o(e^{j\omega_\ell})|^2} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{n_p} \left| 1 + e^{j\frac{\zeta - \omega}{2}(n+1)} \frac{\sin(n\frac{\zeta - \omega}{2})}{\sin(\frac{\zeta - \omega}{2})} \right|^2 \frac{\Phi_v(\zeta)}{\Phi_u(\zeta)} d\zeta \end{aligned}$$

If in addition $\Phi_v(\omega)/\Phi_u(\omega)$ is Lipschitz continuous of order 1 then

$$\lim_{n_p \rightarrow \infty} \lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} \frac{N}{n_p} \text{var}[\hat{G}^f(e^{j\omega})] = \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

Proof: See Appendix D.8. □

Remark

From Theorem 5.3.2, (7.19), (7.20), (7.39), and the fact that the input signal is considered as deterministic, it follows that the estimated transfer function is asymptotically in N normally distributed. Additionally, Theorem 7.4.3 states that asymptotically in n the transfer function estimate is unbiased. For the pulse basis $\mathcal{V}_k(e^{j\omega}) = e^{-j\omega(k+1)}$ we now find from Corollary 7.4.2 that (in suggestive but sloppy notation)

$$\lim_{n_p \rightarrow \infty} \lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} \hat{G}^f(e^{j\omega}) \in \mathcal{N} \left(G_o, \frac{n_p}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)} \right) \quad (7.48)$$

Note the complete similarity of (7.48) with the time domain prediction error results of [113]. However, we need a periodic input signal.

The following theorem addresses the asymptotic bias of the estimated transfer function.

Theorem 7.4.3 *For the estimated transfer function (7.39) there holds for $\Omega_{N_o}^u = \Omega_{N_o}$, $\omega \in [0, 2\pi)$ and $\frac{n_p n}{N_o} \rightarrow 0$ that*

$$\begin{aligned} \lim_{N_o \rightarrow \infty} \left(G_o(e^{j\omega}) - \mathbb{E}[\hat{G}^f(e^{j\omega})] \right) &= G_o(e^{j\omega}) - P(e^{j\omega})\theta_o \\ &= \sum_{k=n}^{\infty} L_k \mathcal{V}_k(e^{j\omega}) \\ \lim_{N_o \rightarrow \infty} |G_o(e^{j\omega}) - \mathbb{E}[\hat{G}^f(e^{j\omega})]| &\leq \mathcal{K} |\mathcal{V}_0(e^{j\omega})| \frac{\eta^n}{1-\eta} \end{aligned}$$

Proof: See Appendix D.9. □

Remarks

1. Theorem 7.4.3 requires that $\Omega_{N_o}^u = \Omega_{N_o}$ and that $N_o \rightarrow \infty$. Theorem 7.4.3 therefore addresses the situation that the input is not periodic and the number of data points goes to infinity.
2. The optimal model order n can be found approximately by employing a numerical search using the asymptotic bias and variance expressions: find the minimum over n of $|\text{Bias}(e^{j\omega}) + \sqrt{\alpha \text{Var}(e^{j\omega})}|$, where α is given by the confidence level. The exact optimal model order can be found by just calculating and comparing the error bounds given by Theorem 5.4.4 or Theorem 5.4.5 for different model orders, where the approximate optimal model order can be used to indicate where to look for the optimal model order.

7.5 Conclusions

We have obtained asymptotic bias and variance expressions for the estimated parameters and transfer function. The bias expressions are asymptotic in the period length of

the input signal, whereas the variance expressions are asymptotic in both the period length of the input signal and the number of periods.

Evaluating the results which we have obtained with respect to input design, model order selection, and the selection of the basis generating system we can summarize as follows.

1. Given a basis generating system, a model order n , and an estimate of $\Phi_v(\omega_k)$, then Theorem 7.4.1 provides valuable information for input design.
Note that Theorem 7.4.3 can be helpful to select a model order n in this situation, and that an estimate of $\Phi_v(\omega_k)$ is given by $\hat{\sigma}_r^2(\hat{G}(e^{j\omega_k})|U^s(e^{j\omega_k})|^2)$, see (5.23).
2. Given a basis generating system, $\Phi_u(\omega_k)$, an estimate of $\Phi_v(\omega_k)$, and an estimate of \mathcal{K} and η of (D.16), then the model order n which minimizes the mean square error error in the estimated transfer function can be found by a numerical search over n using Theorem 7.4.1 and Theorem 7.4.3.
3. Given $\Phi_u(\omega_k)$, an estimate of $\Phi_v(\omega_k)$, and estimates of \mathcal{K} and η of (D.16), then Theorem 7.4.1 and Theorem 7.4.3 can be used to make a selection between a set of candidate basis generating systems.

Additionally, Theorem 7.3.6 shows that, even in the case of undermodelling, the estimated parameters are asymptotically unbiased. Fact 7.3.1 shows that a direct, nonasymptotic, fit in the frequency domain is obtained, which is advantageous because the H_∞ robust control design methodology is frequency domain oriented. For time domain prediction error identification the frequency domain fit can only be described asymptotically in the number of data points, see [106]. Theorem 7.3.7 provides an expression for the expectation of the least squares criterion as the period length of the input signal goes to infinity. This expression shows how the weighting function $W(e^{j\omega})$ can be used to distribute the bias in the transfer function estimate over frequency. Alternatively, the weighting function $W(e^{j\omega})$ can also be used to obtain (an approximation of) the minimum variance estimate, by choosing $W(e^{j\omega_k})$ to be equal to (an estimate of) the inverse variance of $\hat{G}(e^{j\omega_k})$.

Chapter 8

Applications

Summary

In this chapter the model error estimation procedure of Chapter 5 is applied to a laboratory set-up of a variable speed wind turbine system, and to the pick-up mechanism of a compact disk player which operates in closed loop.

For the wind turbine system, the estimated error bounds are validated in two ways. Firstly, the error bounds are used to verify the robustness of a designed controller, indicating that the controller should stabilize the system. Implementing the controller shows that the controller indeed stabilizes the system, and achieves the same amount of performance improvement for the true closed loop as for the designed closed loop. Secondly, the procedure is applied to a nonlinear simulation model of the wind turbine system. The estimated error bounds turn out to be correct. That is, the error bounds are able to capture the undisturbed output of the nonlinear system.

Experiments on the pick-up mechanism of a compact disk player can only be performed when this system is controlled. To handle the resulting closed loop situation we use the coprime factor framework of [153], and the results of Chapter 6. High order models are necessary to accurately describe the coprime factors of the system, and it turns out that the noise contains a periodic component. To obtain values for $\hat{\mathcal{K}}$ and $\hat{\eta}$ which are consistent with the data, we have to allow for a considerable contribution due to undermodelling to the error bound (relative to the variance contribution). This structural contribution in the error bound should account for the error due to the periodic disturbances, in addition to the error due to undermodelling. The resulting error bounds are validated over new data. The estimated model error bounds, together with the estimated bounds on the DFT of the noise, turn out to give a reliable and tight bound for the difference between the DFT of the output of the model and the DFT of the

measured output. Finally, a new robust controller has been designed for the system, while using the estimated uncertainty bounds on the coprime factors in the design phase. This new controller achieves a substantially better performance than the existing one.

Based on these results, we conclude that fairly realistic error bounds are obtained, in spite of the presence of nonlinearities and periodic disturbances, and that the estimated error bounds indeed do provide a valuable basis for high performance robust control design.

8.1 Application to a wind turbine system

8.1.1 Introduction

In this section we will apply the model error estimation technique of Chapter 5 to a laboratory set-up of a wind turbine system. For this system a high performance controller was designed in [14], see also [18, 16]. Using the estimated error bounds we will check whether the controller stabilizes the system, to verify whether the controller can be implemented safely.

Clearly, the test-rig is not linear, not time invariant, and not finite dimensional, and therefore does not satisfy the assumptions under which the error bounds of Chapter 5 are derived. However, in the small area around the operating condition in which the experiments were performed, the test-rig approximately behaves linearly. Therefore, based on the results of Section 5.9, Section 5.10.1, Section 5.10.2, we are fairly confident that the estimated error bounds will provide a realistic indication of the model error for control system design around the operating condition. We will further check this conviction in the following two ways.

- If the estimated error bounds indicate that the controller stabilizes the true system, the controller will be implemented so that we can verify whether the closed loop is indeed stable.
- We will repeat the model error estimation procedure on a nonlinear simulation model of the system, where the undisturbed output is available for verification of the results.

8.1.2 The wind turbine system

We consider the IRFLET test-rig, see [176], which is a laboratory set-up of a variable speed wind turbine system. The layout of the test-rig is given in Fig. 8.1. The controlled DC-machine on the left part of Fig. 8.1 and the flywheel emulate the rotor part of a small wind turbine, enabling experimentation under user definable conditions. The other components in Fig. 8.1 (from left to right) are the transmission, the generator

(a variable speed synchronous machine), and the link with the public grid (a DC-link to establish a decoupling between the varying generator speed and the constant grid frequency). These latter components also appear in a true wind turbine.

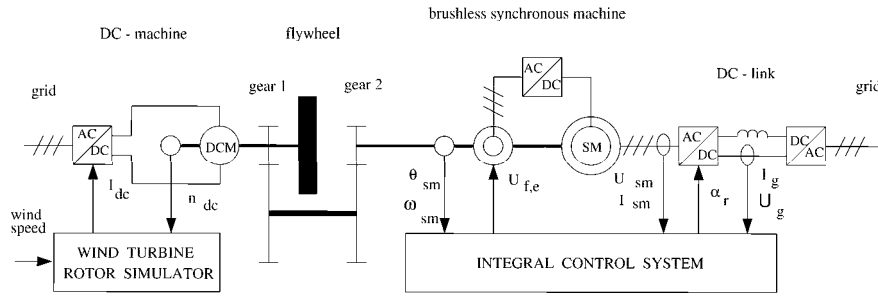


Fig. 8.1: The IRFLET test-rig [176].

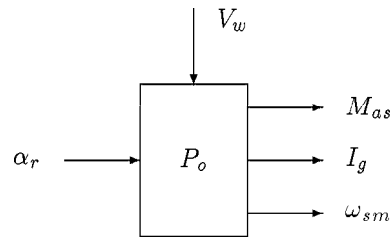


Fig. 8.2: Wind turbine.

The test-rig of Fig. 8.1 is schematically represented in Fig. 8.2, including the input and output variables, with

Vw	V_w	wind velocity	$[m/s]$
alf	α_r	delay angle rectifier	$[rad]$
Mas	M_{as}	rotor shaft torque	$[Nm]$
Ig	I_g	direct current	$[A]$
omg	ω_{sm}	generator speed	$[rpm]$

where the symbols in the first column refer to the labels which we will use in the figures.

Experiments are carried out in one operating condition (full load) of the test-rig, where

wind velocity	12	m/s
field excitation voltage	22	V
delay angle rectifier	0.47	rad

Appropriate anti-aliasing filters are used to measure the data, and a sampling frequency of 100 Hz has been chosen.

The chosen wind disturbance V_w has a flat spectrum up to 10 Hz, and there is no frequency content above 10 Hz. The delay angle α_r is excited by a 5 Hz (clock frequency) pseudo random binary sequence (PRBS). This means that at multiples of 5 Hz the spectrum of the input signal is zero, so that the empirical transfer function estimate (ETFE, see [106]) is unreliable at those frequencies. This excitation signal was repeated every 2000 points, resulting in a periodic input signal.

Employing standard least squares prediction error system identification techniques (see e.g. [110, 106]) parametric models for the test-rig were obtained from the measured data in [18], see also [16]. In Fig. 8.3, Fig. 8.4 and Fig. 8.5 the transfer functions of the individual channels are shown. For each channel the empirical transfer function estimate (ETFE, see [106]) and the estimated parametric model is given. It can be seen that there exists a close resemblance between the parametric models and the ETFE. This holds for both the amplitude as well as the phase.

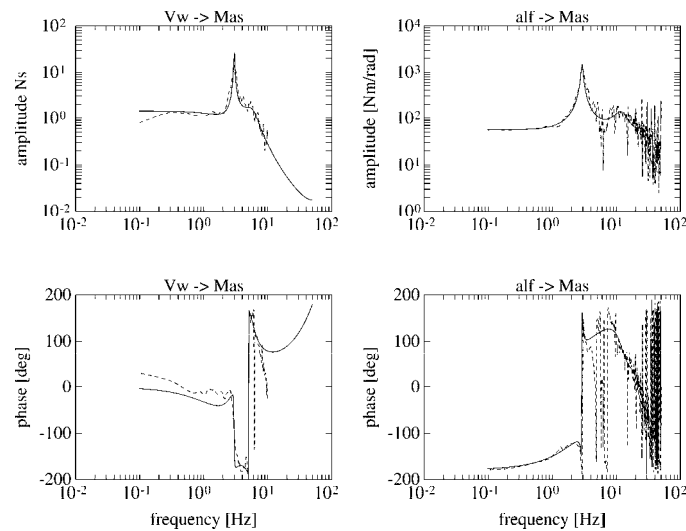


Fig. 8.3: Transfer functions rotor shaft torque, (—) parametric model, (---) ETFE. The titles of the plots refer to the specific inputs and outputs of the elements of the transfer function matrix.

To increase the damping of the mechanical resonance frequencies in the transmission while maintaining a reasonably constant amount of produced electrical energy, a high

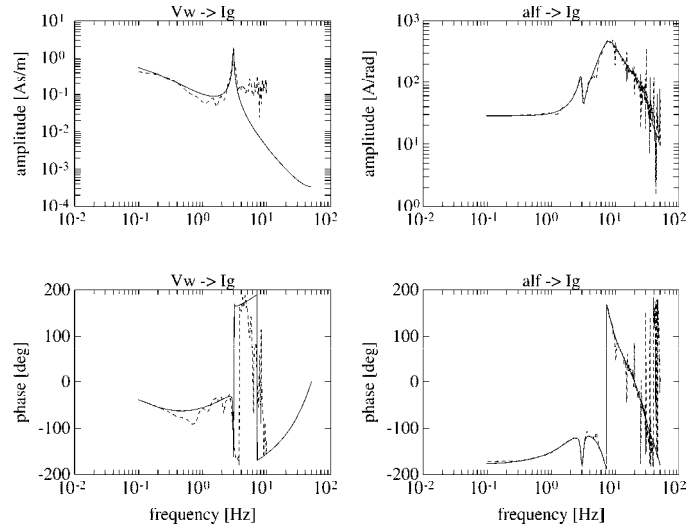


Fig. 8.4: Transfer functions DC-current, (—) parametric model, (---) ETFE. The titles of the plots refer to the specific inputs and outputs.

performance robust controller was designed in [14] (see also [18, 16]) using the identified parametric models. An H_∞ control design method (see [39, 38]) and Hankel norm reduction (see [50]) of the order of the controller were used, resulting in a 25th order controller C .

The feedback configuration is given in Fig. 8.6, where

$$\begin{aligned}
 w \quad \text{exogenous inputs} &= V_w \\
 z \quad \text{controlled outputs} &= M_{as} \\
 u \quad \text{control inputs} &= \alpha_r \\
 y \quad \text{measured outputs} &= \begin{pmatrix} I_g \\ \omega_{sm} \end{pmatrix}
 \end{aligned}$$

As an example of the designed closed loop, the rotor shaft torque reduction is given in Fig. 8.7. It can be seen that the control objective has been met on the model.

8.1.3 Model error estimation

To check whether the controller at least will stabilize the true system, Fig. 8.6 shows that we have to have available a model and a model error bound for the transfer function from α_r to $(I_g \ \omega_{sm})^T$, where V_w acts as a disturbance.

The aim of this section is to find an accurate model, together with a model uncertainty description which provides a *realistic* reflection of the uncertainty about the

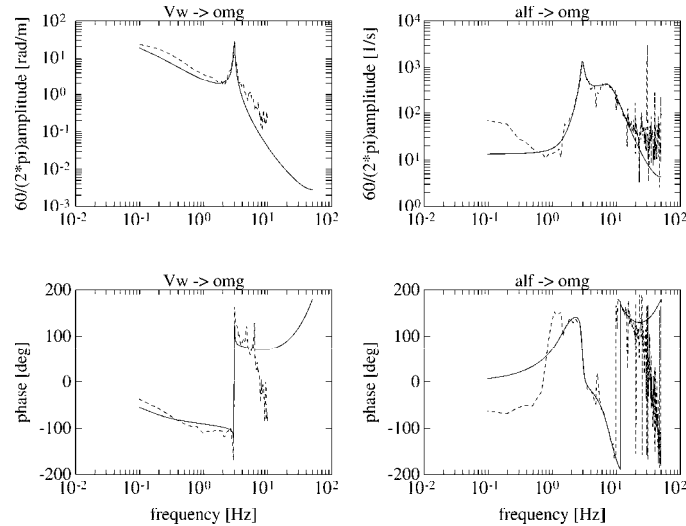


Fig. 8.5: Transfer functions generator speed, (—) parametric model, (---) ETFE. The titles of the plots refer to the specific inputs and outputs.

system which is present in the data, considering the situation where the available data about the system is incomplete and corrupt. That is, we want to find a description of the set of all models which are not highly unlikely to have generated the measured data. In order to achieve this, we will use the model error estimation procedure of Chapter 5.

Clearly, the test-rig is not linear, time invariant, and finite dimensional, and therefore does not satisfy the assumptions under which the error bounds of Chapter 5 are derived. However, in a small area around one operating condition, as used for the experiments, the test-rig approximately behaves linearly. Therefore, in combination with the results of Section 5.9, Section 5.10.1, Section 5.10.2, we are fairly confident that the estimated error bounds will provide a realistic indication of the model error for control system design.

The model uncertainty estimation procedure of Chapter 5 only addresses scalar systems. For the SIMO test-rig we will therefore estimate individual error bounds for the transfer function from α_r to I_g , and the transfer function from α_r to ω_{sm} .

To estimate the error bounds we used the same experimental conditions (wind disturbance, operating point, sampling frequency) as were used to identify the nominal models, see Section 8.1.2. We did however use another excitation signal: a 25 Hz (clock frequency) PRBS instead of a 5 Hz PRBS. This excitation signal is more suitable to estimate the error bounds, since the spectrum of this signal contains less frequency regions in which it is very small. This input signal does yield the same nominal models

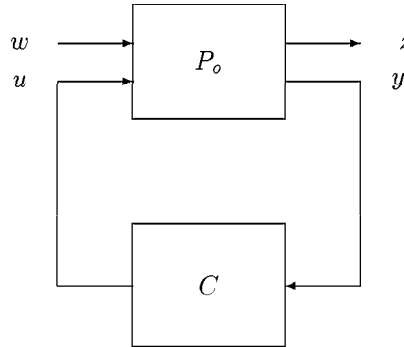


Fig. 8.6: Feedback configuration.

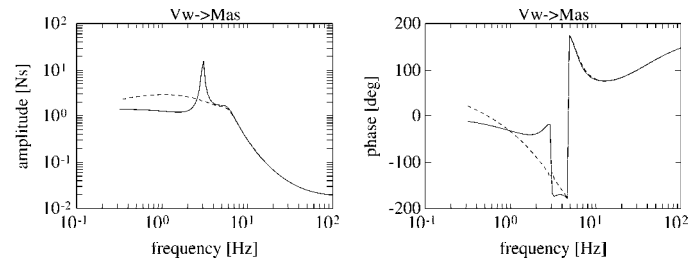


Fig. 8.7: Transfer functions, (—) open loop, (---) designed closed loop, for the feedback configuration of Fig. 8.6. The titles of the plots refer to the specific inputs and outputs.

as the ones of Section 8.1.2.

More specifically, the input signal was chosen to be periodic, using a 25 Hz (clock frequency) PRBS over 2000 points to construct one period. Data was measured over 8 periods. The magnitude of the DFT of the input signal over one period is given in Fig. 8.8. Note that at multiples of 25Hz the spectrum of the input signal is zero.

To estimate the error bounds we had $N = 16000$, $N_o = 2000$ and $N_s = 2400$. The probability level (as given by F_α) for all confidence regions that will be shown is specified to be 99 %.

The nominal models of [18] as given in Section 8.1.2 turned out not to be suitable as basis generating systems. The matching estimated generalized expansion coefficients did not converge to zero at a satisfactory rate. We therefore did try to estimate more accurate models from the measured data in the following way. Firstly, for each of the two transfer functions separately, we used the nominal model of [18] as a basis generating system. The resulting orthonormal basis functions were used to estimate a high order model from the data. This high order model is compared to the ETFE

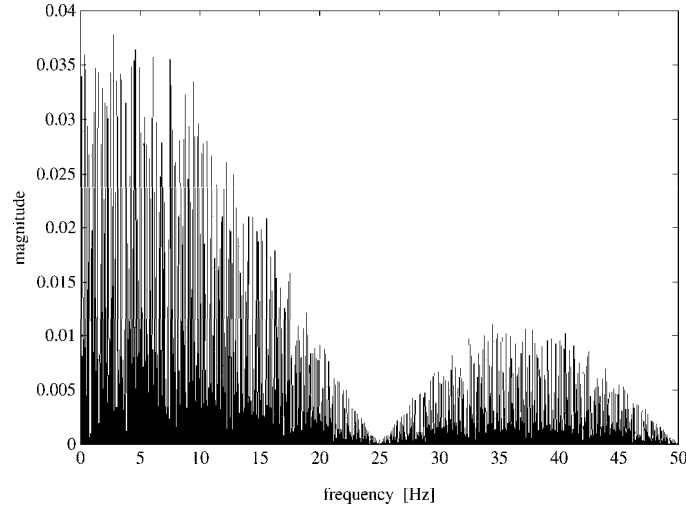


Fig. 8.8: Magnitude of the DFT over one period of the input signal, $|U_i(e^{j\omega_k})|$.

to check whether all significant dynamics are modelled. If not, the model order is increased. To again obtain a low order model, we used model reduction by balancing and truncation (see [128]) or by fitting a low order model on the high order one in the frequency domain using weighted least squares, trying to minimize the infinity norm of the error by adapting the weighting function (i.e. $W_{\text{new}}(\omega_k) = W_{\text{old}}(\omega_k)|G(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})|$ for fitting \hat{G} on G , and visual inspection of the fit in a Nyquist plot to decide when the fit is satisfactory). This lower order model in turn was used as a basis generating system. Iterating in this way we obtained basis generating systems which were satisfactory.

The values for M , ρ , \mathcal{K} , and η are obtained from the data iteratively. Starting from zero, we iterate till the estimated prior information is in accordance with the matching confidence intervals, i.e. till the bounds given by the estimates of the prior information cross the matching confidence intervals only where zero is contained in the confidence interval. This results in $\hat{M} = 100$, $\hat{\rho} = 1.0085$, $\hat{\mathcal{K}} = [150 \ 60 \ 65 \ 60 \ 30 \ 115 \ 55 \ 50]$, and $\hat{\eta} = 0.3$ for the transfer function from α_r to I_g , and $\hat{M} = 70$, $\hat{\rho} = 1.008$, $\hat{\mathcal{K}} = [100 \ 140 \ 110 \ 180]$, and $\hat{\eta} = 0.55$ for the transfer function from α_r to ω_{sm} . These estimated values now are taken as the prior information of Assumption 5.2.1.ii and Assumption 5.2.1.iii.

We will now first give the results for the transfer function from α_r to I_g , and next the results for the transfer function from α_r to ω_{sm} . Figure 8.9 shows the estimated generalized expansion coefficients and the confidence interval of Theorem 5.4.2, using $n_p = 33$ and $W = \text{diag}(1/\hat{\sigma}_r(\hat{G}(e^{j\omega_k})))$, together with the prior information given by

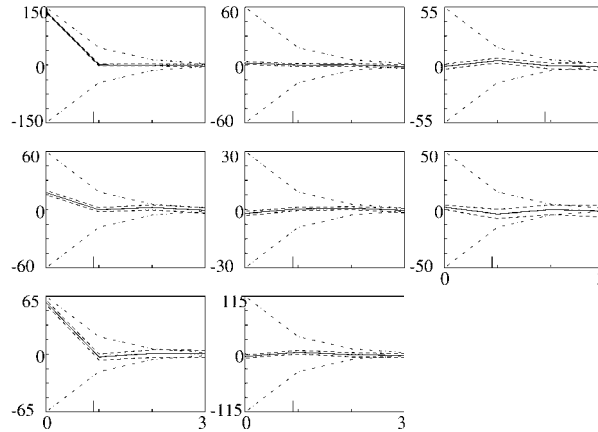


Fig. 8.9: Estimated generalized expansion coefficients (—), confidence interval (---), and estimated prior information (— · —), for the transfer function from α_r to I_g .

$\hat{\mathcal{K}}$ and $\hat{\eta}$. The solid vertical lines in the plots indicate that zero lies in the confidence interval from thereon, the dashed vertical lines indicate that the prior information crosses the estimate. Experience shows that, for the prior information to be realistic, the latter should come at least two coefficients later than the former, and that the tail of the confidence interval which contains zeros should be at least two or three (depending on the preceding behaviour of the confidence interval) coefficients long. Figure 8.10 shows the true impulse response, the prior information given by \hat{M} and $\hat{\rho}$, and the confidence interval of Theorem 5.4.8 using $n_y = 33$ and $\mathcal{W}_y = 1$. Finally, Fig. 8.11 gives the Nyquist plot of the estimated parametric transfer function from α_r to I_g , together with the confidence region of Theorem 5.4.5, using $n_f = 33$ and $\mathcal{W} = 1$.

The results for the transfer function from α_r to ω_{sm} are given in Fig. 8.12, Fig. 8.13 and Fig. 8.14. We used $W = \text{diag}(1/\hat{\sigma}_r(\hat{G}(e^{j\omega_k})))$, $n_p = 29$, $n_y = 29$, $n_f = 29$, and \mathcal{W}_y and \mathcal{W} diagonal matrices with diagonal elements as given in Fig. 8.15.

To illustrate the procedure of estimating the prior information from the data, in Fig. 8.16 values for $\hat{\mathcal{K}}$ and $\hat{\eta}$ are shown which are not consistent with the resulting confidence interval on the estimated generalized expansion coefficients, as indicated by the upper left and lower right plots.

Remarks

- Due to the very large significant length of the impulse responses, see Fig. 8.13

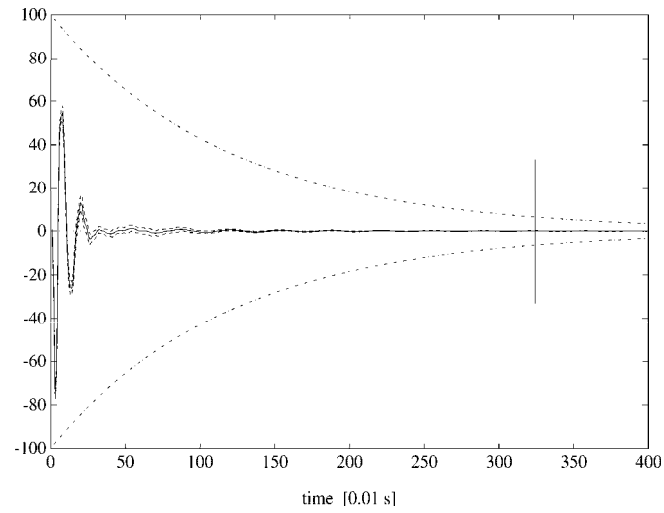


Fig. 8.10: Estimated impulse response (—), confidence interval (---), and estimated prior information (- · -), for the transfer function from α_r to I_g .

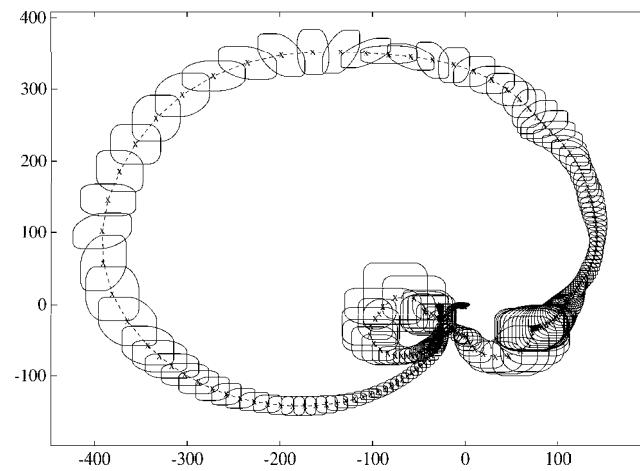


Fig. 8.11: Nyquist plot of the parametric transfer function estimate, and confidence region, for the transfer function from α_r to I_g .

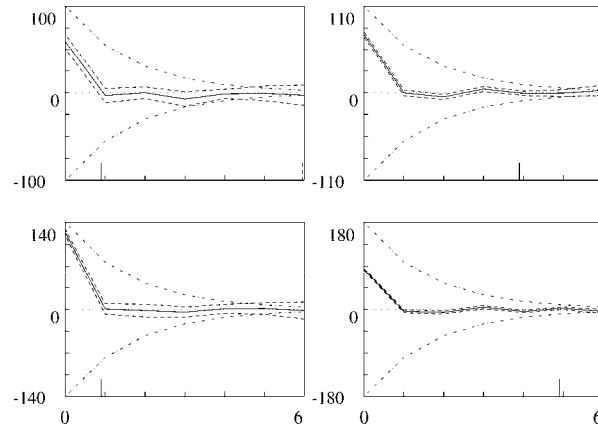


Fig. 8.12: Estimated generalized expansion coefficients (—), confidence interval (— —), and estimated prior information (— ·), for the transfer function from α_r to ω_{sm} .

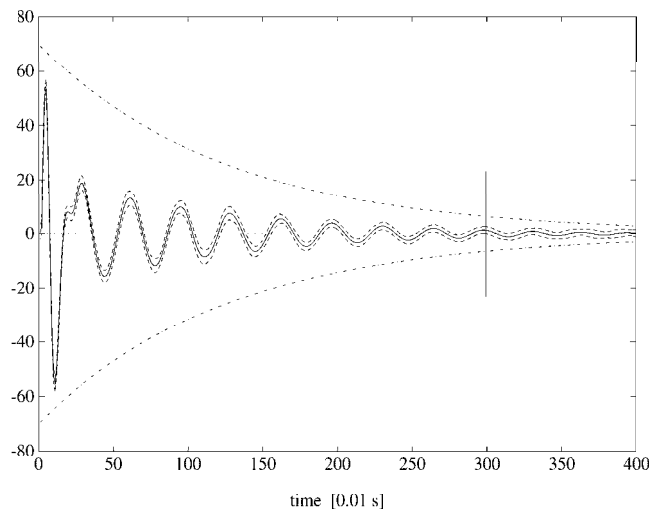


Fig. 8.13: Estimated impulse response (—), confidence interval (— —), and estimated prior information (— ·), for the transfer function from α_r to ω_{sm} .

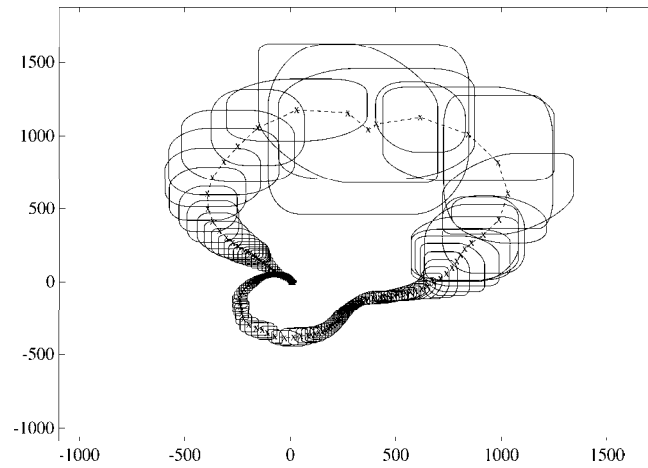


Fig. 8.14: Nyquist plot of the parametric transfer function estimate, and confidence region, for the transfer function from α_r to ω_{sm} .

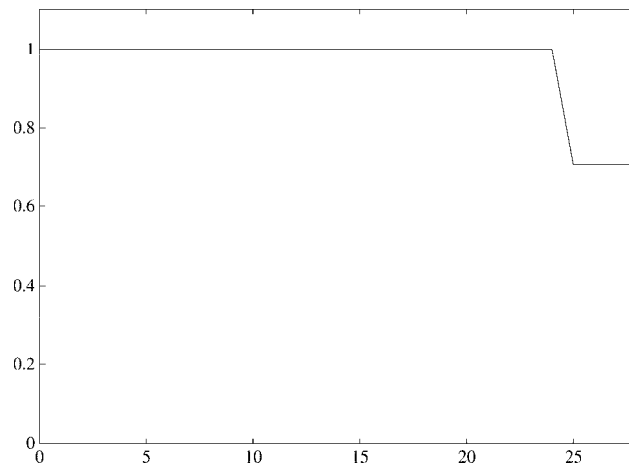


Fig. 8.15: Diagonal elements of the diagonal weighting functions \mathcal{W}_y and \mathcal{W} used for the impulse response estimate and the parametric transfer function estimate, respectively, for the transfer function from α_r to ω_{sm} .

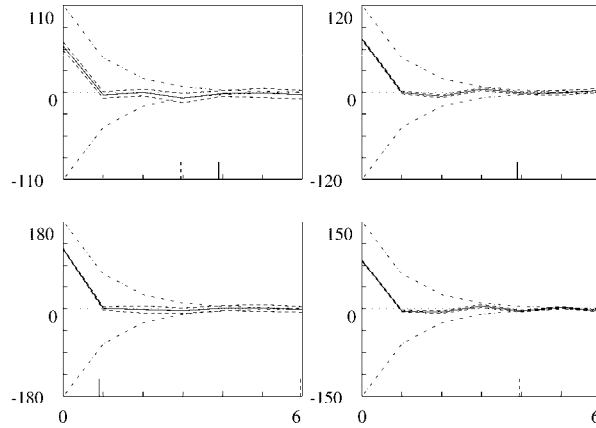


Fig. 8.16: Example of estimated prior information being inconsistent with the resulting confidence intervals on the estimated parameters for the transfer function from α_r to ω_{sm} . Estimated generalized expansion coefficients (—), confidence interval (---), and estimated prior information (· · ·).

and Fig. 8.10, FIR models yield very bad estimates and hence large error bounds for these transfer functions. A specifically tailored set of basis functions are imperative to obtain satisfactory results in such cases.

- It should be noted that the input signal used (a PRBS) is not the best input signal with regard to the uncertainty estimation procedure used, see Section 5.3.3 and Chapter 7, especially Theorem 7.4.1.

8.1.4 Robust stability

The error bounds estimated in the previous section can be represented by an additive uncertainty model Δ_a around the estimated wind turbine model P :

$$P_o = P + \Delta_a$$

Omitting the external signals w and z of Fig. 8.6 and adding the uncertainty description, the feedback configuration is drawn in Fig. 8.17.

According to the small gain theorem (see [190]), the closed loop of Fig. 8.17 is stable provided that

$$\bar{\sigma}((I - CP)^{-1}C\Delta_a) < 1$$

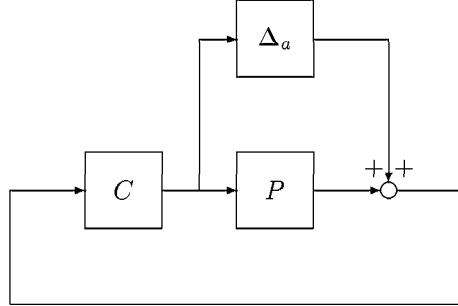


Fig. 8.17: Additive uncertainty representation.

where $\bar{\sigma}$ denotes the maximum singular value. For the estimated (SIMO) model of the wind turbine, the plant can be decomposed as

$$\begin{pmatrix} I_g \\ \omega_{sm} \end{pmatrix} = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \alpha_g$$

the controller as

$$\alpha_g = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{pmatrix} I_g \\ \omega_{sm} \end{pmatrix}$$

while the uncertainty is

$$\Delta_a = \begin{bmatrix} \Delta_{I_g} \\ \Delta_{\omega_{sm}} \end{bmatrix}$$

The transfer function $(I - CP)^{-1}C$ can be rewritten as

$$\begin{bmatrix} (1 - CP)^{-1}C_1 & (1 - CP)^{-1}C_2 \end{bmatrix}$$

Hence, the closed loop remains stable provided that

$$|(1 - CP)^{-1}C_1\Delta_{I_g}| + |(1 - CP)^{-1}C_2\Delta_{\omega_{sm}}| < 1$$

Note that in this step conservativeness could be introduced. The estimated error bound, as measure of the perturbation, contains no phase information. Therefore the following stability condition need to be checked (without adding conservativeness to the previous step)

$$|(1 - CP)^{-1}C_1| |\Delta_{I_g}| + |(1 - CP)^{-1}C_2| |\Delta_{\omega_{sm}}| < 1 \quad (8.1)$$

The graphical representation of the above stability condition is given in Fig. 8.18. Both P_1 , P_2 , as well as $|\Delta_{\omega_{sm}}|$, $|\Delta_{I_g}|$ are taken from Fig. 8.11 and Fig. 8.14. It can be seen in Fig. 8.18 that the stability condition of (8.1) is satisfied. Therefore, given that the estimated error bounds are correct, the real system will be stabilized by the controller. Based on this result, we proceed with the actual implementation of the controller.

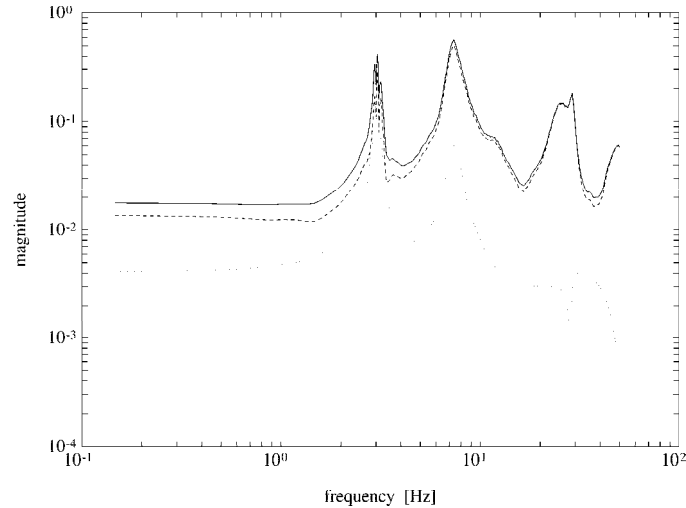


Fig. 8.18: Stability result with respect to additive uncertainty,
 (—) $|(1 - CP)^{-1}C_1| |\Delta I_q| + |(1 - CP)^{-1}C_2| |\Delta \omega_{sm}|$,
 (- -) component $|(1 - CP)^{-1}C_1| |\Delta I_q|$,
 (···) component $|(1 - CP)^{-1}C_2| |\Delta \omega_{sm}|$.

8.1.5 Controller implementation

In this section the predicted results of Section 8.1.4 will be verified with measurements on the test-rig. For this purpose the controller has been implemented and used to control the test-rig. The same wind velocity signal as in Section 8.1.2 and Section 8.1.3 has been applied.

First of all, the controller indeed stabilizes the system, and achieves the same amount of reduction of rotor shaft torque variations for the true closed loop as for the designed closed loop. In Fig. 8.19 the experimentally determined closed loop transfer functions, using the ETFE, and the predicted closed loop transfer functions are given. Compared to Fig. 8.3 or Fig. 8.7 it can be seen that the maximum amplitude of the rotor shaft torque variations are reduced by a factor of 10 in the frequency domain. For the resonance frequency of the rotor shaft, interpreted as a second order system, the controller increases the damping from less than 0.05 to more than 0.5. The figure also shows that the other signals (DC-current and generator speed) are behaving well, and that the control signal (delay angle rectifier) remains small.

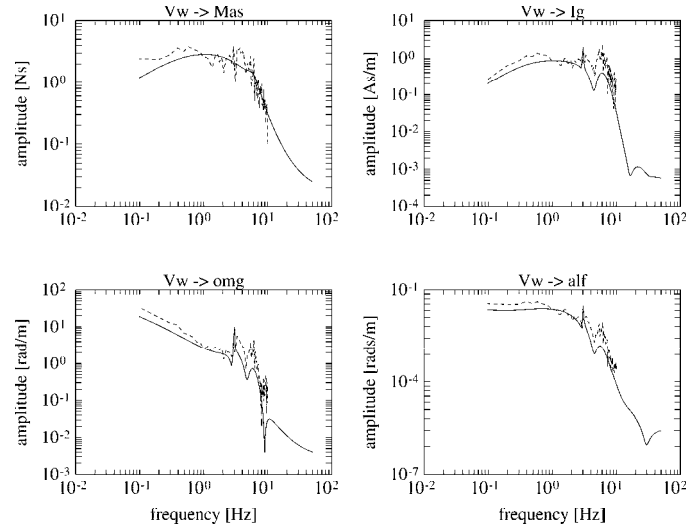


Fig. 8.19: Amplitude part of closed loop transfer functions, (—) designed, (---) ETFE, for the feedback configuration of Fig. 8.6. The titles of the plots refer to the specific inputs and outputs.

8.1.6 Application to nonlinear simulation model

To provide additional verification of the results of the model error estimation procedure of Chapter 5, the procedure is applied to a nonlinear simulation model of the wind turbine system, as supplied by the DUWECS simulation package for wind turbines of [15]. This simulation model is built from first principals, i.e. using the physical laws governing the system, and provides a model which gives a reasonably accurate description of the characteristics of the test-rig of the previous sections (Section 8.1.2 and Section 8.1.3).

The excitation signal α_r is chosen to be periodic, using a 5 Hz PRBS over 2000 points to construct one period, and a white noise wind disturbance V_w is applied to the system. Data was measured over 8 periods.

To estimate the error bounds we did proceed along the same lines as in Section 8.1.3. The results for the transfer function from α_r to I_g are given in Fig. 8.20 and Fig. 8.21. The results for the transfer function from α_r to ω_{sm} are given in Fig. 8.22 and Fig. 8.23.

To validate these results we will check whether the estimated error bounds can indeed capture the undisturbed output of the nonlinear simulation model for an input signal which was not used to estimate the error bounds. In Fig. 8.24 and Fig. 8.25 the true errors (the magnitude of the difference between the true undisturbed output and the output of the model, for I_g and ω_{sm} respectively) are given, together with the confidence intervals of Theorem 5.4.8. For illustration, the true undisturbed output ω_{sm}

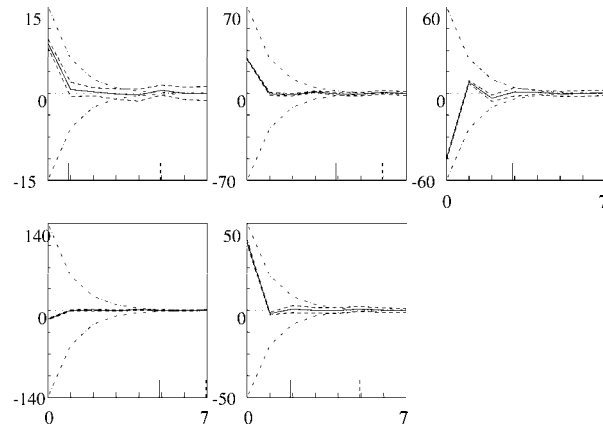


Fig. 8.20: Estimated generalized expansion coefficients (—), confidence interval (---), and estimated prior information (- · -), for the transfer function from α_r to I_g .

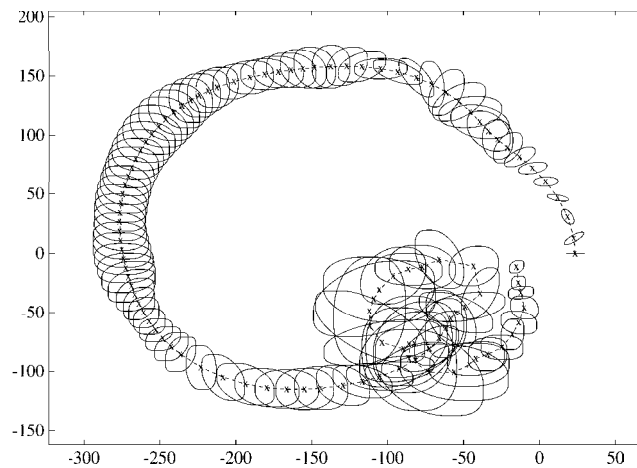


Fig. 8.21: Nyquist plot of the parametric transfer function estimate, and confidence region, for the transfer function from α_r to I_g .

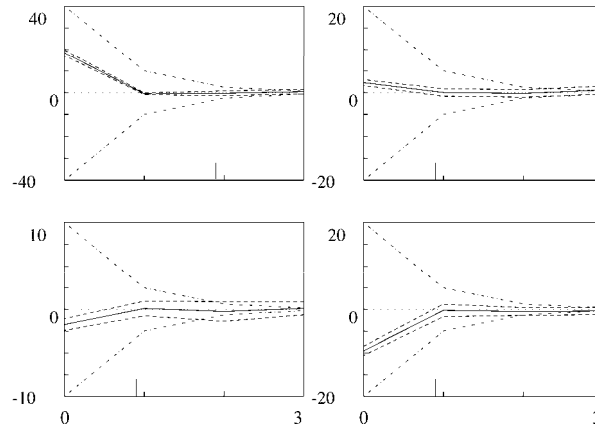


Fig. 8.22: Estimated generalized expansion coefficients (—), confidence interval (---), and estimated prior information (- · -), for the transfer function from α_r to ω_{sm} .

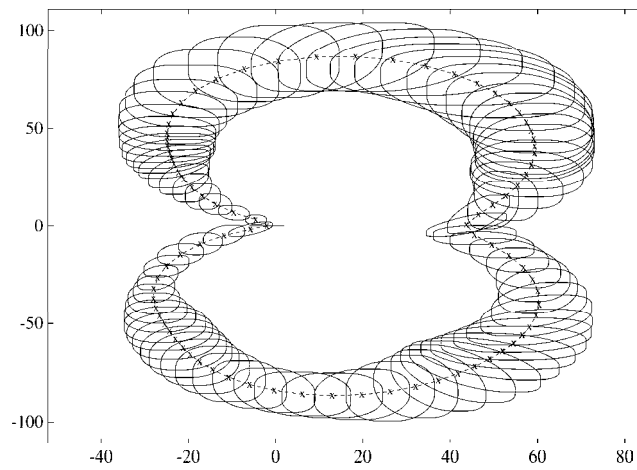


Fig. 8.23: Nyquist plot of the parametric transfer function estimate, and confidence region, for the transfer function from α_r to ω_{sm} .

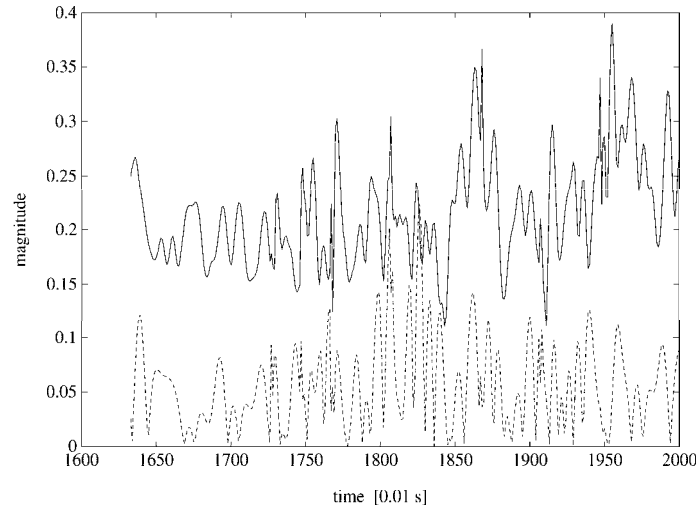


Fig. 8.24: True error $|y_o(\tau) - \hat{y}(\tau)|$ for I_g (---) , and confidence interval (—) .

and the confidence interval are given in Fig. 8.26. Clearly, the estimated error bounds (99 % confidence intervals) on the undisturbed output signals are correct (0.54 % of the samples of the true error are larger than the error bound for I_g , and 0.75 % for ω_{sm}). That is, the error bounds are able to capture the undisturbed output of the nonlinear system.

8.1.7 Conclusions

The model error estimation procedure of Chapter 5 has been applied to a laboratory set-up of a wind turbine system. Error bounds are calculated at a 99 % confidence level, providing evidence that the high performance controller designed in [14] to reduce rotor shaft torque variations should at least stabilize the system. Based on this result the controller is applied to the system. The controller indeed stabilizes the system, and achieves the same amount of reduction of rotor shaft torque variations for the true closed loop as for the designed closed loop. The rotor shaft torque variations are reduced by a factor of 10.

To provide additional verification of the results of the model error estimation procedure, the procedure also has been applied to a nonlinear simulation model of the wind turbine system. The estimated error bounds turn out to be correct. That is, the error bounds are able to capture the undisturbed output of the nonlinear system.

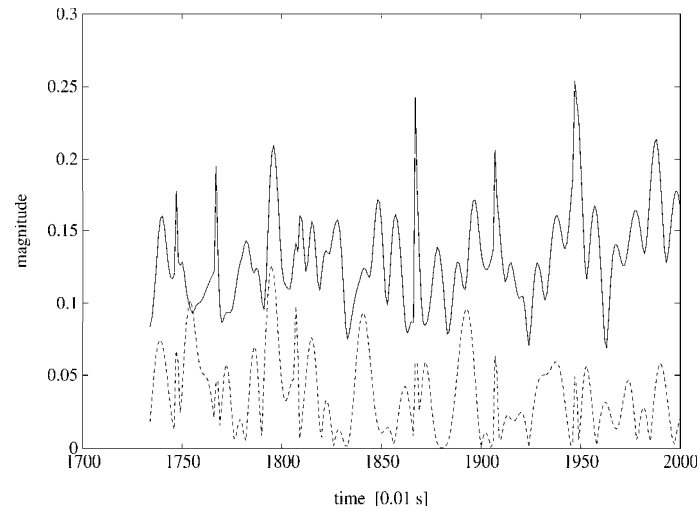


Fig. 8.25: True error $|y_o(\tau) - \hat{y}(\tau)|$ for ω_{sm} (---), and confidence interval (—).

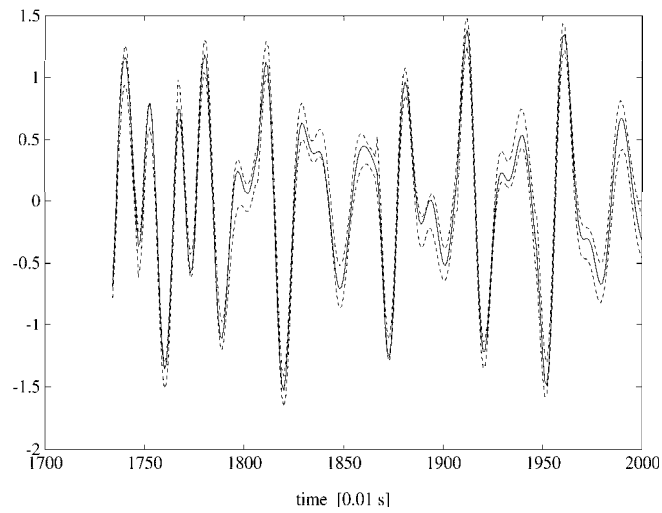


Fig. 8.26: True undisturbed output ω_{sm} (—), and confidence interval (---).

Acknowledgments

The work reported here has been done jointly with Peter Bongers of the Delft University of Technology [16, 33].

We would like to thank Tim van Engelen of the Netherlands Energy Research Foundation for helpful discussions and his cooperation in performing the measurements at the IRFLET test-rig, at ECN, Petten, The Netherlands.

8.2 Application to pick-up mechanism of CD player

8.2.1 Introduction

In this section we will apply the model error estimation procedure of Chapter 5 to the pick-up mechanism of a compact disc (CD) player (the Philips CDM9 mechanism).

A CD player uses an optical decoding device to reproduce high quality audio from a digitally coded signal, recorded as a spiral track on a reflective disc. An increasing amount of equivalent optical devices will be used in portable applications, having severe shock disturbances. The track following properties of a CD player, operating in these conditions, could be improved by designing an enhanced controller. In order to design such a controller, an accurate control relevant model of the pick-up mechanism is required. Additionally, an upper bound on the model error is required, in order to be able to design the controller to be robust against these errors, and to be able to evaluate stability and performance robustness prior to implementing the controller. A linear model, suitable for control design, of the system in the radial servo loop of the pick-up mechanism has been estimated in [175, 30]. The aim of this section is to estimate a bound on the model error.

The pick-up mechanism of a compact disc player is only marginally stable. Moreover, signals reflecting the position information of the pick-up mechanism relative to the disc are available only if the mechanism is controlled to remain in an operating point. Thus, experiments can only be done when this system is controlled. To be able to apply the model error estimation procedure of Chapter 5 in the resulting closed loop situation we will use the results of Chapter 6.

Additionally, utilizing both the estimated model and the estimated model error bounds, an enhanced robust controller for the radial servo loop has been designed in [30], while using the design procedure of [17]. This controller has been implemented, after verifying its robustness, showing a substantial performance improvement over the existing controller.

The whole procedure of identification, model error estimation, and controller design, can be seen as one step in the iterative identification and control design schemes for cautious controller enhancement proposed by [155, 175].

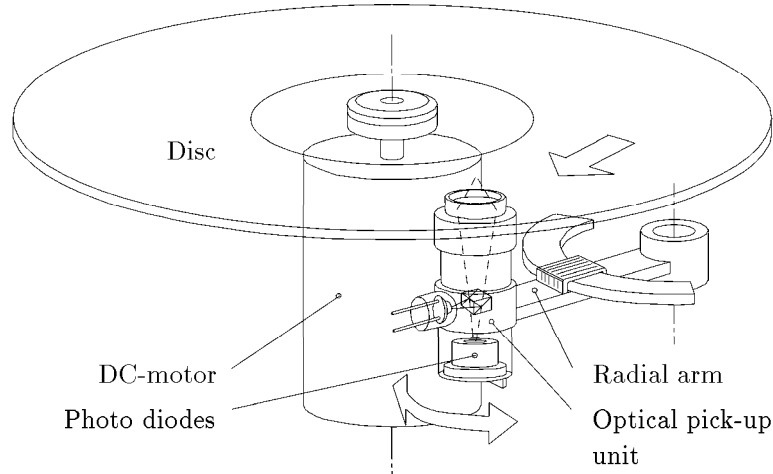


Fig. 8.27: Schematic view of CD mechanism [31].

8.2.2 The compact disc mechanism

In this section we will give a concise description of the compact disc pick-up mechanism, the experimental set-up, and the model obtained in [175, 30].

The CD mechanism consists of a DC-motor for the rotation of the compact disc and a radial arm in order to follow the track on the disc. Furthermore, an optical pick-up unit is mounted on the end of a balanced radial arm to read the digitally coded signal, recorded on the disc. Schematically the CD mechanism is given in Fig. 8.27.

A diode generates a laser beam that passes through a series of lenses in the optical pick-up unit to give a spot on the disc surface. The light reflected from the disc is measured on an array of photo diodes, mounted in the bottom of the optical pick-up unit, yielding the signals required for position error information of the laser spot on the compact disc.

The compact disc mechanism is a feedback controlled system. Following the track on the compact disc involves basically two control loops. Firstly a radial control loop using a permanent magnet/coil system mounted on the radial arm is employed, in order to position the laser spot orthogonal to the track. Secondly a focus control loop using an objective lens suspended by two parallel leaf springs and a permanent magnet/coil system, with the coil mounted in the top of the optical pick-up unit, is employed to focus the laser spot on the disc. In the present configuration, both the radial and focus control loops have been realized by a SISO (single input single output) controller, which consists of a lead-lag element, and a proportional and integrating (PI) action. The closed loop bandwidth is approximately 450 Hz, which is a compromise between several conflicting objectives, see [167].

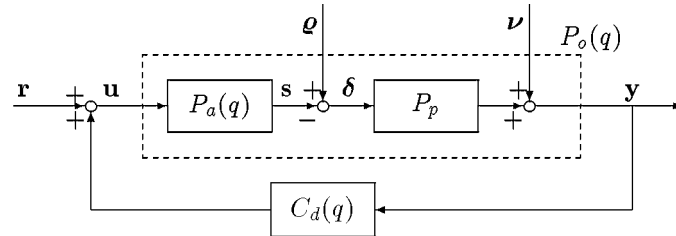


Fig. 8.28: Block diagram of the CD mechanism [31].

In Fig. 8.28 a block diagram of the two control loops is shown. In this figure $P_a(q)$ denotes the transfer function of the radial and focus actuators, P_p the optical pick-up unit, $C_d(q)$ the controller and the system P_o is $P_o(q) = -P_p P_a(q)$. The signals have the following interpretation. The spot position error $\delta(t)$, which is the difference between the track position $\boldsymbol{q}(t)$ and actuator position $\boldsymbol{s}(t)$ in radial and focus direction, generates a (disturbed) error signal $\boldsymbol{y}(t)$ via the optical pick-up unit P_p . This error signal $\boldsymbol{y}(t)$ is led into the controller $C_d(q)$ and feeds the actuators $P_a(q)$ with the input $\boldsymbol{u}(t)$. The signal $\boldsymbol{\nu}(t)$ reflects the disturbance on the error signal $\boldsymbol{y}(t)$.

The absolute track position $\boldsymbol{q}(t)$ and actuator position $\boldsymbol{s}(t)$ cannot be measured and used for identification directly. Only the error signal $\boldsymbol{y}(t)$ and the input $\boldsymbol{u}(t)$ are available. For identification purposes, an external reference signal $\boldsymbol{r}(t)$ can be injected into the control loops to excite the system, as illustrated in Fig. 8.28. This signal $\boldsymbol{r}(t)$ can be specified by the user, and is uncorrelated with the additive noise $\boldsymbol{\nu}(t)$.

While the compact disc mechanism is actually a multivariable system, for this study we will concentrate on the identification and control of the radial servo system only. Moreover, *we will proceed while considering the system P_o to be diagonal*. This is motivated by the fact that the radial and focus servo loop are nearly decoupled, see [42] and [31] where a multivariable model of the system P_o has been identified.

As mentioned, the current closed loop bandwidth of the radial servo loop is approximately 450 Hz. Increasing the bandwidth to 800 Hz to improve the low frequency disturbance rejection leads to excessive peaking of the sensitivity function. In order to design an improved controller a more accurate model is required, together with a characterization of the model error. Following the iterative identification and control design procedures proposed in [155, 175], experiments are performed while using the controller with the increased bandwidth. That is, to arrive at an accurate model which is suitable for control design aiming at a closed loop bandwidth of approximately 800 Hz, experiments should be performed that mimic the desired situation as close as possible. Additionally, error bounds which are estimated under conditions which resemble the desired operating conditions, are to be preferred in view of the extrapolation problem discussed in Section 3.11.

A control relevant linear dynamical model of the system G_o in the radial servo loop (the upper left element in the transfer function matrix P_o) has been identified in [175, 30], employing the coprime factor identification framework developed in [153, 175]. A relatively simple model of G_o was used to construct a suitable filter to generate an auxiliary signal $x(t)$ from $\mathbf{r}(t)$ such that

$$\begin{aligned}y(t) &= N_o(q)x(t) + v_n(t) \\u(t) &= D_o(q)x(t) + v_d(t)\end{aligned}$$

where N_o and D_o constitute a right coprime factorization of G_o (if P_o is diagonal), $G_o = N_o D_o^{-1}$, and where $v_n(t)$ and $v_d(t)$ are disturbance signals which are uncorrelated with $x(t)$. Subsequently, accurate (high order) models of N_o and D_o were estimated using system based orthonormal basis functions, to construct a new signal $x(t)$ which induces a new factorization N_o and D_o which is almost *normalized* right coprime. The use of normalized coprime plant factors has specific advantages both from an identification and control design point of view, see [175] for a short discussion. Finally, using the new signal $x(t)$, low order nominal models of N_o and D_o were estimated using standard least squares prediction error identification techniques, see [106]. Figure 8.29 shows the resulting high order models, and Fig. 8.30 the resulting low order nominal models.

8.2.3 Model error estimation

In this section we will apply the model error estimation procedure of Chapter 5 to the system G_o in the radial servo loop of the compact disc pick-up mechanism. This system contains a near double integrator, and position information of the pick-up mechanism relative to the disc is available only if the mechanism is controlled to remain in an operating point. Hence, experiments can be done only when the system is controlled. To handle the resulting closed loop situation we will use the coprime factor framework of [153] which is discussed in Chapter 6. That is, we will estimate confidence intervals for the factors N_o and D_o in a coprime factorization of the system G_o .

The identification problem now has the following form. Let $u(t)$ and $y(t)$ be the input and output of the system G_o in the radial servo loop (the upper left element of P_o in Fig. 8.28)

$$\begin{aligned}y(t) &= G_o(q)u(t) + v(t) \\u(t) &= C(q)y(t) + r(t)\end{aligned}$$

where $r(t)$ is an external excitation signal (the first component of the reference signal $\mathbf{r}(t)$ in Fig. 8.28), $v(t)$ is a disturbance signal which is uncorrelated with $r(t)$, and C is the controller in the radial servo loop (the upper left element of C_d in Fig. 8.28). Introduce the auxiliary variable $x(t)$ as a filtered version of $r(t)$

$$x(t) = F(q)r(t) \tag{8.2}$$

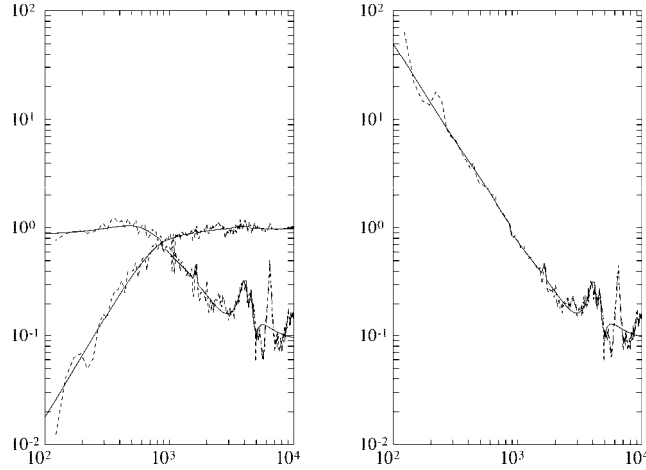


Fig. 8.29: Bode magnitude plots of the high order model [175].
 Left: Identified coprime plant factors \hat{N} and \hat{D} (—) , and
 ETFE of the same factors (- -) .
 Right: Estimated plant model $\hat{N}\hat{D}^{-1}$ (—) , and ETFE (- -) .

where the filter F is chosen as proposed in [175, 30]. We now have (see [175, 30])

$$\begin{aligned} y(t) &= N_o(q)x(t) + v_n(t) \\ u(t) &= D_o(q)x(t) + v_d(t) \end{aligned}$$

where $v_n(t)$ and $v_d(t)$ are disturbance signals which are uncorrelated with $x(t)$, and where N_o and D_o constitute a right coprime factorization of G_o (if P_o is diagonal)

$$G_o = N_o D_o^{-1}$$

We will now estimate models and model error bounds for both N_o and D_o from $u(t)$, $y(t)$ and $x(t)$.

Exciting only the first component $r(t)$ of the reference signal $\mathbf{r}(t)$ in Fig. 8.28, time sequences $u(t)$ and $y(t)$ in the radial control loop were measured, using appropriate anti-aliasing filters and a sampling frequency of 25 kHz. The excitation signal $r(t)$ was chosen to be periodic, using a band limited white noise signal in the frequency domain of interest (100 Hz – 10 kHz) to construct one period. This is the same excitation signal as was used in [175, 30]. The filter F in (8.2) was chosen to be the final filter obtained in the example given in [175].

To estimate the error bounds we have $N = 8192$, $N_o = 1024$ and $N_s = 250000$. Hence, we use data over 8 periods of the excitation signal $r(t)$, which at a sampling

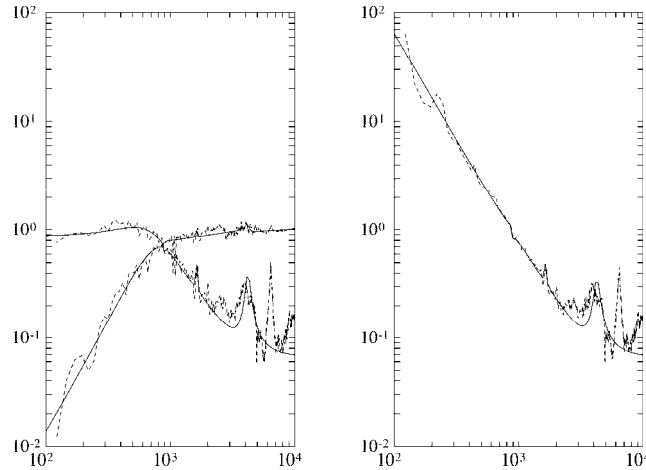


Fig. 8.30: Bode magnitude plots of the low order nominal model [175].
 Left: Identified coprime plant factors \hat{N}_n and \hat{D}_n (—), and
 ETFE of the same factors (---).
 Right: Estimated plant model $\hat{N}_n \hat{D}_n^{-1}$ (—), and ETFE (---).

frequency of 25 kHz concerns a time interval of 0.33 seconds.

The signal $x(t)$ is a filtered version of the periodic excitation signal $r(t)$, see (8.2). Due to the long time interval (given by N_s) relative to the significant length of the impulse response of the filter F in which the periodic external excitation signal $r(t)$ was applied to the closed loop system prior to starting the measurements, the signal $x(t)$ used for identification is almost periodic. However, since $x(t)$ is not really periodic an additional error is introduced with regard to the error bounds developed in Chapter 5 where the input signal is assumed to be periodic. This additional error term can be bounded using Theorem 6.3.3 in Chapter 6. The magnitude of the DFT of the signal $x(t)$ over one period of $r(t)$ is given in Fig. 8.31.

The nominal models of [175] as given in Fig. 8.30 turned out not to be suitable as basis generating systems. The matching estimated generalized expansion coefficients did not converge to zero at a satisfactory rate. We therefore did try to estimate more suitable models from the measured data in the following way. Firstly, for each of the two transfer functions N_o and D_o separately, we used the nominal model of [175] as a basis generating system. The resulting orthonormal basis functions were used to estimate a high order model from the data. This high order model is compared to the ETFE to check whether all significant dynamics are modelled. If not, the model order is increased. To again obtain a low order model, we used model reduction

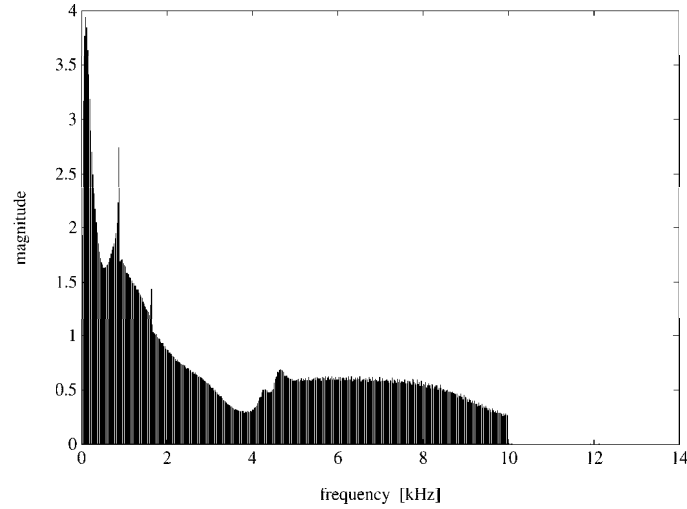


Fig. 8.31: Magnitude of the DFT of $x(t)$ over one period of $r(t)$, $|X_i(e^{j\omega_k})|$.

by balancing and truncation (see [128]) or by fitting a low order model on the high order one in the frequency domain using weighted least squares, trying to minimize the infinity norm of the error by adapting the weighting function (i.e. $W_{\text{new}}(\omega_k) = W_{\text{old}}(\omega_k)|G(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})|$ for fitting \hat{G} on G , and visual inspection of the fit in a Nyquist plot to decide when the fit is satisfactory). This lower order model in turn was used as a basis generating system. Iterating in this way we obtained basis generating systems which perform better than the nominal models of [175], and which provide a reasonable fit on the ETFE.

The values for M , ρ , \mathcal{K} , and η are obtained from the data iteratively. Starting from zero, we iterate till the estimated prior information is in accordance with the matching confidence intervals, i.e. till the bounds given by the estimates of the prior information cross the resulting confidence intervals only where zero is contained in the confidence interval. These estimated values \hat{M} , $\hat{\rho}$, $\hat{\mathcal{K}}$, and $\hat{\eta}$ now are taken as the prior information of Assumption 5.2.1.ii and Assumption 5.2.1.iii.

With regard to the above we have to note that we easily succeeded to find satisfactory values for \hat{M} and $\hat{\rho}$. However, especially for N_o , it was difficult to establish adequate values for $\hat{\mathcal{K}}$ and $\hat{\eta}$. We were able to find reasonable values for $\hat{\mathcal{K}}$ and $\hat{\eta}$ which are consistent with the resulting confidence intervals for the generalized expansion coefficients for a fixed model order, but we had to allow for a considerable contribution due to undermodelling in the error bounds (relative to the variance contribution). However, when the model order is increased to reduce the bound on the undermodelling error, the values of $\hat{\mathcal{K}}$ and $\hat{\eta}$ are again invalidated. This can of course be caused

by incorrect values for $\hat{\mathcal{K}}$ and $\hat{\eta}$, but we were not able to find reasonable values for $\hat{\mathcal{K}}$ and $\hat{\eta}$ for which this did not occur. Thus, especially for N_o , we tend to the opinion that there is no convergence of the estimated generalized expansion coefficients to zero, although all coefficients following the leading ones are small. Based on the similarity of the difficulties to find satisfactory values for $\hat{\mathcal{K}}$ and $\hat{\eta}$ with the observations made in Section 5.9, we suspect that some structural effect is present in the data which can not be captured by a linear time invariant model.

There are basically three sources that can generate a structural effect in the data which can not be captured by a linear time invariant model with stochastic noise: a periodic disturbance which has approximately the same period length as the input signal, nonlinearities in the system, and a system which is periodically time varying with a period length that is approximately equal to the period length of the input signal. The latter is however not at all likely, and time variations in the system also are not expected from the physics of the system. Hence, we suspect that there are nonlinearities in the system, or that a periodic disturbance has affected the data. The presence of nonlinear contributions to the data is however not confirmed by the higher order correlation tests of [10]. The higher order correlations are small, and approximately have the same magnitude as the correlation between the input and the model output error. The autocorrelation of the model output errors does however reveal that there is a strong periodic component in the output errors: the signal and the signal shifted over 500 samples are highly correlated, see Fig. 8.32. Taking the DFT of the autocorrelations shows that for N_o the periodic component contains strong 100, 200 and 300 Hz contributions, and for D_o strong 300, 800 and 1100 Hz contributions.

Together with the observations made in Section 5.9, the presence of the periodic disturbance, which has approximately the same period length as the input signal, explains the lack of convergence of the generalized expansion coefficients, and the resulting difficulty to find adequate values for $\hat{\mathcal{K}}$ and $\hat{\eta}$. According to the results of Section 5.9, the influence of this structural effect in the data, which can not be captured by a linear time invariant model with stochastic noise, should be accounted for by the undermodelling (i.e. the structural) contribution to the error bound. This explains the considerable error due to undermodelling for which we had to allow in order to establish values for $\hat{\mathcal{K}}$ and $\hat{\eta}$ which are consistent with the resulting confidence intervals for the generalized expansion coefficients.

A further confirmation of the presence of a structural effect in the data which can not be captured by a linear time invariant model with stochastic noise, follows from the nonparametric transfer function estimates and their confidence intervals as given by Theorem 5.3.5. The nonparametric estimate of N_o is depicted in Fig. 8.33, together

¹At this point we should have decreased the elements of the weighting matrix W of Section 5.4.2 around these frequencies. However, since this is only an example, we will not do this in order to verify whether reliable error bounds are obtained under this "worst case" disturbance.

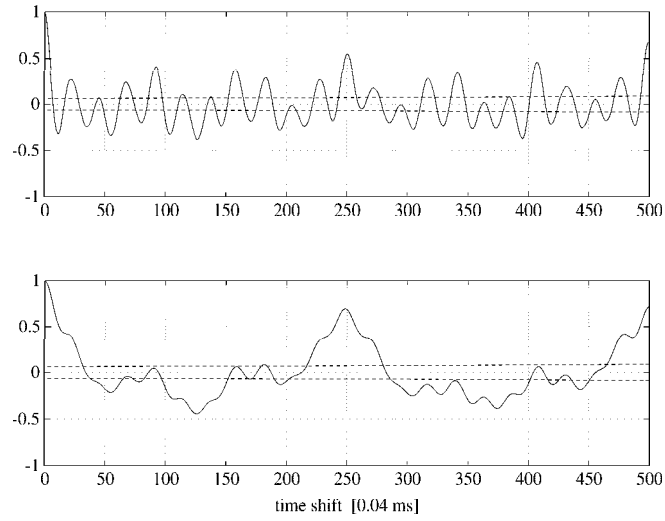


Fig. 8.32: Autocorrelation of the model output error for D_o and N_o respectively.

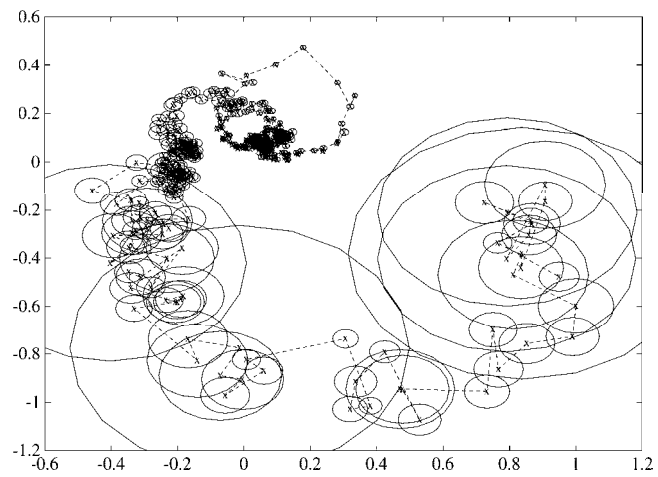


Fig. 8.33: Nyquist plot of nonparametric transfer function estimate for N_o , and confidence region.

with the confidence intervals of Theorem 5.3.5. From the confidence intervals it clearly follows that the system is either of shocking high order, or that there is some structural effect in the data which can not be captured by a linear time invariant model with stochastic noise. Such a structural effect will cause the confidence intervals on the nonparametric estimate to be too small, see Section 5.9. Similar observations follow from the nonparametric estimate of D_o .

In addition to the above one should be aware of the following general fact. When the significant order of a system is high, this implies that for any low order basis generating system the generalized expansion coefficients will only slowly converge to zero. When in addition the variation of the measured signals over the different data segments is small, then high order models are required to achieve the situation where zero lies in the tail of confidence intervals on the estimated generalized expansion coefficients, at a reasonable level of the undermodelling contribution to the confidence intervals. Together with the above described difficulty to find satisfactory values for $\hat{\mathcal{K}}$ and $\hat{\eta}$, this explains the high order models which we will use.

We will now proceed with the model error estimation procedure. The probability level (as given by F_α) for the confidence regions is specified to be 99 %, unless stated otherwise. We will first give the results for D_o , and next the results for N_o . Figure 8.34 shows the estimated generalized expansion coefficients and the confidence interval of Theorem 5.4.2 and Theorem 6.3.3, using $n_p = 111$ and $W = \text{diag}(1/\hat{\sigma}_r(\hat{D}(e^{j\omega_k})))$, together with the prior information given by $\hat{\mathcal{K}}$ and $\hat{\eta}$. The solid vertical lines in the plots indicate that zero lies in the confidence interval from thereon, the dashed vertical lines indicate that the prior information crosses the estimate. Experience shows that, for the prior information to be realistic, the latter should come at least two coefficients later than the former, and that the tail of the confidence interval which contains zero should be at least two or three (depending on the preceding behaviour of the confidence interval) coefficients long. Figure 8.35 shows the estimated impulse response, the prior information given by \hat{M} and $\hat{\rho}$, and the confidence interval of Theorem 5.4.8 and Theorem 6.3.3 using $n_y = 111$ and $\mathcal{W}_y = 1$. The Nyquist plot of the parametric transfer function estimate is given in Fig. 8.36, together with the confidence region of Theorem 5.4.5 and Theorem 6.3.3, using $n_f = 111$ and $\mathcal{W} = 1$. To illustrate the influence of the probability level on the size of the confidence regions, Fig. 8.37 depicts the amplitude of confidence regions for the parametric transfer function estimate at a 90 %, a 99 %, and a 99.9 % probability level. The 99.9 % confidence region is at most a factor two larger than the 90 % confidence region. Finally, Fig. 8.38 gives the Bode plots of the nonparametric estimate, the parametric estimate, and the basis generating system. The parametric estimate very accurately fits on the nonparametric estimate, and is considerably better than the basis generating system.

The results for N_o are given in Fig. 8.39, Fig. 8.40, Fig. 8.41 and Fig. 8.43. We used $W = \text{diag}(1/\hat{\sigma}_r(\hat{N}(e^{j\omega_k})))$, $n_p = 71$, $n_y = 71$, $n_f = 71$, $\mathcal{W}_y = 1$ and $\mathcal{W} = 1$. Figure 8.42 depicts the components of the error bound for the parametric transfer function

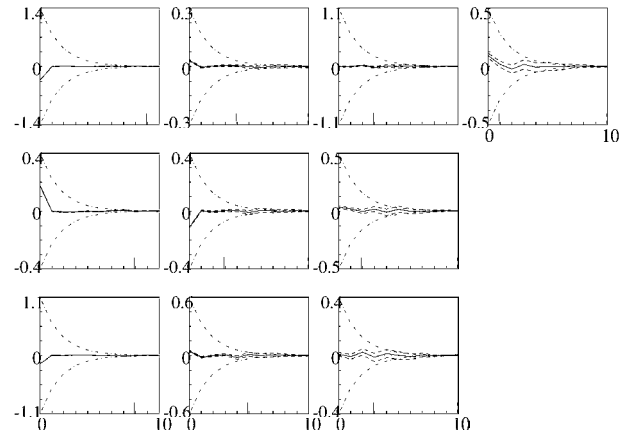


Fig. 8.34: Estimated generalized expansion coefficients (—), confidence interval (---), and estimated prior information $\hat{\mathcal{K}}\hat{\eta}^k$ (-·-), for D_o .

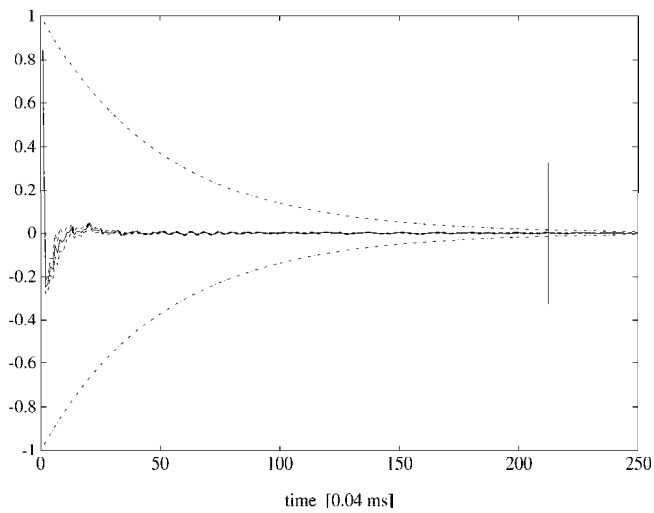


Fig. 8.35: Estimated impulse response (—), confidence interval (---), and estimated prior information $\hat{M}\hat{\rho}^{-k}$ (-·-), for D_o .

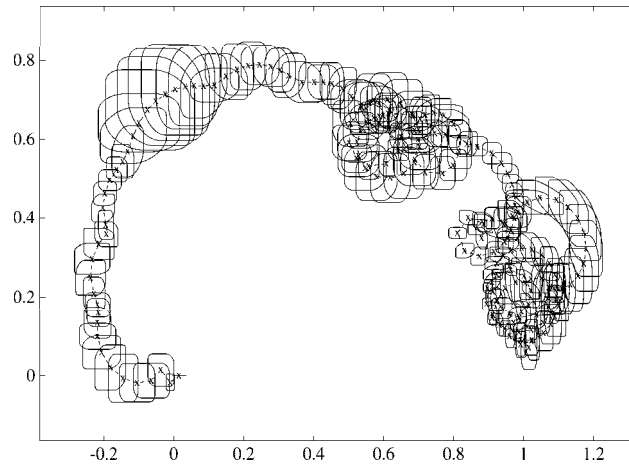


Fig. 8.36: Nyquist plot of parametric transfer function estimate of D_o , and confidence region.

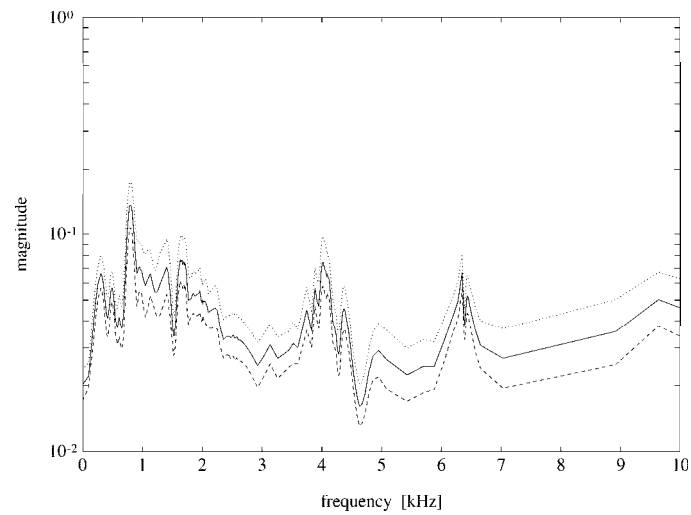


Fig. 8.37: Amplitude of confidence regions for the parametric transfer function estimate at a 90 % (- -), a 99 % (—), and a 99.9 % (· · ·) probability level.

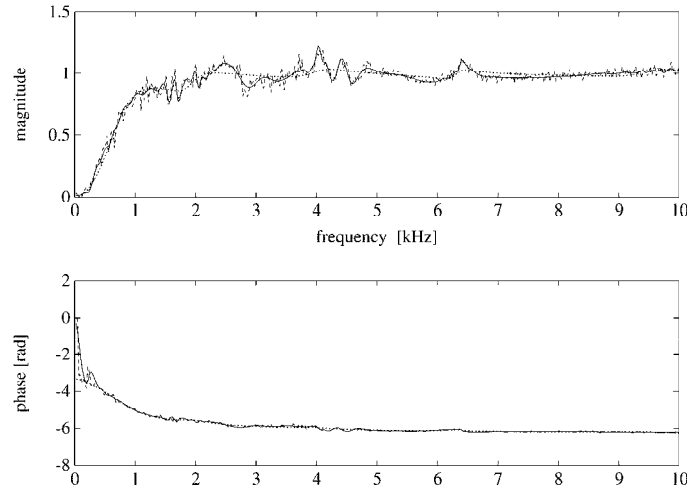


Fig. 8.38: Bode plot of the basis generating system (\cdots), and the nonparametric ($- -$) and the parametric (—) transfer function estimates, for D_o .

estimate, showing the considerable contribution due to undermodelling, relative to the variance contribution.

The nonparametric and parametric transfer function estimates of the system $G_o = N_o D_o^{-1}$ are given in Fig. 8.44. Figure 8.44 shows that the models, although of very high order, do not contain dynamics which are not supported by the data. This is confirmed by the comparison of the nonparametric and parametric transfer function estimates for D_o and N_o individually, see Fig. 8.38 and Fig. 8.43 respectively.

To validate these results we will check whether the estimated error bounds can indeed capture the measured (disturbed) output of the system for an input signal which was not used to estimate the error bounds. In Fig. 8.45, Fig. 8.47, Fig. 8.48 and Fig. 8.49 the real and imaginary part of the DFT of the measured signals $u(t)$ and $y(t)$ are given respectively, together with the confidence intervals of Corollary 5.4.12 and Theorem 6.3.3. For illustration, the real part of the DFT of the measured signal $u(t)$ is given in Fig. 8.46, together with the confidence interval. Clearly, the estimated error bounds (99 % confidence intervals) can account for the observed data.

Finally, a new robust controller has been designed for the system in [30], employing the coprime factor H_∞ control design method of [17]. This control design method optimizes robustness against additive perturbations on coprime factors of the system, and can be used to incorporate the estimated uncertainty bounds on the coprime factors in the control design phase. That is, the estimated uncertainty bounds are used to specify the

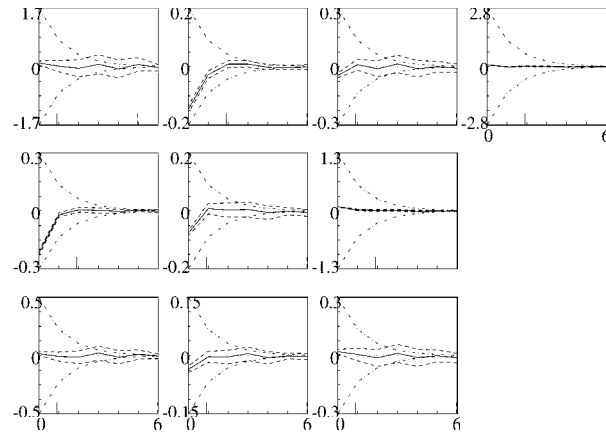


Fig. 8.39: Estimated generalized expansion coefficients (—), confidence interval (---), and estimated prior information (- · -), for N_0 .

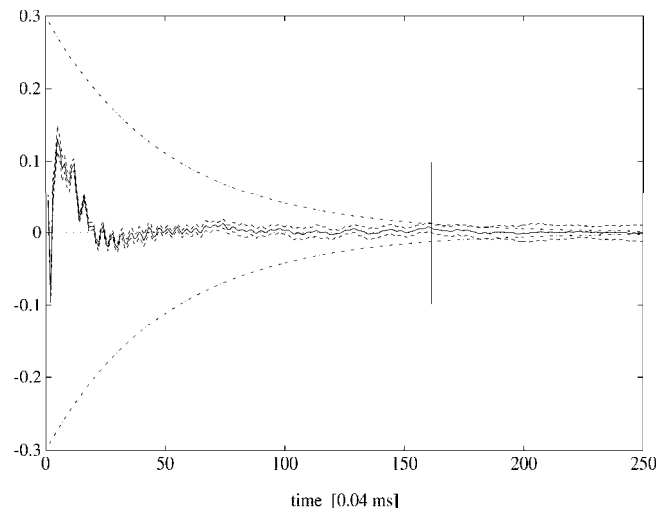


Fig. 8.40: Estimated impulse response (—), confidence interval (---), and estimated prior information (- · -), for N_0 .

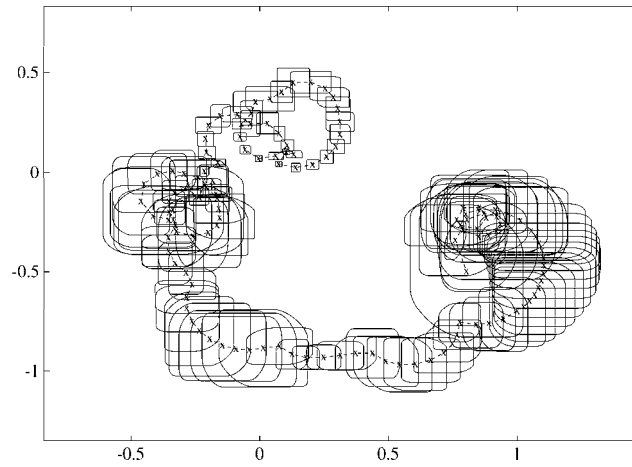


Fig. 8.41: Nyquist plot of parametric transfer function estimate of N_o , and confidence region.

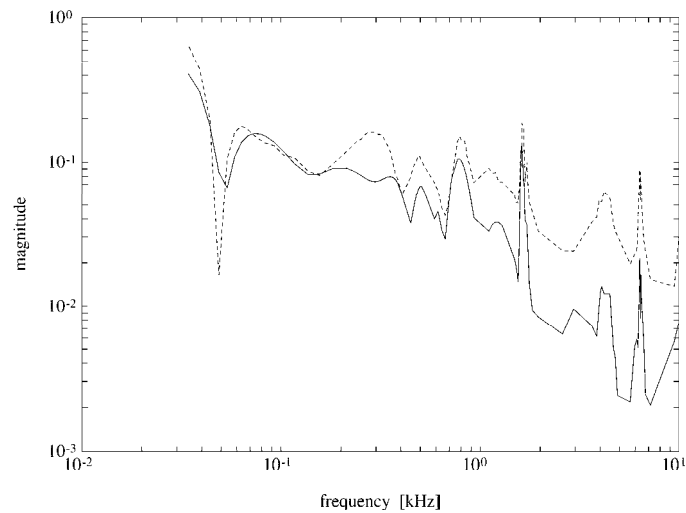


Fig. 8.42: Components of the confidence interval for the parametric transfer function estimate of N_o ; variance (—), undermodelling (---), and unknown past inputs (···). The contribution due to unknown past inputs is too small to be seen (less than 10^{-14}), and there is no error due to weighting (since $\mathcal{W} = 1$).

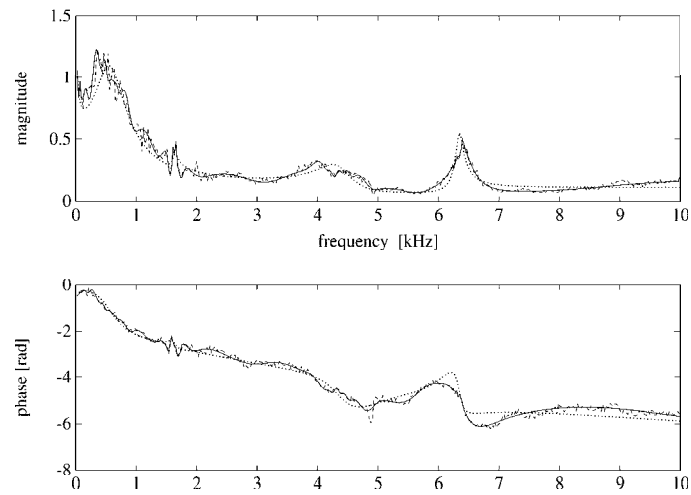


Fig. 8.43: Bode plot of the basis generating system (\cdots), and the nonparametric ($- -$) and the parametric (—) transfer function estimates, for N_o .

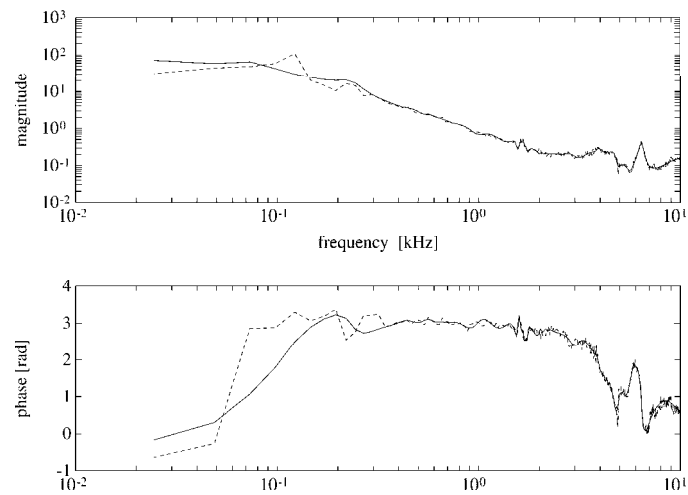


Fig. 8.44: Bode plot of the nonparametric ($- -$) and the parametric (—) transfer function estimates of the system $G_o = N_o D_o^{-1}$.

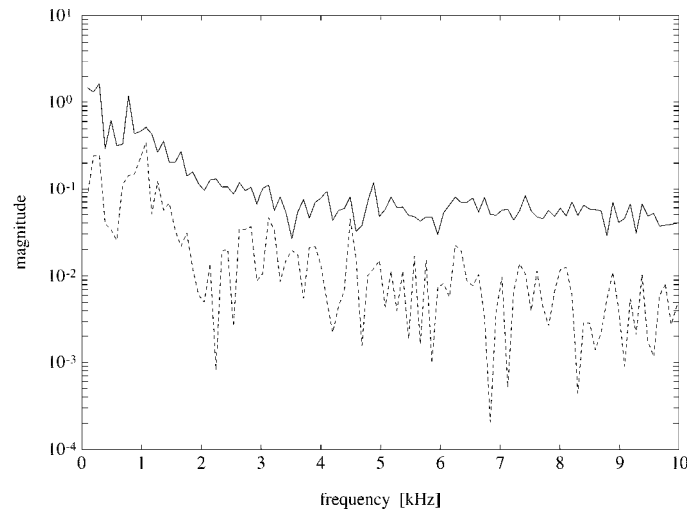


Fig. 8.45: Magnitude of the difference between the real part of the DFT of the measured output and the output of the model over new data $|\text{Re}\{U^q(e^{j\zeta_\ell}) - \tilde{U}^q(e^{j\zeta_\ell})\}|$ (—), and confidence interval (— —).

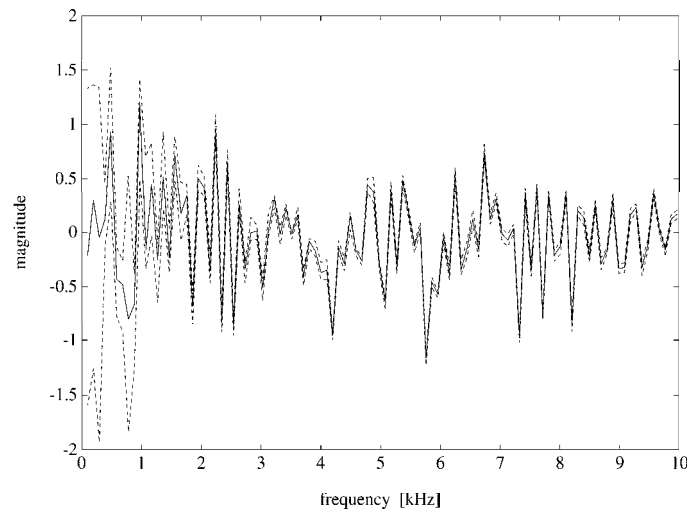


Fig. 8.46: Real part of the DFT of the measured output over new data $\text{Re}\{U^q(e^{j\zeta_\ell})\}$ (—), and confidence interval (— —).

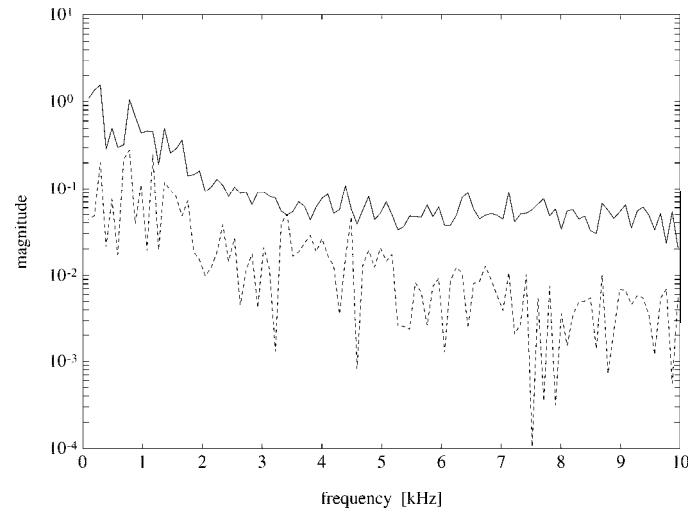


Fig. 8.47: Magnitude of the difference between the imaginary part of the DFT of the measured output and the output of the model over new data $|\text{Im}\{U^q(e^{j\zeta\epsilon}) - \hat{U}^q(e^{j\zeta\epsilon})\}|$ (---), and confidence interval (—).

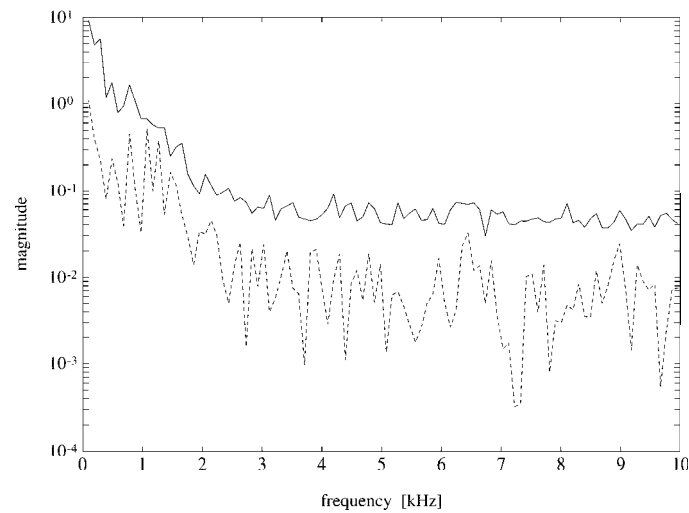


Fig. 8.48: Magnitude of the difference between the real part of the DFT of the measured output and the output of the model over new data $|\text{Re}\{Y^q(e^{j\zeta\epsilon}) - \hat{Y}^q(e^{j\zeta\epsilon})\}|$ (---), and confidence interval (—).

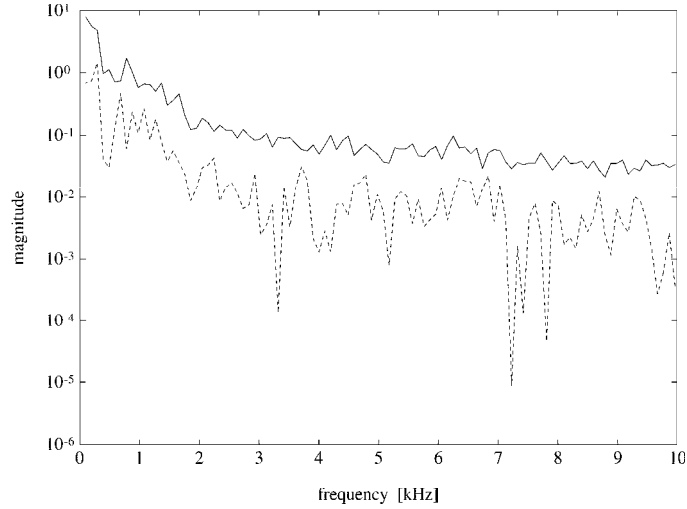


Fig. 8.49: Magnitude of the difference between the imaginary part of the DFT of the measured output and the output of the model over new data $|\text{Im}\{Y^q(e^{j\zeta\epsilon}) - \hat{Y}^q(e^{j\zeta\epsilon})\}|$ (—), and confidence interval (—).

weighting functions that are used in designing the controller. Additionally, using the estimated uncertainty bounds, we can check whether the new controller should stabilize the true system. Let $C = \tilde{D}_c^{-1}\tilde{N}_c$ denote a left coprime factorization of the controller C in the radial servo loop. Let $G = ND^{-1}$ denote a right coprime factorization of a model G . Let $G_o = (N + \Delta_N)(D + \Delta_D)^{-1}$ denote a right coprime factorization of the system G_o . Then (if P_o is diagonal) a sufficient condition for robust stability of the radial servo loop is given in [17] as

$$\left\| [\tilde{N}_c \quad \tilde{D}_c] \begin{bmatrix} \Delta_N \\ \Delta_D \end{bmatrix} \Lambda^{-1} \right\|_{\infty}$$

where

$$\Lambda = \tilde{D}_c D + \tilde{N}_c N$$

This condition can be checked (conservatively) as

$$\left| \frac{\tilde{D}_c}{\Lambda} \right| |\Delta_D| + \left| \frac{\tilde{N}_c}{\Lambda} \right| |\Delta_N| \leq 1 \quad (8.3)$$

while using the estimated bounds on $|\Delta_N|$ and $|\Delta_D|$. The robustness properties of this controller are given in Fig. 8.50, depicting (8.3) for point wise confidence intervals with on the estimated coprime factors with different probability levels. This new controller

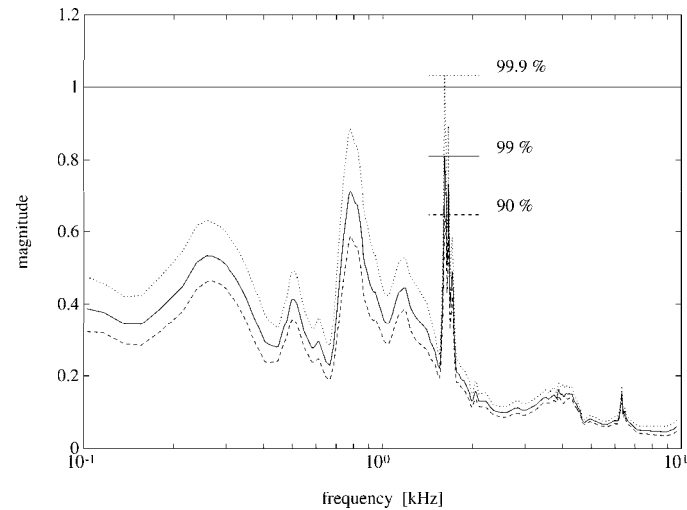


Fig. 8.50: Stability robustness tests of the new controller for point wise confidence intervals on the estimated coprime factors at three probability levels.

achieves a substantially better performance than the existing controller, see Fig. 8.51. A bandwidth of approximately 800 Hz has been achieved without excessive peaking of the sensitivity function, leading to improved disturbance rejection.

8.2.4 Conclusions

The application of the model error estimation procedure of Chapter 5 to the system in the radial servo loop of a compact disc pick-up mechanism was successful.

Experiments can only be performed when this system is controlled. To handle the closed loop situation we used the coprime factor framework of [153], and the results of Chapter 6. Although very high order models are necessary to accurately describe the coprime factors of the system, and in spite of a periodic disturbance which acts on the system, we obtained model error bounds which perform well when validated. That is, together with the estimated bounds on the DFT of the noise, the model error bounds turn out to provide a reliable and tight bound for the error in the DFT of the output of the models when compared to the DFT of the measured output, while using measurement data which was not used to estimate the error bounds. On the basis of these validation results and the experience of Section 5.9 with regard to structural deviations from the case of a linear time invariant system with stochastic noise, we expect the error bounds to provide a realistic indication of the model error.

A new robust controller has been designed for the system, while using the estimated uncertainty bounds on the coprime factors in the design phase. This new controller has

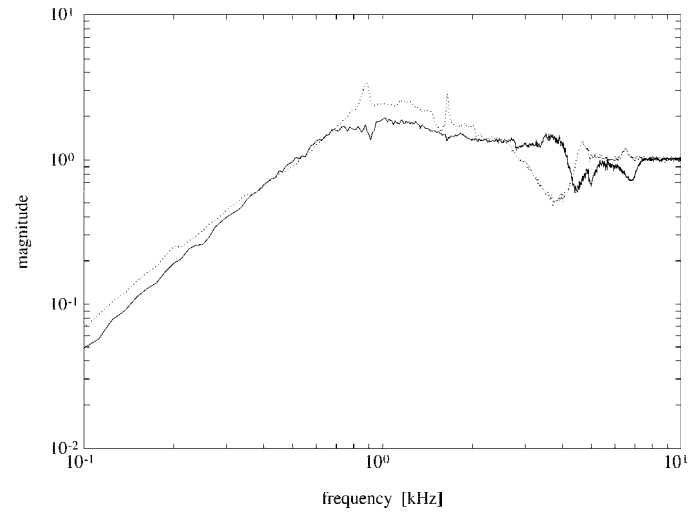


Fig. 8.51: Measured magnitude of the upper left element of the sensitivity matrix $(I - C_d P_o)^{-1}$ for the new (—) and the old (···) controller.

been implemented, and achieves a substantially better performance than the existing controller. We conclude that the estimated error bounds indeed do provide an excellent basis for high performance robust control design.

Acknowledgments

The work reported here has been done jointly with Raymond de Callafon of the Delft University of Technology [30].

We would like to thank Gerrit Schootstra and Maarten Steinbuch of the Philips Research Laboratories, Eindhoven, The Netherlands, for helpful discussions and their cooperation in performing the measurements at the CD player.

Chapter 9

Conclusions and recommendations

9.1 Conclusions

The problem addressed in this thesis was formulated in Chapter 1 and Chapter 3, and reads as follows. Consider the situation where the system is stable, and where the dominant behaviour of the system is linear and time invariant. Given a set of measured pairs of input and output data points, establish the set of all stable linear time invariant systems for which there is indeed evidence in the measured data that these systems could have generated this data. Additionally, solutions to this problem are required to be able to provide models and model error bounds which are suitable for high performance robust control design, and a realistic indication of the deviations between the model and the system still should result when the assumption that the system is linear and time invariant is mildly violated.

This problem was discussed extensively in Chapter 3, and existing literature addressing this problem was evaluated in Chapter 2. Based on these chapters, as a first step a frequency domain approach using the unknown but bounded noise framework was presented in Chapter 4. This approach was improved in Chapter 5 by employing the more appropriate description of the noise as being a realization of a stochastic process. Further developments were presented in Chapter 6, considering unstable systems and closed loop situations, and in Chapter 7, providing asymptotic bias and variance expressions for the estimation procedure proposed in Chapter 5 (frequency domain least squares identification using generalized orthonormal basis functions). Simulation studies were provided in Chapter 5, and two applications were given in Chapter 8.

We conclude that the procedure presented in Chapter 5 indeed is able to provide accurate models together with reliable and tight error bounds, which are suitable to serve as a basis for high performance robust control design. Moreover, essentially no prior information is needed (can be estimated from the data), and only minor assumptions on the noise are required (essentially only a mixing condition).

The only important restriction that is imposed is the assumption that the system

is linear and time invariant. However, by means of simulations and applications it has been made plausible that the procedure is robust to mild violations of the assumptions (linear time invariant system, and stationary stochastic noise). The uncertainty due to the fact that the prior information is estimated from the data is not taken into account. Still, simulations show the probability level of the error bounds to be too low rather than too high. This is due to the fact that the components of the error bound that rely on this prior information are worst case bounds, which is almost always too pessimistic. Finally, the confidence interval on the error due to the noise is only valid asymptotically in the period length of the input signal. However, the procedure displays an excellent nonasymptotic performance in the simulation and application studies.

The basic ideas behind our procedure are the following. A stochastic noise model is used to account for averaging (random) errors, whereas a deterministic framework is used for the undermodelling error to account for structural errors. Data segmentation in combination with a periodic input is used to effectively distinguish between averaging and undermodelling errors, and to estimate the variance of the noise. The estimate of the variance is consistent in the face of undermodelling, and the distribution of the error due to the noise is an asymptotic property of the estimation procedure. A closed form (accounting for the fact that the variance is estimated from the data) confidence interval on the error due to the noise can be established. The prior information also is estimated from the data. As a result, the uncertainty in the data is indeed reflected in the error bounds. Finally, the model set is constructed using system based orthonormal basis functions, in order to enable the identification of high quality models. These ideas appear to be sound, robust, and widely applicable, as is confirmed by the simulation examples for situations which deviate from our assumptions, and practical applications. Topics as model order selection, input design and model validation are also addressed, and an asymptotic analysis is provided for frequency domain transfer function estimation using system based orthonormal basis functions.

The procedure can very well be used in an iterative control design scheme (see Section 2.8), but is not fit for an adaptive control scheme, since we require the input signal to be periodic. However, it appears to be possible to adapt the procedure to allow for quasi-stationary input signals, see Section 9.2.

We can conclude that the problem addressed in this thesis, as formulated in Section 3.1 and Section 3.12, to a large extent has been solved. The approach of Chapter 5 definitely appears to be a powerful one. However, the system is assumed to be linear and time invariant, and the noise to be stochastic and stationary. Although the procedure is robust to mild deviations from these assumptions, this remains to be an idealization of the processes encountered in reality. Tighter and more reliable error bounds should be possible if the problem of nonlinearities and time variations would be addressed in a more structured way. Knowledge from physical modelling appears to be the key ingredient to properly address this topic, see Section 9.2. In our opinion, it

is one of the major challenges and it should be one of the major concerns of the system identification community to address the problem of time varying nonlinear systems, preferably in a physical parameterization. The approach of Chapter 5 also appears to be promising in this more general context, see Section 9.2. Nevertheless, the approximations involved in the linear time invariant framework with stationary stochastic noise are often justified and lead to good results in many cases. The engineering leap of faith (also because the priors are estimated from the data) remains, but has been decreased considerably.

9.2 Recommendations for further research

There are three important topics which need further attention. Firstly, a version of the procedure of Chapter 5 should be established which does not require the input signal to be periodic, so that the procedure is more generally applicable and can be used in an adaptive control scheme. Secondly, it would be advantageous when tighter bounds for the error due to undermodelling could be established. Thirdly, more structure should be imposed on the error due to nonlinearities and time variations, in order to obtain tighter and more reliable error bounds for mildly nonlinear and time varying systems. In the following we will address these three points.

A version of the procedure of Chapter 5 which does not require the input signal to be periodic is desirable in the following two situations. Firstly, in some cases the input signal to the system, or the reference signal to the closed loop in which the system operates, cannot be specified by the user (e.g. process industry). That is, a periodic (external) excitation signal cannot be applied. To avoid confusion, for systems that operate in closed loop it is sufficient if the reference signal (external excitation signal) can be chosen to be periodic, see Chapter 6. Secondly, quite a lot of systems exhibit slow variations in their dynamics (time variations or shifting operating conditions) which are significant with respect to robust high performance control. That is, these variations are large enough to significantly deteriorate the performance. Often these variations are not known in advance, so that an adaptive controller is required to maintain high performance. Now, in an adaptive control scheme, the reference signal will in general not be periodic.

It certainly appears to be possible to develop a time domain analog of the procedure of Chapter 5, without the requirement of a periodic input signal, when the input signal can be modelled as a quasi-stationary stochastic process. Assuming the input signal to be quasi-stationary, and using time domain least squares prediction error identification, the parameter and the transfer function estimates over the different data segments again are asymptotically identically normally distributed, see Section 2.2 and [108, 189, 113, 104, 111]. Thus, we can apply exactly the same procedure as the one that was developed in Chapter 5.

Tighter error bounds for the error due to undermodelling would of course be advantageous. The currently used bounds are often somewhat pessimistic due to their worst case character. On the other hand, these bounds are reliable and appear to contribute substantially to the robustness of the model error estimation procedure with respect to mild violations of the assumptions.

A very effective possibility to reduce the bound on the undermodelling error would be to assume and estimate different rates of convergence for the different elements of the vector sequence of orthonormal basis functions (and to estimate a different number of parameters for the different elements, which would also reduce the variance). This is a trivial extension to our procedure, but the question is whether such an assumption has a sound theoretical basis, and leads to reliable error bounds.

Furthermore, stochastic embedding of the error due to undermodelling, as proposed by [54, 130], deserves consideration. In our opinion, the embedding should address the tail of the generalized expansion coefficients, rather than a separate FIR model as proposed in [54, 130]. However, one certainly should be careful with adopting this approach, as was discussed in Section 3.5.

Finally, it appears to be valuable to make a closer analysis (theoretically or on a simulation level) of the point at which the bounds given by \mathcal{K} and η , or M and ρ are allowed to cross the confidence interval, in order to capture uncertainty about the behaviour of the system as reflected in the experimental data. Such an analysis may indicate that higher rates of convergence are allowed and can be inferred from the confidence intervals on the estimated parameters.

Apart from the above, analytic expressions (or automatic procedures) are desirable for selecting the variables \mathcal{K} and η , or M and ρ in such a way that they are consistent with the resulting error bounds, especially in view of adaptive control applications. These variables can be chosen to be in agreement with the confidence interval for the error due to the noise (smaller than the total error bound), which appears to be less difficult (no iterative procedure necessary).

To achieve tighter and more reliable error bounds when the system is mildly nonlinear or time varying, a more appropriate characterization of the "noise" due to these effects with respect to a linear time invariant description of the system should be pursued. In this thesis we used a combination of a stochastic (averaging) and an unknown but bounded (worst case) description of these effects, as a result of the stochastic description of the noise and the deterministic description of the undermodelling error which we employed.

To achieve better models and tighter and more reliable error bounds when the system exhibits a substantial nonlinear or time varying behaviour, explicit modelling of the nonlinear or time varying behaviour appears to be necessary. In our opinion, this is one of the major challenges of system identification. Ignoring significant effects due to nonlinearities or time variations in the system will lead to conservative error bounds, since we do not know what kind of effect is ignored, and where (what frequency

region) it has its main influence. Therefore, more structure and freedom should be incorporated in the estimation problem, to be able to distinguish between these effects and to specifically account for these effects, instead of just considering them as one large source of errors. Again, if we are not prepared to model these effects (as far as reflected in the data), this will lead to unnecessary large error bounds. Moreover, the error bounds may become unreliable, as was demonstrated in Section 5.10.2 for a system with random time variations.

In this respect, black-box identification alone appears to be inadequate. Black-box identification is possible for linear time invariant systems since they are all governed by a set of linear difference or differential equations. Essentially, only the order of the model has to be chosen. This does not hold for nonlinear or time varying systems. The possibilities are endless, and without at least some prior knowledge of the physical laws governing the system we probably will not be able to identify a satisfactory model. That is, without such prior knowledge the model usually will not at all be parsimonious, leading to an unnecessary large variance of the estimated parameters. Additionally, the undermodelling error usually will be considerable since important aspects of the system probably cannot be captured adequately by the model. In contrast, in our experience grey-box modelling often yields quite satisfactory results, even for highly complex systems. Thus, physical modelling appears to be called for in order to be able to effectively tackle the problem of identifying nonlinear or time varying systems. Moreover, a physical model often is already present from the system design phase, or is desired anyway to provide additional physical insight in order to be able to improve the behaviour of the system. Clearly, the information arising from these models should be used when trying to identify the system.

In our opinion, a well balanced combination grey-box and black-box modelling appears to be called for. That is, one should use a physically motivated model structure to capture the major nonlinear or time varying effects in the data, extended with general linear, nonlinear and time varying elements to account for effects that are not represented adequately in the physically motivated model. Suitable candidates for these additional elements are for instance the usual linear black-box models [106], neural nets [23], and the first order expansion proposed by Hjalmarsson [83] to model the effects of time varying parameters in the system, respectively.

In this context, one should keep in mind that adding more freedom leads to less reliable models if this freedom is not necessary to model the data. Hence, care should be taken as to which effects actually are present in the data, to keep the model parsimonious but flexible enough.

Note that in this thesis, considering the case of (dominantly) linear time invariant systems, the use of physical knowledge (which may be approximative or inaccurate) has been enabled by the use of system based orthonormal basis functions.

Extension of the basic ideas developed in this thesis to nonlinear and time varying systems appears to be possible. That is, the approach of Chapter 5 (mixed averaging

– structural embedding of the modelling errors, and data segmentation to estimate the different error components) certainly appears to be promising in this more general context. The main adaptation that needs to be made is that the model of the system should be generalized and extended as proposed above, and that the parameters of this model should be estimated. Similar as in Chapter 5, the set of estimated parameter vectors over the different data segments can be transformed into a set of estimated output signals. This should allow the specification of an error bound on the output of the system, and thereby the specification of the uncertainty as operator uncertainty on the nonlinear time varying error components. The estimated parameters or output signals on the different data segments again may be asymptotically normally distributed. If not, it may be reasonable to assume them to be normally distributed since according to [87, page 146] and [90, chapter 31] the assumption of normality on the individual observations is less crucial when using an F distribution. Note that the approach of Chapter 5 has been shown to be robust to mild deviations from a linear time invariant undermodelling error and stochastic noise. Thus, if the significant nonlinear time varying effects are adequately captured in the estimated model, the procedure of Chapter 5 should be robust to small remaining nonlinear time varying effects or deterministic noise components.

Of course, it is also possible to apply the approach of Chapter 5 to the difference between the system and a nonlinear or time varying model of the system, but direct application to the system has the advantage of a far better signal to noise ratio.

Besides nonlinear and time varying effects, the problem of deterministic components in the noise also deserves attention. However, this problem appears to be less crucial. Hjalmarsson [83] has shown that using a "mixing" input signal, deterministic disturbances do not differ essentially from stochastic disturbances.

Finally, the ideas presented above also should allow for fairly extensive structured uncertainty models. Robust control (μ analysis and synthesis, see [38, 41, 40, 7]) can deal with structured real, complex, linear, nonlinear, time invariant, and time varying hard bounded uncertainty descriptions. Such extensive uncertainty models do not appear to be realistic, and may not be identifiable at all (nonunique). It definitely seems to be impossible to obtain such structured hard error bounds from experimental data alone. However, when using extensive physical knowledge about the structure of the system, much more should be possible than can be done at present.

Appendix A

Proofs for Chapter 4

A.1 Proof of Theorem 4.3.2

A.1.1 Properties of the N point DFT

To give the proof we have to start by taking a closer look at the properties of the N point DFT, and by dealing with some additional definitions and notation. The periodic continuation of a signal $x(t)$ is denoted by $x^R(t)$

$$x^R(t + kN) = x(t) \quad \text{for } k \in \mathbb{Z}, t \in T^N$$

The N point DFT and inverse DFT are defined in (4.6) and (4.7). A set of N complex orthogonal time domain elementary functions (complex sinewaves) now can be given as

$$\hat{x}_k(t) = \frac{1}{\sqrt{N}} X\left(\frac{2\pi k}{N}\right) e^{j\frac{2\pi k}{N}t} \quad k \in T^N \quad (\text{A.1})$$

There holds

$$\sum_{k=0}^{N-1} \hat{x}_i(t) \hat{x}_j(t) = 0 \quad \text{for } i \neq j$$
$$x(t) = \sum_{k=0}^{N-1} \hat{x}_k(t) \quad \text{for } t \in T^N$$

Note that the elementary functions are also defined outside T^N , and that outside T^N they are given by periodic continuation. Hence, for $t \notin T^N$ the inverse N point DFT gives a periodic continuation

$$x^R(t) = \sum_{k=0}^{N-1} \hat{x}_k(t) \quad \text{for } t \in \mathbb{Z} \quad (\text{A.2})$$

Consider the transformation matrix $W_N \in \mathbb{C}^{N \times N}$

$$W_N = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-j\frac{2\pi}{N}} & \cdots & e^{-j\frac{2\pi(N-1)}{N}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j(N-1)\frac{2\pi}{N}} & \cdots & e^{-j(N-1)\frac{2\pi(N-1)}{N}} \end{bmatrix} \quad (\text{A.3})$$

Note that W_N is an orthonormal matrix: $W_N W_N^* = W_N^* W_N = I$, where W_N^* denotes the complex conjugate transpose of the matrix W_N . The N point DFT can now be seen as a change of basis, where the new orthogonal set of basis functions is given by the columns of the matrix W_N . There holds

$$W_N \begin{bmatrix} \hat{x}_k(0) \\ \hat{x}_k(1) \\ \vdots \\ \hat{x}_k(N-1) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ X(\frac{2\pi k}{N}) \\ \vdots \\ 0 \end{bmatrix} \quad (\text{A.4})$$

where the nonzero element appears in the $(k+1)$ -th row. When a signal is used only over the time interval $T_{N_s}^N$ the DFT is defined according to (4.8), (4.9), and the elementary functions read

$$\begin{aligned} x(t) &= \sum_{k=0}^{N-1} \hat{x}_k^s(t) \quad \text{for } t \in T_{N_s}^N \\ \hat{x}_k^s(t) &= \frac{1}{\sqrt{N}} X^s(\frac{2\pi k}{N}) e^{j\frac{2\pi k}{N}(t-N_s)} \quad \text{for } t \in T_{N_s}^N \end{aligned} \quad (\text{A.5})$$

Finally, the past values of the input signal ($t < 0$) are sometimes denoted as $u^P(t)$ to stress that they are unknown.

A.1.2 Proof

The key observation is that we are able to decompose the input signal $u(t)$ over a measurement interval T^{N+N_s} in the basis W_N .

$$u(t) = \sum_{k=0}^{N-1} \hat{u}_k^s(t) \quad \text{for } t \in T^{N+N_s}$$

This can be done only for partly periodic input signals, see (A.2). For $t \in T^{N+N_s}$ the output now can be written as

$$y(t) = \sum_{i=0}^{\infty} g_o(i) u(t-i) + v(t)$$

$$= \sum_{i=0}^t g_o(i) \sum_{k=0}^{N-1} \hat{u}_k^s(t-i) + \sum_{i=t+1}^{\infty} g_o(i) u^P(t-i) + v(t) \quad (\text{A.6})$$

Note that for an elementary function there holds

$$\begin{aligned} g_o(t) * \hat{u}_k^s(t) &= \sum_{i=0}^{\infty} g_o(i) \hat{u}_k^s(t-i) \\ &= \frac{1}{\sqrt{N}} \sum_{i=0}^{\infty} g_o(i) U^s\left(\frac{2\pi k}{N}\right) e^{j\frac{2\pi k}{N}(t-N_s-i)} \\ &= \frac{1}{\sqrt{N}} U^s\left(\frac{2\pi k}{N}\right) e^{j\frac{2\pi k}{N}(t-N_s)} \sum_{i=0}^{\infty} g_o(i) e^{-j\frac{2\pi k}{N}i} \\ &= \frac{1}{\sqrt{N}} U^s\left(\frac{2\pi k}{N}\right) e^{j\frac{2\pi k}{N}(t-N_s)} G_o\left(\frac{2\pi k}{N}\right) \\ &= G_o\left(\frac{2\pi k}{N}\right) \hat{u}_k^s(t) \end{aligned} \quad (\text{A.7})$$

where $*$ denotes convolution. Hence

$$\begin{aligned} \sum_{i=0}^t g_o(i) \sum_{k=0}^{N-1} \hat{u}_k^s(t-i) &= \sum_{k=0}^{N-1} \sum_{i=0}^t g_o(i) \hat{u}_k^s(t-i) \\ &= \sum_{k=0}^{N-1} \left(\sum_{i=0}^{\infty} g_o(i) \hat{u}_k^s(t-i) - \sum_{i=t+1}^{\infty} g_o(i) \hat{u}_k^s(t-i) \right) \\ &= \sum_{k=0}^{N-1} G_o\left(\frac{2\pi k}{N}\right) \hat{u}_k^s(t) - \sum_{i=t+1}^{\infty} g_o(i) u^R(t-i) \end{aligned}$$

(A.6) now can be written as

$$y(t) = \sum_{k=0}^{N-1} G_o\left(\frac{2\pi k}{N}\right) \hat{u}_k^s(t) + \sum_{i=t+1}^{\infty} g_o(i) [u^P(t-i) - u^R(t-i)] + v(t) \quad (\text{A.8})$$

Define

$$e(t) = \sum_{i=t+1}^{\infty} g_o(i) [u^P(t-i) - u^R(t-i)] \quad (\text{A.9})$$

Writing down (A.8) for all $t \in T_{N_s}^N$, and using (A.5) and (A.9) results in

$$\sum_{k=0}^{N-1} \begin{bmatrix} \hat{y}_k^s(N_s) \\ \hat{y}_k^s(N_s+1) \\ \vdots \\ \hat{y}_k^s(N_s+N-1) \end{bmatrix} = \sum_{k=0}^{N-1} G_o\left(\frac{2\pi k}{N}\right) \begin{bmatrix} \hat{u}_k^s(N_s) \\ \hat{u}_k^s(N_s+1) \\ \vdots \\ \hat{u}_k^s(N_s+N-1) \end{bmatrix} +$$

$$+ \begin{bmatrix} e^{(N_s)} \\ e^{(N_s+1)} \\ \vdots \\ e^{(N_s+N-1)} \end{bmatrix} + \begin{bmatrix} v^{(N_s)} \\ v^{(N_s+1)} \\ \vdots \\ v^{(N_s+N-1)} \end{bmatrix} \quad (\text{A.10})$$

Premultiplying with the $(\ell + 1)$ -th row of W_N and using (A.4) gives

$$Y^s\left(\frac{2\pi\ell}{N}\right) = G_o\left(\frac{2\pi\ell}{N}\right)U^s\left(\frac{2\pi\ell}{N}\right) + E^s\left(\frac{2\pi\ell}{N}\right) + V^s\left(\frac{2\pi\ell}{N}\right)$$

By using the assumptions made on the impulse response and the input signal, an upper bound for $E^s\left(\frac{2\pi\ell}{N}\right)$ can be derived

$$\begin{aligned} |E^s\left(\frac{2\pi\ell}{N}\right)| &= \left| \frac{1}{\sqrt{N}} \sum_{t=N_s}^{N_s+N-1} e^{-j\frac{2\pi\ell}{N}(t-N_s)} \sum_{i=t+1}^{\infty} g_o(i)[u^P(t-i) - u^R(t-i)] \right| \\ &\leq \frac{1}{\sqrt{N}} (\bar{u}^p + \bar{u}) \sum_{t=N_s}^{N_s+N-1} \sum_{i=t+1}^{\infty} |g_o(i)| \\ &\leq \frac{1}{\sqrt{N}} (\bar{u}^p + \bar{u}) \frac{M\rho}{(\rho-1)^2} \rho^{-N_s} (1 - \rho^{-N}) \end{aligned} \quad (\text{A.11})$$

The result now follows by using the assumption made on the noise.

A.2 Proof of Theorem 4.3.5

For a MIMO system having m inputs and p outputs, the vector output ($p \times 1$) is given by

$$y(t) = \sum_{k=0}^{\infty} g_o(k)u(t-k) + v(t)$$

where $g_o(k)$ is the impulse response matrix ($p \times m$) of the system, $u(t)$ is the vector input ($m \times 1$) and $v(t)$ is the vector additive output noise ($p \times 1$). The input is a vector partly periodic signal (each element of the vector input $u(t)$ has the first N_s points equal to the last N_s points). The i -th output $y_{[i]}(t)$ now can be written as

$$\begin{aligned} y_{[i]}(t) &= \sum_{j=1}^m \sum_{k=0}^{\infty} g_{[ij]}(k)u_{[j]}(t-k) + v_{[i]}(t) \\ &= \sum_{j=1}^m \left(\sum_{\ell=0}^{N-1} G_{o[ij]}\left(\frac{2\pi\ell}{N}\right)\hat{u}_{[j]\ell}^s(t) + e_{[ij]}(t) \right) + v_{[i]}(t) \end{aligned}$$

with

$$e_{[ij]}(t) = \sum_{k=t+1}^{\infty} g_{[ij]}(k)[u_{[j]}^P(t-k) - u_{[j]}^R(t-k)]$$

It is now straightforward to show that

$$\begin{aligned} \begin{bmatrix} Y_{[i]}^s(0) \\ Y_{[i]}^s(\frac{2\pi}{N}) \\ \vdots \\ Y_{[i]}^s(\frac{2\pi(N-1)}{N}) \end{bmatrix} &= \sum_{j=1}^m \begin{bmatrix} G_{o[ij]}(0)U_{[j]}^s(0) \\ G_{o[ij]}(\frac{2\pi}{N})U_{[j]}^s(\frac{2\pi}{N}) \\ \vdots \\ G_{o[ij]}(\frac{2\pi(N-1)}{N})U_{[j]}^s(\frac{2\pi(N-1)}{N}) \end{bmatrix} \\ &+ \sum_{j=1}^m \begin{bmatrix} E_{[ij]}^s(0) \\ E_{[ij]}^s(\frac{2\pi}{N}) \\ \vdots \\ E_{[ij]}^s(\frac{2\pi(N-1)}{N}) \end{bmatrix} + \begin{bmatrix} V_{[i]}^s(0) \\ V_{[i]}^s(\frac{2\pi}{N}) \\ \vdots \\ V_{[i]}^s(\frac{2\pi(N-1)}{N}) \end{bmatrix} \end{aligned} \quad (\text{A.12})$$

If the input is in accordance with (4.11) then (A.12) becomes

$$Y_{[i]}^s(\frac{2\pi\ell}{N}) = G_{o[ij]}(\frac{2\pi\ell}{N})U_{[j]}^s(\frac{2\pi\ell}{N}) + \sum_{j=1}^m E_{[ij]}^s(\frac{2\pi\ell}{N}) + V_{[i]}^s(\frac{2\pi\ell}{N})$$

The result now follows from (A.11) and the assumption made on the noise.

A.3 Proof of Proposition 4.4.1

$$\begin{aligned} \left| \frac{d^k G_o(e^{j\omega})}{d\omega^k} \right| &= \left| \frac{d^k}{d\omega^k} \left(\sum_{\ell=0}^{\infty} g_o(\ell) e^{-j\omega\ell} \right) \right| = \left| \sum_{\ell=0}^{\infty} g_o(\ell) (-j\ell)^k e^{-j\omega\ell} \right| \\ &\leq \sum_{\ell=0}^{\infty} \ell^k |g_o(\ell)| \leq M \sum_{\ell=0}^{\infty} \ell^k \rho^{-k} \end{aligned}$$

The use of standard Taylor series concludes the proof.

A.4 Proof of Proposition 4.4.2

Write $F(e^{j\omega}) = G_o(e^{j\omega}) - G_{nom}(e^{j\omega})$. To prove the first inequality of Proposition 4.4.2 we now have to prove that for a scalar complex function $F(e^{j\omega})$ there holds

$$\left| \frac{d^k}{d\omega^k} |F(e^{j\omega})| \right| \leq \left| \frac{d^k F(e^{j\omega})}{d\omega^k} \right| \quad \text{for } k = \{1, 2\}$$

We will only give the proof for $k = 1$ (first derivative), the proof for $k = 2$ (second derivative) is completely analogous. Writing down the first derivative of $|F| = (F^* F)^{\frac{1}{2}}$ gives

$$\frac{d}{d\omega} |F| = \frac{1}{2|F|} \left(\frac{dF^*}{d\omega} F + F^* \frac{dF}{d\omega} \right)$$

Note that for any scalar complex function $F(e^{j\omega})$ there holds $d(F^*)/d\omega = (dF/d\omega)^*$ because

$$\lim_{\omega_2 \downarrow \omega_1} \frac{F^*(e^{j\omega_2}) - F^*(e^{j\omega_1})}{\omega_2 - \omega_1} = \lim_{\omega_2 \downarrow \omega_1} \left(\frac{F(e^{j\omega_2}) - F(e^{j\omega_1})}{\omega_2 - \omega_1} \right)^*$$

Therefore

$$\left| \frac{d}{d\omega} |F(e^{j\omega})| \right| \leq \frac{1}{2|F|} \left(|F| \left| \left(\frac{dF}{d\omega} \right)^* \right| + |F^*| \left| \frac{dF}{d\omega} \right| \right)$$

The the first inequality of Proposition 4.4.2 now follows by noting that for any complex scalar a there holds $|a^*| = |a|$. The second inequality of Proposition 4.4.2 is just the triangle inequality.

Appendix B

Proofs for Chapter 5

B.1 Proof of Lemma 5.3.1

Proof of inequality (5.20): Using (5.15) we have $|S_i| = |R_i|/|U_i|$. The upper bound on $|R_i(e^{j\omega_k})|$ directly follows from the results of Appendix C.1, see Theorem 6.3.3 and (6.9).

Proof of inequality (5.21): From (5.19) and (5.15) we have

$$|S| = \frac{1}{r|U_i|} \left| \sum_{i=1}^r R_i \right|$$

where we used the fact that U_i is independent of i because the input is periodic. Using (5.7), (5.5), (5.6), and the fact that $e^{-j2\pi\ell} = 1$ for $\ell \in \mathbb{Z}$, we find for $k \in T^{N_o}$ and $N = rN_o$

$$\begin{aligned} X^s(e^{j\frac{2\pi k}{N_o}}) &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x(t + N_s)e^{-j\frac{2\pi k}{N_o}t} = \frac{1}{\sqrt{N}} \sum_{i=1}^r \sum_{t=(i-1)N_o}^{iN_o-1} x(t + N_s)e^{-j\frac{2\pi k}{N_o}t} \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^r \sum_{\ell=0}^{N_o-1} x(\ell + (i-1)N_o + N_s)e^{-j\frac{2\pi k}{N_o}(\ell+(i-1)N_o)} \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^r \sum_{\ell=0}^{N_o-1} x_i(\ell)e^{-j\frac{2\pi k}{N_o}\ell} = \frac{1}{\sqrt{r}} \sum_{i=1}^r X_i(e^{j\frac{2\pi k}{N_o}}) \end{aligned} \quad (\text{B.1})$$

Note that it is not assumed that $x(t)$ is periodic. The upper bound on $|R^s(e^{j\omega_k})|$ now follows from the upper bound on $|E^s(e^{j\omega_k})|$ that is given in Appendix C.1 or Appendix A.1, see (6.9) or (A.11).

B.2 Proof of Lemma 5.3.3

Using (5.23) gives

$$\hat{\sigma}_r^2(\tilde{G}) - \hat{\sigma}_r^2(\hat{G}) = \frac{1}{r(r-1)} \sum_{i=1}^r \left(|\tilde{G} - \tilde{G}_i|^2 - |\hat{G} - \hat{G}_i|^2 \right) \quad (\text{B.2})$$

From the fact that $\tilde{G}_i = \hat{G}_i - S_i$ we find

$$\tilde{G} - \tilde{G}_i = \hat{G} - S - \hat{G}_i + S_i = (\hat{G} - \hat{G}_i) + (S_i - S) = A_i + B_i \quad (\text{B.3})$$

Thus

$$|\tilde{G} - \tilde{G}_i|^2 = |A_i + B_i|^2 = |A_i|^2 + A_i B_i^* + B_i A_i^* + |B_i|^2 \quad (\text{B.4})$$

Combining (B.2) and (B.4) gives

$$\hat{\sigma}_r^2(\tilde{G}) - \hat{\sigma}_r^2(\hat{G}) = \frac{1}{r(r-1)} \sum_{i=1}^r A_i B_i^* + B_i A_i^* + |B_i|^2$$

resulting in

$$|\hat{\sigma}_r^2(\tilde{G}) - \hat{\sigma}_r^2(\hat{G})| \leq \frac{1}{r(r-1)} \sum_{i=1}^r 2|A_i||B_i| + |B_i|^2$$

with

$$\begin{aligned} |A_i| &= |\hat{G} - \hat{G}_i| \\ |B_i| &= |S_i - S| = \left| S_i - \frac{1}{r} \sum_{k=1}^r S_k \right| = \left| \frac{r-1}{r} S_i - \frac{1}{r} \sum_{k=1}^{i-1} S_k - \frac{1}{r} \sum_{k=i+1}^r S_k \right| \\ &\leq \frac{r-1}{r} |S_i| + \frac{1}{r} \sum_{k=1}^{i-1} |S_k| + \frac{1}{r} \sum_{k=i+1}^r |S_k| = \frac{r-2}{r} |S_i| + \frac{1}{r} \sum_{k=1}^r |S_k| \end{aligned}$$

which completes the proof.

B.3 Proof of Lemma 5.3.4

We will use the following standard results from probability theory, see e.g. [147].

1. If $x_i \in \mathcal{N}(0, 1)$, $i = 1, \dots, k$, are independent random variables and $z = \sum_{i=1}^k x_i^2$, then $z \in \chi^2(k)$ with mean k and variance $2k$.
2. If $x \in \chi^2(a)$ and $y \in \chi^2(b)$ are independent random variables and $z = \frac{x/a}{y/b}$, then $z \in F(a, b)$.

3. Let $x_i, i = 1, \dots, k$, be a sequence of independent identically distributed real random variables, with $x_i \in \mathcal{N}(\mu, \sigma^2)$. Let $\hat{\mu} = \frac{1}{k} \sum_{i=1}^k x_i$ and let $\hat{\sigma}^2 = \frac{1}{k-1} \sum_{i=1}^k (\hat{\mu} - x_i)^2$. Then $\hat{\mu}$ and $\hat{\sigma}^2$ are independent and $\hat{\mu} \in \mathcal{N}(\mu, \frac{1}{r}\sigma^2)$, $\frac{(k-1)\hat{\sigma}^2}{\sigma^2} \in \chi^2(k-1)$.
4. Uncorrelated jointly normally distributed random variables are independent.
5. Linear combinations of jointly normally distributed random variables are again jointly normally distributed.

The convergence in distribution follows from the Continuous Mapping Theorem, see [145, page 46], which states the following. *Let h be a measurable mapping from \mathbb{R}^k into \mathbb{R}^s . Write C for the set of points in \mathbb{R}^k at which h is continuous. If a sequence x_n of random vectors taking values in \mathbb{R}^k converges in distribution to a random vector x for which $P[x \in C] = 1$, then $h(x_n)$ converges in distribution to $h(x)$.* We will first address the mapping f under consideration, and next we will derive the matching asymptotic distribution. Denote

$$D(e^{j\omega_k}) = \frac{|G_o(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))}$$

and

$$\mathbf{V}(e^{j\omega_k}) = [\operatorname{Re}\{V_1(e^{j\omega_k})\} \dots \operatorname{Re}\{V_r(e^{j\omega_k})\} \quad \operatorname{Im}\{V_1(e^{j\omega_k})\} \dots \operatorname{Im}\{V_r(e^{j\omega_k})\}]^T$$

Clearly we have that $D(e^{j\omega_k}) = f(\mathbf{V}(e^{j\omega_k}))$, where $f(\cdot)$ denotes some function. For a periodic input signal, i.e. $U_i(e^{j\omega_k})$ independent of i , it follows by direct calculation that

$$D = \frac{L(\operatorname{Re}\{V_i\}\operatorname{Re}\{V_m\}, \operatorname{Re}\{V_i\}\operatorname{Im}\{V_m\}, \operatorname{Im}\{V_i\}\operatorname{Im}\{V_m\})}{L(\operatorname{Re}\{V_i\}\operatorname{Re}\{V_m\}, \operatorname{Re}\{V_i\}\operatorname{Im}\{V_m\}, \operatorname{Im}\{V_i\}\operatorname{Im}\{V_m\})}$$

where $L(\cdot)$ denotes some linear combination of its arguments, with real coefficients. Hence $f(\mathbf{V}) : \mathbb{R}^{2r} \rightarrow \mathbb{R}$ is a continuous function for all $\hat{\sigma}_r^2(\tilde{G}) > 0$. Additionally, we have asymptotically in N_o that $P[\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k})) > 0] = 1$ since $V_i(e^{j\omega_k})$ and $V_j(e^{j\omega_k})$, $i \neq j$, are asymptotically independent, see Theorem 5.3.2. From the Continuous Mapping Theorem it now follows that if the vector $\mathbf{V}(e^{j\omega_k})$ converges in distribution to $\mathcal{V}(e^{j\omega_k})$ then $D(e^{j\omega_k}) = f(\mathbf{V}(e^{j\omega_k}))$ converges in distribution to $f(\mathcal{V}(e^{j\omega_k}))$. The distribution of $\mathcal{V}(e^{j\omega_k})$ is completely specified by the asymptotic distribution of \check{V}_{N_o} as given in Theorem 5.3.2. Hence for the derivation of the asymptotic distribution of $D(e^{j\omega_k})$ we can use $\check{V}_{N_o} \in \mathcal{N}(0, \Lambda)$ with Λ as given in Theorem 5.3.2.

Note that $\operatorname{Im}\{\tilde{G}(e^{j\omega_k})\} \equiv 0$ for $\omega_k \in \{0, \pi\}$. We will therefore split the remainder of the proof in two parts. First we will derive the distribution for $\omega_k = 0, \pi$, and next for $\omega_k \neq 0, \pi$. Denote

$$\begin{aligned} B(e^{j\omega_k}) &= \frac{(r-1)\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))}{\operatorname{var}[\tilde{G}(e^{j\omega_k})]} \\ A(e^{j\omega_k}) &= \frac{\tilde{G}(e^{j\omega_k}) - G_o(e^{j\omega_k})}{\sqrt{\operatorname{var}[\tilde{G}(e^{j\omega_k})]}} \end{aligned} \tag{B.5}$$

Case $\omega_k = 0, \pi$

From $\check{V}_{N_o} \in \mathcal{N}(0, \Lambda)$ with Λ as given in Theorem 5.3.2, it follows that the $V_i(e^{j\omega_k})$, $i = 1, \dots, r$, are normally distributed and that they are independent and identically distributed. Therefore $A(e^{j\omega_k}) \in \mathcal{N}(0, 1)$ and $B(e^{j\omega_k}) \in \chi^2(r-1)$. Because $\text{Im}\{A(e^{j\omega_k})\} \equiv 0$ for $\omega_k = 0, \pi$ we have that $A^2 \in \chi^2(1)$. Furthermore $\tilde{G}(e^{j\omega_k})$ and $\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))$ are independent, so that $D(e^{j\omega_k})$ a quotient of two independent χ^2 distributions, which is an F distribution

$$\frac{|A|^2/1}{B/(r-1)} = \frac{|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))} \in F(1, r-1)$$

Case $\omega_k \neq 0, \pi$

From $\check{V}_{N_o} \in \mathcal{N}(0, \Lambda)$ with Λ as given in Theorem 5.3.2, it follows that $A(e^{j\omega_k}) \in \mathcal{N}(0, 1)$. By direct calculation it also follows that $\text{Re}\{\tilde{G}(e^{j\omega_k})\}$ and $\text{Im}\{\tilde{G}(e^{j\omega_k})\}$ have the same variance, and hence that $\text{Re}\{A(e^{j\omega_k})\}$ and $\text{Im}\{A(e^{j\omega_k})\}$ have the same variance, for $\omega_k \neq 0, \pi$. Noting that for any complex random variable z there holds $\sigma^2(z) = \mathbb{E}[z^*z] = \sigma^2(\text{Re}\{z\}) + \sigma^2(\text{Im}\{z\})$ we find that $\text{Re}\{A(e^{j\omega_k})\} \in \mathcal{N}(0, \frac{1}{2})$ and $\text{Im}\{A(e^{j\omega_k})\} \in \mathcal{N}(0, \frac{1}{2})$. Hence $\text{Re}^2\{\sqrt{2}A\} \in \chi^2(1)$ and $\text{Im}^2\{\sqrt{2}A\} \in \chi^2(1)$. Finally, from $\check{V}_{N_o} \in \mathcal{N}(0, \Lambda)$ it follows by direct calculation that $\text{Re}\{A(e^{j\omega_k})\}$ and $\text{Im}\{A(e^{j\omega_k})\}$ are uncorrelated and jointly normally distributed. Hence $\text{Re}\{A(e^{j\omega_k})\}$ and $\text{Im}\{A(e^{j\omega_k})\}$ are independent. The above observations result in

$$2|A|^2 = 2\text{Re}^2\{A\} + 2\text{Im}^2\{A\} \in \chi^2(2)$$

Using a similar argument we find that

$$2B(e^{j\omega_k}) \in \chi^2(2(r-1)) \quad (\text{B.6})$$

Again $\tilde{G}(e^{j\omega_k})$ and $\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))$ are independent, so that

$$\frac{|A|^2/2}{2B/(2(r-1))} = \frac{|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))} \in F(2, 2(r-1))$$

which concludes the proof.

B.4 Proof of Theorem 5.3.5

From Lemma 5.3.4 we find

$$\frac{|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})|^2}{\hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))} \leq \alpha \quad \text{w.p.} \quad \begin{cases} F_\alpha(2, 2(r-1)) & \omega_k \neq 0, \pi \\ F_\alpha(1, r-1) & \omega_k = 0, \pi \end{cases}$$

Hence

$$|G_0(e^{j\omega_k}) - \tilde{G}(e^{j\omega_k})| \leq \sqrt{\alpha \hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))} \quad (\text{B.7})$$

Combining (B.7) and (5.18) gives

$$|G_0(e^{j\omega_k}) - \hat{G}(e^{j\omega_k})| \leq |S(e^{j\omega_k})| + \sqrt{\alpha \hat{\sigma}_r^2(\tilde{G}(e^{j\omega_k}))}$$

Using Lemma 5.3.3 and the triangle inequality

$$\hat{\sigma}_r^2(\tilde{G}) \leq \hat{\sigma}_r^2(\hat{G}) + |\hat{\sigma}_r^2(\tilde{G}) - \hat{\sigma}_r^2(\hat{G})|$$

completes the proof.

B.5 Proof of new elements in Theorem 5.4.2

We will only prove the inequalities for $z^p = \Psi Z$ and $s^p = \Psi S$. The asymptotic distribution follows from (5.45) using a similar argument as in Section B.3, while noting that the estimates $\hat{\theta}_i$ are real so that we do not have additional degrees of freedom due to independent real and imaginary parts. The inequality for $|z^p\langle m \rangle|$ follows as

$$\begin{aligned} z^p\langle m \rangle &= \sum_{\ell=1}^{N_p} \Psi\langle k, \ell \rangle Z(e^{j\zeta_\ell}) \\ &= \sum_{\ell=1}^{N_p} \Psi\langle k, \ell \rangle \sum_{k=\bar{n}_p}^{\infty} L_k \mathcal{V}_k(e^{j\zeta_\ell}) \\ &= \sum_{\ell=1}^{N_p} \Psi\langle k, \ell \rangle \left(\sum_{k=\bar{n}_p}^{\bar{n}_h-1} L_k \mathcal{V}_k(e^{j\zeta_\ell}) + \sum_{k=\bar{n}_h}^{\infty} L_k \mathcal{V}_k(e^{j\zeta_\ell}) \right) \\ &= \sum_{k=\bar{n}_p}^{\bar{n}_h-1} \sum_{\ell=1}^{N_p} \Psi\langle k, \ell \rangle L_k \mathcal{V}_k(e^{j\zeta_\ell}) + \sum_{\ell=1}^{N_p} \Psi\langle k, \ell \rangle \sum_{k=\bar{n}_h}^{\infty} L_k \mathcal{V}_k(e^{j\zeta_\ell}) \\ |z^p\langle m \rangle| &\leq \sum_{k=\bar{n}_p}^{\bar{n}_h-1} |L_k| \left| \sum_{\ell=1}^{N_p} \Psi\langle k, \ell \rangle \mathcal{V}_k(e^{j\zeta_\ell}) \right| + \sum_{\ell=1}^{N_p} |\Psi\langle k, \ell \rangle| \sum_{k=\bar{n}_h}^{\infty} |L_k| |\mathcal{V}_k(e^{j\zeta_\ell})| \\ &\leq \mathcal{K} \sum_{k=\bar{n}_p}^{\bar{n}_h-1} \eta^k \left| \sum_{\ell=1}^{N_p} \Psi\langle k, \ell \rangle \mathcal{V}_k(e^{j\zeta_\ell}) \right| + \mathcal{K} \sum_{\ell=1}^{N_p} |\Psi\langle k, \ell \rangle| |\mathcal{V}_0(e^{j\zeta_\ell})| \sum_{k=\bar{n}_h}^{\infty} \eta^k \\ &\leq \mathcal{K} \sum_{k=\bar{n}_p}^{\bar{n}_h-1} \eta^k \left| \sum_{\ell=1}^{N_p} \Psi\langle k, \ell \rangle \mathcal{V}_k(e^{j\zeta_\ell}) \right| + \mathcal{K} \frac{\eta^{\bar{n}_h}}{1-\eta} \sum_{\ell=1}^{N_p} |\Psi\langle k, \ell \rangle| |\mathcal{V}_0(e^{j\zeta_\ell})| \end{aligned}$$

where we used (E.5), the prior information that $|L_k| \leq \mathcal{K}\eta^k$, and the Taylor series expansion of $1/(1-x)$ with $x = \eta$. The inequality for $|s^p|$ follows directly from

$$s^p\langle m \rangle = \sum_{\ell=1}^{N_p} \Psi\langle m, \ell \rangle S(e^{j\zeta_\ell})$$

B.6 Proof of new elements in Theorem 5.4.4

We will only prove the inequalities for $|\Lambda(e^{j\omega}) - Z^f(e^{j\omega})|$ and $|\Gamma(e^{j\omega})|$. Everything else is similar to the arguments given in Section B.5, Section B.4 and Section B.3. The inequality for $|\Lambda(e^{j\omega}) - Z^f(e^{j\omega})|$ follows as

$$\begin{aligned}
\Lambda(e^{j\omega}) - Z^f(e^{j\omega}) &= \sum_{k=\bar{n}_f}^{\infty} L_k \mathcal{V}_k(e^{j\omega}) - \Upsilon Z \\
&= \sum_{k=\bar{n}_f}^{\infty} L_k \mathcal{V}_k(e^{j\omega}) - \sum_{\ell=1}^{N_p} \Upsilon\langle\ell\rangle \sum_{k=\bar{n}_p}^{\infty} L_k \mathcal{V}_k(e^{j\zeta_\ell}) \\
&= \sum_{k=\bar{n}_f}^{\infty} L_k \mathcal{V}_k(e^{j\omega}) - \sum_{k=\bar{n}_p}^{\infty} L_k \sum_{\ell=1}^{N_p} \Upsilon\langle\ell\rangle \mathcal{V}_k(e^{j\zeta_\ell}) \\
&= \sum_{k=\bar{n}_f}^{\bar{n}_p-1} L_k \mathcal{V}_k(e^{j\omega}) + \sum_{k=\bar{n}_p}^{\bar{n}_h-1} L_k \left(\mathcal{V}_k(e^{j\omega}) - \sum_{\ell=1}^{N_p} \Upsilon\langle\ell\rangle \mathcal{V}_k(e^{j\zeta_\ell}) \right) \\
&\quad + \sum_{k=\bar{n}_h}^{\infty} L_k \left(\mathcal{V}_k(e^{j\omega}) - \sum_{\ell=1}^{N_p} \Upsilon\langle\ell\rangle \mathcal{V}_k(e^{j\zeta_\ell}) \right) \\
|\Lambda(e^{j\omega}) - Z^f(e^{j\omega})| &\leq \mathcal{K} \sum_{k=\bar{n}_f}^{\bar{n}_p-1} \eta^k |\mathcal{V}_k(e^{j\omega})| + \mathcal{K} \sum_{k=\bar{n}_p}^{\bar{n}_h-1} \eta^k \left| \mathcal{V}_k(e^{j\omega}) - \sum_{\ell=1}^{N_p} \Upsilon\langle\ell\rangle \mathcal{V}_k(e^{j\zeta_\ell}) \right| \\
&\quad + \mathcal{K} \frac{\eta^{\bar{n}_h}}{1-\eta} \left(|\mathcal{V}_0(e^{j\omega})| + \sum_{\ell=1}^{N_p} |\Upsilon\langle\ell\rangle| |\mathcal{V}_0(e^{j\zeta_\ell})| \right)
\end{aligned}$$

The inequality for $|\Gamma(e^{j\omega})|$ follows as

$$\begin{aligned}
\Gamma(e^{j\omega}) &= D\mathcal{P}_D + \sum_{m=0}^{\bar{n}_f-1} \mathcal{P}_m L_m^T \\
|\Gamma(e^{j\omega})| &\leq |D| |\mathcal{P}_D| + \sum_{m=0}^{\bar{n}_f-1} |L_m| |\mathcal{P}_m|^T \\
&\leq M |\mathcal{P}_D| + \mathcal{K} \sum_{m=0}^{\bar{n}_f-1} |\mathcal{P}_m|^T \eta^m
\end{aligned} \tag{B.8}$$

B.7 Proof of new elements in Theorem 5.4.5

We will split up the proof in two parts, using

$$\hat{G}^f(e^{j\omega}) - G_o(e^{j\omega}) = \hat{G}^f(e^{j\omega}) - \tilde{G}^f(e^{j\omega}) + \tilde{G}^f(e^{j\omega}) - G_o(e^{j\omega})$$

From (5.51), (5.50), (5.19), (5.55), (5.54) it follows that the difference $\hat{G}^f(e^{j\omega}) - \tilde{G}^f(e^{j\omega})$ equals

$$\hat{G}^f(e^{j\omega}) - \tilde{G}^f(e^{j\omega}) = S^f(e^{j\omega}) + Z^f(e^{j\omega}) - \Lambda(e^{j\omega}) - \Gamma(e^{j\omega})$$

For the real and imaginary parts of the right hand side terms hard error bounds have been established in Theorem 5.4.4, thereby specifying a hard error bound for the left hand side term as a rectangular region in the complex plane.

The term $\tilde{G}^f(e^{j\omega}) - G_o(e^{j\omega})$ can be bounded with a soft error bound as follows. From (5.55), (5.54) it follows that

$$\tilde{G}^f(e^{j\omega}) - G_o(e^{j\omega}) = \Upsilon(e^{j\omega}) \frac{1}{r} \sum_{i=1}^r F_i \quad (\text{B.9})$$

where F_i is given in (5.44), (5.40). From (B.9) and Theorem 5.3.2 it follows that for $\hat{\Sigma}_r^2(\tilde{G}^f(e^{j\omega}))$ as defined in (5.59) there holds asymptotically in N_o , see [87, page 171],

$$\begin{aligned} & \begin{bmatrix} \text{Re}\{\tilde{G}^f(e^{j\omega}) - G_o(e^{j\omega})\} \\ \text{Im}\{\tilde{G}^f(e^{j\omega}) - G_o(e^{j\omega})\} \end{bmatrix}^T \left[\hat{\Sigma}_r^2(\tilde{G}^f(e^{j\omega})) \right]^{-1} \begin{bmatrix} \text{Re}\{\tilde{G}^f(e^{j\omega}) - G_o(e^{j\omega})\} \\ \text{Im}\{\tilde{G}^f(e^{j\omega}) - G_o(e^{j\omega})\} \end{bmatrix} \\ & \leq \frac{2(r-1)}{r-2} \alpha \quad \text{w.p. } F_\alpha(2, r-2) \end{aligned} \quad (\text{B.10})$$

specifying an ellipsoidal confidence region in the complex plane. However, in (B.10) the matrix $\hat{\Sigma}_r^2(\tilde{G}^f(e^{j\omega}))$ is unknown since $\tilde{G}^f(e^{j\omega})$ is unknown, so that we have to replace this matrix with a known matrix to be able to actually calculate the confidence region. Now note that both $\hat{\Sigma}_r^2(\hat{G}^f(e^{j\omega}))$ and $\hat{\Sigma}_r^2(\tilde{G}^f(e^{j\omega}))$ are positive semi-definite by construction, see (5.59). We thus have to find a known matrix $Q \geq 0$ such that the set $\{x \in \mathbb{R}^2 \mid x^T (\hat{\Sigma}_r^2(\hat{G}^f))^{-1} x \leq c\}$ is contained in the set $\{x \in \mathbb{R}^2 \mid x^T Q^{-1} x \leq c\}$. Hence we require that $x^T Q^{-1} x \leq x^T (\hat{\Sigma}_r^2(\hat{G}^f))^{-1} x$. For ease of notation we define $\tilde{\Sigma}(\omega) = \hat{\Sigma}_r^2(\tilde{G}^f(e^{j\omega}))$ and $\hat{\Sigma}(\omega) = \hat{\Sigma}_r^2(\hat{G}^f(e^{j\omega}))$. We will now set out to prove that for

$$Q(\omega) = \begin{bmatrix} \hat{\Sigma}_{11}(\omega) + 2\mathcal{S}^f(\omega) & \hat{\Sigma}_{12}(\omega) \\ \hat{\Sigma}_{21}(\omega) & \hat{\Sigma}_{22}(\omega) + 2\mathcal{S}^f(\omega) \end{bmatrix}$$

there holds

$$x^T Q^{-1}(\omega) x \leq x^T \tilde{\Sigma}^{-1}(\omega) x$$

for all $x \in \mathbb{R}^2$ and $\omega \in [0, 2\pi)$. Firstly, it follows directly from Lemma 5.4.3 that for each entry of $\hat{\Sigma}(\omega) - \tilde{\Sigma}(\omega)$ there holds

$$|\hat{\Sigma}_{k\ell}(\omega) - \tilde{\Sigma}_{k\ell}(\omega)| \leq \mathcal{S}^f(\omega) \quad (\text{B.11})$$

for all $\omega \in [0, 2\pi)$ and $k, \ell = 1, 2$, where $\mathcal{S}^f(\omega)$ is given by (5.57). Secondly, the following equivalence relations are standard results from linear algebra

$$x^T Q^{-1} x \leq x^T \tilde{\Sigma}^{-1} x \Leftrightarrow \tilde{\Sigma}^{-1} - Q^{-1} \geq 0 \Leftrightarrow Q - \tilde{\Sigma} \geq 0$$

Hence we have to prove that $x^T(Q - \tilde{\Sigma})x \geq 0$ for all $x \in \mathbb{R}^2$. It follows that

$$\begin{aligned} & [x_1 \ x_2](Q - \tilde{\Sigma}) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= x_1^2(\hat{\Sigma}_{11} - \tilde{\Sigma}_{11} + 2\mathcal{S}^f) + x_2^2(\hat{\Sigma}_{22} - \tilde{\Sigma}_{22} + 2\mathcal{S}^f) + 2x_1x_2(\hat{\Sigma}_{12} - \tilde{\Sigma}_{12}) \\ &\geq x_1^2(\hat{\Sigma}_{11} - \tilde{\Sigma}_{11} + \mathcal{S}^f) + x_2^2(\hat{\Sigma}_{22} - \tilde{\Sigma}_{22} + \mathcal{S}^f) + (x_1^2 + x_2^2)\mathcal{S}^f - 2|x_1x_2| |\hat{\Sigma}_{12} - \tilde{\Sigma}_{12}| \\ &\geq (x_1^2 + x_2^2 - 2|x_1x_2|)\mathcal{S}^f \end{aligned}$$

where we used (B.11). Now note that for all $x \in \mathbb{R}^2$

$$\begin{aligned} (x_1 + x_2)^2 \geq 0 &\Leftrightarrow x_1^2 + x_2^2 + 2x_1x_2 \geq 0 \\ (x_1 - x_2)^2 \geq 0 &\Leftrightarrow x_1^2 + x_2^2 - 2x_1x_2 \geq 0 \end{aligned}$$

Hence

$$x_1^2 + x_2^2 - 2|x_1x_2| \geq 0$$

which completes the proof.

B.8 Proof of new elements in Theorem 5.4.8

There holds

$$\begin{aligned} \hat{y}(\tau) - y_o(\tau) &= (Q(\tau)\theta_o - y_o(\tau)) + (Q(\tau) \mathcal{W}_y (\hat{\theta} - \theta_o)) + (Q(\tau) (\mathcal{W}_y - I) \theta_o) \\ &= E_1(\tau) + E_2(\tau) + E_3(\tau) \end{aligned}$$

where $E_1(\tau)$ is the error due to undermodelling and unknown past inputs $u_a(\tau)$, $E_2(\tau)$ is due to the error in the estimated parameters, i.e. errors due to undermodelling, past inputs $u(t)$ and noise, and $E_3(\tau)$ is the error due to weighting. For $E_1(\tau)$ we have

$$\begin{aligned} E_1(\tau) &= Du_a(\tau) + \sum_{m=0}^{q-1} u_a(\tau-1-m) \sum_{k=0}^{\bar{n}_y-1} L_k \phi_k(m) \\ &\quad - Du_a(\tau) - \sum_{m=0}^{\infty} u_a(\tau-1-m) \sum_{k=0}^{\infty} L_k \phi_k(m) \\ &= \sum_{m=0}^{q-1} \left(\sum_{k=0}^{\infty} - \sum_{k=\bar{n}_y}^{\infty} \right) u_a(\tau-1-m) L_k \phi_k(m) \\ &\quad - \left(\sum_{m=0}^{q-1} + \sum_{m=q}^{\infty} \right) \sum_{k=0}^{\infty} u_a(\tau-1-m) L_k \phi_k(m) \\ &= - \sum_{m=0}^{q-1} u_a(\tau-1-m) \sum_{k=\bar{n}_y}^{\infty} L_k \phi_k(m) - \sum_{m=q}^{\infty} u_a(\tau-1-m) \sum_{k=0}^{\infty} L_k \phi_k(m) \end{aligned}$$

$$\begin{aligned}
&= - \sum_{m=1}^q u_a(\tau - m) \sum_{k=\bar{n}_y}^{\infty} L_k \phi_k(m-1) - \sum_{m=q+1}^{\infty} u_a(\tau - m) \sum_{k=0}^{\infty} L_k \phi_k(m-1) \\
&= - \sum_{m=1}^q u_a(\tau - m) \sum_{k=\bar{n}_y}^{\infty} L_k \phi_k(m-1) - \sum_{m=q+1}^{\tau} u_a(\tau - m) g_o(m) \\
&\quad - \sum_{m=\tau+1}^{\infty} u_a(\tau - m) g_o(m)
\end{aligned}$$

$$|E_1(\tau)| \leq |E_{11}(\tau)| + M \sum_{m=q+1}^{\tau} |u_a(\tau - m)| \rho^{-m} + \bar{u}_a^p M \frac{\rho^{-\tau}}{\rho - 1}$$

where we used that the impulse response of the system G_o is given by $g_o(m) = \sum_{k=0}^{\infty} L_k \phi_k(m-1)$ for $m > 0$, see (5.69).

$$\begin{aligned}
E_{11}(\tau) &= \sum_{m=1}^q u_a(\tau - m) \sum_{k=\bar{n}_y}^{\infty} L_k \phi_k(m-1) \\
&= \sum_{k=\bar{n}_y}^{\bar{n}_x-1} L_k \sum_{m=1}^q u_a(\tau - m) \phi_k(m-1) + \sum_{m=1}^q u_a(\tau - m) \sum_{k=\bar{n}_x}^{\infty} L_k \phi_k(m-1)
\end{aligned}$$

$$|E_{11}(\tau)| \leq \mathcal{K} \sum_{k=\bar{n}_y}^{\bar{n}_x-1} \eta^k \left| \sum_{m=1}^q u_a(\tau - m) \phi_k(m-1) \right| + |E_{12}(\tau)|$$

Using the Cauchy-Schwarz inequality, we find for $E_{12}(\tau)$

$$|E_{12}(\tau)| \leq \sum_{m=1}^q |u_a(\tau - m)| \sqrt{\sum_{k=\bar{n}_x}^{\infty} \|L_k\|_2^2} \sqrt{\sum_{k=\bar{n}_x}^{\infty} \|\phi_k(m-1)\|_2^2}$$

where

$$\sqrt{\sum_{k=\bar{n}_x}^{\infty} \|L_k\|_2^2} \leq \sqrt{\sum_{k=\bar{n}_x}^{\infty} \|\mathcal{K}\|_2^2 \eta^{2k}} \leq \|\mathcal{K}\|_2 \sqrt{\frac{\eta^{2\bar{n}_x}}{1 - \eta^2}}$$

The fact that the infinite matrix $\Phi(\mathbf{G}_b)$, see (E.11), is orthonormal, yields

$$\begin{aligned}
\sqrt{\sum_{k=0}^{\infty} \|\phi_k(m-1)\|_2^2} &= 1 \\
\sum_{k=\bar{n}_x}^{\infty} \|\phi_k(m-1)\|_2^2 &= 1 - \sum_{k=0}^{\bar{n}_x-1} \|\phi_k(m-1)\|_2^2
\end{aligned}$$

Hence $|E_{12}(\tau)|$ can be bounded as

$$|E_{12}(\tau)| \leq \|\mathcal{K}\|_2 \eta^{\bar{n}_x} \sqrt{\frac{1}{1 - \eta^2}} \sum_{m=1}^q |u_a(\tau - m)| \sqrt{1 - \sum_{k=0}^{\bar{n}_x-1} \|\phi_k(m-1)\|_2^2}$$

For $E_2(\tau)$ we have, see (5.41)

$$E_2(\tau) = Q(\tau) \mathcal{W}_y \Psi \left(Z + S + \frac{1}{r} \sum_{i=1}^r F_i \right) = E_{21}(\tau) + E_{22}(\tau) + E_{23}(\tau)$$

For $E_{23}(\tau)$ a probabilistic error bound is used, for $E_{21}(\tau)$ and $E_{22}(\tau)$ we have

$$E_{21}(\tau) = \Upsilon^y Z = \sum_{\ell=1}^{N_p} \Upsilon^y \langle \ell \rangle Z(e^{j\zeta_\ell})$$

which is similar to the expression for z^p , (B.8), and

$$E_{22}(\tau) = \Upsilon^y S = \sum_{\ell=1}^{N_p} \Upsilon^y \langle \ell \rangle S(e^{j\zeta_\ell})$$

Finally

$$E_3(\tau) = D\mathcal{P}_D^y + \sum_{m=0}^{\bar{n}_y-1} \mathcal{P}_m^y L_m^T$$

which is similar to the expression for $\Gamma(e^{j\omega})$, (B.8).

Appendix C

Proofs for Chapter 6

C.1 Proof of Theorem 6.3.3

We will employ the usual convention that

$$\sum_a^b(\cdot) = 0 \quad \text{if} \quad b < a$$

Using (6.6),(6.7) we find

$$\begin{aligned} Z^s(e^{j\omega_k}) &= \sum_{t=0}^{N-1} z(t + N_s)e^{-j\omega_k t} \\ &= \sum_{t=0}^{N-1} e^{-j\omega_k t} \sum_{p=0}^{\infty} g_o(p)x(t + N_s - p) \\ &= \sum_{t=0}^{N-1} e^{-j\omega_k t} \sum_{p=0}^{\infty} g_o(p) \sum_{m=0}^{\infty} f(m)r(t + N_s - p - m) \\ &= \sum_{t=0}^{N-1} e^{-j\omega_k t} \sum_{p=0}^{\infty} g_o(p)e^{-j\omega_k p} e^{j\omega_k p} \sum_{m=0}^{\infty} f(m)r(t + N_s - p - m)e^{-j\omega_k m} e^{j\omega_k m} \\ &= \sum_{p=0}^{\infty} g_o(p)e^{-j\omega_k p} \sum_{m=0}^{\infty} f(m)e^{-j\omega_k m} \sum_{t=0}^{N-1} r(t + N_s - p - m)e^{-j\omega_k t} e^{j\omega_k p} e^{j\omega_k m} \\ [\ell &= t - p - m] \\ &= \sum_{p=0}^{\infty} g_o(p)e^{-j\omega_k p} \sum_{m=0}^{\infty} f(m)e^{-j\omega_k m} \sum_{\ell=-p-m}^{N-p-m-1} r(\ell + N_s)e^{-j\omega_k \ell} \end{aligned}$$

Because $p \geq 0$ and $m \geq 0$, so that $-p - m \leq 0$, we can write

$$\sum_{\ell=-p-m}^{N-p-m-1} = \sum_{\ell=-p-m}^{-1} + \sum_{\ell=0}^{N-1} - \sum_{\ell=N-p-m}^{N-1}$$

Using $\omega_k = \frac{2\pi k}{N_o}$ and $N = iN_o$, $i, k \in \mathbb{N}$, we find

$$e^{-j\frac{2\pi k}{N_o}(\ell+N)} = e^{-j\frac{2\pi k}{N_o}\ell} e^{-j2\pi ki} = e^{-j\frac{2\pi k}{N_o}\ell} \quad (\text{C.1})$$

Hence we can write

$$\begin{aligned} & \sum_{\ell=-p-m}^{-1} r(\ell + N_s) e^{-j\omega_k \ell} - \sum_{\ell=N-p-m}^{N-1} r(\ell + N_s) e^{-j\omega_k \ell} \\ &= \sum_{\ell=-p-m}^{-1} (r(\ell + N_s) - r(\ell + N_s + N)) e^{-j\omega_k \ell} \end{aligned}$$

Therefore

$$\begin{aligned} Z^s(e^{j\omega_k}) &= G_o(e^{j\omega_k})F(e^{j\omega_k})R^s(e^{j\omega_k}) + E^s(e^{j\omega_k}) \\ &= G_o(e^{j\omega_k})X_p^s(e^{j\omega_k}) + E^s(e^{j\omega_k}) \end{aligned}$$

where

$$E^s(e^{j\omega_k}) = \sum_{p=0}^{\infty} g_o(p) e^{-j\omega_k p} \sum_{m=0}^{\infty} f(m) e^{-j\omega_k m} \sum_{\ell=-p-m}^{-1} (r(\ell + N_s) - r(\ell + N_s + N)) e^{-j\omega_k \ell}$$

Using the fact that $f(m) = 0$ for $m < 0$ and that $-p - m \leq -N_s$ for $m \geq N_s - p$ we find

$$\begin{aligned} E^s(e^{j\omega_k}) &= \sum_{p=0}^{\infty} g_o(p) e^{-j\omega_k p} \sum_{m=0}^{N_s-p} f(m) e^{-j\omega_k m} \sum_{\ell=-p-m}^{-1} (r(\ell + N_s) - r(\ell + N_s + N)) e^{-j\omega_k \ell} \\ &+ \sum_{p=0}^{\infty} g_o(p) e^{-j\omega_k p} \sum_{m=N_s+1-p}^{\infty} f(m) e^{-j\omega_k m} \sum_{\ell=-p-m}^{-N_s-1} (r(\ell + N_s) - r(\ell + N_s + N)) e^{-j\omega_k \ell} \\ &+ \sum_{p=0}^{\infty} g_o(p) e^{-j\omega_k p} \sum_{m=N_s+1-p}^{\infty} f(m) e^{-j\omega_k m} \sum_{\ell=-N_s}^{-1} (r(\ell + N_s) - r(\ell + N_s + N)) e^{-j\omega_k \ell} \end{aligned}$$

Using the fact that $r(t)$ periodic for $t \geq 0$, i.e. $r(t + N_s + iN_o) = r(t + N_s)$ for $t \in T^{N_o}$, $i = 0, 1, \dots, \frac{N}{N_o} - 1$, we find that the first and third left hand side terms are zero, yielding

$$E^s(e^{j\omega_k}) = \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+1-p}^{\infty} f(m) \sum_{\ell=-p-m}^{-N_s-1} (r(\ell + N_s) - r(\ell + N_s + N)) e^{-j\omega_k(p+m+\ell)}$$

$$\begin{aligned}
&= \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+1-p}^{\infty} f(m) \sum_{\ell=-p-m}^{-N_s-1} r(\ell + N_s) e^{-j\omega_k(p+m+\ell)} \\
&\quad - \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+1-p}^{N_s+N-p} f(m) \sum_{\ell=-p-m}^{-N_s-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell)} \\
&\quad - \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+N+1-p}^{\infty} f(m) \sum_{\ell=-p-m}^{-N_s-N-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell)} \\
&\quad - \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+N+1-p}^{\infty} f(m) \sum_{\ell=-N_s-N}^{-N_s-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell)} \\
&= E_1^s(e^{j\omega_k}) - E_2^s(e^{j\omega_k})
\end{aligned}$$

where $E_1^s(e^{j\omega_k})$ contains the unknown past of the external signal $r(t)$, and $E_2^s(e^{j\omega_k})$ the known periodic part of $r(t)$

$$\begin{aligned}
E_1^s(e^{j\omega_k}) &= \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+1-p}^{\infty} f(m) \sum_{\ell=-p-m}^{-N_s-1} r(\ell + N_s) e^{-j\omega_k(p+m+\ell)} \\
&\quad - \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+N+1-p}^{\infty} f(m) \sum_{\ell=-p-m}^{-N_s-N-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell)} \\
E_2^s(e^{j\omega_k}) &= \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+1-p}^{N_s+N-p} f(m) \sum_{\ell=-p-m}^{-N_s-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell)} \\
&\quad + \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+N+1-p}^{\infty} f(m) \sum_{\ell=-N_s-N}^{-N_s-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell)}
\end{aligned}$$

Using $e^{-j\omega_k N} = 1$, see (C.1), the first error term can be written as

$$\begin{aligned}
E_1^s(e^{j\omega_k}) &= \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+1-p}^{\infty} f(m) \sum_{\ell=-p-m-N}^{-N_s-N-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell+N)} \\
&\quad - \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+N+1-p}^{\infty} f(m) \sum_{\ell=-p-m}^{-N_s-N-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell)} \\
&= \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+1-p}^{N_s+N-p} f(m) \sum_{\ell=-p-m}^{-N_s-N-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell)} \\
&\quad + \sum_{p=0}^{\infty} g_o(p) \sum_{m=N_s+1-p}^{\infty} f(m) \sum_{\ell=-p-m-N}^{-p-m-1} r(\ell + N_s + N) e^{-j\omega_k(p+m+\ell)}
\end{aligned}$$

Taking the magnitude gives

$$\begin{aligned}
|E_1^s(e^{j\omega_k})| &\leq \bar{r}^p \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+1-p}^{N_s+N-p} |f(m)|(p+m-N_s-N) \\
&\quad + \bar{r}^p \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+1-p}^{\infty} |f(m)|N \\
&= \bar{r}^p \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+1-p}^{N_s+N-p} |f(m)|(p+m-N_s) \\
&\quad + \bar{r}^p \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+N+1-p}^{\infty} |f(m)|N \\
&= \bar{r}^p \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+1-p}^{\infty} |f(m)|(p+m-N_s) \\
&\quad - \bar{r}^p \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+N+1-p}^{\infty} |f(m)|(p+m-N_s-N) \quad (C.2) \\
&= \bar{r}^p (E_{N_s} - E_{N_s+N})
\end{aligned}$$

Using the fact that $f(m) = 0$ for $m < 0$, we can write E_{N_s} as

$$\begin{aligned}
E_{N_s} &= \sum_{p=0}^{N_s} |g_o(p)| \sum_{m=N_s+1-p}^{\infty} |f(m)|(p+m-N_s) \\
&\quad + \sum_{p=N_s+1}^{\infty} |g_o(p)| \sum_{m=0}^{\infty} |f(m)|(p+m-N_s) \\
&\leq MB \sum_{p=0}^{N_s} \rho^{-p} \sum_{m=N_s+1-p}^{\infty} \mu^{-m}(p+m-N_s) \\
&\quad + MB \sum_{p=N_s+1}^{\infty} \rho^{-p} \sum_{m=0}^{\infty} \mu^{-m}(p+m-N_s) \\
&= MB \sum_{p=0}^{N_s} \rho^{-p} \left(p - N_s + N_s - p + 1 + \frac{1}{\mu - 1} \right) \frac{\mu^{p-N_s}}{\mu - 1} \\
&\quad + MB \sum_{p=N_s+1}^{\infty} \rho^{-p} \left((p - N_s) \frac{\mu}{\mu - 1} + \frac{\mu}{(\mu - 1)^2} \right) \\
&= \frac{MB\mu}{\mu - 1} \left[\frac{1}{\mu - 1} \sum_{p=0}^{N_s} \rho^{-p} \mu^{p-N_s} + \sum_{p=N_s+1}^{\infty} \rho^{-p} \left(p - N_s + \frac{1}{\mu - 1} \right) \right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{MB\mu}{\mu-1} \left[\frac{1}{\mu-1} \left(\mathfrak{U}_{N_s} + \frac{\rho^{-N_s}}{\rho-1} \right) + \rho^{-N_s} \frac{\rho}{(\rho-1)^2} \right] \\
&= \frac{MB\mu}{\mu-1} \left[\frac{\mathfrak{U}_{N_s}}{\mu-1} + \frac{\rho^{-N_s}}{\rho-1} \left(\frac{1}{\mu-1} + \frac{\rho}{\rho-1} \right) \right]
\end{aligned}$$

To find a more elegant expression for

$$\mathfrak{U}_{N_s} = \sum_{p=0}^{N_s} \rho^{-p} \mu^{p-N_s}$$

we have to distinguish three cases: $\rho = \mu$, $\rho > \mu$ and $\rho < \mu$. For $\rho = \mu$ we find

$$\mathfrak{U}_{N_s} = \sum_{p=0}^{N_s} \mu^{-N_s} = (N_s + 1)\mu^{-N_s}$$

For $\rho > \mu$ we find

$$\mathfrak{U}_{N_s} = \sum_{p=0}^{N_s} \left(\frac{\rho}{\mu} \right)^{-p} \mu^{-N_s} = \frac{\rho\mu}{\rho-\mu} (\mu^{-N_s-1} - \rho^{-N_s-1})$$

For $\rho < \mu$ we find

$$\mathfrak{U}_{N_s} = \sum_{p=0}^{N_s} \left(\frac{\mu}{\rho} \right)^{p-N_s} \rho^{-N_s} = \sum_{p=0}^{N_s} \left(\frac{\mu}{\rho} \right)^{-p} \rho^{-N_s} = \frac{\mu\rho}{\mu-\rho} (\rho^{-N_s-1} - \mu^{-N_s-1})$$

For the second error term $E_2^s(e^{j\omega_k})$ we have

$$\begin{aligned}
|E_2^s(e^{j\omega_k})| &\leq \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+1-p}^{N_s+N-p} |f(m)| \sum_{\ell=-p-m}^{-N_s-1} |r(\ell + N_s + N)| \\
&\quad + \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+N+1-p}^{\infty} |f(m)| \sum_{\ell=-N_s-N}^{-N_s-1} |r(\ell + N_s + N)| \\
&= \bar{r} \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+1-p}^{N_s+N-p} |f(m)|(p+m-N_s) \\
&\quad + \bar{r} \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+N+1-p}^{\infty} |f(m)|N \\
&= \bar{r} \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+1-p}^{\infty} |f(m)|(p+m-N_s) \\
&\quad - \bar{r} \sum_{p=0}^{\infty} |g_o(p)| \sum_{m=N_s+N+1-p}^{\infty} |f(m)|(p+m-N_s-N) \\
&= \bar{r}(E_{N_s} - E_{N_s+N})
\end{aligned}$$

where the last equality follows from (C.2).

Finally, the above yields

$$\begin{aligned} |E^s(e^{j\omega_k})| &\leq |E_1^s(e^{j\omega_k})| + |E_2^s(e^{j\omega_k})| \\ &\leq (\bar{r}^p + \bar{r})(E_{N_s} - E_{N_s+N}) \\ &\leq (\bar{r}^p + \bar{r}) \frac{MB\mu}{\mu-1} \left[\frac{\mathcal{U}^s}{\mu-1} + (1 - \rho^{-N}) \frac{\rho^{-N_s}}{\rho-1} \left(\frac{1}{\mu-1} + \frac{\rho}{\rho-1} \right) \right] \end{aligned}$$

where

$$\mathcal{U}^s = \begin{cases} (N_s + 1)\mu^{-N_s} - (N_s + N + 1)\mu^{-N_s-N} & \text{for } \rho = \mu \\ \frac{\rho\mu}{\rho-\mu} [(1 - \mu^{-N})\mu^{-N_s-1} - (1 - \rho^{-N})\rho^{-N_s-1}] & \text{for } \rho > \mu \\ \frac{\mu\rho}{\mu-\rho} [(1 - \rho^{-N})\rho^{-N_s-1} - (1 - \mu^{-N})\mu^{-N_s-1}] & \text{for } \rho < \mu \end{cases}$$

which completes the proof.

Appendix D

Proofs for Chapter 7

D.1 Proof of Lemma 7.3.2

Firstly, see [106, equation (6.31)], there exists a constant $c_1 \in \mathbb{R}$ such that

$$|\mathbb{E}[|V^s(e^{j\omega_k})|^2] - \Phi_v(\omega_k)]| \leq \frac{c_1}{N} \quad \omega_k \in \Omega_{N_o}^u$$

and

$$|\mathbb{E}[V^s(e^{j\omega_k})(V^s(e^{j\omega_\ell}))^*]| \leq \frac{c_1}{N} \quad \omega_k \neq \omega_\ell \quad \omega_k, \omega_\ell \in \Omega_{N_o}^u$$

Secondly, it follows directly from (B.1) that

$$U^s(e^{j\omega_k})(U^s(e^{j\omega_\ell}))^* = \frac{N}{N_o} U_o(e^{j\omega_k})U_o^*(e^{j\omega_\ell})$$

The first statement of the theorem now follows directly.

The second statement of the theorem now follows by noting that the induced two norm of a matrix A is less than or equal to the Frobenius norm, see e.g. [53, page 57],

$$\|A\|_2 \leq \sqrt{\sum_i \sum_k |A_{ik}|^2} \quad (\text{D.1})$$

D.2 Proof of Theorem 7.3.3

The theorem is a direct consequence of (7.32), Lemma 7.3.2, the triangle inequality, and the submultiplicative property of matrix norms

$$\|AB\|_2 \leq \|A\|_2 \|B\|_2 \quad (\text{D.2})$$

where A and B are matrices. Thus, if $\|B - C\|_2 \rightarrow 0$ and $\|A\|_2$ is uniformly bounded, then $\|A(B - C)\|_2 \rightarrow 0$, so that the triangle inequality yields $\|AB\|_2 \rightarrow \|AC\|_2$.

D.3 Proof of Lemma 7.3.4

From (7.26) it follows that

$$\begin{aligned} \Phi^* \Phi &= \begin{bmatrix} 1 & \cdots & 1 \\ \mathcal{V}_0(e^{j\omega_1}) & \cdots & \mathcal{V}_0(e^{j\omega_p}) \\ \vdots & \ddots & \vdots \\ \mathcal{V}_{n-1}(e^{j\omega_1}) & \cdots & \mathcal{V}_{n-1}(e^{j\omega_p}) \end{bmatrix} \begin{bmatrix} 1 & \mathcal{V}_0^*(e^{j\omega_1}) & \cdots & \mathcal{V}_{n-1}^*(e^{j\omega_1}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \mathcal{V}_0^*(e^{j\omega_p}) & \cdots & \mathcal{V}_{n-1}^*(e^{j\omega_p}) \end{bmatrix} \\ &= \sum_{\ell=1}^p \begin{bmatrix} 1 & \mathcal{V}_0^*(e^{j\omega_\ell}) & \cdots & \mathcal{V}_{n-1}^*(e^{j\omega_\ell}) \\ \mathcal{V}_0(e^{j\omega_\ell}) & \mathcal{V}_0(e^{j\omega_\ell})\mathcal{V}_0^*(e^{j\omega_\ell}) & \cdots & \mathcal{V}_0(e^{j\omega_\ell})\mathcal{V}_{n-1}^*(e^{j\omega_\ell}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{V}_{n-1}(e^{j\omega_\ell}) & \mathcal{V}_{n-1}(e^{j\omega_\ell})\mathcal{V}_0^*(e^{j\omega_\ell}) & \cdots & \mathcal{V}_0(e^{j\omega_\ell})\mathcal{V}_{n-1}^*(e^{j\omega_\ell}) \end{bmatrix} \quad (\text{D.3}) \end{aligned}$$

$$= \sum_{\ell=1}^p \phi^*(e^{j\omega_\ell}) \phi(e^{j\omega_\ell}) \quad (\text{D.4})$$

where $\omega_1, \dots, \omega_p$ denote the elements of the set $\Omega_{N_o}^u$, see (7.14), while we used that

$$\begin{aligned} \mathcal{V}_r(e^{j\omega_\ell})\mathcal{V}_s^*(e^{j\omega_\ell}) &= \mathcal{V}_0(e^{j\omega_\ell})\mathcal{V}_0^*(e^{j\omega_\ell})\mathbf{G}_b^r(e^{j\omega_\ell})(\mathbf{G}_b^s(e^{j\omega_\ell}))^* \\ \mathbf{G}_b(e^{j\omega_\ell})\mathbf{G}_b^*(e^{j\omega_\ell}) &= 1 \end{aligned}$$

see (E.3), (E.4) in Appendix E.

To be able to further elaborate (D.3) we first need some additional theory. Firstly, the doubly infinite matrix of basis functions \mathcal{B} given by (7.10) and (7.9) is orthonormal and complete

$$\mathcal{B}^T \mathcal{B} = \mathcal{B} \mathcal{B}^T = I$$

compare (E.11) in Appendix E, for further details we refer to [80, 81, 82]. From (7.9) (compare (E.10) in Appendix E) and the norm preserving inner product for a complete basis in a Hilbert space (Parseval) it follows that

$$\mathcal{B} \mathcal{B}^T = I \Leftrightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{V}_r(e^{j\omega}) \mathcal{V}_s^*(e^{j\omega}) d\omega = \begin{cases} I & r = s \\ 0 & r \neq s \end{cases} \quad (\text{D.5})$$

Secondly, using the inverse z-transform over the unit circle and (7.9), it follows that

$$\int_{-\pi}^{\pi} \mathcal{V}_r(e^{j\omega}) d\omega = b_r(0) = 0 \quad (\text{D.6})$$

Thirdly, for $\Omega_{N_o}^u = \Omega_{N_o}$ we have $p = N_o$ and

$$\frac{1}{N_o} \sum_{\ell=1}^{N_o} X(e^{j\omega_\ell}) \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) d\omega \quad \text{as } N_o \rightarrow \infty \quad (\text{D.7})$$

if $X(z)$ is Riemann integrable over $[-\pi, \pi]$. A sufficient condition for a function $X(z)$ to be Riemann integrable over $[a, b]$ is that $X(z)$ is continuous, except for at most a finite number of points, and bounded in $[a, b]$. In particular, a function $X(z)$ is Riemann integrable over $[a, b]$ when $X(z)$ is continuous over $[a, b]$. Hence $X(e^{j\omega})$ is Riemann integrable over $\omega \in [-\pi, \pi]$ if $X(z)$ represents the transfer function of an exponentially stable system, or the transfer function of a rational and stable system. The mean value theorem of integration states that, if $f : [a, b] \rightarrow \mathbb{R}$ is continuous on $[a, b]$, then there exists a $c \in (a, b)$ such that

$$\int_a^b f(x)dx = f(c)(b - a)$$

Using the mean value theorem of integration, it follows that for the real part of a continuous function $X(e^{j\omega})$ there holds that

$$\begin{aligned} & \left| \frac{1}{N_o} \sum_{\ell=1}^{N_o} \operatorname{Re}\{X(e^{j\omega_\ell})\} - \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Re}\{X(e^{j\omega})\} d\omega \right| \\ &= \left| \sum_{\ell=0}^{N_o-1} \left(\frac{1}{N_o} \operatorname{Re}\{X(e^{j\frac{2\pi\ell}{N_o}})\} - \frac{1}{2\pi} \int_{\frac{\pi(2\ell-1)}{N_o}}^{\frac{\pi(2\ell+1)}{N_o}} \operatorname{Re}\{X(e^{j\omega})\} d\omega \right) \right| \\ &\leq \frac{1}{N_o} \sum_{\ell=0}^{N_o-1} \left| \operatorname{Re}\{X(e^{j\frac{2\pi\ell}{N_o}})\} - \operatorname{Re}\{X(e^{j\zeta_\ell})\} \right| \quad \frac{\pi(2\ell-1)}{N_o} < \zeta_\ell < \frac{\pi(2\ell+1)}{N_o} \\ &\leq \frac{1}{N_o} \sum_{\ell=0}^{N_o-1} \frac{\pi}{N_o} \|X'(e^{j\omega})\|_\infty \\ &\leq \frac{\pi \|X'(e^{j\omega})\|_\infty}{N_o} \end{aligned}$$

where $X'(e^{j\omega}) = dX(e^{j\omega})/d\omega$. A similar result holds for the imaginary part of $X(e^{j\omega})$, so that

$$\left| \frac{1}{N_o} \sum_{\ell=1}^{N_o} X(e^{j\omega_\ell}) - \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) d\omega \right| \leq \sqrt{2} \pi \frac{\|X'(e^{j\omega})\|_\infty}{N_o} \quad (\text{D.8})$$

if $X(e^{j\omega})$ is continuous. Now, the basis generating system G_b is required to be exponentially stable. This implies that the \mathcal{V}_k are exponentially stable when k is finite, see (E.3). Thus, for finite k , we can apply (D.8) to (D.3). Comparing (D.3) and (D.8) it follows that we need to bound

$$\begin{aligned} & \left| \frac{d}{d\omega} \mathcal{V}_r(e^{j\omega}) \mathcal{V}_s^*(e^{j\omega}) \right| \\ &= \left| \frac{d}{d\omega} \mathcal{V}_0(e^{j\omega}) \mathcal{V}_0^*(e^{j\omega}) \mathbf{G}_b^r(e^{j\omega}) (\mathbf{G}_b^s(e^{j\omega}))^* \right| \\ &\leq \left| \frac{d}{d\omega} \mathcal{V}_0(e^{j\omega}) \mathcal{V}_0^*(e^{j\omega}) \right| |\mathbf{G}_b(e^{j\omega})|^{r-s} + |\mathcal{V}_0(e^{j\omega}) \mathcal{V}_0^*(e^{j\omega})| \left| \frac{d}{d\omega} \mathbf{G}_b^{|r-s|}(e^{j\omega}) \right| \end{aligned}$$

$$\begin{aligned}
&\leq \left| \frac{d}{d\omega} \mathcal{V}_0(e^{j\omega}) \mathcal{V}_0^*(e^{j\omega}) \right| + |\mathcal{V}_0(e^{j\omega}) \mathcal{V}_0^*(e^{j\omega})| |r-s| |\mathbf{G}_b(e^{j\omega})|^{r-s-1} \left| \frac{d}{d\omega} \mathbf{G}_b(e^{j\omega}) \right| \\
&\leq \left| \frac{d}{d\omega} \mathcal{V}_0(e^{j\omega}) \mathcal{V}_0^*(e^{j\omega}) \right| + |\mathcal{V}_0(e^{j\omega}) \mathcal{V}_0^*(e^{j\omega})| |r-s| \left| \frac{d}{d\omega} \mathbf{G}_b(e^{j\omega}) \right| \quad (\text{D.9})
\end{aligned}$$

Because we assumed the basis generating system to be exponentially stable, all elements of the matrices and all other terms in (D.9) are uniformly bounded, except $|r-s|$. Considering (D.3) we find that $|r-s| \leq n-1$ if $n \geq 1$. From (D.5), (D.6), (D.7), (D.8) and (D.1) it now follows directly that for $\Omega_{N_o}^u = \Omega_{N_o}$ and $n \geq 1$ there exists constants $c_2, c_3, c_4 \in \mathbb{R}$ such that

$$\begin{aligned}
\left\| \frac{1}{N_o} \Phi^* \Phi - I \right\|_2 &\leq n_p \frac{c_2 + (n-1)c_3}{N_o} \\
&\leq n_p \frac{n}{N_o} c_4
\end{aligned}$$

Finally, for $n = 0$ it follows directly from (D.3) that for $\Omega_{N_o}^u = \Omega_{N_o}$

$$\left\| \frac{1}{N_o} \Phi^* \Phi - I \right\|_2 = 0$$

which concludes the proof.

D.4 Proof of Theorem 7.3.5

First note that if

$$N \text{ cov}[\hat{\theta}] = \frac{1}{N_o} \Phi^* Q \Phi$$

then the results (7.35), (7.36) and (7.37) follow by direct calculation, and (D.7). Hence, to prove the theorem, we have to prove that

$$N \text{ cov}[\hat{\theta}] \rightarrow \frac{1}{N_o} \Phi^* Q \Phi$$

element wise. We will prove the stronger result that

$$\left\| N \text{ cov}[\hat{\theta}] - \frac{1}{N_o} \Phi^* Q \Phi \right\|_2 \rightarrow 0 \quad (\text{D.10})$$

Note that for the induced two norm of a matrix A there holds, see e.g. [53, page 57], that

$$\max_{ik} |A_{ik}| \leq \|A\|_2$$

Using (7.32) we find

$$\left\| N \text{ cov}[\hat{\theta}] - \frac{1}{N_o} \Phi^* Q \Phi \right\|_2$$

$$\begin{aligned}
&= \|N(\Phi^*\Phi)^{-1}\Phi^*\mathbb{E}[\mathbb{F}\mathbb{F}^*]((\Phi^*\Phi)^{-1}\Phi^*)^* - \frac{1}{N_o}\Phi^*Q\Phi\|_2 \\
&= \left\| \frac{1}{N_o} \left(I - \left(I - \left(\frac{1}{N_o}\Phi^*\Phi \right)^{-1} \right) \right) \Phi^* \left(Q - \left(Q - \frac{N}{N_o}\mathbb{E}[\mathbb{F}\mathbb{F}^*] \right) \right) \Phi \right. \\
&\quad \left. \left(I - \left(I - \left(\frac{1}{N_o}\Phi^*\Phi \right)^{-1} \right) \right)^* - \frac{1}{N_o}\Phi^*Q\Phi \right\|_2 \tag{D.11}
\end{aligned}$$

First we will bound $\|I - (\frac{1}{N_o}\Phi^*\Phi)^{-1}\|_2$. For $\|I - X\|_2 < 1$ there holds, see e.g. [53, page 59], that

$$\|I - X^{-1}\|_2 \leq \frac{\|I - X\|_2}{1 - \|I - X\|_2}$$

Using Lemma 7.3.4 it follows that for $c_4 \frac{n_p n}{N_o} < 1$

$$\|I - (\frac{1}{N_o}\Phi^*\Phi)^{-1}\|_2 \leq \frac{c_4 \frac{n_p n}{N_o}}{1 - c_4 \frac{n_p n}{N_o}} \tag{D.12}$$

Lemma 7.3.2 states that

$$\|Q - \frac{N}{N_o}\mathbb{E}[\mathbb{F}\mathbb{F}^*]\|_2 \leq c_1 \frac{N_o}{N} = c_1 \frac{1}{r} \tag{D.13}$$

Additionally

$$\|Q\|_2 = \left\| \frac{\Phi_v(\omega)}{\Phi_u(\omega)} \right\|_\infty$$

and, using the singular value decomposition,

$$\begin{aligned}
\left\| \frac{1}{\sqrt{N_o}}\Phi \right\|_2 &= \sqrt{\left\| \frac{1}{N_o}\Phi^*\Phi \right\|_2} \\
&= \sqrt{\|I - (I - \frac{1}{N_o}\Phi^*\Phi)\|_2} \\
&\leq \sqrt{\|I\|_2 + \|I - \frac{1}{N_o}\Phi^*\Phi\|_2} \\
&\leq \sqrt{1 + c_4 \frac{n_p n}{N_o}} \tag{D.14}
\end{aligned}$$

Straightforward elaboration of (D.11), using (D.2), now gives (D.10).

D.5 Proof of Theorem 7.3.6

From (7.31) we have

$$\begin{aligned}
\mathbb{E}[\hat{\theta}] - \theta_o &= \Psi Z + \Psi S \\
Z(e^{j\omega_\ell}) &= G_o(e^{j\omega_\ell}) - \phi(e^{j\omega_\ell})\theta_o = \sum_{k=n}^{\infty} L_k \mathcal{V}_k(e^{j\omega_\ell}) \tag{D.15}
\end{aligned}$$

Using the prior information that both the system and the basis generating system are exponentially stable, it follows that there exist constants $\mathcal{K} \in \mathbb{R}^{1 \times n_b}$ and $\eta \in \mathbb{R}$ such that

$$|L_k| \leq \mathcal{K} \eta^k \quad \eta < 1 \quad (\text{D.16})$$

Using a standard Taylor series expansion we find

$$\sum_{k=n}^{\infty} |L_k| \leq \mathcal{K} \frac{\eta^n}{1-\eta} \quad (\text{D.17})$$

The two norm of the vector Z now can be bounded as

$$\begin{aligned} \|Z\|_2 &= \sqrt{\sum_{\ell=1}^p \left| \sum_{k=n}^{\infty} L_k \mathcal{V}_k(e^{j\omega_\ell}) \right|^2} \\ &\leq \sqrt{\sum_{\ell=1}^p \left(\mathcal{K} |\mathcal{V}_0(e^{j\omega_\ell})| \frac{\eta^n}{1-\eta} \right)^2} \\ &\leq \sqrt{p} \frac{\eta^n}{1-\eta} \|\mathcal{K} |\mathcal{V}_0(e^{j\omega})|\|_{\infty} \end{aligned} \quad (\text{D.18})$$

where it should be noted that $\mathcal{V}_0 \in H_2^{n_b \times 1}$ and $\mathcal{K} \in \mathbb{R}^{1 \times n_b}$, and that n_b is assumed to be fixed and finite.

Using (D.12), (D.14) and (D.18) it follows that for $\Omega_{N_o}^u = \Omega_{N_o}$

$$\|I - (\frac{1}{N_o} \Phi^* \Phi)^{-1}\|_2 \|\frac{1}{N_o} \Phi^*\|_2 \|Z\|_2 \rightarrow 0 \quad \text{as} \quad \frac{n_p n}{N_o} \rightarrow 0 \quad (\text{D.19})$$

From the condition $\frac{n_p n}{N_o} \rightarrow 0$ and the fact that $n < n_p = n_b n + 1$ it follows that $\frac{n}{\sqrt{N_o}} \rightarrow 0$. Thus $\sqrt{N_o} \gg n$, so that we can split Z in two parts $Z = Z_1 + Z_2$ as

$$\begin{aligned} Z_1(e^{j\omega_\ell}) &= \sum_{k=n}^{\sqrt{p}} L_k \mathcal{V}_k(e^{j\omega_\ell}) \\ Z_2(e^{j\omega_\ell}) &= \sum_{k=\sqrt{p}+1}^{\infty} L_k \mathcal{V}_k(e^{j\omega_\ell}) \end{aligned}$$

for ease of notation we assume that \sqrt{p} is an integer, if this is not the case \sqrt{p} should be rounded to the nearest integer.

For ΨZ we now find from (D.19) that for $\Omega_{N_o}^u = \Omega_{N_o}$ and $\frac{n_p n}{\sqrt{N_o}} \eta^n \rightarrow 0$

$$\lim_{p \rightarrow \infty} \Psi Z = \lim_{p \rightarrow \infty} \frac{1}{p} \Phi^* Z = \lim_{p \rightarrow \infty} \frac{1}{p} \Phi^* Z_1 + \lim_{p \rightarrow \infty} \frac{1}{p} \Phi^* Z_2 \quad (\text{D.20})$$

For the first right hand side term of (D.20) we find

$$\begin{aligned}
\lim_{p \rightarrow \infty} \frac{1}{p} \Phi^* Z_1 &= \lim_{p \rightarrow \infty} \frac{1}{p} \begin{bmatrix} 1 & \cdots & 1 \\ \mathcal{V}_0(e^{-j\omega_1}) & \cdots & \mathcal{V}_0(e^{-j\omega_p}) \\ \vdots & \ddots & \vdots \\ \mathcal{V}_{n-1}(e^{-j\omega_1}) & \cdots & \mathcal{V}_{n-1}(e^{-j\omega_p}) \end{bmatrix} \begin{bmatrix} \sum_{k=n}^{\sqrt{p}} L_k \mathcal{V}_k(e^{j\omega_1}) \\ \vdots \\ \sum_{k=n}^{\sqrt{p}} L_k \mathcal{V}_k(e^{j\omega_p}) \end{bmatrix} \\
&= \lim_{p \rightarrow \infty} \frac{1}{p} \begin{bmatrix} \sum_{\ell=1}^p \sum_{k=n}^{\sqrt{p}} L_k \mathcal{V}_k(e^{j\omega_\ell}) \\ \sum_{\ell=1}^p \mathcal{V}_0(e^{-j\omega_\ell}) \sum_{k=n}^{\sqrt{p}} L_k \mathcal{V}_k(e^{j\omega_\ell}) \\ \vdots \\ \sum_{\ell=1}^p \mathcal{V}_{n-1}(e^{-j\omega_\ell}) \sum_{k=n}^{\sqrt{p}} L_k \mathcal{V}_k(e^{j\omega_\ell}) \end{bmatrix} \\
&= \lim_{p \rightarrow \infty} \frac{1}{p} \begin{bmatrix} \sum_{k=n}^{\sqrt{p}} L_k \sum_{\ell=1}^p \mathcal{V}_k(e^{j\omega_\ell}) \\ \sum_{k=n}^{\sqrt{p}} \left(\sum_{\ell=1}^p \mathcal{V}_0(e^{-j\omega_\ell}) \mathcal{V}_k^T(e^{j\omega_\ell}) \right) L_k^T \\ \vdots \\ \sum_{k=n}^{\sqrt{p}} \left(\sum_{\ell=1}^p \mathcal{V}_{n-1}(e^{-j\omega_\ell}) \mathcal{V}_k^T(e^{j\omega_\ell}) \right) L_k^T \end{bmatrix} \quad (D.21)
\end{aligned}$$

Combining (D.5), (D.6), (D.8) and (D.9) it follows that for $\Omega_{N_o}^u = \Omega_{N_o}$ there exists constants $C_1 \in \mathbb{R}^{n_b \times 1}$ and $C_2 \in \mathbb{R}^{n_b \times n_b}$ such that

$$\left| \frac{1}{p} \sum_{\ell=1}^p \mathcal{V}_s(e^{j\omega_\ell}) \right| \leq \frac{C_1}{\sqrt{p}} \quad \text{for } s \in \mathbb{N} \quad s \leq \sqrt{p} \quad (D.22)$$

$$\left| \frac{1}{p} \sum_{\ell=1}^p \mathcal{V}_r(e^{-j\omega_\ell}) \mathcal{V}_s^T(e^{j\omega_\ell}) \right| \leq \frac{C_2}{\sqrt{p}} \quad \text{for } r, s \in \mathbb{N} \quad r < s \leq \sqrt{p} \quad (D.23)$$

Since n_b is assumed to be fixed and finite, we now find from (D.16) and (D.17) that

$$\lim_{p \rightarrow \infty} \Psi Z_1 = 0$$

Similar to (D.21) we find for the second right hand side term of (D.20)

$$\lim_{p \rightarrow \infty} \frac{1}{p} \Phi^* Z_2 = \lim_{p \rightarrow \infty} \frac{1}{p} \begin{bmatrix} \sum_{k=\sqrt{p}+1}^{\infty} \sum_{\ell=1}^p L_k \mathcal{V}_k(e^{j\omega_\ell}) \\ \sum_{k=\sqrt{p}+1}^{\infty} \sum_{\ell=1}^p \mathcal{V}_0(e^{-j\omega_\ell}) L_k \mathcal{V}_k(e^{j\omega_\ell}) \\ \vdots \\ \sum_{k=\sqrt{p}+1}^{\infty} \sum_{\ell=1}^p \mathcal{V}_{n-1}(e^{-j\omega_\ell}) L_k \mathcal{V}_k(e^{j\omega_\ell}) \end{bmatrix}$$

Using the fact that

$$|\mathcal{V}_k(e^{j\omega})| = |\mathcal{V}_0(e^{j\omega})| |\mathbf{G}_b^k(e^{j\omega})| = |\mathcal{V}_0(e^{j\omega})| \quad (\text{D.24})$$

it follows from (D.16) and (D.17) that

$$\begin{aligned} \left| \lim_{p \rightarrow \infty} \frac{1}{p} \Phi^* Z_2 \right| &\leq \lim_{p \rightarrow \infty} \begin{bmatrix} \sum_{k=\sqrt{p}+1}^{\infty} |L_k| |\mathcal{V}_0(e^{j\omega})|_\infty \\ \sum_{k=\sqrt{p}+1}^{\infty} |\mathcal{V}_0(e^{j\omega})|_\infty |L_k| |\mathcal{V}_0(e^{j\omega})|_\infty \\ \vdots \\ \sum_{k=\sqrt{p}+1}^{\infty} |\mathcal{V}_0(e^{j\omega})|_\infty |L_k| |\mathcal{V}_0(e^{j\omega})|_\infty \end{bmatrix} \\ &\leq \lim_{p \rightarrow \infty} \frac{\eta^{\sqrt{p}+1}}{1-\eta} \begin{bmatrix} \mathcal{K} |\mathcal{V}_0(e^{j\omega})|_\infty \\ |\mathcal{V}_0(e^{j\omega})|_\infty \mathcal{K} |\mathcal{V}_0(e^{j\omega})|_\infty \\ \vdots \\ |\mathcal{V}_0(e^{j\omega})|_\infty \mathcal{K} |\mathcal{V}_0(e^{j\omega})|_\infty \end{bmatrix} \quad (\text{D.25}) \end{aligned}$$

where $|X(\omega)|_\infty$ denotes the vector consisting of the infinity norms of the elements of the vector $X(\omega)$.

Since n_b is assumed to be fixed and finite it follows that

$$\lim_{p \rightarrow \infty} \Psi Z_2 = 0$$

For $S(e^{j\omega_\ell})$ we have, see Lemma 5.3.1,

$$|S(e^{j\omega_\ell})| \leq \frac{\sqrt{N_o}}{N} \frac{\bar{u}^P + \bar{u}}{|U_o(e^{j\omega_\ell})|} \frac{M\rho(1-\rho^{-N})}{(\rho-1)^2} \rho^{-N_s} \quad (\text{D.26})$$

Since $p = N_o \leq N$, it follows that

$$\lim_{p \rightarrow \infty} |S(e^{j\omega_\ell})| = 0 \quad (\text{D.27})$$

Using (D.26) and $p = N_o$ the two norm of the vector \mathbf{S} can be bounded as

$$\|\mathbf{S}\|_2 \leq \frac{N_o}{N} \frac{\bar{u}^P + \bar{u}}{|U_o(e^{j\omega_\ell})|} \frac{M\rho(1 - \rho^{-N})}{(\rho - 1)^2} \rho^{-N_s} \quad (\text{D.28})$$

Using (D.12), (D.14) and (D.28) it follows that for $\Omega_{N_o}^u = \Omega_{N_o}$

$$\|I - (\frac{1}{N_o} \Phi^* \Phi)^{-1}\|_2 \|\frac{1}{N_o} \Phi^*\|_2 \|\mathbf{S}\|_2 \rightarrow 0 \quad \text{as} \quad \frac{n_p n}{N_o} \rightarrow 0$$

Hence

$$\begin{aligned} \lim_{p \rightarrow \infty} \Psi \mathbf{S} &= \lim_{p \rightarrow \infty} \frac{1}{p} \Phi^* \mathbf{S} \\ &= \lim_{p \rightarrow \infty} \frac{1}{p} \begin{bmatrix} \sum_{\ell=1}^p S(e^{j\omega_\ell}) \\ \sum_{\ell=1}^p \mathcal{V}_0(e^{-j\omega_\ell}) S(e^{j\omega_\ell}) \\ \vdots \\ \sum_{\ell=1}^p \mathcal{V}_{n-1}(e^{-j\omega_\ell}) S(e^{j\omega_\ell}) \end{bmatrix} \end{aligned}$$

Using (D.24) and (D.27) we now find

$$\lim_{p \rightarrow \infty} \Psi \mathbf{S} = 0$$

which concludes the proof.

D.6 Proof of Theorem 7.3.7

Using (7.19) and (7.20) we find

$$\begin{aligned} |\hat{G} - \phi\theta|^2 &= |G_o - \phi\theta + \frac{V^s}{U^s} + S|^2 \\ &= |G_o - \phi\theta|^2 + |\frac{V^s}{U^s} + S|^2 + (G_o - \phi\theta)(\frac{V^s}{U^s} + S)^* + (G_o - \phi\theta)^*(\frac{V^s}{U^s} + S) \end{aligned}$$

Since $\mathbb{E}[V^s] = 0$, and $|S(e^{j\omega_\ell})| \rightarrow 0$ as $N_o \rightarrow \infty$, see (D.27), it follows from (7.38) that

$$\theta^* = \arg \min_{\theta} \lim_{p \rightarrow \infty} \frac{1}{p} \sum_{k=1}^p |G_o(e^{j\omega_k}) - \phi(e^{j\omega_k})\theta|^2 |W(e^{j\omega_k})|^2 \quad (\text{D.29})$$

$$= \arg \min_{\theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} |G_o(e^{j\omega}) - \phi(e^{j\omega})\theta|^2 |W(e^{j\omega})|^2 d\omega \quad \text{for } \Omega_{N_o}^u = \Omega_{N_o} \quad (\text{D.30})$$

Further elaboration gives, using (D.15),

$$|G_o - \phi\theta|^2 = |\phi\theta_o - \phi\theta + Z|^2 = |\phi(\theta_o - \theta)|^2 + |Z|^2 + Z^* \phi(\theta_o - \theta) + (\theta_o - \theta)^* \phi^* Z$$

Note that

$$\frac{1}{p} \sum_{k=1}^p \phi^*(e^{j\omega_k}) Z(e^{j\omega_k}) = \frac{1}{p} \Phi^* Z$$

From (D.17), (D.21), (D.22), (D.23), (D.25) and the fact that $\Phi^* Z \in \mathbb{C}^{n_p \times 1}$ it follows directly that for $\Omega_{N_o}^u = \Omega_{N_o}$ there exists a constant $c_5 \in \mathbb{R}$ such that

$$\left\| \frac{1}{p} \Phi^* (Z_1 + Z_2) \right\|_2 \leq \sqrt{n_p} \left(\frac{\eta^n}{\sqrt{p}} + \eta^{\sqrt{p}+1} \right) c_5 \quad (\text{D.31})$$

Direct calculation of the maximum of $\sqrt{n_p} \eta^n$ over n shows that $\sqrt{n_p} \eta^n$ is bounded, if n_b is bounded. Since $n_p \leq p$, it follows that $\sqrt{n_p} \eta^{\sqrt{p}+1} \rightarrow 0$ as $p \rightarrow \infty$. Hence, noting that $Z = Z_1 + Z_2$, we find for $\Omega_{N_o}^u = \Omega_{N_o}$

$$\lim_{p \rightarrow \infty} \left\| \frac{1}{p} \Phi^* Z \right\|_2 = 0$$

Hence, when no weighting function $W(e^{j\omega_k})$ is used, we have for $\Omega_{N_o}^u = \Omega_{N_o}$

$$\begin{aligned} \theta^* &= \arg \min_{\theta} \frac{1}{2\pi} \int_{-\pi}^{\pi} |\phi(e^{j\omega})(\theta_o - \theta)|^2 d\omega \\ &= \arg \min_{\theta} (\theta_o - \theta)^* \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \phi^*(e^{j\omega}) \phi(e^{j\omega}) \right) (\theta_o - \theta) \\ &= \arg \min_{\theta} \|\theta_o - \theta\|_2^2 \\ &= \theta_o \end{aligned}$$

where we used that, see (7.22) and (D.5),

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \phi^*(e^{j\omega}) \phi(e^{j\omega}) = I$$

which concludes the proof.

D.7 Proof of Theorem 7.4.1

To be able to prove the result we need an indication of the magnitude of the two norm of the vector $P(e^{j\omega})$

$$\begin{aligned}\|P(e^{j\omega})\|_2 &= \sqrt{1 + \sum_{k=0}^{n-1} \|\mathcal{V}_k(e^{j\omega})\|_2^2} \\ &= \sqrt{1 + n\|\mathcal{V}_0(e^{j\omega})\|_2^2} \quad (\text{D.32})\end{aligned}$$

$$\begin{aligned}&\leq \sqrt{1 + n n_b c_6} \\ &\leq \sqrt{n_p} c_7 \quad (\text{D.33})\end{aligned}$$

where $c_6, c_7 \in \mathbb{R}$ are constants. Using (D.10) we now find for $\Omega_{N_o}^u = \Omega_{N_o}$ and $\frac{n_p n}{N_o} \rightarrow 0$

$$\begin{aligned}&\lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} \frac{N}{n_p} \text{var}[\hat{G}^f(e^{j\omega})] \\ &= \lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} \frac{N}{n_p} P(e^{j\omega})(\Phi^* \Phi)^{-1} \Phi^* \mathbb{E}[\text{FF}^*] \Phi (\Phi^* \Phi)^{-1} P^*(e^{j\omega}) \\ &= \lim_{N_o \rightarrow \infty} \frac{1}{N_o n_p} P(e^{j\omega}) \Phi^* Q \Phi P^*(e^{j\omega}) \quad (\text{D.34})\end{aligned}$$

From (D.4) and the definition of Q in (7.33) it follows that

$$\begin{aligned}P(e^{j\omega}) \Phi^* Q \Phi P^*(e^{j\omega}) &= P(e^{j\omega}) \left(\sum_{\ell=1}^p Q(e^{j\omega_\ell}) \phi^*(e^{j\omega_\ell}) \phi(e^{j\omega_\ell}) \right) P^*(e^{j\omega}) \\ &= \sum_{\ell=1}^p Q(e^{j\omega_\ell}) |\phi(e^{j\omega_\ell}) P^*(e^{j\omega})|^2\end{aligned}$$

Thus, using (7.22) and (7.40), we now have

$$\lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} \frac{N}{n_p} \text{var}[\hat{G}^f(e^{j\omega})] = \lim_{N_o \rightarrow \infty} \frac{1}{N_o} \sum_{\ell=1}^{N_o} Q(e^{j\omega_\ell}) \frac{1}{n_p} \left| 1 + \sum_{k=0}^{n-1} \mathcal{V}_k^T(e^{j\omega}) \mathcal{V}_k(e^{-j\omega_\ell}) \right|^2$$

Using (7.9) we can write

$$\sum_{k=0}^{n-1} \mathcal{V}_k^T(e^{j\omega}) \mathcal{V}_k(e^{-j\omega_\ell}) = \mathcal{V}_0^T(e^{j\omega}) \mathcal{V}_0(e^{-j\omega_\ell}) \sum_{k=0}^{n-1} [\mathbf{G}_b(e^{j\omega}) \mathbf{G}_b(e^{-j\omega_\ell})]^k$$

Note that for any $x \in \mathbb{C}$ there holds that

$$\frac{x^n - 1}{x - 1} = \frac{(x - 1)x^{n-1} + x^{n-1} - 1}{x - 1} = x^{n-1} + \frac{x^{n-1} - 1}{x - 1} \quad (\text{D.35})$$

Using (D.35) recursively it follows that

$$\frac{x^n - 1}{x - 1} = x^{n-1} + x^{n-2} + \cdots + x + 1 = \sum_{k=0}^{n-1} x^k$$

Also note that

$$\lim_{x \rightarrow 1} \frac{x^n - 1}{x - 1} = n$$

It follows that we can write

$$\sum_{k=0}^{n-1} [\mathbf{G}_b(e^{j\omega}) \mathbf{G}_b(e^{-j\omega_\ell})]^k = \frac{[\mathbf{G}_b(e^{j\omega}) \mathbf{G}_b(e^{-j\omega_\ell})]^n - 1}{\mathbf{G}_b(e^{j\omega}) \mathbf{G}_b(e^{-j\omega_\ell}) - 1}$$

Finally note that

$$|\mathbf{G}_b(e^{j\zeta})|^2 = \mathbf{G}_b(e^{j\zeta}) \mathbf{G}_b(e^{-j\zeta}) = 1$$

so that

$$\mathbf{G}_b(e^{-j\zeta}) = \frac{1}{\mathbf{G}_b(e^{j\zeta})}$$

which concludes the proof of (7.41), (7.42) and (7.43).

To prove (7.4.1) note that from (D.34), Lemma 7.3.4 and (D.33) we find for $Q = I$

$$\lim_{\substack{N_o \rightarrow \infty \\ r \rightarrow \infty}} \frac{N}{n_p} \text{var}[\hat{G}^f(e^{j\omega})] = \lim_{N_o \rightarrow \infty} \frac{1}{n_p} P(e^{j\omega}) P^*(e^{j\omega})$$

Using (D.32) now concludes the proof.

D.8 Proof of Corollary 7.4.2

Using the fact that

$$\sin(\omega) = \frac{e^{j\omega} - e^{-j\omega}}{2j}$$

it follows directly that for $\mathcal{V}_0(e^{j\omega}) = e^{-j\omega}$ and $\mathbf{G}_b(e^{j\omega}) = e^{-j\omega}$

$$\mathcal{P}_n(\omega, \zeta) = \frac{1}{n_p} \left| 1 + e^{j(\zeta - \omega) \frac{n+1}{2}} \frac{\sin(\frac{n}{2}(\zeta - \omega))}{\sin(\frac{1}{2}(\zeta - \omega))} \right|^2$$

The first statement of Corollary 7.4.2 now follows immediately from Theorem 7.4.1.

To prove the second statement of Corollary 7.4.2 first note that, see e.g. [147, pages 418-419],

$$\frac{1}{2\pi n} \frac{\sin^2(n\omega/2)}{\sin^2(\omega/2)} = F_n(\omega)$$

where $F_n(\omega)$ is the Fejer kernel. Denote

$$\tilde{F}_n(\omega) = \frac{1}{\sqrt{2\pi n}} e^{j\omega \frac{n+1}{2}} \frac{\sin(n\omega/2)}{\sin(\omega/2)}$$

so that

$$|\tilde{F}_n(\omega)| = \sqrt{F_n(\omega)}$$

and

$$\mathcal{P}_n(\omega, \zeta) = \left| \sqrt{\frac{1}{n_p}} + \sqrt{\frac{2\pi n}{n_p}} \tilde{F}_n(\zeta - \omega) \right|^2$$

For the Fejer kernel $F_n(\omega)$ there holds, see e.g. [147, pages 400-420], that

$$\begin{aligned} F_n(\omega) &\leq F_n(0) = \frac{n}{2\pi} \\ \lim_{n \rightarrow \infty} F_n(\omega) &= \tilde{\delta}(\omega) \end{aligned}$$

where $\tilde{\delta}(\omega)$ is the Dirac delta. Noting that for the pulse basis $n_p = n + 1$, it follows that as $n \rightarrow \infty$

$$\left| \sqrt{\frac{1}{n_p}} \right| \left| \sqrt{\frac{2\pi n}{n_p}} \tilde{F}_n(\zeta - \omega) \right| = \frac{\sqrt{2\pi n}}{n+1} \sqrt{F_n(\zeta - \omega)} \rightarrow \begin{cases} \frac{n}{n+1} & \zeta = \omega \\ 0 & \zeta \neq \omega \end{cases}$$

Additionally, note that for any $x_1, x_2 \in \mathbb{C}$ there holds that

$$||x_1| - |x_2|| \leq |x_1 + x_2| \leq |x_1| + |x_2|$$

so that

$$|x_1|^2 + |x_2|^2 - 2|x_1||x_2| \leq |x_1 + x_2|^2 \leq |x_1|^2 + |x_2|^2 + 2|x_1||x_2|$$

Since $\Phi_v(\omega)/\Phi_u(\omega)$ is bounded it now follows that

$$\lim_{n_p \rightarrow \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sqrt{\frac{1}{n_p}} + \sqrt{\frac{2\pi n}{n_p}} \tilde{F}_n(\zeta - \omega) \right|^2 \frac{\Phi_v(\zeta)}{\Phi_u(\zeta)} d\zeta = \lim_{n_p \rightarrow \infty} \int_{-\pi}^{\pi} F_n(\zeta - \omega) \frac{\Phi_v(\zeta)}{\Phi_u(\zeta)} d\zeta$$

Again, note that for the pulse basis $n_p = n + 1$. Finally, if $h(\zeta)$ is Lipschitz continuous of order 1 then, see e.g. [147, pages 400-420],

$$\int_{-\pi}^{\pi} h(\zeta) F_n(\zeta - \omega) d\zeta = h(\omega) + \mathcal{O}\left(\frac{\log(n)}{n}\right)$$

which concludes the proof.

D.9 Proof of Theorem 7.4.3

Using (7.39), (7.31) and (D.15) we find

$$\begin{aligned} G_o(e^{j\omega}) - \mathbb{E}[\hat{G}^f(e^{j\omega})] &= G_o(e^{j\omega}) - P(e^{j\omega})\mathbb{E}[\hat{\theta}] \\ &= G_o(e^{j\omega}) - P(e^{j\omega})\theta_o - P(e^{j\omega})\Psi Z - P(e^{j\omega})\Psi S \\ &= Z(e^{j\omega}) - P(e^{j\omega})\Psi Z - P(e^{j\omega})\Psi S \end{aligned}$$

From (D.12), (D.14), (D.18) and (D.33) we find that for $\Omega_{N_o}^u = \Omega_{N_o}$ and $c_4 \frac{n_p n}{N_o} < 1$ there exists a constant $c_8 \in \mathbb{R}$ such that

$$\|P(e^{j\omega})\|_2 \|I - (\frac{1}{N_o} \Phi^* \Phi)^{-1}\|_2 \|\frac{1}{N_o} \Phi^*\|_2 \|Z\|_2 \leq \sqrt{n_p} \eta^n \frac{n_p n}{N_o} c_8 \quad \text{as} \quad \frac{n_p n}{N_o} \rightarrow 0$$

Direct calculation of the maximum of $\sqrt{n_p} \eta^n$ over n shows that $\sqrt{n_p} \eta^n$ is bounded, given that n_b is bounded. Therefore

$$\|P(e^{j\omega})\|_2 \|I - (\frac{1}{N_o} \Phi^* \Phi)^{-1}\|_2 \|\frac{1}{N_o} \Phi^*\|_2 \|Z\|_2 \rightarrow 0 \quad \text{as} \quad \frac{n_p n}{N_o} \rightarrow 0$$

It follows that for $\Omega_{N_o}^u = \Omega_{N_o}$ and $\frac{n_p n}{N_o} \rightarrow 0$ we can write

$$\lim_{p \rightarrow \infty} P(e^{j\omega})\Psi Z = \lim_{p \rightarrow \infty} \frac{1}{p} P(e^{j\omega})\Phi^* Z$$

Using (D.31) and (D.33) we find that there exists a constant $c_9 \in \mathbb{R}$ such that

$$|\frac{1}{p} P(e^{j\omega})\Phi^* Z| \leq n_p \left(\frac{\eta^n}{\sqrt{p}} + \eta^{\sqrt{p}+1} \right) c_9$$

Again, direct calculation of the maximum of $n_p \eta^n$ over n shows that $n_p \eta^n$ is bounded, if n_b is bounded. Since $n_p \leq p$, it follows that $n_p \eta^{\sqrt{p}+1} \rightarrow 0$ as $p \rightarrow \infty$. Hence, for $\Omega_{N_o}^u = \Omega_{N_o}$ and $\frac{n_p n}{N_o} \rightarrow 0$ we find

$$\lim_{p \rightarrow \infty} P(e^{j\omega})\Psi Z = 0$$

From (D.12), (D.14), (D.28) and (D.33) we find that for $\Omega_{N_o}^u = \Omega_{N_o}$ and $c_4 \frac{n_p n}{N_o} < 1$ there exists a constant $c_{10} \in \mathbb{R}$ such that

$$\|P(e^{j\omega})\|_2 \|I - (\frac{1}{N_o} \Phi^* \Phi)^{-1}\|_2 \|\frac{1}{N_o} \Phi^*\|_2 \|S\|_2 \leq \sqrt{\frac{n_p}{N_o}} \frac{n_p n}{N} \rho^{-N_s} c_{10} \quad \text{as} \quad \frac{n_p n}{N_o} \rightarrow 0$$

Since $N \geq N_o \geq n_p$ it follows that

$$\|P(e^{j\omega})\|_2 \|I - (\frac{1}{N_o} \Phi^* \Phi)^{-1}\|_2 \|\frac{1}{N_o} \Phi^*\|_2 \|S\|_2 \rightarrow 0 \quad \text{as} \quad \frac{n_p n}{N_o} \rightarrow 0$$

As a result we can write for $\Omega_{N_o}^u = \Omega_{N_o}$ and $\frac{n_p n}{N_o} \rightarrow 0$

$$\lim_{p \rightarrow \infty} P(e^{j\omega})\Psi S = \lim_{p \rightarrow \infty} \frac{1}{p} P(e^{j\omega})\Phi^* S$$

Using (D.14), (D.28) and (D.33) we find that there exists a constant $c_{11} \in \mathbb{R}$ such that

$$\left| \frac{1}{p} P(e^{j\omega})\Phi^* S \right| \leq \sqrt{\frac{n_p}{N_o} \frac{N_o}{N}} \rho^{-N_s} c_{11} \quad \text{as} \quad \frac{n_p n}{N_o} \rightarrow 0$$

Since $\frac{n_p n}{N_o} \rightarrow 0$ implies that $\frac{n_p}{N_o} \rightarrow 0$ it follows that for $\Omega_{N_o}^u = \Omega_{N_o}$ and $\frac{n_p n}{N_o} \rightarrow 0$

$$\lim_{p \rightarrow \infty} P(e^{j\omega})\Psi S = 0$$

Using (D.15), (D.16) and the fact that

$$|\mathcal{V}_k(e^{j\omega_\ell})| = |\mathcal{V}_0(e^{j\omega_\ell})| |\mathbf{G}_b^k(e^{j\omega})| = |\mathcal{V}_0(e^{j\omega_\ell})|$$

we find for $Z(e^{j\omega})$

$$\begin{aligned} |Z(e^{j\omega})| &\leq \mathcal{K} |\mathcal{V}_0(e^{j\omega})| \sum_{k=n}^{\infty} \eta^k \\ &= \mathcal{K} |\mathcal{V}_0(e^{j\omega})| \frac{\eta^n}{1-\eta} \end{aligned}$$

which concludes the proof.

Appendix E

System based orthonormal basis functions

E.1 Properties

In this section we will discuss a number of properties of the generalized orthonormal basis functions which were introduced by [80]. The specific advantages of these functions will be discussed in the following section. The theory described in this section is due to [80, 81, 82], where multivariable (MIMO) systems are addressed. We will however only consider scalar (SISO) system.

We will denote $(\cdot)^*$ as the complex conjugate transpose of a matrix. \mathbb{N} is the set of nonnegative integers, $\ell_2[0, \infty)$ is the space of squared summable sequences on the time interval \mathbb{N} , $H_2^{p \times q}$ is the field of complex valued matrix functions which are analytic outside the unit disc and squared integrable over the unit circle, and \mathbb{RH}_2 is the field of real rational functions which are analytic outside the unit disc and squared integrable over the unit circle.

A causal stable linear time-invariant finite-dimensional discrete-time system will be represented by its rational transfer function $G \in \mathbb{RH}_2$. A state-space realization (A, B, C, D) of G is stable if all eigenvalues of A lie strictly within the unit circle. If a realization is stable, the controllability Gramian P and observability Gramian Q are defined as the solutions to the Lyapunov equations $APA^* + BB^* = P$ and $A^*QA + C^*C = Q$ respectively. A stable realization is called (internally) balanced if $P = Q = \Sigma$, with $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$, $\sigma_1 \geq \dots \geq \sigma_n$, a diagonal matrix with the positive Hankel singular values as diagonal elements. A stable realization is called input balanced if $P = I$, $Q = \Sigma^2$.

A system $G \in \mathbb{RH}_2$ is called inner if it is stable and satisfies $G^T(z^{-1})G(z) = I$. Since an inner function G is analytic outside and on the unit circle, it has a Laurent series expansion $\sum_{k=0}^{\infty} g(k)z^{-k}$, where $g(k)$ are the Markov parameters. For later use we will formalize a class of inner functions.

Definition E.1.1

$\mathcal{G}_1 := \{ \text{all square inner functions } G \text{ with McMillan degree } > 0 \text{ such that } \|g(0)\|_2 < 1 \}$.

Using the fact that the system $G_o \in \mathbb{RH}_2$ has a state-space representation, its transfer function can be written as

$$G_o(z) = D + C(zI - A)^{-1}B \quad (\text{E.1})$$

Let n_o denote the McMillan degree of G_o . We will denote

$$G_o = (A, B, C, D)$$

meaning that the system G_o has a state-space representation given by the matrices A , B , C , and D . Let

$$G_b = (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) \quad \mathbf{A} \in \mathbb{R}^{n_b \times n_b}, \quad \mathbf{B} \in \mathbb{R}^{n_b \times 1}, \quad \mathbf{C} \in \mathbb{R}^{1 \times n_b}, \quad \mathbf{D} \in \mathbb{R} \quad (\text{E.2})$$

denote a minimal input balanced state-space representation of a causal stable system with McMillan degree n_b . We will call G_b the basis generating system or basis generator. In [81] a procedure is given to find a minimal balanced realization of an inner function $\mathbf{G}_b(z)$ directly from \mathbf{A} and \mathbf{B} .

Proposition E.1.2 ([81]) *Let $G_b \in \mathbb{RH}_2$, with input balanced realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, McMillan degree $n_b > 0$ and $\text{rank}(\mathbf{B}) = 1$. Then*

- (a) *there exist matrices \mathbf{C}, \mathbf{D} such that $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ is a minimal balanced realization of a square inner function $\mathbf{G}_b \in \mathcal{G}_1$.*
- (b) *A realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ has the property mentioned in (a) if and only if*

$$\begin{aligned} \mathbf{C} &= \mathbf{B}^*(I + \mathbf{A}^*)^{-1}(I + \mathbf{A}) \\ \mathbf{D} &= [\mathbf{B}^*(I + \mathbf{A}^*)^{-1}\mathbf{B} - 1] \end{aligned}$$

Proposition E.1.2 implies that the poles of the basis generating system are retained in the corresponding inner function.

The z -transformed orthonormal basis functions, which we will call the generating transfer functions, now can be calculated recursively as

$$\begin{aligned} \mathcal{V}_0(z) &= (zI - \mathbf{A})^{-1}\mathbf{B} & \mathcal{V}_0 &\in H_2^{n_b \times 1} \\ \mathcal{V}_k(z) &= \mathcal{V}_0(z)\mathbf{G}_b^k(z) & \mathbf{G}_b &\in \mathbb{RH}_2 \end{aligned} \quad (\text{E.3})$$

where $\mathbf{G}_b(z)$ is an inner function. Hence

$$|\mathbf{G}_b(z)| = 1 \quad (\text{E.4})$$

$$|\mathcal{V}_k(z)| = |\mathcal{V}_0(z)||\mathbf{G}_b(z)|^k = |\mathcal{V}_0(z)| \quad (\text{E.5})$$

where $|x|$ denotes the vector consisting of the absolute values of each element of the vector x (i.e. $|x|(k) = |x(k)|$).

The system $G_o \in \mathbb{RH}_2$ now can be written as

$$G_o(z) = D + \sum_{k=0}^{\infty} L_k \mathcal{V}_k(z) \quad D \in \mathbb{R}, \quad L_k \in \mathbb{R}^{n_b \times 1}, \quad \mathcal{V}_k \in H_2^{n_b \times 1} \quad (\text{E.6})$$

where D is the feedthrough matrix of the state-space description of G_o c.q. the first markov parameter $g_o(0)$, L_k are the generalized expansion coefficients, i.e. the generalized impulse response, and $\mathcal{V}_k(z)$ are the generating transfer functions. In [81] it is proven that if $G_b = G_o$ then $L_0 = \mathbf{C}$ and $L_k = 0$ for $k > 0$. Consequently, a generalized FIR model of order $m \cdot n_b$ is defined as

$$G(z) = D + \sum_{k=0}^{m-1} L_k \mathcal{V}_k(z) \quad (\text{E.7})$$

where D and L_k , $k = 0, \dots, m-1$, are the parameters of the model.

The convergence result on the generalized expansion coefficients L_k has the following form.

Proposition E.1.3 ([80, lemma 5.4.1 and its proof], [81]) *Let $G_o \in \mathbb{RH}_2$, with McMillan degree n_o and an input balanced realization (A, B, C, D) . Let (\mathbf{A}, \mathbf{B}) be an input balanced pair that generates an inner transfer function $\mathbf{G}_b \in \mathcal{G}_1$ with McMillan degree n_b , leading to an orthonormal basis $\Phi(\mathbf{G}_b)$, see (E.11). Let*

μ_i , $i = 1, \dots, n_o$ denote the eigenvalues of A , and

ϱ_j , $j = 1, \dots, n_b$ denote the eigenvalues of \mathbf{A} .

Then the dynamical system having the L_k as its impulse response has a state-space realization $(X_o, Y_o, Z_o, 0)$ that satisfies

(a) X_o has dimension n_o

(b) X_o has eigenvalues λ_i , $i = 1, \dots, n_o$, that satisfy

$$|\lambda_i| = \prod_{j=1}^{n_b} \left| \frac{\mu_i - \varrho_j}{1 - \mu_i \varrho_j} \right|$$

Note that when the sets of eigenvalues $\{\mu_i\}$, $\{\varrho_j\}$ coincide, then $\lambda_i = 0$, for all i .

When a finite number of expansion coefficients is estimated, it is useful to know a state-space realization of the resulting model.

Proposition E.1.4 ([81]) *Consider the transfer function*

$$\hat{G}(z) = \hat{D} + \sum_{k=0}^{m-1} \hat{L}_k \mathcal{V}_k(z)$$

with $m > 0$ and $\mathcal{V}_k(z)$ the generating transfer functions of an orthonormal basis $\Phi(\mathbf{G}_b)$, where the inner function $\mathbf{G}_b \in \mathcal{G}_1$ has a minimal state-space realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ with dimension $n_b > 0$.

Then $\hat{G}(z)$ has a state-space realization $(\mathbf{A}_m, \mathbf{B}_m, \hat{\mathbf{C}}_m, \hat{\mathbf{D}})$ with state dimension $m \cdot n_b$, where

$$\hat{\mathbf{C}}_m = \begin{bmatrix} \hat{L}_0 & \hat{L}_1 & \cdots & \hat{L}_{m-1} \end{bmatrix}$$

and $\mathbf{A}_1 = \mathbf{A}$, $\mathbf{B}_1 = \mathbf{B}$, and for $m > 1$

$$\mathbf{A}_m = \begin{bmatrix} \mathbf{A} & 0 & \cdots & \cdot & 0 \\ \mathbf{BC} & \mathbf{A} & 0 & \cdot & 0 \\ \mathbf{BDC} & \mathbf{BC} & \cdot & \cdot & 0 \\ \vdots & \vdots & \cdot & \ddots & 0 \\ \mathbf{BD}^{m-2}\mathbf{C} & \mathbf{BD}^{m-1}\mathbf{C} & \cdots & \mathbf{BC} & \mathbf{A} \end{bmatrix} \quad \mathbf{B}_m = \begin{bmatrix} \mathbf{B} \\ \mathbf{BD} \\ \mathbf{BD}^2 \\ \vdots \\ \mathbf{BD}^{m-1} \end{bmatrix}$$

Note that the McMillan degree of this system in general will be large.

The undisturbed output is given as

$$\begin{aligned} y_o(t) &= Du(t) + \sum_{k=0}^{\infty} L_k \mathcal{V}_k(q)u(t) \\ &= Du(t) + \sum_{k=0}^{\infty} L_k \mathcal{X}_k(t) \\ &= Du(t) + \sum_{k=0}^{\infty} \mathcal{X}_k^T(t) L_k^T \end{aligned} \quad (\text{E.8})$$

where

$$\mathcal{X}_k(t) = [\phi_k(0) \quad \phi_k(1) \quad \cdots] \begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \end{bmatrix} \quad \phi_k(m) \in \mathbb{R}^{n_b \times 1} \quad (\text{E.9})$$

and where the (first part of the) ϕ_k can be calculated recursively as

$$\begin{aligned} \phi_0(m) &= \mathcal{V}_0(q)\delta(m+1) = q\mathcal{V}_0(q)\delta(m) \\ \phi_k(m) &= \mathbf{G}_b(q)\phi_{k-1}(m) \end{aligned} \quad (\text{E.10})$$

for $k, m = 0, 1, \dots$. State-space realizations of the transfer functions are

$$\begin{aligned} \mathcal{V}_0(z) &= (\mathbf{A}, \mathbf{B}, I, 0) \\ z\mathcal{V}_0(z) &= (\mathbf{A}, \mathbf{B}, \mathbf{A}, \mathbf{B}) \\ \mathbf{G}_b(z) &= (\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) \end{aligned}$$

The infinite orthonormal matrix $\Phi(\mathbf{G}_b)$

$$\Phi(\mathbf{G}_b) = \begin{bmatrix} \phi_0(0) & \phi_0(1) & \phi_0(2) & \cdots \\ \phi_1(0) & \phi_1(1) & \phi_1(2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (\text{E.11})$$

constitutes a complete orthonormal basis of the signal space $\ell_2[0, \infty)$.

E.2 Examples

Three examples of well known sets of orthonormal functions that are frequently used in the description of linear time invariant dynamical systems, and that occur as special cases of the above framework, are the pulse functions, the Laguerre functions, and the Kautz functions, see e.g. [64, 114, 51, 180]. We will illustrate this for the pulse and Laguerre functions, for the Kautz functions we refer to [81, 82].

Pulse functions

Consider the inner function $\mathbf{G}_b(z) = z^{-1}$, $\mathbf{G}_b \in \mathcal{G}_1$. The generating transfer functions $\mathcal{V}_k(z)$ satisfy $\mathcal{V}_k(z) = z^{-1}\mathbf{G}_b^k(z) = z^{-k-1}$, $k = 0, 1, \dots$. The corresponding set of basis functions $\Phi(\mathbf{G}_b)$ is determined by $\phi_k(t) = \delta(t - k)$ with $\delta(\tau)$ the Kronecker delta function. The inner function \mathbf{G}_b can be realized by the minimal balanced realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = (0, 1, 1, 0)$.

Laguerre functions

Consider the inner function $\mathbf{G}_b(z) = \frac{1-az}{z-a}$, with some $a \in \mathbb{R}$, $|a| < 1$, and denote $\nu = 1 - a^2$. Since $\mathbf{g}_b(0) = -a$ it is clear that $\mathbf{G}_b \in \mathcal{G}_1$. A minimal balanced realization of \mathbf{G}_b is given by $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) = (a, \sqrt{\nu}, \sqrt{\nu}, -a)$. It follows that

$$\begin{aligned} \phi_0(0) &= \sqrt{\nu} \\ \phi_k(0) &= -a\phi_{k-1}(0) \\ \phi_k(i+1) &= a\phi_k(i) + \nu \sum_{j=1}^k (-a)^{j-1} \phi_{k-j}(i) \end{aligned}$$

These equations exactly match the equations that generate the normalized discrete time Laguerre polynomials with discount factor a . The corresponding generating transfer functions $\mathcal{V}_k(z)$ are

$$\mathcal{V}_k(z) = \sqrt{\nu} \frac{(1-az)^k}{(z-a)^{k+1}}$$

which are indeed the generating transfer functions of discrete time Laguerre polynomials.

E.3 Merits

Using the system based orthonormal basis functions introduced by [80] for the system description has the following advantages.

- Any stable system can be used to generate a complete orthonormal basis for the signal space ℓ_2 .
- If the system and the basis generator are both stable a bound $|L_k| \leq \mathcal{K}\eta^k$ with $\eta < 1$ exists.
- A basis generating system that is closer to G_o yields a higher rate of convergence of the parameters to zero, i.e. a higher η in $|L_k| \leq \mathcal{K}\eta^k$.
- A generalized FIR model (E.7) combines an analytic solution to the least squares parameter estimation problem with a good low order approximation of the system and the advantages of an output error (OE) model.

Since any stable system can be used to generate the basis functions, the basis generator can be used to incorporate prior information about the system, as was pointed out by [81]. This prior information need not be correct, it may be uncertain or approximative. However, the more accurate the basis generating system, the more simple the system representation will be (faster convergence of the parameters to zero). The basis generator can e.g. be a nominal model, obtained by any identification technique or physical modelling. A control relevant nominal model therefore can be used as the basis generating system. Control relevant identification procedures are discussed in e.g. [181, 49, 106, 11, 12, 101, 149, 69, 157, 155, 191] and [5, 102, 103, 140, 48, 61, 175]. Note that a faster convergence of the parameters implies that a good approximate description of the system can be obtained using less parameters, so that both the bias and the variance of the estimates can be reduced. An additional merit of the procedure is that it is an extension to existing identification techniques. Any identification method can be used to obtain the basis generating system.

Bibliography

- [1] V.M. Adamyan, D.Z. Arov, and M.G. Krein. Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem. *Math. USSR Sbornik*, 15:31–73, 1971.
- [2] H. Akçay, G. Gu, and P.P. Khargonekar. Identification in H_∞ with non-uniformly spaced frequency response measurements. In *Proc. American Control Conf.*, pages 246–250, Chicago, IL, USA, 24-26 June 1992.
- [3] B.D.O. Anderson and R.L. Kosut. Adaptive robust control: on-line learning. In *Proc. 30th IEEE Conf. on Decision and Control*, pages 297–298, Brighton, England, 1991.
- [4] L.V.R. Arruda and G. Favier. A review and a comparison of robust estimation methods. In *Preprints IFAC/IFORS Symp. on Identification and System Parameter Estimation*, pages 1027–1032, Budapest, Hungary, 8-12 July 1991.
- [5] K.J. Åström. Matching criteria for control and identification. In *Proc. European Control Conf.*, pages 248–251, Groningen, The Netherlands, 28 June - 1 July 1993.
- [6] K.J. Åström and B. Wittenmark. *Adaptive Control*. Addison-Wesley, Reading, MA, USA, 1989.
- [7] G. Balas, J.C. Doyle, K. Glover, A.K. Packard, and R. Smith. *μ -Analysis and Synthesis Toolbox*. The MUSYN Inc., USA, 1991.
- [8] D.S. Bayard. Statistical plant set estimation using Schroeder-phased multisinusoidal input design. In *Proc. American Control Conf.*, pages 2988–2995, Chicago, IL, USA, 24-26 June 1992.
- [9] G. Belforte, B. Bona, and V. Cerone. Parameter estimation algorithms for set-membership description of uncertainty. *Automatica*, 26:887–898, 1990.
- [10] S.A. Billings and W.S.F. Voon. Correlation based model validity test for non-linear models. *Int. J. Control*, 44:235–244, 1986.

- [11] R.R. Bitmead, M. Gevers, and V. Wertz. *Adaptive Optimal Control - Thinking Man's GPC*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1990.
- [12] R.R. Bitmead and Z. Zang. An iterative identification and control strategy. In *Proc. European Control Conf.*, pages 1396–1400, Grenoble, France, 2-5 July 1991.
- [13] P.M.M. Bongers. On a new robust stability margin. In *Recent Advances in Mathematical Theory of Systems, Control Networks and Signal Processing I*, pages 377–382, Proc. 9th Int. Symp. on Mathematical Theory of Networks and Systems, Kobe, Japan, 17-21 June 1991. H. Kimura and S. Kodama (Eds.), Mita Press, Tokyo, Japan, 1992.
- [14] P.M.M. Bongers. Control design and implementation for a wind turbine system. In *Proc. European Control Conf.*, pages 1956–1961, Groningen, The Netherlands, 28 June - 1 July 1993.
- [15] P.M.M. Bongers. Duwecs reference guide v2.0. Technical Report MEMT 26, Delft University of Technology, The Netherlands, 1993.
- [16] P.M.M. Bongers. *Modeling and Identification of Flexible Wind Turbines and a Factorizational Approach to Robust Control*. PhD thesis, Delft University of Technology, The Netherlands, 1994.
- [17] P.M.M. Bongers and O.H. Bosgra. Low order robust H_∞ controller synthesis. In *Proc. 29th IEEE Conf. on Decision and Control*, pages 194–199, Honolulu, HI, USA, 1990.
- [18] P.M.M. Bongers, G.E. van Baars, and S.J. Dijkstra. Load reduction in a wind energy conversion system using an H_∞ controller. Internal Report N-422, Mechanical Engineering, Systems and Control Group, Delft University of Technology, The Netherlands, 1993.
- [19] G.E.P. Box and G.M. Jenkins. *Time Series Analysis, forecasting and control*. Holden-Day, San Francisco, CA, USA, 1970.
- [20] D.R. Brillinger. *Time Series. Data Analysis and Theory*. Holden-Day, San Francisco, CA, USA, expanded edition, 1981.
- [21] J. Chen, C.N. Nett, and M.K.H. Fan. Optimal nonparametric system identification from arbitrary corrupt finite time series: a control-oriented approach. In *Proc. American Control Conf.*, pages 279–285, Chicago, IL, USA, 24-26 June 1992.
- [22] J. Chen, C.N. Nett, and M.K.H. Fan. Worst-case system identification in H_∞ : validation of apriori information, essentially optimal algorithms, and error bounds. In *Proc. American Control Conf.*, pages 251–257, Chicago, IL, USA, 24-26 June 1992.

- [23] S. Chen, S.A. Billings, and P.M. Grant. Recursive hybrid algorithm for non-linear system identification using radial basis function networks. *Int. J. Systems Science*, 1992.
- [24] M.F. Cheung, S. Yurkovich, and K.M. Passion. An optimal volume ellipsoid algorithm for parameter set estimation. In *Proc. 30th IEEE Conf. on Decision and Control*, Brighton, England, 1991.
- [25] C.K. Chui. *Wavelet Analysis and its Applications*, volume I: An Introduction to Wavelets. Academic Press, Boston, USA, 1992.
- [26] P.L. Combettes. The foundations of set theoretic estimation. *Proc. of the IEEE*, 81:182–208, 1993.
- [27] M.A. Dahleh and M.H. Khammash. Controller design for plants with structured uncertainty. *Automatica*, 29:37–56, 1993.
- [28] M.A. Dahleh, E.D. Sontag, D.N.C. Tse, and J.N. Tsitsiklis. Worst-case identification of nonlinear fading memory systems. In *Proc. American Control Conf.*, pages 241–245, Chicago, IL, USA, 24–26 June 1992.
- [29] S. Dasgupta and Y.F. Huang. Asymptotically convergent modified recursive least-squares with data-dependent updating and forgetting factor for systems with bounded noise. *IEEE Trans. Information Theory*, IT-33:383–392, 1987.
- [30] R.A. De Callafon, P.M.J. Van den Hof, and D.K. De Vries. Identification and control of a compact disc mechanism using fractional representations. In *Preprints 10th IFAC Symp. on System Identification*, Copenhagen, Denmark, 4–6 July 1994.
- [31] R.A. De Callafon, P.M.J. Van den Hof, and M. Steinbuch. Control relevant identification of a compact disc pick-up mechanism. In *Proc. 32nd IEEE Conf. on Decision and Control*, pages 2050–2055, San Antonio, TX, USA, 15–17 December 1993.
- [32] D.K. De Vries. Identification for robust control: What is the best approach ? Technical Report MEMT 16, Delft University of Technology, The Netherlands, 1991.
- [33] D.K. De Vries, P.M.M. Bongers, and P.M.J. Van den Hof. Application of estimated error bounds in robust control of a wind energy conversion system. In *Preprints 10th IFAC Symp. on System Identification*, Copenhagen, Denmark, 4–6 July 1994.
- [34] D.K. De Vries and P.M.J. Van den Hof. Quantification of model uncertainty from data: input design, interpolation, and connection with robust control design

- specifications. In *Proc. American Control Conf.*, pages 3170–3175, Chicago, IL, USA, 24-26 June 1992.
- [35] D.K. De Vries and P.M.J. Van den Hof. Quantification of model uncertainty from data. *Int. J. Robust and Nonlinear Control*, 4:301–319, 1994.
- [36] D.K. De Vries and P.M.J. Van den Hof. Quantification of uncertainty in transfer function estimation: a mixed probabilistic – worst-case approach. *Accepted for publication in Automatica*, 1994.
- [37] J.R. Deller. Set membership identification in digital signal processing. *IEEE ASSP Magazine*, 6:4–20, 1989.
- [38] J.C. Doyle. Analysis of feedback systems with structured uncertainties. *Proc. IEE, Pt.D*, 129:242–250, 1982.
- [39] J.C. Doyle, K. Glover, P.P. Khargonekar, and B.A. Francis. State space solutions to standard H_2 and H_∞ control problems. *IEEE Trans. Automatic Control*, 34:831–846, 1989.
- [40] J.C. Doyle, K. Lenz, and A.K. Packard. Design examples using μ -synthesis: space-shuttle lateral axis FCS during reentry. In R.F. Curtain, editor, *Modelling, Robustness and Sensitivity Reduction in Control Systems*, pages 127–154. Springer Verlag, Berlin, 1987.
- [41] J.C. Doyle, J.E. Wall, and G. Stein. Performance and robustness analysis for structured uncertainty. In *Proc. 21st IEEE Conf. on Decision and Control*, pages 629–636, 1982.
- [42] W. Draijer, M. Steinbuch, and O.H. Bosgra. Adaptive control of the radial servo system of a compact disc player. *Automatica*, 28:455–462, 1992.
- [43] P. Eykhoff. *System Identification: Parameter and State Estimation*. Wiley, London, England, 1974.
- [44] E. Fogel. System identification via membership set constraints with energy constrained noise. *IEEE Trans. Automatic Control*, 24:752–758, 1979.
- [45] E. Fogel and Y.F. Huang. On the value of information in system identification – bounded noise case. *Automatica*, 18:229–238, 1982.
- [46] J.B. Garnett. *Bounded Analytic Functions*. Academic Press, New York, London, 1981.
- [47] T.T. Georgiou and M.C. Smith. Optimal robustness in the gap metric. *IEEE Trans. Automatic Control*, 35:673–686, 1990.

- [48] M. Gevers. Towards a joint design of identification and control ? In H.L. Trentelman and J.C. Willems, editors, *Essays on Control: Perspectives in the Theory and its Applications*, pages 111–151. Birkhäuser, Boston, MA, USA, 1993.
- [49] M. Gevers and L. Ljung. Optimal experiment design with respect to the intended model application. *Automatica*, 22:543–554, 1986.
- [50] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L_∞ -error bounds. *Int. J. Control*, 39:1115–1193, 1984.
- [51] K. Glover, J. Lam, and J.R. Partington. Rational approximation of a class of infinite-dimensional systems. i. singular values of hankel operators. *Mathematics of Control, Signals and Systems*, 3:325–344, 1990.
- [52] K. Glover and D. McFarlane. Robust stabilization of normalized coprime factor plant descriptions with H_∞ -bounded uncertainty. *IEEE Trans. Automatic Control*, 34:821–830, 1989.
- [53] G.H. Golub and C.F Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, USA, 2nd edition, 1989.
- [54] G.C. Goodwin, M. Gevers, and B. Ninness. Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Trans. Automatic Control*, 37:913–928, 1992.
- [55] G.C. Goodwin, D.Q. Mayne, and M.E. Salgado. Uncertainty, information and estimation. In *Proc. IFAC Symp. on Adaptive Control and Signal Processing*, pages 39–47, Glasgow, 1989.
- [56] G.C. Goodwin and B. Ninness. Model error quantification for robust control based on quasi-bayesian estimation in closed loop. In *Proc. American Control Conf.*, pages 77–82, Boston, MA, USA, 26-28 June 1991.
- [57] G.C. Goodwin, B. Ninness, and M.E. Salgado. Quantification of uncertainty in estimation. In *Proc. American Control Conf.*, pages 2400–2405, San Diego, USA, 23-25 May 1990.
- [58] G.C. Goodwin and R.L. Payne. *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, New York, USA, 1977.
- [59] G.C. Goodwin and M.E. Salgado. Quantification of uncertainty in estimation using an embedding principle. In *Proc. American Control Conf.*, pages 1416–1421, Pittsburgh, PA, USA, 21-23 June 1989.
- [60] G.C. Goodwin and M.E. Salgado. A stochastic embedding approach for quantifying uncertainty in the estimation of restricted complexity models. *Int. J. of Adaptive Control and Signal Processing*, 3:333–356, 1989.

- [61] S.F. Greabe and G.C. Goodwin. An incremental estimation technique for predicting a bandwidth of robust performance. In *Proc. 32nd IEEE Conf. on Decision and Control*, pages 2260–2265, San Antonio, TX, USA, 15-17 December 1993.
- [62] G. Gu and P.P. Khargonekar. A class of algorithms for identification in H_∞ . *Automatica*, 28:299–312, 1992.
- [63] G. Gu and P.P. Khargonekar. Linear and nonlinear algorithms for identification in H_∞ with error bounds. *IEEE Trans. Automatic Control*, 37:953–963, 1992.
- [64] G. Gu, P.P. Khargonekar, and E.B. Lee. Approximation of infinite-dimensional systems. *IEEE Trans. Automatic Control*, 34:610–618, 1989.
- [65] G. Gu, P.P. Khargonekar, and Y. Li. Robust convergence of two-stage nonlinear algorithms for identification in H_∞ . *Systems and Control Letters*, 18:253–263, 1992.
- [66] I. Gustavsson, L. Ljung, and T. Söderström. Identification of processes in closed loop – identifiability and accuracy aspects. *Automatica*, 13:59–75, 1977.
- [67] R.G. Hakvoort. Worst-case system identification in ℓ_1 : error bounds, optimal models, and model reduction. In *Proc. 31st IEEE Conf. on Decision and Control*, pages 499–504, Tucson, Arizona, USA, 16-18 December 1992.
- [68] R.G. Hakvoort. Worst-case system identification in H_∞ : error bounds, and optimal models. In *Proc. 12th IFAC World Congress*, pages VIII 161–164, Sydney, Australia, 18-23 July 1993.
- [69] R.G. Hakvoort, R.J.P. Schrama, and P.M.J. Van den Hof. Approximate identification in view of LQG feedback design. In *Proc. American Control Conf.*, pages 2824–2828, Chicago, IL, USA, 24-26 June 1992.
- [70] R.G. Hakvoort, P.M.J. Van den Hof, and O.H. Bosgra. Consistent parameter bounding identification using cross-covariance constraints on the noise. In *Proc. 32nd IEEE Conf. on Decision and Control*, pages 2601–2606, San Antonio, TX, USA, 15-17 December 1993.
- [71] F.R. Hansen. *A Fractional Representation Approach to Closed-Loop System Identification and Experiment Design*. PhD thesis, Stanford University, Stanford, CA, USA, 1989.
- [72] F.R. Hansen and G.F. Franklin. On a fractional representation approach to closed-loop experiment design. In *Proc. American Control Conf.*, pages 1319–1320, Atlanta, GA, USA, 15-17 June 1988.

- [73] F.R. Hansen, G.F. Franklin, and R. Kosut. Closed-loop identification via the fractional representation: experiment design. In *Proc. American Control Conf.*, pages 1422–1427, Pittsburgh, PA, USA, 21-23 June 1989.
- [74] A.J. Helmicki, C.A. Jacobson, and C.N. Nett. H_∞ identification of stable LSI systems: a scheme with direct application to controller design. In *Proc. American Control Conf.*, pages 1428–1434, Pittsburgh, PA, USA, 21-23 June 1989.
- [75] A.J. Helmicki, C.A. Jacobson, and C.N. Nett. Identification in H_∞ : a robustly convergent, nonlinear algorithm. In *Proc. American Control Conf.*, pages 386–391, San Diego, USA, 23-25 May 1990.
- [76] A.J. Helmicki, C.A. Jacobson, and C.N. Nett. Identification in H_∞ : linear algorithms. In *Proc. American Control Conf.*, pages 2418–2423, San Diego, USA, 23-25 May 1990.
- [77] A.J. Helmicki, C.A. Jacobson, and C.N. Nett. On zero-order hold equivalents of distributed parameter systems. In *Proc. American Control Conf.*, San Diego, USA, 23-25 May 1990.
- [78] A.J. Helmicki, C.A. Jacobson, and C.N. Nett. Control oriented system identification: a worst-case/deterministic approach in H_∞ . *IEEE Trans. Automatic Control*, 36:1163–1176, 1991.
- [79] A.J. Helmicki, C.A. Jacobson, and C.N. Nett. Least squares methods for H_∞ control-oriented system identification. *IEEE Trans. Automatic Control*, 38:819–826, 1993.
- [80] P.S.C. Heuberger. *On Approximate System Identification with System Based Orthonormal Functions*. PhD thesis, Delft University of Technology, The Netherlands, 1991.
- [81] P.S.C. Heuberger, P.M.J. Van den Hof, and O.H. Bosgra. A generalized orthonormal basis for linear dynamical systems. Internal Report N-404, Mechanical Engineering, Systems and Control Group, Delft University of Technology, The Netherlands, 1992, Revised 1993.
- [82] P.S.C. Heuberger, P.M.J. Van den Hof, and O.H. Bosgra. A generalized orthonormal basis for linear dynamical systems. In *Proc. 32nd IEEE Conf. on Decision and Control*, pages 2850–2855, San Antonio, TX, USA, 15-17 December 1993.
- [83] H. Hjalmarsson. *Aspects on Incomplete Modeling in System Identification*. PhD thesis, Linköping University, Linköping, Sweden, 1993.
- [84] H. Hjalmarsson. Detecting asymptotically non-vanishing model uncertainty. In *Proc. 12th IFAC World Congress*, Sydney, Australia, 18-23 July 1993.

- [85] H. Hjalmarsson and L. Ljung. Estimating model variance in the case of under-modeling. *IEEE Trans. Automatic Control*, 37:1004–1008, 1992.
- [86] G.M. Jenkins and D.G. Watts. *Spectral Analysis and its Applications*. Holden-Day, San Francisco, CA, USA, 1968.
- [87] R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall International, 2nd edition, 1988.
- [88] K.J. Keesman. *A Set-Membership Approach to the Identification and Prediction of Ill-Defined Systems: Application to a Water Quality System*. PhD thesis, Twente University of Technology, The Netherlands, 1989.
- [89] K.J. Keesman and G. Van Straten. Embedding of random scanning and principal component analysis in set-theoretic approach to parameter estimation. In *Proc. 12th IMACS World Congress*, 1988.
- [90] M.G. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume 2. Griffin, London, England, 1961.
- [91] M.H. Khammash and J.B. Pearson. Performance robustness of discrete-time systems with structured uncertainty. *IEEE Trans. Automatic Control*, 36:398–412, 1991.
- [92] R.L. Kosut. Adaptive control via parameter set estimation. *Int. J. of Adaptive Control and Signal Processing*, 2:371–400, 1988.
- [93] R.L. Kosut. Workshop on the Modeling of Uncertainty in Control Systems, Santa Barbara, 18-20 June 1992. Personal communication.
- [94] R.L. Kosut, M. Lau, and S. Boyd. Set-membership identification of systems with parametric and nonparametric uncertainty. *IEEE Trans. Automatic Control*, 37:929–941, 1992.
- [95] J.M. Krause and P.P. Khargonekar. Parametric identification in the presence of non-parametric dynamic uncertainty. *Automatica*, 26:113–123, 1990.
- [96] J.M. Krause and P.P. Khargonekar. Robust parameter adjustment with non-parametric weighted-ball-in- H_∞ uncertainty. *IEEE Trans. Automatic Control*, 35:225–229, 1990.
- [97] R.O. LaMaire. *Robust Time and Frequency Domain Estimation Methods in Adaptive Control*. PhD thesis, Dept. Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1987.
- [98] R.O. LaMaire, L. Valavani, M. Athans, and G. Stein. A frequency-domain estimator for use in adaptive control systems. *Automatica*, 27:23–38, 1991.

- [99] M.K. Lau, S. Boyd, R.L. Kosut, and G.F. Franklin. Robust control design for ellipsoidal plant set. In *Proc. 30th IEEE Conf. on Decision and Control*, pages 291–296, Brighton, England, 1991.
- [100] M.K. Lau, S. Boyd, R.L. Kosut, and G.F. Franklin. A robust control design for FIR plants with parameter set uncertainty. In *Proc. American Control Conf.*, pages 83–88, Boston, MA, USA, 26–28 June 1991.
- [101] W.S. Lee, B.D.O. Anderson, R.L. Kosut, and I.M.Y. Mareels. On adaptive robust control and control-relevant system identification. In *Proc. American Control Conf.*, pages 2834–2841, Chicago, IL, USA, 24–26 June 1992.
- [102] W.S. Lee, B.D.O. Anderson, R.L. Kosut, and I.M.Y. Mareels. A new approach to adaptive robust control. *Int. J. Adaptive Control and Signal Processing*, 7:183–211, 1993.
- [103] W.S. Lee, B.D.O. Anderson, R.L. Kosut, and I.M.Y. Mareels. On robust performance improvement through the windsurfer approach to adaptive robust control. In *Proc. 32nd IEEE Conf. on Decision and Control*, pages 2821–2827, San Antonio, TX, USA, 15–17 December 1993.
- [104] L. Ljung. Asymptotic variance expressions for identified black-box transfer function models. *IEEE Trans. Automatic Control*, 30:834–844, 1985.
- [105] L. Ljung. On the estimation of transfer functions. *Automatica*, 21:677–696, 1985.
- [106] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.
- [107] L. Ljung. Perspectives on the process of identification. In *Proc. 12th IFAC World Congress*, pages V 197–205, Sydney, Australia, 18–23 July 1993.
- [108] L. Ljung and P.E. Caines. Asymptotic normality of prediction error estimation for approximate system models. *Stochastics*, 3:29–46, 1979.
- [109] L. Ljung, J. Sjöberg, and T. McKelvey. On the use of regularization in system identification. In *Proc. 12th IFAC World Congress*, pages VII 381–386, Sydney, Australia, 18–23 July 1993.
- [110] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. M.I.T. Press, Cambridge, MA, USA, 1983.
- [111] L. Ljung and B. Wahlberg. Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra. *Advances in Applied Probability*, 24:412–440, 1992.

- [112] L. Ljung, B. Wahlberg, and H. Hjalmarsson. Model quality: the roles of prior knowledge and data information. In *Proc. 30th IEEE Conf. on Decision and Control*, pages 273–278, Brighton, England, 1991.
- [113] L. Ljung and Z.D. Yuan. Asymptotic properties of black-box identification of transfer functions. *IEEE Trans. Automatic Control*, 30:514–530, 1985.
- [114] P.M. Mäkilä. Approximation of stable systems by Laguerre filters. *Automatica*, 26:333–345, 1990.
- [115] P.M. Mäkilä. On H_∞ identification of stable systems and optimal approximation. *Automatica*, 27:663–676, 1991.
- [116] P.M. Mäkilä. Robust identification and Galois sequences. *Int. J. Control*, 54:1189–1200, 1991.
- [117] P.M. Mäkilä. Worst-case input-output identification. *Int. J. Control*, 56:673–689, 1992.
- [118] S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.
- [119] B.B. Mandelbrot and J.W.V. Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437, 1968.
- [120] D. McFarlane and K. Glover. *Robust Controller Design using Normalized Coprime Factor Plant Descriptions*. Springer Verlag, Berlin, Germany, 1990.
- [121] D. McFarlane, K. Glover, and M. Vidyasagar. Reduced-order controller design using coprime factor model reduction. *IEEE Trans. Automatic Control*, 35:369–373, 1990.
- [122] M. Milanese and R. Tempo. Optimal algorithms theory for robust estimation and prediction. *IEEE Trans. Automatic Control*, 30:864–867, 1985.
- [123] M. Milanese and A. Vicino. Estimation theory for nonlinear models and set membership uncertainty. *Automatica*, 27:403–408, 1991.
- [124] M. Milanese and A. Vicino. Optimal approximation of feasible parameter set in estimation with unknown but bounded noise. In *Preprints IFAC/IFORS Symp. on Identification and System Parameter Estimation*, pages 1365–1369, Budapest, Hungary, 8–12 July 1991.
- [125] M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, 27:997–1009, 1991.

- [126] S.H. Mo and J.P. Norton. Recursive parameter-bounding algorithms which compute polytope bounds. In *Proc. 12th IMACS World Congress*, Paris, France, 1988.
- [127] E.A. Moody. *Truth and Consequence in Mediaeval Logic*. North-Holland Publishing Company, Amsterdam, The Netherlands, 1953.
- [128] B.C. Moore. Principal component analysis in linear systems: controllability, observability and model reduction. *IEEE Trans. Automatic Control*, 26:17–32, 1981.
- [129] M.P. Newlin and R.S. Smith. Model validation and a generalization of μ . In *Proc. 30th IEEE Conf. on Decision and Control*, Brighton, England, 1991.
- [130] B.M. Ninness. *Stochastic and Deterministic Modelling*. PhD thesis, University of Newcastle, Newcastle, New South Wales, Australia, 1993.
- [131] J.P. Norton. *An Introduction to Identification*. Academic Press, London, England, 1986.
- [132] J.P. Norton. Identification and application of bounded-parameter models. *Automatica*, 23:497–507, 1987.
- [133] J.P. Norton. Identification of parameter bounds for ARMAX models from records with bounded noise. *Int. J. Control*, 45:375–390, 1987.
- [134] J.P. Norton and S.H. Mo. Parameter bounding for time varying systems. *Mathematics and Computers in Simulation*, 32:527–534, 1990.
- [135] W. of Ockham. *Summa Totius Logicae*. Venice, Italy, 1508, Obtainable from the Franciscan Institute, St. Bonaventure, NY, USA.
- [136] A.V. Oppenheim and R.W. Schaffer. *Digital Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1975.
- [137] A.V. Oppenheim and R.W. Schaffer. *Discrete-Time Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.
- [138] A. Packard and J.C. Doyle. The complex structured singular value. *Automatica*, 29:71–109, 1993.
- [139] P.J. Parker and R.R. Bitmead. Approximation of stable and unstable systems via frequency response identification. In *Proc. 10th IFAC World Congress*, pages 358–363, 1987.

- [140] A.G. Partanen and R.R. Bitmead. Two stage iterative identification/controller design and direct experimental controller refinement. In *Proc. 32nd IEEE Conf. on Decision and Control*, pages 2833–2838, San Antonio, TX, USA, 15-17 December 1993.
- [141] J.R. Partington. Robust identification and interpolation in H_∞ . *Int. J. Control*, 54:1281–1290, 1991.
- [142] J.R. Partington. Robust identification in H_∞ . *J. Mathematical Analysis and Application*, 166:428–441, 1991.
- [143] H. Piet-Lahanier and E. Walter. Practical implementation of an exact and recursive algorithm for characterizing likelihood sets. In *Proc. 12th IMACS World Congress*, Paris, France, 1988.
- [144] H. Piet-Lahanier and E. Walter. Bounded-error tracking of time-varying parameters. In *Proc. 31st IEEE Conf. on Decision and Control*, pages 66–67, Tucson, Arizona, USA, 16-18 December 1992.
- [145] D. Pollard. *Convergence of Stochastic Processes*. Springer Verlag, Berlin, Germany, 1984.
- [146] K. Poolla, P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal. A time-domain approach to model validation. In *Proc. American Control Conf.*, pages 313–317, Chicago, IL, USA, 24-26 June 1992.
- [147] M.B. Priestley. *Spectral Analysis and Time Series*. Academic Press, London, England, 1981.
- [148] L. Pronzato and E. Walter. Robustness to outliers of bounded-error estimators, consequences on experiment design. In *Preprints IFAC/IFORS Symp. on Identification and System Parameter Estimation*, pages 821–826, Budapest, Hungary, 8-12 July 1991.
- [149] D.E. Rivera, J.F. Pollard, and C.E. García. Control-relevant prefiltering: a systematic design approach and case study. *IEEE Trans. Automatic Control*, 37:964–974, 1992.
- [150] A. Sard. *Linear Approximation*. American Mathematical Society, 1963.
- [151] J. Schoukens, P. Guillaume, and R. Pintelon. Design of multisine excitations. In *Proc. IEE Control 91 Conf.*, Edinburgh, 25-28 March 1991.
- [152] J. Schoukens and R. Pintelon. *Identification of Linear Systems. A Practical Guideline to Accurate Modeling*. Pergamon Press, 1991.

- [153] R.J.P. Schrama. Control-oriented approximate closed-loop identification via fractional representations. In *Proc. American Control Conf.*, pages 719–720, Boston, MA, USA, 26-28 June 1991.
- [154] R.J.P. Schrama. An open-loop solution to the approximate closed-loop identification problem. In *Preprints 9th IFAC/IFORS Symp. on Identification and System Parameter Estimation*, pages 1602–1607, Budapest, Hungary, 8-12 July 1991.
- [155] R.J.P. Schrama. Accurate identification for control: the necessity of an iterative scheme. *IEEE Trans. Automatic Control*, 37:991–994, 1992.
- [156] R.J.P. Schrama. *Approximate Identification and Control Design, with application to a mechanical system*. PhD thesis, Delft University of Technology, The Netherlands, 1992.
- [157] R.J.P. Schrama and P.M.J. Van den Hof. An iterative scheme for identification and control design based on coprime factorizations. In *Proc. American Control Conf.*, pages 2842–2846, Chicago, IL, USA, 24-26 June 1992.
- [158] M.R. Schroeder. Synthesis of low peak-factor signals and binary sequences of low auto-correlation. *IEEE Trans. Information Theory*, 16:85–89, 1970.
- [159] F.C. Schwegge. *Uncertain Dynamic Systems*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1973.
- [160] F.C. Schwegge. Recursive state estimation: unknown but bounded errors and system inputs. *IEEE Trans. Automatic Control*, 13:22–28, 1986.
- [161] R.S. Smith. Model validation and parameter identification for systems in H_∞ and ℓ_1 . In *Proc. American Control Conf.*, pages 2852–2856, Chicago, IL, USA, 24-26 June 1992.
- [162] R.S. Smith and J.C. Doyle. Model invalidation: a connection between robust control and identification. In *Proc. American Control Conf.*, pages 1435–1440, Pittsburgh, PA, USA, 21-23 June 1989.
- [163] R.S. Smith and J.C. Doyle. Towards a methodology for robust parameter identification. In *Proc. American Control Conf.*, pages 2394–2399, San Diego, USA, 23-25 May 1990.
- [164] R.S. Smith and J.C. Doyle. Model invalidation: a connection between robust control and identification. *IEEE Trans. Automatic Control*, 37:942–952, 1992.
- [165] T. Söderström, L. Ljung, and I. Gustavsson. Identifiability conditions for linear multivariable systems operating under feedback. *IEEE Trans. Automatic Control*, 21:837–840, 1976.

- [166] T. Söderström and P. Stoica. *System Identification*. Prentice-Hall, 1989.
- [167] M. Steinbuch, G. Schootstra, and O.H. Bosgra. Robust control of a compact disc player. In *Proc. 31st IEEE Conf. on Decision and Control*, pages 2596–2600, Tucson, Arizona, USA, 16-18 December 1992.
- [168] R. Tempo. Robust estimation and filtering in the presence of bounded noise. *IEEE Trans. Automatic Control*, 33:864–867, 1988.
- [169] T.J.J. Van den Boom. System identification for coprime factor plant description with bounded error. In *Proc. American Control Conf.*, pages 1248–1252, Chicago, IL, USA, 24-26 June 1992.
- [170] T.J.J. Van den Boom. *MIMO System Identification for H_∞ Robust Control, A Frequency Domain Approach with Minimum Error Bounds*. PhD thesis, Eindhoven University of Technology, The Netherlands, 1993.
- [171] T.J.J. Van den Boom. MIMO-systems identification by minimum error bounds. In *Proc. 12th IFAC World Congress*, Sydney, Australia, 18-23 July 1993.
- [172] T.J.J. Van den Boom, M. Klompstra, and A. Daamen. System identification for H_∞ -robust control design. In *Preprints 9th IFAC/IFORS Symp. on Identification and System Parameter Estimation*, pages 1431–1436, Budapest, Hungary, 8-12 July 1991.
- [173] P.M.J. Van den Hof, D.K. De Vries, and P. Schoen. On relaxed delay structure conditions in closed loop identifiability problems. *Automatica*, 28:1047–1050, 1992.
- [174] P.M.J. Van den Hof and P.H.M. Janssen. Some asymptotic properties of multi-variable models identified by equation error techniques. *IEEE Trans. Automatic Control*, 32:89–92, 1987.
- [175] P.M.J. Van den Hof, R.J.P. Schrama, O.H. Bosgra, and R.A. De Callafon. Identification of normalized coprime plant factors for iterative model and controller enhancement. In *Proc. 32nd IEEE Conf. on Decision and Control*, pages 2839–2844, San Antonio, TX, USA, 15-17 December 1993.
- [176] T.G. van Engelen, P.M.M. Bongers, J.T.G. Pierik, and G.E. van Baars. Integral control for damping of mechanical resonance and protection against grid failures: results in a full load operating condition. In *Proc. European Community Wind Energy Conf.*, pages 545–550, 1993.
- [177] S.M. Veres and J.P. Norton. Structure selection for bounded parameter models: consistency conditions and selection criterion. *IEEE Trans. Automatic Control*, 36:474–481, 1991.

- [178] M. Vidyasagar. The graph metric for unstable plants and robustness estimates for feedback stability. *IEEE Trans. Automatic Control*, 29:403–418, 1984.
- [179] M. Vidyasagar. *Control System Synthesis: A Factorization Approach*. M.I.T. Press, Cambridge, MA, USA, 1985.
- [180] B. Wahlberg. System identification using Laguerre models. *IEEE Trans. Automatic Control*, 36:551–562, 1991.
- [181] B. Wahlberg and L. Ljung. Design variables for bias distribution in transfer function estimation. *IEEE Trans. Automatic Control*, 31:134–144, 1986.
- [182] B. Wahlberg and L. Ljung. Hard frequency-domain model error bounds from least-squares like identification techniques. *IEEE Trans. Automatic Control*, 37:900–912, 1992.
- [183] E. Walter and H. Piet-Lahanier. Robust nonlinear parameter estimation in the bounded noise case. In *Proc. 25th IEEE Conf. on Decision and Control*, Athens, Greece, 10–12 December 1986.
- [184] E. Walter and H. Piet-Lahanier. Estimation of parameter bounds from bounded error data: a survey. *Mathematics and Computers in Simulation*, 32:449–468, 1990.
- [185] J.C. Willems. Paradigms and puzzles in the theory of dynamical systems. *IEEE Trans. Automatic Control*, 26:259–294, 1991.
- [186] H.S. Witsenhausen. Sets of possible states of linear systems given perturbed observations. *IEEE Trans. Automatic Control*, 13:556–558, 1968.
- [187] R.C. Younce and C.E. Rohrs. Identification with nonparametric uncertainty. *IEEE Trans. Automatic Control*, 37:715–728, 1992.
- [188] N.J. Young. *An Introduction to Hilbert Space*. University Press, Cambridge, 1988.
- [189] Z.D. Yuan and L. Ljung. Black-box identification of multivariable transfer functions: asymptotic properties and optimal input design. *Int. J. Control*, 40:233–256, 1984.
- [190] G. Zames. Functional analysis applied to nonlinear feedback systems. *IEEE Trans. Circuit Theory*, CT-10:392–404, 1963.
- [191] Z. Zang, R.R. Bitmead, and M. Gevers. Disturbance rejection: on-line refinement of controllers by closed loop modelling. In *Proc. American Control Conf.*, pages 2829–2833, Chicago, IL, USA, 24–26 June 1992.

-
- [192] Y.C. Zhu. Estimation of transfer functions: asymptotic theory and a bound of model uncertainty. *Int. J. Control*, 49:2241–2258, 1989.
- [193] Y.C. Zhu. *Identification and Control of MIMO Industrial Processes: An Integration Approach*. PhD thesis, Dept. Electrical Engineering, Eindhoven University of Technology, The Netherlands, 1990.
- [194] Y.C. Zhu and A.C.P.M. Backx. MIMO process identification for controller design: test signals, nominal model and error bounds. In *Preprints 9th IFAC/IFORS Symp. on Identification and System Parameter Estimation*, pages 1202–1207, Budapest, Hungary, 8–12 July 1991.

Glossary of symbols

Special symbols

$\arg \min_x f(x)$	Value of x that minimizes $f(x)$
$\text{diag}(a_1, \dots, a_n)$	Diagonal $n \times n$ matrix A with $A_{ii} = a_i$ and $A_{ik} = 0$ for $i \neq k$
$\dim(x)$	Dimension of a vector x
D	Direct feedthrough matrix in a state-space description of the system G_o , 78, 243
D	Coprime factor of a model, 140
D_o	Coprime factor of the system, 140, 182
e	Base of the natural logarithm
$e(t)$	Noise generating process, 78, 145
$\mathbb{E}[x]$	Expectation of x
F_i	Error due to the noise over the time interval T_i , $i = 1, 2, \dots, r$, 87
F	Error due to the noise, sample mean of F_i , 147
$F_\alpha(n, d)$	Probability that $x \leq \alpha$ when $x \in F(n, d)$, 83
g_o	Impulse response of the system G_o , 4, 54, 78, 145
G	Model of G_o
G_o	System, 4, 54, 78, 144
\hat{G}_i	Empirical transfer function estimate of G_o over the time interval T_i , $i = 1, 2, \dots, r$, 80, 144
\hat{G}	Nonparametric estimate of G_o , sample mean of \hat{G}_i , 80, 144, 147
\hat{G}^s	Empirical transfer function estimate of G_o over time interval $T_{N_s}^N$, 56
\hat{G}_i^f	Parametric estimate of G_o over the time interval T_i , $i = 1, 2, \dots, r$, 91
\hat{G}^f	Parametric estimate of G_o , sample mean of \hat{G}_i^f , 91

G_b	Basis generating system, 243, 243, 245
\mathbf{G}_b	Inner function obtained from G_b , 145, 243, 243, 245
H_o	Noise filter, 78
I	Identity matrix
I_n	$n \times n$ Identity matrix
$\text{Im}\{x\}$	Imaginary part of x
j	$\sqrt{-1}$
\mathcal{K}, η	Bound on the generalized expansion coefficients of the system G_o , $ L_k \leq \mathcal{K}\eta^k$, 79
L_k	Generalized expansion coefficients, generalized impulse response, 78, 145, 244
M, ρ	Bound on the impulse response of the system G_o , $ g_o(k) \leq M\rho^{-k}$, 55, 79
n_b	McMillan degree of basis generating system G_b , 243
\bar{n}	$(n - 1)/n_b$
N	Number of measured data points, number of data points used for identification, 54, 78, 145
N	Coprime factor of a model, 140
N_o	Period length of the input signal, 56, 78, 145
N_o	Coprime factor of the system, 140, 182, 145
N_s	Length of the time interval used to reduce the effects of unknown past inputs, 54, 56, 78, 145
$P[x]$	Probability of the event x
q	Shift operator, $qx(t) = x(t + 1)$, $q^{-1}x(t) = x(t - 1)$
r	Number of data segments, number of averages, 78, 80, 87, 145
$r(t)$	External excitation signal, 139, 181
$\text{rank}(A)$	Rank of matrix A
$\text{Re}\{x\}$	Real part of x
S_i	Error due to unknown past inputs over the time interval T_i , $i = 1, 2, \dots, r$, 80, 81
S	Error due to unknown past inputs, sample mean of S_i , 80, 81, 147
t	Time index, $t \in \mathbb{Z}$
τ	Time index, $t \in \mathbb{Z}$
T^N	Integer interval $[0, 1, \dots, N - 1]$, 54, 78, 145

$T_{N_s}^N$	Integer interval $[N_s, N_s + 1, \dots, N + N_s - 1]$, 54, 78
T_i	Integer interval $[(i - 1)N_o + N_s, (i - 1)N_o + N_s + 1, \dots, N + (i - 1)N_o + N_s - 1]$, 78
$u(t)$	Input signal to the system G_o , 4, 30, 54, 78, 144
$u_a(\tau)$	Input signal to the system G_o , 98, 102
\bar{u}	Maximum of $ u(t) $ over $t \in T^{N+N_s}$, 54, 79
\bar{u}^p	Bound on $ u(t) $ for $t < 0$, 55, 79
\bar{u}_a^p	Bound on $ u_a(\tau) $ for $\tau < 0$, 98
U, U_i, U^s	Discrete Fourier transforms of the input signal $u(t)$ over different time intervals, 54, 78, 145
$v(t)$	Disturbance signal, additive output noise, 4, 30, 54, 78, 145
V, V_i, V^s	Discrete Fourier transforms of the additive noise $v(t)$ over different time intervals, 54, 78, 145
\bar{V}^s	Bound on $ V^s $, 55
\mathcal{V}_k	Orthonormal basis functions, 78, 145, 243, 228
W	Weighting function in the weighted least squares estimate of the parameter vector $\hat{\theta}$, 87, 89, 152
\mathcal{W}	Weighting function in the parametric transfer function estimate \hat{G}^f , 91, 95
\mathcal{W}_y	Weighting function in the estimate of the undisturbed output signal $\hat{y}(\tau)$, and the DFT of the disturbed output signal $\hat{Y}^q(e^{j\zeta\epsilon})$, 98, 103
$y(t)$	Disturbed output signal of the system G_o , 4, 30, 54, 78, 144
$y_o(\tau)$	Undisturbed output signal of the system G_o , 99
$\hat{y}(\tau)$	Estimate of $y_o(\tau)$, 98
Y, Y_i, Y^s	Discrete Fourier transforms of the disturbed output signal $y(t)$ over different time intervals, 54, 78, 145
Y_o^q	Discrete Fourier transform of $y_o(\tau)$, 103
Y^q	Discrete Fourier transform of $y(\tau)$, 103
\hat{Y}^q	Estimate of Y_o^q , 103
z	Complex variable, $z \in \mathbb{C}$
Z	Undermodelling error, 87, 147
$\delta(k)$	Kronecker delta, $\delta(k) = 1$ for $k = 0$ and $\delta(k) = 0$ for $k \neq 0$
η, \mathcal{K}	Bound on the generalized expansion coefficients of the system G_o , $ L_k \leq \mathcal{K}\eta^k$, 79

θ	Parameter vector
θ_o	True parameter vector, 87, 147
$\hat{\theta}_i$	Estimated parameter vector over the time interval T_i , $i = 1, 2, \dots, r$, 87
$\hat{\theta}$	Estimated parameter vector, sample mean of $\hat{\theta}_i$, 87, 148
ρ, M	Bound on the impulse response of the system G_o , $ g_o(k) \leq M\rho^{-k}$, 55, 79
$\bar{\sigma}(A)$	Maximum singular value of matrix A
$\sigma^2(x)$	Variance of x
$\hat{\sigma}_r^2(x)$	Estimated variance of x , sample variance of x_i , $i = 1, 2, \dots, r$, divided by r , 82, 88
$\sum_{k=a}^b x(k)$	Summation with convention that the sum is zero if $b < a$
$\Phi_u(\omega)$	Power spectrum of the input signal $u(t)$, 146
$\Phi_v(\omega)$	Power spectrum of the noise process $v(t)$
Ψ	Matrix in (weighted) least squares estimate, 87, 148
ω	Frequency variable, $\omega \in [0, 2\pi)$
ω_k	Frequency index, element of a set containing a finite number of frequency points, 59, 79, 80, 147
Ω_{N_o}	Set of frequency points, 79, 146
$\Omega_{N_o}^u$	Set of frequency points, 59, 79, 80, 147
\rightarrow	Tends to
\xrightarrow{d}	Converges in distribution to
$:=$	Is defined as
\in	Is an element of; Is distributed as
\equiv	Is identically equal to

Distributions

$F(n, d)$	Fisher distribution with n degrees of freedom in the numerator and d degrees of freedom in the denominator
$F_\alpha(n, d)$	Probability that $x \leq \alpha$ when $x \in F(n, d)$, 83
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with expectation μ and variance σ^2
$x \in As\mathcal{N}(\mu, \sigma^2)$	The random variable x asymptotically is distributed as $\mathcal{N}(\mu, \sigma^2)$

$\chi^2(d)$ χ^2 distribution with d degrees of freedom

Vectors and matrices

$A\langle i \rangle$	Element i of vector A
$A\langle i, k \rangle$	Element of matrix A at row i and column k
A_{ik}	Element of matrix A at row i and column k
A^T	Transpose of A
A^*	Complex conjugate transpose of A
$A \geq 0$	Matrix A is positive semi-definite
$A \geq B$	Matrix $A - B$ is positive semi-definite
$ A $	Vector or matrix consisting of the absolute values or magnitudes of each element of the vector or matrix A , $ A _{ik} = A_{ik} $, <i>see also</i> $ a $

Norms

$ a $	$\sqrt{a^*a}$ for scalar a , <i>see also</i> $ A $
$\ A\ _2$	Euclidian norm for vector A , $\bar{\sigma}(A)$ for matrix A
$\ A(z)\ _\infty$	$\sup_{\omega} \bar{\sigma}(A(e^{j\omega}))$

Sets

\mathbb{C}	Set of complex numbers
\mathbb{C}^p	Set of p dimensional complex vectors
$\mathbb{C}^{p \times q}$	Set of $p \times q$ complex matrices
\mathbb{N}	Set of natural numbers, $\mathbb{Z} \cap [0, \infty)$
\mathbb{R}	Set of real numbers
\mathbb{R}^p	Set of p dimensional real vectors
$\mathbb{R}^{p \times q}$	Set of $p \times q$ real matrices
\mathbb{Z}	Set of integers
H_2	Set of all complex valued functions which are analytic outside the unit disc and squared integrable over the unit circle
$H_2^{p \times q}$	Set of all $p \times q$ complex valued matrix functions which are analytic outside the unit disc and squared integrable over the unit circle

H_∞	Set of all complex valued functions which are analytic and uniformly bounded outside the unit disc and uniformly bounded on the unit circle
ℓ_1	Set of all absolute summable sequences
ℓ_2	Set of all squared summable sequences
$\ell_2[0, \infty)$	Set of all squared summable sequences on the time interval \mathbb{N}
$\ell_2^{p \times q}[0, \infty)$	Set of all $p \times q$ squared summable sequences on the time interval \mathbb{N}
ℓ_∞	Set of all uniformly bounded sequences
L_2	Set of all complex valued functions which are squared integrable on the unit circle
L_∞	Set of all complex valued functions which are essentially bounded on the unit circle
IRH_2	Set of all real rational functions in H_2
IRH_∞	Set of all real rational functions in H_∞

Acronyms and abbreviations

ARX	Auto Regressive with eXogenous inputs
ARMAX	Auto Regressive Moving Average with eXogenous inputs
DFT	Discrete Fourier Transform
ETFE	Empirical Transfer Function Estimate
FIR	Finite Impulse Response
LTI	Linear Time Invariant
MIMO	Multi Input Multi Output
generalized FIR	Finite generalized Impulse Response, a linear combination of a finite number of generalized orthonormal basis functions, 85, 86, 244
PRBS	Pseudo Random Binary Sequence
SISO	Single Input Single Output
SIMO	Single Input Multi Output
w.p.	with probability

Samenvatting

Dit proefschrift behandelt het probleem van het kwantificeren van modelonzekerheid op grond van experimentele data. Het doel is om op basis van gemeten in- en uitgangssignalen een bovengrens te bepalen voor een relevante norm op het verschil tussen het echte systeem en een geschikt model van dat systeem. Specifieker gesteld richt dit proefschrift zich op de vraag hoe modelfouten gerepresenteerd en gekwantificeerd dienen te worden, teneinde een identifikatieresultaat te verkrijgen dat geschikt is om te dienen als basis voor het ontwerp van een robuuste regelaar met hoge prestaties.

Een kritische evaluatie van de betreffende literatuur laat zien dat een aanzienlijk deel van dit probleem wordt verschoven naar informatie die verondersteld wordt vooraf al aanwezig te zijn, terwijl geen bevredigende procedures bestaan om deze informatie te verkrijgen.

Een eerste aanpak om het probleem op te lossen maakt gebruik van een harde bovengrens op de ruis in het frekwentiedomein als voorkennis, en resulteert in een harde bovengrens op de modelonzekerheid. Eerst wordt een modelonzekerheidsgrens in een eindig aantal frekwentiepunten vastgesteld, waarna deze grens wordt geïnterpoleerd. De verdienste van deze methode is dat het model waarop de onzekerheidsgrens betrekking heeft direct gespecificeerd kan worden door de gebruiker. Een belangrijke nadeel van de methode is echter dat de vereiste harde voorkennis moeilijk verkrijgbaar is.

Gebaseerd op een uitvoerige inventarisatie van mogelijkheden en konsekventies wordt nu voorgesteld om de modelfouten in te bedden in een mengvorm van fouten die uitmiddelen en fouten die verondersteld worden op de meest ongunstige manier te kunnen doorwerken, om accumulaties van zowel toevallige als structurele fouten effectief te kunnen beschrijven in een situatie waar alleen het dominante deel van het systeem lineair en tijdinvariant is. Beargumenteerd wordt dat ruis gezien dient te worden als een stochastisch proces, terwijl ondermodellering beschouwd dient te worden als onbekend maar begrensd.

Gegeven deze keuzen voor het inbedden van de modelonzekerheden wordt vervolgens een aangepaste procedure gepresenteerd voor het schatten van een model tezamen met een grens op de modelonzekerheid vanuit experimentele data. Hierbij wordt gebruik gemaakt van een periodiek ingangssignaal en datasegmentatie om effectief onderscheid te kunnen maken tussen toevallige en structurele fouten. Over iedere periode

van het ingangssignaal wordt een model geschat, hetgeen een verzameling van modellen oplevert. Het gemiddelde over deze verzameling van modellen levert het uiteindelijke model, terwijl de onderlinge verschillen tussen de modellen in de verzameling informatie geven over de onzekerheid in dit uiteindelijke model. Voor het geval van lineaire tijdinvariante systemen met stochastische ruis kan nu, asymptotisch in het aantal datapunten, een betrouwbaarheidsinterval in gesloten vorm verkregen worden voor de onzekerheid ten gevolge van de ruis, en kan een grens op de onzekerheid ten gevolge van ondermodellering worden geschat. In essentie is de enige benodigde voorkennis bij deze aanpak de kwalitatieve informatie dat de ruis zijn eigen verleden vergeet, en dat het systeem lineair tijdinvariant en exponentieel stabiel is. Zowel de onzekerheid ten gevolge van de ruis, als de onzekerheid ten gevolge van ondermodellering worden begrensd vanuit de experimentele data. Bovendien wordt gebruikt gemaakt van een zeer flexibele modelstructuur die gevormd wordt door systeem-gebaseerde orthonormale basisfuncties, zodat parametrische modellen van een hoge kwaliteit geschat kunnen worden, en kan dezelfde aanpak ook in een gesloten lus situatie worden gevolgd. Op basis van simulatievoorbeelden en toepassingen in de praktijk zijn de prestaties van deze procedure bestudeerd, ook onder afwijkingen van het lineaire tijdinvariante geval met stochastische ruis. De geschatte grenzen op de modelonzekerheid blijken strak en betrouwbaar, en robuuste regelaars met een hoge prestatie kunnen worden ontworpen. We konkluderen dat de combinatie van de bovenstaande eigenschappen inderdaad resulteert in modellen en modelonzekerheids grenzen die geschikt zijn om te dienen als basis voor het ontwerp van robuuste regelaars met een hoge prestatie.

Met behulp van het bovenstaande raamwerk kunnen tevens duidelijke indicaties gegeven worden voor de selectie van de modelorde en het ontwerp van het ingangssignaal. Ook resulteert een krachtige methode voor de validatie van modellen. Bovendien wordt een asymptotische analyse gegeven voor frekwentiedomein identifikatie van overdrachtsfuncties die beschreven zijn met behulp van systeem-gebaseerde orthonormale basisfuncties, zoals toegepast in de bovenstaande procedure voor het schatten van grenzen op de modelonzekerheid.

Een belangrijk onderwerp voor verder onderzoek is het aanpassen van de procedure zodat kwasi-stationaire ingangssignalen toegestaan zijn, hetgeen zeker mogelijk lijkt.

Curriculum Vitae

November 3, 1964	Born in Geldrop, The Netherlands
1977–1983	Secondary school: Christelijke Scholengemeenschap Melanchthon, Rotterdam, The Netherlands
1983–1989	Master student at the Delft University of Technology, Faculty of Mechanical Engineering, Delft, The Netherlands; with
1987–1989	Master project at the Philips Research Laboratories, Eindhoven, The Netherlands
1989–1994	Ph.D. research at the Mechanical Engineering Systems and Control Group of the Delft University of Technology, Delft, The Netherlands
April 1994	Military service; defence research at TNO (Netherlands Organization for Applied Scientific Research), Physics and Electronics Laboratory, Den Haag, The Netherlands