

A greedy data collection scheme for linear dynamical systems

Karim Cherifi¹, Pawan Goyal^{2,*}  and Peter Benner²

¹Institut für Mathematik, TU Berlin, MA 4-5, Straße des 17. Juni 136, Berlin D-10623, Germany

²Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, Magdeburg 39106, Germany

*Corresponding author. E-mail: goyalp@mpi-magdeburg.mpg.de

Received: 07 September 2021; Revised: 24 December 2021; Accepted: 21 March 2022

Keywords: Dynamical systems; Loewner framework; machine learning; system identification

Abstract

Mathematical models are essential to analyze and understand the dynamics of complex systems. Recently, data-driven methodologies have gotten a lot of attention which is leveraged by advancements in sensor technology. However, the quality of obtained data plays a vital role in learning a good and reliable model. Therefore, in this paper, we propose an efficient heuristic methodology to collect data both in the frequency domain and the time domain, aiming at having more information gained from limited experimental data than equidistant points. In the frequency domain, the interpolation points are restricted to the imaginary axis as the transfer function can be estimated easily on the imaginary axis. The efficiency of the proposed methodology is illustrated by means of several examples, and its robustness in the presence of noisy data is shown.

Impact Statement


Mathematical modeling describing underlying dynamical behaviors plays essential roles in performing engineering studies such as optimization and control. In Complex processes where constructing models from first principles is more challenging, data-driven dynamical models become a valuable tool. The success of data-driven approaches highly relies on the quality of the data. Unlike typical machine learning applications, for example, image classification, data in engineering applications are somewhat limited. Hence, a design of experiments to collect data should be done with care. This paper presents a scheme to collect data with the aim of (heuristically) maximizing the extraction of information with limited measurements. The scheme can be wrapped on top of any existing data-driven modeling methods in a setting.

1. Introduction

In this paper, our focus is on inferring linear time-invariant systems with m inputs and p outputs of the form:

$$\begin{aligned} \mathbf{E}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t), \end{aligned} \quad (1)$$

from data where $\mathbf{E}, \mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, $\mathbf{C} \in \mathbb{R}^{p \times n}$, and $\mathbf{D} \in \mathbb{R}^{p \times m}$. Using the Laplace transform, the transfer function $\mathbf{H}(s)$ of the system (1) can be obtained and is given as

 This research article was awarded an Open Data badge for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

$$\mathbf{H}(s) = \mathbf{C}(s\mathbf{E} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}, \quad (2)$$

where s represents the Laplace domain variable. Although the results presented in this paper can be applied to multi-input multi-output (MIMO) systems, for simplicity, we will focus only on single-input single-output (SISO) systems. The problem of learning dynamical systems of the form (1) is widely studied and well understood (see, e.g., Mayo and Antoulas, 2007; Drmač et al., 2015a, 2015b; Peherstorfer and Willcox, 2016; Antoulas et al., 2017; Nakatsukasa et al., 2018; Brunton and Kutz, 2019; Yu et al., 2019). A key ingredient to data-driven methodologies is data. It is necessary to provide a good data quality, ensuring high fidelity of the inferred model. Typically, there are two possible ways to collect data, namely in frequency domain and time domain. In the frequency domain, transfer function samples are given, whereas, in the time domain, the output \mathbf{y} is measured for a given input \mathbf{u} . Moreover, these data may be subject to noise. Depending on an experimental setup, we may collect data in either of these domains.

Among several widely used approaches, Loewner-based approaches (Mayo and Antoulas, 2007; Antoulas et al., 2017) to learn dynamical systems have gained increasing popularity. The fundamental idea of the Loewner-based approach is to construct a model that interpolates the given data points in the frequency domain. It has also been extended to time-domain data (Peherstorfer et al., 2017). Furthermore, Antoulas et al. (2017) have shown that if the minimum order of a system that realizes the system is n , then using $2n$ data points, one can construct the realization using the Loewner approach. However, if only data are given, we cannot know a priori the minimal order of a state-space realization that captures the dynamics of the process. Therefore, typically, one aims at collecting $N \gg 2n$ data points, which is then followed by a compression step to determine the minimal order of the state-space model that captures the dynamics of the process. We refer the reader to Mayo and Antoulas (2007) for more details. However, this approach can impose practical constraints if collecting data are expensive. In this case, we aim to gather data points carefully and focus on collecting the data points to obtain as much information (gathered on the imaginary axis) as possible about the underlying dynamics up to tolerance. In a step toward collecting good data, Beattie and Gugercin (2012) presented an approach inspired by iterative rational Krylov algorithm (IRKA) (Gugercin et al., 2008) that focused on using interpolation points that facilitate obtaining a model, minimizing the \mathcal{H}_2 error between the learned model and the unknown ground-truth model. However, again, the approach needs the order of the state-space realization as input, which may not be known in advance. In addition, it is an iterative scheme; thus, the number of transfer function evaluations still becomes significantly large. Furthermore, we mention that there also exist other approaches that aim at determining or collecting measurements greedily but require a ground-truth high-fidelity model (see, e.g., Paul-Dubois-Taine and Amsallem, 2015; Chellappa et al., 2020; Chellappa et al., 2021; Beddig et al., 2022).

On the other hand, our goal is to construct a minimal realization of a process using as few measurement data points as possible up to a tolerance, where we neither know the order of a minimal realization in advance nor have access to any high-fidelity model. Furthermore, one can theoretically consider taking the transfer function measurements on the complex domain, and often, it is known that good, in fact optimal, measurement points are complex as shown in Gugercin et al. (2008) and Beattie and Gugercin (2012). However, from an experimental viewpoint and physical interpretation of a transfer function, we assume that transfer function measurements are taken anywhere on the imaginary $j\omega$ -axis. The motivation of considering measurements on the $j\omega$ -axis are:

- (a) A transfer function at $j\omega$ can be estimated by exciting the system using an input containing sine and cosine functions of the frequency ω . It can be extended to multiple frequencies as well.
- (b) From the control theory perspective, a transfer function can be better interpreted on the $j\omega$ -axis, for example, using the Bode plot. More details can be found in, for example, Ogata (2010).

In this paper, we focus on inferring minimal linear dynamical systems using minimum possible measurement data taken on the imaginary axis up to tolerance, but without knowing the minimal order

of the realization a priori and its realization in any form. For this, we discuss a heuristic approach, allowing us to select adaptively frequency points positioned where the transfer function changes the most. We also discuss the extension of the approach to time-domain data.

The paper is structured as follows. In Section 2, the Loewner framework to learn linear dynamical systems is briefly presented. In Section 3, we present the heuristic approach to select frequency points from the imaginary axis (and transfer function measurements) that contributes the most in reducing the error of the obtained realization. Then, in Section 4, we discuss an extension of the approach when input–output data are collected in the time domain. In Section 5, we illustrate the efficiency of the proposed greedy methodology, and compare it with models that are obtained from IRKA and by taking measurements on a uniform grid. Finally, we conclude the paper with a summary and future research in Section 6.

2. Data-Driven Modeling Based on the Loewner Framework

In this section, we briefly recap the Loewner approach (Mayo and Antoulas, 2007) to construct a realization from given transfer function measurements. To that end, for simplicity, we first write down the underlying realization problem for SISO systems—that is as follows.

Problem 2.1. *Given a set of interpolation points $Z := \{\sigma_1, \dots, \sigma_{2N}\}$ and corresponding transfer function measurements $\mathbf{H}(\sigma_i)$, the goal is to identify a minimal realization whose rational transfer function is denoted by $\widehat{\mathbf{H}}(s)$, satisfying the following interpolation conditions:*

$$\mathbf{H}(\sigma_i) = \widehat{\mathbf{H}}(\sigma_i), \quad i \in \{1, \dots, 2N\}, \tag{3}$$

and the transfer function is given as

$$\widehat{\mathbf{H}}(s) = \widehat{\mathbf{C}} \left(s\widehat{\mathbf{E}} - \widehat{\mathbf{A}} \right)^{-1} \widehat{\mathbf{B}} + \widehat{\mathbf{D}}, \tag{4}$$

where $\widehat{\mathbf{E}}, \widehat{\mathbf{A}} \in \mathbb{R}^{r \times r}$, $\widehat{\mathbf{B}} \in \mathbb{R}^r$, $\widehat{\mathbf{C}} \in \mathbb{R}^r$, and $\widehat{\mathbf{D}} \in \mathbb{R}$ with r being the order of a minimal realization.

This realization problem can be solved using the Loewner approach for given $\widehat{\mathbf{D}}$. For this, we need to split the measurement data into left and right measurements. Let us define the left and right interpolation points as λ_j and μ_j , where $j \in \{1, \dots, N\}$, corresponding transfer function measurements are denoted by $\mathbf{H}(\lambda_j)$ and $\mathbf{H}(\mu_j)$, and $Z = \{\lambda_1, \dots, \lambda_N\} \cup \{\mu_1, \dots, \mu_N\}$. Having these data, in the following, we define the Loewner and shifted Loewner matrices.

Definition 2.1 (Mayo and Antoulas, 2007). *Given left data $(\lambda_j, \mathbf{H}(\lambda_j))$ and right data $(\mu_j, \mathbf{H}(\mu_j))$, the Loewner \mathbb{L} and shifted Loewner \mathbb{L}_s matrices are defined as*

$$\mathbb{L} = \begin{bmatrix} \frac{\mathbf{H}(\lambda_1) - \mathbf{H}(\mu_1)}{\lambda_1 - \mu_1} & \dots & \frac{\mathbf{H}(\lambda_n) - \mathbf{H}(\mu_1)}{\lambda_n - \mu_1} \\ \vdots & \ddots & \vdots \\ \frac{\mathbf{H}(\lambda_1) - \mathbf{H}(\mu_n)}{\lambda_1 - \mu_n} & \dots & \frac{\mathbf{H}(\lambda_n) - \mathbf{H}(\mu_n)}{\lambda_n - \mu_n} \end{bmatrix}, \quad \mathbb{L}_s = \begin{bmatrix} \frac{\lambda_1 \mathbf{H}(\lambda_1) - \mu_1 \mathbf{H}(\mu_1)}{\lambda_1 - \mu_1} & \dots & \frac{\lambda_n \mathbf{H}(\lambda_n) - \mu_1 \mathbf{H}(\mu_1)}{\lambda_n - \mu_1} \\ \vdots & \ddots & \vdots \\ \frac{\lambda_1 \mathbf{H}(\lambda_1) - \mu_n \mathbf{H}(\mu_n)}{\lambda_1 - \mu_n} & \dots & \frac{\lambda_n \mathbf{H}(\lambda_n) - \mu_n \mathbf{H}(\mu_n)}{\lambda_n - \mu_n} \end{bmatrix}.$$

It is proved in Mayo and Antoulas (2007) that based on the Loewner and shifted Loewner matrices, a realization can be constructed that solves Problem 2.1. This result is presented in the following theorem.

Theorem 1 (Mayo and Antoulas, 2007). *Given left data $(\lambda_j, \mathbf{H}(\lambda_j))$ and right data $(\mu_j, \mathbf{H}(\mu_j))$, consider the Loewner and shifted Loewner matrices as defined in Definition 2.1. Moreover, assume that the feed-through term \mathbf{D} is known. Then, an interpolating realization can be constructed as below:*

$$\widehat{\mathbf{E}} = \mathbb{L}, \quad \widehat{\mathbf{A}} = \mathbb{L}_s - \mathbf{1D}\mathbf{1}^T, \quad \widehat{\mathbf{B}} = \mathbf{V} - \mathbf{1D}, \quad \widehat{\mathbf{C}} = \mathbf{W} - \mathbf{D}\mathbf{1}^T, \quad \widehat{\mathbf{D}} = \mathbf{D}, \tag{5}$$

where $\mathbf{1} \in \mathbb{R}^N$ is the vector of ones, $\widehat{\mathbf{E}}, \widehat{\mathbf{A}} \in \mathbb{R}^{n \times n}$, $\widehat{\mathbf{B}} \in \mathbb{R}^n$, $\widehat{\mathbf{C}} \in \mathbb{R}^n$, $\widehat{\mathbf{D}} \in \mathbb{R}$, $\mathbf{V} = [\mathbf{H}(\lambda_1), \dots, \mathbf{H}(\lambda_n)]^T$, and $\mathbf{W} = [\mathbf{H}(\mu_1), \dots, \mathbf{H}(\mu_n)]$. Furthermore, the realization (5) is minimal, assuming \mathbb{L} is of full-rank.

Furthermore, in case the Loewner matrix \mathbb{L} is singular, there exists a lower-order realization of order $\widehat{n} < n$, where $\widehat{n} = \text{rank}(\mathbb{L})$, satisfying the interpolation conditions. To remove the redundant information and obtain the minimal realization, a compression step based on the singular-value decomposition (SVD) of $[\mathbb{L}, \mathbb{L}_s]$ can be employed (see, e.g., Mayo and Antoulas, 2007; Antoulas et al., 2017). We consider here only the minimal realization for systems with an underlying ODE. For high index descriptor systems, one has to be more careful with the definition of minimality (see, e.g., Sokolov, 2006), which might be interesting to investigate as well in the future in detail. For simplicity, we have presented the realization theory using the Loewner approach for SISO systems; however, it can be extended to MIMO systems using the idea of tangential interpolation. We refer to Mayo and Antoulas (2007) and Antoulas et al. (2017) for more details. We also remark that the realization (5) can be complex if data are complex, but if the data are closed under conjugation, then there exists an orthogonal transformation that allows determining the real realization using that orthogonal transformation.

The success of the Loewner approach has been shown in various applications (see, e.g., Ionita and Antoulas, 2014a, 2014b; Poussot-Vassal et al., 2021). Its success lies in the quality of data and assumes that there are enough data for a wide range of frequencies. However, suppose that the data collection process is expensive. In this case, it is essential to make a smart choice to measurements to maximize information about the system with limited data. With this aim, in the following section, we discuss an adaptive scheme to collect measurements to extract more information about the system, as compared to the standard equidistant measurements on the imaginary axis, with a fixed number of measurements.

3. A Greedy-Based Data Collection Scheme

Naturally, the quality of collected data is a key ingredient to learning a good and reliable model, describing the underlying dynamics of a process. Obviously, one can collect as much data as possible if feasible. It can then be followed by inferring a system realization by using, for example, the Loewner approach (Mayo and Antoulas, 2007) as discussed in the previous section. However, a mass collection could be expensive in many scenarios. Therefore, in this section, we discuss a greedy scheme that can guide us to collect the data, so that we can expect to extract as much new information about the dynamical system from every measurement taken on the imaginary axis as possible. In the following, we first note down a corresponding problem.

Problem 3.1. *Identify the underlying state-space dynamical system of a process using a minimal number of transfer function evaluations restricted to the imaginary axis up to a user-defined tolerance.*

To that end, we first assume that we have an initial rough construct of a realization that still can be far from satisfactory. An initial rough model can be constructed using only a few measurement points. We denote the transfer function of the initial realization by $\mathbf{H}_{\text{init}}(s)$. Furthermore, let us denote the transfer function of the unknown ground truth realization by $\mathbf{H}_{\text{true}}(s)$. In this case, we ideally would like to update the model using the point on the $j\omega$ -axis, where the maximum error between the true and initial realized systems occurs; hence, we select the interpolation point that solves the following optimization problem:

$$\sigma_{\text{new}} := \arg \max_{\sigma \in j\omega} \|\mathbf{H}_{\text{true}}(\sigma) - \mathbf{H}_{\text{init}}(\sigma)\|, \tag{6}$$

where $\omega \in \mathbb{R}$. However, to solve the optimization problem (6), we require the transfer function of the ground-truth realization, which is not available. Therefore, we seek for an alternative approach, allowing us to estimate good interpolation points in an iterative procedure. For this, let us assume that we have $2k$ measurement points, to begin with, and denote these measurements by tuples $(\sigma_i, \mathbf{H}(\sigma_i))$, $i \in \{1, \dots, 2k\}$,

where σ_i and $\mathbf{H}(\sigma_i)$ are the interpolation points and the transfer function measurements at these points. Before we proceed further, we define the following notation:

- $\Sigma_i := \{\sigma_1, \dots, \sigma_{2i}\}$.
- The transfer function of the realization constructed using interpolation points Σ_m is denoted by $\mathbf{H}_m(s)$, and the order of the constructed realization is m .

Next, we construct realizations using interpolation points $\Sigma_{k-1} \subset \Sigma_k$ and Σ_k , which are, respectively, denoted by $\mathbf{H}_{k-1}(s)$ and $\mathbf{H}_k(s)$. As discussed earlier, we ideally would add a new measurement for the next $\mathbf{H}_{k+1}(s)$ realization at the frequency that solves the problem (6). Since we do not know \mathbf{H}_{true} , in the following, we discuss a heuristic approach to relax the optimization problem. First note that

$$\begin{aligned} \|\mathbf{H}_{\text{true}}(\sigma) - \mathbf{H}_k(\sigma)\| &= \|\mathbf{H}_{\text{true}}(\sigma) - \mathbf{H}_{k-1}(\sigma) - (\mathbf{H}_k(\sigma) - \mathbf{H}_{k-1}(\sigma))\| \\ &\leq \|\mathbf{H}_{\text{true}}(\sigma) - \mathbf{H}_{k-1}(\sigma)\| + \|\mathbf{H}_k(\sigma) - \mathbf{H}_{k-1}(\sigma)\|. \end{aligned} \tag{7}$$

With unknown $\mathbf{H}_{\text{true}}(\sigma)$, we rather focus on defining our next measurements data based on the second part of (7) with a constraint; that is, we should exclude the regime of already taken measurements points. Hence, we solve an optimization problem to obtain a frequency point at which transfer function measurement needs to be taken to obtain more information about the underlying dynamics:

$$\max_{\sigma \in j\omega} \mathbf{g}(\sigma) \|\mathbf{H}_k(\sigma) - \mathbf{H}_{k-1}(\sigma)\|, \tag{8}$$

where the function $\mathbf{g}(\sigma)$ can be thought of as a mask that aims at excluding the regime of points at which measurements have already been collected. If $\mathbf{g}(\sigma)$ is known, then we can solve (8) to obtain our new measurement points which possibly bring the most information about the system as they would occur at the maximum change in the transfer function from the previous to the next step. It is worth noting that $\|\mathbf{H}_k(\sigma) - \mathbf{H}_{k-1}(\sigma)\|$ could be replaced by an indicator that is cheaper to evaluate than sampling the system and that would reduce the cost of the optimization process (8).

Naturally, the choice of the function $\mathbf{g}(\cdot)$ in (8) plays an important role in determining next measurement points. Therefore, in the following, we discuss a choice of the mask function $\mathbf{g}(\sigma)$. As discussed, a choice of the mask function should be such that it excludes the regime of already considered data points. Among many possible choices, here, we propose the following choice:

$$\mathbf{g}(\sigma) = \prod_{i=1}^{2k} \tilde{\mathbf{g}}(\sigma, \sigma_i),$$

where $\tilde{\mathbf{g}}(\sigma, \sigma_i)$ is defined as

$$\tilde{\mathbf{g}}(\sigma, \sigma_i) = 1 - e^{(-\beta(\log(|\sigma|+\varepsilon) - \log(|\sigma_i|+\varepsilon))^2)}, \tag{9}$$

with $\sigma_i, i \in \{1, \dots, 2k\}$, being already considered measurement points, and ε and β are positive constants and hyperparameters. It can be noticed that $\mathbf{g}(\sigma)$ is zero at all the considered points with small values in their neighborhood in order to favor the exploration over the already chosen interpolation points. The mask function can also be seen as a notch filter at multiple frequencies and β can be seen as the bandwidth of the filter. To illustrate the mask function, we plot the function for $\beta = 0.6$, $\varepsilon = 10^{-15}$, and the notch frequencies $\sigma_i = \{10^{-1}, 10^1, 10^3\}$ in Figure 1. It clearly shows that the mask takes smaller values near the considered frequency points.

Next, we discuss how to further simplify the choice of the function $\mathbf{g}(\sigma)$. First, note that $\mathbf{H}_k(\sigma_i) = \mathbf{H}_{k-1}(\sigma_i)$, for $i \in \{1, \dots, 2k - 2\}$, assuming that the realization is determined by the Loewner approach, and thus, the interpolation at the given measurement points is exact. As a result, we can also consider the optimization problem to determine our next measurement point as follows:

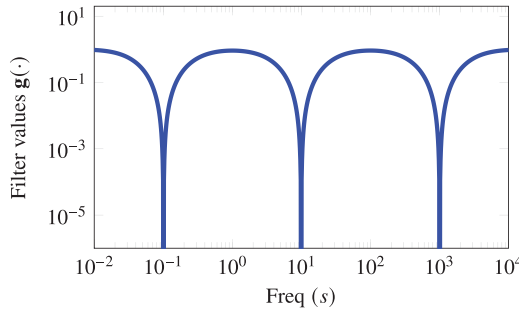


Figure 1. A visual illustration of the filter when the measurements are already taken at the frequencies $\sigma_i = \{10^{-1}, 10^1, 10^3\}$.

$$\sigma_{2k+1} := \arg \max_{\sigma \in j\omega} \prod_{i=2k-1}^{2k} \tilde{g}(\sigma, \sigma_i) \| \mathbf{H}_k(\sigma) - \mathbf{H}_{k-1}(\sigma) \|. \tag{10}$$

In this paper, we utilize the Loewner approach to construct the realization. Therefore, it is preferred to have an even number of measurement points to avoid the construction of a rectangular and imaginary realization. Hence, it is desired to include two new additional points at each step. Consequently, in order to choose one additional point, we solve the following optimization problem once σ_{2k+1} is obtained:

$$\sigma_{2k+2} := \arg \max_{\sigma \in j\omega} \prod_{i=2k-1}^{2k+1} \tilde{g}(\sigma, \sigma_i) \| \mathbf{H}_k(\sigma) - \mathbf{H}_{k-1}(\sigma) \|. \tag{11}$$

As a result, we have two new frequency points σ_{2k+1} and σ_{2k+2} at which we take transfer function measurements. Thus, we update our realization \mathbf{H}_{k+1} by using measurements at $\sigma_i, i \in \{1, \dots, 2k+2\}$. In the SISO case, the norm between the transfer function measurements just becomes the absolute value. The whole procedure is repeated until the error between two iterations is small enough, that is,

$$\max_{s \in j\omega} \| \mathbf{H}_k(s) - \mathbf{H}_{k-1}(s) \| \leq \tau \circ \mathbb{1}. \tag{12}$$

We sketch the whole procedure in Algorithm 1. When the condition (12) is met, the algorithm terminates as we have obtained the minimal realization of the system with respect to the given tolerance.

Moreover, we point out that the algorithm is not sensitive to small changes in the input hyperparameters. For example, a small change in β in the mask function will not significantly impact the convergence of the algorithm as long as the mask function keeps the properties described above. This is discussed more in the numerical section. The other hyperparameter $\tau \circ \mathbb{1}$ needs to be set by the user as a trade-off between the accuracy of the resulting model and the number of interpolation points used.

Algorithm 1. A greedy selection of measurement points in the frequency domain.

Input: A parameter β , and initial measurement points $\Sigma_k := \{\sigma_1, \dots, \sigma_{2k}\}$.

- 1 Determine realizations \mathbf{H}_{k-1} and \mathbf{H}_k by using measurement points $\Sigma_{k-1} \subset \Sigma_k$ and Σ_k , respectively, and by employing the Loewner approach.
- 2 Compute $\text{E r r} := \max_{s \in j\omega} \| \mathbf{H}_k(s) - \mathbf{H}_{k-1}(s) \|$.

```

3 while Err > tol do
4   Compute the new measurement point  $\sigma_{2k+1}$  according to the optimization problem (10).
5   Deduce  $\sigma_{2k+2}$  according to (11).
6   Estimate the transfer function at the measurement points  $\sigma_{2k+1}$  and  $\sigma_{2k+2}$ .
7   Determine a realization using measurement points at  $\Sigma_k \cup \{\sigma_{2k+1}, \sigma_{2k+2}\}$  using the
   Loewner approach.
8    $k \leftarrow k + 1$ .
9   Compute  $\text{Err} := \max_{s \in j\omega} \|\mathbf{H}_k(s) - \mathbf{H}_{k-1}(s)\|$ .

```

Output: A learned model, whose transfer function is given by \mathbf{H}_k .

4. Extension to Time-Domain Data

In the previous section, we have discussed a greedy approach for frequency-domain measurements. Here, we suppose that we have access to the time-domain data instead of the frequency-domain data. In this case, we need to design input to collect time-domain data, so that we can maximize information extraction about the process. In this case, we can also employ Algorithm 1 with a slight modification. First, note that Peherstorfer et al. (2017) have proposed a methodology to realize a state-space model using time-domain data, where frequency-domain measurements are estimated using time-domain data. This is followed by obtaining a realization using the Loewner framework described in Section 2. However, the methodology heavily depends on the choice of input, and the choice should be made in such a way that it allows us to describe the dynamics of the system completely. Hence, in this section, we discuss a suitable choice of inputs. Using an extension of Algorithm 1, one can adaptively choose frequency points composing an input. In practice, time-domain measurements are typically collected at a regular interval; hence, the measurements are discrete. Therefore, the proposed greedy procedure needs to be adapted in a discrete setting. In what follows, we discuss the adaptation that needs to be made to allow us to design an input to extract as much information as possible. Note that the interpolation points are no longer on the s -plane but are rather on the unit circle of the z -plane. Thus, the frequency range on the $j\omega$ -axis needs to be mapped on the unit circle using an appropriate discretization method. The Cayley transform could be used to map between the discrete-time and continuous-time systems. To obtain time-domain measurements, the system is excited with an input spanning a set of interpolation points, and the system response is measured. We design an input $\mathbf{u}_p^{(k)}$ using a sum of sine and cosine functions, that is, the input in the k th step of the algorithm at time $\mathbf{T}_s p$, where \mathbf{T}_s is a sampling time, and p is a nonnegative integer, can be given in the form:

$$\mathbf{u}_p^{(k)} = \frac{1}{K} \sum_{l=1}^2 (1+j) (\cos(\sigma_{2k+l}) + j \sin(\sigma_{2k+l})), \quad (13)$$

where $j := \sqrt{-1}$, $p \in \{0, \dots, K-1\}$, and K is the number of time-domain measurements.

Following the same strategy as in Section 3, the goal is to estimate frequency-domain data, corresponding transfer function estimates, and construct a realization using the interpolation points. However, since only time-domain data are available, one has first to use these data to estimate the frequency measurements.

The complete procedure is as follows. Given an initial realization $\mathbf{H}_{\text{init}}(s)$, the next two interpolation points σ_{2k+1} and σ_{2k+2} are computed as in (10) and (11). Then, we construct the input $\mathbf{u}_p^{(k)}$ (at step k of the algorithm and time step p) as in (13) with the two interpolation points σ_{2k+1} and σ_{2k+2} . We simulate the system to obtain an output having a value of $\mathbf{y}_p^{(k)}$ at the step k of the algorithm and at time step p . Assembling this time-domain data, one can then compute estimates for the transfer function $\mathbf{H}_{\text{true}}(\sigma_{2k+1})$ and $\mathbf{H}_{\text{true}}(\sigma_{2k+2})$. These values are computed using the following least-squares problem (Peherstorfer et al., 2017):

$$\underset{\widehat{\mathbf{H}} \in \mathbb{C}^r}{\operatorname{arg\,min}} \left\| \mathbf{F}\widehat{\mathbf{H}} - \bar{\mathbf{y}} \right\|_2^2, \tag{14}$$

where $\mathbf{F} \in \mathbb{C}^{(K-k_{\min}) \times 2}$ and $\widehat{\mathbf{H}}$ are described below:

$$\mathbf{F} = \begin{bmatrix} \mathbf{U}_1 e^{j\sigma_{2k+1}k_{\min}} & \mathbf{U}_2 e^{j\sigma_{2k+2}k_{\min}} \\ \vdots & \vdots \\ \mathbf{U}_1 e^{j\sigma_{2k+1}(K-1)} & \mathbf{U}_2 e^{j\sigma_{2k+2}(K-1)} \end{bmatrix}, \text{ and } \widehat{\mathbf{H}} = \begin{bmatrix} \widetilde{\mathbf{H}}(\sigma_{2k+1}) \\ \widetilde{\mathbf{H}}(\sigma_{2k+2}) \end{bmatrix}, \tag{15}$$

where $\widetilde{\mathbf{H}}(\sigma)$ is an estimate of $\mathbf{H}_{\text{true}}(\sigma)$, and the output vector is defined as $\bar{\mathbf{y}} := [\mathbf{y}_{k_{\min}}^{(k)}, \dots, \mathbf{y}_{K-1}^{(k)}]^T$ such that $\mathbf{y}_p^{(k)}$ is the measurement of the output at time step p and the step k of the algorithm. \mathbf{U}_1 and \mathbf{U}_2 are the nonzero (discrete) Fourier transform components of the input \mathbf{u} , given in (13), which will be, in fact, at frequencies σ_{2k+1} and σ_{2k+2} , respectively. Moreover, k_{\min} is chosen such that the system reaches a steady state (approximately) after k_{\min} time steps. For a proof of the derivation of the least-squares problem (14), we refer to the discussion in Peherstorfer et al. (2017).

The next steps are then similar to the steps of Algorithm 1. A realization is constructed using all the estimated points at $\Sigma_k \cup \{\sigma_{2k+1}, \sigma_{2k+2}\}$. The new interpolation points are then computed as in (10) and (11). As in the frequency case, we repeat the procedure until the preset tolerance is reached:

$$\max_{\|z\|=1} \|\mathbf{H}_k(z) - \mathbf{H}_{k-1}(z)\| \leq \tau \circ \perp.$$

5. Numerical Experiments

In this section, we illustrate the proposed methodology by means of three examples. The first two examples, namely, the penzl (Penzl, 2006) and the beam (Chahlaoui and Van Dooren (2002)) examples, consider frequency-domain measurements. In the last example, time-domain data are used to construct a realization for an electrical circuit, consisting of resistors, inductors, and capacitors and denoted by RLC, (Gugercin and Antoulas, 2003) with an adaptive choice of measurements. We compare our proposed approach, where we carefully choose measurement points, with the Loewner approach, where logarithmically equidistant measurement points are considered. In both cases, we choose the best splitting method of right and left interpolation points as discussed in Karachalios et al. (2020). In the following, we also note some details. Typically, we are interested in a frequency range $[\omega_{\min}, \omega_{\max}]$, and we consider M points, on a logarithmic scale, from that given range, denoted by $\mathbf{Q} := [q_1, \dots, q_M]$. Whenever a measurement point is taken, we assume it is taken from the set \mathbf{Q} . Moreover, the Loewner framework (Mayo and Antoulas, 2007) is utilized to construct a realization using measurement points and transfer function evaluations. The initial realization is constructed using six approximately logarithmically equidistant interpolation points taken from the set \mathbf{Q} . This is done by computing first six logarithmically equidistant interpolation points $[\widehat{q}_1, \dots, \widehat{q}_6]$ in the frequency range $[\omega_{\min}, \omega_{\max}]$ and then find the interpolation points $[\widetilde{q}_1, \dots, \widetilde{q}_6]$ that are used in the algorithm by solving the following optimization problem:

$$\underset{q_i \in \mathbf{Q}}{\operatorname{arg\,min}} \|\widetilde{q}_i - \widehat{q}_i\|. \tag{16}$$

In each step of the algorithm, the interpolation points are organized into an interlacing manner (Gosea, 2018). The filter parameters in (9) are set to $\beta = 0.6$ and $\varepsilon = 10^{-15}$. The tolerance in Algorithm 1 is set to $\tau \circ \perp = 10^{-8}$. All the numerical experiments are done on an AMD Ryzen 7 PRO 4750 U processor CPU@1.7GHz, up to 8 MB cache, 16 GB RAM, Ubuntu 20.04 LTS, MATLAB® Version 9.8.0.1323502 (R2020a) 64-bit(glnxa64).

5.1. Penzl example

As a first example, we consider the penzl example (Penzl, 2006) of order $N = 1,006$, which is referred to as the FOM example in the subroutine library in systems and control theory (SLICOT) model reduction

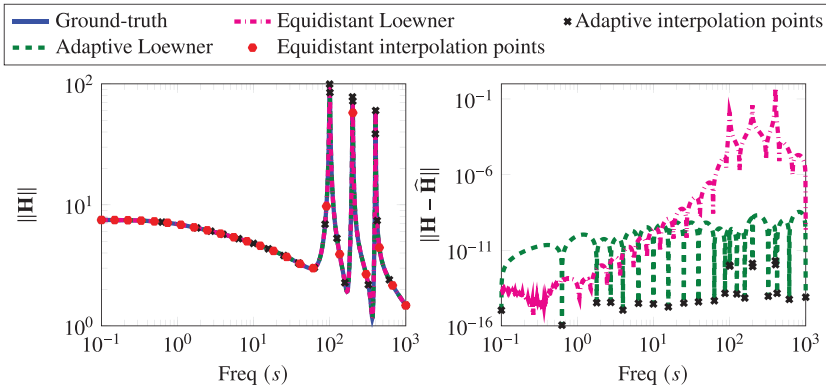


Figure 2. *Penzl example: The Bode plot of the ground-truth, adaptively generated system, and a realization with equidistant points and the corresponding error between them.*

benchmark (Chahlaoui and Van Dooren, 2002). We are interested in the frequency range $[10^{-1}, 10^3]$. Using Algorithm 1, we obtain a realization that on termination uses 24 interpolation points. For comparison, we also construct a realization using 24 logarithmically equidistant interpolation points in the given frequency range. We compare these two learned realizations in Figure 2, where we observe that our greedy scheme to collect measurements clearly outperforms the scheme when measurements are taken equidistantly. Algorithm 1 focuses on the region where the transfer function is involved, and it automatically trends to add more points around the peaks of the transfer function. On the other hand, the realization, constructed using the logarithmically equidistant points, has a larger error as it fails to capture these peaks accurately. Furthermore, in Figure 3, we compare the frequency-limited \mathcal{H}_2 -norm (Wilson, 1970) of the learned and ground-truth systems in terms of the number of interpolation points used. We use the MORLAB software package for the computation of the frequency-limited norm (Benner and Werner, 2021). The figure shows that the frequency-limited \mathcal{H}_2 error decays faster when the measurements are collected adaptively (Algorithm 1).

Next, we compare our method with the IRKA method (Beattie and Gugercin, 2012), as implemented in M-M.E.S.S. (Benner et al., 2021). We highlight that IRKA aims at determining the optimal interpolation points in the \mathcal{H}_2 -norm using an iterative scheme, and these interpolation points can be anywhere in the complex domain. This is in contradiction to our proposed method, where we only focus on the imaginary axis due to practical reasoning. We obtain an optimal model of order 16 using IRKA and compare it with

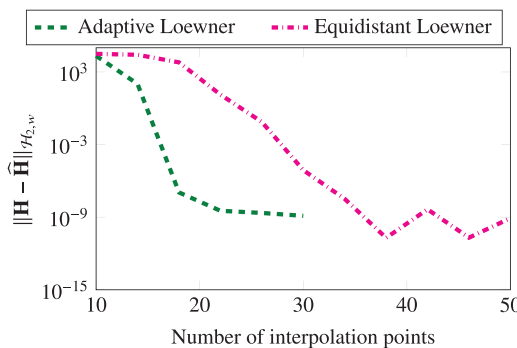


Figure 3. *Penzl example: A comparison between the frequency-limited \mathcal{H}_2 -norm error for the adaptively chosen points and the equidistant points.*

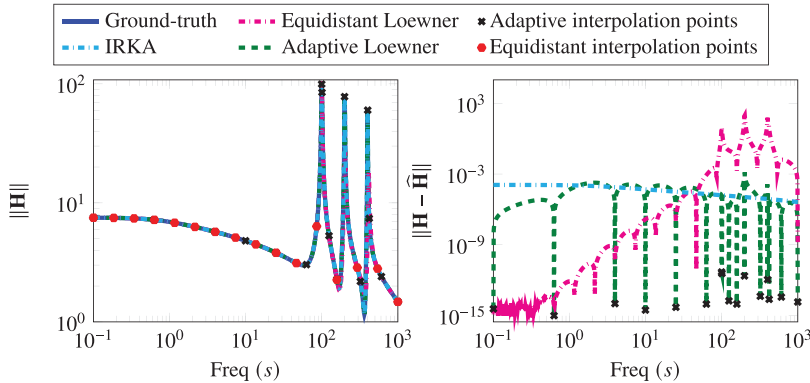


Figure 4. Penzl example: The Bode plots of the ground-truth, adaptively generated system, the realization with equidistant points, and IRKA realization are shown, and the corresponding errors with the ground truth are also presented.

the proposed method and the Loewner approach with equidistant points. The results are shown in Figure 4. Although the IRKA method chooses optimal points, the proposed greedy method to construct a realization compares well with IRKA. However, it is quite worthwhile to note that:

- (a) IRKA requires the order of the realization as an input which is usually not known.
- (b) IRKA chooses interpolation points over the whole complex plane but estimating transfer function at these points in an experiment is not straightforward, and sometimes not even possible.
- (c) Lastly, we highlight that IRKA is an iterative method; hence, it requires repeated evaluations of the transfer function at each iteration. For this example, IRKA took 14 iterations, thus $16 \times 14 = 224$ transfer evaluations. On the other hand, the proposed methodology requires only 16 transfer function evaluations, which are also only on the imaginary axis.

Finally, we study the robustness of Algorithm 1 with respect to the hyperparameter β . We report the quality of the learned realization with respect to the parameter β in Figure 5. We note that Algorithm 1 is quite robust to the parameter, and all learned realizations outperform the one obtained using the logarithmically equidistant measurement points.

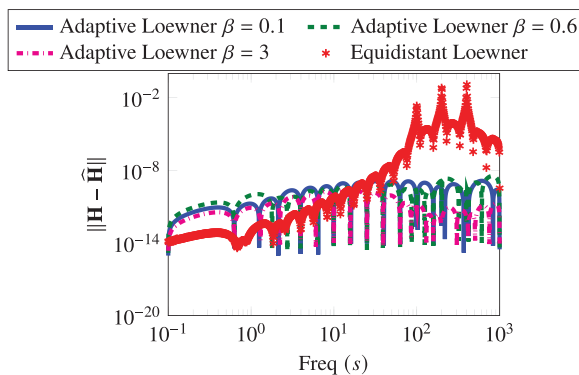


Figure 5. Penzl example: A comparison between the resulting Bode plots for different values of β used to defined the filter $g(\cdot)$ in (9).

Table 1. Penzl example: A comparison of the frequency-limited \mathcal{H}_2 -norm of the error between the ground-truth and realized systems under different levels of noise in the measurement data.

Method	Standard deviation of noise (σ)			
	0	10^{-6}	10^{-5}	10^{-4}
Adaptive	3.287×10^{-9}	1.544×10^{-7}	1.788×10^{-5}	5.886×10^{-4}
Equidistant	7.737×10^{-6}	2.400×10^{-3}	1.400×10^{-2}	3.259×10^{-1}

Measurement noise: Many studies have been made about the robustness of the Loewner framework to noise (see, e.g., Lefteriu et al., 2010; Kergus et al., 2018; Drmač and Peherstorfer, 2020). More generally, some studies have worked on designing identification methods that can better deal with noise (see, e.g., Schwerdtner, 2021). We do a preliminary study and observe the performance of our approach under noisy measurements. We conduct three experiments, where we corrupt the transfer function measurements with Gaussian white noise of standard deviation $\sigma = [10^{-4}, 10^{-5}, 10^{-6}]$, respectively. We compare our approach with a realization based on logarithmically equidistant interpolation points and measurements corrupted with the same noise level as for the adaptive case. The \mathcal{H}_2 -norms of learned and the ground-truth models are compared in Table 1. The table shows that even though the \mathcal{H}_2 error for the adaptive algorithm increases with the level of noise, it yields much better models as compared to the one obtained using the logarithmically equidistant points. This shows that the adaptive method is superior even in the presence of noise in the measurements. It is important to note that the Loewner framework was used to illustrate the application of our method. In the noisy data case, there could be better suitable methods where the proposed adaptive method can be used.

5.2. Beam example

In this example, we consider the beam model from the SLICOT model reduction library (Chahlaoui and Van Dooren, 2002). It comes from the discretization of a PDE (Antoulas et al., 2001). The input is the force applied at one end, and the output is the displacement resulting from the applied force at other end. We consider the frequency range of operation $[10^{-1}, 5]$. Algorithm 1 is applied, and a realization is obtained using 30 interpolation points as shown in Figure 6. As in the first example, we also construct a realization using 30 logarithmically equidistant interpolation points distributed over the chosen range of frequencies. Here again, we notice that most of the interpolation points using the adaptive method are chosen around the areas where most changes are happening. In contrast, measurements taken at logarithmically equidistant points yield a realization having a larger error. In Figure 7, we observe that adaptively chosen interpolation points decrease the \mathcal{H}_2 -norm error faster as compared to the equidistant ones.

RLC circuit

Finally, we consider an RLC circuit with 100 resistors, capacitors, and inductors (Gugercin and Antoulas, 2003; Benner et al., 2020). The frequencies are taken within the range of frequencies $[10^{-2}, 10^3]$. In order to obtain input/output time-domain measurements, a system simulation using the input (13) is performed to obtain the output in each step. Applying the procedure described in Section 4 directly realizes a model of order 12. As in the first example, we consider the realization of a model using 12 logarithmically equidistant interpolation points and their complex conjugates on the unit circle in order to ensure that the resulting system is real as discussed in Peherstorfer et al. (2017). A comparison of the Bode plots is shown in Figure 8. Our algorithm recovers the model with a relatively low error, and the model captures the time-domain response of the ground-truth model. It shows that our method provides a promising direction for the long-standing subject of input choice for time-domain system identification. In addition, since our

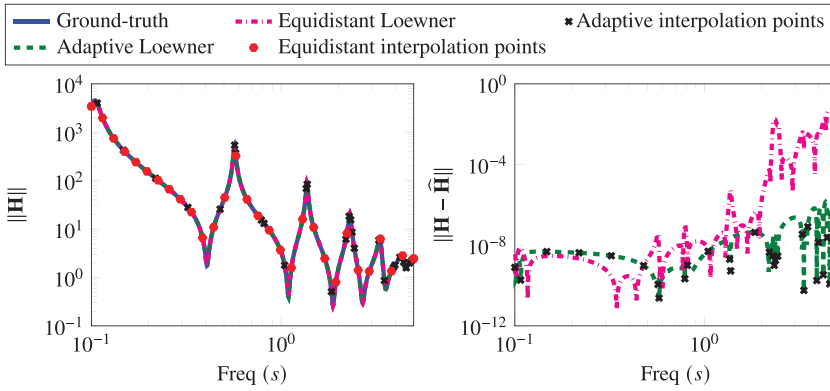


Figure 6. Beam Example: The Bode plot of the ground-truth system, adaptively generated system, and a realized system with equidistant points. The right figure shows the corresponding error between the ground-truth and realized systems.

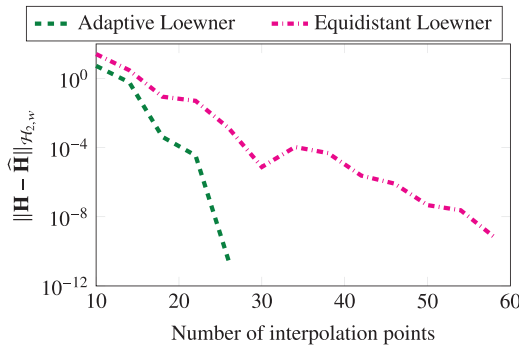


Figure 7. Beam Example: A comparison between the frequency-limited \mathcal{H}_2 norm error for the adaptively chosen points and the equidistant points.

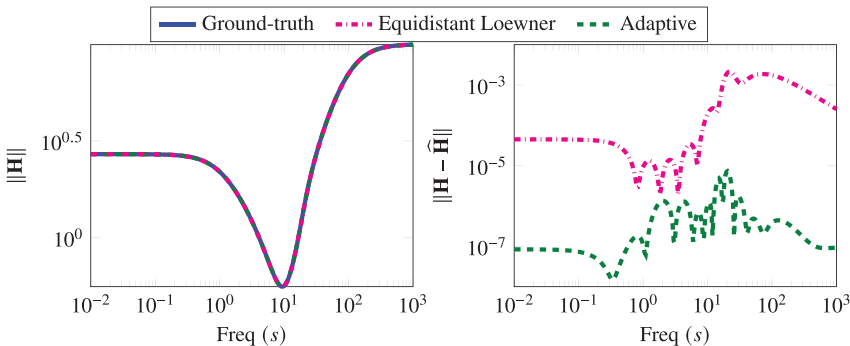


Figure 8. RLC example: A comparison for the Bode plot of the ground-truth and the identified models.

method infers only two transfer function measurements in each step, the size of the matrix F is small. This results in a more numerically stable optimization problem compared to the classical equidistant selection of points case where you have to solve for all the interpolation at once, leading to the matrix F of large size.

6. Conclusion

In this paper, we have presented a purely data-driven realization method that greedily chooses measurement points aiming at maximizing the extraction of the information about system dynamics by restricting measurements points on the imaginary axis. We have discussed both cases where data are taken in frequency and time domains. We have illustrated the efficiency of the proposed methodology by means of three benchmark examples. We have shown that our method performs better as compared to the approach where the measurement data are taken at logarithmically equidistant points in a given frequency range. We have observed that the proposed method can learn a good model without prior knowledge about the system dynamics by greedily choosing interpolation points, which is a key advantage.

Acknowledgment. An initial version of the paper is available on arXiv (arXiv:2107.12950) as a preprint (see Cherifi et al., 2021).

Data Availability Statement. The source code of the implementations, and the scripts to reproduce the presented results, are available on GitHub under the link https://github.com/mpimd-csc/Greedy_Measurement_Scheme.

Author Contributions. Conceptualization: K.C., P.G., and P.B.; Methodology: K.C., P.G., and P.B.; Software: K.C. and P.G.; Supervision: P.G. and P.B.; Writing—original draft: K.C. and P.G.; Writing—review & editing: K.C., P.G., and P.B. All authors approved the final submitted draft.

Funding Statement. This work received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing Interests. The authors declare no competing interests exist.

References

- Antoulas AC, Lefteriu S and Ionita A** (2017) A tutorial introduction to the Loewner framework for model reduction. In Benner P, Cohen A, Ohlberger M and Willcox K (eds), *Model Reduction and Approximation: Theory and Algorithms*. Philadelphia, PA: SIAM, pp. 335–376.
- Antoulas AC, Sorensen DC and Gugercin S** (2001) A survey of model reduction methods for large-scale systems. *Contemporary Mathematics* 280, 193–219.
- Beattie CA and Gugercin S** (2012) Realization-independent \mathcal{H}_2 -approximation. In *51st IEEE Conference on Decision and Control (CDC)*, Hawaii: IEEE pp. 4953–4958.
- Beddig RS, Benner P, Dorschky I, Reis T, Schwerdtner P, Voigt M and Werner SWR** (2022) Structure-preserving model reduction for dissipative mechanical systems. *Preprint*, arXiv:2010.06331.
- Benner P, Goyal P and Van Dooren P** (2020) Identification of port-Hamiltonian systems from frequency response data. *Systems and Control Letters* 143, 104741.
- Benner P, Köhler M and Saak J** (2021) Matrix equations, sparse solvers: M-M.E.S.S.-2.0.1—Philosophy, features and application for (parametric) model order reduction. In Benner P, Breiten T, Faßbender H, Hinze M, Stykel T and Zimmermann R (eds), *Model Reduction of Complex Dynamical Systems*, International Series of Numerical Mathematics, 171. Cham: Birkhäuser, pp. 369–392.
- Benner P and Werner SWR** (2021) Frequency- and time-limited balanced truncation for large-scale second-order systems. *Linear Algebra and its Applications* 623, 68–103. Special issue in honor of P. Van Dooren, Edited by F. Dopico, D. Kressner, N. Mastronardi, V. Mehrmann, and R. Vandebril.
- Brunton SL and Kutz JN** (2019) *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge: Cambridge University Press.
- Chahlaoui Y and Van Dooren P** (2002) A collection of benchmark examples for model reduction of linear time invariant dynamical systems. Technical Report 2002–2, SLICOT Working Note. Available at <http://www.slicot.org/>.
- Chellappa S, Feng L and Benner P** (2020) Adaptive basis construction and improved error estimation for parametric nonlinear dynamical systems. *International Journal for Numerical Methods in Engineering* 121, 5320–5349.
- Chellappa S, Feng L, de la Rubia V and Benner P** (2021) Adaptive interpolatory MOR by learning the error estimator in the parameter domain. In Benner P, Breiten T, Faßbender H, Hinze M, Stykel T and Zimmermann R (eds), *Model Reduction of Complex Dynamical Systems*, International Series of Numerical Mathematics, 171. Cham: Birkhäuser, pp. 97–117.
- Cherifi K, Goyal P and Benner P** (2021) A greedy data collection scheme for linear dynamical systems. *Preprint*, arXiv: 2107.12950. Available at <http://arxiv.org/abs/2107.12950>.
- Drmač Z, Gugercin S and Beattie C** (2015a) Quadrature-based vector fitting for discretized \mathcal{H}_2 approximation. *SIAM Journal on Scientific Computing* 37(2), A625–A652.

- Drmač Z, Gugercin S and Beattie C** (2015b) Vector fitting for matrix-valued rational approximation. *SIAM Journal on Scientific Computing* 37(5), A2346–A2379.
- Drmač Z and Peherstorfer B** (2020) Learning low-dimensional dynamical-system models from noisy frequency-response data with Loewner rational interpolation. *Preprint*, arXiv:1910.00110.
- Gosea IV** (2018) *Model Order Reduction of Linear and Nonlinear Systems in the Loewner Framework*. Ph.D. Thesis, Jacob University Bremen, Bremen.
- Gugercin S and Antoulas AC** (2003) A survey of balancing methods for model reduction. In *Proceedings of the European Control Conference (ECC 2003)*. Cambridge, UK: CD Rom.
- Gugercin S, Antoulas AC and Beattie C** (2008) \mathcal{H}_2 model reduction for large-scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications* 30(2), 609–638.
- Ionita A and Antoulas A** (2014a) Case study: Parametrized reduction using reduced-basis and the Loewner framework. In Quarteroni A and Rozza G (eds), *Reduced Order Methods for Modeling and Computational Reduction*, Vol. 9. Cham: Springer, pp. 51–66.
- Ionita AC and Antoulas AC** (2014b) Data-driven parametrized model reduction in the Loewner framework. *SIAM Journal on Scientific Computing* 36(3), A984–A1007.
- Karachalios DS, Gosea IV and Antoulas AC** (2020) The Loewner framework for system identification and reduction. In Benner P, Grivet-Talocia S, Quarteroni A, Rozza G, Schilders WHA and Silveira LM (eds), *Model Reduction Handbook: Volume I: System- and Data-Driven Methods and Algorithms*, Berlin: Handbook on Model Reduction, 1. De Gruyter.
- Kergus P, Formentin S, Poussot-Vassal C and Demourant F** (2018) Data-driven control design in the Loewner framework: Dealing with stability and noise. In *European Control Conference (ECC)*. Limassol, Cyprus: IEEE, pp. 1704–1709.
- Lefteriu S, Ionita AC and Antoulas AC** (2010) Modeling systems based on noisy frequency and time domain measurements. In Willems JC, Hara S, Ohta Y and Fujioka H (eds), *Perspectives in Mathematical System Theory, Control, and Signal Processing*, Lecture Notes in Control and Information Sciences, Vol. 398. Berlin–Heidelberg: Springer.
- Mayo AJ and Antoulas AC** (2007) A framework for the solution of the generalized realization problem. *Linear Algebra and its Applications* 425(2–3), 634–662.
- Nakatsukasa Y, Sète O and Trefethen LN** (2018) The AAA algorithm for rational approximation. *SIAM Journal on Scientific Computing* 40, A1494–A1522.
- Ogata K** (2010) *Modern Control Engineering*. NJ: Prentice Hall.
- Paul-Dubois-Taine A and Amsalle D** (2015) An adaptive and efficient greedy procedure for the optimal training of parametric reduced-order models. *International Journal for Numerical Methods in Engineering* 102(5), 1262–1292.
- Peherstorfer B, Gugercin S and Willcox K** (2017) Data-driven reduced model construction with time-domain Loewner models. *SIAM Journal on Scientific Computing* 39, A2152–A2178.
- Peherstorfer B and Willcox K** (2016) Data-driven operator inference for non-intrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering* 306, 196–215.
- Penzl T** (2006) Algorithms for model reduction of large dynamical systems. *Linear Algebra and its Applications* 415, 322–343.
- Poussot-Vassal C, Kergus P, Kerhervé F, Sipp D, and Cordier L** (2021). Interpolatory-based data-driven pulsed fluidic actuator control design and experimental validation. *IEEE Transactions on Control Systems Technology*.
- Schwerdtner P** (2021) Port-Hamiltonian system identification from noisy frequency response data. *Preprint*, arXiv:2106.11355.
- Sokolov VI** (2006) *Contributions to the Minimal Realization Problem for Descriptor Systems*. Ph.D. Thesis, Technische Universität Chemnitz, Chemnitz.
- Wilson DA** (1970) Optimum solution of model-reduction problem. *Proceedings of the Institution of Electrical Engineers* 117(6), 1161–1165.
- Yu J, Yan C and Guo M** (2019) Non-intrusive reduced-order modeling for fluid problems: A brief review. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* 233(16), 5896–5912.