Data-driven modeling and complexity reduction for nonlinear systems with stability guarantees

Fahim Shakib

Data-driven modeling and complexity reduction for nonlinear systems with stability guarantees

Fahim Shakib

The research reported in this thesis is part of the research program of the Dutch Institute of Systems and Control (DISC). The author has successfully completed the educational program of the Graduate School DISC.

The work described in this thesis was carried out at the Eindhoven University of Technology.

A catalogue record is available from the Eindhoven University of Technology Library.
ISBN: 978-90-386-5598-7
DOI: 10.6100/d135c018-64ac-46dc-8132-453a3398d19a

Typeset by the author using the pdf LaTeX documentation system.
Cover design: Fons-Jan Luijten & Fahim Shakib
Reproduction: Ipskamp printing, Enschede, the Netherlands

Data-driven modeling and complexity reduction for nonlinear systems with stability guarantees

Proefschrift

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor
Promoties, in het openbaar te verdedigen
op woensdag 23 november 2022 om 16.00 uur

door

Mohammad Fahim Shakib

geboren te Herat, Afghanistan

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. H. Nijmeijer
1e promotor:	prof.dr.ir. N. van de Wouw
2e promotor:	prof.dr. A. Pavlov
co-promotor:	dr. A. Y. Pogromsky
leden:	prof.dr. A. Astolfi (Imperial College London)
	prof.dr.ir. P. M. J. van den Hof
	prof.dr.ir. J. M. A. Scherpen (Rijksuniversiteit Groningen)
	prof.dr.ir. J. Schoukens (Vrije Universiteit Brussel)

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Summary

Data-driven modeling and complexity reduction for nonlinear systems with stability guarantees

Dynamic models of real-life systems are crucial in many engineering applications. Such models enable the prediction of the system response, analysis of the system's dynamic behavior, and controller design. System identification is a data-driven method to construct a model whose simulated response closely matches experimentally obtained data from the system under study. Alternatively, accurate models can be retrieved by first-principle modeling. As first-principle models are often prohibitively complex for the above-mentioned purposes, there is a need for sufficiently accurate but less complex models. Model reduction methods can be used to construct such models of lower complexity based on first-principle models. A great arsenal of techniques exists for both system identification and model reduction of linear models. However, real-life systems often exhibit nonlinear behavior. Hence, predictive nonlinear models are needed. The contributions of this thesis can be split into three parts, of which two parts concern the system identification of nonlinear models and one part concerns the model reduction of nonlinear models.

The first part of the thesis concerns the problem of system identification of nonlinear models. The following two challenges are addressed for this problem. Firstly, even if the nonlinear system exhibits stability properties, it is not guaranteed that the identified model preserves these properties. Consequently, the identified nonlinear model can perform well on the training data but may produce unbounded responses for new inputs. Secondly, the identification problem often boils down to solving a nonlinear optimization problem. Solving such a problem requires the computation of many model responses. Consequently, the computation time of the identification algorithms depends highly on the efficiency of the computation of model responses. This efficiency is critical in time-restrictive applications.

An identification approach for a class of nonlinear models that consist of linear time-invariant (LTI) dynamics placed in feedback with static nonlinearities is developed in this thesis. This approach enforces a strong form of global stability, namely exponential convergence. The convergence property ensures the global exponential stability of time-varying steady-state model responses. Furthermore, a simulation algorithm is devised that significantly improves the computation time from hours to minutes in numerical case studies. The proposed algorithm is analyzed in terms of the time-varying steady-state mismatch between the system and model response. This mismatch is shown to asymptotically drop to zero by increasing the computational expense of the algorithm.

The identification approach is validated experimentally on a mechanical ventilation setup. The results evidence fast and accurate identification of hose and patient lung parameters. These identified parameters can be used for control purposes and the decision-making process of medical personnel operating the mechanical ventilator. Furthermore, the proposed identification approach is validated on benchmark datasets. These results show a trade-off between data fit and model stability. The identification approach thus identifies predictive models, enforces model stability, and is computationally efficient, enabling the fast identification of accurate and stable nonlinear models in time-restrictive applications.

The second part of this thesis concerns the problem of order reduction for nonlinear models. Model order reduction aims to find a sufficiently accurate reduced-order model that replaces the high-order model while preserving key properties. In this problem, it is challenging to preserve model stability, which is among the most important model properties. Furthermore, it is challenging to bound the mismatch between the high-order and reduced-order nonlinear models. Moreover, it is challenging to find the optimal reduced-order model that minimizes the mismatch between the reduced-order and the high-order nonlinear model.

Two challenges for LTI models are addressed as a preliminary step towards dealing with the challenges for nonlinear models. The specific reduction technique employed in this thesis is the time-domain moment matching method. Firstly, the obtained matching property of moment matching in the multivariable LTI case is established. Secondly, additional moment matching parameters are exploited in the LTI case to enforce model stability and to ensure optimality between the high-order and reduced-order models. These results are further exploited to derive an error bound for the mismatch between the reduced-order and the high-order LTI models.

An optimal model order reduction technique for nonlinear models that consist of LTI dynamics placed in feedback with nonlinear dynamics is proposed to address the challenges in the nonlinear case. Hereto, the time-domain moment matching framework is extended to nonlinear models that enjoy a special type of global exponential stability, namely exponential convergence. This extension leads to a natural application of moment matching to this class of nonlinear models. Reduced-order models found by this reduction method preserve the convergence

property and are optimal in terms of a minimized error bound. The proposed methods are validated on various examples, resulting in accurate reduced-order models that preserve the key stability property.

The third part of this thesis concerns the identification of nonlinear models using a reproducing kernel Hilbert space approach. Such a black-box learning method exploits training data to identify models that can predict responses of the original system for new inputs. Despite offering great flexibility, these models lack transparency, which makes it challenging to enforce additional model properties, such as stability. Moreover, a straightforward application of pre-existing methods often yields statistically biased models in the closed-loop case.

To deal with the first challenge, an identification method that enforces exponential convergence on the identified model is proposed. Using benchmark data, it is shown that the nonlinear behavior can be accurately captured while model stability is enforced, whereas models identified using pre-existing methods show instabilities on new inputs. Furthermore, in the closed-loop case, a statistically consistent method for LTI systems is extended to the kernel-based nonlinear case, providing unbiased identification results in numerical examples.

In summary, this thesis presents fundamental approaches for the data-driven modeling and complexity reduction of nonlinear models with stability guarantees that are experimentally and numerically validated.

Contents

1 Introduction

Mathematical modeling is an enabling tool for the design and analysis of engineering systems. This thesis focuses on the data-driven modeling of nonlinear systems and the complexity reduction of nonlinear models, both with a focus on certified stability guarantees. Such stability guarantees enable the safe and generalized usage of models in unseen scenarios.

This chapter starts with the motivation for the modeling of dynamical systems and the concept of a mathematical model in Section 1.1. After that, in Section 1.2, the stability notion of convergence is introduced which guarantees stability and enables generalized usage of mathematical models. Sections 1.3 and 1.4 present state-of-the-art methods for data-driven modeling and complexity reduction, respectively, together with open challenges in these fields. These challenges are embedded in the objectives of this thesis presented in Section 1.5. Section 1.6 describes the contributions of the thesis toward these objectives. Finally, Section 1.7 gives the outline of the remainder of this thesis.

1.1 Mathematical models

Humankind has always tried to understand the surrounding world. This understanding happens through models that we learn and update via interaction with our environment. For a long time, these models were qualitative, predicting outcomes based on our observations, e.g., water flows from a higher elevation to a lower elevation, snow is lingering so it must be freezing, and the trees are moving so it must be windy. Ever since the development of the disciplines of physics and mathematics, quantitative models described by mathematical equations complemented qualitative ones. Such mathematical models provide a deeper understanding of the universe around us and play a central role in science, technology, and society in the modern world. For example, in our solar system, different planets interact with each other and orbit in ways that are now accurately captured by mathematical models. Closer to our everyday use, mathematical modeling has enabled,

Figure 1.1. A surgeon performs robot-assisted surgery using the da Vinci system (figure reprinted from https://www.davincisurgery.com/da-vinci-systems/about-da-vinci-systems).

Figure 1.2. A SpaceX Falcon 9 rocket being prepared for launch at NASA's Kennedy Space Center in Florida, USA (figure reprinted from https://www.nasa.gov/image-feature/spacex-falcon-9-rolls-out-for-saturday-launch).

for example, accurate weather forecasting, which is instrumental for our everyday choices. Mathematical models are thus indispensable to humankind.

In engineering systems, models are, e.g., used for model-based system and control design. From the very start of such design procedures, models allow for predicting the system's future behavior. Based on such predictions, the system performance can be analyzed in a fast way without the need for prototyping, controller deployment, or time-consuming experimentation. Subsequently, system design, control design, or both, can be improved, leading to enhanced performance with fast design cycles. Such engineering systems have been an enabler of technological breakthroughs. For example, ever since 2010, da Vinci medical surgical robots (see Figure 1.1) extend the eye-hand capabilities of surgeons beyond what the human body allows [301]. This has led to higher medical success rates, less patient discomfort, and faster patient recovery [52]. Another example is Flight 20 of Falcon 9, conducted on December 22nd, 2015, which was the first successful vertical landing of the first stage of the Falcon 9 rocket [245] (see Figure 1.2). This milestone achievement is part of the developments toward cost-reduced rocket systems, to be used for launching payloads into orbit and, potentially, for a spacefaring civilization [35]. A third example is the first observation of gravitational waves on September 14th, 2015, using the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) gravitational-wave detector systems [2] (see Figure 1.3). Arguably, this is one of the biggest accomplishments of humankind and marks an exciting new era of astronomy. Like these, there are uncountably many examples of the success of model-based engineering. Without *model-based* system and control design, these achievements would not have been possible.

A model is not the true description of the real-life system, but rather an ab-

Figure 1.3. The Advanced LIGO system in Livingston, USA (left), detected the first gravitational wave, which was subsequently, about seven-thousandths of a second later, confirmed by the Advanced LIGO system in Hanford, USA (right) (figure reprinted from `https://www.ligo.caltech.edu`).

straction that attempts to explain the system's behavior with sufficient accuracy. Here, systems are understood as real-life processes in which different aspects captured in variables interact and produce signals that we observe and call system outputs. Furthermore, systems can often be affected by external stimuli called system inputs, which can sometimes be chosen by us. The interaction between system variables can depend on their history, in which case the system exhibits memory and is called *dynamic*. The ability of a model to make accurate predictions in specific scenarios is an important accuracy measure. However, the usefulness of a model also depends on its ability to predict system outputs for new system inputs and the simplicity of its description. It is thus natural to have different models for the same system, none being true or false, but some being more instrumental for the intended purpose. For example, almost any system exhibits some *nonlinear* behavior. However, a *linear* model for such a nonlinear system, being less accurate but significantly easier to construct and use than a *nonlinear* one, may suffice depending on the intended purpose.

Models can be constructed using first-principle laws such as Newton's laws of motion, conservation of energy laws, thermodynamic laws, and electrostatic laws, to name a few. However, engineering systems are becoming increasingly complex and modeling accuracy requirements are becoming increasingly stringent. Therefore, the derived first-principle models are often too complex for model-based system and control design. For example, testing such a complex model for a new input can be computationally demanding and sometimes even infeasible due to limited computational and data storage capabilities as well as time restrictions. To reduce the computational expense, the full model is often replaced by a reduced model that is less complex but still sufficiently accurate. Besides the reduction in

computational demands, such a reduction step has several other benefits, namely: it facilitates controller design and deployment, and it gives deeper insights into the essence of the dynamics of the full model. In several fields, such as mathematics, mechanics, and systems and control, the practice of finding a reduced model is called *model reduction*.

The increase in system complexity and accuracy requirements often also results in first-principle modeling becoming a too difficult or too time-consuming task. At this stage, it is very tempting to pursue a *data-driven* modeling approach, if the system under study is already realized. Such an approach uses observed inputs and outputs from the system to directly construct compact models that optimally describe this data. Therefore, data-driven modeling is typically fast and enables system modeling if first-principle modeling is infeasible. For example, data-driven modeling provides basic understanding of our brain function, whereas first-principle modeling of our brain function is considered a challenging task [22, 71]. In the systems and control field, this practice of turning datasets into mathematical models is called *system identification*.

In both the data-driven modeling and complexity reduction problems, the identified or reduced model is a virtual representation of the true system. Such models are nowadays used in a large variety of applications, including safety-critical applications such as autonomous driving [165, 318], aircraft design and control [267], surgical robotics [167], clinical decision support [93, 212], and many more. Given the complexity of these applications, the constructed models are in many cases also complex and analysis of their properties is challenging. A popular *static* example with undesired model properties is observed in image classification models, where small perturbations in the visual input (invisible to the human eye), can result in a completely different classification [76, 273, 314]. In such a case, the constructed model does not generalize well to scenarios unseen during the modeling and/or complexity reduction processes. In recent years, there has been an urge to construct models with safe generalization capabilities [104], especially important for safety-critical applications and also for modeling *dynamic* systems [43, 47, 232]. Therefore, constructed models need to be both *predictive* and safely *generalizable* to new inputs. Sometimes predictive capabilities of models are sacrificed to enable safe generalization to new inputs [104].

One way to enhance generalization capabilities is by enforcing additional properties on the identified or reduced model. For example, consider the situation where the output of the system or full model remains more-or-less the same for a small change in the input signal. In such a situation, the sensitivity of the system output with respect to input variations is small. Then, besides finding a predictive model, the found model should also exhibit additional (stability) properties to guarantee such small sensitivity.

Modeling accuracy requirements are becoming increasingly stringent and real-life systems often display some level of nonlinear behavior. Therefore, such nonlinear system characteristics need to be modeled too to be able to predict essential

nonlinear behavior. Consequently, tools for data-driven nonlinear modeling of nonlinear systems and complexity reduction of nonlinear models are essential to facilitate future model-based system and control designs. Furthermore, in the nonlinear context, the call for stability guarantees is even more pronounced because nonlinear models can exhibit complex stability properties. The development of *data-driven nonlinear modeling* and *nonlinear complexity reduction* tools with *stability guarantees* forms the scope of this thesis.

1.2 Convergent nonlinear state-space models

The observed output of a dynamic system depends not only on the current observed input but also on the past observed inputs and outputs. Therefore, dynamic models exhibit a memory of their past. One of the most popular model structures for capturing dynamic behavior is the state-space model structure. State-space models relate observed input variables to observed outputs via latent state variables. The state variables contain all the memory that the model has of its past.

A linear dynamic model lives on a linear (high-dimensional) hyperplane relating its state update linearly to the current state and input. Consequently, in the discrete-time case, the state at the next time step is a linear function of its current state and input. Thanks to linearity, a linear model exhibits the same properties for any input and any state, i.e., its properties hold in a *global* sense. Powerful methods for analysis and design of linear (control) systems are based on the *superposition principle*. This principle states that the response caused by the sum of multiple inputs equals the sum of responses caused by each individual input. This is, for example, exploited in frequency domain analysis and the design of linear control systems [257].

Although easy to use, a linear model is often insufficient to explain nonlinear dynamics. In that case, one has to rely on nonlinear modeling. However, in contrast to a linear model, *local* properties of nonlinear models do generally not hold *globally* and neither does the *superposition principle* hold. Therefore, seeing the model output in a certain part of the input and state space is not sufficient to predict the model output in a different part of the input and state space. For example, for some classes of models, it is yet unclear how many fixed-points the model exhibits [26], let alone finding closed-form expressions for the locations of such fixed points. For such a model, predicting the model output for new inputs, for example constant ones, is a challenging task. Hence, in general, nonlinear models are difficult to generalize toward unseen scenarios.

Models that exhibit certain stability properties can also generalize more reliably to new inputs. Many different stability notions are introduced in the literature for nonlinear models *with inputs*, e.g., input-to-state stability [263], incremental stability [83], contraction analysis [158], and *convergence* [66, 309]. A comparative study on these stability notions can be found in, e.g., [219, 282]. *Convergent* mod-

els exhibit a uniquely defined bounded time-varying *steady-state* solution for any bounded time-varying input. Furthermore, this steady-state solution is globally asymptotically stable (GAS), meaning that all solutions converge to this steady-state solution in forward time, for any initial condition. Hence, for any (unseen) bounded input, convergent dynamics guarantee the following qualitatively predictable behavior of model outputs, namely: (i) model outputs remain bounded; (ii) model outputs corresponding to different initial conditions converge to the steady-state output determined only by the input; and, under minor additional assumptions, (iii) steady-state outputs corresponding to small changes in the input remain close to each other. For these reasons, convergent models are known to exhibit *nonlinear simplicity* [201], not observed in general nonlinear models. It makes convergent models especially useful in solving many analysis and design problems. As this thesis concerns the problem of *model construction* via data-based modeling and complexity reduction techniques, the convergence property can be *enforced* during the model construction via data-driven or model reduction techniques.

In data-driven modeling, enforcing the convergence property is relevant in two cases. If the system under study is convergent, e.g., inferred from experiments, then enforcing convergence is motivated from a property preservation point of view. Alternatively, if it is unknown whether the system is convergent, enforcing convergence might be desired to enhance generalization to new, unseen inputs. In complexity reduction problems, enforcing convergence is motivated from a property preservation point of view [31]. In such a problem, if the full model is known to exhibit convergence, then convergence is desired to be preserved for the reduced model as well.

1.3 Data-driven modeling with stability guarantees

Data-driven modeling or *system identification* is a technique to construct mathematical models for systems using experimental data from the system. The identification cycle is explained in Section 1.3.1. After that, a literature review on differences between methods for nonlinear system identification is presented in Section 1.3.2 on a high abstraction level. More detailed and focused literature surveys are included in the introductory sections of the later chapters.

1.3.1 The four-step identification cycle

The identification cycle first introduced by [312] in the 1950s was a three-step procedure consisting of data collection, model structure selection, and estimation. Later, a validation step was added, complementing the four-step system identification cycle that is used nowadays [152]. The first step amounts to collecting data from the system under study into a dataset, which acts as an information source for the remaining steps. Hereto, identification experiments are conducted with

the real-life system where the input and measured output signals are recorded. The second step is to choose a model class. Such class is described by the relationship between inputs and outputs, e.g., linear vs. nonlinear, and the order of the difference or differential equations that describe the system. The third step is the estimation step, which amounts to choosing the model from the model class that best explains the data according to some criterion. In practice, this comes down to solving a problem that optimizes for the selected error criterion. Finally, the fourth step is model validation. Such a validation tests whether the obtained model is appropriate for its intended use. For example, this can be based on the model's ability to explain new datasets that are experimentally obtained from the system and that were not used in the estimation step, i.e., a cross-validation test. If the model does not pass the validation test, some of the previous steps should be revised and repeated, until an appropriate model is found.

The identification of *linear* systems is a well-established field with many books discussing the field [18, 40, 75, 152, 205, 261]. In some cases, depending on the purpose of the model, a linear model can be sufficient for a nonlinear system. Hereto, the methods in [242] characterize nonlinear distortions, enabling the decision-making process regarding modeling in the linear or nonlinear context. However, dominant *nonlinear* behavior appears in many engineering systems, making the identification problem far more challenging [240]. For example, a linear model class with a sufficiently large state dimension is sufficiently rich to describe the dynamics observed during the identification experiment on linear systems. However, in the nonlinear case, there is no universal model class that is sufficiently rich to capture any system, inevitably resulting in structural modeling errors. Sometimes, structural modeling errors are allowed to avoid overly complex models that are useless for the intended purpose, such as, e.g., control design. Another challenge in the nonlinear case is the handling of process noise, which is nonlinearly distorted in the output measurements, making it challenging to derive uncertainty bounds. A large number of model classes and dedicated identification approaches for nonlinear systems have been proposed in the literature taking into account these challenges. Presenting a complete overview of all these approaches is a daunting task beyond the scope of this thesis. For this reason, the remainder of this section reviews specific model classes relevant to this thesis and particularly relevant to the problem of *data-driven modeling with stability guarantees*. The survey paper [240] is a good starting point for a more complete overview of the nonlinear system identification field.

The model structure of a state-space model is characterized by the functions that describe the state update and output mappings. The amount and type of information used to describe these mappings decide the *shade of gray* of the model [240]. On the one end of the spectrum, a white-box model carefully uses first-principle laws for the parametrization of the mappings. Such a model provides a deep understanding of the laws governing the model equations. On the other end of the spectrum, a black-box modeling approach typically describes the mappings by

a basis function expansion or uses *function estimation* techniques from the field of machine learning. Such black-box models have great expressive capabilities and require few user choices, however, these models lack *interpretability* due to their complicated nature. Between these two ends, several gray shades exist that combine white-box and black-box modeling [240].

Having a model structure at hand, the model parameters should be estimated based on the collected input and output data according to some criteria. The most used criteria are the prediction error minimization (PEM) criterion and the simulation error minimization (SEM) criterion. The former criterion predicts new outputs based on *previous input and outputs*, whereas the latter predicts new outputs *solely on the basis of past inputs*. Construction via the PEM criterion is easier and the constructed models exhibit accurate short-term predictions, however, long-term predictions are typically less accurate. Construction via the SEM criterion is typically more challenging, however, these models do typically exhibit accurate long-term predictions. The PEM or SEM criteria typically lead to a nonlinear optimization problem in the model parameters, which is non-trivial to solve. Often, iterative, gradient-based optimization routines are employed, which, starting from an initial model, only find the best nearby (locally) optimal model. Initial models can often efficiently be found using the best linear approximation (BLA) framework [235, 243]. In general, in data-driven state-space modeling, there is no guarantee that the best model is found.

Given the iterative nature of such identification schemes, many model responses are required to be computed. Despite the fact that most nonlinear systems evolve in continuous-time, there has been a focus on discrete-time nonlinear modeling, motivated by computationally cheaper discrete-time model simulations [236]. In general, model simulation of continuous-time nonlinear models is a difficult task, and assessing the quality of the model simulation is difficult [252]. In the system identification problem, in particular, accurate model simulations are required to distinguish between structural modeling errors and errors due to the numerical simulation algorithm. A unifying framework to analyze errors between sampled system responses and computed model responses of continuous-time nonlinear models is lacking in the literature.

1.3.2 State-of-the-art nonlinear system identification approaches

The class of block-oriented models have been a subject of study for a long time [96, 175, 188, 234, 243, 244, 291, 294, 296, 297]. These models consist of an interconnection of several separate blocks, each exhibiting either linear dynamics or nonlinear statics. Examples include the Wiener and Hammerstein model structures, as well as the nonlinear feedback model structure. Block-oriented models arise naturally in problems with linear dynamics subject to localized nonlinearities, such as nonlinear springs and dampers in mechanical systems, and nonlinear resistors in electrical circuits. Hence, this knowledge of the model structure can be

maintained by block-oriented models using the separation between linear dynamics and nonlinear statics. Furthermore, the BLA framework can be naturally applied for initialization [243] of these models. Although relatively simple, even block-oriented models, especially those with nonlinear feedback, can show instability issues in practice.

The class of polynomial state-space (PNLSS) models is used for black-box modeling in [189, 190] with a range of successful applications [19, 57, 63, 107, 160, 183, 189, 190]. PNLSS models are often initialized by the BLA method and are characterized by monomial basis functions of the state and input, whose parametrization quickly grows for large state orders or large polynomial orders. To address this, polynomial decoupling methods [61, 62, 69, 74] were introduced to reduce the number of parameters. The reduction process replaces multivariate functions with univariate ones, enabling visual inspection of the dominant nonlinearities. Alternatively, [24] proposed tensor networks for modeling a special case of PNLSS models, thereby avoiding an excessively large number of parameters to start with and facilitating faster identification cycles. Although showing excellent expressive capabilities, PNLSS models are prone to instabilities when used for unseen scenarios [60, 61]. To address this challenge, [64, 292] propose multiple shooting methods that can scope through unstable regions of the parameter space. An additional benefit of multiple shooting is the smoothing of the underlying optimization problem, resulting in less risk of getting stuck in local minima [216, 217]. However, the identification of PNLSS with formal stability guarantees is still an open problem.

In light of *enforcing model stability*, another class of black-box modeling approaches related to the PNLSS modeling have been proposed by [38, 159, 280, 281, 284] based on the usage of *convex* tools. These methods enforce *global incremental stability* on the identified models, which is a stability notion for models with input, close to the notion of convergence. The basis functions used for these models include polynomials and trigonometric functions. However, these methods require the availability of a (surrogate) state sequence in addition to inputs and outputs, which can be hard to obtain in practice.

Methods from the *machine learning* community have also been proposed for the identification of nonlinear systems. Examples are the so-called kernel-based methods [54, 202, 203], which include approaches like regularization networks [58], Tychonov regularization problems [97], support vector machines [269], Gaussian process regression [29, 82, 140, 141, 209], and kriging [164]. Under certain conditions, such kernel-based estimators can be shown to be *universal approximators* [58, 202, 266], i.e., can approximate continuous functions arbitrarily well, and enable bias-variance tradeoff via regularization. A particularly well-known class of models is Gaussian process state-space models (GPSSMs) [26, 29, 73, 81, 82, 285]. There are some identification methods for GPSSMs that enforce stability properties on the identified model. For example, [79, 137, 138, 307] present methods that guarantee *global asymptotic stability* of the origin in the *autonomous* case. The method in [255] enforces the identified *autonomous* GPSSM to be locally

contracting, whereas [36] enforces global contraction. Alternatively [285] focuses on a particular kernel function and identifies *autonomous* GPSSMs that exhibit bounded solutions. These methods are tailored to the problem of imitation learning [117], where autonomous models are sufficient and also full-state measurements are available. Enforcing stability on *non-autonomous* GPSSMs, or in the more generic case of kernel-based state-space models, is a challenging and open problem.

A method to estimate latent state data from input-output observations directly is proposed by [298]. Such a state estimate can be further exploited to identify the state-space mapping, such as in [218] for linear parameter varying (LPV) models. This method closely relates to subspace identification methods for linear systems, where the latent state estimation via *canonical correlation analysis* (CCA) [114] is also split from the estimation of the linear model matrices, see, e.g., [147, 293]. However, a statistical analysis of the method proposed by [298] used in an identification setting, such as [218], is lacking in the literature.

A more general class of machine learning methods consists of artificial neural networks [148, 231], such as deep neural network (DNN) models, long short-term memory (LSTM) models, and recurrent neural network (RNN) models. These methods have been explored in the scope of system identification in [80, 155, 166, 176, 256, 272]. For example, [27, 163] uses neural networks to identify encoder functions for latent state estimation and estimation of the dynamics. Despite being universal approximators [23, 194] and, therefore, highly expressive, these models can show extreme sensitivity to inputs changes [53, 273, 299], making them less suitable for generalization toward new inputs. To address the sensitivity to input variations, early work on recurrent neural networks imposes the origin to be a GAS fixed point [271] or imposes local stability with a guaranteed basin of attraction [270]. The method [172] identifies contractive RNN models using constraints on the weights of the network model; [142] identifies *autonomous* DNN models whose origin is globally exponentially stable; and [37] identifies input-to-state stable LSTM models using constraints on the weights of the network model.

More recently, [214] proposed a convex parameterization of incrementally stable models with a prescribed incremental gain using linear matrix inequality (LMI) constraints. Subsequently, [215] proposed a method that imposes contractivity and user-defined integral quadratic constraints (IQCs) on a large class of state-space neural network models, including DNN and RNN models. It is highlighted that the latter, i.e., [215], imposes these properties in an *unconstrained* fashion, making the approach also applicable to large-scale neural network architectures. Moreover, the IQCs enable further, user-defined model robustness guarantees to new inputs. The complexity of neural network models and the number of parameters for their description is controlled by the depth and width of the network layers. The methods in [65, 303] aim at decoupling the neural network as a collection of univariate functions, reducing the number of parameters and enabling visual inspection of these functions.

To summarize, the field of nonlinear data-driven modeling faces the challenges

of (i) constructing expressive classes of models to capture a large class of nonlinear systems; (ii) enforcing model stability to enhance generalization toward unseen, new inputs, especially for safety-critical applications; and (iii) enabling computationally efficient identification for time-restrictive applications, also for continuous-time modeling.

1.4 Complexity reduction with stability preservation

Complexity reduction aims at finding accurate reduced-complexity models for complex high-fidelity models. Section 1.4.1 provides a scope of complexity reduction methods in the systems and control field. After that, Section 1.4.2 reviews some state-of-the-art reduction approaches and points out their technical differences. More detailed and focused literature surveys are included in the introductory sections of the later chapters.

1.4.1 The large scope of complexity reduction

Given the ever-increasing system complexity and modeling requirements, dynamic models of today's engineering systems are often too complex for model-based system and control design. Here, *complexity* is understood as the number of coupled first-order ordinary differential equations (ODEs), which arise naturally by first-principle modeling (and after spatial and temporal discretization of partial differential equations (PDEs)). For highly complex models, model simulation for the prediction of the system output for new inputs can become an infeasible task due to storage limitations, finite computational power, and time restrictions. Furthermore, the complexity of the designed controller by many model-based control techniques, such as the linear quadratic regular (LQR) [6] and $\mathcal{H}_\infty$ control, is of the same complexity as the used model. This can cause serious limitations in control design for large-scale models because computational accuracy reduces due to poor numerical conditioning [184] and due to approximation steps [77, 254]. Moreover, real-time controller deployment can become challenging, e.g., in model-predictive control [192], due to limited storage capabilities and computational power. For these reasons, complex models are often replaced by simpler ones. Recalling from [10], *complexity reduction* or *model reduction* aims at using computationally efficient tools to find less complex models that are accurate with respect to the full one and preserve additional properties such as stability. Furthermore, error bounds, bounding the error of the reduced model with respect to the original one, are instrumental to provide a direct measure of the quality of the reduced model. For example, for linear systems, some methods provide error bounds in the infinity norm of the error transfer-function matrix, giving worst-case error bounds.

In the nonlinear case, both stability preservation and error bounds are crucial. The former describes the qualitative behavior of the reduced model, such

as boundedness and stability of solutions for the convergence property, see Section 1.2. The latter quantitatively describes the mismatch between the full and the reduced model, thereby being a quality measure. A third aspect in the nonlinear context is model structure preservation, which can provide model interpretability. Such preservation enables the analysis and interpretation of individual model blocks and their connection.

1.4.2 State-of-the-art complexity reduction approaches

Model reduction is a well-established subject in the mathematics and systems and control fields [10]. A large array of methods exist for the complexity reduction of linear models, which can roughly be categorized into singular-value-decomposition (SVD) methods and Krylov-projector methods. Approaches in [102, 264] attempted to merge the SVD and Krylov-projector methods.

The SVD methods for linear models include balanced truncation methods [103] and Hankel norm approximation methods [98] together with their extensions to preserve additional properties such as performance criteria [185] and passivity [67, 108]. These methods transform the full model into a balanced state-space realization in which the states are ordered in terms of their observability and controllability properties. From here, difficult to observe and control states are identified and truncated, i.e., omitted from the reduced model. The SVD methods are equipped with *a priori* error bounds and preserve asymptotic stability.

Being instrumental for linear time-invariant (LTI) models, the SVD methods have found extensions to the nonlinear case [84–86, 133, 230]. Analog to the linear case, first a balanced representation of the model is found that is valid in the vicinity of an equilibrium. Subsequently, the state is truncated. Thus, these methods approximate the full model around an equilibrium point [84–86, 230]. Consequently, only local asymptotic stability is guaranteed in the nonlinear case, however, without a region of attraction and error bounds.

Krylov-projector methods for linear models, also called *moment matching* methods, are considered computationally more attractive, especially in conjunction with the Lanczos or Arnoldi procedures [78], and, therefore, also applicable to large-scale models. Here, moments of multivariable LTI models are defined as the coefficients of the Laurent series expansion of the transfer function along so-called *tangential directions* at a complex *interpolation point*, see [10]. The reduction method aims at matching the moments of the reduced-order model to those of the full-order model. The iterative rational Krylov algorithm (IRKA), see [100], iteratively updates the interpolation points and tangential directions, such that locally optimal reduced models are found. However, these methods do not guarantee model stability and neither are these equipped with error bounds.

A connection between the solution of the so-called Sylvester equation and the transfer-function definition of moments [91] has led to the notion of *time-domain* moment matching [17]. Reduced-order models constructed by this method enjoy

parametric freedom that can be exploited to enforce additional model properties, such as stability, passivity, and prescribed pole locations [16], or additional, so-called two-sided, matching conditions [120]. Furthermore, this freedom has been exploited to satisfy first-order optimality conditions [177, 178] in the two-norm of the error transfer function matrix, like the aforementioned IRKA method [100]. Time-domain moment matching methods focus on single-input, single-output (SISO) models. The multiple-input, multiple-output (MIMO) case is not yet fully characterized.

The time-domain interpretation of moment matching has naturally led to a definition of moments for nonlinear models consistent with its LTI counterpart and to reduced-order nonlinear models that achieve moment matching [16, 225, 226]. In these works, the definition of moments for nonlinear models makes use of the center manifold theorem and is, therefore, only defined locally in the neighborhood of the origin. Consequently, the constructed reduced-order model is by definition only an approximation of the full model for inputs close to the specific input used in the reduction process and only in the neighborhood of the origin. In general, an estimate for the size of the neighborhood of the origin is lacking, an error bound is lacking and, by the same token, the reduction methods do not preserve the model structure of the full-order model. More recently, [186] has presented the notion of least squares moment matching for linear and nonlinear models, where time-domain moment matching techniques are employed to approximate an excessive number of moments in a least-squares optimal way. This reduction method provides error bounds for specific classes of inputs.

The time-domain moment matching framework yields great potential as it provides parametric freedom to enforce additional model properties. This freedom can potentially be exploited to achieve optimality in the linear and nonlinear context in some to-be-defined sense. However, a notion of moments of nonlinear models in a *global* setting is lacking. Furthermore, current moment matching techniques do not preserve global stability properties or model structure and are not optimal.

Balancing-based approaches that address the global behavior of the full model, preserve model stability, and provide error bounds are reported in [31–33, 129, 131, 132, 222]. The approaches in [129, 131, 132, 222] use the concepts of generalized and extended differential balancing to truncate less important states of a balanced realization of the full model. These methods preserve the *contraction* property and are equipped with error bounds. The work in [33] relies on the so-called incremental balancing method. The approaches in [31, 32] consider a class of *convergent* nonlinear models that can be decomposed into a feedback interconnection of an LTI subsystem with (dynamic) nonlinear subsystems. Then, only the LTI subsystem is reduced and convergence is preserved by imposing additional conditions. The methods [31–33] provide an a priori error bound. In addition, [31, 32] preserve the structure of the full model to the reduced model. In a similar feedback-interconnected setting, [187] employs *dominance theory* [173] and proposes a structure- and dominance-preserving model reduction technique.

Reduced-order models found by this method can also capture limit cycles and also come with error bounds. The approach in [130] considers the class of monotone models and proposes a method to preserve monotonicity and the input-to-state stability property based on so-called nonlinear dc gain functions. This method is equipped with a *lower* bound on the reduction error. The above-mentioned methods do not pursue optimality. The least-squares Petrov-Galerkin methods in [55, 149] do pursue optimality, as these provide (local) error bounds, which are subsequently minimized.

Data-driven methods, close to some methods reviewed in Section 1.3 on *data-driven modeling*, are also proposed in the literature. Of course, the data-driven methods presented in [38, 280, 284], already introduced in Section 1.3, can also be naturally applied to the model reduction problems to find reduced-order black-box models that preserve incremental stability. Alternatively, the contribution of [1] finds black-box reduced-order models that preserve incremental stability and an incremental gain of the full-order model. Using some error criterion, these methods optimally fit a reduced model to data generated by the full model in simulation for a particular input. The resulting black-box reduced models do not generally preserve the structure of the full model and neither do these models come with error bound.

To summarize, the field of model reduction for *generic* nonlinear models faces the challenges of (i) structure- and stability-preserving reduction for model generalization toward new inputs; (ii) finding error bounds for the quantification of errors between the full and the reduced models; and (iii) optimal reduction to minimize errors between the full and the reduced models.

1.5 Objectives of this thesis

System identification and model reduction are indispensable tools for system modeling. Despite being frequently used, there are still many open challenges, especially for nonlinear systems. This section formulates a list of objectives that together constitute the overarching research objective of this thesis.

> *Develop tools for the computationally efficient identification and optimal model reduction of nonlinear models with stability guarantees.*

1.5.1 Objectives for nonlinear system identification

Computationally efficient identification

The identification of state-space models often boils down to a nonlinear optimization problem that is typically solved using global or gradient-based optimization routines [240]. As these problems are nonlinear in the model parameters, a large

number of evaluations of the error criterion and possibly its gradient with respect to the model parameters are required [193]. Hence, the computational efficiency of such identification depends largely on the efficiency of the computation of model responses and gradient information. For linear models, frequency-domain tools exist for fast computation of *steady-state* model responses for this purpose. For nonlinear models, however, efficient computation of model responses is far less trivial, especially for *continuous-time* models. For this reason, continuous-time models are sometimes replaced by discrete-time ones, because the computation of model responses for the latter is computationally significantly cheaper [236, 240]. Especially in time-critical applications, the fast identification of both continuous-time as well as discrete-time models is an enabler for future developments.

Research objective I. *Enable computationally efficient identification of nonlinear state-space models.*

Identification with stability guarantees

Model construction with stability guarantees enhances model robustness against new unseen inputs as explained in Section 1.2. In particular, enforcing the *convergence* property on identified models guarantees that (i) model solutions remain bounded; (ii) model solutions corresponding to different initial conditions converge to the GAS steady-state output; and (iii) model steady-state solutions corresponding to small changes in the input remain close to each other. The methods in [214, 215, 280, 284] impose related stability notions, namely *incremental stability* or *contractivity*, on *black-box* nonlinear state-space models of certain specific forms. For a more generic model class, for example, models that are derived by first-principle modeling, there is a need for imposing a strong form of model stability during the identification procedure to guarantee and improve robustness against new inputs [154].

Research objective II. *Enable the identification of nonlinear state-space models with inputs with certification of the convergence property.*

Accuracy bounds for model simulation

Numerical model simulation is essential for a wide variety of (nonlinear) systems and control problems such as system identification [152, 205, 240], performance analysis of (control) systems [110, 277, 286], parameter optimization of nonlinear control systems [196, 308], state estimation via (non)linear observers [30], and many more problems. However, even if the model perfectly describes the real-life system, i.e., both are governed by the same set of differential equations, there is a

mismatch between the system response and the computed model. This is due to the fact that real-life systems evolve in the *continuous-time* world whereas their models are simulated in the digital world using numerical simulation algorithms, which have a *discrete-time* nature [15, 252]. Despite being instrumental in practice, the numerical simulation of nonlinear models is often performed without error bounds for such mismatch [252]. Therefore, dedicated simulation techniques are required that come with insightful error bounds to provide an understanding of the key factors governing such errors.

Research objective III. *Establish error bounds for the mismatch between responses of real-life nonlinear systems and responses computed using their models.*

Consistency of closed-loop kernel-based system identification
Given uncertain datasets, consistency is one of the most important and favorable statistical properties of estimators [152]. In some settings, this property is naturally obtained in case of noisy measurement in an open-loop setting. However, in a closed-loop setting, consistency is not obtained if noise is not appropriately handled. The kernel-based identification strategy in [298], which is a nonlinear extension of subspace identification for LTI systems [293], does not explicitly take into account the effect noise. As this identification strategy has favorable properties, such as involving only convex optimization problems, it is important to achieve consistency in both the open-loop and the closed-loop cases.

Research objective IV. *Achieve consistency in data-driven kernel-based state-space modeling.*

1.5.2 Objectives for linear and nonlinear model reduction

Tangential interpolation for MIMO LTI models
Krylov-projector methods or moment matching methods for LTI models are a powerful tool for model reduction with a characteristic transfer-function interpretation [10]. In particular, moments are defined as the coefficients of the Laurent series expansion of the transfer function at complex interpolation points. While in the SISO case, this results in the matching of the transfer functions at the interpolation points, this is not true for the MIMO case. In this case, moment matching achieves matching along the so-called tangential direction of the transfer function matrices. The *time-domain* moment matching technique preserves the matching property for SISO models. However, in general, the *time-domain* moment matching method does not preserve the matching property along tangential directions

for the MIMO case and is thus inconsistent with Krylov-projector reduction techniques.

> **Research objective V.** *Achieve consistency between the time-domain moment matching method and the Krylov-projector method for the reduction of MIMO LTI models.*

Global notion of moments for nonlinear models
The time-domain definition of moments for LTI models has naturally led to a consistent nonlinear counterpart. The current definition of moments in the nonlinear context makes use of the center manifold theorem and is, therefore, only defined locally in the neighborhood of the origin. Consequently, reduced-order models found by the current methods are by definition only an approximation of the full-order model in the neighborhood of the origin. Therefore, there is a need for a definition of moments in the global case and reduction methods that achieve moment matching globally.

> **Research objective VI.** *Extend the moment matching technique for nonlinear models from the local case to the global case.*

Optimal model reduction for linear and nonlinear models
Reduced-order models are used as a replacement for the full-order model when the full-order model is too expensive for its purpose, e.g., analysis or implementation. Therefore, it is desired in both the linear and nonlinear cases that the reduced-order model achieves optimal accuracy. Furthermore, to promote generalizability to new inputs, stability properties, such as convergence, and the model structure of the full model should be preserved. Lastly, error bounds should be derived as these provide a quality measure to form a deeper, quantitative understanding of the capabilities of the reduced model.

> **Research objective VII.** *Enable optimal model reduction for linear and nonlinear models with error bounds whilst preserving the model structure and the convergence property.*

1.6 Contributions of this thesis

This thesis presents novel approaches for the data-driven nonlinear modeling and complexity reduction of nonlinear models that are validated either experimentally,

using benchmark datasets, or via simulation studies. The focus is on guaranteeing a strong form of model stability, namely *convergence*, which provides robustness for the usage of the constructed models in unseen scenarios. In the subsequent sections, the contributions per chapter are detailed for the individual parts of this thesis.

1.6.1 Part I: Identification of convergent Lur'e-type systems

Chapter 2 proposes a parametric identification approach for a class of continuous-time SISO Lur'e-type systems that guarantees the identified Lur'e-type model to be exponentially convergent and uses numerically efficient tools to compute model responses and gradient information to be used for gradient-based optimization. Lur'e-type models can be decomposed in the LTI dynamics that are placed in feedback with a static nonlinearity. The performance and the computational benefits of the proposed identification approach are demonstrated in simulation and experimental case studies. It is shown that the proposed identification method significantly reduces the identification time, whilst enforcing the *convergence* property. In particular, the experimental case study in mechanical ventilation highlights the computational efficiency of this approach, enabling fast identification of both equipment parameters, as well as patient parameters. This chapter contributes toward research objectives I and II.

> **Contribution I.** *A computationally efficient approach to the identification of continuous-time convergent Lur'e-type systems that guarantees that the identified Lur'e-type model is also convergent.*

Chapter 3 proposes an identification approach that can be viewed as the discrete-time counterpart of the approach in Chapter 2 with an extended model class such that MIMO systems with multiple nonlinear functions in the feedback loop can be considered. The identification approach is benchmarked on several datasets. This chapter contributes toward research objectives I and II.

> **Contribution II.** *A computationally efficient approach to the identification of discrete-time multivariable convergent Lur'e-type systems that guarantees that the identified Lur'e-type model is also convergent.*

Chapter 4 analyzes the steady-state mismatch between the response of continuous-time nonlinear Lur'e-type systems and their continuous-time nonlinear Lur'e-type models. Even though real-life systems evolve in the continuous-time world, their measured, sampled responses are only available on a discrete-time grid. Model responses are computed by numerical simulation algorithms which also have a

discrete-time nature. The first result of this chapter bounds the steady-state mismatch between the *measured sampled* system response and the simulated model response. This bound provides a measure of how close the simulated model output is with respect to the sampled measured system output. For the *system identification* problems, in particular, such a mismatch quantification can be exploited to distinguish between structural modeling errors and errors due to inaccurate model simulation. The second result bounds the steady-state mismatch between the *actual* system response and simulated model response, thereby providing insight into the intersample behavior of the system that is not captured by sampling its output response. The bounds are exploited to provide insights into the factors causing the mismatches. Furthermore, methods to reduce the derived bounds are presented. This chapter contributes toward research objective III.

Contribution III. *Error bounds for the mismatch between the steady-state responses of real-life continuous-time nonlinear convergent Lur'e-type systems and the steady-state responses of their continuous-time nonlinear convergent Lur'e-type models computed using a discrete-time numerical simulation algorithm.*

1.6.2 Part II: Moment matching for linear and nonlinear models

Chapter 5 presents a matching property in terms of the MIMO transfer-function matrices of the full- and the reduced-order LTI models for model order reduction by time-domain moment matching. Furthermore, additional conditions on time-domain moment matching parameters are formulated under which the match is consistent with classical Krylov-projector methods. This chapter contributes toward research objective IV.

Contribution IV. *A time-domain moment matching method for MIMO LTI models that achieves a matching property that is consistent with classical Krylov subspace methods.*

Chapter 6 extends the notion of moments of nonlinear models from the local to the global context for a generic class of convergent nonlinear models. Furthermore, this chapter proposes a constructive model reduction approach for convergent Lur'e-type models that preserves the Lur'e-type model structure, inherits the frequency-response function interpretation of LTI moment matching, preserves the convergence property, and is equipped with an a posteriori error bound. The benefits of this approach are illustrated in a numerical case study. This chapter contributes toward research objectives VI and VII.

Contribution V. *The extension of time-domain moment matching to the global case for nonlinear models and a constructive method for the reduction of Lur'e-type models that preserves its structure and the convergence property.*

Chapter 7 presents optimal model reduction techniques for MIMO LTI models and MIMO convergent nonlinear models. In the LTI case, moment matching is achieved, asymptotic stability is preserved, and the approximation error in the $\mathcal{H}_\infty$-norm is minimized. In the nonlinear context, the convergence property is preserved and optimality is achieved by the minimization of an error bound. The effectiveness of the proposed methods is demonstrated in numerical case studies. This chapter contributes towards research objectives V, VI, and VII.

Contribution VI. *A model reduction method for linear and nonlinear models that is structure- and convergence-preserving and optimal in terms of a minimized error bound on the steady-state output mismatch between the full and reduced models.*

1.6.3 Part III: Kernel-based identification of nonlinear systems

Chapter 8 extends a consistent closed-loop identification method for LTI systems to the case of kernel-based identification of nonlinear systems. The proposed three-step approach identifies first the noise sequence, then the state sequence, and finally the state-space mapping. Although not proving overall consistency, each step of the proposed procedure corresponds to a consistent estimate under certain assumptions. Using simulation studies, the performance of the proposed identification strategy is evaluated in both the open-loop and the closed-loop cases. This chapter contributes toward research objective IV.

Contribution VII. *A step-wise consistent approach to the kernel-based state-space identification of open-loop and closed-loop discrete-time multivariable nonlinear systems.*

Chapter 9 presents a kernel-based approach for the state-space identification of nonlinear systems with a focus on enforcing *convergence* on compact sets of the input space and the state space. Certification of the convergence property enables safe and robust generalization toward unseen scenarios. The benefits of the approach are illustrated by means of simulation and benchmark examples. This

chapter contributes toward research objective VIII.

Contribution VIII. *An approach to kernel-based state-space identification of discrete-time nonlinear systems that guarantees that the identified model is convergent.*

1.7 Outline of this thesis

Including this introduction, the thesis consists of ten chapters. Each of the contributions formulated in Section 1.6 is represented by one chapter, where each of these chapters is self-contained and can be read independently. Furthermore, Chapters 2-9 correspond to peer-reviewed publications. The conclusions and recommendations of this thesis are presented in the fourth part of this thesis.

Part I: Identification of convergent Lur'e-type systems

Chapter 2 corresponds to Contribution I and the publications:

- Shakib, M. F., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. (2022). Computationally efficient identification of continuous-time Lur'e-type systems with stability guarantees. *Automatica*, 136, 110012.
- Shakib, M. F., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. (2019). Fast identification of continuous-time Lur'e-type systems with stability certification. *IFAC-PapersOnLine*, 52(16), 227-232.

Chapter 3 corresponds to Contribution II and the publications:

- Shakib, M. F., Vervaet, N., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Fast identification of multivariable discrete-time Lur'e-type systems with stability certification. *Submitted for journal publication.*

Chapter 4 corresponds to Contribution III and the publications:

- Shakib, M. F., Schoukens, J., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Accuracy bounds for the simulation of a class of continuous-time nonlinear models. *In preparation.*

Part II: Moment matching for linear and nonlinear models

Chapter 5 corresponds to Contribution IV and the publications:

- Shakib, M. F., Scarciotti, G., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Time-domain moment matching for multiple-input multiple-output linear time-invariant models. *Provisionally accepted for publication in Automatica.*

Chapter 6 corresponds to Contribution V and the publications:

- Shakib, M. F., Scarciotti, G., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Model reduction by moment matching with preservation of global stability for a class of nonlinear models. *Submitted for journal publication.*
- Shakib, M. F., Scarciotti, G., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. (2021). Model reduction by moment matching for convergent Lur'e-type models. In *Proceedings of the American Control Conference*, (pp. 4449-4454).

Chapter 7 corresponds to Contribution VI and the publications:

- Shakib, M. F., Scarciotti, G., Jungers, M., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. Optimal model reduction by moment matching: the linear and nonlinear feedback cases. *Submitted for journal publication.*
- Shakib, M. F., Scarciotti, G., Jungers, M., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. (2021). Optimal $\mathcal{H}_\infty$ LMI-based model reduction by moment matching for linear time-invariant models. In *Proceedings of the Conference for Decision and Control*, (pp. 6914-6919).

Part III: Kernel-based identification of nonlinear systems

Chapter 8 corresponds to Contribution VII and the publications:

- Shakib, M. F., Tóth, R., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. (2020). State-space kernelized closed-loop identification of nonlinear systems. *IFAC-PapersOnLine*, 53(2), 1126-1131.

Chapter 9 corresponds to Contribution VIII and the publications:

- Shakib, M. F., Tóth, R., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Kernel-based identification of nonlinear state-space models with stability guarantees. *In preparation.*

I

Identification of convergent Lur'e-type systems

2

Computationally efficient identification of continuous-time Lur'e-type systems with stability guarantees

This chapter proposes a parametric system identification approach for a class of continuous-time Lur'e-type systems. Using the Mixed-Time-Frequency algorithm, we show that the steady-state model response and the gradient of the model response with respect to its parameters can be computed in a numerically fast and efficient way, allowing efficient use of global and local optimization methods to solve the identification problem. Furthermore, by enforcing the identified model to be inside the set of convergent models, we certify a stability property of the identified model, which allows for reliable generalized usage of the model, also for other excitation signals than those used to identify the model. The effectiveness and benefits of the proposed approach are demonstrated using a simulation case study. Furthermore, we have experimentally shown that the proposed approach provides fast identification of both medical equipment and patient parameters in mechanical ventilation and, thereby, enables improved patient treatment.

The contents of this chapter are published in: Shakib, M. F., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. (2022). Computationally efficient identification of continuous-time Lur'e-type systems with stability guarantees. *Automatica*, 136, 110012. Preliminary results are published in: Shakib, M. F., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. (2019). Fast identification of continuous-time Lur'e-type systems with stability certification. *IFAC-PapersOnLine*, 52(16), 227-232.

2.1 Introduction

Accurate dynamical models of complex systems are required for model-based controller design, analysis of the dynamic behavior of the system under study, and improved system design. System identification deals with the construction of such dynamic models from observed input and output data. For the class of linear systems, many identification techniques exist [152], together with user-friendly software, see, e.g., [153]. However, many physical systems are essentially nonlinear and, depending on the application, one can choose to model it in a linear framework by neglecting (non-significant) nonlinear phenomena, or can choose to also model the nonlinearities and, therewith, increase the model accuracy at the cost of model complexity.

A major challenge in system identification is to enforce a form of stability on the identified model [154, 240, 280]. Even if it is known that the true system under study enjoys some stability property, it is not guaranteed that the identified model preserves this stability property due to the finite length of the dataset, noisy datasets, or the presence of unmodeled dynamics. Several methods that enforce exponential stability are proposed in the literature for the identification of linear time-invariant (LTI) models [105, 146, 280, 283, 290]. For nonlinear systems, however, the literature on the identification of nonlinear models with some guaranteed form of stability is scarce and the consequences of not enforcing stability are more pronounced and more complex in the nonlinear context [60, 64, 240]. Firstly, it is well-known that nonlinear models can exhibit multiple stable solutions being attractive for different sets of initial conditions [135]. Consequently, the response of the identified model might be close to the measured system response, but might as well be far off or even become unbounded, depending on the initial condition. Secondly, nonlinear models can have a large sensitivity to the excitation signal, and, therefore, their response can be significantly different, or even become unbounded, even for the slightest change in the excitation signal [135]. Such models have poor generalization capabilities to other excitation signals than the ones used during identification. With the instability issue of the identified model in mind, [280, 283] developed a method that imposes global incremental stability for a class of identified black-box nonlinear state-space models, where the incremental stability property avoids the above problems of high sensitivity to input changes and initial conditions. In that approach, the dataset includes the state sequence of the underlying nonlinear system, which is non-trivial and rather case specific to obtain in practice. Other types of stability properties, e.g., asymptotic stability of the origin for zero inputs and input-to-state stability, are enforced in [68, 171, 270, 271, 285], mostly for autonomous black-box nonlinear models. Note, however, that models that enjoy these other types of stability properties can still exhibit a large sensitivity to input changes and initial conditions.

To address the challenge of identifying nonlinear models with a form of incremental stability, we focus on a practically relevant class of nonlinear systems,

namely Lur'e-type systems [135]. These systems are block structured in the sense that all the LTI dynamics are captured in an LTI block and separated from the nonlinearities that are captured in a static nonlinear block in the feedback loop. The class of Lur'e-type systems encompasses so-called Wiener, Hammerstein, and nonlinear feedback systems and, therefore, captures a large class of nonlinear systems. In the system identification literature, Lur'e-type systems are also referred to as nonlinear feedback systems [96], and NonLinear Linear Fractional Representation (NL-LFR) systems [244, 296].

Different approaches for the identification of Lur'e-type systems have been proposed in the literature [115, 175, 188, 234, 243, 244, 291, 294, 296, 300]. What most of these approaches have in common is that, as a final step in the identification procedure, a non-convex cost function has to be minimized using gradient-based optimization routines. Furthermore, all these approaches are formulated in a discrete-time setting. Although continuous-time counterparts of most of these methods can be formulated readily, it is an open problem how to perform the gradient-based minimization of the cost function in the continuous-time setting in a numerically efficient way. More specifically, any gradient-based optimization routine requires the computation of model responses to evaluate the cost function and requires the computation of the gradient of the cost function with respect to the model parameters. While the model simulation of discrete-time nonlinear models boils down to a series of computationally cheap algebraic operations, the standard means of model simulation for continuous-time nonlinear models is numerical forward integration, which is computationally extremely expensive.

In this chapter, we present a parametric identification approach for a class of continuous-time single-input single-output (SISO) Lur'e-type systems that: (i) guarantees the identified model to be exponentially convergent and, (ii) uses numerically efficient tools to compute model responses and gradient information to be used for gradient-based optimization. Point (i) guarantees that the identified model is globally exponentially convergent, which is a property of (non)linear models that guarantees the uniqueness, boundedness, and global exponential stability of the steady-state solution [197]. As a consequence, for any bounded excitation, the identified model 'forgets' its initial condition and, therefore, exhibits a uniquely defined, bounded steady-state solution, which is globally exponentially stable and depends solely on the applied excitation. Furthermore, the identified model enjoys the property that a small variation of the excitation signal results in a small variation of the steady-state output, which adds to the robustness of the identification result. Point (ii) enables system identification of continuous-time nonlinear models in a computationally efficient way. Hereto, the so-called mixed-time-frequency (MTF) algorithm is employed to enable fast computation of steady-state model responses under periodic excitations [196]. The MTF algorithm facilitates the usage of global optimization techniques [156] and also enables efficient computation of the gradient of the underlying cost function with respect to the model parameters, which can be effectively used in any gradient-based optimization routine.

We demonstrate the performance and the computational benefits of the proposed identification approach in two case studies. The first case study is a simulation example, where we identify a black-box nonlinear model. We show that the proposed identification strategy is significantly faster than conventional methods as the computation time is reduced from *hours* to only *minutes*. The second case study is an experimental study on mechanical ventilation, which is used to regulate the breathing of patients in respiratory distress during the nursery of intensive-care and acutely ill surgical patients. Particularly, during the COVID-19 crisis, mechanical ventilation is extensively used to support the breathing of hospitalized patients with severe lung damage. In our longstanding relationship with Demcon Macawi Respiratory Systems, Best, The Netherlands, we mainly focused on the pressure control aspect [116, 211]. In this chapter, we address the identification of the parameters of a first-principle model. The model includes the patient's lung parameters, which reveal important patient health information that is used in the medical decision-making process [5, 145] and for pressure control purposes [116, 275]. We show that the proposed identification method significantly reduces the computation time, which is crucial as it enables faster patient treatment with the aim to avoid negative consequences for the patient's lungs [90, 258] and, saves valuable time for the medical practitioner.

In summary, the main contribution of this chapter is a computationally efficient approach to the identification of continuous-time convergent Lur'e-type systems that guarantees that the identified model is also globally exponentially convergent. We have demonstrated theoretically that this approach is statistically consistent and we have demonstrated numerically that this approach is effective and computationally efficient. Furthermore, we have shown that the developed approach is an enabler for innovation in mechanical ventilation as it allows to identify physical parameters in a fast way, which can subsequently be used for improved patient treatment.

The remainder of this chapter is structured as follows. Section 2.2 introduces the considered class of Lur'e-type systems and recalls sufficient conditions for convergence. Section 2.3 formally poses the identification problem as a constrained optimization problem and also shows the statistical consistency of the estimator. Section 2.4 provides an overview of methods to minimize the cost function, constrained to the set of convergent models. Section 2.5 introduces the MTF algorithm and also shows how the gradient of the cost function is computed accurately and efficiently. Section 2.6 presents the performance of the identification approach in a simulation study. Section 2.7 describes the mechanical ventilation case study. Section 2.8 closes with concluding remarks.

2.2 Convergent Lur'e-type systems

We consider SISO Lur'e-type systems described by

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= \mathcal{A}\boldsymbol{x}(t) + \mathcal{B}\boldsymbol{u}(t) + \mathcal{L}w(t), \\ \boldsymbol{y}(t) &= \mathcal{C}\boldsymbol{x}(t) + \mathcal{D}w(t), \\ \boldsymbol{z}_0(t) &= \mathcal{F}\boldsymbol{x}(t) + \mathcal{G}\boldsymbol{u}(t) + \mathcal{H}w(t), \\ \boldsymbol{u}(t) &= -\phi(\boldsymbol{y}(t), w(t)), \end{aligned} \tag{2.1}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the state vector, $\boldsymbol{y}(t) \in \mathbb{R}$ is the unmeasured feedback signal, $\boldsymbol{z}_0(t) \in \mathbb{R}$ is the output, $w(t) \in \mathbb{R}$ is the user-defined external input, $\phi(\boldsymbol{y}(t), w(t)) : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is a static nonlinear function. We assume that $\phi(0, w) = 0\,\forall\, w \in \mathbb{R}$. Fig. 2.1 depicts the considered Lur'e-type system schematically. Although checking the stability of forced LTI systems is well-understood, checking the stability of (forced) solutions of nonlinear systems is non-trivial. Consequently, enforcing stability on the identified nonlinear model becomes challenging [154]. To address this challenge, we use the notion of convergent dynamics defined as follows.

Definition 2.1 ([197])**.** *System* (2.1) *with input w(t) that is defined and bounded on* $t \in \mathbb{R}$*, is said to be globally exponentially convergent if there exists a solution* $\bar{\boldsymbol{x}}(t)$ *satisfying the following conditions:*

- $\bar{\boldsymbol{x}}(t)$ *is defined and bounded on* $t \in \mathbb{R}$*,*
- $\bar{\boldsymbol{x}}(t)$ *is globally exponentially stable.*

■

A solution $\bar{\boldsymbol{x}}(t)$ of the system (2.1), which is defined for $t \in (t_\star, +\infty)$, is called globally exponentially stable if there exists a $\delta > 0, \alpha > 0$, and $\beta > 0$ such that $|\boldsymbol{x}(t_0) - \bar{\boldsymbol{x}}(t_0)| < \delta$ for any $t_0 \in (t_\star, +\infty)$ implies

$$|\boldsymbol{x}(t) - \bar{\boldsymbol{x}}(t)| \leq \alpha e^{-\beta(t-t_0)} |\boldsymbol{x}(t_0) - \bar{\boldsymbol{x}}(t_0)|\, \forall\, t \geq t_0$$

with $|\cdot|$ the Euclidean norm. The solution $\bar{\boldsymbol{x}}(t)$ is called the *steady-state* solution and depends on the excitation signal $w(t)$. For the considered system (2.1), the following theorem provides sufficient conditions for global exponential convergence.

Theorem 2.1 ([196])**.** *Consider system* (2.1) *and suppose that*

C1 *the matrix* $\mathcal{A}$ *is Hurwitz;*

C2 *there exists a* $K > 0$ *such that the nonlinearity* $\phi(\boldsymbol{y}, w)$ *satisfies* $\left|\frac{\phi(\boldsymbol{y}_2, w) - \phi(\boldsymbol{y}_1, w)}{\boldsymbol{y}_2 - \boldsymbol{y}_1}\right| \leq K$ *for all* $\boldsymbol{y}_1, \boldsymbol{y}_2, w \in \mathbb{R}$*;*

C3 $\sup_{\omega\in\mathbb{R}} |\mathcal{C}(j\omega I - \mathcal{A})^{-1}\mathcal{B}| < \frac{1}{K}$.

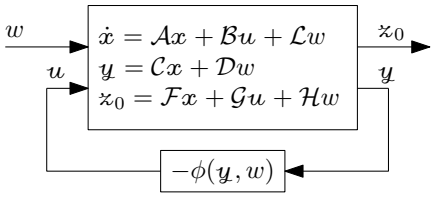


Figure 2.1. Schematic representation of Lur'e-type system (2.1).

Then, for any input $w(t)$ that is defined and bounded on $t \in \mathbb{R}$, system (2.1) *is globally exponentially convergent according to Definition 1. Moreover, any other solution $x(t)$ of system* (2.1) *subject to the same input $w(t)$ satisfies*

$$|x(t) - \bar{x}(t)| \leq \alpha e^{-\beta(t-t_0)} |x(t_0) - \bar{x}(t_0)|, \ \forall t \geq t_0.$$

The constants $\alpha > 0$ and $\beta > 0$ are independent of the input $w(t)$. ▲

For notational convenience, we write exponentially convergent instead of exponentially convergent for any bounded input $w(t)$ from here on. The conditions of Theorem 2.1 are used in Section 2.3 to enforce the identified Lur'e-type model to be globally exponentially convergent according to Definition 2.1. Globally exponentially convergent systems 'forget' their initial condition since, independently of the initial condition, all solutions $x(t)$ converge exponentially to the steady-state solution $\bar{x}(t)$. The conditions of Theorem 1 give the even stronger property of input-to-state convergence, which implies that small variations (in the infinity norm) of the input signal $w(t)$ lead to small variations (in the infinity norm) of the steady-state solution $\bar{x}(t)$, see Theorem 2.2.17 in [200]. Convergent systems also exhibit the property that when excited by a periodic excitation signal $w(t)$ with period-time T, the steady-state solution $\bar{x}(t)$ is also periodic with the same period-time T [197]. In the nonlinear identification literature, often the PISPO (Periodic Input, the same Period Output) system class is considered, which also possesses this periodicity property [235, 243]. This latter property facilitates the use of only steady-state data for identification, which is common practice in nonlinear system identification [240].

2.3 Identification setting

This section formally introduces the identification problem. Hereto, we first introduce the considered model class. After that, we introduce the cost function central to the identification problem and show that the proposed estimator is statistically consistent.

2.3.1 Model parametrization

The considered model class is a copy of (2.1):

$$\begin{aligned}\dot{x}(t,\theta) &= A(\theta)x(t,\theta) + B(\theta)u(t,\theta) + L(\theta)w(t),\\ y(t,\theta) &= C(\theta)x(t,\theta) + D(\theta)w(t),\\ z(t,\theta) &= F(\theta)x(t,\theta) + G(\theta)u(t,\theta) + H(\theta)w(t),\\ u(t,\theta) &= -\varphi(y(t,\theta), w(t), \theta),\end{aligned} \tag{2.2}$$

parametrized by the vector $\theta \in \mathbb{R}^{n_\theta}$. The state dimension n of the true system (2.1) is considered known.

We distinguish between two parametrization approaches, namely (i) the model parametrization following from first-principle modeling; and (ii) a black-box model parametrization. In approach (i), the model parameters of both the linear part and the nonlinear part have a physical meaning. Such model parametrization is used in the experimental case study in Section 2.7. In approach (ii), the parameters do not have a physical meaning. Then, for the state-space matrices, a canonical form or a full parametrization can be taken, in which each element of the state-space matrices is one model parameter. In a black-box approach, the nonlinearity φ is typically parametrized by a set of basis functions as follows:

$$\varphi(y(t,\theta), w(t), \theta) = \sum_{k=1+n_{\text{LTI}}}^{n_\theta} \theta_k f_k(y(t,\theta), w(t)), \tag{2.3}$$

where n_{LTI} is the number of parameters used to parametrize the LTI block. The user-defined basis functions $f_k(y(t,\theta), w(t))$ have associated parameters θ_k. A black-box model parametrization is used in the simulation case study presented in Section 2.6.

One of the goals of this work is to guarantee that the identified model is globally exponentially convergent according to Definition 2.1. To this extent, we define the set $\Theta \subset \mathbb{R}^{n_\theta}$ with a nonempty interior as the set of parameter vectors θ for which model (2.2) satisfies conditions **C1 - C3** in Theorem 2.1 (with $\mathcal{A}, \mathcal{B}, \mathcal{C}, \phi(\underline{y}, w)$ replaced by $A(\theta), B(\theta), C(\theta), \varphi(y, w, \theta)$, respectively). The set Θ serves as a constraint on the parameter vector θ to be respected in the identification process. We close this section by assuming that the to-be-identified true system (2.1) is also globally exponentially convergent, which is a stability property that the developed identification method aims to preserve for the identified model.

Assumption 2.1. *The true system* (2.1) *is in the model class proposed in* (2.2) *with parameter vector* $\theta_0 \in \text{int}\Theta$*, and satisfies conditions* ***C1*** *-* ***C3*** *of Theorem 2.1. Therefore, the true system* (2.1) *is globally exponentially convergent according to Definition 2.1.* ■

Remark 2.1. The nonlinearity φ of model (2.2) should be parametrized carefully in a black-box parametrization approach. In particular, the incremental sector

bound imposed by condition **C2** in Theorem 2.1 should be satisfied globally, i.e., for all $y_1, y_2, w \in \mathbb{R}$. Inappropriate choices for the basis functions f_k in (2.3) can lead to unstable models and to models that exhibit multiple co-existing (stable) steady-state solutions. This is encountered in, e.g., polynomial basis functions [60]. Suitable nonlinearities are, e.g., piecewise linear maps, sigmoids, the arctangent function, and the hyperbolic tangent function. △

2.3.2 Cost function and consistency

Assumption 2.1 guarantees that the true system is globally exponentially convergent. Therefore, when excited by a periodic input $w(t)$ with period-time T, the periodic steady-state output $\bar{x}_0(t)$ of system (2.1) has the same period-time T. In practice, only a discrete version of the measured steady-state output $\bar{x}_0(t)$ is available at sampling instants $t_k := t_0 + kt_s$ with $k = 1, \ldots, N$, where t_s is the sampling interval and $N = PT/t_s$ is the total number of samples in P steady-state periods. We assume the following noise scenario.

Assumption 2.2. *The measured output is $\bar{x}(t_k) = \bar{x}_0(t_k) + e(t_k)$. The discrete-time noise source $e(t_k)$ is zero-mean Gaussian white noise with finite variance σ_e^2 and independent of the applied input signal $w(t_k)$.* ■

By guaranteeing that $\theta \in \Theta$, i.e., the candidate model is also globally exponentially convergent, model (2.2) subject to the same periodic input $w(t)$ results in the periodic steady-state output $\bar{z}(t, \theta)$ with the same period-time T. This leads to the definition of the steady-state error over P periods:

$$\bar{\epsilon}(t_k, \theta) := \bar{z}(t_k, \theta) - \bar{x}(t_k), \quad k = 1, \ldots, N. \tag{2.4}$$

The squared simulation error is then taken as the cost function, which is defined as follows:

$$J_N(\theta) := \frac{1}{N} \sum_{k=1}^{N} \bar{\epsilon}(t_k, \theta)^2. \tag{2.5}$$

Given the dataset $\{w(t_k), \bar{x}(t_k)\}_{k=1}^{N}$, the identification objective is to find the parameter vector $\hat{\theta}_N$ that globally minimizes $J_N(\cdot)$ constrained to the set of convergent models characterized by the parameter set Θ, that is

$$\hat{\theta}_N = \arg\min_{\theta \in \Theta} J_N(\theta). \tag{2.6}$$

To study consistency of the estimator (2.6), define the function $V_N(\cdot)$ as follows:

$$V_N(\theta) := \frac{1}{N} \sum_{k=1}^{N} \left(\bar{z}(t_k, \theta) - \bar{x}_0(t_k) \right)^2. \tag{2.7}$$

This function can be interpreted as a measure of the steady-state mismatch between the model response $\bar{z}(t_k, \theta)$ and the *noiseless* system response $\bar{x}_0(t_k)$. The following persistency of excitation assumption is required.

Assumption 2.3. *The unique global minimum of $V_N(\theta)$ in (2.7) is θ_0.* ■

Assumptions 2.1-2.3 allow formulating the following consistency result.

Theorem 2.2. *Under Assumptions 2.1-2.3, the estimate $\hat{\theta}_N$ in (2.6) converges in probability to θ_0, i.e.,*

$$\lim_{N\to\infty} \mathbb{E}\left[\hat{\theta}_N - \theta_0\right] = 0.$$

▲

Proof. The proof can be found in Appendix A.1. □

Remark 2.2. The model parametrization is non-unique [238, 243] in the sense that a model subject to a similarity transformation of the LTI block, a gain exchange between the LTI block and the static nonlinear block, or a loop-transformation [135] produces the same steady-state output. Therefore, the true parameter vector θ_0 is a set and the consistency claim of Theorem 2.2 should be understood in the sense that $\hat{\theta}_N$ converges in probability to the set θ_0 as $N \to \infty$. △

2.4 Cost function minimization strategy

The identification problem defined in (2.6) is a constrained optimization problem with the cost function in (2.5). As this cost function is generally non-convex, we propose a two-step optimization approach to solve (2.6): (i) initialization and (ii) gradient-based optimization. The result of step (i) should be a set of initial model parameters θ_{init} in the vicinity of θ_0, which globally minimizes the cost function. For applicability of the numerically efficient tools that we present in Section 2.5, the initial model with parameters θ_{init} should satisfy the conditions of Theorem 2.1 and, therefore, be globally exponentially convergent. The gradient-based search of step (ii) is started from the set of initial parameters θ_{init} and results in the parameter set $\hat{\theta}_N$ that corresponds to the nearest minimum of the cost function. Again, for applicability of the numerically efficient tools that we present in Section 2.5, in all optimization iterations of the gradient-based search, the conditions of Theorem 2.1 should be satisfied.

2.4.1 Initialization

We discuss three approaches for the initialization step that yield a set of initial model parameters θ_{init}. Again, we want to stress that the model with parameters θ_{init} should satisfy the conditions of Theorem 2.1.

Physical insight This method is only applicable for models that are derived by first-principle modeling, implying that the parameters of these models represent physical quantities. The user appropriately chooses the initial parameters θ_{init} based on its physical insights. However, it should be kept in mind that the model

with parameters θ_{init} should satisfy the conditions of Theorem 2.1, which can make the initial guess challenging, especially for models with a large number of parameters.

Best Linear Approximation This method relies on estimating the so-called Best Linear Approximation (BLA) of the nonlinear system [235]. Hereto, the MATLAB implementations `N4SID` or `TFEST`, can be used, which can also enforce the $A(\theta)$-matrix of the identified model to be Hurwitz. In addition, by imposing no feedback, i.e., choosing $C(\theta)$ and $D(\theta)$ both 0, $B(\theta) = 0$ or $\varphi(y, w, \theta) = 0$, the initial model satisfies the conditions of Theorem 2.1. The BLA framework has the advantages that it allows for fast estimation of initial model parameters, provides a nonlinear distortion analysis, and gives a rough estimate for the state dimension.

Global Optimization This method employs global optimization routines such as genetic-type, swarm intelligent or Monte-Carlo routines [156]. Although global optimization is typically computationally expensive, we can achieve fast operation using the efficient numerical tools that we introduce in Section 2.5. Global optimization routines typically cluster around the global minimum and provide, therefore, an accurate set of initial model parameters θ_{init} that correspond to a full *nonlinear model* rather than a linear model as in the BLA framework. To limit the search space, (rough) bounds on the parameter values should be known a priori.

Depending on the application at hand, either one of the three methods detailed above can be used to find a set of initial model parameters θ_{init}. In the simulation case study in Section 2.6, we demonstrate the effectiveness of a global optimization routine. In the experimental case study in Section 2.7, we use available knowledge on the physical parameters for initialization. Numerous identification results with the BLA method are presented in the literature, see [243] for an overview.

Remark 2.3. Global optimization can also be viewed as a stand-alone identification procedure without a local gradient-based search as a second step. To view it as a two-step approach, consider step one as a less accurate but global search and step two as a local refinement. △

2.4.2 Gradient-based optimization

After an initial parameter vector θ_{init} in the initialization step is obtained, a gradient-based search can be performed. Hereto, the gradient of the cost function in (2.5) with respect to the parameter vector θ is required, namely:

$$\frac{\partial J_N(\theta)}{\partial \theta} = \frac{2}{N} \sum_{k=1}^{N} \bar{\epsilon}(t_k, \theta) \frac{\partial \bar{\epsilon}(t_k, \theta)}{\partial \theta}. \tag{2.8}$$

In (2.8), the steady-state error $\bar{\epsilon}(t_k, \theta)$ defined in (2.4), as well as the gradient of this steady-state error $\partial \bar{\epsilon}(t_k, \theta)/\partial \theta = \bar{\partial} \bar{z}(t_k, \theta)/\partial \theta$, are required. We show in the next

section that we can compute both $\bar{\epsilon}(t_k, \theta)$ and $\partial\bar{\epsilon}(t_k, \theta)/\partial\theta$ in a computationally efficient and accurate way using the MTF algorithm.

To facilitate the application of the MTF algorithm during a gradient-based search, the candidate model should be convergent in all iterations, i.e., $\theta \in \Theta$. In literature, there are many gradient-based methods[1] that deal with constrained optimization problems [193]. Alternatively, one can introduce an exterior penalty function $\psi(\theta)$, which turns the constrained optimization problem into an unconstrained one, where

$$\psi(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta, \\ \infty & \text{if } \theta \notin \Theta. \end{cases}$$

The modified cost function then reads as follows:

$$J_{N,\mathrm{mod}}(\theta) = J_N(\theta) + \psi(\theta) \tag{2.9}$$

with $J_N(\cdot)$ defined in (2.5). It can be minimized by any gradient-based optimization algorithm (such as the Levenberg-Marquart algorithm) to yield

$$\hat{\theta}_{N,\mathrm{mod}} = \arg\min_{\theta} J_{N,\mathrm{mod}}(\theta). \tag{2.10}$$

The modified cost function $J_{N,\mathrm{mod}}(\cdot)$ is unbounded for models not inside the set of the convergent models, but leaves the cost function $J_N(\cdot)$ unaffected for models inside the set of convergent models. Consequently, as long as the set of convergent models is respected, the gradient of the cost function with respect to the model parameters remains unaffected. In contrast to *interior* penalty functions (also known as barrier functions), the *exterior* penalty function does not result in a bias in the parameters, i.e., $\hat{\theta}_N$ in (2.6) equals $\hat{\theta}_{N,\mathrm{mod}}$ in (2.10).

2.5 Computation of steady-state responses and gradient information

As shown in the previous section, the steady-state output $\bar{z}(t, \theta)$ of the model (2.2) is required at sampling instances t_k in the evaluation of the cost function (2.5) and the computation of its gradient (2.8). First, we introduce the MTF algorithm, which enables fast computation of the steady-state response of convergent Lur'e-type models. After that, we present a method to compute the gradient $\partial\bar{\epsilon}(t, \theta)/\partial\theta$, required in (2.8), again, in a computationally efficient manner using the MTF algorithm.

[1] MATLAB provides built-in solvers such as `fmincon` that can handle constrained optimization problems.

2.5.1 Mixed-time-frequency algorithm

To overcome the time-consuming drawbacks of computing model responses using numerical forward integration, [196] developed the so-called MTF algorithm. If the underlying Lur'e-type model satisfies the conditions stated in Theorem 2.1, then this algorithm computes the steady-state model response $\bar{z}(t,\theta)$ efficiently under a periodic excitation $w(t)$. For notational convenience, the dependency on θ is dropped in this section.

The MTF algorithm is an iterative algorithm. At each iteration i, two mappings are involved:

$$u_{i+1} = \mathcal{F}_{uy} \circ y_i, \tag{2.11a}$$
$$y_{i+1} = \mathcal{F}_{yu} \circ u_{i+1} + \mathcal{F}_{yw} \circ w, \tag{2.11b}$$

where the nonlinear steady-state operator $\mathcal{F}_{uy} \circ y_i \coloneqq -\varphi(y_i, w)$, and the linear steady-state operators $\mathcal{F}_{yu}, \mathcal{F}_{yw}$ map periodic signals u_{i+1} and w to the periodic steady-state output y_{i+1}. The MTF algorithm relies on the fact that the composed operator $\mathcal{F}_{uy} \circ \mathcal{F}_{yu}$ is a contraction operator acting from the space[2] $L_2(T)$ to $L_2(T)$, if the model satisfies the conditions of Theorem 2.1. For computational efficiency, the nonlinear mapping $\mathcal{F}_{uy} \circ y_i$ is evaluated in the time domain, while the linear mappings $\mathcal{F}_{yu}$ and $\mathcal{F}_{yw}$ are evaluated in the frequency domain according to the frequency response functions (FRFs):

$$G_{yu}(j\omega) = C(j\omega I - A)^{-1}B,$$
$$G_{yw}(j\omega) = C(j\omega I - A)^{-1}L + D.$$

The MTF algorithm iteratively evaluates the mappings (2.11b), (2.11a), while transforming the intermediate signals between the time and frequency domains using the (Inverse) Fast Fourier Transform ((I)FFT), truncated to only M frequency contributions. Hereto, denote by $\bar{Y}[m]$ the sequence of Fourier coefficients of $\bar{y}(t)$. Then, the truncated version of $\bar{Y}[m]$, denoted by $\bar{Y}^M[m]$ is defined as follows:

$$\bar{Y}^M[m] =: \begin{cases} \bar{Y}[m] & \text{for } |m| \leq M, \\ 0 & \text{for } |m| > M, \end{cases}$$

and its time-domain representation is denoted by $\bar{y}^M(t)$.

Having the steady-state $\bar{y}(t)$, one can trivially compute $\bar{u}(t)$ in the time domain using $\mathcal{F}_{uy}$ and the output $\bar{z}(t)$ in the frequency domain using the following FRFs:

$$G_{zu}(j\omega) = F(j\omega I - A)^{-1}B + G,$$
$$G_{zw}(j\omega) = F(j\omega I - A)^{-1}L + H.$$

[2]The space $L_2(T)$ denotes the space of piecewise-continuous real-valued T-periodic scalar functions $y(t)$ satisfying $||y||_{L_2} < \infty$, where $||y||_{L_2}^2 \coloneqq \frac{1}{T}\int_0^T |y(t)|^2 dt$.

The following theorem shows that the MTF algorithm converges to the unique 'true' steady-state model response and gives an accuracy bound for the obtained steady-state solution.

Theorem 2.3 ([196])**.** *Under the conditions of Theorem 2.1, for any $M > 0$, there is a unique limit $\bar{y}^M$ to the sequence $y_i, i = 1, 2, \ldots$, resulting from the iterative process with truncation* (2.11a), (2.11b)*. Moreover,*

$$||\bar{y} - \bar{y}^M||_{L_2} \leq \frac{1}{1 - \gamma_{yu}K} \left(\frac{\gamma_{yu}^M \gamma_{yw} K}{1 - \gamma_{yu}K} ||w||_{L_2} + \gamma_{yw} ||w - w^M||_{L_2} \right)$$

with

$$\gamma_{yu} \coloneqq \sup_{m \in \mathbb{Z}} |G_{yu}(jm\omega)|, \quad \gamma_{yu}^M \coloneqq \sup_{m > |M|} |G_{yu}(jm\omega)|, \gamma_{yw} \coloneqq \sup_{m \in \mathbb{Z}} |G_{yw}(jm\omega)|.$$

▲

Theorem 2.3 shows that $\left\|\bar{y} - \bar{y}^M\right\|_{L_2}$ can be made arbitrarily small. Hereto, observe that, for large M, γ_{yu}^M drops to zero (thanks to the transfer from u to y being strictly proper) and that $\left\|w - w^M\right\|_{L_2}$ drops to zero by the Riesz-Fischer theorem, see, e.g., [25]. From here, also $\left\|\bar{z} - \bar{z}^M\right\|_{L_2}$ can be made arbitrarily small. Consider

$$\bar{z} - \bar{z}^M = \mathcal{F}_{zu} \circ (\bar{u} - \bar{u}^M) + \mathcal{F}_{zw} \circ (w - w^M) \tag{2.12}$$

with $\mathcal{F}_{zu}$ and $\mathcal{F}_{zw}$ being linear steady-state operators that map the inputs u and w to the steady-state output z. By the Riesz-Fischer theorem, it is clear that both $\left\|\bar{u} - \bar{u}^M\right\|_{L_2}$ and $\left\|w - w^M\right\|_{L_2}$ converge to zero for large M, which guarantees that $\left\|\bar{z} - \bar{z}^M\right\|_{L_2}$ also converges to zero for large M.

The MTF algorithm is presented in Algorithm 2.1. In lines 1 to 4, the contribution of the excitation signal $w(t)$ in the steady-state output $\bar{y}(t)$ is computed in the frequency domain and transformed into the time domain. After that, inside the while loop, first, the nonlinearity is evaluated in the time domain in line 6 and its output is transformed into the frequency domain in line 7. Next, in lines 8 and 9, the LTI dynamics are evaluated in the frequency domain and the output $y(t)$ is transformed to the time domain. After convergence of the signal $y(t)$, the model output $\bar{z}(t)$ is computed in the frequency domain in line 12 and transformed to the time domain in line 13. The termination criterion is the normalized mismatch between the Fourier coefficient Y_i and those of the previous iteration Y_{i-1}, measured in the ℓ_2 signal norm, which is defined as $||Z||_{\ell_2}^2 \coloneqq \sum_{m=-\infty}^{\infty} |Z[m]|^2$. The MTF algorithm with truncation can be considered as a multiharmonic variant of the describing function method, see [168] for a discussion about the *autonomous* case. The computed steady-state response $\bar{z}(t, \theta)$ can be used together with the measured response $\bar{x}(t)$ to compute the output error (2.4), which can be subsequently used in (2.5) and (2.8) in an optimization routine to solve the underlying identification problem.

Algorithm 2.1 Mixed-Time-Frequency Algorithm

1: Calculate $W[m]$ of $w(t)$ for $|m| \leq M$ using the FFT.
2: Evaluate the LTI dynamics in the frequency domain $Y_0[m] = G_{yw}(jm\omega)W[m]$ for $|m| \leq M$.
3: Compute $y_0(t)$ of $Y_0[m]$ using the IFFT.
4: Set the iteration counter $i = 0$.
5: **while** $||Y_i - Y_{i-1}||_{\ell_2}/||Y_{i-1}||_{\ell_2} > Y^\star$ **do**
6: Evaluate the nonlinearity in the time domain $u_{i+1}(t) = -\varphi(y_i(t), w(t))$.
7: Compute $U_{i+1}[m]$ of $u_{i+1}(t)$ using the FFT.
8: Evaluate the LTI dynamics in the frequency domain $Y_{i+1}[m] = G_{yu}(jm\omega)U_{i+1}[m] + Y_0[m]$ for $|m| \leq M$.
9: Compute $y_{i+1}(t)$ of $Y_{i+1}[m]$ using the IFFT.
10: Set $i = i + 1$.
11: **end**
12: Evaluate the LTI dynamics in the frequency domain $Z[m] = G_{zw}(jm\omega)W[m] + G_{zu}(jm\omega)U_i[m]$ for $|m| \leq M$.
13: Compute $z(t)$ of $Z[m]$ using the IFFT.

2.5.2 Gradient computation

Any gradient-based optimization routine requires the gradient of the cost function with respect to the model parameters as in (2.8). It is well known that this gradient can be computed by simulation of a parameter sensitivity model [134, 272].

The sensitivity model is a continuous-time model with a similar structure as the proposed model structure. Therefore, in other works, its response is computed in the same way as the model response, which is typically done by numerical forward integration. We show that the sensitivity model in our identification problem is again a *convergent* Lur'e-type model satisfying the conditions of Theorem 2.1, which facilitates the use of the efficient MTF algorithm to compute its steady-state response, i.e., the gradient information of the steady-state model response with respect to the parameters. The methodology extends that in [196], where the to-be-optimized parameters only appeared in the static nonlinear block, whereas in the more generic case in (2.2), the parameters appear in both the LTI and nonlinear blocks. The theorem below presents the parameter sensitivity model, where $\bar{x}(t,\theta), \bar{u}(t,\theta)$ and $\bar{y}(t,\theta)$ of (2.2) enter as inputs. Partial derivatives with respect to θ are denoted by subscript θ, e.g., $x_\theta := \partial x/\partial\theta$.

Theorem 2.4. *Consider model* (2.2)*. Under the conditions of Theorem 2.1, if the partial derivative* $\partial\varphi(y,w,\theta)/\partial y$ *exists and is continuous for* $y \in \mathbb{R}$, $\theta \in \Theta$*, and if the partial derivatives* $\varphi_\theta(y,w,\theta)$, $A_\theta(\theta)$, $B_\theta(\theta)$, $C_\theta(\theta)$, $D_\theta(\theta)$, $F_\theta(\theta)$, $G_\theta(\theta)$, $H_\theta(\theta)$, $L_\theta(\theta)$ *exist and are continuous for* $y \in \mathbb{R}$, $\theta \in \Theta$*, then the partial derivative*

$\partial\bar{\epsilon}(t,\theta)/\partial\theta_i$ is the unique T-periodic steady-state output $\bar{z}_{\theta_i}(t,\theta)$ of the model:

$$\begin{aligned}
\dot{x}_{\theta_i}(t,\theta) &= A(\theta)x_{\theta_i}(t,\theta) + B(\theta)\tilde{U}(t,\theta) + W(t,\theta),\\
y_{\theta_i}(t,\theta) &= C(\theta)x_{\theta_i}(t,\theta) + Y(t,\theta),\\
z_{\theta_i}(t,\theta) &= F(\theta)x_{\theta_i}(t,\theta) + G(\theta)\tilde{U}(t,\theta) + Z(t,\theta),\\
\tilde{U}(t,\theta) &= -\varphi_y\left(\bar{y}(t,\theta), w(t), \theta\right) y_{\theta_i}(t,\theta),
\end{aligned} \tag{2.13}$$

where

$$\begin{aligned}
W(t,\theta) =& A_{\theta_i}(\theta)\bar{x}(t,\theta) + B_{\theta_i}(\theta)\bar{u}(t,\theta) + L_{\theta_i}(\theta)w(t) - B(\theta)\varphi_{\theta_i}(\bar{y}(t,\theta), w(t), \theta),\\
Y(t,\theta) =& C_{\theta_i}(\theta)\bar{x}(t,\theta) + D_{\theta_i}(\theta)w(t),\\
Z(t,\theta) =& F_{\theta_i}(\theta)\bar{x}(t,\theta) + G_{\theta_i}(\theta)\bar{u}(t,\theta) + H_{\theta_i}(\theta)w(t) - G(\theta)\varphi_{\theta_i}(\bar{y}(t,\theta), w(t), \theta).
\end{aligned} \tag{2.14}$$

Furthermore, (2.13) *satisfies the conditions of Theorem 2.1 with $\mathcal{A}, \mathcal{B}, \mathcal{C}, \phi(\underline{y}, w)$ replaced by $A(\theta), B(\theta), C(\theta)$, $\varphi_y\left(\bar{y}(t,\theta), w(t), \theta\right) y_{\theta_i}(t,\theta)$, respectively, and is convergent according to Definition 2.1.* ▲

Proof. The proof can be found in Appendix A.2. □

The above theorem shows that the gradient $\partial\bar{\epsilon}(t,\theta)/\partial\theta_i$, for $i = 1, \ldots, n_\theta$, is the steady-state response of a convergent Lur'e-type model satisfying the conditions of Theorem 2.1. Hence, its steady-state response can be computed using the MTF algorithm for every parameter θ_i in θ individually. We emphasize that the gradient information can be computed with arbitrary accuracy using the MTF algorithm, which can positively affect the convergence speed of the gradient-based search used to solve the identification problem in (2.6).

Remark 2.4. Once the steady-state output $\bar{z}(t,\theta)$ of (2.2) is computed by the MTF algorithm, the steady-state state $\bar{x}(t,\theta)$, required in (2.14), is computed in the frequency-domain from the linear part of (2.2). △

2.6 Simulation case study: double mass-spring-damper system

2.6.1 True system and considered model class

To illustrate the effectiveness of the identification approach, we consider a simulation example using the double mass-spring-damper system depicted schematically in Figure 2.2. The dynamics are described by the Lur'e-type system in (2.1) with

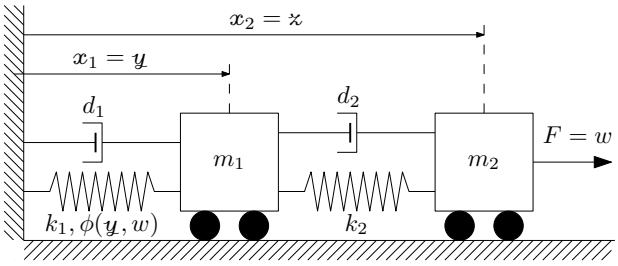


Figure 2.2. Schematic view of the mass-spring-damper system.

the matrices:

$$\mathcal{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{-(k_1+k_2)}{m_1} & \frac{-(d_1+d_2)}{m_1} & \frac{k_2}{m_1} & \frac{d_2}{m_1} \\ 0 & 0 & 0 & 1 \\ \frac{k_2}{m_2} & \frac{d_2}{m_2} & \frac{-k_2}{m_2} & \frac{-d_2}{m_2} \end{bmatrix}, \mathcal{B} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathcal{L} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \mathcal{C} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}^\top,$$
$$\mathcal{F} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}, \mathcal{D} = \mathcal{G} = \mathcal{H} = 0, \tag{2.15}$$

and the nonlinearity:

$$\phi(y, w) = \sum_{i=1}^{3} \alpha_i \tanh(\beta_i y + \gamma_i) - \alpha_i \tanh(\gamma_i). \tag{2.16}$$

The static nonlinearity has the physical interpretation of a nonlinear spring between the fixed earth and the first cart. The system is characterized by masses $m_1 = 0.15$ kg, $m_2 = 0.45$ kg, linear damping constants $d_1 = d_2 = 0.4$ Nm/s, linear spring constants $k_1 = 1000$ N/m, $k_2 = 2200$ N/m, and parameters defining the nonlinear spring $\phi(y, w)$ in (2.16): $\alpha_1 = 1, \alpha_2 = 0.5, \alpha_3 = 1, \beta_1 = 10, \beta_2 = 20, \beta_3 = 20, \gamma_1 = 3, \gamma_2 = -3, \gamma_3 = 0$. The excitation w is the force exerted on the second cart and the output z is the position of the second cart. The Bode magnitude diagram of the transfer functions of the LTI part of the true system and the graph of the nonlinearity are given in Figure 2.4.

The goal is to identify a black-box Lur'e-type model for this system. Hereto, consider the model class in (2.2) with state dimension $n = 4$ and consider a full parametrization for the LTI part, i.e., each element of each model matrix is a model parameter. The nonlinearity is characterized as follows:

$$\varphi(y, w, \theta) = W(\theta)^{[2]^\top} \left(\tanh\left(W(\theta)^{[1]} y + b(\theta) \right) - \tanh(b(\theta)) \frac{1}{2} W(\theta)^{[1]} \right), \tag{2.17}$$

where each element of the vectors $W^{[1]}(\theta), W^{[2]}(\theta), b(\theta) \in \mathbb{R}^3$ is a model parameter. This nonlinear function represents a specific type of a single hidden layer

feedforward neural network with hyperbolic tangent activation functions and a linear output layer [272]. This model is characterized by $n_\theta = 44$ parameters, which we collect in the vector $\theta \in \mathbb{R}^{n_\theta}$. It can be shown that the conditions of Theorem 2.1 are satisfied for any $\theta \in \Theta$ with

$$\Theta \coloneqq \Bigg\{ \theta \in \mathbb{R}^{n_\theta} \mid \lambda\left(A(\theta)\right) \in \mathbb{C}_{<0}, W^{[1]}(\theta), W^{[2]}(\theta) \in \mathbb{R}^3_{\geq 0}, \\ \sup_{\omega \in \mathbb{R}} \left| C(\theta)(j\omega I - A(\theta))^{-1} B(\theta) \right| < \frac{1}{K(\theta)} \Bigg\},$$

where $\lambda(A(\theta))$ denotes the eigenvalues of the matrix $A(\theta)$, $\mathbb{C}_{<0}$ denotes the set of complex number with negative real part and $K(\theta) \coloneqq \frac{1}{2} W(\theta)^{[2]^\top} W(\theta)^{[1]}$. For any $\theta \in \Theta$, the proposed model is globally exponentially convergent according to Definition 2.1. The true system parametrized by θ_0 is in the model class, i.e., $\theta_0 \in \Theta$.

2.6.2 Identification results

The excitation signal w is a random-phase multisine that excites frequencies between 0.5 Hz and 200 Hz with 0.5 Hz spacing and is sampled at 500 Hz, implying a period of $T = 2$ seconds. Following Assumption 2.2, discrete-time white noise is added to the steady-state output $\bar{x}_0$, where the variance σ_e^2 is selected such that a signal-to-noise (SNR) of $20, 40, 60$ and ∞ dB is realized. The SNR of ∞ corresponds to a deterministic setting. To find an initial parameter set θ_{init}, we employ the so-called controlled random search (CRS) [112], which is a global, non-gradient-based optimization routine. In accordance to Remark 2.3, we terminate the CRS prematurely. In the second identification step, we perform a gradient-based search using Matlab's `fmincon` implementation of [45] to solve the constrained optimization problem (2.6).

Figure 2.3 plots the history of the cost J_N over the first 200 optimization iterations. Here, it can be observed that the cost function (non-monotonically) decreases over the iterations. The non-monotonicity is a consequence of avoiding the so-called Maratos effect close to constraint infeasibility [45]. Table 2.1 gives the numerical results. Here, it can be seen that the cost function is successfully minimized to $J_N(\hat{\theta}_N) \approx \sigma_e^2$. For the case of a SNR of 60 dB, Figure 2.4 presents the Bode magnitude diagram of the LTI part and the graph of the nonlinearity of the true system (θ_0), initial model (θ_{init}), and the identified model ($\hat{\theta}_N$). This figure evidences an accurate match between the identified model and the true system. For the 20 dB SNR case, the time-domain response is depicted in Figure 2.5. In the top plot, it can be observed that the steady-state error $\bar{\epsilon}(t, \hat{\theta}_N)$ is significantly smaller than $\bar{\epsilon}(t, \theta_{init})$. The bottom plot displays a validation test, where a new realization of the excitation signal w is used that is 10 times larger in amplitude than the w that is used in identification. This validation test shows that the model accurately

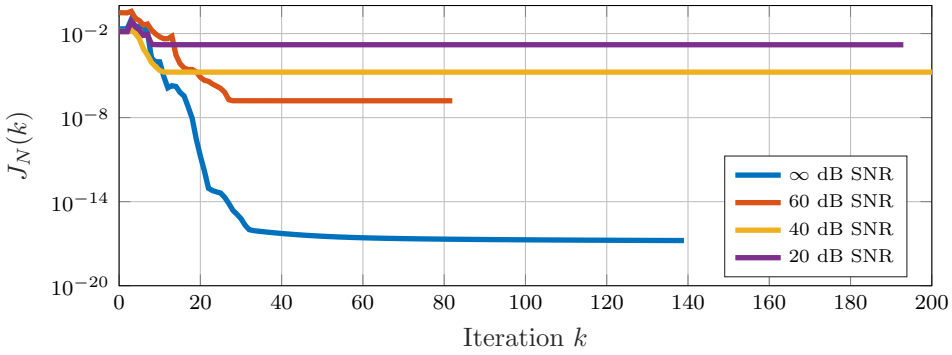


Figure 2.3. The history of J_N over the first 200 iterations k starting from the initial parameter set θ_{init} and terminating at the identified parameter set $\hat{\theta}_N$.

Table 2.1. Identification results for different SNRs.

SNR [dB]	σ_e^2	$J_N(\theta_{init})$	$J_N(\hat{\theta}_N)$
20	1.55×10^{-3}	2.75×10^{-2}	1.62×10^{-3}
40	1.74×10^{-5}	1.45×10^{-2}	1.76×10^{-5}
60	1.57×10^{-7}	3.08×10^{-1}	1.61×10^{-7}
∞	0	2.16×10^{-2}	1.67×10^{-17}

describes the steady-state response of the true system. We would like to note that all identified models are guaranteed globally exponentially convergent and, therefore, these identified models can safely be used for other excitation signals than the one used in identification, without showing instability issues.

2.6.3 Numerical efficiency

To illustrate the numerical efficiency[3] of our method, we consider the deterministic scenario with an SNR of ∞, i.e., $\sigma_e^2 = 0$ in Assumption 2.2. The identification strategy turns out to yield an 'almost' perfect model for the true system evidenced by the cost $J_N(\hat{\theta}_N)$ being minimized up to Matlab's numerical 16-digit precision. In both identification steps, we compute the steady-state model responses $\bar{z}$ and the gradient $\bar{z}_\theta$ by two methods, namely by the MTF algorithm presented in Section 2.5.1 and by numerical forward integration (NFI). To guarantee a similar level of accuracy as in MTF, we simulate NFI for 20 periods and take the last period as steady-state. The result is presented in Table 2.2. It can be observed

[3]The computations of the examples in Section 2.6 and 2.7 are carried out on an Intel Core i7-7700HQ, 2.8GHz processor.

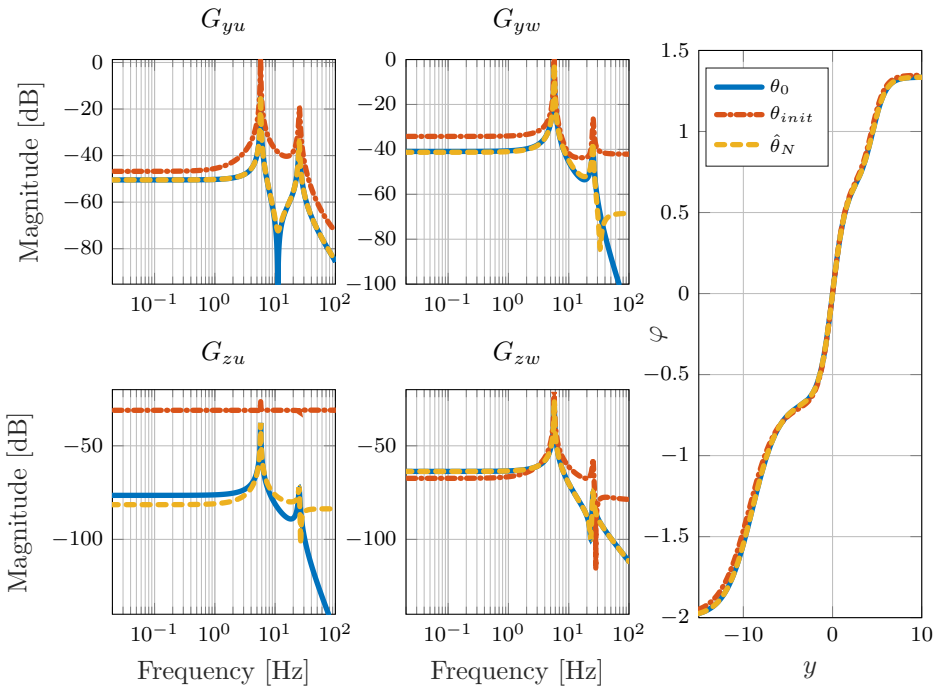


Figure 2.4. Bode magnitude diagrams of the involved transfer functions and the graph of the nonlinearity of the true system (θ_0), initial model (θ_{init}) and identified model ($\hat{\theta}_N$). For visualization, the gain exchange between the LTI block and nonlinearity is fixed by normalization of $W^{[1]}, W^{[2]}$ and a loop transformation is performed [135].

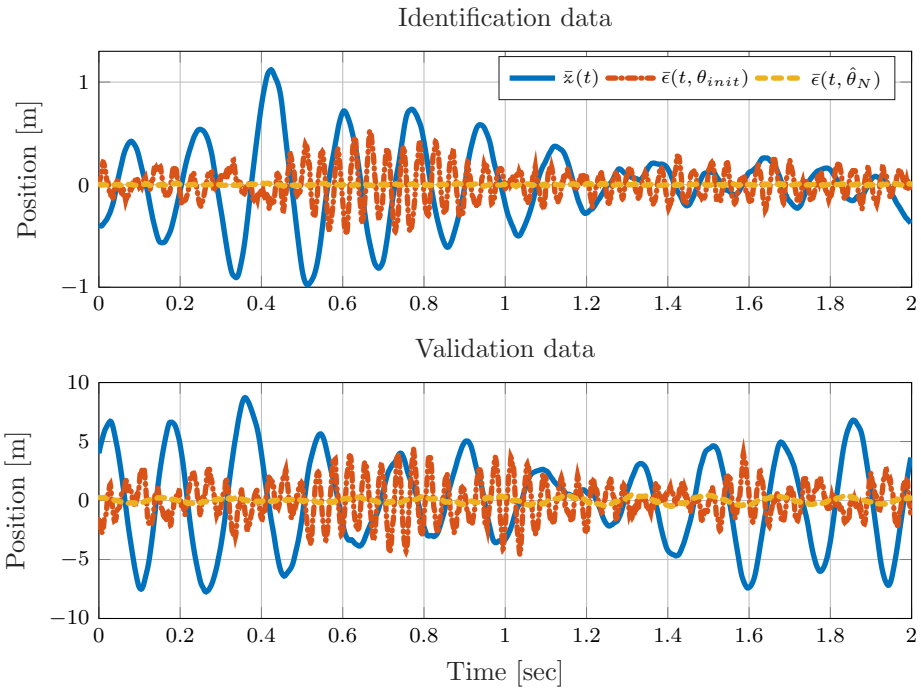


Figure 2.5. Measured $\bar{z}$ and the errors $\bar{\epsilon}$ as defined in (2.4). Top: results for the identification excitation signal. Bottom: results for the validation excitation signal.

Table 2.2. Comparison of computation time between MTF and NFI for both identification steps. The number of computed steady-state model responses is included in the second column.

Step	Model response	MTF [sec]	NFI [sec]
1: Controlled random search	935	32	5927
2: Gradient-based search	8450	196	28878
Total	9385	228	34805

that in the first identification step, where no gradient information is required, the MTF algorithm reduces the computation time from over 1 and a half hours to 32 seconds. In the second identification step, where also gradients have to be computed, the MTF algorithm reduces the computation time from over 9 and a half hours to only 3 minutes and 16 seconds. In total, the computation time is reduced by a factor of approximately 150. Such small computation time enables the application of global optimization routines for initialization and also enables an efficient and effective gradient-based search. Besides highlighting the numerical benefits of our approach, this example also illustrates one of the reasons that the literature around the identification of nonlinear models is mainly focused on the discrete-time setting; namely, computing steady-state model responses and gradient information is computationally prohibitively expensive in the continuous-time case using NFI.

2.7 Experimental case study: mechanical ventilation

The identification strategy developed in this chapter is applied to a mechanical ventilation setup as depicted in Figure 2.6. Mechanical ventilation is used in intensive-care units to assist or stimulate the respiration of patients who are unable to breathe on their own. The treatment quality of mechanical ventilation depends on two aspects. The first aspect is the design of the breathing pressure profile that is fed to the patient [5, 145]. The second aspect is the ability to track the designed breathing pressure profile. Hereto, in the literature, pressure control strategies are proposed to ensure comfortable and stable air flow [39, 116, 211, 275]. In both these aspects, knowledge of the health condition of the patient's lungs is crucial as incorrect treatment can lead to lung damage and, eventually, to patient mortality [90, 258]. Therefore, by identification of a first-principle model, we aim to provide the patient's health information reflected by the patient's lung parameters in a *fast* way.

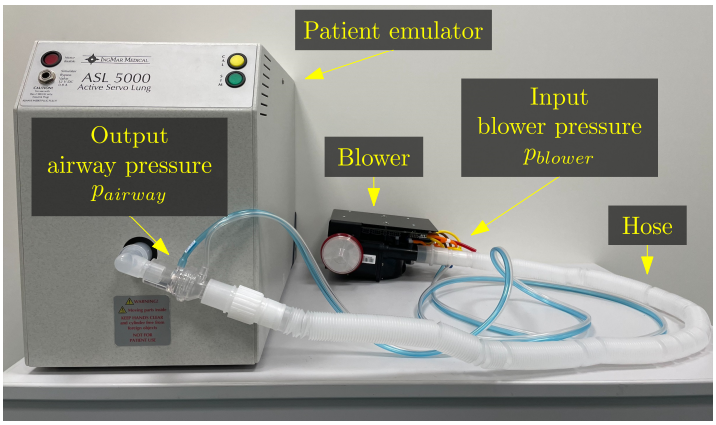


Figure 2.6. Experimental mechanical ventilation setup.

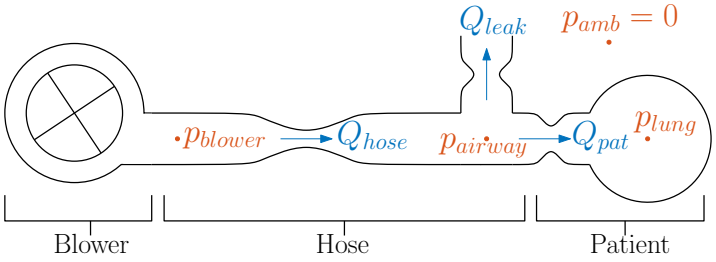


Figure 2.7. Overview of the mechanical ventilation setup.

2.7.1 First-principle modeling and conditions for convergent dynamics

Figure 2.7 gives a schematic view of the setup. Using conservation of flow and taking into account a (signed) quadratic hose resistance, the patient-hose dynamics can be written as follows [109, 287]:

$$\begin{aligned}
\dot{x} &= -\frac{1}{\theta_2\,(\theta_1+\theta_5)}x - 2\frac{\theta_5}{\theta_2\,(\theta_1+\theta_5)}u,\\
y &= -\frac{\theta_5}{\theta_1+\theta_5}x + w,\\
z &= \frac{\theta_5}{\theta_1+\theta_5}x - 2\frac{\theta_1\theta_5}{\theta_1+\theta_5}u,\\
u &= -\varphi(y,\theta),
\end{aligned} \tag{2.18}$$

where the scalar nonlinearity is given by

$$\varphi(y,\theta) = \frac{y}{\theta_3 + \frac{\theta_1\theta_5}{\theta_1+\theta_5} + \sqrt{\left(\theta_3 + \frac{\theta_1\theta_5}{\theta_1+\theta_5}\right)^2 + 4\theta_4|y|}}.$$

The excitation signal w is the pressure p_{blower} generated by the blower, the output z is the airway pressure p_{airway}, the state x represents the lung pressure p_{lung}, the output of the nonlinearity u is a scaled flow Q_{hose} through the hose, and, finally, the input y of the nonlinearity is a scaled pressure drop $p_{airway} - p_{blower}$ over the hose. The intended leakage flow Q_{leak} flushes the system from CO_2-rich exhaled air. The model is parametrized by the lung resistance θ_1 in mbar sec/ml, lung compliance θ_2 in ml/mbar, the linear hose resistance θ_3 in mbar sec/ml, the quadratic hose resistance θ_4 in mbar sec^2/ml^2 and the linear leakage resistance θ_5 in mbar sec/ml.

For any $\alpha > 0$, the following nonlinear scaling of the parameters leaves the steady-state response $\bar{z}$ of model (2.18) unchanged for any bounded excitation w [287]:

$$\theta_{scaled} = \begin{bmatrix} \alpha\theta_1 & \theta_2/\alpha & \alpha\theta_3 & \alpha^2\theta_4 & \alpha\theta_5 \end{bmatrix}^\top.$$

This scaling corresponds to a gain exchange between the LTI part of the model and the nonlinearity φ, see Remark 2.2. Consequently, parameters $\theta_1, \ldots, \theta_5$ cannot be uniquely identified on the basis of w and $\bar{z}$ only. As a remedy, we consider the parameter θ_5 to be known *a priori* and positive in value, since it is only leakage specific, i.e., independent of the patient and the hose, and it can be found relatively easily through calibration. Given the knowledge on θ_5, the to-be-identified parameters are $\theta_1, \ldots, \theta_4$, which are collected in θ. It is shown in [287] that model (2.18) satisfies all conditions of Theorem 2.1 for $\theta_i > 0$, for $i = 1, \ldots, 4$, yielding the set

$$\Theta := \{\theta \in \mathbb{R}^4 \mid \theta_i > 0 \text{ for } i = 1, \ldots, 4\}. \tag{2.19}$$

Therefore, for any $\theta \in \Theta$, model (2.18) is globally exponentially convergent according to Definition 2.1.

2.7.2 Identification experiment

Identification of continuous-time models from discrete data points is a challenging task due to the unobserved intersample behavior of the true system. Analysis tools for errors due to intersample behavior and practical guidelines for the design of the excitation signal are presented [99, 207, 236, 260]. As we are dealing with a medical application, there is no freedom in the design of the excitation signal; it is the signal that is also used during conventional operation. We can only set a standard *target* pressure profile p_{target} for the blower, which is tracked by the blower (using an internal control loop) to realize the pressure p_{blower} at the blower side of the hose, see Figure 2.8. The shape of the blower pressure represents a breathing cycle that

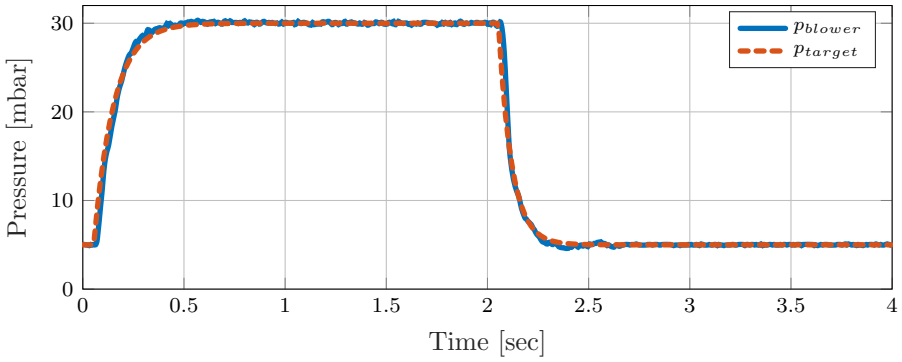


Figure 2.8. Realized (measured) and target breathing pressure.

starts with an inhalation phase and follows by an exhalation phase. The measured pressure p_{blower} is used as the excitation signal $w(t)$ for identification, implying that the internal control loop generating p_{blower} from p_{target} is not part of the to-be-identified dynamics. The period time of p_{target} is $T = 4$ seconds, with a minimum pressure of 5 mbar and a maximum pressure of 30 mbar. Rather than using humans in these experiments, the ASL5000 breathing simulator is utilized, which is specifically designed to emulate the lung behavior of patients. Furthermore, we consider the case where the patient is completely sedated, implying no breathing activity from the patient's side.

We apply 15 periods of the excitation signal depicted in Figure 2.8 and sample the realized blower pressure p_{blower} and the patient's airway pressure p_{airway} uniformly at a sampling frequency of 500 Hz. The average of the last 12 periods of the measured p_{blower} is considered as the periodic input data $w(t)$ and the average of the last 12 periods of p_{airway} is considered as the steady-state output data $\bar{z}(t)$, both with a period time of $T = 4$ seconds. The total experiment time is *only* 1 minute for each experiment, which is of crucial importance in practice as time is costly in this medical application.

The three hose-patient configurations considered are listed in Table 2.3, where the lung parameters θ_1 and θ_2 are set on the ASL5000 unit with an accuracy[4] of 10% and 5%, respectively. The lung resistance θ_1 remains unchanged over the experiments. To cover a set of different configurations, both the lung compliance θ_2 and the leakage resistance θ_5 are changed from experiment to experiment. The same hose with parameters θ_3 and θ_4 is used for all experiments. The considered configurations represent a healthy patient for configurations 1 and 3 and a patient with less stiff lungs for configuration 2.

[4]https://www.ingmarmed.com/product/asl-5000-breathing-simulator/

Table 2.3. Considered experimental cases with uncertain parameters θ_1 and θ_2.

Configuration	$\theta_1 \cdot 10^3$	θ_2	$\theta_3 \cdot 10^3$	$\theta_4 \cdot 10^6$	$\theta_5 \cdot 10^2$
1	$4.5 - 5.5$	$19 - 21$	2.6	1.23	2.8
2	$4.5 - 5.5$	$28.5 - 31.5$	2.6	1.23	2.8
3	$4.5 - 5.5$	$19 - 21$	2.6	1.23	2.1

2.7.3 Identification results

The identification problem in (2.6) is solved using the modified cost function in (2.9) with the constraints (2.19), resulting in $\hat{\theta}_N$. The set of initial parameters reflects an 'average' configuration, see [123, 287]:

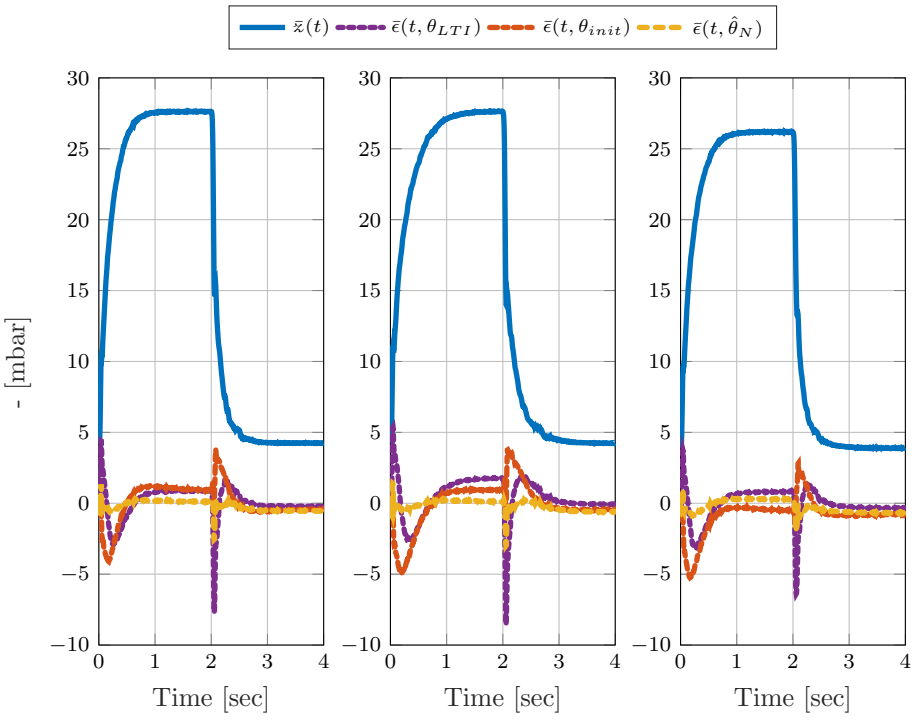
$$\theta_{init} = \begin{bmatrix} 2.75 \cdot 10^{-2} & 26.50 & 2.55 \cdot 10^{-3} & 2.40 \cdot 10^{-6} \end{bmatrix}^\top . \tag{2.20}$$

This initialization method employs the prior knowledge that parameters are typically close to the 'average' values.

The measured response $\bar{z}(t)$ for configurations 1, 2 and 3 is shown in Figures 2.9(a) - 2.9(c), respectively. These figures also depict the steady-state output error $\bar{\epsilon}(t, \theta_{init})$ and $\bar{\epsilon}(t, \hat{\theta}_N)$, produced by the initial and identified model, respectively, with $\bar{\epsilon}$ defined in (2.4). For comparison, also an LTI transfer-function model is identified using the `TFEST` Matlab routine, initialized by the `N4SID` routine. The parameters θ_{LTI} corresponding to the identified LTI models have no physical meaning. The steady-state output error $\bar{\epsilon}(t, \theta_{LTI})$ is also included in the respective figures.

From Figures 2.9(a) - 2.9(c), we draw the following two conclusions. Firstly, the optimization problem (2.6) is successfully solved, as the error $\bar{\epsilon}(t, \hat{\theta}_N)$ of the identified nonlinear model is significantly smaller than the error $\bar{\epsilon}(t, \theta_{init})$ of the initial model. Secondly, a linear model is insufficient to describe the nonlinear dynamics of this system, since the error $\bar{\epsilon}(t, \hat{\theta}_N)$ of the identified nonlinear model is significantly smaller than the error $\bar{\epsilon}(t, \theta_{LTI})$ of the identified linear model. Both these conclusions are confirmed by Table 2.4, which shows that the cost $J_N(\hat{\theta}_N)$ of the identified nonlinear model is significantly smaller than both the costs $J_N(\theta_{init})$ and $J_N(\theta_{LTI})$.

The identified parameters $\hat{\theta}_N$ obtained by solving the optimization problem in (2.6) are presented in Table 2.5. The model (2.18) is a simplification of the system depicted in Figure 2.6. For example, calibration experiments suggest that the leakage component exhibits a quadratic pressure-flow relation, whereas this relation is linear in model (2.18). Furthermore, the static hose model in model (2.18) does not account for hose *dynamics*. Consequently, a part of the unmodeled dynamics is compensated for by the identified model parameters, which is reflected as follows. Firstly, although the same hose is used in all configurations, the results



(a) Configuration 1. **(b)** Configuration 2. **(c)** Configuration 3.

Figure 2.9. Measured $\bar{x}$ and the errors $\bar{\epsilon}$ as defined in (2.4).

Table 2.4. Identification results of the mechanical ventilation application in terms of the value of the cost function (2.5) for the identified black-box LTI model (θ_{LTI}), initial model (θ_{init}) in (2.20) and identified model ($\hat{\theta}_N$) in columns 2-4. Columns 5-6 contain the elapsed computation time in seconds that took to solve the identification problem (2.6).

Configuration	Value of cost function			Computation Time	
	$J_N(\theta_{LTI})$	$J_N(\theta_{init})$	$J_N(\hat{\theta}_N)$	MTF [sec]	NFI [sec]
1	1.377	1.661	0.167	13.3	166.1
2	2.396	2.549	0.173	11.7	149.4
3	1.323	2.177	0.275	12.1	158.2

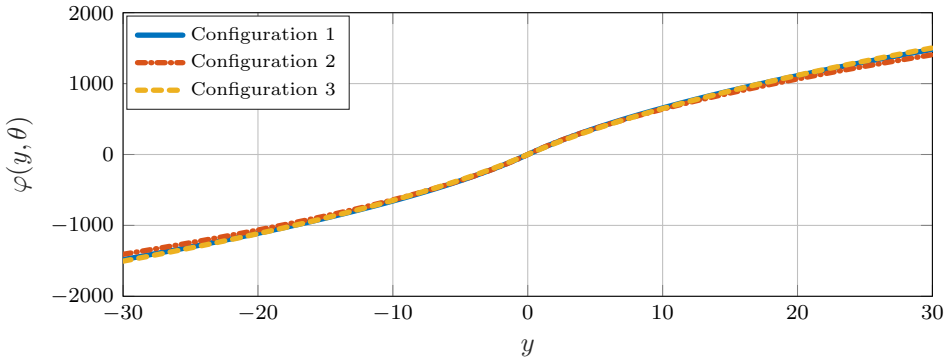


Figure 2.10. Identified nonlinearity $\varphi(y)$ for each configuration in Table 2.5 over the domain of interest in y.

Table 2.5. Identification results of the mechanical ventilation application in terms of the identified parameters $\hat{\theta}_N$ after solving the identification problem (2.6).

Configuration	Identified parameters $\hat{\theta}_N$				
	$\theta_1 \cdot 10^3$	θ_2	$\theta_3 \cdot 10^3$	$\theta_4 \cdot 10^6$	$\theta_5 \cdot 10^2$
1	5.48	23.6	1.05	1.51	2.8
2	5.78	34.6	0.65	1.85	2.8
3	5.43	22.0	1.69	1.32	2.1

in Table 2.5 show that its associated parameters θ_3 and θ_4 are different with respect to each other and also with respect to the ones in Table 2.3. However, the graph of the identified nonlinearities in each configuration in Figure 2.10 shows that all identified models represent the same hose characteristics in the domain of interest. Besides that, the values for θ_3 and θ_4 in Table 2.3 are obtained by a *static* calibration measurement, which is not fully representative in case the hose is used in a *dynamic* identification experiment. Secondly, the identified parameters in Table 2.5 show that θ_1 and θ_2 are estimated slightly too large compared to the preset values as in Table 2.3. We are informed by medical personnel that an accuracy of 15% is sufficient for the lung parameters in the scope of the decision-making process of patient treatment. If the mean of the uncertain θ_1 and θ_2 in Table 2.3 are the true values, then we slightly exceed this accuracy requirement. We would like to note that a bias in model parameters is not a deficiency of the identification approach, but rather a matter of unmodelled dynamics.

2.7.4 Numerical efficiency

To illustrate the computational advantages of the proposed approach, the total computation time required to solve the optimization problem (2.6) is included in Table 2.4 in the column MTF. This table also includes the total computation time required to solve the same optimization problem with numerical forward integration (NFI). To guarantee a similar accuracy level for NFI as for MTF, ten periods of the excitation signal $w(t)$ are applied to the model in the NFI simulations and the last period of $z(t, \theta)$ is taken as the steady-state model output $\bar{z}(t, \theta)$. Table 2.5 shows that the total computation time when using the MTF algorithm is reduced from roughly 2 and a half minutes to only 14 seconds in all considered configurations. The quick availability of the knowledge of the model parameters enables a high treatment quality that can start early.

2.8 Conclusions

This chapter has presented an identification approach for convergent continuous-time Lur'e-type systems. The benefits of the proposed approach are that it (i) guarantees that the identified model preserves the convergence property, (ii) is computationally attractive, and, (iii) applies to a large class of block-oriented feedback systems. In a simulation example, we demonstrated the effectiveness and computational efficiency of our identification approach. Furthermore, in an experimental study on mechanical ventilation in hospitals, the identification approach is effective to identify the parameters of a first-principle model in a fast way, which enables improved patient treatment.

3

Fast identification of discrete-time Lur'e-type systems with stability guarantees

This chapter proposes a computationally efficient approach for the system identification of discrete-time nonlinear Lur'e-type models with stability certification. Lur'e-type models consist of a feedback interconnection of linear time-invariant dynamics and static nonlinearities. By enforcing the identified Lur'e-type model to be inside the set of convergent models, we certify a strong form of stability on the identified model. This certificate allows for reliable usage of model predictions, also for excitation signals other than those used during model training. Furthermore, both steady-state model responses and their gradients with respect to model parameters are computed with a numerical algorithm that is computationally fast. This allows for the effective usage of both global and local optimization methods to solve the system identification problem. The effectiveness and the benefits of the approach are validated via the nonlinear Wiener-Hammerstein benchmark and Silverbox benchmark datasets.

The contents of this chapter are published in: Shakib, M. F., Vervaet, N., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Fast identification of multivariable discrete-time Lur'e-type models with stability certification. *Submitted for journal publication.*

3.1 Introduction

System identification turns experimentally obtained input-output datasets from dynamical systems into compact models described by only a few parameters [152, 205]. Given that the vast majority of physically relevant systems exhibit nonlinear dynamic behavior, there is a strong need for the identification of nonlinear models [240]. Hereto, many nonlinear system identification approaches have been proposed in the literature, among which methods for the family of block-oriented nonlinear models [96, 243]. These models capture nonlinear dynamic behavior by an interconnection of LTI dynamics and static nonlinearities, thereby, enabling a trade-off between model complexity and representation capability of identified models [135, 191]. System identification of Lur'e-type models, see Figure 3.1, is an active research area, see [96, 115, 188, 214, 234, 243, 247, 249, 291, 294, 296].

One of the major challenges in nonlinear system identification is to enforce a form of stability on the identified model, see [154, 249, 280, 283]. Even though stable responses are observed for the true system, the responses of a model resulting from an identification routine may not necessarily obey this property. It is also well-known that nonlinear models can exhibit multiple stable solutions being attractive for different sets of initial conditions [135]. Moreover, the model response may exhibit a large sensitivity to variations in the excitation signal. For example, even though the model explains the identification dataset perfectly well, the model response to an altered excitation (even when it is only slightly different from those used during training) can result in unbounded responses [60, 64]. Such models are not capable of accurately predicting the system response in scenarios other than the exact identification dataset.

Another major challenge in system identification is the computational efficiency of the identification procedure. These kinds of procedures are reported as extremely time-consuming, especially for problems with a large number of model parameters [3, 315]. Typically, the identification problem is cast into a non-convex optimization problem, which is generally solved either by global optimization routines or by gradient-based optimization [193]. Either way, a large number of model responses has to be computed to scan the surface of the objective function central in the optimization problem or to compute gradient (and Hessian) information of the objective function with respect to the model parameters. Therefore, computational efficiency boils down to the efficiency of computing model responses and the efficiency and accuracy of computing gradient information.

Methods that enforce a form of model stability, e.g., asymptotic stability of the origin of *autonomous models*, can be found in [68, 171, 270, 271, 285]. Besides that, methods that enforce a stronger form of model stability and, therefore, deal with the above-listed challenge of enforcing a strong form of model stability, are reported in [214, 247, 249, 280, 283]. The approaches in [280, 283] enforce *incremental stability* on a particular class of identified models. These approaches require a dataset that includes the state of the underlying dynamics, which is

typically hard to obtain. The method in [214] writes recurrent neural network models as Lur'e-type models and describes convex sets of *contractive* models that attain performance criteria expressed via integral quadratic constraints. Lastly, the method in [247, 249] considers *continuous-time* Lur'e-type models and deals with both the above-listed challenges of enforcing model stability, as well as computational efficiency. The identification of *discrete-time* Lur'e-type models in a *computational efficient* manner with a strong form of *model stability* is still an open problem.

The main contribution of this chapter is an identification approach that (i) certifies the identified discrete-time Lur'e-type model with the global exponential convergence property and (ii) proposes efficient numerical tools to compute model responses and exact gradient information, while only using steady-state input and output data. Point (i) guarantees that the identified model exhibits, for any bounded excitation, a bounded and globally exponentially stable steady-state solution [66]. Steady-state responses arise naturally from periodic excitations, commonly used in the identification of (non)linear system [205, 240]. To enforce the convergence property, conditions are presented in the form of linear matrix inequalities (LMIs). Point (ii) enables system identification in a computationally efficient way. Hereto, a novel and computationally efficient numerical algorithm is devised that computes steady-state model responses. This algorithm is the discrete-time counterpart of the so-called mixed time-frequency (MTF) algorithm, presented in [196]. Moreover, a method to efficiently compute exact gradient information is presented. The identification methodology is demonstrated using the Wiener-Hammerstein benchmark dataset and the Silverbox benchmark dataset. The identification approach can be viewed as the discrete-time counterpart of the method proposed in Chapter 2 with an extended model class such that multivariable systems with multiple nonlinear functions can be considered.

The remainder of this section introduces preliminaries and notation. Section 3.2 introduces the considered Lur'e-type model class together with numerically tractable sufficient conditions to assess the convergence property and poses formally the identification problem. Section 3.3 proposes a solution to the identification problem and proposes tools for fast computation of steady-state model responses and exact gradient information. Section 3.4 showcases the performance of the identification method on benchmark datasets. Section 3.5 closes with the concluding remarks of this chapter.

Preliminaries and notation: The imaginary unit $j \coloneqq \sqrt{-1}$. The sets $\mathbb{R}, \mathbb{N}, \mathbb{Z}$, $\mathbb{Z}_{>0}$, and $\mathbb{C}$, denote the set of real, natural, integer, positive integer, and complex numbers, respectively. The $n \times n$ identity and zeros matrix are denoted respectively by I_n and 0_n. The $n \times m$ zeros matrix is denoted by $0_{n\times m}$. The sets $\mathbb{S}^n, \mathbb{D}^n$ respectively, denote the space of $n \times n$ real symmetric matrices and real diagonal matrices. A matrix $P \in \mathbb{S}^n$ is positive (negative) definite, denoted by $P \succ 0$ ($P \prec 0$), if all its eigenvalues have strictly positive (negative) real part. For real vectors $x \in \mathbb{R}^n$, the Euclidean norm is defined by $\|x\|_2 = \sqrt{x^\top x}$, and its P-weighted

norm $\|\cdot\|_P$ for $0 \prec P \in \mathbb{S}^n$ is defined by $\|x\|_P = \sqrt{x^\top P x}$. For scalar valued sequences $g : \mathbb{N} \mapsto \mathbb{R}$, the Banach space $\ell_2(N)$ of all N-periodic sequences having a finite ℓ_2-norm $\|\cdot\|_{\ell_2}$ is such that $\|g\|_{\ell_2}^2 = \sum_{k=1}^{N} |g(k)|^2 < +\infty$. For vector valued sequences $g : \mathbb{N} \mapsto \mathbb{R}^n$, the Banach space $\ell_2^n(N)$ of all N-periodic sequences having a finite ℓ_2^n-norm $\|\cdot\|_{\ell_2^n}$ is such that $\|g\|_{\ell_2^n}^2 = \sum_{i=1}^{n} \|g_i\|_{\ell_2}^2 < +\infty$, where g_i is the $i-$th component of g. Let system Σ define a bounded linear mapping between sequences $u \in \ell_2^{n_u}, y \in \ell_2^{n_y}$, such that $y = \Sigma u$. Then, the system's induced $\mathcal{L}_2$-gain is defined as follows:

$$\|\Sigma\|_{\mathcal{L}_2} \coloneqq \sup_{u \in \ell_2^{n_u} \setminus \{0\}} \frac{\|\Sigma u\|_{\ell_2^{n_y}}}{\|u\|_{\ell_2^{n_u}}}. \tag{3.1}$$

The N-point discrete Fourier transform (DFT) maps from x, a periodic sequence of N complex vectors of dimension n (with $x_i \in \mathbb{C}^n$ being the i-th vector in x) into X, another sequence of N complex vectors of dimension n (with $X_i \in \mathbb{C}^n$ being the i-th vector in X). This transform is denoted by the linear operator $\widehat{\mathcal{F}} : \mathbb{C}^{N\times n} \to \mathbb{C}^{N\times n}$ and its inverse (IDFT) is denoted by $\widehat{\mathcal{F}}^{-1} : \mathbb{C}^{N\times n} \to \mathbb{C}^{N\times n}$. The operators $\widehat{\mathcal{F}}$ and $\widehat{\mathcal{F}}^{-1}$ for the $i-$th vector are $X_i = \widehat{\mathcal{F}} x_i$ and $x_i = \widehat{\mathcal{F}}^{-1} X_i$, respectively, where

$$X_i(k) = \sum_{n=0}^{N-1} x_i(n) e^{-j\frac{2\pi}{N}kn} \qquad \forall k \in \{0, 1, \ldots, N-1\},$$

$$x_i(k) = \frac{1}{N} \sum_{n=0}^{N-1} X_i(n) e^{j\frac{2\pi}{N}kn} \qquad \forall k \in \{0, 1, \ldots, N-1\}.$$

3.2 Identification problem setting

This section starts with the introduction of the considered class of convergent Lur'e-type models. After that, the objective function central to the identification problem is defined. Finally, the identification problem is formalized.

3.2.1 Lur'e-type models

Consider a discrete-time Lur'e-type structure, expressed by the following state-space equations:

$$\Sigma : \begin{cases} x(k+1) &= Ax(k) + Bu(k) + Lw(k), \\ y(k) &= Cx(k) + Dw(k), \\ z(k) &= Fx(k) + Gu(k) + Hw(k), \end{cases} \tag{3.2a}$$

$$\varphi : \quad u(k) = \varphi(y(k), k), \tag{3.2b}$$

where, $x(k) \in R^{n_x}$, $z(k) \in \mathbb{R}^{n_z}$ and $y(k) \in \mathbb{R}^{n_y}$ are the state x, the output z and the nonlinearity input y, respectively, evaluated at the time-instant $k \in \mathbb{Z}$. Moreover, $u(k) \in \mathbb{R}^{n_u}$ and $w(k) \in \mathbb{R}^{n_w}$ are the evaluations of the nonlinearity output

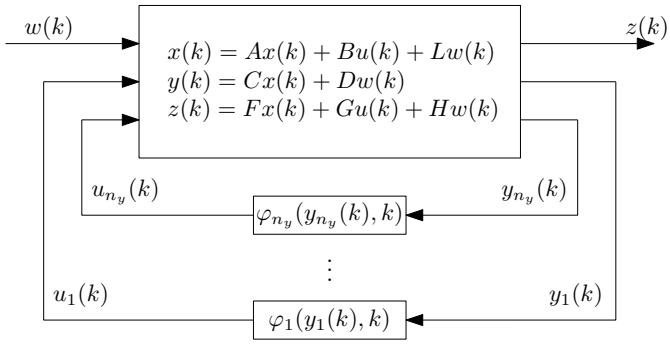


Figure 3.1. Schematic depiction of Lur'e-type model (3.2).

u and of the periodic external excitation w at time $k \in \mathbb{Z}$. The (possibly time-varying) memoryless nonlinearity is considered to be decentralized [135], implying that its i-th output is only a function of its i-th input (with $n_y = n_u$) according to:

$$\varphi(y(k), k) = \begin{bmatrix} \varphi_1(y_1(k), k) & \dots & \varphi_{n_y}(y_{n_y}(k), k) \end{bmatrix}^\top. \tag{3.3}$$

The Lur'e-type model (3.2) can be interpreted as the feedback interconnection depicted in Figure 3.1. Throughout this chapter, a discrete-time Lur'e-type model can be referred to via either (3.2), or the pair $(\Sigma,\ \varphi)$.

For the purpose of model stability, to be defined below, nonlinearity φ is required to satisfy an incremental sector-bound condition, defined as follows.

Definition 3.1. *A decentralized memoryless function $\varphi : \mathbb{R}^{n_y} \times \mathbb{Z} \to \mathbb{R}^{n_y}$ is said to be incrementally sector bounded within bounds $[S_1, S_2]$, for $S_1,\ S_2 \in \mathbb{D}^{n_y}$ with $S_2 - S_1 \succ 0$, if it satisfies the condition*

$$\begin{aligned} &\left[\varphi(y^a, k) - \varphi(y^b, k) - S_1(y^a - y^b)\right]^\top \times \\ &\quad \left[\varphi(y^a, k) - \varphi(y^b, k) - S_2(y^a - y^b)\right] \leq 0, \end{aligned} \tag{3.4}$$

for all $y^a, y^b \in \mathbb{R}^{n_y}$ and $k \in \mathbb{Z}$. ■

For the sake of clarity of exposition, we assume that the nonlinearity satisfies the incremental sector bounds $[-I_{n_y}, I_{n_y}]$, i.e., $S_1 = -I_{n_y}$ and $S_2 = I_{n_y}$. These bounds make sure that each individual nonlinearity $\varphi_i, i \in \{1, 2, \dots, n_y\}$, satisfies the inequality $\left|\frac{\varphi_i(y_i^a, k) - \varphi_i(y_b^i, k)}{y_i^a - y_i^b}\right| \leq 1 \,\forall\, y_i^a \in \mathbb{R}^{n_y}, y_i^b \in \mathbb{R}^{n_y}, k \in \mathbb{Z}$.

Remark 3.1. Any decentralized nonlinearity that satisfies the incremental sector bounds $[S_1, S_2]$, for $S_1,\ S_2 \in \mathbb{D}^{n_y}$ with $S_2 - S_1 \succ 0$ (see Definition 3.1), can be transformed such that it satisfies the incremental sector bounds $[-I_{n_y}, I_{n_y}]$. See [135] for more details on transformations of Lur'e-type models. △

3.2.2 Convergent dynamics

The notion of *exponential convergence* for the class of bounded inputs $\mathcal{W} \coloneqq \{\{w(k)\}_{k\in\mathbb{Z}} | \, w(k) \in \mathbb{R}^{n_w}\}$ is recalled from [199].

Definition 3.2 ([199, Def. 2])**.** *A discrete-time nonlinear model* (3.2) *is said to be globally exponentially convergent for an input class $\mathcal{W}$, if, for every input $w \in \mathcal{W}$:*

- *there exists a unique* steady-state solution $\bar{x}$*, that is defined and bounded on* $\mathbb{Z}$*;*
- $\bar{x}$ *is exponentially stable, i.e., there exists a* $c > 0$ *and* $0 < \rho < 1$*, such that* $|x(k) - \bar{x}(k)| \leq c\rho^{(k-k_0)}|x(k_0) - \bar{x}(k_0)| \, \forall \, k \geq k_0, k_0 \in \mathbb{Z}$*.*

■

All solutions of models with the global exponential convergence (GEC) property exhibit bounded responses that converge to the unique steady-state solution independent from the initial condition. The GEC property can be verified by means of an LMI-based check.

Theorem 3.1. *Consider a Lur'e-type model* $(\Sigma,\ \varphi)$ *as in* (3.2) *with the nonlinearity φ satisfying the incremental sector bounds* $[-I_{n_y}, I_{n_y}]$*. Consider a symmetric positive definite matrix* $P \in \mathbb{S}^{n_x}$ *such that the inequalities:*

$$\begin{aligned} P &\succ 0, \\ \begin{bmatrix} A^\top \\ B^\top \end{bmatrix} P \begin{bmatrix} A^\top \\ B^\top \end{bmatrix}^\top - \begin{bmatrix} P - C^\top C & 0_{n_x \times n_y} \\ \star & I_{n_y} \end{bmatrix} &\prec 0, \end{aligned} \tag{3.5}$$

hold. Then, the Lur'e-type model (3.2) *is GEC for the class of bounded inputs $\mathcal{W}$.*

▲

Proof. The proof can be found in Appendix B.1. □

3.2.3 Identification problem using steady-state data

Consider a $\tilde{N}-$periodic input $\mathbf{w} \in \ell_2^{n_w}(\tilde{N})$ and assume that the data-generating system is GEC. Then, it exhibits a unique steady-state response denoted by $\bar{\mathbf{z}} \in \ell_2^{n_z}(\tilde{N})$, with the same period $\tilde{N}$, see [198, Lemma 1]. Hereafter, boldface fonts are used to emphasize that the data comes from the data-generating system. In practice, the input $\mathbf{w}$ and the steady-state output $\bar{\mathbf{z}}$ may be corrupted by noise. Therefore, an integer number of $\kappa \in \mathbb{Z}_{>0}$ steady-state periods are considered. The $\tilde{N}$ samples from all periods of the input and *steady-state* output are collected in the dataset $\mathcal{D}$, defined as follows:

$$\mathcal{D} \coloneqq \{\mathbf{w}(k), \bar{\mathbf{z}}(k)\}_{k=1}^{N}, \quad N \coloneqq \kappa\tilde{N}. \tag{3.6}$$

System identification on the basis of *steady-state* data is a common practice in the literature, see, e.g., [205, 235, 240]. Amongst others, it allows performing the so-called nonlinear distortion analysis [235] to gain insights into the dominant nonlinearities.

The dataset $\mathcal{D}$ is used to identify a model of the form (3.2). Hereto, consider the parameter vector $\theta \in \mathbb{R}^{n_\theta}$, which fully parametrizes a GEC Lur'e-type candidate model (3.2). The relation between θ and the building blocks of (3.2) depends on the specific problem at hand. One can adopt a black-box modeling approach, in which all entries of the linear block matrices and all nonlinear block parameters represent a model parameter. Contrarily, a white-box modeling routine can impose additional model structure as a result of, e.g., first-principle modeling. We note that model parameters can also appear (in a nonlinear fashion) in the nonlinearity φ.

The set of models that satisfy the conditions of Theorem 3.1 is denoted by $\Theta \subset \mathbb{R}^{n_\theta}$ and defined as follows:

$$\Theta \coloneqq \{\theta \in \mathbb{R}^{n_\theta} \mid (3.5) \text{ is feasible}\}. \tag{3.7}$$

Any candidate model θ from Θ thus obeys the GEC property. The steady-state solution $\bar{z}(\theta) \in \ell_2^{n_z}(N)$ of a candidate model $\theta \in \Theta$ under the excitation by $w \in \ell_2^{n_w}(N)$ is N-periodic, bounded, unique and globally exponentially stable.

If the data-generating system is inside the model class Θ, then the identification goal is to find its parametrization. However, as often this is not the case, a black-box modeling approach is taken where the aim is to find an *accurate* model inside the class Θ for the data-generating system. These goals are formalized by considering the steady-state model output errors with respect to system observations as follows:

$$\bar{\varepsilon}(k,\theta) = \bar{z}(k,\theta) - \bar{\mathbf{z}}(k), \qquad \forall k \in \{1,\dots,N\}. \tag{3.8}$$

Then, the steady-state simulation error criterion is defined according to:

$$J_N(\theta) \coloneqq \frac{1}{N}\sum_{k=1}^{N} \|\bar{\varepsilon}(k,\theta)\|_2^2. \tag{3.9}$$

Finally, we define the system identification problem to be solved in this chapter.

Problem 3.1. *Consider the dataset defined by $\mathcal{D}$ in* (3.6)*. The system identification problem is to solve the optimization problem:*

$$\hat{\theta}_N = \arg\min_{\theta\in\Theta} J_N(\theta), \tag{3.10}$$

where $\theta \in \Theta$ parametrizes the class of GEC models (3.2) *that satisfy the conditions of Theorem 3.1.* △

Remark 3.2. Under the standard assumptions that (i) the system is in the model class, (ii) the input w is persistently exciting, and (iii) there is additive white noise in the measurements $\mathbf{z}$ that is uncorrelated with the input w, it can be shown that the estimator (3.10) is asymptotically efficient. Asymptotically efficient estimators achieve the lowest asymptotic mean squared error possible with a consistent estimator, see [243] for more detail. △

Remark 3.3. The model (3.2) does generally not admit a unique parametrisation [243, 251]. For example, a similarity transformation of the LTI part changes the matrices of the LTI part but does not affect the model output. Therefore, $\hat{\theta}_N$ is a set in Problem 3.1. △

3.3 Computationally efficient solution to the identification problem

Problem 3.1 is a constrained optimization problem. This section reviews optimization methods, after which computationally fast tools are presented to solve the identification problem using such optimization methods.

3.3.1 Constrained optimization

In system identification literature, it is common practice to solve constrained optimization problems by a two-step approach [240]: (i) parameter initialization, (ii) gradient-based optimization.

Step (i): Parameter initialization

The objective of this step is to find an initial parameter vector $\theta_{\text{init}} \in \Theta$, close to the global minimum and inside the set of GEC models characterized by Θ, from which a constrained gradient-based search in the second step can be started. Hereto, we propose two methods being the best-linear approximation (BLA) method and the global optimization method.

Best Linear Approximation

This method relies on estimating the so-called BLA of the nonlinear system [235]. To this end, we approximate the mapping from the excitation $\mathbf{w}$ to the system output $\mathbf{z}$ by an LTI model Σ_{BLA} that admits a state-space representation characterized by the matrices $(A_{BLA},\ B_{BLA},\ C_{BLA},\ D_{BLA})$. The system matrices can be found via the MATLAB routines `N4SID` or `TFEST`, and crucially, these routines can enforce stability of the linear model Σ_{BLA} by placing the eigenvalues of A_{BLA} inside the open unit disc.

The matrices (A, L, F, H) of $\Sigma(\theta_{init})$ are, respectively, selected as the matrices $(A_{BLA}, B_{BLA}, C_{BLA}, D_{BLA})$ of the BLA. The remaining matrices are selected such that there is no (nonlinear) feedback, i.e., either (C, D) are zero matrices, (B, G) are zero matrices, or $\varphi(\cdot, \cdot) = 0$. With this choice, the nonlinear Lur'e-type model $(\Sigma(\theta_{\text{init}}), \varphi(\theta_{\text{init}}))$ becomes identical to the linear model Σ_{BLA} in terms of the input-output behavior. The BLA framework enables fast estimation of initial model parameters, and, additionally, provides a nonlinear distortions analysis together with an estimate for the state dimension [235]. Also, this initialization method was applied successfully for nonlinear system identification in prior research, see [243, 244] and references therein.

Global Optimization

Global optimization routines can be employed to identify an initial parameter vector θ_{init} that is sufficiently close to the global minimum of the objective function. The search can be stopped prematurely to save computational costs since a gradient-based method, starting from this initial point, can thereafter be used to effectively find the closest minimum of the cost function. Global optimization methods typically encompass (sophisticated) random walks through the parameter space. Examples are, among others, Monte-Carlo, genetic-type, and swarm intelligent algorithms [156]. The controlled random search method, as described in [206], was successfully applied as an initialization method for nonlinear system identification of continuous-time models [127, 251, 265]. Generally, global optimization routines are considered computationally expensive [156]. Therefore, a computationally cheap method to evaluate the objective function value, such as the one proposed in Section 3.3.2 below, can be an enabler for the application of global optimization routines.

Step (ii): Gradient-based optimization

Off-the-shelf gradient-based optimizers, such as the one in [56], can be employed to solve the constrained optimization problem in an iterative fashion, starting from the θ_{init} obtained in Step (i). Any gradient-based method requires evaluating the objective function gradient with respect to the model parameters. The gradient of the objective function (3.9) can be expressed as follows:

$$\frac{\partial}{\partial \theta} J_N(\theta) = \frac{2}{N} \sum_{k=1}^{N} \bar{\varepsilon}(k, \theta)^\top \frac{\partial}{\partial \theta} \bar{\varepsilon}(k, \theta), \tag{3.11}$$

where the steady-state mismatch $\bar{\varepsilon}(k, \theta)$ is defined in (3.8). Typically, finite difference approximations of the gradient (3.11) are used, which could lead to an increased number of iterations upon completion of the optimization algorithm. In this chapter, a computationally efficient approach is presented in Section 3.3.2 to compute the exact gradient in (3.11).

3.3.2 Computationally efficient numerical tools

We first propose the discrete-time mixed-time frequency (MTF) algorithm to efficiently compute steady-state responses of Lur'e-type models. After that, we introduce a parameter sensitivity model, the steady-state response of which can again be computed efficiently using the MTF algorithm and can be used to compute exact gradient information. In this section, the dependency on model parameters θ is dropped for notational convenience.

Mixed time-frequency algorithm

In the continuous-time case, [196] proposed the mixed time-frequency (MTF) algorithm for the efficient computation of steady-state model responses of convergent SISO Lur'e-type models. In this section, the MTF algorithm is extended to the class of *multivariable discrete-time* Lur'e-type models.

Before presenting the MTF algorithm, we recall that the asymptotically stable LTI block maps the periodic inputs $w, \bar{u}$ to the periodic steady-state output $\bar{z}$, i.e., $\bar{z} = \mathcal{F}_{zw} w + \mathcal{F}_{zu} \bar{u}$, where the linear operator $\mathcal{F}_{zw}$ is defined for $k \in \mathbb{Z}$ as follows:

$$\left(\mathcal{F}_{zw} w\right)(k) = \sum_{i=-\infty}^{k} F A^{k-i-1} L w(i) + H w(k). \tag{3.12}$$

In a similar fashion, the linear operators $\mathcal{F}_{zu}, \mathcal{F}_{yw}$, and $\mathcal{F}_{yu}$ are defined. Alternatively, the same operation can be defined through the *frequency-domain steady-state operators* $\hat{\mathcal{F}}_{ZW}$ and $\hat{\mathcal{F}}_{ZU}$, where the Fourier coefficients $\bar{Z}$, denoting the Fourier transform of $\bar{z}$, are computed through $\bar{Z} = \hat{\mathcal{F}}_{ZW} W + \hat{\mathcal{F}}_{ZU} \bar{U}$, where W and $\bar{U}$ denote the Fourier transforms of w and $\bar{u}$, respectively. The operator $\hat{\mathcal{F}}_{ZW}$ is defined for all $m \in \{1, \ldots, N\}$ according to

$$\left(\hat{\mathcal{F}}_{ZW} W\right)(m) = \mathcal{G}_{\Sigma_{zw}}\left(e^{j\tilde{\omega}(m-1)}\right) W(m) \tag{3.13}$$

with $\tilde{\omega} \coloneqq 2\pi/N$ and $\mathcal{G}_{\Sigma_{zw}}(\varkappa) \coloneqq F(\varkappa I_{n_x} - A)^{-1} L + H$, $\varkappa \in \mathbb{C}$. The frequency-domain steady-state operators $\hat{\mathcal{F}}_{ZU}, \hat{\mathcal{F}}_{YW}$, and $\hat{\mathcal{F}}_{YU}$ are defined similarly. Finally, the nonlinear operator is defined, in the time domain, as follows:

$$\left(\mathcal{F}_{uy} y\right)(k) = \varphi(y(k), k). \tag{3.14}$$

These steady-state operators can be used in an iterative procedure to efficiently compute steady-state model responses, as presented in the next theorem.

Theorem 3.2. *Consider a periodic excitation $w \in \ell_2^{n_w}(N)$ applied to model (3.2). Under the conditions of Theorem 3.1, the sequence $\left(y_{[i]}\right)_{i \in \mathbb{Z}_{>0}}$ defined by the iterative application of the mappings*

$$u_{[i+1]} = \mathcal{F}_{uy} y_{[i]}, \tag{3.15a}$$

$$y_{[i+1]} = \mathcal{F}_{yu} u_{[i+1]} + \mathcal{F}_{yw} w. \tag{3.15b}$$

with an arbitrary initial condition $y_{[0]} \in \ell_2^{n_y}(N)$, has a unique limit, denoted by $\bar{y}$. Furthermore, this unique limit $\bar{y}$ coincides with the steady-state model output $\bar{y}$ of model $(\Sigma,\ \varphi)$. ▲

Proof. The proof can be found in Appendix B.2. □

This theorem can be used to accurately compute the steady-state solution $\bar{y}$ of a convergent Lur'e-type model via the iterations in (3.15), starting from any $y_{[0]} \in \ell_2^{n_y}(N)$, for example the zero response. Having $\bar{y}$ at hand, one can trivially compute $\bar{z}$, to be used to compute the steady-state error $\bar{\varepsilon}$, as required in (3.8).

To improve the computational efficiency of the implementation of the MTF algorithm, we evaluate the LTI dynamics in *frequency domain* and the nonlinearity in time domain. The intermediate signals are transformed forth and back between time and frequency domains using the DFT and IDFT. The MTF algorithm is summarized in Algorithm 3.1, where the operators $\hat{\mathcal{F}}_{XU}$ and $\hat{\mathcal{F}}_{XW}$ are defined similarly to $\hat{\mathcal{F}}_{ZW}$ in (3.13). As a stopping criterion, the normalized distance between the current and the previous approximation of the Fourier coefficients of the steady-state output $\bar{Y}$ is used, namely:

$$\eta_{[i]} = \frac{\left\|\mathrm{vect}\left(\bar{Y}_{[i]} - \bar{Y}_{[i-1]}\right)\right\|_2}{\left\|\mathrm{vect}\left(\bar{Y}_{[i]}\right)\right\|_2}, \tag{3.16}$$

in which $\mathrm{vect}\,(\cdot)$ is an operator that reshapes all elements of a matrix into a single column vector with a column-major layout. The efficiency of the MTF is highlighted in the numerical studies in Section 3.4.

Efficient computation of gradient information

Parameter sensitivity models are used to compute $\frac{\partial}{\partial\theta}\bar{\varepsilon}(k,\theta)$, required in the evaluation of the gradient in (3.11). By using (3.8), we can further rewrite the partial derivative in (3.11) as:

$$\frac{\partial}{\partial\theta}\bar{\varepsilon}(k,\theta) = \frac{\partial}{\partial\theta}\bar{z}(k,\theta) - \underbrace{\frac{\partial}{\partial\theta}\bar{\mathbf{z}}(k)}_{=0} = \frac{\partial}{\partial\theta}\bar{z}(k,\theta). \tag{3.17}$$

Below, a method for calculation of the partial derivative in the right-hand side of (3.17) is given. For the sake of clarity of presentation and simplified notation, this part is presented for the case of a scalar θ. For a vector θ, the partial derivatives can be directly calculated from the partial derivatives with respect to the components of θ.

Throughout this section, where no ambiguity occurs, the θ and y subscripts denote a partial derivative with respect to the scalar model parameter θ and the nonlinearity input y, respectively. Therefore, one can write $\varphi_\theta(y,k,\theta) \coloneqq \frac{\partial}{\partial\theta}\varphi(y,k,\theta)$ and $\varphi_y(y,k,\theta) \coloneqq \frac{\partial}{\partial y}\varphi(y,k,\theta)$ for the nonlinearity, as well as, e.g., $A_\theta \coloneqq \frac{\partial}{\partial\theta}A(\theta)$

Algorithm 3.1 Mixed-Time-Frequency Algorithm

Input: Lur'e-type model $(\Sigma,\ \varphi)$ whose nonlinearity satisfies the incremental sector bound $[-I_{n_y}, I_{n_y}]$ and the conditions of Theorem 3.1; stopping tolerance $\eta^\star > 0$.

1: Set the iteration counter $i = 0$ and the stopping criterion $\eta_{[0]} = \infty$.
2: Compute W from w using the DFT.
3: Evaluate LTI dynamics in frequency domain (assume $u_{[0]} = 0_{N\times n_u}$): $Y_{[0]} = \hat{\mathcal{F}}_{YW} W$.
4: Compute $y_{[0]}$ from $Y_{[0]}$ using the IDFT.
5: **while** $\eta_{[i]} > \eta^*$ **do**
6: Evaluate the nonlinearity in time domain: $u_{[i+1]} = \hat{\mathcal{F}}_{uy} y_{[i]}$.
7: Compute $U_{[i+1]}$ from $u_{[i+1]}$ using the DFT.
8: Evaluate LTI dynamics in frequency domain: $Y_{[i+1]} = \hat{\mathcal{F}}_{YU} U_{[i+1]} + Y_{[0]}$.
9: Compute $y_{[i+1]}$ from $Y_{[i+1]}$ using the IDFT.
10: set $i = i + 1$.
11: **end**
12: Define $\bar{Y}$ and $\bar{U}$ as $Y_{[i]}$ and $U_{[i]}$, respectively.
13: Evaluate LTI dynamics in frequency domain: $\bar{Z} = \hat{\mathcal{F}}_{ZU}\bar{U} + \hat{\mathcal{F}}_{ZW} W, \quad \bar{X} = \hat{\mathcal{F}}_{XU}\bar{U} + \hat{\mathcal{F}}_{XW} W$.
14: Compute $\bar{u}$, $\bar{y}$, $\bar{z}$, $\bar{x}$ from $\bar{U}$, $\bar{Y}$, $\bar{Z}$, $\bar{X}$ using the IDFT.

Output: Steady-state solution of $(\Sigma,\ \varphi)$: $(\bar{u}, w, \bar{x}, \bar{y}, \bar{z})$.

for a matrix. We impose the following assumption on the partial derivatives of Lur'e-type candidate model components.

Assumption 3.1. *The partial derivates $\varphi_y(y), \varphi_\theta(y)$, A_θ, B_θ, L_θ, C_θ, D_θ, F_θ, G_θ, H_θ exist and are continuous for all $y \in \mathbb{R}^{n_y}$ and $\theta \in \Theta$.* ■

Using this assumption, the notion of a *parameter sensitivity model* including its instrumental properties is formalized in Theorem 3.3 below.

Theorem 3.3. *Consider a Lur'e-type model $(\Sigma,\ \varphi)$ given by* (3.2) *with the system matrices and the nonlinearity dependent on the parameter vector θ such that Assumption 3.1 is satisfied. Let θ be such that the conditions of Theorem 3.1 are satisfied. Moreover, let w be a N-periodic input and $\bar{y}$ be the corresponding N-periodic steady-state output. Define the* parameter sensitivity model $(\Sigma^\theta,\ \varphi^\theta)$ *as*

$$\Sigma^\theta : \begin{cases} x_\theta(k+1) &= Ax_\theta(k) + B\breve{u}^\theta(k) + \breve{w}_1^\theta(k), \\ y_\theta(k) &= Cx_\theta(k) + \breve{w}_2^\theta(k), \\ z_\theta(k) &= Fx_\theta(k) + G\breve{u}^\theta(k) + \breve{w}_3^\theta(k), \end{cases} \tag{3.18a}$$

$$\varphi^\theta : \quad \breve{u}^\theta(k) = \varphi^\theta\big(y_\theta(k), \bar{y}(k), k\big), \tag{3.18b}$$

where $\breve{w}_1^\theta(k), \breve{w}_2^\theta(k)$, *and* $\breve{w}_3^\theta(k)$ *are given by*

$$\begin{aligned}\breve{w}_1^\theta(k) &= A_\theta \bar{x}(k) + B_\theta \bar{u}(k) + L_\theta w(k) + B\varphi_\theta(\bar{y}(k), k),\\ \breve{w}_2^\theta(k) &= C_\theta \bar{x}(k) + D_\theta w(k),\\ \breve{w}_3^\theta(k) &= F_\theta \bar{x}(k) + G_\theta \bar{u}(k) + H_\theta w(k) + G\varphi_\theta(\bar{y}(k), k).\end{aligned}$$

The nonlinearity φ^θ *is incrementally sector bounded with bounds* $[-I_{n_y}, I_{n_y}]$ *and is defined as*

$$\varphi^\theta\big(y_\theta(k), \bar{y}(k), k\big) = \varphi_y\big(\bar{y}(k), k\big) y_\theta(k). \tag{3.19}$$

Then, the parameter sensitivity Lur'e-type model $(\Sigma^\theta, \varphi^\theta)$ *satisfies the following properties:*

- *the conditions of Theorem 3.1 hold, hence* $(\Sigma^\theta, \varphi^\theta)$ *is GEC;*
- *the unique* N*-periodic steady-state model output* $\bar{z}_\theta$ *is such that* $\bar{z}_\theta = \frac{\partial}{\partial \theta}\bar{\varepsilon}$.

▲

Proof. The proof can be found in Appendix B.3. □

The parameter sensitivity model presented in Theorem 3.3 takes as input the steady-state solution of model (3.2), which should be computed in the first step. After that, the steady-state output of the parameter sensitivity model provides the objective function gradient with respect to a single parameter $\theta_i, i \in \{1, \ldots, n_\theta\}$. Therefore, n_θ steady-state solutions of the parameter sensitivity models are to be computed to obtain the gradient information with respect to all parameters. Since each sensitivity model satisfies the assumptions of Theorem 3.3, their steady-state outputs can be computed by means of the MTF algorithm in an accurate and numerically efficient way, see Theorem 3.2 and Algorithm 3.1. In a constrained gradient-based optimization routine, the gradient information can be effectively used to steer towards a minimum in parameter space.

3.4 Numerical examples

This section starts with a simulation example that illustrates the benefits of enforcing the GEC property on the identified model. After that, the identification approach is tested using two benchmark datasets.

3.4.1 Illustrative academic example

Consider the scalar dynamics in the form of equations (3.2) with matrices $A = \theta_1, B = \theta_2, L = \theta_3, C = F = 1, D = G = H = 0$, and the nonlinearity $\varphi(\cdot) = \operatorname{atan}(\cdot)$. The parameters $\theta_1, \ldots, \theta_3$, are collected in the vector $\theta \in \mathbb{R}^3$.

Table 3.1. Parameters and objective function value for the example in Section 3.4.1.

Model	θ_1	θ_2	θ_3	J_N
θ_{true}	0.24	0.75	1	0.2238
θ_{unc}	0.198	0.806	1.005	0.2331
θ_{GEC}	0.203	0.797	1.005	0.2331

The data-generating system is characterized by θ_{true} in Table 3.1 and satisfies the conditions of Theorem 3.1. A periodic input with $N = 1000$ samples is drawn from a zero-mean normal distribution with unit variance. The steady-state output response $\bar{\mathbf{z}}$ is distorted with noise e drawn from a zero-mean normal distribution with a variance that ensures a 10 dB signal-to-noise ratio. Using this dataset, two models are identified. The first model, characterized by θ_{unc}, is identified without enforcing the constraints for GEC, i.e., $\theta_{\text{unc}} \in \mathbb{R}^3$. In this case, the objective function (3.9) is minimized in an unconstrained fashion. The second model, characterized by θ_{GEC}, is identified by solving Problem 3.1 as proposed in this chapter. This model is guaranteed to be GEC, i.e., $\theta_{\text{GEC}} \in \Theta$. The identified model parameters are included in Table 3.1.

Figure 3.2 depicts the steady-state output $\bar{\mathbf{z}}+e$ and the steady-state mismatches $\bar{\varepsilon}(\theta_{\text{unc}})$ and $\bar{\varepsilon}(\theta_{\text{GEC}})$ with $\bar{\varepsilon}$ defined in (3.8). These small mismatches evidence an accurate match based on the steady-state data, which is further confirmed by the objective function value for both models being equally small ($J_N = 0.2331$). However, the response of the model characterized by θ_{unc} converges, depending on the initial condition, to different fixed points for zero input, as evidenced in Figure 3.3. Since a GEC model cannot exhibit multiple fixed points, this model is not GEC. In turn, this implies that for periodic inputs, multiple, stable steady-state responses exist that are attractive for different initial conditions, which is an unfavorable property in the scope of this work. The responses of model θ_{GEC} do converge to the same fixed point, irrespective of the different initial conditions. This example thus shows that even when the data-generating system is GEC, the identified model does not necessarily preserve the GEC property if the conditions for convergence are not enforced.

3.4.2 Wiener-Hammerstein benchmark

The *Wiener-Hammerstein* benchmark, proposed by [239], has been actively studied in the literature, see [27] for an overview. The system consists of two third-order Chebychev filters sandwiching a diode-resistor nonlinearity, see Figure 3.4 for a schematic representation. The dataset consists of 100000 estimation and 88000 test input-output pairs, where the first 1000 samples are treated as transients.

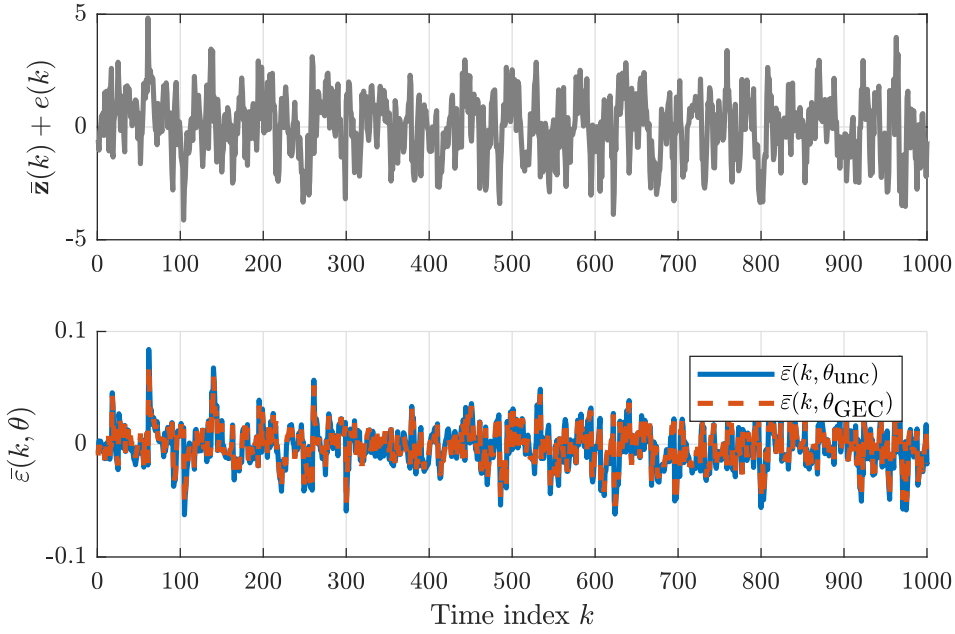


Figure 3.2. Top: the steady-state system output $\bar{\mathbf{z}}(k) + e(k)$. Bottom: the steady-state output error $\bar{\varepsilon}(k, \theta_{\text{unc}})$ and $\bar{\varepsilon}(k, \theta_{\text{GEC}})$ with $\bar{\varepsilon}$ defined in (3.8).

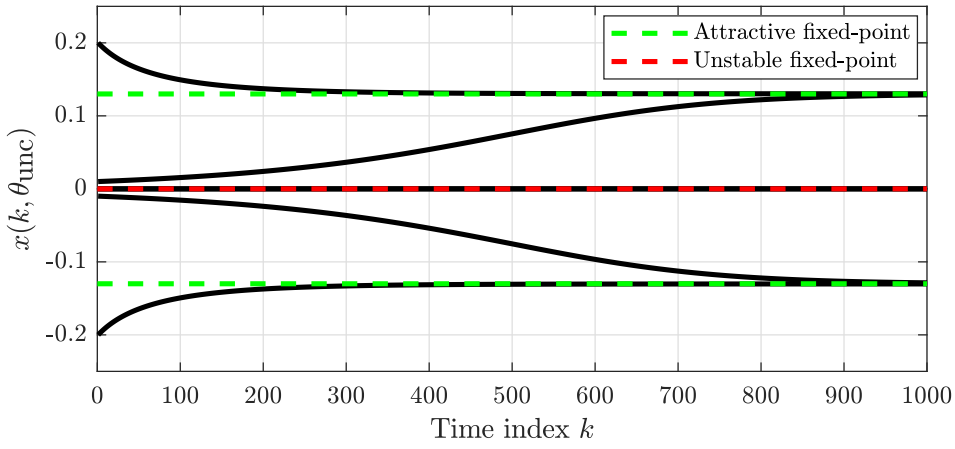


Figure 3.3. Response of the identified model characterized by θ_{unc} for a zero input for different initial conditions $\{-0.2, -0.01, 0, 0.01, 0.2\}$. Depending on the initial condition, the response converges to a different fixed point.

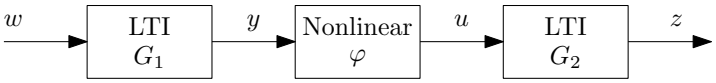


Figure 3.4. Schematic of the Wiener-Hammerstein system.

Model structure and parametrization

The two LTI blocks of the Wiener-Hammerstein model, see Figure 3.4, can be cast into a single LTI model such that output y is the output of the first LTI block, the input u is the input of the second LTI block and the output z is the output of the second LTI block. Each of the third-order LTI blocks is parametrized in a controllability canonical form by seven parameters, yielding $\theta_\Sigma \in \mathbb{R}^{14}$. The nonlinearity, parametrized by $\theta_\varphi = [\alpha, \beta]^\top \in \mathbb{R}^2$, is given by the equation

$$\varphi(y) = -\log(\alpha + \exp(-y)) + \log(\alpha + 1) + \beta y. \tag{3.20}$$

For any $\alpha > 0$ and $\beta \geq 0$, the conditions of Theorem 3.1 boil down to the eigenvalues of the matrix A being located inside the unit disc.

The LTI part of the candidate model is initialized by taking two random stable third-order LTI models (using Matlab's `drss` routine). The parameters of the nonlinearity are initialized by taking a random positive number for α and setting $\beta = 0$.

Identification results

The identification problem is solved by Matlab's `lsqnonlin` implementation of the trust-region-reflective method [56]. The identified nonlinearity is illustrated in Figure 3.6. The measured output $\mathbf{z}$ is plotted together with the error $\varepsilon \coloneqq z(\theta) - \mathbf{z}$ in Figure 3.5. It can be observed that the error ε is close to zero, evidencing an accurately identified model. The root-mean-square error (RMSE) on the test data (excluding transients) is 0.62 mV, where the RMSE is defined as follows:

$$\text{RMSE} \coloneqq \sqrt{\frac{1}{N}\sum_{k=1}^{N} \|\varepsilon(k)\|_2^2}. \tag{3.21}$$

The performance in terms of the RMSE on the test dataset (0.64 mV) is comparable to the performance on the estimation dataset. These accuracy results are in close proximity to the best state-of-the-art results in the literature, see [27] for an overview.

3.4.3 Silverbox benchmark

The *Silverbox system*, proposed by [304], is an electronic implementation of the Duffing oscillator, which can be modeled as a second-order Lur'e-type model with a

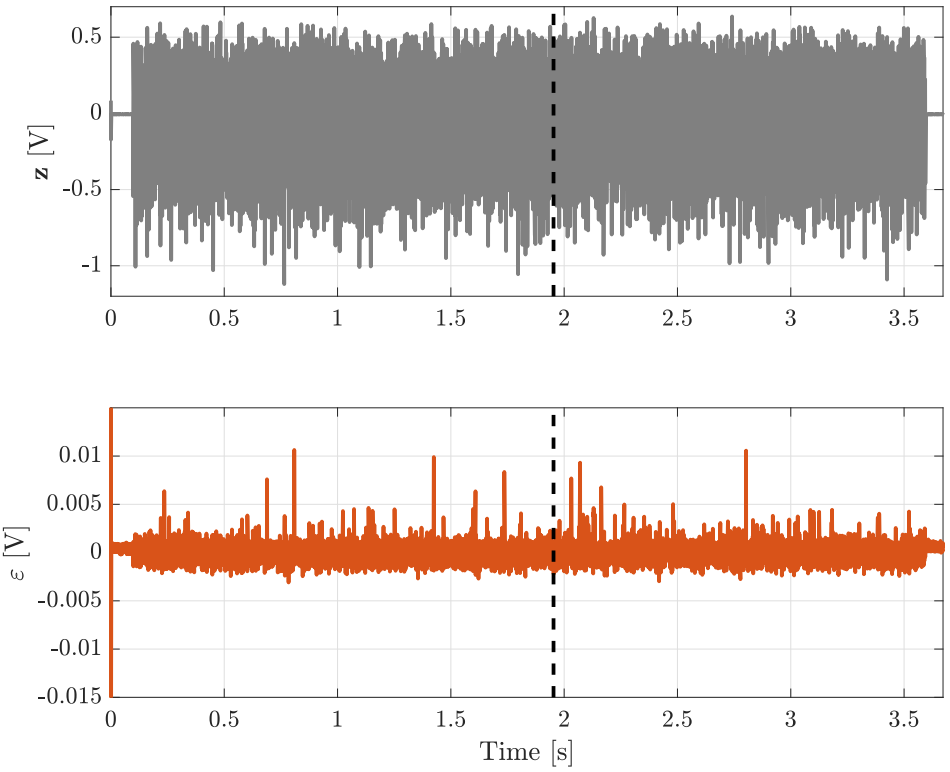


Figure 3.5. Measured output $\mathbf{z}$ (top) and the error ε of the identified model (bottom). The vertical black dashed line represents the split between the estimation data (left) and test data (right).

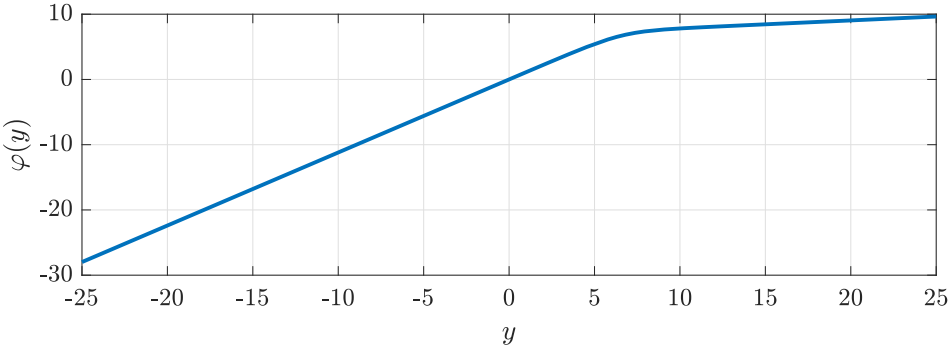


Figure 3.6. Graph of the identified nonlinearity in the Wiener-Hammerstein benchmark.

cubic nonlinearity. The cubic nonlinearity cannot be *globally* incrementally sector bounded, therefore, this system does not fit into the class of models that satisfy the conditions of Theorem 3.1. The goal is to investigate whether an accurate GEC model can be identified for this system by replacing the cubic nonlinearity with a spline nonlinearity that can be *globally* incrementally sector bounded.

Identification data

The scalar excitation signal $\mathbf{w}$ used during the identification experiment is shown in Figure 3.7. The data can be split into a test part (arrow's 'head' in red) and an estimation part (arrow's 'tail' in blue). The test data is a *band-limited Gaussian white noise* sequence with a linear increase in amplitude forward in time. The estimation data consists of ten *random-phase odd multisine* realizations, the last multisine of which is omitted in the identification procedure, consistent with [161]. Note that the excitation amplitude of the test data exceeds the amplitudes seen during training, allowing to test extrapolation properties of the identified models.

Model structure and parameter initialization

Given the prior knowledge of the Silverbox dynamics, a second-order LTI state-space model interconnected via feedback by a SISO nonlinearity is selected. Furthermore, the output z also serves as nonlinearity input y, i.e., $F(\theta) = C(\theta)$, $G(\theta) = 0$ and $H(\theta) = D(\theta)$. Finally, the nonlinearity output u and the external excitation w enter the model according to $B(\theta) = -L(\theta)$. The linear block is parametrized in a controllability canonical form by $\theta_\Sigma \in \mathbb{R}^5$.

Two candidate nonlinearities are proposed, namely:

$$\varphi_{\mathrm{I}}(y) = \sum_{i=1}^{3} k_i y^i, \tag{3.22a}$$

$$\theta_{\varphi_{\mathrm{I}}} = [k_1, k_2, k_3]^\top \in \mathbb{R}^3, \tag{3.22b}$$

$$\varphi_{\mathrm{II}}(y) = \begin{cases} ky^3, & |y| \geq y_\star, \\ 3ky_\star^2 - \operatorname{sign}(y) 2ky_\star^3, & |y| < y_\star, \end{cases} \tag{3.22c}$$

$$\theta_{\varphi_{\mathrm{II}}} = [k, y_\star]^\top \in \mathbb{R}^2. \tag{3.22d}$$

The nonlinearity φ_{I} is a *cubic polynomial*, which fits the prior knowledge on the Silverbox system. However, it does not allow for stability guarantees via Theorem 3.1, since it cannot be globally incrementally sector bounded. The second nonlinearity φ_{II} is a *spline* nonlinearity, which behaves as a pure cubic polynomial within the spline width (including the origin) and is linear outside. For nonnegative constants k and $y_\star$, it is incrementally sector bounded within $[0, \bar{\Omega}]$ with $\bar{\Omega} = 3ky_\star^2$. Loop transformations are used to transform the nonlinearity such that it satisfies the incremental sector bound $[-I_{n_y}, I_{n_y}]$.

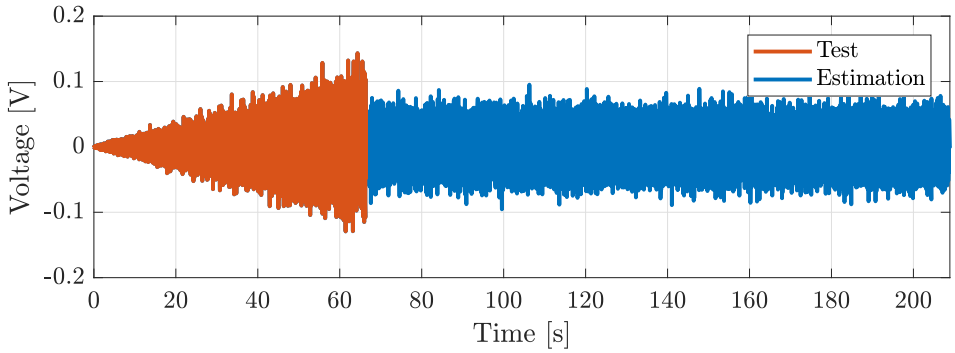


Figure 3.7. Silverbox identification experiment, split between the estimation (blue) and a test dataset (red).

The parameters characterizing the candidate model of a certain structure are expressed by

$$\theta_i = \begin{bmatrix} \theta_\Sigma^\top & \theta_{\varphi_i}^\top \end{bmatrix}^\top, \qquad \forall\, i \in \{\mathrm{I}, \mathrm{II}\}\,. \tag{3.23}$$

The parameters θ_Σ are initialized according to the BLA [235], which is characterized by θ_{init}. The nonlinearity φ_{I} is initialized according to $\theta_{\varphi_{\mathrm{I}}} = 0_3$; and φ_{II} according to $k = 0$ and $y_\star$ a random positive number.

Identification results

Starting from the BLA characterized by θ_{init}, the identification problem is solved by Matlab's `lsqnonlin` implementation of the trust-region-reflective method [56]. For the model structure with the nonlinearity φ_{I}, the GEC property is not enforced and model responses are computed by forward iterating the nonlinear dynamics. The graphs of identified nonlinearities φ_{I} and φ_{II} are depicted in Figure 3.8, where the spline nature of φ_{II} is recognized.

The identification results in terms of the RMSE are presented in Table 3.2, where the RMSE is defined in (3.21) and where the model response is computed by forward iterating the dynamics through the estimation data. The BLA performs the worst, as expected since this model does not include the cubic nonlinear feedback. The model with cubic nonlinearity performs the best, comparable with other identification results in the literature, see [27]. However, this model is not certified with a stability property. The model with the spline nonlinearity performs not as well as the model with the cubic nonlinearity, which is expected as this model does not extrapolate nonlinearly outside the spline width $y_\star$, but it is certified with the GEC property.

The numerical efficiency of our approach relies on two ingredients, namely: (i) the MTF algorithm to compute steady-state model responses; and (ii) exact gradient information computed by the MTF algorithm via the results of Theo-

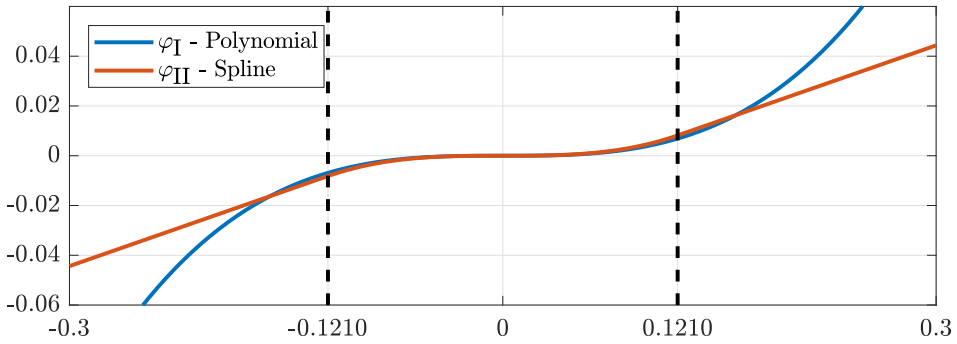


Figure 3.8. Identified polynomial (blue) and spline (red) nonlinearities in the Silverbox example. The spline width $y_\star$ is indicated by the vertical black dashed lines.

Table 3.2. Silverbox identification performance in terms of the RMSE [mV] for the BLA characterized by θ_{init}, the identified nonlinear model with cubic nonlinearity characterized by θ_{I}, and the identified nonlinear model with spline nonlinearity characterized by θ_{II}.

Dataset	θ_{init}	θ_{I}	θ_{II}
Estimation	6.94	0.52	1.14
Test	13.70	0.55	4.28

rem 3.3. The total computation time[1] to find θ_{II} by solving the identification problem using (i) and (ii) is 363 seconds. The total computation time to solve the same problem without (i), but with (ii) is 1288 seconds. In this case, steady-state model responses are computed by forward iterating the dynamics. Solving the same problem with (i), but without (ii) takes 4187 seconds, where gradient information is found by finite-differencing [193]. The total computation time to solve the same problem without (i) and without (ii) is 7500 seconds. In each of these cases, the optimization procedure resulted in approximately the same parameter vector θ_{II}. Our approach thus reduces the computation time by more than 44%, 70%, and 95%, respectively. The computation time can be further reduced to only 175 seconds if, in addition to using (i) and (ii), parallel computing is employed for the computation of the gradient information.

Performance on test data

The performance of the model is quantified by forward iterating the identified model throughout the arrow-'head' excitation (samples 1 - 40585), see Figure 3.7.

[1] On an Intel Core i7-7700HQ, 2.8 GHz processor.

Table 3.3. Elapsed computation time for solving the identification problem on the basis of the estimation dataset using/not using (i) the MTF algorithm and (ii) exact gradient information.

Methods	(i),(ii)	~~(i)~~,(ii)	(i),~~(ii)~~	~~(i)~~,~~(ii)~~
Time [sec]	363	1288	4187	7500

The RMSE, included in Table 3.2, shows that the BLA cannot catch up with the nonlinear models. Model I with cubic nonlinearity is a good representation of the observations. Even in the extrapolation regime, which starts roughly after 50 seconds, only a slight increase in model error is observed, see Figure 3.9. The behavior of the spline model II is in line with expectations. For small excitation magnitudes, the nonlinearity behaves in the cubic regime of the model and the model performance is close to that of model I. For larger excitation magnitudes, the nonlinearity acts in its linear regime inducing model performance loss.

Finally, the estimated models are compared to state-of-the-art models in the literature, see [27]. Model I with cubic nonlinearity (without stability guarantees) accurately models the Silverbox system having its performance measures close to the best results in the literature. Model II with spline nonlinearity is shown to be a significant improvement with respect to the BLA and is the only nonlinear model reported in the literature with stability guarantees. Arguably, the spline model II is rendered in the sweet spot of being relatively accurate and predicting reliably on input sequences that were not seen during model training.

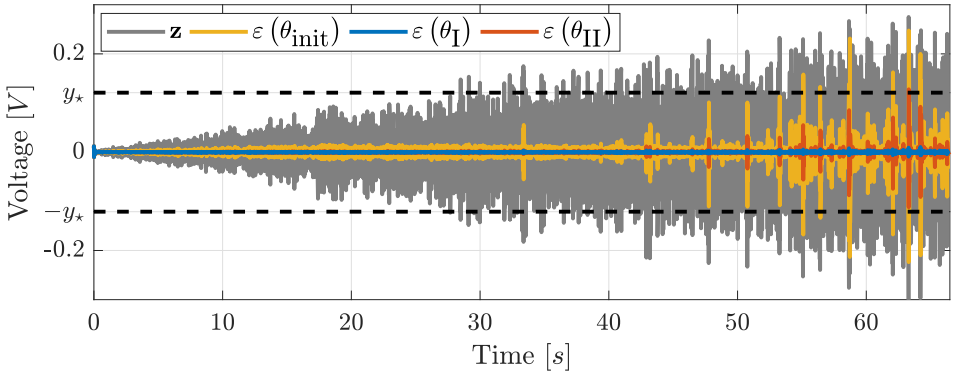


Figure 3.9. The test data $\mathbf{z}$ (blue), the error $\varepsilon(\theta_{\text{init}})$ of the BLA (red), the error $\varepsilon(\theta_{\text{II}})$ of the model with spline nonlinearity (yellow), and the error $\varepsilon(\theta_{\text{I}})$ of the model with cubic nonlinearity (purple). The spline width $y_\star$ is indicated by the black dashed lines.

3.5 Conclusions

This chapter has presented a method for the system identification of discrete-time Lur'e-type models with certification of the convergence property. The convergence property guarantees the stability properties of the identified model and reliable usage of the identified model under generalized inputs. On the basis of sufficient conditions for convergence, we propose a novel numerical algorithm that enables the efficient computation of the steady-state responses of discrete-time MIMO Lur'e-type models. Furthermore, a method to efficiently and exactly compute the gradient of the model responses with respect to the model parameters is presented. These two aspects result in a 95% reduction in the computation time required to solve the identification problem for the nonlinear Silverbox benchmark. Furthermore, the identification approach is validated on the nonlinear Wiener-Hammerstein benchmark.

4

Accuracy bounds for the simulation of a class of continuous-time nonlinear models

Real-life dynamic systems evolve in the continuous-time world whereas their models are simulated in the digital world using discrete-time numerical simulation algorithms. Such model simulation is essential for a wide variety of (nonlinear) systems and control problems such as system identification and performance analysis of (control) systems. Ideally, both the system and the model should produce the exact same response, however, this is typically not the case in practice, even if the model perfectly describes the system. This chapter analyzes the steady-state mismatch between the response of continuous-time, nonlinear Lur'e-type systems and their continuous-time Lur'e-type models whose steady-state response is computed with the so-called mixed time-frequency algorithm. This algorithm has practical importance in a number of performance optimization and identification problems. Firstly, a bound is derived on the steady-state mismatch between the computed model responses and the actual system responses. Secondly, a bound is derived on the steady-state mismatch between the computed model responses and the measured system responses, sampled by an analog-to-digital device. The derived bounds are decomposed into several components, each having a clear interpretation that can be used to reduce the bounds on the mismatch. In a numerical case study, we show that a reduction of the bounds also leads to a reduced mismatch.

The contents of this chapter are published in: Shakib, M. F., Schoukens, J., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Accuracy bounds for the simulation of a class of continuous-time nonlinear models. *In preparation.*

4.1 Introduction

Dynamic models of real-life systems are of crucial importance in many engineering applications as they enable prediction of the real-life system response, analysis of system behavior, and controller design to shape system behavior. Inevitably, most models fail to describe the dynamics of the system perfectly [152], but even if they do, there is a mismatch between the system and the model responses due to the following reasons. Firstly, most real-life, nonlinear systems evolve in continuous time, but their response is only available as a sampled, discrete-time series, implying information loss due to sampling. Secondly, only occasionally the response of a continuous-time model can be found analytically, e.g., the responses of linear time-invariant (LTI) models [135]. For nonlinear models, the response is typically computed by means of numerical integration, which has a discrete-time nature and is, inherently, prone to numerical inaccuracies [50, 252]. Thirdly, unmeasured and unknown disturbances affect the real-life system's inputs and outputs [152], which cannot be taken into account in model response computation. These observations suggest that there is a mismatch between the system response and the computed model response, even if a perfect model for the system is available.

Model response computation is essential for (nonlinear) model predictive control [4, 210], performance analysis of nonlinear (control) systems [110, 277, 286], parameter optimization of nonlinear control systems [196, 308], state estimation via (non)linear observers [30], (nonlinear) system identification [240, 251], and many more problems. Among these problems, [110, 196, 251, 277, 286, 308] require *steady-state* model responses, for which it is important to analyze and quantify the mismatch between the *steady-state* system and model responses. For example, the identification problem requires model responses of candidate models for comparison with the *measured* system response, thereby assessing the quality of the candidate model. Even if the candidate model perfectly describes the real-life system, the aforementioned sources of errors can result in this perfect candidate model being falsely labeled as a poor model.

Previous research has focused on the LTI case, see, e.g., [204, 205, 241]. For nonlinear systems, however, it is far more challenging to quantify the mismatch between the system and model responses. For example, one of the challenges is that nonlinear systems can produce for even the slightest change of the input a response that is wildly different, unbounded, or even undefined [135]. For responses to remain close to each other under input variations, the system should exhibit an incremental stability property, like incremental input-to-state stability [8]. This is even more pronounced given the fact that inputs to the real-life system typically contain an infinite number of harmonics, which are truncated in any computer simulation. Therefore, the inputs that excite the system and the model are different. Hence, the mismatch between the system response and model response can become unbounded and, therefore, unquantifiable. Furthermore, the mismatch depends highly on the particular model simulation algorithm used. Typically, nu-

merical forward integration (NFI) [252] is used, but other simulation algorithms, see, e.g., [15, 48, 50, 135, 195, 246, 276], are also proposed in the literature. These simulation algorithms, however, do not generically come with error bounds on the steady-state behavior.

Nevertheless, [34, 99, 213, 236, 237] analyze approximation errors when a continuous-time nonlinear model is replaced by a discrete-time one, enabling faster model simulation as the NFI simulation method for continuous-time models is computationally expensive. In particular, [99] shows that a continuous-time nonlinear system can be approximated arbitrarily well by a discrete-time nonlinear model with an approximation error over a finite time interval that drops inversely proportional with increasing sampling frequency. An improved result is presented in [236], where the drop rate of the approximation error can be improved by low-pass filtering the input signal before feeding it to the system. In the scope of controller design, [51, 180, 181] formulate conditions under which control design based on an approximate discrete-time model of a continuous-time system also stabilizes the continuous-time system.

In this chapter, we are interested in continuous-time models since these are derived naturally from the physical laws of nature, and their parameters have a physical meaning [94, 207]. To facilitate mismatch quantification, we focus on a practically relevant class of continuous-time nonlinear Lur'e-type systems that enjoy a stability property called *convergence* [200, 309]. Lur'e-type systems consist of LTI dynamics that are placed in feedback with a static nonlinearity. Convergence is a property of (non)linear systems that guarantees the uniqueness, boundedness and global asymptotic stability of the steady-state solution [200]. As a consequence, for any bounded excitation, the convergent model 'forgets' its initial condition and, therefore, exhibits a uniquely defined, bounded steady-state solution, which is globally asymptotic stable and depends solely on the applied input. For this class of models, [196] developed the so-called mixed time-frequency (MTF) algorithm for efficient computation of the unique steady-state model response. This algorithm has been used in system identification [247, 251] and control parameter optimization [196] problems. In these problems, accurate computation of steady-state model responses is essential. Although [196] does provide an upper bound on the mismatch between the steady-state model response and the calculated one, it does not take into account the implementation aspects of the algorithm and the corresponding aliasing errors. Moreover, it does not incorporate the interface of the model to the physical world, which gives rise to the aforementioned sources of errors.

This chapter derives two upper bounds on the steady-state mismatch between the system response and the computed model response by the computationally efficient MTF algorithm [196] for convergent Lur'e-type systems. The first bound is on the steady-state mismatch between the *measured sampled* system response and the model response calculated with the MTF algorithm. This bound provides a measure of how close the simulated model output is with respect to the

sampled measured system output including aliasing effects due to sampling, particularly important in system identification problems [240]. The second bound is on the steady-state mismatch between the *actual* system response and the model response calculated with the MTF algorithm. This bound provides insight into the intersample behavior of the system that is not captured by sampling its output response. We give an interpretation of the individual terms that comprise the bounds and show how these can be reduced. Furthermore, we showcase the usage of the bounds on a nonlinear mechanical system. Finally, we show that the bounds give valuable insights into the factors causing the mismatches and that a reduction of the bounds also results in a reduction of the mismatches.

The main contributions of this chapter are summarized as follows.

- The derivation of an upper bound on the steady-state mismatch between the simulated model response and (i) the *measured sampled* system response; and (ii) the *actual* system response, for the class of convergent Lur'e-type systems.
- A qualitative analysis of the different terms that contribute to these upper bounds and how to reduce the bounds.
- A quantitative analysis in which the results of this chapter are validated in a simulation example.

Firstly, these contributions facilitate parameter selection of the MTF simulation algorithm to control the tradeoff between simulation accuracy and computational load. Secondly, the contributions allow for deciding whether to invest resources to reduce disturbances or improve hardware to facilitate improved interfacing between the continuous-time and the discrete-time worlds.

The remainder of this chapter is outlined as follows. Section 4.2 introduces the notation and describes the preliminaries. Section 4.3 defines the considered dynamics, recalls the MTF-algorithm for model simulation, and formulates the problem setting. Section 4.4 presents the main results of this chapter in the form of the aforementioned bounds and, in addition, presents a qualitative analysis of the bounds. Section 4.5 describes a quantitative analysis and Section 4.6 presents the conclusions of this work.

4.2 Notation and preliminaries

The set of integer, real, and complex numbers are respectively denoted by $\mathbb{Z}, \mathbb{R}, \mathbb{C}$. The unitary complex number $i \coloneqq \sqrt{-1}$.

Continuous-time functions By $L_2(T)$ we denote the space of continuous, real-valued, T-periodic, scalar functions y satisfying $\|y\|_{L_2}^2 \coloneqq \frac{1}{T}\int_0^T |y(t)|^2 \mathrm{d}t < +\infty$. By ℓ_2 we denote the space of complex-valued sequences $Y = \{Y[m]\}_{m\in\mathbb{Z}}$ satisfying $\|Y\|_{\ell_2} < +\infty$, where $\|Y\|_{\ell_2}^2 = \sum_{m\in\mathbb{Z}} |Y[m]|^2$. Both $L_2(T)$ and ℓ_2 are Banach

spaces. The sequence of Fourier coefficients of $y \in L_2(T)$ is denoted by Y. Its elements are given by $Y[m] = \frac{1}{T}\int_0^T y(t)\exp\{-im\omega t\}\mathrm{d}t, m \in \mathbb{Z}$, where $\omega := 2\pi/T$. The inverse Fourier transform is given by $y(t) = \sum_{m\in\mathbb{Z}} Y[m]\exp\{im\omega t\}$. For any $y \in L_2(T)$, let Y be its Fourier coefficients. Then, Parseval's equality holds: $\|y\|_{L_2} = \|Y\|_{\ell_2}$. The truncation operator $(\cdot)_M$ is defined as follows:

$$(Y)_M[m] := \begin{cases} Y[m], & \text{for } -M \le m < M, \\ 0, & \text{else.} \end{cases} \tag{4.1}$$

The truncated version of Y is denoted by Y^M and its time-domain representation is denoted by y^M.

A real-valued T-periodic scalar function y is in the space $C^{k,\alpha}_{K_y}(T)$ of Hölder-α continuous functions for $0 < \alpha \le 1$, if its derivatives up to order $k \ge 0$ are continuous and its k-th order derivative, denoted by $y^{(k)}$, satisfies $|y^{(k)}(t) - y^{(k)}(t+\tau)| \le K_y|\tau|^\alpha$, $\forall\, t \in \mathbb{R}$, and $\forall\, \tau \in \mathbb{R}$, for some finite constant $K_y \ge 0$. In this chapter, we mainly use Hölder-α functions with $\alpha = 1$. The space of function where $k = 0$ and $\alpha = 1$ coincides with the space of Lipschitz continuous functions. Note that $C^{k,\alpha}_K(T) \subset L_2(T)$, for any $T \ge 0, 0 \le K < +\infty, 0 < \alpha \le 1$, and $k \ge 0$.

Discrete-time sequences Discrete sequences are written in bold fonts throughout this chapter. By L_2^M, we denote the space of real-valued, scalar sequences $\mathbf{y}_M := \{\mathbf{y}[k]\}_{k=1}^{2M}$ satisfying $\|\mathbf{y}_M\|^2_{L_2^M} := \frac{1}{2M}\sum_{k=1}^{2M}|\mathbf{y}[k]|^2 < +\infty$. By ℓ_2^M, we denote the space of complex-valued scalar sequences $\mathbf{Y}_M := \{\mathbf{Y}[m]\}_{m\in\mathcal{M}}$ satisfying $\|\mathbf{Y}_M\|^2_{\ell_2^M} := \sum_{m\in\mathcal{M}}|\mathbf{Y}[m]|^2 < +\infty$ with $\mathcal{M} := \{-M, -M+1, \ldots, M-2, M-1\}$. Both L_2^M and ℓ_2^M are Banach spaces. For a given $\mathbf{y}_M$, the elements of the sequence $\mathbf{Y}_M$ are given by the discrete Fourier transform (DFT) $\mathbf{Y}[m] = \sum_{k=1}^{2M}\mathbf{y}[k]\exp\{-i\frac{\pi}{M}m(k-1)\}$, for $m \in \mathcal{M}$. The inverse DFT (IDFT) is given by $\mathbf{y}[k] = \frac{1}{2M}\sum_{m\in\mathcal{M}}\mathbf{Y}[m]\exp\{i\frac{\pi}{M}m(k-1)\}$, for $k = 1, \ldots, 2M$. For any $\mathbf{y}_M \in L_2^M$ and its corresponding Fourier coefficients $\mathbf{Y}_M$, the following equality holds: $\sqrt{2M}\|\mathbf{y}_M\|_{L_2^M} = \|\mathbf{Y}_M\|_{\ell_2^M}$.

Relation between discrete-time sequences and continuous-time signals All non-zero Fourier coefficients of the truncated T-periodic $y^M \in L_2(T)$ can be computed from the time-series $\mathbf{y}_M = \{\mathbf{y}[k]\}_{k=1}^{2M} = \{y^M(t_k)\}_{k=1}^{2M}$, where $2M$ is the length of the time series and $t_k = t_s k$ are the sampling instants with sampling interval $t_s := T/(2M)$. The relation between the Fourier coefficients $\mathbf{Y}_M$ and Y^M is as follows:

$$Y^M[m] = \begin{cases} \frac{1}{2M}\mathbf{Y}[m], & \text{for } m \in \mathcal{M}, \\ 0, & \text{else.} \end{cases} \tag{4.2}$$

Therefore, having the time-series $\mathbf{y}_M$ at hand, the truncated continuous-time function y^M can be computed on $t \in \mathbb{R}$ via the Fourier coefficients relationship in (4.2). Moreover, the relation (4.2) ensures $\|\mathbf{y}_M\|_{L_2^M} = \|y^M\|_{L_2}$.

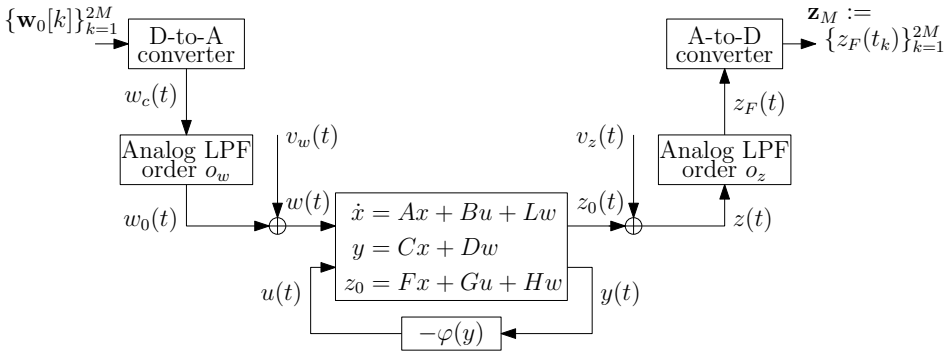


Figure 4.1. Schematic representation of Lur'e-type system (4.3).

4.3 Problem setting

The considered setup is depicted schematically in Fig. 4.1 and its components are detailed in the subsequent sections. The considered problem is to find bounds on the steady-state mismatch between the system response and the computed model response. In Section 4.3.4, we formalize the problem studied in this chapter.

4.3.1 Class of convergent Lur'e-type systems

The starting point of this work is a model that perfectly describes the system under study. The model and system equations coincide and are of single-input single-output (SISO) Lur'e-type form described by

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Lw(t), \\ y(t) &= Cx(t) + Dw(t), \\ z_0(t) &= Fx(t) + Gu(t) + Hw(t), \\ u(t) &= -\varphi(y(t)), \end{aligned} \tag{4.3}$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $y(t) \in \mathbb{R}$ is the nonlinearity input signal, $u(t) \in \mathbb{R}$ is the nonlinearity output signal, $z_0(t) \in \mathbb{R}$ is the output, $z(t) = z_0(t) + v_z(t) \in \mathbb{R}$ is the measured output, $w_0(t) \in \mathbb{R}$ is the realized external input, $w(t) = w_0(t) + v_w(t) \in \mathbb{R}$ is the input to the system, and $\varphi : \mathbb{R} \to \mathbb{R}$ is a static nonlinear function. The signals $v_w(t) \in \mathbb{R}$ and $v_z(t) \in \mathbb{R}$ are unmeasured input and output disturbances, respectively.

It is assumed that the dynamics (4.3) exhibit a strong form of model stability. Hereto, the notion of convergent dynamics is defined according to [200].

Definition 4.1. *The system* (4.3) *is said to be globally exponentially convergent if for every input w that is defined and bounded on $\mathbb{R}$, there exists a solution $\bar{x}$ satisfying the conditions:*

- *$\bar{x}$ is defined and bounded on $\mathbb{R}$,*
- *$\bar{x}$ is globally exponentially stable.*

■

The solution $\bar{x}$ is called the *steady-state* solution and depends on the applied input w. For the considered dynamics (4.3), sufficient conditions for global exponential convergence are recalled next from [196, 309].

Theorem 4.1. *Consider system* (4.3) *and suppose that*

C1 *the matrix A is Hurwitz;*

C2 *there exists a $K_\varphi > 0$ such that the nonlinearity $\varphi(y)$ satisfies $\left|\frac{\varphi(y_2)-\varphi(y_1)}{y_2-y_1}\right| \leq K_\varphi$ for all $y_1 \in \mathbb{R}, y_2 \in \mathbb{R}$.*

C3 $\gamma_{yu} \coloneqq \sup_{\omega\in\mathbb{R}} |C(j\omega I - A)^{-1}B| < \frac{1}{K_\varphi}$.

Then, the system (4.3) *is globally exponentially convergent.* ▲

Assumption 4.1. *System* (4.3) *satisfies conditions **C1** - **C3** of Theorem 4.1 for some constant $K_\varphi \geq 0$ and is, therefore, globally exponentially convergent.* ■

Independent of the initial condition, all solutions x of a convergent system converge to the globally exponentially stable steady-state solution $\bar{x}$, which depends on the input w. Another property of convergent systems is that when excited by a periodic input w with period-time T, the steady-state solution $\bar{x}$ is also periodic with the same period-time T, see [200]. This property implies that the steady-state output $\bar{z}_0$ only contains the harmonics of the periodic input w and its higher harmonics. Finally, the conditions of Theorem 4.1 guarantee a finite incremental gain, which implies that small variations of the input w (in the infinity norm) lead to small variations in the output z (in the infinity norm) [200]. This property is instrumental in quantifying the effect of the disturbances v_w and v_z on the steady-state system response $\bar{z}$. From here onwards, signals corresponding to the system are written with a breve diacritic mark, e.g., the state $\breve{x}$, whereas signals corresponding to the model are written without a breve diacritic mark, e.g., the state x.

Remark 4.1. Assuming that the model perfectly describes the system might seem restrictive, however, in many systems and control-related problems, this assumption is made. For example, in system identification problems, this assumption is commonly used in the Cramér-Rao bound for statistical efficiency [152] and optimal input design problems [268]. △

4.3.2 Interface between the continuous-time system and discrete-time computer

Input signal reconstruction

Let us first discuss the input signal reconstruction of w_0 via the digital-to-analog (D-to-A) converter, see Figure 4.1. In practice, the time-series $\mathbf{w}_{(0,M)} := \{\mathbf{w}_0[k]\}_{k=1}^{2M}$ of the user-defined $T-$periodic w_0 is available. The following assumption formalizes the realization of w_0.

Assumption 4.2. *The T-periodic signal w_c (see Figure 4.1) is realized by a D-to-A converter with the refreshing frequency $f_s := 1/t_s$ from the user-defined time-series $\mathbf{w}_{(0,M)}$ and with t_s the refreshing time. Then, the signal w_c is filtered by a low-pass filter with transfer function $G(s) = (\omega_w/(s+\omega_w))^{o_w}, s \in \mathbb{C}$, of order $o_w > 0$ and with cut-off frequency ω_w to realize the input signal w_0. The constant $0 \leq K_{w_0}^{\circ} < +\infty$ related to the realized signal $w_0 \in C_{K_{w_0}^{\circ}}^{0,1}(T)$ is known. Moreover, the constant $0 \leq K_{w_0}^{\eta M} < +\infty$ related to the truncated signal $w_0^{\eta M} \in C_{K_{w_0}^{\eta M}}^{0,1}(T)$ is known.* ■

A popular D-to-A conversion technique is the so-called zero-order hold (ZOH) reconstruction technique, which holds the sample $\mathbf{w}_0[k]$ for $t \in [t_k, t_{k+1})$. It is well-known that, in addition to the desired harmonics, also unwanted higher harmonics are realized by any D-to-A conversion technique and are present in the periodic signal w_c. Therefore, Assumption 4.2 implies that both w_c and w_0 contain frequency contents beyond the first M harmonics, which are attenuated up to some extent by the low-pass filtering process.

To compute the bounds derived later in this chapter, all Fourier coefficients of w_0 are required. These can be computed analytically from the samples $\mathbf{w}_{(0,M)}$ under the assumption that the D-to-A conversion technique is known and that the order o_w, and cut-off frequency ω_w of the low-pass filter are known.

Remark 4.2. The low-pass filter order o_w is considered strictly larger than 0 by Assumption 4.2. Consequently, the low-passed signal w_0 is Hölder-α continuous with respect to time, even if the output w_c of the D-to-A converter is discontinuous. The condition that the signal w_0 is Hölder-α continuous is required for the main results of this chapter. △

Data acquisition

Next, we discuss the sampling of the output $\breve{z}$ of the system (4.3). In practice, only a discrete version of the measured steady-state response $\breve{\mathbf{z}}$ is available at the sampling instants $t_k := t_0 + kt_s$ with $k = 1, \ldots, 2M$, where t_s is the sampling interval, $2M = T/t_s$ is the total number of samples in *one* steady-state period and $t_0 \in \mathbb{R}$ is the starting time. To attenuate aliasing effects, we low-pass filter the

steady-state signal $\breve{\breve{z}}$ prior to sampling it, see Figure 4.1, which is formalized in the following assumption.

Assumption 4.3. *The steady-state signal $\breve{\breve{z}}$ is filtered by a low-pass filter with transfer function $G(s) = (\omega_z/(s+\omega_z))^{o_z}, s \in \mathbb{C}$, of order $o_z \geq 0$ and with cut-off frequency $\omega_z > 0$ to obtain the steady-state signal $\breve{\breve{z}}_F$. It is assumed that the first $2M$ Fourier coefficients of $\breve{\breve{z}}_F$ are equal to those of $\breve{\breve{z}}$. The signal $\breve{\breve{z}}_F$ is then sampled with sampling frequency $f_s \coloneqq 1/t_s$ to obtain $\breve{\mathbf{z}}_M \coloneqq \{\breve{\breve{z}}_F(t_k)\}_{k=1}^{2M}$.* ■

Remark 4.3. For the sake of clarity, from here on we only cover the case in which the sampling time of the A-to-D converter coincides with the refreshment time of the D-to-A converter, which are both denoted by t_s. △

Remark 4.4. In contrast to Assumption 4.3, the low-pass filter with $o_z \geq 1$ also affects the first $2M$ Fourier coefficients of $\breve{z}_F$. Nevertheless, this assumption is posed as it significantly simplifies the analysis in this chapter. An alternative interpretation is that the effect of low-pass filtering is negligible on the first $2M$ Fourier coefficients of $\breve{z}_f$, which is true for a sufficiently large cut-off frequency ω_z. △

Disturbance setting

The input disturbance v_w affects the input w_0 of the system such that the input to the system is $w = w_0 + v_w$, see Figure 4.1. An example of input disturbances is actuator limitations resulting in a non-perfect realization of the user defined-input w_0. Another example is given by external unmeasurable inputs, such as environmental time-varying factors that excite the system under study.

The output disturbance v_z affects the output $\breve{z}_0$ of the system, such that the available output for measurement is $\breve{z} = \breve{z}_0 + v_z$, see Figure 4.1. This disturbance is typically caused by imperfections in the measurement device. For example, many measurement devices are subject to environmental effects such as temperature. Although sensor calibration reduces these effects to some extent, these effects can not be eliminated. Another example is quantization errors, caused by the finite number of bits used to represent the measured analog signal.

The disturbances are assumed unmeasured; however, these are assumed to be periodic and bounded in norm with a known bound. Furthermore, these disturbances are assumed to admit a finite Lipschitz constant. These assumptions are required to bound the effects of the disturbances in the accuracy bounds derived in this chapter.

Assumption 4.4. *The disturbances $v_w \in L_2(T)$ and $v_z \in L_2(T)$ satisfy $\|v_w\|_{L_2} \leq \Delta_w, \|v_z\|_{L_2} \leq \Delta_z$ for known constant $0 \leq \Delta_w < +\infty$ and $0 \leq \Delta_z < +\infty$.* ■

Assumption 4.5. *The disturbances $v_w \in C^{0,1}_{K_{v_w}}(T)$ and $v_z \in C^{0,1}_{K_{v_z}}(T)$ have known constants $0 \leq K_{v_w} < +\infty$ and $0 \leq K_{v_z} < +\infty$, respectively.* ■

Note that we consider a deterministic disturbance setting and exclude a stochastic setting. In a stochastic setting, the signals v_w and v_z are realizations of stochastic processes that are defined by moments, i.e., mean, variance, and higher-order moments. Stochastic signals with a positive variance do not obey the bound in Assumption 4.4 and, therefore, do not fit in the presented class of disturbances for any finite Δ_w and Δ_z.

4.3.3 Fast computation of steady-state model responses

The steady-state model response is computed on a discrete-time grid with sampling frequency ηf_s, with η a positive integer and $1/f_s$ the sample time in Assumptions 4.2 and 4.3. In each iteration i of the MTF algorithm (the iteration index i is denoted in the subscript in brackets, i.e., $[i]$), the following two mappings are evaluated:

$$\mathbf{u}_{([i+1],\eta M)} = g_{uy}\mathbf{y}_{([i],\eta M)}, \tag{4.4a}$$

$$\mathbf{y}_{([i+1],\eta M)} = g_{yu}\mathbf{u}_{([i+1],\eta M)} + g_{yw}\mathbf{w}_{(0,\eta M)}, \tag{4.4b}$$

where the sequences $\mathbf{w}_{(0,\eta M)} \coloneqq \{w_0^{\eta M}(t_k)\}_{k=1}^{2\eta M}$, $\mathbf{y}_{([i],\eta M)} \coloneqq \{\mathbf{y}_{[i]}[k]\}_{k=1}^{2\eta M}$, and $\mathbf{u}_{([i+1],\eta M)}$ is defined similarly to $\mathbf{y}_{([i+1],\eta M)}$. The nonlinear steady-state operator g_{uy} in (4.4a) is defined as follows:

$$g_{uy}\mathbf{y}_{([i],\eta M)} \coloneqq -\varphi(\mathbf{y}_{([i],\eta M)}), \tag{4.5}$$

whereas the linear steady-state operators g_{yu} and g_{yw} in (4.4b) map $\mathbf{u}_{([i+1],\eta M)}$ and $\mathbf{w}_{(0,\eta M)}$ into $\mathbf{y}_{([i+1],\eta M)}$, respectively. Their frequency-domain counterparts $\hat{g}_{yu}$ and $\hat{g}_{yu}$, respectively, are defined as follows:

$$(\hat{g}_{yu}\mathbf{U}_{\eta M})[m] \coloneqq G_{yu}(j\omega_0 m)\mathbf{U}[m], \tag{4.6a}$$

$$(\hat{g}_{yw}\mathbf{W}_{(0,\eta M)})[m] \coloneqq G_{yw}(j\omega_0 m)\mathbf{W}_0[m], \tag{4.6b}$$

for $m \in \eta\mathcal{M} \coloneqq \{-\eta M, -\eta M + 1, \ldots, \eta M - 2, \eta M - 1\}$, where $\omega_0 = 2\pi/T$ and with the frequency response functions $G_{yu}(j\omega) \coloneqq C(j\omega I - A)^{-1}B$ and $G_{yw}(j\omega) \coloneqq C(j\omega I - A)^{-1}L + D$. The MTF algorithm relies on the fact that the composed operator $g_{uy} \circ g_{yu}$ is a contraction operator acting from the space $L_2^{\eta M}$ to $L_2^{\eta M}$, provided that the model satisfies the conditions of Theorem 4.1.

Theorem 4.2. *Consider the dynamics* (4.3) *and suppose Assumption 4.1 holds. Then, for any integers $\eta > 0$ and $M > 0$, and for any T-periodic input $w_0 \in L_2(T)$, there exists a unique limit $\bar{\mathbf{y}}_{\eta M}$ to the sequence $\mathbf{y}_{([i],\eta M)}, i = 1, 2, \ldots$, resulting from the iterative process* (4.4a) - (4.4b). ▲

Proof. The proof can be found in Appendix C.1. □

For computational efficiency, the nonlinear mapping g_{uy} is evaluated in the time domain, while the two linear mappings are evaluated in the frequency domain. Starting from any initial $\mathbf{y}_{([0],\eta M)}$, e.g., the zero response, the MTF algorithm iteratively evaluates these mappings, while transforming the intermediate signals between the time- and frequency-domain using the (inverse) DFT. Having the steady-state $\bar{\mathbf{y}}_{\eta M}$, one can trivially compute $\bar{\mathbf{u}}_{\eta M}$ using the nonlinear operator g_{uy} and compute the output $\bar{\mathbf{z}}_{\eta M}$ using the frequency response functions $G_{zu}(j\omega) \coloneqq F(j\omega I - A)^{-1}B + G$ and $G_{zw}(j\omega) \coloneqq F(j\omega I - A)^{-1}L + H$. The MTF algorithm is presented in Algorithm 4.1 with the stop function:

$$\epsilon(\mathbf{Y}_{([i],\eta M)}, \mathbf{Y}_{([i-1],\eta M)}) = \frac{\left\|\mathbf{Y}_{([i],\eta M)} - \mathbf{Y}_{([i-1],\eta M)}\right\|_{\ell_2^M}}{\left\|\mathbf{Y}_{([i-1],\eta M)}\right\|_{\ell_2^M}}.$$

This algorithm is closely related to the harmonic balancing method [144, 168]. Furthermore, [46] proposed an algorithm similar to the MTF algorithm, however, without proof of convergence.

Remark 4.5. A different version of the MTF algorithm has been presented in [196] with a proof of contraction. Numerical implementation of that previous version requires choosing additional parameters that affect the accuracy of the steady-state solution. These effects were not studied in [196]. In Algorithm 4.1 and Theorem 4.2 these parameters are explicitly taken into account, which allows deriving the accuracy bounds in Section 4.4. △

Algorithm 4.1 Mixed Time-Frequency Algorithm.

1: Evaluate the LTI dynamics in frequency domain $\left\{\mathbf{Y}_{[0]}[m] = G_{yw}(j\omega_0 m)\mathbf{W}_0[m]\right\}_{m\in\eta\mathcal{M}}$.
2: Compute $\mathbf{y}_{([0],\eta M)}$ from $\mathbf{Y}_{([0],\eta M)}$.
3: Set iteration counter $i = 0$.
4: **while** $\epsilon(\mathbf{Y}_{([i],\eta M)}, \mathbf{Y}_{([i-1],\eta M)}) > Y^\star$ or $i < 1$ **do**
5: Evaluate the nonlinearity in time domain $\{\mathbf{u}_{[i+1]}[k] = -\varphi(\mathbf{y}_{[i]}[k])\}_{k=1}^{2\eta M}$.
6: Compute $\mathbf{U}_{([i+1],\eta M)}$ from $\mathbf{u}_{([i+1],\eta M)}$.
7: Evaluate the LTI dynamics in frequency domain $\left\{\mathbf{Y}_{[i+1]}[m] = G_{yu}(j\omega_0 m)\mathbf{U}_{[i+1]}[m] + Y_{[0]}[m]\right\}_{m\in\eta\mathcal{M}}$.
8: Compute $\mathbf{y}_{([i+1],\eta M)}$ from $\mathbf{Y}_{([i+1],\eta M)}$.
9: Set $i = i + 1$.
10: **end while**
11: Evaluate the LTI dynamics in frequency domain $\left\{\mathbf{Z}[m] = G_{zw}(j\omega_0 m)\mathbf{W}_0[m] + G_{zu}(j\omega_0 m)\mathbf{U}_{[i-1]}[m]\right\}_{m\in\eta\mathcal{M}}$.
12: Compute $\mathbf{z}_{\eta M}$ from $\mathbf{Z}_{\eta M}$.

4.3.4 Problem statement

Several error sources are responsible for a steady-state mismatch between the system and the model responses, even if the nonlinear model perfectly describes the real-life system. Firstly, the MTF Algorithm 4.1 considering only ηM harmonics results in higher harmonics of the input signal w_0 being neglected during simulation; and internal aliasing errors due to the nonlinear feedback φ generating an infinite number of harmonics during model simulation. Secondly, aliasing errors arise due to the sampling of the system output by a finite sampling frequency. Thirdly, unknown and unmeasured input and output disturbances affect the system response but are unavailable for model simulation.

We are interested in the steady-state mismatch between the computed model response $\bar{z}^{\eta M}$ and the *actual* system response $\breve{\bar{z}}$, both available as continuous-time functions. Note that the computed steady-state model response $\bar{z}^{\eta M}$ is obtained by first computing $\bar{\mathbf{z}}_{\eta M}$ on a discrete grid of $2\eta M$ points by means of the MTF algorithm. After that, $\bar{\mathbf{z}}_{\eta M}$ is transformed via (4.2) to its continuous-time counterpart $\bar{z}^{\eta M}$, which only contains ηM harmonics by definition. Moreover, we are also interested in the steady-state mismatch between the truncated model response $\left(\bar{\mathbf{z}}_{\eta M}\right)^M$ and the *measured sampled* system response $\breve{\bar{\mathbf{z}}}_M$, both available at $2M$ discrete time instances. Here, the model response $\bar{\mathbf{z}}_{\eta M}$ is truncated to M harmonics to match the number of $2M$ discrete time instances of the system response.

Problem 4.1. *Consider the dynamics* (4.3)*. Suppose Assumptions 4.1 - 4.4 hold. Find an upper bound for* $\left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2}$*, where* $\breve{\bar{z}}$ *denotes the steady-state system response and* $\bar{z}^{\eta M}$ *denotes the model response computed by the MTF Algorithm 4.1.* △

Problem 4.2. *Consider the dynamics* (4.3)*. Suppose Assumptions 4.1 - 4.5 hold. Find an upper bound for* $\left\|\breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M\right\|_{L_2^M}$*, where* $\breve{\bar{\mathbf{z}}}_M$ *denotes the measured sampled steady-state system response and* $\left(\bar{\mathbf{z}}_{\eta M}\right)^M$ *denotes the truncated model response computed by the MTF Algorithm 4.1.* △

4.4 Bounds on mismatch

The main result is presented in Section 4.4.1 in Theorems 4.3 and 4.4, which, respectively, solve Problems 4.1 and 4.2. After that, in Section 4.4.2, a qualitative analysis that decomposes these bound into several components and analyzes their asymptotic behavior is presented. Finally, in Section 4.4.3, several methods to reduce these bounds are given.

4.4.1 Bound on the steady-state mismatch between the system and the model response

The bounds on the steady-state mismatch between the (measured) system response and the computed model response are presented in Theorems 4.3 and 4.4. The following constants are used with $\omega \coloneqq 2\pi/T$:

$$\gamma_\mu \coloneqq \sup_{m\in\mathbb{Z}} |G_\mu(jm\omega)|, \quad \bar{\gamma}_\mu^\chi \coloneqq \sup_{|m|>\chi} |G_\mu(jm\omega)|, \quad \gamma_\mu^\chi \coloneqq \sup_{|m|\leq\chi} |G_\mu(jm\omega)|, \tag{4.7}$$

where $\mu \in \{yu, yw, zu, zw\}$ and $\chi \in \{M, \eta M\}$. Furthermore, we introduce the following two constants:

$$\gamma_a^{\eta M} \coloneqq \frac{TK_\varphi \gamma_{yw}^{\eta M}}{\sqrt{8}\left(1-\gamma_{yu}^{\eta M}K_\varphi\right)} \sqrt{\zeta(2) - \sum_{m=1}^{\eta M-1} \frac{1}{m^2}}, \tag{4.8a}$$

$$\gamma_F^M \coloneqq \frac{T\omega_z^{o_z}}{\sqrt{8}} \sqrt{\sum_{m=M}^{\infty} \frac{1}{m^2\left(m^2\omega_0^2+\omega_z^2\right)^{o_z}}}, \tag{4.8b}$$

where o_z, ω_z are the low-pass filter parameters in Assumption 4.3. The function $\zeta(s)$ with $s \in \mathbb{C}$ is the so-called Riemann-zeta function defined as follows:

$$\zeta(s) \coloneqq \sum_{n=1}^{\infty} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \dots. \tag{4.9}$$

We only use the function $\zeta(s)$ for $s = 2$, for which the infinite series (4.9) converges to $\pi^2/2$, see [278]. This also guarantees convergence of $\gamma_a^{\eta M}$ in (4.8a). Furthermore, the infinite series in the constant γ_F^M also converges, as shown in the proof of Theorem 4.4 in Appendix C.4.

Prior to presenting the bounds on the mismatch, we first present a technical lemma.

Lemma 4.1. *Consider the dynamics* (4.3) *and suppose Assumptions 4.1 and 4.2 hold. Suppose that* $\bar{y}^{\eta M}$ *is computed by means of the MTF Algorithm 4.1. Then, the magnitude of the Fourier coefficients* $\bar{U}$ *of* $\bar{u} = \varphi(\bar{y}^{\eta M})$*, for* $m \neq 0$*, are upper bounded as follows:*

$$\left|\bar{U}\,[m]\right| \leq \frac{TK_\varphi K_{w_0}^{\eta M}\gamma_{yw}^{\eta M}}{4\,|m|\,(1-\gamma_{yu}^{\eta M}K_\varphi)}, \tag{4.10}$$

where K_φ *is the constant in Assumption 4.1,* $K_{w_0}^{\eta M}$ *the constant in Assumption 4.2, and* $\gamma_{yu}^{\eta M}$ *and* $\gamma_{yw}^{\eta M}$ *are as in* (4.7)*. Furthermore, the upper bound*

$$\left\|\varphi(\bar{y}^{\eta M}) - \left(\varphi\left(\bar{y}^{\eta M}\right)\right)^{\eta M}\right\|_{L_2} \leq \gamma_a^{\eta M} K_{w_0}^{\eta M} \tag{4.11}$$

with $\gamma_a^{\eta M}$ *defined in* (4.8a)*, holds.* ▲

Proof. The proof can be found in Appendix C.2. □

The following result presents the bound on the mismatch between the *actual* system response $\breve{\bar{z}}$ and the model response $\bar{z}^{\eta M}$ calculated with the MTF algorithm. Thereby, it solves Problem 4.1.

Theorem 4.3. *Suppose Assumptions 4.1 - 4.4 hold for system* (4.3) *with the steady-state output response* $\breve{\bar{z}}$*. If the steady-state output of the model* (4.3)*, denoted by* $\bar{z}^{\eta M}$*, is computed using the MTF Algorithm 4.1, then the following upper bound on the* L_2*-norm on the mismatch between* $\breve{\bar{z}}$ *and* $\bar{z}^{\eta M}$ *solves Problem 4.1:*

$$\left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2} \leq \kappa_1 \left\|w_0 - w_0^{\eta M}\right\|_{L_2} + \kappa_2 \Delta_w + \Delta_z + \kappa_3 K_{w_0}^{\eta M} \tag{4.12}$$

with

$$\kappa_1 \coloneqq \frac{K_\varphi \gamma_{zu} \bar{\gamma}_{yw}^{\eta M}}{1 - \gamma_{yu} K_\varphi} + \bar{\gamma}_{zw}^{\eta M}, \quad \kappa_2 \coloneqq \frac{K_\varphi \gamma_{zu} \gamma_{yw}}{1 - \gamma_{yu} K_\varphi} + \gamma_{zw}, \quad \kappa_3 \coloneqq \frac{2\gamma_a^{\eta M}}{1 - \gamma_{yu} K_\varphi},$$

and the constants $\gamma_{yu}, \gamma_{yw}, \gamma_{zw}, \gamma_{zu}, \bar{\gamma}_{yw}^{\eta M}, \bar{\gamma}_{zw}^{\eta M}$*, defined in* (4.7)*, the constant* $\gamma_a^{\eta M}$ *defined in* (4.8a)*,* K_φ *as in Assumption 4.1,* $K_{w_0}^{\eta M}$ *as in Assumption 4.2 and* Δ_w, Δ_z *as in Assumption 4.4. Furthermore, if* $\Delta_w = \Delta_z = 0$*, then:*

$$\lim_{M \to \infty} \left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2} = \lim_{\eta \to \infty} \left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2} = 0. \tag{4.13}$$

▲

Proof. The proof can be found in Appendix C.3. □

Similarly, we present the bound on the mismatch between the *measured* system response and the truncated model response as a solution to Problem 4.2.

Theorem 4.4. *Suppose Assumptions 4.1 - 4.5 hold for system* (4.3) *with the measured steady-state output response* $\breve{\bar{\mathbf{z}}}_M$*. If the truncated steady-state output of the model* (4.3)*, denoted by* $\left(\bar{\mathbf{z}}_{\eta M}\right)^M$*, is computed using the MTF Algorithm 4.1, then the following upper bound on the* L_2^M*-norm on mismatch between* $\breve{\bar{\mathbf{z}}}_M$ *and* $\left(\bar{\mathbf{z}}_{\eta M}\right)^M$ *solves Problem 4.2:*

$$\begin{aligned}\left\|\breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M\right\|_{L_2^M} \leq \xi_1 \left\|w_0 - w_0^{\eta M}\right\|_{L_2} + \xi_2 \Delta_w + \Delta_z \\ + \xi_3 K_{w_0}^{\eta M} + \xi_4 K_{w_0}^{\circ} + \xi_5 K_{v_w} + \xi_6 K_{v_z}\end{aligned} \tag{4.14}$$

with

$$\begin{aligned}\xi_1 &\coloneqq \frac{K_\varphi \gamma_{zu}^M \bar{\gamma}_{yw}^{\eta M}}{1 - \gamma_{yu} K_\varphi}, & \xi_2 &\coloneqq \frac{K_\varphi \gamma_{zu}^M \gamma_{yw}}{1 - \gamma_{yu} K_\varphi} + \gamma_{zw}^M, \\ \xi_3 &\coloneqq \frac{2\gamma_a^{\eta M} \gamma_{zu}^M}{1 - \gamma_{yu} K_\varphi}, & \xi_4 &\coloneqq \gamma_F^M \left(\frac{\gamma_{zu} \gamma_{yw}}{1 - \gamma_{yu} K_\varphi} + \gamma_{zw}\right), \\ \xi_5 &\coloneqq \xi_4, & \xi_6 &\coloneqq \gamma_F^M,\end{aligned}$$

and the constant $\gamma_{yu}, \gamma_{yw}, \gamma_{zu}, \gamma_{zw}, \gamma_{zu}^M, \gamma_{zw}^M, \bar{\gamma}_{yw}^{\eta M}$ *defined in* (4.7), $\gamma_a^{\eta M}$ *defined in* (4.8a), γ_F^M *defined in* (4.8b), K_φ *as in Assumption 4.1,* $K_{w_0}^\circ$ *and* $K_{w_0}^{\eta M}$ *as in Assumption 4.2,* Δ_w, Δ_x *as in Assumption 4.4, and* K_{v_w}, K_{v_z} *as in Assumption 4.5. Furthermore, if* $\Delta_w = \Delta_z = K_{v_w} = K_{v_z} = 0$*, then for any integer* $\eta > 0$*, the condition*

$$\lim_{M\to\infty} \left\| \breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M \right\|_{L_2^M} = 0 \tag{4.15}$$

holds. ▲

Proof. The proof can be found in Appendix C.4. □

Remark 4.6. Increasing the low-pass filter order o_w in Assumption 4.2 results in an increased smoothness of the input signal $w_0(t)$. However, it is not yet clear how to exploit these smoothness properties to derive tighter bounds in Theorems 4.3 and 4.4. △

4.4.2 Qualitative analysis

For each of the individual components of both bounds, an interpretation is presented first. After that, the hypothetical case in which we have an infinitely fast sampling frequency f_s or an infinitely dense time grid in the MTF algorithm is analyzed. Next, the special case for LTI systems is treated. Finally, the results are summarized in Table 4.1 at the end of this section.

Component-level interpretation

The first component in the bounds in (4.12) and (4.14) accounts for the frequency contents in w_0 beyond the first ηM harmonics. These contents are inevitably realized by any D-to-A conversion technique and propagate to the system output, while these are missing in the output computed by the MTF Algorithm 4.1. Furthermore, these contents also influence the system response at the first ηM harmonics due to the nonlinear feedback. The second component accounts for the unknown input disturbance v_w that affects the system output. The third component accounts for the unknown output disturbance, which is directly observed in the system output $\breve{z}$, while its low-passed version is visible in $\breve{z}_F$. The fourth component, related to $K_{w_0}^{\eta M}$, is due to internal aliasing in the MTF Algorithm 4.1 due to the limited number of harmonics that can be considered. The fifth component, related to K_{w_0}, is due to aliasing errors during the sampling of the system output and is, therefore, only present in the bound in Theorem 4.4. The remaining components, related to K_{v_w} and K_{v_z} in the bound in Theorem 4.4 are due to the disturbances v_w and v_z generating contents beyond the first M harmonics in the system output and give rise to aliasing errors during the sampling of the system output $\breve{z}_F$.

Asymptotic analysis

Theorems 4.3 and 4.4 show that the respective mismatches are zero for infinitely large η and/or M and no disturbances. This is an important result, as it shows that faster sampling and/or increased accuracy of MTF Algorithm 4.1 reduces the analyzed mismatches. In this section, we give the interpretation for this asymptotic result and treat the case in which the disturbances are non-zero. Before we do so, we note that for either $\eta \to \infty$ or $M \to \infty$, we have $\|w_0 - w_0^{\eta M}\|_{L_2} = 0$, so the first component in either bound is irrelevant in the asymptotic case and is not further treated below.

The second component is related to the input disturbance v_w. We have the following result:

$$\lim_{\eta\to\infty} \kappa_2 = \lim_{M\to\infty} \kappa_2 = \lim_{M\to\infty} \xi_2 = \gamma_{zw} + \frac{K_\varphi \gamma_{zu} \gamma_{yw}}{1 - \gamma_{yu} K_\varphi}$$

and

$$\lim_{\eta\to\infty} \xi_2 = \gamma_{zw}^M + \frac{K_\varphi \gamma_{zu}^M \gamma_{yw}}{1 - \gamma_{yu} K_\varphi}.$$

In the asymptotic case, there is an error due to the input disturbance v_w, which is expected as it only affects the system output, independent of M or η, but not the computed model output. Similarly, the output disturbance v_z affects the system output, independent of M or η, but not the computed model output.

The remaining components are related to $K_{w_0}^{\eta M}, K_{w_0}^{\circ}, K_{v_w}$ and K_{v_z} and are due to aliasing effects. The following result holds for the components related to $K_{w_0}^{\eta M}$:

$$\lim_{\eta\to\infty} \kappa_3 = \lim_{M\to\infty} \kappa_3 = \lim_{\eta\to\infty} \xi_3 = \lim_{M\to\infty} \xi_3 = 0. \tag{4.16}$$

This result shows that there are no internal aliasing errors in the MTF Algorithm 4.1 if either η or M is increased, which implies that the simulation is performed on a dense time grid. For the components related to $K_{w_0}^{\circ}, K_{v_w}$ and K_{v_z}, the following result holds:

$$\lim_{M\to\infty} \xi_4 = \lim_{M\to\infty} \xi_5 = \lim_{M\to\infty} \xi_6 = 0.$$

This result implies that there are no aliasing errors due to the sampling of the system output with an infinitely large sampling frequency. The coefficients $\xi_4, \ldots, \xi_6$, are independent of η, as aliasing due to the sampling of the system output only depends on the sampling frequency related to M.

The special case of LTI systems

As all the LTI dynamics are captured in the LTI block, the special case of LTI systems is covered by taking a zero nonlinearity, i.e., $K_\varphi = 0$. In this case, it can be trivially verified that in the bound (4.12), $\kappa_1 = \bar{\gamma}_{zw}^{\eta M}, \kappa_2 = \gamma_{zw}$ and $\kappa_3 = 0$.

This implies that there is only a steady-state mismatch between the *actual* system response and the computed model response due to the harmonics beyond ηM in the input w_0 and due to the input and output disturbances. Given the fact that $\kappa_3 = 0$, we conclude that there is no mismatch due to aliasing in the MTF algorithm, which shows that the bound preserves the superposition property of LTI systems. In essence, it guarantees that if the input w_0 contains content beyond the first ηM harmonics, then this content does not influence the first ηM harmonics of the response of the LTI system.

The bound (4.14) between the *measured* system output and the computed model output is treated next. First of all, for $K_\varphi = 0$, it can be trivially verified that $\xi_1 = \xi_3 = 0$ and $\xi_2 = \gamma_{zw}^M$. This implies that there are no internal aliasing errors in the MTF algorithm and the only errors arise due to the input and the output disturbances. The components related to $K_{w_0}^\circ, K_{v_w}$ and K_{v_z}, quantifying the aliasing effects of the sampling of the system output, are as follows for $K_\varphi = 0$:

$$\xi_4 = \xi_5 = \gamma_F^M \left(\gamma_{zu}\gamma_{yw} + \gamma_{zw}\right), \quad \xi_6 = \gamma_F^M. \tag{4.17}$$

This implies that there is a steady-state mismatch between the measured system response and the computed model response due to aliasing effects during the sampling of the system output. The system output possibly contains harmonics larger than M due to the reconstruction of the input signal w_0 and due to the disturbances. For large low-pass filter orders o_z, the constant γ_F^M becomes zero, resulting in the coefficients $\xi_4 = \xi_5 = \xi_6 = 0$. In such a case, there are no aliasing occurs and the only mismatch is due to the disturbances.

4.4.3 Reduction of the mismatch

We present various methods to reduce the bounds in Theorem 4.3 and Theorem 4.4. We demonstrate numerically in the next section that a reduction of the bounds typically also reduces the mismatch.

Increase of MTF parameter η. The integer η controls the time grid used by the MTF Algorithm 4.1. It has been shown in Theorem 4.3 that in the disturbance free case, i.e., $\Delta_v = \Delta_z = 0$, there is no mismatch for $\eta \to \infty$. Therefore, it is expected that an increase in η results in a decrease of the bound in Theorem 4.3. We also expect that the bound in Theorem 4.4 decreases for an increase in η, because the mismatch $\|w_0 - w_0^{\eta M}\|_{L_2}$ becomes smaller and, as shown in (4.16), the coefficient ξ_3 drops to zero for $\eta \to \infty$. A negative consequence of increasing η is that the computational load increases. Thus, the parameter η enables a tradeoff between simulation accuracy and computational load.

Decrease of sampling/refreshing time. Decreasing t_s leads, on the one hand, to capture more harmonics during sampling of the system output, i.e., a larger M, and, on the other hand, to more accurate model simulation, i.e., a larger ηM. In the bound in Theorem 4.4, this translates to a reduction of the constant γ_F^M in (4.8b), implying less aliasing errors due to sampling of the system output. Additionally,

increasing M has the same effect as increasing η, therefore, the arguments of **Increase of MTF parameter** η also apply here.

Decrease of reconstruction errors. Increasing the low-pass filter order o_w or decreasing its cut-off frequency ω_w in Assumption 4.2 results in a decrease of the mismatch $\|w_0 - w_0^{\eta M}\|_{L_2}$. This results subsequently in a reduction of the bounds in both Theorems 4.3 and 4.4. Low-pass filtering of the input signal, prior to feeding it to the system is a well-known practice, also used in [236].

Decrease of output aliasing errors. Aliasing errors are attenuated by the low-pass filtering of the system output. Increasing the low-pass order o_z or decreasing its cut-off frequency ω_z in Assumption 4.3 results in a decrease of γ_F^M in (4.8b) such that $\lim_{o_z \to \infty} \gamma_F^M = 0$. In turn, the coefficients ξ_4, ξ_5, ξ_6 in Theorem 4.4 all reduce for increasing o_z.

Decrease of disturbances. The last method is to attenuate the input disturbance v_w and the output disturbance v_z. For example, one can improve the bandwidth of the actuator or perform a more accurate sensor calibration to reduce these. Moreover, a reduction of the Lipschitz constants of these disturbances, i.e., K_{v_w} and K_{v_z} in Assumption 4.5, results in reduced aliasing errors while sampling the system output.

4.5 Illustrative examples

We consider a mechanical system for which the bounds in Theorems 4.3 and 4.4 are computed in several scenarios. Furthermore, the methods outlined in Section 4.4.3 are applied to reduce the bounds and, evidently, also the mismatch. Finally, the special case of LTI systems is covered.

4.5.1 Nonlinear mass-spring-damper system

We consider a mass-spring-damper system with a nonlinear spring element. The linear dynamics in the form of (4.3) are represented by the following system matrices:

$$\begin{aligned} &A = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}, B = \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}, C = F = \begin{bmatrix} 0 & 1 \end{bmatrix}, \\ &D = G = H = 0, L = \rho B, \end{aligned} \tag{4.18}$$

where $m = 0.1$ kg is the mass, $k = 1000$ N/m is the spring stiffness, $b = 2.5$ N/(msec) is the damper constant, and $\rho = 0.1$ is the actuator gain. The nonlinearity $\varphi(y) = k_2 \text{sign}(y) \max(0, |y| - \delta), \forall\, y \in \mathbb{R}$, is a nonlinear spring with a deadzone characteristic with length $\delta = 10^{-4}$ and slope $k_2 = 500$. We perform a loop transformation such that the transformed nonlinearity, transformed Amatrix, and transformed F matrix, read as $\tilde{\varphi}(y) = \varphi(y) - \frac{k_2}{2} y$, $\tilde{A} \coloneqq A - \frac{k_2}{2} BC$,

Table 4.1. Summary of the bounds in Theorem 4.3 and 4.4, the asymptotic analysis in Section 4.4.2 and the special case of LTI systems in Section 4.4.2. Each column represents a different component of these bounds.

Mismatch	**Input reconstruction**	**Disturbances**	**Aliasing in MTF Algorithm 4.1**	**Aliasing due to output sampling**
$\left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2}$	$\kappa_1 \left\|w_0 - w_0^{\eta M}\right\|_{L_2}$	$\kappa_2 \Delta_w + \Delta_z$	$\kappa_3 K_{w_0}^{\eta M}$	–
$\lim_{\eta\to\infty} \left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2}$ $\lim_{M\to\infty} \left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2}$	0	$\kappa_2 \Delta_w + \Delta_z$	0	–
$\left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2}$ (LTI)	$\bar{\gamma}_{zw}^{\eta M} \left\|w_0 - w_0^{\eta M}\right\|_{L_2}$	$\gamma_{zw} \Delta_w + \Delta_z$	0	–
$\left\|\breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M\right\|_{L_2^M}$	$\xi_1 \left\|w_0 - w_0^{\eta M}\right\|_{L_2}$	$\xi_2 \Delta_w + \Delta_z$	$\xi_3 K_{w_0}^{\eta M}$	$\xi_4 K_{w_0}^{\circ} + \xi_5 K_{v_w} + \xi_6 K_{v_z}$
$\lim_{M\to\infty} \left\|\breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M\right\|_{L_2^M}$	0	$\xi_2 \Delta_w + \Delta_z$	0	0
$\left\|\breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M\right\|_{L_2^M}$ (LTI)	0	$\gamma_{zw}^M \Delta_w + \Delta_z$	0	$\xi_4 K_{w_0}^{\circ} + \xi_5 K_{v_w} + \xi_6 K_{v_z}$

$\tilde{F} \coloneqq F - \frac{k_2}{2}GC$, respectively. The transformed model satisfies all conditions of Theorem 4.1 for $K_\varphi = \frac{k_2}{2}$.

4.5.2 Case study settings

The system is excited by a random phase multisine with a period time $T = 128$ seconds that contains frequencies up to 6.25 Hz. The excitation signal is realized by a ZOH reconstruction filter, running at a refreshing frequency $f_s = 32$ Hz, and is subsequently low-pass filtered according to Assumption 4.2. The low-pass filter order is varied in different scenarios. The constants $K^{\circ}_{w_0}$ and $K^{\eta M}_{w_0}$ are found numerically. The filtered system output $\breve{\breve{z}}_F$ is measured according to Assumption 4.3. Furthermore, both the input and the output disturbances are active. Their Lipschitz constants K_{v_w} and K_{v_z} are computed analytically.

All steady-state model responses are computed by the MTF Algorithm 4.1. The MTF algorithm considers ηM harmonics, in which the parameter $M = 2^{11}$ depends on the sampling frequency f_s via $M = Tf_s/2$, and η is varied. To facilitate the computation of the mismatch between the system and the model responses, the system response is also computed by the MTF algorithm considering M_s harmonics, where $M_s = 2^{19}$. Such large M_s guarantees that the system response is computed with high accuracy, see the analysis in Section 4.4.2. We assume that the computed system response contains no errors with respect to the system response in the continuous-time world.

From here onwards, the steady-state mismatch between the system and the model response in Problem 4.1 is called the *actual* mismatch. Furthermore, the steady-state mismatch between the measured system response and the model response in Problem 4.2 is called the *measured* mismatch.

4.5.3 Numerical results

As a starting point, we select $o_w = 2$ and $o_z = 0$ for the low-pass filter orders in the input and output sides, respectively. Figures 4.2a and 4.3a depicts the change of the bounds, the component of the bounds, and the mismatch as a function of the MTF-parameter η.

We start with the *actual* mismatch for which the results are presented in Figure 4.2a. The component (—✖—), accounting for the errors caused by the harmonics beyond ηM in the input signal, shows a drop over η, which is expected given the analysis in Section 4.4. The components (—○—) and (—□—), related to the disturbances v_w and v_z, are η-independent as these play no role in the MTF algorithm. The component (—▼—), related to internal aliasing errors within the MTF algorithm, drops over η. However, this drop is at a slow rate, which is due to the slow decrease of the constant $\gamma_a^{\eta M}$, i.e., the slow convergence of the Riemann-zeta function. This component dominates the total upper bound (- ◈ -) due to several approximation steps used in the proofs of Lemma 4.1 and Theorem 4.3. The com-

puted mismatch (—+—) is (almost) η-independent, evidencing that it is dominated by the in- and the output disturbances v_w and v_z.

The same conclusions for the components (—×—), (—○—), (—□—), and (—▼—) can be drawn for the *measured* mismatch, for which the results are presented in Figure 4.3a. The additional components (—○—) and (—□—) account for the aliasing errors caused by the disturbances v_w and v_z while sampling of the system output. These components are η-independent as expected from the analysis in Section 4.4. From Figure 4.3a, it is clear that the bound (-◇-) on the *measured* mismatch is dominated by the component (—▼—). The computed mismatch (—+—) is η-independent, implying that it is dominated by the in- and the output disturbances v_w and v_z or the aliasing errors (—▲—) due to the sampling of the system output.

Reduction of disturbances

As a first step, we attenuate the disturbances v_w and v_z. The results are presented in Figures 4.2b and 4.3b.

The results for the *actual* mismatch are depicted in Figure 4.2b. The components (—×—) and (—▼—) are not reduced with respect to the start situation, since these are independent of the disturbances. The disturbance-related components (—○—) and (—□—) are naturally reduced. The bound (-◇-) on the *actual* mismatch is still dominated by (—▼—), which bounds the internal aliasing errors by the MTF algorithm. The computed mismatch (—+—) is significantly reduced and it is now η-dependent, implying that it is either dominated by (—×—) or by (—▼—).

Next, we turn to the *measured* mismatch, for which the results are depicted in Figure 4.3b. The same conclusions as in the *actual* mismatch case can be drawn for the components (—×—), (—○—), (—□—), and (—▼—). The component (—▲—) remains unchanged with respect to the starting situation as it is independent of the disturbances. The disturbance-related components (—○—) and (—□—), accounting for the aliasing errors due to the output sampling, are reduced thanks to their Lipschitz constants also being reduced. The bound (-◇-) on the *measured* mismatch is still dominated by (—▲—), which bounds the aliasing errors due to output sampling. The computed error (—+—) is also reduced, however, still η-independent, implying that it is dominated by the in- and output disturbances v_w and v_z or the aliasing errors (—▲—) due to the sampling of the system output.

Increase of low-pass filter order

Next, we increase the order o_w of the low-pass filter at the input side from 2 to 4. Furthermore, we increase the order of o_z of the low-pass filter at the output side from 0 to 8.

Let us first focus on Figure 4.2c concerning the *actual* mismatch. These results are independent of the increase of o_z, as this parameter does not affect the *actual* steady-state response of the system. In line with the analysis in Section 4.4, an increase of o_w results in the component (—×—) dropping with a larger rate over

η. Furthermore, the component (—▼—), related to aliasing errors within the MTF algorithm, drops slightly thanks to the increase of o_w, also resulting in a slight reduction of the bound (- ◇ -). The computed error (—+—) also reduces, however, not significantly, implying that aliasing errors within the MTF algorithm dominate the error. Note that the disturbance-related components (—○—) and (—□—) remain unaffected by a change of o_w (or o_z), as expected from the analysis in Section 4.4.

Now we turn to Figure 4.3c concerning the *measured* mismatch. The same conclusions as drawn in the *actual* case hold for the components (—×—), (—○—), (—□—), and (—▼—). The component (—▲—), related to aliasing errors due to the sampling of the system output, is reduced significantly as an increase of the low-pass filter order o_z results in more suppression of high-frequency contents prior to sampling. The increase of o_z also results in less aliasing errors due to the disturbances, i.e., the components (—○—) and (—□—). The total bound (- ◇ -), however, is still dominated by (—▲—). The computed mismatch (—+—) drops significantly compared to the previous situation, especially for larger η. In particular, it also becomes η-dependent for small η. However, for large η a constant value is reached, evidencing that the error is dominated by aliasing errors (—▲—) due to sampling of the system output.

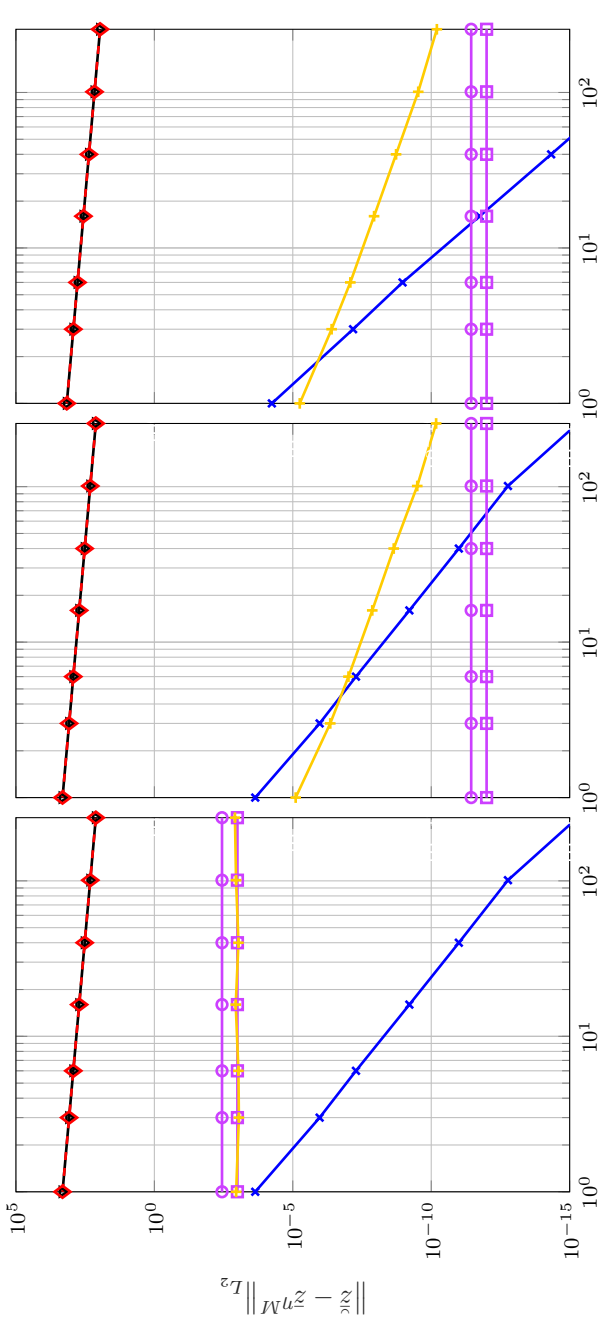
To decrease the *actual* mismatch even further, a larger η could be taken, resulting in a decrease of (—▼—), which is in line with the asymptotic analysis in Section 4.4.2. A larger η results in the more accurate computation of model responses, however, at the expense of computation time. To decrease the *measured* mismatch further, a larger o_w should be taken to decrease (—▲—), representing the aliasing errors due to sampling of the system output. After that, the limiting component will be (—▼—), which can be further reduced by taking a larger η.

The example shows that the bounds (4.12) and (4.14) are conservative, stemming from the conservatism of the conditions in Theorem 4.1 and several worst-case approximations in the derivations of the bounds. Nevertheless, the example also illustrates that a reduction of the bounds (- ◇ -) and its components generally also reduces the mismatch (—+—). This insight, together with the qualitative analysis presented in Section 4.4, gives both a theoretical foundation and intuitive insights for the quantification and reduction of these types of mismatches.

4.5.4 Linear mass-spring-damper system

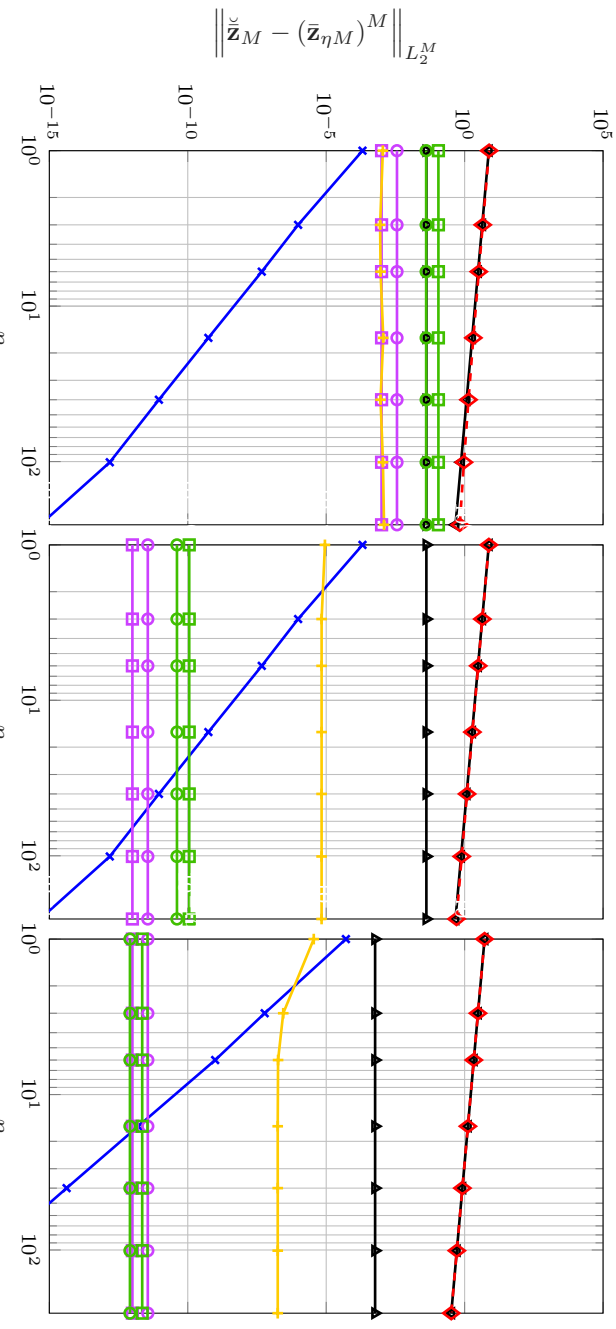
In this example, the parameter k_2 of the nonlinearity is set to 0, such that the considered system is LTI. The conditions of Theorem 4.1 are trivially satisfied and $K_\varphi = 0$ is taken for the computation of the bounds in Theorems 4.3 and 4.4. We present several studies that highlight the features of the derived bounds for LTI systems. In all cases, we consider the steady-state response computed by the MTF algorithm with $M_s = 2^{19}$ harmonics as the 'true' system response $\overset{\approx}{z}$.

We consider the 7 studies that are included in Table 4.2. In all these studies, the sampling frequency f_s is fixed to 32 Hz and the MTF parameter η is fixed to



(a) Start situation. **(b)** Reduction of disturbances. **(c)** Increase of o_w and o_z.

Figure 4.2. Mismatch between the *actual* system response and computed model response. The bound (4.12) (-◇-) is decomposed into $\kappa_1 \left\| w_0 - w_0^{\eta^M} \right\|$ (—×—), $\kappa_2 \Delta_w$ (—○—), Δ_z (—□—) and $\kappa_3 K_{w_0^{\eta^M}}$ (—▸—). The computed error (——) is also depicted.



(a) Start situation. **(b)** Reduction of disturbances. **(c)** Increase of o_w and o_z.

Figure 4.3. Mismatch between the *actual* system response and computed model response. The bound (4.12) (-◇-) is decomposed into $\kappa_1 \left\| w_0 - w_0^{\eta M} \right\|$ (-×-), $\kappa_2 \Delta_w$ (-○-), Δ_z (-□-) and $\kappa_3 K_{w_0^{\eta M}}$ (-▼-). The computed error (-+-) is also depicted.

1. The first three studies, i.e., **A, B** and **C**, consider a disturbance-free situation, i.e., $\Delta_w = \Delta_z = K_{v_w} = K_{v_z} = 0$, for which the bounds in Theorems 4.3 and 4.4 (for $K_\varphi = 0$) reduce to:

$$\left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2} \leq \bar{\gamma}_{zw}^{\eta M} \left\|w_0 - w_0^{\eta M}\right\|_{L_2}, \tag{4.19a}$$

$$\left\|\breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M\right\|_{L_2^M} \leq \gamma_F^M(\gamma_{zu}\gamma_{yw} + \gamma_{zw})K_{w_0}^\circ. \tag{4.19b}$$

The latter four studies, i.e., **D, E, F** and **G**, consider an input-free situation, i.e., $w_0(t) = 0, t \in \mathbb{R}$. The bounds in Theorems 4.3 and 4.4 (for $K_\varphi = 0$) in the input-free situation reduce to:

$$\left\|\breve{\bar{z}} - \bar{z}^{\eta M}\right\|_{L_2} \leq \gamma_{zw}\Delta_w + \Delta_z, \tag{4.20a}$$

$$\left\|\breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M\right\|_{L_2^M} \leq \gamma_{zw}^M\Delta_w + \Delta_z + \gamma_F^M(\gamma_{zu}\gamma_{yw} + \gamma_{zw})K_{v_w} + \gamma_F^M K_{v_z}. \tag{4.20b}$$

Figure 4.4 presents the results of Studies **A, B** and **C**. The order o_w of the low-pass filter at the input side is increased from Study **A** to Study **B**. As a consequence, the *actual* mismatch (—▲—) is relatively large in Study **A** compared to Studies **B** and **C** where o_w is increased. An increase in o_w results in a decrease in $\left\|w_0 - w_0^{\eta M}\right\|_{L_2}$, which is the only component in (4.19a), thus yielding a smaller *actual* mismatch. Going from Study **B** to **C**, only the order o_z of the low-pass filter at the output side is changed, which has no consequence for the *actual* mismatch. The bound (—□—) for the *actual* mismatch shows the same trend and shows to be slightly conservative.

The *measured* mismatch (—o—) and its bound (—×—) also drop when increasing o_w from Study **A** to **B**. This is thanks to the constant $K_{w_0}^\circ$ becoming smaller in (4.19b), resulting in fewer aliasing errors due to output sampling. Furthermore, increasing the order o_z of the low-pass filter at the output side from Study **B** to Study **C** results in the *measured* mismatch (—o—) and its bound (—×—) becoming significantly smaller. This is thanks to the suppression of frequency components beyond the first M harmonic in the system output by the low-pass filter before sampling the system output.

Studies **D** and **E** consider an input free situation, thus the order o_w is irrelevant. These studies consider the same sine wave disturbance v_w and v_z with a frequency below the Nyquist frequency $f_{max} = f_s/2 = 16$ Hz. Since only the order o_z changes from Study **D** to **E**, the *actual* mismatch (—▲—) and its bound (—□—) remain unchanged because these are independent of o_z. In these studies, no aliasing occurs during sampling of the system output, since the system output only contains the same frequency as the disturbances below the Nyquist frequency. Therefore, increasing the order o_z does not have any effect on the *measured* mismatch (—o—). However, its bound (—×—) in (4.20b) decreases, since the constant γ_F^M in (4.8b) becomes smaller for larger o_z.

Table 4.2. Studies **A**, ..., **G**, on the LTI system.

	Study **A**	Study **B**	Study **C**	Study **D**	Study **E**	Study **F**	Study **G**
$w_0(t)$	$w_0(t)$	$w_0(t)$	$w_0(t)$	$w(t) = 0$	$w(t) = 0$	$w(t) = 0$	$w(t) = 0$
o_w	2	4	4	-	-	-	-
o_z	0	0	8	0	8	0	8
$v_w(t) = A_w \sin(2\pi f t)$	$A_w = 0$	$A_w = 0$	$A_w = 0$	$f < f_{\max}$	$f < f_{\max}$	$f > f_{\max}$	$f > f_{\max}$
$v_z(t) = A_z \sin(2\pi f t)$	$A_z = 0$	$A_z = 0$	$A_z = 0$	$f < f_{\max}$	$f < f_{\max}$	$f > f_{\max}$	$f > f_{\max}$

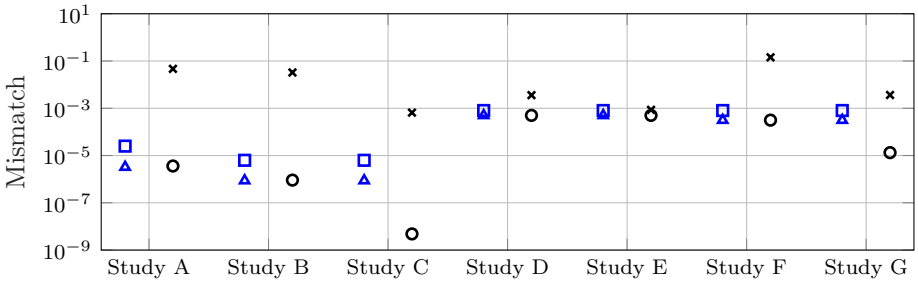


Figure 4.4. Results of Studies **A** - **G** in Table 4.2. The *actual* mismatch represented by (—▲—) together with its upper bound (4.12) represented by (—□—). The *measured* mismatch represented by (—○—) together with its upper bound (4.14) represented by (—×—).

Studies **F** and **G** are analog to Studies **D** and **E**, except that the disturbances now have a frequency above the Nyquist frequency. For the same reasons as before, nothing changes in the *actual* mismatch (—▲—) and its bound (—□—) when increasing o_z from Study **F** to **G**. However, since now aliasing due to output sampling occurs, it can be observed that the *measured* mismatch (—○—) is now reduced when the order o_z of the low-pass filter at the output side is increased. This is thanks to the low-pass filter attenuating the frequency components of the system output beyond the first M harmonics prior to sampling. Similarly, also the bound (—×—) in (4.20b) reduces since the constant γ_F^M in (4.8b) reduces by increasing o_z. The Lipschitz constants K_{v_w} and K_{v_z} are increased for Studies **F** and **G** with respect to Studies **D** and **E** due to the frequency of the disturbances being increased, causing faster variations of the disturbance. Consequently, the bound (—×—) is larger in Studies **F** and **G** compared to **D** and **E** and, therefore, is more conservative.

4.6 Conclusions

Model response computation is required for a variety of systems and control problems. However, the responses computed by numerical simulation algorithms are typically different from the real-life system responses. This chapter has presented for the class of convergent Lur'e-type systems bounds on the steady-state mismatch between the simulated model response computed by means of the MTF simulation algorithm and (i) the *actual* system response; and (ii) the *sampled measured* system response. The bounds can be decomposed into several components, namely, a component that upper bounds the effect of (i) missed higher harmonics of the input signal in the MTF algorithm; (ii) aliasing errors within the MTF algorithm; (iii) aliasing errors due to sampling of the system output; and (iv) a mismatch due to in- and output disturbances being active on the system, but being unavailable

for model simulation. Although conservative, the bounds derived in this chapter provide essential insights that can be used to reduce the mismatch. In a numerical example, we show that reducing the derived bounds typically also results in a reduced mismatch, making the reduction methods presented in this chapter practically valuable. The results in this chapter facilitate the decision-making process regarding simulation parameters and hardware selection, thereby enabling accurate computation of the model responses required for many engineering applications.

II

Moment matching for linear and nonlinear models

5

Time-domain moment matching for multiple-input multiple-output linear time-invariant models

Model reduction by moment matching for linear time-invariant (LTI) models is a reduction technique that has a clear interpretation in the Laplace domain. In particular, for the multiple-input multiple-output (MIMO) LTI case, Krylov subspace methods aim at matching the transfer-function matrix (and possibly its derivatives) of the reduced-order model to the transfer-function matrix of the full-order model along so-called tangential directions at desired interpolation points. A straightforward application of time-domain *moment matching to* MIMO *LTI models does not result in such a match in the transfer-function matrix. In this chapter, we derive a relation between the MIMO transfer-function matrices of the full- and the reduced-order models that follows from the application of time-domain moment matching on MIMO LTI models. This is subsequently exploited to formulate conditions on the parameters of time-domain moment matching under which the transfer-function matrix is matched along tangential directions, thus ensuring consistency with classical Krylov subspace methods.*

The contents of this chapter are published in: Shakib, M. F., Scarciotti, G., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Time-domain moment matching for multiple-input multiple-output linear time-invariant models. *Provisionally accepted for publication in Automatica.*

5.1 Introduction

Moment matching (MM) is a powerful and computationally efficient method to construct reduced-order dynamical models for large-scale systems. Originally, MM techniques were based on Krylov subspaces, which allow for efficient computation of reduced-order models [10, 11, 100]. These methods have the property that in the single-input single-output (SISO) case, the transfer functions of the full- and of the reduced-order model are matched at desired interpolation points in the complex plane. In the multiple-input multiple-output (MIMO) case, the transfer-function matrices are matched along so-called left and/or right tangential directions at desired interpolation points in the complex plane [20, 92, 100].

In [91], the notion of moments was connected to the solution of the so-called Sylvester equation, which gave rise to a novel MM method in a *time-domain* framework [16, 17]. The advantages of time-domain moment matching are twofold, namely: it facilitates parametric freedom, which can be conveniently exploited to enforce properties such as stability and passivity on the reduced-order model [16]; and it has a natural extension to model order reduction for nonlinear models [16, 225, 248].

For SISO models, this *time-domain* framework preserves the property of the classical Krylov methods [10, 11, 100] that the transfer functions of the full- and of the reduced-order model are matched at desired interpolation points in the complex plane [16, 17]. For MIMO models, classical Krylov subspace methods ensure that the transfer-function matrices are matched along the so-called left and/or right tangential directions at desired interpolation points in the complex plane [20, 92, 100]. However, the *time-domain* MM framework does not generally preserve this property in the MIMO case.

In this chapter, we derive explicitly a matching property in terms of the MIMO transfer-function matrices of the full- and the reduced-order model for *time-domain* MM. Subsequently, the derived matching property is exploited to formulate additional conditions on the parameters of time-domain MM under which a match along the tangential directions of the transfer-function matrices is indeed ensured. Finally, a numerical example is provided that illustrates the obtained matching property. The results in this chapter can be used to connect time-domain MM for MIMO models to MM based on Krylov subspace methods by specific parameter selection for time-domain MM.

To summarize, the main contributions of this chapter are as follows.

- We derive a relation between the MIMO transfer-function matrices of the full- and the reduced-order models that follows from the application of time-domain MM on MIMO LTI models.

- We show under which conditions time-domain MM ensures a match of the MIMO transfer-function matrices along the tangential directions, consistent with classical Krylov subspace methods.

The remainder of this chapter is outlined as follows. In Section 2, we present the main results of this chapter, namely a matching property of time-domain MM together with additional conditions to ensure a match that is consistent with classical Krylov subspace methods. Section 3 applies the derived results to a numerical example. Section 4 closes with concluding remarks.

Notation: The set of real numbers is denoted by $\mathbb{R}$ and the set of complex numbers is denoted by $\mathbb{C}$. The empty set is denoted by $\emptyset$. The complex conjugate of a complex number $s \in \mathbb{C}$ is denoted by $\bar{s}$. The $n \times n$ identity matrix is denoted by I_n, the $n \times n$ zero matrix is denoted by 0_n, and the $n \times m$ zero matrix is denoted by $0_{n\times m}$. The eigenvalues of the matrix $A \in \mathbb{C}^{n\times n}$ are denoted by $\sigma(A)$. The transpose (not conjugate transpose) of $A \in \mathbb{C}^{n\times n}$ is denoted by $A^\top$. The imaginary unit $\sqrt{-1}$ is denoted by j.

5.2 Moment matching for MIMO models

The moment matching problem for MIMO LTI models is defined in Section 5.2.1, in which moments are defined through the transfer-function matrix of the LTI model. After that, Section 5.2.2 characterizes moments in the *time-domain* framework. Subsequently, consistent with methods for SISO models, a family of MIMO LTI models is constructed that achieves a generalized matching property. Section 5.2.4 proposes specific selections of *time-domain* MM parameters to ensure a match that is consistent with classical Krylov methods. Finally, Section 5.2.5 presents families of models that achieve zeroth-order moment matching. For brevity, we focus on moments along the right tangential directions and omit the dual of moments along the left tangential directions. The results of this chapter can be readily extended to the case of matching moments along the left tangential directions as well.

5.2.1 The MIMO moment matching problem

Consider the MIMO LTI model

$$\Sigma : \quad \dot{x} = Ax + Bu, \quad y = Cx, \tag{5.1}$$

where $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, y(t) \in \mathbb{R}^p, A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{p\times n}$. The SISO case is retrieved for $m = p = 1$. The transfer-function matrix of this model is

$$K(s) = C(sI - A)^{-1}B, \quad K : \mathbb{C} \to \mathbb{C}^{p\times m}. \tag{5.2}$$

Throughout the rest of this chapter, we assume minimality of model (5.1). The moments of model (5.1) are defined according to [179, Definition 1], which is a MIMO extension of [16, Definition 1].

Definition 5.1. *The k-moment of* (5.1) *at* $s \in \mathbb{C} \setminus \sigma(A)$*, along tangential direction* $\ell \in \mathbb{C}^m$ *is:*

$$\eta_k^{\Sigma}(s,\ell) := \frac{(-1)^k}{k!} \frac{\mathrm{d}^k K(s)}{\mathrm{d}s^k} \ell \in \mathbb{C}^p, \quad k \geq 0, \text{ integer.} \tag{5.3}$$

■

Now consider the MIMO LTI model

$$\hat{\Sigma}: \quad \dot{\xi} = F\xi + Gu, \quad \psi = H\xi, \tag{5.4}$$

where $\xi(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, \psi(t) \in \mathbb{R}^p, F \in \mathbb{R}^{\nu \times \nu}, G \in \mathbb{R}^{\nu \times m}$ and $H \in \mathbb{R}^{p \times \nu}$. The relation between models (5.4) and (5.1) is defined as follows.

Definition 5.2. *Consider the interpolation point* $s \in \mathbb{C} \setminus \sigma(A)$*, the tangential direction* $\ell \in \mathbb{C}^m$*, and an integer* $k \geq 0$*. Model* (5.4) *is called a model of* (5.1) *at* (s, ℓ) *if it achieves moment matching according to*

$$\eta_{\tilde{k}}^{\Sigma}(s,\ell) = \eta_{\tilde{k}}^{\hat{\Sigma}}(s,\ell), \quad \tilde{k} = 0, \ldots, k. \tag{5.5}$$

In addition, if $\nu < n$*, then model* (5.4) *is called a reduced-order model of* (5.1) *at* (s, ℓ)*.* ■

Such matching is analog to tangential Hermite interpolation [10] and is consistent with, e.g., [101].

The problem addressed in this chapter is to find a parametrization for F, G, and H, such that $\hat{\Sigma}$ in (5.4) is a (reduced-order) model of (5.1), according to Definition 5.2. This problem includes the case in which $\nu \geq n$.

5.2.2 Time-domain characterization of moments

The definition of moments in Definition 5.1 relies on the transfer-function matrix. Alternatively, in *time-domain* moment matching, moments are characterized through the solution $\Pi \in \mathbb{C}^{n \times \nu}$ of a Sylvester equation, namely

$$A\Pi + BL = \Pi S, \tag{5.6}$$

where the matrix $L \in \mathbb{C}^{m \times \nu}$ contains information about the tangential directions and the matrix $S \in \mathbb{C}^{\nu \times \nu}$ contains information about the interpolation points. The matrices S and L play a central role in *time-domain* moment matching and are further specified in the remainder of this chapter. Note that under the condition $\sigma(S) \cap \sigma(A) = \emptyset$, the solution $\Pi \in \mathbb{C}^{n \times \nu}$ is unique, see [59].

We distinguish between two cases. The first case deals with repeated *single* complex interpolation points, being the eigenvalues of the matrix S in the *complex-valued Jordan form*. This results in a reduced-order model that consists of complex-valued model matrices. The second case deals with a repeated *pair* of complex conjugated interpolation points, being the eigenvalues of the matrix S in

the *real-valued Jordan form*. This case is the most useful case since it results in reduced-order models that are described by *real-valued* model matrices. Hereafter, we call these cases the *complex-valued Jordan case* and the real-valued Jordan case, respectively. In Remark 5.2, we explain how the general case without any Jordan structure can be treated.

The results for the complex-valued Jordan case are summarized in the following lemma. This lemma is an extension of Lemma 2 in [16] from the SISO case to the MIMO case.

Lemma 5.1. *Consider model (1) with m inputs and p outputs, the interpolation point $s^\star \in \mathbb{C}$, the integer $k \geq 0$ and $\ell_0, \ldots, \ell_k \in \mathbb{C}^m$. Define $\nu \coloneqq k+1$. Suppose $s^\star \notin \sigma(A)$. Then*

$$(C\tilde{\Pi})^\top = \begin{bmatrix} \eta_0(s^\star, \ell_0)^\top \\ \eta_0(s^\star, \ell_1)^\top - \eta_1(s^\star, \ell_0)^\top \\ \vdots \\ \sum_{i=0}^{k-1} (-1)^i \eta_i(s^\star, \ell_{k-1-i})^\top \\ \sum_{i=0}^{k} (-1)^i \eta_i(s^\star, \ell_{k-i})^\top \end{bmatrix} \tag{5.7}$$

where $\eta_i(s^\star, \cdot), i = 0, \ldots, k$, are defined in (5.3) *and where $\tilde{\Pi} \in \mathbb{C}^{n \times \nu}$ is the unique solution of the Sylvester equation*

$$A\tilde{\Pi} + B\tilde{L} = \tilde{\Pi}\Phi_k(s^\star) \tag{5.8}$$

with

$$\begin{aligned} \tilde{L} &\coloneqq \begin{bmatrix} \ell_0, \ell_1, \ldots, \ell_k \end{bmatrix} \in \mathbb{C}^{m \times \nu} \\ \Phi_k(s^\star) &\coloneqq \begin{bmatrix} s^\star & 1 & 0 & \cdots & 0 \\ 0 & s^\star & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & s^\star & 1 \\ 0 & \cdots & \cdots & 0 & s^\star \end{bmatrix} \in \mathbb{C}^{\nu \times \nu}. \end{aligned} \tag{5.9}$$

▲

Proof. The proof can be found in Appendix D.1. □

The results for the real-valued Jordan case are summarized in the next lemma.

Lemma 5.2. *Consider model (1) with m inputs and p outputs, the interpolation point $s^\star = a^\star + jb^\star \in \mathbb{C}$ with $a^\star, b^\star \in \mathbb{R}, b^\star \neq 0$, the integer $k \geq 0$ and*

$\ell_0, \ldots, \ell_{2k+1} \in \mathbb{R}^m$. *Define* $\nu \coloneqq 2(k+1)$. *Suppose* $s^\star \notin \sigma(A)$. *Then,*

$$(C\tilde{\Pi})^\top = \begin{bmatrix} Re(\eta_0(s^\star, \ell_0 + j\ell_1)^\top) \\ Im(\eta_0(s^\star, \ell_0 + j\ell_1)^\top) \\ \sum_{i=0}^{1}(-1)^i Re(\eta_i(s^\star, \ell_{2-2i} + j\ell_{3-2i})^\top) \\ \sum_{i=0}^{1}(-1)^i Im(\eta_i(s^\star, \ell_{2-2i} + j\ell_{3-2i})^\top) \\ \vdots \\ \sum_{i=0}^{k}(-1)^i Re(\eta_i(s^\star, \ell_{2k-2i} + j\ell_{2k-2i+1})^\top) \\ \sum_{i=0}^{k}(-1)^i Im(\eta_i(s^\star, \ell_{2k-2i} + j\ell_{2k-2i+1})^\top) \end{bmatrix}, \tag{5.10}$$

where $\eta_i(s^\star, \cdot), i = 0, \ldots, k$, *are defined in* (5.3) *and where* $\tilde{\Pi} \in \mathbb{R}^{n\times\nu}$ *is the unique solution of the Sylvester equation*

$$A\tilde{\Pi} + B\tilde{L} = \tilde{\Pi}\Psi_k(s^\star) \tag{5.11}$$

with

$$\begin{aligned} \tilde{L} &\coloneqq \begin{bmatrix} \ell_0, \ell_1, \ldots, \ell_{2k+1} \end{bmatrix} \in \mathbb{R}^{m\times\nu} \\ \Psi_k(s^\star) &\coloneqq \begin{bmatrix} \Xi^\star & I_2 & \cdots & 0_2 & 0_2 \\ 0_2 & \Xi^\star & I_2 & \cdots & 0_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0_2 & \cdots & 0_2 & \Xi^\star & I_2 \\ 0_2 & 0_2 & \cdots & 0_2 & \Xi^\star \end{bmatrix} \in \mathbb{R}^{\nu\times\nu}, \\ \Xi^\star &\coloneqq \begin{bmatrix} a^\star & b^\star \\ -b^\star & a^\star \end{bmatrix} \in \mathbb{R}^{2\times 2}. \end{aligned} \tag{5.12}$$

▲

Proof. The proof can be found in Appendix D.2. □

Lemma 5.1 and Lemma 5.2 establish a connection between the elements of $C\tilde{\Pi}$ and the moments in Definition 5.1. Both lemmas carry the same information in the corresponding variable $\tilde{\Pi}$. However, in Lemma 5.2, this information is split into a real and imaginary part, hence, yielding a twice as large number of columns in $\tilde{\Pi} \in \mathbb{R}^{n\times 2(k+1)}$ compared to the complex-valued $\tilde{\Pi} \in \mathbb{C}^{n\times k+1}$ in Lemma 5.1.

Remark 5.1. Despite $\tilde{L}$ being a real-valued matrix in Lemma 5.2, (5.10) shows that the tangential directions are generally complex-valued. △

5.2.3 Generalized family of reduced-order models

In the same fashion as in [16], this section exploits the connection between $C\Pi$ and the moments in Lemmas 5.1 and 5.2 to construct families of reduced-order models. These models, however, achieve a generalized matching property, which

is exploited in Section 5.2.4 to formulate families of models that achieve moment matching in accordance with Definition 5.2. The result for the family of models in the complex-valued Jordan case is summarized in the following theorem.

Theorem 5.1. *Consider the model* (5.1) *with* m *inputs and* p *outputs, the integer* $k \geq 0$*, the constant* $\nu \coloneqq k+1$*, the non-derogatory matrix*[1] $S \in \mathbb{C}^{\nu\times\nu}$ *with* $det(sI - S) = (s - s^\star)^{k+1}$*, where* $s^\star \in \mathbb{C}$ *is the interpolation point, and any matrix* $L \in \mathbb{C}^{m\times\nu}$*. Suppose* $s^\star \notin \sigma(A)$*. Then, for any* $G \in \mathbb{C}^{\nu\times m}$ *such that* $s^\star \notin \sigma(S - GL)$*, model* (5.4) *with* $F \coloneqq S - GL, H \coloneqq C\Pi$*, and with* $\Pi \in \mathbb{C}^{n\times\nu}$ *the unique solution to the Sylvester equation*

$$A\Pi + BL = \Pi S, \tag{5.13}$$

achieves the following match:

$$\sum_{i=0}^{\tilde{k}} (-1)^i \eta_i^{\Sigma}(s^\star, \ell_{\tilde{k}-i}) = \sum_{i=0}^{\tilde{k}} (-1)^i \eta_i^{\hat{\Sigma}}(s^\star, \ell_{\tilde{k}-i}), \tag{5.14}$$

for $\tilde{k} = 0, \dots, k$*, and where* $\ell_i \in \mathbb{C}^m, i = 0, \dots k$*, are the columns of the matrix* LT *with any* $T \in \mathbb{C}^{\nu\times\nu}$ *that is obtained via the Jordan decomposition* $S = T\Phi_k(s^\star)T^{-1}$ *with* $\Phi_k(s^\star) \in \mathbb{C}^{\nu\times\nu}$ *as given in* (5.9). ▲

Proof. Consider the class of models (5.4). Given that $s^\star \notin \sigma(F)$, the Sylvester equation $FP + GL = PS$ has a unique solution $P \in \mathbb{C}^{\nu\times\nu}$. Selecting $P = I_\nu$ results in $F = S - GL$ for any $G \in \mathbb{C}^{\nu\times m}$ such that $\sigma(S - GL) \cap \sigma(S) = \emptyset$. Moreover, selecting $H = C\Pi$ results in the matching $HP = C\Pi$. It is only left to show that the elements in $C\Pi$ are described by (5.14). Hereto, notice that the matrix $S \in \mathbb{C}^{\nu\times\nu}$ contains ν repeated eigenvalues at $s^\star$ and is non-derogatory, which guarantees that there exists a (non-unique) non-singular matrix $T \in \mathbb{C}^{\nu\times\nu}$ that brings $T^{-1}ST$ into the Jordan form as in (5.9), see, e.g., [302]. Therefore, application of Lemma 5.1 with $\Phi_k(s^\star) = T^{-1}ST$ and $\tilde{L} = LT$, gives the relation between the moments (5.14) and the matrix $C\tilde{\Pi}$ with $\tilde{\Pi} \in \mathbb{C}^{n\times\nu}$ the unique solution to the Sylvester equation (5.8). The solution $\Pi \in \mathbb{C}^{n\times\nu}$ of the Sylvester equation in (5.13) is related to $\tilde{\Pi}$ by $\tilde{\Pi} = \Pi T$. Since T is invertible, matching the elements of the matrix $C\Pi$ by the reduced-order model (5.4) guarantees that the linear combination of moments in (5.14) are matched, which completes the proof. □

Similarly to Theorem 5.1, we can construct a family of models for the real-valued Jordan case, which is presented in the following theorem.

Theorem 5.2. *Consider the model* (5.1) *with* m *inputs and* p *outputs, the integer* $k \geq 0$*, the constant* $\nu \coloneqq 2(k+1)$*, the non-derogatory matrix* $S \in \mathbb{R}^{\nu\times\nu}$ *with* $det(sI - S) = ((s - s^\star)(s - \bar{s}^\star))^{k+1}$*, where* $s^\star \in \mathbb{C} \setminus \mathbb{R}$ *is the interpolation point, and any matrix* $L \in \mathbb{R}^{m\times\nu}$*. Suppose* $s^\star \notin \sigma(A)$*. Then, for any* $G \in \mathbb{R}^{\nu\times m}$ *such*

[1] A matrix is non-derogatory if its characteristic and minimal polynomials coincide.

that $s^\star \notin \sigma(S-GL)$*, model* (5.4) *with* $F := S-GL, H = C\Pi$*, and with* $\Pi \in \mathbb{R}^{n\times\nu}$ *the unique solution to the Sylvester equation*

$$A\Pi + BL = \Pi S, \tag{5.15}$$

achieves the following match:

$$\begin{aligned}
\sum_{i=0}^{\tilde{k}}(-1)^i\eta_i^{\Sigma}(s^\star, \ell_{2\tilde{k}-2i} + j\ell_{2\tilde{k}-2i+1}) &= \sum_{i=0}^{\tilde{k}}(-1)^i\eta_i^{\hat{\Sigma}}(s^\star, \ell_{2\tilde{k}-2i} + j\ell_{2\tilde{k}-2i+1}),\\
\sum_{i=0}^{\tilde{k}}(-1)^i\eta_i^{\Sigma}(\bar{s}^\star, \ell_{2\tilde{k}-2i} - j\ell_{2\tilde{k}-2i+1}) &= \sum_{i=0}^{\tilde{k}}(-1)^i\eta_i^{\hat{\Sigma}}(\bar{s}^\star, \ell_{2\tilde{k}-2i} - j\ell_{2\tilde{k}-2i+1}),
\end{aligned} \tag{5.16}$$

for $\tilde{k} = 0, \ldots, k$*, and where* $\ell_i \in \mathbb{R}^m, i = 0, \ldots 2k+1$*, are the columns of the matrix* LT *with any* $T \in \mathbb{R}^{\nu\times\nu}$ *that is obtained via the Jordan decomposition* $S = T\Psi_k(s^\star)T^{-1}$ *with* $\Psi_k(s^\star) \in \mathbb{R}^{\nu\times\nu}$ *as given in* (5.12). ▲

Proof. The proof is analog to the proof of Theorem 5.1 up to the statement that the linear combination of moments at the complex conjugate $\bar{s}^\star$ of $s^\star$ is also matched along the complex conjugate of the tangential direction. This fact is guaranteed by the following property for real rational transfer functions:

$$\overline{\eta_k(s^\star, \ell)} = \eta_k(\bar{s}^\star, \bar{\ell}),$$

for $s^\star \in \mathbb{C}$ and $\ell \in \mathbb{C}^m$. The rest of the proof is omitted for brevity. □

The matrix S being non-derogatory in Theorems 5.1 and 5.2 is required to ensure the existence of a similarity transformation that brings S into the complex-Jordan form in (5.9) or the real-Jordan form in (5.12). The results of Theorems 5.1 and 5.2 are used in Section 5.2.4 to facilitate choices for the pair (S, L) such that application of Theorem 5.1 and 5.2 results in (5.4) being a (reduced-order) model of (5.1), consistent with Definition 5.2. Furthermore, in Section 5.2.5, these theorems are exploited to formulate families of models that achieve zeroth-order moment matching.

Remark 5.2. Lemmas 5.1 and 5.2 can be used as building blocks when multiple distinct interpolation points are considered. Any matrix $S \in \mathbb{C}^{\nu\times\nu}$ can be transformed into a block-diagonal matrix by a similarity transformation, where the block-diagonal terms are either in the complex-valued Jordan form as in (5.9), or in the real-valued Jordan form as in (5.12). Consequently, the matching property can be retrieved for each block-diagonal term by application of the results of Lemma 5.1 and 5.2. △

Remark 5.3. It is emphasized that the tangential directions in Theorems 5.1 and 5.2 are the columns of the matrix LT rather than the columns of the matrix L. Thus, selecting a matrix S that is not in Jordan form also results in a coordinate transformation in the tangential directions. △

5.2.4 Family of models achieving higher-order moment matching

Theorems 5.1 and 5.2 characterize a matching property for time-domain MM for MIMO models in terms of the transfer-function matrices of the full- and of the reduced-order models. However, since there is a summation over moments in (5.14) and (5.16), in general, models found by application of these theorems do not achieve moment matching in accordance with Definition 5.2. In this section, we exploit the results in Theorems 5.1 and 5.2 to arrive at a family of (reduced-order) models that do achieve moment matching according to Definition 5.2. Hereto, we present specific selections of the tangential directions $\ell_0, \ldots, \ell_k$, while restricting S to be either in the complex-valued Jordan form in Theorem 5.1 or in the real-valued Jordan form in Theorem 5.2.

Let us first turn to the case in which the reduced-order model is obtained via the result in Theorem 5.1. In that case, it can be shown that if S is in complex-valued Jordan form, i.e., $S = \Phi_k(s^\star)$, either of the specific choices for $L = \begin{bmatrix} \ell_0 & \ell_1 & \ldots & \ell_k \end{bmatrix} \in \mathbb{C}^{m \times k+1}$:

C1 $\ell := \ell_0 = \cdots = \ell_k \in \mathbb{R}^m$;

C2 $\ell := \ell_0 \in \mathbb{R}^m$ and $\ell_i = 0_{m\times 1}$ for $i = 1, \ldots, k$,

ensures the following match:

$$\eta_{\tilde{k}}^{\Sigma}(s^\star, \ell) = \eta_{\tilde{k}}^{\hat{\Sigma}}(s^\star, \ell), \text{ for } \tilde{k} = 0, \ldots, k. \tag{5.17}$$

Consequently, model (5.4) is a (reduced-order) model of (5.1) at $(s^\star, \ell)$. The transfer-function matrices of the full- and of the reduced-order models and their first k derivatives match along the direction $\ell \in \mathbb{C}^m$ at the interpolation point $s^\star$. If the matrix S is not in Jordan form, then after a similarity transformation (with matrix T, see Theorem 5.1), the columns of the matrix LT should satisfy **C1** or **C2** for (5.4) to be a model of (5.1) in accordance with Definition 5.2.

Next, we treat the case where the reduced-order model is obtained via the result in Theorem 5.2 and S in real-valued Jordan form, i.e., $S = \Psi(s^\star)$. In that case, it can be shown that either of the specific choices $L = \begin{bmatrix} \ell_0 & \ell_1 & \ldots & \ell_{2k+1} \end{bmatrix} \in \mathbb{R}^{m\times 2(k+1)}$:

R1 $\ell := \ell_0 = \cdots = \ell_{2k+1} \in \mathbb{R}^m$;

R2 $\ell := \ell_0 \in \mathbb{R}^m$ and $\ell_i = 0_{m\times 1}$ for $i = 1, \ldots, 2k+1$;

R3 $\ell := \ell_1 \in \mathbb{R}^m$ and $\ell_0 = \ell_i = 0_{m\times 1}$ for $i = 2, \ldots, 2k+1$;

ensures that the transfer-function matrices of the full- and of the reduced-order model match as follows:

$$\eta_{\tilde{k}}^{\Sigma}(s^\star, \ell) = \eta_{\tilde{k}}^{\hat{\Sigma}}(s^\star, \ell), \tag{5.18a}$$

$$\eta_{\tilde{k}}^{\Sigma}(\bar{s}^\star, \ell) = \eta_{\tilde{k}}^{\hat{\Sigma}}(\bar{s}^\star, \ell), \tag{5.18b}$$

for $\tilde{k} = 0, \ldots, k$. Consequently, model (5.4) is a (reduced-order) model of (5.1) at $(s^\star, \ell)$ and $(\bar{s}^\star, \ell)$. If the matrix S is not in Jordan form, then after a similarity transformation (with matrix T, see Theorem 5.2), the columns of the matrix LT should satisfy **R1**, **R2**, or **R3**, for (5.4) to be a model of (5.1) in accordance with Definition 5.2.

The selections of ℓ_i's given above for the complex- and real-valued Jordan cases ensure that the families of models in Theorems 5.1 and 5.2 achieve moment matching as classically intended, i.e., in the sense of Definition 5.2. However, note that Theorem 5.2 can also be used to provide generalized matching conditions with a complex tangential direction. For instance, in the case

R4 $\ell_i = \ell_0$ for i even and $\ell_j = \ell_1$ for j odd, $i, j \leq 2k+1$, i.e., $L = \begin{bmatrix} \ell_0 & \ell_1 & \ldots & \ell_0 & \ell_1 \end{bmatrix} \in \mathbb{R}^{m \times 2(k+1)}$,

the following match is obtained for $\ell := \ell_0 + j\ell_1$:

$$\eta_{\tilde{k}}^{\Sigma}(s^\star, \ell) = \eta_{\tilde{k}}^{\hat{\Sigma}}(s^\star, \ell), \tag{5.19a}$$

$$\eta_{\tilde{k}}^{\Sigma}(\bar{s}^\star, \bar{\ell}) = \eta_{\tilde{k}}^{\hat{\Sigma}}(\bar{s}^\star, \bar{\ell}), \tag{5.19b}$$

for $\tilde{k} = 0, \ldots, k$. Model (5.4) is then a (reduced-order) model of (5.1) at $(s^\star, \ell)$ and $(\bar{s}^\star, \bar{\ell})$ in accordance with Definition 5.2. The tangential direction $\ell \in \mathbb{C}^m$, in this case, is generally complex-valued.

Remark 5.4. Time-domain moment matching for MIMO models was also pursued informally in [119, 121, 179, 224]. In those works, however, the statements for the MIMO case were ambiguously formulated without proof and only for a special case. △

5.2.5 Family of models that achieve zeroth-order moment matching

Finally, in this section, we present a family of models that achieves moment matching for zero moments at multiple interpolation points. This family is described by real-valued matrices, which is the most practical case, and may be viewed as a MIMO counterpart of the results of [16].

Theorem 5.3. *Consider model* (5.1) *with* m *inputs and* p *outputs, the non-derogatory matrix* $S \in \mathbb{R}^{\nu \times \nu}$ *with all distinct eigenvalues, and any matrix* $L \in \mathbb{R}^{m \times \nu}$. *Suppose* $\sigma(A) \cap \sigma(S) = \emptyset$ *and write the eigenvalues* $\sigma(S)$ *of the matrix* S *as follows:*

$$\sigma(S) := \{s_1, \bar{s}_1, \ldots, s_r, \bar{s}_r, s_{r+1}, s_{r+2}, \ldots, s_q\}, \tag{5.20}$$

consisting of r *pairs of complex conjugated eigenvalues and* $q = \nu - 2r$ *real eigen-*

values. Furthermore, consider any non-singular matrix $T \in \mathbb{R}^{\nu\times\nu}$ *such that*

$$\tilde{S} \coloneqq T^{-1}ST = \begin{bmatrix} \Psi_0(s_1) & \cdots & 0_2 & 0_{2\times 1} & \cdots & 0_{2\times 1} \\ \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0_2 & \cdots & \Psi_0(s_r) & 0_{2\times 1} & \cdots & 0_{2\times 1} \\ 0_{1\times 2} & \cdots & 0_{1\times 2} & \Phi_0(s_{r+1}) & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ 0_{1\times 2} & \cdots & 0_{1\times 2} & 0 & \cdots & \Phi_0(s_\nu) \end{bmatrix} \tag{5.21}$$

with $\Phi_0(\cdot)$ *and* $\Psi_0(\cdot)$ *defined in Lemmas 5.1 and 5.2, respectively, Then, for any* $G \in \mathbb{R}^{\nu\times m}$ *such that* $\sigma(S)\cap\sigma(S-GL) = \emptyset$, *model* (5.4) *with* $F \coloneqq S-GL, H = C\Pi$, *and with* $\Pi \in \mathbb{R}^{n\times\nu}$ *the unique solution of the Sylvester equation*

$$A\Pi + BL = \Pi S, \tag{5.22}$$

achieves matching according to

$$\eta_0^{\Sigma}(s_k, \ell_{2k-1} + j\ell_{2k}) = \eta_0^{\hat{\Sigma}}(s_k, \ell_{2k-1} + j\ell_{2k}), \tag{5.23a}$$

$$\eta_0^{\Sigma}(\bar{s}_k, \ell_{2k-1} - j\ell_{2k}) = \eta_0^{\hat{\Sigma}}(\bar{s}_k, \ell_{2k-1} - j\ell_{2k}), \tag{5.23b}$$

for $k = 1, \ldots, r$, *and achieves zeroth-order moment matching according to*

$$\eta_0^{\Sigma}(s_{r+k}, \ell_{2r+k}) = \eta_0^{\hat{\Sigma}}(s_{r+k}, \ell_{2r+k}), \tag{5.24}$$

for $k = 1, \ldots, q$, *where* $\ell_1, \ldots, \ell_\nu \in \mathbb{R}^m$, *are the columns of the matrix* LT, *i.e.,*

$$LT = \begin{bmatrix} \ell_1 & \cdots & \ell_\nu \end{bmatrix}. \tag{5.25}$$

▲

Proof. Notice the block-diagonal structure of the matrix $\tilde{S}$ in (5.21). Then, by the results of Theorems 5.1 and 5.2 for zeroth moments, applied to each block-diagonal element of $\tilde{S}$ and corresponding columns of the matrix LT, the matching properties (5.23) and (5.25) are guaranteed. □

Application of Theorem 5.3 results in model (5.4) being a (reduced-order) model of (5.1) at $\sigma(S)$, in accordance with Definition 5.1, along the tangential directions as given in the theorem. Notice again that the tangential directions are the columns of the matrix LT rather than L, see also Remark 5.3. Analogue to Theorem 5.3, a theorem for complex-valued matrices (S, L) can be formulated.

5.3 Illustrative example

Consider the following 4-th order model with two inputs and one output:

$$K(s) = \frac{\begin{bmatrix} 1, & (s+5) \end{bmatrix}}{(s+1)(s+2)(s+3)(s+4)} \in \mathbb{C}^{1\times 2} \tag{5.26}$$

with $s \in \mathbb{C}$. Denote the transfer function from input 1 to the output by $K_1(s)$ and the transfer function from input 2 to the output by $K_2(s)$. We pick the following matrix S with interpolation points $\sigma(S) = \pm j$ and matrix L:

$$S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{and} \quad L = I_2. \tag{5.27}$$

For this choice of (S, L), the matching property according to **R4** in Section 5.2.4 applies with

$$\ell \coloneqq \begin{bmatrix} 1 \\ j \end{bmatrix}, \tag{5.28}$$

and it can be verified that $C\Pi$ is as follows:

$$C\Pi = \begin{bmatrix} \mathrm{Re}(\mathcal{K}(j)) & \mathrm{Im}(\mathcal{K}(j)) \end{bmatrix} = \begin{bmatrix} \frac{2}{17} & \frac{-1}{34} \end{bmatrix}. \tag{5.29}$$

with $\mathcal{K}(s) = K(s)\ell = K_1(s) + jK_2(s)$. The relation (5.29) holds for any minimal realization of (5.26).

The family of reduced-order models with state dimension $\nu = 2$, parametrized by G is described by (5.4) with $F = S - GL$ and $H = C\Pi$. Here, S and L are given in (5.27) and $C\Pi$ is given in (5.29). For simplicity, we select $G = I_2$, which yields

$$F = S - GL = \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}, \tag{5.30}$$

and satisfies $\sigma(F) \cap \sigma(S) = \emptyset$. The transfer function of the reduced-order model is then as follows:

$$\begin{aligned} \hat{K}(s) &= C\Pi(sI_2 - F)^{-1}I_2 \\ &= \frac{C\Pi}{(s+1)^2 + 1} \begin{bmatrix} s+1 & 1 \\ -1 & s+1 \end{bmatrix}. \end{aligned} \tag{5.31}$$

Using the results in Section 5.2.4, it is expected that the matching property (5.19) is satisfied for $\tilde{k} = 0$, which reads as follows:

$$K(j)\ell = \hat{K}(j)\ell, \tag{5.32a}$$

$$K(-j)\bar{\ell} = \hat{K}(-j)\bar{\ell}. \tag{5.32b}$$

It can indeed be verified that both sides of (5.32a) are equal to $-340 + 170j$ and both sides of (5.32b) are equal to $-340 - 170j$, which shows that the equalities (5.32a) and (5.32b) hold. The Bode magnitude plots of $K(j\omega)\ell$ and $\hat{K}(j\omega)\ell$ in Figure 5.1c evidently show that there is a match at $\omega = 1$ rad/sec. Thus the reduced-order model achieves a generalized matching condition (corresponding to (5.19)) with a complex-valued tangential direction and is a model of (5.1) at (j, ℓ) and $(-j, \bar{\ell})$ according to Definition 5.2.

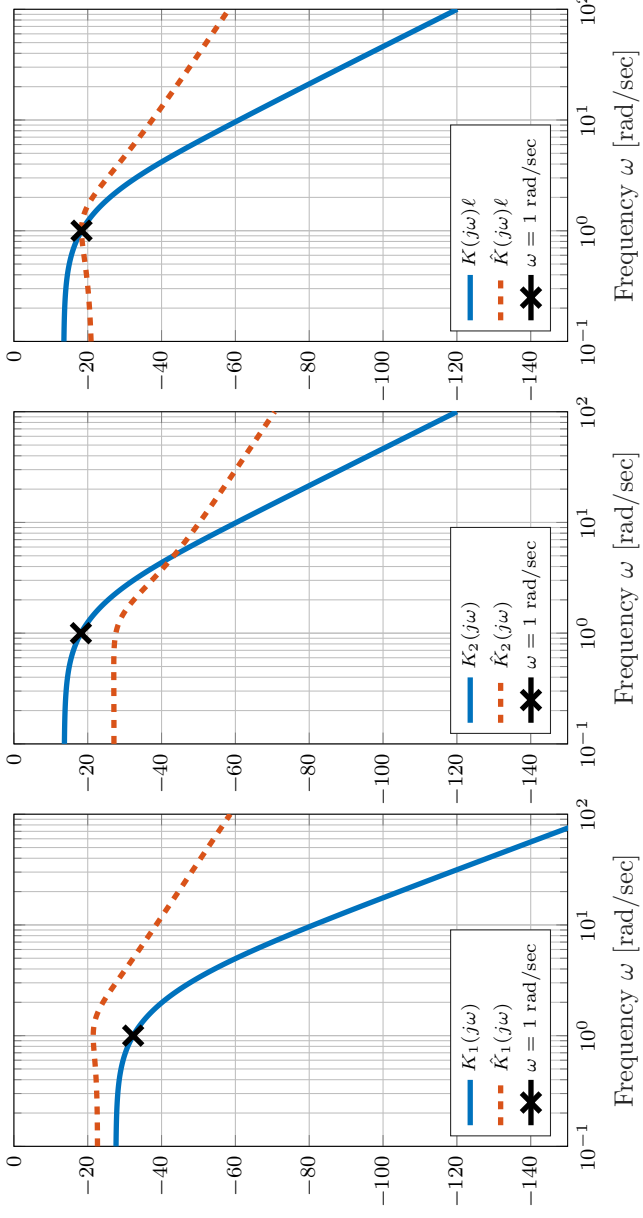


Figure 5.1. (a) Bode magnitude plot of $K_1(j\omega)$ (blue/solid), $\hat{K}_1(j\omega)$ (red/dashed); (b) Bode magnitude plot of $K_2(j\omega)$ (blue/solid), $\hat{K}_2(j\omega)$ (red/dashed); (c) Bode magnitude plot of $K(j\omega)\ell$ (blue/solid), $\hat{K}(j\omega)\ell$ (red/dashed). The black crosses indicate the interpolation frequency $\omega = 1$. A mismatch is observed between $K_i(j\omega)$ and $\hat{K}_i(j\omega)$ at $\omega = 1$ for both $i = 1, 2$. A match is observed between $K(j\omega)\ell$ and $\hat{K}(j\omega)\ell$ at $\omega = 1$.

5.4 Conclusions

This chapter has pointed out that time-domain moment matching applied to MIMO LTI models does not always achieve a match in moments that is consistent with classical moment matching techniques based on Krylov subspaces. To achieve a match that is consistent with classical Krylov subspace methods, the parameters for time-domain moment matching should be carefully chosen. In particular, the matrix characterizing the tangential directions should be chosen with a specific structure. Furthermore, the matrix characterizing the interpolation points should be chosen either in Jordan form or the tangential directions should be corrected by means of a coordinate transformation. The results of this chapter facilitate parameter selection for time-domain moment matching such that the match in the transfer-function matrix is consistent with the match that is achieved by classical Krylov subspace methods.

6

Model reduction by moment matching with preservation of global stability for a class of nonlinear models

Model reduction by time-domain moment matching extends naturally to nonlinear models, where the notion of moments has a local nature stemming from the center manifold theorem. In this chapter, the notion of moments of nonlinear models is extended to the global case and is, subsequently, utilized for model order reduction of convergent Lur'e-type nonlinear models. This model order reduction approach preserves the Lur'e-type model structure, inherits the frequency-response function interpretation of moment matching, preserves the convergence property, and allows deriving an a posteriori error bound. By the grace of the preservation of the convergence property, the reduced-order Lur'e-type model can be reliably used for generalized excitation signals without showing instability issues. In a case study, the reduced-order model accurately matches the moment of the full-order Lur'e-type model and accurately describes the steady-state model response under input variations.

The contents of this chapter are published in: Shakib, M. F., Scarciotti, G., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Model reduction by moment matching with preservation of global stability for a class of nonlinear models. *Submitted for journal publication.* Preliminary results have been presented in: Shakib, M. F., Scarciotti, G., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. (2021, May). Model reduction by moment matching for convergent Lur'e-type models. In *Proceedings of the American Control Conference*, (pp. 4449-4454).

6.1 Introduction

High-fidelity dynamical models of systems are essential in many engineering applications for analysis, prediction, and control design. Such models are typically described by a large number of coupled first-order differential equations. Model simulation of such large dimensional models is computationally expensive and, sometimes, even infeasible due to limited computational and data storage capabilities. To reduce the computational cost and make model simulation feasible, the full-order model is replaced by a reduced-order model which preserves some key properties of the full-order model, e.g., stability properties. Techniques to find a reduced-order model from a full-order model are called model order reduction techniques.

For the class of linear time-invariant (LTI) models, several reduction methods such as balanced truncation [174], Hankel-norm approximations [98] and the interpolation approach [91] have been proposed in the literature. However, most systems are essentially nonlinear and their models are, consequently, nonlinear too. The moment matching approach, categorized in the class of the interpolation approaches, has a natural extension to nonlinear models. Moments of LTI models are defined as the coefficients of the Laurent series expansion of the transfer function at a complex interpolation point, see [10], and the reduction method aims at matching the moments of the reduced-order model to those of the full-order model. In [17], a time-domain interpretation of moment matching is given, exploiting the connection between the solution of the so-called Sylvester equation and the transfer-function definition of moments [91]. This time-domain interpretation has naturally led to the definition of moments for nonlinear models consistent with the definition of moments for LTI models and to reduced-order nonlinear models that achieve moment matching [16, 225, 226]. In these works, the definition of moments for nonlinear models makes use of the center manifold theorem and is, therefore, only defined locally in the neighborhood of the origin. Consequently, the formulated reduced-order model is by definition only an approximation of the full-order model for the same specific input generated by the signal generator and only in the neighborhood of the origin. In general, an estimate for the size of the neighborhood of the origin is lacking, an error bound is lacking and, by the same token, the reduction methods do not preserve the model structure of the full-order model.

In this chapter, the time-domain moment matching approach in [16, 225, 226] is extended to the global case for the class of convergent Lur'e-type nonlinear models. Lur'e-type models, see Figure 6.1, consist of LTI dynamics placed in feedback with a static nonlinearity and arise naturally in problems with localized nonlinearities [135], making them practically relevant. Convergent models exhibit for any bounded input, a bounded and globally asymptotically stable steady-state solution, implying that initial conditions fade out. Furthermore, these models have a finite incremental gain, implying that a small perturbation of the input

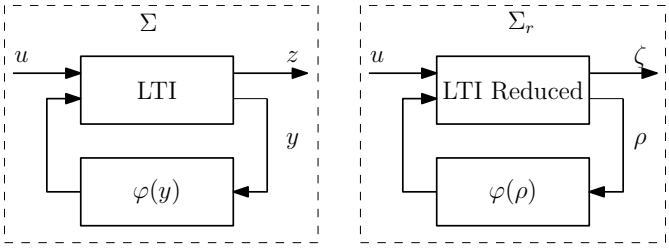


Figure 6.1. Full-order (left) and reduced-order (right) Lur'e-type model. Only the dimension of the LTI block is reduced.

results in a small perturbation of the output. In this chapter, the notion of a moment of a generic nonlinear model is extended from the local context, as in [16, 225, 226], to the global context by replacing the center manifold theorem with a global invariant manifold theorem, assuring the existence of a globally invariant manifold given that the model is convergent as detailed in this chapter.

The reduction method for Lur'e-type models only reduces the state dimension of the LTI block and inherits the static nonlinearity of the full-order model. Consequently, the reduction method preserves the Lur'e-type model structure of the full-order Lur'e-type model. On the one hand, such structure preservation also preserves the physical interpretation of the Lur'e-type model. On the other hand, for the class of Lur'e-type models, a rich array of analysis and controller design tools are available in the literature, see, e.g., [135], which are then compatible with the reduced-order model too. To achieve moment matching and to preserve the Lur'e-type model structure including the static nonlinearity, the frequency response functions (FRFs) of the LTI dynamics should match at an infinite number of interpolation points, which is generally not possible if the model order is reduced. Therefore, the proposed reduction methodology only matches the FRF at a finite number of interpolation points, thereby *approximating* the moments of the full-order model by the moments of the reduced-order model. The proposed reduction method enjoys several benefits besides (Lur'e-type) structure preservation: namely, it preserves the convergence property (which implies preservation of global, as opposed to local, stability properties), it inherits the FRF interpretation of moment matching, and it provides a computable a posteriori error bound. The error bound bounds the L_2-norm of the difference between the moment of the full-order and reduced-order model and also generalizes to the steady-state mismatch between the response of the full-order Lur'e-type model and the response of the reduced-order Lur'e-type model for generalized inputs. Such error bound is important given the fact that moment matching is only achieved in the approximate sense defined in Section 6.2.

To summarize, the main contributions of this chapter are (i) the extension of the notion of moments to the global case for a generic class of convergent nonlin-

ear models; and (ii) a constructive model order reduction approach for convergent Lur'e-type models that preserves the convergence property in addition to preserving the Lur'e-type model structure. Furthermore, this reduction approach has an FRF interpretation and is equipped with a corresponding error bound. The methods developed in this chapter are quantitatively analyzed in a case study on a flexible beam.

Related work includes methods [38, 280] that require solely input, output, and state data generated by the full-order model to construct a reduced-order model that is incrementally stable. The resulting reduced-order model is limited to certain classes of nonlinear models, including polynomial state-space models, thereby does not generally preserve the model structure of the full-order model and lacks an error bound. The methods in [31, 32] consider nonlinear models that can be decomposed into a feedback interconnection of an LTI subsystem with a (dynamic) nonlinear subsystem. These methods reduce only the state dimension of the LTI subsystem and formulate conditions under which the reduced-order nonlinear model preserves the same stability properties of the full-order model. Furthermore, these methods provide an a priori error bound if the reduction technique used for reducing the LTI subsystem provides one. The work in [33] relies on the so-called incremental balancing method and also preserves stability.

The remainder of this chapter is structured as follows. The end of Section 1 introduces the notation used throughout the rest of this chapter. Section 2 extends the notion of moments to the global case for generic convergent nonlinear models and formally introduces the model order reduction problem for convergent Lur'e-type models. Section 3 proposes an approach to the model order reduction of Lur'e-type models. Section 4 describes the results of a case study that illustrates the application and benefits of the proposed model-order reduction approach. Section 5 gives the concluding remarks.

Notation and preliminaries Throughout this chapter, the following notation is used. By $\mathbb{Z}, \mathbb{Z}_{\geq 0}, \mathbb{R}, \mathbb{R}_{\geq 0}, \mathbb{C}$, $\mathbb{C}^0, \mathbb{C}^-$ we respectively denote the set of integers, non-negative integers, real numbers, non-negative real numbers, complex numbers, complex numbers with zero real part and complex numbers with a negative real part. For a vector $x \in \mathbb{R}^n$, we denote the Euclidean norm by $|x| := \sqrt{x^\top x}$. The set of eigenvalues of a matrix $A \in \mathbb{R}^{n\times n}$ is denoted by $\sigma(A)$ and the matrix A is positive (negative) definite, denoted by $A \succ 0\,(A \prec 0)$, if all its eigenvalues are positive (negative). A continuous function $\alpha : [0, a) \to [0, +\infty)$ is said to belong to the class $\mathcal{K}$ if it is strictly increasing and $\alpha(0) = 0$. It is said to belong to the class $\mathcal{K}_\infty$ if $a = +\infty$ and $\alpha(r) \to +\infty$ as $r \to +\infty$. A continuous function $\beta : [0, a) \times [0, +\infty) \to [0, +\infty)$ is said to belong to the class $\mathcal{KL}$ if, for each fixed s, the mapping $\beta(r, s)$ belongs to class $\mathcal{K}$ with respect to r and, for each fixed r, the mapping $\beta(r, s)$ is decreasing with respect to s and $\beta(r, s) \to 0$ as $s \to \infty$. By $L_2(T)$ we denote the space of continuous real-valued T-periodic scalar functions $y(t)$ satisfying $\|y\|_{L_2} < +\infty$, where $\|y\|^2_{L_2} := \frac{1}{T}\int_0^T |y(t)|^2 \mathrm{d}t$ is the L_2-norm.

6.2 Problem statement

Moments of nonlinear models are defined in the literature based on the existence of a solution to a partial differential equation (PDE) which characterizes a center manifold that is only defined locally. Section 6.2.1 introduces a global invariant manifold theorem that replaces the center manifold theorem allowing for defining moments in a global context for the generic class of convergent nonlinear models. Section 6.2.2 presents tractable conditions for Lur'e-type models under which moments are well-defined. Section 6.2.3 motivates that *exact* moment matching can generally not be achieved, leading to the introduction of the *approximate moment matching* problem.

6.2.1 Moments of generic nonlinear models

Consider a single-input, single-output (SISO), continuous-time minimal nonlinear model described by the equations

$$\dot{x} = f(x,u), \quad y = h(x) \tag{6.1}$$

with $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}, y(t) \in \mathbb{R}$, the mapping f locally Lipschitz in x and continuous in u and the mapping h locally Lipschitz in x. Furthermore, consider a signal generator described by the equations

$$\dot{\tau} = s(\tau), \quad u = l(\tau) \tag{6.2}$$

with $\tau(t) \in \mathbb{R}^\nu$, the mappings s and l locally Lipschitz in τ. We assume that the solutions of (6.2) exist on the whole time axis $\mathbb{R}$, which is the case, e.g., if s is globally Lipschitz in τ. Moreover, consider the interconnected model

$$\dot{\tau} = s(\tau), \quad \dot{x} = f(x, l(\tau)), \quad y = h(x). \tag{6.3}$$

Prior to defining the moments, we define the notion of observability.

Definition 6.1 ([226])**.** *The signal generator* (6.2)*, characterized by the pair* (s,l)*, is observable if for any pair of initial conditions* $\tau_a(0) \in \mathbb{R}^\nu$ *and* $\tau_b(0) \in \mathbb{R}^\nu$*, such that* $\tau_a(0) \neq \tau_b(0)$*, the corresponding output trajectories* $l(\tau_a(t))$ *and* $l(\tau_b(t))$ *are such that* $l(\tau_a(t)) - l(\tau_b(t)) \not\equiv 0$*.* ■

Definition 6.2. *Consider the interconnected model* (6.3) *and suppose that the pair* (s,l) *is observable according to Definition 6.1. Suppose that there exists a unique function* $\pi : \mathbb{R}^\nu \to \mathbb{R}^n : \tau \mapsto \pi(\tau)$ *such that the graph*

$$\mathcal{M} \coloneqq \{(\tau, x) : x = \pi(\tau), \tau \in \mathbb{R}^\nu\} \tag{6.4}$$

is invariant with respect to the interconnected model (6.3)*. Then, the function* $h \circ \pi$ *is called the moment of the model* (6.1) *at* (s,l)*.* ■

The definition of moments in Definition 6.2 is consistent with its counterpart for LTI models [226], though different from [16, 225, 226], see Remark 6.1. Different types of sufficient conditions are formulated in the literature for the function π to exist and be unique, see, e.g., [122, 135]. In the scope of this work, we pose two assumptions that guarantee the existence and uniqueness of the function π in Definition 6.2.

Assumption 6.1. *For any $a > 0$, there exists a $b > 0$ such that the initial condition $|\tau(0)| \leq a$ implies that the state evolution $\tau(t) \in \mathbb{R}^\nu$ of the signal generator* (6.2) *satisfies $|\tau(t)| \leq b$ for all $t \in (-\infty, \infty)$. Furthermore, the signal generator* (6.2) *is observable according to Definition 6.1.* ■

Next, we pose an assumption on the stability properties of the nonlinear model (6.1). Hereto, define $\mathcal{U}$ as the set of piecewise continuous functions $u(t) \in \mathbb{R}$ that are defined and bounded on $t \in \mathbb{R}$.

Definition 6.3 ([200])**.** *The model* (6.1) *is said to be globally (uniformly, exponentially) convergent if for every input $u \in \mathcal{U}$, there exists a solution $\bar{x}_u$ to* (6.1) *satisfying the following conditions:*

- *$\bar{x}_u$ is defined and bounded on $t \in \mathbb{R}$,*
- *$\bar{x}_u$ is globally (uniformly asymptotically, exponentially) stable.*

■

The solution $\bar{x}_u$ is called the steady-state solution. The notion of input-to-state convergence is an even stronger stability property and is defined as follows.

Definition 6.4 ([200])**.** *Model* (6.1) *is said to be input-to-state convergent if it is globally uniformly convergent for the class of inputs $\mathcal{U}$ and for every input $u \in \mathcal{U}$, model* (6.1) *is input-to-state stable with respect to the steady-state solution $\bar{x}_u(t)$, i.e., there exist a $\mathcal{KL}$-function $\beta(r, s)$ and a $\mathcal{K}_\infty$-function $\gamma(r)$ such that any solution $x(t)$ of model* (6.1) *corresponding to some input $\hat{u}(t) \coloneqq u(t) + \Delta u(t)$ satisfies*

$$|x(t) - \bar{x}_u(t)| \leq \beta(|x(t_0) - \bar{x}(t_0)|, t - t_0) + \gamma\left(\sup_{t_0 \leq \tau \leq t} |\Delta u(t)|\right), \tag{6.5}$$

for all $t, t_0 \in \mathbb{R}$, $t \geq t_0$. The functions $\beta(r, s)$ and $\gamma(r)$ may depend on the particular input u. ■

Convergent models forget their initial condition and converge to the uniquely defined steady-state solution $\bar{x}_u$. In addition, as evidenced from (6.5), input-to-state convergent models are robust against input variations. Indeed, (6.5) guarantees that in steady state the difference $|x(t) - \bar{x}_u(t)|$ increases monotonically with $|\Delta u(t)|$, implying that a small $|\Delta u(t)|$ results in a small $|x(t) - \bar{x}_u(t)|$.

Assumption 6.2. *The model* (6.1) *is input-to-state convergent according to Definition 6.4.* ■

Assumptions 6.1 and 6.2 guarantee the existence of a *globally* asymptotically stable invariant manifold. This invariant manifold is the counterpart of the center manifold used in previous literature [16, 225, 226].

Lemma 6.1 ([200])**.** *Under Assumptions 6.1 and 6.2, there exists a unique, continuous function π as in Definition 6.2, such that the graph $\mathcal{M}$ in* (6.4) *is invariant with respect to the interconnected model* (6.3)*. Moreover, for every input $u(t) \in \mathcal{U}$, the steady-state solution $\bar{x}_u = \pi(\tau(t))$ is globally uniformly asymptotically stable. Furthermore, if $\pi(\tau)$ is continuously differentiable, i.e., $\pi(\tau) \in C^1$, then $\pi(\tau)$ solves the partial differential equation*

$$\frac{\partial \pi(\tau)}{\partial \tau} s(\tau) = f(\pi(\tau), l(\tau)), \quad \tau(t) \in \mathbb{R}^{\nu}. \tag{6.6}$$

▲

Lemma 6.1 guarantees that moment $h \circ \pi$, see Definition 6.2, is well-defined (also *non-locally*) for nonlinear models that enjoy the input-to-state convergence property. Furthermore, since the graph $\mathcal{M}$ in (6.4) is described by the globally asymptotically stable steady-state solution $\bar{x}_u$, it can be found by computer simulation of the dynamics of the interconnected dynamics (6.3). For example, the graph $\mathcal{M}$ can be computed efficiently using the so-called MTF simulation algorithm [196] for the class of Lur'e-type models. Remark 6.2 below comments on finding reduced-order models that preserve the convergence property. The notion of a moment of a nonlinear model in a global context is employed in the remainder of this chapter to devise a numerically tractable reduction approach for the class of convergent Lur'e-type models.

Remark 6.1. The notion of moments of nonlinear models has been introduced in [16], see also [225, 226]. In these works, under certain assumptions, it was guaranteed that $\pi(\tau)$ is C^1 in the neighborhood of the origin. Subsequently, the definition of moments was based on the solution of the partial differential equation (6.6). However, in this work, Assumptions 6.1 and 6.2 do not guarantee that $\pi(\tau)$ is C^1, but only guarantee that the invariant manifold described by the graph $\mathcal{M}$ in (6.4) exists, is unique and continuous. Therefore, the notion of a moment as in Definition 6.2, is solely based on $\mathcal{M}$ in (6.4), rather than the solution of the PDE (6.6). A definition of moments based on a similar invariant set as in (6.4) was introduced in [227] for linear differential inclusions. △

Remark 6.2. Combining the results in [226] with the insight that the convergence property allows for the well-defined global definition of moments (as in Definition 6.2 of the current chapter), an extension towards a family of reduced-order models can be formulated that achieves moment matching in the global context. To preserve the convergence property, the reduced-order model should satisfy conditions

for convergence, e.g., the so-called Demidovich's condition resulting in input-to-state convergence, see [200, Theorem 2.29]. △

6.2.2 Moments of convergent Lur'e-type models

Section 6.2.1 presents the definition of moments for generic nonlinear models expressed by (6.1). In general, it is a challenging task to verify whether a nonlinear model of the form (6.1) enjoys the convergence property. However, for the class of convergent Lur'e-type models, easy-to-check conditions are proposed in the literature, which we recall here to guarantee the existence and uniqueness of the graph $\mathcal{M}$ in Definition 6.2.

Lur'e-type models consist of a static nonlinear block placed in feedback with an LTI block, see Figure 6.1. The considered Lur'e-type models are described by the following state-space equations:

$$\Sigma: \quad \begin{aligned} \dot{x} &= Ax + B_1 u + B_2 \varphi(y), \\ y &= C_1 x, \\ z &= C_2 x, \end{aligned} \tag{6.7}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}$ is the input, $y(t) \in \mathbb{R}$ is the input to the nonlinear mapping $\varphi : \mathbb{R} \to \mathbb{R}$, $z(t) \in \mathbb{R}$ is the output and $A \in \mathbb{R}^{n\times n}, B_1, B_2 \in \mathbb{R}^{n\times 1}, C_1, C_2 \in \mathbb{R}^{1\times n}$ are model matrices. The associated FRFs are defined as follows:

$$\Phi_{(i,k)}(j\omega) \coloneqq C_i(j\omega I - A)^{-1} B_k, \text{ for } i, k \in \{1, 2\}. \tag{6.8}$$

The following theorem presents conditions for the convergence of the Lur'e-type model (6.7).

Theorem 6.1 ([200])**.** *Consider the model* (6.7)*. Suppose that for some constant* $\gamma > 0$*, the nonlinear function* φ *satisfies the following incremental sector condition:*

$$\left| \frac{\varphi(y_2) - \varphi(y_1)}{y_2 - y_1} \right| \leq \gamma, \quad \forall\, y_1, y_2 \in \mathbb{R}. \tag{6.9}$$

Denote $A_\gamma^- \coloneqq A - \gamma B_2 C_1$ *and* $A^+ \coloneqq A + \gamma B_2 C_1$*. If there exists a* $P = P^\top \succ 0$ *such that*

$$PA_\gamma^- + (A_\gamma^-)^\top P \prec 0 \text{ and } PA_\gamma^+ + (A_\gamma^+)^\top P \prec 0 \tag{6.10}$$

hold, then model (6.7) *is globally exponentially convergent according to Definition 6.3 and input-to-state convergent according to Definition 6.4.* ▲

From here onwards, we simply say that a model is *convergent* to imply that the model is both exponentially convergent according to Definition 6.3 and input-to-state convergent according to Definition 6.4. Since the dimension n of the full-order

model is assumed large, it is not practical to solve the linear matrix inequalities (LMIs) in Theorem 6.1 to verify whether the full-order model is convergent. Alternatively, one can equivalently verify the following three conditions (see, e.g., [200]): 1) the incremental sector condition (6.9); 2) the matrix A being Hurwitz, i.e., $\sigma(A) \in \mathbb{C}^-$; and 3) satisfaction of the following inequality:

$$\sup_{\omega\in\mathbb{R}} \left|\Phi_{(1,2)}(j\omega)\right| = \sup_{\omega\in\mathbb{R}} \left|C_1(j\omega I - A)^{-1}B_2\right| < \frac{1}{\gamma}. \tag{6.11}$$

This latter inequality can be verified graphically, e.g., using the Bode magnitude plot of $\Phi_{(1,2)}(j\omega)$.

The reduction method to be presented in Section 6.3 inherits the FRF interpretation of LTI moment matching thanks to only considering linear signal generators, defined as follows:

$$\dot{\tau} = S\tau, \quad u = L\tau \tag{6.12}$$

with state $\tau(t) \in \mathbb{R}^\nu$, output $u(t) \in \mathbb{R}$ and matrices $S \in \mathbb{R}^{\nu\times\nu}$, $L \in \mathbb{R}^{1\times\nu}$. The following assumption ensures boundedness of trajectories and observability of the states of the signal generator, which implies that Assumption 6.1 is satisfied.

Assumption 6.3. *The matrix S of* (6.12) *has simple eigenvalues that are located on the imaginary axis. In addition, the pair (S, L) is observable.* ■

Finally, the next assumption guarantees the convergence property for the full-order Lur'e-type model (6.7) and implies Assumption 6.2.

Assumption 6.4. *The model* (6.7) *satisfies the conditions of Theorem 6.1 for $\gamma = \gamma^\star$ for some $\gamma^\star > 0$ and is, therefore, convergent.* ■

Since Assumptions 6.1 and 6.2 are implied by Assumptions 6.3 and 6.4, application of Lemma 6.1 guarantees that the moment of the Lur'e-type model is well-defined, i.e., it guarantees the existence of a globally exponentially stable invariant manifold described by $\bar{x}_u(t) = \pi(\tau(t))$ with $\pi : \mathbb{R}^v \to \mathbb{R}^n : \tau \mapsto \pi(\tau)$. The moment of the full-order model (6.7) at (S, L) is denoted by $C_2\pi$ with π as in Definition 6.2. Moments of convergent Lur'e-type models can efficiently be computed using the MTF algorithm [196]. In conclusion, a global definition of moments of nonlinear models is presented and conditions under which these moments are well-defined are given for the class of Lur'e-type models.

6.2.3 Approximate moment matching problem

Model order reduction of Lur'e-type models boils down to reducing the state dimension of the LTI block since all the dynamics are captured therein. Consequently, the proposed reduction method preserves the Lur'e-type structure of the full-order model and inherits the FRF interpretation that is characteristic to moment matching for LTI models. Solely reducing the dimension of the LTI block

results inevitably in the moments of the full-order Lur'e-type model being approximated rather than being exactly matched, which is further explained below.

Consider the class of models Σ_r of order ν:

$$\Sigma_r : \quad \begin{aligned} \dot{\xi} &= F\xi + G_1 u + G_2\varphi(\rho), \\ \rho &= H_1\xi, \\ \zeta &= H_2\xi, \end{aligned} \tag{6.13}$$

where $\xi(t) \in \mathbb{R}^\nu$ is the state, $u(t) \in \mathbb{R}$ is the input, $\rho(t) \in \mathbb{R}$ is the input to the same nonlinear mapping $\varphi : \mathbb{R} \to \mathbb{R}$ as in (6.7), $\zeta(t) \in \mathbb{R}$ is the output and $F \in \mathbb{R}^{\nu\times\nu}, G_1, G_2 \in \mathbb{R}^{\nu\times 1}, H_1, H_2 \in \mathbb{R}^{1\times\nu}$ are model matrices. Note that Σ_r in (6.13) is of the same structure as Σ in (6.7). The moment of this model at (S, L) is denoted by $H_2 p$, where the function p plays the role of the function π in Definition 6.2. The FRFs associated with the LTI part of (6.13) read as:

$$\Gamma_{(i,k)}(j\omega) \coloneqq H_i(j\omega I - F)^{-1}G_k \text{ for } i, k \in \{1, 2\}. \tag{6.14}$$

The main obstacle in achieving moment matching is that the steady-state output $\bar{z}_u$ of the full-order model, related to the moment via $\bar{z}_u = C_2\bar{x}_u = C_2\pi(\tau(t))$, contains typically an infinite number of frequency components. Since the signal u is generated by the signal generator (6.12), its only content is at the frequencies characterized by the interpolation points $\sigma(S)$. However, due to the nonlinear feedback φ, the signal $\varphi(\bar{y}_u)$ can have contents at an infinite number of frequencies. Therefore, to match the steady-state output $\bar{z}_u$ of the full-order model by the steady-state output $\bar{\zeta}_u$ of the reduced-order model, a match should be achieved between the FRFs of the LTI part of the full-order and the reduced-order models at an infinite number of interpolation frequencies. Even though some methods have been proposed to match moments at infinitely many interpolation points, see [226], these cannot be trivially generalized to the current setting. Therefore, in this work, only a finite number of interpolation frequencies are matched in each of the FRFs in (6.8), resulting inevitably in a mismatch in the nonlinear moment.

Having a mismatch between the moment of the full-order and of the reduced-order Lur'e-type model makes it important to derive an error bound for this mismatch. In addition to preserving the convergence property, deriving such an error bound is an important part of the *approximate* moment matching problem.

Problem 6.1. *Consider the full-order Lur'e-type model* (6.7) *with state dimension* n *and the signal generator* (6.12) *with state* $\tau(t) \in \mathbb{R}^\nu$ *characterized by* (S, L) *and suppose* $\nu < n$. *Suppose Assumptions 6.3 and 6.4 hold for some* $\gamma^\star > 0$. *Denote the moment of the full-order Lur'e-type model* (6.7) *at* (S, L) *by* $C_2\pi$.

The approximate moment matching problem is to find matrices F, G_1, G_2, H_1, H_2, *which define the* ν*-th order model* (6.13) *with moment* $H_2 p$, *such that this reduced-order model satisfies the condition in Theorem 6.1, and to find a constant* $0 \leq \varepsilon < +\infty$ *such that the mismatch between the moment of the full-order and reduced-order Lur'e-type model is upper bounded as follows:*

$$\|C_2\pi(\tau) - H_2 p(\tau)\|_{L_2} \leq \varepsilon \|L\tau\|_{L_2}. \tag{6.15}$$

△

The next section presents a reduction method that guarantees a small ε, resulting in a small mismatch between the moments.

Remark 6.3. Other methods, e.g., the method in [16], do achieve moment matching rather than approximate moment matching, for example, by considering a Wiener model, i.e., linear dynamics followed by a static nonlinear output map. However, in those methods, the model structure is generally not preserved. Furthermore, Lur'e-type models represent a broader class of systems than Wiener models do. △

6.3 Solution to the approximate moment matching problem

Section 6.2 presents the approximate moment matching problem in Problem 6.1. This problem is replaced by a constrained optimization problem in Section 6.3.1. Subsequently, a solution is proposed in Section 6.3.2, which also solves Problem 6.1 as presented in Section 6.3.3. An overview of the reduction method closes this section.

6.3.1 Constrained optimization problem formulation

Frequency-domain insights are used to formulate a constrained optimization problem, which, on the one hand, allows finding a constant ε to solve Problem 6.1 and, on the other hand, aims at reducing ε to improve the accuracy of the estimated reduced-order model. Let us first further motivate moment matching at a finite number of interpolation points by the following two properties.

Property 6.1 ([200])**.** *Consider the model* (6.7) *and suppose Assumption 6.4 holds. If the input $u \in L_2(T)$ is periodic with period T, then the corresponding steady-state outputs $\bar{z}_u$ and $\bar{y}_u$ are also periodic with the same period T.* ▲

Property 6.2. *Consider the model* (6.7) *and suppose Assumption 6.4 holds. If the input $u \in L_2(T)$ has a finite Lipschitz constant $0 \leq \ell < +\infty$ with respect to time t, i.e., $|u(t) - u(t+\kappa)| \leq \ell|\kappa| \, \forall \, t, \kappa \in \mathbb{R}$, then the magnitude of the k-th Fourier coefficient of the Fourier series of $\bar{z}_u$ and $\bar{y}_u$ converges to zero according to $O(1/k)$ for $k \neq 0$.* ▲

Proof. The input u having a finite Lipschitz constant $0 \leq \ell < +\infty$ results in each of the steady-state outputs $\bar{z}_u$ and $\bar{y}_u$ of model (6.7) also admitting a finite Lipschitz constant, possibly different from ℓ. Functions that admit a finite Lipschitz constant have a roll-off in the magnitude of the Fourier coefficients according to $O(1/k)$ for $k \neq 0$, see, e.g., [128, Theorem 4.6]. □

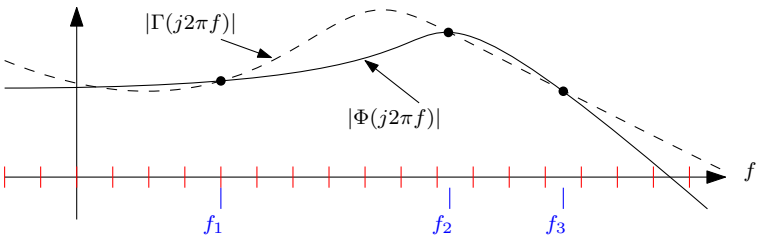


Figure 6.2. The FRFs of the reduced-order and full-order model match at the frequencies in the set $\Omega^0 = \{f_1, f_2, f_3\}$. The mismatch between the FRFs is minimized at the frequencies in Ω^M, here taken on the red grid illustrated in the figure.

Property 6.1 ensures that the signal $\varphi(\bar{y}_u)$ contains the same frequencies as u, but also an infinite number of higher harmonic frequencies. Property 6.2 ensures that the Fourier coefficients corresponding to large frequencies vanish in absolute value, justifying matching only a finite number of frequencies in the corresponding FRFs (note that by Assumption 6.3, the output of the linear harmonic oscillator does admit to a finite Lipschitz constant as required in Property 6.2). Considering these properties, for approximate moment matching, it is beneficial to match the involved FRFs of the LTI block of the Lur'e-type model at the first few harmonics and/or at frequencies corresponding to important model characteristics, such as the 0-Hz frequency or the (anti-)resonance peaks. Hereto, we collect $\eta_{(i,k)}$ user-defined interpolation frequencies in the vector $\Omega^0_{(i,k)} \in \mathbb{R}^{\eta_{(i,k)}}$ for each $i, k \in \{1, 2\}$, where indices i, k refer to the transfer functions (6.8).

To achieve robustness against input variations, we also minimize the mismatch in these FRFs at other frequencies, which has two additional benefits. Firstly, it allows for reducing the mismatch at the harmonic frequencies that are not included as interpolation frequencies in $\Omega^0_{(i,k)}$, resulting in a better approximation of the moment of the Lur'e-type model. Secondly, it adds to the robustness of the reduced-order Lur'e-type model for new inputs that excite different frequencies than the interpolation frequencies. We collect $M_{(i,k)}$ user-defined frequencies in the vector $\Omega^M_{(i,k)} \in \mathbb{R}^{M_{(i,k)}}$ for each $i, k \in \{1, 2\}$, at which the mismatch in FRFs is minimized. An illustration of the sets Ω^0 and Ω^M for a single FRF is presented in Figure 6.2.

Before formally presenting the constrained optimization problem, we define the mismatch in each FRF as follows:

$$\Upsilon_{(i,k)}(j\omega) \coloneqq \Phi_{(i,k)}(j\omega) - \Gamma_{(i,k)}(j\omega), i, k \in \{1, 2\}, \tag{6.16}$$

where $\Phi_{(i,k)}(j\omega)$ and $\Gamma_{(i,k)}(j\omega), i, k \in \{1, 2\}$, are defined in (6.8) and (6.14), re-

spectively. Furthermore, we define the cost function to be minimized:

$$J(F, G_1, G_2, H_1, H_2) \coloneqq \sum_{k=1}^{2} \sum_{i=1}^{2} \sum_{\ell=1}^{M_{(i,k)}} \left| \Upsilon_{(i,k)}(j\omega_{(i,k)}(\ell)) \right|^2, \tag{6.17}$$

where $\omega_{(i,k)}(\ell)$ is the ℓ-th component of $\Omega^0_{(i,k)}$ and $\Upsilon_{(i,k)}$ is defined in (6.16).

Problem 6.2. *Consider the full-order Lur'e-type model* (6.7) *and suppose Assumption 6.4 holds for some* $\gamma^\star > 0$. *Furthermore, consider some given sets of frequencies* $\Omega^0_{(i,k)} \in \mathbb{R}^{\eta_i}, \Omega^M_{(i,k)} \in \mathbb{R}^{M_{(i,k)}}$ *for* $i, k \in \{1, 2\}$. *The reduced-order model* (6.13), *characterized by* F, G_1, G_2, H_1, H_2 *is found by solving the constrained optimization problem:*

$$\min_{F, G_1, G_2, H_1, H_2} \quad J(F, G_1, G_2, H_1, H_2) \tag{6.18a}$$

$$\text{subject to} \quad \Upsilon_{(i,k)}(j\omega) = 0 \, \forall \, \omega \in \Omega^0_{(i,k)}, i, k \in \{1, 2\}, \tag{6.18b}$$

$$Q = Q^\top \succ 0, \tag{6.18c}$$

$$QF^-_{\gamma^\star} + (F^-_{\gamma^\star})^\top Q \prec 0, \tag{6.18d}$$

$$QF^+_{\gamma^\star} + (F^+_{\gamma^\star})^\top Q \prec 0, \tag{6.18e}$$

where $F^-_{\gamma^\star} = F - \gamma^\star G_2 H_1, F^+_{\gamma^\star} = F + \gamma^\star G_2 H_1$ *and* $J(F, G_1, G_2, H_1, H_2)$ *as defined in* (6.17). △

Problem 6.2 has the following interpretation. The minimization of the cost function $J(F, G_1, G_2, H_1, H_2)$ in (6.18a) ensures an optimal fit between the FRFs of the full-order and the reduced-order model, for $i, k \in \{1, 2\}$, at the frequencies $\Omega^M_{(i,k)}$, for $i, k \in \{1, 2\}$. The constraint (6.18b) ensures that the FRF $\Phi_{(i,k)}$ of the full-order model equals to the FRF $\Gamma_{(i,k)}$ of the reduced-order model at the interpolation frequencies $\Omega^i_0, i = 1, 2$, for $i, k \in \{1, 2\}$. The remaining constraints (6.18c) - (6.18e) guarantee that the reduced-order Lur'e-type model (6.13) preserves the convergence property of the full-order Lur'e-type model (6.7) by satisfying the conditions of Theorem 6.1, guaranteeing that a constant ε exists to solve Problem 6.1.

Next, a solution to Problem 6.2 is presented. After that, the connection to Problem 6.1 is made explicit.

Remark 6.4. For the sake of exposition, a frequency-dependent weighting in the cost function (6.17) is not introduced; one could trivially equip the cost function with such weighting. A weighting can be useful to emphasize the importance of a good fit for desired frequency ranges. △

6.3.2 Solution to Problem 6.2

The proposed solution to Problem 6.2 works as follows. Firstly, by application of time-domain moment matching for LTI models, a family of Lur'e-type models

(6.13) is derived such that constraint (6.18b) of Problem 6.2 is satisfied. As shown below, this step yields freedom in parameters G_1, G_2, which is exploited to solve the optimization problem (6.18a) while satisfying the constraints (6.18c) - (6.18e) in order to ensure the convergence property.

Family of reduced-order Lur'e-type models

Let us recall an adapted version of time-domain moment matching for SISO LTI models from [16]. Consider a minimal LTI model described by:

$$\dot{x} = \mathcal{A}x + \mathcal{B}u, \qquad y = \mathcal{C}x, \tag{6.19}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}$ is the input, $y(t) \in \mathbb{R}$ the output, and $\mathcal{A} \in \mathbb{R}^{n\times n}, \mathcal{B} \in \mathbb{R}^{n\times 1}, \mathcal{C} \in \mathbb{R}^{1\times n}$ are model matrices. Consider the reduced-order LTI model

$$\dot{\xi} = \mathcal{F}x + \mathcal{G}u, \qquad y = \mathcal{H}x, \tag{6.20}$$

where $\xi(t) \in \mathbb{R}^\nu$ is the state, $u(t) \in \mathbb{R}$ is the input, $y(t) \in \mathbb{R}$ the output, and $\mathcal{F} \in \mathbb{R}^{\nu\times\nu}, \mathcal{G} \in \mathbb{R}^{\nu\times 1}, \mathcal{H} \in \mathbb{R}^{1\times\nu}$ are model matrices and $\nu < n$. The following theorem borrowed from [16] is specialized to simple interpolation points on the imaginary axis.

Theorem 6.2. *Consider the SISO LTI model* (6.19) *characterized by the matrices* $\mathcal{A}, \mathcal{B}, \mathcal{C}$. *Furthermore, consider the given matrix* $S \in \mathbb{R}^{\nu\times\nu}$ *with simple eigenvalues* $\sigma(S) = \{s_1, \ldots, s_\nu\}$, *the given matrix* $L \in \mathbb{R}^{1\times\nu}$, *and* $\nu < n$. *Suppose Assumption 6.3 holds. Then, for any* $\mathcal{G} \in \mathbb{R}^{\nu\times 1}$ *such that* $\sigma(S) \cap \sigma(S - \mathcal{G}L) = \emptyset$, *the reduced-order model* (6.20) *with matrices* $\mathcal{F} \coloneqq S - \mathcal{G}L, \mathcal{H} \coloneqq \mathcal{C}\Pi$ *with* $\Pi \in \mathbb{R}^{n\times\nu}$ *the unique solution to the Sylvester equation*

$$\mathcal{A}\Pi + \mathcal{B}L = \Pi S, \tag{6.21}$$

matches the 0*-th moments of LTI model* (6.19) *at the eigenvalues of* S, *i.e.,*

$$\mathcal{C}(s_i I - \mathcal{A})^{-1}\mathcal{B} = \mathcal{H}(s_i I - \mathcal{F})^{-1}\mathcal{G}, \quad i = 1, \ldots, \nu.$$

▲

Theorem 6.2 is applied to each of the transfer functions of the LTI part of the Lur'e-type model (6.7) in order to satisfy constraint (6.18b). Hereto, we introduce matrices $S_{(i,k)} \in \mathbb{R}^{\nu_{(i,k)}}$ and $L_{(i,k)} \in \mathbb{R}^{1\times\nu_{(i,k)}}$ for $i, k \in \{1, 2\}$. Furthermore, we introduce the notation $\sigma(S_{(i,k)}) \simeq \Omega^0_{(i,k)}, i, k \in \{1, 2\}$, meaning that if $0 < \alpha \in \Omega^0_{(i,k)}$, then $\pm j\alpha \in \sigma(S_{(i,k)})$ and if $0 \in \Omega^0_{(i,k)}$, then $0 \in \sigma(S_{(i,k)})$. Loosely speaking, the notation $\sigma(S_{(i,k)}) \simeq \Omega^0_{(i,k)}, i, k \in \{1, 2\}$, ensures that the interpolation frequencies in $\Omega^0_{(i,k)}$ are excited by the signal generator. Using this notation, constraint (6.18b) is satisfied by the results presented in the next theorem.

Theorem 6.3. *Consider the LTI part of the Lur'e-type model* (6.7) *characterized by matrices* A, B_1, B_2, C_1, C_2. *Consider the given sets of frequencies* $\Omega^0_{(i,k)} \in \mathbb{R}^{\eta_{(i,k)}}$ *and matrices* $S_{(i,k)} \in \mathbb{R}^{\nu_{(i,k)} \times \nu_{(i,k)}}, L_{(i,k)} \in \mathbb{R}^{1 \times \nu_{(i,k)}}, i, k \in \{1, 2\}$, *and suppose that* $\sigma(S_{(i,k)}) \simeq \Omega^0_{(i,k)}, i, k \in \{1, 2\}$. *Furthermore, suppose all pairs* $(S_{(i,k)}, L_{(i,k)}), i, k \in \{1, 2\}$, *satisfy Assumption 6.3. For each* $i, k \in \{1, 2\}$, *application of Theorem 6.2 with* $(\mathcal{A}, \mathcal{B}, \mathcal{C}) = (A, B_i, C_k)$ *and* $(S, L) = (S_{(i,k)}, L_{(i,k)})$ *results in an LTI model of the form* (6.20) *with matrices*

$$\begin{aligned} \left(F_{(i,k)}, G_{(i,k)}, H_{(i,k)}\right) &\coloneqq (\mathcal{F}, \mathcal{G}, \mathcal{H}) \\ &= \left(S_{(i,k)} - G_{(i,k)} L_{(i,k)}, G_{(i,k)}, C_k \Pi_{(i,k)}\right). \end{aligned} \tag{6.22}$$

Then, under the conditions that $G_{(i,k)} \in \Theta_{(i,k)}, i, k \in \{1, 2\}$, *with*

$$\Theta_{(i,k)} \coloneqq \left\{ G_{(i,k)} \in \mathbb{R}^{\nu_{(i,k)}} \mid \sigma\left(S_{(i,k)}\right) \cap \sigma\left(S_{(i,k)} - G_{(i,k)} L_{(i,k)}\right) = \emptyset \right\}, \tag{6.23}$$

the LTI part of the Lure'-type model (6.13) *characterized by the matrices*

$$F = \mathit{blockdiag}\left(F_{(1,1)}, F_{(1,2)}, F_{(2,1)}, F_{(2,2)}\right), \tag{6.24a}$$

$$G_1 = \begin{bmatrix} G_{(1,1)} \\ 0 \\ G_{(2,1)} \\ 0 \end{bmatrix}, \quad G_2 = \begin{bmatrix} 0 \\ G_{(1,2)} \\ 0 \\ G_{(2,2)} \end{bmatrix}, \tag{6.24b}$$

$$H_1 = \begin{bmatrix} H_{(1,1)} & H_{(1,2)} & 0 & 0 \end{bmatrix}, \tag{6.24c}$$

$$H_2 = \begin{bmatrix} 0 & 0 & H_{(2,1)} & H_{(2,2)} \end{bmatrix}, \tag{6.24d}$$

ensures that $\Upsilon_{(i,k)}(j\omega) = 0$ *for* $\omega \in \Omega^0_{(i,k)}, i, k \in \{1, 2\}$. *Hence, constraint* (6.18b) *of Problem 6.2 is satisfied.* ▲

Proof. The FRFs (6.14) of the LTI part of the Lur'e-type model (6.13) with model matrices (6.24) reads as follows:

$$\Gamma_{(i,k)} = H_{(i,k)} \left(j\omega I - F_{(i,k)}\right)^{-1} G_{(i,k)}, i, k \in \{1, 2\}.$$

For each $i, k \in \{1, 2\}$, by Theorem 6.2, a match is ensured between $\Gamma_{(i,k)}(s)$ and the FRF of the full-order model $\Phi_{(i,k)}(s)$, for all $s \in \sigma(S_{(i,k)})$ if condition (6.23) holds, i.e., $\Upsilon_{(i,k)}(s) = 0$ for all $s \in \sigma(S_{(i,k)})$. Since $\sigma(S_{(i,k)}) \simeq \Omega^0_{(i,k)}, i, k \in \{1, 2\}$, we can conclude that $\Upsilon_{(i,k)}(s) = 0, i, k \in \{1, 2\}$ for all $s \in \sigma(S_{(i,k)})$, hence constraint (6.18b) of Problem 6.2 is satisfied. □

Theorem 6.3 presents the matrices of the LTI block of a family of Lur'e-type models of the form (6.13) that satisfy constraint (6.18b). The family is parametrized by $G_{(i,k)} \in \Theta_{(i,k)}, i, k \in \{1, 2\}$. In the next section, we present a method to find $G_{(i,k)} \in \Theta_{(i,k)}, i, k \in \{1, 2\}$, such that the remaining constraints (6.18c) - (6.18e) are also satisfied, while (6.18a) is minimized.

Remark 6.5. The diagonal structure of the model matrix F in (6.24a) leaves the dynamics decoupled between the four transfer functions. A negative consequence is that when a match is desired in all FRFs at the same frequency, then that same frequency should be included in all $\Omega^0_{(i,k)}, i, k \in \{1, 2\}$, which results in a reduced-order model with a larger-than-needed state dimension. As a remedy, one could apply moment matching for multi-input, multi-output (MIMO) models, since the LTI part of the Lur'e-type model is of MIMO nature. However, current moment matching techniques only allow for moment matching along the so-called tangential directions [10, 28, 100]. As a consequence, not the individual FRFs are matched, but rather their weighted sum, which is undesired in the scope of this work. △

Constrained gradient-based optimization

Let us denote $\theta \coloneqq \{G^{(1,1)}, G^{(1,2)}, G^{(2,1)}, G^{(2,2)}\}$, which contains all to-be-optimized parameters. We note that the parameters θ only appear in the matrix F, G_1 and G_2, see (6.24). Therefore, from here on, we write $F(\theta), G_1(\theta)$ and $G_2(\theta)$ to make their dependency on θ clear. Next, we define $\hat{J}$ as follows:

$$\hat{J}(\theta) \coloneqq J(F(\theta), G_1(\theta), G_2(\theta), H_1, H_2) \tag{6.25}$$

with J as in (6.17). Besides minimizing $\hat{J}$, we aim to preserve the convergence property, which is encoded in the constraints of the following optimization problem:

$$\hat{\theta} = \arg\min_{\theta \in \bar{\Theta}} \hat{J}(\theta), \tag{6.26}$$

where $\bar{\Theta}$ is the set of θ defined as follows:

$$\begin{aligned} \bar{\Theta} = \big\{ \theta \in (\mathbb{R}^{\nu_{(1,1)}} \times \mathbb{R}^{\nu_{(1,2)}} \times \mathbb{R}^{\nu_{(2,1)}} \times \mathbb{R}^{\nu_{(2,2)}}) \mid \exists P = P^\top \succ 0 : \\ PF(\theta)^-_{\gamma^\star} + (F(\theta)^-_{\gamma^\star})^\top P \prec 0, PF(\theta)^+_{\gamma^\star} + (F(\theta)^+_{\gamma^\star})^\top P \prec 0 \, \big\} \end{aligned} \tag{6.27}$$

with $F(\theta)^\pm_{\gamma^\star} \coloneqq F(\theta) \pm \gamma^\star G_2(\theta) H_1$ and $\gamma^\star$ as in Assumption 6.4. We note that $\bar{\Theta} \subset \{\Theta^{(1,1)}, \Theta^{(1,2)}, \Theta^{(2,1)}, \Theta^{(2,2)}\}$, i.e., for any $\theta \in \bar{\Theta}$, the condition (6.23) is satisfied, since satisfaction of (6.27) guarantees that $\sigma(F(\theta)) \in \mathbb{C}^-$, while $\sigma(S_{(i,k)}) \in \mathbb{C}^0$ by Assumption 6.3 for all $i, k \in \{1, 2\}$. One could interpret the minimum in (6.26) as an infimum, for which a numerical solver should then find a sufficiently accurate approximation of that infimum.

The set $\bar{\Theta}$ in (6.27) is derived from the statements in Theorem 6.1 and its LMI constraints are linear in P for fixed θ, hence the constraints in $\bar{\Theta}$ are LMIs. Since $\hat{J}$ in (6.26) is nonlinear in θ, by gradient-based optimization, a local minimum of $\hat{J}$ at $\hat{\theta}$ can be found, which solves the constrained optimization problem (6.26) and, thereby, also solves the constrained optimization problem (6.18a) - (6.18e) in Problem 6.2. To launch the gradient-based search to solve (6.26), an initial convergent reduced-order model is required. A method to find such an initial convergent model is presented next.

Convergent initial reduced-order model

The full-order model satisfies the conditions of Theorem 6.1 by Assumption 6.4. In particular, condition (6.10) is equivalent to bounding the $\mathcal{H}_\infty$ norm of the transfer function of the LTI block of (6.7) from the nonlinearity φ to the output y by the constant $1/\gamma^\star$, see [200], i.e.,

$$\left\|C_1(sI - A)^{-1}B_2\right\|_\infty < \frac{1}{\gamma^\star}, \tag{6.28}$$

where the $\mathcal{H}_\infty$ model norm is defined for stable models, i.e., $\sigma(A) \in \mathbb{C}^-$, as follows:

$$\left\|C_1(sI - A)^{-1}B_2\right\|_\infty \coloneqq \sup_{\omega\in[0,\infty)} |C_1(j\omega I - A)^{-1}B_2|. \tag{6.29}$$

To guarantee convergence of the reduced-order model, it is required that the $\mathcal{H}_\infty$ norm of the transfer function $\Gamma_{(1,2)}$ is also bounded by the same constant $1/\gamma^\star$. Hereto, the following lemma presents a reduced-order LTI model that achieves moment matching and preserves the $\mathcal{H}_\infty$ norm of the full-order LTI model.

Lemma 6.2. *Consider the SISO LTI model* (6.19) *characterized by the matrices* $\mathcal{A}, \mathcal{B}, \mathcal{C}$. *Assume that* $\sigma(\mathcal{A}) \in \mathbb{C}^-$ *and* $\left\|\mathcal{C}(sI - \mathcal{A})^{-1}\mathcal{B}\right\|_\infty < 1/\gamma^\star$, *implying that there exists a matrix* $\bar{P} \succ 0$ *such that the LMIs* (6.10) *are satisfied for* $A = \mathcal{A}, B_2 = \mathcal{B}, C_1 = \mathcal{C}, P = \bar{P}$ *and* $\gamma = \gamma^\star$. *Application of Theorem 6.2 for any matrix* $S \in \mathbb{R}^{\nu\times\nu}$ *and matrix* $L \in \mathbb{R}^{1\times\nu}$ *that satisfy Assumption 6.3 results in a reduced-order model that achieves moment matching at* $\sigma(S)$. *The reduced-order model is of the form* (6.20) *with matrices* $(\mathcal{F}, \mathcal{G}, \mathcal{H}) = (S - \mathcal{G}L, \mathcal{G}, \mathcal{C}\Pi)$ *and* Π *the solution to the Sylvester equation in* (6.21). *Then, for*

$$\mathcal{G} \coloneqq \left(\Pi^\top \bar{P}\Pi\right)^{-1}\Pi^\top \bar{P}\mathcal{B}, \tag{6.30}$$

the matrix $\mathcal{F}$ *is Hurwitz, i.e.,* $\sigma(\mathcal{F}) \in \mathbb{C}^-$, *and the transfer function of the reduced-order model satisfies*

$$\left\|\mathcal{H}(sI - \mathcal{F})^{-1}\mathcal{G}\right\|_\infty < 1/\gamma^\star. \tag{6.31}$$

▲

Proof. The proof can be found in Appendix E.1. □

The following theorem presents an LMI-based method to find a parametrization $\theta^\circ \in \bar{\Theta}$ for the reduced-order model such that all conditions of Theorem 6.1 are satisfied.

Theorem 6.4. *Consider the reduced-order Lur'e-type model* (6.13) *with matrices* (6.24) *and suppose Assumption 6.4 holds for a certain* $\gamma^\star$. *If there exist symmetric positive definite matrices* $P_{(i,k)} = P_{(i,k)}^\top \succ 0 \in \mathbb{R}^{\nu_{(i,k)}\times\nu_{(i,k)}}$, *matrices* $X_{(i,k)} \in \mathbb{R}^{\nu_{(i,k)}}$, *for* $i, k \in \{1, 2\}$, *such that the following two LMIs are satisfied:*

$$\mathcal{L}_{\gamma^\star}^+ + \left(\mathcal{L}_{\gamma^\star}^+\right)^\top \prec 0 \text{ and } \mathcal{L}_{\gamma^\star}^- + \left(\mathcal{L}_{\gamma^\star}^-\right)^\top \prec 0 \tag{6.32}$$

with

$$\mathcal{L}_{\gamma^\star}^{\pm} = \begin{bmatrix} \mathcal{P}_{(1,1)} & 0 & 0 & 0 \\ \pm\gamma^\star X_{(1,2)} H_{(1,1)} & \mathcal{P}_{(1,2)} \pm \gamma^\star \mathcal{X}_{(1,2)} & 0 & 0 \\ 0 & 0 & \mathcal{P}_{(2,1)} & 0 \\ \pm\gamma^\star X_{(2,2)} H_{(1,1)} & \pm\gamma^\star X_{(2,2)} H_{(1,2)} & 0 & \mathcal{P}_{(2,2)} \end{bmatrix}$$

and

$$\begin{aligned} \mathcal{P}_{(i,k)} =& P_{(i,k)} S_{(i,k)} - X_{(i,k)} L_{(i,k)}, \quad i,k \in \{1,2\}, \\ \mathcal{X}_{(1,2)} =& X_{(1,2)} H_{(1,2)}. \end{aligned}$$

Then, the conditions of Theorem 6.1 are satisfied for the reduced-order Lur'e-type model (6.13) *with model matrices* $G_{(i,k)} = P_{(i,k)}^{-1} X_{(i,k)}, i,k \in \{1,2\}$. *Furthermore, condition* (6.23) *is satisfied, i.e.,* $G_{(i,k)} \in \Theta_{(i,k)}$ *for* $i,k \in \{1,2\}$. *Finally, the set of LMIs* (6.32) *is feasible under the stated assumptions.* ▲

Proof. The proof can be found in Appendix E.2. □

The initial model matrices $G^\circ_{(i,k)}, i,k \in \{1,2\}$, found via Theorem 6.4 and collected in θ°, render the reduced-order Lur'e-type model (6.13) convergent. Subsequently, θ° is used to launch a gradient-based search to solve the constrained optimization problem (6.26).

Remark 6.6. The results of Lemma 6.2 can be used beyond the scope of finding an initial convergent Lur'e-type model as presented in Theorem 6.4. For example, these results can be used to compute a $\mathcal{G}$ such that moment matching is achieved for LTI models and the ℓ_2-gain of the full-order model is not exceeded. △

6.3.3 Solution to Problem 6.1

The solution to Problem 6.2 has been presented in Section 6.3.2. This section derives an error bound for the mismatch between the moment of the full-order and the reduced-order model, which is used to compute a constant ε such that Problem 6.1 is solved. Moreover, it is shown that this error bound holds in the more generic case where inputs u are taken from the class of bounded, periodic inputs $L_2(T)$.

Thanks to Assumption 6.4 and by the grace of the fact the reduced-order model is also convergent, a worst-case upper bound on the error between the steady-state outputs $\bar{z}_u$ of the full-order model and $\bar{\zeta}_u$ of the reduced-order model can be formulated. Hereto, the supremum of the mismatch in each FRF is defined as follows:

$$\bar{\Upsilon} \coloneqq \max_{i,k \in \{1,2\}} \sup_{\omega \in \mathbb{R}} |\Upsilon_{(i,k)}(j\omega)|. \tag{6.33}$$

Note that $\bar{\Upsilon}$ thus upper bounds the mismatch in *all* FRFs.

Theorem 6.5. *Consider the full-order model with moment $C_2\pi(\tau)$ and the signal generator (6.12) with state $\tau(t) \in \mathbb{R}^\nu$. Suppose Assumptions 6.3 and 6.4 hold for some $\gamma = \gamma^\star$. Furthermore, consider the reduced-order model with moment $H_2p(\tau)$ following from application of Theorem 6.3 and the optimization problem (6.26). Then, the following mismatch between the moments $C_2\pi(\tau)$ and $H_2p(\tau)$ holds:*

$$\|C_2\pi(\tau) - H_2p(\tau)\|_{L_2} \leq \bar{\Upsilon}\bar{\gamma}\|L\tau\|_{L_2} \tag{6.34}$$

with

$$\bar{\gamma} \coloneqq \left(1 + \frac{\gamma_{\rho u}\gamma^\star}{1 - \gamma_{\rho\varphi}\gamma^\star}\right)\left(1 + \frac{\gamma_{z\varphi}\gamma^\star}{1 - \gamma_{y\varphi}\gamma^\star}\right)$$

and $\bar{\Upsilon}$ as in (6.33) and the constants:

$$\gamma_{\rho u} \coloneqq \sup_{\omega\in\mathbb{R}} |\Gamma_{(1,1)}(j\omega)|, \qquad \gamma_{\rho\varphi} \coloneqq \sup_{\omega\in\mathbb{R}} |\Gamma_{(1,2)}(j\omega)|,$$
$$\gamma_{z\varphi} \coloneqq \sup_{\omega\in\mathbb{R}} |\Phi_{(2,2)}(j\omega)|, \qquad \gamma_{y\varphi} \coloneqq \sup_{\omega\in\mathbb{R}} |\Phi_{(1,2)}(j\omega)|.$$

Thereby, Problem 6.1 is solved for $\varepsilon \coloneqq \bar{\Upsilon}\bar{\gamma}$. Moreover, for any $u \in L_2(T)$, the following upper bound on the steady-state output mismatch holds:

$$\left\|\bar{z}_u - \bar{\zeta}_u\right\|_{L_2} \leq \bar{\Upsilon}\bar{\gamma}\|u\|_{L_2} =: \mathcal{V}_u. \tag{6.35}$$

▲

Proof. The proof can be found in Appendix E.3. □

The error bounds in Theorem 6.5 gives the essential insight that a small mismatch in the FRFs of the LTI blocks results in a small error since the bounds in (6.34) and (6.35) are linear in $\bar{\Upsilon}$. The optimization problem in (6.18a) aims at minimizing $|\Upsilon_{(i,k)}(j\omega)|^2$ on a discrete frequency grid for ω, for all $i, k \in \{1, 2\}$, thereby, also helping to reduce $\bar{\Upsilon}$ in (6.33). The formulated bounds in Theorem 6.5 are likely to be conservative due to (i) the stability conditions in Theorem 6.1 being conservative; (ii) the constant $\bar{\Upsilon}$ holding for any input $u(t) \in L_2(T)$; (iii) usage of the worst-case approximation in the derivation of $\gamma^\star$; and (iv) usage of the triangular inequality, which is conservative by its definition. Below, in Remark 6.7, comments are given on how to obtain tighter bounds. Despite some conservatism, the bounds give the valuable insight that the mismatch in time-domain signals remains bounded for arbitrary input functions selected from the space $L_2(T)$. In a more generic setting without a (global) stability assumption, bounded errors cannot be guaranteed for reduced-order nonlinear models obtained by reduction methods that do not impose a stability property on the reduced-order models, including the time-domain moment matching methods in [16, 225, 226].

Remark 6.7. The constant $\bar{\Upsilon}$ can be made smaller by only evaluating the supremum in (6.33) at the excited frequencies in the corresponding FRF. The constant $\bar{\gamma}$ in (6.34) and (6.35) can be made smaller by redefining the constants $\gamma_{\rho u}, \ldots, \gamma_{y\varphi}$ in Theorem 6.5 such that the supremum is not taken over $\omega \in \mathbb{R}$, but only at the excited frequencies in the corresponding FRF. △

6.3.4 Overview of the reduction method

An overview of the reduction method is presented in Algorithm 6.1 below.

Algorithm 6.1 Model order reduction algorithm

Input: The full-order Lur'e-type model Σ in (6.7), the constant $\gamma^\star$ as in Assumption 6.4 and the sets of frequencies $\Omega^0_{(i,k)}, \Omega^M_{(i,k)}, i, k \in \{1, 2\}$.

1: Define the signal generators (S_i, L_i), $i = 1, 2$, in (6.12) such that $\sigma(S_i) \simeq \Omega^i_0$ and Assumption 6.3 holds for each pair.
2: For each $i, k \in \{1, 2\}$, compute the matrices $C\Pi_{(i,k)}$ from (6.21) with $\mathcal{A} = A, \mathcal{B} = B_i$, and $\mathcal{C} = C_k$.
3: Define the reduced-order model matrices $F(\theta), G_1(\theta), G_2(\theta), H_1$, and H_2, as in (6.24).
4: Compute initial θ° using Theorem 6.4.
5: Using $\Omega^M_{(i,k)}, i, k \in \{1, 2\}$, and θ°, solve the constrained optimization problem (6.26) to find $\hat{\theta}$.

Output: The reduced-order model Σ_r in (6.13) with model matrices $F(\hat{\theta}), G_1(\hat{\theta}), G_2(\hat{\theta}), H_1, H_2$ as in (6.24).

6.4 Case study

6.4.1 Model of a flexible beam

In this case study, a one-sided clamped flexible beam supported by a one-sided spring is considered, see Figure 6.3 for a schematic depiction. The beam has dimensions length × width × height = 2 m × 50 mm × 30 mm and is characterized by its Young's modulus of 200 GPa and density of 7746 kg/m^3. The linear beam dynamics, characterized by partial differential equations, are discretized by the finite-element method on an equidistantly spatial grid of 40 points, leaving a state dimension of $n = 160$ consisting of position coordinates, velocity coordinates, inclination coordinates, and angular velocity coordinates. The beam deflection at the end of the beam is considered as the model output z, see Figure 6.3. Furthermore, the output y is considered as the deflection of the beam at the location of the one-sided spring. Moreover, an external disturbance $u(t)$ acts in the middle of the beam. The Bode magnitude diagram of the LTI part of the full-order LTI model

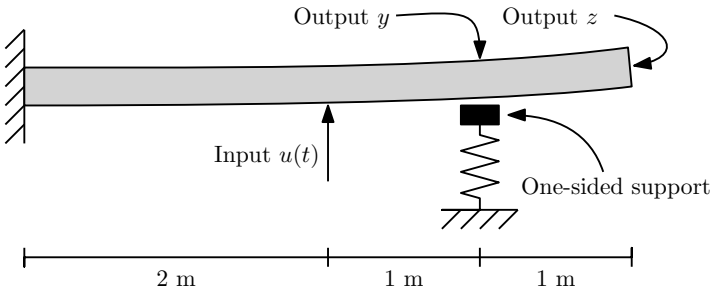


Figure 6.3. One-sided clamped flexible beam supported by a one-sided spring.

is depicted in the solid blue curve in Figure 6.4. The one-sided spring is modeled as

$$\varphi(y) = \gamma \max(0, y) \tag{6.36}$$

with stiffness $\gamma = 7.3{\cdot}10^5$, which is also the maximum slope of the nonlinearity. The model is cast in the Lur'e-type form of (6.7) and is denoted by Σ. The conditions of Theorem 6.1 are satisfied for the aforementioned γ, which guarantees that the full-order model is convergent.

The goal of this example is to use the reduction method described in Algorithm 6.1 to find a reduced-order Lur'e-type model that approximates the moment of the full-order Lur'e-type model. Furthermore, the method in [226] for generic nonlinear models is also applied, which does not enforce any type of model stability. The two models allow comparing the steady-state mismatch between the model responses for different disturbance situations.

6.4.2 Moment matching for generic nonlinear models

In this subsection, we demonstrate the application of the method in [226]. Consider a linear harmonic oscillator $s(\tau) = S\tau, S \in \mathbb{R}^{\nu\times\nu}, \sigma(S) \in \mathbb{C}^0$, and $l(\tau) = L\tau, L \in \mathbb{R}^{1\times\nu}$. This linear signal generator in combination with a function $\delta(\xi) = G \in \mathbb{R}^\nu$ (i.e., independent of ξ) results in the reduced-order dynamics of Wiener form:

$$\begin{aligned} \dot{\xi} &= (S - GL)\xi + Gu, \\ \psi &= h(\pi(\xi)), \end{aligned} \tag{6.37}$$

where the ξ-dynamics are linear and the model output ψ is a static nonlinear map of the state, see [226]. Under the assumption that the pair (S, L) is observable, the eigenvalues of $(S - GL)$ can be placed at any desired location in $\mathbb{C}^-$ by a suitable $G \in \mathbb{R}^\nu$. The eigenvalues of $(S - GL)$ being located in $\mathbb{C}^-$ guarantees that model (6.37) is convergent.

The signal generator is characterized by the following pair (S, L):

$$S = \pi \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 20.4 & 0 & 0 & 0 & 0 \\ 0 & -20.4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 125.2 & 0 & 0 \\ 0 & 0 & 0 & -125.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 360 \\ 0 & 0 & 0 & 0 & 0 & -360 & 0 \end{bmatrix}, \qquad (6.38)$$
$$L = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix},$$

and the initial condition $\tau(0) = L^\top$. This pair of matrices (S, L) satisfy the conditions of Assumption 6.1 and corresponds to the following interpolation points $s = 2\pi \cdot \{0, \pm 10.2j, \pm 62.6j, \pm 180j\}$, yielding a state-dimension of 7 for the reduced-order model. The numerical values for G are found by a pole placement procedure that places the poles of the reduced-order dynamics in (6.37) to a subset of poles of the LTI part of the full-order Lur'e-type model. It is non-trivial to compute the mapping $h \circ \pi$ analytically. Therefore, inspired from [226], steady-state data is generated on an equidistantly time grid by simulation of the full-order model by means of the MTF algorithm, see [196]. This data is subsequently used to estimate a mapping between τ and $\bar{z}_u$ using linear least-squares regression, where the estimated mapping is a second-order polynomial in the elements of τ.

The reduced-order model reads as (6.37) with $h \circ \pi$ replaced by $\widehat{h \circ \pi}$. In the remainder of this section, model (6.37) is called the reduced-order Wiener model. The performance of the reduced-order Wiener model is presented in Section 6.4.4 below.

6.4.3 Moment matching for Lur'e-type models

In this subsection, the application of Algorithm 6.1 for Lur'e-type models is demonstrated. The following interpolation points for the reduction of the LTI part are selected:

$$\begin{aligned} \Omega_0^{(1,1)} &= \begin{bmatrix} 0 & 10.2 & 62.6 & 180 \end{bmatrix}, \\ \Omega_0^{(1,2)} &= \begin{bmatrix} 0 & 10.2 & 64.1 & 180 \end{bmatrix}, \\ \Omega_0^{(2,1)} &= \begin{bmatrix} 0 & 10.2 & 65.7 & 180 \end{bmatrix}, \\ \Omega_0^{(2,2)} &= \begin{bmatrix} 0 & 10.2 & 64.1 & 180 \end{bmatrix}. \end{aligned} \qquad (6.39)$$

The selected frequencies imply a match of the respective FRF at 0-Hz and at the first three largest resonance peaks, see Figure 6.4 where the interpolation frequencies are marked by crosses. Subsequently, the matrices $(S^{(i,k)}, L^{(i,k)}), i, k \in$

$\{1, 2\}$, are defined as follows:

$$\begin{aligned} S^{(i,k)} &\coloneqq \text{blockdiag}\left(0, \Xi_2^{(i,k)}, \ldots, \Xi_{\eta^{(i,k)}}^{(i,k)}\right), \\ \Xi_\ell^{(i,k)} &\coloneqq \begin{bmatrix} 0 & \omega_\ell^{(i,k)} \cdot 2\pi \\ -\omega_\ell^{(i,k)} \cdot 2\pi & 0 \end{bmatrix}, \\ L^{(i,k)} &\coloneqq \begin{bmatrix} 1 & 1 & 0 & \ldots & 1 & 0 \end{bmatrix} \in \mathbb{R}^{\nu^{(i,k)}}, \end{aligned} \tag{6.40}$$

where $\omega_k^{(i,k)}$ is the k-th element of $\Omega_0^{(i,k)} \in \mathbb{R}^{\eta^{(i,k)}}$. Assumption 6.3 holds for all pairs and $\sigma(S^{(i,k)}) \simeq \Omega_0^{(i,k)}, i, k \in \{1, 2\}$. Based on the number of interpolation points, it can be concluded that the state dimension of the reduced-order model is 28. Note that the state dimension of the reduced-order Lur'e-type model is four times larger than the state dimension of the reduced-order Wiener model presented in Section 6.4.2.

The LMIs in Theorem 6.4 are solved to obtain θ°, that, together with (6.24), characterize the initial convergent reduced-order Lur'e-type model Σ_r° in (6.13). The Bode magnitude diagram of Σ_r° is shown in Figure 6.4 in the dotted yellow curve. It can be observed that the respective FRFs of the LTI part of Σ and Σ_r° match at the corresponding interpolation points in (6.39), i.e., constraint (6.18b) is satisfied. However, for other frequencies, there is a significant mismatch.

Next, the constrained optimization problem (6.26) is solved, starting at θ°. Hereto, the sets $\Omega_M^{(i,k)} = 2\pi \cdot 10^\kappa =: \Omega_M$ are defined with $\kappa \in \mathbb{R}^M$ linearly spaced with $M = 1000$ elements between -2 and 5, implying that elements of the set Ω_M are logarithmically spaced between 0.01 Hz and 10 kHz. The resulting $\hat{\theta}$ defines the (final) convergent reduced-order Lur'e-type model Σ_r in (6.13). The Bode magnitude plot of the LTI part of Σ_r is depicted in the dashed red curve in Figure 6.4. With respect to Σ_r°, a significant improvement of the fit of Σ_r to Σ can be observed at almost all frequencies. Furthermore, it can be concluded that an accurate match is obtained up to the frequency corresponding to the third-largest resonance peak. After that frequency, the FRF of the reduced-order model does not match any of the (anti-) resonance peaks.

6.4.4 Performance of reduced-order models

By means of simulations, the quality of the reduced-order Wiener and Lur'e-type model in terms of approximating the steady-state response of the full-order Lur'e-type model is analyzed. Throughout the rest of this section, $\bar{z}$ is called the steady-state output of the full-order model, $\bar{\zeta}_{Wiener}$ is the steady-state output of the reduced-order Wiener model and $\bar{\zeta}_{Lur'e}$ is the steady-state output of the (final) reduced-order Lur'e-type model.

The steady-state output $\bar{z}$ of the interconnected full-order model (6.3) with (S, L) in (6.38) and $\tau(0) = L^\top$ is depicted in Figure 6.5, together with the steady-

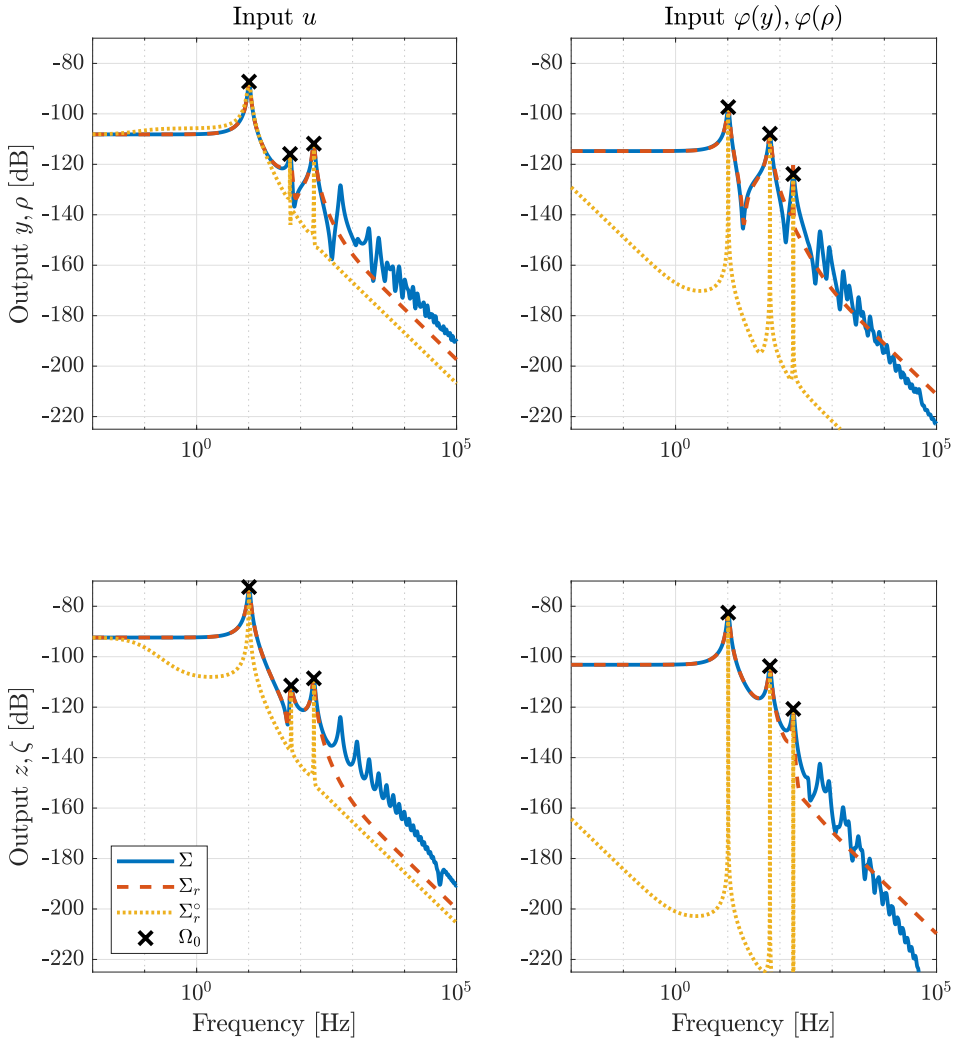


Figure 6.4. Bode magnitude diagram of the LTI parts of the Lur'e-type models Σ (full-order model - solid blue), Σ_r (final reduced-order model - dashed red) and Σ_r° (initial reduced-order model - dotted yellow). The interpolation points $\Omega_0^{(i,k)}, i, k \in \{1, 2\}$ in (6.39) are marked by a cross (the interpolation point at 0 Hz is not visible).

state outputs $\bar{\zeta}_{Wiener}$ and $\bar{\zeta}_{Lur'e}$ of the reduced-order Wiener and Lur'e-type models. It can be observed that the beam deflection is in the order of 100 μm, whereas the approximation error is roughly two orders of magnitude smaller. This concludes that both reduced-order models accurately approximate $\bar{z}$. The same conclusion can be drawn from the column 'training' of Table 6.1, where it can be seen that the approximation error in the L_2-norm is roughly three orders of magnitude smaller than the norm of $\bar{z}$. The best approximation is obtained by the reduced-order Lur'e-type model.

To further investigate the accuracy of the reduced-order models, all models are subject to a test with the following two new inputs:

$$u_1(t) = \sum_{k=1}^{10} \sin\bigl(2\pi k f_0^1 t\bigr), \tag{6.41a}$$

$$u_2(t) = \sum_{k=1}^{10} \sin\bigl(2\pi k f_0^2 t\bigr), \tag{6.41b}$$

where $f_0^1 = 10$ Hz is the base frequency of u_1 and $f_0^2 = 100$ Hz is the base frequency of u_2. The input u_1 excites the beam up to the frequency of 100 Hz, for which an accurate match between the FRFs of the LTI part of the full-order and reduced-order Lur'e-type model is obtained. The input u_2 excites the beam up to the frequency of 1 kHz, which also excites frequencies for which the match in FRF is not accurate. It can be observed in Figure 6.6 that for u_1, the accuracy of the reduced-order Wiener model deteriorated significantly. The accuracy of $\bar{\zeta}_{Lur'e}$, on the other hand, is comparable to the previous test. At this point, it can already be concluded that the reduced-order Lur'e-type model approximates the steady-state output well as long as the input signal excites roughly the same frequencies as those matched in the FRFs. For input u_2, however, the accuracy of the reduced-order Wiener model has a significant mismatch in the 0-Hz component, see in Figure 6.7. The reduced-order Lur'e-type model, however, approximates the steady-state response of the full-order model reasonably well, even for input signals that contain frequencies for which not a good match is obtained in the FRF. The mismatch is quantified in Table 6.1, from which the same conclusions can be drawn.

Table 6.1 includes the upper bound $\mathcal{V}_u$ in (6.35) for the mismatch $\bigl\|\bar{z} - \bar{\zeta}_{Lur'e}\bigr\|_{L_2}$. It can be concluded that this bound is satisfied in all cases and that the bound is extremely conservative for this example.

6.4.5 Discussion

The case study above clearly highlights the benefits of the reduction approach; namely, the reduced-order model preserves the convergence property and provides robustness to input variations. The steady-state mismatch between the response of the full-order and the reduced-order Lur'e-type models depends on the quality of the fit in the FRFs. In particular, if the FRFs are accurately fitted over a

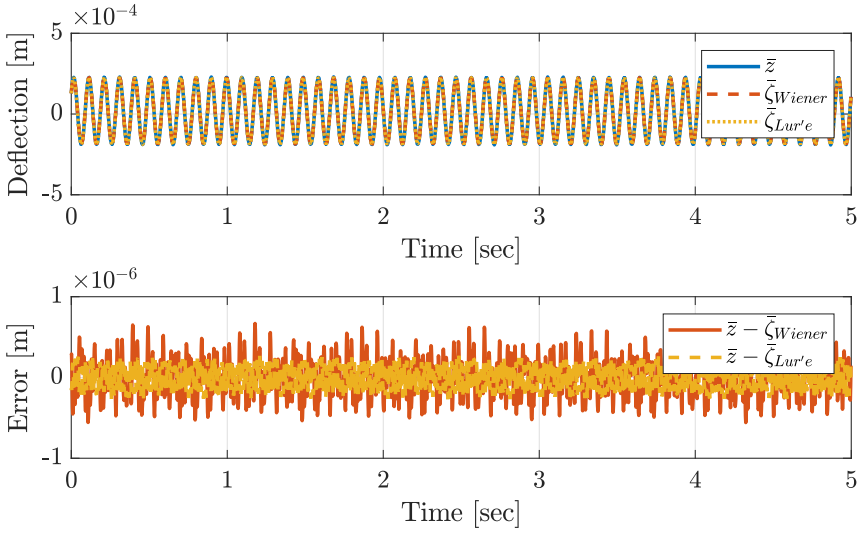


Figure 6.5. Top: steady-state outputs $\bar{z}, \bar{\zeta}_{Wiener}$, and $\bar{\zeta}_{Lur'e}$ subject to the input generated by linear harmonic oscillator characterized by (S, L) in (6.38). Bottom: mismatch between the steady-state outputs $\bar{z}$ and $\bar{\zeta}_{Wiener}$ and between outputs $\bar{z}$ and $\bar{\zeta}_{Lur'e}$.

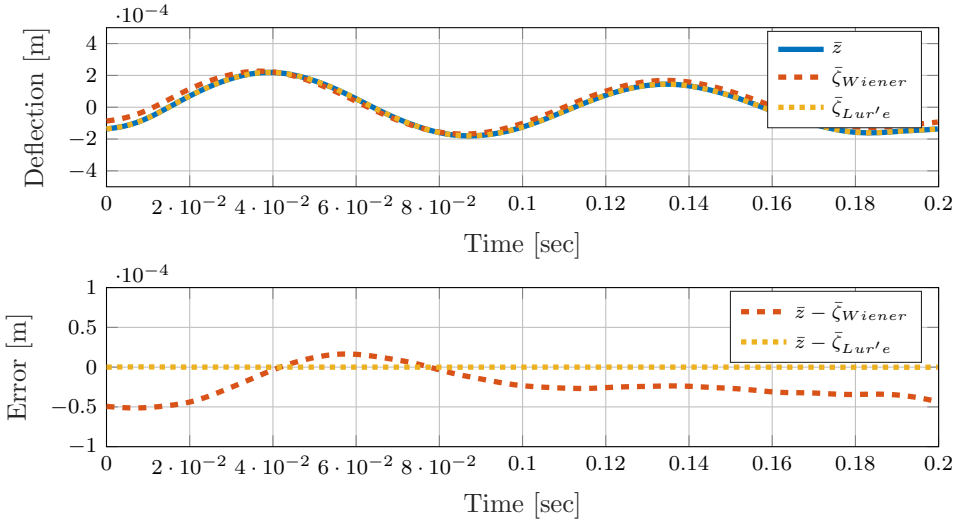


Figure 6.6. Top: steady-state outputs $\bar{z}, \bar{\zeta}_{Wiener}$, and $\bar{\zeta}_{Lur'e}$ subject to input u_1 in (6.41a). Bottom: mismatch between the steady-state outputs $\bar{z}$ and $\bar{\zeta}_{Wiener}$ and between outputs $\bar{z}$ and $\bar{\zeta}_{Lur'e}$.

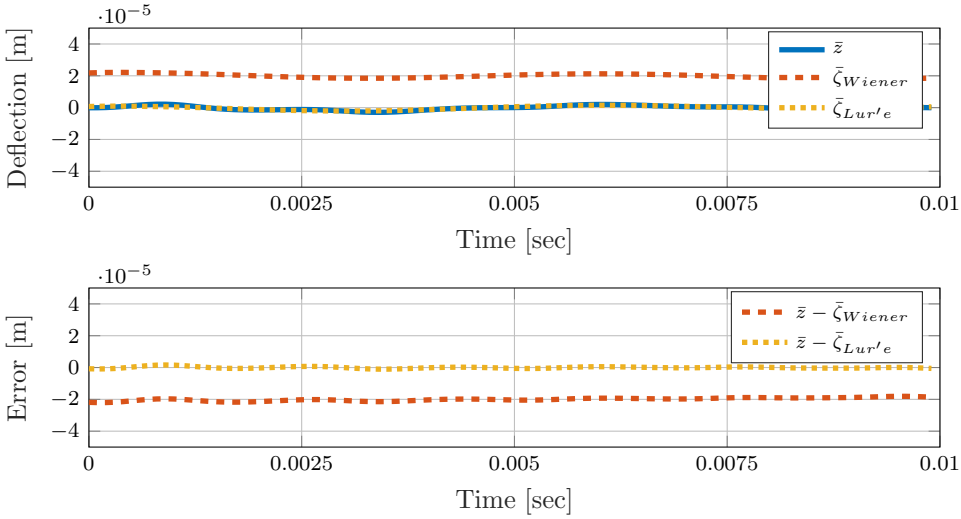


Figure 6.7. Top: steady-state outputs $\bar{z}, \bar{\zeta}_{Wiener}$, and $\bar{\zeta}_{Lur'e}$ subject to input u_2 in (6.41b). Bottom: mismatch between the steady-state outputs $\bar{z}$ and $\bar{\zeta}_{Wiener}$ and between outputs $\bar{z}$ and $\bar{\zeta}_{Lur'e}$.

Table 6.1. Quantitative performance analysis.

	Training	Validation 1 $u_1(t)$	Validation 2 $u_2(t)$
$\|\bar{z}\|_{L_2}$	$1.46 \cdot 10^{-4}$	$1.27 \cdot 10^{-4}$	$1.30 \cdot 10^{-6}$
$\|\bar{z} - \bar{\zeta}_{Wiener}\|_{L_2}$	$2.32 \cdot 10^{-7}$	$2.85 \cdot 10^{-5}$	$2.00 \cdot 10^{-5}$
$\|\bar{z} - \bar{\zeta}_{Lur'e}\|_{L_2}$	$1.36 \cdot 10^{-7}$	$8.47 \cdot 10^{-8}$	$5.29 \cdot 10^{-7}$
$\mathcal{V}_u$	0.30	0.43	0.43

large frequency range, then the reduced-order Lur'e-type models generally approximate the steady-state output accurately, even though only approximate moment matching is achieved (rather than moment matching). Such frequency-domain insights are valuable as in many engineering applications, frequency-domain tools are used for analysis and control design. The model robustness against input variations is generally lacking in other moment matching methods, e.g., the ones in [16, 225, 226]. However, as explained in Remark 6.5, the state dimension of the reduced-order model is typically larger than reduced-order models obtained via methods [16, 225, 226].

6.5 Conclusions

This chapter has extended the notion of moments of nonlinear models from the local case to the global case. Subsequently, this notion is exploited to propose a model order reduction technique by time-domain moment matching for Lur'e-type nonlinear models that enjoy the convergence property. The reduction method approximates the moment of the nonlinear model rather than matching it exactly. Preservation of the convergence property of the full-order model guarantees that the reduced-order model exhibits a bounded and asymptotically stable steady-state response for any bounded input and provides robustness against input variations. Furthermore, it allows deriving a bound on the mismatch between the moment of the full-order and the reduced-order models. Moreover, the Lur'e-type structure of the model is preserved during reduction and the characteristic frequency-domain interpretation of moment matching is inherited. In a numerical case study on a one-sided supported beam, the moment of the reduced-order model matches accurately to the moment of the full-order model. Furthermore, the reduced-order model also accurately captures the steady-state response to generalized inputs.

7

Optimal model reduction by moment matching: the linear and nonlinear feedback cases

This chapter considers the problem of model complexity reduction for both linear models and nonlinear feedback models. We propose an optimal variant of the time-domain moment matching method in which the $\mathcal{H}_\infty$ norm of the error transfer function matrix is minimized. In the linear case, it optimizes the match between the transfer function matrix of the original and reduced-order models. These results are extended for a class of nonlinear feedback models, consisting of linear time-invariant (LTI) dynamics placed in feedback with nonlinear dynamics, where only the order of the LTI dynamics is reduced. In this nonlinear case, the proposed approach aims at minimizing an error bound on the steady-state outputs for any bounded input. Furthermore, the proposed approach preserves both the nonlinear feedback model structure and global stability properties of the full-order nonlinear model. Both the problems for the LTI and the nonlinear cases are formulated as constrained optimization problems with bilinear constraints, solved with a numerical procedure. The effectiveness of the approach is illustrated in the reduction of a structural dynamics model of a linear beam subject to nonlinear support.

The contents of this chapter are published in: Shakib, M. F., Scarciotti, G., Jungers, M., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. Optimal model reduction by moment matching: the linear and nonlinear feedback cases. *Submitted for journal publication.* Preliminary results have been presented in: Shakib, M. F., Scarciotti, G., Jungers, M., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. (2021, December). Optimal $\mathcal{H}_\infty$ LMI-based model reduction by moment matching for linear time-invariant models. In *Proceedings of the Conference for Decision and Control*, (pp. 6914-6919).

7.1 Introduction

Dynamic models enable the prediction of the system response, analysis of the system's dynamic behavior, and controller design. Given the complexity of engineering systems, accurate models are often too complex for the aforementioned purposes. Model order reduction aims at replacing the full-order model with a less complex, reduced-order model. Hereto, it is important to find the reduced-order model that accurately describes the dynamics and that preserves certain properties of the full-order models of which model stability is among the most crucial ones.

Many techniques have been proposed in the literature for linear time-invariant (LTI) models, such as balanced truncation [10], Hankel-norm approximations [98] and the interpolation approach [91]. Optimality has been pursued in the $\mathcal{H}_\infty$ and $\mathcal{H}_2$ norm of the error transfer function matrix. The optimal $\mathcal{H}_\infty$ reduction problem has been posed as a constrained nonlinear optimization problem with bilinear matrix inequalities (BMIs), see [72, 95, 111], following from the Bounded Real Lemma [228]. By fixing certain variables, [95] avoided BMI constraints and wrote an approximate problem using linear matrix inequality (LMI) constraints only. Furthermore, [72] points out the connection between this problem and Hankel-norm approximation. Optimal $\mathcal{H}_2$ reduction problems have also been studied for a long time, pointing out necessary conditions for optimality, see, e.g., [7, 118, 169, 306], developing sophisticated numerical algorithms [21, 44, 88, 113, 150, 310], and providing extensions, e.g., using frequency weighting [106, 317], or the connection to the Sylvester equation [289]. One of the most popular approaches based on the moment matching technique is reported in [100], which uses first-order optimality conditions to perform moment matching until a locally optimal model is found, which is extended in [9] to include a frequency weighting and in [49] to converge faster in exchange for surrogate optimization. In this chapter, we also focus on moment matching techniques.

Reduced-order models constructed by the moment matching method match the *moments* of the reduced-order model to the *moments* of the full-order model, see [10]. For single-input, single-output (SISO) LTI models, moment matching enforces a match in the transfer function at the interpolation points. For multiple-input, multiple-output (MIMO) LTI models, a match in the transfer function matrix along so-called tangential directions is achieved [92]. Unfortunately, in the MIMO case, the match in the individual transfer functions of the transfer function matrix can be poor. Although such MIMO reduced-order models do achieve moment matching along tangential directions, in some cases, for example in the nonlinear feedback case discussed below, an accurate match in each individual transfer function is desired.

The *time-domain* moment matching technique constructs a family of reduced-order models that are parameterized by a small number of parameters, see [16]. These parameters provide freedom to enforce desired model properties, such as

prescribed pole locations; prescribed zero locations; and an L_2-gain constraint [16]. Alternatively, the freedom can be exploited to optimally fit transient model responses [223] or achieve two-sided moment matching [120]. Furthermore, the approaches in [177, 178] exploit the parametric freedom of time-domain moment matching to achieve optimality in the $\mathcal{H}_2$-norm of the error transfer function matrix. Time-domain moment matching extends naturally from LTI models to nonlinear models with a well-defined notion of moments consistent with its LTI counterpart [226]. However, optimality has not been pursued yet in the nonlinear case.

In the first part of this chapter, the parametric freedom in the time-domain moment matching technique for multivariable LTI models is exploited such that optimality in the $\mathcal{H}_\infty$-norm of the error transfer function matrix is achieved. As a result, the peak error in the transfer function matrix is minimized, resulting also in a small error in all individual elements of the transfer functions matrix. The problem is formulated as a constrained optimization problem with bilinear matrix inequality constraints. To numerically solve this problem, a tailored coordinate descent algorithm (CDA) [253] is proposed. Potentially, this CDA prevents from getting stuck prematurely and provides numerical robustness, as is demonstrated in a numerical example.

In the second part of this chapter, the results of the first part on LTI models are extended to the case of nonlinear feedback models consisting of high-order LTI dynamics placed in feedback with low-order nonlinear dynamics, see Figure 7.1. The proposed method preserves the model structure by only reducing the order of the LTI dynamics and inherits the nonlinear dynamics from the full-order model. Unfortunately, this approach results in a mismatch between the moments of the full-order and the reduced-order nonlinear models. However, under a global stability assumption, namely input-to-state convergence [200], an upper bound for this mismatch can be derived, which is in fact more generic as it bounds the $\mathcal{L}_\infty$-signal norm of the mismatch between the steady-state responses of the full-order and the reduced-order nonlinear models for any bounded input signal. Crucially, this upper bound is zero if the $\mathcal{H}_\infty$-norm of the error transfer function matrix of the LTI part is zero, implying that a small $\mathcal{H}_\infty$-norm of the error transfer function matrix of the LTI part results in a small error bound. Therefore, it is exploited to formulate a reduction problem that aims at minimizing the $\mathcal{H}_\infty$-norm of the error transfer function matrix of the LTI part, in addition to preserving the global stability property for the reduced-order nonlinear model.

The literature on *optimal* model reduction in the nonlinear case is rather scarce. The data-based approaches in [55, 149] are an exception, as these provide (local) error bounds, which are subsequently minimized. Two of the challenges in achieving optimality are stability preservation for the reduced-order model and error characterization between the full-order and the reduced-order nonlinear model. To this extent, the approaches in [129, 131, 132, 222] use the concepts of generalized and extended differential balancing to truncate less important states of a

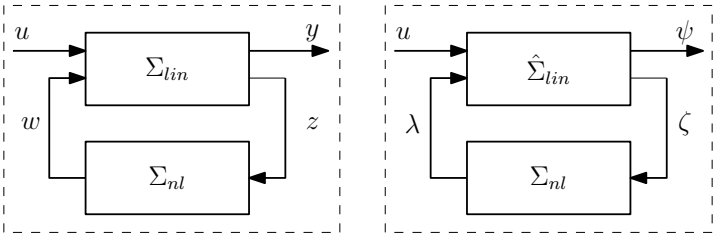


Figure 7.1. Full-order (left) and reduced-order (right) nonlinear feedback model, both consisting of LTI dynamics placed in feedback with nonlinear dynamics. Only the order of the LTI dynamics is reduced.

balanced realization of the full model. These methods preserve the *contraction* stability property and are equipped with error bounds. The balancing methods in [31, 32] study the same class of nonlinear feedback models as in this chapter, preserve model stability, and are equipped with an error bound. However, minimization of the error bound has not been pursued in [31, 32, 129, 131, 132, 222]. In this chapter, the error bound in [31] is minimized to achieve optimality in the context of moment matching. Data-based methods that enforce stability and pursue optimality for *one specific input* are presented in, e.g, [1, 38, 280], also interpreted as *system identification* methods. These methods result in black-box models that do not preserve the model structure and neither are these equipped with error bounds.

To summarize, the main contributions of this chapter are, firstly, a model order reduction technique for multivariable LTI models that matches the moments of the full-order model, preserves model stability, and minimizes the approximation error in the $\mathcal{H}_\infty$-norm and, secondly, a structure-preserving model order reduction technique for multivariable nonlinear feedback models that is stability preserving and that minimizes an upper bound on the error between the steady-state response of the full-order and the reduced-order nonlinear model. Thirdly, a numerical procedure is presented to solve these model reduction problems, which is of independent interest as it can be applied with minor adaptions to a wide variety of optimization problems with BMI constraints. Finally, the effectiveness of the proposed methods for the LTI and nonlinear feedback cases is demonstrated in an example on an actuated beam with nonlinear support.

The remainder of this chapter is structured as follows. Section 7.2 proposes an approach to model order reduction for LTI models, which is extended to the nonlinear feedback case in Section 7.3. Section 7.4 presents a numerical procedure to solve these model order reduction problems. Section 7.5 describes the results of two numerical case studies. Section 7.6 gives the concluding remarks.

Notation: The symbols $\mathbb{R}, \mathbb{C}, \mathbb{C}^0, \mathbb{C}^-$, and $\mathbb{Z}^+$ denote the set of real numbers, complex numbers, complex numbers with zero real part, complex numbers with

negative real part, and positive integers, respectively. The symbol I_n denotes the $n \times n$ identity matrix and the symbol $0_{n\times\nu}$ the zero matrix of dimensions $n \times \nu$. The spectrum of a matrix $A \in \mathbb{R}^{n\times n}$ is denoted by $\sigma(A)$. For a matrix $A \in \mathbb{R}^{n\times n}$, the notation $\mathrm{He}\,(A)$ is short-hand for $A^\top + A$. A symmetric matrix A is called positive (negative) definite, denoted by $A \succ 0$ ($A \prec 0$), if all its eigenvalues are strictly positive (negative). In symmetric block matrices, the symbol $\star$ represents an element that is induced by transposition. The set of symmetric positive definite matrices of dimension $n \times n$ is denoted by $\mathbb{S}_n$. For a vector x, its Euclidean norm is denoted by $|x|$. For a signal x defined on $\mathbb{R}$, the $\mathcal{L}_\infty$ signal norm is denoted by $\|x\|_\infty$ and defined by $\|x\|_\infty := \sup_{\tau\in\mathbb{R}} |x(\tau)|$. Denote by $\mathcal{L}_\infty^m$ the class of functions $u(t) \in \mathbb{R}^m$ that are defined on $\mathbb{R}$ and bounded by $\|u\|_\infty < +\infty$. Functions of class $\mathcal{K}, \mathcal{K}_\infty$, and $\mathcal{KL}$ are defined in accordance with [135]. The identity function id satisfies $\mathrm{id}(s) = s$, $\forall s \in \mathbb{R}$. Consider an LTI model with m inputs, p outputs, and n poles, and denote its transfer function matrix by $\Upsilon(s) \in \mathbb{C}^{p\times m}$ with $s \in \mathbb{C}$. Suppose all its n poles are located in $\mathbb{C}^-$. The $\mathcal{H}_\infty$ norm of Υ is denoted by $\|\Upsilon(s)\|_\infty$ and defined according to [228]: $\|\Upsilon(s)\|_\infty := \sup_{\omega\in[0,\infty)} \bar{\sigma}(\Upsilon(j\omega))$ with $\bar{\sigma}$ the maximum singular value and $j := \sqrt{-1}$ the imaginary unit. Its *peak-to-peak* or $\mathcal{L}_\infty$-induced norm is denoted by $\|\Upsilon(s)\|_1$ and defined according to:

$$\|\Upsilon(s)\|_1 := \sup_{0<\|u\|_\infty<\infty} \frac{\|y\|_\infty}{\|u\|_\infty}. \tag{7.1}$$

Furthermore, the following relation between $\|\Upsilon(s)\|_\infty$ and $\|\Upsilon(s)\|_1$ holds, see [12]:

$$\frac{1}{\sqrt{p}}\|\Upsilon(s)\|_\infty \le \|\Upsilon(s)\|_1 \le (2n+1)\sqrt{m}\|\Upsilon(s)\|_\infty. \tag{7.2}$$

7.2 Optimal model reduction for multivariable LTI models

The considered reduction problem aims at achieving (i) moment matching along tangential directions and (ii) a minimal mismatch in terms of the infinity norm of the transfer function matrix of the error dynamics. This problem is formalized in Section 7.2.1 and a solution is proposed in Section 7.2.2.

7.2.1 The optimal moment matching problem

Consider the class of LTI models described by the following state-space equations:

$$\Sigma(A, B, C): \quad \dot{x} = Ax + Bu, \quad y = Cx, \tag{7.3}$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the input, $y(t) \in \mathbb{R}^p$ is the output, and $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{p\times n}$ are the model matrices. Minimality of model

(7.3) is assumed throughout this chapter. The associated transfer function matrix of model (7.3) is defined as follows:

$$\Phi(s) \coloneqq C(sI_n - A)^{-1}B, \quad s \in \mathbb{C}. \tag{7.4}$$

Based on the transfer function matrix Φ in (7.4), the notion of moments along tangential directions is defined next.

Definition 7.1. *Let $s^\star \in \mathbb{C} \setminus \sigma(A)$. The 0-moment of model (7.3) at $s^\star$ along the right-tangential direction $\ell^\star \in \mathbb{C}^m$ is $\eta_0^\Sigma(s^\star, \ell^\star) \coloneqq \Phi(s^\star)\ell^\star$.* ■

Consider the ν-th-order (reduced-order) model characterized as follows:

$$\hat{\Sigma}(F, G, H): \quad \dot{\xi} = F\xi + Gu, \quad \psi = H\xi \tag{7.5}$$

with $\xi(t) \in \mathbb{R}^\nu, u(t) \in \mathbb{R}^m, \psi(t) \in \mathbb{R}^p$, and $F \in \mathbb{R}^{\nu\times\nu}, G \in \mathbb{R}^{\nu\times m}, H \in \mathbb{R}^{p\times\nu}$. The transfer function matrix of model (7.5) reads as follows:

$$\Gamma(s) \coloneqq H(sI_\nu - F)^{-1}G, \quad s \in \mathbb{C}, \tag{7.6}$$

and its moments at $s^\star \in \mathbb{C}$ along $\ell^\star \in \mathbb{C}^m$ are denoted by $\eta_0^{\hat{\Sigma}}(s^\star, \ell^\star)$. It is assumed throughout this chapter that $\nu < n$. The relation between models (7.5) and (7.3) is characterized next.

Definition 7.2. *Consider the collection of $\kappa \in \mathbb{Z}^+$ distinct interpolation points $s = \{s_i^\star \in \mathbb{C} \setminus (\sigma(A) \cap \sigma(F))\}_{i=1}^\kappa$ and tangential directions $\ell = \{\ell_i^\star \in \mathbb{C}^m\}_{i=1}^\kappa$. Model (7.5) is a reduced-order model of (7.3) at (s, ℓ) if its order ν satisfies $\nu < n$ and if it achieves moment matching according to $\eta_0^{\hat{\Sigma}}(s_i^\star, \ell_i^\star) = \eta_0^\Sigma(s_i^\star, \ell_i^\star)$ for $i = 1, \ldots, \kappa$.* ■

To enable a frequency-response function (FRF) interpretation of moment matching and to limit the complexity of the exposition, only interpolation points on the imaginary axis, real tangential directions, and 0-th moments along the right-tangential directions are considered. The results can be extended to interpolation points in the complex plane, complex tangential directions, higher-order moments, and left-tangential directions.

The problem considered in this section aims at minimizing the $\mathcal{H}_\infty$-norm error of the difference between Φ and Γ. This difference is denoted by Υ and defined as follows:

$$\Upsilon(s) \coloneqq \Phi(s) - \Gamma(s), \quad s \in \mathbb{C}. \tag{7.7}$$

The following stability assumption is used in the problem statement presented next.

Assumption 7.1. *The matrix A of the LTI model (7.3) is Hurwitz, i.e., $\sigma(A) \in \mathbb{C}^-$.* ■

Problem 7.1. *Consider the full-order model* (7.3)*, the reduced-order model* (7.5)*, $\kappa \in \mathbb{Z}^+$ distinct frequencies $\omega_i \geq 0, i = 1, \ldots, \kappa$, and the tangential directions $\ell_i \in \mathbb{R}^m, i = 1, \ldots, \kappa$. Suppose Assumption 7.1 holds. The problem of Optimal $\mathcal{H}_\infty$ model reduction by moment matching consists of solving the following constrained optimization problem:*

$$\min_{F,G,H,\gamma} \quad \gamma \tag{7.8a}$$

$$\text{subject to} \quad \|\Upsilon(s)\|_\infty < \gamma, \tag{7.8b}$$

$$\eta_0^{\Sigma}(\pm j\omega_i, \ell_i) = \eta_0^{\hat{\Sigma}}(\pm j\omega_i, \ell_i), \quad \textit{for } i = 1, \ldots, \kappa. \tag{7.8c}$$

△

Constraint (7.8c) ensures that (7.5) is a (reduced-order) model of (7.3) according to Definition 7.2. Constraint (7.8b) ensures that the $\mathcal{H}_\infty$-norm of the error transfer function matrix remains below γ, and also ensures stability preservation of the reduced-order model as unstable models have an infinitely large $\mathcal{H}_\infty$-norm. Furthermore, a small γ ensures a small $\|\Upsilon(s)\|_\infty$, implying an accurate match between all individual (scalar) transfer functions of the transfer function matrices, over all frequencies.

Remark 7.1. The minimum in Problem 7.1 can be interpreted as the infimum over all values of γ for which the set defined by (7.8b) - (7.8c) is feasible. A numerical solver for (7.8a) - (7.8c) should then find a sufficiently accurate approximation of that infimum. △

Remark 7.2. The interpolation frequencies can be chosen as the zero frequency, resonance frequencies, and other frequencies of interest. The tangential directions, however, are less intuitive to choose as a full characterization between the tangential directions and the optimal solution of Problem 7.1 is lacking. In other moment matching approaches, e.g., [100] and [177], the tangential directions are also part of the optimization variables. △

7.2.2 Parametrization of models satisfying the constraints of Problem 7.1

Reduced-order LTI models achieving moment matching

The set of models that satisfy the constraint (7.8c) of Problem 7.1 is presented in the following theorem. For this purpose, construct the matrices $S \in \mathbb{R}^{\nu \times \nu}$,

$L \in \mathbb{R}^{m\times\nu}$ with $\nu = 2\kappa$ as follows:

$$S \coloneqq \begin{bmatrix} 0 & \omega_1 & \cdots & 0 & 0 \\ -\omega_1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \omega_\kappa \\ 0 & 0 & \cdots & -\omega_\kappa & 0 \end{bmatrix} \tag{7.9a}$$

$$L \coloneqq \begin{bmatrix} \ell_1 & 0_m & \cdots & \ell_\kappa & 0_m \end{bmatrix}, \tag{7.9b}$$

where $\omega_i > 0, \ell_i, i = 1, \ldots, \kappa$, are given in Problem 7.1. The case where $\omega_i = 0$, for an $i \in \{1, \ldots, \kappa\}$, is treated in Remark 7.5 below.

Theorem 7.1. *Consider the full-order model* (7.3) *and the matrices* $S \in \mathbb{R}^{\nu\times\nu}$ *and* $L \in \mathbb{R}^{m\times\nu}$ *as in* (7.9)*. For any* $G \in \mathcal{G}_M$ *with*

$$\mathcal{G}_M \coloneqq \left\{ G \in \mathbb{R}^{\nu\times m} \,:\, \sigma(S) \cap \sigma(S - GL) = \emptyset \right\}, \tag{7.10}$$

the reduced-order model (7.5) *of order* ν *with matrices*

$$F \coloneqq S - GL, \quad H \coloneqq C\Pi, \tag{7.11}$$

where $\Pi \in \mathbb{R}^{n\times\nu}$ *is the unique solution to the equation*

$$\Pi S = A\Pi + BL, \tag{7.12}$$

achieves moment matching according to:

$$\eta_0^{\Sigma}(\pm j\omega_i, \ell_i) = \eta_0^{\hat{\Sigma}}(\pm j\omega_i, \ell_i), \quad i = 1 \ldots \kappa. \tag{7.13}$$

▲

Proof. The proof can be found in Appendix F.1. □

Under specific selection of the tangential directions, the set $\mathcal{G}_M$ presented in Theorem 7.1 contains all the models that satisfy constraint (7.8c), i.e., that achieve moment matching along the right-tangential directions, as stated in the next theorem.

Theorem 7.2. *Consider the matrices* $S \in \mathbb{R}^{\nu\times\nu}$ *and* $L \in \mathbb{R}^{m\times\nu}$ *as in* (7.9) *and suppose that all tangential directions coincide with* ℓ*, i.e.,* $\ell_i = \ell, i \in \{1, \ldots, \kappa\}$*. Furthermore, suppose that the pair* (S, L) *is observable. Then, for any strictly proper transfer function matrix* $W(s) \in \mathbb{C}^{p\times m}$ *that achieves moment matching according to*

$$W(\pm j\omega_i)\ell = \eta_0^{\Sigma}(\pm j\omega_i, \ell), \quad i = 1, \ldots, \kappa, \tag{7.14}$$

there exists a $G \in \mathcal{G}_M$ *such that*

$$W(s)\ell = \Gamma(s)\ell, \quad s \in \mathbb{C}, \tag{7.15}$$

with $\Gamma(s)$ *as in* (7.6). ▲

Proof. The proof can be found in Appendix F.2. □

The set $\mathcal{G}_M$ thus contains *all* reduced-order models that achieve (7.13) in the case where all tangential directions coincide and the pair (S, L) is observable. In the next section, a subset of $\mathcal{G}_M$ is characterized that, in addition to constraint (7.8c), also satisfies the constraint (7.8b) for a fixed γ.

Remark 7.3. Theorem 7.2 is a MIMO extension of [16, Proposition 1]. It is challenging to derive conditions under which a stronger property than (7.15) can be guaranteed. For example, an instrumental property would be the existence of $G \in \mathcal{G}_M$ such that $W(s) = \Gamma(s)$, for any $W(s)$ that achieves (7.14). △

Reduced-order LTI models satisfying all constraints of Problem 7.1

Consider the error dynamics between the original model and the reduced model described in the following form:

$$\begin{aligned} \begin{bmatrix} \dot{x} \\ \dot{\xi} \end{bmatrix} &= \underbrace{\begin{bmatrix} A & 0_{n\times\nu} \\ 0_{\nu\times n} & S - GL \end{bmatrix}}_{\mathcal{A}} \underbrace{\begin{bmatrix} x \\ \xi \end{bmatrix}}_{z} + \underbrace{\begin{bmatrix} B \\ G \end{bmatrix}}_{\mathcal{B}} u, \\ \zeta &= \underbrace{\begin{bmatrix} C & -C\Pi \end{bmatrix}}_{\mathcal{C}} z. \end{aligned} \tag{7.16}$$

Define, for matrices $\mathcal{X} \in \mathbb{S}^{n\times n}, \mathcal{A} \in \mathbb{R}^{n\times n}, \mathcal{B} \in \mathbb{R}^{n\times m}, \mathcal{C} \in \mathbb{R}^{p\times n}$, and constant $\gamma \geq 0$, the following matrix:

$$\mathcal{N}_\gamma(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{X}) \coloneqq \begin{bmatrix} \mathrm{He}\,(\mathcal{X}\mathcal{A}) & \mathcal{X}\mathcal{B} & \mathcal{C}^\top \\ \star & -\gamma I_m & 0_{m\times p} \\ \star & \star & -\gamma I_p \end{bmatrix}. \tag{7.17}$$

The so-called Bounded-Real Lemma is recalled next.

Lemma 7.1 ([228])**.** *Consider a MIMO LTI model characterized by the model matrices $\mathcal{A} \in \mathbb{R}^{n\times n}, \mathcal{B} \in \mathbb{R}^{n\times m}$, and $\mathcal{C} \in \mathbb{R}^{p\times n}$. For any $\gamma > 0$,*

$$\sigma(\mathcal{A}) \in \mathbb{C}^- \quad \textit{and} \quad \left\|\mathcal{C}(sI - \mathcal{A})^{-1}\mathcal{B}\right\|_\infty < \gamma$$

hold true if and only if the following set of LMIs is feasible:

$$\mathcal{N}_\gamma(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{X}) \prec 0, \quad \mathcal{X} \succ 0. \tag{7.18}$$

▲

Proof. The proof can be found in [228]. □

Using Lemma 7.1, the set of models satisfying all constraints of Problem 7.1 is described in the following theorem.

Theorem 7.3. *Consider the full-order model* (7.3) *and the set of reduced-order models* $\hat{\Sigma}(S-GL, G, C\Pi)$ *characterized by* $G \in \mathcal{G}_M$ *in Theorem 7.1. Given a fixed* $\gamma > 0$*, for any* $G \in \mathcal{G}_\gamma \subset \mathcal{G}_M$ *with*

$$\mathcal{G}_\gamma := \left\{ G \in \mathbb{R}^{\nu \times m} : \exists \mathcal{X} \in \mathbb{S}_{n+\nu} : \mathcal{N}_\gamma(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{X}) \prec 0 \right\}, \tag{7.19}$$

the constraints (7.8b) - (7.8c) *of Problem 7.1 are satisfied, where* $\mathcal{N}_\gamma(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{X})$ *is defined in* (7.17) *and where* $\mathcal{A}, \mathcal{B}, \mathcal{C}$ *are as in* (7.16)*. Moreover, if the full-order model is balanced*[1] *with Hankel singular values*[2] $\bar{h}_1 \geq \ldots \geq \bar{h}_\nu \geq \bar{h}_{\nu+1} \geq \ldots \geq \bar{h}_n$*, then the set* $\mathcal{G}_\gamma$ *is empty for* $\gamma < \bar{h}_{\nu+1}$*.* ▲

Proof. The proof can be found in Appendix F.3. □

Using the results of Theorem 7.3, Problem 7.1 can be rewritten as follows:

$$\begin{aligned} \min_{G,\gamma} \quad & \gamma \\ \text{subject to} \quad & G \in \mathcal{G}_\gamma \end{aligned} \tag{7.20}$$

with the set $\mathcal{G}_\gamma$ as in (7.19). Compared to (7.8), the optimization problem (7.20) contains a smaller number of $\nu \times m$ model parameters. The matrix $\mathcal{N}_\gamma(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{X})$ in (7.17) contains products between G and $\mathcal{X}$, and is, therefore, *bilinear* in these matrix variables. Section 7.4 presents a numerical procedure to minimize γ while satisfying the BMI constraints in Theorem 7.3, returning a numerical solution to Problem 7.1.

Remark 7.4. If the full-order model is balanced, then Theorem 7.3 provides a lower bound for γ in constraint (7.8b). Namely, for any triple (F, G, H) (characterizing $\Gamma(s)$), the following bound holds true:

$$\|\Upsilon(s)\|_\infty = \|\Phi(s) - \Gamma(s)\|_\infty \geq \bar{h}_{\nu+1}, \tag{7.21}$$

where $\bar{h}_{\nu+1}$ is the $(\nu+1)^{\text{th}}$ largest Hankel singular value of the full-order model. This bound is likely to be conservative in the scope of moment matching since the triple (F, G, H) are constrained by the moment matching constraint (7.8c) of Problem 7.1. △

Remark 7.5. Suppose $\omega = 0$ is an interpolation frequency (in addition to $\omega_1 > 0, \ldots, \omega_\kappa > 0$) with corresponding tangential direction $\ell_{\kappa+1}$. Then Theorem 7.1 holds with the matrices $\tilde{S} \in \mathbb{R}^{\nu \times \nu}$ and $\tilde{L} \in \mathbb{R}^{m \times \nu}$, given by

$$\tilde{S} = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}, \quad \tilde{L} = \begin{bmatrix} L & \ell_{\kappa+1} \end{bmatrix}, \tag{7.22}$$

playing the role of S and L, where $\nu = 2\kappa + 1$. Theorem 7.3 holds irrespective of $\omega = 0$ being an interpolation frequency or not. △

[1] A model is called balanced if its observability and controllability Gramians are equal and diagonal.

[2] The Hankel singular values of a balanced model are the diagonal entries of the Gramian matrix.

Remark 7.6. In the single input case ($m = 1$), the model set characterized by $G \in \mathcal{G}_\gamma$ in Theorem 7.3 contains *all* models that satisfy the constraints of Problem 7.1 for a given $\gamma \geq 0$ under the condition that the chosen tangential directions $\ell_i \in \mathbb{R}, i \in \{1, \ldots, \kappa\}$, render the pair (S, L) observable. △

7.3 Optimal moment matching for multivariable nonlinear feedback models

This section proposes a moment matching approach for the class of nonlinear models consisting of a feedback interconnection of high-order LTI dynamics with low-order nonlinear dynamics, as depicted in Figure 7.1. The aim is to find a structure-preserving reduced-order model that is optimal in terms of minimizing an error bound and that preserves the input-to-state convergence (ISC) property of the full-order model. The ISC property ensures the existence, uniqueness, and asymptotic stability of steady-state solutions. In addition, it guarantees that the difference between two steady-state responses corresponding to two different inputs is finite. First, the considered class of nonlinear models is introduced and the ISC property is recalled in Section 7.3.1. After that, the notion of moments is defined and the moment matching problem is formalized in Section 7.3.2, followed by a proposed solution strategy in Section 7.3.3.

7.3.1 Class of convergent nonlinear feedback models

The considered models are described by:

$$\Sigma_{lin}: \quad \begin{cases} \dot{x} &= Ax + B_u u + B_w w, \\ z &= C_z x, \\ y &= C_y x, \end{cases} \tag{7.23a}$$

$$\Sigma_{nl}: \quad \begin{cases} \dot{v} &= g(v, z), \\ w &= h(v), \end{cases} \tag{7.23b}$$

where $x(t) \in \mathbb{R}^n$ is the state of the LTI block, $u(t) \in \mathbb{R}^m$ is the input, $y(t) \in \mathbb{R}^p$ is the model output, $z(t) \in \mathbb{R}^q$ is the input to the nonlinear block, $w(t) \in \mathbb{R}^s$ is the output of the nonlinear block, and $A \in \mathbb{R}^{n\times n}, B_u \in \mathbb{R}^{n\times m}, B_w \in \mathbb{R}^{n\times s}, C_z \in \mathbb{R}^{q\times n}, C_y \in \mathbb{R}^{p\times n}$ are model matrices. The state of the nonlinear block is denoted by $v(t) \in \mathbb{R}^r$, and $g : \mathbb{R}^r \times \mathbb{R}^q \to \mathbb{R}^s$, is a mapping locally Lipschitz in v and continuous in z, and $h : \mathbb{R}^r \to \mathbb{R}^s$, is a continuous mapping. Figure 7.1 depicts model (7.23) schematically. The transfer functions, associated with (7.23a), are defined as follows:

$$\Phi_{(i,k)}(s) \coloneqq C_i(sI - A)^{-1}B_k, \quad s \in \mathbb{C}, \tag{7.24}$$

for $i \in \{z, y\}, k \in \{u, w\}$. The overall transfer function matrix collects the individual ones:

$$\Phi(s) := \begin{bmatrix} \Phi_{(z,u)}(s) & \Phi_{(z,w)}(s) \\ \Phi_{(y,u)}(s) & \Phi_{(y,w)}(s) \end{bmatrix}, \quad s \in \mathbb{C}. \tag{7.25}$$

To define moments of the nonlinear model (7.23) in a *global* context, a strong form of model stability is required. Hereto, the convergence and input-to-state convergence property are recalled next for *generic* nonlinear models of the form

$$\dot{x} = f(x, u) \tag{7.26}$$

with $x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m$.

Definition 7.3 ([200]). *The model* (7.26) *is said to be globally uniformly convergent if for every input $u \in \mathcal{L}_\infty^m$, there exists a solution $\bar{x}_u$ to* (7.26) *satisfying the following conditions:*

- *$\bar{x}_u$ is defined and bounded on $\mathbb{R}$.*
- *$\bar{x}_u$ is globally uniformly asymptotically stable.*

■

The solution $\bar{x}_u$ is called the steady-state solution. The input-to-state convergence (ISC) property is a stronger property and is defined as follows.

Definition 7.4 ([200]). *Model* (7.26) *is said to be input-to-state convergent if it is globally uniformly convergent for the class of inputs $\mathcal{L}_\infty^m$ and, for every input $u \in \mathcal{L}_\infty^m$, model* (7.26) *is input-to-state stable with respect to the steady-state solution $\bar{x}_u(t)$, i.e., there exist a $\mathcal{KL}$-function $\beta(r, s)$ and a $\mathcal{K}_\infty$-function $\gamma_{xu}(r)$ such that any solution $x(t)$ corresponding to some input $\hat{u}(t) := u(t) + \Delta u(t), \Delta u \in \mathcal{L}_\infty^m$, satisfies*

$$|x(t) - \bar{x}_u(t)| \leq \beta(|x(t_0) - \bar{x}(t_0)|, t - t_0) + \gamma_{xu}\left(\sup_{t_0 \leq \tau \leq t} |\Delta u(t)|\right) \tag{7.27}$$

for all $t, t_0 \in \mathbb{R}$, $t \geq t_0$. The functions $\beta(r, s)$ and $\gamma_{xu}(r)$ may depend on the particular input u. ■

The function γ_{xu} is called an *incremental gain function* from input u to state x.

Given these definitions, conditions for the uniform convergence and ISC property of the model (7.23) are recalled next. Starting with the linear part of model (7.23), the matrix A being Hurwitz is necessary and sufficient for the ISC property. Therefore, there exist incremental gain functions $\gamma_{xu}, \gamma_{xw} \in \mathcal{K}_\infty$ from inputs u, w to state x, respectively. Furthermore, the steady-state operators $\mathcal{F}_i(u, w) := C_i \bar{x}_{u,w}, i \in \{z, y\}$, are incrementally bounded as follows:

$$\|\mathcal{F}_i(u_1, w_1) - \mathcal{F}_i(u_2, w_2)\|_\infty \leq \chi_{ix}(\gamma_{xu}(\|u_1 - u_2\|_\infty) + \gamma_{xw}(\|w_1 - w_2\|_\infty)) \tag{7.28}$$

for $i \in \{z, y\}$ and for all $u_1, u_2 \in \mathcal{L}^m_\infty, w_1, w_2 \in \mathcal{L}^s_\infty$. The functions $\chi_{zx}, \chi_{yx} \in \mathcal{K}_\infty$ in (7.28) represent incremental bounds on the output equations. For the nonlinear block Σ_{nl} of model (7.23), it is assumed that $h(0) = 0$ and that the incremental bound

$$|h(v_1) - h(v_2)| \leq \chi_{wv}(|v_1 - v_2|) \tag{7.29}$$

holds for all $v_1, v_2 \in \mathbb{R}^r$. Under the assumption that Σ_{nl} is ISC, there exists a class $\mathcal{K}_\infty$ function γ_{vz} such that the nonlinear steady-state operator $\mathcal{F}_{nl} z := \bar{v}_z$ satisfies

$$\|\mathcal{F}_{nl} z_1 - \mathcal{F}_{nl} z_2\|_\infty \leq \gamma_{vz}(\|z_1 - z_2\|_\infty) \tag{7.30}$$

for any $z_1, z_2 \in \mathcal{L}^q_\infty$. Finally, it is assumed that there exist functions $\rho_1, \rho_2 \in \mathcal{K}_\infty$ such that the small-gain condition

$$(\mathrm{id} + \rho_1) \circ \gamma_{xw} \circ \chi_{wv} \circ (\mathrm{id} + \rho_2) \circ \gamma_{vz} \circ \chi_{zx}(r) \leq r, \tag{7.31}$$

holds for all $r \geq 0$. The aforementioned assumptions are sufficient for the ISC property of model (7.23).

Assumption 7.2. *Model Σ_{lin} in* (7.23a) *is asymptotically stable and Σ_{nl} in* (7.23b) *is input-to-state convergent; the function h in* (7.23) *satisfies $h(0) = 0$ and* (7.29)*; and the small-gain condition* (7.31) *is satisfied.* ■

Theorem 7.4. *Suppose that Assumption 7.2 holds. Then, model* (7.23) *is input-to-state convergent.* ▲

Proof. The theorem is a special case of [31, Theorem 2]. □

7.3.2 The optimal nonlinear moment matching problem

Consider $\kappa \in \mathbb{Z}^+$ distinct frequencies $\omega_i > 0, i = 1, \ldots, \kappa$, and the tangential directions $\ell_i \in \mathbb{R}^m, i = 1, \ldots, \kappa$, characterizing the matrices $S \in \mathbb{R}^{\nu \times \nu}, L \in \mathbb{R}^{\nu \times m}$ in (7.9) with $\nu = 2\kappa$. The pair (S, L) forms the signal generator

$$\dot{\tau} = S\tau, \quad u = L\tau \tag{7.32}$$

with $\tau(t) \in \mathbb{R}^\nu$, generating input $u = L\tau$ for model (7.23). Note that by the construction of S, the linear signal generator is marginally stable and produces signals τ and u of class $\mathcal{L}^\nu_\infty$ and $\mathcal{L}^m_\infty$, respectively.

Definition 7.5. *Consider the model* (7.23) *with input $u = L\tau$ driven by signal generator* (7.32) *with the pair (S, L) observable. Suppose there exist unique functions $\pi_1 : \mathbb{R}^\nu \to \mathbb{R}^n : \tau \mapsto \pi_1(\tau)$ and $\pi_2 : \mathbb{R}^\nu \to \mathbb{R}^r : \tau \mapsto \pi_2(\tau)$, such that the graph*

$$\mathcal{M} \coloneqq \{(\tau, x, v) : x = \pi_1(\tau), v = \pi_2(\tau), \tau \in \mathbb{R}^\nu\} \tag{7.33}$$

is invariant. Then, the function $C_y \pi_1$ is called the moment of the model (7.23) *at (S, L).* ■

Given the fact that $\tau \in \mathcal{L}_\infty^\nu$ and $u \in \mathcal{L}_\infty^m$, Assumption 7.2 guarantees the existence and uniqueness of the functions π_1, π_2, see [200, Theorem 4.4]. Furthermore, since the steady-state solution of (7.23) is globally asymptotically stable by Assumption 7.2, the graph $\mathcal{M}$ in (7.33) can be found by numerical simulation of model (7.23) subject to input u generated by the signal generator (7.32). Note that, in contrast to [16, 226], moments as in Definition 7.5 are defined *globally*.

Consider the $(\nu + r)$-th-order model

$$\hat{\Sigma}_{lin}: \quad \begin{cases} \dot{\xi} &= F\xi + G_u u + G_\lambda \lambda, \\ \zeta &= H_\zeta \xi, \\ \psi &= H_\psi \xi, \end{cases} \tag{7.34a}$$

$$\Sigma_{nl}: \quad \begin{cases} \dot{\mu} &= g(\mu, \zeta), \\ \lambda &= h(\mu), \end{cases} \tag{7.34b}$$

where $\xi(t) \in \mathbb{R}^\nu$ is the state of the LTI block, $u(t) \in \mathbb{R}^m$ is the input, $\psi(t) \in \mathbb{R}^p$ is the model output, $\zeta(t) \in \mathbb{R}^q$ is the input to the nonlinear block, $\lambda(t) \in \mathbb{R}^s$ is the output of the nonlinear block, and $F \in \mathbb{R}^{\nu\times\nu}, G_u \in \mathbb{R}^{\nu\times m}, G_\lambda \in \mathbb{R}^{\nu\times s}, H_\zeta \in \mathbb{R}^{q\times\nu}, H_\psi \in \mathbb{R}^{p\times\nu}$ are model matrices. The state of the nonlinear block is denoted by $\mu(t) \in \mathbb{R}^r$. The reduction approach proposed in this chapter only reduces the state dimension of the LTI part, motivated by nonlinear behavior only appearing locally in engineering systems or being described by low-order nonlinear dynamics. Consequently, the same mappings g and h as in model (7.23b) are utilized in the reduced model (7.34). Figure 7.1 depicts model (7.23) schematically. The transfer function matrices associated with the LTI part of (7.34) are denoted by $\Gamma_{(i,k)}$, for $i \in \{\zeta, \psi\}, k \in \{u, \lambda\}$, and are defined similarly to $\Phi_{(i,k)}$ in (7.24). The overall transfer function matrix Γ is defined similarly to Φ in (7.25). The moment of model (7.34) at (S, L) is denoted by $H_\psi p_1$, where the function p_1 plays the role of the function π_1 in Definition 7.5.

The reduction method proposed in this chapter preserves the ISC property for the reduced-order model. Hereto, since the nonlinear part Σ_{nl} of (7.34) is inherited from (7.23), the only conditions for ISC of model (7.34) are: i) the matrix F is Hurwitz; and ii) a small-gain condition of the form (7.31) is satisfied for the reduced-order model. The (incremental) gain functions of the nonlinear part of the reduced-order model can be inherited from the full-order model, i.e., $\gamma_{\mu\zeta}$ and $\chi_{\lambda\mu}$ can be chosen as γ_{vz} and χ_{wv}, respectively. Therefore, the small-gain condition (7.31) can be written as follows for the reduced-order model:

$$(\mathrm{id} + \rho_1) \circ \gamma_{\xi\lambda} \circ \chi_{\lambda\mu} \circ (\mathrm{id} + \rho_2) \circ \gamma_{\mu\zeta} \circ \chi_{\zeta\xi}(r) \leq r, \tag{7.35}$$

where $\chi_{\zeta\xi}$ and $\gamma_{\xi\lambda}$ are the linear (incremental) gain functions of the LTI part of the reduced-order model in (7.34). Thus, the satisfaction of (7.35), together with F being Hurwitz, guarantees the ISC property for the reduced-order model. Furthermore, the matrix F being Hurwitz guarantees the existence of the additional linear (incremental) gain functions $\gamma_{\xi u}, \chi_{\psi\xi} \in \mathcal{K}_\infty$.

Due to the nonlinear feedback, matching the moment $H_\psi p_1$ of the reduced-order model (7.34) to the moment $C_y \pi_1$ of the full-order model (7.23) implies matching the transfer function matrices at an infinite number of interpolation points. As this is non-trivial to achieve, in this work, only a finite number of interpolation points are considered, inevitably resulting in a mismatch between the moments. To address this issue, the next theorem recalls from [31, Theorem 3] an upper bound for the steady-state mismatch between $\bar{y}_u$ and $\bar{\psi}_u$ for any $u \in \mathcal{L}_\infty^m$. This bound also holds for input signals generated by signal generator (7.32), thereby also bounding the mismatch between the moment $H_\psi p_1$ of the reduced-order and the moment $C_y \pi_1$ of the full-order model.

Theorem 7.5. *Consider the model* (7.23) *and suppose that Assumption 7.2 holds. Furthermore, consider the model* (7.34) *and suppose that F is Hurwitz and condition* (7.35) *holds. Then, for any input $u \in \mathcal{L}_\infty^m$, the following steady-state error bound holds:*

$$\left\|\bar{y}_u - \bar{\psi}_u\right\|_\infty \leq \varepsilon_\nu(\|u\|_\infty), \tag{7.36}$$

where

$$\begin{aligned}\varepsilon_\nu(r) \coloneqq &(\varepsilon_{yu} + \varepsilon_{yw} \circ (id + \rho_5^{-1}) \circ \eta_{wu} \\ &+ (\chi_{yx} \circ \gamma_{xw} + \varepsilon_{yw} \circ (id + \rho_5)) \circ \delta_{wu})(r)\end{aligned} \tag{7.37}$$

with $\varepsilon_{ab}(r) \coloneqq r\left\|\Upsilon_{(a,b)}(s)\right\|_1$, $\forall\, r \geq 0$, defined for input/output pairs $a \in \{y, z\}, b \in \{u, w\}$ of the LTI block, and ρ_5 an arbitrary class $\mathcal{K}_\infty$ function. The following functions are used in (7.37)*:*

$$\begin{aligned}\eta_{wu}(r) \coloneqq &\chi_{wv} \circ \gamma_{vz} \circ \eta_{zu}(r), \\ \eta_{zu}(r) \coloneqq &(id - \chi_{zx} \circ (id + \rho_3) \circ \gamma_{xw} \circ \chi_{wv} \circ \gamma_{vz})^{-1} \\ &\circ \chi_{zx} \circ (id + \rho_3^{-1}) \circ \gamma_{xu}(r), \\ \delta_{wu}(r) \coloneqq &\chi_{wv} \circ \gamma_{vz} \circ \delta_{zu}(r), \\ \delta_{zu}(r) \coloneqq &(id - (\chi_{zx} \circ \gamma_{xw} + \varepsilon_{zw} \circ (id + \rho_4)) \circ \chi_{wv} \circ \gamma_{vz})^{-1} \\ &\circ (\varepsilon_{zu} + \varepsilon_{zw} \circ (id + \rho_4^{-1}) \circ \eta_{wu})(r),\end{aligned}$$

where ρ_3 and ρ_4 are arbitrary class $\mathcal{K}_\infty$ functions. ▲

Proof. The proof can be found in [31, Theorem 3]. □

Theorem 7.5 evidences that the error between the LTI blocks in terms of $\|\Upsilon(s)\|_1$ plays an important role in the error bound via the functions $\varepsilon_{yu}, \varepsilon_{yw}, \varepsilon_{zu}$, and ε_{zw}. Using the relation (7.2), for all input/output pairs of the LTI block, the $\mathcal{L}_\infty$-induced norm $\|\Upsilon(s)\|_1$ of the error transfer function matrix Υ can be upper bounded by its $\mathcal{H}_\infty$ norm $\|\Upsilon(s)\|_\infty$ as follows:

$$\|\Upsilon(s)\|_1 \leq \sqrt{m+s}(2(n+\nu)+1)\|\Upsilon(s)\|_\infty, \tag{7.38}$$

where $m+s$ are the number of inputs and $n+\nu$ are the number of poles of the transfer function matrix Υ. Furthermore, in case $\|\Upsilon(s)\|_\infty = 0$, the error bound (7.36) drops to zero, i.e., $\varepsilon_\nu(r) = 0 \,\forall\, r \geq 0$. Besides matching the moments of the linear block, the reduction problem defined next aims at preserving the ISC property and minimizing $\|\Upsilon(s)\|_\infty$, and therewith minimizing an upper bound on $\|\Upsilon(s)\|_1$. The minimization of $\|\Upsilon(s)\|_\infty$ is favored over the direct minimization of $\|\Upsilon(s)\|_1$ because necessary and sufficient matrix inequality conditions for the computation of the former exist in the literature, whereas only sufficient matrix inequality conditions exist for the latter.

Problem 7.2. *Consider the full-order model* (7.23)*, the reduced-order model* (7.34)*, $\kappa \in \mathbb{Z}^+$ distinct frequencies $\omega_i > 0, i = 1, \ldots, \kappa$, the tangential directions $\ell_i \in \mathbb{R}^{m+s}, i = 1, \ldots, \kappa$. Suppose Assumption 7.2 holds. The problem of* Optimal $\mathcal{H}_\infty$ model reduction by approximate moment matching *consists of solving the following constrained optimization problem:*

$$\begin{aligned}
\min_{F, G_u, G_\lambda, H_\zeta, H_\psi, \gamma} \quad & \gamma && \text{(7.39a)}\\
\text{subject to} \quad & \|\Upsilon(s)\|_\infty < \gamma, && \text{(7.39b)}\\
& \textit{Satisfaction of } (7.35), && \text{(7.39c)}\\
& \eta_0^{\Sigma_{lin}}(\pm j\omega_i, \ell_i) = \eta_0^{\hat{\Sigma}_{lin}}(\pm j\omega_i, \ell_i), \quad \textit{for } i = 1, \ldots, \kappa. && \text{(7.39d)}
\end{aligned}$$

△

The objective of Problem 7.2 minimizes $\|\Upsilon(s)\|_\infty$, playing an important role in the upper bound on the reduction error in Theorem 7.5 via (7.38). Constraint (7.39c) guarantees the satisfaction of condition (7.35), resulting in the preservation of the input-to-state convergence property. Constraint (7.39d) guarantees moment matching of the LTI parts.

7.3.3 Parametrization of models satisfying the constraints of Problem 7.2

Reduced-order LTI models achieving moment matching

The tangential direction $\ell_i \in \mathbb{R}^{m+s}, i \in \{1, \ldots, \kappa\}$, in Problem 7.2 include the tangential directions for both the inputs u and w. Application of Theorem 7.1 with $B \coloneqq \begin{bmatrix} B_u & B_w \end{bmatrix} \in \mathbb{R}^{n \times m+s}$ and $C \coloneqq \begin{bmatrix} C_z^\top & C_y^\top \end{bmatrix}^\top \in \mathbb{R}^{p+q \times n}$ results in a set of reduced-order LTI models (7.5) that satisfy constraint (7.39d). This family is described by the set $\tilde{\mathcal{G}}_M$ in Theorem 7.1, recalled here for convenience:

$$\tilde{\mathcal{G}}_M \coloneqq \left\{ G \in \mathbb{R}^{\nu \times m+s} \,:\, \sigma(S) \cap \sigma(S - GL) = \emptyset \right\}. \tag{7.40}$$

The obtained matrix $H = \begin{bmatrix} H_\zeta^\top & H_\psi^\top \end{bmatrix}^\top \in \mathbb{R}^{q+p \times \nu}$ can be decomposed into $H_\zeta = C_z \Pi \in \mathbb{R}^{q \times \nu}$ and $H_\psi = C_y \Pi \in \mathbb{R}^{p \times \nu}$, where $\Pi \in \mathbb{R}^{n \times \nu}$ is the solution to the

Sylvester equation (7.12). Similarly, the matrix $G = \begin{bmatrix} G_u & G_\lambda \end{bmatrix} \in \mathbb{R}^{\nu \times m+s}$ can be decomposed into $G_u \in \mathbb{R}^{\nu \times m}$ and $G_\lambda \in \mathbb{R}^{\nu \times s}$.

Reduced-order models satisfying all constraints of Problem 7.2

Consider the LTI error dynamics described in the following form:

$$\begin{aligned} \begin{bmatrix} \dot{x} \\ \dot{\xi} \end{bmatrix} &= \underbrace{\begin{bmatrix} A & 0_{n\times\nu} \\ 0_{\nu\times n} & S - GL \end{bmatrix}}_{\tilde{\mathcal{A}}} \begin{bmatrix} x \\ \xi \end{bmatrix} + \underbrace{\begin{bmatrix} B_u & B_w \\ G_u & G_\lambda \end{bmatrix}}_{\tilde{\mathcal{B}}} \begin{bmatrix} u & v \end{bmatrix} \\ \begin{bmatrix} e_z \\ e_y \end{bmatrix} &= \underbrace{\begin{bmatrix} C_z & -H_\zeta \\ C_y & -H_\psi \end{bmatrix}}_{\tilde{\mathcal{C}}} \begin{bmatrix} x \\ \xi \end{bmatrix}. \end{aligned} \tag{7.41}$$

It is a challenging problem to enforce constraint (7.39c) during the reduction procedure without introducing additional conservatism. To make the problem tractable and to limit the introduction of additional conservatism, we pose the following assumption.

Assumption 7.3. *Satisfaction of the inequality* $\left\|\Gamma_{(\zeta,\lambda)}(s)\right\|_\infty \leq \left\|\Phi_{(z,w)}(s)\right\|_\infty =: \bar{\gamma}_{zw}$ *implies satisfaction of* (7.39c). ■

Given that the nonlinear block is inherited from the full-order model, upper bounding the $\mathcal{L}_\infty$-induced norm of $\Gamma_{(\zeta,\lambda)}$ by the $\mathcal{L}_\infty$-induced norm of $\Phi_{(z,w)}$ is sufficient for the constraint (7.39c) to hold. Then, Assumption 7.3 replaces the $\mathcal{L}_\infty$-induced norm with the $\mathcal{H}_\infty$ norm. Although the preservation of the $\mathcal{H}_\infty$ norm does not formally imply the preservation of the $\mathcal{L}_\infty$-induced norm, these two norms are closely related to each other, e.g., via relation (7.2). Additional comments on this assumption are presented in Remark 7.7.

For any fixed γ, the following theorem presents a class of models that satisfy all constraints of Problem 7.2.

Theorem 7.6. *Consider the full-order model* (7.26) *and the set of reduced-order models* (7.34) *characterized by* $G \in \tilde{\mathcal{G}}_M$ *with* $\tilde{\mathcal{G}}_M$ *in* (7.40). *If Assumption 7.3 holds, then given a fixed* $\gamma > 0$, *for any* $G \in \tilde{\mathcal{G}}_\gamma \subset \tilde{\mathcal{G}}_M$ *with*

$$\begin{aligned} \tilde{\mathcal{G}}_\gamma := \{ G \in &\mathbb{R}^{\nu\times m+s} : \exists \mathcal{X}_1 \in \mathbb{S}_{n+\nu}, \exists \mathcal{X}_2 \in \mathbb{S}_\nu : \\ &\mathcal{N}_\gamma(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}, \mathcal{X}_1) \prec 0, \\ &\mathcal{N}_{\bar{\gamma}_{zw}}(S - GL, G_\lambda, H_\zeta, \mathcal{X}_2) \prec 0 \}, \end{aligned} \tag{7.42}$$

the constraints (7.39b) - (7.39d) *of Problem 7.2 are satisfied, where matrix* $\mathcal{N}_\gamma$ *is defined in* (7.17), $\mathcal{A}, \mathcal{B}, \mathcal{C}$ *are as in* (7.41), *and* $\bar{\gamma}_{zw}$ *is defined in Assumption 7.3. Furthermore, if the LTI part of the full-order model is balanced with Hankel singular values* $\bar{h}_1 \geq \ldots \geq \bar{h}_\nu \geq \bar{h}_{\nu+1} \geq \ldots \geq \bar{h}_n$, *then the set* $\tilde{\mathcal{G}}_\gamma$ *is empty for* $\gamma < \bar{h}_{\nu+1}$. ▲

Proof. The proof can be found in Appendix F.4. □

The following optimization problem replaces the one in Problem 7.2:

$$\begin{aligned} \min_{G,\gamma} \quad & \gamma \\ \text{subject to} \quad & G \in \tilde{\mathcal{G}}_\gamma. \end{aligned} \tag{7.43}$$

Compared to (7.39), optimization problem (7.43) contains a smaller number of $\nu \times m + s$ model parameters. Note that the set $\tilde{\mathcal{G}}_\gamma$ in (7.42) is characterized by bilinear matrix inequalities due to products between G and $\mathcal{X}_1, \mathcal{X}_2$. A numerical procedure to find the gain matrix G such that γ is minimized is presented in the next section.

Remark 7.7. Assumption 7.3 does not hold in general. Alternative approaches that do guarantee the satisfaction of the constraint (7.39c) are conservative. The first approach is to use the inequality (7.2) to write the following inequalities:

$$\begin{aligned} \frac{1}{\sqrt{q}} \left\| \Phi_{(z,w)}(s) \right\|_\infty &\leq \left\| \Phi_{(z,w)}(s) \right\|_1 \\ \left\| \Gamma_{(\zeta,\lambda)}(s) \right\|_1 &\leq (2\nu+1)\sqrt{s} \left\| \Gamma_{(\zeta,\lambda)}(s) \right\|_\infty, \end{aligned} \tag{7.44}$$

where ν, q, s are the dimensions of $\xi(t) \in \mathbb{R}^\nu, z(t), \zeta(t) \in \mathbb{R}^q$, and $w(t), \lambda(t) \in \mathbb{R}^s$, respectively. In this approach, satisfaction of the inequality $\left\| \Gamma_{(\zeta,\lambda)}(s) \right\|_1 \leq \left\| \Phi_{(z,w)}(s) \right\|_1$ is sufficient for (7.35) to hold and can be guaranteed by the following conservative inequality:

$$\left\| \Gamma_{(\zeta,\lambda)}(s) \right\|_\infty \leq \bar{\kappa} \left\| \Phi_{(z,w)}(s) \right\|_\infty \tag{7.45}$$

with $\bar{\kappa} \coloneqq (2\nu+1)^{-1}(qs)^{-1/2}$. This approach does not allow the preservation of the $\mathcal{H}_\infty$ norm of the full-order model, which is undesired because it does not allow, e.g., for moment matching at the locations of resonance peaks. Furthermore, it can be conservative as the inequality (7.45) may be conservative for the small-gain condition (7.35) to hold. The second approach is to use the sufficient matrix inequality condition for the computation of the $\mathcal{L}_\infty$-induced norm presented in [229], which is also conservative. It is an open question which of these two alternative approaches introduces the most conservatism. The model reduction problem for nonlinear models with the preservation of the ISC property is inherent to conservatism, as is, for example, also the case for the approach in [31]. The conservatism stems from the ISC property being guaranteed by the satisfaction of the small-gain condition (7.31). This small-gain condition relies on the $\mathcal{L}_\infty$-induced norm of the LTI part of the model, the computation of which for a MIMO LTI block is not exact and can only be upper bounded in practice, hence introducing conservatism.

In contrast to these alternative approaches, using Assumption 7.3, constraint (7.35) is written as a necessary and sufficient matrix inequality constraint in (7.42),

which aims at not exceeding the $\mathcal{H}_\infty$ norm of the transfer function $\Phi_{(z,w)}$. Satisfaction of Assumption 7.3 needs to be verified a posteriori and, if this assumption does not hold, then the reduced-order model is not guaranteed to be ISC. It is noted that, in general, preservation of the $\mathcal{L}_\infty$-induced norm of $\Phi_{(z,w)}$ is not required, because, in general, the small-gain condition (7.35) is satisfied with a strict inequality. In the example presented in this chapter (see Section 7.5 below), this $\mathcal{L}_\infty$-induced norm is slightly exceeded after reduction, whereas Assumption 7.3 does hold. If the resulting reduced-order model does not satisfy Assumption 7.3, then one can replace $\bar{\gamma}_{zw}$ by $\bar{\kappa}\bar{\gamma}_{zw}$ with $\bar{\kappa}$ as in (7.45), and solve the optimization problem again. △

7.4 Numerical procedure for solving the $\mathcal{H}_\infty$ moment matching problem

This section proposes a numerical procedure for solving Problems 7.1 and 7.2, which are both written as optimization problems with BMI constraints. A popular method to solve such optimization problems with BMI constraints is the so-called coordinate-descent algorithm (CDA) [253]. The CDA for the constraints as derived in the previous sections is presented in Section 7.4.1. Alternatively, the constraints can be written in the so-called Finsler's form, for which a CDA is presented in Section 7.4.2. These two CDAs are combined to form a single, novel numerical procedure, as presented in Section 7.4.3. Finally, a method to find a feasible starting point, required in both CDAs, is presented in 7.4.4.

7.4.1 CDA for constraints in primal form

The set $\mathcal{G}_\gamma$ in (7.19) is characterized by the existence of an auxiliary matrix $\mathcal{X}$, whereas the set $\tilde{\mathcal{G}}_\gamma$ in (7.42) is characterized by the auxiliary matrix variables $\mathcal{X}_1, \mathcal{X}_2$. With some abuse of notation, the variable $\mathbb{X}$ is used to denote $\mathcal{X}$ in the case of $\mathcal{G}_\gamma$ in (7.19) and to denote $\mathcal{X}_1, \mathcal{X}_2$ in the case of $\tilde{\mathcal{G}}_\gamma$ in (7.42). Using $\mathbb{X}$, both the constraints of $\mathcal{G}_\gamma$ in (7.19) and $\tilde{\mathcal{G}}_\gamma$ in (7.42) can be written in a standard form $\mathcal{R}_p(\mathbb{X}, G, \gamma) \prec 0$, which is called the *primal* form, and where $\mathcal{R}_p(\mathbb{X}, G, \gamma)$ is defined as follows:

$$\mathcal{R}_p(\mathcal{X}, G, \gamma) := \mathcal{M}_1(\mathbb{X}, G, \gamma) + \mathrm{He}\left(\mathcal{M}_2(\mathbb{X})\mathcal{N}^\top(G)\right). \tag{7.46}$$

The matrices $\mathcal{M}_1 \in \mathbb{R}^{\alpha\times\alpha}, \mathcal{M}_2 \in \mathbb{R}^{\alpha\times\beta}, \mathcal{N} \in \mathbb{R}^{\alpha\times\beta}$, and their dimension variables α, β, are defined in Appendix F.5, see (F.9) and (F.10). It is emphasized that these matrices are *linear* in their arguments. Evidently from (7.46), the bilinear terms are due to products between $\mathcal{M}_2$ and $\mathcal{N}$, multiplying $\mathbb{X}$ with G. A CDA iteratively fixes either $\mathbb{X}$ or G to yield an optimization problem with *linear* constraints, and optimizes γ for the other variables, for example, by a bisection search over γ. Even

Algorithm 7.1 CDA for constraints in primal form

Input: Constraints $\mathcal{R}_p \prec 0$ in (7.46), any $\gamma^{[0]} > 0$ and vector $G^{[0]} \in \mathcal{G}_{\gamma^{[0]}}$ (or $G^{[0]} \in \tilde{\mathcal{G}}_{\gamma^{[0]}}$), and an accuracy threshold $\epsilon > 0$.

1: Set iteration index $i = 1$.
2: **while** $(\gamma^{[i-1]} - \gamma^{[i]})/\gamma^{[i]} \geq \epsilon$ **do**
3: Solve the optimization problem

$$\left(\mathbb{X}^{[i]}, \cdot\right) = \arg\min_{\mathbb{X},\gamma} \gamma \text{ subject to } \mathcal{R}_p(\mathbb{X}, G^{[i-1]}, \gamma) \prec 0.$$

4: Solve the optimization problem

$$\left(G^{[i]}, \gamma^{[i]}\right) = \arg\min_{G,\gamma} \gamma \text{ subject to } \mathcal{R}_p(\mathbb{X}^{[i]}, G, \gamma) \prec 0.$$

5: Update $i = i + 1$.
6: **end**

Output: Matrix $G_I = G^{[i-1]}$ and scalar $\gamma_I = \gamma^{[i-1]}$.

without formal a proof of convergence, CDAs work well in practice and guarantee a non-increasing sequence of γ over the iterations [253].

Consider the CDA listed in Algorithm 7.1. In Step 3 of this algorithm, for a given, fixed G, the corresponding constrained optimization problem is solved, where the constraints are LMIs in the variable $\mathbb{X}$. Essentially, this step computes a tight upper bound for $\|\Upsilon(s)\|_\infty$ for the current fixed G. The computed $\mathbb{X}$ in Step 3 is fixed in Step 4 and the corresponding constrained optimization problem is solved for G, where the constraints are now LMIs in G. Algorithm 7.1 requires an initial $G^{[0]}$ and $0 \leq \gamma^{[0]} < +\infty$ such that $G^{[0]} \in \mathcal{G}_{\gamma^{[0]}}$ (or $G^{[0]} \in \tilde{\mathcal{G}}_{\gamma^{[0]}}$). Such initial $G^{[0]}, \gamma^{[0]}$ always exists, as is shown in Section 7.4.4.

Remark 7.8. To reduce the complexity of the exposition, the matrix $\mathbb{X}$ is completely fixed in Step 4 of Algorithm 7.1. However, it is not necessary to fix the complete matrix $\mathbb{X}$, it suffices to only fix its elements that are multiplied by G. △

7.4.2 CDA for constraints in Finsler's form

In the scope of minimizing γ while respecting the constraints of Problems 7.1 and 7.2, a constraint formulation that does not contain products between $\mathbb{X}$ and G is favorable. Thanks to Finsler's Lemma, see, e.g., [41], the constraints are rewritten into the so-called *Finsler's form*: $\mathcal{R}_F(\mathbb{X}, N, G, \gamma) \prec 0$ with $N \in \mathbb{R}^{\alpha+\beta\times\beta}$ a new, additional variable. The matrix $\mathcal{R}_F$ reads as follows:

$$\mathcal{R}_F(\mathbb{X}, N, G, \gamma) := \mathcal{M}(\mathbb{X}, \gamma) + \mathrm{He}\left(\begin{bmatrix}\mathcal{N}(G)\\ -I_\beta\end{bmatrix} N^\top\right), \tag{7.47}$$

Algorithm 7.2 CDA for constraints in Finsler's form

Input: Constraints $\mathcal{R}_F \prec 0$ in (7.47), any $\gamma^{[0]} > 0$ and vector $G^{[0]} \in \mathcal{G}_{\gamma^{[0]}}$ (or $G^{[0]} \in \tilde{\mathcal{G}}_{\gamma^{[0]}}$), and an accuracy threshold $\epsilon > 0$.

1: Set iteration index $i = 1$.
2: **while** $(\gamma^{[i-1]} - \gamma^{[i]})/\gamma^{[i]} \geq \epsilon$ **do**
3: Solve the optimization problem

$$\left(\cdot, N^{[i]}, \cdot\right) = \underset{\mathbb{X},N,\gamma}{\arg\min}\, \gamma \quad \text{subject to } \mathcal{R}_F(\mathbb{X}, N, G^{[i-1]}, \gamma) \prec 0.$$

4: Solve the optimization problem

$$\left(\cdot, G^{[i]}, \gamma^{[i]}\right) = \underset{\mathbb{X},G,\gamma}{\arg\min}\, \gamma \quad \text{subject to } \mathcal{R}_F(\mathbb{X}, N^{[i]}, G, \gamma) \prec 0.$$

5: Update $i = i + 1$.
6: **end**

Output: Matrix $G_{II} = G^{[i-1]}$ and scalar $\gamma_{II} = \gamma^{[i-1]}$.

where

$$\mathcal{M}(\mathbb{X}, \gamma) := \begin{bmatrix} \mathcal{M}_1(\mathbb{X}, \gamma) & \mathcal{M}_2(\mathbb{X}) \\ \star & 0_{\beta\times\beta} \end{bmatrix}. \tag{7.48}$$

Note that the matrix $\mathcal{R}_F(\mathbb{X}, N, G, \gamma)$ does not contain products between $\mathbb{X}$ and G, but instead contains a product between G and the new variable N.

Analogously to Algorithm 7.1, Algorithm 7.2 presents the CDA for iterations in the Finsler's form of the constraints. The largest difference is that in Step 4, the variable N is fixed, leaving $\mathbb{X}$ a decision variable in both Steps 3 and 4. Algorithms 7.1 and 7.2 are combined in the numerical procedure presented in the next section.

Remark 7.9. By the Finsler's Lemma, see, e.g., [41], the constraints in primal and Finsler's form are equivalent, i.e., for any given $\mathbb{X}, G, \gamma$, the matrix inequality $\mathcal{R}_p(\mathbb{X}, G, \gamma) \prec 0$ holds true if and only if an additional matrix N exists such that $\mathcal{R}_F(\mathbb{X}, N, G, \gamma) \prec 0$. △

7.4.3 Combination of CDAs

Algorithms 7.1 and 7.2 enable switching when the respective algorithm cannot decrease γ anymore. For example, suppose that Algorithm 7.1 is performed and returned γ_I and $G_I \in \mathcal{G}_{\gamma_I}$. Then, G_I, γ_I can be used as a starting point for Algorithm 7.2, i.e., $G_{II}^{[0]} = G_I, \gamma_{II}^{[0]} = \gamma_I$. Similarly, after running Algorithm 7.2 and obtaining γ_{II} and $G_{II} \in \mathcal{G}_{\gamma_{II}}$, $G_I^{[0]} = G_{II}, \gamma_I^{[0]} = \gamma_{II}$ can be used as a starting

Algorithm 7.3 Combination of CDAs

Input: Constraints $\mathcal{R}_p \prec 0$ in (7.46), constraints $\mathcal{R}_F \prec 0$ in (7.47) any $\gamma^{[0]} > 0$ and vector $G^{[0]} \in \mathcal{G}_{\gamma^{[0]}}$ (or $G^{[0]} \in \tilde{\mathcal{G}}_{\gamma^{[0]}}$), and an accuracy threshold $\epsilon > 0$.

1: Set iteration index $k = 1$.
2: **while** $(\gamma^{[k-1]} - \gamma^{[k]})/\gamma^{[k]} \geq \epsilon$, **do**
3: Obtain $G_I^{[k]} = G_I$ and $\gamma^{[k]} = \gamma_I$ by running Algorithm 7.1 starting from $G_{II}^{[k-1]}$.
4: Update $k = k + 1$.
5: Obtain $G_{II}^{[k]} = G_{II}$ and $\gamma^{[k]} = \gamma_{II}$ by running Algorithm 7.2 starting from $G_I^{[k-1]}$.
6: Update $k = k + 1$.
7: **end**

Output: Matrix $G = G_{II}^{[k]}$ and scalar $\gamma = \gamma^{[k]}$.

point for Algorithm 7.1. Switching is enabled thanks to the constraints in the primal and Finsler's form being equivalent, see Remark 7.9. The resulting novel numerical procedure is formalized in Algorithm 7.3.

Studying properties of Algorithm 7.3, e.g., convergence and robustness with respect to the starting point, is left for future work, although a preliminary study in Section 7.5, shows that this algorithm can potentially avoid getting stuck prematurely. Furthermore, the switching nature of Algorithm 7.3 provides robustness if either of Algorithms 7.1 or 7.2 runs into numerical problems, especially relevant for the involved high-dimensional matrix inequalities.

7.4.4 Initial $G^{[0]}$

In order to launch Algorithm 7.3, an initial $G^{[0]} \in \mathcal{G}_{\gamma^{[0]}}$ (or $G^{[0]} \in \tilde{\mathcal{G}}_{\gamma^{[0]}}$) and $\gamma^{[0]}$ are required. For the LTI case, any $G^{[0]}$ that renders $\sigma(S - G^{[0]}L) \in \mathbb{C}^-$ guarantees the existence of a $\gamma^{[0]}$. For example, by pole placement, the eigenvalues of $S - G^{[0]}L$ can be placed at any desired location thanks to the observability of the pair (S, L) (by the construction of this pair in (7.9)). Placing the eigenvalues of $S - G^{[0]}L$ to a subset of the eigenvalues of A works well in practice as is demonstrated in the numerical case studies in Section 7.5.

For the nonlinear case, any $G^{[0]}$ that renders $\sigma(S - G^{[0]}L) \in \mathbb{C}^-$ and guarantees $\|\Gamma_{zw}(s)\|_\infty \leq \bar{\gamma}_{zw}$ is valid. Hereto, an LMI feasibility problem is presented in the following theorem.

Theorem 7.7. *Consider the reduced-order model* (7.34) *and suppose Assumptions 7.2 and 7.3 hold. If there exists a matrix* $\mathcal{X} \in \mathbb{S}_\nu$ *and matrices* $Q_1 \in \mathbb{R}^{\nu \times m}, Q_2 \in \mathbb{R}^{\nu \times s}$

such that

$$\begin{bmatrix} He\left(\mathcal{X}S - \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} L\right) & Q_2 & H_\zeta^\top \\ \star & -\bar{\gamma}_{zw} I_s & 0_{s\times q} \\ \star & \star & -\bar{\gamma}_{zw} I_q \end{bmatrix} \prec 0 \tag{7.49}$$

with $\bar{\gamma}_{zw}$ defined in Assumption 7.3, then $(S - G^{[0]}L)$ is Hurwitz and constraint (7.39c) is satisfied for $G^{[0]} = \mathcal{X}^{-1}\begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$. Furthermore, there exists a constant $0 \leq \gamma^{[0]} < +\infty$ such that $G^{[0]} \in \tilde{\mathcal{G}}_{\gamma^{[0]}}$ with $\tilde{\mathcal{G}}_\gamma$ in (7.42). ▲

Proof. The proof can be found in Appendix F.6. □

The results of this theorem can be used to find a feasible starting point in the nonlinear case.

7.5 Numerical case studies

Two numerical case studies are presented. Firstly, a cantilever beam with nonlinear dynamics support is considered. In this case study, the benefits of the reduction approach in Section 7.3 are highlighted. Secondly, an LTI example is given which demonstrates the benefits of the numerical procedure in Algorithm 7.3.

7.5.1 Cantilever beam with nonlinear dynamic support

This case study is inspired by [31]. Consider the flexible beam schematically depicted in Figure 7.2. The beam is clamped rigidly to the fixed world at its left end, whereas it is supported near its other end by the first-order nonlinear dynamics Σ_{nl} reading as follows:

$$\Sigma_{nl}: \quad \dot{v} = -v - v^3 + z, \quad w = \varrho v, \tag{7.50}$$

with $\varrho \geq 0$. The nonlinear dynamics are ISC with (incremental) gain functions $\chi_{wv}(r) = \varrho r$ and $\gamma_{vz}^{-1}(r) = r + \frac{1}{4}r^3$, see [31] for more details. The beam has dimensions length × width × height = 2 m × 50 mm × 30 mm and is characterized by its Young's modulus of 200 GPa and density of 7746 kg/m^3. The beam dynamics are described by a $n = 48$-dimensional LTI model that is obtained by the finite-element method. The deflection at the end of the beam is considered as the model output y, see Figure 7.2. The output z is considered as the deflection of the beam at the location of the nonlinear support. Moreover, an external disturbance $u \in \mathcal{L}_\infty$ acts at the middle of the beam. The LTI dynamics have 2 inputs ($m = 1, s = 1$) and 2 outputs ($p = 1, q = 1$). An upper bound for the composed linear gain function $\chi_{zx} \circ \gamma_{xw}$ is computed using the method described in [220], resulting in

$$\chi_{zx} \circ \gamma_{xw}(r) \leq 2.39 \cdot 10^{-4} r \,\forall\, r \geq 0. \tag{7.51}$$

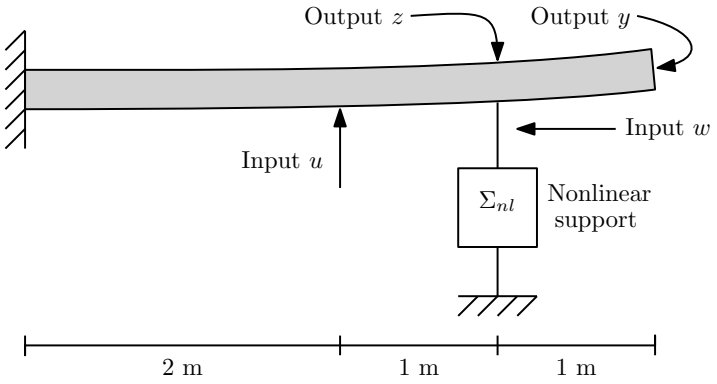


Figure 7.2. One-sided clamped flexible beam supported by a nonlinear dynamic element.

For the choice $\varrho = 3.5 \cdot 10^3$, the conditions of Theorem 7.4 are then satisfied. The interpolation frequencies $\omega_1 = 0$, $\omega_2 = 2\pi 10.2$, and $\omega_3 = 2\pi 64.1$ rad/sec are chosen, corresponding to the zero frequency and the frequencies of the two first resonance peaks. The tangential directions are all chosen as $\ell = \begin{bmatrix} 1 & 1 \end{bmatrix}^\top$, i.e., $\ell_i = \ell, i = 1, 2, 3$.

An initial model, i.e., $G^{[0]}$, is found by placing the poles of $\hat{\Sigma}_{lin}$ at a subset of the poles of Σ_{lin}. The composed gain function $\chi^{[0]}_{\zeta\xi} \circ \gamma^{[0]}_{\xi\lambda}(r) \leq 2.73 \cdot 10^{-4} r$ of the initial reduced-order model has a slightly larger slope than the one of the full-order model in (7.51). Nevertheless, the conditions of Theorem 7.4, in particular (7.35), are satisfied. A solution to Problem 7.2 is obtained by running Algorithm 7.3 starting from $G^{[0]}$, reducing γ from $4.16 \cdot 10^{-5}$ to $5.5 \cdot 10^{-6}$. Based on Remark 7.4, a lower bound for γ in this case study is $2.65 \cdot 10^{-6}$, which is in close proximity of the resulted γ found by Algorithm 7.3. Again, an a posteriori check shows that the conditions of Theorem 7.4 are satisfied for the final reduced-order model with composed gain function $\chi_{\zeta\xi} \circ \gamma_{\xi\lambda}(r) \leq 2.42 \cdot 10^{-4} r$. Therefore, we conclude that Assumption 7.3 is satisfied in this example and, consequently, the ISC property of the full-order model is preserved. In the remainder of this section, the superscript $(\cdot)^\circ$ corresponds to the initial reduced-order model, e.g., $\hat{\Sigma}^\circ_{lin}$. The reduced-order nonlinear feedback model, following from Algorithm 7.3, is referred to as the *final* reduced-order model and denoted by $\hat{\Sigma}_{lin}$.

The Bode magnitude plots of the LTI part of the full-order nonlinear model, the initial $\hat{\Sigma}^\circ_{lin}$ and the final $\hat{\Sigma}_{lin}$ reduced-order nonlinear models are depicted in Figure 7.3. It can be observed that both the initial and final reduced-order models are only accurate for low frequencies. Figure 7.4 depicts the Bode magnitude plot of the LTI part of the error dynamics. It can be concluded that solving Problem 7.2 indeed reduced the peak error in all transfer functions. Furthermore, from the two right subplots, it can be observed that moment matching along the

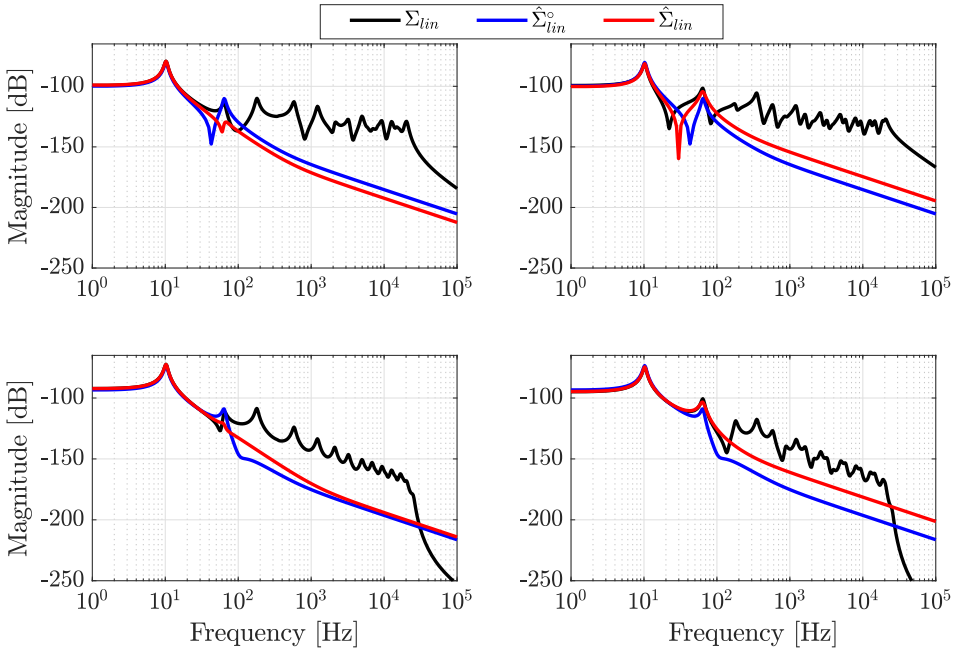


Figure 7.3. Bode magnitude plot of the full-order model Σ_{lin}, initial reduced-order model $\hat{\Sigma}^{\circ}_{lin}$, and final reduced-order model $\hat{\Sigma}_{lin}$.

tangential direction ℓ is achieved since the LTI part of both the initial and final reduced-order models attain zero error for the frequencies ω_2, ω_3.

Next, the time response of the full-order model and the initial and final reduced-order models is analyzed for a block-wave excitation with a frequency 10 Hz. The excitation signal is depicted in the bottom plot of Figure 7.5, whereas the last period of the output response (out of 100 periods) of these models is depicted in the top plot. It is assumed that the model responses are in steady-state in this last period. In the middle plot, it can be seen that the error of the final model is significantly smaller than the error of the initial model. The mismatch $\left\|\bar{y}_u - \bar{\psi}_u\right\|_\infty$ is quantified as 0.135 and 0.019, respectively, for the initial and final reduced-order models. The computation[3] of the time response of the full-order model took 36.4 seconds, whereas the simulation of the final reduced-order model only took 0.5 seconds. This amounts to a reduction of over 98%. A significant reduction in the computation time of model responses is one of the main reasons for performing model reduction.

The error bound in (7.36) for the initial ($\varepsilon^{\circ}_{\nu}$) and final ($\varepsilon_{\nu}$) reduced-order nonlinear models is depicted in Figure 7.6. It can be observed that for small

[3]By numerical forward integration carried out on an Intel Core i7-7700HQ, 2.8 GHz processor.

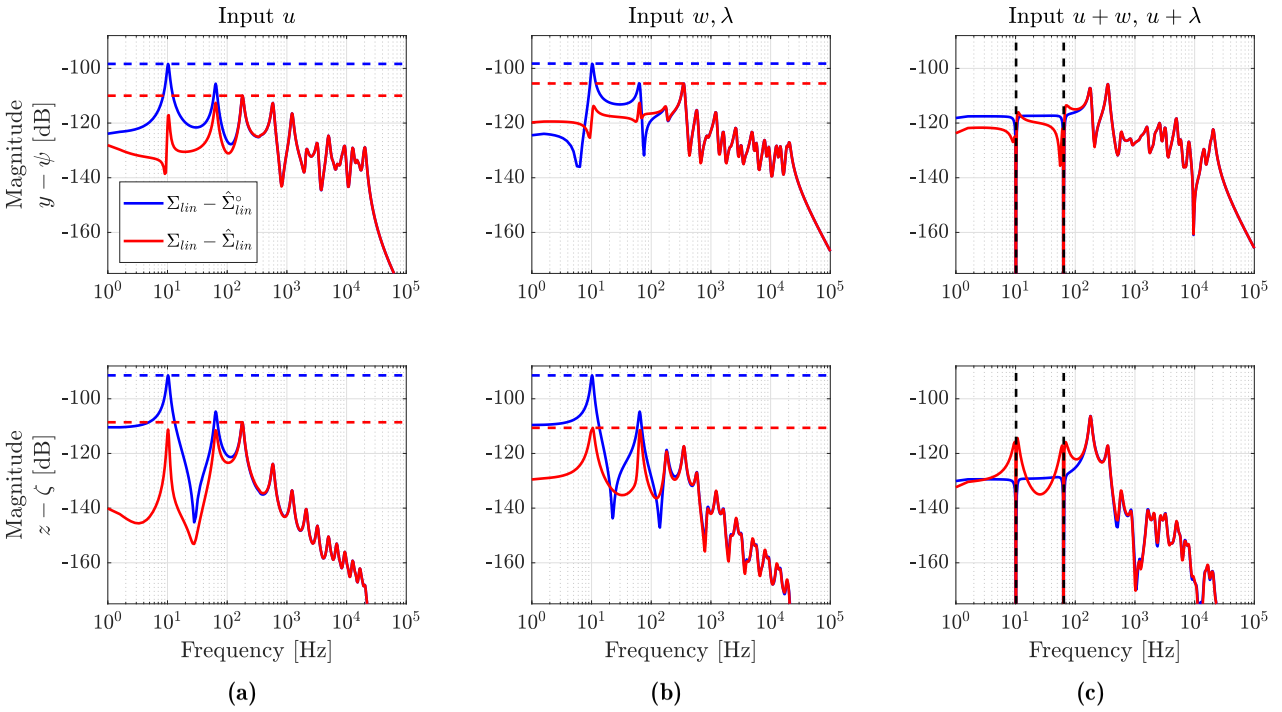


Figure 7.4. Figures 7.4a - 7.4b: Bode magnitude plot of the error dynamics $\Sigma_{lin} - \hat{\Sigma}^{\circ}_{lin}$ and the error dynamics $\Sigma_{lin} - \hat{\Sigma}_{lin}$. Here, the horizontal dashed lines indicate the peak error in the correspondingly colored transfer function. Figure 7.4c: Bode magnitude plot of the error dynamics along the tangential direction $\begin{bmatrix} 1 & 1 \end{bmatrix}^\top$. The vertical black dashed lines in the plots represent the interpolation frequencies ω_2 and ω_3.

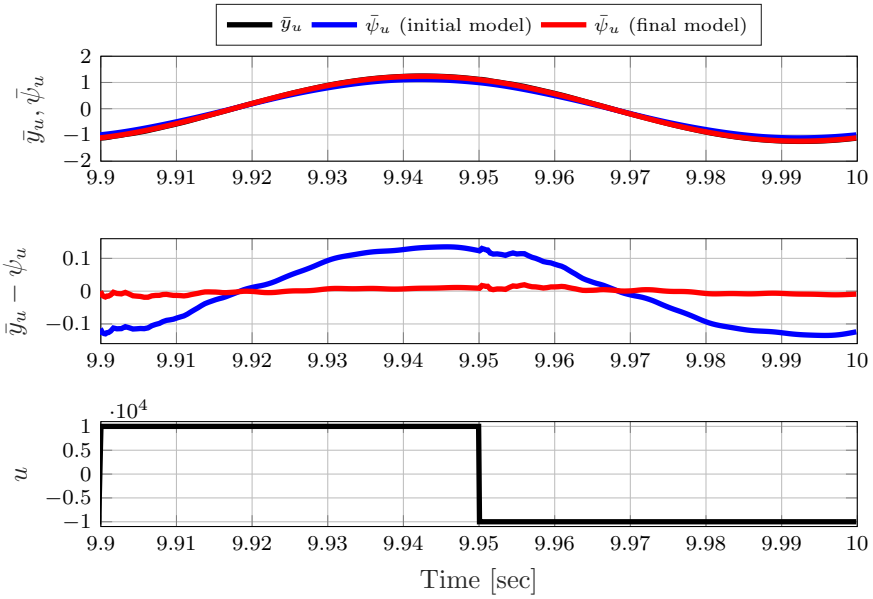


Figure 7.5. Top: one period of the steady-state output response of the full-order, the initial reduced-order, and the final reduced-order nonlinear models. Middle: one period of the steady-state error between the outputs of the full-order and the reduced-order models. Bottom: one period of the applied excitation signal.

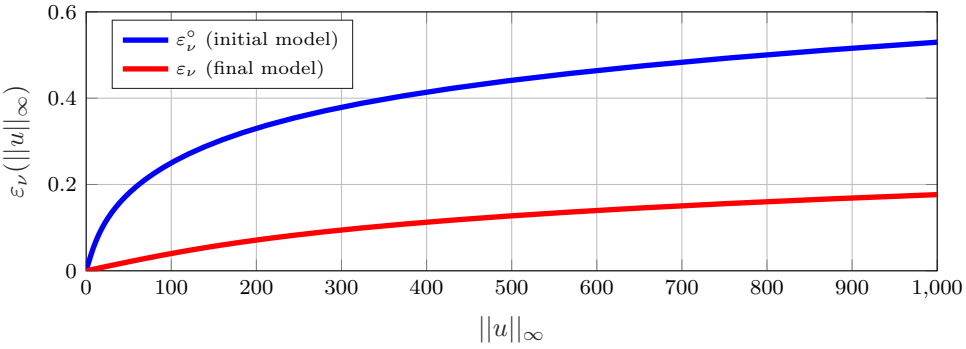


Figure 7.6. Error bound (7.36) computed for the initial model (ε_ν°) and final nonlinear model (ε_ν).

input amplitudes both error bounds are nearly zero. For larger input amplitudes, however, the error bound ε_ν of the final model is significantly smaller than the error bound ε_ν° of the initial model, thanks to solving Problem 7.2. Finally, it can be observed that the error bounds are conservative, due to several approximation steps in the derivations [31]. Nevertheless, this numerical case study shows that reducing the error bound also results in a reduced mismatch in the steady-state responses, as observed in Figures 7.4 and 7.5.

The matrix inequalities characterizing the set $\mathcal{G}_\gamma$ in Theorem 7.3 and $\tilde{\mathcal{G}}_\gamma$ in Theorem 7.6 can be considered high-dimensional as these contain the matrices of the LTI part of the full-order model. Nevertheless, the example in this section took less than 30 minutes to complete on an Intel Core i7-7700HQ, 2.8 GHz processor, successfully reducing the LTI part from a $n = 48$-th order model to a $\nu = 5$-th order model.

7.5.2 LTI example highlighting the benefits of Algorithm 7.3

This example demonstrates the benefits of Algorithm 7.3 on a SISO LTI example. The considered full-order LTI model with transfer function Φ has state dimension $n = 20$ and is already balanced. We apply the reduction method presented in Section 7.2 to find reduced-order models with state dimensions $\nu = 2, 4$, and 6 with corresponding transfer function Γ.

The magnitude of the FRF of the full-order model is depicted in Figure 7.7. The interpolation frequencies are selected according to $\omega_1 = 2\pi 2.24$ rad/sec and $\omega_2 = 2\pi 0.82$ rad/sec to obtain a match at the two largest resonance peaks, and $\omega_3 = 2\pi 0.01$ rad/sec to capture low-frequency behavior, resulting in the interpolation points $\{s_1, \ldots, s_6\} = \{\pm j\omega_1, \pm j\omega_2, \pm j\omega_3\}$. The frequencies $\omega_1, \omega_2, \omega_3$ are marked with a cross in Figure 7.7. For the $\nu = 2$-dimensional reduced-order model, we only match s_1, s_2 corresponding to the largest resonance peak, for the $\nu = 4$-dimensional model, we match s_1, s_2, s_3, s_4 corresponding to the two largest resonance peaks, and for the $\nu = 6$-dimensional model, we match all $s_i, i = 1, \ldots, 6$. For each case of ν, we define the matrix S according to (7.9a) and we take

$$L = \begin{bmatrix} 1 & 0 \end{bmatrix}, L = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}, L = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix},$$

for the $\nu = 2, 4, 6$ cases, respectively.

Next, for each case of ν, we solve equation (7.12) for the unique Π and define F and H according to (7.11), which defines the set $\mathcal{G}_M$. From the set $\mathcal{G}_M$, we select an initial $G^{[0]} \in \mathcal{G}_{\gamma^{[0]}}$ by pole placement. Hereto, we choose the placement locations as a subset of the eigenvalues of A, namely: $p_1 = \{-0.337 \pm 14.053j\}, p_2 = \{-0.798 \pm 5.2129j\}, p_3 = \{-0.225, -10.490\}$. To assess sensitivity to the pole locations of the initial reduced-order model, we run Algorithm 7.3 multiple times for each ν, each time starting with different initial locations as indicated in Table 7.1, where for 'Random', the pole locations are drawn from a normal distribution and mirrored

Table 7.1. Considered studies for the case study in Section 7.5.2.

Study	ν	$h_{\nu+1}$	$\sigma\left(S - G^{[0]}L\right)$	$\gamma^{[0]}$	γ_{final}
1	2	0.777	p_1	1.250	1.195
2			p_2	44.084	1.218
3			p_3	21.348	1.225
4	4	0.321	p_1 and p_2	1.244	0.557
5			p_1 and p_3	11.628	0.550
6			p_2 and p_3	16.236	0.557
7	6	0.038	p_1, p_2 and p_3	0.797	0.066
8			Random	14.943	0.066

to $\mathbb{C}^-$. The approximation error in the $\mathcal{H}_\infty$ norm of the initial model with $G^{[0]}$ is included in Table 7.1 in the column $\gamma^{[0]}$.

Figure 7.7 depicts the magnitude of the FRF Φ of the full-order model and the magnitude of the FRFs Γ of the reduced-order model for studies $1, 4, 7$. The FRFs obtained in Studies 2 and 3, Studies 5 and 6, and Study 8 are left out of this figure as they are similar to the FRFs in Study 1, Study 4, and Study 7, respectively. It can be seen that for any ν, the error $|\Phi(j\omega) - \Gamma(j\omega)|$ is 0 at the corresponding interpolation frequencies ω_1, ω_2, and ω_3, i.e., for the $\nu = 2$ case (Study 1) at ω_1, for the $\nu = 4$ case (Study 4) at ω_1, ω_2, and for the $\nu = 6$ case (Study 7) at $\omega_1, \omega_2, \omega_3$. This is a characteristic property of moment matching in the SISO case. Secondly, it can be observed that increasing ν yields a reduced-order model that significantly more accurately matches the FRF of the full-order model.

The approximation error after application of Algorithm 7.3 is included in Table 7.1 in the column γ_{final}. Firstly, as expected, a larger ν allows for a reduced approximation error thanks to the increased model flexibility. Secondly, it can be seen that in all cases, the approximation error is significantly reduced from $\gamma^{[0]}$ to γ_{final}; in all cases within a factor 2 of the fundamental lower bound $h_{\nu+1}$, i.e., $\gamma_{\text{final}} < 2 \cdot h_{\nu+1}$. Thirdly, it can be seen that in all considered cases, Algorithm 7.3 is robust for the initial starting $G^{[0]}$, as it converges to the same level of approximation error in terms of γ for any of the tested initial starting $G^{[0]}$. Given the fact that the approximation error $\|\Phi(s) - \Gamma(s)\|_{\mathcal{H}_\infty}$ is relatively close to the conservative lower bound $h_{\nu+1}$ in all cases and observing the robustness of Algorithm 7.3 against the initial starting $G^{[0]}$, we conclude that the proposed method is effective. Nevertheless, studying the properties of Algorithm 7.3 in terms of convergence and robustness for the starting point $G^{[0]}$ remains a topic for future work. A second conclusion is that placing the poles of the initial reduced-order model to a subset of $\sigma(A)$ is also effective. For example, placing the poles of the initial reduced-order model at p_1 captures the largest resonance peak.

The iteration history for γ is depicted in Figure 7.8 for each ν. Based on this figure, we would like to highlight two important observations for the considered cases.

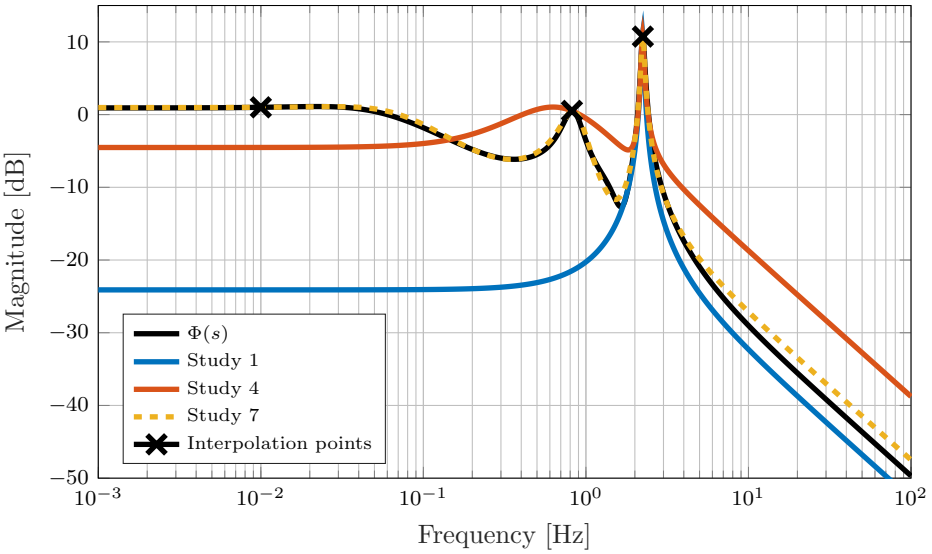


Figure 7.7. Bode magnitude plot of the full-order model Φ and the reduced-order model Γ of Studies 1, 4 and 7 of Table 7.1.

Firstly, if we have a reasonable initial $G^{[0]}$, then the algorithm converges within 100 iterations, for which the computational time is approximately 10 minutes (on an Intel Core i7-7700HQ, 2.8 GHz processor). Secondly, the switching nature of Algorithm 7.3 helps to reduce γ. For example, in Study 8, several switches are made between the CDAs in Algorithms 7.1 and 7.2, when the corresponding algorithm failed to reduced γ further. Thus, in the considered cases, Algorithm 7.3 avoids getting stuck prematurely at suboptimal reduced-order models by switching between the different CDAs.

The switching nature of Algorithm 7.3 is further investigated for the $\nu = 2$ case (reduction to a second-order model) with $G \in \mathbb{R}^{\nu}$. Figure 7.9 depicts the surface of γ over a fine grid of the two entries of G, denoted by $G(1), G(2)$, where flat surface for large γ indicates an unstable reduced-order model. The trajectories of γ in Studies 1, 2 and 3 in Table 7.1 are depicted in Figure 7.9 and the switching points are indicated by yellow circles. From this figure, it is clear that there exists a unique G_{opt} corresponding to γ_{opt} inside the inspected range. It can be observed that the initial point $G^{[0]}$ of Study 2 and Study 3 is chosen relatively close to G_{opt}, which explains the relatively low number of iterations required upon convergence, see Figure 7.8. The initial point $G^{[0]}$ of Study 3, however, is chosen relatively far away and requires, therefore, a relatively large number of iterations to converge. The trajectory of Study 3 also reveals that the search direction is changed significantly at some of the switching points represented by the yellow

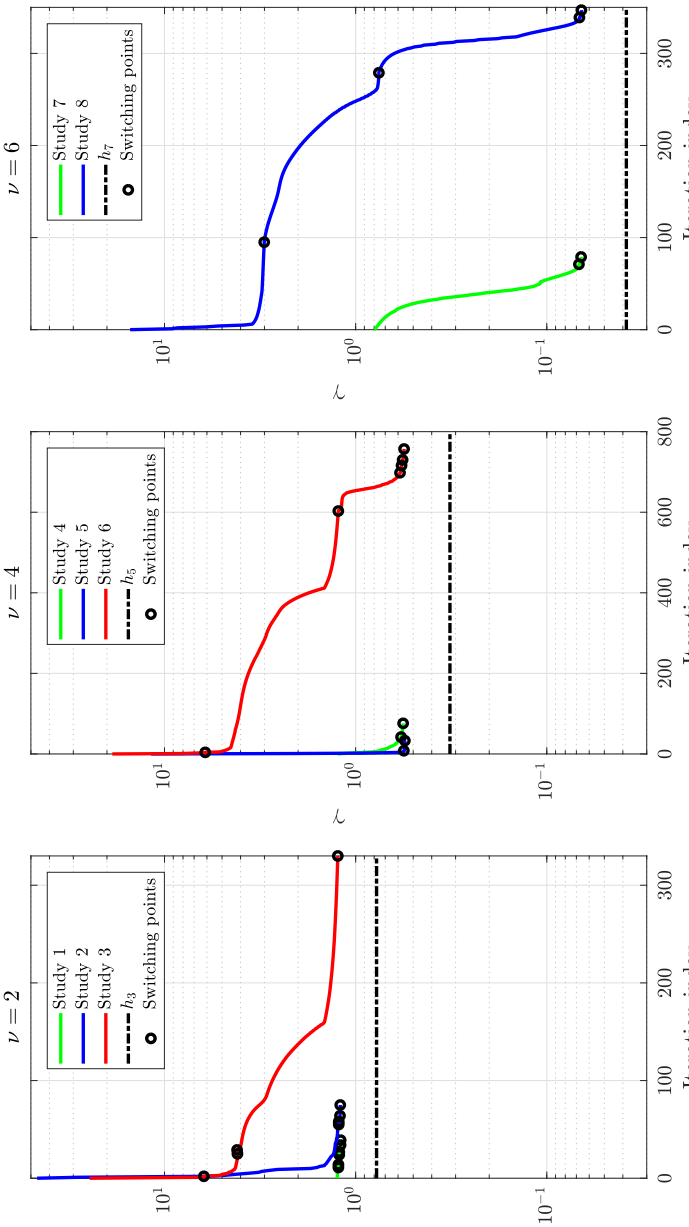


Figure 7.8. Iteration history of γ for the considered studies in Table 7.1 (left: $\nu = 2$, middle: $\nu = 4$, right: $\nu = 6$). The dashed black line corresponds to the lower bound based on the Hankel singular values. The circles correspond to switches in Algorithm 7.3 between Algorithms 7.1 and 7.2.

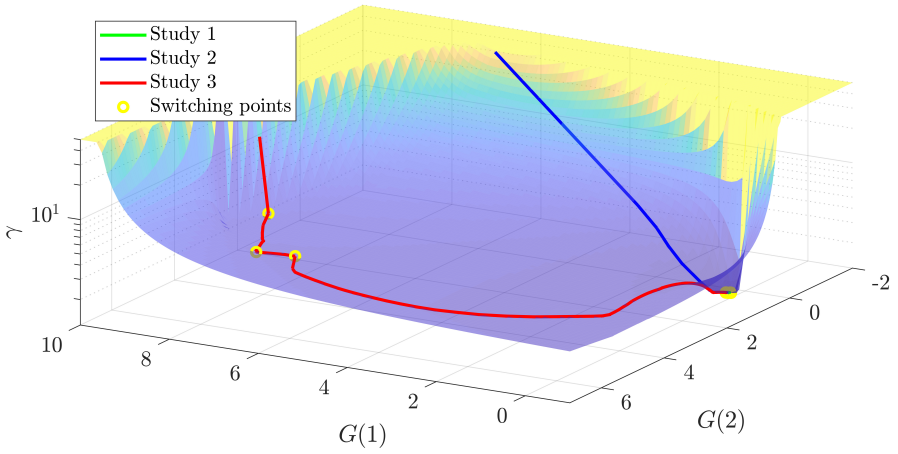


Figure 7.9. Surface of γ over a grid of $G(1)$ and $G(2)$ for the Studies 1, 2 and 3 as considered in Table 7.1. The yellow circles correspond to switches in Algorithm 7.3 between Algorithms 7.1 and 7.2.

circles, enabling further minimization of γ after getting stuck prematurely.

7.6 Conclusions

This chapter has presented a model reduction approach based on time-domain moment matching. In the LTI case, the approach exploits parametric freedom to minimize the $\mathcal{H}_\infty$-norm of the transfer function matrix of the error dynamics in addition to achieving moment matching along tangential directions. The extension to the nonlinear case considers nonlinear models that consist of LTI dynamics that are placed in feedback with nonlinear dynamics and only reduces the state dimension of the LTI dynamics. The proposed approach minimizes an error bound on nonlinear steady-state model responses. Furthermore, both the model structure and the input-to-state convergence stability properties of the full-order nonlinear model are preserved.

III

Kernel-based identification of nonlinear systems

8

State-space kernelized closed-loop identification of nonlinear systems

In this chapter, we propose a kernel-based non-parametric state-space identification approach for open-loop and closed-loop discrete-time nonlinear systems with multiple inputs and multiple outputs. Employing a least squares support vector machine (LS-SVM) approach in a reproducing kernel Hilbert space framework, a nonlinear auto-regressive model with exogenous terms is identified to provide a non-parametric estimate of the innovation noise sequence. Subsequently, this estimate is used to obtain a compatible non-parametric estimate of the state sequence in an unknown basis using kernel canonical correlation analysis. Finally, the estimated state sequence is used together with the estimated innovation noise sequence to find a non-parametric state-space model, again using an LS-SVM approach. The performance of the approach is analyzed in simulation studies with a nonlinear system operating both in open loop and closed loop. The identification approach can be viewed as a nonlinear counterpart of consistent subspace identification techniques for linear time-invariant systems operating in closed loop.

The contents of this chapter are published in: Shakib, M. F., Tóth, R., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. (2020). State-space kernelized closed-loop identification of nonlinear systems. *IFAC-PapersOnLine*, 53(2), 1126-1131.

8.1 Introduction

Identification of nonlinear systems is a challenging and active field of research [54, 243]. One aspect that makes it challenging is the variety of systems, which has led to a large number of model classes proposed in the literature. A generic model class is the class of nonlinear state-space models. This model class is attractive as it is particularly suitable for the parsimonious representation of multiple-input multiple-output (MIMO) systems. Furthermore, many analysis and controller design tools exist for this class of models, see [135].

A discrete-time nonlinear state-space (NL-SS) model is characterized by its state-transition map relating the current state and inputs to the state at the next time stamp, and its output map relating the current state and inputs to the output of the model. As there is a recursion loop in the evolution of the hidden state variable, identification of these mappings is a challenging task [162]. Many NL-SS identification methods exist that are based on direct identification of these mappings using specific parametric model structures. However, the identification corresponds to a computationally-demanding nonlinear optimization problem with the need for efficient initialization and a model parametrization to be provided by the user, see [96, 243]. Alternatively, non-parametric identification techniques exist for the identification of nonlinear input-output models, see [203]. However, for control, NL-SS models are favored over nonlinear input-output models as many design techniques use Lyapunov's second method, which requires a state-space model representation. Unfortunately, state-space realization of nonlinear input-output models is a difficult task with many unsolved problems, see [143].

If the state sequence of the underlying NL-SS model would be available, for example, by means of additional measurements, the identification of the state-space mappings would become a static problem, which simplifies the identification problem significantly. However, the state sequence is often unavailable, whilst obtaining an accurate estimate of it is non-trivial. For linear time-invariant (LTI) systems, subspace techniques, see [147] and [293], can be employed for the estimation of the state sequence. For nonlinear systems, to this extent, [162] proposed a method to approximate the state sequence by, in the first step, identifying a linear model, and, in the next step, minimizing a linear cost function yielding an approximation of the state sequence. Alternatively, [298] proposed a method based on a least-squares support vector machine (LS-SVM) approach to obtain a state sequence, which solves an intersection problem between future and past input/output data using a kernelized version of canonical correlation analysis (CCA).

The method of [298] estimates the state sequence of the underlying NL-SS model. However, in that work, the effect of noise, being inevitably present in practice, was not taken into account, which can result in biased identification results. To this extent, we relate to the consistency property of an estimator, which, loosely speaking, ensures that the true model is recovered if the number of data points tends to infinity [152]. Consistent subspace techniques for LTI systems

operating in closed loop commonly rely on the identification of a consistent noise model first using an auto-regressive model with exogenous inputs (ARX), see, e.g., [288]. Then, in the next steps, the identified noise model is used in the consistent estimation of the state sequence and the state-space matrices (up to a similarity transformation). In particular, we highlight the SSARX technique, see [124], which uses a three-step approach. In the first step, a consistent ARX model is estimated to obtain a one-step-ahead prediction model of the output, which is an aggregated form of the system and its noise model. In the second step, the state sequence is obtained by performing CCA to infer the state sequence as an intersection of future and past input/output data. The third step entails the estimation of the state-space matrices, which is formulated as a problem that is linear in the parameters and solved using linear least squares.

In this chapter, we extend the methodology of SSARX for the identification of LTI systems to the case of nonlinear systems. In the first step, rather than identifying an ARX model, a nonlinear ARX (NARX) predictor model is identified by an LS-SVM approach, which is proven to be consistent, see [58]. The prediction error of this NARX model serves as an estimate of the innovation noise sequence, similar to [170] for the LTI case. Next, in the second step, we use the estimated innovation noise sequence as an additional pseudo-input to estimate the state sequence using the kernelized CCA method presented in [298]. After that, having also an estimate of the state sequence at hand (in an unknown state basis), in the third and final step, we identify the state-transition map and the output map of the NL-SS model non-parametrically, again using an LS-SVM approach. The resulting model is characterized by the state-transition map and output map, both given non-parametrically. An overview of the proposed approach is given in Figure 8.1.

Although we do not give an overall consistency proof, each step of the proposed procedure corresponds to a consistent estimate under the assumption that the true noise and state sequence are provided in the previous steps. Simulation studies, presented in Section 8.5, show that the proposed identification strategy, both in the open-loop and closed-loop case, outperforms the case where noise is not taken into account in the identification process, i.e., direct application of the method of [298]. This simulation result demonstrates that using the estimated innovation noise sequence obtained by the identified NARX model, significantly improves the quality of the identified model. Future work aims at proving rigorously that the proposed method is consistent, even in the closed-loop case.

The remainder of this chapter is organized as follows. In Section 8.2, the identification problem is formally introduced. The concept of function approximation, used in the steps of the identification strategy, is described in Section 8.3. The overall identification approach is presented in Section 8.4. Simulation examples are given in Section 8.5. Section 8.6 presents the conclusions of this chapter.

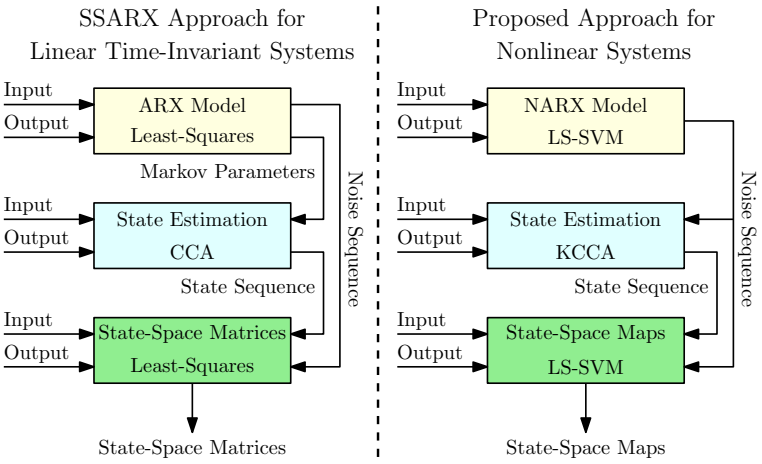


Figure 8.1. Existing SSARX approach [124] for LTI systems (left). Proposed identification approach for nonlinear systems (right).

8.2 Problem formulation and notation

We consider MIMO discrete-time nonlinear systems that can be represented by the following set of first-order difference, i.e., state-space, equations:

$$x_{k+1} = f(x_k, u_k, e_k), \tag{8.1a}$$
$$y_k = h(x_k) + e_k, \tag{8.1b}$$

where, at time instance k, the state is denoted by $x_k \in \mathbb{R}^n$, the input by $u_k \in \mathbb{R}^m$, and the output by $y_k \in \mathbb{R}^l$. The functions f and h are called the state-transition map and output map, respectively. The innovation noise sequence $e_k \in \mathbb{R}^l$, is assumed to be drawn from a zero mean normal distribution with finite diagonal covariance matrix Σ_e. The problem we consider is to identify the state dimension n, to identify the functions f and h non-parametrically, and to give an estimate of the noise variance Σ_e, based on a data-set $\mathcal{D} = \{u_k, y_k\}_{k=1}^N$ generated by (8.1), where N is the number of data points. Conditions imposed on the mappings f and h are given in Section 8.4, where the proposed identification method is presented.

As many systems are unstable by themselves, or only a part of the dynamics of a more complex system is to be identified, we also consider systems that operate in closed loop. However, to avoid the existence of an algebraic loop, i.e., the output not being uniquely determinable, we require the assumption that either the plant or the controller (or both) has (have) no feedthrough. Such an assumption is also commonly adopted in the LTI case, see [288]. Here, for the sake of notational simplicity, we assume that the plant has no direct feedthrough, i.e., the mapping h is not a function of u_k in (8.1b). This implies that the covariance matrix $\mathbb{E}\left\{u_k e_j^\top\right\}$,

where $\mathbb{E}$ denotes the expectation w.r.t. the random variables u_k and e_k, is a zero-matrix if $j > k$, but can be non-zero for $j \leq k$ in the closed-loop setting, which typically results in a bias if the noise is not handled appropriately during identification.

In the sequel, the following notation for vectors of shifted sequences of inputs is used: $\bar{u}_k^d \coloneqq \begin{bmatrix} u_k^\top & \dots & u_{k+d-1}^\top \end{bmatrix}^\top$. Similarly, vectors of shifted outputs and shifted noise are denoted by $\bar{y}_k^d$ and $\bar{e}_k^d$, respectively. Furthermore, iterative evaluations of mapping f with respect to x_k are denoted as follows:

$$\begin{aligned} f_k(x_k) &\coloneqq f(x_k, u_k, e_k) = x_{k+1}, \\ f^d(x_k, \bar{u}_k^d, \bar{e}_k^d) &\coloneqq f_{k+d-1} \circ \dots \circ f_{k+1} \circ f_k(x_k). \end{aligned} \tag{8.2}$$

Finally, the vector of sequential outputs is defined as:

$$\bar{y}_k^d = h^d(x_k, \bar{u}_k^{d-1}, \bar{e}_k^d) \coloneqq \begin{bmatrix} h(x_k) + e_k \\ h \circ f^1(x_k, \bar{u}_k^1, \bar{e}_k^1) + e_{k+1} \\ \vdots \\ h \circ f^{d-1}(x_k, \bar{u}_k^{d-1}, \bar{e}_k^{d-1}) + e_{k+d-1} \end{bmatrix}. \tag{8.3}$$

8.3 Function estimation using an LS-SVM approach

In this section, we describe an approach to non-parametric function estimation. Such function estimation is used in Section 8.4.1 to identify a NARX model and in Section 8.4.3 to identify the mappings f and h of system (8.1).

The core concept of function estimation is to search for a function inside a Hilbert space $\mathcal{H}$, which is equipped with an inner product $\langle \cdot, \cdot \rangle$ and is complete with respect to the induced norm $\|g\|_{\mathcal{H}} \coloneqq \sqrt{\langle g, g \rangle}$. Given the data-set $\mathcal{D} = \{z_k, w_k\}_{i=1}^N$, where $w \in \mathbb{R}$ is the set of observed outputs and $z \in \mathbb{Z} \subseteq \mathbb{R}^{n_z}$ the set of (observed) inputs, the search goal is to find the function g that minimizes the cost functional

$$\min_{g \in \mathcal{H}} \left[\frac{\gamma}{2N} \sum_{i=1}^{N} (w_k - g(z_k))^2 + \frac{1}{2} \|g\|_{\mathcal{H}}^2 \right]. \tag{8.4}$$

In (8.4), the first term penalizes mismatch in data-fit and the second term penalizes function complexity measured in a Hilbert space norm, which acts as regularization with positive regularization parameter γ. This (continuous) regularization parameter can be viewed as the counterpart of the (discrete) tunable parameter that defines, e.g., the model order in classical parametric system identification approaches. As the data-set $\mathcal{D}$ only contains N samples, the optimization problem (8.4) is only well-posed if the search is restricted to the reproducing kernel Hilbert space (RKHS) over $\mathbb{Z}$. This is a space of functions $g : \mathbb{Z} \to \mathbb{R}$ that satisfy the following boundedness criterion:

$$\forall g \in \mathcal{H} \text{ and } \forall z \in \mathbb{Z}, \quad \exists\, 0 \leq c < \infty : |g(z)| < c \|g\|_{\mathcal{H}}. \tag{8.5}$$

The theorem of Moore-Aronszajn, see [14], ensures a one-to-one correspondence between RKHS of functions $\mathcal{H}$ over $\mathbb{Z}$ and symmetric positive semidefinite[1] kernel functions $K : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$. It ensures that for every RKHS $\mathcal{H}$ satisfying (8.5), a unique symmetric reproducing kernel function $K : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ exists which is positive semidefinite and obeys the reproducing property $g(z) = \langle g(\cdot), K(\cdot, z)\rangle$. The optimization problem in (8.4) has a closed-form solution by means of the Representer theorem, see [139]. In particular, for the RKHS $\mathcal{H}$, the minimizer $\hat{g}$ of (8.4) is given by

$$\hat{g}(\cdot) = \sum_{i=1}^{N} \hat{\alpha}_i K_{z_i}(\cdot) \tag{8.6}$$

with $K_{z_i}(\cdot) \coloneqq K(\cdot, z_i)$ and $\hat{\alpha} = \begin{bmatrix}\hat{\alpha}_1 & \dots & \hat{\alpha}_N\end{bmatrix} \in \mathbb{R}^N$ being given by

$$\hat{\alpha} = \left(\frac{1}{N}\mathcal{K}_{zz} + \gamma^{-1} I_N\right)^{-1} \frac{1}{N} W,$$

where $W = \begin{bmatrix}w_1 & \dots & w_N\end{bmatrix}^\top$, $I_N \in \mathbb{R}^{N\times N}$ is an identity matrix and $\mathcal{K}_{zz} \in \mathbb{R}^{N\times N}$ is the Gram matrix defined by $\mathcal{K}_{zz}(i,j) \coloneqq K(z_i, z_j)$. The RKHS optimal function estimator (8.6) is known in the literature as the LS-SVM approach for function estimation.

The quality of the data-fit depends on the selected kernel function $K(\cdot,\cdot)$, the hyper-parameters defining the kernel function, and the regularization parameter. A typical choice is the radial basis function (RBF) kernel: $K(z_i, z_j) = \exp\left\{\left(-\|z_i - z_j\|_2^2/\sigma^2\right)\right\}$, where $\sigma > 0$ is tunable hyper-parameter. The selection of the kernel function is rather case-specific, where, among others, linear, polynomial, rational, spline, and wavelet kernel functions are proposed in the literature, see [233]. The hyper-parameters of the kernel and the regularization parameter γ can be tuned in various ways, e.g., by maximizing the so-called log marginal likelihood function, see [305]. The marginal likelihood function expresses the likelihood that the mapping $\hat{g}$ maps inputs z to observations w.

Remark 8.1. If $g : \mathbb{Z} \to \mathbb{R}^{n_g}$ is multidimensional, i.e., $n_g > 1$, then n_g individual functions $g_{(i)}(\cdot)$, are estimated and concatenated to form the n_g-dimensional function $g(\cdot) = \begin{bmatrix}g_{(1)}(\cdot) & \dots & g_{(n_g)}(\cdot)\end{bmatrix}^\top$. △

8.4 Identification approach

This section presents the three-step identification approach. First, a NARX model is estimated in Section 8.4.1 to obtain an estimate of the innovation noise sequence e_k. Next, the state sequence is estimated in Section 8.4.2. Finally, the mappings f and h are estimated in Section 8.4.3.

[1] $K : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ is positive semidefinite, if $\forall n \in \mathbb{N}$, $\sum_{i=1}^{n}\sum_{j=1}^{n} c_k K(z_k, z_j) c_j \geq 0, \forall \{z_k, c_k\}_{k=1}^{n} \in \mathbb{Z} \times \mathbb{R}$.

8.4.1 Consistent noise sequence estimation

A nonlinear auto-regressive model with exogenous terms (NARX) is used to estimate the innovation noise sequence. A NARX model is an input-output model that can be written in the following form:

$$y_k = f_{\text{NARX}}(z_k) + e_k, \tag{8.7a}$$

$$z_k = \begin{bmatrix} y_{k-1}^\top & \dots & y_{k-p}^\top & u_{k-1}^\top & \dots & u_{k-p}^\top \end{bmatrix}^\top. \tag{8.7b}$$

In the NARX model (8.7a), the output $y_k \in \mathbb{R}^l$ at time k is a function of previous outputs y_{k-i} and inputs u_{k-i} for $i = 1, \dots, p$, where p is the past window length.

Let us rewrite system (8.1) to a NARX model of the form (8.7a). First, notice that (8.1a) can be written in the predictor form by substituting $e_k = y_k - h(x_k)$ in (8.1a):

$$x_{k+1} = f(x_k, u_k, y_k - h(x_k)) =: \tilde{f}(x_k, u_k, y_k), \tag{8.8}$$

Next, following the notation of (8.2), x_{k+p} can be written as $x_{k+p} = \tilde{f}^p(x_k, \bar{u}_k^p, \bar{y}_k^p)$. At this stage, in the LTI case, under the assumption that a stable observer exists, it can be shown that for $p \to \infty$, the effect of x_k in x_{k+p} diminishes completely, see [316]. In fact, this assumption is not only required to transform a state-space model into an ARX model but is also a fundamental assumption in *any* subspace identification algorithm, see [288]. In the context of subspace identification, it is reasonable to take a sufficiently large p, such that the influence of state x_k in x_{k+p} is negligibly small, see [124]. A similar assumption is required in the nonlinear case as well. To this end, we assume that the effect of x_k on $\tilde{f}^p(x_k, \bar{u}_k^p, \bar{y}_k^p)$ is negligible as follows.

Assumption 8.1 (Fading memory)**.** *The NL-SS model* (8.1) *satisfies the following condition:*

$$\begin{aligned} &\forall r_x, \forall r_u, \forall r_y, \forall x_0, \forall u_0, \forall y_0, \forall \epsilon, \exists P \text{ s.t.} \\ &\quad \forall x_k, \tilde{x}_k \in \mathcal{B}_{x_0}^{r_x}, u_k \in \mathcal{B}_{u_0}^{r_u}, y_k \in \mathcal{B}_{y_0}^{r_y}, p > P \text{ ensures} \\ &\qquad \left\| \tilde{f}^p(x_k, \bar{u}_k^p, \bar{y}_k^p) - \tilde{f}^p(\tilde{x}_k, \bar{u}_k^p, \bar{y}_k^p) \right\|_2 \leq \epsilon \end{aligned}$$

with the ball $\mathcal{B}_{z_0}^r := \{z \in \mathbb{R}^{n_r} : \|z - z_0\|_2 < r\}$. ■

This assumption implies that the state can be written as a function of only the p past values of the input and output, up to an error term $\Delta(\epsilon)$:

$$x_k = F(\bar{u}_{k-p}^p, \bar{y}_{k-p}^p) + \Delta(\epsilon),$$

where $F(\bar{u}_{k-p}^p, \bar{y}_{k-p}^p) := \tilde{f}^p(0, \bar{u}_k^p, \bar{y}_k^p)$. By Assumption 8.1, the error $\Delta(\epsilon)$ can be bounded as $\|\Delta(\epsilon)\|_2 \leq \epsilon$, which implies that the finite memory approximation

$$x_k \approx F(\bar{u}_{k-p}^p, \bar{y}_{k-p}^p). \tag{8.9}$$

can be made arbitrarily accurate by taking p sufficiently large. Therefore, in the sequel, we assume that $\Delta(\epsilon) = 0$. Using (8.9) in (8.1b) results in $y_k = h(F(\bar{y}^p_{k-p}, \bar{u}^p_{k-p})) + e_k = f_{\mathrm{NARX}}(z_k) + e_k$, which is of the form (8.7a).

For the estimation of f_{NARX}, the function estimator described in the previous section is employed. After defining the window length p, the kernel function $K_{\mathrm{NARX}}(\cdot,\cdot)$ and having optimized its hyper-parameters and the regularization parameter γ, the estimate $\hat{f}_{\mathrm{NARX}}$ is given by (8.6), where the input data z_k is as defined in (8.7b) and, as output data, $w_k := y_k$ is used. The estimate of the innovation noise sequence is then given by

$$\hat{e}_k = y_k - \hat{f}_{\mathrm{NARX}}(z_k). \tag{8.10}$$

Assuming $\hat{e}$ is zero mean, the empirical estimate for the noise covariance matrix $\hat{\Sigma}_e$ is given by $\hat{\Sigma}_e = \frac{1}{N}\sum_{k=1}^{N} \hat{e}_k \hat{e}_k^\top$.

This type of NARX modeling is analyzed in [58]. There, it is shown that if the input z_k is uncorrelated with e_k, and if some other mild technical conditions on the mapping f hold, then the estimate (8.6) is consistent, i.e.,

$$\lim_{N\to\infty} \mathbb{E}\left\{ \left\| f_{\mathrm{NARX}} - \hat{f}_{\mathrm{NARX}} \right\|^2_{\mathcal{H}_{\mathrm{NARX}}} \right\} = 0. \tag{8.11}$$

As e_k is uncorrelated with z_k (even in the closed-loop case due to the absence of a feedthrough term in $h(x_k)$ in (8.1b)), the estimate of the innovation noise sequence in (8.10) is also consistent, i.e., the true innovation noise sequence is recovered if the number of data points tends to infinity. Based on this estimate, we define the extended data-set $\tilde{\mathcal{D}} := \{u_k, y_k, \hat{e}_k\}_{k=p+1}^{N}$.

The consistent estimation of the innovation noise sequence is key in our three-step approach as it allows us to perform the subsequent two steps, namely, estimation of the state sequence and estimation of the state-space mappings, also in a consistent manner. In the LTI case, it is already shown that the crucial step of estimating a noise model is essential to prove consistency for the subsequent steps in the closed-loop case, see [288]. Although we do not prove that our overall identification procedure is consistent, simulation examples show that the quality of the model fit improves significantly when using the estimate $\hat{e}$ in the next identification steps, compared to the case where the noise sequence is not used.

8.4.2 State sequence estimation

Having the data-set $\tilde{\mathcal{D}}$ at hand, the goal in this section is to provide a method to estimate the state sequence. In this respect, [298] proposed a method to do this based on input and output data. However, in that work, noise was neglected, which could lead to poor model quality in case noise is present, especially in the closed-loop case. Considering the data-set $\tilde{\mathcal{D}}$, we view the estimated innovation noise sequence as an additional pseudo-input. In that way, we can use the method

proposed in [298] to find an estimate of the state sequence. Furthermore, a form of consistency can be proven for this state sequence estimate, assuming that the true innovation noise sequence was obtained in the previous step. What follows is a brief recap of the approach of [298] adapted to our notation.

The CCA method, introduced by [114], is a statistical method to study the linear relations between sets of variables. It is the foundation of the so-called intersection-based subspace algorithms for LTI systems, which rely on the state sequence being a minimal interface between past and future input and output data. In those algorithms, CCA is applied to find the state sequence as a linear combination of the past data, such that it optimally predicts the future data. By introducing a kernel function, linear CCA can be performed in a kernel feature space, hence the name kernel CCA (KCCA). KCCA transforms the sets of variables *nonlinearly* in order to find their maximal correlation.

The basic idea is to, first, write the state x_k as a function of, on the one hand, past inputs and outputs $\bar{\phi}_k^d \coloneqq \left[(\bar{y}_{k-d}^d)^\top \quad (\bar{u}_{k-d}^d)^\top \quad (\bar{e}_{k-d}^d)^\top\right]^\top$ and, on the other hand, future inputs and outputs $\bar{\phi}_{k+d}^d$, i.e., $x_k = \Phi_{pp}\left(\bar{\phi}_k^d\right) = \Phi_{ff}\left(\bar{\phi}_{k+d}^d\right)$ where d is the window length. The existence of mapping $\Phi_{pp}\left(\bar{\phi}_k^d\right)$ is ensured by Assumption 8.1, see (8.9). To support the existence of mapping $\Phi_{ff}\left(\bar{\phi}_{k+d}^d\right)$, the notion of strong local observability is adopted.

Definition 8.1 ([182])**.** *System* (8.1) *is strongly locally observable at* (x_k, u_k, e_k) *if*

$$rank\left(\frac{\partial h^n(x_k, \bar{u}_k^{n-1}, \bar{e}_k^n)}{\partial x_k}\right) = n.$$

■

It is assumed that (8.1) has a fixed-point (x^0, u^0, e^0, y^0), such that $x^0 = f(x^0, u^0, e^0)$ and $y^0 = h(x^0) + e^0$. Next, the following observability assumption is posed.

Assumption 8.2. *System* (8.1) *is strongly locally observable according to Definition 8.1 at the equilibrium* (x^0, u^0, e^0, y^0)*.* ■

Assumption 8.2 ensures that the state x_k can be uniquely determined from a finite sequence of future input vectors $\bar{u}_k^{d-1}, \bar{e}_k^d$ and output vectors $\bar{y}_k^d$, in the neighborhood of the equilibrium.

By formulating a KCCA problem on past data $\bar{\phi}_k^d$ and future data $\bar{\phi}_{k+d}^d$ using an LS-SVM approach, we arrive at the regularized generalized eigenvalue problem (RGEP)

$$\begin{bmatrix} \nu_p \mathcal{K}_{pp} + I & 0 \\ 0 & \nu_f \mathcal{K}_{ff} + I \end{bmatrix} \begin{bmatrix} \eta \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & \mathcal{K}_{ff} \\ \mathcal{K}_{pp} & 0 \end{bmatrix} \begin{bmatrix} \eta \\ \kappa \end{bmatrix} \Lambda, \tag{8.12}$$

where Λ is the diagonal matrix containing the eigenvalues, ν_p and ν_f are the regularization parameters (to be optimized), and $\mathcal{K}_{pp}$ and $\mathcal{K}_{ff}$ are the Gram matrices

whose elements are evaluations of the selected kernel function $K_{pp}(\cdot,\cdot)$ and $K_{ff}(\cdot,\cdot)$ on the past data $\bar{\phi}_k^d$ and future data $\bar{\phi}_{k+d}^d$, respectively. The state dimension n can be estimated by the number of dominant eigenvalues values in the RGEP (8.12). Solving this RGEP results in the canonical vectors η and κ. Subsequently, the estimate for the state sequence obtained from future data is $\hat{x} = \kappa_{1:n}^\top K_{ff}$, where $\kappa_{1:n} = \begin{bmatrix}\kappa_1 & \dots & \kappa_n\end{bmatrix}$ and n is the state dimension. Similarly, the estimate for the state obtained from past data is $\check{x} = \eta_{1:n}^\top K_{pp}$. The sequences $\hat{x}$ and $\check{x}$ are estimates of x in an unknown nonlinearly transformed state basis. With the state estimate at hand, the data-set $\tilde{\mathcal{D}}$ is extended with the estimated $\hat{x}_k$ and denoted by $\hat{\mathcal{D}} = \{u_k, y_k, \hat{e}_k, \hat{x}_k\}_{k=p+d+1}^{N-d}$. The regularization parameters ν_p, ν_f, and the hyper-parameters of the kernel functions $K_{pp}(\cdot,\cdot)$ and $K_{ff}(\cdot,\cdot)$ can again be tuned in various ways. In general, this is a nonlinear optimization problem that can be solved, for example, by a grid search.

Consistency of the estimation of the state sequence via this KCCA approach is claimed by [87] and is understood in the sense that the so-called regularized $\mathcal{F}$-correlation is consistently estimated if the number of data points tends to infinity. Hereto, it is required that the mappings $\Phi_{pp}(\cdot)$ and $\Phi_{ff}(\cdot)$ belong to the RKHS $\mathcal{H}_{pp}$ and $\mathcal{H}_{ff}$, defined through the kernel functions $K_{pp}(\cdot,\cdot)$ and $K_{ff}(\cdot,\cdot)$, respectively. The consistency claim only holds if the true innovation noise sequence is recovered in the previous step.

8.4.3 State transition and output mappings estimation

Once the estimate $\hat{x}$ for the state is available, the identification of the state transition map f and the output map h of (8.1) becomes a static problem. Hereto, again, the function estimator introduced in Section 8.3 is employed to identify the mappings f and h non-parametrically based on the data-set $\hat{\mathcal{D}}$.

For the identification of f, the input $z_k \coloneqq \begin{bmatrix}\hat{x}_k^\top & u_k^\top & \hat{e}_k^\top\end{bmatrix}^\top$ and output $w_k \coloneqq \hat{x}_{k+1}$ are selected. Then, after defining the kernel function $K_f(\cdot,\cdot)$ and tuning its hyper-parameters and the regularization parameter, the identified $\hat{f}$ is given by (8.6). The mapping h is identified using the input $z_k \coloneqq \hat{x}_k$ and the corrected output $w_k \coloneqq y_k - \hat{e}_k$. Again, after defining the kernel function $K_h(\cdot,\cdot)$ and tuning its hyper-parameters and the regularization parameter, the identified output map $\hat{h}$ is given in (8.6).

The identified NL-SS model is then given by

$$x_{k+1} = \hat{f}(x_k, u_k, e_k), \quad y_k = \hat{h}(x_k) + e_k, \tag{8.13}$$

where the mappings $\hat{f}$ and $\hat{h}$ are non-parametric. This non-parametric model is characterized by the data-set $\hat{\mathcal{D}}$, the kernel functions K_f and K_p, their associated hyper-parameters and the regularization parameters used to estimate $\hat{f}$ and $\hat{h}$.

The identification of the mappings f and h is a special form of the identification of the NARX model in Section 8.4.1. Therefore, a similar consistency claim as in

(8.11) can be formulated. Again, for consistency here, it is crucial that in the previous steps the true innovation noise sequence and the true state sequence are recovered.

8.5 Illustrative example

In this section, we assess the performance of the proposed identification approach in a simulation study, both in the open- and closed-loop case. The system under study is inspired by the so-called *logistic map*:

$$x_{k+1} = \frac{1}{2}x_k(1 - x_k) + u_k + e_k, \quad y_k = x_k + e_k. \tag{8.14}$$

In the open-loop case, the input u is selected as a zero-mean normal distribution with variance $\sigma_u^2 = 0.01$. The closed-loop case considers the simple feedback law $u_k = r_k - y_k$, where r_k is the reference trajectory, taken from a zero-mean normal distribution with variance $\sigma_r^2 = 0.01$. The innovation noise sequence e_k is drawn from a zero-mean normal distribution with variance σ_e^2, which is chosen to ensure the prescribed signal-to-noise ratio[2] (SNR) of $\{1, 10, 20\}$ dB. For each SNR, a training and validation data-set containing $N = 1000$ samples starting from a zero initial condition is generated.

To assess the influence of using the estimated innovation noise sequence (obtained by identifying a NARX model), three models are identified on each data-set, namely $\mathcal{M}_0, \mathcal{M}_{\hat{e}}$ and $\mathcal{M}_e$. Model $\mathcal{M}_0$ corresponds to the case where noise is not handled in the identification process, implying that only the second step, state estimation by KCCA, and the third step, identification of the state-space maps, are performed. Model $\mathcal{M}_{\hat{e}}$ corresponds to the case where the three-step identification approach is performed, thus the estimated innovation noise sequence $\hat{e}$ is used to estimate the state sequence and to identify the state-space mappings. Model $\mathcal{M}_e$ corresponds to the case where the NARX model returns the true innovation noise sequence, which would be the case for $N \to \infty$ under the consistency claim (8.11). Thus the true innovation noise sequence e is used in the estimation of the state sequence and the identification of the mappings f and h of model $\mathcal{M}_e$. Performance of each identified model $\mathcal{M}_i$, for $i = \{0, \hat{e}, e\}$, is assessed using the so-called best fit rate (BFR), which is defined as follows:

$$\text{BFR} := 100\% \cdot \max\left(1 - \frac{\|y_k - \hat{y}_k\|_2}{\|y_k - \text{mean}(y_k)\|_2}, 0\right). \tag{8.15}$$

Regarding implementation, the NARX model and the mappings f and h are identified using the Gaussian Process toolbox, see [208]. The KCCA problem

[2] $\text{SNR [dB]} := 10 \cdot \log_{10}\left(\frac{(y^{\text{det}} - \mu^{\text{det}})(y^{\text{det}} - \mu^{\text{det}})^\top}{ee^\top}\right)$ where y_k^{det} is the output of system (1) for $e_k = 0 \,\forall\, k$ and $\mu^{\text{det}} := \text{mean}(y^{\text{det}})$.

(8.12) is solved by the KMBOX-toolbox, see [295]. The regularization parameters of the KCCA problem are specified as $\nu_p = \nu_f = 3000$ for all cases. A polynomial kernel $k(z_i, z_j) = (z_i^\top z_j + c)^\ell$ with order $\ell = 2$ and constant $c = 1$ is used in all identification steps. The window length $p = 2$ for the estimation of the NARX model is selected and the window length $d = 2$ for the estimation of the state sequence is selected.

Table 1 presents the results of the open-loop case, whereas Table 2 presents the results of the closed-loop cases. It can be observed that model $\mathcal{M}_0$ performs the worst, in both the open- and closed-loop cases. This is expected, as during the identification of this model, no information on the innovation noise sequence is used. The identification approach presented in this chapter, yielding the model $M_{\hat{e}}$, shows that estimating the innovation noise sequence using a NARX model indeed improves the model fit quality significantly. However, the quality of the estimated innovation noise sequence determines the quality of the estimated state sequence and, subsequently, the quality of the identified mappings f and h. Therefore, when assuming that the NARX model returns the true innovation noise sequence, as is done in the identification of model M_e, it can be seen that an almost perfect fit is ensured for any SNR, validating the second and third steps of the identification approach. Obtaining a good estimate of the innovation noise sequence is a matter of collecting a sufficiently large dataset. It can also be observed that the BFRs in the closed-loop case are generally better than the BFRs in the open-loop case. This is a result of the feedback in the closed-loop case forcing the variance of u_k to become larger than the variance of u_k in the open-loop case.

For the open-loop case with an SNR of 1, a window of the true innovation noise sequence e and the estimated $\hat{e}$ are depicted in the top plot of Figure 8.2. The bottom plot shows the true response y and the simulated response $\hat{y}$ of model $\mathcal{M}_{\hat{e}}$. For the sake of comparison, also an LTI state-space model is identified on the data-set $\mathcal{D}$, which produces the output $\hat{y}_{\text{LTI}}$, also depicted in the same plot. Clearly, it can be concluded that the nonlinear nature of (8.14) cannot be captured by an LTI model, which is also reflected in the BFR being only 22.2% for the LTI model.

8.6 Conclusions

This chapter has presented a three-step approach to the identification of kernel-based non-parametric nonlinear state-space models for discrete-time nonlinear systems operating both in open and closed loop. In the first step, a NARX model is identified using an LS-SVM approach, which yields an estimate for the noise sequence. In the next step, the noise sequence is used to estimate the state sequence using KCCA. The final step estimates the state-space mappings using, again, an LS-SVM approach. Although we do not give an overall consistency proof, the identification approach relies on consistent estimations in each step. Proving overall

Table 8.1. Open-loop identification results.

Data-Set	Training			Validation		
SNR [dB]	1	10	20	1	10	20
BFR [%] $\mathcal{M}_0$	34.8	64.6	83.5	32.5	62.3	84.8
BFR [%] $\mathcal{M}_{\hat{e}}$	84.4	84.8	98.2	83.9	85.6	97.8
BFR [%] $\mathcal{M}_e$	99.3	99.7	99.5	99.2	99.7	99.5

Table 8.2. Closed-loop identification results.

Data-Set	Training			Validation		
SNR [dB]	1	10	20	1	10	20
BFR [%] $\mathcal{M}_0$	28.1	69.6	88.1	27.7	69.4	88.6
BFR [%] $\mathcal{M}_{\hat{e}}$	94.1	96.6	98.4	93.4	96.9	98.4
BFR [%] $\mathcal{M}_e$	99.8	99.8	99.8	99.8	99.8	99.8

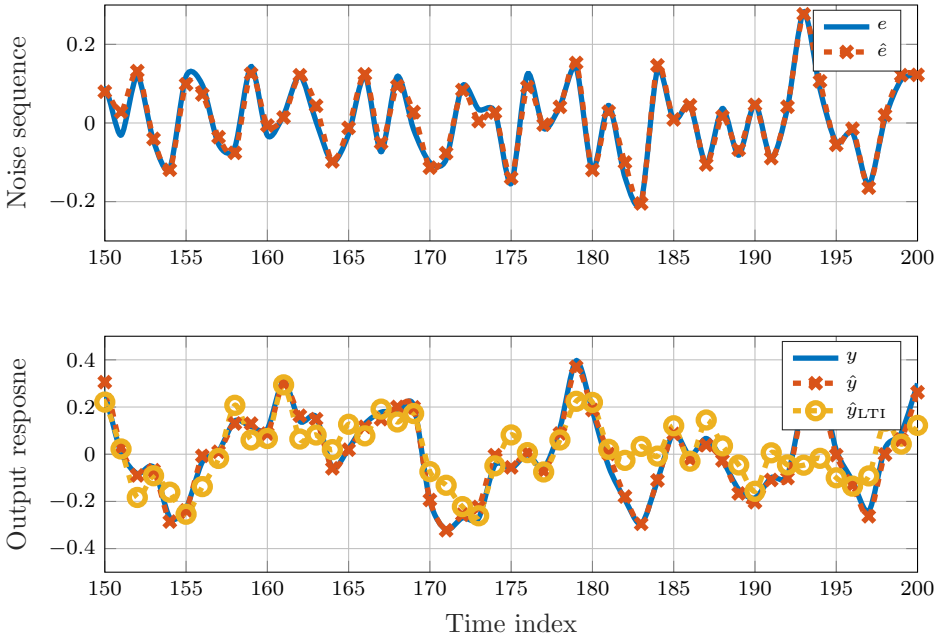


Figure 8.2. Top: true innovation noise sequence e and estimated innovation noise sequence $\hat{e}$. Bottom: true response y, response $\hat{y}$ of nonlinear model $\mathcal{M}_{\hat{e}}$ and response $\hat{y}_{\text{LTI}}$ of the identified LTI model.

consistency is considered a part of future work. In simulation studies, the identification approach obtains accurate predictions on both training and validation data, both in the open-loop and closed-loop cases. The proposed approach can be viewed as a kernel-based non-parametric counterpart for nonlinear systems of the SSARX approach for LTI systems.

9

Kernel-based identification of nonlinear state-space models with stability guarantees

This chapter presents a kernel-based approach for the identification of non-parametric nonlinear state-space models with a focus on enforcing a strong form of model stability on a user-defined compact state and input set. Such a form of model stability provides robustness of the identified models to new inputs unseen during the training phase. The stability concept enforced in this work is called convergence, which guarantees that responses converge to a bounded time-varying steady-state solution and remain inside a user-defined positively invariant set. Using a two-level optimization approach, the optimal model structure in terms of hyper-parameters is selected via a nonlinear optimization problem at the top level, whereas a specific model is selected at the bottom-level problem by solving a convex optimization problem. The benefits of the approach are illustrated using a simulation example and using the Silverbox benchmark dataset.

The contents of this chapter are published in: Shakib, M. F., Tóth, R., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Kernel-based identification of nonlinear state-space models with stability guarantees. *In preparation.*

9.1 Introduction

Given the complexity of today's engineering systems, it can be a challenging task to derive parametric dynamic models from first-principle laws. Therefore, data-based modeling using black-box *non-parametric nonlinear state-space models* has become more and more popular in the recent years [27, 29, 215, 313]. However, non-parametric models often lack interpretability as these are typically described by a large number of model parameters that do not have any physical meaning. In particular, it is difficult for humans to predict the model response to new inputs, given the fact that even a slight input perturbation can result in a wildly different, possibly unbounded, response of the identified nonlinear model. Consequently, there is a trend towards the identification of non-parametric models with *stability guarantees* [70, 125, 215, 221, 274], ensuring, e.g., bounded trajectories. Such stability properties add to the robustness of the identified nonlinear model to inputs that were unseen during training.

A particular class of methods for non-parametric nonlinear state-space modeling is the class of the so-called kernel-based methods [54], which include approaches like regularization networks [58], support vector machines [269]), and Gaussian regression [82, 141]. These methods are extensively used as these provide a high level of modeling flexibility and are, under certain conditions, *universal approximators* [58, 202, 266], i.e., can approximate continuous functions arbitrarily well. However, given the complex function description of kernel-based nonlinear state-space models, analysis of model properties, such as stability, is challenging. To address this challenge, the work in [26] considers *autonomous* kernel-based non-parametric state-space models and analyzes fixed points and their local stability for a selection of kernels. Despite these analysis results, these works also show that analysis of such models is non-trivial as, for example, computing the fixed-points of such a model with the popular squared-exponential kernel can, in general, not be done analytically [26]. There are also identification methods that enforce stability properties on the identified model. For example, the works in [79, 137, 138, 307], present methods to identify *autonomous* kernel-based non-parametric state-space models whose origin is *globally asymptotically stable*. The method in [255] enforces the identified *autonomous* model to be locally contracting, whereas [36]) enforces global contraction. The method in [285] focuses on the squared-exponential kernel and identifies *autonomous* models that exhibit bounded solutions.

The above-mentioned kernel-based state-space modeling methods only deal with *autonomous* systems, i.e., systems without external (time-varying) inputs. As many systems are subject to inputs, the *non-autonomous* modeling counterpart calls for new kernel-based methods that enforce a strong form of model stability on identified *non-autonomous* models. There are several works in the scope of enforcing a strong form of stability on identified *non-autonomous* models. For example, [215] imposes *contraction* on a large class of neural network models, [1, 280, 283] impose *incremental stability* for black-box polynomial state-space

models, and [247, 251] imposes *convergence* for models that can be decomposed into a feedback interconnection between linear dynamics and static nonlinearities. However, none of these methods is directly applicable to kernel-based state-space models.

In this chapter, an identification method for *non-autonomous* kernel-based discrete-time state-space models that enforces the *convergence* stability property is presented. This stability notion guarantees the boundedness, uniqueness, and global asymptotic stability of the steady-state solution [66, 197, 200]. However, in our proposed approach, a *non-global* version of convergence is used, namely convergence on *compact invariant sets*. Such a version is naturally suitable for system identification as system data is only observed in compact sets. The identification problem is cast as a two-level optimization problem. The top-level problem deals with the minimization of the *simulation error* criterion and selects the 'model structure' by tuning, e.g., the kernel hyper-parameters. The bottom-level problem deals with the minimization of the regularized *equation error* criterion. Crucially, at the bottom level, convergence can be enforced in a *convex* manner using linear matrix inequality (LMI) constraints.

Using a simulation example, it is shown that identification without enforcing model stability can result in an unstable model response on a new dataset. Such unfavorable scenarios are excluded using the proposed approach in this chapter. Moreover, the Silverbox benchmark dataset is used to evaluate the predictive performance of the identified model and to highlight again the stability benefits of the proposed approach.

The remainder of this chapter is organized as follows. The identification problem is formally introduced in Section 9.2. A two-level solution to the identification problem is proposed in Section 9.3. The simulation and benchmark study are presented in Section 9.4. Section 9.5 closes with the conclusions of this chapter.

Notation: The symbols $\mathbb{R}, \mathbb{R}_{\geq 0}, \mathbb{C}$, and $\mathbb{Z}$ denote the set of real numbers, non-negative real numbers, complex numbers, and integer numbers, respectively. The symbol I_n denotes the $n \times n$ identity matrix and the symbol 0_n denotes the zero vector of dimensions $n \times 1$. A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called positive (negative) definite, denoted as $A \succ 0$ ($A \prec 0$), if all its eigenvalues are strictly positive (negative). For a vector $x \in \mathbb{R}^n$, its Euclidean norm is denoted by $|x|$. Given a matrix $P \succ 0$ and a vector $x \in \mathbb{R}^n$, $|x|_P$ denotes $\sqrt{x^\top P x}$.

9.2 Problem statement

9.2.1 Data-generating system

Consider MIMO discrete-time nonlinear data-generating systems represented by the following set of first-order difference equations:

$$\tilde{x}_{k+1} = f(\tilde{x}_k, u_k), \tag{9.1a}$$
$$\tilde{y}_k = h(\tilde{x}_k, u_k) + e_k, \tag{9.1b}$$

where, at time instance $k \in \mathbb{Z}$, the state is denoted by $\tilde{x}_k \in \mathbb{R}^n$, the input is denoted by $u_k \in \mathbb{R}^m$ and the output is denoted by $\tilde{y}_k \in \mathbb{R}^l$. The noise $e_k \in \mathbb{R}^l$, is assumed to be independent and identically distributed white with a zero-mean normal distribution with finite diagonal covariance matrix Σ_e, making (9.1) to have an output error (OE) type of noise structure. The mapping f is called the state-transition map and the mapping h is called the output map. Without loss of generality, we assume that the origin is a fixed point for zero input, i.e., $0_n = f(0_n, 0_m)$. System (9.1) is called the data-generating system. The class of bounded inputs $\mathcal{U}_c$ is defined as follows for any constant $c \in \mathbb{R}_{\geq 0}$:

$$\mathcal{U}_c \coloneqq \{\{u_k\}_{k\in\mathbb{Z}} \mid u_k \in U_c, \forall\, k \in \mathbb{Z}\} \tag{9.2}$$

with U_c being a ball around the origin defined as follows:

$$U_c \coloneqq \{u \in \mathbb{R}^m \mid |u| \leq c\}\,. \tag{9.3}$$

Solutions of system (9.1a) for inputs $u \in \mathcal{U}_c$ are all sequences $\{\tilde{x}_k, u_k\}_{k=k_0}^{\infty} \in \mathbb{R}^{n+m}$ that satisfy (9.1a) with $\tilde{x}_0 \in \mathbb{R}^n$. If no confusion arises, such a solution is denoted by $\tilde{x}$.

It is assumed that the data-generating system (9.1a) exhibits a strong form of model stability for any input from an a priori known input class $\mathcal{U}_c$. Hereto, the notion of *global* convergence, defined in [199], is adapted to convergence *on compact sets* and is defined as follows.

Definition 9.1. *The discrete-time nonlinear system* (9.1a) *is said to be exponentially convergent in a set $\mathcal{X} \subset \mathbb{R}^n$ for a class of inputs $\mathcal{U}_c$ if for every $u \in \mathcal{U}_c$:*

- *there exists a solution $\bar{x}$, called the steady-state solution, that is defined and lies in $\mathcal{X}$ for all $k \in \mathbb{Z}$;*
- *the steady-state solution $\bar{x}$ is exponentially stable for any initial condition in $\mathcal{X}$, i.e., there exist scalars $\tau \in \mathbb{R}_{\geq 0}$ and $0 \leq \rho < 1$ such that for any $\tilde{x}_0 \in \mathcal{X}$, the solution $\tilde{x}$ satisfies:*

$$|\tilde{x}_k - \bar{x}_k|^2 \leq \tau\rho^k |\tilde{x}_0 - \bar{x}_0|^2 \,\forall\, k \geq 0. \tag{9.4}$$

■

Note that for the zero input, i.e., $u_k = 0$, for all $k \in \mathbb{Z}$, the origin of a convergent system (9.1a) is an exponentially stable fixed-point for initial conditions in $\mathcal{X}$. Furthermore, for any *time-varying* input sequence from $\mathcal{U}_c$, the exponentially stable steady-state solution $\bar{x}$ is in general also *time-varying*. Consequently, the effect of initial conditions fades out, which can be exploited in an identification setting that uses *steady-state* response data to avoid the dependency on initial conditions, as is done in [251].

Assumption 9.1. *The data-generating system* (9.1a) *is exponentially convergent according to Definition 9.1 on the user-defined convergence region $\mathcal{X}$ for the user-defined class of inputs $\mathcal{U}_c$.* ■

Remark 9.1. With some adaptions, the identification approach proposed in this chapter can be extended to include systems with process noise. In that setting, the identification approach in [250] enables the estimation of the noise realization e and a sample-based estimate of the covariance matrix Σ_e from input-output data directly. Consequently, the noise sequence can be treated as an additional input during the subsequent estimation of f and h. △

9.2.2 Model class

For the estimation of f and h, a reproducing kernel Hilbert space (RKHS) based modeling approach is taken, where basis functions are defined through a so-called *kernel* function. The considered model class is defined as follows:

$$\begin{aligned} Ex_{k+1} = \hat{f}(x_k, u_k) &:= \sum_{i=1}^{N_s} \alpha_i K_i^{\alpha}(x_k, u_k), \\ y_k = \hat{h}(x_k, u_k) &:= \sum_{i=1}^{N_s} \beta_i K_i^{\beta}(x_k, u_k), \end{aligned} \tag{9.5}$$

where the functions $K_i^{\alpha}(x_k, u_k) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $K_i^{\beta}(x_k, u_k) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are *kernel slices*, generated by symmetric, positive definite kernel functions. The weights $\alpha_i \in \mathbb{R}^n, \beta_i \in \mathbb{R}^n$, for all $i \in \{1, \dots N_s\}$, with N_s the number of kernel slices used, and the matrix $E \in \mathbb{R}^{n \times n}$ are model parameters. An example of a popular kernel function is the squared-exponential kernel function defined as follows [233]:

$$K^{\alpha}(z^a, z^b) := \exp\left(-\frac{|z^a - z^b|^2}{2\ell}\right), \tag{9.6}$$

where the kernel width $\ell > 0$ is tunable hyper-parameter and $z^a, z^b \in \mathbb{R}^p$ and p a positive integer. For the problem at hand, the kernel slices in (9.5) are then

defined as follows:

$$K_i^\alpha(x,u) := K^\alpha\left(\begin{bmatrix} x_i^s \\ u_i^s \end{bmatrix}, \begin{bmatrix} x \\ u \end{bmatrix}\right), \tag{9.7}$$

where $x_i^s \in \mathbb{R}^n, u_i^s \in \mathbb{R}^m, i = 1, \ldots, N_s$, are so-called *pseudo inputs* [259] or *inducing variables* [279]. The pseudo inputs are also part of the tunable hyper-parameters and facilitate efficient sparse implementation [259, 279].

The kernel function choice translates to the resulting function space. The identification approach in this chapter can be applied to almost any kernel function, e.g., linear, polynomial, rational, spline, and wavelet kernel function [233], or the squared-exponential kernel function as in (9.6). The only requirement is that the kernel K^α is differentiable with respect to x in the convergence region $\mathcal{X}$. Furthermore, the kernel functions generating the kernel slices in $\hat{f}$ and $\hat{h}$ in (9.5) can be selected independently. The selected kernel functions, together with the choice for kernel hyper-parameters and pseudo-input locations, define a, possibly infinitely large, set of basis functions [14, 139].

9.2.3 Identification problem

The dataset denoted by $\mathcal{D}$ contains N samples of the input $\tilde{u}$, output $\tilde{y}$, and the state $\tilde{x}$ of the data-generating system (9.1), and is defined as follows:

$$\mathcal{D} = \{u_k, \tilde{y}_k, \tilde{x}_k\}_{k=1}^N. \tag{9.8}$$

The availability of the state sequence $\tilde{x}$ is exploited in the solution to the identification problem proposed in Section 9.3. In practice, however, full-state measurements are not always available in addition to input and output measurements. A compatible state sequence can then be estimated using a kernelized version of the canonical correlation method, as outlined in [218, 250, 298]. It is assumed that the data is generated from the zero initial condition.

For the input u from the dataset $\mathcal{D}$, the response of model (9.5) starting from the zero initial condition is denoted by $\{x_k, y_k\}_{k=1}^N$. Based on the model response, the *simulation error* criterion is defined as follows:

$$J_{se} = \frac{1}{N}\sum_{k=1}^{N} \|\tilde{y}_k - y_k\|_2^2. \tag{9.9}$$

Furthermore, the convergence region $\mathcal{X}$ is defined as the convex hyperellipsoidal set:

$$\mathcal{X} := \{x \in \mathbb{R}^n \mid x^\top X x \leq 1\}, \tag{9.10}$$

characterized by the user-defined matrix $0 \prec X \in \mathbb{R}^{n\times n}$. Note that the matrix X can always be chosen such that $\mathcal{X}$ contains the state sequence $\tilde{x}$ in the dataset $\mathcal{D}$, i.e., $\{\tilde{x}_k\}_{k=1}^N \in \mathcal{X}$. However, the user is free to choose the set $\mathcal{X}$ arbitrarily

large to accommodate model extrapolation without instability issues. Based on the simulation error criterion (9.9), the convergence region $\mathcal{X}$, and the user-defined input class $\mathcal{U}_c$, the identification problem is formalized as follows.

Problem 9.1. *Consider the data-set $\mathcal{D}$ in* (9.8)*, the convergence region $\mathcal{X}$ in* (9.10)*, and the class of inputs $\mathcal{U}_c$ in* (9.2)*. Find a model of the form* (9.5) *such that the simulation error criterion J_{se} in* (9.9) *is minimized and such that the identified model* (9.5) *is exponentially convergent on the convergence region $\mathcal{X}$ for the class of inputs $\mathcal{U}_c$.* △

In practice, verifying whether the true system is convergent on $\mathcal{X}$ for the class of input $\mathcal{U}_c$ is a non-trivial task. Therefore, Problem 9.1 can be interpreted in two different ways. Firstly, if the data-generating system is indeed convergent, e.g., inferred by analyzing its responses, then Problem 9.1 aims at preserving this property for the identified model. Secondly, if it is unknown whether the data-generating system is convergent, then Problem 9.1 aims at enforcing the convergence property because it is a favorable model property for the usage of the model, e.g., for reliable model simulation for new inputs.

Problem 9.1 considers the simulation error criterion, which is preferred over, e.g., the equation error criterion, as models identified using the simulation error criterion typically exhibit accurate long-term predictions [240]. However, simulation error minimization comes at the expense of an optimization problem that is nonlinear in its parameters due to the recursive evaluation of the to-be-identified state-transition dynamics. Therefore, a tailored approach to solve Problem 9.1 is presented in the next section.

9.3 Identification via two-level optimization

In this section, the identification problem is formulated as a two-level hierarchical optimization problem. At the top level, the kernel hyper-parameters, regularization parameters, and the pseudo-input locations are optimized using the *simulation error* criterion defined in (9.9). At the bottom level, the weights α_i, β_i for $i \in \{1, \ldots, N_s\}$, and the matrix E are optimized using the *equation error* criterion. The latter problem is formulated as a jointly convex optimization problem in the error criterion as well as in its constraints encoding the convergence property.

9.3.1 Convex identification of convergent models

To facilitate convex identification of convergent models of the form (9.5) at the bottom level, the regularized equation error criterion is introduced together with a convex characterization of the convergence property. Consider the regularized

equation error criterion defined as follows:

$$\begin{aligned} J_{ee} :=& \frac{\gamma_f}{2(N-1)} \sum_{k=1}^{N-1} \left\| E\tilde{x}_{k+1} - \hat{f}(\tilde{x}_k, u_k) \right\|_2^2 + \frac{\gamma_h}{2N} \sum_{k=1}^{N} \left\| \tilde{y}_k - \hat{h}(\tilde{x}_k, u_k) \right\|_2^2 \\ &+ \frac{1}{2} \sum_{i=1}^{n} \left\| \hat{f}_{(i)} \right\|_{\mathcal{H}}^2 + \frac{1}{2} \sum_{i=1}^{p} \left\| \hat{h}_{(i)} \right\|_{\mathcal{H}}^2, \end{aligned} \tag{9.11}$$

where $\|\cdot\|_{\mathcal{H}}^2 = \langle \cdot, \cdot \rangle$ is the squared Hilbert-space norm, defined for functions in a RKHS, and $\hat{f}_{(i)} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the i-th element of the vector-valued function $\hat{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ (and $\hat{h}_{(i)}$ is defined similarly). The RKHS is defined through the kernel function (and its hyper-parameters), see [14] for details. Essentially, the error criterion (9.11) is the one-step-ahead prediction error criterion with additional regularization terms (the last two terms in (9.11)) to avoid overfitting and control the bias-variance trade-off via the regularization-parameters $\gamma_f, \gamma_h \in \mathbb{R}_{\geq 0}$. The model parameters E and $\alpha_i, \beta_i \in \mathbb{R}^n, i = 1, \ldots, N_s$, are collected in the parameter vector $\theta \in \mathbb{R}^{n_\theta}$ with $n_\theta := n^2 + 2nN_s$, where N_s is the number of pseudo inputs. The equation error criterion in (9.11) is quadratic and hence convex in the model parameter vector θ. Using the convex set $\mathcal{X}$ defined in (9.10), an LMI feasibility test for the convergence property on $\mathcal{X}$ for models of the form (9.5) is presented in the next theorem using the notation:

$$A(x, u) \coloneqq \sum_{i=1}^{N_s} \alpha_i \frac{\partial K_i^\alpha}{\partial x}(x, u). \tag{9.12}$$

Theorem 9.1. *Consider model* (9.5)*, the convergence region* $\mathcal{X}$ *in* (9.10)*, and the class of inputs* $\mathcal{U}_c$ *in* (9.2) *defined through* U_c *in* (9.3)*. Assume that* $A(x, u)$ *exists for all* $(x, u) \in (\mathcal{X}, U_c)$*. If there exists a matrix* $P \in \mathbb{R}^{n \times n}$ *such that:*

$$P \succ 0, \tag{9.13a}$$

$$\begin{bmatrix} E + E^\top - P & A^\top(x, u) \\ A(x, u) & P \end{bmatrix} \succ 0, \tag{9.13b}$$

$$\begin{bmatrix} E + E^\top - X & \hat{f}(x, u) \\ \hat{f}^\top(x, u) & 1 \end{bmatrix} \succeq 0, \tag{9.13c}$$

for all $(x, u) \in (\mathcal{X} \times U_c)$*. Then, model* (9.5) *is exponentially convergent in the set* $\mathcal{X}$ *under any inputs from* $\mathcal{U}_c$*. As a consequence, any two solutions starting from* $x_0^a, x_0^b \in \mathcal{X}$*, with the same input* $u \in \mathcal{U}_c$ *will remain in* $\mathcal{X}$*, i.e.,* $x_k^a, x_k^b \in \mathcal{X}$*,* $\forall\, k \geq 0$*. Furthermore, there exist scalars* $\tau \in \mathbb{R}_+$ *and* $0 \leq \rho < 1$ *(defined in the proof in Appendix G.1), such that any two solutions starting from* $x_0^a, x_0^b \in \mathcal{X}$ *with the same input* $u \in \mathcal{U}_c$ *converge exponentially to each other, i.e.,*

$$|x_k^a - x_k^b|^2 \leq \tau \rho^k |x_0^a - x_0^b|^2, \, \forall\, k \geq 0. \tag{9.14}$$

▲

Proof. The proof can be found in Appendix G.1. □

Condition (9.13b) enforces (exponential) *incremental stability* on $(\mathcal{X}, U_c)$, whereas condition (9.13c) enforces *positive invariance* of the set $\mathcal{X}$ for inputs from U_c. The latter property of the set $\mathcal{X}$ for the model (9.5) implies that there exists a solution $\bar{x}$ that lies in $\mathcal{X}$ for all $k \in Z$, see [198, Lemma 2]. As a consequence, (exponential) *convergence* according to Definition 9.1 is guaranteed. The conditions of Theorem 9.1 are convex in the matrix P and the parameter vector θ as these appear linearly in the conditions (9.13).

For most kernel choices, the function $\hat{f}$ in (9.5) is non-convex in x and u. Consequently, the conditions of Theorem 9.1 have to be verified for all $(x, u) \in (\mathcal{X} \times U_c)$. For some specific parametrization of the state-transition map $\hat{f}$ in (9.5), the conditions of Theorem 9.1 can be verified efficiently, for example using a polynomial basis function expansion for $\hat{f}$ and the sum of squares programming techniques, see [1, 280, 283]. Unfortunately, that approach does not apply to kernel-based modeling using a generic class of kernels, as is the considered case in this chapter. To make the conditions computationally tractable, the verification is performed on a grid, where the sets $\mathcal{X}^{\mathrm{g}}, U_c^{\mathrm{g}}$ denote the gridded version of the convergence region $\mathcal{X}$ and the input space U_c, respectively. This is further motivated by the observation that the matrices in the conditions of Theorem 9.1 depend continuously on (x, u). The grid density trades off the risk of violating the constraints (9.13) for some $(x, u) \in (\mathcal{X}, U_c)$ against the numerical complexity.

Besides the parameter vector θ, an additional parameter vector is introduced which collects all remaining parameters. This vector is denoted by $\phi \in \mathbb{R}^{n_\phi}$ with $n_\phi := n_{\mathrm{hyp}} + 2 + (n + m)N_s$, and collects (i) the n_{hyp} number of kernel hyper-parameters; (ii) the regularization parameters γ_f, γ_h in (9.11); and (iii) the N_s number of pseudo-input locations $(x_i^s, u_i^s), i = 1, \ldots, N_s$.

Given the conditions of Theorem 9.1 and the gridded sets $\mathcal{X}^{\mathrm{g}}, U_c^{\mathrm{g}}$, the model set $\Theta(\phi)$ is defined as follows:

$$\Theta(\phi) := \{\theta \in \mathbb{R}^{n_\theta} \mid (9.13) \text{ is feasible} \, \forall \, (x, u) \in (\mathcal{X}^{\mathrm{g}}, U_c^{\mathrm{g}})\}\,. \tag{9.15}$$

The set $\Theta(\phi)$ encodes the convergence property such that any candidate model $\theta \in \Theta(\phi)$ satisfies the conditions of Theorem 9.1 on the grid $(\mathcal{X}^g, U_c^g)$. It is assumed that this grid-based test, for a sufficiently dense grid, gives exponential convergence on the non-grid-based convergence region $\mathcal{X}$ with the non-grid-based input space U_c. The dependency of the set Θ on ϕ is via the constraints (9.13b) and (9.13c), both depending inherently on ϕ. Note that for any $\theta \in \Theta(\phi)$, the quadratic form $x^\top (E + E^\top)x$ is positive definite, guaranteeing the existence of E^{-1}.

With some abuse of notation, the equation error criterion (9.11) is written as $J_{ee}(\theta, \phi)$ with arguments θ and ϕ. Using this notation, for any given choice of

parameters ϕ, the following constrained optimization problem is formulated:

$$\theta_{\text{opt,ee}}(\phi) = \underset{\theta \in \Theta(\phi)}{\arg\min}\, J_{ee}(\theta, \phi), \tag{9.16}$$

which is jointly convex in the parameter vector θ in the criterion as well as in its constraints. The constrained optimization problem (9.16) is called the *bottom-level* optimization problem.

Remark 9.2. For any $\phi \in \mathbb{R}^{n_\phi}$, and any selected kernel function, the set $\Theta(\phi)$ in (9.15) is non-empty because model (9.5) with $E = X, \alpha_i = 0, \beta_i = 0, i = 1, \ldots, N_s$, satisfies the conditions of Theorem 9.1. In fact, such a model is *globally* convergent. △

Remark 9.3. A selection for vector ϕ can be interpreted as a selection of the *model class*. For example, a change in the kernel hyper-parameters or the pseudo-input locations results in different functions generated by the kernels. The vector θ can be interpreted as the parametrization of the *specific model*. A similar two-level optimization problem was also proposed in [136] for system identification. △

9.3.2 Two-level hierarchical optimization problem

The bottom-level convex problem (9.16) is exploited to formulate a two-level optimization problem:

$$\begin{aligned}(\theta_{\text{opt}}, \phi_{\text{opt}}) = \underset{\theta_{\text{opt,ee}} \in \mathbb{R}^{n_\theta}, \phi \in \mathbb{R}^{n_\phi}}{\arg\min}\; & J_{se}(\theta_{\text{opt,ee}}(\phi), \phi) \\ \text{subject to } \theta_{\text{opt,ee}}(\phi) = & \underset{\theta \in \Theta(\phi)}{\arg\min}\, J_{ee}(\theta, \phi).\end{aligned} \tag{9.17}$$

The top-level problem minimizes the simulation error criterion J_{se} in (9.9) for the kernel hyper-parameters, regularization parameters, and pseudo-input locations, all collected in ϕ. As the top-level optimization problem is *unconstrained*, by the grace of Remark 9.2, it can be solved in various ways including global, non-gradient-based optimization such as particle swarm optimization.

The bottom-level problem considers a fixed ϕ, coming from the top level. Given the fact that the equation error criterion J_{ee} in (9.11), as well as the constraints, are jointly convex in θ (and the matrix P in Theorem 9.1), the bottom-level optimization problem can efficiently be solved using, e.g., interior-point methods or the ellipsoid method [42]. As a result, the bottom-level optimization problem can be interpreted as a mapping from $\phi, \mathcal{D}$ to the optimal $\theta_{\text{opt,ee}}$ with a unique $\theta_{\text{opt,ee}}$ for each $\phi, \mathcal{D}$, allowing for efficient implementation.

To initialize the problem, the following *unconstrained* optimization problem is solved:

$$(\theta_{\text{init}}, \phi_{\text{init}}) = \underset{\theta \in \mathbb{R}^{n_\theta}, \phi \in \mathbb{R}^{n_\phi}}{\arg\min}\; J_{ee}(\theta, \phi) \tag{9.18}$$

with the regularized equation error criterion J_{ee} defined in (9.11). Taking a Bayesian view, this problem boils down to Gaussian Process regression [209] and can be efficiently solved using, e.g., the GPML toolbox [208], which allows learning all parameters in ϕ and θ, including the pseudo-input locations using, e.g., the Fully Independent Training Conditional (FITC) approach [259] or the Variational Free Energy (VFE) approach [279]. Initialization results in (i) a small number of optimal pseudo-input locations; and (ii) optimal initial values for hyper and regularization parameters. However, the resulting model does not necessarily satisfy the conditions of Theorem 9.1. Furthermore, this model may not be optimal in terms of the *simulation error* criterion J_{se} defined in (9.9).

Starting from an initial parameter vector ϕ_{init}, the two-level optimization problem (9.17) can be solved effectively using nonlinear optimization. The identified model is guaranteed to be convergent on the user-defined convergence region $\mathcal{X}$ for the class of user-defined inputs $\mathcal{U}_c$. Moreover, since the simulation error forms the top-level criterion, the identified model has optimal long-term prediction capabilities.

9.4 Numerical case studies

This section first presents a *simulation case study* that highlights the benefits of the identification strategy proposed in this chapter. After that, the proposed approach is evaluated using the *Silverbox system* benchmark dataset ([304]). The main contribution of this chapter is the identification of the state-transition mapping with the convergence property enforced. Therefore, the focus of this section is on the identification of the state-transition mapping. The predictive quality of the identified mapping is evaluated using the best fit rate (BFR) of the simulated model response. The BFR is defined as follows:

$$\text{BFR} := 100\% \cdot \max\left(0, 1 - \frac{\|\tilde{x} - x\|_2}{\|\tilde{x} - \text{mean}(\tilde{x})\|_2}\right), \tag{9.19}$$

where $\tilde{x}$ is the (measured) response of the data-generating system and x is the model response computed by the forward simulation using the input from the dataset. The BFR can be computed for both the training and test datasets. For national convenience, the variable ψ is used in the remainder of this section to denote the collection of parameters θ and ϕ.

9.4.1 Implementation details

In both studies, the top-level optimization problem is solved using the controlled-random search algorithm ([206]). The bottom-level optimization problem is solved using `MOSEK` ([13]) in combination with `YALMIP` ([157]). Initialization is performed using the `GPML` toolbox ([208]), where the optimal pseudo-input locations are found

using the FITC method, see ([259]). These pseudo-input locations were not optimized further after initialization.

9.4.2 Demonstration of benefits on a simulation case study

Consider the following data-generating nonlinear system:

$$x_{k+1} = \begin{cases} u_k - 2x_\star^3 + 3x_\star^2 x_k & \text{if } x > x_\star, \\ u_k + x_k^3 & \text{if } |x| \leq x_\star, \\ u_k + 2x_\star^3 + 3x_\star^2 x_k & \text{if } x < -x_\star, \end{cases} \tag{9.20}$$

$$y_k = x_k, \tag{9.21}$$

where $x_\star = 0.5$. The dataset $\mathcal{D}$ of length $N = 100$ is generated using an input sequence drawn from a normal distribution and scaled such that $|u_k| \leq 1$, for all $k \in \{1, \ldots, N\}$, thus $u \in \mathcal{U}_c$ with $c = 1$. Using this dataset, two models are identified using the third-order polynomial kernel and a number of $N_s = 10$ pseudo inputs. Firstly, in a traditional fashion, a model of the form (9.5) is identified in an unconstrained manner according to [209], hence not guaranteed to be convergent. This model is referred to as the *non-convergent* model with estimated parameter vector ψ_{nc} (collecting both θ and ϕ). Secondly, a model of the form (9.5) is identified using the proposed approach in this chapter for the set $\mathcal{X}$ characterized by $X = 1$, i.e., $\mathcal{X} := \{x \in \mathbb{R} \mid |x| \leq 1\}$, and for the class of inputs $\mathcal{U}_c$ as in (9.2) with $c = 1$. This model is referred to as the (optimal) *convergent* model with parameter ψ_c. For the implementation of the LMI constraints characterizing the set $\Theta(\phi)$ in (9.15), the input and state-space are gridded equidistantly to give 5 points in $\mathcal{X}$ and U_c. The identification results are depicted in Figure 9.1. Although both models perform well, the non-convergent model performs slightly better on the estimation data, which is also evident by a best-fit-rate of 95.4% for the non-convergent model versus 91.0% for the convergent model. This accuracy result is expected as the proposed approach in this chapter sacrifices accuracy for guaranteed model stability via the convergence property.

Next, both models are subjected to a new test input that is again drawn from a normal distribution and scaled such that its maximum absolute value is 1. This test input is again from the class of input $\mathcal{U}_c$ with the same constant $c = 1$ as used in the identification step. It can be observed in Figure 9.2 that the response of the identified non-convergent model for this new input grows unbounded, even though the response of the data-generating system remains bounded. The response of the identified convergent model, however, remains within $\mathcal{X}$, i.e., within $|x| \leq 1$, as enforced. The BFR for the convergent model is 78.7% on the test data.

For scalar models, the conditions of Theorem 9.1 can only be guaranteed to hold if the Jacobian of the state-transition mapping $E^{-1}\hat{f}$ with respect to x remains in absolute value below 1 for $(x, u) \in (\mathcal{X}, U_c)$. This necessary condition is violated for the identified non-convergent model for zero input, as can be seen from Figure 9.3. The identified convergent model does satisfy this necessary condition.

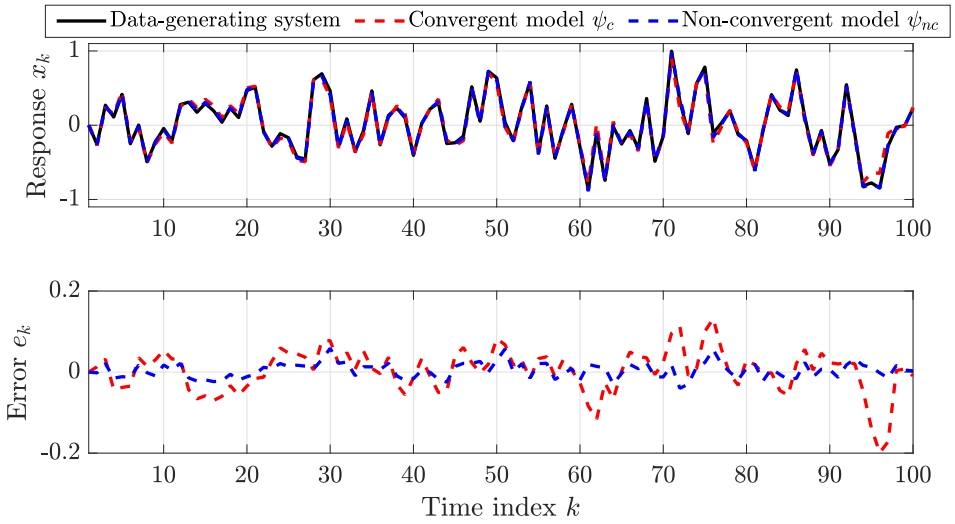


Figure 9.1. Response of the data-generating system together with the simulation error of the identified convergent and non-convergent models.

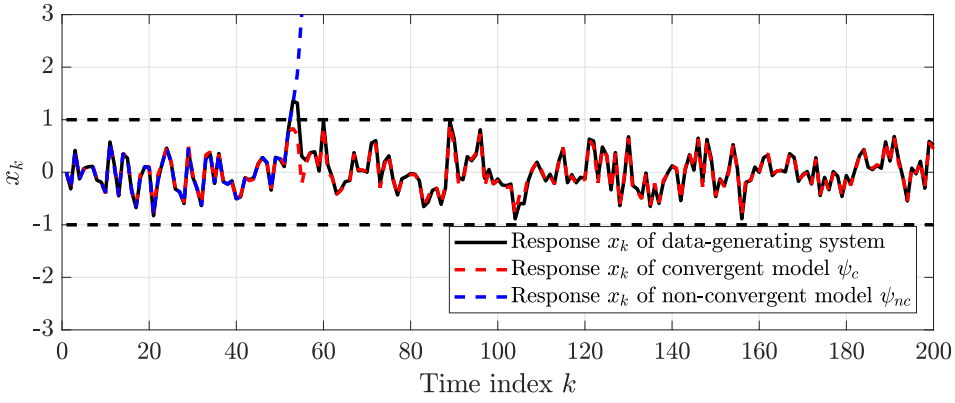


Figure 9.2. Bounded response of the data-generating system for the test input. The response of the identified convergent model remains within $|x| \leq 1$, whereas the response of the identified non-convergent model becomes unbounded.

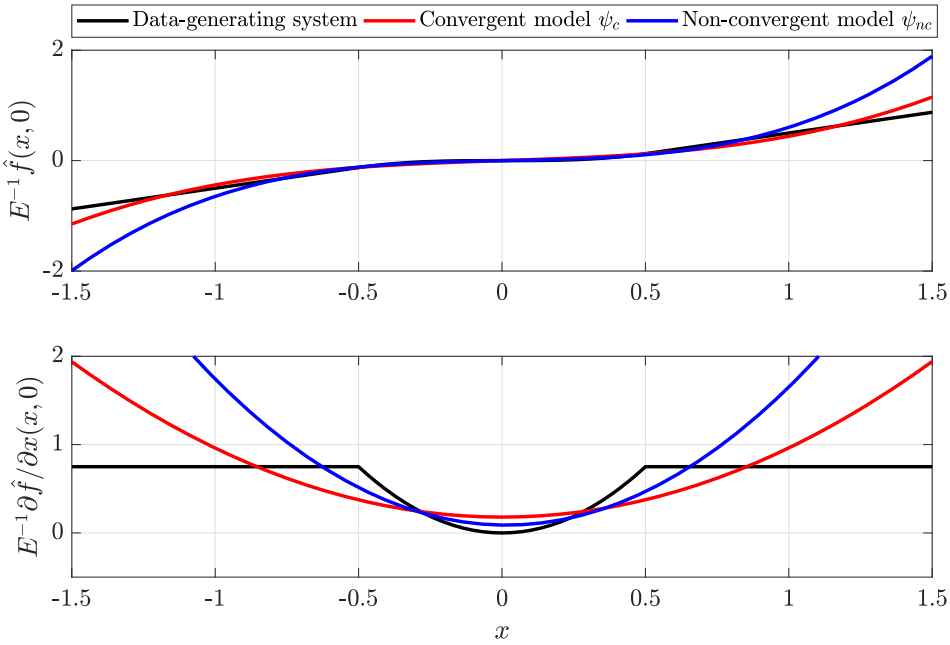


Figure 9.3. The state-transition mapping (top) of the data-generating system, identified convergent model, and identified non-convergent model for zero input, together with their Jacobian (bottom) for zero input.

This example illustrates the potential instability hazards of identified models with traditional approaches, even in a noiseless scenario and without input extrapolation. Furthermore, it highlights the benefits of the proposed approach in this chapter, namely guaranteed bounded model responses inside the user-defined convergence region $\mathcal{X}$ for any input from the user-defined input class $\mathcal{U}_c$. The convergence property thus provides robustness of the identified model to new inputs from $\mathcal{U}_c$.

9.4.3 Silverbox benchmark

The Silverbox system is an electronic implementation of the Duffing oscillator, which can be modeled as a second-order mass-spring-damper system with a nonlinear cubic spring nonlinearity. The benchmark dataset was proposed by ([304]) and has attracted many researchers to test their identification methods, see ([27]) for an overview.

The benchmark dataset consists of input $\tilde{u}$ and output $\tilde{y}$ sequences that can be split into an estimation part and a test part. Notably, the excitation amplitude of the test data exceeds the excitation amplitude of the estimation data, thereby subjecting the identified model to an extrapolation test. Using this dataset, a surrogate state sequence is defined as follows:

$$\tilde{x}_k = \begin{bmatrix} \tilde{y}_k \\ \tilde{y}_k - \tilde{y}_{k-1} \end{bmatrix} \in \mathbb{R}^2, \quad k = 2, \ldots, N, \tag{9.22}$$

where the second state element has the interpretation of the 'velocity'. The model output y is defined as the first state element $x(1)$. Using this dataset, two models are identified using a third-order polynomial kernel and a number of *only* $N_s = 10$ pseudo inputs. Similar to the simulation study in the previous section, firstly, a model of the form (9.5) is identified in an unconstrained manner, referred to as the *non-convergent* model with parameter ψ_{nc}. Secondly, a model of the form (9.5) is identified using the approach in this chapter for the set $\mathcal{X}$ characterized by

$$X := \begin{bmatrix} 9 & -5 \\ -5 & 18 \end{bmatrix}, \tag{9.23}$$

and for the class of inputs $\mathcal{U}_c$ with $c = 0.15$ as in (9.2). Both the state sequence in the estimation and test data lie in the interior of the set $\mathcal{X}$. This identified model is referred to as the *convergent* model and is characterized by the parameters ψ_c. For implementation purposes of the LMI constraints, the input and state space are both gridded. The grid location locations $\mathcal{X}_g$ are depicted in Figure 9.6. The input space is equidistantly gridded at 10 points in U_c. The identification results are depicted in Figure 9.4. The reported BFRs are 83.6% for the convergent model and 91.7% for the non-convergent model. The so-called root-mean-square (rms) error [304], used for benchmarking, is 8.9 mV for the convergent model and 4.5 mV for the non-convergent model. Again, the convergent model is expected to

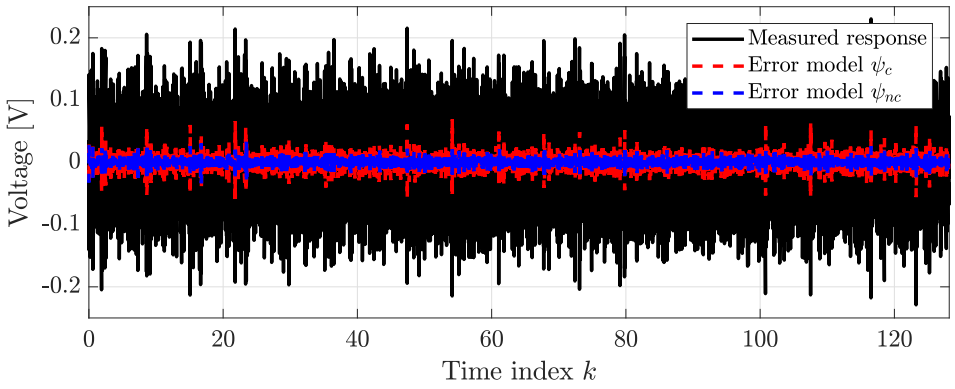


Figure 9.4. Response of the identified models for the estimation data input.

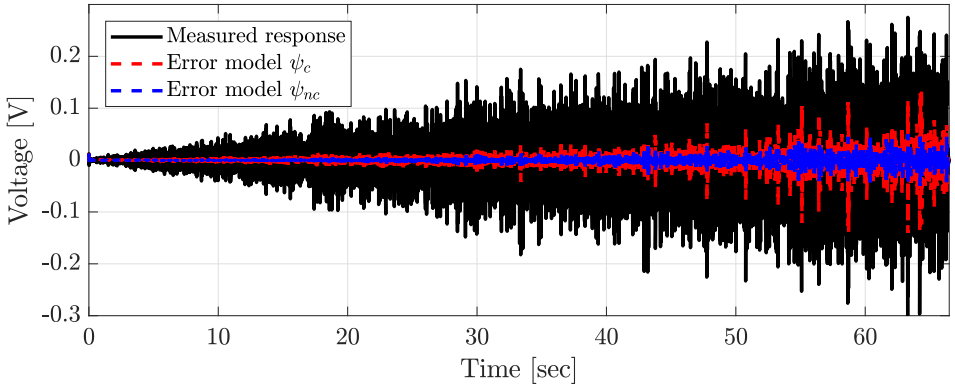


Figure 9.5. Response of the identified models for the test input.

give less accuracy as this model sacrifices accuracy for guaranteed model stability via the convergence property.

The responses of both models are depicted in Figures 9.5 for the test input. Especially for large input amplitudes, both models exhibit large errors with the rms values of 12.3 mV for the convergent model and 6.0 mV for the non-convergent model. The responses of both the convergent and non-convergent model for this test input lie in the interior of the set $\mathcal{X}$, as is observed from the $x(1), x(2)$ plane in Figure 9.6.

Although less accurate in this example, the identified convergent model exhibits additional stability properties, which are highlighted next by a simulation study. As an input, we take a random binary sequence between -0.14 and 0.14, see the top plot of Figure 9.7. These limits are smaller than the amplitudes of the input in the test dataset. The model responses are depicted in Figure 9.7, where it can be

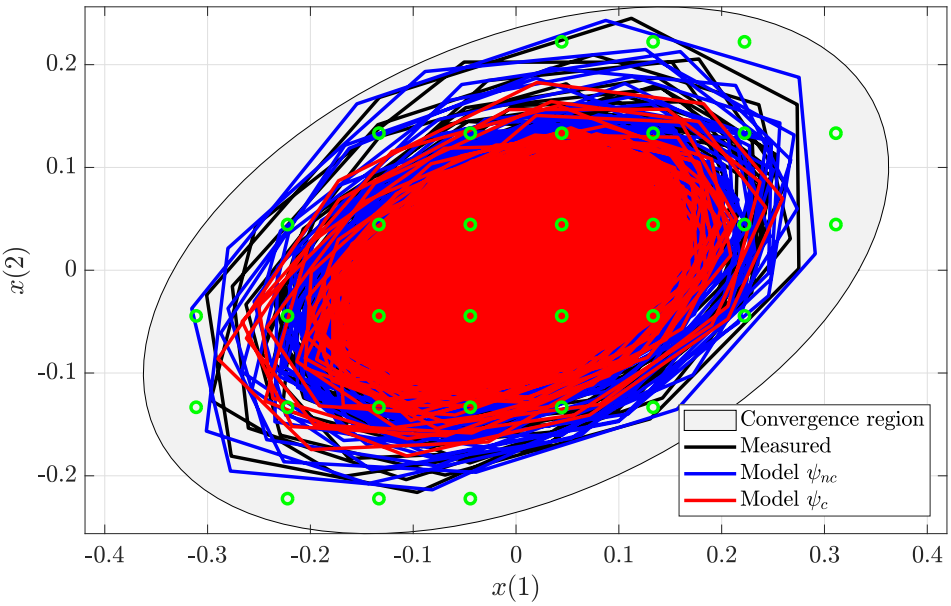


Figure 9.6. The response of the identified models for the test input in the $x(1), x(2)$-plane. The green circles represent the grid points in $\mathcal{X}_g$.

seen that the non-convergent model produces unbounded responses. Thus, even if the non-convergent model produces bounded responses for the test data (see Figure 9.5), there are no guarantees that the model produces bounded responses for new inputs, even if the new inputs do not exceed the input amplitude of the test data. The convergent model, however, produces bounded responses, which also remains inside $\mathcal{X}$ as verified in the $x(1), x(2)$-plane in Figure 9.8.

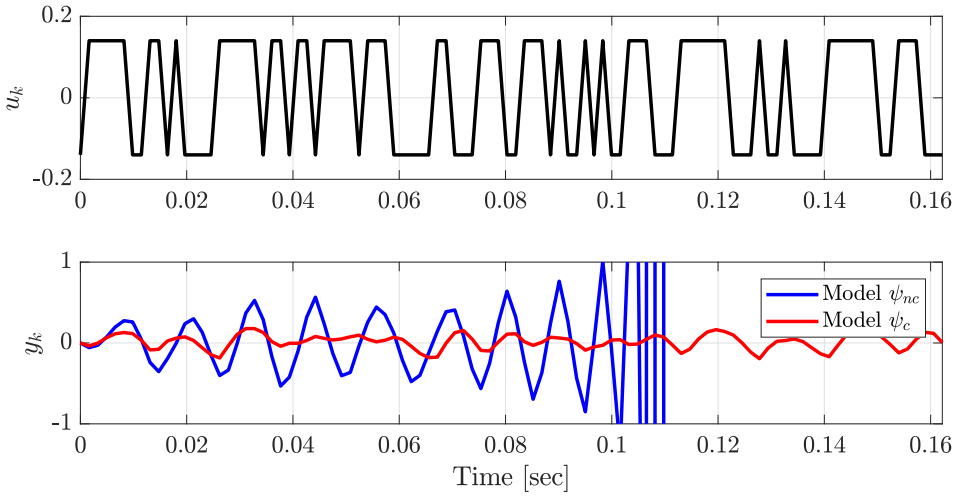


Figure 9.7. The input (top) and response (bottom) of the identified models for a new binary test input. The response of the non-convergent model becomes unbounded, whereas the response of the convergent model remains bounded.

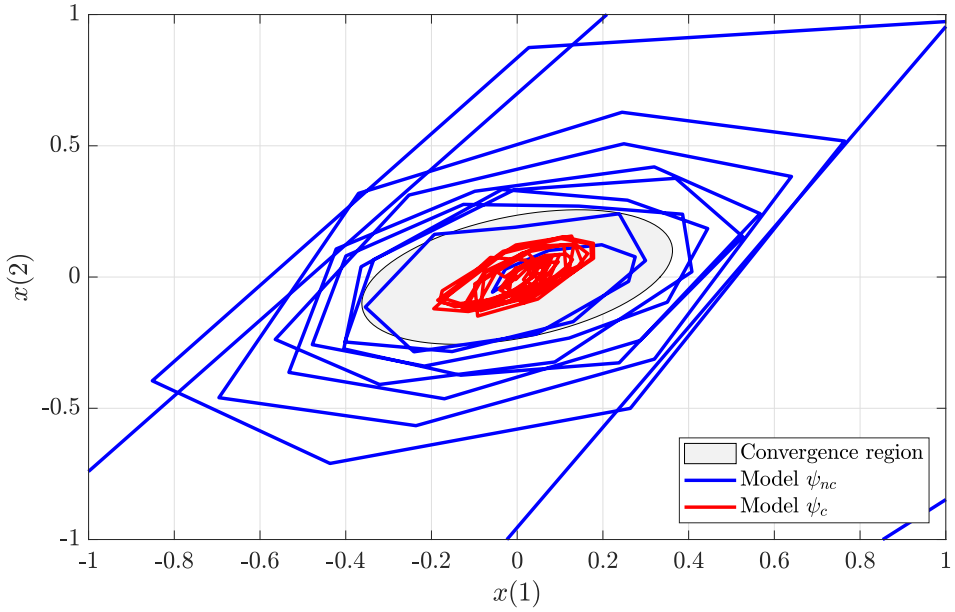


Figure 9.8. The response of the identified models for the binary test input in the $x(1), x(2)$-plane. The response of the identified convergent model remains inside the set $\mathcal{X}$.

9.5 Conclusions

This chapter has presented an approach for the identification of kernel-based nonlinear non-parametric state-space models with stability guarantees. The identified models are guaranteed to remain inside a user-defined positively invariant set for a class of user-defined inputs. Consequently, these models safely generalize toward unseen scenarios, which provides robustness of the identified models to new inputs. The benefits of the approach are illustrated by means of a simulation example and the Silverbox benchmark dataset. In both cases, the identified models generalize in a favorable manner for new inputs, whereas other methods produce unbounded responses for unseen scenarios.

IV

Conclusions and recommendations

10 Conclusions and recommendations

10.1 Conclusions

The data-driven modeling and complexity reduction problems are considered in this thesis. As follows from the literature review in Sections 1.3 and 1.4, the field of nonlinear data-driven modeling faces the challenges of (i) constructing expressive classes of models to capture a large class of nonlinear systems; (ii) enforcing model stability to enhance generalization toward unseen, new inputs, especially for safety-critical applications; and (iii) enabling computationally efficient identification for time-restrictive applications, also for continuous-time modeling. Furthermore, the field of model reduction for *generic* nonlinear models faces the challenges of (i) structure- and stability-preserving reduction for model generalization toward new inputs; (ii) finding error bounds for the quantification of errors between the full and the reduced models; and (iii) optimal reduction to minimize errors between the full and the reduced models. These challenges have formed the overarching research objective of this thesis.

> *Develop tools for the computationally efficient identification and optimal model reduction of nonlinear models with stability guarantees.*

This objective is further split into the Research objectives I - VIII formulated in Section 1.5. The contributions of this thesis toward these research objectives are detailed next.

Chapters 2 and 3 propose identification approaches based on steady-state data for the class of Lur'e-type models. Lur'e-type models can be decomposed into a linear time-invariant (LTI) block that is placed in feedback with a static nonlinearity and fall into the class of *block-oriented models*. The identification approach guar-

antees the identified model to be exponentially convergent. This property ensures the boundedness and global exponential stability of uniquely defined steady-state solutions. Hence, for any (unseen) bounded input, qualitative predictions of model outputs can be given, namely: model responses remain bounded, model responses corresponding to different initial conditions converge to the exponentially stable steady-state response, and model responses corresponding to small changes in the input remain close to each other. This provides robustness to the identified model for unseen inputs. Furthermore, these identification techniques utilize numerically efficient tools to compute model responses and gradient information, enabling fast identification cycles. The approach in Chapter 3 can be viewed as the discrete-time counterpart of the continuous-time approach in Chapter 2 with an extended model class such that multiple-input, multiple-output (MIMO) systems with multiple nonlinear functions can be considered.

The benefits of these approaches are demonstrated in numerical and experimental case studies. The continuous-time approach in Chapter 2 reduces the total identification time from hours to only minutes in simulation studies. Furthermore, this approach is experimentally validated in mechanical ventilation, a technique used to regulate the breathing of patients in respiratory distress during the nursery of intensive-care and acutely-ill surgical patients. In this study, the identification of the parameters of a first-principle model, including the patient's lung parameters, is addressed. The proposed identification method significantly reduces the identification time, thereby enabling faster patient treatment with the aim to avoid negative consequences for the patient's lungs and saving valuable time for the medical practitioner. The identification approach for the discrete-time case in Chapter 3 is demonstrated on the Wiener-Hammerstein benchmark dataset and the Silverbox benchmark dataset. In conclusion, the methods in Chapters 2 and 3 present computationally efficient approaches for the identification of Lur'e-type models with stability guarantees.

Chapter 4 proposes a framework for analyzing the mismatch between steady-state system responses and steady-state model responses caused by model simulation, output sampling, and disturbances. Such a mismatch is inevitable in practice as real-life dynamic systems evolve in the continuous-time world whereas their models are simulated in the digital world using discrete-time numerical simulation algorithms. The model response is then used in a variety of systems and control problems, e.g., the system identification problem in Chapter 2. Two types of mismatches are analyzed for the class of convergent Lur'e-type models, namely (i) the steady-state mismatch between the *sampled measured* system response and the simulated model response; and (ii) the steady-state mismatch between the *actual* system response and the simulated model response. Here, the steady-state model responses of Lur'e-type models are computed using the mixed time-frequency algorithm. The first bound (i) provides a measure of the mismatch between the sampled measured system response and the simulated model response. This bound also addresses aliasing effects due to sampling, which are particularly important

in system identification problems. The second bound (ii) gives essential insights into the intersample behavior that is not captured by sampling the system's output response. For example, in the system identification problem, such an error quantification can be instrumental to distinguish between structural model errors and errors caused by the sampling process or model simulation algorithm. Using a simulation study, valuable insights into the factors causing the mismatches are gained. Furthermore, this study showed that a reduction of the bounds also results in a reduction of the mismatches. This latter insight gives value to the formulated methods to reduce the derived bounds and, thereby, the mismatches. In conclusion, Chapter 4 enables qualitative and quantitative characterization of the steady-state mismatch between system and model responses for the class of convergent Lur'e-type models.

Chapter 5 considers the model order reduction by *time-domain* moment matching problem for MIMO LTI models. Pre-existing Krylov-projector methods obtain a match between the transfer-function matrices of the full-order and of the reduced-order models along so-called *tangential directions*. This chapter derives a matching property of time-domain moment matching for MIMO LTI models and shows that, in general, this property is not consistent with the matching property obtained by pre-existing Krylov-projector methods. To achieve consistency, the derived matching property is exploited to formulate additional conditions on the time-domain moment matching parameters. In conclusion, Chapter 5 derives a matching property of time-domain moment matching for MIMO LTI models and formulates additional conditions to guarantee consistency with Krylov-projector methods.

Chapter 6 extends the notion of moments of nonlinear models from the local to the global case for a generic class of convergent nonlinear models. Hereto, the center manifold theorem used in the local context is replaced by a global invariant manifold theorem for convergent models. Using this global extension, a constructive reduction approach for convergent Lur'e-type models is proposed. In this approach, only the dimension of the LTI block of the Lur'e-type model is reduced by a tailored approach using the LTI time-domain moment matching method. Such a reduction approach is motivated by the fact that many engineering systems exhibit only local and low-order nonlinear behavior. The proposed approach is structure-preserving, inherits the frequency-response function interpretation of the LTI moment matching method, preserves the convergence property, and is equipped with an a posteriori error bound. The preservation of the convergence property, in particular, enables reliable model usage for generalized inputs without showing instability issues. This benefit is highlighted in a case study, where the reduced-order model accurately matches the moment of the full-order Lur'e-type model and accurately describes the steady-state model response under input variations. In conclusion, Chapter 6 presents a global extension of time-domain moment matching for generic convergent nonlinear models and proposes a tailored reduction approach for Lur'e-type nonlinear models.

Chapter 7 presents optimal model order reduction techniques for MIMO LTI models and MIMO nonlinear feedback models. In the LTI case, this approach matches the moments of the full-order model, preserves model stability, and minimizes the approximation error in the $\mathcal{H}_\infty$-norm. The nonlinear case considers *convergent* models that can be decomposed into a feedback interconnection between LTI dynamics and nonlinear dynamics. Then, the LTI reduction technique is further exploited and adapted to reduce the dimension of the LTI dynamics of nonlinear feedback models. This reduction approach is structure-preserving, convergence-preserving, equipped with an error bound between the steady-state response of the full-order and the reduced-order nonlinear models, and minimizes this error bound. Moreover, this chapter proposes a numerical procedure to solve the model reduction problems. This numerical procedure is of independent interest as it can be applied with minor adaptions to a wide variety of optimization problems with bilinear matrix inequality (BMI) constraints. In numerical case studies, this iterative algorithm prevents getting stuck prematurely and effectively finds accurate reduced models, both in the LTI and nonlinear feedback cases. In particular, in the nonlinear feedback case, the found optimal nonlinear model achieves higher accuracy compared to other methods that are not necessarily optimal. In conclusion, Chapter 7 presents optimal model order reduction techniques for MIMO LTI models and MIMO nonlinear feedback models.

Chapter 8 considers the problem of kernel-based identification for closed-loop systems. For this problem, an identification technique is proposed that is inspired by the three-step approach of the consistent SSARX method for the identification of LTI systems that operate in a closed loop. In the first step, rather than identifying an auto-regressive model with exogenous inputs (ARX), a nonlinear ARX (NARX) model is identified by a least-squares support vector machine (LS-SVM) approach. The residual of the fit of the NARX model serves as an estimate of the noise sequence. Next, in the second step, the estimated noise sequence is treated as an additional deterministic input to estimate a state sequence using a kernelized version of canonical correlation analysis. Finally, having also an estimate of the state sequence at hand, in the third step, the state-transition and the output map of the nonlinear state-space model are identified again using an LS-SVM approach. Each of the three steps of the proposed procedure corresponds to a consistent estimate under the assumption that the true noise and state sequence are provided in the previous steps. Using simulation studies, the performance of the proposed identification strategy is analyzed. A Comparison with the case where noise is not taken into account highlights that the proposed approach significantly improves the model quality. In conclusion, Chapter 8 presents a three-step kernel-based identification procedure for closed-loop systems, where each step corresponds to a consistent estimator.

Chapter 9 considers the identification problem of kernel-based state-space models with stability guarantees. The proposed approach enforces the *convergence* property on compact sets using linear matrix inequality (LMI) constraints. This

stability property guarantees that model solutions remain inside a user-defined positively invariant set for user-defined input classes. Consequently, these models safely generalize toward unseen scenarios, which adds to the robustness of the identified models. The benefits of the identification approach are illustrated using simulation examples and the Silverbox benchmark dataset. In both cases, the identified models generalize favorably to new inputs, whereas other methods produce unbounded responses for unseen scenarios. In conclusion, Chapter 9 presents an approach to the kernel-based identification of nonlinear state-space models with stability guarantees.

Summarizing, this thesis presents approaches for data-driven nonlinear modeling and complexity reduction of nonlinear models. The focus is on guaranteeing a strong form of model stability, namely *convergence*.

10.2 Recommendations

This final section presents recommendations for future research directions for data-driven nonlinear modeling and complexity reduction of nonlinear models. The recommendations are grouped per chapter.

Chapters 2 and 3 present numerical tools for the efficient computation of steady-state model responses, instrumental for the system identification problem. Given the increasing interest in neural network models and their long identification times [315], it is recommended to research the applicability of the numerical tools in these chapters to neural network models. For example, in [214, 215], it is shown that a large class of neural network models can be written in a model structure close to the model structure considered in Chapter 3. In particular, the approach in [215] uses *unconstrained* optimization to identify large-scale neural network models with the contractivity property (a global stability notion, similar to convergence) and input-output properties expressed by integral quadratic constraints. It is recommended to merge the unconstrained optimization method from [215], together with the fast model simulation tools presented in Chapter 3 of this thesis to form an identification strategy that is *flexible* and *computationally efficient* for identification problems utilizing *steady-state* data. For the continuous-time counterpart presented in Chapter 2, it is recommended to research the combination of a continuous-time Lur'e-type model with a neural network model to maintain model *interpretability* (obtained from first-principle modeling), enhance modeling *flexibility* (obtained from the neural network), and enhance *identification speed* (thanks to the simulation method in Chapter 2). Furthermore, for enhanced model flexibility, it is recommended to extend the methods in Chapters 2 and 3 to a more generic model class, e.g., where the nonlinear feedback can also be *dynamic*, similar to the model class considered in Chapter 7.

Chapter 4 introduces a unifying framework in which errors between system re-

sponses and model responses due to sampling, model simulation, and disturbances can be analyzed for the class of convergent Lur'e-type models. It is recommended to use the results of this chapter to analyze and reduce errors between model responses and measured system responses for the identification of Lur'e-type systems. Furthermore, as simulation accuracies are of fundamental importance for many other systems- and control-related problems, e.g., model-predictive control, it is recommended to research how this framework can be extended to *other nonlinear model structures* and other *stability concepts*, e.g., input-to-state stability. Moreover, it is recommended to pursue *less conservative error bounds*, e.g., in the direction of Remark 4.6 aiming to exploit the smoothness properties of the excitation signal in the derivations of the bounds. Finally, it is recommended to also account for *modeling errors*. Such modeling errors can already be taken into account via the disturbance signals. However, treating these separately enables utilizing more information, possibly leading to tighter error bounds.

Chapter 5 proposes conditions under which time-domain moment matching for MIMO LTI models is consistent with the Krylov-projector method. The MIMO case is essential for certain applications, for example, the ones in Chapters 6 and 7 of this thesis where a MIMO LTI block is interconnected with static or dynamic *nonlinear* elements. Besides the conditions for consistency, it is recommended to research *other properties* of time-domain moment matching for MIMO LTI models. For example, in the single-input, single-output (SISO) case, tangential directions do not play a role and *completeness* can be shown, i.e., all models achieving moment matching are inside the parametrized set of reduced-order models, see [16, Proposition 1]. However, in the MIMO time-domain moment matching case, *completeness* is only shown partially, see Theorem 7.2, and it is challenging to obtain a stronger completeness claim, see also Remark 7.3. Furthermore, it is also recommended to research the effect of the selection of the tangential directions on additional properties in the MIMO case. For example, an instrumental property would be to select the tangential directions such that matching in the *individual* transfer functions is obtained.

Chapter 6 extends the notion of a moment of a nonlinear model from the local to the global context. Based on this extension, Remark 6.2 states that families of reduced-order nonlinear models can be found similarly as is done in [16]. These reduced models are constructed based on the availability of the mapping between the state of the signal generator and the steady-state state of the full-order nonlinear model. In general, finding an exact representation of this mapping can be challenging. Therefore, it is recommended to research how this mapping can be found analytically or numerically for certain classes of nonlinear models. To *estimate* this mapping, in [225, 226], a basis function expansion is used. However, to guarantee the convergence property for the reduced model, the so-called Demidovich condition should be satisfied for the selected basis functions. It is recommended to research what choices for basis functions are instrumental for complying with the Demidovich condition to efficiently find convergent reduced-order models. A

final recommendation is to utilize the parametric freedom of time-domain moment matching to fit the transient response of the reduced nonlinear model to the transient response of the full nonlinear model, similar to [223] for LTI models. Such a fit aims at accurately capturing the transient response of the full nonlinear by the reduced nonlinear model (in addition to the steady-state response).

Chapter 6 also introduces a constructive reduction method for Lur'e-type models. This method uses a *grid-based* quantification of the mismatch between the LTI part of the full-order and of the reduced-order models via a frequency-domain error criterion. Firstly, it is recommended to complement this error criterion with a frequency-domain weighting in accordance with Remark 6.4, enabling emphasized accuracy demands in certain frequency ranges. Secondly, as a grid-based quantification does not provide optimality guarantees for frequencies not on the grid, it is recommended to exploit the smoothness properties of the frequency-domain error criterion to conclude for frequencies not on the grid. Thirdly, it is recommended to explore alternatives that do provide optimality guarantees for all frequencies, e.g., by the minimization of the $\mathcal{H}_2$- or $\mathcal{H}_\infty$-norm of the error transfer function matrix of the LTI part, similar to the proposed approach in Chapter 7 of this thesis. Finally, as the grid density is a user choice, it is recommended to investigate the effect of grid density on the reduction performance and computational expense.

Chapter 7 proposes an optimal model reduction method for linear and nonlinear models. This approach utilizes BMI constraints to characterize the mismatch between the LTI part of the full and of the reduced models. The main drawback of this approach is that these BMI constraints are high-dimensional, making the approach computationally demanding. It is recommended to research ways to relax the computational demands, especially for large-scale problems. For example, employing a grid-based quantification of the mismatch between the LTI part of the full and reduced models, similar to the one proposed in Chapter 6, may reduce the computational demands. Even though a grid-based quantification does not provide formal accuracy guarantees, such an adaption may be favored for *large-scale* problems. Finally, it is recommended to research the effect of the choice of interpolation points and tangential directions on the accuracy of the reduced nonlinear model.

Chapter 7 also introduces a numerical procedure to solve optimization problems with bilinear constraints. This *combined* algorithm combines two coordinate descent algorithms (CDA). In the numerical examples in this chapter, it is demonstrated that this combined algorithm is robust with respect to the initial starting point and prevents from getting stuck as would have happened using either of the two CDAs independently. Given this potential, it is recommended to research the properties of this algorithm in terms of its *convergence* and its *convergence speed*. Furthermore, this numerical procedure can be applied to a variety of systems- and control-related problems, such as the model reduction by time-domain moment matching (as in Chapter 7 of this thesis), the model matching problem, and the optimal static output feedback problem. It is recommended to research whether

this proposed *combined* algorithm can find better numerical solutions to these problems compared to the two CDAs or other traditional solvers that are available in the literature.

Chapter 8 proposes an identification approach for kernel-based nonlinear models that also handles closed-loop systems. This approach consists of three steps, where each step corresponds to a consistent estimate. However, *overall* consistency has not been proven for the three steps combined and is recommended to be pursued in future research. Furthermore, it is recommended to validate this approach *experimentally*. Choosing a suitable kernel function in an experimental setting can be more challenging compared to the simulation study in Chapter 8.

Chapter 9 introduces an approach for the identification of kernel-based nonlinear models with stability guarantees. This approach uses a *grid-based* LMI feasibility test to enforce the convergence property. It is recommended to research the effect of gridding on the computational complexity and the reliability of enforcing the convergence property. Furthermore, it is recommended to exploit the structure of certain kernels, e.g., the convergence constraints for models using the polynomial kernel may be enforced via the sum of squares programming method, without the need for gridding. This can reduce the computational complexity and improve the reliability of the results. Moreover, it is recommended to extend the approach to the *continuous-time* setting as well. Finally, it is recommended to merge the approaches in Chapters 8 and 9 to form an *overall identification approach* to be validated experimentally.

V

Appendices

A Appendices to Chapter 2

A.1 Proof of Theorem 2.2

Proof. The proof consists of three parts. In the first step, the expected value of the cost function (2.5) with respect to the noise e is analyzed. Under Assumption 2.2, it can be shown that

$$\mathbb{E}\left[J_N(\theta)\right] = V_N(\theta) + \sigma_e^2 \tag{A.1}$$

with $V_N(\theta)$ defined in Assumption 2.3. Since the noise variance σ_e^2 is θ-independent, by Assumption 2.3 (persistency of excitation), a unique global minimum of $\mathbb{E}\left[J_N(\theta)\right]$ exists. By Assumption 2.1 (system in the model class), this unique global minimum is θ_0, as $V_N(\theta_0) = 0$.

In the second step, uniform convergence of $J_N(\theta)$ to $\mathbb{E}\left[J_N(\theta)\right]$, in the mean squared sense, is analyzed. Under Assumption 2.2, the following expression is derived:

$$\mathbb{E}\left[\left(J_N(\theta) - \mathbb{E}\left[J_N(\theta)\right]\right)^2\right] = \frac{2}{N}\sigma_e^4 + \frac{4}{N}\sigma_e^2\gamma, \tag{A.2}$$

where $\gamma := \frac{t_s}{T}\sum_{k=1}^{T/t_s}\left\{\left(\bar{\varkappa}_0(t_k) - \bar{z}(t_k,\theta)\right)^2\right\} < \infty$ is a bounded constant representing the squared error between the bounded measured steady-state response $\bar{\varkappa}(t_k)$ and the bounded steady-state model response $\bar{z}(t_k,\theta)$ over one steady-state period. Since σ_e^2, σ_e^4 are finite by Assumption 2.2 and γ is finite for any $\theta \in \Theta$ by Assumption 2.1 (thus also in a neighborhood of θ_0), the following limit holds:

$$\lim_{N\to\infty} \mathbb{E}\left[\left(J_N(\theta) - \mathbb{E}\left[J_N(\theta)\right]\right)^2\right] = 0. \tag{A.3}$$

In the final step, having (A.1) and (A.3) at hand, along the same lines as the proof of Theorem 2 in [262], it is concluded that $\hat{\theta}_N$ converges in probability to θ_0 as $N \to \infty$, i.e.,

$$\lim_{N\to\infty} \mathbb{E}\left[\hat{\theta}_N - \theta_0\right] = 0.$$

This completes the proof. □

A.2 Proof of Theorem 2.4

Proof. The proof is given for the scalar case of θ and can be repeated analogously for each component of a vector-valued θ. First, consider the following property of Lur'e-type models (2.2) satisfying the conditions of Theorem 2.1.

Property A.1 ([196])**.** *Consider model* (2.2)*. Under the conditions of Theorem 2.1, if* $\theta_1(h)$ *converges to* θ_2 *as* $h \to 0$*, and* T*-periodic* $w_1(t,h)$ *converges to* T*-periodic* $w_2(t)$ *uniformly in* $t \in [t_0, t_0+T)$*, then the corresponding steady-state solution* $\bar{x}_{w_1(t,h)}(t,\theta_1(h))$ *converges to* $\bar{x}_{w_2(t)}(t,\theta_2)$ *uniformly in* $t \in [t_0, t_0+T)$ *as* $h \to 0$*.* ▲

Let us show that for $p(t,\theta,h) := \frac{1}{h}(\bar{x}(t,\theta+h) - \bar{x}(t,\theta))$ there exists the limit $x_\theta(t,\theta) := \partial x/\partial\theta(t,\theta) = \lim_{h\to 0} p(t,\theta,h)$. For notational convenience, we drop the argument t from here on and we define $\theta^+ := \theta + h$.

As follows from the definition of the steady-state solution $\bar{x}(\theta)$ of the model (2.2), for $h \neq 0, p(\theta,h)$ is a T-periodic function satisfying the dynamics

$$\begin{aligned}\frac{dp(\theta,h)}{dt} =& \frac{1}{h}\left(A(\theta^+)\bar{x}(\theta^+) - A(\theta)\bar{x}(\theta)\right) + \\ & \frac{1}{h}\left(-B(\theta^+)\varphi(\bar{y}(\theta^+),w,\theta^+) + B(\theta)\varphi(\bar{y}(\theta),w,\theta)\right) + \frac{1}{h}\left(L(\theta^+) - L(\theta)\right)w,\end{aligned}$$

which can be written as

$$\begin{aligned}\frac{dp(\theta,h)}{dt} =& \frac{1}{h}\left(A(\theta^+)\bar{x}(\theta^+) - A(\theta)\bar{x}(\theta)\right) + \frac{1}{h}\left(A(\theta^+)\bar{x}(\theta) - A(\theta^+)\bar{x}(\theta)\right) + \\ & \frac{1}{h}\left(-B(\theta^+)\varphi(\bar{y}(\theta^+),w,\theta^+) + B(\theta)\varphi(\bar{y}(\theta),w,\theta)\right) + \\ & \frac{1}{h}\left(-B(\theta^+)\varphi(\bar{y}(\theta),w,\theta) + B(\theta^+)\varphi(\bar{y}(\theta),w,\theta)\right) + \\ & \frac{1}{h}\left(-B(\theta^+)\varphi(\bar{y}(\theta^+),w,\theta) + B(\theta^+)\varphi(\bar{y}(\theta^+),w,\theta)\right) + \frac{1}{h}\left(L(\theta^+) - L(\theta)\right)w \\ =& \frac{1}{h}\left(A(\theta^+)\bar{x}(\theta^+) - A(\theta^+)\bar{x}(\theta)\right) + \frac{1}{h}\left(A(\theta^+)\bar{x}(\theta) - A(\theta)\bar{x}(\theta)\right) + \\ & \frac{1}{h}\left(-B(\theta^+)\varphi(\bar{y}(\theta^+),w,\theta^+) + B(\theta^+)\varphi(\bar{y}(\theta^+),w,\theta)\right) + \\ & \frac{1}{h}\left(-B(\theta^+)\varphi(\bar{y}(\theta^+),w,\theta) + B(\theta^+)\varphi(\bar{y}(\theta),w,\theta)\right) + \\ & \frac{1}{h}\left(-B(\theta^+)\varphi(\bar{y}(\theta),w,\theta) + B(\theta)\varphi(\bar{y}(\theta),w,\theta)\right) + \frac{1}{h}\left(L(\theta^+) - L(\theta)\right)w.\end{aligned}$$

The existence of the partial derivatives with respect to θ and the uniform convergence of $p(\theta,h)$ to x_θ is proven in the same way as in [196]. Taking the limit

$\lim_{h\to 0} p(\theta, h)$ and applying Property A.1, allows to write

$$\begin{aligned}\dot{x}_\theta(\theta) =& A(\theta)x_\theta(\theta) + A_\theta(\theta)\bar{x}(\theta) - B_\theta(\theta)\varphi(\bar{y}(\theta), w, \theta)\\ &- B(\theta)\left(\varphi_\theta(\bar{y}(\theta), w, \theta) + \varphi_y(\bar{y}(\theta), w, \theta)y_\theta(\theta)\right) + L_\theta(\theta)w,\end{aligned}$$

where $w, \bar{x}(\theta)$ and $\bar{y}(\theta)$ are considered as inputs. From here, it is straightforward to show

$$\begin{aligned}y_\theta(\theta) =& C(\theta)x_\theta(\theta) + C_\theta(\theta)\bar{x}(\theta) + D_\theta(\theta)w,\\ z_\theta(\theta) =& F(\theta)x_\theta(\theta) + F_\theta(\theta)\bar{x}(\theta) - G_\theta(\theta)\varphi(\bar{y}(\theta), w, \theta)\\ &- G(\theta)\left(\varphi_\theta(\bar{y}(\theta), w, \theta) + \varphi_y(\bar{y}(\theta), w, \theta)y_\theta(\theta)\right),\end{aligned}$$

and present the sensitivity system as in (2.13).

Let us next prove that model (2.13) satisfies conditions **C1 - C3** of Theorem 2.1. Condition **C1** is satisfied since the matrices A, B and C of the LTI block of (2.13) are the same as those of model (2.2). Condition **C2** holds with the same K as for the model (2.2) since $|\partial\varphi/\partial y(\bar{y}(\theta), w, \theta)| \leq K$ for all $\bar{y}(\theta) \in \mathbb{R}, w \in \mathbb{R}$, and $\theta \in \Theta$. Condition **C3** holds automatically since A, B, C, and K remain unchanged. Application of Theorem 2.1 to model (2.13) concludes that for the T-periodic input $[\bar{x}(\theta), w, \bar{y}(\theta)]$, model (2.13) has a unique T-periodic steady-state solution $\bar{x}_\theta(\theta)$. □

B
Appendices to Chapter 3

B.1 Proof of Theorem 3.1

Proof. The proof is split into two parts. In the first part, it is shown that the conditions of Theorem 3.1 imply satisfaction of the conditions of Theorem B.1 in Appendix B.5.1 for global exponential convergence (GEC). These conditions are recalled from [126] for the case where $D = 0_{n_y \times n_w}$. The second part shows that model (3.2) is GEC, also for $D \in \mathbb{R}^{n_y \times n_w}$.

Part 1: Consider model (3.2) with $D = 0_{n_y \times n_w}$ and with the nonlinearity satisfying the incremental sector bounds $[-I_{n_y}, I_{n_y}]$. Consider the loop-transformation $u(k) = \tilde{u}(k) - Cx(k)$, resulting in the nonlinearity satisfying the incremental sector bounds $[0_{n_y}, 2I_{n_y}]$ and the transformed matrix $\tilde{A} \coloneqq A - BC$, see [135]. This transformation can be incorporated in inequality (3.5) by pre- and post-multiplication with the matrix

$$\begin{bmatrix} I_{n_x} & -C^\top \\ 0_{n_y \times n_x} & I_{n_y} \end{bmatrix},$$

and its transpose, respectively, resulting in:

$$\begin{bmatrix} \tilde{A}^\top \\ B^\top \end{bmatrix} P \begin{bmatrix} \tilde{A}^\top \\ B^\top \end{bmatrix}^\top - \begin{bmatrix} P & -C^\top \\ \star & I_{n_u} \end{bmatrix} \prec 0. \tag{B.1}$$

Under the conditions of Theorem 3.1, there exists a $\tilde{P} \succ 0$ such that the matrix inequality (3.5) holds for $P = \tilde{P}$. In turn, this implies that the inequality (B.1) holds for $P = \tilde{P}$. Observe that inequality (B.1) is equivalent to inequality (B.19) in Appendix B.5 if $(A, P_{11}, S_1, \Omega, \tau)$ are replaced by $(\tilde{A}, P, \frac{1}{2}I_{n_y}, 2I_{n_y}, 1)$. Note that $\tau = 1$ is not allowed as $0 < \tau < 1$ in Theorem B.1. By application of [126, Lemma 3], it can be concluded that an additional variable $\tilde{\tau}$ exists, with $0 < \tilde{\tau} < 1$, such that (B.19) holds for $\tau = \tilde{\tau}$. Satisfaction of inequality (B.19) implies satisfaction of inequality (B.20) if $(A, \bar{\Omega}, P_{21}, S_2)$ are replaced by $(\tilde{A}, 2I_{n_y}, P, \frac{1}{2}I_{n_y})$. Since all conditions of Theorem B.1 hold and the transformation $u(k) = \tilde{u}(k) -$

$Cx(k)$ does not affect the GEC property [200], model (3.2) with $D = 0_{n_y \times n_w}$ and whose nonlinearity satisfies the incremental sector bounds $[-I_{n_y}, I_{n_y}]$, is GEC for any bounded input.

Part 2: Given that model (3.2) with $D = 0_{n_y \times n_w}$ satisfies the conditions of Theorem 3.1, it only remains to show that the model (3.2) is also GEC in the generic case with $D \in \mathbb{R}^{n_y \times n_w}$. The statements of Theorem B.1 rely on *global exponential incremental stability* and the *existence of a compact positively invariant set* in which the steady-state dynamics reside. It is sufficient to show that, irrespective of the presence or absence of matrix D, the dynamics obey these two properties.

Consider the incremental Lyapunov function $V(k) \coloneqq \|x_1(k) - x_2(k)\|_{\tilde{P}}^2$. Satisfaction of (3.5) guarantees that such V exists and satisfies

$$V(k+1) - \tilde{\tau} V(k) \leq 0, \quad \forall x_1, x_2 \in \mathbb{R}^{n_x}, k \in \mathbb{Z}, \tag{B.2}$$

for some $0 < \tilde{\tau} < 1$. Since the nonlinearity φ is *globally* incrementally sector bounded, the incremental stability property is invariant in the presence or absence of matrix D. Therefore, V can also be used as an incremental Lyapunov function for the dynamics (3.2) with $D \in \mathbb{R}^{n_y \times n_w}$.

Next, we prove that a compact positively invariant set exists in which the steady-state solution $\bar{x}$ resides. Define the mapping $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \times \mathbb{Z} \to \mathbb{R}^{n_x}$ as follows:

$$f(x, w, k) \coloneqq Ax + B\varphi(Cx + Dw, k) + Lw. \tag{B.3}$$

Consider the following metric:

$$\|f(x, w, k)\|_{\tilde{P}}^2 \leq \|f(x, w, k) - f(0, w, k)\|_{\tilde{P}}^2 + \|f(0, w, k)\|_{\tilde{P}}^2 \tag{B.4}$$

By global incremental stability, inequality (B.2) can be applied as follows:

$$\|f(x, w, k) - f(0, w, k)\|_{\tilde{P}}^2 \leq \tilde{\tau} \|x\|_{\tilde{P}}^2. \tag{B.5}$$

Furthermore, thanks to the incremental sector bounds $[-I_{n_y}, I_{n_y}]$ on the nonlinearity φ, for any bounded input w, there exists a constant $0 \leq C_w < +\infty$ that bounds the term $\|f(0, w, k)\|_{\tilde{P}}^2$ as follows:

$$\begin{aligned} \|f(0, w, k)\|_{\tilde{P}}^2 &= \|B\varphi(Dw, k) + Lw\|_{\tilde{P}}^2 \\ &\leq \|B\varphi(Dw, k)\|_{\tilde{P}}^2 + \|Lw\|_{\tilde{P}}^2 \\ &\leq \|BDw\|_{\tilde{P}}^2 + \|Lw\|_{\tilde{P}}^2 \leq C_w < +\infty. \end{aligned}$$

Then, $\|f(x, w, k)\|_{\tilde{P}}^2 \leq \tilde{\tau} \|x\|_{\tilde{P}}^2 + C_w$ upper bounds (B.4), resulting in the compact positively invariant set

$$\left\{ x \in \mathbb{R}^{n_x} \mid \|x\|_{\tilde{P}}^2 \leq \frac{C_w}{1 - \tilde{\tau}} \right\}. \tag{B.6}$$

Together with the global exponential incremental stability property, this implies GEC for the dynamics (3.2) and for bounded inputs. In conclusion, the conditions of Theorem 3.1 guarantee that model (3.2) with $D = 0_{n_y \times n_w}$ is GEC *(Part I)* and model (3.2) with $D = \mathbb{R}_{n_y \times n_w}$ is GEC *(Part 2)*, which closes the proof. □

B.2 Proof of Theorem 3.2

Proof. Consider two arbitrary nonlinearity input signals $y^a_{[i]},\ y^b_{[i]} \in \ell_2^{n_y}(N)$ defined at an arbitrary iteration $i \geq 0$ of the iterative process (3.15). Application of (3.15) on $y^p_{[i]}$ with $p \in \{a,\ b\}$ results in:

$$y^p_{[i+1]} = \left(\hat{\mathcal{F}}_{yu} \circ \hat{\mathcal{F}}_{uy}\right) y^p_{[i]} + \hat{\mathcal{F}}_{yw} \circ w, \tag{B.7}$$

in which $\hat{\mathcal{F}}_{yu}$ and $\hat{\mathcal{F}}_{\bar{y}w}$ are linear steady-state operators (defined similarly to (3.12)). Furthermore, $\hat{\mathcal{F}}_{uy}$ is the nonlinear steady-state operator defined in (3.14). Consequently, the 'distance' between $y^a_{[i+1]}$ and $y^b_{[i+1]}$ reads

$$\left\| y^a_{[i+1]} - y^b_{[i+1]} \right\|_{\ell_2^{n_y}} = \left\| \left(\hat{\mathcal{F}}_{yu} \circ \hat{\mathcal{F}}_{uy}\right) y^a_{[i]} - \left(\hat{\mathcal{F}}_{yu} \circ \hat{\mathcal{F}}_{uy}\right) y^b_{[i]} \right\|_{\ell_2^{n_y}}. \tag{B.8}$$

This distance can be upper bounded using the following two inequalities:

1. By the incremental sector bounds $[-I_{n_y}, I_{n_y}]$, the following inequality for the nonlinear steady-state operator $\hat{\mathcal{F}}_{uy}$ holds true for any $y_1, y_2 \in \ell_2^{n_y}(N)$:

$$\|u_1 - u_2\|_{\ell_2^{n_u}} = \left\| \hat{\mathcal{F}}_{uy} y_1 - \hat{\mathcal{F}}_{uy} y_2 \right\|_{\ell_2^{n_u}} \leq K_\varphi \|y_1 - y_2\|_{\ell_2^{n_y}} \tag{B.9}$$

   with $K_\varphi = 1$.

2. The following inequality for the linear steady-state operator $\hat{\mathcal{F}}_{yu}$ holds true for any $u_1, u_2 \in \ell_2^{n_u}(N)$:

$$\|y_1 - y_2\|_{\ell_2^{n_y}} = \left\| \hat{\mathcal{F}}_{yu} u_1 - \hat{\mathcal{F}}_{yu} u_2 \right\|_{\ell_2^{n_u}} \leq K_\Sigma \|u_1 - u_2\|_{\ell_2^{n_u}}. \tag{B.10}$$

   Under the conditions of Theorem 3.2, the conditions of Lemma B.1 in Appendix B.4 are satisfied, which guarantees $K_\Sigma = \left\| \mathcal{G}_{\Sigma_{yu}}(\varkappa) \right\|_{\mathcal{H}_\infty} < 1$.

Combining inequalities (B.9) and (B.10) reveals the following bound for any $y_1, y_2 \in \ell_2^{n_y}(N)$:

$$\left\| \left(\hat{\mathcal{F}}_{yu} \circ \hat{\mathcal{F}}_{uy}\right) y_1 - \left(\hat{\mathcal{F}}_{yu} \circ \hat{\mathcal{F}}_{uy}\right) y_2 \right\|_{\ell_2^{n_y}} \leq K_\varphi K_\Sigma \|y_1 - y_2\|_{\ell_2^{n_y}},$$

which is the right-hand side of (B.8) after substituting $y_1 = y^a_{[i]}$ and $y_2 = y^b_{[i]}$. Therefore, we conclude

$$\left\| y^a_{[i+1]} - y^b_{[i+1]} \right\|_{\ell_2^{n_y}} \leq \underbrace{K_\varphi K_\Sigma}_{<1} \left\| y^a_{[i]} - y^b_{[i]} \right\|_{\ell_2^{n_y}}. \tag{B.11}$$

Since $K_\varphi K_\Sigma < 1$, the composed mapping $(\hat{\mathcal{F}}_{yu} \circ \hat{\mathcal{F}}_{uy})$ is a contraction mapping having a unique fixed point $\bar{y}$. Noting that the true steady-state solution $\bar{y}$ of (Σ, φ) is a fixed point of the iterative procedure in (3.15), i.e.,

$$\bar{y} = \left(\hat{\mathcal{F}}_{yu} \circ \hat{\mathcal{F}}_{uy}\right) \bar{y} + \hat{\mathcal{F}}_{yw} w, \tag{B.12}$$

we conclude that the limit of the iterative procedure (3.15), i.e., $y_{[i]}$ for $i \to \infty$, equals the true steady-state solution $\bar{y}$ of (Σ, φ). □

B.3 Proof of Theorem 3.3

Proof. This proof uses the notations as introduced in Section 3.3.2 to improve readability. The product- and chain rule for differentiation are applied to the model (3.2), such that

$$x_\theta(k+1) = A_\theta \bar{x}(k) + A x_\theta(k) + B_\theta \bar{u}(k) + B u_\theta(k) + L_\theta w(k), \tag{B.13a}$$
$$y_\theta(k) = C_\theta \bar{x}(k) + C x_\theta(k) + D_\theta w(k), \tag{B.13b}$$
$$z_\theta(k) = F_\theta \bar{x}(k) + F x_\theta(k) + G_\theta \bar{u}(k) + G u_\theta(k) + H_\theta w(k), \tag{B.13c}$$
$$u_\theta(k) = \varphi_\theta(\bar{y}(k), k) + \varphi_y\big(\bar{y}(k), k\big) y_\theta(k), \tag{B.13d}$$

where it used that $w_\theta = 0$. All partial derivatives exist by Assumption 3.1. After substituting (B.13d) into (B.13a) and (B.13c), we arrive at the parameter sensitivity system $(\Sigma^\theta,\ \varphi^\theta)$ as in Theorem 3.3 with a nonlinearity $\varphi^\theta = \varphi_y\big(\bar{y}(k), k\big) y_\theta(k)$. The incremental sector condition on φ guarantees that $\varphi_y\big(\bar{y}(k), k\big)$ also satisfies $[-I_{n_y}, I_{n_y}]$, guaranteeing that the nonlinearity φ^θ in (3.19) also satisfies the incrementally sector bounds $[-I_{n_y}, I_{n_y}]$.

The sensitivity of steady-state model output $\bar{z}$ with respect to a model parameter θ can thus be described by the steady-state solution of the parameter sensitivity model. Via (3.11) and (3.17), this steady-state output sensitivity is directly related to the objective function gradient with respect to the model parameter. Finally, recall that $(\Sigma,\ \varphi)$ is GEC by verifying the conditions in Theorem 3.1. These conditions only dependent on the variables $\{A,\ B,\ C\}$, and because all these variables are shared between $(\Sigma,\ \varphi)$ and $(\Sigma^\theta,\ \varphi^\theta)$, the parameter sensitivity model is GEC as well. □

B.4 Technical lemma

The proof of Theorem 3.2 (contraction of the MTF algorithm) uses a connection between the conditions in Theorem 3.1 and the $\mathcal{H}_\infty$-norm of the transfer function from the input u to output y of the Lur'e-type model (Σ, φ), as presented in the following lemma. The $\mathcal{H}_\infty$-norm of the transfer function matrix $\mathcal{G}_{\Sigma_{yu}}$ is defined as follows:

$$\left\|\mathcal{G}_{\Sigma_{yu}}(\varkappa)\right\|_{\mathcal{H}_\infty} := \sup_{0\leq\omega<2\pi} \bar{\sigma}\left(\mathcal{G}_{\Sigma_{yu}}(i\omega)\right),$$

where $\bar{\sigma}$ denotes the maximum singular value of a matrix.

Lemma B.1. *Consider a Lur'e-type model* (3.2)*. Under the conditions in Theorem 3.1, the following inequality holds true:*

$$\left\|\mathcal{G}_{\Sigma_{yu}}(\varkappa)\right\|_{\mathcal{H}_\infty} < 1. \tag{B.14}$$

▲

Proof. By application of Lemma B.2 in Appendix B.5 for $\gamma = 1$, the inequality (B.14) is equivalent to the following matrix inequality:

$$\begin{bmatrix} A^\top PA - P & A^\top PB & C^\top \\ \star & B^\top PB - I_{n_u} & 0_{n_y\times n_u} \\ \star & \star & -I_{n_y} \end{bmatrix} \prec 0, \tag{B.15}$$

where A, B, C of model (3.2) parametrize the state-space realization of the transfer function $\mathcal{G}_{\Sigma_{yu}}$. The relation $y = Cx$ is used to manipulate inequality (B.15) by pre- and post-multiplying it with the matrix

$$\begin{bmatrix} I_{n_x} & 0_{n_x\times n_u} & C^\top \\ 0_{n_u\times n_x} & I_{n_u} & 0_{n_u\times n_y} \end{bmatrix} \tag{B.16}$$

and its transpose, respectively. This results in the equivalent inequality:

$$\begin{bmatrix} A^\top PA - P + C^\top C & A^\top PB \\ \star & B^\top PB - I_{n_u} \end{bmatrix} \prec 0. \tag{B.17}$$

Inequality (B.17) is equivalent to inequality (3.5), which is feasible under the conditions of Theorem 3.1. Therefore, under the conditions of Theorem 3.1, all conditions of Lemma B.2 are satisfied for $\gamma = 1$, where this γ is an upper bound as follows:

$$\left\|\mathcal{G}_{\Sigma_{yu}}(\varkappa)\right\|_{\mathcal{H}_\infty} < \gamma = 1, \tag{B.18}$$

which closes the proof. □

B.5 Technical results recalled from the literature

This appendix recalls results used in the proofs of Theorems 3.1 and 3.2.

B.5.1 Convergence characterization [126]

A characterization of the GEC property for models (3.2) without feedthrough from input w to output y, i.e., $D = 0_{n_y \times n_w}$ is given in [126, Theorem 7], for nonlinearities that are sector bounded within $[0_{n_y}, \Omega], 0 \prec \Omega \in \mathbb{D}^{n_y}$ and incrementally sector bounded within $[0, \overline{\Omega}], 0 \prec \overline{\Omega} \in \mathbb{D}^{n_y}$.

Theorem B.1. *Consider the model* (Σ, φ) *in* (3.2) *with* $D = 0_{n_y \times n_w}$. *Consider* P_{11}, $P_{21} \in \mathbb{S}^{n_x}$ *positive definite matrices,* $S_1, S_2 \in \mathbb{D}^{n_y}$ *diagonal positive definite, a positive scalar* $\tau \in (0, 1)$ *such that the following inequalities hold:*

$$\begin{bmatrix} A^\top \\ B^\top \end{bmatrix} P_{11} \begin{bmatrix} A^\top \\ B^\top \end{bmatrix}^\top - \begin{bmatrix} \tau P_{11} & -C^\top S_1 \Omega \\ \star & 2S_1 \end{bmatrix} < 0, \tag{B.19}$$

$$\begin{bmatrix} A^\top \\ B^\top \end{bmatrix} P_{21} \begin{bmatrix} A^\top \\ B^\top \end{bmatrix}^\top - \begin{bmatrix} P_{21} & -C^\top S_2 \overline{\Omega} \\ \star & 2S_2 \end{bmatrix} < 0. \tag{B.20}$$

Then, the Lur'e-type model (3.2) *with* $D = 0_{n_y \times n_w}$ *is GEC for the class of bounded inputs as per Definition 3.2.* ▲

Proof. The proof follows a similar line of reasoning as [126, Theorem 7] and is, therefore, omitted. □

The proof of Theorem 3.1 in Appendix B.1 makes use of the fact that, if $\Omega = \overline{\Omega}$, the satisfaction of (B.20) is implied by satisfaction of (B.19) for any $\tau \in (0, 1)$.

B.5.2 The Bounded Real Lemma [89]

The Bounded Real Lemma relates the $\mathcal{H}_\infty$ norm of a transfer function matrix to the feasibility of a matrix inequality. It is used in the proof of Theorem 3.2 in Appendix B.4.

Lemma B.2 ([89, Lem. 5.1])**.** *Consider a discrete-time LTI model* Σ *whose transfer matrix* $\mathcal{G}_\Sigma(\varkappa)$ *admits a state-space representation* $(A,\ B,\ C,\ D)$, *where* $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times n_u}$, $C \in \mathbb{R}^{n_y \times n_x}$, $D \in \mathbb{R}^{n_y \times n_u}$. *Then, the inequality*

$$\|\mathcal{G}_\Sigma(\varkappa)\|_{\mathcal{H}_\infty} < \gamma$$

holds if and only if there exist $0 \prec P \in \mathbb{S}^{n_x}$ *and* $\gamma \in \mathbb{R}_{>0}$, *such that*

$$\begin{bmatrix} A^\top P A - P & A^\top P B & C^\top \\ \star & B^\top P B - \gamma I_{n_u} & D^\top \\ \star & \star & -\gamma I_{n_y} \end{bmatrix} \prec 0. \tag{B.21}$$

▲

Proof. See [89, Lemma 5.1]. □

C Appendices to Chapter 4

C.1 Proof of Theorem 4.2

Proof. For notational convenience, the superscript L_2^M to the norm operator is omitted in this proof. Operator g_{yu} and frequency-domain counterparts of g_{yu}, g_{yw}, are defined in (4.5) and (4.6), respectively.

Consider two different sequences $\mathbf{y}^{(a)}_{([i+1],\eta M)} \in L_2^{\eta M}$ and $\mathbf{y}^{(b)}_{([i+1],\eta M)} \in L_2^{\eta M}$ with integer $i \geq 0$ and define their difference as follows: $\delta\mathbf{y}_{[i+1]} \coloneqq \mathbf{y}^{(a)}_{([i+1],\eta M)} - \mathbf{y}^{(b)}_{([i+1],\eta M)}$. Application of the step (4.4b) to $\left\|\delta\mathbf{y}_{[i+1]}\right\|$ results in

$$\left\|\delta\mathbf{y}_{[i+1]}\right\| = \left\|g_{yu}\delta\mathbf{u}_{[i]}\right\| \leq \gamma_{yu}\left\|\delta\mathbf{u}_{[i]}\right\|, \tag{C.1}$$

with γ_{yu} defined in (4.7) and $\delta\mathbf{u}_{[i]}$ defined similarly to $\delta\mathbf{y}_{[i+1]}$. Application of step (4.4a) to (C.1) results in

$$\left\|\delta\mathbf{y}_{[i+1]}\right\| \leq \gamma_{yu}\left\|g_{uy}\mathbf{y}^{(a)}_{([i],\eta M)} - g_{uy}\mathbf{y}^{(b)}_{([i],\eta M)}\right\| \leq \gamma_{yu}K_{\varphi}\left\|\delta\mathbf{y}_{[i]}\right\|,$$

where condition ***C2*** of Theorem 4.1 is used in the latter inequality. The iterative mappings in (4.4a) and (4.4b) constitute a contraction mapping, since $\gamma_{yu}K_{\varphi} < 1$ holds true by Assumption 4.1. By the fact that $L_2^{\eta M}$ is a Banach space, we conclude by the Banach fixed-point theorem that there exists a limit $\bar{\mathbf{y}}_{\eta M}$ that satisfies $\bar{\mathbf{y}}_{\eta M} = (g_{yu} \circ g_{uy})\,\bar{\mathbf{y}}_{\eta M} + g_{yw}\mathbf{w}_{(0,\eta M)}$. This limit can be found by the iterative procedure (4.4a) and (4.4b). □

C.2 Proof of Lemma 4.1

Proof. As a first step (i), we derive a bound for the Fourier coefficients $\bar{U}$ of $\bar{u}^{\eta M}$, which depends on properties of $\bar{y}^{\eta M}$. In the second step (ii), we derive the required properties of $\bar{y}^{\eta M}$. In the third step (iii), the results from steps (i) and (ii) are

combined to derive the bound (4.10). In the final step (iv), we formulate the bound in (4.11). Throughout this proof, we make use of the triangular inequality. Furthermore, for $a \in C^{0,1}_{K_a}(T), T \geq 0, 0 \leq K_a < \infty$, the notation $(\delta a)_{t,\tau}$ is used: $(\delta a)_{t,\tau} \coloneqq a(t) - a(t-\tau), t, \tau \in \mathbb{R}$.

(i) Bound on the Fourier coefficients magnitude. The derivation of the bound (4.10) is inspired by [128, p. 26]. Consider the truncated function $\bar{y}^{\eta M} \in C^{0,1}_{K_y^{\eta M}}(T), T \geq 0$ with some $0 \leq K_y^{\eta M} < +\infty$ and $\bar{u} = \varphi(\bar{y}^{\eta M}) \in C^{0,1}_{K_u}(T)$ with $K_u \coloneqq K_\varphi K_y^{\eta M}$. By definition, we have

$$\bar{U}[m] = \frac{1}{T}\int_0^T \bar{u}(t)\exp\{-im\omega t\}\mathrm{d}t, \tag{C.2}$$

$m \in \mathbb{Z}$ with $\omega = \frac{2\pi}{T}$. Thanks to the shifting property of Fourier series, see, .e.g, [128, p. 4], the following manipulation for $m \neq 0$ can be performed:

$$\bar{U}[m] = \frac{-1}{T}\int_0^T \bar{u}(t-\tau)\exp\{-im\omega t\}\mathrm{d}t \tag{C.3}$$

with $\tau \coloneqq T/(2m)$. The average of (C.2) and (C.3) is

$$\bar{U}[m] = \frac{1}{2T}\int_0^T \delta\bar{u}_{t,\tau}\exp\{-im\omega t\}\mathrm{d}t,$$

for $m \neq 0$. Proceeding by taking the absolute value and using Condition ***C2*** of Theorem 4.1, leaves for $m \neq 0$:

$$\left|\bar{U}[m]\right| \leq \frac{K_\varphi}{2T}\int_0^T \left|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right|\mathrm{d}t\ . \tag{C.4}$$

Assuming $\bar{y}^{\eta M} \in C^{0,1}_{K_y^{\eta M}}(T)$ (this property is proven in Step (ii) below), the inequality $|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}| \leq K_y^{\eta M}|\tau|, \tau = T/(2m)$ can be used in (C.4) to write:

$$\left|\bar{U}[m]\right| \leq \frac{K_\varphi K_y^{\eta M}}{4|m|}\int_0^T \mathrm{d}t = \frac{TK_\varphi K_y^{\eta M}}{4|m|}, m \neq 0. \tag{C.5}$$

(ii) Properties of the nonlinearity input signal. Next, we derive an upper bound for the constant $K_y^{\eta M}$, proving that the steady-state signal $\bar{y}^{\eta M}$ is in the space $C^{0,1}_{K_y^{\eta M}}(T)$. Consider the absolute value of the mismatch:

$$\left|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right| \leq \left|g_{yu}\left(\delta\bar{u}^{\eta M}\right)_{t,\tau}\right| + \left|g_{yw}\left(\delta w_0^{\eta M}\right)_{t,\tau}\right|. \tag{C.6}$$

The frequency-domain counterparts of the steady-state operators g_{yu} and g_{yw} are defined in (4.6a) and (4.6b), respectively. It is well-known that (see, e.g., [196])

$$|g_{yu}a| \leq \gamma_{yu}^{\eta M}|a|, \quad |g_{yw}b| \leq \gamma_{yw}^{\eta M}|b|, \tag{C.7}$$

where $\gamma_{yu}^{\eta M}$ and $\gamma_{yw}^{\eta M}$ are defined in (4.7) and $a = \left(\delta\bar{u}^{\eta M}\right)_{t,\tau}, b = \left(\delta w_0^{\eta M}\right)_{t,\tau}$. Inequality (C.7) holds for any $t, \tau \in \mathbb{R}$, where $|g_{yu}a|$ is the evaluation at t, τ of the absolute value of the function resulting from operating g_{yu} on a. The inequalities in (C.7) can be used in (C.6) together with Condition ***C2*** of Theorem 4.1 (by Assumption 4.1) to write:

$$\left|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right| \leq \gamma_{yu}^{\eta M} K_\varphi \left|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right| + \gamma_{yw}^{\eta M} \left|\left(\delta w_0^{\eta M}\right)_{t,\tau}\right|.$$

Collecting terms and using the constant $K_{w_0}^{\eta M}$ of Assumption 4.2 results in:

$$\left|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right| \leq \frac{\gamma_{yw}^{\eta M} K_{w_0}^{\eta M}}{1 - \gamma_{yu}^{\eta M} K_\varphi} |\tau|, \tag{C.8}$$

which is well-defined since the denominator $\gamma_{yu} K_\varphi < 1$ by Assumption 4.1. For condition (C.8) to hold, the constant $K_y^{\eta M}$ must satisfy:

$$0 \leq K_y^{\eta M} \leq \frac{\gamma_{yw}^{\eta M} K_{w_0}^{\eta M}}{1 - \gamma_{yu}^{\eta M} K_\varphi} < +\infty, \tag{C.9}$$

which guarantees that $\bar{y}^{\eta M} \in C^{0,1}_{K_y^{\eta M}}(T)$.

(iii) Bound (4.10)**.** Combining the bound (C.9) on $K_y^{\eta M}$ with the bound (C.5) on the magnitude of $\bar{U}$ results in the derived bound in (4.10).

(iv) Bound (4.11)**.** The definition of truncated signals in (4.1) allows writing $\bar{u}^{\eta M} = \sum_{m=-\eta M}^{\eta M-1} \bar{U}[m] \exp\{im\omega t\}$. Subsequently, we can derive the following bound

$$\left\|\bar{u} - \bar{u}^{\eta M}\right\|_{L_2}^2 \leq 2 \sum_{m=\eta M}^{\infty} |\bar{U}[m]|^2,$$

where $|\bar{U}[m]|^2 = |\bar{U}[-m]|^2, m \in \mathbb{Z}$ is used. Using the upper bound (4.10) for $|\bar{U}[m]|, m \neq 0$, we find that:

$$\begin{aligned}\left\|\bar{u} - \bar{u}^{\eta M}\right\|_{L_2}^2 &\leq \frac{(T K_\varphi K_{w_0}^{\eta M} \gamma_{yw}^{\eta M})^2}{8(1 - \gamma_{yu}^{\eta M} K_\varphi)^2} \sum_{m=\eta M}^{\infty} \frac{1}{m^2} \\ &\leq \frac{(T K_\varphi K_{w_0}^{\eta M} \gamma_{yw}^{\eta M})^2}{8(1 - \gamma_{yu}^{\eta M} K_\varphi)^2} \left(\sum_{m=1}^{\infty} \frac{1}{m^2} - \sum_{m=1}^{\eta M-1} \frac{1}{m^2}\right).\end{aligned}$$

Taking the square root of this expression and noting $\sum_{m=1}^{\infty} \frac{1}{m} = \zeta(2)$ results in (4.11). The Riemann-zeta function $\zeta(2)$, defined in (4.9), *converges* [278]. □

C.3 Proof of Theorem 4.3

Proof. Firstly (i), we introduce an alternative iterative procedure that is equivalent to (4.4a) - (4.4b). Secondly (ii), we derive the bound (4.12). Thirdly (iii), we present an asymptotic analysis supporting the validity of (4.13). For notational convenience, we do not write the superscript L_2 to the norm operator in this proof.

(i) Equivalent iterative procedure. Consider the iterative procedure:

$$u_{[i+1]}^{\eta M} = \breve{\tilde{g}}_{uy} y_{[i]}^{\eta M}, \tag{C.10a}$$

$$y_{[i+1]}^{\eta M} = \tilde{g}_{yu} u_{[i+1]}^{\eta M} + \tilde{g}_{yw} w_0^{\eta M} \tag{C.10b}$$

with the frequency-domain counterpart $\hat{\tilde{g}}_{yu}$ and $\hat{\tilde{g}}_{uy}$ of operators $\tilde{g}_{yu}$ and $\tilde{g}_{yw}$, respectively, defined as follows:

$$\hat{\tilde{g}}_{yu} U[m] \coloneqq G_{yu}(i\omega_0 m) U[m], \tag{C.11a}$$

$$\hat{\tilde{g}}_{yw} W_0[m] \coloneqq G_{yw}(i\omega_0 m) W_0[m], \tag{C.11b}$$

for $m \in \mathbb{Z}$, where $\omega_0 \coloneqq 2\pi/T$. Operators $\hat{\tilde{g}}_{yu}$ and $\hat{\tilde{g}}_{yw}$ work on an infinite number of harmonics as opposed to $\hat{g}_{yu}$ and $\hat{g}_{yw}$ in (4.6a) and (4.6b), respectively, which work only on the first ηM harmonics. Consider the sequences $\mathbf{y}_{([i],\eta M)}, \mathbf{u}_{([i],\eta M)}, \mathbf{w}_{(0,\eta M)}$ and their continuous-time counterparts, obtained via (4.2), denoted by $y_{[i]}^{\eta M}, u_{[i]}^{\eta M}, w_0^{\eta M}$, respectively, containing only ηM harmonics. Then, by linearity of operators $\tilde{g}_{yu}, \tilde{g}_{yw}, g_{yu}$, and g_{yw}, operation (C.10b) is equivalent to operation (4.4b) for the first ηM harmonics.

Next, we show that the operation (C.10a) is equivalent to operation (4.4a) under an appropriate definition of operator $\breve{\tilde{g}}_{uy}$. Operation (4.4a) is by definition equivalent to:

$$\{\mathbf{u}_{([i+1],\eta M)}[k]\}_{k=1}^{2\eta M} = \{-\varphi\left(\mathbf{y}_{([i],\eta M)}[k]\right)\}_{k=1}^{2\eta M}. \tag{C.12}$$

However, taking the continuous-time counterpart $y_{[i]}^{\eta M}$ of $\mathbf{y}_{([i],\eta M)}$ through the nonlinearity results in

$$u_{[i+1]} = \varphi\left(y_{[i]}^{\eta M}\right), \tag{C.13}$$

which now possibly contains an infinite number of harmonics. Observe that the samples $\mathbf{u}_{([i+1],\eta M)}$ obtained via (C.12) are equivalent to the samples $\{u_{[i+1]}(t_k)\}_{k=1}^{2\eta M}$ obtained via (C.13). However, sampling the continuous-time signal $u_{[i+1]}$ gives rise to aliasing. Therefore, we achieve consistency between the first ηM harmonics of $\mathbf{u}_{([i+1],\eta M)}$ and $u_{[i+1]}$, if aliasing is handled:

$$u_{[i+1]}^{\eta M} = \breve{\tilde{g}}_{uy} y_{[i]}^{\eta M} = (\varphi(y_{[i]}^{\eta M}))^{\eta M} + \underbrace{g_{alias} u_{([i],tail)}^{\eta M}}_{a_{[i]}}, \tag{C.14}$$

where $u^{\eta M}_{([i],tail)} = \left(\varphi(y^{\eta M}_{[i]}) - \left(\varphi(y^{\eta M}_{[i]})\right)^{\eta M}\right)$ is the 'tail' of $\varphi(y^{\eta M}_{[i]})$, containing only harmonics beyond ηM. The operator g_{alias} is the aliasing operator [205], which we do not define explicitly. However, below, we do make use of the fact that for any $u^{\eta M}_{([i],tail)} \in L_2(T)$, aliasing does not increase the signal norm, but only folds back (and attenuates) harmonics [205], i.e.,

$$\left\|a_{(i)}\right\| = \left\|g_{alias} u^{\eta M}_{([i],tail)}\right\| \leq \left\|u^{\eta M}_{([i],tail)}\right\|. \tag{C.15}$$

(ii) Derivation of upper bound (4.12)**.** Consider

$$\left\|\breve{\tilde{z}} - \tilde{z}^{\eta M}\right\| = \left\|\tilde{g}_{zu}\left(\breve{\bar{u}} - \bar{u}^{\eta M}\right) + \tilde{g}_{zw}\left(w_0 - w_0^{\eta M} + v_w\right) + v_z\right\| \tag{C.16a}$$

$$\leq \gamma_{zu}\left\|\breve{\bar{u}} - \bar{u}^{\eta M}\right\| + \bar{\gamma}^{\eta M}_{zw}\left\|w_0 - w_0^{\eta M}\right\| + \gamma_{zw}\Delta_w + \Delta_z. \tag{C.16b}$$

Here, the constants Δ_w and Δ_z as in Assumption 4.4 are used to bound $v_w, v_z \in L_2(T)$. The linear operators $\tilde{g}_{zu}$ and $\tilde{g}_{zw}$ are defined similarly to $\tilde{g}_{yu}$ and $\tilde{g}_{yw}$ in (C.11a) and (C.11b), respectively. The constants γ_{zu}, γ_{zw} and $\bar{\gamma}^{\eta M}_{zw}$ are defined in (4.7). Next, we deal with the term $\left\|\breve{\bar{u}} - \bar{u}^{\eta M}\right\|$, where $\breve{\bar{u}} = \varphi(\breve{\bar{y}})$ and the function $\bar{u}^{\eta M} = (\varphi(\bar{y}^{\eta M}))^{\eta M} + \bar{a}$ consistent with (C.14). We write:

$$\begin{aligned}\left\|\breve{\bar{u}} - \bar{u}^{\eta M}\right\| &= \left\|\varphi(\breve{\bar{y}}) - (\varphi(\bar{y}^{\eta M}))^{\eta M} + \bar{a}\right\| \\ &\leq \left\|\varphi(\breve{\bar{y}}) - \varphi(\bar{y}^{\eta M}) + \bar{u}^{\eta M}_{(tail)}\right\| + \left\|\bar{u}^{\eta M}_{(tail)}\right\| \\ &\leq K_\varphi\left\|\breve{\bar{y}} - \bar{y}^{\eta M}\right\| + 2\gamma_a^{\eta M} K_{w_0}^{\eta M},\end{aligned} \tag{C.17}$$

where $\|\bar{a}\|$ is bounded using (C.15). In the last step, the inequality ***C2*** of Theorem 4.1 is used and inequality (4.11) of Lemma 4.1 is applied to upper bound $\bar{u}^{\eta M}_{(tail)}$ by $\gamma_a^{\eta M} K_{w_0}^{\eta M}$ with $\gamma_a^{\eta M}$ as in (4.8a) and $K_{w_0}^{\eta M}$ as in Assumption 4.2.

Next, we derive a bound for $\left\|\breve{\bar{y}} - \bar{y}^{\eta M}\right\|$ as follows:

$$\begin{aligned}\left\|\breve{\bar{y}} - \bar{y}^{\eta M}\right\| \leq& \left\|\tilde{g}_{yu}(\breve{\bar{u}} - \bar{u}^{\eta M})\right\| + \left\|\tilde{g}_{yw}\left(w_0 - w_0^{\eta M}\right)\right\| + \left\|\tilde{g}_{yw} v_w\right\| \\ \leq& \gamma_{yu}\left\|\breve{\bar{u}} - \bar{u}^{\eta M}\right\| + \bar{\gamma}^{\eta M}_{yw}\left\|w_0 - w_0^{\eta M}\right\| + \gamma_{yw}\Delta_w.\end{aligned} \tag{C.18}$$

The constants γ_{yu}, γ_{yw} and $\bar{\gamma}^{\eta M}_{yw}$ are defined in (4.7). Substitution of inequaltiy (C.18) into (C.17) results in:

$$\begin{aligned}\left\|\breve{\bar{u}} - \bar{u}^{\eta M}\right\| \leq& \frac{K_\varphi \bar{\gamma}^{\eta M}_{yw}}{1 - \gamma_{yu} K_\varphi}\left\|w_0 - w_0^{\eta M}\right\| \\ &+ \frac{K_\varphi \gamma_{yw}}{1 - \gamma_{yu} K_\varphi}\Delta_w + \frac{2\gamma_a^{\eta M}}{1 - \gamma_{yu} K_\varphi} K_{w_0}^{\eta M},\end{aligned} \tag{C.19}$$

which is well-defined since the denominator $1 - \gamma_{yu} K_\varphi > 0$ thanks to Assumption 4.1. Substituting inequality (C.19) into (C.16b) yields the bound (4.12).

(iii) Asymptotic analysis. Consider the bound in (4.12) and $\Delta_w = \Delta_z = 0$. In the asymptotic case $\eta \to \infty$ or $M \to \infty$, the first term is zero thanks to $\lim_{\eta\to\infty} \|w_0 - w_0^{\eta M}\| = \lim_{M\to\infty} \|w_0 - w_0^{\eta M}\| = 0$ by the Riesz-Fischer theorem. The only remaining term is then related to $\gamma_a^{\eta M}$, which also vanishes in either of the limits $\eta \to \infty$ or $M \to \infty$, which completes the proof. □

C.4 Proof of Theorem 4.4

Proof. In the first step (i), we present an equivalent upper bound that can be split into two parts, each part further addressed in steps (ii) and (iii). In the fourth step (iv), we present the bound (4.14). In the final step (v), an asymptotic analysis is presented.

(i) Equivalent upper bound. Denote by $\breve{\breve{z}}_S^M$ the continuous-time counterpart of the measured sequence $\breve{\breve{\mathbf{z}}}_M$, which has only contents up to the M-th harmonic. Similarly, $\left(\bar{z}^{\eta M}\right)^M$ denotes the continuous-time counterpart of the computed sequence $\left(\bar{\mathbf{z}}_{\eta M}\right)^M$. Using these definitions, we observe:

$$\left\| \left\{ \breve{\breve{z}}_S^M(t_k) - \left(\bar{z}^{\eta M}\right)^M(t_k) \right\}_{k=1}^{2M} \right\|_{L_2^M} = \left\| \breve{\breve{z}}_S^M - \left(\bar{z}^{\eta M}\right)^M \right\|_{L_2}. \tag{C.20}$$

From here onwards, we only work with continuous-time signals and, therefore, do not write the subscript L_2 for the norm operator anymore for notational convenience.

First, notice that $\breve{\breve{z}}_S^M$ is related to $\breve{\breve{z}}_F$ according to:

$$\breve{\breve{z}}_S^M = \breve{\breve{z}}_F^M + g_{alias}\left(\breve{\breve{z}}_F - \breve{\breve{z}}_F^M\right) \tag{C.21}$$

with g_{alias} the aliasing operator [205]. Using (C.21) in (C.20), we write

$$\left\| \breve{\breve{z}}_F^M + g_{alias}\left(\breve{\breve{z}}_F - \breve{\breve{z}}_F^M\right) - \left(\bar{z}^{\eta M}\right)^M \right\| \leq \left\| \breve{\breve{z}}_F^M - \left(\bar{z}^{\eta M}\right)^M \right\| + \left\| \breve{\breve{z}}_F - \breve{\breve{z}}_F^M \right\|, \tag{C.22}$$

where inequality (C.15) is used to bound the effect of the aliasing operator g_{alias}.

(ii) Mismatch between the first M harmonics. Note that $\breve{\breve{z}}_F^M = \left(\breve{g}_{LPF}\breve{\breve{z}}\right)^M = \breve{\breve{z}}^M$, since by Assumption 4.3, the LPF does not affect the first M harmonic. Therefore, consider the mismatch

$$\begin{aligned}
\left\| \breve{\breve{z}}^M - \left(\bar{z}^{\eta M}\right)^M \right\| &= \left\| v_z^M + \left(\tilde{g}_{zu}\left(\breve{\breve{u}} - \bar{u}^{\eta M}\right)\right)^M + \left(\tilde{g}_{zw}\left(w_0 - w_0^{\eta M} + v_w\right)\right)^M \right\| \\
&\leq \Delta_z + \gamma_{zu}^M \left\| \breve{\breve{u}} - \bar{u}^{\eta M} \right\| + \underbrace{\left\| \left(\tilde{g}_{zw}\left(w_0 - w_0^{\eta M}\right)\right)^M \right\|}_{=0} + \gamma_{zw}^M \Delta_w \\
&\leq \Delta_z + \gamma_{zu}^M \left\| \breve{\breve{u}} - \bar{u}^{\eta M} \right\| + \gamma_{zw}^M \Delta_w.
\end{aligned} \tag{C.23}$$

Substituting the bound (C.19) for $\left\|\breve{\bar{u}} - \bar{u}^{\eta M}\right\|$ results in:

$$\begin{aligned}\left\|\breve{\bar{z}}^{M} - \left(\bar{z}^{\eta M}\right)^{M}\right\| =& \left\|w_0 - w_0^{\eta M}\right\| \left(\frac{K_\varphi \gamma_{zu}^{M} \bar{\gamma}_{yw}^{\eta M}}{1-\gamma_{yu}K_\varphi}\right) \\ &+ K_{w_0}^{\eta M} \frac{2\gamma_a^{\eta M}\gamma_{zu}^{M}}{1-\gamma_{yu}K_\varphi} + \Delta_w \left(\gamma_{zw}^{M} + \frac{K_\varphi \gamma_{zu}^{M}\gamma_{yw}}{1-\gamma_{yu}K_\varphi}\right) + \Delta_z.\end{aligned} \tag{C.24}$$

(iii) Effect of aliasing. Consider the Fourier coefficients $\breve{\bar{Z}}_F[m]$ of $\breve{\bar{z}}_F$ and note that $\left\|\breve{\bar{z}}_F - \breve{\bar{z}}_F^{M}\right\|^2 \leq 2\sum_{m=M}^{\infty} |\breve{\bar{Z}}_F[m]|^2$. Now, using the fact that $\breve{\bar{z}}_F$ is a low-passed version of $\breve{\bar{z}}$, we arrive at:

$$\left\|\breve{\bar{z}}_F - \breve{\bar{z}}_F^{M}\right\|^2 \leq 2\sum_{m=M}^{\infty} |\breve{\bar{Z}}[m]|^2 |G_{LPF}(im\omega_0)|^2. \tag{C.25}$$

Taking similar steps as in the proof of Lemma 4.1 steps (i - iii), we can show that

$$|\breve{\bar{Z}}[m]| \leq \frac{TK_z}{4|m|}, \tag{C.26}$$

where $0 \leq K_z < +\infty$ is such that $\breve{\bar{z}}_F \in C_{K_z}^{0,1}(T)$. Let us find an upper bound for K_z next by analyzing the following mismatch:

$$\begin{aligned}\left|(\delta\breve{\bar{z}})_{t,\tau}\right| &\leq |(\delta v_z)_{t,\tau}| + \left|g_{zu}(\delta\breve{\bar{u}})_{t,\tau}\right| + |g_{zw}(\delta w_0)_{t,\tau}| + |g_{zw}(\delta v_w)_{t,\tau}|, &\text{(C.27a)}\\ &\leq |(\delta v_z)_{t,\tau}| + \gamma_{zu}\left|(\delta\breve{\bar{u}})_{t,\tau}\right| + \gamma_{zw}\left|(\delta w_0)_{t,\tau}\right| + \gamma_{zw}\left|(\delta v_w)_{t,\tau}\right|. &\text{(C.27b)}\end{aligned}$$

We use the fact that $\breve{\bar{u}} = \varphi(\breve{\bar{y}})$ to write:

$$\left|(\delta\breve{\bar{u}})_{t,\tau}\right| \leq K_\varphi \left|(\delta\breve{\bar{y}})_{t,\tau}\right|, \tag{C.28}$$

where

$$\left|(\delta\breve{\bar{y}})_{t,\tau}\right| \leq \gamma_{yu}\left|(\delta\breve{\bar{u}})_{t,\tau}\right| + \gamma_{yw}\left|(\delta w_0)_{t,\tau}\right| + \gamma_{yw}\left|(\delta v_w)_{t,\tau}\right|.$$

Substituting (C.28) in (C.27b) results in

$$\left|(\delta\breve{\bar{z}})_{t,\tau}\right| \leq |(\delta v_z)_{t,\tau}| + \gamma_z\left|(\delta w_0)_{t,\tau}\right| + \gamma_z\left|(\delta v_w)_{t,\tau}\right|$$

with

$$\gamma_z \coloneqq \gamma_{zw} + \frac{\gamma_{zu}\gamma_{yw}}{1-\gamma_{yu}K_\varphi} \tag{C.29}$$

being well-defined since $\gamma_{yu}K_\varphi < 1$ by Assumption 4.1. Next, we use the constants $K_{w_0}^{\circ}$ (in Assumption 4.2) and K_{v_z}, K_{v_w} (in Assumption 4.5) to formulate: $\left|(\delta\breve{\bar{z}})_{t,\tau}\right| \leq \left(K_{v_z} + \gamma_z\left(K_{w_0}^{\circ} + K_{v_w}\right)\right)|\tau|$, which can only be true if

$$0 \leq K_z \leq K_{v_z} + \gamma_z\left(K_{w_0}^{\circ} + K_{v_w}\right), \tag{C.30}$$

proving that $\breve{\breve{z}}_F \in C^{0,1}_{K_z}(T)$.

Now, we proceed by deriving an upper bound for $|G_{LPF}(im\omega_0)|^2$. Hereto, consider

$$|G_{LPF}(im\omega_0)|^2 = \frac{\omega_z^{2o_z}}{(m^2\omega_0^2 + \omega_z^2)^{o_z}}. \tag{C.31}$$

Next, we use the bound (C.26) for $|Z[m]|$ and substitute (C.31) for $|G_{LPF}(im\omega_0)|^2$ in (C.25) to find:

$$\left\|\breve{\breve{z}}_F - \breve{\breve{z}}_F^M\right\|^2 \leq \frac{(TK_z\omega_z^{o_z})^2}{8} \sum_{m=M}^{\infty} \frac{1}{m^2\left(m^2\omega_0^2 + \omega_z^2\right)^{o_z}}. \tag{C.32}$$

Convergence of the infinite series in (C.32) is analyzed as follows. Note that the following inequality holds:

$$\sum_{m=M}^{\infty} \frac{1}{m^2\left(m^2\omega_0^2 + \omega_z^2\right)^{o_z}} \leq \sum_{m=M}^{\infty} \frac{1}{m^{2+2o_z}\omega_0^{2o_z}}. \tag{C.33}$$

The latter infinite series is equivalent to:

$$\frac{\zeta(2(1+o_z))}{\omega_0^{2o_z}} - \sum_{m=1}^{M} \frac{1}{m^{2+2o_z}\omega_0^{2o_z}} \tag{C.34}$$

with $\zeta\left(2(1+o_z)\right)$ the converging Riemann-zeta function, defined in (4.9). By the monotone convergence theorem [311], the infinite series in (C.32) is also convergent.

Finally, we take the square root of (C.32) and use (C.30) to bound the constant K_z to yield:

$$\left\|\breve{\breve{z}}_F - \breve{\breve{z}}_F^M\right\| =: \gamma_F^M\gamma_z\left(K_{w_0}^{\circ} + K_{v_w}\right) + \gamma_F^M K_{v_z} \tag{C.35}$$

with γ_z defined in (C.29) and γ_F^M defined in (4.8b).

(iv) Derivation of upper bound (4.14)**.** Substituting the bounds (C.35) and (C.24) into (C.22) results in:

$$\begin{aligned}\left\|\breve{\breve{z}}_F - \left(\bar{z}^{\eta M}\right)^M\right\| \leq & \left\|w_0 - w_0^{\eta M}\right\|\left(\frac{K_\varphi\gamma_{zu}^M\bar{\gamma}_{yw}^{\eta M}}{1-\gamma_{yu}K_\varphi}\right) + \Delta_w\left(\gamma_{zw}^M + \frac{K_\varphi\gamma_{zu}^M\gamma_{yw}}{1-\gamma_{yu}K_\varphi}\right) + \Delta_z \\ & + K_{w_0}^{\eta M}\frac{2\gamma_a^{\eta M}\gamma_{zu}^M}{1-\gamma_{yu}K_\varphi} + K_{w_0}^{\circ}\gamma_F^M\gamma_z + K_{v_w}\gamma_F^M\gamma_z + K_{v_z}\gamma_F^M,\end{aligned} \tag{C.36}$$

which is equivalent to the upper bound in (4.14).

(v) Asymptotic analysis. Assuming $\Delta_w = \Delta_z = K_{v_w} = K_{v_z} = 0$, the only remaining terms in (C.36) are the terms concerning $\|w_0 - w_0^{\eta M}\|$, $K_{w_0}^{\eta M}$ and $K_{w_0}^{\circ}$.

The first term is zero by the Riesz-Fischer theorem for either η or $M \to \infty$. The second term is also zero for either η or $M \to \infty$, since $\lim_{\eta\to\infty} \gamma_a^{\eta M} = \lim_{M\to\infty} \gamma_a^{\eta M} = 0$. The last term is only zero for $M \to \infty$, i.e., $\lim_{M\to\infty} \gamma_F^M = 0$ (and not for only $\eta \to \infty$). Therefore, in the limit $M \to \infty$, the mismatch is $\lim_{M\to\infty} \left\| \breve{\bar{z}}_F - \left(\bar{z}^{\eta M}\right)^M \right\| = 0$, which is equivalent to (4.15) and completes the proof. □

D

Appendices to Chapter 5

D.1 Proof of Lemma 5.1

Proof. Let $\tilde{\Pi} = \begin{bmatrix}\tilde{\Pi}_0 & \tilde{\Pi}_1 & \cdots & \tilde{\Pi}_k\end{bmatrix} \in \mathbb{C}^{n\times\nu}, \tilde{\Pi}_0, \ldots \tilde{\Pi}_k \in \mathbb{C}^n$, and note that (5.8) can be written as follows:

$$\begin{aligned}
A\tilde{\Pi}_0 + B\ell_0 &= s^\star\tilde{\Pi}_0,\\
A\tilde{\Pi}_1 + B\ell_1 &= s^\star\tilde{\Pi}_1 + \tilde{\Pi}_0,\\
&\vdots\\
A\tilde{\Pi}_k + B\ell_k &= s^\star\tilde{\Pi}_k + \tilde{\Pi}_{k-1}.
\end{aligned}$$

As a result, we find the following equalities:

$$\begin{aligned}
\tilde{\Pi}_0 &= (s^\star I - A)^{-1}B\ell_0,\\
\tilde{\Pi}_1 &= -(s^\star I - A)^{-2}B\ell_0 + (s^\star I - A)^{-1}B\ell_1,\\
&\vdots\\
\tilde{\Pi}_{k-1} &= \sum_{i=0}^{k-1}(-1)^i(s^\star I - A)^{-(i+1)}B\ell_{k-1-i},\\
\tilde{\Pi}_k &= \sum_{i=0}^{k}(-1)^i(s^\star I - A)^{-(i+1)}B\ell_{k-i}.
\end{aligned}$$

Pre-multiplication of $\tilde{\Pi}_0, \ldots, \tilde{\Pi}_k$ by C yields the desired result. □

D.2 Proof of Lemma 5.2

Proof. The goal is to prove the relation between $C\tilde{\Pi}$ and the moments as given in (5.10). Hereto, we write $s^\star = a^\star + jb^\star$. From here onwards, we write respectively a and b instead of $a^\star$ and $b^\star$ for notational convenience.

Let us first prove the case where $a = 0$ and $k = 0$. Partition $\tilde{\Pi} \in \mathbb{R}^{n\times 2}$ as follows:

$$\tilde{\Pi} = \begin{bmatrix} \tilde{\Pi}_0 & \tilde{\Pi}_1 \end{bmatrix} \tag{D.1}$$

with $\Pi_0, \Pi_1 \in \mathbb{R}^n$. Using this partitioning, we find the following two matrix equalities from the Sylvester equation (5.11):

$$A\tilde{\Pi}_0 + B\ell_0 = -b\tilde{\Pi}_1 \tag{D.2a}$$

$$A\tilde{\Pi}_1 + B\ell_1 = b\tilde{\Pi}_0. \tag{D.2b}$$

Multiply (D.2a) by b to obtain

$$Ab\tilde{\Pi}_0 + bB\ell_0 = -b^2\tilde{\Pi}_1. \tag{D.3}$$

Now, substitute (D.2b) into (D.3) to obtain

$$A^2\tilde{\Pi}_1 + AB\ell_1 + bB\ell_0 = -b^2\tilde{\Pi}_1. \tag{D.4}$$

Solving for $\tilde{\Pi}_1$ yields

$$\tilde{\Pi}_1 = -(A^2 + b^2 I_n)^{-1} \left(AB\ell_1 + bB\ell_0\right). \tag{D.5}$$

Similarly, one can obtain

$$\tilde{\Pi}_0 = (A^2 + b^2 I_n)^{-1} \left(bB\ell_1 - AB\ell_0\right). \tag{D.6}$$

Let us now find an expression for $\eta_0(jb, \ell_0 + j\ell_1)$ from Definition 5.1:

$$\begin{aligned} \eta_0(jb, \ell_0 + j\ell_1) &= C(jbI_n - A)^{-1}B(\ell_0 + j\ell_1) \\ &= -C(A - jbI_n)^{-1}(A + jbI_n)^{-1}(A + jbI_n)B(\ell_0 + j\ell_1) \\ &= -C(A^2 + b^2 I_n)^{-1}(A + jbI_n)B(\ell_0 + j\ell_1). \end{aligned} \tag{D.7}$$

It can be verified from (D.7) that $\mathrm{Re}(\eta_0(jb, \ell_0 + j\ell_1))$ equals to $C\tilde{\Pi}_0$ in (D.6) and that $\mathrm{Im}(\eta_0(jb, \ell_0 + j\ell_1))$ equals to $C\tilde{\Pi}_1$ in (D.5), which concludes the proof for the case $a = 0$ and $k = 0$.

For the case where $a \neq 0$ and $k = 0$, it can be shown that the same steps as taken above result in A replaced by $A - aI_n$ in (D.6) and (D.5), which gives expressions for $\tilde{\Pi}_0$ and $\tilde{\Pi}_1$, respectively. Furthermore, it can be shown that in the expression of $\eta_0(a + jb, \ell_0 + j\ell_1)$ in (D.7), again A is replaced $A - aI_n$, which concludes the proof for $a \neq 0$ and $k = 0$.

Let us now turn to the case where $a = 0$ and $k = 1$. Hereto, we partition $\tilde{\Pi} \in \mathbb{R}^{n\times 4}$ as follows:

$$\tilde{\Pi} = \begin{bmatrix} \tilde{\Pi}_0 & \tilde{\Pi}_1 & \tilde{\Pi}_2 & \tilde{\Pi}_3 \end{bmatrix} \tag{D.8}$$

with $\Pi_0, \ldots, \Pi_3 \in \mathbb{R}^n$. Using this partitioning, we find the following four matrix equalities from the Sylvester equation (5.11):

$$A\tilde{\Pi}_0 + B\ell_0 = -b\tilde{\Pi}_1, \tag{D.9a}$$
$$A\tilde{\Pi}_1 + B\ell_1 = b\tilde{\Pi}_0, \tag{D.9b}$$
$$A\tilde{\Pi}_2 + B\ell_2 = \tilde{\Pi}_0 - b\tilde{\Pi}_3, \tag{D.9c}$$
$$A\tilde{\Pi}_3 + B\ell_3 = \tilde{\Pi}_1 + b\tilde{\Pi}_2. \tag{D.9d}$$

The solution to the set of equations (D.9a) and (D.9b) for $\tilde{\Pi}_0$ and $\tilde{\Pi}_1$ is given in (D.6) and (D.5), respectively. Let us find the solution for $\tilde{\Pi}_2$ and $\tilde{\Pi}_3$. Rewrite (D.9c) and (D.9d) as follows:

$$A\tilde{\Pi}_2 + \overline{B\ell_2} = -b\tilde{\Pi}_3, \tag{D.10a}$$
$$A\tilde{\Pi}_3 + \overline{B\ell_3} = b\tilde{\Pi}_2. \tag{D.10b}$$

with $\overline{B\ell_2} = B\ell_2 - \tilde{\Pi}_0$ and $\overline{B\ell_3} \coloneqq B\ell_3 - \tilde{\Pi}_1$. Notice that (D.10a) and (D.10b) have the same form as the set of equations (D.2a) and (D.2b). Therefore, their solution reads as follows (with a pre-multiplication by C):

$$\begin{aligned} C\tilde{\Pi}_2 &= C(A^2 + b^2 I_n)^{-1}\left(bB\ell_3 - b\tilde{\Pi}_1 - AB\ell_2 + A\tilde{\Pi}_0\right) \\ &= Re(\eta_0(jb, \ell_2 + j\ell_3)) \\ &\quad + C(A^2 + b^2 I_n)^{-1}\left(-b\tilde{\Pi}_1 + A\tilde{\Pi}_0\right), \end{aligned} \tag{D.11a}$$

$$\begin{aligned} C\tilde{\Pi}_3 &= -C(A^2 + b^2 I_n)^{-1}\left(AB\ell_3 - A\tilde{\Pi}_1 + bB\ell_2 - b\tilde{\Pi}_0\right) \\ &= Im(\eta_0(jb, \ell_2 + j\ell_3)) \\ &\quad + C(A^2 + b^2 I_n)^{-1}\left(A\tilde{\Pi}_1 + b\tilde{\Pi}_0\right). \end{aligned} \tag{D.11b}$$

It is left to show that the following two equalities hold:

$$\begin{aligned} C(A^2 + b^2 I_n)^{-1}\left(-b\tilde{\Pi}_1 + A\tilde{\Pi}_0\right) = \\ - \mathrm{Re}(\eta_1(jb, \ell_0 + j\ell_1)), \end{aligned} \tag{D.12a}$$

$$\begin{aligned} C(A^2 + b^2 I_n)^{-1}\left(A\tilde{\Pi}_1 + b\tilde{\Pi}_0\right) = \\ - \mathrm{Im}(\eta_1(jb, \ell_0 + j\ell_1)). \end{aligned} \tag{D.12b}$$

Let us start with working out the left-hand side of (D.12a):

$$
\begin{aligned}
&(A^2+b^2I_n)^{-1}\left(-b\tilde{\Pi}_1+A\tilde{\Pi}_0\right) \qquad \text{(D.13)}\\
&\quad=(A^2+b^2I_n)^{-1}b(A^2+b^2I_n)^{-1}\left(AB\ell_1+bB\ell_0\right)\\
&\quad+(A^2+b^2I_n)^{-1}A(A^2+b^2I_n)^{-1}\left(bB\ell_1-AB\ell_0\right)\\
&\quad=(A^2+b^2I_n)^{-1}(A^2+b^2I_n)^{-1}b\left(AB\ell_1+bB\ell_0\right)\\
&\quad+(A^2+b^2I_n)^{-1}(A^2+b^2I_n)^{-1}A\left(bB\ell_1-AB\ell_0\right)\\
&\quad=(A^2+b^2I_n)^{-2}\left(2bAB\ell_1+b^2B\ell_0-A^2B\ell_0\right),
\end{aligned}
$$

where the fact that any matrix A commutes with $(A^2+b^2I_n)^{-1}$ is used and the existence of $(A^2+b^2I_n)^{-1}$ is guaranteed by the assumption that $s^\star=jb\notin\sigma(A)$. Next, we derive an expression for $\eta_1(jb,\ell_0+j\ell_1)$ as follows:

$$
\begin{aligned}
&\eta_1(jb,\ell_0+j\ell_1)=C(jbI_n-A)^{-2}B(\ell_0+j\ell_1)\\
&=C(jbI_n-A)^{-2}(jbI_n+A)^{-2}(jbI_n+A)^2B(\ell_0+j\ell_1)\\
&=C(-b^2I_n-A^2)^{-2}(jbI_n+A)^2B(\ell_0+j\ell_1)\\
&=C(b^2I_n+A^2)^{-2}(A^2+2jbA-b^2I_n)B(\ell_0+j\ell_1)\\
&=C(A^2+b^2I_n)^{-2}(2jbAB+A^2B-b^2B)(\ell_0+j\ell_1).
\end{aligned}
$$

From here, it can be easily verified that the relation in (D.12a) holds. The proof for (D.12b) can be performed in a similar fashion and is omitted here for the sake of brevity.

The proof for $a\neq 0$ and $k=1$ involves the same steps as given above for the case $a\neq 0$ and $k=0$ and, therefore, omitted. The proof for $k>1$ can be continued similarly as above and is omitted for the sake of brevity. □

E

Appendices to Chapter 6

E.1 Proof of Lemma 6.2

Proof. The following notation for real square matrices A is used: $\mathrm{He}(A) \coloneqq A + A^\top$.

The matrix $\mathcal{F} = S - \mathcal{G}L$ is Hurwitz and the inequality in (6.31) holds if and only if there exists a positive definite matrix Q such that

$$\mathrm{He}\left(Q(S - \mathcal{G}L \pm \gamma^\star \mathcal{G}\mathcal{C}\Pi\right) \prec 0. \tag{E.1}$$

Let us show that (E.1) holds for the specific choice $\mathcal{G}$ in (6.30) and $Q \coloneqq \Pi^\top \bar{P}\Pi$. Note that this choice for matrix Q ensures that it is positive definite, i.e., $Q \succ 0$, since Π is full column rank, see, e.g., [121]. Furthermore, note that $Q\mathcal{G} = \Pi^\top \bar{P}\mathcal{B}$. Substitution of this $\mathcal{G}$ and Q in (E.1) results in:

$$\mathrm{He}\left(\Pi^\top \bar{P}\Pi S - \Pi^\top \bar{P}\mathcal{B}L \pm \gamma^\star \Pi^\top \bar{P}\mathcal{B}\mathcal{C}\Pi\right) \prec 0. \tag{E.2}$$

Note that the Sylvester equation (6.21) can be rewritten as:

$$\mathcal{B}L = \Pi S - \mathcal{A}\Pi, \tag{E.3}$$

which is substituted in (E.2) to yield:

$$\begin{aligned} &\mathrm{He}\left(\Pi^\top \bar{P}\Pi S - \Pi^\top \bar{P}(\Pi S - \mathcal{A}\Pi) \pm \gamma^\star \Pi^\top \bar{P}\mathcal{B}\mathcal{C}\Pi\right) \\ &\quad = \mathrm{He}\left(\Pi^\top \bar{P}\mathcal{A}\Pi \pm \gamma^\star \Pi^\top \bar{P}\mathcal{B}\mathcal{C}\Pi\right) \\ &\quad = \mathrm{He}\left(\Pi^\top \bar{P}\left(\mathcal{A} \pm \gamma^\star \mathcal{B}\mathcal{C}\right)\Pi\right) \prec 0. \end{aligned} \tag{E.4}$$

To show that (E.4) holds, first note that since by assumption $\left\|\mathcal{C}(sI - \mathcal{A})^{-1}\mathcal{B}\right\|_\infty < 1/\gamma^\star$, the following LMIs hold

$$\mathrm{He}\left(\bar{P}(\mathcal{A} \pm \gamma^\star \mathcal{B}\mathcal{C})\right) \prec 0 \tag{E.5}$$

for some positive definite matrix $\bar{P}$. Pre- and post-multiplication of (E.5) with $\Pi^\top$ and Π, respectively, concludes that (E.4) is negative definite (since Π is full column rank). Therefore, we conclude that (6.31) must hold and that the matrix $\mathcal{F}$ is Hurwitz. The reduced-order model achieves moment matching since $\sigma(\mathcal{F}) \cap \sigma(S) = \emptyset$, guaranteed by $\sigma(\mathcal{F}) \in \mathbb{C}^-$ while $\sigma(S) \in \mathbb{C}^0$ by Assumption 6.3. □

E.2 Proof of Theorem 6.4

Proof. We first prove that if the LMIs in (6.32) are satisfied, then all the conditions of Theorem 6.1 are satisfied. After that, we prove that there always exists a feasible solution to the LMIs in (6.32).

Condition (6.9) is satisfied for the same $\gamma^\star$ as in Assumption 6.2 thanks to the nonlinear function φ in the reduced-order model (6.13) being the same nonlinear function as in the full-order model in (6.7). The LMI condition in (6.10) with (A, B_2, C_1) replaced by (F, G_2, H_1) in (6.24), respectively, is equivalent to (6.10) for $P = \text{blockdiag}(P_{(1,1)}, P_{(1,2)}, P_{(2,1)}, P_{(2,2)})$ and the change of variables $X_{(i,k)} = P_{(i,k)} G_{(i,k)}, i, k \in \{1, 2\}$. Thus, satisfaction of the LMIs (6.32) is equivalent to satisfaction of the LMIs (6.10) of Theorem 6.1. Therefore, satisfaction of the LMIs (6.32) guarantees that all the reduced-order Lur'e-type model (6.13) with $G_{(i,k)} = P_{(i,k)}^{-1} X_{(i,k)}, i, k \in \{1, 2\}$, satisfies all the conditions of Theorem 6.1.

Satisfaction of LMIs (6.10) guarantees that the matrix F is Hurwitz, i.e., $\sigma(F) \in \mathbb{C}^-$, which in turn guarantees that $G_{(i,k)} \in \Theta_{(i,k)}$ for $i, k \in \{1, 2\}$, since $\sigma(S_{(i,k)}) \in \mathbb{C}^0$, which completes this part of the proof.

Finally, we prove the feasibility of the LMI condition in (6.32). Thanks to the block-triangular structure of $\mathcal{L}_{\gamma^\star}^{\pm}$, for the first $(i, k) = (1, 1)$, third $(i, k) = (2, 1)$ and fourth $(i, k) = (2, 2)$ block diagonal elements, feasibility of (6.32) is guaranteed if and only if there exists a $G_{(i,k)}$ such that $\sigma(S_{(i,k)} - G_{(i,k)} L_{(i,k)}) \in \mathbb{C}^-$. Such a $G_{(i,k)}$ exists by the observability assumption on the pairs $(S_{(i,k)}, L_{(i,k)}), i, k \in \{1, 2\}$. For the second block diagonal element $(i, k) = (1, 2)$, feasibility of (6.32) is equivalent to the existence of a $G_{(1,2)}$ such that the $\mathcal{H}_\infty$ norm of the corresponding transfer function is bounded by $1/\gamma^\star$. Hereto, we apply the results of Lemma 6.2, which shows that there exists a specific $G_{(1,2)}$ such that the $\mathcal{H}_\infty$ norm of the corresponding transfer function is bounded by $1/\gamma^\star$. Since such a $G_{(1,2)}$ exists, we conclude that the LMIs (6.32) are feasible under the stated assumptions. □

E.3 Proof of Theorem 6.5

Proof. We start by recalling some inequalities before proving the theorem. For a single-input–single-output LTI model characterized by (A, B, C) and excited by a T-periodic input $u \in L_2(T)$, if the matrix A is Hurwitz, there exists a unique globally exponentially stable T-periodic steady-state solution $\bar{x}_u(t)$ with the corresponding steady-state output $\bar{y}_u$ with $\bar{y}_u \in L_2(T)$. Hence, this model defines a

linear operator $g_{yu} : L_2(T) \to L_2(T)$ according to

$$g_{yu}u(t) = \bar{y}_u(t). \tag{E.6}$$

The transfer function of model (A, B, C) from input u to output y reads as $G_{yu}(s) \coloneqq C(sI - A)^{-1}B, \quad s \in \mathbb{C}$. We recall from [196] that

$$\|g_{yu}u\|_{L_2} \leq \sup_{m\in\mathbb{Z}} |G_{yu}(jm\omega)| \|u\|_{L_2} \leq \gamma_{yu}\|u\|_{L_2} \tag{E.7}$$

with $\omega \coloneqq 2\pi/T$, $\gamma_{yu} \coloneqq \sup_{\omega\in\mathbb{R}} |G_{yu}(j\omega)|$. For every transfer function of the LTI part of the full-order and reduced-order Lur'e-type model, a linear operator between inputs u, φ and outputs y, z, ρ and ζ can be defined consistent with g_{yu} in (E.6). Then, for any input $u(t) \in L_2(T)$, the following relation holds:

$$\|(g_{yu} - g_{\rho u})u(t)\|_{L_2} \leq \bar{\Upsilon}\|u(t)\|_{L_2}, \tag{E.8}$$

with $\bar{\Upsilon}$ the constant defined in (6.33). Since $\bar{\Upsilon}$ bounds the mismatch in all $\Upsilon_{(i,k)}, i, k \in \{1, 2\}$, similar relations hold also for the other involved FRFs with the same constant $\bar{\Upsilon}$. Furthermore, by the satisfaction of the conditions in Theorem 6.1 (for both the full-order and reduced-order models), the following bounds hold for signals $\bar{y}(t), \bar{\rho}(t) \in L_2(T)$:

$$\|\varphi(\bar{y}(t)) - \varphi(\bar{\rho}(t))\|_{L_2} \leq \gamma^\star \|\bar{y}(t) - \bar{\rho}(t)\|_{L_2}, \tag{E.9a}$$

$$\|\varphi(\bar{y}(t))\|_{L_2} \leq \gamma^\star \|\bar{y}(t)\|_{L_2}, \tag{E.9b}$$

where $\gamma^\star$ is the constant in Assumption 6.4 (incremental sector condition of the nonlinearity). Moreover, satisfaction of the conditions in Theorem 6.1 guarantees that

$$\gamma^\star\gamma_{y\varphi} < 1 \quad \text{and} \quad \gamma^\star\gamma_{\rho\varphi} < 1 \tag{E.10}$$

with $\gamma_{y\varphi}$ and $\gamma_{y\varphi}$ defined in Theorem 6.5.

It is shown in this proof that the bound (6.34) is a special case of the bound (6.35). Therefore, we prove the latter first and show at the end of the proof that the former is implied. Observe that the steady-state response of the full-order model satisfies

$$\bar{y}(t) = g_{y\varphi}\varphi(\bar{y}(t)) + g_{yu}u(t), \tag{E.11a}$$

$$\bar{z}(t) = g_{z\varphi}\varphi(\bar{y}(t)) + g_{zu}u(t), \tag{E.11b}$$

and that the steady-state response of the reduced-order model satisfies

$$\bar{\rho}(t) = g_{\rho\varphi}\varphi(\bar{\rho}(t)) + g_{\rho u}u(t), \tag{E.12a}$$

$$\bar{\zeta}(t) = g_{\zeta\varphi}\varphi(\bar{\rho}(t)) + g_{\zeta u}u(t). \tag{E.12b}$$

Consider the difference:

$$\bar{z}(t) - \bar{\zeta}(t) = g_{z\varphi}\varphi(\bar{y}(t)) - g_{\zeta\varphi}\varphi(\bar{\rho}(t)) + (g_{zu} - g_{\zeta u})u(t), \tag{E.13a}$$

$$\begin{aligned} &= g_{z\varphi}(\varphi(\bar{y}(t)) - \varphi(\bar{\rho}(t))) + (g_{z\varphi} - g_{\zeta\varphi})\varphi(\bar{\rho}(t)) \\ &\quad + (g_{zu} - g_{\zeta u})u(t). \end{aligned} \tag{E.13b}$$

Taking the L_2-norm and using inequalities (E.8) and (E.9a) results in

$$\left\|\bar{z}(t) - \bar{\zeta}(t)\right\|_{L_2} \leq \gamma_{z\varphi}\gamma^\star \|\bar{y}(t) - \bar{\rho}(t)\|_{L_2} + \bar{\Upsilon}\left(\gamma^\star \|\bar{\rho}(t)\|_{L_2} + \|u\|_{L_2}\right) \tag{E.14}$$

with $\gamma_{z\varphi}$ defined in Theorem 6.5.

Next we upper bound the terms $\|\bar{y}(t) - \bar{\rho}(t)\|_{L_2}$ and $\|\bar{\rho}(t)\|_{L_2}$. Consider the difference

$$\bar{y}(t) - \bar{\rho}(t) = g_{y\varphi}\varphi(\bar{y}(t)) - g_{\rho\varphi}\varphi(\bar{\rho}(t)) + (g_{yu} - g_{\rho u})u(t), \tag{E.15}$$

$$\begin{aligned} &= g_{y\varphi}(\varphi(\bar{y}(t)) - \varphi(\bar{\rho}(t))) + (g_{y\varphi} - g_{\rho\varphi})\varphi(\bar{\rho}(t)) \\ &\quad + (g_{yu} - g_{\rho u})u(t). \end{aligned} \tag{E.16}$$

Again, taking the L_2-norm and using inequalities (E.8) and (E.9a) results in

$$\begin{aligned} \|\bar{y}(t) - \bar{\rho}(t)\|_{L_2} &\leq \gamma_{y\varphi}\gamma^\star \|\bar{y}(t) - \bar{\rho}(t)\|_{L_2} + \bar{\Upsilon}\left(\gamma^\star \|\bar{\rho}(t)\|_{L_2} + \|u(t)\|_{L_2}\right), \\ &\leq \frac{\bar{\Upsilon}}{1 - \gamma_{y\varphi}\gamma^\star}\left(\gamma^\star \|\bar{\rho}(t)\|_{L_2} + \|u(t)\|_{L_2}\right). \end{aligned} \tag{E.17}$$

The latter step is allowed since $\gamma_{y\varphi}\gamma^\star < 1$ by (E.10). Lastly, using again inequalities (E.8) and (E.9a), we find:

$$\|\bar{\rho}(t)\|_{L_2} = \|g_{\rho\varphi}\bar{\varphi}(\rho(t)) + g_{\rho u}u(t)\|_{L_2}, \tag{E.18a}$$

$$\leq \gamma_{\rho\varphi}\gamma^\star \|\bar{\rho}(t)\|_{L_2} + \gamma_{\rho u}\|u(t)\|_{L_2}, \tag{E.18b}$$

$$\leq \frac{\gamma_{\rho u}}{1 - \gamma_{\rho\varphi}\gamma^\star}\|u(t)\|_{L_2}, \tag{E.18c}$$

where $\gamma_{\rho u}$ is defined in Theorem 6.5. The latter step is again allowed since $\gamma_{\rho\varphi}\gamma^\star < 1$ by (E.10).

Substitution of (E.18c) into (E.17) and (E.14) and collecting terms yields the bound presented in (6.35), which completes the proof of the bound (6.35).

Finally, we show that the bound (6.34) is a special case of the bound (6.35). Observe that the moment $C_2\pi(\tau)$ coincides with the steady-state response $\bar{z}_u$ of the full-order model and that the moment $H_2p(\tau)$ coincides with the steady-state response $\bar{\zeta}_u$ of the reduced-order model. Therefore, we can replace the left-hand side of (6.35) with the left-hand side of (6.34). Furthermore, substituting $L\tau$ (output of the signal generator) for the input u in (6.35) results in the bound (6.34), which completes the proof. □

F

Appendices to Chapter 7

F.1 Proof of Theorem 7.1

Proof. The Sylvester equation $FP + GL = PS$ has a unique solution $P \in \mathbb{R}^{\nu\times\nu}$ if $\sigma(F) \cap \sigma(S) = \emptyset$. Selecting $P = I_\nu$ results in $F = S - GL$ for any $G \in \mathcal{G}_M$ with $\mathcal{G}_M$ as in (7.10). Moreover, selecting $H = C\Pi$ results in the match $HP = C\Pi$.

It is only left to show that the match $HP = C\Pi$ enforces a match as described by (7.13). The matrix $\Pi \in \mathbb{R}^{n\times\nu}$ is partitioned as follows:

$$\Pi = \begin{bmatrix}\Pi_1 & \Pi_2 & \dots, \Pi_{\nu-1}, \Pi_\nu\end{bmatrix}, \Pi_i \in \mathbb{R}^n, i = 1, \dots, \nu.$$

Using this partitioning, the Sylvester equation (7.12) can be written as the following ν equations:

$$\begin{aligned}
-\omega_1\Pi_2 &= A\Pi_1 + B\ell_1, && \text{(F.1a)}\\
\omega_1\Pi_1 &= A\Pi_2, && \text{(F.1b)}\\
&\vdots && \text{(F.1c)}\\
-\omega_\kappa\Pi_\nu &= A\Pi_{\nu-1} + B\ell_\kappa, && \text{(F.1d)}\\
\omega_\kappa\Pi_{\nu-1} &= A\Pi_\nu. && \text{(F.1e)}
\end{aligned}$$

The solution $\Pi \in \mathbb{R}^{n\times\nu}$ to (F.1) reads as follows:

$$\begin{aligned}
\Pi_1 &= -(A^2 + I_n\omega_1^2)^{-1}AB\ell_1, && \text{(F.2a)}\\
\Pi_2 &= -(A^2 + I_n\omega_1^2)^{-1}\omega_1 B\ell_1, && \text{(F.2b)}\\
&\vdots && \text{(F.2c)}\\
\Pi_{\nu-1} &= -(A^2 + I_n\omega_\kappa^2)^{-1}AB\ell_\kappa, && \text{(F.2d)}\\
\Pi_\nu &= -(A^2 + I_n\omega_\kappa^2)^{-1}\omega_\kappa B\ell_\kappa. && \text{(F.2e)}
\end{aligned}$$

From (F.2), it is trivial to show that

$$C\Pi_1 = \mathrm{Re}(\eta_0^{\Sigma}(j\omega_1, \ell_1)), \tag{F.3a}$$
$$C\Pi_2 = \mathrm{Im}(\eta_0^{\Sigma}(j\omega_1, \ell_1)), \tag{F.3b}$$
$$\vdots \tag{F.3c}$$
$$C\Pi_{\nu-1} = \mathrm{Re}(\eta_0^{\Sigma}(j\omega_\kappa, \ell_\kappa)), \tag{F.3d}$$
$$C\Pi_{\nu} = \mathrm{Im}(\eta_0^{\Sigma}(j\omega_\kappa, \ell_\kappa)). \tag{F.3e}$$

Repeating these derivations for $P \in \mathbb{R}^{\nu\times\nu}$ for the reduced-order model (7.5) results in

$$HP_1 = \mathrm{Re}(\eta_0^{\hat{\Sigma}}(j\omega_1, \ell_1)), \tag{F.4a}$$
$$HP_2 = \mathrm{Im}(\eta_0^{\hat{\Sigma}}(j\omega_1, \ell_1)), \tag{F.4b}$$
$$\vdots \tag{F.4c}$$
$$HP_{\nu-1} = \mathrm{Re}(\eta_0^{\hat{\Sigma}}(j\omega_\kappa, \ell_\kappa)), \tag{F.4d}$$
$$HP_{\nu} = \mathrm{Im}(\eta_0^{\hat{\Sigma}}(j\omega_\kappa, \ell_\kappa)). \tag{F.4e}$$

Thus, $C\Pi = HP$ guarantees moment matching according to

$$\eta_0^{\Sigma}(j\omega_i, \ell_i) = \eta_0^{\hat{\Sigma}}(j\omega_i, \ell_i), \quad \text{for } i = 1, \ldots, \kappa.$$

The transfer function matrices being real and rational ensures that moment matching according to $\eta_0^{\Sigma}(-j\omega_i, \ell_i) = \eta_0^{\hat{\Sigma}}(-j\omega_i, \ell_i), i = 1, \ldots, \kappa$ is also achieved, which completes the proof. □

F.2 Proof of Theorem 7.2

Proof. The transfer function matrix $W(s)$ can be written as $W(s) = \frac{N(s)}{D(s)}$, where $D(s) \in \mathbb{C}$ is a scalar polynomial in s of degree ν and $N(s)$ is a matrix of polynomials not exceeding degree $\nu - 1$ thanks to $W(s)$ being strictly proper. The transfer function matrix of reduced-order models characterized by $G \in \mathcal{G}_M$ can be written as follows:

$$\Gamma(s) = \frac{C\Pi\mathrm{adj}(sI_\nu - (S - GL))G}{\det(sI_\nu - (S - GL))}. \tag{F.5}$$

Define the column $\bar{N}(s) \coloneqq N(s)\ell$ and:

$$\bar{\Gamma}(s) \coloneqq C\Pi\mathrm{adj}(sI_\nu - (S - GL))G\ell. \tag{F.6}$$

It is sufficient to show that there exists a $G \in \mathcal{G}_M$ such that $\det(sI_\nu - (S - GL)) = D(s)$ and $\bar{\Gamma}(s) = \bar{N}(s)$. By observability of the pair (S, L), there exists a $G \in \mathbb{R}^{\nu\times m}$

such that $\det(sI_\nu - (S - GL)) = D(s)$. In the multi-input case ($m > 1$), such G is possibly non-unique. To show that $\bar{\Gamma}(s) = \bar{N}(s)$ for any such G, consider the i-th entry of $\bar{N}(s)$ and the i-th entry of $\bar{\Gamma}(s)$, where $1 \leq i \leq p$ with p the number of outputs. By the restriction that all tangential directions coincide, i.e., $\ell_i = \ell$ for all $i \in \{1, \ldots, \kappa\}$, the scalar polynomials in the i-th entry of $\bar{N}(s)$ and $\bar{\Gamma}(s)$ are constrained at the same ν locations in the complex plane via conditions (7.14) and (7.13), respectively. Since these polynomials are both at most of the degree $\nu - 1$, each of these can only interpolate ν locations. Hence, these polynomials must coincide. This rationale can be repeated to conclude that all entries of $\bar{N}(s)$ and $\bar{\Gamma}(s)$ coincide, which completes the proof. □

F.3 Proof of Theorem 7.3

Proof. Consider the parametrization (7.11) for the reduced-order model (7.5) and the error model (7.16). Direct application of Lemma 7.1 for the given $\gamma > 0$ reveals that $\left\|\mathcal{C}(sI - \mathcal{A})^{-1}\mathcal{B}\right\|_\infty < \gamma$ and $\sigma(\mathcal{A}) \in \mathbb{C}^-$ hold true if and only if the following matrix inequalities hold:

$$\mathcal{N}_\gamma(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{X}) \prec 0, \quad \mathcal{X} \succ 0. \tag{F.7}$$

Note that, thanks to the block-diagonal structure of $\mathcal{A}$ and $\sigma(A) \in \mathbb{C}^-$ by Assumption 7.1, $\sigma(F) = \sigma(S - GL) \in \mathbb{C}^-$ if and only if $\sigma(\mathcal{A}) \in \mathbb{C}^-$. Furthermore, the transfer function $\Upsilon(s)$ satisfies $\Upsilon(s) = \mathcal{C}(sI - \mathcal{A})^{-1}\mathcal{B}, s \in \mathbb{C}$. Therefore, constraints (7.8b) is satisfied for a fixed γ if and only if the inequalities (F.7) are satisfied. Inequalities (F.7) are satisfied for any $G \in \mathcal{G}_\gamma$, guaranteeing that constraints (7.8b) is satisfied. Moreover, the set $\mathcal{G}_\gamma$ is a subset of $\mathcal{G}_M$, since for any $G \in \mathcal{G}_\gamma$, it trivially holds that $\sigma(F) \cap \sigma(S) = \emptyset$, since $\sigma(F) \in \mathbb{C}^-$ and $\sigma(S) \in \mathbb{C}^0$.

Application of [72, Corollary 1] shows that the mismatch $\|\Upsilon(s)\|_\infty$ is lower bounded by $\bar{h}_{\nu+1}$ if Φ and Γ are both stable transfer functions and if the state-space realization of Φ is balanced. Therefore, the set $\mathcal{G}_\gamma$ is empty for $\gamma < \bar{h}_{\nu+1}$, completing the proof. □

F.4 Proof of Theorem 7.6

Proof. Along the same lines of the proof of Theorem 7.3, satisfaction of constraints (7.39b) and (7.39d), and the emptiness of $\tilde{\mathcal{G}}_\gamma$ for $\gamma < h_{\nu+1}$ can be proven. It is only left to show that constraint (7.39c) is satisfied for any $G \in \tilde{\mathcal{G}}_\gamma$. The constraints

$$\mathcal{X}_2 \in \mathbb{S}_{n+\nu}, \mathcal{N}_{\bar{\gamma}_{zw}}(S - GL, G_\lambda, H_\zeta, \mathcal{X}_2) \prec 0, \tag{F.8}$$

ensure via Lemma 7.1 that $\left\|\Gamma_{(\zeta,\lambda)}(s)\right\|_\infty \leq \bar{\gamma}_{zw}$. Finally, Assumption 7.3 guarantees satisfaction of constraint (7.39c). □

F.5 Matrices related to the reduction problem

The constraints in the set $\mathcal{G}_\gamma$ in (7.19) can be written in the standard form (7.46) with

$$\begin{aligned}\mathcal{M}_1(\mathcal{X},\gamma) &\coloneqq \begin{bmatrix} -\mathcal{X} & 0_{n+\nu\times n+\nu+p+m} \\ \star & \mathcal{N}_\gamma\left(\begin{bmatrix} A & 0_{n\times\nu} \\ \star & S \end{bmatrix}, \begin{bmatrix} B \\ 0_{\nu\times m}\end{bmatrix}, \mathcal{C}, \mathcal{X}\right)\end{bmatrix}, \\ \mathcal{M}_2(\mathcal{X}) &\coloneqq \begin{bmatrix} 0_{n+\nu\times n+\nu} \\ \mathcal{X} \\ 0_{m\times n+\nu} \\ 0_{p\times n+\nu}\end{bmatrix}, \mathcal{N}(G) \coloneqq \begin{bmatrix} 0_{n+\nu\times n+\nu} \\ \begin{bmatrix} 0_{n\times n} & 0_{n\times\nu} \\ \star & (-GL)^\top\end{bmatrix} \\ \begin{bmatrix} 0_{m\times n} & G^\top\end{bmatrix} \\ 0_{p\times n+\nu}\end{bmatrix}.\end{aligned} \tag{F.9}$$

Their dimensions read as follows: $\mathcal{M}_1 \in \mathbb{R}^{\alpha\times\alpha}, \mathcal{M}_2 \in \mathbb{R}^{\alpha\times\beta}, \mathcal{N} \in \mathbb{R}^{\alpha\times\beta}$, where $\alpha \coloneqq 2(n+\nu)+m+p$ and $\beta \coloneqq n+\nu$.

The constraints in the set $\tilde{\mathcal{G}}_\gamma$ in (7.42) can be written in the standard form (7.46) with

$$\begin{aligned}&\mathcal{M}_1(\mathcal{X}_1,\mathcal{X}_2,\gamma) \coloneqq \\ &\begin{bmatrix} -\mathcal{X}_1 & 0_{n+\nu\times\nu} & 0_{n+\nu\times n+\nu+m+s+p+q} & 0_{n+\nu\times\nu+s+q} \\ \star & -\mathcal{X}_2 & 0_{\nu\times n+\nu+m+s+p+q} & 0_{\nu\times\nu+s+q} \\ \star & \star & \mathcal{N}_\gamma\left(\begin{bmatrix} A & 0_{n\times\nu} \\ \star & S\end{bmatrix}, \begin{bmatrix} B \\ 0_{\nu\times m+s}\end{bmatrix}, \mathcal{C}, \mathcal{X}_1\right) & 0_{n+\nu+m+s+p+q\times\nu+s+q} \\ \star & \star & \star & \mathcal{N}_{\gamma_{zw}}(S, 0_{\nu\times s}, H_\zeta, \mathcal{X}_2)\end{bmatrix} \\ &\mathcal{M}_2(\mathcal{X}_1,\mathcal{X}_2) \coloneqq \begin{bmatrix} 0_{n+\nu\times n+\nu} & 0_{n+\nu\times\nu} \\ 0_{\nu\times n+\nu} & 0_{\nu\times\nu} \\ \mathcal{X}_1 & 0_{n+\nu\times\nu} \\ 0_{m+s\times n+\nu} & 0_{m+s\times\nu} \\ 0_{p+q\times n+\nu} & 0_{p+q\times\nu} \\ 0_{\nu\times n+\nu} & \mathcal{X}_2 \\ 0_{s\times n+\nu} & 0_{s\times\nu} \\ 0_{q\times n+\nu} & 0_{q\times\nu}\end{bmatrix}, \\ &\mathcal{N}(G) \coloneqq \begin{bmatrix} 0_{n+\nu\times n+\nu} & 0_{n+\nu\times\nu} \\ 0_{\nu\times n+\nu} & 0_{\nu\times\nu} \\ \begin{bmatrix} 0_{n\times n} & 0_{n\times\nu} \\ \star & (-GL)^\top\end{bmatrix} & 0_{n+\nu\times\nu} \\ \begin{bmatrix} 0_{m+s\times n} & G^\top\end{bmatrix} & 0_{m+s\times\nu} \\ 0_{p+q\times n+\nu} & 0_{p+q\times\nu} \\ 0_{\nu\times n+\nu} & (-GL)^\top \\ 0_{s\times n+\nu} & G_\lambda^\top \\ 0_{q\times n+\nu} & 0_{q\times\nu}\end{bmatrix}.\end{aligned} \tag{F.10}$$

Their dimensions read as follows: $\mathcal{M}_1 \in \mathbb{R}^{\alpha\times\alpha}, \mathcal{M}_2 \in \mathbb{R}^{\alpha\times\beta}, \mathcal{N} \in \mathbb{R}^{\alpha\times\beta}$, where $\alpha \coloneqq 2(n+\nu)+2\nu+2(s+1)+m+p$ and $\beta \coloneqq n+2\nu$.

F.6 Proof of Theorem 7.7

Proof. The matrix inequalities in Lemma 7.1 for the transfer function $\Gamma_{(\zeta,\lambda)}$ are, after the application of the change of variables $\begin{bmatrix} Q_1 & Q_2 \end{bmatrix} = X \begin{bmatrix} G_u & G_\lambda \end{bmatrix}$, equivalent to the inequalities (7.49) and $\mathcal{X} \in \mathbb{S}_\nu$. Therefore, if (7.49) and $\mathcal{X} \in \mathbb{S}_\nu$ are satisfied, then $\left\|\Gamma_{(\zeta,\lambda)}(s)\right\|_\infty \leq \bar{\gamma}_{zw}$ and $(S-GL)$ is Hurwitz. The latter, together with Assumption 7.2 (guaranteeing that A is Hurwitz), implies that the linear error dynamics (7.41) are asymptotically stable, guaranteeing the existence of a finite $\gamma^{[0]}$ such that $G \in \tilde{\mathcal{G}}_\gamma^{[0]}$. Furthermore, by Assumption 7.3, constraint (7.39c) is satisfied. □

G Appendices to Chapter 9

G.1 Proof of Theorem 9.1

Proof. Convergence is implied by model (9.5) being incrementally stable on the set $\mathcal{X}$ for the class of inputs $\mathcal{U}_c$ and by the set $\mathcal{X}$ being positively invariant for the class of inputs $\mathcal{U}_c$ with respect to the model dynamics (9.5), see [282, Theorem 13].

Incremental stability: For brevity of the expressions, the following notation is used throughout this proof:

$$\delta x(x^a, x^b) \coloneqq x^a - x^b, \quad \delta f(x^a, x^b, u) \coloneqq f(x^a, u) - f(x^b, u).$$

Consider the incremental Lyapunov function $V(x^a, x^b) \coloneqq |E\delta x(x^a, x^b)|^2_{P^{-1}}$ for an invertible matrix $E \in \mathbb{R}^{n\times n}$ and a positive definite matrix $P \in \mathbb{R}^{n\times n}$. The time increment ΔV of V is given by:

$$\begin{aligned}\Delta V(x^a, x^b, u) &\coloneqq V(f(x^a, u), f(x^b, u)) - V(x^a, x^b)\\ &= |\delta f(x^a, x^b, u)|^2_{P^{-1}} - |E\delta x(x^a, x^b)|^2_{P^{-1}}.\end{aligned}$$

Incremental stability on $\mathcal{X}$ for $\mathcal{U}_c$ is guaranteed by the satisfaction of the following condition:

$$\Delta V(x^a, x^b, u) < 0, \quad \forall\, x^a, x^b \in \mathcal{X}, x^a \neq x^b, u \in U_c. \tag{G.1}$$

Define the function $\phi(\lambda)$ as follows:

$$\begin{aligned}\phi(\lambda) \coloneqq\ & \delta f(x^a, x^b, u)^\top P^{-1} f(x^b + \lambda(x^a - x^b), u)\\ & - \delta x(x^a, x^b)^\top E^\top P^{-1} E(x^b + \lambda(x^a - x^b)).\end{aligned}$$

and notice that $\phi(1) - \phi(0) = \Delta V(x^a, x^b, u)$. By application of the mean-value theorem, the following relation holds for some $\tilde{\lambda} \in [0, 1]$:

$$\begin{aligned}\Delta V(x^a, x^b, u) &= \frac{\partial \phi}{\partial \lambda}(\tilde{\lambda}) \\ &= \delta f(x^a, x^b, u)^\top P^{-1} A(x_{\tilde{\lambda}}, u)\delta x(x^a, x^b) - |E\delta x(x^a, x^b)|^2_{P^{-1}}\end{aligned} \tag{G.2}$$

with $x_{\tilde{\lambda}} = x^b + \tilde{\lambda}(x^a - x^b)$ also being in $\mathcal{X}$ by convexity of $\mathcal{X}$. Using the same reasoning as in the proof of [282, Theorem 14], it can be concluded that

$$\begin{aligned}\delta f(x^a, x^b, u)^\top P^{-1} A(x_{\tilde{\lambda}}, u)\delta x(x^a, x^b) \\ \leq \delta x(x^a, x^b)^\top A(x_{\tilde{\lambda}}, u)^\top P^{-1} A(x_{\tilde{\lambda}}, u)\delta x(x^a, x^b).\end{aligned} \tag{G.3}$$

Using (G.3) on (G.2) results in

$$\Delta V(x^a, x^b, u) \leq |A(x_{\tilde{\lambda}}, u)\delta x(x^a, x^b)|^2_{P^{-1}} - |E\delta x(x^a, x^b)|^2_{P^{-1}}. \tag{G.4}$$

For positive definite matrices P, the following matrix inequality holds:

$$-E^\top P^{-1} E \preceq P - E^\top - E, \tag{G.5}$$

which can be used to upper bound $\Delta V(x^a, x^b, u)$ in (G.4) as follows:

$$\Delta V(x^a, x^b, u) \leq |A(x_{\tilde{\lambda}}, u)\delta x(x^a, x^b)|^2_{P^{-1}} + |\delta x(x^a, x^b)|^2_{P-E-E^\top}. \tag{G.6}$$

Condition (G.1) is then implied if the following matrix inequality holds for all $(x_{\tilde{\lambda}}, u) \in (\mathcal{X}, U_c)$:

$$A^\top(x_{\tilde{\lambda}}, u)P^{-1}A(x_{\tilde{\lambda}}, u) + P - E - E^\top \prec 0. \tag{G.7}$$

Since (G.7) needs to hold for all $(x_{\tilde{\lambda}}, u) \in (\mathcal{X}, U_c)$, it is equivalent to check

$$A^\top(x, u)P^{-1}A(x, u) + P - E - E^\top \prec 0 \tag{G.8}$$

for all $(x, u) \in (\mathcal{X}, U_c)$. Application of the Schur Complement Lemma (see, e.g., [42]) to (G.8) yields the matrix inequality (9.13b), which should hold for all $(x, u) \in (\mathcal{X}, U_c)$. Thus, if condition (9.13b) holds for all $(x, u) \in (\mathcal{X}, U_c)$, then satisfaction of condition (G.1) is implied. Furthermore, satisfaction of condition (9.13b) guarantees positive definitness of the matrix $E + E^\top$, hence E is invertible. This validates the choice for the incremental Lyapunov function V, proving incremental stability of model (9.5) on $(\mathcal{X}, U_c)$.

To show that the bound (9.14) holds, define $Q \coloneqq E^\top P^{-1} E$ and note that (G.1) is equivalent to:

$$|\delta x(x^a_{k+1}, x^b_{k+1})|^2_Q < |\delta x(x^a_k, x^b_k)|^2_Q. \tag{G.9}$$

Because (G.9) is a strict inequality, there exists an $\varepsilon > 0$ such that

$$|\delta x(x^a_{k+1}, x^b_{k+1})|^2_Q \leq \rho |\delta x(x^a_k, x^b_k)|^2_Q \tag{G.10}$$

with $\rho \coloneqq 1 - \varepsilon$. By recursive evaluation and using the min-max theorem [151], we find

$$|\delta x(x^a_k, x^b_k)|^2 \leq \rho^k \frac{\sigma_{\max}(Q)}{\sigma_{\min}(Q)} |\delta x(x^a_0, x^b_0)|^2, \tag{G.11}$$

where $\sigma_{\max}(Q)$ and $\sigma_{\min}(Q)$ denote the maximum and minimum eigenvalues of the positive definite matrix Q, respectively. The bound (G.11) is equivalent to (9.14) for the choice $\tau \coloneqq \sigma_{\max}(Q)/\sigma_{\min}(Q)$, which proves incremental stability.

Positively invariant set: By definition, a set $\mathcal{X}$ is positively invariant with respect to dynamics (9.5) if (and only if) any $x \in \mathcal{X}$ implies that $E^{-1}\hat{f}(x,u) \in \mathcal{X}$ for any $u \in U_c$. Given the definition of the set $\mathcal{X}$ in (9.10), the set $\mathcal{X}$ is positively invariant under any input $u \in \mathcal{U}_c$ if and only if the following condition holds:

$$(E^{-1}\hat{f}(x,u))^\top X E^{-1}\hat{f}(x,u) \leq 1, \tag{G.12}$$

for all $(x,u) \in (\mathcal{X}, U_c)$. By the Schur Complement Lemma, this inequality can be written as follows:

$$\begin{bmatrix} E^\top X^{-1} E & \hat{f}(x,u) \\ \hat{f}^\top(x,u) & 1 \end{bmatrix} \succeq 0 \quad \forall\, (x,u) \in (\mathcal{X}, U_c). \tag{G.13}$$

The term $E^\top X^{-1} E$ is an upper bound for $-X+E^\top+E$, yielding condition (9.13c). Thus, if (9.13c) holds, then satisfaction of (G.12) is implied, guaranteeing that $\mathcal{X}$ is a positively invariant set for the class of inputs $\mathcal{U}_c$.

Using [198, Lemma 2], it be concluded, on the basis of $\mathcal{X}$ being positively invariant, that there exists a solution $\bar{x}$ that is bounded for all $k \in \mathbb{Z}$. This fact, together with the fact that the system is proven to be exponentially incrementally stable implies that (9.4) holds. Now, it can be concluded that the model is exponentially convergent on the set $\mathcal{X}$ for inputs $\mathcal{U}_c$ and that the exponentially stable steady-state solution $\bar{x} \in \mathcal{X}$ for all $k \in \mathbb{Z}$. Hence, the proof is completed. □

Bibliography

[1] M. H. Abbasi, L. Iapichino, W. Schilders, and N. van de Wouw. A data-based stability-preserving model order reduction method for hyperbolic partial differential equations. *Nonlinear Dynamics*, 107(4):3729–3748, 2022.

[2] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016.

[3] J. Adlers and G. Pihl. *Prediction of training time for deep neural networks in TensorFlow*. PhD thesis, KTH, 2018.

[4] F. Allgöwer and A. Zheng. *Nonlinear model predictive control*, volume 26. Birkhäuser, Basel, Switzerland, 2012.

[5] M. B. Amato, M. O. Meade, A. S. Slutsky, L. Brochard, E. L. Costa, et al. Driving pressure and survival in the acute respiratory distress syndrome. *New England Journal of Medicine*, 372(8):747–755, 2015. PMID: 25693014.

[6] B. D. O. Anderson and J. B. Moore. *Optimal Control: Linear Quadratic Methods*. Prentice-Hall International, Inc., Englewood Cliffs, NJ, USA, 1990.

[7] J. Anderson. Geometrical approach to reduction of dynamical systems. In *Proceedings of the Institution of Electrical Engineers*, volume 114, pages 1014–1018, 1967.

[8] D. Angeli. A Lyapunov approach to incremental stability properties. *IEEE Transactions on Automatic Control*, 47(3):410–421, 2002.

[9] B. Anić, C. Beattie, S. Gugercin, and A. C. Antoulas. Interpolatory weighted-$\mathcal{H}_2$ model reduction. *Automatica*, 49(5):1275–1280, 2013.

[10] A. Antoulas. *Approximation of large-scale dynamical systems*. SIAM, Philadelphia, USA, 2005.

[11] A. Antoulas. An overview of approximation methods for large-scale dynamical systems. *Annual reviews in Control*, 29(2):181–190, 2005.

[12] P. Apkarian and D. Noll. Mixed L_1/H_∞-synthesis for L_∞-stability. *International Journal of Robust and Nonlinear Control*, 4(32):2119–2142, 2021.

[13] M. ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019.

[14] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

[15] U. M. Ascher, R. M. M. Mattheij, and R. D. Russell. *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*. Society for Industrial and Applied Mathematics, Philadelphia, USA, 1995.

[16] A. Astolfi. Model reduction by moment matching for linear and nonlinear systems. *IEEE Transactions on Automatic Control*, 55(10):2321–2336, 2010.

[17] A. Astolfi. A new look at model reduction by moment matching for linear systems. In *Proceedings of the 46th Conference on Decision and Control*, pages 4361–4366, 2007.

[18] K. J. Åström. *Introduction to stochastic control theory*. Academic Press, New York, 1970.

[19] N. N. Balaji, S. Lian, M. Scheel, M. R. Brake, P. Tiso, et al. Numerical assessment of polynomial nonlinear state-space and nonlinear-mode models for near-resonant vibrations. *Vibration*, 3(3):22, 2020.

[20] J. A. Ball, I. Gohberg, and L. Rodman. *Interpolation of rational matrix functions*, volume 45 of *Operator Theory: Advances and Applications*. Birkhäuser, Basel, 1990.

[21] L. Baratchart, M. Cardelli, and M. Olivi. Identification and rational l^2 approximation a gradient algorithm. *Automatica*, 27(2):413–417, 1991.

[22] C. I. Bargmann and W. T. Newsome. The brain research through advancing innovative neurotechnologies (brain) initiative and neurology. *JAMA neurology*, 71(6):675–676, 2014.

[23] A. R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information theory*, 39(3):930–945, 1993.

[24] K. Batselier, C.-Y. Ko, and N. Wong. Tensor network subspace identification of polynomial state space models. *Automatica*, 95:187–196, 2018.

[25] R. Beals. *Analysis: an introduction*. Cambridge University Press, Cambridge, 2004.

[26] T. Beckers and S. Hirche. Stability of Gaussian process state space models. In *Proceedings of the European Control Conference*, pages 2275–2281, 2016.

[27] G. Beintema, R. Tóth, and M. Schoukens. Nonlinear state-space identification using deep encoder networks. In *Proceedings of the Machine Learning Research (3rd Annual Learning for Dynamics & Control Conference)*, pages 1–10, 2021.

[28] P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531, 2015.

[29] K. Berntorp. Online Bayesian inference and learning of Gaussian-process state–space models. *Automatica*, 129:109613, 2021.

[30] G. Besançon. *Nonlinear observers and applications*. Springer Verlag, New York, 2007.

[31] B. Besselink, N. van de Wouw, and H. Nijmeijer. Model reduction for a class of convergent nonlinear systems. *IEEE Transactions on Automatic Control*, 57(4):1071–1076, 2011.

[32] B. Besselink, N. van de Wouw, and H. Nijmeijer. Model reduction for nonlinear systems with incremental gain or passivity properties. *Automatica*, 49(4):861–872, 2013.

[33] B. Besselink, N. van de Wouw, J. M. A. Scherpen, and H. Nijmeijer. Model reduction for nonlinear systems by incremental balanced truncation. *IEEE Transactions on Automatic Control*, 59(10):2739–2753, 2014.

[34] S. Billings. *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons, Chichester, West Sussex, United Kingdom, 2013.

[35] L. Blackmore. Autonomous precision landing of space rockets. In *Proceedings of the Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2016 Symposium*, volume 46, pages 15–20. The Bridge Washington, DC, 2016.

[36] C. Blocher, M. Saveriano, and D. Lee. Learning stable dynamical systems using contraction theory. In *Proceedings of the 14th International Conference on Ubiquitous Robots and Ambient Intelligence*, pages 124–129, 2017.

[37] F. Bonassi, E. Terzi, M. Farina, and R. Scattolini. Lstm neural networks: Input to state stability and probabilistic safety verification. In *Proceedings of the Conference on Learning for Dynamics and Control*, pages 85–94, 2020.

[38] B. N. Bond, Z. Mahmood, Y. Li, R. Sredojevic, A. Megretski, et al. Compact modeling of nonlinear analog circuits using system identification via semidefinite programming and incremental stability certification. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 29(8):1149–1162, Aug 2010.

[39] M. Borrello. Modeling and control of systems for critical care ventilation. In *Proceedings of the American Control Conference*, pages 2166–2180, 2005.

[40] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day, 1970.

[41] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994.

[42] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[43] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

[44] A. Bryson Jr and A. Carrier. Second-order algorithm for optimal model order reduction. *Journal of Guidance, Control, and Dynamics*, 13(5):887–892, 1990.

[45] R. H. Byrd, J. C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical programming*, 89(1):149–185, 2000.

[46] T. M. Cameron and J. H. Griffin. An Alternating Frequency/Time Domain Method for Calculating the Steady-State Response of Nonlinear Dynamic Systems. *Journal of Applied Mechanics*, 56(1):149–154, 03 1989.

[47] R. Cao, Y. Lu, and Z. He. System identification method based on interpretable machine learning for unknown aircraft dynamics. *Aerospace Science and Technology*, 126:107593, 2022.

[48] A. Cardona, A. Lerusse, and M. Géradin. Fast Fourier nonlinear vibration analysis. *Computational Mechanics*, 22(2):128–142, 1998.

[49] A. Castagnotto and B. Lohmann. A new framework for $\mathcal{H}_2$-optimal model reduction. *Mathematical and Computer Modelling of Dynamical Systems*, 24(3):236–257, 2018.

[50] F. Cellier and E. Kofman. *Continuous system simulation*. Springer Science & Business Media, New York, 2006.

[51] T. Chen and B. Francis. *Optimal Sampled-Data Control Systems*. Springer-Verlag London, 1995.

[52] C. L. Cheng and C. Rezac. The role of robotics in colorectal surgery. *Bmj*, 360, 2018.

[53] M. Cheng, J. Yi, P.-Y. Chen, H. Zhang, and C.-J. Hsieh. Seq2sick: Evaluating the robustness of sequence-to-sequence models with adversarial examples. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3601–3608, 2020.

[54] A. Chiuso and G. Pillonetto. System identification: A machine learning perspective. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:281–304, 2019.

[55] Y. Choi and K. Carlberg. Space–time least-squares Petrov–Galerkin projection for nonlinear model reduction. *SIAM Journal on Scientific Computing*, 41(1):A26–A58, 2019.

[56] T. F. Coleman and Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization*, 6(2):418–445, 1996.

[57] S. B. Cooper, K. Tiels, B. Titurus, and D. Di Maio. Polynomial nonlinear state space identification of an aero-engine structure. *Computers & Structures*, 238:106299, 2020.

[58] G. De Nicolao and G. F. Trecate. Consistent identification of NARX models via regularization networks. *IEEE Transactions on Automatic Control*, 44(11):2045–2049, 1999.

[59] E. de Souza and S. P. Bhattacharyya. Controllability, observability and the solution of $AX - XB = C$. *Linear Algebra and its Applications*, 39:167–188, 1981.

[60] J. Decuyper, T. De Troyer, M. Runacres, K. Tiels, and J. Schoukens. Nonlinear state-space modelling of the kinematics of an oscillating circular cylinder in a fluid flow. *Mechanical Systems and Signal Processing*, 98:209–230, 2018.

[61] J. Decuyper, K. Tiels, M. Runacres, and J. Schoukens. Retrieving highly structured models starting from black-box nonlinear state-space models using polynomial decoupling. *Mechanical Systems and Signal Processing*, 146:106966, 2021.

[62] J. Decuyper, P. Dreesen, J. Schoukens, M. C. Runacres, and K. Tiels. Decoupling multivariate polynomials for nonlinear state-space models. *IEEE Control Systems Letters*, 3(3):745–750, 2019.

[63] J. Decuyper, J.-P. Noël, T. De Troyer, M. Runacres, and J. Schoukens. Polynomial nonlinear state-space modelling of vortex-induced vibrations: black-box vs grey-

box approach. In *Proceedings of the International Forum on Aeroelasticity and Structural Dynamics*, 2017.

[64] J. Decuyper, M. C. Runacres, J. Schoukens, and K. Tiels. Tuning nonlinear state-space models using unconstrained multiple shooting. *IFAC-PapersOnLine*, 53(2):334–340, 2020.

[65] J. Decuyper, K. Tiels, S. Weiland, M. C. Runacres, and J. Schoukens. Decoupling multivariate functions using a nonparametric filtered tensor decomposition. *Mechanical Systems and Signal Processing*, 179:109328, 2022.

[66] B. P. Demidovich. Lectures on Stability Theory. Technical report, Nauka, Moscow, 1967.

[67] U. Desai and D. Pal. A transformation approach to stochastic model reduction. *IEEE Transactions on Automatic Control*, 29(12):1097–1100, 1984.

[68] F. J. Doyle, R. K. Pearson, and B. A. Ogunnaike. *Identification and control using Volterra models*. Springer London, 2002.

[69] P. Dreesen, M. Ishteva, and J. Schoukens. Decoupling multivariate polynomials using first-order information and tensor decompositions. *SIAM Journal on Matrix analysis and Applications*, 36(2):864–879, 2015.

[70] J. Duan, Y. Ou, J. Hu, Z. Wang, S. Jin, and C. Xu. Fast and stable learning of dynamical systems based on extreme learning machine. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(6):1175–1185, 2017.

[71] S. Dura-Bernal, B. A. Suter, P. Gleeson, M. Cantarelli, A. Quintana, et al. Netpyne, a tool for data-driven multiscale modeling of brain circuits. *Elife*, 8:e44494, 2019.

[72] Y. Ebihara and T. Hagiwara. On $\mathcal{H}_\infty$ model reduction using LMIs. *IEEE Transactions on Automatic Control*, 49(7):1187–1191, 2004.

[73] S. Eleftheriadis, T. Nicholson, M. Deisenroth, and J. Hensman. Identification of Gaussian process state space models. *Advances in neural information processing systems*, 30, 2017.

[74] A. F. Esfahani, P. Dreesen, K. Tiels, J.-P. Noël, and J. Schoukens. Parameter reduction in nonlinear state-space identification of hysteresis. *Mechanical Systems and Signal Processing*, 104:884–895, 2018.

[75] P. Eykhoff. *System identification: parameter and state estimation*. Wiley-Interscience, London, 1974.

[76] M. Fazlyab, A. Robey, H. Hassani, M. Morari, and G. Pappas. Efficient and accurate estimation of Lipschitz constants for deep neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

[77] F. Feitzinger, T. Hylla, and E. W. Sachs. Inexact kleinman–newton method for Riccati equations. *SIAM Journal on Matrix Analysis and Applications*, 31(2):272–288, 2009.

[78] P. Feldmann and R. W. Freund. Efficient linear circuit analysis by Padé approximation via the Lanczos process. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 14(5):639–649, 1995.

[79] N. Figueroa and A. Billard. Locally active globally stable dynamical systems:

Theory, learning, and experiments. *The International Journal of Robotics Research*, 41(3):312–347, 2022.

[80] M. Forgione and D. Piga. dynonet: A neural network architecture for learning dynamical systems. *International Journal of Adaptive Control and Signal Processing*, 35(4):612–626, 2021.

[81] R. Frigola, Y. Chen, and C. E. Rasmussen. Variational Gaussian process state-space models. *Advances in neural information processing systems*, 27, 2014.

[82] R. Frigola, F. Lindsten, T. B. Schön, and C. E. Rasmussen. Bayesian inference and learning in Gaussian process state-space models with particle MCMC. *Advances in neural information processing systems*, 26, 2013.

[83] V. Fromion, S. Monaco, and D. Normand-Cyrot. Asymptotic properties of incrementally stable systems. *IEEE Transactions on Automatic Control*, 41(5):721–723, 1996.

[84] K. Fujimoto and J. M. A. Scherpen. Nonlinear input-normal realizations based on the differential eigenstructure of Hankel operators. *IEEE Transactions on Automatic Control*, 50(1):2–18, 2005.

[85] K. Fujimoto and J. M. Scherpen. Balanced realization and model order reduction for nonlinear systems based on singular value analysis. *SIAM Journal on Control and Optimization*, 48(7):4591–4623, 2010.

[86] K. Fujimoto and D. Tsubakino. Computation of nonlinear balanced realization and model reduction based on Taylor series expansion. *Systems & Control Letters*, 57(4):283–289, 2008.

[87] K. Fukumizu, F. R. Bach, and A. Gretton. Statistical consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research*, 8(Feb):361–383, 2007.

[88] P. Fulcheri and M. Olivi. Matrix rational h_2 approximation: a gradient algorithm based on Schur analysis. *SIAM Journal on Control and Optimization*, 36(6):2103–2127, 1998.

[89] P. Gahinet and P. Apkarian. A linear matrix inequality approach to H_∞ control. *International Journal of Robust and Nonlinear Control*, 4(4):421–448, jan 1994.

[90] O. Gajic, S. I. Dara, J. L. Mendez, A. O. Adesanya, E. Festic, et al. Ventilator-associated lung injury in patients without acute lung injury at the onset of mechanical ventilation. *Critical care medicine*, 32(9):1817–1824, 2004.

[91] K. Gallivan, A. Vandendorpe, and P. Van Dooren. Sylvester equations and projection-based model reduction. *Journal of Computational and Applied Mathematics*, 162(1):213–229, 2004.

[92] K. Gallivan, A. Vandendorpe, and P. Van Dooren. Model reduction of mimo systems via tangential interpolation. *SIAM Journal on Matrix Analysis and Applications*, 26(2):328–349, 2004.

[93] A. X. Garg, N. K. J. Adhikari, H. McDonald, M. P. Rosas-Arellano, P. J. Devereaux, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *Jama*, 293(10):1223–1238, 2005.

[94] H. Garnier and L. Wang. *Identification of Continuous-time Models from Sampled Data*. Springer Verlag London, 2008.

[95] J. C. Geromel, R. G. Egas, and F. R. R. Kawaoka. h_∞ model reduction with application to flexible systems. *IEEE Transactions on Automatic Control*, 50(3):402–406, 2005.

[96] F. Giri and E.-W. Bai. *Block-oriented nonlinear system identification*, volume 1. Springer–Verlag London, 2010.

[97] F. Girosi. An equivalence between sparse approximation and support vector machines. *Neural computation*, 10(6):1455–1480, 1998.

[98] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds. *International journal of control*, 39(6):1115–1193, 1984.

[99] G. C. Goodwin, J. C. Aguero, M. E. Cea Garridos, M. E. Salgado, and J. I. Yuz. Sampling and sampled-data models: The interface between the continuous world and digital algorithms. *IEEE Control Systems Magazine*, 33(5):34–53, 2013.

[100] S. Gugercin, A. C. Antoulas, and C. Beattie. $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM journal on matrix analysis and applications*, 30(2):609–638, 2008.

[101] S. Gugercin, R. V. Polyuga, C. Beattie, and A. Van Der Schaft. Structure-preserving tangential interpolation for model reduction of port-Hamiltonian systems. *Automatica*, 48(9):1963–1974, 2012.

[102] S. Gugercin. An iterative SVD-Krylov based method for model reduction of large-scale dynamical systems. *Linear Algebra and its Applications*, 428(8-9):1964–1986, 2008.

[103] S. Gugercin and A. C. Antoulas. A survey of model reduction by balanced truncation and some new results. *International Journal of Control*, 77(8):748–766, 2004.

[104] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42, 2018.

[105] R. G. HAKVOORT and P. M. J. V. D. HOF. Frequency domain curve fitting with maximum amplitude criterion and guaranteed stability. *International Journal of Control*, 60(5):809–825, 1994.

[106] Y. Halevi. Frequency weighted model reduction via optimal projection. In *Proceedings of the 29th conference on decision and control*, pages 2906–2911, 1990.

[107] G. Harnack, L. Lauwers, R. Pintelon, and J. Schoukens. Identification of nonlinear feedback systems using a structured polynomial nonlinear state space model. *IFAC Proceedings Volumes*, 42(10):332–337, 2009.

[108] P. Harshavardhana, E. Jonckheere, and L. Silverman. Stochastic balancing and approximation-stability and minimality. *IEEE Transactions on Automatic Control*, 29(8):744–746, 1984.

[109] F. Heck. Design and verification of a self-learning control approach for a respiratory module. Master's thesis, Eindhoven University of Technology, 2018.

[110] M. Heertjes, F. Cremers, M. Rieck, and M. Steinbuch. Nonlinear control of optical

storage drives with improved shock performance. *Control Engineering Practice*, 13(10):1295–1305, 2005.

[111] A. Helmersson. Model reduction using LMIs. In *Proceedings of the 33rd Conference on Decision and Control*, pages 3217–3222, 1994.

[112] E. Hendrix, P. Ortigosa, and I. García. On success rates for controlled random search. *Journal of Global Optimization*, 21(3):239–263, 2001.

[113] J. M. Hokanson and C. C. Magruder. $\mathcal{H}_2$-optimal model reduction using projected nonlinear least squares. *SIAM Journal on Scientific Computing*, 42(6):A4017–A4045, 2020.

[114] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

[115] P. Hu and F. Ding. Multistage least squares based iterative estimation for feedback nonlinear systems with moving average noises using the hierarchical identification principle. *Nonlinear Dynamics*, 73(1-2):583–592, 2013.

[116] B. Hunnekens, S. Kamps, and N. van de Wouw. Variable-gain control for respiratory systems. *IEEE Transactions on Control Systems Technology*, 2018.

[117] A. Hussein, M. M. Gaber, E. Elyan, and C. Jayne. Imitation learning: A survey of learning methods. *ACM Computing Surveys (CSUR)*, 50(2):1–35, 2017.

[118] D. C. Hyland and D. S. Bernstein. The optimal projection equations for model reduction and the relationships among the methods of wilson, skelton, and moore. *IEEE Transactions on Automatic Control*, 1985.

[119] T. C. Ionescu, O. V. Iftime, and I. Necoara. Optimal time-domain moment matching with partial placement of poles and zeros. In *Proceedings of the European Control Conference*, pages 1769–1774, 2020.

[120] T. Ionescu. Two-sided time-domain moment matching for linear systems. *IEEE Transactions on Automatic Control*, 61(9):2632–2637, 2015.

[121] T. C. Ionescu, A. Astolfi, and P. Colaneri. Families of moment matching based, low order approximations for linear systems. *Systems & Control Letters*, 64:47–56, 2014.

[122] A. Isidori. *Nonlinear control systems*. Springer, Berlin, 1995.

[123] ISO Central Secretary. Medical electrical equipment — Part 2-12: Particular requirements for basic safety and essential performance of critical care ventilators. Standard ISO 80601-2-12:2020, International Organization for Standardization, Geneva, CH, February 2020.

[124] M. Jansson. Subspace identification and ARX modeling. *IFAC Proceedings Volumes*, 36(16):1585–1590, 2003.

[125] S. Jin, Z. Wang, Y. Ou, and W. Feng. Learning accurate and stable dynamical system under manifold immersion and submersion. *IEEE Transactions on Neural Networks and Learning Systems*, 30(12):3598–3610, 2019.

[126] M. Jungers, N. van de Wouw, and M. F. Shakib. Discrete-time convergent nonlinear systems [Submitted]. *Transactions of Automatic Control*, 2022.

[127] G. C. Karras, C. P. Bechlioulis, M. Leonetti, N. Palomeras, P. Kormushev, et al. On-line identification of autonomous underwater vehicles through global derivative-

free optimization. In *Proceedings of the International Conference on Intelligent Robots and Systems*, pages 3859–3864, 2013.

[128] Y. Katznelson. *An introduction to harmonic analysis*. Cambridge: Cambridge University Press, 2004.

[129] Y. Kawano. Controller reduction for nonlinear systems by generalized differential balancing. *IEEE Transactions on Automatic Control*, 2021.

[130] Y. Kawano, B. Besselink, J. M. A. Scherpen, and M. Cao. Data-driven model reduction of monotone systems by nonlinear dc gains. *IEEE Transactions on Automatic Control*, 65(5):2094–2106, 2019.

[131] Y. Kawano and J. Scherpen. Model reduction by generalized differential balancing. In *Mathematical Control Theory I*, pages 349–362. Springer, Cham, Switzerland, 2015.

[132] Y. Kawano and J. M. A. Scherpen. Model reduction by differential balancing based on nonlinear hankel operators. *IEEE Transactions on Automatic Control*, 62(7):3293–3308, 2016.

[133] Y. Kawano and J. M. A. Scherpen. Empirical differential balancing for nonlinear systems. *IFAC-PapersOnLine*, 50(1):6326–6331, 2017.

[134] K. J. Keesman. *System identification: an introduction*. Springer London, 2011.

[135] H. K. Khalil. Nonlinear systems. *Prentice-Hall, New Jersey*, 1996.

[136] D. Khandelwal. *Automating data-driven modelling of dynamical systems: an evolutionary computation approach*. Springer Nature, Cham, Switzerland, 2022.

[137] S. M. Khansari-Zadeh and A. Billard. Learning control Lyapunov function to ensure stability of dynamical system-based robot reaching motions. *Robotics and Autonomous Systems*, 62(6):752–765, 2014.

[138] M. Khosravi and R. S. Smith. Nonlinear system identification with prior knowledge on the region of attraction. *IEEE Control Systems Letters*, 5(3):1091–1096, 2020.

[139] G. Kimeldorf and G. Wahba. Some results on Tchebycheffian spline functions. *Journal of mathematical analysis and applications*, 33(1):82–95, 1971.

[140] J. Kocijan. *Modelling and control of dynamic systems using Gaussian process models*. Springer Cham, Switzerland, 2016.

[141] J. Kocijan, A. Girard, B. Banko, and R. Murray-Smith. Dynamic systems identification with Gaussian processes. *Mathematical and Computer Modelling of Dynamical Systems*, 11(4):411–424, 2005.

[142] J. Z. Kolter and G. Manek. Learning stable deep dynamics models. *Advances in neural information processing systems*, 32, 2019.

[143] Ü. Kotta, K. Schlacher, and M. Tõnso. Relaxing realizability conditions for discrete-time nonlinear systems. *Automatica*, 58:67–71, 2015.

[144] K. Kundert and A. Sangiovanni-Vincentelli. Simulation of nonlinear circuits in the frequency domain. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 5(4):521–535, 1986.

[145] B. Lachmann. Open up the lung and keep the lung open. *Intensive care medicine*, 18(6):319–321, 1992.

[146] S. Lacy and D. Bernstein. Subspace identification with guaranteed stability using constrained optimization. *IEEE Transactions on Automatic Control*, 48(7):1259–1263, 2003.

[147] W. E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *Proceedings of the 29th Conference on Decision and Control*, pages 596–604, 1990.

[148] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[149] K. Lee and K. T. Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics*, 404:108973, 2020.

[150] A. Lepschy, G. A. Mian, G. Pinato, and U. Viaro. Rational L_2 approximation: a non-gradient algorithm. In *Proceedings of the 30th Conference on Decision and Control*, pages 2321–2323, 1991.

[151] E. H. Lieb and M. Loss. *Analysis*. American Mathematical Society, Providence, RI, 2001.

[152] L. Ljung. *System identification: theory for the user*. Prentice-Hall, Upper Saddle River, NJ, second edition, 1999.

[153] L. Ljung. *System identification toolbox: User's guide*. Natick, MA: The MathWorks, Inc., 2003.

[154] L. Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.

[155] L. Ljung, C. Andersson, K. Tiels, and T. B. Schön. Deep learning and system identification. *IFAC-PapersOnLine*, 53(2):1175–1181, 2020.

[156] M. Locatelli and F. Schoen. *Global optimization: theory, algorithms, and applications*, volume 15. SIAM, Philadelphia, 2013.

[157] J. Lofberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *Proceedings of the IEEE international conference on robotics and automation*, pages 284–289, 2004.

[158] W. Lohmiller and J. J. E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998.

[159] I. R. Manchester, M. M. Tobenkin, and J. Wang. Identification of nonlinear systems with stable oscillations. In *Proceedings of the 50th Conference on Decision and Control and European Control Conference*, pages 5792–5797, 2011.

[160] A. Marconato, M. Schoukens, K. Tiels, W. D. Widanage, A. Abu-Rmileh, and J. Schoukens. Comparison of several data-driven non-linear system identification methods on a simplified glucoregulatory system example. *IET Control Theory & Applications*, 8(17):1921–1930, 2014.

[161] A. Marconato, J. Sjöberg, J. Suykens, and J. Schoukens. Identification of the Silverbox benchmark using nonlinear state-space models. In *IFAC Proceedings Volumes (IFAC-PapersOnline)*, volume 16, pages 632–637. IFAC Secretariat, 2012.

[162] A. Marconato, J. Sjöberg, J. A. Suykens, and J. Schoukens. Improved initialization

for nonlinear state-space modeling. *IEEE Transactions on Instrumentation and Measurement*, 63(4):972–980, 2013.

[163] D. Masti and A. Bemporad. Learning nonlinear state–space models using autoencoders. *Automatica*, 129:109666, 2021.

[164] G. Matheron. The intrinsic random functions and their applications. *Advances in applied probability*, 5(3):439–468, 1973.

[165] M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner. *Autonomous driving: technical, legal and social aspects*. Springer Berlin, Germany, 2016.

[166] B. Mavkov, M. Forgione, and D. Piga. Integrated neural networks for nonlinear continuous-time system identification. *IEEE Control Systems Letters*, 4(4):851–856, 2020.

[167] H. Mayer, F. Gomez, D. Wierstra, I. Nagy, A. Knoll, and J. Schmidhuber. A system for robotic heart surgery that learns to tie knots using recurrent neural networks. *Advanced Robotics*, 22(13-14):1521–1537, 2008.

[168] A. Mees. The describing function matrix. *IMA Journal of Applied Mathematics*, 10(1):49–67, 1972.

[169] L. Meier and D. Luenberger. Approximation of linear constant systems. *IEEE Transactions on Automatic Control*, 12(5):585–588, 1967.

[170] G. Mercère, I. Markovsky, and J. A. Ramos. Innovation-based subspace identification in open- and closed-loop. In *Proceedings of the 55th Conference on Decision and Control*, pages 2951–2956, 2016.

[171] M. Milanese and C. Novara. Model quality in identification of nonlinear systems. *IEEE Transactions on Automatic Control*, 50(10):1606–1611, 2005.

[172] J. Miller and M. Hardt. Stable recurrent models. *arXiv preprint arXiv:1805.10369*, 2018.

[173] F. A. Miranda-Villatoro, F. Forni, and R. J. Sepulchre. Analysis of Lur'e dominant systems in the frequency domain. *Automatica*, 98:76–85, 2018.

[174] B. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, 1981.

[175] A. V. Mulders, L. Vanbeylen, and K. Usevich. Identification of a block-structured model with several sources of nonlinearity. In *Proceedings of the European Control Conference*, pages 1717–1722, 2014.

[176] K. S. Narendra and K. Parthasarathy. Identification and control of dynamical systems using neural networks. *IEEE Transactions on neural networks*, 1(1):4–27, 1990.

[177] I. Necoara and T. Ionescu. Optimal H_2 moment matching-based model reduction for linear systems by (non) convex optimization. *arXiv preprint arXiv:1811.07409*, 2018.

[178] I. Necoara and T. Ionescu. Parameter selection for best H_2 moment matching-based model approximation through gradient optimization. In *Proceedings of the 18th European Control Conference*, pages 2301–2306, 2019.

[179] I. Necoara and T. C. Ionescu. H_2 model reduction of linear network systems by moment matching and optimization. *IEEE Transactions on Automatic Control*, 65(12):5328–5335, 2020.

[180] D. Nešić and A. R. Teel. A framework for stabilization of nonlinear sampled-data systems based on their approximate discrete-time models. *IEEE Transactions on Automatic Control*, 49(7):1103–1122, 2004.

[181] D. Nešić, A. Teel, and P. Kokotović. Sufficient conditions for stabilization of sampled-data nonlinear systems via discrete-time approximations. *Systems & Control Letters*, 38(4-5):259–270, 1999.

[182] H. Nijmeijer. Observability of autonomous discrete time non-linear systems: a geometric approach. *International journal of control*, 36(5):867–874, 1982.

[183] J.-P. Noël, A. F. Esfahani, G. Kerschen, and J. Schoukens. A nonlinear state-space approach to hysteresis identification. *Mechanical Systems and Signal Processing*, 84:171–184, 2017.

[184] G. Obinata and B. D. Anderson. *Model reduction for control system design*. Springer–Verlag, New York, 2001.

[185] P. C. Opdenacker and E. A. Jonckheere. A contraction mapping preserving balanced reduction scheme and its infinity norm error bounds. *IEEE Transactions on Circuits and Systems*, 35(2):184–189, 1988.

[186] A. Padoan. Model reduction by least squares moment matching for linear and nonlinear systems. *arXiv preprint arXiv:2110.06072*, 2021.

[187] A. Padoan, F. Forni, and R. Sepulchre. Model reduction of dominant feedback systems. *Automatica*, 130:109695, 2021.

[188] J. Paduart, G. Horvath, and J. Schoukens. Fast identification of systems with nonlinear feedback. *IFAC Proceedings Volumes*, 37(13):381–385, 2004.

[189] J. Paduart, L. Lauwers, R. Pintelon, and J. Schoukens. Identification of a wiener–hammerstein system using the polynomial nonlinear state space approach. *Control Engineering Practice*, 20(11):1133–1139, 2012.

[190] J. Paduart, L. Lauwers, J. Swevers, K. Smolders, J. Schoukens, and R. Pintelon. Identification of nonlinear systems using polynomial nonlinear state space models. *Automatica*, 46(4):647–656, 2010.

[191] G. Palm. On representation and approximation of nonlinear systems - Part II: Discrete time. *Biological Cybernetics*, 34(1):49–52, sep 1979.

[192] G. Pannocchia, J. B. Rawlings, and S. J. Wright. Fast, large-scale model predictive control by partial enumeration. *Automatica*, 43(5):852–860, 2007.

[193] P. Y. Papalambros and D. J. Wilde. *Principles of Optimal Design: Modeling and Computation*. Cambridge University Press, Cambridge, third edition, 2017.

[194] J. Park and I. W. Sandberg. Universal approximation using radial-basis-function networks. *Neural computation*, 3(2):246–257, 1991.

[195] T. Parker and L. Chua. *Practical numerical algorithms for chaotic systems*. Springer Science & Business Media, New York, NY, 2012.

[196] A. Pavlov, B. G. Hunnekens, N. v. Wouw, and H. Nijmeijer. Steady-state performance optimization for nonlinear control systems of Lur'e type. *Automatica*, 49(7):2087–2097, 2013.

[197] A. Pavlov, A. Pogromsky, N. van de Wouw, and H. Nijmeijer. Convergent dynamics, a tribute to Boris Pavlovich Demidovich. *Systems & Control Letters*, 52(3):257 – 261, 2004.

[198] A. Pavlov and N. van de Wouw. Convergent discrete-time nonlinear systems: The case of PWA systems. In *Proceedings of the American Control Conference*, pages 3452–3457, 2008.

[199] A. Pavlov and N. van de Wouw. Steady-state analysis and regulation of discrete-time nonlinear systems. *IEEE Transactions on Automatic Control*, 57(7):1793–1798, 2012.

[200] A. Pavlov, N. van de Wouw, and H. Nijmeijer. *Uniform output regulation of nonlinear systems: a convergent dynamics approach*. Springer Science & Business Media, Birkhäuser Boston, 2006.

[201] A. Pavlov and N. v. d. Wouw. Convergent systems: nonlinear simplicity. In *Nonlinear systems*, pages 51–77. Springer, Cham, Switzerland, 2017.

[202] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3):657–682, 2014.

[203] G. Pillonetto, M. H. Quang, and A. Chiuso. A new kernel-based approach for nonlinear system identification. *IEEE Transactions on Automatic Control*, 56(12):2825–2840, 2011.

[204] R. Pintelon, J. Schoukens, and Y. Rolain. Frequency-domain approach to continuous-time system identification: Some practical aspects. In *Identification of continuous-time models from sampled data*, pages 215–248. Springer, London, 2008.

[205] R. Pintelon and J. Schoukens. *System identification: a frequency domain approach*. Hoboken, New Jersey, USA: Wiley & Sons Inc., 2012.

[206] W. L. Price. A controlled random search procedure for global optimisation. *The Computer Journal*, 20(4):367–370, 1977.

[207] G. Rao and H. Unbehauen. Identification of continuous-time systems. *IEE Proceedings - Control Theory and Applications*, 153:185–220(35), March 2006.

[208] C. E. Rasmussen and H. Nickisch. Gaussian processes for machine learning (gpml) toolbox. *Journal of machine learning research*, 11(Nov):3011–3015, 2010.

[209] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, USA, 2005.

[210] J. B. Rawlings, D. Q. Mayne, and M. M. Diehl. *Model predictive control: theory, computation, and design*. Nob Hill Publishing Madison, Santa Barbara, California, 2nd edition, 2017.

[211] J. Reinders, B. Hunnekens, F. Heck, T. Oomen, and N. van de Wouw. Adaptive control for mechanical ventilation for improved pressure support. *IEEE Transactions on Control Systems Technology*, 2020.

[212] J. Reinders, B. Hunnekens, N. van de Wouw, and T. Oomen. Noninvasive breathing effort estimation of mechanically ventilated patients using sparse optimization. *IEEE Open Journal of Control Systems*, 2022.

[213] R. Relan and J. Schoukens. Recursive discrete-time models for continuous-time systems under band-limited assumptions. *IEEE Transactions on Instrumentation and Measurement*, 65(3):713–723, 2016.

[214] M. Revay, R. Wang, and I. R. Manchester. A convex parameterization of robust recurrent neural networks. *IEEE Control Systems Letters*, 5(4):1363–1368, 2020.

[215] M. Revay, R. Wang, and I. R. Manchester. Recurrent equilibrium networks: Unconstrained learning of stable and robust dynamical models. In *Proceedings of the Conference on Decision and Control*, pages 2282–2287, 2021.

[216] A. H. Ribeiro and L. A. Aguirre. Shooting methods for parameter estimation of output error models. *IFAC-PapersOnLine*, 50(1):13998–14003, 2017.

[217] A. H. Ribeiro, K. Tiels, J. Umenberger, T. B. Schön, and L. A. Aguirre. On the smoothness of nonlinear system identification. *Automatica*, 121:109158, 2020.

[218] S. Z. Rizvi, J. M. Velni, F. Abbasi, R. Tóth, and N. Meskin. State-space LPV model identification using kernelized machine learning. *Automatica*, 88:38–47, 2018.

[219] B. S. Rüffer, N. van de Wouw, and M. Mueller. Convergent systems vs. incremental stability. *Systems & Control Letters*, 62(3):277–285, 2013.

[220] N. K. Rutland and P. G. Lane. Computing the 1-norm of the impulse response of linear time-invariant systems. *Systems & control letters*, 26(3):211–221, 1995.

[221] I. Salehi, G. Rotithor, G. Yao, and A. P. Dani. Dynamical system learning using extreme learning machines with safety and stability guarantees. *International Journal of Adaptive Control and Signal Processing*, 35(6):894–914, 2021.

[222] A. Sarkar and J. M. A. Scherpen. Extended differential balancing for nonlinear dynamical systems. *IEEE Control Systems Letters*, 2022.

[223] G. Scarciotti, Z. Jiang, and A. Astolfi. Data-driven constrained optimal model reduction. *European Journal of Control*, 53:68–78, 2020.

[224] G. Scarciotti. Low computational complexity model reduction of power systems with preservation of physical characteristics. *IEEE Transactions on Power Systems*, 32(1):743–752, 2016.

[225] G. Scarciotti and A. Astolfi. Data-driven model reduction by moment matching for linear and nonlinear systems. *Automatica*, 79:340–351, 2017.

[226] G. Scarciotti and A. Astolfi. Nonlinear model reduction by moment matching. *Foundations and Trends® in Systems and Control*, 4(3-4):224–409, 2017.

[227] G. Scarciotti, A. R. Teel, and A. Astolfi. Model reduction for linear differential inclusions: moment-set and time-variance. In *Proceedings of the American Control Conference*, pages 3483–3487, 2017.

[228] C. Scherer. *The Riccati inequality and state-space $\mathcal{H}_\infty$-optimal control.* PhD thesis, Universitat Wurzburg, Germany, 1990.

[229] C. Scherer and S. Weiland. *Linear matrix inequalities in control*, pages 24/1–24/30. CRC Press, 2011.

[230] J. M. A. Scherpen. Balancing for nonlinear systems. *Systems & Control Letters*, 21(2):143–153, 1993.

[231] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[232] M. Schmidt and H. Lipson. Distilling free-form natural laws from experimental data. *science*, 324(5923):81–85, 2009.

[233] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, Cambridge, USA, 2001.

[234] J. Schoukens, J. Nemeth, P. Crama, Y. Rolain, and R. Pintelon. Fast approximate identification of nonlinear systems. *Automatica*, 39(7):1267–1274, 2003.

[235] J. Schoukens, R. Pintelon, R. Dobrowiecki, and Y. Rolain. Identification of linear systems with nonlinear distortions. *Automatica*, 41(3):491–504, 2005.

[236] J. Schoukens, R. Relan, and M. Schoukens. Discrete time approximation of continuous time nonlinear state space models. *IFAC-PapersOnLine*, 50(1):8339–8346, 2017.

[237] J. Schoukens, M. Vaes, A. Esfahani, and R. Relan. Challenges in the identification of discrete time block oriented models for continuous time nonlinear systems. *IFAC-PapersOnLine*, 48(28):596–601, 2015.

[238] J. Schoukens, L. Gommé, W. Van Moer, and Y. Rolain. Identification of a block-structured nonlinear feedback system, applied to a microwave crystal detector. *IEEE Transactions on Instrumentation and Measurement*, 57(8):1734–1740, 2008.

[239] J. Schoukens and L. Ljung. Wiener-Hammerstein benchmark. In *Proceedings of the 15th IFAC Symposium on System Identification*, 2009.

[240] J. Schoukens and L. Ljung. Nonlinear system identification: A user-oriented road map. *IEEE Control Systems Magazine*, 39(6):28–99, 2019.

[241] J. Schoukens, R. Pintelon, and H. van Hamme. Identification of linear dynamic systems using piecewise constant excitations: use, misuse and alternatives. *Automatica*, 30(7):1153–1169, 1994.

[242] J. Schoukens, M. Vaes, and R. Pintelon. Linear system identification in a nonlinear setting: Nonparametric analysis of the nonlinear distortions and their impact on the best linear approximation. *IEEE Control Systems Magazine*, 36(3):38–69, 2016.

[243] M. Schoukens and K. Tiels. Identification of block-oriented nonlinear systems starting from linear approximations: A survey. *Automatica*, 85:272–292, 2017.

[244] M. Schoukens and R. Tóth. On the initialization of nonlinear LFR model identification with the best linear approximation. *IFAC-PapersOnLine*, 53(2):310–315, 2020.

[245] E. Seedhouse. Falcon 9 and Falcon Heavy. In *SpaceX*, pages 71–93. Springer, 2022.

[246] A. Semlyen and A. Medina. Computation of the periodic steady state in systems with nonlinear components using a hybrid time and frequency domain methodology. *IEEE Transactions on Power Systems*, 10(3):1498–1504, 1995.

[247] M. F. Shakib, A. Pogromsky, A. Pavlov, and N. van de Wouw. Fast identification of continuous-time Lur'e-type systems with stability certification. *IFAC-PapersOnLine*, 52(16):227–232, 2019.

[248] M. F. Shakib, G. Scarciotti, A. Pogromsky, A. Pavlov, and N. van de Wouw. Model reduction by moment matching for convergent Lur'e-type models. In *Proceedings of the American Control Conference*, pages 4459–4464, 2021.

[249] M. F. Shakib, G. Scarciotti, M. Jungers, A. Pogromsky, A. Pavlov, and N. van de Wouw. Optimal $\mathcal{H}_\infty$ LMI-Based Model Reduction by Moment Matching for Linear Time-Invariant Models. In *Proceedings of the 60th Conference on Decision and Control*, pages 6914–6919, 2021.

[250] M. F. Shakib, R. Tóth, A. Pogromsky, A. Pavlov, and N. van de Wouw. State-space kernelized closed-loop identification of nonlinear systems. *IFAC-PapersOnLine*, 53(2):1126–1131, 2020.

[251] M. F. Shakib, A. Y. Pogromsky, A. Pavlov, and N. van de Wouw. Computationally efficient identification of continuous-time Lur'e-type systems with stability guarantees. *Automatica*, 136:110012, 2022.

[252] L. Shampine, I. Gladwell, L. Shampine, and S. Thompson. *Solving ODEs with MATLAB*. Cambridge: Cambridge University Press, 2003.

[253] E. Simon, P. R-Ayerbe, C. Stoica, D. Dumur, and V. Wertz. LMIs-based coordinate descent method for solving BMIs in control design. *IFAC Proceedings Volumes*, 44(1):10180–10186, 2011.

[254] V. Simoncini. Analysis of the rational Krylov subspace projection method for large-scale algebraic Riccati equations. *SIAM Journal on Matrix Analysis and Applications*, 37(4):1655–1674, 2016.

[255] V. Sindhwani, S. Tu, and M. Khansari. Learning contracting vector fields for stable imitation learning. *arXiv preprint arXiv:1804.04878*, 2018.

[256] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, et al. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724, 1995.

[257] S. Skogestad and I. Postlethwaite. *Multivariable feedback control: analysis and design*, volume 2. Wiley New York, 2007.

[258] A. S. Slutsky and V. M. Ranieri. Ventilator-induced lung injury. *New England Journal of Medicine*, 369(22):2126–2136, 2013.

[259] E. Snelson and Z. Ghahramani. Sparse Gaussian processes using pseudo-inputs. *Advances in neural information processing systems*, 18, 2005.

[260] T. Soderstrom, H. Fan, B. Carlsson, and S. Bigi. Least squares parameter estimation of continuous-time ARX models from discrete-time data. *IEEE Transactions on Automatic Control*, 42(5):659–673, 1997.

[261] T. Söderström and P. Stoica. *System Identification*. Prentice-Hall, New York, 1989.

[262] T. Söderström. Convergence properties of the generalised least squares identitication method. *Automatica*, 10(6):617–626, 1974.

[263] E. D. Sontag. On the input-to-state stability property. *European Journal of Control*, 1:24–36, 1995.

[264] D. Sorensen and A. Antoulas. The Sylvester equation and approximate balanced reduction. *Linear Algebra and its Applications*, 351-352:671–700, 2002. Fourth Special Issue on Linear Systems and Control.

[265] S. L. Speetjens, J. D. Stigter, and G. van Straten. Physics-based model for a water-saving greenhouse. *Biosystems Engineering*, 105(2):149–159, feb 2010.

[266] I. Steinwart. On the influence of the kernel on the consistency of support vector machines. *Journal of machine learning research*, 2(Nov):67–93, 2001.

[267] B. L. Stevens, F. L. Lewis, and E. N. Johnson. *Aircraft control and simulation: dynamics, controls design, and autonomous systems*. John Wiley & Sons, Hoboken, NJ, USA, 2015.

[268] J. D. Stigter and K. J. Keesman. Optimal parametric sensitivity control of a fed-batch reactor. *Automatica*, 40(8):1459–1464, 2004.

[269] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. P. L. Vandewalle. *Least squares support vector machines*. World scientific Singapore, 2002.

[270] J. Suykens, B. de Moor, and J. Vandewalle. Robust local stability of multilayer recurrent neural networks. *IEEE Transactions on Neural Networks*, 11(1):222–229, 2000.

[271] J. Suykens, J. Vandewalle, and B. De Moor. Nl_q theory: checking and imposing stability of recurrent neural networks for nonlinear modeling. *IEEE Transactions on Signal Processing*, 45(11):2682–2691, 1997.

[272] J. A. K. Suykens, B. L. R. De Moor, and J. Vandewalle. Nonlinear system identification using neural state space models, applicable to robust control design. *International Journal of Control*, 62(1):129–152, 1995.

[273] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, et al. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[274] N. Takeishi and Y. Kawahara. Learning dynamics models with stable invariant sets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 9782–9790, 2021.

[275] F. Tehrani, M. Rogers, T. Lo, T. Malinowski, S. Afuwape, et al. A dual closed-loop control system for mechanical ventilation. *Journal of clinical monitoring and computing*, 18(2):111–129, 2004.

[276] N. Telang and L. Hunt. Frequency domain computations for nonlinear steady-state solutions. *IEEE transactions on signal processing*, 49(8):1728–1733, 2001.

[277] J. Thompson and H. Stewart. *Nonlinear dynamics and chaos: geometrical methods for engineers and scientists*. Wiley New York, 1986.

[278] E. C. Titchmarsh. *The theory of the Riemann zeta-function*. Oxford University Press, New York, 2nd edition, 1986.

[279] M. Titsias. Variational learning of inducing variables in sparse Gaussian processes. In *Proceedings of the international conference on Artificial intelligence and statistics*, pages 567–574, 2009.

[280] M. M. Tobenkin, I. R. Manchester, and A. Megretski. Convex parameterizations and fidelity bounds for nonlinear identification and reduced-order modelling. *IEEE Transactions on Automatic Control*, 62(7):3679–3686, 2017.

[281] M. M. Tobenkin, I. R. Manchester, J. Wang, A. Megretski, and R. Tedrake. Convex

optimization in identification of stable non-linear state space models. In *Proceedings of the 49th Conference on Decision and Control (CDC)*, pages 7232–7237, 2010.

[282] D. N. Tran, B. S. Rüffer, and C. M. Kellett. Convergence properties for discrete-time nonlinear systems. *IEEE Transactions on Automatic Control*, 64(8):3415–3422, 2018.

[283] J. Umenberger and I. R. Manchester. Specialized interior point algorithm for stable nonlinear system identification. *IEEE Transactions on Automatic Control*, 2018.

[284] J. Umenberger and I. R. Manchester. Specialized interior point algorithm for stable nonlinear system identification. *CoRR*, abs/1803.01066, 2018.

[285] J. Umlauft, A. Lederer, and S. Hirche. Learning stable Gaussian process state space models. In *Proceedings of the American Control Conference*, pages 1499–1504, 2017.

[286] D. van Campen, R. Fey, F. van Liempt, and A. d. Kraker. Steady-state behaviour of a solar array system with elastic stops. In *Proceedings of the Symposium on New Applications of Nonlinear and Chaotic Dynamics in Mechanics*, pages 303–312, 1999.

[287] L. van de Kamp. *Nonlinear System Identification in Mechanical Ventilation*. Bachelor's thesis, Eindhoven University of Technology, 2019.

[288] G. Van der Veen, J.-W. van Wingerden, M. Bergamasco, M. Lovera, and M. Verhaegen. Closed-loop subspace identification methods: an overview. *IET Control Theory & Applications*, 7(10):1339–1358, 2013.

[289] P. Van Dooren, K. A. Gallivan, and P.-A. Absil. $\mathcal{H}_2$-optimal model reduction of MIMO systems. *Applied Mathematics Letters*, 21(12):1267–1273, 2008.

[290] T. Van Gestel, J. Suykens, P. Van Dooren, and B. De Moor. Identification of stable models in subspace identification by using regularization. *IEEE Transactions on Automatic Control*, 46(9):1416–1420, 2001.

[291] A. Van Mulders, J. Schoukens, and L. Vanbeylen. Identification of systems with localised nonlinearity: From state-space to block-structured models. *Automatica*, 49(5):1392–1396, 2013.

[292] A. Van Mulders, J. Schoukens, M. Volckaert, and M. Diehl. Two nonlinear optimization methods for black box identification compared. *Automatica*, 46(10):1675–1681, 2010.

[293] P. van Overschee and B. de Moor. *Subspace identification for linear systems: Theory-Implementation-Applications*. Kluwer Academic Press, Dordrecht, The Netherlands, 1996.

[294] T. Van Pelt and D. Bernstein. Non-linear system identification using Hammerstein and non-linear feedback models with piecewise linear static maps. *International Journal of Control*, 74(18):1807–1823, 2001.

[295] S. Van Vaerenbergh. *Kernel methods for nonlinear identification, equalization and separation of signals*. PhD thesis, University of Cantabria, February 2010. Software available at `https://github.com/steven2358/kmbox`.

[296] L. Vanbeylen. Nonlinear LFR block-oriented model: Potential benefits and improved, user-friendly identification method. *IEEE Transactions on Instrumentation and Measurement*, 62(12):3374–3383, 2013.

[297] G. Vandersteen and J. Schoukens. Measurement and identification of nonlinear systems consisting of linear dynamic blocks and one static nonlinearity. *IEEE Transactions on Automatic Control*, 44(6):1266–1271, 1999.

[298] V. Verdult, J. A. Suykens, J. Boets, I. Goethals, and B. De Moor. Least squares support vector machines for kernel CCA in nonlinear state-space identification. In *Proceedings of the 16th international symposium on Mathematical Theory of Networks and Systems*, pages 1–11, 2004.

[299] H. Verrelst, K. Van Acker, J. Suykens, N. Motmans, B. De Moor, and J. Vandewalle. Application of nl_q neural control theory to a ball and beam system. *European journal of control*, 4(2):148–157, 1998.

[300] Z. Wang, Y. Shen, D. Wu, and Z. Ji. Hierarchical least squares algorithms for nonlinear feedback system modeling. *Journal of the Franklin Institute*, 353(10):2258–2269, 2016.

[301] P. A. Weber. Telerobotic-assisted laparoscopic right and sigmoid colectomies for benign disease. *Diseases of the colon & rectum*, 45(12):1689–1696, 2002.

[302] S. H. Weintraub. Jordan canonical form: theory and practice. *Synthesis Lectures on Mathematics and Statistics*, 2(1):1–108, 2009.

[303] D. T. Westwick, J. Decuyper, and J. Schoukens. Decoupling non-polynomial functions: A neural network example. *IFAC-PapersOnLine*, 54(7):667–672, 2021.

[304] T. Wigren and J. Schoukens. Data for benchmarking in nonlinear system identification. *Technical Reports from the department of Information Technology*, 6:2013–006, 2013.

[305] C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, USA, 2006.

[306] D. Wilson. Optimum solution of model-reduction problem. In *Proceedings of the Institution of Electrical Engineers*, volume 117, pages 1161–1165, 1970.

[307] W. Xiao, A. Lederer, and S. Hirche. Learning stable nonparametric dynamical systems with Gaussian process regression. *IFAC-PapersOnLine*, 53(2):1194–1199, 2020.

[308] Z. Xiao and X. Jing. Frequency-domain analysis and design of linear feedback of nonlinear systems and applications in vehicle suspensions. *IEEE/ASME Transactions on Mechatronics*, 21(1):506–517, 2016.

[309] V. Yakubovich. Matrix inequalities method in stability theory for nonlinear control systems: I. absolute stability of forced vibrations. *Automation and remote control*, 7:905–917, 1964.

[310] W.-Y. Yan and J. Lam. An approximate approach to $\mathcal{H}_2$ optimal model reduction. *IEEE Transactions on Automatic Control*, 44(7):1341–1358, 1999.

[311] J. Yeh. *Real analysis: theory of measure and integration*. World Scientific Publishing Company, Singapore, second edition, 2006.

[312] L. Zadeh. On the identification problem. *IRE Transactions on Circuit Theory*, 3(4):277–281, 1956.

[313] Y. Zhang, L. Cheng, H. Li, and R. Cao. Learning accurate and stable point-

to-point motions: A dynamic system approach. *IEEE Robotics and Automation Letters*, 2022.

[314] S. Zheng, Y. Song, T. Leung, and I. Goodfellow. Improving the robustness of deep neural networks via stability training. In *Proceedings of the conference on computer vision and pattern recognition*, pages 4480–4488, 2016.

[315] H. Zhu, M. Akrout, B. Zheng, A. Pelegris, A. Jayarajan, et al. Benchmarking and analyzing deep neural network training. In *Proceedings of the International Symposium on Workload Characterization*, pages 88–100, 2018.

[316] Y. Zhu. *Black-box identification of MIMO transfer functions: asymptotic properties of prediction error models*. EUT report. E, Fac. of Electrical Engineering. Technische Universiteit Eindhoven, Eindhoven, the Netherlands, 1987.

[317] U. Zulfiqar, V. Sreeram, M. Ilyas Ahmad, and X. Du. Frequency-weighted $\mathcal{H}_2$-pseudo-optimal model order reduction. *IMA Journal of Mathematical Control and Information*, 38(2):622–653, 2021.

[318] A. Zyner, S. Worrall, J. Ward, and E. Nebot. Long short term memory for driver intent prediction. In *Proceedings of the Intelligent Vehicles Symposium*, pages 1484–1489, 2017.

Samenvatting

Data-gedreven modellering en complexiteitsreductie voor niet-lineaire systemen met stabiliteitsgaranties

Dynamische modellen van fysieke systemen zijn cruciaal in veel engineering toepassingen. Dergelijke modellen maken de voorspelling van de systeemrespons, analyse van het dynamische gedrag van het systeem en het ontwerp van regelstrategieën mogelijk. Systeemidentificatie is een data-gebaseerde methode om een model te construeren waarvan de gesimuleerde respons nauw aansluit bij experimenteel verkregen data van het bestudeerde systeem. Als alternatief kunnen nauwkeurige modellen worden geconstrueerd door eerste-principemodellering. Omdat eerste-principemodellen vaak te complex zijn voor de bovengenoemde doeleinden, is er behoefte aan voldoende nauwkeurige, maar minder complexe modellen. Modelreductiemethoden kunnen worden gebruikt om dergelijke modellen met een lage complexiteit te construeren op basis van complexe eerste-principemodellen. Er bestaat een groot arsenaal aan technieken voor zowel systeemidentificatie als modelreductie van lineaire modellen. Fysieke systemen vertonen echter vaak niet-lineair gedrag. Daarom zijn voorspellende niet-lineaire modellen nodig. De bijdragen van dit proefschrift kunnen worden opgesplitst in drie delen, waarvan twee delen de systeemidentificatie van niet-lineaire modellen betreffen en één deel de modelreductie van niet-lineaire modellen betreft.

Het eerste deel van het proefschrift betreft het probleem van systeemidentificatie van niet-lineaire modellen. De volgende twee uitdagingen worden aangepakt voor dit probleem. Ten eerste, zelfs als het niet-lineaire systeem stabiliteitseigenschappen geniet, is het niet gegarandeerd dat het geïdentificeerde model deze eigenschappen behoudt. Bijgevolg kan het geïdentificeerde niet-lineaire model goed presteren op de trainingsdata, maar kan het onbegrensde responsen produceren voor nieuwe excitaties. Ten tweede komt het identificatieprobleem vaak neer op het oplossen van een niet-lineair optimalisatieprobleem. Het oplossen van

een dergelijk probleem vereist de berekening van veel modelresponsen. Daardoor hangt de rekentijd van het identificatiealgoritme sterk af van de efficiëntie van de berekening van modelresponsen. Deze efficiëntie is van cruciaal belang in tijdbeperkende toepassingen.

Een identificatiemethode ontwikkeld voor een klasse van niet-lineaire modellen die bestaan uit lineaire tijd-invariante (LTI) dynamica in feedback met statische niet-lineariteiten. Deze methode dwingt een sterke vorm van globale stabiliteit af, namelijk exponentiële convergentie. De convergentie-eigenschap zorgt voor de globale exponentiële stabiliteit van de tijd-variërende steady-state modelresponsen. Verder is er een simulatie-algoritme bedacht dat de rekentijd significant verbetert van uren naar minuten in numerieke casestudies. Verder is het voorgestelde simulatie-algoritme geanalyseerd in termen van het tijd-variërende steady-state verschil tussen de systeem- en modelrespons. Dit verschil blijkt asymptotisch tot nul te dalen door de rekenkosten van het algoritme te verhogen.

De identificatiemethode is experimenteel gevalideerd op een mechanische ventilatieopstelling. De resultaten bewijzen een snelle en nauwkeurige identificatie van de slang parameters en patiënt longparameters. Deze geïdentificeerde parameters kunnen worden gebruikt voor regeldoeleinden en het besluitvormingsproces van medisch personeel dat de mechanische ventilator bedient. Bovendien is de voorgestelde identificatiemethode gevalideerd op benchmarkdatasets. Deze resultaten laten een wisselwerking zien tussen de data fit en modelstabiliteit. De identificatiebenadering identificeert dus voorspellende modellen, dwingt modelstabiliteit af en is rekenkundig efficiënt, waardoor de snelle identificatie van nauwkeurige en stabiele niet-lineaire modellen in tijdbeperkende toepassingen mogelijk is.

Het tweede deel van dit proefschrift betreft het probleem van complexiteitsreductie voor niet-lineaire modellen. Modelreductie heeft als doel een voldoende nauwkeurig gereduceerd model te vinden dat het complexe model vervangt met behoud van belangrijke eigenschappen. In dit probleem is het een uitdaging om de modelstabiliteit te behouden, wat één van de belangrijkste modeleigenschappen is. Daarnaast is het een uitdaging om het verschil tussen de niet-lineaire modellen van hoge en lage complexiteit te begrenzen. Als laatste is het een uitdaging om het optimale gereduceerde model te vinden dat het verschil tussen het niet-lineaire model van lage en hoge complexiteit minimaliseert.

Twee uitdagingen voor LTI-modellen zijn behandeld als een voorbereidende stap naar het aanpakken van de uitdagingen voor niet-lineaire modellen. De specifieke reductietechniek gebruikt in dit proefschrift is de tijd-domein 'moment matching' methode. Als eerste is de verkregen matching-eigenschap van moment matching voor multivariabele LTI modellen afgeleid. Ten tweede zijn in het geval van LTI modellen de aanvullende moment matching parameters benut om modelstabiliteit af te dwingen en om de optimaliteit tussen de modellen van hoge en lage complexiteit te verzekeren. Deze resultaten zijn verder benut om een foutgrens af te leiden voor het verschil tussen de LTI modellen van lage en hoge complexiteit.

Een optimale modelreductiemethode voor niet-lineaire modellen die bestaan uit

LTI dynamica in feedback met niet-lineaire dynamica is ontwikkeld om de uitdagingen voor niet-lineaire modellen aan te pakken. Hiertoe wordt het tijd-domein moment matching raamwerk uitgebreid tot niet-lineaire modellen die genieten van een speciaal type globale exponentiële stabiliteit, namelijk exponentiële convergentie. Deze uitbreiding leidt tot een natuurlijke toepassing van moment matching op deze klasse van niet-lineaire modellen. Modellen met een gereduceerde complexiteit die door deze reductiemethode worden gevonden behouden de convergentie-eigenschap en zijn optimaal in termen van een minimale foutgrens. De voorgestelde methoden zijn gevalideerd op verschillende voorbeelden, wat resulteert in nauwkeurige modellen met een gereduceerde complexiteit die de stabiliteitseigenschap behouden.

Het derde deel van dit proefschrift betreft de identificatie van niet-lineaire modellen met behulp van een 'reproducing kernel Hilbert space'-benadering. Een dergelijke zwarte-doos-leermethode maakt gebruik van trainingsdata om modellen te identificeren die de respons van het oorspronkelijke systeem op nieuwe excitaties kunnen voorspellen. Ondanks dat ze een grote flexibiliteit bieden, missen deze modellen transparantie, wat het een uitdaging maakt om extra modeleigenschappen, zoals modelstabiliteit, af te dwingen. Bovendien levert een directe toepassing van al bestaande methoden vaak statistisch afwijkende modellen op wanneer de data in een gesloten lus experiment is verkregen.

Om de eerste uitdaging aan te pakken is een identificatiemethode voorgesteld die exponentiële convergentie op het geïdentificeerde model afdwingt. Met behulp van een benchmarkdataset is aangetoond dat het niet-lineaire gedrag nauwkeurig kan worden vastgelegd en de modelstabiliteit afgedwongen kan worden, terwijl modellen die zijn geïdentificeerd met behulp van al bestaande methoden instabiliteit vertonen voor nieuwe excitaties. Bovendien wordt een statistisch consistente methode voor de identificatie van LTI systemen die in gesloten lus opereren uitgebreid naar kernel-gebaseerde identificatie van niet-lineaire modellen, wat statisch consistente identificatieresultaten oplevert in numerieke voorbeelden.

Samenvattend presenteert dit proefschrift fundamentele methodes voor datagedreven modellering en complexiteitsreductie van niet-lineaire modellen met stabiliteitsgaranties. Deze methodes zijn experimenteel en numeriek gevalideerd.

List of publications

Peer-reviewed journal articles

- Shakib, M. F., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. (2022). Computationally efficient identification of continuous-time Lur'e-type systems with stability guarantees. *Automatica*, 136.

- Shakib, M. F., Detournay, E., & van de Wouw, N. (2019). Nonlinear dynamic modeling and analysis of borehole propagation for directional drilling. *International Journal of Non-Linear Mechanics*, 113, 178-201.

- Shakib, M. F., Scarciotti, G., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Time-domain moment matching for multiple-input multiple-output linear time-invariant models. *Provisionally accepted for publication in Automatica.*

- Jungers, M. Shakib, M. F., & van de Wouw, N. Discrete-time convergent nonlinear systems. *Submitted for journal publication.*

- Shakib, M. F., Scarciotti, G., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Model reduction by moment matching with preservation of global stability for a class of nonlinear models. *Submitted for journal publication.*

- Shakib, M. F., Vervaet, N., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Fast identification of multivariable discrete-time Lur'e-type systems with stability certification. *Submitted for journal publication.*

- Shakib, M. F., Scarciotti, G., Jungers, M., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. Optimal model reduction by moment matching: the linear and nonlinear feedback cases. *Submitted for journal publication.*

- Shakib, M. F., Schoukens, J., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. Accuracy bounds for the simulation of a class of continuous-time nonlinear models. *In preparation.*

- Amiri. A., Shakib, M. F., Arteaga, I. L., & van de Wouw, N. Nonlinear system identification of models for electro-active liquid crystal polymer coatings. *In preparation.*

Peer-reviewed articles in conference proceedings

- Shakib, M. F., Scarciotti, G., Jungers, M., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. (2021). Optimal $\mathcal{H}_\infty$ LMI-based model reduction by moment matching for linear time-invariant models. In *Proceedings of the Conference for Decision and Control*, (pp. 6914-6919).

- Shakib, M. F., Scarciotti, G., Pogromsky, A.Y., Pavlov, A., & van de Wouw, N. (2021). Model reduction by moment matching for convergent Lur'e-type models. In *Proceedings of the American Control Conference*, (pp. 4449-4454).

- Shakib, M. F., Tóth, R., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. (2020). State-space kernelized closed-loop identification of nonlinear systems. *IFAC-PapersOnLine*, 53(2), 1126-1131.

- Shakib, M. F., Pogromsky, A. Y., Pavlov, A., & van de Wouw, N. (2019). Fast identification of continuous-time Lur'e-type systems with stability certification. *IFAC-PapersOnLine*, 52(16), 227-232.

- Shakib, M. F., Detournay, E., & van de Wouw, N. (2019). Delay complementarity modeling for dynamic analysis of directional drilling. In *Proceedings of the American Control Conference*, (pp. 5209-5214).

- Shakib, F., Detournay, E., & van de Wouw, N. (2018). Nonlinear modeling of directional drilling. *Nonlinear Dynamics and Control of Deep Drilling Systems.*

Dankwoord

Dit proefschrift staat symbool voor het einde van mijn tijd als promovendi aan de TU/e. Een promotietraject is lang waarin je op zowel academisch als persoonlijk vlak veel ups en downs meemaakt. Hierbij prijs ik me gelukkig dat ik fijne mensen om me heen heb die me hierbij bijgestaan hebben. Jullie bijdragen aan dit proefschrift zijn onmisbaar en daar wil ik jullie graag voor bedanken.

Allereerst wil ik graag mijn promoteren bedanken. Nathan, je hebt me alle vrijheid gegeven om het onderzoek te vormen naar mijn wens, je hebt me gefaciliteerd in het opzetten van samenwerkingen en je hebt me betrokken bij het onderwijs. Ook heb je me geholpen en initiatieven genomen om een geschikte postdoc positie voor mij te vinden en plannen te maken voor de verdere toekomst. Buiten dit allemaal om ben je ook zeer prettig in de omgang en heb je me ook zowel op academisch als persoonlijk vlak altijd gesteund. Een fijnere begeleider had ik me niet kunnen wensen. Graag wil ik je hartelijk bedanken voor alles dat je voor me gedaan hebt.

Sasha, thanks a lot for your supervision and for being my go-to person for discussions. Your door was always open, which I enjoyed very much. I appreciate your eye for mathematical rigor and have learned a lot from you. Also, many thanks for sharing your experience in academia and helping me with making career plans.

Alexey, thank you for being involved in this track. Even though I have only met you a couple of times in real life, you were an indispensable part of the team. My research builds upon the blocks you once laid; therefore, it was very fruitful and comforting to have you on the team. Thanks for all your input and all the nice discussions.

I would like to thank the committee members Alessandro Astolfi, Jacquelien Scherpen, Johan Schoukens, and Paul van den Hof for reading, assessing, and providing feedback on my dissertation, and for taking part in my defense committee. I am also grateful for your contributions to the symposium before the defense ceremony.

I have had the pleasure of collaborating and interacting with several great people. Marc Jungers, thank you for your contributions to my thesis and for all the interesting discussions we had in the past years. I enjoy doing research with

you a lot. Furthermore, thanks a lot for hosting me in Nancy and for giving me an in-depth historical city tour of Nancy. Giordano Scarciotti, I enjoyed our collaboration very much. Thank you for all your contributions to my thesis and for your assistance in finding a postdoc position for me. I am looking forward to interacting with you more closely at Imperial College. Johan Schoukens, hartstikke bedankt voor al je bijdragen aan dit proefschrift. Ik vond het erg prettig en erg leerzaam om onderzoek met je te bedrijven. Roland Tóth, thanks a lot for all the interesting discussions and for being part of my research trajectory. I want to thank you for all your contributions to this thesis. Furthermore, many thanks for the career planning advice. Bram Hunnekens, graag wil ik jou, en daarbij ook Demcon macawi respiratory systems, bedanken voor de medewerking aan de experimenten van hoofdstuk twee. Finally, I would like to thank Jean-Philippe Noël, Maarten Schoukens, and Koen Tiels for the interesting discussions and your interest in my research.

Het onderzoek in dit proefschrift is mede tot stand gekomen door de hulp van studenten. Ik wil graag Lars, Calvin, Tim, Maxim en Harmen bedanken voor de interessante discussies, leerzame momenten en jullie bijdragen aan dit proefschrift. Niels, jij verdient een speciaal plekje hier, want jouw werk heeft een grote bijdrage geleverd aan hoofdstuk drie, waarvoor ik je enorm dankbaar ben.

Ook wil ik graag Geertje en Anouk bedanken voor het op jullie nemen van allerlei organisatorische taken zoals de D&C groepsuitjes, het maken van afspraken in overvolle agenda's en de herinneringen om af en toe vrij te nemen. Jullie hebben het onderzoek in dit proefschrift gefaciliteerd en jullie bijdrage is dus groot. Geertje, ook erg bedankt voor je bijdrage aan het organiseren van het symposium voorafgaand aan mijn verdediging.

A big thank you to all the (former) colleagues of the D&C and CST sections, and in particular, Andy, Bardia, Brandon, Chyannie, Daniel, Frans, Jijju, Lars (2x), Luuk, Maurice, Masahiro, Max, Mohammad, Nard, Nic, Rishi, Ricky, Roeland, Sajad, and Wouter, for the nice work ambiance, coffee breaks, Friday afternoon drinks, lunch walks, and the nice times during conferences. Thank you for making my time at the TU/e so enjoyable! Mijn buurman Noud verdient een extra speciaal plekje hier, mede door het ontstaan van begrippen zoals BOTjes, BOFjes, Zimbabwe 10 en de ochtend/middag/late-afternoon sessies. Anahita, you deserve a special spot too for the enjoyable collaboration over the years. I would also like to thank the members of our (former) Futsal team, Atletico Bitterballen, for uniting people from the D&C, CST, and CS sections. I have enjoyed a lot playing with all of you and getting the necessary breaks from work. Finally, I would like to thank Daniel, Jijju, and Roeland for the nice collaboration in organizing the D&C colloquium series.

Graag wil ik ook mijn kantoorgenoten bedanken. Joey en Camiel, het was heel speciaal om het promotietraject tegelijk met jullie te starten in -1.137 en jullie te zien promoveren. Het doet me goed om te zien dat jullie het naar je zin hebben bij jullie nieuwe werkgevers. Ik heb genoten van jullie aanwezigheid,

de vele ontspannende activiteiten tijdens werkuren (zoals CurveFever), de leuke cadeautjes bij verjaardagen (zoals de yogamat met mijn afbeelding erop) en de leuke momenten buiten het kantoor. Joey, graag wil ik je ook bedanken voor je bijdrage aan de experimenten van Hoofdstuk twee. Camiel, jou heb ik bijna dagelijks gezien op kantoor. Een van je beste ideeën tijdens je promotie was misschien wel het introduceren van de kantoorgitaar (onze buren vonden dat een minder goed idee), wat de werksfeer nog leuker maakte. Ook wil ik je hartstikke bedanken voor je steun in de laatste fase van mijn promotietraject, dit heeft een groot verschil gemaakt. Benn, in de korte tijd dat we een kantoor hebben gedeeld, heb ik je leren kennen als een fijne en enthousiaste kantoorgenoot. Ik hoop dat je veel plezier in je project hebt en wil je heel veel succes wensen met de rest van je promotietraject.

Ook wil ik al mijn vrienden bedanken. In het bijzonder wil ik graag Bart, Jeroen en Tom bedanken voor jullie vriendschap die nu al meer dan twintig jaar duurt en voor de nodige gezelligheid en afleiding. Fons, jou wil ik ook graag speciaal bedanken voor je vriendschap en voor het in leven laten komen van mijn kaftideeën met je geweldige illustratorvaardigheden. Huub, bedankt voor het aangaan van de Eden Outdoors uitdaging met mij! De potjes FIFA waren een essentieel onderdeel van zowel Eden Outdoors als de afronding van dit proefschrift. Melanie, jij mag natuurlijk ook niet ontbreken in deze lijst. Het beginnen met gitaarspelen is een van mijn beste beslissingen geweest in de afgelopen jaren. Bedankt voor je begeleiding daarbij en je fijne lessen (die door enthousiasme soms veel langer doorliepen dan gepland). Ook heeft voetbalclub Merefeldia een grote rol gespeeld in mijn leven. Hierbij wil ik alle personen die betrokken waren bij de club en de teams waarin ik gespeeld heb, in het bijzonder Jeroen, Stijn, en Thomas, bedanken voor de gezelligheid en sportiviteit (ook buiten het voetballen om).

Padar jan en madar jan, ik zal jullie altijd dankbaar zijn voor jullie beslissing en moed om naar Nederland te vertrekken. Dankzij die beslissing, jullie onvoorwaardelijke steun en liefde, en onze warme thuishaven sta ik waar ik nu sta. Ik heb alles aan jullie te danken. Mijn lieve zussen Arefah jan en Shakofah jan, en mijn lieve broertje Nadim jan wil ik ook graag bedanken; zonder jullie steun, liefde, gevraagde (en ongevraagde) adviezen, en gezelligheid was dit boekje niet tot stand gekomen. Ook wil ik graag de familie Leurs bedanken voor de warme ontvangst, de mooie momenten samen en jullie interesse in mij en mijn onderzoek.

Als allerlaatste wil ik graag mijn allerliefste Nina bedanken. Twee jaar geleden liep je zomaar mijn leven binnen op een donkere kletsnatte zondag. Sindsdien heb je de zon laten schijnen, mijn leven met geluk gevuld, en mijn hart met liefde gevuld. Bedankt voor je voortdurende steun, je onuitputtelijke geduld, en bovenal, het geluk en liefde dat je me biedt. Lieve Nina, ik hou van je en ik heb heel veel zin in alle avonturen die de toekomst op ons pad brengt.

Fahim Shakib
Eindhoven, October, 2022

About the author

Fahim Shakib was born on September 5, 1992, in Herat, Afghanistan. After finishing his secondary education in 2010 at the Philips van Horne, Weert, the Netherlands, he studied mechatronics at the Fontys University of Applied Sciences, Eindhoven, the Netherlands, where he received his Bachelor of Applied Science degree (cum laude) in 2014. As part of his bachelor's curriculum, he performed internships at GEA Food Solutions, Weert, the Netherlands, and at MI-Partners, Eindhoven, the Netherlands. He continued specializing in the field of Systems & Control in the Mechanical Engineering department of the Eindhoven University of Technology (TU/e), Eindhoven, the Netherlands, and received his Master of Science degree in 2017 (cum laude). As part of his master's curriculum, he performed an internship at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway, under the supervision of prof. Anton Shiriaev (NTNU). His master's thesis project focused on the nonlinear modeling and dynamic analysis of directional drilling processes and was conducted in collaboration with the University of Minnesota (UMN) in Minneapolis, USA, under the supervision of prof.dr.ir. Nathan van de Wouw (TU/e) and prof.dr. Emmanuel Detournay (UMN).

In 2018, Fahim started his Ph.D. research within the Dynamics and Control group at the Mechanical Engineering department of the TU/e under the supervision of prof.dr.ir. Nathan van de Wouw (TU/e), dr. Alexander Pogromsky (TU/e), and prof.dr. Alexey Pavlov (NTNU). During his Ph.D., Fahim was a visiting researcher at the Centre national de la recherche scientifique (CNRS) in Nancy, France (3 weeks in 2022), as part of his collaboration with dr. Marc Jungers. His research is centered around data-driven modeling and complexity reduction for nonlinear systems with stability guarantees. The main results of his research are presented in this thesis.

Stellingen behorende bij het proefschrift

Data-driven modeling and complexity reduction for nonlinear systems with stability guarantees

1. Het is relatief eenvoudig om een model te construeren dat de experimentele data van het systeem nabootst. Het is een grotere uitdaging om een model te construeren dat het systeem verklaart. *Dit proefschrift.*

2. Niet alleen het toevoegen van fysica-gebaseerde voorkennis aan het identificatieprobleem verbetert de interpreteerbaarheid van data-gebaseerde modellen. Ook het garanderen van systeemeigenschappen, bijvoorbeeld stabiliteit, verbetert de interpreteerbaarheid. *Dit proefschrift.*

3. Data-gebaseerde niet-lineaire modellen kunnen instabiel gedrag vertonen als gevolg van zowel extrapolatie als interpolatie. *Dit proefschrift, Hoofdstuk 9.*

4. Begrippen zoals convergentie en incrementele stabiliteit zijn kernbegrippen voor het uitbreiden van lokale model reductie technieken met stabiliteitsbehoud naar de globale context. *Dit proefschrift.*

5. Identificatie algoritmes worden voornamelijk geijkt op basis van de kwaliteit van voorspellingen van het geïdentificeerde model. Echter zijn andere criteria, zoals de rekenbelasting, geheugenverbruik, implementeerbaarheid en bruikbaarheid van het algoritme, ook onmisbaar voor een gebalanceerde vergelijking van identificatie algoritmes.

6. Het eenzijdig nastreven van bepaalde academische prestatie-indicatoren, zoals bijvoorbeeld het aantal citaties, kunnen op de lange termijn leiden tot fysieke en mentale onbalans en demotivatie bij wetenschappers. De sleutel tot creatieve, wetenschappelijke prestaties ligt echter in het behouden van fysieke en mentale balans.

7. Regelen is de theorie van het aanpassen van gedrag naar onze wens. Echter, in de praktijk is regelen een illusie.

8. Onderzoekers zullen tevreden moeten zijn met het gegeven dat de meeste onderzoeksvragen onbeantwoord blijven.

9. Bewustwording van onze insignificante invloed op het verloop van het universum brengt verlichting in de Ph.D. ervaring. *Gebaseerd op het boek 4000 Weken door Oliver Burkeman.*

10. Voetbalwedstrijden zouden met effectieve speeltijd afgewerkt moeten worden voor een eerlijker wedstrijd- en competitieverloop en voor minder ergernis op het veld.

Eindhoven, November 2022 Fahim Shakib

Propositions accompanying the thesis

Data-driven modeling and complexity reduction for nonlinear systems with stability guarantees

1. It is relatively simple to construct a model that mimics the experimental data from the system. It is more challenging to construct a model that explains the system. *This thesis.*

2. Not only the addition of physics-based prior knowledge to the identification problem enhances the interpretability of data-based models. Also guaranteeing system properties, for example, stability, enhances interpretability. *This thesis.*

3. Data-based nonlinear models can exhibit unstable behavior as a result of both extrapolation and interpolation. *This thesis, Chapter 9.*

4. Concepts such as convergence and incremental stability are core concepts for extending local model reduction techniques with stability preservation to the global context. *This thesis.*

5. Identification algorithms are mainly benchmarked in terms of the predictive quality of the identified model. However, other criteria such as the computational load, memory storage, implementability, and usability of the algorithm are also indispensable for a balanced comparison of identification algorithms.

6. The one-sided pursuit of certain academic performance indicators such as the number of citations, for example, can lead to physical and mental imbalance and demotivation for scientists in the long run. The key to creative, scientific performance, however, lies in maintaining a physical and mental balance.

7. To control is the theory of adapting behavior to our desire. However, in practice, control is an illusion.

8. Researchers will have to be satisfied with the fact that most research questions remain unanswered.

9. Awareness of our insignificant influence on the course of the universe brings relief to the Ph.D. experience. *Based on the book Four Thousand Weeks by Oliver Burkeman.*

10. Soccer matches should be played with effective playing time for a fairer course of games and competitions and fewer on-field annoyances.

Eindhoven, November 2022 Fahim Shakib

Erratum for the Ph.D. thesis: Data-driven modeling and complexity reduction for nonlinear systems with stability guarantees, Fahim Shakib, 14–11–2022

After publishing the thesis entitled 'Data-driven modeling and complexity reduction for nonlinear systems with stability guarantees', the author has noticed that some corrections regarding a part of the text in Chapter 4 and Appendix C are in order. The current erratum presents these corrections. The corrections are of technical nature and do not influence the main conclusions of Chapter 4, nor do these influence Figures 4.2 - 4.4 corresponding to the presented examples in Chapter 4. What follows is a list of errata.

1. **The following text on page 88:**

$$\xi_4 \coloneqq \gamma_F^M \left(\frac{\gamma_{zu}\gamma_{yw}}{1-\gamma_{yu}K_\varphi} + \gamma_{zw} \right)$$

should be replaced by:

$$\xi_4 \coloneqq \gamma_F^M \left(\frac{K_\varphi\gamma_{zu}\gamma_{yw}}{1-\gamma_{yu}K_\varphi} + \gamma_{zw} \right)$$

2. **The following text on page 91:**

$$\xi_4 = \xi_5 = \gamma_F^M \left(\gamma_{zu}\gamma_{yw} + \gamma_{zw} \right), \quad \xi_6 = \gamma_F^M .$$

should be replaced by:

$$\xi_4 = \xi_5 = \gamma_F^M \gamma_{zw}, \quad \xi_6 = \gamma_F^M .$$

3. **The following text on page 99:**

$$\left\| \breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M \right\|_{L_2^M} \leq \gamma_F^M (\gamma_{zu}\gamma_{yw} + \gamma_{zw}) K_{w_0}^\circ . \tag{4.19b}$$

should be replaced by:

$$\left\| \breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M \right\|_{L_2^M} \leq \gamma_F^M \gamma_{zw} K_{w_0}^\circ . \tag{4.19b}$$

4. **The following text on page 99:**

$$\left\| \breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M \right\|_{L_2^M} \leq \gamma_{zw}^M \Delta_w + \Delta_z + \gamma_F^M (\gamma_{zu}\gamma_{yw} + \gamma_{zw}) K_{v_w} + \gamma_F^M K_{v_z} . \tag{4.20b}$$

should be replaced by:

$$\left\| \breve{\bar{\mathbf{z}}}_M - \left(\bar{\mathbf{z}}_{\eta M}\right)^M \right\|_{L_2^M} \leq \gamma_{zw}^M \Delta_w + \Delta_z + \gamma_F^M \gamma_{zw} K_{v_w} + \gamma_F^M K_{v_z} . \tag{4.20b}$$

5. **The following text on page 238:** Consider the truncated function $\bar{y}^{\eta M} \in C^{0,1}_{K_y^{\eta M}}(T)$, $T \geq 0$ with some $0 \leq K_y^{\eta M} < +\infty$ and $\bar{u} = \varphi(\bar{y}^{\eta M}) \in C^{0,1}_{K_u}(T)$ with $K_u \coloneqq K_\varphi K_y^{\eta M}$.

   should be replaced by: Consider the truncated function $\bar{y}^{\eta M} \in L_2(T)$ and $\bar{u} = \varphi(\bar{y}^{\eta M}) \in L_2(T)$ for some $T \geq 0$.

6. **The following text on pages 238 – 239:** Assuming $\bar{y}^{\eta M} \in C^{0,1}_{K_y^{\eta M}}(T)$ (this property is proven in Step (ii) below), the inequality $|\left(\delta \bar{y}^{\eta M}\right)_{t,\tau}| \leq K_y^{\eta M}|\tau|, \tau = T/(2m)$ can be used in (C.4) to write:

$$\left|\bar{U}\left[m\right]\right| \leq \frac{K_\varphi K_y^{\eta M}}{4\left|m\right|} \int_0^T \mathrm{d}t = \frac{T K_\varphi K_y^{\eta M}}{4\left|m\right|}, m \neq 0. \tag{C.5}$$

   (ii) Properties of the nonlinearity input signal. Next, we derive an upper bound for the constant $K_y^{\eta M}$, proving that the steady-state signal $\bar{y}^{\eta M}$ is in the space $C^{0,1}_{K_y^{\eta M}}(T)$. Consider the absolute value of the mismatch:

$$\left|\left(\delta \bar{y}^{\eta M}\right)_{t,\tau}\right| \leq \left|g_{yu}\left(\delta \bar{u}^{\eta M}\right)_{t,\tau}\right| + \left|g_{yw}\left(\delta w_0^{\eta M}\right)_{t,\tau}\right|. \tag{C.6}$$

   The frequency-domain counterparts of the steady-state operators g_{yu} and g_{yw} are defined in (4.6a) and (4.6b), respectively. It is well-known that (see, e.g., [196])

$$\left|g_{yu}a\right| \leq \gamma_{yu}^{\eta M}\left|a\right|, \quad \left|g_{yw}b\right| \leq \gamma_{yw}^{\eta M}\left|b\right|, \tag{C.7}$$

   where $\gamma_{yu}^{\eta M}$ and $\gamma_{yw}^{\eta M}$ are defined in (4.7) and $a = \left(\delta \bar{u}^{\eta M}\right)_{t,\tau}, b = \left(\delta w_0^{\eta M}\right)_{t,\tau}$. Inequality (C.7) holds for any $t, \tau \in \mathbb{R}$, where $|g_{yu}a|$ is the evaluation at t, τ of the absolute value of the function resulting from operating g_{yu} on a. The inequalities in (C.7) can be used in (C.6) together with Condition ***C2*** of Theorem 4.1 (by Assumption 4.1) to write:

$$\left|\left(\delta \bar{y}^{\eta M}\right)_{t,\tau}\right| \leq \gamma_{yu}^{\eta M} K_\varphi \left|\left(\delta \bar{y}^{\eta M}\right)_{t,\tau}\right| + \gamma_{yw}^{\eta M}\left|\left(\delta w_0^{\eta M}\right)_{t,\tau}\right|.$$

   Collecting terms and using the constant $K_{w_0}^{\eta M}$ of Assumption 4.2 results in:

$$\left|\left(\delta \bar{y}^{\eta M}\right)_{t,\tau}\right| \leq \frac{\gamma_{yw}^{\eta M} K_{w_0}^{\eta M}}{1 - \gamma_{yu}^{\eta M} K_\varphi}|\tau|, \tag{C.8}$$

   which is well-defined since the denominator $\gamma_{yu}K_\varphi < 1$ by Assumption 4.1. For condition (C.8) to hold, the constant $K_y^{\eta M}$ must satisfy:

$$0 \leq K_y^{\eta M} \leq \frac{\gamma_{yw}^{\eta M} K_{w_0}^{\eta M}}{1 - \gamma_{yu}^{\eta M} K_\varphi} < +\infty, \tag{C.9}$$

which guarantees that $\bar{y}^{\eta M} \in C^{0,1}_{K_y^{\eta M}}(T)$.

(iii) Bound (4.10). Combining the bound (C.9) on $K_y^{\eta M}$ with the bound (C.5) on the magnitude of $\bar{U}$ results in the derived bound in (4.10).

should be replaced by: The square of the magnitude of the Fourier coefficients $\bar{U}$ in (C.4) can be upper bounded as follows for $m \neq 0$:

$$\left|\bar{U}\left[m\right]\right|^2 \leq \frac{K_\varphi^2}{4T^2}\left(\int_0^T \left|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right| \mathrm{d}t\right)^2 \leq \frac{K_\varphi^2}{4T}\int_0^T \left|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right|^2 \mathrm{d}t\,, \tag{C.5}$$

where for functions $f \in L_2(T)$ on a closed interval $[a,\, b]$, the special case

$$\left(\int_a^b f(x)\mathrm{d}x\right)^2 \leq (b-a)\int_a^b \left(f(x)\right)^2 \mathrm{d}x$$

of the Cauchy-Schwarz inequality[1] is used. Note that by definition of the L_2-norm (see the notation in Section 4.2), the inequality in (C.5) can equivalently be written as follows for $m \neq 0$:

$$\left|\bar{U}\left[m\right]\right|^2 \leq \frac{K_\varphi^2}{4}\left\|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right\|_{L_2}^2. \tag{C.6}$$

(ii) Properties of the nonlinearity input signal. The L_2-norm of $\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}$ can be related to the L_2-norm of the truncated input $\left(\delta w_0^{\eta M}\right)_{t,\tau}$ of the Lur'e-type model. Hereto, consider:

$$\begin{aligned}
\left\|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right\|_{L_2} \leq& \left\|g_{yu}\left(\delta\bar{u}^{\eta M}\right)_{t,\tau}\right\|_{L_2} + \left\|g_{yw}\left(\delta w_0^{\eta M}\right)_{t,\tau}\right\|_{L_2} && \text{(C.7a)}\\
\leq& \gamma_{yu}^{\eta M} K_\varphi \left\|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right\|_{L_2} + \gamma_{yw}^{\eta M}\left\|\left(\delta w_0^{\eta M}\right)_{t,\tau}\right\|_{L_2} && \text{(C.7b)}\\
\Rightarrow \left\|\left(\delta\bar{y}^{\eta M}\right)_{t,\tau}\right\|_{L_2} \leq& \frac{\gamma_{yw}^{\eta M}}{1-\gamma_{yu}^{\eta M}K_\varphi}\left\|\left(\delta w_0^{\eta M}\right)_{t,\tau}\right\|_{L_2}, && \text{(C.7c)}
\end{aligned}$$

where the denominator $1-\gamma_{yu}^{\eta M}K_\varphi > 0$ by Assumption 4.1. In turn, the L_2-norm of $\left(\delta w_0^{\eta M}\right)_{t,\tau}$ can be upper bounded as follows:

$$\left\|\left(\delta w_0^{\eta M}\right)_{t,\tau}\right\|_{L_2}^2 = \frac{1}{T}\int_0^T \left|\left(\delta w_0^{\eta M}\right)_{t,\tau}\right|^2 \mathrm{d}t \leq \left(K_{w_0}^{\eta M}\right)^2 |\tau|^2, \tag{C.8}$$

[1] See Theorem 15.25 of K.G. Binmore. *Mathematical analysis: a straightforward approach*, Cambridge University Press, UK, 1977.

where the Lipschitz constant $K_{w_0}^{\eta M}$ of the input signal $w_0^{\eta M}$ with respect to time is used (as in Assumption 4.2). Using (the square root of) (C.8) in (C.7c) results in

$$\left\| \left(\delta \bar{y}^{\eta M}\right)_{t,\tau} \right\|_{L_2} \leq \frac{\gamma_{yw}^{\eta M} K_{w_0}^{\eta M}}{1 - \gamma_{yu}^{\eta M} K_\varphi} |\tau|.$$

(iii) Bound (4.10). Combining the latter bound with the (square root of the) bound (C.6) on the magnitude of $\bar{U}$ results in

$$\left|\bar{U}\,[m]\right| \leq \frac{K_\varphi K_{w_0}^{\eta M} \gamma_{yw}^{\eta M}}{2(1 - \gamma_{yu}^{\eta M} K_\varphi)} |\tau|, \tag{C.9}$$

for $m \neq 0$. After substitution of $\tau = T/(2m)$, the bound in (4.10) is found.

7. **The following text on pages 243 – 244:** Taking similar steps as in the proof of Lemma 4.1 steps (i - iii), we can show that

$$|\breve{\bar{Z}}[m]| \leq \frac{T K_z}{4|m|}, \tag{C.26}$$

where $0 \leq K_z < +\infty$ is such that $\breve{\bar{z}}_F \in C_{K_z}^{0,1}(T)$. Let us find an upper bound for K_z next by analyzing the following mismatch:

$$\begin{aligned} \left|(\delta \breve{\bar{z}})_{t,\tau}\right| &\leq \left|(\delta v_z)_{t,\tau}\right| + \left|g_{zu}(\delta \breve{\bar{u}})_{t,\tau}\right| + \left|g_{zw}(\delta w_0)_{t,\tau}\right| + \left|g_{zw}(\delta v_w)_{t,\tau}\right|, && \text{(C.27a)} \\ &\leq \left|(\delta v_z)_{t,\tau}\right| + \gamma_{zu}\left|(\delta \breve{\bar{u}})_{t,\tau}\right| + \gamma_{zw}\left|(\delta w_0)_{t,\tau}\right| + \gamma_{zw}\left|(\delta v_w)_{t,\tau}\right|. && \text{(C.27b)} \end{aligned}$$

We use the fact that $\breve{\bar{u}} = \varphi(\breve{\bar{y}})$ to write:

$$\left|(\delta \breve{\bar{u}})_{t,\tau}\right| \leq K_\varphi \left|(\delta \breve{\bar{y}})_{t,\tau}\right|, \tag{C.28}$$

where

$$\left|(\delta \breve{\bar{y}})_{t,\tau}\right| \leq \gamma_{yu}\left|(\delta \breve{\bar{u}})_{t,\tau}\right| + \gamma_{yw}\left|(\delta w_0)_{t,\tau}\right| + \gamma_{yw}\left|(\delta v_w)_{t,\tau}\right|.$$

Substituting (C.28) in (C.27b) results in

$$\left|(\delta \breve{\bar{z}})_{t,\tau}\right| \leq \left|(\delta v_z)_{t,\tau}\right| + \gamma_z\left|(\delta w_0)_{t,\tau}\right| + \gamma_z\left|(\delta v_w)_{t,\tau}\right|$$

with

$$\gamma_z \coloneqq \gamma_{zw} + \frac{\gamma_{zu}\gamma_{yw}}{1 - \gamma_{yu}K_\varphi} \tag{C.29}$$

being well-defined since $\gamma_{yu}K_\varphi < 1$ by Assumption 4.1. Next, we use the constants $K_{w_0}^\circ$ (in Assumption 4.2) and K_{v_z}, K_{v_w} (in Assumption 4.5) to formulate: $\left|(\delta \breve{\bar{z}})_{t,\tau}\right| \leq \left(K_{v_z} + \gamma_z\left(K_{w_0}^\circ + K_{v_w}\right)\right)|\tau|$, which can only be true if

$$0 \leq K_z \leq K_{v_z} + \gamma_z\left(K_{w_0}^\circ + K_{v_w}\right), \tag{C.30}$$

proving that $\breve{\breve{z}}_F \in C^{0,1}_{K_z}(T)$.

should be replaced by: Taking similar steps as in the proof of Lemma 4.1 steps (i - ii), we can show that, for $m \neq 0$, the following inequality holds true

$$|\breve{\breve{Z}}[m]|^2 \leq \frac{1}{4}\big\|(\delta\breve{\breve{z}})_{t,\tau}\big\|^2_{L_2}. \tag{C.26}$$

Let us find an upper bound for the L_2-norm of $(\delta\breve{\breve{z}})_{t,\tau}$ as follows:

$$\begin{aligned}\big\|(\delta\breve{\breve{z}})_{t,\tau}\big\|_{L_2} &\leq \big\|(\delta v_z)_{t,\tau} + g_{zu}(\delta\breve{\breve{u}})_{t,\tau} + g_{zw}\left((\delta w_0)_{t,\tau} + (\delta v_w)_{t,\tau}\right)\big\|_{L_2} \\ &\leq \|(\delta v_z)_{t,\tau}\|_{L_2} + \gamma_{zu}\big\|(\delta\breve{\breve{u}})_{t,\tau}\big\|_{L_2} \\ &\qquad + \gamma_{zw}\|(\delta w_0)_{t,\tau}\|_{L_2} + \gamma_{zw}\|(\delta v_w)_{t,\tau}\|_{L_2}.\end{aligned} \tag{C.27}$$

By the incremental gain of the nonlinearity φ (Condition ***C2*** of Theorem 4.1 (by Assumption 4.1)), the inequality

$$\big\|(\delta\breve{\breve{u}})_{t,\tau}\big\|_{L_2} \leq K_\varphi\big\|(\delta\breve{\breve{y}})_{t,\tau}\big\|_{L_2}$$

can be used together with the inequality

$$\big\|(\delta\breve{\breve{y}})_{t,\tau}\big\|_{L_2} \leq \gamma_{yu}\big\|(\delta\breve{\breve{u}})_{t,\tau}\big\|_{L_2} + \gamma_{yw}\|(\delta w_0)_{t,\tau}\|_{L_2} + \gamma_{yw}\|(\delta v_w)_{t,\tau}\|_{L_2}$$

to write

$$\big\|(\delta\breve{\breve{u}})_{t,\tau}\big\|_{L_2} \leq \frac{K_\varphi\gamma_{yw}}{1 - K_\varphi\gamma_{yu}}\left(\|(\delta w_0)_{t,\tau}\|_{L_2} + \|(\delta v_w)_{t,\tau}\|_{L_2}\right). \tag{C.28}$$

Substituting (C.28) in (C.27) results in

$$\big\|(\delta\breve{\breve{z}})_{t,\tau}\big\|_{L_2} \leq \|(\delta v_z)_{t,\tau}\|_{L_2} + \gamma_z\|(\delta w_0)_{t,\tau}\|_{L_2} + \gamma_z\|(\delta v_w)_{t,\tau}\|_{L_2}$$

with

$$\gamma_z \coloneqq \gamma_{zw} + \frac{K_\varphi\gamma_{zu}\gamma_{yw}}{1 - \gamma_{yu}K_\varphi} \tag{C.29}$$

being well-defined since $\gamma_{yu}K_\varphi < 1$ by Assumption 4.1. Next, using the constants $K^\circ_{w_0}$ (in Assumption 4.2) and K_{v_z}, K_{v_w} (in Assumption 4.5) allows writing $\big\|(\delta\breve{\breve{z}})_{t,\tau}\big\|_{L_2} \leq K_z|\tau|$ with $K_z \coloneqq K_{v_z} + \gamma_z\left(K^\circ_{w_0} + K_{v_w}\right)$. Using this inequality in (the square root of) (C.26) and substituting $\tau = T/(2m)$ results in

$$|\breve{\breve{Z}}[m]| \leq \frac{TK_z}{4|m|}, \quad \forall\, m \neq 0. \tag{C.30}$$